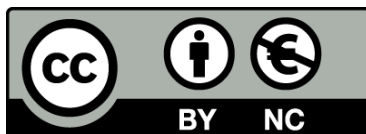




Desenvolupament i aplicació de metodologies analítiques i quimiomètriques per a l'estudi de la contaminació atmosfèrica

Marta Alier Pedemonte



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 3.0. Spain License.**

DESENVOLUPAMENT I APLICACIÓ DE METODOLOGIES ANALÍTIQUES I QUIMIOMÈTRIQUES PER A L'ESTUDI DE LA CONTAMINACIÓ ATMOSFÈRICA

Marta Alier Pedemonte



Programa de doctorat:
Química Analítica del medi ambient i la pol·lució

**DESENVOLUPAMENT I APLICACIÓ DE
METODOLOGIES ANALÍTIQUES I QUIMIOMÈTRIQUES
PER A L'ESTUDI DE LA CONTAMINACIÓ
ATMOSFÈRICA**



Marta Alier Pedemonte

Director:

Dr. Romà Tauler Ferré (Institut de Diagnòstic Ambiental i Estudis de l'Aigua -
Consell Superior d'Investigacions Científiques, IDAEA -CSIC)

Tutora:

Dra. Anna de Juan Capdevila (Universitat de Barcelona)

El Dr. Romà Tauler Ferré, professor d'Investigació del Departament de Química Ambiental de l'Institut de Diagnòstic Ambiental i Estudis de l'Aigua (IDAEA) adscrit al Consell Superior d'Investigacions Científiques (CSIC)

FA CONSTAR:

que la present memòria titulada “DESENVOLUPAMENT I APLICACIÓ DE METODOLOGIES ANALÍTIQUES I QUIMIOMÈTRIQVES PER A L'ESTUDI DE LA CONTAMINACIÓ ATMOSFÈRICA” ha estat realitzada sota la seva direcció per la Sra. Marta Alier Pedemonte en el Departament de Química Ambiental de l'Institut de Diagnòstic Ambiental i Estudis de l'Aigua (IDAEA) adscrit al Consell Superior d'Investigacions Científiques (CSIC).

Barcelona, Març 2014

Dr. Romà Tauler Ferré

Professor d'Investigació del Departament de Química Ambiental de l'Institut de Diagnòstic Ambiental i Estudis de l'Aigua (IDAEA) adscrit al Consell Superior d'Investigacions Científiques (CSIC)

A la meva família

*“Tots som molt ignorants. El que ocorre és
que no tots ignorem les mateixes coses.”*

Albert Einstein

AGRAÏMENTS

Aquesta Tesi no hagués estat possible sense l'ajuda de diverses persones, i per tant, vull agrair

- Al meu director, el Dr. Romà Tauler, per haver-me donat la oportunitat de realitzar aquesta Tesi sota la seva direcció. Hem passat moments bons i dolents, però finalment, hem aconseguit arribar al final. Gràcies per tot el que m'has ensenyat i la paciència que has tingut durant aquests anys.
- A la Dra. Anna de Juan per haver acceptat ser la meva tutora durant aquesta Tesi, i per l'ajuda que m'ha donat durant la mateixa, però també molt abans quan vaig començar a saber què era la quimiometria en la carrera i per què servia.
- A tots els companys del grup de quimiometria del IDAEA (tant els que portem uns anys com els que us heu incorporat darrerament), els del grup de quimiometria de la UB, i a tots els que han vingut a fer estades, per fer un molt bon ambient de treball i sempre tenir algú del grup que et pot ajudar i explicar coses.
- Al Dr. Barend L. Van Drooge, per tota l'ajuda que m'ha donat en l'anàlisi dels orgànics. Barend, gràcies per la paciència al laboratori en explicar-me cada tècnica. Gràcies per contestar sempre les meves preguntes amb bona cara i explicar-me les coses tan clarament. Treballar amb tu al laboratori i amb l'ordinador ha estat molt fàcil i he après molt.
- Al Dr. Joan O. Grimalt per deixar-me utilitzar el seu laboratori, i a la gent que he conegut en ell, com la Marta, el Barend, l'Anuar,... per deixar-me un tros del vostre espai i ajudar-me en tot el que he

necessitat. Gràcies Marta, pels bons moments que hem passat juntes, i pels que ens queden...

- A la Generalitat de Catalunya, per proporcionar-nos les bases de dades utilitzades en la primera part d'aquesta tesi, i als integrants dels projectes SAPUSS i CalNex per haver-nos deixat una part de les seves mostres, i poder realitzar la segona part d'aquesta Tesi intercanviant idees i coneixements.
- Al Ministerio de Economía y Competitividad d'Espanya per concedir els projectes CTQ2009-11572 i CTQ2012-38616-C02-01 i al CSIC pel contracte 201080E033 per realitzar aquesta Tesi.
- Finalment, agrair a la meva família: els meus pares, el Jose i família, les meves germanes, marits i nens, i a tots els meus amics i amigues... tots els dies que hem compartit i que han fet que aquests anys de Tesi hagin passat en molt bona companyia.

ÍNDEX

Resum	11
1. Objectius i estructura de la Tesi	13
1.1. Objectius	15
1.2. Estructura de la Tesi	17
1.3. Relació dels treballs científics presentats a la memòria	19
2. Introducció	21
2.1 Contaminació atmosfèrica	25
2.2 Quimiometria	32
2.3 Contaminants ambientals estudiats en aquesta Tesi	37
2.3.1 Unitats de concentració	53
2.3.2 Legislació	54
3. Part experimental	59
3.1 Contaminants de les bases de dades de la Generalitat	61
3.1.1 Estructura de les dades ambientals	63
3.2 Contaminants orgànics en filtres de material particulat analitzats al laboratori	65
3.2.1 Presa de mostra de l'aerosol	65
3.2.2 Tractament de mostra	68
3.2.3 Anàlisi per GC-MS	69
3.2.4 Estructura de les dades cromatogràfiques	71
4. Tractament de les dades experimentals	73
4.1 Estructura de les dades	75
4.2 Tractaments preliminars	79
4.3 Mètodes quimiomètrics	85

4.3.1 Anàlisi de Components Principals (PCA)	87
4.3.2 Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS)	89
4.3.3 PARAFAC	107
4.3.4 TUCKER3	110
5. Resultats i discussió	113
5.1 Bloc 1 - Contaminants de les bases de dades de la Generalitat	115
5.1.1 Discussió de resultats	165
a) Aplicació dels mètodes quimiomètrics PCA i MCR-ALS a dades ambientals de grans dimensions i posterior interpretació dels perfils resolts en base a les seves variacions geogràfiques i temporals	165
b) Aplicació de les restriccions de trilinearitat i interacció entre components en el mètode MCR-ALS a dades ambientals de diferent complexitat, i comparació amb mètodes equivalents	168
c) Aplicació del mètode MCR-ALS a dades incompletes de contaminació per NO i O ₃	176
5.2 Bloc 2 - Contaminants orgànics en mostres de material particulat	181
5.2.1 Discussió de resultats	219
a) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Barcelona	219
b) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Los Angeles i comparació amb els resultats obtinguts a Barcelona	226
6. Conclusions	233
7. Referències	239

RESUM

En aquesta Tesi s'han estudiat les fonts d'emissió de diversos contaminants atmosfèrics, principalment a l'àrea de Barcelona. Els contaminants estudiats han estat l'òxid de nitrogen (NO), compost tòxic que s'emet en grans quantitats pels vehicles a causa de la combustió, l'ozó, que té naturalesa oxidant, i es troba molt relacionat amb la presència d'òxids de nitrogen, i diversos contaminants orgànics presents en el material particulat atmosfèric. Entre els contaminants orgànics estudiats hi ha alguns de més tòxics com són els hidrocarburs aromàtics policíclics (PAH), i d'altres presents al medi, però amb menor impacte sobre la salut, com són els àcids dicarboxílics o els productes derivats de l'oxidació de l'isoprè o l' α -pinene.

Les taules de dades corresponents a les concentracions de NO i O₃ s'han obtingut a partir de les bases de dades del departament de medi ambient de la Generalitat de Catalunya. Els contaminants orgànics s'han determinat en mostres de material particulat obtingudes de campanyes de seguiment ambiental a les ciutats de Barcelona i Los Angeles. Aquests compostos han estat determinats experimentalment al laboratori, mitjançant la seva extracció i anàlisi amb cromatografia de gasos acoblada a espectrometria de masses.

L'ús de la quimiometria ha permès interpretar i simplificar la gran quantitat de dades obtingudes en els estudis de seguiment ambiental realitzats en aquesta Tesi. Depenent de la informació que s'ha volgut obtenir en cada estudi, les taules de dades s'han disposat en matrius de dades augmentades en dos direccions (*two-way*) i en alguns casos també en estructures en tres direccions (*three-way*). S'han aplicat diferents mètodes quimiomètrics, entre els que ha destacat el mètode de resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS). L'aplicació d'aquest mètode amb diferents tipus de restriccions com la no-negativitat i la trilinearitat ha permès estimar les fonts d'emissió i de distribució geogràfica i temporal dels contaminants en les regions estudiades.

S'ha descrit amb detall els perfils de variació temporal (diària i anual) de les fonts d'emissió de NO i O₃ a la ciutat de Barcelona i a diferents regions de Catalunya. S'ha trobat que les principals fonts d'emissió de contaminants atmosfèrics són antropogèniques, i s'ha confirmat que el tràfic de vehicles és un dels principals emissors dels contaminants estudiats. Altres fonts de contaminació de compostos orgànics trobades en material particulat han estat la crema de biomassa, els productes d'oxidació de l'isoprè i l' α -pinene, i els àcids dicarboxílics, que poden tenir tant orígens biogènics com antropogènics.

Capítol 1

OBJECTIUS I ESTRUCTURA DE LA TESI

1.1. Objectius

L'objectiu general d'aquesta Tesi ha estat el desenvolupament i l'aplicació de metodologies analítiques i quimiomètriques per a l'estudi de la contaminació atmosfèrica, principalment a la ciutat de Barcelona. La interpretació de la gran quantitat de dades obtingudes en l'anàlisi dels contaminants atmosfèrics ha fet necessària l'aplicació de mètodes quimiomètrics.

Es descriuen a continuació els diferents objectius específics d'aquesta Tesi, separats per objectius de tipus quimiomètric i de tipus ambiental.

Objectius de tipus quimiomètric

- Desenvolupament i aplicació del mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS) a l'anàlisi de dades de contaminació atmosfèrica i a l'estudi de la distribució de fonts de contaminació (*source apportionment*)
- Adaptació del mètode MCR-ALS i d'altres mètodes quimiomètrics a l'anàlisi simultània de múltiples taules o matrius de dades disposades en dues o més direccions (*two-way, three-way, multi-way data*).
- Aplicació de les restriccions de trilinearitat i d'interacció entre components en el mètode MCR-ALS en estudis de contaminació atmosfèrica. Interpretació dels resultats obtinguts a partir d'aquestes restriccions i comparació d'aquests resultats amb els obtinguts amb altres mètodes relacionats (PARAFAC i TUCKER3).
- Desenvolupament d'una nova versió del mètode MCR-ALS que permet treballar amb conjunts de dades incompletes, on alguns dels blocs de dades es troben absents.

Objectius de tipus ambiental

- Estudi i interpretació de les variacions geogràfiques i temporals de la contaminació atmosfèrica de l'aire per òxid de nitrogen i ozó a Catalunya.
- Desenvolupament de mètodes analítics d'identificació i de determinació experimental de la concentració de contaminants orgànics en mostres de material particulat a Barcelona i Los Angeles. Desenvolupament dels procediments corresponents d'extracció, derivatització i posterior anàlisi amb cromatografia de gasos acoblada a espectrometria de masses.
- Comparació de les fonts de contaminació atmosfèrica en regions distants geogràficament, amb característiques climàtiques semblants com són Barcelona i Los Angeles.

1.2. Estructura de la Tesi

La present memòria es presenta en forma d'una part introductòria, una part experimental, una part sobre el tractament quimiomètric de les dades, una part de discussió dels resultats obtinguts, on s'inclouen els articles científics publicats, i per últim, les referències emprades en les diferents parts. La present Tesi consta de set capítols que es descriuen a continuació.

- 1) En el primer capítol es presenten els objectius i es detalla l'estructura i la relació dels treballs científics de la present memòria.
- 2) En el segon capítol es fa una introducció sobre la contaminació atmosfèrica, la quimiometria, i es descriuen els compostos estudiats. Entre els compostos estudiats es troben l'òxid de nitrogen, l'ozó, i compostos orgànics com els hidrocarburs aromàtics policíclics (PAH) i els àcids dicarboxílics entre d'altres. També es detalla la legislació que afecta a aquests contaminants.
- 3) En el tercer capítol es recullen els procediments experimentals emprats per a l'estudi dels contaminants estudiats en aquesta Tesi detallant la presa de mostra, el tractament de la mostra i el seu anàlisi.
- 4) En el quart capítol es descriuen els mètodes quimiomètrics utilitzats en la present Tesi per al tractament de les dades, incloent les diferents estructures en que es poden disposar les taules de dades, així com els pretractaments més usuals de les mateixes.
- 5) En el cinquè capítol es descriuen els resultats obtinguts en els treballs realitzats en aquesta Tesi, en dos blocs que agrupen aquests treballs per temàtiques. En el primer bloc es mostren els resultats dels estudis dels contaminants en mostres d'aire obtinguts a partir de les xarxes de seguiment de la contaminació atmosfèrica de la Generalitat de Catalunya. En el segon bloc es mostren els resultats dels estudis realitzats sobre contaminants orgànics analitzats en mostres de material particulat.

- 6) En el sisè capítol es recullen les conclusions generals més importants de la present Tesi.
- 7) Finalment, en el setè i últim capítol es recullen totes les referències emprades en aquesta Tesi.

1.3. Relació dels treballs científics presentats en aquesta Tesi doctoral

Article 1. M. Alier, M. Felipe-Sotelo, I. Hernández, R. Tauler; *Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods*. *Analytica Chimica Acta* 642 (2009) 77–88

Article 2. Marta Alier, Mónica Felipe-Sotelo, Isabel Hernández, Romà Tauler; *Trilinearity and component interaction constraints in the Multivariate Curve Resolution investigation of NO and O₃ pollution in Barcelona*. *Analytical Bioanalytical Chemistry* (2011) 399:2015–2029

Article 3. Marta Alier, Romà Tauler; *Multivariate Curve Resolution of incomplete data multisets*. *Chemometrics and Intelligent Laboratory Systems* 127 (2013) 17–28

Article 4. Marta Alier, Barend L. van Drooge, Manuel Dall'Osto, Xavier Querol, Joan O. Grimalt, Romà Tauler; *Source apportionment of submicron organic aerosol at an urban background and road site in Barcelona (Spain) during SAPUSS*. *Atmospheric Chemistry and Physics* 13 (2013) 10353-10371.

Article 5. M. Alier, Y.-H. Lin, J.D. Surratt, M. Dall'Osto, R. Tauler, J.O. Grimalt, B.L. van Drooge; *On the origin of water-soluble organic tracer compounds in fine aerosols in two cities: the case of Los Angeles and Barcelona*. *Environmental Science and Pollution Research* (2014) DOI: 10.1007/s11356-013-2460-9.

Capítol 2

INTRODUCCIÓ

La gran quantitat de contaminants produïts principalment per l'home, ha fet que l'interès per la contaminació atmosfèrica hagi augmentat força en els últims anys. Aquests contaminants van començar a aparèixer quan l'home va trencar l'equilibri ecològic en el seu desenvolupament. Els canvis produïts per aquest desenvolupament van millorar de forma notable la qualitat de vida de l'home, però van crear situacions noves, com l'efecte hivernacle, la destrucció de la capa de l'ozó o la creació de noves substàncies que han introduït noves fonts de contaminació, i han modificat el medi, tant aquàtic, terrestre com atmosfèric. La quantitat de contaminants presents en el medi ambient ha anat creixent exponencialment i ha arribat a totes les regions del planeta, ja que la contaminació es dispersa per tots els compartiments ambientals. Amb el temps s'ha començat a prendre consciència de que aquests canvis afecten a la salut del ecosistema i a la nostra pròpia, de manera que s'han començat a aplicar mesures de control d'aquests contaminants.

En els anys 70 es va crear un dels organismes més importants en quant a la protecció ambiental, com és la EPA (*Environmental Protection Agency*) i va ser també quan es va crear una llei federal, la US CAA (*US Clean Air Act*) que regulava les emissions en l'aire de fonts estacionàries i mòbils. Aquesta llei va autoritzar a la EPA a establir uns estàndards de qualitat de l'aire NAAQS (*National Ambient Air Quality Standards*) per protegir la salut i el benestar públic (incloent danys a animals, la vegetació, els camps de cultiu i els edificis) i regular les emissions de contaminants perillosos a l'aire. Aquesta llista constava de sis contaminants (monòxid de carboni, plom, òxids de nitrogen, ozó, òxids de sofre i material particulat) que es van ampliar amb els anys, i els nivells acceptats dels quals s'han anat reduint amb el temps. En la Unió Europea, s'han regulat els contaminants atmosfèrics amb la creació de la *EU Thematic Strategy on Air Pollution* amb nivells encara més restrictius que els EEUU, i al 1979 es va firmar el Conveni de Ginebra sobre la contaminació atmosfèrica transfronterera. Aquest conveni va establir un marc de col·laboració intergovernamental per protegir la salut i el medi ambient contra la contaminació atmosfèrica que pot afectar a diferents països, amb l'elaboració de polítiques adequades, intercanvi d'informació, realització d'activitats

d'investigació i vigilància. Al 1994 es va crear l'Agència Europea de Medi Ambient (AEMA), la qual confecciona conjunts de dades europees a partir de la informació recopilada en els estats membres i elabora indicadors i informes sobre l'estat del medi ambient.

La contaminació atmosfèrica és un problema cada vegada més preocupant. A nivell mundial, la Organització Mundial de la Salut (OMS, ó WHO en les seves sigles en anglès, *World Health Organization*) ha estimat que dos milions de persones, la meitat dels quals són de països desenvolupats, moren cada any degut a la pol·lució de l'aire. Entre aquesta pol·lució, el principal factor de risc és la matèria particulada (PM), seguida de l'ozó. La OMS, recomana reduir l'emissió dels gasos que produeixen canvi climàtic, i els límits recomanats per l'ozó i el diòxid de sofre. També, en la revisió de les guies de la qualitat de l'aire (*Air Quality Guidelines, AQGs*), es creu que si es redueixen els nivells d'un particular tipus de contaminant com és el PM₁₀ (material particulat de menys de 10µm), es podrien reduir el nombre de morts deguts a infeccions respiratòries, enfermetats del cor, o càncer de pulmó, en un 15% en les ciutats més contaminades. En la unió Europea, s'ha estimat que la matèria particulada més petita (PM_{2.5}) causa una disminució estadística de 8.6 mesos de l'esperança de vida. La OMS està treballant amb tots els països per millorar aquesta situació i arribar als estàndards de qualitat de l'aire. Hi ha molts països que no tenen regulacions en la pol·lució de l'aire (ciutats asiàtiques i llatinoamericanes), ni tenen els nivells de pol·lució monitoritzats, i aquesta situació s'ha d'anar canviant amb el temps, encara que els problemes de salut no es veuen només en les ciutats més contaminades, sinó que també en ciutats relativament netes (Europa i Nord-Amèrica).

La contaminació atmosfèrica és un tema complex i en el següent apartat es descriuran diferents conceptes relatius a aquest tema. En aquesta Tesi s'han fet estudis regionals de contaminants atmosfèrics com l'òxid de nitrogen i l'ozó, i amb contaminants orgànics (hidrocarburs aromàtics policíclics (PAH), àcids dicarboxílics, derivats de l'isoprè, levoglucosan, etc.) els quals tenen diferents nivells de complexitat i també es descriuran a continuació. L'estudi

d'aquests contaminants s'ha fet amb mètodes quimiomètrics, pel que també es farà una introducció a aquesta disciplina.

2.1. Contaminació atmosfèrica

La contaminació atmosfèrica és la presència en l'aire de matèries gasoses, líquides o sòlides o formes d'energia que impliquin risc per les persones, animals i naturalesa. Aquests contaminants es dispersen i transporten per tota l'atmosfera degut als diferents fenòmens meteorològics. Per que una substància sigui considerada com a contaminant, no és necessari que la seva identitat sigui diferent de la de qualsevol dels compostos naturals de l'aire, és suficient amb que la seva proporció no sigui la natural (Orozco, 2003). Habitualment, el terme contaminació atmosfèrica acostuma a utilitzar-se per fer referència a la ocasionada per fonts de naturalesa antropogènica, però les alteracions de la composició de l'aire poden tenir un doble origen:

- *Natural*: erupcions volcàniques, meteorits, focs, incendis...
- *Artificial o antropogènica*: provocada per activitats humanes (bàsicament les combustions de hidrocarburs fòssils en el transport, i els processos industrials).

Una vegada emesos els contaminants a la atmosfera poden sofrir transformacions químiques que alterin la seva naturalesa. Per això, quan es classifiquen els contaminants atmosfèrics es distingeixen dos tipus, primaris i secundaris (Spiro, 2004; Baird, 2001). La denominació de contaminants primaris s'aplica a aquells agents que s'emeten de forma directa a l'atmosfera, mentre que, sota el terme de secundaris, s'engloben aquelles espècies que no s'emeten com a tal a l'atmosfera, sinó que es formen en ella per interaccions entre contaminants primaris, components naturals de l'atmosfera i radiacions incidents en la mateixa. Hi ha diferents grups de contaminants atmosfèrics entre els que hi ha els òxids de carboni, els òxids de sofre, els òxids de nitrogen, els compostos orgànics volàtils, el material particulat i els aerosols, els compostos oxidants i altres.

A continuació es farà una descripció dels diferents grups de compostos fent un especial èmfasi en els que han estat tractats en aquesta Tesi, com són els òxids de nitrogen, l'ozó, els compostos orgànics volàtils, les partícules i aerosols.

Òxids de nitrogen

Els òxids de nitrogen inclouen l'òxid nítric (NO), el diòxid de nitrogen (NO_2) i l'òxid nítrós (N_2O). L'òxid nítric i el diòxid de nitrogen s'acostumen a considerar de forma conjunta amb la denominació NO_x . Són contaminants primaris tòxics i amb molta transcendència en problemes de contaminació. L'òxid nítrós és inert a la troposfera però contribueix a l'efecte hivernacle. Els NO_x tenen fonamentalment un origen antropogènic, on els majors emissors d'aquests contaminants són el transport i altres tipus de processos on tenen lloc reaccions de combustió a elevades temperatures. El que s'emet amb més quantitat és el NO, que s'oxida ràpidament a NO_2 per acció de l'ozó. El diòxid de nitrogen és, en presència d'hidrocarburs i llum ultraviolada, la principal font d'ozó troposfèric i d'aerosols de nitrats, que constitueixen una fracció important de la massa del $\text{PM}_{2.5}$ de l'aire ambient. Habitualment es mesuren en l'aire ambient les concentracions de NO_2 però no les d'altres contaminants derivats de la combustió, pel que és adequat mantenir un límit prudent de NO_2 . També la contaminació de NO_x emesa pels vehicles és un tema d'interès, ja que els cotxes dièsel són els que més NO_x emeten. Per exemple, en la nova normativa Euro 6 que entrarà en vigor al setembre de 2014 s'han establert una reducció dels nivells de NO_x pels vehicles dièsel de passatgers, lleugers i semilleugers, que passen dels 180 mg/km acceptats en l'actual Euro 5 a 80 mg/km en la futura Euro 6. Aquesta normativa serà especialment determinant en països com Espanya, on per exemple a Barcelona, els cotxes dièsel són el 75% del total de vehicles (Reche, 2012).

En aquesta Tesi s'ha estudiat la variació geogràfica i temporal de l'òxid de nitrogen en una determinada zona geogràfica (Catalunya) durant diversos anys. Els resultats més importants d'aquest estudi s'explicaran en la secció de resultats.

Compostos orgànics volàtils

Aquest grup inclou diferents compostos volàtils d'origen natural o antropogènic com el metà (CH_4), els hidrocarburs o l'isoprè i els seus derivats. El metà és un contaminant primari, que és molt abundant i que es forma de manera natural en diverses reaccions anaeròbiques del metabolisme. El metà no produeix problemes en la salut ni en els éssers vius, però influeix de manera significativa en l'efecte hivernacle i en les reaccions estratosfèriques. En l'atmosfera hi ha presents molts altres hidrocarburs, procedents de fenòmens naturals i originats per activitats humanes, sobretot relacionades amb l'extracció, refinatge i l'ús del petroli i els seus derivats. Els seus efectes sobre la salut són variables. Alguns no presenten problemes però altres són tòxics, poden afectar al sistema respiratori, i poden produir càncer. Entre aquests es troben els hidrocarburs aromàtics policíclics o PAH, alguns dels quals s'han estudiat en aquesta Tesi i dels que en parlarem amb més detall més endavant. Els VOCs s'han estudiat àmpliament, sobretot en la última dècada, i s'estudien tant en ambients tancats, controlant per exemple els nivells d'exposició durant diversos anys (Matysik, 2013), com a l'atmosfera, analitzant per exemple les diferents fonts d'emissió dels mateixos (McCarthy, 2013).

Partícules i aerosols

La paraula aerosol s'utilitza per referir-se a una barreja heterogènia de partícules sòlides o líquides suspeses en un gas. El terme aerosol es refereix tant a les partícules sòlides com al gas, i la seva notació com a material particulat (PM) s'utilitza per referir-se exclusivament a les partícules. La presència de les partícules en la atmosfera, així com la seva posterior deposició, pot generar efectes en el clima, en els ecosistemes i en els éssers vius (Dockery, 1996; Arimoto, 2001). El terme PM_{10} es refereix a partícules que tenen un diàmetre aerodinàmic de $10 \mu\text{m}$ o menys (partícules grosses), el terme $\text{PM}_{2.5}$ és per les partícules amb un diàmetre de $2.5 \mu\text{m}$ o menor, el terme PM_1 si les partícules són de diàmetre $1 \mu\text{m}$ o menor (partícules fines), i així successivament s'utilitzen altres termes amb altres valors numèrics (veure Figura 2.1). Nombrosos estudis epidemiològics afirmen que existeix una correlació significativa entre la exposició al material particulat atmosfèric i

diversos efectes adversos per la salut (Künzli, 2000; Brunekreef, 1997). Els efectes que pot provocar la inhalació de material particulat depenen de diferents factors, però un dels més importants és el de la mida de la partícula. Les partícules més perilloses són les de diàmetre $<2.5 \mu\text{m}$, perquè aquestes poden arribar a la cavitat alveolar i provocar greus afeccions a la salut. El material particulat atmosfèric també pot afectar els ecosistemes quan es diposita sobre la superfície terrestre i la regulació del clima, ja que les partícules en suspensió actuen com a nuclis de condensació en la formació de núvols (Mészáros, 1999) i les partícules afecten també a la visibilitat atmosfèrica (WHO, 2002). Al principi es van legislar només els nivells permesos de PM_{10} ($40 \mu\text{g}/\text{m}^3$ en l'actualitat com a límit de promig anual), però com es va veure que les partícules fines eren encara més perilloses, en els últims anys s'ha legislat també els límits permesos de $\text{PM}_{2.5}$ ($25 \mu\text{g}/\text{m}^3$ en el límit com a promig anual).

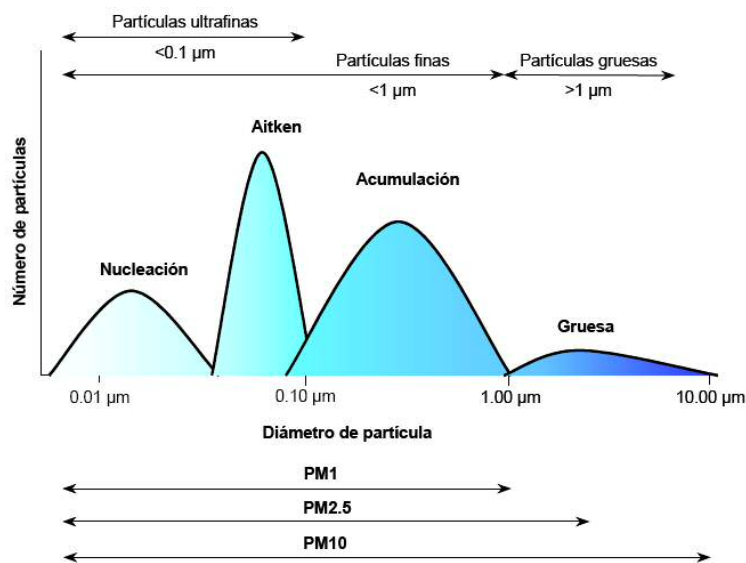


Figura 2.1. Distribució del nombre de partícules en funció del diàmetre

Els aerosols atmosfèrics estan formats per compostos inorgànics solubles en aigua (com el sulfat i el nitrat), i per una part de pols mineral insoluble en aigua i de compostos de carboni. Aquests compostos de carboni poden estar, a la seva vegada, constituïts per compostos orgànics solubles i insolubles, i per carboni elemental. L'aerosol orgànic (OA) pot constituir entre un 20-90% de l'aerosol fi (material particulat amb diàmetre inferior de $2.5 \mu\text{m}$)

(Kanakidou, 2005; Zhang, 2007). Els aerosols orgànics poden ser emesos directament a la atmosfera com a emissions primàries de fonts antropogèniques (activitats humanes) i de fonts naturals. Hi ha moltes fonts d'emissió que poden contribuir als aerosols orgànics primaris, com poden ser la crema de biomassa (De Gouw, 2009) i dels biofuels (Kanakidou, 2005), el cuinat d'aliments (Rogge, 1991), la pols del carrer, o els residus de les rodes dels vehicles, frens, etc. (Rogge, 1993a). A més, les emissions directes de plantes i els residus d'agricultura també poden produir aerosols orgànics primaris (Rogge, 1998). Tradicionalment, s'havia assumit que l'aerosol orgànic primari era una petita fracció de l'aerosol orgànic, especialment en àrees urbanes, no obstant, recentment s'ha suggerit que els aerosols orgànics primaris poden constituir una fracció més gran del que es suposava fins ara (Donahue, 2009).

Els aerosols orgànics també es poden transformar a l'atmosfera donant lloc als aerosols secundaris a partir de processos físico-químics que involucren altres aerosols, gasos o partícules atmosfèriques amb precursors d'emissions naturals o antropogèniques (Edney, 2005; Sullivan, 2006a; Warneck, 1988). Aquests aerosols orgànics secundaris (SOA) tenen propietats físiques i químiques diferents de les seves partícules primàries precursoras (Donahue, 2009) i es poden modificar degut a la presència de diferents oxidants atmosfèrics. Els aerosols orgànics secundaris consisteixen en una barreja d'espècies orgàniques oxigenades que depenen del grau de processament de l'aerosol de l'atmosfera. Els seus mecanismes de formació i evolució encara no estan del tot clars (Hallquist, 2009). Aquests aerosols poden provenir de productes d'oxidació de diferents compostos orgànics volàtils antropogènics o biogènics (Volkamer, 2006; Claeys, 2004a; Hamilton, 2009; Hallquist, 2009). Es considera que els hidrocarburs, alquens i els monoterpens constitueixen la major font global de l'aerosol orgànic secundari (Engelhart, 2008) i els hidrocarburs aromàtics la major font de l'aerosol orgànic urbà. El isoprè (2-metil-1,3-butadiè, C_5H_8), és un compost orgànic volàtil biogènic molt abundant, que s'ha estudiat extensament com a precursor per a la formació d'aerosols orgànics secundaris (Henze, 2006). Un dels camins de producció d'aerosols orgànics secundaris és l'oxidació de compostos orgànics volàtils biogènics o

antropogènics amb compostos com el ozó, el radical hidroxil ($\text{OH}\cdot$) i el radical nitrat ($\text{NO}_3\cdot$) (Kroll, 2008). També, els compostos orgànics volàtils gasosos es poden absorbir en les gotes dels núvols i experimentar la seva oxidació química en fase aquosa (Kanakidou, 2005). Els compostos altament volàtils queden exclusivament a la fase gas, mentre que els productes no volàtils s'acumulen en la fase particulada. Els productes de volatilitat intermèdia, anomenats compostos semi-volàtils, estableixen un equilibri de partició entre la fase gas i la fase particulada. Aquest equilibri depèn de múltiples factors, entre els quals s'inclou la pressió de vapor de saturació del compost, la temperatura ambient, i la quantitat de massa orgànica-aquosa d'aerosol on es pot produir la partició (Pankow, 1994).

L'aerosol orgànic secundari està format per compostos orgànics que tenen un ampli ventall de grups funcionals, on s'inclouen els àcids mono i dicarboxílics, els compostos carbonílics i els alcohols. Aquests compostos són tant hidrofòbics com hidrofílics (Sullivan, 2006b). El interval de pesos moleculars en els aerosols orgànics secundaris és molt ample, depenent de la localització, de l'estació de l'any o de la font de l'aerosol entre d'altres. Una vegada formats, els aerosols orgànics secundaris s'envelleixen per oxidació, tant amb O_3 com amb radicals OH (Rudich, 2007). Amb aquest envelliment decreix la seva volatilitat, i també es pot alterar la seva densitat i higroscopicitat.

L'interès per la matèria particulada ha crescut considerablement en la última dècada, com ho demostren la quantitat d'articles relacionats amb aquest tema (més de 1000 publicacions anuals des del 2004). Els compostos més analitzats han estat típicament inorgànics, com en un estudi on s'han analitzat els metalls presents en el material particulat en una zona urbano-industrial (Mbengue, 2014), però en l'actualitat es comença a analitzar també la fracció orgànica. Per exemple en un estudi es va analitzar els compostos nitro-orgànics en filtres de PM_{10} en una zona rural (Kahnt, 2013). En aquesta Tesi s'han estudiat compostos orgànics primaris i secundaris presents en els aerosols, els quals es descriuran amb detall més endavant.

Compostos Oxidants (l'ozó)

L'ozó és la substància principal d'aquest grup. S'ha de diferenciar entre l'ozó estratosfèric, que és imprescindible per a la vida del planeta ja que absorbeix les radiacions ultraviolades del sol, i l'ozó troposfèric, que és el que es troba a la superfície de la terra i és un important contaminant secundari. L'ozó estratosfèric va ser un punt de màxima preocupació (cap al 1987) quan es va produir un forat en la seva capa a causa dels CFCs (clorofluorocarbonats) que s'utilitzaven com a fluids refrigerants. El problema es va començar a solucionar amb el Protocol de Montreal, en el que es va firmar el compromís de reduir a la meitat la producció de CFC's en un període de 10 anys. En l'actualitat el problema es considera solucionat, degut a la prohibició dels productes causants, que s'han substituït per altres. L'ozó troposfèric es forma per reaccions induïdes per la llum solar (reaccions fotoquímiques) en les que participen principalment els òxids de nitrogen i els hidrocarburs presents a l'aire. Aquest fenomen rep el nom d'smog fotoquímic, i és bastant habitual a ciutats que tenen elevat tràfic i condicions càlides amb sol. L'ozó s'ha estudiat freqüentment degut a que és un contaminant important, pot causar mortalitat prematura (Nawahda, 2013), i també pels diferents problemes que ha creat en el medi ambient, però el seu estudi continua sent d'actualitat, degut per exemple a que els nivells permesos d'ozó es superen en moltes ciutats europees (Gerosa, 2007). En aquesta Tesi s'ha estudiat la variació temporal de l'ozó a Barcelona durant diversos anys, així com la seva relació complementària amb l'òxid de nitrogen, amb mètodes quimiomètrics.

Altres compostos

Entre els contaminants més estudiats tenim el diòxid de carboni (CO₂), que és un contaminant primari amb gran importància degut a la seva participació en l'efecte climàtic i l'escalfament global del planeta. El diòxid de carboni es troba present a l'atmosfera de forma natural, no és tòxic i té un paper important en el cicle natural del carboni, pel que no s'hauria de considerar una substància que contamina, però l'efecte hivernacle és un tema d'actualitat, pel que els estudis relacionats amb el CO₂ són habituals. El diòxid de sofre, SO₂, també és un important contaminant primari, que és irritant i tòxic,

i s'emet a l'atmosfera per activitats humanes com la metal·lúrgia i també de manera natural en l'activitat volcànica. El diòxid de sofre s'allibera en molts processos de combustió ja que combustibles com el carbó, el petroli, el dièsel o el gas natural contenen una quantitat de compostos amb sofre, però el seu ús s'ha limitat en el transport, principalment a la Unió Europea. El triòxid de sofre, SO_3 , és un contaminant conegut per reaccionar amb l'aigua produint la formació d'àcid sulfúric, H_2SO_4 , i l'anomenada pluja àcida que produeix importants problemes sobre la salut dels organismes biològics, la reproducció dels peixos i amfibis, la corrosió de metalls, etc.

Altres contaminants importants són les substàncies radioactives. Alguns isòtops radioactius com el radó 222, iode 131, cesi 137 i cesi 134 són emesos a l'atmosfera com a gasos o com a partícules en suspensió. Normalment es troben en concentracions baixes i no suposen perill, però en concentracions relativament altes (molt baixes en valor absolut) poden provocar càncer i afectar a la reproducció de les persones humanes i dels éssers vius. La seva presència a l'atmosfera pot ser deguda a fenòmens naturals però també a activitats humanes com les centrals nuclears, les armes nuclears, i a algunes pràctiques mèdiques.

Finalment dins d'aquesta categoria es poden incloure alguns paràmetres físics que no són substàncies químiques, però que també formen part de la contaminació atmosfèrica. Entre aquests paràmetres tenim la temperatura, que pot tenir importància en algunes zones, el soroll, que pot causar disminució de la capacitat auditiva o la contaminació electromagnètica que pot arribar a alterar el metabolisme cel·lular.

2.2 Quimiometria

La quimiometria és una disciplina que utilitza les matemàtiques, l'estadística i la lògica formal per dissenyar i seleccionar procediments experimentals òptims, i per obtenir la màxima informació i coneixement químic a partir de l'anàlisi de dades químiques (Massart, 1997). Encara que les metodologies estadístiques com l'ajust de corbes o el control estadístic es van

utilitzar en química analítica molt abans, no va ser fins el 1971 quan Svante Wold va introduir el terme quimiometria (Brereton, 1990). Una mica més tard, Bruce Kowalski amb col.laboració amb Wold van formar la Societat Internacional de Quimiometria (Hopke, 2003). Amb l'arribada dels ordinadors als laboratoris analítics i l'increment de les dades generades per instrumentació analítica cada vegada més sofisticada, la disciplina de la Quimiometria va adquirir força al llarg dels anys 80. És així com van aparèixer dues revistes especialitzades en quimiometria, *Chemometrics and Intelligent Laboratory Systems* (1986) i *Journal of Chemometrics* (1987), que van consolidar definitivament aquesta disciplina.

La quantitat d'informació disponible a partir de les mesures analítiques s'ha incrementat molt en els darrers anys, de manera s'han necessitat mètodes més complexos per resoldre la informació contenida en els estudis realitzats. La quimiometria ha passat a ésser una eina molt important, ja que ens permet processar i interpretar la gran quantitat d'informació química contenida en les dades analítiques obtingudes (Hanrahan, 2009). Dins dels procediments quimiomètrics, l'anàlisi estadístic multivariant continua guanyant acceptació entre els químics com ho demostra la seva creixent aparició en la literatura rellevant de les diferents disciplines químiques, especialment de la química analítica, i de les diferents disciplines aplicades que se'n deriven. Això és degut a la varietat i complexitat de les dades instrumentals que produeixen actualment els instruments moderns i al desenvolupament d'una gran quantitat d'algorismes matemàtics moderns disponibles pels químics analítics en aplicacions ambientals, biològiques o en l'anàlisi d'aliments (Escandar, 2007).

Les noves tècniques analítiques, cada vegada més sensibles i sofisticades, proporcionen una enorme quantitat de dades de diferents tipus (cinètiques, farmacèutiques, proteòmiques, metabolòmiques, etc.) que es poden classificar segons la seva complexitat i mida. Així les dades són d'ordre zero, si com a resultat d'una mesura sobre una mostra es té un valor escalar (es produeix una resposta individual per mostra). Són en canvi de primer ordre, si es té com a resultat d'una mesura sobre una mostra, un vector (per exemple un espectre). Són de segon ordre si es té com a resultat d'una mesura sobre una mostra, una taula o matriu de dades (per exemple utilitzant un instrument

analític que registra espectres d'excitació-emissió o acoblant dos instruments de primer ordre, com LC-DAD o CG-MS). Són de tercer ordre, si es té com a resultat d'una mesura sobre una mostra, un conjunt de taules o matrius de dades, o el que és el mateix, un cub de dades. Depenent de les dades que tinguem en cada cas particular, es poden utilitzar uns mètodes quimiomètrics o uns altres (Escandar, 2007).

Per dades d'ordre zero s'utilitza habitualment un ajust lineal per mínims quadrats, que es coneix com a calibració univariant, i que requereix selectivitat total per l'analít d'interès (Danzer, 1998). Per dades de primer ordre s'ha de compensar la falta de selectivitat amb algorismes matemàtics eficients per extraure el senyal que faci possible predir la concentració d'un constituent en presència d'altres senyals interferents. Per l'anàlisi de dades d'aquest tipus s'han desenvolupat mètodes com el *Principal Component Regression* (PCR) o el *Partial Least Squares* (PLS) (Hanrahan, 2009). Quan les dades es comporten de manera no lineal respecte de la concentració de l'analít, llavors es poden utilitzar altres mètodes com el *Artificial Neural Network* (ANN) o el *Least-Squares Support Vector Machine* (LS-SVM) (Suykens, 2009). En tots els casos es necessita la construcció d'un conjunt de mostres de calibració representatives de la variabilitat de les mostres desconegudes, i realitzar la validació del model obtingut amb un conjunt de validació apropiat (preferiblement no emprat per al calibratge del model) i representatiu del que seran posteriorment les mostres de predicció. Per dades de segon ordre i superiors, hi ha altres mètodes més sofisticats com el *Parallel Factor Analysis* (PARAFAC, Bro, 1997b), el *Generalized Rank Annihilation Method* (GRAM, Sanchez, 1986), el *Direct Trilinear Decomposition* (DTLD, Sanchez, 1990) o el *Multivariate Curve Resolution-Alternating Least Squares* (MCR-ALS, Tauler, 1995a), que són especialment rellevants per l'anàlisi de barreges complexes. Altres mètodes molt utilitzats per analitzar dades amb estructures de segon ordre, són l'anàlisi de components principals (*Principal Component Analysis*, PCA, Jolliffe, 2002), o en general els que es deriven de la família de mètodes anomenats d'anàlisi de factors (*Factor Analysis*, FA, Hopke, 2009a), els mètodes de reconeixement de pautes (*Pattern Recognition*, Hopke, 2009b) o els mètodes d'anàlisi d'agrupacions (*Cluster Analysis* (CA, Bratchell, 1989) o de

discriminació (*Discrimant Analysis* DA, Lavine, 2009) entre d'altres. Per dades d'ordres superiors s'utilitzen també els mètodes com PARAFAC, TUCKER3 (Tucker, 1966) o MCR-ALS.

Tots aquests mètodes tenen aplicacions molt diverses, entre les que destaquen les aplicacions medioambientals (per qualsevol dels compartiments ambientals, ja siguin l'aigua, l'aire o els sòls) que són el motiu d'estudi d'aquesta Tesi. Aquesta Tesi s'ha centrat en l'estudi de la contaminació de l'aire i una part molt important de la qual ha estat dedicada a la modelització i l'estudi de les fonts de contaminació, pel que en el següent apartat es farà una introducció als models de distribució proporcional de fonts de contaminació (*source apportionment*).

Models de distribució (proporcional) de fonts

En el camp de les ciències atmosfèriques els models de distribució (proporcional) de fonts (*source apportionment*) intenten reconstruir els impactes sobre un determinat lloc geogràfic de les emissions de les diferents fonts de contaminants atmosfèrics, basant-se en mesures ambientals registrades en diferents localitzacions o estacions de presa de mostra. L'assumpció bàsica és que la variació total associada als canvis de concentració dels contaminants pot ser descrita amb un nombre reduït de fonts de contaminació d'origens diversos (industrial, agrícola, antropogènica...) definits per perfils que descriuen la seva composició química i la seva distribució (o contribució) temporal i/o geogràfica. Les tècniques de distribució de fonts són eines importants en el disseny de programes de control efectiu d'emissions per reduir la pol·lució per partícules d'aire. Les possibilitats dels estudis de distribució de fonts es poden veure incrementades amb l'ús simultani de múltiples traçadors coneguts com a constituents dels perfils químics de composició de determinades fonts de contaminació.

El nombre de tècniques existents per a l'estudi de la distribució de fonts és relativament llarg i inclou mètodes com l'anàlisi de components principals (*Principal Component Analysis*, PCA, Jolliffe, 2002), el balanç de masses químic (*Chemical Mass Balance*, CMB, US-EPA, 1987) o la factorització positiva de matrius (*Positive Matrix Factorization*, PMF, Paatero, 1994). Aquestes

tècniques requereixen diferents graus de coneixement de les fonts de contaminació. En aquesta Tesi es proposa i s'utilitza el mètode de resolució multivariant de corbes per mínims quadrats alternats, MCR-ALS (Multivariate Curve Resolution Alternating Least Squares; Tauler, 1995a; Tauler, 1995b; Jaumot, 2005), que ja ha estat utilitzat prèviament en estudis de distribució de fonts en sistemes aquàtics (Salau, 1997; Terrado, 2009) i que s'explicarà en detall en l'apartat 4.4.2 de la present Tesi.

El CMB necessita un coneixement previ detallat de la naturalesa química de totes les fonts possibles i dels seus perfils d'emissió i en canvi el PCA, PMF i MCR-ALS no precisen d'aquest coneixement i les fonts s'estimen durant la mateixa anàlisi de dades. El PCA utilitza restriccions matemàtiques fortes (ortogonalitat, normalització) que li permeten obtenir una solució única a partir de la qual es pot fer una interpretació ambiental a partir dels perfils (*scores* i *loadings*) dels components que descriuen la naturalesa de les fonts de contaminació. Aquesta interpretació és no obstant de vegades difícil (degut a les restriccions matemàtiques imposades per PCA en la seva obtenció) però la seva aplicació és senzilla i proporciona solucions sense ambigüitat. El mètode CMB requereix un coneixement previ de les fonts d'emissió i per tant, és incapaç de descobrir quines són les fonts de contaminació no prèviament conegudes. El PMF integra restriccions naturals com la no-negativitat i estimacions de les incerteses en una optimització no lineal dels perfils de distribució i de composició de fonts. El fet que apliqui restriccions més naturals que el PCA té per una banda l'avantatge de proporcionar solucions físicament més similars a les vertaderes, però en canvi no presenta sempre l'avantatge de solucions úniques com les de PCA. El mètode MCR-ALS (descriu amb detall en l'apartat 4.4.2) es basa en una optimització per mínims quadrats alternats sota restriccions de no-negativitat i produeix perfils més interpretables que PCA i anàlegs al PMF (Tauler, 2009; Staminirova, 2011), però a partir d'un algorisme matemàtic més senzill i flexible que aquest últim. A l'apartat 4 es fa una descripció detallada dels mètodes de determinació de distribució de fonts (*source apportionment*) de contaminació emprats en aquesta Tesi.

2.3 Contaminants ambientals estudiats en aquesta Tesi

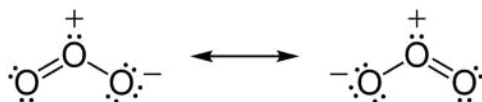
En aquesta Tesi s'han estudiat tant contaminants primaris com secundaris. Per una part s'ha estudiat la contaminació atmosfèrica per òxid de nitrogen i per ozó, a partir de dades obtingudes en campanyes de seguiment de la qualitat atmosfèrica del Departament de Medi Ambient de la Generalitat de Catalunya. Per una altra banda, s'ha estudiat la contaminació atmosfèrica per compostos orgànics obtinguts a partir de la seva extracció en mostres de material particulat ($PM_{2.5}$ i PM_1) i analitzats posteriorment al laboratori per cromatografia de gasos acoblada a espectrometria de masses.

En les primeres etapes d'aquesta Tesi es va fer un estudi sobre les variacions geogràfiques i temporals de la contaminació per NO a Catalunya. La gran quantitat disponible de dades de diferents anys en diferents estacions, va permetre l'aplicació de diversos mètodes quimiomètrics (principalment MCR-ALS) i interpretar la seva informació. Això va permetre la preparació d'una primera publicació, on es comparaven les dades obtingudes en diferents estacions de mesura. Es van poder investigar les relacions que existien entre elles i estudiar els perfils de variació temporal de les fonts principals de contaminació detectades. Com també es disposava de dades de concentracions d'ozó en alguns anys coincidents, es va realitzar l'estudi de les relacions complementàries entre aquest compost ozó i el NO. En aquest estudi es va avançar en la implementació de diferents restriccions aplicables al mètode MCR-ALS, com la interacció entre components, que s'explica en detall en la secció 4.4.2.4 de l'apartat 4 d'aquesta Tesi. Aquesta nova restricció permetia millorar la interpretació de les fonts o pautes de contaminació quan presentaven algun dels seus perfils temporals comuns. També es van comparar els resultats obtinguts per MCR-ALS amb els obtinguts a partir d'altres mètodes quimiomètrics equivalents com són els mètodes PARAFAC i el TUCKER3 (veure seccions 4.4.3 i 4.4.4 de l'apartat 4). A partir d'aquests resultats es va preparar una segona publicació. A partir dels resultats obtinguts, es va poder fer una interpretació molt detallada dels mecanismes

complementaris de formació i destrucció de O_3 i de NO i de la seva diferent incidència a escala diària (horària), estacional i interanual.

Com només s'havia pogut treballar simultàniament amb les dades de NO i O_3 que tenien anys coincidents, es va desenvolupar una nova versió del programa MCR-ALS que permetés treballar amb dades incompletes, és a dir amb dades dels dos contaminants que presentessin anys coincidents i anys no coincidents. En el cas estudiat, aquestes dades incompletes corresponien doncs a anys sencers dels quals no es disposava de dades. El mètode desenvolupat ha permès optimitzar simultàniament les matrius de dades de O_3 i de NO encara que no coincideixin en el seu any d'adquisició, de forma combinada amb aquelles en les quals si que hi havia aquesta coincidència. Aquests resultats han estat publicats recentment (Article 3) i s'explicaran amb detall a la secció 5.

A continuació es mostren les estructures del NO i l' O_3 .



massa molecular 48 g mol^{-1}

Les característiques més destacades de l'ozó i del NO ja s'han explicat en els apartats anteriors.

L'estudi de contaminants orgànics en filtres de material particulat va sorgir a partir de la col·laboració en el projecte SAPUSS (Solving Aerosol Problems by Using Synergistic Strategies, FP7-PEOPLE-2009-IEF-254773, Dall'Osto, 2013). En aquest projecte s'ha estudiat la composició de l'aerosol de la ciutat de Barcelona i s'han determinat els processos principals (fonts

d'emissió) responsables de la seva composició. Es va fer un seguiment de la qualitat de l'aire de la ciutat de Barcelona durant 30 dies entre els mesos de setembre i octubre de l'any 2010 utilitzant diferents tècniques de presa de mostra i d'anàlisis "on-line" i "off-line". Es van recollir mostres de material particulat (PM_{10} , $PM_{2.5}$, PM_1) en diverses estacions situades a la ciutat de Barcelona. En el cas particular d'aquesta Tesi, s'han analitzat mostres de PM_1 en dues localitzacions de la ciutat de Barcelona (Torre Girona (UB) i carrer Urgell (RS)). En el capítol 3 d'aquesta memòria, s'explica el procediment experimental analític utilitzat per a la determinació dels compostos orgànics amb detall. S'ha emprat la cromatografia de gasos acoblada a l'espectrometria de masses per a determinar la concentració cada 12h de 35 compostos en les dues localitzacions durant 26 dies. Per poder interpretar simultàniament i millor tota la informació adquirida a partir d'aquestes anàlisis s'han utilitzat mètodes estadístics i d'anàlisi multivariant de dades, també anomenats en el cas de dades químiques, mètodes quimiomètrics. S'han utilitzat mètodes com el PCA (Anàlisi de Components Principals) i el MCR-ALS (Resolució Multivariant de Corbes per mínims quadrats alternats), els quals permeten la resolució i interpretació de les possibles pautes, tendències o fonts de contaminació i la determinació de la seva distribució geogràfica i temporal. Amb aquests mètodes s'ha pogut deduir i comparar la qualitat atmosfèrica de l'aire de les estacions de presa de mostra investigades i estimar les influències de les possibles fonts d'emissió de PM_1 a Barcelona en el període estudiat. Amb els resultats obtinguts en aquest estudi s'ha fet una publicació científica (Article 4), la qual s'explicarà amb més detall en la secció de resultats.

L'estudi de contaminants orgànics urbans a partir de filtres de material particulat es va continuar amb la col·laboració amb el projecte CalNex 2010 (California Nexus, *Research at the Nexus of Air Quality and Climate Change*) en el qual es van mesurar mostres d'aire a Pasadena (Los Angeles). Aquesta campanya es va centrar en l'estudi de la química nocturna i en la formació de l'aerosol orgànic secundari durant un mes, entre maig i juny de 2010. En el nostre cas es van analitzar filtres de $PM_{2.5}$ dels compostos orgànics solubles en aigua amb el mateix procediment anterior, que s'explicarà amb detall al capítol 3 d'aquesta memòria. S'han utilitzat els mètodes quimiomètrics per resoldre i

interpretar les possibles fonts de contaminació orgànica de la ciutat de Los Angeles. Els resultats obtinguts a Los Angeles s'han comparat amb els obtinguts a Barcelona per investigar les seves similituds i diferències, en relació a les fonts d'emissió entre les dues localitzacions. Les dues ciutats tenen un clima mediterrani semblant caracteritzat per elevades radiacions solars i freqüents condicions atmosfèriques anticiclòniques, per la qual cosa resulta especialment interessant aquesta comparació a pesar de la seva gran distància geogràfica. També ha estat especialment interessant l'aplicació simultània del mètode MCR-ALS al conjunt de dades format pels resultats obtinguts en les dues ciutats. Amb aquests resultats s'ha publicat recentment un article en una revista científica (Article 5).

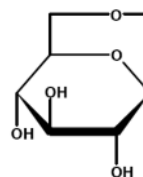
Els contaminants orgànics identificats i analitzats cromatogràficament en aquests dos treballs es descriuen a continuació.

Galactosan, Mannosan, Levoglucosan

Aquests compostos s'obtenen de la degradació tèrmica de la cel·lulosa i hemicel·lulosa (Simoneit, 2002; Reche, 2012) i es poden utilitzar com a marcadors de la crema de biomassa. Formen una part important de la fracció de carboni orgànic soluble en els aerosols atmosfèrics i tenen importància en la formació de núvols i precipitacions i en el clima en general, entre altres efectes. Degut a la crema de fusta per a calefacció de les llars, les concentracions de levoglucosan i dels seus estereoisòmers, mannosan i galactosan, són generalment elevades durant el hivern. Encara que el levoglucosan és susceptible d'oxidació durant el seu transport atmosfèric, (Hennigan, 2011; Kessler, 2010) s'ha observat també a baixes concentracions a l'estiu a Barcelona (van Drooge, 2012a). El levoglucosan és una glucosa deshidratada amb un grup funcional cetona i és un traçador específic de la crema de biomassa. Les emissions de manosan i galactosan són menors que les de levoglucosan.

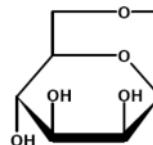
Levoglucosan

$C_6H_{10}O_5$ (1,6-anhidro-beta-D-glucopiranososa)
 massa molecular $162.14 \text{ g mol}^{-1}$



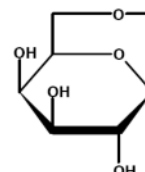
Galactosan

$C_6H_{10}O_5$ (1,6-anhidro-beta-D-galactosa)
 massa molecular $162.14 \text{ g mol}^{-1}$



Mannosan

$C_6H_{10}O_5$ (1,6-anhidro-beta-D-manopiranososa)
 massa molecular $162.14 \text{ g mol}^{-1}$



Hidrocarburs aromàtics policíclics

Els hidrocarburs aromàtics policíclics (PAH) són un grup nombrós de substàncies químiques que són derivats polimèrics del benzè. Van ser els primers agents químics reconeguts com a causants de tumors malignes en humans, degut a la seva facilitat d'interacció amb el material genètic cel·lular. Els PAH es formen bàsicament quan la matèria orgànica es sotmet a elevada temperatura durant suficient temps. Es poden produir per fonts naturals com les erupcions volcàniques o incendis forestals, però a l'actualitat en el món desenvolupat, són més importants els produïts pels motors de combustió dels vehicles motoritzats, per les calefaccions domèstiques o per les plantes industrials. També l'hàbit de fumar o els aliments massa torrats per efecte del foc, poden fer incrementar l'exposició a aquests compostos fins superar els nivells de risc (Vives, 2001).

Els hidrocarburs aromàtics policíclics constitueixen una família de compostos que es troba àmpliament distribuïda en el medi ambient i que es caracteritzen per contenir dos o més anells de benzè units entre sí. La majoria contenen només àtoms de carboni i d'hidrogen i són estructures polinuclears de tipus aromàtic. Els PAH són substàncies lipòfiles, tendència que s'incrementa amb l'augment de la seva massa molecular. Són inestables fotoquímicament, degradant-se amb la llum (Albers, 1995). La toxicitat de molts PAH acostuma a

ser baixa, però són importants els efectes que produeixen a mig i llarg termini. Existeixen centenars de compostos del tipus PAH i en molts d'ells s'ha demostrat una activitat carcinogènica, mutagènica o immunosupressora (Koss, 1999), però no es bioacumulen, ja que es metabolitzen per diferents vies i els seus metabòlits s'eliminen per la orina. En general, els PAH de baix pes molecular són els que més s'absorbeixen. La genotoxicitat dels PAH depèn en gran mesura de la seva estructura. L'exposició humana a aquests compostos es deu a la inhalació o ingestió, encara que també és possible per via cutània. Després de la seva absorció es distribueix per diversos òrgans i teixits, especialment els rics en lípids. Un cop incorporats a l'organisme, experimenten una oxidació enzimàtica convertint-se en epòxids i dihidrodiols. Aquestes espècies químiques constitueixen la forma genotòxicament activa dels PAH, que poden formar adductes covalents amb les proteïnes i amb els àcids nucleics cel·lulars. Els adductes amb ADN poden donar lloc a mutacions genètiques, amb possibles conseqüències tumorals malignes per als individus exposats, sense oblidar el risc de malformacions en embrions i fetus. A la Taula 2.1 es mostren els efectes carcinogènics, genotòxics i mutagènics d'alguns PAH.

<i>PAHs</i>	<i>Carcinogenicidad</i>	<i>Genotoxicidad</i>	<i>Mutagenicidad</i>
Fenantreno	I	L	+
Antraceno	N	N	-
Pireno	N	L	+
Benzo[fluoreno]	I	I	?
Benzo[a]antraceno	S	S	+
Benzo [e]pireno	I	L	+
Benzo[a]pireno	S	S	+
Dibenz[a]antraceno	S	S	+
Benzo[ghi]perileno	I	I	+
Dibenzopirenos	S	I	+
2-Nitronaftaleno	N	L	-
1-Nitropireno	I	S	+

(S= suficiente; I= insuficiente; N= no carcinogénico; L= limitados.)
Mutagenicidad (Test de Ames): + (positivo); - (negativo); ? (inconcluso).

Taula 2.1. Dades relatives als efectes carcinogènics, genotòxics i mutagènics d'alguns PAHs.

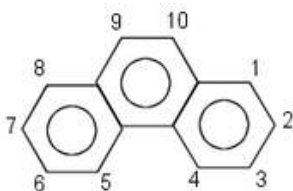
Els PAH es formen durant la combustió incompleta del petroli i dels seus derivats, però també en la combustió incompleta de tot tipus de matèria

orgànica com el carbó, la fusta, el tabac o la vegetació en general (Albers, 1995). De fet són els components principals de l'emissió dels motors dièsel i de la crema de biomassa (Schauer, 1999 i 2001). La piròlisi de la matèria orgànica és un procés molt estès, el que explica la ampla distribució dels PAH en tots els compartiments ambientals (aire, sòls, aigües, sediments i teixits biològics) i en els aliments (carn i peix fumats, cereals, sucre, olis, greixos, llets i derivats làctics, sobre tot). Les concentracions de PAH en el medi ambient poden arribar a ésser importants. Aquests compostos s'han trobat en l'aire de les grans ciutats a concentracions elevades degut a la densitat de vehicles o fonts directes de combustió, però també es troben en les aigües, sediments i neu de llacs d'alta muntanya (Carrera, 2001; Vilanova, 2001). Gràcies a la seva relativa volatilitat i capacitat d'adsorció a les partícules atmosfèriques, els PAH es transporten contínuament a través de l'aire, i podent ser dipositats en llocs remots a gran distància.

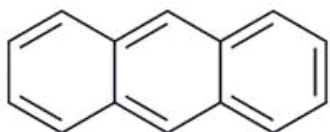
A continuació es descriuen els usos d'alguns PAH que es troben freqüentment en les campanyes de seguiment ambiental. L'antracè s'utilitza en la producció d'antraquinona, la qual s'utilitza com a matèria primera per a la fabricació de colorants ràpids, per diluir conservants de la fusta, i en la producció de fibres sintètiques, plàstics i monocristalls. El fenantrè s'utilitza en la fabricació de colorants i explosius, i en la síntesi de fàrmacs. El fluorantè s'utilitza com a material de revestiment per protegir l'interior de les tuberies d'aigua potable d'acer i els tancs d'emmagatzematge (Stellman, 1998; Lawerys, 1994). El benzo[a]pirè i el crisè es troben en el quitrà d'hulla, en els gasos d'escapament dels vehicles i també es poden trobar en altes concentracions als fums de tabac.

Els PAH que s'han estudiat en aquesta Tesi són:

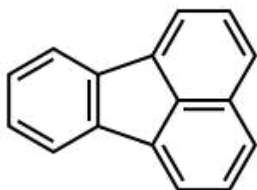
Fenantrè, $C_{14}H_{10}$, massa molecular 178,23 g/mol



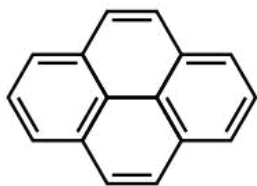
Antracè, $C_{14}H_{10}$, massa molecular 178,23 g/mol



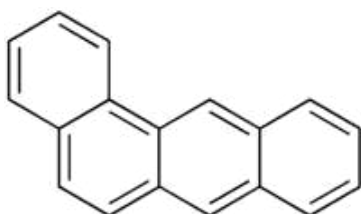
Fluorantè, $C_{16}H_{10}$, massa molecular 202,26 g/mol



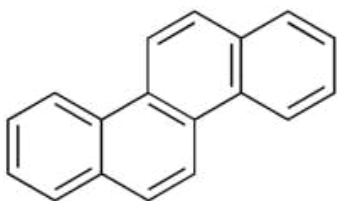
Pirè, $C_{16}H_{10}$, massa molecular 202,25 g/mol



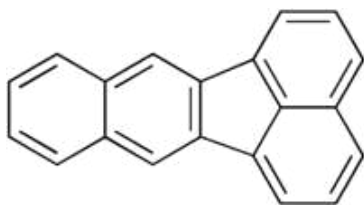
Benz[a]antracè, $C_{18}H_{12}$, massa molecular 228,29 g/mol



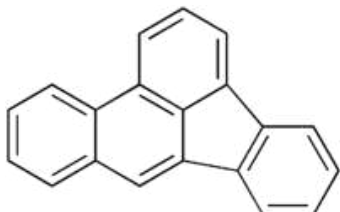
Crisè, $C_{18}H_{12}$, massa molecular 228,28 g/mol



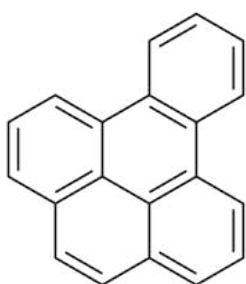
Benzo[k]flourantè, $C_{20}H_{12}$, massa molecular 252,31 g/mol



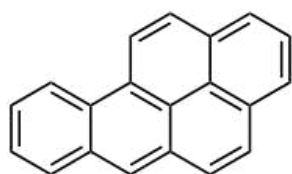
Benzo[b]fluorantè, $C_{20}H_{12}$, massa molecular 252,31 g/mol



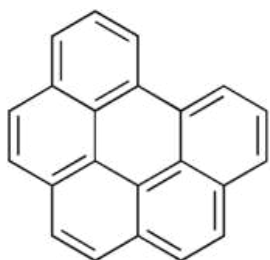
Benzo[e]pirè, $C_{20}H_{12}$, massa molecular 252,31 g/mol



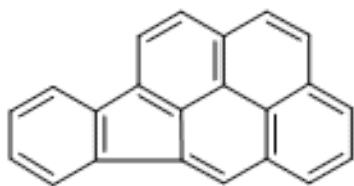
Benzo[a]pirè, $C_{20}H_{12}$, massa molecular 252,31 g/mol



Benzo[ghi]perilè, $C_{22}H_{12}$, massa molecular 276,33 g/mol



Indeno[cd]pirè, $C_{22}H_{12}$, massa molecular 276,33 g/mol



Coronè, $C_{24}H_{12}$, massa molecular 300,35 g/mol



Hopans

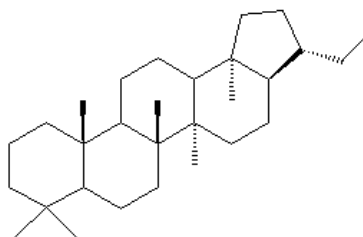
Els hopans són triterpens pentacíclics que contenen entre 27 i 35 àtoms de carboni en una estructura naftalènica composta de quatre anells de sis carbonis i un anell de cinc carbonis. Procedeixen de precursors amb grups funcionals presents en les bactèries. Els hopanoids estan presents en les membranes de molts procariotes, tant aeròbics com anaeròbics i la seva funció sembla ser la millora de l'estabilitat de les parets cel·lulars. Els hopans amb la configuració 17a(H), 21b(H) (en el interval C_{27} - C_{35}) són característics del petroli (Hester, 2008), i s'utilitzen com a marcadors moleculars dels olis minerals. La seva presència pot estar relacionada amb residus d'olis de lubricació no cremats en els vehicles a motor, especialment en els cotxes dièsel (Schauer, 2007).

Els hopans que s'han estudiat en aquesta Tesi són el 17a(H)21b(H)-29-norhopà i el 17a(H)21b(H)-hopà, que tenen l'estructura següent i difereixen en un metil i la posició d'aquest:

17 α (H), 21 β (H)-29-norhopà

C₂₉H₅₀

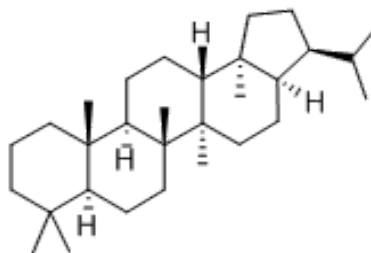
massa molecular 398.71 g/mol



17 α (H), 21 β (H)-hopà

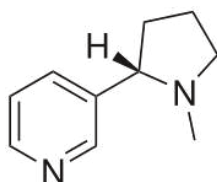
C₃₀H₅₂

massa molecular 412.73 g/mol



Nicotina

La nicotina és un alcaloid que està present a elevades concentracions en el fum de tabac. De fet, el fum del tabac és una de les fonts petites que contribueixen al material particulat orgànic fi. La nicotina és un possible traçador del fum de tabac i està distribuïda entre la fase gas i la fase particulada. Encara que generalment sota condicions ambientals (degut a la seva relativa elevada volatilitat) està principalment en la fase gas, es pot detectar a nivells traça en els filtres de material particulat (Bi, 2005). La nicotina associada a la fase gas és susceptible de fotodegradació (Eatough, 1989) i per tant no és un traçador quantitatiu fiable, encara que es pot utilitzar com un traçador qualitatiu. Si es detecta nicotina en la matèria particulada fina urbana, s'ha de considerar com a indicació de que el fum de tabac és troba present. En canvi, la seva quantificació acurada s'hauria de basar en traçadors més estables.



Nicotina

C₁₀H₁₄N₂

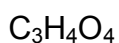
massa molecular 162.23 g mol⁻¹

Àcids carboxílics, dicarboxílics i altres

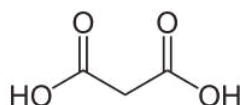
Els àcids dicarboxílics són un grup important dels compostos orgànics identificats en els aerosols atmosfèrics (Rogge, 1993b). La seva contribució a la quantitat total de partícules de carboni va des del 1 i el 3 % en les àrees urbanes i semiurbanes fins al 10% en medis marins remots (Kawamura, 1999). Els diàcids carboxílics són altament solubles en aigua i tenen el potencial de modificar les propietats higroscòpiques de les partícules atmosfèriques, i afecten a la seva mida (Cruz, 1998). Els àcids dicarboxílics provenen de diferents fonts, entre les quals hi ha les emissions primàries de combustibles fòssils, la crema de biomassa (Simoneit, 1986), el fum de tabac (Rogge, 1994), així com la oxidació fotoquímica de precursors orgànics tant d'origen antropogènic com biogènic (Chebbi, 1996). Entre aquests àcids dicarboxílics hi ha els de cadena curta com l'oxàlic, el màlic, el succínic o el glutàric. Entre els de cadena més llarga tenim els àcids azelaic, pimèlic i subèric, que són productes d'oxidació dels àcids grassos insaturats, com l'àcid oleic (Kawamura, 1987). Malgrat els avenços que s'han fet per esclarir els tipus de fonts dels àcids dicarboxílics, encara queda per esbrinar la seva presència i formació en les partícules (Kerminen, 2000).

Els compostos estudiats són:

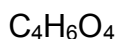
Àcid malònic (àcid propanodioic)



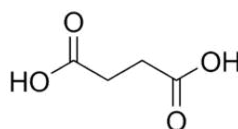
massa molecular 104,06 g/mol



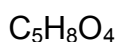
Àcid succínic (àcid butanodioic)



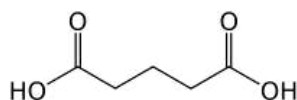
massa molecular 118,09 g/mol



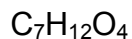
Àcid glutàric (àcid pentanodioic)



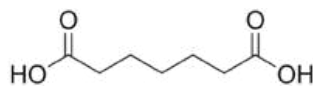
massa molecular 132,12 g/mol



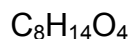
Àcid pimelic (àcid heptanodioic)



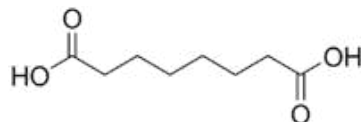
massa molecular 160,17 g/mol



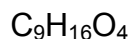
Àcid subèric (àcid octanodioic)



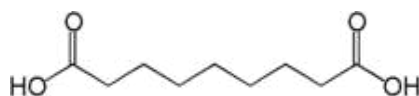
massa molecular 174,20 g/mol



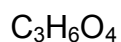
Àcid azelaic (àcid nonanodioic)



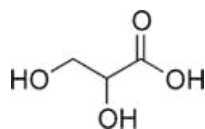
massa molecular 188,22 g/mol



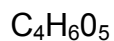
Àcid glicèric (àcid 2,3-dihidroxiopropanoic)



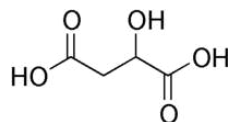
massa molecular 106,08 g/mol



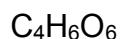
Àcid màlic (àcid hidroxibutanodioic)



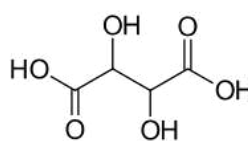
massa molecular 134,09 g/mol



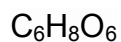
Àcid tartàric (àcid 2,3-dihidroxibutanodioic)



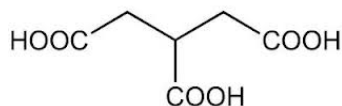
massa molecular 150,087 g/mol



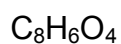
Àcid tricarbàllític (1,2,3-propantricarboxílic acid)



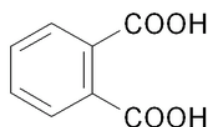
massa molecular 176 g mol⁻¹



Àcid ftàlic (àcid benzè-1,2-dicarboxílic)



massa molecular 166,14 g/mol



Productes d'oxidació del α -pinene

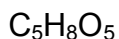
El α -pinene (2,6,6-trimetilbiciclo [3,1,1] hept-2-è), és un compost orgànic de la classe dels terpens i és un dels dos isòmers del pinene. Els terpens (formats per unitats d'isoprè (2-metil-1,3-butadiè)) són principalment emesos per les coníferes i el α -pinene és un dels emesos en quantitats més grans. Està ben establert que els terpens donen lloc a l'aerosol orgànic secundari a través de reaccions amb l'ozó i els radicals NO_3 (Claeys, 2007). En anàlisis recents duts a terme amb α -pinene en cambres de fums s'han observat la formació de productes altament oxidats (Edney, 2003; Kubatova, 2000) que també s'han detectat en mostres de material particulat. El *cis*-pinonic acid, 3-hidroxi-glutaric i 3-metil-1,2,3-butanetricarboxylic acid (MBTCA) han estat identificats tots ells en els aerosols atmosfèrics i estan relacionats amb la oxidació fotoquímica de compostos biogènics volàtils del α -pinene (Claeys, 2007; Szmigielski, 2007).

L'àcid *cis*-pinònic és un producte de primera generació de la oxidació dels monoterpens (α i β -pinens) per ozó o pel radical OH (O'Dowd, 2002). En alguns estudis s'ha trobat àcid *cis*-pinònic en aerosols en les immediacions de boscos confirmant així la seva importància en aerosols biogènics (Cheng, 2004). També estudis recents de laboratori han trobat l'àcid *cis*-pinònic com a indicador de l'edat de l'aerosol atmosfèric, i es considera per tant que és un bon traçador de l'aerosol orgànic biogènic secundari (Jimenez, 2009).

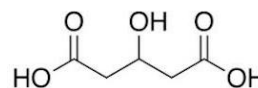
La formació de l'àcid 3-hidroxi-glutàric i del MBTCA es pot explicar a partir de la reacció d'oxidació en fase gas de l'àcid *cis*-pinonic, amb la participació del radical OH (Szmigielski, 2007). També es pot formar per irradiació del α -pinene en presència de NO_x (Edney, 2003). La presència d'aquests compostos es va detectar per primera vegada en mostres d'aerosol de l'Amazònia i de Bèlgica (Kubatova, 2000).

Els compostos derivats de la oxidació del α -pinene que s'han estudiat són:

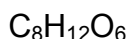
Àcid 3-hidroxi-glutàric (àcid 3-hidroxi-pentanodioic)



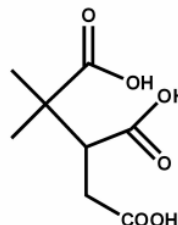
massa molecular 148,11 g mol⁻¹



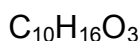
Àcid 3-metil-1,2,3-butantricarboxílic (MBTCA)



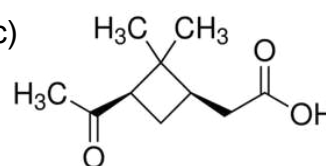
massa molecular 204,12 g mol⁻¹



Àcid cis-pinònic (àcid 3-acetil-2,2-dimetilciclobutaneacetic)



massa molecular 184,23 g/mol



Productes d'oxidació de l'isoprè

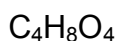
Hi ha diversos compostos que són traçadors d'isoprè en l'aerosol orgànic secundari, alguns d'ells han estat descoberts recentment com els 2-metiltetrols (2-metiltreitol i 2-metileritrol) i l'àcid 2-metilglicèric (Claeys, 2004a; Hallquist, 2009). La fotooxidació del isoprè és una font important del aerosol orgànic secundari (SOA), encara que no totes les emissions de isoprè es transformen en SOA (Claeys, 2004b). Els 2-metiltetrols es poden obtenir de la reacció d'isoprè amb peròxid d'hidrogen, el qual que es pot formar a l'atmosfera per combinació de peroxiradicals (Penkett, 1979), en presència d'àcid sulfúric com a catalitzador (Claeys, 2004a). L'àcid 2-metilglicèric es produeix a partir de la fotooxidació ràpida del isoprè emès localment. Altres traçadors de SOA del isoprè són, a més dels 2-metiltetrols i l'àcid 2-metilglicèric, els C₅ alquetriols (2-metil-1,3,4-trihidroxibutè (cis i trans) i 3-metil-2,3,4-trihidroxibutè) (Edney, 2005, Wang, 2005). Els C₅ alquetriols es poden explicar per la obertura, catalitzada amb àcid, d'un anell epoxidiol derivat del isoprè, com el 1,2-epoxi-2-metil-3,4-dihidroxibutà o el 1,2-dihidroxibutà-2-metil-3,4-epoxibutà.

Els compostos orgànics solubles en aigua com els C₅ alquetriols, poden millorar la capacitat dels aerosols atmosfèrics per actuar com a nuclis de

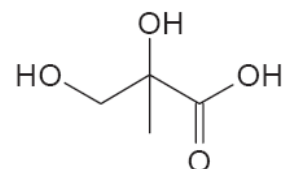
condensació de núvols, i que per tant són importants en la formació de boires i en la disminució de la visibilitat en àrees forestals (Wang, 2005).

Els compostos derivats de la oxidació de l'isoprè que s'han estudiat són:

Àcid 2-metilglicèric (àcid 2,3-dihidroxi-2-metil-propanoic)

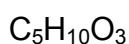


massa molecular 120,10 g/mol

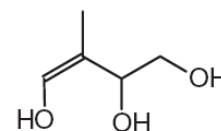
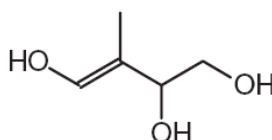


C_5 alquetriols

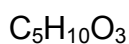
2-metil-1,3,4-trihidroxi-1-butè (cis i trans)



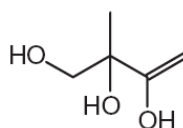
massa molecular 118 g/mol



3-metil-2,3,4-trihidroxi-1-butè

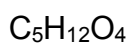


massa molecular 118 g/mol

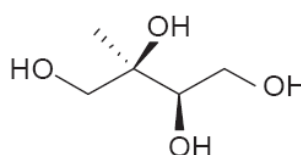


2-metiltetrols (2-metil-1,2,3,4-tetrahidroxibutà)

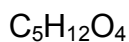
2-metiltreitol



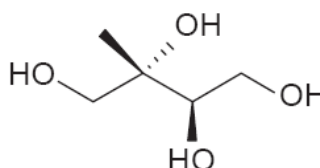
massa molecular 136.15 g/mol



2-metileritritol



massa molecular 136.15 g/mol



Altres compostos analitzats

Altres compostos com l'àcid oleic i els n-alcans s'han detectat però no s'han pogut emprar en aquesta Tesi perquè sortien en els blancs i les seves concentracions no eren fiables. Els blancs han de sortir nets dels compostos analitzats, i si no és el cas no es poden quantificar adequadament ja que no es

pot saber si la seva presència és deguda a que estan a l'aire o a alguna part del procediment de la presa de mostra.

Els àcids grassos insaturats constitueixen una important fracció dels àcids grassos. L'àcid oleic (9-octadecenoic) té un doble enllaç que el fa susceptible de ser atacat per ozó, i per això pot donar lloc a l'aerosol orgànic oxidat a l'atmosfera (Vesna, 2009). En els aerosols marins continentals s'han trobat àcids grassos insaturats, sent l'àcid oleic un dels més trobats. L'àcid oleic s'ha utilitzat com a marcador de la oxidació atmosfèrica de materials orgànics degut a la seva prevalència en la matèria particulada troposfèrica (Rogge, 1991).

Els n-alcans arriben a l'atmosfera per fonts naturals i antropogèniques. Algunes fonts biogèniques de partícules de n-alcans inclouen plantes, pol·len, microorganismes i insectes (Kadowaki, 1994). Els n-alcans també venen de la utilització de combustibles fòssils i olis lubricants i deriven de la seva combustió incompleta (Lin, 2004). S'han trobat en aerosols recollits en àrees marines, àrees netes continentals i en àrees urbanes (Simoneit, 1991). Aquest grup de compostos no polars és bastant estable químicament, i per tant, la seva quantificació és relativament fàcil en comparació amb la dels altres compostos orgànics. Els alcans tenen unes pautes de distribució peculiars i s'han utilitzat els seus perfils per explicar l'origen dels aerosols orgànics atmosfèrics (Schauer, 2000). En els alcans que provenen de fonts biològiques, predominen els hidrocarburs amb un nombre de carbonis imparell. Els alcans que provenen de fuels fòssils com el petroli i del carbó no presenten cap tendència respecte al nombre d'àtoms de carboni.

2.3.1 Unitats de concentració

La concentració dels diversos agents contaminants en l'atmosfera s'expressa habitualment amb dos tipus d'unitats:

- *Unitats volum/volum*, cm^3/m^3 o mm^3/m^3
- *Unitats massa/volum*, mg/m^3 , $\mu\text{g}/\text{m}^3$ o ng/m^3

S'ha de tenir sempre en compte que quan es considera la contaminació atmosfèrica, els valors de ppm (o ppb) són diferents dels mg/dm³ (o µg/dm³). Per convertir unes unitats en les altres, s'ha de tenir en compte les equacions d'estat dels gasos. En parlar de concentracions de contaminants en l'atmosfera és necessari diferenciar els valors d'emissió dels d'immissió:

Emissió: és la concentració de contaminants que emet un focus determinat. Es mesura a la sortida del focus emissor.

Immissió: és la concentració de contaminants present en el sinus d'una atmosfera determinada i, per tant, és a aquests valors als que es troben exposats els éssers vius.

Els valors de les concentracions corresponents a l'emissió o a la immissió acostumen a ser completament diferents. Els valors d'immissió són conseqüència no només dels valors d'emissió, sinó també de tots els fenòmens de barreja, transport, deposició i transformació que experimenten els contaminants una vegada emesos a l'atmosfera.

2.3.2 Legislació

Els nivells d'immissió que estableix la legislació han de ser tals que garanteixin que l'aire tingui una qualitat acceptable, encara que sabem que sempre hi haurà un cert grau de contaminació present en qualsevol mesura experimental. La vigilància dels nivells d'immissió ha de fer-se en els nuclis urbans, especialment en aquells punts situats en un entorn pròxim a polígons industrials. Els grans focus emissors tenen l'obligació de tenir la seva pròpia xarxa de vigilància dels valors d'immissió del seu entorn geogràfic i d'aportar les dades a l'Administració competent. Per vigilar els valors d'immissió existeixen diverses xarxes de vigilància. Les xarxes tenen com a finalitat conèixer els nivells de contaminants a que es veuen sotmesos els ciutadans, la naturalesa o el patrimoni historicoartístic. Els estudis de la Organització Mundial de la Salut (OMS), i les investigacions realitzades sobre vegetació i medi

natural, fan que cada vegada els valors admesos de concentracions d'agents contaminants siguin inferiors i, que les legislacions cada cop siguin més exigents i restrictives.

El *Reial Decret 102/2011* desenvolupa els aspectes relacionats amb la qualitat de l'aire i de la protecció de l'atmosfera de la *Llei 34/2007*, de 15 de novembre, i transposa la nova *Directiva europea 2008/50/CE*, tot integrant els reials decrets anteriorment aprovats. Per tant, la legislació de referència per l'avaluació de la qualitat de l'aire és la *Llei 34/2007* i la del *Reial Decret 102/2011* en les quals s'estableixen diferents tipus de paràmetres:

- *Valor límit*: és un nivell que es fixa a partir dels coneixements científics existents, i que té per finalitat prevenir i reduir els efectes nocius dels contaminants per a la salut humana, el medi ambient en el seu conjunt i els altres béns de qualsevol naturalesa. Han de ser respectats en tot el territori i, en cas de superar-se, s'han d'estudiar les causes que han originat aquesta superació i s'han d'establir actuacions adequades per evitar la seva repetició. S'estableixen tant valors límit diaris com anuals i, en ocasions, es distingeix entre el valor límit per a la protecció humana i els valors límit per a la protecció de la vegetació o dels ecosistemes.
- *Nivell crític*: nivell fixat a partir dels coneixements científics existents per sobre dels quals poden produir-se efectes nocius per alguns receptors com les plantes, els arbres o els ecosistemes naturals però no per l'home.
- *Valor objectiu*: nivell d'un contaminant que s'haurà d'assolir, en la mesura que sigui possible, en un moment determinat per evitar, prevenir o reduir els efectes nocius sobre la salut humana, el medi ambient en el seu conjunt i altres béns de qualsevol naturalesa.
- *Valors guia*: valors que tenen la consideració d'objectius a assolir a més llarg termini i que seran indicatius d'aire net, sense contaminació. Si la concentració de contaminants d'una determinada zona està per sota del valor límit però per sobre del valor guia, la seva tendència serà cap a la reducció, i cap a la implantació en la zona considerada de nous processos de desenvolupament de tecnologies netes.

- *Llindar d'informació*: nivell d'un contaminant a partir del qual una exposició de breu duració suposa un risc per a la salut humana de grups de població especialment vulnerables. Les administracions competents han de subministrar una informació immediata i apropiada sobre aquests valors llindars.
- *Llindar d'alerta*: nivell a partir del qual una exposició de breu duració suposa un risc per la salut humana i que afecta al conjunt de la població. Estableix concentracions de contaminants que obliguen a la presa immediata de mesures per part de les administracions competents per tal de finalitzar el problema.

A continuació es mostren els valors dels paràmetres de qualitat de l'aire vigents per els contaminants estudiats en aquesta Tesi:

Òxids de nitrogen (NO₂ i NO_x)

- Valor límit horari (1h) per a la protecció de la salut humana: 200 µg/m³ de NO₂ (no podrà superar-se en més de 18 ocasions per any civil)
- Valor límit anual (1 any civil) per a la protecció de la salut humana: 40 µg/m³ de NO₂
- Nivell crític (1 any civil): 30 µg/m³ de NO_x (es consideren només les dades de les estacions representatives dels ecosistemes naturals i de la vegetació que cal protegir)
- Llindar d'alerta (1h): 400 µg/m³ si es supera durant tres hores consecutives en llocs representatius de la qualitat d'aire d'una zona que sigui de com a mínim, 100 km² o d'una zona o aglomeració sencera, prenent la que sigui superfície que sigui menor.

Ozó

- Valor objectiu (màxim de les mitjanes 8-horàries del dia) per a la protecció de la salut humana: 120 µg/m³ (no es podrà superar més de 25 vegades per cada any civil de mitjana en un període de 3 anys).

- Valor objectiu per a la protecció de la vegetació: $18000 \mu\text{g}/\text{m}^3 \times \text{h}$ de mitjana en un període de 5 anys (es baixarà a $6000 \mu\text{g}/\text{m}^3$ a partir de 01/01/2020).
- Llindar d'informació (mitjana horària): $180 \mu\text{g}/\text{m}^3$
- Llindar d'alerta (mitjana horària): $240 \mu\text{g}/\text{m}^3$

Hidrocarburs aromàtics policíclics

Es pren com a valor de concentració referència pel control dels PAH totals, el valor de la concentració del benzo[a]pirè.

- Valor objectiu pel benzo[a]pirè en aire ambient en la fracció PM_{10} : $1\text{ng}/\text{m}^3$ com a mitjana durant un any natural (aquest valor s'ha de complir a partir del 1 de gener de 2013).

Material particulat

Partícules en suspensió de diàmetre inferior a $10 \mu\text{m}$ (PM_{10}):

- Valor límit diari (24h) per a la protecció de la salut humana: $50 \mu\text{g}/\text{m}^3$ (no es podrà superar més de 35 ocasions per any)
- Valor límit anual (1 any civil): $40 \mu\text{g}/\text{m}^3$

Partícules en suspensió de diàmetre inferior a $2.5 \mu\text{m}$ ($\text{PM}_{2.5}$):

- Valor objectiu anual (1 any civil): $25 \mu\text{g}/\text{m}^3$ (en vigor des del 1 de gener de 2010)
- Valor límit anual (1 any civil): $25 \mu\text{g}/\text{m}^3$ en vigor a partir del 1 de gener de 2015 (es baixarà a $20 \mu\text{g}/\text{m}^3$ a partir del 1 de gener de 2020).

Capítol 3

PART EXPERIMENTAL

La part experimental portada a terme en aquesta Tesi es divideix en dues parts. Per una banda hi ha el treball amb contaminants obtinguts a partir de les bases de dades de la Generalitat de Catalunya, i per l'altra hi ha el treball experimental amb contaminants orgànics obtinguts en filtres de material particulat i analitzats per nosaltres al laboratori mitjançant GC-MS.

3.1 Contaminants de les bases de dades de la Generalitat

En aquesta Tesi s'han analitzat les bases de dades obtingudes a partir de campanyes de seguiment i vigilància de la contaminació ambiental del Departament de Mediambient de la Generalitat de Catalunya. La Generalitat estableix programes de vigilància de la qualitat de l'aire per assegurar-se del compliment de la legislació en vigor i en cas d'haver desviacions, detectar-les i dissenyar els plans d'actuació precisos per aconseguir l'adequació de l'aire que respirem dins les condicions preestablertes en les directrius mediambientals. Aquests programes de vigilància estan dissenyats de forma que permetin valorar els nivells de contaminació existents en cada moment, controlar els episodis puntuals de contaminació que es produeixin, estudiar la migració de contaminants i fer un seguiment de l'evolució de la contaminació al llarg del temps.

