



European
Commission

J R C R E F E R E N C E R E P O R T S



European Guide on Air Pollution Source Apportionment with Receptor Models

Claudio A. Belis, Bo R. Larsen, Fulvio Amato,
Imad El Haddad, Olivier Favez, Roy M. Harrison,
Philip K. Hopke, Silvia Nava, Pentti Paatero,
André Prévôt, Ulrich Quass, Roberta Vecchi,
Mar Viana

2014

Report EUR 26080 EN

Joint
Research
Centre

European Commission

Joint Research Centre
Institute for Environment and Sustainability

Contact information

Claudio A. Belis
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 441, 21027 Ispra (VA), Italy
E-mail: claudio.belis@jrc.ec.europa.eu
Tel.: +39 0332 78 6644
Fax: +39 0332 78 5236

<http://www.jrc.ec.europa.eu>

This publication is a Reference Report by the Joint Research Centre
of the European Commission.

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission
is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers to your questions about the European Union
Freephone number (*): 00 800 6 7 8 9 10 11

(* Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server <http://europa.eu/>.

JRC83309

EUR 26080 EN

ISBN 978-92-79-32513-7 (pdf) ISBN 978-92-79-32514-4 (print)
ISSN 1831-9424 (online) ISSN 1018-5593 (print)

doi: 10.2788/9307

Luxembourg: Publications Office of the European Union, 2014

© European Union, 2014

Reproduction is authorised provided the source is acknowledged.

Printed in Italy

European Guide on Air Pollution Source Apportionment with Receptor Models

Drafting committee

Claudio A. Belis¹, chair and editor

*Bo R. Larsen², Fulvio Amato^{3,4}, Imad El Haddad⁵, Olivier Favez⁶,
Roy M. Harrison⁷, Philip K. Hopke⁸, Silvia Nava⁹, Pentti Paatero¹⁰,
André Prévôt⁵, Ulrich Quass¹¹, Roberta Vecchi¹², Mar Viana³*

¹ European Commission – Joint Research Centre, Institute for Environment and Sustainability. Ispra, Italy.

² European Commission – Joint Research Centre, Institute for Health and Consumer Protection. Ispra, Italy.

³ Institute of Environmental Assessment and Water Research (IDAEA), Barcelona, Spain

⁴ TNO, Built Environment and Geosciences, Dept Air Quality and Climate, Utrecht, The Netherlands

⁵ Laboratory of Atmospheric Chemistry (LAC), Paul Scherrer Institut, Villigen, Switzerland

⁶ Institut National de l'Environnement Industriel et des Risques (INERIS), Verneuil-en-Halatte, France

⁷ National Centre for Atmospheric Science, University of Birmingham, Birmingham, UK

⁸ Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY, USA

⁹ Department of Physics, Università degli Studi di Firenze and INFN-Firenze, Sesto Fiorentino, Italy

¹⁰ Department of Physics, University of Helsinki, Helsinki, Finland

¹¹ Air Quality and Sustainable nanotechnology (IUTA), D- Duisburg, Germany

¹² Department of Physics, Università degli Studi di Milano and INFN, Milan, Italy

TABLE OF CONTENTS

JRC REFERENCE REPORT – POLICY SUMMARY	5
SUMMARY	9
PART A: INTRODUCTION TO SOURCE APPORTIONMENT WITH RECEPTOR MODELS	13
PART B: HARMONISED RECEPTOR MODEL PROTOCOL	21
B1. PRELIMINARY EVALUATION OF THE STUDY AREA	21
B2. DEFINING A METHODOLOGICAL FRAMEWORK	23
B3. EXPERIMENT DESIGN - CRITERIA FOR SITE AND SPECIES SELECTION AND ESTIMATION OF MINIMUM NUMBER OF SAMPLES	25
B4. DATA COLLECTION / FIELD WORK / CHEMICAL ANALYSIS	31
B5. KNOWING YOUR DATASET: BASIC STATISTICS	35
B6. PRELIMINARY DATA QUALITY CHECKS	39
B7. INPUT DATA UNCERTAINTY CALCULATION	43
B8. CHEMICAL MASS BALANCE MODELS	47
B9. FACTOR ANALYSIS I: SELECTION OF THE NUMBER OF FACTORS AND DEALING WITH ROTATIONAL AMBIGUITY (PMF)	51
B10. FACTOR ANALYSIS II: EVALUATION OF SOURCE CONTRIBUTION ESTIMATION AND MODEL PERFORMANCE INDICATORS	55
B11. FACTOR ANALYSIS III: CRITERIA FOR FACTOR ASSIGNMENT	59
B12. TESTS FOR MODEL PERFORMANCE VALIDATION	61
B13. REPORTING RESULTS AND METHODOLOGY	65
PART C: ADVANCED MODELS	67
C1. WIND AND TRAJECTORY ANALYSIS IN SOURCE APPORTIONMENT	67
C2. THE USE OF PMF and ME-2 IN AEROSOL MASS SPECTROMETER DATA PROCESSING	71
C3. THE AETHALOMETER MODEL	75
C4. APPORTIONMENT OF THE PM CARBONACEOUS FRACTION: RADIOCARBON AND TRACER ANALYSIS	79
C5. CONSTRAINED AND EXPANDED MODELS IN FACTOR ANALYSIS	85

JRC REFERENCE REPORT – POLICY SUMMARY

Title of the Reference Report

EUROPEAN GUIDE ON POLLUTION SOURCE IDENTIFICATION WITH RECEPTOR MODELS

Authors

Claudio A. Belis, European Commission - JRC, Bo R. Larsen, European Commission - JRC, Fulvio Amato, IDAEA, Imad El Haddad, Paul Scherrer Institute, Olivier Favez, INERIS, Roy Harrison, University of Birmingham, Philip K. Hopke, Clarkson University, Silvia Nava, University of Florence, Pentti Paatero, University of Helsinki, André Prévôt, Paul Scherrer Institute, Ulrich Quass, IUTA, Roberta Vecchi, University of Milan, Mar Viana, IDAEA

Keywords

pollution sources, harmonisation, receptor models, air quality directives, pollution abatement measures, thematic strategy on air pollution

Introduction

This report contains a guide and a European harmonised protocol prepared within the framework of the JRC initiative for the harmonisation of source apportionment with receptor models. This initiative was launched in collaboration with the European networks in the field of air quality modelling (FAIRMODE) and measurements (AQUILA). The initiative also includes a review of the methodologies used in Europe for source identification and intercomparison exercises for the quantitative assessment of the performance of source apportionment models.

The document, drafted by a group of international experts, is organised following the logical sequence of steps to be carried out in a source apportionment study. Sections with increasing levels of complexity make it accessible to readers with different degrees of familiarity with this topic, from air quality managers to air pollution experts and modellers. It has been conceived as a reference document that includes tutorials, technical recommendations and check lists.

EU policy context and importance of the issue

The abatement of pollution at its source is one of the overarching principles of the Thematic Strategy on Air Pollution (TSAP; Dir. 2008/50/EC, preamble). Reliable and quantitative information on pollution sources is essential for the implementation of the Air Quality Directives (AQD: Dir. 2008/50/EC and Dir. 2004/107/EC). For instance, pollution source information is required for identifying whether exceedances are due to natural sources or to road salting and sanding (arts. 20 and 21), preparing air quality plans (Annex XV A), quantifying transboundary pollution (Annex IV A), informing the public (Annex XVI) and, in the past, for demonstrating eligibility for postponement of PM₁₀ and NO₂ limit value attainment (COM/2008/403).

Source Apportionment (SA) is the identification of ambient air pollution sources and the quantification of their contribution to pollution levels. This task can be accomplished using different approaches: emission inventories, source-oriented models and receptor-oriented models.

Goals/objectives of the report

The objective of this document is to disseminate and promote the best available methodologies for source identification using receptor models, and to harmonise their application across Europe.

In addition, it aims at making results of source apportionment studies more accessible to experts involved in the development and assessment of pollution source abatement measures.

Methodology

Receptor models (RMs) apportion the measured mass of an atmospheric pollutant at a given site to its emission sources by solving a mass balance equation. These models have the advantage of providing information derived from real-world measurements, including

estimations of output uncertainty. However, their applicability to very reactive species is limited. RMs are extensively used for source contribution quantification at local and regional scales all over the world. In the past decade, the number of scientific publications and applications in this field has been increasing steadily, and tools have been developed with constantly improving capabilities in terms of source resolution and the accuracy of source contribution quantification (Belis et al., 2013).

The protocol presented in this document focuses on the most commonly used RMs: Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF). The CMB model is a 'least squares' model which estimates source contributions on the basis of the emissions' chemical composition (fingerprints) and the concentration of pollutants. The PMF model is based on uncertainty-weighted factor analysis which relies on pollutant measurements.

In addition, to promote the development and application of state-of-the-art methodologies, a section is also included on innovative and advanced methods, most of which are under continuous development. This section comprises trajectory and wind-based models, constrained and expanded models, the Aethalometer model, and models based on advanced spectrometric measurements and isotopic analyses.

Key results, deliverables, key messages

Due to the complexity of source apportionment studies, it is essential to support the final results with an appropriate description of the methodological choices available and with documentation of the objective qualitative or quantitative information that supports expert decisions. In this way, reviewers and final users (e.g. air quality managers) are provided with the elements they need to assess the relevance of the study, and other modellers have the possibility to reproduce the same approach. It is essential that only methodologies fulfilling quality standards that are in line with the objectives of the study are adopted. To that end, the information about models' performance collected in the above-mentioned intercomparison exercises provides the necessary complement to the procedures described in this document. These exercises have demonstrated that RMs provide quantitative estimations of contributions by source category with at most 50% uncertainty (Karagulian et al., 2012). It follows that SA studies that are consistent with the present protocol, in particular with regard to the quality assurance steps, can claim state-of-the-art performance in line with that observed in European-wide intercomparison exercises.

Real/potential impact and benefits to customers, users, and stakeholders

1. Quantitative estimations of pollution sources obtained with reliable and harmonised methods across Europe constitute a fundamental input for the different actors involved in the implementation of the Air Quality Directives at the local and regional scales. The present document helps to streamline the technical criteria required for accomplishing such a complex task according to the best available standards, with a view to improving the transparency and comparability of results obtained by different practitioners in different areas of Europe.
2. Expected benefits of the report for different target groups:
 - the report is intended to be a reference for practitioners, providing them with clear and widely accepted criteria for model execution and the interpretation of results;
 - final users of pollution source data, such as authorities involved in air quality management, would have access to transparent and comparable information obtained with known quality standards that can be used as input data in scenario- or cost-benefit analyses;
 - harmonisation would have a positive impact on the quality and comparability of data reported by Member States to the Commission under the scheme for reciprocal exchange of information and reporting on ambient air quality (Commission Implementing Decision 2011/850/EU);
 - the report is an information dissemination tool for air quality managers and atmospheric scientists that are not familiar with this methodology.
3. In addition, the synergy between the harmonised technical protocol and the intercomparison exercises provides the basis for the continuous improvement of source identification approaches in order to keep abreast of the scientific developments in this field.
4. All the methodologies for source identification have strengths and limitations. Considering that RMs deliver independent estimates of source contributions at a given site, they can also be used for the validation of other methodologies such as emission inventories and air quality models.

References

2011/850/EU. Commission implementing decision of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality. (notified under document C(2011) 9068)(2011/850/EU).

Belis C. A., Karagulian F., Larsen B. R., Hopke P. K., 2013. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmospheric Environment* 69, 94-108.

COM(2008) 403. Communication from the Commission on notifications of postponements of attainment deadlines and exemptions from the obligation to apply certain limit values pursuant to Article 22 of Directive 2008/50/EC on ambient air quality and cleaner air for Europe.

Karagulian F., Belis C. A., Borowiak A., 2012. Results of the European Intercomparison exercise for receptor models 2011-2012. Part I. JRC Report EUR 25727 EN.

Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air.

SUMMARY

Abatement of pollution at its source is one of the overarching principles of the Thematic Strategy on Air Pollution (TSAP; Dir. 2008/50/EC, preamble). Reliable and quantitative information on pollution sources is essential for the implementation of the Air Quality Directives (AQD: Dir. 2008/50/EC and Dir. 2004/107/EC). For instance, pollution source information is required for identifying whether exceedances are due to natural sources or to road salting and sanding (arts. 20 and 21), preparing air quality plans (Annex XV A), quantifying trans-boundary pollution (Annex IV A), informing the public (Annex XVI), and demonstrating eligibility for the postponement of PM₁₀ and NO₂ limit value attainment (COM/2008/403).

Source Apportionment (SA) is the practice of deriving information about pollution sources and the amount they contribute to ambient air pollution levels. This task can be accomplished using three main approaches: emission inventories, source-oriented models and receptor-oriented models. The objective of this document is to present the receptor-oriented methodology, explaining its role in the identification of sources with particular reference to particulate matter, and to describe the best practices for the available and emerging methodologies with a view to promoting their harmonisation across Europe.

Receptor-oriented models (also known as receptor models (RMs)) apportion the measured mass of an atmospheric pollutant at a given site, called the receptor, to its emission sources by using multivariate analysis to solve a mass balance equation. These tools have the advantage of providing information derived from real-world measurements, including estimations of output uncertainty. However, there are limitations in their application to very reactive species. RMs are extensively used for the quantification of source contributions at local and regional scales all over the world. In the past decade, the number of scientific publications and applications in this field has been increasing steadily, and tools have been developed with improved capabilities in terms of source resolution and the accuracy of source contribution quantification.

This report is the result of the work of a group of international experts carried out within the framework of the JRC initiative for the harmonisation of source apportionment with receptor models. This initiative was launched in collaboration with the European networks in the field of air quality modelling (FAIRMODE) and measurements (AQUILA). The initiative also includes a review of the methodologies used in Europe for source identification, and intercomparison exercises for the quantitative assessment of the performance of SA models.

The structure of this document follows the logical sequence of steps to be carried out in an SA study. The organisation of the report in sections of increasing levels of complexity makes it accessible to readers with different degrees of familiarity with this topic: from air quality managers to air pollution experts and modelers. The report has been conceived as a reference document that includes tutorials, technical recommendations and check lists. However, it is not intended to substitute practitioners' experience and competence, which can only be acquired through training and working under the supervision of experts.

The core part of the report focuses on the most commonly used RMs: Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF) models. The CMB model is a 'least squares' model which estimates source contributions on the basis of the chemical fingerprints of the source and the concentration of pollutants. The PMF model is based on uncertainty-weighted factor analysis which relies on pollutant measurements.

In addition, to promote the development and application of state-of-the-art methodologies, a section is also included on innovative and advanced methods, most of which are under continuous development. This section comprises trajectory and wind-based models, constrained and expanded models, the Aethalometer model, and models based on advanced spectrometric measurements and isotopic analyses.

Due to the high number of variables to be considered, SA studies are complex. Therefore, it is essential to support the final results with an appropriate description of the methodological choices made and documentation of the qualitative or quantitative information that supports expert decisions. In this way, reviewers and final users, such as local air quality managers, are provided with the elements they need to assess the relevance of every study, and other modellers have the possibility to reproduce the methodology.

Moreover, it is essential that only methodologies fulfilling quality standards that are in line with the objectives of the study are adopted. To that end, the information about models' performance collected in the above-mentioned intercomparison exercises provides the necessary complement to the procedures described in this document. These exercises have demonstrated that RMs provide quantitative estimations of contributions by source categories that are consistent with a 50% standard uncertainty criterion. It follows that SA studies consistent with the present protocol, especially with the steps concerning quality assurance, can claim state-of-the-art performance supported by European-wide intercomparison exercises.

CHAPTER AUTHORS

Part A: C. A. Belis, P.K. Hopke
 Part B:
 Chapter B1. P. K. Hopke, C. A. Belis
 Chapter B2. C. A. Belis
 Chapter B3. C. A. Belis, U. Quass
 Chapter B4. C. A. Belis, O. Favez
 Chapter B5. C. A. Belis, B. R. Larsen, U. Quass
 Chapter B6. C. A. Belis, M. Viana, U. Quass, P. Paatero, R. M. Harrison
 Chapter B7. C. A. Belis, B. R. Larsen, P. Paatero
 Chapter B8. I. El Haddad, R. M. Harrison
 Chapter B9. R. Vecchi, P. Paatero, U. Quass
 Chapter B10. M. Viana, R. Vecchi
 Chapter B11. R. Vecchi
 Chapter B12. F. Amato, B. R. Larsen, P. Paatero
 Chapter B13. C. A. Belis, B. R. Larsen
 Part C:
 Chapter C1. U. Quass, F. Amato
 Chapter C2. A. Prevôt, C. A. Belis
 Chapter C3. O. Favez
 Chapter C4. S. Nava
 Chapter C5. F. Amato

ACKNOWLEDGEMENTS

The authors are grateful to F. Karagulian and G. Calzolari for their contributions to Chapter B7 and C5, respectively. Many thanks to G. Mulhern for proofreading the drafts, and to E. Vignati, M. Bressi, L. Bontoux and D. Buzica for reviewing and commenting on the manuscript.

GLOSSARY

Chemical mass balance (CMB): models that solve the mass balance equation using effective variance least square. These are applied when the number and composition of sources are known.

Degrees of freedom: the number of independent observations minus the number of parameters estimated using them.

Factor: an independent theoretical variable calculated by linearly combining many measured dependent variables in order to describe their relationship patterns.

Factor analytical methods: multivariate techniques which do not require information on the number and composition of sources in the model input. In this document, factor analysis (FA) refers to techniques without intrinsic constraints.

Factor/source: the pollution-emitting entity identified in an SA study. Depending on the type of model used, the output may be a factor (multivariate analysis type) or a source (CMB type).

Kronecker product (denoted by \otimes): an operation performed on two matrices which, unlike the classical matrix multiplication, does not impose limitations on the dimension of the matrices being multiplied.

Multivariate analysis: methods used to deal with datasets consisting of several measurements (variables) for each object (sample unit).

Positive matrix factorization (PMF): a specific type of factor analytical method which uses experimental uncertainty for scaling matrix elements and constrains factor elements to be non-negative.

PM₁₀, PM_{2.5}: particulate matter with aerodynamic diameter equal to or less than 10 and 2.5 micrometres, respectively.

Receptor models (RMs): methodology to apportion the measured mass of air pollutants in one or more sites to their emission sources by solving a mass balance equation using multivariate analysis.

Source: a source of air pollution is any human activity or natural process that causes pollutants to be released into the atmosphere.

Source apportionment (SA): the practice of deriving information about pollution sources and the amount they emit.

Source category: a group of sources that emit pollutants with similar chemical composition and time trends.

Source contribution estimate (SCE): quantitative output of an RM expressed as mass ($\mu\text{g m}^{-3}$) that represents the amount of a pollutant that can be attributed to a specific source or source category.

Source profile or fingerprint: the average relative chemical composition of the particulate matter deriving from a pollution source, commonly expressed as the ratio between the mass of every species to the total PM mass.

ACRONYMS

AMS: Aerosol Mass Spectrometer
 ACSM: Aerosol Chemical Speciation Monitor
 APS: aerodynamic particle sizer
 BDL: below the detection limit
 CEN: European Committee for Standardisation
 CC: carbonatic carbon
 CTMs: chemical transport models
 DL: detection limit
 DRUM/RDI: Davis rotating-drum Universal-size-cut Monitoring impactor
 EC: elemental carbon
 EMEP: European Monitoring and Evaluation Programme
 EPA: Environmental Protection Agency (US)
 GC-MS: gas chromatography coupled with mass spectrometry
 GF-AAS: graphite furnace - atomic absorption spectrometry
 HPLC: high-performance liquid chromatography
 IC: ion chromatography
 ICP-MS: inductively coupled plasma- mass spectrometry
 LS: least squares
 LOD: limit of detection
 OC: organic carbon
 OM: organic matter
 OPC: optical particle counter
 PAHs: polyaromatic hydrocarbons
 PBL: planetary boundary layer
 PIXE: particle-induced X-ray emission
 PM: particulate matter
 POC: primary organic carbon
 RM: receptor model
 SA: source apportionment
 SMPS: scanning mobility particle sizer,
 SOC: secondary organic carbon
 TC: total carbon
 TOR: thermo optical reflectance
 TOT: thermal optical transmission
 VOCs: volatile organic compounds
 XRF: energy dispersive X-ray fluorescence

RECEPTOR MODEL ACRONYMS

APCFA: absolute principal components factor analysis
 APCA: absolute principal component analysis
 CMB: chemical mass balance
 COPREM: constrained physical receptor model
 CPF: conditional probability function
 FA: factor analysis
 ME-2: multilinear engine version 2
 NWR: non-parametric wind regression
 PCA: principal components analysis
 PMF: positive matrix factorization
 PDRM: pseudo deterministic receptor model
 PSCF: potential source contribution function
 SoFi: Source Finder
 SQTBA: simplified quantitative transport bias analysis
 TSA: trajectory sector analysis
 TRMB: trajectory mass balance
 TMBR: trajectory mass balance regression

PART A: INTRODUCTION TO SOURCE APPORTIONMENT WITH RECEPTOR MODELS

European Guide and Harmonised Receptor Model Protocol: driving elements

The objective of this document is to disseminate and promote the best available operating procedures for source apportionment (SA) with receptor models (RMs) and to harmonise their application across Europe.

The target audience is:

- practitioners involved in the model execution and in the interpretation of results,
- air quality managers interested in the output of RMs for the design of abatement measures,
- air quality experts and atmospheric scientists not familiar with this methodology.

The structure of this document follows a logical sequence of steps to be carried out in an SA study, with different levels of complexity accessible to readers with different levels of expertise.

This document has been conceived as a guide (including tutorials, technical recommendations and check lists) that provides relevant references to the original information sources. However, it is not meant to be comprehensive, nor intended to substitute experience and competence. Although the guide aims to promote the highest quality standards, it is subject to the intrinsic limitations of any SA methodology, which lie in the fact that the “true” contribution of sources to atmospheric pollution at a given point cannot be measured directly.

Organisation of the Guide

This document is the result of the collaboration of leading European experts in the field of atmospheric pollution with the support of the P. K. Hopke of Clarkson University, New York, United States. It is structured in three parts.

Part A. Introduction to source apportionment with RMs describes the basic elements of SA and RMs.

Part B. Harmonised Receptor Model Protocol (hereafter referred to as ‘the Protocol’) is the core of the document. It contains a description of the steps to be taken in carrying out the most common and widespread RM techniques, with particular reference to Chemical Mass Balance and Factor Analysis.

Part C. Advanced Models describes innovative and advanced methods, most of which are under continuous development. It also includes methods which, although they have been available for a long time, have not yet been exploited to their full potential.

Identification of pollution sources

Source Apportionment (SA) is the practice of deriving information about pollution sources and the amount they contribute to ambient air pollution levels.

Information on pollution sources is essential to the design of air quality policies and, therefore, SA is required explicitly or implicitly for the implementation of the Air Quality Directives (Dir. 2008/50/EC and Dir. 2004/107/EC). Activities for which identification of pollution sources is relevant include:

- Drawing up action plans
- Assessment of the effectiveness of abatement measures (before and after)
- Application for the postponement of attaining limit values (PM_{10} , NO_2)
- Quantification of pollution arising from:
 - long-range transport
 - transboundary transport
 - natural sources
 - winter sanding and salting

- Identification of sources of pollutants that are of particular interest, e.g. polycyclic aromatic hydrocarbons (PAHs), ozone precursor hydrocarbons, elemental carbon (black carbon).

Different approaches are used to determine and quantify the impacts of air pollution sources on air quality. Commonly used SA techniques are:

- Explorative methods
- Emission inventories
- Inverse modelling
- Artificial neural networks
- Lagrangian models
- Gaussian models
- Eulerian models
- Receptor models

Exploratory methods use simple mathematical relationships and a number of assumptions to achieve a preliminary estimation of the source contribution.

Emission inventories are detailed compilations of the emissions from all source categories in a certain geographical area and within a specific year. Emissions are estimated by multiplying the intensity of each relevant activity (activity rate) by a pollutant-dependent proportionality constant (emission factor).

- In inverse modelling, air quality model parameters are estimated by fitting the model to the observations. The inverse technique consists of a least squares optimisation with an objective function defined as the sum of squared deviations between modelled and observed concentrations.

- Artificial neural networks (ANN) are sets of interconnected simple processing elements (artificial neurons) which can exhibit complex global behaviour. In order to produce a desired signal flow, algorithms designed to modulate the weights of the connections in the network are applied.

- Lagrangian models use a moving frame of reference to describe the trajectories of single or multiple particles as they move in the atmosphere.

- Gaussian plume models assume that turbulent dispersion can be described using a Gaussian distribution profile. This type of model is often used to estimate emissions from industrial sources.

- Eulerian models encompass equations of motion, chemistry and other physical processes that are solved at points arranged on a 3D grid.

Often, the terms ‘dispersion models’ or ‘source oriented models’ are used to refer to the latter three categories. Nevertheless, there are relevant differences in how these models are applied for source identification purposes.

Figure A.1. Schematic representation of the different methods for source identification.

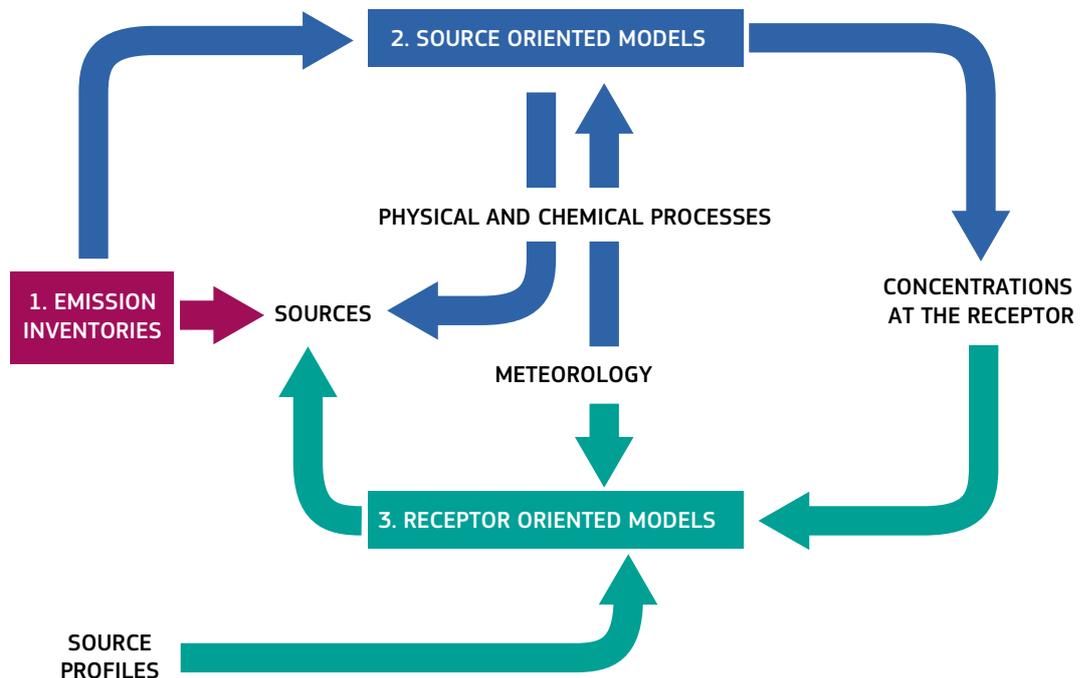


Table A1.
Main characteristics
of RMs.

Use measured concentrations at the receptor (sampling site)
Make reference to the chemical mass balance principle
Are based on the solution of multilinear equations
At the first step do not consider physical and chemical processes, but evolved hybrid models can process additional information to constrain rotational uncertainty
Do not depend on emission inventories; source profiles (fingerprints) are required by certain kinds of RMs
Do not require complex meteorological and chemical processors
Require low computational intensity
Their application with reactive species requires correcting terms
Mainly used on particulate matter (PM) and seldom on hydrocarbons and inorganic gases
Appropriate for urban and regional scales

Receptor models (RMs) focus on the properties of the ambient environment at the point of impact, as opposed to the source-oriented dispersion models which account for transport, dilution, and other processes that take place between the source and the sampling or receptor site (Figure A1).

What are receptor models (RMs)?

The fundamental principle of receptor modelling is that mass conservation between the emission source and the study site can be assumed, and a mass balance analysis can be used to identify and apportion sources of atmospheric pollutants. Table A.1 summarises the main characteristics of RMs.

RMs identify sources by solving the following mass balance equation:

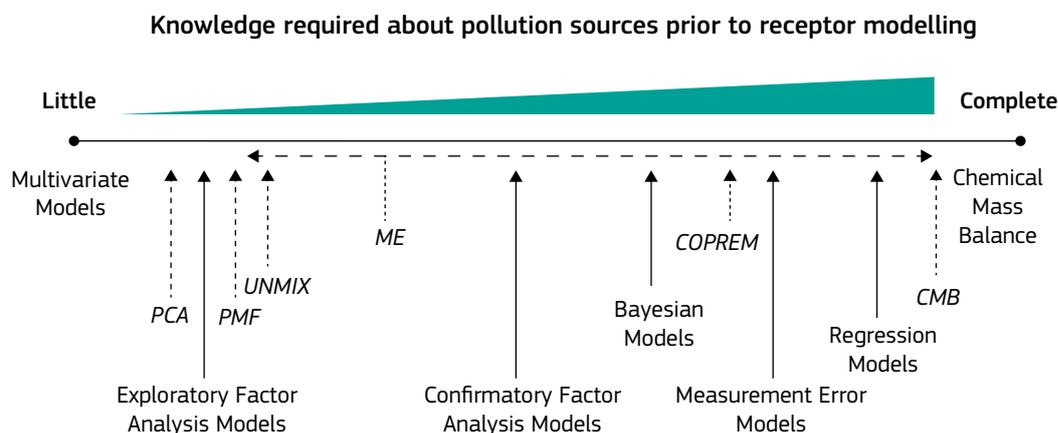
$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (\text{A.1})$$

where x_{ij} is the concentration of the j^{th} species in the i^{th} sample, g_{ik} the contribution of k^{th} source to the i^{th} sample, f_{kj} the concentration of the j^{th} species in the k^{th} source, and e_{ij} is the residual (i.e. the difference between the measured and fitted value) term.

In order to find the solution, a dataset with a rather large amount of data consisting of chemical constituents (such as elemental concentrations) gathered from a number of observations (samples) is required. The larger the data matrix, the higher the chances that the model will identify distinct factors that can be identified as sources.

If the number and nature (composition profiles/fingerprints) of the sources in the study area are known (f_{kj}), then the only unknown term of equation (A.1) is the mass contribution of each source to each sample, g_{ik} . Solving the mass balance equation in this way was first independently suggested by Winchester and Nifong (1971) and by Miller et al. (1972). The problem is typically solved using an effective-variance least-squares approach that is now generally referred to as the chemical mass balance (CMB) model (Watson, 1979, 1984). Since then, many models and methodologies have been developed and are still under continuous evolution. RMs have been traditionally classified into those which explicitly use information about the emission fingerprints (described above) and those which do not use any *a priori* information on source chemical profiles (factor analysis methods).

Figure A.2. Approaches to estimate pollution sources with RMs (from Viana et al., 2008)



The main types of RMs are presented in Figure A.2 and Table A.2. A more detailed description and discussion of the most common RMs can be found in Watson et al. (2008), Viana et al. (2008), Hopke (2010), and Belis et al. (2013).

In the US, RMs are officially recognised and promoted as tools for air quality management (US-EPA SCRAM). Dedicated monitoring networks exist and a number of tools were developed and are freely distributed by the US-EPA. RMs are also used extensively in Europe, although the lack of a common approach and documented performance limits their application to air quality policy.

The role of RMs in the identification of pollution sources

Within the activities of the Forum for Air Quality Modelling in Europe (FAIRMODE) group on “Contribution of natural sources and source apportionment”, two surveys were carried out on the type and frequency of modelling tools that are used in Europe for source apportionment (Fragkou et al., 2012). The most recent of these surveys collected information on the use

of models for the source apportionment of regulated pollutants and on the procedures used to evaluate the applied methodologies. The use of the different tools for source identification ranged from less than 20% for Gaussian models to almost 60% for receptor models (Figure A.3). Lagrangian (e.g. Lagrangian particle dispersion models) and Trajectory models were less frequently used and always complementary to other models. The use of CFD models was only reported in one case.

A study by Viana and co-authors carried out an overview of source apportionment studies in Europe from 1987 to 2007 by compiling meta-data on 71 studies (see Table 1 page 831 of Viana et al., 2008) based on a questionnaire and existing publications.

According to this study, PCA was the most frequently used model up to 2005 (30% of the studies), followed by the ‘Lenschow approach’ or incremental concentrations approach (11%) and back-trajectory analysis (11%). An increase in the use of PMF (13%) and the mass balance analysis of chemical components (19%) was observed from 2006 onwards.

Table A.2. Types of RM (adapted from Belis et al., 2013).

Type	Examples
Exploratory methods	Enrichment factor, tracer method, incremental approach
Chemical Mass Balance	EPA CMB 8.2
Eigenvector-based models	PCA, UNMIX
Factor analysis without constraints	FA, APCFA
Positive matrix factorization	PMF2, EPA PMF v3
Hybrid trajectory-based models	CPF, PSCF
Hybrid expanded models	PMF solved with ME-2, COPREM

Legend: CMB, chemical mass balance; PCA, principal components analysis; FA, factor analysis; APCFA, absolute principal component factor analysis; PMF, positive matrix factorization; ME, multilinear engine; CPF, conditional probability function; PSCF, potential source contribution function; COPREM: Constrained physical receptor model.

