Polluter group specific emission optimisation for regional air quality analyses using four-dimensional variational data assimilation

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

vorgelegt von

Pascal Martin Backes

aus Erkelenz

Jülich, 2023

Berichterstatter: Prof. Dr. H. Fuchs
Prof. Dr. S. Crewell

Tag der mündlichen Prüfung: 12.07.2023

Abstract

Pollutants in the atmosphere, such as nitrogen oxides and particulate matter, pose a threat to the environment and human health. In addition to natural sources, anthropogenic emissions contribute significantly to air pollution. Since emission rates cannot be measured directly, their estimates provided by research institutes and national environmental agencies are subject to considerable uncertainty. However, for accurate air quality forecasts using atmospheric chemical transport models such as the regional European Air pollution Dispersion - Inverse Model (EURAD-IM), reliable emission data are crucial. To correct the emission data of inventories based on observations of trace gas and aerosol concentrations in the atmosphere, the EURAD-IM comprises a four dimensional variational data assimilation system (4D-Var) that allows for simultaneous optimisation of initial concentrations and species-dependent emission corrections. In order to improve the knowledge about the sources of air pollution, in this work, a new approach is developed and implemented in the data assimilation system of the EURAD-IM to correct emissions individually for source categories such as road transport, industry and agriculture. For the distinction between the emissions of different source categories, the new approach exploits the spatial separation of emission sources of different categories as well as their characteristic diurnal emission profiles and chemical compositions. Assuming a fixed chemical composition of the emissions of the source categories within the grid cells, a full correlation between the emission corrections of the different chemical species is introduced. Furthermore, an anisotropic diffusion operator is implemented that increases the spatial correlation between the road traffic emission corrections of the grid cells along roads. To investigate the ability of the new development to distinguish between emissions of different sectors, two different types of simulations are performed. In identical twin experiments based on synthetic observations, scenarios with increased industrial and agricultural emissions and a simultaneous decrease in road transport emissions are simulated. The data assimilation system based on the new approach is able to reproduce the emission changes in the experiments for large parts of the model domain through the determined sector specific emission corrections. Furthermore, a study is performed in which the emissions within a two-week period in North Rhine-Westphalia are analysed using real observation data. It is shown that in this scenario a distinction of industrial and power plant emissions versus road transport emissions is possible through the sector specific emission optimisation. Moreover, changes in agricultural emissions can be specified due to their high NH_3 fraction. For all observed species, i.e. O_3, NO_2, SO_2, PM_{10} and $PM_{2.5}$, the agreement of the simulated with the observed concentrations is comparable to that of a reference simulation using the current EURAD-IM data assimilation system. An improvement of the results is expected through additional

observation data, especially of CO and CO_2 concentrations.

Kurzzusammenfassung

In der Atmosphäre enthaltene Schadstoffe wie Stickoxide und Feinstaub stellen eine Gefahr für die Umwelt und die menschliche Gesundheit dar. Neben natürlichen Quellen tragen anthropogene Emissionen wesentlich zur Luftverschmutzung bei. Da Emissionsraten nicht direkt gemessen werden können, sind deren Schätzungen, die von Forschungsinstituten und nationalen Umweltbehörden bereitgestellt werden, mit erheblichen Unsicherheiten verbunden. Für präzise Vorhersagen der Luftqualität mithilfe von atmosphärischen Chemietransportmodellen wie dem regionalen European Air pollution Dispersion - Inverse Model (EURAD-IM) sind verlässliche Emissionsdaten jedoch entscheidend. Zur Korrektur der Emissionsdaten von Inventaren auf Grundlage von Beobachtungen von Spurengas- und Aerosolkonzentrationen in der Atmosphäre beinhaltet das EURAD-IM ein vier-dimensionales variationelles (4D-Var) Datenassimilationssystem, das eine gleichzeitige Optimierung von Anfangskonzentrationen und speziesabhängigen Emissionskorrekturen ermöglicht. Um die Erkenntnisse über die Quellen von Luftverschmutzung zu verbessern, wird in dieser Arbeit ein neuer Ansatz entwickelt und in das Datenassimilierungssystem des EURAD-IM implementiert, um Emissionen individuell für Quellenkategorien wie Straßenverkehr, Industrie und Landwirtschaft zu korrigieren. Für die Unterscheidung zwischen den Emissionen verschiedener Quellenkategorien nutzt der neue Ansatz die räumliche Trennung von Emissionsquellen verschiedener Kategorien sowie deren charakteristische tageszeitliche Emissionsprofile und chemische Zusammensetzungen. Unter der Annahme einer feststehenden chemischen Zusammensetzung der Emissionen der Quellenkategorien innerhalb der Gitterzellen wird eine vollständige Korrelation zwischen den Emissionskorrekturen der verschiedenen chemischen Spezies eingeführt. Darüber hinaus ist ein anisotroper Diffusionsoperator implementiert, der entlang von Straßen die räumliche Korrelation zwischen den Straßenverkehrs-Emissionskorrekturen der Gitterzellen erhöht. Um die Fähigkeit der neuen Entwicklung, zwischen Emissionen verschiedener Sektoren zu unterscheiden, zu untersuchen, werden zwei verschiedene Arten von Simulationen durchgeführt. In identischen Zwillingsexperimenten auf Basis von synthetischen Beobachtungen werden Szenarien mit erhöhten industriellen und landwirtschaftlichen Emissionen bei gleichzeitiger Reduzierung der Straßenverkehrsemissionen simuliert. Das Datenassimilationssystem auf Basis des neuen Ansatzes ist in der Lage, die Emissionsänderungen in den Experimenten für weite Teile des Modellgebiets durch die ermittelten sektorspezifischen Emissionskorrekturen wiedergegeben. Zudem wird eine Studie durchgeführt, in dem die Emissionen innerhalb eines zweiwöchigen Zeitraums in Nordrhein-Westfalen anhand realer Beobachtungsdaten analysiert werden. Es wird gezeigt, dass in diesem Szenario eine Unterscheidung von Industrie- und Kraftwerksemissionen gegenüber Straßentransportemissionen durch die sektorspezifische Emission-

soptimierung möglich ist. Zudem können Änderungen landwirtschaftlicher Emissionen wegen ihres hohen NH_3-Anteils spezifiziert werden. Für alle beobachteten Spezies, das heißt O_3, NO_2, SO_2, PM_{10} und $PM_{2.5}$, ist die Übereinstimmung der simulierten mit den beobachteten Konzentrationen vergleichbar mit der einer Referenzsimulation unter Benutzung des aktuellen Datenassimilationssystems des EURAD-IM. Eine Verbesserung der Ergebnisse wird durch Einbeziehung zusätzlicher Beobachtungsdaten, insbesondere von CO und CO_2-Konzentrationen, erwartet.

Contents

List of Figures

List of Tables

Acronyms

PM_{10} particulate matter with a diameter smaller than 10 micrometres. 1, 52

$PM_{2.5}$ particulate matter with a diameter smaller than 2.5 micrometres. 1, 52

3D-Var three-dimensional variational data assimilation. ix, 8, 9, 11

4D-Var four-dimensional variational data assimilation. ix, 2–4, 8–11, 14, 18–20, 66, 76

CAMS Copernicus Atmosphere Monitoring Service. 27, 30

CDIAC Carbon Dioxide Information Analysis Center. 1

CEIP Centre on Emission Inventories and Projections. 30

CTM Chemistry Transport Models. 1, 10

EAC4 European Centre for Medium-Range Weather Forecasts Atmospheric Composition Reanalysis 4. 27

EDGAR Emissions Database for Global Atmospheric Research. 1

EEA European Environment Agency. 38

EEM EURAD emission module. 30

Eionet European Environment Information and Observation Network. 38

EMEP European Monitoring and Evaluation Programme. 38

EnKF Ensemble Kalman Filter. 3, 8, 10

EURAD European Air pollution Dispersion Model. 11, 27

EURAD-IM European Air pollution Dispersion-Inverse Model. ix, xiii, xix, 4–6, 11, 12, 14, 18–21, 27–30, 33, 34, 38, 44, 60, 66, 81, 83, 102

GEIA Global Emissions Initiative. 34

GNFR Gridding Nomenclature for Reporting. x, xi, xiii, xix, 30, 31, 34–37, 46–48, 58, 61–64, 67, 69, 77

GOME-2 Global Ozone Monitoring Experiment-2. 3

GRETA Gridding Emission Tool for ArcGIS. x, 30, 32

L-BFGS Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm. 17

MEGAN Model of Emissions of Gases and Aerosols from Nature. 37, 38

MOPITT Measurements of Pollution in the Troposphere. 2

NRW North Rhine-Westphalia. x, xi, xiii, xv, xvii, xviii, 30–33, 38, 40, 42, 43, 45–47, 53, 57, 58, 60, 61, 63, 64, 70, 77, 78, 85, 98, 99, 101, 103, 105

OMI Ozone Monitoring Instrument. 3

OSM Open Street Map. x, 22, 23

PBL planetary boundary layer. 16

WHO World Health Organisation. 1

WRF Weather Research and Forecasting Model. 28

ZSE Central System Emissions of Germany. 30

1. Introduction

Air pollution by particulate matter (PM), nitrogen oxides (NO_x), ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO) and other trace gases is harmful to human health. Several studies found evidence that these pollutants are associated with cardiovascular disease, respiratory disease and lung cancer (Chen and Hoek, 2020; Lee et al., 2020; Orellano et al., 2020; Huangfu and Atkinson, 2020; Zheng et al., 2021; Orellano et al., 2021). Following Fuller et al. (2022), in 2019, 4.5 million deaths were associated with outdoor air pollution. The World Health Organisation (WHO) published a new air quality guideline in 2021, with a recommended annual average concentration below 15 µg/m^3 for particulate matter with a diameter smaller than 10 micrometres (PM_{10}), 5 µg/m^3 for particulate matter with a diameter smaller than 2.5 micrometres ($PM_{2.5}$), and 10 µg/m^3 for nitrogen dioxide (NO_2). However, 90 % of the global population is exposed to average annual $PM_{2.5}$ concentrations above 10 µg/m^3, which is the limit recommended in the WHO guideline of 2006 (Shaddick et al., 2020).

Both natural and anthropogenic emission sources significantly drive air pollution. Besides biogenic emissions from plants, natural emissions can be caused by wildfires (Requia et al., 2021) or extreme events like volcanic eruptions (Reikard, 2019). Industrial factories, power plants, road traffic and agriculture are strong anthropogenic pollution sources. To understand, investigate and forecast the impact of emissions on air quality, atmospheric Chemistry Transport Models (CTM) are an important tool to simulate past, present and future scenarios. CTMs predict concentrations of trace gases and aerosols, using emission data, meteorological data, as well as models for atmospheric chemistry and aerosol formation.

A crucial input to atmospheric chemistry transport models and at the same time one of the most important sources of uncertainty in these models are the anthropogenic and biogenic emission data. These data, in particular the anthropogenic emissions, are in general collected and composed in emission inventories. Many emission inventories are based on annual anthropogenic emission data for different chemical species and polluter groups. Polluter groups denote categories for emission sources, for example industry, road traffic or agriculture. The annual national emission totals are provided by research institutes and national environmental agencies, e.g. the Umweltbundesamt (German Environment Agency) in Germany (Schneider et al., 2016). Following the studies of Solazzo et al. (2021) about the Emissions Database for Global Atmospheric Research (EDGAR) and of Andres et al. (2016) about the gridded uncertainty of emissions caused by fossil fuel combustion in the Carbon Dioxide Information Analysis Center (CDIAC) emission maps, emission uncertainties have two main sources: The

first is the appraisal of annual and national emission totals. This includes assumptions about the relationship between the activity level and emissions of pollution sources. For example, the NO_x emissions from road vehicles per litre of fuel consumption are subject to a large uncertainty. The second source of uncertainty is due to the temporal and spatial distribution of emissions. Emission inventories commonly rely on general distribution functions, which are based on statistical knowledge and therefore often deviate from the real emission situation. Hence, to improve the model predictions of CTMs, there is a need for reducing the uncertainty of their emission input.

Several methods have already been developed to improve emission data for trace gases and aerosols. In this thesis, a new approach is introduced that allows for a polluter group specific correction of inventory emissions using observations of pollutant concentrations. In comparison, many other approaches do not distinguish between different source categories when correcting inventory emissions. In the following, different methods for an improvement of emission estimates are discussed, including polluter group specific and non-polluter group specific approaches.

Methods of emission estimation can be classified into bottom-up and top-down approaches, as defined in e.g. Cheewaphongphan et al. (2019). Bottom-up approaches derive emissions from activity data using assumptions about the relationship between the activity level and emissions of pollution sources. In contrast, top-down approaches use observed concentrations of trace gases and aerosols for the estimation of emissions. Top-down approaches can be based on data assimilation methods. These methods combine model a priori knowledge with the information of observations, as explained by Kalnay (2002b). With data assimilation, the initial values of a simulation and/or model parameters can be optimised such that the agreement between model predictions and observations is improved. In many atmospheric chemistry transport models, data assimilation systems are implemented that are able to optimise the emissions of trace gases and aerosols. The advantage of emission optimisation by data assimilation methods over bottom-up approaches is that assumptions about emission sources described in the previous paragraph can be avoided.

Several atmospheric chemistry transport models use the four-dimensional variational data assimilation (4D-Var) technique for a correction of inventory emissions. For example, it is implemented in the European Air pollution Dispersion-Inverse Model (EURAD-IM) by Elbern et al. (2007). This was the first approach with a joint optimisation of emissions and initial concentrations for several trace gas and aerosol species. In a case study, a significant improvement in the prediction of ozone concentrations was achieved. Other 4D-Var data assimilation systems correct inventory emissions without performing an optimisation of initial concentrations. For example, Wang et al. (2021) corrected black carbon (BC) emissions in Northern China in 2016, resulting in 20 to 60 % lower emissions compared to the inventory. Jiang et al. (2015) assimilated CO concentrations from Measurements of Pollution in the Troposphere (MOPITT) satellite retrievals. The results suggest a reduction of CO emissions in the tropics

and the subtropics, but an increase in the extratropics. Qu et al. (2017) developed a hybrid model which combines the 4D-Var technique with mass balance methods to determine NO_x emissions using satellite observations. A 4D-Var data assimilation for the correction of NH_3 emissions in Europe in 2016 was performed by Cao et al. (2022). The resulting posterior emissions showed a stronger peak in springtime than the prior emission estimates, leading to an improvement in the seasonality of the predicted NH_3 concentrations.

Another data assimilation method that can be used for the correction of inventory emissions in atmospheric chemistry transport models is the Ensemble Kalman Filter (EnKF). Application examples are the assimilation of NO_2 satellite retrievals (Miyazaki et al., 2012, 2017) and CO ground observations (Jia et al., 2022) to correct NO_x and CO emissions.

In all models and studies mentioned above, the corrections for the inventory emissions do not distinguish between different polluter groups. However, such a distinction has the potential to increase our knowledge about the sources of air pollution. In recent years, several approaches deriving polluter group specific emissions from observations of trace gas and aerosol concentrations were developed and discussed.

In many studies, polluter group specific emissions are directly estimated from observations, mostly from tropospheric columns of trace gases measured by satellite instruments. For example, Lin et al. (2010) developed a method exploiting the diurnal cycle of NO_x emissions from traffic, industry, power plants and heating in China. For the polluter group specific emission estimations, NO_2 satellite retrievals from measurements by the Global Ozone Monitoring Experiment-2 (GOME-2) and Ozone Monitoring Instrument (OMI) (both with a polar orbit) are used, which have different passing times over China (~10 a.m. versus ~2 p.m.). Liu et al. (2016) determined power plant emissions in China considering plumes of power plants with large capacities. These approaches have the advantage of a low computing effort compared to data assimilation methods. On the other hand, some important aspects of the system are not taken into account, especially concerning the chemical reactions, the transport as well as wet and dry deposition.

The advantage of data assimilation systems is that they are based on atmospheric chemistry transport models that include all processes relevant for the evolution of pollutant concentrations. Yuan et al. (2022) developed an approach based on the EnKF technique in which inventory emissions of NO_x and CO from seven polluter groups in Beijing are corrected. It assumes a homogeneous spatial distribution of correction factors. However, as mentioned before, the spatial distribution of emissions is a significant source of uncertainty of inventories. There are also 4D-Var based methods developed for a polluter group specific correction of inventory emissions. Hooghiemstra et al. (2011) corrected global CO emissions with a 4D-Var system for a period of two years, distinguishing between anthropogenic, natural and biomass

burning emissions. Meirink et al. (2008) calculated CH_4 emission corrections for 11 different source categories in 2003. Both studies are limited to determining emission corrections for one chemical compound with a long lifetime. Moreover, they do not include information on the annual cycle of emissions, and the spatial resolution of the models is too coarse (far above $1\,^\circ \times 1\,^\circ$) to resolve local emission sources.

The 4D-Var-based approach developed in the framework of this thesis has the aim for correcting inventory emissions

- for individual polluter groups (denoted as sectors in the following),
- with a high spatial resolution of up to 1 km × 1 km,
- on the basis of NO_2, O_3, SO_2, CO, PM_{10} and $PM_{2.5}$ observations and
- taking (photo-)chemical reactions, aerosol dynamics, transport by advection and diffusion as well as wet and dry deposition into account.

The existing 4D-Var data assimilation system of the EURAD-IM fulfills all listed criteria except for the first. For this reason, the new development for sector specific emission corrections is implemented in the EURAD-IM. The data assimilation system is extended to split the emission corrections into the sectors included in the emission data. In the studies of this work, the Gridding Nomenclature for Reporting (GNFR) categorisation with 12 sectors is used (Granier et al., 2019). To correctly determine sector specific emission corrections, the data assimilation system needs to exploit the differences between the sectors. The sectors differ with regard to the spatial distribution, chemical composition and diurnal profile of emissions. In order to investigate the ability of the new development to distinguish between emissions from different sectors, identical twin experiments as well as a real-case study using ground observations are performed.

This work is structured as follows. In Chapter 2, the principles of data assimilation are explained and different techniques are discussed. Chapter 3 introduces the current EURAD-IM model, including the forward model and the 4D-Var-based emission optimisation. In Chapter 4, the approach of this work allowing for a sector specific optimisation of emissions is presented, emphasising the differences to the current data assimilation system of the EURAD-IM. Chapter 5 deals with the input data of the studies in this work. They include the initial and boundary conditions, the meteorological data, the emission data and the ground station observations. Because of their special importance for the sector specific emission optimisation, the observational data in North Rhine-Westphalia and the assumptions about the emissions are discussed in detail. Identical twin experiments including three different emission scenarios are the topic of Chapter 6. The performance of the new approach in this idealised case with synthetic observations is discussed on the basis of the determined sector specific emission corrections. Chapter 7 presents an experiment in which real ground-based

observations for a two-week period of in March 2016 are assimilated. The emission corrections determined by the sector specific emission optimisation as well as the agreement between simulated and observed pollutant concentrations are investigated in detail. Furthermore, the results are compared to those of a reference simulation using the current data assimilation system of the EURAD-IM. At the end of the thesis (Chapter 8), conclusions concerning the potential and limits of the new approach are drawn. Moreover, further simulations, studies and improvements are suggested.

2. Data assimilation

Data assimilation methods have the aim of improving the agreement between simulated and observed quantities of a system (Kalnay, 2002b). In order to improve the agreement, the initial values of a simulation, in the following referred to as initial state, and model parameters can be optimised. In atmospheric chemistry transport models (CTM), the emission rates of trace gases and aerosols are parameters that have a crucial impact on the simulated concentrations. In addition, the initial state plays an important role, which denotes the concentrations $\mathbf{x_0}$ of pollutants at the beginning of a simulation. Both the initial state and the emission rates are subject to a high uncertainty. For this reason, a joint optimisation of the initial state and emission rates is implemented in the EURAD-IM to improve the simulated pollutant concentrations.

A simple data assimilation approach used in early weather forecast models is the interpolation of observations to the grid cells to estimate the initial state. However, it is not suitable for models with high spatial resolution, since the degree of freedom is orders of magnitude higher than the number of available observations. The degree of freedom denotes the number of grid cells multiplied by the number of prognostic variables. In addition, an optimisation of pollutant emissions is not possible with this approach.

Advanced data assimilation techniques

More advanced data assimilation methods combine observational data with the information of model forecasts. They are more appropriate for highly under-determined problems than the interpolation of observations. On the example of an initial state optimisation, the general procedure of data assimilation can be described by the following steps (Kalnay, 2002b):

1. First, a forecast is performed on the basis of a first guess of the initial state $\mathbf{x}_{0,B}$. This forecast is denoted as the background forecast and $\mathbf{x}_{0,B}$ is denoted as the initial state of the background. In the case of atmospheric chemistry transport models, $\mathbf{x}_{0,B}$ are the initial concentrations of the background. For this, e.g. the final concentrations of a previous simulation can be used.

2. The observational increments

$$\mathbf{y}^O - \mathcal{H}(\mathbf{x}_{0,B}) \tag{2.1}$$

are calculated, indicating the discrepancy between model forecast and observations. $\mathbf{y}^O$ is the observation vector containing all selected observations and is also denoted as the observational state. In the case of chemistry transport models, measurements of pollutant concentrations, e.g. from ground stations or satellite instruments, are used as observations. $\mathcal{H}$ is the observation operator used to calculate the model equivalents $\mathcal{H}(\mathbf{x})$ of the observations, thus allowing for a comparison between model results and observations.

3. The initial state is optimised such that the observational increments are reduced. The resulting initial state is denoted as the initial state $\mathbf{x}_{0,A}$ of the analysis. The optimisation can be described by a weighting operator $\mathbf{W}$:

$$\mathbf{x}_{0,A} = \mathbf{x}_{0,B} + \mathbf{W}[\mathbf{y}^O - \mathcal{H}(\mathbf{x}_{0,B})] \tag{2.2}$$

4. On the basis of $\mathbf{x}_{0,A}$, a forecast is performed, which is denoted as the analysis forecast.

Several techniques were developed for the optimisation of initial states and/or model parameters, which is partially discussed in the following.

Ensemble-based data assimilation

Some data assimilation methods are ensemble-based. An ensemble consists of a certain number of forecasts where the initial state $\mathbf{x}_0$ (and/or model parameters, e.g. emissions) of a system is perturbed. Ideally, the ensemble spread induced by the perturbations reflects the forecast uncertainty. Ensemble-based data assimilation techniques such as the Ensemble Kalman Filter (EnKF) (Evensen, 1994, 2009; Gillijns et al., 2006) combine the information of observations with forecast ensembles to estimate the optimal initial state and/or model parameters. Because of the relatively simple implementation, ensemble-based data assimilation methods are commonly used in meteorological models. They contribute significantly to the reduction of the uncertainties of weather forecasts (Kalnay et al., 2006; Slingo and Palmer, 2011). The EnKF is also used in several atmospheric chemistry transport models for the correction of inventory emissions (Jia et al., 2022; Miyazaki et al., 2012, 2017).

Variational data assimilation

In contrast, variational data assimilation approaches estimate the analysis state of a system by minimising a cost function J (Rabier and Liu, 2003). The cost function is derived from the Bayes' Theorem on conditional probability:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x}) \cdot P(\mathbf{x})}{P(\mathbf{y})} , \tag{2.3}$$

where $P(\mathbf{x})$ is the prior probability distribution for the state of the considered system and $P(\mathbf{y})$ the probability distribution for the observational state. $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{x}|\mathbf{y})$ denote conditional probabilities. Gaussian distributions are assumed for $P(\mathbf{x})$ and $P(\mathbf{y}|\mathbf{x})$. Variational data assimilation seeks to find the state $\mathbf{x}$ with the highest probability density of $P(\mathbf{x}|\mathbf{y})$.

3D-Var

The three-dimensional variational data assimilation (3D-Var) has the aim of finding the best fit between the background state and observations at a certain time. The cost function

$$\begin{aligned} J(\mathbf{x}) = &\frac{1}{2} \cdot (\mathbf{x} - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_B) \\ &+ \frac{1}{2} \cdot (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})) \end{aligned} \tag{2.4}$$

derived from the Bayes' Theorem is minimised in 3D-Var (Kalnay, 2002a). $\mathbf{x}$ denotes the state, i.e. all simulated variables which are arguments of the cost function. $\mathbf{x}_B$ denotes the background state, i.e. the first guess for the state $\mathbf{x}$. $\mathbf{B}$ denotes the background error covariance matrix (Section 3.2) and $\mathbf{R}$ the observation error covariance matrix (Section 5.4). The term

$$J_B = \frac{1}{2} \cdot (\mathbf{x} - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_B) \tag{2.5}$$

is also referred to as the background costs J_B and the term

$$J_O = \frac{1}{2} \cdot (\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})) \tag{2.6}$$

is called the observational costs. 3D-Var data assimilation is used in meteorological forecast models (Andersson et al., 1998), ocean models (Li et al., 2008) and also in chemistry transport models, e.g. EURAD-IM, EMEP, MOCAGE and SILAM (Collin, 2020). It has the advantage of a low computational demand compared to EnKF and 4D-Var (described below). A major disadvantage of 3D-Var is that it does not take the dynamics of a system into account, since only the state at a certain time is considered. Furthermore, no parameter optimisation is possible. Additional information about the 3D-Var data assimilation method can be found in Kalnay (2002a) and Fisher (2002).

4D-Var

The mentioned disadvantages of 3D-Var can be solved by including the temporal dimension, i.e. optimising the agreement between model and observations for a whole time window, the so-called assimilation window. The observational costs J_O are summed over all time steps in the assimilation window. This is the principle of the

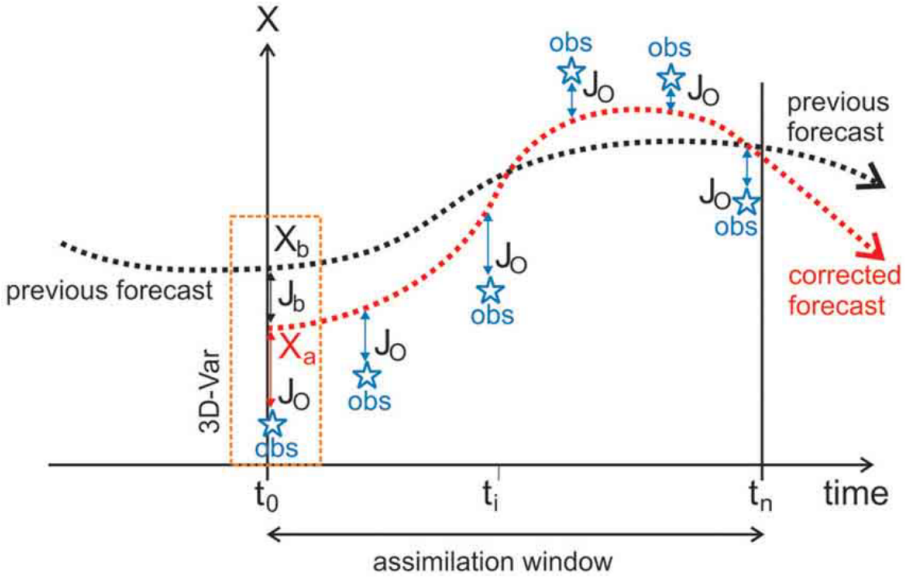

Figure 2.1.: Illustration of the principle of 4D-Var. Observations are depicted as blue stars, the background forecast as black dotted line and the analysis forecast as red dotted line. The differences between observations and model forecast contribute to the observational costs (Eq. 2.6) and are indicated by J_O. In 4D-Var, the initial value (value at time t_0) is optimised such that the observational increments (Eq. 2.1) of the whole assimilation window are minimised. Thus, in comparison with 3D-Var, 4D-Var takes dynamics of the system into account. Source: Lahoz and Schneider (2014).

four-dimensional variational data assimilation method (4D-Var) illustrated in Fig. 2.1. Both meteorological and atmospheric chemistry transport models use this method. 4D-Var data assimilation systems are also implemented in non-atmospheric models, for example in ocean models (Lee et al., 2018) and solar models (Sacha Brun et al., 2019). Initial state optimisation with 4D-Var is described by the following cost function (Bouttier and Courtier, 1999):

$$\begin{aligned} J(\mathbf{x}_0) = &\frac{1}{2} \cdot (\mathbf{x}_0 - \mathbf{x}_{0,B})^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_{0,B}) \\ &+ \frac{1}{2} \cdot \sum_{i=0}^{N} (\mathbf{y}(t_i) - \mathbf{H}(t_i)\mathbf{x}(t_i))^T \mathbf{R}^{-1} (\mathbf{y}(t_i) - \mathbf{H}(t_i)\mathbf{x}(t_i)) \ . \end{aligned} \tag{2.7}$$

$\mathbf{H}$ denotes the tangent linear observation operator. $\mathbf{x}_0$ is the initial state and $\mathbf{x}_{0,B}$ the initial state of the background. Compared with the 3D-Var cost function, the major difference is the summation of the observational increments over all time steps of the assimilation window. A typical assimilation window in weather prediction models is 6 h (Kalnay, 2002b). In the studies of this work (Chapters 6 and 7), an interval of 24 h is selected to determine daily emission corrections.

4D-Var data assimilation systems iteratively optimise the initial state and model parameters. An iteration consists of a forward run of the CTM, an adjoint run and the minimisation procedure. The adjoint or "backward" run is of particular importance for the 4D-Var data assimilation. In this step, the gradient ∇J of the cost function with respect to the initial values and the model parameters, is calculated. The gradient is necessary to determine the initial state and model parameters of the next iteration in the minimisation procedure. The adjoint model contains the equations of the CTM in adjoint form.

On the one hand, the effort required for the adjoint model is a disadvantage of the 4D-Var data assimilation technique in comparison with the EnKF. On the other hand, there is also an important advantage related to the adjoint model: In the 4D-Var optimisation of the initial state and model parameters, relevant processes such as chemical reactions and transport of pollutants have a higher influence than in the EnKF. For this reason, the new approach towards sector specific emission optimisation presented in this thesis is based on the 4D-Var data assimilation method.

3. The atmospheric chemistry transport model EURAD-IM

The European Air pollution Dispersion-Inverse Model (EURAD-IM) (Elbern et al., 1997, 2007) as a further development of the European Air pollution Dispersion Model (EURAD) (Hass et al., 1995) is a regional scale Eulerian chemistry transport model, which predicts concentrations of trace gases and aerosols in the troposphere. Furthermore, 3D-Var and 4D-Var data assimilation systems are implemented. The 4D-Var data assimilation system allows for a joint optimisation of emissions and initial concentrations. The EURAD-IM contributes to the daily air quality forecasts for Europe (CAMS2_40) of the Copernicus Atmosphere Monitoring Service (Collin, 2020). In this chapter, the forward model (Section 3.1) and the 4D-Var data assimilation system (Section 3.2) of the current EURAD-IM are described.

3.1. Forward model

Fig. 3.1 shows a simplified flow chart of the EURAD-IM forward model. The model input contains information about the meteorological situation, terrestrial data, anthropogenic and biogenic emission data as well as initial and boundary values of trace gas and aerosol concentrations. The input components are topic of Chapter 5. These input fields are used to calculate the transformation and dispersion of trace gases and aerosols. For this, advection, diffusion, (photo-)chemical reactions, aerosol dynamics and dry and wet deposition are taken into account. The output fields include predicted concentrations of aerosol and trace gas species.

The initial state $\mathbf{x}(t = 0)$ is integrated forward in discrete time steps Δt. At each time step, the following operator sequence is applied:

$$\mathbf{x}(t + \Delta t) = T_h D_h T_v D_v C D_v T_v D_h T_h \, \mathbf{x}(t), \tag{3.1}$$

where T denotes the vertical (v) and horizontal (h) transport operator, and D the operator for vertical and horizontal diffusion. C represents all state transformations due to the (photo-)chemical reactions, aerosol dynamics, emissions as well as dry and wet deposition processes (Elbern et al., 2007).

Applying the Lambert conformal conic projection, the model domain is horizontally subdivided into squares of equal size. In the studies presented in this thesis, the

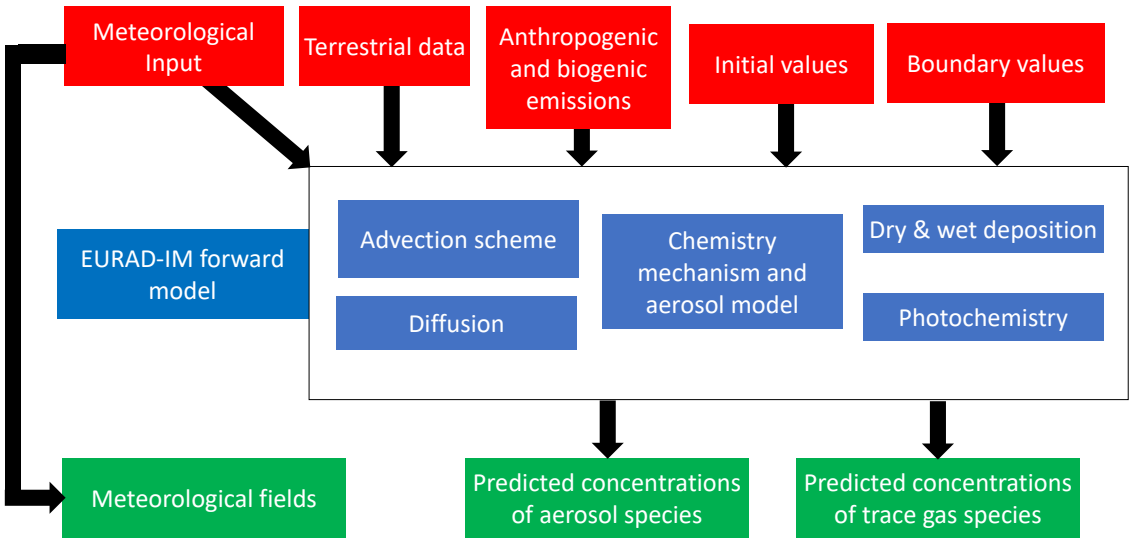

Figure 3.1.: Simplified flow chart of the EURAD-IM forward model. The input fields (red), processes represented in the model (blue) and output fields (green) are shown.

horizontal resolution is 1 km × 1 km. The 30 vertical layers are defined on σ-coordinates, where

$$\sigma = \frac{p - p_T}{p_S - p_T}. \tag{3.2}$$

Here, p is the pressure of the respective model layer, $p_T = 100\,\text{hPa}$ is the pressure at the top of the domain and p_S describes the surface pressure.

3.2. 4D-Var data assimilation system

With the 4D-Var data assimilation system developed by Elbern et al. (2007), emissions and initial concentrations can be optimised. The emissions are optimised for each emitted chemical species with the same spatial resolution as the model grid. A data assimilation cycle starts with background emissions ϵ_B in the first iteration. In the following iterations, correction factors e for these emissions are determined. The emission correction factors are defined as the ratio of the corrected emissions ϵ and

the background emissions:

$$e(i,j,k,l) = \frac{\epsilon(i,j,k,l)}{\epsilon_B(i,j,k,l)} \ , \tag{3.3}$$

where i and j denote the horizontal indices of the grid cells, k the layer of the grid cell and l the index of the species. Note that e is not a function of time. This means that the emission correction factors are constant in time for the entire assimilation window.

Cost function and gradient

For a joint optimisation of emission correction factors and initial concentrations, the following cost function has to be minimised:

$$\begin{aligned} J(\mathbf{x}_0, \mathbf{u}) = & \frac{1}{2} \cdot (\mathbf{x}_0 - \mathbf{x}_{0,B})^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_{0,B}) \\ & + \frac{1}{2} \cdot \mathbf{u}^T \mathbf{K}^{-1} \mathbf{u} \\ & + \frac{1}{2} \cdot \sum_{i=0}^{N} (\, \mathbf{y}(t_i) - \mathbf{H}(t_i)\, \mathbf{x}(t_i)\,)^T \mathbf{R}^{-1} (\, \mathbf{y}(t_i) - \mathbf{H}(t_i)\, \mathbf{x}(t_i)) \ , \end{aligned} \tag{3.4}$$

where $\mathbf{u}$ denotes the vector of the logarithm of the emission correction factors and $\mathbf{K}$ the emission error covariance matrix. Besides the background costs and the observational costs, the cost function includes emission correction costs

$$J_K = \frac{1}{2} \cdot \mathbf{u}^T \mathbf{K}^{-1} \mathbf{u} \ . \tag{3.5}$$

Since $\mathbf{K}$ and $\mathbf{B}$ are high-dimensional, an explicit representation and calculation of their inverse is not feasible. Therefore, an incremental formulation of the cost function, following Weaver and Courtier (2001), is used:

$$J(\mathbf{v}, \mathbf{w}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2} \cdot \sum_{i=0}^{N} (\, \mathbf{d}(t_i) - \mathbf{H}(t_i)\, \delta\mathbf{x}(t_i)\,)^T \mathbf{R}^{-1} (\, \mathbf{d}(t_i) - \mathbf{H}(t_i)\, \delta\mathbf{x}(t_i)) \ , \tag{3.6}$$

where $\delta\mathbf{x}(t_i) = \mathbf{x}(t_i) - \mathbf{x}_B(t_i)$ and $\mathbf{d}$ is the so-called innovation vector:

$$\mathbf{d}\,(t_i) = \mathbf{y}^O(t_i) - \mathcal{H}(t_i)\, \mathbf{x}_B(t_i), \tag{3.7}$$

and $\mathbf{v}$ and $\mathbf{w}$ are defined by

$$\mathbf{v} := \mathbf{B}^{-1/2} (\mathbf{x}_0 - \mathbf{x}_{0,B}),\ \mathbf{w} := \mathbf{K}^{-1/2}\, \delta\mathbf{u}, \delta\mathbf{u} := \ln(\mathbf{e}) \ , \tag{3.8}$$

where $\mathbf{e}$ are the emission correction factors. For the calculation of the initial state and emission correction factors, the gradient of the cost function has to be determined. Its

formula is given by

$$\nabla_{(\mathbf{v},\mathbf{w})^T} J = \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} - \begin{pmatrix} \mathbf{B}^{1/2} & 0 \\ 0 & \mathbf{K}^{1/2} \end{pmatrix} \times \sum_{i=1}^{N} \mathbf{M}^T(t_0, t_i)\, \mathbf{H}^T\, \mathbf{R}^{-1} \left(\mathbf{d}(t_i) - \mathbf{H}\delta\mathbf{x}(t_i)\right) , \quad (3.9)$$

where $\mathbf{H}^T$ is the adjoint of the tangent linear observation operator and $\mathbf{M}^T$ is the adjoint of the tangent linear model operator (Elbern et al., 2007). $\mathbf{M}^T(t_0, t_i)$ propagates the observational increments (Eq. 2.1) backward from time t_i to t_0. For this, an adjoint model run is performed.

Data assimilation cycle

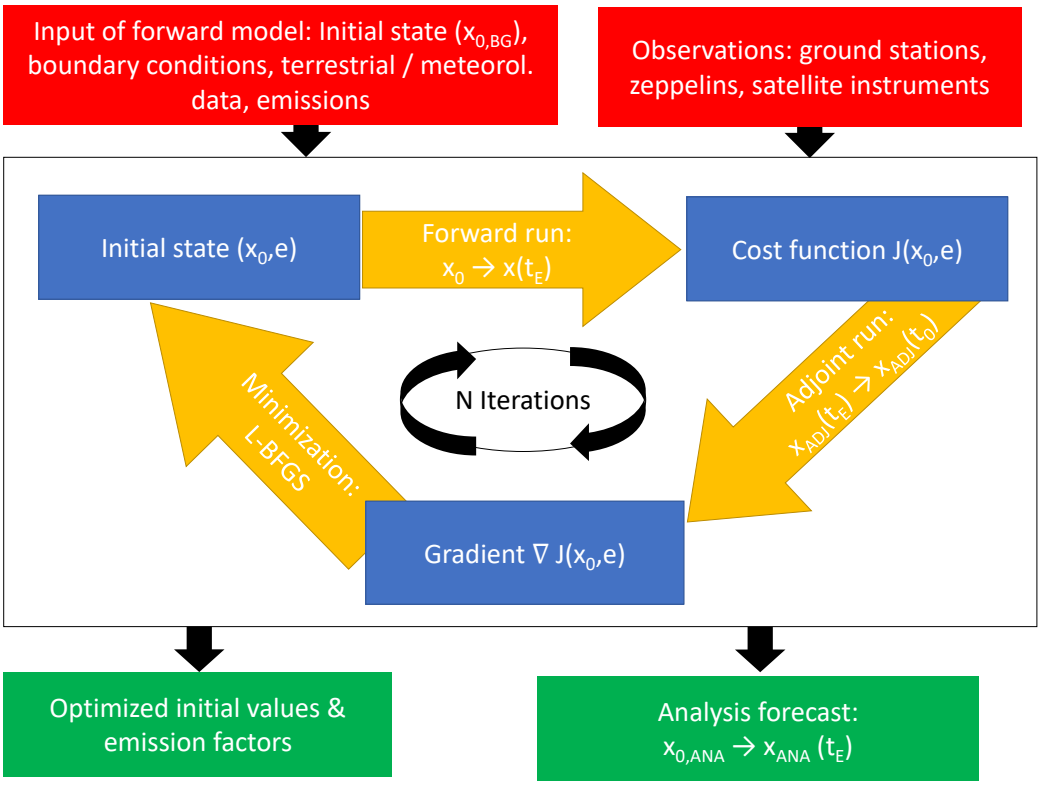

Figure 3.2.: Flowchart of the 4D-Var data assimilation cycle implemented in the EURAD-IM, including the input data (red), process steps (yellow), intermediate results (blue) and output (green). The initial state is denoted as $\mathbf{x}_0$, the state at the end of the assimilation window as $\mathbf{x}(t_E)$, the adjoint state as $\mathbf{x}_{ADJ}$, the state of the analysis as $\mathbf{x}_{ANA}$, the emission correction factors as $\mathbf{e}$, the cost function as J and the assimilation interval as $[t_0,t_E]$.

