



Estimating ammonia emissions using an adjoint-free 4D-Var approach

Using data assimilation to combine the LOTOS-EUROS model with LML and synthetic IRS satellite observations

Marjolein Leegwater

Estimating ammonia emissions using an adjoint-free 4D-Var approach

Using data assimilation to combine the LOTOS-EUROS
model with LML and synthetic IRS satellite observations

by

Marjolein Leegwater

to obtain the degree of Master of Science

in Applied Mathematics

specialisation Computational Science and Engineering

at the Delft University of Technology,

to be defended publicly on Wednesday July 28, 2021 at 15:00 PM.

Student number: 4549864
Project duration: December 1, 2020 – July 28, 2021
Thesis committee: Prof. dr. ir. A. W. Heemink, TU Delft, responsible professor
Dr. ir. A. J. Segers, TNO, daily supervisor
Dr. N. V. Budko, TU Delft, member of thesis committee

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

Abstract

Ammonia (NH_3) is an important chemical compound in the nitrogen cycle. Ammonia is an essential nutrient and an important part of fertilizer, which in soil leads to increased growth of crops. However, the current excess of NH_3 emissions is a hazard to environmental and human health. While ammonia emissions need to be reduced, the emission estimates are still highly uncertain.

In this study, a method is developed to combine the chemical transport model LOTOS-EUROS with measured ammonia concentrations to improve the ammonia emission estimates in the Netherlands and the surrounding regions. The measurements used are generated by the miniDOAS instruments on LML stations and the IRS instrument on board the future MTG-S satellite.

The proposed method is an adjoint-free 4D-Var method. The 4D-Var method aims to retrieve the emission parameters for which the LOTOS-EUROS model determines NH_3 concentrations that resemble the measurements while keeping the emissions fairly similar to the original inventories. A linear approximate model has been developed, which uses the near-linearity of the NH_3 concentrations in terms of the NH_3 emissions. When using the approximate model, the 4D-Var method can be solved without using an adjoint model, making the method adjoint-free. Subsequently, the 4D-Var cost function is rewritten to allow the emission parameters to have a log-normal prior distribution. A maximum likelihood approach is developed to estimate the parameters of both the prior distribution of the emissions and the likelihood of the measured observations. Last, a preconditioner in reduced space has been considered, to estimate emissions on a fine spatial resolution, while keeping the computational cost feasible. This preconditioner uses the property that the emission parameters are correlated in space.

First, the methodology has been tested in an identical twin experiment where the emissions vary only in time and strength, using the LML observations. It was concluded that the methodology worked well for short periods (less than 30 days), but the results were dominated by observational noise. When using the real observations in the 4D-Var method, the results seem unrealistic

Second, the method has been tested in an identical twin experiment where emissions vary in space, as well as in time and strength. Here, synthetic IRS observations of the future MTG-S satellite are used. The optimized emissions did resemble the true emissions of the twin experiment. Observational noise appeared to no longer be an issue. However, the results were not perfect. The regions with the highest emission increase appeared to be underestimated, low emission areas appeared to have large relative estimation errors, and estimates for coastal regions seemed to be incorrect. Hence, on a local scale, the emission estimates can be imperfect, but overall the adjoint-free 4D-Var method does greatly improve the emission inventories.

Preface

This study has been executed as a master thesis for the master Applied Mathematics at the Delft University of Technology. During my studies, I followed the specialization Computational Science and Engineering. For this project, I worked as an intern at TNO (Netherlands Organisation for Applied Scientific Research) at the department Climate, Air, and Sustainability. I worked at TNO from September 2020 until July 2021. For the first three months, I worked on a different project, where I developed the chemically reduced model for ammonia which is used in this thesis. As of December, I have been working on this data assimilation study.

I would like to start by thanking my daily supervisor at TNO: Arjo Segers. Thank you for the help and guidance, the enthusiasm, and the confidence during my year at TNO. Second, I would like to thank Arnold Heemink for helping me find this thesis project and for all the feedback on my work. Third, I want to thank Enrico Dammers from TNO for his large contribution to the project and his enthusiasm during all the meetings. Last, I would like to mention my appreciation for the CAS department. I encountered a very friendly, helpful, and inspiring group of people that would always be welcoming during meetings, make time to answer my questions, and make me feel appreciated for the work I was doing.

*Marjolein Leegwater
Delft, July 2021*

Contents

1	Introduction	3
1.1	Estimating ammonia concentration	4
1.1.1	Chemical transport model	4
1.1.2	Measuring instruments	4
1.1.3	Data Assimilation	4
1.2	Uncertainty in the application of manure and fertilizer	5
1.3	Aim of the study	5
1.4	Research questions	5
1.5	Overview of the report	6
2	LOTOS-EUROS Model	7
2.1	General characteristics of the LOTOS-EUROS model	7
2.2	Discretization of the continuity equation	7
2.3	Observation operator	8
2.4	Simulated processes	8
2.4.1	Emissions	8
2.4.2	Chemistry	9
2.4.3	Dry deposition	9
2.4.4	Wet deposition	10
2.5	Chemically reduced model	10
3	Measurement instruments	13
3.1	MiniDOAS instrument of LML network	13
3.1.1	Description	13
3.1.2	Data selection	14
3.1.3	Uncertainty of LML observations	14
3.1.4	LML observation simulated by LOTOS-EUROS model	14
3.2	IRS instrument on board the MTG-S satellite	14
3.2.1	Description	14
3.2.2	Uncertainty of IRS observations	15
3.2.3	IRS observations simulated by LOTOS-EUROS model	15
4	Data Assimilation: adjoint-free 4D-Var method	17
4.1	Problem description	17
4.2	Data assimilation methods	18
4.3	Bayesian approach to 4D-Var method	19
4.4	Adjoint-free 4D-Var method for Gaussian distributions	20
4.4.1	Definition of the cost function and gradient	20
4.4.2	Approximate linear state to make the 4D-Var method adjoint free	21
4.4.3	Finding the optimal parameter vector	23
4.4.4	Covariance of the optimal solution	23
4.5	Adjoint-free 4D-Var method for log-normal prior distributions	23
4.5.1	Log-normal multiplication factors	24
4.5.2	Approximate linear state, cost function, and gradient	24
4.5.3	Finding the optimal parameter vector	25
4.6	Adjoint-free 4D-Var method when using a preconditioned state	25
4.6.1	Preconditioned state: Gaussian prior distribution	26
4.6.2	Preconditioned state: log-normal prior distribution	28
4.6.3	Reduce size: Eigenvalue decomposition	29

4.7	Maximum likelihood method	31
4.7.1	Objective of maximum likelihood method	31
4.7.2	Determine the distribution of observations, given the parameters	32
4.7.3	Minimization of the Log-Likelihood function	32
4.8	Reducing computational cost of the ensembles	34
4.9	Summary	34
5	Results	37
5.1	Concept of twin experiment	37
5.1.1	Twin experiment procedure	38
5.2	Experiment 1: LML observations Zegveld.	38
5.2.1	Problem description	38
5.2.2	Parameterization	39
5.2.3	Results	40
5.2.3.1	Results twin experiment	41
5.2.3.2	Results real LML data	42
5.2.3.3	Distribution of multiplication factors	43
5.2.4	Conclusion	44
5.3	Experiment 2: LML observations Netherlands	45
5.3.1	Problem description	45
5.3.2	Parameterization	46
5.3.3	Results	47
5.3.3.1	Results twin experiment	47
5.3.3.2	Results real LML data	48
5.3.3.3	Distribution of multiplication factors	49
5.3.4	Conclusion	49
5.4	Experiment 3: IRS observations	50
5.4.1	Problem description	50
5.4.2	Parameterization	50
5.4.3	Results	52
5.4.4	Conclusion	53
6	Conclusion	57
7	Recommendations for further research	59

List of Symbols

$\hat{\theta}$	Optimal approximation of some parameter vector of θ
$\hat{\mathbf{x}}$	Optimal approximation of state vector of \mathbf{x}
β	Parameter vector with a multivariate Gaussian distribution
γ	Parameter vector with a multivariate log-normal distribution
μ_θ	First parameter of the distribution of parameter vector θ (Mean for $\theta \sim \mathcal{N}(\mu_\theta, \mathbf{B}_\theta)$)
θ^b	Background parameter vector of some parameter vector θ
θ^{true}	True values of some parameter vector θ
\mathcal{H}	General observation operator
\mathbf{B}_θ	Second parameter of the distribution of parameter vector θ (Covariance for $\theta \sim \mathcal{N}(\mu_\theta, \mathbf{B}_\theta)$)
\mathbf{C}_θ	Correlation matrix of parameter vector θ
\mathbf{E}_θ	Jacobian of state \mathbf{x} with respect to some parameter vector θ
\mathbf{H}	Linear observation operator
\mathbf{R}	Error covariance
\mathbf{R}_{obs}	Observation error covariance
\mathbf{R}_{repr}	Representation error covariance
\mathbf{S}_θ	Diagonal matrix formed from the standard deviation field of θ
\mathbf{w}	Preconditioned parameter vector
\mathbf{x}	State vector
\mathbf{x}^b	Background state vector
\mathbf{x}^{true}	True state vector
\mathbf{y}	Vector with observations
J	4DVAR cost function
N_{MF}	Number of multiplication factors to be determined (size of the parameter vector)
N_O	Number of time instances for which observations are performed

1

Introduction

Ammonia (NH_3) is a chemical compound present in the atmosphere and is an important tracer in the nitrogen cycle. Ammonia is an essential nutrient and an important part of fertilizer, which in soil leads to increased growth of crops. The most important origin of ammonia emissions is agriculture in the form of animal manure and fertilizer (TNO 2019). An excess of ammonia can have negative impacts, both on biodiversity and human health. The most evident downside is that when there is too much ammonia in soil, certain species, such as grasses and nettles, grow so well, that they suppress other plants that would flourish when there are fewer nutrients in the soil (Compendium voor de Leefomgeving 2018). When the suppressed plants disappear in a certain region, animals that depend on the plant also die. In this way, a surplus of ammonia in soil leads to a reduction in biodiversity. Another risk of ammonia is that it can form particulate matter (PM) by chemical reactions in the atmosphere (Baek, Aneja, and Tong 2004). Inhalation of PM can be harmful to human health, especially for people with respiratory and cardiovascular diseases (Gezondheidsraad 2018). Furthermore, ammonia can cause eutrophication of water bodies, acidification, and it can form N_2O in soil, which is an important greenhouse gas and therefore contributes to global warming.

Although ammonia emissions in the Netherlands have been decreasing since 1990 due to national and European policies (CBS 2020), still too much is emitted into the atmosphere. This causes a surplus of nitrogen deposition in for example "Natura 2000" areas, a European network of protected nature areas. In the Netherlands, there are 160 Natura 2000 areas, of which 130 are sensitive to nitrogen (Figure 1.1a). A critical deposition value has been determined for each nature reserve. If this threshold is exceeded, there is a risk that the habitat will be affected significantly. The advisory board on nitrogen problems (Dutch: Adviescollege stikstofproblematiek), led by J. W. Remkes, concluded that as of 2020, the critical deposition value was exceeded in 118 out of the 130 nitrogen sensitive Natura 2000 areas (Remkes, Dijk, et al. 2020).

In 2015, the policy framework "Programma Aanpak Stikstof" (PAS) (*Programma Aanpak Stikstof* n.d.) was introduced by the Ministry of Economic Affairs and the Ministry of Infrastructure and the Environment. The aim of the program was to reduce nitrogen emissions, including ammonia emissions, and to reduce the already present excess of nitrogen in nature reserves. The PAS was also used to grant permits for new activities that would cause additional nitrogen deposition in certain nature reserves. However, in 2019 it was ruled that the PAS could not be used to grant those permits. This led to the so-called nitrogen crisis in the Netherlands (Dutch: stikstofcrisis). The issuing of permits had to stop, which caused many construction and infrastructure projects to be halted, to stop nitrogen emissions. Drastic measures were taken to cut nitrogen emissions as proposed by the advisory board (Remkes, Dijkgraaf, et al. 2019). These measures led to large protests of farmers, dominating the Dutch news.

To move out of the nitrogen crisis, a new approach to reduce nitrogen emissions is needed. At the moment, the advisory committee states that a reduction of 50% of ammonia emissions is needed by 2030, to bring the ammonia deposition in the Natura 2000-areas below the critical value (Remkes, Dijk, et al. 2020). A key element to reduce the ammonia emissions efficiently is getting a better understanding

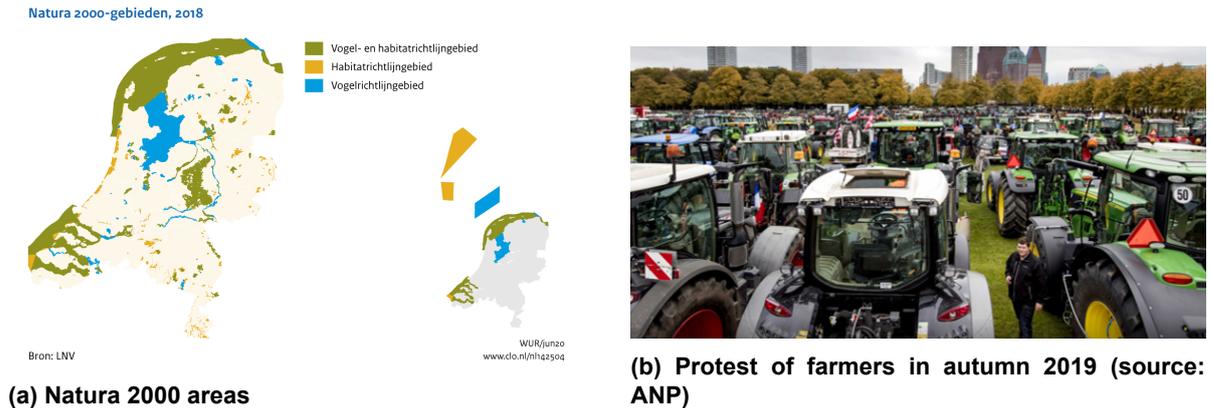


Figure 1.1: Nitrogen crisis in The Netherlands.

of what is causing the emissions and which emissions sources lead to nitrogen deposition in Natura 2000 areas.

1.1. Estimating ammonia concentration

To estimate atmospheric ammonia concentrations, simulations by chemical transport models and observations by measuring instruments are used. Both chemical transport models and measuring instruments have their disadvantages. Measuring equipment is expensive, the number of measurements is usually limited, and measurements can be biased and be influenced by local sources. Similarly, chemical transport models are expensive to develop and run, and contain many errors as well.

1.1.1. Chemical transport model

In this study, the chemical transport model LOTOS-EUROS is used. The LOTOS-EUROS model is developed by a consortium of institutes, including TNO (Netherlands Organization for Applied Scientific Research), RIVM (National Institute for Public Health and the Environment), and KNMI (Royal Netherlands Meteorological Institute). The model calculates the concentrations and depositions of several chemical tracers, given data on emissions and meteorological conditions. The LOTOS-EUROS model can be used for several applications, such as air quality forecasting, emission modeling, and deposition modeling (*LOTOS-EUROS applications website 2020*).

1.1.2. Measuring instruments

There are two types of measuring instruments considered in this report. The first type is the miniDoas, located at stations of the LML (Landelijk Meetnet Luchtkwaliteit) network of the RIVM. The miniDoas measures hourly surface concentrations of NH_3 at 5 locations in the Netherlands (in 2017). The second type of measuring instrument is the future InfraRed Sounder (IRS) on the MTG-S satellite. The MTG-S satellite is planned to be launched in 2023. Hence, the IRS is not operational yet. Once launched, the IRS instrument will measure vertical column densities of NH_3 on a $4\text{km} \times 4\text{km}$ resolution.

1.1.3. Data Assimilation

In this study, a data assimilation method is developed that combines the LOTOS-EUROS model simulations with observations from the LML system and the future IRS instrument. Data assimilation is a mathematical field where data of theoretical (numerical) mathematical models are combined with observational data. Objectives of data assimilation methods can be to determine the optimal state of a system, estimate initial conditions, or determine model parameters. When input parameters of the numerical model are estimated, such as initial conditions or model parameters, the problem is called an inverse problem.

There are two types of data assimilation methods: variational methods, such as the 4D-Var method, and stochastic methods, such as the Kalman filter. In this study, the choice has been made to use a 4D-Var method to estimate emission parameters. The advantage of the 4D-Var method for this application is that it uses observations over a longer time range, obeys the conservation of mass, and has can deal with a large number of observations, while barely increasing computational cost. These differences between variational and stochastic models will be described more elaborately in section 4.2.

1.2. Uncertainty in the application of manure and fertilizer

The ammonia concentrations, determined by the LOTOS-EUROS model, differ from the measurements. One of the causes of the difference is that the LOTOS-EUROS model requires estimates of ammonia emissions, which are highly uncertain. The emission data used in the LOTOS-EUROS model approximate the true emissions, but the data set is not perfect. One of the major uncertainties in the used emission inventories is the application of animal manure and fertilizer to soil in the spring period. The true timing of the application is not exactly known but only estimated using the approach of (Skj  th et al. 2011), adapted by (Hendriks et al. 2016). The estimated emissions due to the application of fertilizer depend on local production methods, crop types, and meteorological conditions. However, these modeled emissions are based on many assumptions and are expected to differ from the actual ammonia emission. This study focuses on two of the uncertainties in the emission inventory: the moment in time when manure or fertilizer is applied, and the amount of ammonia emitted during application.

1.3. Aim of the study

The aim of this study is to develop a method to combine the chemical transport model LOTOS-EUROS with measured ammonia concentrations to improve the ammonia emission estimates in the Netherlands and the surrounding regions. A method is developed to improve the time profiles and strength of the emission estimates for different parts within the considered region. As the to be estimated parameters are input variables of the LOTOS-EUROS model, the problem to solve is an inverse modeling problem.

The concentrations determined by the LOTOS-EUROS model and the measured ammonia concentrations are combined using the data assimilation method 4D-Var, resulting in the optimal emission parameters. The 4D-Var method aims to retrieve the emission parameters for which the LOTOS-EUROS model determines NH_3 concentrations that resemble the measurements, while keeping the emissions fairly similar to the original inventories. To do so, uncertainties of both the measurements and the emission inventories are taken into account. An important part of this study is to quantify the uncertainty of the emission inventory and the measured concentrations, using a maximum likelihood approach.

The common procedure to solve the 4D-Var problem is to minimize a cost function by using an iterative gradient-based optimization method, where an adjoint model is used to determine the gradient. However, for the extensive LOTOS-EUROS model, no adjoint model is available. Developing one would be cumbersome as it would take a lot of programming effort to make it and to maintain it. Instead, an efficient adjoint-free gradient-based 4D-Var method is developed for which the results resemble the results of the original 4D-Var method.

This adjoint-free 4D-Var method uses an ensemble of runs of the LOTOS-EUROS model. One of the limiting factors in this study is that running the LOTOS-EUROS model is very time-consuming. Running the model for a month takes approximately 12-18 hours. The number of times for which the LOTOS-EUROS model is executed should be limited in order to obtain results within a reasonable time frame. Many design choices are made such that the number of model runs remains small.

1.4. Research questions

This study focuses on estimating NH_3 emissions using the data assimilation method 4D-Var. The aim is to design and test the methodology. The estimated NH_3 emission estimates are not quantitatively interpreted. This study aims to answer the following research questions.

Research question 1: Could an efficient adjoint-free 4D-Var method be designed to estimate time-varying ammonia emissions in the Netherlands using the observations of the LML stations?

The first objective of this research is to design a mathematical method that estimates daily NH_3 emissions, using the observations of the 5 LML stations in the Netherlands. As running the LOTOS-EUROS model is time-consuming, certain simplifications have been made, to keep the computational cost of the methodology feasible. Also, adjustments have been made to the commonly used 4D-Var cost function, to allow parameters to have a log-normal distribution, instead of a Gaussian distribution. The developed method is tested in twin-experiments, to assess the performance of the methodology. Additionally, the method is used to generate estimates for the NH_3 emissions, using the real observations of the LML systems.

Research question 2: Could an efficient adjoint-free 4D-Var method be designed to estimate time-varying ammonia emissions in the Netherlands on a fine spatial resolution, using synthetic observations of the future IRS instrument on the MTG-S satellite.

The second objective of this research is to improve the design by incorporating spatial variability for the daily NH_3 emission estimates. This procedure should be able to estimate the NH_3 emissions on a relatively fine grid. For this procedure, an increased number of observations from the IRS instrument is used. As the number of emission parameters increases due to the spatial variability, as well as the number of observations, this design focuses on limiting the computational complexity and the amount of computer memory needed. As the IRS instrument is not operational yet, the designed method is tested in a twin-experiment only.

1.5. Overview of the report

This thesis starts with introducing the chemical transport model LOTOS-EUROS in chapter 2. Second, the two measuring instruments, the miniDOAS on the LML stations, and the IRS instrument onboard the MTG-S satellite will be discussed in chapter 3. The main part of this thesis will consist of chapter 4, the development of the adjoint-free 4D-Var data assimilation method. The methodology will be tested in three experiments that are discussed in chapter 5. This thesis will end with discussing the main conclusions in chapter 6, followed by recommendations for further research in chapter 7.

2

LOTOS-EUROS Model

In this chapter, the LOTOS-EUROS model will be introduced. Initially, the general model characteristics will be stated and mathematically formulated in a partial differential equation. Second, the discretization of the equation will be discussed. Also, the concept of an observation operator will be introduced. After that, a more detailed description of several processes will be given. A more elaborate explanation of certain aspects of the LOTOS-EUROS model can be found in (Manders et al. 2017), and in the *Reference Guide of the LOTOS-EUROS model* (Manders-Groot, A.J. Segers, and Jonkers 2019). Last, a chemically reduced version of the LOTOS-EUROS model will be discussed. This chemically reduced model has lower computational complexity than the full LOTOS-EUROS model.

2.1. General characteristics of the LOTOS-EUROS model

The LOTOS-EUROS model is a chemical transport model, used to calculate concentrations and depositions of trace gasses and aerosols in the lower layers of the atmosphere (Manders et al. 2017; Manders-Groot, A.J. Segers, and Jonkers 2019). The main input to the model are estimates of emissions and meteorological data. The processes included in the model are, among others, transport, emissions, chemistry, dry deposition (uptake by vegetation), wet deposition (wash-out by precipitation), and sedimentation. There are two types of transport considered in the model: advection (due to wind) and vertical diffusion. Concentrations in the model are expressed as volume mixing ratios in ppb for trace gasses, and mass concentrations in $\mu\text{g}/\text{m}^3$ for aerosols. The concentrations of tracers, denoted as C , depend on these processes according to the continuity equation:

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

Here, U , V , and W represent the large-scale wind components in the west-east, south-north, and vertical direction respectively; K_z is the vertical turbulent diffusion coefficient, R denotes the source or sink due to chemistry, Q the source due to emissions, and D and W denote the sink due to dry and wet deposition. The transport due to wind is represented by the term:

$$U \frac{\partial C}{\partial x} + V \frac{\partial C}{\partial y} + W \frac{\partial C}{\partial z},$$

and the transport due to vertical diffusion is represented by the term:

$$\frac{\partial}{\partial z} \left(K_z \frac{\partial C}{\partial z} \right).$$

2.2. Discretization of the continuity equation

To solve equation (2.1) numerically, the equation is discretized in time and space. For every grid cell and every tracer in the model, the concentration is determined for every time step. Let $\mathbf{x}(t_k)$ denote the

vector that contains the concentration of every tracer for every grid cell at time t_k . Then the discretized model can be denoted as:

$$\mathbf{x}(t_{k+1}) = \mathbf{M}_{k+1}(\mathbf{x}(t_k), \theta), \quad k = 0, 1, 2, \dots \quad (2.2)$$

Here, \mathbf{M} is a non-linear operator that performs a discrete time step by solving equation (2.1), and θ is a parameter vector. In this study, the parameter vector θ will be related to the emission parameters. The general parameter vector is denoted by θ , but if the context of a specific problem is defined, the parameter vector might also be denoted as β , γ , or \mathbf{w} .

The simulation time step (the time between t_k and t_{k+1}) is determined by the model itself and varies during the model run. The time steps are limited by the CFL-criterion which states that a parcel of air should not cross a complete grid cell within one time step. The output is saved at a frequency configured by the user. In this study, the output frequency is hourly.

2.3. Observation operator

The LOTOS-EUROS model estimates the concentrations for each chemical tracer for all time steps and all grid cells. Measurements, on the other hand, could be concentrations at a specific location (for example the LML-stations), or vertical column densities (for the IRS instrument). To compare the model and the measurements, a function should be made that maps the concentrations per grid cell in the LOTOS-EUROS model to something resembling the true measurements. This mapping operator \mathcal{H}_k takes as input the concentrations of the model at a certain time t_k and generates a simulated observation \mathbf{y}_k^{sim} :

$$\mathbf{y}_k^{sim} = \mathcal{H}_k(\mathbf{x}(t_k), \theta), \quad k = 0, 1, 2, \dots \quad (2.3)$$

In chapter 3, the operator \mathcal{H}_k will be described for both the LML-stations and the IRS instrument. The simulated observations \mathbf{y}_k^{sim} are expected to differ from the actual observations (denoted as \mathbf{y}^o) due to errors in the model, measurement noise of the instrument, and errors in the mapping \mathcal{H}_k . If this error is denoted as ϵ_k , the simulated observations can be linked to the actual observations as:

$$\mathbf{y}_k^o = \mathbf{y}_k^{sim} + \epsilon_k = \mathcal{H}(\mathbf{x}(t_k), \theta) + \epsilon_k. \quad (2.4)$$

2.4. Simulated processes

In this section, a more detailed description of several processes of the LOTOS-EUROS model will be given.

2.4.1. Emissions

There are several sources of ammonia emission. In the Netherlands, approximately 88% of the ammonia emissions are released by agriculture (TNO 2019). The most important agricultural emissions are animal manure and fertilizer. The other ammonia emissions are mainly released by households, industry, and road traffic. In other parts of the world, the main emission sources can be different. Some additional important ammonia emission sources are biomass burning (wildfires, fires as part of deforestation, biomass burning) (Andreae and Merlet 2001), and local animal colonies (for example seals (Riddick et al. 2012)).