Percentage of model type used for SA by different EU countries

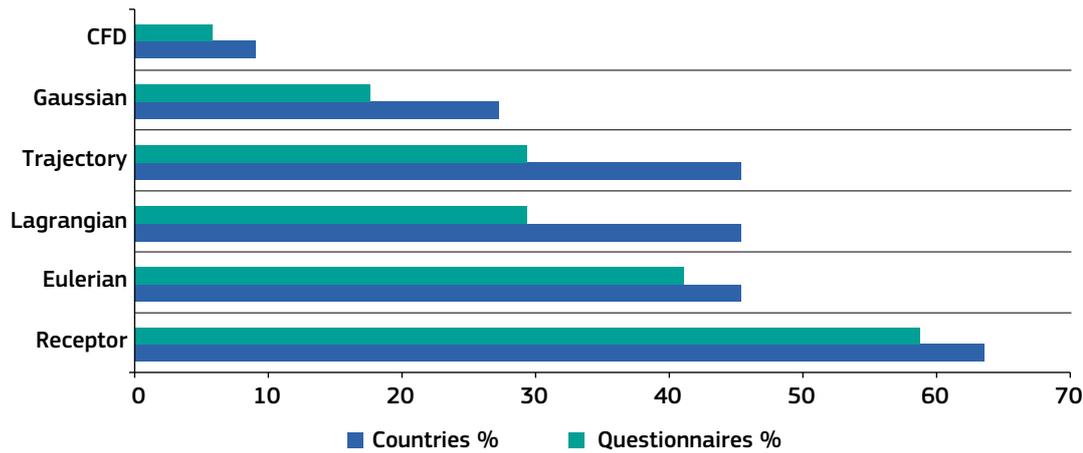


Figure A.3: Percentage of model types used for SA by different EU countries (from Fragkou et al., 2012).

PM₁₀ was the preferred target metric (46%) followed by PM_{2.5} (33%) and coarse fraction (PM_{2.5-10}; 9%). The majority of the studies were carried out in urban background locations (53% of the studies) while industrial or kerbside sites represented 11% and 20% of the studies, respectively.

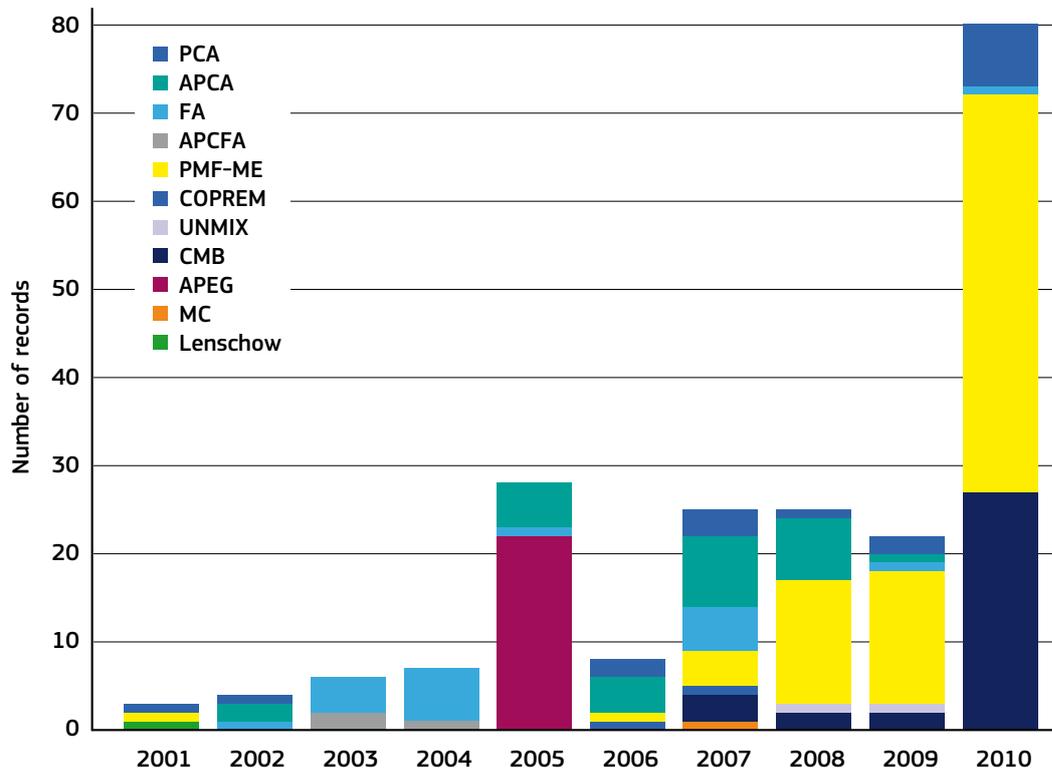
Overall, a generally good spatial coverage of SA studies over Europe, especially regarding the northern, south-eastern and south-western dimensions, was observed.

In this review, four main source categories across Europe were identified:

- Traffic sources, characterised by Carbon/Fe/Ba/Zn/Cu, often including road dust;
- Mineral/crustal matter sources with Al/Si/Ca/Fe as distinctive components;
- Sea-salt, sea-spray and marine sources associated with high Na/Cl/Mg concentrations;
- Regional-scale pollution and long-range transboundary anthropogenic pollution sources rich in either vanadium/nickel/sulphate or sulphate/nitrate/ammonium.

A survey on the use of receptor models (RMs) for particulate matter (PM) source apportionment in Europe between 2001 and 2010, including 79 studies and 243 reported records (Karagulian and Belis, 2012), found evidence of a dramatic increase in the number of scientific publications on this topic during the past decade and an increasing number of ready-to-use tools (Figure A.4). The highest rate of increase in the number of studies coincides with the entry into force of the limit value for PM₁₀ (1999/30/EC) and the target value for PM_{2.5}. About 60% of the studies were carried out in urban background sites, 16% in source-oriented sites (sites mainly affected by a single source), and 15% in rural sites.

Figure A.4. Time trend of RM studies in Europe between 2001 and 2010 (from Karagulian & Belis, 2012).



In contrast with the tendency observed between 1987 and 2005, the majority of the studies were performed using Positive Matrix Factorization and Chemical Mass Balance models in the period 2001-2010 (Figure A.4).

Most of the studies were conducted in Spain, Italy and the UK. Many recent studies completed or in progress were also carried out in France.

A detailed meta-analysis of data available from previous studies is presented in the most recent review of source identification studies, which covers the period until 2012 (Belis et al., 2013). In order to compare all the SA results and to attain useful conclusions, sources have been pooled into six major categories covering those most frequently observed in the individual studies: Sea/Road Salt, Crustal/Mineral Dust, Secondary Inorganic Aerosol (SIA), Traffic, Point Sources and Biomass Burning. In addition, residential heating by coal (or coal substitutes) combustion proved to be a major PM pollution source in many areas of the new EU Member States. Residential coal combustion in small stoves and boilers has also been found to be a main source of PM_{10} and benzo(a)pyrene in certain areas of Europe (Junninen et al., 2009).

The main results of the above-mentioned review show that the field of receptor models is developing swiftly, with Positive Matrix Factorization and Chemical Mass Balance (which are the most used models) evolving towards tools with refined uncertainty treatment.

The review demonstrates that, aside from mineral dust and sea/road salt, PM_{10} and $PM_{2.5}$ derive from the same sources. Secondary pollution deriving from gas-to-particle conversion is the main PM mass and particulate organic carbon source. Therefore, in order to reduce the concentration of these pollutants it is necessary to abate the sources of secondary inorganic aerosol deriving mainly from traffic emissions and agriculture. Primary emissions from traffic and biomass burning have also been identified as causes of exceedances, especially during the cold seasons.

The review stresses the need for long-term speciated PM datasets and the characterisation of source fingerprints to further improve source identification studies. In addition, harmonisation of the different approaches would facilitate the interpretation and comparison of the results and their application in the design of abatement measures.

When to use receptor models (RMs)?

The application of RMs requires quantitative data on air pollutant concentrations, good knowledge about atmospheric processes, good command of the chemical nature of the source emissions, and competence in the use of computational tools.

RMs have mainly been used to apportion airborne particulate matter sources. Therefore, the protocol presented in this report will mostly deal with this type of pollutant. However, it is

also possible to use this methodology on volatile organic compounds (VOCs), polyaromatic hydrocarbons (PAHs), inorganic gaseous pollutants and particle size distribution.

If very little information is available on the study area or if skilled staff are not available for running the standard applications, exploratory methods can be used to obtain a preliminary picture of the most relevant sources. Nevertheless, in order to achieve more accurate estimations of the source contributions and their uncertainties, a well-designed study is necessary, including field work, laboratory analyses for the chemical characterisation and data processing with standard tools.

Hybrid trajectory-based methods provide information about the geographical origin of pollutants. Advanced tools such as hybrid expanded models introduce *a priori* physical constraints in the model or combine different types of data (e.g. chemical and physical parameters, meteorology), making it possible to identify sources with small contributions and to better resolve similar or collinear ones.

Moreover, RMs can be used in combination with independent methodologies (e.g. emission inventories, chemical transport models (CTMs)) to achieve more robust estimations by mutual validation of the outputs.

Harmonisation of receptor models

Different methodologies for identifying sources are available. However, it is difficult to establish to what extent a methodology is appropriate for a specific purpose and to quantitatively express the reliability of the results. This is mainly because the actual source contributions at a specific point are unknown. In addition, the techniques used by experts with different backgrounds need to be harmonised so as to make the results of the different studies comparable. In order to address the challenges related to the use of modelling techniques in estimating pollution sources, the JRC launched an initiative in 2010 for the harmonisation of RMs used to identify pollution sources in Europe (Figure A.5, <http://source-apportionment.jrc.ec.europa.eu/>).

The initiative, which involved experts from many European countries, consisted of three main activities:

- reviewing RMs studies in Europe,
- organising European-wide intercomparison exercises for RMs and,
- developing a European harmonised technical protocol for RMs.

The initiative contributes to the activity of the Forum on Air Quality Modelling in Europe (FAIRMODE) Working Group on Source Apportionment.

The information about model performances collected in the above-mentioned intercomparison exercises provides the necessary complement to the procedures described in this document. These exercises have demonstrated that RMs provide quantitative estimations of the contributions of source categories with 50% or lower standard uncertainty (Karagulian et al., 2012).

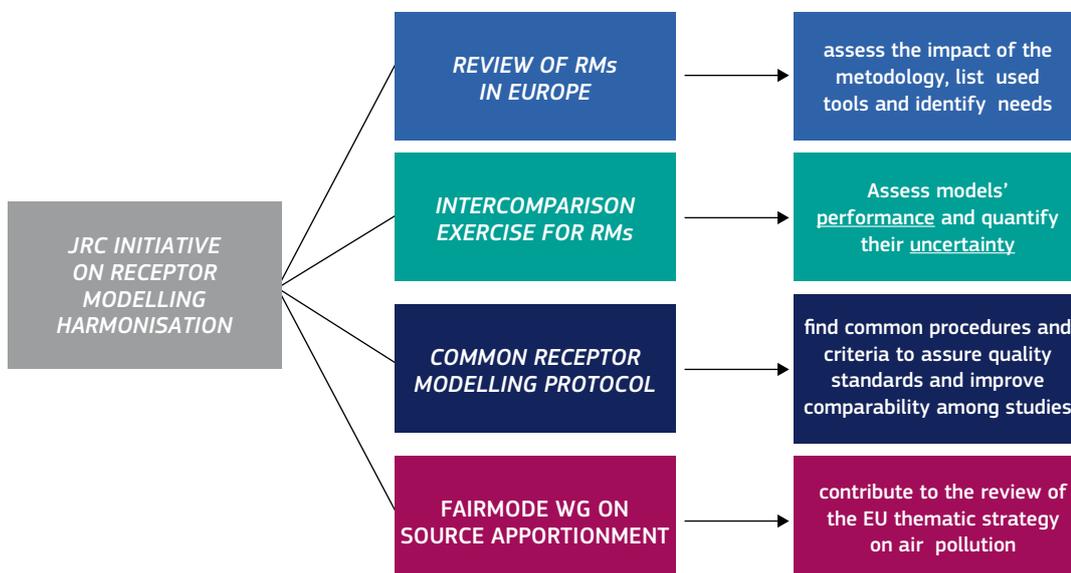


Figure A.5. JRC initiative for RM harmonisation.

References

- Belis C.A., Karagulian F., Larsen B.R., Hopke P. K., 2013. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmospheric Environment* 69, 94-108.
- Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.
- Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air.
- Fragkou, E., Douros I., Moussiopoulos N., Belis C. A., 2012. "Current Trends in the use of Models for Source Apportionment of Air Pollutants in Europe." *International Journal of Environment and Pollution* 50 (1-4), 363-375.
- Hopke, P. K., 2010. The application of receptor modeling to air quality data. *Pollution Atmospherique*, Special Issue 91-109.
- Junninen, H., Mønster, J., Rey, M., Cancelinha, J., Douglas, K., Duane, M., Forcina, V., Müller, A., Lagler, F., Marelli, L., Borowiak, A., Niedzialek, J., Paradiz, B., Mira-Salama, D., Jimenez, J., Hansen, U., Astorga, C., Stanczyk, K., Viana, M., Querol, X., Duvall, R.M., Norris, G.A., Tsakovski, S., Wählin, P., Horak, J., Larsen, B.R., 2009. Quantifying the impact of residential heating on the urban air quality in a typical European coal combustion region. *Environmental Science and Technology* 43, 7964-7970.
- Karagulian F. Belis C.A., 2012. Enhancing Source Apportionment with receptor models to Foster the Air Quality Directive Implementation. *International Journal of Environmental Pollution* 50, 190-199.
- Karagulian F., Belis C.A., Borowiak A., 2012. Results of the European Intercomparison exercise for receptor models 2011-2012. Part I. JRC Scientific and Policy Reports. Report EUR 25727 EN. ISBN 978-92-79-28130-3.
- Miller, M. S., Friedlander S. K., and Hidy G. M., 1972. A Chemical Element Balance for the Pasadena Aerosol. *Journal of Colloid and Interface Science* 39 (1), 165-176.
- US-EPA Support Center for Regulatory Atmospheric Modelling (<http://www.epa.gov/scram001/aqmindex.htm>).
- Viana M., Kuhlbusch T.A.J., Querol X., Alastuey A., Harrison R.M., Hopke P.K., Winiwarter W., Vallius M., Szidat S., Prévôt A.S.H., Hueglin C., Bloemen H., Wählin P., Vecchi R., Miranda A.I., Kasper-Giebl A., Maenhaut W., Hitzemberger R., 2008. Source apportionment of particulate matter in Europe: A review of methods and results. *Aerosol Science* 39, 827-849.
- Watson, J.G., 1979. Chemical element balance receptor model methodology for assessing the sources of fine and total suspended particulate matter in Portland, Oregon. Oregon Graduate Center. PhD Thesis.
- Watson, J.G., 1984. Overview of Receptor Model Principles. *Air Pollution Control Association Journal*, Vol. 34, No.6, June 1984
- Watson J. G., Chen L.-W. A., Chow J. C., Doraiswamy P., Lowenthal D. H., 2008. Source Apportionment: Findings from the U.S. Supersites Program. *Journal of the Air and Waste Management Association* 58, 265-288.
- Winchester, J. W., Nifong G. D., 1971. Water Pollution in Lake Michigan by Trace Elements from Pollution Aerosol Fallout. *Water, Air, and Soil Pollution* 1 (1), 50-64.

PART B: HARMONISED RECEPTOR MODEL PROTOCOL

B1. PRELIMINARY EVALUATION OF THE STUDY AREA

Collection of available data on atmospheric pollution

A sound source apportionment study requires careful preparation. The most important task in this step is the collection of all the relevant existing information about atmospheric pollution in the area under examination or in areas with similar characteristics. Bibliographic research should concentrate on both scientific publications and reports issued or sponsored by official bodies in charge of environmental monitoring that concern:

- emission inventories with a level of detail appropriate to the study (at least municipality or town level),
- local source profiles,
- time series at different time resolutions (daily, yearly averages) and daily profiles of pollutant levels and exceedances of legal thresholds,
- spatial distribution of pollutants, hotspots,
- meteorology at local and synoptic scale,
- previous source apportionment studies.

This step is essential to understand the nature and number of sources and the factors influencing pollutant dispersion (e.g. advection) and transformation (e.g. gas-to-particle processes).

The preliminary evaluation will be of great help in defining the objectives of the project and in planning the experimental work. To that end, it is also recommended that the local authorities be interviewed to understand the kind of information on pollution sources for air quality assessment and planning they are interested in, gather information on the measures that have been proposed or implemented, and understand the limitations they have encountered in their enforcement.

Description of the physical system

In addition to having a good conceptual understanding of the sources in the study area, it is important to understand the physical nature of the system. The topography, natural or artificial, has a significant influence on the local source-receptor relationships (e.g. Chow et al., 2007; Belis et al., 2008), and a lack of understanding of the physical system can lead to problems in interpreting and understanding the source apportionment results. A number of aspects of the physical system should be identified and incorporated into the planning and execution of a project as well as in the analysis of the subsequently generated data, in particular:

marine salt	industrial emissions
crustal material	secondary ammonium sulphate
road dust	secondary ammonium nitrate
gasoline vehicle exhaust	biomass burning / wood burning
diesel vehicle exhaust	maritime transport
power plants	secondary organic aerosol

Table B1. Common source categories of particulate matter (PM) in ambient air.

More details on the most common sources of PM in Europe can be found in Viana et al. (2008) and Belis et al. (2013)

- mountain / valley terrain,
- tall buildings,
- water bodies,
- local source complexes (grouped sources),
- isolated local sources,
- major transportation information,
- prevailing wind directions,
- distant sources.

Obstructions interfere with the direction of wind flows. People generally live in low-lying areas and thus often occupy valleys surrounded by obstructions (hills or mountains) that limit the wind directions to those found within the valley. Mountains can give rise to day-time upslope winds and night-time downslope winds. Tall buildings produce urban street canyons or block specific wind directions so that local meteorological measurements can be biased away from the actual wind directions. Water bodies also affect air flow locally (e.g. the influence of land-sea breezes). Thus, understanding the geography as well as the natural and anthropogenic topography will be important in understanding source/receptor relationships for a given site.

Sources can be contained in an industrial area that is well-delimited and/or isolated from other major sources. In the case of complex sources, the emitted pollutants come from roughly the same location and, if the temporal patterns of emissions are similar from multiple sources, methods that use the covariation of measured chemical species to identify specific source types will be confused by the simultaneous variation of the receptor-site impacts of emissions from disparate sources. Isolated sources can provide the opportunity to carry out some local sampling in areas known to be highly affected by that specific source and thereby get an indication of the nature of that source. For any source, it is important to understand the nature of the activities being conducted at the site and thus, what materials are likely to be released to the environment.

Transportation systems are sources of particles and other pollutants. Vehicles with combustion engines clearly produce significant tailpipe emissions along with emissions from tyre and brake wear, re-suspension of road dust, and other related materials. Electrified systems such as trams, trains and electric buses also produce particulate emissions from the ablation

of the runners that pick up the electricity from the wires and transfer it to the moving vehicle. There may also be ablation from the steel wheels rolling and stopping on the steel rails. The location of highways and other transportation systems, the nature of the vehicles operating in the vicinity of the sampler, their operating pattern (highway speed, stop-and-go, etc.) and the prevailing wind directions may all influence the measurements at the receptor site.

Prevailing wind directions determine the probability of emitted materials being transported to the measurement site. Sources with low probability wind directions are unlikely to make a large impact on a site (on the long-term average) even if they are significant emitters of the measured pollutant(s).

Although primary emissions are diluted over time and distance, secondary pollutants, e.g. produced by gas-to-particle conversion processes, can increase the concentrations over relatively long distances, particularly for species such as secondary sulphate and secondary organic aerosols that take time to form in the atmosphere.

References

Belis C.A., Karagulian F., Larsen B.R., Hopke P.K., 2013. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmospheric Environment* 69, 94-108.

Belis C. A., Gianelle V., De Stefani G., Colombi C. and Magnani T., 2008. The influence of orography and meteorology on the spatial distribution of pollutants in the alpine city of Sondrio (Northern Italy). *Chemical Engineering Transactions* 16, 153-160. <http://www.aidic.it/cet/08/16/021.pdf>

Chow, J.C., Watson, J.G., Chen, L.-A. Magliano, K., 2007. Particulate matter measurement in central California, *Air and Waste Management Association - Symposium on Air Quality Measurement Methods and Technology 2007*, pp. 179-183.

Viana M., Kuhlbusch T.A.J., Querol X., Alastuey A., Harrison R.M., Hopke P.K., Winiwarter W., Vallius M., Szidat S., Prévôt A.S.H., Hueglin C., Bloemen H., Wählin P., Vecchi R., Miranda A.I., Kasper-Giebl A., Maenhaut W., Hitzenberger R., 2008. Source apportionment of particulate matter in Europe: A review of methods and results. *Aerosol Science* 39, 827-849.

B2. DEFINING A METHODOLOGICAL FRAMEWORK

Source apportionment studies should be planned in advance according to:

- the preliminary evaluation (section B1),
- the objectives of the study,
- the available resources (funds, staff skills, time),
- the model and software to be used,
- the input data source (already available or data collection is needed),
- the required qualification of the operator and training needs.

Appropriate study planning prevents or reduces the risk of collecting useless information, missing relevant information for model execution or data interpretation, using resources inefficiently and/or building up a delay with respect to the scheduled deadlines. Useful advice for the definition of the methodological framework is available from Kim Oanh et al. (2009), Johnson et al. (2011), Watson et al. (2002) and Watson et al. (2008). The adoption of a quality management system (QMS) for the project could be useful to identify and document procedures, deliverables, responsibilities and deadlines (e.g. ISO 9001:2008).

The preliminary evaluation achieved in the previous step provides the basis for defining the objectives of the study.

At this point, the expert shall define the main questions he/she intends to answer. Subsequently, the main questions are translated into operational hypotheses and how the experimental work will contribute to test those hypotheses is clearly explained.

The objectives of the study must be in line with the available resources in terms of equipment, staff, and software. It is important to evaluate whether the required technical skills are present in the team, to make sure there is access

to the technical and methodological information and, if possible, to collaborate with experts in institutions with demonstrated expertise in the field of source apportionment that can provide professional advice.

Selecting the type of model early in the planning process is also important as the kind of information to be collected depends on the model input variables:

- a chemical mass balance (CMB) model requires local source profiles as input;
- principal components analysis (PCA) and factor analysis do not require source profiles as input, but do require a very good knowledge of the study area in order to be able to interpret the output factors in terms of source categories;
- positive matrix factorization (PMF) and CMB models need an uncertainty estimation for each data entry;
- advanced models also process other types of data: e.g. meteorological variables, trajectories, day of the week, size distribution.

The choice of the model should take into account the fact that running more than one model on the same dataset can mutually validate their outputs and lead to more robust results. This may require additional time and skills.

If the input data for the selected model is not available (as is almost always the case) it is necessary to plan field activity in order to collect information on the ambient concentration of the pollutants of interest and the chemical profiles of local sources (see section B3). You may also need to collect meteorological data if this kind of information is not available close to your study site.

References

- ISO 9001:2008. Quality management systems – Requirements. http://www.iso.org/iso/home/standards/management-standards/iso_9000.htm
- Johnson T. M, Guttikunda S., Wells G. J., Artaxo P., Bond T. C., Russell A. G., Watson J. G., West J., 2011. Tools for Improving Air Quality Management. Formal Report 339/11. Energy Sector Management Assistance Program. Washington DC, USA.
- Kim Oanh N.T., Pongkiatkul P., Upadhyay N., Hopke P.K., 2009. Designing ambient particulate matter monitoring program for source apportionment study by receptor modeling. *Atmospheric Environment* 43(21), 3334-44.
- Watson J. G. , Zhu T., Chow J. C., Engelbrecht J., Fujita E. M., Wilson W. E., 2002. Receptor modeling application framework for particle source apportionment. *Chemosphere* 49 1093–1136.
- Watson J. G., Chen L.-W. A., Chow J. C., Doraiswamy P., Lowenthal D. H., 2008. Source Apportionment: Findings from the U.S. Supersites Program. *Journal of the Air and Waste Management Association* 58, 265–288

B3. EXPERIMENT DESIGN - CRITERIA FOR SITE AND SPECIES SELECTION AND ESTIMATION OF MINIMUM NUMBER OF SAMPLES

Site Selection

For source apportionment, sites representative of the mixture of sources in a given area are preferable to sites influenced by specific sources. To establish the number and location of sources, it is necessary to study emission source distribution, wind roses and typical dispersion patterns (upwind, downwind of major sources). According to Kim Oanh et al. (2009), several sites are required to represent the different sub airsheds in a city.

Stack height, temperature, mechanical buoyancy, and temporal variation of emissions are important pieces of information for point sources. Additional information to evaluate the distribution of pollutants can be obtained from basic meteorological parameters and the levels of primary gaseous pollutants (Kim Oanh et al., 2009).

Representativeness of monitoring sites and heterogeneity of the study areas can be tested using geostatistical methods. These techniques assess the relationship between the difference of concentrations in and distance between different sites by fitting functions known as “variograms” (Clark & Harper, 2002; Kim et al., 2005, Hwang et al., 2008, Lagudu et al.; 2011, Kumar et al., 2012).

In order to obtain estimations of source contributions in an area, a combination of multiple sites with the same or different characteristics is commonly used. In the incremental or ‘Lenschow’ approach, the differences in contributions from traffic, urban background and rural or regional backgrounds is used to estimate sources. A more complex option is the combination of independent source contribution estimations for different sites (e.g. Larsen et al., 2012). The orientation of sites according to the main wind directions makes it possible to assess the contributions from medium- to long-range transport (e.g. AIRPARIF and LSCE, 2012).

Species selection

The chemical species to include in the analysis should be selected according to the study objectives, the site characteristics and expected sources, taking into account the available human, technical, and financial resources.

Since RMs have mainly been used to apportion sources of airborne particulate matter, this document focuses on this type of pollutant (Table B3.1). Nevertheless, this methodology has also been used on datasets containing volatile organic compounds (VOCs; e.g. Elbir et al., 2007; Lanz et al., 2009; Niedojadlo et al., 2007), polycyclic aromatic hydrocarbons (PAHs; e.g. Belis et al., 2012; Hanedar et al., 2011; Mari et al., 2010; Okuda et al., 2010) and inorganic gaseous pollutants (e.g. Ogulei et al., 2006).

Table B3.1 Examples of input data for source apportionment with RMs

Ions	sulphate, nitrate, ammonium, chloride, Na ⁺ , Mg ⁺⁺ , K ⁺ , Ca ⁺⁺
Carbonaceous fractions	Total carbon (TC), elemental carbon (EC)/organic carbon (OC) total or fractions obtained in every analytical step
Elements	Na, Mg, Al, Si, P, S, Cl, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Zr, Mo, Rh, Pd, Ag, Cd, Sn, Sb, Te, I, Cs, Ba, La, W, Au, Hg, Pb
Organic markers	n-alkanes, alkanolic (carboxylic) acids (especially fatty acids), aromatic carboxylic acids, levoglucosan/mannosan, PAHs, hopanes, resin acids, syringols, cholesterol
Aerosol size distribution	scanning mobility particle sizer (SMPS), optical particle counter (OPC), aerodynamic particle sizer (APS), cascade impactors, streakers, Davis rotating-drum Universal-size-cut Monitoring impactor (DRUM/RDI)
Mass fragments (m/z) concentrations	obtained with Aerosol Mass Spectrometer (AMS) or Aerosol Chemical Speciation Monitor (ACSM) techniques and used to apportion the organic fraction (see section C2).
Optical properties	absorption coefficients to apportion C _{ff} * and C _{wb} *, light scattering at multiple wavelengths (see section C3).
Isotopic ratios	¹⁴ C/ ¹² C ratios to apportion fossil C and recent C (see section C5)
Radon	indicator of planetary boundary layer (PBL) mixing and long-range pollution transport

* C_{ff}: carbonaceous fraction deriving from fossil fuel and C_{wb}: carbonaceous fraction deriving from wood burning.

Chemical species that are difficult to analyse or that yield anomalous values (commonly referred to as “weak elements” in PMF) tend to result in physically meaningless factors (Huang et al., 1999). For that reason, certain authors recommend that species considered unsuitable as source tracers be excluded. According to Ito et al. (2004), species that are not indicative of any source, or that are indicative of sources which are not relevant to the objectives of the study, can be discarded. However, the exclusion of species may lead to a loss of relevant information if we consider that the concomitant variation of a set of species could be indicative of a source even though none of them is exclusively emitted by that source. Quite often, analytical protocols such as those of X-ray fluorescence (XRF) or gas chromatography coupled with mass spectrometry (GC-MS) are able to provide multiple species output at little or no additional cost. The opportunity to take advantage of these “additional” species should not be ignored.

Using a reduced number of species could limit the number of sources that can be identified. Many multivariate methods like PMF and CMB are sensitive to collinearity. Increasing the number of species may help to reduce the collinearity between different source or factor profiles, thereby increasing the number of sources that can potentially be resolved.

In order to prevent double mass counting, redundant species should be avoided. This could be the case with sulphur (S) and sulphate, between elements and their corresponding cations or between organic carbon / elemental carbon (OC/EC) and total carbon (TC). However, soluble potassium (K) can sometimes be a

useful indicator of biomass burning and thus, soluble and insoluble K can both be included in the model where insoluble K = total K – soluble K. More generally, if the two species are proportional to each other throughout the dataset, then it does not matter which one is used. However, if they do not track each other, a better separation of sources could be achieved by keeping both species in the dataset during the analysis. Double mass counting should be corrected at a later stage by retaining only one of the species in the computed factor profiles.

The traditional approach in receptor models relies on a basic set of chemical species that represents most of the particulate mass such as major ions (sulphate, nitrate, and ammonium) and the carbonaceous fraction (total organic carbon (TOC), OC/EC) plus a number of elements whose absolute and relative concentrations or specific ratios are used to identify sources (Miller et al., 1972). Although organic matter constitutes a considerable share of PM and has relevant influence on the physical and chemical properties and effects of the aerosol on health, the analytical techniques used in the past were not suitable for describing this fraction in full. The development of mass spectrometry made it possible to determine and identify organic compounds that are characteristic of certain sources called **molecular markers**. For example:

- levoglucosan, methoxyphenols (Simoneit, 2002) and syringol are markers for **biomass burning**,

- hopanes and steranes for **vehicle emissions** (Cass, 1998; Schauer et al., 2002),
- cholesterol and fatty acids for **cooking emissions** (Chow et al., 2007; Zhao et al., 2007, Schauer et al., 1999),
- benzene, di-, tri and tetra carboxylic acids, phthalates, branched ketones for **secondary organic aerosols** (Jaeckels et al., 2007; Subramanian et al., 2007).

The inclusion of molecular markers in the set of species is often desirable but requires specific sampling and analytical techniques (Wang et al., 2012). Moreover, the development and availability of instruments to measure the optical properties of the aerosol (light scattering, light absorption) and its size distribution has led to studies in which this information is combined with the chemical composition in order to better constrain the sources on the basis of their properties and the processes that pollutants undergo in the atmosphere.

Mass concentration or number concentrations in particle size bins can be used as species together with chemical species (e.g. Gu et al. 2011, Pere-Trepat et al., 2007; Pey et al., 2009; Zhou et al., 2005). For example, Ogulei et al. (2006) reported NO_3^- associated with particles larger than those associated with SO_4^{2-} , and ultra-fine particles (UFP) associated with gasoline and diesel exhausts but not with the burning of vegetation.

Number and frequency of samples

From the mathematical point of view, CMB can be executed with just one sample. In practice, many samples are required to obtain results that are representative of the variety of conditions in the study area, including the variability of sources over time. By contrast, multivariate techniques only work properly with large numbers of samples as input. According to the EPA PMF v3 User Guide (Norris et al., 2008), this method is often used on speciated $\text{PM}_{2.5}$ datasets with over 100 samples. At least 100 samples of 24-hour data of at least 20 species are recommended by Brown & Hafner (2005). Johnson et al. (2001) claim that at least 50 chemically characterised ambient samples are required for running multivariate models. According to Henry et al. (1984), the minimum number of samples (N) is the one that yields a ratio between degrees of freedom (D) and number of variables (V) that is higher than 60, while the optimal is one that leads to values above 100, according to the following equation:

$$D/V = N - (V/2 - 1.5)$$

On the other hand, Thurston and Spengler (1985) propose that the number of samples should exceed the number of variables by at least a factor of three.

In practice, the minimum number of samples required to detect the latent variables cannot be established a priori as it depends on the amount of information contained in the dataset. If the relative contribution of sources were the same in all samples, analysing new samples would not add any new information to the model. Therefore, there should be enough samples to catch the variability of the sources, including samples where some sources are absent or negligible.

Small datasets simultaneously obtained for several sites may be made usable for PMF analysis by pooling them into one matrix. This can be done, for example, for several urban background sites within a city (Xie et al., 2012), a larger region (Belis et al., 2011; Larsen et al., 2012; Mooibroek et al., 2009) or even for different site types to assess contributions of common sources (natural sources, Beuck et al., 2011).

In filter-based systems, the most common configuration is the collection of 24-hour samples. This is in part due to the requirements of reference gravimetric methods for the determination of the PM mass. In addition, a 24-hour period is considered to be representative of all the sources occurring in one day-to-night cycle and hence an appropriate unit for data elaboration. A practical reason for selecting 24-hour sampling also derives from the need to collect enough PM for chemical analysis. This limitation is especially true for low-volume samplers when PM levels are low such as in areas that are located far from the sources or in seasons during which the main sources are not active. In urban areas, four- to six-hour sampling times usually allow for the collection of enough material for major component analyses (e.g. Vecchi et al., 2009; Bernardoni et al., 2011). This configuration provides the opportunity to detect the daily trend of most sources, making their identification with receptor modelling more feasible. With high-volume samplers, two- to four-hour samples can be sufficient.