Actualment, existeixen diferents xarxes per a la mesura i vigilància de les concentracions de contaminants atmosfèrics, amb diferents objectius i àmbits d'actuació. Entre elles, tenim les xarxes a escala mundial per veure les concentracions de diferents compostos no reactius de l'aire després del seu transport a llarga distància, les xarxes a escala regional que proporcionen informació sobre la concentració i deposició dels contaminants atmosfèrics, estimant els fluxos a través de les fronteres de les diferents nacions, i per últim, les xarxes a escala nacional o autonòmica, per mesurar els nivells de contaminació urbana i/o en grans instal·lacions industrials. A Espanya, les diferents Comunitats Autònomes i també alguns ajuntaments tenen les seves pròpies xarxes, constituïdes per estacions manuals o automatitzades que aporten dades mitjanes diàries o horàries, respectivament. Per establir el

nombre i ubicació dels punts de presa de mostra han de tenir-se en compte diversos aspectes, entre els quals cal destacar els següents:

- Existència de focus emissors, fixos, mòbils i difusos.
- Característiques topogràfiques i urbanístiques de la zona.
- Característiques meteorològiques de la zona.
- Dades demogràfiques
- Localització de zones sensibles.
- Demanda social.

A Catalunya hi ha la Xarxa de Vigilància i Prevenció de la Contaminació Atmosfèrica (XVPCA) on es detecten nivells d'immissió dels contaminants principals. Aquesta xarxa té estacions en diferents localitzacions de Catalunya depenent de les característiques meteorològiques de cada zona com poden ser ubicacions de costa i ubicacions de muntanya, i tant en àrees urbanes com d'agricultura. Aquesta xarxa va començar el seu funcionament a l'any 1983 i ha anat canviant per adaptar-se als canvis de normativa i a les necessitats de cada territori (www20.gencat.cat/portal/site/mediambient). Va començar amb poques estacions, però aquestes han anat creixent fins arribar en l'actualitat a 82 estacions automàtiques. En aquestes estacions es mesuren diferents contaminants com són SO₂, H₂S, NO, NO₂, O₃ i CO, i també variables meteorològiques com temperatura, humitat relativa, radiació solar, pressió, precipitacions i velocitat del vent.

Les dades de que es van disposar per fer aquesta Tesi estaven repartides en 50 estacions repartides durant tota Catalunya durant els anys del 2000 al 2006. S'han utilitzat principalment dades de concentracions de NO i O₃ obtingudes entre els anys 2000 i 2006 en diferents estacions. Les concentracions d'ozó es van mesurar per fotometria UV amb analitzadors automàtics d'acord amb la normativa ISO 13964:1998. Les concentracions de NO es van mesurar d'acord amb la norma UNE 77212:1993 equivalent a la ISO 7993:1985 mitjançant analitzadors automàtics de quimioluminescència. Les mesures i els procediments estan establerts per la legislació Espanyola segons el Real Decret 1073/2002. Els dos analitzadors automàtics, tan el de NO com el de O₃ donen mesures cada minut i es guarden en períodes semihoraris.

D'aquestes dades el Departament de Medi Ambient fa una anàlisi i validació junt amb el seu emmagatzemament. Ells reben i coordinen totes les dades dels diferents contaminants obtingudes en les diferents estacions de la xarxa.

3.1.1 Estructura de les dades ambientals

De la Xarxa de Vigilància i Prevenció de la Contaminació Atmosfèrica de la Generalitat s'han extret les dades de concentració ($\mu\text{g}/\text{m}^3$) de NO i O₃ obtingudes durant cada mitja hora de cada dia durant diferents anys en diferents estacions. Les taules de dades s'han arranjat en fulls Excel com el que es veu a continuació en la Taula 3.1

Fecha	Badalona	L'Hospitalet	Sant Celoni	Sant Cugat	Sabadell	Granollers	Bcn Eixample	Bcn Gracia	Sant Andreu	Bcn Poblenou	Sant Adrià
01/01/2004 00:30	3	38	8	9	13	14	41	50	48	31	88
01/01/2004 01:00	7	22	13	10	6	15	33	20	23	17	90
01/01/2004 01:30	2	19	6	8	7	14	27	24	12	3	89
01/01/2004 02:00	1	26	7	9	7	12	24	32	14	14	95
01/01/2004 02:30	2	27	7	9	8	12	27	36	10	17	102
01/01/2004 03:00	2	19	6	9	13	13	25	36	14	24	92
01/01/2004 03:30	5	24	5	11	10	13	37	45	21	22	117
01/01/2004 04:00	1	12	4	11	7	11	15	40	31	22	121
01/01/2004 04:30	12	29	4	16	10	13	13	13	50	32	119
01/01/2004 05:00	21	25	4	23	29	11	16	1	49	35	109
01/01/2004 05:30	27	23	4	21	37	11	27	1	52	31	119
01/01/2004 06:00	23	25	4	24	35	13	33	3	52	29	131
01/01/2004 06:30	36	24	3	23	31	10	29	2	44	25	128
01/01/2004 07:00	22	23	4	24	36	9	43	1	42	24	123
01/01/2004 07:30	26	25	5	25	31	10	34	24	35	29	155
01/01/2004 08:00	21	21	5	28	30	10	48	23	34	36	163
01/01/2004 08:30	23	22	5	27	41	10	34	9	36	27	162
01/01/2004 09:00	21	21	6	26	41	11	45	4	35	24	143

Taula 3.1. Mesures de concentració d'ozó ($\mu\text{g}/\text{m}^3$) en diferents estacions

Això vol dir que per cada compost i cada estació de presa de mostra es té una columna de concentracions amb 17520 valors, equivalents a 48 mesures diàries durant els 365 dies de l'any. Depenent de la informació que es vulgui obtenir, es treballarà directament amb les columnes (vectors) d'aquesta taula o arrançant les dades en una matriu. Per realitzar una millor interpretació dels perfils diaris i anuals les dades es poden reagrupar per tenir en les columnes els diferents valors semihoraris de cada dia (48) i en les files els 365 dies de l'any (Figura 3.1). D'aquesta manera es podran observar directament els perfils diaris i anuals.

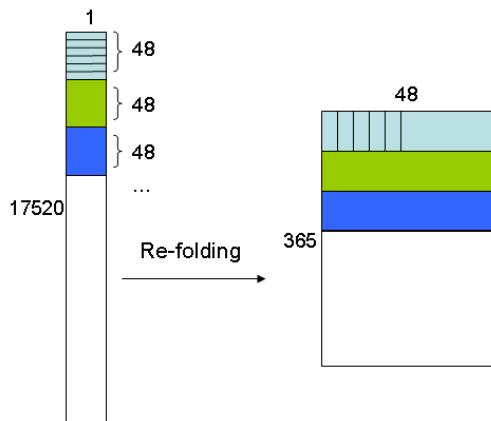


Figura 3.1. Reagrupament dels valors semihoraris de concentració d'un compost individual. A les files hi ha els valors diaris d'un any (365 valors), i a les columnes els valors semihoraris de cada dia (48 valors). Les dimensions de la taula o matriu de dades corresponent serà 365x48.

Les taules de dades dels diferents anys o estacions també es podran analitzar conjuntament (Figura 3.2) amb diferents tipus d'estructures que s'explicaran en detall en l'apartat 4.1. En la Figura 3.2 es mostra l'estructura conjunta de taules de dades de diferents anys que s'han reagrupat prèviament com es mostrava a la figura anterior.

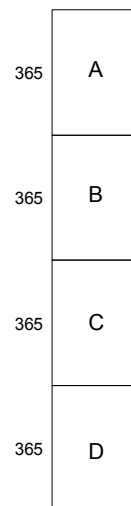


Figura 3.2. Estructura conjunta de diversos anys amb matrius prèviament reagrupades.

3.2 Contaminants orgànics en filtres de material particulat analitzats al laboratori

En aquesta Tesi s'han analitzat contaminants orgànics obtinguts en filtres de material particulat a les ciutats de Barcelona i Los Angeles. Aquests filtres provenen de dues campanyes diferents, SAPUSS i CalNex, com ja s'ha explicat anteriorment en la secció 2.3.2. En el projecte SAPUSS (a Barcelona) s'han analitzat 103 mostres de filtres de material particulat submicrònic (PM_{1}), en dues localitzacions durant el 22 de setembre al 18 d'octubre de 2010. En el projecte CalNex (a Los Angeles) s'han analitzat 69 filtres de $PM_{2.5}$ en una localització del 15 de maig al 16 de Juny de 2010. En tots els casos també es van recollir i analitzar blancs de camp que van rebre el mateix tractament que les mostres, és a dir, són filtres que han seguit el mateix procés que les mostres amb la única diferència que no s'ha pres mostra amb ells.

3.2.1 Presa de mostra de l'aerosol

Les dues localitzacions on s'ha fet la presa de mostra, Los Angeles i Barcelona, tenen un clima similar, caracteritzat per una elevada radiació solar i freqüents condicions anticiclòniques, que permeten l'acumulació d'aerosols primaris i secundaris. A més, els dos llocs es troben sota la intensa influència de les emissions primàries urbanes de vehicles motoritzats.

A Barcelona, totes les mostres de filtres es van recollir en filtres de fibra de quars (Tissuquartz™ Filters, 2500 QAT-UP, Pall Life Sciences) prèviament escalfats a l'estufa a 450°C tota la nit per eliminar contaminants orgànics, de 15 cm de diàmetre, preparats amb un dispositiu de presa de mostres d'alt volum PM_{1} (Digitel-DH80; 30 m³/h), com el que es pot veure a la Figura 3.3. Les mostres es van recollir en intervals de 12 h de 9:00 a 21:00 i de 21:00 a 9:00, en hora local.

Les 103 mostres recollides es van emmagatzemar en un congelador a -18°C fins la seva anàlisi química. La presa de mostra es va realitzar en dues localitzacions de Barcelona. La primera és una localització urbana amb contribucions d'emissions de fons, *Urban Background site*, (estació UB, 41.3899°N; 2.1161°E; 80 m.a.s.l.) situada al nord-oest de Barcelona, en un petit

parc a prop de la zona universitària, a una distància d'uns 500 m d'una de les avingudes amb més tràfic de la ciutat (Diagonal). L'altre és un punt de trànsit intens, *Road Site*, (estació RS; 41.3884°N; 2.1500°E; 40 m.a.s.l.) localitzada a 4 m del carrer Urgell al costat d'un pàrking i de la sortida del metro d'Hospital Clínic. Aquesta localització també està a prop d'un carrer molt transitat de quatre carrils de vehicles, un de bicis i un de vianants amb una mitjana de pas de vehicles de 17000 per dia. Les dues estacions estaven separades 3 km entre elles (Dall'Osto, 2013).



Figura 3.3 Captador d'alt volum Digitel-DH80, 30 m³/h

A Pasadena (Califòrnia), les mostres de filtres es van recollir en filtres de fibra de quars (Tissuquartz™ Filters, 2500 QAT-UP, Pall Life Sciences), prèviament escalfats durant 24h a 550°C, de 20.3 x 25.4 cm, i preparats per a una presa de mostra d'alt volum PM_{2.5} (TE-6001, 60 m³/h). A més dels filtres de 23h (des de mitjanit fins a les 23h del dia posterior en hora local) també es van recollir 6 filtres de manera intensiva, en set dies preseleccionats (22, 25, 30 i 31

de maig, i 4, 7 i 10 de juny). En aquestes preses de mostra intensives, es va recollir un filtre durant els següents períodes de temps: 0:00-6:00, 6:00-9:00, 9:00-12:00, 12:00-15:00, 15:00-18:00, i 18:00-23:00, sempre en hora local. Les 69 mostres de filtres recollides, com la que es pot veure a la Figura 3.4, es van emmagatzemar a -18°C en un congelador fins a la seva anàlisi química. La presa de mostra de Los Angeles es va realitzar en una localització urbana (*urban background site*) a Pasadena (LA; 34.1406°N ; 118.1225°W ; 240 m.a.s.l.). Aquesta estació es trobava a 17 km al nord-est del centre de Los Angeles i a 44 km de l'àrea del port de Long-Beach.



Figura 3.4. Filtre de material particulat

Barcelona es troba en la conca mediterrània occidental i és una de les ciutats més poblades d'Europa, amb al voltant de 1.7 milions d'habitants i una àrea metropolitana de 5 milions d'habitants ($1.5 \text{ hab}/\text{km}^2$). Barcelona es caracteritza per una elevada proporció de motocicletes (30% del total de vehicles), vehicles pesats i una elevada proporció en l'ús de vehicles privats (60% del total de vehicles) dels quals el 75% són dièsel (Dades bàsiques de mobilitat, 2011). Les activitats relacionades amb el tràfic constitueixen les principals fonts d'emissió primàries en aquesta àrea (Reche, 2012). En la conca de LA les fonts d'emissió principals són també els vehicles motoritzats, que estan formats principalment per vehicles industrials lleugers i vehicles de benzina (Bahreini, 2012). Els vehicles dièsel representen només una petita fracció (<5%) que està formada principalment per vehicles industrials pesats (<http://www.arb.ca.gov/ei/emissiondata.htm>). Aquesta diferent composició de vehicles pot representar diferències en la composició dels aerosols formada per compostos orgànics secundaris trobats en els dos llocs, ja que els motors de benzina formen més compostos orgànics secundaris que els motors dièsel (Bahreini, 2012).

3.2.2 Tractament de mostra

Una vuitena o quarta part de cada filtre va ser extreta per ultrasons amb una mescla de diclorometà i metanol (2:1 (v/v); 3 x 5 mL; Merck, Germany) durant 15 minuts. Abans de l'extracció es van afegir a les mostres i als blancs, 25 µL d'estàndard de levoglucosan deuterat (LG-D₇, Cambridge Isotopic Laboratories, UK), àcid succínic deuterat (SA-D₄, Sigma Aldrich, Germany), alcà deuterat (n-C₂₄D₅₀, Cambridge Isotopic Laboratories, UK), i una barreja d'estàndards d'hidrocarburs aromàtics policíclics deuterats (anthracene-D₁₀, benz[a]anthracene-D₁₂, benzo[k]fluoranthene-D₁₂ i benzo[ghi]perylene-D₁₂ (Dr. Ehrenstorfer)), per així poder avaluar les possibles pèrdues produïdes durant el procediment analític. Els extractes es van filtrar amb una xeringa acoblada a un filtre de membrana de tefló de 0,45 µm (Whatman, USA) i així eliminar les partícules insolubles. Aquests extractes es van concentrar al rotavapor fins a 1 mL.

Una vegada obtinguts als extractes s'han seguit dos procediments, un per analitzar els compostos polars i un altre per analitzar els compostos neutres. Els compostos polars són compostos orgànics solubles en aigua, i la seva anàlisi és relativament ràpida. L'anàlisi de compostos polars es va realitzar tant en les mostres de Barcelona com en les de Los Angeles. El procediment per analitzar els compostos neutres és molt més lent i costós ja que s'ha de fer una columna de separació per cada mostra. En aquest cas, els compostos neutres només es van analitzar en les mostres de Barcelona.

Per analitzar la part polar, com per exemple els polisacàrids anhidres, els àcids, els poliols i la nicotina, el procediment analític utilitzat va ser el mateix que el descrit anteriorment per altres autors (El Haddad, 2011; Medeiros, 2007; van Drooge, 2012a). Alíquotes de 25 µL dels extractes es van evaporar sota corrent de N₂ fins a sequedat. Es va procedir a la seva derivatització en èsters metilsililats mitjançant la reacció amb 25 µL de bis(trimetilsilil)trifluoroacetamida (BSFTA) (Supelco, USA) i 10 µL de piridina (Merck, Germany) a 70°C durant 3h. Abans de la injecció al cromatògraf de gasos, s'afegeixen a més, 25 µL de

l'estàndard intern de pirè-D₁₀, per la correcció de possibles canvis de resposta durant la seva anàlisi.

Per analitzar la part neutra (hidrocarburs aromàtics policíclics i hopans), l'extracte resultant de l'anàlisi dels compostos polars, es va evaporar quasi fins a sequedat sota corrent de nitrogen i es va redissoldre en 0.5 mL d'una mescla hexà : diclormetà (9:1 v/v) (Merck, Germany). Llavors es va passar per una columna cromatogràfica d'adsorció (com la que es veu a la Figura 3.5) amb 1g d'òxid d'alumini (Merck, Germany) que havia estat prèviament activat a 120°C durant tota la nit. Els analits es van eluir amb 4 mL de hexà : diclormetà (9:1 v/v) i 4 mL de hexà : diclometà (1:2 v/v) i es van recollir conjuntament. Les fraccions recollides es van concentrar sota corrent de N₂ fins a 50µL. Abans de la injecció al cromatògraf de gasos, també es van afegir 25 µL d'estàndard intern de pirè-D₁₀.



Figura 3.5. Columna cromatogràfica d'adsorció amb òxid d'alumini

3.2.3 Anàlisi per GC-MS

L'anàlisi de les mostres de filtres s'ha realitzat en un cromatògraf de gasos acoblat a un espectròmetre de masses (GC-MS Thermo Trace GC Ultra – DSQ II), com el que es veu a la Figura 3.6, equipat amb una columna capil·lar de 60 m (HP-5MS 0,25-mm x 0,25-µm).

La separació dels diferents compostos de les mostres es va començar a una temperatura de 60°C, mantenint-la durant 1 min, i llavors s'incrementava

fins a 120°C a 12°C/min i fins a 310°C a 4°C/min on es mantenia durant 10 min. Les temperatures de l'injector, la font d'ions, la línia de quadrupol i de transferència van ser 280°C, 200°C, 150°C i 280°C, respectivament. El gas portador era Heli a un cabal de 0,9 mL/s i la detecció es va fer en mode d'impacte electrònic (70 eV) operant en *fullscan* (m/z 50 – 650) per l'anàlisi dels compostos polars, i en mode *SIM* (Selective Ion Monitoring) per a l'anàlisi dels compostos neutres, ja que la seva concentració era molt més baixa.



Figura 3.6. GC-MS (Thermo Trace GC Ultra – DSQ II)

Els diferents compostos es van identificar pels seus ions corresponents (que apareixen a la Taula 3.2), i els seus temps de retenció, prenent com a referència els patrons deuterats afegits mencionats anteriorment (àcid succínic-D₄, levoglucosan-D₇, benz[a]antracè-D₁₂, benzo[k]fluorantè-D₁₂ i benzo[ghi]perilè-D₁₂).

Els pics cromatogràfics d'alguns dels compostos com el 3-hidroxiàcid glutàric, el MBTCA, el C5 alquen triols i els 2-metiltetrols, dels quals no es tenien els seus estàndards disponibles van ser identificats per comparació amb els patrons de fragmentació d'espectrometria de masses trobats a la literatura i a les llibreries corresponents (Claeys, 2007; Kourtchev, 2005; Cleemens, 2007).

La quantificació es va portar a terme mitjançant el càlcul de les concentracions a partir de corbes de calibració externa, i tenint en compte les recuperacions dels estàndards deuterats prèviament esmentats així com la de

l'estàndard intern, i les concentracions dels blancs. El límit de quantificació (LOQ) en unitats de concentració es va calcular a partir de la divisió dels valors més baixos mesurats en les corbes de calibració pels volums de les fraccions analitzades.

Compostos	Ío (m/z)
fenantrè	178
antracè	178
antracè-D ₁₀	188
fluorantè	202
pirè	202
benz[a]antracè	228
benz[a]antracè-D ₁₂	240
crisè	228
benzofluorantens	252
beno[k]fluorantè-D ₁₂	264
benzo[e]pirè	252
benzo[a]pirè	252
indeno[123cd]pirè	276
benzo[ghi]perilè	276
benzo[ghi]perilè-D ₁₂	288
coronè	300
17 α (H)21 β (H)-29-norhopà	191
17 α (H)21 β (H)-hopà	191
nicotina	84
levoglucosan	204
levoglucosan-D ₇	206
galactosan	217
mannosan	204
àcid malònic	233
àcid succínic	247
àcid succínic-D ₄	251
àcid glutàric	261
àcid pimèlic	289
àcid subèric	303
àcid azelaic	317
àcid glicèric	292
àcid màlic	233
àcid tartàric	292
àcid tricarbàlltic	377
àcid ftàlic	295
àcid 3-hidroxi glutàric	349
àcid 3-metil-1,2,3-butantricarboxílic (MBTCA)	405
àcid cis -pinonic	171
C5 alquen triols	231
àcid 2-metilglicèric (2MGA)	219
2-metiltreitol	219
2-metileritritol	219

Taula 3.2. Ions característics utilitzats en el GC-MS

3.2.4 Estructura de les dades cromatogràfiques

A partir de l'anàlisi per GC-MS dels diferents filtres de material particulat s'obtingueren les concentracions en ng/m³ dels diferents compostos orgànics esmentats anteriorment en el apartat 2.3.2. D'aquesta manera es van obtenir

matrius de dades, en les que a les seves columnes hi ha els diferents compostos analitzats i a les seves files, les mostres analitzades en els diferents dies. Igual que s'ha explicat abans per als contaminants en bases de dades de la Generalitat, les matrius de dades corresponents a les dues localitzacions de Barcelona, es poden analitzar conjuntament a partir de la matriu augmentada (en la direcció de les columnes, una a sobre de l'altre), ja que les dues matrius de dades individuals tenen en comú els compostos analitzats. Pel cas de les mostres analitzades a Los Angeles, on només s'han considerat els compostos de la part polar de l'extracte, es pot fer la seva anàlisi individual o també conjuntament amb les dades corresponents de Barcelona (mateixos compostos analitzats). En l'apartat 4.1 es descriurà amb més detall les estructures i les agrupacions possibles de les diferents matrius.

Capítol 4

TRACTAMENT DE LES DADES EXPERIMENTALS

4.1 Estructura de les dades

La informació que es pot obtenir a partir de l'aplicació dels mètodes quimiomètrics depèn de com s'han estructurat les dades experimentals. Les dades ambientals s'organitzen generalment en taules de dades on a les seves columnes hi ha les variables mesurades, i en les seves files hi ha les mostres o objectes sobre les quals s'han mesurat aquestes variables. Aquesta taula de dades formarà una matriu de dades, com la que es mostra a la Figura 4.1. Els conjunts de dades organitzats en taules o matrius de dades, tenen doncs dues direccions o modes de mesura (*two-way or two-mode data sets*).

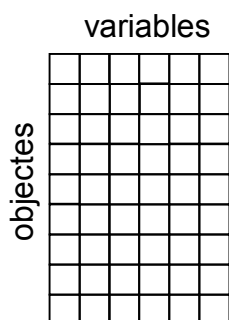


Figura 4.1. Conjunt de dades en dues direccions

Quan es tenen múltiples conjunts de dades que han proporcionat cadascun d'ells una taula o matriu de dades, es poden agrupar de diferents maneres. En el cas que els dos modes o direccions d'aquestes taules o matrius de dades siguin coincidents (mateix número de files i de columnes), el conjunt de dades passarà a tenir tres direccions o modes i es podran agrupar en un cub o en un paral·lelepípede de dades (*three-way or three-mode data sets*) (Figura 4.2). En cas de tenir més direccions o modes de mesura, les estructures de dades corresponents s'anomenen generalment dades multidireccionals o multimode (*multiway or multimode data*) (Smilde, 2004).

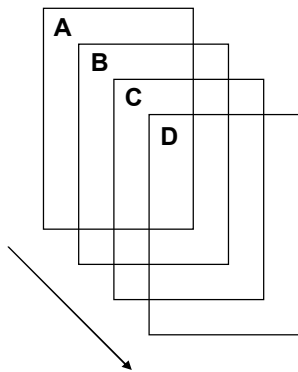


Figura 4.2. Agrupació de múltiples conjunts de dades en un de cub o paral·lelepípede

De tota manera, quan les taules o matrius de dades corresponents a l'anàlisi del mateix sistema ambiental no són coincidents entre si en una de les dues direccions o modes (diferent nombre de mostres o diferent nombre de variables en les diferents matrius de dades), l'agrupació (anàlisi conjunta) de les taules de dades individuals encara es pot fer a partir de la seva augmentació en taules o matrius més grans, utilitzant la direcció o mode que tenen en comú. Òbviament, aquest procediment també es pot utilitzar per taules o matrius de dades que tenen les dues direccions o modes iguals, és a dir que tenen el mateix nombre de files o de columnes. En aquesta Tesi, quan es tracta d'analitzar simultàniament conjunts múltiples de dades, s'utilitzarà fonamentalment el procediment d'augmentació de matrius. La denominació de 'dades multiconjunt' (*multiset data*) s'utilitzarà en general per múltiples conjunts de dades, siguin o no coincidents en totes les seves direccions o modes de mesura.

L'augmentació de matrius consisteix en concatenar les diferents matrius de dades en la direcció de les files (una al costat de l'altre, Figura 4.3a) o en la direcció de les columnes (una a sota de l'altre, Figura 4.3b). En química ambiental és fàcil generar dades multidireccionals (*multiway*) i/o multiconjunt (*multiset*) quan es mesuren simultàniament moltes variables sobre conjunts grans de mostres en condicions diferents. Els avenços instrumentals tant en sensibilitat com en precisió han incrementat de forma considerable la quantitat de variables i d'informació disponible en els estudis ambientals. Cada vegada

són més comuns instruments capaços de generar grans quantitats de dades en tres o més direccions a partir de l'anàlisi d'una sola mostra individual, com és el cas GC/GC/MS (Parastar, 2013) o el GC/MS/MS (Ferrer, 2013). També en la última dècada ha augmentat de forma considerable l'interès per la fusió de dades (*data fusion*) on es combina informació obtenida per diferents tècniques instrumentals (Vera, 2011).

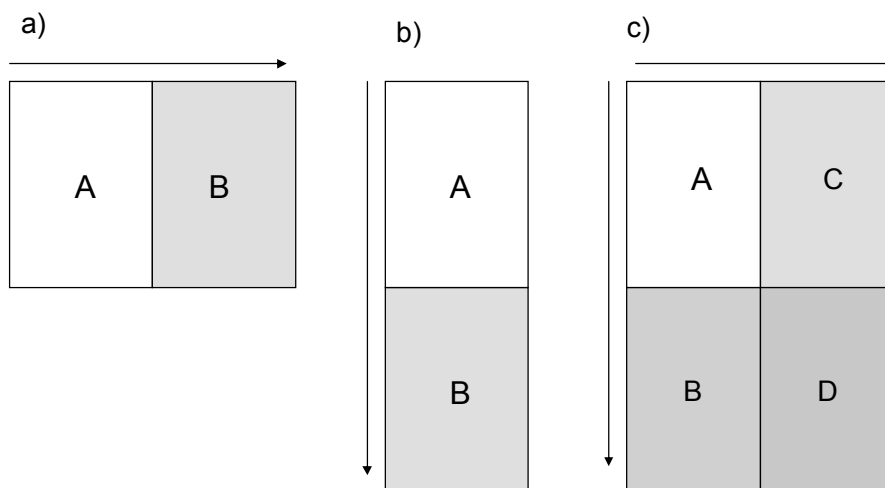


Figura 4.3. Agrupacions de múltiples conjunts de dades en dues direccions

Agrupacions de dades en una única taula o matriu de dades

El cas més senzill d'agrupació de dades correspon al cas on tenim una única taula de dades. Les dades s'agruparan en una matriu **D** de dimensions I (objectes) \times J (variables).

Agrupacions de dades en múltiples taules o matrius de dades

Un segon cas és aquell en el que es tenen dos o més taules de dades relacionades en alguna de les seves dues direccions o modes. En aquest cas, depenent de si les dades tenen en comú els objectes (files) o les variables (columnes), es pot crear una nova matriu augmentada de dades, posant una matriu al costat de l'altra (augmentació en la direcció de les files de la matriu, Figura 4.3a) o una matriu sota de l'altra (augmentació en la direcció de les columnes de la matriu, Figura 4.3b). Si les dades tenen en comú tant els objectes (files), com les variables (columnes), llavors les matrius es poden

agrupar de manera combinada (augmentació en la direcció de les files i de les columnes (Figura 4.3c)). Aquesta augmentació de matrius es pot estendre fàcilment a moltes matrius de dades, que tenen o no les mateixes dimensions en les seves dues direccions. És imprescindible, això sí, que com a mínim tinguin en comú una de les seves dues direccions o modes.

Agrupacions de dades en tres (*three-way*) o més direccions

El cas més senzill d'estructures de dades en tres direccions (*three-way*) o modes, és aquell en el que es tenen taules de dades en forma de cub de dades. En aquesta situació, el conjunt de dades agrupades en les diferents capes del cub han de ser matrius de dades d'igual mida (igual nombre de files i columnes), o el que és el mateix, han de tenir en comú les mostres o objectes (files) i les variables (columnes), de manera que es pugui formar el cub o paral·lelepípede (Figura 4.2) de dades. Si es té més d'un cub de dades llavors han de coincidir els objectes (files) i les variables (columnes) dels diferents cubs, i aquests també es poden posar un al costat o a sota de l'altre (Figura 4.4).



Figura 4.4. Agrupació de múltiples conjunts de dades en diversos cubs o paral·lelepípedes

Evidentment, les dades que es poden agrupar en tres o més direccions o modes (per exemple en forma de cub), poden també agrupar-se en matrius augmentades per files i per columnes, o segons els dos modes simultàniament, tal com s'ha indicat en l'apartat anterior.

4.2 Tractaments preliminars

Els tractaments preliminars, més correntment anomenats mètodes de pretractament de dades, inclouen aquells tractaments que es fan al conjunt de les dades, abans d'aplicar els mètodes quimiomètrics escollits, per així millorar la seva anàlisi i interpretació posterior. Aquests pretractaments de les dades inclouen un primer grup de tractaments emprats per a solucionar problemes associats a l'obtenció de les dades, com poden ser la pèrdua o absència de determinats valors de les dades (*missings*) o la consideració de valors censurats o per sota del límit de detecció instrumental. Un segon tipus de pretractaments, inclouen aquells que són necessaris per poder aplicar correctament els mètodes quimiomètrics escollits i possibilitar una interpretació correcta dels resultats. Alguns d'aquests pretractaments de dades poden millorar substancialment la informació que es pot extraure després d'aplicar els diferents mètodes quimiomètrics.

Pretractament de conjunts de dades amb valors absents (*missing values*)

Un dels primers problemes a resoldre quan es treballa amb conjunts de dades multivariants està relacionat amb l'absència d'alguns valors (*missings*) a les taules de dades ambientals. Aquests són valors que no es coneixen o que es troben en blanc, on els valors de les concentracions dels compostos considerats no han estat o no han pogut ser analitzats. El tema dels valors absents o 'buits' és un tema que es va ja considerar dins de la quimiometria al voltant dels anys 90, utilitzant per exemple l'algorisme K-NN (*k-nearest neighbour*) per fer l'estimació d'aquests valors absents (Todeschini, 1990), però encara avui és un tema recurrent en molts treballs. Alguns mètodes per afrontar el problema dels valors absents, són els mètodes d'imputació simple o múltiple, que utilitzen algorismes iteratius (Arteaga, 2009; Walczak, 2001a,b).

En aquesta Tesi, quan el nombre de valors absents és petit i aquests valors estan repartits aleatòriament per les taules de dades, es fa una estimació

dels mateixos, i els valors absents s'han substituït per estimacions d'ells calculades. Per estimar els valors absents s'ha utilitzat la funció de MATLAB *mdcheck* disponible en la PLS 3.5 Toolbox i posteriors (Eigenvector research Ltd., Manson, WA, USA). Aquesta funció porta a terme la estimació dels valors absents reemplaçant-los en primer lloc per zeros i calculant un model PCA. A continuació, els zeros són reemplaçats pels nous valors predits pel model PCA i es torna a calcular un nou model PCA. El procés es repeteix fins que els valors estimats per PCA dels valors absents convergeixen a uns mateixos valors. Una indicació del possible nombre de components adequats per a fer l'estimació dels valors absents és comprovar que les seves estimacions no donin valors negatius per sobre ajust (tot ajustant el soroll) de les dades. En la funció *mdcheck* es poden canviar diferents opcions per ajustar el resultat obtingut. Entre aquestes opcions es troben la quantitat màxima de valors absents que pot haver presents, el nombre de components principals, el criteri de convergència utilitzat i el algorisme utilitzat en la substitució. En el nostre cas, les condicions que van donar millors resultats van ser quantitats relatives màximes de dades perdudes entre 0.5 i 0.9, cinc components principals com a màxim, 0.1 com a criteri de convergència i l'algorisme PCA *nipals*.

Un cop feta l'estimació dels valors absents, s'ha de comprovar que les estimacions realitzades no són discordants i si aquestes són plausibles. En el cas que un valor absent, després d'haver provat diferents opcions per a la seva substitució, correspongués a un valor sobresortint o discordant (per exemple un valor negatiu), la seva estimació i predicció s'hauria de fer amb uns altres paràmetres. Quan el nombre de valors absents té una disposició sistemàtica o per blocs dins de la matriu de dades, com en el cas de files senceres sense dades en les taules de dades, l'estimació dels buits no és possible ni recomanable. En els casos de valors absents en files senceres, s'han reemplaçat aquests valors absents per les mitjanes corresponents a cada columna de les taules de dades. Per blocs sencers de valors absents s'ha treballat amb una estratègia de modelització que permet treballar amb aquests valors absents sense la necessitat d'estimar-los (Apartat 4.4.2, Aplicació del mètode MCR-ALS a dades incompletes).

Pretractament de conjunts de dades amb valors censurats

S'anomenen valors censurats aquells que només es coneixen que estan per sobre o per sota d'un determinat valor, però no es coneix el seu valor exacte. En química analítica, el valor censurat més habitual és el valor censurat per l'esquerra, és a dir, aquell que es troba fora de l'interval de mesura d'un instrument, és a dir per sota del seu límit de detecció. Per resoldre aquest problema hi ha diverses opcions, entre les quals es troba l'estimació estadística, la substitució per zeros, la substitució pel límit de detecció o la substitució per la meitat del límit de detecció, però l'elecció d'una opció o una altra dependrà de cada cas concret (Jain, 2008). Per exemple, en un estudi amb dades geoquímiques subterrànies reals i simulades es va demostrar que una opció bona era la de substitució del valor censurat per la meitat el valor del límit de detecció, sempre que no hi hagués més d'un 30% de valors censurats (Farnham, 2002).

Preprocessament de dades

Degut als avenços recents en química analítica, hi ha hagut un gran increment en la quantitat de dades que es produeixen i de la seva complexitat. El preprocessament és un primer pas important en l'anàlisi de dades, ja que es transformen unes dades on s'han eliminat variacions que no són d'interès, com per exemple, possibles artefactes experimentals o instrumentals (soroll i variacions sistemàtiques no relacionades amb l'objecte de l'anàlisi). Si el preprocessament no es fa de la manera adequada, es poden introduir variacions i efectes no desitjats, i per tant fer un bon preprocessament de dades és un pas crític que influirà en els resultats obtinguts i en la seva interpretació. Normalment, el preprocessament de les dades segueix diferents passos, i cadascun d'aquests passos eliminarà un artefacte particular, de manera que l'ordre com s'apliquin pot influir el resultat final. L'elecció òptima del mètode o de la combinació de mètodes de preprocessament, depèn de les

propietats de les dades i de l'objectiu de l'anàlisi. No existeix doncs, un únic mètode de pretractament de dades que pugui ser generalitzat a totes les situacions possibles de dades ambientals.

Hi ha mètodes de preprocessament molt diferents (Zeaiter, 2010), com la normalització (per reduir la influència de variació no desitjada entre observacions), l'escalat (per eliminar les possibles diferències d'escala entre els valors de diferents variables), el centrat (per eliminar o ometre la part constant de les variables), el suavitzat (per fer les dades menys abruptes), la correcció de la línia base (per eliminar desplaçaments constants o pendents) o l'alineament de pics i deformació o *warping* (si es tenen diferències de sincronització temporal en els cromatogrames o canvis de forma) entre d'altres. Aquests mètodes es poden aplicar a dades en dues o més direccions, i hi ha mètodes de pretractament que s'han desenvolupat especialment per dades multi-way (Engel, 2013), com per exemple el *Parametric Time Warping* (PTW, Eilers, 2004), que és un mètode d'alineació global o el *Correlation Optimized Warping* (COW, Skov, 2006) i el *icoshift* (Savorani, 2010), que s'apliquen per alineaments locals a vectors de dades (en una direcció). Aquests últims també s'han adaptat per alinear dades en les dos direccions de LC-MS i GC-MS al mateix moment (Christin, 2008; Bloemberg, 2010).

A continuació s'explicaran els pretractaments de centrat i escalat, entre els quals l'escalat ha estat molt utilitzat en aquesta Tesi. Aquestes operacions es poden realitzar segons les files o segons les columnes de la matriu de dades. Els mètodes emprats més freqüentment en aquesta Tesi, han estat els realitzats segons la direcció de les columnes de la matriu de dades, és a dir segons les variables o paràmetres mesurats. Aquests pretractaments es poden aplicar a dades en dues o tres direccions, els últims dels quals poden ser més complexos perquè es poden emprar i combinar de diferents maneres.

Centrat de les dades experimentals

En el centrat de les dades experimentals, a cada valor original d'una variable determinada se li resta el valor de la seva mitjana (de tots els valors mesurats de la variable considerada). D'aquesta manera, s'aconsegueix que cada variable de la nova matriu de dades quedi amb una mitjana igual a zero.

El centrat amb la mitjana de les dades realitza una translació de l'origen de coordenades de les dades des de zero fins al valor de la mitjana de les dades. El centrat és útil per visualitzar les variacions respecte al valor mitjà, ometent la informació d'aquelles variables que no varien i dels valors constants o *offset*. En alguns casos pot produir també efectes no desitjats, com per exemple la generació de valors negatius de les variables. El centrat es pot veure com una projecció de les dades en un espai on els valors constants o *offsets* s'han eliminat. Els errors també es trobaran centrats perquè també seran projectats en aquest espai. Quan hi ha presència de dades perdudes, centrar traient la mitjana de les files o les columnes no és adequat per eliminar els valors constants o *offsets*. En aquests casos els *offsets* s'han d'eliminar simultàniament amb l'ajust de la part bilineal (Bro, 2003).

Per dades en dues direccions, el centrat es pot fer en una direcció (la de les files), en l'altra direcció (la de les columnes) o en les dues de manera seqüencial. Si es fa en una sola direcció rep el nom de centrat simple, i si es fa en dues direccions de manera seqüencial, llavors rep el nom de centrat doble. En el cas del centrat doble el resultat serà el mateix independentment de per quina direcció comencem en centrat, i tant la mitjana de les columnes com la de les files serà zero.

Per dades en tres direccions o més, les formes de procedir són les mateixes que en el centrat en dues direccions, però en aquest cas els *offsets* poden estar en tres direccions diferents. Si es vol fer un centrat en una direcció d'un cub de dades (tres direccions), es pot desplegar aquest cub per tenir unes dades en dues direccions, i llavors aquestes dades es poden tractar com s'ha explicat anteriorment.

Escalat de les dades experimentals

En les dades escalades, cada valor original de la variable es divideix per un factor de escala, que habitualment és la seva desviació estàndard (de tots els valors mesurats d'aquesta variable). Les variables escalades presenten una distribució més homogènia que les fa més fàcilment comparables entre si, però

l'escalat no canvia l'estructura del model. L'escalat és especialment important en aquells casos on les variables estan mesurades en escales diferents o en unitats diferents, ja que un model de mínims quadrats es basa en una descripció de la variació, implicant que variables amb elevada variació són importants. La desviació estàndard de les variables un cop escalades és, per totes, igual a 1, de manera que s'iguali la seva importància relativa. No obstant, cal anar en compte en el cas de tenir variables en les que la majoria dels seus valors siguin molt baixos i constants, ja que aleshores el seu escalat pot augmentar de manera artificial la seva importància relativa portant a resultats i conclusions errònies. També és pot important com es porta a terme l'escalat, per no introduir una estructura artificial a les dades. Es poden escalar les files o les columnes, però si es vol escalar tant les files com les columnes, llavors la situació es complica perquè escalar en un direcció afecta a l'escala de l'altra direcció. En alguns casos es pot utilitzar l'escalat iteratiu, però no funciona en totes les situacions (Harshman, 1984). Si volem aplicar l'escalat a dades en tres direccions, enlloc d'ajustar el model a les dades originals, s'ajusta el model a unes dades transformades habitualment per una matriu d'escalat diagonal en la direcció de les variables que volem escalar. Això significa que totes les matrius s'han d'escalar pel mateix valor en l'anàlisi en tres direccions.

Autoescalat de les dades experimentals

En les dades autoescalades s'aplica simultàniament el centrat i l'escalat de les variables. És a dir, a cada valor original de la variable se li resta el valor de la mitjana dels valors d'aquesta variable i després es divideix per la desviació estàndard dels valors d'aquesta variable. La distribució dels valors de les variables obtingudes en aplicar el tractament d'autoescalat és similar al cas de l'escalat, però al mateix temps les dades experimenten una translació del seu origen degut al centrat amb la mitjana de les dades, i la seva nova mitjana és igual a zero per totes les variables. S'eliminen d'aquesta manera els efectes constants o *offset* i s'igualen les variables a una mateixa escala de variació.

En aquest cas també és important en quina direcció s'apliquen els diferents pretractaments i en l'ordre en què s'efectuen. Per exemple, l'escalat

en una direcció altera el centrat en la mateixa direcció però no en les altres direccions (Ten Berge, 1989), però en canvi, el centrat en una direcció altera l'escalat en totes les direccions, excepte en alguns casos concrets. Un d'aquests casos especials és l'escalat de les columnes i el centrat de les files, ja que aquest centrat no altera l'escalat de les columnes. L'ordre dels preprocessaments s'ha d'escollir depenent dels propòsits de l'anàlisi. Per exemple, pels models de descomposició és important treure els *offsets*, pel que el centrat s'hauria de fer abans que l'escalat. En regressió, és important que els *scores* estiguin centrats, pel que en aquest cas és millor que el centrat es faci després de l'escalat (Smilde, 2004).

4.3 Mètodes quimiomètrics

Depenent de l'estructura que presentin les dades a analitzar, es podran aplicar diferents tipus de models i mètodes quimiomètrics. Per a dades organitzades en dues direccions (*two-way data*), com per exemple les taules o matrius de dades obtingudes en estudis de seguiment ambiental, es poden aplicar mètodes basats en models de descomposició bilineal de dades, que es poden descriure amb l'equació

$$\mathbf{d}_{ij} = \sum_{n=1}^N \mathbf{u}_{in} \mathbf{v}_{nj} + \mathbf{e}_{ij} \quad \text{Equació 1}$$

En aquesta equació, \mathbf{d}_{ij} fa referència al valor experimental mesurat per la mostra i del paràmetre o variable j . En aquesta descomposició bilineal \mathbf{u}_{in} i \mathbf{v}_{nj} són els factors de la descomposició (anomenats generalment *scores* i *loadings*) corresponents a la mostra i , variable j , pel component n . El nombre total de components de la descomposició és N i \mathbf{e}_{ij} és la contribució no explicada després de considerar aquests N components, per a la mostra i , paràmetre o variable j .

En forma matricial, l'equació anterior es pot expressar com:

$$\mathbf{D}=\mathbf{UV}^T+\mathbf{E}$$

Equació 2

On \mathbf{D} és la taula o matriu de totes les mesures realitzades per $i=1\dots I$ mostres (files de la matriu de dades \mathbf{D}) i $j=1\dots J$ (columnes de la matriu de dades \mathbf{D}). Les dimensions per tant d'aquesta matriu de dades seran $\mathbf{D}(I,J)$. Aleshores la matriu de *scores* \mathbf{U} tindrà per dimensions $\mathbf{U}(I,N)$ i la matriu de *loadings* \mathbf{V}^T tindrà les dimensions $\mathbf{V}^T(N,J)$. $\mathbf{E}(I,J)$ serà la matriu residual que conté la variància no explicada pel model descrit a l'equació 2. Alguns dels mètodes que utilitzen aquest model bilineal de descomposició d'una matriu de dades són el mètode d'Anàlisi de Components Principals (PCA) i el mètode de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS).

El model bilineal mostrat en l'equació 1 es pot estendre a l'anàlisi simultani de múltiples matrius o taules de dades, que també es poden agrupar en tres direccions o cubs i paral·lelepípedes de dades (veure apartat 4.1 d'estructura de dades...). Aquesta extensió pot realitzar-se en alguns casos amb l'anomenat model trilineal, en el que la descomposició de les dades experimentals s'estén a les tres direccions de les dades tal com es mostra a continuació (equació 3)

$$d_{ijk} = \sum_{n=1}^N x_{in} y_{jn} z_{kn} + e_{ijk}$$

Equació 3

que en forma matricial es pot expressar com:

$$\mathbf{D}_k=\mathbf{XZ}_k\mathbf{Y}^T+\mathbf{E}_k.$$

Equació 4

En l'equació 3, d_{ijk} fa referència a la matriu, cub o paral·lelepípede de dades mesurades, amb i objectes, j variables i k condicions, i x_{in} i y_{jn} són les contribucions dels objectes i variables en cada situació particular o condició k . En el model trilineal, cada porció de les dades en tres direccions, està descomposat amb els mateixos x_{in} i y_{jn} i difereix només en les seves quantitats

relatives. En la notació matricial \mathbf{D}_k és la matriu de mesures en les diferents condicions, \mathbf{X} són els *loadings* en el mode 1, \mathbf{Y}^T són els *loadings* en el mode 2, \mathbf{Z}_k són els *loadings* en el tercer mode, i \mathbf{E}_k és el soroll en les diferents condicions no explicat pel model. Alguns mètodes que utilitzen el model trilineal de descomposició d'un cub de dades són PARAFAC, TUCKER3 o MCR-ALS, els quals es descriuran en les següents seccions.

En aquesta Tesi s'ha treballat principalment amb els mètodes PCA i MCR-ALS, i en alguns casos també amb PARAFAC i amb TUCKER3, els qual s'explicaran a continuació.

4.3.1 Anàlisi de Components Principals (PCA)

L'anàlisi de components principals (*Principal Component Analysis*, PCA) (Jolliffe, 2002; Wold, 1987) és un mètode d'estadística multivariant molt emprat en diferents àmbits científics i tecnològics, i en particular, també freqüentment utilitzat en l'anàlisi de taules de dades obtingudes en estudis de seguiment i monitoratge ambiental. El PCA és una tècnica de reducció de la complexitat de les dades, que projecta els objectes i variables en un espai de menors dimensions. Els gràfics d'aquestes projeccions es poden utilitzar amb propòsits exploratoris (Smilde, 2004). El PCA està basat en una descomposició bilineal del conjunt de dades originals segons les equacions 1 i 2. Les principals característiques de PCA és que aplica les restriccions d'ortogonalitat, màxima variància explicada i normalització dels *loadings*. El fet d'aplicar les restriccions d'ortogonalitat i màxima variància, permet donar una solució única, però per contra aquestes solucions poden no tenir significat físic.

El PCA parteix de la hipòtesi que en el grup de dades original existeixen un conjunt reduït de factors, components dominants (fonts de variació) o variables latents que influeixen de manera important sobre la variància observada en les dades. Generalment, una part important de la informació continguda en les variables mesurades experimentalment és redundant (correlacionada amb les altres variables) o bé irrellevant. El PCA permet

transformar aquest conjunt de variables en un nou conjunt de variables no correlacionades entre si, que és més senzill d'interpretar, fent més evident la informació rellevant i separar-la de la part que correspon al soroll o error experimental. Aquestes noves variables, factors o components principals són ortogonals i per tant el que explica un component, no ho explica l'altre. A més, en PCA, els components es troben ordenats segons la seva quantitat de variància explicada, de manera que el primer component explica més variància que el segon i així successivament. Per aquest motiu la informació més rellevant del conjunt de dades es troba sempre concentrada en els primers components principals generats, i els components posteriors podran ser omesos sense una pèrdua significativa de la variància o d'informació rellevant. Aquest fet és molt útil per a la interpretació de les fonts de variació experimental i de les seves característiques principals.

El mètode més utilitzat per a la descomposició de PCA és la descomposició en valors singulars (Singular Value Decomposition, SVD, Golub, 1996) ja que és un mètode acurat i ràpid. Amb els avenços informàtics d'avui en dia el càlcul d'un nombre elevat de components principals generalment ja no comporta cap problema de temps de càlcul. Independentment del mètode emprat per a calcular els components principals. L'aspecte més important en PCA és decidir quin és el número adequat de components o factors que descriuen la matriu \mathbf{D} . Una primera forma de veure quants components s'han d'incloure a l'anàlisi, és mitjançant l'observació de la mida dels valors propis associats a aquests components, ja que aquests indiquen la magnitud o quantitat de variància retinguda per cada component considerat. Tant els valors propis com la magnitud de variància retinguda per cada component disminueix en afegir més components fins a un nivell petit i pràcticament constant. El nombre de components N s'escull fins que l'addició d'un nou component ja no proporciona informació addicional rellevant dins del context del problema estudiat.

PCA està dissenyat per a proporcionar una millor visualització de la variància de les dades. Les representacions amb PCA permeten veure els possibles agrupaments (clústers) de les dades, identificar els *outliers* i trobar les causes de variació que hi ha darrere el que s'observa experimentalment. La

representació de les variables en els nous eixos definits pels *loadings*, dóna informació sobre la correlació entre les variables i sobre quines variables són les més importants i afecten més al model. La representació de les mostres en el nou espai definit pels nous components o factors dóna els *scores*. Aquestes representacions permeten observar com s'agrupen les mostres segons la seva similitud i detectar la presència de valors aberrants o extrems. Així doncs, la representació dels *scores* genera el mapa de les mostres i la representació dels *loadings* genera el mapa de les variables en el nou espai vectorial definit pels components principals. En el cas de dades ambientals, a partir dels *loadings* s'identifiquen les fonts de contaminació, i a partir dels *scores* se'n dedueix la seva distribució geogràfica i temporal sobre les mostres. L'anàlisi de PCA ajuda a l'exploració i interpretació de dades multivariants, aplicant restriccions d'ortogonalitat i de màxima variància, però els resultats de PCA no sempre són fàcilment interpretables en termes ambientals.

Una decisió important que s'ha de prendre abans de fer el PCA és si cal utilitzar les dades originals o cal fer un pretractament de les dades com els que hem explicat prèviament a la secció 4.2. Si una variable té una variància molt més gran que les altres, llavors aquesta variable pot condicionar molt l'anàlisi (veure mètodes de pretractament de les dades).

4.3.2 Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS)

La resolució multivariant de corbes (*Multivariate Curve Resolution*, MCR) descriu una família de procediments quimiomètrics utilitzats amb la finalitat d'identificar i resoldre les contribucions existents en un conjunt de dades. L'ús del MCR s'ha estès a diversos àmbits de la química, com l'espectroscòpia d'infraroig (Ruckebush, 2003), la cromatografia (Gargallo, 1996), la ressonància magnètica nuclear (Jaumot, 2004), l'anàlisi d'imatges hiperespectrals (Piqueras, 2011), la química ambiental (Terrado, 2009), la voltametria (Esteban, 2000), etc. El mètode MCR permet resoldre problemes relacionats amb la identificació i determinació analítica de compostos objectiu en presència

d'interferències desconegudes, a partir de l'anomenada petjada dactilar (*fingerprint*) dels constituents d'una mostra complexa (de Juan, 2013). També s'ha utilitzat per la determinació de biodièsel en barreges de biodièsel i dièsel convencional amb GC/GC-FID (Mogollon, 2013) o en la fusió de dades espectroscòpiques i cromatogràfiques per a la interpretació de processos de degradació (Mas, 2011). Altres aplicacions són per exemple, l'estudi de la fotodegradació cinètica de la melatonina (de Luca, 2013) o la resolució de cromatogrames obtinguts en cromatografia de gasos de dues dimensions (GCxGC-TOFMS) de mostres complexes com extractes de petroli (Parastar, 2013).

El mètode MCR es basa en una descomposició bilineal equivalent a la del mètode PCA (equació 1). El PCA utilitza restriccions d'ortogonalitat, de màxima variància explicada i de normalització dels *loadings*, i dóna solucions úniques, però aquestes solucions no són directament interpretables des d'un punt de vista físic, ja que els perfils (vectors) que s'obtenen pels *scores* i *loadings* per cada component són en realitat una combinació lineal dels perfils (espectrals, instrumentals, de concentració, elució,...) dels components que causen realment la variància observada. En MCR la finalitat és recuperar aquests perfils i poder fer la interpretació dels resultats directament. MCR utilitza restriccions més naturals que PCA, com per exemple la no-negativitat dels perfils espectrals i de concentració, enlloc de la seva ortogonalitat. El preu que s'ha de pagar no obstant per aquesta aplicació de restriccions més naturals i per tenir solucions més interpretables, és que les solucions no són úniques i existeix una certa ambigüïtat associada a elles (Tauler, 1995b).

La resolució en els mètodes MCR es pot fer de manera iterativa o no iterativa. En els mètodes no iteratius s'utilitzen les idees de selectivitat i aniquilació de rang en mètodes com *Rank Annihilation Evolving Factor Analysis* (RAEFA, Gampp, 1987), *Window Factor Analysis* (WFA, Malinowski, 1992) i *Heuristic Evolving Latent Projections* (HELP, Kvalheim, 1992). Entre els mètodes iteratius destaquen els procediments iteratius de resolució per mínims quadrats alternats. Aquests mètodes de resolució iteratius resolen el model bilineal a través d'una optimització iterativa per mínims quadrats alternats de U

i/o V^T . Aquests mètodes de resolució per iteracions comencen el procés d'optimització amb unes estimacions inicials que són posteriorment modificades de manera iterativa sota l'acció de diferents restriccions. Les restriccions forcen als factors U i/o V^T a obeir unes propietats químiques o matemàtiques prèviament definides. Les restriccions constitueixen, al mateix temps, una manera d'introduir informació matemàtica o química que es coneix prèviament, per tal d'assegurar l'obtenció de solucions més interpretables i pròximes a les reals. Entre aquests mètodes iteratius, destaca el mètode de resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS, Tauler, 1995a,b). Aquest mètode utilitza una implementació específica de mínims quadrats alternats (ALS) que és la que s'ha utilitzat en aquesta Tesi.

El mètode de MCR-ALS descomposa la matriu de dades D mitjançant un algorisme de mínims quadrats alternats, que minimitza la següent equació pels *scores* U i *loadings* V^T desconeguts sota restriccions adjents

$$\min_{U \text{ or } V^T} \| \hat{D}_{\text{PCA}} - \hat{U} \hat{V}^T \| \quad \text{Equació 5}$$

on \hat{D}_{PCA} és la matriu de dades reproduïda (filtrada o projectada) per PCA pel nombre de components considerat. \hat{U} i \hat{V}^T són respectivament les estimacions dels factors (*scores* i *loadings* en PCA) matriu U i V^T obtinguts durant la optimització per ALS. Aquestes dues matrius obtingudes per MCR-ALS reben noms diferents depenen del context d'aplicació, per exemple U es pot referir a perfils de variació de concentració amb el temps, el pH, d'elució cromatogràfica, etc. i V^T poden ser els perfils espectrals, instrumentals, de resposta, etc. En el cas de dades de seguiment ambiental U es pot referir a perfils de distribució de les fonts de contaminació i V^T pot contenir els perfils de composició química d'aquestes fonts de contaminació (Terrado, 2009).

En les cas de no aplicar restriccions (*constraints*) durant el procés d'optimització iteratiu (ALS), les solucions de l'equació 5 es descriuen amb les equacions 6 i 7 respectivament

$$\hat{\mathbf{U}} = \hat{\mathbf{D}}_{\text{PCA}} (\hat{\mathbf{V}}^{\text{T}})^+ \quad \text{Equació 6}$$

i

$$\hat{\mathbf{V}}^{\text{T}} = (\hat{\mathbf{U}})^+ \hat{\mathbf{D}}_{\text{PCA}}, \quad \text{Equació 7}$$

on $(\hat{\mathbf{V}}^{\text{T}})^+$ i $(\hat{\mathbf{U}})^+$ són les estimacions de les pseudoinverses de les matrius \mathbf{V}^{T} i \mathbf{U} (Golub, 1996). En l'algorisme de MCR-ALS, aquestes dues equacions es resolten sota un conjunt de restriccions pels perfils de la matriu en els factors \mathbf{U} i \mathbf{V}^{T} , com la no-negativitat, i altres que s'explicaran més endavant (de Juan, 2003; Tauler, 1995 a i b; Jaumot, 2005). En aquest cas el problema es pot resoldre rigorosament amb l'algorisme d'optimització de mínims quadrats ràpids no-negatius *fnnls* (*fast non-negative least squares*) (Bro, 1997a) expressat com a les equacions 8 i 9.

$$\hat{\mathbf{U}} = \text{fnnls}(\hat{\mathbf{D}}_{\text{PCA}}, \hat{\mathbf{V}}^{\text{T}}) \quad \text{Equació 8}$$

$$\hat{\mathbf{V}}^{\text{T}} = \text{fnnls}(\hat{\mathbf{D}}_{\text{PCA}}, \hat{\mathbf{U}}^{\text{T}}) \quad \text{Equació 9}$$

Altres restriccions que es poden aplicar són la unimodalitat, la de sistema tancat o de balanç de matèria, la de selectivitat i rang local (Tauler, 1995b), i en el cas de l'anàlisi de dades multidireccionals, la de trilinearitat (Tauler, 1995b, 1998) o la interacció entre components (Peré-Trepat, 2007; Article 2 de la present Tesi). Aquestes restriccions es poden aplicar de manera flexible i independent per cada component, i separatament si és necessari. L'èxit de l'aplicació del mètode MCR-ALS es troba precisament en la selecció i aplicació adequada d'aquestes restriccions. Una evolució recent del mètode MCR-ALS ha estat l'avaluació de l'extensió de les ambigüitats rotacionals de les solucions obtingudes per MCR-ALS amb l'algorisme MCR-BANDS (Tauler 2001, Jaumot, 2010) i els efectes de considerar les incerteses en les dades experimentals (Wentzell, 2006; Dadashi, 2012 i 2013) per evitar la seva propagació en els paràmetres del model (en aquest cas \mathbf{U} i \mathbf{V}^{T}), utilitzant per exemple l'aproximació de mínims quadrats alternats sospesats (MCR-WALS) o de forma

anàloga, fent la projecció inicial de les dades utilitzant un algorisme de màxima versemblança d'anàlisi de components principals (MLPCA (Wentzell, 1997), enlloc de l'algorisme clàssic de PCA.

En tots els mètodes de MCR, un primer pas és proposar un nombre inicial de components, N , que contribueix a la variància de les dades. Aquest nombre de components es pot escollir amb PCA. En el mètode MCR, les variàncies dels components individuals es solapen i els perfils resolts no són ortogonals com són en PCA (Jolliffe, 2002; Vandeginste, 1998). La suma de les variàncies explicades per cada component individual de MCR és habitualment més gran que la variància explicada simultàniament en el model conjunt. Això no ocorre en canvi en PCA, on aquesta suma és igual a la variància del model conjunt. Els perfils de composició i de distribució medioambientals rarament són ortogonals i es solapen, i per això no són directament interpretables a partir de l'aplicació de PCA.

L'algorisme MCR-ALS també necessita de les estimacions inicials d'una de les dues matrius \mathbf{U} o \mathbf{V}^T per tal de començar el procés iteratiu d'optimització. Existeixen múltiples possibilitats que permeten generar bones estimacions inicials i disminuir els temps computacionals i minimitzar els problemes de convergència. Un dels mètodes quimiomètrics més emprats per obtenir estimacions inicials d'un conjunt de dades és el mètode SIMPLISMA (*SIMPL*e-*to-use Iterative Self-Modeling Analysis*), que selecciona les columnes o files més pures (selectives) de la matriu \mathbf{D} (Windig, 1992 i 2005) com a estimacions respectivament dels perfils dels components en les matrius \mathbf{V}^T o \mathbf{U} respectivament. El càlcul de les estimacions inicials en MCR-ALS està moltes vegades inspirat en aquest mètode.

L'avaluació de la qualitat del modelatge de les dades en MCR-ALS es fa utilitzant dos paràmetres, que són el percentatge de la falta d'ajust (*lack of fit, lof %*) i les variàncies explicades ($R^2\%$)

$$lof\% = 100 \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m e_{i,j}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{i,j}^2}}, \quad e_{i,j} = x_{i,j} - \hat{x}_{i,j} \quad \text{Equació 10}$$

$$R^2\% = 100 \frac{\sum_{i=1}^n \sum_{j=1}^m x_{i,j}^2 - \sum_{i=1}^n \sum_{j=1}^m e_{i,j}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{i,j}^2} \quad \text{Equació 11}$$

L'avaluació de la qualitat dels perfils dels components recuperats per MCR-ALS només és possible estrictament quan els perfils són prèviament coneguts com en el cas de dades simulades preparades per posar a prova un procediment proposat. Els paràmetres utilitzats en aquest cas seran, per una banda el coeficient de correlació entre els perfils coneguts (\mathbf{x} i \mathbf{y}) i els estimats (r^2) i per l'altra, l'angle (α) entre ells calculat com el valor del cosinus que és igual a aquest valor de r^2 .

$$r^2 = \cos\alpha = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad \text{Equació 12}$$

$$\alpha = \arccos(r^2) \quad \text{Equació 13}$$

Restriccions

Les restriccions són propietats químiques o matemàtiques generals que els perfils en les matrius de factors \mathbf{U} i/o \mathbf{V}^T han de complir, les quals actuen dirigint el procediment d'optimització iteratiu cap a una solució amb significat físic i interpretable en termes ambientals. La utilització adequada de restriccions és una condició imprescindible en tots els procediments de resolució multivariant que permet disminuir l'ambigüitat de les solucions finals obtingudes. Les restriccions poden ser classificades d'acord amb la seva naturalesa o bé d'acord amb la seva utilització. Una restricció o condició pot ser imposada de dues maneres: forçant el perfil (espectral, de concentració, ...) d'un component determinat (o alguns dels seus elements) a ser igual a determinats valors preestablerts (restriccions d'igualtat, *equality constraints*), o

bé forçar-lo a ser superior o inferior a ells (restriccions de desigualtat, *inequality constraints*).

Entre les diferents restriccions aplicables del mètode MCR-ALS per dades ambientals destaquen la no negativitat i la normalització dels *loadings* (\mathbf{V}^T), i en alguns casos la trilinearitat. Totes tres es descriuran breument a continuació. Hi ha altres restriccions que són utilitzades en altres camps de la química analítica, com són la unimodalitat, la de sistema tancat (balanç de matèria) o les de selectivitat o rang local (Tauler, 1995b i 2001). Algunes restriccions com la trilinearitat i la de selectivitat/rang local, poden eliminar totalment l'ambigüitat rotacional associades a un o a tots els perfils resolts dels components. És important, a l'hora d'aplicar qualsevol restricció, considerar que aquesta estigui plenament justificada tant des d'un punt de vista matemàtic com des d'un punt de vista químic ambiental. El fet d'utilitzar una restricció inadequada, pot donar lloc a resultats erronis i pitjors que si aquesta no s'apliqués. El procediment recomanat és el d'anar introduint les restriccions de manera gradual durant el procés de resolució, i anar comprovant l'efecte que té la seva introducció a partir de la millora de l'ajust obtinguda. Quan l'aplicació de les restriccions porta a solucions que no són raonables des d'un punt de vista fisicoquímic o ambiental, caldrà comprovar si la seva aplicació ha estat encertada o és massa estricta, o també si cal ampliar els nivells de tolerància per evitar el seu incompliment.

a) No-negativitat

La restricció de no-negativitat és probablement la restricció més emprada en els mètodes MCR. En el cas de dades ambientals, la restricció de no negativitat es pot aplicar sobre els perfils de composició (\mathbf{V}^T) i contribució/distribució (\mathbf{U}) de les fonts de contaminació, ja que en aquests casos només tenen sentit solucions positives. Si a un factor se li aplica la restricció de no negativitat, els seus elements només poden ser positius (o zero). Hi ha diferents formes d'aplicar aquesta restricció de no negativitat, com ara la substitució directa dels valors negatius per zero durant els diferents passos iteratius o mètodes més rigorosos, basats en procediments per mínims quadrats no negatius (*nonnegative least squares*, nnls, Lawson, 1974) i les

seves variants més ràpides, com per exemple *fast non-negative least squares*, *fnnls* (Bro, 1997a). En aquesta Tesi s'ha utilitzat l'algorisme basat en aquest darrer mètode de mínims quadrats no negatius ràpids (*fnnls*).

b) Normalització dels loadings

La normalització es realitza per evitar les ambigüitats d'escala (Tauler, 1995b i 2001). A conseqüència d'aquesta ambigüitat, els valors dels *scores* poden incrementar-se arbitràriament de manera il·limitada multiplicant-los per un escalar arbitrari i al mateix temps els valors dels *loadings* corresponents al mateix component es poden fer decreïxer per la mateixa quantitat dividint-los pel mateix valor escalar, de manera que no es canvi el resultat de la seva multiplicació. Quan es fa la normalització dels *loadings* es divideixen tots els valors del perfil d'un component determinat per la seva norma (arrel quadrada de la suma dels quadrats de tots els elements del perfil), forçant per tant que els perfils de tots els components tinguin la mateixa norma igual a la unitat o a un valor constant.

MCR-ALS aplicat a dades en tres direccions (three-way)

Quan la matriu de dades presenta una estructura de dades més complexa com a conseqüència de l'anàlisi simultània de diverses taules o matrius de dades, com per exemple en el cas de quan es realitza diverses campanyes de seguiment ambiental en diferents punts de presa de mostra i durant diferents dies, mesos o anys, es poden agrupar aquestes dades en tres direccions (*ways*) o modes (*modes*). En MCR-ALS, es pot treballar en dades en tres direccions mitjançant la utilització de matrius de dades augmentades, com s'ha explicat anteriorment. La descomposició d'una matriu de dades augmentada es porta a terme amb un model bilineal anàleg al que s'utilitza en el cas de l'anàlisi d'una matriu de dades individual, però ara els resultats també es podran mostrar com en els models trilineals mostrant els *loadings* en els tres modes o direccions **X**, **Y** i **Z** (Figura 4.5).

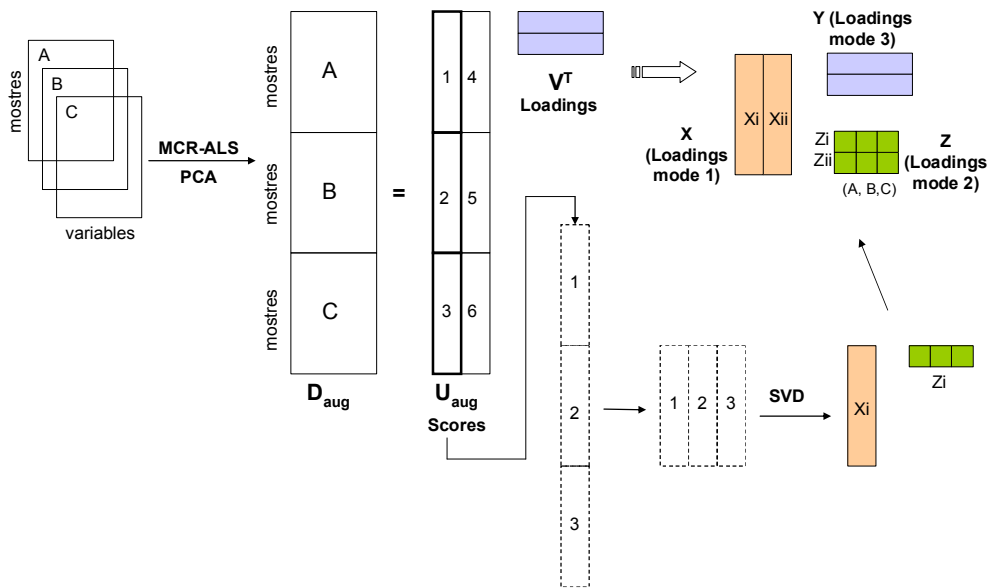


Figura 4.5. Descomposició bilineal per PCA o MCR-ALS d'una matriu augmentada seguida del replegament dels scores per recuperar els perfils dels *loadings* en els tres modes.

Quan es realitza l'anàlisi de dades en tres direccions o tres modes, ja no s'utilitza la denominació de *scores* i *loadings*, sinó només de *loadings* en el primer, segon i tercer modes. La denominació d'*scores* s'utilitza preferiblement només en el cas de dades en dues direccions (*two-way data*). La possible obtenció de *loadings* en tres modes independentment de si s'ha aplicat la restricció de trilinearitat o no, permet simplificar i generar un nous resultats similars als que s'obtenen quan s'apliquen mètodes basats en models trilineals, permetent la comparació amb altres mètodes multidireccionals (*multiway*). La única restricció per obtenir els perfils en els tres modes es que el tamany de les matrius de dades individuals analitzades simultàniament tinguin el mateix nombre de files i columnes, és adir una estructura de dades en tres direccions.

A la Figura 4.5 es mostra el procediment emprat per obtenir els perfils en els tres modes. Cada component de la matriu augmentada U_{aug} es pot replegar en una nova matriu segon els seus dos modes que pot ser descomposada en els seus valors singulars. Considerant el primer valor singular, es generen dos vectors, els qual són una estimació dels *loadings* en el primer i tercer mode. El procés es repeteix per cadascun dels components. La nova matriu de *loadings*

en el primer mode (\mathbf{X}) (Figura 4.5) descriu com els components resolts (fonts/pautes de contaminació) es distribueixen entre les mostres i la nova matriu de *loadings* en el tercer mode (\mathbf{Z}) descriu com són les diferències entre les diferents matrius de dades, és a dir entre les diferents campanyes de seguiment ambiental, per exemple en diferents dies, mesos o anys. Els *loadings* en el segon mode (\mathbf{Y}) (Figura 4.5) en canvi tenen el mateix significat que els obtinguts en la descomposició bilineal (\mathbf{V}^T), és a dir donen els perfils de composició de les fonts/pautes de contaminació.

S'ha de diferenciar entre l'aplicació d'un model bilineal i representar els resultats en tres direccions o modes, que és el que s'ha mostrat anteriorment, o aplicar un model trilineal i representar també els resultats en tres direccions o modes, que és el que s'explicarà en el següent apartat. Encara que el sistema es comporti de manera bilineal, pot ser útil representar-lo en la forma trilineal (amb tres modes), ja que d'aquesta manera podrem comparar els resultats obtinguts amb els obtinguts amb altres mètodes basats en els models trilineals com el PARAFAC (Bro, 1997b) o el TUCKER3 (Smilde, 2004). Representar els resultats en tres modes no vol dir ja que el sistema sigui trilineal (la majoria de processos químics i ambientals no són trilineals), però pot ser útil representar dades no trilineals en els tres modes com a procediment exploratori de dades, de la mateixa manera que es poden aplicar mètodes bilineals com el PCA per aproximar i visualitzar sobretot el comportament de sistemes de dades no-bilineals.

Aplicació de la restricció de trilinearitat al mètode MCR-ALS

Aquesta restricció està relacionada amb l'anàlisi simultània de conjunts o matrius múltiples de dades. L'aplicació d'aquesta restricció implica l'assumpció de que les dades tenen una estructura tridireccional (*three-way*) o en tres modes (veure apartat 4.1 de l'estructura de les dades), i que es poden descriure adequadament amb un model trilineal (Equació 3). Per analitzar dades en tres direccions per MCR-ALS, es construeix primer la matriu augmentada corresponent, formada per la concatenació dels diferents blocs o

matrius de dades (dades en dos direccions o *two-way*) en la direcció d'interès, generalment en la direcció d'augmentació segons les columnes (veure 4.1) L'aplicació d'aquesta restricció en MCR-ALS per dades que compleixen aquestes característiques força els perfils dels *scores* resolts de cada component corresponents a cada una de les diferents matrius constituents de la matriu augmentada a tenir la mateixa forma (Tauler, 1998). Aquesta restricció es podrà doncs aplicar quan les diferents matrius analitzades tenen el mateix nombre de files i de columnes. El model trilineal es pot també resoldre amb el mètode anomenat de PARAFAC (apartat 4.4.3). En MCR-ALS, aquesta restricció es pot aplicar de forma flexible per a tots els components o a només per aquells components que compleixin aquesta restricció. En general, les dades químiques compleixen el model trilineal només en alguns casos particulars, com per exemple en les dades d'emissió/excitació de fluorescència molecular, on la forma de l'espectre d'emissió d'un determinat compost químic és la mateixa per a qualsevol longitud d'ona d'excitació. També es pot trobar en algun casos de dades ambientals, com per exemple quan els perfils de variació dels components en els dos primers modes (mostres i variables) es mantenen invariants en el tercer mode, que pot ser per exemple el temps (diàriament, mensualment, anualment,...).

La restricció de model trilineal total o parcial s'aplica durant l'optimització iterativa (Tauler, 1998; Article 2 de la present Tesi) dins del procediment de mínims quadrats alternats (ALS). A la Figura 4.6 es pot veure gràficament com es porta a terme aquesta implementació. Quan s'aplica la restricció de trilinearitat, s'afegeix un pas a l'optimització ALS en el que cada perfil individual (columna) de la matriu augmentada de *scores* \mathbf{U}_{aug} és forçat a satisfer la condició de trilinearitat. El mateix procediment utilitzat i explicat anteriorment per recuperar els *loadings* en els dos modes (primer i tercer modes) a partir dels *scores* augmentats del model bilineal s'aplica ara durant la optimització per ALS.

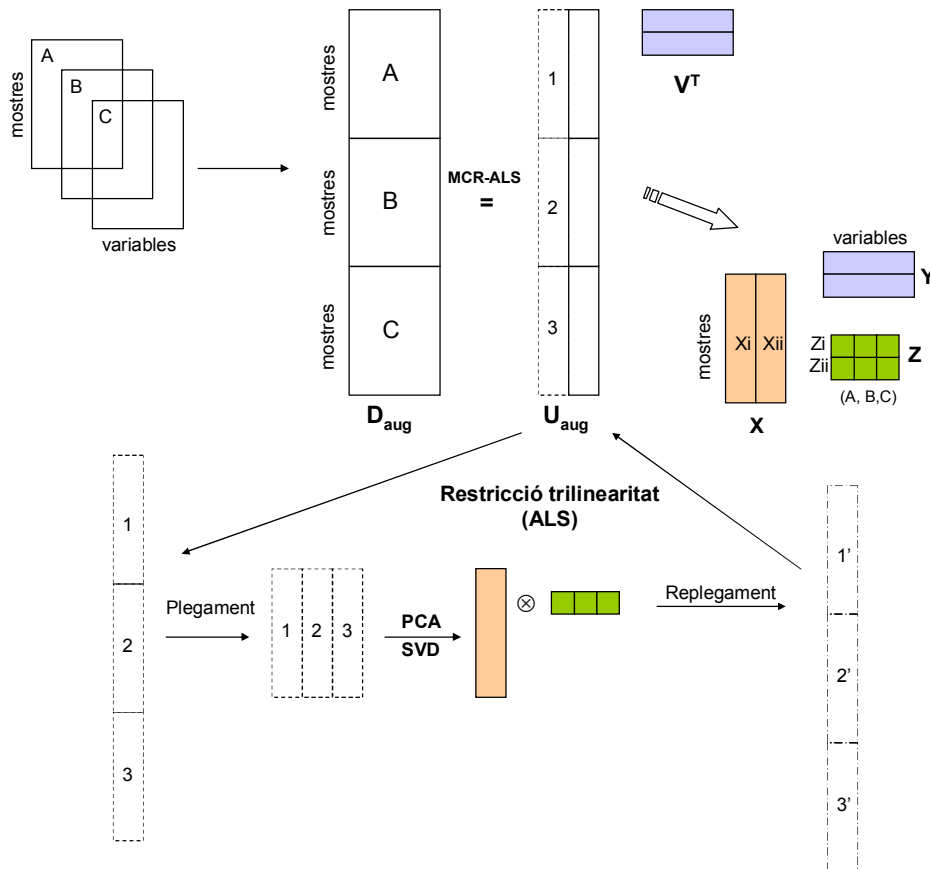


Figura 4.6. Descomposició per MCR-ALS d'una matriu augmentada amb restricció de trilinearitat.

Com es mostra a la Figura 4.6, en cada iteració ALS, cada columna de la matriu augmentada de *scores* U_{aug} és apropiadament plegada per donar lloc a una nova matriu amb les dimensions del primer i tercer modes, és a dir per exemple amb un número de files igual al número de mostres i un nombre de columnes igual al nombre de variables temporals (dies, mesos, anys...). El primer component de la descomposició en valors singulars d'aquesta matriu donarà els *loadings* en el primer i tercer mode respectivament per cada component considerat tal com es mostra a la Figura 4.6. El producte de kronecker (Smilde, 2004) d'aquests dos nous vectors *loadings* dona lloc aleshores a un nou vector augmentat de *scores*, que compleix amb la condició de trilinearitat i que substitueix la columna corresponent de la matriu de *scores* U_{aug} . Quan aquesta restricció és inserida durant cada pas del procés iteratiu d'optimització ALS, aquesta restricció força a que la forma del vector *loadings* en el primer mode sigui sempre igual. La variació d'intensitat o escala estarà

expressada en els *loadings* del tercer mode (\mathbf{Z}). Els resultats obtinguts amb l'aplicació d'aquesta restricció en MCR-ALS seran equivalents amb els obtinguts amb el mètode PARAFAC, que s'explicarà més avall.

Una de les principals avantatges de l'aplicació de la restricció de trilinealitat en el mètode MCR-ALS en front el PARAFAC, és que en MCR la restricció de trilinearitat s'aplica de manera independent per a cada component i que no és necessari aplicar-la de manera simultània a tots els perfils resolts. Aquesta restricció és diferent de l'aplicada en PARAFAC, on tots els components resolts han de satisfer forçosament la condició de trilinearitat.

Aplicació de la restricció de interacció entre components al mètode MCR-ALS

Aquesta restricció està relacionada amb l'anterior i també s'aplica al cas de dades tridireccionals (*three-way*). A diferència del cas del model trilineal estricte, en aquest cas es permet tenir un nombre diferent de components per cada un dels tres modes de l'estructura de dades tridireccionals (cub de dades). Aquest tipus de comportament es pot resoldre amb els models anomenats Tucker (Tucker, 1966), els quals també es descriuran més avall. L'aplicació de la restricció d'interacció entre components permet resoldre casos senzills de models TUCKER3 restringits (Smilde, 2004). En MCR-ALS s'aplica quan diferents components presenten un mateix perfil en algun dels modes, i permet reduir el nombre de components en aquell mode. D'aquesta manera es simplifiquen els resultats obtinguts sense perdre informació.

En el model PARAFAC trilineal, hi ha un perfil individual per cada component en cada mode. En canvi en MCR-ALS amb interacció de components (com també en el model TUCKER3), hi ha completa llibertat en la correspondència dels perfils entre els diferents components i modes, on per exemple, dos components poden interaccionar i tenir el mateix perfil en un dels modes.

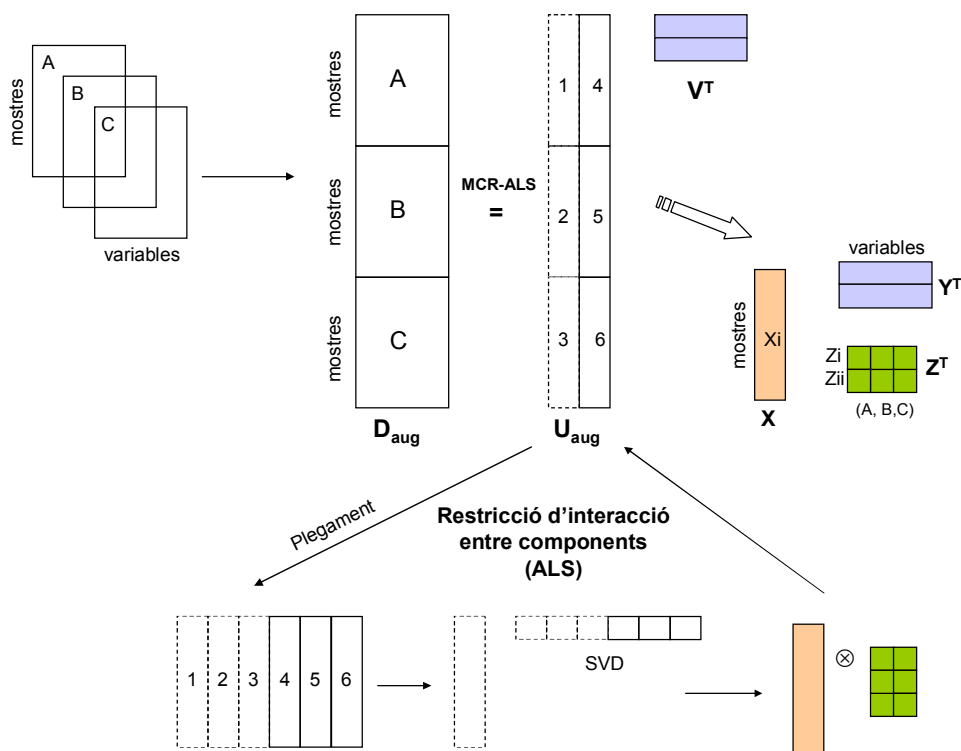


Figura 4.7. Descomposició per MCR-ALS d'una matriu augmentada amb restricció d'interacció entre components.

En la Figura 4.7 es mostra gràficament el funcionament d'aquesta restricció d'interacció entre components en el mètode MCR-ALS. En l'exemple que es mostra en la figura, dues columnes dels perfils de la matriu augmentada dels scores U_{aug} es pleguen per donar una nova matriu que es descomposarà per SVD, i només el primer component serà considerat. Mentre que en el cas d'aplicació de la restricció de trilinealitat, aquest plegament i descomposició només es realitzava per una de les columnes de la matriu U_{aug} , ara aquest procés es realitza per dues o més columnes de la matriu U_{aug} . Aquesta descomposició donarà directament un sol perfil del *loading* en el primer mode (X) i un perfil allargat del *loading* en el tercer mode (Z) que es pot desdoblar en dos. En aquest cas per tant s'obtidria un sol perfil en el primer mode i dos perfils en el segon (Y) i tercer modes. El producte de Kronecker dels vectors del primer i tercer mode donarà de nou la estimació dels nous perfils augmentats de scores en la matriu U_{aug} que compleixen amb la restricció imposada d'interacció entre components i que s'inserten com a noves estimacions d'ells durant la optimització iterativa ALS.

Aplicació del mètode MCR-ALS a dades incompletes

El mètode MCR-ALS s'utilitza habitualment amb blocs de dades complets que donen matrius augmentades en alguna de les dues direccions (files o columnes) en les que no hi falta cap de les matrius individuals corresponents. En aquest apartat, s'explica una nova variant del mètode MCR-ALS que permet analitzar matrius augmentades de dades incompletes, on algunes de les matrius individuals de dades (blocs) són absents.

En el cas estudiat en aquesta Tesi, les matrius augmentades de dades incompletes són dades sobre l'evolució temporal de dos contaminants diferents, però en les que els anys disponibles no eren sempre els mateixos per als dos contaminants. Com es pot veure a la Figura 4.8, per un contaminant es disposava de dades dels anys A, B, C i D, i per l'altre es disposava de dades dels anys C, D, E i F. Només hi havia dades comunes pels dos contaminants en dos dels sis anys investigats. Amb la nova versió del mètode MCR-ALS presentada en aquesta Tesi, es permet l'anàlisi conjunta dels dos contaminants de tots els anys disponibles, tant si eren comunes per els dos contaminants, com si només s'havien realitzat per a un d'ells. Per poder analitzar les matrius de dades incompletes es van substituir els blocs de dades absents per 'NaN' (Not a Number) en la notació utilitzada per MATLAB (The Mathworks, Natick, USA).

Tal com es mostra a la Figura 4.8, les dades incompletes es van estructurar en una matriu super-aumentada en les dues direccions o modes (augmentació simultània de files i columnes), que es poden subdividir en tres submatrius diferents: \mathbf{D}_1 és la submatriu super-aumentada en la direcció de les files i les columnes corresponents a les dades completes, és a dir amb les concentracions dels anys on tots dos contaminants varen ésser analitzats (C1, C2, D1 i D2), \mathbf{D}_2 és la submatriu augmentada en la direcció de les columnes amb les dades de concentració disponibles d'un contaminant (C1, D1, E i F), i \mathbf{D}_3 és la submatriu augmentada en la direcció de les columnes amb les dades de concentració disponibles per l'altre contaminant (A, B, C2 i D2).

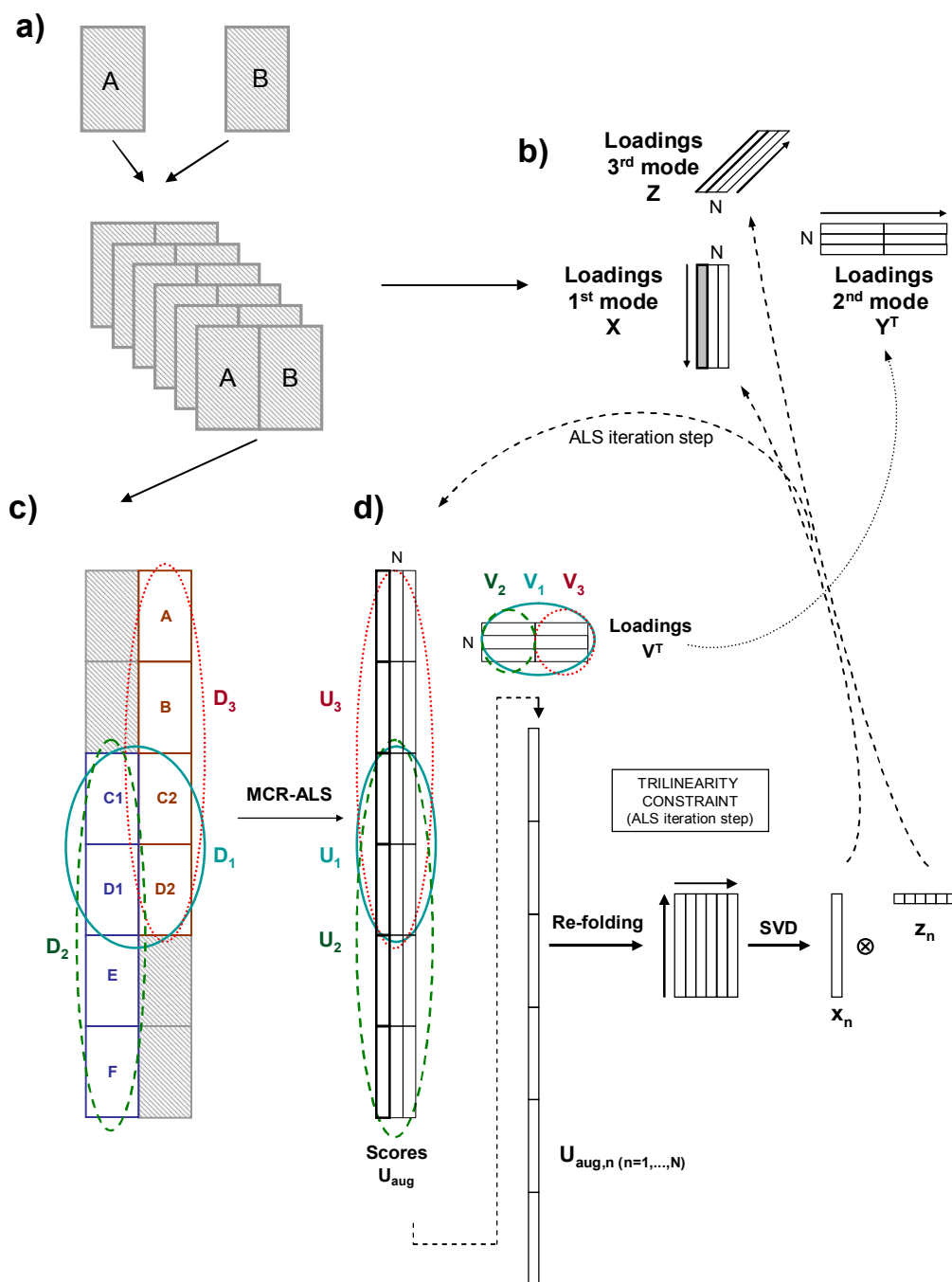


Figura 4.8. Estructura de les dades incompletes i la seva descomposició mitjançant MCR-ALS utilitzant la restricció de trilinearitat.

