Inverse Modelling for Tropospheric Chemical State

Estimation by 4-Dimensional Variational Data Assimilation

from Routinely and Campaign Platforms

Inaugural – Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch–Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Zoi Paschalidi

aus Athen, Griechenland

Köln, 2015

Supervisor: P.D. Dr. H. Elbern
Prof. Dr. A. Wahner

Date of oral examination: 30 October 2015

Abstract

The use of models to analyse the complex atmospheric processes deals with a lot of uncertainties of key parameters, most prominently the emission rates along with the initial values. Especially for campaigns, the ability to analyse the processes is dependent on the most precise state of the variables needed for the parametrisation of the system, notably the emissions. Using its 4D-var and inverse modelling scheme, EURAD-IM is able to provide optimised initial state variables, highlighting here the optimisation of the emission factors for the gas phase. In order to address how emission plumes can be identified, real case studies were applied with EURAD-IM to study the dispersion of the anthropogenic emissions over regional to urban surface within the planetary boundary layer. To achieve a high analysis skill, some features needed to be updated and also some new to be developed and added to the model setup. The online calculation of emissions in the CTM has been coupled to the joint optimisation of initial values and emission rates, providing detailed spatial and temporal emission distribution per source and per grid cell. Moreover, improvement of the background emission factor error covariances is achieved by including additional correlations between the emitted species and rectifying the standard deviations of the emission factors, optimising, thus, the precondition of the minimisation problem. In addition, assimilation of the comprehensive observational set of PEGASOS project during campaign in the polluted region of Po valley, performed to study the vertical structure of the atmosphere, analysing the mixing within the PBL and to validate the airborne data influence on the model performance. Quality control of the assimilation procedure, obtained by χ^2-validation and comparison with independent observations, shows successful minimisation performance of the 4D-var algorithm. The case study analysis concluded to optimised emission factors, addressing also the lack of representativity of observations (such as NO_2) by the high resolution of the nesting technique - up to 1 km. The more in depth assessment of the vertical mixing in the PBL, by the assimilation of the airborne campaign measurements, captures fairly clear the temporal variations of the emission patterns, the influence of the inverse temperature on the concentrations of pollutants and the layered structure of the PBL. In general, considerable improvement of the forecast quality of the model is achieved and the system's ability to provide improved chemical consistent simulation results throughout the PBL, mainly in urban polluted regions, is demonstrated. Besides, the assimilation of airborne measurements could be applied to shed light to the development of the PBL and the processes there in, information that is shown to miss when no campaign data are available.

Kurzzusammenfassung

Modelle zur Analyse komplexer atmosphärischer Prozesse beinhalten verschiedene Fehlerquellen, wie zum Beispiel Emissionsraten oder Anfangswerte. Insbesondere die modellgestützte Prozessanalyse bei Messkampagnen hängt von der bestmöglichen Abschätzung von Modellparametern ab, zu denen im Fall chemischer Transportmodelle die Emissionsraten zählen. Das EURAD-IM System bietet die Möglichkeit einer kombinierten Optimierung von Anfangswerten und Emissionsfaktoren gasförmiger Spurenstoffe mit Hilfe vierdimensionaler variationeller Datenassimilation. Der Fragestellung, inwiefern Emissionsfahnen identifiziert werden können, wird mit Hilfe des EURAD-IM Systems anhand von Fallbeispielen nachgegangen, indem die Ausbreitung von anthropogenen Emissionen über ländlichen und städtischen Gebieten innerhalb der planetaren Grenzschicht untersucht wird. Um die Modellanalyse zu verbessern, war es notwendig einige Modellteile zu aktualisieren, andere neu zu entwickeln und einzugliedern. Hier ist zum Beispiel die Kopplung der Berechnung räumlich und zeitlich hochaufgelöster Emissionsraten im CTM an die gemeinsame Optimierung von Anfangswerten und Emissionsfaktoren zu nennen. Des Weiteren wurde die Hintergrundfehlerkovarianzmatrix für Emissionen neu aufgestellt, indem Korrelationen zwischen emittierten Stoffen ergänzt, sowie Standardabweichungen der Emissionsfaktoren angepasst wurden. Die umfassenden Bobachtungsdaten der Zeppelin Messkampagne des PEGASOS Projekts in der Po-Ebene wurden genutzt, um die vertikale Struktur in der planetaren Grenzschicht zu analysieren. Eine Qualitätskontrolle des Datenassimilationsverfahrens wurde über einen χ^2-Test und den Vergleich von Analysergebnissen mit unabhängigen Beobachtungen erfolgreich durchgeführt. Die Analyse der PEGASOS-Fallstudie mit hoher horizontaler Auflösung bis zu 1 km ergibt optimierte Emissionsfaktoren und erhöht deutlich die Representativität der Beobachtungsdaten (insbesondere für NO_2). Die Assimilationsgestützte Untersuchung der vertikalen Mischung in der planetaren Grenzschicht durch die Assimilation von Zeppelinmessungen erfasst sehr exakt die zeitliche Emissionsstruktur, den Einfluss von Inversionen auf die Konzentration von Schadstoffen und die mehrschichtige Struktur der Grenzschicht. Durch den Einsatz der kombinierten Optimierung von Anfangswerten und Emissionsfaktoren wird eine beachtliche Verbesserung der Vorhersagequalität des Modells erreicht und die Fähigkeit des Systems zu chemisch konsistenten Simulationen über die gesamte vertikale Ausdehnung der planetaren Grenzschicht bewiesen. Weiterhin kann die Assimilation von luftchemischen Messdaten der PEGASOS Kampagne für ein besseres Verständnis der Entwicklung der planetaren Grenzschicht und der darin stattfindenden Prozesse genutzt werden.

Contents

List of Figures

List of Tables

Wissenschaft:
Es ist nicht ihr Ziel,
der unendlichen Weisheit eine Tür zu öffnen,
sondern eine Grenze zu setzen dem unendlichen Irrtum.

Bertolt Brecht

Chapter 1

Introduction

In atmospheric chemistry the use of numerical models aims to represent accurately and consistently the state of the atmosphere, a challenging task for assessing the air quality. As an ubiquitous air quality monitoring is impossible, air quality modelling is the necessary substitute to study and predict the atmospheric processes [*Jacobson*, 2000]. Strong motivation consists the determination of the interactions between regional climate change and atmospheric composition regarding emissions [*Brasseur et al.*, 2001; *Meleux et al.*, 2007]. The dispersion of the emissions over the surface up to the planetary boundary layer (PBL) is of great research interest, not only because of their impact on air quality [*Stull*, 1988; *Ribeiro et al.*, 2013], but also in human's health [*Pierce et al.*, 2010; *Zhang et al.*, 2012]

For the purposes of air quality modelling, advection-diffusion-reaction models are used to predict concentrations of pollutants, taking into account allocation of their sources [*Beychok*, 2005]. The models include the human knowledge of physical laws and chemical reactions that take place in the atmosphere [*Sandu et al.*, 2005]. However, only this information is not enough to depict the state of the atmosphere [*Turner*, 1994]. As an independent source of information, observations play a very important role and can therefore be used to ameliorate the simulation results, improving the model's forecast [*Sandu and Chai*, 2011; *Lahoz and Schneider*, 2014]. Today there is an increase of available information and air quality observations [*USEPA*; *MACC*; *EEA*]. The different types of observational instruments (i.e., ground or airborne based instruments, satellite retrievals) provide a huge amount of measurements, that can give a detailed insight into the state of the atmosphere [*Fishman et al.*, 2008; *Martin*, 2008; *Inness et al.*, 2015; *Elbern et al.*, 2014]. Additional advantages offer the data collected by campaigns (i.e., flight campaigns with aeroplanes, Zeppelins and balloons), as they give a more complete picture of the air quality in elev-

ated heights, in special areas of interest [*Chai et al.*, 2006; *Hamburger et al.*, 2011].

Although benefiting from high accuracy and abundance, atmospheric observations are spatially inhomogeneously distributed [*Carmichael et al.*, 2008]. Additionally, many instrument sites are not representative for an area of the size of a regional model's grid cell [*Buizza et al.*, 2007]. These facts result in giving the model misleading information about the air conditions of the area under consideration. Typical examples are predominantly NO_2 measurements; the observation sites are usually located in populated areas, or close to large streets and point sources of NO_x emissions. Therefore, they cannot always represent the state of NO_2 of a whole grid cell [*Elbern et al.*, 2007*b*]. Another challenge is that there are cases where species are not measured directly, fact that leads to lack of detailed information. For instance, in case of aerosol observations, not only it is their optical thickness that is measured instead of their individual composition, but also their size integrated lumped mass is given in μg [*Lyamani et al.*, 2012; *Sandrini et al.*, 2014; *Pilinis et al.*, 2014].

The maximum gain of information can be retrieved by the balanced combination of the various types of information on the state of the atmosphere (i.e., observations) and the knowledge of the physical and chemical processes (i.e., models) [*Lahoz et al.*, 2010]. This is achieved by data assimilation methods [*Courtier*, 1997]. Data assimilation is the discipline that fills in the gaps of the discrete observations, in both space and time, by the use of models [*Brasseur and Jacob*, 2013]. Thus, data assimilation uses the laws of the system's evolution in space and time, coded in a numerical prognostic model, as well as a previous most probable estimate of the state, in order to resume how information varies between the discrete set of observations [*Daley*, 1991].

Since the late '90s, spatio-temporal variational data assimilation algorithms are used in meteorology for operational weather forecasting [*Navon*, 2009]. The progress in developing comprehensive Chemistry Transport Models (CTM) for both the stratosphere and the troposphere has motivated the enhancement of air quality forecast services and long term model simulations. Later, the four dimensional variational (4D-var) data assimilation methods are used in atmospheric chemistry. The ill-conditioned problem of representing the atmospheric state in a chemical consistent way can be handled only by advanced data assimilation methods, that apply the technique of inverse modelling. *Fisher and Lary* [1995] show the usefulness of the temporal data assimilation in atmospheric chemistry with a strongly simplified box model simulation for stratosphere. *Elbern et al.* [1997] uses the 4D-var technique for troposphere and the gas phase mechanism of the second generation Regional Acid Deposition Model (RADM2, [*Stockwell et al.*, 1990]), in the EURopean Air pollution Dispersion

model system (EURAD). Later, *Elbern and Schmidt* [1999] extend the regional EURAD to the first available complete adjoint CTM for the troposphere and the first real world application of a high ozone concentration case is presented in *Elbern and Schmidt* [2001].

Data assimilation typically aims to improve the prognostic state variables at initial time. These variables represent critical factors for an improved analysis, which have large uncertainties and so they have to be optimised [*Kalnay*, 2003]. The initial values of the chemical species are primarily one of these parameters. Yet, especially in troposphere there are several other processes which have significant impact on the temporal evolution of the system; *Hanna et al.* [1998; 2001]; *Schmidt and Martin* [2003] are some of the studies which give a thorough assessment of the uncertainties of various input model parameters.

Incontrovertibly, emissions have a large forcing impact on tropospheric chemistry. The emission rates have great influence on the trace gas concentrations. In addition, the possible strong exchange between the PBL and the free troposphere can robustly influence large model areas by long range transport. Moreover, emission rates are subject to huge uncertainties, especially in areas exposed to air quality problems, since the CTM's hourly emission rates are calculated on the basis of past annual emission estimates. Hence, temporal evaluation patterns are not available.

The need for further studies on the emission issue has urged a lot of researchers to work on the estimation of source and sinks. Starting with *Newsam and Enting* [1988] and *Enting and Newsam* [1990], that address the distribution of sources and sinks of CO_2 by the inverse of the diffusion equation, the research goes on to variational approaches [*Kaminski et al.*, 1999*a*;*b*]. *Kaminski et al.* [2002] assimilates CO_2 observations and achieves more realistic flux simulations. The same aim follows *Engelen et al.* [2004], combining the 4D-var algorithm with satellite CO_2 data. Starting with the use of variational data assimilation for source estimation [*Issartel*, 2003; *Hourdin et al.*, 2006], *Bocquet* [2005*a*;*b*] proceeds with the estimation of position, time and strength of emission sources.

Up to this point, the regularisation of the ill-conditioned problem of the atmospheric transport inversion is achieved by a priori information on the spatial and temporal distribution of sources and sinks derived from emission inventories. Hence the interest of inverse modelling turns on using observations to optimise the inventories themselves, in order to reduce their uncertainties. *Muller and Stavrakou* [2005] assess emission rates of continental scales, by assimilation of CO tropospheric column retrievals, while *Chai et al.* [2009] scales NO_x emission factors using SCIAMACHY tropospehric NO_2, as well as *Quélo et al.* [2005] performs one year of observations and simulates them to retrieve

NO_x emission and their time distribution. Further work on optimisation of CO emission estimates by 4D-var is presented in *Yumimoto and Uno* [2006] for a regional model and in *Hooghiemstra et al.* [2011] for a global one. Moreover, data assimilation techniques have been applied in estimating biogenic and aerosol emissions [*Dubovik et al.*, 2008; *Koohkan et al.*, 2013; *Li et al.*, 2013; *Sofiev et al.*, 2015], where also rise 4D-var applications [*Yumimoto et al.*, 2007; *Meirink et al.*, 2008; *Nieradzik*, 2011]. The capabilities of the emission optimisation have additionally been studied in case of special emission episodes, as Chernobyl explosion, in 1986 [*Bocquet*, 2012] and the accident in Fukushima, in 2011 [*Winiarek et al.*, 2014].

In chemical data assimilation there are applications that optimise the initial values jointly with emission factors for selected species. For instance, representative of these studies is the inverse modelling application of *Pison et al.* [2007] for NO_x emission and concentrations using campaign data, or of *Miyazaki et al.* [2012] for assimilation of NO_2, O_3, CO and HNO_3 satellite data in a global model, as well as the ensemble Kalman filter of *Tang et al.* [2011] for NO_x and VOCs initial conditions and emissions. Though, the first time that the 4D-var method is implemented for a complete set of emitted species is in *Elbern et al.* [2000]. Here an identical twin experiment is adopted to show that the inversion technique can be applied not only for the initial concentration values, but simultaneously for the emission rates and, thus, quantitatively determinate emission patterns. Based on these results, a real world experiment is presented in *Elbern et al.* [2007*a*] for the estimation of pollutant precursor sources by the joint optimisation of initial values and emission factors, using a variety of observations, including campaign data.

The results of scientific studies in the field come to the point that the optimisation of the initial concentration values together with the emission rates lead to an improved forecast of the chemical model evolution. Nonetheless, a good surface forecast is often critically dependent on accurate estimates of surface fluxes, so the concise overview of the atmosphere's chemistry is still an important issue to be addressed. The emissions of the human activity on the surface influence the chemical reactions and more over the interaction of the atmospheric components impacts the air quality. Therefore, it is necessary that models provide updated emission maps, by optimising the emission factors, in order also to identify emission plumes.

Tackling these issues, 4D-var gains information on the sources of pollutants and on the general chemical state of the atmosphere within the model, by the assimilation of all available measurements. Certainly, ground based in-situ measurements have the advantage of being abundant and covering large areas, a fact that is important for the assimilation result. However, the essential

information of the atmospheric state in higher vertical levels within the PBL, is only reached by airborne campaign data. Since flight campaigns are difficult and expensive to be frequently organised, a more in depth assessment of the role of the PBL and its importance for skilful simulations is hardly evaluated by appropriate observations. Therefore, it is of high value to estimate to what extend the high resolved campaign data improve the model's performance and result in better air quality forecast.

Consequently, the main intention of the present thesis is first to demonstrate the feasibility of estimating the athropogenic emissions over urban areas, by 4D-var techniques in a regional dispersion model, the EURAD-IM [*Elbern et al.*, 1997; *Elbern and Schmidt*, 1999]. To this end, extensions and improvements to the EURAD-IM system are developed in the scope of this work, namely the coupling of the online calculation of emissions to the joint optimisation of initial values and emission rates, as well as the improvement of the emission factor background error covariances, by expansion of the correlation of the emitted species, the standard deviations for the emission factors, and technical updates to the model performance. To evince the model's ability to optimise emissions and determine emission patterns, three case studies are simulated. The chosen configuration aims to deal with the lack of representativity of observations. The evaluation of the analysis results is twofold; a χ^2 validation and a comparison with independent observations take place.

Second, the high quality airborne data of the Pan-European Gas-AerSOLs-climate interaction Study (PEGASOS) in July 2012 are assimilated to assess their impact on the model's performance. Furthermore, the atmosphere's vertical structure is studied, analysing the vertical mixing in the PBL. The most of the trace gas conversion takes place in higher levels up to the PBL, where CO, NO, NO_2 contribute to OH-sink [*Stull*, 1988]. The rich campaign measurements of OH reactivity and vertical profiles [*Jäger*, 2014] motivate the study by 4D-var of the vertical distribution of emissions and thus the evolution of the PBL.

This thesis is structured as follows: chapter 2 gives an overview of the 4D-var data assimilation method for the joint optimisation of initial species concentrations and emission rates. The features of the EURAD-IM system are discussed in chapter 3, highlighting the improvements developed in this work, while in chapter 4 the available observational data are presented, together with the description of the observation error covariance matrix. Finally, chapter 5 describes the characteristics of the case studies and the results of their analysis. The main findings are summarised in chapter 6.

Chapter 2

Chemistry Data Assimilation by 4D-var

The 4D-var data assimilation method is able to use the information of all kind of measurements; direct or remote sensing observation as well as satellite retrievals, can be inserted into the model, even if they are scattered in space and time. The 4D-var assimilation algorithm propagates all the available information forward and backward in time. By this way the model's simulation is fitted to the set of observations, distributed in a predefined time interval. So the model is able to calculate the *analysis state*; the state that lays in the minimum distance between model and observations, where the consistency of the system is guaranteed. Fig. 2.1 illustrates the aforementioned assimilation procedure that takes place in the 4D-var problem. Regarding the multiple advantages of the method, the maintenance of the physical and chemical consistency of the problem has to be highlighted. That is why the information is propagated to regions with poor data and the analysis includes observed and non-observed constituents. Within 4D-var data assimilation the analysis problem is formulated as a minimisation problem by the variational calculus. It is necessary to define an objective function, or with other words a *cost function*, that calculates the distances between the model simulation and the observation, during a predefined time interval.

2.1 Optimisation of initial values and emission factors

The joint optimisation of both the initial chemical values and the emission rates enables the model trajectory to fit to the 'truth' state. By this way, the

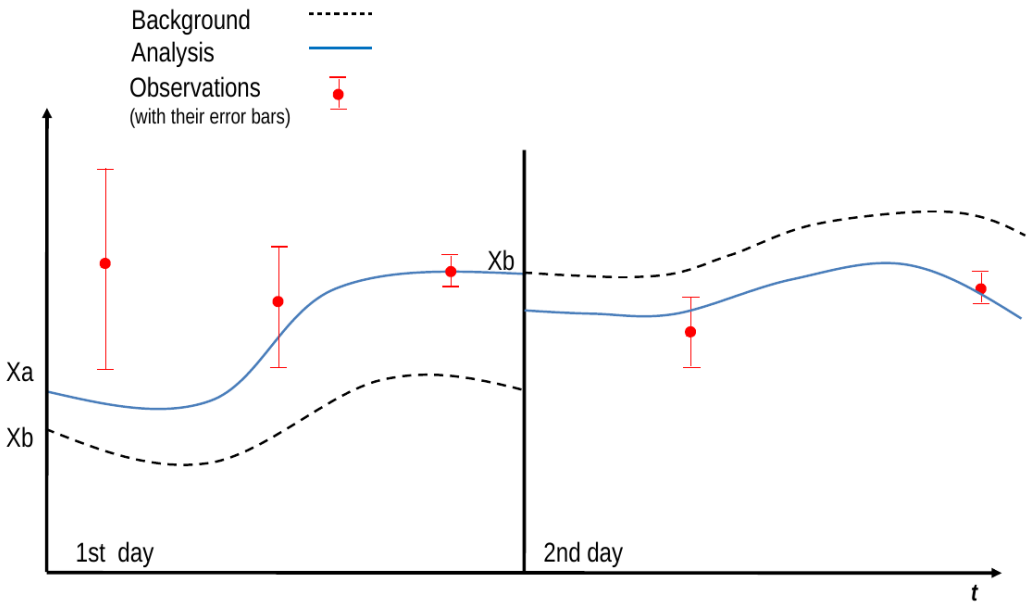

Figure 2.1: *The 4D-var assimilation procedure. During the first day of interest the forward model calculates the atmospheric state (background run - black dashed line), based on the a priori information for the initial states, and the initial emission rates, the background x_b. Taking into account the observations with their error bars, the adjoint integration follows. It calculates the atmospheric state backward in time, resulting in an updated deviation from the background state and an emission correction factor, the analysis x_a. Now with these optimal data taken as model input, the model trajectory is recalculated, providing the best estimation of the truth state and an a priori initial state for the next day(analysis run - blue solid line).*

biases of the separate optimisation of the two parameters vanish and the model is benefited from their different individual advantages. The initial values play an important role in the beginning of the optimization, mainly when no air is advected from other observed area. The emission rates optimization influence in long term and shows the correct direction to fit the observations until the end of the simulation. Consequently, an optimal analysis result cannot be given from the separate optimisation of only the initial values or the emission rates, but from the joint optimization of both of them.

In the CTM the emission data strength from the inventories are accordingly treated in order to shape a more useful form of information for the optimisation problem, giving the *emission rates*. Within the model, the emission rates differ for each species l, at each location (i, j, k) and each time t, fact that

leads the stepwise emission variation to result in an extremely ill posed inversion problem. In order to deal with this problem, the degree of freedom of the emission rate space state should be reduced drastically. Thus it is the diurnal profile shape of the emissions, which is generally better known, that it is considered as strong constraint. By this way, the model calculates and follows the evolution of the emission rates based on these profiles, such that the calculated data and the initial ones differ only by the amplitude of the diurnal profiles. So now only these amplitudes are taken as control parameters to the optimisation problem, by the *emission factors*. The *emission factor* is a representative value that attempts to relate the quantity of a pollutant released to the atmosphere with an activity associated with the emission process. This means that the emission factors vary for the different species emitted in various locations, dependent on the emission source.

The theoretical background of the 4D-var data assimilation technique is presented below, based on the notations recommended in *Ide et al.* [1997].

2.2 Inversion of the variational model

In order to define the cost function J, an *a priori* or *first guess* of the state variables is given. The chemical state vector $\boldsymbol{x}_b(t) \in \mathbb{R}^N$ denote the *background* field of the model, with dimension N of the phase space portion for the chemical constituents. The vector $\boldsymbol{x}_b(t)$ represents the best estimate of the current state $\boldsymbol{x}(t)$, prior the use of observations at time t. The background field is obtained either by some climatological files, a short range forecast based on the analysis of the previous day, or a result of a previous 4D-var assimilation, as it is done in the current study. There is also given the *emission rates* $\boldsymbol{e}_b(t) \in \mathbb{R}^E$, with E the dimension of partial phase space of the emission rates. The vector $\boldsymbol{e}_b(t)$ denotes the initial information about the emissions before the model's calculation of the emission rates $\boldsymbol{e}(t)$, at time t. The background emission rates are usually taken from inventories - for the present study they are based on TNO-MACC II [*Kuenen et al.*, 2014]. Further, the *observations* $\boldsymbol{y}^o(t) \in \mathbb{R}^{M(t)}$ are given, where $M(t)$ is the dimension of the observation space during the time interval $[t_0, t_N]$. In general, the observation space is much smaller than the model space, $M(t) << N$, making the interpolation of the observations a main challenge in the 4D-var assimilation algorithm. The assimilation window, i.e., the time interval of the performed simulations, for the configuration in the present study, is 24 hours, a full day assimilation.

Let $\boldsymbol{x}_t(t)$ be the true state of the chemical conditions of air at time t, our

goal is to end up to an analysis vector $\boldsymbol{x}_a(t)$, such that it is as close to the true state as possible, $\boldsymbol{x}_a(t) \rightarrow \boldsymbol{x}_t(t), \forall t \in [t_0, t_N]$, that is the most probable state.

So, under the assumption that the model is deterministic, i.e. it is not influenced by stochastic processes, it can be generally written that the true model state $\boldsymbol{x}_t$ follows

$$\boldsymbol{x}_i = \mathcal{M}_{[t_0,t_i]}(\boldsymbol{e}_{[t_0,t_i]})\boldsymbol{x}_{i-1} \tag{2.1}$$

where $\mathcal{M}_i$ is a nonlinear model operator that integrates from time t_{i-1} to time t_i.

Based on this principle, our CTM, including the emissions, is given by

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{M}_i(\mathbf{x}(t_i), \mathbf{e}(t_i)). \tag{2.2}$$

Here the model operator $\mathcal{M}$ depends on both the chemical conditions and the emission rates at time $t_i \in [t_0, t_N]$.

The Eq. (2.2) shows that the state variable $\mathbf{x}(t)$ at time t is uniquely defined by $\mathcal{M}$ and the emission estimate $\mathbf{e}(\tau)$, $\tau \in [t_0, t]$, given an ever fixed initial state $\mathbf{x}(t_0)$. Which means that the model evolves in time as

$$x_i = \mathcal{M}_i(e_i)x_{i-1} \tag{2.3}$$

For simplicity, it is noted $x_i \equiv x(t_i)$, $e_i \equiv e(t_i)$, $\forall i = 1, ..., N-1$ such that $t_i \in [t_0, t_N]$. It is concluded, then, that a perturbation $\delta\mathbf{e}$, $\delta\mathbf{e}(t) = \mathbf{e}(t) - \mathbf{e}_b(t)$, on $\mathbf{e}(t)$ produces a perturbation $\delta\mathbf{x}$, $\delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_b(t)$, on $\mathbf{x}(t)$, which evolves via the tangent linear form of Eq. (2.2),

$$\frac{d\delta\mathbf{x}}{dt} = \mathbf{M}'_{[t_0,dt]}(\delta\mathbf{x}, \delta\mathbf{e}), \tag{2.4}$$

where it is considered that $\delta\mathbf{x}(t_0) = 0$ and $\mathbf{M}'$ is the linearised model operator, the *tangent linear model*, which will be discussed below.

2.3 Treatment of emission rates

As it is mentioned above, the strong constraint for the diurnal profile shape of the emissions leads to the optimisation of the emission factors. Simultaneously, the ill-posed inverse problem of emissions needs that the positive definiteness of emission is maintained throughout all the optimisation procedure. Therefore, to satisfy these requirements for the emission factors, a positive

function that describes them should be defined. For each species l and each location (i, j, k) in the model domain, a function $\mathbf{u} = \mathbf{u}(e) = ln(\mathbf{e})$ is introduced, such that

$$\begin{aligned}\delta\mathbf{u}(i,j,k,l) &= \mathbf{u}(e, e_b) \\ &= ln\left(e\left(i,j,k,l\right)\right) - ln\left(e_b\left(i,j,k,l\right)\right) \\ &= ln\left(\frac{e\left(i,j,k,l\right)}{e_b\left(i,j,k,l\right)}\right).\end{aligned} \tag{2.5}$$

It is noticed that the given definition allows $\delta\mathbf{u}(i, j, k, l)$ to be constant and possitive definite during the whole assimilation window.

So, if it is also set $diag\left(U\left(i,j,k,l\right)\right) := exp\left(\delta\mathbf{u}\left(i,j,k\right)\right) - 1$, then $\delta\mathbf{e} = U\mathbf{e}_b$ and now Eq. (2.4) is written as

$$\frac{d\delta\mathbf{x}}{dt} = \mathbf{M}'(\delta\mathbf{x}, Ue_b), \tag{2.6}$$

which lets the emission information be time independent in the time interval. It is underlined that the tangent linear model $\mathbf{M}'$ includes two components, the tangent linear model with respect to the initial values, $\mathbf{M}'_{i,iv}$, and the equivalent with respect to the emission factors, $\mathbf{M}'_{i,ef}$, such that $\mathbf{M}'_i = \begin{Bmatrix}\mathbf{M}'_{i,iv} \\ \mathbf{M}'_{i,ef}\end{Bmatrix} = \begin{Bmatrix}\dfrac{\partial\mathcal{M}_{i,iv}}{\partial x_i} \\ \dfrac{\partial\mathcal{M}_{i,ef}}{\partial u}\end{Bmatrix}$.

Consequently, Eq. (2.5) and Eq. (2.6) let the optimisation of $\delta\mathbf{e}$, through $\delta\mathbf{u}$, result in a *correction factor*

$$f\left(i,j,k,l\right) = \frac{e\left(i,j,k,l\right)}{e_b\left(i,j,k,l\right)}, \tag{2.7}$$

which depends on the location (i, j, k, l) of the model's grid, but it is independent of time t_i, within the assimilation window. So, the Eq. (2.6) and Eq. (2.7) imply that the emission rates, $\mathbf{e}$, are described as the background knowledge of the emission rates, $\mathbf{e}_b$, scaled by the emission factor f, $\mathbf{e}(t_i) = f \cdot \mathbf{e}_b(t_i)$, for every time $t_i \in [t_0, t_N]$. Thus now the problem of the optimisation of the emission rates is equivalent with the optimisation of the emission correction factors, for a defined set (i, j, k, l).

2.4 The cost function of the minimisation problem

The distance between the model states can be defined, considering also the emissions, and the observations, at time t, with the help of the cost function. The cost function J, reflects here the imposition of a strong constraint [*Sasaki*, 1970], and consists of three parts: the background cost of the initial state of the chemical constituents J_{iv}, the observational cost J_{ob}, and the cost of the emission rates J_{ef},

$$\begin{aligned} J\left(\mathbf{x}(t), \mathbf{e}(t)\right) &= J_{iv} + J_{ob} + J_{ef} \\ &= \frac{1}{2} \int_{t_0}^{t} \left[x(t) - x_b(t)\right]^T \mathbf{B}^{-1} \left[x(t) - x_b(t)\right] dt \\ &+ \frac{1}{2} \sum_{i=0}^{N} \left(\left[H_i M_i \left(x(t)\right) - y_i\right]^T \mathbf{R}^{-1} \left[H_i M_i \left(x(t)\right) - y_i\right] \right) \\ &+ \frac{1}{2} \int_{t_0}^{t} \left[e(t) - e_b(t)\right]^T \mathbf{K}^{-1} \left[e(t) - e_b(t)\right] dt, \end{aligned} \tag{2.8}$$

where, the superscripts T and -1 denote the transpose and the inverse of a matrix, respectively, $H(t) \in \mathbb{R}^{M(t)\times N}$ is a forward observation operator that maps from model space to observation space and produces the model equivalents of observations, given the time t, $\mathbf{B} \in \mathbb{R}^{N\times N}$ is the error covariance matrix of the background values of the chemical constituents, $\mathbf{R} \in \mathbb{R}^{M(t)\times M(t)}$ is the error covariance matrix of the observations, $\mathbf{K} \in \mathbb{R}^{E\times E}$ is the error covariance matrix of the emissions. A more detailed discussion about the formation and the properties of the matrices $\mathbf{B}$ and $\mathbf{K}$ is given in Chapter 3, whereas for $\mathbf{R}$ in Chapter 4.