In the LOTOS-EUROS model, there are several sources of emissions considered. By default, the following categories are used in the LOTOS-EUROS model: Anthropogenic sources (related to human activity), biogenic sources (mainly trees and bacteria in soil), sea-spray sources (source of sea salt), dust sources, and forest fires (Manders-Groot, A.J. Segers, and Jonkers 2019). For simulation of ammonia concentrations in the Netherlands, only the anthropogenic sources are relevant. Therefore, the LOTOS-EUROS model with only anthropogenic emission sources is considered in this report.

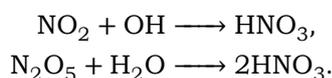
The anthropogenic emission inventories have been compiled by TNO as part of the Copernicus Atmosphere Monitoring Service (CAMS). In this study, version 2.2 is used (Kuenen et al. 2014) to model most emissions. This database, also called an emission inventory, contains emission data of several

sectors such as industry, transport, and solvents. For the sector agriculture, a different emission inventory version, version 4.1, is used. This inventory also contains the modeled NH_3 emissions with an improved timing for fertilizer application based on the approach of (Skjøth et al. 2011), adapted by (Hendriks et al. 2016).

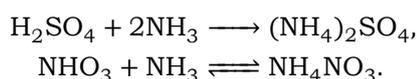
2.4.2. Chemistry

Chemical conversion implies that a tracer in the air reacts with another tracer. This can be both a source and a sink for tracers. In the LOTOS-EUROS model, there are three types of chemistry: Gas-phase chemistry, cloud chemistry, and aerosol chemistry. As cloud chemistry barely influences NH_3 concentrations, only the gas-phase chemistry and aerosol chemistry will be discussed in this report.

The gas-phase chemistry is modeled in the LOTOS-EUROS model using a modified version of the Carbon Bond-IV Mechanism (CBM-IV) (Manders-Groot, A.J. Segers, and Jonkers 2019). The scheme includes 33 species and 104 reactions, including 14 photolytic reactions (Whitten, Hogo, and Killus 1980). These reactions can be found in Appendix B of the LOTOS-EUROS reference guide (Manders-Groot, A.J. Segers, and Jonkers 2019). NH_3 itself is not one of the species in CBM-IV. However, some of the species in CBM-IV have an impact on NH_3 through the aerosol chemistry described below. An important process with regard to NH_3 in the gas-phase chemistry is the production of HNO_3 out of NO_2 or N_2O_5 . HNO_3 plays an important role in the aerosol chemistry. Important chemical reactions with respect to the production of HNO_3 of the gas-phase chemistry are:



Another type of chemistry is aerosol chemistry, or actually the secondary inorganic gas-aerosol-phase equilibrium. In the LOTOS-EUROS model, this is implemented using the ISORROPIA-II module (Manders-Groot, A.J. Segers, and Jonkers 2019). This module describes the equilibrium between the concentrations of the gasses HNO_3 , H_2SO_4 , and NH_3 , and the aerosols NH_4NO_3 and $(\text{NH}_4)_2\text{SO}_4$ in presence of aerosol water (Fountoukis and Nenes 2007). Ammonia (NH_3) reacts with sulfuric acid (H_2SO_4), forming particulate ammonium sulfate ($(\text{NH}_4)_2\text{SO}_4$) (Baek, Aneja, and Tong 2004). This reaction is dominant and irreversible. Furthermore, there is an important equilibrium reaction, where ammonia (NH_3) reacts with nitric acid (HNO_3), forming particulate ammonium nitrate (NH_4NO_3). This equilibrium depends on the temperature and the relative humidity, and the initial concentrations of the three tracers. The chemical reactions of the aerosol chemistry can be written as:



2.4.3. Dry deposition

Tracers can be removed from the air by dry deposition. Dry deposition represents the uptake of tracers from the atmosphere by the surface of the earth: vegetation, soil, and water surfaces can all absorb trace gases and aerosols. The largest part of the nitrogen deposition in populated areas with intensive animal housing, such as the Netherlands, is caused by the dry deposition of ammonia (TNO 2019; Pitcairn et al. 1998; Walker et al. 2008).

For most tracers, the exchange is modeled to be one-directional, e.g. from atmosphere to vegetation (Wichink Kruit et al. 2012). The extent to which a tracer is removed from the atmosphere due to dry deposition depends on vegetation and surface properties, as well as the meteorological conditions.

However, for NH_3 bi-directional exchange is possible. Hence, NH_3 can be both emitted and absorbed by plants, soil, and water surfaces (Johnson et al. 2008; Sutton, Schjorring, and Wyers 1995). Whether ammonia is emitted or deposited depends on the compensation point. This value is defined as the NH_3 concentration in the atmosphere at which no net NH_3 exchange takes place between the surface and the air close to the surface. If the ammonia concentration close to the surface is higher than the compensation point, deposition will occur, and if the ammonia concentration is low, the surface will emit NH_3 . This compensation point depends on many factors. One of those factors is the average SO_2

concentration close to the surface that was present during the previous days. Additionally, the dry deposition depends on the ratio of the currently present SO_2 and NH_3 . Due to these two influences, the ammonia dry deposition depends on the SO_2 concentration.

2.4.4. Wet deposition

Tracers could also be removed from the atmosphere by wet deposition. Wet deposition consists of two processes. First, species can be absorbed by water in the clouds, which is called in-cloud scavenging. Alternatively, rain droplets can absorb and collide with species, whilst falling. This process is called below-cloud scavenging. An important property of ammonia is that it is highly soluble. Therefore, ammonia is removed effectively by wet deposition.

2.5. Chemically reduced model

In this study, the LOTOS-EUROS model is run several times with different ammonia emissions. Such an ensemble of model runs with varied emission inventories is used to determine how the observed concentrations of ammonia would change if the emissions would be different. Unfortunately, running such an ensemble is very time-consuming. It would be desirable if the LOTOS-EUROS model would need less computational time.

Of all the tracers and aerosols in the LOTOS-EUROS model, only the concentrations of ammonia are relevant for this study. To save computational time, a reduced version of the LOTOS-EUROS model could be used, that focuses on modeling NH_3 concentrations only. Such a reduced version of the LOTOS-EUROS model has been made in a previous study (Leegwater 2020). It has been determined which processes in the LOTOS-EUROS model are relevant for determining the ammonia concentrations, and which processes are negligible. A chemically reduced model is made, which is about three times faster than the full LOTOS-EUROS model, while still obtaining accurate results for the ammonia concentrations.

The reduced model is made to run an ensemble with slightly different ammonia emissions. Before running this ensemble of the reduced model, the full LOTOS-EUROS model is executed once, to save certain features and results. These features and results are subsequently used as input values for the reduced model.

In the reduced model, the concentrations of only four tracers are determined, whereas the full LOTOS-EUROS model considers over 50 tracers (gas phase plus aerosols). The considered species are NH_3 , HNO_3 , NH_4^+ , and NO_3^- . The tracers HNO_3 , NH_4^+ , and NO_3^- are included since changes in NH_3 cause changes in the concentrations of HNO_3 , NH_4^+ , and NO_3^- and vice versa. The reduced model determines the processes of transport, emissions, and deposition for only these four tracers. The processes have been adapted for the reduced model:

- Chemistry needs to be modeled in the reduced model. When modeling the chemistry, many additional tracers become involved. In the aerosol chemistry, the species SO_4^{2+} is relevant. However, the SO_4^{2+} concentration is independent of the NH_3 concentrations. Therefore, the SO_4^{2+} concentrations do not need to be determined by the reduced model but can be imported from the background run of the full LOTOS-EUROS model.
- Also, gas-phase chemistry is considered. NH_3 is not a tracer in the gas-phase chemistry, but one of the other modeled species, HNO_3 , is highly influenced by the gas-phase chemistry. The gas-phase chemistry considers many tracers, and the process is computationally expensive. To save time, the production of HNO_3 in the gas-phase chemistry is saved in the full model run and imported in the reduced model version, where it is added to the HNO_3 concentration.
- Last, the concentrations of SO_2 and the average concentration of SO_2 at the surface of the previous month (denoted as SO_2ave) are saved in the full model and imported in the reduced model. These values are used to model the influence of SO_2 on the dry deposition of NH_3 , described in section 2.4.3.

This reduced version with four tracers, all containing nitrogen, is also called the "N4-model". A visualization of the "N4-model" is shown in Figure 2.1. It is also visualized that the reduced model can be run

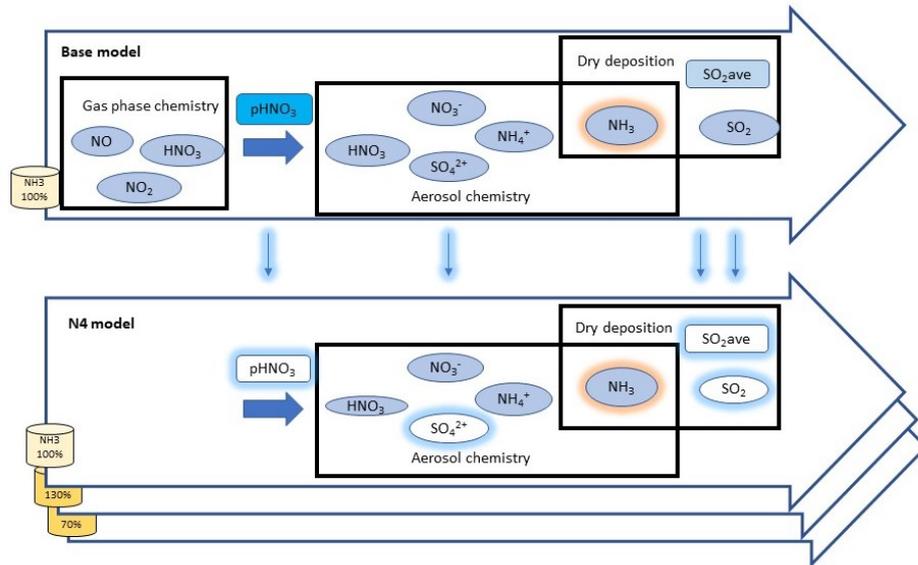


Figure 2.1: Visualization of the chemically reduced version (also called "N4-model") of the LOTOS-EUROS model

several times with different emissions, each time importing the same data from the full LOTOS-EUROS model.

3

Measurement instruments

In this chapter, the two measurement instruments considered in this study will be described. First, the MiniDOAS instrument of the LML network will be described in section 3.1. After that, in section 3.2, the IRS instrument hosted on board the future MTG-S satellite will be discussed. The sections will focus on the properties of the systems, the uncertainties in the observations, and the observation operator used to simulate observations with the LOTOS-EUROS model.

3.1. MiniDOAS instrument of LML network

3.1.1. Description

The first measurement instrument considered, is the *mini Differential Optical Absorption Spectroscopy* (miniDOAS) instrument. The miniDOAS instrument is built by the National Institute for Public Health and the Environment (RIVM). As of 2016, there are six miniDoas instruments in the Netherlands that measure hourly NH_3 concentrations. Five out of six instruments were operational in the spring period of 2017, the year used for experiments in this study. The instruments are part of the Dutch National Air Quality Network (LML, Dutch: Landelijk Meetnet Luchtkwaliteit). Within the LML network, concentrations of NH_3 are measured, as well as the concentrations and wet deposition of other pollutants. A picture of a miniDOAS on an LML station is shown in Figure 3.1a.



(a) MiniDOAS (source: RIVM)



(b) Artist's impression of MTG-S satellite of SENTINEL-4 mission (source: ESA)

Figure 3.1: Measurement instruments for NH_3 .

The miniDOAS instrument aims a light source at a retroreflector. The retroreflector directs the light back to a receiver. Subsequently, the intensity of this reflected light is measured. As ammonia absorbs light of a certain wavelength, the reflected light can be compared to a reference light spectrum, to see how much light is absorbed by the atmosphere in the lightpath as a function of the wavelength. The absorbed amount of light is then used to determine the ammonia concentration between the light

source, the retroreflector, and the receiver. A more extensive description of the miniDoas instrument can be found in (Volten et al. 2012; Berkhout et al. 2017).

3.1.2. Data selection

In this study, one of the objectives is to use the measurements of miniDOAS instruments to estimate the ammonia emissions in the surrounding regions. However, when the ammonia concentrations are strongly influenced by local ammonia sources, the measurements might not be representative for larger regions. Especially under conditions with lower mixing layer heights, for example, during the night, there is only a limited amount of mixing of NH_3 . In this case, emitted ammonia remains close to the emission source leading to high local concentrations. These can differ substantially from the estimated NH_3 concentrations of the LOTOS-EUROS model, which are average values in grid cells of several kilometers. Vice versa, the model can also overestimate the NH_3 observation in smaller low emission zones/natural areas. To avoid the possibility that the measurements are highly influenced by local sources, the LML observations are only used when the height of the atmospheric mixed layer is higher than 350m.

3.1.3. Uncertainty of LML observations

The miniDOAS instrument measures the NH_3 concentrations, but the measurements contain uncertainties. Possible causes for the incorrect measurements are errors in the retrieval algorithm or lamp artifacts. (Berkhout et al. 2017) estimated the uncertainties in the measurements to be around $0.25 \mu\text{g}/\text{m}^3$. In (Dammers, Schaap, et al. 2017), additional uncertainties related to dark current are considered, resulting in random errors with a standard deviation of $0.4 \mu\text{g}/\text{m}^3$.

3.1.4. LML observation simulated by LOTOS-EUROS model

To compare the LOTOS-EUROS model and the measurements of the LML stations, an operator \mathcal{H}_k , needs to be defined that maps the concentrations of the LOTOS-EUROS model to a resemblance of the LML-station observations, as described in section 2.3. In the case of LML stations, the measurements at time t_k are approximated by the NH_3 concentration of the LOTOS-EUROS model at time t_k for the grid cell containing the LML station. Then, the operator \mathcal{H}_k is a matrix with as many rows as there are LML-stations, and as many columns as there are grid cells. This linear operator \mathbf{H}_k has for each row all zeros, except for the index corresponding to the grid cell that contains the LML station, where the value is one. Hence, for each LML station s for which an observation is available at time k , and when the atmospheric mixed layer is at least 350m, the matrix \mathbf{H}_k contains a row:

$$[\mathbf{H}_k]_s = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0]. \quad (3.1)$$

3.2. IRS instrument on board the MTG-S satellite

3.2.1. Description

The second measurement instrument that is considered is the InfraRed Sounder (IRS). The IRS instrument is one of the two instruments on board the future Meteosat Third Generation Sounding satellites (MTG-S) as part of the SENTINEL-4 mission. MTG-S is currently planned to be launched at the end of 2023. Hence, the instrument is not operational yet. An artist's impression of the MTG-S satellite can be found in Figure 3.1b. The objective of the SENTINEL-4 mission is to provide data to monitor the amount of trace gasses and aerosols in the atmosphere over Europe. The SENTINEL-4 mission is part of the European earth observation program Copernicus and is developed in collaboration with the European Space Agency (ESA), the European Environment Agency (EEA), and the European Union, in support of the Copernicus Atmosphere Monitoring Service (CAMS).

The IRS instrument is capable of observing multiple chemical tracers, including NH_3 , as well as on meteorological properties such as cloud coverage. The NH_3 column densities are expected to be retrieved at a high spatial resolution of $4\text{km} \times 4\text{km}$ at nadir. MTG-S is a geostationary satellite, implying that the satellite always views the earth from the same perspective. This means that the IRS instrument will provide hourly observations for all of Europe and the northern part of Africa.

3.2.2. Uncertainty of IRS observations

Satellite observations have several sources of uncertainty. As the IRS instrument is not operational yet, no validation of the observations and their uncertainty can be made yet. In this study, the accuracy of NH_3 retrieval algorithm is considered to be the same as for the Cross-track Infrared Sounder (CrIS) following (Dammers, Shephard, et al. 2017).

The measurement errors for NH_3 satellite observations can have various causes such as instrument noise, presence of interfering species such as H_2O , CO_2 , and O_3 , smoothing errors in the retrieval approach. Some errors are systematic and others are random. In this study, only random errors are considered. Systematic errors are neglected. In this study, the same standard deviation for the random observation error is used as derived by (Dammers, Shephard, et al. 2017) for CrIS, which has comparable performance as the IRS although at a coarser spatial resolution. This results in an uncertainty of:

- $5.3 \cdot 10^{15}$ molec/cm² when the measured NH_3 column densities are higher than 10^{16} molec/cm²;
- $4.1 \cdot 10^{15}$ molec/cm² when the measured NH_3 column densities are lower than 10^{16} molec/cm².

Biases of up to 30% of the measured concentration were found for the CrIS retrieval. However, in this study, biases are not considered. It would however be an interesting experiment for the future to test how the 4D-Var method developed in this study would perform when bias is added to the observations, as biases are expected to happen in reality as well.

3.2.3. IRS observations simulated by LOTOS-EUROS model

Also for the IRS measurements an observation operator $\mathcal{H}(\mathbf{x}, \theta)$ should be defined, using the state \mathbf{x} of the LOTOS-EUROS model. The operator is first used to generate synthetic IRS measurements using a (perturbed) LOTOS-EUROS run, and later also as part of the 4D-var inversion to simulate the measurements from model states.

In this study, the CAMS Satellite Operator (CSO) toolbox is used as implementation for the observation operator. The CSO toolbox is developed as part of the Copernicus Atmospheric Monitoring Service (CAMS). CSO is currently used as an observation operator for the TROPOMI instrument on the Sentinel-5p satellite, but can also be adjusted for hourly NH_3 concentration of the IRS instrument.

For real retrievals, the simulation should be formed by convolution of the concentration profile with a so-called averaging kernel that accounts for height-dependend measurement sensitivities. In this study with synthetic data, however, the averaging kernels are ignored, and the basic simulated column densities are used to mimic retrievals. These column densities are retrieved for the expected footprints of the IRS instrument. Only the retrievals that are generated between 04:00 and 20:00 are considered to be of sufficient quality to use; although also nighttime observations will be available, these are likely to be of much lower accuracy.

One of the limiting factors of the quality of satellite observations is cloud coverage. The instrument cannot measure NH_3 under the cloud and for conditions with a high fraction of cloud cover. In that case, most of the observations of the IRS instrument are expected to be of insufficient quality. However, for sake of simplicity, the optimistic and arguably unrealistic case is considered in this study where there are clear-sky conditions for the entire time period.

4

Data Assimilation: adjoint-free 4D-Var method

In this chapter, the adjoint-free 4D-Var data assimilation will be described. First, the choice for parameters to estimate will be described in section 4.1. Second, the motivation for choosing the 4D-Var method will be given in section 4.2. In section 4.3, an introduction of the 4D-Var method for a general case will be given. After that, different versions of adjoint-free 4D-Var methods will be derived. In section 4.4, the 4D-Var method will be described under the assumption that both the parameters and the observations are samples from a Gaussian distribution. In section 4.5, an extension will be discussed where the parameters are assumed to be samples from a log-normal distribution. In section 4.6, an extension of the 4D-Var method will be discussed for cases where the number of parameters is very large and the first two methods become infeasible in terms of computational cost. In section 4.7, a maximum likelihood method will be described that estimates parameters of the 4D-Var cost functions. Section 4.8 discusses two ways to reduce the computational cost of the 4D-Var methods. Last, section 4.9 gives a summary of all the steps of the adjoint-free 4D-Var procedure.

4.1. Problem description

The LOTOS-EUROS model approximates NH_3 concentrations, but the estimated values differ from the measurements. In Figure 4.1, time series can be found with the measured NH_3 concentrations at the LML station Zegveld in blue and the estimated concentrations of the LOTOS-EUROS model in orange, for the month of March 2017. It can be seen that the measurements are initially relatively similar to the model estimates, but as of March 13th, large differences start to occur. Also, it seems that both the modeled and observed NH_3 concentrations increase and decrease for approximately the same days, but the rate of increase and decrease could differ strongly.

The difference between the LOTOS-EUROS model and observations is for a large part due to uncertainties in the LOTOS-EUROS model. Uncertainties are for example present in model parameters such as meteorological input data, chemical reaction parameters, and emission inventories, but also in the implementation of the model processes. In this study, we focus on uncertainties in the ammonia emission estimates. Errors in the emission inventories can be caused by several uncertainties, such as the emission strength, the temporal profile, the location, or the height of the emission. The uncertainties in the model and more specific in the emissions are visualized in Figure 4.2.

As described in section 1.2, uncertainty in the used emission inventory is mainly related to the application of manure and fertilizer. Errors in the modeled emission timing and emission strength could explain why the difference in NH_3 concentrations of the model and the LML station in Figure 4.1 is very different for each day. In this study, a data assimilation method will be used to improve these emission parameters such that the LOTOS-EUROS model will better resemble the observational data.

In this study, the objective is to estimate daily Multiplication Factors (MF) as parameters. The origi-

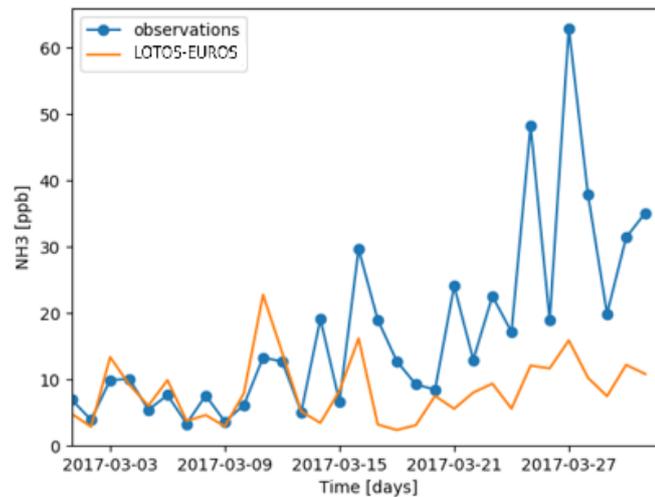


Figure 4.1: Daily average NH_3 concentrations. The observations of LML station are blue and the estimated concentration of LOTOS-EUROS model are orange

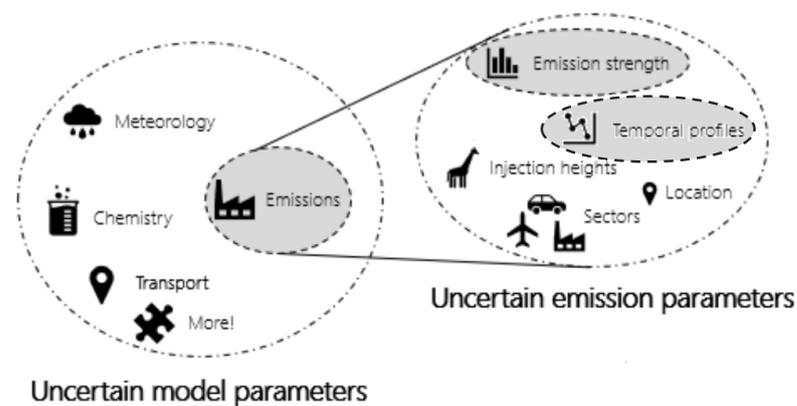


Figure 4.2: Uncertain parameters in the LOTOS-EUROS model.

nal emission inventory, described in section 2.4.1, will be multiplied by these multiplication factors. The approach with multiplication factors is based on the study of (Zijlker 2020). This approach with multiplication factors conserves important properties of the original emission inventory, such as the location of the major emission sources. The parameter vector θ will contain different multiplication factors for each day, possibly varying for different parts of the domain. In this way, the emission strengths of the original emission inventory can be improved in time, improving the temporal profiles, and optionally in space as well. When all multiplication factors are exactly one, the original emission inventory is retrieved.

The multiplication factors are the only estimated parameters in this study. The other sources of uncertainty shown in Figure 4.2, are not taken into account. When using the data assimilation approaches described in the remainder of this chapter, the method might incorrectly adjust the MF's to compensate for other model errors. This might lead to incorrect estimations of the multiplication factors. It is important to keep that in mind when analyzing results for real-life measurements.

4.2. Data assimilation methods

The book of (Lewis, Lakshmivaran, and Dhall 2006) contains an extensive explanation of data assimilation methods and their properties. Data assimilation methods are divided into two categories: variational methods for deterministic models, and statistical methods for stochastic models. In this section, the variational method 4D-Var and the statistical method Ensemble Kalman filter, will be com-

pared. It will be decided that the 4D-Var method suits the objective of this study better.

The variational data assimilation method 4D-Var assumes that the considered model is deterministic. In this way, certain physical properties such as the conservation of mass are satisfied. The 4D-Var method minimizes a cost function over a certain time range, which is useful as it allows observations made before and after a certain time t to influence parameters at time t . This property of the 4D-Var method is useful when NH_3 is emitted at time t , transported to a measuring system during a period Δt , and is then measured in the "future" $t + \Delta t$. Also, the 4D-Var method can optimize many parameters for different time instances simultaneously, taking into account the dependencies between the parameters. In this study, the parameters (emission MF's) will not be able to change in time. To still incorporate time variations, a separate MF is considered for each day. A benefit of the 4D-Var method is that it works well when the number of observations is large, as the computational time barely increases for large data sets. The major downside of the 4D-Var method is that it is generally solved using an iterative gradient-based optimization method, where an adjoint model is used to determine the gradient. However, such an adjoint model is not available for the extensive LOTOS-EUROS model, and developing one would be cumbersome as it would take a lot of programming effort to make it and to maintain it.

The second category, statistical data assimilation methods, assumes that the considered model is stochastic. Random noise is added to the model state \mathbf{x} of equation (2.2), to model the uncertainties in Figure 4.2. When this noise is Gaussian, a statistical method that could be used is the Kalman filter. In the case of the ensemble Kalman filter, an ensemble of model runs is used to quantify how uncertainties influence the NH_3 concentrations in time. As random noise is added to the state, the system no longer obeys laws of physics, such as conservation of mass. Additionally, the state no longer needs to correspond to a run of the LOTOS-EUROS model for some emission inventory. The Kalman filter uses its observations sequentially, meaning that parameters are adjusted immediately when new observations are available. This leads to different parameters for each time instance for which measurements are available. For constant or slowly-varying parameters, the Kalman filter is not a good method. In this case, the Kalman filter state could become biased or even diverge from the true state (Vermeulen and Heemink 2006). Also, the computational complexity of the Kalman filter can increase too much when the number of observations becomes large.

