MEGAPOLI Scientific Report 10-22
Evaluation of Source Apportionment Methods
MEGAPOLI Deliverable 4.6

Co-Editors:
Nicolas Moussiopoulos, John Douros, George Tsegas

Contributing Co-Authors:
Spyros N. Pandis, Christos Fountoukis, Lia Fragkou, Eleftherios Chourdakis, Vassilis Akylas

Difference map for $O_3$ concentrations calculated with baseline and the zeroed out traffic emissions

Thessaloniki, 2010
Colophon

Serial title:
MEGAPOLI Project Scientific Report 10-22

Title:
Evaluation of Source Apportionment Methods

Subtitle:
MEGAPOLI Deliverable 4.6

Editors:
Nicolas Moussiopoulos, John Douros, George Tsegas

Contributing Author(s):
Spyros N. Pandis, Christos Fountoukis, Lia Fragkou, Eleftherios Chourdakis, Vassilis Akylas

Responsible institution(s):
Laboratory of Heat Transfer and Environmental Engineering (LHTEE)
Aristotle University of Thessaloniki (AUTH)

Language:
English

Keywords:
Source apportionment, particulate matter, receptor modelling

Url:

Digital ISBN:
978-87-92731-03-6

MEGAPOLI:
MEGAPOLI-25-REP-2010-12

Website:
www.megapoli.info

Copyright:
FP7 EC MEGAPOLI Project
Content:

1 Source apportionment modelling methodologies in the EU member states ........................................4
  1.1 Introduction ................................................................................................................................4
  1.2 Methodologies and current practice in source apportionment ...................................................5
    1.2.1 Receptor models ..................................................................................................................5
    1.2.2 Dispersion models ..............................................................................................................5
    1.2.3 Hybrid approaches ..............................................................................................................8
    1.2.4 Model validation and estimation of uncertainty .................................................................8
    1.2.5 Current practice in source apportionment in EU countries .................................................9

2 Application of 3D chemical transport model (PMCAMx) in the European domain ......................10
  2.1 Introduction ..............................................................................................................................10
  2.2 PMCAMx-2008 description .....................................................................................................11
  2.3 Model application ....................................................................................................................12
  2.4 Results and discussion .............................................................................................................14

3 Analysis of Paris 2009 Campaign Data using Receptor Modelling ................................................21
  3.1 Theoretical background ...........................................................................................................21
    3.1.1 Principal Component Analysis (PCA) ..............................................................................21
    3.1.2 Nonnegative Matrix Factorisation (NMF) ........................................................................22
    3.1.3 Caveats and limitations .....................................................................................................22
  3.2 The Paris Measurement Network and the 2009 Campaign ......................................................23
  3.3 Measurement analysis using PCA ............................................................................................24
  3.4 Measurement analysis using NMF ...........................................................................................26
  3.5 PMF using other tools ..............................................................................................................27

4. Use of Dispersion Modelling for Supporting Receptor-based Source Apportionment .....................30
  4.1 Simulations of Pollutant Dispersion in the Greater Paris Area ................................................30
  4.2 PCA based on zero-out modelling ...........................................................................................31
  4.3 Nonnegative Matrix Factorisation (NMF) ...............................................................................37
  4.4 Comparison of PCA and NMF .................................................................................................39
  4.5 Trajectory modelling ................................................................................................................40

5. Conclusions ....................................................................................................................................43

Acknowledgements ............................................................................................................................44

References..........................................................................................................................................45

Previous MEGAPOLI reports............................................................................................................51
Abstract
This report aims at describing results that have been produced in the frame of WP4 – Task 4.4 on source apportionment and the identification and quantification of relevant source contributions. The report starts with a review of source apportionment modelling methodologies used in the EU member states, aiming to describe the current state-of-play regarding source apportionment efforts. At the same time it seeks to identify shortcomings in present methodologies as well as emerging trends in research in order to suggest possible courses towards the objectives set out by the task. The core of the report focuses on model applications that took place in the framework of MEGAPOLI, aiming to explore an integrated approach in source apportionment, involving innovative combinations of standard receptor modelling and state-of-the-art CTMs. This survey begins with an application of the PMCAMx chemical transport model in the European domain, which included an implementation of a new mechanism for primary organic aerosol volatility and further oxidation of secondary organic aerosol. A combined approach is then introduced, where dispersion modelling is used for supporting the receptor-based source apportionment of the Paris 2009 Campaign Data.

1 Source apportionment modelling methodologies in the EU member states

1.1 Introduction
Pollutant levels exceeding EU limit values for the protection of health and the environment constitute a major concern of most European cities. This is, for example, reflected in the significant number of EU countries that have applied for a postponement of attainment of PM$_{10}$ limits until 2011, according to Article 22 of the Air Quality Directive (AQD) 2008/50/EU. Efficient air quality management is required in order to ensure that the legal limits are not exceeded and that the economical and social costs of poor air quality are controlled and minimised. The first step to ensure the application of successful mitigation measures and the development of appropriate pollution abatement national strategies is the accurate identification of pollution sources and of their individual contributions to the ambient pollutant concentrations. To this end, a wide range of modelling methodologies has been proposed and applied for the apportionment of atmospheric pollutants, including receptor-based models, methods based on the application of advanced chemical dispersion modelling, as well as a combination of both. Nevertheless, efforts to systematically evaluate and compare the performance of alternative methodologies, as well as estimate their intrinsic uncertainties have been scarce (see e.g. Favez et al., 2010; Viana et al., 2008b; Miller et al., 2002). Systematic evaluation efforts could serve as a foundation for the development of a harmonised approach in source apportionment, which could then be applied for a range of pollutants, source types and horizontal scales. It is worth mentioning that, within the frame of the Forum for Air Quality Modelling in Europe (FAIRMODE – http://fairmode.ew.eea.europa.eu), a working subgroup on the “Contribution of natural sources and source apportionment” has been formed, aiming to provide useful guidance and suggest best modelling practices and quality assurance procedures for member countries, in order to promote harmonised model use for source apportionment in the EU (Fragkou et al., 2010). In general, the use of modelling tools in combination with measurements is recommended for the purpose of source apportionment, as monitoring of the contributions from all emitting sources in an air quality management zone or agglomeration would be unrealistic and economically unfeasible.
1.2 Methodologies and current practice in source apportionment

At a first stage, source apportionment may be based on the observation and analysis of monitoring data, through examining their correlation with relevant meteorological parameters and subtracting of regional background levels from the urban background and hot-spot concentrations, in order to determine the importance of local sources. A similar methodology can be used to quantify natural contributions, however in this case regional background levels are subtracted from those measured at the urban and traffic stations of interest for a specific period of days. The occurrence of concentration peaks of measurements simultaneously at different stations which are representative for a variety of horizontal scales can indicate an episode due to transboundary pollutant transport or due to an accidental release. Although the importance of monitoring data for screening purposes and for an initial air quality assessment cannot be disputed, their use is subject to limitations, mainly due to issues of spatial and temporal representativity compromised by the increased costs associated with adequate coverage and reliability. Therefore, source apportionment based solely on the use of monitoring data is usually applied for limited time periods, for which high resolution measurements can be obtained, for example from dedicated campaigns (Gélenescér et al., 2007), such as the one performed in the framework of WP3 of the MEGAPOLI project (http://megapoli.dmi.dk/maininfo/wp3.html).

Source apportionment for primary species is relatively simple to obtain due to the linearity that characterises source-receptor relationships (Yarwood et al., 2007). This is not the case, however, for secondary pollutants, where complex chemical interactions between sources and species may not allow for the interpretation of pollutant concentrations in direct correlation to emissions from a particular source. Therefore, chemical dispersion grid models which account for a number of chemical reactions between atmospheric pollutant species and between sources are the scientifically suitable tools for investigating source contributions for secondary pollutants. The use of both receptor models and dispersion models for source apportionment is described in the following paragraphs.

1.2.1 Receptor models

The fundamental principle of receptor modelling is that mass conservation is assumed and a mass balance analysis can be used to determine and apportion ambient pollutant concentrations to individual emitting sources. A mass balance equation is written to account for all chemical species identified in the filter samples analysed, as contributions from independent sources. Receptor models represent a statistical evaluation of ambient measurements at different times and locations, thus the selection of the appropriate method depends on prior knowledge on the sources and source profiles. If the sources are known and detailed information on source profiles is available, Chemical Mass Balance (CMB) models can be applied, whereas in case the sources are unknown and there is limited information on source profiles, Principal Component Analysis (PCA) and Positive Matrix Factorisation (PMF) or Nonnegative Matrix Factorisation (NMF) methods are preferred. Apart from the different Chemical Mass Balance modelling tools, receptor models can also be categorised as Aerosol Evolution and Equilibrium models, which estimate how reduction in one precursor will affect PM end-products, and as Back Trajectory models, which identify the origin of polluted air masses usually transported to the receptor from long distance sources. A review of European publications, which were reported in the questionnaires submitted by EU countries within the frame of the COST 633 Action, revealed that PCA was the most frequently used modelling method for source apportionment, as it appeared in 30% of the studies, while back-trajectory analysis was represented in 11% of the studies (Viana et al., 2008a). Other receptor models were also frequently used, such as PMF (8%), CMB (7%) and mass balance analysis (7%).

1.2.2 Dispersion models

In contrast to the receptor models that use ambient concentrations as inputs to calculate source contributions, dispersion models use source emissions as input data to calculate ambient concentra-
tions. Eulerian and Lagrangian dispersion models describe the chemical and physical atmospheric processes in order to predict pollutant concentrations and can be applied at different spatial scales. Gaussian dispersion models are source-oriented models that are useful for source apportionment purposes as they characterise atmospheric processes by dispersing a pollutant directly emitted from the source of interest at selected downwind receptor locations. Gaussian simulations allow for the assessment of contribution from a new source to the ambient pollution load, however multiple sources can also be considered in a single simulation. Recently, several EU countries have relied on the combined use of available emission data and a dispersion model to estimate source contributions. For example, the hybrid Swedish AIRVIRO dispersion model has been applied in a number of European cities including Prague, Riga, Vilnius and Tallinn. The Gaussian ADMS-urban model has been used for source apportionment and for the evaluation of proposed emission reduction measures in Poland, within the frame of an air quality management project in Cracow between 2005 and 2006 (Adamczyk et al., 2007). Eulerian dispersion modelling systems have been used recently in Mediterranean member states to assess the Saharan dust contribution to ambient particle concentrations. In the studies by Astitha et al. (2005) and Kallos et al. (2006), the SKIRON/ETA dispersion forecasting system was applied for urban Mediterranean regions, while Rodríguez et al. (2001) have combined SKIRON results with back-trajectory analysis to determine the proportion of Sahara-induced exceedances with respect to the total annual exceedances in Southern Spain. However, grid dispersion models are not developed to directly provide information on source apportionment, because the impact of all sources has been combined to calculate the total pollutant concentration in each cell. Thus, two main approaches in the use of grid chemical dispersion models for source apportionment purposes have emerged, which fall into two general categories: sensitivity analysis and reactive tracers.