Fig. 3.2 shows a schematic representation of the 4D-Var assimilation cycle, which consists of several iterations with a forward run, adjoint run and minimisation procedure. An iteration of the assimilation cycle starts with the initial state and the

emission correction factors determined in the previous iteration. In the case of the first iteration, the background initial state $\mathbf{x}_{0,B}$ and background emission correction factors $\mathbf{e}_B$ are used. In the forward run, the model state is integrated from t_0 to t_E, containing all processes described in Chapter 3. Furthermore, the cost function is calculated, following Eq. 3.6. The gradient of the cost function with respect to the initial state and the emission correction factors (Eq. 3.3) is computed in the adjoint model run. The adjoint model operator $\mathbf{M}^T$ is applied, which includes the equations of the forward model in adjoint form. The initial state and emission correction factors for the next iteration are calculated in the minimisation procedure. After performing a certain number of iterations, the iteration with the lowest cost function value (Eq. 3.4) is selected to determine the optimised initial values and emission correction factors. This optimised state is used for the analysis forecast.

Minimisation procedure

In the following, the minimisation procedure is described in detail. First, the gradient with respect to the transformed variables $\mathbf{v}$ and $\mathbf{w}$ (Eq. 3.8) is determined. In order to calculate $\nabla_{(\mathbf{v},\mathbf{w})^T} J$ (Eq. 3.9), $\mathbf{B}^{1/2}$ and $\mathbf{K}^{1/2}$ (described in the following) have to be applied to the adjoint initial state and emission correction factors. Both the background error covariance matrix $\mathbf{B}$ and the emission error covariance matrix $\mathbf{K}$ are factorised in the following way:

$$\mathbf{K} = \mathbf{\Sigma}\,\mathbf{C}^{1/2}\,\mathbf{C}^{T/2}\,\mathbf{\Sigma} \tag{3.10}$$

with the diagonal error matrix $\mathbf{\Sigma}$ and the correlation matrix $\mathbf{C}$ (Elbern et al., 2007). In the emission error covariance matrix $\mathbf{K}$, the entries of $\mathbf{\Sigma}$ are species-dependent. For the background error covariance matrix $\mathbf{B}$, a minimum absolute error for each species and a relative error increasing with height are defined in Elbern et al. (2007).

The correlation matrix $\mathbf{C}$ is defined by

$$\mathbf{C}^{1/2} = \mathbf{\Lambda}\mathbf{L}^{1/2}\mathbf{W}^{-1/2}\mathbf{\Omega}\ , \tag{3.11}$$

where $\mathbf{L}$ is the diffusion operator, $\mathbf{W}$ the geometry matrix, $\mathbf{\Lambda}$ the normalisation matrix and $\mathbf{\Omega}$ the emission error correlation matrix. Note that $\mathbf{\Omega}$ is contained in $\mathbf{K}$, but not in $\mathbf{B}$. $\mathbf{L}$ produces a spatial correlation between the emission correction factors and initial value corrections, respectively, based on the discrete solution of the diffusion equation. Assuming an isotropic diffusion, Schwinger (2006) determined the following formula:

$$\mathbf{L}^{1/2} = \{\{\mathbf{I} + \kappa_v\,\Delta t\,\mathbf{D}_v\}\{\mathbf{I} + \kappa_h\,\Delta t\,\mathbf{D}_h\}\}^{M/2}\ . \tag{3.12}$$

$\mathbf{I}$ denotes the unit matrix, Δt the diffusion time step, M the number of diffusion steps, $\kappa_{h/v}$ the horizontal/vertical diffusion coefficient and $D_{h/v}$ the discretised Laplacian. M and Δt are calculated such that the stability criteria of the diffusion schemes are fulfilled. The diffusion coefficients are derived from the following formula (Elbern

et al., 2007):

$$L_{h/v} = \sqrt{2\kappa_{h/v} t_{h/v}} \ . \tag{3.13}$$

$L_{h/v}$ denote the horizontal and vertical diffusion lengths. In all considered model versions, L_h is 2.5 km for the surface, 10 km for the planetary boundary layer (PBL) and 20 km for the model top. L_v are equal to the Eddy diffusion coefficients. Note that $\mathbf{K}$ only contains horizontal diffusion in the first layer. The different heights of the grid cells due to the σ-coordinates have to be taken into account in the correlation matrix $\mathbf{C}$. For this, the metric $\mathbf{W}$ is introduced. This is a diagonal matrix which contains the vertical grid elements Δz of the model grid. The normalisation matrix $\mathbf{\Lambda}$ is a diagonal matrix with

$$\lambda_l = \frac{1}{\sqrt{t_l}}, \ t_l = \mathbf{e}_l^T \mathbf{L} \mathbf{W}^{-1} \mathbf{e}_l \tag{3.14}$$

as diagonal entries, where $\mathbf{e}_l$ denote the unit vectors. $\mathbf{\Lambda}$ normalises $\mathbf{B}$ and $\mathbf{K}$ in such a way that their effective standard deviations are in accordance with the errors specified in $\mathbf{\Sigma}$ (Weaver and Courtier, 2001).

		SO2	SO4	NO2	NO	ALD	HCHO	NH3	HC3	HC5	HC8	ETH	CO	ETE	OLT	OLI	TOL	XYL	KET	ISO	DIEN	ORA2	LIM	GLY	API
Sulfur dioxide	SO2	100	40	12	12	0.4	0.4	0.4	0.4	0.4	0.4	0.4	9	0.4	0.4	0.4	0.4	0.4	0.4	0	0.4	0.4	0	0.4	0
Sulfate	SO4		100	12	12	0.4	0.4	0.4	0.4	0.4	0.4	0.4	9	0.4	0.4	0.4	0.4	0.4	0.4	0	0.4	0.4	0	0.4	0
Nitrogen dioxide	NO2			100	26	8	8	7	8	8	8	8	12	8	8	8	7	7	8	0.4	8	5	0.4	8	0.4
Nitrogen oxide	NO				100	8	8	7	8	8	8	8	12	8	8	8	7	7	8	0.4	8	5	0.4	8	0.4
Higher aldehydes	ALD					100	26	5	6	6	6	6	5	10	6	6	5	5	26	0.3	6	5	0.3	26	0.3
Formaldehyde	HCHO						100	5	6	6	6	6	5	10	6	6	5	5	26	0.3	6	5	0.3	26	0.3
Ammonia	NH3							100	5	5	5	5	5	5	5	5	5	5	5	0.3	5	5	0.3	5	0.3
Hydrocarbons I *	HC3								100	23	23	23	5	18	15	15	6	6	6	0.3	18	8	0.3	6	0.3
Hydrocarbons II **	HC5									100	23	23	5	18	15	15	6	6	6	0.8	18	8	0.8	6	0.8
Hydrocarbons III ***	HC8										100	23	5	18	15	15	6	6	6	0.8	18	8	0.8	6	0.8
Ethane	ETH											100	5	18	15	15	6	6	6	0.8	18	8	0.8	6	0.8
Carbon monoxide	CO												100	6	5	5	5	5	5	0.3	5	5	0.3	5	0.3
Ethene	ETE													100	23	23	6	6	10	1	18	5	1	10	1
Terminal alkenes	OLT														100	27	5	5	6	15	5	10	15	6	15
Internal alkenes	OLI															100	5	5	6	15	5	10	15	6	15
Aromatics I ****	TOL																100	24	5	0.3	5	5	0.3	5	0.3
Aromatics II *****	XYL																	100	5	0.3	5	5	0.3	5	0.3
Ketones	KET																		100	0.3	6	5	0.3	26	0.3
Isoprene	ISO																			100	0.8	18	40	0.4	40
Anthropogenic dienes	DIEN																				100	5	15	6	15
Acetic acid / higher acids	ORA2																					100	15	5	15
Cyclic diene-terpenes	LIM																						100	0.3	40
Glyoxal	GLY																							100	0.3
α-pinene	API																								100

>10% >1% < 1%

* Alkanes, alcohols, esters & alkynes with OH rate constant less than 3.4 x 10^{-12} cm^3 s^{-1}
** Alkanes, alcohols, esters & alkynes with OH rate constant between 3.4 x 10^{-12} and 6.8 x 10^{-12} cm^3 s^{-1}
*** Alkanes, alcohols, esters & alkynes with OH rate constant higher than 6.8 x 10^{-12} cm^3 s^{-1}
**** Toluene & less reactive aromatics
***** Xylene & more reactive aromatics

Figure 3.3.: Upper triangle of the emission error correlation matrix $\mathbf{\Omega}$ for the gas phase species used in the current EURAD-IM. For each combination of emitted species, a fixed correlation is defined. High correlations are assumed for combinations of species with a similar origin or if they are chemically related to each other. Source: Paschalidi (2015)

.

The emission error correlation matrix $\mathbf{\Omega}$ introduces correlations between the emission correction factors of the chemical species. For each combination of species, a fixed correlation between their emission correction factors is defined. The emission error correlation matrix of the current EURAD-IM model was defined by Paschalidi (2015). The upper triangle of the matrix for the gas phase species is shown in Fig. 3.3. It follows the assumption that there is a relatively high correlation between species, if they have a similar origin or if they are chemically related to each other. This applies, for example, to combinations of NO, NO_2 and SO_2 with correlations above 10 %. If this is not the case, a low correlation is defined for a combination of species. For example, no correlation is assumed between emissions of NO and isoprene, since no biogenic source that emits isoprene also emits NO.

The state vector $(\mathbf{v}, \mathbf{w})_k$ of the current iteration and the gradient $\nabla_{(\mathbf{v},\mathbf{w})^T} J$ are input to the minimisation. The cost function is minimised using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) (Liu and Nocedal, 1989). The L-BFGS calculates the new state vector $((\mathbf{v}, \mathbf{w})_{k+1})$. Since the cost function is minimised with respect to the transformed variables, a back-transformation of $\mathbf{v}$ and $\mathbf{w}$ to $\mathbf{x}_0$ and $\mathbf{u}$ is necessary after the minimisation:

$$\mathbf{x}_0 = \mathbf{B}^{1/2}\,\mathbf{v} + \mathbf{x}_{0,B},\ \mathbf{u} = \mathbf{K}^{1/2}\,\mathbf{w}\ . \tag{3.15}$$

$\mathbf{x}_0$ and $\mathbf{e} = \exp(\mathbf{u})$ are the initial values and emission correction factors of the next iteration.

4. Theory on sector specific optimisation of emissions

In this chapter, the new approach for a sector specific optimisation of emissions developed in the framework of this thesis is presented. Sectors denote source categories such as road traffic, industry or agriculture. There are different categorisations for emission sources, e.g. the Gridding Nomenclature for Reporting (GNFR) used in the studies of this work. The approach is also applicable to other categorisations. To allow for a sector specific optimisation of emissions, the 4D-Var data assimilation system of the EURAD-IM (Elbern et al., 2007) is extended. In the following sections, the theoretical basis of this method is introduced (Section 4.1) and two additional modifications implemented in the data assimilation system are described (Sections 4.2 and 4.3).

4.1. Theoretical basis

To achieve a sector specific optimisation of emissions, the vector of emission correction factors $\mathbf{e}$ is extended so that individual correction factors for different emission sectors are included. $\mathbf{e}$ consists of N_s subvectors $\mathbf{e}_s$, each representing the emission correction factors of a sector s, where N_s is the number of sectors considered. The dimension of each vector $\mathbf{e}_s$ is equal to that of the total vector of emission correction factors $\mathbf{e}$ defined in Elbern et al. (2007): number of grid cells × number of emitted species.

The cost function J of the current 4D-Var data assimilation system in the EURAD-IM (Eq. 3.4) is extended to

$$\begin{aligned} J(\mathbf{x}_0, \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{N_s}) = & \frac{1}{2} \cdot (\mathbf{x}_0 - \mathbf{x}_{0,B})^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_{0,B}) \\ & + \frac{1}{2} \cdot \sum_{s=1}^{N_s} \mathbf{u}_s^T \mathbf{K}_s^{-1} \mathbf{u}_s \\ & + \frac{1}{2} \cdot \sum_{i=0}^{N} (\mathbf{y}(t_i) - \mathbf{H}(t_i)\mathbf{x}(t_i))^T \mathbf{R}^{-1} (\mathbf{y}(t_i) - \mathbf{H}(t_i)\mathbf{x}(t_i)) \end{aligned} \tag{4.1}$$

where $\mathbf{u}_s$ is the vector containing the logarithmic emission correction factors of emission sector s and $\mathbf{K}_s$ is the emission background error covariance matrix for sector s. The cost function J is extended such that it is a function of the emission correction factors of individual emission sectors (Eq. 4.1). This enables the 4D-Var data assimilation

system to correct the emissions of the considered sectors separately.

In the forward model, the total emissions ϵ_{tot} in time step t_i are calculated as the sum of the corrected emissions for each sector. For this, the sector specific emission correction factors are used:

$$\epsilon_{tot}(i,j,k,l,t_i) = \sum_{s=1}^{N_s} e(i,j,k,l,s) \cdot \epsilon_B(i,j,k,l,s,t_i) \ , \tag{4.2}$$

where i and j denote the horizontal indices of the grid cell, k the vertical index of the grid cell, l the index of the emitted species, s the index of the emission sector, and ϵ_B the background emissions.

As in the current EURAD-IM without sector specification (Section 3.2), the adjoint model contains the equations of the forward model in adjoint form. In the adjoint model run, the vectors of the adjoint concentrations, adjoint emission rates (**ade**) and adjoint emission correction factors (**adef**) are calculated. In contrast to the EURAD-IM without sector specification, the adjoint emission correction factors are sector specific in the new approach. Accordingly, the vector **adef** contains N_s subvectors, each related to an emission sector s. The adjoint emission rates are computed in the adjoint chemistry and aerosol modules. These are needed to determine the adjoint emission correction factors. In each adjoint time step t_i, an adjoint emission correction factor ($adef(i,j,k,l,s,t_i)$) is calculated for each grid cell, emitted species and sector using the following formula:

$$adef(i,j,k,l,s,t_i) = adef(i,j,k,l,s,t_{i+1}) + ade(i,j,k,l,t_i) \cdot \epsilon_B(i,j,k,l,s,t_i) \ . \tag{4.3}$$

The vector **adef** is integrated backwards in time from the final time t_E to the start t_0 of the assimilation window. $\mathbf{adef}(t_0)$ corresponds to the gradient $\nabla J_{\mathbf{u}}$ of the cost function with respect to the emission correction factors. Thus, $\nabla J_{\mathbf{u}(i,j,k,l,s)}$ reflects the sensitivity of the observational increments (Chapter 2) to the sector specific emission correction factor $e(i,j,k,l,s)$. For this reason, the sector specific gradients are investigated in the studies (Chapters 6 and 7).

The formula for the gradient of the cost function ($\nabla_{(\mathbf{v},\mathbf{w})^T} J$) with respect to the modified variables $\mathbf{v}$ and $\mathbf{w}$ is equal to that of the current 4D-Var data assimilation system of the EURAD-IM (Eq. 3.9). However, the augmentation of the state vector by sector specific emission correction factors requires changes in the transformed vector $\mathbf{w}$. Here, it is defined by the increments $\delta\mathbf{u_s}$ of the sector specific emission correction factors:

$$\mathbf{w} := \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \dots \\ \mathbf{w}_{N_s} \end{pmatrix}, \ \mathbf{w}_s := \mathbf{K}_s^{-1/2} \delta\mathbf{u}_s, \tag{4.4}$$

with sector specific emission error covariance matrices $\mathbf{K}_s$. The full emission error covariance matrix $\mathbf{K}$ is block diagonal with the matrices $\mathbf{K}_s$ as block elements. Without applying additional modifications (Sections 4.2 and 4.3), all matrices $\mathbf{K}_s$ are defined in the same way as $\mathbf{K}$ in the current 4D-Var data assimilation system of the EURAD-IM (Eq. 3.10).

4.2. Modification of the emission error correlation matrix

The characteristic chemical composition of emission sectors is crucial for the distinction between their emissions through the sector specific emission optimisation. According to the calculation of the adjoint emission correction factors (Eq. 4.3), the sector specific emission gradient for a species strongly depends on the emission strength of the species in the sector. This way, the data assimilation system uses the different chemical composition of the sectors to distinguish between their emissions.

Regarding real emissions, there is a high correlation between the species within the sectors. The emission correction factors determined in the current EURAD-IM 4D-Var data assimilation system are species-dependent. In the case of the sector specific approach, species-dependent emission correction factors would lead to a high variation of the emission correction factors within a sector, contradicting the assumption of a high correlation between species. Further, the degree of freedom of the optimisation algorithm would increase drastically, which might hamper finding the optimum solution.

The modification described in this subsection, which is developed for the approach for sector specific emission optimisation, is based on two strong assumptions:

- If the (real) emissions of a sector in a grid cell increase or decrease, all species are equally affected, i.e. the chemical composition does not change.
- The chemical composition of the sectors, which is specific in each grid cell, is correctly estimated in the emission inventory.

However, there are cases in which the assumptions do not apply, for example:

- There is a high dependence of NO_x emissions from light-duty vehicles on the outdoor temperature (Grange et al., 2019). This does not apply to the CO emissions. The chemical composition of road transport emissions thus changes with higher temperature towards a lower NO_x/CO ratio. The inventory used for the studies of this work does not take this into account. However, strong deviations in the composition are only expected at temperatures above 15 °C, which did not occur during the simulation period of the studies (10 to 23 March 2016).
- Applications of waste air purifications, e.g. in industrial facilities and power plants, have a large impact on the chemical composition of the emissions. This is

for example the reason for the significant decrease in SO_2 emissions from power plants in the last decades (Smith et al., 2011). Emission inventories represent such historical trends. However, the efficiency of the applied methods can vary between facilities even if they belong to the same sector (Asif et al., 2022). This is a source of uncertainty concerning the estimated chemical composition of emissions.

Following the assumptions above, for each sector and grid cell, a full correlation between the emissions of all species is introduced. Consequently, all species in a sector get the same emission correction factor. This is expected to improve the exploitation of the chemical composition for the distinction between the sectors by the data assimilation system. Furthermore, the degree of freedom of the emission correction factors is reduced.

In order to achieve a full correlation, the emission error correlation matrix $\mathbf{\Omega}$ (Fig. 3.3) contained in the emission error covariance matrices ($\mathbf{K}_s$) is modified. The new matrix $\mathbf{\Omega}$ has the property that $\mathbf{\Omega}\,\mathbf{\Omega}^T$ is a matrix in which all entries are equal to 1. This is necessary because both $\mathbf{\Omega}^T$ and $\mathbf{\Omega}$ are applied in the minimisation procedure. Following this definition, the emission correlation matrix $\mathbf{\Omega}$ has the form

$$\mathbf{\Omega} = \frac{1}{\sqrt{N}} \cdot \mathbf{O} \ . \tag{4.5}$$

N is the number of emitted gas phase and aerosol species considered in the data assimilation and $\mathbf{O}$ denotes a matrix with the dimension of $N \times N$. All entries of $\mathbf{O}$ are equal to 1.

4.3. Anisotropic correlation of emission correction factors along roads

In the current 4D-Var data assimilation system of the EURAD-IM, an isotropic horizontal diffusion operator $\mathbf{L}_h$ is applied to the emission correction factors (Eq. 3.12). It is based on the assumption of an isotropic spatial correlation of emission correction factors. This assumption applies to all emission correction factors in the same way. The approach for sector specific emission optimisation includes a modification of the horizontal diffusion of the road transport emission correction factors. A high spatial correlation of emission correction factors along roads is assumed for the road transport sector. In Fig. 4.1, the principle of anisotropic diffusion of emission correction factors along roads is illustrated and compared to the isotropic diffusion approach of the current EURAD-IM.

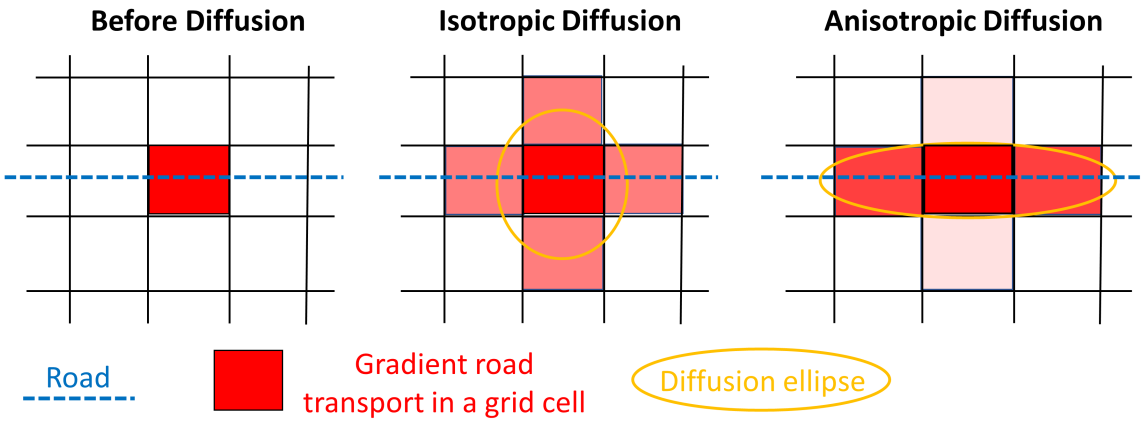

Figure 4.1.: Illustration of the principle of anisotropic diffusion along roads for emission correction factors of road transport and their gradient. A case is considered in which a non-zero cost function gradient with respect to the emission correction factors for road transport is determined for a grid cell in the adjoint run (left picture). This grid cell contains a road running in the x-direction. If isotropic diffusion is applied as in the current EURAD-IM data assimilation system, all neighbour grid cells in the x-direction and y-direction receive equal gradients after the diffusion (second picture). In contrast, the anisotropic diffusion along roads induces a stronger gradient in the neighbour grid cells in the x-direction than in the neighbour grid cells in the y-direction (right picture). This means that the gradient is stretched along the road.

The increased correlation along roads is implemented with a special diffusion operator for the road transport sector $\mathbf{L}_{h,roadtr}$. $\mathbf{L}_{h,roadtr}$ causes a stronger diffusion in the direction of roads than perpendicular to those. This means that the emission correction factors for road transport and the related cost function gradient are stretched along roads. Perpendicular to the direction of the roads, they are shrunk.

Information about the location of roads within the model domain is taken from Open Street Map (OSM) (OpenStreetMap, 2023). All roads are considered that are classified by OSM as motorways, trunk roads, primary roads and secondary roads. For the approach of anisotropic correlation of emission correction factors along roads, a road field ρ is introduced in which these roads are aggregated. In the road field, each grid cell is assigned to the road of the highest category contained in the grid cell. With the road field, the strength of the stretching along roads is determined, as described below. It is assumed that the correlation of emission correction factors along roads is higher for roads of a higher category. For instance, if a car drives on a motorway, it will pass a larger number of consecutive grid cells than a car driving on a secondary road. The consequence is a higher anisotropic correlation of road transport emissions along a motorway than along a secondary road. For this reason, a grid cell containing

a motorway is assigned a higher road field value than a grid cell that contains only secondary roads.

Using OSM information, six categories of grid cells were defined, each assigned a road field value ρ. For grid cells without motorways, trunk roads, primary roads and secondary roads (as classified by OSM), the value of the road field is set to $\rho = 0.1$. If a grid cell contains a road of the highest defined category, the value is set to $\rho = 15$. Between the highest and the lowest category, four categories are defined. The following road field values are assigned to them: 1, 2, 5 and 10. Note that these categories deviate from the official classification of roads, e.g. into "Autobahn", "Bundesstraße", "Landesstraße" and "Kreisstraße" in Germany. The road field of the North Rhine-Westphalian domain is shown in Fig. 4.2.

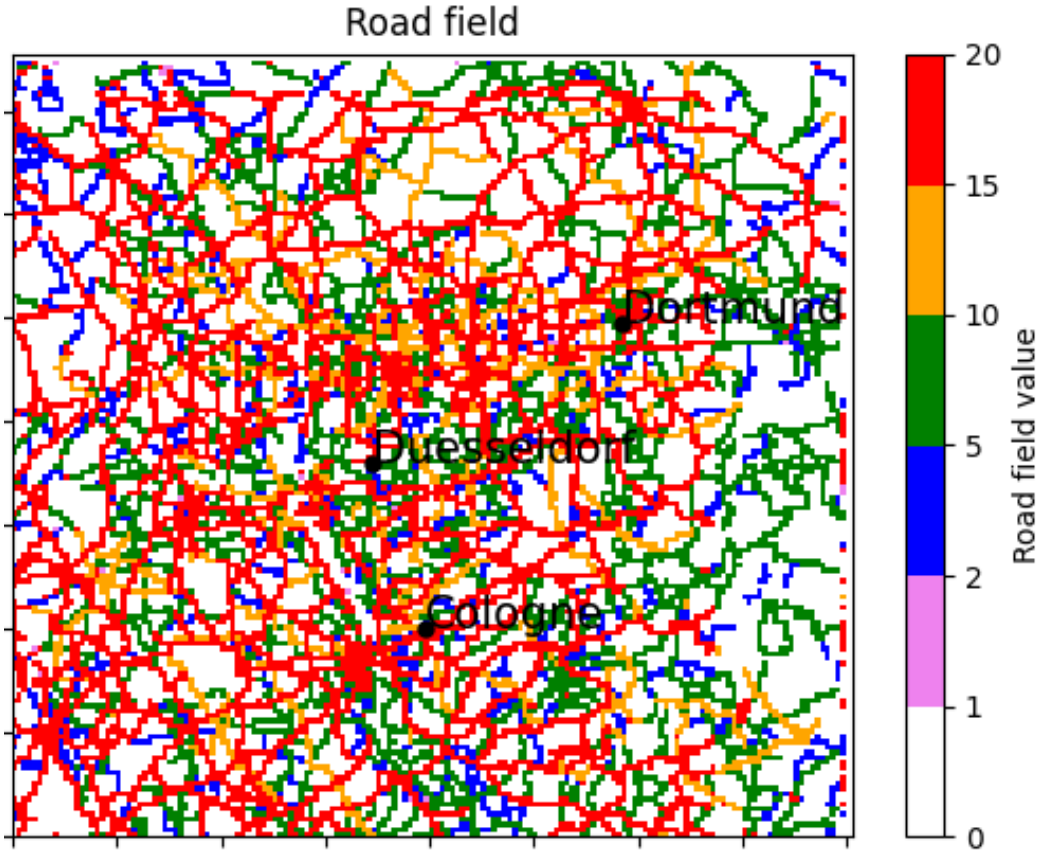

Figure 4.2.: Road field ρ for the domain in North Rhine-Westphalia with a horizontal resolution of 1 km × 1 km. The minimum value is 0.1 (white) and corresponds to grid cells without motorways, trunk roads, primary roads and secondary roads (as classified by OSM). The maximum value is 15 (red), which corresponds to grid cells that contain at least one road of the highest defined category.

In the following, the calculation of the anisotropic diffusion along roads is described. The road field ρ is used to define the direction and the strength of the stretching of the emission correction factors. The direction is perpendicular to the two-sided gradient $\nabla\rho$ of the road field. Its x-component $\nabla_x\rho$ is calculated by averaging the difference to the right and left neighbour grid cells:

$$\nabla_x\rho\,(i,j) \;=\; \frac{[\,\rho(i,j)-\rho(i+1,j)\,]+[\,\rho(i,j)-\rho(i-1,j)\,]}{2}\;, \tag{4.6}$$

where i and j denote the horizontal indices of the grid cells. Its y-component direction, $\nabla_y\rho$, is calculated by averaging the difference to the upper and lower neighbour grid cells:

$$\nabla_y\rho\,(i,j) \;=\; \frac{[\,\rho(i,j)-\rho(i,j+1)\,]+[\,\rho(i,j)-\rho(i,j-1)\,]}{2}\;. \tag{4.7}$$

For the grid cells at the edge of the domain, no road field gradients are calculated.

In the following, the stretching angle and stretching factors are defined. These indicate the direction of the stretching ($\alpha_{stretch}$), the strength of the stretching along the road field (S_1) and the strength of the shrinking perpendicular to the road field (S_2). The variables $\alpha_{stretch}$, S_1 and S_2 can be interpreted as parameters of a diffusion ellipse, as illustrated in Fig. 4.3. First, the normalised total gradient

$$\nabla_{tot}\rho \;=\; \frac{\sqrt{(\nabla_x\rho)^2+(\nabla_y\rho)^2}}{\max\left(\sqrt{(\nabla_x\rho)^2+(\nabla_y\rho)^2}\right)} \tag{4.8}$$

is determined for each grid cell. The maximum total gradient in the domain, $\max\left(\sqrt{(\nabla_x\rho)^2+(\nabla_y\rho)^2}\right)$, is used as the normalisation factor. This means that the gradients $\nabla_{tot}\rho$ are normalised such that the maximum total gradient is equal to 1. In the next step, the stretching factors in the stretching direction (S_1) and in the shrinking direction (S_2) as well as the stretching angle ($\alpha_{stretch}$) are determined for each grid cell:

$$S_1 \;=\; 1+\nabla_{tot}\rho\cdot(S_{max}-1)\;;\; S_2 \;=\; \frac{1}{S_1}\;;\; \alpha_{stretch} \;=\; -\arccos\!\left(\frac{\nabla_y\rho}{\nabla_{tot}\rho}\right). \tag{4.9}$$

S_{max} denotes the maximum stretching factor specified in the configuration of the data assimilation. In the studies in this work, the value of S_{max} is set to 100. $\alpha_{stretch}$ is perpendicular to the gradient $\nabla\rho$. The formula for S_1 is designed to have a minimum value of 1 if the total gradient of the road field ($\nabla_{tot}\rho$) is equal to 0. In this case, an isotropic diffusion is applied. The maximum value S_{max} is obtained for S_1 if $\nabla_{tot}\rho$ is equal to 1.

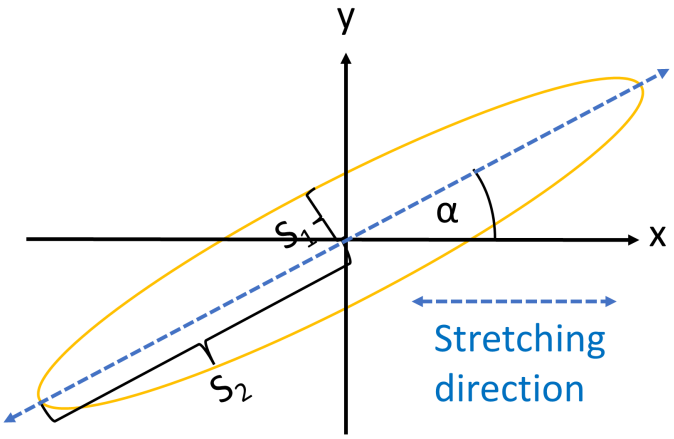

Figure 4.3.: Illustration of the stretching factors S_1 and S_2 and the stretching angle $\alpha_{stretch}$. The anisotropic diffusion along roads can be described by a diffusion ellipse. The stretching along the road field (S_1) corresponds to the semi-major axis of the diffusion ellipse. The shrinking perpendicular to the road field (S_2) corresponds to the semi-major axis of the ellipse. The stretching angle $\alpha_{stretch}$ corresponds to the angle of rotation of the ellipse with respect to the considered coordinate system and determines the direction of the stretching.

From the geometry of ellipses, three stretching factors in the x-direction, y-direction and xy-direction are derived so that the stretching is along the road field ρ:

$$S_x = S_1 \cdot \cos(\alpha_{stretch})^2 + S_2 \cdot \sin(\alpha_{stretch})^2 \tag{4.10}$$

$$S_y = S_1 \cdot \sin(\alpha_{stretch})^2 + S_2 \cdot \cos(\alpha_{stretch})^2 \tag{4.11}$$

$$S_{xy} = (S_1 - S_2) \cdot \sin(\alpha_{stretch}) \cdot \cos(\alpha_{stretch}), \tag{4.12}$$

following the spherical correlation modelling described by Weaver and Courtier (2001), and the approach of Schwinger (2006) for an anisotropic diffusion stretched along a potential vorticity field. The stretching factors (Eq. 4.10, 4.11, 4.12) only apply to the lowest layer of the model grid. For all layers above the lowest, S_x and S_y are 1 and S_{xy} is 0.

The horizontal diffusion operator $\mathbf{L}_{h,roadtr}$, has the following form (Schwinger, 2006):

$$\mathbf{L}_{h,roadtr} = \mathbf{I} + \kappa_h \cdot \Delta t \cdot div\,(\mathbf{S}\, grad_h)\,;\ \mathbf{S} = \begin{pmatrix} S_x & -S_{xy} \\ -S_{xy} & S_y \end{pmatrix}. \tag{4.13}$$

$\mathbf{I}$ is the unit matrix, κ_h the horizontal diffusion coefficient, Δt the time interval of the diffusion step and $grad_h$ the horizontal gradient operator. $\mathbf{L}_{h,roadtr}$ is applied to the cost function gradient with respect to the emission correction factors for road transport.

As a consequence of the change of the diffusion operator, also a specific normalisation matrix for the road transport sector ($\mathbf{\Lambda}_{roadtr}$) has to be calculated (Eq. 3.14). Both matrices, $\mathbf{L}_{h,roadtr}$ and $\mathbf{\Lambda}_{roadtr}$, are part of the emission error covariance matrix for road transport ($\mathbf{K}_{roadtr}$).

Although the spatial correlation of road transport emissions along roads is a reasonable assumption, the exact properties of this correlation in the real world, such as the correlation lengths, are not known. This lack of information contributes to the uncertainty of the data assimilation results. Nevertheless, an improvement of the results can be expected from the application of the anisotropic diffusion operator ($\mathbf{L}_{h,roadtr}$) compared to the isotropic diffusion operator, since the isotropic diffusion operator does not contain any information about the spatial correlation of road transport emissions along roads.

5. Model input

This chapter describes all model inputs used in the studies, including the initial and boundary conditions (Section 5.1), the meteorological data (Section 5.2), the emissions (Section 5.3) and the observations (Section 5.4). The assumptions made for the emission data are discussed in detail, since they have a crucial impact on the results of the sector specific optimisation of emissions in the studies of this work (Chapters 6 and 7).

5.1. Initial and boundary conditions

For predictions of pollutant concentrations and estimations of emission correction factors with high spatial resolution, nesting is applied. Large model domains, e.g. containing whole continents, necessarily have a coarse resolution to enable computationally affordable simulations on supercomputers such as JUWELS (Alvarez, 2021) and JURECA (Thörnig, 2021). With the nesting technique, high resolution domains are sequentially embedded in coarse domains. A one-way multiple-nesting has already been implemented in the EURAD model (Jakobs et al., 1995).

Nesting sequence for the simulations

The following nesting sequence, depicted in Fig. 5.1, is used for the studies presented in Chapters 6 and 7. The boundary values of the European domain with a horizontal resolution of 15 km are taken from the European Centre for Medium-Range Weather Forecasts Atmospheric Composition Reanalysis 4 (EAC4) (Inness et al., 2019) of the Copernicus Atmosphere Monitoring Service (CAMS). The daughter of the European domain is the Central European domain with a resolution of 5 km. For both the 15 km and the 5 km domain, analysis data from EURAD-IM simulations are available. All simulations presented in Chapter 6 and 7 are performed on a domain with 1 km resolution. The latter is the daughter of the Central European domain and contains most parts of North Rhine-Westphalia with the river Rhine, which flows through Cologne and Düsseldorf. The region is selected due to high emissions from industry and road traffic, which allows for the investigation of the distinction between industrial and traffic emissions by the approach of this work (Chapter 4).

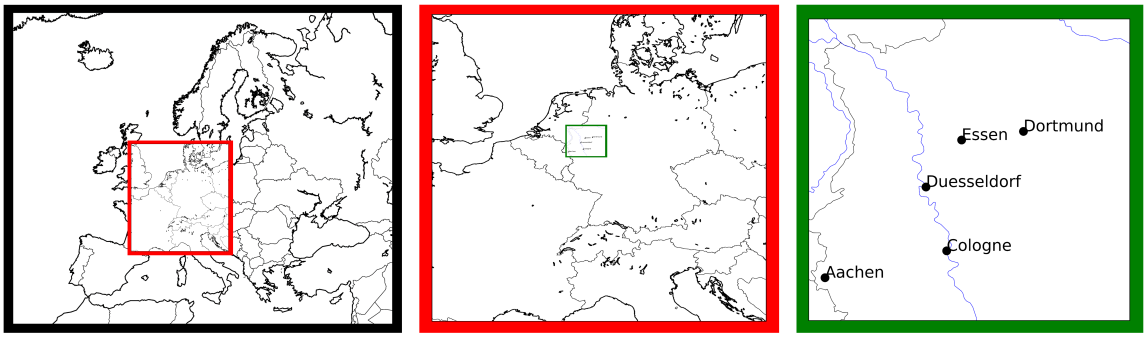

Figure 5.1.: Nesting sequence for the simulations discussed in Chapters 6 and 7: the European domain (15 km × 15 km resolution, black), the Central European (5 km × 5 km domain resolution, red) and the North Rhine-Westphalian domain (1 km × 1 km resolution, green).

Spin-up, initial state and background emission correction factors for the simulations

For all simulations in the 1 km domain (Chapters 6 and 7), the spin-up period is two days. In this period, model runs without data assimilation are performed. The first day of the spin-up itself starts with interpolated initial values of the 5 km mother domain. After the spin-up, daily data assimilation cycles are performed to determine emission correction factors (Eq. 3.3) for each day of the simulation period. The background emission correction factors ($\mathbf{e}_B$) of the first day of a data assimilation period are equal to 1 for all species and sectors, i.e. the simulation starts with uncorrected emission data. For the following days, the data assimilation cycle uses the emission correction factors of the analysis ($\mathbf{e}_A$) of the previous day as background emission correction factors ($\mathbf{e}_B$). The initial state of the background ($\mathbf{x}_{0,B}$) is the analysis state of the simulation of the previous day at 24 UTC ($\mathbf{x}_{t_E,A}$).

5.2. Meteorological fields

For simulations with the EURAD-IM (Section 3), meteorological input data are necessary. They are generated by separate forecast runs of the Weather Research and Forecasting Model (WRF), version 3.7 (Skamarock et al., 2008). The WRF is a nonhydrostatic mesoscale Eulerian model, using Cumulus parametrisations and taking into account surface physics, planetary boundary layer physics and atmospheric radiation physics. Since separate WRF simulations are performed for the EURAD-IM input, EURAD-IM predictions of the concentrations of atmospheric pollutants, e.g. aerosols, have no impact on the meteorological fields predicted by the WRF model. As in the EURAD-IM, a one-way nesting technique is used in the WRF model, with the same nesting sequence as described in Section 5.1.

5.3. Emission data

As mentioned in the introduction, emissions are crucial input data for atmospheric chemistry transport models. In the simulations of this work (Chapters 6 and 7), both anthropogenic and biogenic emissions are taken into account, which are calculated separately in the model. Both have a crucial influence on the concentrations of trace gases and aerosols. Biogenic sources emit mainly non-methane volatile organic compounds (NMVOCs). However, the observation dataset used in the studies of this work does not include concentrations of NMVOCs. Therefore, no reliable corrections for biogenic emissions can be expected in these studies in contrast to anthropogenic emissions. For this reason, this section mainly focuses on the anthropogenic emission data used in the simulations (Chapters 6 and 7). In the following, the annual emission totals, spatial distributions, chemical compositions, height profiles and temporal distributions of the emissions of the Gridding Nomenclature for Reporting (GNFR) sectors (Granier et al., 2019) are discussed.