Because of the fact that the 4D-Var method uses observations over a longer time range, obeys the conservation of mass, and is suitable for a large number of observations, the choice has been made to use the 4D-Var method to estimate NH_3 emission parameters. The major downside of the 4D-Var method, the need for an adjoint model, is overcome by developing an adjoint-free version of the 4D-Var method.

4.3. Bayesian approach to 4D-Var method

In this section, a statistical analysis of the 4D-Var cost function will be described. The cost function will be derived for some general parameter vector θ . In the remainder of this report, several other parameter vectors β , γ , and \mathbf{w} will be considered, all related to the multiplication factors, but with different statistical prior distributions. This section describes the 4D-Var cost function for a general probability density function.

The 4D-Var method considers all observations in a certain time window. Given these observations \mathbf{y}^o and a background parameter estimate θ^b , and their probability density function, the 4D-Var method aims to find the parameter vector $\hat{\theta}$ which is the most probable. Hence, the objective of the 4D-Var method is to let the optimal parameter vector $\hat{\theta}$ be

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} p(\theta|\mathbf{y}^o, \theta^b). \quad (4.1)$$

Bayes' theorem states that the following relationship holds:

$$p(\theta|\mathbf{y}^o, \theta^b) \propto p(\mathbf{y}^o|\theta, \theta^b) p(\theta|\theta^b). \quad (4.2)$$

Hence, once the prior distribution $p(\theta|\theta^b)$ and the likelihood of the observations $p(\mathbf{y}^o|\theta, \theta^b)$, are determined, the 4D-Var objective is defined.

Instead of maximizing the posterior distribution, one can also minimize the opposite of the logarithm of the posterior distribution. This is done to simplify the calculations with the function. Combining the objective (4.1), the equivalence in equation (4.2), taking the opposite of the logarithm of the distribution, and neglecting constant terms leads to a new objective:

Find the parameter vector $\hat{\theta}$ for which

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} (-\log p(\theta|\theta^b) - \log p(\mathbf{y}^o|\theta, \theta^b)). \quad (4.3)$$

The function between the parentheses is generally referred to as the *cost function*. In the subsequent sections, several choices for the parameters θ and their prior distributions will be discussed. Additionally, it will be discussed how the cost functions are minimized and how the uncertainty in the optimal solution can be determined.

4.4. Adjoint-free 4D-Var method for Gaussian distributions

In this section, a 4D-Var approach will be described to optimize a parameter vector consisting of multiplication factors, denoted by β . As the first step, the easiest and most common assumption for the prior distribution and observation likelihood is used: both are samples from Gaussian distributions. The background parameters β^b are defined to be $\beta^b = \mathbf{1}$, which results into emissions that are equal to the inventory.

First, the 4D-Var cost function for the Gaussian parameter β will be derived. Second, a linearization of the model state will be discussed. This approximate linear state is needed to estimate several terms in the 4D-Var cost function within a feasible amount of time. Because of this linearization, an adjoint model as usually required to solve the 4D-var problem is not needed. Once the value and gradient of the 4D-Var cost function can be determined, the optimal parameter vector β that minimizes the 4D-Var cost function can be determined. Last, some remarks are made regarding the accuracy and uncertainty of the optimally estimated parameter vector $\hat{\beta}$.

4.4.1. Definition of the cost function and gradient

First, the cost function of the 4D-Var method will be derived. To determine the cost function in equation (4.3), the distributions $p(\beta|\beta^b)$ and $p(\mathbf{y}^o|\beta, \beta^b)$ have to be defined.

Under the assumption that both the prior distribution and the observation likelihood are Gaussian, we obtain for the prior distribution of the multiplication factors that:

$$\beta \sim \mathcal{N}(\beta^b + \mu_\beta, \mathbf{B}_\beta). \quad (4.4)$$

When $\mu_\beta = \mathbf{0}$, the multiplication factors have as expected value the background parameters β^b . When $\mu_\beta \neq \mathbf{0}$, the multiplication factors are biased. \mathbf{B}_β represents the error covariance of β . The values and parameterization of the bias μ_β and the covariance \mathbf{B}_β will be defined in chapter 5 for different experiments.

The observations \mathbf{y}^o are simulated by the LOTOS-EUROS model as \mathbf{y}^{sim} of equation (2.3). As discussed earlier in section 2.3, the measured observations are generally different:

$$\mathbf{y}^o = \mathcal{H}(\mathbf{x}, \beta) + \epsilon. \quad (4.5)$$

This error ϵ is assumed to be unbiased and Gaussian, such that:

$$\mathbf{y}^o - \mathcal{H}(\mathbf{x}, \beta) = \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \quad (4.6)$$

Here, \mathbf{R} is the error covariance of the observation representation.

Now, as the probability density functions $p(\beta|\beta^b)$ and $p(\mathbf{y}^o|\beta, \beta^b)$ are known for the Gaussian distributions, the posterior distribution of β can be calculated. In this context, the objective (4.3) becomes:

Find the parameter vector $\hat{\beta}$, that minimizes the cost function

$$J(\beta) = \frac{1}{2}(\beta - \beta^b - \mu_\beta)^T \mathbf{B}_\beta^{-1}(\beta - \beta^b - \mu_\beta) + \frac{1}{2} \sum_{m=1}^{N_O} (\mathbf{y}_m^O - \mathcal{H}_m(\mathbf{x}_m, \beta))^T \mathbf{R}_m^{-1}(\mathbf{y}_m^O - \mathcal{H}_m(\mathbf{x}_m, \beta)). \quad (4.7)$$

Here, N_O denotes the number of time instances for which observations are available. If a parameter vector $\hat{\beta}$ minimizes the cost function (4.7), then the following holds:

$$\nabla_\beta J(\hat{\beta}) = \mathbf{0}. \quad (4.8)$$

The minimum of $J(\beta)$ can sometimes be found analytically by solving $\nabla_\beta J(\beta) = 0$ for β . Alternatively, if the solution cannot be determined analytically, numerical solvers can be used to find β . Most numerical solvers use an iterative procedure in which the gradient $\nabla_\beta J(\beta)$ is used. Hence, in both cases it is convenient to have an expression of the gradient. The gradient of the cost function $J(\beta)$, as defined in equation (4.7), is:

$$\nabla_\beta J(\beta) = \mathbf{B}_\beta^{-1}(\beta - \beta^b - \mu_\beta) + \sum_{m=1}^{N_O} \left(\frac{\partial \mathcal{H}_m}{\partial \beta} \right)^T \mathbf{R}_m^{-1}(\mathcal{H}_m(\mathbf{x}_m, \beta) - \mathbf{y}_m^O). \quad (4.9)$$

4.4.2. Approximate linear state to make the 4D-Var method adjoint free

To determine the cost $J(\beta)$ and its gradient, the simulated observations $\mathcal{H}_k(\mathbf{x}_k, \beta)$ and their gradient towards parameter changes $\frac{\partial \mathcal{H}_k}{\partial \beta}$ need to be determined. The simulated observations $\mathcal{H}_k(\mathbf{x}_k, \beta)$ can be determined by the LOTOS-EUROS model. However, it takes a lot of computational effort to determine $\mathcal{H}_k(\mathbf{x}_k, \beta)$ for several β in an iterative procedure. Calculation of the adjoint operator $\frac{\partial \mathcal{H}_k}{\partial \beta}$ would be even more difficult, as the NH_3 concentrations, and hence the operator $\mathcal{H}_k(\mathbf{x}_k, \beta)$ are nonlinear in terms of the emission parameters β .

The most common approach to estimate the gradient of the model is by using an adjoint model. This method is described by (Courant and Hilbert 1953). However, an adjoint model is currently not available for the LOTOS-EUROS model. Instead, this study develops an adjoint-free 4D-Var method, similar to the method described in (Zijlker 2020). An explicit expression of $\mathcal{H}_k(\mathbf{x}_k, \beta)$, in terms of the parameter vector β will be made, such that both $\mathcal{H}_k(\mathbf{x}_k, \beta)$ and $\frac{\partial \mathcal{H}_k}{\partial \beta}$ are easily determined given a certain β .

Both observation operators in this study (for the LML stations and for the IRS instrument of the MTG-S satellite) are linear in the state \mathbf{x}_k :

$$\mathcal{H}_k(\mathbf{x}_k, \beta) = \mathbf{H}_k \mathbf{x}_k. \quad (4.10)$$

Because of this linearity, finding an expression of $\mathcal{H}_k(\mathbf{x}_k, \beta)$ in terms of β reduces to finding an expression of the NH_3 concentrations \mathbf{x}_k in terms of β .

To estimate the state, an approach based on the method of (Zijlker 2020) is used. There, the state is first approximated by the Trajectory Piecewise Linearization (TPWL). The LOTOS-EUROS model, $\mathbf{x}_k = M_k(\mathbf{x}_{k-1}, \beta)$, is linearized around the background state \mathbf{x}^b and background parameter β^b . The background state \mathbf{x}^b is the determined state of the LOTOS-EUROS model when using the background parameter vector β^b . This linearization leads to:

$$\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^b \approx \left[\frac{\partial M_{k+1}}{\partial \beta}(\mathbf{x}_k^b, \beta^b) \right] (\beta - \beta^b) + \left[\frac{\partial M_{k+1}}{\partial \mathbf{x}}(\mathbf{x}_k^b, \beta^b) \right] (\mathbf{x}_k - \mathbf{x}_k^b). \quad (4.11)$$

The computational effort of this linearization are dominated by determining the Jacobian matrices. As both the parameter vector β and especially the state \mathbf{x} can be high-dimensional, the Jacobians can be

very large in size. Approximating of the derivative with respect to the state \mathbf{x} can be inaccurate or computationally infeasible. Fortunately, the study of (Zijlker 2020) showed that the derivative of the model with respect to the state \mathbf{x} could be neglected for the relatively short-lived trace gas NO_2 . As NH_3 is also a short-lived trace gas, it is assumed that in this linear approximate state, the derivative of the model with respect to \mathbf{x} can be neglected as well. Now, it follows that the state \mathbf{x} can be expressed in terms of β , once the background state \mathbf{x}^b , background parameter β^b and the Jacobian matrix $\frac{\partial M_k}{\partial \beta}(\mathbf{x}_{k-1}^b, \beta^b)$ are known, as:

$$\mathbf{x}_k = \mathbf{x}_k^b + \left[\frac{\partial M_k}{\partial \beta}(\mathbf{x}_{k-1}^b, \beta^b) \right] (\beta - \beta^b) = \mathbf{x}_k^b + [\mathbf{E}_\beta]_k (\beta - \beta^b). \quad (4.12)$$

Here:

$$[\mathbf{E}_\beta]_k = \frac{\partial M_k}{\partial \beta}(\mathbf{x}_{k-1}^b, \beta^b). \quad (4.13)$$

This Jacobian is independent of β . Therefore, it only has to be determined once in the process of finding the optimal parameters $\hat{\beta}$. If the minimum of the cost function $J(\beta)$ were to be found analytically, the matrix $[\mathbf{E}_\beta]_k$ would become a constant matrix, and if the cost function were to be found using a numerical procedure, this matrix would be the same for each iteration.

The Jacobian $[\mathbf{E}_\beta]_k$ can not be determined explicitly, as the LOTOS-EUROS model depends on the multiplication factors in a complicated manner. Therefore, the Jacobian is approximated using a finite difference approach. The i 'th column of the matrix is approximated as:

$$[\mathbf{E}_\beta]_{k,i} = \left[\frac{\partial M_k}{\partial \beta}(\mathbf{x}_{k-1}^b, \beta^b) \right]_i \approx \frac{M_k(\mathbf{x}_{k-1}^b, \beta^b + \Delta\beta_i \mathbf{E}_i) - M_k(\mathbf{x}_{k-1}^b, \beta^b)}{\Delta\beta_i}. \quad (4.14)$$

Experiments have shown that this linear approximation of the state in terms of the multiplication factors β , defined in equation (4.12) is accurate, also for large deviations. This is because the LOTOS-EUROS model is approximately linear with respect to emissions. An increase in emissions at one location is expected to result in a proportional increase in the concentration of the ammonia plume carried downwind from the emission source.

Most nonlinear terms of the LOTOS-EUROS model are related to chemistry and the balance between dry emission and dry deposition. As a result of this linear approximate state, these nonlinear terms are simplified. However, this does in general not seem to cause major errors. Alternatively, if larger errors in the estimate of $\hat{\beta}$ would occur due to the nonlinear terms, one could also perform multiple optimization loops: first find the optimal parameter vector $\hat{\beta}$ using the linear approximate state, based on the emission inventory of the LOTOS-EUROS model with background parameter $\beta^b = \mathbf{1}$. Then, perform a second optimization iteration with as background parameter $\beta^b = \hat{\beta}$ to find a second estimate for the parameter vector $\hat{\beta}$. This can be repeated several times. However, in this study, the linear approximate state is considered to be accurate enough after one optimization loop.

Now that the state \mathbf{x} can be expressed in terms of the parameter vector β , the observation operator can be expressed in terms of β as well:

$$\mathcal{H}_k(\mathbf{x}_k, \beta) = \mathbf{H}_k \mathbf{x}_k = \mathbf{H}_k \mathbf{x}_k^b + \mathbf{H}_k [\mathbf{E}_\beta]_k (\beta - \beta^b). \quad (4.15)$$

The adjoint operator becomes:

$$\left(\frac{\partial \mathcal{H}_k}{\partial \beta} \right)^T = (\mathbf{H}_k [\mathbf{E}_\beta]_k)^T. \quad (4.16)$$

The computational effort of the 4D-Var method now gets dominated by determining the Jacobian matrix $[\mathbf{E}_\beta]_k$, as for each element of β a model run needs to be performed. Let N_{MF} number of multiplication factors, i.e. the length of the parameter vector β . In total, there will be $N_{MF} + 1$ model runs: one run to determine the background state \mathbf{x}^b and one run for each of the N_{MF} columns of $[\mathbf{E}_\beta]_k$.

4.4.3. Finding the optimal parameter vector

Now that all terms in the cost function in equation (4.7) and the gradient in equation (4.9) are defined, the function can be minimized. A convenient result from the assumption of Gaussian distributions is that all terms in the gradient in equation (4.9) are linear. Substitution of the approximate observations from equation (4.15) and the adjoint operator in equation (4.16) in the gradient of the cost function in equation (4.9) leads to:

$$\begin{aligned} \nabla_{\beta} J(\beta) &= \mathbf{B}^{-1}(\beta - \beta^b - \mu_{\beta}) \\ &+ \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_{\beta}]_m)^T \mathbf{R}_m^{-1} \left(\mathbf{H}_m \mathbf{x}_m^b + \mathbf{H}_m[\mathbf{E}_{\beta}]_m(\beta - \beta^b) - \mathbf{y}_m^o \right). \end{aligned} \quad (4.17)$$

If the gradient is set to be equal to zero, this becomes a linear system in terms of β and can therefore be solved analytically. Define:

$$\begin{aligned} \mathbf{A}_{\beta} &= \mathbf{B}^{-1} + \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_{\beta}]_m)^T \mathbf{R}_m^{-1} (\mathbf{H}_m[\mathbf{E}_{\beta}]_m) \\ \mathbf{c}_{\beta} &= \mathbf{B}^{-1} \beta^b + \mathbf{B}^{-1} \mu_{\beta} + \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_{\beta}]_m)^T \mathbf{R}_m^{-1} \left(-\mathbf{H}_m \mathbf{x}_m^b + \mathbf{H}_m[\mathbf{E}_{\beta}]_m \beta^b + \mathbf{y}_m^o \right). \end{aligned} \quad (4.18)$$

Then:

$$\nabla_{\beta} J(\hat{\beta}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}_{\beta} \hat{\beta} = \mathbf{c}_{\beta} \quad \Leftrightarrow \quad \hat{\beta} = \mathbf{A}_{\beta}^{-1} \mathbf{c}_{\beta}. \quad (4.19)$$

By solving the system, the optimal parameter vector $\hat{\beta}$ is found.

4.4.4. Covariance of the optimal solution

It could be useful have the uncertainty of the optimal parameter $\hat{\beta}$ quantified. The covariance of the posterior distribution of $\hat{\beta}$ is equal to the inverse of the Hessian of the cost function $J(\beta)$. Using the notation of matrix \mathbf{A}_{β} and vector \mathbf{c}_{β} from the equations in system (4.18), the gradient of the cost function can be written as:

$$\nabla_{\beta} J(\beta) = \mathbf{A}_{\beta} \beta - \mathbf{c}_{\beta}. \quad (4.20)$$

With this, the Hessian becomes:

$$\nabla_{\beta} (\nabla_{\beta} J(\beta)) = \nabla_{\beta} (\mathbf{A}_{\beta} \beta - \mathbf{c}_{\beta}) = \mathbf{A}_{\beta}. \quad (4.21)$$

Hence, the covariance of the posterior distribution of $\hat{\beta}$ is \mathbf{A}_{β}^{-1} . As for any covariance matrix, the diagonal of \mathbf{A}_{β}^{-1} will contain the variances for each of the multiplication factors. These variances can be used to determine a confidence interval for the estimated multiplication factors.

From the definition of \mathbf{A}_{β} in equation (4.18), it can be seen that norm of the diagonal elements of matrix \mathbf{A}_{β} becomes larger when the number of measurements increases, when $\mathbf{H}_k[\mathbf{E}_{\beta}]_k$ contains large values (meaning that perturbing emissions results in large changes in measured concentrations), or when the entries of \mathbf{R}_k^{-1} are large (the errors in observations are small). If the diagonal elements of \mathbf{A}_{β} are high valued, \mathbf{A}_{β}^{-1} has low values diagonal elements, which means that the uncertainty in the optimally estimated parameters is small. In this case, the optimal multiplication factors $\hat{\beta}$ are expected to be relatively close to the true multiplication factors β^{true} .

4.5. Adjoint-free 4D-Var method for log-normal prior distributions

In the previous section, the 4D-Var method for multiplication factors with a Gaussian distribution was discussed. However, if one considers multiplication factors for ammonia emissions, this distribution might not be appropriate. As this project aims to find large positive deviations in the emission factors, caused by the irregular spreading of manure or fertilizer, a log-normal distribution for the multiplication factors seems more realistic. A log-normal distribution assumes an equal probability of increasing or

decreasing emissions, similar to a Gaussian distribution, but for the log-normal distribution, the MF's are only allowed to take positive values and are more likely to have some high peak values. An example of the log-normal distribution $\text{LogN}(0, 1)$ is shown in Figure 4.3.

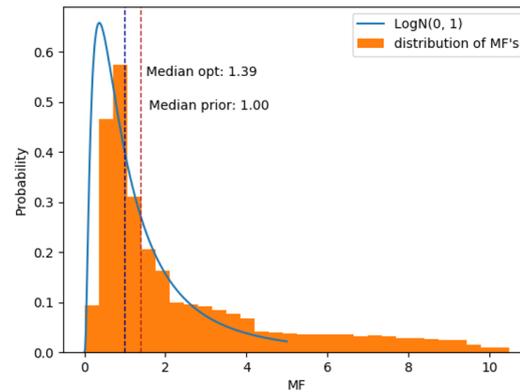


Figure 4.3: Log-normal distribution

4.5.1. Log-normal multiplication factors

Let the multiplication factors with a log-normal distribution be denoted by γ . An issue with the 4D-Var method is now that the usual cost function (4.7) is based on Gaussian distributions. Also, the log-normally distributed γ are defined to be positive, adding a constraint to the problem. To incorporate the log-normality of γ , the following property of the log-normal distribution is used. If:

$$\gamma - \gamma^b \sim \text{LogN}(\mu_\gamma, \mathbf{B}_\gamma), \quad (4.22)$$

then the parameter:

$$\beta = \log(\gamma) \quad (4.23)$$

has a Gaussian distribution with mean $\beta^b + \mu_\gamma$ and covariance \mathbf{B}_γ . Hence, $\beta - \beta^b \sim \mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma)$. Here, the background parameter vector $\gamma^b = \mathbf{1}$, and $\beta^b = \mathbf{0}$, such that $e^{\beta^b} = \gamma^b = \mathbf{1}$. Instead of expressing the cost function in terms of the multiplication factors γ , we now let the cost function be a function of $\beta = \log(\gamma)$. If the dependence of $\mathcal{H}_m(\mathbf{x}_m, \gamma)$ and its gradient with respect to β can be expressed in terms of β , instead of the multiplication factors γ , the 4D-Var method can be used to determine the optimal $\hat{\beta}$. Then, the optimal multiplication factors can be calculated as $\hat{\gamma} = \exp \hat{\beta}$. Since the exponential function will always evaluate to positive numbers, the constraint of positive $\hat{\gamma}$ is automatically satisfied.

4.5.2. Approximate linear state, cost function, and gradient

In section 4.4.2, it was described that the state x could be approximated by a linear approximation of the model in terms of the multiplication factors. This approximation appeared to be accurate, as the LOTOS-EUROS model is nearly linear with respect to ammonia emissions. However, in this section, the transformation has been made from the multiplication factor parameter vector γ to the parameter vector $\beta = \log(\gamma)$. To still use the state approximation from equation (4.12), a small adjustment has to be made. One would like the state to be linear in the multiplication factors γ :

$$\mathbf{x}_k = \mathbf{x}_k^b + [\mathbf{E}_\gamma]_k (\gamma - \gamma^b). \quad (4.24)$$

Here, the matrix:

$$[\mathbf{E}_\gamma]_k = \frac{\partial M_{k+1}}{\partial \gamma}(\mathbf{x}_k^b, \gamma^b),$$

is the same matrix as $[\mathbf{E}_\beta]_k$ in equation (4.13), with the only difference being that the multiplication factors were denoted as β in section 4.4.2 and are now called γ . Apart from notation, the matrices are identical. However, γ should not be called explicitly in the cost function, as the cost function should depend on β only. Therefore, the reverse transformation $\gamma = \exp \beta$ should be substituted. This results in a new approximate state:

$$\mathbf{x}_k = \mathbf{x}_k^b + [\mathbf{E}_\gamma]_k (e^\beta - e^{\beta^b}). \quad (4.25)$$

Since the parameter vector β has a Gaussian distribution, the cost function in equation (4.7) and the gradient in equation (4.9) are valid for this problem. The difference with the previous section is that the estimated observations $\mathcal{H}(\mathbf{x}, \beta)$ and the adjoint operator $\left(\frac{\partial \mathcal{H}_m}{\partial \beta}\right)^T$ in the cost function and gradient are different. For the approximate state (4.25), it is obtained that:

$$\mathcal{H}(\mathbf{x}, \beta) = \mathbf{H}\mathbf{x}_k^b + \mathbf{H}[\mathbf{E}_\gamma]_k(e^\beta - e^{\beta^b}) \quad (4.26)$$

$$\left(\frac{\partial \mathcal{H}}{\partial \beta}\right)^T = \left(\mathbf{H}[\mathbf{E}_\gamma]_k \text{diag}(e^\beta)\right)^T \quad (4.27)$$

It follows that the gradient of the 4D-Var problem becomes:

$$\begin{aligned} \nabla_\beta J(\beta) &= \mathbf{B}_\gamma^{-1}(\beta - \beta^b - \mu_\gamma) \\ &+ \sum_{m=1}^{N_0} \left(\mathbf{H}_m[\mathbf{E}_\gamma]_m \text{diag}(e^\beta)\right)^T \mathbf{R}_m^{-1} \left(\mathbf{H}_m \mathbf{x}_m^b + \mathbf{H}_m[\mathbf{E}_\gamma]_m(e^\beta - e^{\beta^b}) - \mathbf{y}_m^o\right). \end{aligned} \quad (4.28)$$

4.5.3. Finding the optimal parameter vector

The gradient of the cost function (4.28) is now non-linear in terms of β . Because of the non-linearity, it is no longer possible to solve the system $\nabla_\beta J(\beta) = \mathbf{0}$ analytically. To still find the parameter vector β that minimizes the cost function, a numerical optimizer is used.

There are many numerical optimizers available, all having their advantages and disadvantages. In this study, the minimizer of the SciPy module in Python using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is used. A detailed description of the BFGS algorithm and its properties can be found in the book of (Nocedal and Wright 2006a). This iterative method is made for solving unconstrained nonlinear optimization problems. As this algorithm is a Quasi-Newton method, fast or even super-linear convergence could occur. Because the method uses the gradient of the cost function as an input value, the method needs fewer iterations. An additional benefit of the method is that it returns an approximation of the inverse of the Hessian of the cost function. As mentioned in section 4.4.4, this inverse Hessian of the optimal Gaussian parameter $\hat{\beta}$ can be used to quantify the uncertainty of the optimal estimate. The fact that the inverse Hessian is returned is considered an advantage when the size of the parameter vector β is small (less than 1000). However, when the parameter vector becomes large, the size of the inverse Hessian matrix will become huge as it increases with order n^2 . This may result in computer memory issues. However, if memory would be an issue, the limited-memory version of the algorithm, the L-BFGS-B method (L: limited memory, B: bounded), can be used, which does not return an approximate of the inverse Hessian matrix. However, in this study the BFGS algorithm is used, as saving the Hessian matrix does not result in memory issues.

The used BFGS algorithm returns an approximation of the inverse Hessian of the system, which can, analogously to section 4.4.4 be used to determine the standard deviations of the elements in the Gaussian parameter vector $\hat{\beta}$. Those standard deviations can be used to determine the two-sigma bounds for the parameters of $\hat{\beta}$. However, we are not interested in the values for $\hat{\beta}$, but in the values of the multiplication factors $\hat{\gamma} = \exp(\hat{\beta})$. If a confidence interval for $\hat{\beta}$ is known to be $(\hat{\beta} - 2\sigma; \hat{\beta} + 2\sigma)$, then a confidence interval for the MF's $\gamma = \exp(\beta)$ is:

$$\left(\exp(\hat{\beta} - 2\sigma); \exp(\hat{\beta} + 2\sigma)\right).$$

4.6. Adjoint-free 4D-Var method when using a preconditioned state

If the size of the to be optimized parameter vector becomes very large, it often becomes too computationally expensive to analyze the effect of a perturbation of every single element of some parameter vector θ by running the LOTOS-EUROS model. In the previous two cases, such an ensemble run was needed to calculate the matrices \mathbf{E}_β and \mathbf{E}_γ . Furthermore, computations with (possibly non-diagonal) matrix \mathbf{B}_θ^{-1} can be computationally expensive if the size of the parameter vector, and consequently \mathbf{B}_θ^{-1}

is very large.

