CO_2 State and Flux Estimation by Four-Dimensional Variational Data Assimilation

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

vorgelegt von

Johannes Klimpt

aus Eisenach

Köln, 2016

Berichterstatter: PD. Dr. H. Elbern
Prof. Dr. S. Crewell

Tag der mündlichen Prüfung: 02. November 2016

Abstract

The four-dimensional variational data assimilation system EURAD-IM is extended for the realisation of CO_2 inversions. These analysis techniques seek to optimise CO_2 surface-atmosphere fluxes by the use of atmospheric CO_2 concentration observations and atmospheric transport models. In this work, CO_2 surface-atmosphere fluxes are optimised jointly with atmospheric CO_2 concentrations. The optimisation is characterised by a high degree of freedom of the optimisation space and by the use of frequent observations. The system traces location, time, and strength of surface-atmosphere CO_2 fluxes due to their weak signal in the atmospheric concentration. This enables new possibilities for the determination of sinks and sources of atmospheric CO_2. In order to increase the sensitivity of the assimilation system for small surface-atmosphere fluxes, the modelling of the background error variances is improved. Numerical experiments with synthetic observations demonstrate the benefits of the joint optimisation of CO_2 concentrations and fluxes compared to optimising fluxes only. A long-term simulation of June 2012 is executed to evaluate the extended data assimilation system. A significant improvement of analysed concentration time-series is obtained. The analysed fields of CO_2 surface-atmosphere fluxes show spatially inhomogeneous structures close to surface measurement stations. This reveals the high requirements on the modelled CO_2 fluxes and their error characterisation for this optimisation approach, in order to control the high degree of freedom of the optimisation space. The ability of the system to trace back and improve biogenic fluxes is shown with the aid of several examples.

Kurzzusammenfassung

Das vierdimensionale variationelle Datenassimilationssytem EURAD-IM wurde erweitert für die Realisierung von CO_2-Inversionen. Diese Analysetechniken optimieren CO_2-Oberflächen-Atmosphärenflüsse durch die Verwendung von Messungen der CO_2-Konzentration in der Atmosphäre und von atmosphärischen Transportmodellen. In dieser Arbeit werden CO_2-Fflüsse und CO_2-Konzentrationen gemeinsam optimiert. Die Optimierung ist gekennzeichnet durch einen hohen Freiheitsgrad des Optimierungsraums und durch die Verwendung zeitlich hoch aufgelöster Beobachtungen. Das Analysesystem kann den Ort, die Zeit und die Größe von CO_2-Oberflächen-Atmosphärenflüssen aufgrund ihres Fußabrucks in den atmosphärischen CO_2-Konzentrationen zurückverfolgen. Dies eröffnet neue Möglichkeiten zur Bestimmung von CO_2-Quellen und -Senken. Die Optimierung der CO_2-Oberflächen-Atmosphärenflüsse wird von der absoluten Größe der betreffenden Flüsse beeinflusst. Aus diesem Grund werden die Hintergrundfehler der Flüsse in Abhängigkeit der Flussgröße modelliert, wodurch die Analyse von kleinen Flüssen verbessert wird. Numerische Experimente mit synthetischen Beobachtungen beweisen die Vorteile der gemeinsamen Optimierung von CO_2-Konzentrationen und -Flüssen gegenüber der alleinigen Optimierung von CO_2-Flüssen. Um das erweiterte Datenassimilationssytem zu testen wurde eine Langzeitstudie für den Juni 2012 durchgeführt. Die Analyse bewirkt eine deutliche Verbesserung der modellierten CO_2-Konzentrationen. Die optimierten CO_2-Flüsse sind räumlich inhomogen in der Nähe von Messstationen, die sehr dicht am Boden sind. Für eine verbesserte Analyse mit der verwendeten Optimierungsstrategie müssen die a priori Flüsse und ihre Fehlercharakterisierung sehr hohe Ansprüche erfüllen. Die Fähigkeit des Assimilationssystems biogene Flüsse zurückzuverfolgen und zu verbessern, wird anhand mehrerer Beispiele gezeigt.

Contents

1 Introduction **1**

1.1 The carbon cycle and its impact on climate 1

1.2 State of the art inverse modelling of carbon fluxes 2

1.2.1 CO_2 inversion from global to regional scale 2

1.2.2 Bottom-up approach 4

1.2.3 Top-down approach 5

1.3 Objectives of this thesis 8

2 Theory of Data Assimilation **11**

2.1 Mathematical formulation of the Problem 12

2.1.1 Notation 13

2.1.2 Assumptions 14

2.1.3 Bayesian formulation and the cost function 16

2.2 Overview of CO_2 DA concepts 17

2.2.1 Sequential DA 19

2.2.2 Variational methods 19

3 Model description **23**

3.1 Meteorological model WRF 23

3.2 Land surface model CLM 24

3.2.1 Modelling of photosynthesis and leaf respiration . . . 25

3.2.2 Modelling of soil respiration 26

3.3 Anthropogenic CO_2 emission inventory 27

3.4 4D-Var system EURAD-IM 28

3.4.1 Forward model of EURAD-IM for CO_2 29

3.4.2 Adjoint EURAD-IM model 31
3.4.3 Preconditioning of the cost function and minimisation 32
3.4.4 Modelling of background error covariance matrices . . 33

4 Model set-up and observations 37
4.1 Model set-up 37
4.1.1 Model domain and nesting configuration 37
4.1.2 Different CO_2 tracers in EURAD-IM 38
4.2 Observations 40
4.2.1 Available measurements 40
4.2.2 Observation error covariance matrices 42

5 Simulation results 43
5.1 Identical twin experiments 43
5.1.1 Experiments with one-dimensional version of EURAD-IM 44
5.1.2 Experiments with EURAD-IM 5.8.1 48
5.2 Real case study for TR32 domain 55
5.2.1 5 km domain 56
5.2.2 1 km domain 66

6 Summary and Outlook 75
6.1 Summary 75
6.2 Outlook 77

A 79
A.1 The adjoint model $\mathbf{M}^{\mathrm{T}}$ 79
A.2 Derivation of the diagonal of $\mathbf{K}$ 80

Bibliography 81

Danksagung 99

List of Figures

2.1 Diurnal time series of photosynthesis. 12
2.2 Classification of data assimilation methods. 18
2.3 Principle of joint state and flux estimation. 21

3.1 Overview of used models and data. 24
3.2 Biogenic fluxes at 5 km resolution. 28
3.3 Anthropogenic emissions from power generation and road traffic. 29
3.4 Comparison of non-monotone and Walcek advection scheme. 31

4.1 Domain decomposition configuration. 38
4.2 Time series of six tagged CO_2 tracers and observations at station Rollesbroich. 39

5.1 Overview of the experiment setup of Sect. 5.1.1.1. 45
5.2 Influence of the variance of **K** on the flux factor optimisation. 45
5.3 Identical twin experiment assessing the ill-posedness of joint optimisation. 47
5.4 Initial value correction for different vertical layers. 50
5.5 Comparison of flux factors obtained by only flux factor analysis and joint analysis. 51
5.6 Comparison of cost functions for joint optimisation and only flux factor optimisation. 52
5.7 Time series of CO_2 concentration for nature, background, joint analysis, and flux factor analysis run. 53
5.8 Vertical profiles of CO_2 concentration for nature, background, joint analysis, and flux factor analysis run. . . . 54

5.9 RMSE and bias of background run and analysis runs with respect to concentration observations. 57
5.10 Concentration time series for 5 km domain for June 2012. . . 59
5.11 CO_2 tracer of soil respiration, anthropogenic emission, and photosynthesis in the area of the Cabauw tower. 61
5.12 BmA for anthropogenic emissions and soil respiration in the area of the Cabauw tower. 61
5.13 RMSE and bias of background run and analysis fluxes with respect to NEE observations. 62
5.14 Analysed flux factor at 5 km domain for 14 June 2012. . . . 63
5.15 Vertical cross section of different tagged CO_2 tracers. 64
5.16 BmA of different tagged CO_2 tracers on 14 June 2012. . . . 65
5.17 NEE time series for 5 km domain for June 2012. 67
5.18 Mean biogenic CO_2 fluxes for June 2012. 68
5.19 Bias and RMSE of background run and analysis runs with respect concentration and flux observations for 1 km domain. 69
5.20 Concentration time series for 1 km domain for June 2012. . . 71
5.21 NEE time series for 1 km domain for June 2012. 72
5.22 Analysed flux factor at 1 km domain for 14 June 2012. . . . 73
5.23 Analysed flux factors for Weisweiler, Niederaußem, Frimmersdorf, and Neurath for June 2012. 74

List of Tables

4.1 Used ground stations in this work. 41

5.1 Analysed flux factor (ff) in dependence of $\sqrt{\mathbf{K}_{r,r}}$. 46

5.2 RMSE and bias of analysed flux factors for configuration **S1** and **S2**. 46

5.3 Configurations for identical twin experiment of Sect. 5.1.2. . 48

5.4 Configurations of stations whose data is used for assimilation or evaluation. 56

5.5 Emissions of Weisweiler, Niederaußem, Frimmersdorf, and Neurath for June 2012. 74

List of Acronyms

3D-Var	Three-Dimensional Variational
4D-Var	Four-Dimensional Variational
a.g.l.	above ground layer
BECM	Background Error Covariance Matrix
BLUE	Best Linear Unbiased Estimate
BmA	Background minus Analysis
CLM	Community Land Model
CTM	Chemistry Transport Model
DA	Data Assimilation
ECMWF	European Centre for Medium-Range Weather Forecasts
EURAD-IM	European Air Pollution and Dispersion-Inverse Model
ff	Flux Factor
iv	Initial Value
GHG	Greenhouse Gases
LAI	Leaf Area Index
L-BFGS	Limited-Memory Broyden Fletcher Goldfarb Shanno
NEE	Net Ecosystem Exchange
NmA	Nature minus Analysis
NWP	Numerical Weather Prediction
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
RMSE	Root Mean Square Error
TNO	Toegepast Natuurwetenschappelijk Onderzoek
WRF	Weather Research and Forecast Model

Chapter 1

Introduction

1.1 The carbon cycle and its impact on climate

Carbon is by definition the essential element of organic life on earth. The carbon cycle, the movement of carbon among the reservoirs biosphere, pedosphere, geosphere, hydrosphere, and atmosphere, comprises a sequence of processes that are decisive for living organisms and influences particularly the climate of the earth. Carbon in the earth's atmosphere occurs mainly as carbon dioxide (CO_2), methane (CH_4), and carbon monoxide (CO). In the context of a warming climate system, these **g**reen**h**ouse **g**ases (GHG) have gained broad attention over the previous decades and urge for a better understanding of the carbon cycle to provide predictions of future changes. Carbon dioxide is the second most important GHG after water vapour. There is clear evidence that the increase in its atmospheric concentration over the past 200 years from ∼280 ppmV to more than 400 ppmV today [*IPCC*, 2014], is mainly caused by anthropogenic emissions and land use change. This increase in CO_2 is unprecedented over the last 650,000 years [*Keeling et al.*, 2005]. On a global scale, approximately 30% of the anthropogenic emissions are removed by oceans [*Wanninkhof et al.*, 2013] as well as 30% by terrestrial biosphere [*Sitch et al.*, 2013], the latter showing much larger variability and uncertainty [*Jung et al.*, 2011], representing also a critical uncertainty in the global carbon cycle.

1.2 State of the art inverse modelling of carbon fluxes

1.2.1 CO_2 inversion from global to regional scale

Global CO_2 estimation

Interpretation of atmospheric CO_2 observations with the help of transport models to derive estimates about the global carbon budget dates back to the late 1980s [*Enting and Mansbridge*, 1989; *Tans et al.*, 1989] and is commonly called CO_2 inversion. Early studies indicated that the northern hemisphere acts as a net terrestrial carbon sink, based on the investigation of global atmospheric CO_2 distribution or isotope and oxygen measurements [*Tans et al.*, 1990; *Ciais et al.*, 1995; *Keeling et al.*, 1996]. Detailed comparisons [*Law et al.*, 1996; *Denning et al.*, 1999*a*; *Gurney et al.*, 2002] of different global inversion studies showed an agreement that the northern hemisphere is a substantial net sink of $2.0 - 3.5$ PgC/yr. Nevertheless the uncertainties for the continents remained large as *Gurney et al.* [2002] concluded that "consensus has not yet been reached regarding the size and distribution of regional carbon fluxes obtained". On a global scale there is continuous progress in estimating the net terrestrial carbon flux for the past decades. *Houghton* [2001; 2007]; *Canadell et al.* [2007]; *Sarmiento et al.* [2010]; *Le Quéré et al.* [2015] estimate a global carbon sink ranging from 0.20 PgC/yr to 0.27 PgC/yr before 1988 and 1.15 PgC/yr and 4.70 PgC/yr for later years.