La descomposició bilinear de les tres matrius de dades es fa simultàniament utilitzant l'algorisme de mínims quadrats alternats ALS adaptat al càlcul de tres funcions d'error, corresponents a cadascuna de les submatrius esmentades \mathbf{D}_1 , \mathbf{D}_2 i \mathbf{D}_3 . En l'algorisme ALS proposat, aquestes tres funcions s'optimitzen simultàniament (sumant les tres funcions d'error individuals en una funció d'error total) per intentar uns valors conjunts òptims dels *scores* i *loadings* que permetin descriuen adequadament tots els anys analitzats. En aquesta descomposició bilinear dels *scores*, la matriu \mathbf{U}_1 té el mateix nombre de files que la matriu \mathbf{D}_1 , la matriu \mathbf{U}_2 té les mateixes files que \mathbf{D}_2 , i la matriu \mathbf{U}_3 té les mateixes files que \mathbf{D}_3 . En la descomposició bilinear de les tres matrius de dades, els *loadings* resultats \mathbf{V}_1 tenen les mateixes columnes que \mathbf{D}_1 , \mathbf{V}_2 té les mateixes columnes que \mathbf{D}_2 i \mathbf{V}_3 té les mateixes columnes que \mathbf{D}_3 (veure Figura 4.8). Com es pot veure en la Figura 4.8d, la implementació de l'algorisme ALS per conjunts de dades incompletes té aspectes diferents de l'algorisme general ALS utilitzat per conjunts de dades completes. En el càlcul de la matriu dels *scores* en el conjunt de dades incompletes es poden trobar tres situacions diferents (\mathbf{U}_1 , \mathbf{U}_2 i \mathbf{U}_3) i en cada cas el càlcul inclou la multiplicació de la matriu augmentada apropiada (\mathbf{D}_1 , \mathbf{D}_2 o \mathbf{D}_3) per la pseudoïnversa de la corresponent matriu de *loadings* (\mathbf{V}_1 , \mathbf{V}_2 o \mathbf{V}_3). En cada iteració de la optimització ALS global, les següents equacions del subconjunt ALS s'han de resoldre:

Per la matriu de dades \mathbf{D}_1

Càlcul del factor matriu \mathbf{U}_1 de l'estimació de $\hat{\mathbf{V}}_1^T$ per:

$$\min_{U_1, \text{restriccions}} \left\| \hat{\mathbf{D}}_{1,PCA} - \hat{\mathbf{U}}_1 \hat{\mathbf{V}}_1^T \right\| \hat{\mathbf{U}}_1 = \text{fnls}(\hat{\mathbf{D}}_{1,PCA}, \hat{\mathbf{V}}_1^T) \quad \text{Equació 14}$$

Càlcul del factor matriu \mathbf{V}_1^T de l'estimació de $\hat{\mathbf{U}}_1$ per:

$$\min_{\hat{\mathbf{V}}_1^T, \text{restriccions}} \left\| \hat{\mathbf{D}}_{1,PCA} - \hat{\mathbf{U}}_1 \hat{\mathbf{V}}_1^T \right\| \hat{\mathbf{V}}_1^T = \text{fnls}(\hat{\mathbf{D}}_{1,PCA}, \hat{\mathbf{U}}_1^T) \quad \text{Equació 15}$$

El càlcul de la funció d'error per mínims quadrats corresponent a la matriu de dades \mathbf{D}_1 es defineix a partir dels residuals calculats utilitzant les estimacions de $\hat{\mathbf{U}}_1$ i $\hat{\mathbf{V}}_1^T$ segons l'equació

$$ssq_1 = \sum_{i=1}^{I_1} \sum_{j=1}^{J_1} \left(d_{1,i,j}^2 - \sum_{n=1}^N \hat{u}_{1,i,n} \hat{v}_{1,j,n} \right) \quad \text{Equació 16}$$

De la mateixa manera, per a les altres matrius \mathbf{D}_2 i \mathbf{D}_3 es poden escriure les equacions corresponents i les seves funcions d'error per mínims quadrats corresponents, donant respectivament $\hat{\mathbf{U}}_2$, $\hat{\mathbf{V}}_2^T$, ssq_2 , $\hat{\mathbf{U}}_3$, $\hat{\mathbf{V}}_3^T$ i ssq_3 . La matriu final $\hat{\mathbf{U}}$ s'obté llavors a partir de la juxtaposició de les tres matrius d'scores obtingudes separatament

$$\hat{\mathbf{U}} = \{ \hat{\mathbf{U}}_1; \hat{\mathbf{U}}_2; \hat{\mathbf{U}}_3 \} \quad \text{Equació 17}$$

Quan hi ha coincidència en les files, la matriu $\hat{\mathbf{U}}_1$ prevaleix sobre $\hat{\mathbf{U}}_2$ o $\hat{\mathbf{U}}_3$ perquè \mathbf{D}_1 és la matriu conté la informació que en principi és més fiable, ja que s'ha obtingut a partir de l'anàlisi de les dades dels dos contaminants, és a dir amb la part conjunta completa de les dades (veure Figura 4.8c). Per altra banda, la matriu final de *loadings* és calcula a partir de l'expressió:

$$\hat{\mathbf{V}}^T = 1/2 \left(\hat{\mathbf{V}}_1^T + \left[\hat{\mathbf{V}}_2^T, \hat{\mathbf{V}}_3^T \right] \right) \quad \text{Equació 18}$$

On $\left[\hat{\mathbf{V}}_2^T, \hat{\mathbf{V}}_3^T \right]$ és el factor matriu augmentat en la direcció de les files, format per la concatenació segons les files de $\hat{\mathbf{V}}_2^T$ i $\hat{\mathbf{V}}_3^T$, tenint en compte la màxima informació (no només la de \mathbf{D}_1 , sinó també la de les altres dues matrius \mathbf{D}_2 i \mathbf{D}_3).

La funció que finalment és optimitzada per ALS és

$$ssq = ssq_1 + ssq_2 + ssq_3 \quad \text{Equació 19}$$

$$\sigma = \sqrt{\frac{ssq}{(I_1 J_1 + I_2 J_2 + I_3 J_3)}} \quad \text{Equació 20}$$

on I_1 , J_1 , I_2 , J_2 , I_3 i J_3 són respectivament el nombre de files i columnes de les tres matrius augmentades \mathbf{D}_1 , \mathbf{D}_2 i \mathbf{D}_3 .

En l'equació 20, sigma σ , dona l'arrel quadrada mitjana de la suma dels quadrats residuals (ssq). El criteri de convergència de l'optimització es produeix quan la millora en els canvis dels valors de sigma entre iteracions és per sota d'un determinat valor, normalment 0.1%. Aquest valor es pot modificar depenent del criteri de convergència de la optimització que es vulgui assolir. Normalment per propòsits exploratoris s'utilitza un valor més elevat (per exemple 1%), i quan ja s'ha trobat un bon model, es proven valors més baixos per si s'obté alguna millora de les solucions obtingudes.

En la Figura 4.8d també es pot veure com es podrien obtenir també els perfils en tres modes (\mathbf{X} , \mathbf{Y} i \mathbf{Z}), on ara el tercer mode correspondria als dos tipus de contaminant. Aquesta disposició permetria també (al menys en teoria) l'aplicació de la restricció de trilinearitat. La manera com s'aplicaria la restricció de trilinearitat a dades incompletes seria similar a la que s'ha explicat anteriorment. Això implica que el perfil d'*scores* de cada component corresponent a cada matriu individual constituent de la matriu augmentada és forçat a ser invariant (tenir la mateixa forma) i només difereixi en un factor d'escala. Finalment i com succeeix en els casos que ja s'ha comentat abans, es podran també representar els resultats obtinguts a partir dels *loadings* en les tres direccions o modes.

4.3.3 PARAFAC

El mètode PARAFAC (*Parallel Factor Analysis*) és basa en un model trilineal que va ser introduït per Harshman (1970) als anys 70 i també independentment per Carroll & Chang (1970) sota el nom de Canonical Decomposition (CANDECOMP). La idea bàsica del model trilineal és utilitzar

els mateixos factors per descriure la variació en diverses matrius simultànies encara que amb diferents coeficients de pes per cada matriu.

El model PARAFAC es pot introduir generalitzant la descomposició de valors singulars. Per unes dades $\mathbf{D}(I \times J \times K)$ en tres direccions (cub de dades) en el model trilineal es fixen dos conjunts de paràmetres (per exemple x i y), i un element del cub de dades, s'expressa com a funció lineal dels paràmetres restants (per exemple z).

$$d_{ijk} = \sum_{n=1}^N g_{nmn} x_{in} y_{jn} z_{kn} + e_{ij} \quad \text{Equació 21}$$

i en notació matricial com

$$\mathbf{D}_k = \mathbf{XZ}_k\mathbf{Y}^T + \mathbf{E}_k = z_{k1}\mathbf{x}_1\mathbf{y}_1^T + \dots + z_{kN}\mathbf{x}_N\mathbf{y}_N^T + \mathbf{E}_k \quad \text{Equació 22}$$

On \mathbf{Z}_k és la matriu diagonal amb les k files de \mathbf{Z} en la seva diagonal (elements z_{k1}, \dots, z_{kN}); \mathbf{x}_n , \mathbf{y}_n són les n columnes de \mathbf{X} i \mathbf{Y} , respectivament i el terme residual $\mathbf{E}_k (I \times J)$ és definit de manera similar a \mathbf{D}_k . Cada \mathbf{D}_k es modela utilitzant els mateixos components \mathbf{X} i \mathbf{Y} , però amb diferents pesos, representats per \mathbf{Z}_k . El model PARAFAC dóna unes estimacions úniques, és a dir, \mathbf{X} , \mathbf{Y} i \mathbf{Z} no es poden canviar sense canviar els residuals (no hi ha ambigüitat rotacional). En la Figura 4.9 es pot veure la descomposició del model PARAFAC.

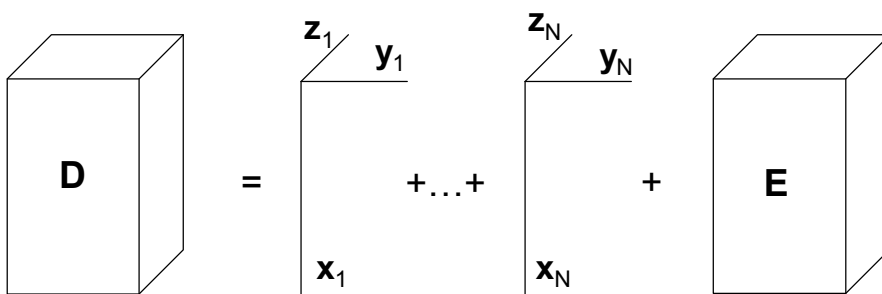


Figura 4.9. Descomposició del model PARAFAC amb N components.

L'anàlisi de PARAFAC d'un cub o paral.lelepípede de dades \mathbf{D} genera directament les matrius de *loadings* en els tres modes o direccions \mathbf{X} , \mathbf{Y} i \mathbf{Z} . La matriu \mathbf{X} (*loadings* en el primer mode) conté la contribució dels components en el primer mode (generalment les mostres). La matriu \mathbf{Y} (*loadings* en el segon

mode) conté la contribució dels components en el segon mode (generalment les variables o paràmetres mesurats sobre les mostres). La matriu **Z** (*loadings* en el tercer mode) conté la contribució dels components en el tercer mode (generalment les condicions diferents sota les quals s'han fet les mesures de les variables en les diferents mostres). En la descomposició realitzada per PARAFAC, el nombre de components en els tres modes és el mateix, N . Com ja s'ha comentat abans, les solucions obtingudes amb el model PARAFAC sota unes determinades restriccions són úniques i no presenten ambigüitats rotacionals (Smilde, 2004). La resolució del model trilineal per PARAFAC es pot realitzar amb diferent tipus de restriccions, com ara d'ortogonalitat o de no-negativitat.

Els paràmetres **X**, **Y** i **Z** es poden estimar amb diferents algorismes, i s'han d'estimar simultàniament quan aquests no són ortogonals i per tant no són independents entre ells. Els algorismes que s'utilitzen per ajustar el model PARAFAC es basen (com en el cas de MCR-ALS) generalment en els mínims quadrats alternats (ALS) (Smilde, 2004), i rebent llavors el model el nom de PARAFAC-ALS. El sistema ALS té els seus avantatges perquè és un algorisme simple d'implementar, és simple incorporar-li restriccions (com la ortogonalitat, la no-negativitat, la unimodalitat, etc. (Bro, 1997b; Harshman, 1970)), i garanteix la convergència. No obstant, a vegades és lent. L'elecció d'unes bones estimacions inicials pot accelerar l'algorisme PARAFAC-ALS i disminuir els risc de convergir en un mínim local. La descripció dels algorismes ALS emprats en el model PARAFAC surt de l'àmbit d'aquesta Tesi, pel que no s'explicarà aquí, però es pot trobar una descripció detallada dels mateixos al llibre *Multi-way analysis* (Smilde, 2004).

L'ús del PARAFAC va començar dins de l'àrea de la psicometria, com per exemple en l'estudi de les diferents freqüències de la veu segons la posició de la llengua o els llavis amb diagrames de raigs x (Ladefoged, 1978) o en l'estudi de les respostes emocionals produïdes per diferents estímuls (Cole, 1985), però el seu ús s'ha estès a molts àmbits de la ciència i de la química. En aquest darrer cas de la química, ha trobat especialment ús en l'anàlisi de dades d'espectrometria d'excitació-emissió (fluorescència). La seva utilització en altres àmbits com en la cromatografia multidimensional o en la cromatografia

hifenada amb detecció espectromètrica (UV-DA o MS) és més controvertida deguda a la necessitat del compliment del model trilineal. Exemples d'aplicació recent són la caracterització dels components del petroli en un vessament al golf de Mèxic al 2010 (Zhou, 2013) o la caracterització de la matèria orgànica dissolta en aigua en una conca subtropical en diferents règims hidrològics (Hong, 2012).

4.3.4 TUCKER3

El mètode Tucker deu el seu nom a L. Tucker, que va introduir la descomposició d'un conjunt de dades agrupades en tres direccions en tres matrius de *loadings* i un bloc o matriu nucli (*core*) (Tucker, 1966). Més tard, l'any 1980 Kroonenberg i de Leeuw van crear un algorisme de càlcul basat en una optimització per mínims quadrats alternats (ALS, *alternating least squares*) sota restriccions d'ortogonalitat (Kroonenberg, 1980 i 1983).

El model TUCKER3 es pot explicar amb l'equació

$$d_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R x_{ip} y_{jq} z_{kr} g_{pqr} + e_{ijk} \quad \text{Equació 23}$$

on d_{ijk} representa l'element ijk en el conjunt de les dades; x_{ip} , y_{jq} i z_{kr} són els elements típics de les matrius de loadings \mathbf{X} ($I \times P$), \mathbf{Y} ($J \times Q$) i \mathbf{Z} ($K \times R$), g_{pqr} és l'element típic del que s'anomena nucli \mathbf{G} (cub de dimensions reduïdes $P \times Q \times R$), i e_{ijk} és el terme de residuals \mathbf{E} ($I \times J \times K$).

En la Figura 4.10 es pot veure la descomposició gràfica del model TUCKER3.

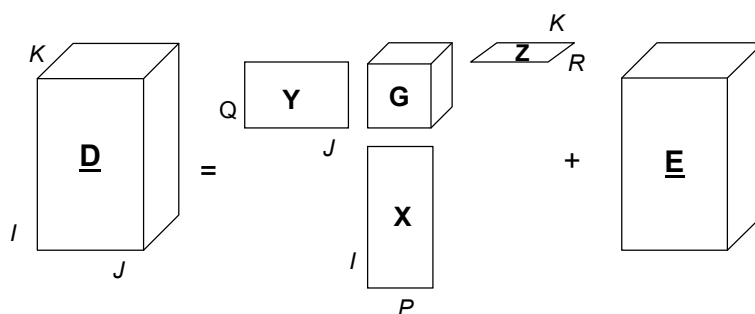


Figura 4.10. Descomposició gràfica de model TUCKER3.

En el model TUCKER3, la notació (P , Q , R) s'utilitza per indicar que el model té P , Q , R factors en els tres modes diferents. A diferència de l'equació 21, explicada anteriorment en PARAFAC, no hi ha cap requeriment del model per tenir el mateix nombre de components en X i Y . Un nucli no diagonal significa explícitament que en el model hi ha interaccions entre els factors, i aquesta és una propietat dels models Tucker en general. El model PARAFAC és idèntic a un model TUCKER3 on el nucli s'ha fixat per que tots els elements fora de la superdiagonal siguin zero, en canvi, el nucli d'un model TUCKER3 tindrà els elements fora de la diagonal diferents de zero. En l'anàlisi de components principals tradicional, els vectors de *loadings* només poden interaccionar entre ells si pertanyen al mateix component i en modes diferents. Per exemple, el segon vector (perfil) d'*scores* interacciona amb el segon vector (perfil) de *loadings* amb una magnitud definida pel segon valor singular. En el model TUCKER3 tots els vectors poden interaccionar entre si. Per exemple, el primer vector dels *loadings* pot interaccionar amb el tercer vector dels *loadings* amb una magnitud definida per l'element g_{13} . El model TUCKER3 té la versatilitat de permetre estudiar casos en els que el nombre de components no és el mateix en els diferents modes. En aquest sentit, el model TUCKER3 és flexible i permet modelar situacions en les que es produeixen interaccions entre els components (similar a la restricció d'interacció entre components que s'ha explicat anteriorment per MCR). En els casos on es produeix interacció entre components i el nombre de components no és el mateix en els diferents modes, el model rep el nom de TUCKER3 restringit, que és el que s'ha utilitzat en aquesta Tesi.

Per resoldre el model TUCKER3 es poden utilitzar diferents algorismes, però el més emprat és el TUCKALS3 que realitza l'ajust per mínims quadrats alternats amb vectors de *loadings* ortonormals (Kroonenberg, 1980). Més tard van aparèixer altres algorismes més eficients i algorismes que permetien models amb ordres més elevats i que utilitzen altres tipus de restriccions, com la no-negativitat (Andersson, 1998; Ten Berge, 1987; Morup, 2006). El mètode TUCKER3 s'ha utilitzat en diferents àrees com la cromatografia (de Ligny, 1984), l'anàlisi ambiental (Gemperline, 1992) o en l'anàlisi de la percepció personal (Van der Kloot, 1985). Més recentment s'ha utilitzat per exemple en l'avaluació de l'estabilitat química de diferents metalls en residus del tractament d'aigües residuals (Pardo, 2013) o en el modelatge i interpretació en dades ambiental en tres direccions adquirides en la presa de mostra de material particulat (PM₁₀) en quatre localitzacions en una regió d'Austria (Astel, 2010).

Capítol 5

RESULTATS I DISCUSSIÓ

5.1 BLOC 1 - Contaminants de les bases de dades de la Generalitat

En aquest bloc es mostren els resultats obtinguts en els estudis dels contaminants obtinguts a partir de les xarxes de seguiment de la contaminació atmosfèrica de la Generalitat de Catalunya. L'aplicació dels mètodes quimiomètrics com el PCA i el MCR-ALS han permès obtenir perfils (*scores*) d'informació geogràfica i temporal segons el tipus d'augmentació de matrius individuals. S'han aplicat diferents tipus de restriccions en el mètode MCR-ALS, per millorar la interpretació dels resultats, i s'ha desenvolupat una nova versió del citat mètode MCR-ALS que permet treballar amb dades multibloc incompletes.

Article 1 - M. Alier, M. Felipe-Sotelo, I. Hernández, R. Tauler; *Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods*. *Analytica Chimica Acta* 642 (2009) 77–88

En aquest treball es van analitzar dades de concentracions d'òxids de nitrogen obtingudes en diferents estacions de Catalunya durant el període comprès entre els anys 2001 i 2006. Per una banda es van analitzar dades de NO obtingudes simultàniament en diferents estacions i durant diversos anys. Els valors de concentració de NO obtinguts per cada any en diferents estacions de Catalunya, es van disposar en forma de vectors (columnes), posant els diferents vectors un al costat de l'altre per les diferents estacions i anys disponibles. En aquest estudi es va utilitzar un model bilineal i els resultats es van mostrar en dos modes. Per una altra banda, també es van analitzar dades de tres estacions concretes, Barcelona-Gràcia (anys 2001, 2002, 2005 i 2006), Bellver de Cerdanya (anys 2004, 2005 i 2006) i Tarragona-Sant Salvador (anys 2004, 2005 i 2006). En els tres casos es van analitzar les dades de NO utilitzant matrius augmentades en la direcció de les columnes pels diferents anys, emprant un model trilineal i mostrant els

resultats en tres modes. Cal destacar que la utilització de la restricció de model trilineal (trilinearitat) permet resumir de forma considerable la informació contenida en els resultats i facilitar-ne la seva interpretació.

Article 2 - Marta Alier, Mónica Felipe-Sotelo, Isabel Hernández, Romà Tauler; *Trilinearity and component interaction constraints in the Multivariate Curve Resolution investigation of NO and O₃ pollution in Barcelona*. Analytical Bioanalytical Chemistry (2011) 399:2015–2029

En aquest treball l'estudi es va centrar en l'estudi de només una estació de presa de mostres, a Barcelona en barri de Gràcia, però en la que es van analitzar de forma individual i conjuntament dades de concentració de dos compostos químics, de NO i O₃. L'anàlisi conjunta d'aquests dos compostos es va realitzar a partir de diferents tipus d'augmentacions, que van permetre a més l'aplicació de la restricció de trilinearitat i la nova restricció d'interacció entre components. Aquesta última restricció, igual que la de trilinearitat, permetia obtenir resultats més d'acord amb la situació real on s'han de resoldre un nombre de components diferents en els diferents modes de mesura. En tots els casos es van utilitzar models bilineals i trilineals i els resultats es van mostrar en tres modes. En particular es van obtenir models satisfactoris on el nombre de components que explicaven la variació anual, diària i interanual eren tres per l'estudi individual de NO i pels conjunts de NO i O₃, i dos per l'estudi del O₃.

Informació suplementària de l'Article 2 - Marta Alier, Mónica Felipe-Sotelo, Isabel Hernández, Romà Tauler; *Trilinearity and component interaction constraints in the Multivariate Curve Resolution investigation of NO and O₃ pollution in Barcelona*. Analytical Bioanalytical Chemistry, Electronic Supplementary Information.

En aquesta informació suplementària s'afegeixen quatre figures més dels models emprats i dels resultats en els tres modes obtinguts en cada cas.

Es mostren dues figures pels resultats de PCA, i dues figures més pels resultats de l'aplicació dels mètodes de PARAFAC-ALS i de TUCKER3-ALS restringit, respectivament, per les anàlisis realitzades independentment pel NO i pel O₃.

Article 3 - Marta Alier, Romà Tauler; *Multivariate Curve Resolution of incomplete data multisets*. Chemometrics and Intelligent Laboratory Systems 127 (2013) 17–28.

En aquest treball es va desenvolupar una nova versió del mètode MCR-ALS per treballar amb diferents blocs de dades incomplets. Aquestes dades incompletes provenien de l'estudi conjunt de NO i O₃ en l'estació de Barcelona-Gràcia, on les dades d'aquests contaminants no estaven disponibles per els mateixos anys, formant així blocs sencers sense dades. Per comprovar el funcionament d'aquesta nova versió del mètode MCR-ALS es van utilitzar tant dades reals com simulades. Aquestes dades simulades es van crear amb perfils de NO i O₃ coneguts i afegint diversos nivells d'error. En aquesta nova versió del mètode MCR-ALS també es van aplicar diferents restriccions com la no-negativitat i la trilinearitat, aplicant respectivament, models bilineals i trilineals. Els resultats de l'aplicació del mètode MCR-ALS s'han mostrat en tots els casos en els tres modes.



Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods

M. Alier^a, M. Felipe-Sotelo^{b,*}, I. Hernández^c, R. Tauler^a^a Department of Environmental Chemistry, IIQAB-CSIC, C/Jordi Girona Salgado, 18–26, 08034 Barcelona, Spain^b Department of Chemistry, ETSEA, University of Lleida, Av. Alcalde Rovira Roure, 177, 25198 Lleida, Spain^c Department of Environment, Generalitat de Catalunya, Av. Diagonal, 523–525, 08029 Barcelona, Spain

ARTICLE INFO

Article history:

Received 29 September 2008

Received in revised form 26 January 2009

Accepted 29 January 2009

Available online 11 February 2009

Keywords:

Chemometrics

Principal component analysis

Multivariate curve resolution

Nitric oxide pollution

ABSTRACT

Multivariate data analysis methods are applied to study of the geographical and temporal distribution of nitric oxide (NO) in Catalonia (North-East Spain), measured during the period 2001–2006 in 50 sampling stations. Principal component analysis (PCA) and Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) were applied for that purpose. The simultaneous analysis of NO data from sampling stations showed that its geographical distribution was rather uniform during the period considered. When three individual sampling stations were considered (two urban sites and one rural location), three different temporal patterns were resolved, with marked daily-night changes mainly attributed to traffic and also, important winter–summer seasonal variations. A decreasing trend in the levels of NO has been also observed in recent years. Comparison with nitrogen dioxide (NO₂) profiles shows that the daily variation is quite similar to the NO variation, however NO₂ displays very little oscillations along the seasons and no reduction of its concentration was observed in the last years, contrasting with NO tendencies.

The use of MCR-ALS is confirmed to be a useful method to improve interpretability in atmospheric contamination studies. The use of non-negativity and trilinearity constraints is shown to provide improved interpretations of the different contamination patterns in environmental terms.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Nitrogen oxides (NO_x), a mixture of nitric oxide (NO) and nitrogen dioxide (NO₂), are produced from natural sources, motor vehicles and other fuel combustion processes. Although nitric oxide does not significantly affect human health, there is a concern on the NO emissions and levels, since this is one of the main compounds involved in the formation of ground-level ozone [1], which can trigger serious respiratory problems. NO can react to form nitrate particles, acid aerosols, as well as NO₂, which also cause respiratory problems and it contributes to the formation of acid rain and atmospheric particles (see for instance <http://www.epa.gov/air/urbanair/nox/> [43] for further information). Spanish legislation on air quality establishes as 1-h limit for human health protection 200 μg m⁻³ NO₂ and NO_x and an alert threshold of 400 μg m⁻³ (both values expressed as NO₂) (RD 1073/2002 and RD 717/1987). Moreover, the odd oxides of nitrogen among other nitrogen gases such as NH₃, are considered indirect greenhouse gases. After several transformations in the atmosphere,

they return to the ground level in gaseous or liquid form and they can then undergo further transformations involving the formation and emission of the direct greenhouse gas nitrous oxide (N₂O) [2]. NO is involved in complex photochemical transformations and transport processes in the atmosphere once it is emitted, which means that problems associated with this pollutant are not confined to areas where it is emitted. Consequently, controlling NO_x is often most effective if it is done at a regional scale, rather than focussing on sources in only a particular local area. Better understanding of the regional scale of pollutants distribution and variation is necessary to design efficient regulatory programs, and it will provide more options [3]. The work presented by Kaiser et al. [4] on the monitoring of NO_x, CO and O₃ is a good example of how pollutants are transported from industrial and highly populated areas in continental Europe and can affect remote locations in the Alps. The transboundary nature of the atmospheric pollution generates a need for methodologies that tackle the problem from a global perspective.

Monitoring studies of environmental pollution have become more important in the recent years, from local to planetary scale, due to noticeable changes in human activities and their effects on public health and climate. Researchers make many efforts to find out the best way to analyse pollutants, to establish their spatial and temporal distributions, and to identify and characterise their

* Corresponding author. Tel.: +34 934006100; fax: +34 932045904.

E-mail addresses: mfsqam@cid.csic.es, mfelipe@quimica.udl.cat (M. Felipe-Sotelo).

sources. The final aim of these studies is to drag out useful conclusions for the reduction of emissions and design of minimization and remediation strategies. From the several methodologies available for the evaluation of environmental data, chemometric tools show powerful capabilities, because they allow to work with large scale data matrices, they do not usually require long computing times, and which also is very important, they allow rising useful conclusions, even in the case that the underlying chemical and physical processes are not completely understood or known [5]. Thus, the use of chemometric models and methods can give simpler and more approachable solutions.

Some works have applied multivariate approaches to the study of atmospheric pollution. Gao [6] has applied a functional analysis of variance (FANOVA) on O_3/NO_x measurements in order to characterise their diurnal cycles. Principal components analysis has been used quite often in atmospheric pollution research in order to establish seasonal variations [7], geographical distributions, dependence with meteorological conditions [8] and atmospheric chemical reactions. However, factors found by principal component analysis (PCA) do not often have straightforward or unique interpretations. In order to overcome this drawback of PCA when applied to atmospheric data, some authors have applied different type of factor rotations [9,10] to clarify the meaning of factors.

In the particular case of source apportionment studies, different techniques have been proposed for the analysis of the various factors contributing to contamination of the atmosphere, including principal component analysis [10], and alternative methods and tools such as: Unmix [11], Positive Matrix Factorization (PMF) [12], the Multilinear Engine (ME) [13], and Multivariate Curve Resolution Alternating Least Squares (MCR-ALS). In a previous work [14], ground-level ozone pollution was investigated over the same geographical area, and the application of the MCR-ALS method permitted the successful identification of three different sub-regions within Catalonia, where O_3 concentrations showed distinct spatial and temporal patterns. Different daily, weekly and annual patterns were obtained for each sub-region. The differences between these sub-regions could be partly explained considering the titration effect of NO_x emitted by car exhausts, apart from other local meteorological effects. Representation of MCR-ALS resolved profiles by GIS software helped to visualise more clearly these three sub-regions.

In the present work, the research is focussed on promoting and extending the use of chemometric tools, in particular of the MCR-ALS method, for the evaluation of atmospheric pollutants data. Non-negativity and trilinearity constrains are implemented for the resolutions of the MCR-ALS models. The geographical and temporal patterns of nitric oxide in Catalonia (NE, Spain) are investigated, covering the period from 2001 to 2006. For this purpose, PCA and MCR-ALS were applied to a large data set obtained in a continuous atmospheric monitoring network, and their results are presented and compared. Additionally, the NO data from three sampling stations were analysed individually; a highly populated urban location, medium size urban site and a rural site were selected to assess the sources of NO and their temporal profiles in each location. In some cases, the results obtained for NO are also compared with NO_2 and O_3 in order to achieve a better understanding of their mutual effects.

2. Data and methods

2.1. Geography, climate and sources of NO_x

Catalonia is located in the NE of Spain, it covers a total area of 32,000 km² and its population has recently overpassed more than 7 million inhabitants (the density of population is twice as much

as in the rest of Spain). Regarding geography, the mountain range of the Pyrenees in the North forms a long strip running 230 km from the Val d'Aran to the sea. The axial Pyrenees have steep sides and high peaks (about 3000 m high) as well as some lower areas. The Catalan Mediterranean System consists of range of mountains that run parallel to the coast and between them a plain area that holds the most densely populated regions. An extensive plain area lays at the eastern end on the River Ebre, and includes a series of basins formed by erosion, high plateaus and the alluvial plains that are largely used for growing crops. The sheer coastline consists of 580 km of steep cliffs interspersed with hidden or sandy coves depending on the proximity of the mountains. This morphological diversity leads to a wide variation in climate. The Pyrenees and nearby areas have climates typical of high mountainous regions, with minimum temperatures below 0 °C, annual rainfall of above 1000 mm and heavy snowfall in the winter. Coastal areas have mild, temperate weather, with temperatures that rise and rainfall that decreases as one moves from north to south. Inland, far from the sea, the climate is typical of continental Mediterranean regions, with cold winters and very hot summers. Further information on the geography and climate can be found in, <http://www.gencat.net/> [44] and <http://mediambient.gencat.cat> [45].

The main source of air pollution, and in particular NO_x , is traffic, although there are also a number of small industrial activities contributing to air pollution. It is remarkable the size of the fleet of private cars in Catalonia, which arrives to ca. 1,000,000 vehicles only in Barcelona. In addition to the traffic mobility within the community, it should be noted that Catalonia is the main gate for transport of goods from the Iberian Peninsula to the continental Europe (and vice versa). The location of this community, its growing population and active economy make this case of study of particular interest.

2.2. Experimental data

Catalonia has a network of atmospheric pollution control sampling stations operated by the Department of Environment of the Generalitat (Catalan Autonomous Government). NO concentrations are measured according to the method UNE 77212:1993, equivalent to ISO 7993:1985, performed by automatically operated chemiluminescence analysers. These analysers provide measurements every minute which are saved in half-hourly period. The measurements and the procedures are precisely established by Spanish legislation (Real Decreto 1073/2002). See <http://mediambient.gencat.cat> for further information about the Catalanian atmospheric monitoring network [45].

Half-hourly NO concentrations were collected from 1 January 2001 to 31 December 2006 in 50 automated sampling stations spread over Catalonia (see sample sites locations in Fig. 1), including different kinds of environments, from coastal to mountain sites, and from urban to agricultural areas. This database was provided by the Department of Environment of the Generalitat and it was transferred into MATLAB 6.5 (The Mathworks, Natick, USA) for subsequent multivariate data analysis.

2.3. Multivariate data analysis

In the present work, PCA and MCR-ALS are proposed to evaluate NO data at a regional scale. Among the multivariate methods available, PCA is one of the most popular tools for the evaluation of atmospheric data [8,10,15]. The PCA method is based on a bilinear orthogonal decomposition of the original data set. In matrix form, it can be represented as $\mathbf{D}_{(I \times J)} = \mathbf{U}_{(I \times N)} \mathbf{V}^T_{(N \times J)} + \mathbf{E}_{(I \times J)}$, where \mathbf{D} is the original data array, with I rows (observations) and J columns (sampling stations); \mathbf{U} is the matrix of orthogonal scores of dimensions $I \times N$, where N is the reduced number of components; \mathbf{V}^T is the matrix of orthonormal loadings with dimensions $N \times J$ and \mathbf{E}

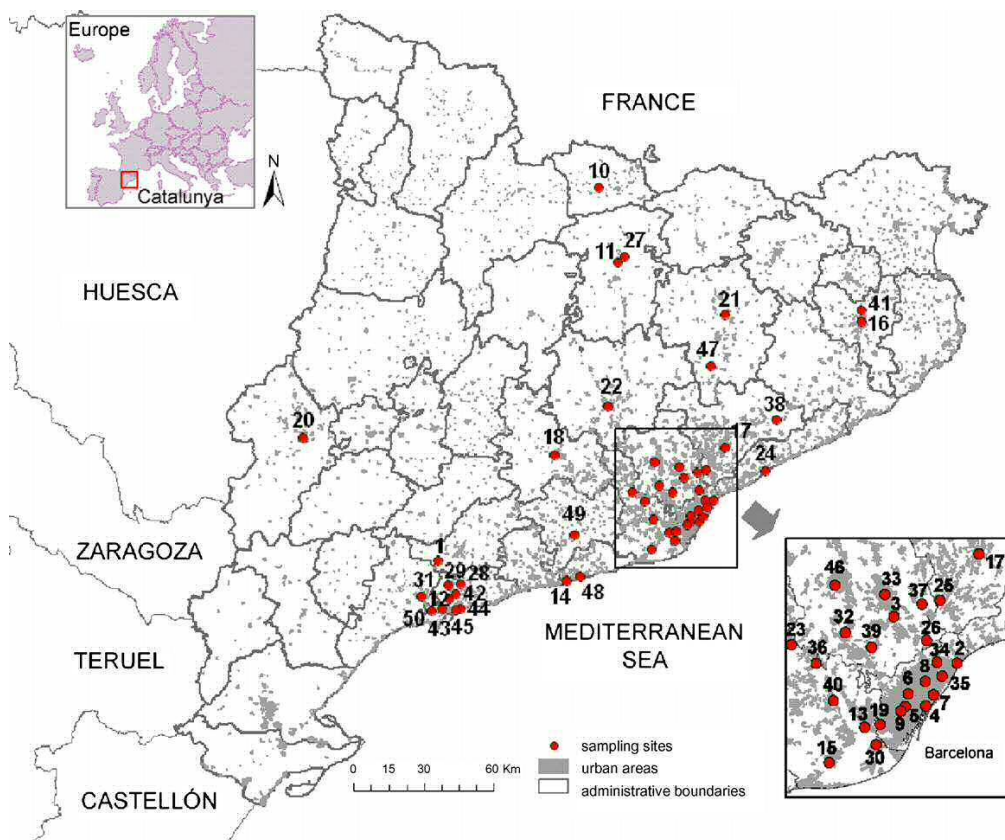


Fig. 1. Location of sampling stations.

is the matrix of residuals not modelled by the N principal components (see [16,17] for more details). The loadings (matrix \mathbf{V}^T) give the correlation among the variables in each PC, and therefore, they allow for the investigation of the common variability patterns of the original variables. PCA bilinear data decomposition is performed under orthogonal constraints, loadings normalisation and accounting for maximum explained variance for the successive extracted components. Under these constraints, solutions obtained by PCA are unique and they are identified without rotational or intensity ambiguities. Constraints like orthogonality and maximum explained variance, which are very helpful for data exploration and interpretation, may not be easily translated to environmental interpretation.

The use of MCR-ALS is proposed as a useful alternative to improve environmental interpretability. Although MCR-ALS has been usually applied for the resolution of multiple components in unresolved mixtures investigated by spectroscopic means [18], it has been also proved its ability to resolve and model average temporal and geographical distribution profiles from environmental monitoring studies. One of the main advantages of MCR-ALS is that it decomposes the data matrix by applying more natural constraints than PCA, such as non-negativity, and thus, environmental results can be interpreted more straightforwardly. The drawbacks of MCR-ALS are however that solutions may not be unique (due to rotation ambiguities), as they are in PCA, and caution should be taken to check the reliability of the obtained solutions [19,20], and another possible drawback is that MCR-ALS needs good \mathbf{U} or \mathbf{V}^T initial estimates. However MCR-ALS solutions are more easily interpretable and they resemble more the true variance sources.

The bilinear decomposition in MCR methods is also performed according to the same expression already given for PCA. The MCR-ALS method decomposes the same data matrix using an Alternating Least Squares algorithm, which in the case of not applying any constraints is described by the two equations $\mathbf{U} = \mathbf{D}(\mathbf{V}^T)^+$ and $\mathbf{V}^T = (\mathbf{U})^+ \mathbf{D}$, where $(\mathbf{V}^T)^+$ and $(\mathbf{U})^+$ are the current estimations of the pseudo-inverse of \mathbf{V}^T and \mathbf{U} [21]. These two equations are however in the MCR-ALS algorithm solved under a set of constraints over the factor matrices \mathbf{U} and \mathbf{V}^T , such as non-negativity, unimodality, closure, trilinearity or selectivity, as it is described elsewhere in more detail [18,22–24]. Recently a new approach has been proposed to take into account uncertainties in measured data [25] using an appropriate weighted least squares approach (MCR-WALS), however this approach was not been applied in the present study. The first step in the ALS optimisation is the determination of the number of components contributing to \mathbf{D} (this could be done by PCA). Following, initial estimates of either \mathbf{U} or \mathbf{V}^T must be found. In the case of \mathbf{U} , this can be done by evolving factor analysis. An estimation of \mathbf{V}^T can be given by the set of “purest” rows in \mathbf{D} (those rows which are more dissimilar to the rest). Given \mathbf{D} and \mathbf{U} , least squares and constraints are applied to calculate \mathbf{V}^T ($\mathbf{V}^T = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{D}$; $\mathbf{V}^T = \mathbf{U}^+ \mathbf{D}$). In the present work, non-negativity and trilinear constraints were applied [26]. In the next step, given \mathbf{D} and \mathbf{U}^T , least squares and constraints are applied to calculate \mathbf{U} ($\mathbf{U} = \mathbf{D} \mathbf{V}^T \mathbf{V}^T{}^{-1}$; $\mathbf{U} = \mathbf{D} (\mathbf{V}^T)^+$, and with the new estimate of \mathbf{U} , a new matrix \mathbf{V}^T can be calculated. Convergence is achieved when in two consecutive iterative cycles, relative differences in standard deviations of the residuals between experimental and ALS calculated data values were less than a previously selected threshold value, usually chosen as 0.1%

[27]. Complex data arrays ordered in more than two modes, ways or directions [28,29] can be arranged in an augmented two-way data matrix by concatenation of the individual data matrices. Resolved profiles using a bilinear model data decomposition can be then transformed afterwards to recover the profiles in each of the three modes [30]. Thus, one of the aims of this work is to extend the use of MCR-ALS to large environmental data bases and also to prove its suitability for this type of data and problems. In the present work, in addition to the non-negativity constraints used in previous studies of atmospheric pollutant [14], trilinearity is implemented to resolve the MCR-ALS model and the results compared.

PCA models were developed using MATLAB PLS 3.5 Toolbox (Eigenvector research Ltd., Manson, WA, USA), whereas MCR-ALS studies were performed using a software graphical interface [27], freely available on the web (<http://www.ub.es/gesq/mcr/mcr.htm> [46]).

In a first step, missing data were estimated by PCA imputation. The function implemented MATLAB PLS 3.5 Toolbox applies an iterative algorithm using as first estimates the mean value of the available data. This allows for developing a PCA model a re-estimation of the missing data. The convergence criteria as well as the maximum number of principal components and the desired fraction of variance captured by the PCA model must be set by the user. Stations with less than 90% annual data available, due to temporal malfunction of the automated measuring system were removed from this study.

To study the geographical and temporal variability of NO concentrations, multivariate data analysis was first performed for each year separately (matrices, D_{2001} ($17,520 \times 14$), D_{2002} ($17,520 \times 9$), D_{2003} ($17,520 \times 15$), D_{2004} ($17,568 \times 13$), D_{2005} ($17,520 \times 36$) and D_{2006} ($17,520 \times 39$) (the first number indicates the total number of half-hourly NO measures and the second number indicated the number of sampling stations monitored in that particular year) and also simultaneously for the 6-year period (note that the numbers of sampling stations available were different each year). In the later case, the data were arranged in a large row-wise augmented data matrix $D_{row-aug}$ of size ($17,520 \times 126$), where the columns corresponded to sampling stations in each year and the rows to the half-hourly NO concentration (see Fig. 2a). In the present study, column-wise data matrix augmentation was not possible, due to the lack of consistency in the distribution of the sampling sites throughout the different years.

PCA was applied then to the individual data matrices and also to the row-wise augmented data matrix $D_{row-aug}$. Afterwards, MCR-ALS models were also developed after column scaling, and applying non-negativity constraints. The new PCA or MCR loadings obtained after multivariate analysis (V^T) gave information about those sampling locations where NO followed a similar variation pattern throughout the 6 years under study. On the other hand, the scores matrix U gave temporal information on daily and seasonal variation profiles for each component. A proper refolding of scores can be performed to recover time-averaged scores for each component (Fig. 2b). For each augmented scores profile, a new scores matrix of size 48×365 is obtained, where the 365 columns correspond to the different days and the 48 rows to the scores for each half an hour of each day. When averaging was performed for the 48 row values of each column, a mean score value was obtained for every day, and thus the seasonal and annual variations for each individual component can be deduced. On the other side, when averages were calculated along the 365 column values of the 48 rows, the daily profiles of NO concentrations were obtained for every considered component. Alternatively, the singular value decomposition (SVD) can be applied to each of these score matrices of every component and allows the recovery of the main seasonal and daily profiles. Note that, as non-negativity constraints were applied, loading and score values are directly physically interpretable in terms of NO concentration.

In order to investigate NO sources and their temporal patterns more deeply, three sampling stations were investigated in detail over the different years; Barcelona-Gràcia (6), Bellver de Cerdanya (10) and Tarragona-Sant Salvador (42) (see locations in Fig. 1). They were selected according to the results obtained in a previous work [14]. As markedly different O_3 temporal profiles were observed for these locations, some distinctive NO patterns and/or sources are expected. Some studies in NO_2 were also carried out in these locations. For each of the sampling sites, NO (or alternatively, NO_2) data were arranged appropriately to build up the corresponding column-wise augmented data matrix D_{aug} , as shown in Fig. 3. During the ALS optimisation, every column of the augmented matrix of scores can be constrained independently to fulfil the trilinearity condition. Each of the MCR-ALS resolved components identified by this model are interpreted as a main pattern of NO_x concentration variation in the considered sampling site, differentiated by their particular emission temporal patterns. The loadings matrix, or loadings in

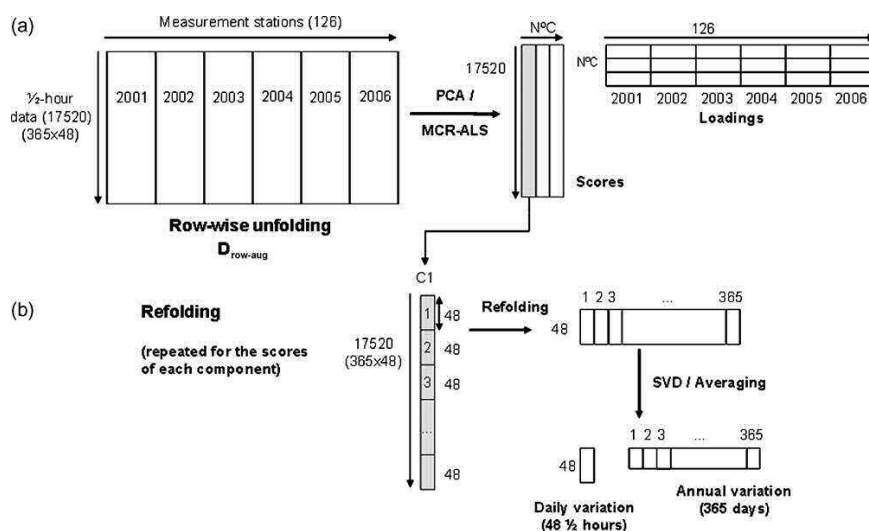


Fig. 2. Data arrangement. (a) Building of $D_{row-aug}$ by row-wise matrix augmentation, where columns correspond to each of the sampling sites available every year and rows correspond to 1/2-h NO concentration values. (b) Refolding and SVD analysis of the scores profile of every component to obtain the daily and annual variation trends.

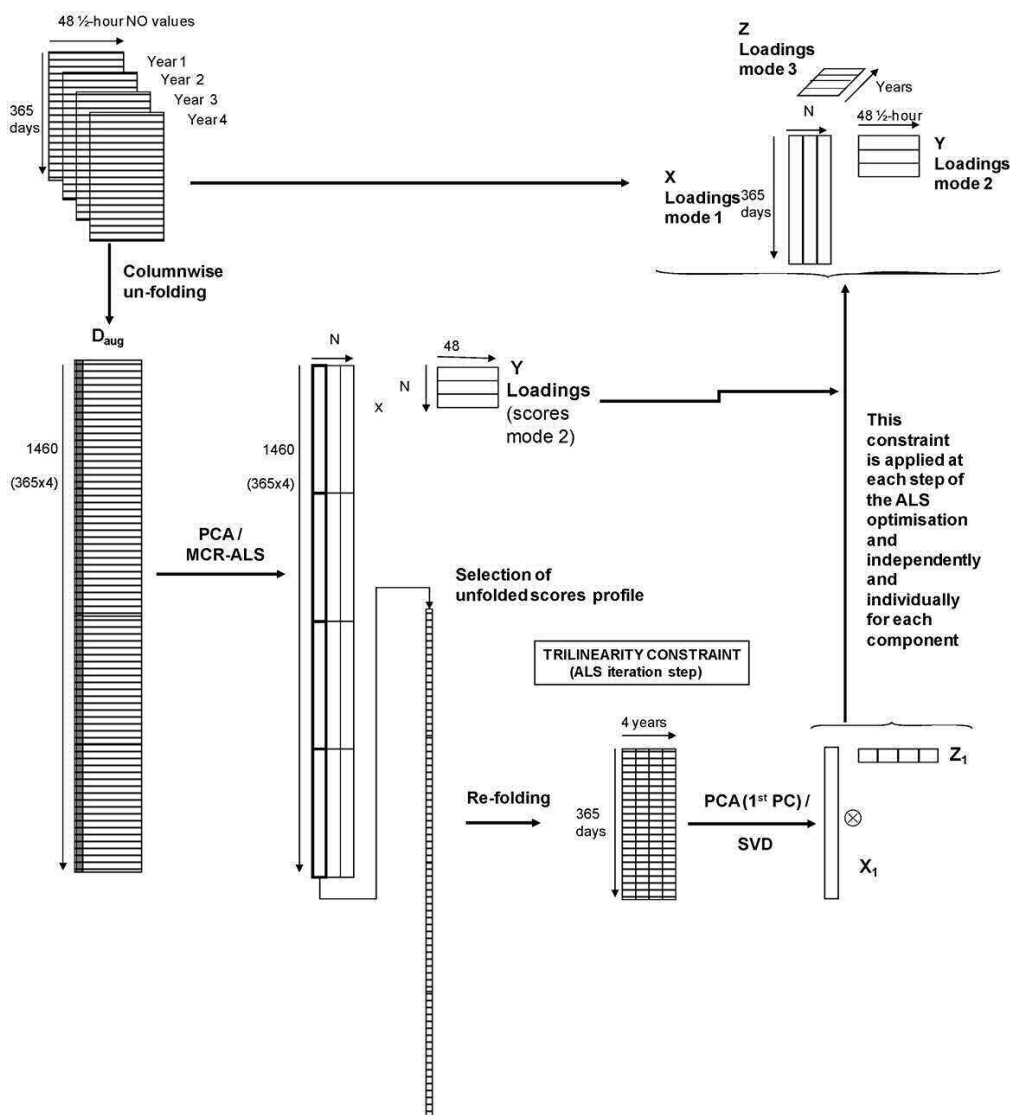


Fig. 3. MCR-ALS (or PCA) bilinear decomposition with trilinearity constraints of the column-wise augmented data matrix D_{aug} , corresponding to NO data from the individual station of Barcelona-Gràcia (years 2001, 2001, 2005 and 2006) (left) and further scores refolding and SVD analysis in order to recover the profiles in the three modes (right).

mode 2, Y^T in Fig. 3, gives the daily variation profile of each of the NO_x sources identified. In the scores MCR-ALS factor matrix, X , annual and inter-annual information are apparently intermixed, and consequently a new rearrangement is necessary to recover independently the temporal patterns. In Fig. 3 (right), it is shown how for each of the NO_x components, the scores column is selected, re-folded and decomposed by SVD. Using this simple procedure, the annual (vector-matrix X) and the inter-annual (vector-matrix Z) profiles are recovered for each NO_x resolved component.

It must be mentioned that, since O_3 data are available in addition to NO_x concentrations, column-wise augmented data matrices O_{aug} and N_{aug} could be built for ozone and NO_2 , respectively, analogue to D_{aug} for NO. In order to extend the proposed methodology to simultaneous analysis of NO, NO_2 and O_3 , the three matrices D_{aug} , O_{aug} and N_{aug} could be assembled to build a row-wise augmented matrix including values for the three variables. Results are not included for the sake of brevity.

3. Results and discussion

The application of some basic statistics to the whole data set is shown in Table 1 (see locations in Fig. 1), including annual averages, standard deviations and maximum 1/2-h concentrations. Annual averages ranged between 2.2 and $131 \mu\text{g m}^{-3}$. The lower year averages ($<5 \mu\text{g m}^{-3}$) corresponded to areas with moderate anthropogenic influence, like the stations placed in the south of Catalonia, around Tarragona – Alcover (1), Constantí (12) and Perafort (28) – and also some rural areas in the northeast – Bellver (10), La Nou (27) and Tona (47) – (Fig. 1). Spanish legislation establishes as annual threshold for human health protection an averaged concentration of $40 \mu\text{g m}^{-3}$ for NO_2 , which would correspond to $28 \mu\text{g m}^{-3}$ of NO (RD 1073/2002 and RD 717/1987). The higher values ($>50 \mu\text{g m}^{-3}$) were registered at some locations in the metropolitan area of Barcelona, locations 5, 6, 8, 25, 26, 33, 37 and 46 in Table 1 and Fig. 1. Regarding the variability of the measured NO values in one particu-

lar station and measured as standard deviations, it can be observed that they varied in a wide range from ± 4.3 to $\pm 342 \mu\text{g m}^{-3}$. Again, the sampling stations located in the metropolitan area of Barcelona showed the highest standard deviations, whilst the sites in rural areas (1) and around Tarragona (12 and 28) presented the lowest variability in NO.

3.1. Geographical and temporal variation of NO levels in Catalonia

In order to study the temporal and geographical distribution of NO concentrations in Catalonia, the data of all sampling stations during the period from 2001 to 2006 were analysed simultaneously by PCA and MCR-ALS. The data were arranged as shown in Fig. 2a in a row-wise augmented data matrix $\mathbf{D}_{\text{row-aug}}$. The other possible column-wise data matrix augmentation as in [14] was not possible in this case, due to the fact that sampling sites in the data base were not evenly distributed during the whole period (note in Table 1 that the number of sampling sites available is not the same for all the years).

PCA was carried out applying column autoscaling as data pre-treatment. Two components were chosen, which accounted only for ca. 33% of the variance (Table 2). Despite the low variance explained, addition of more components was discarded since it did not provide new information (scores and loadings profiles did not give new interpretable features). The variance explained is lower than 70%, the value commonly accepted as acceptable in environmental studies. As it can be seen in Table 1, the set of sampling points lacks geographical homogeneity and temporal continuity. The low consistency of the database could be cause of the low percentage of variance explained. Representations of the loadings for PC1 versus PC2 in Fig. 4, show that PC1 differentiates the sampling sites according to their NO concentration levels. Locations with high traffic in the metropolitan area of Barcelona exhibited high positive loading values of PC1 (Fig. 4a). The site of El Prat de Llobregat (30) gave a low contribution of PC1, but it had a high loading value of PC2 instead. This sampling point belongs to the metropolitan area of Barcelona, where the international airport of El Prat is situated, consequently this point may register a different pattern on NO emissions, and it could differ from the road traffic daily trends followed by most of the other locations of the Barcelona metropolitan area. When searching for year-to-year differences (Fig. 4b), no clear year patterns were found. Only years 2005 and 2006 seemed to have some higher PC1 loadings than the other, but this conclusion may be obscured by the fact that the number of sampling stations available during these 2 years was higher. When scores were re-folded and averaged to obtain daily and annual average variations for each component, no clear conclusions could be extracted either, since the positive and negative contributions of each sampling site to the whole component impaired a clear interpretation of these profiles.

Therefore, MCR-ALS of matrix $\mathbf{D}_{\text{row-aug}}$ with non-negativity constraints was also attempted looking for an improved interpretation of the results. In this case, no data treatment was applied and only two components were used to account for approximately 56% of

the data variance. Note that PCA without any data pre-treatment (no autoscaling) explained the same percentage of data variance (see variances explained in Table 2). MCR-ALS loadings of the two components show how component 1 included mainly the sampling sites of year 2006, whereas component 2 explained basically the remaining years investigated, from 2001 to 2005 (both included). When MCR-ALS scores matrix was re-folded as indicated in Fig. 2b, daily and annual variation for each component could be investigated, shown in Fig. 5. Note that since non-negativity constraints were applied, the average scores can be interpreted directly in terms of NO concentrations variation. For both components the daily and annual patterns were very similar; therefore the final model basically has captured the differences in NO concentration patterns of variation between year 2006 and the rest. The daily profiles for both components resolved by MCR-ALS (Fig. 5a) showed two maxima; one between 8:00 and 9:00 am and the other, lower and wider, around 21:00. Similar profiles have been also reported in other studies for other urban locations [31–34]. The first peak appeared during the rush traffic hour and it could be caused mainly by fuel combustion by vehicles. Sunlight triggers a succession of photochemical reactions, which include the reaction of NO to give NO₂ and depletion of O₃. NO concentration decreases until it reached a minimum at approximately 16:00. This observation agrees with the results obtained in a previous work on the variation of O₃ in Catalonia [14]; the NO morning maximum overlaps with a minimum in O₃ levels, and as O₃ arrives to its maximum between 14:00 and 16:00, NO decreases, achieving its daily minimum value. After that, NO begins to accumulate again, reaching a new maximum during the evening [34]. As the photochemical reactions are not promoted during the evening, when the solar radiation and temperature are lower, this could lead to an accumulation of NO. This secondary maximum is also induced by evening traffic emissions, and the accumulation of NO in the lower mixing layer height. Just slight differences were found between the daily profiles for components 1 and 2; the daily variation patterns are very similar but it can be observed in Fig. 5a that the average NO for C1 (whose main contribution was year 2006) was lower than for C2 (remaining period from 2001 to 2005).

Regarding the seasonal variation of NO concentrations (Fig. 5b), it can be observed that its highest levels are registered during late autumn and winter. Several causes can justify these trends; one of them being the temperature variations associated to different seasons. In spring and summer the higher levels of solar radiation and temperatures promote photochemical reactions and the decreasing of NO concentration levels. Moreover, the stronger vertical atmospheric mixing in summer helps the mixture and dispersion of pollutants and therefore contributes to the lower levels of NO [33]. Finally, another reason can be the increase of NO emissions during winter due to the use of heating [34] and heavier traffic in cities compared to the vacations period. Similar patterns have also been observed in previous studies in other locations by other authors [32–37].

Table 2

Explained variances for the different models developed for the row-wise augmented matrix $\mathbf{D}_{\text{row-aug}}$ (all the NO concentrations collected in Catalonia from 2001 to 2006).

	Pre-treatment	Number of components.	% variance captured	Total % variance captured ^a
PCA	Autoscaling	1	25.9	25.9
		2	7.24	33.1
PCA	None	1	50.9	50.2
		2	5.4	56.3
MCR-ALS (non-negativity)	None	1	32.0	–
		2	36.2	56.2

^a Explained variances in MCR are not accumulative.

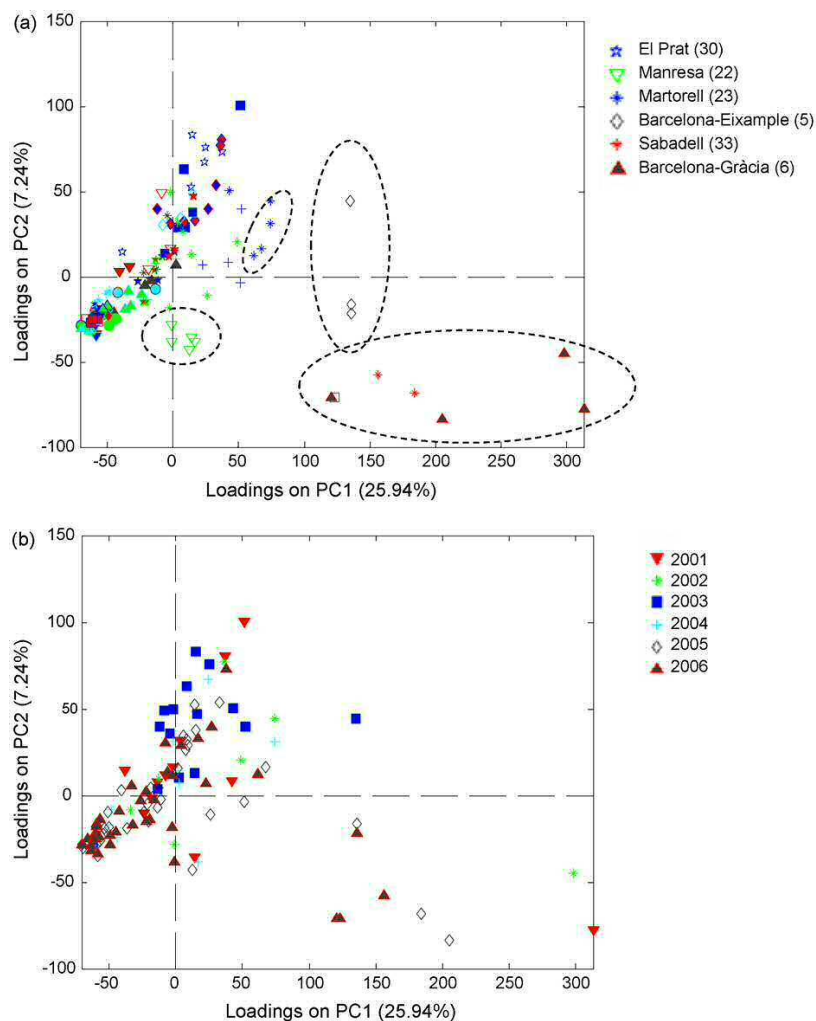


Fig. 4. Loading plots for the PCA model (autoscaling) of the row-wise augmented data matrix $D_{row-aug}$. PC2 loadings versus PC1, where (a) classes with different symbols correspond to different sampling locations and (b) classes with different symbols correspond to different years.

3.2. Sources and temporal patterns of NO in Catalonia: three cases of study

Multivariate models were developed in order to assess NO sources and their temporal patterns. In this study, models were developed for three individual sampling stations. In a previous paper on the distribution of O_3 in Catalonia [14], three sub-regions were identified, in which the O_3 followed distinctive patterns. Each of the location considered here belongs to one of the sub-regions, and their NO profiles are compared with those of O_3 and NO_2 , when possible. The cases of study selected were; (i) Barcelona-Gràcia, a populated and popular quarter of the city (site 6 on the map, Fig. 1), (ii) Bellver de Cerdanya (10), a rural area in the Pyrenees, and (iii) Tarragona-Sant Salvador (42), a location in medium size city, under the influence of a petrol refinery.

The year-averaged NO concentrations in Barcelona-Gràcia ranged from 40 to 79 $\mu\text{g m}^{-3}$ during the period 2001–2006 (Table 1), and a clear anthropological influence on the patterns of NO_x sources was expected. The 1/2-h NO values for this sampling site were arranged as shown in Fig. 3. Due to number of missing data (more than 10%), the study was limited to years 2001, 2002, 2005 and 2006.

PCA was applied first to the column-wise augmented data matrix D_{aug} as exploratory tool, in order to assess the correct number of components and to look for the main patterns of orthogonal variation. In Table 3, it can be seen that 3 components accounted for 77% and 90% of the variance for autoscaled and raw data, respectively. MCR-ALS was then applied to D_{aug} using 3 components and with non-negativity constraints. The model captured also 90% of the variance of the raw data (Table 3, note that explained variances in MCR models are not accumulative). To investigate the possible trilinear structure of data, the scores of each of the components were re-folded into 365×4 matrices and SVD analysed. Since only the first singular value was significant, it could be concluded that the 3 component score profiles (NO sources) behaved quite similarly throughout the period under study (for the sake of brevity, full details of these models are omitted) and therefore, the data could be well approximated using a trilinear model. Hence, MCR-ALS with non-negativity and trilinearity constraints was applied. The procedure is fully described in [38] and depicted in Fig. 3. The scores re-folding procedure allowed for recovering of the daily (Y , loadings mode 2), annual (X , loadings mode 1) and inter-annual (Z , loadings mode 3) profiles of each of the NO concentration patterns. Note than in the trilinear decompositions, the resolved profiles in

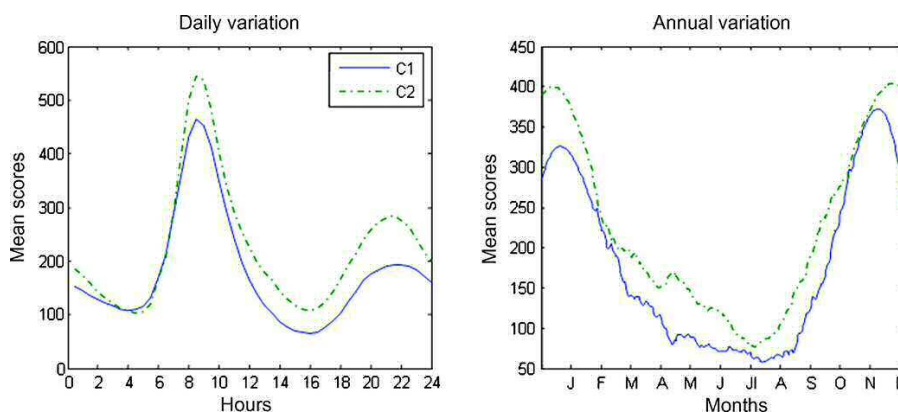


Fig. 5. Averaged scores after refolding of the MCR-ALS model (non-negativity on $D_{row-aug}$): (a) daily variation and (b) annual trends of components C1 and C2.

the three modes are always called loadings, whereas in the bilinear decompositions, the loadings in modes 1 and 3 are intermixed in the scores profiles. As non-negativity constraints were applied, the resolved averaged loadings in the three modes using a trilinear model could be interpreted straightforwardly in terms of NO concentration values.

Using the trilinear constraint, the resulting MCR-ALS model explained ca. 70% of the total variance (see Table 3, the percentage of overlapping variance of the 3 components was approximately only of 19%), which is lower than the variance explained by the model without trilinearity. Fig. 6 shows the average loadings in the three modes for the 3 components, i.e. NO sources identified.

According to the literature [39], soil emissions and lightning are the main natural sources of NO_x into the atmosphere, being soil emissions predominantly in the form of NO. The natural emission of NO is markedly seasonal, and it reaches maximum rates in summer (up to 14% of the anthropogenic emissions, and even more in areas that have high agricultural activities). As it can be seen in Fig. 6a–c, none of the NO sources identified reaches a maximum in summer, consequently natural or biogenic NO sources can be discarded as the major sources in this sampling site.

Barrero et al. [40] have reported the daily cycles of NO in an urban atmosphere (Bilbao, North Spain). They observed that the daily emissions of NO had different patterns in each season. The summer profile showed a clear maximum at early hours at 8:00 am, and then it reached a constant level until the following day. During winter the morning peak was up to three times higher than in summer, and moreover, a second maximum was observed at 20:00.

The evening maximum was half of the morning peak. The results obtained in this work for Barcelona–Gràcia agreed with the patterns observed by Barrero et al. [40] and also with a heavily industrialised area in the south of Spain (Campo de Gibraltar [8] and in Athens basin (Greece) [32].

In the present study for Barcelona–Gràcia, the second and third components of the MCR-ALS models seem to account for the seasonal differences in the daily profile, and they can be clearly attributed to traffic contributions, since they correspond to the afternoon and morning rush hours, respectively (see Fig. 6e and f). During the summer the main contribution to atmospheric NO is due to the morning rush hour (Fig. 6c and f), whereas the contribution of the evening traffic is rather low (Fig. 6b and e). This could be attributed to the fact that during spring and summer the high temperatures and solar radiation, in combination with lower precipitation promote the formation of radicals at noon, which accumulate during the afternoon and evening. Consequently during the evening rush hour the NO is quickly removed by the radicals accumulated. On the contrary, in autumn and winter the formation of radicals is not favoured by the meteorological conditions, and therefore, NO can accumulate, resulting in a higher morning peak (as concluded from Fig. 6c and f) and a second maximum at 20:00. Regarding the first component, it can be seen in Fig. 6a and b that it represents a NO contribution during the night, and this input is almost constant throughout the year. This could be due to the fact that during the night photochemical reaction are not expected, and consequently the differences between summer and winter are not so marked. A slight increment during the winter of this night

Table 3

Explained variances for the different models developed for the D_{aug} matrix, corresponding to NO data from Barcelona–Gràcia (years 2001, 2002, 2005 and 2006).

	Pre-treatment	Number of components	% variance captured	Total % variance captured ^a
PCA	Autoscaling	1	48.9	48.9
		2	20.0	68.9
		3	8.0	76.9
PCA	None	1	78.5	78.5
		2	6.6	85.0
		3	5.1	90.1
MCR-ALS (non-negativity)	None	1	17.4	–
		2	40.8	–
		3	51.7	90.0
MCR-ALS (non-negativity and trilinearity.)	None	1	8.9	–
		2	33.8	–
		3	46.1	69.9

^a Explained variances in MCR are not accumulative.

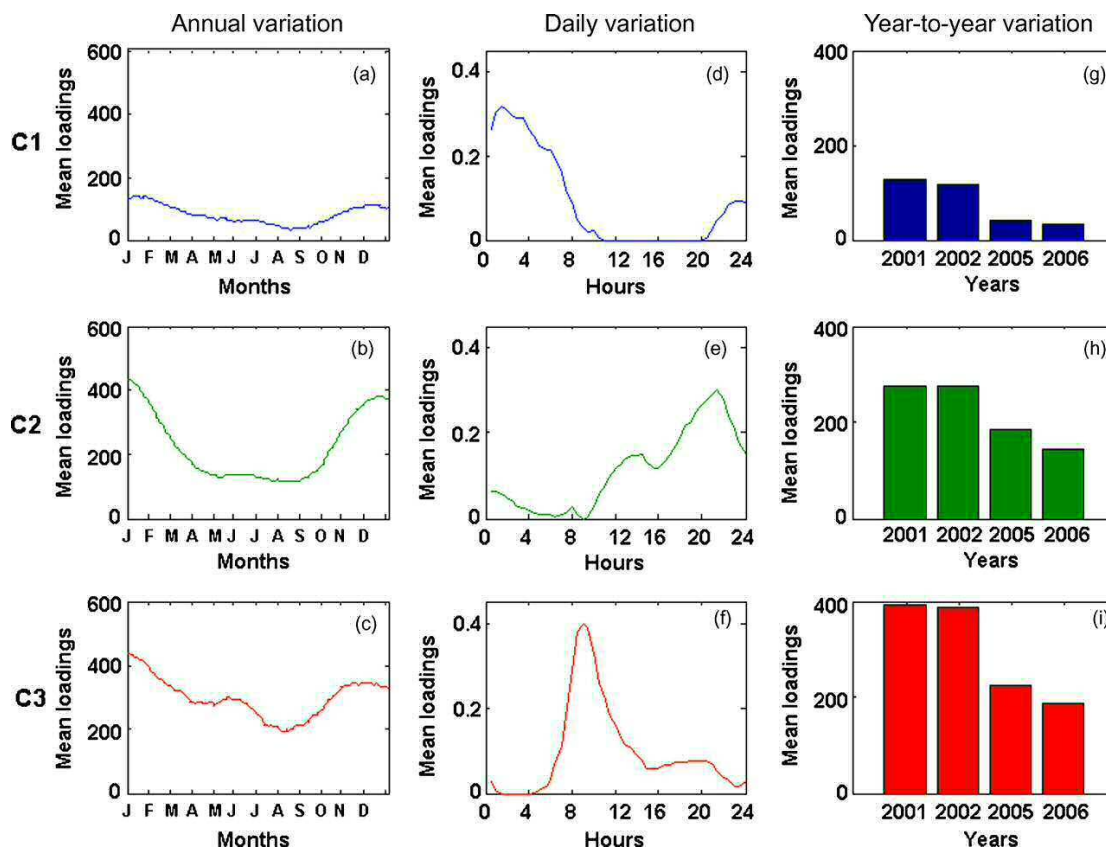


Fig. 6. MCR-ALS resolved profiles for the three resolved components using non-negativity and trilinearity constraints for NO concentration variations in Barcelona-Gràcia sampling station. (a–c) seasonal; (d–f) daily; (g–i) inter-annual trends.

contribution could be attributed to the use of fuel-fed heating during the evening. Moreover this contribution is much lower than the morning and afternoon inputs.

The role of NO_x ($\text{NO} + \text{NO}_2$) in the O_3 production is complicated. Whilst NO_x ($\text{NO} + \text{NO}_2$) acts as O_3 precursor according to the reactions $\text{NO}_2 + h\nu \rightarrow \text{NO} + \text{O}(^3\text{P})$ and $\text{O}(^3\text{P}) + \text{O}_2 + \text{M} \rightarrow \text{O}_3 + \text{M}$, it also destructs ozone following the reaction $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$ (where M represents a third-body molecule) [41]. In the case of high NO/NO_2 mixing ratios, O_3 is destroyed by injection of fresh NO, whereas low NO/NO_2 ratios result in decreasing of titration as accumulation of O_3 [42], VOC- or NO_x -limited situations, respectively [41]. Hence, when these profiles of NO daily and seasonal variations are compared with those obtained for O_3 in urban areas [14], it was observed that NO variation clearly opposed O_3 daily oscillation. In urban sites in Catalonia, O_3 showed a daily minimum that overlapped with maximum concentration of NO between 7:00 and 9:00 am, whereas O_3 maximum between 14:00 and 16:00 h corresponded quite well with the minimum in NO concentration; when maximum solar radiation and temperatures are reached. A study of NO_2 in this location, analogous to the model applied to NO (column-wise matrix augmentation, and MCR-ALS with non-negativity and trilinearity) showed that NO_2 daily variation in Barcelona-Gràcia followed similar pattern to NO. Full study of NO_2 in Catalonia was not possible due to the lack of consistency of the data base, and models were only made for comparison purposes. NO_2 daily profile showed a clear maximum in the morning, from 8:00 to 10:00 am, and a minimum around 16:00, overlapping with those peaks of NO. However, the daily oscillations of NO_2 were smooth and not as pronounced as observed for nitric oxide. As the NO/NO_2 ratio in the

early hour is high, this justifies the depletion in O_3 caused by the introduction of fresh NO in the troposphere. At noon the NO/NO_2 decreases (NO levels fall much faster than NO_2) causing the increment of O_3 . Similar NO, NO_2 and O_3 tendencies have been observed in other urban sites [32]. Regarding seasonal cycles, NO_2 show little changes along the year, whereas NO clearly decreased during spring and summer, when the O_3 concentrations achieved their maximum values [14].

When year-to-year variation was considered (Fig. 6g–i), it could be clearly seen that there was an outstanding reduction trend of NO emission during the time period under study. This encouraging trend has been also observed in other hot-spots of Europe, such as Athens basin [32]. The gap between 2002 and 2005 was due to the absence of data for 2003 and 2004. If these data were included the decreasing trend would have been smoother.

Day-of-week effects have been reported previously for several pollutants, NO among them [6,33]. The main cause of this effect is the reduction of traffic. Moreover, during the weekends there is reduction of particulate matter emissions from heavy-duty trucks and industries. This causes the increment of solar radiation at ground level and consequently the formation of radicals that react with NO, favouring its depletion [33]. It can be observed in Fig. 7 that mean loadings for the components 3 and 2, which corresponded to NO maximum levels during the morning and evening rush hours, respectively, decreased markedly during the weekend. This tendency confirms the anthropogenic origin of these contributions. On the contrary, component C3, the night contribution, showed a slight increment during the weekend. The night accumulation of NO dur-

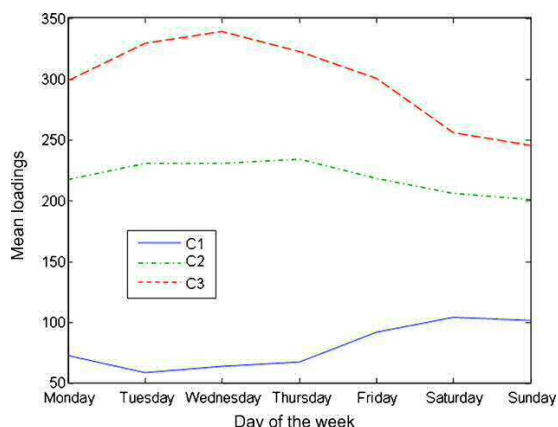


Fig. 7. Average scores (MCR-ALS) representing weekly NO variation of components C1, C2 and C3 in Barcelona-Gràcia.

ing the weekend could be due to the reduction in the emission of other atmospheric pollutants that act as NO scavengers.

Regarding the NO patterns in the rural location, Bellver de Cerdanya (10, in Fig. 1) showed (Table 1) an average concentration of $2.7 \mu\text{g m}^{-3}$ from 2004 to 2006, one of the lowest NO levels registered in Catalonia. Data collected during this period were arranged as depicted in Fig. 3. The application of MCR-ALS with non-negativity and trilinearity constraints gave a three-component model very similar to the ones obtained for Barcelona-Gràcia (65% variance captured). The NO contributions identified in this rural area followed the same profiles as the ones resolved in the first case of study of this section. None of them showed a maximum in summer, which is characteristic of natural NO sources [39]. Consequently, it could be concluded, no natural NO sources could be identified in this rural location or they were negligible (if any) and the three contributions were mainly anthropogenic. It is worth noting that, even though the NO concentrations in Bellver de Cerdanya were much lower than in the urban locations, the temporal profiles followed the same trends. The similarities between both locations regarded also the NO₂ variation, showing a daily profile with two maxima around 9:00 am and 20:00 pm, and a minimum registered from 14:00 to 16:00, that overlapped with the NO profiles.

Tarragona-Sant Salvador (42) showed a moderate concentration of NO that ranged from 8.6 to $9.9 \mu\text{g m}^{-3}$ during the period 2004–2006. The MCR-ALS model (non-negativity and trilinearity) explained 54% of the variance with 2 components. They followed the same profiles as C2 and C3 and for Barcelona-Gràcia in Fig. 6e and f. The contribution during the night was not resolved.

In both Bellver de Cerdanya and Tarragona-Sant Salvador weekend-effects, were observed, with a decrease in NO concentration during Saturday and Sunday, although not so marked as the NO depletion observed in Barcelona-Gràcia.

The fact that NO profiles in three very different locations showed such similar tendencies proved the homogeneity of tropospheric NO in Catalonia, which had been already suggested by the results obtained in the previous section, when analysing **D_{row-aug}**. In a previous work on the O₃ [14], clear differences were observed within the territory of Catalonia. In the present work it has been shown that these differences cannot be entirely attributed to the titration effects of NO and NO₂, since these two pollutants seem to be rather uniform in all the territory. The PCA study of the variables influencing O₃ concentration suggested that other parameters such as wind regimen, humidity and particulate matter influenced the formation/depletion rate of O₃.

4. Conclusions

The application of multivariate analysis to large data sets acquired in current monitoring studies by governmental and environmental agencies and stored in large public data basis allows working with a large number of variables and makes easier its interpretation without the need of more involved mechanistic models or time-consuming calculations. Multivariate analysis and chemometric methods greatly help to understand and interpret better the trends and patterns of environmental behaviour of main pollutants in a particular geographical area over time.

When the geographical and temporal patterns of NO in Catalonia were investigated, it was observed that PCA models succeeded in grouping the sampling sites according to their concentration NO levels, but the interpretation of the seasonal and daily trends and patterns of NO concentration levels was more difficult. The application of Multivariate Resolution techniques like MCR-ALS with non-negativity constraints to the simultaneous analysis of data in all the sampling points available provided insight in the temporal and geographical distributions of NO in Catalonia. Additionally, when MCR-ALS applied to three individual sampling sites (two urban locations and one rural site), it could be observed that the three locations showed very similar NO contributions, and their temporal profiles could be easily interpreted in terms of the traffic patterns and photochemical reactions in the atmosphere.

Acknowledgements

The authors acknowledge the Spanish Ministry of Education for the award of the research project CTQ-2006-15052 and the Generalitat of Catalunya, Department of Environment for providing the data base. The authors also acknowledge Ms M. Terrado MSc for her help with Geographical Information Systems (GIS).

References

- [1] Y. Itano, H. Bandow, N. Takenaka, Y. Saitoh, A. Asayama, J. Fukuyama, *Sci. Total Environ.* 379 (2007) 46.
- [2] O.T. Denmead, D. Chen, D.W.T. Griffith, Z.M. Loh, M. Bai, T. Naylor, *Aust. J. Exp. Agric.* 48 (2008) 213.
- [3] M.S. Bergin, J.S. Shih, A.J. Krupnick, J.W. Boylan, J.G. Wilkinson, M.T. Odman, A.G. Russell, *Environ. Sci. Technol.* 41 (2007) 4677.
- [4] A. Kaiser, H. Schellfing, W. Spangl, A. Weiss, S. Gilge, W. Fricke, L. Ries, D. Cemas, B. Jesenovec, *Atmos. Environ.* 41 (2007) 9273.
- [5] M.L. Thompson, J. Reynolds, L.H. Cox, P. Guttorp, P.D. Sampson, *Atmos. Environ.* 35 (2001) 617.
- [6] H.O. Gao, *Transport. Res. Part D: Transport Environ.* 12 (2007) 292.
- [7] S.C. Chang, C.T. Lee, *J. Environ. Manage.* 86 (2008) 627.
- [8] F.J. González Gallero, M. Galán Vallejo, A. Umbría, J. Baena Gervilla, *Environ. Monit. Assess.* 119 (2006) 405.
- [9] T. Ozeki, T. Ihara, N. Ogawa, *Chemometr. Intell. Lab. Syst.* 82 (2006) 15.
- [10] M.H. Shah, N. Shaheen, *J. Hazard. Mater.* 147 (2007) 759.
- [11] R.C. Henry, B.M. Kim, *Chemometr. Intell. Lab. Syst.* 8 (1990) 205.
- [12] P. Paatero, *Chemometr. Intell. Lab. Syst.* 38 (1997) 223.
- [13] P. Paatero, *J. Comput. Graph. Stat.* 8 (1999) 854.
- [14] M. Felipe-Sotelo, L. Gustems, I. Hernández, M. Terrado, R. Tauler, *Atmos. Environ.* 40 (2006) 7421.
- [15] R. Guardani, J.L. Aguiar, C.A.O. Nascimento, C.I.V. Lavaca, Y. Yanagi, *J. Air Waste Manage. Assoc.* 53 (2003) 553.
- [16] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer-Verlag, New York, USA, 2002.
- [17] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, 2nd ed., Elsevier Science B.V., The Netherlands, 1998.
- [18] A. de Juan, R. Tauler, *Anal. Chim. Acta* 500 (2003) 195.
- [19] J. Jaumot, R. Gargallo, R. Tauler, *J. Chemometr.* 18 (2004) 327.
- [20] R. Tauler, *J. Chemometr.* 15 (2001) 627.
- [21] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, USA, 1989.
- [22] R. Tauler, *Chemometr. Intell. Lab. Syst.* 30 (1995) 133.
- [23] R. Tauler, A. Smilde, B. Kowalski, *J. Chemometr.* 9 (1995) 31.
- [24] F. Berbel, E. Kapoya, J.M. Díaz-Cruz, C. Arino, M. Esteban, R. Tauler, *Electroanalysis* 15 (2003) 499.
- [25] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, *BMC Bioinformatics* 7 (2006) 343.

- [26] E. Pere-Trepat, A. Ginebreda, R. Tauler, *Chemometr. Intell. Lab. Syst.* 88 (2007) 69.
- [27] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, *Chemometr. Intell. Lab. Syst.* 76 (2005) 101.
- [28] A.K. Smilde, R. Bro, P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*, 1st ed., John Wiley & Sons Chichester, United Kingdom, 2004.
- [29] A. de Juan, R. Tauler, J. Chemometr. 15 (2001) 749.
- [30] R. Tauler, S. Lacorte, M. Guillamón, R. Céspedes, P. Viana, D. Barceló, *Environ. Tox. Chem.* 23 (2004) 565.
- [31] D. Pudasainee, B. Sapkota, M.L. Shrestha, A. Kaga, A. Kondo, Y. Inoue, *Atmos. Environ.* 40 (2006) 8081.
- [32] N. Hatzianastassiou, B.D. Katsoulis, B. Antakis, *Environ. Monit. Assess.* 128 (2007) 447.
- [33] J. Zabalza, D. Ogulei, D. Elustondo, J.M. Santamaría, A. Alastuey, X. Querol, P.K. Hopke, *Environ. Monit. Assess.* 134 (2007) 137.
- [34] J. Tu, Z.G. Xia, H.S. Wang, W.Q. Li, *Atmos. Res.* 85 (2007) 310.
- [35] W.K. Jo, J.H. Park, *Chemosphere* 59 (2005) 1557.
- [36] S. Saito, I. Nagao, H. Tanaka, *Atmos. Environ.* 36 (2002) 1277.
- [37] R.L. Raddatz, J.D. Cummine, *Boundary-Layer Meteorol.* 99 (2001) 411.
- [38] R. Bro, S. DeJong, J. Chemometr. 11 (1997) 393.
- [39] M. Trainer, D.D. Parrish, P.D. Goldan, J. Roberts, F.C. Fehsenfeld, *Atmos. Environ.* 34 (2000) 2045.
- [40] M.A. Barrero, J.O. Grimalt, L. Cantón, *Chemometr. Intell. Lab. Syst.* 80 (2006) 67.
- [41] Y. Sadanaga, S. Shibata, M. Hamana, N. Takenaka, H. Bandow, *Atmos. Environ.* 42 (2008) 4708.
- [42] K.K. Yeung, W.L. Chang, R.E. Newell, W. Hu, G.L. Gregory, *HKMets Bull.* 6 (1996) 3.

Web references

- [43] Environmental Protection Agency. <http://www.epa.gov/air/urbanair/nox/>.
- [44] Generalitat de Catalunya. <http://www.gencat.net/>.
- [45] Catalanian atmospheric monitoring network. <http://mediambient.gencat.net>.
- [46] MCR-ALS graphical interface for MATLAB. <http://www.ub.es/gesq/mcr/mcr.htm>.

Anal Bioanal Chem (2011) 399:2015–2029
DOI 10.1007/s00216-010-4458-1

ORIGINAL PAPER

Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O₃ pollution in Barcelona

Marta Alier · Mónica Felipe · Isabel Hernández · Romà Tauler

Received: 31 July 2010 / Revised: 17 November 2010 / Accepted: 18 November 2010 / Published online: 21 December 2010
© Springer-Verlag 2010

Abstract Multiway and multiset data analysis extensions of the multivariate curve resolution alternating least squares (MCR-ALS) method are proposed for the investigation of the temporal distribution of the pollution by nitric oxide (NO) and ozone (O₃) in one sampling station in the urban centre of Barcelona (Catalonia, Spain), during the years 2000–2006. Different specific studies were performed considering the annual and pluriannual contamination by these two contaminants, individually or in combination using different data matrix augmentation strategies and multiway and multiset data analysis models. Daily, hourly and annual profiles were estimated describing different patterns and summarising the main contamination processes. The daily and night trends found were mainly attributed to traffic and photochemical processes favoured by light

radiation. Moreover, winter–summer seasonal trends were also clearly detected and their changes over different years assessed. The extension MCR-ALS method to multiset data analysis using different constraints like non-negativity, trilinearity and interaction among components is confirmed to be a powerful method to improve the interpretability of the different contamination patterns in atmospheric contamination studies.

Keywords Chemometrics · Multiway data analysis methods · Multiset data analysis methods · Multivariate curve resolution alternating least squares · Air pollution · Nitric oxide · Ozone

Published in the special issue of *Chemometrics (VII Colloquium Chemiometricum Mediterraneum)* with Guest Editors Marcelo Blanco, Juan M. Bosque-Sendra and Luis Cuadros-Rodríguez

Electronic supplementary material The online version of this article (doi:10.1007/s00216-010-4458-1) contains supplementary material, which is available to authorized users.

M. Alier · R. Tauler (✉)
Department of Environmental Chemistry, IDAEA-CSIC,
C/Jordi Girona Salgado, 18-26,
08034 Barcelona, Spain
e-mail: Roma.Tauler@idaea.csic.es

M. Felipe
Department of Chemistry, Loughborough University,
Leicestershire, UK LE11 3TU

I. Hernández
Department of Environment, Generalitat de Catalunya,
Av. Diagonal, 523-525,
08029 Barcelona, Spain

Introduction

Multivariate curve resolution–alternating least squares (MCR-ALS) has become a popular chemometric method for data analysis [1, 2]. On one side, this recognition is due to the great variety of data sets that can be analysed by this method; essentially, any multicomponent system giving one or multiple data tables or data matrices which can be described by a bilinear model. These multicomponent systems include all kinds of chemical processes and mixtures (e.g., chemical reactions, industrial processes, chromatographic runs, spectroscopic images or environmental data, among others) measured and/or monitored using multivariate responses and sensors. Another reason for the general acceptance of MCR-ALS is its remarkable flexibility to analyse multiple data matrices simultaneously and the diversity of constraints that can be applied to improve the resolution results.

MCR methods have been extended to the analysis of three-way and multiway data sets and to other multidata set arrangements [1]. Resolution ambiguities and rank deficiency [3] problems frequently inherent to bilinear modelling of two-way data sets can be reduced significantly and even eliminated when these more complex data structures containing richer information (like multiway data sets) are analysed by the extension of multivariate curve resolution methods [4].

Bearing in mind that the basic idea of MCR methods is the correct resolution of the pure component profiles from their mixtures, it has to be recognised that there are other multiway methods, such as PARAFAC [5], or those based on Tucker models, that in many circumstances can also be used for resolution purposes, although they do not fall into the classical MCR denomination. In this sense, it is interesting to emphasise that the typical application of MCR to simultaneous analysis of multiple data sets via column- or row-wise augmented data matrices can be easily extended to impose multilinear models, such as those based on trilinear [4] or Tucker models [6] during the ALS optimization process, in the form of constraints. As a particularity of the extension of the MCR-ALS algorithm, these constraints can be applied either to the whole set of component profiles or selectively to only some of them. These constraints are very strong, and they can eliminate the rotation ambiguities inherent to bilinear modelling [3], and provide sounder chemical knowledge linked to physicochemical parameters of interest.

The present work is focussed on promoting and extending the use of multivariate curve resolution alternating least squares method, including trilinear and component interaction constraints, for the investigation of NO and O₃ air pollution. Temporal patterns of O₃ and NO have been studied in detail in one station of the centre of Barcelona (Gràcia) for some years in the period comprised between 2000 and 2006, even though not all the years were finally included because of the lack of compatibility between years. The Barcelona-Gràcia measurement station was chosen for its location in a central area of Barcelona and for its known high levels of pollution compared to other areas. In a previous work, O₃ and NO in Catalonia were already studied separately for a larger number of stations [7, 8].

NO and O₃ have correlated sources and reactivities [8, 9]. NO is mainly emitted by car exhausts, and it reacts with O₃ giving NO₂, which acts as an O₃ sink. On the contrary, in inner rural areas with low anthropogenic activities, this titration effect of NO emissions is not observed, resulting in higher average O₃ concentrations than in urban areas. Such a phenomenon has been already reported for O₃ concentration in other regions of Spain [9]. In the case of coastal sites, air flows can cause disturbances in vertical stability and moreover, deposition of O₃ on the sea surface is slow

[10]. Consequently, medium-sized cities at the coast would exhibit O₃ concentrations and variability similar to rural inner areas. Nitrogen oxides (NO_x), a mixture of nitric oxide (NO) and nitrogen dioxide (NO₂), are produced from natural sources, motor vehicles and other fuel combustion processes. Although nitric oxide does not significantly affect human health, ground-level ozone can trigger serious respiratory problems [11]. NO can react to form nitrate particles, acid aerosols, as well as NO₂, which also cause respiratory problems and it contributes to the formation of acid rain and atmospheric particles [12]. The short-term exposure to concentrations of ozone can temporarily affect the lungs, respiratory tract and eyes. It can also increase susceptibility to respiratory allergens. The long-term exposure to relatively low ozone concentrations can cause a reduction in the lung function. Spanish legislation on air quality establishes as 1-h limit for human health protection 200 µgm⁻³ NO₂ and NO_x and an alert threshold of 400 µgm⁻³ (both values expressed as NO₂; [13], RD 1073/2002 and RD 717/1987). In the case of O₃, the human health protection threshold is 120 µgm⁻³ ([13], Directive 2002/3/ECE and RD 1796/2003). The main source of air pollution, and in particular NO_x, is traffic, although there are also a number of small industrial activities contributing to air pollution. It is remarkable the size of the fleet of private cars in Catalonia, which arrives to ca. 1,000,000 vehicles only in Barcelona.