To reduce the number of computations, a preconditioner is used. The idea is to consider only the main patterns of the parameter vector θ . Instead of determining the effect of a perturbation of each instance, only the effect of perturbations of the main patterns needs to be determined. A state transformation is used, such that the optimal parameter vector is found in the reduced space. Once this optimal reduced state vector is found, the optimal parameter $\hat{\theta}$ can be estimated using a reverse transformation. In this way, a good approximation of the entire state can be made while reducing the computational complexity tremendously.

Before going into the details of how to determine the main patterns of some parameter vector θ , the 4D-Var method will be discussed for this new transformed parameter vector. Section 4.6.1 considers the case where the multiplication factors have a Gaussian prior distribution, section 4.6.2 considers the case with the log-normal prior distribution, and section 4.6.3 described the concept of the main patterns of θ .

4.6.1. Preconditioned state: Gaussian prior distribution

First, the case is discussed where the multiplication factors have a Gaussian distribution, as in section 4.4. The multiplication factors are denoted by β . Then, as before:

$$\beta - \beta^b \sim \mathcal{N}(\mu_\beta, \mathbf{B}_\beta). \quad (4.29)$$

The preconditioner here is chosen to be the same as in (Arjo Segers, Tokaya, and Houweling 2020; Meirink, Eskes, and Goede 2006). The preconditioned state \mathbf{w} is defined as:

$$\mathbf{w} = \mathbf{B}_\beta^{-1/2}(\beta - \beta^b). \quad (4.30)$$

Given a certain \mathbf{w} , the multiplication factors β can be determined using the reverse transformation:

$$\beta = \beta^b + \mathbf{B}_\beta^{1/2}\mathbf{w}. \quad (4.31)$$

Here, the multiplication factors β are equal to the sum of the background parameters β^b and a linear combination of the columns of $\mathbf{B}_\beta^{1/2}$, where \mathbf{w} contains the coefficients of the linear combination. The parameter vector \mathbf{w} is in reduced space \mathbb{R}^m , and $\mathbf{B}_\beta^{1/2} \in \mathbb{R}^{N_{MF} \times m}$, with $m \ll N_{MF}$. The matrix $\mathbf{B}_\beta^{1/2}$ is a reduced size approximation of the root of the matrix \mathbf{B}_β , such that:

$$\mathbf{B}_\beta^{1/2}\mathbf{B}_\beta^{T/2} \approx \mathbf{B}_\beta. \quad (4.32)$$

Instead of finding estimates for the potentially very large parameter vector β , a method will be described to determine the values of the reduced state parameter vector \mathbf{w} .

One of the important properties of the transformation in equation (4.30) is that it is linear in $(\beta - \beta^b)$. Because of this linearity, all terms containing the matrix \mathbf{E}_β cancel out when expressing the observation operator $\mathcal{H}_m(\mathbf{x}_m, \mathbf{w})$ from equation (4.15) in terms of \mathbf{w} :

$$\begin{aligned} \mathcal{H}_m(\mathbf{x}_m, \mathbf{w}) &= \mathbf{H}(\mathbf{x}^b + \mathbf{E}_\beta(\beta - \beta^b)) \\ &= \mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}_\beta\mathbf{B}_\beta^{1/2}\mathbf{w} \\ &= \mathbf{H}\mathbf{x}^b + \mathbf{H}\frac{\partial M}{\partial \mathbf{w}}\mathbf{w} \end{aligned} \quad (4.33)$$

$$= \mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}_\mathbf{w}\mathbf{w} \quad (4.34)$$

Here, it was used that:

$$\mathbf{E}_\mathbf{w} = \frac{\partial M}{\partial \mathbf{w}} = \frac{\partial M}{\partial \beta} \frac{\partial \beta}{\partial \mathbf{w}} = \mathbf{E}_\beta \mathbf{B}_\beta^{1/2}. \quad (4.35)$$

To determine $\mathcal{H}_m(\mathbf{x}_m, \mathbf{w})$, it is now no longer needed to determine \mathbf{E}_β for which an ensemble of size N_{MF} was needed. Only $\mathbf{E}_\mathbf{w}$ has to be determined, where \mathbf{w} is of reduced size m . The Jacobian $\mathbf{E}_\mathbf{w}$ is

determined using a finite difference approach, similar to the Jacobian \mathbf{E}_β in equation (4.14). For the i 'th column, the matrix is approximated as:

$$[\mathbf{E}_w]_i = \left[\frac{\partial M}{\partial \mathbf{w}}(\mathbf{x}^b, \mathbf{w}^b) \right]_i = \frac{M(\mathbf{x}, \mathbf{w}^b + \Delta w_i \mathbf{e}_i) - M(\mathbf{x}, \mathbf{w}^b)}{\Delta w_i}. \quad (4.36)$$

This is where the computational effort is reduced, as only $m + 1 \ll N_{MF} + 1$ model runs are needed. Once \mathbf{E}_w is known, calculating $\mathcal{H}_m(\mathbf{x}_m, \mathbf{w})$ and its adjoint $\left(\frac{\partial \mathcal{H}_m}{\partial \mathbf{w}}\right)^T = (\mathbf{H}\mathbf{E}_w)^T$ is computationally cheap.

By substituting the reverse transformation from equation (4.31) and the observation operator from equation (4.33) in equation (4.7), the cost function in terms of the preconditioned state \mathbf{w} becomes:

$$J(\mathbf{w}) = \frac{1}{2}(\mathbf{w} - \mathbf{B}_\beta^{-1/2} \mu_\beta)^T (\mathbf{w} - \mathbf{B}_\beta^{-1/2} \mu_\beta) + \frac{1}{2} \sum_{m=1}^{N_0} (\mathbf{y}_m^o - \mathcal{H}_m(\mathbf{x}_m, \mathbf{w}))^T \mathbf{R}_m^{-1} (\mathbf{y}_m^o - \mathcal{H}_m(\mathbf{x}_m, \mathbf{w})). \quad (4.37)$$

Here, it was used that for a covariance matrix \mathbf{B}_β it holds that

$$\mathbf{B}_\beta = \mathbf{B}_\beta^{1/2} \mathbf{B}_\beta^{T/2}, \quad \text{and hence} \quad \mathbf{B}_\beta^{-1} = \mathbf{B}_\beta^{-T/2} \mathbf{B}_\beta^{-1/2}.$$

It follows that the gradient of $J(\mathbf{w})$ with respect to \mathbf{w} becomes

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{w} - \mathbf{B}_\beta^{-1/2} \mu_\beta + \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_w]_m)^T \mathbf{R}_m^{-1} (\mathcal{H}_m(\mathbf{x}_m, \mathbf{w}) - \mathbf{y}_m^o) \quad (4.38)$$

In the case where $\mu_\beta = \mathbf{0}$, all terms containing \mathbf{B}_β cancel out, eliminating the computations with a possibly very large matrix \mathbf{B}_β . For very large \mathbf{B}_β , it might also take too much computer memory to save the entire matrix \mathbf{B}_β , making computations difficult. If $\mu_\beta \neq \mathbf{0}$, the computations regarding \mathbf{B}_β are only relevant for projecting the vector μ_β on the new reduced space according to the transformation $\mu_w = \mathbf{B}_\beta^{-1/2} \mu_\beta$. The gradient in equation (4.38) is a linear function in \mathbf{w} . Hence, the optimal parameter vector $\hat{\mathbf{w}}$ that minimizes the 4D-Var cost function can be found analytically by solving the system:

$$\nabla_{\mathbf{w}} J(\hat{\mathbf{w}}) = \mathbf{0}. \quad (4.39)$$

Let Matrix $\mathbf{A}_w \in \mathbb{R}^{m \times m}$, and vector $\mathbf{c}_w \in \mathbb{R}^m$ be defined as:

$$\mathbf{A}_w = I + \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_w]_m)^T \mathbf{R}_m^{-1} (\mathbf{H}_m[\mathbf{E}_w]_m),$$

$$\mathbf{c}_w = \mathbf{B}_\beta^{-1/2} \mu_\beta + \sum_{m=1}^{N_0} (\mathbf{H}_m[\mathbf{E}_w]_m)^T \mathbf{R}_m^{-1} (-\mathbf{H}_m^b + \mathbf{y}_m^o). \quad (4.40)$$

Then:

$$\nabla_{\mathbf{w}} J(\hat{\mathbf{w}}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}_w \hat{\mathbf{w}} = \mathbf{c}_w \quad \Leftrightarrow \quad \hat{\mathbf{w}} = \mathbf{A}_w^{-1} \mathbf{c}_w. \quad (4.41)$$

By solving the system, the optimal parameter vector $\hat{\mathbf{w}}$ is found. Similar as before in section 4.4.4, the covariance of the posterior distribution of $\hat{\mathbf{w}}$ is the inverse of the Hessian matrix: \mathbf{A}_w^{-1} . The distribution of \mathbf{w} can be used to derive the distribution of the multiplication factors β . When:

$$\mathbf{w} \sim \mathcal{N}(\hat{\mathbf{w}}, \mathbf{A}_w^{-1}), \quad (4.42)$$

then for the reverse transformation:

$$\beta = \beta^b + \mathbf{B}_\beta^{1/2} \mathbf{w} \sim \mathcal{N}(\beta^b + \mathbf{B}_\beta^{1/2} \hat{\mathbf{w}}, \mathbf{B}_\beta^{1/2} \mathbf{A}_w^{-1} \mathbf{B}_\beta^{T/2}). \quad (4.43)$$

This distribution for β can be used to quantify the uncertainty in the optimally estimated multiplication factors $\hat{\beta}$.

4.6.2. Preconditioned state: log-normal prior distribution

In this section, the preconditioner methodology for large parameter vectors is extended to the case where the multiplication factors have a log-normal distribution. These log-normal multiplication factors, similar to section 4.5, are denoted by γ . As before, in section 4.5, the parameter vector to optimize does not contain the multiplication factors γ , but contains the auxiliary Gaussian parameters $\beta = \log(\gamma)$. Then:

$$\beta - \beta^b \sim \mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma). \quad (4.44)$$

The first decision to be made is to chose a preconditioner. Here, the chosen preconditioned state \mathbf{w} is the same as before in equation (4.30), only now expressed in the auxiliary parameter β instead of the multiplication factors γ :

$$\mathbf{w} = \mathbf{B}_\gamma^{-1/2}(\gamma - \gamma^b) = \mathbf{B}_\gamma^{-1/2}(e^\beta - e^{\beta^b}). \quad (4.45)$$

The corresponding reverse transformation is:

$$\beta = \log(e^{\beta^b} + \mathbf{B}_\gamma^{1/2} \mathbf{w}). \quad (4.46)$$

Note that this reverse transformation is defined only when $e^{\beta^b} + \mathbf{B}_\gamma^{1/2} \mathbf{w} > 0$, i.e. the multiplication factors γ , are positive. Again, the parameter vector \mathbf{w} is in reduced space \mathbb{R}^m and $\mathbf{B}_\gamma^{1/2} \in \mathbb{R}^{N_{MF} \times m}$, with $m \ll N_{MF}$. Substituting the preconditioner in observation operator equation (4.26) for log-normal multiplication factors results in the same linear observation operator $\mathcal{H}_m(\mathbf{x}_m, \mathbf{w})$ as for the Gaussian case in equation (4.33), where all terms with \mathbf{E}_γ cancel out:

$$\begin{aligned} \mathcal{H}_m(\mathbf{x}_m, \mathbf{w}) &= \mathbf{H}(\mathbf{x}^b + \mathbf{E}_\gamma(e^\beta - e^{\beta^b})) \\ &= \mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}_\gamma \mathbf{B}_\gamma^{1/2} \mathbf{w} \\ &= \mathbf{H}\mathbf{x}^b + \mathbf{H} \frac{\partial M}{\partial \mathbf{w}} \mathbf{w} = \mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}_w \mathbf{w} \end{aligned} \quad (4.47)$$

Here, it was used that:

$$\mathbf{E}_w = \frac{\partial M}{\partial \mathbf{w}} = \frac{\partial M}{\partial \gamma} \frac{\partial \gamma}{\partial \mathbf{w}} = \mathbf{E}_\gamma \mathbf{B}_\gamma^{1/2}. \quad (4.48)$$

Again, it is no longer needed to approximate the Jacobian matrix \mathbf{E}_γ using the finite difference approximation with an ensemble of size N_{MF} . To estimate \mathbf{E}_w as in equation (4.36), only $m + 1$ runs of the LOTOS-EUROS model are needed.

By substituting the reverse transformation from equation (4.46) and the observation operator from equation (4.47) in equation (4.7) and (4.9), the cost function and gradient in terms of the preconditioned state \mathbf{w} become:

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2} \left(\mathbf{B}^{-1/2} \left(\log(e^{\beta^b} + \mathbf{B}^{1/2} \mathbf{w}) - \beta^b - \mu \right) \right)^T \left(\mathbf{B}^{-1/2} \left(\log(e^{\beta^b} + \mathbf{B}^{1/2} \mathbf{w}) - \beta^b - \mu \right) \right) \\ &\quad + \frac{1}{2} \sum_{m=1}^{N_0} (\mathbf{y}_m^o - \mathcal{H}_m(\mathbf{x}_m, \mathbf{w}))^T \mathbf{R}_m^{-1} (\mathbf{y}_m^o - \mathcal{H}_m(\mathbf{x}_m, \mathbf{w})), \end{aligned} \quad (4.49)$$

$$\begin{aligned} \nabla_{\mathbf{w}} J &= \left(\mathbf{B}^{1/2} \right)^T \text{diag} \left(\frac{1}{(e^{\beta^b})_i + (\mathbf{B}^{1/2} \mathbf{w})_i} \right) \mathbf{B}^{-T/2} \mathbf{B}^{-1/2} \left(\log(e^{\beta^b} + \mathbf{B}^{1/2} \mathbf{w}) - \beta^b - \mu \right) \\ &\quad + \sum_{m=1}^{N_0} (\mathbf{H}\mathbf{E}_w)^T \mathbf{R}_m^{-1} \left(\mathcal{H}_m(\mathbf{x}_m, \mathbf{w}) - \mathbf{y}_m^o \right) \end{aligned} \quad (4.50)$$

Here it was again used that:

$$\mathbf{B}_\gamma^{-1} = \mathbf{B}_\gamma^{-T/2} \mathbf{B}_\gamma^{-1/2}.$$

This is done, because determining the inverse of the $N_{MF} \times N_{MF}$ sized matrix \mathbf{B}_γ^{-1} can be time consuming for non-diagonal matrices. However, determining $\mathbf{B}_\gamma^{-1/2}$ in reduced form is relatively simple as will be described in section 4.6.3. A relevant remark is that the cost function (4.49) and gradient (4.50) are not

defined for all values of \mathbf{w} . The logarithms in both the functions are only defined for positive values. This is related to the property of the log-normal distribution, stating that the multiplication factors $\gamma = \exp(\beta)$ are defined to be positive. However, there are parameter vectors \mathbf{w} for which this constraint is not satisfied. Hence, an additional constraint is added to the problem, requiring the multiplication factors $\gamma = \exp(\beta)$ to be positive as:

$$e^{\beta^b} + \mathbf{B}^{1/2} \mathbf{w} > 0. \quad (4.51)$$

The nonlinear cost function of equation (4.49), in combination with the constraint of equation (4.51) cannot be minimized analytically. Instead, a numerical minimizer from the SciPy module in Python is used. The algorithm for this problem uses Sequential Least Squares Programming (SLSQP). This method has been chosen as linear constraints can be added, and because it uses the gradient of the cost function. The previously mentioned algorithm L-BFGS-B can also set bounds on individual parameters, but the SLSQP has the advantage that constraints can be considered for linear combinations of the parameters, as is needed according to equation (4.51). The SLSQP algorithm is also a quasi-newton method using BFGS updates. More about the SLSQP algorithm can be found in the book of (Nocedal and Wright 2006b).

A disadvantage of the SLSQP algorithm is that it does not return an approximation of the Hessian inverse, needed to estimate the uncertainty of the optimal parameters $\hat{\mathbf{w}}$. Also, for constrained cost functions in terms of the non-Gaussian parameter \mathbf{w} , it does not hold that the covariance of $\hat{\mathbf{w}}$ is the inverse of the Hessian matrix. Hence, even if the Hessian would be returned, it could not be used to quantify the uncertainty as easily as before.

4.6.3. Reduce size: Eigenvalue decomposition

In this section, it will be described how for some parameter vector θ , its covariance matrix $\mathbf{B}_\theta^{1/2}$, of size $N_{MF} \times N_{MF}$ will be reduced to an approximation of size $N_{MF} \times m$, with $m \ll N_{MF}$ such that:

$$\mathbf{B}_\theta^{1/2} \mathbf{B}_\theta^{T/2} \approx \mathbf{B}_\theta. \quad (4.52)$$

This section is based on the formulation by (Arjo Segers, Tokaya, and Houweling 2020). First, define the covariance matrix \mathbf{B}_θ as:

$$\mathbf{B}_\theta = \mathbf{S}_\theta \mathbf{C}_\theta \mathbf{S}_\theta. \quad (4.53)$$

Here \mathbf{S}_θ is a diagonal matrix formed from a standard deviation field, and \mathbf{C}_θ holds the correlations between elements of \mathbf{B}_θ . Now, use an eigenvalue decomposition of \mathbf{C}_θ :

$$\mathbf{C}_\theta = \mathbf{Q}_\theta \Lambda_\theta \mathbf{Q}_\theta^T. \quad (4.54)$$

Here \mathbf{Q}_θ holds the orthonormal eigenvectors of \mathbf{C}_θ as columns, and Λ_θ is a diagonal matrix with the corresponding eigenvalues. Since \mathbf{B}_θ is a covariance matrix and therefore positive-definite, its eigenvalue decomposition exists and its eigenvalues are strictly positive. Now, \mathbf{B}_θ and subsequently $\mathbf{B}_\theta^{1/2}$ can be rewritten as:

$$\begin{aligned} \mathbf{B}_\theta &= \mathbf{S}_\theta \mathbf{Q}_\theta \Lambda_\theta \mathbf{Q}_\theta^T \mathbf{S}_\theta \\ \mathbf{B}_\theta^{1/2} &= \mathbf{S}_\theta \mathbf{Q}_\theta \Lambda_\theta^{1/2}. \end{aligned} \quad (4.55)$$

This matrix $\mathbf{B}_\theta^{1/2}$ becomes of reduced size, when only the main patterns of the correlation are considered. This can be done by only considering the m columns for $\mathbf{B}_\theta^{1/2}$ that correspond to the largest m eigenvalues of \mathbf{B}_θ , and omit the other columns. A common criterion to choose the number of columns m is to choose the largest m eigenvalues $\lambda_1, \dots, \lambda_m$ such that:

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^{N_{MF}} \lambda_i} \approx 0.95. \quad (4.56)$$

This represents that approximately 95% of the energy of \mathbf{B}_θ is contained when retrieving \mathbf{B}_θ from the reduced matrix $\mathbf{B}_\theta^{1/2}$ as in equation (4.52). It appears that for highly correlated parameter vectors, the

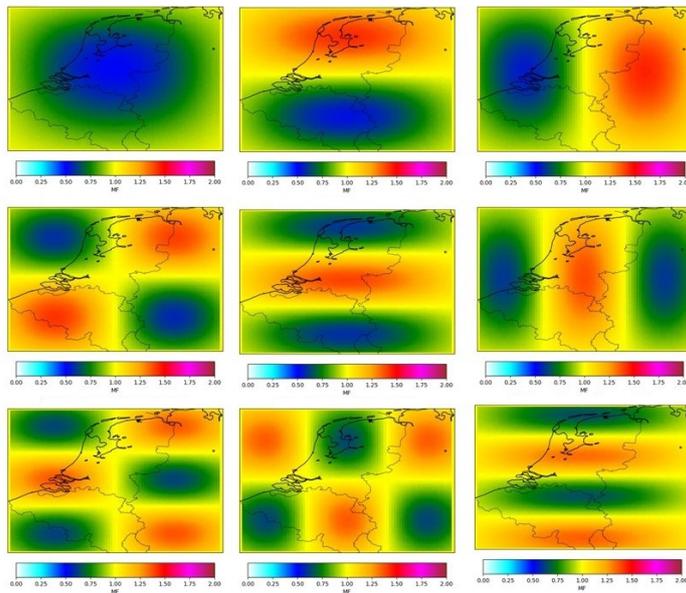
values of the eigenvalues decrease very rapidly, and hence, m is relatively small. For almost independent parameter vectors, most eigenvalues are close to the average value, resulting in a large value for m .

In the 4D-Var cost function (4.49) and its gradient (4.49) for the log-normal multiplication factors γ , not only the reduced matrix $\mathbf{B}^{1/2}$, but also a reduced version of its inverse $\mathbf{B}^{-1/2}$ is used. Using the properties of the eigenvalue decomposition leads to a simple formulation for the inverse of $\mathbf{B}^{1/2}$:

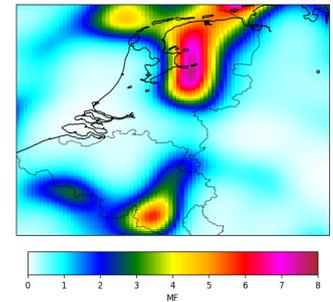
$$\mathbf{B}^{-1/2} = (\mathbf{S}_\theta \mathbf{Q}_\theta \Lambda_\theta^{1/2})^{-1} = \Lambda_\theta^{-1/2} \mathbf{Q}_\theta^T \mathbf{S}_\theta^{-1}. \quad (4.57)$$

This holds, since \mathbf{Q}_θ is an orthogonal matrix and hence $\mathbf{Q}_\theta^{-1} = \mathbf{Q}_\theta^T$. Computing the inverse of \mathbf{S}_θ^{-1} and $\Lambda_\theta^{-1/2}$ can be done analytically as the matrices are diagonal.

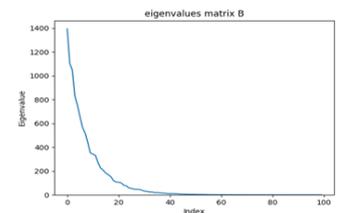
To get a better understanding of what the eigenvalue decomposition and the preconditioner represent, some visualizations are made, using the problem definition of the third experiment described in section 5.4. In the experiment, multiplication factors are defined for each grid cell in the domain. They are correlated in space, according to equation (5.9). An example of such an MF field is shown in Figure 4.4b. The columns of $\mathbf{B}^{1/2}$ corresponding to the largest eigenvalues can be determined. The eigenvectors are visualized in Figure 4.4a. Each eigenvector represents a characteristic pattern of MF's, for example, lower in just the center (top left), or higher in the north and lower in the south (top middle), etc. The reverse transformation of the 4D-Var method in equation (4.31) for the Gaussian method and the exponent of the reverse transformation in equation (4.46) for the log-normal case show that the multiplication factors in Figure 4.4b are approximated by adding a linear combination of these characteristic patterns to the background parameter vector $\mathbf{1}$. The preconditioned state \mathbf{w} holds the coefficients of this linear combination and contains positive and negative weights that should be applied to the individual patterns. When more patterns are considered, the MF field in Figure 4.4b can be approximated more closely. However, it also takes time to run the LOTOS-EUROS model to determine how a change in emissions would influence the measurement. In Figure 4.4c, a plot is made with the largest 100 eigenvalues in consecutive order from large to small. It can be seen that there is a small number of large eigenvalues, and they converge rapidly to zero. In this case, the first 40 eigenvalues contain 93% of the energy of the system. In the experiment $m = 40$ is used, reducing the size of the parameter vector and hence the number of LOTOS-EUROS model runs from 8640 to 40 per day.



(a) Columns of $\mathbf{B}^{1/2}$ corresponding to the largest 9 eigenvalues



(b) Example of MF field



(c) Largest 100 eigenvalues in order from large to small

Figure 4.4: Illustration of preconditioning using dominant patterns.

4.7. Maximum likelihood method

To use one of the 4D-Var methods, the probability distributions of the parameter vector and of the observation likelihood have to be determined. For the prior distribution of the parameter vector, the following two cases were considered:

$$\beta - \beta^b \sim \mathcal{N}(\mu_\beta, \mathbf{B}_\beta), \quad \text{or} \quad (4.58)$$

$$\gamma - \gamma^b \sim \text{LogN}(\mu_\gamma, \mathbf{B}_\gamma). \quad (4.59)$$

For the likelihood of the observation, a Gaussian distribution was assumed:

$$\mathbf{y}^o - \mathcal{H}(\mathbf{x}^{true}, \theta) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \quad (4.60)$$

The means μ and the covariances \mathbf{B} and \mathbf{R} of these distribution need to be defined. In practice these vectors and matrices are formulated using a limited number of parameters, for example a certain constant value or a characteristic length scale. In this section, a maximum likelihood method will be described to estimate these parameters.

In this chapter, first the general objective of the maximum likelihood method will be described. Secondly, the objective of the maximum likelihood will be made suitable for the probability density functions considered in this study. Lastly, the process of finding the parameter that maximizes the maximum likelihood function will be described in section 4.7.3. This chapter is based on the work of (Hausaire 2020).