Error sources for global inversions

Since surface fluxes are mostly inferred from atmospheric CO_2 observations, rather than fluxes itself, it is not surprising that the proper modelling of atmospheric transport is one of the greatest challenges. On large scales, additionally to turbulence, many processes are parameterized e.g. convection, precipitation, and the characterisation of the boundary layer. Errors of different transport models do not only cause a spread between inversion studies but may also introduce a systematic bias. *Stephens et al.* [2007] corrected for systematic errors of the vertical gradient found in earlier studies and estimated the carbon sink of the northern hemisphere to be 1.5 PgC/yr weaker. Another transport problem is linked to the strong seasonal and diurnal cycle of terrestrial CO_2 fluxes. If these variations of the CO_2 fluxes are not accounted for properly by atmospheric transport models, "rectification" errors occur, leading to substantial uncertainties [*Denning et al.*, 1999*b*; *Engelen et al.*, 2002].

The coarse resolution of a priori fluxes used in global inversions leads to aggregation errors [*Kaminski et al.*, 2001], as spatial patterns of CO_2 fluxes are not sufficiently resolved. In dependence of the resolution is also the representativity error, since inversions implicitly assume that in situ observations can be used as an average CO_2 distribution for one model grid box. This assumption becomes questionable e.g. over mountainous terrain with small scale turbulence in combination with coarse resolutions. *Gerbig et al.* [2003*a*] first quantified the representativity error and suggested horizontal resolutions smaller than 30 km to resolve atmospheric CO_2 dispersion in the boundary layer.
Observations of atmospheric CO_2 are often based on in situ measurements from surface networks, ships, or aircraft measurements. Historically, CO_2 measurement towers have been deployed such that atmospheric CO_2 background concentrations are well represented and the influence of local terrestrial fluxes and anthropogenic emissions is minimised. Mountain summits and small islands were often selected as measurement sites [*Fan et al.*, 1998]. As global inversions often used quite large temporal resolutions (≥ 1 day) they were not capable to analyse the diurnal variability of CO_2 surface fluxes from the biosphere. Hence the contribution to the understanding of the underlying processes of the carbon cycle is limited: "It has been difficult for the two communities [of atmospheric inversions and process models] to even disagree meaningfully" [*Kaminski et al.*, 2002].

Benefits of regional inversions

The resolution of CO_2 inversions, which are denoted as regional, ranges from a few degrees to a few kilometres [*Gerbig et al.*, 2003*a*;*b*; *Matross et al.*, 2006; *Lauvaux et al.*, 2008; 2009*b*;*a*; *Rödenbeck et al.*, 2009; *Peters et al.*, 2010; *Gourdji et al.*, 2010; *Broquet et al.*, 2011]. One of the main advantages of inversions with smaller scales is the usage of better resolved atmospheric measurements. A large part of the information in the signature of the inert CO_2 observed in the atmospheric boundary layer, can only be resolved with relatively small spatio-temporal resolutions [*Van der Molen and Dolman*, 2007]. Contrarily to global inversions, the results of regional top-down inversions can almost directly be compared with observations at the surface, often called bottom-up measurements (see the following sections for a discussion of different approaches). Modelling at this "missing scale", between global inversions and bottom-up models, was a large step towards uncertainty reduction of carbon fluxes, because independent data streams could be compared, which also helped to improve different approaches [*Lauvaux et al.*, 2008].
Top-down methods are used to identify errors in biogeochemical models

[*Peylin et al.*, 2005; *Desai*, 2010] or to validate inventories [*Levin and Karstens*, 2007; *Desai et al.*, 2011], making use of different kinds of observations. On the other hand, surface flux estimates from bottom-up approaches enable better a priori fluxes needed for detailed atmospheric inversions [*Wang et al.*, 2009; *Turner et al.*, 2011]. Finally, regional scale inversion techniques combine many sources of information to gain insight to the driving processes of the carbon cycle, and thus deliver more reliable regional budgets on a national level to check national emission reductions [*IPCC*, 2014].

Requirements for regional inversions

To solve atmospheric inversion problems successfully, measurements have to fulfil high-quality requirements. The precision of atmospheric CO_2 measurements has to be high in comparison to other trace gases, since the chemical inert CO_2 is just dispersed in the atmosphere and surface fluxes add to a high background value. Similarly, just CO_2 net flux can be measured, consisting of the juxtaposed larger fluxes of respiration and photosynthesis, usually inferred by eddy covariance methods. Concentration and flux measurements are discussed in more detail in Sect. 1.2.2 and 1.2.3. The largest part of observations used in this thesis is provided by the TR32 (http://tr32new.uni-koeln.de/) project "Patterns in Soil-Vegetation-Atmosphere-Systems".
Precise information about the surface, orography, changes in vegetation cover, and many other parameters, which influence the wind patterns and the atmospheric boundary layer height, are necessary as well. Requirements for transport models due to high horizontal resolution are discussed in Sect. 1.2.3. *Rodgers* [2000] states that with increasing resolution in CO_2 inversions their uncertainty increases, due to ill-conditioning of the inverse problem and increasing uncertainty of atmospheric transport.

1.2.2 Bottom-up approach

Bottom-up approaches can be classified into two broad categories: process-based models and empirical models based on upscaling. Process-based biogeochemical models simulate carbon exchange between ecosystem and atmosphere by considering the underlying biochemical mechanisms of carbon fluxes (e.g photosynthesis, autotrophic/heterotrophic respiration) for different vegetation types. Early developments of biogeochemistry models rested on many simplifying assumptions [*Potter et al.*, 1993; *Ruimy et al.*, 1996] and have been substantially refined to numerous sophisticated models (e.g. Biosphere Energy-Transfer Hydrology scheme (BETHY) [*Knorr and Heimann*, 2001], Community Land Model (CLM) [*Oleson et al.*, 2010],

Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) [*Krinner et al.*, 2005]), which are discussed e.g in *Huntzinger et al.* [2012]; *Keenan et al.* [2012]. Although the understanding of the biogeochemistry of the carbon cycle increases with process-based models, there are still large variations of different models for short and long-term predictions [*Schwalm et al.*, 2010; *Dietze et al.*, 2011], weakening the confidence in current performance.

Empirical models use inventory data-sets of anthropogenic emissions [*Kuenen et al.*, 2014], (*http://carboeurope.ier.uni-stuttgart.de*), observed fluxes from trees and land use change [*McKinley et al.*, 2011; *Houghton*, 2010], and environmental factors like air temperature, radiation, and vegetation parameters [*Beer et al.*, 2010; *Mahadevan et al.*, 2008].
Improving design of direct measurement techniques of carbon fluxes is an ongoing active research area. Over land, the eddy correlation method [*Deacon*, 1959; *Aubinet et al.*, 2000] is the most established method. To realise measurements with this method several requirements have to be fulfilled (e.g. fully developed turbulence, negligible density fluctuations [*Foken and Wichura*, 1996]), leading to problems especially due to uncertain small scale layering and under lateral flow [*Papale and Valentini*, 2003; *Alfieri et al.*, 2011]. Another drawback is the small spatial representativeness of these measurements (<1 km^2). Although there are currently more than 550 recording flux towers (*http://fluxnet.ornl.gov*), the heterogeneity of the biosphere hampers an accurate upscaling. In general, it can be stated that uncertainty of bottom-up approaches increases with larger resolutions, as more extrapolation is required.

1.2.3 Top-down approach

Inverse modelling with the top-down approach exploits variations of atmospheric CO_2 observations to quantify surface-atmosphere CO_2 fluxes with an atmosphere transport model. The main method, Bayesian synthesis inversion, starts with a priori surface fluxes as an input for atmospheric transport models to simulate concentrations for a set of observed locations. The goal is to find a maximum a posteriori distribution of surface fluxes with respect to the a priori information and the model-data mismatch, usually under the assumption of multivariate normal distributions.

Observations

One difficulty already mentioned in Sect. 1.2.1 is the sparseness of accurate measurements. Networks for CO_2 tower measurements include e.g Fluxnet

(*http://fluxnet.ornl.gov*), NOAA (*http://www.esrl.noaa.gov/gmd/dv/site/-site_table2.php*), or the CarboEurope atmospheric database (*http://www.ce-atmosphere.lsce.ipsl.fr/database/index_database.html*).
Although these measurements are used by almost all top-down inversion studies they do not cover equally all regions. Recent programmes as **G**lobal **A**tmospheric **W**atch (GAW) improved the situation for Asia [*Thompson et al.*, 2016]. Regular aircraft measurements are limited to a few regions, e.g CARIBIC (Civil Aircraft for the Regular Investigation of the atmosphere on the basis of an Instrument Container) and CONTRAIL (Comprehensive Observation Network for TRace gases by AIrLiner) [*Crevoisier et al.*, 2010; *Niwa et al.*, 2012], but allow a more detailed examination of vertical gradients of atmospheric CO_2 [*Stephens et al.*, 2007; *Kretschmer et al.*, 2014]. To overcome sparse in situ measurements a promising approach is the use of remote sensing instruments like the Thermal and Near infrared Sensor for Carbon Observation (TANSO) on Greenhouse gases Observing SATellite (GOSAT, *http://http://www.gosat.nies.go.jp/*) [*Hamazaki et al.*, 2004], the Atmospheric Infrared Sounder (AIRS, *http://airs.jpl.nasa.gov*) on Aqua [*Aumann et al.*, 2003], the Infrared Atmospheric Sounding Interferometer (IASI, *http://smsc.cnes.fr/IASI*) on Met-Op-1, the former SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY) on EnviSAT, and the Orbiting Carbon Observatory-2 (OCO-2, *http://oco.jpl.nasa.gov*).

Difficulties and unsolved problems

The weak signal of terrestrial CO_2 fluxes in atmospheric observations is used to identify the correct time, location and amount of the fluxes with the usage of atmospheric transport models. In combination with the mentioned limited amount of observations, the CO_2 top-down inversion becomes underdetermined (the degree of freedom of terrestrial fluxes is larger than that of the available observations) and ill-posed (several combinations of fluxes exist, differing in time, location, or amount, which are in accordance with the available observations). The ill-posed character of the inversion procedure can be mitigated by additional information to surface-atmosphere fluxes, introduced from other tracers or the energy budget [*Tolk et al.*, 2009], and the combination with bottom-up approaches [*Baker et al.*, 2008; *Andres et al.*, 2012; *Gourdji*, 2013; *Zhu et al.*, 2014].
Besides the modelled surface-atmosphere fluxes and measurements themselves, the correct specification of their respective uncertainties is highly important [*Engelen et al.*, 2002; *Broquet et al.*, 2013]. The proper estimation of the covariance between a priori fluxes and the model-data mismatch

has a comparable influence as the a priori information and the data themselves [*Lauvaux et al.*, 2009*b*]. Different estimation and modelling strategies for the covariance matrices exist, usually taken from **n**umerical **w**eather **p**rediction (NWP) methods.
The influence of the atmospheric transport model as a part of the inversion procedure is often investigated with numerous inter-comparison studies for global and regional inversions. The transport model should be with low numeric diffusion, monotone, and numerical efficient. The uncertainty introduced by transport modelling is discussed in detail in *Lauvaux et al.* [2009*b*]. Several other factors are known to influence the inversion results, e.g. the strategy to update fluxes [*Tolk et al.*, 2011], initial and boundary conditions [*Lauvaux et al.*, 2008; *Peylin et al.*, 2005], and if different parameters are optimised jointly or step-wise [*Peylin et al.*, 2016].

From batch inversions to data assimilation

The simplest approach to solve the inversion problem, which links CO_2 surface fluxes and the CO_2 atmospheric measurements with a transport model, is to calculate a set of linear equations, performing the inversion in one step. This method requires running the atmospheric transport model either once per observation or once per estimated flux region per period. Thus, this inversion method is also called batch inversion [*Gelb*, 1974] or matrix formulation and is historically widely used, in global inversions. It is well known that the dimension of the matrices to be inverted is defined either by the flux space or the observation space [*Rodgers*, 2000], in order to decrease the most time consuming calculation step. Still the increasing amount of observations and spatio-temporal resolution makes this method computationally infeasible, which brought time stepping approaches into focus, used in NWP for several decades under the banner of data assimilation. Advanced data assimilation for stratospheric [*Austin*, 1992; *Fisher and Lary*, 1995] and tropospheric chemistry [*Elbern et al.*, 1997], were applied quite early compared to data assimilation of CO_2 flux estimation based on the sequential [*Bruhwiler et al.*, 2005; *Peters et al.*, 2005] or variational method [*Chevallier et al.*, 2005; *Rödenbeck*, 2005]. One of the main advantages is the feasibility to include more observations from different sources into the system. This helps to decrease the under-determination of the inverse problem [*Rayner*, 2010]. Finally, assimilation allows also for predictions with optimised fluxes.
The **four-d**imensional **var**iational assimilation (4D-Var) method requires the development of the adjoint model, which allows to use information available in the data backwards in time, see Sect. 2.2.2. The Kalman filter includes a posterior covariance calculation, which allows a detailed evalu-

ation of the analysis error, but has additional challenges due to increased computing time. Both methods are discussed in detail in Chapter 2.