The sensitivity analysis approach investigates concentration–emission responses by exploring the sensitivity of model results to changes in emissions input. The relevant methodologies are conceptually simple to apply and are directly related to air quality planning and management, as they allow for a realistic consideration of the effects on air quality of planned developments and public works. The following methodologies are included in the sensitivity analysis approach:

- **Brute Force or Direct Method**: It is the most straightforward sensitivity method for quantifying the response of pollutant concentrations at receptors to perturbations at sources. The main stages in the methodology include setting up and running a model Base Case simulation, repeating the simulation with perturbed emissions from a particular source, and comparing the two simulation results. The difference in the concentration output between the base case and the sensitivity case accounts for the effect of changing that emissions source on air pollutant concentrations at all receptor sites. This method is simple and can be applied to any model, but may not always be practical because computational cost increases linearly with the cases of emission perturbations to examine. Accuracy also may be an issue for the brute force method for the smaller concentration changes between the simulations, as for small input changes the output change may be influenced in a large degree by numerical precision or model “noise” (Koo et al., 2009).

- **Zero-out modelling**: The zero-out method can be considered as an extreme case variation of the brute force method, as emissions from a particular source are changed (i.e. reduced) by 100%. Thus, the zero-out method is used to quantify the change in the pollutant concentration that would occur if the relevant source is removed (Koo et al., 2009). Similarly to the brute force approach, zero-out modelling is simple to apply with any model but is not practical for investigating the effect of many sources, because a complete model run is required for each source. Other disadvantages of this method are that results may be inaccurate in the case of small emission source and when indirect effects, such as complex chemical reactions, are present.

- **Decoupled direct method (DDM)**: The DDM is a frequently used method to examine the effect of emission changes on calculated concentrations. The DDM is conceptually similar to the brute force method but is directly implemented in the host model, i.e. sensitivity equations
derived from the governing equations of the atmospheric processes modelled are directly solved. Therefore, one of its drawbacks is that the implementation is technically challenging and large computer memory is required to examine many sensitivities simultaneously (Koo et al., 2009). One alternative that is numerically more efficient is the DDM-3D variation that uses less rigorous numerical algorithms to solve time evolution of the chemistry sensitivity equations than those used to solve concentrations, although this may cause potential inconsistencies between sensitivities and concentrations. Thus, DDM-3D has been implemented in regularly used grid dispersion models, such as CAMQ (Napelenok et al., 2006) and CAMx (Koo et al., 2009). The DDM has potential advantages of greater efficiency and accuracy relative to the brute force method, as shown in the findings of several inter-comparison studies of sensitivity analysis methodologies (Cohan et al., 2005; Martien et al., 2006; Napelenok et al., 2006).

Although the sensitivity analysis methodologies constitute a simple and powerful tool in assessing the importance of the relative contribution of selected sources, they are usually computationally expensive and their accuracy and validity largely depends on the base case scenario. Numerical noise may also contaminate the results of sensitivity analysis approaches and they also do not provide source apportionment for non-linear systems. Most specifically, sensitivity analysis methods may not be able to provide a realistic response of the concentrations of secondary pollutants, generated by nonlinear interactions of various precursor substances, to precursor emission changes. This is particularly important for examining future ozone, nitrogen dioxide or particulate matter concentrations, as a result of changes in precursor emissions from a specific source. Sensitivity analysis results may be liable to errors due to “cross-sensitivity” interactions between the impacts of multiple emission sources for the pollutant species of interest, i.e. when the sensitivity of ambient concentrations of a pollutant to emissions from one source is influenced by the emission rate of another source (Cohan et al., 2005). In these cases, Reactive Tracer Methods may be more suitable. In this approach, new tracers or tags (as extra species) and new algorithms are added to air quality models to track the chemical conversions, transport and fate of pollutants from selected emissions sources. Computational efficiency is achieved through using computed changes in bulk species concentrations to account for the changes for tagged species within individual atmospheric processes. The conceptual idea behind this methodology is that a standard grid dispersion model calculates concentrations in each grid cell for a particular species that has many sources and the predicted total concentration of this species is the concentration due to all sources. A reactive tracer is assigned to each source, so that the sum of the reactive tracers will equal total concentration of the species under consideration. Numerical algorithms for solving the reactive tracer concentrations should be developed to support this equality (Yarwood et al., 2007). Three main methodologies that fall under this category of Reactive Tracers are regularly applied in grid dispersion models:

- **PM Source Apportionment Technology (PSAT):** In PSAT, tracer algorithms are formulated to include tracers for all sources and not only for a single source of interest, as in other Reactive Tracer methods. The method is used to apportion primary PM, secondary PM and gaseous precursors to secondary PM among different source categories and source regions. PSAT has been implemented in the “Comprehensive Air quality Model with extensions” (CAMx), and its performance regarding source apportionment has been compared to DDM (Koo et al., 2009). The PSAT and DDM were applied in the same regional modelling framework to assess the model responses to various brute force emission reductions by 100% and 20%, for several primary and secondary pollutants. Although none of the methods can be considered to be more accurate, the results are comparable for pollutants that are linearly related to emissions but present considerable differences when nonlinearity and/or indirect effects are involved.

- **Tagged Species Source Apportionment (TSSA):** TSAA is a similar source apportionment methodology to PSAT. The main difference is that TSSA explicitly solves tagged species using the same algorithms as the host model for physical atmospheric processes like advection and diffusion, i.e. an “online” approach is adopted. This method is relatively straightforward for non-reactive species, when one tracer is added for each source, but more complicated for
chemical reactive species, such as NO\textsubscript{x}, and secondary particulates, as the chemical reactions that convert a tracer between different gaseous species and from a gaseous species to PM have also to be traced. Unlike PSAT which includes tracers for all sources, TSAA can be used to estimate contributions of emissions either from a single source or from selected classes of sources, grouped by source category and/or by region. Similarly to PSAT, TSAA is useful for evaluating the contribution of small sources that cannot be evaluated with “brute force” sensitivity methodologies, and the results between the two methods are quite similar.

- **Source Oriented External Mixture (SOEM):** SOEM is a more rigorous approach than PSAT and TSAA, as each tagged species is separately simulated through all physical and chemical atmospheric processes. It is therefore potentially more accurate but also more computationally demanding than the other Reactive Tracer methods. In the case of secondary PM and related gases, chemical change is accounted for by expanding the chemical mechanism to treat different source regions/categories as separate precursor and product species. Despite of this potential advantage of SOEM, in a study by Yarwood et al. (2007) which evaluated different methods for PM source apportionment, SOEM and PSAT produced similar results, when implemented in the CAMx dispersion model. In the following paragraphs, some other examples on the use of receptor and dispersion models in EU member countries for source apportionment are described.

### 1.2.3 Hybrid approaches

In an attempt to overcome the deficiencies of each approach hybrid approaches have been proposed, attempting to integrate the current technology of receptor models, which begin with pollutant information monitored at a receptor point, with the transport and dispersion models, which start with the source characteristics, in order to define the apportionment of monitored data to their sources and estimate the uncertainty involved. Hybrid models use not only the concentration measurements as input data, but in addition source emission rates and atmospheric dispersion/transformation information characteristic of dispersion models. By incorporating this additional information, hybrid modelling allows for improved source apportionment estimates and consideration of problems that are inaccessible in terms of classical receptor modelling. In a relevant study on PM\textsubscript{10} source apportionment, a simple Gaussian model has been coupled with a CMB receptor model, formulated to break down pollutant contributions according to independent meteorological periods (Haupt, 2005). Another case where the combination of receptor and dispersion models is recommended is for control strategy development, particularly if the emission inventory used in a dispersion model is marginally adequate (Hopke, 1991). Receptor models can also be used to evaluate the results of dispersion models by scrutinizing the inventory assumed in the dispersion calculations.

### 1.2.4 Model validation and estimation of uncertainty

The EU official guidelines suggest that models are the scientifically relevant tools to be used also for source apportionment (2008/50/EU). Thus, the models applied for source apportionment have to be tested and assessed in order to ensure that they meet certain quality objectives before considered to be suitable for regulatory use. Particularly in the case of receptor statistical models, any solution does not by itself guarantee physical reality and thus a validation procedure is essential to ensure reliable outcomes. For this purpose, statistical receptor models are preferably used in conjunction with dispersion models, especially to justify emission reduction measures on different source types. As for all modelling tools used for regulatory purposes, total model uncertainty has to be considered for models used in source apportionment, including model uncertainties, emission inventory uncertainties and uncertainties relating to meteorological variability during air pollution episodes. In general, model validation is performed by means of comparing model results to measurements. Monitoring data accuracy and coverage has to be ensured in this case, thus it is recommended that data from dedicated monitoring campaigns are used for this purpose, such as in the study by Simpson and Yttri (2009) for Switzerland, Sweden and Norway. A number of recent European publica-
tions focus on the intercomparison of different receptor models applied to a single data set. In most cases there is considerable disagreement between results from different modelling tools, due to the different underlying theoretical assumptions. The recent study by Favez et al. (2010) compares the performance of CMB and PMF receptor models in determining the contribution of wood burning organic aerosols to the total organic fraction of the field samples obtained from Grenoble, France. It was found that the CMB model overestimated the wood burning contribution, probably due to the loss of semi-volatile compounds from sources to the receptor site. In this way, model intercomparison can provide useful information on model accuracy and reliability, particularly if model results are also evaluated against measurement data. Thus, the use of more than one receptor modelling approach is a useful validation tool when apportioning sources of the ambient aerosol, revealing model limitations for specific pollutants, spatial scales and applications. Through similar exercises, hybrid models or combined model application may emerge as innovative solutions to reduce uncertainty. In a relevant recent study, three commonly used receptor modelling techniques (PCA, PMF and CMB) were tested for the same PM$_{10}$ data set obtained from an industrial area in Spain (Viana et al., 2008b). The comparisons of the results between the different models and validation against measurement data evidenced a good agreement regarding source identification; however, larger differences were obtained regarding the quantification of source contributions. The analysis of the results suggested that the combined use of factor analysis techniques (PCA, PMF) to identify emission sources and the subsequent application of CMB for more detailed source apportionment and quantification of contributions would provide an appropriate methodology to overcome model limitations.

