Estimating NOx emissions using S5-P TROPOMI

An adjoint-free 4DVAR approach

by

Tammo Zijlker

to obtain the degree of Master of Science

in Applied Mathematics,

specialisation Computational Science & Engineering

at Delft University of Technology,

to be defended publicly on *August 11, 2020*.

Student number:	4425375	
Project duration:	December 9, 2019 – August 11, 2020	
Thesis committee:	Prof. dr. ir. A. W. Heemink	TU Delft, responsible professor
	Dr. ir. A. J. Segers,	TNO, daily supervisor
	Prof. dr. ir. M. Verlaan	TU Delft, member of thesis committee
	Prof. dr. ir. C. Vuik	TU Delft, member of thesis committee

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft Delft University of Technology

TNO

Abstract

This study aims to improve estimates of NO_x emission strengths by assimilation of TROPOMI satellite retrievals in the LOTOS-EUROS chemical transport model. Nitrogen oxides (NO and NO_2) play a pivotal role in atmospheric chemistry, are an important source of air pollution and contribute to nitrogen deposition over vulnerable natural areas. Therefore, it is paramount to have accurate estimates of emissions.

Emissions are parameterised by multiplicative correction factors for NO_x emission strengths from existing inventories. Optimal estimates of the correction factors are calculated by assimilation of TROPOMI NO_2 retrievals in LOTOS-EUROS.

This study proposes an adjoint-free approach to solving the 4DVAR data assimilation problem. Due to the near linearity of LOTOS-EUROS with respect to NO_2, an approximate model is proposed that calculates NO_2 concentrations from the background state and a linear combination of the influences of the parameters on the state. This approximate model is calculated from an ensemble of LOTOS-EUROS simulations with perturbed parameters. After substitution of the approximate model in the 4DVAR cost function, it is quadratic and the minimum can be calculated directly. For this approximate cost function, the optimal estimate of the parameters and the covariance of this estimation can be obtained with negligible computational costs.

Twin experiments, where synthetic satellite observations are assimilated, show that the adjoint-free 4DVAR method is able to accurately minimise the cost function. Errors in estimated parameters are in agreement with the covariance calculated for the estimate. It was also shown that by using domain decomposition, it is possible to generate the approximate model from fewer simulations of LOTOS-EUROS and thereby increasing the computational efficiency of the method.

In experiments using TROPOMI NO_2 retrievals, the method performs well when modelled plumes align with the retrievals. However, differences between modelled plumes and retrievals, that are resolved by the high resolution of the TROPOMI instrument, may strongly hamper results as the method is only able to correct the intensity of the plumes but not their positions. This leads to an underestimation of NO_x emission strengths.

Further research is required to handle differences of plume positions in LOTOS-EUROS and TROPOMI retrievals to apply the method to actual TROPOMI retrievals. In addition, more research into domain decomposition may further increase computational efficiency of the method.

Keywords — TROPOMI, Inverse Modelling, Data Assimilation, LOTOS-EUROS, Emission, NO_2

Preface

This thesis is part of the MSc programme of applied mathematics, in the specialisation of Computational Science & Engineering at Delft University of Technology. During this project, I worked as an intern at the department of Climate, Air and Sustainability of TNO (Netherlands Organisation for Applied Scientific Research). The project lasted from December 2019 until August 2020.

First of all, I would like to thank the CAS department for allowing me to work on this project during my graduation. Many thanks to all the colleagues at TNO for the warm welcome at the department, the many lunch meetings and coffee breaks. I would also like to thank Arjo Segers, my daily supervisor at TNO, for his help and guidance during the project. Already having switched to Skype long before the Coronavirus arrived in Holland, I count myself lucky to receive such great supervision during the pandemic. Finally, I would like to thank Arnold Heemink for his help during orientation on thesis projects, his enthusiasm and feedback during the project.

Enjoy the read!

Tammo Zijlker
Delft, August 2019

Contents

List of Symbols

$\mathbf{A}$	Averaging kernel
$\mathbf{B}$	Background covariance matrix
$\boldsymbol{\beta}$	Parameter vector
$\boldsymbol{\beta}^b$	Background parameter vector
$\hat{\boldsymbol{\beta}}$	Estimated parameter vector
$\boldsymbol{\beta}_{\text{true}}$	Twin experiment true parameter vector
$\mathbf{E}$	Jacobian of state $\boldsymbol{x}$ with respect to parameters $\boldsymbol{\beta}$
$\boldsymbol{\epsilon}$	Observation error
$\boldsymbol{\eta}$	Reduced order model parameter vector
$\mathbf{H}_m$	Observation operator at time t_m
J	4DVAR cost function
M	Air-mass factor
$\mathbf{M}$	State evolution function
M^{strat}	Air-mass factor stratosphere
M^{trop}	Air-mass factor troposphere
n	Dimension of state vector
$n_{\text{nlev}}^{\text{CTM}}$	Number of vertical levels of the chemical transport model
$n_{\text{nlev}}^{\text{TROP}}$	Number of vertical levels of TROPOMI averaging kernel
N_o	Number of observation times
N_p	Number of parameters
$N_{\text{p,red}}$	Dimension of reduced parameter vector
n_{red}	Dimension of reduced state vector
N_s	NO_2 slant column density
N_s^{strat}	NO_2 stratospheric slant column density
N_s^{trop}	NO_2 tropospheric slant column density
N_v	NO_2 vertical column density
N_v^{strat}	NO_2 stratospheric vertical column density
N_v^{trop}	NO_2 tropospheric vertical column density
$\boldsymbol{\Phi}_\beta$	Parameter POD projection matrix
$\boldsymbol{\Phi}_x$	State POD projection matrix
$\mathbf{R}_m$	Observation error covariance matrix at t_m
$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$	Approximate covariance of estimate
$\mathbf{x}$	State vector
$\mathbf{X}$	Snapshot matrix
$\mathbf{x}^b$	Background state vector trajectory
$\hat{\mathbf{x}}$	Approximate model state vector
$\mathbf{y}$	Observation vector
$\mathbf{z}$	Reduced order model state vector

Acronyms

4DVAR	Four-Dimensional variational data assimilation
AMF	Air-mass factor
CAMS	Copernicus Atmosphere Monitoring Service
CBM-IV	Carbon Bond Mechanism-IV
CTM	Chemical transport model
DOAS	Differential Optical Absorption Spectroscopy
EC	European Commission
ESA	European Space Agency
LOTOS-EUROS	LOng Term Ozone Simulation - EURopean Operational Smog model
MF	Multiplication Factor
NSO	Netherlands Space Organisation
POD	Proper Orthogonal Decomposition
S5P	Sentinel-5 Precursor
SVD	Singular Value Decomposition
TNO	Dutch organisation for Applied Scientific Research
TPWL	Trajectory Piece-wise Linearisation
TROPOMI	Tropospheric Monitoring Instrument
VCD	Vertical Column Density
VMR	Volume Mixing Ratio

1

Introduction

Nitrogen oxides (NO and NO_2, aggregated as NO_x) are trace gases that play a pivotal role in atmospheric chemistry. NO is the main oxide of nitrogen formed during high-temperature combustion by oxidation of either atmospheric nitrogen (N_2) or nitrogen in the fuel. Small quantities of NO_2 are produced by these combustion processes too, but NO_2 is mainly formed by oxidation of NO in the atmosphere (Seinfeld and Pandis 2016).

In figure 1.1, the main chemical reactions of nitrogen compounds are illustrated. NO_2 plays an important role in the deposition of nitrogen to soil after oxidation to HNO_3. The ecological effects of nitrogen deposition cause significant damage to ecosystems. The Dutch government is forced by law to limit or compensate for this damage. NO_x is also important in the production of ozone in the troposphere. During daytime, tropospheric ozone is mainly produced by photolysis of NO_2. Ozone could cause severe injuries to human health and also harms vegetation. NO_2 itself is also an important source of air pollution. Research has indicated long-term exposure to NO_2 can be associated with non-accidental and cause-specific mortality in the Dutch population above the age of 30 (Fischer et al. 2015). For all these reasons, NO_x emissions should be kept as low as possible and emission reduction emission measures are widely implemented.

Over the past year, nitrogen has made it to the headlines of the Dutch news. Nitrogen deposition on vulnerable natural areas exceeds thresholds established by law. Policies proposed by the Dutch government were rejected in court, forcing the Dutch government to take more drastic measures to cut nitrogen emissions to limit the nitrogen deposition on vulnerable nature areas. Such measures have been proposed in the report “Niet alles kan” (Remkes et al. 2019).

These developments have also raised interest further for atmospheric modelling, emission and deposition studies. The report “Meer meten, robuuster rekenen” (Hordijk et al. 2020), released June 15, 2020, gives a review of the current state of emission and deposition measurement and calculation techniques that support policy-making. One of the main recommendations for improving the emission and deposition modelling reported is making use of satellite retrievals of trace gasses to validate and improve models. This is exactly what this study is about. Recent developments in satellite technology, culminating in the launch of the TROPOMI satellite have brought a steady, high-quality stream of measurement data that can be used in data assimilation for improvement of atmospheric quality models.

1.1. Study aim

The purpose of this study is to estimate emission parameters using satellite data. The parameters considered are the emission strengths of NO_x. Accurate estimates of air pollutant emission such as NO_x are fundamental for air quality modelling. Recent developments in satellite technology, culminating in the launch of the Tropospheric Monitoring Instrument (TROPOMI) instrument on the Sentinel-5 Precursor (S5P) satellite have brought a wealth of high-resolution measurement data of NO_2 and other

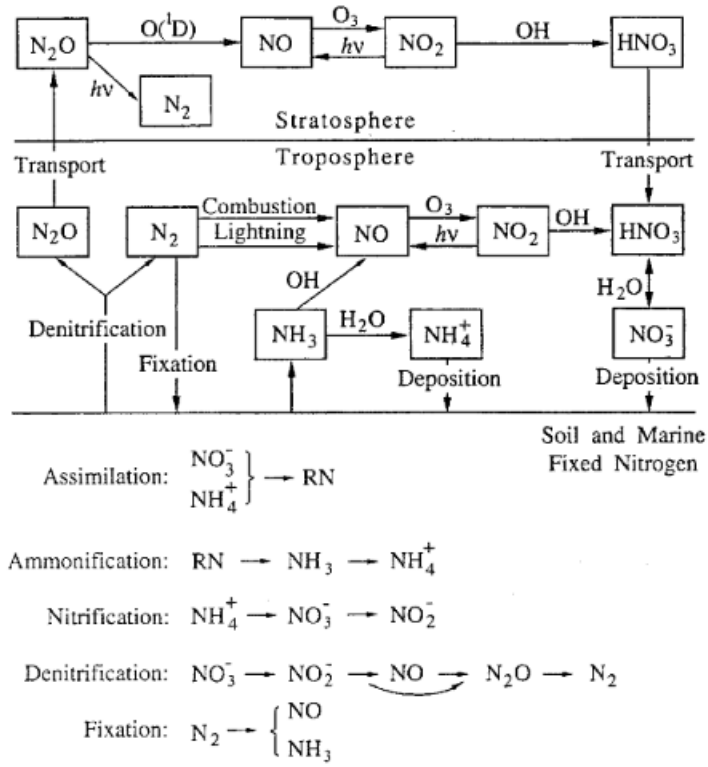


Figure 1.1: Processes in the atmospheric cycle of nitrogen compounds. A species written over an arrow signifies reaction with the species from which the arrow originates obtained from Seinfeld and Pandis 2016.

atmospheric trace gasses. These recent developments give rise to great opportunities for improving the estimation of emission strengths (Bocquet et al. 2015).

Time-averaged columns of NO_2 can relatively easily indicate trends in emissions. However, local meteorological features that dictate the spread of NO_2 are highly variable in time. A Chemical transport model (CTM) is necessary to accurately link the actual emissions to the satellite observations. This study makes use of the LOng Term Ozone Simulation - EURopean Operational Smog model (LOTOS-EUROS) air quality model. The estimated parameters are input variables of the atmospheric transport model and, therefore, the problem to solve is an inverse modelling problem. A data assimilation scheme will be developed to perform this parameter estimation task. A variational method, 4DVAR will be used. One specific methodological aim of this study is to develop an assimilation scheme that solves the 4DVAR problem in an adjoint-free way.

1.2. Atmospheric chemistry model LOTOS-EUROS

This study uses the LOTOS-EUROS model. LOTOS and EUROS were developed in the 1990s by the Dutch organisation for Applied Scientific Research (TNO) and the Environmental Assessment Agency of the Dutch National Institute for Public Health and the Environment (RIVM/MNP) and have been used as models for the study of the atmospheric distribution of photo-oxidants. The two models were combined in 2004 to give LOTOS-EUROS version 1.0, used by TNO, RIVM and the Royal Dutch Meteorological Institute (KNMI). The open-source 2.0 version of the model was released in 2016 to extend the community of users.

LOTOS-EUROS has been used for a wide range of applications supporting scientific research, regulatory programmes and air quality forecasts. At present, the LOTOS-EUROS model is used in operational air quality forecasts over Europe (Marécal et al. 2015) and The Netherlands (De Ruyter de Wildt et al. 2011). Chapter 2 gives more background information on the LOTOS-EUROS and the setup of the model in this study.

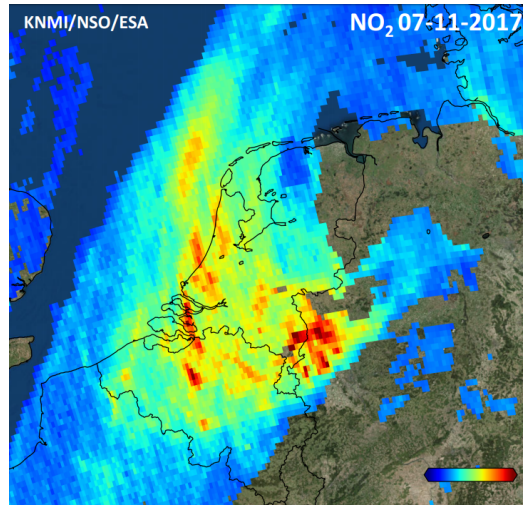


Figure 1.2: Early NO_2 retrievals over the Netherlands by the TROPOMI instrument. Source: www.tropomi.eu

1.3. Satellite observations

In this thesis, observational data from the TROPOMI instrument will be used. TROPOMI was launched as payload onboard the Sentinel 5-Precursor satellite on October 13, 2017. The main objective of the Sentinel-5P mission is to perform atmospheric measurements with a high spatio-temporal resolution. These measurements are to be used for air quality, ozone and UV radiation, as well as climate monitoring and forecasting. Its mission is part of the earth observatory programme Copernicus which is managed by the European Commission (EC) and the European Space Agency (ESA). The TROPOMI instrument was funded by the Netherlands Space Organisation (NSO) and ESA. It was developed by TNO and Airbus DS Netherlands. The KNMI and SRON are the principal investigators (Veefkind et al. 2012).

The TROPOMI instrument provides measurements of atmospheric trace gases and of cloud and aerosol properties at an unprecedented spatial resolution of approximately 5.5×3.5 km^2 (approx. 7×3.5 km^2 before August 6, 2019). Early results revealed that the instrument was able to resolve individual NO_2 plumes over the Netherlands. In figure 1.2, an example of these early, revolutionary results are shown. The abundance in data has sparked research interest in model validation and inverse modelling. The NO_2 data, for example, are being used for model validation (Ialongo et al. 2020) and to derive emissions (Lorente et al. 2019; Wang et al. 2019). Methane columns from TROPOMI have also helped to identify large leaks at oil-producing sites in the U.S. (Zhang et al. 2020). Chapter 3 gives more background information on TROPOMI and satellite observation of atmospheric trace gasses.

1.4. NO_2 as a proxy for CO_2

The applications for improved estimation of NO_x emissions are abundant. One specific application is to investigate whether the co-emission of NO_x and CO_2 may be used to improve anthropogenic (caused by human activity) CO_2 emission estimates. Anthropogenic CO_2 emission is mainly caused by the burning of fossil fuels. Along with these chemical processes, NO_x is produced. NO_x is a far shorter-lived compound than CO_2. Effects of anthropogenic CO_2 emissions are often negligible against the background concentrations, which makes it hard to measure by satellites. Using the co-emission of CO_2 and NO_2 could help to locate and estimate the CO_2 emissions. Curier et al. 2016, show such a procedure. In Reuter et al. 2019, the authors show a method to enhance the plumes of CO_2, observed by the OCO-2 satellite by making use of NO_2 retrievals from TROPOMI. The inversion system developed in this study could be used for the assimilation of data provided by new CO_2 satellites.

ESA is planning the launch of a satellite constellation (CO2M) that will provide measurements of CO_2 for the EU Copernicus program[1]. This CO2M satellite constellation is supposed to measure CO_2 columns at 2 km×2 km with a precision of 0.7 ppm (Meijer 2019). To locate CO_2 plumes, NO_2 measurements can

[1] www.che-project.eu/

be used, as they are a good tracer of anthropogenic carbon emissions (Reuter et al. 2019). The CO2M satellite would also be equipped with sensors for NO_2.

1.5. Data assimilation

This study applies data assimilation to provide estimates of NO_x emission strengths. In data assimilation, observational data are combined with numerical models to give the best estimate of model parameters. The field of data assimilation was initially developed in numerical weather prediction. Broadly speaking, data assimilation methods may be classified into two categories. Deterministic, *variational* methods such as 4DVAR treat the assimilation as an optimal control problem. A cost function based on the misfit between observations and numerical models is minimised. *Stochastic*, filtering methods such as the Kalman filter and ensemble methods statistically estimate parameters recursively.

This thesis investigates the application of variational methods on the assimilation of satellite data. For solving the Four-Dimensional variational data assimilation (4DVAR) problem, a well-known method is to use an iterative, gradient-based optimisation technique, where the gradient is calculated using an adjoint model. This method has proven to be highly efficient for data assimilation. The computational costs of an increase in the number of parameters do not strongly increase if the number of parameters to be estimated grows. This is of high importance for data assimilation schemes as computational resources are always limited and the systems are often extremely high-dimensional. The main drawback of this method is its dependency on the implementation of the adjoint model. This task requires large programming and especially maintenance efforts, that are often not feasible.

Solving the variational data assimilation problem without the adjoint method is possible. In this study, such an adjoint-free, gradient-based method is proposed. This method is non-intrusive: the large programming effort associated with the development of an adjoint model is not required. Moreover, it is attractive because of its scalability of the number of parameters with respect to the computational costs required. The highest computation costs are related to the creation of an ensemble of model simulations. However, using domain decomposition strategy, increasing the number of parameters does not lead to a proportional increase of model evaluations. Chapter 4 provides a detailed treatment of the theory of data assimilation and the methodology proposed for the assimilation of satellite data in this study.

1.6. Research questions

This study aims to improve estimates of NO_x emissions through data assimilation of TROPOMI NO_2 retrievals. The following research questions have to be answered to reach the objective.

Research question 1: How can an adjoint-free 4DVAR procedure be designed to assimilate synthetic TROPOMI retrievals in a twin experiment?

The focus of this thesis is the development of adjoint-free data assimilation method that can be used for the estimation of emission parameters. The proposed method will be tested in a twin experiment. Synthetic satellite NO_2 retrievals are generated from a model run with known emission parameters. An adjoint-free 4DVAR assimilation scheme should estimate these parameters correctly. This experiment should provide a proof of concept for the methodology. As in a twin experiment, the true value of the parameter estimated is known, it is possible to assess the performance of the data assimilation.

Research question 2: How can domain decomposition methods be used to make the procedure scalable in the number of parameters compared to the required number of full order model runs?

When the adjoint method is used to solve the 4DVAR problem, an increase in the number of parameters does not lead to an increase in the number of model simulations needed to calculate the gradient of the cost function, although the number of iterations might grow. However, when using an adjoint-free method, such as finite differences, to calculate the gradient, the number of full order model runs quickly becomes computationally unfeasible.

This thesis investigates whether, through domain decomposition, the efficiency of the method can be increased.

Research question 3: Can the procedure assimilate TROPOMI retrievals to improve estimates of anthopogenic NO_x emission strengths?

Experiments with actual TROPOMI retrievals have to be conducted to test the methodology on real data. Other sources of uncertainty, not captured by the parameters, chosen to be estimated, might have to be dealt with. Such sources of uncertainty could arise from priors in the numerical model, for example, meteorological conditions, temporal profiles of emissions and injection heights of the emissions. Thought will have to be given to the selection of observations to assimilate too.

1.7. Structure

This thesis introduces the LOTOS-EUROS chemical transport model and its state-space representation in Chapter 2. An overview of satellite observations of atmospheric trace gasses is presented in Chapter 3. Chapter 4 is concerned with the mathematical theory of data assimilation. In this chapter, the adjoint-free method for solving the 4DVAR problem investigated in this thesis is introduced. Chapter 5 presents the results of several twin experiments that have been conducted to validate the method. In Chapter 6, experiments are conducted with actual TROPOMI data. Chapter 7 presents the main conclusions of this thesis, followed by recommendations for further research.

2

LOTOS-EUROS

This chapter gives an introduction to the LOTOS-EUROS air quality model used in this study. Firstly, the general model characteristics are described. Secondly, the emission module used in the model is described in more detail, since this study focuses specifically on emissions. Finally, the state space representation of a numerical model is introduced. With the state space representation of LOTOS-EUROS, uniform mathematical notation of the data assimilation theory discussed in Chapter 4 is possible.

2.1. Model formulation

The history and main applications of LOTOS-EUROS were introduced in Section 1.2. This section provides a more technical description of LOTOS-EUROS, adapted from Segers et al. 2019.

2.1.1. Simulated compounds

Concentrations of various atmospheric trace gasses and aerosols, denoted C, can be simulated in LOTOS-EUROS. Some groups of compounds are coupled through chemical reactions such as oxidants, but others can be calculated independently, such as dust. As this thesis focuses on NO_2, the compounds that interact with it through the Carbon Bond Mechanism-IV (CBM-IV) scheme are important. A list of these compounds can be found in Segers et al. 2019.

An aerosol is a suspension of fine particles or droplets in the air. The model distinguishes different aerosol types based on their chemical composition. For NO_2, the nitrate NO_3 aerosol is relevant, since nitrate is one of its oxidation products. Whereas the lifetime of NO_2 is rather short 1-12 hours (Stavrakou et al. 2013), the lifetime of nitrate is longer. This provides a mechanism for transport of nitrogen N emitted as NO_x over longer distances.

A different group of tracers frequently modelled with LOTOS-EUROS is 'primary aerosol'. Primary aerosol consists of PM2.5, PM10-2.5, elementary carbon, particulate organic matter, sea-salt, and dust. The simulation of primary aerosol can be performed independently from the other compounds in LOTOS-EUROS as it is not chemically coupled to other them.

2.1.2. Simulated processes

LOTOS-EUROS simulates the concentrations of atmospheric trace gasses and aerosols. The processes incorporated into the model are:

- **Emissions**

  Emissions are the primary source of trace gases and aerosols in the atmosphere. As this thesis focuses on estimation of emissions, they play a pivotal role in this project. Section 2.2 will provide more details about the emission module in LOTOS-EUROS.

- **Transport**

  Transport of tracers and aerosols is implemented using advection by wind in 3 dimensions, entrainment and vertical diffusion. Advection, transport by wind, is driven by meteorological wind fields that are updated every 3 hours from an offline database. Entrainment results from the changing vertical structure in the atmosphere during the day, which impacts on the vertical levels defined in the model. After updating the vertical structure of the model, tracer concentrations are redistributed by linear interpolation. Entrainment only plays a role when vertical levels are varied in the model. In this study, fixed vertical levels are used. These are defined by the meteorology input data.

- **Chemistry**

  Atmospheric chemistry is highly complex. Hundreds of organic and inorganic compounds interact through thousands of reactions. Modelling each of those explicitly is not feasible. Therefore, the LOTOS-EUROS model has adopted the CBM-IV. This scheme lumps different compounds together and reduces the number of reactions considered to 81.

- **Dry and wet deposition**

  Dry and wet deposition of gasses and aerosols represent an important sink of compounds modelled in LOTOS-EUROS. Wet deposition is the process of tracers getting caught by droplets of water in the atmosphere and eventually being deposited on the surface by gravity. Dry deposition of particles and gasses is a process through which particles and gasses deposit themselves on surfaces directly.

2.1.3. Continuity equation

C denotes the concentration of a trace gas or aerosol. For every concentration C, the processes involved can be described using the following continuity equation:

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} + V\frac{\partial C}{\partial y} + W\frac{\partial C}{\partial z} = \frac{\partial}{\partial z}(K_z\frac{\partial C}{\partial z}) + E + R + Q - D - W \tag{2.1}$$

The large-scale wind components in the west-east, south-north and vertical direction are represented by U,V and W respectively. K_z is the vertical turbulent diffusion coefficient. E denotes entrainment due to variations in layer height. In this study, fixed vertical levels are used so entrainment is disabled. R represents a source or sink term in chemistry. Q is the contribution by emissions. D and W describe loss terms due to dry and wet deposition respectively.

It is important to note that in Equation (2.1), the only non-linear contribution to the partial differential equation (PDE) originate in chemistry. This non-linearity is due to reactions with other tracers. All other processes are effectively linear operators in terms of concentrations. Thus, doubling the initial concentration at some time t_1 will lead to a doubled concentration at the next time step t_2.

2.1.4. Horizontal grid and vertical levels

Continuity Equation (2.1) is discretized in space and time to solve it numerically. The standard horizontal grid resolution of LOTOS-EUROS is 0.5° longitude and 0.25° latitude. Increasing this resolution by a factor 5 is feasible.

Several options are available for defining the vertical structure of the model. In this study, the vertical levels are defined by the meteorological input. Concentrations of tracers are calculated in 10 vertical layers. Concentrations in 5 layers on top of these 10 vertical layers are calculated from global boundary conditions. In total, the LOTOS-EUROS model has 15 vertical layers in its output.

2.1.5. Time steps

The LOTOS-EUROS model typically has an hourly output. However, concentrations are calculated in smaller, intermediate time steps. These intermediate time steps are limited by advection. Within one time step, an air parcel should not cross a complete grid cell.

2.2. Emission database

The emission module of LOTOS-EUROS simulates the introduction of trace gasses and aerosols into the atmosphere from various sources. The module considers different sources:

- **Anthropogenic sources**

  Sources of tracers and aerosols related to human activity are a key factor in emissions. In this study anthropogenic NO_2 emissions from power plants, industrial facilities and other sources play a key role. Therefore, Subsection 2.2.1 gives more details on anthropogenic NO_x emissions.

- **Biogenic sources**

  Emission of various tracers results from various natural processes in the biosphere, these are called biogenic sources. Biogenic NO_x emissions are a source of NO_2. It is mainly produced by bacteria in soil. Biogenic NOx emission is usually only relevant in remote places without human activity since it is quickly exceeded by anthropogenic emissions.

- **Sea-spray sources**

  Through wind and wave action, sea-salt is released from the sea into the atmosphere.

- **Dust sources**

  Emission of dust from the surface can be caused by various processes such as wind erosion and re-suspension of dust by traffic and agricultural activities.

- **Forest fires**

  Forest fires cause extremely high emissions of trace gasses and aerosols. Due to their irregular spatial and temporal occurrence, a dedicated inventory for forest fires is used. Forest fires can add significant quantities of NO_x to the atmosphere in densely forested areas.

2.2.1. Anthropogenic NO_x emissions

Anthropogenic emissions in LOTOS-EUROS are obtained from databases called inventories. These inventories are available from the European Union-funded project Copernicus Atmosphere Monitoring Service (CAMS)[1]. The version used in this study is the TNO/CAMS v2.2 inventory. The following paragraphs show how NO_x emissions are calculated in LOTOS-EUROS from this database. The calculation and the structure of the TNO/CAMS database are the same for the anthropogenic emission of other compounds.

Source categories

Emissions are attributed to 16 different source categories, listed in Table 2.1. Not all of these source categories account for NO_x emissions.

Composition

Emissions of NO and NO_2 are aggregated as NO_x in the database. A constant ratio of NO and NO_2 in the emitted NO_x is assumed. Most of the NO_x (97%) is emitted as NO, only a small fraction (3%) is emitted as NO_2.

Resolution and spatial distribution

The database distinguishes area and point sources. Area emission sources are reported spatially on a regular grid with a resolution of $1/10° \times 1/20°$. An example of an area source is the emission from road transport. This is defined based on maps of road and traffic density. The emission map of these sources shows the road network at the resolution of the inventory, which is detailed enough to resolve the highway network. Such an emission map is shown in Figure 5.4a . Point sources are e.g. power plants and large industrial stacks. They are positioned in the emission database at their actual location.

[1] https://atmosphere.copernicus.eu/

Category	Short Code
Public Power	A
Industry	B
Other stationary combustion	C
Volatile substances	D
Solvents	E
Road transport exhaust gasoline	F1
Road transport exhaust diesel	F2
Road transport exhaust LPG gas	F3
Road transport non-exhaust	F4
Shipping	G
Aviation	H
Off-road	I
Waste	J
Agricultural livestock	K
Agricultural other	L

Table 2.1: Categories of anthropogenic emission in the CAMS database. N.B. Not all these categories account for NO_x emission.