4.7.1. Objective of maximum likelihood method

The concept of the maximum likelihood method is to find a set of parameters describing a stochastic distribution such that a set of observations \mathbf{y} is most probable. In this study, this set of parameters to be estimated define the mean μ_θ and covariance matrices \mathbf{B}_θ and \mathbf{R} . The optimal mean and covariance matrices are denoted by $\hat{\mu}_\theta$, $\hat{\mathbf{B}}_\theta$ and $\hat{\mathbf{R}}$. The objective of the maximum likelihood method is to estimate these parameters, and to subsequently use them in the 4D-Var method. The maximum likelihood problem can mathematically be formulated as:

Find $\hat{\mu}_\theta$, $\hat{\mathbf{B}}_\theta$, and $\hat{\mathbf{R}}$, such that

$$\hat{\mu}_\theta, \hat{\mathbf{B}}_\theta, \hat{\mathbf{R}} = \underset{(\mu_\theta, \mathbf{B}_\theta, \mathbf{R})}{\text{argmax}} p(\mu_\theta, \mathbf{B}_\theta, \mathbf{R} | \mathbf{y}). \quad (4.61)$$

Now, using Bayes theorem for continuous random variables, the problem (4.61) becomes equivalent to:

$$\hat{\mu}_\theta, \hat{\mathbf{B}}_\theta, \hat{\mathbf{R}} = \underset{(\mu_\theta, \mathbf{B}_\theta, \mathbf{R})}{\text{argmax}} \frac{p(\mathbf{y} | \mu_\theta, \mathbf{B}_\theta, \mathbf{R}) p(\mu_\theta, \mathbf{B}_\theta, \mathbf{R})}{p(\mathbf{y})} = \underset{(\mu_\theta, \mathbf{B}_\theta, \mathbf{R})}{\text{argmax}} p(\mathbf{y} | \mu_\theta, \mathbf{B}_\theta, \mathbf{R}) p(\mu_\theta) p(\mathbf{B}_\theta) p(\mathbf{R}). \quad (4.62)$$

For μ_θ a Gaussian distribution is assumed in order to bound their values:

$$\mu_\theta \sim \mathcal{N}(0, \mathbf{M})$$

Without this assumption, the estimate for μ_θ can become unrealistically large in regions where the effect on the measurements \mathbf{y} is very small. In this study, the the value $\mathbf{M} = \mathbf{I}$ has been chosen. The probability density function $p(\mu_\theta)$ is then known. For the covariances \mathbf{B}_θ and \mathbf{R} , all values are equally likely, such that $p(\mathbf{B}_\theta)$ and $p(\mathbf{R})$ are constants. Then, the terms $p(\mathbf{B}_\theta)$ and $p(\mathbf{R})$ can be omitted from equation (4.62). Now, if the distribution $p(\mathbf{y} | \mu_\theta, \mathbf{B}_\theta, \mathbf{R})$ is known, the optimal values for μ_θ , \mathbf{B}_θ , and \mathbf{R} can be determined.

It is cumbersome to let the maximum likelihood method determine μ_θ , \mathbf{B}_θ , and \mathbf{R} entirely. The vector μ_θ , and matrices \mathbf{B}_θ and \mathbf{R} become large if the number of parameters of θ is large. Additionally, the matrices \mathbf{B}_θ and \mathbf{R} are covariance matrices and therefore have to satisfy many constraints. Instead, the mean μ_θ and covariances \mathbf{B}_θ and \mathbf{R} are parameterized using a relatively small number of parameters. In chapter 5, the vector μ_θ and matrices \mathbf{B}_θ and \mathbf{R} are defined for several problem statements. In those definitions, the vector μ_θ and matrices \mathbf{B}_θ and \mathbf{R} will depend only on a small number of scalar parameters. The maximum likelihood method is used to find these scalar parameters.

4.7.2. Determine the distribution of observations, given the parameters

In this section, the distribution of the observations \mathbf{y} , given the parameters μ_θ , \mathbf{B}_θ , and \mathbf{R} is determined. Once this distribution is known, the probability density function $p(\mathbf{y}|\mu_\theta, \mathbf{B}_\theta, \mathbf{R})$ can be substituted in problem (4.62).

First of all, the case is considered where the multiplication factors are from a Gaussian distribution as in equation (4.58), with the residual distributed as in equation (4.60). When using that the true state \mathbf{x}^{true} is approximated by equation (4.12) for the Gaussian multiplication factors resulting in the observation operator as in equation (4.15), an expression for $\mathbf{y} - \mathbf{H}\mathbf{x}^b$ can be retrieved. For the Gaussian multiplication factors this becomes:

$$\begin{aligned}\mathbf{y} - \mathbf{H}\mathbf{x}^b &= \mathbf{y} - \mathbf{H}(\mathbf{x}^{true} - [\mathbf{E}_\beta](\beta^{true} - \beta^b)) \\ &= (\mathbf{y} - \mathbf{H}\mathbf{x}^{true}) + \mathbf{H}[\mathbf{E}_\beta](\beta^{true} - \beta^b).\end{aligned}\quad (4.63)$$

This is an affine transformation of two normally distributed vectors $(\mathbf{y} - \mathbf{H}\mathbf{x}^{true})$ and $(\beta^{true} - \beta^b)$ of which the distributions are known. Hence:

$$\begin{aligned}\mathbf{y} - \mathbf{H}\mathbf{x}^b &\sim \mathcal{N}(\mathbf{H}[\mathbf{E}_\beta]\mu_\beta, \mathbf{R} + (\mathbf{H}[\mathbf{E}_\beta])\mathbf{B}_\beta(\mathbf{H}[\mathbf{E}_\beta])^T) \\ \mathbf{y} &\sim \mathcal{N}(\mathbf{H}\mathbf{x}^b + \mathbf{H}[\mathbf{E}_\beta]\mu_\beta, \mathbf{R} + (\mathbf{H}[\mathbf{E}_\beta])\mathbf{B}_\beta(\mathbf{H}[\mathbf{E}_\beta])^T)\end{aligned}\quad (4.64)$$

Now, given μ_β , \mathbf{B}_β , and \mathbf{R} , the probability density function of \mathbf{y} can be determined.

Alternatively, for the log-normally multiplication factors γ as in equation (4.59), with auxiliary parameter $\beta = \log(\gamma)$, for which

$$\beta - \beta^b \sim \mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma), \quad (4.65)$$

the linear approximate state of equation (4.25) and the observation operator of equation (4.26) are used. For the log-normally distributed MF's, the residual vector can be expressed as:

$$\begin{aligned}\mathbf{y} - \mathbf{H}\mathbf{x}^b &= \mathbf{y} - \mathbf{H}(\mathbf{x}^{true} - [\mathbf{E}_\gamma](e^{\beta^{true}} - e^{\beta^b})) \\ &= (\mathbf{y} - \mathbf{H}\mathbf{x}^{true}) + \mathbf{H}[\mathbf{E}_\gamma](e^{\beta^{true}} - e^{\beta^b}).\end{aligned}\quad (4.66)$$

This is no longer a linear expression in Gaussian distributions, as $e^{\beta^{true}}$ has a log-normal distribution. Therefore, the distributions cannot be added easily. The choice has been made to linearize the exponents in the expressions:

$$e^{\beta^{true}} \approx \mathbf{1} + \beta^{true}, \quad (4.67)$$

$$e^{\beta^b} \approx \mathbf{1} + \beta^b. \quad (4.68)$$

Then, equation (4.66) becomes:

$$\mathbf{y} - \mathbf{H}\mathbf{x}^b \approx (\mathbf{y} - \mathbf{H}\mathbf{x}^{true}) + \mathbf{H}[\mathbf{E}_\gamma](\beta^{true} - \beta^b). \quad (4.69)$$

This expression consist of Gaussian parameters and constants only. As $[\mathbf{E}_\gamma] = [\mathbf{E}_\beta]$, this expression is the exact same as for the Gaussian case in equation (4.63). In this case, the distribution of \mathbf{y} is identical to the distribution in equation (4.64). Subsequently, the probability density function $p(\mathbf{y}|\mu_\theta, \mathbf{B}_\theta, \mathbf{R})p(\mu)$ is the same for both the Gaussian case and the log-normal case. This results in a similar objective (4.62) for both the Gaussian and the log-normal problem. As the functions are identical, the parameters that minimizes the objective function are identical as well. Hence, for the optimal parameters it holds that $\hat{\mathbf{B}}_\beta = \hat{\mathbf{B}}_\gamma$, $\hat{\mu}_\beta = \hat{\mu}_\gamma$, while $\hat{\mathbf{R}}$ remains the same since it is independent of the choice for a Gaussian or the log-normal distribution for the multiplication factors. In the remainder of this section, the subscripts for \mathbf{B} and μ will therefore be omitted.

4.7.3. Minimization of the Log-Likelihood function

Now that the probability density functions of $p(\mathbf{y}|\mu, \mathbf{B}, \mathbf{R})$ and $p(\mu)$ are known, as both distributions are Gaussian, the optimization problem (4.62) is defined. The next step is to find the parameters μ , \mathbf{B} , and

\mathbf{R} that maximize the objective.

Instead of maximizing the likelihood function (4.62), one can also minimize the opposite of its logarithm. We use that for a variable θ^{test} with a multivariate normal distribution with mean μ^{test} and covariance Σ^{test} , the opposite of the log-likelihood function, multiplied with two, is:

$$l(\theta^{test}|\mu^{test}, \Sigma^{test}) = (\theta^{test} - \mu^{test})^T [\Sigma^{test}]^{-1} (\theta^{test} - \mu^{test}) + \log |\Sigma^{test}| + C^{test}. \quad (4.70)$$

Similarly, for $\mathbf{y} \sim \mathcal{N}(\mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{E}\mu, \mathbf{R} + (\mathbf{H}\mathbf{E})\mathbf{B}(\mathbf{H}\mathbf{E})^T)$ and $\mu \sim \mathcal{N}(0, \mathbf{M})$, the opposite of the logarithm of likelihood function (4.62), multiplied with two, is

$$l(\mu, \mathbf{B}, \mathbf{R}|\mathbf{y}^o) = (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b - \mathbf{H}\mathbf{E}\mu)^T [\mathbf{R} + (\mathbf{H}\mathbf{E})\mathbf{B}(\mathbf{H}\mathbf{E})^T]^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^b - \mathbf{H}\mathbf{E}\mu) + \log |\mathbf{R} + (\mathbf{H}\mathbf{E})\mathbf{B}(\mathbf{H}\mathbf{E})^T| + \mu^T \mathbf{M}^{-1} \mu + C. \quad (4.71)$$

This is the function that will be minimized. Due to several complicated expressions, such as the logarithm of a determinant, this function cannot be minimized analytically. Therefore, a numerical minimizer is used to find the values for $\hat{\mu}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{R}}$ that minimize the expression.

To minimize the log-likelihood function (4.71), the optimization algorithm L-BFGS-B of the SciPy module in Python is used. The L-BFGS-B method is a numerical minimizer, similar to the BFGS method as described in section 4.5.3. Differences are that the L-BFGS-B uses a limited amount of computer memory (L) and that the method can deal with bounded variables (B). The latter difference, the option to deal with bounded variables, is the reason why the L-BFGS-B is preferred over the ordinary BFGS method to minimize the log-likelihood function. In chapter 5, where the vector and matrices μ , \mathbf{B} , and \mathbf{R} are parameterized, it is decided that some of the to be estimated parameters are standard deviations. Standard deviations are constrained to be positive and the L-BFGS-B minimizer can handle such constraints.

4.8. Reducing computational cost of the ensembles

In the adjoint-free 4D-Var method, an ensemble of runs of the LOTOS-EUROS model is executed. Running this ensemble is the most time-consuming part of the optimization procedure. An important part of this study is to keep the computational cost of running the ensemble low.

A first way to reduce the computational cost is to use the chemically reduced model, described in section 2.5. Before the reduced model can be used, the background run, which determines \mathbf{x}^b , is executed using the full LOTOS-EUROS model. Then, the ensemble members needed to determine the matrix \mathbf{E}_θ are run using the reduced N4-model, for which output values of the full LOTOS-EUROS run are imported. When using the reduced model, the computational time is reduced by a factor of three.

A second reduction of the cost is related to the fact that NH_3 is a relatively short-lived trace gas. After NH_3 is emitted, most of the NH_3 is chemically converted or deposited within a few hours. After two days, the fraction of the emitted NH_3 left in the atmosphere is considered to be negligible. To determine the approximate linear state, the effect of perturbing each daily multiplication factor needs to be determined for the entire time period, but due to the short lifetime, each perturbation only affects the state for three days. To reduce computations, the influence of perturbing one daily MF has not been determined for the entire time period, but only for the day of the increased emission and the subsequent two days. For other days, the contribution to the linear approximate state is assumed to be zero. Now, for each MF in space, only three ensemble members are needed to determine the effect of perturbations for all daily MF's in the time period. The first ensemble member perturbs the MF's on days 1, 4, 7, etc., the second ensemble member perturbs the MF's on days 2, 5, 8, etc., and the third ensemble member perturbs the MF's on days 3, 6, 9, etc. This results in a major reduction in ensemble members needed, especially when longer time periods are considered.

4.9. Summary

This section does summarize the adjoint-free 4D-Var method procedure, used to estimate multiplication factors for NH_3 emissions. In the next chapter, several different experiments following the procedure will be described. The algorithm consist of the following steps:

1. **Problem definition:** The first step is to define the parameter estimation problem by choosing the simulation domain, the grid resolution, and the time range.
2. **Observations:** A choice has to be made which observations \mathbf{y}^o to use. It has to be decided which measurement instruments to use and which observations are of sufficient quality.
3. **Multiplication factors:** The spatial variation of the multiplication factors has to be defined. Also, it should be decided whether to assume a Gaussian or a log-normal prior distribution for the multiplication factors. The corresponding mean μ and the covariance \mathbf{B} of the prior distribution should also be parameterized.
4. **Choose which method to use:** Depending on the size and the prior distribution of the multiplication, one of the four developed adjoint-free 4D-Var methods has to be chosen. An overview of the methods and their properties can be found in Table 4.1.
5. **(For a preconditioned method) Determine the matrix $\mathbf{B}^{1/2}$ of reduced size:** First, the eigenvalue decomposition of \mathbf{B} should be determined. Then, $m \ll N_{MF}$, the number of columns of the reduced matrix $\mathbf{B}^{1/2}$ and the size of the preconditioned state \mathbf{w} , should be chosen.
6. **Calculation of the linear approximate model:** The linear approximate model is determined by performing an ensemble of LOTOS-EUROS model runs. First, a background run with parameters θ^b of the full LOTOS-EUROS model is performed. Subsequently, an ensemble of chemically reduced model runs is used to determine the matrix \mathbf{E}_θ using the finite difference approach. For the standard 4D-Var methods, equation (4.14) is used to determine \mathbf{E}_θ , for which an ensemble of size N_{MF} is needed. For the preconditioned 4D-Var methods, equation (4.36), for which an ensemble of size $m \ll N_{MF}$ is needed. The computational cost of running the ensemble can be reduced, as described in section 4.8.

7. **(For a twin experiment) Generate observations:** In the case of a twin experiment, true multiplication factors θ^{true} should be chosen. These multiplication factors are used to generate synthetic observations, as will be described in section 5.1.1.
8. **Use the maximum likelihood method:** The parameterization of vector μ and matrices \mathbf{B} and \mathbf{R} should depend on a small number of scalar parameters. The maximum likelihood method, described in section 4.7, should be used to estimate these scalar parameters.
9. **Minimize the cost function:** Now that all terms in the 4D-Var cost function are defined, the cost function can be minimized using a gradient-based approach. Depending on the chosen method, the function can be minimized analytically by solving the system in equation (4.19) or equation (4.41), or iteratively with gradient-based a numerical minimizer. When possible, the posterior covariance should be determined to quantify the uncertainty of the optimal parameters.

	How to solve	Additional constraints	Negative MF's	Posterior covariance	Suitable for large parameter vectors	Optimizer used
Gaussian standard	Analytical	no	yes	yes	no	-
Log-normal standard	Numerical	no	no	yes	no	BFGS
Gaussian prec.	Analytical	no	yes	yes	yes	-
Log-normal prec.	Numerical	yes	no	no	yes	SLSQP

Table 4.1: Overview of the differences of the 4D-Var methods

In this study, the computational cost of the method is dominated by step 6, the calculation of the linear approximate model, as an ensemble of LOTOS-EUROS runs is needed. For the third experiment of this study, in section 5.4, it takes about 6 hours to run the full LOTOS-EUROS model from February 25th until April 1st using 8 CPUs of a High Performance Computing (HPC) cluster. Subsequently, for an ensemble member, it takes 12-18 hours to run the chemically reduced LOTOS-EUROS model using 1 CPU of the HPC cluster. Luckily, multiple ensemble members can run simultaneously.

For the standard 4D-Var methods, the additional steps of the method are of negligible cost. However, for the preconditioned 4D-Var methods, there are more time-consuming steps. First, in step 5, the eigenvalue decomposition of matrix B of size $N_{MF} \times N_{MF}$ has to be determined. The computational cost of determining the eigenvalue decomposition increases with order n^3 . Furthermore, in step 9, when solving the 4D-Var method for the log-normal preconditioned method, the preconditioned state needs to satisfy all N_{MF} constraints in equation (4.51). The numerical minimizer needs additional time to find a parameter vector that obeys all the constraints. This makes minimizing the 4D-Var cost function more time-consuming.

In this study, the computational cost for determining the eigenvalue decomposition and minimizing the log-normal preconditioned cost function is still negligible in comparison to running a large LOTOS-EUROS ensemble, but those steps become problematic when the number of multiplication factors increases.

5

Results

In this chapter, three experiments that aim to estimate NH_3 emission multiplication factors will be described and the results will be discussed. The experiments use the data assimilation methodology described in chapter 4. As this study aims to develop a methodology, the focus is on validating the data assimilation methodology and testing the made assumptions. At this stage, the retrieved parameters and corresponding NH_3 emission estimates are of less importance.

First, the general concept of an identical twin experiment will be described in section 5.1. As the real NH_3 emissions and the emission parameters are unknown, it is difficult to validate the obtained optimal result. To still validate the data assimilation methodology, the method will be tested in a twin experiment, before testing the method using the real measurements.

Subsequently, the three experiments and their results will be discussed. Initially, in section 5.2, a simple configuration using only one measurement site will be used. The performance and limitations of the data assimilation method will be analyzed. Based on the limitations of the method, a more elaborate second configuration will be used in section 5.3. In this experiment, multiple measuring sites will be considered. Also, spatially varying multiplication factors will be introduced. Finally, in section 5.4, the most extensive but also most realistic experiment will be performed. Here, the IRS observations will be used to estimate multiplication factors on a much finer spatial resolution.

5.1. Concept of twin experiment

The aim of this study is to develop a data assimilation method to estimate the NH_3 emissions in the Netherlands using observations. In this real-life problem, a lot of uncertainty is present. First, it is unknown what the true NH_3 emissions have been for some selected period. Therefore, it is difficult to validate whether the estimated emissions resemble the true emissions or not. Second, if it would be known that there are differences between the true and estimated NH_3 emissions, it is not known if this is caused by errors or limitations of the data assimilation method or something else. Alternative, these errors could be caused because the LOTOS-EUROS model is not a perfect representation of reality (even with perfect emission inventories), or that the described experiment is a simplification of reality.

However, this study aims to find the accuracy and limitations of the data assimilation method. To do so, it would be desired to know the true NH_3 emissions and to know that all errors in the estimates are caused by the limitations of the 4D-Var method. This is not possible in reality, but a hypothetical case can be considered. This hypothetical test scenario is considered in a twin experiment. A twin experiment can be seen as a best-case scenario. If the 4D-Var method already makes errors in the twin experiment, it will definitely make errors for the real-life case when many more uncertainties are present.

5.1.1. Twin experiment procedure

The concept of the twin experiment is to generate self made synthetic observations, instead of using the real observations. In this study, the self made observations are based on concentrations determined by the LOTOS-EUROS model \mathbf{x}^{true} for known emission parameters θ^{true} . The run to determine \mathbf{x}^{true} is called the nature-run. The synthetic observations are determined by the observation operator $\mathbf{y}_k^{sym} = \mathcal{H}(\mathbf{x}^{true}, \theta^{true})$. Measurement errors can be added to the synthetic observations. Generally, this noise is a sample of the assumed error distribution $\epsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. The 4D-Var method is used to retrieve estimates for the emission parameters $\hat{\theta}$, using the self made observations \mathbf{y}_k^{sym} , instead of the true observations. It can be analyzed whether the 4D-Var method estimates emission parameters $\hat{\theta}$ that are close to the known, 'true' emission parameters θ^{true} . To quantify the uncertainty of the optimal solution $\hat{\theta}$, the posterior covariance of the optimized parameters $\hat{\theta}$ is sometimes available as part of the optimizer output (see Table 4.1). If the adjoint-free 4D-Var method works well, most of the parameters θ^{true} should be contained in the 95% confidence interval of $\hat{\theta}$.

This identical twin experiment is not a proper test for determining how well the method would work when real-life observations are used, as the measurements are too perfect. Most uncertainties visualized in Figure 4.2 are neglected. A more realistic way to test how well the 4D-Var method works is to generate the simulated synthetic observations \mathbf{y}_k^{sym} with a different chemical transport model. In this case, the optimized parameter vector $\hat{\theta}$ and the true parameter vector θ^{true} are both known and can be compared when more uncertainties are considered. However, this study does not perform such experiments, and they are left for future study.

5.2. Experiment 1: LML observations Zegveld

5.2.1. Problem description

For the first and simplest experiment, one single LML station in Zegveld is used to optimize emissions. The station is located at 4.84°E, 52.14°N. The station is represented by a black dot in Figure 5.1. The station Zegveld has been chosen for this experiment, as a sufficient amount of measuring data was available, and because of its location. The station Zegveld is located in a region where there is both agriculture nearby, and influence from the industrial areas of the Netherlands.

The grid used for this problem is a relatively small grid, centered around the LML station Zegveld. The grid extends from 4.35°E to 5.35°E and from 51.65°N to 52.65°N, where the resolution is 0.05° longitude and 0.025° latitude. This is equivalent to cells of approximately 3.4 km × 2.8 km. A visualization of the horizontal domain with grid cells can be found in Figure 5.1. The black dot represents the location of the LML station. Vertically, the domain is divided into 12 vertical layers. The first experiment is performed for the months March to May of 2017. The LOTOS-EUROS model has been run, starting on February 25th to minimize the effects of the initial state. All ammonia concentrations in this section are given in the unit parts per billion (ppb).

The objectives of this first experiment are:

1. to test if the linear approximation of state is accurate enough to approximate the true perturbed concentrations of the nature-run;
2. to test if the assumption that the influence of a perturbation can be neglected after two days;
3. to quantify the uncertainty in the optimal result: is one measuring station enough to estimate the surrounding emissions?
4. to analyze what the differences in the result are between the Gaussian 4D-Var method and the log-normal 4D-Var method;
5. to analyze if the retrieved MFs seem realistic when the 4D-Var method uses the true LML observations;
6. to analyze what type of distribution the multiplication factors obtain when the 4D-Var method uses the true LML observations.

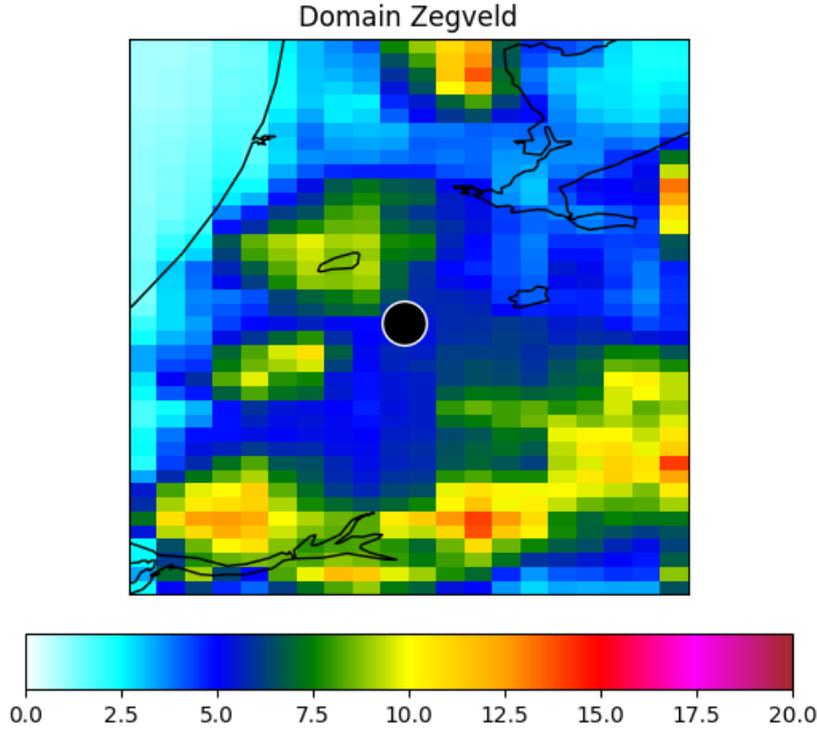


Figure 5.1: LOTOS-EUROS domain around Zegveld. The black dot represents the location of the LML-station of Zegveld. The colors represent the ammonia surface concentrations (ppb).

5.2.2. Parameterization

In this first experiment, there is one multiplication factor for each day. This multiplication factor is constant over the entire domain. This results in a total of 92 multiplication factors, as there are 92 days in March, April, and May. There are two cases considered: the case where the multiplication factors have a Gaussian distribution, $\beta \sim \mathcal{N}(\mu_\beta, \mathbf{B}_\beta)$, and the case where the multiplication factors have a log-normal distribution, $\gamma \sim \text{LogN}(\mu_\gamma, \mathbf{B}_\gamma)$, with corresponding auxiliary parameter $\log(\gamma) = \beta \sim \mathcal{N}(\mu_\beta, \mathbf{B}_\beta)$. As was derived in the section 4.7, the maximum likelihood method estimates the same values for the parameter μ_β as for μ_γ and the same values for \mathbf{B}_β as for \mathbf{B}_γ , independent of whether a Gaussian or a log-normal distribution is assumed. Therefore, the parameters are defined identically. The parameters will be denoted as μ , and \mathbf{B} .

For this experiment, the following assumptions are made. First, it is assumed that the multiplication factors are independent in time which implies that the covariance matrix \mathbf{B} is diagonal. Second, the distribution of $\beta_k - \beta_k^b$ is assumed to be the same for all days k . This means that for all multiplication factors, $\beta_k - \beta_k^b$ have the same expected value μ and that matrix \mathbf{B} has the same variance σ_B^2 on its diagonal entries $\mathbf{B}_{k,k}$ for each day k . Mathematically, this is formulated as:

$$\mu = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \mathbf{B} = \sigma_B^2 \mathbf{1}_{d \times d}. \quad (5.1)$$

Here σ_B represents the standard deviation of $\beta - \beta^b$. The parameter d denotes the number of days for which a multiplication factor is determined. In this case $d = 92$, the number of days in March, April, and May.