1.2.5 Current practice in source apportionment in EU countries

A review of methodologies used in source apportionment for regulatory purposes in the EU reveals that receptor modelling is by far the most popular method used by the member states (MSs), whereas a combination of receptor and dispersion modelling is also represented in a lower but significant percentage of MSs. Source apportionment based solely in dispersion modelling was reported for only one MS (Fragkou et al., 2010).

The increased use of dispersion models and of the combination of models for the quantification of natural contributions (as e.g. in the case of Saharan dust) is reflected in the technical reports submitted by EU countries in support of their application for postponement of attaining PM$_{10}$ limit values. As the Directive allows for a 3-year extension of the attainment deadline in case that exceedances can be explained due to contributions from natural sources, adverse climatic conditions and specific dispersion characteristics, the applicant EU countries had to develop a reliable methodology to: (a) confirm that a significant number of exceedances or high annual mean concentrations was due to natural sources, (b) to quantify the proportion of these exceedances and (c) to determine the extent to which the different natural sources were responsible by estimating the PM$_{10}$ concentrations resulting from their relevant emissions.

It is of great interest to note the high percentage of member countries (71%) that have applied a combination of modelling approaches for source apportionment (see table 1.1, adapted from Fragkou et al., 2010). In several cases, Eulerian dispersion models were complemented by Lagrangian trajectory models to account for transboundary contributions in the case of Cyprus, Portugal and Spain (natural transboundary contributions), and Belgium and Austria (anthropogenic transboundary contributions). In Greece and Italy, Eulerian dispersion models have been used to account for transboundary transport of polluted air masses in combination with statistical receptor models for source attribution of both local/national and long-distance sources, while in Slovakia a Gaussian model was used for air quality assessment complemented by a Eulerian Chemical Transport Model (CTM) to assess transboundary contribution. Slovakia and Poland were the only countries to account for resuspension using the EPA emissions modelling approach, which requires input information on traffic characteristics, dust load on the road and street geometry.
### Table 1.1: Modelling tools used for source apportionment by different EU countries for the purposes of preparing the time extension reports.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Number of countries</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagrangian</td>
<td>7</td>
<td>41</td>
</tr>
<tr>
<td>Eulerian</td>
<td>10</td>
<td>59</td>
</tr>
<tr>
<td>Trajectory</td>
<td>7</td>
<td>41</td>
</tr>
<tr>
<td>Receptor</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>Gaussian</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>CFD</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Combination of models</td>
<td>12</td>
<td>71</td>
</tr>
</tbody>
</table>

1 Percentages do not add up to 100%, as many countries used more than one model type

Several countries have verified the model results against available measurements within the frame of the application, while the majority of the models used by the member countries for source apportionment are extensively validated in the literature. In some cases, such as for the United Kingdom, France and Portugal, model validation was explicitly described. The report submitted by the United Kingdom explains the use of a Volatile Correction Model to calibrate and validate the model results by applying appropriate scaling factors prior to the comparison with measurements. Portugal refers to the use of the “Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance” (*ASTM, 2005*) that has been consulted to validate the prognostic meteorological and air pollution TAPM modelling system. Finally, France applied the Eulerian CTM modelling system PREV’AIR to estimate transboundary and natural contributions, including on-line verification procedures.

## 2 Application of 3D chemical transport model (PMCAMx) in the European domain

PMCAMx-2008, a detailed three dimensional chemical transport model (CTM), is applied for the first time in the European domain to simulate the mass concentration and chemical composition of particulate matter (PM) during the periods of May 2008 and February/March 2009. The model includes a state-of-the-art organic aerosol module which is based on the volatility basis set framework treating both primary and secondary organic components to be semivolatile and photochemically reactive. The model shows that much of the traditionally thought as primary organic aerosol (OA) emissions is actually evaporating to produce low-volatility organic vapours which are the source (through photochemical aging) of a substantial amount of oxygenated OA that is distributed not only in urban and suburban areas but rural regions as well.

### 2.1 Introduction

Sulfate and organics are the major components of sub-micrometer particles (PM$_1$) in most locations throughout the world. Organic particulate matter, originating from many different natural and anthropogenic sources and processes, is the least understood component of atmospheric aerosols. More than 50% of the atmospheric fine aerosol mass can be comprised of organic compounds at continental mid-latitudes and as high as 90% in tropical forested areas (*Andreae and Crutzen, 1997; Roberts et al., 2001; Kanakidou et al., 2005*). Organic aerosol (OA) is generally categorized into two types; primary organic aerosol (POA) which is injected into the atmosphere in the particulate phase and secondary organic aerosol (SOA) which is emitted as volatile organic compounds (VOCs) in the gas phase and then reacts and condenses in the particulate phase. In order to gain more insight into sources and processes of organic aerosol, quantification, characterization and
speciation of organic aerosol is needed which was until recently hindered by analytical difficulties (Kanakidou et al., 2005). For instance, conventional techniques (e.g. GC-MS) can only be used for the speciation of a small fraction of the OA mass. Within the last decade, several new methods have emerged that can analyze and quantify the different types of OA present in ambient aerosol. Among several measurement techniques, the Aerosol Mass Spectrometer (AMS) is the most commonly used to measure the size-resolved mass concentration and corresponding total mass spectrum of organic aerosols with a time resolution of minutes (Jayne et al., 2000; Zhang et al., 2005a; Takегawa et al., 2005). Information about processes or sources contributing to the OA levels can be provided from the Positive Matrix Factorization (PMF) method (Paatero and Tapper, 1994; Paatero, 1997; Lanz et al., 2007, 2009; Ng et al., 2009) or the custom principal component analysis (Zhang et al., 2005b) of the AMS measurements. These methods allow a classification of the OA into two different types based on their different temporal and mass spectral signatures; the hydro-carbon-like organic aerosol (HOA) and the oxygenated organic aerosol (OOA) which together usually account for all the OA mass measured by the AMS (Zhang et al., 2005b). HOA correlates with fossil fuel fresh POA in urban areas while OOA correlates with secondary OA. In order to better understand the role of organic aerosol, regional and global air quality and climate models need to be able to correctly predict the concentrations of these different types of OA. Chemical transport models have traditionally treated POA as non-volatile and non-reactive and SOA as semi-volatile products of VOCs’ oxidation (Kanakidou et al., 2005). However, several recent studies have shown that POA is semi-volatile (Lipsky and Robinson, 2006; Robinson et al., 2007; Huffman et al., 2009a, 2009b) while SOA can be seriously underpredicted by models that use traditional SOA precursors and do not take into account further oxidation of these compounds (Baltensperger et al., 2005; Ng et al., 2006; Hildebrandt et al., 2009; Dzepina et al., 2009). Lane et al. (2008a) implemented the volatility basis set framework (Donahue et al., 2006; Stanier et al., 2008) in PMCAMx for the SOA components using four surrogate species with saturation concentrations of 1, 10, 100, and 1000 μg m⁻³ at 300 K, and employed a chemical aging mechanism which shifts mass down by one volatility bin. Shrivastava et al. (2008), using the chemical transport model PMCAMx, reassigned the nonvolatile POA to volatility bins and incorporated the chemical aging of the gas-phase POA by assuming reaction with the hydroxyl radical OH, thus treating POA as semi-volatile and reactive. Tsimpidi et al. (2010) combined the above two approaches for POA and SOA in a single CTM (PMCAMx-2008) and applied it to the polluted area of Mexico City for a week during April 2003. The model was found to reproduce well the AMS measurements of HOA and OOA, however the need for a larger simulation domain was highlighted in order to test whether the regional OA can be predicted with current SOA parameterizations. Murphy and Pandis (2009) applied PMCAMx-2008 in the Eastern United States and evaluated the new modelling framework against daily ambient measurements from the EPA STN (Speciated Trends Network) and the IMPROVE (Interagency Monitoring of Protected Visual Environments) data as well as hourly data from PAQS (Pittsburgh Air Quality Study). Unfortunately, AMS data were not available for the evaluation of the predictions of the model. In this work PMCAMx-2008 is applied for the first time in the European domain to simulate the chemical composition and mass of the major PM₁ components during the months of May 2008 and March 2009 during which intensive campaigns of measurements were performed.

2.2 PMCAMx-2008 description

PMCAMx-2008 (Murphy and Pandis, 2009; Tsimpidi et al., 2010; Karydis et al., 2010) uses the framework of the CAMx air quality model (Environ, 2003) describing the processes of horizontal and vertical advection, horizontal and vertical dispersion, wet and dry deposition, and gas-phase chemistry. For the aerosol processes, three detailed aerosol modules are used. The approach of Fahey and Pandis (2001) is used for the simulation of aqueous-phase chemistry. The inorganic aerosol growth is described in Gaydos et al., (2003) and Koo et al., (2003), and the secondary
organic aerosol (SOA) formation and growth in Koo et al. (2003). These aerosol modules use a sectional approach to dynamically track the size evolution of the aerosol mass across 10 size sections spanning from 40 nm to 40 μm. The aerosol species modelled include sulfate, nitrate, ammonium, sodium, chloride, potassium, calcium, magnesium, elemental carbon, primary and secondary organics. The chemical mechanism used in the gas-phase chemistry is based on the SAPRC99 mechanism (Carter, 2000; Environ, 2003). The mechanism includes 211 reactions of 56 gases and 18 radicals.

Three options are available in PMCAMx-2008 for the simulation of inorganic aerosol growth. The most computationally efficient approach is the bulk equilibrium approach, which assumes equilibrium between the bulk inorganic aerosol and gas phase. At a given time step the amount of each species transferred between the gas and aerosol phases is determined by applying the aerosol thermodynamic equilibrium model ISORROPIA (Nenes et al., 1998) and is then distributed over the aerosol size sections by using weighting factors for each size section based on their surface area (Pandis et al., 1993). The second approach (hybrid approach) assumes equilibrium for the fine particles (<1 μm) and solves the mass transfer differential equations for the coarse particles (Caspaldo et al., 2000). The most accurate but computationally demanding method is the dynamic approach where mass transfer is simulated explicitly for all particles (Pilinis et al., 2000). In this work we use the bulk equilibrium approach since we are focusing on the model’s performance for fine particles.