Anthropogenic emissions

In EURAD-IM, the calculation of anthropogenic emissions is based on national emission totals. They are reported for many different trace gases and aerosols that are harmful to the climate, the environment and to human health. For the input of the EURAD-IM, the emission data for five trace gas species (CO, NH_3, NO_x, SO_x, NMVOC) and two aerosol species (PM_{10}, $PM_{2.5}$) are used.

The emissions are broken down into 12 GNFR sectors, listed in Table 5.1. A list of the processes included in each GNFR sector can be found, for example, in Schneider et al. (2016). The public power sector contains all processes of power plants for public electricity and heat production. The industry sector includes all processes that cause emissions in the manufacturing industry (not only stationary combustion), e.g. in the steel industry, the chemical industry and the food industry. The main emission sources of the sector of other stationary combustion are residential combustion plants. The fugitives sector includes emission sources related to coal mining, fuel exploitation and solid fuel transformation. The solvents sector contains coating applications, printing and the use of degreasing agents, chemical products and fungicides. The road transport sector includes the combustion of fuels by passenger cars, light and heavy duty vehicles as well as other processes that cause emissions, such as road abrasion, brake and tyre wear and evaporation of gasoline. The main emission sources of the shipping sector and the aviation sector are ships and aircraft. The off-road traffic sector includes off-road vehicles used in agriculture, forestry, manufacturing industries or gardening. Major emission sources of the waste sector are plants for the incineration of industrial, clinical and municipal waste. The livestock agriculture sector (also referred to as agriculture I sector) includes emissions related to livestock farming, e.g. dairy cows and pigs. The Other agriculture sector (also referred to as non-livestock agriculture or agriculture II sector) includes agricultural emission sources that are not related

to livestock farming, but e.g. to farm-level and off-farm transport of agricultural products.

Table 5.1.: GNFR sectors in the emission inventory used as an input of the EURAD-IM (Granier et al., 2019).

GNFR category	Category name
A	Public power
B	Industry
C	Other stationary combustion
D	Fugitives
E	Solvents
F	Road transport
G	Shipping
H	Aviation
I	Off-road traffic
J	Waste
K	Livestock agriculture (Agriculture I)
L	Other (non-livestock) agriculture (Agriculture II)

For the studies that are discussed in this thesis, gridded data for Germany and Europe are necessary. This applies to the NRW domain as well as to the mother domains which are necessary for the boundary conditions. For Germany, the Federal Environmental Agency (UBA) provides emission data in the resolution of each model domain. The data are processed by the GRETA tool (Schneider et al., 2016). This tool extracts the total German emissions from the Central System Emissions of Germany (ZSE). The emission totals are spatially distributed to point sources (mainly power plants and industrial factories), line sources (especially for traffic emissions) and area sources. A detailed description can be found in Schneider et al. (2016). The emission data for other European countries are based on the inventory of the Copernicus Atmosphere Monitoring Service (CAMS) (Kuenen et al., 2018). The CAMS dataset has a resolution of $0.05° \times 0.1° \approx 6$ km $\times$ 6 km in Central Europe. The national emission sums of European countries, reported to the Centre on Emission Inventories and Projections (CEIP), are used. Further information about the CAMS dataset can be found in Kuenen et al. (2018). To apply this data in the EURAD-IM, the EURAD emission module (EEM) preprocessor (Memmesheimer et al., 1991) projects the CAMS emissions onto the NRW model grid, using land use and land cover information of Open Street Map (OpenStreetMap, 2023) and CORINE Land Cover 2012 (Copernicus, 2023a). These data are used for the parts of Belgium and the Netherlands, which are contained in the NRW domain (Fig. 5.1).

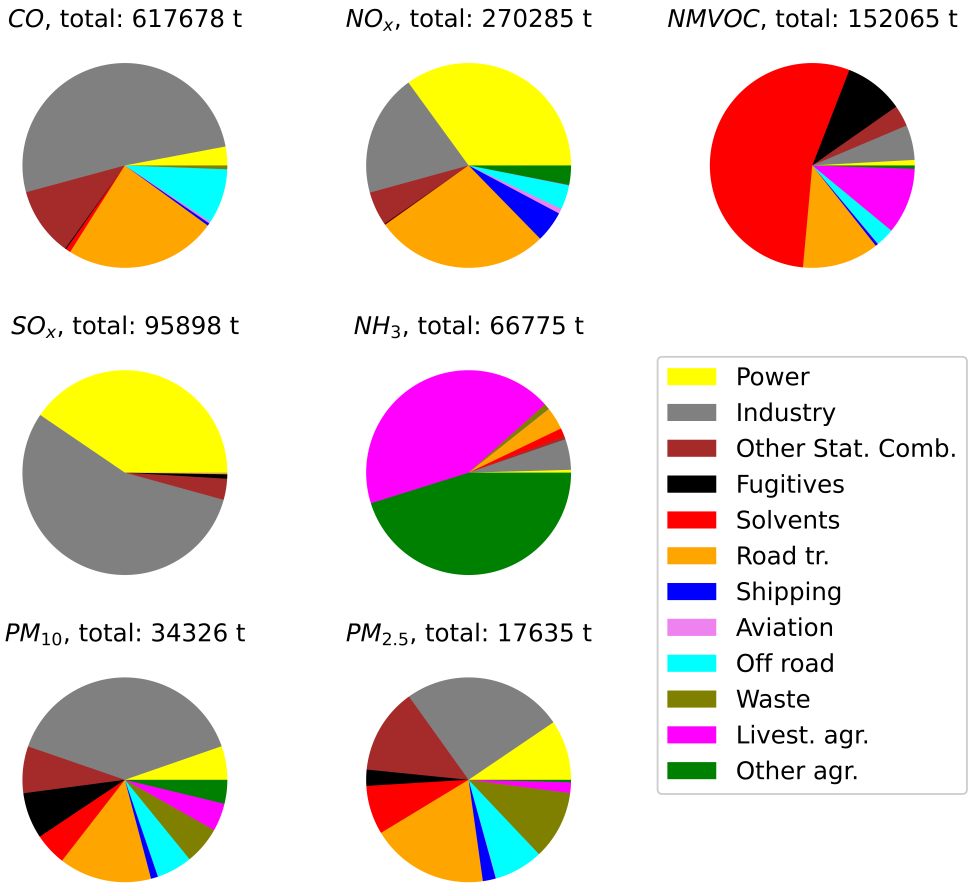

Figure 5.2.: Annual total emissions per species in the NRW domain and their distribution among the GNFR sectors. NMVOC denotes non-methane volatile organic compounds.

Fig. 5.2 gives an overview of the total annual emissions of the species for each GNFR sector in the NRW domain, provided by the emission inventory. Approximately, one third of the NO_x emissions originate from road transport and roughly one third from the public power sector. In contrast, the industry sector is the dominant anthropogenic source of CO and PM_{10} within the NRW domain. More than 90 % of the SO_x emissions are from the industry and public power sector. The most important anthropogenic source of non-methane volatile organic compounds NMVOCs is the solvents sector. About 90 % of the NH_3 emissions are attributed to agricultural activities. The strongest $PM_{2.5}$ emitters are industry and road transport. Thus, power plants, industrial factories, road vehicles and agriculture are the main anthropogenic polluters in NRW according to the emission data.

The chemical composition of the emissions from the public power, industry, road transport and both agriculture sectors is depicted in Fig. 5.3. The average chemical composition of the inventory emissions in the NRW domain for the year 2016 is

considered. Note that the chemical composition is specific in each grid cell and thus may differ from the average values shown in Fig. 5.3. The sectors differ significantly in the ratio of SO_x/NO_x. Compared to the road transport (<0.01) and the agricultural sectors (0), the industry (1.02) and public power (0.41) sectors have a high SO_x/NO_x ratio. The sectors also show large differences in the CO/NO_x ratio. The ratio is highest in the industry sector (6.1), followed by the road transport sector (2.0) and the public power sector (0.19). The two agricultural sectors, on the other hand, are characterised by a high NH_3 fraction in the total emissions of the inventory. The NH_3 fraction is 74% in the livestock agriculture sector and 62% in the agriculture II sector. The livestock agriculture and non-livestock agriculture sectors differ mainly in the fractions of NO_x (0.3% vs. 20%) and non-methane volatile organic compounds (34% vs. 2%).

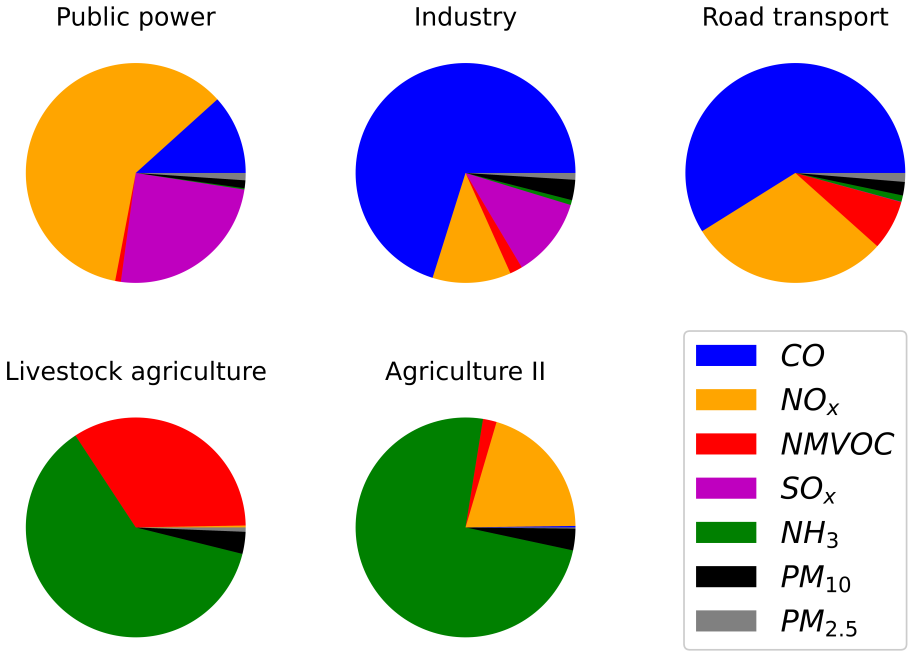

Figure 5.3.: Chemical composition of the emissions from the public power, industry, road transport and both agriculture sectors. The average chemical composition of the inventory emissions in the NRW domain for the year 2016 based on the inventory of the Copernicus Atmosphere Monitoring Service (CAMS) (Kuenen et al., 2018) and the GRETA tool (Schneider et al., 2016) of the Federal Environmental Agency (UBA) is depicted for each sector. Note that the chemical composition in the emission data is specific in each grid cell and thus may differ from the average values shown in this figure.

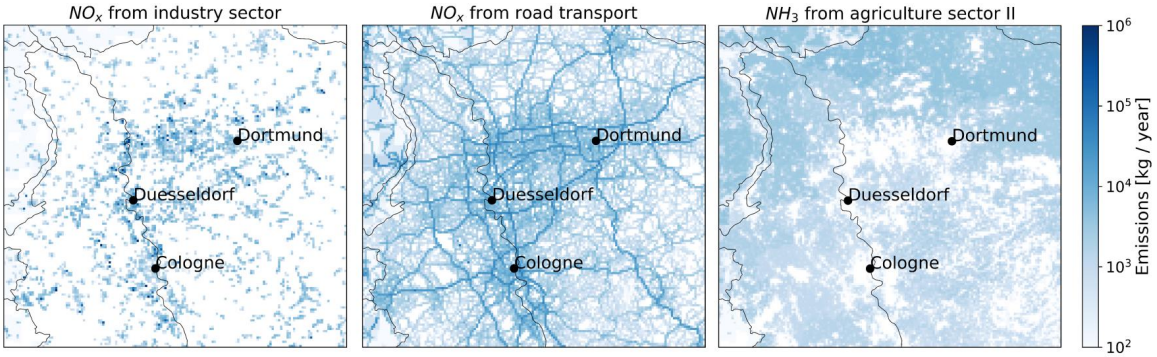

Figure 5.4.: Spatial distribution of yearly emissions in the NRW domain, depicted in a logarithmic scale. Left: NO_x emissions of the industry sector. Centre: NO_x emissions of the road transport sector. Right: NH_3 emissions of the non-livestock agriculture sector.

Fig. 5.4 shows exemplarily the spatial distribution of the industrial, road traffic and non-livestock agriculture emissions. The industrial emission sources are mostly located along the Rhine river Cologne and Düsseldorf, as well as in the area west of Dortmund. A characteristic feature of industrial are the point sources with large emissions, primarily located in the Ruhr area. A large part of the road transport emissions originates from the motorways and the national highways. They are distributed all over the NRW domain, but hotspots can be seen in the region between the cities of Cologne, Düsseldorf and Dortmund. The emission sources of the non-livestock agriculture are located in non-inhabited areas. Strong agricultural emissions occur in the north of the NRW domain.

For each sector, a fixed height profile of emissions is assumed in the emission data. The same distributions as in the CAMS air quality forecasts (Collin, 2020) are used. The effective emission heights are considered taking into account plume rise and buoyancy effects described by Briggs (1984). In the public power and the industry sector, different emission height profiles are assumed for point and area sources. The relative weights of emissions per model layer in the EURAD-IM are depicted in Fig. 5.5. Most sectors, such as the road transport sector, only emit below 20 m height and thus in the two lowest model layers, respectively. The public power sector emits highest up to ~1000 m, followed by the waste, the industry and the shipping sector. The differences between the distributions for the public power, the waste and the industry sector result mainly from the statistical distributions of the stack heights of power plants, waste incineration plants and industrial facilities. The emission height profile for the shipping sector differs from that of other sectors, e.g. the road transport sector, due to the exhaust height of large ships.

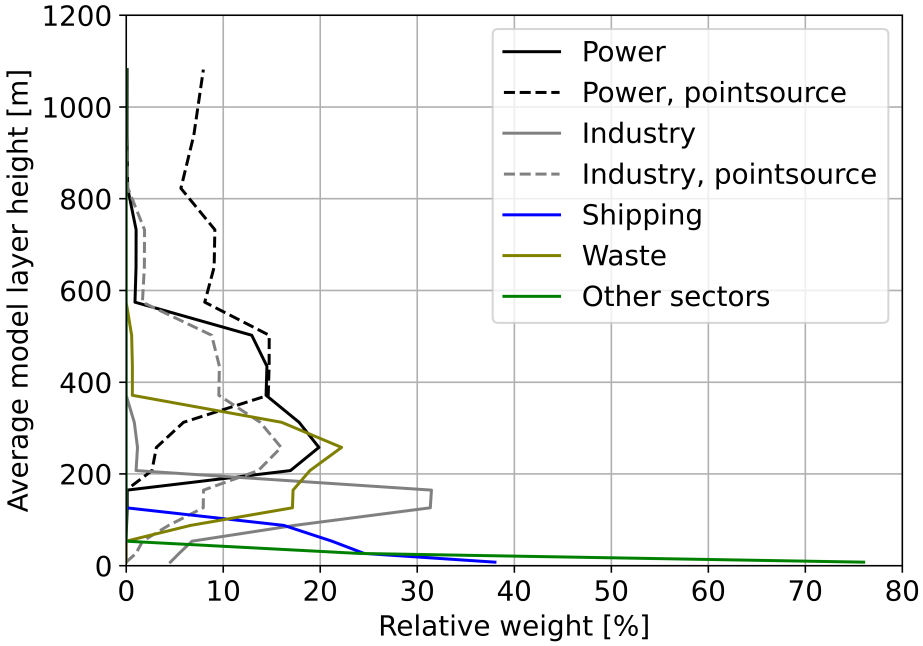

Figure 5.5.: Distributions of the effective emission heights of the sectors. The x values indicate the relative weights of the emissions in the respective model layers. Note that the public power and the industry sector contain area and point sources to which different emission height profiles are assigned.

The species considered in the emission data for the anthropogenic sectors differ from the emitted species defined in the EURAD-IM. For this reason, the emissions of NO_x, NMVOC, SO_x, PM_{10} and $PM_{2.5}$ have to be distributed among the model species. For this, the same branching factors are used as in the CAMS forecasts (Collin, 2020). They are fixed for each emission sector, i.e. they apply to all grid cells and do not depend on the season or time of day. The NO_x emissions consist of NO and NO_2. The ratio of NO_2/NO_x is 20% for the road transport sector, 10% for the aviation sector, 10% for the off-road traffic sector and 5% for all other sectors. For the SO_x emissions, it is assumed that these consist of 4% H_2SO_4 and 96% SO_2 in all sectors. The NMVOC emissions of the GNFR sectors are distributed among 25 NMVOC groups defined by the Global Emissions Initiative (GEIA), described in Huang et al. (2017). Fixed branching factors for each sector are used. The emissions of the GEIA NMVOC groups are assigned to the 18 emitted NMVOC species defined in the EURAD-IM (see e.g. Fig. 3.3). Both the PM_{10} and $PM_{2.5}$ emissions are distributed among five aerosol species. These include elemental carbon, organic carbon, sulphate, sodium and other minerals, each with a diameter of 2.5-10 µm (PM_{10}) or <2.5 µm ($PM_{2.5}$), respectively.

The EURAD-IM model requires hourly emission data. In order to calculate these, annual, weekly and daily emission profiles are used, based on the description of Denier van der Gon et al. (2011), with a modification made to match the GNFR categories. These distributions are discussed in the following.

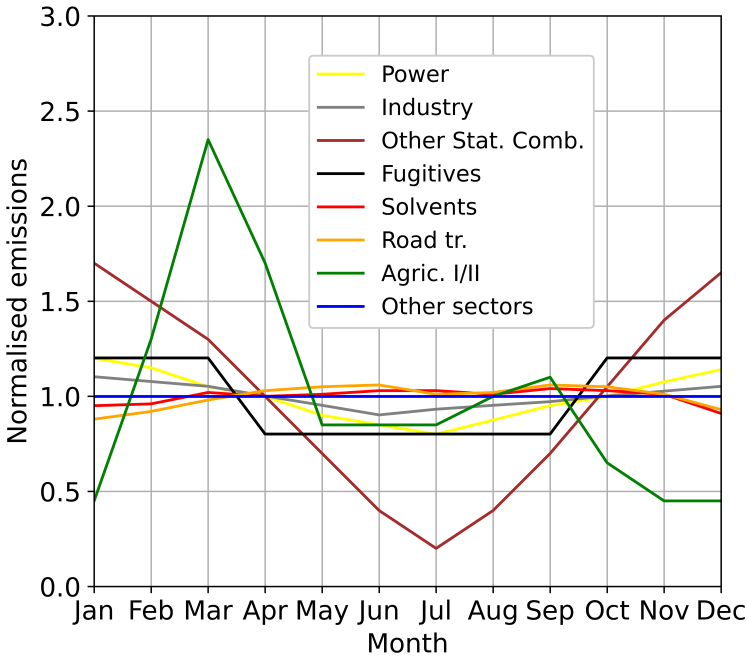

Figure 5.6.: Monthly emission profiles of the GNFR sectors based on the description of Denier van der Gon et al. (2011). These are used to calculate the monthly emissions from the annual emissions. The emissions are normalised to the average value per month.

The monthly emission profiles are used to calculate the monthly emissions from the annual emissions. The distributions normalised to the average value per month are shown in Fig. 5.6. Most GNFR sectors have normalised values between 0.8 and 1.2 for all months, which means that the seasonal variation in their emissions is small. The agricultural sectors have the strongest variations, with low emissions in January, November and December (normalised value of 0.5) and a peak in March (2.3). This reflects the variations in agricultural activity. The sector of other stationary combustion shows a large difference between the emissions in summer and winter months with a maximum in January and December (normalised value of 1.7) and a minimum in July (0.2). This is due to the seasonal heating requirements.

The weekly emission profiles are used to calculate the daily emissions from the monthly emissions, taking into account differences between the days of the week. The distributions normalised to the average value per day are shown in Fig. 5.7. Most sectors are assumed to have the same emissions on all days of the week. Only the solvent, the road transport, the other stationary combustion and the public power sectors deviate from a flat distribution with lower emissions on the weekends. This is

due to less work-related emissions on weekend days.

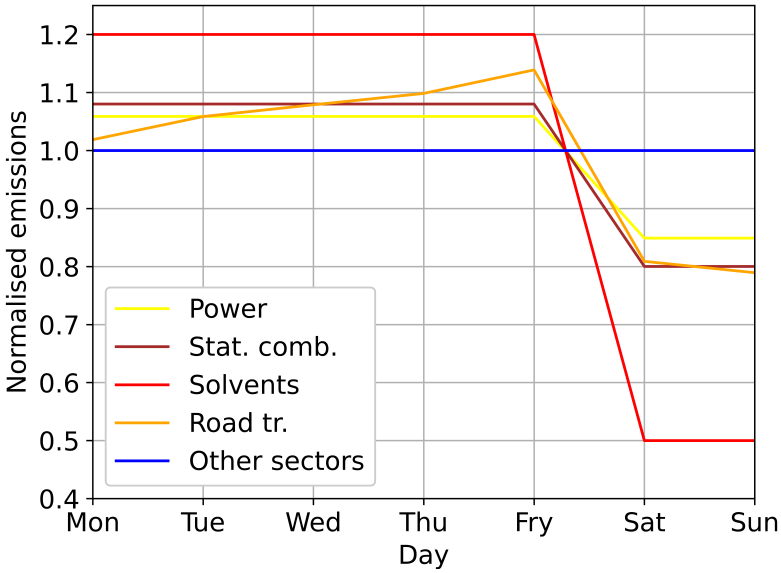

Figure 5.7.: Weekly emission profiles of the GNFR sectors based on the description of Denier van der Gon et al. (2011). These are used to calculate the daily emissions from the monthly emissions, taking into account differences between the days of the week. The emissions are normalised to the average value per day.

The diurnal emission profiles are used to calculate the hourly emissions from the daily emissions. The distributions normalised to the average value per hour are shown in see Fig. 5.8. The fugitives, the shipping, the aviation, the off-road traffic and the waste sector have flat diurnal profiles. The reason for this is a lack of information on the real diurnal profiles of these sectors. The road transport sector is characterised by peaks in the rush hours around 8 h and 17 h (normalised values of 1.9 and 2.1) and negligible emissions between 1 h and 5 h. The sector of other stationary combustion shows a similar distribution with maximum emissions around 8 h (normalised value of 1.5). This profile can be explained by reduction of heating at night. Emissions from solvents and agriculture are also assumed to be significantly higher during the day than at night, according to typical working hours. In comparison, the diurnal profiles of the public power and the industry sector are less pronounced than for the previously mentioned sectors. This is due to night shifts in industrial companies and the resulting electricity demand during the night.

Note that the sector specific diurnal emission profiles are static in the data assimilation, i.e. they can not be changed by the sector specific optimisation of emissions described in Chapter 4. In contrast, the monthly (Fig. 5.6) and weekly (Fig. 5.7) emission profiles as well as the spatial distribution of emissions (Fig. 5.4) can be corrected for each

GNFR sector with this approach.

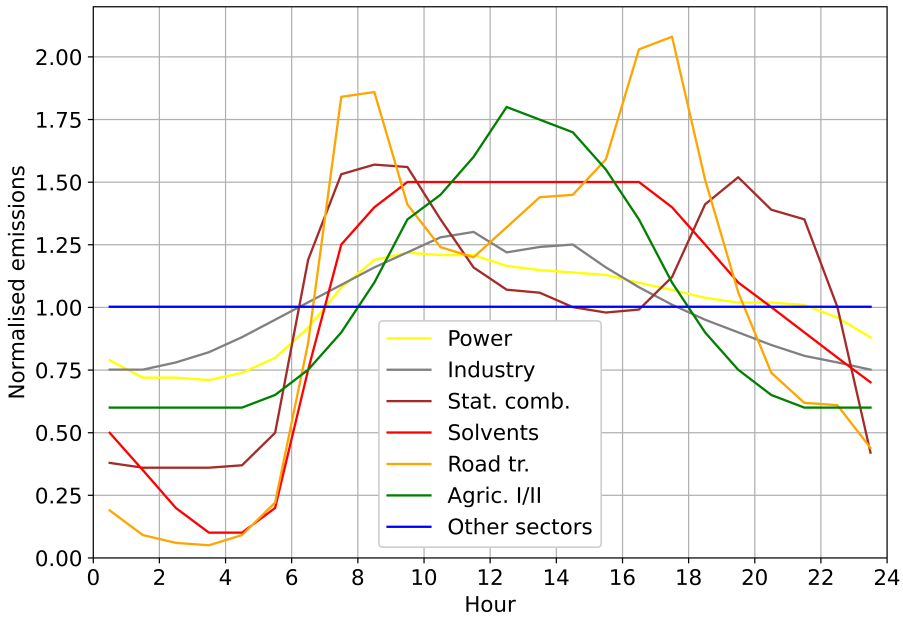

Figure 5.8.: Diurnal emission profiles of the GNFR sectors based on the description of Denier van der Gon et al. (2011). These are used to calculate the hourly emissions from the daily emissions. The emissions are normalised to the average value per hour.

Emissions from natural sources

Biogenic sources are essential for the understanding of atmospheric chemistry, since they are strong emitters of NMVOCs, which are crucial for the O_3 chemistry. In the EURAD-IM, biogenic emissions are calculated using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2012). MEGAN includes 147 compounds - CO and 146 NMVOCs - and estimates the emission rates on the basis of the leaf area index, solar radiation, soil moisture, temperature and CO_2 concentration. Since biogenic emissions are strongly dependent on these parameters, no fixed monthly or diurnal emission profiles are applied in MEGAN. The emissions are determined using activity factors depending on environmental parameters, and emission factors defined for 16 different vegetation types.

Following a MEGAN simulation for the year 2000 discussed in Guenther et al. (2012), ~3% of the total global CO emissions originate from biogenic sources. In contrast, the major part of the global NMVOCs emissions originates from biogenic sources. Isoprene is estimated to contribute with ~50% to the global biogenic NMVOC emissions. The contribution of monoterpenes such as α-pinene, β-pinene and limonene to the global NMVOC emissions is ~15% in the simulation for the year 2000 (Guenther et al., 2012).

Other NMVOC species with high global emissions included in MEGAN are methanol, acetone, ethanol, acetaldehyde, ethene and propene. More information on MEGAN can be found in Kaiser et al. (2012).

In addition to biogenic emissions, emissions from wildfires, soils and sea salt are calculated in the EURAD-IM. These are not described further here because they are zero or negligible in the NRW domain.

In the studies of this work (Chapters 6 and 7), a biogenic emission sector is considered in addition to 12 anthropogenic GNFR sectors, which in principle allows for an optimisation of biogenic emissions. However, no reliable corrections can be expected for biogenic emissions, as the observational data used in the studies do not include NMVOC concentrations.

5.4. Ground observations in North Rhine-Westphalia

In the studies of this work (Chapters 6 and 7), ground station observations are assimilated. For this, validated hourly observational data for the CAMS forecasts (Collin, 2020) are used. The data were extracted from the European Environment Information and Observation Network (Eionet) (Copernicus, 2023b) of the European Environment Agency (EEA) and from the European Monitoring and Evaluation Programme (EMEP) (Tørseth et al., 2012). Most stations in the NRW domain are located in the German federal states North Rhine-Westphalia and Rhineland-Palatinate, where the environmental agencies provide the data of their operational measurements.

Observation error covariance matrix

For the calculation of the costs J and its gradient, an observation error covariance matrix $\mathbf{R}$ is defined (Eq. 4.1). $\mathbf{R}$ has the dimension of $N \times N$, where N is the number of observations in the assimilation window. In the case of ground observations, $\mathbf{R}$ is a diagonal matrix. The diagonal elements (R_{ii}) are given by the error variances of the observations. They are calculated using the following formula:

$$R_{ii} = \sigma_{i,meas}^2 + \sigma_{i,repr}^2 \ . \tag{5.1}$$

$\sigma_{i,meas}$ denotes the measurement error and $\sigma_{i,repr}$ the representativeness error of the observation i.

The measurement error σ_{meas} is related to the accuracy of the measured concentrations, which depends on the measurement instrument. In the case of the observational dataset used in the studies of this work, no information is available on the measurement uncertainties of individual ground stations. Consequently, possible differences

in terms of measurement quality can not be taken into account in the assimilation of observations. The estimation of the measurement errors (σ_{meas}) in the studies of this work is based on the quality assessment of observational data by Mohnen (1999). The estimated values of σ_{meas} depend on the species and the measured concentrations (y^O). For each species, a minimum ($\sigma_{meas}^{abs,min}$) and a relative (σ^{rel}) measurement error are defined so that the formula for σ_{meas} is

$$\sigma_{meas} = \max(\sigma_{meas}^{abs,min}, \sigma^{rel} \times y^O) \ . \tag{5.2}$$

The species-dependent values for $\sigma_{meas}^{abs,min}$ and σ^{rel} are given in Table 5.2.

Table 5.2.: Species-dependent values for the minimum absolute error ($\sigma_{meas}^{abs,min}$) and the relative error (σ^{rel}) used for the calculation of the measurement error (σ_{meas}) of observations (Eq. 5.2), following Mohnen (1999). In this table, only the species included in the observational dataset for the studies in this work are listed.

Species	$\sigma_{meas}^{abs,min}$	σ^{rel}
NO_2	1.5 ppbV	15 %
O_3	1.5 ppbV	10 %
SO_2	1.5 ppbV	15 %
CO	30 ppbV	15 %
PM_{10}	5 µg m^{-3}	15 %
$PM_{2.5}$	3 µg m^{-3}	15 %

The representativeness error σ_{repr} is related to the deviation of the average concentration in the grid cell of the observation from the concentration at the location of the measurement. This deviation is relevant for the comparison of simulated and observed concentrations. σ_{repr} depends on the observed species, the location of the ground observation station as well as on the resolution of the model grid (Δx). Each ground observation station is assigned a location type such as "urban" or "rural", which characterises the location of the station. For example, "urban" locations are characterised by a strong impact of local emissions (e.g. from traffic or industry) on the pollutant concentrations in contrast to "rural" locations. For each location type, a representativeness length (L_{repr}) is specified, which can be interpreted as an influence radius with respect to pollutant concentrations. In addition, a species-dependent absolute representativeness error (σ_{repr}^{abs}) is specified. σ_{repr}^{abs} is multiplied by the square root of the ratio between the grid resolution (Δx) and the representativeness length (L_{repr}) to calculate the representativeness error (σ_{repr}) following Elbern et al. (2007):

$$\sigma_{repr} = \sqrt{\frac{\Delta x}{L_{repr}}} \times \sigma_{repr}^{abs} \ . \tag{5.3}$$

Table 5.3.: Values for the representativeness length (L_{repr}) depending on the location type of the observation station and for the absolute error (σ^{abs}) depending on the species, both according to Elbern et al. (2007). Both quantities are used to calculate the representativeness error of the observations (Eq. 5.3). In this table, only the species included in the observational dataset for the studies in this work are listed.

Station type	L_{repr}
Remote	20 km
Rural	10 km
Suburban	4 km
Urban	2 km
Traffic	1 km
Unknown	3 km

Species	σ^{abs}
NO_2	1.8 ppbV
O_3	2.0 ppbV
SO_2	1.5 ppbV
CO	30 ppbV
PM_{10}	3 $\mu g\,m^{-3}$
$PM_{2.5}$	2 $\mu g\,m^{-3}$

In the studies of this work, the values for σ^{abs}_{repr} and L_{repr} specified in Elbern et al. (2007) are used, which are listed in Table 5.3.

Locations of the ground observation stations

In total, the NRW domain contains 20 NO_2, 17 O_3, 7 SO_2, one CO, 15 PM_{10} and 6 $PM_{2.5}$ observation stations used for the data assimilation in the studies of this work (also referred to as assimilation stations). The locations of the assimilation stations are shown in Fig. 5.9. There are also observational data of so-called validation stations (not depicted in Fig. 5.9) which are not assimilated in the simulations. They are used for the evaluation of the data assimilation results (Chapter 7). In Table 5.4, the number of assimilation and validation stations per location type are indicated for each observed species. NO_2, O_3 and PM_{10} are observed by at least 25 stations (Fig. 5.9 and Table 5.4), whereas significantly fewer ground observation stations measure SO_2 and $PM_{2.5}$ concentrations. Concentrations of CO, an important component of industrial emissions, are measured by only one station in the northwest of the domain. Most of the stations, especially those deploying NO_2, O_3 and PM_{10} measurements, are located in the Rhineland near the cities Düsseldorf and Cologne or in the Ruhr area.

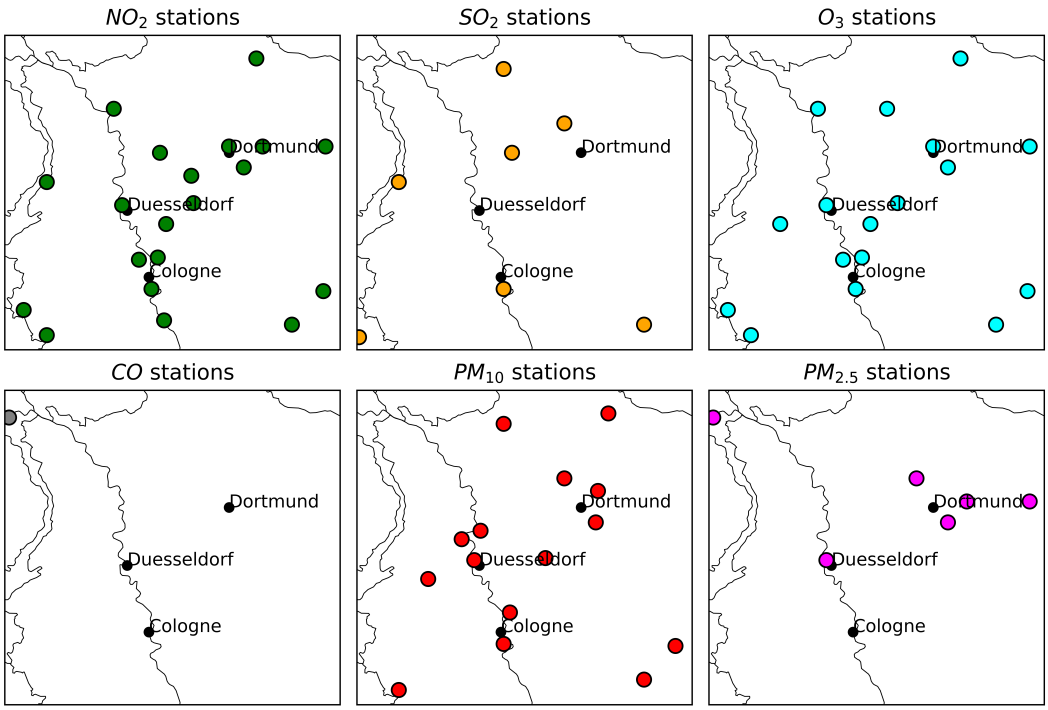

Figure 5.9.: Ground observation stations in the NRW domain used for the data assimilation in the studies of this work (Chapters 6 and 7). The observational dataset for the studies also contains data from stations that are not used for data assimilation, but for the validation of the data assimilation results. The locations of these stations are not shown in this figure.

Concerning the location type (Table 5.4), most of the NO_2, O_3 and PM observations are situated in urban and suburban areas. This corresponds with a low representativeness length of 2 km and 4 km, respectively. Nevertheless, with a resolution of 1 km, the representativeness errors are relatively small: e.g. $\sigma_{repr} = 1$ ppbV in urban and $\sigma_{repr} = 0.7$ ppbV in suburban locations for NO_2 and smaller for the other observed species. The high representativeness of observations is an important advantage of the high horizontal model resolution selected in the studies of this work (Chapters 6 and 7).

Table 5.4.: Number of ground observation stations per location type that are used for assimilation ("assim.") and validation ("valid."). The stations are assigned the location types urban, suburban, rural, rural-regional and rural-nearcity. The location type is related to the representativeness error of the observations. For example, observations of a station in the "urban" category are assumed to have a higher representativeness error than observations of a station in the "rural" category (Eq. 5.3 and Table 5.3).

Location / type	NO_2	O_3	SO_2	CO	PM_{10}	$PM_{2.5}$
Assimilation total	20	17	7	1	15	6
Urban (assim.)	8	8	1	1	5	3
Suburban (assim.)	7	5	2	-	6	2
Rural (assim.)	1	1	2	-	1	-
Rural-regional (assim.)	2	1	-	-	1	-
Rural-nearcity (assim.)	2	2	2	-	2	1
Validation total	10	10	1	-	11	6
Urban (valid.)	3	4	-	-	4	2
Suburban (valid.)	4	3	1	-	6	4
Rural (valid.)	2	2	-	-	-	-
Rural-regional (valid.)	-	-	-	-	-	-
Rural-nearcity (valid.)	1	1	-	-	1	-

Availability of hourly observational data

For most of the stations in the NRW domain, the hourly observational dataset is not complete. The observed concentration of a chemical species at a certain station at a certain hour is missing if no measurements were performed at that time or if the data was rejected according to quality management criteria. The percentage of available data strongly varies between the ground stations. Table 5.5 contains statistics about the data availability between 10 March 2016, 0 UTC and 24 March 2016, 0 UTC. This period corresponds with the simulation period of the study discussed in Chapter 7. The average percentage of available data is higher than 85 % for NO_2, O_3, PM_{10} and $PM_{2.5}$ and no station is below 50 %. The data availability is significantly worse for SO_2 with a mean of 50.7 % and two stations below 10 %. Furthermore, no CO observations are available.

Table 5.5.: Statistics about the availability of the hourly observational data from the ground stations between 10 March 2016, 0 UTC and 24 March 2016, 0 UTC. An availability of 100% means that for each hour in this period an observed concentration of the respective species is available.

	NO_2	O_3	SO_2	CO	PM_{10}	$PM_{2.5}$
Mean of stations	86.1 %	86.1 %	50.7 %	0 %	97.7 %	89.5 %
Standard deviation of stations	5.9 %	4.5 %	27.6 %	-	2.9 %	14.3 %
Median of stations	86.8 %	86 %	63.1 %	-	98.8 %	96 %
Maximum of stations	100 %	96.1 %	81.5 %	-	100 %	100 %
Minimum of stations	67.6 %	74.2 %	5.9 %	-	89.9 %	52.4 %

Conclusions with regard to sector specific emission optimisation

With regard to the sector specific emission optimisation (Section 4), the following aspects of the observational dataset are of particular importance:

- Many O_3 observations are available. These can be used by the data assimilation system to improve the O_3 concentrations in the model. Due to the chemical relationship between O_3, NO_2 and NO, accurate O_3 concentrations in the model are essential for reliable corrections of NO_x emissions.
- Since road transport mainly contributes to the NO_x emissions in NRW (Section 5.3), observed NO_2 concentrations are strongly related to emissions of this sector. The expectation is that the good observational situation of NO_2 leads to reliable road transport emission corrections.
- Many ground stations observe aerosol concentrations. They are related to emissions of several emission sectors, where the industry sector is the most important emitter (Section 5.3). Thus, the aerosol observations can potentially contribute to emission corrections of several sectors.
- The relatively poor observational situation of SO_2 with only 7 assimilation stations and a large amount of missing data (Table 5.5) may limit the reliability of emission corrections for the sectors public power, industry and other stationary combustion.
- CO observations would deliver additional information about the industry sector, the strongest CO emitter, but are not available.
- No observations are available for non-methane volatile organic compounds emitted mainly from biogenic sources and solvents.
- NH_3 concentrations are not observed, which may limit the reliability of emission corrections for the two agricultural sectors emitting large amounts of NH_3.

6. Case studies based on identical twin experiments

In the studies presented in this chapter, the sector specific emission optimisation is applied to artificial emission scenarios with the assumption of zero model and observation errors. The purpose of this study is to investigate the potential and limits of the sector specific emission optimisation in distinguishing between emissions of different sectors under idealised, but realistic, conditions. For this, so-called identical twin experiments are performed, consisting of nature runs and data assimilation runs. Nature runs are simple forward runs of a model in which model parameters and/or initial values are perturbed. In the study of this chapter, the inventory emissions of different sectors are perturbed. For the investigation of the sector specific emission optimisation, synthetic observations are extracted from the nature runs. For both the nature runs and 4D-Var data assimilation runs, the EURAD-IM forward model described in Section 3.1 and the input fields described in Chapter 5 are used, except for the emission perturbation in the nature runs. The aim of the assimilation is to reproduce the perturbed nature run's emissions using the sector specific 4D-Var optimisation (Chapter 4). After a description of the experimental setup (Section 6.1), the resulting emission correction factors are discussed (Section 6.2). Moreover, the impact of the anisotropic diffusion operator for road transport emission corrections is investigated (Section 6.3).