The 4D-var data assimilation algorithm implies that the better the initial conditions introduced to the model are, the better the forecast and analysis result are. So the optimum of the atmospheric state at time t_0, the *initial time*, is needed to be identified. Then, since the goal is the joint optimisation of the initial chemical values $\mathbf{x}_b(t_0)$ and the emission rates $\mathbf{e}_b$, both parameters should be combined in a common vector, by an appropriate scaling, based on the equivalent perturbations $\delta\mathbf{x}(t_0)$ and $\delta\mathbf{e}$. To do so, it is set $\delta\mathbf{z} := (\delta\mathbf{x}(t_0), \delta\mathbf{u})^T$, with $\mathbf{u} = \mathbf{u}(e)$ as described above. Now $\delta\mathbf{z}$ is the full control parameter of the model evolution.

Then Eq. (2.8) can be written as

$$\begin{aligned} J(\delta\mathbf{z}) &= J_{iv} + J_{ob} + J_{ef} \\ &= \frac{1}{2}\left[\delta\mathbf{x}(t_0)\right]^T \mathbf{B}^{-1}\delta\mathbf{x}(t_0) \\ &+ \frac{1}{2}\sum_{i=0}^{N}\left[d(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i)\right]^T \mathbf{R}^{-1}\left[d(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i)\right] \\ &+ \frac{1}{2}\left[\delta\mathbf{u}\right]^T \mathbf{K}^{-1}\delta\mathbf{u}, \end{aligned} \tag{2.9}$$

where $\mathbf{H}(t)$ is the linearised approximation of the forward observation operator H, that $\mathbf{H}(t) = H'(t)$ and d_t is the innovation vector $d(t) = y^o(t) - H(t)x_b(t)$, i.e. the differences between the observations and the corresponding model equivalent state $x_b(t)$, at time t.

The desired optimum state for $\delta\mathbf{z}$ is nothing more than a Best Linear Unbiased Estimate (BLUE) of the initial conditions and the emission rates for the ill conditioned problem, with respect to all observations during the assimilation window, the *analysis* state

$$\delta\mathbf{z}_\alpha = \left(\delta\mathbf{x}_\alpha, \delta\mathbf{u}_\alpha\right)^T. \tag{2.10}$$

The $\delta\mathbf{z}_\alpha$ is given by the minimisation of the $J(\delta\mathbf{z})$. For doing so the gradient of J with respect to the joint chemical state and emission rate variable $\delta\mathbf{z}$ has to be determined. This is derived by

$$\nabla_{\mathbf{z}} J = \mathbf{B}^{-1}\delta\mathbf{x}(t_0) - \sum_{t_i=t_0}^{t_N} \mathbf{H}^T(t_i)\mathbf{M}_i^*\mathbf{R}^{-1}(\mathbf{d}(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i)) + \mathbf{K}^{-1}\delta\mathbf{u} \tag{2.11}$$

where $\mathbf{H}^T$ is the adjoint of the tangent linear observation operator $\mathbf{H}$, and $\mathbf{M}_i^*$ is the adjoint of the tangent linear of the model, such that

$$\mathbf{M}^*_{[t_0,t_i]} = \begin{Bmatrix} \mathbf{M}^*_{[t_0,t_i],iv} \\ \mathbf{M}^*_{[t_0,t_i],ef} \end{Bmatrix} = \begin{Bmatrix} \prod_{j=i}^{0} \mathbf{M}^T_{j,iv} \\ \prod_{j=i}^{0} \mathbf{M}^T_{j,ef} \end{Bmatrix}.$$

It is underlined that when using partial derivatives with respect to elements of a discretised representation, as it is done here, the adjoint of a matrix is equivalent to its transpose. Moreover, $\mathbf{B}^{-1}$, $\mathbf{R}^{-1}$, $\mathbf{K}^{-1}$ are the inverse matrices of $\mathbf{B}$, $\mathbf{R}$, $\mathbf{K}$ respectively.

2.5 The precondition of the cost function

Formally, a joint covariance matrix $\begin{pmatrix} \mathbf{B} & corr \\ corr & \mathbf{K} \end{pmatrix}$ can be defined. This joint covariance matrix is able to describe the correlations between the concentration levels of the emitted species and their emission rates around the areas of sources and also after the chemical and transport procedures that take place in the low troposphere. Furthermore the $\begin{pmatrix} \mathbf{B} & corr \\ corr & \mathbf{K} \end{pmatrix}$ can include the cross-correlations between the initial values and the emission rates. However, the current study omits the aforementioned correlations. The reason behind this approach is on the one hand the poor background statistical information that are so far available, and on the other hand the high computational demands for the minimisation of the highly non-linear Eq. (2.9). Considering that only matrix $\mathbf{B}$ has $\mathcal{O}(10^6)$ order of magnitude, the numerical treatment of $\mathbf{B}$ and $\mathbf{K}$ is prohibitive.

In order to alleviate the computational expenses an incremental form of the 4D-var problem is needed. By preconditioning the cost function, the full non-linear problem is transformed to a sequence of linearised functions, avoiding ill-conditioning. *Elbern and Schmidt* [2001] present a singular value decomposition for $\mathbf{B}$ and moreover, *Elbern et al.* [2007*a*], an efficient precondition by the diffusion approach [*Weaver and Courtier*, 2001] of the minimisation problem for the joint estimation of emission rates and chemical states. The latter is adopted here to the 4D-var assimilation algorithm, as it improves its computational behaviour.

The precondition is based on the transformation of the optimisation parameters by square roots of $\mathbf{B}$ and $\mathbf{K}$, such that $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$ and $\mathbf{K} = \mathbf{K}^{1/2}\mathbf{K}^{T/2}$ [*Courtier*, 1997]. The new variables now can be defined

$$\begin{aligned} \mathbf{v} &:= \mathbf{B}^{-1/2}\delta\mathbf{x} \\ \mathbf{w} &:= \mathbf{K}^{-1/2}\delta\mathbf{u}. \end{aligned} \tag{2.12}$$

Then the cost function Eq. (2.1) can be written as

$$J(\mathbf{v},\mathbf{w}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\sum_{t_i=t_0}^{t_N}(\mathbf{d}(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i))^T\mathbf{R}^{-1}(\mathbf{d}(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i)) \tag{2.13}$$

and the gradient (2.11) as

$$\nabla_{(v,w)^T} J = \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} - \begin{bmatrix} \mathbf{B}^{T/2} & 0 \\ 0 & \mathbf{K}^{T/2} \end{bmatrix} \sum_{i=0}^{N} \widetilde{\mathbf{M}}_i^T \mathbf{H}(t_i)\mathbf{R}^{-1}(\mathbf{d}(t_i) - \mathbf{H}(t_i)\delta\mathbf{x}(t_i)). \tag{2.14}$$

The precondition of the cost function results in an easier calculation of its gradient, which is the main benefit for the 4D-var algorithm. Now only the $\mathbf{B}^{T/2}$ and $\mathbf{K}^{T/2}$ are needed, avoiding the more demanding $\mathbf{B}^{-1}$ and $\mathbf{K}^{-1}$ and the ill-conditioning introduced by any formulation of $\mathbf{B}$ and $\mathbf{K}$. Moreover the transformation back to the increments $\delta\mathbf{x}$ and $\delta\mathbf{u}$ from Eq. (2.12) uses just the $\mathbf{B}^{1/2}$ and $\mathbf{K}^{1/2}$, giving the optimised state variables

$$\begin{aligned} \delta\mathbf{x} = \mathbf{B}^{1/2}\mathbf{v} &\Longleftrightarrow \mathbf{x}(t_0) = \mathbf{B}^{1/2}\mathbf{v} + \mathbf{x}_b(t_0) \\ \delta\mathbf{u} = \mathbf{K}^{1/2}\mathbf{w} &\Longleftrightarrow \mathbf{e}(t_0) = \mathbf{K}^{1/2}\mathbf{w} + \mathbf{e}_b(t_0). \end{aligned} \tag{2.15}$$

Chapter 3

The EURAD-IM Data Assimilation System

For the purposes of atmospheric modelling an as reliable and comprehensive as possible picture of the chemical processes is necessary. The chances that this is achieved gets always better due to the more and more advanced observational data and the model development. In case of the analysis of the tropospheric state, it is shown that the variational methods provide strong tools for this purpose [*Lahoz et al.*, 2010]. Based on this principle, the EURopean Air pollution Dispersion - Inverse Model (EURAD-IM) is developed, at the Rhenish Institute of Environmental Research at the University of Cologne (Rheinischen Institut für Umweltforschung an der Universität zu Köln - RIU, `http://www.eurad.uni-koeln.de/`). The roots of the EURAD-IM go back to the Regional Acid Deposition Model RADM2 [*Chang et al.*, 1987], which is succeeded by the EURAD chemistry transport model, developed in RIU [*Hass et al.*, 1995]. EURAD is further expanded, with the development of its adjoint version, into EURAD-IM [*Elbern et al.*, 1997; *Elbern and Schmidt*, 1999].

EURAD-IM is a Eulerian meso-scale CTM, that involves advection, diffusion, chemical transformation, wet and dry deposition and sedimentation of tropospheric trace gases and aerosols [*Memmesheimer et al.*, 2004]. It includes both three and four dimensional variational chemical data assimilation and it is able to run in nesting mode. The meteorological driver that is applied in the EURAD-IM is the Weather Research and Forecasting Model (WRF) [*Skamarock et al.*, 2008].

A wide range of applications is covered by EURAD-IM for air quality studies, aiming to understand atmospheric chemistry and air pollution. So far it has taken part in several recent pollution studies [*Marécal et al.*, 2015; *Monteiro et al.*, 2013; *Zyryanov et al.*, 2012; *Monteiro et al.*, 2012; *Elbern et al.*,

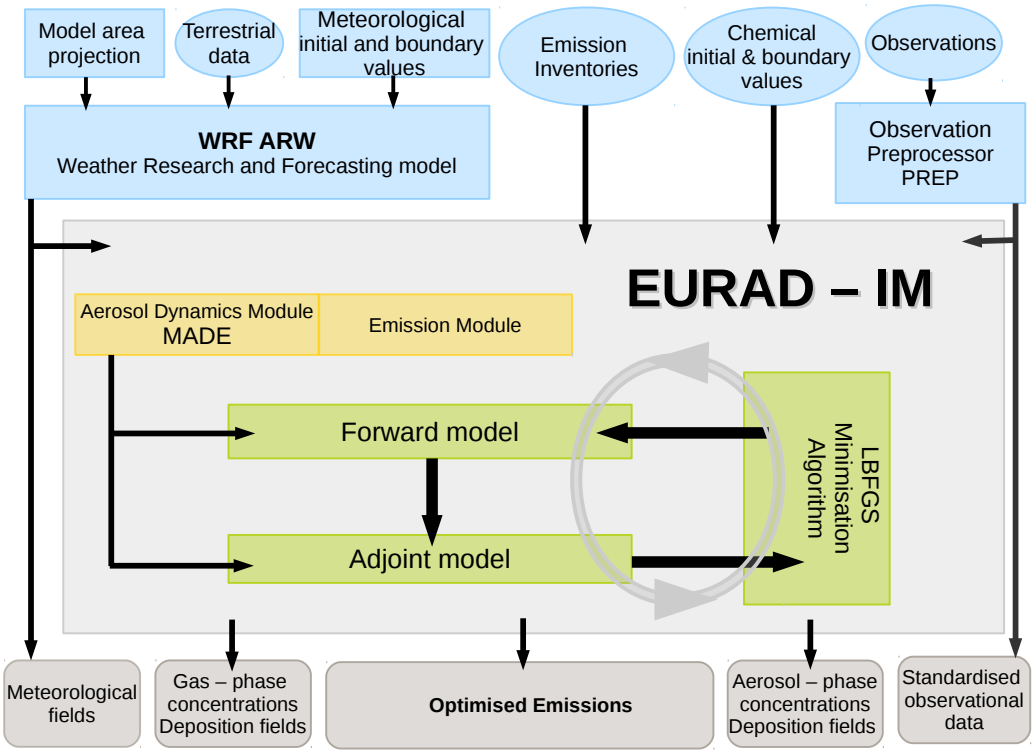

Figure 3.1: *The flow of the procedures within the EURAD-IM system.*

2011; *Kanakidou et al.*, 2011]. On the one hand the model is used for long term simulations focusing on the emission directives, including aerosols [*Li et al.*, 2013], while on the other hand it provides operational forecasts for scientific field campaigns (e.g., PEGASOS campaing), also in the framework of analysis and monitoring.

The EURAD-IM 4D-var data assimilation system consists of four parts:

- the EURAD-IM CTM and its adjoint,
- the formulation of background error covariance matrices for both the initial states and the emissions, as well as their treatment to the minimisation problem,
- the observational basis and its related error covariance matrix,
- the minimisation, including the transformation for the preconditioning.

A flowchart of all model contributors and procedures is shown in Fig. 3.1. Firstly, the meteorological, boundary, observational and emission input parameters are necessary. All the needed meteorological data are provided by WRF model. The meteorological driver works independently from the CTM, so there is no feedback from the latter to the WRF. Another offline pre-process for the observational data takes place, the observation preprocessor (PREP), in order to provide the observations information in a suitable format for the CTM. The raw emission data from the inventories, categorised by emission source, are inserted directly to the CTM, for further processing. For every first simulation of the model, the chemical initial values are given by climatological information, whereas, in any further simulation, they are provided by the previously accomplished model output.

After the initialisation, it is the evolution of the model state to be calculated and compared with the available observations. To this end, the CTM uses two modules, the online Emission Module and the Modal Aerosol Dynamics module for Europe (MADE) [*Ackermann et al.*, 1998]. The former, updated and further developed for the purposes of the current thesis, calculates the emission rates for each species across the domain, based on the raw emission data from the inventories [*Kuenen et al.*, 2014], while the latter provides information on the aerosol size distribution and chemical composition. The forward model of the EURAD-IM calculates the chemical states and provides the cost function. Here the model states of each time step are saved in temporal files for later use. Now the adjoint CTM is to be calculated. It performs a subsequent adjoint model integration backward in time, where it recalls the previously saved files, in order to reproduce the state of the model during the forward run, after comparing with the observations. This procedure leads to the calculation of the gradient of the cost function with respect to the control parameters. Finally, the cost function, as well as its gradient and the initial values of the control parameters are sent into the Limited-memory Broyden−Fletcher−Goldfarb−ShannoL minimisation algorithm (L-BFGS) [*Nocedal*, 1980; *Liu and Nocedal*, 1989], which delivers the desired new optimised control values. The global minimum of the cost function is not feasible to be found after only one minimisation step, due to the tangent linear approximation of the gradient. Instead, the system is applied iteratively, until the minimisation of the cost function from one step to the next one is lower than a certain threshold or until a maximum number of iterations is reached.

There is a large variety of EURAD-IM's outputs, available for further analysis and development, which are governed by the optimised emissions. The control parameters that are optimised by the aforementioned procedure are

the initial concentrations of the assimilated species and the emission factors. After a 4D-var run the improvement of the control parameters leads to an improved, accurate knowledge of the emission patterns and the evolution of the emission fluxes. A closer look is given in Chapter 6, where analysis of special features of the model's output is performed, based on real case studies.

In the present Chapter, a more detailed description of the features of EURAD's inverse model is presented. Firstly, an insight into the forward and the adjoint model is given. Then the crucial part of the background error covariance matrices is discussed, highlighting the innovated features of this study, and last, a description of the minimisation procedure is given.

3.1 The EURAD-IM Chemistry Transport Model and its Adjoint

3.1.1 The EURAD-IM CTM

The EURAD-IM CTM calculates the transport, diffusion and gas transformation for the tropospheric trace gases. The amount of treated species and their chemical reactions depend on the selection of chemistry mechanism. For the purposes of the current work, the selected chemistry mechanism includes 60 chemical species with 147 reactions [*Stockwell et al.*, 1997; *Schell et al.*, 2001]. Thus, a set of partial differential equations is solved by the EURAD-IM CTM, in an Eulerian framework, for the tendency of all the different species included into the mechanism. The set of partial differential equations follows

$$\frac{\partial c_i}{\partial t} = -\nabla\left(\boldsymbol{v}c_i\right) + \nabla\left(\rho\mathbf{G}\nabla\frac{c_i}{\rho}\right) + A_i + E_i - D_i, \tag{3.1}$$

where c_i is the concentration of species i, such that i=1,...,U the number of species in the mechanism, $\boldsymbol{v}$ is the wind velocity and ρ the air density. What is more, $\mathbf{G}$ a symmetric eddy diffusivity tensor, A_i the chemical transformation term for gas phase species i, while E_i the emission rate of species i and lastly, D_i is the deposition rate of species i, where $D_i = \boldsymbol{u}_i^d c_i$ and $\boldsymbol{u}_i^d$ denotes the deposition velocity for species c_i.

The chemical transformation term A_i, considering a production rate P_i and

a loss rate L_i of species i, is written as

$$\begin{aligned} A_i &= P_i + L_i \\ &= \sum_{r=1}^{R} \left(k(r) \left(s_i(r_+) - s_i(r_-) \right) \prod_{j=1}^{U} c_j^{s_j(r_-)} \right), \end{aligned} \tag{3.2}$$

where R signifies the number of chemical reactions, s $\in \mathbb{N}_0$ is a stoichiometric coefficient and *k(r)* denotes the reaction rate of reaction r; either being productive (r_+) or destructive (r_-) for species i.

In order to proceed with the solution of the partial differential Eq. (3.1), a symmetric operator splitting of the dynamic procedures takes place, which includes the chemistry solver module $\mathcal{A}$. This technique has been proved beneficial as it minimises systematic biases [*Sandu and Zhang*, 2007]. So the operator splitting scheme, when stepping from time t to time $t+\Delta t$ (*Yanenko* [1971]; *McRae et al.* [1982]), can be described by

$$x^{t+\Delta t} = T_h T_v D_v \mathcal{A} D_v T_v T_h x^t, \tag{3.3}$$

where T, D denote transport and diffusion operators in horizontal (h) or vertical (v) direction, respectively. T and D are applied for one half of the model's advective time step before the chemistry mechanism and one half after it. The gas phase chemistry module $\mathcal{A}$ includes the parametrisation of the emission sources, whereas the vertical diffusion term D_v includes the deposition processes. The gas phase dry deposition modelling follows the method proposed by *Zhang et al.* [2003], while the dry deposition of aerosol species is treated size dependent, using the resistance model of *Petroff and Zhang* [2010]. Moreover, the wet deposition of gases and aerosols is derived from the cloud model in the Environmental Protection Agency (United States EPA) Models-3 Community Multi-scale Air Quality (CMAQ) modelling system [*Roselle and Binkowski*, 1999].

The wind driven transport is calculated by the positive definite advection scheme of *Bott* [1989]. For the parametrisation of the vertical sub-grid-scale turbulent transport an eddy diffusion approach is used. Here the calculation of the vertical eddy diffusion coefficients is based on the specific turbulent structure in the individual regimes of the planetary boundary layer (PBL), according to the PBL height and the Monin-Obukhov length [*Holtslag and Nieuwstadt*, 1986]. Further, the diffusion equation is solved by a semi-implicit scheme, following Crank-Nicholson [*Crank and Nicolson*, 1947].

Chemistry mechanism and solver

The decision of the appropriate chemistry mechanism in chemistry transport models is a challenging issue, as on the one hand it should satisfy the

demand to restrict the computational costs and on the other hand to deliver a good representation of the tropospheric chemical processes. In EURAD-IM these demands are fulfilled by the Regional Atmospheric Chemistry Mechanism (RACM) [*Stockwell et al.*, 1997], and its extension, for the formation of secondary organic aerosols (SOA), including the production of low-volatility products and their subsequent gas/particle partitioning [*Schell et al.*, 2001]. Based on physical and chemical principles [*Strader et al.*, 1999; *Bowman et al.*, 1997; *Lurmann et al.*, 1997; *Pandis et al.*, 1992], the coupled RACM-SOA mechanism not only represents the gas phase and photolysis reactions, but also links the gas phase precursor species with the SOA formation, determining also the contributions of biogenic and anthropogenic components of the total SOA. The present work presents the first application of RACM-SOA in real case studies.

The gas phase system is described by a set of stiff ordinary differential equations, as in Eq. (3.1). *Sandu et al.* [1996] show that for the solution of this kind of systems it is better to use an implicit method. The system of Eq. (3.3) is solved numerically by a stage-2 Rosenbrock solver [*Sandu et al.*, 2003; *Sandu and Sander*, 2006] and its adjoint [*Elbern et al.*, 1997; *Strunk*, 2006]. The Rosenbrock method has implicit stable regions and, what is more, the evaluation of the solver's step size, Δt, only needs to fulfil the user's accuracy requirements.

The photolysis frequencies are derived using the Fast Tropospheric Ultraviolet-Visible (FTUV) model, according to *Tie et al.* [2003]. Therein, the radiative transfer model is based on the TUV developed by *Madronich and Weller* [1990].

Aerosols

EURAD-IM simulates the bidirectional transfer between gas phase and aerosols. The information on the aerosol size distribution and chemical composition is gained by the MADE. To this end, the Fully Equivalent Operational Model (FEOM) version is applied, which is refined by the High Dimensional Model Representation (HDMR) technique [*Rabitz et al.*, 1999; *Nieradzik*, 2005], of an accurate mole fraction based thermodynamic model [*Friese and Ebel*, 2010]. The simulation of the SOA formation is accomplished by the updated Secondary Organic Aerosol Model (SOGRAM) coupled with MADE [*Li et al.*, 2013]. It should be underlined that the assimilation of aerosols is out of the scopes of the present work and so it is not performed.

Initial model state

At the beginning of a simulation, EURAD-IM uses climatological data to set up the initial values. The initial configuration, thus, includes a seasonal

mean concentration of long-lived species dependent on latitude and height for the first forward model run. Since this initial time lacks of better knowledge, inflow boundary values are defined in the same way as the initial model state, or in some other applications of EURAD-IM they can come from a global CTM. All later simulation runs let the model output of a previously operated run to be considered as the initial model state. After each run, the gained information optimises the model parameters. By using these data as initial values for following simulations, the improvement and the consistency of the system are ensured.

Emissions

The EURAD-IM treats both anthropogenic and biogenic emissions [*Sofiev et al.*, 2015; *Li et al.*, 2013; *Strunk*, 2006; *Nieradzik and Elbern*, 2006]. The Emission Module within the EURAD-IM system produces improved emission data, with high temporal resolution and European coverage. It calculates online emissions by converting the annual emission rates, received by the inventories, to hourly emission rates, with the use of temporal and spatial allocation factors. Moreover, a vertical distribution of the emission rates of each emitted species takes places, based on the source of the emissions and the type of the point sources.

A new development of the present work is the coupling of the adjoint code for the emission estimation of EURAD-IM CTM with the new online Emission Module. This update increase the accuracy of the assimilation result, due to the detailed emission distribution in time and height, per source and per grid box that takes place. A short description of the emission processes also before the simulation runs of the EURAD-IM CTM follows.

For anthropogenic emissions, the TNO-MACC-II emission inventory of the year 2009 is used, with 7 km $\times$ 7 km horizontal resolution [*Kuenen et al.*, 2014]. The Model of Emissions of Gases and Aerosols from Nature (MEGAN) [*Guenther et al.*, 2012] calculates the biogenic emission input for EURAD-IM. Additionally, emissions from fires are taken into account using the Global Fire Assimilation System, GFASv1.1, product [*Kaiser et al.*, 2012], daily available with $0.1^o \times 0.1^o$ resolution.

The raw emission data from the inventories are available as annual emission rates, measuring the average amount of specific pollutants discharged into the atmosphere by a specific process (fuel/equipment/source and counted in Mg/year for every country). In case of the anthropogenic emissions, the 7 emission rates of NO_x, SO_x, CO, NH_3, particular matter ($PM_{2.5}$ and PM_{10}), as well as non-methane volatile organic compound (NMVOC) are provided for

the European emission domain at $0.125^o \times 0.0625^o$ longitude-latitude resolution, subdivided into the 10 anthropogenic source-sectors (SNAP codes, Table 3.1). The spatial disaggregation of these data is done based on the origin of the emissions. For instance, the data coming form industrial areas are covered by Corine Land Cover dataset [*Stjernholm*, 2009], whereas those that come from the road network by OpenStreetMap [*Haklay and Weber*, 2008]. The disaggregation outputs are emission files in $0.0625^o \times 0.0625^o$ resolution, including data which are no longer equally distributed in a grid cell, but over the emission source area. Only a small percentage of emission remains as a background emission. Eventually, this aggregated information is converted to a suitable for the EURAD-IM CTM format, using a GIS system. Here the ARC-GIS [*ArcGIS*] is used to adjust the projection and the resolution in the CTM's set up. After all, the input to the Emission Module of the CTM is a set of 70 files; a file with aggregation information for each one of the 7 aforementioned species and for each SNAP category of emission sources. Moreover, in order to have the most accurate emission estimation, three more files are provided to the Emission Module, giving information of the country code and the time-zone of each cell of the grid, as well as the point sources that are included.

SNAP	Sector name
1	Energy industries
2	Non-industrial combustion
34	Industry (combustion + processes)
5	Extraction and distribution of fossil fuels
6	Product use
7	Road transport
8	Non-road transport and other mobile sources
9	Waste treatment
10	Agriculture

Table 3.1: *Definition of the SNAP source categories. SNAP 3 and 4 are merged to SNAP 34 [Kuenen et al., 2014].*

Grid specifications

EURAD-IM is a mesoscale-α model, that operates on the integration domains with a boundary length of about 5000 km. In this study, the horizontal

grid is based on Lambert conformal conic projection, centred at 54^oN latitude and 12.5^oE longitude.

The horizontal grid structure of the model is defined by the Arakawa C grid stencil [*Arakawa and Lamb*, 1977]. The horizontal wind components (u,v) are staggered to all other variables of the model. This staggering allows horizontal flux terms to be evaluated relatively accurate.

A staggering of variables is also used in the vertical grid. Vertically the model includes 23 layers. The terrain follows the σ-coordinates defined as

$$\sigma_k = \frac{p_k - p_{top}}{p_{bot} - p_{top}}, \tag{3.4}$$

where p_{top} and p_{bot} is the model pressure at model top and bottom, respectively, and p_k is the pressure of the k^{th} model layer. The lowest model layer represents 38 m height, whereas the top model is defined by the isobaric level of 100 hPa and roughly corresponds to 16 km height.

The nesting technique in EURAD-IM

In large grids with low horizontal resolution, there is poor representativity of many observations of important air quality related species, as it is already mentioned in Chapter 2.

This issue is addressed into EURAD-IM system by the application of the adjoint nesting technique. The idea is the following: The simulation starts with a large but not very high resolved domain, the coarse or mother domain. The initial and boundary values of this grid are based on climatological data, or, in other cases, on a global CTM. For a more detailed study, the coarse domain of the model can be restructured in a horizontal sub-domain, with less horizontal expansion but higher resolution, the nest or daughter domain. The boundary values are interpolated from the mother domain's respective grid cells after the analysis, resulting in more reasonable data with the temporal resolution of the coarse grid. This happens because the analysis of the coarse grid for the current day already includes all the available information about the system, as it has been gained comparing with the observations of the same day, and so these boundary values are superior to any forecast resting on an analysis from the previous day.

As far as the initial data for the nest is concerned, they should include information of high resolution. That is why they are taken from precedent simulation of the same nest domain. Here, physical and chemical features, which are necessary for the nest, may not be identified by the coarse grid simulation, because of its lower resolution.

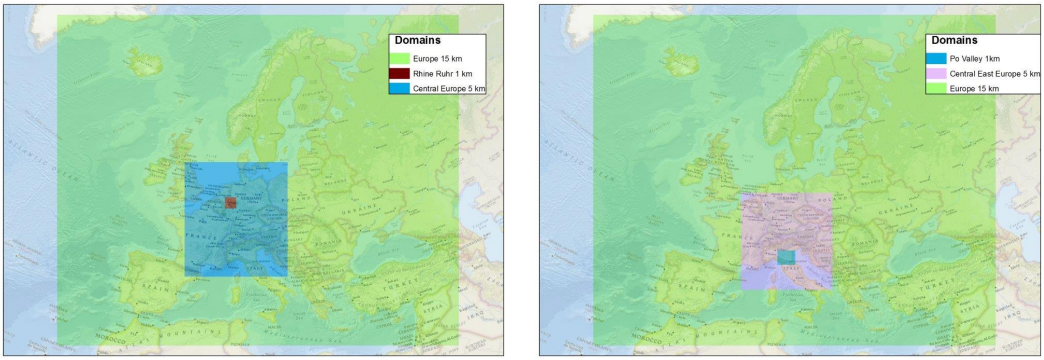

Figure 3.2: *The nesting sequence in EURAD-IM for different case studies. The coarse grid is Europe with 15 km resolution and includes two nest domains with 5 km and 1 km resolution. On the left, these are central Europe and the Rhine-Ruhr area in Germany, respectively. On the right, accordingly, the two nests are the central-eastern Europe and the area of Po valley in Italy.*

After the end of the simulation, this nest can become the mother domain for a new nesting domain, with even higher horizontal resolution, and so on.

The use of optimised initial model states for the model integration lead to improved forecast skills. In case of initial value optimisation, attention should also be paid in the size definition of the domains. In 4D-var the information of an observation is integrated backward in time and this results in correction of the model state at initial time. However, if the grid size is not large enough, or the meteorological conditions are not taken into account (i.e., strong winds), this resulting corrected position may be outside of the domain and in that case the gained information is lost. So, a balance between the domain's size and the assimilation windows should always be kept.

The nesting technique in EURAD-IM is described in *Jakobs et al.* [1995] and further extensions are discussed in *Strunk* [2006].

In the present work the horizontal resolution ranges from 15km to 1km, using three domains for the analysis of different episodes within the case studies of interest. In all of these cases the mother domain covers the whole European area, as well as parts from north Africa and west Asia, with a 15 km resolved grid. The first sequence of nests includes central Europe, covering Germany, France, Austria and parts from Poland and Italy in a 5 km resolution domain. A nest for the Rhine-Ruhr area in North Rhine Westphalia, in Germany, with 1 km resolution is applied. A second pair of nests includes a 5 km resolved domain over Italy, Germany, Austria, as well as an 1 km resolved domain, focused on the area of the Po valley in northern Italy. The aforementioned

nesting sequence of domains is shown in Fig. 3.2.

3.1.2 The adjoint EURAD-IM CTM

In order to develop the adjoint of the EURAD-IM model, the evolution of the concentrations over time is necessary, taking into account the transport, diffusion, deposition and gas transformation processes of the chemical species.

Thus, the adjoint calculation either is provided by the numerical solvers of the forward model Eq. (3.1), or by their adjoint differential equations, after the application of the variational calculus, such ass

$$\begin{aligned}\frac{\partial \delta c_i^*}{\partial t} &= -\boldsymbol{v}\nabla\delta c_i^* - \frac{1}{\rho}\nabla\left(\rho\mathbf{G}\nabla\delta c_i^*\right) + D_i^* + A_i^* \\ &= -\boldsymbol{v}\nabla\delta c_i^* - \frac{1}{\rho}\nabla\left(\rho\mathbf{G}\nabla\delta c_i^*\right) + \boldsymbol{u}_i^d c_i^* \\ &\quad + \sum_{r=1}^{R}\left(k\left(r\right)\frac{s_i\left(r_-\right)}{c_i}\prod_{j=1}^{U} c_j^{s_j(r_-)}\sum_{n=1}^{U}\left[s_n\left(r_+\right) - s_n\left(r_-\right)\right]\delta c_n^*\right). \qquad (3.5)\end{aligned}$$

Here δc_i^* is the adjoint variable of c_i. It should be noticed that the emission strength is independent of the species concentration c_i and so the term E_i of 3.1 does not take part into the adjoint formulation.