Despite all the progress that has been made, the problem of analysing and predicting terrestrial carbon fluxes remains challenging. Obvious obstacles include the spatial heterogeneity of the land use and the different time scales at which biospheric surface-atmosphere fluxes differ. Firstly, the European land use is highly fragmented, actively managed by humans, and often embedded in densely populated areas, which hampers reliable biogenic or anthropogenic flux inventories. Secondly, CO_2 uptake and sequestration by plants has a strong diurnal and seasonal cycle as well as large inter-annual variability [*Goulden et al.*, 1996; *Hollinger et al.*, 2004; *Urbanski et al.*, 2007; *Stoy et al.*, 2009]. Finally, CO_2 release from soil to the atmosphere happens on timescales ranging from seconds to several decades [*Ryan and Law*, 2005]. Therefore, the uncertainty of inferred terrestrial CO_2 fluxes at high spatio-temporal scale remains large, requiring further progress on inversion strategies.

1.3 Objectives of this thesis

This work aims to optimise the estimation of anthropogenic emissions, photosynthesis, and biogenic respiration for each grid cell using the 4D-Var system **EUR**opean **A**ir pollution **D**ispersion-**I**nverse **M**odel (EURAD-IM). Additionally, the initial states of the atmospheric CO_2 concentration are optimised jointly with CO_2 fluxes. The goal is to address uncertainty of modelled CO_2 concentration values introduced by uncertain surface-atmosphere fluxes *and* transport modelling during the previous optimisation period properly, by adjusting the initial CO_2 concentration. This allows a more detailed consideration of uncertainty, but the increased ill-posedness of the problem has to be controlled.
The potential and limits of an advanced model-data-fusion technique are investigated to determine to what extent atmospheric CO_2 in situ data with half-hourly temporal resolution can be exploited to analyse CO_2 surface-atmosphere fluxes. As a result, CO_2 inversion is making progress by a broader access to atmospheric CO_2 measurements, in the context of increasing spatio-temporal resolution in CO_2 inversions. The ability of the system to trace back surface-atmosphere fluxes due to their signal in the atmospheric concentration offers new opportunities for the determination of CO_2 sinks and sources.

The main task of this work is the extension of the EURAD-IM for CO_2

inversions to be able to optimise parameters for anthropogenic emissions, photosynthesis, biogenic respiration, and initial concentration jointly for each grid cell. For this reason, a biogeochemical model (CLM) is modified in order to provide a priori CO_2 fluxes. To handle the high demands on the transport model, the absolute monotone Walcek advection scheme [*Walcek*, 2000] with low numeric diffusion is implemented and investigated. It enables the analysis of small CO_2 flux footprints in the atmospheric background and provides a better understanding of observed CO_2 time series. The adjoint of the advection scheme is implemented with the same desirable properties. A stringent relation between the adjoint variable with respect to the initial values and the surface fluxes is derived. Based on this derivation a new calculation of the error variances for the CO_2 surface fluxes is developed and implemented into EURAD-IM. This serves for an improved optimisation of fluxes to restrict the ill-posed nature of the inverse problem. Finally, the EURAD-IM is modified in order to analyse three fluxes by the usage of atmospheric observations from one species.

The work is structured as follows: Chapter 2 describes the theory of the data assimilation method 4D-Var. In Chapter 3 the modified and applied model system of this work is explained in detail. The model set-up and used observations are described in Chapter 4. Results of the numerical experiments with synthetic and real data are presented in Chapter 5 and final conclusions are given in Chapter 6.

Chapter 2

Theory of Data Assimilation

This chapter gives a short overview of the **d**ata **a**ssimilation (DA) technique **four-d**imensional **var**iational assimilation (4D-Var), which is the most important theoretical concept used in this work. DA concepts for meteorology comprise a vast set of methods to analyse and predict the atmosphere, such that a complete overview of this topic is far beyond the scope of this work, as there exists also no clear definition for DA, see e.g. *Rodgers* [2000]; *Lahoz et al.* [2010] for an overview.
In general, DA is regarded as an analysis technique to combine different information sources in an "optimal way" to describe a dynamical system (in this work the atmosphere) as accurate as possible, usually with the help of a numerical model, which discretises the atmosphere in space and time. The most important information sources used are

(i) the physical knowledge about the dynamical, physical, chemical, and radiative processes, represented by ordinary or partial differential equations (ODE's/PDE's), which are discretised in a numerical model,
(ii) measurements, which are samples of the atmosphere with a certain spatial and temporal scale in dependence of the used observation technique,
(iii) prior knowledge of the current state of the atmosphere, using the information available up to this time.

As stated, this information is combined in an "optimal way", what implicitly means that all information/data sources have an error. The correct specification of corresponding errors is a central topic in DA. Thus, mathematically all data has to be interpreted in a probabilistic sense as random variables.

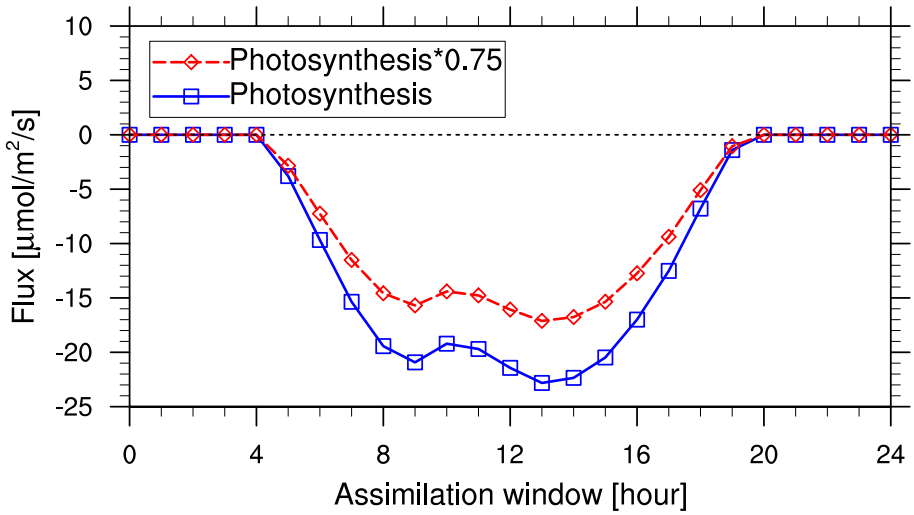


Figure 2.1: *The diurnal time series of photosynthesis (blue line), calculated with WRF-CLM at 24 July 2012 and spatially averaged. The rescaled photosynthesis (red line) preserves the diurnal profile, but changes the absolute amount.*

2.1 Mathematical formulation of the Problem

DA in atmospheric sciences was first applied and developed in NWP, which is regarded as an initial value problem, which is the principal parameter for optimisation in this area. However, for CO_2 the largest uncertainty is introduced by surface-atmosphere fluxes and thus are the main parameters which are optimised in CO_2 inversions. This study seeks to optimise two parameters jointly [*Elbern et al.*, 2007]:

I. the initial atmospheric CO_2 concentration values (iv)
II. a flux factor (ff), which scales the a priori surface-atmosphere CO_2 fluxes for one day, changing the absolute amount of the flux but leaving the diurnal profile unchanged.

The CO_2 surface fluxes are separated into anthropogenic emissions, biogenic respiration (both sources for the atmosphere), and photosynthesis (atmospheric sink). Thus three ff's are optimised. Figure 2.1 visualises the concept of ff optimisation, showing the diurnal cycle of the three CO_2 fluxes. The main idea [*Elbern et al.*, 2000] of optimising the ff's is to reduce the degree of freedom of the subspace of flux rates, by not optimising the fluxes at each time step t_i. Rather it is pointed out that, due to the good knowledge of the diurnal cycle of fluxes (compared to the knowledge

of the absolute amount), one efficient parameter to optimise is their diurnal amplitude. This concept decreases substantially the amount of optimisation parameter such that the problem is less ill-posed than the flux optimisation for each time step [*Elbern et al.*, 2007].

2.1.1 Notation

The same notation as in [*Klimpt et al.*, 2016] will be used in the following. Let $[t_0, t_1, \ldots, t_N]$ be the time steps of the simulated time period $[t_0, t_N]$. Subscripts will always be used as time indices, $(\cdot)_i$ refers to t_i, $(\cdot)_{i+1/2}$ to $(t_i + t_{i+1})/2$, and $(\cdot)_{i,j}$ to the interval $[t_i, t_j]$. $n \in \mathbb{N}$ is the dimension of the model space and $p_i \in \mathbb{N}$ $(i = 0, \ldots, N)$ the dimension of the observation space at t_i. The following notation will be used with small bold letters indicating vectors and capital bold letters indicating matrices:

$\mathbf{x} \in \mathbb{R}^n$:	model state vec.	$\mathbf{y}_i \in \mathbb{R}^{p_i}$:	observation vec.
$\mathbf{f} \in \mathbb{R}^n$:	flux factor vec.	$\mathbf{U} \in \mathbb{R}^{n \times n}$:	fluxes
$(\cdot)^{\mathrm{b}}$:	background state	$(\cdot)^{\mathrm{t}}$:	true state
$(\cdot)^{\mathrm{iv}}$:	initial value	$(\cdot)^{\mathrm{ff}}$:	flux factor
$(\cdot)^{\mathrm{o}}$:	observation		
$\boldsymbol{\epsilon} \in \mathbb{R}^n$ or $\mathbb{R}^{p_i}$:	error term	$\mathcal{M}$:	Model
$\mathcal{H}_i : \mathbb{R}^{2n} \to \mathbb{R}^{p_i}$:	observation operator	$\mathbf{B} \in \mathbb{R}^{n \times n}$:	background error covariance matrix
$\mathbf{R}_i \in \mathbb{R}^{p_i \times p_i}$:	observation error covariance matrix	$\mathbf{K} \in \mathbb{R}^{n \times n}$:	flux factor covariance matrix.

For the sake of clarity the flux factor $\mathbf{f}$ is defined to have the same dimension n as the model state vector $\mathbf{x}$, although $\mathbf{f}$ is formally an element of $\mathbb{R}^{3n}$ since three fluxes are regarded (thus formally are $\mathbf{U}, \mathbf{K} \in \mathbb{R}^{3n \times 3n}$ and $\mathcal{H}_i : \mathbb{R}^{4n} \to \mathbb{R}^{p_i}$). The flux factor vector $\mathbf{f}$ scales the a priori knowledge of the background flux $\mathbf{U}^{\mathrm{b}}$ by an optimisation factor per grid point and per flux type. For notational convenience $\mathbf{U}^{\mathrm{b}} \in \mathbb{R}^{n \times n}$ is defined to be a diagonal matrix. The fluxes are thus ($\mathrm{diag}(\mathbf{a})$ designates a diagonal matrix with entries of the vector $\mathbf{a}$)

$$\mathbf{U}_{i+1/2} = \mathrm{diag}(\mathbf{U}^{\mathrm{b}}_{i+1/2}\,\mathbf{f}) \in \mathbb{R}^{n \times n}, \quad \forall\, i = 0, \ldots, N-1. \tag{2.1}$$

The actual parameter of optimisation will be $\mathbf{g} := \ln(\mathbf{f})$ due to two reasons: (1) The transformation results in the same partial costs for flux factors $\mathbf{g}$

and $\mathbf{1}/\mathbf{g}$. (2) Since $\mathbf{f} > \mathbf{0}$ it can be described by a log-normal distribution, resulting in a Gaussian probability density function for $\mathbf{g}$. *Fletcher and Zupanski* [2006] and *Fletcher* [2010] describe 4D-Var systems with hybrid Gaussian and log-normal distribution in general.
An important step is to use increments of $\mathbf{x}$ and $\mathbf{g}$ [*Courtier et al.*, 1994] as they enable an efficient discretisation (see Section 2.2.2):

$$\delta\mathbf{x}_0 := \mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}, \quad \delta\mathbf{g} := \mathbf{g} - \mathbf{g}^{\mathrm{b}}. \tag{2.2}$$

The function $\mathcal{H}_i$ allows to compare observations $\mathbf{y}_i$ with the model equivalent $\mathcal{H}_i(\mathbf{x}_i^{\mathrm{b}}, \mathbf{g}^{\mathrm{b}})$. Therefore the innovation vector is defined

$$\mathbf{d}_i := \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i^{\mathrm{b}}, \mathbf{g}^{\mathrm{b}}) = \mathbf{y}_i - \mathcal{H}_i\mathcal{M}_{0,i}(\mathbf{x}_0^{\mathrm{b}}, \mathbf{g}^{\mathrm{b}}). \tag{2.3}$$

2.1.2 Assumptions

True values and perfect model

We assume the existence of probabilistic "true" vectors $\mathbf{x}_i^{\mathrm{t}} \in \mathbb{R}^n, \mathbf{g}^{\mathrm{t}} \in \mathbb{R}^n$ and $\mathbf{y}_i^{\mathrm{t}} \in \mathbb{R}^{p_i}$ for $i = 1, \ldots, N$. Our knowledge is represented by deterministic background and observation vectors, which deviate from the true values by random error vectors for the model state $\boldsymbol{\epsilon}^{\mathrm{iv}}$, the flux factors $\boldsymbol{\epsilon}^{\mathrm{ff}}$, and the observations $\boldsymbol{\epsilon}_i^{\mathrm{o}}$ respectively:

$$\boldsymbol{\epsilon}^{\mathrm{iv}} := \mathbf{x}_0^{\mathrm{b}} - \mathbf{x}_0^{\mathrm{t}}, \qquad \boldsymbol{\epsilon}^{\mathrm{ff}} := \mathbf{g}^{\mathrm{b}} - \mathbf{g}^{\mathrm{t}}, \qquad \boldsymbol{\epsilon}_i^{\mathrm{o}} := \mathcal{H}_i(\mathbf{x}_i^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) - \mathbf{y}_i. \tag{2.4}$$