Temporal distribution
The emission data are stored as a spatially distributed, yearly averaged flux. However, emissions vary at different time scales. In LOTOS-EUROS, time factors are used to generate realistic temporal emission profiles. Per source category, time profiles for month-in-the-year, day-in-the-week and hour-of-the-day are available. The profiles are on average 1 so the total emission corresponds to the yearly average, but they can be higher or lower based on for example the season; (winter versus summer) or rush hours. These factors are part of the inventories.

Vertical emission profiles
Some categories of emissions only take place at the surface, such as road transport and agriculture. However, other emissions have a non-trivial vertical distribution; for example, industrial stacks inject emissions at higher altitudes. For each source category, an average vertical emission profile is defined Emissions are distributed over these vertical profiles in LOTOS-EUROS.

2.3. State space representation

A state space representation is a mathematical way of representing a physical system. When the field of data assimilation was developed, many different ways of notation were introduced. Ide and Ghil 1997 introduced self-consistent notation that has been adopted by the community since. In this section, the state space representation of LOTOS-EUROS and the mathematical notation used throughout this thesis is introduced.

2.3.1. State space

The aim of a state space representation is to describe the evolution of the state vector $\mathbf{x}$. $\mathbb{R}^n$, the n-dimensional Euclidean space, denotes the state space of a dynamic model. The state vector $\mathbf{x} \in \mathbb{R}^n$ contains the quantities simulated in the model. $\mathbf{x}$ is obtained through discretisation of the continuity Equation (2.1), and contains the concentrations of all tracers and aerosols, at all grid cells defined in the LOTOS-EUROS. For example, surface concentrations of NO_2 are a part of the state vector. Such surface concentrations are illustrated in Figure 2.1. The state vector $\mathbf{x}$ can be extremely high-dimensional, in the order of millions.

The temporal evolution of the state vector is described with:

$$\mathbf{x}(t_{k+1}) = \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta}), \qquad k = 1, 2, \ldots \tag{2.2}$$

In this equation, $\mathbf{M}$ is a non-linear operator that performs a discrete time step by solving Equation (2.1). The operator $\mathbf{M}$ thus represents the full LOTOS-EUROS model in the state space representation.

Parameter vector $\boldsymbol{\beta} \in \mathbb{R}^{N_p}$ controls the temporal evolution of the system, with N_p denoting the number of parameters. In this thesis, $\boldsymbol{\beta}$ contains emission parameters. The content of $\boldsymbol{\beta}$ will be defined in Section 4.1 and depends on the chosen application. Given a parameter vector $\boldsymbol{\beta}$, the evolution of the state vector over time, $\mathbf{x}_1, \mathbf{x}_2, \ldots$ can be calculated. In data assimilation, this evolution is called the *trajectory* of $\mathbf{x}$.

2.3.2. Observation space

Real observations of the system $\mathbf{y}_m^o$ at time t_m are often available. In the context of LOTOS-EUROS, examples of such observations are ground station measurements, deposition measurements and of course the TROPOMI NO_2 retrievals. In the state space representation of a model, observations are defined as a vector in the n_o-dimensional observation space $\mathbb{R}^{n_o}$.

It is often necessary to compare the model results to these observations. The model equivalent of the observation, called the *simulated observation* $\mathbf{y}_m$, is defined through observation operator $\mathbf{H}_m$. It projects the state vector on the observation space:

$$\mathbf{y}_m = \mathbf{H}_m\left(\mathbf{x}(t_m), \boldsymbol{\beta}\right), \qquad m = 1, 2, \ldots \tag{2.3}$$

$\mathbf{H}_m$ can be non-linear and depend on the parameters $\boldsymbol{\beta}$ studied. In this thesis, the observations are satellite observed NO2 columns and $\mathbf{H}_m$ is therefore associated with satellite retrieval. In Section 3.5, the construction of $\mathbf{H}_m$ for satellite retrievals is described in detail. $\mathbf{H}_m$ will turn out to be a linear operator for this application.

With a perfect model and highly accurate corresponding state $\mathbf{x}$, due to representation errors, a real observation can differ still differ from its simulated counterpart. Differences may arise from measurement noise in the instrument, errors in the retrieval and representation errors. Representation errors are related to the use of a state with discrete elements representing an average concentration in a

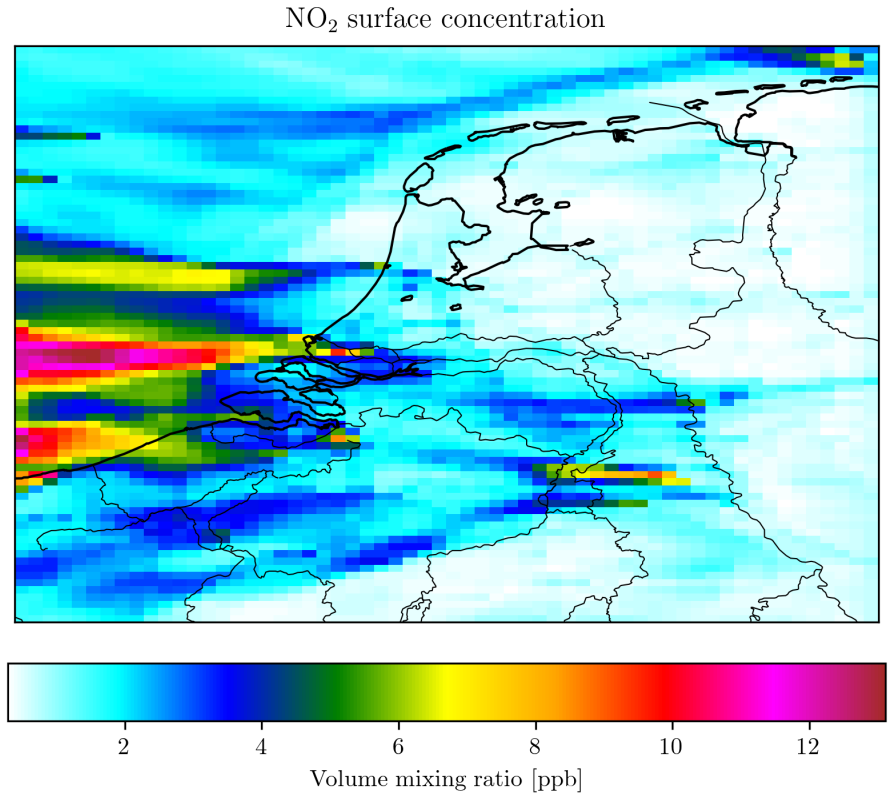


Figure 2.1: Surface concentrations of NO_2 on July 2, 2018 at 10:00 calculated by LOTOS-EUROS. These surface concentrations are part of the state vector $\mathbf{x}$

large grid cell, while in reality concentrations could show strong variations even at small spatial scales. Real observations of the system are therefore linked to the state vector through the following equation:

$$\mathbf{y}_m^o = \mathbf{y}_m + \boldsymbol{\epsilon}_m = \mathbf{H}_m\left(\mathbf{x}(t_m), \boldsymbol{\beta}\right) + \boldsymbol{\epsilon}_m, \qquad m = 1, \ldots, N_o \tag{2.4}$$

where $\boldsymbol{\epsilon}_m$ denotes the observation error at time t_m.

Together, state vector $\mathbf{x}$, parameters $\boldsymbol{\beta}$, the state evolution Equation (2.2) and measurement Equation (2.4) define the state space representation of LOTOS-EUROS in this thesis. This compact form is especially suitable for the discussion of the data assimilation methodology in Chapter 4.

3

TROPOMI NO_2

This chapter introduces the TROPOMI instrument, its measurements and its NO_2 retrieval algorithm. It also describes the concept of averaging kernels, that is essential for assimilation of TROPOMI data into CTMs. Finally, the observation operator that was developed to link LOTOS-EUROS simulations to the TROPOMI retrievals is described.

3.1. The TROPOMI instrument

TROPOMI is an instrument onboard the Copernicus Sentinel-5 Precursor satellite. Figure 3.1 shows an artist's impression of the satellite. The S5P satellite is the first atmospheric composition Sentinel, launched on 13 October 2017, planned for a seven-year mission (Veefkind et al. 2012). S5P is a low earth orbit polar satellite. Because of the polar orbit and wide swath of the scanner, near-daily global coverage is achieved.

TROPOMI measures light in the ultraviolet (UV), the visible (VIS), the near-infrared (NIR) and the shortwave infrared (SWIR) light. In this spectrum, TROPOMI is capable of measuring important atmospheric constituents such as nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO), methane (CH_4), formaldehyde (CH_2O), aerosol properties and clouds.

Figure 3.1: Artist's impression of the Sentinel 5-Precursor satellite

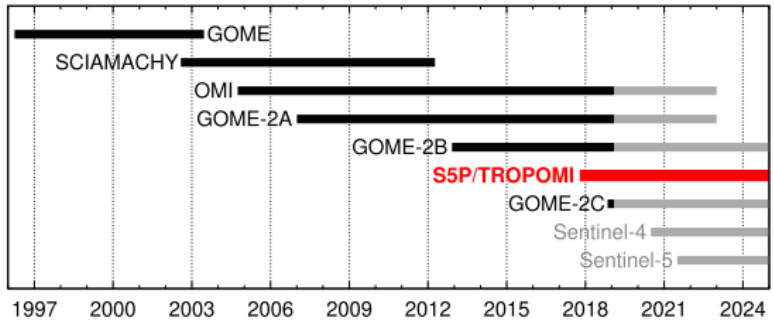


Figure 3.2: Timeline of European UV/Vis backscatter satellite instruments that retrieve tropospheric and stratospheric NO_2. Obtained from Van Geffen et al. 2019.

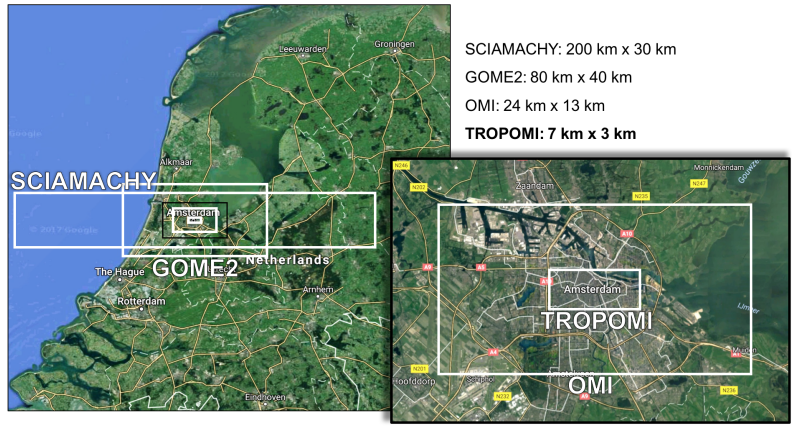


Figure 3.3: Overview of subsequent NO_2 monitoring instrument resolutions. From August 6, 2019 TROPOMI operational resolution was improved to 5.6 km × 3.5 km Source: https://www.tudelft.nl/en/2017/citg/grs/what-makes-tropomi-special/

3.2. NO_2 satellite retrieval history

TROPOMI is not the first instrument to monitor tropospheric concentrations of NO_2. A variety of remote sensing instruments, ground-based, in-situ (such as in weather balloons) and satellite-based - have been and are currently being developed.

Figure 3.2 shows the timeline of the missions of several satellites that monitor tropospheric NO_2. These instruments have provided continuous measurements over the last 25 years. TROPOMI was launched as the precursor to the Sentinel 5 mission to bridge the gap between the retirement of OMI and launch of the Sentinel 5 mission. Each generation of instruments brought improvements in measured spectra, resolution and signal-to-noise ratios. By receiving wider spectra of light, it is possible to measure additional trace gasses in the atmosphere. Increased measurement resolution makes it possible to investigate smaller spatial scales.

Figure 3.3 illustrates the different resolutions of consecutive instruments. From August 6, 2019, onwards, the operational resolution of the TROPOMI instrument in the along-track direction was improved from 7.2 km to 5.6 km. This resolution is truly revolutionary and enables investigation at city block scale. In terms of emission measurement, point sources of NO_x, such as powerplants can now be resolved.

3.3. TROPOMI NO_2 retrieval algorithm

This section gives a brief overview of the TROPOMI NO_2 retrieval algorithm, adapted from Van Geffen et al. 2019. Satellite retrieval of atmospheric trace gasses is complex. High-level overview and a basic understanding of retrieval methodology are necessary to appreciate the benefits and limitations of the data being used for assimilation purposes.

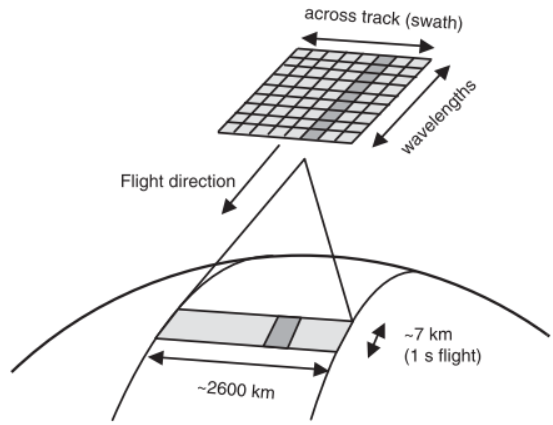


Figure 3.4: TROPOMI measurement principle, obtained from Veefkind et al. 2012. The dark-grey ground pixel is imaged on the two-dimensional detector as a spectrum. All ground pixels in the 2600 km swath are measured at the same time.

The algorithm consists of a three-step procedure.

1. Retrieval of the total NO_2 slant column density using the Differential Optical Absorption Spectroscopy (DOAS) method.
2. Separation of the total slant column density into a stratospheric and tropospheric part, based on information from a data assimilation system.
3. Conversion of slant tropospheric and slant stratospheric column densities into Vertical Column Density (VCD) using Air-mass factors (AMFs).

3.3.1. Slant column density retrieval with DOAS method

TROPOMI uses the DOAS method measurement principle. DOAS stands for Differential Optical Absorption Spectroscopy. In Figure 3.4, the measurement principle of TROPOMI is illustrated. The satellite is pointed towards the earth and measures light reflected from the earth and its atmosphere. The extinction due to scattering and absorbing of photons by tracers changes the reflected spectrum compared to the direct sunlight reaching the satellite. A radiance model is fitted to these different received spectra for the calculation of the NO_2 concentration. This is the DOAS method. The effective absorption by NO_2 along this path is represented as the slant column density of NO_2, N_s.

3.3.2. Separation of slant column density

The light measured in the spectrometer follows a slant trajectory from the sun through the earth's atmosphere and back to the satellite. The DOAS method assumes concentrations of NO_2 to be constant along that trajectory. The path passes through both stratosphere and troposphere. From a meteorological point of view, processes in the stratosphere and troposphere are very different. It is necessary to separate the slant column into a stratospheric (N_s^{strat}) and a tropospheric (N_s^{trop}) part.

Over heavily polluted areas, most of the NO_2 measured in the slant column will be contained in the troposphere. However, in remote areas such as oceans, stratospheric NO_2 is dominant in the slant column density.

Separating the slant column density into a stratospheric and tropospheric part is carried out by a data assimilation method, based on the Kalman filter technique. The observations are assimilated in the TM5-MP chemical transport model. This method relies on observations over remote areas with little tropospheric NO_2 to get accurate estimates of stratospheric slant column density (N_s^{strat}). As total reactive nitrogen (NO_y) is fairly well-conserved in the stratosphere, with relatively small sources and sinks, information of stratospheric NO_2 can be used in the model over long time periods. Stratospheric winds will transport the analysis results over remote areas to the polluted areas. In that way, a good estimate of stratospheric NO_2 can be given over areas with heavy tropospheric pollution as well.

After data assimilation, the stratospheric column is estimated. The tropospheric column is calculated by subtracting the stratospheric column from the total column.

$$N_s^{\text{trop}} = N_s - N_s^{\text{strat}} \tag{3.1}$$

For detailed information, the reader is referred to Van Geffen et al. 2019. It is important to note this step in the retrieval algorithm incorporates prior information on NO_2 columns from the TM5-MP model.

3.3.3. Conversion to vertical column density

The last step of the retrieval algorithm is to convert the slant column densities to vertical columns. This is important because the vertical column can easily be compared to model results. Slant columns are converted to vertical columns happens by Air-mass factors (AMFs), denoted by the symbol M. The AMF links the slant column density to the vertical density in the following way:

$$N_v = \frac{N_s}{M} \tag{3.2}$$

For calculation of the AMFs, prior information is again needed. The AMF depends on the vertical profile of the trace gas and is written as:

$$M = \frac{\sum_l m_l n_l c_l}{\sum_l n_l} \tag{3.3}$$

The total AMF can thus be expressed as a linear combination of altitude-dependent AMFs m_l, that describes the vertically resolved sensitivity for NO_2 in layer l. n_l denotes the NO_2 column density in layer l and c_l denotes a temperature correction term which is not further considered in this thesis. The altitude dependent AMFs are based on retrieval parameters such as the satellite viewing geometry, surface albedo, surface pressure and cloud cover. These AMFs are calculated from a radiative transfer model and are therefore used as prior information in the retrieval algorithm.

Finally, AMFs can be calculated for the troposphere (M^{trop}) and the stratosphere (M^{strat}) by limiting the sum in Equation (3.3) to the troposphere and stratosphere layers respectively. Using these factors, slant columns N_s^{strat} and N_s^{trop} are be converted to vertical columns N_v^{strat} and N_v^{trop}.

3.4. Averaging kernels

For data assimilation purposes, satellite trace gas retrievals need to be compared to results from a numerical model. The sensitivity of satellite instruments to tracer densities are strongly height dependent in the troposphere as a result of meteorological conditions. Therefore, large systematic errors could be introduced if information about these vertical profiles is not incorporated. Column densities, integrated over the layers of a numerical model cannot be directly compared to N_v^{trop} as supplied by TROPOMI. Rodgers 1976 introduced the concept of averaging kernels to enable these comparisons. Eskes and Boersma 2003 published a useful description of averaging kernels for optically thin absorbers such as NO_2.

The TROPOMI data product provides the averaging kernel **A** for each ground pixel. The averaging kernel is defined as the altitude-dependent AMFs ratioed by the total air-mass factor M. Figure 3.5 illustrates several averaging kernels. The three kernels, denoted **(a)**-**(c)** all show different sensitivities of the total vertical NO_2 column for modelled vertical profiles. For example, if thick cloud cover is present with the top at 800 hPa, the satellite is not able to measure any NO_2 below this level (kernel **(c)**). This illustrates the importance of the averaging kernel for comparison of model results to TROPOMI data. The model approximation of the tropospheric NO_2 column to compare to TROPOMI is the following:

$$N_v^{\text{trop}} = \mathbf{ASx}_m \tag{3.4}$$

In this equation, **A** denotes the TROPOMI averaging kernel operator, $\mathbf{x}_m$ denotes a vector containing the NO_2 concentrations of one vertical profile from the model, matrix **S**, denotes an operator that executes a mass-conserving vertical interpolation of the model profile $\mathbf{x}_m$ to the TROPOMI layers and when applicable, a change in units to mlc/m^2.

The TROPOMI averaging kernel is defined at 34 vertical levels $l = 1,\ldots,n_{\text{nlev}}^{\text{TROP}}$. The CTM used for simulations may have a different number of vertical levels $j = 1,\ldots,n_{\text{nlev}}^{\text{CTM}}$, defined at different pressure

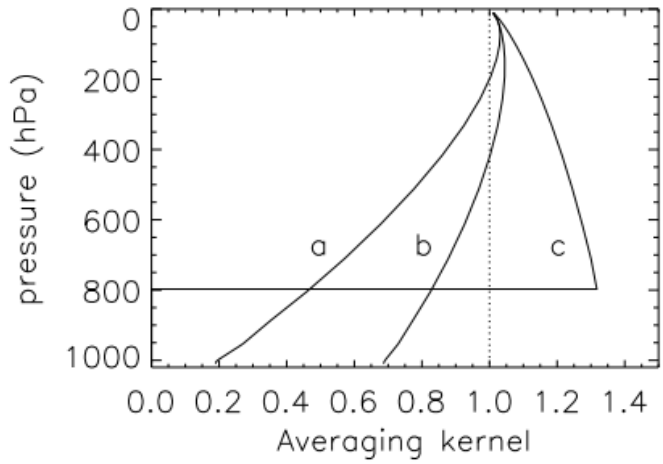


Figure 3.5: Example of averaging kernels: **(a)** clear pixel with a surface albedo of 0.02, **(b)** clear pixel with a surface albedo of 0.15, **(c)** pixel over a clouded area with a cloud top at 800hPa. Obtained from Eskes and Boersma 2003.

levels. In that case, vertical interpolation of the model layers is carried out before applying the averaging kernel **A**. For level l of the TROPOMI averaging kernel **A**, the l^{th} row of matrix S, accounts for the applied vertical interpolation. Each entry of S_{lj}, corresponds to the fraction of the model layer j is in TROPOMI level l. Moreover, the units used in the CTM may be different from the TROPOMI retrieval (mlc/m^2). The conversion of units is discussed in the next section.

3.5. Observation operator

This section describes the definition of the observation operator for TROPOMI. The observation operator was introduced in Subsection 2.3.2.

Observations are continually taken while TROPOMI orbits the earth. For comparison to model results, timestamps are rounded to the nearest hour. At all rounded timestamps, TROPOMI pixels are retrieved over the computation grid, measurement operator $\mathbf{H}_m$ is calculated. $\mathbf{H}_m$ is a matrix with dimension $N_{\text{pixel}} \times N_{\text{state}}$, where N_{pixel} denotes the number of pixels in the retrieval $\mathbf{y}_m$ at time t_m and N_{state} denotes the dimension of the state vector. This matrix is very large and needs significant storage resources, but by using its sparsity when storing the matrix this issue is overcome. For each TROPOMI pixel, the vertical NO_2 profile above the closest model horizontal grid cell is used for comparison. As the TROPOMI pixel footprints do not equal the model grid cells, horizontal averaging of every model grid cell that overlaps in some part with the TROPOMI pixels would be the most rigorous way to calculate the model observations, but that has not been implemented yet.

Every row of $\mathbf{H}_m$ corresponds to the observation of one single pixel. The non-zero coefficients of this row correspond to the vector **AS** in Equation (3.4). Therefore, entries of matrix **S** (dimensions $(n_{\text{nlev}}^{\text{TROP}} \times n_{\text{nlev}}^{\text{CTM}})$) have to be calculated. This is carried out per row, in two steps.

1. Conversion factors from concentrations (ppb) to vertical column densities (1×10^{15}mlc/cm^2) are calculated for the model's layers.

   In the CTM, concentrations of tracers are calculated and stored in ppb. For each vertical model layer, a factor is calculated that converts these Volume Mixing Ratios (VMRs) of NO_2 in ppb to a vertical column density measured in [1×10^{15}mlc/cm^2] [1]. This is carried out using the following equation.

[1] Reporting of VCD in SI-units would be done in [mol/m^2]. However, the units [1×10^{15}mlc/cm^2] are more commonly used for NO_2 retrievals.

$$\begin{aligned}
\text{VCD} &= \frac{1}{1\times10^{9}} \quad \text{VMR} \quad N_A \quad \frac{1}{M_{\text{air}}} \quad \text{VCD}_{\text{air}} \quad \frac{1}{1\times10^{4}} \frac{1}{1\times10^{15}} \\
\left[\frac{1\times10^{15}\,\text{mlc}}{\text{cm}^2}\right] &= \left[\frac{1}{\text{ppb}}\right]\left[\frac{\text{mol}\,NO_2\,\text{ppb}}{\text{mol air}}\right]\left[\frac{\text{mlc}}{\text{mol}}\right]\left[\frac{\text{mol air}}{\text{kg air}}\right]\left[\frac{\text{kg air}}{\text{m}^2}\right]\left[\frac{\text{m}^2}{\text{cm}^2}\right]\frac{1}{1\times10^{15}}
\end{aligned} \tag{3.5}$$

In this equation N_A denotes the Avogadro number, M_{air} denotes the molar mass of air, VCD_{air} denotes the air vertical column density of a certain vertical layer which can be approximated under the hydrostatic assumption from the pressure difference ΔP between the top and bottom of the vertical layer:

$$\text{VCD}_{\text{air}} = \frac{-\Delta P}{g}. \tag{3.6}$$

2. The entries of **S** corresponding to the vertical interpolation of the model's layers to the TROPOMI are calculated.

   The vertical layers in the numerical model need to be vertically interpolated to levels of the TROPOMI averaging kernel **A**, defined at 34 layers in the TM5-MP model. Interpolation of the averaging kernel is performed using pressure levels as reference height. A hypothetical model layer that is defined from 700 hPa to 600 hPa is divided over the TROPOMI layers in and around that level.

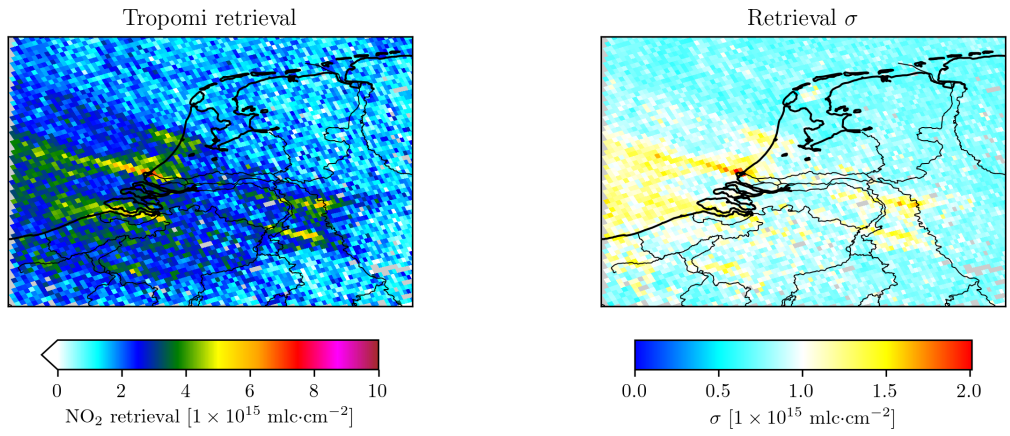
After calculation of the unit conversion factors and the vertical interpolation matrix, matrix **S** is completely defined. **AS** corresponds to the non-zero row entries of the pixel in $\mathbf{H}_m$ under consideration. These are entered in the correct place to correspond to the correct entries in the state vector $\mathbf{x}_m$.

3.6. Uncertainties and quality flags

For data assimilation purposes, it is paramount to have estimates of the accuracy of the observations. Van Geffen et al. 2019 describe in detail how uncertainty arises in the retrieval process. Basically, the total error arises by propagation of errors made in the three steps of the retrieval algorithm. These are slant column errors, errors in the separation of slant columns into stratospheric and tropospheric parts, errors in tropospheric air-mass factors. The total error is reported in the TROPOMI data product. A rough estimate of the measurement error σ of the vertical tropospheric column N_v^{trop} is given by:

$$\sigma = 0.5\times10^{15}\text{mlc/cm}^2 + [0.2\,\text{to}\,0.5]\cdot N_v^{\text{trop}} \tag{3.7}$$

For each retrieved pixel, a quality value q for the retrieval is reported. This value ranges from 0 to 1. It mainly depends on cloud presence over the pixel. Van Geffen et al. 2019 advise to use pixels with a quality value of 0.52 and above for data assimilation purposes. In this study, this threshold is used to include TROPOMI retrievals in data assimilation.



(a) **(b)**

Figure 3.6: Illustration of the TROPOMI retrieval error σ. **(a)** TROPOMI retrieval **(b)** TROPOMI error estimate σ.

4

Data assimilation theory & methodology

This chapter describes the mathematical theory of data assimilation. In Section 4.1, the parameter estimation problem is formulated. Section 4.2 gives an overview of methods within the field of data assimilation and the choice for variational methods in this study is motivated. In Section 4.3, the variational 4DVAR problem is stated along with the 'standard' adjoint-based solution strategy. Next, Section 4.4 describes adjoint-free methods for solving the 4DVAR problem. A general approach based on linearisation and reduced order modelling, proposed by Vermeulen and Heemink 2006 is discussed in Section 4.5. Section 4.6 discusses the application of this approach to the context of this study and derives an approximate linear model for LOTOS-EUROS. Section 4.7 shows how the 4DVAR problem can be solved using the approximate model for LOTOS-EUROS. Section 4.8 derives an approximation to the covariance of the estimated parameters. Section 4.9 synthesises the ideas in this chapter to a proposed methodology for solving the data assimilation problem in this study. Section 4.10 discusses how domain decomposition can be used to increase the computational efficiency of the model.

This chapter uses the notation from the state-space representation of LOTOS-EUROS that was introduced in Section 2.3.

4.1. Parameterisation and problem statement

The parameter vector $\boldsymbol{\beta}$ was introduced in Section 2.3. This section formulates its exact definition, based on which the data assimilation problem, addressed in this study, is stated.