However, it is already discussed that the changes of the species concentrations over time at the forward code are efficiently described by the operator splitting scheme Eq. (3.3). The adjoint EURAD-IM CTM is developed by this approach, where the operators T_h, T_v,D_v and $\mathcal{A}$ are replaced by their adjoint counterparts.

For the construction of the adjoint chemistry the Kinetic Pre-Processor KPP [*Damian-Iordache*, 1996] is adopted. Within EURAD-IM the coding of the chemistry solver is based on KKP Version 2.1 by *Sandu and Sander* [2006] and its correctness is verified by the proposed method of *Chao and Chang* [1992].

3.2 Background error covariance matrices

In Chapter 2, where the formulation of the minimisation problem is given, the background field of the model, $\boldsymbol{x}_b(t)$, is defined as the representation of the best estimate of the current atmospheric state, before any use of observations

at time t. Similarly, the background emission rates, $\boldsymbol{e}_b(t)$, denote the initial information about the emissions, based on the emission inventory, before the model's calculation of the emission rates at the same time. As with all information, the background state is prone to errors, which must be accounted by the variational data assimilation. The most convenient way to deal with this issue is through the background error covariance statistics, which in case of initial species concentrations is represented by matrix $\mathbf{B}$, in the cost function Eq. (2.9), whereas in case of emission rates and observations by matrices $\mathbf{K}$ and $\mathbf{R}$, respectively.

The estimation and the construction of a realistic representation of the background error covariance matrices is one of the challenges of 4D-var data assimilation. An accurate estimation of the background error covariance matrices takes advantage of single observations to spread information in the model state, even if the observations have sparse spatial coverage.

The background error covariance matrices are symmetric and positive definite matrices. Generally, their diagonal elements represent the variances, which are the auto-covariances between grid points corresponding to a particular model variable. The off-diagonal elements represent the covariances, namely the remaining elements, that represent the cross-covariances between grid points, corresponding to different model variables. The first consists the univariate part of the matrices and the second the multivariate one; both necessary in finding the minimum of the cost function.

However, this construction implies that, for the exact definition of the background error covariance matrices, the knowledge of the true state of the atmosphere at all times and everywhere in the model's domain is required. Certainly, this is not possible. The background error covariance matrices cannot be calculated completely, thus, they have to be estimated and modelled as approximations to the true covariances of background error.

Within the 4D-var algorithm of EURAD-IM, two assumptions are necessary to make attainable the calculation of the covariances. Firstly, it is supposed that there are no correlations between the concentration of the initial values and the emission rates. Secondly, the correlations between the initial concentrations and the observations are also omitted. That is why the 4D-var problem includes the three independent covariance matrices, $\mathbf{B}$, $\mathbf{R}$, $\mathbf{K}$. In the following, the description and estimation of matrices $\mathbf{B}$ and $\mathbf{K}$ takes place, whereas, for matrix $\mathbf{R}$ this is done in Chapter 4.

3.2.1 The Background Error Covariance Matrix for the initial values

The error covariance matrix of the background values of the chemical constituents is $\mathbf{B} \in \mathbb{R}^{N \times N}$, defined for the dimension N^2 of the model space, which in EURAD-IM is of $\mathcal{O}(10^{12})$. Matrix $\mathbf{B}$ is symmetric with all of its eigenvalues being positive, and contains the correlations between the different components of the state vector $\boldsymbol{x}(t)$, as well as the weights of the accuracy of the preliminary knowledge. Yet, the high storage requirements forbids the complete estimation of $\mathbf{B}$. *Elbern and Schmidt* [2001] deal first with this issue, limiting $\mathbf{B}$ to a few principal components only, such as leading singular vectors. Still, the preconditioning of the minimisation procedure Eq. (2.12) requires the square root of the covariance matrix and so a proper covariance model operator is more preferred than a full matrix. Hence the simplified implementation of a sequence of operators for $\mathbf{B}$ is used, as it is proposed in *Weaver and Courtier* [2001].

For this, $\mathbf{B}$ is decomposed in a diagonal matrix Σ, where it contains the estimates of the background standard error deviations, and a correlation matrix $\mathbf{C}$, such that

$$\mathbf{B} = \Sigma \; \mathbf{C} \; \Sigma. \tag{3.6}$$

Following the transformation of $\mathbf{B}$ in Chapter 2.5 and a suitable factorisation of the correlation matrix $\mathbf{C}$, the former equation gives

$$\begin{aligned} \mathbf{B} &= \mathbf{B}^{1/2}\mathbf{B}^{T/2} \\ &= \Sigma \; \mathbf{C}^{1/2} \; \mathbf{C}^{T/2} \; \Sigma. \end{aligned} \tag{3.7}$$

The background error deviations of matrix Σ are assumed to be dependent on species and height. Namely, the deviations are based on the degree of variability of the individual constituents and on the increasing height of the species concentrations calculated by the model; the higher the species concentration, the smaller the confidence in knowledge on the chemical state of the specific species [*Elbern et al.*, 2007*a*].

The correlation matrix $\mathbf{C}$ is equivalent with a sequence of operators that include the diffusion information, as it is implemented in *Strunk* [2006] and further defined in *Elbern et al.* [2007*a*]. In the latter, the operator splitting scheme of $\mathbf{C}$, for the calculation of Eq. (3.7), is given

$$\mathbf{C}^{1/2} = \Lambda \mathbf{L}_v^{1/2} \mathbf{L}_h^{1/2} \mathbf{W}^{1/2}, \tag{3.8}$$

where, Λ is a normalisation operator, $\mathbf{L}_{h,v}$ are the horizontal and vertical diffusion operators, respectively, and $\mathbf{W}$ is a diagonal matrix of correction factors needed for the grid (it is account for the changing heights of the grid cells due to the application of σ-coordinates). Λ is a diagonal matrix needed to counteract the flattening due to the action of the diffusion operators $\mathbf{L}$. The formulation of $\mathbf{C}$ matrix in Eq. (3.8) allows the diffusion characteristics, i.e., the correlation length (L), to be designed by the diffusion coefficients (κ) and the integration time (T), and play the essential role in calculation the covariances $\mathbf{B}$. *Elbern et al.* [2007*a*] defines the radii of influence and the appropriate time stepping for the horizontal and vertical diffusion. Here it is assumed that the horizontal radii of influence are defined as

$$L = \sqrt{2\kappa T}. \tag{3.9}$$

3.2.2 The Background Error Covariance Matrix for the emission factors

The background emission rate covariance matrix $\mathbf{K} \in \mathbb{R}^{E\times E}$ is defined for the dimension E of partial phase space of the emission rates and describes the emission rate covariances between emitted species at each location. As with matrix $\mathbf{B}$, a suitable and accurate $\mathbf{K}$ optimises the precondition of the minimisation problem. However, in contrast with $\mathbf{B}$, for the construction of the background emission rate covariance matrix two important assumptions are made; the spatial correlations between the emission rates are not taken into account and, further, all the multivariate correlation between the emitted species in one grid cell maintain the same over the whole domain. However, given relative information both assumptions can be relaxed.

Following the decomposition of matrix $\mathbf{B}$, the emission factor error covariance matrix $\mathbf{K}$ is factorised as

$$\mathbf{K} = \Gamma\ \mathbf{D}^{1/2}\ \mathbf{D}^{T/2}\ \Gamma, \tag{3.10}$$

where Γ is a diagonal matrix containing the standard error deviations of the emission factors and $\mathbf{D}$ is the background error correlation matrix. In the followings, a closer look to both of them is taken.

In Chapter 2.3 a detailed view of what are the emission factors and how are they estimated is given, leading to Eq. (2.7). The model, in the beginning of the assimilation, initialises all the emission factors by a factor equal to one. Based on this information, the standard deviations of Γ matrix actually depict the default errors bars of the emission factors of the emitted species.

Presently, emitted species such as SO_2, NO and NH_3 are better evaluated than others, for instance CO, and thus their standard deviations can be better assigned. This fact is highlighted in *Strunk* [2006] and the entries of matrix Γ are given. However, in the present thesis is concluded that larger error bars are needed for the background emission factors, giving the opportunity to the 4D-var algorithm to take into account the observation information. The content of the updated matrix Γ, for the current configuration of the model, is shown in the Table 3.2.

Species	Standard deviations ($\delta\mathbf{u} = \ln f$)
SO_2	10.6
NH_3	10.6
NO	5.3
others	13.9

Table 3.2: *The standard deviations of the emission factors for different kind of species; the content of the diagonal matrix* Γ.

The positive definite background error correlation matrix **D** contains the correlations between emitted species. How they are correlated is a matter of the annual amount of emissions, documented by the TNO inventory and, furthermore, of their chemical evolution. The emission inventory provides the annual amount of NO_x, SO_x, and VOC's, where they are divided into different polluter groups, i.e., emission of anthropogenic or biogenic origin. The matrix **D** contains the percentage of the correlations between the emitted species, that the model takes into account for the contraction of the emission factor background error covariance matrix, as it is shown in Fig. 3.3. Here, the definition of **D** in *Elbern et al.* [2007*a*] is further expanded by the current thesis. Five species of biogenic origin are added in EURAD-IM, making matrix **D**, and consequently the background error covariance matrix **K**, more representative of the tropospheric chemical procedures. Figure 3.3 illustrates the upper triangle sub-matrix of **D**, containing 24 emitted species (NO_x, SO_x and VOC's). High correlations (larger than 10%) are achieved either between species that have the similar origin (i.e., SO_x - NO_x by anthropogenic sources), or between species that are chemically related to each other (i.e., NO_2 - NO) and between families of VOC's (i.e., ethylene (ETE) and olefine (OLT or OLI)). On the other hand, low correlation rates (lower than 1%) have species that come from different sources, as for example isoprene (ISO) and SO_x; the first is emitted by plants and the second is a result of the human activity. Finally, reasonably correlated species are those of biogenic origin that come from different types

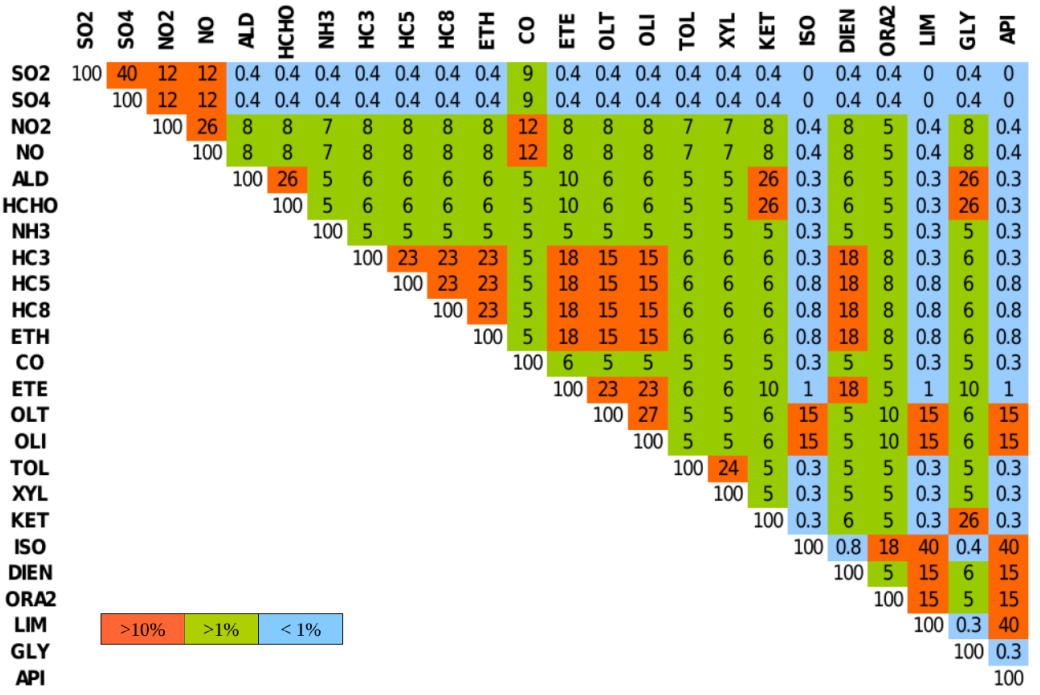

Figure 3.3: *The background error correlation matrix* ***D*** *includes the correlations between 24 emitted species. High correlated species, more than 10%, are given in red, species with correlations between 1% and 10% in green, and in blue are represented correlations less than 1%.*

of vegetation.

It is important that the background error covariance matrices are accurately defined. In case of the emission factors and matrix **K**, continuous efforts are made to improve the estimation of the background error correlation matrix, in collaboration with the emission experts of TNO. Since in reality the correlations between the emitted species can variate within the domain, a future update could also include the spatial variations of the emissions and further, the construction of matrix **D** dependent on the SNAP source categories.

3.3 The minimisation

The cost function of the 4D-var algorithm depicts the distance between the model's output and the observation state of the atmosphere. The optimal result comes from the minimisation of this distance. For so, the calculation of

the gradient of the the cost function with respect to the control parameters, the initial values of the chemical constituents and the emission factors, is needed.

The precondition of the cost function as it is discussed in Chapter 2.5, provides a computationally less expensive way to proceed with the problem. The transformation of the optimisation parameters of Eq. (2.12) includes, for both the initial concentrations and the emission rates, the deviation from the background state, $\boldsymbol{x}_b(t)$ and $\boldsymbol{e}_b(t)$, respectively. In the beginning of the assimilation, the aforementioned deviations, $\delta\mathbf{x}$ and $\delta\mathbf{u}$, are set to zero, as the first guess is identical to the background state. Then the initial state is updated by the analysed state values. Next, it is the turn of the cost function Eq. (2.13) to be calculated, as well as its gradient Eq. (2.14), resulting after the backward (i.e., adjoint) integration. The state that minimises the cost function is found by solving the equations $\nabla_{(v,w)^T} J = 0$. Yet, the size of the problem requires a numerical method to approximate the solution. Thus, the EURAD-IM 4D-var system uses a parallelised version of the L-BFGS optimisation algorithm [*Nocedal*, 1980; *Liu and Nocedal*, 1989].

The Limited memory BFGS optimisation algorithm belongs to the family of quasi-Newton methods that is based on the Broyden - Fletcher - Goldfarb - Shanno (BFGS) algorithm using a limited amount of computer memory. In L-BFGS the Hessian matrix H of the classical Newton method,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \boldsymbol{H}^{-1}(\mathbf{x}_k)\nabla_{(v,w)} J,$$

is approximated iteratively, where L-BFGS stores only a few vectors that represent the approximation implicitly.

The minimisation procedure into EURAD-IM algorithm contains a sequence of steps, which are iteratively repeated until the optimum state is reached. The forward model run calculates the cost function, based on an a priori or updated state of the optimisation variables, as described above. Then the adjoint integration follows, calculating the transformed gradient Eq. (2.14). For this calculation the horizontal and vertical diffusion is applied, for the half of the time steps defined in Eq. (3.9). Further, it is the turn of the L-BFGS to get as input values the a priori or updated state of the optimisation variables, the cost function and its gradient, and then to calculate the optimal $(\mathbf{v}, \mathbf{w})$, which are saved for the following iteration run. Now the transformation back to $(\delta\mathbf{x}, \delta\mathbf{u})$ by the Eq. (2.15) takes place, giving the improved variables $(\mathbf{x}, \mathbf{e})$.

The procedure stops when the iteration limit is reached (defined by the user) or when the cost function cannot be minimised more and, thus, the total minimum is achieved.

Chapter 4

Observations

The present chapter deals with the treatment of the observed information within the EURAD-IM. The characteristics of the available observations assimilated in this work and the way they are ingested in the model by observation operators is discussed, as well as the role of the observation error covariance matrix.

4.1 The available observational data

Despite the great improvement of the observation methods, the measured snapshots of the atmospheric air are sparse, especially when the spatial resolution of a model is considered. A way to deal with this issue, is the combination of abundant sets of different kinds of observation, which can give a more complete picture of the puzzle of the atmospheric processes, if available.

Within the scope of this thesis, in-situ, satellite and airborne observations are combined by 4D-var assimilations. In-situ and satellite observations are assimilated in case of an aestival high ozone and a hibernal high stagnant air episodes, as it is presented in Chapter 5. Moreover, the benefits of campaign data are studied in case of the PEGASOS flight campaign, in combination with ground based data and satellite retrieved information. Table 4.1 shows the observational data that are used for each case study. In the current section, a description of the observational data assimilated and the characteristics of the measurements are given.

	Summertime ozone episode 2010	Wintertime aerosol episode 2012	PEGASOS flight campaign 2012
EEA in-situ sations (NO, NO_2, SO_2, O_3, CO)	✓	✓	✓
GOME-2 (NO_2)	✓	✓	✓
SCIAMACHY (NO_2)	✓		
IASI (O_3)	✓	✓	
OMI (NO_2)	✓	✓	✓
MOPITT (CO)	✓	✓	
PEGASOS airborne data (NO, NO_2, O_3, CO)			✓

Table 4.1: *The type and the origin of the assimilated observational data for each case study.*

4.1.1 In-situ observations

The provider of the surface in-situ observations assimilated in this study is the European Environmental Agency (EEA - `http://www.eea.europa.eu`). The EEA data base contains air quality monitoring data and information submitted by all countries of the European Union and some EEA potential candidate countries. The air quality database consists of multi-annual time series of air quality measurement data and their statistics for a representative selection of stations and for a number of pollutants. Further, it includes meta-information on the involved monitoring networks, their station characteristics and their measurements.

The density of the network of ground stations depends on each country and the geographical location, being partly dense, partly coarse. The center of Europe is well covered by in-situ stations from EEA, so for the domains under considerations in this thesis there is a large data base of ground observations. All types of stations (i.e., background, industrial, traffic stations) are included, summing up to roughly 13.000 observations for an average day, among which

the 37% is NO_2 observations, the 30% is O_3, the 20% is NO and the 13% is SO_2 observations. In general, the NO_x data are available in half-hourly frequency, whereas the rest of the data in hourly frequency. The in-situ network also provides particular matter observations, $PM_{2.5}$ and PM_{10}. However, these data are not assimilated in the current work.

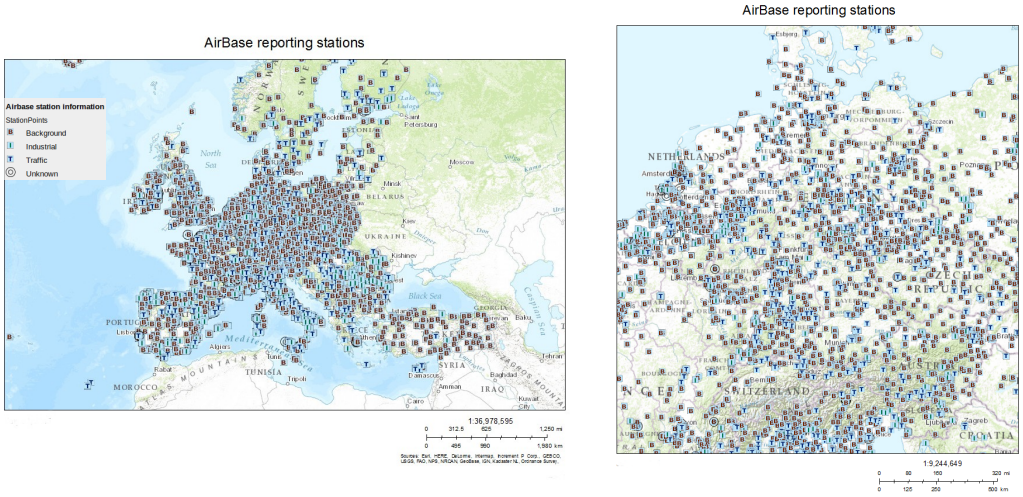

Figure 4.1: *The coverage of ground stations in Europe (left) and Germany (right) by the Airbase data; a dense network of different kind of observation stations; background, industrial, traffic (source:http://acm.eionet.europa.eu/databases/airbase).*

4.1.2 Satellite observations

Tropospheric satellite retrievals for gas phase species are considered to enhance the assimilation result, supplementing the ground network of stations during the three case studies of interest.

For the satellite data in use, a short description for each satellite instrument and its observed values follows.

Measurement Of Pollution In The Troposphere - MOPITT
MOPITT [*Deeter et al.*, 2010; 2013] is an instrument flying on NASA's Earth Observing System Terra spacecraft, sensitive to free tropospheric CO (Fig. 4.2). MOPITT is the first satellite sensor to use gas correlation spectroscopy calculating total column observations and profiles of CO in the lower atmosphere. The high spatial resolution, of 22 km at nadir, allows MOPITT to

focus on the distribution, transport, sources, and sinks of CO in troposphere. Hence, MOPITT CO data have been used in numerous research works for the estimation of CO emissions and analysis of the tropospheric chemical composition (e.g., *Miyazaki et al.* [2015]; *Hooghiemstra et al.* [2012; 2011]).

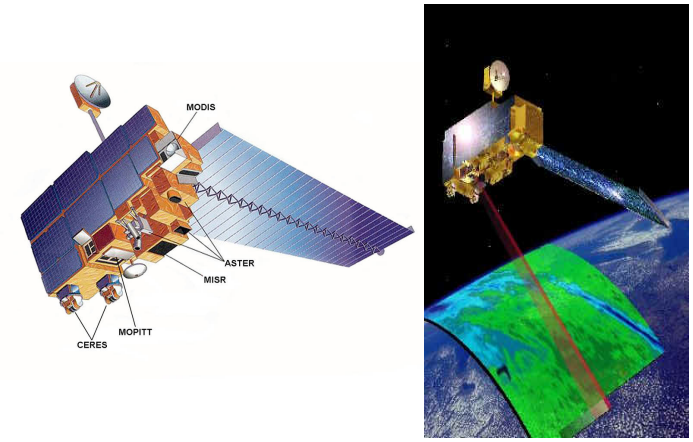

Figure 4.2: *On the left, the Terra spacecraft with its different remote sensing instruments. On the right, the path track of MOPITT instrument on board of the Terra. Source:* `http://terra.nasa.gov`

Ozone Monitoring Instrument - OMI

OMI [*Levelt et al.*, 2006] is a key instrument on Earth Observing System (EOS) Aura (Fig. 4.3), contributed by the Netherlands's Agency for Aerospace Programs in collaboration with the Finnish Meteorological Institute, for monitoring the recovery of the ozone layer in response to the phase out chlorofluorocarbons (CFCs). OMI measures both direct and atmospehre-backscatter sunlight in the ultraviolet and visible radiances, fact that improves the accuracy and precision of the measured total ozone amounts. It also measures BrO, formaldehyde, and OClO which all play a role in chemistry of the stratosphere and troposphere, as well as SO_2, and aerosols in Near Real Time (NRT). However, its most common used measurements is NO_2 [*Miyazaki et al.*, 2015; *Inness et al.*, 2015; *Pierce et al.*, 2010], as OMI can detect regional variability [*Bechle et al.*, 2013].

Global Ozone Monitoring Experiment2 - GOME-2

The GOME-2 [*Callies et al.*, 2000] is one of the new-generation European instruments (Fig. 4.4) carried on MetOp satellite. It provides information

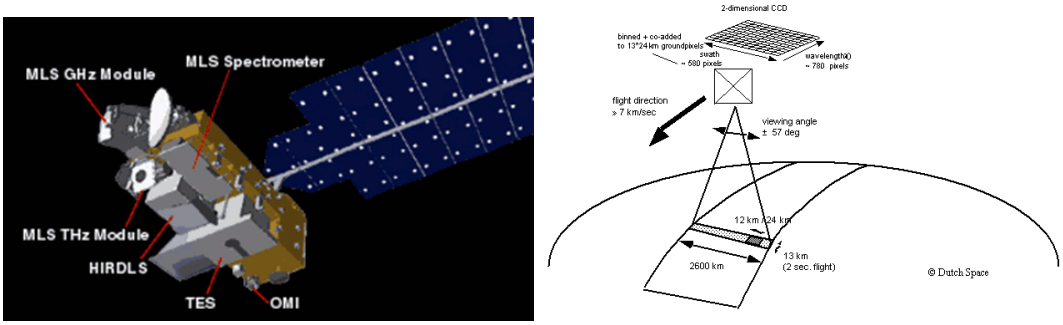

Figure 4.3: *On the left, the Aura spacecraft and the instruments attached on it. On the right, the principle of OMI. Source:* `http://www.nasa.gov/mission_pages/aura/spacecraft`

for traces gases and aerosols, though, of high importance are the GOME-2 profiles of O_3 and NO_2, as they are representative of the lowermost 50 km of the Earth's atmosphere. GOME-2 NO_2 tropospheric column concentrations have been used in many recent studies, as *Inness et al.* [2015], *Cuesta et al.* [2013], *Miyazaki et al.* [2012], *Richter et al.* [2011].

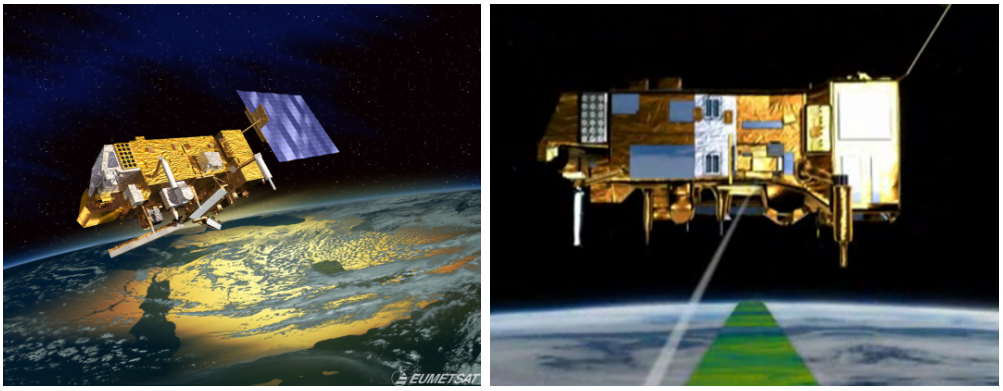

Figure 4.4: *On the left, the MetOp satellite, while on the right, the scanning principle of GOME-2 instrument. Source:* `http://www.esa.int/Our_Activities/Observing_the_Earth/The_Living_Planet_Programme/Meteorological_missions/MetOp/About_GOME-2`

Infrared Atmospheric Sounding Interferometer - IASI

IASI [*Clerbaux et al.*, 2009] is a hyperspectral infrared sounder residing on the European Space Agency's MetOp series of polar orbiting satellites (Fig. 4.5), for the purpose of supporting Numerical Weather prediction. IASI measures in the infrared part of the electromagnetic spectrum at a horizontal resolution of

12 km over a swath width of about 2.200 km, as well as volcanic ash detection. Water Vapor, O_3, CO, CO_2 and CH_4 are the high resolved observed constituents. IASI is a pioneer sensor sensitive to tropospheric O_3 [*Emili et al.*, 2014], important for air quality studies and inverse modelling applications [*Bocquet et al.*, 2015; *Safieddine et al.*, 2014; *Cuesta et al.*, 2013; *Cooper et al.*, 2014; *Barret et al.*, 2011].

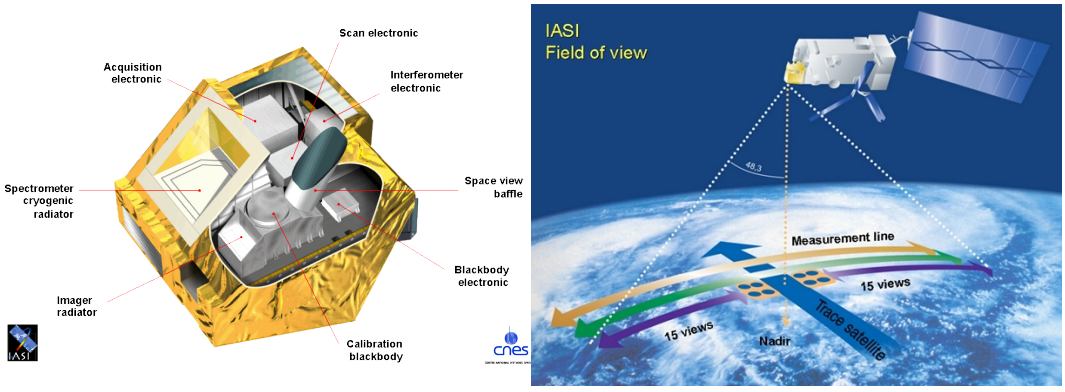

Figure 4.5: *On the left, the anatomy of IASI remote sensing instrument, on the right, its field of view.* `http://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Metop/MetopDesign/IASI/index.html`

SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY - SCIAMACHY

SCIAMACHY [*Bovensmann et al.*, 1999] was an imaging spectrometer (Fig. 4.6), aboard the Environmental Satellite (ENVISAT), during the years 2007-2012, launched by ESA. SCIAMACHY's high resolution and wide wavelength range, was the key to detect many different trace gases despite low concentrations, as well as clouds and aerosols. Applications of data assimilation focus on the use of the NO_2 tropospheric columns from SCIAMACHY [*Inness et al.*, 2015; *Zyryanov et al.*, 2012; *Chai et al.*, 2009; *Lahoz et al.*, 2010; *Elbern et al.*, 2007*b*].

4.1.3 Air-borne campaign observations

While routine data provide continue information about the air quality in the location of the station, campaign data can be the extended insight to this

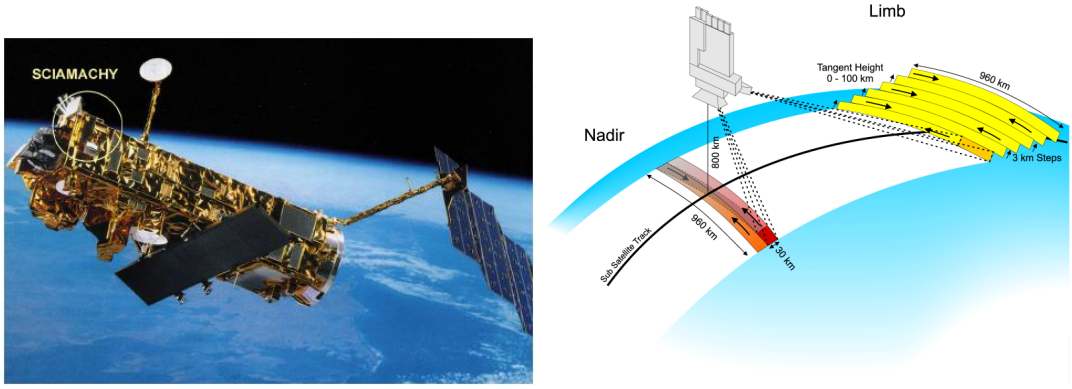

Figure 4.6: *On the left, the SCIAMACHY aboard ENVISAT, on the right, its field of view. Source:* `http://www.sciamachy.org/products/index.php?species=NO2`

information [*Zhang et al.*, 2012]. Airborne observations are provided by different kind of sources, such as aircrafts, helicopters, Zeppelins and balloons. Among them, the Zeppelin NT has a pioneering role in the collection of high quality airborne data, acting as a measuring platform to investigate the troposphere. The main advantage of the Zeppelin NT is its ability to fly with low velocity in low altitudes collecting measurements over land. Thus it is ideal for characterising the PBL in terms of height profiles, local chemical processes but also dynamic processes in the PBL evolution.