6.1. Experimental setup

Selection of the region

The simulations are performed for the North Rhine-Westphalian domain with a horizontal resolution of 1 km (Section 5.1). The region was selected because of the high emissions from the road transport sector in large cities as well as on motorways, and because of the high industrial emissions (Fig. 5.4). Furthermore, agricultural areas are included with significant emissions (Fig. 5.2 and 5.6). Note that it is important for the investigation of the sector specific optimisation of emissions that the domain covers high emissions from several sectors.

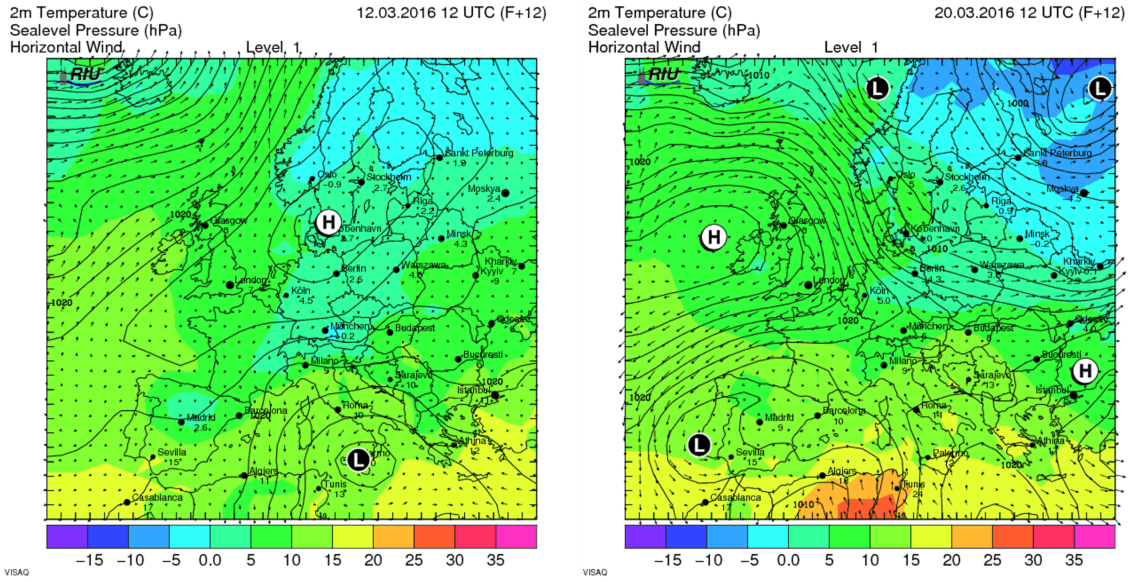

Figure 6.1.: Meteorological conditions of the simulation periods, shown for 12 March 2016 (left) and 20 March 2016 (right) as examples. The colour scale indicates the temperature 2 m above ground. Source: `http://www.eurad.uni-koeln.de/`.

Simulation period and meteorological situation

A simulation period from 10 March to 16 March 2016 is selected. For each of the seven days, consecutive nature runs and data assimilation cycles are performed with an assimilation window of 24 h. The meteorological conditions are suitable for optimising emission correction factors as discussed in the following. Furthermore, the year 2016 was predetermined for this study because of its involvement in a larger project. As the simulation period of the study discussed in Chapter 7 is from 10 to 23 March 2016, the meteorological situation from 10 to 23 March 2016 is described here. In Fig. 6.1, the 2 m temperature, the sealevel pressure and the wind fields are shown for 12 March and 20 March 2016, 12 UTC, as these days are representative for the whole period. On 12 March 2016, the meteorological situation is characterised by a stable anticyclone covering large parts of central Europe and Scandinavia, moving westward in the following days. This is accompanied with easterly winds and no precipitation in the model region. The meteorological situation on 12 March is representative for the period from 10 March to 17 March 2016. In the following days, the meteorological situation changed with a cyclone over Scandinavia on 20 March 2016. At this time, the anticyclone mentioned before was located over the British islands. This led to NNW wind directions and to moderate local precipitation (>1 mm per day) in the NRW domain from 20 March to 23 March 2016. The main reasons for the selection of the simulation period are the following. Firstly, the average wind speed in the domain at a height of 10 m did not exceed 6 m/s during the simulation period, except on 16 March. These relatively low wind speeds limit the transport of the gradient of

the sector specific emission correction factors (explanation in Section 4.1) outside the domain. This reduces the loss of information due to transport. The second reason is the low precipitation in the whole domain, especially between 10 March and 19 March (< 1 mm per day). Due to this, the wet deposition of aerosols, which is a non-linear effect and may complicate the interpretation of the simulation results, played a minor role.

Emission scenarios

Three identical twin experiments with different emission scenarios are discussed in this chapter. In Fig. 6.2, the total emissions of the species and GNFR sectors in the NRW domain for the simulation period are depicted for the emission inventory and for the three experiments. The experiments are described below:

1. In experiment I, the emissions of the industry sector are increased by a factor of 2, whereas the road transport emissions are decreased by a factor of 0.5 compared to the emission inventory. The road transport sector is characterised by high NO_x emissions, while the industry sector is the main emitter of SO_x, CO and PM_{10} in the NRW domain (Fig. 6.2).

2. In experiment II, the non-livestock agricultural emissions are increased by a factor of 2, whereas the road transport emissions are decreased by a factor of 0.5 compared to the emission inventory. The non-livestock agriculture sector is characterised by large emissions in March (Fig. 5.6). They contain a significant amount of NO_x and nearly half of the NH_3 emissions in the domain (Fig. 6.2). It is important to note that NH_3 is not directly observed, but takes part in the formation of secondary aerosols. Thus, NH_3 emissions impact the concentrations of observed species.

3. In experiment III, the non-livestock agricultural emissions are increased by a factor of 1.3 and the road transport emissions are decreased by a factor of 0.5 compared to the emission inventory. With this scenario, the sensitivity of the sector specific emission correction to moderate emission perturbations is investigated.

The mentioned emission perturbations are applied in the nature runs of the respective experiments on all seven days of the simulation period.

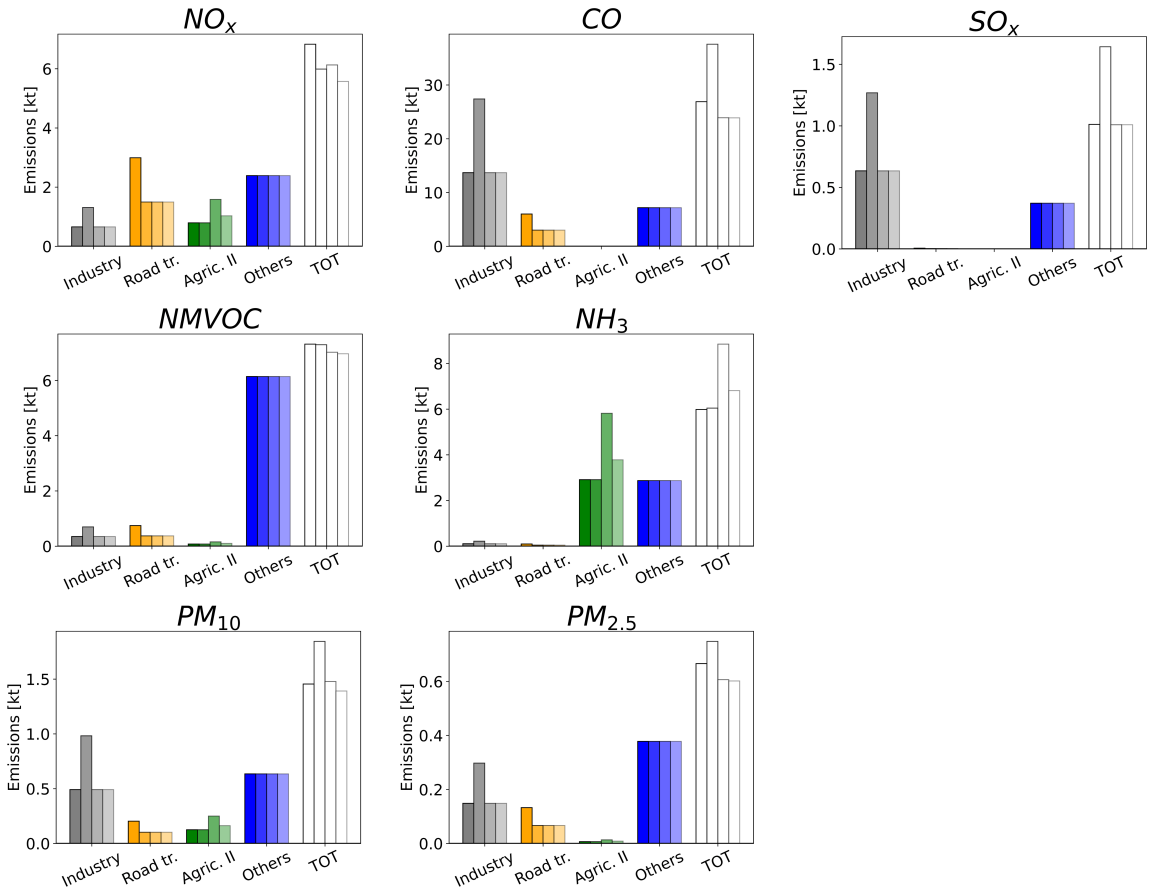

Figure 6.2.: Bar charts of the NO_x, SO_x, CO, NH_3, PM_{10} and $PM_{2.5}$ emissions during the period from 10 March to 16 March 2016 in the NRW domain. The emissions of the industry, road transport ("road tr.") and non-livestock agricultural ("agric. II") sectors as well as the emission sum of the other anthropogenic GNFR sectors (Table 5.1) and the emission sum of all GNFR sectors ("TOT") are shown. For each sector and sector group, the inventory emissions are depicted as first bar, the perturbated emissions in experiment I as second bar, the perturbated emissions in experiment II as third bar and the perturbated emissions in experiment III as fourth bar.

Configuration of the model and data assimilation system

The data assimilation system is configured such that emission correction factors are determined for each GNFR sector (Table 5.1) and for a sector of natural emissions. Both the full correlation between the emission correction factors of all species (Section 4.2) and the anisotropic diffusion of road transport emission correction factors (Section 4.3) are applied in the experiments, unless otherwise stated. The emission errors in the emission error covariance matrix $\mathbf{K}$ (Eq. 3.10) are set to $\Sigma_E = 8.3$. The value of Σ_E is the same for all trace gas and aerosol species in order to achieve equal emission correction factors for all species within the sectors and grid cells. The maximum

number of iterations per data assimilation cycle is set to 15. Besides the aspects mentioned in this paragraph, the configuration of the model and data assimilation system follows the descriptions in the previous chapters.

Observations and observation operator

The locations of the synthetic observations correspond to those of the assimilation stations shown in Fig. 5.9. From the nature runs, hourly synthetic observational data are extracted for each station. Note that this represents an idealised observation situation. For real observations, the availability of hourly observation data is less than 100% (Table 5.5). The observation operator $\mathcal{H}(t)$ interpolates the model concentrations to the location of the ground observation stations.

6.2. Emission corrections and cost reduction

The resulting emission correction factors for all experiments are shown in two types of figures. In Fig. 6.3, the time evolution of the spatially averaged emission correction factors during the simulation period is depicted for each GNFR sector (Table 5.1). The spatial distributions of the emission correction factors for the road transport, industry, non-livestock agriculture, off-road, aviation and fugitives sectors averaged over the simulation period are shown in Fig. 6.4, 6.5 and 6.6. Due to the full correlation of species, the emission correction factors apply to all species. In the following, the behaviour of the emission corrections is discussed for each GNFR sector.

Considering the road transport sector, it is noticeable that in all experiments the emissions are decreased, which is in accordance with the emission perturbation of this sector. However, on average the corrections are significantly weaker than the emission perturbation factor of 0.5 in the nature runs. As indicated in Fig. 6.3, the minimum correction factor is 0.65 on 13 March 2016 in experiment III. The correction factors on the other days and for the other experiments are higher, with up to 0.78. One reason for average emission correction factors significantly above 0.5 is the observation situation: As indicated by the spatial distribution of the road transport sector, the corrections are small in parts of the southeast of the domain where no observations are available (Fig. 5.9). In addition, emission perturbations of other sectors partially affect the emission correction of the road transport sector, especially in experiment II. Fig. 6.5 shows that the strong positive emission perturbation of the agriculture sector can affect the corrections of the road transport emissions. For example, in areas north of Dortmund and east of Cologne, positive emission corrections for the road transport sector are determined.

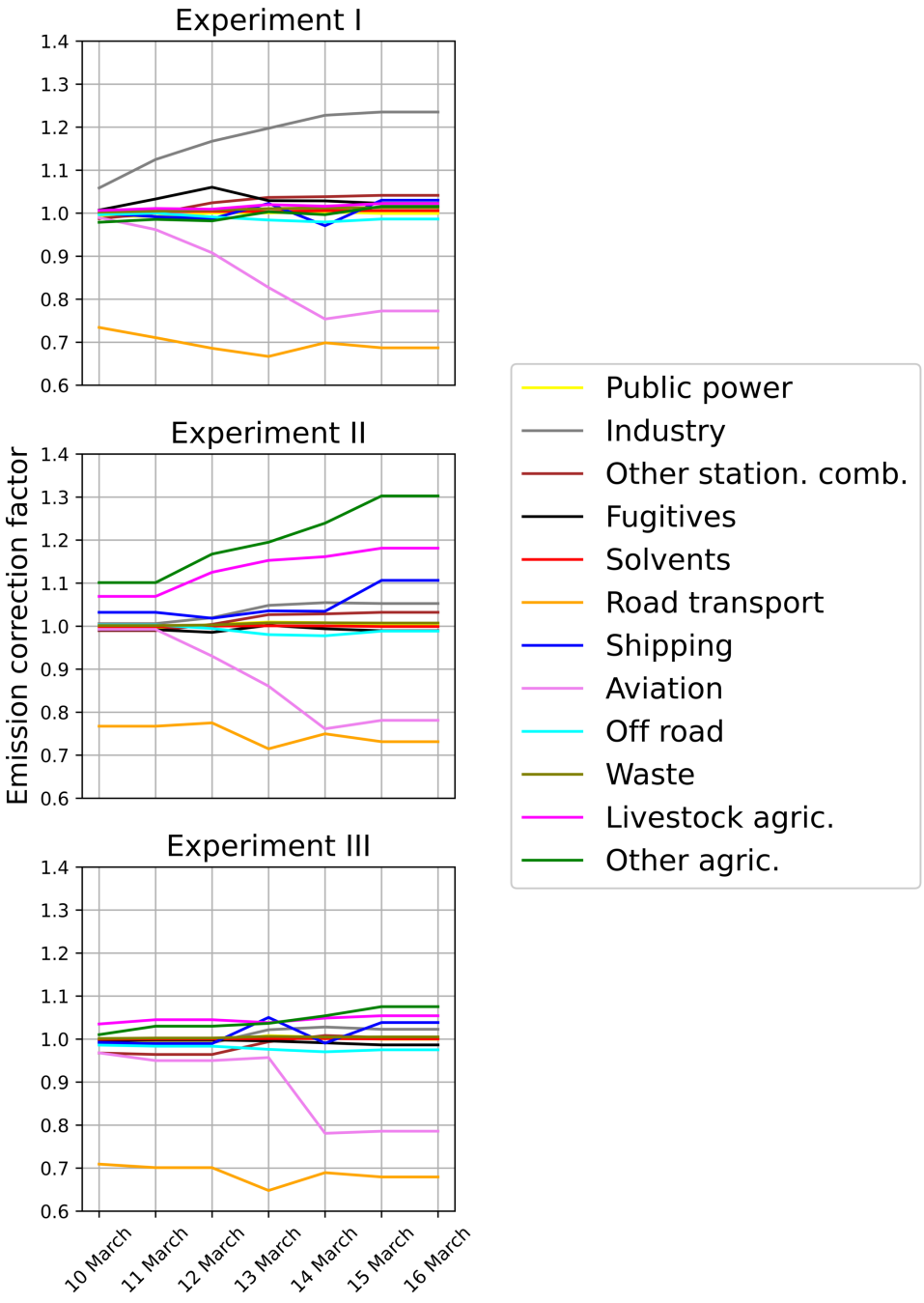

Figure 6.3.: Spatially averaged emission correction factors in experiment I, experiment II and experiment III for each day of the simulation period. The true emission correction factors according to the emission perturbation in the nature runs in experiment I are 2 for the industry, 0.5 for the road transport and 1 for all other sectors. The true correction factors in experiment II are 2 for the non-livestock agriculture, 0.5 for the road transport and 1 for all other sectors. The true correction factors in experiment III are 1.3 for the non-livestock agriculture, 0.5 for the road transport and 1 for all other sectors.

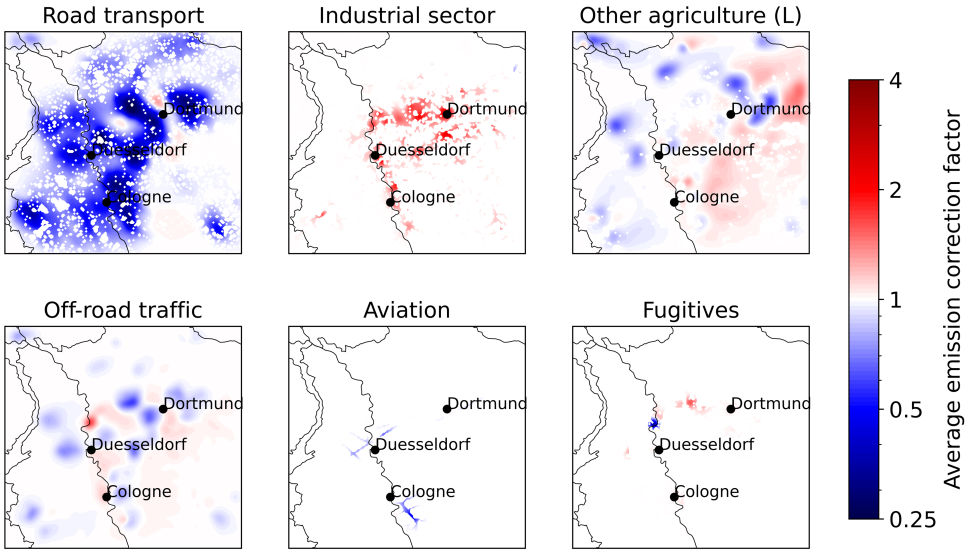

Figure 6.4.: Spatial distribution of emission correction factors averaged over the simulation period in experiment I. The true emission correction factors according to the perturbation in the nature runs in experiment I are 2 for the industry, 0.5 for the road transport and 1 for all other sectors. Areas with white colour mean either an emission correction factor of 1 or no emissions.

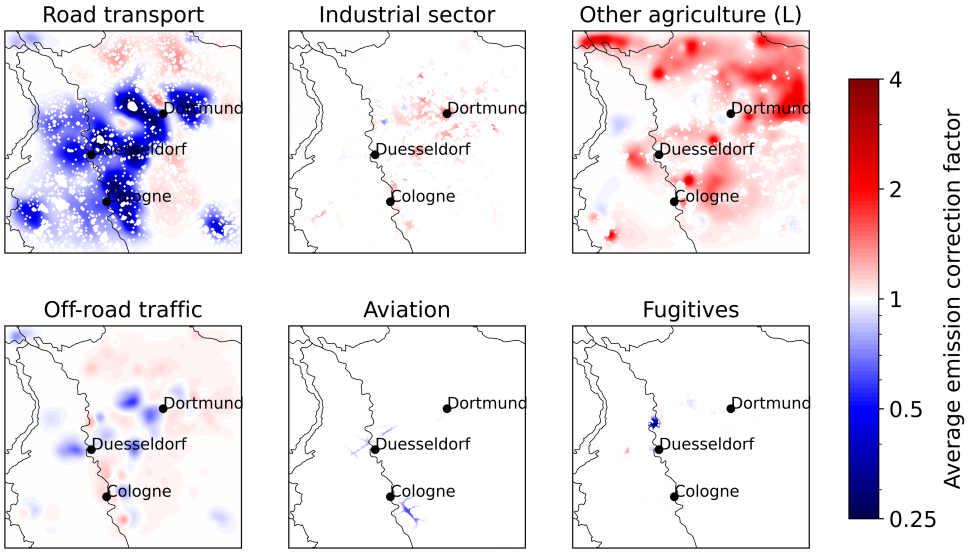

Figure 6.5.: Spatial distribution of emission correction factors averaged over the simulation period in experiment II. The true emission correction factors according to the perturbation in the nature runs in experiment II are 2 for the non-livestock agriculture, 0.5 for the road transport and 1 for all other sectors. Areas with white colour mean either an emission correction factor of 1 or no emissions.

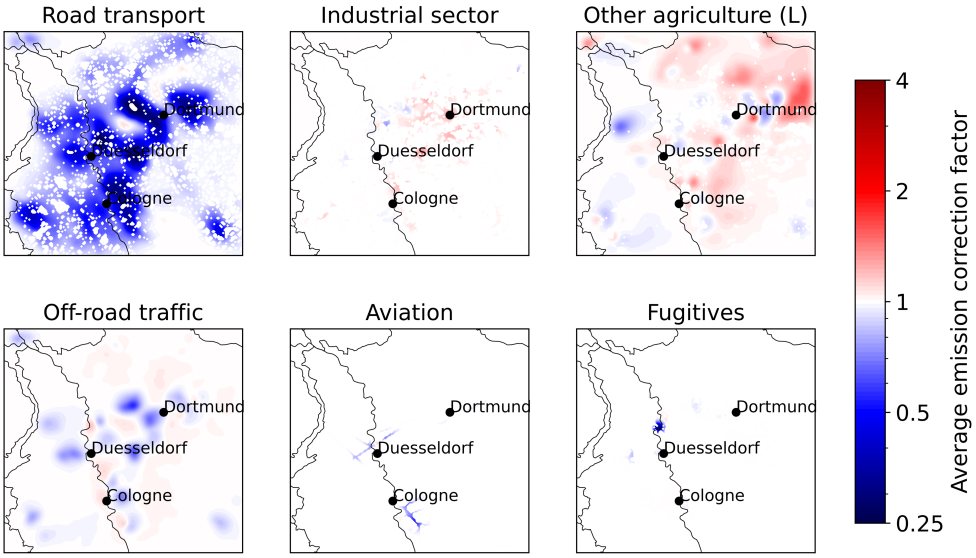

Figure 6.6.: Spatial distribution of emission correction factors averaged over the simulation period in experiment III. The true emission correction factors according to the perturbation in the nature runs in experiment III are 1.3 for the non-livestock agriculture, 0.5 for the road transport and 1 for all other sectors. Areas with white colour mean either an emission correction factor of 1 or no emissions.

The negative emission perturbations of the road transport sector also affect the emission correction of sectors with low NO_x emissions. Fig. 6.4, 6.5 and 6.6 show that this applies to the off-road traffic, aviation and fugitives sectors. The reason for this is that sources of these sectors are located near roads where the emissions are reduced due to the perturbation. This reduction is wrongly attributed also to the off-road traffic, aviation and fugitives sectors by the data assimilation system. The aviation sector is the most affected sector due to the spatial proximity of important airports (Cologne-Bonn and Düsseldorf) with large local NO_x emissions and strong road transport emissions. As indicated in Fig. 6.3, in all experiments, the average correction factor for the aviation sector is below 0.8 from 14 March to 16 March. In the case of the fugitives sector, the emission corrections are affected by reduced aerosol emissions of the road transport sector. In an area north of Düsseldorf (Fig. 6.4, 6.5 and 6.6), this leads to negative emission corrections of this sector in all experiments.

With regard to the industry sector, Fig. 6.4 and 6.3 show that the data assimilation system is able to reflect perturbations of industrial emissions. In experiment I, emission correction factors above 1 for the industry sector are determined in the area between the cities Cologne, Düsseldorf and Dortmund. In this area, also strong road transport emissions occur (Fig. 5.4) which are reduced in experiment I. This means that the model is able to distinguish between emission perturbations of the industry and the road transport sector. This is due to the ability of the model to exploit the

different chemical compositions of the sectors. The industry sector is the dominant SO_x and aerosol emitter in the domain, whereas the road transport emissions contain less aerosols and nearly no SO_x. As a consequence, observational increments (Eq. 2.1) of SO_2, PM_{10} and $PM_{2.5}$ are correctly attributed to the industry sector. However, the average emission correction factor for the industry sector is only 1.06 on 10 March (Fig. 6.3), which is small compared to the (true) perturbation factor of 2 for this sector. It is increased successively by the data assimilation system to reach the maximum of 1.23 on 15 March 2016. The slow increase of the emission correction factors for the industry sector can be explained by the relatively small sensitivity of the costs (Eq. 4.1) to SO_x and aerosol emissions compared to NO_x emissions. Since the emission correction factors for the industry sector are mainly related to the relatively small observational increments (Eq. 2.1) of SO_2, PM_{10} and $PM_{2.5}$, the emission correction factors for this sector are only increased slowly. In comparison, the emission correction factors for the road transport sector are mainly related to observational increments of NO_2 and O_3, which are very sensitive to NO_x emissions. This leads to strong emission corrections for the road transport sector already on 10 March.

Perturbations of non-livestock agricultural emissions can affect corrections for the industry sector. In experiment II with a perturbation factor of 2 for the non-livestock agricultural sector, emission correction factors of up to 1.06 on average are determined for the industry sector. The reasons for this are the NO_x emissions of the non-livestock agriculture sector (Fig. 5.3) and the relatively flat diurnal emission cycle of both sectors (Fig. 5.8).

The data assimilation system is also able to reflect perturbations of the non-livestock agriculture sector. In experiment II with an emission perturbation factor of 2 for the non-livestock agriculture sector, the agricultural emissions are increased by the optimisation. This applies also to areas in the proximity to large cities like Cologne and Düsseldorf with high emissions of other sectors (Fig. 6.5). In experiment III with a perturbation factor of only 1.3 for the non-livestock sector, emission increases on spatial average are determined for this sector. A maximum emission correction factor of 1.08 is reached on 15 March (Fig. 6.3). Like for the industrial emission corrections in experiment I, the agriculture emission corrections in experiments II and III are successively increased during the simulation period (Fig. 6.3). In experiment II, a maximum of 1.3 is reached, whereas the highest value in experiment III is 1.08. The main reason for the slow increase of the emission corrections for the non-livestock agriculture sector is the relatively low sensitivity of the costs (Eq. 4.1) to NH_3 emissions caused by the agricultural sectors.

The ability of the model to separate between the road transport and the non-livestock agriculture sector has two main reasons. Firstly, the large NH_3 emissions of the non-livestock agriculture sector cause increased aerosol concentrations, which are indirectly observed as enhanced PM_{10} and $PM_{2.5}$ concentrations. The adjoint model is able to project the resulting observation - model discrepancy to the main cause, which is the

NH_3 emissions. Furthermore, the difference between the diurnal emission profile of the agriculture and the road traffic assumed by the inventory is exploited. In contrast to the road transport sector, the agriculture sector has significant emissions at night (Fig. 5.8). Thus, in experiments II and III, the observed NO_2 concentrations between 0 UTC and 5 UTC are higher than the simulated NO_2 concentrations, which leads to observational increments (Eq. 2.1) at night. These observational increments are correctly attributed to emissions from the non-livestock agriculture sector by the data assimilation system.

The corrections for the livestock agriculture and shipping sectors are affected by the perturbation of the non-livestock agriculture emissions in experiments II and III. The emission corrections for the livestock agriculture sector have the highest correlation with the corrections for the non-livestock agriculture sector. The maximum spatially averaged emission correction factor for the livestock agriculture sector is 1.18 in experiment II and 1.06 in experiment III (Fig. 6.3). The reasons are that both sectors emit a high amount of NH_3 (Fig. 5.2) and that there is no spatial separation between the sectors.

The spatially averaged emission correction factors for the sector of other stationary combustion are up to 1.05 in experiment I and up to 1.03 in experiment II (Fig. 6.3), although the emissions were not perturbed. For this sector, the data assimilation system determines moderate local emission increases and decreases in all experiments (not shown in the figures). This is also valid for the public power sector. This shows that the emission corrections for both the public power sector and the other stationary combustion sector can be affected by perturbations of road transport, industry and non-livestock agriculture emissions.

Small corrections are determined for the solvents and waste sectors, to which also no emission perturbations are applied in the experiments. The spatially averaged emission correction factors of these sectors are between 0.98 and 1.02 for all experiments on all days of the simulation period (Fig. 6.3). The reason for this is that both sectors have only low emissions of observed species in the NRW domain. Solvents emit a large amount of non-methane NMVOCs which are not included in the synthetic observations. In contrast, the contribution of the solvents sector to the total emissions of NO_x, CO, SO_x, NH_3, PM_{10} and $PM_{2.5}$ in the domain is less than 10% (Fig. 5.2). The contribution of the waste sector to the total emissions in the domain is less than 1% for NO_x, CO, SO_x, NMVOC and NH_3, about 5% for PM_{10} and about 10% for $PM_{2.5}$ (Fig. 5.2).

Fig. 6.7 shows the observational costs (Eq. 3.4) of the background and the analysis during the simulation period for each measured species. The background costs before the optimisation as well as the analysis costs after the optimisation are depicted. A reduction of the analysis costs compared to the background costs corresponds to an improvement of the agreement between simulated and observed concentrations by the sector specific emission optimisation. NO_2 and O_3 are the species with the highest

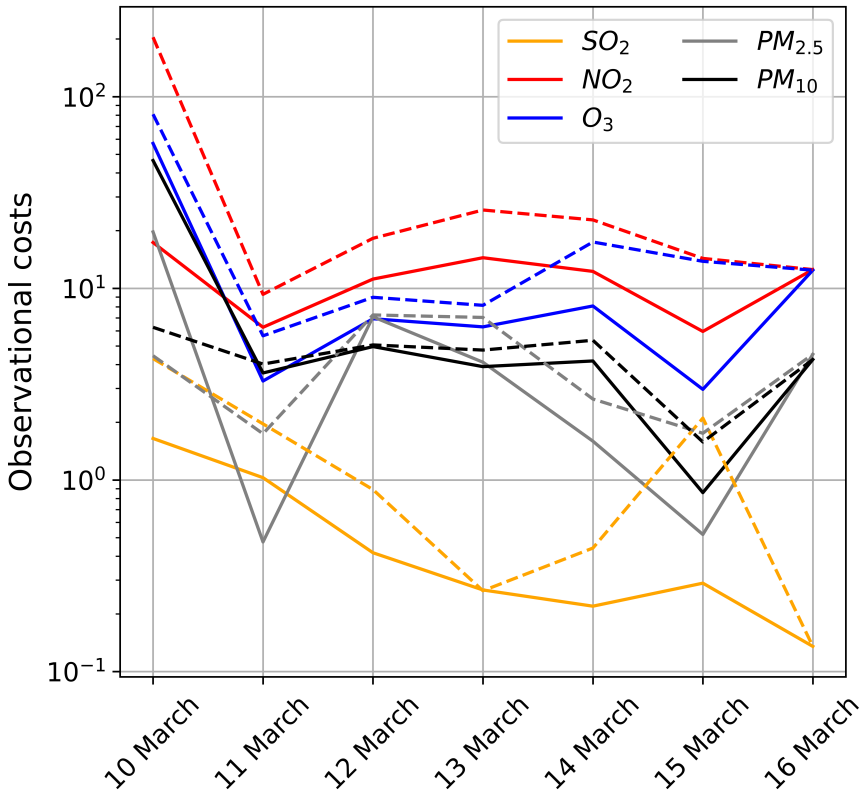

Figure 6.7.: Observational costs (Eq. 2.6) of the observed species during the simulation period in experiment I with emission perturbation factors of 2 for the industry and 0.5 for the road transport sector. Analysis costs are depicted as solid lines, background costs as dashed lines. If the analysis costs are lower than the background costs of a species, the sector specific optimisation of emissions has improved the agreement between simulated and observed concentrations of this species.

background costs. This reflects the perturbation of the NO_x emissions in this experiment, which causes changes in the NO_2 and O_3 concentrations. In addition, the good observational situation for both species leads to high observational costs. Compared to the effect of the perturbed NO_x emissions on the NO_2 and O_3 concentrations, the impact of the perturbed aerosol emissions on the PM_{10} and $PM_{2.5}$ concentrations is lower and thus causes lower observational costs. The SO_2 costs are low due to the observational situation (Section 5.4).

In most cases, the analysis costs are lower than the background costs (Fig. 6.7), which means that the costs are reduced by the sector specific emission optimisation. The NO_2 costs are decreased by 90% on the first day, due to the strong emission corrections in the road transport sector. In contrast, the PM_{10} and $PM_{2.5}$ costs on 10 March are increased by the emission optimisation. The reason for this is that the total aerosol

emissions are decreased by the optimisation, whereas the emission perturbations in the nature run lead to an increase in total aerosol emissions (Fig. 6.2). This is due to a weak increase in industrial emissions by the emission optimisation on 10 March (average emission correction factor of 1.06), while the road transport emissions are strongly reduced to an average emission correction factor of 0.73 (Fig. 6.3). The O_3 and SO_2 costs are reduced, but less than the NO_2 costs. The weaker cost reduction can be explained by the weak corrections of the industrial emissions on 10 March (Fig. 6.3). The average emission correction factor increases until 14 March 2016, as discussed before. The optimisation leads to a cost reduction, i.e. lower analysis than background costs, for all species from 11 March to 15 March 2016. This corresponds to an improvement of the simulated concentrations. An exception with no difference between background and analysis costs is the 16 March 2016. Note that the variations in the cost reduction also depend on the meteorological conditions, especially on the wind direction and wind speed. In experiments II and III, the observational costs behave similarly (Fig. A.1 and A.2 in the appendix).

6.3. Impact of the anisotropic diffusion operator on the corrections of the road transport sector

In the previously discussed one-week simulation, the anisotropic diffusion operator (Section 4.3) was applied to the emission correction factors for the road transport sector and to the related cost function gradient ($\nabla_{\mathbf{w},road}J$). In order to investigate the effect of the anisotropic diffusion, a second data assimilation cycle was performed for the 10 March 2016 without using the anisotropic diffusion operator. Fig. 6.8 shows the gradient $\nabla_{\mathbf{w},road}J$ of the first iteration for both data assimilation runs. The emission correction factors of the analysis iteration are depicted in Fig. 6.9. Note that the gradient in general dictates the strength and direction of emission corrections and that a negative gradient is related to positive emission corrections.

The maximum of the gradient ($\nabla_{\mathbf{w},road}J$) in the domain is similar in both data assimilation runs with a value of about 8. This is a consequence of the normalisation by the matrix $\mathbf{\Lambda}$ or $\mathbf{\Lambda}_{roadtr.}$, respectively (Sections 3.2 and 4.3). But the spatial spread is increased by the anisotropic diffusion. In the experiment with isotropic diffusion, the gradient field around Cologne and Dortmund contains clear lines and spots along large roads and highways and at observation stations. In contrast, when selecting the anisotropic diffusion along roads, a strong gradient stretches across large parts of the southern Rhineland and of the eastern Ruhr area. The spatial distributions of the analysis emission correction factors behave similarly in both experiments, but with a significantly weaker decrease in maximum when applying the anisotropic diffusion. The total emission correction for road transport is stronger when the anisotropic diffusion operator is applied. The spatially averaged emission correction factors are 0.73 in the experiment with application of anisotropic diffusion and 0.76 in the experiment without

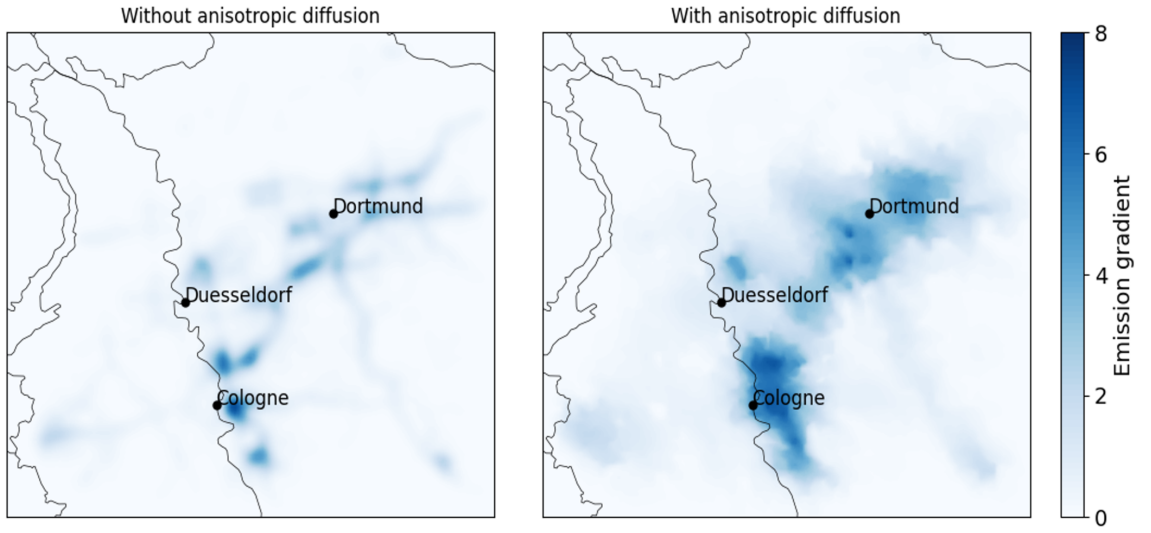

Figure 6.8.: Gradient with respect to the emission corrections for the road transport sector ($\nabla_{\mathbf{w},road}J$) without (left) and with (right) anisotropic diffusion along roads (Section 4.3) in experiment I on 10 March 2016. The emission perturbation factors in the nature run of this experiment are 2 for the industry sector and 0.5 for the road transport sector.

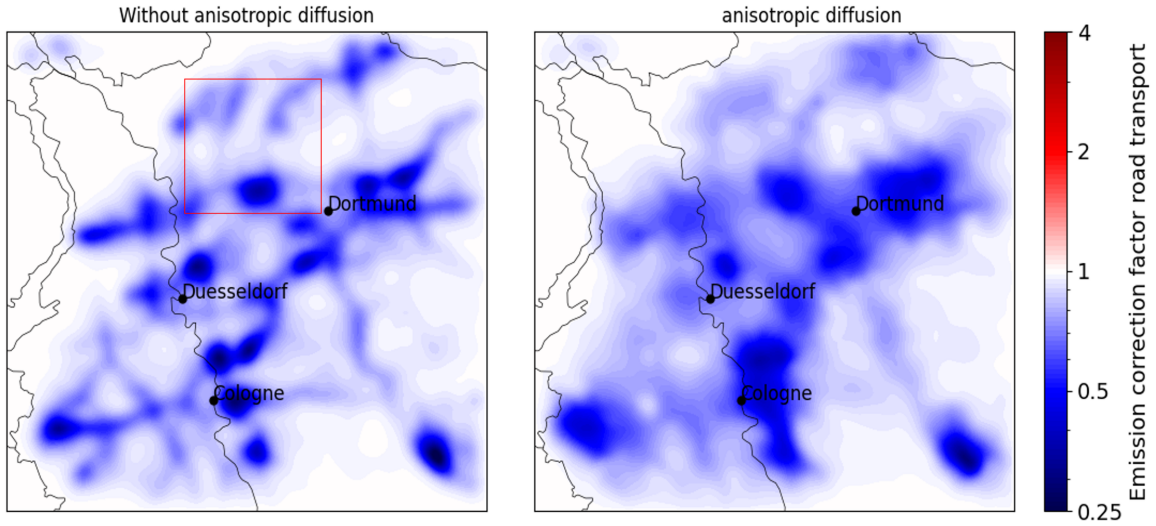

Figure 6.9.: Emission correction factors of the road transport sector without (left) and with (right) anisotropic diffusion along roads (Section 4.3) in experiment I on 10 March 2016. The emission perturbation factors in the nature run of this experiment are 2 for the industry sector and 0.5 for the road transport sector. The area shown in Fig. 6.10 is marked as a red rectangle in this figure.

application of anisotropic diffusion. In contrast, the minimum emission correction factor in the domain is lower in the experiment without application of anisotropic diffusion (0.22) than in the experiment with application of anisotropic diffusion (0.29).