Higher time resolutions can be achieved using semi-continuous systems for chemical analysis: particle-into-liquid samplers (PILS), semi-continuous elements in aerosol systems (SEAS), monitoring instrument for aerosols and gasses (MARGA), semicontinuous EC/OC, with resolutions ranging from a few minutes to one hour (see section B4). Streakers or DRUM/RDI samplers also provide the opportunity to select the time resolution of the analysis on size-resolved

samples. Physical parameters associated with particle size or optical properties (scattering, absorption) can be obtained with time resolutions close to a minute or less.

Time resolutions in the order of seconds and minutes can be obtained with online aerosol mass spectrometers (Pratt and Prather, 2011; Drewnick, 2012).

References

AIRPARIF and LSCE, 2012. Etude de la contribution des sources des particules en Ile-de-France- Rapport Final, Edited, 171 pp. AIRPARIF, Paris. English version: http://www.airparif.asso.fr/_pdf/publications/rapport-particules-anglais-120829.pdf

Belis, C. A., Cancelinha, J., Duane, M., Forcina, V., Pedroni, V., Passarella, R., Tanet, G., Douglas, K., Piazzalunga, A., Bolzacchini, E., Sangiorgi, G., Perrone, M. G., Ferrero, L., Fermo, P., Larsen, B. R., 2011. Sources for PM air pollution in the Po Plain, Italy: I. Critical comparison of methods for estimating biomass burning contributions to benzo(a)pyrene. *Atmospheric Environment* 45, 7266-7275

Bernardoni, V., Vecchi, R., Valli, G., Piazzalunga, A., Fermo, P., 2011. PM₁₀ source apportionment in Milan (Italy) using time-resolved data. *The Science of the Total Environment* 409, 4788-4795

Beuck H., Quass U., Klemm O., Kuhlbusch T.A.J., 2011. Assessment of sea salt and mineral dust contributions to PM₁₀ in NW Germany using tracer models and positive matrix factorization, *Atmospheric Environment* 45, 5813-5821

Brown S.G., Hafner H., 2005. Multivariate Receptor Modeling Workbook. Sonoma Technology Inc. Petaluma, CA. U.S.A. 236 pp. http://www.sonomatechdata.com/sti_workbooks/

Cass, G. R. 1998. Organic Molecular Tracers for Particulate Air Pollution Sources. TrAC - *Trends in Analytical Chemistry* 17 (6): 356-366.

Chow, J.C., Watson, J.G., Lowenthal, D.H., Chen, L.-W.A., Zielinska, B., Mazzoleni, L.R., Magliano, K.L., 2007. Evaluation of organic markers for chemical mass balance source apportionment at the Fresno supersite. *Atmos. Chem. Phys.* 7(7), 1741-2754.

Chow, J.C., 1995. Critical review: Measurement methods to determine compliance with ambient air quality standards for suspended particles. *Journal of the Air and Waste Management Association.* 45(5):320-382.

Clark I., Harper W., 2002. Practical Geostatistics, 2000. Ecosse North America Llc. Columbus. Ohio, USA. 342pp ISBN 0-9703317-0-3

Drewnick, F. Speciation Analysis in On-line Aerosol Mass Spectrometry. 2012. *Analytical and Bioanalytical Chemistry* 404:2127-2131.

Elbir, T., Cetin, B., Cetin, E., Bayram, A., Odabasi, M., 2007. Characterization of volatile organic compounds (VOCs) and their sources in the air of Izmir, Turkey. *Environmental Monitoring and Assessment* 133, 149-160

Gu, J.W., Pitz, M., Schnelle-Kreis, J., Diemer, J., Reller, A., Zimmermann, R., Soentgen, J., Stoelzel, M., Wichmann, H. E., Peters, A., Cyrys, J., 2011. Source apportionment of ambient particles: Comparison of positive matrix factorization analysis applied to particle size distribution and chemical composition data. *Atmospheric Environment* 45, 1849-1857.

Hanedar, A., Alp, K., Kaynak, B., Baek, J., Avsar, E., Odman, M. T., 2011. Concentrations and sources of PAHs at three stations in Istanbul, Turkey. *Atmospheric Research* 99, 391-399

Henry R.C., Lewis C.W., Hopke P.K., Williamson H.J., 1984. Review of receptor model fundamentals. *Atmospheric Environment* 18(8), 1507-1515.

Huang S., Rahn K.A., Arimoto R., 1999. Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode island. *Atmospheric Environment* 33(14), 2169-85.

Hwang I.J., Hopke P.K., Pinto J.P., 2008. Source Apportionment and Spatial Distributions of Coarse Particles During the Regional Air Pollution Study (RAPS), *Environ. Sci. Technol.* 42, 3524-3530.

Ito K, Xue N, Thurston G., 2004. Spatial variation of PM_{2.5} chemical species and source-apportioned mass concentrations in New York city. *Atmospheric Environment* 38(31), 5269-82.

Jaekels, J. M., Bae M. -S, Schauer J. J., 2007. Positive Matrix Factorization (PMF) Analysis of Molecular Marker Measurements to Quantify the Sources of Organic Aerosols. *Environmental Science and Technology* 41 (16), 5763-5769.

Johnson T.M., Suttikunda S., Wells G.J., Artaxo, P. Bond T. C., Russell A.G., Watson J.G., West J., 2011. Tools for improving Air Quality Management REPORT339/11 ESMAP.

- Kim E., Hopke P.K., Pinto J.P., Wilson W.E., 2005. Spatial Variability of Fine Particle Mass, Components, and Source Contributions During the Regional Air Pollution Study in St. Louis, *Environmental Science and Technology* 39, 4172-4179.
- Kim Oanh N.T., Pongkiatkul P., Upadhyay N., Hopke P.K., 2009. Designing ambient particulate matter monitoring program for source apportionment study by receptor modeling. *Atmospheric Environment* 43(21), 3334-44.
- Kumar P., Hopke P.K., Raja S., Casuccio G., Lersch T.L., West R.R., 2012. Characterization and Heterogeneity of Coarse Particles Across An Urban Area, *Atmospheric Environment* 46, 339-459
- Lagudu U.R.K., Raja S., Hopke P.K., Chalupa D.C., Utell M.J., Casuccio G., Lersch T.L., West R.R., 2011. Heterogeneity of Coarse Particles in an Urban Area, *Environmental Science and Technology* 45, 3288-3296.
- Lanz, V. A., Henne, S., Staehelin, J., Hueglin, C., Vollmer, M. K., Steinbacher, M., Buchmann, B., and Reimann, S., 2009. Statistical analysis of anthropogenic non-methane VOC variability at a European background location (Jungfrauoch, Switzerland). *Atmospheric Chemistry and Physics* 9, 3445-3459.
- Larsen, B.R., Gilardoni S., Stenström K., Niedzialek J., Jimenez J., Belis C. A., 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. *Atmospheric Environment* 50, 203-213.
- Mari, M., Harrison R. M., Schuhmacher M., Domingo J. L., Pongpiachan S., 2010. Inferences Over the Sources and Processes Affecting Polycyclic Aromatic Hydrocarbons in the Atmosphere Derived from Measured Data. *Science of the Total Environment* 408 (11), 2387-2393.
- Miller, M.S., S.K. Friedlander and G.M. Hidy, 1972. A chemical element balance for the Pasadena aerosol, *Journal of Colloid Interface Science* 39, 65-176.
- Mooibroek D., Schaap M., Weijers E.P., Hoogerbrugge R., 2011. Source apportionment and spatial variability of PM_{2.5} using measurements at five sites in the Netherlands. *Atmospheric Environment* 45: 4180-4191.
- Niedojadlo, A., Becker, K. H., Kurtenbach, R., Wiesen, P., 2007. The contribution of traffic and solvent use to the total NMVOC emission in a German city derived from measurements and CMB modelling. *Atmospheric Environment* 41, 7108-7126
- Norris G., Vedantham R., Wade K., Brown S., Prouty J., Foley C., 2008. EPA Positive Matrix Factorization (PMF 3.0) Fundamentals & User Guide. EPA 600/R-08/108.
- Ogulei D., Hopke P.K., Zhou L., Patrick Pancras J., Nair N., Ondov J.M., 2006. Source apportionment of Baltimore aerosol from combined size distribution and chemical composition data. *Atmospheric Environment* 40(SUPPL. 2), 396-410.
- Okuda, T., Okamoto, K., Tanaka, S., Shen, Z., Han, Y., and Huo, Z., 2010. Measurement and source identification of polycyclic aromatic hydrocarbons (PAHs) in the aerosol in Xi'an, China, by using automated column chromatography and applying positive matrix factorization (PMF). *Science of the Total Environment* 408, 1909-1914
- Pey, J., Querol, X., Alastuey, A., Rodriguez, S., Putaud, J.P., Van Dingenen, R., 2009. Source apportionment of urban fine and ultra-fine particle number concentration in a Western Mediterranean city. *Atmospheric Environment* 43 (29), 4407-4415.
- Pratt, K. A., Prather, K. A., 2012. Mass spectrometry of atmospheric aerosols—Recent developments and applications. Part II: On-line mass spectrometry techniques. *Mass Spectrometry Review* 31: 17-48.
- Pere-Trepat E., Kim E., Paatero P., Hopke P.K., 2007. Source apportionment of time and size resolved ambient particulate matter measured with a rotating DRUM impactor. *Atmospheric Environment* 41 (28), 5921-5933.
- Schauer, J. J., Kleeman M. J., Cass G. R., Simoneit B. R. T., 1999. Measurement of Emissions from Air Pollution Sources. 1. C1 through C29 Organic Compounds from Meat Charbroiling. *Environmental Science and Technology* 33 (10), 1566-1577.
- Schauer, J. J., Kleeman M. J., Cass G. R., Simoneit B. R. T., 2002. Measurement of Emissions from Air Pollution Sources. 5. C 1 - C 32 Organic Compounds from Gasoline-Powered Motor Vehicles. *Environmental Science and Technology* 36 (6), 1169-1180.

- Simoneit B.R.T., 2002. Biomass burning - A review of organic tracers for smoke from incomplete combustion. *Applied Geochemistry* 17(3), 129-62.
- Subramanian, R.; Donahue, N. M.; Bernardo-Bricker, A.; Rogge, W. F.; Robinson, A. L., 2007. Insights into the primary-secondary and regional-local contributions to organic aerosol and PM_{2.5} mass in Pittsburgh, Pennsylvania. *Atmospheric Environment* 41, 7414-7433.
- Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment* 19, 9-25.
- Vecchi R, Bernardoni V, Fermo P, Lucarelli F, Mazzei F, Nava S, Prati P, Piazzalunga A, Valli G., 2009. 4-hours resolution data to study PM₁₀ in a "hot spot" area in Europe. *Environmental Monitoring and Assessment* 154(1-4), 283-300.
- Wang, Y. Hopke, P.K., Xia, X., Rattigan, O., Chalupa, D.C., Utell, M.J., 2012. Source apportionment of airborne particulate matter using inorganic and organic species as tracers. *Atmospheric Environment* 55, 525-532.
- Xie, M., Coons, T.L. Hemann, J-G., Dutton S.J., Milford, J.B. Peel. J.L. Miller S.L Kim, S-Y. Vedal, S., Sheppard, L., Hannigan, M.P., 2012. Intra-urban spatial variability and uncertainty assessment of PM_{2.5} sources based on carbonaceous species, *Atmospheric Environment* 60, 305-315.
- Zhao Y, Hu M, Slanina S, Zhang Y., 2007. Chemical compositions of fine particulate organic matter emitted from chinese cooking. *Environmental Science and Technology* 41(1), 99-105.
- Zhou, L., Hopke P. K., Stanier C. O., Pandis S. N., Ondov J. M., Pancras J. P., 2005. Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares and positive matrix factorization, *Journal of Geophysical. Research* 110, D07S18 1-15.

B4. DATA COLLECTION / FIELD WORK / CHEMICAL ANALYSIS

The analytical techniques are selected on the basis of: particulate matter (PM) concentrations, required detection limits (DLs) and uncertainties, costs, access to laboratory facilities, and time resolution of the processes under study.

Sampling systems

Offline chemical analysis of PM is commonly performed using filter-based methods.

Different size fractions and sampling flow rates are available. Since PM_{10} and $PM_{2.5}$ are regulated under Directive 2008/50/EC, reference methods exist (EN 12341 and EN 14907, currently under revision) and most experience and types of instruments are available in Europe (Lagler et al., 2011). The advantage of using these methods is that data can be compared with those in a wide number of sites. On the other hand, the current European legislation focuses more on total PM mass concentration than on the analysis of its chemical composition (analysis of major carbon fractions and ions are requested only for few rural/remote sites). Therefore, reference methods are not always the most appropriate for source apportionment. In the US there are samplers that are specially designed for PM speciation: “RAAS” (Andersen), “MASS” (URG), “SASS” (Met One), “Partisol 2300” (Thermo), among others (Solomon et al., 2000).

The high-volume polyurethane foam (PUF) sampler, which has a large volumetric flow (hundreds $L\ min^{-1}$), may be used in parallel with low-volume PM samplers to collect samples of semi-volatile organic compounds (SVOCs) in both PM and gaseous phases (Kim Oanh et al., 2009).

Filter choice

The selection of filters is guided by the following criteria: limited artefacts, compatibility with the analytical techniques, no interactions with the sample, low level of impurities, and high efficiency.

Commonly used filter matrices are pure quartz, coated quartz and Teflon, nylon, polycarbonate, glass fibre and cellulose esters. For a detailed discussion, see Chow (1995).

Significant differences are possible between sampling systems for organic carbon and nitrates due to loss of nitrate or either deposition or loss of organic carbon. In order to test the influence of deposition and loss of semi-volatile compounds in filter-based methods, relatively complex sampling systems equipped with denuders and double filters (front filter and backup filter) are required (e.g. Subramanian et al., 2004). Unlike CEN standards, the EMEP protocols recommend the application of these methodologies for the limitation and/or the estimation of positive and negative sampling artefacts.

It is worth mentioning that only quartz fibre filters are suitable for the determination of ions, elements, and carbonaceous fractions (organic, elemental) on the same sampling support, as carbonaceous aerosols have to be analysed at elevated temperatures.

Most common analytical techniques

Organic carbon and elemental carbon (OC, EC), either total or by single temperature steps, are commonly measured using thermal-optical methods. These methods take advantage of the different behaviour of the various carbonaceous fractions (i.e. OC, EC, and carbonate carbon) when exposed to elevated temperatures and to light. OC evolves at lower temperatures than EC while the latter absorbs more light than the former. The main differences between the existing thermal-optical methods (e.g. “NIOSH”, “IMPROVE” and “EUSAAR”) rely mainly on the temperature programs and on the devices used for optical measurements: Thermal Optical Transmission (TOT) or Thermo Optical Reflectance (TOR). More information can be found in Chow et al. (2004) and Cavalli et al. (2010). Since OC/EC analyses are required by the European Air Quality Directive 2008/50/EC, a standardised procedure is currently under preparation by the working group 35 of the CEN technical committee 264.

The methods most used for anions and cations are ion chromatography (IC) or automated colorimetric analysis. Also for these compounds, a standardised procedure is currently under preparation by the working group 34 of the CEN technical committee 264.

For inorganic elements, inductively coupled plasma - mass spectrometry (ICP-MS) and graphite furnace - atomic absorption spectrometry (GF-AAS) which are the reference methods for the determination of metals (As, Cd, Ni and Pb) in PM₁₀ (Standard EN 14902). Although some alteration cannot be excluded due to vacuum and slight heating (Yatkin and Gerboles, 2013), energy dispersive X-ray fluorescence (XRF) is commonly used in source apportionment because it covers many elements (from Na to U), does not require sample pre-treatment and does not destroy the samples. It also has good accuracy and repeatability, and automation of the analysis makes it possible to treat high numbers of samples with reduced costs. A similar technique, particle-induced X-ray emission (PIXE) is also suitable. Differences in detection limits (DL), when compared to XRF, are due to intrinsic features of the two techniques, such

as different ionisation cross-sections for photons or protons and differences in the intensities of the continuous background (Calzolari et al., 2008). PIXE is more powerful than XRF in analysing very small samples (i.e. size-segregated samples, high time-resolution samples or those collected in remote areas). The main limitation is due to the availability of beam time at the accelerator facility where PIXE analysis can be carried out.

Organic compounds

PAH levels in PM₁₀ are regulated under Directive 2004/107/EC. The application of ISO standard 12884 is recommended but there is no reference method in this case. Either GC-MS or high-performance liquid chromatography (HPLC) methods are used for these compounds. Offline GC-MS is used to characterise a wide range of organic compounds (see table B3.1). More recently, the sensitivity of thermal desorption GC-MS methods has improved and, when combined with in situ derivation, enables the identification of polar and non-polar components (Laskin et al., 2012).

Advanced User Box

Online monitoring techniques for PM composition

In online systems, sampling and analysis are integrated into a single instrument.

The determination of PM mass using online systems relies mostly on two operating principles: Tapered element oscillating microbalance (TEOM) and beta gauges.

Examples of online analytical instrumentation for ions are:

- a) Particle-Into-Liquid Sampler (PILS) developed by the Georgia Institute of Technology (Weber et al., 2001).
- b) Monitoring instrument for aerosols and gasses (MARGA; Khylstov et al., 1995).

For the analysis of elements the following methods are available:

- a) Semi-continuous elements in aerosol system (SEAS), developed by University of Maryland (Kidwell and Ondov, 2001), and b) a modified version "SEAS II" and "KSEAS" (Lee et al., 2011).

A field-deployable system also exists to collect airborne particles and apply semi-continuous XRF analysis to the samples (Xact 620, Cooper Environmental Services).

For carbonaceous fractions, a semi-continuous OC/EC analyser is available from Sunset (Bauer et al., 2009), although the measurements are not fully comparable with those obtained with the offline method (Belz et al., 2012).

Optical techniques for monitoring the aerosol light absorbance, such as the Multi-Angle Absorption Photometer (Petzold and Schönlinner, 2004) or the Aethalometers (Hansen et al. 1984), are commonly used to estimate the carbon absorbing part of the aerosol, commonly known as Black Carbon.

Finally, as described in section C.2, organic mass spectra can be obtained routinely using Aerosol Mass Spectrometers (AMS) or Aerosol Chemical Speciation Monitors (ACSM).

There is a variety of techniques for the determination of anhydrosugars (e.g. levoglucosan). GC-MS-based methods have been extensively used, but different types of HPLC and IC techniques have also been proposed (Schkolnik and Rudich, 2006).

Local source profiles (fingerprints)

Chemical fingerprints of local sources are of utmost importance in SA studies. The characterisation of the most relevant sources in the study area should be included in the work programme. Considering that local source characterisation is resource consuming, it is possible to adopt fingerprints available from previous work in similar areas or obtained from source profile repositories (e.g. SPECIATE, US-EPA; <http://www.epa.gov/ttnchie1/software/speciate/>). The sample collection varies from source to source. For pollutants deriving from combustion processes, samples collected directly from the stack or exhaust at temperatures much higher than that of ambient air may lead to biases due to the absence of the condensed fraction in the particulate phase. To overcome this pitfall, it is possible to dilute the emissions with a known volume of clean air. An alternative is to sample the plume at a distance that allows the effluent to dilute and cool down to near ambient temperatures.

Source-oriented monitoring stations can be used to characterise the source emissions if periods in which other sources influence the sample are excluded from the analysis. Characterisation of mobile sources can be obtained with samples collected in the lab (e.g. Montero et al., 2010; Adam et al., 2011), on the road (e.g. Georgios et al., 2012) or in tunnel experiments (e.g. El Haddad et al., 2009).

Re-suspension of road dust and contributions deriving from industrial dust can be estimated by sampling deposits directly from the ground (Amato et al., 2009, Ashbaugh et al., 2003; Colombi et al., 2010). Samples of vehicle parts' wear (tyres, brakes, clutch) can also be obtained directly by abrasion in the laboratory (e.g. Sjödim et al., 2010).

References

Adam T., A. Farfaletti, L. Montero, G. Martini, U. Manfredi, B. R. Larsen, G. De Santi, A. Krasenbrink, C. Astorga., 2010. Chemical characterization of emissions from modern two-stroke mopeds complying with legislative regulation in Europe (EURO-2). *Environmental Science and Technology* 44, 505-512.

Amato F., Pandolfi M., Viana M., Querol X., Alastuey A., Moreno T., 2009. Spatial and chemical patterns of PM10 in road dust deposited in urban environment. *Atmospheric Environment*, 43 (9) 1650-1659.

Ashbaugh, L.L., Carvacho, O.F., Brown, M.S., Chow, J.C., Watson, J.G., Magliano, K.L., 2003. Soil sample collection and analysis for the Fugitive Dust Characterization Study. *Atmospheric Environment*, 37(9-10),1163- 1173.

Bauer JJ, Yu XY, Cary R, Laulainen N, Berkowitz C., 2009. Characterization of the sunset semi-continuous carbon aerosol analyzer. *Journal of the Air and Waste Management Association* 59(7), 826-33.

Belz G., A. Cinieri, P.R. Dambruoso, B.E. Daresta, G. de Gennaro, A. Giove, G. Miglietta, R.M. Nacci and C. Tortorella. 2012. Comparison between offline and semi-continuous analyzers for aerosol organic and elemental carbon detection. European Aerosol Conference 2012 Granada. Abstract A-WG08S1P20.

Calzolari, G., Chiari, M., Lucarelli, F., Mazzei, F., Nava, S., Prati, P., Valli, G., Vecchi, R., 2008. PIXE and XRF analysis of particulate matter samples: an inter-laboratory comparison. *Nuclear Instruments and Methods in Physics Research B*, 266, 2401-2404.

Cavalli, F., Viana, M., Yttri, K.E., Genberg, J., Putaud, J.P., 2010. Toward a standardised thermal-optical protocol for measuring atmospheric organic and elemental carbon: the EUSAAR protocol. *Atmospheric Measurement Techniques* 3, 79-89.

Chow, J.C., 1995. Critical review: Measurement methods to determine compliance with ambient air quality standards for suspended particles. *Journal of the Air and Waste Management Association*, 45(5):320-382.

Chow J., J. Watson, Chen L.W. A, Patrickarnott W., Moosmüller H., 2004. Equivalence of Elemental Carbon by Thermal/Optical Reflectance and Transmittance with Different Temperature Protocols, *Environmental Science and Technology* 38, 4414-4422.

Cooper Environmental Services.
<http://cooperenvironmental.com>

El Haddad, I., Marchand, N., Dron, J., Temime-Roussel, B., Quivet, E., Wortham, H., Jaffrezo, J. L. Baduel, C. Voisin, D., Besombes, J. L., Gille, G., 2009. Comprehensive primary particulate organic characterization of vehicular exhaust emissions in France, *Atmospheric Environment* 39, 6190-6198.

- EN 12341:1998: Air Quality – Determination of the PM₁₀ fraction of suspended particulate matter – Reference method and field test procedure to demonstrate reference equivalence of measurement methods, European Standard, CEN, Brussels
- EN 14907:2005: Ambient air quality – Standard gravimetric measurement method for the determination of the PM_{2.5} mass fraction of suspended particulate matter, European Standard, CEN, Brussels
- Georgios F., Martini G., Manfredi U., Marotta A., Krasenbrink A., Maffioletti F., Terenghi R., Colombo M., 2012. Assessment of on-road emissions of four Euro V diesel and CNG waste collection trucks for supporting air-quality improvement initiatives in the city of Milan. *Science of the Total Environment* 426, 65–72.
- Hansen, A. D. A., Rosen H., Novakov T. 1984. "The Aethalometer - an Instrument for the Real-Time Measurement of Optical Absorption by Aerosol Particles." *Science of the Total Environment*, 36 (C): 191-196.
- Khlystov, A., Wyers, G.P., Slanina, J., 1995. The steam-jet aerosol collector. *Atmospheric Environment* 29, 2229-2234.
- Kidwell, C.B., Ondov, J.M., 2001. Development and evaluation of a prototype system for collecting sub-hourly ambient aerosol for chemical analysis. *Aerosol Science and Technology* 35, 596–601.
- Kim Oanh N.T., Pongkiatkul P., Upadhyay N., Hopke P.K., 2009. Designing ambient particulate matter monitoring program for source apportionment study by receptor modeling. *Atmospheric Environment* 43 (21), 3334-44.
- Lagler F., Belis C.A., Borowiak A., 2011. A Quality Assurance and Control Program for PM_{2.5} and PM₁₀ measurements in European Air Quality Monitoring Networks. EUR Report EUR 24851 EN. <http://bookshop.europa.eu/en/a-quality-assurance-and-control-program-for-pm2.5-and-pm10-measurements-in-european-air-quality-monitoring-networks-pbLBN24851/>
- Laskin, A., Laskin J., Nizkorodov S. A., 2012. Mass Spectrometric Approaches for Chemical Characterisation of Atmospheric Aerosols: Critical Review of the most Recent Advances. *Environmental Chemistry* 9 (3), 163-189.
- Lee D S., Lee B., Eom J. W., 2011. A compact semi-continuous atmospheric aerosol sampler for elemental analysis: A preliminary result. *Atmospheric Pollution Research* 2, 506-512
- Montero L., Duane M., Manfredi U., Astorga C., Carriero M.; Martini G., Krasenbrink A., Larsen B.R., 2010. Hydrocarbon Emission Fingerprints from Contemporary Vehicle/Engine Technologies with Conventional and new Fuels. *Atmospheric Environment* 44, 2167-2175.
- Petzold, A. Schönlinner, M., 2004. Multi-angle absorption photometry- A new method for the measurement of aerosol light absorption and atmospheric black carbon, *Journal of Aerosol Science*, 35, 421–441.
- Schkolnik G. and Y. Rudich, 2006. Detection and quantification of levoglucosan in atmospheric aerosols: A review. *Analytical and Bioanalytical Chemistry* 385, 26–33.
- Sjödim A., Ferm M., Björk A., Rahmberg M., Gudmundsson A., Swietlicki E., Johanson C., 2010. Wear particles from road traffic – a field, laboratory and modeling study. Final Report. Swedish Environmental Research Institute. 97 pp.
- Solomon P. A., Mitchell W., Tolocka M., Norris G., Gemmill D., Wiener R., Vanderpool R., Murdoch R., Natarajan S., Hardison E., 2000. Evaluation of PM_{2.5} Chemical Speciation Samplers for Use in the EPA National PM_{2.5} Chemical Speciation Network. EPA Rep. EPA-454/R-01-005, Off. of Air Qual. Plann. and Stand., Research Triangle Park, N. C.
- SPECIATE (<http://www.epa.gov/ttn/chief/software/speciate/>)
- Subramanian R., Khlystov A.Y, Cabada J.C., Robinson A.L., 2004. Positive and Negative Artifacts in Particulate Organic Carbon Measurements with Denuded and Undenuded Sampler Configurations. *Aerosol Science and Technology*, 38:27–48
- Weber, R.J., Orsini, D., Duan, Y., Lee, Y.-N., Klotz, P., Brechtel, F., 2001. A particle-into liquid collector for rapid measurements of aerosol chemical composition. *Aerosol Science and Technology* 35, 718–727.
- Yatkin S. and M. Gerboles. 2013. Investigation of the mass loss of water soluble ions caused by EDXRF analysis of PM₁₀ loaded filters. *Atmospheric Environment* 73, 159-168.

B5. KNOWING YOUR DATASET: BASIC STATISTICS

Before starting any kind of data treatment, it is good practice to make some summary plots and run some simple tests to gain an overview of the relationships between variables and how they change from sample to sample. Many commercial and free software applications are available that can carry out routine statistical tests (e.g. Statistica, Matlab, R, SPSS).

Central and dispersion statistics

Box and whisker plots are useful to visualise central values of your variables (mean, median) and the dispersion of your data around the central values (quartiles, minimum and maximum values).

Check the statistical distribution that best describes your data

Quite often the air pollution data can be better described using a log-normal distribution rather than a normal one. Many statistical tests assume that data is normally distributed even though small deviations from normality are acceptable. In order to better assess the results of standard statistical tests, knowing the statistical distribution of your data could be useful. Box and whisker plots give a visual overview

of the data spread that enables a preliminary assessment of the distribution. Visual tests of normality such as histograms, probability plots and normal probability plots are also useful. For an in-depth evaluation, normality tests such as Kolmogorov-Smirnov (Massey, 1951) or Shapiro-Wilk W (Shapiro and Wilk, 1965; Royston, 1982) can be applied.

Correlation matrices

The correlation between variables can be visually assessed using scatter plots. This is particularly useful to identify anomalous data points (suspected outliers) that may affect the correlation. However, when many variables are involved the use of correlation matrices reporting the Pearson correlation coefficient (r) and related statistics for every possible pair of variables is a useful exploratory technique, provided the influence of outliers has been evaluated (see below).

Linear regression

Ordinary Least Squares regression is the simplest and quickest technique to more in-depth exploration of the association between two variables. The evaluation of the curve parameters

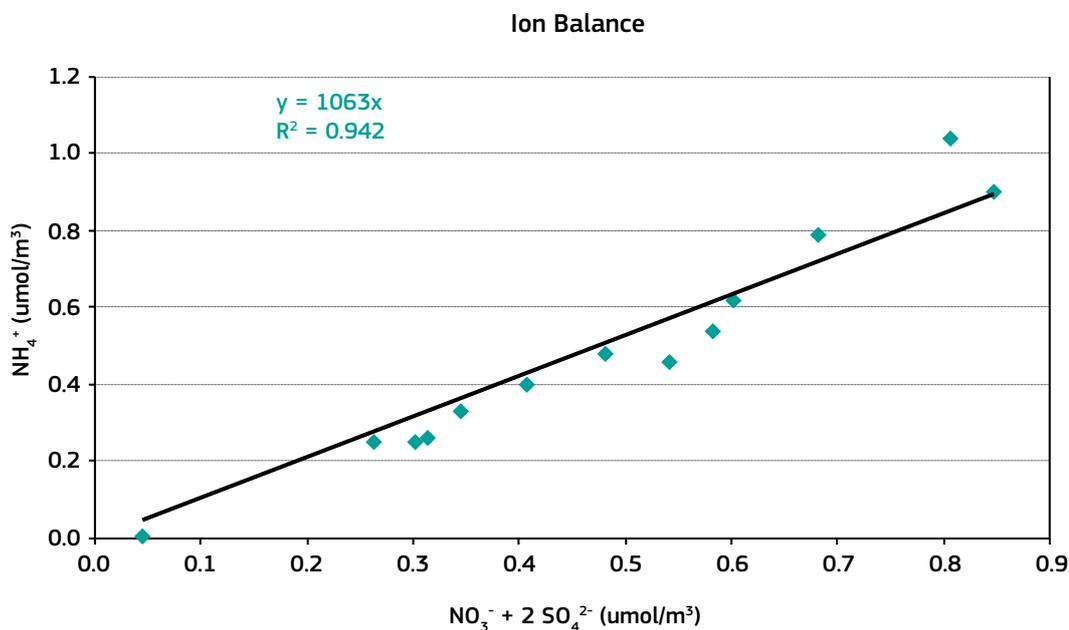
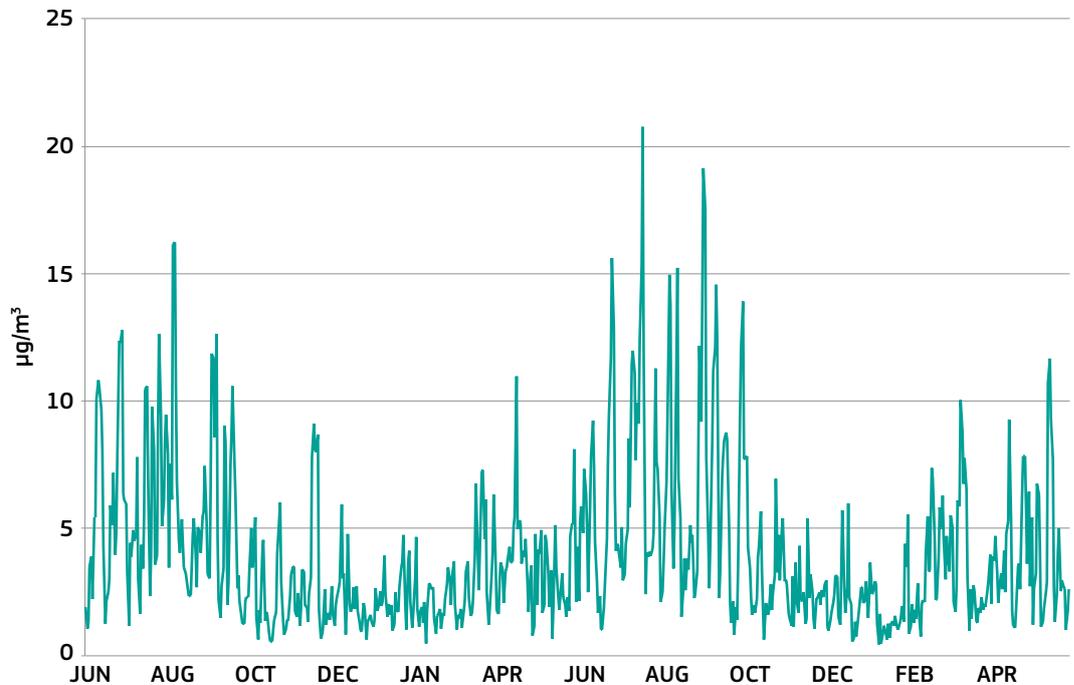


Figure B5.1 Linear regression to test ion balance in PM (B. Larsen, unpublished).