Based on the approach of Shrivastava et al. (2008), primary organic aerosol in PMCAMx-2008 is assumed to be semivolatile. Nine surrogate POA species are used, with effective saturation concentrations at 298 K ranging from $10^{-2}$ to $10^{6}$ μg m$^{-3}$. POA is simulated in the model in three types, “fresh” (unoxidized) POA, oxidized POA (OPOA) and any aged organic aerosol transported into the domain from the boundary conditions. SOA consists of organic aerosol of anthropogenic (ASOA) and biogenic (BSOA) origin. The SOA volatility basis-set approach (Lane et al., 2008a) used in the current version of the model includes four SOA species for each VOC with 4 volatility bins (1, 10, 100, 1000 μg m$^{-3}$). Laboratory results from recent smog-chamber experiments (Hildebrandt et al., 2009; Ng et al., 2006) have been used to update the SOA module in PMCAMx in order to include anthropogenic aerosol yields. The SOA module incorporates NOx-dependent SOA yields (Lane et al., 2008b) which are based on an assumed density of 1.5 g cm$^{-3}$. The model treats all organic species (primary and secondary) as chemically reactive. Chemical aging through gas-phase oxidation of OA vapours is modelled using a gas-phase OH reaction with a rate constant of $4 \times 10^{11}$ cm$^3$ molec$^{-1}$ s$^{-1}$ (Atkinson and Arey, 2003). Each reaction is assumed to decrease the volatility of the vapour material by a factor of 10.

### 2.3 Model application for MEGAPOLI

The PMCAMx-2008 modelling domain covers a 5400 × 5832 km$^2$ region in Europe with 36 × 36 km grid resolution and 14 vertical layers covering approximately 6 km. PMCAMx-2008 was set to perform simulations on a rotated polar stereographic map projection. The first two days of each simulation were excluded from the analysis to limit the effect of the initial conditions on the results. Concentrations of the major PM$_1$ species at the boundaries of the domain are shown in table 2.1 (for the period of May 2008), representing background concentrations with each domain side having different boundary conditions. The boundary condition organic aerosol (BC-OA) is expected to consist of both SOA and oxidized POA. Here we assume that the BC-OA is all oxidized and half of it is biogenic OA and the other half oxidized primary OA (Kanakidou et al., 2005; Farina et al., 2010). All concentrations reported here are under ambient temperature and pressure conditions.

The necessary inputs to the model include horizontal wind components, vertical diffusivity, temperature, pressure, water vapour, clouds and rainfall. The meteorological model WRF (Weather Research and Forecasting; Skamarock et al., 2005) was used to create the above inputs. WRF was driven by static geographical data and dynamic meteorological data (near real-time and historical
data generated by the Global Forecast System (1×1 deg)). 27 sigma-p layers up to 0.1 bar were used in the vertical dimension. Each layer of PMCAMx-2008 is aligned with the layers used in WRF. The WRF May 2008 run was periodically re-initialized (every 3 days) to ensure accuracy in the corresponding fields that are used as inputs in PMCAMx-2008.

Table 2.1: Aerosol concentrations (in μg m⁻³) at the boundaries of the domain.

<table>
<thead>
<tr>
<th>Species</th>
<th>Boundary</th>
<th>North</th>
<th>South</th>
<th>West</th>
<th>East</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sulfate</td>
<td>1</td>
<td>0.8</td>
<td>0.2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ammonium</td>
<td>0.4</td>
<td>0.08</td>
<td>0.05</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Nitrate</td>
<td>0.1</td>
<td>0.02</td>
<td>0.01</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>0.001</td>
<td>0.005</td>
<td>0.09</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Chloride</td>
<td>0.002</td>
<td>0.01</td>
<td>0.1</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Anthropogenic and biogenic hourly emission gridded fields were developed for the European domain for gases and primary particulate matter. Volatile organic compounds are split based on the SAPRC 99 chemical mechanism. Anthropogenic gas emissions that were used to develop the gridded fields include land emissions from the GEMS dataset (Visschedijk et al., 2007) as well as international shipping emissions. Anthropogenic particulate matter mass emissions of organic and elemental carbon are based on the Pan-European Carbonaceous Aerosol Inventory. A variety of emission sources are identified in the two inventories, including industrial, domestic, agricultural and traffic. Three different datasets are combined in order to produce the biogenic gridded emissions for the model. Emissions from ecosystems are produced by MEGAN (Model of Emissions of Gases and Aerosols from Nature) (Guenther et al., 2006). MEGAN inputs include the leaf area index, the plant functional type and the emission factors while the weather data needed for MEGAN are provided from the WRF model. Since sea surface covers a considerable portion of the domain, the marine aerosol model developed by O’Dowd et al. (2008) has been used to estimate mass fluxes for both accumulation and coarse mode including the organic aerosol fraction. Wind speed data from WRF and chlorophyll-a concentrations are the inputs needed for the marine aerosol model. Wildfire emissions from May 2008 were also included. Table 2.2 shows a summary of the emission masses (in ktons/month) from the different sources for the European domain during May 2008. The OA emissions in PMCAMx-2008 were distributed by volatility using the volatility distributions of Tsimpidi et al. (2010).

Table 2.2: Emission mass totals (in ktons/month) for the European domain during May 2008.

<table>
<thead>
<tr>
<th>Species</th>
<th>CO</th>
<th>NOₓ</th>
<th>SO₂</th>
<th>NH₃</th>
<th>NMVOC</th>
<th>NOₓ⁺</th>
<th>Sulfate</th>
<th>NH₄⁺</th>
<th>EC</th>
<th>OC</th>
<th>Na⁺</th>
<th>Cl⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropogenic (MEGAN)</td>
<td>3682</td>
<td>1461</td>
<td>1675</td>
<td>501</td>
<td>1291</td>
<td></td>
<td></td>
<td>59</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biogenic (MEGAN)</td>
<td>439</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td>2879</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biogenic (Sea salt)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1294</td>
<td>43</td>
<td>4951</td>
<td>8900</td>
<td></td>
</tr>
<tr>
<td>Biogenic (Fires)</td>
<td>768</td>
<td>22</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>16</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4 Results and discussion

Figure 2.1 shows the PMCAMx-2008 average ground-level concentrations for PM$_1$ “fresh” primary OA, oxidized POA, anthropogenic SOA, biogenic SOA, elemental carbon and total mass of organic aerosol, during May 2008 while figure 2.2 shows the predicted average fractional contribution of each one of the OA components to total OA. Fresh POA concentrations are predicted to be relatively high (up to 8 $\mu$g m$^{-3}$) in various urban and suburban areas mainly in central and Northern Europe and drop dramatically (less than 0.2 $\mu$g m$^{-3}$) in the rest of the domain. Less than 7% of total PM$_1$ OA is predicted to consist of fresh POA on average over the entire domain with a maximum of 71% in the area of St. Petersburg, northwest of Russia. OPOA concentrations are predicted to be higher than the fresh POA in the modelling domain. Oxidized POA is predicted to be the dominant OA component during this period with an average contribution of 54% to the total predicted OA over the domain while almost 60% of the total oxygenated OA consists of OPOA. Anthropogenic SOA concentrations are predicted to be generally low in the modelling domain (less than 0.5 $\mu$g m$^{-3}$ in most areas). On average biogenic SOA contributes approximately one third of the total oxygenated OA (sum of OPOA, SOA and BC-OA) in the modelling domain. The model predicts an almost uniform distribution of aged organic aerosol transported into the domain from the boundaries (~0.5 $\mu$g m$^{-3}$). An average of 0.2 $\mu$g m$^{-3}$ is predicted for PM$_1$ EC concentrations in the domain, although much higher values are predicted locally (up to 4 $\mu$g m$^{-3}$) in urban areas. Among the PM$_1$ species, organic matter is the dominant PM$_1$ species in central and northern Europe with a peak value of 10 $\mu$g m$^{-3}$ (in north-western Russia). The average monthly contribution of OA to total PM$_1$ mass was predicted to range between 50 and 60% in central and northern Europe and drops to 20-30% at the rest of the domain.
Figure 2.1: Ground-level concentration predictions averaged over the entire simulation period (1-29 May 2008) for PM$_1$ (a) “fresh” POA, (b) oxidized POA, (c) anthropogenic SOA, (d) biogenic SOA, (e) elemental carbon and (f) total organic aerosol (in μg m$^{-3}$). Different scales are used.

The Mediterranean region is the only area in Europe where sulfate concentrations are predicted to be much higher than the OA values among the PM$_1$ species. A peak value is predicted in the area
south of Malta where sulfate concentration is more than 2 times larger than the OA value. In central and northern Europe however, PMCAMx-2008 predicts the opposite with an average OA to sulfate mass ratio of 2.5 and a maximum at the Paris area in France where the PM$_1$ OA mass is predicted to be 4 times greater than the PM$_1$ sulfate levels. During March 2009 (figures 2.3 and 2.4), PMCAMx-2008 predicts a higher contribution of fresh POA to total PM$_1$ OA (approximately 12%) compared to the May period, on average over the entire domain. Oxidized POA is predicted to be the dominant OA component during this wintertime period with an average contribution of 46% to the total predicted PM$_1$ OA.

Figure 2.2 : Predicted average ground-level fractional contribution of a) anthropogenic SOA, b) biogenic SOA, c) oxidized POA, and, d) fresh primary POA to PM$_1$ total OA during May 2008. Different scales are used.

The modelling results can be used further to explore the predicted contributions of various sources to total organic aerosol concentrations at specific cities. Figures 2.5 and 2.6 show the average diurnal concentration profile of all the OA components as predicted by PMCAMx-2008 at 5 Mega-cities in Europe, during May 2008 and March 2009 respectively. During May 2008, the model predicts a relatively flat average diurnal profile in London, Athens, Po Valley and Ruhr area, and a strong diurnal variation of OA concentrations in Paris. The predicted source contributions to OA in Paris show a strong diurnal variation of fresh POA with increases during the rush hours and relatively flat profiles for the other OA components. Fresh POA is predicted to be the dominant OA component in Paris with an average contribution to total OA of more than 50%. During the winter period a more distinct diurnal variation is predicted in most of the cities following the diurnal variation of emission rates mainly due to combustion processes.

The variety of trends indicates that conclusions made about the organic aerosol diurnal profile in Paris cannot be extrapolated to all (urban) sites in the domain. In general, the relatively flat diurnal profiles that the model tends to predict in many areas, is surprising considering the expected diurnal
variations due to emissions, photochemical activity and temperature changes and is in disagreement with older modelling efforts (Pun et al., 2003). These results indicate that the average diurnal profile of the OA concentrations can be used, if there are corresponding available measurements, as a tool for the evaluation of the performance of CTMs given the uncertainties in day to day OA and VOC emissions.