Some recent works have applied different multivariate approaches to the study of atmospheric pollution by O₃/NO. Gao [14] has applied a functional analysis of variance on O₃/NO_x measurements in order to characterise their diurnal cycles. Principal components analysis has been used quite often in atmospheric pollution research in order to establish seasonal variations [15], geographical distributions, dependence with meteorological conditions [16] and atmospheric chemical reactions. However, factors found by principal component analysis (PCA) do not often have straightforward or unique interpretations. In order to overcome this drawback of PCA when applied to atmospheric data, some authors have applied different type of factor rotations [17] to clarify the meaning of factors. In the present work, the extension of MCR-ALS to investigate multiway data is proposed to evaluate simultaneously the NO and O₃ contamination multiset and data sets as an alternative to well-known methods like PCA. Although MCR-ALS has been usually applied for the resolution of multiple components in unresolved mixtures investigated by spectroscopic means, it has been also proved its ability to resolve and model average temporal and geographical distribution profiles from environmental monitoring studies. One of the main advantages of MCR-ALS is that it decomposes the data matrix by applying more natural constraints than PCA, such as non-negativity, and thus, environmental results can be interpreted more straightforwardly.

Data

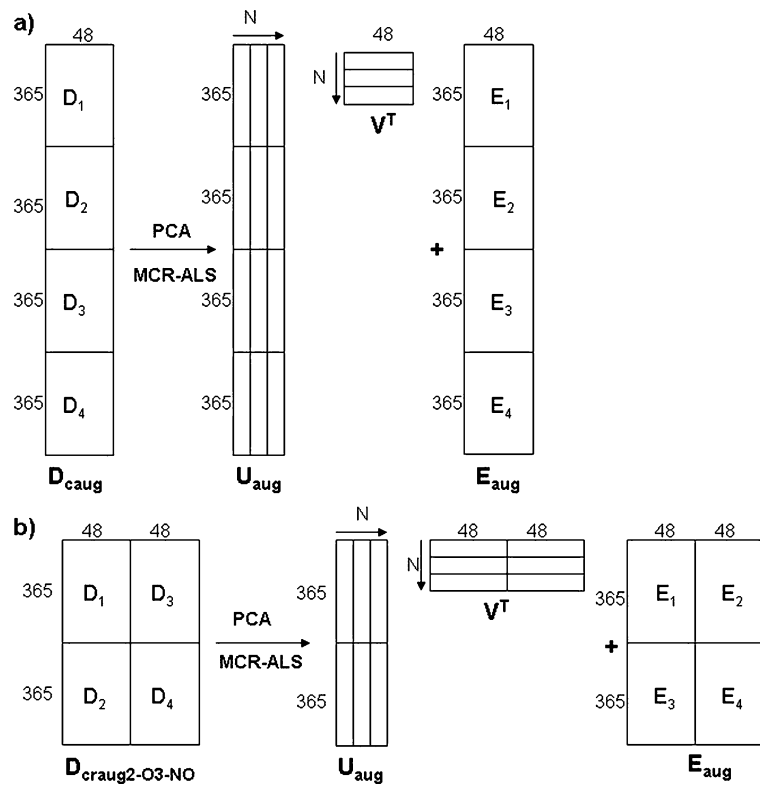
Experimental data

Half-hourly NO and O₃ concentrations were measured from 1 January 2000 to 31 December 2006 in one sampling station of the city of Barcelona, in the district of Gràcia. This station was selected because it is close to the centre of the city and it is highly populated. Only years 2001, 2002, 2005 and 2006 were used for NO and years 2000, 2001, 2002 and 2003 for O₃ due to an excess of missing data in the other years. Ozone concentrations were measured by means of UV photometry according to ISO 13964:1998 [18], and they were also acquired by automatically operated analysers. These analysers provide ozone measurements every minute and they are saved in half-hourly periods. NO concentrations were measured according to the UNE method 77212:1993 (equivalent to ISO 7993:1985) [19], using automatically operated chemiluminescence analysers. See [13] for further information about the Catalanian atmospheric monitoring network. This database was transferred into MATLAB 7.0 (The Mathworks, Natick, USA) for subsequent multivariate data analysis. Before the analysis, missing data were estimated by PCA imputation by means of the MATLAB PLS 3.5 Toolbox.

Data arrangement

Experimental data were stored in half-hour periods of every day during every year. This gave a data vector of 17,520 values per year for every measured compound (O₃ or NO) in the considered measurement station. A proper folding of this very long data vector gives a new data matrix with 365 rows corresponding to the different days of the year and with 48 columns corresponding to the half-an-hour values of every day. In the present work, three different studies have been performed. In a first study, NO concentration measures have been investigated during 4 years (2001, 2002, 2005, 2006) using a column-wise augmented data matrix $D_{\text{caug-NO}}$ (1,460×48) as shown in Fig. 1a. In a second study, O₃ concentration measures have been investigated during 4 years (2000, 2001, 2002, 2003), using a column-wise augmented data matrix $D_{\text{caug-O}_3}$ (1,460×48) as shown in Fig. 1a. And finally, in a third study, the combination of NO and O₃ concentration measures have been analysed simultaneously for 2 years (2001, 2002) using two different data arrangements, a column-wise augmented data matrix similar to the other studies, $D_{\text{caug1-O}_3\text{-NO}}$ (1,460×48; Fig. 1a), and a super (row- and column wise) augmented matrix $D_{\text{craug2-O}_3\text{-NO}}$ (730×96), as it is shown in Fig. 1b. In order to make the

Fig. 1 PCA and MCR-ALS bilinear decompositions of augmented data matrices. **a** Bilinear decomposition of column-wise augmented data matrices. In NO column-wise augmented data matrix $D_{\text{caug-NO}}$, D_1 is year 2001, D_2 is year 2002, D_3 is year 2005 and D_4 is year 2006. In O₃ column-wise augmented data matrix $D_{\text{caug-O}_3}$, D_1 is year 2000, D_2 is year 2001, D_3 is year 2002 and D_4 is year 2003. In O₃ and NO column-wise augmented data matrix $D_{\text{caug1-O}_3\text{-NO}}$, D_1 is year 2001 (O₃), D_2 is year 2002 (O₃), D_3 is year 2001 (NO) and D_4 is year 2002 (NO). **b** Bilinear decomposition of super (row- and column-wise) augmented data matrices. In O₃ and NO super (row- and column-wise) augmented data matrix $D_{\text{craug2-O}_3\text{-NO}}$, D_1 is year 2001 (O₃), D_2 is year 2002 (O₃), D_3 is year 2001 (NO) and D_4 is year 2002 (NO)



scales comparable of the two types of concentration measures for NO and O₃, in the augmented and super-augmented data matrix of NO and O₃, the values of O₃ concentration were multiplied by 3 before their inclusion in the data matrix augmentation. In this supraugmented data matrix, the 2 years were first column-wise augmented for O₃ and NO matrices and then they were row-wise augmented.

In compact (MATLAB notation) the different matrix augmentations can be written as:

$$\text{Column - wise: } D_{\text{caug}} = \begin{pmatrix} \mathbf{D1} \\ \mathbf{D2} \\ \mathbf{D3} \\ \mathbf{D4} \end{pmatrix} = [\mathbf{D1}; \mathbf{D2}; \mathbf{D3}; \mathbf{D4}] \quad (1)$$

$$\begin{aligned} \text{Row - wise: } D_{\text{raug}} &= (\mathbf{D1}\mathbf{D2}\mathbf{D3}\mathbf{D4}) \\ &= [\mathbf{D1}, \mathbf{D2}, \mathbf{D3}, \mathbf{D4}] \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Column and row - wise: } D_{\text{craug}} &= \begin{pmatrix} \mathbf{D1,O3} & \mathbf{D1,NO} \\ \mathbf{D2,O3} & \mathbf{D2,NO} \end{pmatrix} \\ &= [\mathbf{D1,O3}, \mathbf{D1,NO}; \mathbf{D2,O3}, \mathbf{D2,NO}] \end{aligned} \quad (3)$$

Methods

Multivariate data analysis methods

PCA [20], MCR [1, 2], PARAFAC [5] and restricted Tucker3 [5] multivariate data analysis methods have been used in this work to analyse the different experimental data sets under study. The focus of this work is, however, the application of the multivariate curve resolution–alternating least squares method and of its extensions to analyse different type of trilinear and multiset data structures [1].

PCA and MCR are based on the bilinear decomposition given by the two following equations:

$$d_{ij} = \sum_{n=1}^N u_{in} v_{nj} + e_{ij} \quad (4)$$

$$D = UV^T + E \quad (5)$$

where $\mathbf{D}(I,J)$ is the original data array collecting all d_{ij} individual measures (of O₃ or NO), with $i=1,\dots,I$ rows (number of day observations during the year, $I=365$) and $j=1,\dots,J$ columns (every half-an-hour observation during

1 day, $J=48$). This is a linear mixture model where the measured values in the data matrix d_{ij} are the result of the sum of a reduced number of contributions $n=1,\dots,N$, from different sources plus an unknown amount of unexplained variance (noise) e_{ij} collected in the error matrix $\mathbf{E}(I,J)$. Each of these individual contributions can be described by the product of two factors, one related with the variance along the rows, u_{in} and another one related with the variance along the columns, v_{nj} . In the PCA bilinear decomposition, $\mathbf{U}(I,N)$ is the matrix of orthogonal scores, u_{ij} , where N is the reduced number of components or factors considered in the analysis; $\mathbf{V}^T(N,J)$ is the matrix of orthonormal loadings, v_{nj} , and $\mathbf{E}(I,J)$ is the matrix of residuals, e_{ij} , not modelled by the N principal components. These N components are obtained under the additional constraint of maximum explained variance and they are ranked accordingly (see [20, 21] for more details). Thus, a reduced set of \mathbf{N} ($N \ll J$ or I) principal components (PCs) are selected that contain most of the relevant variance (not noise) in the original matrix. Orthogonality, normalisation and maximum explained variance constraints in PCA provide unique solutions of the bilinear model in Eq. 4 which are then used for data exploration and interpretation. However, PCs derived from PCA are not always easily interpreted in environmental terms. PCA matrix decomposition of Eq. 4, is usually performed using a singular value decomposition algorithm [20].

MCR methods are based on the same bilinear model of Eqs. 4 and 5 and they decompose the same data matrix. In the MCR-ALS method, this decomposition is performed using an alternating least squares algorithm, which minimises the least squares error function Q defined by

$$Q^2 = \sum_{i=1}^m \sum_{j=1}^n (d_{ij} - \hat{d}_{ij})^2 \quad (6)$$

$$\frac{\partial Q^2}{\partial V^T} = 0, \frac{\partial Q^2}{\partial U} = 0 \quad (7)$$

Which is expressed by the two equations

$$\min_{V^T} \left\| \hat{\mathbf{D}}_{N,PCA} - UV^T \right\| \quad (8)$$

$$\min_U \left\| \hat{\mathbf{D}}_{N,PCA} - UV^T \right\| \quad (9)$$

where $\hat{\mathbf{D}}_{N,PCA}$ is the PCA reproduced (and filtered) data matrix for N components. In case of not applying any

constraints, Eqs. 8 and 9 have as a solution the two following matrix equations

$$\begin{aligned} \hat{\mathbf{V}}^T &= (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U} \hat{\mathbf{D}}_{n,PCA} = \mathbf{U}^+ \hat{\mathbf{D}}_{n,PCA} \\ \hat{\mathbf{U}} &= \hat{\mathbf{D}}_{n,PCA} \mathbf{V} (\mathbf{V}^T \mathbf{V})^{-1} = \hat{\mathbf{D}}_{n,PCA} (\mathbf{V}^T)^+ \end{aligned} \tag{10}$$

where $(\mathbf{V}^T)^+$ and $(\mathbf{U})^+$ are the current estimations of the pseudoinverse of \mathbf{V}^T and \mathbf{U} [20]. However in the MCR-ALS algorithm, these two equations are solved iteratively under a set of constraints applied over the factor matrices \mathbf{U} and \mathbf{V}^T , such as non-negativity, unimodality, closure, or selectivity, as it is described elsewhere [3, 21, 22]. In the case of MCR-ALS, \mathbf{U} and \mathbf{V}^T factor matrices are not orthogonal, but they should be full rank (invertible) and fulfil the required constraints.

The first step in the MCR-ALS method is the determination of the number of components N contributing to \mathbf{D} variance. This could be done in the same way as for PCA. Next, initial estimates of either \mathbf{U} or \mathbf{V}^T are obtained. An estimation of \mathbf{U} or \mathbf{V}^T can be obtained from the set of “purest” columns or rows of matrix \mathbf{D} [23] (those columns or rows of the original data matrix \mathbf{D} which are more dissimilar and more orthogonal to the rest). Equations 8 and 9 are then solved iteratively under the stipulated constraints until convergence is achieved (i.e. when in two consecutive iterative cycles, relative differences in standard deviations of the residuals between experimental and ALS calculated data values are less than a previously selected threshold value).

One of the main advantages of MCR-ALS is that it decomposes the data matrix applying more natural constraints than PCA, such as non-negativity, and thus, environmental results can be interpreted more straightforwardly. The drawback of MCR-ALS however, is that solutions may not be unique, in contrast to PCA, and caution should be taken to check the reliability of the obtained solutions [1, 2]. Another difference between PCA and MCR-ALS is that in MCR-ALS the variances of the components are overlapped compared to the orthogonal variances of PCA. The sum of the explained variances from every individual MCR-ALS component is usually larger than the explained variance by all components when they are considered simultaneously in the model. Furthermore, in MCR-ALS the variances of the components are not ranked (permutation ambiguity) according to the amount of explained variance. Therefore, PCA and MCR-ALS methods can be both applied in a complimentary way.

The extension of PCA and MCR-ALS to the simultaneous analysis of the multiple data sets investigated in this work can be written in different ways according to the data mode (direction) kept in common. For instance in the case of the column-wise data matrix augmentation used to examine the

multiyear (4 years) changes in concentration of NO or O₃, the extension of the bilinear model is written as

$$\begin{aligned} D_{caug} &= \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} V^T + \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix} \\ &= U_{aug} V^T + E_{aug} \end{aligned} \tag{11}$$

Which in more compact (MATLAB notation) form can be also written as:

$$\begin{aligned} D_{caug} &= [D_1; D_2; D_3; D_4] \\ &= [U_1; U_2; U_3; U_4] V^T + [E_1; E_2; E_3; E_4] \\ &= U_{aug} V^T + E_{aug} \end{aligned} \tag{12}$$

In the case of the row- and column-wise matrix augmentation used to examine the changes in the concentration of O₃ and NO simultaneously for 2 years

$$\begin{aligned} D_{craug2-O_3-NO} &= \begin{pmatrix} D_{1,O_3} & D_{1,NO} \\ D_{2,O_3} & D_{2,NO} \end{pmatrix} \\ &= \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} (V_{O_3} \quad V_{NO}) + \begin{pmatrix} E_{1,O_3} & E_{1,NO} \\ E_{2,O_3} & E_{2,NO} \end{pmatrix} \end{aligned} \tag{13}$$

Which in compact (MATLAB notation) form can be written as:

$$\begin{aligned} D_{craug2-O_3-NO} &= [D_{1,O_3}, D_{1,NO}; D_{2,O_3}, D_{2,NO}] \\ &= [U_1; U_2] [V_{O_3}^T, V_{NO}^T] \\ &\quad + [E_{1,O_3}, E_{1,NO}; E_{2,O_3}, E_{2,NO}] \\ &= U_{aug} V_{aug}^T + E_{aug} \end{aligned} \tag{14}$$

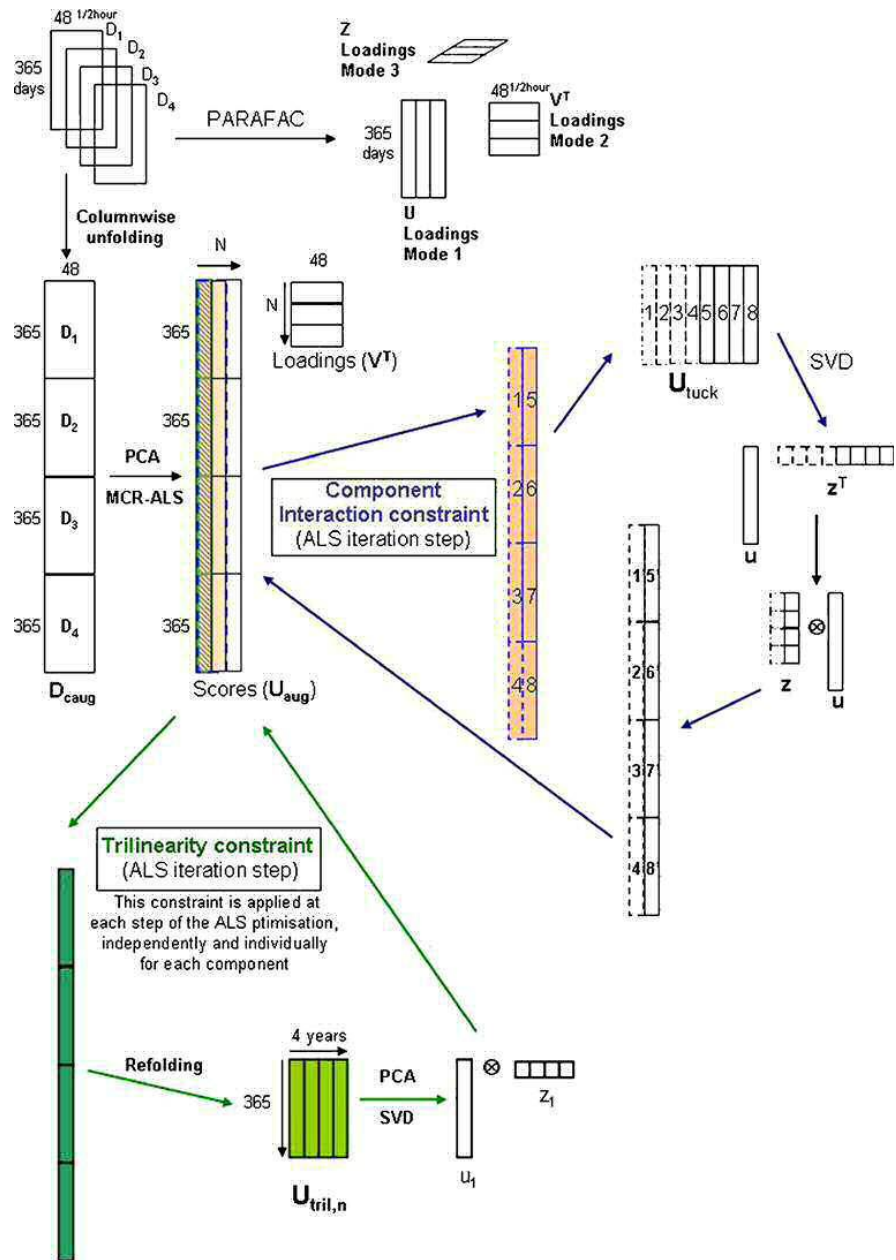
In Fig. 1, these different bilinear decompositions are shown for the different data matrix arrangements investigated in this work. Whereas for the 1 year (individual) data matrix of dimensions $\mathbf{D}(I, J)$, $I=365$ days and $J=48$ half-hours, a single scores (row) factor matrix $\mathbf{U}(I, N)$ with dimensions number of days in the rows by number of components in the columns is obtained, in the case of the column-wise (multiple years) augmented data matrices $\mathbf{U}_{aug}(I \times K, J)$, the scores matrix has dimensions of number of years (K) times number of days ($I=365$) in the rows by number of components, N , in the columns. From the scores matrix (\mathbf{U}_{aug}), the daily trends in one or several years (inter-annual) can be deduced for each component. The loadings matrix has dimensions of number of components, N , by half-hourly measurements (48). In the loadings matrices (\mathbf{V}^T), the hourly trends within a day of every component can be investigated.

In the scores factor matrix, U_{aug} , the concentration changes over one or several years (intra- and inter-annual information) are intermixed and consequently a new rearrangement is necessary to recover independently the temporal patterns corresponding to the changes over different years. In Fig. 2, this is shown. For every resolved component, its scores column vector can be selected, re-folded and analysed by SVD considering only the first component forcing ALS to resolve a single profile that fulfils the trilinear model (see also below in the “Trilinear

models in multivariate curve resolution” section for the trilinearity constraint in MCR-ALS). Using this simple procedure, the annual (vector–matrix U) and the inter-annual (vector–matrix Z) profiles are unmixed for each resolved component. From the intra-annual variations in U , it is also then possible to study the weekly, monthly and seasonally trends over years.

In the super (row- and column wise) augmented data matrix $D_{craug2-O3-NO}$, (see Fig. 1b) the bilinear decomposition is similar to previous cases but with some particular-

Fig. 2 Trilinearity and component interaction constraints in the MCR-ALS algorithm



ities. Now the number of rows in the scores matrix is equal to the number of years \times 365 and the number of columns is equal to the number of components, N . In the scores matrix the information about the two different years are still visible. In the loadings matrix, the number of rows is equal to the number of components and the number of columns is equal to 48 (half hours) \times 2. In this loadings matrix, the information about O₃ and NO is separated.

Trilinear models in multivariate curve resolution

The general expression of the trilinear model, to describe the decomposition of a three-way (cube) data set, is given (in element way) by Eq. 15 as

$$d_{ijk} = \sum_{n=1}^N u_{in}v_{jn}z_{kn} + e_{ijk} \quad (15)$$

where d_{ijk} represents the ijk th element in the three-way data set ($i=1, \dots, I, j=1, \dots, J$, and $k=1, \dots, K$), n is the number of components (chemical rank) common to the three modes ($n=1, \dots, N$), u_{in} , v_{jn} , and z_{kn} are the elements of \mathbf{U} , \mathbf{V}^T , and \mathbf{Z} factor matrices (loadings or component profiles in the three modes), and e_{ijk} is the residual term (part of the data not explained by the model). This trilinear model can be also described in a matrix way for each slice of the data cube as follows:

$$\mathbf{D}_k = \mathbf{U}\mathbf{Z}_k\mathbf{V}^T + \mathbf{E}_k, \mathbf{D}(I, J, K) = \{\mathbf{D}_k = 1, \dots, K\} \quad (16)$$

Other tensor notations for the trilinear model decomposition of the data cube can be found elsewhere [5].

As described by the matrix notation in Eq. (16), \mathbf{Z}_k will be the only factor matrix in the trilinear model that changes in the decomposition of the different matrices \mathbf{D}_k . This means that \mathbf{U} and \mathbf{V}^T matrices are the same for the simultaneously analysed matrices (slices of the cube) \mathbf{D}_k . This is a very strong constraint that enforces trilinear decompositions to give unique solutions under mild conditions [5]. However, this constraint is only fulfilled by true variance sources in some cases (it is not general).

MCR-ALS can be easily adapted to the analysis of three-way data sets formed by data matrices with the same row and column sizes (data cubes) and take advantage of their structure if they follow the premises of the trilinear model. Because of the inherent freedom in the modelling of the profiles of the augmented \mathbf{U}_{aug} matrix, the so-called trilinear structure is incorporated as an optional constraint [4] during the ALS optimization of the \mathbf{U}_{aug} profiles (see Fig. 2). An outstanding aspect of the trilinearity constraint in ALS is that the application of this constraint is performed separately and optionally for the profiles of every compo-

nent. Its implementation during ALS is easy and it provides fast solutions.

As shown in Fig. 2, when this trilinear constraint is applied, profiles of the same component in the different \mathbf{U}_{aug} matrices are forced to have a common shape during each iteration of the ALS optimization and only differ in a scaling factor. To do this, they are first placed and joined one beside the other to form a new one-component profile matrix, $\mathbf{U}_{\text{tril},n}$, with I rows (number of rows of each data matrix) and K columns (number of matrices simultaneously analysed), where n indicates what component is considered. For a full trilinear model, every component gives an $\mathbf{U}_{\text{tril},n}$ matrix ($n=1, \dots, N$), which is decomposed by SVD and only the first component is considered (implying that for this component, the shape of its year profile is exactly equal for the different years)

$$\mathbf{U}_{\text{tril},n}(I, K) = \mathbf{u}(1, 1)\mathbf{z}^T(1, K) \quad (17)$$

where \mathbf{u} now is the one column vector, which contains the common profile of the considered component (n) in the different K matrices, and \mathbf{z}^T is a row vector with the relative amounts of this profile in the different K matrices. After this, the full \mathbf{U}_{aug} augmented matrix is rebuilt and updated. \mathbf{u} and \mathbf{z} profiles will give the current estimation of the first and third mode loading profiles for the considered component in the trilinear model. Apart from forcing the shape of the loading vectors of the same component in the first mode to be the same, this procedure captures the relative intensity variation of this component in the loadings of the third mode, showing its scale differences in the different data matrices simultaneously investigated. This is precisely what trilinear models described by Eqs. 15 and 16 are considering explicitly. Therefore, results obtained by MCR-ALS with the trilinearity constraint [4] applied to all the components of the system will give similar results as the application of other trilinear model-based methods (such as the PARAFAC method, see Ref. [5]) when applied to the same system. An additional and obvious advantage of the trilinearity constraint implemented in MCR-ALS is that it may be applied independently and optionally to each component of the data set, giving more flexibility to data analysis, allowing for full trilinear models, and for partial trilinear models.

A test of data departure from the trilinear model can be deduced from differences between the fit obtained when the data is forced to fulfil the trilinear model compared to the fit obtained when data are only forced to fulfil the bilinear model (see in the “Results and discussion” section). When the data fulfils the trilinear model, both fits should be rather similar, apart from the slight overfitting tendency of bilinear models.

Component interaction models in multivariate curve resolution

The general expression for TUCKER3 models [5, 24] used to describe the decomposition of three-way data sets (element-wise) is given by Eq. 9:

$$d_{ijk} = \sum_{p=1}^{N_p} \sum_{q=1}^{N_q} \sum_{r=1}^{N_r} g_{pqr} u_{ip} v_{jq} z_{kr} + e_{ijk} \quad (18)$$

where d_{ijk} represents the ijk th element in the three-way data set, u_{ip} , v_{jq} , and z_{kr} are the elements in $\mathbf{U}(I, N_p)$, $\mathbf{V}(J, N_q)$, and $\mathbf{Z}(K, N_r)$ factor matrices (loadings in the three modes) used to reconstruct the d_{ijk} element of $\mathbf{D}(I, J, K)$, e_{ijk} is the residual term in $\mathbf{E}(I, J, K)$, and N_p , N_q , and N_r are the number of components considered in each of the three modes, which are not necessarily equal as in the trilinear model. g_{pqr} is the pqr element of the core array $\mathbf{G}(N_p, N_q, N_r)$, where the non null elements are spread out in different manners depending on each particular data set. In the PARAFAC trilinear model, there is a single profile per each component in each mode, whereas the TUCKER3 model has complete freedom in the correspondence of profiles among different components and modes. It may happen that two components interact and have the same profile in one of the modes. Matrix algebra representations of Eq. 18 need the introduction of tensor products such as the Kronecker product [5, 25] and this is not given here for brevity.

MCR-ALS can be adapted to fulfil simple restricted TUCKER3 models. In accordance with the TUCKER3 model described by Eq. (18), a different number of components in each of the three modes of the three-way data set are allowed as well as the interaction among them. Figure 2 shows graphically how this can be achieved. In the example shown in this figure, two of the column profiles of the augmented \mathbf{U}_{aug} matrix, are unfolded to give the row-wise augmented scores matrix \mathbf{U}_{tuck} with number of rows equal to the number of rows in the individual matrices (I) and with the number of columns equal to the two component times the number of matrices simultaneously analysed ($2 \times K$). This \mathbf{U}_{tuck} folded matrix containing the profiles of the two components is then decomposed by SVD and only the first component is considered

$$\mathbf{U}_{\text{tuck}} \mathbf{1}, 2 \times K \quad \mathbf{u} \mathbf{i}, 1 \quad \mathbf{z}^T \mathbf{1}, 2 \times K \quad 19$$

This decomposition gives directly the loading profile in the first mode $\mathbf{u}(I, 1)$, and the loading profile (scaling factor) in the third mode $\mathbf{z}^T(1, 2 \times K)$; Fig. 2). Note that in this case only one profile (one loading vector) will be recovered in the first mode. On the contrary, two profiles (two loading vectors) are recovered for the third mode, as well as for the second mode, in the \mathbf{V}^T matrix. The appropriate Kronecker product [5, 25] of these new loading vectors allows the estimation of the two new augmented profile vectors, which

will substitute the corresponding two columns of the \mathbf{U}_{aug} matrix. When this component interaction constraint is inserted during each step of the ALS iterative optimization procedure, the results obtained are analogous to those obtained by a TUCKER3 model with one component in the first mode and two components in the other two modes.

Summarising, three-way data set decompositions using either PARAFAC or TUCKER3 models yield explicitly loadings in three modes (\mathbf{U} , \mathbf{V}^T , and \mathbf{Z}). Instead, column-wise (similarly row-wise) augmented data matrix decompositions using bilinear models, such as PCA or MCR-ALS, give only loadings in the second mode (\mathbf{V}^T matrix), whereas the loadings in the other two modes (\mathbf{U} and \mathbf{Z}) are mixed in the augmented scores matrix (\mathbf{U}_{aug} matrix). Using the SVD analysis on the suitable augmented factor profiles, these loadings in the first (\mathbf{U}) and third (\mathbf{Z}) modes can be recovered and the results compared with those obtained by PARAFAC and TUCKER3 model-based methods.

Results and discussion

Data description

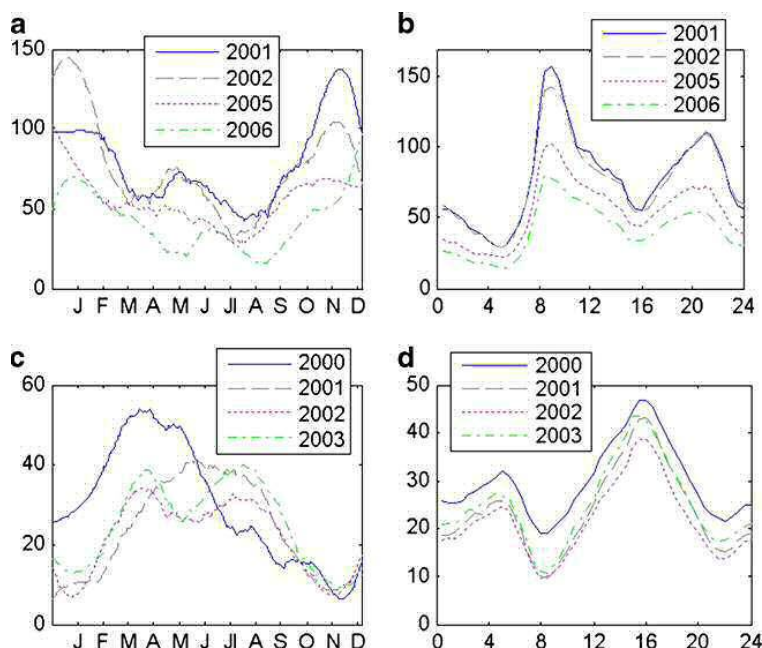
Daily and hourly averages were calculated for NO and O₃ concentration data (see Fig. 3). A Savitzky–Golay smoothing algorithm [26] was applied to these daily averages to facilitate the identification of seasonal patterns. When the averages of the 48 half-an-hour values of every year data matrix were calculated, daily profiles (365 values) were obtained. When averages were calculated for the 365 daily values of every year data matrix, half-hourly profiles (48 values) were obtained. NO averages ranged between 20 and 160 $\mu\text{g m}^{-3}$ and O₃ averages ranged between 10 and 55 $\mu\text{g m}^{-3}$.

In the daily profiles (Fig. 3a,c), the seasonal variation of NO concentration showed that the highest levels were registered during late autumn and winter, while the highest average values in the seasonal profiles of O₃ were registered in spring and summer.

In the hourly profiles (Fig. 3b,d), two maxima were clearly shown for NO concentration; one between 8:00 and 10:00 am and the other, lower and wider, around 21:00 hours. In the O₃ hourly profile, a marked minimum at 8:00 am was shown, coincident with the maximum NO concentration. O₃ hourly profiles showed also a maximum around 16:00, coincident with the NO minimum between the two maximum, and a minimum at 21:00 hours, coincident with the NO maximum.

Similar profiles have also been reported in previous studies for other urban locations [27–32] and with our previous results in Catalonia [8] and in Barcelona-Gràcia sampling station [7].

Fig. 3 **a** Daily and **b** hourly averages of NO concentration ($\mu\text{g m}^{-3}$) for the years 2001, 2002, 2005 and 2006; **c** daily and **d** hourly averages of O₃ measured concentrations ($\mu\text{g m}^{-3}$) for the years 2000, 2001, 2002 and 2003



Preliminary PCA and SVD results

The initial estimation of the number of components to be included in PCA and MCR analysis was obtained by SVD analysis of the different data sets previously described. A first selection of the number of components was simply chosen by visual observation of the plots of the sizes of the singular values and taking into account the explained variances. In all cases, PCA models were calculated with the proposed number of components as well as with a slightly higher and lower number of components. The number finally selected took into account the data fit and the interpretability of the resolved loading and score profiles.

In Table 1, explained variances for the different data arrangements and different number of components are given for PCA and MCR without any pre-treatment (see [electronic supplementary material](#)).

In order to check for the adequacy of the trilinear model to analyse the different data sets under study, SVD analysis was applied also to the two possible data matrix augmentation arrangements for every data set under study, column- and row-wise (Table 1). The comparison of the singular values obtained in each case showed that the trilinearity was not strictly fulfilled for $\mathbf{D}_{\text{caug-NO}}$ and $\mathbf{D}_{\text{caug-O}_3}$, and was not adequate for $\mathbf{D}_{\text{caug1-O}_3\text{-NO}}$ and $\mathbf{D}_{\text{craug2-O}_3\text{-NO}}$.

MCR-ALS results

Results of MCR-ALS will be shown separately for the three performed data arrangements, $\mathbf{D}_{\text{caug-NO}}$, $\mathbf{D}_{\text{caug-O}_3}$,

$\mathbf{D}_{\text{caug1-O}_3\text{-NO}}$ and $\mathbf{D}_{\text{craug2-O}_3\text{-NO}}$, previously described in section “Data arrangement” and Fig. 1.

NO study ($\mathbf{D}_{\text{caug-NO}}$ data matrix)

MCR-ALS analysis of matrix $\mathbf{D}_{\text{caug-NO}}$ was performed using three components, non-negativity and loading normalisation constraints, explaining a 90.1% of the total variance (Table 1). Note that whereas the total variance explained by MCR-ALS was practically equal to the variance explained by PCA, the sum of the variances explained by the individual MCR components increases up to 109.9%, with a variance overlap of 19.8% ($109.9 - 90.1 = 19.8\%$).

Loading and score profiles resolved by MCR-ALS are given in Fig. 4. In the hourly trends (loadings in second mode, see Fig. 4d–f), three NO profiles were resolved. The third component showed a peak around 9:00 hours which appeared during the rush traffic hour and it is assigned to fuel combustion by vehicles. NO concentration decreases until it reached a minimum at approximately 16:00. When O₃ reaches its maximum concentration, NO decreases, achieving its daily minimum value. Sunlight triggers a succession of photochemical reactions, which include the reaction of NO with O₃ to give NO₂ producing the depletion of O₃. After that, NO begins to accumulate again, reaching a new maximum during the evening [27], which appeared in the second component (Fig. 4e). An additional profile in the first component describing the NO night behaviour (constant along the year) was also resolved (Fig. 4d). This is attributed to the fact that during the night,

Table 1 SVD decomposition and explained variances for the different data matrices and models

	SVD		Explained variances			
	Column-wise ^a	Row-wise ^a	PCA ^b	MCR-ALS ^c	MCR-ALS ^d	MCR-ALS ^e
D_{caug-NO}^f (3 comp)	2.21E+04	2.02E+04	90.20%	90.10%	69.90%	65.50%
	6.39E+03	6.67E+03	PC1=78.5%	C1=17.4%	C1=8.9%	C1=11.1%
	5.63E+03	5.70E+03	PC2=6.6%	C2=40.8%	C2=33.8%	C2=34.9%
	3.59E+03	4.20E+03	PC3=5.1%	C3=51.7%	C3=46.1%	C3=36.6%
	2.89E+03	3.96E+03			PARAFAC=76.5%	TUCKER3 ₍₁₃₃₎ =65.7%
D_{caug-O3}^f (2 comp)	8.20E+03	7.72E+03	91.00%	91.00%	76.10%	73.40%
	2.55E+03	2.12E+03	PC1=83%	C1=66.9%	C1=53.5%	C1=57.6%
	1.21E+03	1.55E+03	PC2=8%	C2=32.9%	C2=44.2%	C2=24.5%
	1.16E+03	1.39E+03			PARAFAC=78.6%	TUCKER3 ₍₁₂₂₎ =73.5%
	8.50E+02	1.34E+03				
D_{caug1-O3-NO}^f (3 comp)	2.35E+04	2.12E+04	88.80%	88.60%	70.00%	–
	8.75E+03	1.04E+04	PC1=71.6%	C1=42.4%	C1=37.8%	
	7.51E+03	6.88E+03	PC2=9.9%	C2=19.9%	C2=17.7%	
	5.01E+03	4.96E+03	PC3=7.3%	C3=43.4%	C3=26.5%	
	2.97E+03	4.40E+03			PARAFAC=76.6%	
D_{caug2-O3-NO}^f (3 comp)	2.16E+04	2.24E+04	84.80%	83.30%	73.70%	–
	1.20E+04	8.03E+03	PC1=60.6%	C1=34.1%	C1=16.0%	
	6.64E+03	7.10E+03	PC2=18.6%	C2=38.2%	C2=38.4%	
	5.24E+03	5.60E+03	PC3=5.7%	C3=26.5%	C3=33.0%	
	3.81E+03	4.91E+03			PARAFAC=71.8%	

^a Augmentation mode constraint^b PCA for raw data^c MCR-ALS for raw data with non-negativity constraint^d MCR-ALS for raw data with non-negativity and trilinearity constraint^e MCR-ALS for raw data with non-negativity and interaction constraint^f Augmented data matrices (see Fig. 1 and explanation in section [Data arrangement](#))

photochemical reactions are not produced and this could lead to the accumulation of NO concentration. Consequently, the differences between summer and winter are not so marked. Other possible explanation of this night component could be attributed to the use of fuel-fed heating during the evening and the increase of traffic in the night of the weekends [7].

Loadings in the first mode described the NO trends along the year. The same seasonal daily trends (Fig. 4a–c) can be distinguished for the three components along the 4 years, with a pronounced U shape (intense minimums in April and in August). Several causes can justify these seasonal trends. One of them is the temperature variation associated to different seasons. In spring and summer the higher levels of solar radiation and temperatures promote photochemical reactions and the decreasing of NO concentration levels because of their reaction with maximum concentration levels of O₃. Another reason for the observed trends is the increase of NO emissions during winter due to the use of heating [27] and heavier traffic in cities compared to the summer and Easter vacations period. Loadings in the

third mode were describing the inter-annual trends (Fig. 4g–i) showing the tendency towards lower levels of NO concentration during the last years of the study (2005 and 2006).

MCR-ALS analysis for **D_{caug-NO}** was then performed using the trilinearity constraints (see method section “[Trilinear models in multivariate curve resolution](#)” and Fig. 2). In this case, the total explained variance decreased down to 69.9% (see Table 1). Loading profiles in the three modes were rather similar to those previously described without using the trilinearity constraint (see Fig. 4) and to those obtained by PARAFAC with non-negativity constraints (see Table 1 and supplementary information). PARAFAC core consistency (50.8%) showed that the model with three components departed significantly from the trilinear model, although the shapes of the resolved profiles were close to those obtained by bilinear MCR-ALS. The little difference between the explained variance in PARAFAC (76.5%) and MCR-ALS (69.9%) was due to the fact that the system is not fully trilinear and the two methods differ in the way they apply trilinearity.

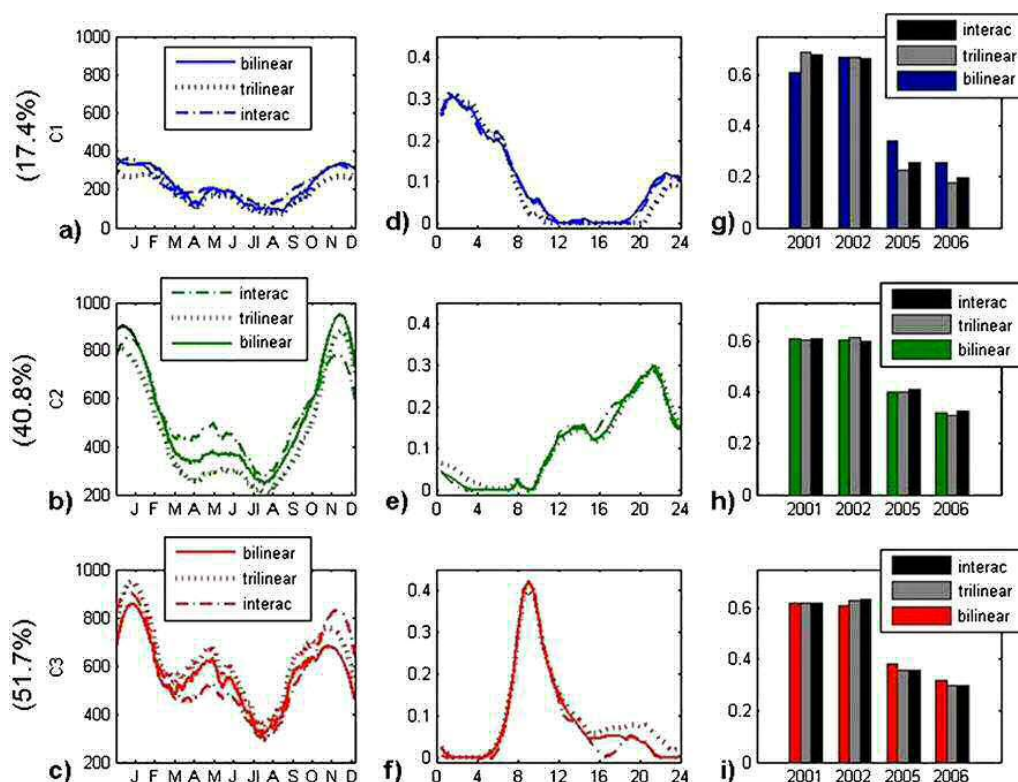


Fig. 4 MCR-ALS profiles for three components using non-negativity (*continuous line*), non-negativity and trilinearity (*dotted line*) and non-negativity and interaction constraints (*dashed line*) for NO augmented data matrix ($D_{\text{caug-NO}}$, see Fig. 2). The shown explained variances are

for the trilinear representation of the bilinear solution (see Fig. 2 and explanation in section “Multivariate data analysis methods”). **a–c** daily (within a year); **d–f** hourly (within a day) and **g–i** inter-annual profiles

Possible interaction between the three profiles in the first mode (133 model) was tested using the interaction constraint described in “Component interaction models in multivariate curve resolution” section and Fig. 2 for MCR-ALS. Using a restricted TUCKER3 model [5] with this interaction constraint for $D_{\text{caug-NO}}$ matrix, the total explained variance by MCR-ALS was 65.5% (Table 1) with a variance reduction from the trilinear model of only 4.4%. TUCKER3 of the same data using the same 133 model and applying non-negativity gave a 65.7% of explained variance. However, night profile could not be resolved appropriately by TUCKER3 (see supplementary information). The profiles of all the components on the first mode are forced to be the same and the interpretation of within a year profiles greatly simplified with this single profile. Therefore, this reduced model is considered to be more easily interpretable of the main yearly trends of NO contamination in the station under study.

O₃ study ($D_{\text{caug-O}_3}$ data matrix)

MCR-ALS analysis of matrix $D_{\text{caug-O}_3}$ using two components and non-negativity constraints, gave a total explained

variance of 91.0%. There is little variance overlap of 8.8% between the two components (see Table 1), which means that they did explain rather independent sources of variance in this case. When trilinearity constraints were also applied, the total explained variance was 76.1%, with a decreasing of the total explained variance of around 15%.

Loading and scores profiles obtained by MCR-ALS for $D_{\text{caug-O}_3}$ using non-negativity and using or not using trilinearity constraints are shown in Fig. 5. In both cases, they were similar. In the daily profiles (loadings in first mode), the seasonal trends of the two resolved components (Fig. 5a–b) showed clearly one peak maximum which is coincident with the maximum solar radiation registered in summer and spring months and practically reversed to the U shape shown for NO seasonal profiles (see Fig. 4).

MCR hourly resolved profiles (within a day, second mode loading profiles, Fig. 5c), for the first component (with the highest explained variance) showed a clear maximum around 16:00 hours (14:00 solar hours), coincident with the increase of solar radiation. Again, there was practically no difference in the shape of the profiles obtained when the trilinearity constraint was either applied

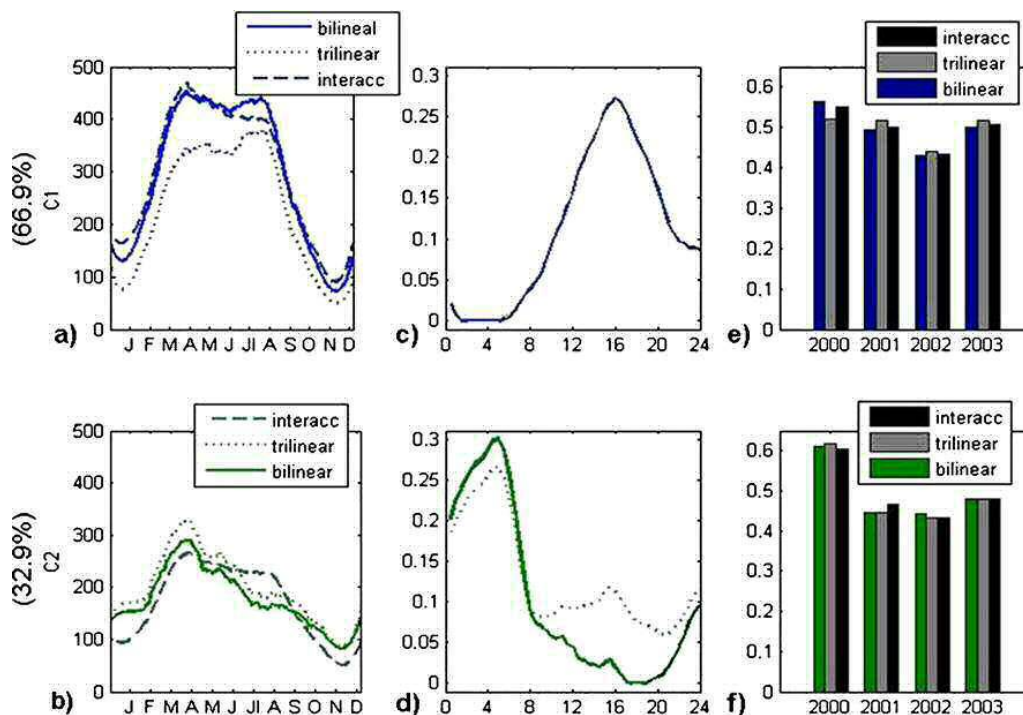


Fig. 5 MCR-ALS profiles for two components using non-negativity (continuous line), non-negativity and trilinearity (dotted line) and non-negativity and interaction constraints (dashed line) for O_3 augmented data matrix ($D_{\text{caug-}O_3}$, see Fig. 2). Shown explained variances are for

the trilinear representation of the bilinear solution (see Fig. 2 and explanation in section “Multivariate data analysis methods”). **a–b** daily (within a year); **c–d** hourly (within a day) and **e–f** inter-annual profiles

or not. The MCR hourly resolved profile for the second component (Fig. 5d) shows a different maximum corresponding to a second increase of O_3 concentration during the night. This pattern of O_3 increase in absence of solar radiation during the night can be attributed to the titration effect of NO emitted by car exhaust. At night, traffic decreases, NO emissions are reduced and consequently O_3 concentration increase. Inter-annual profiles (third mode loadings profile, Fig. 5e–f), showed that O_3 contamination patterns practically have remained rather constant during the years of study.

In the results obtained by PARAFAC for $D_{\text{caug-}O_3}$ matrix (Table 1), the core consistency test (92.6%) indicated that the trilinear model with two components was rather consistent. For that reason, explained variances in PARAFAC (78.6%) and MCR-ALS (76.1%) were closer. Resolved loading profiles in the three modes (see supplementary information) were practically equal to those obtained by MCR-ALS using the trilinearity constraint.

However, since resolved seasonal profiles (loadings in first mode) were rather similar, interaction of the two profiles in this first mode was attempted. MCR-ALS results of this model 122, using the interaction constraint previously described (see

“Component interaction models in multivariate curve resolution” section and Fig. 2) gave a total explained variance of 73.4% (Table 1). Loading profiles are shown in Fig. 5. Results obtained by restricted TUCKER3 applying non-negativity gave practically the same explained variance (73.5%) and similar profiles (see supplementary information) to these obtained by MCR-ALS with the interaction constraint (see Table 1). However, using this interaction constraint a better reduction and summary of the complexity of the model was achieved. First mode is now described by a single profile (seasonal trends) without significant loss of information.

O_3 and NO simultaneous study ($D_{\text{caug1-}O_3\text{-NO}}$ and $D_{\text{craug2-}O_3\text{-NO}}$ data matrices)

Possible correlation between O_3 and NO data sets and interaction effects between these two pollutants in the sampling station under study were investigated using the data arrangements described in Section Data arrangement and Fig. 1 for $D_{\text{caug1-}O_3\text{-NO}}$ and $D_{\text{craug2-}O_3\text{-NO}}$ data matrices. Three components were finally selected in both cases.

MCR-ALS analysis of $D_{\text{caug1-O}_3\text{-NO}}$ with only non-negativity constraints gave a total explained variance of 88.6% (Table 1). When the trilinearity constraint was applied, the total explained variance decreased down to 70.0%. This decrease of 18.6% of the explained variance confirmed the poor trilinear behaviour of the system, also seen previously in the different sizes of the singular values obtained by SVD (Table 1) and in the poor PARAFAC consistency test (51.3%). Figure 6 shows that the first component is mostly describing *NO pollution*, the second component is mostly describing *O₃ pollution*, and the third component, is describing the possible *interaction between O₃ and NO pollution*. This last third component has mixed contributions of both contaminants (Fig. 6i) but it is mainly influenced by O₃ (in Fig. 7d this interaction is seen better). Seasonal and hourly trends are similar to those previously described in sections *NO study* ($D_{\text{caug-NO}}$ data matrix) and *O₃ study* ($D_{\text{caug-O}_3}$ data matrix). Since all components showed profiles with different shapes, component interaction constraint (TUCKER3 models) was not possible in this case.

MCR-ALS analysis of $D_{\text{craug2-O}_3\text{-NO}}$ with non-negativity constraints explained 83.3% variance (Table 1). When the trilinearity constraint was applied, the total explained variance decreased down to 73.7%. MCR-ALS resolved profiles are given in Fig. 7 with and without the trilinearity constraint. Loadings of the first component in second mode (Fig. 7a), showed again the interaction between NO and O₃. The changes along the year of this first component (first mode loadings) with a strong contribution of the two chemical compounds (second mode loadings) are rather constant independently of sun light radiation effects whereas in the second component only appeared the NO contribution, and in the third component only appeared the O₃ contribution. In the hourly and daily profiles (loadings in first and second modes, Fig. 7b–c and e–f) of these two components, the sun light and seasonal effects on O₃ and NO contamination are described, confirming what was previously encountered in their separate analysis in Figs. 4 and 5. From the between (inter) year profiles (third mode loading profiles), the contamination by NO (Fig. 7h) remained rather constant for these 2 years (2001 and

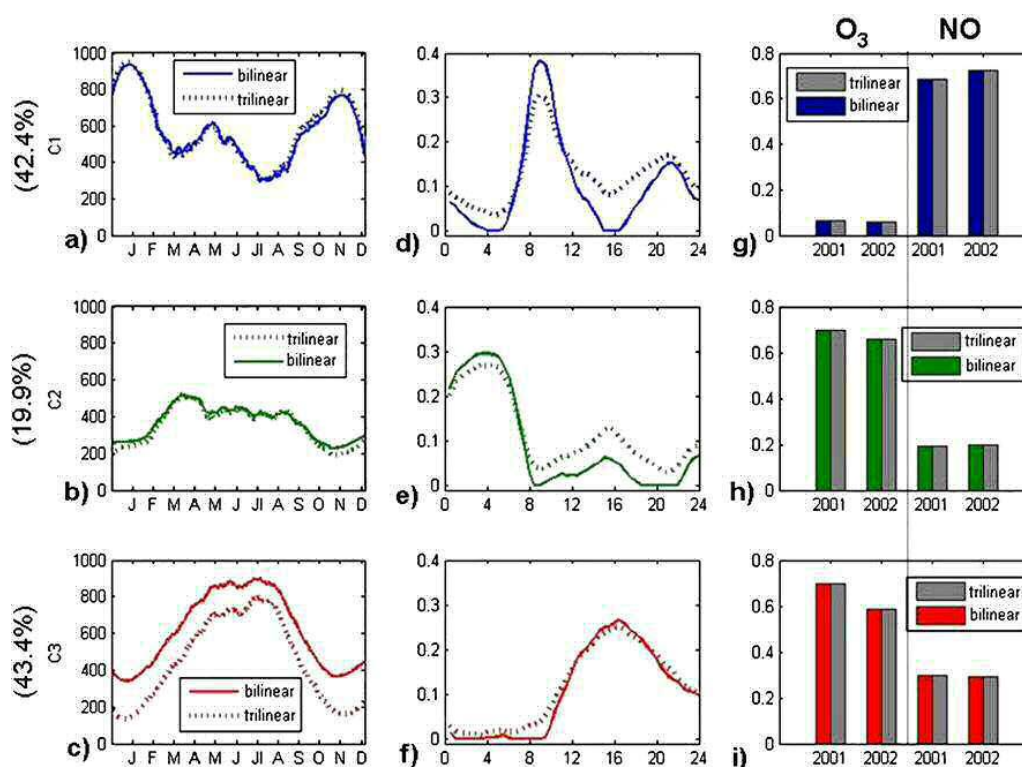


Fig. 6 MCR-ALS profiles for three components using non-negativity (continuous line) and non-negativity and trilinearity constraints (dotted line) for NO and O₃ augmented data matrix ($D_{\text{craug1-O}_3\text{-NO}}$, see Fig. 2). Shown explained variances are for the trilinear

representation of the bilinear solution (see Fig. 2 and explanation in section “Multivariate data analysis methods”). a–c daily (within a year); d–f hourly (within a day) and g–h yearly profiles (between years)

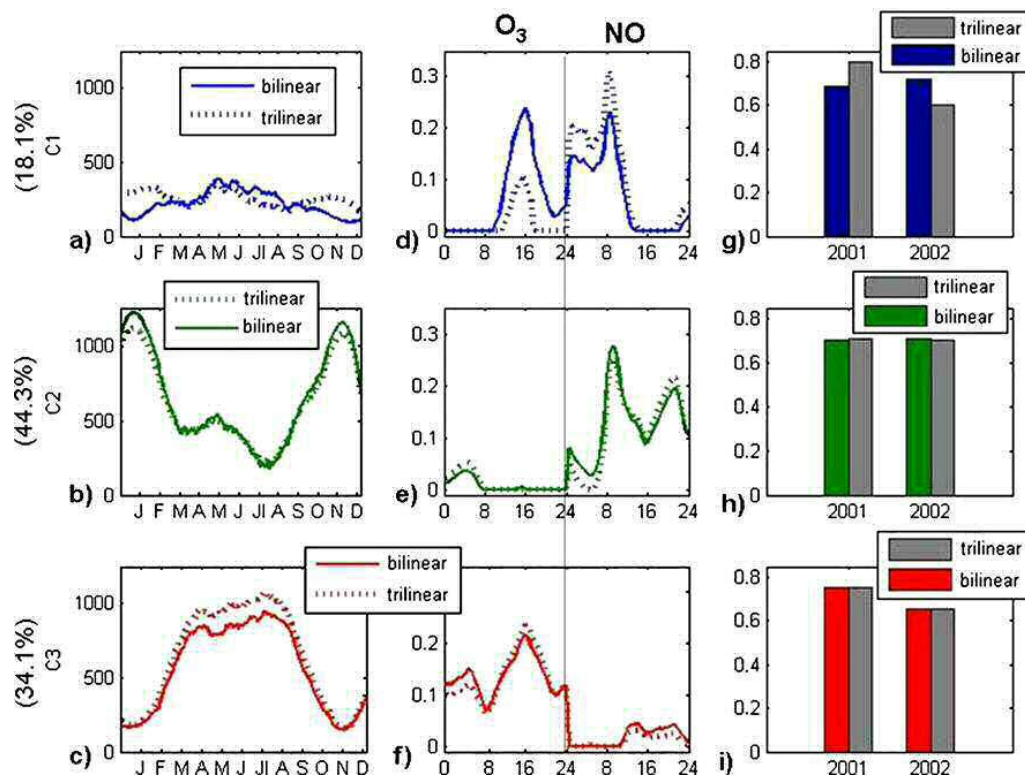


Fig. 7 MCR-ALS profiles for three components using non-negativity (continuous line) and non-negativity and trilinearity constraints (dotted line) for NO and O₃ augmented data matrix ($D_{\text{caug2-O}_3\text{-NO}}$, see Fig. 2). Shown explained variances are for the trilinear

representation of the bilinear solution (see Fig. 2 and explanation in section *Multivariate data analysis methods*). **a–c** daily (within a year); **d–f** hourly (within a day) and **g–h** yearly profiles (between years)

2002), but the contamination by O₃ (Fig. 7i) was apparently lower in the second year (2002). In this case also, interaction constraints were not either applied for the same reason as before for $D_{\text{caug1-O}_3\text{-NO}}$: profiles of the components in the three modes were rather different in all cases.

Conclusions

MCR-ALS with different constraints like trilinearity and component interaction produced results analogous to well-established methods like PARAFAC and restricted TUCKER3 model-based methods in the analysis of environmental multiway data sets. The advantages offered by the application of these constraints are, on one side, the better and simpler interpretability of the profiles (more summarised and concise) and on the other side the fact that especially in the case of PARAFAC the resolved profiles are unique. An additional advantage of the

trilinearity constraint implemented in MCR-ALS is that it may be applied independently and optionally to each component of the data set, giving more flexibility to data analysis, allowing for full trilinear models, and for partial trilinear models.

Resolved MCR profiles using these constraints have been shown to describe adequately the different patterns and evolution of O₃ and NO contamination during the different hours of the day and during the different seasons of the year (hourly and daily variations). The more important hourly O₃ resolved MCR profile showed a maximum around 16:00 hours and the more important hourly NO profile showed two maximums, one at 9:00 hours and another around 20:00 hours. Interaction between O₃ and NO profiles is also demonstrated by the complimentary shapes of the profiles of these two compounds. The O₃ profile maximum corresponded to a profile minimum of NO. The complimentary relation between NO and O₃ was also seen in the mixed behaviour of some components.

Acknowledgements The authors acknowledge the Spanish Ministry of Education for the award of the research project *CTQ2009-11572* and the Generalitat of Catalunya, Department of Environment for providing the different databases.

References

- Rutan SC, de Juan A, Tauler R, Vol.2, Chapter 2.15 and Tauler R, Maeder M, de Juan A, Vol.2, Chapter 2.24 in : Brown SD, Tauler R, Walczak B (2009) Comprehensive chemometrics: chemical and biochemical data analysis four-volume set. Elsevier
- Tauler R, de Juan A, Chapter 11, in: Gemperline P (2006) Practical guide to chemometrics. Taylor & Francis Group, LLC
- Tauler R, Smilde A, Kowalski B (1995) Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *J Chemometr* 9:31
- Tauler R, Marqués I, Casassas E (1998) Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths. *J Chemometr* 12:55–75
- Smilde A, Bro R, Geladi P (2004) Multi-way analysis: applications in the chemical sciences. Wiley, West Sussex
- Peré-Trepat E, Ginebreda A, Tauler R (2007) Comparison of different multiway methods for the analysis of geographical metal distributions in fish, sediments and river waters in Catalonia. *Chemometr Intell Lab Syst* 88:69–83
- Alier M, Felipe-Sotelo M, Hernández I, Tauler R (2009) Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods. *Anal Chim Acta* 642:77–88
- Felipe-Sotelo M, Gustems L, Hernández I, Terrado M, Tauler R (2006) Investigation of geographical and temporal distribution of tropospheric ozone in Catalonia (North-East Spain) during the period 2000–2004 using multivariate data analysis methods. *Atmos Environ* 40:7421
- Álvarez E, de Pablo F, Tomás C, Rivas L (2000) Spatial and temporal variability and ground-level ozone in Castilla-León (Spain). *Int J Biometeorol* 44:44–51
- Ribas A, Peñuelas J (2004) Temporal patterns of surface ozone levels in different habitats of the North Western Mediterranean basin. *Atmos Environ* 38:985–992
- Itano Y, Bandow H, Takenaka N, Saitoh Y, Asayama A, Fukuyama J (2007) Impact of NO_x reduction on long-term ozone trends in an urban atmosphere. *Sci Total Environ* 379:46
- <http://www.epa.gov/air/urbanair/nox>
- <http://mediambient.gencat.net>
- Gao HO (2007) Day of week effects on diurnal ozone/NO_x cycles and transportation emissions in Southern California. *Transport Res Part D: Transport Environ* 12:292
- Chang SC, Lee CT (2008) Evaluation of the temporal variations of air quality in Taipei City, Taiwan, from 1994 to 2003. *J Environ Manage* 86:627
- González Gallero FJ, Galán Vallejo M, Umbría A, Baena Gervilla J (2006) Multivariate statistical analysis of meteorological and air pollution data in the ‘Campo de Gibraltar’ region, Spain. *Environ Monit Assess* 119:405
- Ozeki T, Ihara T, Ogawa N (2006) Study of pollutants in precipitation (rain and snow) transported long distance to west coasts of Japan Islands using oblique rotation factor analysis with partially non-negative constraint. *Chemometr Intell Lab Syst* 82:15
- ISO 13964:1998; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=23528
- ISO 7993:1985; UNE 77212:1993; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=14995; http://www.aenor.es/desarrollo/normalizacion/normas/buscador_normas.asp
- Jolliffe IT (2002) Principal component analysis. Springer, New York
- Vandeginste BGM, Massart DL, Buydens LMC, de Jong S, Lewi PJ, Smeyers-Verbeke J (1998) Handbook of chemometrics and qualimetrics. Elsevier, The Netherlands
- Golub GH, Van Loan CF (1989) Matrix computations. John Hopkins University Press, Baltimore
- de Juan A, Tauler R (2003) Chemometrics applied to unravel multicomponent processes and mixtures—revisiting latest trends in multivariate resolution. *Anal Chim Acta* 500:195
- Tauler R (1995) Multivariate curve resolution applied to second order data. *Chemometr Intell Lab Syst* 30:133
- Windig W, Guilment J (1991) Interactive self-modeling mixture analysis. *Anal Chem* 63:1425–1432
- Kroonenberg P (1983) Three mode principal component analysis. DSWO, Leiden
- An BDS (1995) Introduction to tensor products with applications to multiway data analysis. *Chemom Intell Lab Syst* 28:229–237
- Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627
- Tu J, Xia ZG, Wang HS, Li WQ (2007) Temporal variations in surface ozone and its precursors and meteorological effects at an urban site in China. *Atmos Res* 85:310
- Pudasainee D, Sapkota B, Shrestha ML, Kaga A, Kondo A, Inoue Y (2006) Ground level ozone concentrations and its association with NO_x and meteorological parameters in Kathmandu valley, Nepal. *Atmos Environ* 40:8081
- Zabalza J, Ogulei D, Elustondo D, Santamaría JM, Alastuey A, Querol X, Hopke PK (2007) Study of urban atmospheric pollution in Navarre (Northern Spain). *Environ Monit Assess* 134:137
- Jo WK, Park JH (2005) Characteristics of roadside air pollution in Korean metropolitan city (Daegu) over last 5 to 6 years: temporal variations, standard exceedances, and dependence on meteorological conditions. *Chemosphere* 59:1557
- Saito S, Nagao I, Tanaka H (2002) Relationship of NO_x and NMHC to photochemical O₃ production in a coastal and metropolitan areas of Japan. *Atmos Environ* 36:1277
- Raddatz RL, Cummine JD (2001) Temporal surface ozone patterns in urban Manitoba, Canada. *Boundary-Layer Meteorol* 99:411

Analytical and Bioanalytical Chemistry

Electronic Supplementary information

Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O₃ pollution in Barcelona

Marta Alier, Mónica Felipe, Isabel Hernández and Romà Tauler

Comment to Fig. S1. PCA profiles for NO augmented data matrix $D_{\text{caug-NO}}$ were more difficult to interpret than MCR-ALS profiles given in Figure 4, because negative values had no meaning in terms of environmental pollution. Hourly profiles in the second and third principal components resemble resolved MCR-ALS for the morning and night NO profiles (see Figure 4)

Comment to Fig. S2. NO resolved profiles using PARAFAC and TUCKER3 (133 model) methods in the analysis of augmented data matrix $D_{\text{caug-NO}}$ can be compared with those obtained by MCR-ALS analysis of the same data matrix applying trilinearity and component interaction constraints (see Figure 4). Resolved daily profiles were similar for the three methods, except for hourly profiles. MCR-ALS hourly profiles revealed a different separate night trend (Figure 4d)

Comment to Fig. S3. Interpretation of PCA profiles for O₃ augmented data matrix $D_{\text{caug-O}_3}$ were again more difficult to interpret than those obtained by MCR-ALS (see Figure 5). Hourly profile of the second principal component resemble MCR-ALS night profile (see Figure 5)

Comment to Fig. S4. O₃ resolved profiles using PARAFAC and TUCKER3 (122 model) methods in the analysis of augmented data matrix $D_{\text{caug-O}_3}$ can be compared with those obtained by MCR-ALS analysis of the same data matrix applying trilinearity and component interaction constraints (see Figure 5). Resolved daily profiles were similar for the three methods. PARAFAC and TUCKER3 (122 model) second component hourly profiles showed differences (Figure 11d), and also with the corresponding MCR-ALS profile that revealed again a separate night trend (Figure 5d)

Fig. S1. PCA profiles using three components for NO augmented data matrix ($\mathbf{D}_{\text{caug-NO}}$, see Fig.2). (a–c) daily (within a year); (d–f) hourly (within a day) and (g–h) inter-annual profiles

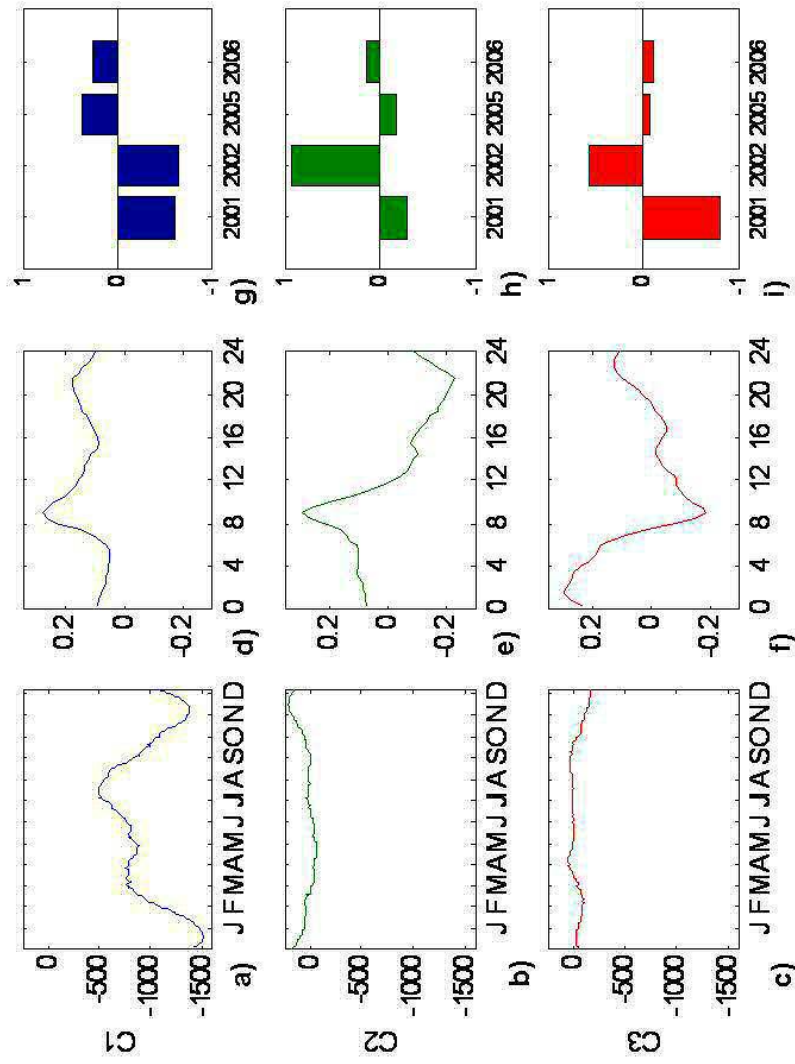


Fig. S2. PARAFAC (continuous line) and TUCKER3 (model 133, dotted line) profiles using three components (using non-negativity) for NO augmented data matrix ($D_{\text{caug-NO}}$, see Fig.2). (a–c) daily (within a year); (d–f) hourly (within a day) and (g–i) inter-annual profiles

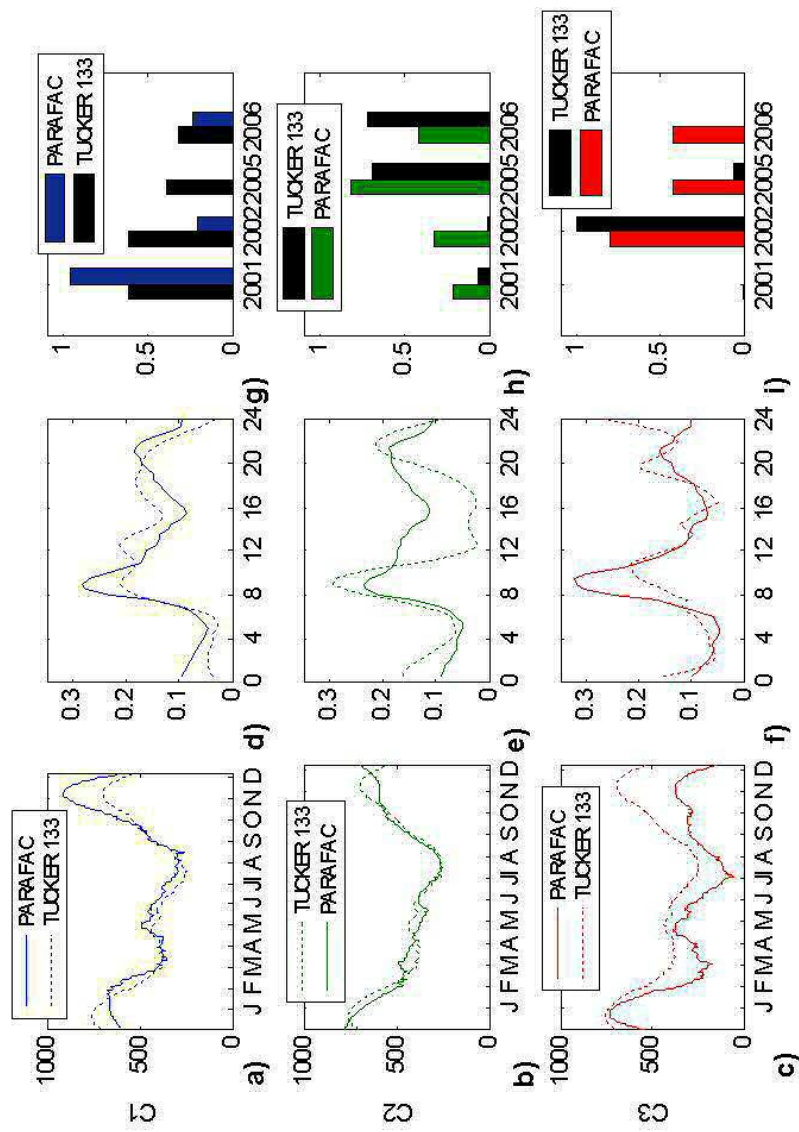


Fig. S3. PCA profiles using two components for O_3 augmented data matrix ($D_{\text{aug-}03}$, see Fig.2). (a–b) daily (within a year); (c–d) hourly (within a day) and (e–f) inter-annual profiles

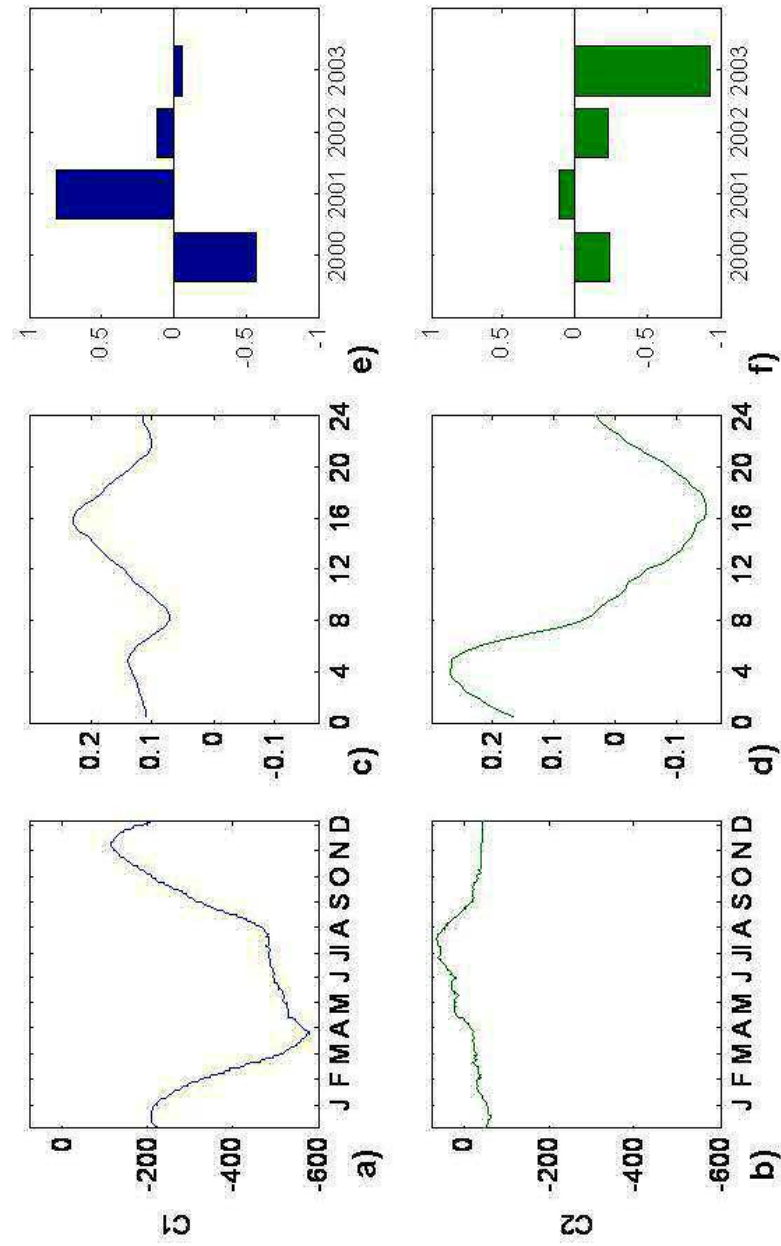
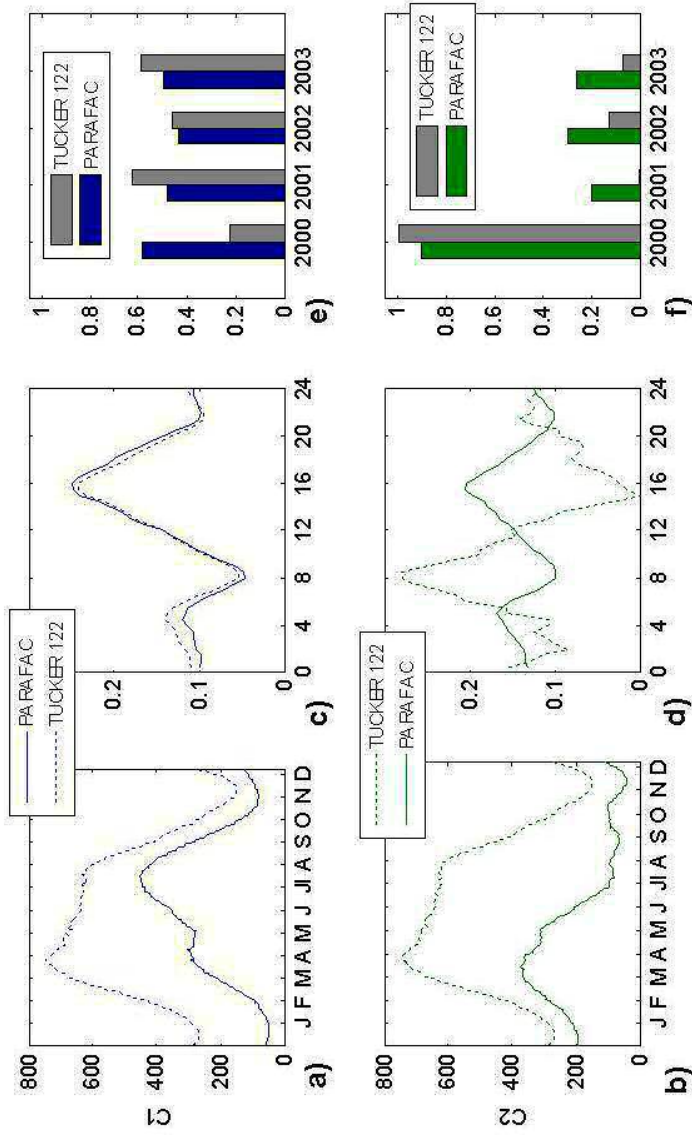


Fig. S4. PARAFAC (continuous line) and TUCKER3 (model 133, dotted line) profiles MCR-ALS profiles using two components (using non-negativity) for O₃ augmented data matrix ($D_{\text{caug-03}}$, see Fig.2). (a–b) daily (within a year); (c–d) hourly (within a day) and (e–f) inter-annual profiles





Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolabMultivariate Curve Resolution of incomplete data multisets[☆]Marta Alier, Romà Tauler^{*}

Department of Environmental Chemistry, IDAEA-CSIC, C/Jordi Girona, 18-26, 08034, Barcelona, Spain

ARTICLE INFO

Article history:

Received 5 February 2013
Received in revised form 7 May 2013
Accepted 15 May 2013
Available online 23 May 2013

Keywords:

Incomplete data multisets
Multivariate Curve Resolution
Alternating Least Squares
Data multiset
Multiway data

ABSTRACT

In this paper the application of the Multivariate Curve Resolution Alternating Least Squares method (MCR-ALS) to incomplete data multisets is explored. The experimental incomplete data multiset studied in this work is taken from a previous multiannual atmospheric monitoring study of the changes of ozone and nitrogen oxide concentrations in an air quality sampling station located in the city of Barcelona, in which some of the individual data sets were missing. Based on the preliminary results obtained in this study, new data multisets, complete and incomplete, with different levels of noise were simulated and analysed by a new variant of the MCR-ALS method which optimises a combined error function including all possible complete data subsets derived from the original incomplete data multiset. Conclusions are drawn about the effects of data completeness on the results obtained for different noise levels and on the viability of trilinear models.

© 2013 Published by Elsevier B.V.

1. Introduction

The problem of incomplete data sets is ubiquitous in many application areas like network traffic analysis, bibliometrics, social network analysis, chemometrics and computer vision [1,2]. These missing data can appear due to many reasons such as for instance malfunctioning of the analytical techniques or instruments, electrical disconnections, wrong or limited data acquisition planning strategies or loss of a signal due to contamination with high amounts of mechanical noise [3]. Imputation of missing values can be useful as long as the amount of missing data is small, however performance degrades for large amounts of missing data [4] or when this is systematically produced. The problem of incomplete data can be very difficult for chemometric analysis, and here we have developed a new Multivariate Curve Resolution method to analyse incomplete data multisets where some of the data subsets (full data matrices) are missing. Data multiset (or multiblock) is composed by multiple data sets (or multiple data blocks). Each individual data set or data block has its own entity and it has a regular structure. Incomplete data multiset (or incomplete data multiblock) refer to cases where some of these individual data sets (or data blocks) are missing. For instance, in the case under study, each individual data set (data block) refers to one year ozone or NO data measurements, giving a data matrix of the same dimensions (same number of rows and columns). If one of these individual data sets, either O₃ or NO (but not

both) is missing, the resulting data structure is named incomplete data multiset. Matrix augmentation strategies will give then data structures with missing complete data sets (blocks, submatrices, see Fig. 1c).

Multivariate Curve Resolution (MCR) methods are now well accepted as useful techniques in chemometrics to investigate general mixture analysis problems in chemistry, as in source apportionment studies in atmospheric sciences [5], in component spectra unmixing in hyperspectral imaging [6], in resolving species and analyte concentration and spectra profiles in analytical chemistry [7] or in factor analysis problems in general [8]. The goal of all MCR methods is the separation of the pure component contributions from their unknown mixture. The minimum assumption in MCR is the fulfilment of a bilinear model under natural constraints such as non-negativity, unimodality or closure, in contraposition to the stronger mathematical constraints like orthogonality (in Principal Component Analysis, [9]) or independence (in Independent Component Analysis [10]). Constraints applied in MCR methods, however, do not assure the estimation of unique solutions and rotation ambiguities can be present [11–13]. These ambiguities can be overcome when higher structured multiway data sets are available and approaches exploiting these structures by means of trilinear or multilinear models are applied [14,15]. In Chemistry, this is sometimes a too strong requirement which is only fulfilled in some circumstances, like in excitation-emission spectroscopy [16]. A more general situation is to have data structures where the multilinear (trilinear) models are fulfilled only partially by some of the components of the system or where interactions between the different components of the system may exist [17,18]. The Multivariate Curve Resolution method presented in this work allows analysing incomplete data multisets fulfilling or not, or partly, multilinear models.

From its initial development, MCR-ALS has been used for the analysis of multiset and multiway data [19]. In fact, most of the successful

[☆] This paper was presented at the TRICAP 2012 meeting held in Bruges in June of 2012.

^{*} Corresponding author.

E-mail addresses: Marta.Alier@idaea.csic.es (M. Alier), Roma.Tauler@idaea.csic.es (R. Tauler).

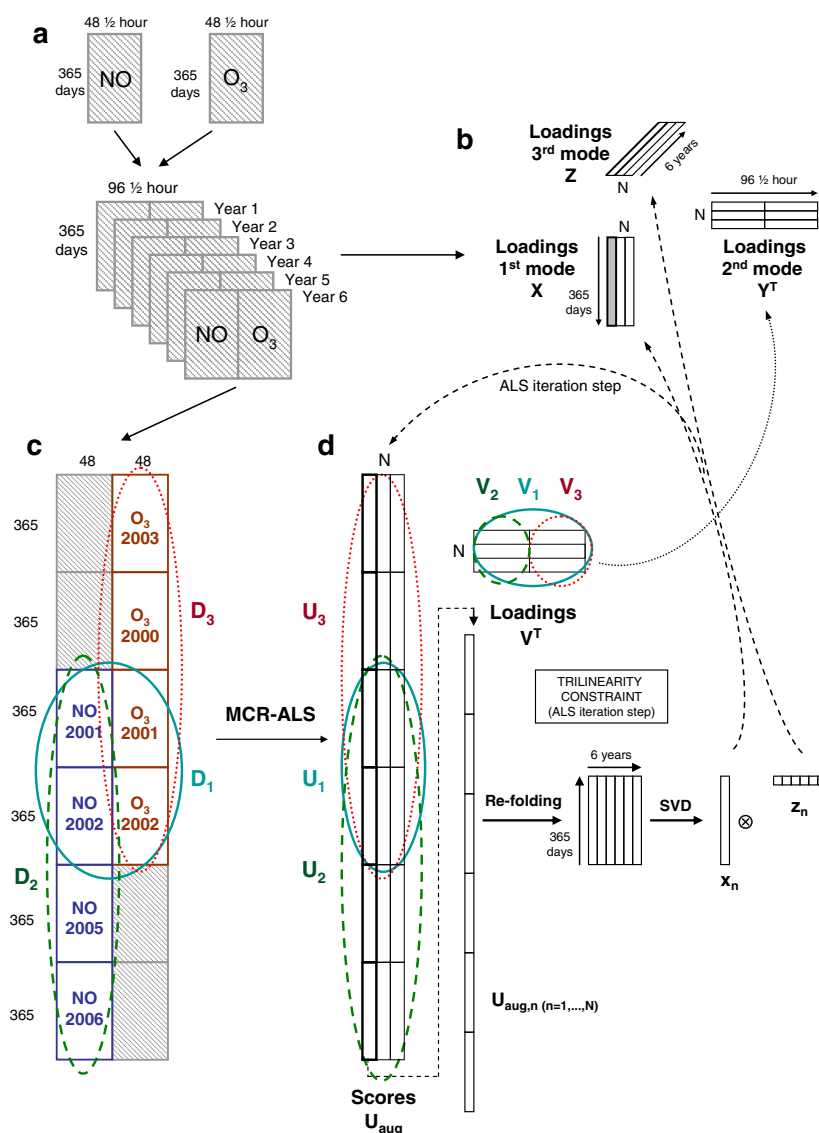


Fig. 1. Incomplete data multiset arrangement and its MCR-ALS bilinear decomposition using the trilinearity constraint. Recovery of the three mode profiles \mathbf{X} , \mathbf{Y} and \mathbf{Z} , from bilinear augmented scores \mathbf{U}_{aug} and loadings \mathbf{V}^T . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

applications of MCR-ALS are related to the analysis of data multisets. Rotation ambiguities associated to MCR methods can be considerably reduced or even eliminated when data multiset are analysed using this procedure. Data multisets can be arranged in different type of row-wise, column-wise or row- and column-wise augmented matrices. Constraints like trilinearity [14,15], quadrilinearity [20] or interaction between components [17,18] have been proposed and implemented successfully in the MCR-ALS method to analyse and take advantage of cases where data fulfil these higher ordered structures, with analogous results to those obtained by standard multiway methods like PARAFAC [21] or Tucker [22] multilinear decompositions. In all these situations however, it has been always assumed that the data have a complete structure (without missing data sets). In contrast, in the present paper, we present a new MCR-ALS procedure to analyse incomplete data multisets, where some data sets or data blocks are missing.

2. Data

In the present paper, incomplete experimental and simulated data multisets have been used to test the application of MCR-ALS on them. The experimental data used in this work consisted on ozone (O_3) and nitrogen oxide (NO) concentrations recorded from the network of atmospheric pollution control sampling stations operated by the Department of Environment of the Catalan Autonomous Government [23,18]. In this network, different compounds and environmental parameters are analysed every half-hour during the whole day and in different stations. The Catalan Autonomous Government can add or remove stations due to practical and administrative reasons. Failures of stations and/or a malfunction of them can cause absence of measurements for hours, days or months. In this work, we have applied MCR to the resulting incomplete data multisets in which several data matrices

were missing. Taking as a model the NO and O₃ experimental data multisets and their MCR-ALS results, new similar simulated data multisets were built to test the developed procedure. The experimentally available data for the selected station sets covered the years 2001, 2002, 2005 and 2006, for NO and 2000, 2001, 2002 and 2003 for O₃. Years 2000 and 2003 were missing for NO, and years 2005 and 2006 were missing for O₃.

2.1. Experimental data multisets

In the atmospheric monitoring network of Catalonia, different parameters such as temperature, relative humidity, wind direction, solar radiation and several pollutants such as nitrogen oxides, carbon dioxide or sulphur dioxide are measured daily during the whole year in different stations spread over the territory (see <http://mediambient.gencat.net> for further information).

In this study, half-hourly NO and O₃ concentrations measured from January 1st, 2000 to December 31st, 2006, in a single sampling station located in the city of Barcelona (district of Gràcia) were used. This station was selected because it is close to the centre of the city and it is highly populated. Ozone concentrations were measured by means of UV photometry according to ISO 13964:1998 [24], and they were also acquired by automatically operated analysers. These analysers provide ozone measurements every minute and they are saved in half-hourly periods. NO concentrations were measured according to the UNE method 77212:1993 (equivalent to ISO 7993:1985) [25], using automatically operated chemiluminescence analysers.

In a previous study [18] these two data sets were analysed using MCR-ALS, first separately for NO and O₃, using a column-wise data matrix arrangement and then jointly, using a row- and column-wise super data matrix arrangement. However in the later case, the simultaneous NO and O₃ analysis could be only performed for the common years of recording (2001 and 2002) of the two chemicals. For the NO and O₃ data, years 2000 and 2003, and 2005 and 2006 were missing respectively. Discussions of these results are given in detail in [18].