Figure B5.2 Seasonal time trends in sulphates at Saint Louis supersite (Lee et al., 2006).



(intercept and slope and the determination coefficient (r^2)) provides useful preliminary information to describe the (linear) relationship between the variables considered. Precaution should also be taken concerning outliers.

Time trends

Plotting time trends of the variables makes it possible to identify regular patterns in data (e.g. seasonality, influence of the day of the week) or extraordinary events that probably indicate the impact of specific sources influencing the study area for short periods (e.g. Saharan events, wild fires). In addition, when hourly data are available, characteristic daily profiles of certain species can be used to identify specific sources (e.g. a peak of traffic markers during rush hours).

Outliers

Values that do not follow the distribution of data with similar characteristics are referred to as outliers. They may reflect genuine properties of the studied system or derive from measurement errors or anomalies that are not relevant for the model. Outliers can be extreme values or values with unusual relationships with other variables (e.g. ratio).

In some statistics software applications, values above or below the quartiles at a distance of 1.5 - 2 times the interquartile range (height of the box) are labelled as outliers. Of the analytical tests to identify outliers (in normal distributions), the most commonly used is the Grubbs test. This is based on the difference between the mean of the sample and the most extreme or the two most extreme data values, considering

the standard deviation (Grubbs, 1950, 1969). These tests help the practitioner to decide whether these data provide useful information on sources or whether they only introduce noise into the model. It is good practice to report the outliers excluded from the analysis and the reason for their exclusion.

Identify samples of special interest

In source apportionment (SA), it is important to distinguish between a one-off event with a unique profile and an episode that occurs due to the increased contribution from a source with a known profile that is already present in other samples. The analysis of ancillary data is useful to investigate the possible causes of anomalous samples identified with the previous techniques. Meteorological variables such as wind direction, precipitation, or extraordinary events such as forest fires, fireworks, or volcano eruptions may influence the levels of the studied pollutant for short periods.

Spatial distribution

Spatial patterns can be only assessed when many sites are available. At this stage of the study, it can be checked whether the spatial variations of the chemical and physical properties of the aerosol are coherent with geographical gradients in variables that influence the emission of concentrations of atmospheric pollutants (e.g. NaCl is expected to be higher in sites close to the coast; e.g. Schaap et al., 2010).

Ratio-ratio scatter plotting

By representing the concentration (in ambient PM) of two receptor species in a scatter plot, descriptive information can be obtained for an SA dataset in which few sources (or source types) contribute to these species. The data points in the plot will be distributed in an ordered manner between edges, delimited by the emission factors of these pairs of species for each source (type). The advantage of visualising concentrations of receptor compounds in PM that are normalised to concentrations of reference compounds (e.g. EC) in two-dimensional scatter plots was first demonstrated by Robinson et al. (2006).

References

Belis, C. A., Cancelinha, J., Duane, M., Forcina, V., Pedroni, V., Passarella, R., Tanet, G., Douglas, K., Piazzalunga, A., Bolzacchini, E., Sangiorgi, G., Perrone, M. G., Ferrero, L., Fermo, P., and Larsen, B. R., 2011. Sources for PM air pollution in the Po Plain, Italy: I. Critical comparison of methods for estimating biomass burning contributions to benzo(a)pyrene. *Atmospheric Environment* 45, 7266-7275.

Grubbs F. E., 1950. Sample Criteria for Testing Outlying Observations, *Annals of Math. Statistics* 21 27-58.

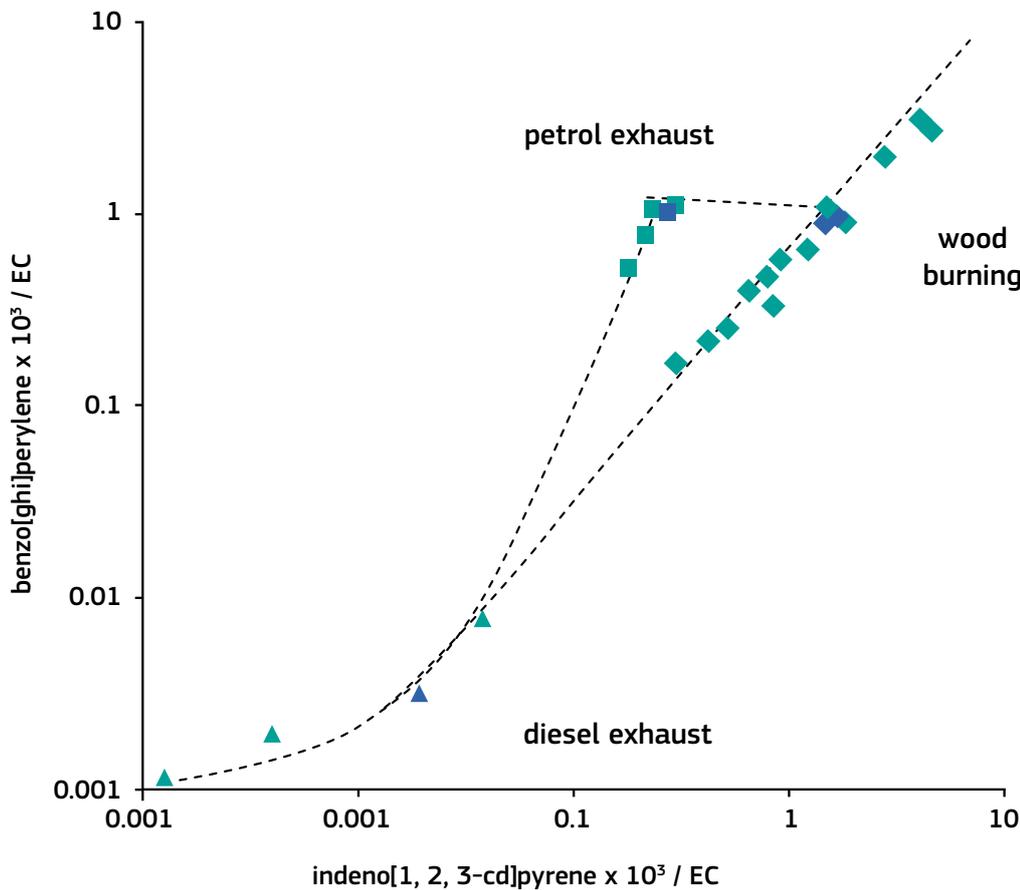


Figure B5.3. Ratio-ratio plots using data on B(ghi)P, Ind(1,2,3)P and EC to visualise the potential contribution of three source scenarios for ambient PM.

Emission source data – triangles: diesel exhausts; squares: petrol exhausts; rhombs: wood burning (from Belis et al., 2011)

- Grubbs F. E., 1969. Procedures for Detecting Outlying Observations in Samples, *Technometrics* 11(1), 13-14.
- Lee, J. H., Hopke, P. K., and Turner, J. R., 2006. Source identification of airborne PM_{2.5} at the St. Louis-Midwest Supersite. *Journal of Geophysical Research D: Atmospheres* 111.
- Massey, F. J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46, 68-78.
- Robinson, A. L., Subramanian, R., Donahue, N. M., Rogge, W. F., 2006. Source Apportionment of Molecular Markers and Organic Aerosol. Polycyclic Aromatic Hydrocarbons and Methodology for Data Visualization, *Environmental Science & Technology* 24, 7803-7810.
- Royston, J. P., 1982. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Applied Statistics* 31(2), 115-124.
- Schaap, M., Weijers, E.P., Mooibroek, D., Nguyen, L., Hoogerbrugge, R., 2010. Composition and origin of Particulate Matter in the Netherlands Results from the Dutch Research Programme on Particulate Matter. Report 500099007/2010. PBL
- Shapiro S. S., Wilk. M. B., 1965. An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52(3/4), 591-611.

B6. PRELIMINARY DATA QUALITY CHECKS

Missing values

In order to run a multivariate analysis, the entries in the dataset must comply with minimum requirements. This applies in particular for factor analysis. It is a common misconception that negative or zero concentrations are harmful for factor analyses. If a true value is zero or near zero, then there is a probability that the corresponding measured value will be negative. Such negative values should be kept in the dataset. They may be rejected only if their confidence interval does not include zero, which would obviously indicate a measurement error. If negative values are truncated to zero, then a modelling error is caused, and the data becomes biased. ‘Least squares’ (LS) methods are not appropriate for such kind of data. Unfortunately, some measurement techniques are not able to produce unbiased near-zero values. How to deal with such biased values is still an open question. The most promising approach in Multilinear Engine 2 (ME-2) seems to be to use error model code -16 (see box below).

Zero or negative *uncertainties* have no physical meaning and therefore should be excluded from the input file or replaced by reasonable values. Since it is not possible to perform the analysis when empty cells are present in the input data matrix, missing values should be handled in advance by the operator. The simplest choice is to cancel the row (sample) or the column (species) from the input matrix. However, this may cause the loss of important information. An alternative approach is to substitute missing values with estimated values, such as the mean, the median or the geometric mean of the measured concentrations of the species, in all the samples of that particular study site (Polissar et al., 1998). The procedure by Polissar et al. (1998) is often used without testing its validity for any given dataset. Scientists should find, for every dataset, the most suitable uncertainties of the substituted values to avoid distorting the model. In the EPA Unmix 6.0 receptor model, for instance, there is an automated subroutine that substitutes missing values using the maximum and

minimum ratios of the variable for which the value is missing (EPA Unmix 6.0 user manual). It should be noted that for any receptor model, the more missing values are reconstituted, the greater the uncertainty of the source contribution estimates. As a rule of thumb, missing values substituted for a given species should not be more than half of the samples (Brown & Hafner, 2005).

Values below the detection Limit

Values below the detection limit (BDL) of the analytical method should be used if they are available. If values are not provided by the laboratory they can be substituted either by zero (or by a value sufficiently close to zero), by the detection limit itself or by a fraction of the detection limit. The most common practice is to substitute BDL values with half of the detection limit (Polissar et al., 1998). Substitution of BDL values only makes sense if the number of values above the detection limit of that species in the dataset is sufficient to provide information about sources. It has been suggested that it is only worth including species that present more than 50% of BDL values in the data treatment if the signal-to-noise ratio is reasonable (see below) or the species is a tracer (Brown & Hafner, 2005).

Note: the official nomenclature (IUPAC, 1997-2006) defines *detection limit* (DL) as the minimum value that can be distinguished from the blanks, and *limit of detection* (LOD) as the blank value plus a multiple of the standard deviation of this measurement. In analytical chemistry, LOD is commonly considered to be the lowest analyte concentration at which detection is feasible; measurements below that value are reported as “<LOD”.

Advanced User Box

There are cases in which the substitution of many values below the detection limit in several trace species creates an artificial factor containing trace species with a characteristic pattern. This “ghost factor” is generated by the model to fit the substituted values in all those species in which they occur simultaneously. When using ME-2-based analysis tools, it is possible to avoid such distortions by using the special error model code -16 (Paatero, 2000) for all substituted data points. This code stipulates that all fitted values below the detection limit are to be considered a perfect fit, with Q contribution (see chapter B9) equal to zero. This alternative is not yet implemented in version 3.0 of the US - EPA Positive Matrix Factorization (PMF), but it may be used when controlling ME-2 using home-made scripts.

Signal-to-noise

The signal-to-noise ratio (S/N) is defined as the power ratio between a desired signal (S, meaningful information) and the background noise (N, unwanted signal).

In receptor model analysis this can be interpreted as the relationship between concentrations (x) and uncertainties (s) (Paatero and Hopke, 2003):

$$\left(\frac{S}{N}\right)_j = \sqrt{\frac{\sum_{i=1}^n x_{ij}^2}{\sum_{i=1}^n s_{ij}^2}} \quad (\text{B6.1})$$

In the EPA PMF v3, the equation is even stricter and considers only the portion of the concentration that exceeds the uncertainty (EPA-PMF 3.0 User Guide):

$$\left(\frac{S}{N}\right)_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - s_{ij})^2}{\sum_{i=1}^n s_{ij}^2}} \quad (\text{B6.2})$$

During the European RM intercomparison (Karagulian et al., 2012), it was discovered that both of the above-mentioned equations for S/N fail totally if a species contains strongly down-weighted values or if different matrix rows contain different scaling factors, e.g. some in mg and others in μg . The next version of the EPA PMF will contain an improved expression that should work well for all kinds of data, even those where different rows have different

scaling factors. When using this new expression, the numerical limits for weak/bad/good discrimination must be changed from the customary values shown below.

The signal-to-noise ratio is useful for classifying variables according to the information they supply for the source identification analysis. According to Paatero and Hopke (2003), variables with signal-to-noise ratios below 0.2 (bad) are to be excluded from the analysis, while variables where the ratio falls between 0.2 and 2.0 (weak) are suitable for the analysis. However, it is recommended that such variables be down-weighted by a factor of 1/2 or 1/3.

Mass closure and ion balance

Preliminary tests exist to match the masses or the electric charges of species. In the first case, mass closure is accomplished by comparing the mass of particulate matter (PM) to the sum of the masses of the major chemical components. For this calculation, organic carbon (OC) is to be converted into organic matter (OM) using an empirical coefficient that normally ranges from 1.4 to 2.1 (e.g. Turpin and Lim, 2001). The mass of crustal fraction must also be estimated from elements, as these are frequently present as oxides or carbonates. Therefore, the mass of the missing oxygen and carbon atoms should be added. The following empirical equations have been proposed to estimate these kinds of materials, by accounting for unmeasured oxides in minerals (Watson et al., 2002; Malm & Hand, 2007):

$$\text{Geological} = 1.89\text{Al} + 2.14\text{Si} + 1.4\text{Ca} + 1.43\text{Fe} \quad (\text{B6.3})$$

$$\text{Soil} = 2.2\text{Al} + 2.49\text{Si} + 1.94\text{Ti} + 1.63\text{Ca} + 2.42\text{Fe} \quad (\text{B6.4})$$

Commonly, the mass of PM, determined with the gravimetric method, is higher than the sum of the chemical components. This can be explained in different ways: a) not all the relevant chemical components have been determined; b) the mass measurement includes water adsorbed to particles that is not quantified in the chemical analyses; c) the selected coefficient for converting OC to OM is not optimal for the study area; d) the elements that have been assumed to be present as oxides and carbonate have not been taken into consideration.

By comparing the sum of anion equivalents with the sum of cation equivalents, it is possible to assess departure from neutrality, and plotting values in a graph helps to identify

samples with an atypical ionic composition. The most common ionic species in PM are inorganic cations (ammonium, sodium, potassium, calcium, and magnesium) and inorganic anions (sulphate, nitrate, chloride, and carbonate). Among the organic acids, the most relevant anions are those deriving from oxalic, malonic, succinic, formic, and acetic acids (Chebbi & Carlier, 1996).

It is also possible to develop simplified mass closure models which provide an excellent check on the consistency of data from individual samples. An example is the Pragmatic Mass Closure Model (Harrison et al., 2003) which uses simple empirical parameterisations to account for the measured mass of particles in terms of a small number of analytical variables. Although such a model might be expected to be site-specific, it has proved to be transferable between sites, years and particle size fractions (Yin and Harrison, 2008). However, caution should be exercised in the application of the model to sites with entirely different pollution traits. Simple empirical corrections should be feasible in such cases.

Analysis of consistency in time and space

In order to populate a dataset with an appropriate number of samples it may be necessary to collect data for more than one year. However, species and other variables collected during different years may show different relationships. In order to check these patterns before running the analysis, scatter plots to look for edges (Henry, 2003) or time trend plots are useful. Changes in sampling methodologies or analytical techniques may create disruption in time series that must be duly taken into account during data elaboration. Comparing time series from different sites is helpful to detect anomalous patterns. Nevertheless, it must be considered that different monitoring networks may have different instrumentation (e.g. different inlets, different operation principles) or different data treatment protocols.

If several receptor sites have been operated near each other, e.g. within one city area, then it may be useful to soft-constrain regional factors (more details on constrained models in section C4), such as secondary sulphate, in order to have similar factors at all sites. In this way, a significant part of rotational uncertainty may be avoided. It should be kept in mind that the secondary sulphate G factor often has no rotation-limiting zero values and hence is prone to rotations if no constraints are applied.

References

- Brown S.G., Hafner H., 2005. *Multivariate Receptor Modeling Workbook*. Sonoma Technology Inc. Petaluma, CA. U.S.A. 236 pp. http://www.sonomatechdata.com/sti_workbooks/
- Chebbi, A., Carlier, P., 1996. Carboxylic acids in the troposphere, occurrence, sources, and sinks: a review. *Atmospheric Environment* 30, 4233–4249.
- Harrison, R.M., Jones, A.M. Lawrence, R.G., 2003. A Pragmatic Mass Closure Model for Airborne Particulate Matter at Urban Background and Roadside Sites. *Atmospheric Environment* 37, 4927–4933.
- Henry R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. *Chemometrics and Intelligent Laboratory Systems* 65, 179–89.
- IUPAC. *Compendium of Chemical Terminology*, 2nd ed. (the “Gold Book”). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford, 1997. XML on-line corrected version: <http://goldbook.iupac.org>, 2006 created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook.
- Karagulian F., Belis C. A., Borowiak A., 2012. Results of the European Intercomparison exercise for receptor models 2011–2012. Part I. JRC Report EUR 25727 EN.
- Norris G., Vedantham R., K. Wade, S. Brown, J. Prouty, C. Foley, 2008. *EPA Positive Matrix Factorization (PMF 3.0) Fundamentals & User Guide*. EPA 600/R-08/108.
- Norris G, Vedantham R., Duvall R., Henry R. C., 2007. *EPA Unmix 6.0 Fundamentals & User Guide*. EPA/600/R-07/089.
- Malm W., Hand J.L., 2007. An examination of the physical and optical properties of the aerosols collected in the IMPROVE program. *Atmospheric Environment* 41, 3407–3427
- Paatero P., Hopke P.K., 2003. Discarding or downweighting high-noise variables in factor analytic models. *Analytica Chimica Acta* 490 277–289.

- Polissar, A. V., Hopke P. K., Paatero P., Malm W. C., and Sisler J. F., 1998. Atmospheric aerosol over Alaska 2. Elemental composition and sources, *Journal of Geophysical Research D: Atmospheres*, 103(D15), 19045-19057.
- Turpin, B.J., Lim, H-J., 2001. Species Contributions to PM_{2.5} Mass Concentrations: Revisiting Common Assumptions for Estimating Organic Mass. *Aerosol Science and Technology* 35, 602–610.
- Watson J. G., Zhu T., Chow J. C., Engelbrecht J., Fujita E. M., Wilson W. E., 2002. Receptor modeling application framework for particle source apportionment. *Chemosphere* 49, 1093–1136.
- Yin, J., Harrison, R.M., 2008. Pragmatic Mass Closure Study for PM₁₀, PM_{2.5} and PM₁₀ at Roadside, Urban Background and Rural Sites. *Atmospheric Environment*. 42, 980-988.

B7. INPUT DATA UNCERTAINTY CALCULATION

Uncertainty is the quantitative estimation of the quality of a measurement that makes it possible to compare results among themselves and with reference values (Joint Committee for Guides in Metrology (JCGM) 100:2008). Estimating the uncertainty of measurements is a common practice in analytical chemistry and physics that is performed routinely according to international criteria laid down in standards and implemented in reference methods. In analytical chemistry, uncertainty is evaluated both as the standard deviation of repeated observations and by comparison with reference materials.

In source apportionment, analytical uncertainty is important since the most commonly used models, like PMF and CMB, require the uncertainty of the species concentrations as input data in order to find the solution and the uncertainty of the output.

In PMF analysis, uncertainty estimation is particularly critical because every entry is weighted according to its uncertainty. Although analytical uncertainty estimation is an important step of receptor modelling, it must be noted that it is only one component of the overall input data uncertainty required by receptor models (Polissar et al., 1998). Other contributions to the overall uncertainty include flow rate uncertainty, between-sampler uncertainty and other unidentified noise.

Moreover, not all components of overall uncertainty behave equally. In PMF input, only the components of uncertainty that are capable of generating residuals, i.e. components that will increase the Q value of the fit (see section B9), should be included. Flow rate uncertainty is a prime example: flow rate variations influence all values on a matrix row by the same multiplier, hence causing no increase of residuals e_{ij} (equation A.1). Flow rate uncertainty, and other similar uncertainties, should bypass the PMF stage and be attached directly to the computed G factor elements (equation A.1). In addition to analytic uncertainty, modelling errors (e.g. variation of source profiles with time, chemical transformations during transport from source

to receptor) also cause residuals in PMF modelling. Expected contributions from modelling errors must also be accounted for in the PMF input data uncertainties. There is no fixed rule for such contributions. To begin with, it is reasonable to include 10% of each data value as a provision for modelling errors. When experience is accumulated, this numerical coefficient may be adjusted. However, this additional uncertainty must always be reported in publications so that the work is reproducible.

The operator also needs to attribute an uncertainty to missing values and to values below the detection limit. That uncertainty is normally higher than that of measured values. Polissar et al. (1998) set the uncertainty of values below the detection limit to 5/6 of the detection limit, while the uncertainty of missing values is by convention set at four times the geometric mean. This convention has no general statistical basis. For some datasets, significantly larger uncertainty values are needed for missing values.

Sometimes the attribution of uncertainties may be achieved by a trial-and-error process that aims to obtain the best model fit which is evaluated using Q values (see paragraph B9), scatterplots, distribution of residuals and results from multiple regressions (e.g. Polissar et al., 2001).

When dealing with databases in which single entry uncertainties are unavailable or are inappropriate for modelling purposes, the global input data uncertainties may be estimated using equation-based approaches, which rely on the species detection limit (DL), empirical constants (k), species concentration (C) and/or the coefficient of variation (CV) (Reff et al., 2007, Karagulian & Belis, 2012).

Analytical uncertainty can be estimated by the linear regression described in equation B7.1 where σ_a is the uncertainty of the analytical procedure, m is the mass of the analyte, and σ_0 and α are fitting parameters (Anttila et al., 1995):

$$\sigma_a^2 = \sigma_0^2 + (\alpha m)^2 \quad (\text{B7.1})$$

In the estimated fractional uncertainties (EFU) method the error structures (s_{ij}) are (Kim and Hopke, 2005):

$$S_{ij} = [DL]/3 + kx \quad (B7.2)$$

When no empirical constants are used other than the DL and coefficient of variation (CV), the analytical uncertainty is (Chow et al., 2007):

$$\sigma_{i,t}^a = \sqrt{DL_i^2 + (CV_i \times C_{i,t})^2} \quad (B7.3)$$

Sampling contributes to the uncertainty of measured values due to sampling volume uncertainty, selective effect and other artefacts caused by the sampler inlet, and losses due to sample transport and conservation. These contributions can be assessed with field tests (e.g. collocated measurements and comparison with reference instrumentation and techniques). In the case of destructive analysis of the filters where PM is collected, the procedure of subtracting blank filter (different from sampled ones) concentrations is an additional source of uncertainty. Sampling and blank subtraction uncertainties have been incorporated into the input data uncertainty by Amato et al. (2009):

$$\sigma_A^2 = \sigma_a^2 + \sigma_{BLK}^2 \quad (B7.4)$$

$$\sigma_{ij}^2 = \sqrt{\frac{\sigma_A^2}{V_i^2} + (\beta x_{ij})^2} \quad (B7.5)$$

where the standard deviation of species concentrations in blank filters (σ_{BLK}), the sampled volume (V_i) and a coefficient (β) are used to account for the additional uncertainty sources.

Input data uncertainties can also be estimated with the PMF2 software. This is a more complex procedure that uses three codes, C1, C2 and C3, the error model and the arrays T, U and V (Paatero, 2004).

CMB uses source profiles as input data with associated uncertainty estimation. When source profiles are too similar, CMB may be not able to find a solution (collinearity). In order to prevent problems related to collinearity, sources with similar chemical composition are either combined into source categories / composite profiles or only one profile is incorporated in the analysis while the other is dropped. The uncertainty of the composite is obtained by propagation of the uncertainty of the pooled single profiles (Watson, 2004). However, this may not fully account for the variety of similar sources in the study area and their variability over time.

In order to deal with the variability of source profiles, initial model runs often contain many profiles, and a sensibility test should be carried out to assess their influence on the precision and stability of the source contribution estimates.

A default value of zero with a standard deviation equal to the analytical detection limit may be assigned to a species of a source profile if that species is known to be absent from that source (Watson, 2004).

Metals are excellent receptor species given the assumption that such receptor species do not chemically react or physically repartition during transport from source to receptor. As such, metals have been used from the very beginning of receptor modelling activities (e.g. Hopke et al., 1991). However, in the search for specific receptor species for different combustion sources (also called molecular markers), the use of organic chemical compounds has grown popular in modern source apportionment studies, even though this class of compounds often comes into conflict with the above-mentioned assumption. An interesting utilisation of uncertainty data for the inclusion of semi-volatile and photo-chemically reactive species in CMB and PMF has been developed and adopted by Latella et al. (2005), Junninen et al. (2009), Belis et al. (2011) and Larsen et al. (2012). In these studies, methods are described for using information on volatility to account for the re-partitioning processes (PAHs; semi-volatile organic carbon fraction) and photochemical degradation (hydrocarbons, levoglucosan) from source to receptor. This information has served as error-input to CMB and PMF for the error weighting in the statistical procedures.

References

- Amato, F., Pandolfi M., Escrig A., Querol X., Alastuey A., Pey J., Perez N., Hopke P. K., 2009. Quantifying road dust re-suspension in urban environment by Multilinear Engine: A comparison with PMF2, *Atmospheric Environment* 43(17), 2770-2780.
- Anttila, P., Paatero P., Tappert U., Jarvinen O., 1995. Source identification of bulk wet deposition in Finland by positive matrix factorization, *Atmospheric Environment* 29(14), 1705-1718.
- Belis, C. A., Cancelinha, J., Duane, M., Forcina, V., Pedroni, V., Passarella, R., Tanet, G., Douglas, K., Piazzalunga, A., Bolzacchini, E., Sangiorgi, G., Perrone, M. G., Ferrero, L., Fermo, P., and Larsen, B. R., 2011. Sources for PM air pollution in the Po Plain, Italy: I. Critical comparison of methods for estimating biomass burning contributions to benzo(a)pyrene. *Atmospheric Environment* 45, 7266-7275.
- Chow, J. C., Watson J. G., Chen L. W. A., Chang M. C. O., Robinson N. F., Trimble D., Kohl S., 2007. The IMPROVE_A temperature protocol for thermal/optical carbon analysis: Maintaining consistency with a long-term database. *Journal of the Air and Waste Management Association* 57(9), 1014-1023.
- Hopke P.K. (ed.), 1991. Receptor Modelling for air quality management. Data handling in science and technology- volume 7. Elsevier. ISBN 0-444-88218-9
- Joint Committee for Guides in Metrology, 2008. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML.
- Karagulian F., Belis C.A, 2012. Enhancing Source Apportionment with receptor models to Foster the Air Quality Directive Implementation. *International Journal of Environmental Pollution*. 50: 190-199.
- Larsen, B.R., Gilardoni S., Stenström K., Niedzialek J., Jimenez J., Belis C. A., 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. *Atmospheric Environment* 50, 203-213.
- Latella, A, Stani G., Cobelli L., Duane M., Junninen H., Astorga C., Larsen B.R., 2005. Semicontinuous GC analysis and receptor modelling for source apportionment of ozone precursor hydrocarbons in Bresso, Milan, 2003. *Journal of Chromatography A* 1071, 29–39.
- Kim, Y. J., Hopke P. K., 2005. Estimation of Organic Carbon Blank Values and Error Structures of the Speciation Trends Network Data for Source Apportionment. *Journal of the Air and Waste Management Association* 55, 1190-1199.
- Paatero P., 2004. User's guide for positive matrix factorization programs PMF2 and PMF3, Parts 1 and 2. University of Helsinki, Helsinki, Finland.
- Polissar, A. V., Hopke P. K., Paatero P., Malm W. C., Sisler J. F., 1998. Atmospheric aerosol over Alaska 2. Elemental composition and sources. *Journal of Geophysical Research D: Atmospheres* 103(D15), 19045-19057.
- Polissar A.V., Hopke P.K. Poirot R.L., 2001. Atmospheric aerosol over Vermont: chemical composition and sources. *Environmental Science and Technology*. 35, 4604-4621.
- Reff, A., Eberly, S.I., Bhave, P.V., 2007. Receptor modelling of ambient particulate matter data using Positive Matrix Factorization: review of existing methods. *Journal of Air & Waste Management Association* 57, 146-154.
- Watson, J. G., 2004. Protocol for Applying and Validating the CMB model to PM2.5 and VOCs, US EPA, USA. Report EPA-451/R-04-001. 173 pp.

B8. CHEMICAL MASS BALANCE MODELS

Chemical Mass Balance (CMB) is based on the mass conservation of individual chemical species: ions, elements, and organic compounds, which are commonly referred to as markers. In the mass conservation equations, deriving from the general equation A.1 (here the original notation was kept to facilitate consultation of the references and our notation is reported between parentheses), known concentrations $C_{ik}(x_{ij})$ of specific species at a receptor site k are written as the product of unknown source contributions $s_{jk}(g_{jk})$ and known source profiles $a_{ij}(f_{kj})$ (Cooper et al., 1984; Watson et al., 1998). $a_{ij}(f_{kj})$ are the fractional abundances of the species in the source emissions, commonly expressed by the ratios between the species and the $PM_{2.5}$ or organic carbon mass. The mass conservation equations for each species emitted from m (p) sources can be written as follows:

$$C_{ik} = \sum_{j=1}^m a_{ij} s_{jk} \quad (\text{B8.1})$$

In practice, the set of linear equations generated by equation B8.1 is solved with an effective variance-weighted least square method using the EPA-CMB8.2 software. Note that although equation B8.1 is similar to equation A.1 (in this case $a_{ij}(f_{kj})$ are known values), the model is conceived for one sample per site and has no residual term.

Friedlander (1973) proposed a modified version of equation B8.1 that included a coefficient, a_{ij} , that accounted for changes in the profile values for specific species in transit. However, the current practice is to apportion the primary material that has not changed between source and receptor, so this coefficient is set to 1. The remaining quantities of reactive species such as ammonium, nitrate, sulphate, and organic carbon are then indirectly apportioned to secondary sources. Accordingly, the species used as fitting species are strictly of primary origin. They must be (i) stable during atmospheric transport (i.e. low volatility and moderately reactive), (ii) accurately determined at the receptor site and (iii) reported for all source profiles considered

in the model. The number of fitting species has to exceed the number m of emission sources. The first attempts to solve the mass balance equation were based on tracer compounds (ideally one for each source, e.g. Miller, 1972). Since inorganic compounds rarely derive from a single source, this approach gave way to another that considers a higher number of species than sources. This latter approach was fully developed in the Chemical Mass Balance Model as described by Watson et al. (1997), among others. More recently, the identification of organic compounds that can be used as tracers for specific sources or types of sources (e.g. Schauer, 1999a and b) led to a combination of both approaches, i.e. containing more species than sources but including some organic species (tracer or markers) deriving from unique sources (e.g. Chow, 2007; Subramanian, 2006).

The main strength of the CMB model is that, unlike other statistical receptor models (e.g. PMF), it does not require a large dataset and theoretically equation B8.1 can be solved for an individual sample (see section B3). Moreover, unlike factor analysis techniques, the CMB output does not require additional identification of the contributing sources/factors, as the profiles are selected *a priori* for well-defined sources.

However, the most important issue generally encountered in CMB modelling is the selection of the source profiles that best represent the aerosol collected at the receptor site. This selection relies heavily on two implicit assumptions:

- (i) The aggregate emissions from a given source class are well represented by an average source profile with well-known a_{ij} ratios.
- (ii) All the major primary sources of the species are included in the model.

With most commonly measured species for particles (e.g. ions, elements, carbon and organic compounds) and common source types, approximately four to eight primary source classes are linearly independent and can thus

be apportioned by the CMB. These conventionally comprise traffic emissions which are often separated between diesel and gasoline combustion engines, biomass burning, vegetative detritus, cooking emissions and dust (e.g. Zheng et al., 2006a and b; Sheesley et al., 2007; Docherty et al., 2008; Stone et al., 2008; Favez et al., 2010). Additional profiles can also be selected to specifically represent the area of study, including coal burning (Rutter et al., 2009), metal smelting (El Haddad et al., 2011), metallurgical coke production (Subramanian et al., 2007; El Haddad et al., 2011) and shipping/heavy fuel oil combustion (Minguillon et al., 2008; El Haddad et al., 2011).

Currently in the literature, there are a great number of profiles and composite profiles for the major primary sources (e.g. more than 50 profiles for traffic emissions and more than 40 profiles for biomass smoke). A comparison of these profiles reveals significant variations in emissions depending on the fuel type and combustion conditions, rendering the choice between these profiles very complex. Subramanian et al. (2007) show that library profiles may not always reflect the properties of a specific source in a given study area.

To achieve CMB analysis and validation, a number of steps must be followed.

First, for each source, several profiles and composite profiles have to be selected based on the specificity of the study area (e.g. harbour, industries, wood or coal burning, predominance of diesel cars, etc.) and the species concentrations at the receptor site. Examining diagnostic ratios between species can help to eliminate outlier profiles (Robinson et al., 2006a, b and c; El Haddad et al., 2011). Constructing composite profiles from available data (Sheesley et al., 2007; Favez et al., 2010) or developing new source profiles through real world measurements, for instance tunnel experiments (e.g. Phuleria et al., 2006; El Haddad et al., 2009) and open fires (e.g. Lee et al., 2005), are also common practices carried out to better represent the emissions in the study area.