**Figure 2.3**: Ground-level concentration predictions averaged over the entire simulation period (February 25th – March 24th, 2009) for PM$_1$ (a) “fresh” POA, (b) oxidized POA, (c) anthropogenic SOA, (d) biogenic SOA, (e) elemental carbon and (f) total organic aerosol (in µg m$^{-3}$). Different scales are used.
Figure 2.4: Predicted average ground-level fractional contribution of a) anthropogenic SOA, b) biogenic SOA, c) oxidized POA, and, d) fresh primary POA to PM$_1$ total OA during March 2009. Different scales are used.

Diurnal PM$_1$ OA distribution in Paris during May 2008
Figure 2.5: Average diurnal profile of PM$_1$ OA components in 5 European Megacities during May 2008. Different scales are used.
3 Analysis of Paris 2009 Campaign Data using Receptor Modelling

3.1 Theoretical background

The fundamental principle of receptor modelling is that mass conservation is assumed and a mass balance analysis can be used to determine and apportion ambient pollutant concentrations to individual emitting sources. Receptor models represent a statistical evaluation of ambient measurements at different times and locations, thus the selection of the appropriate method depends on prior knowledge on the sources and source profiles. If the sources are known and detailed information on source profiles is available, CMB models can be applied, whereas in case the sources are unknown and there is limited information on source profiles, PCA and PMF or NMF methods are preferred. The following sections contain a brief description of the PCA and NMF methods.

3.1.1 Principal Component Analysis (PCA)

PCA is often used as an exploratory tool to identify the major sources of air pollutant emissions and to select statistically independent source tracers. The PCA/APCS (Thurston & Spengler, 1985) approach presents a source apportionment technique which requires a minimum of inputs regarding source characteristics, but provides quantitative information regarding both source profiles and their impacts.

PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or the singular value decomposition of a data matrix, usually after mean centering the data. This statistical technique can be applied to a set of variables in order to reduce their dimensionality. That is, to replace a large set of intercorrelated variables with a smaller number of independent variables. These new variables (components) are derived from the original variables, and are simply linear combinations of those variables.

If the original variables are correlated with each other, a large portion of the variance in the dataset can be explained with a small number of factors. Note that some non-material variables such as wind speed and solar radiation can be included as factors (Chan & Mozurkewich, 2007). A technique such as PCA does not directly identify sources; it only identifies groups of correlated variables. This method produces loadings and scores. Loadings are composed of the Principal Components (PCs) which are lineal combinations of the original variables, while the scores contain information about the projections onto the PCs, instead of the original variables.

One of the downsides of this method is the lack of constraints on the values of either the loadings or scores, besides the requirement of the resulting components to be orthogonal. This can result in
negative loadings and scores, namely giving physically unreasonable output. However this can be moderated after applying the Varimax rotation. Varimax rotation is a change of coordinates used in principal component analysis that maximizes the sum of the variance of the squared loadings. It was suggested by Kaiser (1958) and is a popular scheme for orthogonal rotation. The main idea of the rotation is that "for each factor, high loadings (correlations) will result for a few variables; the rest will be near zero." After applying the rotation, it is typical that for each component all scores (the amounts of the component present) that are not near zero have the same sign; these can be chosen to be positive.

3.1.2 Nonnegative Matrix Factorisation (NMF)

The NMF model developed by Paatero and Tapper (1994) is a method based on solving the factor analysis problem by least squares using a data point weighing method, which uses estimates of the data uncertainties to provide optimum data point scaling. The solution is forced to be nonnegative through the use of a penalty function.

The factor analysis model is given by:

$$X = GF + E,$$

where $X$ is a matrix of observed data, $G$ is the unknown left-hand factor matrix, $F$ is the unknown right-hand factor matrix, and $E$ is a matrix of residuals.

The difference between the model and the data is:

$$E = X - GF$$

The least-squares problem can then be transformed into a minimization problem:

$$\min_{G,F} Q(X, \sigma, G, F)$$

where

$$Q = \left( \frac{(X - GF)}{\sigma} \right)_{F,G}^2 = \sum_i \sum_j \left( \frac{e_{ij}}{\sigma_{ij}} \right)^2, \quad e_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}$$

with $\sigma$ a known estimated uncertainty for each $x_{ij}$ value.

One important advantage of the NMF method is the ability to handle missing data and data below detection limits by adjusting the corresponding error estimates of these data points. These problematic data are usually replaced by mean concentration and half of the detection limit values, respectively, and large error estimates are being used for such values. For the determined values, one should adjust the effective error estimates so that they include all errors that produce a deviation between the fit and the observed value.

Usually, the final set of values of the error estimates are obtained by running the NMF program many times with different sets of error estimates and applying the trial and error method until a substantially improved fit to the data is achieved. The non negativity requirement in NMF analysis decreases the rotational freedom, and in some cases, the result is unique with no rotational freedom. (Polissar et al., 2001)

3.1.3 Caveats and limitations

In receptor modelling approaches like PCA and PMF, determining the number of sources contributing to urban pollutant concentrations and their interpretation is not straightforward. In applying these approaches more than one valid solution can usually be obtained, and no definitive, objective
or statistical indicator can be calculated to indicate which of these solutions (e.g. the number of components or sources the original data can be reduced to) best represents reality. Furthermore, any such objective output regarding an optimum solution must be interpreted in light of uncertainties in the measurement data and variations in environmental conditions, such as meteorology and chemical processing in the atmosphere and variations in actual emission characteristics (Lee et al., 2003). Therefore, determining the number and nature of sources that explain the observed concentrations, inherently involves some amount of subjectivity and requires experience with respect to application and interpretation of the method in use and the characteristics of the domain of interest (e.g., general knowledge or inventory-based information of the types of local and regional sources expected to contribute to each pollutant concentration). Alternatively, modelling techniques such as the one proposed here may be used to provide aid to the user in determining the number and interpretation of sources.

3.2 The Paris Measurement Network and the 2009 Campaign

Within the framework of MEGAPOLI WP3, a measurement campaign was carried out in order to improve the quantification of primary and secondary organic aerosol sources, and their relation to gaseous precursors for the Greater Paris area. Paris is the capital and largest city in France at the heart of the Île-de-France region. It is located in northern France by the river Seine, in a relatively flat area that contains some hills, of which the highest is at 195 metres above sea level. The city of Paris is one of the most populated urban areas in Europe, with an estimated population of 2.2 millions and a population density of 20807 /km² within its administrative limits. The Greater Paris area has a population of around 12 millions with population density of 815.3 /km². With central latitude at 48.867 degrees north and central longitude at 2.333 degrees east, Paris has a typical Western European oceanic climate, which is a mild and moderately wet climate that is very often affected by the North Atlantic Current.

The Paris air quality monitoring network, AIRPARIF, consists of 65 stations and provides real time information of air pollution levels in the Greater Paris area. Specifically in regard to particulate pollution, concentrations of particulate matter with aerodynamic diameter smaller than 2.5 μm (PM$_{2.5}$) have been continuously measured for the past 10 years, using a conventional on-line automatic system. The analysis of these measurements for all stations shows that the yearly mean concentration of PM$_{2.5}$ had a stable value of $\sim15 \, \mu g/m^3$ for the period between 2000 and 2006. However, after 2007, new measurement hardware was installed and used that allowed a better determination of the semi-volatile species, which indicated an annual PM$_{2.5}$ mean value that was higher by 10 $\mu g/m^3$, reaching the limit of 25 $\mu g/m^3$ as defined in the EU Directive 2008/50/EC.
Therefore, accurate and detailed source apportionment of the particulate pollution is required that will guide effective air quality strategies in order to reduce pollution levels.

During the MEGAPOLI campaign three additional monitoring sites, one urban and two sub-urban, were set up. The urban station was located at the Laboratoire d’Hygiène de Paris (LHPV), in the 13th District of Paris and one of the sub-urban stations was at Golf de la Poudrière in the north-east region of the city. But for the purpose of this work package the measurements of the SIRTA station were used and its summer campaign. The SIRTA sub-urban measurement site was located in the south-west region of the Greater Paris area, with co-ordinates 2.2025 degrees east and 48.714167 degrees north. The station was equipped with Aerosol Mass Spectrometer (AMS) for PM$_1$ speciation, filters for the chemical analysis of PM$_{2.5}$ organic species, aethalometer for aerosol measurements, but it also included instruments for measuring inorganic salts and many gaseous species. The values of basic meteorological parameters were also recorded. A brief description of the equipment, species and time resolution of the summer campaign is given at Table 3.1.

Table 3.1: Brief description of the summer measurement campaign

<table>
<thead>
<tr>
<th>Site</th>
<th>Instrument</th>
<th>description</th>
<th>Start date</th>
<th>Stop date</th>
<th>Time resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIRTA</td>
<td>AMS</td>
<td>size resolved chemistry in PM$_1$ OA, SO$_4$, NO$_3$, NH$_4$, Cl</td>
<td>30/6/2009 8:45</td>
<td>31/7/2009 12:22</td>
<td>10 min</td>
</tr>
<tr>
<td>Filter (DA80 in PM$_{2.5}$)</td>
<td>Organic tracers in PM$_{2.5}$ (OC, EC, TC, Cl, NO$_3$, Mali, SO$_4$, Ox, Na, NH$_4$, K, Mg, Ca, Levo)</td>
<td>1/7/2009 7:00</td>
<td>30/7/2009 7:00</td>
<td>~10 hours</td>
<td></td>
</tr>
<tr>
<td>PILS-IC</td>
<td>Inorganic Salts (chloride, sulfate, nitrate, oxalate)</td>
<td>2/7/2009 1:34</td>
<td>1/8/2009 10:57</td>
<td>8 min</td>
<td></td>
</tr>
<tr>
<td>Aethalometer 7-L</td>
<td>Absorption coeff (BC)</td>
<td>2/7/2009 19:10</td>
<td>31/7/2009 2:58</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>IR analyser</td>
<td>CO</td>
<td>1/7/2009 00:00</td>
<td></td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>AC31M, UV analyser</td>
<td>O$_3$, NO, NO$_2$, NO$_x$</td>
<td>23/6/2009 15:22</td>
<td>31/7/2009 23:59</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>HPLC-UV</td>
<td>HONO</td>
<td>9/7/2009 14:33</td>
<td>30/7/2009 6:40</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>Luminol Chemiluminescence</td>
<td>NO$_2$</td>
<td>1/7/2009 7:21</td>
<td>30/7/2009 7:10</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>GC-ECD</td>
<td>PAN</td>
<td>1/7/2009 00:00</td>
<td>30/7/2009 23:45</td>
<td>15 min</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weather data</td>
<td>1/7/2009 00:00</td>
<td>31/7/2009 23:00</td>
<td>1 hour</td>
<td></td>
</tr>
</tbody>
</table>

3.3 Measurement analysis using PCA

The PCA methodology was applied to measurements conducted at the SIRTA station for the period between July 1st and July 31st 2009. A total of 20 gas and particle species, measured by the different instruments listed in table 3.1, were selected for the analysis, namely O$_3$, NO, NO$_2$, CO, SO$_4$, NO$_3$, black carbon (BC), benzene, HONO, NH$_4$, peroxyacetyl nitrate (PAN), organic PM, chloride, oxalate, iso-butane, iso-pentane, isoprene, n-butane, n-pentane and hexane.
PCA offers a degree of flexibility by allowing the user to select the number of “factors” (also known as “components”) upon which the dataset will be projected. For all subsequent analyses a total of 6 factors have been selected, representing a balance between accounting for a sufficient part of the dataset’s total variance and at the same time limiting the number of factors so that only meaningful correlations between pollutants are retained.