Further, the perfect model assumption will be used

$$\mathbf{x}_i^{\mathrm{t}} = \mathcal{M}_{i-1,i}(\mathbf{x}_{i-1}^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}). \tag{2.5}$$

It follows directly that the determination of $\mathbf{x}_i^{\mathrm{t}}$ reduces to the determination of $\mathbf{x}_0^{\mathrm{t}}$ and $\mathbf{g}^{\mathrm{t}}$. To emphasise this fact the notation $\mathcal{G} := \mathcal{H} \circ \mathcal{M}$ is used:

$$\mathcal{G}_i(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) := \mathcal{H}_i\mathcal{M}_{i-1,i} \cdots \mathcal{M}_{0,1}(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) = \mathcal{H}_i(\mathbf{x}_i^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) \quad \forall i = 1, \ldots N\,.$$

Gaussian distribution and linearity

The random error vectors are assumed unbiased and determine the covariance matrices:

$$\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{iv}}] = 0, \qquad \mathbf{B} := \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{iv}}(\boldsymbol{\epsilon}^{\mathrm{iv}})^{\mathrm{T}}\right], \tag{2.6}$$

$$\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{ff}}] = 0, \qquad \mathbf{K} := \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{ff}}(\boldsymbol{\epsilon}^{\mathrm{ff}})^{\mathrm{T}}\right], \tag{2.7}$$

$$\mathbb{E}[\boldsymbol{\epsilon}_i^{\mathrm{o}}] = 0, \qquad \mathbf{R}_i := \mathbb{E}\left[\boldsymbol{\epsilon}_i^{\mathrm{o}}(\boldsymbol{\epsilon}_i^{\mathrm{o}})^{\mathrm{T}}\right], \forall i, 1, \ldots, N. \tag{2.8}$$

Further, Gaussian distribution is claimed for the error vectors $\boldsymbol{\epsilon}^{\mathrm{iv}}$, $\boldsymbol{\epsilon}^{\mathrm{ff}}$, $\boldsymbol{\epsilon}^{\mathrm{o}}$.

Remark 1 *The true vectors* $\mathbf{x}^t = \mathbf{x}_0^b + \boldsymbol{\epsilon}^{iv}$, $\mathbf{g}^t = \mathbf{g}^b + \boldsymbol{\epsilon}^{ff}$ *as they are defined in this work, have to be understood as random vectors. The interpretation in this work is, that the exact realisation of these true vectors can never be known, but their probability distribution. This should not be confused with the "true" value of a state variable of the atmosphere, which is a deterministic value by definition.*

$\mathcal{N}(\mathbf{m}, \mathbf{A})$ designates the multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{A}$:

$$\boldsymbol{\epsilon}^{\mathrm{iv}} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}), \quad \boldsymbol{\epsilon}^{\mathrm{ff}} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}), \quad \boldsymbol{\epsilon}_i^{\mathrm{o}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i), \tag{2.9}$$

$$p(\mathbf{x}_0^{\mathrm{t}}) = \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_0^{\mathrm{t}} - \mathbf{x}_0^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0^{\mathrm{t}} - \mathbf{x}_0^{\mathrm{b}})\right], \tag{2.10}$$

$$p(\mathbf{g}^{\mathrm{t}}) = \frac{1}{(2\pi)^{n/2}|\mathbf{K}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{g}^{\mathrm{t}} - \mathbf{g}^{\mathrm{b}})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{g}^{\mathrm{t}} - \mathbf{g}^{\mathrm{b}})\right], \tag{2.11}$$

$$p(\mathbf{y}_i \mid \mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) = \frac{1}{(2\pi)^{p_i/2}|\mathbf{R}_i|^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_i - \mathcal{G}_i(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}})\right)^{\mathrm{T}} \mathbf{R}_i^{-1}\left(\mathbf{y}_i - \mathcal{G}_i(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}})\right)\right]. \tag{2.12}$$

Because of $\mathbf{g}^{\mathrm{t}} := \ln(\mathbf{f}^{\mathrm{t}})$, actually $\mathbf{f}^{\mathrm{t}}$ is assumed to have a multivariate log-normal distribution $\mathbf{f}^{\mathrm{t}} \sim \mathrm{LN}(\mathbf{f}^{\mathrm{b}}, \mathbf{K}')$ [*Fletcher*, 2010]. It will be crucial for the incremental formulation of the cost function, that there exist tangent linear approximations $\mathbf{M}$ and $\mathbf{H}$ of $\mathcal{M}$ and $\mathcal{H}$ (and as a consequence $\mathbf{G} := \mathbf{H} \circ \mathbf{M}$ of $\mathcal{G} := \mathcal{H} \circ \mathcal{M}$) for small perturbations of $\mathbf{x}$ and $\mathbf{g}$ respectively:

$$\begin{aligned}
\mathcal{M}_{i,i+1}(\mathbf{x}_i + \delta\mathbf{x}_i, \mathbf{g} + \delta\mathbf{g}) &\approx \mathcal{M}_{i,i+1}(\mathbf{x}_i, \mathbf{g}) + \mathbf{M}_{i,i+1}\begin{pmatrix}\delta\mathbf{x}_0\\ \delta\mathbf{g}\end{pmatrix},\\
\mathcal{H}_i(\mathbf{x}_i + \delta\mathbf{x}_i, \mathbf{g} + \delta\mathbf{g}) &\approx \mathcal{H}_i(\mathbf{x}_i, \mathbf{g}) + \mathbf{H}_i\begin{pmatrix}\delta\mathbf{x}_0\\ \delta\mathbf{g}\end{pmatrix},\\
\mathcal{G}_i(\mathbf{x}_0 + \delta\mathbf{x}_0, \mathbf{g} + \delta\mathbf{g}) &\approx \mathcal{G}_i(\mathbf{x}_0, \mathbf{g}) + \mathbf{G}_i\begin{pmatrix}\delta\mathbf{x}_0\\ \delta\mathbf{g}\end{pmatrix}
\end{aligned} \tag{2.13}$$

This assumption must be well verified [*Vukicevic*, 1991; *Courtier et al.*, 1994; *Bouttier and Courtier*, 2002] and will be used for our purposes in combination with Eqs. (2.2) & (2.3):

$$\mathbf{y}_i - \mathcal{G}_i(\mathbf{x}_0, \mathbf{g}) \approx \mathbf{d}_i - \mathbf{G}_i\begin{pmatrix}\delta\mathbf{x}_0\\ \delta\mathbf{g}\end{pmatrix}. \tag{2.14}$$

Non-correlation

All random error vectors are assumed to be mutually uncorrelated, for all $i, j = 0, \ldots, N, i \neq j$ holds:

$$\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{iv}}(\boldsymbol{\epsilon}^{\mathrm{ff}})^{\mathrm{T}}] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{iv}}(\boldsymbol{\epsilon}_i^{\mathrm{o}})^{\mathrm{T}}] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{ff}}(\boldsymbol{\epsilon}_i^{\mathrm{o}})^{\mathrm{T}}] = \mathbf{0},\ \mathbb{E}[\boldsymbol{\epsilon}_j^{\mathrm{o}}(\boldsymbol{\epsilon}_i^{\mathrm{o}})^{\mathrm{T}}] = \mathbf{0}. \tag{2.15}$$

Uncorrelated random variables which are Gaussian distributed are even independent.

2.1.3 Bayesian formulation and the cost function

Given the a priori distributions for $\mathbf{x}_0^{\mathrm{t}}$ and $\mathbf{g}^{\mathrm{t}}$ (Eqs. (2.10) & (2.11)) and the conditional probability of $\mathbf{y}$ given $\mathbf{x}_0^{\mathrm{t}}$ and $\mathbf{g}^{\mathrm{t}}$ (Eq. (2.12)), the quantity of interest can be derived now, the a posteriori distribution. Applying the Bayesian rule (BR) and the independence (ind.) one gets

$$
\begin{aligned}
& p(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}} \mid \mathbf{y}_1, \ldots, \mathbf{y}_N) && (2.16)\\
& \overset{\text{(BR)}}{=} \frac{p(\mathbf{y}_1, \ldots, \mathbf{y}_N \mid \mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) \cdot p(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}})}{p(\mathbf{y}_1, \ldots, \mathbf{y}_N)} && (2.17)\\
& \overset{\text{(ind.)}}{=} \frac{\prod_{i=1}^{N} p(\mathbf{y}_i \mid \mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}}) \cdot p(\mathbf{x}_0^{\mathrm{t}})\, p(\mathbf{g}^{\mathrm{t}})}{p(\mathbf{y}_1, \ldots, \mathbf{y}_N)} && (2.18)\\
& \propto \exp\Big[-\frac{1}{2}\Big((\mathbf{x}_0^{\mathrm{t}} - \mathbf{x}_0^{\mathrm{b}})^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_0^{\mathrm{t}} - \mathbf{x}_0^{\mathrm{b}}) + (\mathbf{g}^{\mathrm{t}} - \mathbf{g}^{\mathrm{b}})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{g}^{\mathrm{t}} - \mathbf{g}^{\mathrm{b}}) && (2.19)\\
& \qquad + \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathcal{G}_i(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}})\right)^{\mathrm{T}} \mathbf{R}_i^{-1} \left(\mathbf{y}_i - \mathcal{G}_i(\mathbf{x}_0^{\mathrm{t}}, \mathbf{g}^{\mathrm{t}})\right) \Big)\Big] \\
& \overset{(2.14)}{\approx} \exp\Big[-\frac{1}{2}\, \delta\mathbf{x}_0^{\mathrm{T}} \mathbf{B}^{-1} \delta\mathbf{x}_0 - \frac{1}{2}\, \delta\mathbf{g}^{\mathrm{T}} \mathbf{K}^{-1} \delta\mathbf{g} && (2.20)\\
& \qquad - \frac{1}{2} \sum_{i=1}^{N} \left[\mathbf{d}_i - \mathbf{G}_i \begin{pmatrix} \delta\mathbf{x}_0 \\ \delta\mathbf{g} \end{pmatrix} \right]^{\mathrm{T}} \mathbf{R}_i^{-1} \left[\mathbf{d}_i - \mathbf{G}_i \begin{pmatrix} \delta\mathbf{x}_0 \\ \delta\mathbf{g} \end{pmatrix} \right] \Big] \\
& \propto p(\delta\mathbf{x}_0, \delta\mathbf{g} \mid \mathbf{d}_1, \ldots, \mathbf{d}_N)), && (2.21)
\end{aligned}
$$

which is again a Gaussian distribution. The cost function is now defined as the negative logarithm of the a posteriori distribution.

$$
\begin{aligned}
\mathcal{J}(\delta\mathbf{x}_0, \delta\mathbf{g}) &:= -\log(p(\delta\mathbf{x}_0, \delta\mathbf{g} \mid \mathbf{d}_1, \ldots, \mathbf{d}_N)) \\
&= \frac{1}{2}\, \delta\mathbf{x}_0^{\mathrm{T}} \mathbf{B}^{-1} \delta\mathbf{x}_0 + \frac{1}{2}\, \delta\mathbf{g}^{\mathrm{T}} \mathbf{K}^{-1} \delta\mathbf{g} + \\
&\quad \frac{1}{2} \sum_{i=0}^{N} \left[\mathbf{d}_i - \mathbf{G}_i \begin{pmatrix} \delta\mathbf{x}_0 \\ \delta\mathbf{g} \end{pmatrix} \right]^{\mathrm{T}} \mathbf{R}_i^{-1} \left[\mathbf{d}_i - \mathbf{G}_i \begin{pmatrix} \delta\mathbf{x}_0 \\ \delta\mathbf{g} \end{pmatrix} \right]. \qquad (2.22)
\end{aligned}
$$