Second, the model is run repeatedly, including different combinations of the selected profiles. Based on the quality of the CMB solutions, the best combinations can be selected. The sensitivity of the results to the choice of the profiles and the related uncertainties can be assessed. As a quality control check of the CMB calculation, statistical performance measures include the use of R-square (target 0.8–1.0), chi-square (target 0–4.0) and the species' calculated-to-measured ratios (target $0.5 < C/M < 2$), as indicators of the goodness of fit (Watson et al., 1998).

If the CMB solutions do not meet these criteria, it would mean that one of the two aforementioned assumptions is transgressed (i.e. non-representative or missing profiles).

The CMB also provides the uncertainties of the source contribution estimates by propagating the uncertainty estimates of the receptor data and source profiles (entered as input by the operator) through the effective-variance least squares calculations. Their magnitudes are a function of the uncertainties in the input data and of the amount of collinearity (i.e. degree of similarity) among source profiles. Two or three times the standard error may be taken as an upper limit of the source contribution.

Third, CMB is often applied to the carbonaceous component of PM and, if the results are combined with those of other analytes using a simple mass closure approach, this can be a valuable check on data quality (e.g. Yin et al., 2010). A further useful check is whether the concentration of organic carbon unaccounted for in a CMB model and assumed to be secondary in origin can be compared with independent estimates of secondary organic carbon derived using the elemental carbon tracer method as reported by Yin et al. (2010).

When available, soluble potassium, water-soluble organic carbon, radiocarbon and Aethalometer measurements can also help corroborate the CMB outputs, especially in the case of high contributions from secondary organic aerosols and biomass burning organic aerosols (e.g. Docherty et al., 2008; Favez et al., 2010).

References

- Cooper, J.A., Watson, J.G., Huntzicker, J.J., 1984. The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model, *Atmospheric Environment* 18, 1347-1355.
- Chow, J. C., Watson J. G., Lowenthal D. H., Chen L. W. A., Zielinska B., Mazzoleni L. R., Magliano K. L., 2007. Evaluation of Organic Markers for Chemical Mass Balance Source Apportionment at the Fresno Supersite. *Atmospheric Chemistry and Physics* 7 (7), 1741-1754.
- Docherty, K. S., Stone, E. A., Ulbrich, I. M., DeCarlo, P. F., Snyder, D. C., Schauer, J. J., Peltier, R. E., Weber, R. J., Murphy, S. M., Seinfeld, J. H., Grover, B. D., Eatough, D. J., Jimenez, J. L., 2008. Apportionment of Primary and Secondary Organic Aerosols in Southern California during the 2005 Study of Organic Aerosols in Riverside (SOAR-1), *Environmental Science & Technology* 20, 7655-7662.

- El Haddad, I., Marchand, N., Dron, J., Temime-Roussel, B., Quivet, E., Wortham, H., Jaffrezo, J. L., Baduel, C., Voisin, D., Besombes, J. L., Gille, G., 2009. Comprehensive primary particulate organic characterization of vehicular exhaust emissions in France, *Atmospheric Environment* 39, 6190-6198.
- El Haddad, I., Marchand, N., Wortham, H., Piot, C., Besombes, J.-L., Cosic, J., Chauvel, C., Armengaud, A., Robin, D., Jaffrezo, J.-L., 2011. Primary sources of PM_{2.5} organic aerosol in an industrial Mediterranean city, Marseille, *Atmospheric Chemistry and Physics* 11, 2039–2058, doi:10.5194/acp-11-2039-2011.
- Favez, O., Haddad, I. E., Piot, C., Boréave, A., Abidi, E., Marchand, N., Jaffrezo, J.-L., Besombes, J.-L., Personnaz, M.-B., Sciare, J., Wortham, H., George, C., D'Anna, B., 2010. Inter-comparison of source apportionment models for the estimation of wood burning aerosols during wintertime in an Alpine city (Grenoble, France). *Atmospheric Chemistry and Physics discussion*, 559-613.
- Friedlander, S.K., 1973. Chemical Element Balances and Identification of Air Pollution Sources. *Environmental Science and Technology* 7, 235-240.
- Lee, S., Baumann, K., Schauer, J. J., Sheesley, R. J., Naeher, L. P., Meinardi, S., Blake, D. R., Edgerton, E. S., Russell, A. G., Clements, M., 2005. Gaseous and Particulate Emissions from Prescribed Burning in Georgia, *Environmental Science & Technology* 23, 9049-9056.
- Miller, M.S., Friedlander S.K., Hiddy G.M.. 1972. A chemical element balance for the Pasadena aerosol. *Journal of Colloidal Interface Science* 39, 65-176.
- Minguillon, M. C., Arhami, M., Schauer, J. J., Sioutas, C., 2008. Seasonal and spatial variations of sources of fine and quasi-ultrafine particulate matter in neighbourhoods near the Los Angeles-Long Beach harbor, *Atmospheric Environment* 42, 7317–7328.
- Phuleria, H. C., Geller, M. D., Fine, P. M., Sioutas, C., 2006. Size-resolved emissions of organic tracers from light-and heavy-duty vehicles measured in a California roadway tunnel. *Environmental Science & Technology* 13, 4109-4118.
- Robinson, A. L., Subramanian, R., Donahue, N. M., Rogge, W. F., 2006a. Source Apportionment of Molecular Markers and Organic Aerosol. Polycyclic Aromatic Hydrocarbons and Methodology for Data Visualization. *Environmental Science & Technology* 24, 7803-7810.
- Robinson, A. L., Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., Rogge, W. F., 2006b. Source Apportionment of Molecular Markers and Organic Aerosol. 2. Biomass Smoke. *Environmental Science & Technology* 24, 7811-7819,
- Robinson, A. L., Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., Rogge, W. F., 2006c. Source Apportionment of Molecular Markers and Organic Aerosol. 3. Food Cooking Emissions, *Environmental Science & Technology* 24, 7820-7827.
- Rutter, A. P., Snyder, D. G., Schauer, J. J., Deminter, J., Shelton, B., 2009. Sensitivity and Bias of Molecular Marker-Based Aerosol Source Apportionment Models to Small Contributions of Coal Combustion Soot. *Environmental Science & Technology* 20, 7770-7777.
- Schauer, J. J., Kleeman M. J., Cass G. R., and Simoneit B. R. T., 1999a. Measurement of Emissions from Air Pollution Sources. 1. C1 through C29 Organic Compounds from Meat Charbroiling." *Environmental Science and Technology* 33 (10), 1566-1577.
- Schauer, J. J., M. J. Kleeman, G. R. Cass, and B. R. T. Simoneit. 1999b. Measurement of Emissions from Air Pollution Sources. 2. C 1 through C 30 Organic Compounds from Medium Duty Diesel Trucks. *Environmental Science and Technology* 33 (10), 1578-1587.
- Sheesley, R. J., Schauer, J. J., Zheng, M., Wang, B., 2007. Sensitivity of molecular marker-based CMB models to biomass burning source profiles. *Atmospheric Environment* 39, 9050-9063.
- Stone, E. A., Snyder, D. C., Sheesley, R. J., Sullivan, A. P., Weber, R. J., Schauer, J. J., 2008. Source apportionment of fine organic aerosol in Mexico City during the MILAGRO experiment 2006. *Atmospheric Chemistry and Physics* 5, 1249-1259.

- Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., Rogge, W. F., Robinson, A. L., 2006. 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* 40 8002–8019
- Subramanian, R., Donahue, N. M., Bernardo-Bricker, A., Rogge, W. F., Robinson, A. L., 2007. Insights into the primary-secondary and regional-local contributions to organic aerosol and PM_{2.5} mass in Pittsburgh, Pennsylvania. *Atmospheric Environment* 41, 7414-7433
- Watson, J. G., Robinson, N. F., Fujita, E. M., Chow, J. C., Pace, T. G., Lewis, C., Coulter, T. 1998. CMB8 Protocol for Applying and Validating PM_{2.5} and VOCs, US EPA, USA.
- Watson, J.G., Robinson, N.F., Lewis, C.W., Coulter, C.T., Chow, J.C., Fujita, E.M., Lowenthal, D.H., Conner, T.L., Henry, R.C., Willis, R.D., 1997. Chemical Mass Balance Receptor Model Version 8 (CMB) User's Manual. Reno. NV, US.
- Yin, J., Harrison, R.M., Chen, Q., Rutter, A. and Schauer, J.J., 2010. Source Apportionment of Fine Particles at Urban Background and Rural Sites in the UK Atmosphere, *Atmospheric Environment* 44, 841-851.
- Zheng, M., Hagler, G. S. W., Ke, L., Bergin, M. H., Wang, F., Louie, P. K. K., Salmon, L., Sin, D. W. M., Yu, J. Z., Schauer, J. J., 2006a. Composition and sources of carbonaceous aerosols at three contrasting sites in Hong Kong. *Journal of Geophysical Research-Atmospheres*, D20.
- Zheng, M., Ke, L., Edgerton, E. S., Schauer, J. J., Dong, M. Y., Russell, A. G., 2006b. Spatial distribution of carbonaceous aerosol in the southeastern United States using molecular markers and carbon isotope data. *Journal of Geophysical Research-Atmospheres*, D10.

B9. FACTOR ANALYSIS I: SELECTION OF THE NUMBER OF FACTORS AND DEALING WITH ROTATIONAL AMBIGUITY (PMF)

The goal of Positive Matrix Factorization (PMF) - like any other multivariate receptor model (RM) - is to identify a number of factors p , the species profile f of each source, and the amount of mass g contributed by each factor to each individual sample (equation A.1).

PMF is an advanced factor analysis technique based on the work of Paatero and Tapper (1994); it uses realistic error estimates to weigh data values and imposes non-negativity constraints in the factor computational process. Briefly, it is a weighted least squares fit, with weights based on the known standard uncertainties of the element concentrations in the data matrix. The factor model PMF can be written as:

$$X = G \cdot F + E \quad (\text{B9.1})$$

where X is the known n by m matrix of the m -measured chemical species in n samples. G is an n by p matrix of source contributions to the samples (time variations of factors scores). F is a p by m matrix of source compositions (source profiles). G and F are factor matrices to be determined, and E is defined as a residual matrix i.e. the difference between the measurement X and the model $Y = G \cdot F$ as a function of G and F .

Two common programs solve the PMF problem described above: PMF2 (Paatero, 2010) and the multilinear engine (ME) platform (Paatero, 1999) that is used in the EPA PMF v3 tool.

It is well known that factor analysis can give a number of possible solutions, all mathematically correct. The choice of the best solution in PMF analysis, e.g. the **number of factors** that best represent the real case under study, shall be supported by quantitative indicators (Hopke, 2000; Reff et al., 2007).

Examining the Q-value

The Q-value is a goodness of fit parameter, the evaluation of which may give useful indications when the data-point uncertainties are well determined.

The theoretical Q-value is approximately equal to the number of degrees of freedom or to the total number of good data points in the input data array minus the total number of fitted factor elements. If the errors are properly estimated, it can be seen that fitting each good (not weak) data point in such a way that the fitted value falls within the estimated error value, contributes a value of approximately 1 to the Q-value. Contributions resulting from fitting downweighted weak values are usually so small that they may be ignored.

The theoretical Q-value can be approximated by the user as $nm - p(n+m)$, where n is the number of species, m is the number of samples in the dataset, and p is the number of factors fitted by the model (Paatero and Hopke, 2009).

It is useful to look at the changes in the Q-value as additional factors are calculated. After an appropriate number of factors are included in the fit, additional factors will not result in further significant improvements in the Q-value.

It should be noted that the absolute level of Q-values depends strongly on the assumed uncertainties. Usually, it is not recommended to change uncertainties just to get closer to the theoretical Q-value (Brown and Hafner, 2005). If uncertainties have been adjusted so as to produce a reasonable Q-value, then the Q-value can no longer be considered a goodness of fit indicator (Paatero, 2010). However, the differences of Q-values obtained with different numbers of factors are useful indicators even with adjusted uncertainties. If introducing another factor lowers the Q-value only by the number of additional factor elements, then the introduced factor should be rejected.

Useful information can be retrieved by comparing the theoretical Q-value to Q(true) and Q(robust) values, which are output by each run of the EPA-PMF. Q(robust) is calculated by excluding outliers and the Q(true) includes all points. Solutions where Q(true) is 1.5 times greater than Q(robust) may indicate that the model is inconsistently modelling the data. Outliers may be causing this, and can be downweighted by the user so that they have less influence in the model (Brown and Hafner, 2005; Paatero, 2010). Weak variables (i.e. species with low S/N values as defined in paragraph B6) may also be downweighted.

A good fit of the data is characterised by values for Q(robust) and Q(true) that are near to the theoretical Q-value calculated by the user (Brown and Hafner, 2005).

Examining the scaled residuals

The scaled residual is the ratio of the PMF-modelled residual e_{ij} to the input uncertainty σ_{ij} :

$$\frac{e_{ij}}{\sigma_{ij}} = \frac{x_{ij} - \sum_{k=1}^p g_{ik} f_{kj}}{\sigma_{ij}} \quad (\text{B9.2})$$

In PMF analysis, plotting the scaled residuals is also useful in choosing the final number of factors. These residuals should be symmetrically distributed within a range of -3 to +3 (and preferably less). If the scaled residuals are especially large ($\ll -3$ or $\gg +3$) for certain variables, then one may consider that perhaps the uncertainties specified for these variables are too small. If the scaled residuals are especially small (close to zero) for one variable, then either overly large uncertainties have been specified or this variable is explained by a unique factor. It may be acceptable to have a unique factor for a specific variable, but it must make physical/chemical sense for the problem under consideration. A spurious unique factor may arise if uncertainties that are too small are specified for a species. Too many very narrow distributions suggest the presence of too many factors such that the solution is fitting the errors rather than the concentration values. A strong skewness in the scaled residual plots suggests that the fit is not correct and that other solutions should be sought.

Examining the regression parameters

If in the original dataset there is a good mass closure (i.e. the sum of the mass of the single chemical components is close to the gravimetric mass), the “external mass” method - i.e. where the PM mass is not included in the data array analysed by PMF - can be applied. In this case, the measured mass is regressed against the estimated source contribution values. If the

regression produces negative parameters, then too many factors have been included in the solution (Kim et al., 2003), or a strong source does not emit any of the measured species and hence is not represented in any factor but only in PM mass.

The regression parameters can be also used to obtain the scaled source/factor profiles. Once the source profiles are scaled, they can be summed and it can be determined whether the sum of a source/factor profile exceeds 100% (within a 20% tolerance level to account for errors). If this is the case, too few factors may have been chosen (Kim et al., 2003; Hopke, pers. comm.).

Examining the species/mass reconstruction

The appropriateness of the chosen solution can be also assessed by looking at the mass/species reconstruction, which should improve when approaching the best solution.

In the EPA-PMF, there is a regression analysis of the variable with its reconstructed values that provides some measure of the fit to the measurements. However, these regressions are unweighted and, thus, values that are below the detection limit or are missing have a large influence on the results and can produce degraded r^2 values (see chapter B5). To overcome this issue, regressions with weighted values should be calculated manually.

Examining the IM and IS parameters

The maximum individual column mean (IM) and the maximum individual column standard deviation (IS) parameters can be also used to identify the number of factors in a PMF. When the number of factors increases to a critical value, the IM and IS values will drop dramatically (Lee et al., 1999).

Examining multiple solutions

It is essential to perform the PMF analysis several times (typically 20) to be certain that the same solution is obtained. A test for the best selection of the number of factors is that one does not obtain multiple solutions or obtains at most one alternative solution. With greater or fewer factors than the optimum, multiple solutions are more often obtained.

In general, any bilinear factor analysis has **rotational ambiguity**. In other words, there is no unique solution even though there is a global minimum in the ‘least squares’ fitting process. The addition of constraints can reduce the rotational freedom in the system, but non-negativity alone does not generally result in a unique

solution. One of the key features of PMF is that the rotations are part of the fitting process and are not applied after the extraction of the factors, as is done in eigenvector-based methods.

Controlling rotations by the FPEAK value

FPEAK is a parameter used to explore the rotational ambiguity of a PMF solution *a posteriori*. Assigning positive or negative FPEAK values produces rotations of which the suitability is assessed by observing the changes of the Q-value and the G and F factors. The mathematically optimum solution in PMF is FPEAK=0.0. Therefore, in the absence of any other consideration such as G-space plots (see below), and unless there is a substantial improvement in the interpretability of the profiles, the best fit is given by FPEAK = 0.0.

Controlling rotations by imposing external information

PMF rotations can also be controlled by imposing external information on the solution.

Fkey and Gkey constraints consist of binding individual elements of the F and G matrices, respectively, to zero. If specific values of profiles or time series are known to be zero, then it is possible to force the solution toward zero for those values through appropriate settings of the rotational tools Fkey and Gkey values in PMF2. Controls can be applied through modifying the script in Multilinear Engine-2 (ME-2) applications (see chapter C4). Additional flexibility in applying external constraints will be available in future releases of the EPA PMF.

Controlling rotations by examining G-space plots

G-space plots are source contribution scatter plots for pairs of factors (Paatero et al., 2005). When factors are plotted in this way, unrealistic rotations appear as oblique edges that define correlation between the factors. Edges are well-defined straight borders between regions that are densely populated with points and regions where no points occur. With a correct rotation, the limiting edges usually coincide with, or are parallel to, the axes.

Inspection of the plots helps choose a realistic rotation, but one must bear in mind the fact that specific physical situations might occur where there is an oblique edge even though a realistic rotation has been achieved. As an example, correlations could be induced by collocation of the sources coupled with meteorology (i.e. emissions from independent sources may have a certain degree of correlation due to the influence of meteorology).

It must be also emphasised that the presence of aligned edges in G-plots of factors does not necessarily guarantee that a unique rotation has been found. Although uniqueness will probably be the case in most practical situations, there may be special cases in which the rotational uniqueness does not hold.

References:

- Brown S.G, Hafner H.R., 2005. Multivariate Receptor Modeling Workbook. <http://www.sonomatechdata.com/>
- Hopke P.K., 2000. A guide to Positive Matrix Factorization, <http://www.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf>
- Kim E., Larson T.V., Hopke P.K., Slaughter C., Sheppard L.E., Claiborn C., 2003. Source identification of PM_{2.5} in an arid Northwest U.S. city by positive matrix factorization. *Atmospheric Research* 66, 291-305.
- Lee et al., 1999. Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, *Atmospheric Environment* 33, 3201-3212
- Paatero P., Tapper U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111-126.
- Paatero P., 1999. The multilinear engine - A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Graphical Statistics* 8, 854-888.
- Paatero P., Hopke, P.K., Begum, B.A., Biswas, S.K., 2005. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution, *Atmospheric Environment* 39: 193-201
- Paatero P. and Hopke P.K., 2009. Rotational tools for factor analytic models. *J. Chemometrics* 23, 91-100
- Paatero P., 2010. User's guide for Potive Matrix Factorization programs PMF2 and PMF3, Part 1: tutorial. Available @ <http://www.helsinki.fi/~paatero/PMF/pmf2.zip>
- Reff, A., Eberly, S.I., Bhave, P.V., 2007. Receptor modelling of ambient particulate matter data using Positive Matrix Factorization: review of existing methods. *Journal of Air & Waste Management Association* 57, 146-154.

B10. FACTOR ANALYSIS II: EVALUATION OF SOURCE CONTRIBUTION ESTIMATION AND MODEL PERFORMANCE INDICATORS

Principal Component Analysis – Multilinear Regression

Different techniques exist to carry out source contribution estimations by performing multi-linear regression of the principal components versus the total PM mass: APCS (Absolute Principal Component Scores; Thurston and Spengler, 1985), APCA (Absolute Principal Component Analysis; e.g. Swietlicki and Krejci, 1996), and PCA-MLR (Principal Component Analysis – Multilinear Regression; e.g. Tauler et al., 2008).

In the following, these techniques are referred to as APCA.

This analysis may be carried out using numerous statistical software packages, many of them freely available. The computation of source contributions with APCA is characterised by:

- no specific software required
- fast source identification

- relatively time-consuming source contribution estimation.

However, this analysis suffers from three strong limitations:

1. Given that non-negativity constraints are not included in APCA, negative regression coefficients may be obtained. As a result, the output could show negative source contributions (in terms of mass). Two different approaches are generally used to solve this issue: including the resulting negative mass concentrations in the final result of APCA, even though this has no physical meaning, or eliminating the negative values by replacing them with zero or an empty cell. Evidently, the results obtained after the application of one or the other approach may vary largely. Thus, in the absence of consensus regarding the issue of negative regression coefficients, APCA solutions may be prone to high subjectivity and lack of comparability.

APCA		PMF		CMB	
Sources	% PM ₁₀	Sources	% PM ₁₀	Sources	% PM ₁₀
Individual sources					
Clay	31	Clay	16	Clay	41
Industrial#1	15	Industrial	16	Industrial#1	4
Industrial#2	2			Industrial#2	2
Vehicular	10	Vehicular	10	Vehicular	13
Regional+marine	34	Regional+marine	23	Regional	18
		Regional SO ₄ ²⁻	25	Marine	3
				Soil	12
Undetermined	8	Undetermined	10	Undetermined	7
Grouped sources					
Mining&Industry	48	Mining&Industry	32	Mining&Industry	47
Vehicular	10	Vehicular	10	Vehicular	13
Regional	34	Regional	48	Regional	33

Table B10.1. Sources and source contributions obtained during a receptor model intercomparison (Viana et al., 2008).

Source contributions are divided into the sources obtained directly by the receptor models, and grouped into three main source categories to facilitate the intercomparison of results.

Table B10.2. Number of solutions obtained during a receptor model intercomparison (Viana et al., 2008), and average absolute errors (AAEs) for each of the solutions computed using APCA, PMF and CMB.

Number of factors	APCA		PMF		CMB	
	AAE (%)	AAE _{max} (%)	AAE (%)	AAE _{max} (%)	AAE (%)	AAE _{max} (%)
3	18	77	19	89	67	85
4	NS	NS	17	81	13	34
5	14	69	16	73	11	34
6	NS	NS	19	69	11-36*	25-68*
7	NS	NS	16	69	11	20
8	NS	NS	15	68	NS	NS

* range depending on which source profiles are excluded.

2. The number of sources resolved by APCA is generally lower than that obtained with other models (e.g. PMF or CMB), and quantification of source contributions may not be as precise (Table B10.1).

3. APCA has lower flexibility with respect to PMF and CMB, with fewer valid solutions being produced (Table B10.2).

Several model performance indicators are available when applying APCA. These are relatively simple to use in order to assess:

- Correlation between modelled and measured PM mass.
- Chemical mass closure: sum of the estimated source contributions, and comparison with total measured PM mass.
- Average absolute error (AAE): the average of the absolute percentage differences between the estimated and experimental PM mass data, when different numbers of sources or factors are considered (Chow et al., 2007, Table B10.2).
- Correlation between modelled and measured known sources: the most commonly used source for this test is the marine source, calculated as the sum of the chemically determined Na and Cl in study areas with no other major sources of these elements.

In addition, model performance may also be tested by:

- Understanding the data and the solution: Does it make sense? Is it reasonable?
- Combination with other tools (e.g. back-trajectory analysis)
- Combination with other models (PMF to refine the quantification of source contributions and potentially obtain uncertainty estimates)

- Correlation with EU guideline methods (SEC(2011) 208) for natural aerosol sources (e.g. African dust, sea-spray).

In conclusion:

- APCA is an exploratory receptor modelling tool for urban air quality management, i.e. for the design of air pollution mitigation strategies, because of:
 - Fast source identification
 - The size of datasets required could potentially be available from air quality monitoring networks, from at least one selected station/network
 - Easy to interpret model performance indicators, e.g. average absolute errors (AAE)
 - However, the quantification of source contributions is rather inaccurate, and source contributions (in terms of mass) are subject to uncertainty due to the absence of the non-negativity constraint.
- In scientific applications, APCA should mainly be used to obtain a preliminary picture of the possible contribution sources, as a preparatory step for the use of more advanced models (PMF, CMB, etc.).

PMF (Positive Matrix Factorization)

In PMF, Q values indicate how well the model fits the input data. Q(robust) is calculated by excluding outliers while Q(true) is calculated including all data points. The expected (theoretical) Q is $nm - p(n+m)$, where n is the number of species, m is the number of samples in the dataset, and p is the number of factors fitted by the model (see the EPA PMF v3 User Guide, Norris et al., 2008).

An alternative estimation distinguishes weak from good species as follows (Brown and Hafner, 2005):

$$Q = (\# \text{ samples} * \# \text{ good species}) + [(\# \text{ samples} * \# \text{ weak species})/3] - (\# \text{ samples} * \# \text{ factors being estimated})$$

In addition, a number of diagnostic tests are embedded in the EPA PMF v3 software to evaluate the runs: residual analysis, observed vs predicted scatter plot and time series, combined plots of profiles and contributions and box plots to summarise the distribution of the contributions, G-space plots (or G-plots) and factor pie charts. All of these plots are described in the EPA PMF v3 User Guide (Norris et al., 2008).

A number of elements can contribute to the uncertainty in the solutions modelled by PMF, including temporal variation of particulate matter (PM) source profiles, measurement error, sampling variability, and intrinsic limitations in the modelling process, such as rotational ambiguity and incorrectly specified number of factors (see section B9).

In PMF2, it is possible to estimate uncertainties in the F and G matrices (eq. 9.1) using the process originally described by Roscoe and Hopke (1981) and described in detail by Malinowski (1991). The errors in the elements of one matrix are estimated based on the errors in the ambient concentration values, assuming that the other matrix is error-free. Each matrix (F or G) is treated similarly in such a way that an uncertainty value is associated with each element of the matrix.

The standard deviation of the source contribution estimates (SCE) of every factor in all the samples can be used as an estimation of the uncertainty of the average SCEs.

Bootstrapping (available in EPA PMF v3) can be used to determine the precision of PMF profiles by calculating the standard deviation (assuming normality) or various percentiles of factor profiles (F-matrix values) from numerous bootstrap runs. Nevertheless, to obtain a better representation of the component of uncertainty associated with rotational ambiguity, an improved error estimation scheme has been proposed by Paatero et al. (2013) and will be available in the new release of the EPA-PMF(version 5). The new scheme combines bootstrapping and a “displacement” technique based on the controlled perturbation of factor elements.

References:

Brown S.G, Hafner H.R., 2005. Multivariate Receptor Modeling Workbook. <http://www.sonomatechdata.com/>

Chow, J.C.; Watson, J.G.; Chen, L.-W.A.; Chang, M.-C.O.; Robinson, N.F.; Trimble, D.L.; and Kohl, S.D., 2007. The IMPROVE_A temperature protocol for thermal/ optical carbon analysis: Maintaining consistency with a long-term data base. *Journal of Air & Waste Management Association* 57:1014-1023.

Malinowski E.R., 1991. Factor Analysis in Chemistry, Wiley, New York, 2nd Ed.

Norris G., Vedantham R., Wade K., Brown S., Prouty J., Foley C., 2008. EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. EPA 600/R-08/108.

Roscoe B.A., Hopke P.K., 1981. Error Estimation of Factor Loadings and Scores Obtained with Target Transformation Factor Analysis. *Analytica Chimica Acta* 132:89-97.

Paatero P., Eberly S., Brown S.G., Norris, G.A. 2013. Methods for estimating uncertainty in factor analytic solutions. *Atmospheric Measurement Techniques Discussions*, 6, 7593-7631.

SEC(2011) 2008. COMMISSION STAFF WORKING PAPER. Establishing guidelines for demonstration and subtraction of exceedances attributable to natural sources under the Directive 2008/50/EC on ambient air quality and cleaner air for Europe.

Swietlicki, E., Krejci, R., 1996. Source characterisation of the Central European atmospheric aerosol using multivariate statistical methods. Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms 109-110, 519-525.

Tauler, R., Paatero, P., Henry, R.C., Spiegelman, C., Park, E.S., Poirot, R.L., Viana, M., Querol, X., Hopke, P.K., 2008. Identification, Resolution and Apportionment of Contamination Sources, vol. 3, pp. 269-284.

Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmospheric Environment* 19, 9-25.

Viana, M., Pandolfi, M., Minguillón, M. C., Querol, X., Alastuey, A., Monfort, E., Celades, I., 2008. Inter-comparison of receptor models for PM source apportionment: Case study in an industrial area. *Atmospheric Environment* 42, 3820-3832.

B11. FACTOR ANALYSIS III: CRITERIA FOR FACTOR ASSIGNMENT

The most subjective and least quantifiable step in applying PMF for source apportionment is the assignment of identities to the factors chosen as the final solution. It is important for the data analyst to know what types of sources are present in the study area. However, even in cases where there are good emission inventories, there can be situations where a source cannot be identified (Hwang and Hopke, 2006). In addition, atmospheric processes may result in multiple factors such as summer and winter secondary sulphate, or in producing sufficiently collinear sources that an irresolvable mixture of source profiles is obtained. Thus, profiles have to be interpreted with both knowledge of the study area and a background in atmospheric science. For that reason, any choice concerning the correspondence between source categories and factors must be supported by objective and quantitative tests.

High shares of a source marker in a factor profile may be used for a preliminary source attribution. However, further evidence is required for confirmation of this initial hypothesis.

Proposed steps to support factor assignment:

- Compare the obtained factor profiles with those reported in previously published PMF studies (the comparison can be performed either visually or numerically using, for instance, the Pearson coefficient);
- Search the literature for measured PM source profiles with characteristics similar to the factor profiles in the F-matrix;
- Search for measured PM source profiles in relevant databases (e.g. SPECIATE);
- Identify the source by comparing certain species ratios (also referred to as “enrichment factors”) in PMF source/factor profiles to the same ratios in measured PM source profiles (see also section B12);
- Perform local and/or regional source sampling along with the ambient PM sampling to develop source profiles needed to identify PMF profiles;
- Look at temporal patterns for “expected” behaviours (e.g. the largest contributions of a source believed to be residential wood burning should likely occur during winter months); plots of contributions over time can be inspected in order to look for daily, weekly, seasonal, and yearly oscillations of source contributions. Mean source contributions by season and by day of the week (weekend versus weekday) should also be examined (see also section B12).

It should be noted that when source profiles are not independent (i.e. there is severe collinearity) it is difficult to separate their contributions. In this case, additional chemical/physical information is needed to improve source segregation. Nevertheless, sources can clearly be separated for a sufficiently low level of collinearity and precision in the input data. In spectrochemical problems, good factors can be obtained despite quite severe collinearity. However, the collinearity inflates the uncertainties of the values (Cheng et al., 1988).

Advanced User Box

Auxiliary analyses can be used to aid in the identification of PMF factors: e.g. contribution of wind roses, conditional probability function, potential source contribution function, cluster analysis, and residence time analysis are some techniques for analysing wind or backward trajectories (see section C1).

References:

Cheng, M. -D, Hopke P. K., Jennings D. E., 1988. The Effects of Measurement Errors, Collinearity and their Interactions on Aerosol Source Apportionment Computations. *Chemometrics and Intelligent Laboratory Systems* 4 (3), 239-250.

Hwang I.J. Hopke P.K., 2006. Comparison of Source Apportionments of PM_{2.5} at Two San Jose Speciation Trends Network Sites. *Journal of Air & Waste Management Association* 56: 1287-1300

B12. TESTS FOR MODEL PERFORMANCE VALIDATION

The fundamental, natural physical constraints that must be fulfilled in any source apportionment study are as follows (Hopke, 2010):

- The original data must be reproduced by the model; the model must explain the observations;
- The predicted source compositions must be non-negative; a source cannot have a negative elemental concentration (slightly negative values are acceptable provided zero is in the confidence interval);
- The predicted source contributions to the aerosol must all be non-negative; a source cannot emit negative mass;
- The sum of the predicted elemental mass contributions for each source must be less than or equal to the total measured mass for each element; the whole is greater than or equal to the sum of its parts.

The assignment of a source factor to a source type (or source category) is a critical step in factor analysis. Therefore, it is important to carry out sensitivity tests that assess the variability of the results because of different combinations of sources and/or species in the model (Watson et al., 2008). Several diagnostics are available to evaluate the receptor model results.

Advanced User Box

Actually, ME 2 allows a certain degree of negativity in the source/factor contributions for the sake of better rotational uniqueness (Norris et al., 2009).

Ratios

Unique source tracers are rare, therefore elemental and/or molecular ratios have often been used to trace similar sources, such as combustion processes or mineral sources, for example. In factor analysis techniques, the resolved

factor profiles are often evaluated by comparing relative amounts of elements/compounds with those expected in relevant sources (Galarneau, 2008). Robinson et al. (2006a, b and c) demonstrated that the ratio of marker species in a source profile, when compared with those from the same and/or different source types and from ambient samples, helps to interpret the source variability and identify the most important sources in a region. However, one should bear in mind that the two assumptions of unique ratios among sources and conservative ratios in the atmosphere are not always met in reality. Also, the species examined should have similar reactivities with respect to atmospheric oxidants and solar radiation and similar particle size distributions in order to exclude differences in particle scavenging by precipitation or particle dry deposition (Galarneau, 2008).