LOTOS-EUROS simulates atmospheric processes as accurately as possible. However, the model is not perfect. One source of errors is the uncertainty in the model's parameters, such as the meteorology inputs, chemistry parameterisation, emissions and many more. Figure 4.1 illustrates the uncertainty of parameters in LOTOS-EUROS. Assimilating information from observations in the model is a way of finding better estimates of these parameters.

This thesis focuses on improving estimates of emission parameters, specifically anthropogenic NO_x emissions. In densely populated areas, these are the main sources of NO_x. Subsection 2.2.1 discussed the emission module of LOTOS-EUROS and the anthropogenic emission inventory. Uncertainty from anthropogenic emissions can be further broken down. Total emission strength, emission time profiles, spatial distribution of area sources and injection heights of emissions are important factors to calculate the emissions that all deal with uncertainty. The output of the emission module of LOTOS-EUROS is an emission strength per hour that is spatially distributed and subdivided into the source categories from Table 2.1. In this thesis, the aim is to estimate Multiplication Factors (MFs) for these emission strengths using satellite observations. No parameterisation of other emission uncertainties such as its temporal distribution is used. As the satellite retrievals are only available around 12:00 a.m. daily, variations of emissions within a day cannot be estimated from these retrievals.

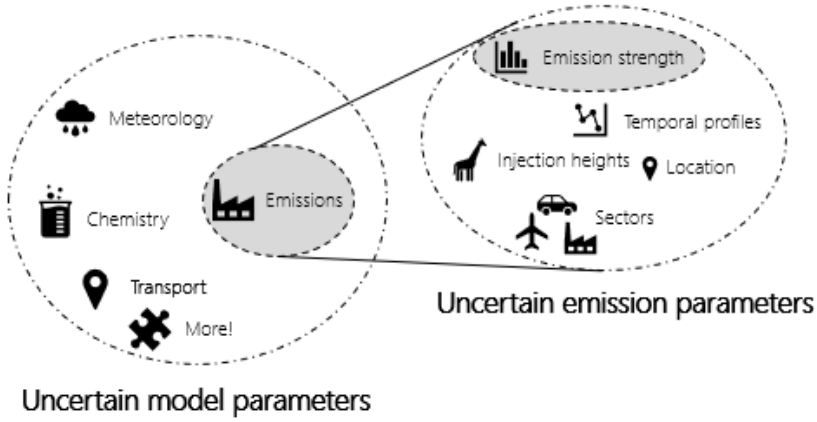


Figure 4.1: Realm of model parameter uncertainty. Many sources of uncertainty may cause the model to be imperfect. This thesis focuses specifically on improving emission strength estimates by assimilating satellite observations.

The parameter vector $\boldsymbol{\beta}$ is defined to contain multiplication factors that correct the emission strength from the inventories. This method of parameterising uncertainty in the emission strength is useful because it maintains prior information from the anthropogenic emission inventories. The starting point of the data assimilation is the so-called background parameter vector $\boldsymbol{\beta}^b$ that is equal to $\mathbf{1}$, which corresponds to the anthropogenic emissions from the inventories.

The approach of using MFs provides a flexible way of defining emission uncertainty. Jin et al. 2018 used the same approach for the parametrisation of uncertainty in dust emissions. MFs can be defined as spatially varying or uniform over the domain and be applied to specific source categories such as road transport only or to the total anthropogenic emissions. In principle any combination is possible. For example, known emission hotspots such as power plants and large industrial sites may receive individual corrections factors and to area sources such as road transport, a factor that is uniform over the domain may be applied.

By restricting parameter vector $\boldsymbol{\beta}$ to emission strengths, other sources of uncertainty from Figure 4.1 are not taken into account. Consequently, emission MFs may compensate for other errors in the model. This may lead to incorrect estimates of emissions parameters. This should be kept in mind when interpreting the multiplication factors.

With $\boldsymbol{\beta}$ defined, the parameter estimation problem statement is complete. The objective of this thesis is to assimilate TROPOMI NO_2 observations in LOTOS-EUROS to obtain better estimates of emission strengths through the MFs in parameter vector $\boldsymbol{\beta}$. The best estimate of $\boldsymbol{\beta}$ given the data will be referred to as $\hat{\boldsymbol{\beta}}$.

4.2. Methods for solving the data assimilation problem

Lewis et al. 2006 categorised data assimilation methods in two groups: variational methods and statistical methods.

Variational methods assume that the underlying model is deterministic. A model is deterministic if no randomness is involved in the evolution of the state. Variational methods treat the parameter estimation problem as an optimisation problem. A cost function $J(\boldsymbol{\beta})$ is defined that penalises the mismatch between model simulations and observations and the deviation of the parameters from their a-priori value. The best estimate of the parameter vector $\hat{\boldsymbol{\beta}}$ is the vector $\boldsymbol{\beta}$ that minimises J. A well-known example is the 4DVAR algorithm, discussed in Section 4.3.

The *statistical* approach assumes that the underlying model is stochastic and adds noise to the state evolution Equation (2.2). Under the assumption of Gaussian error distributions, this approach reduces to that of the well-known Kalman filter. In the Kalman filter algorithm, data are assimilated sequentially.

An improved estimate of the model state is reached by a weighted average of the prior and observation information.

Recently, many data assimilation methods have been proposed that combine variational and statistical methods. Bannister 2017 gives a good overview of such hybrid methods. This thesis focuses on the application of variational methods on the parameter estimation problem. This choice is made because for two reasons.

- **Constant-in-time parameters**

  Estimation of parameters (in this thesis: MFs) that are constant in time fits very well into the variational problem formulation. In the stochastic approach, parameters would have to be appended to the state vector to estimate them. As states are inherently supposed to be (slowly) time-varying in the stochastic approach, the same then applies to the parameters. Extra smoothing or averaging steps are necessary to calculate the best constant estimate. Moreover, for constant parameters, the Kalman filter is known to be biased and could even diverge (Vermeulen and Heemink 2006).

- **Variational methods respect model physics**

  The second advantage of variational methods is that the optimal estimated trajectory corresponds to a run of a model simulation. The recursive estimations made in the Kalman filter allow for the trajectory of a state to have shocks and permit that the analysis trajectory is not the result of a model simulation. As the aim of this thesis is to estimate actual emission strengths, the physical link between emissions and measured NO_2 concentrations is extremely important. This physical link essentially means conservation of mass, which is not violated in the variational approach.

4.3. The 4DVAR problem

This section states the 4DVAR problem and the canonical adjoint-based way of solving it. The 4DVAR cost function J is defined as:

$$\begin{aligned} J(\boldsymbol{\beta}) = &\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}^b)^T\mathbf{B}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^b) \\ &+\frac{1}{2}\sum_{m=1}^{N_o}\left(\mathbf{y}_m^o-\mathbf{H}_m(\mathbf{x}(t_m),\boldsymbol{\beta})\right)^T{\mathbf{R}_m}^{-1}\left(\mathbf{y}_m^o-\mathbf{H}_m(\mathbf{x}(t_m),\boldsymbol{\beta})\right) \end{aligned} \tag{4.1}$$

$\boldsymbol{\beta}^b$ denotes the background (prior) parameter vector. The background parameter vector contains the prior value of $\boldsymbol{\beta}$. In this thesis, $\boldsymbol{\beta}^b = \mathbf{1}$, resulting in emissions that are equal to the original inventory. $\mathbf{B}$ is the prior covariance matrix of $\boldsymbol{\beta}$. In Chapters 5 and 6 that describe the experiments, $\mathbf{B}$ will be defined. $\mathbf{y}_m^o$ are the observations at time t_m. In this thesis, the observations are TROPOMI NO_2 retrievals. $\mathbf{R}_m$ is the covariance matrix of the observation error $\boldsymbol{\epsilon}_m$ in Equation (2.4).

Minimisation of Equation (4.1) is the objective of the 4DVAR method. The estimated parameter vector $\hat{\boldsymbol{\beta}}$ is therefore defined as:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{N_p}}{\arg\min}\, J(\boldsymbol{\beta}) = \{\mathbf{v}\in\mathbb{R}^{N_p} | \forall \mathbf{w}\in\mathbb{R}^{N_p} : J(\mathbf{v}) \leq J(\mathbf{w})\} \tag{4.2}$$

It is important to note that minimising Equation (4.1) is a constrained optimisation problem. The trajectory of state vector $\mathbf{x}$ is constrained by Equation (2.2), through which the dependence of the trajectory on the parameters $\boldsymbol{\beta}$ is included.

4.3.1. Gradient-based solution strategies

Solving the 4DVAR problem poses a significant mathematical and computational challenge. Solution strategies are usually gradient-based. In an iterative procedure, the gradient $\nabla_{\boldsymbol{\beta}} J(\boldsymbol{\beta})$ is obtained for the latest estimate of $\boldsymbol{\beta}$ and a new estimate of $\boldsymbol{\beta}$ is found using an optimisation strategy such as steepest descent. Most computational costs in solving the 4DVAR problem are associated with obtaining the gradient of J.

4.3.2. The adjoint method of obtaining the gradient

Courant and Hilbert 1953 introduced the so-called adjoint approach of obtaining the gradient of the cost function J. In this section, this method, adapted from Lewis et al. 2006, will be discussed.

In this method, firstly, the minimisation problem in Equation (4.1) is reformulated as an unconstrained minimisation problem by Lagrange multipliers $\boldsymbol{\lambda}(t_k)$. Without loss of generality, it is assumed that observations are available at each timestep t_k. Gradient $\nabla_{\boldsymbol{\beta}} J$ can be calculated to be:

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) =& \mathbf{B}(\boldsymbol{\beta}-\boldsymbol{\beta}^b) - \sum_{k=1}^{K}\left(\frac{\partial \mathbf{M}_k(\mathbf{x}(t_{k-1}),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{\lambda}(t_k) - \\ & \sum_{k=1}^{K}\left(\frac{\partial \mathbf{H}_k(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^T \mathbf{R}_k^{-1}\left(\mathbf{y}_k^o - \mathbf{H}_k(\mathbf{x}(t_k),\boldsymbol{\beta})\right). \end{aligned} \tag{4.3}$$

In this equation, $\boldsymbol{\lambda}(t_k)$ are the Langrange multipliers that satisfy the *adjoint* equations:

$$\begin{aligned}\boldsymbol{\lambda}(t_k) =& \left(\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}\right)^T \boldsymbol{\lambda}(t_{k+1}) + \\ & \left(\frac{\partial \mathbf{H}_k(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}\right)^T \mathbf{R}_k^{-1}\left(\mathbf{y}_k^o - \mathbf{H}_k(\mathbf{x}(t_k),\boldsymbol{\beta})\right) \end{aligned} \tag{4.4}$$

In this equation, $k = K,\ldots,1$ and $\boldsymbol{\lambda}(t_{K+1}) = 0$. This recursion is solved backwards in time.

Calculation of the gradient using the adjoint method is done in two steps. The first step is a forward model run during which the trajectory and the Jacobians are stored. The second step is a run of the adoint model, Equation (4.4) backwards in time to obtain $\boldsymbol{\lambda}(t_k)$. After this, the ∇J can be calculated using Equation (4.3).

The adjoint approach is very attractive computationally. For calculation of the gradient, one forward model and one adjoint model run suffice, irrespective of the number of parameters. Increasing the number of parameters would likely mean more gradient calculations and updates of $\boldsymbol{\beta}$ to arrive at the minimum of J, but the costs of computing a gradient remain more or less the same. If the adjoint model is available, this would be the preferred method of solving the 4DVAR problem.

However, for many numerical models, the adjoint model is not available. The transposed Jacobians that form the adjoint model correspond to linearisation of every model step and reversing the time step. Creating and maintaining it requires a tremendous programming effort. The source code of the model would have to be accessed for that purpose. Numerical simulation models such as LOTOS-EUROS consist of thousands of lines of code and are often written by several persons over many years. Calculation of the tangent and adjoint model needs to be done line by line which requires a lot of time and detailed knowledge of the code. An approach to solving the 4DVAR problem that includes accessing the model source code is called *intrusive*.

4.4. Adjoint-free, gradient-based 4DVAR solution strategies

As implementation of the adjoint of a large, non-linear model is often not feasible, other ways of obtaining the gradient have been developed. In this thesis, such methods to solve the 4DVAR problem will be referred to as "adjoint-free, gradient-based methods".

4.4.1. Finite differences

The most basic method of approximating the gradient of J around some $\tilde{\boldsymbol{\beta}}$ would be to calculate it by finite differences. The i^{th} component of $\nabla_{\boldsymbol{\beta}} J(\tilde{\boldsymbol{\beta}})$ can be approximated by:

$$\nabla J(\tilde{\boldsymbol{\beta}})_i = \frac{\partial J(\tilde{\boldsymbol{\beta}})}{\partial \beta_i} \approx \frac{J(\tilde{\boldsymbol{\beta}} + \Delta\beta_i \mathbf{e}_i) - J(\tilde{\boldsymbol{\beta}})}{\Delta\beta_i}, \qquad i = 1,\ldots,N_p \tag{4.5}$$

for some perturbation $\Delta\beta_i > 0$. $\mathbf{e}_i$ is the i^{th} column of the $N_p \times N_p$ identity matrix. This way of calculating ∇J requires N_p+1 simulations of the model: one model run using the base parameters $\tilde{\boldsymbol{\beta}}$ and N_p runs for perturbation of each parameter. Increasing the number of parameters will increase the computational

costs of this method greatly. More perturbation runs are required for calculation of the gradient and the computation needs to be repeated many times to find the minimum of the cost function in an iterative procedure. Therefore, this method is not feasible in situations where the number of parameters is large and single model runs are very costly.

4.4.2. Gradients from approximate models

Another approach to make the adjoint method feasible is to replace the original model by an approximation for which the adjoint model is available. An incremental approach was proposed by Courtier et al. 1994, who replaced a weather forecast model by a linear approximate model. In the incremental approach, the modelled state contains perturbations around a base trajectory that was computed using the full model. This approximate model used simpler physics. It was, however, necessary to develop and implement this linear model from scratch (Courtier et al. 1994).

4.4.3. Reduced order model approach

Vermeulen and Heemink 2006 introduced a new approach to variational data assimilation that also makes use of approximate models. Rather than developing it from first principles or deducing the full tangent linear models, they used an ensemble of forward models to deduce this approximate model. From this ensemble of forward models, Proper Orthogonal Decomposition (POD) patterns were calculated that form a subspace of the full state space. A reduced-order model was developed by projecting the original model onto this subspace. An advantage of this method is that it is non-intrusive to the model's source code. This approach was reported to have comparable computational efficiency to the adjoint method. The method was successfully applied to geological inverse problems (Xiao et al. 2019), (Kaleta et al. 2011) and to shallow-water flow modelling (Altaf et al. 2009). This approach was chosen as starting point in this study.

4.5. Construction of an approximate, reduced order model

This section presents the approach by Vermeulen and Heemink 2006 of developing a linear approximation of the original model. The adjoint of this approximate, linear model is used for solving the 4DVAR problem.

4.5.1. Linearisation of the original model

The approach starts with linearisation of the original model $\mathbf{M}_k$ around the background trajectory $\mathbf{x}_k^b$ and background parameter vector $\boldsymbol{\beta}^b$. The background trajectory $\mathbf{x}_k^b$ is calculated using the original model with background parameters $\boldsymbol{\beta}^b$. The difference between a trajectory x and the background is then approximated by:

$$\begin{aligned}\mathbf{x}(t_{k+1}) - \mathbf{x}^b(t_{k+1}) \approx & \frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta})}{\partial \mathbf{x}(t_k)} (\mathbf{x}(t_k) - \mathbf{x}^b(t_k)) \\ & + \frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \boldsymbol{\beta}^b)\end{aligned} \tag{4.6}$$

In this equation, both Jacobian matrices are evaluated at $(\mathbf{x}^b(t_k), \boldsymbol{\beta}^b)$. This linearisation is called the Trajectory Piece-wise Linearisation (TPWL). This linearisation corresponds to the tangent of the original model that was necessary to build the adjoint model. These Jacobians are generally not available; if they were, the adjoint model is would be available as well.

Rather than deriving the Jacobians in Equation (4.6) from the model source code, as is done for derivation of the adjoint model, they could theoretically be estimated from forward model runs where the state and parameters are calculated from the background values plus perturbations Δx_i and $\Delta \beta_i$ respectively. The i_{th} columns of the Jacobians are approximated as:

$$\left[\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta})}{\partial \mathbf{x}(t_k)} \right]_i \approx \frac{\mathbf{M}_{k+1}(\mathbf{x}(t_k) + \Delta x_i \mathbf{e}_i, \boldsymbol{\beta}) - \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta})}{\Delta x_i} \tag{4.7}$$

and

$$\left[\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]_i \approx \frac{\mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta}+\Delta\beta_i\mathbf{e}_i)-\mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\Delta\beta_i} \tag{4.8}$$

In Equation (4.7), $i = 1,\ldots,n$ and n is the dimension of the state. In Equation (4.8), $i = 1,\ldots,N_p$ and N_p is the length of the parameter vector.

Xiao et al. 2019 chose a different strategy for approximating the Jacobians in Equation (4.6) using Radial Basis Functions. With this approach, an interpolation model is formed from an ensemble of state vectors and the derivative of this model is used as an estimate for the Jacobians.

Both approaches, however, need to work around the problem that $\boldsymbol{\beta}$ and especially $\mathbf{x}$ can be extremely high-dimensional. The large number of full model runs needed to calculate all components of the Jacobians is not feasible. Therefore, Equations (4.7) and (4.8) cannot be used to approximate the Jacobians.

The innovative approach by Vermeulen and Heemink 2006 was lowering the dimension of both $\mathbf{x}-\mathbf{x}^b$ and $\boldsymbol{\beta}$ by using a reduced-order modelling approach based on Proper Orthogonal Decomposition (POD). After the dimension reduction of $\boldsymbol{\beta}$ and $\mathbf{x}$, a feasible number of model runs was required to approximate the Jacobians in reduced dimensions. The following section describes the construction of a reduced-order approximation of Equation (4.6). Subsection 4.5.3 shows how, by making use of reduced-order modelling, the calculation of the Jacobians becomes feasible. The description is adapted from the comprehensive and detailed description of the method in Kaleta et al. 2011, that was was applied to a geological parameter estimation problem.

4.5.2. POD-based reduced order modelling

In the POD dimension reduction approach, only the main patterns in $\mathbf{x}-\mathbf{x}^b$ are used in the approximate model. These patterns form the orthogonal projection matrix $\boldsymbol{\Phi}_x \in \mathbb{R}^{(n\times n_{\text{red}})}$, where n denotes the length of the state vector $\mathbf{x}$ and n_{red} the reduced state dimension. $\mathbf{x}-\mathbf{x}^b$ is projected on a lower dimensional subspace $\mathbb{R}^{n_{\text{red}}}$. The reduced state vector $\mathbf{z}$ is defined as:

$$\mathbf{x}(t_k)-\mathbf{x}^b(t_k) = \boldsymbol{\Phi}_x\,\mathbf{z}(t_k) \quad \Rightarrow \quad \mathbf{z}(t_k) = \boldsymbol{\Phi}_x{}^T\,(\mathbf{x}(t_k)-\mathbf{x}^b(t_k)) \tag{4.9}$$

In a similar fashion, the parameter vector $\boldsymbol{\beta}$ can be approximated in reduced dimension $\mathbb{R}^{N_{\text{p,red}}}$ by $\boldsymbol{\eta}$:

$$\boldsymbol{\beta}-\boldsymbol{\beta}^b = \boldsymbol{\Phi}_\beta\boldsymbol{\eta} \quad \Rightarrow \quad \boldsymbol{\eta} = \boldsymbol{\Phi}_\beta{}^T(\boldsymbol{\beta}-\boldsymbol{\beta}^b) \tag{4.10}$$

In nearly linear systems, it is often possible to represent $\mathbf{x}$ in a much lower dimension by means of POD. In the case of modelling NO_2 in LOTOS-EUROS, this is therefore expected to be applicable.

The reduction of the parameter space depends heavily on the chosen parametrisation. Xiao et al. 2019 considered spatially varying permeability patterns of geological formations and was able to significantly reduce the dimension of the parameter vector with acceptable loss of resolution. Jin et al. 2018 used spatially varying MFs to parameterise dust emissions and also successfully reduced the parameter vector dimension.

Method of snapshots

To find the dominant patterns in the state space that define the projection matrix $\boldsymbol{\Phi}_x$, the so-called method of *snapshots* is used. Matrix $\mathbf{X}$ is formed, where the columns consist of snapshots of the model state. These snapshots, $\mathbf{x}_{j,k}$, are meant to capture the variations in the model induced by the parameters. They are defined as:

$$\mathbf{x}_{j,k} = \mathbf{M}_k(\mathbf{x}_j(t_{k-1}),\boldsymbol{\beta}_j)-\mathbf{M}_k(\mathbf{x}_j(t_{k-1}),\boldsymbol{\beta}^b), \quad j = 1,\ldots,N_{\text{runs}} \tag{4.11}$$

The snapshots are computed from N_{runs} model runs with different parameter settings $\boldsymbol{\beta}_j$. Snapshot matrix $\mathbf{X}$ is defined as:

$$\mathbf{X} = \{\mathbf{x}_{1,1},\ldots,\mathbf{x}_{1,K},\ldots,\mathbf{x}_{N_{runs},1},\ldots,\mathbf{x}_{N_{runs},K}\} \tag{4.12}$$

The size of $\mathbf{X}$ is $n\times s$, where s is the number of snapshots. K denotes the number of time-steps of the model at which snapshots are taken. These snapshots define a subspace of the full state space $\mathbb{R}^n$. This

subspace should capture the variation of the model induced by the parameters. Therefore, in each of the N_{runs}, another parameter vector $\boldsymbol{\beta}_j$ is used. How N_{runs} and $\boldsymbol{\beta}_j$ should be chosen is discussed below. POD is used to calculate a basis of this subspace to efficiently describe it in a relatively low dimension.

The dominant, orthogonal patters $\boldsymbol{\phi}_i$ in $\mathbf{X}$ form this basis and correspond to the eigenvectors that solve the eigenvalue problem:

$$(\mathbf{X}\mathbf{X}^T)\boldsymbol{\phi}_i = \lambda\boldsymbol{\phi}_i. \tag{4.13}$$

This eigenvalue problem may be solved by calculating the Singular Value Decomposition (SVD) of $\mathbf{X}$:

$$\mathbf{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{4.14}$$

$\boldsymbol{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{V} \in \mathbb{R}^{s\times s}$ are orthonormal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times s}$ is pseudo-diagonal matrix with the singular values arranged in decreasing order on its diagonal. It can be shown that the eigenvectors of $\mathbf{X}\mathbf{X}^T$ are equal to $\boldsymbol{U}$ and that the eigenvalues are the squares of the diagonal of $\boldsymbol{\Sigma}$.

By an energy criterion, n_{red} eigenvectors corresponding to the largest eigenvalues can be chosen from $\boldsymbol{U}$. The eigenvalue λ_i, that corresponds to eigenvector $\boldsymbol{\phi}_i$, provides a measure α_i for the energy associated with $\boldsymbol{\phi}_i$ in $\mathbf{U}$:

$$\alpha_i = \frac{\lambda_i}{\sum_{i=1}^{s}\lambda_i}. \tag{4.15}$$

By choosing n_{red} in a way that

$$\sum_{i=1}^{n_{\text{red}}} \alpha_i > 0.95, \tag{4.16}$$

the eigenvectors $\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{n_{\text{red}}}$ represent 95% of the energy contained in $\mathbf{X}$. The chosen threshold defines the trade-off between dimension reduction and the quality of the low-dimensional representation of the state.

4.5.3. Constructing the reduced order model

After the derivation of projection matrices $\boldsymbol{\Phi}_x$ and $\boldsymbol{\Phi}_\beta$ from a model ensemble, $\mathbf{x}$ and $\boldsymbol{\beta}$ can be represented in reduced dimension using Equations (4.9) and (4.10). It is then possible to reduce the TPWL model in Equation (4.6). This reduced model becomes:

$$\mathbf{z}(t_{k+1}) = \left(\boldsymbol{\Phi}_x{}^T \frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}\boldsymbol{\Phi}_x\right)\mathbf{z}(t_k) + \left(\boldsymbol{\Phi}_x{}^T \frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\boldsymbol{\Phi}_\beta\right)\boldsymbol{\eta} \tag{4.17}$$

Using Equation (4.9), the reduced state vector $\mathbf{z}$ can always be transformed into the full state space. In Equation (4.17), the Jacobian matrices of the original model are still unknown. However, a much more efficient computational procedure can now be used to estimate them compared to the approach in Equations (4.7) and (4.8).

In Equations (4.7) and (4.8), every element of $\mathbf{x}$ and every element of $\boldsymbol{\beta}$ is perturbed, a total of $n+N_p$ perturbations. The advantage of the reduced form of the TPWL model in Equation (4.17) is that it does not require all these perturbations. Due to the projection on the reduced model space, only perturbations in this reduced space are needed.

The i^{th} columns of the respective Jacobians are approximated by:

$$\left[\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}\boldsymbol{\Phi}_x\right]_i \approx \frac{\mathbf{M}_{k+1}(\mathbf{x}(t_k)+\Delta\phi_x\boldsymbol{\phi}_{x,i},\boldsymbol{\beta}) - \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\Delta\phi_x} \tag{4.18}$$

and

$$\left[\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\boldsymbol{\Phi}_\beta\right]_i \approx \frac{\mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta}+\Delta\phi_\beta\boldsymbol{\phi}_{\beta,i}) - \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\Delta\phi_\beta} \tag{4.19}$$

Here, because $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}\boldsymbol{\Phi}_x$ is a $n \times n_{\text{red}}$ matrix, only n_{red} perturbations need to be done for approximating the Jacobian associated with the state. Instead of perturbing the state $\mathbf{x}(t_k)$ by $\Delta x_i\mathbf{e}_i$ (where $i = 1,\ldots,n$) as in Equation (4.7), only perturbations are $\Delta\phi_x\boldsymbol{\phi}_{x,i}$ (where $i = 1,\ldots,n_{\text{red}}$ need to be carried out. $\boldsymbol{\phi}_{x,i}$ is the i^{th} column of projection matrix $\boldsymbol{\Phi}_x$.

Similarly, only $N_{\text{p,red}}$ perturbations have to be carried out for approximating the Jacobian with respect to the parameters. It is at this point that the added value of reduced-order modelling can be appreciated.

After the projection matrices $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_\beta$ are determined and calculations of the reduced order Jacobians in Equations (4.18) and (4.19) are finished, the approximate linear model in Equation (4.17) is completely defined. The adjoint model of this linear approximation is available. The adjoint method, discussed in Subsection 4.3.2 may then be used to find an approximation of the gradient of cost function $J(\boldsymbol{\beta})$.

4.6. Defining the approximate linear model for LOTOS-EUROS

The methodology of finding a linear, reduced order approximate model, that was described generically in Section 4.5, could be applied to any state space representation of a numerical model. Such applications can be environmental modelling, weather prediction, geophysical applications and many more. In this section, the application of the methodology from Section 4.5 to the specific context of this study is discussed.

4.6.1. State vector definition

When the state space representation of LOTOS-EUROS was introduced in Section 2.3, the state vector $\mathbf{x}$ was defined to contain concentrations at all positions of all tracers simulated in LOTOS-EUROS. It is necessary to keep track of all tracers in the state vector because tracers can interact through chemical reactions. Model operator $\mathbf{M}$ needs to calculate these reactions for each timestep, and therefore all tracers have to be stored in $\mathbf{x}$.

In this study, the focus is on NO_x emissions and the only measurements available are of NO_2. When developing an approximate model for data assimilation purposes, the reactions of NO_2 with other tracers are not modelled explicitly. Reactions with other tracers are modelled in the LOTOS-EUROS runs from which the approximate model is developed, but the approximate model state only contains NO_2 concentrations. In twin experiments, this approximation will be proven to be valid. Therefore, state vector $\mathbf{x}$ is defined to *only contain the NO_2 concentrations* calculated by LOTOS-EUROS.

4.6.2. Trajectory Piece-Wise Linearisation of LOTOS-EUROS

The TPWL model in Equation (4.6) consists of two types of linearisations. The Jacobian $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}$, dictates how small perturbations of the state vector at time t_k translate into perturbations at time t_{k+1}. Similarly, the Jacobian $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ dictates how perturbations of the parameter vector $\boldsymbol{\beta}$ translate to perturbations of the state vector at time t_{k+1}.

As NO_2 is a relatively short-lived trace gas (1-12 hours) (Stavrakou et al. 2013), it can be assumed that the state of a previous day will have little influence on the state of the following day. TROPOMI observations are available every day around noon. For the 4DVAR problem, accurate estimates of the state are required only at the times of observations. As a result, the linearisation of the time-evolution of the state through matrix $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}$ may not be relevant for an approximate model to accurately calculate the state at the times of observation t_m.