The residuals at time k are assumed to be normally distributed:

$$\mathbf{y}_k^o - \mathbf{H}_k \hat{\mathbf{x}}_k \sim \mathcal{N}(0, \mathbf{R}_k)$$

The error covariance \mathbf{R}_k is assumed to consist of two terms that cause uncertainty: the observation error of the LML system, $\mathbf{R}_{obs,k}$ and a representation error, $\mathbf{R}_{repr,k}$. The representation error denotes the uncertainty in the simulation of the measurement from a model state, due to the fact that the measurements have been carried out at one location, whereas the LOTOS-EUROS model gives averages concentrations for the grid cell. Local sources can influence the measurements leading to a larger representation error. In this experiment, there is only one measuring station considered, and therefore $\mathbf{R}_{obs,k}$ and $\mathbf{R}_{repr,k}$ (and therefore \mathbf{R}_k) are all scalar positive numbers. It is assumed that the error covariances are independent of each other in time.

The standard deviation of the observation errors of the LML system ($\sqrt{\mathbf{R}_{obs,k}}$) are assumed to consist of a constant term for the measuring system and a term that is a fraction of the measurements:

$$\mathbf{R}_{obs,k} = (0.57 + r_{obs}y_k^0)^2.$$

The constant term of the standard deviation is approximately $0.4 \mu\text{g}/\text{m}^3 = 0.57\text{ppb}$ (Dammers, Schaap, et al. 2017). The second term, the fraction of the measurements, implies that if the ammonia concentrations are higher, the uncertainty in the measurements gets larger as well. This fraction is denoted by r_{obs} and is chosen to be 0.025, meaning that the standard deviation at time k is increased by 2.5% of the measured concentration y_k^0 at time k . The variance, $\mathbf{R}_{obs,k}$, is the square of the standard deviation. The representation error covariance $\mathbf{R}_{repr,k}$ is chosen to be constant in time, with standard deviation r_{repr} :

$$\mathbf{R}_{repr} = r_{repr}^2. \quad (5.2)$$

Alternative choices for the representation could be made, such as letting the representation error depend on:

- observed concentrations (large influence of local sources results generally in large observed concentrations);
- wind direction and wind speed (larger representation error for low wind speeds with hardly mixing, or larger representation error when ammonia from a particular local source is blown towards the measuring system);
- the spatial gradient of the ammonia concentrations in the LOTOS-EUROS model (larger representation error if the concentrations vary much in the area);
- the variability of the observations in time (on certain days the measurements fluctuating between 50 and 200 ppb in a few hours; this is likely to happen due to local sources).

However, it is very complicated to quantify to which extend these factors do influence the representation error as there is a lot of uncertainty involved. Especially if one would consider many observation sites, it would be cumbersome to determine such a factorization for each site separately. Therefore, the choice has been made to keep the problem simple, and assume r_{repr} to be a constant.

Adding the observation error covariance and the representation error covariance results in the total error covariance:

$$\mathbf{R}_k = \mathbf{R}_{obs,k} + \mathbf{R}_{repr} = (0.57 + r_{obs}y_k^0)^2 + r_{repr}^2. \quad (5.3)$$

In the parametrizations in equation (5.1) and equation (5.3), there are three parameters that have not yet been defined: μ , σ_B , and r_{repr} . These three numbers are determined using the maximum likelihood method as described in section 4.7. The retrieved values are $\sigma_B = 4.16$, $\mu = 0.14$, and $r_{repr} = 10.42\text{ppb}$. As $\mu = 0.14$, it is implied that on average all emissions should be increased by 14% to make the LOTOS-EUROS model better match the observations. This is considered to be an important change, motivating the decision to assume the nonzero μ .

5.2.3. Results

Now that the first experiment is defined, the optimal multiplication factors can be retrieved using the 4D-Var method. The 4D-Var method is described in section 4.4 for the multiplication factors with a Gaussian distribution, and in section 4.5 for the multiplication factors with a log-normal distribution. First, the 4D-Var method is tested in a twin experiment, to test the first four objectives of this experiment. For these

first three objectives, it is convenient to know the true values of the concentrations and the multiplication factors. Second, the 4D-Var method is used to estimate the MFs based on the true LML measurements. These results are used to again comment on the third and fourth objective, and additionally discuss the fifth objective. Last, the posterior distribution of the optimal MFs is analyzed, as needed for the sixth objective.

5.2.3.1. Results twin experiment

First of all, we focus on objectives 1 and 2. If the 4D-Var method uses very noisy observations, it becomes difficult to determine which extend errors for the estimated MFs are caused by the assumed simplifications of objectives 1 and 2 or whether errors are caused by noise in the observations. Keeping this in mind, the 4D-Var method is first tested using the synthetic observations from the nature-run with no noise added. For this test, only the case where the multiplication factors have a Gaussian prior distribution is considered.

Results of the twin experiment, without added noise, can be found in Figure 5.2. The parameter values in the nature-run are arbitrarily chosen. Figure 5.2a contains time series of the true multiplication factors (dashed line) and the estimated multiplication factors (solid line). Figure 5.2b contains the corresponding NH_3 concentrations of the background run (orange), the optimal estimation of the concentrations (green), and the synthetic observations of the nature-run (blue).

In Figure 5.2a, it becomes clear that the 4D-Var method is able to reproduce the true MFs very well for March and the beginning of April. The adjoint-free 4D-Var method, using the linear approximate state, does result in an accurate estimation of the peak where the true MFs have values around 3 at the end of March and when the true MF decreases to 2.2. The true and optimized MFs, as well as the observations and the optimized NH_3 concentrations, are initially very similar. This indicates that the linear approximation of the NH_3 concentrations is accurate enough to approximate the true perturbed concentrations of the nature-run, and that initially the assumption holds that the influence of a perturbation can be neglected after two days.

However, at the end of April and in May, differences start to occur. Especially the part in May where the true MFs are one is interesting, as the emissions in the nature-run are identical to those of background run. It was assumed that influences of perturbations on NH_3 concentrations can be neglected after two days, which would result in similar NH_3 concentrations in the observations and the background run in May. However, if one looks at the NH_3 concentrations of the observations in May in Figure 5.2b, it appears that the observations are higher than the background run. The cause for those increased concentrations is an increase in dry emission. Due to higher NH_3 concentrations in March and April, vegetation and water surfaces have absorbed more ammonia, leading to higher dry emissions in May. Hence, looking at a period of several months the assumption that the influence of a perturbation can be neglected after two days does not hold. Instead of estimating the MFs to be 1 in May, the 4D-Var method incorrectly overestimates the anthropogenic emissions to compensate for the unmodeled increased dry emissions.

Now that the error due to the linear approximation of the state and due to neglecting the influence of perturbations on NH_3 concentrations after two days are analyzed, the additional error due to the uncertainty in the observations can be analyzed. Also, the Gaussian 4D-var method and log-normal 4D-Var method will be compared. In Figure 5.3a and 5.3b, time series of the estimated multiplication factors are shown. Figure 5.3a contains the result for the Gaussian multiplication factors and Figure 5.3b the results for the log-normal multiplication factors. Figure 5.3c and 5.3d contain the corresponding NH_3 concentrations of the background run (orange), the optimal estimation of the concentrations (green), and the synthetic observations of the nature run (blue).

It can be seen in Figure 5.3a and 5.3b that for both methods the estimated MFs (solid line), on average, follow the true MFs (dashed line), but the estimates are very noisy, especially in comparison with the results without noise in Figure 5.2. For the first 16 days, the emissions in both the nature-run and the background run are identical, so the true ammonia concentrations should be similar. However, as noise is added to the simulated observations, differences in concentration, and subsequently differences between the true and the estimated MFs occur, leading to inaccurate results. Hence, the noise

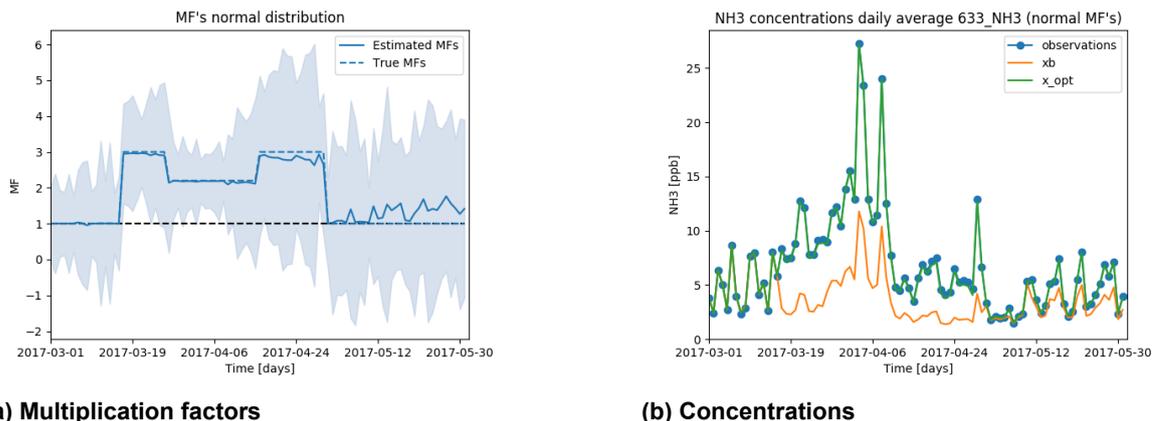


Figure 5.2: Multiplication factors and concentrations in time for the nature run when no noise is added to the observations. The case where the multiplication factors have a Gaussian distribution is considered.

in the observations results in significant errors in the estimated MFs.

When comparing the Gaussian and the log-normal method, it can be seen that both methods retrieve similar optimized multiplication factors. A noteworthy difference between the two methods is that for May 8th, the estimated MF is approximately -1 for the Gaussian case, whereas it is close to zero in the log-normal case. A multiplication factor of -1 would imply that all anthropogenic emission sources would become sinks for one day, which is unlikely as this can only be caused by deposition. The negative MFs cause ammonia concentrations to decrease or even become negative. In real life, it can happen that some NH_3 observations are negative due to measurement errors, but this should not be modeled, as concentrations can only be positive. In Figure 5.3c, the observed and estimated concentrations are indeed negative for May 8th. In the log-normal case, the estimated concentrations on May 8th in Figure 5.3d, remain positive, even when the observed concentrations are negative.

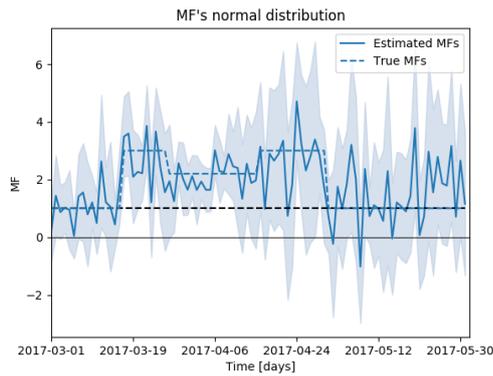
The major difference between the two methods is the 95% uncertainty interval. The Gaussian method has a relatively smaller confidence region, which sometimes contains negative MFs. The log-normal method does not allow MFs to be negative but does contain much higher values in its confidence interval, up to almost 25, whereas the range of the Gaussian method does not go higher than 7. For both cases, the true MFs, represented by the dashed line, are always contained in the 95% confidence interval of the estimated multiplication factors.

5.2.3.2. Results real LML data

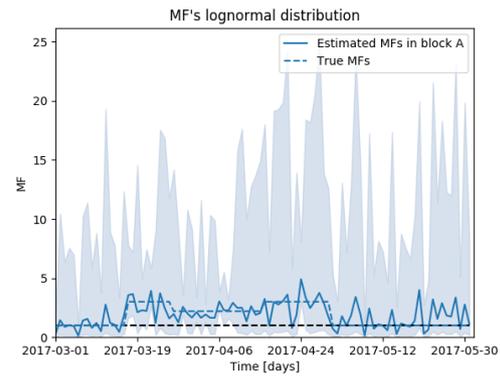
In this section, the results for the first problem statement using the actual measurements of the LML station in Zegveld will be discussed. The aim of this test is to look into the uncertainty of the optimal results, analyze the differences between the Gaussian and the log-normal method, and analyze if the results seem realistic. In this realistic experiment, it is unknown what the true emissions are, which makes it difficult to verify the obtained results.

In Figure 5.4, the time series of the estimated MFs are plotted. It can be seen that the Gaussian method and the log-normal method give quite similar results for the optimal MFs. Both methods allow large peaks to occur, but the MFs for the log-normal method can be slightly higher than for the Gaussian method (16 instead of 15). As before, the two methods give a different 95% uncertainty interval. This 95% uncertainty interval can be very different per day. For some days, for example at the end of March, the results are relatively accurate, but for others, such as in mid-April, the results are very uncertain. This uncertainty in April is due to the fact that there are no observations available for these days.

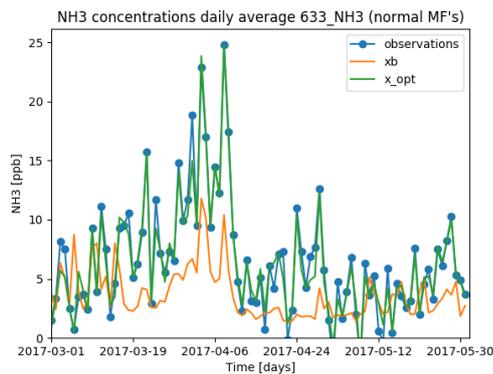
Overall, the results do not seem to be very realistic. It is for example remarkable that the multiplication factors can obtain values up to 15. It might be unrealistic that the emissions inventories contain



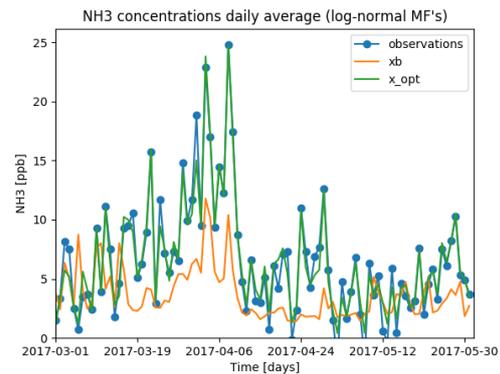
(a) Multiplication factors β with a Gaussian prior distribution.



(b) Multiplication factors γ with a log-normal prior distribution.



(c) Concentrations when the multiplication factors β have a Gaussian prior distribution.



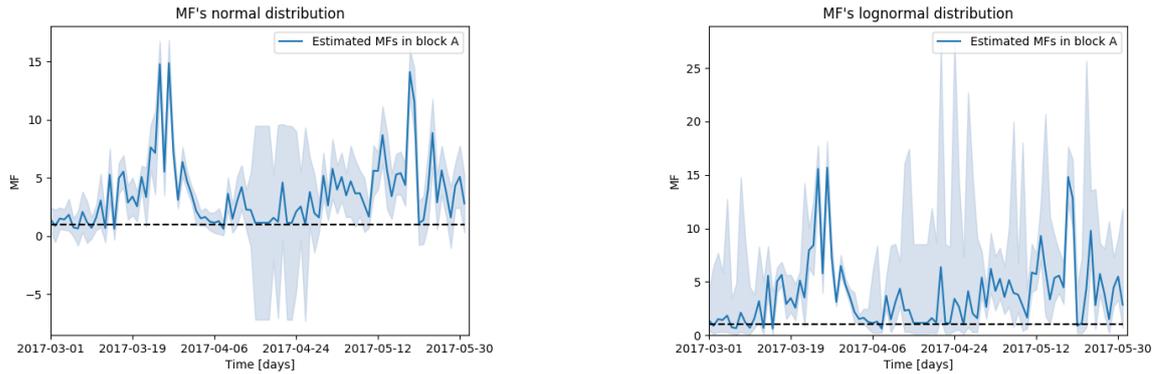
(d) Concentrations when the multiplication factors γ have a log-normal prior distribution.

Figure 5.3: Figure (a) and (b) contain the multiplication factors in time for the nature run. Figure (c) and (d) contain the daily average ammonia concentrations at the location of the LML station in Zegveld for the nature run.

such large errors for multiple days. Also, the MFs differ very much from day to day, which might not be realistic. In fact, if manure and fertilizer are applied, it should increase the emissions in a region for several consecutive days. Especially the peak at the end of May, where the MF goes from 13 to 1 in one day seems unrealistic. As the results seem unrealistic, a more elaborate problem statement needs to be defined. Recommendations for a next experiment will be stated in section 5.2.4, which will be incorporated in the second experiment in section 5.3.

5.2.3.3. Distribution of multiplication factors

The last objective was to determine which distribution does characterize the uncertainties in the multiplication factors the best: the Gaussian or the log-normal distribution. To analyze the distributions, histograms have been made of the optimally estimated multiplication factors $\hat{\beta}$ and $\hat{\gamma}$ in Figure 5.5. Additionally, the prior assumed distribution, resulting from the maximum likelihood method, is plotted in blue. In both cases, the histogram does not properly resemble the prior distribution. This implies that the prior distribution determined by the maximum likelihood method might not be a realistic prior distribution of the true multiplication factors. An important thing to note is that in the Gaussian case the mean μ_{β} , and in the log-normal case the median $e^{\mu_{\gamma}}$ of the multiplication factors are much lower for the prior distribution than for the estimated parameters. In this setting, both prior distributions do generally underestimate the MFs. This indicates that a prior distribution with a larger value μ would better approximate the distribution of the MFs. In both plots 5.5a and 5.5b, the histograms do indeed seem to follow some sort of log-normal distribution shape, but with a very different parameterization than retrieved from the maximum likelihood method. The multiplication factors seem to follow a log-

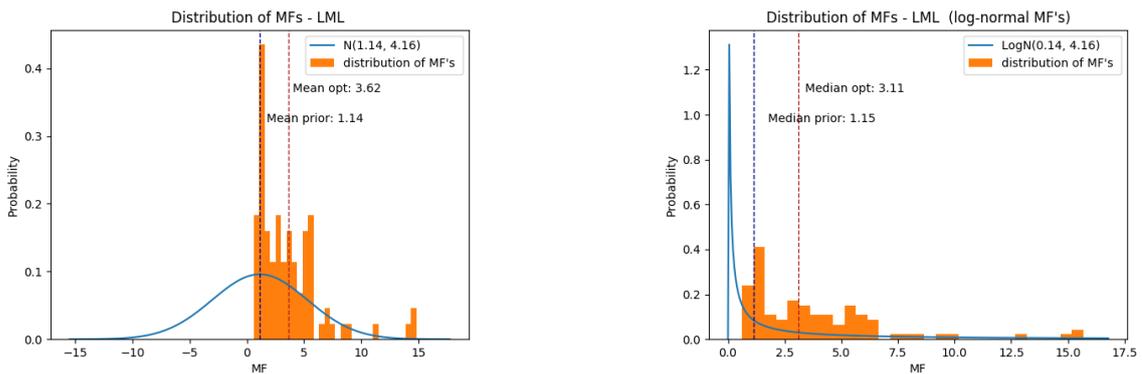


(a) Multiplication factors β with a Gaussian prior distribution.

(b) Multiplication factors γ with a log-normal prior distribution.

Figure 5.4: Multiplication factors in time using the actual LML measurements. The solid line represents the optimal estimate of the multiplication factors $\hat{\beta}$ or $\hat{\gamma}$, and the faded region represents the 95% confidence interval of the estimated multiplication factors.

normal distribution with a larger parameter μ and a smaller covariance **B**. This would result in higher emission estimates, with less uncertainty for the estimated MFs.



(a) Distribution of optimal multiplication factors $\hat{\beta}$ with a Gaussian prior distribution

(b) Distribution of optimal multiplication factors $\hat{\gamma}$ with a log-normal prior distribution

Figure 5.5: Histograms representing the probability density function of the optimal multiplication factors $\hat{\beta}$ and $\hat{\gamma}$ with the prior assumed distribution plotted in blue.

5.2.4. Conclusion

Summarizing, from the twin experiments, it is learned that the linear approximation of the state is sufficiently accurate. Further, the influence of a perturbation can be neglected after two days, if one looks at a short time period, but not if one looks at a time period of several months because of effects on deposition. Furthermore, it was concluded that the results contain a lot of noise. The estimated MFs can differ much from the true multiplication factors in the twin experiment. The Gaussian and log-normal method generate similar MFs, but different uncertainty intervals. Occasionally, the Gaussian formulation can generate negative MFs, which is unrealistic.

The experiment with the real LML data again shows noisy and uncertain estimated MFs, which seem unrealistic. In addition, the experiment showed that the optimized parameters show a distribution that has characteristics similar to the log-normal distribution, although the resemblance to the log-normal distribution parameterized by the maximum likelihood method is not good.

Recommendations, resulting from the current limitations of the model, are to optimize over a shorter time period. If one is interested in longer time periods, it is recommended to first optimize for the first few weeks and run the model for a longer period with the improved emissions for the first few weeks, to get a better estimate for the dry emission in the subsequent time period. Second, it is recommended to add more measurement stations to decrease the uncertainty in the estimates. Third, it is recommended to use the formulation with log-normal prior distribution, based on the log-normal distribution shaped histograms of the optimized MFs, in combination with the property that the log-normal distribution does not allow negative MFs. Last, it would be recommended to use a larger domain. There are multiple days for which the LOTOS-EUROS model does underestimate the measured observations eminently (resulting in MFs of 15). As a relatively small domain is considered in this experiment, it could be the case that relevant emission sources outside of this domain are neglected.

5.3. Experiment 2: LML observations Netherlands

5.3.1. Problem description

As a second experiment, there are four adjustments made based on the recommendations of the previous experiments. First, more LML observation stations are considered. In the spring period of 2017, five LML stations were operating in the Netherlands: Vredepeel, Wekerom, Zegveld, Wieringerwerf, and Valthermond. The stations are represented by black dots in Figure 5.6a.

Second, as recommended, but also because these five stations are spread across the country, a larger domain is considered. The grid used for this problem contains all of the Netherlands. The grid is from 3.15°E to 7.55°E by 50.65°N to 53.70°N, where the grid resolution is again 0.05° longitude and 0.025° latitude. This is equivalent to cells of approximately 3.4 km × 2.8 km. As before, the domain is divided vertically into 12 vertical layers. Figure 5.6a shows a visualization of the horizontal domain with grid cells. The five black dots denote the locations of the LML stations.

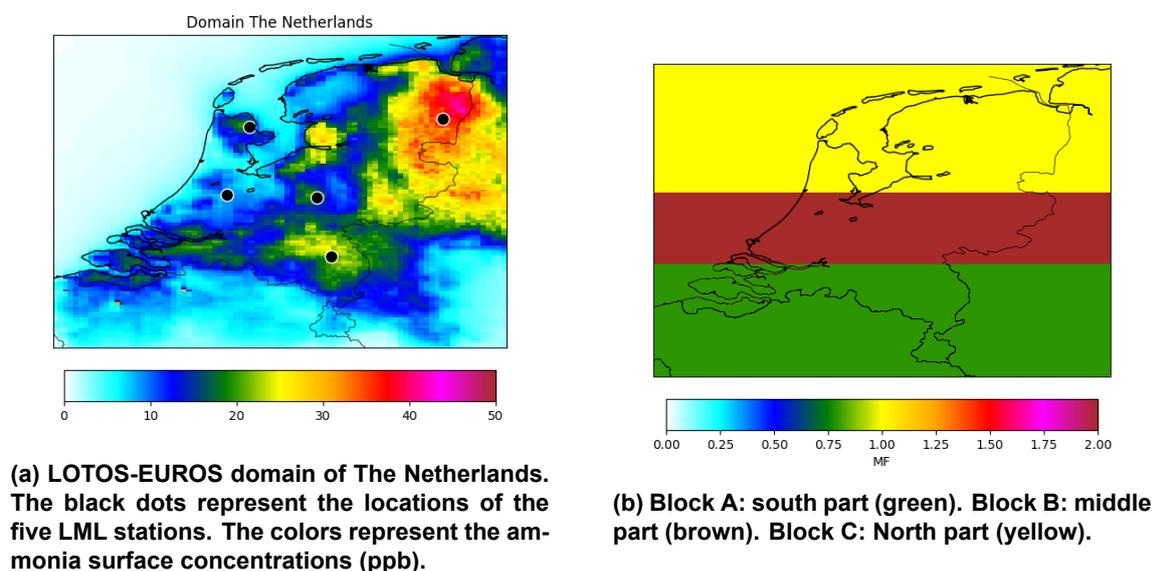


Figure 5.6

Third, this second experiment is performed for the month of March in 2017 only, such that the difference in dry emission due to varied NH_3 emissions remains limited. The LOTOS-EUROS model has been run, starting on February 25th to minimize the effects of the initial state.

The fourth and last modification in this experiment is that the multiplication factors can be different for various parts of the domain. Due to this spatial variation, the emission estimates can be adjusted

for the separate regions.

The objectives of this second experiment are:

1. to test whether it is possible to retrieve spatial variability in the multiplication factors in a twin experiment;
2. to test if the accuracy in the optimal estimates of the multiplication factors are increased once observations of multiple stations are added in a twin experiment;
3. to determine whether spatial variability does occur according to the real LML observations;
4. to test whether realistic and accurate emissions are estimated when using the real LML observations.

5.3.2. Parameterization

In this experiment, the domain is divided into three blocks: block A consisting of the south of the domain, block B consisting of the middle of the domain, and block C consisting of the north of the domain. The blocks are shown in Figure 5.6b. For each block, there is one multiplication factor for each day of March. This results in a total of 93 multiplication factors. In this experiment, only the case where the multiplication factors have a log-normal distribution is considered:

$$\gamma \sim \text{Log}\mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma),$$

with corresponding auxiliary parameter:

$$\beta = \log(\gamma) \sim \mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma)$$

As in the previous experiment, the multiplication factors are assumed to be independent of each other, and the distribution out of which the multiplication factors are a sample is the same for all days k . However, the parameters of the log-normal distribution vary per block. Each block has its own expected value μ_A , μ_B , or μ_C , and variance $\sigma_{B,A}$, $\sigma_{B,B}$, or $\sigma_{B,C}$. Then, the parameter vector μ and covariance matrix \mathbf{B} become:

$$\mu = \begin{bmatrix} \mu_A \\ \vdots \\ \mu_A \\ \mu_B \\ \vdots \\ \mu_C \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \sigma_{B,A} \mathbf{I}_{d \times d} & 0 & 0 \\ 0 & \sigma_{B,B} \mathbf{I}_{d \times d} & 0 \\ 0 & 0 & \sigma_{B,C} \mathbf{I}_{d \times d} \end{bmatrix}. \quad (5.4)$$

Here, the number of days $d = 31$.