Analysis results were obtained both without and with a Varimax rotation (Kaiser, 1958) being applied to the loading matrix. In figure 3.2, the loadings of the primary gas species over the first few principal components, corresponding to the relative mass contribution of each pollutant together with the scores of each factor as a function of time, are shown. The percentage figures represent the fraction of total variance in the sample explained by each factor.

The first component, with strong correlations of NO, NO₂, NO₃, BC and VOCs clearly represents contributions from traffic emissions while the second component accumulates almost all of the ozone contribution and can be related with the evolution of secondary gas species. The third component accounts for a marginal 10% of total variance and contains a strong contribution of SO₄ and NH₄, suggesting a correlation with secondary inorganic aerosols. The rest of the components account to much less than 10% of total variance and cannot be easily identified with specific source categories.

Varimax rotation induces a negligible redistribution of the loadings in the first two components, as shown in figure 3.3; however CO contribution is now mainly concentrated in the third component.
3.4 Measurement analysis using NMF

Compared to the results of the PCA, the Nonnegative Matrix Factorisation analysis for the SIRTA measurements produces a more detailed picture of the various source contributions. For example, even in the case of unrotated loadings (Figure 3.4, left) it is obvious that the first factor encompasses much of the traffic contributions, as evident by the strong correlations of CO, primary NOx and VOCs. The second component accumulates a large contribution of SO4 and NH4, also including a significant part of O3 indicating correlation with secondary pollutants and having an increased variance of 23%. The third component corresponds to the “ozone” component of the PCA results, nevertheless O3 correlations are split between the second, third and fifth components. Finally, from the temporal evolution of component scores (figure 3.4, right) it can be concluded that factors 5 and 6 represent possible transient effects which cannot be easily identified with specific source types.

The introduction of Varimax rotation again helps to disambiguate the correlations of major pollutant species in the first few factors, while at the same time increases the variance accounted for by the same factors (Figure 3.5). The “traffic” factor now accounts for 52% of total variance instead of 26% of the unrotated case, while the “transient” factors 5 and 6 are eliminated. On the other hand, the CO contribution is now split between the first “traffic” factor and a new “CO factor” which accounts for 20% of total variance. The appearance of this new factor delegates the main O3 contribution towards the third factor, which has a similar constitution with the corresponding factor of the unrotated case.
In conclusion, the analysis results from unrotated NMF appear to provide the clearer picture, although the interpretation of higher order factors is generally difficult. It also appears that although the SIRTA location does not lie in the vicinity of major traffic emission sources, it nevertheless is strongly influenced by traffic plumes. To a lesser extent, the SIRTA site is also influenced by secondary pollutants which are formed in the urban area of Paris.

3.5 PMF using other tools

The United States Environmental Protection Agency (EPA) provides a software product that applies PMF for evaluation purposes. The EPA PMF 3.0 is using the multilinear engine version 2 (ME2) to converge the object function Q to a minimum value (Paatero and Tapper, 1994; Paatero, 1999). Additional rotation to the base model results can be applied to achieve best fitting results. This is achieved by entering a user-specified parameter \( F_{peak} \) that pulls the object function Q to extreme values, zero or unity. Also the factor profile matrix can be pulled to extreme values by the use of a control matrix called Fkey (Lee et al., 1999).

Many studies had been published using PMF for source apportionment. The variety of locations and their sources used in the past shows that there are no serious limitation in the use of the method by local conditions. For example, PMF had been applied for PM2.5 at Seoul (Heo et al., 2008) and Toronto (Lee et al., 2003), PM10 for Bogota (Vargas et al., 2009) for PM1 for Italian cities and many more cases varying in topography, emissions, population density and meteorology.

The application of PMF for the Paris data has been carried out in order to evaluate the source apportionment method discussed in previous sections. To tone with the other methods and because of the PMF results led there, the six factors solution is being used. The uncertainty values were calculated by a built-in equation. Inputs of the equation-based calculation of the uncertainties were the detection limit of the species and the error fraction. Low error fraction was used for all species. Furthermore, additional uncertainty was used for every chemical species individually and for the whole case. According to the guidelines of user guide (Norris et al., 2008), and after several tries the residual analysis of the results led to the increase of the uncertainty for several species (Black carbon, benzene, hexane, nitrous acid, isoprene and oxalate) and the overall modelling uncertainty. The explanation of the results demands expertise in the specific field in order to specify the number of factor needed and the meaning of its factor. The base model results are shown in Figure 3.6 and the results after the Fkey rotation in Figure 3.7. Red squares refer at the left vertical axis which designates the percentage the particular species that is contributing to the specific factor. The coloured bar indicates the mass concentration in logarithmic scale (right vertical axis).

The first factor is the secondary sulfate factor and is dominated by sulfate (SO₄) and ammonium. Several studies stated that this factor has a seasonal trend with a summer peak suggesting that the formation of secondary sulfate was enhanced by the increased photochemical reactivity during the summer. The presence of black carbon was expected in a factor of sulfates. This factor is a major mass contributor to Particulate Matter. The second factor it is dominated mainly by oxalate. With the presence of isoprene, PAN and organics strongly indicates factor correlated with biogenic emissions. The third factor is the secondary nitrates. It is another factor of the secondary pollutants which the major contributor is nitrates, with the presence of ammonium. Black carbon and sulfate are present in this factor, without the seasonal trend of the latter as in the first factor. The contribution of chloride reveals a marine origin. The fourth factor can be characterized as the ozone factor. Ozone is scattered among other factors also so its contribution to this factor is not as massive as expected but still is major. The fifth factor is traffic. The major contributor is nitric oxide with the presence of nitric dioxide and carbon monoxide. The complete absence of ozone fits with characterization of the factor. The last factor is mainly composed by nitrous acid. Also there is a good correlation with NO₂ and NO₃ suggesting that this factor is describing the NOx cycle. Further analysis could be performed at the above base results. Additional rotation was applied with the use of Fkey matrix. The main purpose was to contain all the carbon monoxide, a typical traffic
pollutant, to a single factor, the traffic one. After the application, the main contributors of each factor were not influenced, not altering the main characterization of its factor. Furthermore, the correlations stated above were enhanced, like the increase of NO in the traffic factor or the increase of ozone in the fourth factor, which strengthen the assumptions made before, as well as, the need for the additional rotation.

**Figure 3.6**: Factor profiles produced by the base PMF model.
Figure 3.7: Factor profiles produced by the PMF model after Fkey rotation.
4. Use of Dispersion Modelling for Supporting Receptor-based Source Apportionment

A series of model simulations were performed for the summer 2009 Paris case. This application had three main targets: the first was to undertake a consistent determination of the main emission sources which are affecting the pollutant concentration fields in the study area. The second target was to assess the performance of receptor-based models in analyzing concentration data both provided by dispersion models as well as obtained from measurements. A third target was to present a common framework for the application of receptor-based and dispersion models, in order to provide support in the analysis of observational data for source apportionment purposes.

4.1 Simulations of Pollutant Dispersion in the Greater Paris Area

The non-hydrostatic meteorological mesoscale model MEMO (Moussiopoulos, 1995; Kunz and Moussiopoulos, 1997) was used in order to provide an approximation of the wind flow during the simulation period as well as to determine the necessary meteorological input for the MARS-aero dispersion and chemical transformation model (Moussiopoulos et al., 1997; Arvanitis et al., 2001). Two concentric computational grids were defined in a nested configuration: a coarse grid with an extent of 300×300 km² and resolution of 5 km and a fine grid extending to 50×50 km² with a resolution of 500 m, covering the central areas of Paris (figure 4.1, table 4.1). Initial and lateral boundary conditions for the outer grid were obtained from meteorological sounding data.

![Figure 4.1](image1)

**Table 4.1**: MEMO and MARS-aero computational domains for the Paris case

<table>
<thead>
<tr>
<th>Domain</th>
<th>Dimensions (number of cells)</th>
<th>Resolution (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MEMO</td>
<td>MARS-aero</td>
</tr>
<tr>
<td>Coarse</td>
<td>60×60</td>
<td>-</td>
</tr>
<tr>
<td>Fine</td>
<td>50×50</td>
<td></td>
</tr>
</tbody>
</table>

Dispersion modelling offers the possibility to quantify the effect of reduced sectoral or source-specific emissions on the calculated concentration fields. In zero-out modelling, emissions from a
particular source or source category are reduced by 100%, thus quantifying the change in the pollutant concentration that would occur if the relevant source(s) were removed. Figure 4.2 illustrates the reduction on PM$_{10}$ concentrations in the Paris 2009 case caused by zeroing-out the traffic emissions (SNAP sector 7). In the left part of the same figure the subsequent reduction on local ozone levels, due to a reduction in NO$_x$ titration, is also illustrated.

![Figure 4.2](image)

*Figure 4.2: Concentration difference map for O$_3$ (left) and PM$_{10}$ (right) between the baseline and the zero-traffic emission cases.*

### 4.2 PCA based on zero-out modelling

Concentration time-series calculated using the MARS-aero model for a period of three days, have been analysed by means of the Principal Component Analysis method. A total of ten gas species together with an aggregate PM$_{10}$ species were considered during the analysis of model data. As in previous applications, a total of six components were selected for each analysis.