If this is indeed the case, the first term of Equation (4.6) may be omitted. The whole reduced order modelling approach for the state, discussed in Subsection 4.5.2, was only considered for making the calculation of $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}$ feasible. Omitting $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}$ in Equation (4.6) means that the dimension reduction of the state $\mathbf{x} \in \mathbb{R}^n$ to $\mathbf{z} \in \mathbb{R}^{n_{\text{red}}}$ is no longer required.

When $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \mathbf{x}(t_k)}$ is omitted from Equation (4.6), only the linearisation of the state around the background state $\mathbf{x}^b$ with respect to $\boldsymbol{\beta}$, described by $\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ is left. The following equation remains:

$$\mathbf{x}(t_{k+1}) - \mathbf{x}^b(t_{k+1}) \approx \frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k),\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \boldsymbol{\beta}^b) \tag{4.20}$$

As mentioned, in this equation, the Jacobian is evaluated in $(\mathbf{x}^b(t_k), \boldsymbol{\beta}^b)$. On closer inspection, the time-stepping nature of the approximate TPWL model in Equation (4.6) is abandoned in Equation (4.20). To calculate the state $\mathbf{x}(t_{k+1})$, the previous state $\mathbf{x}(t_k)$ is no longer needed. Rearranging this yields:

$$\mathbf{x}(t_{k+1}) \approx \mathbf{x}^b(t_{k+1}) + \left.\frac{\partial \mathbf{M}_{k+1}(\mathbf{x}(t_k), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\mathbf{x}(t_k)=\mathbf{x}^b(t_k), \boldsymbol{\beta}=\boldsymbol{\beta}^b} (\boldsymbol{\beta} - \boldsymbol{\beta}^b) \tag{4.21}$$

The Jacobian matrix in Equation (4.21) will play an important role in this thesis and will be referred to as $\mathbf{E}_k$. As the timestepping nature of the approximate model is lost, the state at time t_k will be described instead of time t_{k+1}.

$$\mathbf{E}_k = \left.\frac{\partial \mathbf{M}_k(\mathbf{x}(t_{k-1}), \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\mathbf{x}(t_{k-1})=\mathbf{x}^b(t_{k-1}), \boldsymbol{\beta}=\boldsymbol{\beta}^b} \tag{4.22}$$

A compact notation of Equation (4.21) can now be given:

$$\mathbf{x}(t_k) \approx \mathbf{x}^b(t_k) + \mathbf{E}_k(\boldsymbol{\beta} - \boldsymbol{\beta}^b) \tag{4.23}$$

4.6.3. Calculating the approximate model

The approximate model in Equation (4.23) is completely defined by $\mathbf{E}_k$. This Jacobian matrix is not available explicitly from the model. However, it can be approximated by the finite-difference procedure in Subsection 4.5.1. In Equation (4.8), this finite difference approximation is defined mathematically. Several model runs are needed to calculate $\mathbf{E}_k$. $N_p + 1$ forward model runs are needed, one with background parameters $\boldsymbol{\beta}^b$ and N_p where the parameters are consecutively pertubed by $\Delta\beta_i$.

4.6.4. Reducing the parameter space

It is important to note that the reduction of the state vector using projection matrix $\boldsymbol{\Phi}_x$ is no longer necessary when using the approximation in Equation (4.23). However, reduction of the parameter vector $\boldsymbol{\beta}$, described in Subsection 4.5.2 is still possible and could greatly improve the computational efficiency of the derivation of Jacobian $\mathbf{E}_k$. If the parameter space is reduced to $\mathbb{R}^{N_{\text{p,red}}}$, only $N_{\text{p,red}}+1$ perturbation runs are needed instead of $N_p + 1$.

For the chosen parametrisation in this study for the NO_x emissions, reduction of the parameter space will not bring benefit. Dimension reduction is suitable for global, continuously spatially varying patterns. Jin et al. 2018 successfully reduced large-scale parameter patterns for dust emissions and Xiao et al. 2019 successfully reduced spatially varying geological permeability patterns. The MFs for NO_x emission strengths in this study do not have redundancy that can be exploited for order reduction. Each MF is defined explicitly and no correlations between the MFs are present.

4.6.5. The final approximate model in layman's terms

The theory of finding an approximate model for LOTOS-EUROS in Sections 4.5 and 4.6 was discussed in a quite formal, mathematical way. This section provides a more intuitive description of the final approximate model of Equation (4.23) in layman's terms.

The approximate model for LOTOS-EUROS serves the following purpose: giving a good description of the influence of the parameters $\boldsymbol{\beta}$ on the NO_2 concentrations in the atmosphere that are calculated by LOTOS-EUROS and are observed by TROPOMI. For any choice of parameters, LOTOS-EUROS may be run and calculate the concentrations. However, every run of LOTOS-EUROS is costly and time-consuming. This depends on the settings of LOTOS-EUROS but as a rule of thumb, simulation of a week takes about an hour on the high-performance computing cluster that was available for this study. For optimisation of the parameters in the 4DVAR problem, many parameter settings need to be tried. This created the need for an approximate model that calculates the concentrations based on chosen parameters much quicker.

The key to the derivation of this approximate model is to think about the way the parameters influence the model output. One of these parameters, β_1, for example, can change the NO_x emission strength from a power plant. Intuitively, it is clear that increasing that parameter and thereby the NO_x emission from the plant increases the NO_2 concentrations in the exhaust plume carried downwind from the plant.

A LOTOS-EUROS is a nearly linear model with respect to NO_2, it can be expected that a certain increase in emissions results in a proportional increase in the NO_2 concentrations in the plume from the power plant. This is exactly what Equation (4.23) shows. The NO_2 concentrations $\mathbf{x}$ at time t_k are approximately

the NO_2 concentrations with the background emissions plus something extra caused by the increased emission from the power plant.

To calculate the increased concentrations ofNO_2, two runs of LOTOS-EUROS are done. One where the emission of the power plant was as the default emissions ($\beta_1 = 1$) and one where the emission was increased, e.g. $\beta_1 = 1.5$. The difference in concentrations between these runs at time t_k is stored in the first column of matrix $\mathbf{E}_k$.

It is now possible to estimate the NO_2 concentrations if for example $\beta_1 = 1.2$. This is just the NO_2 concentrations resulting from the background emissions plus 40% of the difference between the concentrations with $\beta_1 = 1.5$ and $\beta_1 = 1$. This holds for all parameters $\beta_i, i = 1,\ldots,N_p$ that are defined. Figure 5.5 illustrates such plumes of NO_2 that are altered by the parameters.

4.7. Approximate 4DVAR optimisation

The approximate model for LOTOS-EUROS derived in Subsection 4.6.2 and stated in Equation (4.23) turns out to be very convenient for solving the 4DVAR problem. In this section, the optimisation of the 4DVAR cost function J using the approximate model is discussed.

The right-hand side of Equation (4.23) is denoted as $\hat{\mathbf{x}}$. This is the approximate model:

$$\hat{\mathbf{x}}(t_k) = \mathbf{x}^b(t_k) + \mathbf{E}_k(\boldsymbol{\beta} - \boldsymbol{\beta}^b) \tag{4.24}$$

Substitution of $\mathbf{x}$ by $\hat{\mathbf{x}}$ in the 4DVAR cost function, Equation (4.1), yields:

$$\begin{aligned} J(\boldsymbol{\beta}) \approx{}& \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^b)^T \mathbf{B}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^b) \\ &+ \frac{1}{2}\sum_{m=1}^{N_o} (\mathbf{y}_m^o - \mathbf{H}_m\hat{\mathbf{x}}(t_m))^T {\mathbf{R}_m}^{-1}(\mathbf{y}_m^o - \mathbf{H}_m\hat{\mathbf{x}}(t_m)) \end{aligned} \tag{4.25}$$

where $\mathbf{H}_m$ is the linear observation operator that corresponds to the averaging kernel of TROPOMI that was described in Section 3.5.

The substitution of $\mathbf{x}$ by $\hat{\mathbf{x}}$ greatly simplifies optimisation of the 4DVAR cost funtion J. Minimising Equation (4.1) is a constrained optimisation problem, where trajectory $\mathbf{x}$ has to satisfy the LOTOS-EUROS model, Equation (2.2). However, in Equation (4.25) $\hat{\mathbf{x}}$ is explicitly available from Equation (4.24) and the problem is no longer constrained.

A closer look at Equation (4.25) reveals that it is of a quadratic form in terms of $\boldsymbol{\beta}$. Therefore, it is relatively easy to calculate the gradient. The gradient can be calculated by:

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} J ={}& \mathbf{B}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^b) \\ &- \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \left[\mathbf{y}_m^o - \mathbf{H}_m[\mathbf{x}^b(t_m) + \mathbf{E}_m(\boldsymbol{\beta} - \boldsymbol{\beta}^b)]\right] \end{aligned} \tag{4.26}$$

Equation (4.26) can be used for some gradient-based optimisation strategy for the 4DVAR problem. However, functional $J(\boldsymbol{\beta})$ reaches its minimum for $\boldsymbol{\beta}$ when $\nabla_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) = \mathbf{0}$. This condition is equivalent to solving a system $\mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ where:

$$\mathbf{A} = \mathbf{B}^{-1} + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m \tag{4.27}$$

$$\mathbf{b} = \mathbf{B}^{-1}\boldsymbol{\beta}^b + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \left[\mathbf{y}_m^o - \mathbf{H}_m\mathbf{x}^b(t_m) + \mathbf{E}^m\boldsymbol{\beta}^b\right] \tag{4.28}$$

Using these matrices, the following result is found:

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{b}. \tag{4.29}$$

This linear system is of size $N_p \times N_p$. In practice, the inverse matrix $\mathbf{A}^{-1}$ is never calculated explicitly but a numerical solver for the system $\mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ is used.

$\hat{\boldsymbol{\beta}}$ minimises the approximate cost function in Equation (4.25). It, therefore, is a suboptimal solution of the original 4DVAR problem. In theory, it can be necessary to repeat the assimilation procedure multiple times where, in the approximate model in Equation (4.24), $\hat{\boldsymbol{\beta}}$ is used as the new background parameter vector $\boldsymbol{\beta}^b$. In twin experiments, the necessity of such iterations will be investigated.

4.8. An approximation of the estimate covariance

The approximate 4DVAR method from Section 4.7 only provides an estimate $\hat{\boldsymbol{\beta}}$, based on the observations $\mathbf{y}_m^o$. However, the observations $\mathbf{y}_m^o$ contain measurement errors and also the representation by the model has uncertainties; the combined observation error is described by $\boldsymbol{\epsilon}_m$. It would be valuable to quantify how the observation error influences the estimate $\hat{\boldsymbol{\beta}}$. In this section, an approximation of the covariance of $\hat{\boldsymbol{\beta}}$ is derived. This can be used to quantify the uncertainty in $\hat{\boldsymbol{\beta}}$.

$\hat{\boldsymbol{\beta}}$ can be calculated using Equation (4.29). The full equation is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^{-1} + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m)^{-1} \left(\mathbf{B}^{-1} \boldsymbol{\beta}^b + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} (\mathbf{y}_m^o - \mathbf{H}_m \mathbf{x}^b(t_m) + \mathbf{H}_m \mathbf{E}_m \boldsymbol{\beta}^b) \right) \tag{4.30}$$

Four assumptions have to be made to approximate the covariance of $\hat{\boldsymbol{\beta}}$.

1. There is some true parameter vector $\boldsymbol{\beta}_{\text{true}}$, of which the value is known.
2. The observations $\mathbf{y}_m^o$ can be simulated with a model run using $\boldsymbol{\beta}_{\text{true}}$ and one particular sample of the observation error. These observations are generated using the observation operator Equation (2.4):
$$\mathbf{y}_m^o = \mathbf{H}_m \mathbf{x}(t_m) + \boldsymbol{\epsilon}_m, \qquad m = 1, \ldots, N_o. \tag{4.31}$$
3. The observation error $\boldsymbol{\epsilon}_m$, in the equation above, is multivariate Gaussian, independent in time.
$$\boldsymbol{\epsilon}_m \sim N(\mathbf{0}, \mathbf{R}_m) \tag{4.32}$$
In this study, the covariance matrix $\mathbf{R}_m$ is assumed to be a diagonal matrix with $\boldsymbol{\sigma}_m^2$ on its diagonal, where $\boldsymbol{\sigma}_m^2$ is supplied in the TROPOMI data product.
4. The fourth assumption is that LOTOS-EUROS can be approximated using the approximate model in Equation (4.23).

The first three assumptions are satisfied in twin experiments, described in Chapter 5, that will be used to test the assimilation methodology. In experiments with real TROPOMI data, the assumptions are violated. However, in experiments with real TROPOMI retrievals, knowledge of the accuracy under idealised conditions can still be valuable. The fourth assumption has to be tested in experiments, but this assumption is already fundamental to approximate 4DVAR method described in the previous section. If this assumption is violated, the approximate 4DVAR method itself will not work, thereby rendering the derivation in this section useless.

The approximate model plays an important role in the derivation of the covariance of $\hat{\boldsymbol{\beta}}$. The following approximation of the observations can be made using the approximate model in Equation (4.23):

$$\mathbf{y}_m^o \approx \mathbf{H}_m \left(\mathbf{x}^b(t_m) + \mathbf{E}_m (\boldsymbol{\beta}_{\text{true}} - \boldsymbol{\beta}^b) \right) + \boldsymbol{\epsilon}_m, \qquad m = 1, \ldots, N_o. \tag{4.33}$$

Using this approximation of the observations, Equation (4.30) can be rewritten to:

$$\begin{aligned} \hat{\boldsymbol{\beta}} \approx (\mathbf{B}^{-1} + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m)^{-1} \Bigg(& \mathbf{B}^{-1} \boldsymbol{\beta}^b + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m \boldsymbol{\beta}_{\text{true}} \\ & + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m \boldsymbol{\epsilon}_m \Bigg) \end{aligned} \tag{4.34}$$

This closed-form approximation of $\hat{\boldsymbol{\beta}}$ can be used to estimate the statistical properties of $\hat{\boldsymbol{\beta}}$. In Equation (4.34), $\boldsymbol{\epsilon}_m$ is a Gaussian random variable. Equation (4.34) is an affine transformation of $\boldsymbol{\epsilon}_m$. It follows that the approximation of $\hat{\boldsymbol{\beta}}$ is a Gaussian random variable as well. For a multivariate Gaussian variable $\mathbf{v} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, it can be shown that $\mathbf{w} = \mathbf{u} + \mathbf{B}\mathbf{v}$ is also multivariate Gaussian with mean $\mathbf{u}$ and covariance $\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T$.

Using the assumption that $\boldsymbol{\epsilon}_m \sim N(\mathbf{0}, \mathbf{R}_m)$, the approximation to $\hat{\boldsymbol{\beta}}$ in Equation (4.34) is a Gaussian random variable with mean $\boldsymbol{\mu}_{\hat{\boldsymbol{\beta}}}$:

$$\boldsymbol{\mu}_{\hat{\boldsymbol{\beta}}} = (\mathbf{B}^{-1} + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m)^{-1} \left(\mathbf{B}^{-1} \boldsymbol{\beta}^b + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m \boldsymbol{\beta}_{\text{true}} \right) \tag{4.35}$$

and covariance $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \sum_{m=1}^{N_o} \mathbf{D}_m \mathbf{R}_m \mathbf{D}_m^T \tag{4.36}$$

In this equation, $\mathbf{D}_m$ is used to shorten notation. $\mathbf{D}_m$ equals:

$$(\mathbf{B}^{-1} + \sum_{m=1}^{N_o} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m)^{-1} \mathbf{E}_m^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{E}_m. \tag{4.37}$$

Equation (4.36) gives an approximation of the covariance of $\hat{\boldsymbol{\beta}}$. It describes how uncertainty in the observations induced by $\boldsymbol{\epsilon}_m$ influences $\hat{\boldsymbol{\beta}}$. This can be very useful to quantify the uncertainty in $\hat{\boldsymbol{\beta}}$.

From Equation (4.35), it follows that $\boldsymbol{\mu}_{\hat{\boldsymbol{\beta}}}$ is a weighted average of $\boldsymbol{\beta}^b$ and $\boldsymbol{\beta}_{\text{true}}$. This weighting is a result of the background and observational part of the cost function. In the case that observation covariance matrices $\mathbf{R}_m$ are small compared to the background covariance $\mathbf{B}$, the observations are most important in the cost function and $\boldsymbol{\mu}_{\hat{\boldsymbol{\beta}}}$ approaches $\boldsymbol{\beta}_{\text{true}}$. In the case that observation covariance matrices $\mathbf{R}_m$ are large compared to the background covariance $\mathbf{B}$, the prior knowledge about $\boldsymbol{\beta}$ is most important and, therefore, $\boldsymbol{\mu}_{\hat{\boldsymbol{\beta}}}$ approaches $\boldsymbol{\beta}^b$. This result is consistent with the intuition of the assimilation problem.

The derivation of these results relies on the assumptions mentioned above. In twin experiments, those assumptions hold and can be used to assess the accuracy with which the parameters can be estimated based on the satellite data. If $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ is large, the measurement error limits the accuracy of the estimation of the parameters. The actual difference $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\text{true}}$ may be be larger than $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ predicts, because of errors in the approximate model on which the 4DVAR methodology of Section 4.7 is based.

4.9. Proposed adjoint-free 4DVAR data assimilation method for LOTOS-EUROS

This chapter is concluded by a description of the proposed method of assimilating TROPOMI NO_2 retrievals.

1. **Problem definition**

   The parameter estimation problem is defined by choosing the time range, simulation domain, LOTOS-EUROS setting and the parameterisation. The parameterisation $\boldsymbol{\beta}$ defines MFs applied to anthropogenic NO_x emissions for which the best estimate is found based on the TROPOMI retrievals.

2. **Calculation of the observation operators $\mathbf{H}_m$**

   In the assimilation period, N_o TROPOMI NO_2 retrievals are available at times $t_m, m = 1, \ldots, N_o$. From the averaging kernels defined at these times, the observation operators $\mathbf{H}_m$ need to be calculated as described in Section 3.4.

3. **Calculation of approximate model**

   The approximate model was defined by Equation (4.23). At all observation times t_m, matrix $\mathbf{E}_m$ needs to be calculated. For this, an ensemble of $N_p + 1$ model runs is needed: one with default emissions ($\boldsymbol{\beta} = \mathbf{1}$) and N_p where each parameter β_i is perturbed by $\Delta\beta_i$. Equation (4.8) describes how the columns of $\mathbf{E}_m$ can be calculated from the ensemble.

4. **Calculating the linear system for $\boldsymbol{\beta}$**

   When matrices $\mathbf{H}_m$ and $\mathbf{E}_m$ have been calculated, it is possible to form the linear system in Equation (4.29).

5. **Solve the system for $\hat{\boldsymbol{\beta}}$**

   The solution of Equation (4.29) is $\hat{\boldsymbol{\beta}}$.

6. **Calculate the approximate covariance of $\hat{\boldsymbol{\beta}}$**

   An approximation of the covariance $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ of $\hat{\boldsymbol{\beta}}$ can be calculated using Equation (4.36). This gives an indication of the uncertainty of $\hat{\boldsymbol{\beta}}$ induced by the measurement error in the satellite observations.

7. **(Optionally) perform a second optimisation loop**

   $\hat{\boldsymbol{\beta}}$ obtained in step 5 is optimal for the approximate 4DVAR cost function in Equation (4.25). It may be necessary to repeat the procedure from step 3, taking $\hat{\boldsymbol{\beta}}$ as background parameter vector $\boldsymbol{\beta}^b$. The need for multiple optimisation loops will be investigated in experiments described in Chapter 5.

A schematic overview of the method is given in Figure 4.2

The computational effort of the method is dominated by the calculation of the approximate model. N_p+1 runs of the original model are necessary. The other steps in the method are of negligible cost.

As the number of parameters increases, the computational effort of generating the approximate model increases proportionally. This is prohibitive for a large number of parameters. It is possible to create a more efficient procedure for calculating the approximate model by making use of domain decomposition and by reducing chemistry calculations. This will be discussed in the following sections.

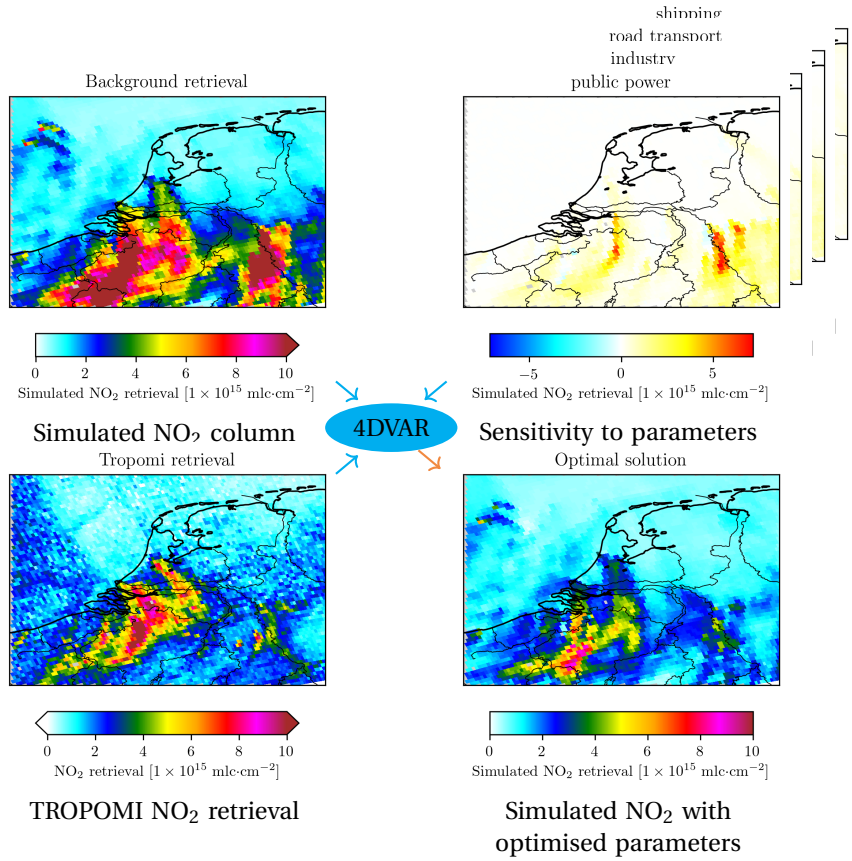


Figure 4.2: Schematic illustration of the proposed assimilation method described.

4.10. Increasing computational efficiency by domain decomposition

In this section, a method is proposed to calculate the approximate model more efficiently by using domain decomposition. Until now, the calculation of each column of $\mathbf{E}_m$ requires one LOTOS-EUROS run with background parameters and N_p runs where exactly one parameter is perturbed one by one. By using domain decomposition, it is possible to calculate multiple columns of $\mathbf{E}_m$ simultaneously from a run where multiple parameters in different subdomains are perturbed at the same time.

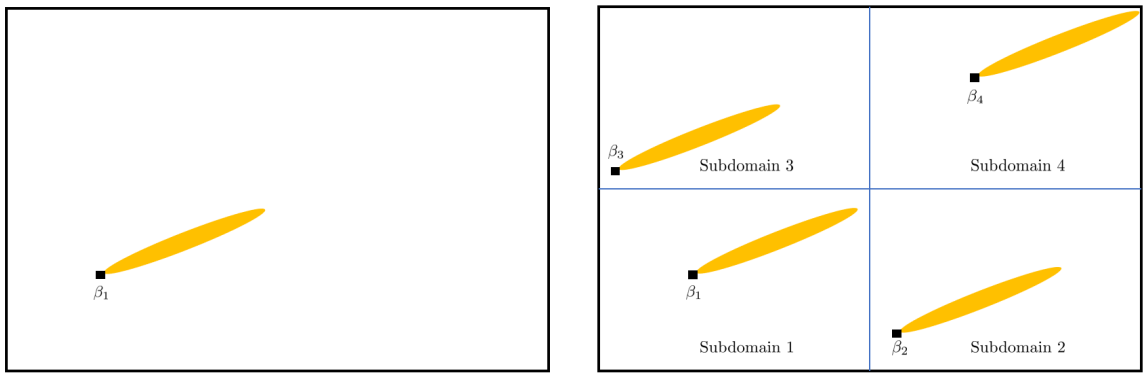
In domain decomposition, the computational grid is subdivided into multiple subdomains. Emission parameters that are defined in one subdomain may not have any effect in neighbouring subdomains. For the example of a MF for an emission hotspot, this is the case when the NO_2 plume that propagates away from the source does not enter other subdomains. In the same LOTOS-EUROS run that this parameter is perturbed, another parameter in another subdomain can be perturbed.

This is illustrated in Figure 4.3. In one LOTOS-EUROS run, parameters β_1 - β_4 are perturbed simultaneously. The difference between the LOTOS-EUROS run with background parameters is visible in the plumes in Figure 4.3b. If the subdomains are chosen as shown in the figure, only the effect of one parameter is present, within each subdomain. The 4 columns of $\mathbf{E}_m$ that correspond to $\beta_1 - \beta 4$ can be calculated on the subdomains only.

In theory, calculating the effect of changing multiple parameters from one LOTOS-EUROS run is possible whenever these effects are non-overlapping. The domain decomposition can be defined in various ways. This will be studied in detail in a twin experiment, described in Section 5.5.

4.11. Increasing computational efficiency by reducing chemistry

The domain decomposition method proposed in the previous section reduces the number of LOTOS-EUROS simulations needed to calculate the approximate model. Another way of increasing computational efficiency of the method is to decrease the costs of obtaining the ensemble of simulations. In the LOTOS-EUROS runs that form the ensemble, only the NO_x emission strengths are perturbed. Although the concentrations of some tracers are affected through chemical reactions with NO_2, the effect for many other tracers may be minor. It would, therefore, be attractive to only simulate all tracers in LOTOS-EUROS once and store the results. These results may be used in the runs where the perturbations of NO_x emission strengths are done. During these perturbation runs, fewer tracers have to be simulated, drastically improving the efficiency of obtaining the ensemble. This idea is not investigated in this study but is interesting to explore in further research.



(a) **(b)**

Figure 4.3: Illustration of domain decomposition. **(a)** The plume from the emission hotspot that is estimated by β_1 does not travel through the entire domain. **(b)** After defining subdomains 1-4, it is possible to perturb parameters $\beta_1 - \beta_4$ simultaneously in a LOTOS-EUROS run.

5

Twin experiments

This chapter describes three twin experiments that were performed to test the data assimilation methodology proposed in Section 4.9. Section 5.1 introduces the general concept and goal of twin experiments. Sections 5.3 to 5.5 will discuss the three twin experiments conducted and their results.

5.1. Twin experiments

The goal of a twin experiment in the context of data assimilation is to assess the performance of the methodology used. In Section 4.9, a methodology was proposed to estimate NO_x emission strengths from satellite retrievals. Using actual TROPOMI retrievals, this methodology would return some estimated emission parameters $\hat{\boldsymbol{\beta}}$. However, there is no rigorous way to assess how well the methodology works as the true value of the parameters is unknown. Furthermore, as described in Section 4.1, there are other sources of uncertainty and errors in LOTOS-EUROS apart from the NO_x emission strengths. These uncertainties and errors are not covered by the parameters. In experiments with TROPOMI data, this can cause incorrectly estimated parameters.

The purpose of a twin experiment is to work around the problems that arise in real-life applications to test the methodology. In a twin experiment, a setting is created where the true value of the parameters $\boldsymbol{\beta}_{\text{true}}$ and the errors introduced in the observations are known. The accuracy of the methodology can then be assessed.

In Section 4.1, the approach of parameterising NO_x emissions by multiplication factors was introduced. In this study, $\boldsymbol{\beta}$ consists of MFs for the NO_x emission strength. These MFs can be defined in various ways. In the three twin experiments in this thesis, three different parameterisations i.e. definitions for the MFs will be used.

In the twin experiments, synthetic satellite observations will be used instead of the TROPOMI retrievals. These synthetic observations are calculated from a LOTOS-EUROS run with known parameters $\boldsymbol{\beta}_{\text{true}}$ and perturbed by a random error. This run will be called the *truth run*. The process of creating synthetic observations is discussed in the next subsection.

The twin experiments described in this chapter answer three questions:

1. *How well can parameters be estimated from the synthetic data in the twin experiment?*

2. *How well does the 4DVAR optimisation procedure using the approximate model work?*

3. *Does the development of the approximate model using domain decomposition introduce extra errors?*

Twin experiments 1 and 2 address the first two questions. Twin experiment 3 studies the third question.

5.2. Creating synthetic TROPOMI observations

The synthetic observations used in twin experiments are generated from the *truth* run. A LOTOS-EUROS simulation is performed with known parameters $\boldsymbol{\beta}_{\text{true}}$, the resulting trajectory is $\mathbf{x}_{\text{true}}$. At all times t_m that TROPOMI retrievals are available, the simulated retrieval can be calculated using observation operator $\mathbf{H}_m$. The construction of the observation operator was discussed in Section 3.5.