The distribution of the residual is similar to the distribution described in the previous experiment. The only difference is that there are now five stations instead of one. The error covariances \mathbf{R}_k , $\mathbf{R}_{obs,k}$, and $\mathbf{R}_{repr,k}$ are now matrices in $\mathbb{R}^{5 \times 5}$. It is assumed that both the observation errors and the representation errors are independent in time and space, i.e. $\mathbf{R}_{obs,k}$ and $\mathbf{R}_{repr,k}$ are diagonal matrices. Let $\mathbf{y}_{k,s}^O$ denote the observation at time instance k at LML station s , then, analogous to the previous experiment:

$$\mathbf{R}_{obs,k} = \begin{bmatrix} (0.57 + r_{obs} \mathbf{y}_{k,1}^O)^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & (0.57 + r_{obs} \mathbf{y}_{k,5}^O)^2 \end{bmatrix}. \quad (5.5)$$

The representation error covariances are again constant in time. For each station s a standard deviation $r_{repr,s}$ can be chosen. The covariance matrix then becomes:

$$\mathbf{R}_{repr} = \begin{bmatrix} (r_{repr,1})^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & (r_{repr,5})^2 \end{bmatrix}. \quad (5.6)$$

The total observation representation error covariance matrix is then:

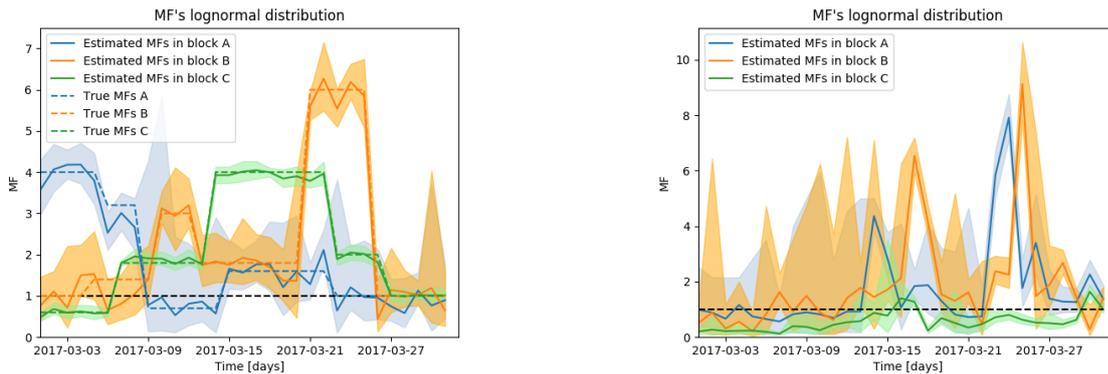
$$\mathbf{R}_k = \mathbf{R}_{obs,k} + \mathbf{R}_{repr} \quad (5.7)$$

If the number of stations is large, it can be a lot of work to estimate the parameters $r_{repr,s}$ for all stations using the maximum likelihood approach of section 4.7. To limit the number of parameters $r_{repr,s}$, it can be assumed that measuring stations in comparable environments (such as urban, agriculture, road, rural) have the same r_{repr} . In this experiment, it is assumed that the stations in the north (Wieringerwerf and Valthermond) have the same r_{repr} and that the stations in the middle (Zegveld and Wekerom) have the same r_{repr} . Hence, there is a total of 3 representation variables r_{repr} ; one for the stations within each block.

The maximum likelihood method gives for block A $\mu_A = 0.184$, $\sigma_{B,A} = 1.686$, and for the measuring station Vredepeel in block A, $r_{repr} = 20.46$. For block B, $\mu_B = 0.191$, $\sigma_{B,B} = 1.764$, and for the measuring stations in block B, Zegveld and Wekerom, $r_{repr} = 12.40$. For block C, $\mu_C = -0.415$, $\sigma_{B,C} = 0.419$, and for the measuring stations in block C, Wieringerwerf and Valthermond, $r_{repr} = 6.268$. It becomes clear that the distribution of both the emission inventories and the representation error of the LML stations vary a lot per block.

5.3.3. Results

The optimal multiplication factors $\hat{\gamma}$ are determined using the 4D-Var method described in section 4.5. In Figure 5.7, the results are plotted. The results for the observations of the twin experiment can be found in Figure 5.7a and the results when using the real LML observations can be found in Figure 5.7b. Time series of the optimal estimated $\hat{\gamma}$ for the different blocks are represented by the solid lines and the time series for the true multiplication factors γ^{true} in the twin experiment is represented by the dotted line in Figure 5.7a. The faded regions represent the corresponding 95% confidence intervals of the estimated multiplication factors.



(a) Multiplication factors γ using synthetic observations from the nature run

(b) Multiplication factors γ using the real observations of all 5 LML stations

Figure 5.7: Multiplication factors for the three blocks in time using observations for all five observation stations. The solid line represents the optimal estimate of the multiplication factors $\hat{\gamma}$, and the faded region represents the 95% confidence interval of the estimated multiplication factors. In case of the nature run, the dashed line represents the true multiplication factors γ^{true} .

5.3.3.1. Results twin experiment

For this twin experiment, the multiplication factors are arbitrarily chosen. In comparison to the nature run in the previous experiment, higher peaks are included. Also, multiplication factors smaller than 1 are included. It can be seen in Figure 5.7a that the estimated MFs do resemble the true MFs quite accurately, especially compared to the results in the previous experiment in Figure 5.3b. The true MFs are generally contained in or close to the boundaries of the 95% percent confidence interval. It can be concluded that the 4D-Var method can retrieve the multiplication factors for the emissions in different blocks, even when the emissions increase for one block and decrease for another block. Hence the 4D-Var method is able to retrieve spatial variability in the multiplication factors. To test this was the first

objective of the experiment.

The second objective of this experiment is related to the accuracy of the optimal estimates. In Figure 5.7a it can be seen that the uncertainty in the optimal estimates is much smaller in this experiment than for the previous twin experiment for Zegveld in Figure 5.3b. This is caused by the increased number of observations from 5 measuring stations, as well as due to the smaller uncertainty in the covariance matrix \mathbf{B} . In the previous experiment with Zegveld only, the maximum likelihood method estimated the standard deviation $\sigma_B = 4.16$, whereas for this experiment the maximum likelihood method estimates much smaller standard deviations: $\sigma_{B,A} = 1.686$, $\sigma_{B,B} = 1.764$, and $\sigma_{B,C} = 0.419$. Especially in block C, where both the uncertainty in the multiplication factor and the uncertainty due to the representation error is small, the method considers the optimized results to be very accurate. It can be concluded that for this experiment with a larger domain, with the multiplication factors determined for three blocks, and with 5 measuring stations, the accuracy of the estimates of the multiplication factors is increased.

5.3.3.2. Results real LML data

In Figure 5.7b, the results are plotted for an experiment using the real LML observations. It appears that the retrieved MFs can be very different for the three blocks on certain days. Generally, the MFs for block A and block B are distributed around one, with a few peaks that occur on different days per block, whereas block C estimates much lower MFs for almost all days. This indicates that the MFs might indeed be different for different parts of the Netherlands. Hence, according to the observations, spatial variability does occur. Allowing the MFs to vary for different regions is therefore considered to be a realistic and important addition to the problem statement.

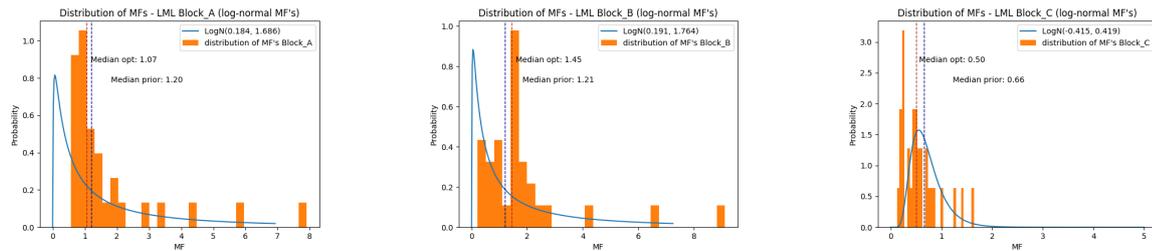
However, this division of the MFs into three blocks is arbitrary and not very realistic. It is for example unlikely that on March 25th exactly all emissions in Block B were underestimated by a factor of 10, whereas all emissions in block C were overestimated by a factor of 2. Especially, if the multiplication factors represent changes in emission due to the application of fertilizer, it is unlikely that the changes in MFs close to the boundary between the blocks B and C are so different. The division of the domain into three blocks is very coarse. A smoother, possibly correlated, parameterization of MFs would be more realistic.

Furthermore, if these results are compared to the results of the previous experiment around Zegveld in Figure 5.4b, it appears that the previously estimated, unrealistically large multiplication factors of 15 at the end of March are no longer estimated. Large peaks in the estimated MFs still occur in Figure 5.7b, but they are less high (now up to 9 instead of 15 before), and they occur on different days. This improvement is made due to the increased number of observation sites since not all stations do measure increased concentrations for the same days. Additionally, using a larger domain that also contains emission sources further away from Zegveld does reduce the peaks in the MFs. When using this larger domain, the modeled NH_3 concentrations at the location of Zegveld in the background run are higher and do already look more like the observations of the LML station. It might be even better to consider a larger domain, containing more emission sources surrounding the Netherlands. Especially the LML-station Valthermond in the northeast is located relatively close to the boundary of the domain, but there are many ammonia emission sources in the surrounding area (in Germany) that are not contained in our domain.

Last, it can be seen that the uncertainty of the estimated MFs in Figure 5.7b is smaller than it was for the previous experiment in Figure 5.4b. However, the uncertainty in the optimal solution can still be very large. Considering 5 LML measurement systems does increase the accuracy of the method, but it would be beneficial if more measurements were available. To get accurate results, the number of observations should be larger than the number of parameters. Especially when the observations contain a lot of noise, more observations are needed. In this experiment, three MF are determined per day, while the five LML stations have several observations for most of the days. When using only the LML stations, the number of parameters to estimate should not be much larger, as otherwise, observational noise will dominate the results.

5.3.3.3. Distribution of multiplication factors

In Figure 5.8, histograms have been made of the optimally estimated multiplication factors for the three blocks, based on the true LML observations. Additionally, the three different prior distributions are plotted in blue as well. For all three blocks A, B, and C, the histograms seem to follow the prior distribution much better than before in Figure 5.5b. For all blocks, the prior median is relatively close to the median of the optimized MFs. Furthermore, the distributions in all three figures seem quite log-normal. For block A and B, there is a small number of high MFs but the majority is close to the median value. For block C, there are no high MFs estimated as expected based on the prior distribution. For block A, the MFs are never close to zero, leading to some discrepancy between the optimized and the prior assumed MFs. However, overall these histograms confirm the assumption that a log-normal prior distribution is a better fit than a normal distribution would be.



(a) Distribution multiplication factors block A

(b) Distribution multiplication factors block B

(c) Distribution multiplication factors block C

Figure 5.8: Distribution MFs for different blocks

5.3.4. Conclusion

From the nature-run of this problem, it can be concluded that the 4D-Var method can estimate the multiplication factors for three blocks, based on observations of 5 LML stations. Even when the emissions are increased for some blocks, and are decreased for others, the method reconstructs the MFs quite well. Also, in comparison with the previous experiment in Zegveld, the uncertainty in the estimated multiplication factors has decreased a lot.

In the experiment with the real data for the five LML stations, it was found that the MFs can indeed differ quite a lot for the three blocks. Unfortunately, the results are still not considered to be very realistic. Also, the accuracy of the method has improved, but the optimized MFs still contain a lot of uncertainty.

Recommendations resulting from this experiment are the following three:

- First, a more realistic spatial variability of the MFs should be made. Now that it is found that the MFs are likely to vary depending on the location, it would be recommended to make this parameterization more realistic. This would include increasing the number of regions/blocks for which the MFs can be varied and possibly also define a spatial correlation between those regions to avoid an MF of 10 in one region and an MF of 0.5 in a neighboring region.
- Second, it would be recommended to increase the number of observation sites, possibly using a different type of measuring instrument such as the IRS instrument. If the number of spatial regions for which the MFs can vary is increased, it will be difficult to measure the effects of all these different MFs in space, using only 5 LML-stations. It would for example be difficult to obtain results about the MFs for Zeeland, using only LML-stations that are more than 50km away. Also, the additional number of measuring sites would help to reduce the still quite large uncertainty in the MF estimates.
- Last of all, an even larger domain, containing more emission sources surrounding the Netherlands should also be considered.

5.4. Experiment 3: IRS observations

5.4.1. Problem description

In this last experiment, the main goal is to improve the spatial variability of the multiplication factors in a realistic manner, as recommended in the previous experiment. The aim is to estimate a multiplication factor for each grid cell in the domain.

The previous experiment recommended increasing the number of observation sites. Instead of using the limited number of measurements of the LML station at only a small number of locations, this experiment uses the observations of the IRS instrument on the MTG-S satellite. In this experiment, it will be tested how well the 4D-Var method works in retrieving multiplication factors for each grid cell, using the observations of the IRS instrument. As the MTG-S satellite is not operational yet, this experiment will only use synthetic observations from a nature-run.

For this problem, a larger domain is considered such that all major sources that influence the Netherlands are included. Therefore, Belgium and a larger part of the west of Germany are included as well. The new grid is from 2.55°E to 8.55°E by 49.65°N to 53.70°N, where the grid resolution is 0.075° longitude and 0.0375° latitude. This is equivalent to cells of approximately 5.1 km × 4.2 km. In total, the grid consists of $80 \times 108 = 8640$ grid cells. A visualization of the horizontal domain with grid cells can be found in Figure 5.9. Vertically, the domain is divided into 12 vertical layers. Observations now consist of NH₃ column densities in units of 1e15 molecules per cm² (1e15 molec/cm²). Synthetic IRS observations are produced using a LOTOS-EUROS nature run. In this experiment, the MFs are determined from March 1st till March 10th. The LOTOS-EUROS model has been run, starting on February 25th to minimize the effects of the initial state.

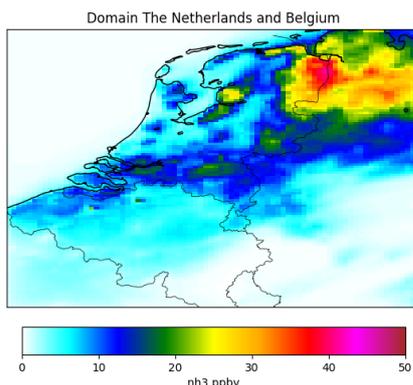


Figure 5.9: LOTOS-EUROS domain Netherlands, Belgium, and west part of Germany.

The objectives of this third experiment are

1. to test to which accuracy can the MFs be estimated.
2. to find out for which locations the 4D-Var method approximates the true emissions well, and for which locations are estimations are inaccurate.

5.4.2. Parameterization

This experiment aims to estimate one multiplication factor per grid cell per day. In this case, the number of parameters to estimate would be 8640 MFs per day × 10 days = 86.400. It is not feasible to run the LOTOS-EUROS model for each parameter separately. Therefore, in this experiment, a preconditioner of reduced size is used, as described in section 4.6. The multiplication factors are assumed to have a log-normal distribution, $\gamma \sim \text{Log}\mathcal{N}(\mu_\gamma, \mathbf{B}_\gamma)$, so the 4D-Var method described in section 4.6.2 is used. The auxiliary parameter to optimize is $\beta = \log(\gamma) \sim \mathcal{N}(\mu_\beta, \mathbf{B}_\beta)$. In this experiment, the first 8640 elements of the parameter vector β represent all MFs in the domain for day 1, the second 8640 elements represent all MFs for day 2, etc. Each grid cell i can have its own mean μ_i , but the mean is constant

for the entire period. Hence:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{8640} \\ \mu_1 \\ \vdots \\ \mu_{8640} \end{bmatrix}. \quad (5.8)$$

To efficiently reduce the size of the preconditioner \mathbf{w} , a correlation between the MFs needs to be defined. In this experiment, it is assumed that the MFs are correlated in space, but are independent in time. For each day, the $80 \times 108 = 8640$ MFs are strongly correlated to the MFs of nearby grid cells, and hardly correlated to the MFs of grid cells far away. For the application of fertilizer, this is a realistic correlation as generally most farmers in a certain region apply their fertilizer at the same time because of similar growing seasons and weather conditions. The correlation between two grid cells i and j for some day k is defined by the relation:

$$\mathbf{C}_{i,j} = \exp\left(-\frac{1}{2} \left(\frac{d(i,j)}{L}\right)^2\right). \quad (5.9)$$

Here $d(i,j)$ denotes the distance between the center of cell i and j and L is the correlation length scale. This correlation is the same as the spatial correlation in (Arjo Segers, Tokaya, and Houweling 2020) for methane emissions. For a larger L , the MFs are stronger correlated. If all these values $\mathbf{C}_{i,j}$ are put into a correlation matrix \mathbf{C}_k for day k , and if a matrix \mathbf{S}_k with a standard deviations on its diagonal for parameter $\beta_{i,k}$ for each grid cell i is defined on day k , then the covariance matrix \mathbf{B} for the parameter vector β of day k is defined as:

$$\mathbf{B}_k = \mathbf{S}_k \mathbf{C}_k \mathbf{S}_k. \quad (5.10)$$

If one looks at the entire parameter vector β with MFs for all days, the covariance matrix is:

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{B}_d \end{pmatrix}. \quad (5.11)$$

It is assumed that \mathbf{B}_k is the same for all days k .

In this section, the observations of the IRS instrument of the MTG-S satellite are used. The error in those measurements are chosen as described in section 3.2.2, where the standard deviation is $5.3e15$ molec/cm² when the measured NH₃ column densities are higher than $1e16$ molec/cm² and the standard deviation is $4.1e15$ molec/cm² when the measured NH₃ column densities are lower than $1e16$ molec/cm². This error is chosen to be independent of the meteorological properties, such as cloud coverage and the height of the mixing layer.

As values for the parameters, the choice has been made for $\mathbf{S} = \mathbf{I}$, $L = 50$ km, and $\boldsymbol{\mu} = \mathbf{0}$. The maximum likelihood method would not be suitable to use for this problem, especially not to estimate L , because the ensemble members that describe how the concentrations of NH₃ depend on changes in \mathbf{w} , \mathbf{E}_w , depend on the length scale L . For each time that the log-likelihood function is evaluated for a different L , a new $\mathbf{H}\mathbf{E}_w$, based on a new ensemble would need to be determined. Furthermore, the eigenvalue decomposition of \mathbf{B} would need to be determined after each update of L , which is also computationally costly.

In this experiment, the value of $L = 50$ km is chosen, because for this value, the MFs can vary a lot in space, but the number of main patterns of the matrix B and the size of the preconditioned state, m remains relatively small. For $L = 50$, the first 40 eigenvectors of the correlations matrix B_k for each day k , contain 93% of the energy of the system, which approximates the balance in equation (4.56). For the ensemble of model runs, 3 ensemble members are needed per main pattern, as described in section 4.8. This leads to an ensemble of size 120, which is needed to determine the linear approximate state.

5.4.3. Results

In this experiment, the optimal multiplication factors have been estimated for the period of March 1st to March 10th. For the nature-run, the true MFs are chosen to be a sample of the prior distribution:

$$\gamma^{true} \sim \text{LogN}(\mu, \mathbf{B}). \quad (5.12)$$

The results for March 3th will be discussed in this section. These are representative for the other days. In Figure 5.10, the multiplication factors are shown. Figure 5.10a shows the background parameters γ^b , Figure 5.10b shows the optimally estimated MFs $\hat{\gamma}$, Figure 5.10c shows the true MFs that are used in the nature-run, γ^{true} , and Figure 5.10d shows the difference between the estimated and the true MFs.

It can be seen that above the land surface the estimated and true MFs do show a lot of similarities, especially, compared to the initial MFs. Both figures 5.10b and 5.10c have a peak in the MFs above Friesland and Flevoland and have increased emissions in the south and east of Belgium. However, the extreme values of the peaks are not accurately retrieved by the 4D-Var method but underestimated. In Flevoland, a peak of 6 is estimated, whereas the true peak had values close to 8, and in the southeast of Belgium, a peak of 3 to 4 is found instead of a peak of 6. In Figure 5.10d it can be seen that these underestimated peak regions are surrounded by regions where the MFs are overestimated. On the other hand, the regions where the MFs decrease or remain fairly similar are estimated correctly, provided that these regions are not very close to the high MF areas. The 4D-Var method retrieves decreased MFs for Germany and the west of the Netherlands.

Above the water surfaces, the 4D-Var method does not retrieve the MFs accurately. Above the North Sea, there is a huge peak in the estimated MFs, which does not occur at all in the true MF plot. Also, above the IJsselmeer, there is a relatively large error in MFs.

This error above the water surfaces has to do with the amount of anthropogenic NH_3 emissions. Figure 5.11 shows a map of the daily average NH_3 emissions for March 3th. Figure 5.11a shows the original emission inventory, described in section 2.4.1, Figure 5.11b shows the estimated emissions (emission inventory times the optimally estimated MFs), Figure 5.11c shows the emissions used in the nature run (emission inventory times the true MFs), and Figure 5.11d shows the differences between the estimated and the true emission inventories. It can be seen that the optimized emission does resemble the true emission inventory much better than the original inventory did. Also, it can be seen that for both the estimated and true emission inventories, the emissions above the North Sea and the IJsselmeer are zero. At those locations, no ammonia is emitted by anthropogenic sources. When the original emissions are zero, the emission inventories cannot be changed there, even for a very large multiplication factor, as multiplication with zero remains zero for all MFs. Hence, the incorrectly estimated MFs above the North Sea and the IJsselmeer do not influence the emission inventories above water surfaces.

However, in Figure 5.11d, it can be seen that the emissions are generally overestimated near coastal areas and for the area surrounding the IJsselmeer. Due to the spatial correlation in MFs, increased MFs above water cause slightly increased MFs close to water. Also, the emission sources close to water surfaces are used to compensate for the differences between the modeled concentrations and the measurements above the sea, caused by measurement noise and limitations of the 4D-Var method.

Additionally, it can be seen that for low emission regions, the MFs are more often incorrect, but that does not influence the emission inventories as much. This is caused as for low emission regions, changes in emissions are difficult to measure, and improving the MFs does barely decrease the 4D-Var cost.

In Figure 5.12, the NH_3 column densities are plotted for one hour on March 16. The figures have been made for multiple hours, but only the figures at 16:00 are included in this report, as they all show similar results. In Figure 5.12a, the densities of the background state are plotted, in Figure 5.12b the estimated densities, in Figure 5.12c, the noisy NH_3 observations resulting from the nature run, and in Figure 5.12d the difference in concentration between the noisy observations and the estimated densities. The NH_3 column densities resulting from the estimated MFs look quite similar to the true observations, apart from all the noise, especially, compared to the initial column densities. The difference plot of Figure 5.12d shows that a noisy map that somewhat resembles a Gaussian noise sample. This indicates that

the differences in estimated and measured concentrations seem to be mainly caused by noise in the observations. However, there seems to be some correlation in the NH_3 estimation error. It can for example be seen that in Flevoland, where the emission estimates are underestimated, the NH_3 column densities are also underestimated for a large group of neighboring pixels, and that the NH_3 column densities are overestimated in North-Holland, which could be caused by the increased emissions close to the IJsselmeer.

It has also been tested what would happen if the experiment was performed without adding noise to the observations. In that test, it was confirmed that there are indeed regions where both emissions and NH_3 column densities are overestimated and regions where both emissions and NH_3 column densities are underestimated. The locations of the errors in NH_3 column densities are close to, but not identical to the regions with incorrectly estimated emissions, as the atmospheric NH_3 is transported by advection and diffusion. If the effect of incorrect NH_3 emissions is indeed measurable, there might be adjustments possible for the adjoint-free 4D-Var method to increase the emission estimates even more.

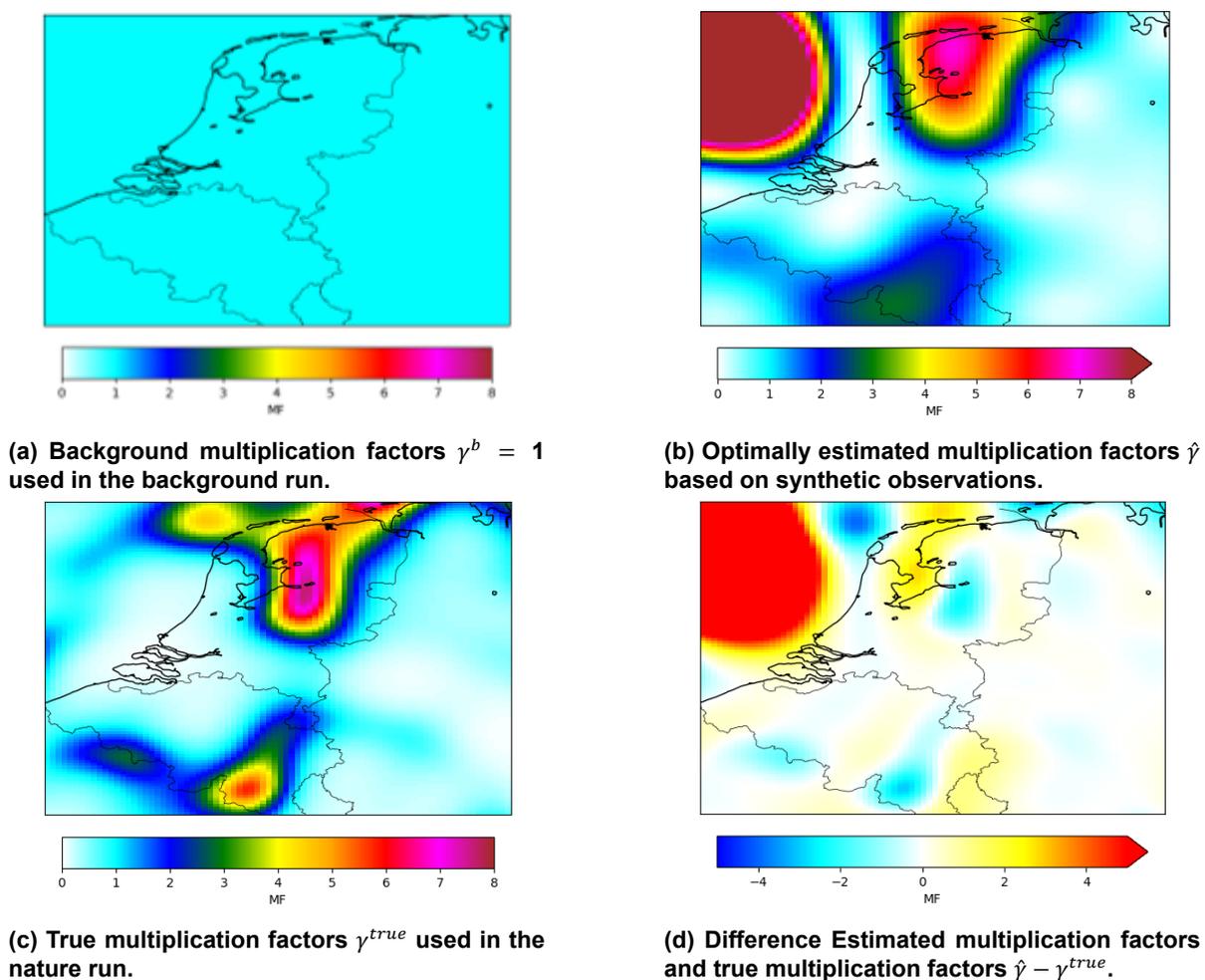


Figure 5.10: Multiplication factors on March 3th.