2.2. Simulated data multisets

In order to do the validation of the MCR-ALS algorithm, profiles resolved by this method should be compared with the correct ones. This is only possible when simulated data are generated using with previously known profiles. In this work, these previously known profiles were taken from those obtained in a previous MCR-ALS analysis of experimental data [18]. Once the whole simulated data multiset was generated with these previously known profiles, some of the data sets included in the data set were eliminated, and the incomplete data set was created.

In our previous analysis, experimental data sets corresponding to years in common (2001 and 2002) for NO and O₃ were investigated (**D1** in Fig. 1). Before the analysis, experimental data were smoothed with a Savitzky–Golay filter [26]. The profiles of the three components in the three modes (see below in Section 3.3) were obtained from MCR-ALS analysis using non-negativity and trilinearity constraints. The resolved first mode profile was smoothed again to remove the noise propagated during the MCR-ALS analysis of the experimental data. In order to generate the complete simulated data multisets, a third mode profile with the contribution of the six years (2000, 2001, 2002, 2003, 2005 and 2006) for every component was generated following the trend observed for the available years in a previous study [18]. This simulated third mode was then appropriately multiplied by the first mode profile obtained before to obtain the simulated scores for each individual data set. Finally, these simulated two-way scores were multiplied by the loadings resolved by the MCR-ALS analysis of the experimental data to generate the complete individual data sets for the six data matrices of dimensions 365 (days of the year) by 48

(half an hour measures over every day). Random normal noise with standard deviation at different levels (0.1%, 5%, 10% and 20% of the mean of simulated data) was added to these individual matrices. The complete data set with all year matrices included and with the different levels of noise was then converted to the incomplete data multiset by eliminating the individual matrices for some of the years. Thereby, the incomplete simulated data multiset was built to be equivalent in structure to the incomplete experimental data multiset described above.

2.3. Data arrangement

All individual data matrices for NO and O₃, experimental and simulated had dimensions 365 × 48 (year days × half-hourly measurements). Before the data multiset was arranged, O₃ matrices were multiplied by 3 to equate the data scale of both pollutants. Other traditional pre-treatments (like equal variance scaling) to remove scale differences were not possible due to the incomplete data structure. For the incomplete experimental data multiset, the individual data matrices were initially smoothed as previously described for simulated data in Section 2.2. In Fig. 1, the scheme used for data arrangement and data analysis is summarised. In the top left, the different individual data matrices for NO and O₃ were appended to form the different one year row-wise augmented data matrices. When the data multiset is complete, these different year data matrices can be arranged in three directions, ways or modes forming a data cube as shown in the Fig. 1b. The first mode of the data cube (rows of the data matrices) corresponds to the annual profiles (within year daily values), i.e. values for 365 days, and it shows the monthly and seasonal changes. The second mode (columns of the data matrices) corresponds to the within day values, 48 1/2 h, for the two chemicals, NO and O₃, in total 96 values. This mode will show the hourly daily changes. And finally, in the third mode, the interannual or between year (different slices or data matrices) changes (for six years) will be shown. The three-way or three-mode decomposition of this data cube gives three loading matrices corresponding to first mode annual loadings (**X** matrix), second mode daily loadings (**Y** matrix) and third mode interannual loadings (**Z**). For MCR-ALS analysis, a super-augmented data matrix is constructed with NO and O₃ data. First, NO and O₃ data were arranged independently in two column-wise augmented data matrices and then they were appended row-wise (see Fig. 1). This super-augmented (row- and column-wise) data matrix has incomplete data for NO (years 2000 and 2003 absent) and for O₃ (years 2005 and 2006 absent).

3. Method

3.1. MCR-ALS method

MCR-ALS is primarily based on a bilinear model which implies the decomposition of the data arranged in a data matrix expressed as

$$\mathbf{D} = \mathbf{UV}^T + \mathbf{E} \quad 1$$

where **D** is the data matrix with I rows and J columns; **U** is the matrix of scores of dimensions (I,N), where N is the considered number of components in the bilinear matrix decomposition; **V**^T is the matrix of loadings with dimensions N × J and **E** is the matrix of residuals not modelled by the considered N components and with the same dimensions as the data matrix **D**.

Principal Component Analysis [9] performs this bilinear data decomposition of Eq. (1) using the constraints of orthogonality (on **U** and **V**^T), profiles normalisation (on **V**^T) and maximum explained variance for every consecutive component sequentially extracted. Although the mathematical properties of these constraints in PCA provide unique solutions, these solutions have no direct physical meaning and they

are in fact linear combination of the underlying components causing the observed data variance. In MCR-ALS, the goal is precisely to get these underlying components by means of more natural constraints, like non-negativity. Components resolved by MCR-ALS are physically sounder than those obtained by PCA, whose profiles may be difficult to be interpreted directly. Although the application of more natural constraints in MCR-ALS provides better solutions from a physical point of view, they are not unique in general and a set of feasible solutions can be obtained instead [27]. The MCR-ALS method decomposes the data matrix \mathbf{D} by means of an Alternating Least Squares algorithm, which minimizes iteratively the following equation for the two unknowns \mathbf{U} and \mathbf{V}^T

$$\min_{\mathbf{U}, \mathbf{V}^T} \|\hat{\mathbf{D}}_{\text{PCA}} - \hat{\mathbf{U}}\hat{\mathbf{V}}^T\| \quad 2$$

where $\hat{\mathbf{D}}_{\text{PCA}}$ is the PCA reproduced (filtered) data matrix for the considered number of components. $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}^T$ are respectively the current estimations of \mathbf{U} and \mathbf{V}^T factor matrices obtained during the ALS optimisation. In case of not applying constraints, the solutions of Eq. (2) are described by Eqs. (3) and (4) respectively

$$\hat{\mathbf{U}} = \hat{\mathbf{D}}_{\text{PCA}}(\hat{\mathbf{V}}^T)^+ \quad 3$$

and

$$\hat{\mathbf{V}}^T = (\hat{\mathbf{U}})^+ \hat{\mathbf{D}}_{\text{PCA}}, \quad 4$$

where $(\hat{\mathbf{V}}^T)^+$ and $(\hat{\mathbf{U}})^+$ are the current estimations of the pseudoinverses of \mathbf{V}^T and \mathbf{U} [28]. In the MCR-ALS algorithm, these two equations are however solved under a set of constraints over the profiles in the two factor matrices \mathbf{U} and \mathbf{V}^T , such as non-negativity as it has been described elsewhere in different papers [7,27,29,30]. In this case, the problem may be solved rigorously by means of a fast non-negative least squares (fnls) optimisation algorithm [31] expressed as in Eqs. (5) and (6):

$$\hat{\mathbf{U}} = \text{fnls}(\hat{\mathbf{D}}_{\text{PCA}}, \mathbf{V}^T) \quad 5$$

$$\hat{\mathbf{V}}^T = \text{fnls}(\hat{\mathbf{D}}_{\text{PCA}}, \hat{\mathbf{U}}^T). \quad 6$$

Other constraints can also be applied such as unimodality, closure, local rank, correlation, trilinearity (see below), component interaction or physical model constraints [27,30]. These constraints can be applied in a flexible way and independently for every component, separately if necessary. The art and success of the application of the MCR-ALS method is precisely in the proper selection and application of constraints. Recent developments of the MCR-ALS method include the evaluation of the extent of rotation ambiguities in MCR solutions (MCR-BANDS, [12]) and the effects of considering uncertainties in measured data [32–34] using a weighted Alternating Least Squares approach (MCR-WALS). In all MCR methods, the first step is to propose the initial number of components, N , contributing to data variance. This can be done from the variances explained by PCA for different number of components. In MCR methods, individual component variances do overlap and the resolved profiles are not orthogonal as they are in PCA [9,35]. The sum of the explained variances of every individual MCR-ALS component is usually larger than the explained variance by all components when they are considered simultaneously in the MCR model. Natural source profiles in the environment are rarely orthogonal and they do overlap, therefore components resolved by MCR methods can have a physically sounder interpretation (closer to the true one) than PCA resolved ones. Further on, initial estimates of either \mathbf{U} or \mathbf{V}^T can be easily obtained from the set of "purest" columns or rows of matrix \mathbf{D} , [36].

Evaluation of the quality of data modelling in MCR-ALS is performed using the two parameters, percentage of lack of fit (lof%) and explained variances ($R^2\%$)

$$\text{lof}\% = 100 \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m e_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2}}, \quad e_{ij} = x_{ij} - \hat{x}_{ij} \quad 7$$

$$R^2\% = 100 \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - \sum_{i=1}^n \sum_{j=1}^m e_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2} \quad 8$$

Evaluation of the quality in the recovery of the component profiles is possible in case they are known as in the case of simulated data prepared to test the proposed procedure. The parameters used in this case are on one side the correlation coefficient between estimated and known profiles (r^2) and the angle between them calculated as the one whose cosines value is equal to this r^2 value.

$$r^2 = \cos\alpha = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad 9$$

Where \mathbf{x} and \mathbf{y} are the two profiles to be compared

$$\alpha = \arccos(r^2) \quad 10$$

3.2. Application of the MCR-ALS method to an incomplete data multiset

In Fig. 1(c, d) the proposed MCR-ALS procedure is described for the analysis of the incomplete environmental data multiset investigated in this work. Every individual single data matrix corresponds to one year measurements either of ozone (O_3) or of nitrogen oxide (NO), with I rows equal to 365 day observations and J columns equal to 48 half-an-hour observations during one day. As previously stated in the data section, this incomplete data multiset was arranged in a super-augmented data matrix which has some empty matrices. This incomplete data multiset is constituted by three complete data submatrices: \mathbf{D}_1 is the row- and column-wise augmented matrix with the data of years in common (2001 and 2002) for NO and O_3 , \mathbf{D}_2 is the column-wise augmented data matrix with the data of the years where NO concentrations were available (years 2001, 2002, 2005 and 2006), and \mathbf{D}_3 is the column-wise augmented data matrix with the data of the years where O_3 concentrations were available (years 2000, 2001, 2002 and 2003). Note that in this arrangement of the incomplete data multiset shown in Fig. 1c, the years order has been slightly changed to keep data continuity.

The bilinear decomposition of these three data matrices is performed simultaneously using an Alternating Least Squares algorithm which gives three least squares error functions, one for each of them. In the proposed ALS algorithm these three functions are optimised simultaneously (summing the three individual error functions in a total error function) to find out an optimal estimation of scores and loadings profiles describing all the analysed years. In this bilinear decomposition the scores matrix \mathbf{U}_1 has the same rows as \mathbf{D}_1 (the NO- O_3 row- and column-wise superaugmented data matrix), \mathbf{U}_2 has the same rows as \mathbf{D}_2 (the NO column-wise augmented matrix) and \mathbf{U}_3 has the same rows as \mathbf{D}_3 (the O_3 column-wise augmented matrix). In the augmented data matrices corresponding to O_3 and NO, the resolved scores will include for O_3 years 2003 and 2000, and for NO years 2005 and 2006. In the bilinear decomposition of the three data matrices, the resolved loadings matrix \mathbf{V}_1 has the same columns as \mathbf{D}_1 describing simultaneously NO and O_3 loadings, whereas \mathbf{V}_2 has the same columns as \mathbf{D}_2 describing NO loadings and \mathbf{V}_3 has the same columns as \mathbf{D}_3 , describing O_3 loadings. Note that in theory, the non-measured empty matrices could be imputed from the finally obtained results.

According to Fig. 1d, in the implementation of the ALS algorithm for incomplete data multisets, some aspects are different to those common to the general ALS algorithm for the analysis of complete data multisets. In the calculation of the score factor matrices from the incomplete data set, three different situations are encountered: a) evaluation of the scores \mathbf{U}_1 from data matrix \mathbf{D}_1 which accounts for the data corresponding only to the common NO and O_3 data (circles in arrow blue continuous line for common years 2001 and 2002 in Fig. 1c), (b) evaluation of the scores \mathbf{U}_2 from data matrix \mathbf{D}_2 which accounts only for the data corresponding to NO data (circles in green dashed line for years 2001, 2002, 2005 and 2006 in Fig. 1c) and c) evaluation of the scores \mathbf{U}_3 from data matrix \mathbf{D}_3 which accounts only for the data corresponding to O_3 data (circles in red dashed line for years 2000, 2001, 2002, and 2003 in Fig. 1c). In every case, the calculation involves the multiplication of the appropriate augmented data matrix, i.e. \mathbf{D}_1 , \mathbf{D}_2 or \mathbf{D}_3 , by the pseudoinverse of the corresponding loading matrix, i.e. \mathbf{V}_1 , \mathbf{V}_2 and \mathbf{V}_3 . Observe in Fig. 1d that the appropriate rows and columns of the incomplete data matrix were selected to form the three new rectangular invertible data submatrices and observe also that columns of the corresponding \mathbf{V}^T loadings matrices have the same number of columns as these rectangular submatrices. In summary, at each iteration of the ALS global optimisation, the following subset of ALS equations would be solved:

For data matrix \mathbf{D}_1

Calculation of factor matrix \mathbf{U}_1 from the current estimation of $\hat{\mathbf{V}}_1^T$ by:

$$\min_{\mathbf{U}_1, \text{constraints}} \|\hat{\mathbf{D}}_{1,PCA} - \hat{\mathbf{U}}_1 \hat{\mathbf{V}}_1^T\|_{\hat{\mathbf{U}}_1} = \text{fnmls}(\hat{\mathbf{D}}_{1,PCA}, \hat{\mathbf{V}}_1^T) \quad 11$$

Calculation of factor matrix \mathbf{V}_1^T from the current estimation of $\hat{\mathbf{U}}_1$ by

$$\min_{\mathbf{V}_1^T, \text{constraints}} \|\hat{\mathbf{D}}_{1,PCA} - \hat{\mathbf{U}}_1 \hat{\mathbf{V}}_1^T\|_{\hat{\mathbf{V}}_1^T} = \text{fnmls}(\hat{\mathbf{D}}_{1,PCA}, \hat{\mathbf{U}}_1^T) \quad 12$$

Calculation of the least squares error function corresponding to data matrix \mathbf{D}_1 to be minimised which is defined from the residuals calculated using the current estimations of $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{V}}_1^T$

$$\text{ssq}_1 = \sum_{i=1}^{I_1} \sum_{j=1}^{J_1} \left(d_{1,ij}^2 - \sum_{n=1}^N \hat{u}_{1,i,n} \hat{v}_{1,j,n} \right) \quad 13$$

Analogous Equations can be written for the least squares error functions corresponding to the other matrices \mathbf{D}_2 and \mathbf{D}_3 given respectively $\hat{\mathbf{U}}_2, \hat{\mathbf{V}}_2^T$, ssq_2 , $\hat{\mathbf{U}}_3, \hat{\mathbf{V}}_3^T$ and ssq_3 estimations. The final $\hat{\mathbf{U}}$ matrix is then obtained by column-wise juxtaposition of the three scores matrices

$$\hat{\mathbf{U}} = \{ \hat{\mathbf{U}}_1; \hat{\mathbf{U}}_2; \hat{\mathbf{U}}_3 \} \quad 14$$

where, when there are coincidence in the rows, matrix $\hat{\mathbf{U}}_1$ always prevails over $\hat{\mathbf{U}}_2$ or $\hat{\mathbf{U}}_3$ because \mathbf{D}_1 is the matrix that should contain more reliable information as it was obtained with the complete joint column vector space (see Fig. 1c). On the other hand the corresponding final loading matrix

$$\hat{\mathbf{V}}^T = \frac{1}{2} (\hat{\mathbf{V}}_1^T + [\hat{\mathbf{V}}_2^T, \hat{\mathbf{V}}_3^T]) \quad 15$$

where $[\hat{\mathbf{V}}_2^T, \hat{\mathbf{V}}_3^T]$ is the row-wise augmented factor matrix formed by concatenation of $\hat{\mathbf{V}}_2^T$ and $\hat{\mathbf{V}}_3^T$, taking into account maximum information (not only from \mathbf{D}_1 , but also the one from the other two matrices \mathbf{D}_2 and \mathbf{D}_3).

The function which is finally optimised by ALS is

$$\text{ssq} = \text{ssq}_1 + \text{ssq}_2 + \text{ssq}_3 \quad 16$$

From which the sigma value, σ

$$\sigma = \sqrt{\frac{\text{ssq}}{(I_1 J_1 + I_2 J_2 + I_3 J_3)}} \quad 17$$

where $I_1, J_1, I_2, J_2, I_3,$ and J_3 are respectively the number of rows and columns of the three augmented data matrices $\mathbf{D}_1, \mathbf{D}_2$ and \mathbf{D}_3 . In Eq. (17), sigma or σ , gives the root mean square of the sum of residual squares, ssq . The optimisation convergence criterion is when improvement on relative changes on the sigma value between iterations is below a threshold value, usually 0.1%. This value may be modified depending on the stage of the optimisation. Usually at the beginning of a particular study a higher value is used (i.e. 1%) for exploratory purposes. In contrast, once a good model has been found, lower values are attempted to see whether there is any appreciable improvement in the obtained solutions.

In Fig. 1d, the way the profiles in three modes are recovered from MCR-ALS resolved scores and loadings is shown. When the trilinearity constraint is applied, MCR-ALS score profiles are iteratively forced to conform the trilinear model. This implies that the score profiles of every component corresponding to each individual data matrix are forced to be invariant (same shape) and only differ in a scaling factor. This is precisely what the scheme in Fig. 1d describes for one of the components ($\mathbf{u}_{\text{aug},n}, n = 1 \dots N$). Once this single component score profile is appropriately refolded in a data matrix with the same number of columns as matrices (data slices) simultaneously analysed, its first singular component decomposition gives the invariant or common part of them in \mathbf{x}_n and the scaling factor corresponding to each individual matrix in \mathbf{z}_n . If these two vectors are multiplied using the Kronecker tensor product [22], the corresponding augmented score vector $\mathbf{u}_{\text{aug},n}$ is updated in \mathbf{U}_{aug} . This operation can be performed for every component $n = 1, \dots, N$, and the full \mathbf{U}_{aug} scores matrix will have all components conforming with the trilinear model. Once the ALS procedure converges, the profiles of the components in the three modes can be easily recovered as also indicated in Fig. 1b and d. Note that the loadings \mathbf{V}^T obtained by MCR-ALS give directly the \mathbf{Y}^T profiles in the second mode. In three-way (or three-mode) data analysis, the profiles in the three modes are usually simply called 'loadings' in first, second and third modes, and the 'scores' denomination is preferably used in the case of two-way data analysis for the profiles in the rows (samples) first mode. In any case, either if the trilinearity constraint previously described is applied or not during the ALS bilinear decomposition, the profiles in the three modes can always be recovered as indicated in Fig. 1b and they can be easily compared with those obtained using other multiway methods. The only restriction needed to get these profiles in the three modes is that the sizes of the individual data matrices simultaneously analysed have the same number of rows and columns, or what is the same that the different data modes are consistent and conform with the higher mode data structure (a data cube in the case of three-way or three-mode data). In the case of applying MCR-ALS without the trilinearity constraint (only bilinear modelling), the recovered profiles in three modes can still approximated as shown in Fig. 1b and d, where the relation between profiles in \mathbf{U}, \mathbf{V}^T and \mathbf{X}, \mathbf{Y} and \mathbf{Z} is given, as well as how they can be obtained one from the other.

In previous works [17,18], we have shown that when the trilinear constraint is implemented during the ALS optimisation, the corresponding loadings in the first and third mode are equivalent to those obtained by PARAFAC or other methods [21,22]. We have also described variants of this method to handle possible interaction of component profiles in some modes, giving results equivalent to the application of restricted Tucker model based methods [22]. One of the advantages of the implementation of this method is that it may be applied in a flexible and optional way for only some of the components of the bilinear decomposition, allowing for the postulation of mixed multilinear models (e.g. partially bilinear and trilinear). See

Table 1

MCR-ALS for experimental and complete and incomplete simulated data multisets with different noise levels (see Section 2.2). Lof (%) lack of fit defined in Eq. (7), R^2 (%) explained variances defined in Eq. (8), r^2 correlation coefficients between profiles (Eq. (9)) and angle between profiles (Eq. (10)).

						Sim. Load.1 st mode (X)		Sim. Load 2 nd mode (Y ^T)	
				Lof (%)	R2 (%)	r2	angle	r2	angle
Experimental data	Incomplete matrix	(non-neg)	Comp.1	57.99	66.38	-	-	-	-
			Comp.2	77.89	39.33	-	-	-	-
			Comp.3	88.06	22.46	-	-	-	-
			total	7.63	99.42	-	-	-	-
		(non-neg + tril 110)	Comp.1	51.27	73.72	-	-	-	-
			Comp.2	81.11	34.22	-	-	-	-
			Comp.3	95.38	9.04	-	-	-	-
			total	19.69	96.12	-	-	-	-
Simulated data	Complete matrix	error 0.1% (non-neg)	Comp.1	64.59	58.28	0.9991	2.457	0.9944	6.072
			Comp.2	75.99	42.24	0.9963	4.922	0.9869	9.276
			Comp.3	87.91	22.72	0.8886	27.299	0.9927	6.922
			total	0.08	99.99	-	-	-	-
		(non-neg + tril)	Comp.1	62.69	60.69	0.9998	1.041	0.9999	0.895
			Comp.2	78.4	38.54	1	0.005	1	0.268
			Comp.3	87.95	22.64	0.9991	2.459	0.9983	3.347
			total	0.11	99.99	-	-	-	-
		error 20% (non-neg)	Comp.1	61.66	61.98	0.9999	0.8	0.9985	3.156
			Comp.2	81.27	33.95	0.9954	5.5	0.9997	1.395
			Comp.3	86.85	24.57	0.999	2.524	0.9894	8.361
			total	15.97	97.45	-	-	-	-
	(non-neg + tril)	Comp.1	62.89	60.45	0.9999	0.626	0.9999	0.675	
		Comp.2	79.49	36.81	0.9998	1.22	0.9999	0.9036	
		Comp.3	89.05	20.7	0.9992	2.319	0.9991	2.471	
		total	16.19	97.38	-	-	-	-	
	Incomplete matrix	error 0.1% (non-neg)	Comp.1	61.99	61.57	0.9989	2.697	0.9895	8.326
			Comp.2	79.47	36.84	0.9912	7.611	0.9869	9.277
			Comp.3	85.34	27.17	0.8508	31.707	0.9761	12.549
			total	0.15	99.99	-	-	-	-
		(non-neg + tril 110)	Comp.1	60.15	63.82	0.9901	8.077	0.9859	9.6473
			Comp.2	80.69	34.89	0.9689	14.317	0.9734	13.251
			Comp.3	86.43	25.3	0.8745	29.019	0.8481	31.993
			total	5.09	99.74	-	-	-	-
error 5% (non-neg)		Comp.1	63.11	60.17	0.9985	3.163	0.9866	9.407	
		Comp.2	79.48	36.83	0.9871	9.218	0.9871	9.211	
		Comp.3	83.88	29.64	0.8353	33.353	0.9676	14.623	
		total	4.13	99.83	-	-	-	-	
(non-neg + tril 110)	Comp.1	57.05	67.46	0.9989	2.724	0.9967	4.6566		
	Comp.2	78.3	38.69	0.9907	7.828	0.9815	11.027		
	Comp.3	94.33	11.02	0.9196	23.13	0.9957	5.298		
	total	7.78	99.39	-	-	-	-		
error 10% (non-neg)	Comp.1	55.88	68.78	0.9899	8.1607	0.9969	4.497		
	Comp.2	88.08	22.43	0.965	15.206	0.9957	5.298		
	Comp.3	82.43	32.05	0.9848	10.004	0.8775	28.66		
	total	7.95	99.37	-	-	-	-		
(non-neg + tril)	Comp.1	54.53	70.26	0.9975	4.087	0.9896	8.273		
	Comp.2	84.34	28.86	0.9598	16.311	0.9967	4.634		
	Comp.3	90.45	18.19	0.9817	10.992	0.9592	16.414		
	total	11.68	98.64	-	-	-	-		
error 20% (non-neg)	Comp.1	56.05	68.59	0.9835	10.428	0.9956	5.4		
	Comp.2	88.31	22.02	0.9496	18.269	0.9834	10.457		
	Comp.3	83.36	30.51	0.9808	11.234	0.8962	26.342		
	total	15.75	97.52	-	-	-	-		
(non-neg+ tril)	Comp.1	55.57	69.12	0.9893	8.409	0.9949	5.79		
	Comp.2	82.34	32.2	0.9888	8.596	0.9894	8.335		
	Comp.3	95.4	8.98	0.9363	20.555	0.9847	10.039		
	total	20.12	95.95	-	-	-	-		

references [17,18] for more details about this implementation and its advantages.

All data analyses were performed under MATLAB (The Mathworks, Natick, USA). MCR-ALS toolbox is freely available at www.mcrals.info and the new MCR-ALS command line MATLAB program to work with incomplete data is a recent home made version only available at present under request to one of the authors (RT). To identify incomplete data years in MATLAB environment for both contaminants, their values were substituted for NaN (Not a Number).

4. Results and discussion

In this work, the MCR-ALS method for incomplete data has been developed and tested for two types of incomplete data multisets, one experimental and another one simulated. A summary of the obtained results is given in Table 1. In the first lines of the Table, results are given for the incomplete experimental data multiset and in the next lines, results are given for the different simulated data multisets. As previously explained, the simulated data multiset was built with the profiles obtained from the MCR analysis of the experimental one. For the simulated data multisets, results are given for the corresponding complete and incomplete data multisets. In the case of the simulated incomplete data multiset, results are shown for four different levels of added random noise (0.1%, 5%, 10% and 20%), whereas for the corresponding complete data set, results are shown only for two noise levels (0.1% and 20%). Comparison of these results obtained for the two simulated data multisets allowed for a better discussion of the effects of data completeness, noise and rotation ambiguities. In all cases, simulated and experimental, results are given considering the fulfilment of both the bilinear and trilinear models under non-negativity constraints. Frobenius normalisation of second mode loadings (\mathbf{V}^T or \mathbf{Y}) was applied in all cases. In the columns of Table 1, results are given for data lack of fit (lof%, Eq. (7)) and explained variances (R^2 %, Eq. (8)). These two values are given individually for every component separately and for the full model considering all them together (see below discussion). Moreover, in the last four columns of Table 1, recovery of first and second mode loadings profiles are evaluated using correlation coefficients (r^2) between the MCR recovered profiles and the correct ones (used for the data simulation) and using angles between them (calculated from the arc whose cosines is equal to the r^2 values). These values are especially useful to evaluate also the presence of rotation ambiguities. Values close to one for r^2 and close to zero for the angles confirm that the recovery of the first and second mode loadings profiles are good with low rotation ambiguity, whereas r^2 far from one and large angles meant bad recoveries (high rotation ambiguity) of the profiles. In the Table, recovery values are only given for the first and second mode loadings, \mathbf{X} and \mathbf{Y}^T respectively in Fig. 1b (and not shown for intermixed profiles \mathbf{U}_{aug} and for the third mode loadings \mathbf{Z}). These recovery values could not be calculated for the experimental data multiset, because the correct true profiles were not known.

In Fig. 2, the different behaviour obtained during the ALS optimisation of the error function (defined as in Eqs. (16) and (17)) for the different cases under study in this work is shown graphically. In all cases the \log_{10} of the error function is plotted for scale convenience. In the bilinear model case (for each data set, at its left part of the Fig. 2), the behaviour of the optimisation is the expected one with a fast decrease at the beginning and a slowly decrease afterwards until convergence. This behaviour is the same for complete and for incomplete data multisets. Convergence was faster for experimental data than for simulated data with random noise due to easier achievement of the convergence threshold condition in case of real noise cases. When noise is very low and homocedastic (from simulated data), ALS convergence close to the minimum is slow because at each iteration an improvement will be still possible until reach practically the very small noise level. This does not happen with real data, where sources of noise are higher and not

often homocedastic. In the case of the complete data multisets, the behaviour of the optimisation function were also the expected one when the trilinear constraint was applied (left lower part of the Figure), either at 0.1% or at 20% noise levels. In contrast in the case of the analysis of the incomplete data multisets, the implementation of the trilinearity constraint produced unstable results during the optimisation of the residuals. Although three-mode profiles could still be recovered, the behaviour of the log of the error function was unstable with sudden decreasing and increasing alternating intervals. A first conclusion is therefore that the implementation of the trilinear constraint in the case of incomplete data sets resulted more difficult (although still possible) than for the bilinear case.

4.1. Experimental data

Results of application of MCR-ALS to the incomplete experimental data multiset for three components using the bilinear model and non-negativity and loadings normalisation constraints are shown first (Table 1 and Fig. 3). Three components were chosen because of the results of our previous study, where the experimental data were examined separately for the three data arrays \mathbf{D}_1 , \mathbf{D}_2 and \mathbf{D}_3 shown in Fig. 1c [18]. This number of components finally selected corresponded to those giving better results in the explanation of the complete data matrix (\mathbf{D}_1 , in Fig. 1). Trilinearity constraint was considered for the first and second components, and not for the third component (trilinear model noted as 110 in Table 1). If this is also applied to the third component, rank deficiency problems were encountered. In Fig. 2 upper left, plots of the log of the error function (Eq. (14)) optimised during ALS iterations for the experimental data multiset are shown. For a convergence criterion of 0.1% of relative change of the standard deviation of the residuals, convergence was achieved at the iteration 129 when the bilinear model was applied. In contrast when for the same system the trilinear model (110) was forced, the behaviour of the log of the error function was anomalous and not decreasing smoothly with its lowest value already attained at the third iteration, stabilising afterwards at a higher value. This reflects the difficulty of the trilinear model implementation for the incomplete data multiset, in contrast to the bilinear model, which showed the expected behaviour. In terms of total explained variances, they were 99.42% for the bilinear model and 96.12% for the trilinear model respectively (Table 1), and in terms of lack of fit values, imposing trilinearity to first and second components makes worse the lack fit from 7.63% to 19.69%.

In Fig. 3, the profiles finally obtained under both types of models, bilinear (green colour) and trilinear (red colour), under non-negativity and normalisation constraints are shown. As explained in the Method section, the augmented scores \mathbf{U}_{aug} can be rearranged as three-way loadings in first and third modes, \mathbf{X} and \mathbf{Z} , respectively, and the two-way loadings \mathbf{V}^T are the three-way loadings in second mode, \mathbf{Y} (see Fig. 1). As it can be seen in Fig. 3, there were little differences in the shape of resolved profiles either when the bilinear or the trilinear models were applied. This means that in spite of the differences in data fitting and convergence for the two approaches the results in terms of the shapes of the profiles were similar and rather consistent. The profiles obtained for the first (\mathbf{X}) and second modes (\mathbf{Y}) of the first and second components explaining a large amount of data variance (Table 1) were very close to those described and published in our previous work for the results of the complete \mathbf{D}_1 data matrix [18].

Since the goal of this work was testing the proposed extension of the MCR-ALS method to incomplete data multisets and not the environmental interpretation of the results obtained in their study (this was already done in previous paper [18]), only a brief explanation of them is given here. The resolved profiles are similar to those showed in our previous papers [18,23] and other studies [37]. Experimental data values already reflected that the highest concentrations for ozone were found in spring and summer, coinciding with the highest solar radiation and that the opposite behaviour was obtained for NO

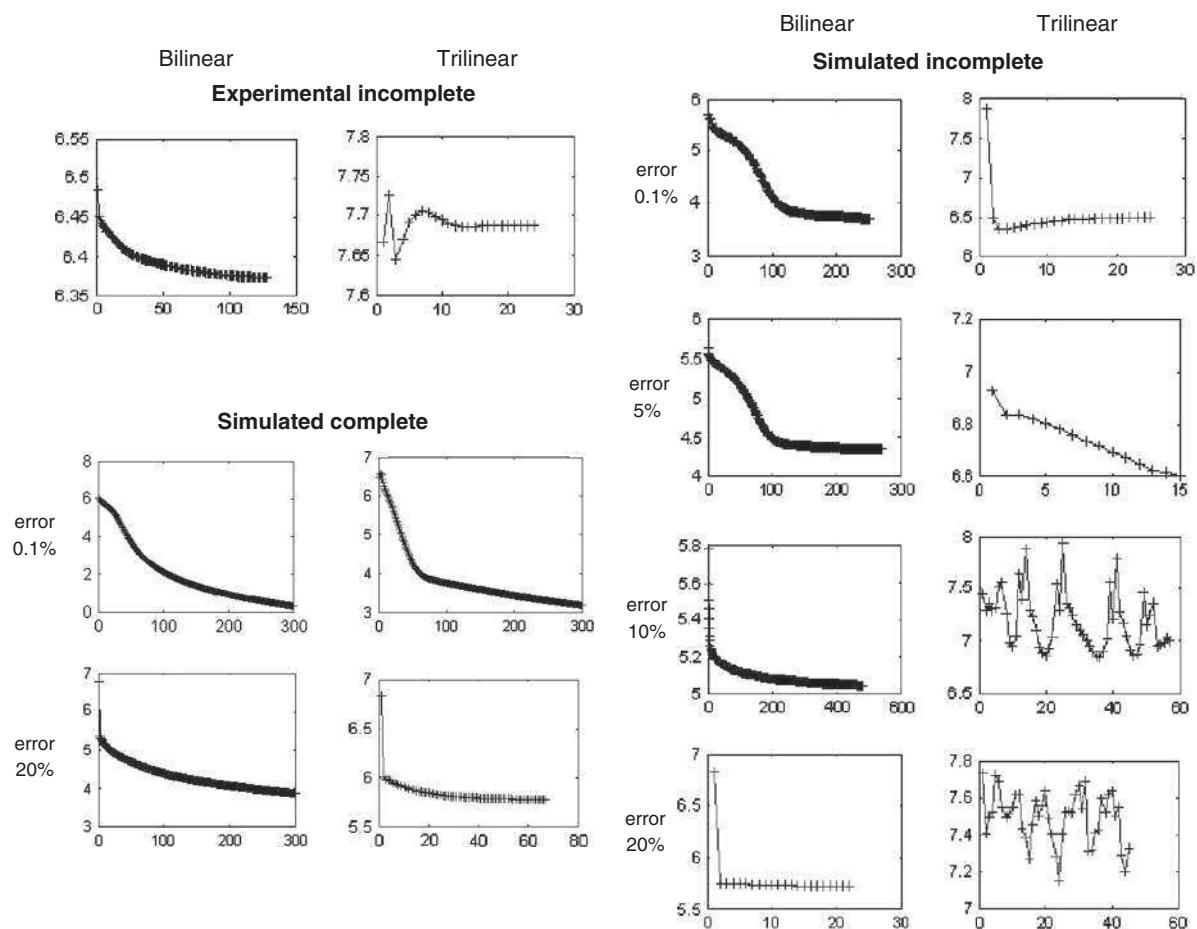


Fig. 2. Optimisation pathway of MCR-ALS log of error function for the different data multisets studied in this work using different noise levels and models (bilinear or trilinear). Y axis gives $\log(u)$, $u = u_1 + u_2 + u_3$ (Eq. (14)). X axis gives number of MCR-ALS iterations.

concentrations, because of its reaction and interaction with ozone. In the daily variations (hourly trend), the maximum of ozone was at 16.00 h (14.00 h solar time) coinciding also with the highest levels of solar radiation during the day. For NO concentration levels, their maximum concentration values were found during rush traffic hours, around 8.00 h in the morning and 19.00 in the evening. There is not clear trend for the possible variation of these tendencies over different years. In spring and summer the higher levels of solar radiation and temperatures promote photochemical reactions and the decreasing of NO concentration levels because of their reaction with maximum concentration levels of O_3 . Another factor to consider explaining the observed trends is the increase of NO emissions during winter due to the use of heating and heavier traffic in cities compared to the summer and Easter vacations periods.

In the first MCR resolved component given in the upper part of Fig. 3, the contribution of O_3 predominates over the contribution of NO, as it is clear from second mode loading profiles, (V^T or Y) which show the hourly variation during the day separately for the NO (left) and O_3 (right). The right part of this loading profile has a maximum around 16:00 h (14:00 solar hour) which is coincident with the maximum high solar radiation at this hour, and a minimum at the traffic rush hour (8:00 h), precisely when NO (left part) has its maximum. This profile therefore described the interaction between NO and O_3 , as it is confirmed below when is compared with the profiles of component 2. The first component of first mode loadings (X) reflect

mostly the seasonal trends in solar radiation, in agreement with the observed higher concentration levels of O_3 , and lower concentration levels of NO in spring and summer. The inter-annual trends given in third mode profile (Z) do not show clear changes over years.

In the second MCR resolved component given in the middle part of Fig. 3, NO contribution predominates (left part of Y) over O_3 contribution (right part of Y). On the other side two clear maxima are observed in second component X profile coincident with NO highest concentration levels registered during late autumn and winter, when ozone concentrations were much lower due to lower solar radiation. In the third mode profile, a decrease of the contribution of NO is observed for the last two years of study.

Finally, in the third MCR resolved component given in Fig. 3, the profile shows a small minimum, less pronounced in comparison with that found in the other component profiles. Second mode profiles of second and third components (Y), show a peak maximum in the NO part around 21:00 h and the second component also around 9.00 hours (morning) which correspond to rush traffic hours assigned to fuel combustion by vehicles. The small presence of NO in the first component and of O_3 in the second and third components shows again the interaction (inverse correlation) between NO and O_3 concentration profiles. Sunlight triggers a succession of photochemical reactions, which include the reaction of NO with O_3 to give NO_2 producing the depletion of O_3 . When O_3 reaches its maximum concentration, NO

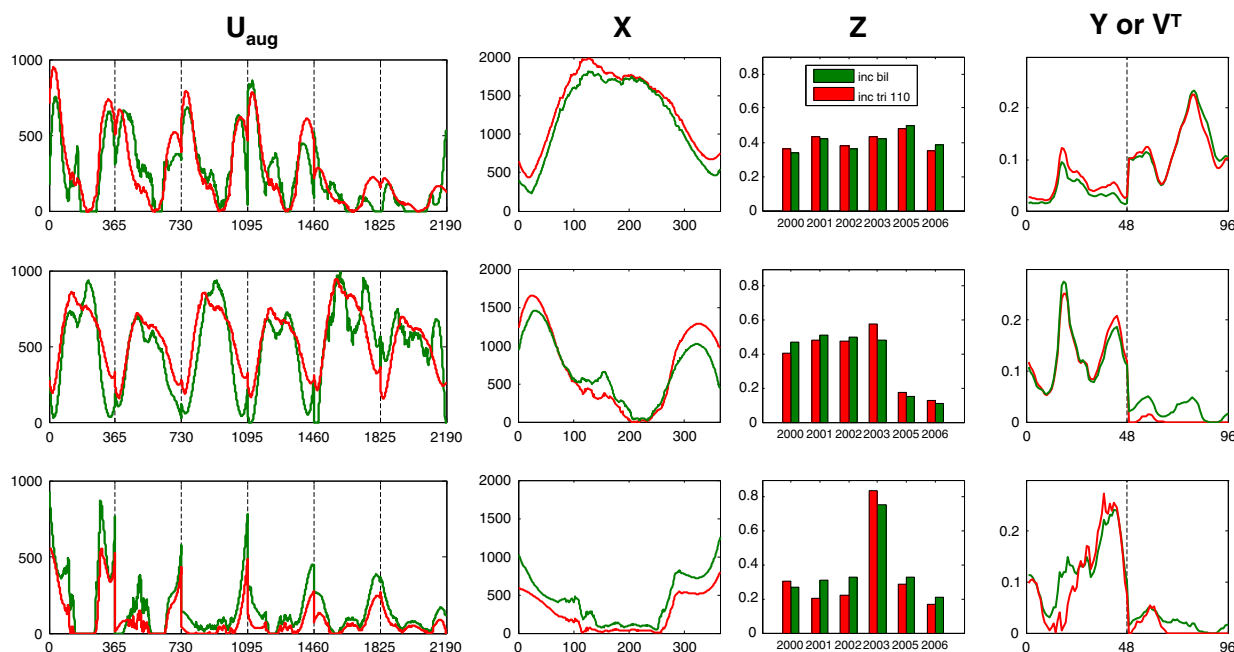


Fig. 3. MCR-ALS resolved profiles for incomplete experimental data applying non-negativity constraints (green line), and non-negativity and trilinearity (in first and second components) constraints (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

decreases, achieving its daily minimum value. After that, NO begins to accumulate again, reaching a new maximum during the evening. Finally, loadings in the third mode describing the inter-annual profiles showed a constant trend during the different years of the study, with an outlying high value in 2003.

4.2. Simulated data

4.2.1. Simulated complete data multisets

To compare the peculiarities of working with incomplete data multisets compared to complete data multisets, MCR-ALS was first used to analyse the complete simulated data multiset, including NO and O₃ matrices for all the same years, and using different noise levels (see Table 1). Only two noise levels are given in this case, 0.1% (nearly no noise) and 20% (high noise level). As explained in the data section, the complete data multiset was generated under the premises of the trilinear model. Lack of fit and R² figures given in Table 1 show the expected tendency according to noise levels. In this case, the implementation of trilinearity implied only a small increase of lack of fit values compared to those obtained when only the bilinear model was applied (from 0.08% to 0.11% and from 15.97% to 16.19%). The same can be said for R² values, proving the good behaviour of MCR-ALS under these two model cases, either bilinear or trilinear (trilinearity constraint). In Fig. 2, the behaviour of the log of the error function, either using the bilinear or the trilinear model, and for the two noise levels is as expected with a slow decreasing tendency which stops depending on the convergence criterion (0.01% in the Figure). Changes in the resolved profiles were negligible (non-visible) already when the convergence criterion for the relative standard deviation of the residuals is lower than 1%.

As shown in Fig. 4 for the case of highest noise level of 20%, profiles obtained either assuming only the bilinear non-negativity or the trilinear non-negativity models were again practically the same and coincident with theoretical ones used for the data simulation. This is also confirmed in Table 1, where all r² are close to 1 and all the angles are very low. This means that MCR-ALS bilinear analysis (with non-negativity constraints)

of the simulated complete data multisets already eliminated most of rotation ambiguities associated to this model and that noise effects on the shapes of the profiles were very low in this case.

4.2.2. Simulated incomplete data multisets

To further investigate the effects of data completeness, noise level and model structure (either bilinear or trilinear), the proposed MCR-ALS method was applied to the simulated incomplete data multisets previously described at four noise levels, 0.1%, 5%, 10% and 20%. A summary of these results are shown in the lower half part of Table 1. MCR-ALS was applied considering three components and using non-negativity with and without (optional) the trilinearity constraint, apart from loadings normalisation. At noise levels of 0.1% and 5%, trilinearity constraint was considered for the first and second components, and not for the third component (trilinear model noted as 110). Like for the experimental data case, if this were also applied to the third component, rank deficiency problems were encountered due to the change of rank structure when some data blocks were removed from the complete data set. At noise levels 10% and 20% the trilinearity constraint could be applied to all three components.

In Fig. 2, the behaviour of the log of the error function, either using the bilinear or the trilinear model, and for the different noise levels is shown (right side of the Figure). A different behaviour is encountered depending on the added noise level and on the application of the bilinear or trilinear models. In the case of the application of the bilinear model the behaviour of log of the error function along the optimisation was the expected one, with a faster decrease during the first part of the optimisation and a slow convergence in the second part of the optimisation which stops depending on the noise level and convergence criterion. Interestingly, convergence (at the 0.01% relative change of the standard deviation of the residuals between consecutive iterations) was faster at higher noise levels. This confirms the observation of slow convergence of ALS algorithms for low noise data (homocedastic in case of simulated data) compared to higher noise for real data. This slower convergence may be explained by the impossibility to accomplish any improvement of the optimised parameters when the error function already reaches

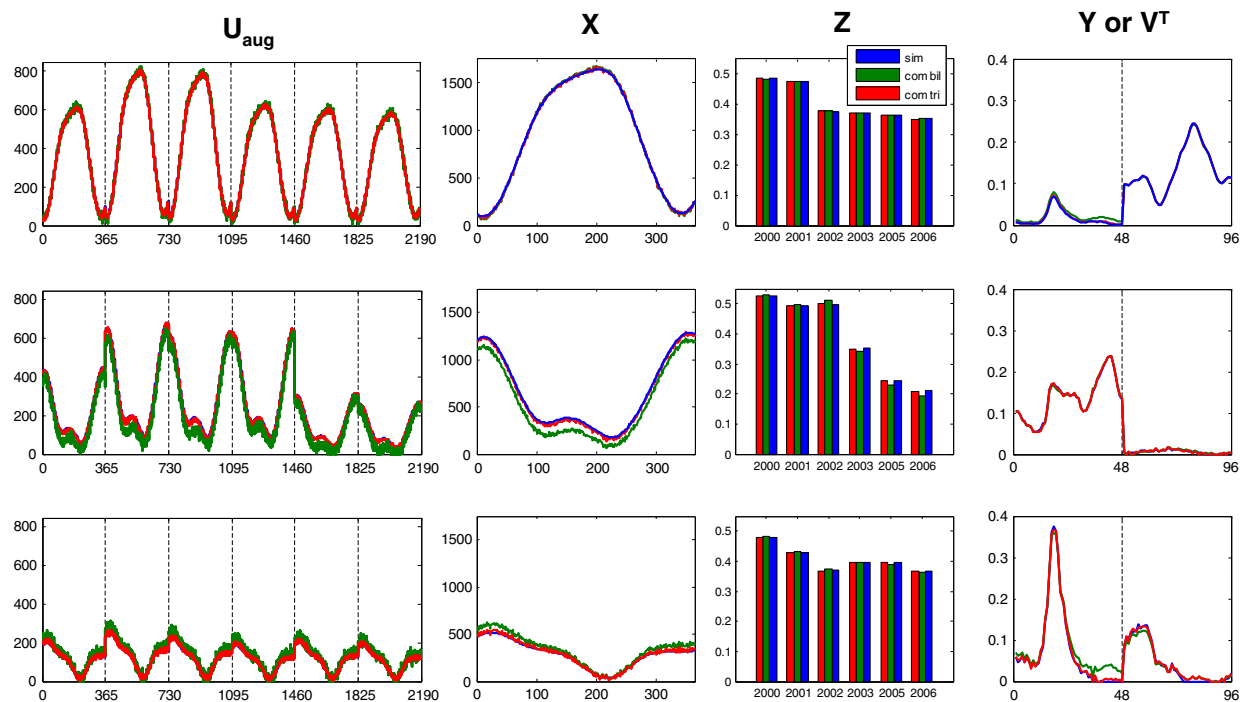


Fig. 4. MCR-ALS resolved profiles for complete simulated data with 20% error applying non-negativity constraints (green line), and non-negativity and trilinearity (in all components) constraints (red line). Blue lines are the profiles used in the simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

the noise level. When the latter is very low (or zero for pure data), the optimisation can still produce a very small change in the parameters which makes decrease the error function which in absence of noise can reach the machine accuracy of the computer used [38]. On the contrary, the log of the error function in the case of applying the trilinear model is very unstable, especially for the higher noise levels investigated, 10% and 20%, with oscillating stages of the log of the error function with fast decreases and fast increases. Usually the minimum value in all cases is obtained at the earliest iterations, from which the function increases (divergence) or oscillates. We interpret these results considering the higher difficulty of the implementation of the trilinear model for the incomplete data multiset. Even though the procedure described in this work allows for the estimation of the profiles in the three modes using the trilinearity algorithm implemented in MCR-ALS (note that this is completely different to the way the trilinear model is solved in methods like PARAFAC [21]), the results are unstable, especially for the cases where noise level was high. When these results are compared with those obtained previously for the bilinear model for the incomplete data multiset and for the trilinear model for the complete data multiset, it is concluded that the observed instability is due to the difficulty in the implementation of the trilinear constraint for incomplete data multisets. Some work is pursued in this direction to further investigate the problem.

In Table 1, lack of fit and R^2 values are given for the different simulated incomplete data multisets at the different noise levels and models (bilinear or trilinear). In the case of the bilinear model, the results follow the expected behaviour, with lack of fit values increasing at the level of noise level. However in the case of the application of the trilinear model, and similarly to what was happening in the analysis of the experimental incomplete data multiset, the fit for the trilinear model resulted significantly worse, even though the data were simulated under the premises of the fulfilment of the trilinear model. This situation agrees also with the anomalous behaviour of the minimisation of the error function in the case of the trilinear model.

In Figs. 5 and 6, MCR-ALS resolved profiles for the simulated incomplete data multisets at the added noise levels of 5% and 20% are shown. MCR-ALS resolved profiles obtained at other noise levels (0.1% and 10%) are not shown graphically for brevity. Blue lines are the profiles used in the simulation (the correct ones), the green lines are the MCR-ALS resolved profiles assuming only the bilinear model and non-negativity constraints and the red lines are the resolved profiles assuming the trilinear model and non-negativity constraints.

In both cases of high noise levels 5% (Fig. 5) and 20% (Fig. 6), the recovery of profiles can be considered rather good, especially in the complete data subregions (corresponding to \mathbf{U}_1 scores). In the case of application of the bilinear model, the effect of noise propagation is however clearly seen in the regions of \mathbf{U}_{aug} corresponding to \mathbf{U}_2 and \mathbf{U}_3 score matrices (green lines) and also in a lower extent in the corresponding averaged \mathbf{X} first mode loadings. For the first component, which mainly was reflecting ozone concentration changes (see \mathbf{V}^T and \mathbf{Y} profiles and discussion of experimental data multiset in Section 4.1), a noisy pattern is appearing in the last part of the two-way augmented scores profile, related to the year subsets where the information about ozone concentration was absent. For the second component, which now is mostly related to NO, the noisy pattern is in the first part of the recovered two-way augmented scores profile related to the year subsets where NO concentrations were missing. For the third component, the noisy pattern is seen for both parts of the profile, at the beginning and end, which reflects that in the recovery of this profile, both ozone and NO subsets were important. These effects are not noticed so much in the profiles in second mode (\mathbf{V}^T and \mathbf{Y}), which always are more stable to noise perturbations. This agrees with the fact that their estimation always involves a much more overdetermined system of equations. The augmented data matrices have always more rows (365 days \times number of years) than columns (96 half hours). This noise propagation pattern is less important for the case of the trilinear constraint implementation (red line). When this constraint was applied, although unstable algorithmically, MCR-ALS recovered loading profiles (red lines) were less

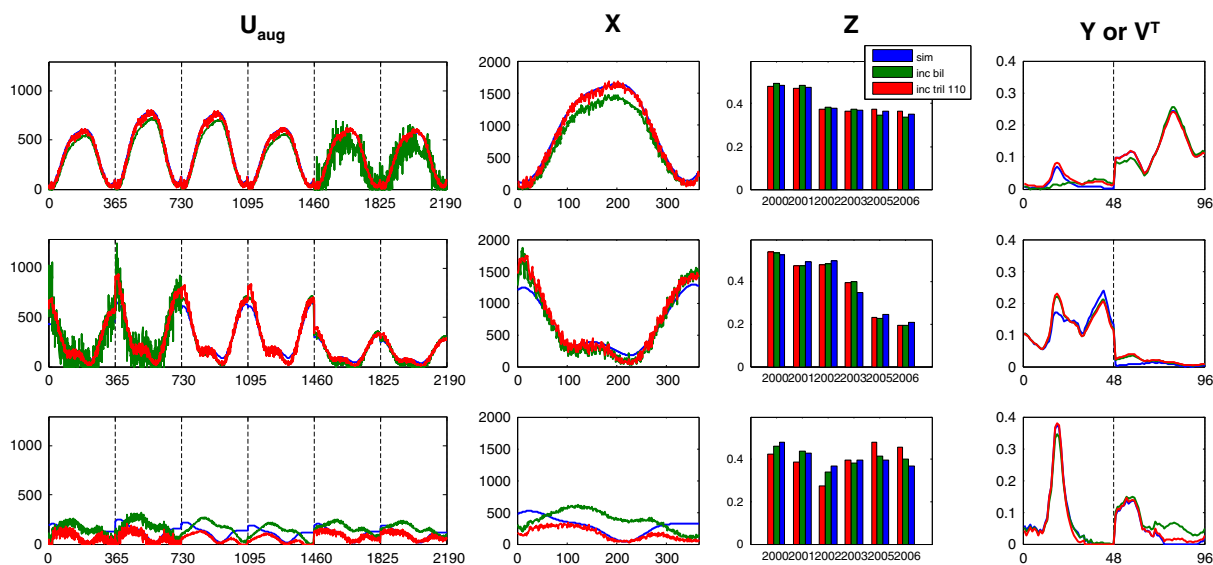


Fig. 5. MCR-ALS resolved profiles for incomplete simulated data with 5% error applying non-negativity constraints (green line), and non-negativity and trilinearity (in first and second components) constraints (red line). Blue lines are the profiles used in the simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

perturbed by noise propagation, which agrees with the higher robustness to noise propagation of trilinear models compared to bilinear models. A contradictory situation is then encountered, on one hand the implementation of the trilinearity constraint is more unstable than employing only the assumption of bilinearity, but on the other hand, overfitting effects at higher noise levels (20%) are more pronounced for the latter (bilinear model) than for the former (trilinear model). In any case and as

previously said also, the recovery of the profiles for the third component in the different modes was always the more difficult one. This is due to the larger rotation ambiguity associated to this component. Although profile recoveries measured by r^2 and angle values in Table 1 become a little worse when noise level increases and trilinearity constraint was applied, no clear conclusions can be derived in this case due to the intricate relation between noise propagation and rotation ambiguity.

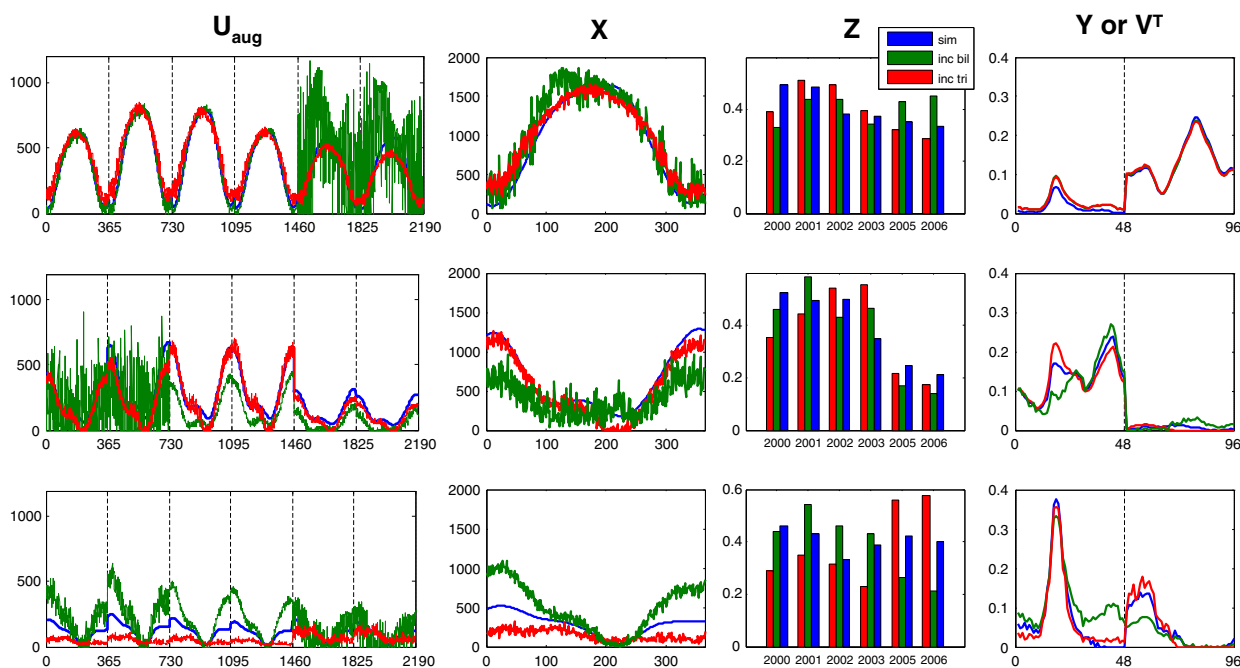


Fig. 6. MCR-ALS resolved profiles for incomplete simulated data with error 20% applying non-negativity constraints (green line), and non-negativity and trilinearity (in all components) constraints (red line). Blue lines are the profiles used in the simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

5. Conclusions

A new MCR-ALS method is presented to cope with incomplete data multisets. The method is successfully tested in the analysis of different simulated incomplete data multisets at different noise levels (up to 20%), using either bilinear or trilinear models, with comparable results to those obtained for complete data multisets. With higher noise levels, caution should be taken for the noise propagation effects on the resolved profiles, especially for those profile subregions corresponding to incomplete data subsets. For incomplete data sets, application of trilinearity constraints may produce unstable behaviour of the optimised error function, although the quality of the recovered profiles have reduced error propagation effects due to the trilinear constraint more stringent condition. Testing the proposed method for experimental incomplete data multisets, confirmed that the method was working properly although caution should be applied in the case of trilinear constraint implementation.

Acknowledgements

We acknowledge Ministerio de Economía y Competitividad, Spain, for the grant to the research project CTQ2012-38616-C02-01. One of the authors also acknowledges CSIC project grant 201080E033 for her PhD contract.

References

- [1] B. Walczak, D.L. Massart, Dealing with missing data: part I, *Chemometrics and Intelligent Laboratory Systems* 58 (2001) 15–27.
- [2] B. Walczak, D.L. Massart, Dealing with missing data: part II, *Chemometrics and Intelligent Laboratory Systems* 58 (2001) 29–42.
- [3] E. Acar, D.M. Dunlavy, T.G. Kolda, M. Morup, Scalable tensor factorizations for incomplete data, *Chemometrics and Intelligent Laboratory Systems* 106 (2011) 41–56.
- [4] A. Ruhe, Numerical Computation of Principal Components when Several Observations are Missing, Tech. Rep. UMINF-48-74, Depart. of Information Processing, Institute of Mathematics and Statistics, University of Umea, Umea, Sweden, 1974.
- [5] R. Tauler, M. Viana, X. Querol, A. Alastuey, R.M. Flight, P.D. Wentzell, P.K. Hopke, Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, *Atmospheric Environment* 43 (2009) 3989.
- [6] S. Piqueras, L. Duponchel, R. Tauler, A. de Juan, Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares (MCR-ALS), *Analytica Chimica Acta* 705 (2011) 182–192.
- [7] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution, *Analytica Chimica Acta* 500 (2003) 195.
- [8] E.R. Malinowski, D.G. Howery, *Factor Analysis in Chemistry*, 3rd edn Wiley, New York, US, 2002.
- [9] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer-Verlag, New York, USA, 2002.
- [10] H. Parastar, M. Jalali-Heravi, R. Tauler, Is independent component analysis appropriate for multivariate resolution in analytical chemistry? *Trends in Analytical Chemistry* 31 (2011) 134–143.
- [11] E. Spjøtvoll, H. Martens, R. Volden, Restricted least squares estimation of the spectra and concentration of two unknown constituents available in mixtures, *Technometrics* 24 (1982) 173–180.
- [12] J. Jaumot, R. Tauler, MCR-BANDS: a user friendly MATLAB program for the evaluation of rotation ambiguities in multivariate curve resolution, *Chemometrics and Intelligent Laboratory Systems* 103 (2010) 96–107.
- [13] H. Abdollahi, R. Tauler, Uniqueness and rotation ambiguities in multivariate curve resolution methods, *Chemometrics and Intelligent Laboratory Systems* 108 (2011) 100–111.
- [14] R. Tauler, I. Marqués, E. Casassas, Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths, *Journal of Chemometrics* 12 (1998) 55–75.
- [15] A. de Juan, R. Tauler, Comparison of three-way resolution methods for nontrilinear chemical data sets, *Journal of Chemometrics* 15 (2001) 749–772.
- [16] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, *Journal of Chemometrics* 15 (2001) 749–772.
- [17] E. Peré-Trepat, A. Ginebreda, R. Tauler, Comparison of different multiway methods for the analysis of geographical metal distributions in fish, sediments and river waters in Catalonia, *Chemometrics and Intelligent Laboratory Systems* 88 (2007) 69–83.
- [18] M. Alier, M. Felipe, I. Hernández, R. Tauler, Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O₃ pollution in Barcelona, *Analytical and Bioanalytical Chemistry* 399 (2011) 2015–2029.
- [19] R. Tauler, M. Maeder, A. de Juan, Multiset data analysis, *Extended Multivariate Curve Resolution*, *Comprehensive Chemometrics*, vol. 2, Elsevier, 2009, (Chapter 2.24).
- [20] A. Malik, R. Tauler, Extension and application of Multivariate Curve Resolution-Alternating Least Squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India-A case study, 2013 (submitted for publication).
- [21] R. Bro, PARAFAC. Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems* 38 (1997) 149–171.
- [22] A. Smilde, R. Bro, P. Geladi, *Multi-Way Analysis: Applications in the Chemical Sciences*, Wiley, West Sussex, 2004.
- [23] M. Alier, M. Felipe-Sotelo, I. Hernández, R. Tauler, Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods, *Analytica Chimica Acta* 642 (2009) 77–88.
- [24] ISO 13964, http://www.iso.org/iso/iso_catalogue/catalogue_detail.htm?csnumber=23528.
- [25] ISO 7993:1985; UNE 77212:1993; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=14995; <http://www.aenor.es/desarrollo/normas/normas/buscadornormas.asp>.
- [26] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry* 36 (1964) 1627.
- [27] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, 3-way data analysis and ambiguity in multivariate curve resolution, *Journal of Chemometrics* 9 (1995) 31–58.
- [28] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Second ed. John Hopkins University Press, Baltimore, USA, 1989.
- [29] R. Tauler, Multivariate curve resolution applied to second order data, *Chemometrics and Intelligent Laboratory Systems* 30 (1995) 133.
- [30] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics and Intelligent Laboratory Systems* 76 (2005) 101–110.
- [31] R. Bro, S. de Jong, A fast non-negativity-constrained least squares algorithm, *Journal of Chemometrics* 11 (1997) 393–401.
- [32] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner- Washburne, Multivariate curve resolution of time course microarray data, *BMC Bioinformatics* 7 (2006) 343.
- [33] M. Dadashi, H. Abdollahi, R. Tauler, Maximum likelihood principal component analysis as initial projection step in multivariate curve resolution analysis of noisy data, *Chemometrics and Intelligent Laboratory Systems* 118 (2012) 33–40.
- [34] M. Dadashi, H. Abdollahi, R. Tauler, Application of maximum likelihood multivariate curve resolution to noisy data sets, *Journal of Chemometrics* 27 (2013) 34–41.
- [35] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, 2nd ed. Elsevier Science B.V., The Netherlands, 1998.
- [36] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Analytical Chemistry* 63 (1991) 1425–1432.
- [37] M.I. Khoder, Diurnal, seasonal and weekdays-weekends variations of ground level ozone concentrations in an urban area in greater Cairo, *Environmental Monitoring and Assessment* 149 (2009) 349–362.
- [38] R. Tauler, Comments on a recently published paper 'some surprising properties of multivariate curve resolution-alternating least squares (MCR-ALS) algorithms', *Journal of Chemometrics* 24 (2) (2010) 87–90.

5.1.1 DISCUSSIÓ DE RESULTATS

En aquesta secció es presenta la discussió general dels resultats obtinguts en els tres articles continguts en aquest bloc. Aquests resultats s'agruparan segons tres objectius:

- a) Aplicació dels mètodes quimiomètrics PCA i MCR-ALS a dades ambientals de grans dimensions i posterior interpretació dels perfils resolts en base a les seves variacions geogràfiques i temporals
- b) Aplicació de les restriccions de trilinearitat i interacció entre components en el mètode MCR-ALS a dades ambientals de diferent complexitat, i comparació amb mètodes equivalents
- c) Aplicació del mètode MCR-ALS a dades incompletes de contaminació per NO i O₃

a) Aplicació dels mètodes quimiomètrics PCA i MCR-ALS a dades ambientals de grans dimensions i posterior interpretació dels perfils resolts en base a les seves variacions geogràfiques i temporals

La quimiometria és una eina molt potent per treballar amb grans conjunts de dades multivariants ambientals, sobretot quan la complexitat de la informació que contenen és molt gran. En aquesta Tesi, l'aplicació de la quimiometria ha estat una part important que ha permès extreure informació de les dades de que disposàvem.

Un exemple de l'ús dels mètodes quimiomètrics amb dades de grans dimensions es pot veure en el estudi que es publica a l'Article 1, on sense l'ús d'aquesta eina la informació obtinguda en les dades hauria estat difícil de desxifrar. En aquest estudi es disposaven de dades semi-horàries de diferents contaminants durant diversos anys i en diferents estacions. Si es té en compte

que per cada any de mesura de cada contaminant en cada estació concreta tenim un vector de 17520 valors, això ens dona una idea del gran volum de dades que s'ha pogut tractar gràcies als mètodes quimiomètrics.

En aquest primer treball es van utilitzar dades de NO obtingudes durant sis anys i en diferents estacions disposades per tota Catalunya. En la Taula 1 de l'Article 1 es poden veure la mitjana, la desviació estàndard i la concentració màxima semi-horària de les concentracions de NO ($\mu\text{g m}^{-3}$) obtingudes en els anys del 2001 al 2006 en les diferents estacions. Les concentracions més baixes es van trobar en àrees al sud de Catalunya de moderada influència antropogènica, i els valors més elevats es van trobar en l'àrea metropolitana de Barcelona.

Les dades de concentració de NO es van disposar en forma de gran matriu de dades de dimensions 17520 files per 126 columnes. Les 126 columnes corresponien a les 126 estacions de mesura disponibles durant els anys del 2001 al 2006 i les 17520 files corresponien a totes les mesures realitzades en una estació de mesura determinada. Aquest valor 17520 s'obtenia a partir de les mesures semihoràries (48 per dia), durant tots els dies de l'any (365 dies). Com a pas previ abans de les anàlisis quimiomètriques, les possibles dades perdudes ('forats' en la matriu de dades) es van estimar com s'ha explicat en el apartat 4.2.1.

Aquesta matriu de grans dimensions (17520x126) es va analitzar primer per PCA i després per MCR-ALS. Els resultats d'aquestes anàlisis permeten investigar les variacions entre les diferents estacions i les variacions al llarg del temps. PCA i MCR-ALS permeten simplificar la informació contenida en aquesta gran matriu de dades a partir d'un nombre petit de perfils de variació geogràfica i temporal. El resultat de PCA permet investigar les principals pautes de variació contingudes en les dades analitzades, però MCR-ALS pot ajudar a interpretar millor aquesta informació ja que utilitza restriccions més naturals que les que utilitza el PCA com ara la no-negativitat dels perfils resolts enlloc de la seva ortogonalitat com el PCA.

El PCA permet investigar el nombre adequat de components relacionats amb les fonts o pautes de variació principals de les dades analitzades. En el

cas de l'exemple d'aquest treball, es va escollir un model de dos components, amb una variància explicada del 33%. Tot i no ser una variància gaire gran, aquesta ja explicava una part considerable de la informació i l'addició de més components no aportava nova informació. Aquests dos components permetien separar les mostres segons la localització de les diferents estacions. Es van poder distingir diversos grups, però el més significatiu va ser per mostres amb un nivell de contaminació més alt corresponents a les estacions del centre de Barcelona, on el tràfic i l'activitat antropogènica és més important. En l'apartat 3.1 i en la Figura 4 de l'Article 1 es mostra una explicació detallada d'aquests resultats. L'aplicació posterior del mètode MCR-ALS, va permetre una millor interpretació de la informació contenida en les dades. Els perfils dels *scores* obtinguts en aquesta última anàlisi, es reagrupen posant a les files els 48 valors semihoraris de cada dia, i a les columnes els 365 dies de l'any (veure Figura 2b de l'Article 1). D'aquesta manera es poden obtenir també els perfils de variació temporals mitjans diaris i anuals de la variació del NO. Aquests resultats es poden observar a la Figura 5 i s'expliquen detalladament a l'apartat 3.1 de l'Article 1. Breument, en els perfils de variació diària es van observar dos màxims, el primer entre les 8.00 i les 9.00 del matí, i el segon, més petit i ample al voltant de les 20:00h, que es corresponen amb hores punta, on el tràfic és molt intens. El perfil de variació anual té forma de U, on els nivells més alts de NO es troben a la tardor i a l'hivern, degut a les menors temperatures presents en aquestes dues èpoques de l'any. A la primavera i a l'estiu, les elevades temperatures afavoreixen l'augment dels nivells d'ozó, el qual fa disminuir el nivell de NO, degut a la reacció fotoquímica del NO amb el O₃ per produir NO₂, cosa que succeeix a la inversa en la tardor i l'hivern.

Com a conclusió d'aquest estudi es demostra que l'aplicació dels mètodes quimiomètrics PCA i MCR-ALS, ha permès obtenir informació de la distribució geogràfica (estacions de mesura distribuïdes en el territori) i temporal (horària, diària, mensual, estacional i anual) dels perfils de variació de contaminació per NO a Catalunya. Aquesta informació hauria estat més difícil d'expressar i resumir sense la utilització d'aquest tipus d'eines, degut al gran nombre de dades que s'analitzaven simultàniament en aquest estudi.

b) Aplicació de les restriccions de trilinearitat i interacció entre components en el mètode MCR-ALS a dades ambientals de diferent complexitat, i comparació amb mètodes equivalents

El treball anterior en el que han estat analitzades dades de diferents estacions de mesura ha permès investigar els perfils o pautes de variació geogràfica d'un determinat territori, però si es vol aprofundir en les particularitats específiques d'una determinada localització, és millor seguir una altra estratègia d'arranjament de les dades. En aquest exemple, s'ha realitzat l'estudi principalment d'una única estació de mesura, com és l'estació de Barcelona-Gràcia. Es va escollir aquesta estació per la seva disposició en el centre de Barcelona i per disposar de dades més complertes al llarg del temps que en d'altres estacions. El fet de treballar amb una sola estació ha possibilitat l'aplicació de diferents restriccions com la de trilinearitat i la d'interacció entre components, que seria molt més difícil si es barreguessin les dades de diferents localitzacions, ja que és molt difícil que disposin d'un grau d'homogeneïtat suficient i permetin tenir estructures de dades en més de dues direccions.

Estudi de la contaminació per NO en l'estació de Barcelona-Gràcia

Aquest estudi es va realitzar amb les dades de concentració per NO de diferents anys en l'estació de Barcelona-Gràcia. Es van utilitzar dades dels anys 2001, 2002, 2005 i 2006, degut a que en els altres anys hi havia un elevat nombre de dades perdudes (més del 10%). Les dades obtingudes cada any es van agrupar en una matriu de dimensions 365 dies x 48 mesures semi-horàries, pels 4 anys mesurats. Les matrius corresponents als quatre diferents anys disponibles, es van agrupar en una nova matriu augmentada en la direcció de les columnes (veure Figura 3 de l'Article 1). A partir de l'anàlisi exploratori amb PCA, es va escollir un nombre de components igual a tres. Posteriorment, es va fer l'anàlisi amb MCR-ALS, però en aquest cas utilitzant

restriccions de no-negativitat i trilinearitat (model trilineal). Aquesta darrera restricció permet simplificar els resultats i dona resultats únics (sense ambigüitats rotacionals). En aquest cas, es pot obtenir un perfil únic per cada component en els tres modes de mesura: perfils de variació diària (segons les hores del dia), de variació (inter)anual (diferents dies de l'any, que mostren les variacions mensuals i estacionals) i de variació interanual (entre els diferents anys).

Els perfils de variació en els tres modes (variació diària, anual i interanual) i la seva descripció detallada obtinguda a partir de MCR-ALS, utilitzant les restriccions de no-negativitat i de trilinearitat es mostren en la Figura 6 i a l'apartat 3.2 de l'Article 1. Els perfils que es van trobar en aquest estudi coincideixen amb els trobats en altres estudis previs, com per exemple a l'estudi fet amb mostres del nord d'Espanya, a la ciutat de Bilbao (Barrero, 2006) o també a l'estudi realitzat al sud d'Espanya, al Camp de Gibraltar (González, 2006).

Breument, pel segon i tercer component els perfils resolts van ser similars als trobats en l'estudi anterior realitzat per un nombre elevat d'estacions a Catalunya. És a dir, el perfil anual (primer mode) presentava els nivells més alts a la tardor i a l'hivern, i en el perfil diari (segon mode), els màxims nivells de NO es trobaven en les hores punta (al voltant de les 9.00h i al voltant de les 20.00h). Per altra banda el primer component, que és el que conté menor variància però amb nova informació, mostra un perfil de variació de la concentració de NO que augmenta durant la nit. Aquest increment durant la nit, s'explica probablement perquè durant la nit no es produeixen reaccions fotoquímiques i es pot produir l'acumulació de concentracions de NO, o també pot ser degut a l'ús de calefaccions de combustible més emprades durant la nit, principalment a la tardor i a l'hivern. Aquesta variació diària nocturna no presenta un perfil estacional com el que s'observa en la variació anual (primer mode) dels altres dos components, ja que la variació anual d'aquest component presenta un perfil bastant constant al llarg de l'any. En referència a la variació interanual (tercer mode), es va observar un descens de les emissions de NO durant els últims anys de l'estudi, que també s'ha observat en altres ciutats europees com Atenes (Hatzianastassiou, 2007).

Per poder observar millor les diferències entre els dies de la setmana, es va realitzar un nou arranjament als *scores*, disposant aquests valors en columnes ordenats pels dies de la setmana. D'aquesta manera es té una matriu on a les files hi ha el nombre de setmanes i a les columnes els diferents dies de la setmana. Aquest reagrupament confirma el comportament diferent dels dies laborables i els caps de setmana (veure Figura 7 de l'Article 1). Es va observar una clara reducció dels nivells de contaminació per NO durant els caps de setmana, degut principalment a una reducció en el tràfic, confirmant que l'origen d'aquestes emissions és antropogènic. El component que presentava un increment durant la nit, va ser el únic que durant la variació setmanal va presentar un increment de NO durant els caps de setmana, degut probablement al increment del tràfic nocturn en aquests dies.

Els nivells de contaminació de NO a l'estació de Barcelona-Gràcia també es van comparar amb els de les altres dues estacions menys urbanes, com van ser les de Bellver de Cerdanya i la de Tarragona-Sant Salvador. En l'estació de Bellver de Cerdanya, considerada com una localització típicament rural apartada de les contribucions antropogèniques d'una gran ciutat, es van analitzar les dades de NO entre els anys 2004 i 2006. En l'estació de Tarragona-Sant Salvador, considerada com una localització semiurbana amb nivells intermedis de NO, també es van analitzar dades de NO entre els anys 2004 i 2006. En els dos casos, l'anàlisi amb MCR-ALS aplicant restriccions de no-negativitat i trilinearitat, va proporcionar resultats que mostraven components similars als obtinguts en l'estació de Barcelona-Gràcia, però amb intensitats inferiors, que vol dir contaminació inferior per NO. Aquest resultat confirmaria que no hi hauria fonts naturals de NO a la zona geogràfica estudiada o que aquestes són negligibles, i que per tant les principals fonts de contaminació per NO a Catalunya són d'origen antropogènic.

Amb les mateixes dades de NO en l'estació de Barcelona-Gràcia es va fer un nou estudi ampliant les condicions utilitzades en l'estudi anterior. En aquest nou estudi, a més de la restricció de trilinearitat, també es va utilitzar la restricció d'interacció entre components, i es van comparar el resultats amb

els obtinguts amb dos mètodes àmpliament emprats a la bibliografia com són PARAFAC (Bro, 1997b; Harshman, 1970) i TUCKER3 (Smilde, 2004; Tucker, 1966; Andersson, 1998). Aquests mètodes i el funcionament de les restriccions de trilinearitat i d'interacció entre components (amb un nombre de components diferent en els diferents modes de mesura), s'han explicat en detall en l'apartat 4.4 de la part metodològica d'aquesta Tesi.

En la Taula 1 i a la Figura 4 de l'Article 2, es poden veure les variàncies explicades pels diferents mètodes i models emprats, així com els perfils obtinguts en els tres modes en l'anàlisi realitzat per MCR-ALS utilitzant les restriccions de no-negativitat i alternativament la de trilinearitat o la de interacció entre components. En el cas de la restricció que possibilita la interacció entre components, el model millor ha estat el 133, que permet que el perfil en el primer mode (variació anual) sigui el mateix pels tres components resolts. Els perfils obtinguts són molt similars als obtinguts en l'estudi anterior, però en aquest cas, a més del perfil resolt amb la restricció de trilinearitat, també es mostra el perfil bilineal i el perfil obtingut amb la restricció d'interacció entre components. També s'han comparat els resultats obtinguts amb MCR-ALS amb els obtinguts a partir de l'aplicació dels altres mètodes PARAFAC i TUCKER. És a dir, els resultats d'aplicació del mètode MCR-ALS amb la restricció de trilinearitat s'han comparat amb els obtinguts a partir del mètode PARAFAC, i els obtinguts a partir de l'aplicació del mètode MCR-ALS amb la restricció d'interacció entre components s'han comparat amb els obtinguts a partir del mètode TUCKER3. Aquesta comparació de resultats es pot trobar en la Figura S2 de la Informació Suplementària adjunta a l'Article 2. En el cas del PARAFAC la consistència del nucli o *core consistency* va indicar que les dades no podien explicar-se de forma adequada a partir d'un model totalment trilineal encara que els perfils obtinguts fossin semblants als obtinguts per MCR-ALS (tant bilineals com trilineals). La diferència entre les variàncies dels dos models pot ser precisament deguda a que com les dades no s'expliquen bé pel model trilineal, l'aplicació dels dos mètodes que utilitzen estratègies diferents donen resultats diferents. En el cas de la comparació dels resultats obtinguts a partir de l'aplicació de MCR-ALS amb la restricció

d'interacció respecte als obtinguts a partir de l'aplicació del mètode TUCKER3-ALS, els resultats i variàncies explicades han estat pràcticament els mateixos, amb un perfil de variació en el primer mode (perfil anual) similar en els dos casos, que és on es produeix la interacció pels tres components resultants.

Estudi de la contaminació per O₃ en l'estació de Barcelona-Gràcia

En el cas de l'estudi de la contaminació per O₃ a la mateixa estació de Gràcia, les mesures diàries, anuals i interanuals s'han agrupat d'una manera similar a la utilitzada en l'últim estudi explicat pel NO. En aquest cas, les dades de O₃ de les que es disposava eren les corresponents als anys 2000, 2001, 2002 i 2003. Com en el cas de l'estudi de la contaminació per NO, s'ha fet l'anàlisi per PCA i per MCR-ALS aplicant les restriccions de no-negativitat, trilinearitat i d'interacció entre components i també s'ha fet la seva corresponent comparació amb els resultats obtinguts a partir de l'aplicació dels mètodes PARAFAC i TUCKER3. En aquest cas, però, només han calgut dos components per explicar la informació (91% de variància explicada per PCA) contenida en les dades. A la Taula 1 i a la Figura 5 de l'Article 2 es poden veure les variàncies obtingudes pels diferents mètodes i els perfils de variació obtinguts per MCR-ALS en els diferents modes de mesura. En aquest cas, la restricció d'interacció entre components ha estat per un model 122, on com abans és el perfil en el primer mode de mesura (variació anual) el que serà comú per a tots dos components. En la Figura S4 de la Informació Suplementària adjunta de l'Article 2 es poden veure també els perfils corresponents obtinguts pels mètodes PARAFAC (model 2 2 2) i TUCKER3 restringit (model 1 2 2).

A partir de la forma dels perfils de variació diària de O₃ (segon mode del primer component en la Figura 5 de l'Article 2) es pot observar un màxim d'intensitat que coincideix amb el màxim de radiació solar que es produeix durant el dia. Si es compara amb el perfil diari obtingut per el NO (Figura 4 de l'Article 2), es pot comprovar que el màxim d'intensitat del perfil diari de variació de O₃ coincideix amb el mínim del perfil diari corresponent de NO.

Això és degut a la reacció fotoquímica ja mencionada anteriorment, que hi ha entre aquests dos compostos. Aquesta relació inversa també es pot observar en els perfils anuals (primer mode de la Figura 4 i Figura 5 de l'Article 2), on per l'ozó les màximes intensitats del perfil de variació anual es troben a l'estiu i a la primavera, mentre que en el perfil de variació anual de NO es troben a la tardor i a l'hivern. En aquest cas, també es resol un perfil de variació nocturna pel O₃ (segon mode del segon component de la Figura 5 de l'Article 2) que pot ser degut a que la baixada de tràfic que es produeix durant la nit i que fa que es mantinguin els nivells de O₃ adquirits durant el dia. Si es comprova la variació en el perfil anual d'aquest component nocturn, la tendència de valors més elevats a l'estiu i primavera resulta ser menys pronunciada que en el cas del perfil produït durant el dia.

Els perfils obtinguts quan s'apliquen les restriccions de trilinearitat i interacció entre components al mètode MCR-ALS (Figura 5 de l'Article 2) donen resultats similars. En l'aplicació del mètode PARAFAC, la consistència del nucli (*core consistency*) era bastant bona (92.6%) indicant que la variació de les concentracions de O₃ tenen un comportament més trilineal que no pas el NO. Els perfils obtinguts pels mètodes PARAFAC i TUCKER3 (Figura S4 de la Informació Suplementària de l'Article 2) són molt similars en el primer i segon mode del primer component, sent també similars als obtinguts amb MCR-ALS amb les diferents restriccions. En el cas de la variació interanual (tercer mode) del primer component, els mètodes PARAFAC i MCR-ALS amb diferents restriccions donen perfils similars, però el de TUCKER3 restringit (model 122) dóna una intensitat menor per l'any 2000. El perfil diari (segon mode) del segon component obtingut per PARAFAC difereix de l'obtingut pel mètode TUCKER3 restringit (model 122) i també difereix del perfil obtingut per MCR-ALS. Aquest diferent perfil diari del segon component pot ser degut a alguna diferència de comportament en l'any 2000. En el tercer mode del segon component (Figura S4 de la Informació Suplementària de l'Article 2) es pot observar que l'any 2000 predomina en aquest perfil, com ja s'havia vist anteriorment el seu comportament diferent. Aquest comportament diferent de l'any 2000 fa que el perfil resolt per cada mètode sigui diferent, degut a que utilitzen algorismes de resolució diferents.

Estudi conjunt de la contaminació per NO i O₃ en l'estació de Barcelona-Gràcia

L'estudi conjunt de la contaminació per NO i O₃ en l'estació de Barcelona-Gràcia, es va fer pels anys on eren coincidents les mesures realitzades en la mateixa estació, durant els anys 2001 i el 2002. En aquest estudi conjunt es volia posar de manifest més clarament les relacions d'interacció que existeixen entre aquestes dos compostos. Es van fer dos tipus d'arranjament de les dades, una primera fent l'augmentació dels conjunts de dades en la direcció de les columnes ($D_{\text{craug1-O3-NO}}$, de dimensions $(365 \times 4) \times 48$) i l'altra, en la direcció de les files i les columnes simultàniament ($D_{\text{craug2-O3-NO}}$, de dimensions $(365 \times 2) \times (48 \times 2)$). L'esquema d'aquestes augmentacions de matrius es mostra en la Figura 1 de l'Article 2, i també s'ha explicat anteriorment a l'apartat 4.1. (Estructura de les dades). Depenent de la disposició de les matrius de dades es va poder obtenir informació on s'emfatitzin aspectes diferents. En els dos casos, es van aplicar restriccions de no-negativitat i trilinearitat en el mètode MCR-ALS (88.6% i 70.0% de variància explicada respectivament) i es va escollir un model amb tres components. En aquest cas, la restricció d'interacció entre components no va ser possible degut a que ara els perfils de variació anuals (primer mode de les Figures 6 i 7 de l'Article 2) resultaven ser diferents per cada component.

En l'anàlisi per MCR-ALS de la matriu $D_{\text{craug1-O3-NO}}$ (Figura 6 de l'Article 2), es va veure que el comportament dels diferents components resolts era poc trilineal, la qual cosa es va confirmar mitjançant l'anàlisi SVD de la matriu augmentades per columnes i també per la poca consistència del nucli (51.3%) trobat quan s'aplicava el mètode PARAFAC amb tres components. El primer component resolts per MCR-ALS descriu una pauta de variació fonamentalment relacionada amb la pol·lució per NO (42.4% i 37.8% de variàncies explicades pel model bilineal i trilineal respectivament), el segon descriu majoritàriament la pol·lució per O₃ (19.9% i 17.7% de variàncies explicades pel model bilineal i trilineal respectivament) i el tercer descriu la possible interacció entre el NO i el O₃ (26.5% i 33.0% de variàncies explicades

pel model bilineal i trilineal respectivament). Aquesta interacció es mostra sobretot en el tercer mode del tercer component, que es on apareixen conjuntament el O_3 i el NO, on es pot veure que els dos compostos contribueixen a aquest component.

Els resultats obtinguts per MCR-ALS en l'anàlisi de la matriu $D_{\text{craug2-O3-NO}}$ es mostren en la Figura 7 de l'Article 2. En aquest cas, la interacció entre el NO i el O_3 s'observa en el segon mode del primer component, que és on apareixen conjuntament els perfils diaris del NO i del O_3 . En el primer mode del primer component també es pot veure que aquesta relació és bastant constant al llarg de l'any. Cal esmentar, que en tots dos casos ($D_{\text{craug1-O3-NO}}$ i $D_{\text{craug2-O3-NO}}$, Figures 6 i 7 de l'Article 2) es pot veure que la contaminació per NO es va mantenir constant durant els anys 2001 i 2002, i en canvi, la contaminació per O_3 va disminuir lleugerament en l'any 2002.

Com a conclusió dels estudis realitzats es pot concloure que les eines quimiomètriques proposades han permès fer una descripció adequada dels perfils de contaminació per NO i O_3 en l'estació de Gràcia durant el període estudiat. Aquests perfils s'han estudiat de manera individual o combinada i amb diferents restriccions depenent de la informació que es volia treure. En l'estudi del NO es pot concloure que la procedència de la contaminació per NO a Barcelona té un origen principalment antropogènic, ja que els perfils trobats encaixen perfectament amb els patrons de tràfic urbà. Aquests patrons també s'han trobat en altres localitzacions menys urbanes com és la de Bellver de Cerdanya, encara que amb nivells menors. En l'estudi del O_3 s'ha vist que l'evolució d'aquest contaminant està molt influïda per la radiació solar, mostrant els seus nivells més alts a l'estiu i primavera, però que aquests nivells es modulen i depenen de la seva interacció amb el NO. En aquest sentit, cal remarcar que s'han resolt els perfils complementaris de contaminació d'aquests dos contaminants tant en els estudis individuals com en els estudis conjunts. A més ha estat possible la resolució dels possibles perfils d'interacció entre el NO i el O_3 .

c) Aplicació del mètode MCR-ALS a dades incompletes de contaminació per NO i O₃

Com s'ha explicat anteriorment, els conjunts de dades de NO i O₃ només coincidien en els anys 2001 i 2002. Només en aquests casos era possible fer l'anàlisi conjunta de la contaminació per NO i O₃. Per incorporar tota la informació disponible en un mateix estudi, es va proposar una nova versió del mètode MCR-ALS que permet treballar amb aquest tipus de dades incompletes (on només alguns blocs són coincidents en el temps) i extrapolar les estimacions fetes per tots els anys d'estudi. Una descripció detallada d'aquest mètode s'ha explicat a la secció 4.4.2.5.

Per poder comprovar el mètode proposat es va primer treballar amb dades simulades per així poder avaluar millor la fiabilitat del mètode desenvolupat. Les dades simulades es van crear a partir de perfils coneguts, que es corresponien amb els obtinguts anteriorment per la matriu $D_{\text{craug2-NO-O}_3}$. Es van escollir aquests perfils perquè són els corresponents a la part comuna dels dos compostos (NO i O₃) pels anys estudiats (anys 2001 i 2002, veure Figura 4.9 en l'apartat 4.4.2.5) i que es van obtenir en l'estudi anterior. Es van crear uns nous perfils que tenien la mateixa forma que aquests però amb diferents intensitats, per tots els anys en els que es volia fer l'estudi (2000, 2001, 2002, 2003, 2005 i 2006). Es van afegir diferents nivells de soroll (0.1%, 5%, 10% i 20%) a les matrius de dades individuals de cada any. A partir d'aquestes matrius simulades senceres o completes, es va crear la matriu augmentada completa i la matriu augmentada incompleta on alguns dels blocs s'havien eliminat (Figura 4.9).

En la Taula 1 de l'Article 3 es mostren les variàncies finalment obtingudes ($R^2\%$) i la falta d'ajust (Lof%) quan s'escull diferent nombre de components en l'anàlisi per MCR-ALS aplicant restriccions de no-negativitat i trilinearitat, tant per la matrius experimental incompleta, com per les matrius augmentades completes i incompletes simulades i per els diferents nivells de soroll. En el cas de les matrius de dades simulades, també es mostra la

comparació dels perfils en el primer i segon mode resolts per MCR-ALS amb els que es van emprar per a la simulació de les dades. Els resultats d'aquesta comparació es fa a partir del coeficient de correlació (r^2) de l'angle entre els perfils MCR-ALS i els emprats per a la simulació de les dades. Aquestes valors són especialment útils per avaluar la presència d'ambigüitats rotacionals. Valors pròxims a 1 per r^2 i pròxims a zero pels angles confirmen una bona recuperació dels perfils en el primer i tercer mode, mentre que valors de r^2 distants de 1 i angles grans, impliquen males recuperacions amb elevades ambigüitat rotacional.