One of the first uses of the elemental ratio was proposed by Juntto and Paatero (1994) who compared the Na/Cl ratio in PMF factors with sea-water composition. Liu et al. (2003) showed that their long-range transported dust profiles correlated well with standard reference Chinese desert dust, with the exception of enrichment in sulphate. Hien et al. (2004) used several ratios to distinguish between Local Burning and Long-Range transport aerosols. Hien et al. (2005) used different Ca/Si ratios to separate coal fly ash from soil dust. Lanz et al. (2007) calculated ratios of the modelled primary organic aerosols (POA) and measured primary pollutants such as elemental carbon (EC), nitrogen oxides (NO_x), and carbon monoxide (CO), finding good agreement with literature values. Organic and inorganic ratio evolutions have been also examined as a function of photochemical age of aerosols (DeCarlo et al., 2010).

Residuals

The distribution of residuals (the percentage of all scaled residuals in a given bin, 0.5 for example) should be investigated in order to verify how well the model fits each species. If a species has many large-scaled residuals or displays a non-normal curve, it may be an

indication of a poor fit. A well-modelled species instead shows normally distributed residuals within the range +3 and -3.

In weighted 'least squares' analysis, the distribution of residuals can vary substantially with the different values of the variables (species). Therefore, weighted residuals (Graybill and Iyer, 1994) must be used in graphical residual analysis, so that the plots can be interpreted as usual. This must be taken into account when evaluating EPA PMF v3 default unweighted residual graphs.

The scale of the histogram chart (y-axis) is important. Setting the maximum values as the maximum value of each species is helpful when examining individual species and the shape of their distributions. If the Y-axis maximum is fixed at 100%, a comparison between species can easily be made. Species with residuals beyond +3 and -3 need to be further evaluated by comparing the observed vs modelled concentrations by means of scatter plots and/or time series. Large positive scaled residuals may indicate that the model is not fitting the species or that the species is present in an infrequent source. Species that do not have a strong correlation between observed and modelled values or have poorly modelled peak values should be evaluated by the user to determine if they should be downweighted or excluded from the model.

Other useful statistics when comparing observed vs modelled values are the coefficient of determination (r^2), intercept, slope, and normal residual (EPA PMF v3 User Guide; Norris et al., 2008).

Advanced User Box

Residuals can also be compared between different runs of one model (different starting points). The sum of squared difference between residuals of a pair of runs can be used (as in EPA PMF v3) as a diagnostic of different solutions (rather than mere rotations of the same solution).

In EPA PMF v5.0, it is possible to examine the ratio of observed to expected Q-values using Q/Qexp charts. This is an efficient way to understand the residuals of the PMF solution, and in particular, what samples and/or species were not well modelled (i.e. have values greater than 2).

The Kolmogorov-Smirnov test can be used to determine whether the residuals are normally distributed. If the test indicates that the residuals are not normally distributed, the user should visually inspect the histogram for outlying residuals. A very narrow (leptokurtic) distribution of residuals suggests that species are fitted too well and may be an indicator of "ghost factors", which can explain most of the variation of one species (Amato and Hopke, 2012).

Time trends

Source strengths are often time-dependent due to the influence of atmospheric processes (nucleation, volatilisation, transport, etc.), meteorological parameters (solar radiation, humidity, precipitation, etc.), and variation in human activity (intra-day, day-to-day). As a result, the source contributions will also change over time, and this variation is a suitable diagnostic for evaluating interpretations of factor profiles.

Some programs such as EPA PMF v3 already implement tools for a quick check of the seasonal and weekday/weekend variation of factor contributions. However, the user can further explore their time variability in relation to concentrations of gaseous pollutants such as SO₂, CO and NO_x for combustion sources (Zhou et al., 2005; Yue et al., 2008; Brown et al., 2012), O_x (O₃+NO₂) for secondary sources (Huang et al., 2010), and NH₃ for agricultural activities (Eatough et al., 2010). In some cases, factor analysis can couple different pollutant categories in a unique dataset; for example, Pey et al. (2009) combined the size distribution of aerosols, meteorological parameters, gaseous pollutants and chemical speciation of PM_{2.5} to carry out a PCA analysis.

A posteriori wind direction analysis

A simple but reliable method is to plot source contributions in a polar scatter plot in such a way that wind direction determines the angle and source contribution determines the radius of each plotted point. Such a plot shows at a glance the general behaviour of wind-directional dependence. Also, an overview of the individual points is helpful, as one or two high-concentration points cannot distort the picture, as discussed above. Additional information, such as winter/summer classification, may be indicated by using different colours when plotting winter and summer source contribution points. See section C1 for a more detailed discussion of wind direction analysis techniques.

Overall uncertainty

The output from source apportionment (SA) consists of source contribution estimates (SCEs) with a definite uncertainty. Special efforts must be taken by the SA scientist to analyse and communicate this uncertainty. Most receptor models compute the uncertainty of the output. However, in cases where results derive from more than one SA technique, the computation of the uncertainty of the combined SCEs is not straightforward. Larsen et al. (2012) have recently demonstrated how probabilistic uncertainty characterisation by Monte Carlo simulations yielding probability distributions can be used to combine results deriving from CMB, PMF, and emission factor analysis. The advantage of this approach is that it generates the uncertainty of the combined SCEs as well as essential data for sensitivity analysis.

Advanced User Box

RMs output uncertainty derives from both inaccuracy in the input data and model assumptions and ambiguities (Karagulian and Belis, 2012). Monte-Carlo probabilistic methods (such as bootstrapping) are suitable for estimating the random component of the output uncertainty in factor analytical methods. On the other hand, according to a study on the error estimation methods implemented in EPA PMF v.5 (Paatero et al., 2013), analysis of controlled perturbations of the F and G matrices elements (displacements) is most appropriate for estimating the rotational uncertainty (which is a non-random component).

If the number of factors is large, as is typical when analysing speciated data, rotational uncertainty is often the leading cause of uncertainty in results, and relying on Monte-Carlo methods may produce error intervals that are much too narrow. On the other hand, if there is only a small number of rotation-limiting zero values in true time series (G) factors, customary bootstrapping may also lead to uncertainties that are much too large. Whenever the resampling process happens to eliminate such zero values, rotational uncertainty may increase dramatically, and bootstrapped results may deviate dramatically from the original full-data results.

The new methodology is promising especially when used in combination with Monte Carlo tests. Nevertheless, more experience is needed on its application to real-world datasets.

References

- Amato, F. and Hopke, P.K., 2012. Source Apportionment of the Ambient PM_{2.5} in East St. Louis using Constrained Positive Matrix Factorization. *Atmospheric Environment* 46, 329-337
- Brown, S.G., Lee, T., Norris, G.A., Roberts, P.T., Collett, J.L., Paatero, P., Worsnop, D.R., 2012. Receptor modeling of near-roadway aerosol mass spectrometer data in Las Vegas, Nevada, with EPA PMF. *Atmospheric Chemistry and Physics* 12, 309-325
- Decarlo, P.F., Ulbrich, I.M., Crouse, J., De Foy, B., Dunlea, E.J., Aiken, A.C., Knapp, D., (...), Jimenez, J.L., 2010. Investigation of the sources and processing of organic aerosol over the Central Mexican Plateau from aircraft measurements during MILAGRO. *Atmospheric Chemistry and Physics* 10 (12), 5257-5280.
- Eatough, D.J., Kuprov, R., Hansen, J.C., Olsen, N., 2010. PMF apportionment of primary and secondary fine particulate material in the Salt Lake Valley: Winter 2009. Proceedings of the Air and Waste Management Association's Annual Conference and Exhibition, AWMA 8, 6258-6263.
- Galarnau, E., 2008. Source specificity and atmospheric processing of airborne PAHs: Implications for source apportionment. *Atmospheric Environment* 42, 8139-8149
- Graybill, F.A., and H.K. Iyer, 1994. Regression Analysis: Concepts and Applications. Duxbury Press, Belmont, California.
- Hien, P.D., Bac, V.T., Thinh, N.T.H., 2004. PMF receptor modelling of fine and coarse PM₁₀ in air masses governing monsoon conditions in Hanoi, northern Vietnam. *Atmospheric Environment* 38 (2), 189-201.
- Hien, P.D., Bac, V.T., Thinh, N.T.H., 2005. Investigation of sulfate and nitrate formation on mineral dust particles by receptor modelling. *Atmospheric Environment* 39 (38), 7231-7239.
- Hopke, P. K., 2010. The application of receptor modeling to air quality data. *Pollution Atmospheric special issue* 91-109
- Huang, X., Zhao, Q., He, L., Hu, M., Bian, Q., Xue, L., Zhang, Y., 2010. Identification of secondary organic aerosols based on aerosol mass spectrometry. *Science China Chemistry* 53 (12), 2593-2599

- Juntto, S. and Paatero, P., 1994. Analysis of daily precipitation data by positive matrix factorization. *Environmetrics* 5 (2), 127-144,
- Karagulian F. and Belis C.A., 2012. Enhancing Source Apportionment with receptor models to Foster the Air Quality Directive Implementation. *International Journal of Environmental Pollution* 50, 190-199.
- Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C., Prevot, A.S.H., 2007. Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra. *Atmospheric Chemistry and Physics* 7 (6), 1503-1522.
- Larsen, B.R., Gilardoni S., Stenström K., Niedzialek J., Jimenez J., Belis C.A., 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. *Atmospheric Environment* 50, 203-213.
- Liu, W., Hopke, P.K., Van Curen, R.A., 2003. Origins of Fine Aerosol Mass in the Western United States Using Positive Matrix Factorization. *Journal of Geophysical Research- Atmospheres* 108 (D23), D04716.
- Norris G., Vedantham R., K. Wade, S. Brown, J. Prouty, C. Foley, 2008. EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. EPA 600/R-08/108.
- Norris G., Vedantham, R., Wade D., Zahn P., Brown S., Paatero P., Eberly S. Foley C., 2009. Guidance Document for PMF Applications with Multilinear Engine. EPA 600/r-09-032.
- Paatero P., Eberly S., Brown S.G., Norris, G.A. 2013. Methods for estimating uncertainty in factor analytic solutions. *Atmospheric Measurement Techniques Discussions*, 6, 7593-7631.
- Pey, J., Querol, X., Alastuey, A., Rodriguez, S., Putaud, J.P., Van Dingenen, R., 2009. Source apportionment of urban fine and ultra-fine particle number concentration in a Western Mediterranean city. *Atmospheric Environment* 43 (29), 4407-4415.
- Robinson, A.L., Subramanian, R., Donahue, N.M., Bernardo-Bricker, A., Rogge, W.F., 2006. Source Apportionment of Molecular Markers and Organic Aerosols. 1. Polycyclic Aromatic Hydrocarbons and Methodology for Data Visualization. *Environmental Science and Technology* 40, 7803-7810.
- Robinson, A.L., Subramanian, R., Donahue, N.M., Bernardo-Bricker, A., Rogge, W.F., 2006. Source Apportionment of Molecular Markers and Organic Aerosol, 2. Biomass Smoke. *Environmental Science and Technology* 40, 7811-7819.
- Robinson, A.L., Subramanian, R., Donahue, N.M., Bernardo-Bricker, A., Rogge, W.F., 2006. Source Apportionment of Molecular Markers and Organic Aerosol. 3. Food Cooking Emissions. *Environmental Science and Technology* 40, 7820-7827.
- Watson, J.G., Chen L.-W.A., Chow J.C., Doraiswamy P., Lowenthal D.H., 2008. Source Apportionment: Findings from the U.S. Supersites Program. *Journal of Air & Waste Management* 58, 265-288.
- Yue, W., Stolzel, M., Cyrus, J., Pitz, M., Heinrich, J., Kreyling, W.G., Wichmann, H.-E., (...), Hopke, P.K., 2008. Source apportionment of ambient fine particle size distribution using positive matrix factorization in Erfurt, Germany. *Science of the Total Environment* 398 (1-3), 133-144.
- Zhou, L., Kim, E., Hopke, P.K., Stanier, C., Pandis, S.N., 2005. Mining airborne particulate size distribution data by positive matrix factorization. *Journal of Geophysical Research D: Atmospheres* 110 (7), art. no. D07S19, 1-15

B13. REPORTING RESULTS AND METHODOLOGY

Due to the large number of variables to be considered, source apportionment (SA) studies are complex. They often require adaptation of existing methods to the specific problem or the development of tailor-made solutions. In addition, there are many steps in which decisions have to be taken by the modeller. Therefore, it is essential to support the final results with an appropriate description of the methodological choices and documentation of the objective qualitative or quantitative information that support expert decisions. In this way, reviewers and final users are provided with the elements to assess the relevance of the study and other modellers get a chance to reproduce the methodology. If the results are reported in a peer-reviewed scientific journal, much detail can be provided as supplementary material that most journals now support.

The present protocol has been conceived as a reference document that cannot substitute for experience and competence. For that reason, documented participation of experts in training and intercomparisons should be promoted in order to develop and demonstrate individual and institutional capacities.

Although this protocol aims at promoting the highest quality standards, it has to deal with the intrinsic limitation of any SA study: the “true” contribution of sources to atmospheric pollution at a given point cannot be directly measured.

SA studies can be considered as being consistent with the present protocol if they comply with the following requisites:

- The results are described according to the steps proposed in sections B1- B12. When limited space is available, these technical notes should be allocated in an annex or, in the case of scientific journals, as supporting information or supplementary material.
- Expert decisions are described and evidence of the objective information (e.g. quantitative tests, sensitivity analysis, external information) that supports them is provided. This point is essential for critical steps such as the selection of source profiles in chemical mass balance modelling, and the identification of the number of sources and factor assignment in factor analysis.
- The documentation includes the references of the source profiles used as input or to validate factor assignment.
- The model and version used are clearly reported and the modifications adopted for the specific case well described.
- The quantitative uncertainty of the output is estimated and reported using the techniques described in the present document or other robust methodologies available in the literature. Sources of uncertainty that cannot be quantified should be acknowledged, bearing in mind that both inaccuracy in the input data and model assumptions and ambiguities contribute to the total uncertainty budget in receptor models.
- Estimation of overall uncertainty and validation is achieved by comparing outputs from independent models on the same dataset and/or using permutation or displacement techniques.
- Sensitivity analysis is performed to demonstrate that there are no substantial deviations from the mass conservation assumption.
- Only solutions that implement the quality assurance steps described in this guide can claim state-of-the-art performance supported by community-wide intercomparison exercises.

PART C: ADVANCED MODELS

C1. WIND AND TRAJECTORY ANALYSIS IN SOURCE APPORTIONMENT

Introduction

Source apportionment results are frequently complemented by procedures to identify the direction of air masses with high pollution levels or where certain compounds of interest come from (table C1.1). For low- to medium-spatial scales this can be done by, for example, wind rose analysis (see section B12). However, medium and long-range transport may be better assessed using backward trajectories calculated with a suitable dispersion model (Stohl, 1998).

Wind Direction Analysis

The potential location of emission sources or the origin of polluted air masses can be investigated *a posteriori*, once the source contributions are already obtained. As a starting point, simple concentration roses (polar plots of sector-averaged wind contributions) can be used. The conditional probability function (CPF; Ashbaugh et al., 1985) is a common tool used to analyse point source impacts from varying wind directions using the source contribution estimates from receptor models coupled with the wind direction values measured on site (Kim et al., 2003). When particulate matter (PM) measurements are performed over 24 hours, the same daily source contribution is assigned to each hour of a given day in order to match to the hourly wind data. The conditional probability function (CPF) estimates the probability that a given source contribution from a given wind direction will exceed a predetermined threshold criterion. It is defined as:

$$CPF_{\Delta\theta} = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \quad (\text{B12.1})$$

where $m_{\Delta\theta}$ is the number of occurrences from wind sector $\Delta\theta$ that exceeded the threshold criterion, and $n_{\Delta\theta}$ is the total amount of data from the same wind sector. Typically, 12 sectors are used ($\Delta\theta = 15$ degrees) and calm wind periods are excluded due to the isotropic behaviour of wind vane under calm winds. The threshold criterion should be chosen based on sensitivity tests

ANALYSIS OF WIND DIRECTION
Conditional probability function (CPF)
Non-parametric wind regression (NWR)
Pseudo deterministic receptor model (PDRM)
ANALYSIS OF BACKWARD TRAJECTORIES
Trajectory sector analysis (TSA)
Potential source contribution function (PSCF)
Simplified quantitative transport bias analysis (SQTBA)
Trajectory mass balance (TRMB) or TRMB regression (TMBR)

Table C1.1. Hybrid trajectory-based receptor models (based on Belis et al., 2013)

with several different percentiles of the fractional contribution from each source. A commonly used threshold is the 75th percentile (e. g. Amato and Hopke, 2012; Jeong et al., 2011; Kim et al., 2004).

The sources are likely to be located based on the wind directions that have high conditional probability values. A large number of papers have been published on the application of these approaches to the receptor modelling problem (Zhao and Hopke, 2006; Kim and Hopke, 2004, among others). However, Zhou et al. (2004) showed that the conditional probability function can provide misleading results when many directions are used with very few (or no) wind occurrences and when the distribution of concentrations is far from normal. The non-parametric regression analysis technique is an alternative that can be used to locate sources. In this technique, the relationship of the contribution and wind direction is determined by kernel regression and confidence intervals are also given (Henry et al., 2002; Henry, 2002). The expected concentration C at θ is computed by:

$$\bar{C}(\theta, \Delta\theta) = \frac{\sum_{i=1}^n K((\theta - W_i) / \Delta\theta) C_i}{\sum_{i=1}^n K((\theta - W_i) / \Delta\theta)} \quad (\text{B12.2})$$

where K is a Gaussian kernel function, W_i and C_i are the wind direction and concentration of the i^{th} sample respectively, and $\Delta\theta$ is the smoothing parameter, the only adjustable parameter in nonparametric regression (Zhou et al., 2004). Kim and Hopke (2004) showed that conditional probability function and nonparametric regression provided very similar results for many cases.

Backward Trajectory Analysis

In a source apportionment study, back trajectories can be used either to pre-select datasets for analysis (e.g. in cases where specific sources and source regions are of major interest) or, as is most frequently the case, to check the plausibility of identified sources/processes and to get information about their geographical distribution and locations.

In addition to the models operated commercially by national weather forecast organisations, there is a variety of research-oriented models available in Europe which allow back-trajectory plots to be produced, e.g. FLEXPART (NILU); REM-CALGRID (TRUMF), EURAD (RIU). However, such trajectories can be obtained only through the research groups or companies operating these models and usually have to be paid for. Therefore, the most widely used tool is the NOAA HYSPLIT model (NOAA Atmospheric Research Laboratory; Draxler, 2012) which is available free of charge to the scientific community. Trajectories can be calculated on demand via a web application (Rolph, 2012)

or locally after downloading the program package. While the first approach allows for rapid feedback, limitations have been imposed to reduce computational activities on the NOAA Atmospheric Research Laboratory servers. Hence, local installation is recommended for routine use.

The HYSPLIT model can process different meteorological file types that may also be downloaded via the program. Global data assimilation system files are the standard meteorological files that can be used in Europe, and have a spatial resolution of 1 degree longitude and latitude. More detailed information can be found on the NOAA Atmospheric Research Laboratory website.

In a basic approach, several trajectories are calculated for each day, varying the time of arrival and height above ground level for backward trajectory periods of usually 3 or 4 days.

To get a more temporally representative picture of the regions associated with, for example, episodes with high PM levels, computation of trajectory data is needed for longer periods (up to several years) and multiple sites. The trajectory cloud obtained can be further processed using statistical methods such as clustering (Stohl et al., 2002) to identify the most relevant types of air mass transport to the sites or areas under consideration. Such multiple-trajectory processing also reduces the uncertainties of single trajectory processing, which increase considerably with greater distances.

Figure C1.1. 96-hour back trajectory calculated with HYSPLIT 4.9 for a day with high mineral dust concentration (left; Draxler, 2012), pointing to arid source regions close to the Caspian Sea (right; Abasova, 2010)

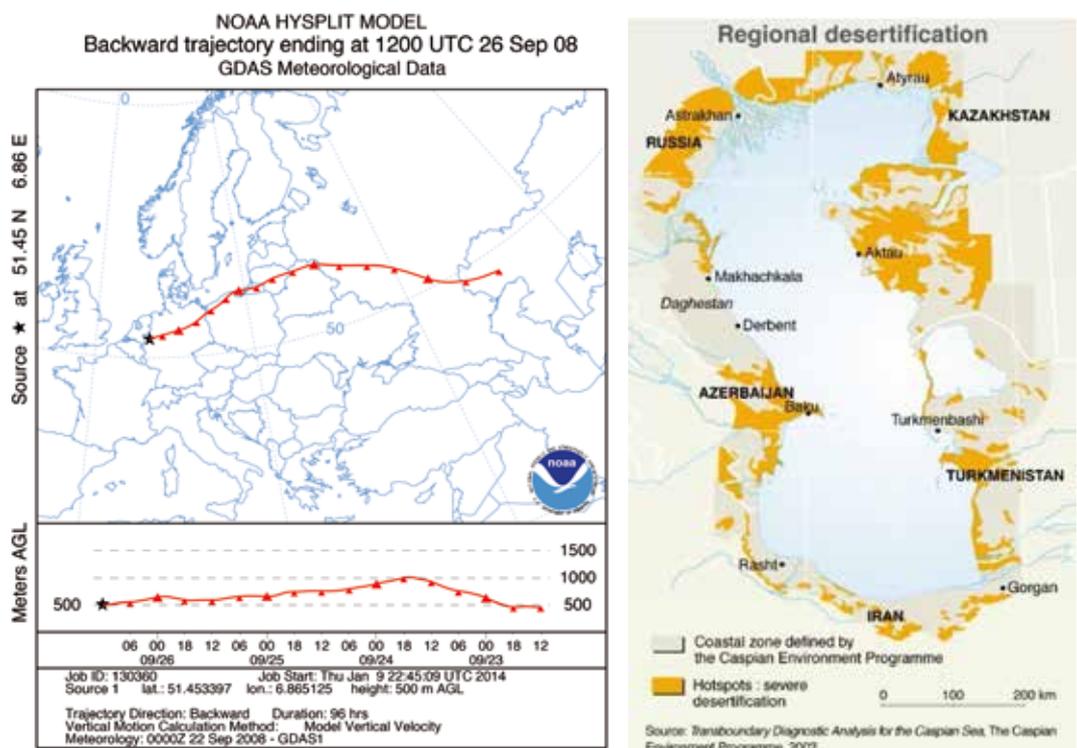
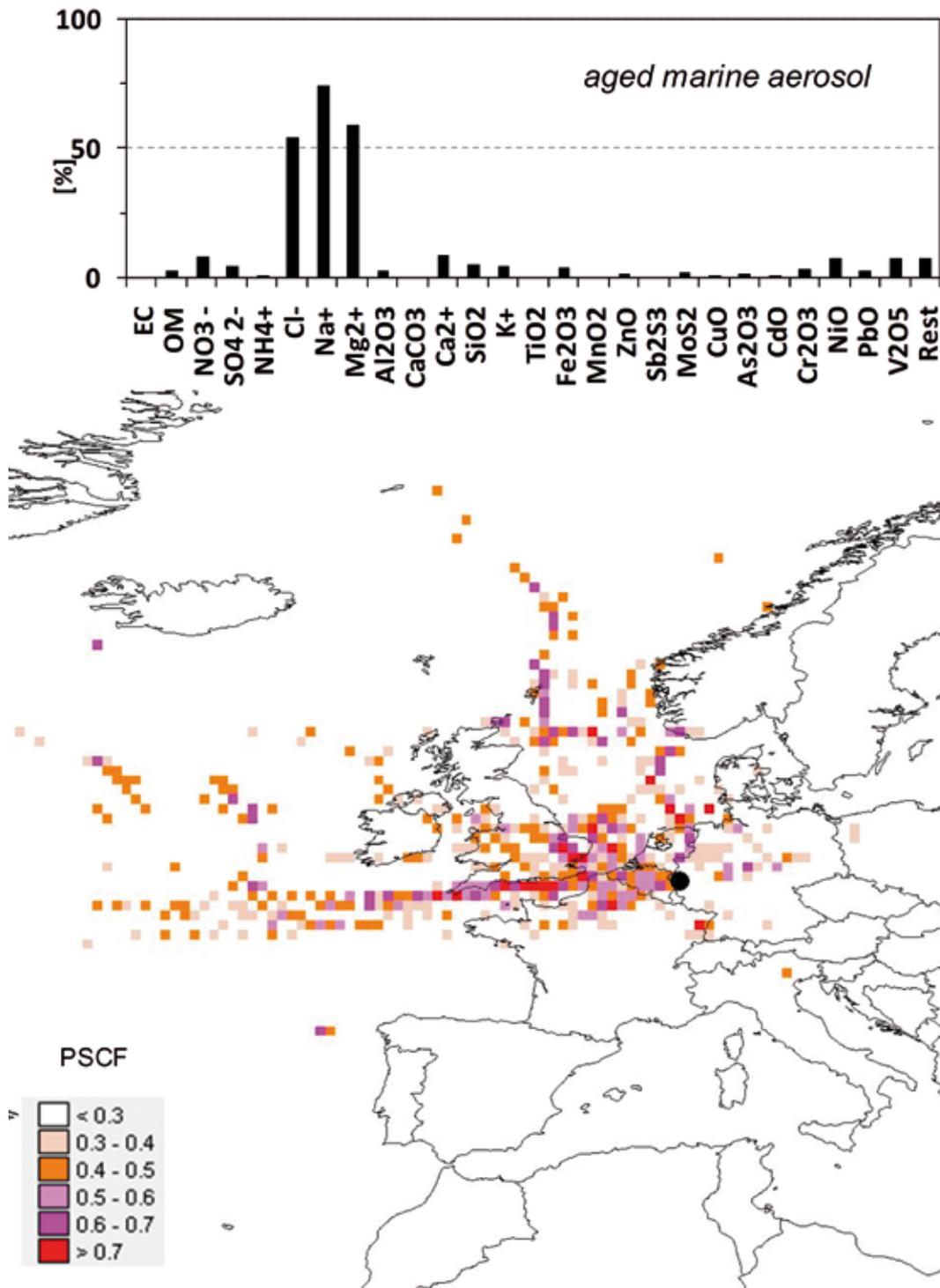


Figure C1.2. PMF source profile denoted as "Aged marine aerosol" and corresponding potential source contribution function plot for factor contributions > 75 percentile (Beuck et al., 2011)



PSCF plot computed with TrajStat software (Wang et al., 2009)

The most common procedures use six-hour increments in arrival times to cover a 24-hour period. However, detailed time resolutions such as one-hour increments are also possible (note: the basic meteorological models are run on a three-hour time resolution). Single-day calculations have proven to be particularly useful in the case of short-term dust events caused, for example, by long-range dust intrusions from arid regions or wildfires. An example is given in figure C1.1, which shows straight air mass advection from Eastern Europe which carried

mineral dust probably from arid regions close to the Caspian Sea (Beuck et al., 2011, Abasova, 2010).

Moreover, advanced evaluation methods exist to apportion PM levels measured at the receptor site to the trajectory segments using analysis of backward trajectories (Table C1.1). An example of an analysis of the potential source contribution function (PSCF) is shown in Figure C1.2.

References:

- Abbasova T., 2010. Detection and analysis of changes in desertification in the Caspian Sea Region. Master Thesis. Stockholm University, Faculty of Science, Department of Physical Geography and Quaternary Geology (INK), Stockholm University. <http://su.diva-portal.org/smash/record.jsf?searchId=1&pid=diva2:354905>; access 10.01.2012
- Amato, F., Hopke, P. K., 2012. Source apportionment of the ambient PM_{2.5} across St. Louis using constrained positive matrix factorization, *Atmospheric Environment*, 46(0), 329–337, doi:10.1016/j.atmosenv.2011.09.062.
- ARL: Air Resources Laboratory, <http://www.arl.noaa.gov/>
- Ashbaugh, L. L., Malm, W. C., Sadeh, W. Z., 1985. A Residence Time Probability Analysis of Sulfur Concentrations at Grand Canyon National Park. *Atmospheric Environment - Part A General Topics* 19 (8),1263-1270.
- Belis C.A., Karagulian, F., Larsen, B.R., Hopke, P. K., 2013. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmospheric Environment* 69, 94-108.
- Beuck, H., Quass, U., Klemm, O., Kuhlbusch, T.A.J., 2011. Assessment of sea salt and mineral dust contributions to PM₁₀ in NW Germany using tracer models and positive matrix factorization. *Atmospheric Environment* 45, 5813-5821
- Draxler, R.R. and Rolph, G.D., 2012. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (<http://ready.arl.noaa.gov/HYSPLIT.php>). NOAA Air Resources Laboratory, Silver Spring, MD.
- Henry, R.C., 2002. Multivariate receptor models—current practice and future trends. *Chemometrics and Intelligent Laboratory Systems* 60, 43–48.
- Henry, R.C., Changa, Y-S., Spiegelman, C.H., 2002. Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. *Atmospheric Environment* 36, 2237–2244. NILU: <http://transport.nilu.no/flexpart>
- Jeong, C. H., McGuire, M. L., Herod, D., Dann, T., Dabek-Zlotorzynska, E., Wang, D., Ding, L., Celo, V., Mathieu, D., Evans, G., 2011. Receptor model based identification of PM_{2.5} sources in Canadian cities, *Atmospheric Pollution Research*, 2(2), 158–171.
- Kim, E., Hopke, P. K., Larson, T. V., Covert, D. S., 2004. Analysis of ambient particle size distributions using unmix and positive matrix factorization, *Environmental science & technology*, 38(1), 202–209.
- Kim, E. and Hopke, P.K., 2004. Comparison between conditional probability function and nonparametric regression for fine particle source directions. *Atmospheric Environment* 38 (28), 4667-4673
- Kim, E., Hopke, P.K., Paatero, P., Edgeton, E.S., 2003. Incorporation of Parametric Factors into Multilinear Receptor Models Studies of Atlanta Aerosol. *Atmospheric Environment* 37, 5009-5021
- RIU: the EUROpean Air Pollution Dispersion (EURAD) Project: http://www.eurad.uni-koeln.de/index_e.html?modell/eurad_descr_e.html
- Rolph, G.D., 2012. Real-time Environmental Applications and Display sYstem (READY) Website (<http://ready.arl.noaa.gov>). NOAA Air Resources Laboratory, Silver Spring, MD.
- Stohl A., 1998. Computation accuracy and applications of trajectories - A review and bibliography. *Atmospheric Environment* 32, 947-966.
- Stohl A., Eckhardt S., Forster C., James P., Spichtinger N., Seibert P., 2002. A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements. *Atmospheric Environment* 36, 4635-4648.
- TRUMF (Troposphärische UmweltForschung): TRAMPER Trajectories: <http://www.geo.fu-berlin.de/en/met/ag/trumf/Trajektorien/index.html>
- Wang Y.Q., Zhang X. Y., Draxler R.R., 2009. TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data. *Environmental Modelling & Software* 24, 938. download of product from <http://www.meteothinker.com/TrajStatProduct.html>
- Zhao, W., Hopke, P.K., 2006. Source investigation for ambient PM 2.5 in Indianapolis, IN. *Aerosol Science and Technology* 40 (10), 898-909.
- Zeng Y., Hopke P. K., 1989. A Study of the Sources of Acid Precipitation in Ontario, Canada. *Atmospheric Environment* 23, 1499-1509.
- Zhou, L., Hopke, P.K., Paatero, P., Ondov, J.M., Pancras, J.P., Pekney, N.J., Davidson C.I., 2004. Advanced factor analysis for multiple time resolution aerosol composition data. *Atmospheric Environment* 38 (29), 4909-4920.

C2. THE USE OF PMF and ME-2 IN AEROSOL MASS SPECTROMETER DATA PROCESSING

The aerosol mass spectrometer (AMS), developed by Aerodyne Research Inc. (ARI), Massachusetts, has been designed to provide real-time quantitative information on size-resolved mass concentrations for volatile and semi-volatile components present in/on ambient aerosol particles (Jayne et al., 2000). The AMS is designed to provide quantitative composition information on ensembles of particles, with limited single particle information. The instrument combines standard vacuum and mass spectrometric techniques with recently developed aerosol sampling techniques. A schematic representation of the AMS is shown in figure C2.1.

The AMS consists of three main parts: an aerosol inlet, a particle sizing chamber, and a particle composition detection section. The different sections are separated by small apertures and differentially pumped. The aerosol inlet samples a flow of $1.5 \text{ cm}^3 \text{ s}^{-1}$ and focuses particles into a narrow beam ($\sim 1 \text{ mm}$ diameter). Size-dependent particle velocities created by expansion into vacuum are used to determine particle size through a particle time-of-flight measurement. Detection is performed by directing the particle beam onto a resistively heated roughened surface under high vacuum ($\sim 10^{-7}$ Torr).

Upon impact, the volatile and semi-volatile components in/on the particles flash vaporise. The vaporisation source is integrally coupled to an electron impact ioniser at the entrance of a quadrupole mass spectrometer. The instrument's electronics are coupled to a computer for real-time instrument control and data acquisition, analysis, and display. Because most molecules undergo extensive fragmentation, the AMS spectra provide information on the bulk organic aerosol with limited molecular detail. The AMS has revolutionised aerosol research concerning atmospheric processes involving aerosols, and provides quantitative information on organic aerosol sources and components at high time resolution without filter sampling issues and extrapolation of small marker concentrations to the bulk (Jimenez et al., 2009). More than 500 research papers using the technique have been published since 2005. In this section, the attention is focused on the organic composition of the aerosol. According to this fraction it is grouped into hydrocarbon-like organic aerosol (often mostly from traffic), oxygenated organic aerosol (mostly secondary organic aerosol), cooking organic aerosol, biomass burning/domestic wood burning aerosol and other components (Table C2.1).