The PEGASOS campaign

In 2012 a Zeppelin NT embarked on the Pan-European gas-aerosols-climate interaction study (PEGASOS), a large scale integrating project with 26 European partner organizations [*Pandis*, 2010; *FORTH/ICE-HT*]. The objective of the PEGASOS campaign is the better understanding on the interactions of anthropogenic and biogenic components of the atmosphere and their impact on air quality and climate change. The campaign was divided in two parts where the first took place in Netherlands from 19 May to 27 May 2012, followed by a southbound trip to Italy's Po-valley from 18 June to 13 July. The second part was a two-month trip from Friedrichshafen to Finland in spring 2013.

For the scientific experiments of the PEGASOS campaign a special platform was trapped to the top of the Zeppelin NT and more instruments were installed inside the gondola (Fig. 4.7). This configuration could hold a permanent equipment and also different cabin lay outs depending the research question for each flight. So a unique data set with comprehensive air-quality measurements could be created, including photochemistry trace gases, HO_x

Figure 4.7: *The Zeppelin NT during the PEGASOS campaign. There are highlighted the sensors (black) for meteorological parameters and the installed measuring instruments in the gondola (green) and the top platform (red). Source: Gomm [2014].*

radical concentration, secondary organic aerosols, particulate matter, actinic flux density, as well as meteorological parameters and inflight aviation data (Table 4.2) [*Jäger*, 2014].

The present study focuses on the data retrieved in the area of Po valley during 10-12 July 2012, assimilating airborne measurements of NO, NO_2, O_3 and CO.

Po valley is located in northern Italy close to Bologna and is considered as a sensible hot spot, being a major populated area and subject to strong anthropogenic pollution [*Finardi et al.*, 2014]. The regional background ozone concentrations often exceed the limit values, because of high anthropogenic emissions, abundant solar radiation and stagnant air layering that characterise the area [*Liu et al.*, 2007].

During the PEGASOS campaign the airship Zeppelin flew for 25 days over the Po valley, from 18.06.2012 - 13.07.2012. Though, a special focus is placed on the 12^{th} of July. At this date the airship performed a series of near-surface vertical flights in a spiral mode, starting at 50 m and reaching roughly the 750 m above sea level, in the morning between 04:30 UTC and 08:45 UTC. The flights were performed near the field site of San Pietro Capofiume (SPC, $44^o41^{'}$N, $11^o38^{'}$E), with urban background characteristics. Thus, the airborne measurements can be compared with the in-situ observations and provide information for the vertical mixing. The meteorological conditions assess the collection of an extensive set of measurements, which includes concentrations of HCHO (formaldehyde), VOC's, OH, HO_2, NO, NO_2, O_3, CO, HONO and also OH reactivity, particle concentration/size distribution, solar actinic flux densities, temperature, pressure, relative humidity, 3-D wind [*Li et al.*, 2014].

Platform	Measurements
Top platform equipment	OH, HO_2, OH lifetime, actinic flux
Cabin: permanent equipment	NO_x, O_3, CO, particle size distributions, actinix flux, meteorological data
Cabin: photochemistry layout	HONO, formaldehyde, VOCs
Cabin: SOA layout	chemical aerosol composition, hydroscopicity, soot, VOCs
Cabin: nucleation layout	natural ions, distribution of natural ions

Table 4.2: *The total amount of measured species from all kind of equipment during PEGASOS campaign.*

These high quality data and mainly the OH measurements and reactivity make the 12^{th} of July special day of the campaign. The analysis of these data sheds light on crucial issues for the air quality, as the investigation of HONO [*Li et al.*, 2014] and HCHO [*Kaiser et al.*, 2015] sources, and the distribution of trace gases in the PBL.

The vertical flights of the Zeppelin NT cross the PBL (Fig. 4.8) providing a clear picture of its chemical composition. For better understanding of the chemical processes that take place in PBL and the information percieved by the PEGASOS airborne measurements, it is essential to discuss in more depths the development and the layering of the PBL.

The planetary boundary layer (PBL)

The PBL extends upward from the surface to a height that ranges from 100 to 3000 m. In a high pressure system, the PBL is thinner and has a well defined structure that evolves with the diurnal cycle [*Stull*, 1988]. In Fig. 4.8 the main parts of this structure are shown; the mixing layer (ML), the residual layer (RL), the nocturnal boundary layer (NBL). The bottom of the boundary layer, called surface layer (SL), has a well mixed maximum height of 50 m during a whole diurnal cycle and contains high concentrations of primary emissions.

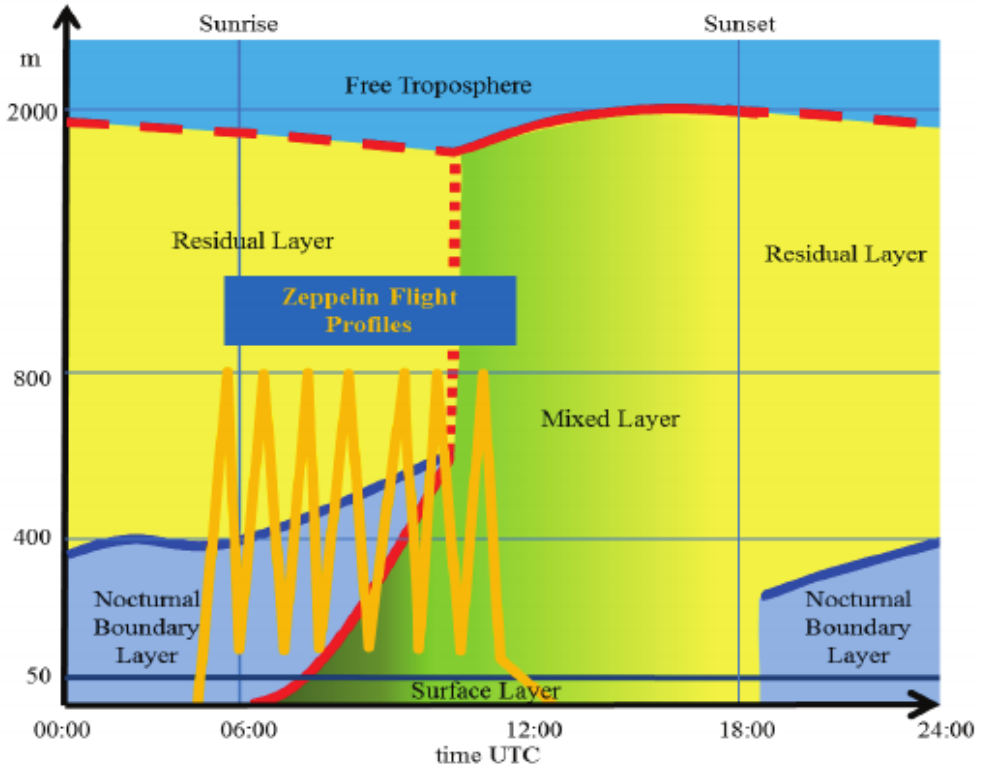

Figure 4.8: *The evolution of the planetary boundary layer (PBL) in 24h, including the Zeppelin flight profiles. Source: Jäger [2014].*

On the other hand, the free troposphere, i.e., the air above the PBL, is not influenced by the procedures on the surface.

Half an hour after sunrise the turbulent ML starts to grow, because of the surface heating of the solar radiation. The warm air rises from ground, resulting in intense mixing, until the late afternoon, where ML reaches its maximum depth. In the high pressure system of the current episode, the pollutants are trapped below ML. Subsequently, half hour before the sunset the RL is formed, without having direct contact with the ground, at 400 m - 2000 m height. The air in RL has almost the same composition as in ML, as only same oxidation or deposition takes place, mainly by O_3 and NO_3. Simultaneously, after the sunset the surface radiative cooling begins and the NBL is formed, up to 400 m height. NBL is characterised by stable air, due to low winds, which are enhanced as the layer develops, resulting in potential mixing through the NBL.

4.2 The observation operator

The observation operator $H(t)$ maps the model output, for specific space and time coordinates, to the observation space, in order to compare it with the available measurements. In the typical case of differing locations of grid points for an observation, this means that the observation operator interpolates the gridded information to the measurement location. In case of remote sensing data, H calculates the model equivalent of the observations, by radiative transfer calculations.

In Chapter 2.4, where the minimisation problem is presented, the cost function with respect to the control parameters includes the linearised approximation of the forward observation operator, $\mathbf{H}(t)$. Also, the gradient of the cost function holds the adjoint of the tangent linear observation operator $\mathbf{H}^T$. This approach guarantees that the cost function (Eq. (2.13)) is quadratic and therefore it holds a unique minimum [*Lahoz and Schneider*, 2014].

Another key point is that for every different kind of observational data, a different observation operator should be available. Hence, the design of the appropriate operator is demanding and the ability to introduce the coded knowledge of the observation into the model depends on its success. The key idea for the construction of an observation operator is the simulation of instrument's observation, by mapping the inputs in model space, and taking into account the physics of the measurement and the characteristics of the instrument. For instance, the observation operator of satellite measurements uses the averaging kernel information in observation space, which is the matrix that contains the sensitivity of the retrieval result with respect to small changes in the modelled trace gas distribution [*Rodgers*, 2000; *Strunk*, 2006].

Observations can contain information at smaller scales than the model can resolve. Thus, the errors of representativity are the result of small scale information in observations being incorrectly represented in the model.

4.3 The observation error covariance matrix

The observation error covariance matrix has a key role in the calculation of the observational cost of (2.13) and its adjoint (2.14). It is defined as $\mathbf{R} \in \mathbb{R}^{M(t)\times M(t)}$, where $M(t)$ is the dimension of the observation space at time t. $\mathbf{R}$ is given by the sum of two other covariance matrices; the measurement error covariance matrix and the representativeness error covariance matrix. Within the EURAD-IM, the observation error covariance matrix is assumed to be diagonal, which means that the observation errors are not correlated.

Thus, the diagonal elements of $\mathbf{R}$ are given as sum of the measurements and representativeness variances,

$$R_{ii} = r_{i,meas} + r_{i,repr}, \quad i = 1, ..., n_{obs}. \tag{4.1}$$

The *measurement error covariance matrix* contains the estimates of the statistical performance of the measuring instruments. It is implied that the institutions which collect and process the observed data are responsible for the validity of the measurement error covariance matrix. Nevertheless, it is often that this is not the case as, for example, ground based observations miss specification of the standard deviations. Due to these gaps of observation information, a minimum absolute (ε^{abs}_{min}) and relative error (ε^{rel}) for each species observation (y^o) should be introduced, giving the measurement error in any case,

$$r_{meas} = max\left(\varepsilon^{abs}_{min}, \varepsilon^{rel} \times y^o\right). \tag{4.2}$$

Consequently, the measurement error covariance matrix is constructed following the scheme exposed in *Mohnen* [1999].

The *representativeness error covariance matrix* holds the information of how representative is an observation for the volume-average value. The representativeness of an observation depends on the grid resolution and on the characteristics of the location, which means whether a location is remote, rural, suburban, urban or traffic. For example, CO observations, from a ground station close to a point source, do not represent the level of the CO concentrations for a whole low resolved grid. This issue is addressed by the increase of the grid resolution and the application of the nesting technique in 4D-var assimilation algorithm. Hence, the coding of the representativeness error includes a characteristic absolute error for each measured species (ε^{abs}), which is scaled by a factor that depends on the grid resolution (Δx) and a characteristic representativeness length of influence for each station type (L_{repr}), as

$$r_{repr} = \sqrt{\frac{\Delta x}{L_{repr}}} \times \varepsilon^{abs}. \tag{4.3}$$

Obviously, the error of representativity is reduced by increasing the model's resolution or by increasing the observation length scale.

In the present work, high resolution grids are applied, starting with a 15 km resolution for the mother domain and ending with a 1km resolution for the finest nest, reducing the representativity error. Moreover, both the characteristic absolute error for each measured species, ε^{abs}, and the characteristic representativeness radius of influence for each station type, are calculated following *Elbern et al.* [2007*a*].

Chapter 5

Analysis of special case studies

The 4D-var optimisation algorithm is applied for three different case studies and the main results are presented in the current chapter. An insight into the goals, the special meteorological conditions of each episode and the characteristics of the system's set up are discussed and evaluated.

In the beginning, two special real cases are studied, an aestival high ozone and a high hibernal stagnant air episode. Both are used to introduce the main features of the EURAD-IM system, evaluating firstly, the newly developed online adjoint emission factors optimisation, using a rich observational basis of ground and satellite measurement and secondly, the ability of resulting in updated emissions. For more comprehensive statistics, the assimilation procedure is applied for several days in a row, giving the opportunity for a better understanding of the system's performance over longer periods. In addition, the model's ability to identify special episode situations and the dispersion of elevated pollutants concentrations is estimated.

Further, assimilations for selected days of the PEGASOS flight campaign in North Italy are performed, motivated by twofold objectives. Firstly, the 4D-var data assimilation analysis in a high resolved grid over polluted areas, such as Po valley is evaluated. Secondly, the new perspective that the airborne data provide to the model's performance is studied, in terms of distinguishing emission patterns, investigating the vertical distribution of trace gases and the PBL's dynamics, given the scientific campaign measurements.

The identification of the aforementioned case studies is motivated by the objectives of the Regional Climate Change project REKLIM - Topic 5 ('Atmospheric composition and climate: Interactions from global to regional scales', `http://www.reklim.de/en/topics/topic-5.html`) and the research activities in the frame of Monitoring Atmospheric Composition and Climate (MACC,

https://www.gmes-atmosphere.eu/), regarding the determination of interactions between regional climate change and air composition influenced by emissions.

Validating the assimilation results, two different methods are applied; the minimisation performance of the 4D-var algorithm is assessed by χ^2-test and by comparison with independent observations withheld from the assimilation procedure. The high computational needs for the episodes simulations are handled by technical develops and updates of the 4D-var algorithm. Moreover, all the simulation runs are performed in the Jülich Supercomputing Centre (JSC) of the Research Center Jülich (http://www.fz-juelich.de/ias/jsc/EN/Home/home_node.html).

5.1 Aestival high ozone episode

The tropospheric ozone is a 'secondary' pollutant formed by complex photochemical reactions. The ozone production is mainly driven by emissions of primary pollutants, such as NO_x, CO and VOCs. It is the third most important gas in its contribution to the global greenhouse effect, after CO_2 and CH_4 [*Solomon et al.*, 2007], and a powerful oxidising agent. Ozone has a lifetime average of about two weeks, that makes its transport possible for several thousands of kilometres in the free troposphere, contributing to the PBL pollution and enhancing its adverse effects on humans' health and vegetation. Elevated ozone concentrations most frequently occur near large urban areas, as the main sectors that emit ozone precursors are road transport, industry, power and heat generation plants.

An ozone episode is defined as a period, usually a few days up to three weeks, with high ozone concentrations, characterised by daily exceedances of the threshold set to protect human health [*European Parliament and Council of the European Union*, 2008]. This is the case for the summer of 2010, that was characterised by a long period with frequent exceedances of the threshold of 120 $\mu g/m^3$ maximum daily eight-hour mean ozone concentrations, during the warm meteorological conditions between the 24^{th} of June and the 22^{nd} of July [*EEA Technical report*, 2011].

The long lasting episode triggers the interest of the atmospheric modelling for simulations with the 4D-var method. The joint optimisation of initial values and emission rates provides a well estimated representation of the quality state of the atmosphere, when assimilating both the emitted species and ozone. By this way, more information about the surface ozone formation around urban areas is gained.

The main focus of the current case study is the period from 07.07.2010 until 12.07.2010, where the highest temperatures are recorded together with the highest percentage of threshold exceedances, as it is highlighted in Fig. 5.1.

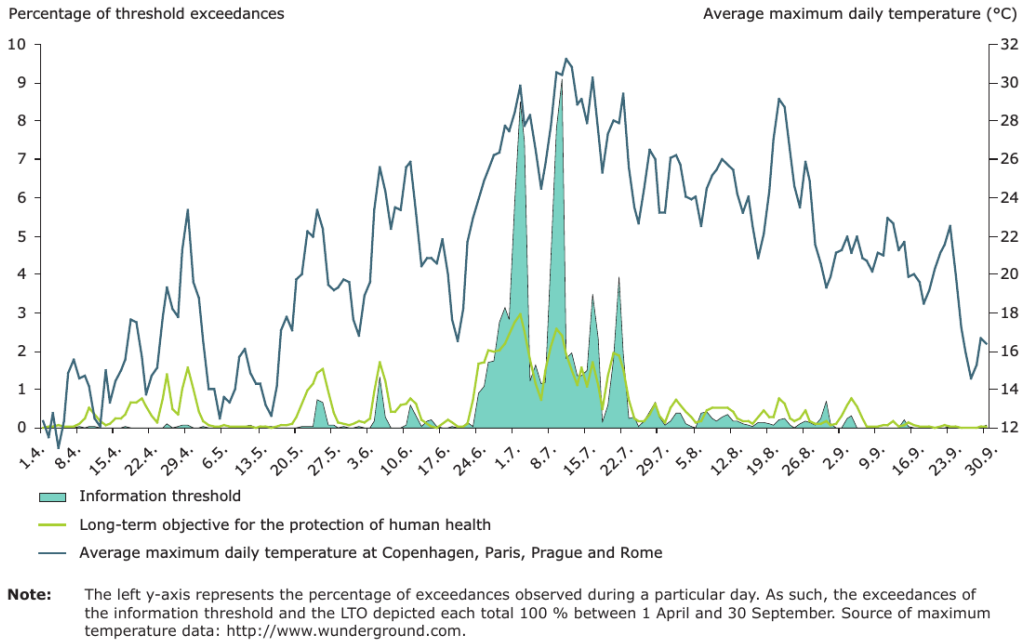

Figure 5.1: *The distribution of daily ozone exceedances for the entire continent of Europe and the maximum temperatures observed in four European capital cities (Copenhagen, Paris, Prague and Rome). Source: EEA Technical report [2011].*

5.1.1 The meteorological conditions

The ozone concentrations do not depend only on precursor emissions, but also on the meteorological conditions. Sunlight and high pressure systems are necessary conditions for enhancing the tropospheric ozone levels. Moreover, the airmass stagnation over areas with high emissions leads to ozone accumulation.

During summer 2010, and more precisely the dates under consideration, high insolation and temperatures occurred for long period. The Mediterranean countries were affected, recording the most widespread concentrations of ozone. However, the same meteorological conditions were also experienced in western and central Europe, fact that is not a typical behaviour for these regions.

The very low horizontal wind trapped the polluted airmass, increasing the concentrations of the emitted pollutants, mainly around urban areas.

Fig. 5.2 highlights the meteorological conditions during the 11.07.2010, a representative day of the case study. The high pressure system over Europe caused high temperatures, more than $27°C$, and low wind speed. Additionally, the low humidity conditions enhanced the stability of the low troposphere air, trapping the concentration of pollutants, highlighting ozone.

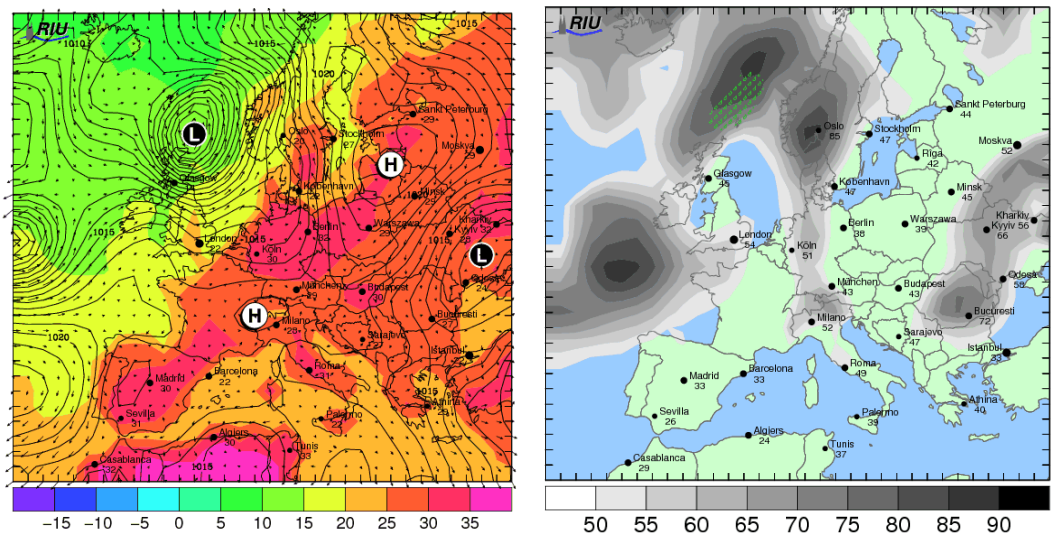

Figure 5.2: *The meteorological condition, for the surface level, during the 5^{th} day of the episode, representative for the average of the episode's days. Left panel: 2m temperature (°C, indicated in the color bar), sea-level pressure (hPa) and horizontal wind (m/s). Right panel: percentage of relative humidity (%, indicated in the color bar) and tendency of rain (precipitation >6mm, green triangles). Source:* `http://www.eurad.uni-koeln.de/`.

5.1.2 The observational data and the model set up

The observational basis consists of ground based and satellite measurements (Table 4.1). EEA in-situ data of NO, NO_2, O_3, SO_2 and CO are combined with the NO_2 tropospheric columns of GOME-2, SCIAMACHY and OMI, as well as with O_3 and CO observations from IASI and MOPITT, respectively. In other words, a full observational set of emitted species and products is assimilated. The episode simulation is performed for the 15 km resolved European grid, in which the locations of the in-situ stations network is shown in Fig. 5.3.

For the period 07.-12.07.2010 the data assimilation procedure is applied with

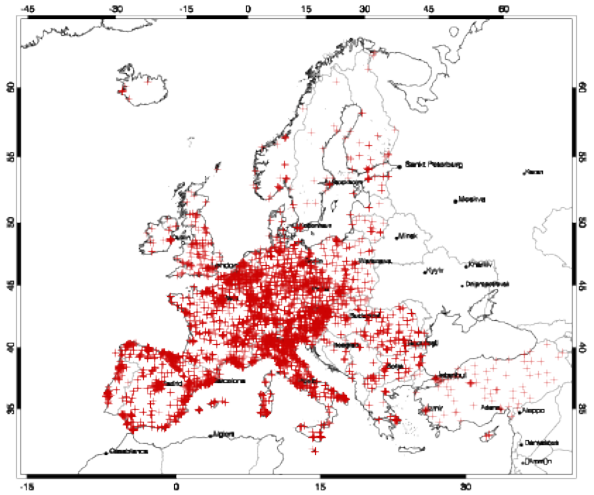

Figure 5.3: *The network of in-situ stations for the domain of interest, the 15 km resolved European grid.*

an assimilation window of 24 hours. The motivation for selecting a long length window lays on the diurnal profile of the emission, that is taken as a strong constrain. Joint optimisation simulations are performed, improving both the initial species concentration values and the emission factors. Moreover, for validation reasons, control runs are performed without any data assimilation, including the background information of the EURAD-IM.

5.1.3 The analysis results

Assimilation of routinely data

Performing 4D-var runs for 5 days continuously leads to optimised initial data for the following day every time and so to an improved analysis state for it. The fifth day of the assimilation achieves higher minimisation of the cost function than the first day. Fig. 5.4 illustrates the last statement, showing the iterative reduction of the cost function for each of the 5 days of the assimilation. The analysis every day benefits from the optimised result of the previous day and starts with already reduced values for the cost function, that are further minimised.

The validity of the aforementioned assimilation result has to be checked, thus a χ^2-test is adopted. As it is discussed in *Talagrand* [2010], given the unbiased innovation vector

$$d = y - \mathbf{H}x$$

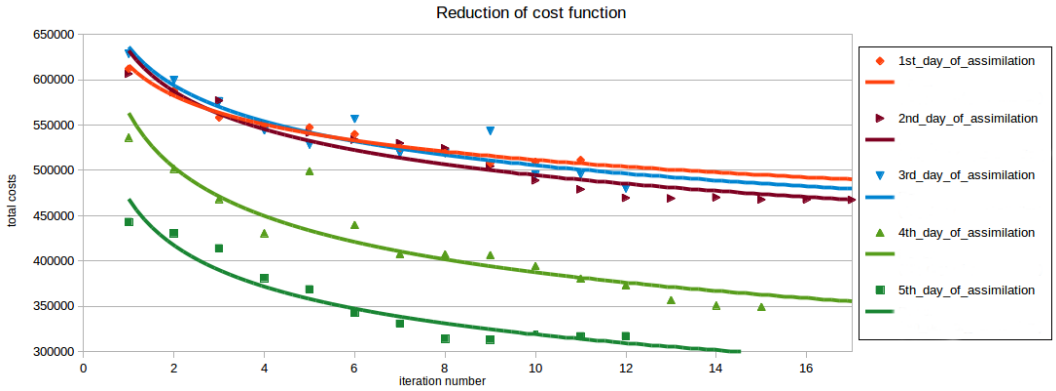

Figure 5.4: *The absolute reduction of the cost function for the 5 assimilated days.*

the value of the cost function at its minimum reads

$$J_{min} = \frac{1}{2}\sum_{i=1}^{N} d_i^2, \tag{5.1}$$

where N is the total number of observations. Likewise, the expectation of J_{min} in a consistent system is

$$\mathcal{E}[J_{min}] = \frac{N}{2}. \tag{5.2}$$

Eq. (5.2) provides the check for the consistency of the assimilation system, so the assimilation result is evaluated whether it satisfies this condition.

The average over the 5 assimilated days gives a mean cost of 46×10^4 and a halved number of observations of 33×10^4. Namely, the expectation of J_{min} is larger than $N/2$, which implies a slight underestimation of the observation errors. However, as it is shown in a series of χ^2-test applications [*Elbern et al.*, 2007*a*; *Chapnik et al.*, 2006; *Muccino et al.*, 2004], systems that have gone through extended operational validations and tuning usually show differences of at most few units, of the $\mathcal{E}[J_{min}]$ and its theoretical value $N/2$. With this in mind, it can be concluded that the resulting difference from the χ^2-validation is not significant of an inconsistent system. Although modifications of the cost function's error statistics are a matter of further improvement, the consistency of the assimilation system is guaranteed.

The optimised analysis by 4D-var results in improved calculation of the species concentration, either the emitted ones or their products. The over- or underestimation of the model's background calculations is corrected by the joint optimisation run. Fig. 5.5, presents the differences between the analysis

result and the model run without any assimilation - i.e., the background - for the simulation of the first and the fifth day of the ozone episode, on surface level, for O_3 (first row) and NO_2 concentrations (second row). These figures underline that the analysis provides improved data compared with the model for the whole assimilated period. On the first day of the assimilation the analysis is improved by the observations around areas with high emission sources, while on the fifth day of assimilation for the whole domain. In case of O_3 this correction is more extended than in case of NO_2, which is an expected result, as the 15 km resolved grid is not representative for NO_2 observations.

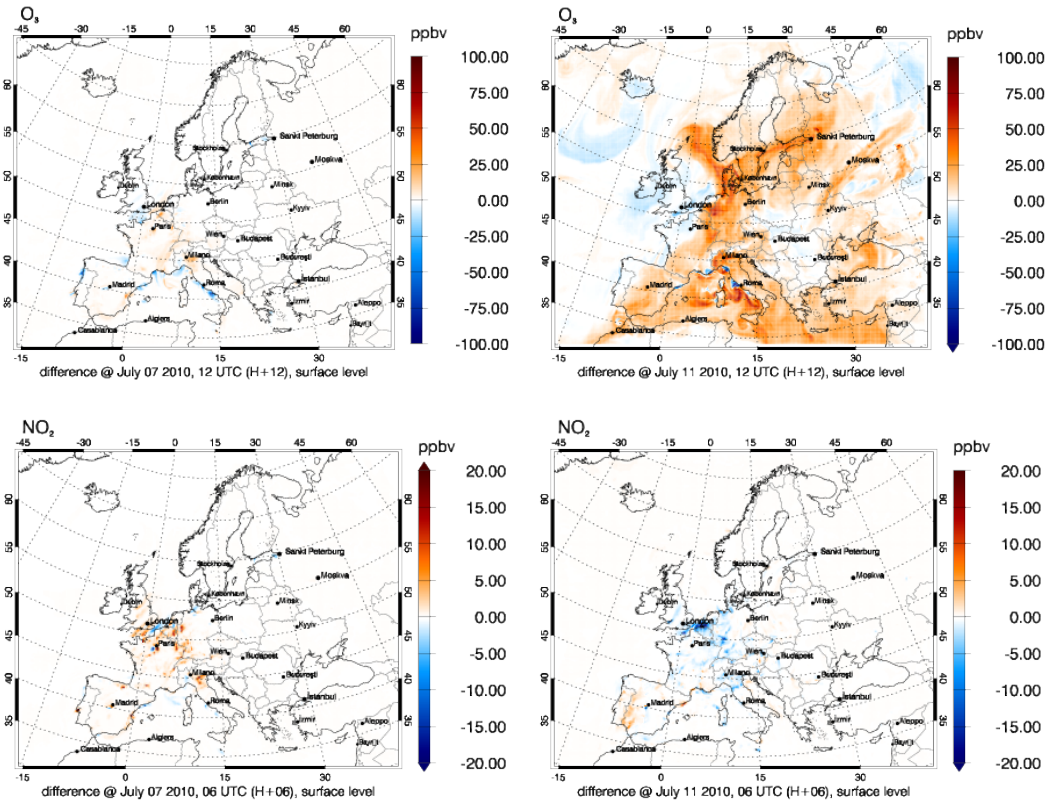

Figure 5.5: *Differences between the analysis and the model (analysis-minus-background) for O_3 during noon at 12:00 UTC (first row) and NO_2 during morning rush hour at 06:00 UTC (second row) in the 15km resolution European domain, at the first (left panels - 07.07.2010) and the fifth (right panels - 11.07.2010) day of assimilation of the ozone episode.*

A main matter under consideration is whether the model is able to optimise the emission factors and calculate emission patterns. The 4D-var algorithm of EURAD-IM is shown to provide optimised emission data. The emission

factors, set equal to one in the beginning of the assimilation, are improved at the end of the first day of the episode. After 5 assimilation days there is a significant correction of the emission factors for all four assimilated emitted species. Fig. 5.6 presents the correction of the emission factors of NO_2, NO, SO_2 and CO.