En la Figura 2 de l'Article 3 es pot observar gràficament el comportament de l'optimització per ALS de la funció d'error en escala logarítmica pels diferents casos d'aquest estudi. En els casos de models bilineals (tant per dades completes com per dades incompletes) hi ha una gran disminució de la funció d'error al principi i després un decreixement més lent fins a la convergència. La convergència és més ràpida per les dades experimentals que per les dades simulades, ja que és més fàcil assolir el criteri de convergència en el casos de soroll real (la funció ja no disminueix més que un determinat valor escollit com a llindar). En el cas de l'aplicació de la restricció de trilinearitat, el comportament és inestable per les dades incompletes, ja que aplicar aquesta restricció en absència de blocs de dades és molt més difícil que en el cas del model bilineal.

Els resultats de l'aplicació del mètode MCR-ALS a la matriu augmentada incompleta de dades experimentals es mostren a la Figura 3 de l'Article 3. En aquest cas, a més del model bilineal, també es va aplicar el model trilineal sobre primer i segon components (model trilineal 110), ja que si s'aplicava també al tercer s'obtenien deficiències de rang. Els perfils obtinguts pel primer i segon components van ser molt similars als obtinguts a l'estudi anterior, on només es va analitzar la part central de les dades (anys 2001 i 2002). El tercer component va mostrar un perfil diari on només sortia un dels màxims de NO, però aquest perfil corresponia principalment a l'any 2003 i tenia una variància

explicada menor. La descripció d'aquests perfils ja s'ha realitzat en apartats anteriors, i també es pot veure en detall en el apartat 4.1 de l'Article 3.

Per les dades simulades, es van analitzar les matrius augmentades de dades completes i incompletes i es van comparar els seus resultats. En el cas de les matrius completes es va utilitzar només dos nivells d'error, un de petit, del 0.1% i un de gran, del 20%. En la Figura 4 de l'Article 3 es pot observar com per a la matriu completa, amb un 20% d'error, els perfils obtinguts aplicant MCR-ALS són pràcticament idèntics als utilitzats en la simulació tant en el model bilineal com en el trilineal. Com era d'esperar el mètode ha funcionat perfectament per la matriu completa, i s'eliminen pràcticament totes les ambigüitats rotacionals (model trilineal) i els efectes del soroll. El logaritme de la funció d'error presenta l'evolució esperada, amb una baixada fins que s'atura dependent del criteri de convergència.

Per a la matriu augmentada simulada incompleta es van utilitzar quatre nivells d'error per veure com afectava aquest nivell als resultats. Els quatre nivells d'error utilitzats van ser 0.1%, 5%, 10% i 20%. La restricció de trilinearitat només es va poder aplicar al primer i segon component per a les matrius incompletes amb error 0.1% i 5%, ja que quan s'aplicava al tercer component es produïa el problema de deficiència de rang. Aquest problema es produeix quan alguns blocs de dades es treuen de la matriu completa. El comportament del logaritme de la funció d'error (Figura 2 de l'Article 3) va dependre del grau d'error afegit a cada matriu. Es va observar una convergència més ràpida amb nivells d'error més elevats, confirmant la major dificultat d'assoliment de la convergència pels nivells de soroll més baixos. En el cas de l'aplicació del model trilineal, especialment amb alts nivells de soroll, es produeixen fortes oscil·lacions en el logaritme de la funció d'error. Normalment el valor mínim s'aconsegueix en les primeres iteracions, a partir de les quals la funció d'error oscil·la, amb augments i disminucions successives. Un possible explicació d'aquest fet es troba en la dificultat d'implementació del model trilineal en el cas de les dades incompletes, però aquest aspecte s'hauria d'investigar més a fons per conèixer més exactament la causa concreta d'aquest comportament.

En les Figures 5 i 6 de l'Article 3 es donen els perfils resultats per MCR-ALS per a la matriu augmentada de dades incompletes en els casos d'error afegit del 5% i del 20% respectivament. En els dos casos, la recuperació dels perfils és bastant bona, sobretot en les regions completes. Quan s'aplica el model bilineal, l'efecte de la propagació del soroll es veu clarament en les regions on les dades estaven incompletes, principalment en els perfils dels *scores* (\mathbf{U}_{aug}). En el model trilineal, encara i tenir un algorisme inestable, la propagació del soroll resulta ser molt inferior, degut a la major robustesa dels models trilineals en comparació amb els models bilineals.

Com a conclusió d'aquest estudi, es pot dir que la nova versió del mètode MCR-ALS per dades incompletes dona resultats útils encara que lògicament pitjors als obtinguts per matrius completes. Quan els nivells de soroll són elevats, s'ha de vigilar l'efecte de la propagació del soroll, especialment en les zones o blocs de dades incompletes. L'aplicació del model trilineal pot resultar inestable, amb oscil·lacions en l'evolució de la funció d'error durant l'optimització ALS, encara que aquest model permet reduir l'efecte de la propagació del soroll.

5.2 BLOC 2 - Contaminants orgànics en filtres de material particulat

En aquest apartat es mostren els resultats dels estudis realitzats sobre contaminants orgànics analitzats en mostres de material particulat. S'han analitzat experimentalment al laboratori mostres de material particulat a Barcelona i a Los Angeles amb cromatografia de gasos acoblada a espectrometria de masses. Les dades obtingudes s'han analitzat amb mètodes quimiomètrics per intentar fer una interpretació ambiental de la seva presència i de les seves concentracions en l'aire, resoldre les seves possibles fonts d'emissió i també comparar les característiques de les diferents localitzacions estudiades.

Article 4 - Marta Alier, Barend L. van Drooge, Manuel Dall'Osto, Xavier Querol, Joan O. Grimalt, Romà Tauler; *Source apportionment of submicron organic aerosol at an urban background and road site in Barcelona (Spain) during SAPUSS*, Atmospheric Chemistry and Physics 13 (2013) 10353–10371.

En aquest treball s'han analitzat per cromatografia de gasos acoblada a espectrometria de masses, diversos contaminants orgànics recollits en filtres de material particulat en dues localitzacions de la ciutat de Barcelona. Entre els contaminants orgànics estudiats es troben els hidrocarburs aromàtics policíclics (PAH), el levoglucosan, els àcids dicarboxílics o alguns derivats de l'isoprè, els quals tenen orígens tant antropogènics com naturals. Els resultats obtinguts s'han analitzat també amb mètodes quimiomètrics com el MCR-ALS per resoldre les possibles fonts de contaminació en el període i localització estudiada.

Informació suplementària de l'Article 4 - Marta Alier, Barend L. van Drooge, Manuel Dall'Osto, Xavier Querol, Joan O. Grimalt, Romà Tauler; *Source*

apportionment of submicron organic aerosol at an urban background and road site in Barcelona (Spain) during SAPUSS.

En aquesta Informació Suplementària s'ha afegit una taula amb un model de set components, enlloc del de sis components escollit a l'Article 4, on es resumeixen les característiques d'aquests components. També s'ha afegit una figura on es mostren els *scores* i els *loadings* del model de set components.

Article 5 - M. Alier, M. Dall'Osto, Y.-H. Lin, J.D. Surratt, R. Tauler, J.O. Grimalt, B.L. van Drooge; *On the origin of water-soluble organic tracer compounds in fine aerosols in two cities: the case of Los Angeles and Barcelona*, Environmental Science and Pollution Research (2014), DOI 10.1007/s11356-013-2460-9.

En aquest treball s'han analitzat filtres de material particulat per cromatografia de gasos acoblada a espectrometria de masses de mostres recollides a les ciutats de Barcelona i Los Angeles. En aquest cas els contaminants orgànics analitzats han estat compostos polars com els àcids dicarboxílics, el levoglucosan o derivats de l'isoprè. Els resultats obtinguts s'han analitzat per MCR-ALS per resoldre les possibles font de d'emissió d'aquests contaminants en aquesta zona i s'han comparat amb els resultats obtinguts a la ciutat de Barcelona.

Atmos. Chem. Phys., 13, 10353–10371, 2013
www.atmos-chem-phys.net/13/10353/2013/
doi:10.5194/acp-13-10353-2013
© Author(s) 2013. CC Attribution 3.0 License.



Atmospheric
Chemistry
and Physics
Open Access

Source apportionment of submicron organic aerosol at an urban background and a road site in Barcelona (Spain) during SAPUSS

M. Alier¹, B. L. van Drooge¹, M. Dall'Osto^{1,*}, X. Querol¹, J. O. Grimalt¹, and R. Tauler¹

¹Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), C/Jordi Girona, 18–26, 08034, Barcelona, Catalonia, Spain

* now at: Institut de Ciències del Mar (ICM-CSIC), Pg. Marítim de la Barceloneta 37–49, 08003 Barcelona, Spain

Correspondence to: B. L. van Drooge (barend.vandrooge@idaea.csic.es)

Received: 8 March 2013 – Published in Atmos. Chem. Phys. Discuss.: 24 April 2013

Revised: 10 September 2013 – Accepted: 12 September 2013 – Published: 24 October 2013

Abstract. This study investigates the contribution of potential sources to the submicron (PM₁) organic aerosol (OA) simultaneously detected at an urban background (UB) and a road site (RS) in Barcelona during the 30 days of the intensive field campaign of SAPUSS (Solving Aerosol Problems by Using Synergistic Strategies, September–October 2010). A total of 103 filters at 12 h sampling time resolution were collected at both sites. Thirty-six neutral and polar organic compounds of known emission sources and photo-chemical transformation processes were analyzed by gas chromatography–mass spectrometry (GC-MS). The concentrations of the trace chemical compounds analyzed are herein presented and discussed.

Additionally, OA source apportionment was performed by multivariate curve resolution–alternating least squares (MCR-ALS) and six OA components were identified at both sites: two were of primary anthropogenic OA origin and three of secondary OA origin, while a sixth one was not clearly defined. Primary organics from emissions of local anthropogenic activities (urban primary organic aerosol, or POA Urban), mainly traffic emissions but also cigarette smoke, contributed 43 % (1.5 µg OC m⁻³) and 18 % (0.4 µg OC m⁻³) to OA at RS and UB, respectively. A secondary primary source – biomass burning (BBOA) – was found in all the samples (average values 7 % RS; 12 % UB; 0.3 µg OC m⁻³), but this component was substantially contributing to OA only when the sampling sites were under influence of regional air mass circulation (REG.). Three secondary organic aerosol (SOA) components (describing overall 60 % of the variance) were observed in the urban ambient PM₁. Products of isoprene oxidation (SOA ISO) – i.e.

2-methylglyceric acid, C₅ alkene triols and 2-methyltetrols – showed the highest abundance at both sites when the city was under influence of inland air masses. The overall concentrations of SOA ISO were similar at both sites (0.4 and 0.3 µg m⁻³, or 16 % and 7 %, at UB and RS, respectively). By contrast, a SOA biogenic component attributed to α-pinene oxidation (SOA BIO PIN) presented average concentrations of 0.5 µg m⁻³ at UB (24 % of OA) and 0.2 µg m⁻³ at RS (7 %), respectively, suggesting that this SOA component did not impact the two monitoring sites at the same level. A clear anti-correlation was observed between SOA ISO and SOA PIN during nucleation days, surprisingly suggesting that some of the growth of urban freshly nucleating particles may be driven by biogenic α-pinene oxidation products but inhibited by isoprene organic compounds. A third SOA component was formed by a mixture of aged anthropogenic and biogenic secondary organic compounds (SOA Aged) that accumulated under stagnant atmospheric conditions, contributing for 12 % to OA at RS (0.4 µg OC m⁻³) and for 18 % at UB (0.4 µg OC m⁻³).

A sixth component, formed by C₇–C₉ dicarboxylic acids and detected especially during daytime, was called “urban oxygenated organic aerosol” (OOA Urban) due to its high abundance at urban RS (23 %; 0.8 µg OC m⁻³) vs. UB (10 %; 0.2 µg OC m⁻³), with a well-defined daytime maximum. This temporal trend and geographical differentiation suggests that local anthropogenic sources were determining this component. However, the changes of these organic molecules were also influenced by the air mass trajectories, indicating that atmospheric conditions have an influence on this component, although the specific origin on this component remains

10354

M. Alier et al.: Source apportionment of submicron organic aerosol

unclear. It points to a secondary organic component driven by primary urban sources including cooking and traffic (mainly gasoline) activities.

1 Introduction

Atmospheric aerosols' influence on the atmospheric visibility (Watson, 2002) is relevant in climate forcing (Forster et al., 2007) and has several adverse effects on human health (Brunekreef et al., 2005). They contain a significant and variable fraction of organic material, ranging from 20 % to 90 % of the submicron ($< 1 \mu\text{m}$ in particle size) particulate matter (PM_{10}) mass (Kanakidou et al., 2005). Such organic material divides into two broad categories termed primary and secondary. Primary organic aerosols (POA) in urban areas are emitted from combustion sources, including heavy- and light-duty vehicles, wood smoke, cooking activities, industries and others. Such primary particles can be modified in the presence of various atmospheric oxidants, such as OH radical, O_3 and NO_x , (Donahue et al., 2009), yielding more oxygenated products that change their original physico-chemical properties. By contrast, secondary organic aerosols (SOA) are formed from reactions of both anthropogenic and biogenic volatile organic compounds (VOC). They are mainly composed of oxygenated organic species (Hallquist et al., 2009). Overall, the knowledge about sources, fate and mutual interaction of gas phase and aerosol organics – of which it is estimated that there are between 10 000 and 100 000 different compounds – is still limited (Schauer et al., 2007; Goldstein and Galbally, 2007; Bi et al., 2008). Particularly, the composition and behaviour of SOA in the atmosphere are not completely well known. For example, several studies showed that diluted emissions from diesel emissions or biomass burning produce large quantities of SOA (Grieshop et al., 2009; Sage et al., 2008), while other field measurements showed the enhanced formation of SOA from gasoline emissions over diesel emissions (Bahreini et al., 2012).

Such complexity of the atmospheric aerosol system is pushing the development of measurement techniques and instrumentation (Pratt and Prather, 2012). In this regard, the aerosol mass spectrometer (AMS) has recently revolutionized the field of atmospheric science because it allows chemical speciation, sizing and mass detection of submicron non-refractory PM at high time resolution (Canagaratna et al., 2007). This technique has demonstrated that a large part of the organic aerosol is highly oxidized and probably formed by SOA (Jimenez et al., 2009; Kanakidou et al., 2005). Recently, a lot of emphasis has been put on the importance of cooking activities for ambient air quality (Allan et al., 2010; Mohr et al., 2012).

However, a main disadvantage of such on-line technique is the limited organic aerosol speciation provided (Pratt and

Prather, 2012). By contrast, the combination of off-line filter techniques in combination with gas chromatography–mass spectrometry (GC-MS) allows a wider organic compound speciation, although covering a very small fraction (often $< 5\%$) of the organic aerosol mass detected. However, the relatively small number of tracer compounds analyzed can be used as input for chemometric methods with the aim to re-construct the emissions from different sources of atmospheric components (Schauer et al., 2007; Stone et al., 2009). The number of existing source apportionment techniques applied on off-line filter data is relatively large, including principal component analysis (PCA; Jolliffe, 2002), chemical mass balance (CMB; US-EPA, 1987) and positive matrix factorization (PMF; Paatero and Tapper, 1994). PCA uses orthogonal constraints, and environmental interpretation of score and loading profiles is sometimes cumbersome. CMB requires the previous knowledge of the emission source profiles, and therefore it is unable to identify unknown sources. PMF integrates natural constraints like non-negativity and uncertainty estimations in a rigorous non-linear optimization of the distribution and composition source profiles. In this work we deploy the multivariate curve resolution–alternating least squares (MCR-ALS) method (Tauler et al., 1995; Tauler, 1995; Jaumot et al., 2005) that has previously been applied for environmental source apportionment (Salau et al., 1997; Terrado et al., 2009). MCR-ALS is based on an alternating linear least squares optimization under non-negativity constraints which produces physically better profiles than PCA, and it has been shown to produce analogous results to PMF (Tauler et al., 2009; Staminirova et al., 2011).

The objective of this study is to characterize the PM_{10} organic aerosol component sampled during the SAPUSS intensive field study conducted in Barcelona (Spain) during the period September–October 2010. This manuscript is part of the SAPUSS special issue (Dall'Osto et al., 2013b). The current study is unique in regards of a large data set of more than 100 PM_{10} filters collected simultaneously at an urban background site (UB) and an urban road site (RS). The urban study area of Barcelona has some unique features. It is characterized by one of the highest density of vehicles in Europe as well as a densely populated city center. Moreover, its geographical position (western Mediterranean basin) favors photo-chemical reactions and accumulation of secondary aerosols (Querol et al., 2009; Perez et al., 2010; Pey et al., 2009).

Past studies at an urban background site in Barcelona (same UB location used in this SAPUSS study) showed that a large fraction (56 %) of the OA consisted of oxygenated OA (Mohr et al., 2012). Moreover, radioactive carbon analysis showed that about 60 % of the organic carbonaceous atmospheric matter in this background site has a non-fossil origin (Minguillón et al., 2011). However, there is evidence that biomass burning contributes only a small part ($\sim 15\%$) of the urban background OA, which is mainly related to the winter season or wild fire events (Mohr et al., 2012; Reche

M. Alier et al.: Source apportionment of submicron organic aerosol

10355

et al., 2012; van Drooge et al., 2012a). This gives way to the hypothesis that other non-fossil sources should account for an important part of the OA. Indeed, Mohr et al. (2012) attributed 17% of the OA to primary food cooking emissions after AMS analysis in the UB site. Nevertheless, there is still a lack of information upon the chemical speciation of this source. Furthermore, the contribution of different types of SOA of both anthropogenic and biogenic origins is still highly uncertain in the urban area of Barcelona.

In summary, the simultaneous analysis of selected organic tracer compounds analyzed from the PM₁ filter samples collected at the UB and at the RS sites is herein presented. Furthermore, a source apportionment of the OA is also carried out and discussed.

2 Materials and methods

2.1 Sampling sites

Detailed information on the urban sites studied in Barcelona is given elsewhere (Dall'Osto et al., 2013b). The urban background site (UB; 41°23'15" N; 02°07'05" E; 80 m.a.s.l.) was situated at the western periphery of the city center in a small park. The road site (RS; 41°23'18" N; 02°09'00" E; 40 m.a.s.l.) was located in a car park next to Urgell Street within the square-grid street network (Eixample) (Fig. 1). The road, which cuts the city from southeast to northwest, is a street canyon composed by a two-way cycling path and a one-way four-lane vehicle road. Vehicle intensity for the month of measurements was about 17 000 vehicles per day. The two monitoring sites are located within 3 km from each other.

The meteorological data were measured in the sampling sites by means of automatic weather stations (Dall'Osto et al., 2013b). Air mass back trajectories were calculated daily during the sampling campaign. Four main air mass scenarios were identified during the SAPUSS field study (Dall'Osto et al., 2013b): Atlantic advection (2 periods; 7 days), North African – west (2 periods; 3 days), North African – east (1 period; 3 days), and Regional (2 periods; 8 days).

2.2 Organic compounds analyzed

The applied extraction method and instrumental analysis offers a selective sight on the total mass of the ambient air PM₁. The usage of organic solvent mixture allows the extraction of a wide range of organic compounds with different polarities. Moreover, the instrumental analysis allows the detection of semi-volatile organics. During the years many of these compounds have been detected in emissions and ambient air samples, and therefore certain organic compounds can be used as tracers for sources and atmospheric processes. Table 1 shows the 36 organic compounds that were analyzed in the present study. They were selected based on their representativeness



Fig. 1. Location of the sampling sites in Barcelona (source: Google Earth).

of primary organic aerosol emission sources and secondary organic aerosol formation. Briefly:

1. Polycyclic aromatic hydrocarbons (PAH, 12 in total). They are toxic components of fossil fuels and primary products of incomplete combustion of fossil fuels and biomass (Rogge et al., 1993; Schauer et al., 2007).
2. Hopanes (17(H) α -21(H) β -29-norhopane and 17(H) α -21(H) β -hopane). They are molecular markers for mineral oils, and their presence can be related to unburned lubricating oil residues from primary vehicles emissions (Rogge et al., 1993; Schauer et al., 2007).
3. Nicotine. This alkaloid is present in high concentrations in environmental tobacco smoke. Although it is mainly present in the gas phase due to its relatively high volatility, it can be detected at trace levels on PM filter samples (Rogge et al., 1994; Bi et al., 2005).
4. Levoglucosan, galactosan and mannosan. They are monosaccharide anhydrides generated by thermal alteration of cellulose and hemi-cellulose. They are emitted in large quantities during biomass burning (Simoneit et al., 2002; Fine et al., 2004).
5. Dicarboxylic acids (DCA), hydroxy-DCA and aromatic-DCA (11 in total, Table 1). They can be emitted from various primary sources (mobile emission, meat cooking, etc.) although atmospheric photo-chemical formation is probably the main source (Jang et al., 1997; Kerminen et al., 2000; Heald et al., 2010; Sheesley et al., 2010; Paulot et al., 2011).

10356

M. Alier et al.: Source apportionment of submicron organic aerosol

Table 1. Summary of the concentrations (ng m^{-3}) of analyzed organic tracer compounds. The results indicated in bold showed significant (t test, $p < 0.05$) differences between sites, day and night, or between weekday (WD) and weekends/holidays (WE).

		UB		RS		UB vs. RS		ratio day/night		ratio WD/WE	
		mean sd	(min–max)	mean sd	(min–max)	ratio	r^2	UB	RS	UB	RS
polycyclic aromatic hydrocarbons	phenanthrene	0.05 ± 0.02	(0.01–0.12)	0.09 ± 0.03	(0.04–0.17)	0.6	0.5	0.8	0.9	1.6	1.4
	anthracene	0.01 ± 0.00	(0.00–0.02)	0.02 ± 0.01	(0.01–0.04)	0.5	0.4	1.0	1.1	1.5	1.6
	fluoranthene	0.06 ± 0.03	(0.02–0.14)	0.13 ± 0.04	(0.06–0.21)	0.5	0.4	0.8	0.9	1.7	1.4
	pyrene	0.08 ± 0.05	(0.02–0.22)	0.20 ± 0.07	(0.10–0.38)	0.4	0.4	0.8	0.9	1.9	1.5
	benz[a]anthracene	0.05 ± 0.03	(0.02–0.12)	0.13 ± 0.05	(0.04–0.24)	0.4	0.4	0.9	1.3	1.7	1.8
	chrysene	0.07 ± 0.03	(0.02–0.14)	0.16 ± 0.07	(0.05–0.31)	0.4	0.5	0.9	1.2	1.6	1.7
	benzofluoranthenes	0.13 ± 0.07	(0.03–0.37)	0.17 ± 0.09	(0.03–0.48)	0.7	0.3	1.0	1.2	1.3	1.5
	benzo[e]pyrene	0.15 ± 0.08	(0.05–0.37)	0.33 ± 0.16	(0.07–0.73)	0.4	0.5	0.9	1.3	1.7	1.9
	benzo[a]pyrene	0.06 ± 0.04	(0.01–0.20)	0.18 ± 0.09	(0.03–0.40)	0.4	0.4	0.9	1.3	1.6	2.1
	indeno[123cd]pyrene	0.04 ± 0.03	(0.01–0.12)	0.11 ± 0.06	(0.02–0.27)	0.4	0.2	0.9	1.2	1.0	1.5
	benzo[ghi]perylene	0.12 ± 0.05	(0.04–0.26)	0.30 ± 0.13	(0.07–0.56)	0.4	0.4	0.9	1.4	1.4	1.9
	coronene	0.04 ± 0.02	(0.01–0.08)	0.09 ± 0.03	(0.03–0.17)	0.5	0.2	1.0	1.4	1.2	1.5
	hopanes	17a(H)21β(H)-29-norhopane	0.52 ± 0.22	(0.20–1.11)	0.73 ± 0.37	(0.20–1.70)	0.7	0.5	1.1	1.1	1.5
17a(H)21β(H)-hopane		0.46 ± 0.22	(0.19–1.16)	0.79 ± 0.41	(0.20–1.88)	0.6	0.5	1.0	1.1	1.6	2.0
alkaloid	nicotine	7.0 ± 5.6	(0.5–20.0)	58.1 ± 39.8	(9.2–164)	0.1	0.4	1.3	1.5	1.8	2.5
anhydro-sugars	levoglucosan	9.4 ± 9.4	(1.7–40.0)	5.3 ± 5.8	(0.6–30.5)	1.8	0.8	0.9	0.9	0.7	0.8
	galactosan	1.3 ± 1.6	(0.2–7.1)	0.7 ± 0.9	(0.1–4.4)	1.8	0.8	0.9	0.9	0.6	0.9
	mannosan	1.2 ± 1.4	(0.2–5.6)	0.6 ± 0.7	(0.1–3.8)	1.9	0.8	0.8	0.9	0.6	0.7
acids, polyols, methyltetrols	malonic acid C ₃	1.5 ± 1.5	(0.3–8.1)	1.6 ± 1.0	(0.4–4.9)	1.0	0.5	1.1	1.3	0.9	1.2
	succinic acid C ₄	7.3 ± 4.8	(3.1–24.7)	6.2 ± 3.0	(2.5–14.8)	1.2	0.3	1.1	1.0	0.9	0.9
	glutaric acid C ₅	2.0 ± 1.3	(0.6–7.5)	1.5 ± 0.7	(0.6–3.2)	1.3	0.3	1.3	1.2	1.0	1.0
	pimelic acid C ₇	0.6 ± 0.3	(0.2–2.1)	1.3 ± 0.6	(0.6–3.7)	0.5	0.5	1.3	1.6	0.9	1.0
	suberic acid C ₈	0.9 ± 0.5	(0.4–3.0)	1.7 ± 0.7	(0.7–3.9)	0.5	0.2	1.2	1.5	1.0	1.0
	azelaic acid C ₉	2.9 ± 1.3	(1.0–7.3)	6.1 ± 2.6	(2.4–13.0)	0.5	0.2	1.2	1.7	1.0	1.2
	glyceric acid	3.2 ± 2.9	(0.7–15.2)	3.4 ± 2.0	(1.1–9.8)	0.9	0.5	1.0	1.1	0.9	1.1
	malic acid	14.2 ± 14.3	(1.2–72.7)	9.9 ± 8.5	(1.2–33.6)	1.4	0.7	1.2	1.0	1.0	0.9
	tartaric acid	2.1 ± 2.8	(0.0–12.2)	1.8 ± 1.7	(0.1–7.0)	1.2	0.6	1.5	1.2	0.9	0.9
	tricarballic acid	3.9 ± 3.3	(0.3–15.3)	3.0 ± 2.4	(0.4–11.6)	1.3	0.6	1.2	1.2	1.1	0.9
	phthalic acid	3.9 ± 2.4	(0.9–11.4)	3.6 ± 1.6	(1.3–9.3)	1.1	0.7	1.4	1.2	1.1	1.1
	3-hydroxyglutaric acid	4.5 ± 3.5	(0.5–20.2)	3.1 ± 2.1	(0.8–8.6)	1.5	0.7	1.1	1.0	1.0	1.1
	MBTCA	5.5 ± 4.5	(1.0–23.5)	4.1 ± 2.4	(0.7–10.3)	1.4	0.6	1.3	1.3	1.2	1.2
	cis-pinonic acid	15.4 ± 8.4	(3.6–49.7)	8.0 ± 2.9	(3.1–18.4)	1.9	0.6	1.0	1.1	1.0	1.0
	C ₅ -alkene triols	1.0 ± 0.7	(0.1–3.0)	0.7 ± 0.5	(0.1–2.2)	1.4	0.7	1.1	1.0	1.2	1.4
	2-methylglyceric acid	2.1 ± 1.3	(0.4–6.4)	1.9 ± 1.1	(0.6–4.5)	1.1	0.5	1.2	1.1	1.1	1.3
	2-methylthreitol	2.8 ± 1.6	(0.6–7.8)	1.4 ± 0.9	(0.1–3.8)	1.9	0.6	1.1	1.0	1.2	1.1
2-methylerythritol	6.5 ± 4.1	(1.7–20.7)	3.5 ± 2.4	(0.6–9.5)	1.9	0.6	1.1	1.0	1.2	1.2	

6. Cis-pinonic acid, 3-hydroxyglutaric acid and 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA). They have been identified in aerosols and related to the photo-chemical oxidation of biogenic volatile compounds such as α -pinene (Claeys et al., 2007; Szmigielski et al., 2007). 3-MBTCA is formed by OH-initiated oxidation of cis-pinonic acid (Szmigielski et al., 2007) and was first detected in aerosol samples from Amazonia and Belgium (Kubatova et al., 2000). It is worth noting that the most abundant monoterpene in the study area presented in this manuscript is α -pinene (Seco et al., 2011).

7. 2-methylglyceric acid and polyols, such as C₅ alkene triols and 2-methyltetrols. They have been related to isoprene oxidation (Claeys et al., 2004; Hallquist et

al., 2009), a major volatile organic compound emitted from land vegetation.

2.3 Analytical procedure

Digitel-DH80 HiVol samplers were used to collect 103 PM₁ filters (Pall Life Sciences; TISSUEQUARTZ 2500QAT-UP), each encompassing 12 h sampling hours, between 09:00 and 21:00 (UTC), at a sampling rate of 30 m³ h⁻¹. More information can be found elsewhere (Dall'Osto et al., 2013b). All filters were preheated at 450 °C overnight before sampling. A 1/8 section of each filter was preserved for analysis of organic compounds. This section was ultrasonically extracted with (2 : 1, v/v) dichloromethane : methanol (3 × 5 mL; Merck, Germany) for 15 min. Before extraction 25 μ L of the surrogate standards levoglucosan-d₇, n-C₂₄d₅₀

(Cambridge Isotopic Laboratories, UK), succinic acid- d_4 (Sigma Aldrich), anthracene- d_{10} , benz[a]anthracene- d_{12} , benzo[k]fluoranthene- d_{12} and benzo[ghi]perylene- d_{12} (Dr. Ehrenstorfer) were added. The extracts were filtered over 0.45 μm teflon membrane filters (Whatman) in order to remove insoluble particles. Then, they were concentrated to 1 mL under a gentle N_2 -gas stream.

Anhydro saccharides, acids, polyols and nicotine were analyzed using a procedure that was similar to others described previously (El Haddad et al., 2011; Medeiros and Medeiros, 2007; van Drooge et al., 2012a). Briefly, an aliquot of the extract (25 μL) was evaporated under a gentle N_2 stream until dryness. Then, 25 μL of bis(trimethylsilyl)trifluoroacetamide (BSFTA)+trimethylchlorosilane (99:1) (Supelco) and 10 μL of pyridine (Merck) were added for derivatization of the saccharides, acids and polyols to their trimethylsilyl esters at 70 $^\circ\text{C}$ during 1 h. Before injection into a gas chromatograph coupled to a mass spectrometer (GC-MS), 25 μL of the internal standard, pyrene- d_{10} , were added to the vial.

For the analysis of polycyclic aromatic hydrocarbons and hopanes, the remaining extract was evaporated to almost dryness under a gentle N_2 -gas stream and re-dissolved in 0.5 mL hexane+dichloromethane (9:1 v/v) (Merck, Germany). Then, it was cleaned up by adsorption column chromatography on 1 g of aluminum oxide (Merck, Germany) that was activated overnight at 120 $^\circ\text{C}$. The analytes were eluted with 4 mL of (9:1 v/v) hexane:dichloromethane and 4 mL of (1:2 v/v) hexane:dichloromethane, respectively (Merck, Germany). The collected fraction was concentrated under a gentle N_2 -gas stream to 50 μL , and the internal standard, pyrene- d_{10} , was added before injection into GC.

Sample extracts were injected into a Thermo GC/MS (Thermo Trace GC Ultra – DSQ II) equipped with a 60 m fused capillary column (HP-5MS 0.25 mm \times 25 μm film thickness). The oven temperature program started at 60 $^\circ\text{C}$ held for 1 min, and then it was heated to 120 $^\circ\text{C}$ at 12 $^\circ\text{C min}^{-1}$ and to 310 $^\circ\text{C}$ at 4 $^\circ\text{C min}^{-1}$ where it was held for 10 min. The injector, ion source, quadrupole and transfer line temperatures were 280 $^\circ\text{C}$, 200 $^\circ\text{C}$, 150 $^\circ\text{C}$ and 280 $^\circ\text{C}$, respectively. Helium was used as a carrier gas at 0.9 mL s^{-1} . The MS selective detector was operated in fullscan (m/z 50–650) and electron impact (70 eV) ionization modes.

Besides comparison of retention times, levoglucosan and mannosan were identified with ion m/z 204, galactosan with ion m/z 217 and nicotine with ion m/z 84. Acids, polyols and 2-methyltetrosols were identified with the following ions: malonic acid (m/z 233), succinic acid (m/z 247), glutaric acid (m/z 261), pimelic acid (m/z 289), suberic acid (m/z 303), azelaic acid (m/z 317), glyceric acid (m/z 292), malic acid (m/z 233), tartaric acid (m/z 292), phthalic acid (m/z 295), tricarballic acid (m/z 377), cis-pinonic acid (m/z 171), 3-hydroxyglutaric acid (m/z 349), 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) (m/z 405),

2-methylglyceric acid (m/z 219), C_5 -alkene triols (m/z 231), 2-methylthreitol and 2-methylerythritol (m/z 219). Quantification was performed with the external standard calibration curves. All concentrations were corrected by the recoveries of the surrogate standard succinic acid (m/z 251) and levoglucosan- d_7 (m/z 206). No standards were available for 3-hydroxyglutaric acid, MBTCA, 2-methylglyceric acid, C_5 alkene triols, 2-methylthreitol and 2-methylerythritol. Their chromatographic peaks were identified by comparison of their mass spectra to literature and library data (Claeys et al., 2007; Kourtchev et al., 2005; Clements and Seinfeld, 2007), and they were quantified with the calibration curve of succinic acid. Application of the calibration curves of other standards leads to lower concentrations with a maximum variation of factor of 3. Therefore, caution should be taken when comparing these results with those from other studies.

PAH were identified by retention time comparison of the peaks with the following ions in SIM mode: phenanthrene (m/z 178), anthracene (m/z 178), fluoranthene (m/z 202), pyrene (m/z 202), benz[a]anthracene (m/z 228), chrysene+triphenylene (m/z 228), benzo[b]fluoranthene (m/z 252), benzo[k]fluoranthene (m/z 252), benzo[e]pyrene (m/z 252), benzo[a]pyrene (m/z 252), indeno[1,2,3-cd]pyrene (m/z 276), benzo[ghi]perylene (m/z 276) and coronene (m/z 300). 17(H) α -21(H) β -29-Norhopane and 17(H) α -21(H) β -hopane were identified in the m/z 191 mass fragmentogram and the corresponding retention times. Quantification was also performed by the external standard method, and the calculated concentrations were corrected for the recoveries of the abovementioned surrogates: anthracene- d_{10} (m/z 188), benz[a]anthracene- d_{12} (m/z 240), benzo[k]fluoranthene- d_{12} (m/z 264) and benzo[ghi]perylene- D_{12} (m/z 288).

In all cases the recoveries of the surrogate standards were higher than 70%. Field blank levels were between 1% and 30% of the sample levels. All concentrations were corrected for blank levels. Limits of quantification (LOQ) were calculated by dividing the lowest measured levels in the standard calibration curves by the volumes of the analyzed sample fraction. These were 0.1 ng m^{-3} for the anhydrosaccharides, 0.06 ng m^{-3} for the acids and 0.005 ng m^{-3} for PAHs and hopanes.

2.4 Chemometrics

MCR-ALS method has been successfully used in the analysis of environmental data sets, both in contamination studies of surface waters (Terrado et al, 2009) and in air source apportionment studies (Salau et al., 1997; Tauler et al., 2009). One of the main advantages of MCR-ALS is that it decomposes the data matrix by applying more natural constraints than PCA, such as non-negativity, and, therefore, interpretation of results is more straightforward.

MCR-ALS is based on a bilinear decomposition of the original data set. In matrix form, it is expressed as $\mathbf{D}(I \times J)$

10358

M. Alier et al.: Source apportionment of submicron organic aerosol

$= \mathbf{U}(I \times N)\mathbf{V}^T (N \times J) + \mathbf{E}(I \times J)$, where \mathbf{D} is the original data array, with I rows (samples) and J columns (compounds); \mathbf{U} is the matrix of scores of dimensions $I \times N$, where N is the reduced number of components; \mathbf{V}^T is the matrix of loadings with dimensions $N \times J$; and \mathbf{E} is the matrix of residuals not modeled by the N components. The MCR-ALS method decomposes the data matrix using an alternating least squares algorithm under a set of constraints such as non-negativity, unimodality, closure, trilinearity or selectivity (Tauler et al., 1995; Tauler, 1995; Jaumot et al., 2005). The variance explained by the different components is overlapped and not orthogonal (non-overlapped) as in PCA (Jolliffe, 2002). Natural sources in the environment are rarely orthogonal; thus the MCR methods may have a physically sounder interpretation than orthogonal database decomposition methods.

For the multivariate data analysis, the experimental data were stored in 12 h periods between 22 September 2010 and 18 October 2010 at the two sites. This generated data matrices of 52 and 51 samples for UB and RS stations, and the number of compounds analyzed in each was 35. Indeno[123-cd]pyrene was finally excluded from the databases because 20% of the sample concentrations were below LOQ. The merged matrix (52×35 for UB and 51×35 for RS) allows a better comparison of potential source apportionment profiles than the individual matrices. The observed sources are representative of the two monitoring sites. The joint data set was imported into MATLAB 7.4 (The Mathworks, Natick, USA) for subsequent calculations using MATLAB PLS 5.8 Toolbox (Eigenvector Research Inc, Masson, WA, USA).

3 Results

3.1 Chemical analysis of PM₁ filter samples

Means, standard deviations ($\pm 1\sigma$), maximum and minimum concentration values of the compounds analyzed are described in Table 1.

3.1.1 PAH

The total concentrations of the 12 PAHs analyzed were two times higher at the RS (1.91 ng m^{-3}) than at the UB site (0.89 ng m^{-3}), reflecting a stronger influence of residues from fossil fuel vehicles at the former site. This difference was similar to that observed for the concentrations of NO_x (46 vs. $33 \text{ } \mu\text{g m}^{-3}$) and elemental carbon (1.4 vs. $0.7 \text{ } \mu\text{g m}^{-3}$) (Dall'Osto et al., 2013b). Despite these differences in concentrations, there was a significant, moderately positive correlation between the PAH concentrations at the two sites ($r^2 = 0.55$, $P < 0.05$). This can be interpreted to reflect the atmospheric conditions affecting the distributions of primary atmospheric pollutants in the urban area under study.

The average isomeric ratios fluoranthene vs. pyrene (Fl/(Fl+Pyr)), benz[a]anthracene vs. chrysene

(BaA/(BaA+Chr)), benzo[e]pyrene vs. benzo[a]pyrene (BeP/(BeP+BaP)), and indeno[123cd]pyrene vs. benzo[ghi]perylene (IP/(IP+BgP)) – 0.42 ± 0.05 , 0.43 ± 0.03 , 0.68 ± 0.06 and 0.26 ± 0.09 , respectively – were very similar among all analyzed samples at both sites. These ratios are consistent with a dominance of fossil fuel combustion inputs from traffic related activities (Galarneau et al., 2008).

At the RS, the detected PAH of high molecular weight (PAH_{high}, $m/z > 228$; see Table 1) showed higher concentrations during daytime than at nighttime (1.3 , $P < 0.05$), whereas no difference was found for PAH of low molecular weight (PAH_{low}, $m/z > 228$; see Table 1). By contrast, at the UB site no day/night difference was found for all the PAH detected. Several factors could be responsible for the day/night pattern of PAH concentrations at the RS site, including traffic intensity, vehicle fleet (e.g. diesel, gasoline) and meteorological conditions (e.g. temperature). Although the traffic intensity showed a marked day/night trend (ratio 2.4), the PAH concentrations did not correlate with such intensity. Such low correlation may be due to gas to particle partitioning of PAH compounds, which has a stronger influence on the more volatile PAH_{low} (Lohmann and Lammel, 2004).

The observed \sum PAH concentrations in Barcelona were in the range of the ones observed in other urban areas in the USA (Ning et al., 2007; Subramanian et al., 2006) and Europe (Křumal et al., 2013, and references therein), although wintertime concentrations in those site are generally higher, due to enhanced fuel combustion for domestic heating. The relative composition of the PAH compounds in the ambient air PM in Barcelona was found enriched with lower molecular weight PAH (PAH_{low}; phenanthrene until pyrene; $m/z < 228$), whereas the PM₁ in the urban areas in the USA were more enriched with higher molecular weight PAH (PAH_{high}; benz[anthracene until coronene; $m/z > 228$). These differences between the urban locations could be due to the higher proportion of diesel engine vehicles over gasoline vehicle in Barcelona in comparison to the sites in the USA, where gasoline vehicles dominate (Rogge et al., 1993; Zielinska et al., 2004).

Finally, it is worth noting that the PAH concentrations were about a factor 1.5 higher during weekdays in comparison to weekends/holidays days at both sites (Table 1), and this factor was found to be similar to the one of vehicle intensities near the two monitoring sites (1.4; Dall'Osto et al., 2013a).

3.1.2 Hopanes

Total hopanes concentrations were 0.75 ng m^{-3} and 0.50 ng m^{-3} at RS and at the UB sites, respectively (Table 1), reflecting – likewise for the PAH concentrations – the strong influence of fossil fuel vehicle combustion at the former site. The observed \sum hopane concentrations in BCN

fit into the range of the concentrations observed in urban areas in the USA and Europe (Subramanian et al., 2006; Ning et al., 2007; Kfomal et al., 2013). Although hopanes are susceptible to oxidation in the atmosphere (Robinson et al., 2006a; Lambe et al., 2009), the depletion is probably small due to the very short distance between the sampling site and emission source. Temporal trends of hopanes were also found to be moderately correlated between the two monitoring sites ($r^2 = 0.65$, $P < 0.05$). There were no differences between day and night periods, likewise for the PAH_{low} concentrations. The relationship between PAH_{low} and hopanes reflects the primary emissions of diesel vehicles (Rogge et al., 1993; Zielinska et al., 2004). Higher concentrations were observed during weekdays in comparison to weekends/holidays (factor 1.6 at UB and 2 at RS, respectively; Table 1).

3.1.3 Nicotine

This alkaloid was found in all filter samples. Much higher concentrations (58 ng m^{-3}) were observed at the RS site than at UB (7 ng m^{-3}). This eight-fold difference points to a very strong outdoor cigarette consumption near RS, which is situated next to a busy street and an exit of an underground metro station. The UB sampling station is situated in a local park and is away from streets and pedestrians. It is worth noting that Spain ranks among the countries of highest cigarette consumption in the European Union and in the world (WHO, 2010). High outdoor gas-phase nicotine concentrations were previously measured in Barcelona in summer 2010 between 0.5 and $1.5 \mu\text{g m}^{-3}$ (Sureda et al., 2012).

Nicotine concentrations showed a moderate correlation ($r^2 = 0.4$, $P < 0.05$) between the two monitoring sites, suggesting that the whole urban atmosphere of Barcelona was generally influenced by outdoor tobacco smoke. At RS nicotine concentrations had a day/night ratio of 1.5 ($P < 0.05$) at the RS, whereas a statistical difference was not found at the UB site. Significantly higher nicotine concentrations were detected during weekdays than on weekends/holidays (ratio of 1.8 and 2.5 at UB and RS, respectively, $P < 0.05$; Table 1). This indicates that the nicotine concentrations are more affected by anthropogenic working activities during daytime of the weekdays rather than leisure ones occurring during weekends. This result is in line with other SAPUSS studies of single particle aerosol time-of-flight mass spectrometry showing that nicotine concentrations correlate with traffic activities and not with gastronomic ones (Dall'Osto et al., 2013b).

3.1.4 Anhydro saccharides

In this group of compounds, mannosan and galactosan showed low concentrations (about $2\text{--}4 \text{ ng m}^{-3}$), and also the levoglucosan concentrations were relatively low (9 and 5 ng m^{-3} at UB and RS, respectively; Table 1). These lev-

els are similar to previous studies in the same region during the summer season (van Drooge et al., 2012b) and in European background sites with low influence of biomass burning (Puxbaum et al., 2007). In contrast, higher levoglucosan concentrations were found at the UB site in former studies conducted during wintertime ($35\text{--}60 \text{ ng m}^{-3}$, Reche et al., 2012; van Drooge et al., 2012a). In the present study, these high levels were only observed at both sites at the end of the sampling period (about 35 ng m^{-3} , 15–20 October 2010). This increase was coincident with the legal permission of burning biomass waste in regional background fields from 15 October and onwards (DOGC 1995) as well as with a stagnant regional anticyclonic conditions (Dall'Osto et al., 2013b). The temporal concentration changes of the anhydro saccharides compounds at both sites were very similar ($r^2 = 0.88$, $P < 0.05$).

However, the higher levels found at the UB site relative to the RS site (ratio 1.9; $P < 0.05$) are likely due to the higher exposure of the UB site from regional sources because the RS site is located within a network of streets that reduce natural ventilation. This difference between the two stations was more defined in the last period of the sampling when higher concentrations were detected.

3.1.5 Dicarboxylic acids (DCA), hydroxy-DCA and aromatic-DCA

Mean concentrations of individual DCA compounds were found ranging between 1 and 14 ng m^{-3} and between 1 and 10 ng m^{-3} at the UB and RS, respectively. The highest individual mean concentrations at both sites were observed for malic acid (14 ng m^{-3} at UB; 10 ng m^{-3} at RS) and succinic acid (7 ng m^{-3} at UB; 6 ng m^{-3} at RS). This range of concentrations is similar to those measured during wintertime at the UB site (Van Drooge et al., 2012a). The higher ratio of malic acid/succinic acid was previously observed in other sites influenced by anthropogenic emission sources (Yang et al., 2008a). Phthalic acid – a secondary organic tracer for fossil fuel combustion – showed similar mean concentrations at both sites (4 ng m^{-3} , Table 1). As regards the temporal trends of the DCA concentrations during SAPUSS, the majority of them were found similar between the two monitoring sites ($0.2 < r^2 < 0.8$, $P < 0.05$; Table 1).

Most DCA concentrations were found slightly higher at the UB site relative to the RS one (UB/RS about 1.2), with the exception of the $C_7\text{--}C_9$ dicarboxylic acids (pimelic, suberic and azelaic acids), which were found to possess a UB/RS ratio of 0.5. Moreover, in contrast to any of the other analyzed compounds, $C_7\text{--}C_9$ dicarboxylic acids showed a clear daytime maximum in their concentrations, especially at the RS site (Table 1). Among the $C_7\text{--}C_9$ DCA analyzed, azelaic acid showed the highest concentrations (3 ng m^{-3} and 6 ng m^{-3} at UB and RS, respectively). The presence of azelaic acid in ambient air PM has been related to fast oxidation of unsaturated fatty acids, i.e. oleic acid, by ozone (Gogou et al.,

10360

M. Alier et al.: Source apportionment of submicron organic aerosol

1994; Moise and Rudich, 2002). The oxidation of biogenic precursors can also be an important source of C₇–C₉ dicarboxylic acids (Kawamura and Gagosian, 1987; Stephanou and Stratigakis, 1993; Mochida et al., 2003; Zhang et al., 2010; Ho et al., 2011). Oleic acid in the urban atmosphere may have many sources, such as food cooking, traffic and biogenic sources, including the marine environment (Fang et al., 2002; Schauer et al., 2002; Robinson et al., 2006b). In the present study, oleic acid was identified in all samples (including field blank filters) at low concentrations (<2 ng m⁻³). The lack of substantial oleic acid concentrations in the samples may be due to the chemical instability of this unsaturated fatty acid and/or a fast oxidation to other compounds in the urban atmosphere. Further considerations on possible sources and atmospheric processes responsible for the elevated azelaic acid concentrations detected are given in Sect. 3.2, where an aerosol source apportionment is attempted.

3.1.6 α -pinene SOA tracers: cis-pinonic acid, 3-hydroxyglutaric acid and MBTCA

Cis-pinonic acid, 3-hydroxyglutaric acid and MBTCA, the three biogenic SOA tracers for α -pinene oxidation, were detected in all samples. Cis-pinonic acid was found in two-times higher concentrations at the UB sites (15 ng m⁻³) than at the RS site (8 ng m⁻³). The other two tracer compounds (3-hydroxyglutaric acid and MBTCA) showed lower average concentrations (3–5 ng m⁻³) (Table 1) but still statistically higher at UB. All the observed SOA pinene tracer concentrations are similar to the ones measured by El Haddad et al. (2011) in a nearby French Mediterranean urban area. There was no substantial correlation between the temporal concentration trends of cis-pinonic acid and the other two tracer compounds ($r^2 = 0.1$), whereas there was a very good correlation between 3-hydroxyglutaric acid and MBTCA in the two different monitoring sites ($r^2 = 0.65$; 0.75, $P < 0.05$). The same correlations between these three biogenic SOA tracers were recently discussed by Ding et al. (2011). Oxidation of α -pinene – emitted mainly by coniferous trees – forms pinic acid and cis-pinonic acid. Previous studies proposed that cis-pinonic acid and pinic acid could further photo-degrade to MBTCA (Claeys et al., 2007; Szmigielski et al., 2007). Although MBTCA has other precursors apart from α -pinene (Szmigielski et al., 2007), the significant correlations among MBTCA and 3-hydroxyglutaric acid suggest that α -pinene is probably its major precursor.

The ratio of cis-pinonic acid to MBTCA (P/M) can be employed to estimate the aging of α -pinene SOA (Ding et al., 2011). The P/M ratios at the UB and at the RS were 4.0 and 2.8, respectively, indicating that overall the α -pinene SOA tracers were relatively fresh during our campaign. It is important to note that cis-pinonic acid was found to possess the highest ratio between the two sites (UB/RS = 1.9) of all the 36 detected organic compounds in this study, implying this

fresh SOA oxidation product may be further oxidized in the more urban city center (RS site).

3.1.7 Isoprene SOA tracers: C₅-alkene triols, 2-methylglyceric acid, 2-methylthreitol and 2-methylerythritol

It is generally considered that the main atmospheric source of isoprenoid VOCs is the direct emission from vegetation (Guenter et al., 1995). However, tailpipe sources may also contribute a non-negligible amount of isoprene in urban areas (Borbon et al., 2001; Park et al., 2011). Indeed, previous studies in the metropolitan region of Barcelona described the human-related sources of isoprene (Seco et al., 2013). However, whilst the predominately anthropogenic origin is mainly seen in winter months, during the summer ones (including September and October, this study) a mainly biogenic origin is apportioned (Filella and Penuelas, 2006).

During SAPUSS, all the isoprene-related markers' concentrations were found higher at the UB site than the more anthropogenically influenced RS site (Table 1). Whilst it is not possible to uniquely attribute isoprene markers only to biogenic sources, previous urban air quality studies suggest that the majority of isoprene SOA should be of biogenic origin. For example, Langford et al. (2010) concluded that as much as 80 % of isoprene in London has biogenic sources. Schneidmesser et al. (2011) concluded that isoprene does not have a very important role in atmospheric chemistry in London or Paris. However, given the unknown anthropogenic contribution, we leave this SOA isoprene aerosol source without a clear biogenic labeling.

During this study, C₅-alkene triols concentrations (about 1 ng m⁻³) were slightly lower than those reported in other European sites (Kourtchev et al., 2005) and much lower than the levels of 300 ng m⁻³ in sites located in the southeastern USA (Lin et al., 2012). The 2-methylglyceric acid (2-MGA) concentrations were in the same range as those observed in Marseille (about 2 ng m⁻³; El Haddad et al., 2011), whereas the levels of 2-methyltetrols (2-MT) were higher in the present study (about 2 ng m⁻³). 2-methylerythritol concentrations were 2.5 times higher than 2-methylthreitol concentrations. They were highly correlated within each other at both sites ($r^2 = 0.9$; $P < 0.05$), as previously observed in other studies (Edney et al., 2005; Ion et al., 2005; El Haddad et al., 2011). Overall there were good correlations between C₅-alkene triols and \sum 2-methyltetrols ($r^2 = 0.8$; $P < 0.05$), as well as between 2-MGA and C₅-alkene triols, and \sum 2-methyltetrols ($r^2 = 0.6$ and 0.7, respectively; $P < 0.05$). Overall, there was no day/night or weekdays/weekend significant difference at both monitoring sites for all the SOA isoprene markers. The concentrations of 2-MGA at UB were not significantly different than those observed at RS, whereas the 2-MT concentrations were about 2 times higher at UB.

Although both 2-MT and 2-MGA originate from the photo-oxidation of isoprene, NO_x concentrations and aerosol

acidity (Surratt et al., 2010) may affect their formation processes and consequently their relative abundances. Moreover, relative abundances of these organic compounds can evolve as they can undergo further photo-chemical processing or heterogeneous reactions upon their formation (Robinson et al., 2007). 2-MT are formed by photo-oxidation of isoprene under NO-limited conditions (Lin et al., 2012), while it is known that 2-MGA can also be formed by the further oxidation of volatile methacryloylperoxynitrate (MPAN), a major second-generation product of isoprene oxidation under initially high-NO conditions (Surratt et al., 2010). Recent work by Lin et al. (2013) found that 2-MGA is directly formed from methacrylic acid epoxide (MAE), which is a gas-phase oxidation product of MPAN. In summary, it is likely that the different traffic conditions encountered at the two monitoring sites, with higher NO at RS than UB, can impact the concentrations of 2-MGA and 2-MT.

3.2 Multivariate data analysis

Principal component analysis (PCA) was applied on a matrix of scaled concentrations to explore the amount of variance that could be explained by a reduced number of components and to identify the most determinant associations amongst the organic tracer compounds analyzed in this study. A six-component solution was selected from the analysis involving a 96.5 % of explained variance. Addition of a seventh component (97.2 % of explained variance) did not show any further relevant environmental information.

Following the optimal number of components obtained by PCA, MCR-ALS was applied to the scaled column-wise data matrix (i.e. without mean-centering) and non-negativity constraints, resulting in six components which accounted for 96.2 % of the total data variance.

Although many compounds were distributed between different MCR-ALS components for many of the chemical compounds, an insight of the contributions of the different OA sources can be seen in Fig. 2. The average organic carbon concentrations (OC, determined separately by the Sunset EC/OC instrument; Dall'Osto et al., 2013b) were found to be $3.5 \mu\text{g m}^{-3}$ at the RS and $2.1 \mu\text{g m}^{-3}$ at the UB site (Dall'Osto et al., 2013b). There was a very good correlation between the total OC concentrations of PM_{10} and \sum scores of component for each sampled day ($r^2 = 0.9$, $P < 0.05$), although between 2 and 6 % of the total OC mass was chromatographically resolved. The excellent correlation indicates that the selected organic components described in this study can be representative for the contributions to the organic fraction sampled during the SAPUSS project. The percentages of OA contribution of each of the six components obtained were calculated from the relative contribution of the individual MCR-ALS resolved component with respect to the \sum score of the six components. The mass contribution of OC per volume for each component (Fig. 2) was calculated by multiplying the relative contributions with the OC concentrations from

Dall'Osto et al. (2013b). The following sections discuss the unique features of the six OA sources identified, as well as the temporal trend of the scores. The percentages presented in brackets after the molecular markers represent the proportion of such marker among all six MCR-ALS profiles, and not the percentage of such marker among the total signal of the individual MCR-ALS profile.

3.2.1 Urban Primary Organic Aerosol (POA Urban)

This component is defined by strong contributions from PAH (43 %), hopanes (43 %) and nicotine (79 %) and accounts for 32 % of the explained variance. The relatively larger contribution to the component of LMW PAH in combination with the hopanes points to diesel vehicles as an important emission source. Among the six components this is the one showing the best correlation with traffic intensity ($r^2 = 0.5$), pointing to traffic as a major contributor to this component. This statement is further supported by the good correlation between this component and other traffic indicators, such as NO_x ($r^2 = 0.8$), $\text{EC}_{\text{PM}_{10}}$ ($r^2 = 0.7$) and preliminary AMS-PMF data on the hydrocarbon-like organic carbon (HOA; $r^2 = 0.8$) that were analyzed for the sampling sites.

The temporal changes of this source between the two sites were correlated ($r^2 = 0.5$, $P < 0.05$). POA Urban accounted for 0.35 and $1.5 \mu\text{g m}^{-3}$ OC at UB (18 % of total OC) and RS (43 % of OC), respectively. This observation is consistent with a previous study (Mohr et al., 2012) in which 16 % of the OA was attributed to primary organic aerosol at UB. In the urban area of Marseille along the Mediterranean coast a similar concentration of about $0.9 \mu\text{g OC m}^{-3}$ (17 %) was attributed to primary urban emissions (El Hadded et al., 2011) based on organic tracer analysis, while the primary emissions in Paris ranged between 0.2 and $0.5 \mu\text{g OC m}^{-3}$, depending on the used method.

POA Urban was higher during weekdays, especially on Wednesday and Thursday, whilst the lowest values were recorded during weekends and holidays (Fig. 2). There was no clear day–night fluctuation for this component, which is in accordance with the absence of such fluctuation in the LMW PAH and hopanes. The scores of this component are nearly equally divided among the four main air mass scenarios (Fig. 3), suggesting that it is less dependent on regional air mass transport and that local emissions are mainly responsible for the presence of such primary atmospheric pollutants in the urban area.

3.2.2 Primary regional biomass burning organic aerosol (BBOA Regional)

This component is essentially formed by levoglucosan (68 %), mannosan (75 %) and galactosan (79 %), several markers of primary biomass burning. The component also contains an important part of glyceric acid (45 %), which could be emitted after biomass burning or formed during

10362

M. Alier et al.: Source apportionment of submicron organic aerosol

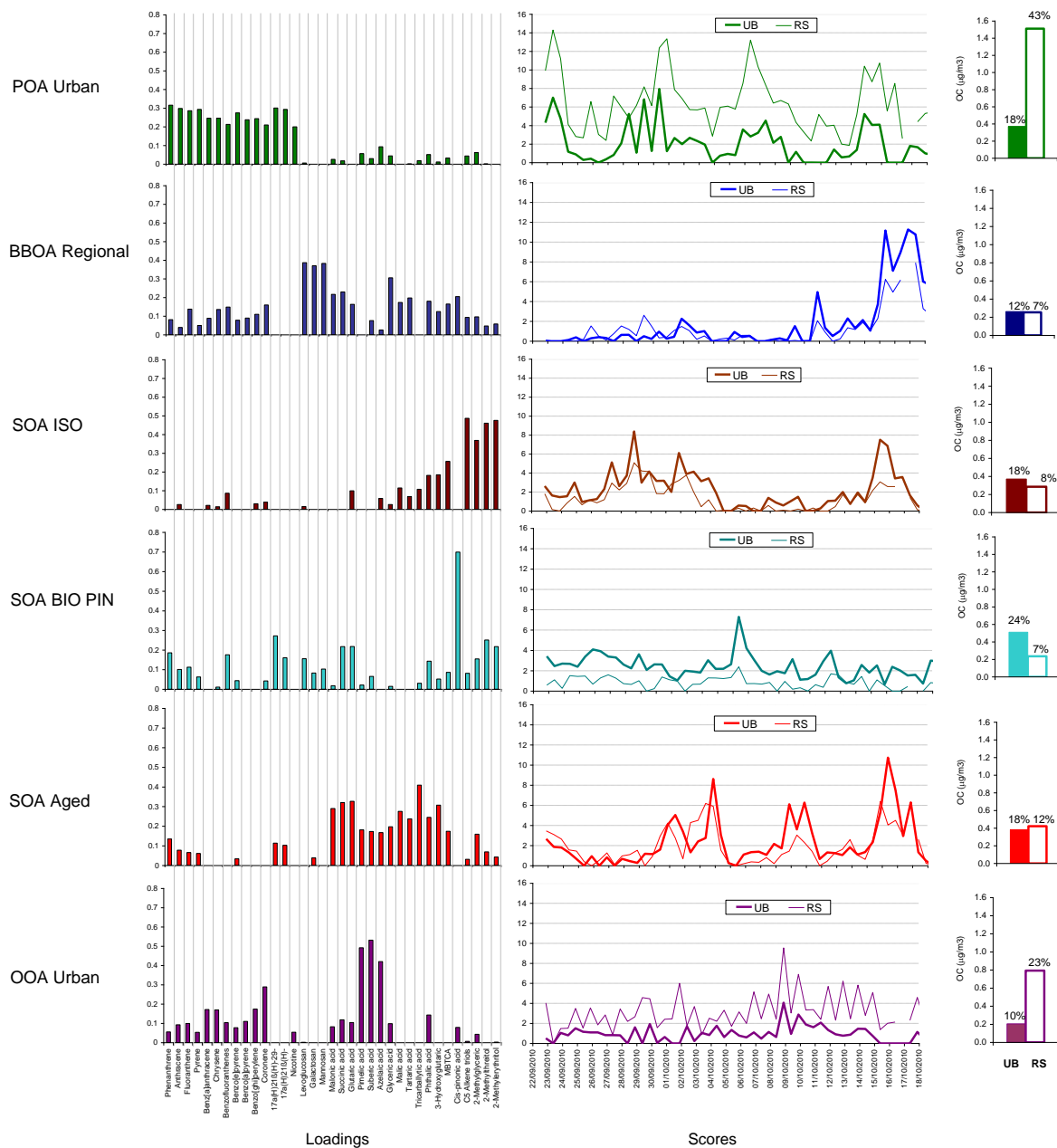


Fig. 2. MCR-ALS resolved profiles applying non-negativity constraints. Column on the left shows the loading organic compound composition of the six selected components. Column in the middle shows the temporal trend of the score values of the six components. Column on the right shows the relative contribution of the scores (%) as well as the estimated mass contribution of organic carbon to volume ($\mu\text{g OC m}^{-3}$).

atmospheric transport. Moderate contributions (about 20 % of the total signal) from other DCAs and PAH are also observed (Fig. 2).

This component accounted for 15 % of the total variance and its values in the two sites were strongly correlated ($r^2 = 0.85$). The highest score values in both sites were observed in the last sampling period of the SAPPUSS campaign (15–

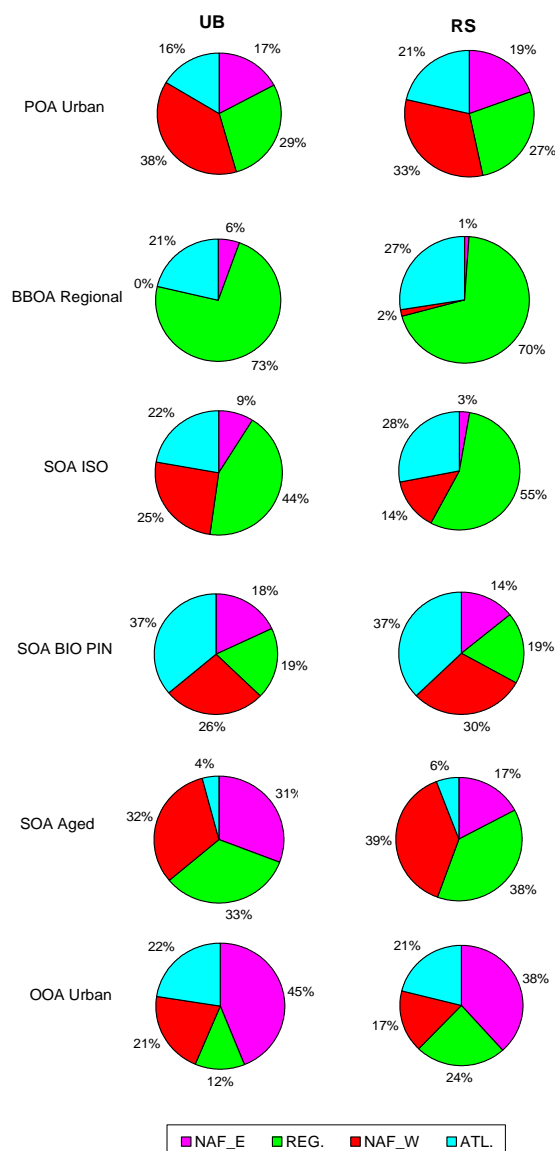


Fig. 3. Relative scores of the components in UB and RS in relation to the different air mass scenarios; NAF_W: North African air mass from the west; Atl.: Atlantic advection; REG.: regional air circulation; NAF_E: North African from the east over the Mediterranean Sea (Dall'Osto et al., 2012a).

17 October, Fig. 2), coinciding with regional recirculation air masses (Fig. 3, 70 %) as well as with the permission to burn litter from fields surrounding the urban area of Barcelona (DOGC, 1995).

Although the relative contribution of BBOA was higher at UB (12 %) than at RS (7 %), both monitoring sites presented roughly similar concentrations (about $0.3 \mu\text{g m}^{-3}$).

Mohr et al. (2012) also found a moderate contribution of 11 % for BBOA in winter 2009 by AMS analysis, which was correlated with off-line filter analysis of levoglucosan (Van Drooge et al., 2012a). The contribution of BBOA was very similar to the $0.4 \mu\text{g m}^{-3}$ estimated for the UB site after year-round measurements in 2009 (Reche et al., 2012). These contributions are in the lower range of those estimated for other European sites ($0.04\text{--}3.1 \mu\text{g OC m}^{-3}$; Puxbaum et al., 2007), where the higher concentrations are observed in sites that are directly exposed to local biomass burning. It is important to remember that organic tracers for biomass burning may be susceptible to oxidation (Hennigan et al., 2011; Hoffmann et al., 2010; Kessler et al., 2010). However, the analysis of inorganic species during the SAPUSS study (Dall'Osto et al., 2013a) also found a similar apportionment at the two monitoring sites, although slightly higher at the UB site.

3.2.3 Isoprene secondary organic aerosol (SOA ISO)

C_5 alkene triols (65 %), 2-MGA (42 %) and the 2-MTs (60 %) represent the SOA tracers of isoprene oxidation, and they are grouped in one component that accounts for 11 % of the explained variance. Based on the high NO levels in the city center (Dall'Osto et al., 2013b), one would expect larger differences between the two monitoring sites as seen for some of the isoprene markers (see Sect. 3.1.7). Nevertheless, the overall contributions of SOA ISO were similar at both sites (0.4 and $0.3 \mu\text{g m}^{-3}$ at UB and RS, respectively), suggesting mainly a regional origin. Indeed, about 50 % (Fig. 3) of the concentrations detected were recorded during regional air masses, while considerably lower contributions were observed when the air mass came from the Mediterranean Sea. The regional apportionment of the SOA ISO is also reflected in the temporal trend of this component (Fig. 2), which was found to be very similar between the two sites ($r^2 = 0.7$, $P < 0.05$).

3.2.4 Biogenic α -pinene secondary organic aerosol (SOA BIO PIN)

Another component is dominated by a strong signal associated with cis-pinonic acid (72 %), a first-generation product of α -pinene oxidation and minor contributions from 2-methyltetrols (24 %), succinic acid and glutaric acid (22 %). Hopanes were also found to contribute to this aerosol source (35 % of the total hopanes signal) and could be related to biogenic hopanoid precursors produced by microbiota (Yan et al., 2008). Other second-generation products of α -pinene oxidation (i.e. 3-hydroxyglutaric acid and MBTCA) did not contribute to this component, implying it is relatively fresh (see Sect. 3.1.6). The SOA BIO PIN component accounted for 10 % of the explained variance, with average concentrations estimated to be $0.5 \mu\text{g m}^{-3}$ at UB (24 % of OA) and $0.2 \mu\text{g m}^{-3}$ at RS (7 %), respectively. The temporal trends at both sites were poorly correlated ($r^2 = 0.3$, $P < 0.05$),

10364

M. Alier et al.: Source apportionment of submicron organic aerosol

suggesting this SOA component did not impact the two monitoring site at the same level. This could be due to different reasons, including the fact that α -pinene SOA probably undergoes further oxidation in the city center. This is in contrast to the previous SOA described (SOA ISO) where no difference was found between the two monitoring sites.

The present study shows a remarkable temporal trend (Fig. 2): the highest SOA BIO PIN scores at both monitoring sites were recorded on 5 October 2010, when a strong urban new particle formation (NPF) event was detected (Dall'Osto et al., 2013c). Furthermore, high levels can also be seen on 25–27 September when other minor NPF events were detected. Cis-pinonic acid has been linked to nucleation processes in forested areas as a first step in the formation of aerosols from organic vapors (O'Dowd et al., 2002; Laaksonen et al., 2008). Further considerations on the matter are discussed later in Sect. 4.

3.2.5 Aged secondary organic aerosol (SOA Aged)

SOA tracer compounds of both anthropogenic and biogenic origin, including C₃–C₅ DCA (40%), C₇–C₉ DCA (20%), phthalic acid (26%), hydroxy-DCA (45%) and tricarballic acid (75%), were defining this component. It is interesting to note the absence of cis-pinonic signal, suggesting that this SOA component is more aged than the SOA BIO PIN. Indeed, 3-hydroxyglutaric acid and MBTCA can be seen in this SOA Aged profile (Fig. 2) as products from further photochemical reactions of cis-pinonic acid (Szmigielski et al., 2007). The higher contribution of C₃–C₅ DCA over C₇–C₉ DCA also indicates the more aged character of this component, since the higher-carbon-numbered DCA tend to photochemically degrade to smaller DCA (Yang et al., 2008b). Further support from the high degree of “aging” of this component also comes from the trend seen for the isoprene SOA markers, with the 2-MGA score being the strongest among all isoprene markers. In other words, when comparing this SOA Aged component with the SOA ISO one (Fig. 2), whilst 2-MGA is found to be the least important tracer among the four main ones describing this component, the opposite trend is seen for SOA Aged. Interestingly, the SOA Aged component shows the best correlation ($r^2 = 0.6$) with the OOA component from preliminary AMS-PMF analysis among all other components.

This component accounted for 18% of the explained variance and was related to SOA tracers. Similar concentrations ($0.4 \mu\text{g m}^{-3}$ OC) were found at both sites (18% of OA at UB; 12% at RS). There was a good correlation between the scores of the two sampling sites ($r^2 = 0.5$), suggesting this SOA component is impacting the two sites at a similar level. The temporal trend of this component (Fig. 2) was air mass dependent, with lowest values observed during windy North Atlantic advections (Fig. 3).

3.2.6 Oxidized organic aerosol of mainly urban origin (OOA Urban)

Contrary to the previously described five OA sources, it is not known at this stage if this OOA Urban source has a primary and/or a secondary component; hence the name does not include any specification in this regard. The unique feature of this aerosol source is the strong association with pimelic (65%), suberic (61%) and azelaic (55%) acids (C₇–C₉ DCA). It is also interesting to note the presence of PAHs (21% of total PAH signal), with particular evidence for coronene (39%, strongest signal of all six sources) and minor ones from benzo[ghi]perylene (32%) and benz[a]anthracene (33%). This source has also the peculiarity of containing the remaining 20% of the nicotine signal, which in this study is mainly apportioned by POA Urban (80%; see Sect. 3.2.1 and Fig. 2). OOA Urban was found to have an average concentration of $0.8 \mu\text{g m}^{-3}$ (23%) at the RS site and of $0.2 \mu\text{g m}^{-3}$ (10%) at the UB site. A temporal correlation of this OA source was not found between the two monitoring sites. The consistently higher abundance of this component at the city center site (RS) with respect to the urban background site (UB), as well as a daytime maximum (Fig. 2), suggest that local urban emissions sources were involved. The strong PAH_{high} component (benzo[ghi]pyrene and coronene) suggests that motor vehicle emissions, previously associated mainly with primary gasoline-powered vehicles, could be involved (Rogge et al., 1993; Zielinska et al., 2004). On the other hand, C₇–C₉ DCAs are often associated with oxidation of unsaturated fatty acids (e.g. oleic acid; Moise et al., 2002). A major source for primary emissions of oleic acid in an urban environment is food cooking (Schauer et al., 2002; Robinson et al., 2006b), so this source could also be related to the one reported by Mohr et al. (2012), attributing 17% of the atmospheric organic PM₁ due to primary food cooking emissions in the same location. The OOA Urban component was found to correlate only moderately with preliminary PMF results from the HR-ToF-AMS analysis. A weak correlation with cooking organic aerosol (COA, $r^2 = 0.45$) was found only at the RS but not at the UB site, suggesting that OOA Urban is a mixture of urban sources.

Nevertheless, a complex air mass dependency of this source at both monitoring sites suggests that this component is also related to atmospheric processes. Figure 3 shows that at both sites the higher scores of OOA Urban were observed under North African air masses from the Mediterranean Sea (NAF_East, with maximum at both monitoring sites between 7 and 10 October, Fig. 2). During this period, the concentrations at the two monitoring sites were highly correlated ($r^2 = 0.9$) and the increase of the OOA Urban coincided with the increase of SOA Aged. However, in two other REG. periods of elevated SOA Aged concentrations (29 September–3 October and 14–17 October) the OOA Urban component was found at relatively low concentrations and not correlated between the two monitoring sites. Whilst during NAF_East air

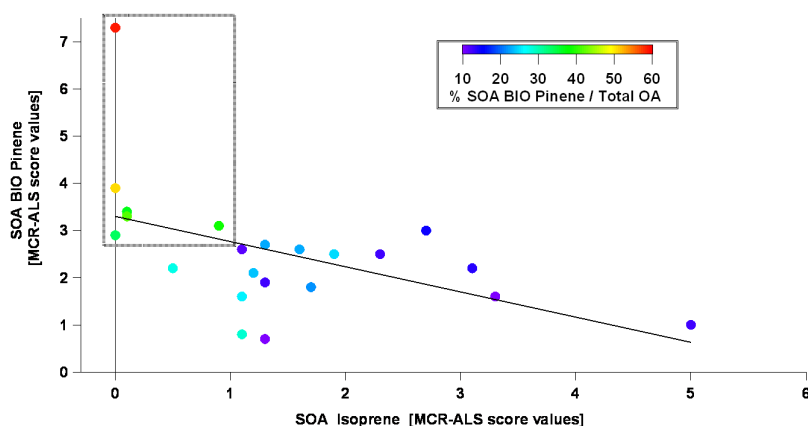


Fig. 4. MCR-ALS scores values for SOA isoprene and SOA BIO pinene for daytime samples (09:00–21:00) collected during SAPUSS. Please note the rectangle in figure includes the five nucleation days seen during SAPUSS (25, 26, 27 September, 5 and 18 October 2010).

mass type the nitrate concentrations were at minimum, the opposite scenarios was found for the REG. air mass types (Dall’Osto et al., 2013a). Yang et al. (2008b) previously reported that the nitrate ions play a major role in accelerating the photo-oxidation of azelaic acid by increasing OH concentration. This could be a possible explanation of why OOA Urban can accumulate during “sulphate high–nitrate low” air masses (NAF_East) whereas much lower concentrations are found during “sulphate low–nitrate high” air masses (REG.) at both sites.

In summary, it seems that certain atmospheric conditions are responsible for the abundance of OOA Urban OA source. This aerosol component is the least characterized of the six ones apportioned in this study. It points to compounds related to gasoline vehicle emissions and cooking activities, and the different behaviors under different air masses suggest its abundance depend on the atmospheric conditions encountered during its emission.

4 Implications and conclusions

The application of MCR-ALS on a database generated from off-line GC-MS analysis of 36 organic tracer compounds in 103 PM₁ filter samples collected simultaneously in two sites from Barcelona during 30 successive days allowed the identification of six organic aerosol components in the two sites that explained 96 % of the total database variance. The specific study of these components provides information on the environmental sources and processes responsible for the organic aerosol constituents.

Two of the aerosol components (POA Urban and BBOA REG.) could be linked directly to primary anthropogenic emission sources. At the RS site, the primary emissions from anthropogenic sources explain 43 % of the variance of the

OA, whereas only 28 % at the UB site. Overall, the regional BBOA component had little influence on the air quality in the city.

The sum of the three SOA components (SOA Aged, SOA ISO and SOA BIO PIN) contributed 60 % of the OA at the UB, in line with the 55 % calculated by Mohr et al. (2012) in the same location during winter 2009. The relatively high contribution of SOA in the urban background of Barcelona is related to the high frequency of anticyclonic weather conditions, allowing aged aerosols to accumulate. At the RS, the SOA component was found to be only 27 %. Two of them were represented by products of isoprene and α -pinene oxidation, while the third was composted by a mixture of more oxidized compounds from anthropogenic and biogenic origin. Most of the SOA formed from biogenic sources are believed to be due to isoprene and monoterpene emissions because of their high emission rates from various vegetated surfaces. Biogenic contributions to SOA were previously expected to be relatively low in urban areas. However, the complex emissions in urban settings have been shown to accelerate oxidation of biogenic VOCs and thus production of biogenic SOA (Goldstein and Galbally, 2007; Liao et al., 2007). Our study suggests that the overall biogenic SOA component at the UB and RS was at least 42 % and 15 %, respectively, with a ISO/PIN ratio of about 3.

An unexpected sticking result of this work is the relationship between SOA ISO, SOA BIO PIN and the NPF events during the SAPUSS project. Briefly, a clustering analysis of aerosol size distributions data collected at a UB site in Barcelona revealed that ultrafine particles often fail to grow above 10 nm (Dall’Osto et al., 2012). However, NPF originating within the city center can grow to larger sizes while being transported away from the city to the urban–regional background (Dall’Osto et al., 2013c). Previous laboratory results showed that the high reactivity of isoprene

10366

M. Alier et al.: Source apportionment of submicron organic aerosol

with OH leads to suppression of new particle formation (Kiehnler-Scharr et al., 2009). Large isoprene emissions can also suppress NPF formation in forests although the underlying mechanism for the suppression is unclear (Kanawade et al., 2011). Although the role of biogenic VOC in limiting oxidant levels is established, the influence of OH remains uncertain (Taraborrelli et al., 2012).

Figure 4 summarizes a key finding of our study during the SAPUSS project. A clear anti correlation between SOA ISO and SOA BIO PIN can be seen for the plotted PM₁ daytime samples. Additionally, the five days characterized by NPF events are all distributed in the top left part of the panel. In other words, our results suggest that large isoprene emissions can suppress NPF formation in urban areas. Furthermore, the very high percentages of SOA BIO PIN in the PM₁ mass detected during the NPF events suggest that biogenic compounds could be partially responsible for the growth of ultrafine particles in the urban area of Barcelona. Such results go in the same direction of recent findings in the same area showing that the majority of OA is of non-fossil origin (Minguillón et al., 2011).

The third SOA component (SOA Aged) contains a mixture of both biogenic and anthropogenic SOA tracers. Therefore, it was not possible to estimate the contribution of non-fossil and fossil organic carbon in this component. By considering BBOA REG., SOA BIO PIN of non-fossil origin, this was about 37 % at the UB ($0.8 \mu\text{g m}^{-3}$) and 15 % at the RS ($0.5 \mu\text{g m}^{-3}$). It is likely that SOA ISO is mainly biogenic during the studied period although a possible influence of an anthropogenic contribution in the urban sites could not be excluded. Nevertheless, adding SOA ISO to the non-fossil OA would lead to contributions of 54 % and 22 % at UB and RS, respectively. The relatively large contribution of non-fossil SOA in this study is in agreement with the isotopic carbon data from former studies in the same urban area (Minguillón et al., 2011).

What remains unclear in the present study is the exact meaning of the OOA Urban component. It is composed mainly of C₇–C₉ dicarboxylic acids, with contributions of PAH_{high} compounds. Our results suggest that the presence of this aerosol source of anthropogenic origin depends on a number of unknown conditions. When analyzing the MCR-ALS seven-component solution (see Supplement), two new components appear. POA Urban 1 (PAH_{high} and nicotine) and POA Urban 2 (PAH_{low} and hopanes) somehow take away all the primary OA component from OOA Urban, leaving it with only the C₇–C₉ signature (Fig. S1). Unfortunately, the solution with seven components worsens all other external correlations with all the other existing factors, and therefore it cannot be chosen as the correct one. Nevertheless, this suggests that both POA Urban and OOA Urban somehow hide some unknown OA sources or processes not distinguishable in this study. The 12 h sampling time resolution used in this study does not allow us to fully apportion OOA Urban to traffic or cooking emissions. Although the linkage of this

component to traffic cannot be excluded, the chemical composition formed by the C₇–C₉ dicarboxylic acids suggests that cooking activities could play an important role in this component. Nevertheless, the component is represented only by the less volatile products of the oxidation of unsaturated fatty acids (Moise and Rudich, 2002). The air mass dependence of the component indicates that its abundance cannot be directly linked to any emission source, but could be related to oxidation processes in the urban center.

This study reveals that 43–66 % and 18–28 % of the OA detected at the RS and UB, respectively, have an urban origin, leaving a substantial contribution of biogenic components in the studied urban area. However, the urban atmosphere seems to be responsible for the transformation of OA (Donahue et al., 2009), creating difficulties in fully assigning OA sources with the traditional definitions of primary and secondary organic aerosol. In future publications of this ACP special issue, data from collocated and independent SAPUSS measurements will be compared with the apportionment of the OA sources obtained during this study (GC-MS/MCR-ALS), in order to increase the insights of the potential sources and transformation processes in the urban Mediterranean area of Barcelona.

Acknowledgements. Technical assistance from R. Chalier and D. Fanjul is acknowledged. Financial support for this study was provided by the Marie Curie FP7 SAPUSS (FP7-PEOPLE-2009-IEF, Project number 254773), and previously supported by research projects from the D. G. de Calidad y Evaluación Ambiental (Spanish Ministry of the Environment) and the Plan Nacional de I+D (Spanish Ministry of Science and Innovation) CGL2010-19464-VAMOS, CTQ2009-11572 and CTQ2009-377-14777-C02-01-AERTRANS). The SAPUSS team is acknowledged.

Edited by: A. S. H. Prevot

References

- Allan, J. D., Williams, P. I., Morgan, W. T., Martin, C. L., Flynn, M. J., Lee, J., Nemitz, E., Phillips, G. J., Gallagher, M. W., and Coe, H.: Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two UK cities, *Atmos. Chem. Phys.*, 10, 647–668, doi:10.5194/acp-10-647-2010, 2010.
- Bahreini, R., Middlebrook, A. M., De Gouw, J. A., Warneke, C., Trainer, M., Brock, C. A., Stark, H., Brown, S. S., Dube, W. P., Gilman, J. B., Hall, K., Holloway, J. S., Kuster, W. C., Perring, A. E., Prevot, A. S. H., Schwarz, J. P., Spackman, J. R., Sztand, S., Wagner, N. L., Weber, R. J., Zotter, P., and Parrish, D. D.: Gasoline emissions dominate over diesel in formation of secondary organic aerosol mass, *Geophys. Res. Lett.*, 39, L06805, doi:10.1029/2011GL050718, 2012.
- Bi, X., Sheng, G., Feng, Y., Fu, J., and Xie, J.: Gas- and particulate-phase specific tracer and toxic organic compounds in environmental tobacco smoke, *Chemosphere*, 61, 1512–1522, 2005.

- Bi, X., Simoneit, B. R. T., Sheng, G., Ma, S., and Fu, J.: Composition and major sources of organic compounds in urban aerosols, *Atmos. Res.*, 88, 256–265, 2008.
- Borbon, A., Fontaine, H., Veillerot, M., Locoge, N., Galloo, J. C., and Guillermo, R.: An investigation into traffic-related fraction of isoprene at an urban location, *Atmos. Environ.*, 35, 3749–3760, 2001.
- Brunekreef, B.: Forsberg, B.: Epidemiological evidence of effects of coarse airborne particles on health, *Eur. Respir. J.*, 26, 309–318, 2005.
- Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, 2007.
- Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., Cafmeyer, J., Guyon, P., Andreae, M. O., Artaxo, P., and Maenhaut, W.: Formation of secondary organic aerosols through photooxidation of isoprene, *Science*, 303, 1173–1176, 2004.
- Claeys, M., Szmigielski, R., Kourtchev, I., Van der Veken, P., Vermeylen, R., Maenhaut, W., Jaoui, M., Kleindienst, T., Lewandowski, M., Offenberg, J., and Edney, E.: Hydroxydicarboxylic Acids: Markers for Secondary Organic Aerosol from the Photooxidation of α -Pinene, *Environ. Sci. Technol.*, 41, 1628–1634, 2007.
- Clements, A. L. and Seinfeld, J. H.: Detection and quantification of 2-methyltetrols in ambient aerosol in the southeastern United States, *Atmos. Environ.*, 41, 1825–1830, 2007.
- Dall’Osto, M., Beddows, D. C. S., Pey, J., Rodriguez, S., Alastuey, A., Harrison, R. M., and Querol, X.: Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain, *Atmos. Chem. Phys.*, 12, 10693–10707, doi:10.5194/acp-12-10693-2012, 2012.
- Dall’Osto, M., Querol, X., Amato, F., Karanasiou, A., Lucarelli, F., Nava, S., Calzolari, G., and Chiari, M.: Hourly elemental concentrations in PM_{2.5} aerosols sampled simultaneously at urban background and road site during SAPUSS – diurnal variations and PMF receptor modelling, *Atmos. Chem. Phys.*, 13, 4375–4392, doi:10.5194/acp-13-4375-2013, 2013a.
- Dall’Osto, M., Querol, X., Alastuey, A., Minguillon, M. C., Alier, M., Amato, F., Brines, M., Cusack, M., Grimalt, J. O., Karanasiou, A., Moreno, T., Pandolfi, M., Pey, J., Reche, C., Ripoll, A., Tauler, R., Van Drooge, B. L., Viana, M., Harrison, R. M., Gietl, J., Beddows, D., Bloss, W., O’Dowd, C., Ceburnis, D., Martucci, G., Ng, N. L., Worsnop, D., Wenger, J., Mc Gillicuddy, E., Sodeau, J., Healy, R., Lucarelli, F., Nava, S., Jimenez, J. L., Gomez Moreno, F., Artinano, B., Prévôt, A. S. H., Pfaffenberger, L., Frey, S., Wilsenack, F., Casabona, D., Jiménez-Guerrero, P., Gross, D., and Cots, N.: Presenting SAPUSS: Solving Aerosol Problem by Using Synergistic Strategies in Barcelona, Spain, *Atmos. Chem. Phys.*, 13, 8991–9019, doi:10.5194/acp-13-8991-2013, 2013b.
- Dall’Osto, M., Querol, X., Alastuey, A., O’Dowd, C., Harrison, R. M., Wenger, J., and Gómez-Moreno, F. J.: On the spatial distribution and evolution of ultrafine particles in Barcelona, *Atmos. Chem. Phys.*, 13, 741–759, doi:10.5194/acp-13-741-2013, 2013c.
- Ding, X., Wang, X.-M., and Zheng, M.: The influence of temperature and aerosol acidity on biogenic secondary organic aerosol tracers: observations at a rural site in the central Pearl River Delta region, South China, *Atmos. Environ.*, 45, 1303–1311, 2011.
- DOG 1995: Forest fire prevention, Catalonia, Ordinance 64/1995.
- Donahue, N. M., Robinson, A. L., and Pandis, S. N.: Atmospheric organic particulate matter: From smoke to secondary organic aerosol, *Atmos. Environ.*, 43, 94–106, 2009.
- Edney, E. O., Kleindienst, T. E., Jaoui, M., Lewandowski, M., Offenberg, J. H., Wang, W., and Cleays, M.: Formation of 2-methyltetrols and 2-methylglyceric acid in secondary aerosol from laboratory irradiated isoprene/NO_x/SO₂/air mixture and their detection in ambient air PM_{2.5} samples collected in the eastern United States, *Atmos. Environ.*, 39, 5281–5289, 2005.
- El Haddad, I., Marchand, N., Temime-Roussel, B., Wortham, H., Piot, C., Besombes, J.-L., Baduel, C., Voisin, D., Armengaud, A., and Jaffrezo, J.-L.: Insights into the secondary fraction of the organic aerosol in a Mediterranean urban area: Marseille, *Atmos. Chem. Phys.*, 11, 2059–2079, doi:10.5194/acp-11-2059-2011, 2011.
- Fang, J., Kawamura, K., Ishimura, Y., and Matsumoto, K.: Carbon isotope composition of fatty acids in the marine aerosols from the Western North Pacific: Implication for the source and atmospheric transport, *Environ. Sci. Technol.*, 62, 2598–2604, 2002.
- Filella, I. and Penuelas, J.: Daily, weekly, and seasonal time courses of VOC concentrations in a semi-urban area near Barcelona, *Atmos. Environ.*, 40, 7752–7769, 2006.
- Fine, P. M., Cass, G. R., and Simoneit, B. R. T.: Chemical characterization of fine particle emissions from the fireplace combustion of wood types grown in the Midwestern and Western United States, *Environ. Eng. Sci.*, 21, 387–409, 2004.
- Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in Atmospheric Constituents and in Radiative Forcing. In: IPCC, Climate Change 2007: The Physical Science Base. Contribution of Working Group I to the fourth Assessment Report of the IPCC, ISBN: 978 0521 88009-1, 2007.
- Galarneau, E. Source specificity and atmospheric processing of airborne PAHs: Implications for source apportionment, *Atmos. Environ.*, 42, 8139–8149, 2008.
- Gogou, A., Stephanou, E. G., Stratigakis, N., Grimalt, J. O., Simo, R., Aceves, M., and Albaiges, J.: Differences in lipid and organic salt constituents of aerosols from Eastern and Western Mediterranean coastal cities, *Atmos. Environ.*, 28, 1301–1310, 1994.
- Goldstein, A. H. and Galbally, I. E.: Known and unexplored organic constituents in the earth’s atmosphere, *Environ. Sci. Technol.*, 41, 1514–1521, 2007.
- Grieshop, A. P., Donahue, N. M., and Robinson, A. L.: Laboratory investigation of photochemical oxidation of organic aerosol from wood fires 2: analysis of aerosol mass spectrometer data, *Atmos. Chem. Phys.*, 9, 2227–2240, doi:10.5194/acp-9-2227-2009, 2009.
- Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W. A. Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A Global-Model of Natural Volatile Organic-Compound Emissions, *J. Geophys. Res.*, 100, 8873–8892, 1995.