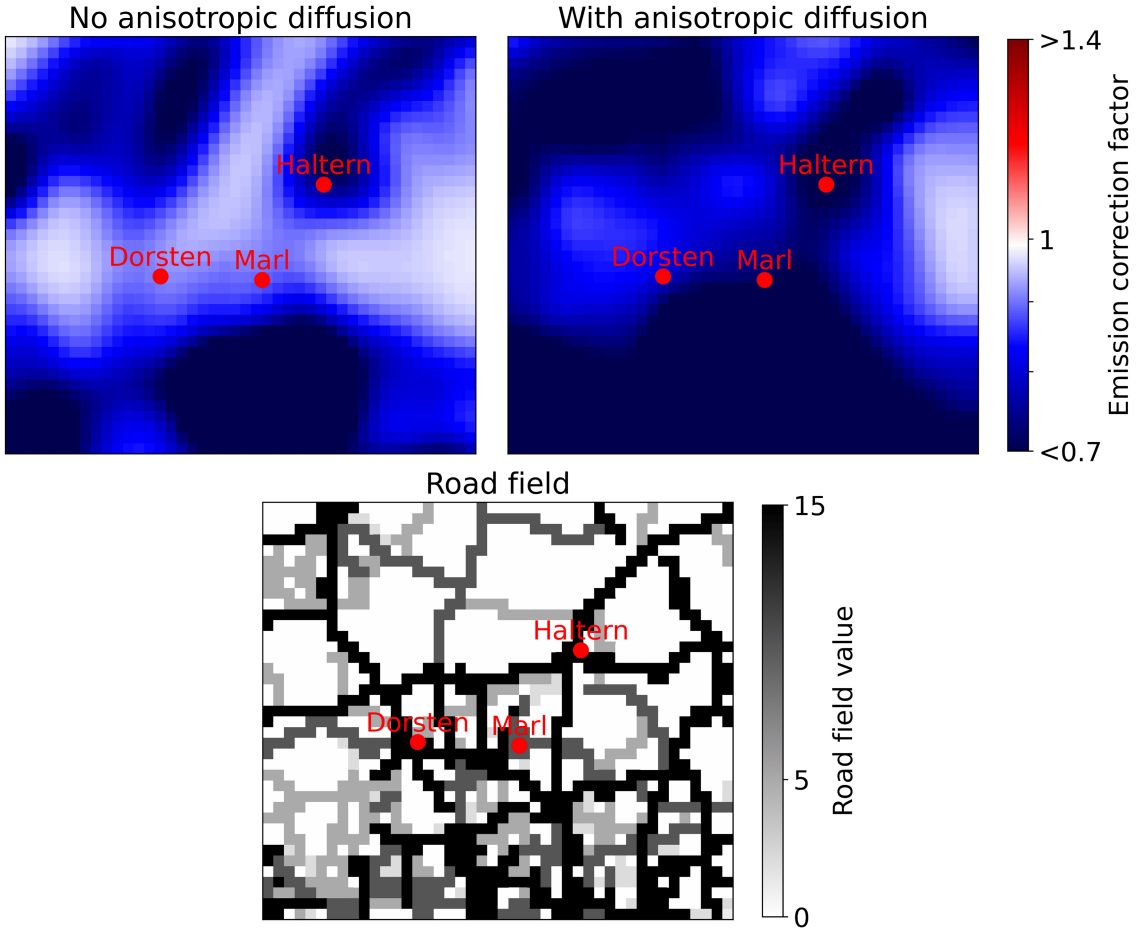

Figure 6.10.: Emission correction factors of the road transport sector without (left) and with (right) anisotropic diffusion along roads on 10 March 2016. In addition, the road field used for the anisotropic diffusion along roads (Section 4.3) is depicted. A 45 km × 40 km area in the NRW domain is shown, which is marked in Fig. 6.9. The emission perturbation factors in the nature run of this experiment are 2 for the industry sector and 0.5 for the road transport sector.

Fig. 6.10 shows the emission correction factors for road transport and the road field ρ (Section 4.3) for a 45 km × 40 km area contained in the NRW domain and located northwest of Dortmund. The area is marked in Fig. 6.9. The road density in this area is relatively low, e.g. compared to the area between the cities Cologne, Düsseldorf and Dortmund. In contrast to areas with a high road density, structures of the road field can be partially identified in the emission correction factors for the area depicted

in Fig. 6.10. One example is the road between Marl and Haltern with a road field value of 15, where the emission corrections are significantly stronger in the experiment with application of anisotropic diffusion than in the experiment without application of anisotropic diffusion. Another example is a road north of Haltern running in western direction, where emission corrections can be observed in the experiment with anisotropic diffusion, but not in the experiment without anisotropic diffusion.

To conclude, in regions with a dense road network such as between the cities Cologne, Düsseldorf and Dortmund, the main impact of the anisotropic diffusion operator is a wider distribution of the road transport emission corrections. In the case of this study, this leads to a slightly higher total emission correction in the model domain. Only in regions characterised by a low road density, structures of the road field used for the anisotropic diffusion along roads can be observed in the spatial distribution of the emission correction factors.

6.4. Conclusions of the identical twin experiments

In the scenarios of the identical twin experiments, the new approach for sector specific emission optimisation implemented in the EURAD-IM (Chapter 4) is able to reflect emission changes of individual GNFR sectors. Distinction between sectors is possible, if there is a spatial separation between their sources or if there are substantial differences between the diurnal cycle or the chemical composition of their emissions. As an additional condition, a high emission strength of the sectors is necessary. Furthermore, the pollutants emitted from the sources need to be transported to an observation station, which is a general requirement for emission corrections and depends on both the observational situation and the wind direction. In the experiments in this chapter, a distinction between decreased road transport and increased industrial emissions is achieved for large parts of the NRW domain. There is also an efficient distinction between non-livestock agricultural and road transport emissions, especially in the case of a high emission perturbation. With the optimisation of the emission correction factors, also the observational costs, especially of NO_2 and O_3, are strongly reduced. The cost reduction reflects a substantial improvement of the simulated concentrations by the sector specific emission optimisation. An additional simulation shows an effective increase of the spatial correlation of road transport emission corrections through the application of the anisotropic diffusion operator, which is part of the new development (Section 4).

Overall, the experiments demonstrate the ability of the presented approach for sector specific emission optimisation. The method of identical twin experiments is suitible for this purpose, since it allows for a comparison between the emission corrections determined in the optimisation and the emission perturbations applied in the nature runs. Another aspect of the presented identical twin experiments is the idealised conditions with zero model and observation uncertainties. In the following, three important

sources of uncertainty in the sector specific emission corrections are described that are not included in the identical twin experiments:

- The synthetic observations in experiments I, II and III are "perfect". In contrast, real observations of trace gases and aerosols contain statistical and systematic measurement errors, as described in Section 5.4. In addition, real observations are not fully representative of the grid cell of the model domain.
- Predictions of atmospheric chemistry transport models contain uncertainties e.g. due to their incomplete chemistry mechanism, their aerosol model, their diffusion model and their meteorological input data (Section 5.2). In identical twin experiments, these model uncertainties are not included since the synthetic observations are based on the same model as the data assimilation. However, in an assimilation of real observations, they might play an important role.
- In the presented experiments, the sector specific diurnal profiles and chemical compositions in the emission data (Section 5.3) are assumed to be valid. Real emissions can significantly deviate from these assumptions. This source of uncertainty is of particular importance for the separation between the sectors.

All these aspects affect the sector specific emission optimisation in the case of real observations. Thus, also studies based on real observations, as presented in the next chapter, are necessary to investigate the behaviour, potential and limits in presence of the uncertainties mentioned above.

7. Simulation using ground station observations in North Rhine-Westphalia

The following chapter presents a simulation in which real observations of ground stations in NRW are assimilated. As described in the last chapter, the aim is to consider a real emission scenario where all sources of uncertainties are included. The uncertainties are due to the observations, the model and its input as well as the assumptions on the height profiles and diurnal cycles of the emissions from the Gridding Nomenclature for Reporting (GNFR) sectors. In contrast to the identical twin experiments discussed in Chapter 6, the true emission corrections for the sectors are unknown. Thus, the studies in the two chapters should be regarded as complementary. This chapter includes a brief description of the simulation configuration (Section 7.1), an investigation of the sector specific emission corrections and the related cost function gradients (Section 7.2) as well as of the agreement between observed and simulated concentrations (Section 7.3), and ends with conclusions from the results (Section 7.4).

7.1. Configuration of the simulation

The configuration of this simulation is similar to the identical twin experiments except for the usage of real ground observations, described in Section 5.4. Furthermore, the simulation period is two weeks instead of one week, in order to investigate the behaviour of the sector specific data assimilation under two different meteorological conditions. The change in wind direction from 17 March (between east and northeast) to 18 March 2016 (between north and northwest) has a crucial impact on the transport of atmospheric pollutants (Section 6.1).

Regarding the configuration of the data assimilation system in this study, both the emission correction factors and the initial concentrations of the pollutants are optimised in contrast to the studies discussed in Chapter 6. In addition, a limit value for emission correction factors is set for all grid cells and sectors: $e \leq 15$. This avoids unrealistically high emission corrections.

The results of the sector specific emission optimisation (Chapter 4) are compared to a reference simulation using the current 4D-Var data assimilation system of the EURAD-IM (Section 3.2), also referred to as "4D-VarREF" in this chapter. The

4D-VarREF simulation is performed for the same period and region and uses the same input data. The differences of the two simulations in the emission correction factors and in the time series of the simulated concentrations will be discussed in the following sections.

7.2. Emission correction factors and gradients

As mentioned in Section 3.2, emission correction factors determined by 4D-Var data assimilation are a consequence of the cost function gradient with respect to the emission correction factors ($\nabla_{\mathbf{w}} J$). In the sector specific 4D-Var data assimilation, a gradient field for each GNFR sector (Table 5.1) exists. The gradients of the species are fully correlated within the sectors in the minimisation procedure (Section 4.2). Therefore, the gradients after minimisation contain no information about the single species. In contrast, the gradients in the adjoint run, which is performed before the minimisation procedure (Fig. 3.2), are species-dependent. For this reason, the sector- and species-dependent gradients after the first adjoint run of the assimilation cycle are discussed in the following.

Figure 7.1 displays the relative contributions of NO, NO_2, SO_2, NH_3 and aerosols with a diameter between 2.5 and 10 µm (PM_{coarse}) to the total gradient of each sector for the simulation period. For each combination of species l and sector s, the absolute values of the respective gradient are summed up over the grid cells and the days of the simulation period. The share $\mathcal{S}_{l,s}$ of species l to the total gradient of a sector s is determined using the following formula:

$$\mathcal{S}_{l,s} = \frac{\sum_{i,j,d} |\nabla J_{u(i,j,l,s,d)|}}{\sum_{i,j,d,l} |\nabla J_{u(i,j,l,s,d)|}} \ . \tag{7.1}$$

$J_{\mathbf{u}}$ is the cost function gradient after the first adjoint run of the assimilation cycle with respect to the emission correction factors. i and j denote the horizontal indices of a grid cell and d denotes the index of the day. For road transport, shipping, aviation and off-road traffic, the NO gradient is most important, followed by the NO_2 gradient. The reason for this is that the observational increments (Eq. 2.1) of NO_2 are mostly traced back to NO emissions. The gradients of other species have only a small impact on the emission corrections. The gradients of the public power, the industry and the stationary combustion sectors are also dominated by NO. For public power and industry, SO_2 plays a minor role, despite their high SO_x emissions. This is due to the low number of SO_2 observations compared to the number of NO_2 observations. This leads to a low sensitivity of the cost function value (J) to SO_x emissions in this simulation. NH_3 is the most important emitted species for emission corrections of the livestock agriculture sector, although no NH_3 observations are assimilated. The reason for this is that NH_3 takes part in the formation of secondary aerosols. Thus, NH_3 emissions have an impact on the aerosol concentrations observed in the NRW

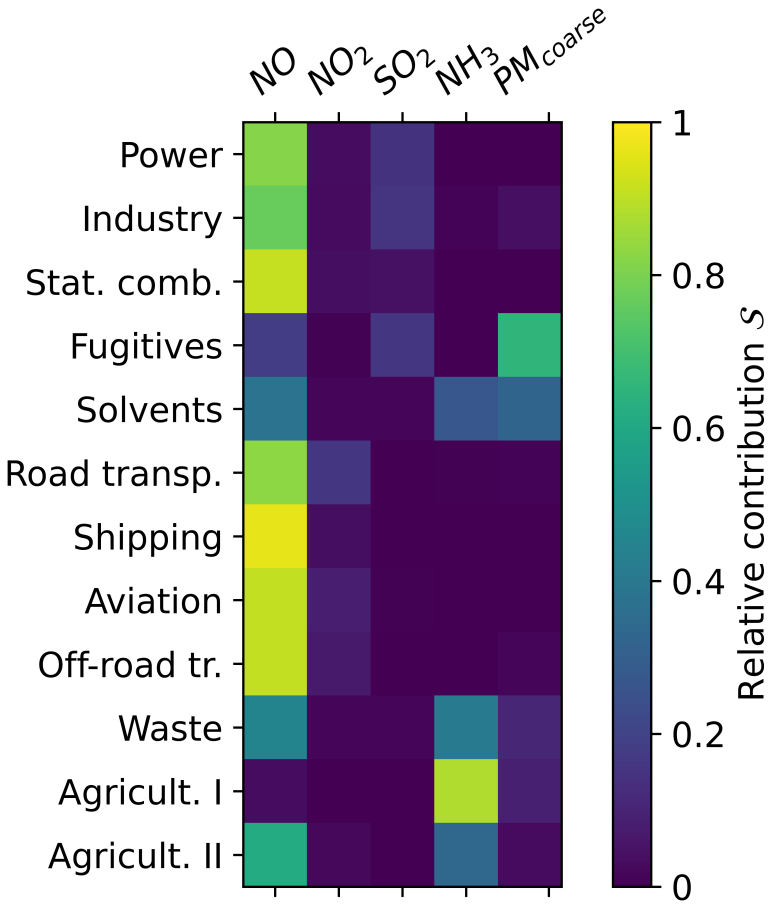

Figure 7.1.: Contributions of the NO, NO_2, SO_2, NH_3 and PM_{coarse} (aerosols with a diameter between 2.5 and 10 µm) gradients to the total gradient of the GNFR sectors. The absolute values of the sectors' and species' gradients after the first adjoint model run are summed up over all grid cells of the model domain and over all days of the simulation period. The values are normalised such that the sum of the gradients of all species is 1 for each sector (Eq. 7.1).

domain. In the case of the non-livestock agriculture sector, both NO and NH_3 have an impact on the gradient. The PM_{coarse} gradient is only important for the sectors of the fugitives and the solvents, due to the low gradient of the other species. Overall, Fig. 7.1 indicates a dominant impact of NO_2 observations on the determined emission corrections in this simulation.

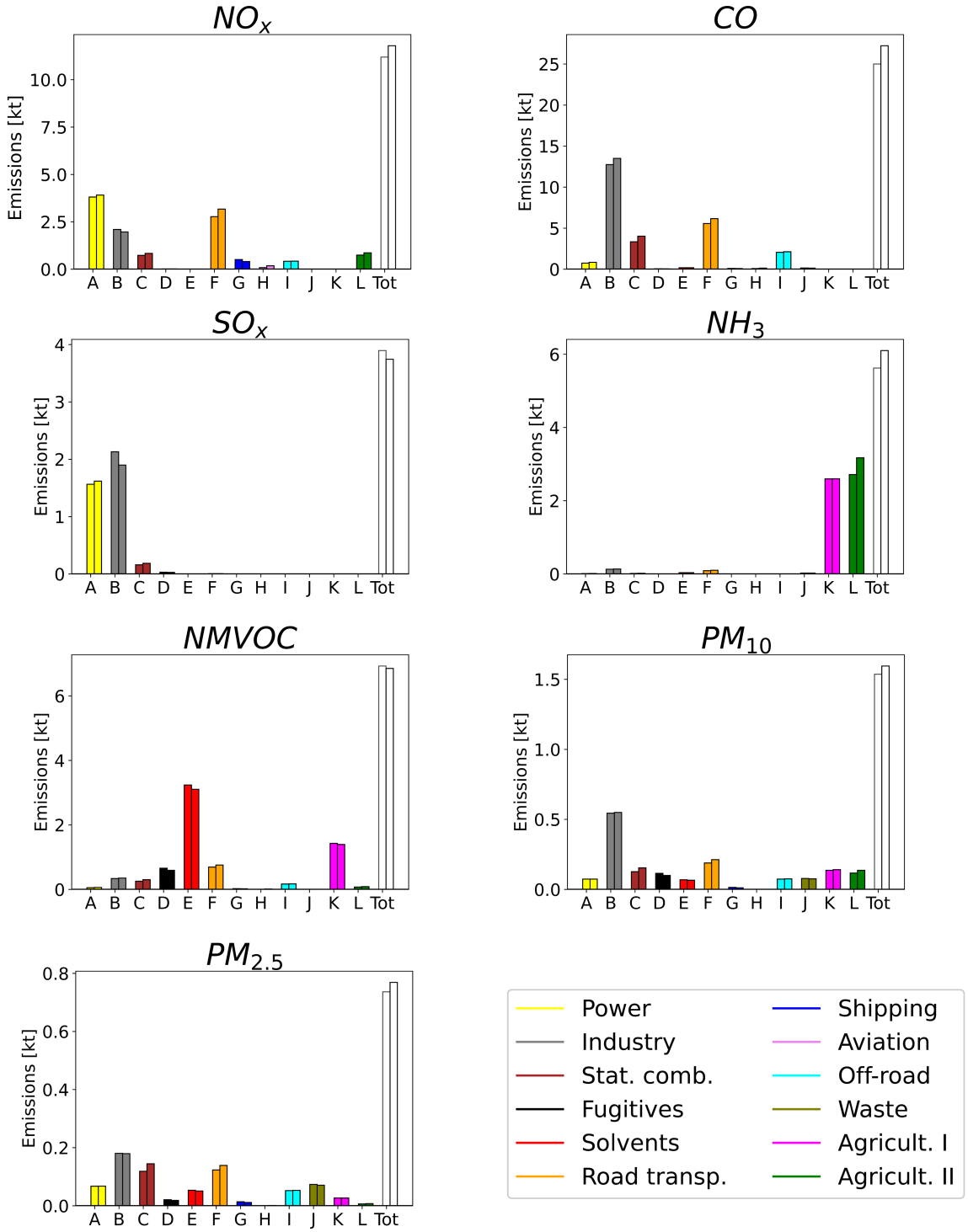

Figure 7.2.: Bar charts of the NO_x, SO_x, CO, NH_3, PM_{10} and $PM_{2.5}$ emissions in the NRW domain for the simulation period. The inventory emissions for each GNFR sector (Table 5.1) are depicted as first bar, the corrected emissions as second bar.

The total emissions of CO, NO_x, SO_x, NH_3, non-methane volatile organic compounds (NMVOC), PM_{10} and $PM_{2.5}$ from each GNFR sector (Table 5.1) during the simulation period are shown in Fig. 7.2. Both the inventory emissions and the corrected emissions after the sector specific optimisation are depicted. The total emission corrections for the species in the NRW domain for the whole simulation period are relatively small. The relative emission correction summed over all sectors is strongest for CO with an enhancement of 9%. The main contributors to the increase of CO emissions are the public power sector with 13%, the road transport sector with 11% and the other stationary combustion sector with 20% relative increase. The NO_x emissions are increased by 5%, with the highest increases in the road transport sector (+14%) and the non-livestock agriculture sector (16%). The SO_x emissions are decreased by 4%, with a slight increase in the public power sector (+4%) and a moderate decrease in the industry sector (-11%). The NH_3 emissions are increased by 9% due to the emission corrections in the non-livestock agriculture sector (+17% NH_3). The solvents sector (-4% NMVOC) contributes to a slight decrease of the total NMVOC emissions by 1%. Both the total PM_{10} and $PM_{2.5}$ emissions are increased by 4%, which is mainly caused by the emission corrections for other stationary combustion (+22% PM_{10} and $PM_{2.5}$) and road transport (+13% PM_{10} and $PM_{2.5}$).

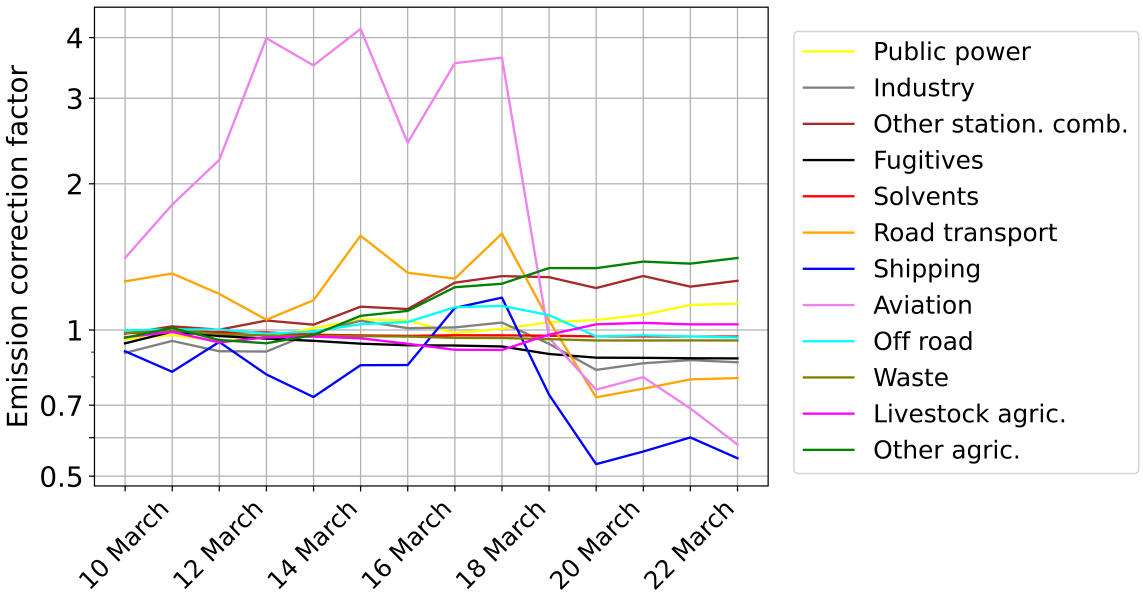

Figure 7.3.: Time evolution of the spatially averaged sector specific emission correction factors for the GNFR sectors (Table 5.1) during the simulation period.

The time evolution of the spatially averaged sector specific emission correction factors is shown in Fig. 7.3. For this, the ratio of the total optimised emissions and the total inventory emissions in the domain is calculated for each sector and each day of the simulation period.

Until the 18 March 2016, the road transport emissions are increased with a maximum correction factor of 1.5 on 18 March, but strongly decrease after the change of the main wind direction and reach a minimum value of 0.75 on 20 March. A change in wind direction generally results in an impact of other emission sources on the observed concentrations at the ground observation stations. In this case, this means that the data assimilation system tends to increase the inventory emissions east of the observation stations, while it tends to decrease the emissions west of the stations due to the observed concentrations. The emission corrections for the road transport sector are mainly driven by observational increments of NO_2 because of the high NO_x emissions from road transport. The shipping sector shows a similar behaviour, but the average emission correction factors are around 1.0 until 18 March and are strongly decreased below 0.6 from 20 to 23 March.

For the aviation sector, the emission corrections are strongly increased during the first four days of the simulation period. A maximum emission correction factor of 4.2 is reached on 15 March, which is by far higher than that of the other sectors. A probable reason for this is that increased NO_x emissions from road transport affect the correction of aviation emissions near Cologne/Bonn Airport. A similar behaviour could be observed in the identical twin experiments in which the aviation emissions were decreased by the data assimilation system although they were not perturbed (Fig. 6.3). Corresponding figures depicting the emission correction factors, gradients and NO_2 time series from 11 March to 13 March in the vicinity of the airport can be found in the appendix (Section B.4). After the change of the main wind direction on 18 March, the emission correction factors show values below 1.

The power, other stationary combustion and non-livestock agriculture sector show a slow, but relatively continuous increase of emission corrections during the simulation period, with maximum values between 1.2 and 1.4 on 23 March. In contrast to the other sectors, the values do not decrease on 18 March and 19 March.

The average emission correction factor for the industry sector is 0.9 on the first day of the simulation period. The value slightly increases until it reaches a maximum on 15 March (1.1) and decreases from 18 March to 20 March to a minimum (0.85).

The emissions of fugitives are continuously decreased by the sector specific optimisation to a value of 0.9 on 23 March. The fugitives sector mainly emits NMVOCs, which are not observed, and aerosols (Fig. 5.2). Hence, the decrease of the emissions from the fugitives sector is corresponds to a general decrease of PM_{10} and $PM_{2.5}$ emissions in this simulation. The reason for this decrease are the observed aerosol concentrations, which are mostly above the simulated aerosol concentrations. This is further discussed in Subsection 7.3.3.

The emissions of the other sectors are only slightly corrected on spatial average. Their average correction factors are between 0.8 and 1.2 on each day of the simulation

period. In the case of the waste and the off-road sector, this is due to relatively low emissions of all species compared to the emissions of other sectors. In contrast, the solvents sector is the strongest anthropogenic NMVOC emitter. However, no NMVOC observations are assimilated, so that there are only small corrections for the solvents sector. The emission corrections for the livestock agriculture sector are approximately 1 on spatial average, but moderate local emission increases and decreases are determined for this sector (Fig. 7.5).

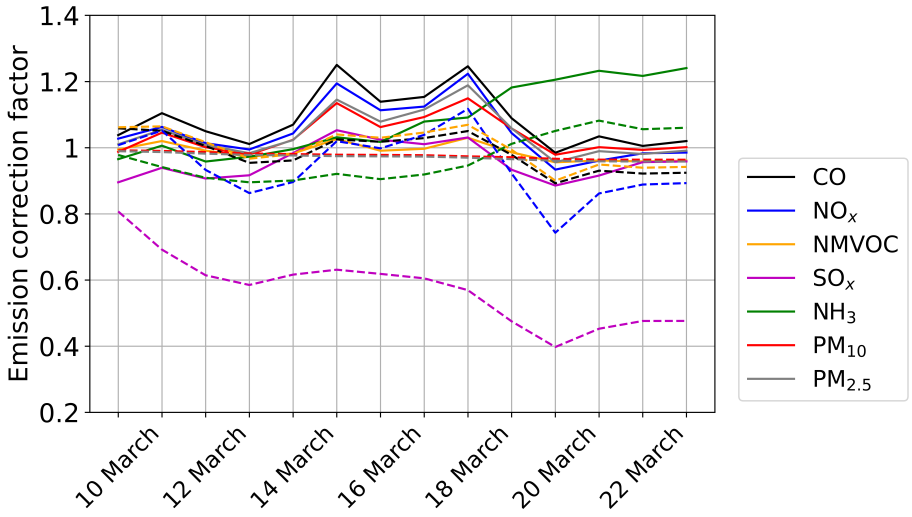

Figure 7.4.: Comparison of the time evolution of the emission correction factors for NO_x, SO_x, CO, NH_3, PM_{10} and $PM_{2.5}$, averaged over the model domain and the sectors. The sector specific optimisation (Chapter 4) is represented by solid lines, the optimisation by the current data assimilation system of the EURAD-IM (Section 3.2) by dashed lines.

The temporal evolution of the spatially averaged emission correction factors for NO_x, SO_x, CO, NH_3, PM_{10} and $PM_{2.5}$, averaged over the emission sectors, is shown in Fig. 7.4. The sector specific optimisation (Chapter 4) is compared to the optimisation by the current data assimilation system of the EURAD-IM (Section 3.2), denoted as "4D-VarREF" in the following. The total emission correction factors e_L of the species (Fig. 7.4) are calculated using the following formula:

$$e_L = \frac{\sum_{i,j,k,l,s} e_{i,j,l,s} \times \epsilon_{i,j,k,l,s}}{\sum_{i,j,k,l,s} \epsilon_{i,j,k,l,s}} \tag{7.2}$$

where ϵ are the daily emissions of the inventory, e the emission correction factors determined by the 4D-Var data assimilation. i and j denote the horizontal indices of the grid cells, k the layer of the grid cell, l the index of the species in the lumped species L and s the index of the sector. Note that in the 4D-VarREF data assimilation, e does not depend on the sector and in the sector specific data assimilation, e does

not depend on the species.

As expected, the emission corrections determined by the sector specific optimisation show large correlations between the species, with the exception of NMVOC with low corrections and NH_3 from agricultural activities. The time evolution of the emission corrections for CO, NO_x, PM_{10} and $PM_{2.5}$ are similar to that of the road transport sector (Fig. 7.3 and 7.4). This is due to the dominance of this sector in the domain.

The largest differences between the sector specific optimisation and the 4D-VarREF optimisation can be seen in the SO_x corrections. The 4D-VarREF optimisation decreases the SO_x emissions by up to 60%. This reflects the large difference between the simulated and observed concentrations. In contrast, the sector specific optimisation decreases the SO_x emissions by only ~10 % from 10 March to 13 March and from 19 March to 23 March. In between these periods, the emissions are slightly increased by the sector specific optimisation with a maximum of 1.05 on 15 March. The different emission corrections determined by the 4D-VarREF and the sector specific optimisation are a result of the full correlation between all species within the GNFR sectors (Section 4.2), which is only applied in the sector specific optimisation. The SO_x corrections in the sector specific optimisation are mainly driven by the NO gradient, as further discussed in Subsection 7.3.2. For all species except NMVOC, the sector specific optimisation determines higher emission corrections than the 4D-VarREF optimisation. For the species CO, PM_{10} and $PM_{2.5}$, this is a consequence of the full correlation between the corrections of all species within the sectors (Section 4.2) in the sector specific optimisation. Especially the positive emission corrections for road transport and other stationary combustion contribute to the positive overall emission corrections for NO_x, CO, PM_{10} and $PM_{2.5}$ in the sector specific optimisation (Fig. 7.2). As previously discussed, the corrections for both sectors are mainly driven by observational increments of NO_2. The positive NH_3 emission corrections determined by the sector specific optimisation correspond to the relatively high emission corrections for the non-livestock agriculture sector.

The spatial distribution of the total absolute emission corrections in the simulation period determined by the sector specific optimisation is depicted for each GNFR sector (Table 5.1) in Fig. 7.5. The absolute emission corrections are defined by

$$ec_{i,j,L,s} = \sum_{k,l,d} \left(e_{i,j,s,d} - 1 \right) \times \epsilon_{i,j,k,l,s,d} \ , \tag{7.3}$$

where ec is the absolute emission correction in kg, ϵ the daily emissions of the inventory and e the emission correction factor. i and j denote the horizontal indices of the grid cell, k the vertical grid cell and d the index of the day.

The spatial distribution of emission corrections for the road transport sector is heterogeneous. Road transport emissions are increased by the sector specific optimisation

in the southeast of the domain and decreased in the area between the cities Cologne, Düsseldorf and Dortmund. As previously discussed, the emission corrections for the road transport sector are driven by observational increments (Eq. 2.1) of NO_2. The concentrations at the ground stations are influenced by different emission sources in the periods before and after the change in the main wind direction from east to west from 18 March 2016 to 19 March 2016. Positive emission corrections for the road transport sector in the southeast of the domain are related to observed NO_2 concentrations above the simulated concentrations from 10 March to 18 March. This is particularly the case at the three ground stations near Cologne (Fig. 5.9). In contrast, the concentrations at the three ground stations near Cologne are below the simulated concentrations from 19 March to 23 March. For this reason, the data assimilation system determines negative emission corrections for road transport for the area west of Cologne. Thus, the change in meteorological conditions during the simulation period is beneficial for determining reliable emission corrections in most areas of the domain.

The spatial distributions of emission corrections for other stationary combustion, shipping, aviation and off-road traffic show similarities to that for road transport. Examples are the negative emission corrections for the shipping sector on the Rhine north of Düsseldorf and the positive corrections for other stationary combustion and off-road traffic emissions east of Cologne as well as southeast of Dortmund. Also, emissions from aviation southeast of Cologne are strongly increased by the sector specific optimisation. It should be noted that the corrections for the sectors of other stationary combustion, shipping, aviation and off-road traffic are subject to large uncertainty. The results of the identical twin experiments (Chapter 6) indicate that emission changes in the road transport sector can affect the emission corrections for these sectors (Fig. 6.3).

For the public power sector, positive emission corrections are determined in most parts of the domain. This also applies to the Rhine area, where strong negative corrections for the road transport sector are obtained by the sector specific optimisation. Negative emission corrections for public power are only determined west of Cologne and in a small area northwest of Dortmund. This means that the data assimilation system distinguishes between increased emissions from public power and decreased emissions from road transport in the area between Cologne, Düsseldorf and Dortmund. For the distinction between the two sectors, the different diurnal emission profile is exploited, which is flatter for the public power sector than for the road transport sector (Fig. 5.8). Observational increments (Eq. 2.1) of NO_2 between 0 h and 6 h local time are attributed to emission changes in the public power sector rather than to emission changes in the road transport sector in the sector specific optimisation.

The industry sector displays a similar distribution of emission corrections as the road transport sector. However, the distribution also shows characteristic features, for example the positive emission corrections east of Düsseldorf. For this area, the sector specific optimisation distinguishes between increased industrial and decreased

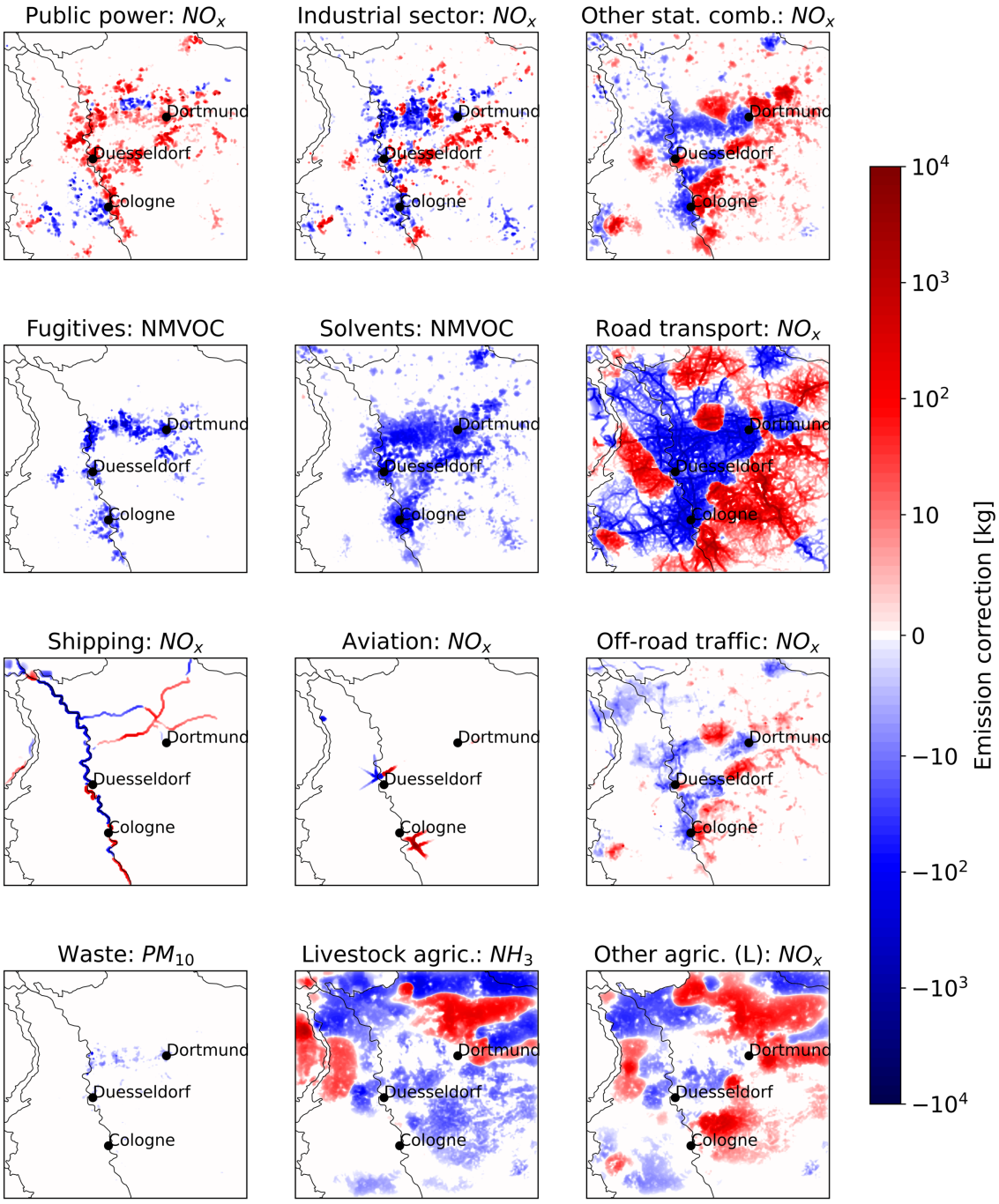

Figure 7.5.: Spatial distribution of the absolute emission corrections for the simulation period for each GNFR sector (Table 5.1). If a sector has moderate or high NO_x emissions, the emission corrections for NO_x are shown. For each of the four sectors with low or zero NO_x emissions, the emission corrections are depicted for a species that is characteristic for the respective sector.

road transport emissions. Due to the flatter diurnal profile of industrial emissions, positive observational increments at night are not attributed to emission increases in the road transport sector, but in the industry sector (and in the public power sector, refer to the previous paragraph). This is discussed for the period from 11 March to 14 March 2016 in the appendix (Section B.6).

The spatial distributions of the emission corrections for the agricultural sectors are also heterogeneous. Both agricultural sectors are strong NH_3 sources (Fig. 5.2). The emission corrections for livestock agriculture are mainly related to observational increments of species that are chemically coupled to NH_3. For the corrections of non-livestock agriculture emissions, also observational increments of NO_2 are important due to the NO_x emissions from this sector (Fig. 5.2). This leads to differences in the spatial distribution of the emission corrections for the two agricultural sectors, especially in an area east of Cologne. In this area, high positive emission corrections are determined for the non-livestock agriculture sector, similar as for the road transport sector. A probable reason for this is that the emission corrections for the non-livestock agriculture sector are affected by high road transport emissions in this area. A further discussion of this behaviour can be found in the appendix (Section B.5).

The emission corrections for fugitives, solvents and waste are zero or negative in the entire domain. This has the following reason: The three sectors emit a significant amount of aerosols, but a low amount of NO_x (Fig. 5.2). Hence, aerosol observations have a significant impact on the emission corrections for fugitives and solvents. In the entire domain, most aerosol observations are below the simulated concentrations. This leads to negative emission corrections for the fugitives, the solvents and the waste sector. In the case of the waste sector, PM_{10} emission corrections above 10 kg per grid cell during the simulation period are only determined for some grid cells north of Duesseldorf. This is due to the low emissions from the waste sector in all other parts of the domain, resulting in small absolute emission corrections for this sector.

A comparison of the spatial distributions of the NO_x emission corrections summed over all sectors determined by the sector specific and the 4D-VarREF optimisation can be found in the appendix (Fig. B.1). In the 4D-VarREF optimisation, only small corrections are obtained for the area between the cities Cologne, Düsseldorf and Dortmund and for the southeast of the NRW domain in contrast to the corrections determined by the sector specific optimisation. The main reason for this is the large emission corrections for the road transport sector in the sector specific optimisation (Fig. 7.5).

7.3. Agreement between simulated and observed concentrations

The aim of optimising the initial values and emission correction factors is to reduce the discrepancy between model and observations. As previously described (Chapter 2), the final step of a data assimilation cycle is the analysis forecast (also referred to as "analysis" in this section) using the optimised initial concentrations and emission corrections. In this section, the agreement between observed and simulated pollutant concentrations is investigated. The time series of the NO_2, O_3, SO_2 and PM_{10} concentrations averaged for all validation stations are depicted and discussed in the following subsections. The observed concentrations are compared to the simulated concentrations of the analysis of the sector specific optimisation, of the analysis of the 4D-VarREF optimisation and of a reference run without optimisation of emissions and initial concentrations. In addition, the following statistics are included in the figures shown in this section:

- The correlation coefficient of observed and simulated concentrations ("Corr") defined by $\frac{\sum_{k=1}^{N}(x_i-\overline{x})\,(y_i-\overline{y})}{\sqrt{\sum_{k=1}^{N}(x_i-\overline{x})^2\,(y_i-\overline{y})^2}}$ with the simulated concentrations x_i, the mean $\overline{x}$ of the simulated concentrations, the observed concentrations y_i, the mean $\overline{y}$ of the observed concentrations and N the number of observations,
- the bias in µg/m³ as the difference of the mean of simulated and observed concentrations defined by $\frac{\sum_{k=1}^{N} x_i-y_i}{N}$,
- the unbiased root mean squared error ("RMSE") in µg/m³ defined by $\sqrt{\frac{1}{N}\sum_{k=1}^{N}[(x_i-\overline{x})(y_i-\overline{y})]^2}$,
- the normalised mean bias ("NMB") defined by $\frac{\sum_{k=1}^{N} x_i-y_i}{\sum_{k=1}^{N} y_i}$ and
- the normalised mean error ("NME") defined by $\frac{\sum_{k=1}^{N} |x_i-y_i|}{\sum_{k=1}^{N} y_i}$.