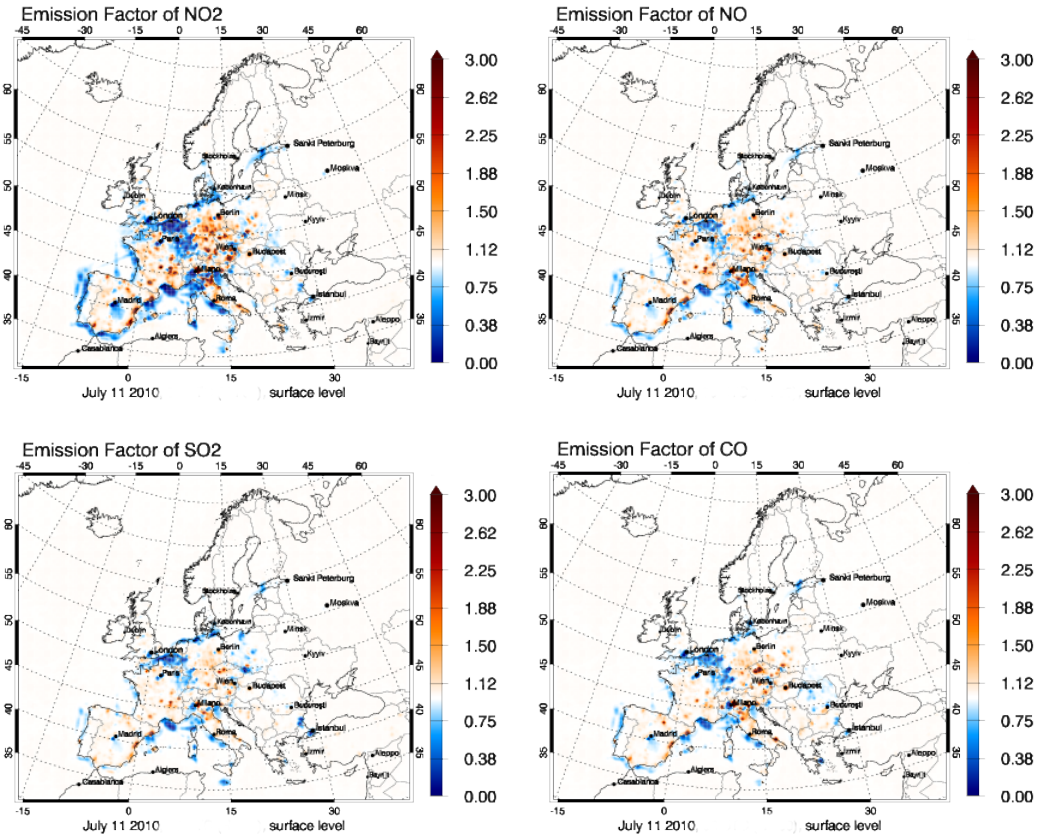

Figure 5.6: *Correction of emission factors after five days of assimilation for the high aestival ozone episode on 11.07.2010. First row: correction of NO_2 (left panel) and NO (right panel) emission factors, second row: correction of SO_2 (left panel) and CO (right panel) emission factors.*

For instance, in case of NO, the model calculates higher concentrations in areas with rich emissions, such as large urban centres, industrial areas and seaports, thus, the optimised emission factors are in general reduced. This result is also well captured for the other emitted species, implying the influence of the inverse temperature and the stagnant conditions in the atmosphere to the elevated pollutants concentrations in the surface layer. The analysis gives more accurate concentration fields for precursor and product species (Fig. 5.7).

The elevated O_3 concentrations dominate around Europe, with averaged values higher than 90 $\mu g/m^3$ during the afternoon hours. The assimilation also captures the night-time depletion of O_3, resulting from the strong anthropogenic emissions.

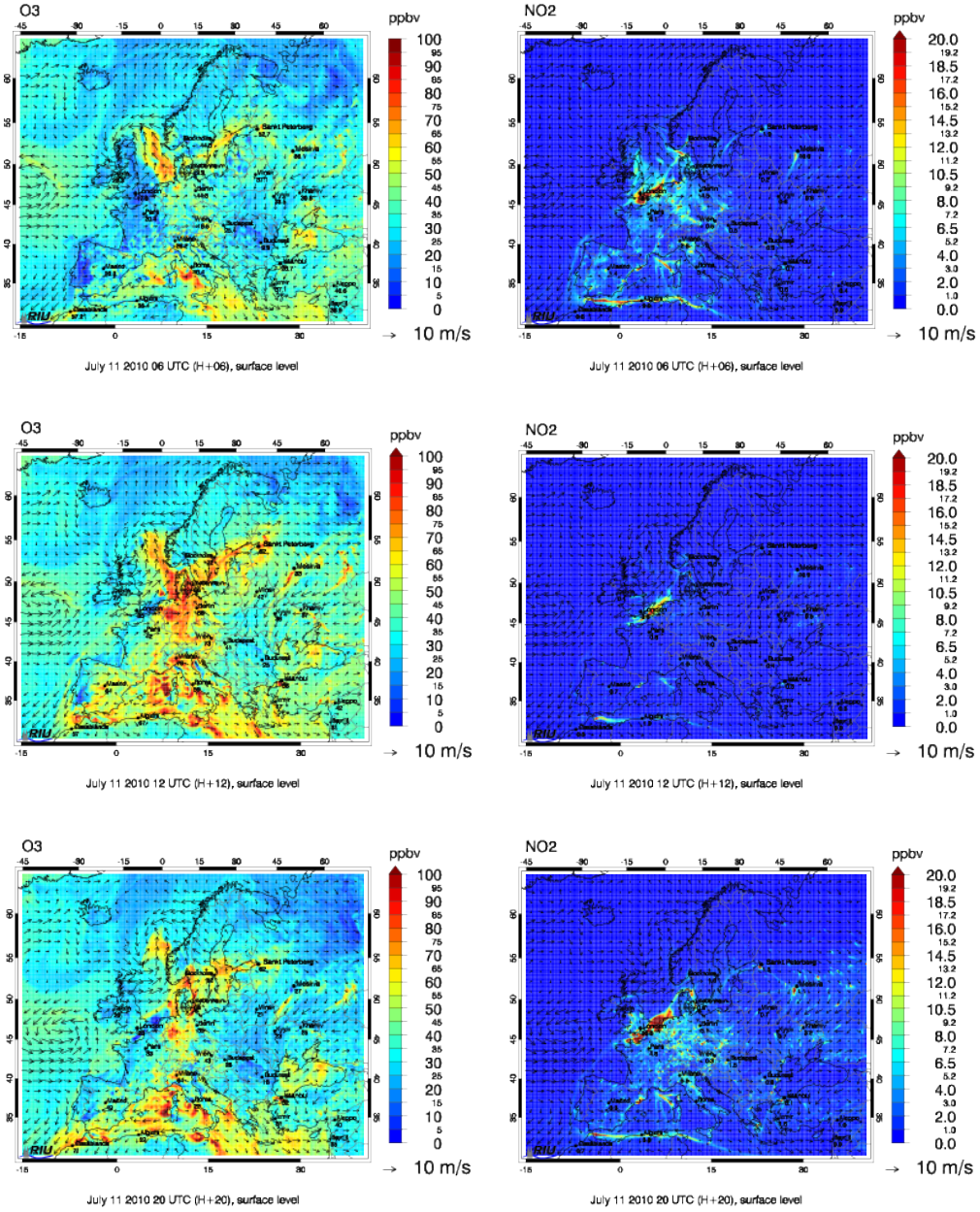

Figure 5.7: *The calculated concentrations of O_3 (at 12:00 UTC) and NO_2 (at 06:00 UTC) during the 11.07.2010.*

The model's behaviour can be also recognized in the time series of the analysis and the background run. Fig. 5.8 depicts average time series of O_3, of the total 2004 assimilated stations in the whole European domain, which have recorded O_3 observations, during the fifth day of the episode. Here the assimilation of O_3 gives a satisfying analysis result that fits better to the observations than the background run does.

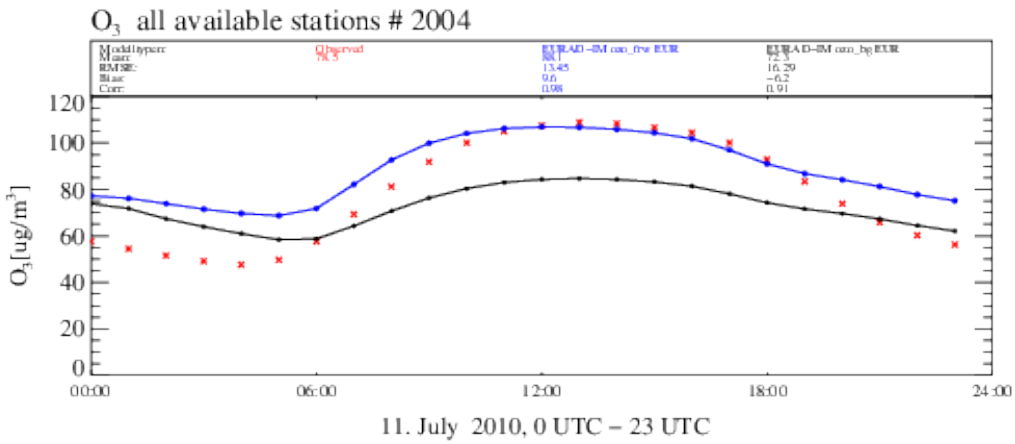

Figure 5.8: *Time series of the average observed and assimilated O_3 concentrations, from all assimilated stations in the coarse domain for 11.07.2010;* **black***: background, red: observations, blue: both initial value and emission factor optimization.*

An example of times series over individual stations is given in Fig. 5.9, for a station in Augsburg in Germany and in El Prat in Span, in case of O_3 observations. The analysis of O_3, as product of the nitrogen circle, benefits from the assimilation of O_3 and NO_2 observations together, as well as of the joint optimisation of initial values and emission factors. It is the optimised emission factors of the ozone precursors that enhance the successful result. That gets clearer mainly after the 6 first hours of the assimilation window, were the influence of the O_3 initial concentrations optimisation is not so strong any more.

Conclusions

Overall, the minimisation performance of the model after several days of assimilation is proved to be satisfied and the χ^2-validation of the optimised result successful. These facts can conclude to improved forecast abilities of the model. Moreover, the first application of the newly developed emission factor

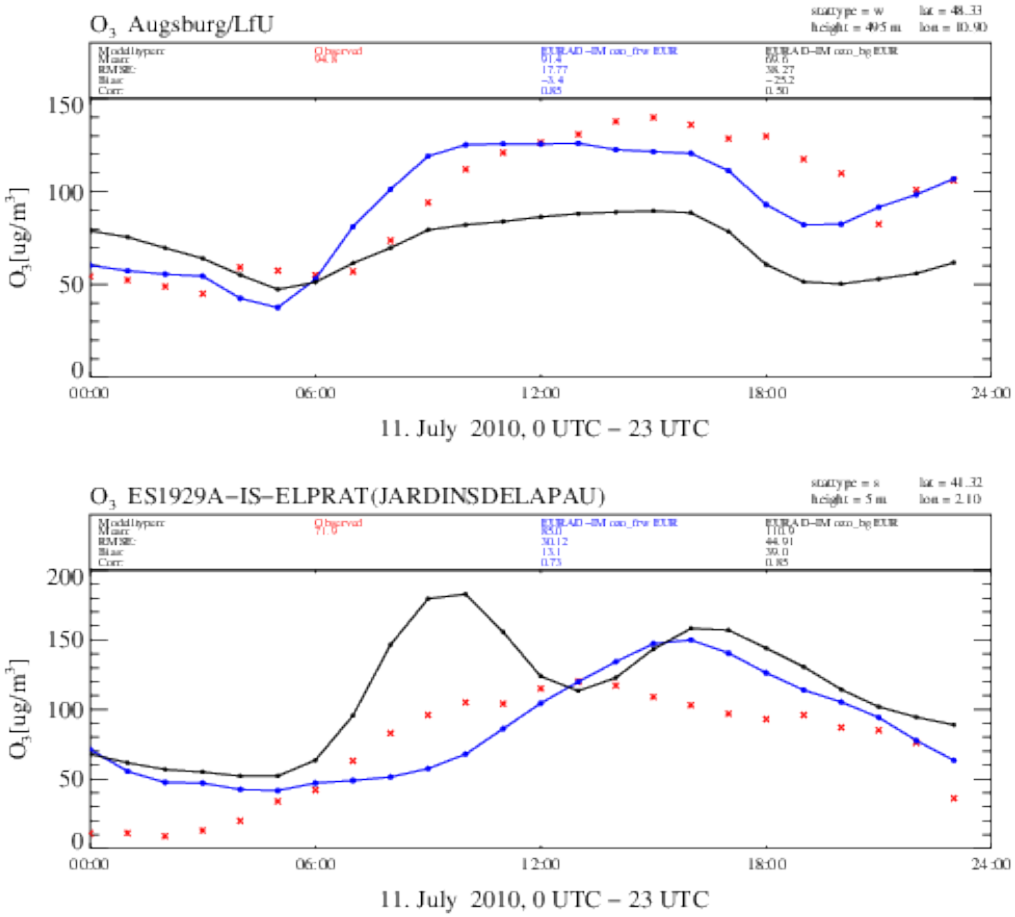

Figure 5.9: *Time series of O_3 observation in case of individual urban stations in Augsburg - Germany (upper panel) and in El Prat - Spain (down panel) on 11.07.2010;* **black***: background, red: observations, blue: joint initial value and emission factor optimization.*

error covariance matrix results in optimised emission factors and so improved concentration fields for precursor and product assimilated species.

5.2 Hibernal high stagnant air episode

Urban areas usually suffer from elevated concentrations of atmospheric aerosols and some gaseous pollutants, during specific meteorological conditions, like low temperature inversions, high pressure systems and low wind [*Choi et al.*, 2008]. During the winter of 2012, such an episode of long lasting high

stagnant air concentrations took place in Europe, influencing the quality of the air in the surface layer for numerous urban centres. Here data from the main days of the episode are assimilated (15.01.2012 - 21.01.2012), using ground based in-situ and satellite measurement. It is underlined that aerosols observations are not assimilated in the current work, as the main focus is on gaseous pollutants that under such extreme episodes can undergo pronounced increase and thus threat human heath [*Lyamani et al.*, 2012]. Further, the optimisation of emission factors is able to provide more information on the open question whether this kind of events are forced by increased anthropogenic emissions or by extreme meteorological conditions, notably low inversion layers.

5.2.1 The meteorological conditions

The extreme cold winter 2012 in Europe [*Sillmann et al.*, 2013] enhanced the concentration level of particulates, because of the prevailing low level stagnant atmospheric conditions. Fig. 5.10 highlights the meteorological conditions over Europe for 15.01.2012, representative day for the whole episode.

The high pressure system over western Europe caused clear sky and very low temperatures. The anticyclone induced low wind-speeds and intuition of vertical exchange. The warm air over the cooler air acted as a lid, preventing upward movements and thus the pollutants were trapped near the ground.

5.2.2 The observational data and the model set up

Both ground based and satellite measurements for species in gas phase are assimilated in this case study (Table 4.1). The network of the EEA in-situ data (Fig. 5.3) is combined with retrievals from GOME-2, OMI, IASI and MOPITT, for NO, NO_2, O_3, SO_2 and CO.

The 4D-var inverse modelling technique is applied for 7 days of the episode, leading to optimised initial conditions for each following day. The initial concentrations of the species are optimised together with the emission factors for a 24-hours assimilation window. In order to control the analysis results, runs without data assimilation are performed for the same time period.

5.2.3 The analysis results

Assimilation of routinely data
The assimilation of the present case study helps to assess the model behaviour

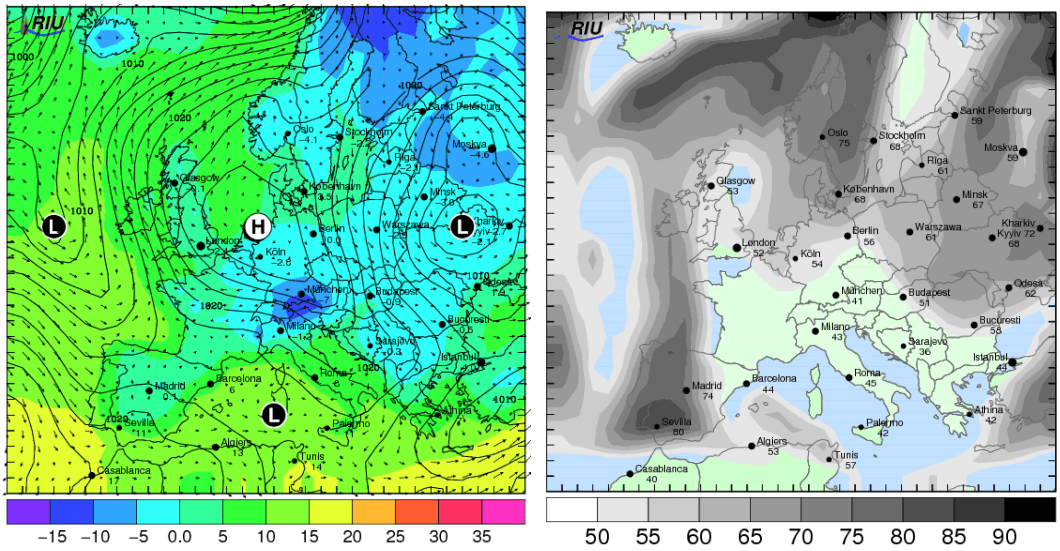

Figure 5.10: *The meteorological conditions of 15.01.2012, representative for the average of the episode's days. Left panel: 2m temperature (°C, indicated in the color bar), sea-level pressure (hPa) and horizontal wind (m/s). Right panel: percentage of relative humidity (%, indicated in the color bar) and tendency of rain (precipitation >6mm, green triangles). Source:* `http://www.eurad.uni-koeln.de/`.

for the joint optimisation of the initial values and the emission rates. In general, NO_2 concentrations are underestimated by the model, especially during wintertime. There is a dual cause to this issue. First of all, the low spatial representativeness of NO_2 observations for the model's grid is strong. This is because the majority of stations that observe NO_2 is located in urban areas, or close to large streets and highways, which do not represent the state of a grid or a domain. Further, there are uncertainties on what it is measured from the observational instruments, remote sensing and in-situ. In reality, in addition to NO_x, the measured information for nitrogen oxides may also include NO_y, including HNO_3 concentrations. This information is used afterwards to calculate NO and NO_2 measurements. Hence, this procedure is biased and prompt to end up with higher observed NO_2 concentrations.

So far these two matters are difficult to be controlled by the CTMs. However, 4D-var within the EURAD-IM proved to be able to reduce these unwanted results. For instance, after 7 days of assimilation, the underestimation of NO_2 background concentration is significantly corrected towards the observations in the analysis result. This can be seen in Fig. 5.11, where the differences between the analysis and the background run are shown, for the morning rush hour of the 7^{th} day of the assimilation. This correction is larger

around urban areas with high emissions and point sources, where also correction of the NO_2 emission factors are analysed.

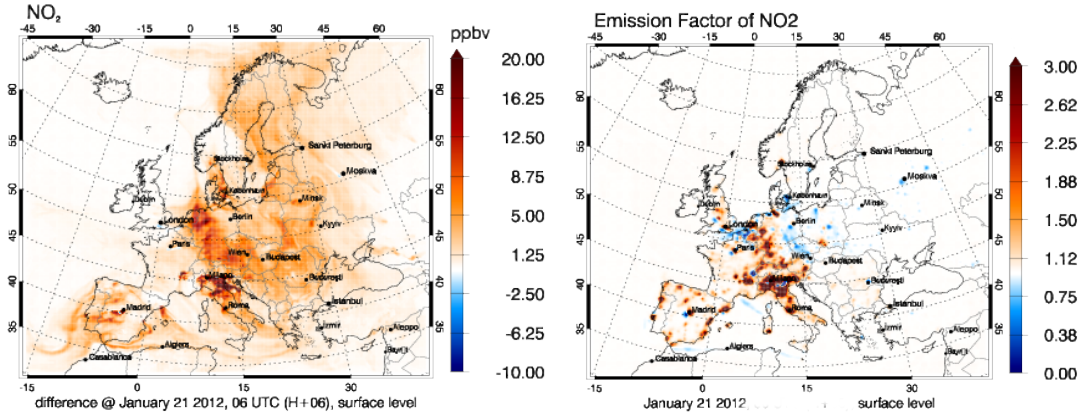

Figure 5.11: *Differences between the analysis and the model (analysis-minus-background) for NO_2 concentrations in the 15 km resolution European domain, during morning rush hour (06:00 UTC), after 7 days of assimilation on 21.01.2012 (left panel) and the correction of the NO_2 emission factors for the same day (right panel).*

The optimised information gained by the analysis run is also visible in time series. In Fig. 5.12 the time series of the averaged assimilated NO_2 observations from 493 ground stations in Italy verify the improvement of the model in the direction of the observations, addressing the initial underestimation of NO_2 concentrations. The success of this correction during the whole assimilation window is a result of the joint optimisation of initial values and emission factors, which also provide improved initial conditions from the previous day assimilation.

The issue of representativeness of NO_2 measurements, prevents the analysis for higher minimisation of the discrepancies. However, in the time series of individual stations close to emission point sources the optimisation is well achieved. For instance, Fig. 5.13 shows time series plots of NO_2 observations of two individual ground stations in Cottbus in Germany and in Sankt Georgen im Lavanttal in Austria. Both observation stations are located in areas where there are large lignite-fire power stations. Here the analysis result diminish the differences between the model and the observations, fitting well to the observational sets. This result indicates the influence of the error covariances to the analysis; both the background and the emission factor error covariances benefit from the optimised initial data from the previous days' simulation and allow the analysis to get closer to the observed data, being in agreement with

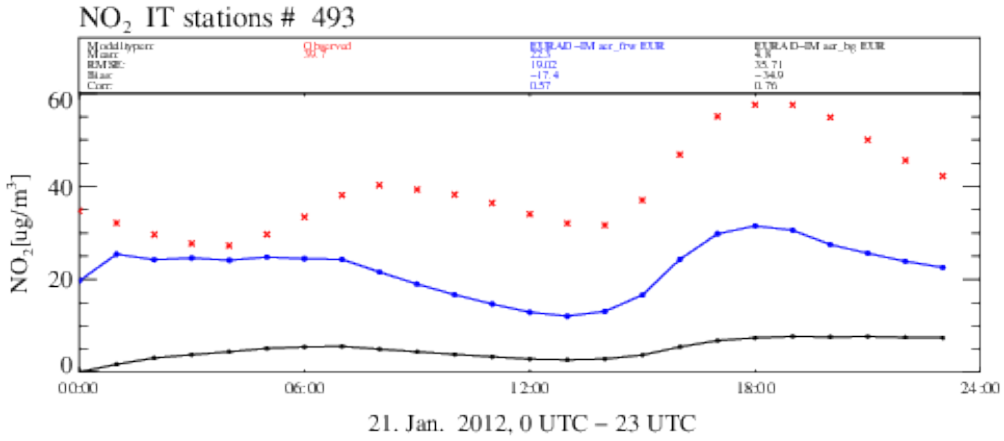

Figure 5.12: *Time series of the average assimilated NO_2 observation of 493 stations in Italy on 21.01.2012.* **black***: background, red: observations, blue: both initial value and emission factor optimization*

the elevated emissions in these areas.

A comparison of the data assimilation skill on the coarse grid with independent observations is presented for validation. Fig. 5.14 includes time series of independent observations of NO_2 and O_3, comparing the analysis result and the model background. Measurements from EBAS stations (`http://ebas.nilu.no/`) are withdrawn from the assimilation procedure to evaluate the analysis result. Here the output of representative stations from Spain (Niembro and Donana) and Macedonia (Lazaropole) are presented. In case of NO_2, the model calculates lower concentrations than the observations, whereas the analysis tries to minimise these discrepancies. The optimised result gives elevated initial concentrations compared with the observations, fact that comes from higher measurements of NO_2 of the previous day. Looking at the O_3 time series, the analysis optimises the background calculations. Although the first hours of the assimilation the analysis gets lower values than the model, this is soon corrected. A main reason for this behaviour is the joint optimisation of initial values and emission rates that is performed in EURAD-IM, assimilating together precursor and product, NO_2 and O_3 observations.

Generally, the assimilation period is characterised by higher emissions than provided by TNO-MACC-II inventory. In Fig. 5.15 emission factors of NO_2, NO, CO and SO_2 are presented for the coarse grid the 7^{th} assimilation day. The optimised initial values and emission factors from the long term simulations are shown to be elevated, verifying the influence of the meteorological conditions

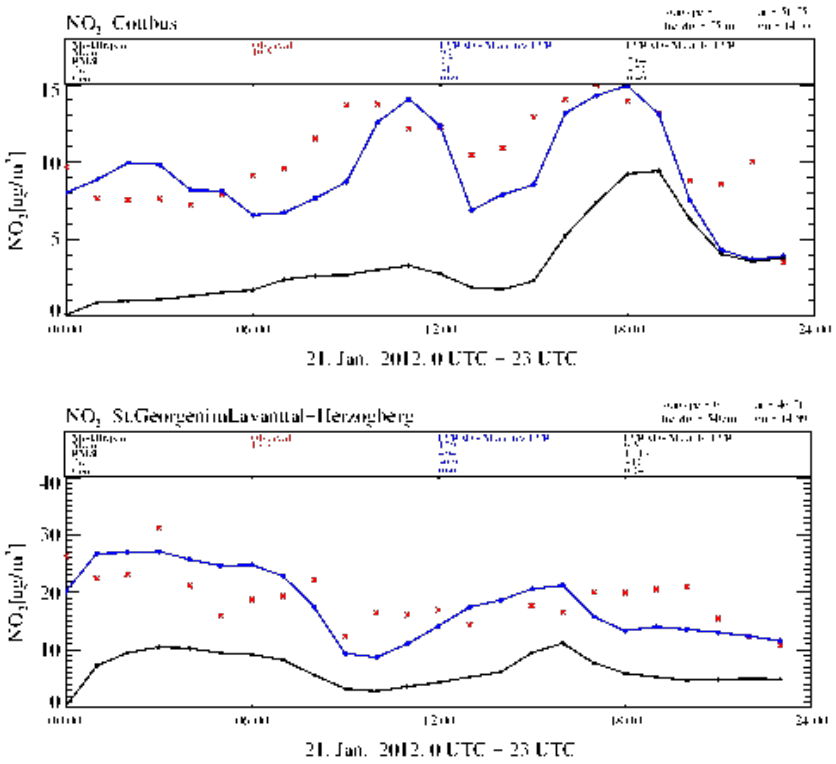

Figure 5.13: *Time series of NO_2 observation, on 21.01.2012, in case of individual stations close to emission point sources in Cottbus - Germany (upper panel) and in Sankt Georgen im Lavanttal - Austria (lower panel).* **black***: background, red: observations, blue: both initial value and emission factor optimization - best analysis.*

that enhance them.

Conclusions

In brief, the analysis of the stagnant winter episode let the evaluation of the model performance, regarding the optimisation of the error covariances, by the joint optimisation of the initial values and the emission factors. The comparison of the assimilation result with independent observations is proved to be successful. Therefore, improved emission fields can be provided by the analysis. The elevated emission rates, influenced by the meteorological conditions, can presumably maintained for longer periods, as it is concluded by the 7 days assimilation.

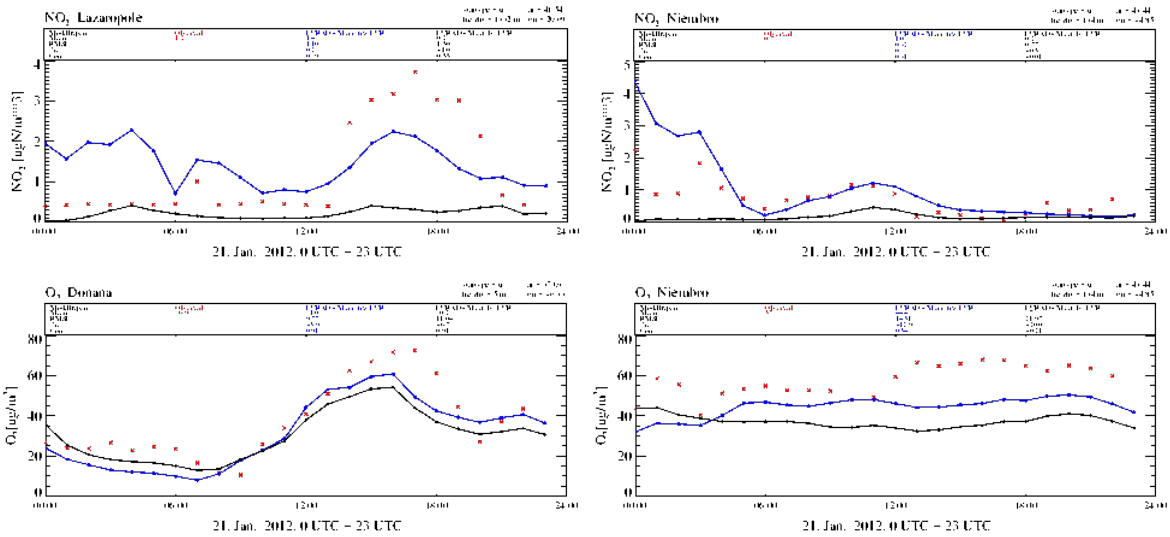

Figure 5.14: *Time series on 21.01.2012; evaluation of the analysis against independent observations, for NO_2 over the ground stations of Lazaropole in Macedonia and Niembro in Spain and for O_3 over the ground stations of Donana and Niembro in Spain;* **black***: background, red: observations, blue: both initial value and emission factor optimization.*

5.3 PEGASOS flight campaign

The main stimulus of the PEGASOS flight campaign was to shed light on local production and transport processes in lower troposphere, especially focused on aerosols and the hydroxyl radical OH, the so called "detergent of the atmosphere". The campaign offered large data sets of high quality measurements for the polluted area Po valley, that help to study and understand the emission patterns and the chemical consistency in this region [*Li et al.*, 2014]. However the airborne measurements are spatially and temporally limited and that raise questions on how they can support the assimilation analysis result in a CTM. To this end, the current study applies the 4D-var data assimilation method within the EURAD-IM for three days of the campaign, 10.-12.07.2012, to determine whether and how the campaign data are able to indicate corrections to the model analysis and thus to improve the air quality forecast. Further, the optimisation of emission factors in nested grids of 1km resolution is studied, addressing the issue of the representativity of observations, such as NO_2. Taking advantage of the high quality airborne campaign measurements, special focus is given on the analysis of the vertical mixing in the PBL with EURAD-IM, studying the the atmosphere's vertical structure.