10368

M. Alier et al.: Source apportionment of submicron organic aerosol

- Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, Th. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, *Atmos. Chem. Phys.*, 9, 5155–5236, doi:10.5194/acp-9-5155-2009, 2009.
- Heald, C. L., Kroll, J. H., Jimenez, J. L., Docherty, K. S., DeCarlo, P. F., Aiken, A. C., Chen, Q., Martin, S. T., Farmer, D. K., Artaxo, P., and Weinheimer, A. J.: A simplified description of organic aerosol composition and implications for atmospheric aging, *Geophys. Res. Lett.*, 37, 8, L08803, doi:10.1029/2010GL042737, 2010.
- Hennigan, C. J., Miracolo, M. A., Engelhart, G. J., May, A. A., Presto, A. A., Lee, T., Sullivan, A. P., McMeeking, G. R., Coe, H., Wold, C. E., Hao, W.-M., Gilman, J. B., Kuster, W. C., de Gouw, J., Schichtel, B. A., J. L. Collett Jr., Kreidenweis, S. M., and Robinson, A. L.: Chemical and physical transformations of organic aerosol from the photo-oxidation of open biomass burning emissions in an environmental chamber, *Atmos. Chem. Phys.*, 11, 7669–7686, doi:10.5194/acp-11-7669-2011, 2011.
- Ho, K. F., Ho, S. S. H., Lee, S. C., Kawamura, K., Zou, S. C., Cao, J. J., and Xu, H. M.: Summer and winter variations of dicarboxylic acids, fatty acids and benzoic acid in PM_{2.5} in Pearl Delta River Region, China, *Atmos. Chem. Phys.*, 11, 2197–2208, doi:10.5194/acp-11-2197-2011, 2011.
- Hoffmann, D., Tilgner, A., Iinuma, Y., and Herrmann, H.: Atmospheric stability of levoglucosan: A detailed laboratory and modeling study, *Environ. Sci. Technol.*, 44, 694–699, 2010.
- Ion, A. C., Vermeylen, R., Kourchev, I., Cafmeyer, J., Chi, X., Gelecsér, A., Maenhaut, W., and Claeys, M.: Polar organic compounds in rural PM_{2.5} aerosols from K-pusztá, Hungary, during a 2003 summer field campaign: Sources and diel variations, *Atmos. Chem. Phys.*, 5, 1805–1814, doi:10.5194/acp-5-1805-2005, 2005.
- Jang, M. and McDow, S. R.: Products of benz[*a*]anthracene photodegradation in the presence of known organic constituents of atmospheric aerosols, *Environ. Sci. Technol.*, 31, 1046–1053, 1997.
- Jaumot, J., Gargallo, R., de Juan, A., and Tauler, R.: A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometr. Intell. Lab. Syst.*, 76, 101–110, 2005.
- Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, *Science*, 326, 1525–1529, 2009.
- Jolliffe, I. T.: *Principal Component Analysis*, 2nd ed., Springer-Verlag, New York, USA, 2002.
- Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, 5, 1053–1123, doi:10.5194/acp-5-1053-2005, 2005.
- Kanawade, V. P., Jobson, B. T., Guenther, A. B., Erupe, M. E., Pressley, S. N., Tripathi, S. N., and Lee, S.-H.: Isoprene suppression of new particle formation in a mixed deciduous forest, *Atmos. Chem. Phys.*, 11, 6013–6027, doi:10.5194/acp-11-6013-2011, 2011.
- Kawamura, K. and Gagosian, R. B.: Implications of ω -Oxocarboxylic acids in the remote marine atmosphere for photo-oxidation of unsaturated fatty acids, *Nature*, 325, 330–332, 1987.
- Kerminen, V. M., Virkkula, A., Hillamo, R., Wexler, A. S., and Kulmala, M.: Secondary organics and atmospheric cloud condensation nuclei production, *J. Geophys. Res.*, 105, 9255–9264, 2000.
- Kessler, S. H., Smith, J. D., Che, D. L., Worsnop, D. R., Wilson, K. R., and Kroll, J. H.: Chemical sinks of organic aerosol: Kinetics and products of the heterogeneous oxidation of erythritol and levoglucosan, *Environ. Sci. Technol.*, 44, 7005–7010, 2010.
- Kiendler-Scharr, A., Wildt, J., Dal Maso, M., Hohaus, T., Kleist, E., Mentel, T. F., Tillmann, R., Uerlings, R., Schurr, U., and Wahner, A.: New particle formation in forests inhibited by isoprene emissions, *Nature*, 461, 381–384, 2009.
- Kourchev, I., Ruuskanen, T., Maenhaut, W., Kulmala, M., and Claeys, M.: Observation of 2-methyltetrols and related photo-oxidation products of isoprene in boreal forest aerosols from Hyytiälä, Finland, *Atmos. Chem. Phys.*, 5, 2761–2770, doi:10.5194/acp-5-2761-2005, 2005.
- Křůmal, K., Milkuška, P., and Večeřa, Z.: Polycyclic aromatic hydrocarbons and hopanes in PM₁ aerosols in urban areas, *Atmos. Environ.*, 67, 27–37, 2013.
- Kubatova, A., Vermeylen, R., Claeys, M., Cafmeyer, J., Maenhaut, W., Roberts, G., and Artaxo, P.: Carbonaceous aerosol characterization in the Amazon basin, Brazil: novel dicarboxylic acids and related compounds, *Atmos. Environ.*, 34, 5037–5051, 2000.
- Laaksonen, A., Kulmala, M., O'Dowd, C. D., Joutsensaari, J., Vaattovaara, P., Mikkonen, S., Lehtinen, K. E. J., Sogacheva, L., Dal Maso, M., Aalto, P., Petäjä, T., Sogachev, A., Yoon, Y. J., Lihavainen, H., Nilsson, D., Facchini, M. C., Cavalli, F., Fuzzi, S., Hoffmann, T., Arnold, F., Hanke, M., Sellegri, K., Umann, B., Junkermann, W., Coe, H., Allan, J. D., Alfarra, M. R., Worsnop, D. R., Riekkola, M.-L., Hyötyläinen, T., and Viisanen, Y.: The role of VOC oxidation products in continental new particle formation, *Atmos. Chem. Phys.*, 8, 2657–2665, doi:10.5194/acp-8-2657-2008, 2008.
- Lambe, A. T., Miracolo, M. A., Hennigan, C. J., Robinson, A. L., and Donahue, N. M.: Effective rate constants and uptake coefficients for the reactions of organic molecular markers (n-alkanes, hopanes, and steranes) in motor oil and diesel primary

- organic aerosols with hydroxyl radicals, *Environ. Sci. Technol.*, 43, 8794–8800, 2009.
- Langford, B., Nemitz, E., House, E., Phillips, G. J., Famulari, D., Davison, B., Hopkins, J. R., Lewis, A. C., and Hewitt, C. N.: Fluxes and concentrations of volatile organic compounds above central London, UK, *Atmos. Chem. Phys.*, 10, 627–645, doi:10.5194/acp-10-627-2010, 2010.
- Liao, H., Henze, D. K., Seinfeld, J. H., Wu, S., and Mickley, L. J.: Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, *J. Geophys. Res.*, 112, D06201, doi:10.1029/2006JD007813, 2007.
- Lin, Y.-H., Zhang, Z., Docherty, K. S., Zhang, H., Budisulistiorini, S. H., Rubitschun, C. L., Shaw, S. L., Knipping, E. M., Edgerton, E. S., Kleindienst, T. E., Gold, A., and Surratt, J. D.: Isoprene epoxydiols as precursors to secondary organic aerosol formation: Acid-catalyzed reactive uptake studies with authentic compounds, *Environ. Sci. Technol.*, 46, 250–258, 2012.
- Lin, Y.-H., Zhang, H., Pye, H. O. T., Zhang, Z., Marth, W. J., Park, S., Arashiro, M., Cui, T., Budisulistiorini, S. H., Sexton, K. G., Vizuete, W., Xie, Y., Luecken, D. J., Piletic, I. R., Edney, E. O., Bartolotti, L. J., Gold, A., and Surratt, J. D.: Epoxide as a precursor to secondary organic aerosol formation from isoprene photooxidation in the presence of nitrogen oxides, *Proc. Natl. Acad. Sci. USA*, 110, 6718–6723, 2013.
- Lohmann, R. and Lammel, G.: Adsorptive and absorptive contributions to the gas-particle partitioning of polycyclic aromatic hydrocarbons: State of knowledge and recommended parametrization for modeling, *Environ. Sci. Technol.*, 38, 3793–3803, 2004.
- Medeiros, P. M. and Simoneit, B. R. T.: Analysis of sugars in environmental samples by gas chromatography-mass spectrometry, *J. Chromatogr. A*, 1141, 271–278, 2007.
- Minguilón, M. C., Perron, N., Querol, X., Szidat, S., Fahrni, S. M., Alastuey, A., Jimenez, J. L., Mohr, C., Ortega, A. M., Day, D. A., Lanz, V. A., Wacker, L., Reche, C., Cusack, M., Amato, F., Kiss, G., Hoffer, A., Decesari, S., Moretti, F., Hillamo, R., Teinilä, K., Seco, R., Peñuelas, J., Metzger, A., Schallhart, S., Müller, M., Hansel, A., Burkhardt, J. F., Baltensperger, U., and Prévôt, A. S. H.: Fossil versus contemporary sources of fine elemental and organic carbonaceous particulate matter during the DAURE campaign in Northeast Spain, *Atmos. Chem. Phys.*, 11, 12067–12084, doi:10.5194/acp-11-12067-2011, 2011.
- Mochida, M., Kawabata, A., Kawamura, K., Hatsushika, H., and Yamazaki, K.: Seasonal variation and origins of dicarboxylic acids in the marine atmosphere over the western North Pacific, *J. Geophys. Res.-Atmos.*, 108, 4193, doi:10.1029/2002JD002355, 2003.
- Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., and Prévôt, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, *Atmos. Chem. Phys.*, 12, 1649–1665, doi:10.5194/acp-12-1649-2012, 2012.
- Moise, T. and Rudich, Y.: Reactive Uptake of Ozone by Aerosol-Associated Unsaturated Fatty Acids: Kinetics, Mechanism, and Products, *J. Phys. Chem. A*, 106, 669–6476, 2002.
- Ning, Z., Geller, M. D., Moore, K. F., Sheesley, R., Schauer, J. J., and Sioutas, C.: Daily variation in chemical characteristics of urban ultrafine aerosols and inference of their sources, *Environ. Sci. Technol.*, 41, 6000–6006, 2007.
- O'Dowd, C. D., Aalto, P., Hämeri, K., Kulmala, M., and Hoffmann, T.: Atmospheric particles from organic vapours, *Nature*, 416, 497–498, 2002.
- Paatero, P. and Tapper, U.: Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, 5, 111–126, 1994.
- Paulot, F., Wunch, D., Crouse, J. D., Toon, G. C., Millet, D. B., DeCarlo, P. F., Vigouroux, C., Deutscher, N. M., González Abad, G., Notholt, J., Warneke, T., Hannigan, J. W., Warneke, C., de Gouw, J. A., Dunlea, E. J., De Mazière, M., Griffith, D. W. T., Bernath, P., Jimenez, J. L., and Wennberg, P. O.: Importance of secondary sources in the atmospheric budgets of formic and acetic acids, *Atmos. Chem. Phys.*, 11, 1989–2013, doi:10.5194/acp-11-1989-2011, 2011.
- Park, C., Schade, G. W., and Boedeker, I.: Characteristics of the flux of isoprene and its oxidation products in an urban area, *J. Geophys. Res.-Atmos.*, 116, D21303, doi:10.1029/2011JD015856, 2011.
- Pérez, N., Pey, J., Cusack, M., Reche, C., Querol, X., Alastuey, A., and Viana, M.: Variability of particle number, black carbon and PM₁₀, PM_{2.5} and PM₁ levels and speciation: influence of road traffic emissions on urban air quality, *Aerosol Sci. Technol.*, 44, 487–499, 2010.
- Pey, J., Querol, X., and Alastuey, A.: Variations of levels and composition of PM₁₀ and PM_{2.5} at an insular site in the Western Mediterranean, *Atmos. Res.*, 94, 285–299, 2009.
- Pratt, K. A. and Prather, K. A.: Mass spectrometry of atmospheric aerosols – Recent developments and applications, Part II: On-line mass spectrometry techniques, *Mass Spectrom. Rev.*, 31, 17–48, 2012.
- Puxbaum, H., Caseiro, A., Sánchez-Ochoa, A., Kasper-Giebl, A., Cleays, M., Gelencsér, A., Legrand, M., Preunkert, S., and Pio, C.: Levoglucosan levels at background sites in Europe for assessing the impact of biomass combustion on the European aerosol background, *J. Geophys. Res.*, 112, D23S05, doi:10.1029/2006JD008114, 2007.
- Querol, X., Alastuey, A., Moreno, T., Viana, M. M., Castillo, S., Pey, J., Rodríguez, S., Artiñano, B., Salvador, P., Sánchez, M., García Dos Santos, S., Herce Garraleta, M. D., Fernandez-Patier, R., Moreno-Grau, S., Negral, L., Minguilón, M. C., Monfort, E., Sanz, M. J., Palomo-Marín, R., Pinilla-Gil, E., Cuevas, E., de la Rosa, J., and Sánchez de la Campa, A.: Spatial and temporal variations in airborne particulate matter (PM₁₀ and PM_{2.5}) across Spain 1999–2005, *Atmos. Environ.*, 42, 3964–3979, 2009.
- Reche, C., Moreno, T., Amato, F., Viana, M., van Drooge, B.L., Chuang, H., Bérubé, K., Jones, T., Alastuey, A., and Querol, X.: A multidisciplinary approach to characterise exposure risk and toxicological effects of PM₁₀ and PM_{2.5} samples in urban environments, *Ecotox. Environ. Safe.*, 78, 327–335, 2012.
- Robinson, A. L., Donahue, N. M., and Rogge, W. F.: Photochemical oxidation and changes in molecular composition of organic aerosol in the regional context, *J. Geophys. Res.*, 111, D03302, doi:10.1029/2005JD006265, 2006a.
- Robinson, A. L., Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., and Rogge, W. F.: Source apportionment of molecular markers and organic aerosol – 3. food cooking emissions, *Environ. Sci. Technol.*, 40, 7820–7827, 2006b.

10370

M. Alier et al.: Source apportionment of submicron organic aerosol

- Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking organic aerosols: semivolatile emissions and photochemical aging, *Science*, 315, 1259–1262, 2007.
- Rogge, W. F., Hildemann, L. M., Mazurek, M. A., Cass, G. R., and Simoneit, B. R. T.: Sources of fine organic aerosol. 2. Noncatalyst and catalyst-equipped automobiles and heavy duty diesel trucks, *Environ. Sci. Technol.*, 27, 636–651, 1993.
- Rogge, W.F., Hildemann, L.M., Marurek, M.A., Glen, R., Cass, G. R., Simoneit, B.R.: Sources of Fine Organic Aerosol. 6. Cigarette Smoke in the Urban Atmosphere, *Environ. Sci. Technol.*, 28, 1375–1388, 1994.
- Sage, A. M., Weitkamp, E. A., Robinson, A. L., and Donahue, N. M.: Evolving mass spectra of the oxidized component of organic aerosol: results from aerosol mass spectrometer analyses of aged diesel emissions, *Atmos. Chem. Phys.*, 8, 1139–1152, doi:10.5194/acp-8-1139-2008, 2008.
- Salau, J. S. I., Tauler, R., Bayona, J. M., and Tolosa, I.: Input Characterization of Sedimentary Organic Contaminants and Molecular Markers in the Northwestern Mediterranean Sea by Exploratory Data Analysis, *Environ. Sci. Technol.* 31, 3482–3490, 1997.
- Schauer, J. J., Kleemann, M. J., Cass, G. R., and Simoneit, B. R. T.: Measurements of emissions from air pollution sources. 5. C1–C32 organic compounds from gasoline-powered motor vehicles, *Environ. Sci. Technol.*, 36, 1169–1180, 2002.
- Schauer, J. J., Rogge, W. F., Hildemann, L. M., Mazurek, M. A., and Cass, G. R.: Source apportionment of airborne particulate matter using organic compounds as tracers, *Atmos. Environ.*, 41, S241–S259, 2007.
- Schneidemesser, E. V., Monks, P. S., Gros, V., Gauduin, J., and Sanchez, O.: How important is biogenic isoprene in an urban environment? A study in London and Paris, *Geophys. Res. Lett.*, 38, L19804, doi:10.1029/2011GL048647, 2011.
- Seco, R., Peñuelas, J., Filella, I., Llusia, J., Molowny-Horas, R., Schallhart, S., Metzger, A., Müller, M., and Hansel, A.: Contrasting winter and summer VOC mixing ratios at a forest site in the Western Mediterranean Basin: the effect of local biogenic emissions, *Atmos. Chem. Phys.*, 11, 13161–13179, doi:10.5194/acp-11-13161-2011, 2011.
- Seco, R., Peñuelas, J., Filella, I., Llusia, J., Schallhart, S., Metzger, A., Müller, M., and Hansel, A.: Volatile organic compounds in the western Mediterranean basin: urban and rural winter measurements during the DAURE campaign, *Atmos. Chem. Phys.*, 13, 4291–4306, doi:10.5194/acp-13-4291-2013, 2013.
- Sheesley, R. J., Deminter, J. T., Meiritz, M., Snyder, D. C., and Schauer, J. J.: Temporal trends in motor vehicle and secondary organic tracers using in situ methylation thermal desorption GCMS, *Environ. Sci. Technol.*, 44, 9398–9404, 2010.
- Simoneit, B. R. T.: Biomass burning – a review of organic tracers for smoke from incomplete combustion, *Appl. Geochem.*, 17, 129–162, 2002.
- Stanimirova, I., Tauler, R., and Walczak, B.: A Comparison of Positive Matrix Factorization and the Weighted Multivariate Curve Resolution Method. Application to Environmental Data, *Environ. Sci. Technol.*, 45, 10102–10110, 2011.
- Stephanou, E. G. and Stratigakis, N.: Oxocarboxylic and alpha,omega-dicarboxylic acids – photooxidation products of biogenic unsaturated fatty-acids present in urban aerosols, *Environ. Sci. Technol.*, 27, 1403–1407, 1993.
- Stone, E. A., Zhou, J., Snyder, D. C., Rutter, A. P., Mieritz, M., and Schauer, J. J.: A comparison of summertime secondary organic aerosol source contributions at contrasting urban locations, *Environ. Sci. Technol.*, 43, 3448–3454, 2009.
- Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., Rogge, W. F., and Robinson, A. L.: Contribution of motor vehicle emissions to organic carbon and fine particle mass in Pittsburgh, Pennsylvania: Effects of varying source profiles and seasonal trends in ambient marker concentrations, *Atmos. Environ.*, 40, 8002–8019, 2006.
- Sureda, X., Martinez-Sanchez, J. M., Lopez, M. J., Fu, M., Agüero, F., Salto, E., Nebot, M., and Fernandez, E.: Secondhand smoke levels in public building main entrances: outdoor and indoor PM_{2.5} assessment, *Tobacco Control* 21, 543–548, 2012.
- Surratt, J. D., Chan, A. W. H., Eddingsaas, N. C., Chan, M. N., Loza, C. L., Kwan, A. L., Hersey, S. P., Flagan, R. C., Wennberg, P. O., and Seinfeld, J. H.: Reactive intermediates revealed in secondary organic aerosol formation from isoprene, *P. Natl. Acad. Sci. USA*, 107, 6640–6645, 2010.
- Szmigielski, R., Surratt, J.D., Gómez-González, Y., Van der Veken, P., Kourtchev, I., Vermeylen, R., Blockhuys, F., Jaoui, M., Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., Edney, E. O., Seinfeld, J. H., Maenhaut, W., and Claeys, M.: 3-Methyl-1,2,3-butanetricarboxylic acid: an atmospheric tracer for terpene secondary organic aerosol, *Geophys. Res. Lett.*, 34, L24811, doi:10.1029/2007GL031338, 2007.
- Taraborrelli, D., Lawrence, M. G., Crowley, J. N., Dillon, T. J., Gromov, S., Groß, C. B. M., Vereecken, L., and Lelieveld, J.: Hydroxyl radical buffered by isoprene oxidation over tropical forests, *Nature Geosci.*, 5, 190–193, 2012.
- Tauler, R.: Multivariate curve resolution applied to second order data, *Chemom. Intell. Lab. Syst.*, 30, 133–146, 1995. 43, 3989–3997, 2009.
- Tauler, R., Smilde, A., and Kowalski, B.: Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution, *J. Chemometr.*, 9, 31–58, 1995.
- Tauler, R., Viana, M., Querol, X., Alastuey, A., Flight, R. M., Wentzell, P. D., and Hopke, P. K.: Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, *Atmos. Environ.*,
- Terrado, M., Barceló, D., and Tauler, R.: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns in surface water, *Environ. Sci. Technol.*, 43, 5321–5326, 2009.
- US-EPA, Protocol for applying and validating the CMB model (ed. Office for Air Quality Planning and Standards), 1987.
- van Drooge, B. L., Crusack, M., Reche, C., Mohr, C., Alastuey, A., Querol, X., Prevot, A., Day, D. A., Jimenez, J. L., and Grimalt, J. O.: Molecular marker characterization of the organic composition of submicron aerosols from Mediterranean urban and rural environments under contrasting meteorological conditions, *Atmos. Environ.*, 61, 482–489, 2012a.
- van Drooge, B. L., Lopez, J., and Grimalt, J. O.: Influences of natural emission sources (wildfires and Saharan dust) on the urban organic aerosol in Barcelona (Western Mediterranean Basin) during a PM event, *Environ. Sci. Pollut. Res.*, 19, 4159–4167, 2012b.

M. Alier et al.: Source apportionment of submicron organic aerosol**10371**

- Watson, J. G.: Visibility: Science and regulation, *J. AirWaste Manage.*, 52, 628–713, 2002.
- WHO: http://www.who.int/whosis/whostat/EN_WHS10_Full.pdf, last access: December 2010.
- Yan, B., Zheng, M., Hu, Y. T., Lee, S., Kim, H. K., and Russell, A. G.: Organic composition of carbonaceous aerosols in an aged prescribed fire plume, *Atmos. Chem. Phys.*, 8, 6381–6394, doi:10.5194/acp-8-6381-2008, 2008.
- Yang, L., Ray, M. B., and Yu, L. E.: Photooxidation of dicarboxylic acids – Part II: Kinetics, intermediates and field observations, *Atmos. Environ.*, 42, 868–880, 2008a.
- Yang, L., Ray, M. B., and Yu, L. E.: Photooxidation of dicarboxylic acids – Part I: Effects of inorganic ions on degradation of azelaic acid, *Atmos. Environ.*, 42, 856–867, 2008b.
- Zhang, Y. Y., Müller, L., Winterhalter, R., Moortgat, G. K., Hoffmann, T., and Pöschl, U.: Seasonal cycle and temperature dependence of pinene oxidation products, dicarboxylic acids and nitrophenols in fine and coarse air particulate matter, *Atmos. Chem. Phys.*, 10, 7859–7873, doi:10.5194/acp-10-7859-2010, 2010.
- Zielinska, B., Sagebiel, J., McDonald, J. D., Whitley, K., and Lawson, D. R.: Emission rates and comparative chemical composition from selected in-use diesel and gasoline-fueled vehicles, *J. Air Waste Manage. Assoc.*, 54, 1138–1150, 2004.

1 **SUPPORTING INFORMATION**

2

3 **Source apportionment of submicron organic aerosol at an urban background**

4 **and road site in Barcelona during SAPUSS**

5

6 Marta Alier, Barend L. van Drooge, Manuel Dall'Osto, Xavier Querol, Joan O. Grimalt,

7 Romà Tauler

8

9 Department of Environmental Chemistry, Institute of Environmental Assessment and

10 Water Research (IDAEA-CSIC), C/Jordi Girona, 18-26, 08034, Barcelona, Spain

11

12 Table SI 1. Seven solution component with all 35 compounds

13

14

Seven solution component factor	Comment
1 POA Urban 1	It maintains the features of the six factor solution POA Urban, with a main PAH (PAH_high) structure.
2 POA Urban 2	POA Urban 2 (new component). This is a splitting of POA Urban 1. The merging of Urban POA 1 and Urban POA 2 from the seven solution component give the urban POA of the 6 component ($R^2=1$). This solution is a splitting of the POA Urban.
3 OOA Urban	OOA Urban. This solution is very constant over the 6 factor solution. The regression line is 1:1, but only the C_7 - C_9 signature is conserved in the loadings of the organic compounds, and the PAH-hopanes structure is lost.
4 SOA BIO ISO	This component is really stable, has same temporal trend and same R^2 and same factor profile with the 6 and 7 factor solution. There is almost no difference between the 6 and 7 factor solution.
5 SOA BIO PIN	Main component is still cis-pinonic acid, but the second one major one is a PAH. The temporal trend of this component is also less clear.
6 BBOA Reg.	This solution is similar to the 6 factor one. However, there is generally lower conc. ($0.2 \mu\text{g m}^{-3}$) than previous studies (0.3 - $0.4 \mu\text{g m}^{-3}$). Overall the RS-UB correlation are the same ($R^2=0.88$). However, the correlation between RS and UB for the last period is much better ($R^2=0.95$) for the 6 component than the 7 one ($R^2<0.7$). The 6 factor solution is better.
7 SOA aged	This component maintains the feature of the aged SOA. However, the difference in conc. between RS and UB is very high. The 6 factor solution is better.

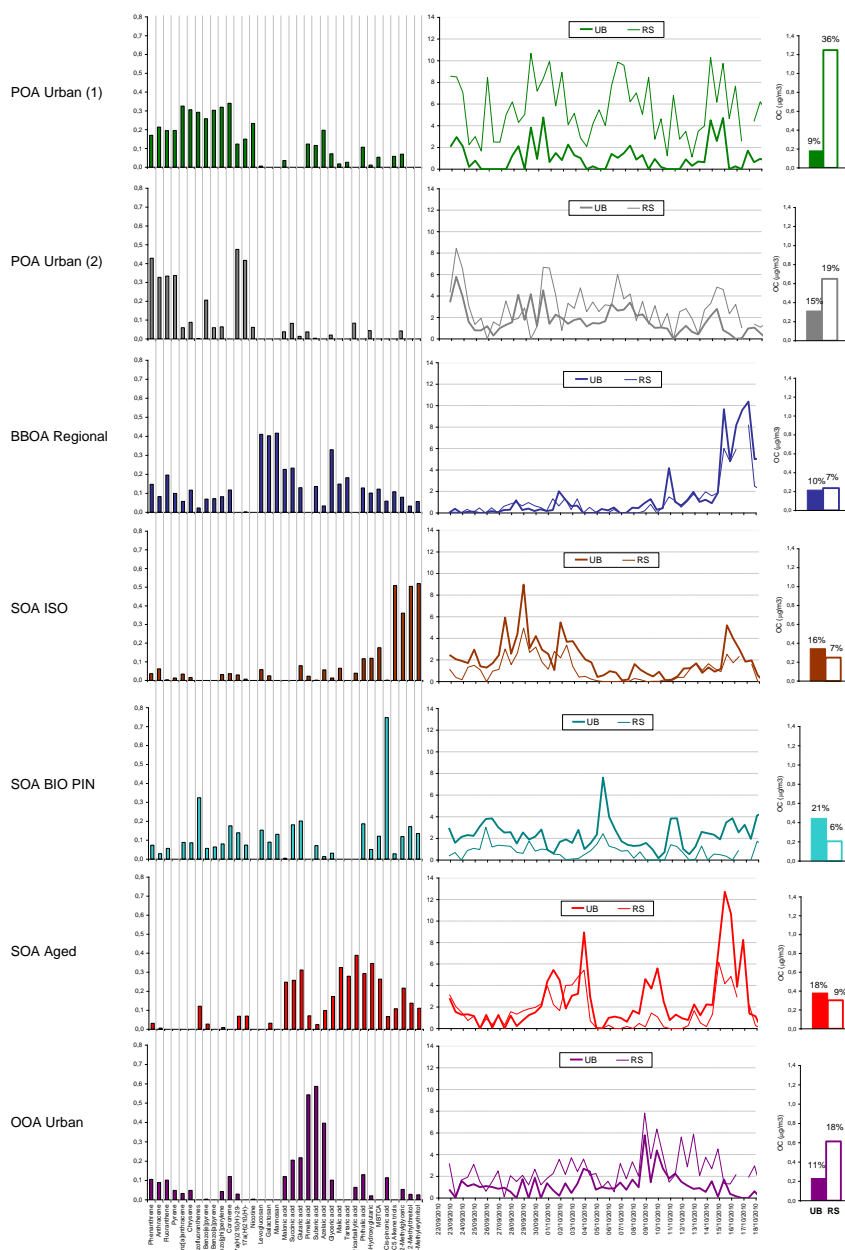
15

16 Figure SI 1. Organic markers loadings, temporal trends and concentrations for the

17 seven solution

18

19



20

21 Figure S11. MCR-ALS resolved profiles applying non-negativity constraints. Column on
 22 the left shows the loading organic compound composition of the six selected
 23 components. Column in the middle shows the temporal trend of the score values of the
 24 six components. Column on the right shows the relative contribution of the scores (%)
 25 as well as the estimated mass contribution of organic carbon to volume ($\mu\text{g OCm}^{-3}$).

On the origin of water-soluble organic tracer compounds in fine aerosols in two cities: the case of Los Angeles and Barcelona

M. Alier · M. Dall'Osto · Y.-H. Lin · J. D. Surratt ·
R. Tauler · J. O. Grimalt · B. L. van Drooge

Received: 14 September 2013 / Accepted: 12 December 2013
© Springer-Verlag Berlin Heidelberg 2014

Abstract Water-soluble organic compounds (WSOCs), represented by anhydro-saccharides, dicarboxylic acids, and polyols, were analyzed by gas chromatography interfaced to mass spectrometry in extracts from 103 PM₁ and 22 PM_{2.5} filter samples collected in an urban background and road site in Barcelona (Spain) and an urban background site in Los Angeles (USA), respectively, during 1-month intensive sampling campaigns in 2010. Both locations have similar Mediterranean climates, with relatively high solar radiation and frequent anti-cyclonic conditions, and are influenced by a complex mixture of emission sources. Multivariate curve resolution-alternating least squares analyses were applied on the database in order to resolve differences and similarities in WSOC compositions in the studied sites. Five consistent clusters for the analyzed compounds were obtained, representing primary regional biomass burning organic carbon, three secondary organic components (aged SOC, isoprene SOC, and α -pinene SOC), and a less clear component, called urban oxygenated organic carbon. This last component is probably influenced by in situ urban activities, such as food cooking and traffic emissions and oxidation processes.

Keywords PM · WSOC · Secondary organic aerosol · Dicarboxylic acids · Anhydro-saccharides

Introduction

Organic carbon (OC) accounts for a major fraction of the atmospheric fine particulate matter (PM_{2.5}) and water-soluble organic compounds (WSOC) are important contributors to OC (Hersey et al. 2011). Biomass burning is an important primary emission source for WSOC (Fine et al. 2004). Nevertheless, in the absence of biomass burning, most of the WSOC is thought to be derived from secondary organics of which many are oxygenated compounds, such as dicarboxylic acids (DCA) (Hallquist et al. 2009). Although DCA are emitted in small quantities from traffic and vegetation it is expected that the majority of these oxidized compounds are formed in the atmosphere after photochemical transformation of volatile and semi-volatile organic compounds from nonfossil (e.g., vegetation) as well as fossil origins (e.g., fossil fuel combustion) (Heald et al. 2010; Kleindienst et al. 2012; Paulot et al. 2011). The importance of oxidized organics is emphasized by their large contribution (40–90 %) to the total the organic fraction in PM_{2.5}, especially in polluted areas (Jimenez et al. 2009).

The composition and evolution of secondary organic carbon (SOC) in the atmosphere is still not well known and prone to controversy (Hallquist et al. 2009). For example, several studies showed that diluted emissions from diesel emissions or biomass burning produce large quantities of SOC, whereas field measurements in the Los Angeles (LA) basin indicated an enhanced formation of SOC from gasoline emissions over diesel emissions (Bahreini et al. 2012). Although the net emissions of volatile organic compounds (VOCs) from vehicles have been reduced by a factor of two in LA over the past decades (despite that the number of vehicles has increased), it is still thought to be a dominant source (Warneke et al. 2012).

Responsible editor: Gerhard Lammel

M. Alier · R. Tauler · J. O. Grimalt · B. L. van Drooge (✉)
Department of Environmental Chemistry, Institute for Environmental
Assessment and Water (IDAEA-CSIC), Barcelona, Spain
e-mail: barend.vandrooge@idaea.csic.es

Y.-H. Lin · J. D. Surratt
Department of Environmental Sciences and Engineering, Gillings
School of Global Public Health, University of North Carolina,
Chapel Hill, NC, USA

Present Address:

M. D. Osto
Institut de Ciències del Mar (ICM-CSIC), Pg. Marítim de la
Barceloneta 37–49, 08003 Barcelona, Spain

Nevertheless, the contributions of oxygenated OC from non-fossil sources, such as biogenic VOCs emitted from vegetation, are still uncertain. Recent results from radioactive carbon (^{14}C) analysis of urban aerosols collected from both Barcelona (BCN) and LA showed that non-fossil sources contribute for about 50 % of the fine OC aerosol (Minguillón et al. 2011; Bahreini et al. 2012). Moreover, food-cooking activities seem to contribute substantially to OC in urban areas. In both LA and BCN, these contributions were estimated to be around 15 % of the OC based on online high-resolution aerosol mass spectrometry (HR-AMS) analysis (Hayes et al. 2013; Mohr et al. 2012).

The comparison between LA and BCN is of interest as both urban areas have similar Mediterranean climates, which are characterized by high solar radiation and frequent anti-cyclonic atmospheric conditions, resulting in accumulation of primary and secondary aerosols. Moreover, both urban areas are under influence of intensive primary urban emissions from vehicles. The vehicle composition in the LA basin is dominated by gasoline-based engines (Bahreini et al. 2012; Hayes et al. 2013), whereas around 75 % of the vehicles in BCN have diesel engines and 30 % of the total vehicles are motorcycles (Reche et al. 2012). Previous results from the same sampling campaigns as the ones in the present study have shown that the urban primary emissions contributed for 12 % to the organic aerosol in LA (Hayes et al. 2013), which was similar to the 18 % estimated in the urban background site in BCN, while this contribution in the traffic intensive road site was 43 %.

As the number of chemical species is very large in fine aerosol, it is almost impossible to obtain a comprehensive chemical characterization of organic aerosol. As a result, the analysis of tracer organic compounds can provide insight into the origin and evolution of the organic aerosol if these compounds are related to source emissions and photooxidation processes. The tracer compounds measured in this study are:

- Levoglucosan, galactosan, and mannosan are monosaccharide anhydrides that are thermal alteration products of cellulose and hemi-cellulose, respectively (Simoneit 2002). Levoglucosan is emitted in large quantities during biomass burning (Fine et al. 2004), therefore, easily identified in PM samples.
- DCA, hydroxy-DCA and aromatic-DCA have received much attention because of their role in affecting the global climate and their value as organic tracers for secondary organic aerosol formation (Heald et al. 2010; Moise and Rudich 2002; Paulot et al. 2011; Yang et al. 2008).
- *Cis*-pinonic acid, 3-hydroglutaric acid and 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) have been identified in fine aerosols and related to the photochemical oxidation of α -pinene (Claeys et al. 2007; Szmigielski et al. 2007), which is the most abundant monoterpene in the studied area derived from biogenic emissions. MBTCA is formed by OH-initiated oxidation of *cis*-pinonic acid (Szmigielski et al. 2007).

- 2-methylglyceric acid and polyols, such as C_5 -alkene triols and 2-methyltetrols, have been related to isoprene oxidation (Claeys et al. 2004; Hallquist et al. 2009), a major VOC emitted from terrestrial vegetation.

The selection of these compounds allows analyzing the influence of the different potential sources (anthropogenic vs. biogenic; fossil vs. nonfossil) and oxidation processes on organic aerosol formation. Besides the chemical analysis of these WSOC constituents in the filter extracts from LA and BCN, the relationships among these chemical organic species was further investigated using multivariate curve resolution-alternating least squares (MCR-ALS) analyses on data matrices. MCR-ALS is based on a bilinear decomposition as it is described elsewhere in more detail (Tauler et al. 1995; Jaumot et al. 2005). One of the main advantages of MCR-ALS is that decomposes the data matrix by applying natural constraints, such as non-negativity, and thus, results can be interpreted more straightforwardly. MCR-ALS has been successfully used in the analysis of environmental data sets in air source apportionment studies (Tauler et al. 2009; Alier et al. 2013).

Methods and materials

Sampling sites in BCN

Two sites were sampled in BCN from 22 September 2010 to 18 October 2010 as part of the Solving Aerosol Problems by Using Synergistic Strategies project (SAPUSS). The urban background site (BCN-UB; 41.3899° N; 2.1161° E; 80 m above sea level (a.s.l.)) was located in a residential area, whereas the road site (BCN-RS; 41.3884° N; 2.1500° E; 40 m a.s.l.) was located in Urgell Street within the square-grid street network of BCN's city center. More detailed information is described elsewhere (Dall'Osto et al. 2013).

Sampling site in LA

The sampling site in the LA Basin was located in Pasadena, CA (LA; 34.1408° N; 118.1223° W; 230 m a.s.l.), where continuous gas- and aerosol-phase sampling occurred from 15 May to 16 June 2010 as part of the CalNex (California Research at the Nexus of Air Quality and Climate Change) field study (Ryerson et al. 2013). This site is 18 km NE from downtown LA and 44 km from the Long Beach harbor area. Based on the characteristic of this site it is comparable to the background site in BCN.

Sampling details

In BCN, all filter samples were collected on pre-baked (450 °C overnight) quartz fiber filters (Tissuquartz™ Filters,

2500 QAT-UP, Pall Life Sciences), allowing for high-volume PM₁ sampling (Digitel-DH80; 30 m³/h). Samples were collected in 12-h sample interval from 9:00 to 21:00 and 21:00 to 9:00 local time. In LA, filter samples were collected on quartz fiber filters (Tissuquartz™ Filters, 2500 QAT-UP, Pall Life Sciences), allowing for high-volume PM_{2.5} (TE-6001; 60 m³/h) of 23-h integrated filter samples from midnight to 11 PM local time.

Filter samples collected in both locations from each day were stored in a freezer at −18 °C until chemical analysis. Field blanks were collected every 7–10 days by placing a pre-fired quartz fiber filter into the sampler for 15 min before removing and storing in the same manner as the field samples. Analysis of both field and lab blanks showed no significant organic contaminants on pre-fired quartz filters, indicating that all of the organic compounds characterized in the present study are due to the aerosol collected from LA and BCN.

Organic tracer analysis

A detailed description of the analytical procedure can be found elsewhere (Alier et al. 2013). Briefly, a fraction (1/4 or 1/8 parts) of the blank and sample filter was ultrasonically extracted in a mixture of (2:1, v/v) dichloromethane and methanol (3×15 ml, Merck, Germany). After extraction, the extracts were filtered, concentrated by rotovap and evaporated under a gentle stream of N₂ until dryness. WSOCs were derivatized to their methylsilylate esters by adding *bis*-(trimethylsilyl)trifluoroacetamide+trimethylchlorosilane (99:1; Supelco, USA) and pyridine (Merck, Germany). A Thermo gas chromatography interfaced to mass spectrometry (GC/MS; Thermo Trace GC Ultra-DSQ II) equipped with a 60-m fused capillary column (HP-5MS 0.25-mm×25-μm film thickness) was used for analysis operating in full scan (*m/z* 50–650) and electron ionization (70 eV) modes. Compounds were identified with their characteristic ions at the corresponding retention time in authentic standards or by comparison of mass spectrometric fragmentation patterns from literature and library data in the case of 3-hydroxyglutaric acid, 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA), 2-methylglyceric acid, C₅-alkene triols, 2-methylthreitol, and 2-methylerythritol (Claeys et al. 2007). Quantification was performed by calculating the concentrations using external standard calibration curves of authentic standard. Here, the chromatographic area of the ion of an analyte was divided by the area of internal standard, so a calibration curve can be formed with the known amounts of the standards. Then, the area of the analyte in the sample, divided by the area of the internal standard, was introduced to the regression line of the calibration curve in order to obtain the amount of the analyte in the extract (corrected by the known amount of internal standard). To correct for losses during the analytical procedure, the amounts of the analytes

are corrected by the recoveries of the surrogate standards (succinic acid-d₄ (Sigma Aldrich) and levoglucosan-d₇ (Cambridge Isotopic Laboratories)) that were added to the filter samples before extraction.

Data arrangement

Experimental data obtained from GC/MS were arranged in a data matrix for every analyzed station and used for chemometric analysis in the MCR-ALS method. In the two BCN sites, the samples were stored in two separate data matrices in 12-h periods from 22 September to 18 October 2010. In LA site, samples were stored in a single data matrix in 24-h periods from 15 May to 12 June 2010. The samples of three locations were analyzed individually and simultaneously. Experimental data from the different samples (excluding field blanks) are set in the matrix rows and the measured compounds in the matrix columns. This gave a data matrix of 52 rows (samples) and 17 columns (measured compounds) for BCN-UB station, a data matrix of 51×17 for BCN-RS station and a data matrix of 22×17 for LA station. As the three data matrices (stations) had the same number of columns (analyzed compounds), the three individual data matrices were also arranged in a column-wise augmented data matrix with dimensions 125×17.

All data arrangements and pretreatments were performed into MATLAB 7.4 (The Mathworks, Natick, USA) for subsequent multivariate data analysis using MATLAB PLS 5.8 Toolbox (Eigenvector Research Inc, Masson WA, USA).

Results and discussion

WSOC analysis

Table 1 shows the mean, minimum and maximum concentrations of WSOC constituents (or tracers) in the two locations after analyses of the extracts by GC/MS.

Anhydro-saccharides

Levoglucosan, mannosan and galactosan, as organic tracers for biomass burning, showed low concentrations, with mean levels for levoglucosan of 14, 9, and 5 ng/m³ in LA, BCN-UB, and BCN-RS, respectively (Table 1). The levoglucosan levels were comparable to those found in background sites where biomass burning has little influence on the air quality (Puxbaum et al. 2007). Ratios between the compounds have been used in the past to indicate the biomass source of the combustion (Fine et al. 2004, Schmidt et al. 2008). The ratios of levoglucosan to mannosan were very consistent during the samplings and similar in all sites (9.8±1.7 in LA, 9.6±3.4 in BCN-UB, and 9.2±2.1 in BCN-RS). Based on combustion

Table 1 Mean (minimum–maximum) concentrations (nanograms per cubic meter) of analyzed WSOC in PM samples from LA and BCN (number of samples (N))

	LA (N=22)	BCN-UB (N=52)	BCN-RS (N=51)
Galactosan	1.4 (0.6–2.5)	1.3 (0.2–7.1)	0.7 (0.1–4.4)
Mannosan	1.5 (0.7–3.1)	1.2 (0.2–5.6)	0.6 (0.1–3.8)
Levoglucozan	14.1 (7.1–24.0)	9.4 (1.7–40.0)	5.3 (0.6–30.5)
Succinic acid	14.4 (4.2–38.4)	7.3 (3.1–24.7)	6.2 (2.5–14.8)
Glutaric acid	5.9 (1.8–13.9)	2.0 (0.6–7.5)	1.5 (0.6–3.2)
Pimelic acid	1.0 (0.5–1.8)	0.6 (0.2–2.1)	1.3 (0.6–3.7)
Suberic acid	0.9 (0.3–1.8)	0.9 (0.4–3.0)	1.7 (0.7–3.9)
Azelaic acid	5.1 (2.5–12.1)	2.9 (1.0–7.3)	6.1 (2.4–13.0)
Malic acid	19.4 (4.0–47.1)	14.2 (1.2–72.7)	9.9 (1.2–33.6)
Phthalic acid	5.2 (0.7–13.2)	3.9 (0.9–11.4)	3.6 (1.3–9.3)
<i>Cis</i> -pinonic acid	4.8 (2.2–11.3)	15.4 (3.6–49.7)	8.0 (3.1–18.4)
3-hydroxyglutaric acid	6.2 (1.1–14.1)	4.5 (0.5–20.2)	3.1 (0.8–8.6)
MBTCA	1.2 (0.3–2.7)	5.5 (1.0–23.5)	4.1 (0.7–10.3)
C5-alkene triols	5.7 (1.5–12.6)	1.0 (0.1–3.0)	0.7 (0.1–2.2)
2-methylglyceric acid	6.2 (1.1–14.1)	2.1 (0.4–6.4)	1.9 (0.6–4.5)
2-methylthreitol	3.7 (1.4–7.9)	2.8 (0.6–7.8)	1.5 (0.4–3.8)
2-methylerythritol	9.3 (3.1–22.4)	6.5 (1.7–20.7)	3.5 (0.6–9.5)

experiments using different wood types, where softwood and hardwood ratios are around 4 ± 1 and 18 ± 5 , respectively (Fine et al. 2004; Schmidt et al. 2008), the observed ratios in the present study indicate a mixture of soft- and hardwood.

Both in LA and BCN, wood combustion is not a local energy source and the observed concentrations are representative for influence of regional biomass combustion. This is further indicated by the good correlation between the temporal concentration trends in the two sites in BCN ($r^2=0.8$). Here, higher concentrations were observed in the last part of the sampling campaign (Fig. 1), which coincidence with the period when open fires in fields for biomass waste burning are legally allowed in the region. In LA, the concentrations showed fluctuations between 10 and 20 ng/m^3 , but there was a decreasing trend along the sampling period (Fig. 1; $r^2=0.3$). The small influence of biomass burning in LA was also observed in the HR-AMS analysis during the CalNex campaign (Hayes et al. 2013). In the HR-AMS analyses, the biomass burning signal was below detection limit and probably mixed with secondary organic species, indicating the origin of the biomass burning aerosol was the result of long-range transport.

The 12-h resolution samples in BCN did not show any clear difference between day and nighttime samples for the concentrations of anhydro-saccharides, although during the end of the sampling period, slightly higher levels were observed during the nighttime, which probably indicates the influx of biomass burning influenced air masses from the inland by the land-breeze during this period.

Despite the detection of the anhydro-saccharides in the PM samples in both locations, as well as the indications that these

compounds are sampled in aerosols after long-rang atmospheric transport, it is important to remember that these organic tracers for biomass burning may be susceptible to oxidation (Hennigan et al. 2011; Hoffmann et al. 2010).

DCA, hydroxy-DCA, and aromatic-DCA

The mean concentration of these compounds ranged between 1 and 19 ng/m^3 in LA, 1–14 ng/m^3 in BCN-UB, and 1–10 ng/m^3 in BCN-RS (Table 1), with the highest individual compound concentrations observed for malic acid (73 ng/m^3 in BCN and 47 ng/m^3 in LA) and succinic acid (24 ng/m^3 in BCN and 38 ng/m^3 in LA). The higher levels of malic acid (hydroxy-succinic acid) than its parent compound, succinic acid, was also observed in other sites that were influenced by anthropogenic emission sources although a non-fossil origin of these compounds cannot be excluded (Yang et al. 2008, and references therein). Malic acid concentrations were similar to those measured in winter in the urban area of BCN (van Drooge et al. 2012) and comparable to other European sites, such as K-Puszt, Hungary (38 ng/m^3) (Ion et al. 2005) and Jülich, Germany (39 ng/m^3) (Kourtchev et al. 2008).

Phthalic acid, a possible oxidation product of gas-phase polycyclic aromatic hydrocarbons from fossil fuel combustion, had similar concentrations in the three sites ($\sim 4 \text{ ng}/\text{m}^3$; Table 1). Phthalic acid concentrations in LA were consistent with those obtained from GC-ITMS analyses by Kleindienst et al. (2012) in collocated filter samples collected from the CalNex study. Moreover, they identified 2-methylphthalic acid in their sample extracts at concentrations around 3 times lower than phthalic acid, but with a similar temporal trend.

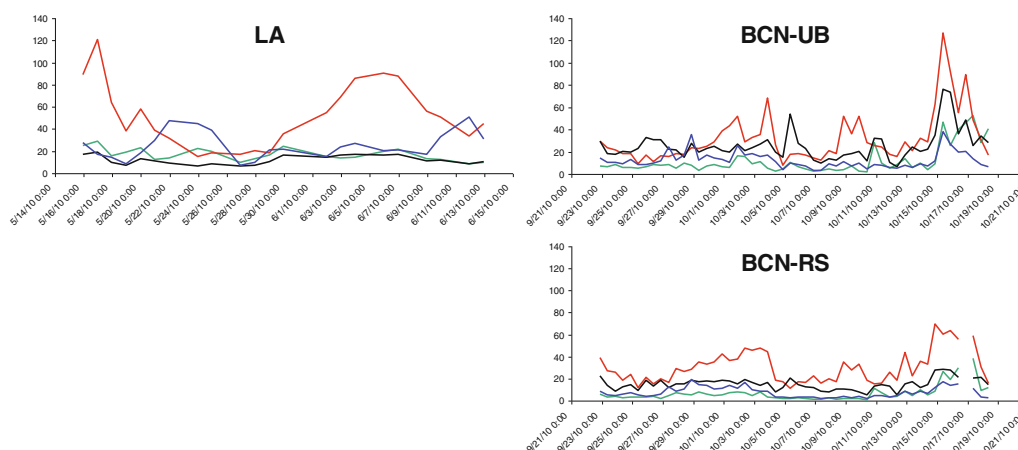


Fig. 1 Temporal trends of concentrations (nanograms per cubic meter) of the $\sum_{\text{anhydro-saccharides}}$ (green), $\sum_{\text{dicarboxylic acids}}$ (red), \sum_{SOC} tracers for α -pinene (black), and \sum_{SOC} tracers for isoprene (blue) in the three sampling sites

Posterior analysis of the GC/MS chromatograms in our study confirms the presence of 2-methylphthalic acid in similar concentrations (LA=1.5 ng/m³; BCN-UB and BCN-RS=0.9 ng/m³) as Kleindienst et al. (2012) and showing good correlations with phthalic acid ($r^2=0.8$ in LA; 0.7 in BCN-UB and 0.4 in BCN-RS). In the 12 h sampling resolution in BCN, the daytime samples also showed higher concentrations than the nighttime samples, indicating the formation of these products in the city. Kleindienst et al. (2012) discussed the application of phthalic acid (and isomers) for the source apportionment of SOC contribution from PAH emissions (i.e., naphthalene oxidation). The temporal multiday trend in LA as well as the similar concentrations in two sites in BCN ($r^2=0.7$; slope=1.1) agrees with their statement that primary emissions of phthalic acid are probably very small in comparison to secondary formation. Nevertheless, the weaker regression coefficients in BCN compared with the one in LA as well as the smaller slopes in BCN between 2-methylphthalic acid vs. phthalic acid (0.29 in LA and 0.18 in BCN), may indicate that the apportion of primary phthalic acid is higher in BCN than in LA. Conversely, in the two locations phthalic acid did not show any difference between weekend and weekdays and highest multiday concentrations were observed during regional circulation of air masses (e.g., between 2 and 6 June in LA and between 15 and 17 October in BCN (Fig. 1) (Dall'Osto et al. 2013; Hayes et al. 2013; Thompson et al. 2012)).

These results indicate that most of the detected phthalic acid (and other DCAs, such as succinic acid, glutaric acid and malic acid) were SOC, which were relatively rapidly formed in the atmosphere after oxidation of precursors. Under stagnant atmospheric condition these compound accumulate in the urban atmosphere. This could lead to relatively higher day than nighttime levels, but not necessarily to higher weekday concentrations.

Longer-chained DCAs (C₇–C₉), i.e., pimelic acid, suberic acid, and azelaic acid, seemed to behave differently from the shorter DCAs and their isomers (including aromatic DCAs). For example, the C₇–C₉ concentrations were two times higher in the urban center of BCN-RS compared with the background site BCN-UB. In all sites, azelaic acid concentrations were always higher than pimelic acid and suberic acid (Table 1), with good correlations between the temporal trends ($r^2\sim 0.7$). In BCN-RS the concentrations of these compounds showed a clear daytime maximum and nighttime minimum, while this was less clear in BCN-UB. The presence of azelaic acid in ambient PM_{2.5} has been related to fast oxidation of unsaturated fatty acids, i.e., oleic acid, by ozone (Moise and Rudich 2002). Oleic acid itself may have many sources in the urban atmosphere, such as food cooking, traffic and non-fossil sources, including the marine environment (Schauer et al. 2002). In the present study, this possible precursor for azelaic acid was detected only in low concentrations (<2 ng/m³; including in field blanks). It is very well possible that the C₇–C₉ DCAs are related to rapid oxidation processes within the urban atmosphere that are not yet well characterized. An indication for this hypothesis is the high concentrations in the city center (BCN-RS, 6.1 ng/m³) in comparison to the background site in BCN (BCN-UB, 2.9 ng/m³). Nevertheless, azelaic acid itself is also susceptible to relatively fast oxidation in the presence of OH radicals to form shorter chained DCAs. As discussed in detail by Alier et al. (2013), in BCN there was a linkage of these compounds and primary organics typically emitted by gasoline engine vehicles as well as air mass dependency, with higher concentrations under low-nitrate conditions.

Biogenic SOC tracers for α -pinene

The biogenic SOC tracers for α -pinene oxidation, which includes *cis*-pinonic acid, 3-hydroxyglutaric acid (3-HGA)

and MBTCA, had average concentrations of 5, 6, and 1 ng/m³ in LA, 15, 5, and 6 ng/m³ in BCN-UB, and 8, 3, and 4 ng/m³ in BCN-RS, respectively (Table 1). The LA levels were in the lower range as those observed in July–August in Riverside, situated east within the LA Basin ($\sum_{\alpha\text{-pinene SOA}} \sim 38\text{--}60$ ng/m³) (Stone et al. 2009), which may indicate seasonal dependence of the concentrations of these compounds, since the present study was conducted in May–June. On the other hand the present concentrations were very similar to the ones measured during June–July in urban sites in Marseille, along the Mediterranean coast in France (15, 4, and 5 ng/m³, for *cis*-pinonic acid, 3-HGA, and MBTCA, respectively; El Haddad et al. 2011). The higher concentrations in BCN-UB could be caused by the proximity of this site the coniferous forest, whereas the lower *cis*-pinonic concentrations in BCN-RS may reflect further oxidation of this compound in the urban atmosphere. Studies performed in forested areas showed slightly higher levels of MBTCA, such as K-Pusztá, Hungary (12 ng/m³) (Kourtchev et al. 2009), and Jülich, Germany (7 ng/m³) (Kourtchev et al. 2008). The α -pinene SOC tracers showed different diurnal and multiday trends. While 3-HGA and MBTCA were highly correlated in the two locations ($r^2=0.8$ in LA and 0.7 in BCN), *cis*-pinonic acid did not show any correlation with 3-HGA and MBTCA. *Cis*-pinonic acid is a first-generation product of α -pinene oxidation, whereas MBTCA is formed by OH-initiated oxidation of *cis*-pinonic acid (Szmigielski et al. 2007) and 3-HGA is also thought to be a further-generation oxidation product (Claeys et al. 2007). Therefore, this possibly explains why there was no correlation observed between *cis*-pinonic acid and the other two oxidation products. Despite the substantial correlations between 3-HGA and MBTCA in both sites, the ratio between the compounds in LA was 5.1 ± 1.7 and 0.8 ± 0.4 in BCN. It is unclear what mechanisms are behind these differences, but it is possible the time of sampling (May–June for LA; September–October for BCN) influences these ratios.

SOC tracers for isoprene

The SOC tracers for isoprene, which includes C₅-alkene triols, 2-methylglyceric acid (2-MGA), 2-methylthreitol and 2-methylerythritol (2-MT), had average concentrations of 6, 6, 4, and 9 ng/m³ in LA, 1, 2, 3, and 7 ng/m³ in BCN-UB and 1, 2, 1, and 4 ng/m³ in BCN-RS, respectively (Table 1). Similar concentrations were found for Lewandowski et al. (2013) in a study done also at the CalNex campaign for 2-methylglyceric acid (3 ng/m³), 2-methylthreitol (1 ng/m³), and 2-methylerythritol (3 ng/m³). C₅-alkene triols are known isoprene SOC tracers and have been reported at levels around 3 ng/m³ in other European sites (Kourtchev et al. 2008) and at very high levels around 500 ng/m³ in sites located in the southeastern USA (Lin et al. 2013a). Similarly high concentrations of 2-MT (570 ng/m³) were observed in southeastern

U.S. during late summer (Lin et al. 2013a). 2-MT and C₅-alkene triols are formed by the photooxidation (i.e., OH-initiated oxidation) of isoprene under NO-limited conditions. 2-MGA is formed by the further oxidation of volatile methacryloylperoxynitrate (MPAN), which is a major second-generation product of isoprene oxidation under initially high-NO conditions (Surratt et al. 2010). Recent work by Lin et al. (2013b) found that 2-MGA is directly formed from methacrylic acid epoxide, which is a gas-phase oxidation product of MPAN. The 2-MGA concentrations were in the same range as those observed in Marseille, France (El Haddad et al. 2011), while the levels of 2-MT were higher in the present study. In the present study, both locations had 2-methylerythritol concentrations that were ~2.5 times higher than 2-methylthreitol concentrations, but the two compounds had the same temporal trend ($r^2=1$). This ratio between these isomers of the 2-methyltetrols has also been observed in other studies (El Haddad et al. 2011; Ion et al. 2005).

Source apportionment of WSOC

MCR-ALS was applied to the individual and column-wise augmented data matrices (see “Data arrangement”) in order to investigate the similarities and differences of the analyzed WSOCs in both LA and BCN. In all cases, non-negativity and loadings normalization constraints were applied. Loading normalization allow the comparison of the compounds inside one component and also between different components. The total explained variances for the individual analyses were 98.7, 97.7, and 97.8 % for LA, BCN-UB, and BCN-RS sites, respectively, and 95.9 % for the augmented data matrix, when the three sites were analyzed simultaneously. This latter analysis emphasizes the composition of the analyzed compounds that the three stations had in common. The explained variance of the obtained components was normalized to 100 % in order to compare the contributions (score values) of the components to the WSOC in the three sites.

Five consistent components (Fig. 2) were identified in the three sites for the individual and augmented database analysis. The five resolved components could be described as potential WSOC sources: one from regional biomass burning and the other four from oxidation processes (SOC). A primary urban organic carbon (POC) component, related to traffic, was not obtained in this study. However, previous studies from the same sampling campaigns showed that 12 % of the organic matter in LA was attributed to primary emissions that were related to traffic, whereas this was 18 % in the urban background site (BCN-UB) and 43 % in the traffic intensive road site (BCN-RS) (Alier et al. 2013; Hayes et al. 2013). These percentages were applied here in order to estimate the contribution of the obtained WSOC components to the mean OC concentration. The mean OC concentration in LA was 3.3 $\mu\text{g}/\text{m}^3$, while this was 2.1 $\mu\text{g}/\text{m}^3$ in BCN-UB and 3.5 $\mu\text{g}/\text{m}^3$ in

BCN-RS. The OC concentrations versus the MCR-ALS sum of scores were correlated in both locations ($r^2=0.82$ in LA and 0.24 in BCN ($p<0.01$)). The lower correlation in BCN was due to the relatively high POC to OC contributions. Hence, the correlation coefficient improved when the POC was included ($r^2=0.75$). The correlation coefficient in LA was high and remained high if the POC contributions were added. These findings suggest that the obtained WSOC components in this study together with the contributions of POC from previous studies covered most of the OC in the two locations. In this way it was estimated that about 88 % of the OC was attributed to WSOC ($2.8 \mu\text{g OC}/\text{m}^3$), while this was 82 % in BCN-UB ($1.7 \mu\text{g OC}/\text{m}^3$) and 57 % in BCN-RS ($2.0 \mu\text{g OC}/\text{m}^3$). In the following paragraphs, the five resolved WSOC components are presented and their contributions to OC are discussed.

Regional biomass burning

The first component (Fig. 2a) was related to biomass burning and represented by levoglucosan, mannosan, and galactosan in all sites. The influence of regional biomass burning on the WSOC was 17 % in LA, with a decreasing trend towards the end of the sampling period (Fig. 3). In BCN (9–16 % of WSOC), the score values increase only in the last part of the sampling period, related to the period when open fires in fields for biomass waste burning are legally allowed in the region (Fig. 3). The loadings of DCA in this component were similarly low in all sites (Fig. 2), indicating a clear separation of biomass burning contributions and those from other sources and/or processes.

These estimates showed that in LA about $0.5 \mu\text{g}/\text{m}^3$ of OC corresponded to “biomass burning organic carbon” (BBOC; 15 % of OC), while this was $0.3 \mu\text{g}/\text{m}^3$ in BCN-UB (13 % of OC) and $0.2 \mu\text{g}/\text{m}^3$ in BCN-RS (5 % of OC) (Fig. 4). The BBOC contribution in LA is higher than the <5 % estimated in aerosol mass spectrometry (AMS) analysis by Hayes et al. (2013). The low contributions in these analysis was possibly caused by a generally low biomass burning signal (i.e. low levoglucosan concentrations) in combination with the high abundance of secondary organics in the LA atmosphere.

Aged secondary organic carbon

The second component (Fig. 2b), which accounted for 28 % of the WSOC in LA and about 15 % in BCN, was composed mainly of oxygenated organic aerosol tracer compounds, such as short-chained diacids (e.g., succinic acid and glutaric acid), phthalic acid, and more oxygenated compounds from biogenic origin, such as malic acid, 3-HGA, and MTBCA.

Phthalic acid, possibly formed in several steps by oxidation of naphthalene (Kautzman et al. 2010), can be related to processed aerosols. Interestingly, in both locations, phthalic acid

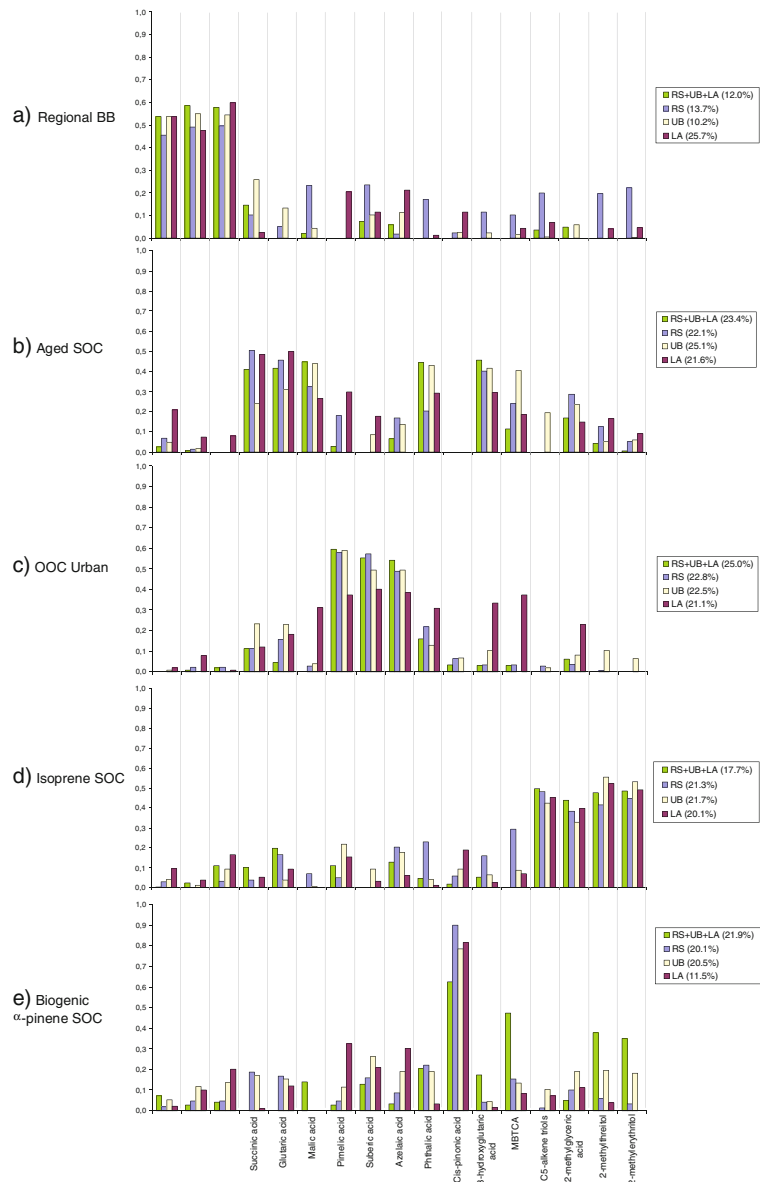
showed good correlations with MBTCA and 3-HGA ($r^2>0.5$); two products of further α -pinene oxidation. In all sites, there was a 1:1 relationship between 3-HGA and phthalic acid. Conversely, MBTCA showed five times lower concentrations compared with phthalic acid in LA, while in BCN the concentrations of MBTCA were slightly higher than phthalic acid concentrations. Succinic, glutaric, and malic acid also showed similar relationships between 3-HGA and MBTCA in the two locations. As mentioned before, it could be possible that the sampling performance in different seasons (late spring in LA and early fall in BCN) may have an influence on the formation pathways of α -pinene SOC formation.

Generally, the compounds represented in this component were products of further oxidation processes and can therefore be linked to processed air from urban and biogenic emissions in relation with circulation of air masses within the studied urban areas. In LA and BCN, the temporal trend of the Aged SOC component (Fig. 3) was very similar to the trend of the low volatile oxygenated organic aerosols (LV-OOA) measured by AMS (Hayes et al. 2013), indicating the high oxidation state of these aerosols. Moreover, at both sites, the highest abundance of this component (Fig. 3) coincides with generally high aerosol loadings during the sampling campaigns, between 3–5 Oct and 14–16 Oct in BCN, and 4–7 June in LA (Dall’Osto et al. 2013; Hayes et al. 2013; Thompson et al. 2012). This component contributed to about 30 % of the OC ($0.8 \mu\text{g OC}/\text{m}^3$), whereas this was about 8–13 % ($0.3 \mu\text{g OC}/\text{m}^3$) in BCN (RS and UB, respectively) (Fig. 4).

Oxidized organic carbon of mainly urban origin

The third component (Fig. 2c), which accounted for 16 % of the WSOC in LA, 19 % in BCN-UB and 50 % in BCN-RS, was mainly composed of longer-chained DCA (C_7 – C_9). There was a clear separation between these longer-chained DCAs and shorter ones (succinic acid and glutaric acid) in the augmented data base, while this separation was less clear for the phthalic acid, malic acid, 3-HGA, and MBTCA in the case of LA when the database was analyzed separately (Fig. 2c). The presence of these oxidized compounds in this component in LA may be related to a more equilibrated influence of potential sources/processes to this component in this site compared with the sites in BCN. In the BCN-RS there was a strong day–night fluctuation of the component (Fig. 3c) and the contribution to OC ($1.0 \mu\text{g OC}/\text{m}^3$) was significantly higher than the one in BCN-UB ($0.3 \mu\text{g OC}/\text{m}^3$) (Fig. 4), indicating that this component could be related to specific in-site urban oxidation processes, which may include cooking activities (see “DCA, hydroxy-DCA, and aromatic-DCA”). In the urban background site of BCN-UB this day–night fluctuation is less pronounced (Fig. 3c), whereas in LA the major compounds (C_7 – C_9 DCAs) mix up with the “Aged SOC” compounds (Fig. 2b, 3b). Therefore, it seems that the “oxidized organic

Fig. 2 MCR-ALS loadings of components for individual analysis in BCN-RS, BCN-UB, LA, and simultaneous analysis (RS+UB+LA) with the relative contribution to the overall variation in the legends. **a** Regional biomass burning, **b** aged secondary organic carbon, **c** oxygenated organic carbon urban, **d** isoprene secondary organic carbon, and **e** biogenic α -pinene secondary organic carbon



carbon urban” component was formed after rapid oxidation in the urban atmosphere and then accumulates in the background areas. However, the exact precursor and formation pathway remains unclear. In a previous study in BCN (Alier et al. 2013), the abundance of this component was related to primary urban emissions (i.e. higher molecular weight PAH), but there was also an air mass dependence.

In LA, the temporal trend of this component showed similarities with the cooking influenced organic aerosol, although it also correlated with the LV-OOA component (Hayes et al. 2013). Therefore, further study is necessary to elucidate the origin of this component.

Isoprene SOC

The fourth component (Fig. 2d) accounted for about 36 % of the WSOC in LA and 6–11 % in BCN (RS and UB, respectively), and was composed of the products of isoprene oxidation; specifically, the C₅-alkene triols, 2-methyltetrols, and 2-methylglyceric acid.

Overall, there were substantial correlations between C₅-alkene triols and the 2-MT ($r^2=0.8$ in BCN and 0.7 in LA). A similar good correlation was observed for 2-MGA in BCN ($r^2=0.6$ and 0.7, respectively), however weaker in LA ($r^2=0.3$ and 0.5, respectively). Although the results in this study

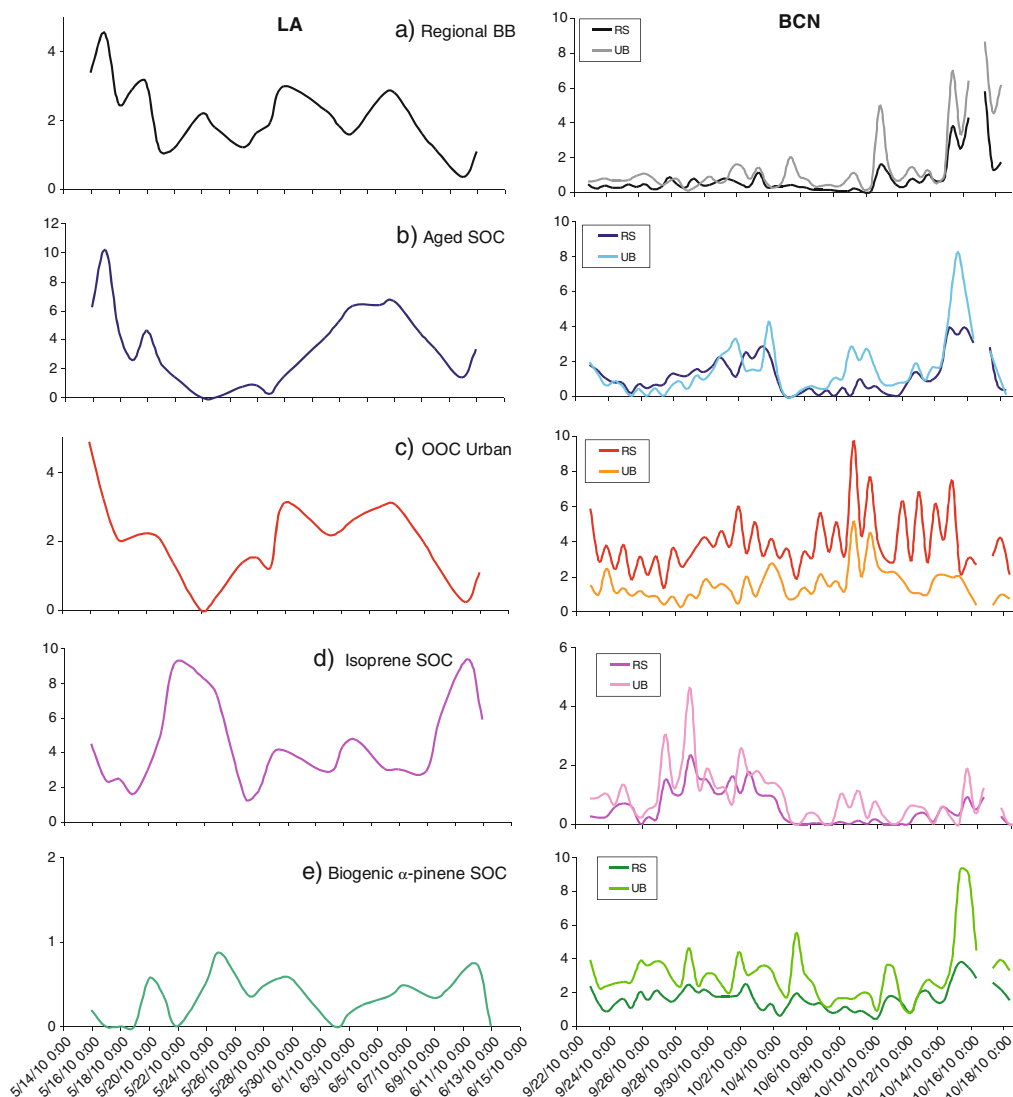


Fig. 3 Simultaneous MCR-ALS score resolved profiles for the LA (*left*) and BCN sites (*right*) using non-negativity constraints

suggest that the isoprene oxidation products have similar origins, the weaker correlations with 2-MGA points to different transformation pathway, which were discussed in detail in “SOC tracers for isoprene”: 2-MT is formed under NO-limited conditions. Despite the slightly lower NO concentration in LA ($4 \mu\text{g}/\text{m}^3$) compared with the ones in BCN (5 and $8 \mu\text{g}/\text{m}^3$ in UB and RS, respectively), there was no correlation between 2-MT and NO (Dall’Osto et al. 2013; Pollack et al. 2012). Overall, the temporal trends of this component (Fig. 2d) does not coincident with the highest bulk load for PM and organic aerosols (Dall’Osto et al. 2013; Thompson et al. 2012) nor with any of the AMS components obtained by Hayes et al. 2013. This may indicate that these secondary organics from isoprene oxidation probably do not contribute substantially to

the urban OC here. However, the substantial contribution of 31 % to OC of this component in LA ($1.0 \mu\text{g OC}/\text{m}^3$; Fig. 4) suggest the contrary, that at least in this site the influence of isoprene oxidation is important.

Biogenic α -pinene SOC

The fifth component (Fig. 2e) was composed of *cis*-pinonic acid, and contributed to 4 % of the WSOC in LA, whereas this was 21–37 % in BCN (RS and UB, respectively). The separation of this compound from all the other analyzed compounds indicates that a unique source or process and specific conditions must be involved.

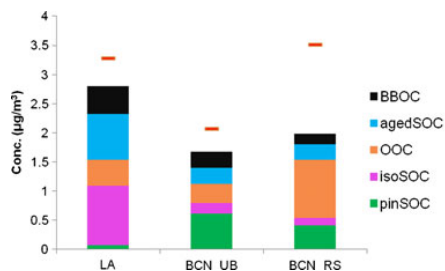


Fig. 4 Mean contributions of WSOC components (micrograms per cubic meter) to OC for the studied periods in LA, BCN-UB, and BCN-RS. The red line above the columns indicates the mean OC concentration (micrograms per cubic meter)

The locations investigated in the present study did not show correlations between the concentrations of *cis*-pinonic acid and 3-HGA or MBTCA ($r^2 < 0.2$). The absence of these correlations can be explained by the different formation modes of these compounds as mentioned in “Biogenic SOC tracers for α -pinene”: *cis*-pinonic acid is a first-generation product of pinene oxidation, while MBTCA and 3-HGA are probably formed from further reaction of *cis*-pinonic acid in the presence of NO_x . However, no correlations were observed between these parameters in any of the sites. Nevertheless, the ratio between *cis*-pinonic acid and MBTCA showed peak ratios (>10) in BCN on 25 September and 5 October, when urban new particle formation events were detected (Dall’Osto et al. 2013). *Cis*-pinonic acid has been linked to nucleation processes in forested areas as a first step in the formation of aerosols from organic vapors (O’Dowd et al. 2002; Laaksonen et al. 2008). In LA, the ratio between *cis*-pinonic acid and MBTCA were highest (>10) between 24 and 29 May, coinciding with a period of generally low aerosol loadings (Hayes et al. 2013).

Implications and conclusions

GC/MS with prior derivatization allows for the identification of several known organic tracer compounds for WSOC in PM filter samples. Moreover, the MCR-ALS was successfully applied on the WSOC dataset from LA and BCN, obtaining five consistent components in these urban atmospheres. The estimated contributions of WSOC to OC were 88 % in LA, 82 % in BCN-UB and 57 % in BCN-RS. In both locations, the biomass burning contributions to WSOC were low but could be separated from components that were related to oxidation processes, from biogenic and anthropogenic origin. These components made up about two thirds of the OC is related to SOC in LA and the urban background site in BCN, while this was half of the OC in BCN road site. Fresh biogenic SOA species, such as *cis*-pinonic from α -pinene oxidation and 2-methyltetrols from the oxidation of isoprene, showed different temporal concentration trends in comparison to further-

generation oxidation products (aged SOC), including those from biogenic emissions as well as vehicle emissions. Isoprene oxidation was more important in LA than in BCN. Although this could be related to the different periods of sampling, the influence of oxidants, such as NO_x , can also play a role and should be studied in more detail. The aged products from biogenic and anthropogenic origin clustered in another component that was strongly related to the organic aerosol loadings in the studied areas. Interestingly, the component represented by longer-chained C_7 – C_9 DCA that are probably related to urban activities and in situ oxidation processes, may also involve specific atmospheric conditions. This later component was dominant in the road site in BCN, and further study should elucidate the role of primary sources (cooking and traffic), secondary aerosol processing, and atmospheric conditions on this component.

Acknowledgments The authors thank Xavier Querol, Manuel Dall’Osto, Joost de Gouw, Jochem Stutz, Jason Surratt, John Seinfeld, and Jose Luis Jimenez, the organizers of the SAPUSS and CalNex-LA campaigns, for their service to the community. Technical assistance from R. Chalier and D. Fanjul is acknowledged. Financial support for this study was provided by projects: CTQ2009-11572, SAPUSS (FP7-PEOPLE-2009-IEF, Project number 254773) and AEROTRANS (CTQ2009-377 14777-C02-01).

References

- Alier M, van Drooge BL, Dall’Osto M, Querol X, Grimalt JO, Tauler R (2013) Source apportionment of submicron organic aerosol at an urban background and a road site in Barcelona (Spain) during SAPUSS. *Atmos Chem Phys* 13:10353–10371
- Bahreini R, Middlebrook AM, de Gouw JA, Wameke C, Trainer M, Brock CA, Stark H, Brown SS, Dube WP, Gilman JB, Hall K, Holloway JS, Kuster WC, Perring AE, Prevot ASH, Schwartz JP, Spackman JR, Szidat S, Wagner NL, Weber RJ, Zotter P, Parrish DD (2012) Gasoline emissions dominate over diesel in formation of secondary organic aerosol mass. *Geophys Res Lett* 39, L06805
- Claeys M, Graham B, Vas G, Wang W, Vermeylen R, Pashynska V, Cafmeyer J, Guyon P, Andreae MO, Artaxo P, Maenhaut W (2004) Formation of secondary organic aerosols through photooxidation of isoprene. *Science* 303:1173–1176
- Claeys M, Szmigielski R, Kourtev, Van der Veken P, Vermeylen R, Maenhaut W, Jaoui M, Kleindienst TE, Lewandowski M, Offenberger JH, Edney EO (2007) Hydroxydicarboxylic acids: markers for secondary organic aerosol from the photooxidation of α -pinene. *Environ Sci Technol* 41:1628–1634
- Dall’Osto M, Querol X, Alastuey A, Minguillon MC, Alier M, Amato F, Brines M, Cusak M, Grimalt JO, Karanasiou A, Moreno T, Pandolfi M, Pey J, Reche C, Ripoll A, Tauler R, Van Drooge BL, Viana M, Harrison RM, Gietl J, Beddows D, Bloss W, O’Dowd C, Ceburnis D, Martucci G, Ng S, Worsnop D, Wenger J, Mc Gillcuddy E, Sudou J, Healy R, Lucarelli F, Nava S, Jimenez JL, Gomez Moreno F, Artinano B, Prevot AS, Pfaffenberger L, Frey S, Wilsenack F, Casabona D, Jimenez-Guerrero P, Gross D, Cotz N (2013) Presenting SAPUSS: Solving aerosol problem by using synergistic strategies at Barcelona, Spain. *Atmos Chem Phys* 13: 8991–9019

- El Haddad I, Marchand N, Temime-Roussel B, Wortham H, Piot C, Besombes JL, Baduel C, Voisin D, Armengaud A, Jaffrezo JL (2011) Insights into the secondary fraction of the organic aerosol in a Mediterranean urban area: Marseille. *Atmos Chem Phys* 11: 2059–2079
- Fine PM, Cass GR, Simoneit BRT (2004) Chemical characterization of fine particle emissions from the wood stove combustion of prevalent United States tree species. *Environ Eng Sci* 21:705–721
- Hallquist M, Wenger JC, Baltensperger U, Rudich Y, Simpson D, Claeys M, Dommen J, Donahue NM, George C, Goldstein AH, Hamilton JF, Herrmann H, Hoffmann T, Iinuma Y, Jang M, Jenkin ME, Jimenez JL, Kiendler-Scharr A, Maenhaut W, McFiggans G, Mentel TF, Monod A, Prévôt ASH, Seinfeld JH, Surratt JD, Szmigielski R, Wildt J (2009) The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmos Chem Phys* 9:5155–5236
- Hayes PL, Ortega AM, Cubison MJ, Froyd KD, Zhao Y, Cliff SS, Hu WW, Toohey DW, Flynn JH, Lefter BL, Grossberg N, Alvarez S, Rappenglück B, Taylor JW, Allan JD, Holloway JS, Gilman JB, Kuster WC, de Gouw JA, Massoli P, Zhang X, Liu J, Weber RJ, Corrigan AL, Russell LM, Isaacman G, Worton DR, Kreisberg NM, Goldstein AH, Thalman R, Waxman EM, Volkamer R, Lin YH, Surratt JD, Kleindienst TE, Offenberg JH, Dusanter S, Griffith S, Stevens PS, Brioude J, Angevine WM, Jimenez JL (2013) Organic aerosol composition and sources in Pasadena, California during the 2010 CalNex Campaign. *J Geophys Res Atmos* 118:1–25
- Heald CL, Kroll JH, Jimenez JL, Docherty KS, DeCarlo PF, Aiken AC, Chen Q, Martin ST, Farmer DK, Artaxo P, Weinheimer AJ (2010) A simplified description of organic aerosol composition and implications for atmospheric aging. *Geophys Res Lett* 37, L08803
- Hennigan CJ, Miracolo MA, Engelhart GJ, May AA, Presto AA, Lee T, Sullivan AP, McMeeeking GR, Coe H, Wold CE, Hao WM, Gilman JB, Kuster WC, de Gouw J, Schichtel BA, Jr JC, Kreidenweis SM, Robinson AL (2011) Chemical and physical transformations of organic aerosol from the photo-oxidation of open biomass burning emissions in an environmental chamber. *Atmos Chem Phys* 11: 7669–7686
- Hersey SP, Craven JS, Schilling KA, Metcalf AR, Sorooshian A, Chan MN, Flagan RC, Seinfeld JH (2011) The Pasadena aerosol characterization observatory (PACO): chemical and physical analysis of the western Los Angeles Basin aerosol. *Atmos Chem Phys* 11: 7417–7443
- Hoffmann D, Tilgner A, Iinuma Y, Herrmann H (2010) Atmospheric stability of levoglucosan: a detailed laboratory and modeling study. *Environ Sci Technol* 44:694–699
- Ion AC, Vermeylen R, Kourtev I, Cafmeyer J, Chi X, Gelencsér A et al (2005) Polar organic compounds in rural PM_{2.5} aerosols from K-puszta, Hungary, during a 2003 summer field campaign: sources and diel variations. *Atmos Chem Phys* 5:1805–1814
- Jaumot J, Gargallo R, de Juan A, Tauler R (2005) A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom Intell Lab Syst* 76:101–110
- Jimenez JL, Canagaratna MR, Donahue NM et al (2009) Evolution of organic aerosols in the atmosphere. *Science* 326:1525–1529
- Kautzman KE, Surratt JD, Chan MN, Chan AWH, Hersey SP, Chabra PS, Dalleska NF, Wennberg PO, Flagan RC, Seinfeld JH (2010) Chemical composition of gas- and particle-phase products from photooxidation of naphthalene. *J Phys Chem* 114:913–934
- Kleindienst TE, Jaoui M, Lewandowski M, Offenberg JH, Docherty KS (2012) The formation of SOA and chemical tracer compounds from the photooxidation of naphthalene and its methyl analogs in the presence and absence of nitrogen oxides. *Atmos Chem Phys* 12: 8711–8726
- Kourtev I, Warnke J, Maenhaut W, Hoffmann T, Claeys M (2008) Polar organic marker compounds in PM_{2.5} aerosol from a mixed forest site in western Germany. *Chemosphere* 73:1309–1314
- Kourtev I, Copolovici L, Claeys M, Maenhaut W (2009) Characterization of aerosols at the forested site in central Europe. *Environ Sci Technol* 43:4665–4671
- Laaksonen A, Kulmala M, O'Dowd CD, Joutsensaari J, Vaattovaara P, Mikkonen S, Lehtinen KEJ, Sogacheva L, Dal Maso M, Aalto P, Petäjä T, Sogachev A, Yoon YJ, Lihavainen H, Nilsson D, Facchini MC, Cavalli F, Fuzzi S, Hoffmann T, Arnold F, Hanke M, Sellegri K, Umann B, Junkermann W, Coe H, Allan JD, Alfarra MR, Worsnop DR, Riekkola ML, Hyötyläinen T, Viisanen Y (2008) The role of VOC oxidation products in continental new particle formation. *Atmos Chem Phys* 8:2657–2665
- Lewandowski M, Piletic IR, Kleindienst TE, Offenberg JH, Beaver MR, Jaoui M, Docherty KS, Edney EO (2013) Secondary organic aerosol characterisation at field sites across the United States during the spring–summer period. *Int J Environ Anal Chem*. doi:10.1080/03067319.2013.803545
- Lin YH, Knipping EM, Edgerton ES, Shaw SL, Surratt JD (2013a) Investigating the influences of SO₂ and NH₃ levels on isoprene-derived secondary organic aerosol formation using conditional sampling approaches. *Atmos Chem Phys Discuss* 13:3095–3134
- Lin YH, Zhang H, Pye HOT, Zhang Z, Marth WJ, Park S, Arashiro M, Cui T, Budisulistiorini SH, Sexton KG, Vizuete W, Xie Y, Luecken DJ, Piletic IR, Edney EO, Bartolotti LJ, Gold A, Surratt JD (2013b) Epoxide as a precursor to secondary organic aerosol formation from isoprene photooxidation in the presence of nitrogen oxides. *Proc Natl Acad Sci U S A* 110:6718–6723
- Minguillón MC, Perron N, Querol X, Szidat S, Fahmi SM, Alastuey A, Jimenez JL, Mohr C, Ortega AM, Day DA, Lanz VA, Wacker L, Reche C, Cusack M, Amato F, Kiss G, Hoffer A, Decesari S, Moretti F, Hillamo R, Teinil K, Seco R, Peñuelas J, Metzger A, Schallhart S, Müller M, Hansel A, Burkhardt JF, Baltensperger U, Prevot ASH (2011) Fossil versus contemporary sources of fine elemental and organic carbonaceous particulate matter during the DAURE campaign in Northeast Spain. *Atmos Chem Phys* 11:12067–12084
- Mohr M, DeCarlo PF, Heringa MF, Chirico R, Slowik JG, Richter R, Reche C, Alastuey A, Querol X, Seco R, Peñuelas J, Jimenez JL, Crippa M, Zimmermann R, Baltensperger U, Prevot ASH (2012) Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data. *Atmos Chem Phys* 12:1649–1665
- Moise T, Rudich Y (2002) Reactive uptake of ozone by aerosol-associated unsaturated fatty acids: kinetics, mechanism, and products. *J Phys Chem A* 106(27):6469–6476
- O'Dowd CD, Aalto P, Hämeri K, Kulmala M, Hoffmann T (2002) Atmospheric particles from organic vapours. *Nature* 416:497
- Paulot F, Wunch D, Crouse JD, Toon GC, Millet DB, DeCarlo PF, Vigouroux C, Deutscher NM, González Abad G, Notholt J, Warneke T, Hannigan JW, Arneke C, de Gouw JA, Dunlea EJ, De Mazière M, Griffith DWT, Bernath P, Jimenez JL, Wennberg PO (2011) Importance of secondary sources in the atmospheric budgets of formic and acetic acids. *Atmos Chem Phys* 11:1989–2013
- Pollack IB, Ryerson TB, Trainer M et al (2012) Airborne and ground-based observations of a weekend effect in ozone, precursors, and oxidation products in the California South Coast Air Basin. *J Geophys Res* 117, D00V05. doi:10.1029/2011JD016772
- Puxbaum H, Caseiro A, Sánchez-Ochoa A, Kasper-Giebl A, Claeys M, Gelencsér A, Legrand M, Preunkert S, Pio C (2007) Levoglucosan levels at background sites in Europe for assessing the impact of biomass combustion on the European aerosol background. *J Geophys Res* 112:D23S05
- Reche C, Moreno T, Amato F, Viana M, van Drooge BL, Chuang H, Bérubé K, Jones T, Alastuey A, Querol X (2012) A multidisciplinary approach to characterise exposure risk and toxicological effects of PM₁₀ and PM_{2.5} samples in urban environments. *Ecotoxicol Environ Safe* 78:327–335

- Ryerson TB, Andrews AE, Angevine WM, Bates TS, Brock CA, Cairns B, Cohen RC, Cooper OR, de Gouw JA, Fehsenfeld FC, Ferrare RA, Fischer ML, Flagan RC, Goldstein AH, Hair JW, Hardesty RM, Hostetler CA, Jimenez JL, Langford AO, McCauley E, McKeen SA, Molina LT, Nenes A, Oltmans SJ, Parrish DD, Pederson JR, Pierce PK, Quinn PK, Seinfeld JH, Senff CJ, Sorooshian A, Stutz J, Surratt JD, Trainer M, Volkamer R, Williams EJ, Wofsy SC (2013) The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex) field study. *Am Geophys Union*. doi:10.1002/jgrd.50331
- Schauer JJ, Kleemann MJ, Cass GR, Simoneit BRT (2002) Measurements of emissions from air pollution sources, 5. C1–C32 organic compounds from gasoline-powered motor vehicles. *Environ Sci Technol* 36:1169–1180
- Simoneit BRT (2002) Biomass burning—a review of organic tracers for smoke from incomplete combustion. *Appl Geochem* 17:129–162
- Schmidt C, Marr IL, Caseiro A, Kotianová P, Berner A, Bauer H, Kasper-Geibl PH (2008) Chemical characterisation of fine particle emissions from wood stove combustion of common woods growing in mid-European Alpine regions. *Atmos Environ* 42:126–141
- Stone EA, Zhou J, Snyder DC, Rutter AP, Mieritz M, Schauer JJ (2009) A comparison of summertime secondary organic aerosol source contributions at contrasting urban locations. *Environ Sci Technol* 43:3448–3454
- Surratt JD, Chan AWH, Eddingsaas NC, Chan MN, Loza CL, Kwan AJ, Hersey SP, Flagan RC, Wennberg PO, Seinfeld JH (2010) Reactive intermediates revealed in secondary organic aerosol formation from isoprene. *Proc Natl Acad Sci U S A* 107:6640–6645
- Szmigielski R, Surratt JD, Gómez-González Y, Van der Veken P, Kourtchev I, Vermeylen R et al (2007) 3-methyl-1,2,3-butanetricarboxylic acid: an atmospheric tracer for terpene secondary organic aerosol. *Geophys Res Lett* 34, L24811. doi:10.1029/2007GL03133
- Tauler R, Smilde A, Kowalski B (1995) Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *J Chemom* 9:31–58
- Tauler R, Viana M, Querol X, Alastuey A, Flight RM, Wentzell PD, Hopke PK (2009) Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies. *Atmos Environ* 43:3989
- Thompson JE, Hayes PL, Jimenez JL, Adachi K, Zhang X, Lui J, Weber RJ, Buseck PR (2012) Aerosol optical properties at Pasadena, CA during CalNex 2010. *Atmos Environ* 55:190–200
- van Drooge BL, Cusack M, Reche C, Mohr C, Alastuey A, Querol X, Prevot A, Day DA, Jimenez JL, Grimalt JO (2012) Molecular marker characterization of the organic composition of submicron aerosols from Mediterranean urban and rural environments under contrasting meteorological conditions. *Atmos Environ* 61:482–489
- Warneke C, de Gouw JA, Holloway JS, Peischl J, Ryerson TB, Atlas E, Blake D, Trainer M, Parrish DD (2012) Multiyear trends in volatile organic compounds in Los Angeles, California: five decades of decreasing emissions. *J Geophys Res* 117, doi:10.1029/2012JD017899
- Yang L, Ray MB, Yu LE (2008) Photooxidation of dicarboxylic acids—part II: kinetics, intermediates and field observations. *Atmos Environ* 42:868–880

5.2.1 DISCUCCIÓ DE RESULTATS

En aquesta secció es presenta la discussió general dels dos articles continguts en aquest bloc. Aquests resultats s'agruparan segons dos objectius:

- a) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Barcelona
- b) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Los Angeles i comparació amb els resultats obtinguts a Barcelona

a) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Barcelona

En aquest estudi s'han analitzat filtres de material particulat de mida sub-micrònica (PM_{10}), en dues localitzacions de la ciutat de Barcelona. S'ha realitzat una extracció i un posterior anàlisi amb cromatografia de gasos acoblada a espectrometria de masses, com ja s'ha explicat amb detall a la secció 3.2 de la present Tesi.