The aim is to find $\mathbf{x}_0^{\mathrm{a}}$ and $\mathbf{g}^{\mathrm{a}}$, which maximise the a posteriori distribution

$$
(\mathbf{x}_0^{\mathrm{a}}, \mathbf{g}^{\mathrm{a}}) := \arg\max_{\mathbf{x}_0, \mathbf{g}} \big(p(\mathbf{x}_0, \mathbf{g} \mid \mathbf{y}_1, \ldots, \mathbf{y}_N)\big), \qquad (2.23)
$$

or equivalently minimise $\mathcal{J}$. Thus $\mathbf{x}_0^{\mathrm{a}}$ and $\mathbf{g}^{\mathrm{a}}$ can be calculated by setting the gradient of $\mathcal{J}$ to $\mathbf{0}$

$$\nabla\mathcal{J}(\delta\mathbf{x}_0,\delta\mathbf{g}) = \begin{pmatrix}\mathbf{B}^{-1}\delta\mathbf{x}_0\\ \mathbf{K}^{-1}\delta\mathbf{g}\end{pmatrix} - \sum_{i=0}^{N}\mathbf{G}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{G}_i\begin{pmatrix}\delta\mathbf{x}_0\\ \delta\mathbf{g}\end{pmatrix}\right], \tag{2.24}$$

where $\mathbf{G}_i^{\mathrm{T}} = (\mathbf{M}_{0,i})^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}$ is the adjoint operator, with $(\mathbf{M}_{0,i})^{\mathrm{T}}$ modelling backwards in time from t_i to t_0. Since the construction of the adjoint model uses the derivative with respect to the optimisation parameter, we have to distinguish between the adjoint model related to initial values $\mathbf{x}_0$ and flux factors $\mathbf{g}$. See Appendix A.1 for a detailed derivation of $(\mathbf{M}_{0,i})^{\mathrm{T}}$ with respect to the initial values and flux factors.
Calculating $\mathbf{x}_0^{\mathrm{a}}$ and $\mathbf{g}^{\mathrm{a}}$ explicitly gives

$$\begin{pmatrix}\mathbf{x}_0^{\mathrm{a}}\\ \mathbf{g}^{\mathrm{a}}\end{pmatrix} = \begin{pmatrix}\mathbf{x}_0^{\mathrm{b}}\\ \mathbf{g}^{\mathrm{b}}\end{pmatrix} + \widetilde{\mathbf{K}}\hat{\mathbf{d}}, \quad \hat{\mathbf{d}} := (\mathbf{d}_0^T,\ldots,\mathbf{d}_N^T)^T, \tag{2.25}$$

with $\widetilde{\mathbf{K}}$ being the gain (sometimes also Kalman gain) matrix

$$\widetilde{\mathbf{K}} = \left(\begin{pmatrix}\mathbf{B}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}^{-1}\end{pmatrix} + \hat{\mathbf{G}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{G}}\right)^{-1}\hat{\mathbf{G}}^T\hat{\mathbf{R}}^{-1} \tag{2.26}$$

$$\overset{(*)}{=} \begin{pmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}\end{pmatrix}\hat{\mathbf{G}}^T\left(\hat{\mathbf{G}}\begin{pmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{K}\end{pmatrix}\hat{\mathbf{G}}^T + \hat{\mathbf{R}}\right)^{-1}, \tag{2.27}$$

$$\hat{\mathbf{G}} := \begin{pmatrix}\mathbf{H}_0\\ \mathbf{H}_1\mathbf{M}_{0,1}\\ \vdots\\ \mathbf{H}_N\mathbf{M}_{0,N}\end{pmatrix}, \ \hat{\mathbf{R}} := \begin{pmatrix}\mathbf{R}_0 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_N\end{pmatrix}, \tag{2.28}$$

here $(*)$ uses the "Sherman-Morrison-Woodbury formula" to calculate the inverse matrix. The analysis values $\mathbf{x}^{\mathrm{a}}$ and $\mathbf{g}^{\mathrm{a}}$ are called "**B**est **L**inear **U**nbiased **E**stimator" (BLUE). It is called the "best" estimate, because it minimises the variance. The linearity is clear from Eq. (2.25) and unbiased is a consequence of the assumption that the random error vectors $\boldsymbol{\epsilon}$ are unbiased.

2.2 Overview of CO_2 DA concepts

A short overview of the variety of methods to solve the BLUE, the unbiased linear combination of a priori knowledge and observations with minimal

variance, is given. As already mentioned in Section 1.2.3 one widely used simple method is the so- called batch inversion, which solves Eq. (2.25) in one step. If the observation error matrix $\hat{\mathbf{R}}$ can be inverted with low numerical costs (e.g. $\hat{\mathbf{R}}$ is diagonal) and the optimisation state space is small, the gain matrix $\widetilde{\mathbf{K}}$ can be calculated with Eq. (2.26). Contrarily, if the observation space is quite small compared to the model space, it is preferential to use Eq. (2.27) to calculate $\widetilde{\mathbf{K}}$. However, for large model state and observation dimensions, the inversion of the matrices becomes infeasible and the cost function has to be minimised by the use of more efficient minimisation methods.

Two basic approaches exist for DA: sequential methods, using observations made in the past of the analysis and variational assimilation, taking also observations from the future into account. Another classification distinguishes between the processing of observations: intermittent analysis, processing at each time point, or continuous analysis, processing the observations for longer periods [*Bouttier and Courtier*, 2002]. Figure 2.2 used from *Bouttier and Courtier* [2002] visualises the principal differences.

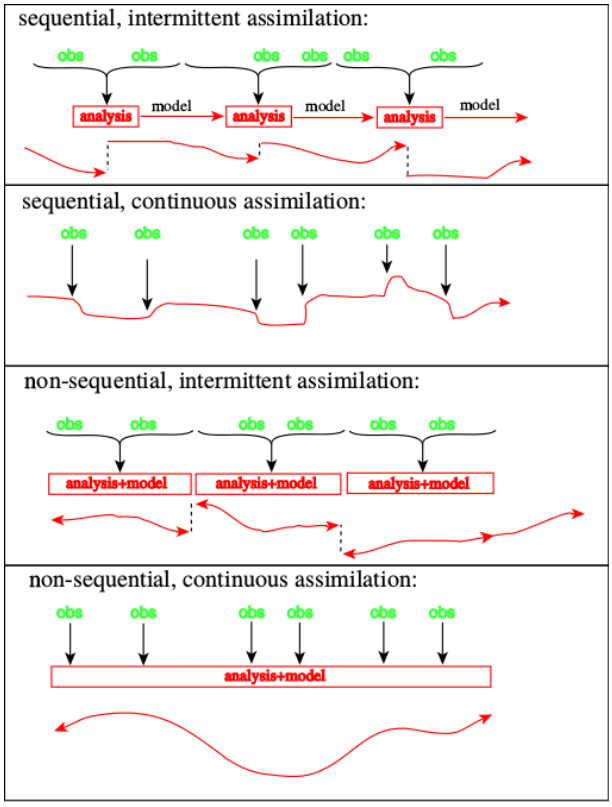


Figure 2.2: *Classification of DA methods from Bouttier and Courtier [2002].*

2.2.1 Sequential DA

The **K**alman **f**ilter (KF) (panel one of Fig. 2.2) [*Kalman*, 1960] was designed for linear problems and is an intermittent method, solving a variation of Eq. (2.25) which uses background error covariance matrices at each time step (the formulae is written for one optimisation parameter only)

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{b}} + \mathbf{K}_i \left(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\right), \tag{2.29}$$

$$\mathbf{K}_i = \mathbf{B}_i\mathbf{H}_i^T(\mathbf{H}_i\mathbf{B}_i\mathbf{H}_i^T + \mathbf{R}_k)^{-1}, \quad \mathbf{x}_{i+1}^{\mathrm{b}} = \mathcal{M}_{i,i+1}(\mathbf{x}_i^{\mathrm{a}}). \tag{2.30}$$

The **e**xtended KF (EKF) [*Kalman and Bucy*, 1961] allows also nonlinear operators $\mathcal{H}$ and $\mathcal{M}$. KF and EKF give also the analysis error covariance matrices for each time step but cannot be practically implemented due to the large dimensions in real applications, which make an inversion of the error matrices impracticable.
A significant step forward towards computational feasibility for real applications was the development of the **en**semble KF (EnKF) by *Evensen* [1994]. It provides an approximation method for the error covariance estimation, based on Monte Carlo methods. Instead of using background error covariance matrices with the size of the model space ($\mathbf{B} \in \mathbb{R}^{n\times n}$), a rank reduction to the size of an ensemble of differing model runs is performed, which can be chosen much smaller than n for atmospheric applications. This method is applicable to non-linear optimisation and several variations of this technique exist, using deterministic or stochastic filters, see *Evensen* [2009] for a detailed review of methods. As the EnKF is easy to implement and is applicable to a wide range of problems, it is used in nearly all fields of earth science. The first usage of EnKF methods for CO_2 flux inversion was as recently as 2005 [*Peters et al.*, 2005].
The **K**alman **S**moother (KS), (panel two of Fig. 2.2) is described in estimation theory books, e.g. *Gelb* [1974] and is a generalisation of the KF which assimilates also future observations. It is a classic result that the KS is equivalent to the weak constraint four-dimensional variational method if the same information statistics are used [*Ménard and Daley*, 1996].

2.2.2 Variational methods

Variational DA [*Sasaki*, 1970*a*;*b*; *Lorenc*, 1986; *Talagrand and Courtier*, 1987] comprises a class of methods which uses optimisation parameters explicitly as minimiser for the cost function Eq. (2.22). The calculation of the gain matrix $\widetilde{\mathbf{K}}$ is avoided, instead the minimum of the cost function $\mathcal{J}$ (Eq. (2.22)) is sought iteratively with the gradient of $\mathcal{J}$ (Eq. (2.24)).

The minimisation is done efficiently with the help of the adjoint approach [*Errico*, 1997] using $\mathbf{M}^T$ and $\mathbf{H}^T$: *Talagrand* [2010] points out that the determination of $\nabla\mathcal{J}$ requires one forward run of the model ($\mathbf{x}_N = \mathcal{M}_{0,N}(\mathbf{x}_0)$) and one backwards run of the adjoint model $\mathbf{M}^T$. It does take at most four times more arithmetic operations for the calculation of $\nabla\mathcal{J}$, than the calculation of $\mathcal{J}$. "It is this fact" [the efficient calculation of $\nabla\mathcal{J}$ with the adjoint method] "that made variational assimilation possible at all in the first place" [*Talagrand*, 2010]. Quasi-Newton methods approximate the inverse Hessian of $\mathcal{J}$ with a sequence of its gradient and minimise poorly conditioned problems much faster than methods of steepest descent.

Three-dimensional **var**iational assimilation (3D-Var) (panel three of Fig. 2.2) comprises the observations of a certain time interval at one time point, which is a reasonable assumption if the interval is short enough or the driving processes are almost stationary [*Daley and Barker*, 2001]. The time dependency for the cost function (Eq. (2.22)) is eliminated and the development of an adjoint model is not necessary.

The 4D-Var method (panel four of Fig. 2.2) minimises the misfit between a sequence of model states and observations for a certain assimilation window. "As such, and contrary [...] to sequential algorithms for assimilation, it propagates the information contained in the data both forward and backwards in time" [*Talagrand*, 2010]. Thus the usage of the adjoint model allows an estimation of the influence of observations on the analysis [*Cardinali et al.*, 2004], as well as on the forecast [*Langland and Baker*, 2004]. The development of the incremental approach (Eq. (2.2)) by *Courtier et al.* [1994] is of great importance for the practical implementation of 4D-Var. The idea is to simplify the dynamical model $\mathcal{M}$ not directly, but its tangent linear version (Eq. (2.13)) to facilitate the calculation of the adjoint model. "It is the incremental method which, after the adjoint method, makes variational assimilation feasible" [*Talagrand*, 2010].

Variational assimilation was first used for operational forecasts at the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Klinker et al.*, 2000] and improved predictions remarkably. Meteorological services of France, Japan, United Kingdom, Canada, and China followed. Variational assimilation is applied in oceanography [*Thacker and Long*, 1988; *Bennett*, 1992], surface hydrology [*Calvet et al.*, 1998], seismology [*Tromp et al.*, 2005], atmospheric chemistry [*Fisher and Lary*, 1995; *Elbern et al.*, 1997], and CO_2 flux inversion [*Chevallier et al.*, 2005; *Engelen and McNally*, 2005].

Several properties make the 4D-Var method attractive for such a wide field

of applications. 4D-Var acts as smoother during the assimilation window [*Thepaut and Courtier*, 1991]. Abandoning the perfect model assumption allows to account for model errors, which is denoted as weak constraint 4D-Var [*Trémolet*, 2007] and allows the use of longer assimilation windows. An extension to non-Gaussian distributions [*Fletcher*, 2010] allows the handling of a larger class of problems. *Elbern et al.* [2007] developed the combined estimation of initial values and flux factors. The principal benefit of this method is shown in Fig. 2.3. The main drawback of the

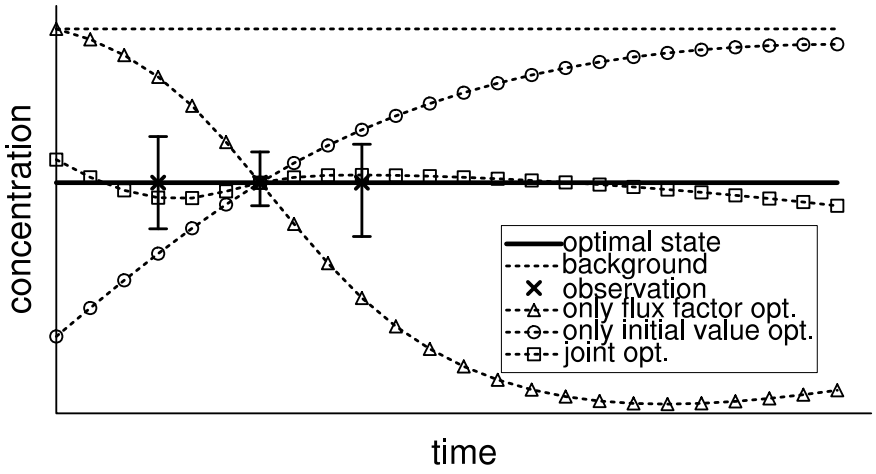


Figure 2.3: *Given the background run (dashed line without markers) and the observations (crosses) three analysis runs are compared with respect to the optimal state (solid line). Optimising only initial values (dashed line with circles) leads to strong forcing at the start of the assimilation window. Optimising only flux rates (dashed line with triangles) can not change the initial values and may force the analysis to give more weight at the starting time than desirable. The joint optimisation (dashed line with quadrats) combines the benefits of both approaches, leading to a smoother analysis.*

4D-Var method is the large effort for development and testing of the adjoint model [*Talagrand*, 2010], which leaves only a few groups using 4D-Var compared to EnKF. The reward of this effort is a consistent analysis of the CO_2 dispersion, by quantifying the amount of CO_2 sinks and sources backwards in time. This cannot be achieved by KF methods, only the KS is theoretical equivalent to weak constraint 4D-Var.