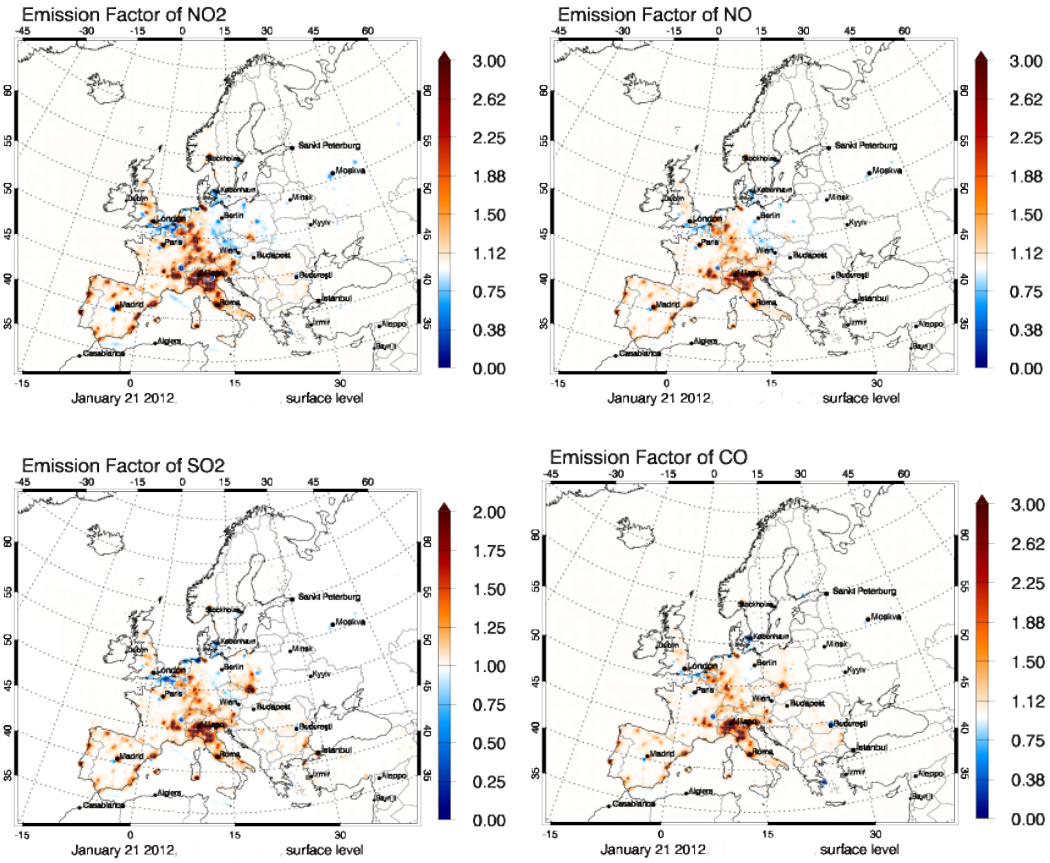

Figure 5.15: *Emission factors after seven days of assimilation, on 21.01.2012. First row: correction of NO_2 (left panel) and NO (right panel) emission factors, second row: correction of SO_2 (left panel) and CO (right panel) emission factors.*

5.3.1 The meteorological conditions

During the PEGASOS campaign in summer 2012, the airship Zeppelin NT flew for 25 days over the Po valley in North Italy, from 18.06.2012 to 13.07.2012. During this period, a high pressure system raised from central Europe to Eastern-Southern Europe, characterised by weak winds and stability.

On 12.07.2012, the main day of interest here, similar conditions maintained. Especially in the area of the Po valley, the temperature remained in high levels, with mean value the $28°C$ and, simultaneously, the mean relative humidity was 46% (Fig. 5.16). The temperature inversion prevent the upward movement of air, trapping the emissions in low levels. The area which is most influenced by such conditions is the lower layer of the atmosphere, namely the PBL.

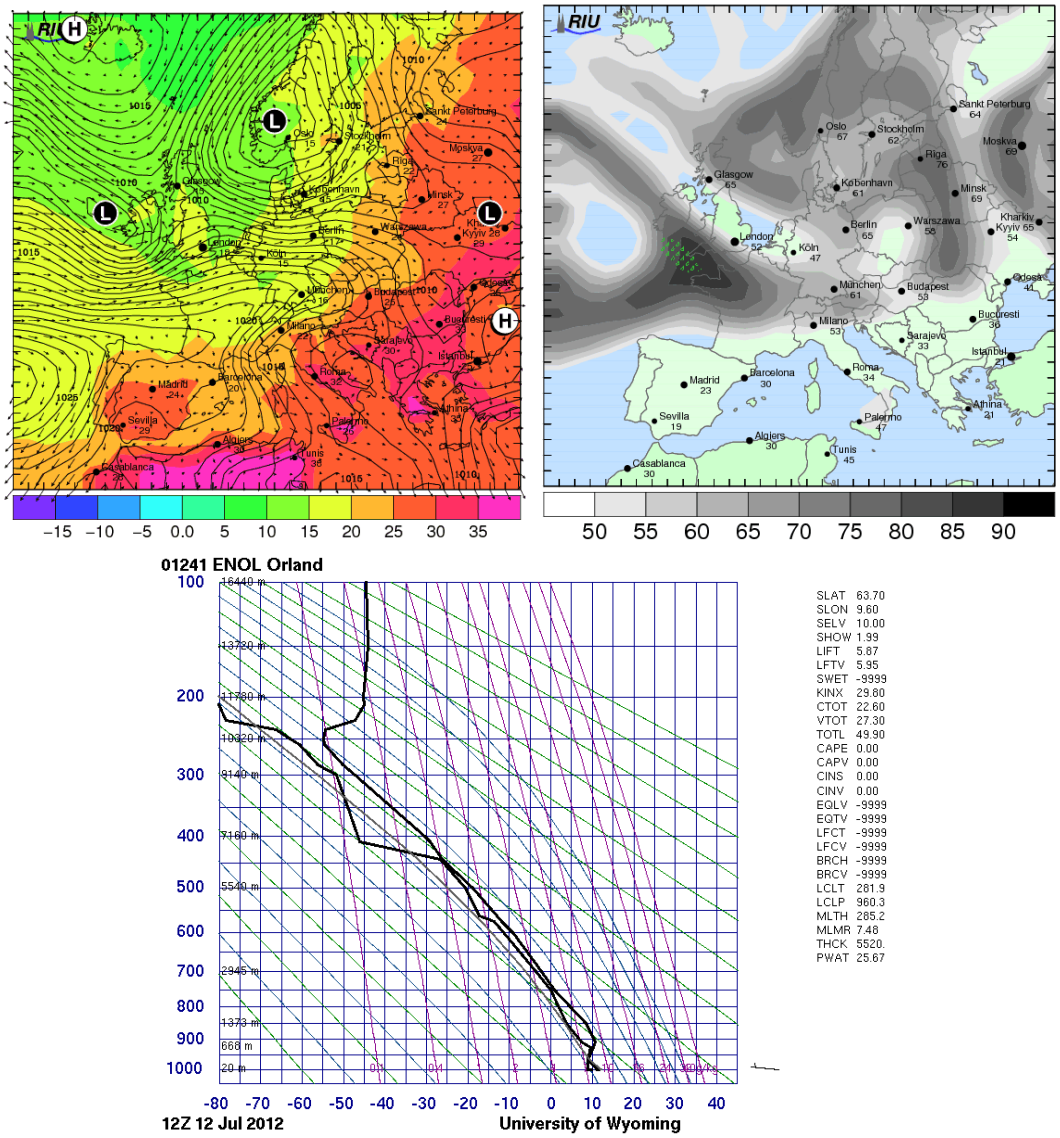

Figure 5.16: *The meteorological conditions on 12.07.2012, 12 UTC. Left panel: 2m temperature (°C, indicated in the color bar), sea-level pressure (hPa) and horizontal wind (m/s). Right panel: percentage of relative humidity (%, indicated in the color bar) and tendency of rain (precipitation >6mm, green triangles). Source:* `http://www.eurad.uni-koeln.de/`*. Low panel: temperature inversion on 12.07.2012. Source:* `http://weather.uwyo.edu`*.*

5.3.2 The observational data and the model set up

The observational basis consists of routinely ground based and satellite measurements, enhanced by the airborne measurements from PEGASOS cam-

paign (Table 4.1). The campaign data sets include vertical profiles of NO, NO_2, O_3 and CO, as the flight pattern was focused on the dynamics of the PBL and cross sections through the Po valley (Fig. 4.8 - yellow solid line). The vertical profiles cross all three layers within the PBL, providing valuable information about the evolution of constituents into the PBL.

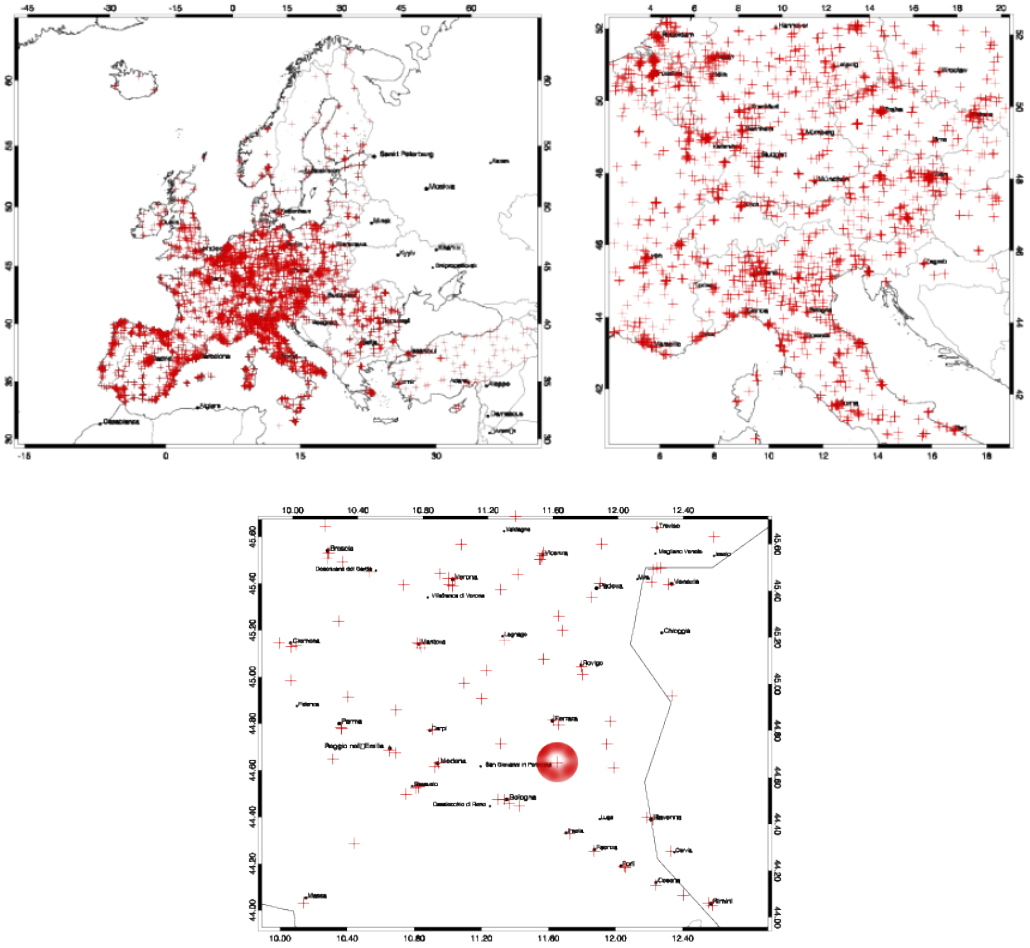

Figure 5.17: *The network of the EEA in-situ stations for the domains of interest: the 15 km resolved coarse grid (upper left), the 5km resolved first nest (upper right) and the 1km resolved second nest (lower), highlighted in the red circle the station of SPC.*

The Zeppelin's flight was coordinated with existing EEA ground measuring stations in the region (Fig. 5.17), so that to compare the different measured data [*Kaiser et al.*, 2015]. In Po valley the profile flights were performed above the rural supersite of San Pietro Capofiume (SPC).

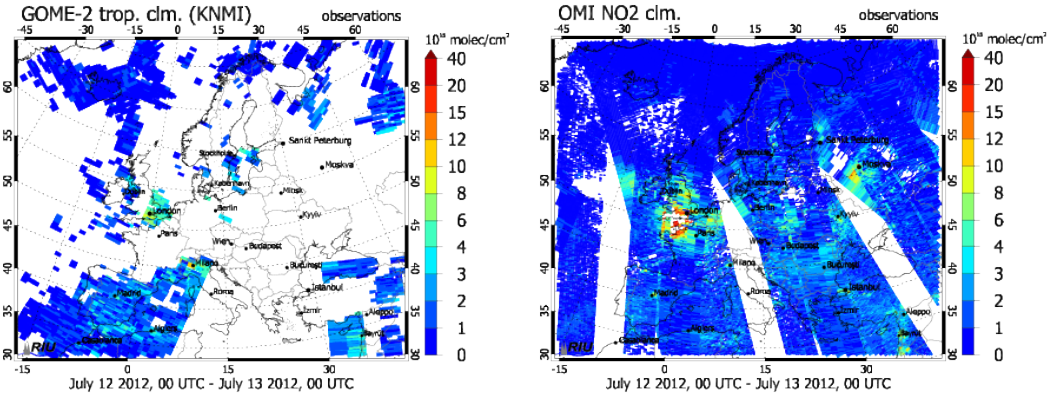

Figure 5.18: *NO_2 tropospheric column measurements from OMI over the first and the second nest, on 12.07.2012.*

In contrast with GOME-2, the measurements of NO_2 tropospheric columns from OMI during the campaign cannot contribute to the analysis result for the Po valley, as the satellite tracks are not over this region (Fig. 5.18). Though, the assimilation of the OMI data for the the coarse grid and the first nest contributes in optimised initial state variables for the finest second nest over the main area of interest.

A sequence of different case studies analysis were performed, with respect to the different kind of measured data. The first case study includes the assimilation of in-situ ground based and satellite observations. Here the assimilation period endures three days, 10-12.07.2012. For the second case, the scientific observations from the PEGASOS campaign are assimilated and the 12.07.2012 is studied. In all simulations, joint optimisation runs are performed, improving both the initial species concentrations and the emission factors, over all three domains of interest, for a 24-hour assimilation window. Moreover, for validation reasons, control runs are performed without any data assimilation, consisting the background field, namely the first guess of the model.

5.3.3 The analysis results

Assimilation of routinely data

A three-day assimilation of the ground based and satellite observations is presented as a reference run. Moreover, the accuracy of the model is assessed by high resolving grids, together with the joint optimisation of initial values and emission factors.

Fig. 5.19 underlines the benefit from the joint optimisation, illustrating the differences between the analysis result and the model's first guess in case of NO_2 and O_3 concentrations, for all three resolutions in the surface layer, during the first day of assimilation, the 10.07.2012. In case of NO_2 the morning rush hour at 06:00 UTC is shown, whereas for O_3 the noon-peak, at 12:00 UTC. Regarding the coarse grid, the fact that the 15 km resolution is not representative for the NO_2 observations is depicted with small differences of the analysis from the background, mainly in areas with large emission. For instance, the port of Marseille in south France is a well know large point source of emissions, that is why the first guess of the model tends to overestimate the NO_2 concentrations in this area and the analysis tries to correct this behaviour. Going down to the first nest, the 5 km resolution helps to identify emission plumes close to urban areas. Here, the analysis around the area of Marseille's port clarifies better patterns regarding the correction of NO_2 and O_3 concentrations, as the observational sets provide more accurate information. This information is even more detailed for the second nest, as the domain is 1 km resolved and so emission patterns can be clarified, diminishing also the problem of the representativity of NO_2 observations. For instance, the influence in the pollutants concentrations of the elevated car emissions is shown to be missed in less resolved grids. The analysis improves the background and calculated higher NO_2 concentrations, identifying the Italian highway that crosses the south-west domain, as well as a part of the rest smaller road network in the area.

Generally, an underestimation of NO_2 concentrations is dominant for all three resolutions, whereas the meteorological conditions are reflected in the high assimilated NO_2 concentrations. The temperature inversion let the warm air overlying colder air as a lid, preventing the upward movement and raising the pollutants concentrations. On the other hand, regarding O_3, the background of the model highly overestimates the O_3 concentrations in polluted areas, as Po valley. This fact is addressed by the strong decrease of the analysed concentrations. Being chemically consistent, the model analysis influence also the concentrations of the emitted species, and this is better shown by the correction of their emission factors.

As a result of the assimilation procedure, the analysed emission factors of NO_2 and CO are given in Fig. 5.20. The emission factors are initialised to one for the European coarse grid and optimised further on. The nesting technique guarantees that each nest gets as initial information the resulted improved emission factors from its mother domain, refining the previous determined scaling factors. Thus, looking in the finest domain covering Po valley, the different optimisation stages can be recognised, accounting for grid cells of

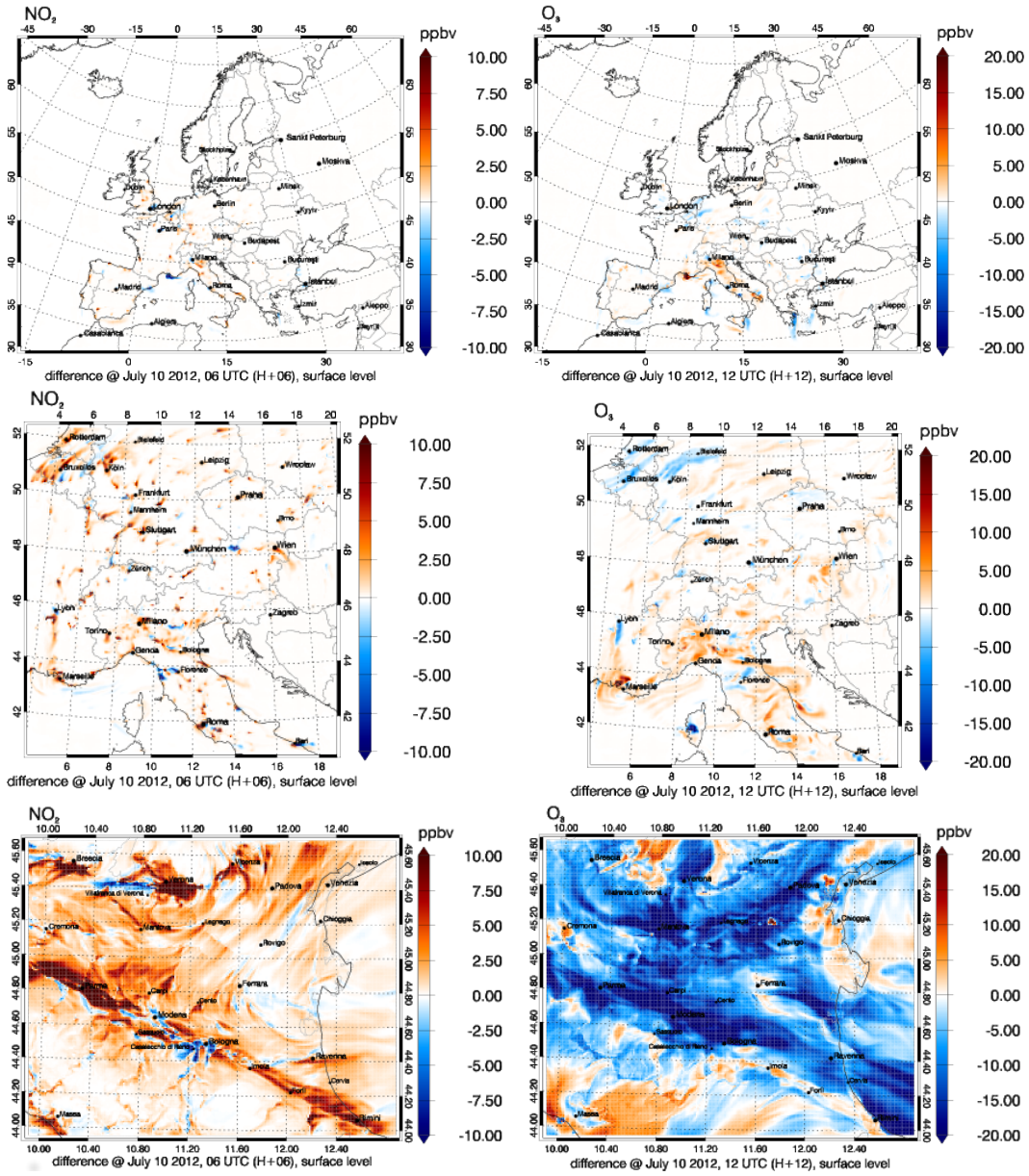

Figure 5.19: *Differences between the analysis and the background (analysis-minus-background) for NO_2 during morning rush hour (06:00 UTC) and O_3 during noon (12:00) for the coarse grid of 15 km resolution (first row), its first nest of 5 km resolution (second row) and the final nest of 1 km resolution (third row), in 10.07.2012, first day from a 3-days simulation.*

the mother domain and the nests. The correction of the emission factors achieved here is able to shed light on the dispersion of urban emissions in

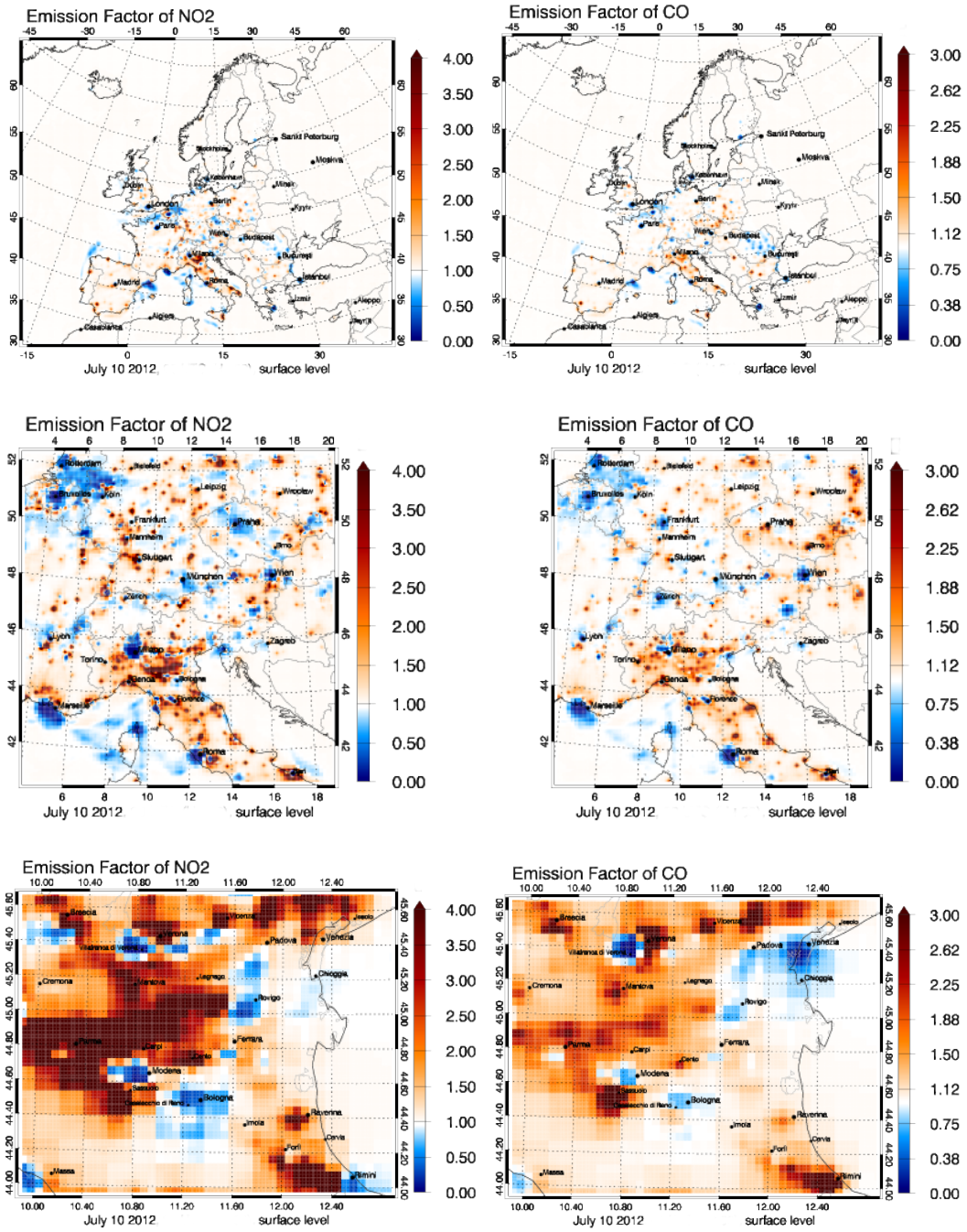

Figure 5.20: *The emission factors of NO_2 and CO for the coarse grid of 15 km resolution (first row), its first nest of 5 km resolution (second row) and the final nest of 1 km resolution (third row), in 10.07.2012.*

neighbouring areas. The increased optimised NO_2 emission factors, up to a factor of four, correct the general underestimation of the model for the NO_2 concentrations.The correction is larger in areas with high pollution levels. For instance, for the second nest, the CO emission factors around the region of Modena (central) or of Rimini (low left) can result in identifying the spread of CO emission plumes in these areas, whereas the NO_2 emission factors in identifying the highway network.

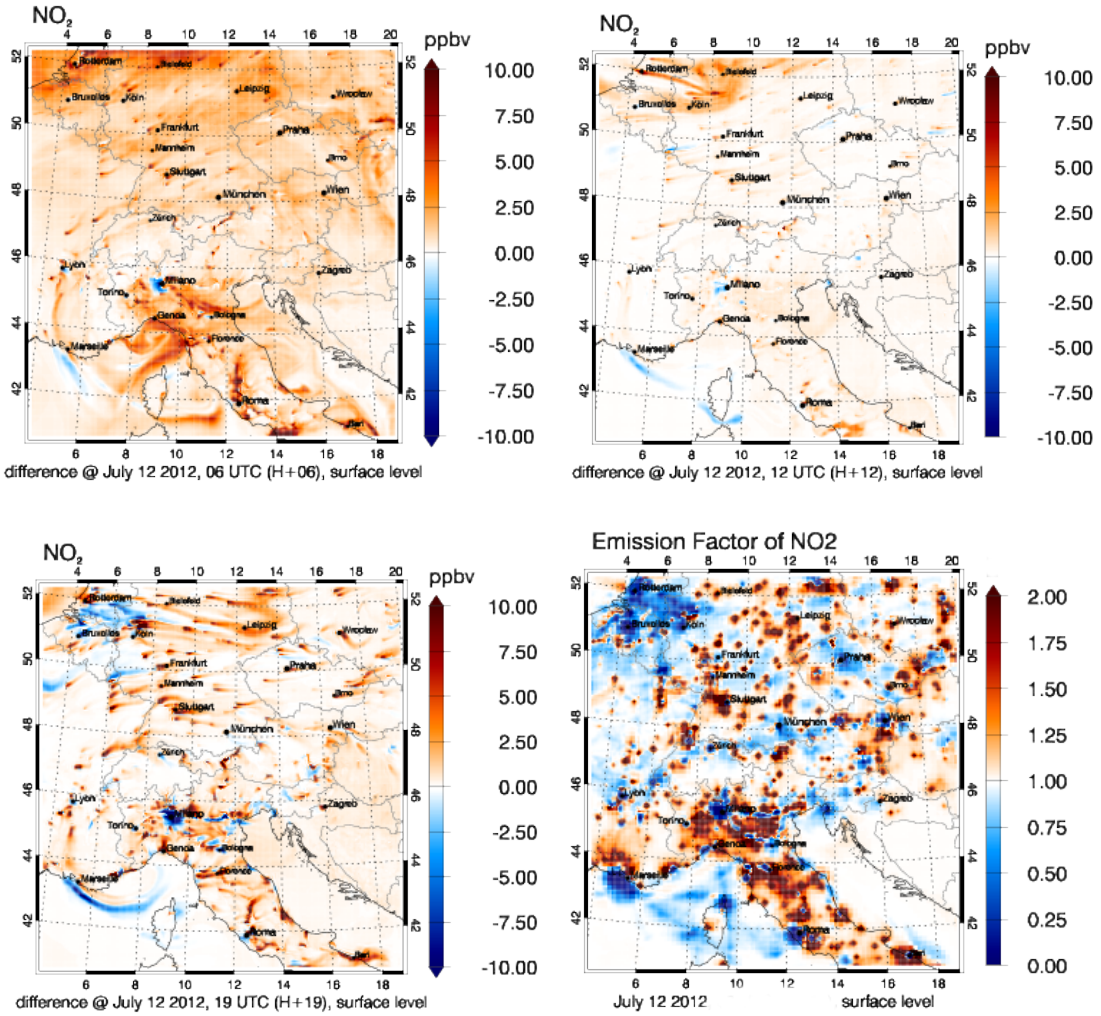

Figure 5.21: *Differences between the analysis and the background (analysis-minus-background) for NO_2 in the first nest (5 km res.), in 12.07.2012, third day from a 3-days simulation - the morning rush hour, 06:00 UTC (upper left), the noon low, 12:00 UTC (upper right) and the afternoon peak, 19:00 (lower left). The NO_2 optimised emission factors for the same day (lower right).*

The simulation run lasts for three days of the episode, ending in 12.07.2012. Evaluating the model's performance after two days of assimilation, Fig. 5.21 depicts the differences between the analysis and the background of the NO_2 concentrations, during three different times within the day, for the first nest. The optimised fields are smoother than the first day of the episode (Fig. 5.19 middle right panel), spread along the whole grid, not only around the sources of high NO_2 emissions, influenced by the optimised initial variable state of the previous day's assimilation. When this procedure is combined with high resolution nesting, the transferred information is more accurate and clear, enhancing the assimilation result. The resulted optimised emission factors induce more detailed and sharp emission patterns over urban centres and polluted areas, as the Po valley region, concluding in updated concentration fields.

Validating the analysis output for each resolution, a comparison with independent observations takes place. Fig. 5.22 depicts the time series of NO_2 and O_3 concentrations for the analysis result of the three different domains against measurements of two stations in Italy that are not included into the assimilation procedure. For both the emitted NO_2 and its product O_3 better optimisation is achieved for the finest nest grid (blue curve). In case of NO_2 the good representativeness of NO_2 observations for the 1 km resolved grid is obvious, fact that also influence the analysis of O_3, since the system maintains its chemical consistency. The afternoon peak of NO_2 analysed concentrations in case of the second nest (blue curve) presumably comes from the assimilation observations from other stations over areas with higher traffic activity, fact that is not possible to be recognised in case of the less resolved grid (green and black curves).

Assimilation of campaign data

The second sequence of the case study runs includes the assimilations only of the PEGASOS airborne observations, in 12.07.2012 during 04:30 - 08:45 UTC, over the SPC supersite. The assimilation of the high resolved campaign data provides a distinct representation of the chemical state of the air up to the PBL. The campaign comprises timely high resolving measurements, recording the changes of the pollutant concentrations. Thus, for the model's simulation the temporal resolution is maintained high, so that to take advantage of the campaign data and be able to identify the details of the evolution of the tropospheric constituents. Besides, species like NO have short life time after they are emitted and they fast react with others. The coarse grid runs with 300 sec of temporal resolution, whereas the first and second nests with 100 sec and 25 sec, respectively.