5.4.4. Conclusion

It can be concluded that the 4D-Var method improves the emission inventory almost everywhere. The regions are retrieved where the emissions increase due to large MFs. However, the optimized values of the multiplication factors are not always perfect. The highest values can be underestimated and regions surrounding the highest peaks can incorrectly compensate for this by overestimating the MFs.

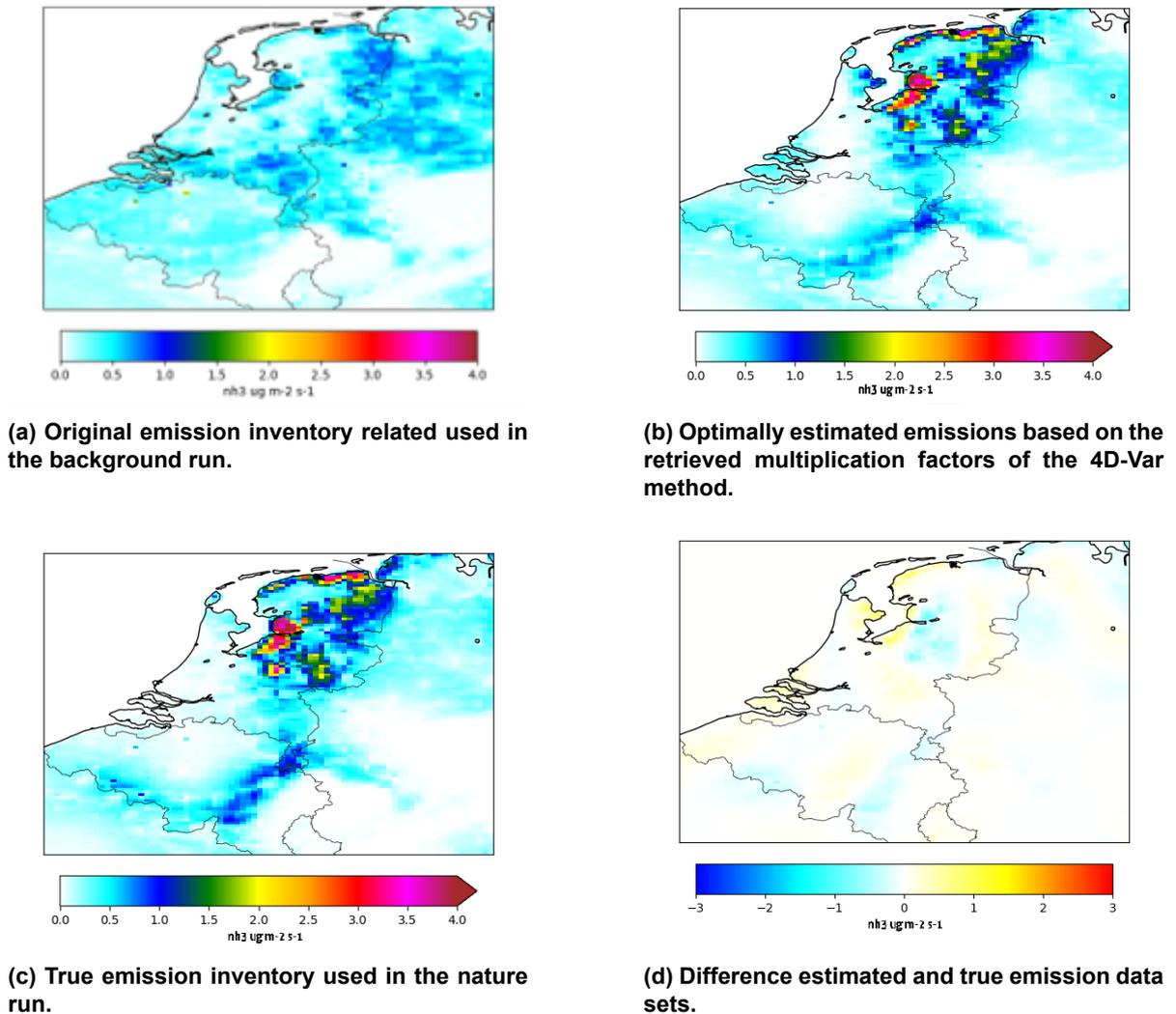
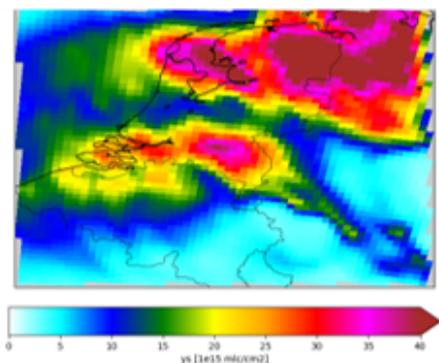


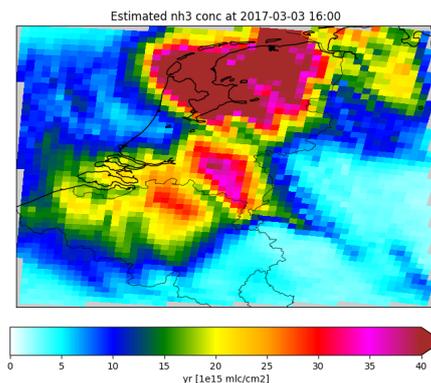
Figure 5.11: Emission inventories daily average of March 3th.

In regions where there are no anthropogenic emissions, such as above water surfaces, the estimated MFs can be very different, but that is not an issue as multiplication factors only cause increased emissions if there actually are emissions. However, MFs in the regions surrounding the zero-emission regions can incorrectly increase as well. In low emission regions, it can be difficult to estimate the MFs correctly, as these regions barely influence the NH_3 column densities and subsequently barely influence the cost of the 4D-Var function.

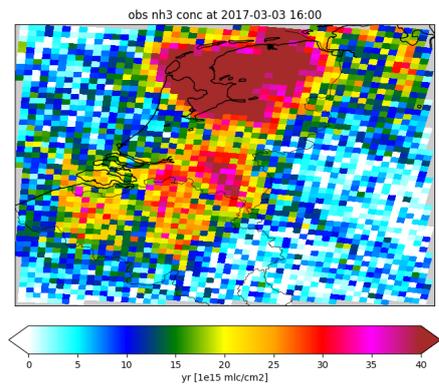
The errors in the estimated multiplication factors are mainly expected to be caused by the accuracy that is lost when using a preconditioner, and by the unobservability of the emission variations of low emission areas. It appears that reducing the measurement noise does barely influence the optimally estimated MFs. This happens because the measurement noise is unbiased and the number of measurements is large, both in time and in space.



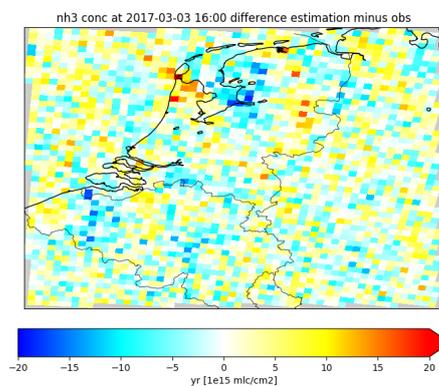
(a) Initial NH_3 column densities made in the background run



(b) Optimally estimated NH_3 column densities based on the retrieved multiplication factors of the 4D-Var method.



(c) True noisy observations of NH_3 column densities made in the nature run.



(d) Difference estimated and "true" NH_3 column densities.

Figure 5.12: Hourly NH_3 column densities on March 3th at 16:00.

6

Conclusion

In this study, adjoint-free 4D-Var methods to improve the ammonia emission estimates for the Netherlands have been developed, implemented, and tested. The adjoint-free 4D-Var methods aim to find an ammonia emission inventory for which simulation with the LOTOS-EUROS model would better resemble observational data. Measurements of two instrument types have been considered: the miniDOAS instrument of LML systems and the IRS instrument on the MTG-S satellite. Four different versions of the 4D-Var method have been developed, all having their advantages related to the accuracy of the result and the computational cost. Conclusions of the study will be presented by answering the research questions of section 1.4

Research question 1: Could an efficient adjoint-free 4D-Var method be designed to estimate time-varying ammonia emissions in the Netherlands using the observations of the LML stations?

This study started with developing an adjoint-free version of the 4D-Var method to estimate time-varying multiplication factors for ammonia emissions. Minimizing the 4D-Var cost function without using an adjoint model was done by approximating the NH_3 concentrations by a linear approximate model in terms of the multiplication factors. As the NH_3 emissions, and hence the multiplication factors, have a nearly linear effect on the NH_3 concentrations and observations, this approximate model was found to be very accurate. Some computational cost is related to determining the linear approximate model as an ensemble of LOTOS-EUROS model runs is needed. However, the size of the ensemble could be reduced a lot, when using the fact that NH_3 is a relatively short-lived trace gas. Also, the ensemble could be run using a chemically reduced version of the LOTOS-EUROS model, which decreases the computational time by a factor of three. In the case of MFs varying in time only, the ensemble size could be limited to 3, and if the domain is divided into three regions with their own MFs, the ensemble size could be limited to 9, both proceeded by one full-chemistry run. This part of the method proved to be very time-efficient.

With this linear approximate model, two adjoint-free 4D-Var methods were developed to estimate emission multiplication factors; one where it was assumed that the multiplication factors were a sample of a Gaussian distribution and one where the multiplication factors were a sample of a log-normal distribution. Both 4D-Var cost functions were minimized using a gradient-based approach. For the model version with the Gaussian prior distribution, the gradient of the cost function became a linear expression in terms of the parameters. Because of that, the multiplication factors that minimized the 4D-Var cost function were found analytically. For the model version with the log-normal prior distribution, the gradient became a nonlinear expression. Now, a numerical optimizer was needed to minimize the cost function. This takes some additional computational effort, but overall the computational cost was negligible in comparison to running the LOTOS-EUROS ensemble.

The 4D-Var cost functions contain parameters related to the distribution of the MFs and the observation likelihood. The maximum likelihood approach was used to quantify a bias parameter for the MFs and the standard deviations of the MFs and the representation error of the observations. According to the maximum likelihood method, both the original MFs and the observations contain large uncertainties.

Once the two efficient adjoint-free 4D-Var methods were developed, they were tested initially in identical twin experiments using synthetic observations. It was found that the methodology worked well for short periods (less than 30 days). However, when observational noise was added to the synthetic measurements, the estimated MFs became noisy as well. It was found that the uncertainty decreased when using 5 measuring stations instead of one. However, the results were still dominated by observational noise.

Last, the methodology was tested using real LML observations. Both 4D-Var methods retrieved the optimal MFs, but the obtained result appeared to be unrealistic. It was found that the MFs obtain a log-normal distribution, which indicates that the log-normal method should give the most realistic result. Hence, if computational time allows it, the log-normal method would be preferred over the Gaussian method. Also, experiments with different MFs for different regions were performed. They showed that the emission estimates should not only be improved in time, but also in space. However, as there are only 5 LML stations, the LML observations could not be used to retrieve a realistic parameterization for spatially varying MFs.

Research question 2: Could an efficient adjoint-free 4D-Var method be designed to estimate time-varying ammonia emissions in the Netherlands on a fine spatial resolution, using synthetic observations of the IRS instrument on the MTG-S satellite.

The objective of the second part of the study was to include a realistic parameterization for the spatial variance of the MFs. To better observe the spatial distribution of atmospheric NH_3 , synthetic observations of the not yet operational IRS instrument on the MTG-S satellite were considered.

In this experiment, daily multiplications are determined for each grid cell. This leads to an enormous amount of parameters to estimate. To keep the computational cost feasible, a preconditioner in reduced space has been considered. This preconditioner uses the property that the MFs are correlated in space. The concept is to consider only the main patterns of the MFs' spatial variability. The linear approximate state, the cost function, and its gradient can now be expressed in terms of a linear combination of the small number of main patterns, instead of the very large parameter vector containing all MFs. In this way, the adjoint-free 4D-Var method is made computationally efficient, while obtaining accurate results. For determining the main patterns, an eigenvalue decomposition of a very large matrix has to be determined, which can take some time. However, in this study, most of the computational effort is still related to running a LOTOS-EUROS ensemble to make the linear approximate state. For this problem, an ensemble of size 120 was needed, preceded by one full-chemistry run.

Again, two versions of the preconditioned adjoint-free 4D-Var methods were developed; one where the MFs have a Gaussian prior distribution and one where the MFs have a log-normal distribution. The Gaussian case can again be solved analytically with negligible computational cost. The log-normal case is more computational costly, especially since an additional positivity constraint for the large number of MFs is added to the 4D-Var objective. However, the optimization cost was negligible in comparison to running the LOTOS-EUROS model 121 times.

The preconditioned adjoint-free 4D-Var methods were tested in an identical twin experiment. The optimized inventory did resemble the inventory used to generate the synthetic observations much better than the original inventory. The 4D-Var method did retrieve the regions where the MFs were highest. Also, unbiased observational noise appeared to barely influence the results. However, the results were not perfect. The regions with the highest MF peaks were generally smooth out over a larger region. Also, for low emission areas, the errors in the estimated MFs were relatively large, as the IRS instrument has difficulties observing the relatively small variations in NH_3 concentrations. However, this does barely affect the total NH_3 emissions. Furthermore, the method has difficulties estimating the emissions close to zero-emission areas such as coastal areas. Hence, when zooming in to a local scale, the emission estimates can be imperfect, but overall the adjoint-free 4D-Var methods do greatly improve the emission inventories.



Recommendations for further research

The developed adjoint-free 4D-Var methodologies show great potential, as multiplication factors are retrieved accurately in a reasonable amount of time. However, before the methods can be used for improving ammonia emission inventories, some further research should be done.

- First, it should be analyzed how well the method works when more errors of the model are involved. Model errors, resulting from for example imperfect meteorological input data or incorrect chemistry, are expected to cause the errors in the estimated MFs. In the study of (Zijlker 2020), it was for a similar 4D-Var method shown that results become unrealistic when plumes in the model are misaligned with the plumes of the observations. Research should be done where a twin experiment is performed where the observations are not generated by the LOTOS-EUROS model, but by a different chemical transport model. In this way, it can be tested how the method behaves when more model errors are made, while still being able to compare the estimated MFs to known true MFs.
- A second improvement would be to consider a more realistic satellite operator for $\mathcal{H}(\mathbf{x}, \theta)$. The operator used for this study is easy to implement and great for showing the potential of the 4D-Var methodology, but it is missing some properties of real satellite retrievals. Once the map $\mathcal{H}(\mathbf{x}, \theta)$ is improved, more research could be performed regarding the required quality of the IRS observations to estimate emissions at different spatial or temporal scales. In reality, observations are often not generated when the cloud density is high. Furthermore, satellite observations tend to be biased, as for example shown for NH_3 observations from the CrIS instrument (Dammers, Shephard, et al. 2017). It is important to test how missing data and biased data would influence the performance of the inversion algorithm.
- Third, this study determines a multiplication factor for all NH_3 emissions, whereas the MFs are expected to depend on the type of emission source. For example, if some MFs represent emissions due to fertilizer application, then the MFs should not be used for the neighboring NH_3 emissions from animal housing.
- Fourth, additional types of measurements could be included. Additional NH_3 measurements are for example from the MAN network, generating monthly average NH_3 surface concentrations, and the CrIS and IASI instruments on board satellites, measuring NH_3 column densities once per day. It would even be possible to include measurements of other chemical tracers as well. Measurements of concentrations and deposition rates of NO_3 and NH_4^+ are available and are dependent on NH_3 emission.
- Last, when all aspects of the method are considered to be realistic, the method can be used for the real IRS observational data. The method is expected to perform worse than for the identical twin experiment. Unfortunately, testing this will not be possible until the MTG-S satellite is launched.

Bibliography

- Andreae, M. O. and P. Merlet (2001). "Emission of trace gases and aerosols from biomass burning". In: *Global Biogeochemical Cycles*. ISSN: 08866236. DOI: 10.1029/2000GB001382.
- Baek, Bok Haeng, Viney P. Aneja, and Quansong Tong (2004). "Chemical coupling between ammonia, acid gases, and fine particles". In: *Environmental Pollution*. ISSN: 02697491. DOI: 10.1016/j.envpol.2003.09.022.
- Berkhout, Augustinus J.C. et al. (2017). "Replacing the AMOR with the miniDOAS in the ammonia monitoring network in the Netherlands". In: *Atmospheric Measurement Techniques*. ISSN: 18678548. DOI: 10.5194/amt-10-4099-2017.
- CBS (Sept. 2020). "Emissies van luchtverontreinigende stoffen volgens NEC-richtlijnen". In: URL: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70947ned/table?ts=1605023105914>.
- Compendium voor de Leefomgeving (2018). "Kwaliteit en trend stikstofbeschikbaarheid ecosystemen in 2017". In: URL: <https://www.clo.nl/indicatoren/nl2045-overschrijding-stikstofdepositie-natuur>.
- Courant, R. and D. Hilbert (1953). *Methods of Mathematical Physics*. Vol. 1. Interscience Publishers. DOI: <https://doi.org/10.1002/qj.49708034534>.
- Dammers, Enrico, M. Schaap, et al. (2017). "Measuring atmospheric ammonia with remote sensing campaign: Part 1 – Characterisation of vertical ammonia concentration profile in the centre of The Netherlands". In: *Atmospheric Environment*. ISSN: 18732844. DOI: 10.1016/j.atmosenv.2017.08.067.
- Dammers, Enrico, Mark W. Shephard, et al. (2017). "Validation of the CrIS fast physical NH₃ retrieval with ground-based FTIR". In: *Atmospheric Measurement Techniques*. ISSN: 18678548. DOI: 10.5194/amt-10-2645-2017.
- Fountoukis, C. and A. Nenes (2007). "ISORROPIAII: A computationally efficient thermodynamic equilibrium model for K⁺-Ca²⁺-Mg²⁺-NH₄⁺-Na⁺-SO₄²⁻-NO₃⁻-Cl-H₂O aerosols". In: *Atmospheric Chemistry and Physics*. ISSN: 16807324. DOI: 10.5194/acp-7-4639-2007.
- Gezondheidsraad (Jan. 2018). *Gezondheidseffecten luchtverontreiniging. Achtergronddocument bij: Gezondheidswinst door schonere lucht*. Tech. rep.
- Haussaire, Jean-Matthieu (2020). *Maximum Likelihood for FLEXVAR*. Tech. rep.
- Hendriks, C. et al. (2016). "Ammonia emission time profiles based on manure transport data improve ammonia modelling across north western Europe". In: *Atmospheric Environment*. ISSN: 18732844. DOI: 10.1016/j.atmosenv.2016.01.043.
- Johnson, Martin T. et al. (2008). "Field observations of the ocean-atmosphere exchange of ammonia: Fundamental importance of temperature as revealed by a comparison of high and low latitudes". In: *Global Biogeochemical Cycles*. ISSN: 08866236. DOI: 10.1029/2007GB003039.
- Kuenen, J. J.P. et al. (2014). "TNO-MACC-II emission inventory; A multi-year (2003-2009) consistent high-resolution European emission inventory for air quality modelling". In: *Atmospheric Chemistry and Physics*. ISSN: 16807324. DOI: 10.5194/acp-14-10963-2014.
- Leegwater, M. (2020). *Chemical model reduction of the LOTOS-EUROS model for ammonia*. Tech. rep.
- Lewis, John M., S. Lakshmiarahan, and Sudarshan Dhall (2006). *Dynamic Data Assimilation*. Cambridge University Press. ISBN: 978-0-521-85155-8. DOI: 10.1017/cbo9780511526480. URL: <https://www.cambridge.org/core/books/dynamic-data-assimilation/91BD4736A2CD99B575002A89>
- LOTOS-EUROS applications website (Nov. 2020). URL: <https://lotos-euros.tno.nl/applications/>.
- Manders, A. M. M. et al. (2017). "Curriculum vitae of the LOTOS-EUROS (v2.0) chemistry transport model". In: *Geoscientific Model Development* 10.11, pp. 4145–4173. DOI: 10.5194/gmd-10-4145-2017. URL: <https://gmd.copernicus.org/articles/10/4145/2017/>.
- Manders-Groot, A.M.M., A.J. Segers, and S. Jonkers (2019). *LOTOS-EUROS v2.2.000 Reference Guide*.
- Meirink, J F, H J Eskes, and A P H Goede (2006). *Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling*. Tech. rep., pp. 1275–1292. URL: www.atmos-chem-phys.net/6/1275/2006/.

- Nocedal, Jorge and Stephen J. Wright (2006a). "Quasi-Newton Methods". In: *Numerical Optimization*. Ed. by Thomas V. Mikosch, Sidney I. Resnick, and Stephen M. Robinson. Second edi. Springer. Chap. 6: Quasi-N, pp. 135–163. ISBN: 978-0-387-30303-1. DOI: <https://doi.org/10.1007/978-0-387-40065-5>.
- (2006b). "Sequential Quadratic Programming". In: *Numerical Optimization*. Ed. by Thomas V. Mikosch, Sidney I. Resnick, and Stephen M. Robinson. Second edi. Springer. Chap. 18: Sequen, pp. 529–562. ISBN: 978-0-387-30303-1. DOI: <https://doi.org/10.1007/978-0-387-40065-5>.
- Pitcairn, C. E.R. et al. (1998). "The relationship between nitrogen deposition, species composition and foliar nitrogen concentrations in woodland flora in the vicinity of livestock farms". In: *Environmental Pollution*. DOI: 10.1016/S0269-7491(98)80013-4.
- Programma Aanpak Stikstof (n.d.). *2015-2021 zoals gewijzigd na partiële herziening op 18 december 2017*. Tech. rep.
- Remkes, J.W., J.J. van Dijk, et al. (2020). *Niet alles kan overal. Eindadvies over structurele aanpak Adviescollege Stikstofproblematiek*. Tech. rep.
- Remkes, J.W., E. Dijkgraaf, et al. (2019). *Niet alles kan. Eerste advies van het Adviescollege Stikstofproblematiek. Aanbevelingen voor korte termijn*. Tech. rep.
- Riddick, S. N. et al. (2012). "The global distribution of ammonia emissions from seabird colonies". In: *Atmospheric Environment*. ISSN: 13522310. DOI: 10.1016/j.atmosenv.2012.02.052.
- Segers, Arjo, Janot Tokaya, and Sander Houweling (2020). *Description of the CH4 Inversion Production Chain*. Tech. rep. Copernicus Atmosphere Monitoring Service.
- Skjøth, C. A. et al. (2011). "Spatial and temporal variations in ammonia emissions $\hat{g}\hat{c}$ " a freely accessible model code for Europe". In: *Atmospheric Chemistry and Physics*. ISSN: 16807316. DOI: 10.5194/acp-11-5221-2011.
- Sutton, M. A., J. K. Schjorring, and G. P. Wyers (1995). "Plant-atmosphere exchange of ammonia". In: *Philosophical Transactions - Royal Society of London, A*. ISSN: 0962-8428. DOI: 10.1098/rsta.1995.0033.
- TNO (Aug. 2019). *FACTSHEET EMISSIES EN DEPOSITIE VAN STIKSTOF IN NEDERLAND*. Tech. rep.
- Vermeulen, Peter T.M. and A. W. Heemink (Oct. 2006). "Model-reduced variational data assimilation". In: *Monthly Weather Review* 134.10, pp. 2888–2899. ISSN: 00270644. DOI: 10.1175/MWR3209.1. URL: <https://journals.ametsoc.org/view/journals/mwre/134/10/mwr3209.1.xml>.
- Volten, H et al. (2012). "Atmospheric Measurement Techniques Two instruments based on differential optical absorption spectroscopy (DOAS) to measure accurate ammonia concentrations in the atmosphere". In: *Atmos. Meas. Tech* 5, pp. 413–427. DOI: 10.5194/amt-5-413-2012. URL: www.atmos-meas-tech.net/5/413/2012/.
- Walker, John et al. (2008). "Inferential model estimates of ammonia dry deposition in the vicinity of a swine production facility". In: *Atmospheric Environment*. ISSN: 13522310. DOI: 10.1016/j.atmosenv.2007.06.004.
- Whitten, Gary Z., Henry Hogo, and James P. Killus (1980). "The carbon-bond mechanism: A condensed kinetic mechanism for photochemical smog". In: *Environmental Science and Technology*. ISSN: 15205851. DOI: 10.1021/es60166a008.
- Wichink Kruit, R. J. et al. (2012). "Modeling the distribution of ammonia across Europe including bi-directional surface-atmosphere exchange". In: *Biogeosciences*. ISSN: 17264170. DOI: 10.5194/bg-9-5261-2012.
- Zijlker, T. (2020). *Estimating NOx emissions using S5-P TROPOMI An adjoint-free 4DVAR approach*. Tech. rep.