Chapter 3

Model description

This chapter gives an overview of the used models in this work, illustrated in Fig. 3.1. The main model is the 4D-Var system EURAD-IM (Sect. 3.4), optimising CO_2 fluxes and concentration values. It requires meteorological input from WRF (Sect. 3.1) to simulate the atmospheric transport. Biogenic fluxes are provided by the land surface model CLM (Sect. 3.2) and anthropogenic emissions by a database of TNO (Sect. 3.3). The simulations are compared with observations to derive optimised fluxes and initial concentration values.

3.1 Meteorological model WRF

The **W**eather **R**esearch and **F**orecasting (WRF) model version 3.6.1 [*Skamarock et al.*, 2008] is used as the meteorological driver for the EURAD-IM in this work. WRF is a mesoscale NWP system which uses fully compressible non-hydrostatic Euler equations in flux-form. In this study WRF processes six-hourly ECMWF reanalysis data [*Dee et al.*, 2011], interpolated to $0.225° \times 0.225°$ grid resolution. The terrain-following vertical σ-coordinates [*Laprise*, 1992] are used up to a height of 100 hPa in 23 layers. Several preprocessing systems are included in WRF, which prepare external data and the ECMWF reanalysis for the numerical processing. To process MODIS **L**eaf **A**rea **I**ndex (LAI) and land use categories with the land surface model CLM (see next Section), minor changes in five subroutines were necessary. The most important prognostic variables for this work are hourly horizontal and vertical wind components, moisture, perturbation potential temperature, and perturbation geopotential.

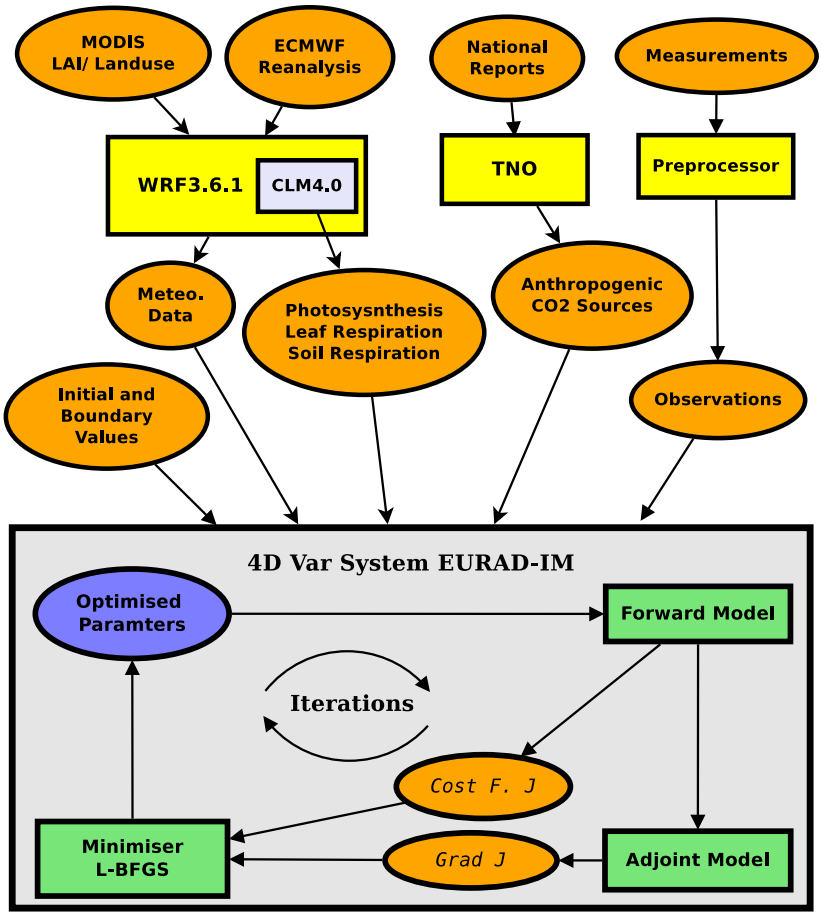


Figure 3.1: *Overview of the used models and data in this work. Models are presented as rectangles, data as ellipsoids.*

3.2 Land surface model CLM

The **C**ommunity **L**and **M**odel (CLM) 4.0 [*Oleson et al.*, 2010] is the land surface model of the **C**ommunity **C**limate **S**ystem **M**odel (CCSM), which combines models for the atmosphere, land surface, oceans, and glaciers to simulate the earth's climate. The CLM is a combination of three land models simulating the biophysical processes for soil-snow-vegetation columns. In addition to climate simulations, it is also suitable for regional applications with high temporal resolution. The version CLM 4.0 is available as land surface scheme in WRF, but the modules **D**ynamic **G**lobal **V**egetation **M**odel (DGVM) and **C**arbon-**N**itrogen (CN) module cannot be applied directly. These two modules allow for the enhanced modelling of carbon processes in soil and biosphere, focusing on long time periods (months to decades).

In this work several changes in the source code are implemented to allow the usage of MODIS LAI and land use categories as the standard-version uses simplified table values. Leaf respiration and soil respiration are additionally implemented in this work according to *Collatz et al.* [1991] (implementation taken from Markus Übel) and to *Lloyd and Taylor* [1994] respectively. The main concept of CLM is the calculation of the corresponding biological, chemical, and physical processes at different hierarchical levels, which define soil properties and plant physiology up to the stomatal structure of single leaves.

3.2.1 Modelling of photosynthesis and leaf respiration

Photosynthesis is a biochemical process which converts low energy molecules in combination with sunlight into high energy biomolecules. It is performed by plants, most algae, and cyanobacteria and is globally the largest flux of CO_2 from the atmosphere into ecosystems. In the first step photosynthesis consists of the transformation of light energy with chlorophyll into chemical energy. In the second step a light-independent reaction (Calvin Cycle) converts CO_2 and H_2O into carbohydrates and oxygen. The net carbon dioxide assimilation at leaf-level is determined by photosynthesis (CO_2 flux into the leaf), photorespiration, and mitochondrial respiration (CO_2 flux out of the leaf) [*Bernacchi et al.*, 2013].
Farquhar et al. [1980] described a mechanical leaf model, which played a fundamental role (more than 5000 citations up to now) in understanding and quantifying photosynthesis of C_3 plants (about 95 % of all land plants). This model is used in CLM to calculate photosynthesis. The classification of plants into C_3 and C_4 (e.g. sugarcane, maize) species distinguishes the stomata behaviour for different climatological situations. In this work only C_3 plants are regarded. These are predominant in temperate zones and perform less photosynthesis than C_4 plants under warm and dry conditions to avoid drying out.
The model of *Farquhar et al.* [1980] limits the assimilation of CO_2 by leaves either through the (i) available amount of radiation w_j or the (ii) enzyme **R**ib**U**lose-1,5-**BI**pho**S**phate-/**C**arboxylase/-**O**xygenase (RUBISCO) w_c. (iii) *Sharkey* [1985] and *Collatz et al.* [1991] determined another limit w_e, the rate at which inorganic phosphate is utilised. The photosynthesis is the minimum of these three quantities

$$A = \min(w_j, w_c, w_e). \tag{3.1}$$

The light limited rate w_j (i) depends on the CO_2 concentration inside the leaf C_i, the compensation point Γ^* (the concentration at which photosynthetic carbon uptake is equal to photorespiratory CO_2 release), the **P**hotosynthetic **A**ctive **R**adiation I_{PAR} [Wm^{-2}], and the quantum efficiency α $\left[\frac{\mathrm{mol(CO_2)}}{\mathrm{mol(photons)}}\right]$

$$w_j = \frac{4.6\,(C_i - \Gamma^*)\,I_{PAR}\,\alpha}{C_i + 2\Gamma^*}. \tag{3.2}$$

The RUBISCO limited photosynthesis w_c (ii) depends also on the internal O_2 concentration O_i, the Michaelis-Menten constants [*Dowd and Riggs*, 1965] for CO_2 and O_2 K_c, and K_o, and on the maximum rate of carboxylation $V_{c,max}$

$$w_c = \frac{V_{c,max}\,(C_i - \Gamma^*)}{C_i + K_c(1 + O_i/K_o)}. \tag{3.3}$$

w_e (iii) is simply estimated to be

$$w_e = 0.5\,V_{c,max}. \tag{3.4}$$

The proper estimation of the parameters used in Eqs (3.2)-(3.4), which are highly dependent on temperature, is decisive for the mechanistic leaf model. See *Bernacchi et al.* [2013] for an overview of the different estimation approaches. Finally, the net CO_2 leaf assimilation A_n subtracts the mitochondrial leaf respiration $R_d = 0.015\,V_{c,max}$ from the photosynthesis

$$A_n = A - R_d = \min(w_j, w_c, w_e) - 0.015\,V_{c,max}. \tag{3.5}$$

The upscaling of CLM from leaf-level to larger areas uses the LAI and distinguishes between sunlit and shaded leaves in the processing. The simulated photosynthesis and leaf respiration rates show thus similar pattern like the MODIS LAI, see Figure 3.2.

3.2.2 Modelling of soil respiration

This work parameterises the heterotrophic soil respiration with a simple temperature function introduced by *Lloyd and Taylor* [1994] which is equal for all soil types. The cited work showed that the best unbiased fit to 15 different data sets was obtained with an Arrhenius type equation, which

includes temperature dependence of the activation energy (E_0). The respiration is obtained by

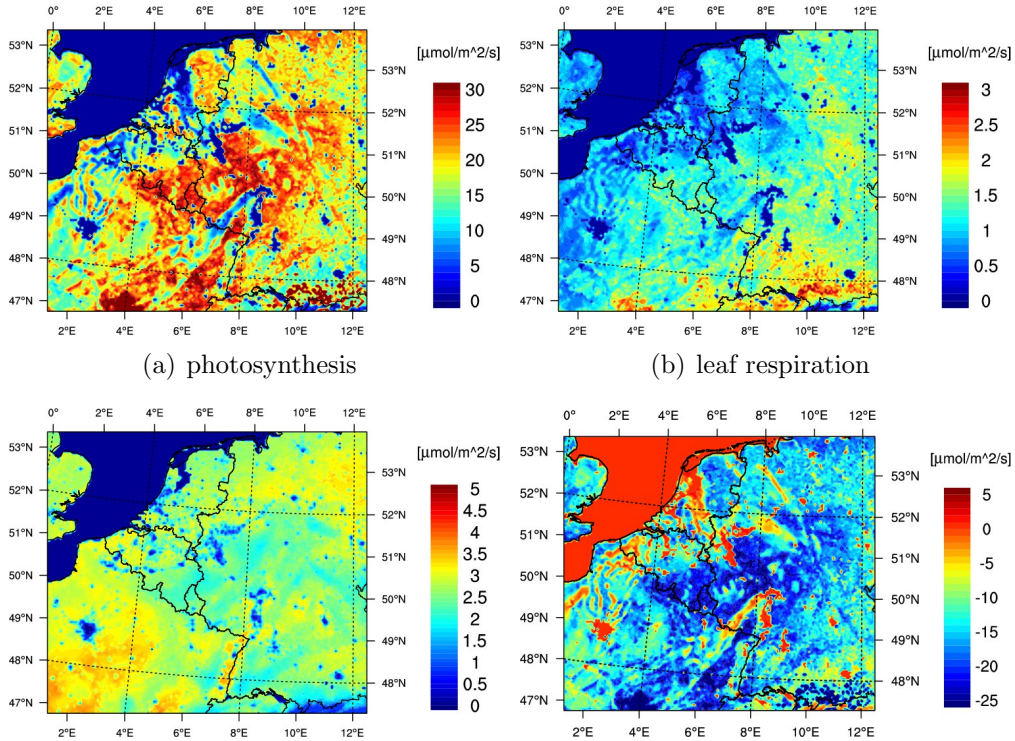
$$R = R_{10} \exp\left\{\frac{E_0}{\mathcal{R}}\left(\frac{1}{56.15} - \frac{1}{T - T_0}\right)\right\}, \tag{3.6}$$

with R_{10} being the respiration rate at reference temperature 10°C, $\mathcal{R}$ the universal gas constant, T_0 equals 227°K and T the soil temperature. The temperature of soil layer four is used in the CLM model (15 layers in total), which lies approximately in 12 cm depth and has a thickness of 7.5 cm. The parameter R_{10} is one optimisation parameter of the 4D-Var analysis, since the flux factor scales the respiration linearly.