*Table 4.2: AIRPARIF measurement locations used in the analysis of the dispersion model results.*

<table>
<thead>
<tr>
<th>STATION</th>
<th>STATION EUROPEAN CODE</th>
<th>LONGITUDE (degrees)</th>
<th>LATITUDE (degrees)</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aubervilliers</td>
<td>FR04018</td>
<td>2.385278</td>
<td>48.903611</td>
<td>urban background</td>
</tr>
<tr>
<td>Versailles</td>
<td>FR04029</td>
<td>2.130555</td>
<td>48.799168</td>
<td>urban background</td>
</tr>
</tbody>
</table>

The PCA analysis was applied in time-series of ground-level pollutant concentrations calculated for three locations in the Paris Greater Area, corresponding to the locations of two AIRPARIF background stations (Table 4.2) and the SIRTA site of the 2009 measurement campaign. For this analysis, eleven pollutants species were considered, namely ozone, NO, NO$_2$, CO, SO$_2$, NO$_x$, CH$_4$, toluene, formaldehyde, organic nitrate and aggregate PM$_{10}$. In Figure 4.3 the component loadings, corresponding to the relative mass contribution of each pollutant in the 6 factors is shown, together with the scores of each factor as a function of time.

The first component accounts for almost 50% of total variance and is dominated by contributions characteristic of traffic emissions: positively correlated NO, NO$_2$, CO and SO$_2$ combined with a strong presence of PM$_{10}$. The corresponding time-series of scores for this component indicates a strong similarity with the bimodal diurnal pattern of traffic emissions (as used in the model calcula-
tions), which is characterised by a large peak between 08:00 and 09:00 and a smaller secondary peak around 20:00. The second component includes strong positive correlations of ozone and NO$_3$ while the typical diurnal profile of the corresponding scores indicates a consistent maximum in the afternoon hours which in turn is compatible with the formation of secondary pollutants. From the remaining four components accounting for less than 10% each, we can note the strong contribution of methane in the third component.

Comparing the PCA loadings for the SIRTA site that were produced with the aid of the dispersion model (Figure 4.3) and the ones that were produced using the measurement data (Figure 3.2), some visible similarities can be noted, including the accumulation of typical traffic-related pollutants such as NO, NO$_2$ and CO in the first component as well as the robust presence of O$_3$ in the second component. Apart from the differences in speciation, the apparent correspondence between the main components of both cases shows that PCA provides a reasonable representation of atmospheric composition at specific sites. This correspondence will allow us to interpret and validate characteristics of the measurement analysis through appropriate treatment of the dispersion modelling results.

Results of a similar analysis of model results, this time calculated with zero traffic emissions (SNAP sector 7) are shown in Figure 4.4. In the case of the Aubervilliers station, the exclusion of traffic emissions from the dispersion model induces a mixing effect on the first component: species from the “traffic” component of the full-emission case (NO, NO$_2$, SO$_2$) still appear positively correlated in the new first component but they are mixed with negatively correlated species (O$_3$, NO$_3$) originally appearing in the “secondary” component of the full emission case. In the case of the SIRTA and Versailles stations, the masking of traffic emissions significantly reduces the contribution of the suspected “traffic” component (from 58% and 61% to 48% and 53%, respectively) without altering the pollutant loadings in the first two components, thus rendering evident the validity of our initial assumption regarding the characterisation of the principal components. As a general remark, the omission of traffic sources seems to reduce the model’s ability to resolve a consistent mixture of components for all stations. By observing e.g. the time evolution of the 1st component’s scores in the case of the SIRTA station (Figure 4.4), it is evident that a physical interpretation even for the most strongly-correlated components becomes difficult in the zeroed-out emission case.

Nevertheless, the picture that emerges from the above results is that the analysis of dispersion modelling results from runs, performed with zeroed-out emissions of specific emission sectors, provides a more robust guide for validating initial assumptions of factor characterisations compared to analyses based solely on measurements. This is important as it can be applied based on relatively short and computationally inexpensive model runs in order to provide valuable insight into measurement analysis. Of course, the method outlined above could not be easily extended to cases of factors that are not associated with primary emissions. In the majority of cases, however, it remains a powerful tool for the identification of the majority of factors, which by itself can usually lead to a clearer apportionment of source types.

A similar analysis was carried out using the same set of calculated time-series, this time including a rotation step for the 6 principal components, by means of the Varimax algorithm. The rotation tends to maximise projections of eigenvector components along the coordinate directions, hence it is expected that the loadings matrix will be dominated by strong correlations for a small number of species.

As evident in Figure 4.5, in the case of the first (“traffic”) component, the rotation emphasizes the contribution of CO while the main contributions of NO and NO$_2$ are split in several components. In the case of the Aubervilliers station, it is also notable that the time profile of component 3 scores follows the profile of component 1 within a ~2 hour delay, suggesting a strong correlation with NO
and NO₂ concentrations, respectively. No such effect can be discerned in the case of SIRTA and Versailles, mainly due to the much reduced improvement caused by the Varimax rotation in the component projections at these stations.

**Figure 4.3:** PCA scores and unrotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the full emissions scenario.
Figure 4.4: PCA scores and unrotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the zero-traffic emissions scenario.
Figure 4.5: PCA scores and Varimax-rotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the full emissions scenario.
Figure 4.6: PCA scores and Varimax-rotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the zero-traffic emissions scenario.

Using the same rotation methodology, results from the zero-traffic emission case were analysed. A systematic reduction on the power of the first principal components can be observed while the contributions of traffic-related time profiles can no longer be associated with a single component.
4.3 Nonnegative Matrix Factorisation (NMF)

The NMF formulation was applied on the same set of concentration time series calculated with MARS-aero. Varimax rotation was applied in all analyses in order to minimise overlapping correlations between different factors and to accumulate strong loadings on the first few factors.

Figure 4.7: Factor scores and Varimax-rotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the full emissions scenario.
The NMF analysis of calculated time-series gives very similar results with the PCA analysis for the same stations. It is reasonable to assume the first factor represents traffic emissions, accounting for between 44% and 67% of total variance, while secondary photochemical contributions are represented by the second (in the case of SIRTA and Versailles) or third factor in the case of Aubervilliers, accounting for between 4% and 28%.

**Figure 4.8**: Factor scores and Varimax-rotated loadings obtained from concentration time-series calculated for the locations of the three AIRPARIF stations using the zero-traffic emissions scenario.
As in the case of PCA, visible similarities can be discerned in the loadings for the SIRTA site produced from the dispersion model (Figure 4.7) and the ones that were produced using the measurement data (Figure 3.4), with traffic-related pollutants appearing in first component. Performing NMF on the zero-out traffic calculations allows us again to positively identify contributions of traffic emissions and quantify their influences on the daily concentration patterns registered at each station.

It is again worth noting that in the zero-traffic emission case, the power of the suspected traffic factor is significantly reduced, indicating that our initial interpretation was correct.

### 4.4 Comparison of PCA and NMF

Both PCA and NMF attempt to apportion sources usually on the basis of observations and can be both run without prior knowledge of the source profiles. In most cases, in fact, major sources do not have a standard emission pattern for all constituent species. Examples of such sources include internal combustion engines, both diesel and gasoline, and other like cooking operations and smoking (Hildemann, Markowski et al. 1991). Source signatures that change with time are a limitation for this and other types of receptor models. It has also been suggested that factor analysis attempts to get more information out of atmospheric data than is really there (Henry, 1987).

PCA and NMF are principally the same technique, since the objective is to decompose the input matrix with an iterative method, minimizing the object function by the least squares. The do however have some important differences (Table 4.3). One of the most central differences is that PCA does not have constraints on the values of either the component loadings or scores as mentioned above, but requires that the resulting components be orthogonal, while NMF requires component loadings and scores to be non-negative without an orthogonality requirement. The lack of a non-negativity requirement in PCA has the potential of giving physically unreasonable results in the form of negative values for quantities that have to be non-negative. A possible interpretation for the negative loadings can be that of a representation of physically meaningful anti-correlations in the data and an effective non-negativity constraint for absolute PCA scores can be Varimax rotation. On the other hand, to gain freedom from the orthogonality condition, NMF uses an iterative method to fit the component loadings and scores to the measurements. This makes NMF numerically more difficult to implement than PCA.

An advantage of NMF has been that it is designed to permit the weighting of data (Chan & Mozurkewich, 2007), while PCA usually substitutes values less than the detection limit by one half of the detection limit (Harrison et al., 1996; Buhamra et al., 1998), with this being an arbitrary decision.

### Table 4.3: Comparison of PCA and NMF characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>PCA</th>
<th>NMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can run without source compositions as inputs</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Reactive compounds can be used as tracers</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Can include measurement uncertainties</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Can handle missing or below-detection-limit data</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Constrains solutions to be &gt; 0</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Must determine optimal number of “factors”</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Must interpret “factors” to identify sources</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>May encounter problems with co-linearity</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Adapted from Lynn M. Hildemann (http://www.epa.gov/osp/presentations/airtox/hildeman.pdf)
Huang et al. (1999) performed NMF and PCA on an aerosol composition data set and concluded that the two techniques can provide indistinguishable results. They also found that, to obtain meaningful results, the inclusion of appropriate input data and appropriate usage of the method are more important than the specific method used, outlining the need for supportive tools together with the core method.

4.5 Trajectory modelling

Air pollution levels depend to a considerable extent on local anthropogenic emissions. However, in some areas large amounts of pollutants can be emitted from various sources of natural or anthropogenic origin, which are transported from regions outside the domain of interest. Trajectory modelling constitutes an efficient aid for source apportionment calculations and source attribution analyses, mainly for the latter case. Back trajectory analyses use interpolated measured or modelled meteorological fields to estimate the most likely central path over geographical areas that provided air to a receptor at a given time. This method essentially follows a parcel of air backwards in hourly steps for a specified length of time.

The choice of the basic model parameters affects the trajectories generated and the final attribution analyses based on them. In particular, trajectories tend to become increasingly uncertain the further back in time they are used. Besides, the impact of the receptor height on an individual trajectory is also important. Low-ending trajectories represent air parcels nearer ground level. High-ending trajectories may represent more accurate boundary layer flow above the local terrain. Trajectory heights are not constant throughout the trajectory duration and often vary considerably from the receptor height selected. As a consequence, trajectories generated for the same location and time but with different receptor heights may indicate significantly different flow patterns for part of or the entire trajectory.