It would not be realistic to use these simulated retrievals in data assimilation directly. As discussed in Section 2.3 noise and representation errors always distort observations. Therefore, the simulated retrievals are also perturbed by noise to create synthetic observations.

In this study, the noise used to perturb the simulated retrievals to create the synthetic observations is generated using the standard deviation reported for TROPOMI retrievals. In this way, the observation error in the twin experiments is comparable to the TROPOMI retrievals, under the assumption that the supplied error characteristics are correct. An illustration of the process is given in Figure 5.1.

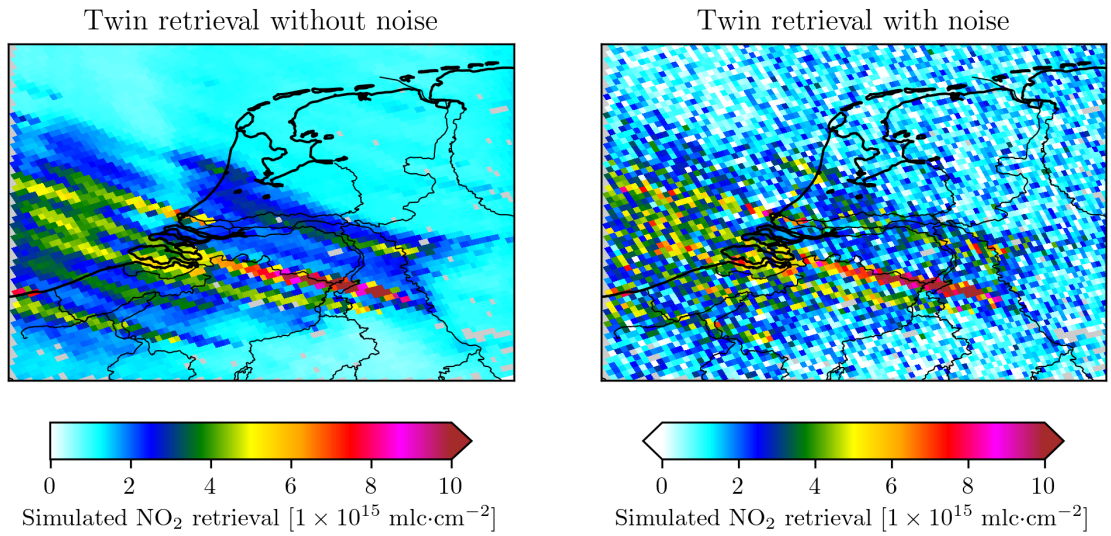
The creation of the synthetic observation $\mathbf{y}^o_m$ at time t_m is described in the following equation:

$$\mathbf{y}^o_m = \mathbf{H}_m \mathbf{x}_{\text{true}}(t_m) + \boldsymbol{\epsilon}_m \tag{5.1}$$

In this equation, the random additive error $\boldsymbol{\epsilon}_m$ is a sample from a Gaussian distribution:

$$\boldsymbol{\epsilon}_m \sim N(\mathbf{0}, \text{diag}(\boldsymbol{\sigma}^2_m)) \tag{5.2}$$

The standard deviation of the error, $\boldsymbol{\sigma}_m$, is taken from the TROPOMI product. The assumption of independence of the error between pixels may be too strong. In reality, the retrieval error of nearby pixels is likely to be correlated. However, no information on this is supplied in the TROPOMI product. Therefore, $\boldsymbol{\epsilon}_m$ is modelled as independent Gaussian in the twin experiments.



(a) (b)

Figure 5.1: Creation of synthetic observations for twin experiments on July, 1, 2018. **(a)** Simulated retrieval from truth run. **(b)** Truth run disturbed by random additive noise that satisfies the TROPOMI noise statistic.

5.3. Twin experiment 1 - blocks

This section describes the first twin experiment that was conducted. It was created as a first test case in this study on which methodologies could be tested.

5.3.1. Test region & time

A grid is defined with coordinates $[2°, 8°] \times [50°, 54°]$ with a resolution of $0.1° \times 0.05°$. This translates to a resolution of 7×6 km. A total of $60 \times 80 = 4.800$ horizontal grid cells are used. With 10 vertical layers, this yields a total of 48.000 grid points where NO2 concentrations are calculated. The assimilation period is set to July 1-7, 2018. For this period, sufficient cloud-free TROPOMI NO_2 retrievals over Benelux area are available.

5.3.2. Parameterisation

In the experiment, five MFs for NO_x emission are chosen, $\boldsymbol{\beta} \in \mathbb{R}^5$. Each parameter will correct the total NO_x emission in a square subdomain. These block fields are chosen over the large NO_2 emitters that are the ports of Antwerp, Rotterdam, the steel factories of IJmuiden and the Amsterdam metropolitan area, the southern part of the Ruhr area and the northern part of the Ruhr area. Figure 5.3d illustrates these emission hotspots and the NO_2 plumes that propagate from them, simulated by LOTOS-EUROS. $\boldsymbol{\beta}$ is modelled as an independent Gaussian distributed variable. $\boldsymbol{\beta} \sim N(\mathbf{1}, \sigma^2\mathbf{I})$. Different samples of $\boldsymbol{\beta}^b$ for $\sigma = 0.3$ are shown in figure 5.2.

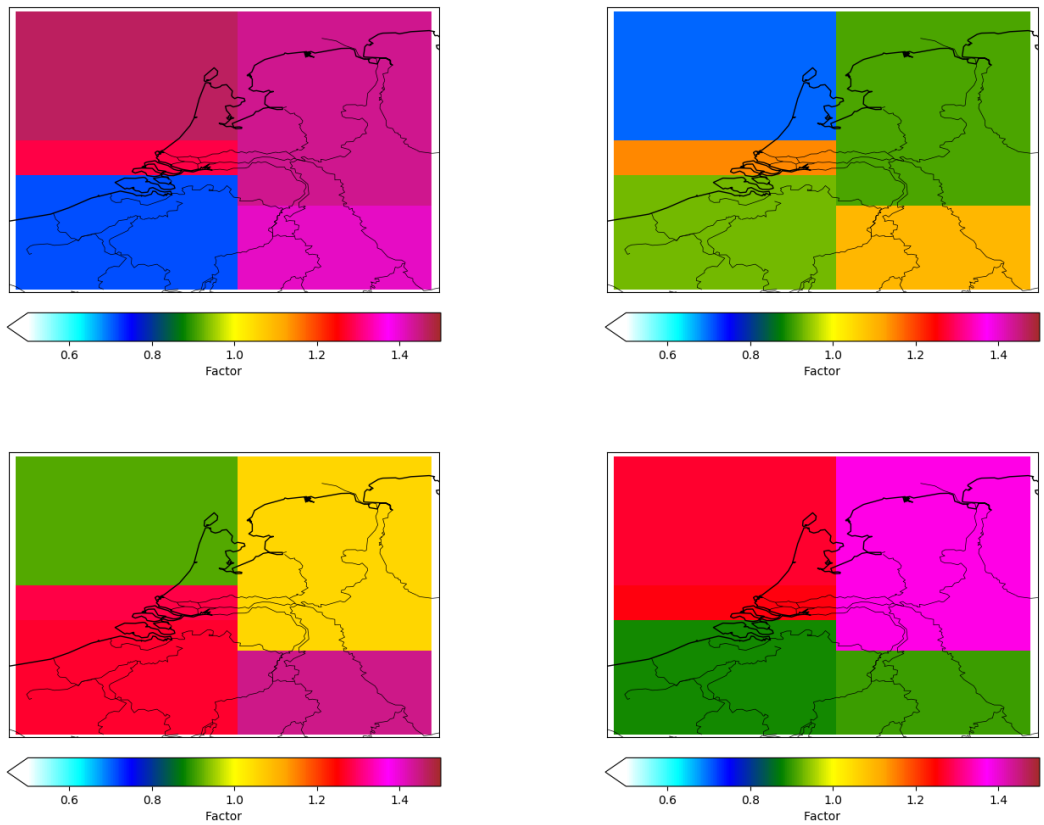


Figure 5.2: Examples of spatial MFs of NO_x emissions in the first twin experiment.

5.3.3. Results and discussion

The results are reported visually in Figure 5.3. On the first day of the assimilation window, July 1, 2018, the synthetic satellite observation is shown in Figure 5.3a. In Figure 5.3b, the retrieval corresponding to a model run with the estimated parameters is shown. Figure 5.3c shows the difference between the synthetic observation and the retrieval corresponding to estimated parameters. Figure 5.3d shows the

MFs	True value	Estimated value	Error	$\Sigma_{\hat{\beta}}$
Block 1 (Antwerp)	0.8	0.8078	0.98%	0.0027 (0.3325%)
Block 2 (Rotterdam)	1.5	1.4996	−0.02%	0.0051 (0.3422%)
Block 3 (Amsterdam)	1.1	1.1015	0.14%	0.0060 (0.5456%)
Block 4 (Southern Ruhr area)	0.9	0.9687	7.63%	0.0015 (0.1623%)
Block 5 (Northern Ruhr area)	1.3	1.3013	0.10%	0.0032 (0.2468%)

Table 5.1: Results of twin experiment 1 with assimilation window July 1, 2018, until July 7, 2018. The blocks contain the areas between parentheses. The error is calculated as absolute difference between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\text{true}}$. The entries under $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ are square root of the diagonal of the estimated covariance matrix.

background retrieval i.e. the simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$. The difference between the background retrieval and the retrieval with estimated parameters is visible.

The results of the first twin experiment are reported in Table 5.1. From this table, it can be seen that all factors are estimated within 1% of their true value, except for the MF of block 4. These results are promising and show that the methodology is successful for this parameterisation of emission.

An estimate for the covariance of $\hat{\boldsymbol{\beta}}$ was also computed using Equation (4.34). The estimated sensitivity of the estimated parameters to the measurement error is small (< 1%) using this assimilation window of one week. The square root of the diagonal of $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ is reported in the table. This is an estimate of the standard deviation in each parameter.

Twin experiment 1 shows that the methodology can be used to estimate the parameters $\boldsymbol{\beta}_{\text{true}}$. This implicitly means that the approximate model used in the 4DVAR assimilation works well enough for its purpose. Of course, other questions arise that are related to this experiment. How does the length of the assimilation window influence the accuracy of the estimation? How well is the 4DVAR cost function approximated using the approximate model? These questions are answered in the second twin experiment, where a more sophisticated parameterisation is implemented.

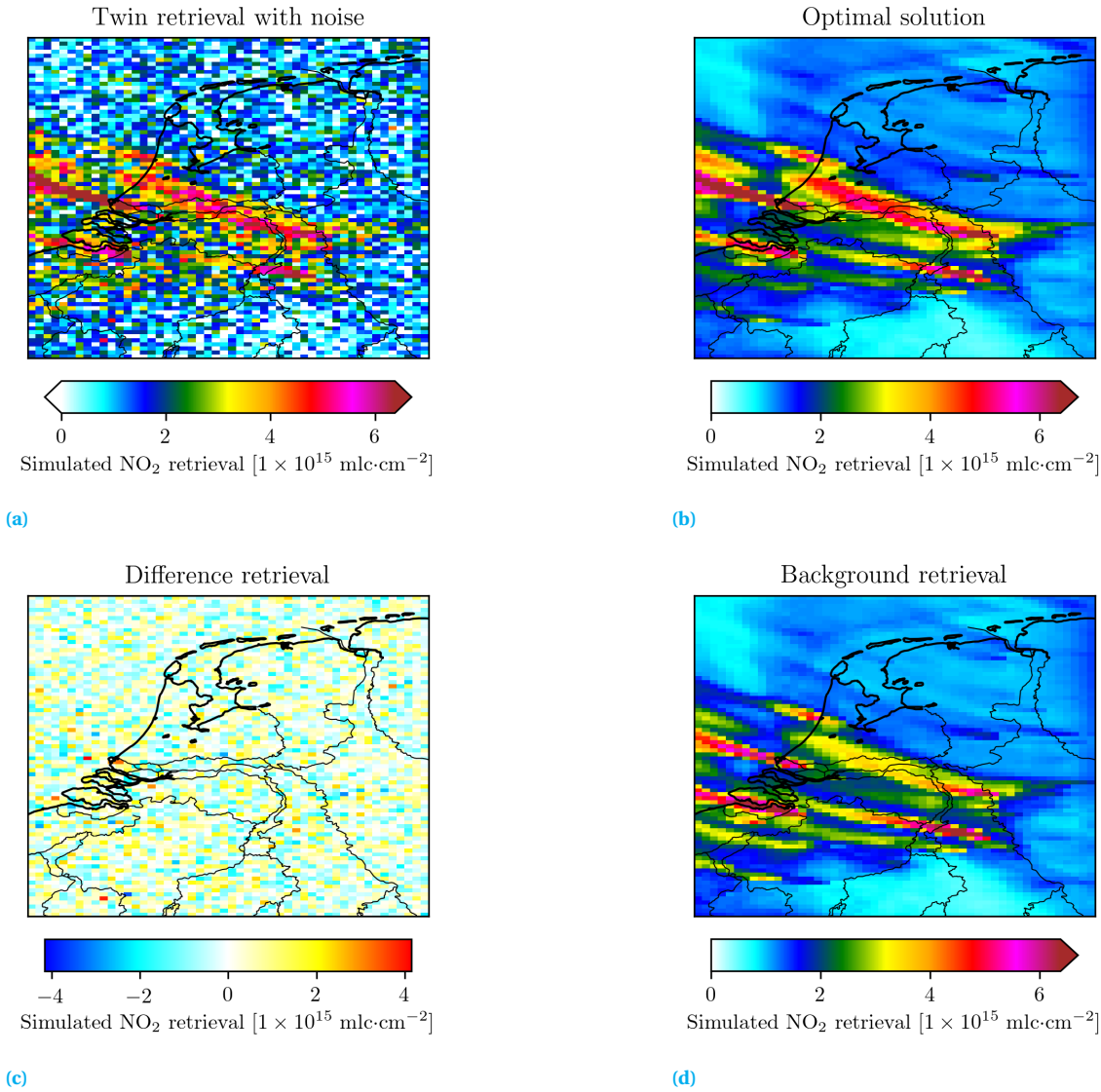


Figure 5.3: Results of twin experiment 1 at 2018-07-01 12:00. **(a)** Synthetic satellite observation as described in Section 5.2 **(b)** Simulated retrieval with optimal parameters **(c)** Difference between twin retrieval and optimal solution **(d)** Simulated retrieval with background parameters $\boldsymbol{\beta}_b = \mathbf{1}$.

5.4. Twin experiment 2 - Source categories and emission hotspots

This section describes the setup and results of the second twin experiment that was conducted. In this experiment, a more sophisticated parameterisation of emissions was used. A more detailed analysis of the performance of the assimilation methodology using the approximate model is discussed.

5.4.1. Test region & time

Similar to the first experiment, a grid is chosen with coordinates $[2°, 8°] \times [50°, 54°]$ and a resolution of $0.1° \times 0.05°$. This is equivalent to a resolution of 7×6 km. A total of $60 \times 80 = 4.800$ horizontal grid cells are used. With 10 vertical layers, this yields a total of 48.000 grid points where NO2 concentrations are calculated.

Two assimilation windows are used. The first is the assimilation period is July 1, 2018, until July 7, 2018. In this period, sufficient cloud-free TROPOMI retrievals are available over the Benelux area. In a second experiment, the assimilation window only consists of July 1, 2018, to investigate the performance of the assimilation using the retrievals of only one day.

5.4.2. Parameterisation

In this experiment, MFs apply to specific NO_x emission categories and the total emission of predetermined emission hotspots. This parameterisation corresponds closely to the emission inventory. As described in Section 2.2, anthropogenic NO_x emissions are classified into different categories. Table 2.1 lists the categories of anthropogenic emissions that are distinguished in the *CAMS v2.2* inventory. A useful experiment could be to specify MFs that apply to these emission categories. The data assimilation scheme can then be used to calculate whether emission from specific categories is too high or too low using the satellite data. In the same way, it is interesting to estimate the total emission from an emission hotspot such as the port of Rotterdam's industrial area. To achieve this, MFs are defined that apply to all grid cells that correspond to this area. The definition of MFs for categories and emission hotspots is discussed in the following subsections.

Multiplication factors for emission categories

In this twin experiment, MFs are applied to the five dominant categories of NO_x emission. These are public power, industry, road transport for four different fuel types (F1-F4), shipping and off-road vehicles. This selection was made based on their magnitude. No thorough selection criterion was developed since the goal of this twin experiment is to test whether factors for such categories could be estimated with the current procedure. Although emissions from road transport are subdivided into four categories, one MF is used for the total road transport emission. As the spatial and temporal emission patterns of the four categories are identical, it is impossible to distinguish the four subcategories using satellite data.

MFs for emission hotspots

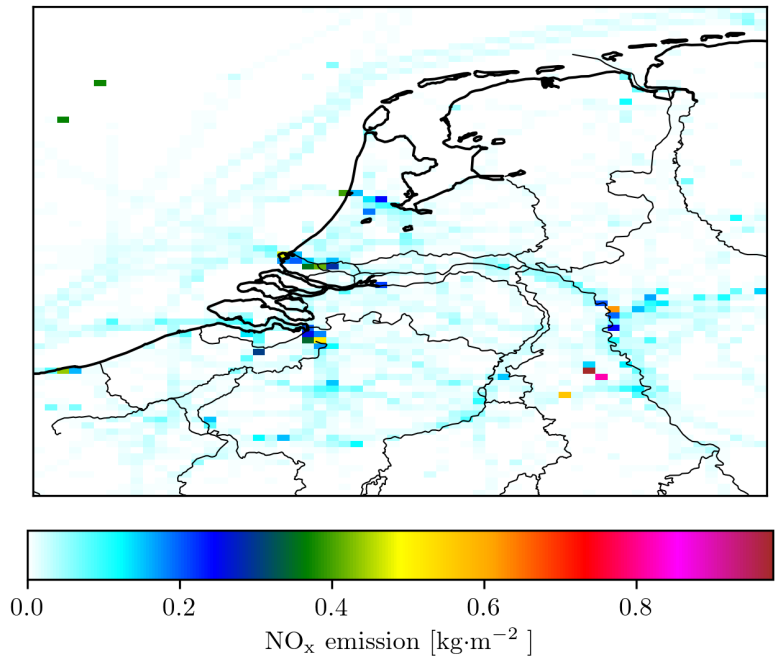
From the emission inventories, it is found that the majority of NO_x emission is spatially concentrated. Figure 5.4a shows the aggregated NO_x emission from July 1, 2018, until July 7 2018. In this figure, some clear hotspots of NO_x emissions are visible.

In this experiment, hotspots were defined as grid cells for which the NO_x emission exceeds $0.2\,\mathrm{kg\,m^{-2}}$ in one week. For these grid cells, a MF was defined that applies to the total anthropogenic NO_x emission. To ensure observability, selected cells that are close together share the same MF. In this experiment, grid cells that are at most 3 cells apart are clustered together. This definition results in 8 clusters of grid cells that receive a MF. These clusters are shown in Figure 5.4b.

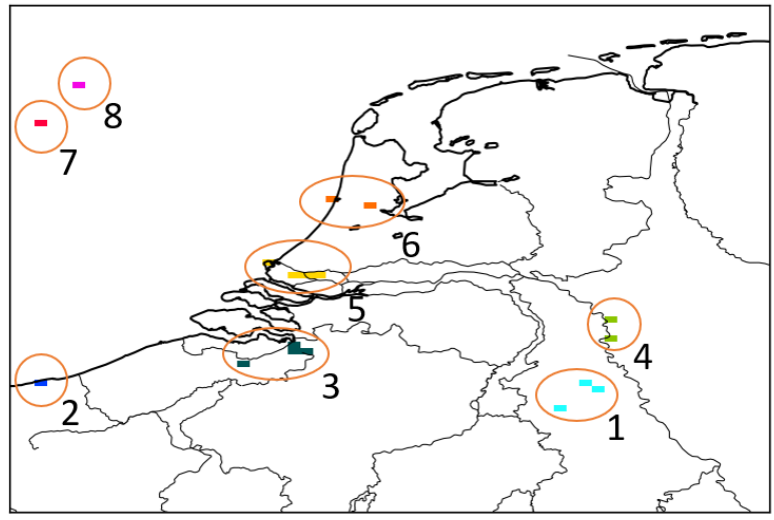
5.4.3. Approximate model

To perform data assimilation using the approximate 4DVAR method outlined in section 4.9, the approximate model needs to be constructed. Columns of the matrix $\mathbf{E}_m$ have to be calculated. These are the derivatives of the state $\mathbf{x}(t_m)$ with respect to $\boldsymbol{\beta}$. Their calculation was defined in Equation (4.8).

On July 1, 2018, at 12:00, four calculated derivatives, after multiplying them with the observation operator $\mathbf{H}_m$ are shown in figure 5.5.



(a)



(b)

Figure 5.4: (a) Total of anthropogenic NO_x emission from July 1, 2018, until July 7, 2018 obtained from CAMS v2.2 database. (b) Clustered grid cells that represent the emission hotspots. Selected grid cells have a NO_x emission of at least $0.2\,kg\,m^{-2}$ in one week. The clusters numbers are used to report the results of the experiments.

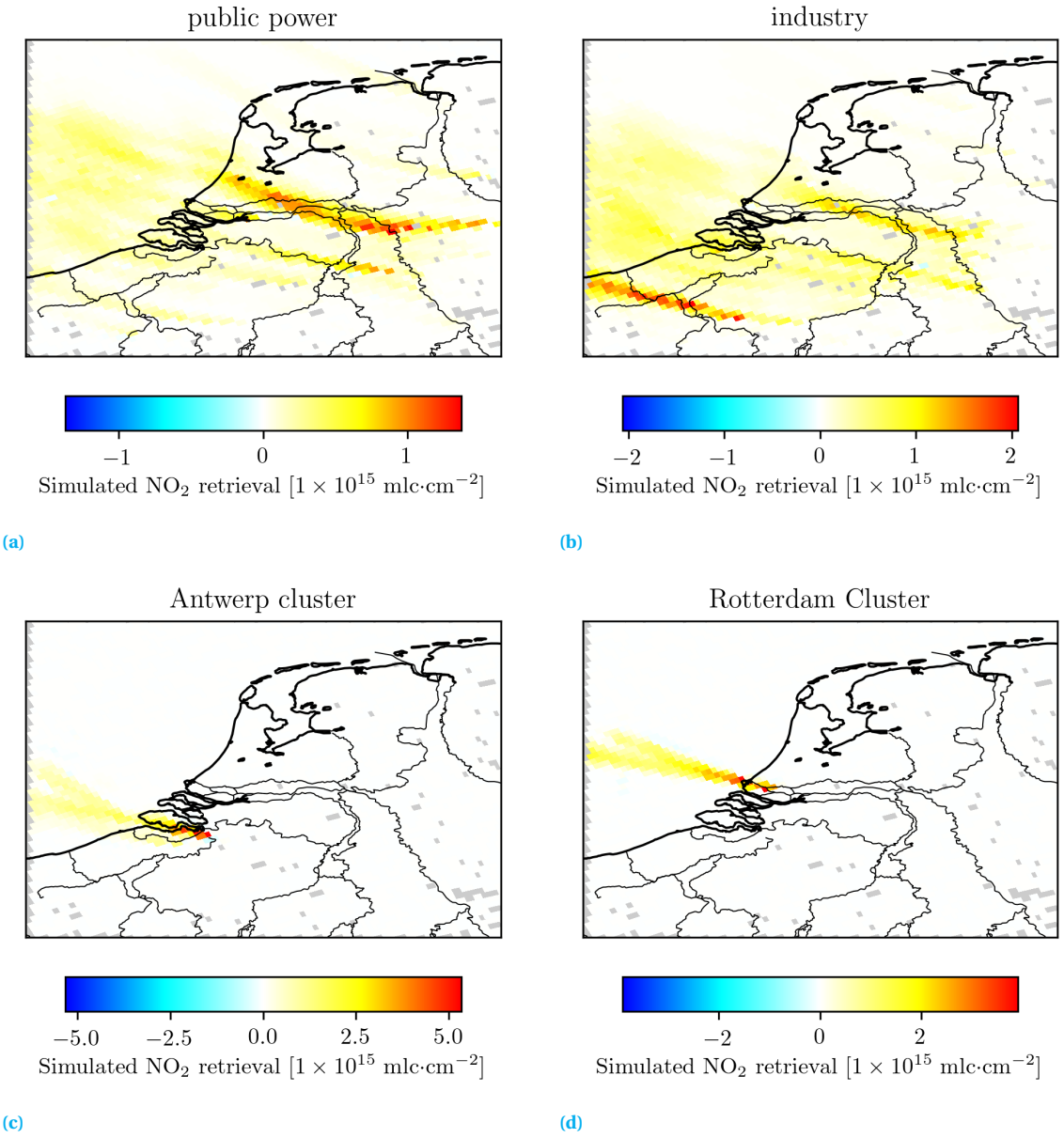


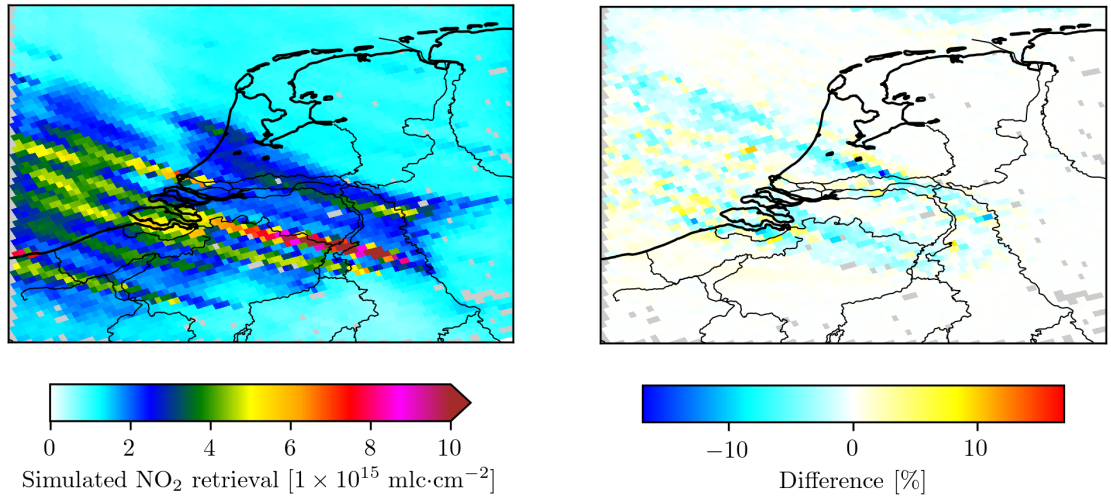
Figure 5.5: Approximate derivative of satellite retrieval at 2018-07-01 12:00 with respect to the parameters of NO_x emission. This derivative at t_m w.r.t β_i can be calculated as $(\mathbf{H}_m\mathbf{E}_m)_i$. **(a) - (b)**: factors for categories **(c)-(d)**: factor for clustered point source Rotterdam and Antwerp port area.

Computational effort required

For the estimation of 13 parameters in this twin experiment, the computational effort of obtaining the approximate model consists of 14 LOTOS-EUROS runs. Costs of calculating the derivatives from the output of the 14 runs are negligible.

Approximate model performance

To estimate the minimum of the exact 4DVAR cost function (4.1), the approximate cost function (4.25) is minimised. This will only yield an acceptable performance if the approximate model can accurately replace the LOTOS-EUROS model. To test this, the simulated satellite observations from to the LOTOS-EUROS run with the true parameters $\boldsymbol{\beta}_{\text{true}}$ can be compared to the satellite observations simulated by the approximate model for parameters $\boldsymbol{\beta}_{\text{true}}$.



(a) (b)

Figure 5.6: Twin experiment 2. Illustration of approximate model accuracy on July 1, 2018 at 12:00. **(a)** Simulated retrieval from $\mathbf{x}_{\text{true}}$. **(b)** Difference between simulated retrieval from truth run and simulated retrieval from the approximate model with $\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{true}}$.

Figure 5.6 shows the accuracy of the approximate model on July 1, 2018, at 12:00. On average, in the assimilation window of one week, the mean absolute error of the simulated retrievals from the approximate model for parameters $\boldsymbol{\beta}_{\text{true}}$ was found to be 2%. This error is significantly less than the uncertainty of the retrieval that the TROPOMI product reports. Therefore, the approximate model is likely to be sufficiently accurate to use in the data assimilation methodology. This is in line with the results from twin experiment 1 where the parameters were estimated correctly using the approximate model. In the next section, the approximate cost function will be compared to the exact cost function for the estimated parameters $\hat{\boldsymbol{\beta}}$ as a second test of the approximate model performance.

5.4.4. Results and discussion

For the assimilation of one week of observations, the results of July 1, 2018, are shown in Figure 5.7. The synthetic satellite observation is shown in Figure 5.7a. In Figure 5.7b, the retrieval corresponding to a model run with the estimated parameters is shown. Figure 5.7c shows the difference between the synthetic observation and the retrieval corresponding to estimated parameters. Figure 5.7d shows the background retrieval i.e. the simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$. The difference between the background retrieval and the retrieval with estimated parameters is visible in the NO_2 plumes.