Figure 3.2 shows the three simulated biogenic fluxes and the net ecosystem exchange (NEE), as the sum of soil and leaf respiration minus photosynthesis.

3.3 Anthropogenic CO_2 emission inventory

Anthropogenic emissions, introduced as a first guess for inversion, are taken from **T**oegepast **N**atuurwetenschappelijk **O**nderzoek (TNO) inventory (H. Denier van der Gon, personal communication, May 2013), following the methodology outlined in *Kuenen et al.* [2014] and *Pouliot et al.* [2012] for air pollutants. European national reports per source sector are evaluated, afterwards a gap filling and an error correction are executed. TNO provides point and area (1/8° × 1/16° lon-lat horizontal resolution) sources of CO_2 split into ten source categories over Europe for the years 2000-2011. A correction factor from 2011 to 2012 (the modelled time period of this work is June 2012) is applied. Annual values are disaggregated for each source category in dependence of month, day and hour (see also *Memmesheimer et al.* [1995] for details). Hourly values are interpolated to model time step which is 300 s, 120 s and 60 s in dependence of the model resolution (15, 5, and 1 km horizontal resolution). Besides time-profiles, emissions are also splitted vertically, in accordance to plume-rise estimations for different source categories [*EMEP*, 2013]. Area sources are scaled down horizontally to 5 km and 1 km resolution, using land cover and infrastructure information with the **G**eographic **I**nformation **S**ystem (GIS), shown in Figure 3.3 for the CO_2 emissions from road traffic.



(a) photosynthesis (b) leaf respiration

(c) soil respiration (d) NEE

Figure 3.2: *Biogenic fluxes modelled with CLM 4.0. Fluxes are temporal mean values of June 2012 at 12 UTC. The horizontal resolution is 5* km*. Photosynthesis and leaf respiration reflect the spatial structure of the LAI, while the soil respiration is in dependence of the shallow soil temperature. Photosynthesis is opposed to the other fluxes.*

3.4 4D-Var system EURAD-IM

The **EUR**opean **A**ir pollution **D**ispersion-**I**nverse **M**odel (EURAD-IM) [*Elbern et al.*, 2007; *Goris and Elbern*, 2015], is an Eulerian mesoscale non-hydrostatic chemistry transport model, for the simulation of gas phase, aerosols and pollen. It is used for operational analyses in 3D-Var mode and for campaign analyses in 4D-Var mode. A nesting configuration with 15 , 5, and 1 km horizontal resolution is used in order to derive enhanced boundary values. Detailed information about the nested model domains and the set-up of initial and boundary values can be found in Chapter 4.

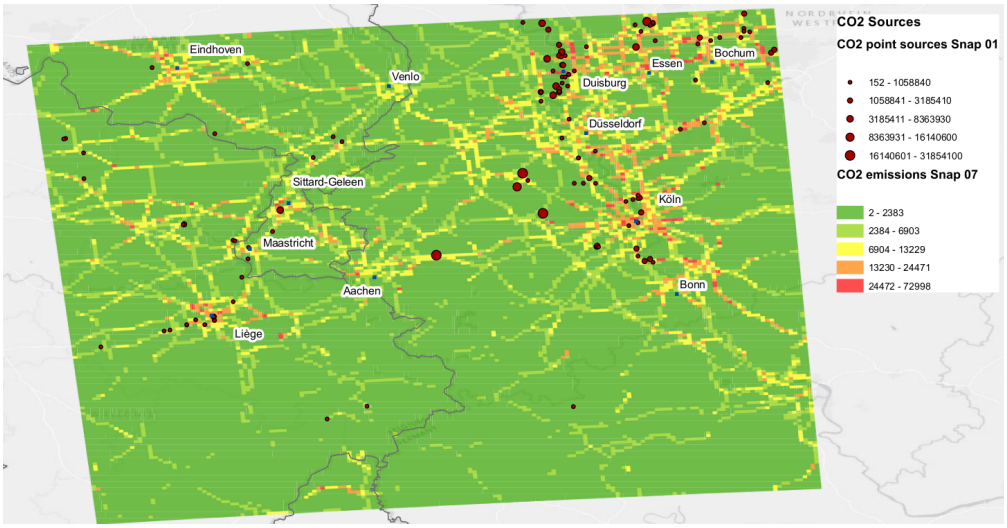


Figure 3.3: *Map of CO_2 emissions including source category 1 (power generation, brown dots) and source category 7 (road traffic, green to red area plots). The emissions from road traffic are scaled down to 1* km *horizontal resolution. The emissions from power generation are given as point sources. The largest power plants depicted are Weisweiler, Niederaußem, Frimmersdorf, and Neurath (all brown coal) located within the Aachen, Köln, Düsseldorf triangle. Figure courtesy of Erna Bem.*

3.4.1 Forward model of EURAD-IM for CO_2

The transport-diffusion equation describes the dispersion of the mixing ratio of CO_2 as a passive tracer $C(\mathbf{r}, t)$ in dependence of spatial location $\mathbf{r} \in \mathbb{R}^3$ and time $t \in [t_0, t_N]$. It can be written as [*Elbern and Schmidt*, 2001]

$$\frac{\partial C}{\partial t} + \underbrace{(\mathbf{u} \cdot \nabla)C}_{\text{advection}} - \underbrace{\frac{1}{\varrho}\nabla\ (\varrho \mathbf{K}^{\mathrm{e}} \nabla C)}_{\text{turbulent diffusion}} = \underbrace{e}_{\text{flux}} . \tag{3.7}$$

The equation includes terms for advection, turbulent diffusion, and additional fluxes. Here $\mathbf{u}$ designates the wind field, ϱ the air density and $\mathbf{K}^{\mathrm{e}}$ is the eddy diffusion tensor. Sinks and sources to the atmosphere, consisting of anthropogenic emissions, biogenic respiration, and photosynthesis, are aggregated to the flux e.

Physical and chemical processes in EURAD-IM are implemented by an operator split method, due to different numerical characteristics of the major transport processes advection (hyperbolic) and turbulent diffusion (parabolic) [*Yanenko*, 1971]. The operators for advection and diffusion are processed for time periods t_i to $t_{i+1/2}$ and $t_{i+1/2}$ to t_{i+1} while fluxes are inserted

instantaneously at time points $t_{i+1/2}$. A symmetric order of the operators $\mathbf{A}^1_{i,i+1/2} \in \mathbb{R}^{n\times n}$ (advection in sequence x-y-z), $\mathbf{A}^2_{i+1/2,i+1} \in \mathbb{R}^{n\times n}$ (advection in sequence z-y-x), $\mathbf{D}_{i,i+1/2}, \mathbf{D}_{i+1/2,i+1} \in \mathbb{R}^{n\times n}$ (turbulent diffusion), and $\mathbf{F}^{\mathbf{f}}_{i+1/2} \in \mathbb{R}^{n\times n}$ (fluxes) is chosen to suppress numerical approximation errors:

$$\mathbf{x}_{i+1} = \mathbf{A}^2_{i+1/2,i+1}\mathbf{D}_{i+1/2,i+1}\mathbf{F}^{\mathbf{f}}_{i+1/2}\mathbf{D}_{i,i+1/2}\mathbf{A}^1_{i,i+1/2}\mathbf{x}_i. \tag{3.8}$$

The flux operator $\mathbf{F}^{\mathbf{f}}$ is defined by the background flux $\mathbf{U}^{\mathrm{b}}$ times the flux factor $\mathbf{f}$ ($\triangle t := t_{i+1} - t_i$)

$$\mathbf{F}^{\mathbf{f}}_{i+1/2}(\mathbf{D}_{i,i+1/2}\mathbf{A}^1_{i,i+1/2}\mathbf{x}_i) := (\mathbf{D}_{i,i+1/2}\mathbf{A}^1_{i,i+1/2}\mathbf{x}_i) + \triangle t\mathbf{U}^{\mathrm{b}}_{i+1/2}\,\mathbf{f}. \tag{3.9}$$

It may again be noted, that $\mathbf{U}^{\mathrm{b}}_{i+1/2}\,\mathbf{f}$ represents three fluxes, each with an independent flux factor for anthropogenic emissions, photosynthesis, and respiration respectively.

Remark 2 *The correct writing for the flux operator and Eq.* (3.9) *is (*$\boldsymbol{F}^{\mathbf{f}}$ *is not a linear operator in Eq.* (3.8) *and* (3.9)*)*

$$\widetilde{\boldsymbol{F}}_{i+1/2} := \begin{pmatrix} \boldsymbol{I} & \triangle t\,\mathbf{U}^b_{i+1/2} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \tag{3.10}$$

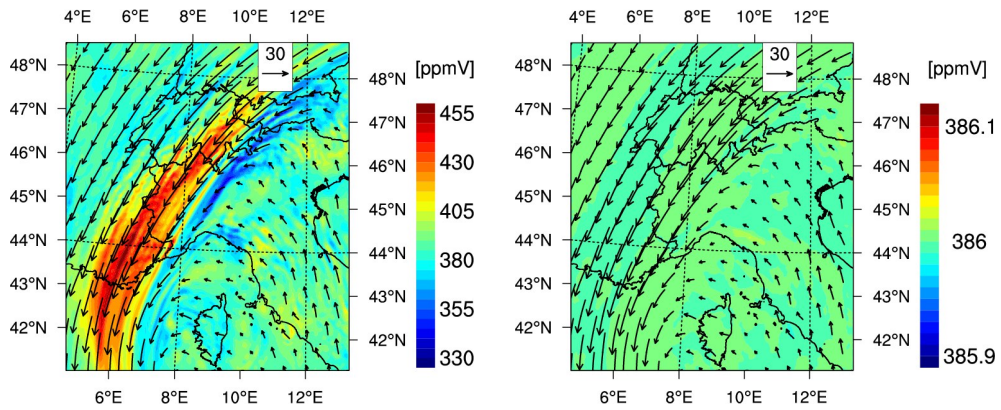
$$\widetilde{\boldsymbol{F}}_{i+1/2}\begin{pmatrix} \boldsymbol{D}_{i,i+1/2}\boldsymbol{A}^1_{i,i+1/2}\mathbf{x}_i \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_{i,i+1/2}\boldsymbol{A}^1_{i,i+1/2}\mathbf{x}_i + \triangle t\mathbf{U}^b_{i+1/2}\,\mathbf{f} \\ \mathbf{f} \end{pmatrix}. \tag{3.11}$$

An extension to three fluxes is straightforward with this notation, compare also with Eq. (A.3) *in Appendix A.1.*

In the case of strong spatial gradients of the wind speed, numerical experiments performed in this work with a non-monotone advection scheme, display strong deterioration of CO_2 mixing ratio, see Fig. 3.4. To account for this problem, the Walcek scheme [*Walcek*, 2000] is implemented in this work to solve the advection process. This scheme specifies dimensional dependent densities, which allows splitting of the forward equation into one-dimensional realisations. The Walcek scheme solves the transport equation for one dimension

$$\frac{\partial C}{\partial t} = -\frac{\partial\,(u\cdot C)}{\partial x} \tag{3.12}$$

based on the van Leer algorithm [*Van Leer*, 1977] with monotonic constraints and a flux adjustment around local mixing ratio extrema. Therefore



(a) non monotone advection scheme (b) Walcek advection scheme

Figure 3.4: *Zoom of the 5* km *horizontal resolution grid, started with constant mixing ratio after four hour run-time at vertical layer 19 (*$\sim$ 4775 − 6100 m*) barely influenced from surface fluxes. The wind field is normalised to 30* m/s.

mass is aggregated numerically around the extrema and numerical diffusion is reduced. The Walcek scheme is absolute monotone, positive definite, of order two, produces only low numeric diffusion, and is numerical efficient. Although an exact parallelisation of the Walcek scheme is numerically not efficient, a good approximation with a small overlap of two grid cells is implemented in this work, which preserves the numerical efficiency.