The assimilation regionally shows optimised results of the species concentra-

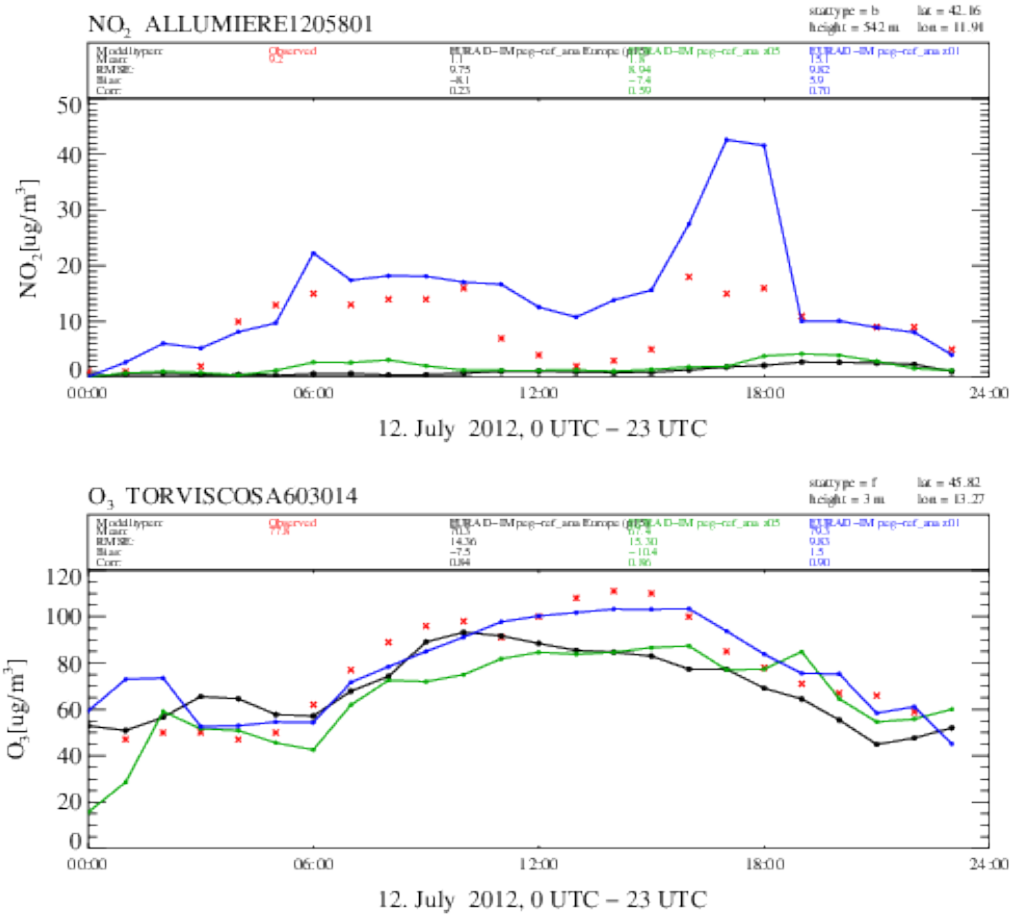

Figure 5.22: *Time series of NO_2 (upper panel) and O_3 (lower panel) observations of the three different domains, for two stations of in-situ non-assimilated observations, Allumiere and Torviscosa in Italy. The analysis result for each domain represents the joint assimilation of initial values and emission factors for all given in situ and satellite observations for each domain. With red are given the non-assimilated observations, with* ***black*** *is given the analysis after the assimilation of the mother European domain (15 km resolution), with green is given the analysis after the assimilation of the first nest of north Italy (5 km resolution) and with blue is given the analysis after the assimilation of the second nest over the area of Po valley in Italy (1 km resolution).*

tions and the emissions. Of coarse, it is not possible that the measured profiles over a specific area can optimise the initial variables for the whole coarse grid, sized 15 km × 15 km. However, even in the coarse grid, for the area of SPC the airborne measurements give a more detailed insight than the routine data, being able to optimise the initial values and the emission rates. For instance,

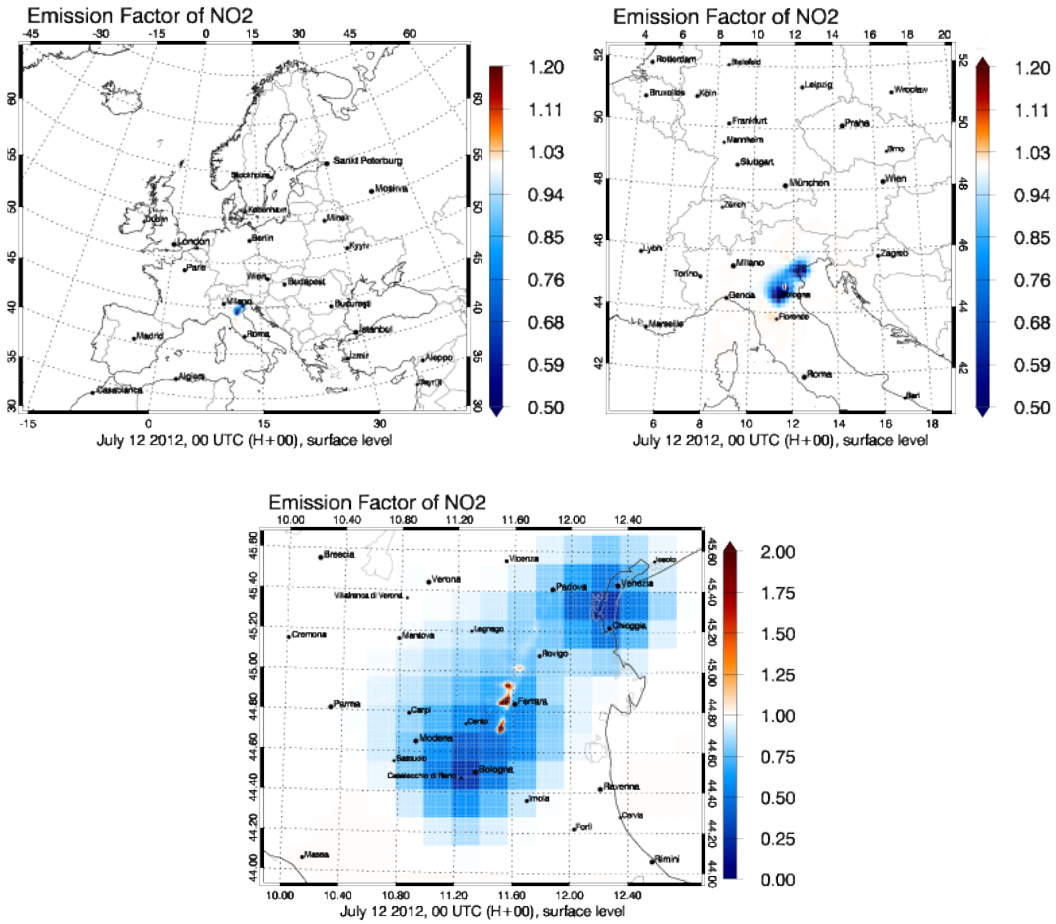

Figure 5.23: *The emission factors of NO_2 for the coarse grid of 15km resolution (left panel), its first nest of 5km resolution (middle panel) and the final nest of 1km resolution (right panel), resulted by the assimilation of the PEGASOS airborne data, on 12.02.2012.*

Fig. 5.23 illustrates the NO_2 emission factors at the surface level for all three domains. Here the scientific observation data from the PEGASOS campaign are efficient to clarify detailed emission patterns, which are not received by the analysis of the routine data. The optimised emission factors for the finest nest in Po valley identify improved characteristics for the model's estimation of NO_2 emissions in the area of the Zeppelins flight track for the surface layer.

Having vertical profiles, let analyse the gas phase concentrations of the model simulations. The model's behaviour is depicted in Hovmøller plots, for NO_2 (Fig. 5.24), CO and O_3 (Fig. 5.25) over the ground observation station of SPC. Here it should be mentioned that for plotting simplicity the PBL

is represented as a solid black line, where the ML exist from sunrise until sunset, 'falling' then directly to NBL, leaving outside of the black curve, up to its highest peak, the RL. However, the model calculates the PBL height according to the meteorological data input from WRF.

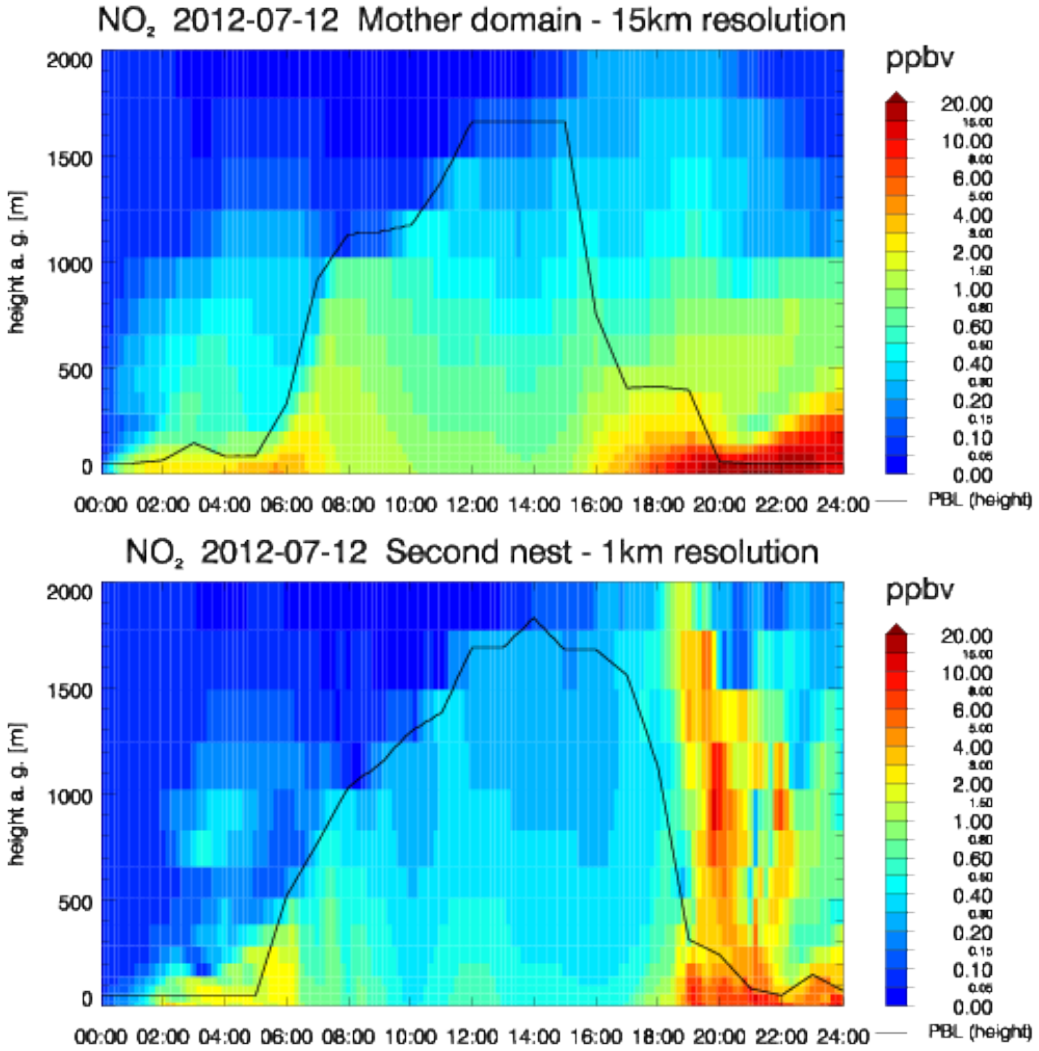

Figure 5.24: *Hovmøller plot: time series of the vertical NO_2 concentrations over the ground station of San Pietro Capofiume in Po valley; the European domain - 15 km resolved grid and 300 sec temporal resolution (left panel), the second nest over Po valley - 1 km resolved grid and 25 sec temporal resolution (right panel).*

In Fig. 5.24 the time series of NO_2 vertical concentrations are depicted in two cases; the assimilation in the 15 km resolution mother domain with 300 sec temporal resolution (first row) and the 1 km resolution second nest with 25 sec

temporal resolution (second row). The analysis of the model captures fairly clear the temporal variations of the emission patterns, here in case of NO_2. The morning and afternoon rush hour elevated NO_2 concentrations are depicted, more pronounced for the 300 sec temporal resolution of the coarse grid. The stable meteorological conditions trap the emissions in the NBL, increasing the NO_2 concentrations in the lower troposphere, fact that is represented in the analysis of the model.

Furthermore, the model output at the finer domain, with also higher temporal resolution, shows elevated NO_2 concentrations in the afternoon, expanded vertically through the PBL. The positive temperature gradient of the NBL causes convection and so vertical exchange, which is well captured by the assimilation result. In the less resolved case, this information is concentrated at the surface layer rather than represented in vertical details, as mixing processes, mainly the NO reactions that form NO_2, are fast with high reaction velocity.

Since the model analysis result is evaluated, a comparison with the assimilated air-borne observations from the PEGASOS campaign should take place. In Fig. 5.25 the time series for the vertical concentrations of NO_2 and O_3, as they are assimilated in the second nest (1 km spatial - 25 sec temporal resolution) are presented. The Zeppelin flight profiles during the campaign are included together with the measured concentrations of NO_2 and O_3, respectively.

In case of NO_2, during the beginning of the flight, close to the surface, high concentrations are measured, fact that captures the higher NO_2 emissions due to the morning rush hour. Compared with the analysed concentrations (background colour), there is a match with them and the air-borne data in upper altitudes, 500-700 m, from the beginning of the flight until around 8:00, as well as at lower altitudes, close to 300 m until 6:00. On the other hand, in the model's ML, the observed NO_2 concentrations are higher than the analysed ones, up to 300-400 m. The Zeppelin's observations capture clearly the layered structure of the PBL [*Li et al.*, 2014], though, differently than the model. In other words, the campaign data underline that the model calculates higher the PBL than it is measured. Although there is a correction of the analysed concentrations towards the observations, this does not influence the calculation of the PBL by the model. Thus, during the model analysis the mixing takes place in higher altitudes, fact that it is not verified by the observed information.

Looking at O_3 vertical time series of the analysed values, from sunset until sunrise, above 600 m height, the concentrations are stable and higher than in lower heights. This implies the existence of RL, which is not influenced by

emission or deposition processes at the surface. The lower O_3 concentrations in heights up to 600 m indicate the NBL, that later evolves in ML and arises the O_3 concentrations. During sunrise the O_3 in the NBL decreases, due to the reaction with the emitted NO and the NO_2 that start to insert into the troposphere, and further after 08:00 UTC the ML is characterised by stable high O_3 concentrations. During afternoon, the convection in NBL and the photochemical processing let NO react with O_3 producing NO_2, fact that it is illustrated in the lower afternoon concentrations of O_3 and the same time the elevated concentrations of NO_2. This layering of the PBL is also captured by the measured O_3 values during the PEGASOS campaign, but in different heights than in the model, identifying the ML much lower. There is a great matching between observed and analysed concentration around 5:00 - 8:00 up to roughly 400 m. Here, in the model's ML, the model analysis is benefited by the assimilation of the air-borne data, correcting the background. However, the higher measured data during the beginning of the flight and in heights higher than 400 m, implies a lower RL. Furthermore, the air-borne measurements show that the noon O_3 peak starts later than is analysed by the model.

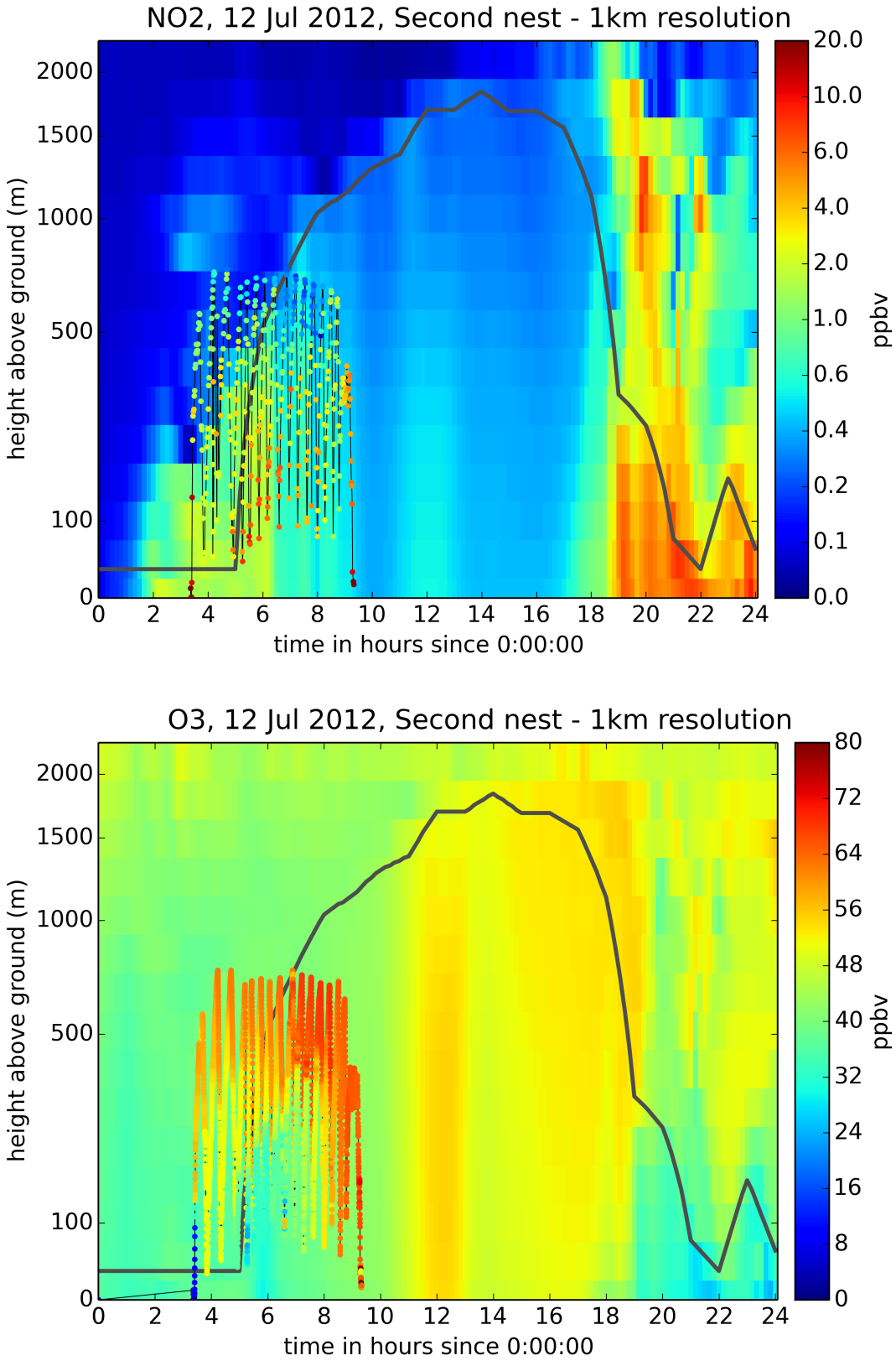

Figure 5.25: *Hovmøller plot: time series of the vertical NO and O_3 concentrations over the ground station of San Pietro Capofiume in Po valley for the second nest over Po valley - 1 km grid resolution and 25 sec temporal resolution.*

Conclusions

Summarising, the high resolution nesting technique is shown to face the representativity of NO_2 observations in the finest grid, being able to identify traffic emissions and more accurate emission patterns. Furthermore, since the temperature inversion increased the pollutants concentration in low levels, it is concluded that an improved parametrisation of the model is needed for the PBL area. To this end, the assimilation of the scientific campaign data is proved to enhance the model analysis for the vertical perspective of gas phase concentrations. The simulation results could capture the temporal variations of the emission patterns and could detect the layered structure of the PBL, benefiting by the assimilation of the airborne measurements.

Chapter 6

Summary and Conclusions

The present thesis intents to demonstrate the feasibility of the 4D-var inverse modelling to estimate the fate of anthropogenic emissions over urban areas in real case scenarios and to study the vertical distribution of concentrations in the PBL with the use of campaign data. The current version of the EURAD-IM system has been developed by a complete revision of the precursor version and many features have been improved to the assimilation procedure, in order to enhance the model performance.

The configuration of EURAD-IM system relating to emission rate optimisation was further developed, including the necessary technical updates to boost the model performance. In the CTM, the online calculation of emissions has been coupled to the joint optimisation of initial values and emission rates, increasing the accuracy of the assimilation result, due to the detailed emission distribution in time and height per source and per grid box. Furthermore, the emission factor background error covariances have been improved, leading to optimisation of the precondition to the minimisation problem. Additional correlations between emitted species were introduced, using polluter group based statistics. Moreover, rectified estimations of the standard deviations for the emission factors were provided, taking into account error statistics. As a result, the scaling emission factors estimation becomes more accurate and is maintained in realistic ranges, giving updated emission maps.

The first application of the EURAD-IM with the updated configuration takes place studying real case episodes. The joint optimisation of initial concentrations of the chemical constituents and emission rates has been successfully employed for a 24 hours assimilation window, letting the better known diurnal emission profiles be considered as strong constraint. In addition, the use of the nesting technique enables the CTM and its adjoint to telescope from regional scale down to urban scale in high resolution. Nests starting from 15

km and reaching the 1 km resolution in specific areas of interest, are applied for each case study. This configuration is proved to address efficiently the lack of representativity of observations. The latter result has been especially valuable for nitrogen oxide observations, for which perceptible optimisation is achieved, highlighted in the finest grids. Assessing the cost function evolution, a twofold evaluation is performed, by χ^2-validation and comparison with independent observations withheld from the assimilation procedure. The results are successful for the minimisation performance of the 4D-var algorithm. The study of an aestival high ozone and a high hibernal stagnant air episodes conclude that the EURAD-IM is able to optimise emission factors and estimate concentration distributions of emitted species and their products. Hence, improved analyses could be demonstrated by optimising the forecast skills of the model.

In the course of this work, a more in depth assessment of the vertical mixing in the PBL took place, by the assimilation of the high quality airborne data of the PEGASOS campaign in July 2012. The analysis of the vertical perspective of the gas phase concentrations captures fairly clear the temporal variation of the emission patterns. The influence by the meteorological conditions and the inverse layer on the concentrations of pollutants was proved by the assimilation of the campaign data. The layered structure of the PBL is detected in the vertical distribution of the analysed trace gases, benefiting from the clear view of the layering captured by the Zeppelin's observations. It is concluded that the rich campaign measurements have been essential to the model analysis as they give a more detailed insight than the routine data to the horizontal and vertical dispersion of the emissions in polluted areas. Moreover, the assimilation of airborne measurements could be applied to shed light to the development of the PBL and the processes there in, information that is missed when no campaign data are available. All in all, the campaign data can be a key for corrections and further improvements of the model's behaviour in forecasting and analysing the air quality in areas of interest.

In summary, the system demonstrated the ability to provide improved chemical consistent simulation results throughout the PBL, from polluted urban environments up to free troposphere, contributing greatly to air quality forecast and analysis. However, since the computational cost for such simulations is large, further technical updates in the model performance are needed. Future applications should include the transfer between the PBL and the free troposphere for a fully detailed tropospheric study of emission distribution and their potential global impact. More improvements to the EURAD-IM simulations could be achieved including the spatial variations of the correlations between the emitted species, SNAP-wise optimisation and improved chemical covariances. Besides, a complete study on the estimation of the anthropogenic

emissions and further campaign data should definitely include the assimilation of anthropogenic aerosols.

Bibliography

Ackermann, I. J. I. J., H. Hass, M. Memmesheimer, A. Ebel, F. S. F. S. Binkowski and U. Shankar, Modal aerosol dynamics model for Europe: Development and first applications, *Atmospheric Environment*, *32*, (17), 2981–2999, 1998.

Arakawa, A. and V. Lamb, Computational design for the basic dynamical processes of UCLA general circulation model, *Math. Comp. Phys.*, *17*, 174–264, 1977.

ArcGIS, ArcGIS 10.3 for Desktop system requirements. http://desktop.arcgis.com/en/.

Barret, B., E. Le Flochmoen, B. Sauvage, E. Pavelin, M. Matricardi and J. P. Cammas, The detection of post-monsoon tropospheric ozone variability over south Asia using IASI data, *Atmospheric Chemistry and Physics*, *11*, (18), 9533–9548, 2011.

Bechle, M. J., D. B. Millet and Julian D. Marshall, Remote sensing of exposure to NO 2 : Satellite versus ground-based measurement in a large urban area, *Atmospheric Environment*, *69*, 345–353, 2013.

Beychok, M. R., *Fundamentals of Stack Gas Dispersion*, Milton R. Beychok, 2005.

Bocquet, M., Reconstruction of an atmospheric tracer source using the principle of maximum entropy. II: Applications, *Quarterly Journal of the Royal Meteorological Society*, *131*, (610), 2209–2223, 2005*a*.

Bocquet, M., Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, *Quarterly Journal of the Royal Meteorological Society*, *131*,, (No. 610), 2191–2208, 2005*b*.

Bocquet, M., Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, *138*, (664, A), 664–681, 2012.

Bocquet, M., H. Elbern, H. Eskes, M. Hirtl, R. Žabkar, G. R. Carmichael, J. Flemming, A. Inness, M. Pagowski, J. L. Pérez Camaño, P. E. Saide, R. San Jose, M. Sofiev, J. Vira, A. Baklanov, C. Carnevale, G. Grell and C. Seigneur, Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, *Atmospheric Chemistry and Physics*, *15*, (10), 5325–5358, 2015.

Bott, A., A positive definite scheme obtained by nonlinear renormalization of the advective fluxes, *Monthly Weather Review*, *117*, (11), 2626–2632, 1989.

Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance and A. P. H. Goede, SCIAMACHY: mission objectives and measurement modes, *J. Atmos. Sci.*, *56*, 127–150, 1999.

Bowman, F. M., J. R. Odum and J. H. Seinfeld, Mathematical Model for Gas-Particle Partitionin G of Secondary Organic Aerosols, *Atmospheric Environment*, *31*, (23), 3921–3931, 1997.

Brasseur, G. and D. Jacob, *Mathematical Modeling of Atmospheric Chemistry*, G.P. Brasseur, 2013.

Brasseur, G., J.-F. Müller, X. X. Tie and L. Horowitz, Tropospheric ozone and climate: past, present and future, in *Present and Future of Modeling Global Environmental Change: Toward Integrated Modeling*, edited by T. Matsuno and H. Kida, 63–75, TERRAPUB, 2001.

Buizza, R., C. Cardinali, G. Kelly and J. N. Thepaut, The value of targeted observations, in *ECMWF Newsletter*, vol 111, 11–20, ECMWF, 2007.

Callies, J., E. Corpaccioli, M. Eisinger, A. Hahne and A. Lefebvre, GOME-2 - Metop's second-generation sensor for operational ozone monitoring, *ESA Bulletin-European Space Agency*, *102*, (may), 28–36, 2000.

Carmichael, G. R., A. Sandu, T. Chai, D. N. Daescu, E. M. Constantinescu and Y. Tang, Predicting air quality: Improvements through advanced methods to integrate models and measurements, *Journal of Computational Physics*, *227*, (7), 3540–3571, 2008.

Chai, T., G. R. Carmichael, A. Sandu, Y. Tang and D. N. Daescu, Chemical data assimilation of Transport and Chemical Evolution over the

Pacific (TRACE-P) aircraft measurements, *Journal of Geophysical Research: Atmospheres*, *111*, (2), 1–18, 2006.

Chai, T., G. R. Carmichael, Y. Tang, A. Sandu, A. Heckel, A. Richter and J. P. Burrows, Regional NOx emission inversion through a four-dimensional variational approach using SCIAMACHY tropospheric NO2 column observations, *Atmospheric Environment*, *43*, (32), 5046–5055, 2009.

Chang, J. S., R. A. Brost, I. S. A. Isaksen, S. Madronich, P. Middleton, W. R. Stockwell and C. J. Walcek, A three-dimensional Eulerian acid deposition model: Physical concepts and formulation, *Journal of Geophysical Research*, *92*, (D12), 14681–14700, 1987.

Chao, W. C. and L. P. Chang, Development of a four-dimensional variational analysis system using the adjoint method at GLA Part I: Dynamics, *Month. Weath. Rev.*, *120*, (8), 1661–1673, 1992.

Chapnik, B., G. Desroziers, F. Rabier and O. Talagrand, Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, *Quarterly Journal of the Royal Meteorological Society*, *132*, (615), 543–565, 2006.

Choi, H., Y. H. Zhang and K. H. Kim, Sudden high concentration of TSP affected by atmospheric boundary layer in Seoul metropolitan area during duststorm period, *Environment International*, *34*, (5), 635–647, 2008.

Clerbaux, C., A. Boynard, L. Clarisse, M. George, J. Hadji-Lazaro, H. Herbin, D. Hurtmans, M. Pommier, A. Razavi, S. Turquety, C. Wespes and P.-F. Coheur, Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, *Atmospheric Chemistry and Physics*, *9*, 6041–6054, 2009.

Cooper, M., R. V. Martin, C. Wespes, P.-F. Coheur, C. Clerbaux and L. T. Murray, Tropospheric nitric acid columns from the IASI satellite instrument interpreted with a chemical transport model: Implications for parameterizations of nitric oxide production by lightning, *Journal of Geophysical Research*, 68–79, 2014.

Courtier, P., Dual formulation of four-dimensional variational assimilation, *Q. J. R. Meteorol. Soc.*, *123*, (544), 2449–2461, 1997.

Crank, J. and P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type,

Mathematical Proceedings of the Cambridge Philosophical Society, *43*, (01), 50–67, 1947.

Cuesta, J., M. Eremenko, X. Liu, G. Dufour, Z. Cai, M. Höpfner, T. Von Clarmann, P. Sellitto, G. Foret, B. Gaubert, M. Beekmann, J. Orphal, K. Chance, R. Spurr and J. M. Flaud, Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements over Europe, *Atmospheric Chemistry and Physics*, *13*, (19), 9675–9693, 2013.

Daley, R., *Atmospheric Data Analysis*, vol 2, Cambridge University Press, Cambridge, 1991.

Damian-Iordache, V., *KPP -Chemistry Simulation Development Environment*, Ph.D. Thesis, The University of Iowa, 1996.

Deeter, M. N., D. P. Edwards, J. C. Gille, L. K. Emmons, G. Francis, S.-P. Ho, D. Mao, D. Masters, H. Worden, J. R. Drummond and P. C. Novelli, The MOPITT version 4 CO product: Algorithm enhancements, validation, and long-term stability, *Journal of Geophysical Research*, *115*, (D7), 1–13, 2010.

Deeter, M. N., S. Martínez-Alonso, D. P. Edwards, L. K. Emmons, J. C. Gille, H. M. Worden, J. V. Pittman, B. C. Daube and S. C. Wofsy, Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000-2011, *Journal of Geophysical Research: Atmospheres*, *118*, (12), 6710–6725, 2013.

Dubovik, O., T. Lapyonok, Y. J. Kaufman, M. Chin, P. Ginoux, R. A. Kahn and A. Sinyuk, Retrieving global aerosol sources from satellites using inverse modeling, *Atmospheric Chemistry and Physics*, *8*, (2), 209–250, 2008.