After constructing the approximate model, the assimilation procedure was carried out. Table 5.2 presents the results using the assimilation window of July 1, until July 7, 2018. Table 5.3 presents the results using the assimilation window of only July 1, 2018. The numbers show that the method can estimate the parameters within 10% of their true value if one week of observations is used. When only the observations on July 1 are assimilated, larger errors in $\hat{\boldsymbol{\beta}}$ are found.

These larger errors are also reflected in the approximated covariance of $\hat{\boldsymbol{\beta}}$. The approximate covariance $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ was calculated using Equation (4.36). The square roots diagonal entries of $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ give an approximation of the uncertainty $\hat{\boldsymbol{\beta}}$. It indicates the measurability of the parameters given the data and the model. When assimilating only one day of observations, the estimated covariance of the parameters is significantly higher. The parameters that correspond to relatively small sources of NO_x such as off-road traffic show have higher uncertainty after assimilation. A parameter that corresponds to relatively larger sources of NO_x such as the port area of Rotterdam (cluster 5) has lower uncertainty. This result is expected: the effect of parameter changes of strong emitters can be measured better in noisy satellite retrievals.

Comparisons of $\hat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_{\text{true}}$ as shown in Tables 5.2 and 5.3 give an indication how accurately the parameters $\boldsymbol{\beta}_{\text{true}}$ can be estimated from the data. It does not describe how well the proposed data assimilation methodology using an approximate model can reduce the cost function. These results can be found in Table 5.4. Four values of the cost function are important. $J(\boldsymbol{\beta}^b)$ is the value of the cost function with background parameters. This is the starting point of the assimilation. $J(\boldsymbol{\beta}_{\text{true}})$ is the value of the cost function calculated from a LOTOS-EUROS run with the true parameters. This can be seen as the optimal value that can be reached during the optimisation. $\tilde{J}(\hat{\boldsymbol{\beta}})$ is the value of the approximate cost function, calculated using Equation (4.25). $J(\hat{\boldsymbol{\beta}})$ is the value of the cost function calculated from a LOTOS-EUROS run with the estimated parameters.

Two important conclusions can be drawn from the results in Table 5.4. The approximate cost fuction value $\tilde{J}(\hat{\boldsymbol{\beta}})$ is very close to the exact cost function $J(\hat{\boldsymbol{\beta}})$. For both assimilation windows, the difference is well less than 1%. The first conclusion, therefore, is that the approximate cost function accurately approximates the exact 4DVAR cost function for the estimated parameters $\hat{\boldsymbol{\beta}}$. The data assimilation method minimises the cost function to a value that is very close to $J(\boldsymbol{\beta}_{\text{true}})$, the difference is under 0.5% for both assimilation windows. For the assimilation window of only July 1, 2018, $J(\hat{\boldsymbol{\beta}})$ is even slightly lower than $J(\boldsymbol{\beta}_{\text{true}})$. This means that the minimisation of the cost function using the approximate model performs well. Additional optimisation loops are therefore not necessary. This is a positive result for the computational cost of the assimilation method. Additional optimisation loops would require additional full model runs to recalculate the approximate model.

For the assimilation window of July 1, 2018, the cost function with estimate parameters $\hat{\boldsymbol{\beta}}$ is slightly lower than the cost function with true parameters $\boldsymbol{\beta}_{\text{true}}$. This means the optimisation of the cost function by the approximate 4DVAR method was successful. Any other optimisation strategy of the 4DVAR cost function discussed in Section 4.2 could find approximately the same optimum. However, as can be seen in Table 5.3, $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\text{true}}$ are significantly different for some parameters. This error can, therefore, be attributed to the error in the measurements. As reflected by the estimated covariance of $\hat{\boldsymbol{\beta}}$, assimilating one day of observations leaves quite some posterior uncertainty for $\hat{\boldsymbol{\beta}}$.

MFs	**True value**	**Estimated value**	**Error**	$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$
Public Power	0.887	0.9260	4.4392%	0.0212 (2.39%)
Industry	1.511	1.4998	-0.7655%	0.0166 (1.10%)
Road transport (F1-F4)	1.076	1.0899	1.2696%	0.0094 (0.87%)
Shipping	0.479	0.5202	8.6788%	0.0069 (1.45%)
Off-road	0.9	0.8043	-10.6512%	0.0663 (7.37%)
Cluster 1	2.106	2.1095	0.1712%	0.0098 (0.46%)
Cluster 2	1.413	1.3801	-2.3296%	0.0259 (1.84%)
Cluster 3	0.7	0.6825	-2.4661%	0.0105 (1.50%)
Cluster 4	0.497	0.4877	-1.8468%	0.0228 (4.58%)
Cluster 5	1.203	1.1824	-1.7318%	0.0124 (1.03%)
Cluster 6	0.582	0.5853	0.5682%	0.0263 (4.52%)
Cluster 7	1.648	1.6411	-0.4355%	0.0271 (1.64%)
Cluster 8	0.787	0.7958	1.0771%	0.0209 (2.65%)

Table 5.2: Results of twin experiment 2 with assimilation window July 1, 2018, until July 7, 2018. The error is calculated as absolute difference between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\text{true}}$. The entries under $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ are the values on the diagonal of the estimated covariance matrix.

MFs	**True value**	**Estimated value**	**Error**	$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$
Public Power	0.887	1.012	14.117%	0.0953 (10.75%)
Industry	1.511	1.624	7.477%	0.0836 (5.53%)
Road transport (F1-F4)	1.076	1.063	-1.192%	0.0373 (3.46%)
Shipping	0.479	0.399	-16.614%	0.0472 (9.85%)
Off-road	0.900	0.772	-14.212%	0.2994 (33.26%)
Cluster 1	2.106	2.141	1.676%	0.0294 (1.40%)
Cluster 2	1.413	1.563	10.644%	0.1917 (13.57%)
Cluster 3	0.700	0.737	5.383%	0.0537 (7.68%)
Cluster 4	0.497	0.396	-20.290%	0.0883 (17.77%)
Cluster 5	1.203	1.303	8.322%	0.0674 (5.60%)
Cluster 6	0.582	0.576	-1.073%	0.1317 (22.63%)
Cluster 7	1.648	1.320	-19.926%	0.2901 (17.60%)
Cluster 8	0.787	0.418	-46.916%	0.2555 (32.45%)

Table 5.3: Results of twin experiment 2 with assimilation window July 1, 2018. The error is calculated as absolute difference between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\text{true}}$. The entries under $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ are the values on the diagonal of the estimated covariance matrix.

Cost function		Assimilation window	
		July 1, 2018	July 1, 2018 - July 7, 2018
Cost function with background parameters	$J(\boldsymbol{\beta}^b)$	9.572×10^3	7.296×10^4
Cost function with true parameters	$J(\boldsymbol{\beta}_{\text{true}})$	7.083×10^3	4.678×10^4
Approximate cost function with estimated parameters	$\tilde{J}(\hat{\boldsymbol{\beta}})$	7.079×10^3	4.713×10^4
Cost function with estimated parameters	$J(\hat{\boldsymbol{\beta}})$	7.077×10^3	4.684×10^4

Table 5.4: 4DVAR cost function values in twin experiment 2. The cost function value with $\boldsymbol{\beta}^b$, $\boldsymbol{\beta}_{\text{true}}$ and $\hat{\boldsymbol{\beta}}$ are calculated from a LOTOS-EUROS run with those parameters. The approximate cost function with parameters $\hat{\boldsymbol{\beta}}$ is calculated using the approximate model, Equation (4.25).

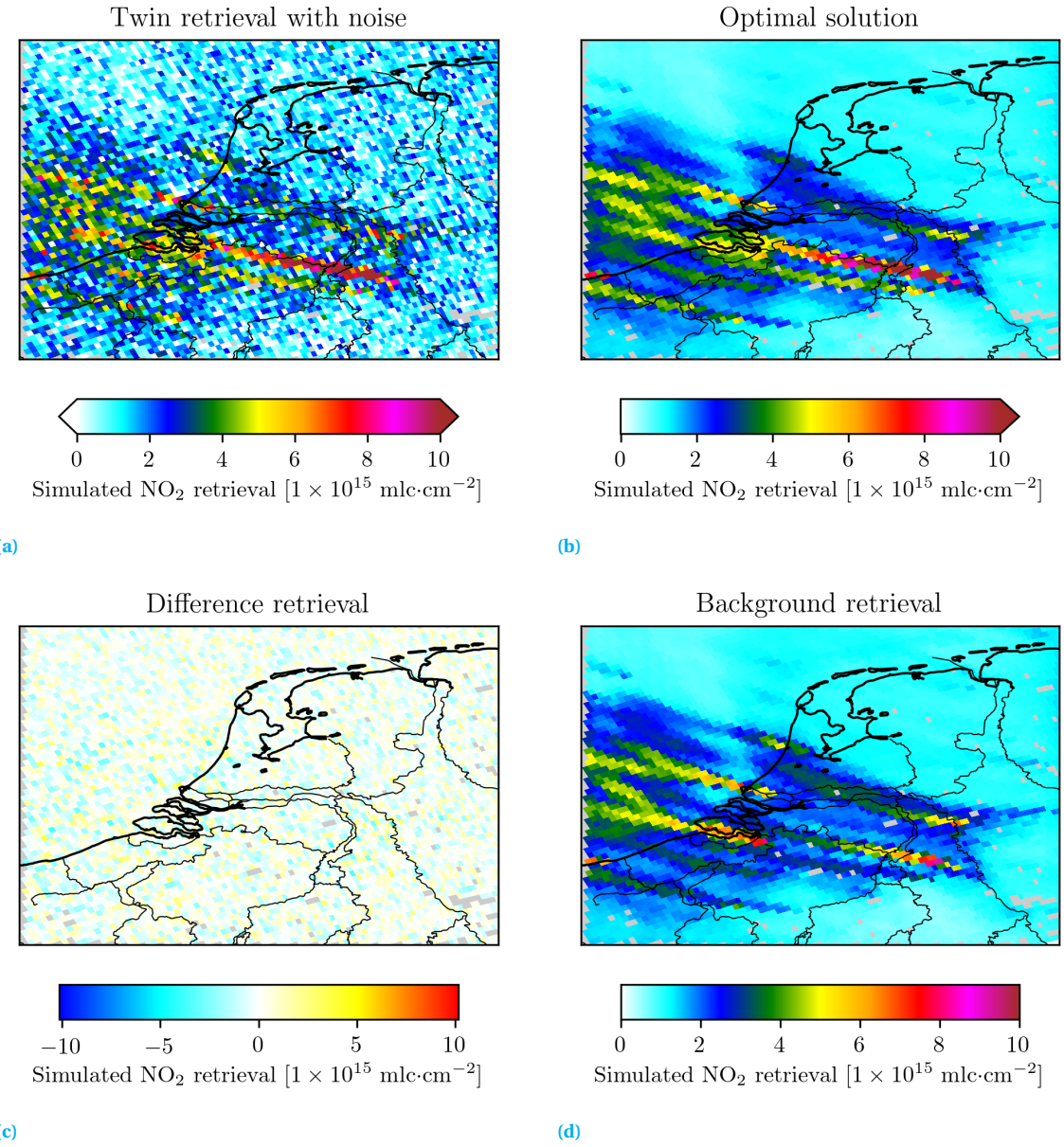


Figure 5.7: Twin experiment 2 results at 2018-07-01 12:00. **(a)** Synthetic satellite observation **(b)** Optimal solution **(c)** Difference between twin retrieval and optimal solution **(d)** Retrieval with background parameters $\boldsymbol{\beta}_b = \mathbf{1}$.

5.5. Twin experiment 3 - Domain decomposition

This section describes the setup and results of the third twin experiment that was performed. This experiment aims to increase the computational efficiency of calculating the approximate model through domain decomposition as described in Section 4.10. For the construction of the approximate model, the computational grid is decomposed in subdomains on which the approximate model is calculated. In one LOTOS-EUROS run, parameters in multiple subdomains are perturbed simultaneously. In this way, fewer runs of LOTOS-EUROS are necessary for the calculation of the approximate model.

In twin experiments, wrong estimates of $\boldsymbol{\beta}$ after data assimilation are either a result of errors in the approximate model that is calculated or due to the noise used to perturb the synthetic observations. To distinguish these sources of error, the synthetic observations generated in this experiment are not perturbed by noise. Resulting errors in the estimates of $\boldsymbol{\beta}$ then have to be the result of errors in the approximate model.

5.5.1. Test region & time

A grid is chosen with coordinates $[1°, 14°] \times [48°, 54°]$ and resolution $0.1° \times 0.05°$. This grid is illustrated in Figure 5.8. This translates to a resolution of 7×6 km. This domain is larger than in the first two twin experiments. The assimilation period is again July 1, 2018, until July 7, 2018.

Parametrisation

In this experiment, MFs are applied to emission hotspots according to the procedure in the second twin experiment. Figure 5.8, shows the total emission from July 1, 2018, until July 7, 2018. The domain is decomposed into 4 subdomains. The boundary between the subdomains is at 8° longitude and 51° latitude. Within each of these domains, the largest 5 emission hotspots get a multiplication factor. The subdomains and the clusters of grid cells that get a multiplication factor are shown in Figure 5.9.

In total, 20 parameters have to be estimated in this twin experiment. The parameters are modelled as independent Gaussian, $\boldsymbol{\beta} \sim N(\mathbf{1}, \sigma^2 \mathbf{I})$, where $\sigma = 0.2$.

Approximate model calculation

Using the subdomain decomposition method as outlined in Section 4.10, the calculation of the approximate model can be done more efficiently. Several perturbation strategies are investigated in this twin experiment.

- **One-by-one perturbation**

  Parameters are perturbed one-by-one. This corresponds to the original strategy of calculating the approximate model. This strategy was also performed in this experiment for comparison to more efficient strategies. This requires a total of 20 perturbation runs and one base run to calculate the approximate model.

- **Simultaneous perturbation**

  In each subdomain, one parameter is perturbed simultaneously. This requires a total of 5 perturbation runs and one base run to calculate the approximate model.

- **Diagonal perturbation**

  One parameter is perturbed in domains 1&4 and 2&3 at the same time. This requires a total of 10 perturbation runs and one base run to calculate the approximate model.

After the base and perturbation runs of LOTOS-EUROS, the approximate model is calculated by Equation (4.8). The derivatives are restricted to the subdomains on which the parameters are defined.

Errors in the calculated approximate model using domain decomposition may arise from two origins. Firstly, the parameters from one domain can influence other subdomains. Secondly, the limitation of the approximate model to the subdomains, since parts of the plume that start in one domain and then reach another may be truncated. These errors are illustrated in Figure 5.10. Both the approximate model in subdomain 1 and 3 have errors. In domain 3, not only the effect of changing β_3 is present, but also the effect of changing β_1. The effect in domain 3 is only attributed to β_3. In domain 1, by restricting computation of the approximate model to domain 1, part of the effect of $\beta 1$ is not calculated.

The results of the assimilation experiment are shown in Table 5.5.

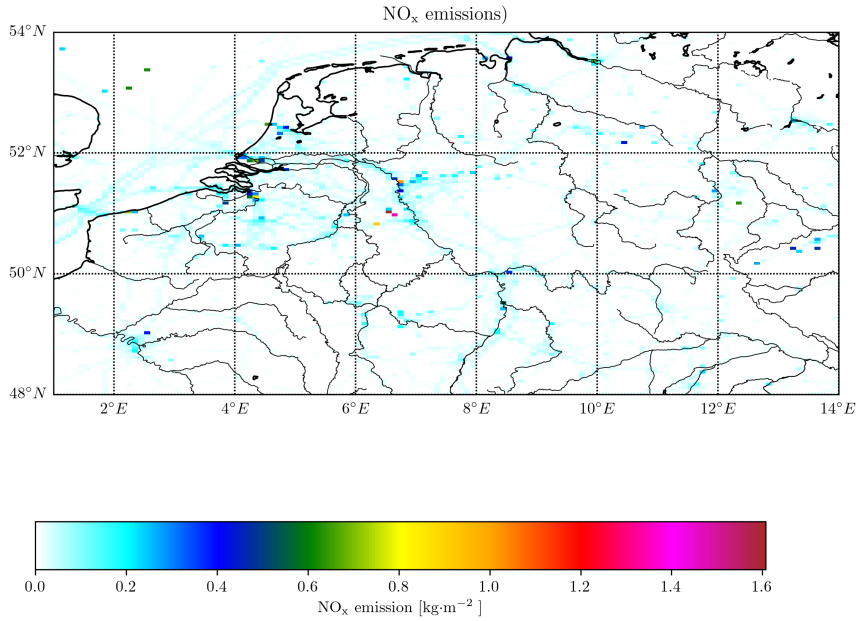
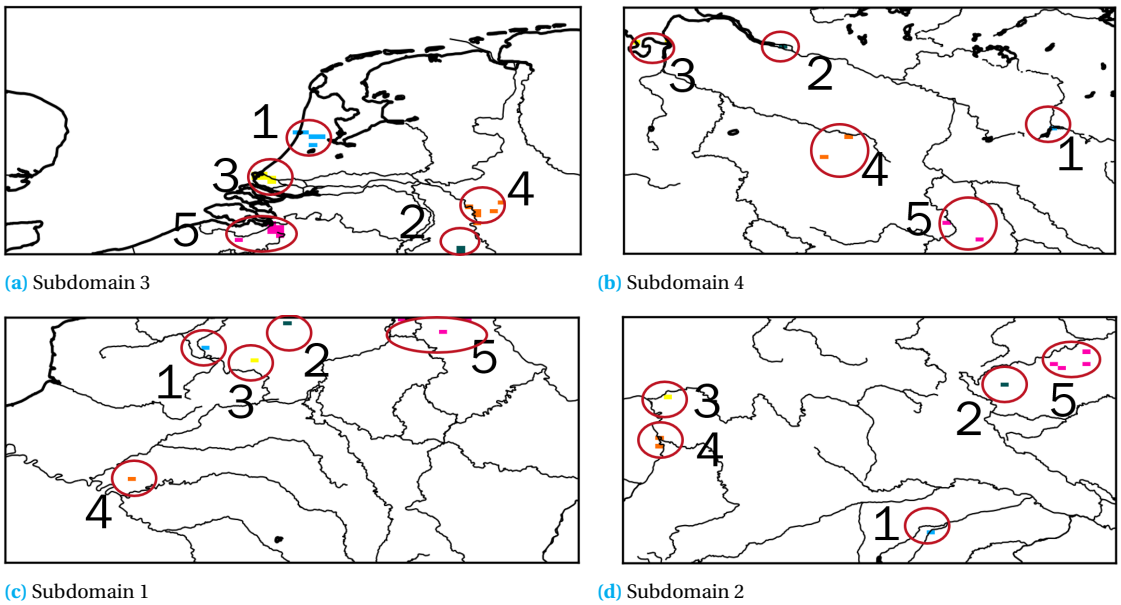


Figure 5.8: Total NO_x emission from July 1, 2018, until July 7, 2018 on the grid of twin experiment 3



(a) Subdomain 3

(b) Subdomain 4

(c) Subdomain 1

(d) Subdomain 2

Figure 5.9: Domain decomposition of the third twin experiment and clusters of gridcells that receive a MF.

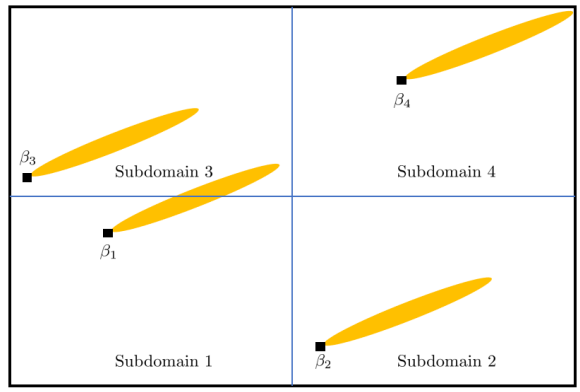


Figure 5.10: Illustration of a scenario where errors occur in the calculation of the approximate model using domain decomposition

5.5.2. Results and discussion

The columns under **original method** shows that calculating the approximate model without domain decomposition is accurate. The original method correctly estimates the parameters.

However, when restricting the derivatives of parameters to their respective subdomains errors arise. Using the one-by-one perturbation strategy, the only error made in the calculation of the approximate model is the limitation of the calculated derivatives to the subdomains. Significant errors arise for the clusters in domains 1 and 3 that are close to the boundaries of these domains. Other MFs that are defined more centrally in the subdomains are accurately estimated. It can be concluded that using a domain decomposition where plumes cross subdomain boundaries should be prevented.

Table 5.5 shows that more efficient perturbation strategies generally do not further deteriorate the results. The errors in $\hat{\boldsymbol{\beta}}$ do not increase significantly using these strategies. The exception is the MF for cluster 2 in domain 2. Both using the diagonal and simultaneous perturbation methods results in a 10% error in the estimate of the MF. This can be explained by the influence of the very large emission hotspot corresponding to cluster 2 in domain 3. This MF is perturbed simultaneously with cluster 2 in domain 2 in both the diagonal and simultaneous perturbation method. The plume from this hotspot reaches domain 2, resulting in an error in the approximate model.

In conclusion, domain decomposition increases the computational efficiency of the method. Errors mainly occur due to the limitation of the approximate model to the subdomains. To prevent this, the subdomains should be chosen in such a way that specified MFs are not close to the downwind boundaries. In hindsight, the grid chosen for this twin experiment was too small and the densely populated Benelux area is not suitable for domain decomposition. On larger grids, domain decomposition can be used to increase the efficiency of calculating the approximate model.

Another possible approach is to use a more flexible definition of the subdomains on which the derivatives are calculated instead of the static domain decomposition chosen in this experiment. In Figure 5.10, the plumes that originate from β_1 crosses the boundary of the subdomain 3. However, as the plumes do not overlap, it is possible to define the area on which the derivatives of the state with respect to β_1 and β_3 are calculated to capture each plume correctly. This definition may vary in time to accommodate for changing meteorological conditions. The only requirement for the perturbation strategy is that the plumes from sources that are perturbed simultaneously do not overlap.

In this twin experiment, by using 4 subdomains, it was possible to perturb 4 parameters simultaneously for calculation of the approximate model. This means that the number of perturbation runs needed was reduced by a factor 4. In general, when increasing the number of parameters N_p, the number of perturbation runs needed without domain decomposition is N_p. If computational resources become critical due to a larger number of parameters used, domain decomposition may increase computational efficiency. Theoretically, any NO_x sources for which the derivative states are spatially separable during the assimilation period can be perturbed simultaneously.

MFs		**True value**	**Original method**		**Subdomain one-by-one**		**Subdomain diagonal**		**Subdomain simultaneous**	
			$\hat{\boldsymbol{\beta}}$	Error	$\hat{\boldsymbol{\beta}}$	Error	$\hat{\boldsymbol{\beta}}$	Error	$\hat{\boldsymbol{\beta}}$	Error
Domain 1	Cluster 1	1.325	1.322	0%	1.390	5%	1.388	5%	1.307	-1%
	Cluster 2	0.878	0.879	0%	1.396	59%	1.409	61%	0.594	-32%
	Cluster 3	0.894	0.898	0%	1.032	15%	1.036	16%	0.938	5%
	Cluster 4	0.785	0.803	2%	0.812	3%	0.812	3%	0.903	15%
	Cluster 5	1.173	1.174	0%	0.980	-16%	0.981	-16%	1.223	4%
Domain 2	Cluster 1	0.540	0.555	3%	0.549	2%	0.547	1%	0.545	1%
	Cluster 2	1.349	1.345	0%	1.345	0%	1.211	-10%	1.212	-10%
	Cluster 3	0.848	0.856	1%	0.841	-1%	0.839	-1%	0.845	0%
	Cluster 4	1.064	1.065	0%	1.050	-1%	1.028	-3%	1.036	-3%
	Cluster 5	0.950	0.952	0%	0.954	0%	0.956	1%	0.979	3%
Domain 3	Cluster 1	1.292	1.291	0%	1.290	0%	1.290	0%	1.291	0%
	Cluster 2	0.588	0.590	0%	0.671	14%	0.671	14%	0.549	-7%
	Cluster 3	0.936	0.935	0%	0.935	0%	0.935	0%	0.934	0%
	Cluster 4	0.923	0.925	0%	0.930	1%	0.930	1%	0.923	0%
	Cluster 5	1.227	1.226	0%	1.234	1%	1.234	1%	1.223	0%
Domain 4	Cluster 1	0.780	0.797	2%	0.790	1%	0.790	1%	0.803	3%
	Cluster 2	0.966	0.967	0%	0.966	0%	0.966	0%	0.938	-3%
	Cluster 3	0.824	0.832	1%	0.832	1%	0.832	1%	0.836	1%
	Cluster 4	1.008	1.009	0%	1.006	0%	1.006	0%	0.978	-3%
	Cluster 5	1.117	1.112	0%	1.118	0%	1.082	-3%	1.097	-2%

Table 5.5: The results of twin experiment 3. The parameters and their true value are shown in columns **MFs** and **True Value**. The column **Original method** shows the result of the twin experiment where the approximate model is calculated by the original method as in the first two twin experiments. The columns **Subdomain** show the results of the experiment where the approximate models are calculated on and limited to the subdomains. Three perturbation strategies are studied, **one-by-one** perturbs one parameter at a time,**diagonal** perturbs parameters in domains 1&4 and 2&3 simultaneously and **simultaneous** perturbs parameters is all subdomains at the same time.

6

Assimilation of TROPOMI data

This chapter describes experiments with the assimilation of actual TROPOMI NO_2 retrievals using the methodology discussed in Chapter 4. The experiments aim to estimate MFs for the NO_x emission strengths.

Two experiments will be described in this section. The first experiment estimates a MF for one cluster of point sources in the Amsterdam metropolitan area. The second experiment estimates MFs for 5 emission categories and 8 clusters of point sources. The parameterisation in this experiment is equal to twin experiment 2 from Chapter 5.

6.1. Test region & time

The same LOTOS-EUROS setup is used for both experiments. A grid is chosen with coordinates $[2°, 8°] \times [50°, 54°]$ and a resolution of $0.1° \times 0.05°$ (about 7×6 km). A total of $60 \times 80 = 4.800$ horizontal grid cells are used. With 10 vertical layers, this yields a total of 48.000 grid points where NO_2 concentrations are calculated. The time span of the base model run is the two weeks from June 23 until July 7, 2018. One week was added to the assimilation window of the twin experiment. In this week, from June 23, until June 30, 2018, some clouded days are present. The TROPOMI retrievals are shown in Appendix A.

Three assimilation windows will be investigated. The first is the full two-week window from June 23 until July 7, 2018. Also, correction factors are estimated for the individual days June 25 and July 1. July 1, 2018, is a cloud-free day on which NO_2 plumes are resolved with high detail in the TROPOMI retrievals. On June 25, 2018, plumes are less clear in the TROPOMI retrievals because of lower wind speeds.

6.2. Experiment 1 - Estimation of a single cluster of point sources

This section describes the parameterisation and results of the first experiment.

6.2.1. Parameterisation

In the first experiment, a correction factor for the emission strength of NO_x is estimated for a cluster of point sources in the Amsterdam metropolitan area. The emission from the city centre and the Tata Steel plants at IJmuiden are perturbed with one, common factor. These grid cells are highlighted in figure 6.1. The prior distribution of $\boldsymbol{\beta}$ is Gaussian with mean 1 and standard deviation 0.3.

6.2.2. Results

For three different assimilation periods, the optimal MF found is reported in table 6.1. From this table, it is clear that the data assimilation procedure is flawed. The negative correction factors found for the window from June 23, 2018, until July 7, 2018, and for only June 25, 2018, are non-physical solutions. They would correspond to a sink of NO_x. However, these negative values are mathematically valid solutions of the approximate 4DVAR optimisation problem (4.25).

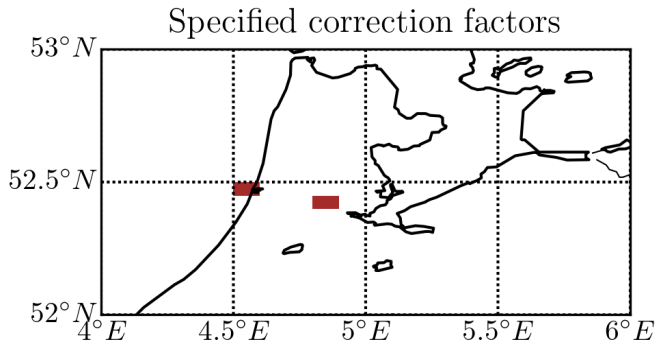


Figure 6.1: Grid cells for which the emission strength is corrected in the first experiment. One MF corrects the NO_x emission strength from both grid cells.