In the framework of the present study, the Hybrid-Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model was applied for a full calendar month period, i.e. July 2009, which coincides with the summer campaign period of MEGAPOLI for the Paris Area. HYSPLIT is a commonly used trajectory model developed by the National Oceanic and Atmospheric Administration’s (NOAA) Air Resources Laboratory (ARL) which uses archived 3-dimensional meteorological fields generated from observations and short-term forecasts. As regards the meteorological conditions occurring in Paris during July 2009, with a monthly average temperature above normal by 0.7 °C, Paris had a fairly mixed weather in July. In the first week the meteorological conditions were typical for this particular period of the year. On the other hand, in the rest of the month cooler times alternated with heat waves. Besides, rainfall was above the average normal levels during the whole of the month, while the sunshine was rather deficient. A trajectory with duration of 96 hours (4 days) backward in time was chosen as an appropriate period in order to accurately analyse the model results. Besides, in order to have a better overview of the large scale air mass flow on the air pollution levels of Paris, three different receptor heights were chosen for each model application, namely 500 m, 1500 m and 2500 m. As a result, the application of HYSPLIT for each day of the period under consideration generated three discrete back trajectory plots, one for each of the aforementioned end heights.

These back trajectories were examined in order to investigate the origin of the dominant air mass flow which affects Paris for each day of July 2009. The origin of the flows was classified into three major categories, namely “Atlantic”, “North” and “Non Adective”. The first one corresponds to air masses which apparently come from the Atlantic Ocean, the second one corresponds to air masses originating from Scandinavia or the North Sea, while the last one corresponds to air mass flows with a starting point located nearby the boundaries of France. No occasion of air masses from the South or East were identified. Figure 1.1 shows typical trajectory plots for each one of the aforementioned categories.
Table 4.4 presents the total number of days which are subsumed into each origin category and the respective percentages. As evidenced from table 1, during the period of July 2009 Paris was dominated by air mass flows originating from the Atlantic Ocean. Almost 68% of the back trajectory plots, which equals a total of 21 days, displayed oceanic air mass transport over the Atlantic with a direction to Paris. On the other hand, about 20% of the days were classified as “Non Advective” and for only a relatively small percent of 12% of the days the analysis of the trajectory plots indicated that air quality status in Paris was affected by air masses coming from the North. In figure 2, the time-series of the origin of the dominant air mass flow which affects Paris for each day of the period of July 2009 is presented. Those results are incompatible with the absence of episodic situations during the period of this study.
Table 4.4: Total number of days which are subsumed into each origin category and respective percentages.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Number of Days</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlantic</td>
<td>21</td>
<td>67.74%</td>
</tr>
<tr>
<td>Non advective</td>
<td>6</td>
<td>19.35%</td>
</tr>
<tr>
<td>North</td>
<td>4</td>
<td>12.91%</td>
</tr>
</tbody>
</table>

Figure 4.10: Time-series of the origin of the dominant air mass flow which affects Paris for each day of July 2009.
5. Conclusions

Efficient air quality management requires an accurate identification of pollution sources and of their individual contributions to the ambient pollutant concentrations. To this end, many modelling methodologies have been proposed and applied for the apportionment of atmospheric pollutants, including receptor-based models, methods based on the application of advanced chemical dispersion modelling, as well as a combination of both. It was of interest to this MEGAPOLI task to explore combined use of receptor modelling techniques and chemical transport models.

The dispersion model PMCAMx-2008 was applied for the European region in order to investigate the formation and oxidation of organic aerosol. The implementation of the new approach for primary organic aerosol volatility and the further oxidation of secondary organic aerosol were tested. Model results revealed that the new module acts like a source of oxygenated organic aerosol in urban and suburban regions and to a smaller degree for rural areas. These results indicate that the average diurnal profile of the organic aerosol concentrations can be used, if there are corresponding available measurements, as a tool for the evaluation of the performance of CTMs given the uncertainties in day to day organic aerosol and VOC emissions.

Despite their popularity and success in analysing multi-dimensional measurement datasets, many receptor modelling techniques present challenging problems in the interpretation of their results. One way to deal with these issues is to supplement the factor analysis of measurements with methods based on dispersion simulations. In such an approach, model results which were calculated using selected emissions sources were analysed in parallel with measurements, helping to identify the relevant factor contributions and enabling their interpretation as specific source types. The approach was applied using a zero-out method for traffic emissions aiming to ascertain the attribution and provide a quantification of the relevant factor contributions in the analysed measurement data. A key benefit of the use of this approach is that it can be applied with a limited computational cost as the runs required for a successful interpretation need only cover limited time periods. Further validation will be required to investigate the suitability of the approach in the analysis of diverse measurement datasets in conjunction with different model configurations.
Acknowledgements

We thank the MEGAPOLI campaign team for measurement data and especially for the summer campaign at SIRTA station. In particular, contributions by the following persons are acknowledged:

A. Borbon¹, J. Sciare³, V. Gros³, A. Colomb⁵, J. L. Jaffrézo⁶, J Cozic⁶, V. Michoud¹, M. Lopez⁴, V. Kaza⁴n¹, M. Crippa², J. C. Dupont⁷, S. Kukui⁸, A. Schmidt⁹

¹ Laboratoire InterUniversitaire des Systèmes Atmosphériques (LISA), Université Paris Est et 7, CNRS, Créteil, France,
² Paul Scherrer Institut, Villigen, Switzerland,
³ Laboratoire des Sciences du Climat et de l’Environnement (LSCE), Gif sur Yvette, France,
⁴ Max-Planck-Institute for Chemistry, Mainz, Germany,
⁵ Laboratoire de Glaciologie et Géophysique de l’Environnement, Grenoble, France,
⁶ Laboratoire de Météorologie Physique, Clermont-Ferrand, France,
⁷ SIRTA/IPSL, Palaiseau, France,
⁸ Laboratoire Atmosphères, Milieux, Observations Spatiales, Paris, France,
⁹ Universität Duisburg-Essen), Germany

The research leading to these results has received funding from the European Union’s Seventh Framework Programme FP/2007-2011 within the project MEGAPOLI, grant agreement n°212520.
References


Halmer G., Tsegas, G., Douros, I. and Moussiopoulos, N. (YYYY): Using a coupled meteorological and chemistry transport model to evaluate the impact of the aerosol direct effect on meteorology and concentration fields in Paris, XXXXXXX, YYY, XXX-YYY.


Atmospheric Environment, 39, 7181-7189.


Previous MEGAPOLI reports

Previous reports from the FP7 EC MEGAPOLI Project can be found at:
http://www.megapoli.info/

http://megapoli.dmi.dk/publ/MEGAPOLI_sr09-01.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr09-02.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr09-03.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-01.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-02.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-03.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-04.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-05.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-06.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-07.pdf

Scientific Report 10-08, MEGAPOLI-11-REP-2010-03, 19p, ISBN: 978-87-993898-1-0
http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-08.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-09.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-10.pdf


http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-12.pdf


http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-16.pdf

http://megapoli.dmi.dk/publ/MEGAPOLI_sr10-17.pdf


MEGAPOLI

Megacities: Emissions, urban, regional and Global Atmospheric POLLution and climate effects, and Integrated tools for assessment and mitigation

EC FP7 Collaborative Project
2008-2011

Theme 6: Environment (including climate change)
Sub-Area: ENV-2007.1.2.1:
Megacities and regional hot-spots air quality and climate

MEGAPOLI Project web-site
http://www.megapoli.info

MEGAPOLI Project Office
Danish Meteorological Institute (DMI)
Lyngbyvej 100, DK-2100
Copenhagen, Denmark
E-mail: alb@dni.dk
Phone: +45-3915-7441
Fax: +45-3915-7400

MEGAPOLI Project Partners

- **DMI** - Danish Meteorological Institute (Denmark)
  - Contact Persons: Prof. Alexander Baklanov (coordinator), Dr. Alexander Mahura (manager)
- **FORTH** - Foundation for Research and Technology, Hellas and University of Patras (Greece)
  - Prof. Spyros Pandis (vice-coordinator)
- **MPIC** - Max Planck Institute for Chemistry (Germany)
  - Dr. Mark Lawrence (vice-coordinator)
- **ARIANET** Consulting (Italy)
  - Dr. Sandro Finardi
- **AUTH** - Aristotle University Thessaloniki (Greece)
  - Prof. Nicolas Moussiopoulos
- **CNRS** - Centre National de Recherche Scientifique (incl. LISA, LaMP, LSCE, GAME, LGGE) (France)
  - Dr. Matthias Beekmann
- **FMI** - Finnish Meteorological Institute (Finland)
  - Prof. Jaakko Kukkonen
- **JRC** - Joint Research Center (Italy)
  - Dr. Stefano Galmarini
- **ICTP** - International Centre for Theoretical Physics (Italy)
  - Prof. Filippo Giorgi
- **KCL** - King’s College London (UK)
  - Prof. Sue Grimmond
- **NERSC** - Nansen Environmental and Remote Sensing Center (Norway)
  - Dr. Igor Esau
- **NILU** - Norwegian Institute for Air Research (Norway)
  - Prof. Andreas Stohl
- **PSI** - Paul Scherrer Institute (Switzerland)
  - Prof. Urs Baltensperger
- **TNO** - Built Environment and Geosciences (The Netherlands)
  - Prof. Peter Builijes
- **MetO** - UK MetOffice (UK)
  - Dr. Bill Collins
- **UHam** - University of Hamburg (Germany)
  - Prof. Heinke Schluenzen
- **UHel** - University of Helsinki (Finland)
  - Prof. Markku Kulmala
- **UH-CAIR** - University of Hertfordshire, Centre for Atmospheric and Instrumentation Research (UK)
  - Prof. Ranjeet Sokhi
- **USTUTT** - University of Stuttgart (Germany)
  - Prof. Rainer Friedrich
- **WMO** - World Meteorological Organization (Switzerland)
  - Dr. Liisa Jalkanen
- **CUNI** - Charles University Prague (Czech Republic)
  - Dr. Tomas Halenka
- **IFT** - Institute of Tropospheric Research (Germany)
  - Prof. Alfred Wiedensohler
- **UCam** - Centre for Atmospheric Science, University of Cambridge (UK)
  - Prof. John Pyle

Work Packages

WP1: Emissions
  (H. Denier van der Gon, P. Builijes)
WP2: Megacity features
  (S. Grimmond, I. Esau)
WP3: Megacity plume case study
  (M. Beekmann, U. Baltensperger)
WP4: Megacity air quality
  (N. Moussiopoulos)
WP5: Regional and global atmospheric composition
  (J. Kukkonen, A. Stohl)
WP6: Regional and global climate impacts
  (W. Collins, F. Giorgii)
WP7: Integrated tools and implementation
  (R. Sokhi, H. Schlunzen)
WP8: Mitigation, policy options and impact assessment
  (R. Friedrich, D. van den Hout)
WP9: Dissemination and Coordination
  (A. Baklanov, M. Lawrence, S. Pandis)