EEA, European Environmental Agency, http://www.eea.europa.eu/.

EEA Technical report, Air pollution by ozone across Europe during summer 2010, technical Report No 6/2011, European Environmental Agency, Copenhagen, 2011.

Elbern, H. and H. Schmidt, A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, *Journal of Geophysical Research*, *104*, (D15), 18,518–583,598, 1999.

Elbern, H. and H. Schmidt, Ozone Episode Analysis by Four-Dimensional Variational Chemistry Data Assimilation, *Journal of Geophysical Research*, *D4*, (106), 3569–3590, 2001.

Elbern, H., H. Schmidt and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *Journal of Geophysical Research*, *102*, (D13), 15,967–15,985, 1997.

Elbern, H., H. Schmidt, O. Talagrand and A. Ebel, 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environmental Modelling and Software*, *15*, (6-7 SPEC. ISS), 539–548, 2000.

Elbern, H., A. Strunk, H. Schmidt and O. Talagrand, Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmospheric Chemistry and Physics*, *7*, (1), 3749–3769, 2007*a*.

Elbern, H., A. Strunk, H. Schmidt and O. Talagrand, Chemical weather analysis optimisation with emission impact estimation using nested four-dimensional variational chemistry data assimilation, in *Air, Water and Soil Quality Modelling for Risk and Impact Assessment*, edited by A. Ebel and T. Davitashvili, 315–325, Springer, 2007*b*.

Elbern, H., A. Strunk, E. Friese and L. Nieradzik, Assessment of Source/Receptor Relations by Inverse Modelling and Chemical Data Assimilation, in *Persistent Pollution Past, Present and Future*, edited by M. Quante, R. Ebinghaus, and G. Flöser, School of Environmental Research - Helmholtz-Zentrum Geesthacht, 1st Auflage, 2011.

Elbern, H., A. Benedetti, F. Chevallier, R. Engelen, J. Flemming, A. Inness, J. Kaiser, S. Massart and Coworkers, Assimilation of Satellite Data for Atmospheric Composition, in *ECMWF Seminar*, 2014.

Emili, E., B. Barret, S. Massart, E. Le Flochmoen, A. Piacentini, L. El Amraoui, O. Pannekoucke and D. Cariolle, Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical transport model, *Atmospheric Chemistry and Physics*, *14*, (1), 177–198, 2014.

Engelen, R., E. Andersson, F. Chevallier, A. Hollingsworth, M. Matricardi, A. P. McNally, J. N. Thépaut and P. D. Watts, Estimating atmospheric CO2 from advanced infrared satellite radiances within an operational 4D-Var data assimilation system: Methodology and first results, *Journal of Geophysical Research*, *109*, (D19309), 2004.

Enting, I. G. and G. N. Newsam, Inverse problems in atmospheric constituent studies: II. Sources in the free atmosphere, *Inverse Problems*, *6*, (3), 349–362, 1990.

European Parliament and Council of the European Union, Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, *Official Journal of the European Communities*, 1–43, 2008.

Finardi, S., C. Silibello, A. D'Allura and P. Radice, Analysis of pollutants exchange between the Po Valley and the surrounding European region, *Urban Climate*, 2014.

Fisher, M. and D. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc.*, *121*, (527), 1681–1704, 1995.

Fishman, J., K. W. Bowman, J. P. Burrows, A. Richter, K. V. Chance, D. P. Edwards, R. V. Martin, G. a. Morris, R. B. Pierce, J. R. Ziemke, J. a. Al-Saadi, J. K. Creilson, T. K. Schaack and A. M. Thompson, Remote sensing of tropospheric pollution from space, *Bulletin of the American Meteorological Society*, *89*, (6), 805–821, 2008.

FORTH/ICE-HT, PEGASOS - Home page http://pegasos.iceht.forth.gr/.

Friese, E. and A. Ebel, Temperature dependent thermodynamic model of the system H(+)-NH(+)-Na(+)-SO-NO-Cl-HO., *J. Phys. Chem. A*, *114*, (43), 11595–11631, 2010.

Gomm, S., *Luftgestützte Messung von HOx-Radikalkonzentrationen mittels Laser-induzierter Fluoreszenz auf einem Zeppelin NT: Untersuchung der atmosphärischen Oxidationsstärke der unteren Troposphäre*, Ph.D. Thesis, Forscungszentrum Juelich GmbH, Troposphere (IEK-8), 2014.

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons and X. Wang, The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions, *Geosci. Model Dev.*, *5*, 1471–1492, 2012.

Haklay, M. and P. Weber, OpenStreetMap: User-Generated Street Maps, *IEEE Pervasive Computing*, *7*, (4), 12–18, 2008.

Hamburger, T., G. McMeeking, A. Minikin, W. Birmili, M. Dall'Osto, C. O'Dowd, H. Flentje, B. Henzing, H. Junninen, A. Kristensson, G. De Leeuw, A. Stohl, J. F. Burkhart, H. Coe, R. Krejci and A. Petzold, Overview of the synoptic and pollution situation over Europe during the EUCAARI-LONGREX field campaign, *Atmospheric Chemistry and Physics*, *11*, (3), 1065–1082, 2011.

Hanna, S., J. Chang and M. Fernau, Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, *Atmospheric Environment*, *32*, 3619–3628, 1998.

Hanna, S. R., Z. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau and D. A. Hansen, Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmospheric Environment*, *35*, (5), 891–903, 2001.

Hass, H., H. J. Jakobs and M. Memmesheimer, Analysis of a regional model (EURAD) near surface gas concentration predictions using observations from networks, *Met. Atmos. Phys.*, *57*, 173–200, 1995.

Holtslag, A. A. M. and F. T. M. Nieuwstadt, Scaling the atmospheric boundary layer, *Boundary-Layer Meteorology*, *36*, (1-2), 201–209, 1986.

Hooghiemstra, P. B., M. C. Krol, J. F. Meirink, P. Bergamaschi, G. R. Van Der Werf, P. C. Novelli, I. Aben and T. Röckmann, Optimizing global CO emission estimates using a four-dimensional variational data assimilation system and surface network observations, *Atmospheric Chemistry and Physics*, *11*, (10), 4705–4723, 2011.

Hooghiemstra, P. B., M. C. Krol, P. Bergamaschi, a. T. J. De Laat, G. R. Van Der Werf, P. C. Novelli, M. N. Deeter, I. Aben and T. Rckmann, Comparing optimized CO emission estimates using MOPITT or NOAA surface network observations, *Journal of Geophysical Research: Atmospheres*, *117*, (6), 1–23, 2012.

Hourdin, F., O. Talagrand and A. Idelkadi, Eulerian backtracking of atmospheric tracers. II: Numerical aspects, *Q. J. R. Meteorol. Soc.*, *132*, 585–603, 2006.

Ide, K., P. Courtier, M. Ghil and A. C. Lorenc, Unifed notation for data assimilation: operational, sequential and variational, *J. Metrorol. Soc. Japan*, *75*, (1{B}), 181–189, 1997.

Inness, A., A.-M. Blechschmidt, I. Bouarar, S. Chabrillat, M. Crepulja, R. J. Engelen, H. Eskes, J. Flemming, A. Gaudel, F. Hendrick, V. Huijnen, L. Jones, J. Kapsomenakis, E. Katragkou, A. Keppens, B. Langerock, M. de Mazière, D. Melas, M. Parrington, V. H. Peuch, M. Razinger, A. Richter, M. G. Schultz, M. Suttie, V. Thouret, M. Vrekoussis, A. Wagner and C. Zerefos, Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS, *Atmospheric Chemistry and Physics*, *15*, (9), 5275–5303, 2015.

Issartel, J. P., Rebuilding sources of linear tracers after atmospheric concentration measurements, *Atmospheric Chemistry and Physics*, *3*, 2111–2125, 2003.

Jacobson, M. Z., *Fundamentals of Atmospheric Modeling*, Cambridge, 2000.

Jäger, J., *Airborne VOC measurements on board the Zeppelin NT during the PEGASOS campaigns in 2012 deploying the improved Fast-GC-MSD System*, Ph.D. Thesis, Forscungszentrum Juelich GmbH, Troposphere (IEK-8), 2014.

Jakobs, J. H., H. Feldmann, H. Hass and M. Memmesheimer, The Use of Nested Models for Air Pollution Studies: An Application of the EURAD Model to a SANA Episode, *American Meteorological Society*, *34*, (6), 1301–1319, 1995.

Kaiser, J. W., A. Heil, M. O. Andreae, A. Benedetti, N. Chubarova, L. Jones, J. J. Morcrette, M. Razinger, M. G. Schultz, M. Suttie and G. R. Van Der Werf, Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, *Biogeosciences*, *9*, 527–554, 2012.

Kaiser, J., G. M. Wolfe, B. Bohn, S. Broch, H. Fuchs, L. N. Ganzeveld, S. Gomm, R. Häseler, A. Hofzumahaus, F. Holland, J. Jäger, X. Li, I. Lohse, K. Lu, a. S. H. Prévôt, F. Rohrer, R. Wegener, R. Wolf, T. F. Mentel, A. Kiendler-Scharr, A. Wahner and F. N. Keutsch, Evidence for an unidentified non-photochemical ground-level source of formaldehyde in the Po Valley with potential implications for ozone production, *Atmospheric Chemistry and Physics*, *15*, (3), 1289–1298, 2015.

Kalnay, E., *Atmospheric modeling, data assimilation, and predictability*, vol 54, Cambridge Univ. Press, 2003.

Kaminski, T., M. Heimann and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport: 1. Adjoint model and Jacobian matrix, *Journal of Geophysical Research*, *104*, (D15), 18535–18553, 1999*a*.

Kaminski, T., M. Heimann and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO2 in the 1980s, *Journal of Geophysical Research*, *104*, (D15), 18555–18581, 1999*b*.

Kaminski, T., W. Knorr, P. Rayner and M. Heimann, Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Glob. Biogeochem. Cycles*, *16*, (4), 2002.

Kanakidou, M., M. Dameris, H. Elbern, M. Beekmann, I. Konovalov, L. Nieradzik, A. Strunk and M. Krol, Synergistic use of retrieved trace constituents distributions and numerical modelling, in *The remote sensing of tropospheric composition from space*, edited by J. Burrows, U. Platt, and P. Borrell, Springer, 2011.

Koohkan, M. R., M. Bocquet, Y. Roustan, Y. Kim and C. Seigneur, Estimation of volatile organic compound emissions for Europe using data assimilation, *Atmospheric Chemistry and Physics*, *13*, (12), 5887–5905, 2013.

Kuenen, J. J. P., A. J. H. Visschedijk, M. Jozwicka and H. A. C. Denier van der Gon, TNO-MACC_II emission inventory: a multi-year (20032009) consistent high-resolution European emission inventory for air quality modelling, *Atmospheric Chemistry and Physics*, *14*, 10963–10976, 2014.

Lahoz, W. A. and P. Schneider, Data assimilation: making sense of Earth Observation, *Frontiers in Environmental Science*, *2*, (May), 1–28, 2014.

Lahoz, W. A., B. V. Khattatov and R. Ménard, *Data Assimilation*, Springer, 2010.

Levelt, P., E. Hilsenrath, G. Leppelmeier, G. van den Oord, P. Bhartia, Tamminen, J. de Haan and J. Veefkind, Science Objectives of the Ozone Monitoring Instrument, *IEEE Transactions on Geoscience and Remote Sensing*, *44*, (5), 1199–1208, 2006.

Li, Y. P. P., H. Elbern, K. D. D. Lu, E. Friese, A. Kiendler-Scharr, T. F. Mentel, X. S. S. Wang, A. Wahner and Y. H. H. Zhang, Updated aerosol module and its application to simulate secondary organic

aerosols during IMPACT campaign May 2008, *Atmospheric Chemistry and Physics*, *13*, (13), 6289–6304, 2013.

Li, X., F. Rohrer, A. Hofzumahaus, T. Brauers, R. Häseler, B. Bohn, S. Broch, H. Fuchs, S. Gomm, F. Holland, J. Jäger, J. Kaiser, F. N. Keutsch, I. Lohse, K. Lu, R. Tillmann, R. Wegener, G. M. Wolfe, T. F. Mentel, A. Kiendler-Scharr and A. Wahner, Missing gas-phase source of HONO inferred from Zeppelin measurements in the troposphere, *Science*, *344*, (April), 292–297, 2014.

Liu, D. C. and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Math. Programming*, *45*, 503–528, 1989.

Liu, L., S. Andreani-Aksoyoglu, J. Keller, C. Ordóñez, W. Junkermann, C. Hak, G. O. Braathen, S. Reimann, C. Astorga-Llorens, M. Schultz, a. S. H. Prévôt and I. S. a. Isaksen, A photochemical modeling study of ozone and formaldehyde generation and budget in the Po basin, *Journal of Geophysical Research: Atmospheres*, *112*, (22), 1–16, 2007.

Lurmann, F., A. Wexler, S. Pandis, S. Musarra, N. Kumar and J. Seinfeld, Modelling Urban and Regional Aerosols - II. Application To California's South Coast Air Basin, *Atmospheric Environment*, *31*, (17), 2695–2715, 1997.

Lyamani, H., J. Fernández-Gálvez, D. Pérez-Ramírez, A. Valenzuela, M. Antón, I. Alados, G. Titos, F. J. Olmo and L. Alados-Arboledas, Aerosol properties over two urban sites in South Spain during an extended stagnation episode in winter season, *Atmospheric Environment*, *62*, (SEPTEMBER 2012), 424–432, 2012.

MACC, Monitoring Atmospheric Composition and Climate, http://www.gmes-atmosphere.eu/.

Madronich, S. and G. Weller, Numerical integration errors in calculated tropospheric photodissociation rate coefficients, *Journal of Atmospheric Chemistry*, *10*, (3), 289–300, 1990.

Marécal, V., V.-H. Peuch, C. Andersson, S. Andersson, J. Arteta, M. Beekmann, A. Benedictow, R. Bergström, B. Bessagnet, A. Cansado, F. Chéroux, A. Colette, A. Coman, R. L. Curier, H. a. C. Denier van der Gon, A. Drouin, H. Elbern, E. Emili,

R. J. Engelen, H. J. Eskes, G. Foret, E. Friese, M. Gauss, C. Giannaros, J. Guth, M. Joly, E. Jaumouillé, B. Josse, N. Kadygrov, J. W. Kaiser, K. Krajsek, J. Kuenen, U. Kumar, N. Liora, E. Lopez, L. Malherbe, I. Martinez, D. Melas, F. Meleux, L. Menut, P. Moinat, T. Morales, J. Parmentier, A. Piacentini, M. Plu, A. Poupkou, S. Queguiner, L. Robertson, L. Rouïl, M. Schaap, A. Segers, M. Sofiev, M. Thomas, R. Timmermans, Á. Valdebenito, P. van Velthoven, R. van Versendaal, J. Vira and A. Ung, A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, *Geoscientific Model Development Discussions*, *8*, (3), 2739–2806, 2015.

Martin, R. V., Satellite remote sensing of surface air quality, *Atmospheric Environment*, *42*, (34), 7823–7843, 2008.

McRae, G., W. Goodin and J. Seinfeld, Numerical solution of the atmospheric diffusion equation for chemically reacting flows, *Journal of Computational Physics*, *45*, 1–42, 1982.

Meirink, J. F., P. Bergamaschi and M. C. Krol, Technical Note: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, *Atmospheric Chemistry and Physics Discussions*, *8*, (3), 12023–12052, 2008.

Meleux, F., F. Solmon and F. Giorgi, Increase in summer European ozone amounts due to climate change, *Atmospheric Environment*, *41*, (35), 7577–7587, 2007.

Memmesheimer, M., E. Friese, A. Ebel, H. Jakobs, H. Feldmann, C. Kessler and G. Piekorz, Long-term simulations of particulate matter in Europe on different scales using sequential nesting of a regional model, *Int. J. Environment and Pollution*, *22*, (1-2), 108–132, 2004.

Miyazaki, K., H. J. Eskes, K. Sudo, M. Takigawa, M. van Weele and K. F. Boersma, Simultaneous assimilation of satellite NO2, O-3, CO, and HNO3 data for the analysis of tropospheric chemical composition and emissions, *Atmospheric Chemistry and Physics*, *12*, (20), 9545–9579, 2012.

Miyazaki, K., H. J. Eskes and K. Sudo, A tropospheric chemistry reanalysis for the years 20052012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data, *Atmospheric Chemistry and Physics*, *15*, (14), 8315–8348, 2015.

Mohnen, V., Data Quality Assessment - an overview for the TRACT 16./17. September 1992 field intensive, Report to the coordinator of tfs-lt1, IFU Garmisch-Partenkirchen, 1999.

Monteiro, A., A. Strunk, A. Carvalho, O. Tchepel, a. I. Miranda, C. Borrego, S. Saavedra, A. Rodríguez, J. Souto, J. Casares, E. Friese, H. Elbern and A. C. O. T. A. I. M. C. B. S. S. A. R. J. S. J. C. E. F. H. E. A. Monteiro A. Strunk, Investigating a high ozone episode in a rural mountain site, *Environmental Pollution*, *162*, 176–189, 2012.

Monteiro, A., I. Ribeiro, O. Tchepel, E. Sá, J. Ferreira, A. Carvalho, V. Martins, A. Strunk, S. Galmarini, H. Elbern, M. Schaap, P. Builtjes, A. I. Miranda and C. Borrego, Bias Correction Techniques to Improve Air Quality Ensemble Predictions: Focus on O3 and PM Over Portugal, *Environ. Model. Assess.*, *18*, 533–546, 2013.

Muccino, J. C., N. F. Hubele and A. F. Bennett, Significance testing for variational assimilation, *Quarterly Journal of the Royal Meteorological Society*, *130*, (600), 1815–1838, 2004.

Muller, J. F. and T. Stavrakou, Inversion of CO and NOx emissions using the adjoint of the IMAGES model, *Atmospheric Chemistry and Physics*, *5*, 1157–1186, 2005.

Navon, I. M., Data Assimilation for numerical weather prediction: a review, in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, edited by S.-K. Park, chapter Data Assim, 21–65, Springer, 2009.

Newsam, G. N. and I. G. Enting, Inverse problems in atmospheric constituent studies. I. Determination of surface sources under a diffusive transport approximation, *Inverse Problems*, *4*, (4), 1037–1054, 1988.

Nieradzik, L., Application of a High Dimensional Model Representation on the atmospheric aerosol module MADE of the EURAD-CTM, Master Thesis, Institut für Geophysik und Meteorologie der Universität zu Köln, 2005.

Nieradzik, L., *Four-Dimensional Variational Assimilation of Aerosol Data from In-situ and Remote Sensing Platforms*, Ph.D. Thesis, University of Cologne, 2011.

Nieradzik, L. P. and H. Elbern, Variational assimilation of combined satellite retrieved and in situ aerosol data in an advanced chemistry transport

model, in *Proceedings of the ESA Atmospheric Science Conference*, ESA, ESA-ESRIN, Frascati, 2006.

Nocedal, J., Updating Quasi-Newton Matrices With Limited Storage, *Math. Comput.*, *35*, (151), 773–782, 1980.

Pandis, S., Description of Work PEGASOS, *European Comission, Framework Programme 7 (FP7-ENV-2010.1.1.2-1)*, 2010.

Pandis, S. N., R. A. Harley, G. R. Cass and J. H. Seinfeld, Secondary organic aerosol formation and transport, *Atmospheric Environment. Part A. General Topics*, *26*, (13), 2269–2282, 1992.

Petroff, A. and L. Zhang, Development and validation of a size-resolved particle dry deposition scheme for application in aerosol transport models, *Geoscientific Model Development*, *3*, (2), 753–769, 2010.

Pierce, T., C. Hogrefe, S. Trivikrama Rao, P. S. Porter and J. Y. Ku, Dynamic evaluation of a regional air quality model: Assessing the emissions-induced weekly ozone cycle, *Atmospheric Environment*, *44*, (29), 3583–3596, 2010.

Pilinis, C., P. E. Charalampidis, N. Mihalopoulos and S. N. Pandis, Contribution of particulate water to the measured aerosol optical properties of aged aerosol, *Atmospheric Environment*, *82*, 144–153, 2014.

Pison, I., L. Menut and G. Bergametti, Inverse modeling of surface NOx anthropogenic emission fluxes in the Paris area during the Air Pollution Over Paris Region (ESQUIF) campaign, *Journal of Geophysical Research: Atmospheres*, *112*, (24), 1–15, 2007.

Quélo, D., V. Mallet and B. Sportisse, Inverse Modeling of NOx Emissions at Regional Scale over Northern France. Preliminary Investigation of the Second-Order Sensitivity, *Journal of Geophysical Research*, *110*, (D24310), 2005.

Rabitz, H., Ö. F. Ali, J. Shorter and K. Shim, Efficient input-output model representations, *Computer Physics Communications*, *117*, (1), 11–20, 1999.

Ribeiro, I., Monteiro A., F. A.P., M. A.C., L. M., B. C. and M. A.I., Air quality modelling as a supplementary assessment method in the framework of the European air quality Directive, *15th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. 6-9 May 2013, Madrid, Spain*, 270–274, 2013.

Richter, A., M. Begoin, A. Hilboll and J. P. Burrows, An improved NO2 retrieval for the GOME-2 satellite instrument, *Atmospheric Measurement Techniques*, *4*, (6), 1147–1159, 2011.

Rodgers, C. D., *Inverse Methods for Atmospheric Sounding*, World Scientific, 2000.

Roselle, S. and F. Binkowski, Cloud Dynamics and Chemistry, in *Science Algorithms of the EPA Models–3 Community Multiscale Air Quality (CMAQ) Modeling System, EPA 600/R-99-030*, edited by D. Byun and J. Ching, U.S. Environ. Prot. Agency, Research Triangle Park, N.C., 1999.

Safieddine, S., A. Boynard, P.-F. Coheur, D. Hurtmans, G. Pfister, B. Quennehen, J. L. Thomas, J.-C. Raut, K. S. Law, Z. Klimont, J. Hadji-Lazaro, M. George and C. Clerbaux, Summertime tropospheric ozone assessment over the Mediterranean region using the thermal infrared IASI/MetOp sounder and the WRF-Chem model, *Atmospheric Chemistry and Physics Discussions*, *14*, (8), 12377–12408, 2014.

Sandrini, S., S. Fuzzi, A. Piazzalunga, P. Prati, P. Bonasoni, F. Cavalli, M. C. Bove, M. Calvello, D. Cappelletti, C. Colombi, D. Contini, G. de Gennaro, A. Di Gilio, P. Fermo, L. Ferrero, V. Gianelle, M. Giugliano, P. Ielpo, G. Lonati, A. Marinoni, D. Massabò, U. Molteni, B. Moroni, G. Pavese, C. Perrino, M. G. Perrone, M. R. Perrone, J.-P. Putaud, T. Sargolini, R. Vecchi and S. Gilardoni, Spatial and seasonal variability of carbonaceous aerosol across Italy, *Atmospheric Environment*, *99*, 587–598, 2014.

Sandu, A. and T. Chai, Chemical data assimilation-An overview, *Atmosphere*, *2*, (3), 426–463, 2011.

Sandu, A. and R. Sander, Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmospheric Chemistry and Physics*, *6*, (5), 187–195, 2006.

Sandu, A. and L. Zhang, Discrete second order adjoints in atmospheric chemical transport modeling, *Journal of Computational Physics*, *227*, (12), 5949–5983, 2007.

Sandu, A., F. Potra, G. Carmichael and V. Damian, Efficient implementation of fully implicit methods for atmospheric chemical kinetics, *Journal of Computational Physics*, *129*, 101–110, 1996.

Sandu, A., D. N. Daescu and G. R. Carmichael, Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I - Theory and software tools, *Atmospheric Environment*, *37*, (36), 5083–5096, 2003.

Sandu, A., D. N. Daescu, G. R. Carmichael and T. Chai, Adjoint sensitivity analysis of regional air quality models, *Journal of Computational Physics*, *204*, (1), 222–252, 2005.

Sasaki, Y., Some Basic Formalisms in Numerical Variational Analysis, *Monthly Weather Review*, *98*, (12), 875–883, 1970.

Schell, B., I. J. Ackermann, H. Hass, F. S. Binkowski and A. Ebel, Modeling the formation of secondary organic aerosol within a comprehensive air quality model system, *Journal of Geophysical Research*, *106*, (D22), 28,275–28,293, 2001.

Schmidt, H. and D. Martin, Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale, *Journal of Geophysical Research*, *108*, (D17), 2003.

Sillmann, J., L. Pozzoli, E. Vignati, S. Kloster and J. Feichter, Aerosol effect on climate extremes in Europe under different future scenarios, *Geophysical Research Letters*, *40*, (10), 2290–2295, 2013.

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X.-Y. Huang, W. Wang and J. Powers, *A Description of the Advanced Research WRF Version 3*, Boulder, 2008.

Sofiev, M., U. Berger, M. Prank, J. Vira, J. Arteta, J. Belmonte, K.-C. Bergmann, F. Chéroux, H. Elbern, E. Friese, C. Galan, R. Gehrig, D. Khvorostyanov, R. Kranenburg, U. Kumar, V. Marécal, F. Meleux, L. Menut, A.-M. Pessi, L. Robertson, O. Ritenberga, V. Rodinkova, A. Saarto, A. Segers, E. Severova, I. Sauliene, P. Siljamo, B. M. Steensen, E. Teinemaa, M. Thibaudon and V.-H. Peuch, MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe, *Atmospheric Chemistry and Physics*, *15*, (14), 8115–8130, 2015.

Solomon, S., D. Qin, M. Manning, R. B. Alley, T. Berntsen, N. L. Bindoff, Z. Chen, A. Chidthaisong, J. M. Gregory, G. C. Hegerl, M. Heimann, B. Hewitson, B. J. Hoskins, F. Joos, J. Jouzel, V. Kattsov, U. Lohmann, T. Matsuno, M. Molina, N. Nicholls, J. Overpeck, G. Raga, V. Ramaswamy, J. Ren, M. Rusticucci,

R. Somerville, T. F. Stocker, P. Whetton, D. Wood and R. a. Wratt, *2007: Technical Summary. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge,, United Kingdom and New York, NY, USA, 2007.

Stjernholm, M., CORINE Land Cover 2006, technical report, Aarhus University, Denmark, 2009.

Stockwell, W. R., P. Middleton and J. S. Chang, The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *Journal of Geophysical Research*, *95*, (D10), 16343–16367, 1990.

Stockwell, W. R., F. Kirchner, M. Kuhn and S. Seefeld, A new mechanism for regional atmospheric chemistry modeling, *Journal of Geophysical Research*, *102*, (D22), 25847, 1997.

Strader, R., F. Lurmann and S. N. Pandis, Evaluation of secondary organic aerosol formation in winter, *Atmospheric Environment*, *33*, (29), 4849–4863, 1999.

Strunk, A., *Tropospheric Chemical State Estimation by Four-Dimensional Variational Data Assimilation on Nested Grids*, Ph.D. Thesis, Institut für Geophysik und Meteorologie der Universität zu Köln, 2006.

Stull, R. B., *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers, Ordrecht I Boston I London, 1988.

Talagrand, O., Evaluation of assimilation algorithms, in *Data Assimilation: Making Sense of Observations*, 217–240, Springer Berlin Heidelberg, 2010.

Tang, X., J. Zhu, Z. F. Wang and A. Gbaguidi, Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions, *Atmospheric Chemistry and Physics*, *11*, (24), 12901–12916, 2011.

Tie, X., S. Madronich, Stacy Walters, R. Zhang, P. Rasch and W. Collins, Effect of clouds on photolysis and oxidants in the troposphere, *Journal of Geophysical Research*, *108*, (D20), 2003.

Turner, D. B., *Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling, Second Edition*, CRC Press, 1994.

USEPA, United States Environmental Protection Agency http://www.epa.gov.

Weaver, A. and P. Courtier, Correlation Modelling on the Sphere Using a Generalized Diffusion Equation, *Q. J. R. Meteorol. Soc.*, *127*, (575), 1815–1846, 2001.

Winiarek, V., M. Bocquet, N. Duhanyan, Y. Roustan, O. Saunier and A. Mathieu, Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations, *Atmospheric Environment*, *82*, 268–279, 2014.

Yanenko, N., *The Method of Fractional Steps*, Springer, 1971.

Yumimoto, K. and I. Uno, Adjoint inverse modeling of CO emissions over Eastern Asia using four-dimensional variational data assimilation, *Atmospheric Environment*, *40*, (35), 6836–6845, 2006.

Yumimoto, K., I. Uno, N. Sugimoto, A. Shimizu and S. Satake, Adjoint inverse modeling of dust emission and transport over East Asia, *Geophysical Research Letters*, *34*, (8), 1–6, 2007.

Zhang, L., J. R. Brook and R. Vet, A revised parameterization for gaseous dry deposition in air-quality models, *Atmospheric Chemistry and Physics*, *3*, (2), 2067–2082, 2003.

Zhang, Y., M. Bocquet, V. Mallet, C. Seigneur and A. Baklanov, Real-time air quality forecasting, Part II: State of the science, current research needs, and future prospects, *Atmospheric Environment*, *60*, 656–676, 2012.

Zyryanov, D., G. Foret, M. Eremenko, M. Beekmann, J. P. Cammas, M. D'Isidoro, H. Elbern, J. Flemming, E. Friese, I. Kioutsioutkis, A. Maurizi, D. Melas, F. Meleux, L. Menut, P. Moinat, V. H. Peuch, A. Poupkou, M. Razinger, M. Schultz, O. Stein, a. M. Suttie, A. Valdebenito, C. Zerefos, G. Dufour, G. Bergametti and J. M. Flaud, 3-D evaluation of tropospheric ozone simulations by an ensemble of regional Chemistry Transport Model, *Atmospheric Chemistry and Physics*, *12*, (7), 3219–3240, 2012.

Acknowledgements

A very demanding but beautiful route into the scientific research is accomplished by this dissertation and I can't but acknowledge the generous support that I have received by my entourage.

I acknowledge the assistance and advice I received from my supervisor PD Dr. Hendrik Elbern during all the last years that we work together, his 'open door' policy and his comments while proof-reading this work. But most of all I acknowledge his support and his confidence in my work and abilities.

Dealing with the challenging field of modelling is straight connected with the unwavering support of Dr. Elmar Friese. I appreciate his help and advice during all the difficulties that I faced until the very end of this work.

Furthermore I would like to thank my colleagues Erna Bern, Philipp Franke, Dr. Luise Fröhlich, Ketevan Kasradze, Johannes Klimpt, Anne Caroline Lange and Dr. Isabel Ribeiro for their offer to contribute in ameliorating this manuscript by their comments, suggestions and translating skills.

By all means, I gratefully acknowledge *die Liebe Euradler*, all my wonderful colleges in RIU that since my first day in the Institute they welcomed me and they always made me feel like home. Being generous supporters in all of my moves, they have been my German family.

Last but not least, I would like to thank Ioannis Tsoukakis for being always there for me, encouraging every step on my way. Without his endless care and support this milestone wouldn't have been accomplished.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschliesslich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.
Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD. Dr. H. Elbern betreut worden.