7.3.1. NO_2 and O_3

The averaged time series of the NO_2 concentrations at the validation stations are shown in Fig. 7.6. The bias is negative for the analysis of the sector specific optimisation with -0.9 µg/m³ and -2 µg/m³ for the reference run without optimisation. In contrast, a positive bias is calculated for the analysis of the 4D-VarREF optimisation (2 µg/m³). It is paradoxical that the sector specific optimisation determines higher NO_x emissions (Fig. 7.4) and at the same time leads to lower analysis concentrations for NO_2 at the validation stations than the 4D-VarREF optimisation. The major reason for this is a stronger negative correction of initial NO_2 concentrations by the sector specific optimisation than by the 4D-VarREF optimisation, particularly on 19 March and 20 March. The initial value corrections for both simulations on both days are shown

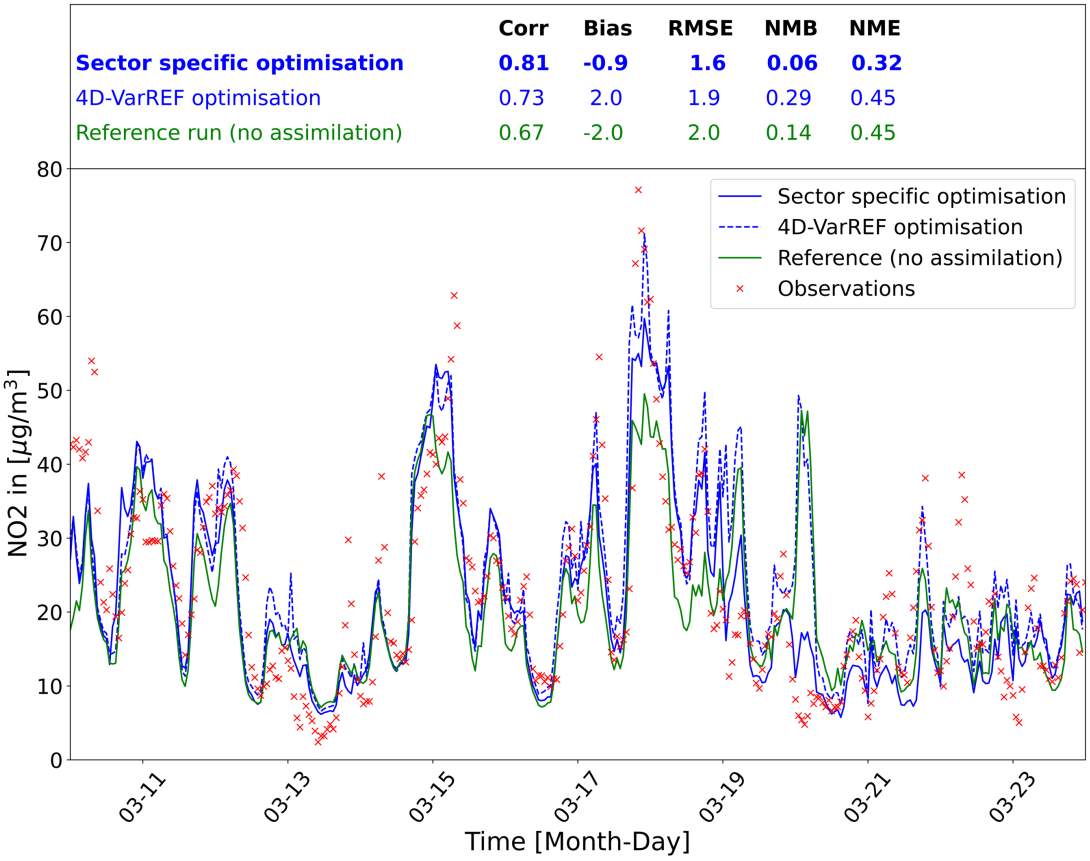

Figure 7.6.: Time series of NO_2 concentrations averaged over all validation stations. The observations are depicted as red crosses, the reference run without data assimilation as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line. The correlation ("Corr"), the bias in µg/m³, the unbiased root mean squared error ("RMSE") in µg/m³, the normalised mean bias ("NMB") and the normalised mean error ("NME") - each with respect to the observations - are given for the time series of the three simulations.

in the appendix (Section B.2, Fig. B.2, B.3 and B.4). Compared to the reference simulation without optimisation, the correlation between model and observations is higher and the unbiased root mean squared error is lower for the analyses of both the sector specific and the 4D-VarREF optimisation. The improvement of the simulated NO_2 concentrations by the sector specific optimisation results in a correlation coefficient of 81%. In comparison, for the analysis of the 4D-VarREF optimisation, the correlation coefficient is significantly lower with a value of 73%. The differences between the analyses of the sector specific optimisation and the 4D-VarREF optimisation are most remarkable on 19 March and 20 March. On 20 March, around 0 UTC, the NO_2 concentrations of the analysis of the 4D-VarREF optimisation and of the reference

simulation without optimisation show a peak with a value of 50 µg/m^3, while the observed concentrations are lower than 10 µg/m^3. In contrast, the NO_2 concentrations of the analysis of the sector specific optimisation are much lower with a maximum value of 18 µg/m^3 at that time, which is more consistent with the observations. As mentioned in the previous paragraph, this difference is due to a stronger initial value correction in the sector specific optimisation than in the 4D-VarREF optimisation (Fig. B.2 and B.3 in the appendix). The better agreement with the observed concentrations on 19 March and 20 March leads to a lower unbiased root mean squared error for the analysis of the sector specific optimisation (1.6 µg/m^3) than for the analysis of the 4D-VarREF optimisation (1.9 µg/m^3).

The comparison of the averaged O_3 time series at the validation stations displayed in Fig. 7.7 reveals similar characteristics as the time series for NO_2. A major reason for this is that NO_2 and O_3 concentrations are strongly coupled by their photostationary state. The bias for the analysis of the sector specific optimisation (3.8 µg/m^3) is higher than that for the reference simulation without optimisation (2.9 µg/m^3). In contrast, the bias for the analysis of the 4D-VarREF optimisation (2.7 µg/m^3) is slightly lower than for the reference simulation. The improvement of the correlation and the unbiased RMSE is greater for the sector specific optimisation than for the 4D-VarREF optimisation (91 % vs. 88 % correlation and 1.8 µg/m^3 vs. 2 µg/m^3 unbiased RMSE). The differences in the bias and unbiased RMSE between the analysis of the sector specific optimisation and the analysis of the 4D-VarREF optimisation are mainly due to the discrepancy between their time series on 19 March and 20 March. On both days, the analysis of the sector specific optimisation is more consistent with the observed O_3 concentrations than the analysis of the 4D-VarREF optimisation. The reason for the difference between the two analyses is a strong positive correction of the initial O_3 concentrations on 19 March and 20 March in the sector specific optimisation. A comparison of the initial value corrections for O_3 by the sector specific and the 4D-VarREF optimisation can be found in the appendix (Fig. B.4).

To conclude, the improvement of the agreement between simulated and observed NO_2 and O_3 concentrations by the sector specific optimisation is comparable to that of the 4D-VarREF optimisation. The differences on 19 and 20 March are mainly due a stronger initial value correction by the sector specific optimisation than by the 4D-VarREF optimisation. This is not characteristic for the sector specific optimisation, since the strength of initial value corrections can be adjusted in the configuration of both data assimilation systems.

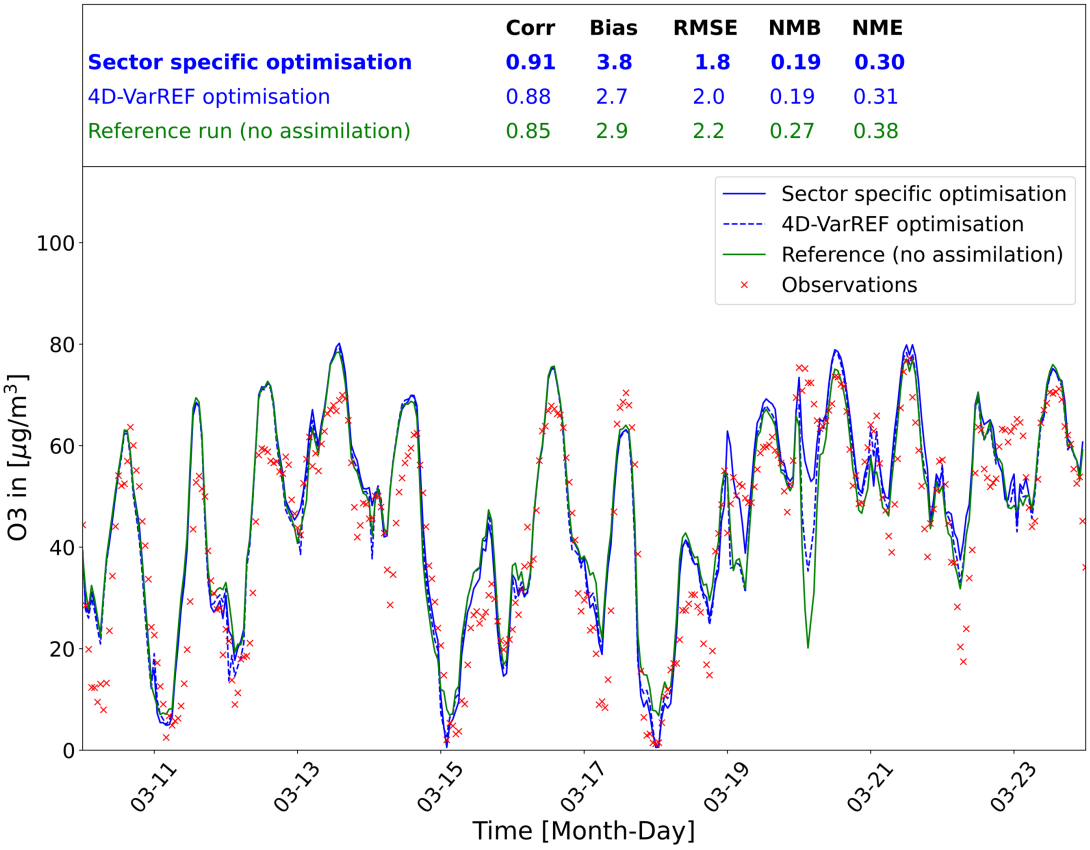

Figure 7.7.: Time series of O_3 concentrations averaged over all validation stations. The observations are depicted as red crosses, the reference run without data assimilation as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line. The correlation ("Corr"), the bias in µg/m³, the unbiased root mean squared error ("RMSE") in µg/m³, the normalised mean bias ("NMB") and the normalised mean error ("NME") - each with respect to the observations - are given for the time series of the three simulations.

7.3.2. SO_2

The time series of SO_2 concentrations at the validation station are shown in Fig. 7.8. In contrast to NO_2 and O_3, for the SO_2 concentrations, there is a large discrepancy between each of the three simulations and the observations. This is reflected by high values for the normalised mean bias (NMB) and the normalised mean error (NME) as well as by low correlation coefficients for all simulations. The bias for the reference simulation without data assimilation is 6 µg/m³. The bias for the 4D-VarREF optimisation (2.4 µg/m³) and the sector specific optimisation (3.3 µg/m³) is lower compared to the reference simulation. The discrepancy between simulated and observed SO_2 concentrations suggest that the SO_x emissions are strongly overestimated in the emis-

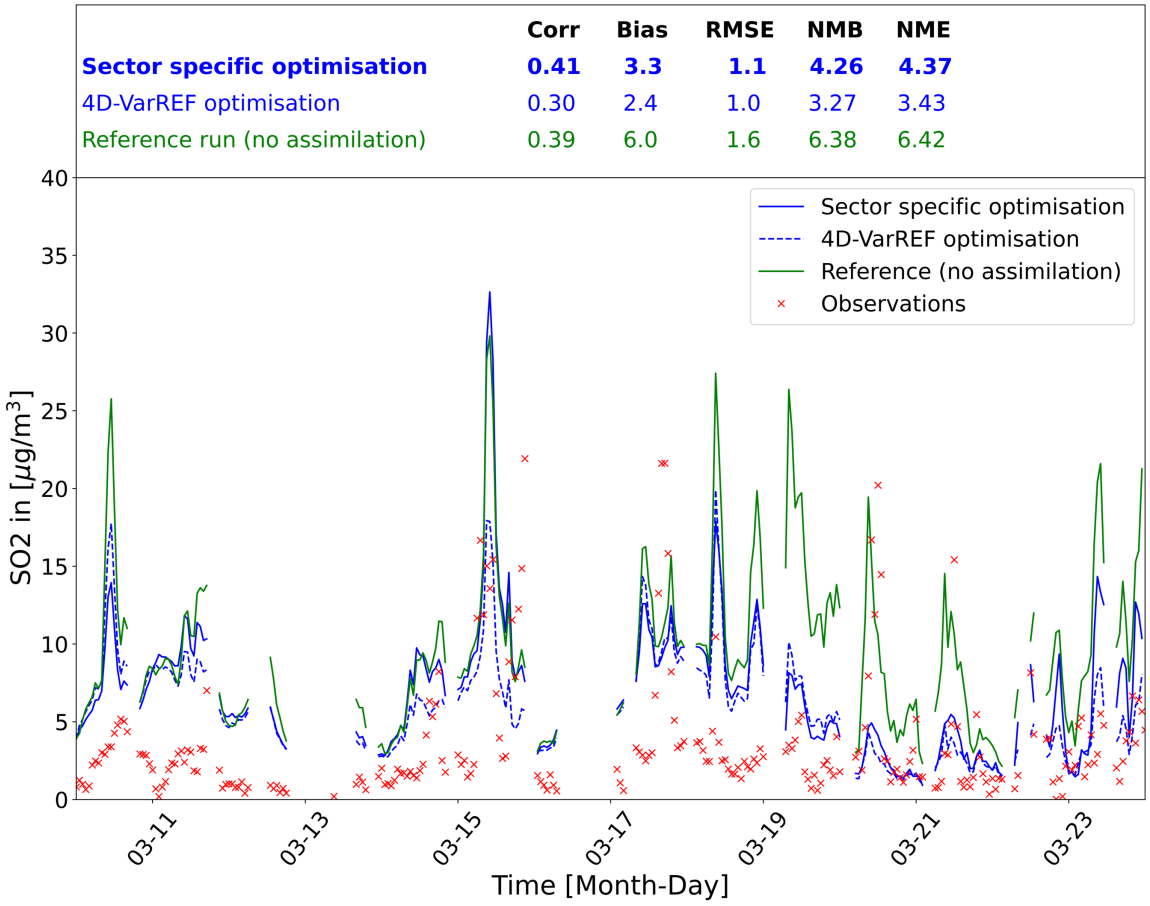

Figure 7.8.: Time series of SO_2 concentrations at the validation station. The observations are depicted as red crosses, the reference run without data assimilation as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line. The correlation ("Corr"), the bias in µg/m^3, the unbiased root mean squared error ("RMSE") in µg/m^3, the normalised mean bias ("NMB") and the normalised mean error ("NME") - each with respect to the observations - are given for the time series of the three simulations. Note that there is only 1 validation station for SO_2 and that the hourly observational dataset for this station is incomplete (Table 5.5).

sion inventory. Another reason for the bias could be incorrect assumptions about the vertical profile of emissions (Fig. 5.5) from the public power and the industry sector. If the effective emission height assumed in the inventory is too low, the simulated SO_2 concentrations are higher than the observed SO_2 concentrations, particularly in the vicinity of emission sources. The 4D-VarREF optimisation reduces the bias by a strong decrease of the total daily SO_x emissions by up to 60% (Fig. 7.4). In the sector specific optimisation, the emission corrections for SO_x are much weaker than in the 4D-VarREF optimisation. Nevertheless, the bias of the SO_2 concentrations is

significantly smaller for the analysis of the sector specific optimisation than for the reference simulation without data assimilation.

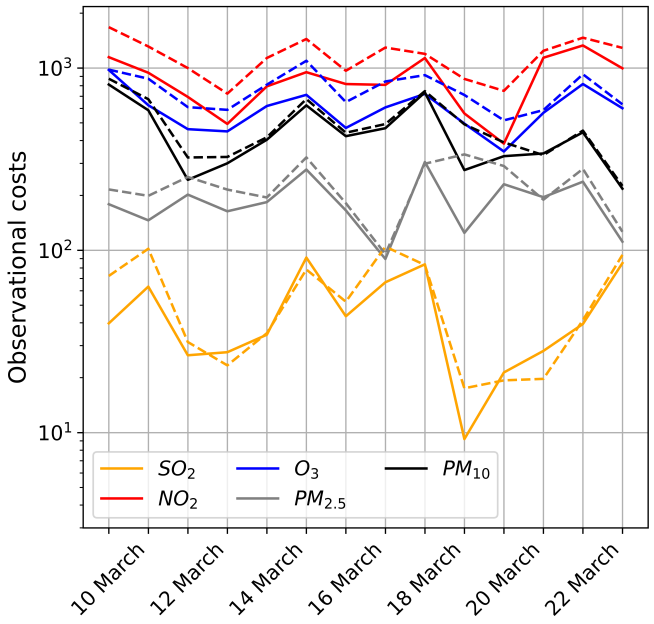

Figure 7.9.: Development of the observational costs (Eq. 2.6) of the analysis during the simulation period from 10 March to 23 March. The background costs are depicted as dashed lines, the analysis costs as solid lines.

Regarding the statistical values (Fig. 7.8), the agreement between simulated and observed concentrations is worse for SO_2 than for NO_2 and O_3. This also applies to both simulations using data assimilation. The reason for this is the low observational costs of SO_2 in comparison to all other observed species (Fig. 7.9). The low costs are due to the small number of observations and a high relative observation error assumed in the model (Tables 5.2, 5.3 and Fig. 5.9). In order to minimise the total costs, the 4D-Var data assimilation system tends to optimise the species with the highest observational costs more strongly. The bias is larger for the analysis of the sector specific optimisation than for the analysis of the 4D-VarREF simulation because of the weaker SO_x emission corrections obtained in the sector specific optimisation. As discussed in Section 7.2, this is a consequence of the full correlation between the SO_x and the NO_x emission corrections within the sectors (Section 4.2).

After 19 March, the SO_2 concentrations are lower with an average of approximately 5 µg/m^3 and less fluctuating. The discrepancy on these days is strongly reduced by both the sector specific and the 4D-VarREF optimisation compared to the reference

simulation without data assimilation. In order to investigate the reason for this in the case of the sector specific optimisation, a closer look is taken at the SO_x emission corrections from 18 to 20 March.

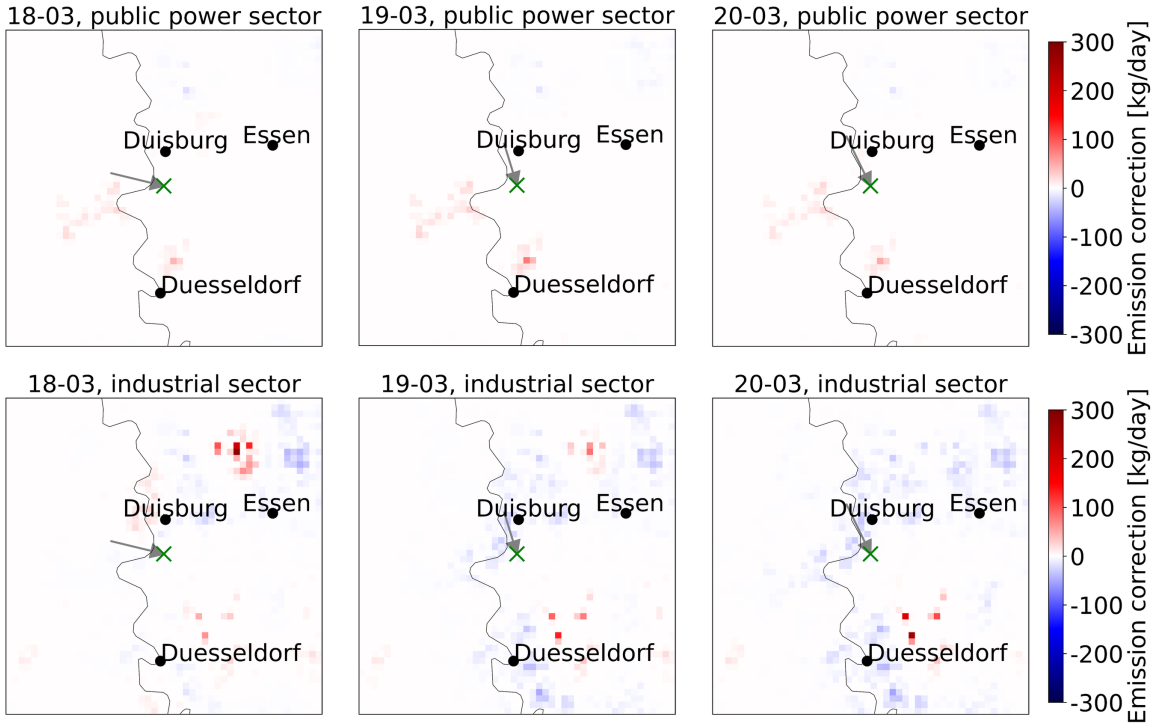

Figure 7.10.: Absolute SO_x emission corrections for the public power and the industry sector from 18 March to 20 March obtained in the sector specific optimisation. The sum of the emissions corrections over all model layers is plotted. A 51 km × 51 km part of the NRW domain is shown, with the SO_2 validation station (green cross) in the center. The wind direction at 12 UTC is displayed as a grey arrow.

The public power, industry and stationary combustion sectors are the only GNFR sectors with moderate or high SO_x emissions in NRW (Fig. 5.2). Fig. 7.10 shows the absolute SO_x emission corrections from 18 to 20 March for the public power and the industry sector in the area around the validation station obtained in the sector specific optimisation. The emission corrections for the sector of other stationary combustion are negligible in the investigated area and period and are therefore not shown in Fig. 7.10. The emission corrections for the public power sector are mostly positive with two exceptions in the areas north of Duisburg and north of Essen and are similar for all three days. In contrast, the emission corrections for the industry sector decrease during the three days. The strong emission increase north of Duisburg on 18 March change to a strong emission decrease on 19 March. This applies also to the corrections northwest of Essen. The wind direction changes from west-northwest on 18 March to north-northwest on 19 and 20 March. Thus, on 19 and 20 March, emissions from

industrial sources near Duisburg impact the SO_2 concentrations observed at the validation station. On 18 March, west of the validation station, the power sector has positive and the industry sector has negative corrections. The result is a low total SO_x emission correction in this area. Therefore, the reduction of the simulated SO_2 concentrations is small. In contrast, on 19 and 20 March, strong negative emission corrections for the industry sector north of the validation station reduce the simulated SO_2 concentrations at the validation station. The consequence is a better agreement between the simulated and observed concentrations on 19 and 20 March for the analysis of the sector specific optimisation compared to the reference simulation without data assimilation.

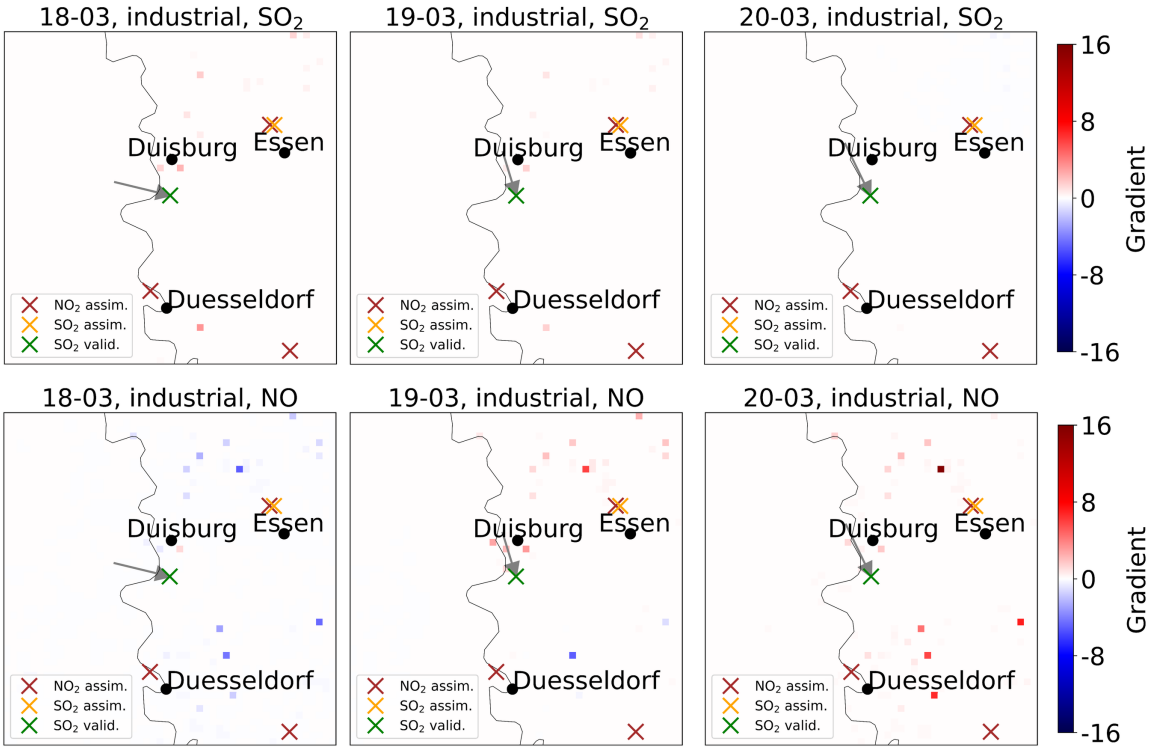

Figure 7.11.: SO_2 and NO emission gradient of the industry sector from 18 to 20 March after the first adjoint run. Note that a negative gradient corresponds to emission increases and vise-versa. A part of the NRW domain with 51 km × 51 km is shown, containing the SO_2 validation station (green cross), three NO_2 assimilation stations (brown crosses) and one SO_2 assimilation station (orange cross). The wind direction at 12 UTC is displayed as a grey arrow.

The corrections for the industry sector are a consequence of the NO and SO_2 emission gradients for this sector, which are depicted in Fig. 7.11. In this figure, the gradient after the first adjoint model run is shown. Note that most grid cells have negligible gradients since there are low or zero industrial emissions. For 18 March, a slightly

positive SO_2 gradient is calculated for some grid cells. The maximum value is 2.3 in the investigated area. However, the negative NO gradient is stronger than with a minimum value of -5 in the depicted area. The correlation of both species in the sector specific optimisation (Section 4.2) leads to positive emission corrections for the industry sector. On 19 March, the SO_2 gradient decreases, but the NO gradient changes its sign from negative to positive. On 20 March, the NO gradient increases further. Consequently, negative emission corrections for the industry sector are determined on both days, which strongly decreases the simulated SO_2 concentrations at the validation station. Thus, the investigated case illustrates the dominant impact of the NO gradient and the minor impact of the SO_2 gradient on the emission corrections for the industry sector.

To conclude, the agreement between simulated and observed SO_2 concentrations is strongly improved by the sector specific optimisation in comparison to the reference simulation without data assimilation. However, a significant positive bias remains after the optimisation. The main reason for this is the low sensitivity of the cost function value to the SO_x emissions compared to its sensitivity to NO_x emissions. A possible approach to further improve the simulated SO_2 concentrations in future developments could be to determine separate correction factors for SO_x emissions and emissions of other species in the case of the industry and the public power sector. This can be achieved by modifications of the emission error correlation matrices of the two sectors (Section 4.2).

7.3.3. Aerosols

The time series of the PM_{10} concentrations averaged over all validation stations are shown in Fig. 7.12. On average, the simulated PM_{10} concentrations are higher than the observations. The bias for the reference simulation without data assimilation is 1.9 µg/m^3. It is larger for the analysis of the sector specific optimisation (3 µg/m^3) and for the analysis of the 4D-VarREF optimisation (2.6 µg/m^3). The correlation between the simulated and observed concentrations is high for all simulations with values of 93-94%. Both the unbiased root mean squared error and the normalised mean error (NME) are slightly worse for both simulations using data assimilation than for the reference run without data assimilation.

The main reason for the higher bias in the simulations with data assimilation in comparison to the reference run is the strong positive initial value corrections of aerosols. This can be illustrated by average diurnal cycles at three assimilation stations as examples (Fig. 7.13). From 0 to 8 UTC, the concentrations of the analyses of both the sector specific and the 4D-VarREF optimisation are higher than that of the reference simulation without data assimilation. This indicates high initial value corrections of PM_{10} at the three stations, also in the case that the observed values are lower than the results of the reference simulation.

The following statements can be made about the impact of the aerosol emission

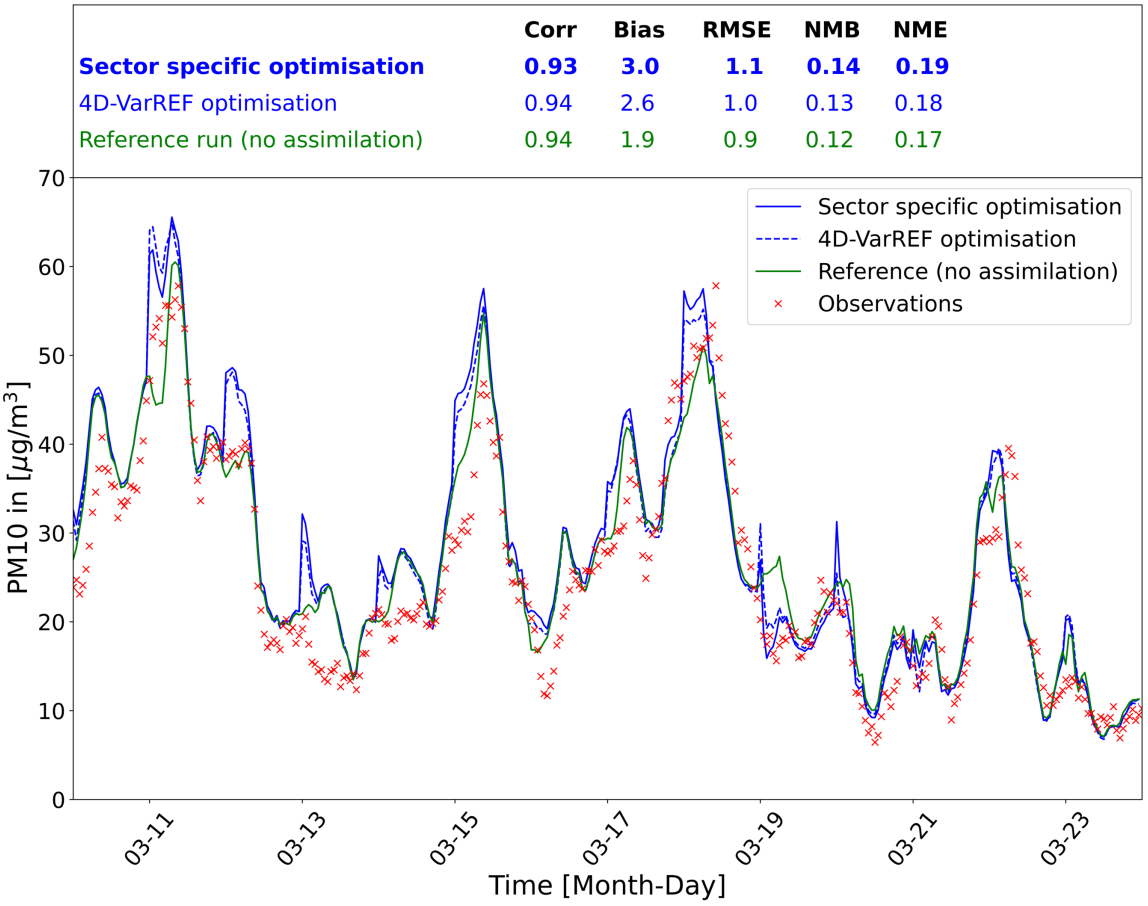

Figure 7.12.: Time series of PM_{10} concentrations averaged over all validation stations. The observations are depicted as red crosses, the reference run without data assimilation as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line. The correlation ("Corr"), the bias in µg/m^3, the unbiased root mean squared error ("RMSE") in µg/m^3, the normalised mean bias ("NMB") and the normalised mean error ("NME") - each with respect to the observations - are given for the time series of the three simulations.

corrections. Despite the high observational costs for PM_{10} (Fig. 7.9), the aerosol gradients are small in both the sector specific and the 4D-VarREF optimisation. As a consequence, only small corrections for aerosol emissions are determined by the 4D-VarREF optimisation (Fig. 7.4), which have a minor influence on the simulated concentrations. In the sector specific optimisation, the total aerosol emissions of the simulation period are increased due to the full correlation between the aerosol and the NO_x emission corrections within the sectors (Section 4.2). This leads to a slightly higher bias for the analysis of the sector specific optimisation than for the analysis of the 4D-VarREF optimisation.

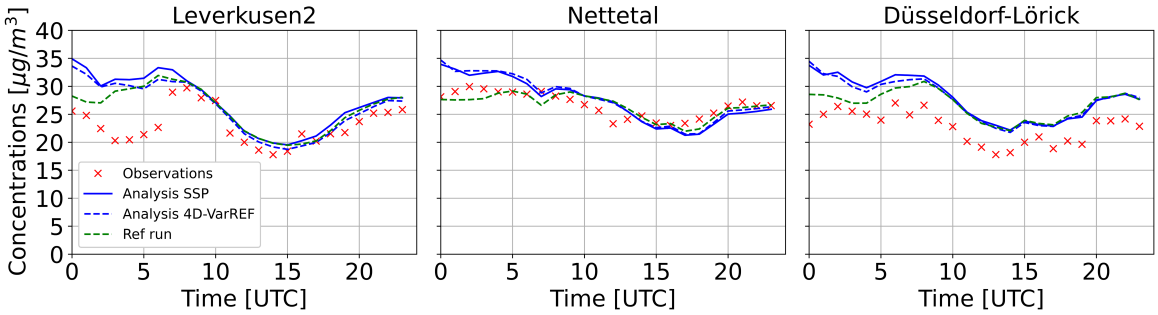

Figure 7.13.: Average diurnal cycle of simulated and observed PM_{10} concentrations during the simulation period at three assimilation stations. The observations are depicted as red crosses, the reference run as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line.

The time series of $PM_{2.5}$ concentrations show similar characteristics as the time series for PM_{10}. They are displayed in the appendix (Fig. B.5).

To conclude, the agreement between simulated and observed PM_{10} concentrations is already high in the reference simulation of the EURAD-IM without data assimilation. Neither the sector specific nor the 4D-VarREF optimisation can improve the simulated concentrations. This is mainly due to incorrect initial value corrections for aerosols, which can also be observed in other simulations of the EURAD-IM. The reason for this should be investigated, which is beyond the scope of this work.

7.4. Conclusions of the simulation using ground observations

In this Chapter, the behaviour, potential and limits of the approach developed in the framework of this thesis have been investigated on the basis of a simulation using real ground observations. In this study, the sector specific emission optimisation is able to significantly improve the agreement between simulated and observed NO_2, O_3 and SO_2 concentrations compared to a simulation without data assimilation. However, there is no substantial difference to the concentrations obtained in a reference simulation using the current data assimilation system of the EURAD-IM. The major advantage of the new development is the gain of information on the emission strengths of individual polluter groups. The sector specific emission optimisation is able to distinguish between agricultural and road transport emissions using the high NH_3 emissions from agriculture. Furthermore, the data assimilation system determines positive emission corrections for the public power sector in large parts of the domain, where negative emission corrections for the road transport sector are obtained at the same time.

For the distinction between the emissions from the two sectors, the different diurnal emission profile of the two sectors is efficiently exploited by the data assimilation system. Similar statements can be made for the distinction between the emissions from the road transport and the industry sector. In this study, the emission corrections for the public power, industry and road transport sectors are mainly driven by observational increments of NO_2 and O_3, which are mostly attributed to NO_x emissions. The additional use of observations of other species, e.g. CO and CO_2, could also enable the data assimilation system to distinguish between the emissions of the three sectors on the basis of their chemical composition. This should be investigated in future studies.

8. Conclusion and outlook

In the framework of this thesis, a new data assimilation approach was developed that allows for the correction of emission data of trace gases and aerosols for individual source categories. The main purpose of this development is to obtain information on the contribution of individual source categories such as road transport and industry to air pollution. In comparison to other approaches for sector specific emission correction, e.g. Meirink et al. (2008) and Hooghiemstra et al. (2011), the development of this work is suitable for the assimilation of observations of NO_2 and O_3 concentrations with a high sensitivity to regional emissions.

The new approach was implemented in the complex 4D-Var data assimilation system of the EURAD-IM. The new data assimilation system includes individual emission correction factors for the Gridding Nomenclature for Reporting (GNFR) sectors and the calculation of adjoint sector specific emission correction factors in the adjoint model run. In addition, two modifications are introduced to improve the results of the sector specific optimisation. The first is based on the assumption of a constant chemical composition of the emissions from the sectors. This leads to equal emission correction factors for all species within the sectors and aims to reduce the degree of freedom of the emission correction factors. The second modification increases the spatial correlation of the emission corrections for the road transport sector along roads. This corresponds to the assumption of a high correlation of real traffic emissions along roads of a high category such as motorways.

The studies discussed in this thesis have the purpose of investigating the ability of the new approach to distinguish between emissions from different GNFR sectors. The domain selected for the studies covers large parts of North Rhine-Westphalia. In this area, emissions from the industry, public power and road transport sectors are predominant. The simulation period in March 2016 was characterised by high agricultural emissions. Two types of studies were performed that should be regarded as complementary. In the identical twin experiments, artificial emission scenarios with emission changes in the road transport, industry and non-livestock agriculture sectors were simulated. Synthetic observations were generated in simulations with these emission scenarios as input for the data assimilation. The advantage of these studies is that the true emission changes are known. Hence, the emission corrections determined in the sector specific optimisation can be compared with the true corrections to evaluate the performance of the new method. Another aspect of the identical twin experiments is that no sources of uncertainty related to the model and the observations are included in the simulated scenarios. However, these uncertainty sources are not

negligible if real observations are assimilated. For this reason, a two-week simulation using ground observations was performed in addition to the identical twin experiments.

All studies discussed in this thesis demonstrate the ability of the new data assimilation approach to distinguish between emissions from different sectors that contribute substantially to air pollution. In an identical twin experiment, an emission scenario with an increase in industrial emissions and a simultaneous decrease in road transport emissions is correctly reflected by the data assimilation system. In another identical twin experiment, increased non-livestock agriculture emissions are distinguished from decreased road transport emissions. In the simulation assimilating real ground observations, positive emission corrections are determined for the public power sector in a large area of the model domain, where negative corrections are obtained for road transport emissions. For the distinction between the agriculture sectors and other sectors, the high agricultural NH_3 emissions are exploited by the data assimilation system. Public power and industrial emissions are distinguished from road transport emissions mainly on the basis of the different diurnal emission profiles. For road transport, low emissions between 0 h and 6 h local time are assumed in the emission data, while industrial and public power emissions are characterised by a flat diurnal profile. Therefore, differences between observed and simulated concentrations at night are mainly attributed to emission changes in the industry and the public power sector.

Although there are substantial differences in the chemical composition of the emissions from road transport, industry and public power, the data assimilation system cannot efficiently exploit them in the simulations. High SO_x emissions are characteristic of the industry and the public power sector, in contrast to the road transport sector. However, the sensitivity of the data assimilation system to SO_x emissions is low due to the small number of SO_2 observations in the dataset compared to NO_2 and O_3 observations. Both the industry and the road transport sector emit high amounts of CO, in contrast to the public power sector. This difference cannot be exploited for a distinction, since no CO observations are assimilated. With the observational dataset used in the simulations, NO_2 and O_3 observations, which are sensitive to NO_x emissions, have a dominant impact on the emission corrections. This hampers the data assimilation system to distinguish between emissions from different sectors on the basis of their chemical composition.