Correction factors	**Assimilation window**		
	25-6-2018	1-7-2018	23-6-2018 - 7-7-2018
Amsterdam & Tata	−1.1257	1.7203	−1.0653
Cost function reduction	4.73%	0.42%	4.25%

Table 6.1: Results of assimilation experiment with the parametrisation as outlined in Section 6.2. The reduction of the cost function is calculated as the difference between the approximate cost function (Equation (4.25)) with estimated parameters $\hat{\boldsymbol{\beta}}$ and background parameters $\boldsymbol{\beta}^b$.

Figure 6.2 visually presents the result of the one-day window of July 1, 2018. In Figure 6.2a, the TROPOMI retrieval shows clear NO_2 plumes propagating to the northwest. The simulated retrieval with background parameters in Figure 6.2d shows a similar plume from the Amsterdam metropolitan area and IJmuiden. In this assimilation window, the plumes in the simulated and TROPOMI retrievals look quite similar. The single emission parameter in this experiment was estimated higher than its background value. Figure 6.2b shows the retrieval calculated with the estimated parameter of 1.7203. Especially further offshore the fit of the simulated retrieval to the TROPOMI retrieval is improved.

The assimilation using the one-day window of June 25, 2018, is further investigated in figure 6.3. In figure 6.3a, the TROPOMI retrieval shows a clear NO_2 plume propagating to the southeast. The simulated retrieval with background parameters in figure 6.3d shows a slightly wider NO_2 plume that travels in a more southerly direction. Figure 6.4 illustrates this difference in plume location. Figure 6.4a shows the TROPOMI retrieval and figure 6.4b shows the derivative of the simulated retrieval with respect to the parametrisation. Given the parametrisation, the only degree of freedom allowed in the assimilation is the NO_x emission strength from the two grid cells highlighted in figure 6.1. The NO_2 plumes from the grid cells propagate along a slightly different trajectory. Because of the high resolution of the TROPOMI retrievals, this difference is distinguishable. Furthermore, the tail of the modelled plume partly overlaps with the NO_2 plumes from Rotterdam and Antwerp. The NO_2 concentrations in these plumes are higher in the model than in the TROPOMI retrievals. The only way to correct for this difference is to lower the emission strength from the Amsterdam metropolitan area and the Tata steel plants. This explains the unrealistic, negative value estimate for the parameter in this experiment.

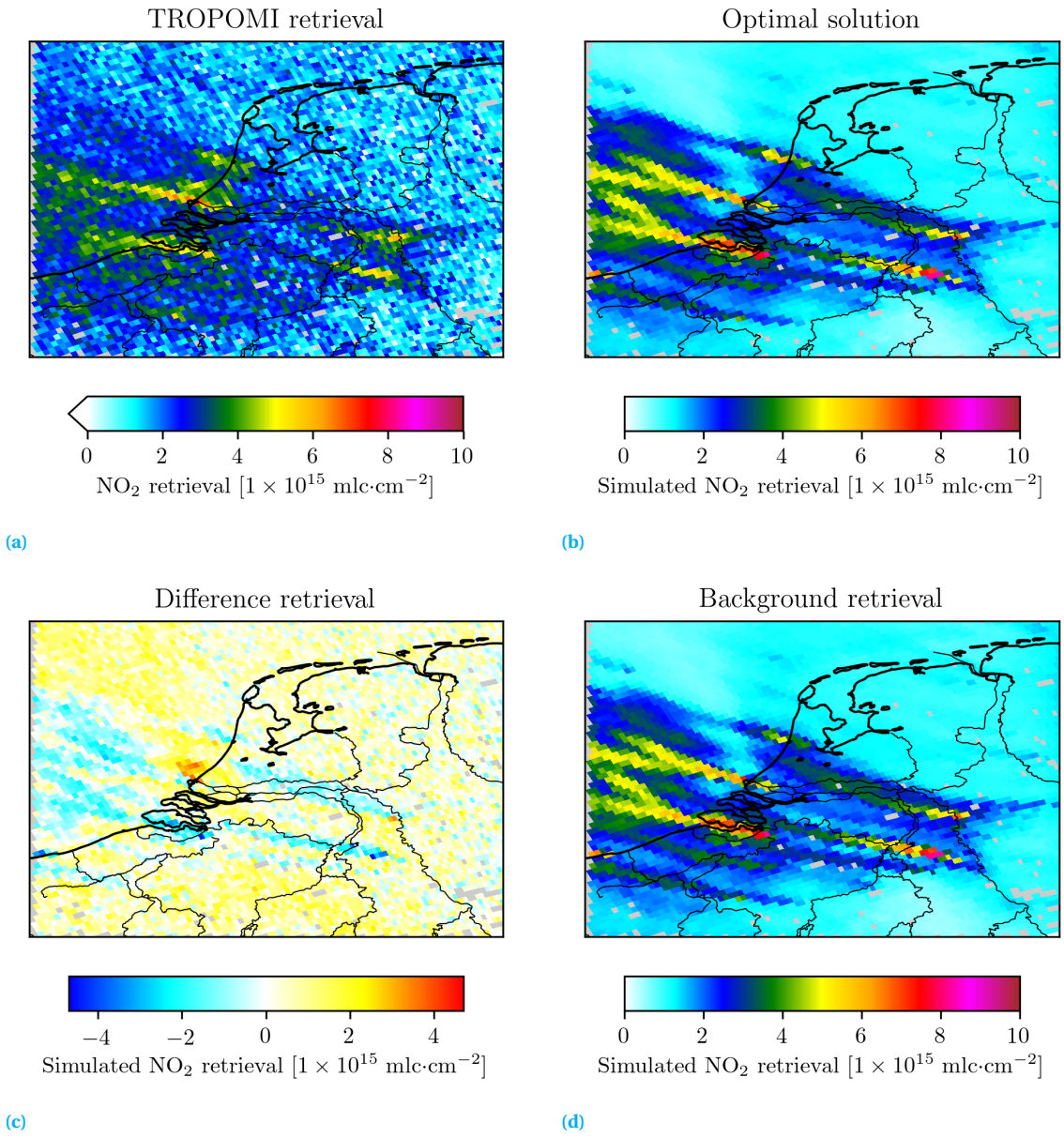


Figure 6.2: TROPOMI experiment 1 results on July 1, 2018 at 12:00. **(a)** TROPOMI retrieval **(b)** Simulated retrieval with estimated parameters $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ **(c)** Difference between TROPOMI retrieval and optimal solution **(d)** Simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$

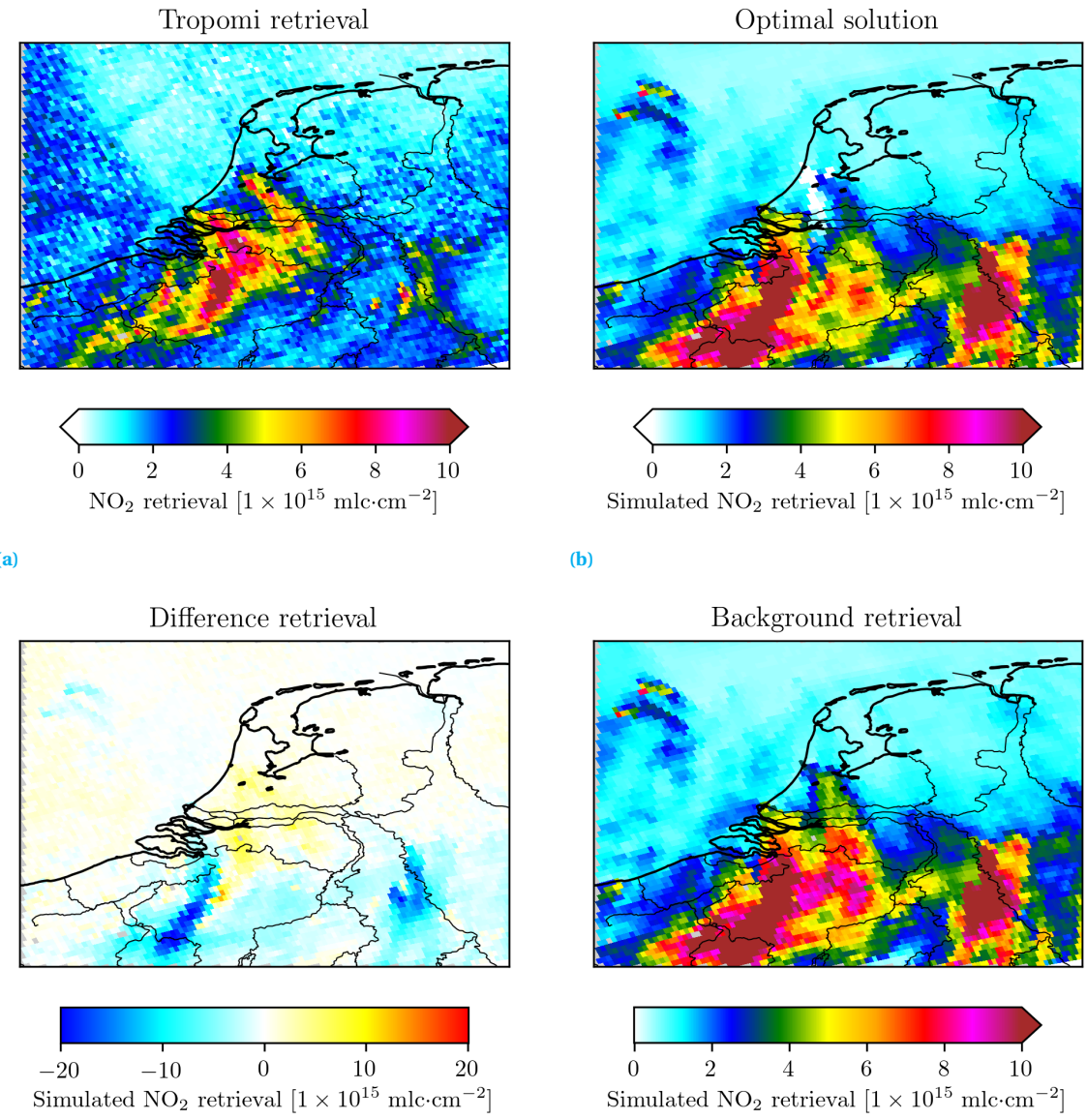
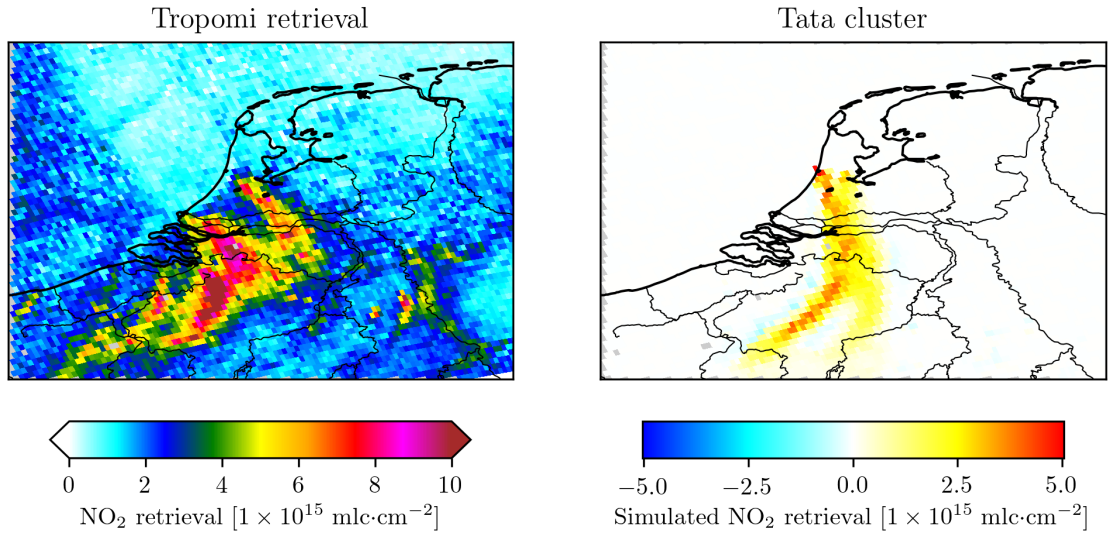


Figure 6.3: TROPOMI experiment 1 results on June 25, 2018, at 12:00. **(a)** TROPOMI retrieval **(b)** Simulated retrieval with estimated parameters $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ **(c)** Difference between TROPOMI retrieval and optimal solution **(d)** Simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$



(a) TROPOMI retrieval at 2018-06-25 12:00

(b) Derivative of simulated retrieval with respect to correction factor

Figure 6.4: Comparison of TROPOMI retrieval and calculated derivative from the LOTOS-EUROS model

6.3. Experiment 2 - Multiple clusters of points sources and categories

This section describes the parameterisation and results of the first experiment where 13 MFs are estimated.

6.3.1. Parameterisation

The same parametrisation was used for the emission factors as in the second twin experiment, described in Section 5.4. NO_x emissions strengths from 5 categories (industry, public power, road transport, shipping and off-road transport) and 8 clusters of point sources are corrected by a MF. Figure 5.4 illustrates the definition of the 8 clusters of point sources for which the emissions are estimated. The prior distribution of the parameters is Gaussian: $\boldsymbol{\beta}^b \sim N(\mathbf{1}, 0.3^2\boldsymbol{I})$. In the cost function, therefore, $\mathbf{B} = 0.3^2\boldsymbol{I}$.

6.3.2. Results

For the three assimilation windows, the estimated MFs are presented in Table 6.2. Figures 6.5 and 6.6 graphically present the results of the assimilation on July 1, 2018, and June 25, 2018, respectively.

In comparison to the results to Table 6.1, the reduction of the cost function is larger for each assimilation window. This is an indication of a better fit of the simulated retrieval using the estimated correction factors than in the first experiment. This logically follows from the higher degree of freedom in the parametrisation. However, non-physical, negative estimates for the factors are still found. Furthermore, the TROPOMI retrieval in Figure 6.6a and the simulated retrieval with the estimated correction factors in Figure 6.6b visually still show significant differences. For example, the NO_2 plume from Amsterdam and IJmuiden is underestimated, but less severely than in the first assimilation experiment. The configuration of more degrees of freedom partly avoids that emissions from certain grid cells are not used to compensate for other emissions that cannot be perturbed, as was the case in the first experiment.

As shown by Figure 6.6, the assimilation procedure is not able to correct a spatial mismatch between modelled plumes and plumes present in the TROPOMI retrievals. It is only possible to correct the intensity of the modelled plumes through the MFs in the parameterisation. Therefore, significant differences remain between the simulated retrievals with optimised parameters and the TROPOMI retrievals in this experiment.

Correction factors	**Assimilation window**		
	25-6-2018	1-7-2018	23-6-2018 - 7-7-2018
Public power	0.4767	0.2568	0.7608
Industry	0.2923	0.4094	0.6554
Road transport	0.5065	1.1056	0.9269
Shipping	1.0680	1.0095	0.8385
Off-road vehicles	−0.2461	5.2912	0.5832
Cluster 1	0.0189	0.5787	0.4301
Cluster 2	0.8391	0.0626	0.3497
Cluster 3	−0.6395	0.2403	−0.1389
Cluster 4	−0.1348	0.5435	0.3530
Cluster 5	0.3791	0.4557	−0.0076
Cluster 6	0.9776	1.2288	0.2934
Cluster 7	0.6794	0.9538	0.2855
Cluster 8	0.2692	1.7213	0.1127
Cost function reduction	77.42%	12.30%	42.13%

Table 6.2: Results of assimilation experiment with the parametrisation as outlined in Section 6.3. The reduction of the cost function is calculated as the difference between the approximate cost function (Equation (4.25)) with estimated parameters $\hat{\boldsymbol{\beta}}$ and background parameters $\boldsymbol{\beta}^b$.

Analysis of the cost funtion

The value of 4DVAR cost function has been reduced by 77.42% for the assimilation window of June 25, 2018. Figure 6.7 provides more details of the observational part of the cost function. The observational part of the cost function is the second term of Equation (4.1). It is associated with the mismatch between real observations and simulated observations.

Figure 6.7a shows the observational part of the approximate cost function (Equation (4.25)) with the estimated parameters $\hat{\boldsymbol{\beta}}$. Figure 6.7b shows the observational part of the cost function with background parameters $\boldsymbol{\beta}^b$. Figure 6.7c shows the percentual change of the cost function in each pixel between the cost function with background parameters and the approximate cost function with estimated parameters. Figure 6.7d shows how the share of each pixel in the total improvement of the cost function. This is the ratio of the cost function improvement per pixel and the total cost function improvement.

From Figure 6.7d, it can be seen that the most significant reductions of the cost function are spatially concentrated around large plumes found in the TROPOMI retrievals. Figure 6.7c shows that although the overall cost function decreases, the reduction is not uniform throughout the domain. In some places, the cost function increases to allow more substantial decreases in the cost function elsewhere. This weighting is a result of the chosen 4DVAR cost criterion that has to be minimised.

The maps show that when plumes in LOTOS-EUROS and TROPOMI retrievals have different positions, the reduction of the cost function does not necessarily imply a better fit of the model to the data.

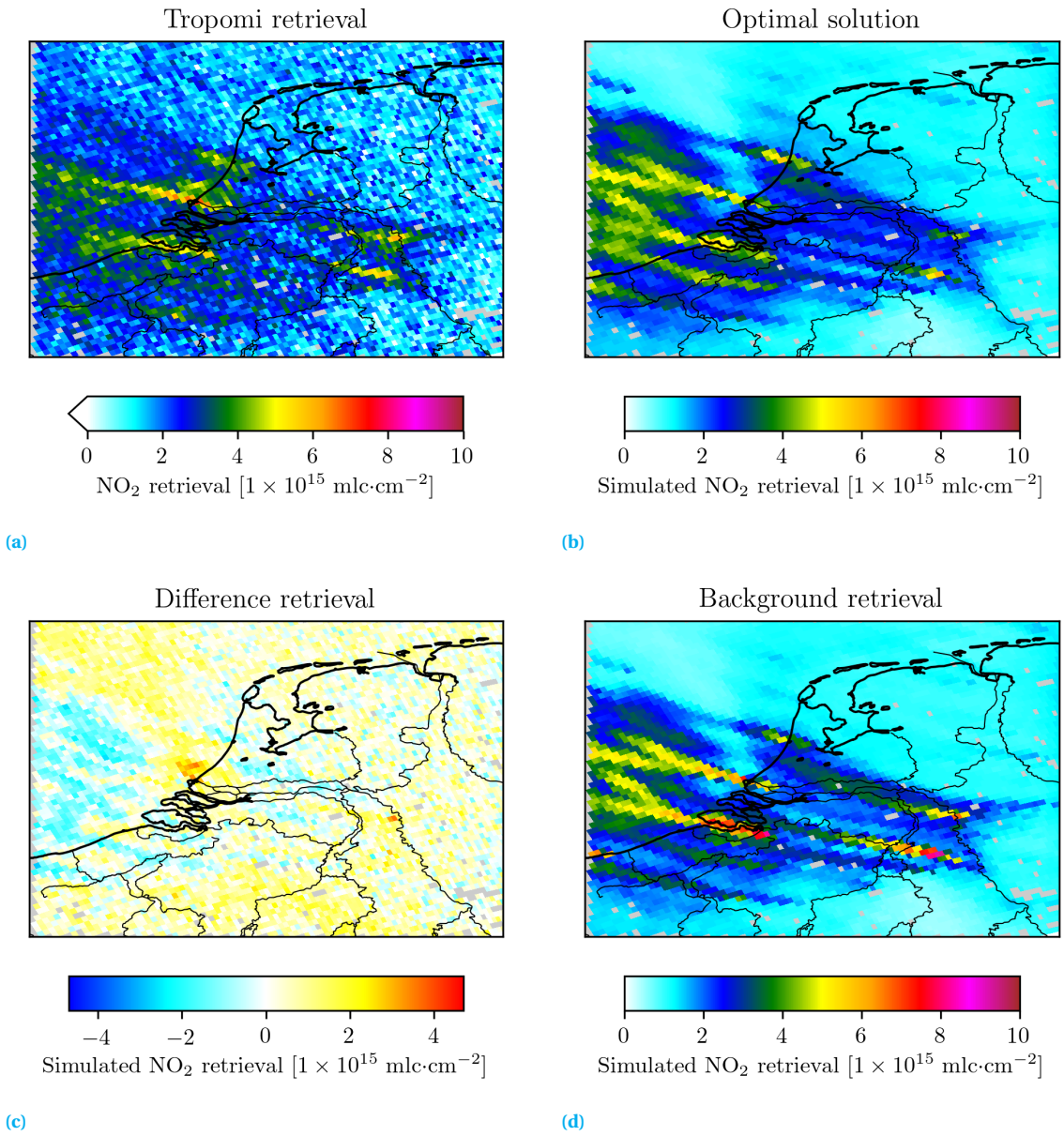


Figure 6.5: TROPOMI experiment 2 results on July 1, 2018, at 12:00. **(a)** TROPOMI retrieval **(b)** Simulated retrieval with estimated parameters $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ **(c)** Difference between TROPOMI retrieval and optimal solution **(d)** Simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$

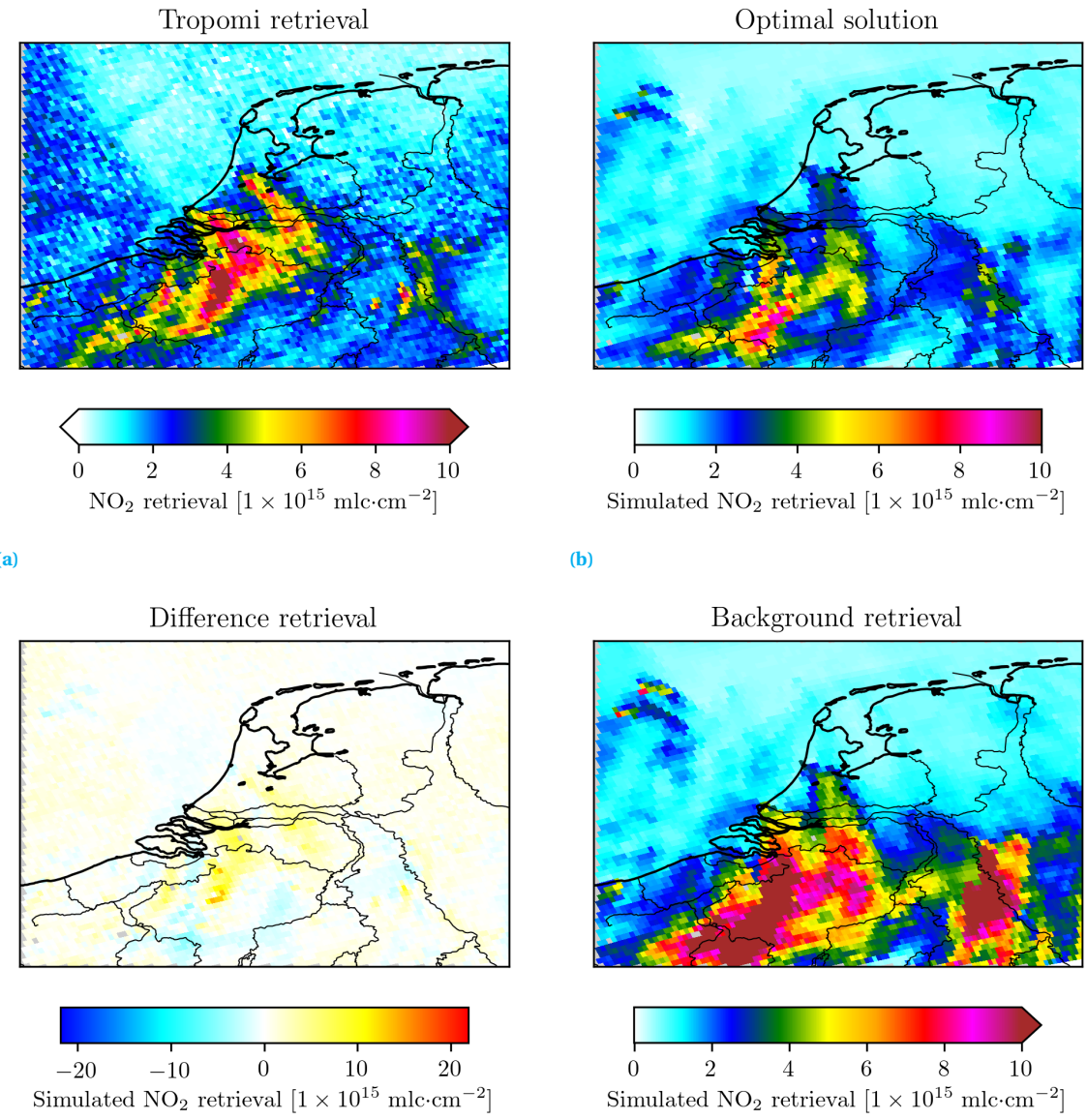
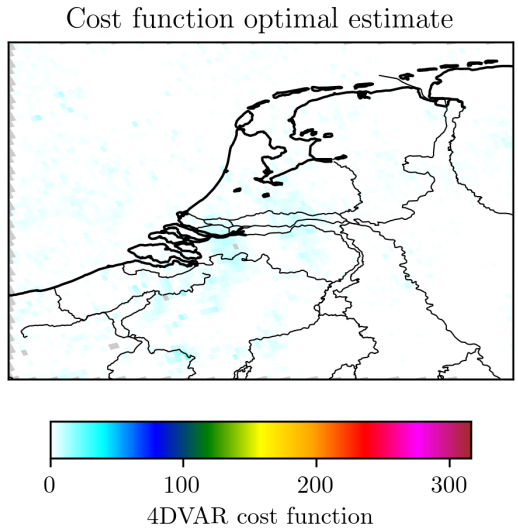
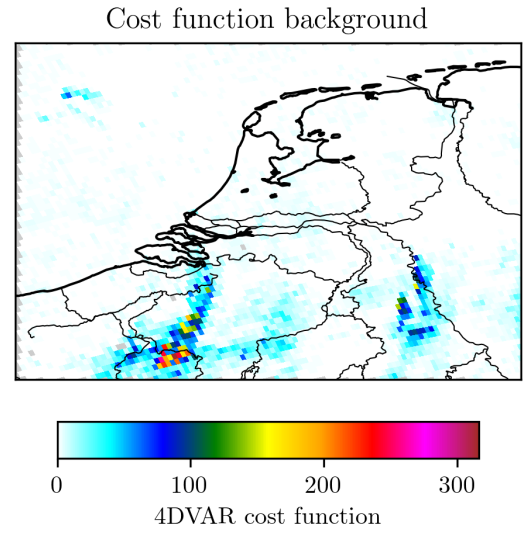


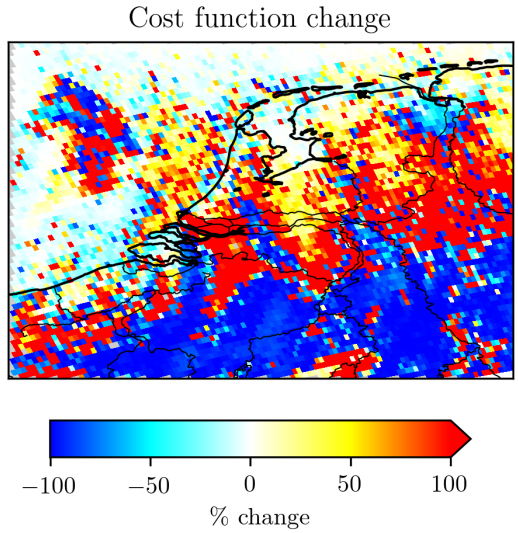
Figure 6.6: TROPOMI experiment 2 results on June 25, 2018, at 12:00. **(a)** TROPOMI retrieval **(b)** Simulated retrieval with estimated parameters $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ **(c)** Difference between TROPOMI retrieval and optimal solution **(d)** Simulated retrieval with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$



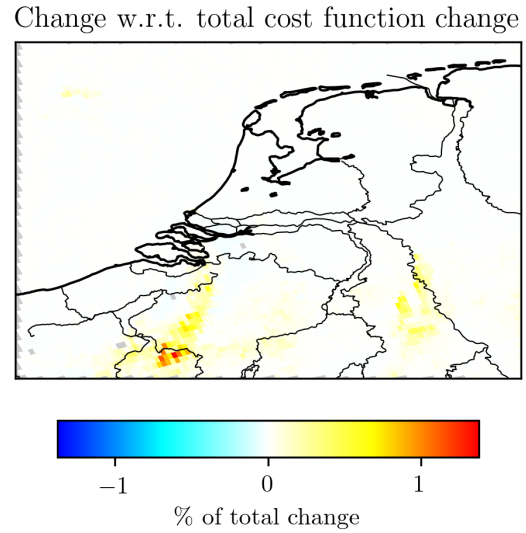
(a) Approximate 4DVAR cost function with estimated parameters $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$



(b) 4DVAR cost function with background parameters $\boldsymbol{\beta}^b = \mathbf{1}$



(c) Percentage change of cost function at each observation footprint



(d) Percentage of total cost function change at each observation footprint

Figure 6.7: Distribution of the spatial part of the 4DVAR cost function and changes after optimisation.