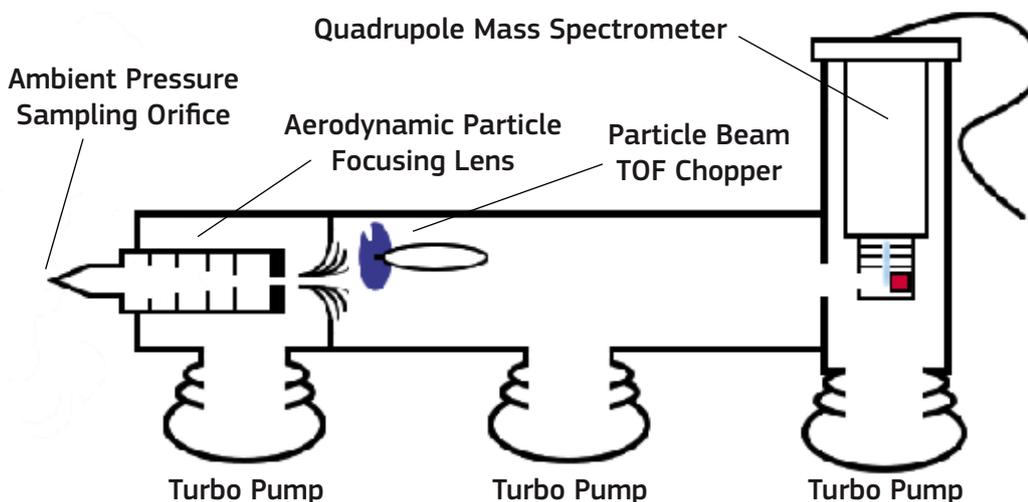


Figure C2.1. Schematic description of the AMS source: <http://cires.colorado.edu/~jjose/ams.html#Aerodyne-AMS>

Table C2.1. Commonly reported factors using AMS (adapted from Zhang et al., 2011)

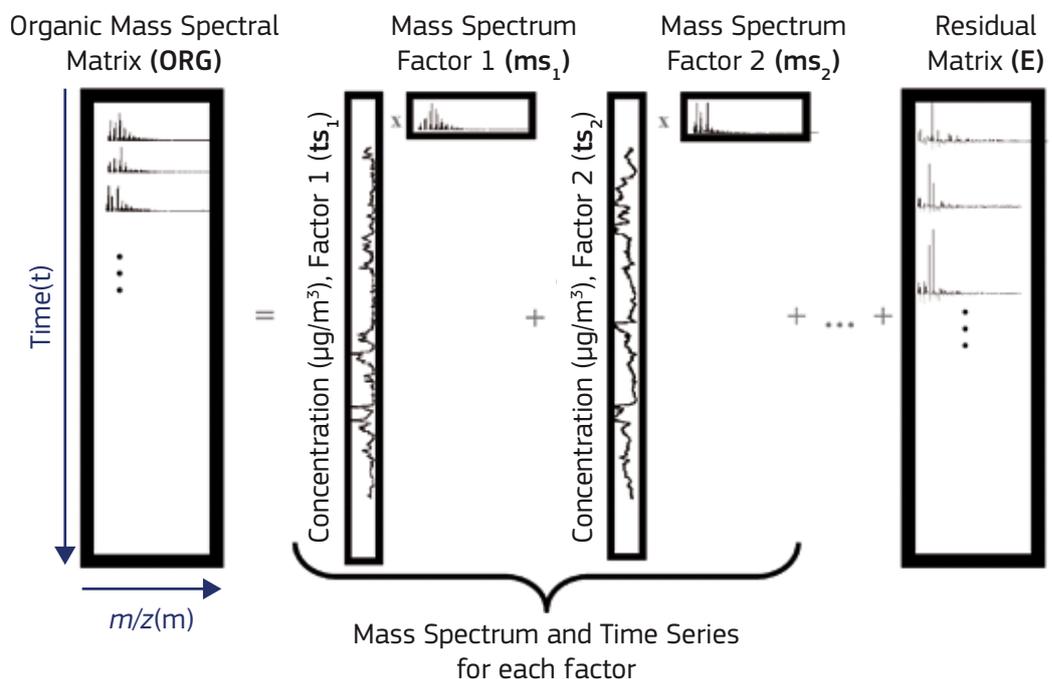
Factor name	Description
1. LV-OOA: low volatility oxygenated organic aerosol	Oxygen-to-carbon atomic ratio O/C = 0.63, often correlates with sulphate
2. SV-OOA: semi-volatile oxygenated organic aerosol	(O/C = 0.38) it correlates better with ammonium nitrate and chloride than with LV-OOA
3. NOA: nitrogen-enriched organic aerosol	Higher N/C ratio than other organic aerosol components
4. COA: cooking-related organic aerosol	Spectral features similar to those of particulate organic aerosol from cooking emissions and a distinctive diurnal pattern peaking during lunch and dinner times
5. HOA: hydrocarbon-like organic aerosol deriving from fossil fuel combustion	Given its low O/C ratio (0.06) and good correlation with primary combustion emission species, for example NO _x and elemental carbon
6. BBOA (or WBOA): particulate organic aerosol from biomass/wood burning	Spectral features similar to those from wood burning emissions, often with high evening contribution in areas where domestic heating is fuelled using wood. High correlation with levoglucosan or BC _{wb} from the Aethalometer model

The first version of the quadrupole-AMS has been available for 10 years (Jayne et al., 2002), the high-resolution AMS (De Carlo et al., 2006) for six years and the ACSM (Aerosol chemical speciation monitor) for around two years. The use of AMS is established and issues such as its composition-dependent collection efficiency are now characterised and can be taken into account. However, the AMS is very labour intensive to run over long time periods, while the ACSM is specifically designed for long-term monitoring (Ng et al., 2011). A first dedicated network of ACSM instruments in Europe was started in 2012 (<http://www.psi.ch/acsm-stations/acsm-and-emep-stations>). Some features of the AMS are still evolving, also allowing for ionisation techniques other than electron impact. The calibration is performed using ammonium nitrate. For organic matter, it would be advantageous to define

certain compounds or develop certified reference materials (CRMs) to ensure comparability of the organic mass spectra between different instruments. However, the current retrieved factors are already rather robust.

The AMS records the temporal variations of the composition and concentration of the organic aerosol in the form of a mass spectral matrix denoted "ORG" that usually comprises thousands of ensemble spectra with mass-to-charge ratios (m/z) of organic fragments acquired with a time resolution of seconds to minutes (Figure C2.2). Multivariate factor analysis is applied to deconvolute the observed ORG matrix into unique factors. Factor analysis of the data matrices from quantitative instruments usually involves solving a mass conservation model expressed as a two-dimensional bilinear equation. In the past, the solution to the equation has

Figure C2.2 Schematic representation of an ORG matrix (from Zhang et al., 2011 adapted from Ulrich et al., 2009)



been found using different methods: the custom principal component analysis (CPCA) method and multiple component analysis (MCA) (Zhang et al., 2005). More recently, Lanz et al. (2007) applied PMF for the first time on an AMS dataset acquired in Zurich, Switzerland.

A dedicated PMF tool programmed in the numerical computing environment IGOR Pro and a recent database make it possible to perform the analysis in a more standardised way and to compare different mass spectra from different solutions from places around the world. The programming language Multilinear Engine 2 (see section C4) was used by Lanz et al. (2008) to perform hybrid CMB-PMF analyses. In the study carried out in Zurich, hydrocarbon-like organic aerosol (HOA) was fixed to a certain degree while additional factors were freely obtained as in PMF. A new

IGOR interface called Source Finder (SoFi) has been developed by the Paul Scherrer Institute (Canonaco et al., 2013) to run analyses with ME-2, anywhere between CMB and PMF, for AMS and ACSM data. Solutions for filter-based measurements and other kind of data are under development (<http://www.psi.ch/acsm-stations/me-2>).

The guidelines for selecting the best solutions reported in table C2.2, adjusted for AMS spectra, are also useful for traditional source apportionment studies.

The family of mass spectrometric techniques for the analysis of aerosols has evolved swiftly in the past decade, reaching a degree of specialisation and diversification that makes it suitable either for the study of atmospheric processes and for long-term monitoring.

1. Calculate data and error matrices.
2. Further data and error treatment.
2a. Apply minimum error.
2b. Remove anomalous spikes, if desired.
2c. Smooth data, if desired.
2d. Downweight data with low signal-to-noise ratios.
2e. Downweight repeated information ($m/z = 44$ and related m/z values).
3. Run PMF for a range of factors (P) and random starts (seeds). Examine the ratio between the observed and expected Q (Q/Q_{exp}) vs P. A steep change in slope indicates the minimum P to be considered.
3a. Examine results from different random starts for each P. Sort results by Q/Q_{exp} values and compare the factors in each. Are there multiple types of results (representing local minima in the solution space)? If not, proceed to use SEED 0. If yes, can any of the result types be excluded because the factors are not physically meaningful? Proceed, exploring seeds that have results with physically meaningful factors.
3b. Try to determine the optimum number of factors by examining multiple criteria: <ul style="list-style-type: none"> • Look for correlations between factor time series and time series of external tracers. • Look for correlations between factor time series and time series of individual m/z values or ions. • Consider factor diurnal profiles, meteorological data, etc. • Examine factor mass spectra for tracer ions and fragmentation patterns. • Look for signs of “split” factors, considering the correlation of mass spectra and time series of factors in the same solution. After identifying factors that may have split, explore solutions with more factors to check for new, physically meaningful factors.
3c. Examine solution Q contributions and residuals. Do the residuals and Q values summed to form time series or mass spectra show periods or m/z values that do not fit well? Is this because the solution needs more factors, because the data do not fit the model of constant spectra for a given component, or because of instrumental drift, etc.? Are the distributions of the scaled residuals (x_{ij}/σ_{ij}) for each m/z approximately Gaussian, centred around 0, with a reasonable standard deviation?
4. For the best solution chosen from step 3, run PMF for a range of FPEAKs (Paatero, 2004; Norris, et al., 2008) such that the range of Q/Q_{exp} values is at least 3% above the minimum Q/Q_{exp} .
4a. Exclude from further consideration solutions that have unrealistic mass spectra and/or time series.
4b. Does changing the FPEAK change the solution in a way that would change the interpretation of the factors from step 3, or do these solutions just represent rotational ambiguity in the solution? If the interpretation changes, choose the most representative solution and support this choice. If the differences represent rotational ambiguity, choose the solution at FPEAK = 0.
5. Conduct bootstrapping analysis on the final solution from step 4 to estimate uncertainty in the solutions. *
6. Make and examine key diagnostic plots.
6a. Q/Q_{exp} vs varying P.

Table C2.2 Steps for preparing and choosing the best solution from PMF analysis of AMS datasets (adapted from Zhang et al., 2011)

6b. Q/Qexp vs FPEAK for the best P.
6c. Fractions of OA factors vs FPEAK for the best P.
6d. Correlations among PMF factors for the best P.
6e. The box and whisker plots of scaled residuals as a function of m/z for the best P.
6f. The time series of the measured OA concentration and the reconstructed organic mass (= sum of all factors) for the best P.
6g. The variations of the residual (= measured – reconstructed) of the fit as a function of time.
6h. The time series and mass spectra of total residuals and Q contribution for the best P solution.
6i. Comparisons of the P – 1, P, and P + 1 solutions for the acceptable FPEAK, where P is the best solution.

* Uncertainty estimate for each factor element should be obtained as the larger of the two values: variation caused by FPEAK variation, and variation in bootstrapped results (see chapter B10).

References

Canonaco F., Crippa M., Slowik J. G., Baltensperger U., Prévôt A. S. H. 2013, SoFi, an Igor based interface for the efficient use of the generalized multilinear engine (ME-2) for source apportionment: application to aerosol mass spectrometer data. *Atmospheric Measurement Techniques Discussions*, 6, 6409–6443.

Jayne, J.T., D.C. Leard, X. Zhang, P. Davidovits, K.A. Smith, C.E. Kolb, D.R. Worsnop, 2000. Development of an Aerosol Mass Spectrometer for Size and Composition. Analysis of Submicron Particles. *Aerosol Science Technology* 33 49–70.

Jimenez, J.L., M.R. Canagaratna, N.M. Donahue, A.S.H. Prevot, Q. Zhang, J.H. Kroll, P.F. DeCarlo, J.D. Allan, H. Coe, N.L. Ng, A.C. Aiken, K.D. Docherty, I.M. Ulbrich, A.P. Grieshop, A.L. Robinson, J. Duplissy, J. D. Smith, K.R. Wilson, V.A. Lanz, C. Hueglin, Y.L. Sun, J. Tian, A. Laaksonen, T. Raatikainen, Rautiainen, P. Vaattovaara, M. Ehn, M. Kulmala, J.M. Tomlinson, D.R. Collins, M.J. Cubison, E.J. Dunlea, J.A. Huffman, T.B. Onasch, M.R. Alfarra, P.I. Williams, K. Bower, Y. Kondo, J. Schneider, F. Drewnick, S. Borrmann, S. Weimer, K. Demerjian, D. Salcedo, L. Cottrell, R. Griffin, A. Takami, T. Miyoshi, S. Hatakeyama, A. Shimono, J.Y. Sun, Y.M. Zhang, K. Dzepina, J.R. Kimmel, D. Sueper, J.T. Jayne, S.C. Herndon, A.M. Trimborn, L.R. Williams, E.C. Wood, C.E. Kolb, A.M. Middlebrook, U. Baltensperger, D.R. Worsnop. 2009. Evolution of Organic Aerosols in the Atmosphere. *Science* 326, 1525–1529.

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

Lanz, V.A., M.R. Alfarra, U. Baltensperger, B. Buchmann, C. Hueglin, S. Szidat, M.N. Wehrli, L. Wacker, S. Weimer, A. Caseiro, H. Puxbaum, A.S.H. Prevot, 2008. Source Attribution of Submicron Organic Aerosols During Wintertime Inversions

by Advanced Factor Analysis of Aerosol Mass Spectra. *Environmental Science & Technology* 42(1), 214–220.

Middlebrook A.M., R. Bahreini, J.L. Jimenez, M.R. Canagaratna, 2012. Evaluation of Composition-Dependent Collection Efficiencies for the Aerodyne Aerosol Mass Spectrometer using Field Data. *Aerosol Science and Technology* 46, 258–271,

Ng, N.L., S.C. Herndon, A. Trimborn, M.R. Canagaratna, P.L. Croteau, T.B. Onasch, D. Sueper, D.R. Worsnop, Q. Zhang, Y.L. Sun, J.T. Jayne., 2011. An Aerosol Chemical Speciation Monitor (ACSM) for Routine Monitoring of the Composition and Mass Concentrations of Ambient Aerosol. *Aerosol Science and Technology* 45(7), 780–794.

Norris G., Vedantham R., K. Wade, S. Brown, J. Prouty, C. Foley, 2008. EPA Positive Matrix Factorization (PMF 3.0) Fundamentals & User Guide. EPA 600/R-08/108.

Paatero P., 2004. User's guide for positive matrix factorization programs PMF2 and PMF3, Parts 1 and 2. University of Helsinki, Helsinki, Finland.

Ulbrich I.M., M.R. Canagaratna, Q. Zhang, D.R. Worsnop, J.L. Jimenez., 2009. Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data. *Atmospheric Chemistry and Physics* 9(9), 2891–2918.

Zhang, Q., M.R. Alfarra, D.R. Worsnop, J.D. Allan, H. Coe, M.R. Canagaratna, J.L. Jimenez, 2005. Deconvolution and Quantification of Hydrocarbon-like and Oxygenated Organic Aerosols Based on Aerosol Mass Spectrometry. *Environmental Science & Technology* 39, 4938–4952.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Ulbrich, I. M., Ng, N. L., Worsnop, D. R., & Sun, Y., 2011. Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. *Analytical and Bioanalytical Chemistry* 401(10), 3045–3067.

C3. THE AETHALOMETER MODEL

The Aethalometer instrument was originally developed to quantify light absorption by elemental carbon, which is considered to be the predominant light-absorbing aerosol species at visible wavelengths (Hansen et al., 1984). However, several studies recently pointed out that organic carbon significantly absorbs light in the ultraviolet wavelengths and less significantly going into the visible (e.g. Kirchstetter et al., 2004). This fraction, known as brown carbon for its light brownish colour, includes tar materials from smouldering fires or solid fuel combustion, pyrolysis products from biomass burning and humic-like substances from soil or biogenic emissions (Feng et al., 2013).

Light absorption by aerosols is usually parameterised as proportional to $\lambda^{-\alpha}$, where λ is the light wavelength and α represents the Ångström absorption exponent. While the spectral dependence of elemental carbon light absorption is low ($\alpha \sim 1$, Bond and Bergstrom, 2006), brown carbon exhibits a much higher Ångström absorption exponent (up to 7, see e.g. Hoffer et al., 2006). Based on these differences in optical properties, a growing number of studies recently used multi-wavelength Aethalometers to detect and/or apportion wood burning carbonaceous aerosols in ambient air (e.g. Jeong et al.,

2004; Sandradewi et al., 2008a, 2008b; Yang et al., 2009; Favez et al., 2009, 2010; Sciare et al., 2011).

Some of the most recent works proposed methodologies where total carbonaceous material (CM_{total}) could be primarily considered as the sum of brown-carbon-containing carbonaceous material (i.e. CM_{wb} here), non-brown-carbon-containing carbonaceous material originating from fossil fuel combustion (CM_{ff}), and non-combustion organic aerosol (CM_{other}), as follows:

$$CM_{total} = CM_{ff} + CM_{wb} + CM_{other} = C_1 \times b_{abs,ff,950nm} + C_2 \times b_{abs,wb,470nm} + C_3 \tag{C3.1}$$

where $b_{abs,ff,950nm}$ represents the absorption coefficient of CM_{ff} at 950nm, $b_{abs,wb,470nm}$ represents the absorption coefficient of CM_{wb} at 470 nm, C_1 and C_2 relate the light absorption to the particulate mass of both sources (Figure C3.1), and C_3 corresponds to the amount of non-combustion organic aerosol (assumed here to have a negligible light absorption capacity).

It should be noted that CM_{ff} comprises traffic emissions as well as carbonaceous aerosols

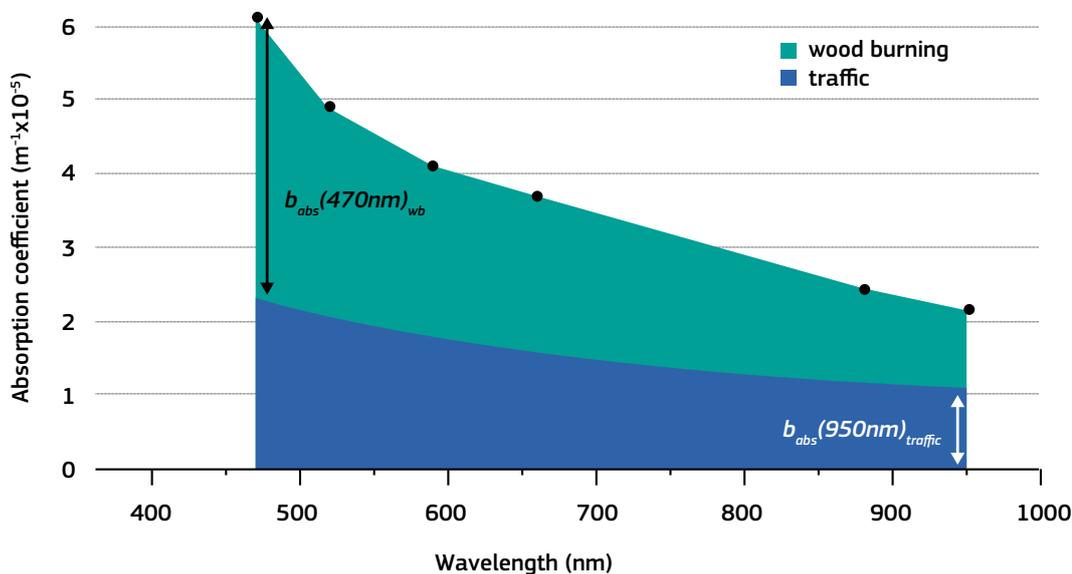


Figure C3.1 Graphical representation of the apportionment of light absorption between wood burning and traffic sources (from Sandradewi et al., 2008b).

originating from fuel oil and natural gas combustion, but excludes coal-burning organic aerosol. Indeed, the latter was shown to significantly absorb light at near UV wavelengths (e.g. Yang et al., 2009) and may thus interfere with $b_{\text{abs,wb},470\text{nm}}$. Another limitation of this approach might be the presence of mineral dust particles (notably containing iron oxides), that also absorb light at near UV wavelengths (Fialho et al., 2006) and should thus be considered carefully.

The development of receptor models based on multi-wavelength light absorption is still in the early stages and is subject to continuous improvements and to trials in various use applications. In particular, different methodologies are currently proposed to resolve equation C3.1, using for instance universal or site-specific C_1 and C_2 constants. On the other hand, it should also be kept in mind that these methodologies are very sensitive to initial conditions (and especially to the chosen Ångström absorption exponent), which leads to high uncertainties. This is the reason why users usually perform (and give results of) sensitivity tests with wide ranges for these initial conditions (see e.g. Favez et al., 2010 and Sciare et al., 2011), and suggest that the results of these sensitivity tests be considered as the total uncertainties of the model outputs. Finally, it should be mentioned that, due to the methodology used by Aethalometers (filter-based measurement), absorption coefficients directly obtained from these instruments are affected by various sampling and analytical artefacts (mostly referred to as multiple scattering and shadowing effects) which need to be carefully corrected prior to any data treatment (Collaud Coen et al., 2010 and references therein).

Recently, Wang et al. (2012a and b) included DeltaC (the difference in Aethalometer BC measured at 370 nm and that measured at 880 nm) in their PMF analyses of data from Rochester, NY. With the typical collection of elements, ions, organic carbon and elemental carbon, the addition of DeltaC provided a clear resolution of biomass burning from traffic sources (Wang et al., 2012a). In an analysis including molecular markers, the DeltaC was observed primarily in the biomass burning factor along with levoglucosan (Wang et al., 2012b).

The considerable increase in measurements carried out using Aethalometers associated with the interest in the potential impacts of elemental carbon on climate and on health, makes this technique an interesting resource for improving the understanding of aerosol sources, with particular reference to biomass burning.

Care is however recommended in the interpretation of data from the Aethalometer because of the non-specific nature of its measurements (Harrison et al., 2013).

References

- Bond, T.C., Bergstrom, R.W., 2006. Light absorption by carbonaceous particles: an investigative review. *Aerosol Science Technology* 40, 27-67
- Collaud Coen, M., Weingartner, E., Apituley, A., Ceburnis, D., Fierz-Schmidhauser, R., Flentje, H., Henzing J.S., Jennings, S.G., Moerman, M., Petzold, A., Schmid, O. Baltensperger, U., 2010. Minimizing light absorption measurement artifacts of the Aethalometer: evaluation of five correction algorithms. *Atmospheric Measurement Techniques* 3, 457-474.
- Favez, O., Cachier, H., Sciare, J., Sarda-Estève, R., Martinon, L., 2009. Evidence for a significant contribution of wood burning aerosols to PM_{2.5} during the winter season in Paris, France. *Atmospheric Environment* 43, 3640-3644.
- Favez, O., El-Haddad, I., Piot, C., Boreave, A., Abidi, E., Marchand, N., Jaffrezou, J.-L., Besombes, J.-L., Personnaz, M.-B., Sciare, J., Wortham, H. D'Anna, B., 2010. Inter-comparison of source apportionment models for the estimation of wood burning aerosols during wintertime in an Alpine city (Grenoble, France). *Atmospheric Physics and Chemistry* 10, 5295-5314.
- Feng Y., Ramanathan V., Kotamarthi V. R., 2013. Brown carbon: a significant atmospheric absorber of solar radiation? *Atmospheric Chemistry and Physics*, 13, 8607-8621.
- Fialho, P., Freitas, M.C., Barata, F., Viera, B., Hansen, A.D.A., Honrath, R.E., 2006. The Aethalometer calibration and determination of iron concentration in dust aerosols, *Journal of Aerosol Science* 37, 1497-1506.
- Hansen, A.D.A., Rosen, H., Novakov, T., 1984. The Aethalometer – An instrument for the real-time measurement of optical absorption by aerosol particles. *Science of the Total Environment* 36, 191-196.
- Harrison R.M., Beddows, D.C.S., Jones, A.M., Calvo, A., Alves, C. and Pio, C.A., 2013. An evaluation of some issues regarding the use of Aethalometers to measure woodsmoke concentrations. *Atmospheric Environment* 80, 540-548.

- Hoffer, A., Gelencsér, A., Guyon, P., Kiss, G., Schmid, O., Frank, G.P., Artaxo, P., Andreae, M.O., 2006. Optical properties of humic-like substances (HULIS) in biomass-burning aerosols. *Atmospheric Chemistry and Physics* 6, 3563-3570.
- Jeong, C.-H., Hopke, P.K., Kim, E., Lee, D.-W., 2004. The comparison between thermal-optical transmittance elemental carbon and Aethalometer black carbon measured at multiple monitoring sites. *Atmospheric Environment* 38, 5193-5204.
- Kirchstetter, T.W., Novakok, T., Hobbs, P.V., 2004. Evidence that the spectral dependence of light absorption by aerosols is affected by organic carbon. *Journal of Geophysical Research* 109, D21208.
- Sandradewi, J., Prévôt, A.S.H., Weingartner, E., Schmidhauser, R., Gysel M., Baltensperger, U., 2008a. A study of wood burning and traffic aerosols in an Alpine valley using a multi-wavelength, Aethalometer. *Atmospheric Environment* 42, 101-112.
- Sandradewi, J., Prévôt, A.S.H., Szidat, S., Perron, N., Alfarra, M.R., Lanz, V.A., Weingartner, E., and Baltensperger, U., 2008b. Using aerosol light absorption measurements for the quantitative determination of wood burning and traffic emission contributions to particulate matter. *Environmental Science and Technology* 42, 3316-3323.
- Sciare, J., d'Argouges, O., Sarda-Esteve, R., Gaimoz, C., Dolgorouky, C., Bonnaire, N., Favez, O., Bonsang, B., and Gros, V., 2011. Large contribution of water-insoluble secondary organic aerosols in the Paris region (France). *Journal of Geophysical Research* 116, D22203.
- Weingartner, E., Saathoff, H., Schnaiter, M., Streit, N., Bitnar, B., Baltensperger, U., 2003. Absorption of light by soot particles: determination of the absorption coefficient by means of Aethalometers. *Journal of Aerosol Science* 34, 1445-1463.
- Yang, M., Howell, S.G., Zhuang, J., Huebert, B.J., 2009. Attribution of aerosol light absorption to black carbon, brown carbon and dust in China - interpretations of atmospheric measurements during EAST-AIRE. *Atmospheric Chemistry and Physics* 9, 2035-2050.
- Wang, Y., Hopke, P.K., Xia, X., Rattigan, O., Chalupa, D.C., Utell, M.J., 2012a. Multiple year black carbon measurements and source apportionment using Delta-C in Rochester, NY. *Journal of Air and Waste Management Association* 62, 880-887.
- Wang, Y., Hopke, P.K., Xia, X., Rattigan, O., Chalupa, D.C., Utell, M.J., 2012b. Source apportionment of airborne particulate matter using inorganic and organic species as tracers. *Atmospheric Environment* 55: 525-532.

C4. APPORTIONMENT OF THE PM CARBONACEOUS FRACTION: RADIOCARBON AND TRACER ANALYSIS

The carbonaceous fraction is one of the main components of particulate matter (PM). The study of carbonaceous aerosol is important because of its adverse effects on health (Highwood & Kinnersley, 2006; Mauderly & Chow, 2008), air quality (Putaud et al., 2004; Turpin & Huntzicker, 1995; Vecchi et al., 2008; among others), visibility (Watson, 2002), cultural heritage (Bonazza et al., 2005), and the Earth's radiation balance (IPCC, 2007).

Total carbon (TC) in atmospheric aerosols consists of two main fractions: elemental (EC) and organic (OC) carbon. EC is produced by the incomplete combustion of fossil and biomass fuels in an oxygen-poor environment (Chow et al., 2001). It is the most refractory carbon fraction and the most efficient solar-light absorber. OC is contained in organic matter which is composed of thousands of chemical constituents belonging to many compound classes, for which complete characterisation is extremely difficult. Carbonatic carbon (CC), that is the carbon contained in carbonates, may also be present, however its contribution to total carbon may be considered negligible in most European areas, with few exceptions (Perrone et al., 2011; Cuccia et al., 2011; Yubero et al., 2011).

While EC is exclusively produced by direct combustion emissions, OC may derive from primary sources (primary organic carbon, POC), such as fossil-fuel combustion, biomass burning and bioaerosol emissions, as well as from the atmospheric gas-to-particle conversion of other pollutants through condensation processes (driven by temperature and dilution effects) and oxidation processes (secondary organic carbon, SOC). Since most of the emitted POC is semivolatile and some gas-to-particle processes take place shortly after emission, some authors consider the distinction between POC and SOC to be obsolete. In this document, this terminology is used for coherence with the reviewed literature and to emphasise the difference between sources and processes, which is relevant for the development of abatement measures. The lack of direct chemical analysis

methods for the determination of either POC or SOC led to the development of different indirect approaches, of which the most widespread is the method based on variations of measured OC/EC ratios. In this approach, elemental carbon is assumed to be a conservative tracer for primary combustion-generated OC emissions, and SOC simply appears as an increase in the OC/EC ratio relative to that of the primary OC/EC ratio (Turpin and Huntzicker, 1995).

Large uncertainties still affect emission inventories of carbonaceous particles. Monks et al. (2009) reviewed global emission estimates: uncertainties up to factors 3.4 and 80 are reported for primary and secondary carbonaceous particles, respectively. The highest uncertainties still concern natural emissions.

The reactivity, volatility and to some extent the hygroscopicity of compounds in the OC fraction, also including main source tracers (such as levoglucosan), may compromise the basic assumptions for receptor models and strongly increase difficulties and uncertainties in source apportionment.

In this context, the use of "inert" tracers, such as the $^{14}\text{C}/^{12}\text{C}$ isotopic ratio, may be of great help. Radiocarbon measurement of TC is a good tool for fossil/non-fossil source separation (Currie, 2000 and the literature cited therein; Hildemann et al., 1994). The main principle may be briefly explained as follows. "Modern" carbon from biomass contains a constant proportion of radioactive ^{14}C , giving a $^{14}\text{C}/^{12}\text{C}$ isotopic ratio of 1. ^{14}C decays with a radiocarbon half-life of 5730 years, which means that none is left in fossil fuels, which are millions of years old. Therefore, as the fraction of modern carbon (fm) is zero for fossil fuels, and as fm should be 1 for modern materials, it is possible to estimate the proportion of fossil and non-fossil fuels that led to a particular level of total carbon in the atmosphere by looking at the value of the $^{14}\text{C}/^{12}\text{C}$ isotopic ratio. Actually, nuclear tests in the 1950s increased the $^{14}\text{C}/^{12}\text{C}$ ratio in the atmosphere by up to a factor of 2 in the

early 1960s. Values have been slowly decreasing since then and f_m is now approaching 1: the trend of the ^{14}C content in the atmosphere can be found in Levin et al. (2010). The excess of ^{14}C in the atmosphere led to the increase of $^{14}\text{C}/^{12}\text{C}$ ratio in biological material and must be taken into account when apportioning modern sources.

However, the sole use of radiocarbon measurements on total carbon only allows for a separation between modern and fossil contributions. This simple division is not enough to apportion natural and anthropogenic sources since modern carbon could result from natural emissions as well as from wood/biomass burning and other anthropogenic activities (such as cooking). To overcome this limitation, Szidat et al. (2004; 2006) proposed performing radiocarbon measurements of OC and EC separately. In this way, EC may be directly apportioned between fossil-fuel combustion and biomass burning, and the fossil-fuel combustion contribution to OC may be also directly obtained; provided that the OC/EC emission ratio for wood/biomass burning is known. This model is limited by the uncertainty regarding the knowledge of the OC/EC emission ratio for wood/biomass burning and by the difficulty in the assessment of the secondary contribution of this source (Szidat et al., 2009), as the OC/EC emission ratio measured at the source cannot correctly account for secondary aerosol formation. Moreover, this method requires an effective physical isolation of the two carbonaceous fractions (Andersson et al., 2011; Bernardoni et al., 2013; Calzolari et al., 2011;

Heal et al., 2011; Szidat et al., 2004, 2009; Zencak et al., 2007; Zhang et al., 2012), which are operationally defined quantities. Indeed, the analytical separation of OC from EC using thermal protocols is ambiguous because part of the OC can pyrolyse, especially in an oxygen-poor atmosphere, and some of this EC can evolve in the presence of oxygen (Watson et al., 2005). It is also noteworthy that water-soluble organic carbon (WSOC) is particularly prone to pyrolysis and that soluble inorganic compounds can catalyse EC pre-combustion (Chow et al., 2001; Novakov and Corrigan, 1995; Wang et al., 2010; Yu et al., 2002).

Recent literature has attempted a natural/anthropogenic source apportionment, coupling ^{14}C measurements of TC with the analysis of other micro and macro tracers (Gelencsér et al., 2007; Gilardoni et al., 2011; Holden et al., 2011; May et al., 2009; Yttri et al., 2011a, 2011b). A number of tracers and emission factors have been employed in these studies: levoglucosan as tracer for biomass combustion together with OC/levoglucosan and OC/EC emission ratios; carbon monoxide as tracer for primary fossil-fuel combustion together with the OC/EC emission ratio; cellulose for plant debris together with the OC/cellulose ratio; arabinol and mannitol saccharide concentrations as tracers of fungal spores. In these papers, marker concentrations, emission ratios and their uncertainties were used to estimate possible ranges of source contributions identified by modelling techniques.