A la Taula 1 de l'Article 4 es mostren les concentracions mitjanes, desviacions estàndard, màximes i mínimes del conjunt dels diferents dies i compostos analitzats durant aquest estudi. Els compostos s'han agrupat per famílies: hidrocarburs aromàtics policíclics, hopans, alcaloid, sucres anhidres i àcids, poliols o metiltetrols.

Les concentracions dels 12 PAH analitzats van ser dues vegades més alts a la estació RS que a l'estació UB, reflectint una forta influència dels residus provinents dels vehicles amb fuels fòssils. La correlació positiva entre

les dues localitzacions ($r^2=0.55$) UB i RS, tot i la diferència de concentracions, reflecteix les condicions atmosfèriques que afecten a la distribució de contaminants atmosfèrics primaris en l'àrea estudiada. En la estació RS, els PAH detectats amb alt pes molecular ($m/z > 228$) van mostrar concentracions més altes durant el dia que durant la nit, mentre que aquesta diferència no es va veure entre els PAH de baix pes molecular. En la estació UB, no es van detectar diferències dia/nit pels PAH detectats. Els factors responsables d'aquesta variació dia/nit en l'estació RS poden ser la intensitat del tràfic, la flota de vehicles (dièsel o benzina) o les condicions meteorològiques (com la temperatura). El sumatori de les concentracions de PAH a Barcelona es va trobar en el mateix interval que els observats en altres àrees urbanes dels EEUU (Ning, 2007; Subramanian, 2006). No obstant, a Barcelona, la composició relativa dels compostos PAH en l'aire estava enriquida amb els PAH de baix pes molecular, mentre que àrees urbanes dels EEUU estaven més enriquides en els PAH d'elevat pes molecular. Aquesta diferència entre dues localitzacions urbanes pot ser degut a l'elevada proporció de vehicles amb motor dièsel a Barcelona, en comparació amb les localitzacions a EEUU, on dominen els vehicles de benzina (Rogge, 1993c; Zielinska, 2004).

En relació als hopans, les concentracions trobades van ser similars a les dues estacions, mostrant la forta influència de la combustió de fuels fòssils dels vehicles en la zona estudiada. La relació entre els PAH de baix pes molecular amb els hopans, reflecteix les emissions primàries dels vehicles dièsel. No es van trobar diferències entre els períodes dia i nit, però sí es van observar concentracions més elevades entre els dies laborables en comparació amb els caps de setmana o vacances.

Les concentracions de nicotina van ser molt més elevades a la estació RS en comparació a l'estació UB. Això pot ser degut a que l'estació RS està molt propera a un carrer amb molt moviment de vehicles i persones i també d'una sortida del metro. És important mencionar que Espanya es troba entre els països on el consum de cigarretes és més elevat dins la Unió Europea i al món (WHO, 2010). Es van trobar concentracions més elevades durant el dia que durant la nit, i també durant els dies laborables. Aquest fet indica l'elevada relació de la nicotina amb les jornades de treball.

Els anhidrosacàrids mannosan i galactosan es van trobar en baixes concentracions en les dues estacions, tot i que la concentració de levoglucosan va ser més elevada especialment a la última part de l'estudi. La influència de la crema de biomassa és petita en aquesta regió tot i el seu augment en la última part de l'estudi, degut al començament del permís de cremar a partir del 15 d'Octubre i també degut a unes condicions anticiclòniques regionals en aquest període (DOGC, 1995; Dall'Osto, 2013).

Les concentracions dels àcids dicarboxílics es van trobar lleugerament superiors a l'estació UB en comparació a l'estació RS, amb l'excepció dels àcids dicarboxílics llargs (C₇-C₉), que van tenir concentracions més elevades a RS. Entre aquests àcids dicarboxílics llargs, el que va presentar concentracions més elevades va ser l'àcid azelaic, que està relacionat amb l'oxidació d'àcids grassos insaturats, com l'àcid oleic, per l'ozó (Gogou, 1994; Moise, 2002). L'àcid oleic en atmosferes urbanes pot provenir de diferents fonts, com el cuinat d'aliments (Fang, 2002; Robinson, 2006) o el tràfic, però la oxidació de precursors biogènics també pot ser una font important de d'àcids dicarboxílics llargs (Mochida, 2003; Zhang, 2010; Ho, 2011).

Els tres traçadors biogènics per l'aerosol orgànic secundari (SOA) de l'oxidació de l' α -pinene, àcid *cis*-pinònic, àcid 3-hidroxiglutàric i àcid 3-metil-1,2,3-butantricarboxílic (MBTCA), van mostrar concentracions més elevades a l'estació UB, però similars a les trobades per El Haddad (2011) a l'àrea urbana mediterrània francesa. L'àcid *cis*-pinònic no va presentar correlació amb els àcids 3-hidroxiglutàric i MBTCA, però el àcid 3-hidroxiglutàric i el MBTCA sí van presentar una correlació significativa entre ells. Aquest fet és degut a que la tendència temporal de l'àcid *cis*-pinònic és diferent de la obtinguda pels altres dos traçadors i el seu origen pot ser diferent (Ding, 2011; Szmigielski, 2007).

Les emissions des de la vegetació es consideren les majors fonts de compostos orgànics volàtils derivats dels isoprenoids (Guenter, 1995), però també els tubs d'escapament aporten una quantitat d'isoprè no negligible en àrees urbanes (Borbon, 2001; Park, 2011). Estudis previs en la regió metropolitana de Barcelona han descrit fonts d'isoprè antropogèniques, encara que aquestes s'han vist principalment al hivern (Seco, 2012). En aquest estudi, tots els marcadors relacionats amb el isoprè (C₅-alquetriols, àcid 2-

metilglicèric i metiltetrols) han presentat concentracions més elevades en la estació UB, encara que la estació RS és la que té més influència antropogènica. També s'ha de tenir en compte, que aquest estudi s'ha realitzat durant els mesos de setembre i octubre, amb temperatures encara bastant més elevades que al hivern.

L'anàlisi multivariant de les dades ens ha permès en aquest cas observar més fàcilment la relació que hi ha entre els diferents compostos amb mètodes com el PCA i el MCR-ALS. Com ja s'ha mencionat anteriorment, el PCA ens ha servit per veure quanta variància pot explicar el model, i el MCR-ALS ens ha facilitat la interpretació d'aquesta informació. Per tant, amb PCA es va escollir un model amb sis components i amb MCR-ALS es van interpretar els resultats obtinguts. Es va obtenir una molt bona correlació ($r^2=0.9$) entre la concentració total de carboni orgànic present en els filtres de PM_{10} i el sumatori dels *scores* del model obtingut amb MCR-ALS, indicant que els compostos orgànics seleccionats en aquest estudi són representatius de la contribució de la fracció orgànica de les mostres analitzades durant aquesta campanya. En l'apartat 3.2 de l'Article 4 es pot veure una descripció detallada dels components obtinguts, que es descriuen més breument a continuació. En la Figura 2 de l'Article 4 es mostren els sis components obtinguts amb l'anàlisi MCR-ALS aplicant restriccions de no-negativitat i normalització dels *loadings*. En la Figura 3 de l'Article 4 es mostra, per cada component, el percentatge de cada tipus d'aire (escenaris) determinats en aquesta campanya (Dall'Osto, 2013).

El primer component (Figura 2 de l'Article 4, *POA Urban (aerosol orgànic primari urbà)*) té una elevada contribució dels PAH, hopans i nicotina. La seva correlació amb la intensitat del tràfic ha estat bona i per tant es pot dir que correspon a l'aerosol orgànic primari urbà. Els nivells més alts es van trobar durant els dies de la setmana, especialment el dimecres i el dijous, mentre que els caps de setmana i festius van presentar nivells molt més baixos. Els *scores* d'aquest component estan pràcticament igual dividits entre els quatre escenaris de masses d'aire (Figura 3 de l'Article 4), suggerint que és poc dependent del

transport de masses d'aires regionals i les emissions locals són les responsables de la contaminació atmosfèrica primària en l'àrea urbana.

El segon component (Figura 2 de l'Article 4, *BBOA Reg. (aerosol orgànic per crema de biomassa regional)*) està format essencialment per levoglucosan, galactosan i manosan, que són marcadors de crema de biomassa primaris. Els valors més elevats d'aquest component es van observar en el últim període de la campanya de mostreig, coincidint amb masses d'aire regionals recirculades (Figura 3 de l'Article 4), així com amb el començament del permís de crema de matolls en l'àrea de Barcelona. La contribució trobada per la crema de biomassa va ser similar a la trobada en altres estudis a Barcelona (Reche, 2012) i també a l'anàlisi d'espècies inorgàniques realitzat en la mateixa campanya (Dall'Osto, 2013).

Els traçadors de l'oxidació de l'isoprè (C_5 -alquetriols, àcid 2-metilglicèric i metiltetrols) formen el tercer component (Figura 2 de l'Article 4, *SOA ISO (aerosol orgànic secundari derivat del isoprè)*). Les contribucions observades van ser similars a les dues localitzacions, suggerint un origen local, tot i que en el 50% dels casos es van detectar aires de masses regionals (Figura 3 de l'Article 4).

El quart component (Figura 2 de l'Article 4, *SOA BIO PIN (aerosol orgànic secundari biogènic derivat de l' α -pinene)*) està dominat principalment per l'àcid *cis*-pinònic, encara que hi ha petites contribucions dels metiltetrols i dels hopans. Els perfils temporals en les dues localitzacions UB i RS és diferent, suggerint que aquest component no influeix amb el mateix nivell en les dues localitzacions. Això pot ser degut a que l'aerosol orgànic secundari de l' α -pinene sofreix una oxidació addicional en el centre de la ciutat. La tendència temporal remarcable en aquest component es va registrar al 5 d'Octubre, que va coincidir amb una nova formació urbana de partícules que podia estar relacionada amb processos de nucleació en àrees forestals com a primer pas de la formació d'aerosols a partir de vapors orgànics (Dall'Osto, 2013; O'Dowd, 2002; Laaksonen, 2008).

En el cinquè component (Figura 2 de l'Article 4, *SOA Aged (aerosol orgànic secundari antic)*) es troben traçadors de l'aerosol orgànic secundari

tant d'origen antropogènic com biogènic com són els àcids dicarboxílics C₃-C₅, els hidroxiàcids dicarboxílics, l'àcid tricarbòllític, i en menor mesura els àcids dicarboxílics C₇-C₉ i l'àcid ftàlic. L'àcid *cis*-pinònic no està present en aquest component, suggerint que aquest component és més antic i oxidat que els anteriors. La presència més gran dels àcids dicarboxílics curts (C₃-C₅) en relació als àcids dicarboxílics més llargs (C₇-C₉), també indica el caràcter més antic d'aquest component, ja que els àcids dicarboxílics amb elevat nombre de carbonis tendeixen a degradar-se fotoquímicament a àcids dicarboxílics més petits (Yang, 2008). Similars concentracions i tendències es van trobar en les dues localitzacions, suggerint que aquest component impacta de manera similar en les mateixes. Aquest component va mostrar dependència amb les masses d'aire, amb valors més baixos en presència d'adveccions nord-atlàniques (Figura 3 de l'Article 4).

El sisè component (Figura 2 de l'Article 4, *OOA Urban (aerosol orgànic oxidat urbà)*) està format pels àcids dicarboxílics de cadena llarga (C₇-C₉), i també hi ha una petita presència de PAH i de nicotina. No està clar si aquest component té un origen primari o secundari, però té un origen urbà, ja que el nivell és més gran per l'estació RS, i no hi ha correlació entre les dues estacions. L'abundància d'aquest component en el centre de la ciutat, i el seu màxim durant el dia, fa pensar que hi ha fonts d'emissió urbanes locals implicades en la seva formació. Els àcids dicarboxílics de cadena llarga tenen relació amb la oxidació dels àcids grassos insaturats, com pot ser l'àcid oleic. Una de les fonts primàries més importants de l'àcid oleic en ambients urbans és el cuinat d'aliments (Robinson, 2006; Mohr, 2012). Una dependència complexa de les masses d'aire en aquest component (Figura 3 de l'Article 4) suggereix que els compostos presents en aquest component estan relacionats amb processos atmosfèrics. Els *scores* més elevats d'aquest component es van observar sota masses d'aire nord Africà del mar mediterrani, però també hi ha un increment d'aquest component coincident amb l'augment de l'aerosol secundari oxidat (cinquè component), i una disminució en dos períodes amb aires regionals.

En la Taula SI1 i la Figura SI1 de la Informació Suplementària de l'Article 4, es poden observar els diferents components escollits per un model de set components. La majoria dels components són pràcticament iguals dels trobats anteriorment pel model de sis components, amb la diferència que aquí el primer component apareix subdividit en dos. El primer component és molt semblant al de la solució de sis components, però aquí presenta menys importància dels hopans, que s'han traslladat al nou component, juntament amb PAH de més baix pes molecular. Aquest nou component sembla emfatitzar el tràfic degut als vehicles dièsel, que són una part important dels vehicles de Barcelona.

Com a conclusió d'aquest apartat, es pot dir que l'aplicació del mètode MCR-ALS a la base de dades de concentració obtinguda amb l'anàlisi GC-MS de diferents compostos orgànics, ens ha permès obtenir sis fonts d'emissió d'aerosol orgànic. Entre aquests s'han trobat dos directament relacionats amb emissions antropogèniques primàries (primer i segon components), que han estat més importants a l'estació de RS. La suma dels tres components SOA, tercer, quart i cinquè, han contribuït al 60% de l'aerosol orgànic a l'estació UB, similar al 55% trobat per Mohr en la mateixa localització (Mohr, 2012), mentre que aquest percentatge va ser 27% per la estació RS. Dos d'aquests components estan representats per productes d'oxidació del isoprè i del α -pinene, mentre que el tercer està compostat per una barreja de compostos amb origen antropogènic i biogènic. Les contribucions biogèniques al SOA, s'esperaven baixes en àrees urbanes, però s'ha demostrat que les emissions complexes urbanes poden accelerar la oxidació dels compostos orgànics volàtils (Goldstein, 2007; Liao, 2007). Respecte al sisè component, compostat principalment pels àcids dicarboxílics de cadena llarga C₇-C₉, els resultats suggereixen que té origen antropogènic i que depèn de processos oxidatius desconeguts en el centre urbà, però es necessitarien més dades per especificar la procedència d'aquestes emissions.

b) Determinació analítica i quimiomètrica de compostos orgànics en filtres de material particulat a la ciutat de Los Angeles i comparació amb els resultats obtinguts a Barcelona

A Los Angeles es va seguir un procediment semblant a l'utilitzat per Barcelona, però en aquest cas només es va disposar d'una estació i només es van analitzar els compostos orgànics polars, el procediment analític dels quals ja s'ha explicat anteriorment a la secció 3.2. de la present Tesi. Tant la ciutat de Barcelona com la de Los Angeles, són dues ciutats urbanes amb climes mediterranis, caracteritzades per unes elevades radiacions solars i freqüents condicions atmosfèriques anticiclòniques, el que ens ha permès una comparació interessant. Una de les diferències entre les dues ciutats és la composició dels vehicles, ja que a Los Angeles la majoria de vehicles són de benzina (Bahreini, 2012; Hayes, 2013), mentre que a Barcelona el 75% dels vehicles són dièsel i el 30% dels vehicles són motocicletes (Reche, 2012).

A la Taula 1 de l'Article 5 es poden veure les mitjanes (mínim-màxim) de les concentracions (ng/m^3) en els compostos polars analitzats en les tres estacions (una a Los Angeles i dues a Barcelona). Els traçadors orgànics per la crema de biomassa, levoglucosan, galactosan i mannosan van mostrar concentracions baixes en les tres localitzacions, ja que ni a Barcelona ni a Los Angeles la combustió per fusta és una font local d'energia important. Pel que fa als àcids dicarboxílics, tant a LA com a BCN les concentracions individuals més elevades es van trobar per l'àcid màlic i l'àcid succínic, amb valors comparables d'àcid màlic en un estudi realitzat al hivern a Barcelona (Van Drooge, 2012) i en altres estudis realitzats en altres localitzacions europees (Ion, 2005; Kourtchev, 2008). Les concentracions d'àcid succínic i àcid màlic van ser superiors a LA en comparació a les obtingudes per BCN, mentre que l'àcid ftàlic va tenir concentracions més comparables en les tres estacions. No es van observar diferències de concentració en l'àcid ftàlic entre els dies laborables i els caps de setmana, i les concentracions més elevades es van trobar durant la circulació de masses d'aire regionals. Això indica que la majoria de l'àcid ftàlic detectat (i altres àcids dicarboxílics com el succínic, el

glutàric o el màlic) són secundaris i es formen relativament ràpid en l'atmosfera en presència de radicals hidroxil.

Els àcids dicarboxílics de cadena llarga (C₇-C₉), àcid subèric, pimèlic i azelaic, tenen un comportament diferent dels àcids dicarboxílics curts, ja que aquests primers tenen concentracions més elevades en el centre de la ciutat. En el tres casos, tant a BCN com a LA, l'àcid azelaic és el que té una concentració més elevada. La presència de l'àcid azelaic en el aerosol està relacionat amb la ràpida oxidació d'àcids grassos insaturats, com per exemple l'àcid oleic, que té moltes fonts urbanes com el cuinat d'aliments, el tràfic, o altres fonts no fòssils, com els ambients marins (Schauer, 2002). Els traçadors biogènics pel carboni orgànic secundari, àcid *cis*-pinònic, àcid 3-hidroxi-glutàric, i MBTCA, van mostrar concentracions similars a les obtingudes a Marsella (El Haddad, 2011) per Barcelona, i en canvi per LA, les concentracions trobades van ser més petites que les trobades a Riverside (Stone, 2009). Això pot ser degut a la dependència estacional de la concentració d'aquestes compostos, ja que no totes les mesures es van realitzar exactament en els mateixos mesos. La relació de l'àcid 3-hidroxi-glutàric amb el MBTCA queda confirmat per les tres estacions ($r^2=0.8$ a LA i 0.7 a BCN), però l'àcid *cis*-pinònic no mostra aquesta relació. Això és degut a que l'àcid *cis*-pinònic és un producte de primera generació en la oxidació de l' α -pinene, mentre que el MBTCA es forma a partir de la oxidació del àcid *cis*-pinònic amb OH, i l'àcid 3-hidroxi-glutàric també és un producte d'oxidació generat més lentament.

Els traçadors de l'isoprè, C₅-alquetriols, àcid 2-metilglicèric i metiltetrols, han presentat valors més elevats a l'estació LA, però similars als trobats per Lewandowski (2013) en un estudi realitzat durant la mateixa campanya. Els metiltetrols i C₅-alquetriols es formen per la fotooxidació de l'isoprè amb condicions limitades de NO, mentre que l'àcid 2-metilglicèric es forma per la oxidació del metacriloperoxinitrat (MPAN) volàtil, que és un producte de segona generació de l'oxidació de l'isoprè (Surrat, 2010).

El mètode MCR-ALS es va aplicar individualment i de forma conjunta a les matrius de dades obtingudes a Barcelona i a Los Angeles per veure les diferents fonts d'emissió presents en aquestes zones. En tots els casos, el

mètode MCR-ALS es va portar a terme aplicant restriccions de no-negativitat i normalització dels *loadings*. L'anàlisi conjunta (anàlisi simultània en una matriu augmentada en la direcció de les columnes) de les bases de dades de les dues ciutats es va realitzar per emfatitzar les parts que tenen en comú les tres estacions. A les Figures 2 i 3 de l'Article 5 es poden veure els *scores* i *loadings* dels diferents components resolts amb el mètode MCR-ALS així com les seves variàncies explicades. En la Figura 2 es poden veure els *loadings* pels estudis individuals i conjunt de manera simultània. En la Figura 3 es mostren els *scores* de l'anàlisi conjunt. També s'han calculat les contribucions de cada component respecte a les mitjanes de concentració de OC corresponents a la part polar (WSOC) analitzada que són $2,8 \mu\text{g OC}/\text{m}^3$, $1,7 \mu\text{g}/\text{m}^3$ i $2,0 \mu\text{g}/\text{m}^3$ respectivament per LA, BCN-UB i BCN-RS (Figura 4 de l'Article 5).

Per la ciutat de Los Angeles, es disposava també de 7 períodes intensius (montres de 3h) per estudiar les diferències i similituds entre les tendències diàries. El mètode MCR-ALS també es va aplicar a aquesta matriu de dades intensives, i es van escollir tres components per explicar les peculiaritats diàries dels compostos orgànics polars analitzats. En aquest cas, la variància explicada total va ser del 95.8%. A la Figura 5.2.1 es poden veure els *scores* i *loadings* obtinguts amb el mètode MCR-ALS aplicat a les dades intensives de LA. Aquesta Figura no es va publicar a l'Article 5 per motius d'espai, i per això es mostra aquí, en la present Tesi.

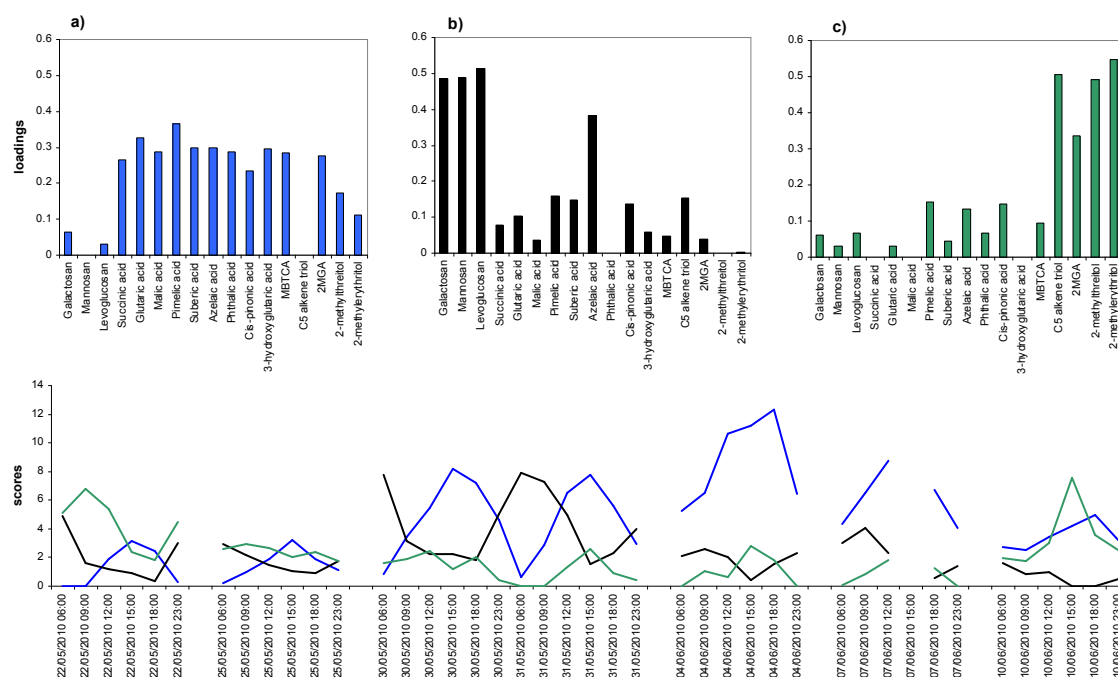


Figura 5.2.1. Scores i *loadings* obtinguts amb l'anàlisi MCR-ALS dels compostos traçadors orgànics polars en la campanya intensiva a LA en temps local. a) Carboni Orgànic Secundari, blau, b) Crema de biomassa regional (negre), c) Carboni Orgànic secundari derivat del isoprè (verd).

Tant en els anàlisis individuals com en l'anàlisi conjunt per MCR-ALS de les matrius de dades obtingudes a Barcelona i a Los Angeles, es va escollir un model de cinc component per descriure les diferents fonts d'emissió dels contaminants orgànics estudiats. El fet de que dues ciutats amb dos estudis independents tinguin fonts d'emissió similars demostra que realment aquestes dues ciutats tenen una contaminació atmosfèrica orgànica comparable degut a les seves condicions meteorològiques similars. Les variàncies explicades totals van ser 98.7, 97.7 i 97.8% pels anàlisis individuals de LA, BCN-UB i BCN-RS respectivament, i 95.9% per l'anàlisi conjunt.

El primer component o font d'emissió trobada (Figura 2a de l'Article 5, *Regional BB* (crema de biomassa regional)) separa clarament les contribucions regionals de crema de biomassa, ja que està format quasi exclusivament pel levoglucosan, mannosan i galactosan. Tant a LA (17% de WSOC) com a BCN (9-16% de WSOC) els perfils observats són bastant

constants (Figura 3a de l'Article 5). A LA es va veure una baixada a la part final del mostreig, mentre que a BCN a la part final del mostreig es va observar un increment d'aquest component. En l'anàlisi intensiu de LA (Figura 5.2.1b) els nivells més alts de crema de biomassa es van observar a la nit (0.00h-6.00h) i els nivells més baixos es van observar a primeres hores de la tarda (15.00h-18.00h).

El segon component o font d'emissió (Figura 2b de l'Article 5, *Aged SOC* (carboni orgànic secundari antic)) està format principalment pels traçadors d'aerosol orgànic oxigenats, com els àcids dicarboxílics de cadena curta i altres més oxigenats d'origen biogènic, com l'àcid màlic, l'àcid 3-hidroxi-glutàric o el MBTCA. Els compostos representats en aquest component (28% de WSOC en LA i 15% a BCN) són productes de processos d'oxidació relacionats amb aire processat d'emissions urbanes i biogèniques en relació amb la circulació de masses d'aire en les zones estudiades. Les tendències d'aquest component són similars a les obtingudes per l'aerosol orgànic oxigenat de baixa volatilitat obtingut per AMS en les dues localitzacions (Dall'Osto, 2013; Hayes, 2013) indicant l'alt estat d'oxidació d'aquests aerosols. En el mostreig intensiu de LA, els cicles diürns d'aquests compostos oxigenats (Figura 5.2.1a) van tenir intensitats màximes entre les 12.00h i les 15.00h. En les tres localitzacions i al mostreig intensiu de LA, els majors nivells d'aquest component corresponen als dies amb més quantitat d'aerosols durant les diferents campanyes.

El tercer component o font d'emissió (Figura 2c de l'Article 5, *OO Urban* (carboni orgànic oxigenat urbà)) està compost pels àcids dicarboxílics de cadena llarga (C_7-C_9). Aquest component és el que presenta més diferències entre BCN (19% de WSOC a BCN-UB i 50% a BCN-RS) i LA (16% de WSOC). A BCN, en aquest component només apareixen els àcids dicarboxílics C_7-C_9 , mentre que a LA, també apareixen l'àcid ftàlic, l'àcid màlic, l'àcid 3-hidroxi-glutàric i el MBTCA. Les majors contribucions d'aquests compostos oxidats a LA pot estar relacionat amb una influència més equilibrada dels potencials processos/fons d'aquest component en comparació amb les de BCN. En l'estació RS hi ha una forta fluctuació dia-nit d'aquest component que pot estar relacionat amb emissions urbanes

específiques i processos d'oxidació. En l'estació UB aquesta fluctuació és menys pronunciada, mentre que a LA els àcids C₇-C₉ estan més barrejats amb els àcids més curts que són més antics i oxidats. Per tant, aquest component és més fresc a BCN, mentre que és més antic (oxigenat) a LA.

El quart component o font d'emissió (Figura 2d de l'Article 5, *Isoprene* SOC (carboni orgànic secundari del isoprè)) està compost a les tres localitzacions pels productes de la oxidació del isoprè, C₅-alquetriols, 2-metiltetrols i 2-metilglicèric àcid. Els C₅-alquetriols i els metiltetrols van mostrar bones correlacions tant per BCN (6-11% de WSOC) com per LA (36% de WSOC). La correlació d'aquests compostos amb el àcid 2-metilglicèric també va ser bona en BCN, però més dèbil en LA. Aquest fet suggereix que, encara que tots aquests productes derivats de la oxidació del isoprè tenen orígens similars, es formen amb camins diferents, com ja s'ha comentat anteriorment. En el mostreig intensiu per LA (Figura 5.2.1c) també es veu el comportament una mica diferent de l'àcid 2-metilglicèric, ja que té una intensitat una mica menor que els altres derivats del isoprè, i a més apareix amb més intensitat al component relacionat amb els compostos oxidats més antics.

El cinquè component o font d'emissió (Figura 2e de l'Article 5, *Biogenic α -pinene* SOC (carboni orgànic secundari biogènic de l' α -pinene)) està format principalment per l'àcid *cis*-pinònic. La separació d'aquest compost dels altres compostos analitzats indica que prové d'una font (procés) produïda amb condicions específiques. Els nivells més alts d'àcid *cis*-pinònic es van observar tant a BCN (21-37% de WSOC) com a LA (4% de WSOC) sota la influència de condicions d'advecció caracteritzades per baixes humitats relatives. La correlació negativa de l'àcid *cis*-pinònic i la humitat relativa suggereix que les reaccions en fase aquosa juguen un paper important en el camí de formació de l'àcid *cis*-pinònic, com ja s'ha observat en experiments de laboratori (Fick, 2002). El increment de l'àcid *cis*-pinònic en les localitzacions estudiades pot estar relacionat amb el carboni orgànic secundari biogènic fresc en una de les primeres etapes de formació de l'aerosol a partir de vapors (O'Dowd, 2002).

Com a conclusió d'aquest apartat podem dir que la cromatografia de gasos acoblada a l'espectrometria de masses ens ha permès obtenir diversos compostos traçadors orgànics polars en filtres de material particulat que ens han proporcionat informació important sobre les localitzacions estudiades. L'aplicació del mètode MCR-ALS ens ha permès trobar cinc components relatius a les diverses fonts de contaminació de les zones estudiades. Entre aquestes fonts s'han trobat processos d'oxidació d'orígens biogènics i antropogènics que han estat similars en dues ciutats molt separades entre sí, però que tenen condicions meteorològiques semblants, com són Barcelona i Los Angeles a Califòrnia (USA).

Capítol 6

CONCLUSIONS

Conclusions de tipus quimiomètric

1. L'aplicació dels mètodes d'Anàlisi de Components Principals (PCA) i de Resolució Multivariant de Corbes per Mínims Quadrats Alternats (MCR-ALS) a taules de dades ambientals agrupades en diferents estructures ha permès obtenir i interpretar perfils de distribució de fonts de contaminació atmosfèrica (*source apportionment*).

2. L'aplicació de les restriccions de trilinearitat i interacció entre components en el mètode MCR-ALS ha permès resoldre els diferents perfils diaris i anuals d'emissió d'òxid de nitrogen i d'ozó a la ciutat Barcelona. La restricció de trilinearitat ha permès simplificar la interpretació dels perfils de distribució estacionals de les fonts d'aquests contaminants en diferents anys. La restricció d'interacció entre components ha simplificat també en alguns casos la interpretació dels perfils interanuals, reduint el nombre de perfils a interpretar en el mode on s'ha aplicat aquesta restricció.

3. S'han comparat els resultats obtinguts aplicant les restriccions de trilinearitat i d'interacció de components en el mètode MCR-ALS amb els resultats equivalents obtinguts pels mètodes PARAFAC i TUCKER3, respectivament. S'han obtingut resultats comparables, amb perfils semblants i diferències petites.

4. S'ha desenvolupat una nova versió del mètode MCR-ALS que ha permès treballar amb dades incompletes. Aquesta nova versió del mètode ha permès analitzar dades simulades amb diferents nivells d'error, els resultats de les quals s'han comparat amb els obtinguts amb dades completes. S'ha aplicat el model trilineal en aquestes dades incompletes, reduint els efectes de propagació del soroll.

Conclusions de tipus ambiental

5. S'ha obtingut la distribució geogràfica i temporal dels perfils d'emissió de NO en diferents estacions de Catalunya durant diferents anys (2001-2006). S'ha comprovat que els perfils de NO estan fortament influïts pel tràfic intens que es troba en localitzacions urbanes com Barcelona, mostrant dos màxims diaris en les hores punta del matí i de la tarda, i mostrant uns majors nivells de NO a l'hivern i a la tardor. Aquests perfils diaris i estacionals s'han observat en menor intensitat també en zones urbanes petites i més rurals. S'ha confirmat que les emissions de NO tenen un origen principalment antropogènic.

6. S'ha obtingut la distribució geogràfica i temporal dels perfils d'emissió de O₃ en una estació de Barcelona durant diferents anys (2000-2003). S'ha confirmat que els perfils diaris d'emissió d'O₃ estan molt correlacionats amb la intensitat de la radiació solar, mostrant els nivells més alts al voltant de les 16h del dia, i en els perfils estacionals aquests màxims són a la primavera i a l'estiu. També s'ha observat que els perfils d'emissió del O₃ estan inversament correlacionats amb els perfils d'emissió del NO.

7. L'aplicació del mètode MCR-ALS a les taules de dades de NO i O₃ de forma conjunta (anys coincidents) en una mateixa estació de Barcelona, ha confirmat clarament la complementarietat dels perfils del NO i el O₃, visualitzant fàcilment la correlació inversa entre aquests dos contaminants en els seus perfils temporals (diaris i estacionals).

8. S'ha desenvolupat un procediment analític de determinació de compostos orgànics en material particulat mitjançant extracció, derivatització i posterior anàlisi per cromatografia de gasos acoblada a espectrometria de masses, que ha permès detectar i quantificar 36 compostos orgànics diferents tant polars com neutres.

9. El mètode MCR-ALS ha permès estimar les possibles fonts d'emissió de compostos orgànics analitzats en material particulat a Barcelona. S'han trobat sis fonts diferents d'emissió dels constituents de l'aerosol orgànic, entre les quals hi ha dues fonts primàries d'origen antropogènic, com són la crema de biomassa i el tràfic, i altres quatre fonts relacionades amb l'aerosol orgànic secundari, que poden tenir origen biogènic o també antropogènic. En aquests quatre fonts hi ha contaminants provinents de l'oxidació de l'isoprè, contaminants provinents de l'oxidació de l' α -pinene, àcids dicarboxílics de cadena llarga provinents de fonts urbanes i àcids dicarboxílics de cadena curta oxidats provinents de fonts antropogèniques i biogèniques.

10. S'han comparat els nivells de 17 contaminants orgànics polars en material particulat analitzats a Los Angeles i a Barcelona. Els nivells trobats tenen petites diferències en les concentracions dels diferents compostos.

11. S'han comparat les fonts d'emissió de contaminants orgànics a Los Angeles i Barcelona, obtingudes mitjançant el mètode quimiomètric MCR-ALS. S'han resolt cinc fonts d'emissió que són coincidents en les dues ciutats. Aquestes fonts d'emissió han estat assignades a: crema de biomassa, contaminants provinents de l'oxidació de l'isoprè, contaminants provinents de l'oxidació de l' α -pinene, àcids dicarboxílics de cadena llarga provinents de fonts urbanes i àcids dicarboxílics de cadena curta oxidats provinents de fonts antropogèniques i biogèniques. Aquests resultats s'han obtingut tant en les anàlisis individuals de les tres estacions (dos a Barcelona i una a Los Angeles), com en la seva anàlisi conjunta.

Capítol 7

REFERÈNCIES

Albers, P.H.: *Handbook of Ecotoxicology*, Lewis Publishers, Boca Raton, **1995**, pp. 330-355.

Andersson, C.A.; Bro, R.: Improving the speed of multiway algorithms: Part I. Tucker3. *Chemometrics and Intelligent Laboratory Systems* 42 (**1998**) 93-103.

Andrade, J.M.; Fresco, P.; Muniategui, S.; Prada, D.: Comparison of oil spillages using mid-IR indexes and 3-way procrustes rotation, matrix-augmented principal components analysis and parallel factor analysis, *Talanta* 77 (**2008**) 2, 863-869.

Arimoto, R.: Eolian dust and climate: relationships to sources, tropospheric chemistry, transport and deposition. *Earth-Science Reviews* 54 (**2001**) 29-42.

Arteaga, F.; Ferrer-Riquelme, A.J.: Missing Data, *Comprehensive Chemometrics*, Ed. Elsevier, **2009**, volum 3, 285-314.

Astel, A.; Simeonov, V.; Bauer, H.; Puxbaum, H.: Multidimensional modeling of aerosol monitoring data, *Environmental Pollution* 158 (**2010**) 10, 3201-3208.

Baird, C.: *Química ambiental*; Ed. Reverté S. A., **2001**

Bahreini, R.; Middlebrook, A.M.; de Gouw, J.A.; Warneke, C.; Trainer, M.; Brock, C.A.; Stark, H.; Brown, S.S.; Dube, W.P.; Gilman, J.B.; Hall, K.; Holloway, J.S.; Kuster, W.C.; Perring, A.E.; Prevot, A.S.H.; Schwartz, J.P.; Spackman, J.R.; Szidat, S.; Wagner, N.L.; Weber, R.J.; Zotter, P.; Parrish, D.D.: Gasoline emissions dominate over diesel in formation of secondary organic aerosol mass. *Geophysical Research Letters* (**2012**) 39, L06805.

Bi, X.; Sheng, G.; Feng, Y.; Fu, J.; Xie, J.: Gas and particulate phase specific tracer and toxic organic compounds in environmental tobacco smoke, *Chemosphere* 61 (**2005**) 1512-1522.

Bloemberg, T.G.; Gerretzen, J.; Wouters, H.J.P.; Gloerich, J.; van Dael, M.; Wessels, H.J.C.T.; van den Heuvel, L.P.; Eilers, P.H.C.; Buydens, L.M.C.; Wehrens, R.: Improved parametric time warping for proteomics, *Chemometric and Intelligent Laboratory Systems* 104 (2010) 65–74.

Borbon, A.; Fontaine, H.; Veillerot, M.; Locoge, N.; Galloo, J. C.; Guillermo, R.: An investigation into traffic-related fraction of isoprene at an urban location, *Atmospheric Environment* 35 (2001), 5, 3749–3760.

Bratchell, N.: Cluster analysis, *Chemometrics and Intelligent Laboratory Systems* 6 (1989) 2, 105-125.

Brereton, R.G.: *Chemometrics: Applications of mathematics and statistics to laboratory systems*. New York: Ellis Horwood, 1990.

Bro, R.; de Jong, S.: A fast non-negativity-constrained least squares algorithm, *Journal of Chemometrics* 11 (1997a) 393–401.

Bro, R.: PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* 38 (1997b) 149-171.

Brunekreef, B.; Janssen, N.A.H.; Harssema, H.; Knape, M.; Vliet, P.V.: Air Pollution from truck traffic and lung function in children living near motorways. *Epidemiology* 8 (1997), 8.

Carrera, G.; Fernández, P.; Vilanova, R.M.; Grimalt, J.O.: Persistent organic pollutants in snow from European high mountain areas, *Atmospheric Environment* 35 (2001) 2, 245–254.

Chebbi, A.; Carlier, P.: Carboxylic acids in the troposphere, occurrence, sources and sinks. A review. *Atmospheric Environment* 30 (1996), 24, 4233-4249.

- Cheng**, Y.; Li, S.-M.; Leithead, A.; Brickell, P.C; Leaitch, W.R.: Characterizations of cis-pinonic acid and n-fattyacids on fine aerosols in the Lower Fraser Valley during Pacific 2001 Air Quality Study, *Atmospheric Environment* 38 (2004) 789–5800.
- Claeys**, M.; Wang, W.; Ion, A.C.; Kourtchev, I.; Gelencsér, A.; Maenhaut, W.: Formation of secondary organic aerosols from isoprene and its gas-phase oxidation products through reaction with hydrogen peroxide, *Atmospheric Environment* 38 (2004a) 4093–4098.
- Claeys**, M.; Graham, B.; Vas, G.; Wang, W.; Vermeylen, R.; Pashynska, V.; Cafmeyer, J.; Guyon, P.; Andreae, M.O.; Artaxo, P.; Maenhaut, W.: Formation of Secondary Organic Aerosols through Photooxidation of Isoprene, *Science* 303 (2004b) 5661: 1173-76.
- Claeys**, M.; Szmigielski, R.; Kourtchev, I.; Van Der Veken, P.; Vermeylen, R.; Maenhaut, W.; Jaoui, M.; Kleindienst, T.E.; Lewandowski, M.; Offenberg, J.H.; Edney, E.O.: Hydroxydicarboxylic Acids: Markers for Secondary Organic Aerosol from the Photooxidation of α -Pinene, *Environmental Science & Technology* 41 (2007)1628-1634.
- Cleemens**, A.L.; Seinfeld, J.H.: Detection and quantification of 2-methyltetrols in ambient aerosol in the southeastern United States, *Atmospheric Environment* 41 (2007)1825-1830.
- Cole**, H.W.; Ray, W.J.: EEG correlates of emotional tasks related to attentional demands, *International Journal of Psychophysiology* 3 (1985) 1, 33-41.
- Christin**, C.; Smilde, A.K.; Hoefsloot, H.C.; Suits, F.; Bischoff, R.; Horvatovich, P.L.: Optimized time alignment algorithm for LC-MS data: correlation optimized warping using component detection algorithm-selected mass chromatograms, *Analytical Chemistry* 80 (2008) 7012–7021.

Cruz, C. and Pandis, S. N.: The effect of organic coatings on the cloud condensation nuclei activation of inorganic atmospheric aerosol. *Journal of Geophysical Research* 103 (1998) 13,111-123.

Dadashi, M; Abdollahi, H; Tauler, R.: Maximum Likelihood Principal Component Analysis as initial projection step in Multivariate Curve Resolution analysis of noisy data, *Chemometrics and Intelligent Laboratory Systems* 118 (2012) 33–40.

Dadashi, M; Abdollahi, H.; Tauler, R.: Application of maximum likelihood multivariate curve resolution to noisy data sets, *Journal of Chemometrics* 27 (2013) 34-41.

Dades bàsiques de mobilitat, 2001. Ajuntament de Barcelona.

<http://w110.bcn.cat/Mobilitat/Continguts/Documents/Fitxers/dadesbasiques2011complert.pdf>

Dall'Osto, M.; Querol, X.; Alastuey, A.; Minguillon, M. C.; Alier, M.; Amato, F.; Brines, M.; Cusak, M.; Grimalt, J. O.; Karanasiou, A.; Moreno, T.; Pandolfi, M.; Pey, J.; Reche, C.; Ripoll, A.; Tauler, R.; Van Drooge, B. L.; Viana, M.; Harrison, R.M.; Gietl, J.; Beddows, D.; Bloss, W.; O'Dowd, C.; Ceburnis, D.; Martucci, G.; Ng, S.; Worsnop, D.; Wenger, J.; Mc Gillcuddy, E.; Sudou, J.; Healy, R.; Lucarelli, F.; Nava, S.; Jimenez, J.L.; Gomez Moreno, F.; Artinano, B.; Prevot, A.S.; Pfaffenberger, L.; Frey, S.; Wilsenack, F.; Casabona, D.; Jimenez-Guerrero, P.; Gross, D.; Cotz, N.: Presenting SAPUSS: Solving Aerosol Problem by using synergistic strategies at Barcelona, Spain, *Atmospheric Chemistry and Physics* 13 (2013) 8991–9019.

Danzer, K; Currie, L.A.: Guidelines for calibration in analytical chemistry, *Pure and applied chemistry* 70 (1998) 993-1014.

de Gouw, J.; Jimenez, J.L.: Organic Aerosols in the Earth's Atmosphere, *Environmental Science & Technology* 43 (2009) 20:7614-18.

de Juan, A; Tauler, R.: Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution, *Analytica Chimica Acta* 500 (2003) 195.

de Juan, A. ; Mas, S.: Multivariate Curve Resolution Methods for Food Chemistry, *Data Handling in Science and Technology* 28 (2013) 235-263.

de Ligny, C.L.; Spanjer, M.; van Houwelingen, J.C.; Weesie, H.M.: Three-mode factor analysis of data on retention in normal-phase high-performance liquid chromatography, *Journal of Chromatography* 301 (1984) 311–324.

de Luca, M.; Tauler, R.; Ioele, G.; Ragno, G.: Study of photodegradation kinetics of melatonin by multivariate curve resolution (MCR) with estimation of feasible band boundaries, *Drug Testing and Analysis* 5 (2013) 96-102.

Ding, X.; Wang, X.-M.; Zheng, M.: The influence of temperature and aerosol acidity on biogenic secondary organic aerosol tracers: observations at a rural site in the central Pearl River Delta region, South China, *Atmospheric Environment* 45 (2011) 1303–1311.

Dockery, D.; Pope, A.: Epidemiology of acute health effects: Summary of timeseries studied. *Particles in our air: concentration and health effects* (1996) (ed. Spengler J.D. Wilson R.), pp. 123-147. Harvard University Press.

DOGC 1995: Forest fire prevention, Catalunya, Ordinance 64/1995.

Donahue, N. M.; Robinson, A.L.; Pandis, S.N.: Atmospheric Organic Particulate Matter: From Smoke to Secondary Organic Aerosol, *Atmospheric Environment* 43 (2009) 1: 94-106.

Eatough, D. J.; Benner, C. L.; Bayona, J. M.; Richards, G.; Lamb, J. D.; Lee, M. L.: Chemical composition of environmental tobacco smoke. 1. Gas-phase acids and bases, *Environmental Science & Technology*. 23 (1989) 679-687.

Edney, E. O.; Kleindienst, T. E.; Conner, T. S.; McIver, C. D.; Corse, E. W.; Weathers, W. S.: Polar organic oxygenates in PM_{2.5} at a southeastern site in the United States. *Atmospheric Environment* 37 (2003) 3947-3965.

Edney, E. O.; Kleindienst, T.E.; Lewandowski, M.; Offenberg, J.H.; Wang, W.; Claeys, M.: Formation of 2-Methyl Tetrols and 2-Methylglyceric Acid in Secondary Organic Aerosol from Laboratory Irradiated Isoprene/NO_x/SO₂/Air Mixtures and Their Detection in Ambient PM_{2.5} Samples Collected in the Eastern United States. *Atmospheric Environment* 39 (2005) 29: 5281-89.

Eilers, P.H.C.: Parametric time warping, *Anal. Chem.* 76 (2004) 404–411.

El Haddad, I., Marchand, N., Temime-Roussel, B., Wortham, H., Piot, C., Besombes, J.-L., Baduel, C., Voisin, D., Armengaud, A., and Jaffrezo, J.-L.: Insights into the secondary fraction of the organic aerosol in a Mediterranean urban area: Marseille, *Atmospheric Chemistry and Physics* 11 (2011) 2059-2079.

Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J.J.; Downey, G.; Blanchet, L.; Buydens, L.M.C.: Breaking with trends in pre-processing? (Review), *Trends in Analytical Chemistry* 50 (2013) 96-106.

Engelhart, G. J.; Asa-Awaku, A.; Nenes, A.; Pandis, S.N.: CCN Activity and Droplet Growth Kinetics of Fresh and Aged Monoterpene Secondary Organic Aerosol. *Atmospheric Chemistry and Physics* 8 14 (2008): 3937-49

Escandar, G.M.; Faber, N.M.; Goicoechea, H.C.; Muñoz de la Peña, A.; Olivieri, A.C.; Poppi, R.J.: Second- and third-order multivariate calibration: data, algorithms and applications, *Trends in Analytical Chemistry* 26 (2007) 7.

Esteban, M.; Ariño, C.; Díaz-Cruz, J.M.; Díaz-Cruz, M.S.; Tauler, R.: Multivariate curve resolution with alternating least squares optimisation: A

soft-modelling approach to metal complexation studies by voltammetric techniques, *Trends in Analytical Chemistry* 19 (2000) 1, 49-61.

Fang, J.; Kawamura, K.; Ishimura, Y.; Matsumoto, K.: Carbon isotope composition of fatty acids in the marine aerosols from the Western North Pacific: Implication for the source and atmospheric transport, *Environmental Science and Technology* 62 (2002) 2598–2604.

Farnham, I.M.; Singh, A.K.; Stetzenbach, K.J.; Johannesson, K.H.: Treatment of nondetects in multivariate analysis of groundwater geochemistry data, *Chemometrics and Intelligent Laboratory Systems* 60 (2002) 265– 281.

Ferrer, I.; Thurman, E.M.: GC-MS-MS for the analysis of phytoestrogens in the environment, *Comprehensive Analytical Chemistry* 61 (2013) 231-251.

Fick, J.; Pommer, L.; Nilsson, C.; Andersson, B.: Effect of OH radicals, relative humidity, and time on the composition of the products formed in the ozonolysis of α -pinene. *Atmospheric Environment* 37 (2003) 4087–4096.

Gampp, H.; Maeder, M.; Meyer, Ch.; Zuberhler, A.D.: Quantification of a known component in an unknown mixture, *Analytica Chimica Acta* 193 (1987) 287.

Gargallo, R.; Tauler, R.; Cuesta-Sánchez, F.; Massart, D.L.: Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation, *Trends in Analytical Chemistry* 15 (1996) 7, 279-286.

Gemperline, P.J.; Miller, K.H.; West, T.; Weinstein, E.; Hamilton, J.C.; Bray, J.T.: Principal component analysis, trace elements and blue crab shell disease, *Analytical Chemistry* 64 (1992). 523–532.

Gerosa, G.; Ferretti, M.; Bussotti, F.; Rocchini, D.: Estimates of ozone AOT40 from passive sampling in forest sites in South-Western Europe, *Environmental Pollution* 145 (2007) 629-635.

Gogou, A.; Stephanou, E. G.; Stratigakis, N.; Grimalt, J. O.; Simo, R.; Aceves, M.; Albaiges, J.: Differences in lipid and organic salt constituents of aerosols from Eastern and Western Mediterranean coastal cities, *Atmospheric Environment* 28 (1994)1301–1310.

Goldstein, A. H.; Galbally, I. E.: Known and unexplored organic constituents in the earth's atmosphere, *Environmental Science & Technology* 41 (2007) 1514–1521.

Golub, GH; Loan, CFV.: Matrix computations, John Hopkins University Press., Baltimore, USA, 1996

Guenther, A.; Hewitt, C. N.; Erickson, D.; Fall, R.; Geron, C.; Graedel, T.; Harley, P.; Klinger, L.; Lerdau, M.; McKay, W. A.; Pierce, T.; Scholes, B.; Steinbrecher, R.; Tallamraju, R.; Taylor, J.; Zimmerman, P.: A global model of natural volatile organic compound emissions. *Journal of Geophysical Research-Atmospheres* 100 (1995) (D5) 8873-8892.

Hallquist, M.; Wenger, J. C.; Baltensperger, U.; Rudich, Y.; Simpson, D.; Claeys, M.; Dommen, J.; Donahue, N.M.; George, C.; Goldstein, A.H.; Hamilton, J.F.; Herrmann, H.; Hoffmann, T.; Iinuma, Y.; Jang, M.; Jenkin, M.E.; Jimenez, J.L.; Kiendler-Scharr, A.; Maenhaut, W.; McFiggans, G.; Mente, Th.F.; Monod, A.; Prevot, A.S.H.; Seinfeld, J.H; Surratt, J.D.; Szmigielski, R.; Wildt, J.: The Formation, Properties and Impact of Secondary Organic Aerosol: Current and Emerging Issues. *Atmospheric Chemistry and Physics* 9 14 (2009): 5155-236.

Hamilton, J. F.; Lewis, A.C.; Carey, T.J.; Wenger, J.C.; Borrás i Garcia, E.; Muñoz, A.: Reactive Oxidation Products Promote Secondary Organic Aerosol

Formation from Green Leaf Volatiles. *Atmospheric Chemistry and Physics* 9 11 (2009): 3815-23

Hanrahan G.: Environmental chemometrics. Principles and modern applications. CRC Pres, Taylor & Francis Group, 2009.

Harshman, R.A.: Foundations of the PARAFAC procedure: models and conditions for an 'exploratory' multimodal factor analysis. *UCLA Working Papers Phonet.* 16 (1970) 1-84.

Harshman, R.A.; Lundy, M.E.: Data preprocessing and the extended PARAFAC model, en *Research methods for multimode data analysis*, Law, H.G.; Snyder, C.W.; Hattie, J.; Mc Donald, R.P. (Eds), Praeger, New York, 1984, pàg. 216-284.

Hatzianastassiou, N; Katsoulis, B.D.; Antakis, B.: Extreme nitrogen oxide and ozone concentrations in athens atmosphere in relation to meteorological conditions. *Environmental Monitoring and Assessment* 128 (2007) 1-3, 447-464.

Hayes, P.L.; Ortega, A.M.; Cubison, M.J.; Froyd, K.D.; Zhao, Y.; Cliff, S.S.; Hu, W.W.; Toohey, D.W.; Flynn, J.H.; Lefer, B.L.; Grossberg, N.; Alvarez, S.; Rappenglück, B.; Taylor, J.W.; Allan, J.D.; Holloway, J.S.; Gilman, J.B.; Kuster, W.C.; de Gouw, J.A.; Massoli, P.; Zhang, X.; Liu, J.; Weber, R.J.; Corrigan, A.L.; Russell, L.M.; Isaacman, G.; Worton, D.R.; Kreisberg, N.M.; Goldstein, A.H.; Thalman, R.; Waxman, E.M.; Volkamer, R.; Lin, Y.H.; Surratt, J.D.; Kleindienst, T.E.; Offenberg, J.H.; Dusanter, S.; Griffith, S.; Stevens, P.S.; Brioude, J.; Angevine, W.M.; Jimenez, J.L.: Organic Aerosol Composition and Sources in Pasadena, California during the 2010 CalNex Campaign. *Journal of Geophysical Research: Atmospheres* 118 (2013) 1–25.

Hennigan, C. J.; Miracolo, M. A.; Engelhart, G. J.; May, A.A.; Presto, A.A.; Lee, T.; Sullivan, A.P.; McMeeking, G.R.; Coe, H.; Wold, C.E.; Hao, W.-M.; Gilman, J.B.; Kuster, W.C.; De Gouw, J.; Schichtel, B.A.; Collett Jr., J.L.; Kreidenweis,

S.M.; Robinson, A.L.; Chemical and physical transformations of organic aerosol from the photo-oxidation of open biomass burning emissions in an environmental chamber, *Atmospheric Chemistry and Physics* 11 (2011) 15, 7669-7686.

Henze, D. K., Seinfeld, J.H.: Global Secondary Organic Aerosol From Isoprene Oxidation. *Geophysical Research Letters* 33 (2006) 9.

Hester, R.E; Harrison, R.M.: Environmental Forensics. *Issues in Environmental Science and Technology*, 26 (2008) 1350-7583.

Ho, K. F.; Ho, S. S. H.; Lee, S. C.; Kawamura, K.; Zou, S. C.; Cao, J. J.; Xu, H. M.: Summer and winter variations of dicarboxylic acids, fatty acids and benzoic acid in PM_{2.5} in Pearl Delta River Region, China, *Atmospheric Chemistry and Physics* (2011) 11, 2197–2208.

Hong, H.; Yang, L.; Guo, W.; Wang, F.; Yu, X.: Characterization of dissolved organic matter under contrasting hydrologic regimes in a subtropical watershed using PARAFAC model, *Biogeochemistry* 109 (2013) 1-3, 163-174.

Hopke, P.K.: The evolution of chemometrics, *Analytica Chimica Acta* 500 (2003) 365-377.

Hopke, P.K.: Factor Analysis, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Ed. Elsevier, 2009a, Capítol 4.03.6, 64.

Hopke, P.K.: Pattern Recognition, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Ed. Elsevier, 2009b, Capítol 4.03.2, 56.

Ion, A.C.; Vermeylen, R.; Kourtchev, I.; Cafmeyer, J.; Chi, X.; Gelencsér, A.; Maenhaut, W.; Claeys, M.: Polar organic compounds in rural PM_{2.5} aerosols from K-pusztá, Hungary, during a 2003 summer field campaign: sources and diel variations. *Atmospheric Chemistry and Physics* 5 (2005) 1805–14.

ISO 13964:1998; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=23528

ISO 7993:1985; UNE 77212:1993; http://www.iso.org/iso/catalogue/catalogue_tc/catalogue_detail.htm?csnumber=14995 ;
http://www.aenor.es/desarrollo/normalizacion/normas/buscador_normas.asp

Jain, R.B.; Caudill, S.P.; Wang, R.Y.; Monsell, E.: Evaluation of maximum likelihood procedures to estimate left censored observations, *Analytical Chemistry* 80 (2008) 4, 1124-1132.

Jaumot, J.; Marchán, V.; Gargallo, R.; Grandas, A.; Tauler, R.: Multivariate Curve Resolution applied to the analysis and resolution of two - dimensional [1H, 15N]-NMR reaction spectra, *Analytical Chemistry* 76 (2004) 7094 – 7101.

Jaumot, J., Gargallo, R., de Juan, A., Tauler, R.: A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics and Intelligent Laboratory Systems* 76 (2005) 101–110.

Jaumot, J.; Tauler, R.: MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, *Chemometrics and Intelligent Laboratory Systems* 103 (2010) 96–107.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K.,

Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, *Science*, 326 (**2009**) 1525–1529.

Jolliffe, I. T.: *Principal Component Analysis*, 2nd Edn., Springer-Verlag, New York, USA, **2002**.

Kadowaki, S.: Characterization of Carbonaceous Aerosols in the Nagoya Urban Area. 2. Behavior and Origin of Particulate n-Alkanes, *Environmental Science & Technology* 28 (**1994**)129-135.

Kahnt, A.; Behrouzi, S.; Vermeylen, R.; Safi Shalamzari, M.; Vercauteren, J.; Roekens, E.; Claeys, M.; Maenhaut, W.: One-year study of nitro-organic compounds and their relation to wood burning in PM₁₀aerosol from a rural site in Belgium, *Atmospheric Environment* 81 (**2013**) 561-568.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., Wilson, J.: Organic aerosol and global climate modelling: a review. *Atmospheric Chemistry and Physics* 5 (**2005**) 1053–1123.

Kawamura, K.; Gagosian, R. B.: Implications of ω -Oxocarboxylic acids in the remote marine atmosphere for photo-oxidation of unsaturated fatty acids, *Nature* 325 (**1987**) 330–332.

Kawamura, K. Sakaguchi F.: Molecular distributions of water soluble dicarboxylic acids in marine aerosols over the Pacific Ocean including tropics. *Journal of Geophysical Research* 104 (**1999**) 3501-3509.

Kerminen, V.-M.; Ojanen, C.; Pakkanen, T.; Hillamo, R.; Aurela, M.; Merilä-Kinen, J.: Low-molecular-weight dicarboxylic acids in an urban and rural atmosphere, *Journal of Aerosol Science* 31 (2000)3:349-362.

Kessler, S. H.; Smith, J. D.; Che, D. L.; Worsnop, D. R.; Wilson, K. R.; Kroll, J. H., Chemical sinks of organic aerosol: Kinetics and products of the heterogeneous oxidation of erythritol and levoglucosan. *Environmental Science and Technology* 44 (2010) 7005-7010.

Koss G, Tesseraux I, Toxicology, Academic Press, San Diego, (1999) pp. 603-644.

Kourtchev, I.; Ruuskanen, T.; Maenhaut, W.; Kulmala, M.; Claeys, M.: Observation of 2-methyltetrols and related photo-oxidation products of isoprene in boreal forest aerosols from Hyytiälä, Finland, *Atmospheric Chemistry and Physics* 5 (2005) 2761-2770.

Kourtchev, I.; Warnke, J.; Maenhaut, W.; Hoffmann, T.; Claeys, M.: Polar organic marker compounds in PM_{2.5} aerosol from a mixed forest site in western Germany. *Chemosphere* 73 (2008)1309–14.

Kroll, J. H., Seinfeld, J.H.: Chemistry of Secondary Organic Aerosol: Formation and Evolution of Low-Volatility Organics in the Atmosphere. *Atmospheric Environment* 42 (2008)16: 3593-624.

Kroonenberg, P.M.; de Leeuw, J.: Principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika*, 45 (1980) 69–97.

Kroonenberg, P. M.: Three-Mode Principal Component Analysis; DSWO Press: Leiden, 1983.

Kubatova, A., Vermeylen, R., Claeys, M., Cafmeyer, J., Maenhaut, W., Roberts, G., and Artaxo, P.: Carbonaceous aerosol characterization in the Amazon basin, Brazil: novel dicarboxylic acids and related compounds, *Atmospheric Environment* 34 (2000) 5037-5051.

Künzli N., Kaier R., Medina S., Studnicka M., Chanel O., Filliger P., Herry M., Jr F.H., Puybonnieux-Textier V., Quénel P., Schneider J., Seethaler R., Vergnaud J.C. y Sommer H., Public health impact of outdoor and traffic related air pollution: a European assessment. *The Lancet* 356 (2000), 795-801.

Kvalheim, O.M.; Liang, Y.Z.: Heuristic evolving latent projections: Resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution, *Analytical Chemistry* 64 (1992) 936

Laaksonen, A.; Kulmala, M.; O'Dowd, C. D.; Joutsensaari, J.; Vaattovaara, P.; Mikkonen, S.; Lehtinen, K. E. J.; Sogacheva, L.; Dal Maso, M.; Aalto, P.; Petaja, T.; Sogachev, A.; Yoon, Y.J.; Lihavainen, H.; Nilsson, D.; Facchini, M. C.; Cavalli, F.; Fuzzi, S.; Hoffmann, T.; Arnold, F.; Hanke, M.; Sellegri, K.; Umann, B.; Junkermann, W.; Coe, H.; Allan, J. D.; Alfarra, M.R.; Worsnop, D. R.; Riekkola, M.-L.; Hyotylainen, T.; Viisanen, Y.: The role of VOC oxidation products in continental new particle formation, *Atmospheric Chemistry and Physics* 8 (2008) 2657–2665.

Ladefoged, P.; Harshman, R.; Goldstein, L.; Rice, L.: Generating vocal tract shapes from formant frequencies, *Journal of the Acoustical Society of America* 64 (1978) 4, 1027-1035.

Lavine, B.K.: Statistical Discriminant Analysis, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Ed. Elsevier, 2009, Capítol 3.16.

Lawerys, RR.: Cánceres de Origen Profesional. Toxicología Industrial e Intoxicaciones Profesionales. Barcelona, Madrid: Masson S.A.; (1994), p. 553-76.

Lawson, C.L.; Hanson, R.J.: Solving least-squares problems, 1st ed., Prentice-Hall, Englewood Cliffs, NJ, US, **1974**

Lewandowski, M.; Piletic, I.R.; Kleindienst, T.E.; Offenberg, J.H.; Beaver, M.R.; Jaoui, M.; Docherty, K.S.; Edney, E.O.: Secondary organic aerosol characterisation at field sites across the United States during the spring–summer period. *International Journal of Environmental Analytical Chemistry* (**2013**) DOI: 10.1080/03067319.2013.803545.

Liao, H.; Henze, D. K.; Seinfeld, J. H.; Wu, S.; Mickley, L. J.: Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, *Journal of Geophysical Research* 112 (**2007**) D06201.

Lin, J.J.; Lee, L.-C.: Characterization of n-alkanes in urban submicron aerosol particles (PM₁), *Atmospheric Environment* 38 (**2004**) 2983-2991.

Malinowski, E.R.: Window factor analysis: Theoretical derivation and application to flow injection analysis data , *Journal of Chemometrics* 6 (**1992**) 29-40.

Mas, S.; Tauler, R.; de Juan, A.: Chromatographic and spectroscopic data fusion analysis for interpretation of photodegradation processes, *Journal of Chromatography A* 1218 (**2011**) 51, 9260-9268.

Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J.: *Handbook of Chemometrics and Qualimetrics, Part A*. Amsterdam (**1997**) Elsevier Science.

Matysik, S.; Opitz, P.; Herbarth, O.: Long-term trend of indoor volatile organic compounds (VOC), *WIT Transactions on Ecology and the Environment* 174 (**2013**) 65-73.

Mbengue, S.; Alleman, L.Y.; Flament, P.: Size-distributed metallic elements in submicronic and ultrafine atmospheric particles from urban and industrial areas in northern France, *Atmospheric Research* 135-136 (2014) 35-47.

McCarthy, M.C.; Aklilu, Y.-A.; Brown, S.G.; Lyder, D.A.: Source apportionment of volatile organic compounds measured in Edmonton, Alberta, *Atmospheric Environment* 81 (2013) 504-516.

Medeiros, P. M.; Simoneit, B.R.T.: Analysis of sugars in environmental samples by gas chromatography-mass spectrometry, *Journal of Chromatography A* 1141 (2007) 271-278.

Mochida, M.; Kawabata, A.; Kawamura, K.; Hatsushika, H. Yamazaki, K.: Seasonal variation and origins of dicarboxylic acids in the marine atmosphere over the western North Pacific, *Journal of Geophysical Research: Atmospheres* 108 (2003) 4193.

Mogollon, N.G.S.; Ribeiro, F.A.D.L.; Lopez, M.M.; Hantao, L.W.; Poppi, R.J.; Augusto, F.: Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution, *Analytica Chimica Acta* 796 (2013) 130-136.

Mohr, C.; DeCarlo, P. F.; Heringa, M. F.; Chirico, R.; Slowik, J. G.; Richter, R.; Reche, C.; Alastuey, A.; Querol, X.; Seco, R.; Peñuelas, J.; Jiménez, J. L.; Crippa, M.; Zimmermann, R.; Baltensperger, U.; Prevot, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, *Atmospheric Chemistry and Physics* 12 (2012) 1649–1665.

Moise, T; Rudich, Y.: Reactive Uptake of Ozone by Aerosol-Associated Unsaturated Fatty Acids: Kinetics, Mechanism, and Products, *Journal of Physical Chemistry A* (2002) 106, 469–6476.

Morup, M.; Hansen, L. K.; Arnfred, S. M.: Sparse Higher Order Non-Negative Matrix Factorization, 4658; IMM, Danish Technical University: Lyngby, Denmark, **2006**.

Nawahda, A.; Yamashita, K.; Ohara, T.; Kurokawa, J.; Ohizumi, T.; Chen, F.; Akimoto, H.: Premature mortality in Japan due to ozone, *Atmospheric Environment* 81 (**2013**) 538-545.

Ning, Z.; Geller, M. D.; Moore, K. F. Sheesley, R.; Schauer, J. J.; Sioutas, C.: Daily variation in chemical characteristics of urban ultrafine aerosols and inference of their sources, *Environmental Science and Technologies* (**2007**) 41, 6000–6006.

O'Dowd, C.D.; Aalto, P.; Hämeri, K.; Kulmala, M.; Hoffmann, T.: Atmospheric particles from organic vapours, *Nature*, 416 (**2002**) 497.

Orozco, C; Pérez, A.; González, M^aN.; Rodríguez, F.J.; Alfayate, J.M.: Contaminación ambiental, una visión desde la química; Ed. Thomson, **2003**

Paatero, P. and Tapper, U.: Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics* 5 (**1994**) 111–126.

Pankow, J. F., An absorption model of gas-particle partitioning of organic compounds in the atmosphere. *Atmospheric Environment*, 28 (**1994a**) 2, 185-188.

Parastar, H.; Radović, J.R.; Bayona, J.M.; Tauler, R.: Solving chromatographic challenges in comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry using multivariate curve resolution-alternating least squares ABC Highlights: Authored by Rising Stars and Top Experts, *Analytical and Bioanalytical Chemistry* 405 (**2013**) 19, 6235-6249.

Pardo, R.; Vega, M.; Barrado, E.; Castrillejo, Y.; Sánchez, I.: Three-way principal component analysis as a tool to evaluate the chemical stability of metal bearing residues from wastewater treatment by the ferrite process, *Journal of Hazardous Materials* 262 (2013) 71-82.

Park, C.; Schade, G. W.; Boedeker, I.: Characteristics of the flux of isoprene and its oxidation products in an urban area, *Journal of Geophysical Research: Atmospheres* 116 (2011) D21303.

Penkett, S.A., Jones, B.M., Brice, K.A., Eggleton, A.E.: The importance of atmospheric ozone and hydrogen peroxide in oxidizing sulfur dioxide in cloud and rain water. *Atmospheric Environment* 13 (1979) 123–137.

Peré-Trepat, E.; Ginebreda, A.; Tauler, R.; Comparison of different multiway methods for the analysis of geographical metal distributions in fish, sediments and river waters in Catalonia, *Chemometrics and Intelligent Laboratory Systems* 88 (2007) 1, 69-83.

Piqueras, S.; Duponchel, L.; Tauler, R.; De Juan, A.: Resolution and segmentation of hyperspectral biomedical images by Multivariate Curve Resolution-Alternating Least Squares, *Analytica Chimica Acta* 705 (2011)1-2, 182-192.

Reche, C.; Moreno, T.; Amato, F.; Viana, M.; van Drooge, B.L.; Chuang, H.; Bérubé, K.; Jones, T.; Alastuey, A.; Querol, X. A multidisciplinary approach to characterise exposure risk and toxicological effects of PM₁₀ and PM_{2.5} samples in urban environments. *Ecotoxicology and Environmental Safety* 78 (2012) 327-335.

Robinson, A. L.; Subramanian, R.; Donahue, N. M.; Bernardo-Bricker, A.; Rogge, W. F.: Source apportionment of molecular markers and organic aerosol – 3. food cooking emissions, *Environmental Science & Technology* 40 (2006) 7820–7827.

Rogge, W. F., Hildemann, L.M.; Mazurek, M.A.; Cass; G.R.; Simoneit, B.R.T.: Sources of Fine Organic Aerosol. 1. Charbroilers and Meat Cooking Operations. *Environmental Science & Technology* 25 (1991) 6: 1112-25.

Rogge, W. F., Hildemann, L.M.; Mazurek, M.A.; Cass; G.R.; Simoneit, B.R.T.: Sources of Fine Organic Aerosol. 3. Road Dust, Tire Debris, and Organometallic Brake Lining Dust - Roads as Sources and Sinks, *Environmental Science & Technology* 27 (1993a) 9: L1892-904.

Rogge, W. Mazurek, M. A., Hildemann, L. M, Cass, G. R. and Simoneit, B. R. T.: Quantification of urban organic aerosols at molecular level: identification, abundance and seasonal variation. *Atmospheric Environment* 27A (1993b) 1309-1330.

Rogge, W. F.; Hildemann, L. M.; Mazurek, M. A.; Cass, G. R.; Simoneit, B. R. T.: Sources of fine organic aerosol. 2. Noncatalyst and catalyst-equipped automobiles and heavy duty diesel trucks, *Environmental Science & Technology* (1993c) 27, 636–651.

Rogge, W.F.; Hildemann, L.M.; Marurek, M.A.; Glen, R.; Cass, G.R.; Simoneit, B.R.: Sources of Fine Organic Aerosol. 6. Cigarette Smoke in the Urban Atmosphere, *Environmental Science and Technology* 28 (1994) 1375-1388.

Rogge, W. F., Hildemann, L.M.; Mazurek, M.A.; Cass; G.R.; Simoneit, B.R.T.: Sources of Fine Organic Aerosol. 9. Pine, Oak and Synthetic Log Combustion in Residential Fireplaces. *Environmental Science & Technology* 32 1 (1998): 13-22.

Ruckebusch, C.; Duponchel, L.; Sombret, B.; Huvenne, J.P.; Saurina, J.: Time-Resolved Step-Scan FT-IR Spectroscopy: Focus on Multivariate Curve Resolution, *Journal of Chemical Information and Computer Sciences* 43 (2003) 6, 1966-1973.

Rudich, Y.; Donahue, N. M.; Mentel, T. F.: Aging of organic aerosol: Bridging the gap between laboratory and field studies. *Annual Review of Physical Chemistry* 58 (2007), 321-352.

Salau, J. S. I., Tauler, R., Bayona, J. M., and Tolosa, I.: Input Characterization of Sedimentary Organic Contaminants and Molecular Markers in the Northwestern Mediterranean Sea by Exploratory Data Analysis, *Environmental Science & Technology* 31 (1997) 3482–3490.

Sánchez, E.; Kowalski, B.R.: Generalized rank annihilation factor analysis [8], *Analytical Chemistry* 58 (1986) 2, 496-499.

Sánchez, E.; Kowalski, B.R.: Tensorial resolution: a direct trilinear decomposition, *Journal of chemometric*, 4 (1990) 29-45.

Savorani, F.; Tomasi, G.; Engelsen, S.B.: Icoshift: a versatile tool for the rapid alignment of 1D NMR spectra, *Journal of Magnetic Resonance* 202 (2010) 190–202.

Schauer, J. J., Kleeman, M. J., Cass, G. R., and Simoneit, B. R. T.: Measurement of emissions from air pollution sources. 2. C- 1 through C-30 organic compounds from medium duty diesel trucks, *Environmental Science & Technology* 33 (1999) 1578–1587.

Schauer, J. J., Kleeman, M. J., Cass, G. R., and Simoneit, B. R. T.: Measurement of emissions from air pollution sources. 3. C-1-C- 29 organic compounds from fireplace combustion of wood, *Environmental Science & Technology* 35 (2001) 1716–1728.

Schauer, J.J.; Kleemann, M.J.; Cass, G.R.; Simoneit, B.R.T.: Measurements of emissions from air pollution sources, 5. C1-C32 organic compounds from

gasoline-powered motor vehicles, *Environmental Science & Technology* 36 (2002) 1169-1180.

Schauer, J. J.; Rogge, W. F.; Hildemann, L. M.; Mazurek, M. A.; Cass, G. R.: Source apportionment of airborne particulate matter using organic compounds as tracers, *Atmospheric Environment* (2007) S241-S259.

Seco, R.; Peñuelas, J.; Filella, I.; Llusia, J.; Schallhart, S.; Metzger, A.; Muller, M.; Hansel, A.: Volatile organic compounds in the Western Mediterranean Basin: urban and rural winter measurements during the DAURE campaign, *Atmospheric Chemistry and Physics Discussions* 12 (2012) 30909–30950.

Simoneit, B.R.T.: Characterization of organic constituents in aerosols in relation to their origin and transport, *International Journal of Environmental Analytical Chemistry* 23 (1986) 207-237.

Simoneit, B.R.T; Guoying, S.; Chen, X.; Fu, J.; Zhang, J., Molecular marker study of extractable organic matter in aerosols from urban areas of China, *Atmospheric Environment*, 25 (1991) 2111-2129

Simoneit, B.R.T. Biomass burning – a review of organic tracers for smoke from incomplete combustion, *Applied Geochemistry* 17 (2002) 129-162.

Skov, T.; van den Berg, F.; Tomasi, G.; Bro, R.: Automated alignment of chromatographic data, *Journal of Chemometrics* 20 (2006) 484–497.

Smilde, A.; Bro, R.; Geladi, P. Multi-way analysis. John Wiley & Sons Ltd. Chichester, England, 2004.

Spiro, T.G; Stigliani, W.N. Química Medioambiental; Ed. Pearson, Prentice Hall, 2004

Stanimirova, I.; Tauler, R.; Walczak, B.: A Comparison of Positive Matrix Factorization and the Weighted Multivariate Curve Resolution Method. Application to Environmental Data, *Environmental Science and Technology* 45 (2011) 10102–10110.

Stellman, J.M.; McCan, M: Hidrocarburos poliaromáticos. Enciclopedia de Salud y Seguridad en el Trabajo. 3ª edición. Madrid: Organización Internacional del Trabajo. Ministerio de Trabajo y Asuntos Sociales - Subdirección General de Publicaciones. 1998. p. 310-9.

Stone, E. A.; Zhou, J.; Snyder, D. C.; Rutter, A. P.; Mieritz, M.; Schauer, J. J.: A comparison of summertime secondary organic aerosol source contributions at contrasting urban locations, *Environmental Science and Technology* 43 (2009) 3448–3454.

Subramanian, R.; Donahue, N. M.; Bernardo-Bricker, A.; Rogge, W. F.; Robinson, A. L.: Contribution of motor vehicle emissions to organic carbon and fine particle mass in Pittsburgh, Pennsylvania: Effects of varying source profiles and seasonal trends in ambient marker concentrations, *Atmospheric Environment* (2006) 40, 8002–8019.

Sullivan, A. P.; Peltier, R.E.; Brock, C.A.; Gouw, J.A.; Holloway, J.S.; Warneke, C; Wollny, A.G.; Weber, R.J.: Airborne Measurements of Carbonaceous Aerosol Soluble in Water over Northeastern United States: Method Development and an Investigation into Water-Soluble Organic Carbon Sources. *Journal of Geophysical Research-Atmospheres* 111 D23 (2006a). DOI: 10.1029/2006jd007072

Sullivan, A. P.; Weber, R. J.: Chemical characterization of the ambient organic aerosol soluble in water: 1. Isolation of hydrophobic and hydrophilic fractions with a XAD-8 resin. *Journal of Geophysical Research-Atmospheres*, 111 (2006b) (D5), 9.

Surratt, J.D.; Chan, A.W.H.; Eddingsaas, N.C.; Chan, M.N.; Loza, C.L.; Kwan, A.J.; Hersey, S.P.; Flagan, R.C.; Wennberg, P.O.; Seinfeld, J.H.: Reactive intermediates revealed in secondary organic aerosol formation from isoprene. *Proceedings of the National Academy of Sciences of USA* 107 (2010) 6640-6645.

Suykens, J.: Least Squares Support Vector Machines: Regression and Classification, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Ed. Elsevier, 2009, Capítol 3.11, 445.

Szmigielski, R.; Surratt, J.D.; Gómez-González, Y.; Van der Veken, P.; Kourtchev, I.; Vermeylen, R.; Blockhuys, F.; Jaoui, M.; Kleindienst, T.E.; Lewandowski, M.; Offenberg, J.H.; Edney, E.O.; Seinfeld, J.H.; Maenhaut, W.; Claeys, M.: 3-Methyl-1,2,3-butanetricarboxylic acid: an atmospheric tracer for terpene secondary organic aerosol, *Geophysical Research Letters* 34 (2007) L24811, doi:10.1029/2007GL03133,.

Tauler, R.: Multivariate curve resolution applied to second order data, *Chemometrics and Intelligent Laboratory Systems* 30 (1995a) 133–146.

Tauler, R., Smilde, A., and Kowalski, B.: Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution, *Journal of Chemometrics* 9 (1995b) 31–58.

Tauler, R.; Marques, I.; Cassassas, E. Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths. *Journal of Chemometrics* 12 (1998) 55.

Tauler, R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics* 15 (2001) 627–646.

Tauler, R., Viana, M., Querol, X., Alastuey, A., Flight, R. M., Wentzell, P. D., and Hopke, P. K.: Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, *Atmospheric Environment* 43 (2009) 31–58.

Ten Berge, J.M.F.; de Leuw, J.; Kroonenberg, P.M.: Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika* 52 (1987) 183-191.

Ten Berge, J.M.F.: Convergence of PARAFAC preprocessing procedures and the Deming-Stephan method of iterative proportional fitting, en *Multiway data analysis*, Coppi, R.; Bolasco, S. (Eds), Elsevier, Amsterdam, 1998, pàg. 53-63.

Terrado, M., Barcelo, D., Tauler, R.: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns in surface water, *Environmental Science & Technology* 43 (2009), 5321–5326.

Todeschini, R.: Weighted k-nearest neighbour method for the calculation of missing values, *Chemometrics and Intelligent Laboratory Systems* 9 (1990) 2, 201-205.

Tucker, L.R. Some mathematical notes on 3-mode factor analysis. *Psychometrika* 31 (1966) 279

US-EPA: Protocol for applying and validating the CMB model, edited by: Office for Air Quality Planning and Standards, 1987

Vandeginste, B.G.M; Massart, D.L; Buydens, L.M.C.; de Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J.: *Handbook of Chemometrics and Qualimetrics*, 2nd ed., Elsevier Science B.V, The Netherlands, 1998.

van der Kloot, W.A.; Kroonenberg, P.M.: External analysis with three-mode principal component analysis, *Psychometrika* 50 (1985) 479–494.

van Drooge, B.L.; Crusack, M.; Reche, C.; Mohr, C.; Alastuey, A.; Querol, X.; Prevot, A.; Day, D.A.; Jimenez, J.L.; Grimalt, J.O.: Molecular marker characterization of the organic composition of submicron aerosols from Mediterranean urban and rural environments under contrasting meteorological conditions, *Atmospheric Environment* 61 (2012), 482-489.

Vera, L.; Aceña, L.; Guasch, J.; Boqué, R.; Mestres, M.; Busto, O.: Discrimination and sensory description of beers through data fusion, *Talanta* 87 (2011) 136–142.

Vesna, O.; Sax, M.; Kalberer, M.; Gaschen, A.; Ammann, M., Product study of oleic acid ozonolysis as function of humidity, *Atmospheric Environment*, 43 (2009) 3662-3669.

Vilanova, R.; Fernández, P.; Martínez, C.; Grimalt, J.O. Polycyclic aromatic hydrocarbons in remote mountain lake waters. *Water Research* 35 (2001) 3916-3926.

Vives, I.; Grimalt, J.O.; Guitart R.: Los hidrocarburos aromáticos policíclicos y la salud humana. *Apuntes de Ciencia y Tecnología*. Nº 3, septiembre 2001. p. 45-51.

Volkamer, R.; Jimenez, J.L.; Dzepina, K.; Salcedo, D.; San Martini, F.M.; Molina, L.T.; Worsnop, D.R.; Molina, M.J.: Secondary organic aerosol formation from anthropogenic air pollution: rapid and higher than expected. *Geophysical Research Letters* 33 (2006) L17811.

Walczak, B.; Massart, D.L.: Dealing with missing data: Part I, *Chemometrics and Intelligent Laboratory Systems* 58 (2001a) 1, 15-27.

Walczak, B.; Massart, D.L.: Dealing with missing data: Part II, *Chemometrics and Intelligent Laboratory Systems* 58 (2001b) 1, 29-42.

Wang, W.; Kourtchev, I.; Graham, B.; Cafmeyer, J.; Maenhaut, W.; Claeys, M., Characterization of oxygenated derivatives of isoprene related to 2-methyltetrols in Amazonian aerosols using trimethylsilylation and gas chromatography/ion trap mass spectrometry, *Rapid Communications in Mass Spectrometry* 19 (2005) 1343–1351.

Warneck P. Chemistry of the natural atmosphere. International Geophysics Series. Wiley & Sons. (1988) Vol. 41. Academy Press, pp. 757.

Wentzell, P.D.; Andrews, D.T.; Hamilton, D.C.; Faber, K.; Kowalski, B.R.: Maximum likelihood principal component analysis, *Journal of Chemometrics* 11 (1997) 4, 339–366.

Wentzell, P.D; Karakach, T.K.; Roy, S; Martinez, M.J.; Allen, C.P.; Werner-Washburne, M. Multivariate curve resolution of time course microarray data, *BMC Bioinformatics* 7 (2006) 343.

WHO (2002) Guidelines for concentration and exposure-response measurement of fine and ultra fine particulate matter for use in epidemiological studies. WHO Dietrich Schwela, Lidia Morawska, QUT, Dimitrios Kotzias, EC JRC. Published on behalf of the European Commission.

WHO (2010) World Health Statistics, WHO Library Cataloguing-in-Publication Data, http://www.who.int/whosis/whostat/EN_WHS10_Full.pdf

Windig, W.; Stephenson, D.A.: Self-modeling mixture analysis of second-derivative near-infrared spectral data using the SIMPLISMA approach, *Analytical Chemistry* 64 (22) (1992) 2735-2742.

Windig, W.; Gallagher, N.B.; Shaver, J.M.; Wise, B.M.: A new approach for interactive self-modeling mixture analysis, *Chemometric and Intelligent Laboratory Systems* 77 (2005) 85-96.

Wold, S; Esbensen, K; Geladi, P.: Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems* 2 (1987) 37-52.

Yang, L.; Ray, M. B.; Yu, L. E.: Photooxidation of dicarboxylic acids – Part I: Effects of inorganic ions on degradation of azelaic acid, *Atmospheric Environment* 42 (2008) 856–867.

Zeiter, M.; Rutledge, D.: Preprocessing Methods (Capítol 3.04), *Comprehensive Chemometrics, Volum 3*, 2010, 121-231

Zhang, Q.; Zhang, Q.; Jimenez, J.L.; Canagaratna, M.R.; Allan, J.D.; Coe, H.; Ulbrich, I.; Alfarra, M.R.; Takami, A.; Middlebrook, A.M.; Sun, Y.L.; Dzepina, K.; Dunlea, E.; Docherty, K.; DeCarlo, P.F.; Salcedo, D.; Onasch, T.; Jayne, J.T.; Miyoshi, T.; Shimono, A.; Hatakeyama, S.; Takegawa, N.; Kondo, Y.; Schneider, J.; Drewnick, F.; Borrmann, S.; Weimer, S.; Demerjian, K.; Williams, P.; Bower, K.; Bahreini, R.; Cottrell, L.; Griffin, R.J.; Rautiainen, J.; Sun, J.Y.; Zhang, Y.M.; Worsnop, D.R.: Ubiquity and Dominance of Oxygenated Species in Organic Aerosols in Anthropogenically-Influenced Northern Hemisphere Midlatitudes, *Geophysical Research Letters* 34 (2007) 13.

Zhang, Y. Y.; Muller, L. Winterhalter, R.; Moortgat, G. K.; Hoffmann, T.; Poschl, U.: Seasonal cycle and temperature dependence of pinene oxidation products, dicarboxylic acids and nitrophenols in fine and coarse air particulate matter, *Atmospheric Chemistry and Physics* (2010) 10, 7859–7873.

Zielinska, B.; Sagebiel, J.; McDonald, J. D.; Whitley, K.; Lawson, D. R.: Emission rates and comparative chemical composition from selected in-use diesel and gasoline-fueled vehicles, *Journal of Air Waste Management* (2004) 54, 1138–1150.

Zhou, Z.; Guo, L.; Shiller, A.M.; Lohrenz, S.E.; Asper, V.L.; Osburn, C.L.: Characterization of oil components from the Deepwater Horizon oil spill in the Gulf of Mexico using fluorescence EEM and PARAFAC techniques, *Marine Chemistry* 148 (2013) 10-21.