In the following, suggestions for next steps to further develop and investigate the sector specific emission optimisation are described. Future improvements should aim at a more efficient exploitation of the chemical composition of the emissions of the sectors by the data assimilation system. An enhanced impact of SO_2 and aerosol observations on the determined emission corrections might be beneficial in particular for the distinction between industry, public power and road transport emissions. This could be achieved by applying species-dependent scaling factors to the emission gradients. The setting of such scaling factors could be sector-dependent, e.g. based on the chemical composition of the sectors. In addition, exceptions to the full correlation between all

species within the sectors could be introduced in the emission error correlation matrix. For example, a low correlation between SO_x and all other emitted species could be specified in the matrix, which would enable a separate optimisation of SO_x emissions within the sectors. This would probably improve the agreement between simulated and observed SO_2 concentrations.

As previously discussed, the observational dataset used for the data assimilation has a crucial impact on the performance of the sector specific emission optimisation. The ground observations in NRW are well suited for the correction of road transport emissions due to the high number of NO_2 observations and the high sensitivity to ground level emissions. CO observations could be beneficial for the distinction between road transport and public power emissions, since the CO/NO_x ratio is significantly higher for road transport than for public power emissions. Also the assimilation of CO_2 observations might improve the results. In addition, observations from zeppelin campaigns could be advantageous, as a high fraction of public power and industrial emissions occurs at the flight altitudes of zeppelins.

In future studies, the performance and behaviour of the sector specific emission optimisation under different conditions should be investigated. Meteorological parameters such as wind speed, precipitation and cloud cover have a crucial impact on atmospheric chemistry and thus affect the data assimilation results. Therefore, simulations under different meteorological conditions, e.g. with a comparison between the seasons, should be performed. Further studies for different regions would be beneficial to study the distinction between the sectors by the data assimilation system in different emission situations. The domain of Lower Saxony would be suitable for such a study. Compared to North Rhine-Westphalia, the agricultural activity in Lower Saxony is significantly higher, while industrial emissions are negligible in many areas. Furthermore, the sensitivity to a coarser resolution of the model grid could be investigated.

In summary, the studies presented in this thesis showed the ability of the new development for sector specific emission optimisation to effectively distinguish between emissions of different source categories. This applies in particular to the sectors of public power, industry, road transport and agriculture. A further improvement of the performance could be achieved by using additional observational data, e.g. of CO and CO_2.

General acknowledgements

The author gratefully acknowledges the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS (Alvarez, 2021) at the Jülich Supercomputing Centre (JSC).

The author gratefully acknowledges the computing time granted through JARA on the supercomputer JURECA (Thörnig, 2021) at Forschungszentrum Jülich.

The author acknowledges the Federal Environment Agency (Umweltbundesamt) for funding this work within the framework of project FKZ 3717512520.

A. Appendix: Supplements to the identical twin experiments

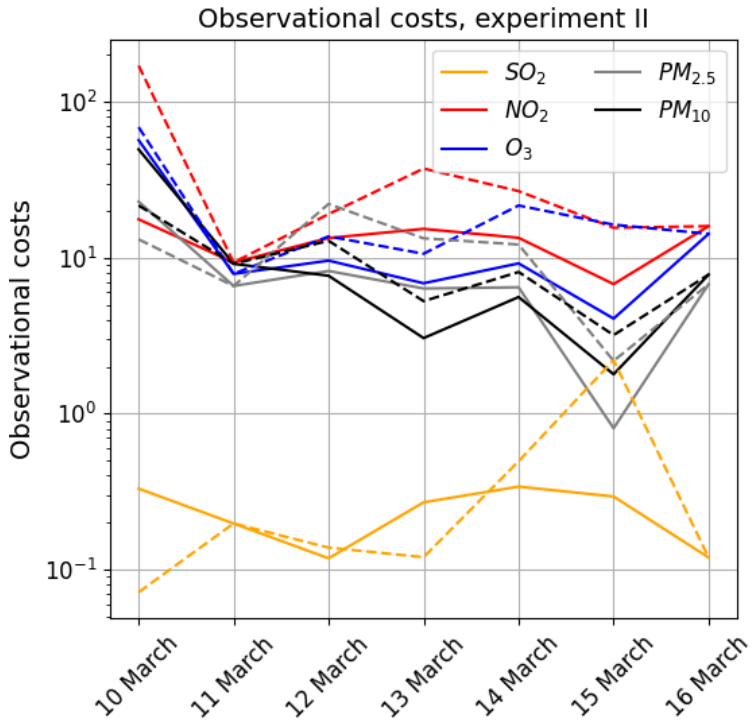

Figure A.1.: Observational costs (Eq. 2.6) of the observed species during the simulation period in experiment II with emission perturbation factors of 2 for the non-livestock agriculture and 0.5 for the road transport sector. Analysis costs are depicted as solid lines, background costs as dashed lines. If the analysis costs are lower than the background costs of a species, the sector specific optimisation of emissions has improved the agreement between simulated and observed concentrations of this species.

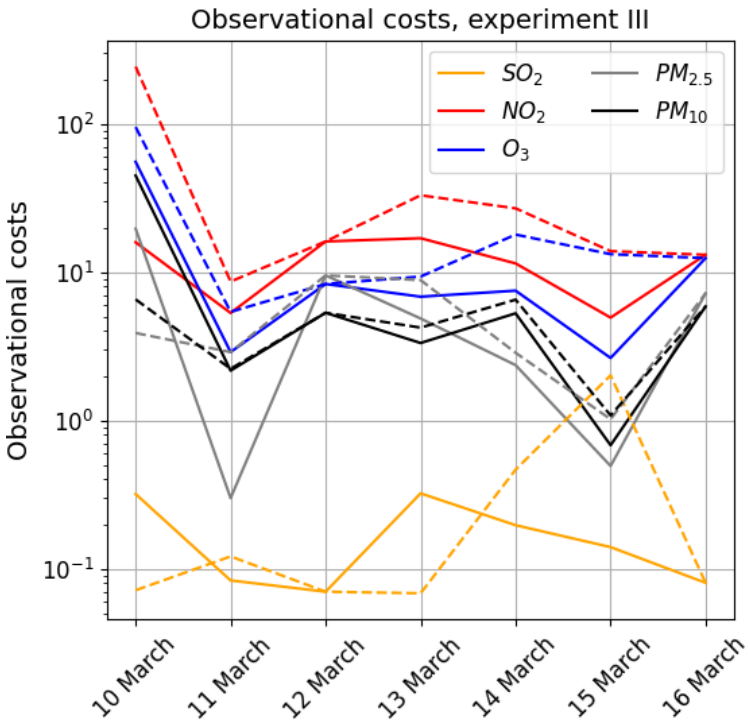

Figure A.2.: Observational costs (Eq. 2.6) of the observed species during the simulation period in experiment III with emission perturbation factors of 1.3 for the non-livestock agriculture and 0.5 for the road transport sector. Analysis costs are depicted as solid lines, background costs as dashed lines. If the analysis costs are lower than the background costs of a species, the sector specific optimisation of emissions has improved the agreement between simulated and observed concentrations of this species.

B. Appendix: Supplements to the simulation using real observations

B.1. Spatial distribution of emission corrections

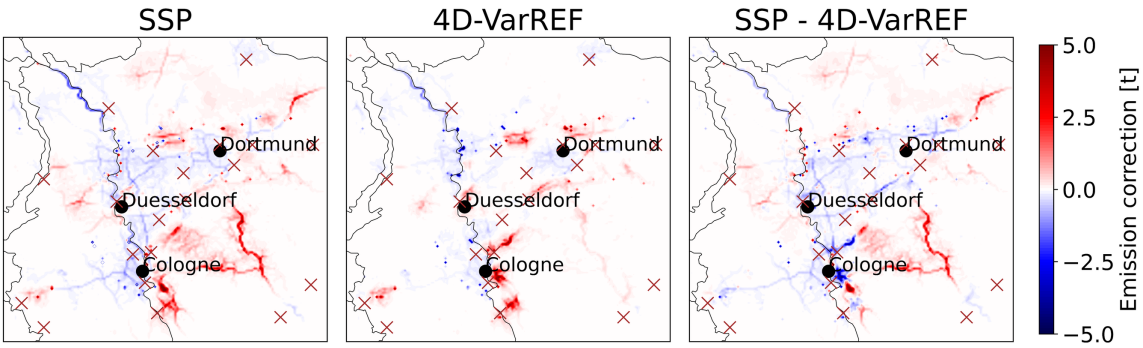

Figure B.1.: Absolute NO_x emission corrections for the whole simulation period from 10 to 23 March 2016 of the 4D-VarREF and the sector specific optimisation. The sum over all sectors is shown. In addition, the difference between the two optimisations, denoted as "SSP - 4D-VarREF", is depicted. The locations of the NO_2 assimilation stations are depicted as brown crosses.

B.2. Initial value corrections of NO_2

As stated in Subsection 7.3.1, on 19 March and 20 March, the NO_2 and O_3 concentrations of the analysis of the sector specific optimisation agree better with the observations than the concentrations of the analysis of the 4D-VarREF optimisation. The main reason for this is the initial value corrections, which are stronger in the sector specific optimisation than in the 4D-VarREF optimisation.

Fig. B.2 shows the initial NO_2 mixing ratios of the background (Section 2) and the analysis of the sector specific and the 4D-VarREF optimisation and their difference on 19 March 2016. The difference between analysis and background concentrations ("ANA-BG") corresponds to the correction of the initial mixing ratios. The figure highlights the different behaviour of the sector specific and the 4D-VarREF optimisation concerning the initial value corrections. The NO_2 concentrations on 19 March, 0 UTC, show a high heterogeneity, which is characteristic for a compound with a short lifetime. In large parts of the north of the domain, the mixing ratios are below 3 ppb, while they exceed a value of 20 ppb south of Cologne characterised by high NO_x emissions. The initial mixing ratios of the background are different for the two simulations. They are lower in the Rhine area south of Cologne and higher in some other areas, e.g. east of Cologne, in the simulation using the sector specific optimisation than in the simulation using the 4D-VarREF optimisation. This is a result of the differences between the emission corrections for the previous day determined by the sector specific and the 4D-VarREF optimisation. The initial value corrections of NO_2 are mostly negative for both simulations. The initial value corrections determined by the sector specific optimisation are strongest between Cologne and Düsseldorf, while they are less than 4 ppb for the 4D-VarREF optimisation in the same area. Overall, the initial value corrections on 19 March and 20 March (Fig. B.3) are significantly stronger in the sector specific optimisation than in the 4D-VarREF optimisation. This also applies to the initial O_3 maxing ratios, which are strongly increased by the sector specific optimisation. This is shown in Fig. B.4.

To conclude, the better agreement between simulated and observed concentrations of NO_2 and O_3 concentrations for the analysis of the sector specific optimisation than for the analysis of the 4D-VarREF optimisation on 19 March and 20 March is due to stronger initial value corrections in the simulations. The strength of the initial value correction is not characteristic for the sector specific optimisation, as it can be adjusted in the configuration of the data assimilation system (e.g. by changing the errors specified in the background error covariance matrix $\mathbf{B}$, see Eq. 3.10).

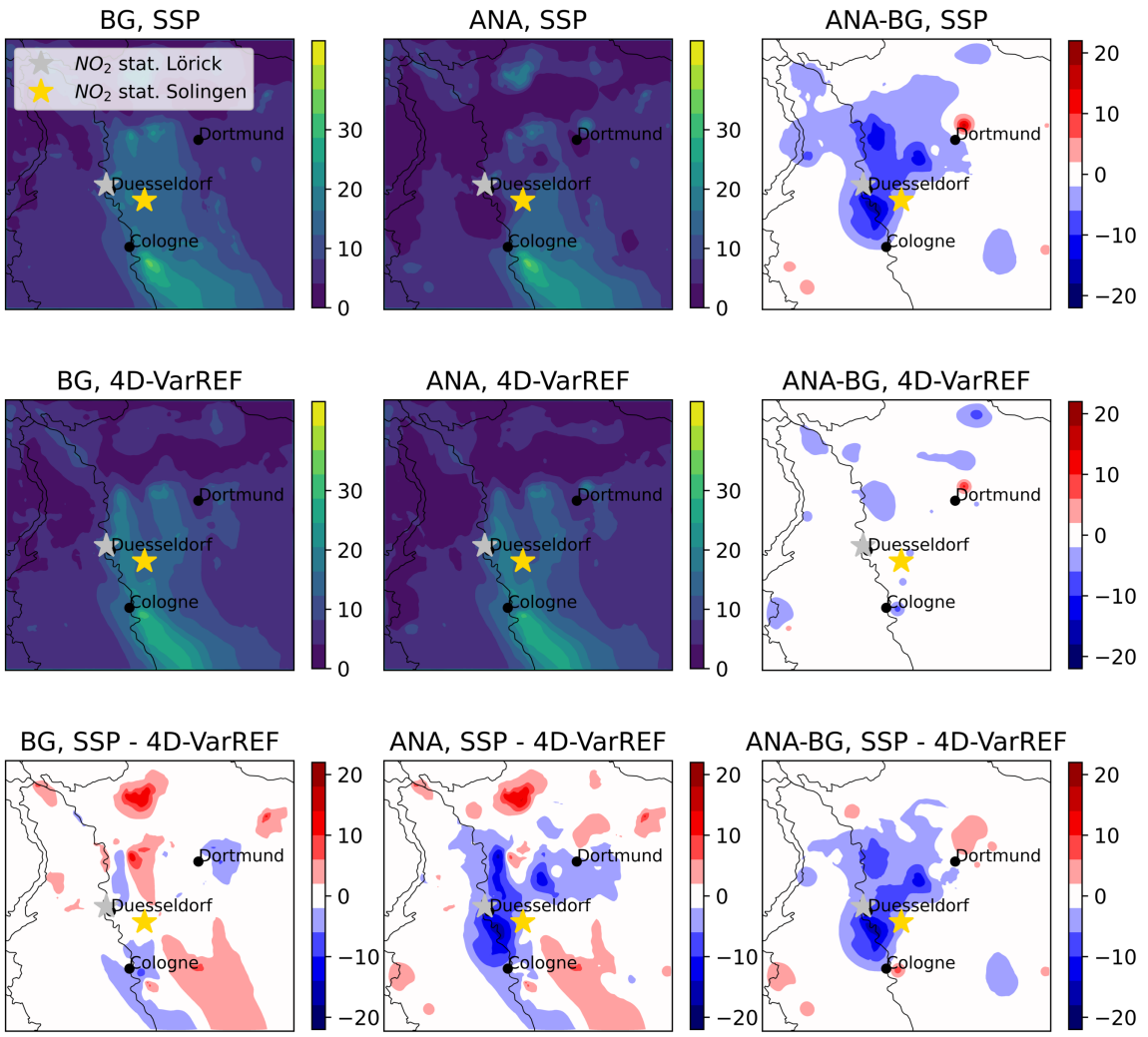

Figure B.2.: Simulated initial NO_2 mixing ratios in the lowest model layer on 19 March 2016 in ppb. They are shown for the background (BG) and the analysis (ANA) of the 4D-VarREF and the sector specific (SSP) optimisation. ANA-BG denotes the difference between analysis and background, which corresponds to the initial value correction. SSP - 4D-VarREF denotes the difference between the SSP and the 4D-VarREF simulation. "SSP-normal" denotes the difference between the simulation using the sector specific optimisation and the simulation using the current data assimilation system of the EURAD-IM (4D-VarREF). The locations of the stations in Solingen and Düsseldorf-Lörick are depicted as gold and silver stars.

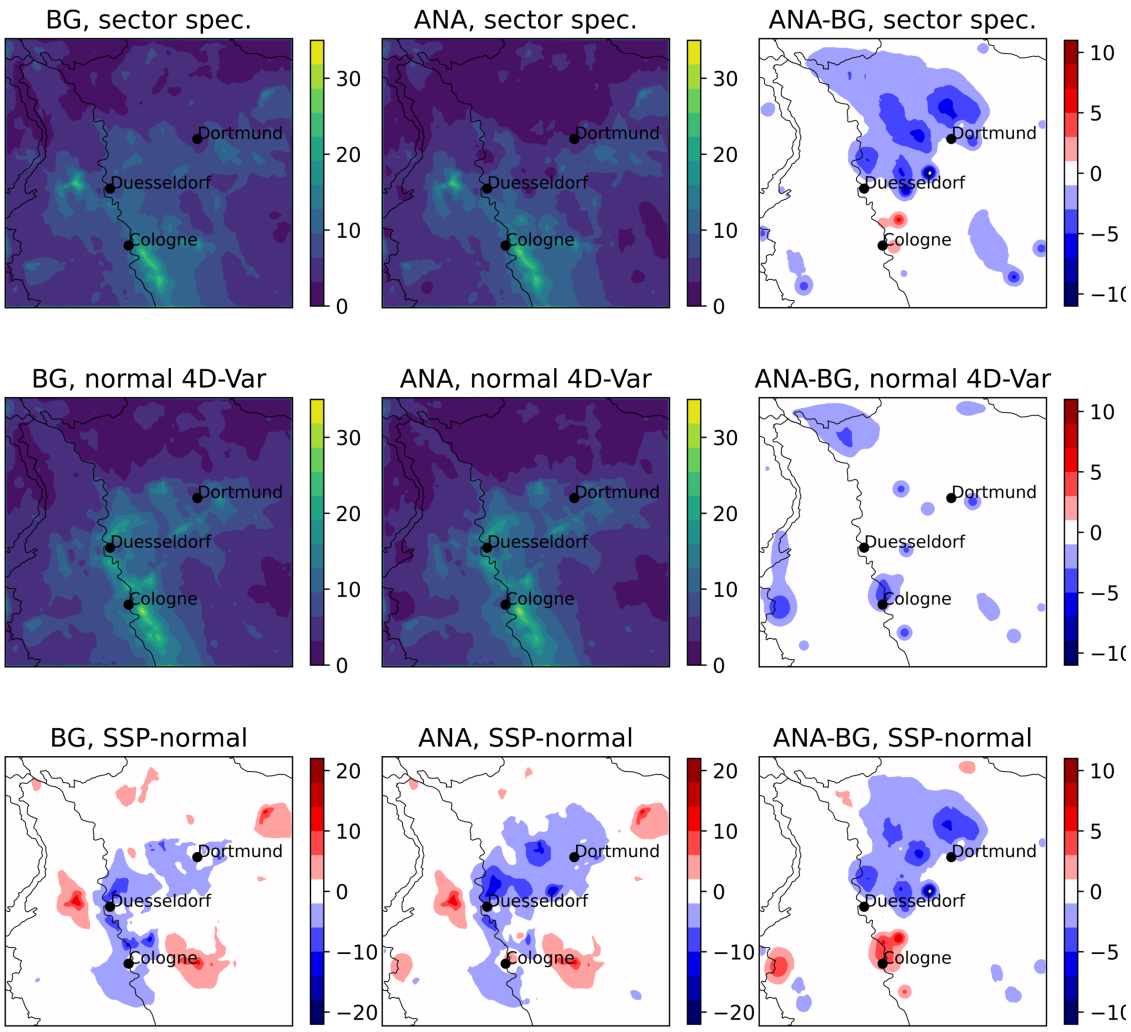

Figure B.3.: Simulated initial NO_2 mixing ratios in the lowest model layer on 20 March 2016 in ppb. They are shown for the background (BG) and the analysis (ANA) of the 4D-VarREF and the sector specific (SSP) optimisation. ANA-BG means the difference between analysis and background, which corresponds to the initial value correction. "SSP-normal" denotes the difference between the simulation using the sector specific optimisation and the simulation using the current data assimilation system of the EURAD-IM (4D-VarREF).

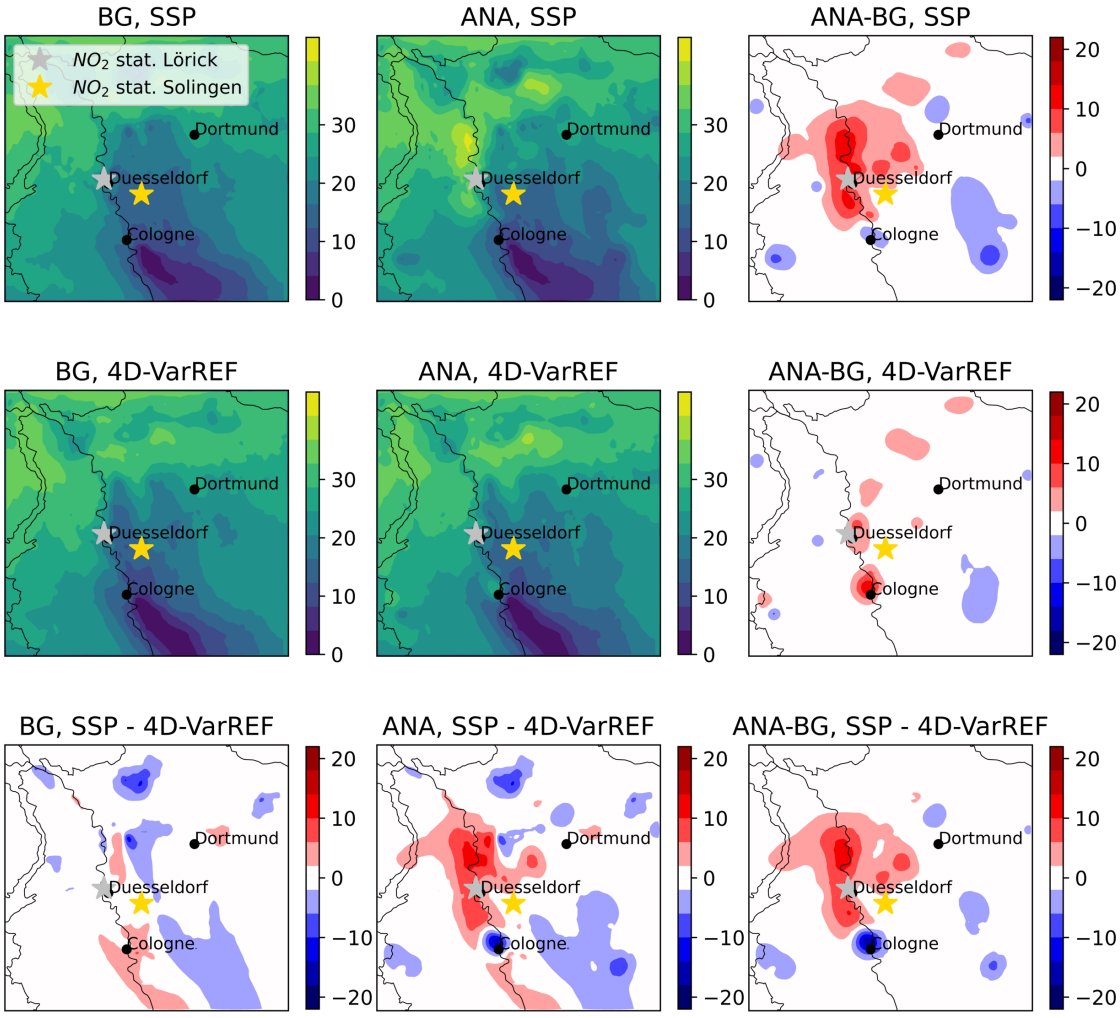

Figure B.4.: Simulated initial O_3 mixing ratios in the lowest model layer on 19 March 2016 in ppb. They are shown for the background (BG) and the analysis (ANA) of the 4D-VarREF and the sector specific (SSP) optimisation. ANA-BG means the difference between analysis and background, which corresponds to the initial value correction. "SSP-normal" denotes the difference between the simulation using the sector specific optimisation and the simulation using the current data assimilation system of the EURAD-IM (4D-VarREF).

B.3. Time series of PM2.5

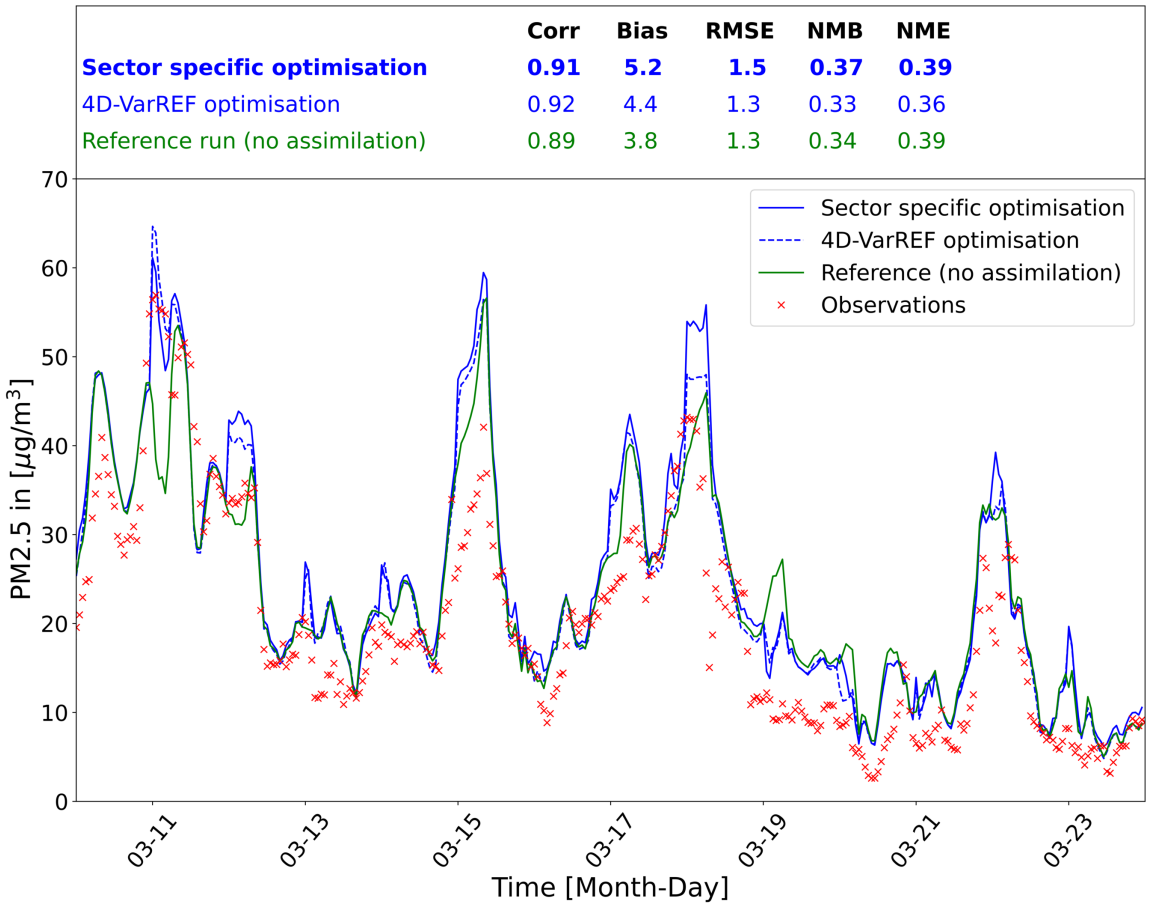

Figure B.5.: Time series of $PM_{2.5}$ concentrations averaged over all validation stations. The observations are depicted as red crosses, the reference run without data assimilation as green solid line, the analysis of the sector specific optimisation as blue solid line and the analysis of the 4D-VarREF optimisation as blue dashed line. The correlation ("Corr"), the bias in µg/m³, the unbiased root mean squared error ("RMSE") in µg/m³, the normalised mean bias ("NMB") and the normalised mean error ("NME") - each with respect to the observations - are given for the time series of the three simulations.

B.4. High emission corrections of the aviation sector

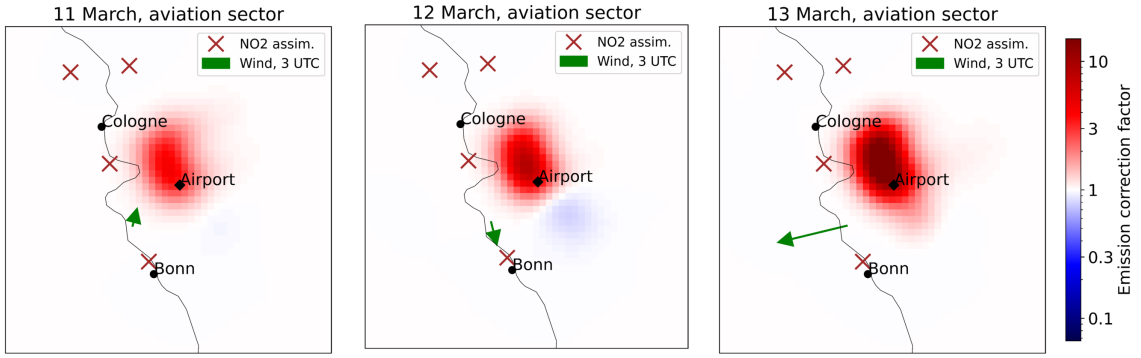

Figure B.6.: Emission correction factors of the aviation sector from 11 to 13 March. A 51 km × 51 km part of the NRW domain is shown, with Cologne/Bonn Airport in the centre. The wind direction at 3 UTC is depicted as a green arrow, the NO_2 assimilation stations as a brown cross.

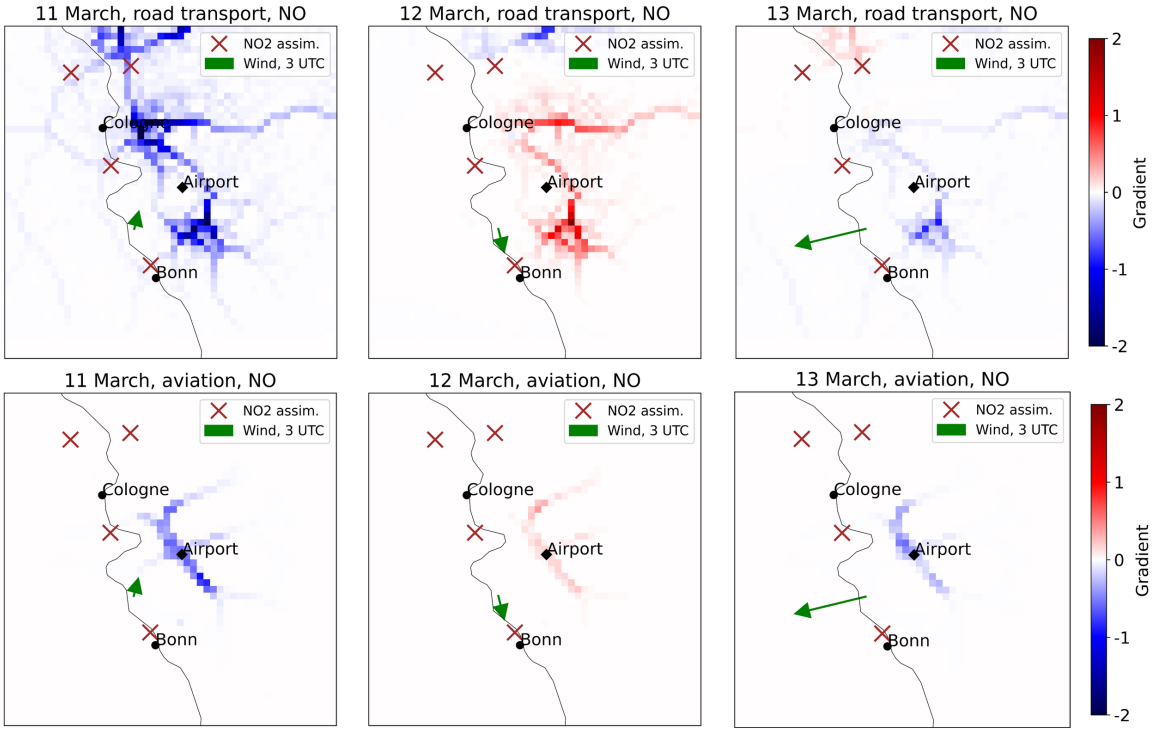

Figure B.7.: First iteration NO gradient of the aviation sector from 11 to 13 March. A 51 km × 51 km part of the NRW domain is shown, with Cologne/Bonn Airport in the centre. The wind direction at 3 UTC is depicted as a green arrow, the NO_2 assimilation stations as a brown cross.

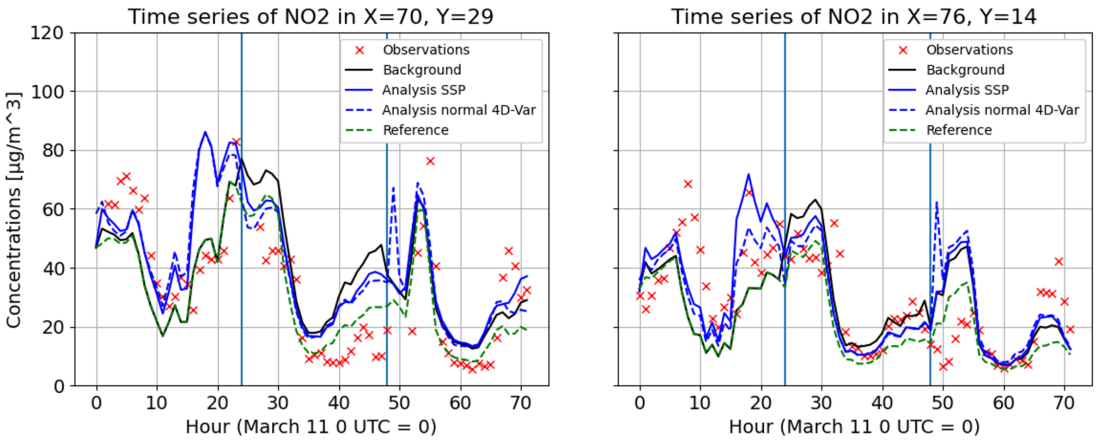

Figure B.8.: Time series of NO_2 concentrations from 11 to 13 March at two assimilation stations ~10 km north north-west of Cologne and in Bonn. The observations are depicted as red crosses, the reference simulation without data assimilation in green, the background in black, the SSP analysis as blue solid line and the 4D-VarREF as blue dashed line. The dashed blue line shows the 4D-VarREF concentrations.

B.5. Enhancement of agriculture II emissions

As stated in Section 7.2, large emission enhancements for the agriculture II sector in an area northeast of Cologne are determined by the sector specific optimisation. Furthermore, the spatially averaged emission correction factor for the domain of up to 1.4 does not decrease after 18 March 2016, in contrast to many other sectors (Fig. 7.4). In the following discussion, the evolution of the emission correction factors from 13 March to 16 March is investigated. A 51 km × 51 km area northeast of Cologne is considered. This area contains 5 NO_2 and 3 PM_{10} assimilation stations. Wind directions from east prevail during this period, so that the emissions are transported from the investigated source region mainly to the observation stations in Chorweiler and Leverkusen north of Cologne.

Fig. B.9 shows the emissions of the agriculture II and road transport sectors on 14 March as an representative example for the simulation period. The emissions from the agriculture II sector are homogeneously distributed in areas with low population.

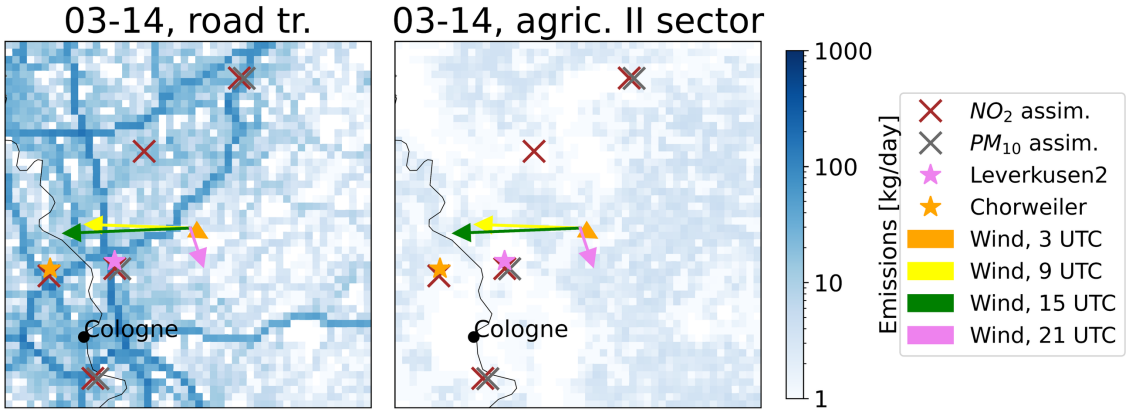

Figure B.9.: NO_x emissions on 14 March shown for the agriculture II and the road transport. A part of the NRW domain with a high agricultural emission correction is displayed. This cut-out includes 51 km × 51 km. The wind directions at 3, 9, 15 and 21 UTC are depicted as a arrows. The locations of the stations in Chorweiler and Leverkusen are depicted as stars. The locations of the other NO_2 assimilation stations are depicted as brown crosses, the locations of PM_{10} assimilation as grey crosses.

In Fig. B.10, the time evolution of the emission correction factors are depicted. The emission correction factors of the agriculture II sector strongly increase from 13 March to 16 March 2016. The emission corrections of the road transport are characterised by strong reductions in the western and enhancements in the eastern part of the investigated area. The emission reduction north of Cologne weakens from 13 March to 14 March. The enhancement of road transport emissions in the eastern part decreases

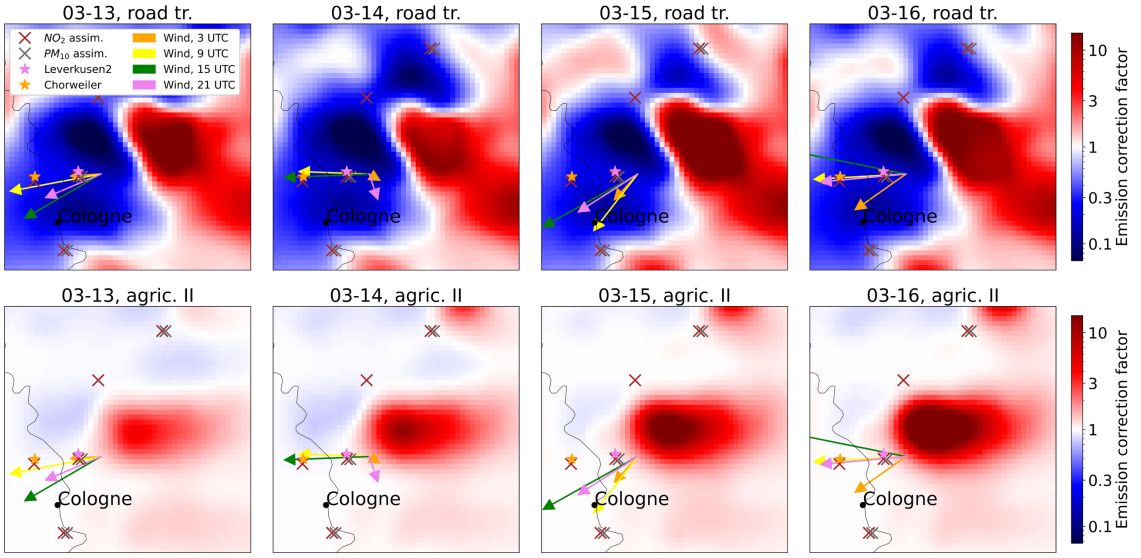

Figure B.10.: Emission correction factors for the agriculture II and the road transport sector from 13 March to 16 March. A part of the NRW domain with a high agricultural emission correction is displayed. This cut-out includes 51 km × 51 km. The wind directions at 3, 9, 15 and 21 UTC are displayed as arrows. The locations of observation the stations in Chorweiler and Leverkusen are depicted as stars, the locations of the other NO_2 assimilation stations as brown crosses and the locations of PM_{10} assimilation as grey crosses.

from 13 March to 14 March, but strongly increases from 14 March to 15 March.