Table C4.1. Schematic representation of the carbonaceous fractions derived by combining radiocarbon measurements with organic markers in Gelencsér et al. (2007).

carbonaceous	organic/elemental	primary/secondary	fossil, biomass, burning, biogenic	how is it estimated?
Total carbon (TC)	Elemental carbon (EC)	(only primary)	Fossil fuel (EC_{FF})	by subtracting EC_{BB} from measured EC
			Biomass burning (EC_{BB})	from OC_{BB} and the OC/EC emission ratio for wood burning
	Organic carbon (OC)	Primary organic carbon (POC)	Fossil fuel (OC_{FF})	from EC_{FF} and the OC/EC ratio for fossil-fuel combustion
			Biomass burning (OC_{BB})	from levoglucosan and the OC/levoglucosan ratio for wood burning
			Biogenic (OC_{BIO})	derived from cellulose and the OC/cellulose emission ratio
		Secondary organic carbon (SOC)	Fossil fuel (SOC_{FF})	using the radiocarbon measurement of TC
			Biomass burning (SOC_{BB})	

In the work by Gelencsér et al. (2007), measurements of EC, OC, levoglucosan, cellulose and fm(total carbon) are used for TC apportionment in the following basic classes (Table C4.1): EC from fossil fuel combustion (EC_{FF}) and from biomass burning (EC_{BB}), POC from fossil fuel combustion (OC_{FF}), biomass burning (OC_{BB}) and biogenic sources (OC_{BIO}), and SOC from precursors emitted by fossil and non-fossil sources. Briefly, OC_{BIO} is derived from cellulose and the OC/cellulose emission ratio; OC_{BB} from levoglucosan and the OC/levoglucosan ratio for wood burning; EC_{BB} from OC_{BB} and the OC/EC emission ratio for wood burning; EC_{FF} by subtracting EC_{BB} from measured EC; OC_{FF} from EC_{FF} and the OC/EC ratio for fossil-fuel combustion. SOC is grouped as fossil and non-fossil using the radiocarbon measurement of TC.

This method involves many steps, each of which has substantial uncertainty, mainly due to the high variability of emission ratios: to tackle the multitude of possible combinations of these uncertainty parameters, a statistical approach, the Latin-hypercube sampling method, was used. A very similar approach is used by Gilardoni et al. (2011). In this case, however, the estimation of the primary biogenic OC contribution is based on the measurements of arabinol and mannitol saccharides, assuming that fungal spores dominate primary biological aerosol particles in $PM_{2.5}$. A Quasi-Montecarlo approach is used to determine the most probable contributions of the different source categories and the associated uncertainties.

In other works (Bernardoni et al., 2013; Minguillon et al., 2011; Heal et al., 2011; Aiken et al., 2010), source apportionment based on fm measurements of OC and EC fractions is reinforced/compared with other approaches, such as the levoglucosan-tracer method, PMF-Aerosol Mass Spectrometry, PMF-offline filter dataset and Chemical Mass Balance of Organic Molecular Markers. In Bernardoni et al. (2013) and Minguillon et al. (2011), the fossil OC is also apportioned into primary and secondary contributions; moreover, a further apportionment of the non-fossil OC is performed in order to take into account modern contributions to OC other than wood burning and biogenic aerosols (e.g. cooking, tyre wear). In particular, in Bernardoni et al. (2013) the fossil fuel combustion contribution to OC is split into primary and secondary components using the EC-tracer method (Turpin and Huntzicher, 1995) which uses PMF-derived emission ratios, while the non-fossil OC is grouped into primary OC_{BB} , secondary OC_{BB} , OC_{BIO} and OC from other urban modern sources, according to emission ratios and literature approaches.

It should be noted that these recent probabilistic uncertainty characterisations have demonstrated that results obtained with such trace-based methods may have high uncertainties (Larsen et al, 2012).

Finally, it is also worth mentioning that radiocarbon analyses are extremely time-consuming and expensive, due to the procedures for sample preparation and to the need for an Accelerator Mass Spectrometry system to determine the radiocarbon concentration. Such features limit the number of samples that can be characterised and, therefore, the representativeness of the obtained data. Nevertheless, some steps to overcome such limitations have been recently taken by coupling a commercial EC/OC analyser with an Accelerator Mass Spectrometry system (Perron et al., 2010).

References

- Aiken A. C., de Foy B., Wiedinmyer C., DeCarlo P. F., Ulbrich I. M., Wehrli M. N., Szidat S., Prevot A. S. H., Noda J., Wacker L., Volkamer R., Fortner E., Wang J., Laskin A., Shutthananda V., Zheng J., Zhang R., Paredes-Miranda G., Arnott W. P., Molina L. T., Sosa G., Querol X., Jimenez J. L., 2010. Mexico city aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) – Part 2: Analysis of the biomass burning contribution and the non-fossil carbon fraction. *Atmospheric Chemistry and Physics* 10, 5315–5341.
- Andersson, A., Sheesley, R.J., Krusa, M., Johansson, C., Gustafsson, O., 2011. 14C-based source assessment of soot aerosols in Stockholm and the Swedish EMEP-Aspvreten regional background site. *Atmospheric Environment* 45, 215–222.
- Bernardoni V., Calzolari G., Chiari M., Fedi M., Lucarelli F., Nava S., Piazzalunga A., Riccobono F., Taccetti F., Valli G. & Vecchi R., 2013. Radiocarbon analysis on organic and elemental carbon in aerosol samples and source apportionment at an urban site in Northern Italy. *Journal of Aerosol Science* 56, 88–99.
- Bonazza, A., Sabbioni, C., & Ghedini, N., 2005. Quantitative data on carbon fractions in interpretation of black crusts and soiling on European built heritage. *Atmospheric Environment* 39, 2607–2618.
- Calzolari G., Bernardoni V., Chiari M., Fedi M., Lucarelli F., Nava S., Riccobono F., Taccetti F., Valli G., Vecchi R., 2011. The new sample preparation line for radiocarbon measurements on atmospheric aerosol at LABEC. *Nuclear Instruments and Methods B* 269, 203–208.

- Chow, J.C., Watson, J.G., Crow, D., Lowenthal, D.H., & Merrifield, T., 2001. Comparison of IMPROVE and NIOSH carbon measurements. *Aerosol Science and Technology* 34, 23–34.
- Cuccia, E., Piazzalunga, A., Bernardoni, V., Brambilla, L., Fermo, P., Massabò, D., Molteni, U., Prati, P., Valli, G., & Vecchi, R., 2011. Carbonate measurements in PM₁₀ near the marble quarries of Carrara (Italy) by infrared spectroscopy (FT-IR) and source apportionment by positive matrix factorization (PMF). *Atmospheric Environment* 45, 6481–6487.
- Currie, L.A., 2000. Evolution and multidisciplinary frontiers of 14C aerosol science. *Radiocarbon* 42, 115–126.
- Gelencsér, A., May, B., Simpson, D., Sánchez-Ochoa, A., Kasper-Giebl, A., Puxbaum, H., Caseiro, A., Pio, C., & Legrand, M., 2007. Source apportionment of PM_{2.5} organic aerosol over Europe: primary/secondary, natural/anthropogenic, and fossil/biogenic origin. *Journal of Geophysical Research* 112, D23S04.
- Gilardoni, S., Vignati, E., Cavalli, F., Putaud, J.P., Larsen B.R., Karl, M., Stenstrom, K., Gemberg, J., Henne, S., & Dentener, F., 2011. Better constraints on sources of carbonaceous aerosols using a combined 14C-macro tracer analysis in a European rural background site. *Atmospheric Chemistry and Physics* 11, 5685–5700.
- Heal, M.R., Naysmith, P., Cook, G.T., Xu, S., Ravento's Duran, T., & Harrison, R.M., 2011. Application of 14C analyses to source apportionment of carbonaceous PM_{2.5} in the UK. *Atmospheric Environment* 45, 2341–2348.
- Holden, A.S., Sullivan, A.P., Munchak, L.A., Kreidenweis, S.M., Schichtel, B.A., Malm, W.C., & Collet, J.L., Jr., 2011. Determining contributions of biomass burning and other sources to fine particle contemporary carbon in the western United States. *Atmospheric Environment* 45, 1986–1993.
- Hildemann, L.M., Klinedinst, D.B., Klouda, G.A., Currie, L.A., & Cass, G.R., 1994. Sources of urban contemporary carbon aerosol. *Environmental Science and Technology* 28, 1565–1575.
- Highwood, E.J., & Kinnerson, R.P., 2006. When smoke gets in our eyes: the multiple impacts of atmospheric black carbon on climate, air quality and health. *Environment International* 32, 560–566.
- IPCC 2007. In: S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, & H.L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, UK and New York, NY, USA.
- Larsen B.R., Gilardoni S., Stenström K., Niedzialek J., Jimenez J., Belis C.A., 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. *Atmospheric Environment* 50, 203–213.
- Levin, I., Naegler, T., Kromer, B., Diehl, M., Francey, R.J., Gomez-Pelaez, A.J., Steele, P., Wagenbach, D., Weller, R., & Worthy, D.E., 2010. Observations and modelling of the global distribution and long-term trend of atmospheric 14CO₂. *Tellus B* 62B, 26–46.
- Mauderly, J.L., Chow, J.C., 2008. Health effects of organic aerosols. *Inhalation Toxicology* 20, 257–288.
- May, B., Wagenbach, D., Hammer, S., Steier, P., Puxbaum, H., & Pio, C., 2009. The anthropogenic influence on carbonaceous aerosol in the European background. *Tellus* 61B, 464–472.
- Minguillon, 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., Teinila, K., Seco, R., Penuelas, J., Metzger, A., Schallhart, S., Müller, M., Hansel, A., Burkhart, J.F., Baltensperger, U., & Prévot, A.S.H., 2011. Fossil versus contemporary sources of fine elemental and organic carbonaceous particulate matter during the DAURE campaign in Northeast Spain. *Atmospheric Chemistry and Physics* 11, 12067–12084.
- Monks, P.S., Granier, C., Fuzzi, S., Stohl, A., Williams, M.L., Akimoto, H., Amanni, M., Baklanov, A., Baltensperger, U., Bey, I., Blake, N., Blake, R.S., Carslaw, K., Cooper, O.R., Dentener, F., Fowler, D., Fragkou, E., Frost, G.J., Generoso, S., Ginoux, P., Grewet, V., Guenther, A., Hansson, H.C., Hennew, S., Hjorth, J., Hofzumahaus, A., Huntrieser, H., Isaksen, I.S.A., Jenkin, M.E., Kaiser, J., Kanakidou, M., Klimont, Z., Kulmala, M., Laj, P., Lawrence, M.G., Lee, J.D., Liousse, C., Maione, M., McFiggans, G., Metzger, A., Mieville, A., Moussiopoulos, N., Orlando, J.J., O'Dowd, C.D., Palmer, P.I., Parrish, D.D., Petzold, A., Platt, U., Poschl, U., Prévot, A.S.H., Reeves, C.E., Reimann, S., Rudich, Y., Sellegri, K., Steinbrecher, R., Simpson, D., ten Brink, H., Theloke, J., van der Werf, G.R., Vautard, R., Vestreng, R., Vlachokostas, C., von Glasow, R., 2009. Atmospheric composition change—global and regional air quality. *Atmospheric*

Environment 43, 5268–5350.

Novakov, T., & Corrigan, C.E., 1995. Thermal characterization of biomass smoke particles. *Mikrochimica Acta* 119, 157–166.

Perron N., Szidat S., Fahrni S., Ruff M., Wacker L., Prévot A.S.H, Baltensperger U., 2010. Towards on-line ¹⁴C analysis of carbonaceous aerosol fractions. *Radiocarbon* 52, 761–768.

Perrone, M.R., Piazzalunga, A., Prato, M., Carofalo, I., 2011. Composition of fine and coarse particles in a coastal site of the Central Mediterranean: carbonaceous specie contributions. *Atmospheric Environment* 45, 7470–7477.

Piazzalunga, A., Belis, C., Bernardoni, V., Cazzuli, O., Fermo, P., Valli, G., Vecchi, R., 2011. Estimates of wood burning contribution to PM by the macrotracer method using tailored emission factors. *Atmospheric Environment* 45, 6642–6649.

Putaud, J.P., Raes, F., Van Dingenen, R., Brüggemann, E., Facchini, M.C., Decesari, S., Fuzzi, S., Gehrig, R., Hüglin, C., Laj, P., Lorbeer, G., Maenhaut, W., Mihalopoulos, N., Müller, K., Querol, X., Rodriguez, S., Schneider, J., Spindler, G., ten Brink, H., Tørseth, K., Wiedensholer, A., 2004. A European aerosol phenomenology-2: chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. *Atmospheric Environment* 38, 2579–2595.

Szidat, S., Jenk, T.M., Gaggeler, H.W., Synal, H.-A., Hajdas, I., Bonani, G., Saurer, M., 2004. THEODORE, a two-step heating system for the EC/OC determination of radiocarbon (¹⁴C) in the environment. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 223–224, 829–836.

Szidat, S., Jenk, T.M., Synal, H.-A., Kalberer, M., Wacker, L., Hajdas, I., Kasper-Giebl, A., Baltensperger, U., (2006). Contributions of fossil fuel, biomassburning, and biogenic emissions to carbonaceous aerosols in Zurich as traced by ¹⁴C. *Journal of Geophysical Research* 111, D07206.

Szidat, S., Ruff, M., Perron, N., Wacker, L., Synal, H.-A., Hallquist, M., Shannigrahi, A.S., Yttri, K.E., Dye, C., Simpson, D., 2009. Fossil and non-fossil sources of organic carbon (OC) and elemental carbon (EC) in Goteborg, Sweden. *Atmospheric Chemistry and Physics* 9, 1521–1535.

Turpin, B.J., Huntzicker, J.J., 1995. Identification of secondary organic aerosol episodes and

quantitation of primary and secondary organic aerosol concentrations during SCAQS. *Atmospheric Environment* 29, 3527–3544.

Vecchi R., Chiari M., D'Alessandro A., Fermo P., Lucarelli F., Mazzei F., Nava S., Piazzalunga A., Prati P., Silvani F., Valli G., 2008. A mass closure and PMF source apportionment study on the sub-micron sized aerosol fraction at urban sites in Italy. *Atmospheric Environment* 42, 2240–2253.

Wang, Y., Chung, A., & Paulson, S.E., 2010. The effect of metal salts on quantification of elemental and organic carbon in diesel exhaust particles using thermal-optical evolved gas analysis. *Atmospheric Chemistry and Physics* 10, 11447–11457.

Watson, J.G., 2002. Visibility: science and regulation. *Journal of Air & Waste Management Association* 52, 628–713.

Watson, J.G., Chow, J.C., & Chen, L.-W.A., 2005. Summary of organic and elemental carbon/black carbon analysis methods and intercomparisons. *Aerosol and Air Quality Research* 5, 65–102.

Yttri, K.E., Simpson, D., Nøjgaard, J.K., Kristensen, K., Genberg, J., Stenstrom, K., Swietlicki, E., Hillamo, R., Aurela, M., Bauer, H., Offenberg, J.H., Jaoui, M., Dye, C., Eckhardt, S., Burkhart, J.F., Stohl, A., & Glasius, M., 2011a. Source apportionment of the summer time carbonaceous aerosol at Nordic rural background sites. *Atmospheric Chemistry and Physics* 11, 13339–13357.

Yttri, K.E., Simpson, D., Stenstrom, K., Puxbaum, H., & Svendby, T., 2011b. Source apportionment of the carbonaceous aerosol in Norway—quantitative estimates based on ¹⁴C, thermal-optical and organic tracer analysis. *Atmospheric Chemistry and Physics* 11, 9375–9394.

Yu, J.Z., Xu, J.H., & Yang, H., 2002. Charring characteristics of atmospheric organic particulate matter in thermal analysis. *Environmental Science and Technology* 36, 754–761.

Yubero, E., Carratalá, A., Crespo, J., Nicolás, J., Santacatalina, M., Nava, S., Lucarelli, F., & Chiari, M., 2011. PM₁₀ source apportionment in the surroundings of the San Vicente del Raspeig cement plant complex in southeastern Spain. *Environmental Science and Pollution Research* 18, 64–74.

Zencak, Z., Elmquist, M., & Gustafsson, O., 2007. Quantification and radiocarbon source apportionment of black carbon in

atmospheric aerosols using the CTO-375 method. *Atmospheric Environment* 41, 7895–7906.

Zhang, Y.L., Perron, N., Ciobanu, V. G., Zotter P., Minguillon M. C., Wacker, L., Prevot A.S.H., Baltensperger U., & Szidat, S., 2012. On the isolation of OC and EC and the optimal strategy of radiocarbon-based source apportionment of carbonaceous aerosols, *Atmospheric Chemistry and Physics* 12, 10841–10856.

C5. CONSTRAINED AND EXPANDED MODELS IN FACTOR ANALYSIS

Common Positive Matrix Factorization (PMF) analysis fits the data into a limited bilinear model. However, PMF also permits the development of more complex mathematical models to account for physical and chemical information when fitting the data. “Constrained” and “Expanded” PMF models represent the advanced tools in receptor modelling, and efforts are being made to improve and increase their capabilities. Today they feature in less than 10% of published studies, but this percentage is likely to increase in the near future.

Since this is a new field of research, the terminology is still evolving in the literature and the distinction between Constrained and Expanded models is not very well defined yet. In this document, Constrained models are considered a subcategory of Expanded (or Extended) models. More precisely, Constrained models are those in which additional constraints are introduced (in most cases after an initial run, the results of which are used as a starting point), while Expanded models are those in which the customary bilinear equation is augmented by another more complicated set of equations, depending on the aims of the study.

Constrained PMF

By definition, the Positive Matrix Factorization model is a weighted least squares analysis where the object function is minimised under the constraint that all or some of the elements of G and F are constrained to non-negative values (Paatero, 1997). Therefore, all PMF studies are constrained. Nevertheless, recent literature uses the term ‘constrained’ to refer to more complicated PMF models, where the constraint is not limited to non-negativity.

Different types of constraint can be implemented in PMF, but they must all derive from some a priori knowledge of the user about the system that is to be modelled. This knowledge can be of physical or chemical origin (Amato et al., 2009). Physical constraints can relate, for example, to the mass conservation principle

(e.g. the sum of factor profiles cannot exceed unity; the lower the particle size, the lower the source contribution, etc.). On the other hand, chemical information is associated with source profiles. The relative abundance of some elements/compounds may already be known and can represent valuable information for the model in order to find a better solution, reducing the number of possible alternatives (the ‘rotational ambiguity’). Another example of a priori knowledge is the information about periods during which a specific point source is not operative. These data can be useful constraints to drive the model towards a more realistic solution by setting the emission of that source to zero.

The choice of the program to use in performing a Constrained PMF depends on the type of constraint to be used:

- PMF2 implements only the F_{key} and G_{key} constraints, which consist of binding individual elements of the F and G matrices, respectively, to zero. $G_{key_{ik}}$ and $F_{key_{kj}}$ are two matrices of the same shapes as G and F respectively. They are applied *a posteriori* (in a ‘continuation run’ that takes place after the base run) and each element of the matrix with a key value >1 is bound to zero, with an increased strength of the bond for higher key values. It is not possible to bind elements to non-zero values. Both constraints are imposed regardless of changes in the Q value, i.e. they are considered to be “hard” constraints.
- ME-2 (Multilinear Engine) is a special-purpose programming language, which allows for the incorporation of any additional constraints that are introduced by the user into the script (Paatero, 1999; Paatero and Hopke, 2009; Amato et al., 2009; Amato and Hopke, 2012). The constraints can be introduced in terms of pulling equations, upper/lower limits and fixed values. Pulling equations are

weighted by uncertainties, which express the confidence of the user in the equation. A lower uncertainty corresponds to a harder pulling effect. Each pulling equation is converted into an auxiliary term of the object function to be minimised.

- EPA PMF v5.0 includes a user-friendly interface for introducing constraints in several ways:
 - Ratios of F elements (e.g. $Si_{soil}/Al_{soil}=3.2$)
 - Mass Balance between F elements of the same or different factors (e.g. $Fe_{diesel}=2.5 Fe_{gasoline}$)
 - Custom expression, where the user can build any kind of equation on F and G elements about which he/she is confident (e.g. $G_{smelter}=0$ from August to December 2008)

When F and/or G elements are set to zero or confined to upper/lower limits, the constraints are “hard” or imposed without regard to the change in the Q value. Equations constraining variables towards a value, upwards or downwards, are classified as “soft” pulling, and their strength (based on the confidence of the user) is expressed by the limit of change allowed in the Q value. A higher dQ will determine a harder constraint.

Once the constraints are applied in a continuation run, the user should look at the deviations in the results between the two model runs and examine the impact of the constraints on:

- Achievement of the target values (within the uncertainty range in the case of ME-2)
- The increase of dQ
- Correlations between factor profiles and reference source profiles
- Changes in G-space plots
- Possible distortions in all factors and source contributions
- Possible factor swaps, so that identities of factors have changed. Such swaps cause the constraints to act on physical factor(s) that are different from those originally intended, so the constraints are meaningless. For details, see Paatero et al. (2013).

Sensitivity tests, carried out modifying the strength of pulling equations can be useful for a comprehensive evaluation of the model output (Viana et al., 2009; Brown et al., 2012).

Expanded PMF

As already mentioned, the Multilinear Engine (ME-2) has been used to constrain PMF profiles and contributions. However, the flexible structure of ME-2 makes it suitable for solving any other complex problems such as expanded models. ME-2 has been applied to several datasets for multiple purposes.

In general, the expanded models were found to give similar source contributions and source profiles when compared with the original PMF analyses, but also provided information associated with the meteorological and temporal conditions. In some cases, the expanded model provided additional resolution of sources: Kim et al. (2003) were able to resolve diesel and gasoline emissions using the expanded model when they had been unable to do so with the basic bilinear factor analysis model. However, much more equivocal results were obtained for Washington, DC relative to the prior PMF analyses (Begum et al., 2005).

Zhao et al. (2004) developed a novel factor analysis model, in which the normal chemical mass balance model was augmented by a parallel equation that accounted for wind speed and direction, temperature, and weekend/weekday effects. The model was fitted with a multilinear engine (ME) to provide identification and apportionment of the VOC sources in Houston during the Texas Air Quality Study (TexAQ5) 2000. The analysis determined the profiles and contributions of nine sources and the corresponding wind speed, wind direction, temperature, and weekend factors. The reasonableness of these results suggested the high resolving power of the expanded factor analysis model for source apportionment, but also provides novel and effective auxiliary information for more specific source identification. This study demonstrates the feasibility of the expanded model to identify sources in complex VOC systems.

Zhou et al. (2009) developed an expanded model to investigate the effect of wind direction, wind speed, seasons, and weekdays/weekends in the Cleveland (Ohio, USA) area. The expanded model and PMF2 produced essentially the same results with only minor differences being observed between the two sets of profiles and contributions. Thus, the addition of meteorological and temporal parameters to the model did not improve the source resolution. Zhao and Hopke (2006) followed a similar approach in Indianapolis and conclude that PMF coupled with a posteriori back-trajectory analysis (such as CPF, PSCF, seasonal variation analysis, and weekday/weekend variation analysis) yields comparable results to expanded factor analysis and is simpler to employ.

New monitoring technologies permit the measurement of a variety of chemical species with time resolution as high as 10 minutes to one hour. However, most species are still measured with longer integration periods such as several hours to a day. Traditional factor analysis techniques (PCA and customary PMF) are unable to analyse datasets consisting of different time scale measurements. Zhou et al. (2004) developed an expanded PMF model which can use each data value (of a mixed time-resolution dataset) in its original time schedule without averaging or interpolation. Averaging the high time-resolution data leads to a loss of valuable temporal information, while interpolating the low time-resolution data produces unreliable high-resolution series. The contribution series are smoothed by the regularisation of auxiliary equations especially for sources containing very little high-resolution species. Similar study designs were followed by Ogulei et al. (2005) and Zhao et al. (2004).

An expanded receptor model was applied to identify and apportion the PM_{2.5} sources that were common to three different environments (personal, indoor and outdoor) and to which asthmatic children were exposed (Zhao et al., 2007). Two types of sources (factors), external and internal, were defined – the external sources were left free to contribute to all three environments while the internal sources were constrained to only contribute to the personal and indoor samples. The expanded receptor model was expressed as:

$$x_{ijdt} = \sum_{p=1}^N g_{ipdt} f_{jp} + \sum_{p=N+1}^{N+H} g_{ipdt} f_{jp}$$

(t=1 for personal; t=2 for indoor)

$$x_{jdt} = \sum_{p=1}^N g_{pdt} f_{jp}$$

(t=3 for outdoor)

where i is the individual index, j the species index, d the sampling date index, t the type index, N the number of external sources, and H the number of internal sources. x_{jdt} denotes the concentration of species j in the sample of type t collected by subject i on date d , g_{ipdt} denotes the contribution of source p to the sample of type t collected by subject i on date d , and f_{jp} denotes the relative concentration of species j in source p . Further information is available in Hopke et al. (2003) and Zhao et al. (2006).

This approach was able to resolve four external (sulphate, soil, nitrate and traffic) and three internal (chlorine-based cleaning, cooking, tobacco) sources. Strict bilinear (PMF2) and trilinear models (PMF3) were also applied to

indoor-outdoor-personal samples (Larson et al., 2006).

Pere-Trepat et al. (2007) analysed data that combined particle size and composition data using an expanded PMF model to permit the extraction of maximal information from size-segregated aerosol composition data. This three-way model accounts for the variation in the composition of the source emissions in the different size ranges (three-stage DRUM impactor; Pere-Trepat et al., 2007). The data are three-way in that their size and composition are measured over time. Three-way data have been also analysed by means of the stricter trilinear PARAFAC model (Yakovleva et al., 1999; Hopke et al., 2003), which does not offer the flexibility of ME-2.

With the ME-2 approach (Pere-Trepat et al., 2007), each profile is a matrix of $m \times n$ dimension where m is the number of measured variables and n the number of measured size fractions. The profiles are then a three-dimensional array of source by composition by size. For each source (factor), there is a vector of mass contributions, so combining them produces a matrix whose dimensions are defined by the number of sampling days by the number of sources (factors).

This model evolved from the Tucker 1 model (Tucker, 1966). This model is logically a two-way model, but is organised as a three-way array with data also in a three-way array, X . The main equation of the model is as follows:

$$\bar{X} = A \otimes \bar{B} + \bar{E}$$

$\bar{X}(i,j,k)$ is the three-way array of observed data, \otimes represents a Kronecker product of the $\bar{B}(i,j,k)$ with the contribution matrix $A(i,p)$ and $\bar{E}(i,j,k)$ is a three-way array of residuals (Pere-Trepat et al., 2007).

Developing new models using ME-2

As already mentioned, new models can be developed by modifying existing ME-2 scripts, or by writing entirely new ones. Due to the difficulties normally encountered in debugging new scripts, practitioners are advised to use existing code as much as possible. To that end, developers are encouraged to obtain information on existing script material, preferably in the early stages of their work. In this way, they also contribute to guiding the future development of ME-2 in directions that are most useful for the further development of the source apportionment methods.

References

- Amato, F., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Pey, J., Perez, N., Hopke, P.K., 2009. Quantifying road dust resuspension in urban environment by multilinear engine: a comparison with PMF2. *Atmospheric Environment* 43, 2770–2780.
- Amato, F., Hopke, P.K., 2012. Source Apportionment of the Ambient PM2.5 in East St. Louis using Constrained Positive Matrix Factorization. *Atmospheric Environment* 46, 329–337.
- Begum, B.A., Hopke, P.K., Zhao, W., 2005. Source Identification of Fine Particles in Washington DC by Expanded Factor Analysis Modeling. *Environmental Science and Technology* 55, 227–240.
- Brown, S.G., Lee, T., Norris, G.A., Roberts, P.T., Collett, J.L., Paatero, P., Worsnop, D.R., 2012. Receptor modeling of near-roadway aerosol mass spectrometer data in Las Vegas, Nevada, with EPA PMF. *Atmospheric Chemistry and Physics* 12, 309–325.
- Hopke, P.K., Ramadan, Z., Paatero, P., Norris, G., Landis, M., Williams, R., Lewis, C.W., 2003. Receptor modeling of ambient and personal exposure samples: 1998 Baltimore particulate matter epidemiology–exposure study. *Atmospheric Environment* 37, 3289–3302.
- Kim, E., Hopke, P.K., Paatero, P., Edgeton, E.S., 2003. Incorporation of Parametric Factors into Multilinear Receptor Models Studies of Atlanta Aerosol. *Atmospheric Environment* 37, 5009–5021.
- Larson, T., Gould, T., Simpson, C., Liu, L.-J.S., Claiborn, C., Lewtas, J., 2004. Source apportionment of indoor, outdoor, and personal PM2.5 in Seattle, Washington, using positive matrix factorization. *Journal of the Air and Waste Management Association* 54 (9), 1175–1187.
- Ogulei, D., Hopke, P.K., Zhou, L., Paatero, P., Park, S.S., Ondov, J.M., 2005. Receptor modeling for multiple time resolved species: The Baltimore supersite. *Atmospheric Environment* 39 (20), 3751–3762.
- Paatero, P., 1997. Least squares formulation of robust non negativity factor analysis. *Chemometrics and Intelligent Laboratory Systems* 37, 23–35.
- Paatero, P., 1999. The multilinear engine: a table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics* 8, 854–888.
- Paatero, P., Hopke, P.K., 2009. Rotational tools for factor analytic models. *Journal of Chemometrics* 23 (2), 91–100.
- Paatero, P., Eberly, S., Brown, S.G., Norris, G.A., 2013. Methods for estimating uncertainty in factor analytic solutions. *Atmospheric Measurement Techniques Discussions*, 6, 7593–7631.
- Pere-Trepat, E., Kim, E., Paatero, P., Hopke, P.K., 2007. Source apportionment of time and size resolved ambient particulate matter measured with a rotating DRUM impactor. *Atmospheric Environment* 41 (28), 5921–5933.
- Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 279–311.
- Viana, M., Amato, F., Querol, X., Alastuey, A., Moreno, T., García, S., Herce, D., Fernandez, R., 2009. Chemical tracers of particulate emissions from commercial shipping. *Environmental Science and Technology* 43 (19), 7472–7477.
- Yakovleva, E., Hopke, P.K., Wallace, L., 1999. Receptor modeling assessment of PTEAM data. *Environmental Science and Technology* 33, 3645–3652.
- Zhao, W., Hopke, P.K., Norris, G., Williams, R., Paatero, P., 2006. Source apportionment and analysis on ambient and personal exposure samples with a combined receptor model and an adaptive blank estimation strategy. *Atmospheric Environment* 40, 3788–3801.
- Zhao, W., Hopke, P.K., Gelfand, E.W., Rabinovitch, N., 2007. Use of an expanded receptor model for personal exposure analysis in schoolchildren with asthma. *Atmospheric Environment* 41 (19), 4084–4096.
- Zhao, W., Hopke, P.K., Karl, T., 2004. Source Identification of Volatile Organic Compounds in Houston, Texas. *Environmental Science and Technology* 38 (5), 1338–1347.
- Zhao, W., Hopke, P., 2006. Source investigation for ambient PM 2.5 in Indianapolis, IN. *Aerosol Science and Technology* 40 (10), 898–909.
- Zhou, L., Hopke, P.K., Zhao, W., 2009. Source apportionment of airborne particulate matter for the speciation trends network site in Cleveland, OH. *Journal of the Air and Waste Management Association* 59 (3), 321–331.
- Zhou, L., Hopke, P.K., Paatero, P., Ondov, J.M., Pancras, J.P., Pekney, N.J., Davidson, C.I., 2004. Advanced factor analysis for multiple time resolution aerosol composition data. *Atmospheric Environment* 38 (29), 4909–4920.

European Commission

EUR 26080 – Joint Research Centre – Institute for Environment and Sustainability

Title: European Guide on Air Pollution Source Identification with Receptor Models

Authors: Claudio A. Belis, Bo R. Larsen, Fulvio Amato, Imad El Haddad, Olivier Favez, Roy M.Harrison, Philip K. Hopke, Silvia Nava, Pentti Paatero, André Prévôt, Ulrich Quass, Roberta Vecchi, Mar Viana

Luxembourg: Publications Office of the European Union

2014 – 88 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series - ISSN 1831-9424 (online) ISSN 1018-5593 (print)

ISBN 978-92-79-32513-7 (pdf) ISBN 978-92-79-32514-4 (print)

doi: 10.2788/9307

Abstract

This report contains a guide and a European harmonised protocol for the identification of air pollution sources using receptor models. The document aims at disseminating and promoting the best available methodologies for source identification and at harmonising their application across Europe. It was developed by a committee of leading experts within the framework of the JRC initiative for the harmonisation of source apportionment that has been launched in collaboration with the European networks in the field of air quality modelling (FAIRMODE) and measurements (AQUILA).

The protocol has been conceived as a reference document that includes tutorials, technical recommendations and check lists connected to the most up-to-date and rigorous scientific standards. As a guide, it is structured in sections with increasing levels of complexity that make it accessible to readers with different degrees of familiarity with this topic, from air quality managers to air pollution experts and modellers.

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

*Serving society
Stimulating innovation
Supporting legislation*



Publications Office

ISBN 978-92-79-32513-7



9 789279 325137

doi:10.2788/9307