6.4. Conclusions of TROPOMI experiments

This section discusses the conclusions of the two experiments using TROPOMI retrievals that were conducted in this chapter. In general, the data assimilation methodology works in the sense that it technically optimizes the emission strength parameterised by parameter vector $\boldsymbol{\beta}$ through minimisation of the 4DVAR cost function. In all experiments, a reduction of the cost function was possible. However, the results are only realistic when the simulated and observed plumes align rather well, as is the case on July 1, 2018, in the experiments. There are two important reasons why the current method cannot always accurately estimate the NO_x emission strength using real data.

- **Differences of plume positions in the spatio-temporal domain**

  Slight differences between plume positions in the spatio-temporal domain, that can be resolved by TROPOMI's high-resolution satellite product, could strongly hamper the results. Because of the chosen parametrisation, the data assimilation method is only able to change the intensity of the

plumes. Whenever modelled and observed plumes differ in location, modelled plumes are most likely compared to TROPOMI pixels with low NO_2 concentrations in the observational part of the cost function. To minimise the cost function, the only solution is to drastically lower the emission parameters. This leads to very low, sometimes negative values of $\hat{\boldsymbol{\beta}}$ estimated in the experiments with TROPOMI retrievals. The use of an increased number of observations by lengthening the assimilation window will not solve the problem. At every time in the assimilation window that plumes dot not match, the method will underestimate $\hat{\boldsymbol{\beta}}$. There is no such mechanism that results in an overestimation of $\hat{\boldsymbol{\beta}}$. Therefore, the current method consistently underestimates the MFs for the emission strength in experiments with TROPOMI retrievals.

- **Incomplete parameterisation of NO_x emission strengths**

  The second source of bias in the estimated parameters $\hat{\boldsymbol{\beta}}$ arises when plumes of different sources overlap. This happened in the first experiment on June 25, 2018, where only the emission strength from a cluster of point sources in the Amsterdam metropolitan area is parameterised. The plume from this cluster partially overlaps with the plumes from Antwerp and Rotterdam. Concentrations in this plumes are higher in the model than in TROPOMI retrievals. The MF for the cluster of point sources around Amsterdam is used to compensate for higher emission strengths from other sources that are not parameterised. Therefore, the emission parameter estimated does not reflect its true value.

The two issues discussed above need to be addressed to be able to apply the data assimilation method using actual TROPOMI retrievals.

The issue of incomplete parameterisation can be solved relatively easy by making sure that the emission strength of relevant sources of NO_x is included in the parameterisation, as done in the second experiment in this chapter.

The problem that arises due to different plume positions in the spatio-temporal domain is harder to solve. One possible solution is to rigorously select the retrievals used in the assimilation more strictly. Restricting the assimilation to retrievals that match the modelled plumes will overcome the issue. However, cherry-picking of observations does not provide a satisfactory solution, giving the impression that the rejected observations are somehow wrong, whereas it is the model that is unable to accurately match the TROPOMI retrievals.

Another solution may be to define more degrees of freedom in the data assimilation. As discussed in Section 4.1, uncertainty is only parameterised for the emission strengths is this study. As illustrated in Figure 4.1, there are other sources of uncertainty in LOTOS-EUROS. Position errors in the spatio-temporal domain may occur because of these uncertainties. The main contributor to this is probably the meteorological input. It may also result from incorrect injection heights, temporal profiles of emissions or the numerical implementation of transport in the model. Assimilating more parameters in LOTOS-EUROS may result in better model fits the TROPOMI retrievals. However, this will complicate the data assimilation. Increasing the number of estimated parameters will increase the computational cost of the data assimilation. Furthermore, the data assimilation scheme developed in this study relies heavily on the development of an approximate model. The influence of other parameters may not be modelled in the same way as in Equation (4.23). If this is the case, the assimilation methodology of this study can no longer be used. Therefore, this solution is impractical.

Another approach would be to somehow match the modelled and observed plumes before comparison. One way to do this is by making use of image morphing techniques. The key in these methods is to first correct the position error of the plumes by distorting the grid. After the distortion of the grid, the model plumes should align with observed plumes. The distortion is followed by data assimilation in which the intensity of the plume is corrected. This method was for example applied in geological parameter estimation problems by Lawniczak 2012. Further research is needed to cope with the position errors of the NO_2 plumes in the data assimilation scheme.

7

Conclusions and recommendations

7.1. Conclusions

Accurate estimates of NO_x emissions are fundamental to atmospheric modelling of air pollutants. High-resolution satellite measurements of NO_2 that are available from the TROPOMI instrument are expected to contribute to model validation and emission estimation. The purpose of this study is to estimate NO_x emission strengths using variational data assimilation techniques. The conclusions of this study are presented by answering the research questions posed in Section 1.6

Research question 1: How can an adjoint-free 4DVAR procedure be designed to assimilate synthetic TROPOMI retrievals in a twin experiment?

- The starting point of this study was the model-reduced variational data assimilation method proposed by Vermeulen and Heemink 2006. The key idea of this method is calculating a reduced-order, linear approximation of a model from which the adjoint can be used to approximate the gradient of the 4DVAR cost function. For the LOTOS-EUROS model, the parameters $\boldsymbol{\beta}$ (multiplication factors for NO_x emission strengths) have a nearly linear effect on the state vector. This linearity was utilized to propose a simpler approximate model compared to Vermeulen and Heemink 2006. The effect of the parameters on the state is calculated from an ensemble of one LOTOS-EUROS simulation with background parameters and N_p simulations of LOTOS-EUROS where the parameters are perturbed sequentially. The approximate model proposed for LOTOS-EUROS calculates the state from the background state plus the effects of perturbed parameters. After substitution of this approximate model in the 4DVAR cost function J, it becomes quadratic and an explicit formulation of its gradient is available. The estimated parameters $\hat{\boldsymbol{\beta}}$ are calculated by solving the linear system that follows from setting the gradient equal to zero. Furthermore, it is possible to derive an approximation of the covariance of $\hat{\boldsymbol{\beta}}$ that indicates the measurability of the parameters.
- Twin experiments demonstrated that the assimilation method, based on the approximate model, optimizes the 4DVAR cost function correctly. The cost function using the approximate model for estimated parameters $\hat{\boldsymbol{\beta}}$ is very close to the cost of the original cost function given true parameters $\hat{\boldsymbol{\beta}}$ ($< 0.5\%$).
- Depending on the assimilation window, the method is able to accurately estimate the parameters in twin experiments. Based on one week of mostly cloud-free retrievals in July 2018, emission strengths for 8 large emission hotspots in the Benelux area and 5 categories of emissions are estimated with errors of at most 10%. The errors are consistent with the covariance of $\hat{\boldsymbol{\beta}}$ that is calculated.
- The computational costs of this method are $N_p + 1$ full runs of the LOTOS-EUROS model for the calculation of the approximate model and negligible costs for optimisation of the 4DVAR cost functions and calculation of the covariance of $\hat{\boldsymbol{\beta}}$. Furthermore, the method proved to be

very efficient during development and testing. Once the ensemble of model runs to calculate the approximate model is generated, experiments can be repeated with few computational costs during development.

Research question 2: How can domain decomposition methods be used to make the procedure scalable in the number of parameters compared to the required number of full order model runs?

- The method developed in this study needs one extra full run of LOTOS-EUROS for every additional parameter estimated. For a large number of parameters, this computational burden will be prohibitive. Through domain decomposition, the efficiency of the method can be increased. A twin experiment showed that by using a static domain decomposition, it is possible to perturb parameters in different subdomains simultaneously and calculate the approximate model on the subdomains. If computational resources are prohibitive, domain decomposition can thus significantly increase the efficiency of the method.
- For some parameters, domain decomposition induced large errors in the approximate model due to two reasons. Plumes of NO_2 that propagate from one subdomain into another are truncated during the calculation of the approximate model on the original subdomain and disturb calculation of the approximate model on the subdomain they enter. The errors in the approximate model resulted in large errors (> 50%) in estimated parameters and should, therefore, be prevented by choosing the domain decomposition such, that plumes do not cross the subdomain boundaries.

Research question 3: Can the procedure assimilate TROPOMI retrievals to improve estimates of anthopogenic NO_x emission strengths?

- The method is technically able to optimize emission parameters using actual TROPOMI retrievals. However, results are only reliable when simulated and measured plumes are well aligned.
- If positional differences between modelled and measured plumes occur due to the high resolution of TROPOMI retrievals, this leads to incorrect estimates of emissions, especially for the emission from emission hotspots. Due to the parameterisation, the assimilation method is only able to correct intensities of modelled plumes and cannot solve positional errors. Although the cost function is improved in all cases, simulated retrievals with optimised parameters often do not match the TROPOMI retrievals. Therefore, the method cannot be used to accurately estimate NO_x emission strengths in cases where positional errors in plumes are present. This issue has to be solved before the method can be used in an operational procedure.
- Furthermore, it is important to make sure that all relevant sources of NO_x are covered by parameters to prevent MFs from overcompensating for emissions that are not covered by the parameterisation.

7.2. Recommendations for further research

Further research is required before the method is to be used for improving emission strengths estimates in inventories by assimilating TROPOMI retrievals.

- It is crucial to handle positional errors of modelled plumes. Experiments showed that misalignment of TROPOMI and modelled plumes leads to significant underestimation of emissions. Grid distortion methods could be a way to correct these positional errors, followed by the correction of the intensity of the plumes by the data assimilation method.
- To increase the computational efficiency of the method, further development of domain decomposition techniques is necessary. In theory, sources that have non-overlapping plumes can be perturbed simultaneously. The subdomains used in the approximate model calculation may be chosen to be time-varying.

- The method's computational efficiency may also be increased by storing concentrations of some other tracers than NO_2 during the run with background parameters and using these in the LOTOS-EUROS runs where the parameters are perturbed. This will significantly decrease the numerical cost of generating the ensemble of model runs needed to calculate the approximate model. Further research should investigate how much computational costs decrease and how much error is introduced by this reduction of simulated compounds.
- Once the method can be used to accurately estimate NO_x emissions, it may be interesting to use these estimations for calculation of anthropogenic CO_2 emissions. As NO_x is a co-emitted species of CO_2, it may be possible to improve estimates of CO_2 emissions by using the ratio of NO_x and CO_2 emissions and the estimated NO_x emissions.

Bibliography

Altaf, M. U., Heemink, A. W., & Verlaan, M. (2009). Inverse shallow-water flow modeling using model reduction. *International Journal for Multiscale Computational Engineering*, *7*(6), 577–594. https://doi.org/10.1615/IntJMultCompEng.v7.i6.70

Bannister, R. N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 607–633. https://doi.org/10.1002/qj.2982

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Aabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., & Seigneur, C. (2015). Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics*, *15*(10), 5325–5358. https://doi.org/10.5194/acp-15-5325-2015

Courant, R., & Hilbert, D. (1953). *Methods of Mathematical Physics*. Wiley & Sons. https://doi.org/10.1002/9783527617210

Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1367–1387. https://doi.org/10.1002/qj.49712051912

Curier, R. L., Kranenburg, R., Jozwicka, M., Timmermans, R., & van der Gon, H. D. (2016). Estimation of anthropogenic CO2 emission from ozone monitoring instrument tropospheric NO2 columns using chemistry transport modelling over north western Europe. *Springer Proceedings in Complexity*, (2), 587–591. https://doi.org/10.1007/978-3-319-24478-5{_}95

De Ruyter de Wildt, M., Eskes, H., Manders, A., Sauter, F., Schaap, M., Swart, D., & van Velthoven, P. (2011). Six-day PM10 air quality forecasts for the Netherlands with the chemistry transport model Lotos-Euros. *Atmospheric Environment*, *45*(31), 5586–5594. https://doi.org/10.1016/j.atmosenv.2011.04.049

Eskes, H. J., & Boersma, K. F. (2003). Averaging kernels for DOAS total-column satellite retrievals. *Atmospheric Chemistry and Physics*, *3*(5), 1285–1291. https://doi.org/10.5194/acp-3-1285-2003

Fischer, P. H., Marra, M., Ameling, C. B., Hoek, G., Beelen, R., De Hoogh, K., Breugelmans, O., Kruize, H., Janssen, N. A., & Houthuijs, D. (2015). Air pollution and mortality in seven million adults: The dutch environmental longitudinal study (DUELS). *Environmental Health Perspectives*, *123*(7), 697–704. https://doi.org/10.1289/ehp.1408254

Hordijk, L., Erisman, J. W., Eskes, H., Hanekamp, J. C., Levelt, P. F., Schaap, M., & De Vries, W. (2020). *Meer meten, robuuster rekenen - Eindrapport Adviescollege Meten en Berekenen Stikstof* (tech. rep.).

Ialongo, I., Virta, H., Eskes, H., Hovila, J., & Douros, J. (2020). Comparison of TROPOMI/Sentinel-5 Precursor NO2 observations with ground-based measurements in Helsinki. *Atmospheric Measurement Techniques*, *13*(1), 205–218. https://doi.org/10.5194/amt-13-205-2020

Ide, B. K., & Ghil, M. (1997). Uni ed Notation for Data Assimilation : Operational , Sequential and Variational P . Courtier A . C . Lorenc. *75*(1), 1–20.

Jin, J., Lin, H. X., Heemink, A., & Segers, A. (2018). Spatially varying parameter estimation for dust emissions using reduced-tangent-linearization 4DVar. *Atmospheric Environment*, *187*(June), 358–373. https://doi.org/10.1016/j.atmosenv.2018.05.060

Kaleta, M. P., Hanea, R. G., Heemink, A. W., & Jansen, J. D. (2011). *Model-reduced gradient-based history matching* (Vol. 15). https://doi.org/10.1007/s10596-010-9203-5

Lawniczak, W. (2012). *Feature-based estimation for applications in geosciences* (Doctoral dissertation). TU Delft.

Lewis, J. M., Lakshmivarahan, S., & Dhall, S. (2006). *Dynamic Data Assimilation : A Least Squares Approach*. Cambridge University Press.

Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., van Geffen, J. H., de Zeeuw, M. B., Denier van der Gon, H. A., Beirle, S., & Krol, M. C. (2019). Quantification of nitrogen oxides emissions from

build-up of pollution over Paris with TROPOMI. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-56428-5

Marécal, V., Peuch, V. H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Van Der Gon, H. A., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., … Ung, A. (2015). A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geoscientific Model Development*, *8*(9), 2777–2813. https://doi.org/10.5194/gmd-8-2777-2015

Meijer, Y. (2019). *Copernicus CO2 Monitoring Mission Requirements Document* (tech. rep.). ESA. https://esamultimedia.esa.int/docs/EarthObservation/CO2M_MRD_v2.0_Issued20190927.pdf

Remkes, Dijkgraaf, Freriks, Gerbrandy, Maij, Nijhof, Post, Rabbinge, & Schol. (2019). *Niet alles kan - Eerste advies Adviescollege Stikstofproblematiek* (tech. rep.). https://doi.org/https://edepot.wur.nl/501323

Reuter, M., Buchwitz, M., Schneising, O., Krautwurst, S., O'Dell, C. W., Richter, A., Bovensmann, H., & Burrows, J. P. (2019). Towards monitoring localized CO2 emissions from space: Co-located regional CO2 and NO2 enhancements observed by the OCO-2 and S5P satellites. *Atmospheric Chemistry and Physics*, *19*(14), 9371–9383. https://doi.org/10.5194/acp-19-9371-2019

Rodgers, C. D. (1976). Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Reviews of Geophysics*, *14*(4), 609–624. https://doi.org/10.1029/RG014i004p00609

Segers, A., Manders, A., & Kranenburg, R. (2019). LOTOS-EUROS User Guide v2.2.000.

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics* (3rd ed.). John Wiley & Sons, Incorporated.

Stavrakou, T., Müller, J. F., Boersma, K. F., Van Der A., R. J., Kurokawa, J., Ohara, T., & Zhang, Q. (2013). Key chemical NOx sink uncertainties and how they influence top-down emissions of nitrogen oxides. *Atmospheric Chemistry and Physics*, *13*(17), 9057–9082. https://doi.org/10.5194/acp-13-9057-2013

Van Geffen, J. H. G. M., Eskes, H. J., Boersma, K. F., Maasakkers, J. D., & Veefkind, J. P. (2019). TROPOMI ATBD of the total and tropospheric NO2 data products. *S5p/TROPOMI*, (1.4.0), 1–76. https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., … Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, *120*(2012), 70–83. https://doi.org/10.1016/j.rse.2011.09.027

Vermeulen, M., & Heemink, A. (2006). Model-Reduced Variational Data Assimilation. *Monthly Weather Review*, (Jazwinski 1970), 2888–2899.

Wang, Y., Wang, J., Xu, X., Henze, D., & Qu, Z. (2019). Inverse modeling of SO2 and NOx emissions over China using multi-sensor satellite data: 1. formulation and sensitivity analysis. *Atmospheric Chemistry and Physics Discussions*, (2), 1–41. https://doi.org/10.5194/acp-2019-879

Xiao, C., Leeuwenburgh, O., Lin, H. X., & Heemink, A. (2019). Non-intrusive subdomain POD-TPWL for reservoir history matching. *Computational Geosciences*, *23*(3), 537–565. https://doi.org/10.1007/s10596-018-9803-z

Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J. D., Sadavarte, P., Lyon, D., Nesser, H., Sulprizio, M. P., Varon, D. J., Zhang, R., Houweling, S., Zavala-araiza, D., Alvarez, R. A., Lorente, A., Hamburg, S. P., Aben, I., & Jacob, D. J. (2020). Quantifying methane emissions from the largest oil-producing basin in the United States from space. *Science Advances*, *2007*(April), 1–10.

A

TROPOMI retrievals July 1-7, 2018

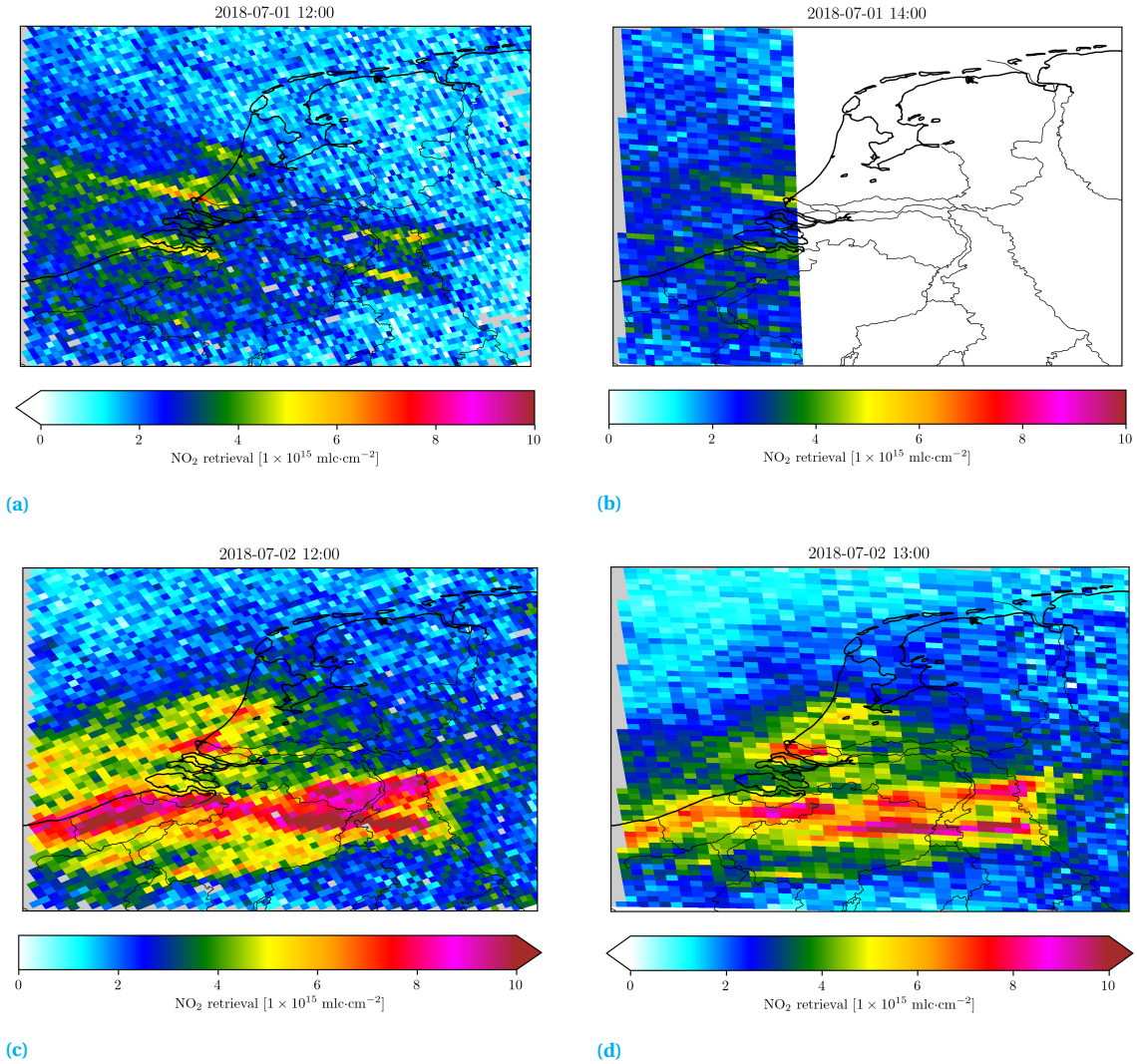
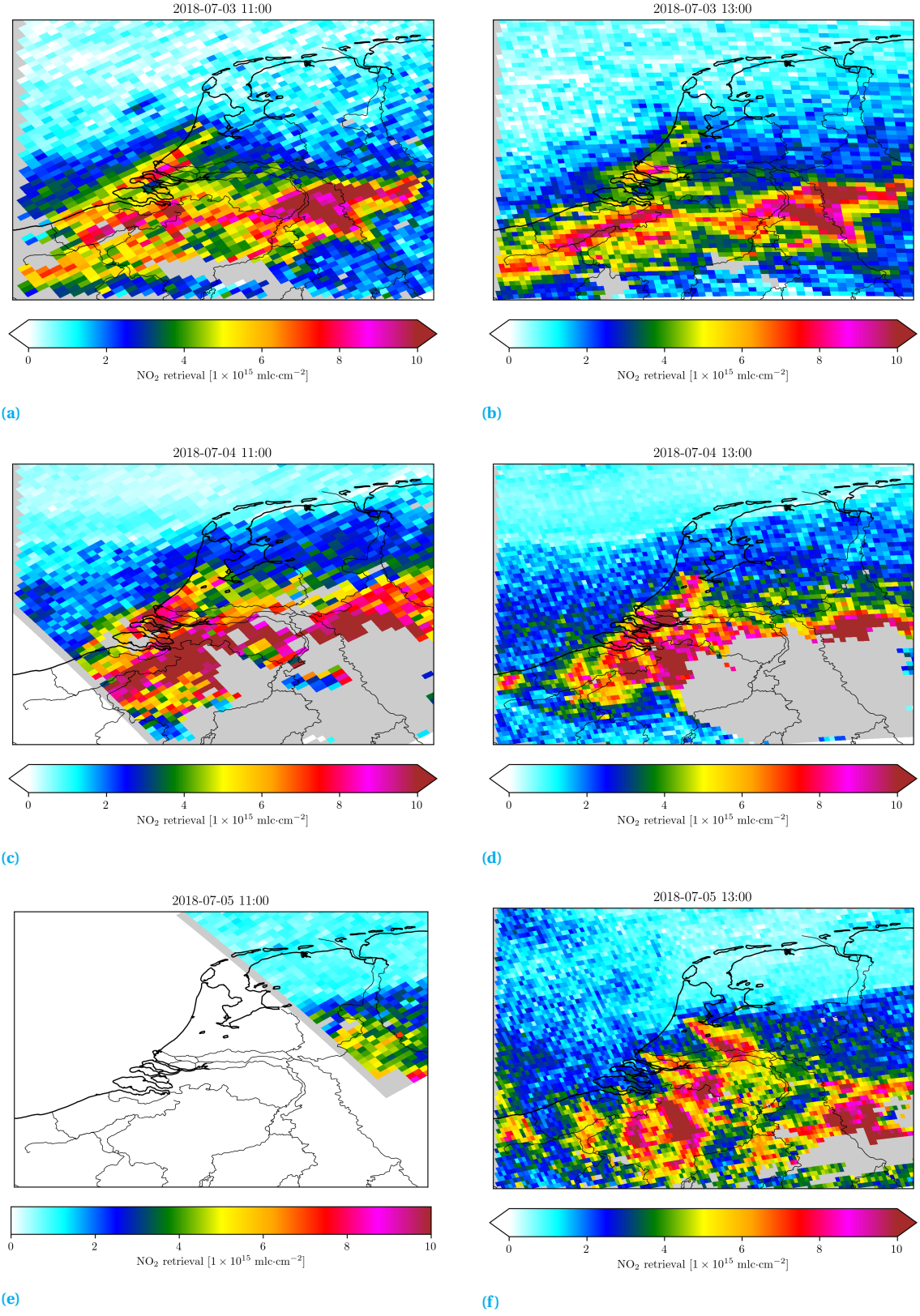
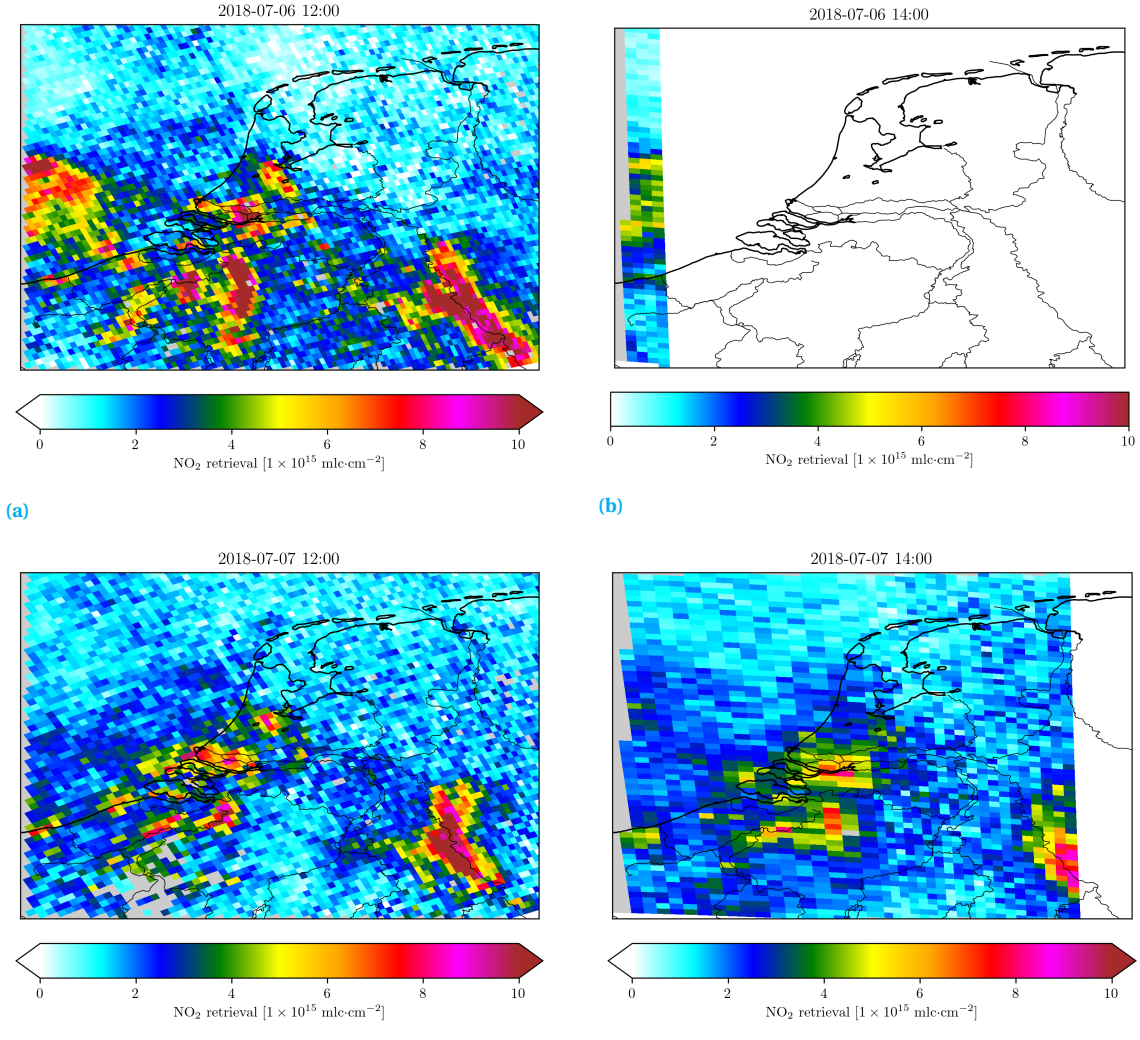


Figure A.1: Tropomi retrievals



(a) (b) (c) (d) (e) (f)

Figure A.2: Tropomi retrievals



(a) (b)

(c) (d)

Figure A.3: Tropomi retrievals