The NO_2 concentrations at the stations in Chorweiler and Leverkusen are shown in Fig. B.11. The locations of the stations are depicted in Fig. B.10. At the station in Chorweiler, on 14 March from 5 UTC to 17 UTC, on 15 March from 7 UTC to 13 UTC and on 16 March from 7 UTC to 14 UTC, the concentrations of all simulations are below the observations. From 14 March, 21 UTC to 15 March, 5 UTC, the simulated concentrations are above the observed concentrations. The simulated concentrations deviate significantly from each other in the last hours of 14 March. At that time, the deviation from the observations is higher for the 4D-VarREF simulation with up to 35 µg/m^3 than that of the SSP concentrations with up to 14 µg/m^3. Also at the station in Leverkusen, the difference between observed and simulated concentrations is large. On 14 March from 1 UTC to 8 UTC and on 16 March from 22 to 23 UTC, the concentrations of the analysis of the sector specific optimisation are up to 43 µg/m^3 lower than the observations. The NO_2 deviations between model and observations are related to a high overprediction of O_3 concentrations at the same time. This discrepancy could be caused by unrepresented ozone-consuming chemical reactions with VOCs. Another possible reason could be a high representativeness error due to the location of the observation station approximately 400 m west of a motorway.

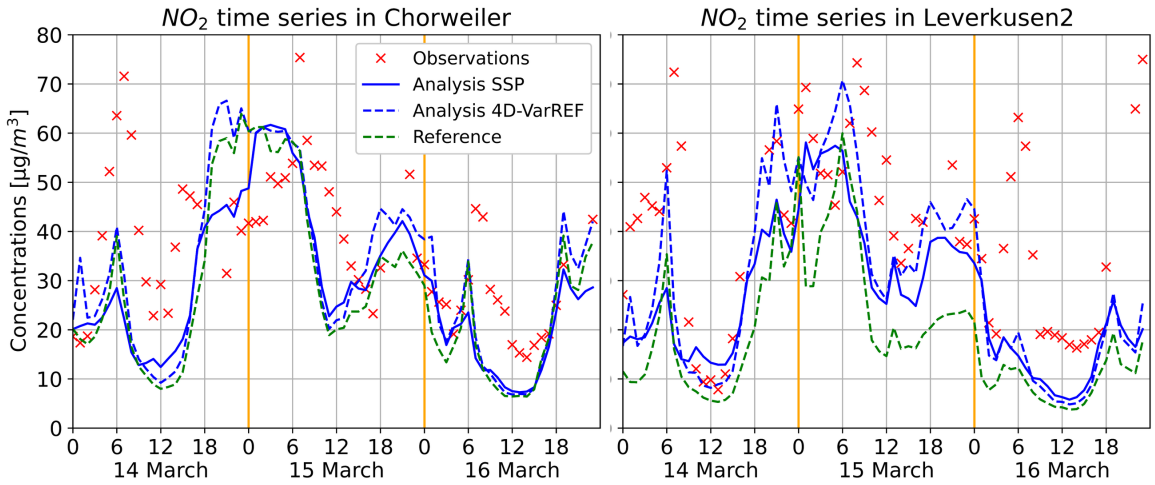

Figure B.11.: Time series of NO_2 concentrations from 14 March to 16 March at the observation stations in Chorweiler and Leverkusen. The observations are depicted as red crosses and the reference run as green dashed line. The dashed blue line shows the concentrations of the 4D-VarREF analysis, while the analysis of the sector specific optimisation is shown as a solid blue line.

The NO gradient fields of the agriculture II and the road transport sector as well as the NH_3 gradient fields of the agriculture II sector are depicted in Fig. B.12. On 15 March and 16 March, a slightly positive NH_3 gradient can be seen in the center of the considered area, while it is negligible on 14 March. Thus, the NH_3 gradient does not cause the increase of the emission corrections of the agriculture II sector. The negative NO gradient is significantly stronger than the positive NH_3 gradient, especially on 15 March and 16 March. It is a consequence of the deviation between observed and simulated NO_2 concentrations at the stations in Chorweiler and Leverkusen, mentioned above. As a result, positive emission corrections for the agriculture II sector are obtained by the sector specific optimisation. The NO gradient for road transport is also mostly negative. It is particularly strong near the negative NO gradient for the agriculture II sector. The NO gradient for the road transport sector is strongest on 15 March due to the east-northeast wind, which transports the emissions of the investigated area to the considered observation stations. The gradient decrease on 16 March is caused by the changed wind direction from east-northeast at 3 UTC to east at 9 UTC. In contrast, the gradient of the agriculture II sector does not significantly decrease on that day. This can be explained by the different diurnal emission cycles of the agriculture II and the road transport sector. Low traffic emissions are assumed around 3 UTC (Fig. 5.8), when the east-northeast wind transports the emissions from the investigated source area to the station in Leverkusen. The agriculture II sector has a flatter diurnal emission profile than the road transport sector and thus emits a moderate amount of NO_x at that time. As a consequence, the NO gradient with

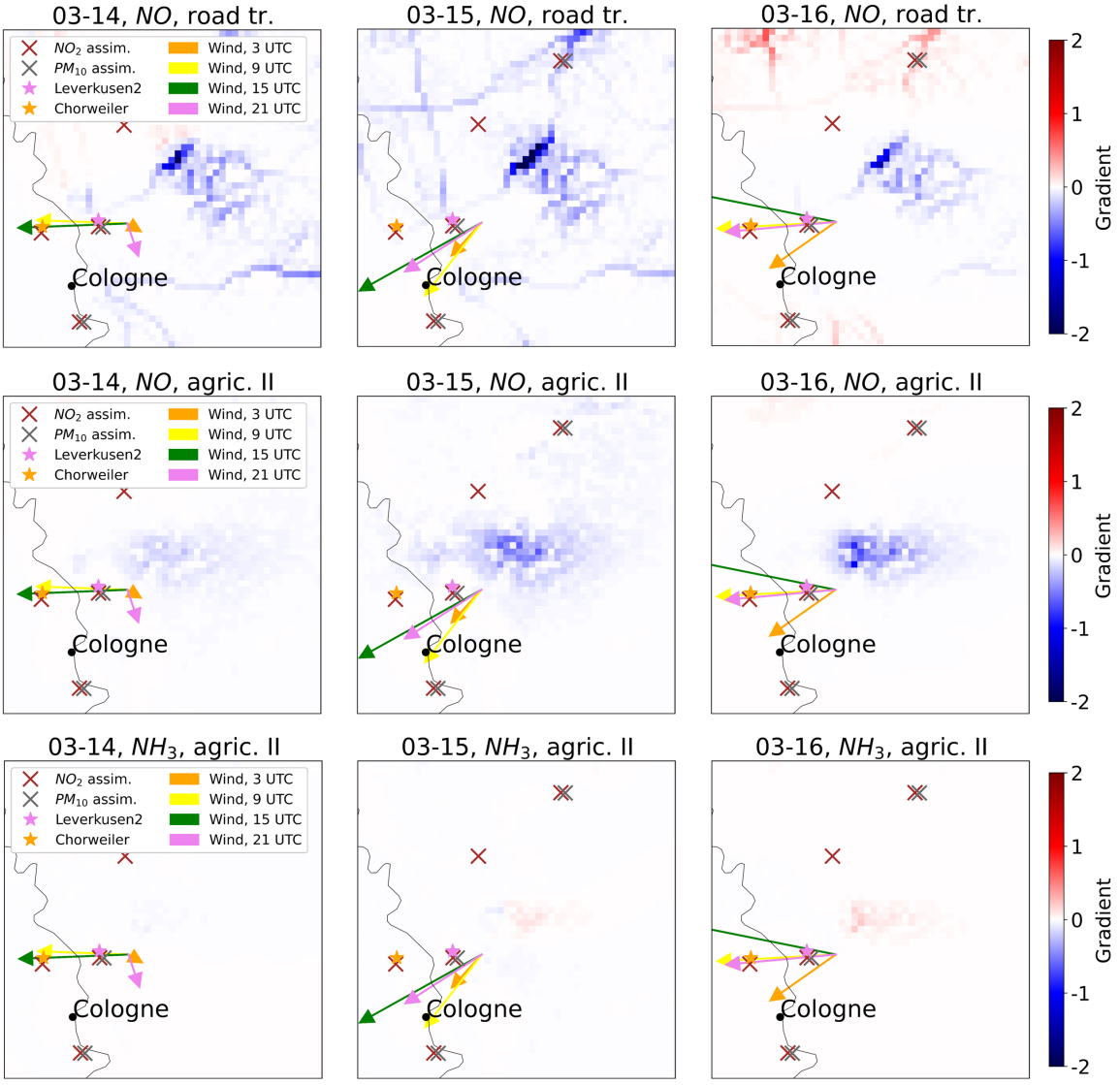

Figure B.12.: First iteration NO and NH_3 gradients of the agriculture II and the road transport sector from 14 March to 16 March. A part of the NRW domain with a high agricultural emission correction is displayed. This cut-out includes 51 km × 51 km. The wind directions at 3, 9, 15 and 21 UTC are displayed as arrows. The locations of the observation stations in Chorweiler and Leverkusen are depicted as stars, the locations of the other NO_2 assimilation stations as brown crosses and the locations of PM_{10} assimilation stations as grey crosses.

respect to the road transport sector is lower on 16 March than on 15 March in contrast to that of the agriculture II sector.

This section revealed two reasons for the strong corrections of agricultural emis-

sions. The first is the generally large difference between observed and simulated NO_2 concentrations at the stations in Chorweiler and Leverkusen. The second is the particularly large discrepancy between observed and simulated NO_2 concentrations in the first hours of 16 March. These increments are attributed to agricultural emissions by the data assimilation. This demonstrates the ability of the sector specific data assimilation to exploit the different diurnal profiles of the sectors. However, the emission corrections for the agriculture II sector probably compensate for incorrect assumptions in the model or input or for observation uncertainty. For example, the diurnal profile of the sectors could deviate from real emissions, e.g. the night-time traffic emissions could be underestimated. Another reason for the difference between model and observations could be unrepresented chemical processes in the EURAD-IM model with an impact on the simulated NO_2 concentrations. In addition, the representativeness error at the observation station in Leverkusen could play an important role. Note that all these uncertainty sources also apply to emission optimisations with the current data assimilation system of the EURAD-IM (Section 3.2) and with data assimilation systems of other models.

Similar statements can be made about the reasons for the large emission corrections for the aviation sector at the airport Cologne/Bonn. The corresponding figures can be found in Section B.4.

B.6. Distinction between road transport and industrial emissions

In NRW, the road transport and the industry sector are two strong pollution sources, with different properties. In the identical twin experiments (Chapter 6), it is shown that the sector specific optimisation is able to effectively distinguish between increasing industrial and decreasing traffic emissions. In this section, this is further investigated for the real-case study (Chapter 7). A 51 km × 71 km part of the domain around Cologne and Düsseldorf is considered in the period from 11 March to 14 March. This part contains 8 NO_2 observation stations. Winds from northeasterly directions prevail. In the selected part of the NRW domain and during the selected period, emission increases for the industry sector and at the same time emission decreases for the road transport sector are obtained by the sector specific optimisation.

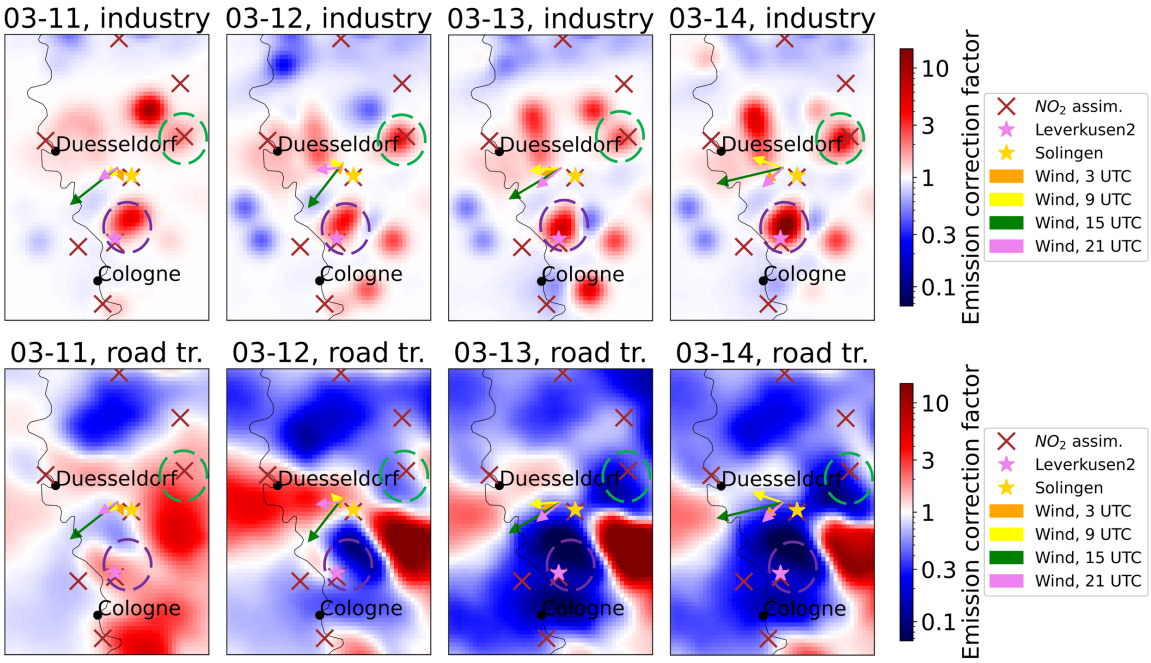

Figure B.13.: Emission correction factors of the industry and the road transport sector from 11 March to 14 March. A part of the NRW domain with 51 km × 71 km is shown. The wind directions at 3, 9, 15 and 21 UTC are displayed as arrows. The locations of the observation stations in Leverkusen and Solingen are depicted as violet and gold stars, the locations of the other NO_2 assimilation stations as brown crosses, area 1 as a violet and area 2 as a green circle.

Fig. B.13 displays the emission correction factors obtained by the sector specific optimisation. On average, the industrial emissions in the considered part are enhanced by the sector specific optimisation. There are three areas where the corrections signifi-

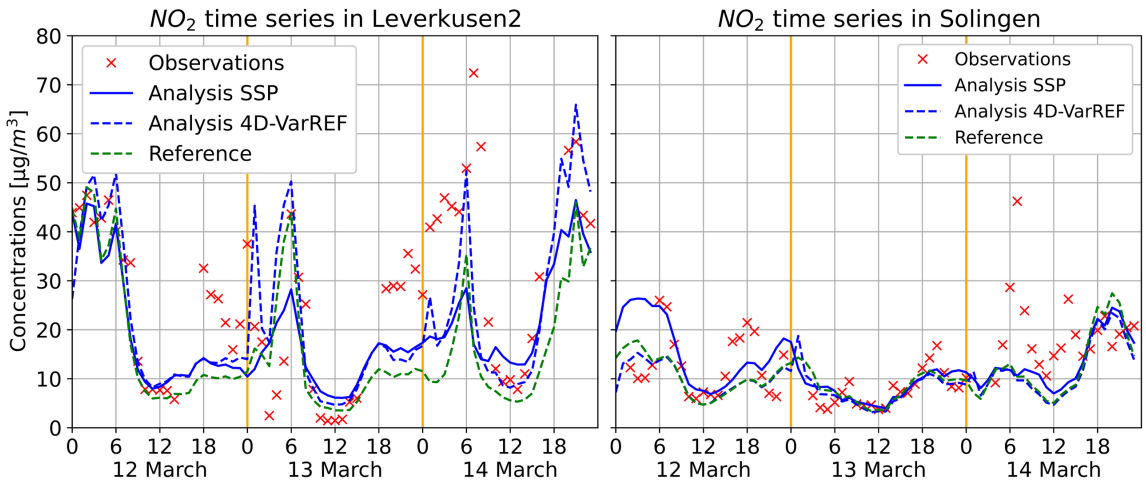

Figure B.14.: Time series of NO_2 concentrations at the observation stations in Leverkusen and Solingen from 12 March to 14 March. The observations are depicted as red crosses and the reference run as green dashed line. The dashed blue line shows the concentrations of the analysis of the 4D-VarREF optimisation, while the analysis of the sector specific optimisation is shown as solid line.

cantly increase during the investigated period, especially from 13 March to 14 March. West of the Rhine river and in the northern part, the emissions are decreasing. The distribution of the emission corrections for the road transport sector are more variable in time than those for the industry sector. Most notable are the changes north of Cologne from emission increases on 11 March to strong emission decreases on the following three days. This means that there are areas in the investigated part of the NRW domain, where emissions of the industry sector are increased from 11 March to 14 March and at the same time emissions of the road transport sector are decreased. In the following, two areas with this behaviour are discussed.

One area (”area 1”) is located north of the NO_2 observation station in Leverkusen, and the second area (”area 2”) approximately 20 km north-east of the observation station in Solingen. The locations of both areas and observation stations are depicted in Fig. B.13. The stations are selected so that the emissions in area 1 and area 2 have a significant impact on the NO_2 concentrations observed at the stations. The time series of the observed concentrations at both stations are shown in Fig. B.14. At the station in Leverkusen, the concentrations of the reference simulation without data assimilation are significantly below the observed concentrations on average. Examples are the lower simulated concentrations from 12 March, 18 UTC to 13 March, 0 UTC (up to 25 µg/m^3) and from 13 March, 19 UTC to 14 March, 7 UTC (up to 35 µg/m^3). The analysis concentrations of the sector specific optimisation are slightly higher than those of the reference simulation without data assimilation. The time series at the station in Solingen show that the concentrations of the reference simulation are

significantly lower than the observations on 12 March from 6 UTC to 8 UTC and from 16 UTC to 19 UTC (up to 12 µg/m^3 in both intervals) and on 14 March from 5 UTC to 15 UTC (up to 36 µg/m^3). In all three intervals, the simulated concentrations are significantly higher in the analysis of the sector specific optimisation than in the reference simulation without data assimilation.

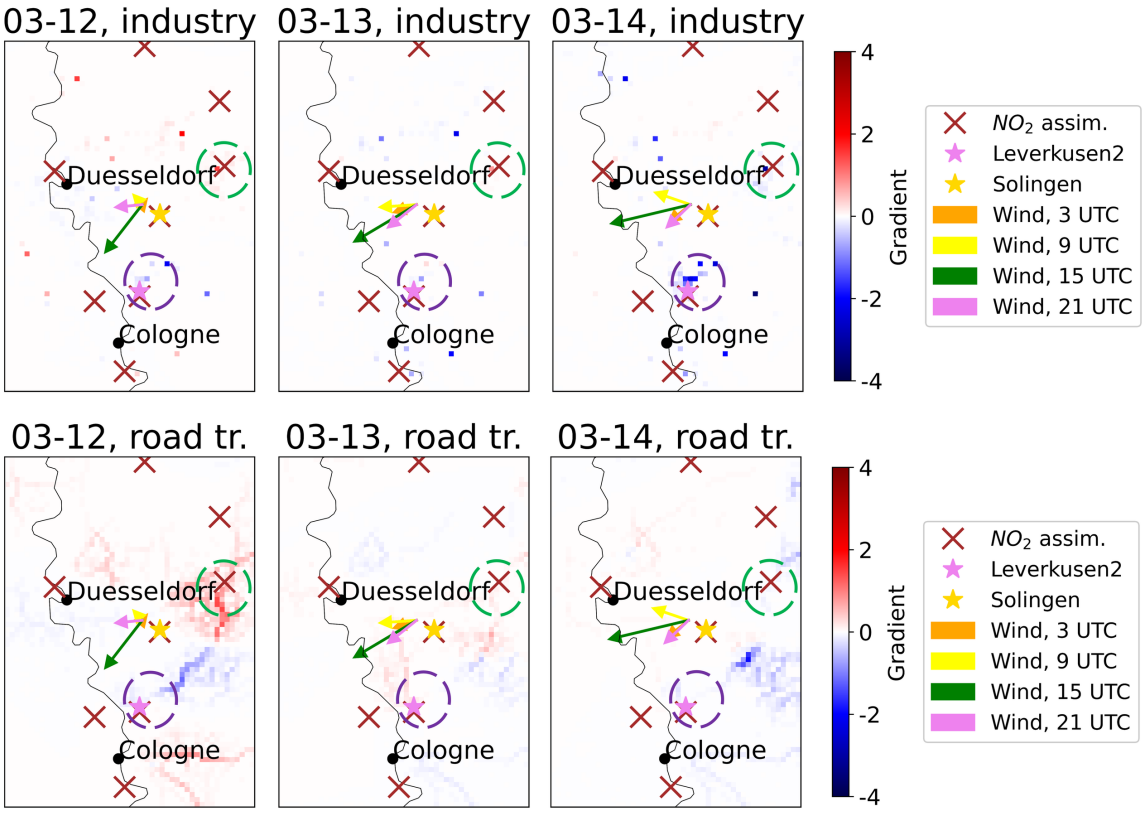

Figure B.15.: First iteration NO gradients of the industry and the road transport sectors from 12 March to 14 March. A part of the NRW domain with 51 km × 71 km is shown. The wind directions at 3, 9, 15 and 21 UTC are displayed as arrows. The locations of the observation stations Leverkusen and Solingen are depicted as violet and gold stars, the locations of the other NO_2 assimilation stations as brown crosses, area 1 as a violet and area 2 as a green circle.

The NO gradients of both sectors after the first adjoint run are shown in Fig. B.15. As discussed in previous sections, the gradient of other species play a minor role. In area 1, the road traffic gradient is slightly negative on 12 March and 14 March and slightly positive on 13 March. The gradient of the industry sector is negative from 12 March to 14 March, but strongest on 14 March, resulting in an emission enhancement in area 1. This reflects the difference of up to 35 µg/m^3 between observed and Ref concentrations in Leverkusen in the first hours on 14 March. In comparison, the negative gradient of

the road transport sector in area 1 is significantly weaker on that day. This is due to the assumed diurnal emission cycles of the sectors. While low traffic emissions are assumed from 1 UTC to 6 UTC, the industry sector is characterised by a flatter diurnal emission profile in comparison to the road transport sector (Fig. 5.8). In area 2, on 12 March and 13 March, the gradients of both sectors are positive, with higher values on 12 March. On 14 March, the road transport gradient is negligible. Nevertheless, the road transport emissions are significantly decreased after the first iteration. The reason for this is that the gradient changes its sign in the following iterations. In contrast, there is a significant negative gradient. As a consequence, the industrial emissions are increased in area 2.

To conclude, in the sector specific data assimilation, the spatial separation of the industrial and the traffic emission sources is effectively exploited. In addition, there are areas with positive emission corrections for the industry sector and at the same time negative emission corrections for the road transport sector. The investigated case demonstrates the ability of the data assimilation system to exploit the different diurnal emission profiles of the road transport and the industry sector. The same applies to the distinction between emissions from road transport and public power. The chemical composition of the emissions plays a minor role, due to the lack of CO observations and weak SO_2 and PM_{10} gradients. The assimilation of CO and CO_2 observations could improve the distinction between road transport, public power and industrial emissions.

Bibliography

Alvarez, D.: JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre, Journal of large-scale research facilities, 7, https://doi.org/10.17815/jlsrf-7-183, 2021.

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Viterbo, P., Cardinali, C., and Thépaut, J.-N.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results, Q. J. R. Meteorol. Soc., 124, 1831–1860, https://doi.org/10.1002/qj.49712455004, 1998.

Andres, R. J., Boden, T. A., and Higdon, D. M.: Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example, Atmos. Chem. Phys., 16, 14 979–14 995, https://doi.org/10.5194/acp-16-14979-2016, 2016.

Asif, Z., Chen, Z., Wang, H., and Zhu, Y.: Update on air pollution control strategies for coal-fired power plants, Clean Technol. Environ. Policy, 24, 2329–2347, https://doi.org/10.1007/s10098-022-02328-8, 2022.

Bouttier, F. and Courtier, P.: Data assimilation concepts and methods. Meteorological training course lecture series. ECMWF, 1999.

Briggs, G. A.: Plume rise and buoyancy effects, Atmospheric science and power production, 327, 366, 1984.

Cao, H., Henze, D., Zhu, L., Shephard, M., Cady-Pereira, K., Dammers, E., Sitwell, M., Heath, N., Lonsdale, C., Bash, J., Miyazaki, K., Flechard, C., Fauvel, Y., Kruit, R., Feigenspan, S., Bruemmer, C., Schrader, F., Twigg, M., Leeson, S., and Capps, S.: 4D-Var Inversion of European NH3 Emissions Using CrIS NH3 Measurements and GEOS-Chem Adjoint With Bi-Directional and Uni-Directional Flux Schemes, J. Geophys. Res. Atmos., 127, https://doi.org/10.1029/2021JD035687, 2022.

Cheewaphongphan, P., Chatani, S., and Saigusa, N.: Exploring Gaps between Bottom-Up and Top-Down Emission Estimates Based on Uncertainties in Multiple Emission Inventories: A Case Study on CH4 Emissions in China, Sustainability, 11, 2054, https://doi.org/10.3390/su11072054, 2019.

Chen, J. and Hoek, G.: Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis, Environ. Int., 143, 105 974, https://doi.org/10.1016/j.envint.2020.105974, 2020.

Collin, G.: Regional Production, Updated Documentation Covering All Regional Operational Systems and the ENSEMBLE: Following U2 Upgrade, Tech. rep., METEO-FRANCE, URL https://atmosphere.copernicus.eu/sites/default/files/2020-09/CAMS50_2018SC2_D2.0.2-U2_Models_documentation_202003_v2.pdf, 2020.

Copernicus: CLC 2012, URL https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012, 2023a.

Copernicus: In situ observations, URL https://atmosphere.copernicus.eu/situ-observations, 2023b.

Denier van der Gon, H. A. C., Hendriks, C., Kuenen, J., Segers, A., and Visschedijk, A. J. H.: Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns, Tech. rep., TNO, Princetonlaan 6, 3584 CB Utrecht, The Netherlands, URL https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC_TNO_del_1_3_v2.pdf, 2011.

Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys. Res. Atmos., 102, 15 967–15 985, https://doi.org/10.1029/97JD01213, 1997.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, https://doi.org/10.5194/acp-7-3749-2007, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res. Oceans, 99, 10 143–10 162, https://doi.org/10.1029/94JC00572, 1994.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE, Control Syst., 29, 83 – 104, https://doi.org/10.1109/MCS.2009.932223, 2009.

Fisher, M.: Assimilation techniques (3): 3dVar, Meteorological Training Course Lecture Series, URL https://www.ecmwf.int/node/16932, 2002.

Fuller, R., Landrigan, P., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., and Yan, C.: Pollution and health: a progress update, Lancet Planet. Health, 6, e535–e547, https://doi.org/10.1016/S2542-5196(22)00090-0, 2022.

Gillijns, S., Barrero Mendoza, O., Chandrasekar, J., De Moor, B., Bernstein, D., and Ridley, A.: What is the ensemble Kalman filter and how well does it work?, in: Proc. Am. Control Conf., vol. 1–12, pp. 4448–4453, https://doi.org/10.1109/ACC.2006.1657419, 2006.

Grange, S. K., Farren, N. J., Vaughan, A. R., Rose, R. A., and Carslaw, D. C.: Strong Temperature Dependence for Light-Duty Diesel Vehicle NOx Emissions, Environ. Sci. Technol., 53, 6587–6596, https://doi.org/10.1021/acs.est.9b01024, pMID: 31094196, 2019.

Granier, C., Darras, S., Denier van der Gon, H., Jana, D., Elguindi, N., Bo, G., Michael, G., Marc, G., Jalkanen, J.-P., Kuenen, J., Liousse, C., Quack, B., Simpson, D., and Sindelarova, K.: The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version), Research report, Copernicus Atmosphere Monitoring Service, https://doi.org/10.24380/d0bn-kx16, 2019.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: 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, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Hass, H., Jakobs, H. J., and Memmesheimer, M.: Analysis of a regional model (EURAD) near surface gas concentration predictions using observations from networks, Meteorol. Atmos. Phys., 57, 173–201, https://doi.org/10.1007/BF01044160, 1995.

Hooghiemstra, P. B., Krol, M. C., Meirink, J. F., Bergamaschi, P., van der Werf, G. R., Novelli, P. C., Aben, I., and Röckmann, T.: Optimizing global CO emission estimates using a four-dimensional variational data assimilation system and surface network observations, Atmos. Chem. Phys., 11, 4705–4723, https://doi.org/10.5194/acp-11-4705-2011, 2011.

Huang, G., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., Guizzardi, D., Muntean, M., Schaaf, E., and Friedrich, R.: Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970–2012, Atmos. Chem. Phys., 17, 7683–7701, https://doi.org/10.5194/acp-17-7683-2017, 2017.

Huangfu, P. and Atkinson, R.: Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis, Environ. Int., 144, 105 998, https://doi.org/10.1016/j.envint.2020.105998, 2020.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Jakobs, H. J., Feldmann, H., Hass, H., and Memmesheimer, M.: The Use of Nested Models for Air Pollution Studies: An Application of the EURAD Model to a

SANA Episode, J. Appl. Meteorol. Climatol., 34, 1301 – 1319, https://doi.org/10.1175/1520-0450(1995)034<1301:TUONMF>2.0.CO;2, 1995.

Jia, G., Huang, Z., Tang, X., Ou, J., Lu, M., Xu, Y., Zhong, Z., Sha, Q., Wu, H., Zheng, C., Deng, T., Chen, D., He, M., and Zheng, J.: A meteorologically adjusted ensemble Kalman filter approach for inversing daily emissions: A case study in the Pearl River Delta, China, J. Environ. Sci., 114, 233–248, https://doi.org/10.1016/j.jes.2021.08.048, 2022.

Jiang, Z., Jones, D., Worden, H., and Henze, D.: Sensitivity of top-down CO source estimates to the modeled vertical structure in atmospheric CO, Atmos. Chem. Phys., 15, 1521–1537, https://doi.org/10.5194/acp-15-1521-2015, 2015.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, https://doi.org/10.5194/bg-9-527-2012, 2012.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, pp. 168–177, Cambridge University Press, https://doi.org/10.1017/CBO9780511802270, 2002a.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, pp. 12–15, Cambridge University Press, https://doi.org/10.1017/CBO9780511802270, 2002b.

Kalnay, E., Hunt, B., Ott, E., and Szunyogh, I.: Ensemble forecasting and data assimilation: two problems with the same solution?, p. 157–180, Cambridge University Press, https://doi.org/10.1017/CBO9780511617652.008, 2006.

Kuenen, J., Dellaert, S., Visschedijk, A., Jonkers, S., and Denier van der Gon, H.: D81.1.1.1 : European emissions dataset (2016), Tech. rep., TNO, 2018.

Lahoz, W. and Schneider, P.: Data Assimilation: Making Sense of Earth Observation, Front. Environ. Sci., 2, 1–28, https://doi.org/10.3389/fenvs.2014.00016, 2014.

Lee, J., Kim, T., Pang, I.-C., and Moon, J.-H.: 4DVAR Data Assimilation with the Regional Ocean Modeling System (ROMS): Impact on the Water Mass Distributions in the Yellow Sea, Ocean Sci. J., 53, 165–178, 2018.

Lee, K. K., Spath, N., Miller, M. R., Mills, N. L., and Shah, A. S.: Short-term exposure to carbon monoxide and myocardial infarction: A systematic review and meta-analysis, Environ. Int., 143, 105 901, https://doi.org/10.1016/j.envint.2020.105901, 2020.

Li, Z., Chao, Y., McWilliams, J., and Ide, K.: A Three-Dimensional Variational Data Assimilation Scheme for the Regional Ocean Modeling System, J. Atmos. Ocean. Technol., 25, 2074–2090, https://doi.org/10.1175/2008JTECHO594.1, 2008.

Lin, J.-T., McElroy, M. B., and Boersma, K. F.: Constraint of anthropogenic NO_x emissions in China from different sectors: a new methodology using multiple satellite retrievals, Atmos. Chem. Phys., 10, 63–78, https://doi.org/10.5194/acp-10-63-2010, 2010.

Liu, D. C. and Nocedal, J.: On the Limited Memory BFGS Method for Large Scale Optimization, Math. Program., 45, 503–528, https://doi.org/https://doi.org/10.1007/BF01589116, 1989.

Liu, F., Zhang, Q., van der A, R., Zheng, B., Tong, D., Yan, L., Zheng, Y., and He, K.: Recent reduction in NOx emissions over China: Synthesis of satellite observations and emission inventories, Environ. Res. Lett., 11, 114 002, https://doi.org/10.1088/1748-9326/11/11/114002, 2016.

Meirink, J. F., Bergamaschi, P., and Krol, M.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: Method and comparison with synthesis inversion, Atmos. Chem. Phys., 8, 6341–6353, https://doi.org/10.5194/acp-8-6341-2008, 2008.

Memmesheimer, M., Tippke, J., Ebel, A., Hass, H., Jakobs, H. J., and Laube, M.: On the use of EMEP emission inventories for European scale air pollution modeling with the EURAD model, in: Proceedings of the EMEP workshop on Photooxidant Modelling for Long-Range Transport in Relation to Abatement Strategies, pp. 307–324, 1991.

Miyazaki, K., Eskes, H. J., and Sudo, K.: Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 12, 2263–2288, https://doi.org/10.5194/acp-12-2263-2012, 2012.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807–837, https://doi.org/10.5194/acp-17-807-2017, 2017.

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.

OpenStreetMap: Planet OSM, URL `https://planet.osm.org`, 2023.

Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., and Ciapponi, A.: Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis, Environ. Int., 142, 105 876, https://doi.org/10.1016/j.envint.2020.105876, 2020.

Orellano, P., Reynoso, J., and Quaranta, N.: Short-term exposure to sulphur dioxide (SO2) and all-cause and respiratory mortality: A systematic review and meta-analysis, Environ. Int., 150, 106 434, https://doi.org/10.1016/j.envint.2021.106434, 2021.

Paschalidi, Z.: Inverse modelling for the tropospheric chemical state estimation by 4-dimensional variational data assimilation from routinely and campaign platforms, Ph.D. thesis, University of Cologne, 2015.

Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X., Wang, J., and Keller, M.: Monthly top-down NOx emissions for China (2005–2012): A hybrid inversion method and trend analysis, J. Geophys. Res. Atmos., 122, 4600–4625, https://doi.org/ https://doi.org/10.1002/2016JD025852, 2017.

Rabier, F. and Liu, Z.: Variational data assimilation: theory and overview, in: Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Reading, UK, September 8–12, pp. 29–43, 2003.

Reikard, G.: Volcanic emissions and air pollution: Forecasts from time series models, Atmospheric Environment: X, 1, 100 001, https://doi.org/https://doi.org/10.1016/j.aeaoa.2018.100001, 2019.

Requia, W., Amini, H., Mukherjee, R., Gold, D., and Schwartz, J.: Health impacts of wildfire-related air pollution in Brazil: a nationwide study of more than 2 million hospital admissions between 2008 and 2018, Nature Communications, 12, https://doi.org/10.1038/s41467-021-26822-7, 2021.

Sacha Brun, A., Pui Hung, C., Fournier, A., Jouve, L., Talagrand, O., Strugarek, A., and Hazra, S.: A solar cycle 25 prediction based on 4D-var data assimilation approach, Proc. Int. Astron. Union, 15, 138–146, https://doi.org/ 10.1017/S1743921320003993, 2019.

Schneider, C., Pelzer, M., Toenges-Schuller, N., Nacken, M., and Niederau, A.: ArcGIS basierte Lösung zur detaillierten, deutschlandweiten Verteilung (Gridding) nationaler Emissionsjahreswerte auf Basis des Inventars zur Emissionsberichterstattung, Umweltbundesamt, Dessau-Roßlau, Germany, 2016.

Schwinger, J.: Four-dimensional variational data assimilation for estimation of the atmospheric chemical state from the tropopause to the lower mesosphere, Ph.D. thesis, University of Cologne, 2006.

Shaddick, G., Thomas, M., Mudu, P., Ruggeri, G., and Gumy, S.: Half the world's population are exposed to increasing air pollution, NPJ Clim. Atmos. Sci., 3, 23, https://doi.org/10.1038/s41612-020-0124-2, 2020.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., and Powers, J. G.: A Description of the Advanced Research WRF Version 3,

Tech. Rep. NCAR/TN-475+STR, University Corporation for Atmospheric Research, https://doi.org/10.5065/D68S4MVH, 2008.

Slingo, J. and Palmer, T.: Uncertainty in Weather and Climate Prediction, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 369, 4751–67, https://doi.org/10.1098/rsta.2011.0161, 2011.

Smith, S. J., van Aardenne, J., Klimont, Z., Andres, R. J., Volke, A., and Delgado Arias, S.: Anthropogenic sulfur dioxide emissions: 1850–2005, Atmos. Chem. Phys., 11, 1101–1116, https://doi.org/10.5194/acp-11-1101-2011, 2011.

Solazzo, E., Crippa, M., Guizzardi, D., Muntean, M., Choulga, M., and Janssens-Maenhout, G.: Uncertainties in the Emissions Database for Global Atmospheric Research (EDGAR) emission inventory of greenhouse gases, Atmos. Chem. Phys., 21, 5655–5683, https://doi.org/10.5194/acp-21-5655-2021, 2021.

Thörnig, P.: JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre, Journal of large-scale research facilities, 7, https://doi.org/10.17815/jlsrf-7-182, 2021.

Tørseth, K., Aas, W., Breivik, K., Fjæraa, A. M., Fiebig, M., Hjellbrekke, A. G., Lund Myhre, C., Solberg, S., and Yttri, K. E.: Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972 - 2009, Atmos. Chem. Phys., 12, 5447–5481, https://doi.org/10.5194/acp-12-5447-2012, 2012.

Wang, C., An, X., Hou, Q., Sun, Z., Li, Y., and Li, J.: Development of four-dimensional variational assimilation system based on the GRAPES–CUACE adjoint model (GRAPES–CUACE-4D-Var V1.0) and its application in emission inversion, Geosci. Model Dev., 14, 337–350, https://doi.org/10.5194/gmd-14-337-2021, 2021.

Weaver, A. and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation, Q. J. R. Meteorol. Soc., 127, 1815 – 1846, https://doi.org/10.1002/qj.49712757518, 2001.

Yuan, L., Popoola, O. A. M., Hood, C., Carruthers, D., Jones, R. L., Sun, H. Z., Liu, H., Zhang, Q., and Archibald, A. T.: Improving NO_x emission estimates in Beijing using network observations and a perturbed emissions ensemble, Atmos. Chem. Phys., 22, 8617–8637, https://doi.org/10.5194/acp-22-8617-2022, 2022.

Zheng, X., Orellano, P., Lin, H., Jiang, M., and Guan, W.: Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: A systematic review and meta-analysis, Environ. Int., 150, 106 435, https://doi.org/10.1016/j.envint.2021.106435, 2021.

Persönliche Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die meine Promotionsarbeit begleitet und mich dabei unterstützt haben.

Zunächst möchte ich meinen Dank an beide Betreuer*innen meiner Promotion richten. Prof. Dr. Astrid Kiendler-Scharr bin ich dankbar für die Vergabe des interessanten Themas, die freundlichen und konstruktiven Gespräche sowie die immer kooperative Zusammenarbeit. Aus sehr traurigen Gründen konnte sie die Betreuung nicht zu Ende führen. Prof. Dr. Hendrik Fuchs übernahm diese Aufgabe und arbeitete sich sehr kurzfristig in das Thema meiner Promotionsarbeit ein. Seine Anmerkungen im Rahmen des Korrekturlesens dieser Dissertation waren sehr hilfreich.

Prof. Dr. Susanne Crewell möchte ich für ihr Interesse an meiner Arbeit und ihre Bereitschaft zur Übernahme des Zweitgutachtens danken.

Mein besonderer Dank gilt Anne Caroline Lange und Philipp Franke für ihre tatkräftige Unterstützung vom ersten bis zum letzten Tag. Unter anderem haben sie mir durch ihre Erklärungen und Anregungen die Einarbeitung in das EURAD-IM-Modell und die Datenassimilation erleichtert und mit mir viele produktive Diskussionen über die Entwicklung der Methode, die Gestaltung der Studien sowie die Interpretation der Simulationsergebnisse geführt. Ihre Expertise und ihr Engagement waren auch beim Schreiben dieser Arbeit sehr hilfreich.

Auch Felix Wieser und Prof. Dr. Andreas Wahner haben mit Kommentaren und Vorschlägen zur Dissertation großartige Unterstützung geleistet, wofür ich mich ebenfalls sehr bedanken möchte.

Ich erinnere mich gerne an alle Kolleg*innen am Institut IEK-8 im Forschungszentrum Jülich und bedanke mich für die interessierte, hilfsbereite und freundliche Atmosphäre. Aufgrund der eingeschränkten Präsenz während der Corona-Pandemie war die gemeinsame Zeit am Institut leider sehr kurz.

Nicht zuletzt sage ich auch Danke an meine Familie und meine Freunde. Sie haben mir während meiner Promotionsarbeit eine angenehme Zeit bereitet und mich auch in anstrengenden und schwierigen Zeiten unterstützt.

Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

14.04.2023 Backes

Datum, Pascal Martin Backes