3.4.2 Adjoint EURAD-IM model

The adjoint model $\mathbf{M}^{\mathrm{T}}$ discretises the adjoint of the forward PDE, Eq. (3.7). Let C^* designate the adjoint variable, then the adjoint PDE has the form

$$-\frac{\partial C^*}{\partial t} - \nabla(\mathbf{u}C^*) - \nabla\left(\varrho\mathbf{K}^{\mathrm{e}}\nabla\frac{C^*}{\varrho}\right) = \mathbf{0}. \tag{3.13}$$

The flux factor $\mathbf{f} = \mathrm{e}^{\mathbf{g}}$ remains constant during the application of the forward model. The adjoint model $(\mathbf{M})^{\mathrm{T}}$ instead, changes the adjoint flux factor $\mathbf{f}^*$ during the background run. In contrast to Eq. (3.8), $\mathbf{x}_i^*$ *and* $\mathbf{f}_i^*$ change with the innovation forcing. The EURAD-IM backward model applies again an operator splitting to solve Eq. (3.13):

$$(\mathbf{M}_{i,i+1})^{\mathrm{T}}\begin{pmatrix}\mathbf{x}_{i+1}^*\\ \mathbf{f}_{i+1}^*\end{pmatrix} = \begin{pmatrix}(\mathbf{T}_{i,i+1})^{\mathrm{T}}\mathbf{x}_{i+1}^*\\ \triangle t\,\mathbf{U}_{i+1/2}^{\mathrm{b}}\,(\mathbf{T}_{i+1/2,i+1})^{\mathrm{T}}\mathbf{x}_{i+1}^* + \mathbf{f}_{i+1}^*\end{pmatrix}, \tag{3.14}$$

with $\mathbf{T}_{i,i+1}$ merging the transport operators from time t_i to t_{i+1}

$$\begin{aligned} \mathbf{T}_{i,i+1} &:= \mathbf{A}^2_{i+1/2,i+1}\mathbf{D}_{i+1/2,i+1}\mathbf{D}_{i,i+1/2}\mathbf{A}^1_{i,i+1/2}, \\ \mathbf{T}_{i+1/2,i+1} &:= \mathbf{A}^2_{i+1/2,i+1}\mathbf{D}_{i+1/2,i+1}. \end{aligned} \tag{3.15}$$

See Appendix A.1 for a detailed derivation of Eq. (3.14).
The adjoint diffusion operator $\mathbf{D}^{\mathrm{T}}$ and adjoint flux operator $(\mathbf{F}^{\mathbf{f}})^{\mathrm{T}}$ are constructed with the automatic differentiation tool Tapenade [*Hascoet and Pascual*, 2013] from the discretisation of the forward PDE Eq. (3.7): The derivatives of the functions specified by $\mathbf{D}$ and $\mathbf{F}^{\mathbf{f}}$ are calculated applying the chain rule to achieve the tangent linear of the forward operators. Afterwards, the linearised code is transposed resulting in $\mathbf{D}^{\mathrm{T}}$ and $(\mathbf{F}^{\mathbf{f}})^{\mathrm{T}}$.
To construct the adjoint of the advection operators $(\mathbf{A}^1)^{\mathrm{T}}$ and $(\mathbf{A}^2)^{\mathrm{T}}$ a different approach has been chosen. The Walcek scheme is not differentiable due to the monotonicity and automatic differentiation is not directly applicable to obtain the adjoint code. Following the suggestion of *Gou and Sandu* [2011], the continuous adjoint for advection has been used. The adjoint PDE of Eq. (3.12)

$$-\frac{\partial C^*}{\partial t} = \frac{\partial\left(u \cdot C^*\right)}{\partial x} \tag{3.16}$$

can also be approximated by the forward Walcek scheme, if the reverse winds and a rescaling for C^* is used. For 4D-Var optimisation problems *Gou and Sandu* [2011] also suggest to construct adjoint routines for diffusion and advection as it is done here: automatic differentiation for diffusion and the continuous adjoint for advection.

3.4.3 Preconditioning of the cost function and minimisation

The background error covariance matrices (BECM) $\mathbf{B}$ and $\mathbf{K}$ are designed to capture the spatial correlation between initial values and surface fluxes respectively. Increasing radii of influence leads to larger condition numbers of $\mathbf{B}$, which make efficient minimisation more difficult. Using the increments $\delta\mathbf{x}_0$ and $\delta\mathbf{g}$ multiplied with the inverse square roots of $\mathbf{B}$ and $\mathbf{K}$, a

formulation of $\mathcal{J}$ equivalent to Eq. (2.22) is possible:

$$\begin{aligned}
\mathbf{v} &:= \mathbf{B}^{-1/2}\delta\mathbf{x}_0 = \mathbf{B}^{-1/2}(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}) \text{ and } \mathbf{w} := \mathbf{K}^{-1/2}\delta\mathbf{g} = \mathbf{K}^{-1/2}(\mathbf{g} - \mathbf{g}^{\mathrm{b}}), \\
\mathcal{J}(\mathbf{v},\mathbf{w}) &= \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \\
&\quad \frac{1}{2}\sum_{i=0}^{N}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right]^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right].
\end{aligned} \tag{3.17}$$

The gradient of $\mathcal{J}$ with respect to $\mathbf{v}$ and $\mathbf{w}$ reads then

$$\nabla\mathcal{J}(\mathbf{v},\mathbf{w}) = \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathbf{K}^{\mathrm{T}/2}\end{pmatrix}\sum_{i=0}^{N}(\mathbf{M}_{0,i})^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right]. \tag{3.18}$$

One iteration cycle of the 4D-Var system EURAD-IM to jointly analyse initial values and flux factors consists of:

1. The forward run, solving Eq. (3.7). Our a priori knowledge $\mathbf{x}_0^{\mathrm{b}}$, $\mathbf{g}^{\mathrm{b}}$ is used as first guess, allowing to apply only $\mathbf{B}^{1/2}$, $\mathbf{K}^{1/2}$ and $\mathbf{B}^{\mathrm{T}/2}$, $\mathbf{K}^{\mathrm{T}/2}$.

2. The calculation of the gradient of $\mathcal{J}$ during the backward time loop and its preconditioning, resulting in:
$\begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathbf{K}^{\mathrm{T}/2}\end{pmatrix}\sum_{i=0}^{N}(\mathbf{M}_{0,i})^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right]$.

3. The application of limited memory quasi-Newton minimiser L-BFGS [*Liu and Nocedal*, 1989; *Nocedal*, 1980], modified for parallel usage, which calculates a new vector $(\mathbf{v}^{\mathrm{T}}, \mathbf{w}^{\mathrm{T}})^{\mathrm{T}}$ and saves it for the next iteration.

4. The determination of improved initial states $\mathbf{x}_0 = \mathbf{B}^{1/2}\mathbf{v} - \mathbf{x}_0^{\mathrm{b}}$ and flux factors $\mathbf{g} = \mathbf{K}^{1/2}\mathbf{w} - \mathbf{g}^{\mathrm{b}}$ that can now be used for step 1.

Steps 1 to 4 are repeated until the convergence criterion is fulfilled. As stop criterion, the two most recent evaluations of the cost function must differ by less than 1.0 E−6 for all model domains.

3.4.4 Modelling of background error covariance matrices

Both BECM's $\mathbf{B}$ and $\mathbf{K}$ are modelled using the diffusion approach. Following *Weaver and Courtier* [2001] a **g**eneralised **d**iffusion **e**quation (GDE) is

formulated with a polynomial in the Laplacian, which is self-adjoint. This allows a factorisation of $\mathbf{B}$ into $\mathbf{B}^{1/2}\mathbf{B}^{\mathrm{T}/2}$ and analogue for $\mathbf{K}$. Anisotropic and inhomogeneous spatial radii of influence can be modelled in this manner as well.

3.4.4.1 Initial value background error covariance matrix

Modelling of $\mathbf{B}$ follows the method in *Elbern et al.* [2007] and is given in short form here for convenience. The background error standard deviation depends on a constant c^{iv} and the initial concentration (r specifies the location):

$$\sqrt{(\mathbf{B}_{r,r})} = \mathbf{x}_0^{\mathrm{b}}(r) \cdot c^{\mathrm{iv}}. \tag{3.19}$$

Elbern et al. [2007] use an additional height dependent factor which increases the error exponentially with model height. This factor is not applicable for CO_2, as the lower layers, influenced by surface fluxes, are much more heterogeneous than better mixed high model layers.
The decomposition of $\mathbf{B}$ (and analogue for $\mathbf{K}$) is given by

$$\begin{aligned}
\mathbf{B} &= \Sigma\mathbf{C}\Sigma && (3.20)\\
&\overset{(@)}{=} \Sigma\mathbf{C}^{1/2}\mathbf{C}^{\mathrm{T}/2}\Sigma && (3.21)\\
&= (\Sigma\Lambda\mathbf{L}_v^{1/2}\mathbf{L}_h^{1/2}\mathbf{W}^{-1/2})(\mathbf{W}^{-1/2}\mathbf{L}_h^{\mathrm{T}/2}\mathbf{L}_v^{\mathrm{T}/2}\Lambda\Sigma) && (3.22)\\
&= \mathbf{B}^{1/2}\mathbf{B}^{\mathrm{T}/2}, && (3.23)
\end{aligned}$$

where Σ is the diagonal matrix of standard deviations, $\mathbf{C}$ is the correlation matrix with Λ a diagonal normalisation matrix and $\mathbf{L}_{v,h}$ the vertical and horizontal diffusion operator. $\mathbf{W}$ is a diagonal matrix accounting for the variable extension of each vertical layer $k = 1, \ldots, k_{\max} = 23$. To accomplish the factorisation of the correlation matrix (@), it is crucial that the Laplacian, which describes the diffusion operator, is self-adjoint.
Due to the incremental formulation of the cost function, solely a multiplication of $\mathbf{B}^{1/2}$, $\mathbf{K}^{1/2}$ and $\mathbf{B}^{\mathrm{T}/2}$, $\mathbf{K}^{\mathrm{T}/2}$ is needed. The diffusion approach provides the background error covariances as operators. One crucial advantage of this approach consists in obtaining the square root and its transposed of $\mathbf{B}$ by adjustment of two defining parameters for the diffusion operator, the diffusion coefficient κ and the integration time T. Using half of the integration time yields the square root of $\mathbf{B}$.
The horizontal influence radius L is approximated in *Weaver and Courtier*

[2001] as

$$L^2 \approx 2\kappa T, \tag{3.24}$$

stating for example that two model grid points with distance L have a mutual correlation of $\exp(-1/2)$. Various influence radii for initial values, anthropogenic emissions, and biogenic fluxes have been implemented. The horizontal influence radii for initial values and emissions increase linearly with model height with a kink at the PBL [*Elbern et al.*, 2007]. The vertical correlation length is estimated by Eq. (3.24) with the vertical diffusion coefficient κ^v which is calculated from WRF. This ensures that a well mixed boundary layer is reflected by the vertical correlation length without additional computational effort [*Elbern et al.*, 2007].

3.4.4.2 Flux factor background error standard deviation

In the following the variance modelling for the flux factor BECM $\mathbf{K}$, developed in this work, is presented. Joint optimisation of initial values and flux factors requires a distinction between the adjoint models with respect to these two optimisation parameters. Although the flux factor $\mathbf{f} = \exp(\mathbf{g})$ is constant during one iteration, the adjoint flux factor $\mathbf{f}^*$ changes with $(\mathbf{M})^{\mathrm{T}}$ depending on the background flux strength $\mathbf{U}^{\mathrm{b}}$. The influence of $\mathbf{U}^{\mathrm{b}}$ on $\nabla\mathcal{J}$ is investigated hereinafter. Using Eq. (3.14) iteratively and the transport operator $\mathbf{T}$ (Eq. (3.15)), we derive

$$(\mathbf{M}_{0,i})^{\mathrm{T}} \begin{pmatrix} \mathbf{x}_i^* \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} (\mathbf{T}_{0,i})^{\mathrm{T}} \mathbf{x}_i^* \\ \sum_{j=1}^{i} \triangle t \, \mathbf{U}^{\mathrm{b}}_{j-1/2} \, (\mathbf{T}_{j-1/2,i})^{\mathrm{T}} \mathbf{x}_i^* \end{pmatrix}, \tag{3.25}$$

with an equidistant time interval $[t_0, \ldots, t_N]$: $\triangle t := t_{i+1} - t_i$ for $i = 0, \ldots, N-1$. See Appendix A.1 for a detailed derivation. Thus, the adjoint model with respect to the flux factor is proportional to $\mathbf{U}^{\mathrm{b}} \, (\mathbf{T})^{\mathrm{T}}$.
With regard to Eq. (3.18) the quantity $\mathbf{K}^{\mathrm{T}/2}\mathbf{U}^{\mathrm{b}}(\mathbf{T})^{\mathrm{T}}$ is proportional to the observational part of the gradient of the preconditioned cost function. We want to construct the diagonal of $\mathbf{K}$ such that we achieve the following properties of the 4D-Var system:

1. Priority of optimisation of fluxes is ordered according to their strength.
2. Satisfactory sensitivity of the optimisation system to fluxes with small absolute amount.
3. Robustness of the system.

A compromise between counteracting properties 2 and 3 has to be accomplished. Identical twin experiments (see Sect. 5.1.1.1) demonstrate that with a constant diagonal of $\mathbf{K}$, the 4D-Var system EURAD-IM shows a tendency to optimise only the largest fluxes, but is too conservative for smaller fluxes and property 2 is not satisfied. This especially holds true for anthropogenic emissions, dominated by few very large point sources.
Using a specific construction for the standard deviation of $\mathbf{K}$ in dependence of $\widehat{\mathbf{U}^{\mathrm{b}}} := \sum_{i=1}^{N} \mathbf{U}^{\mathrm{b}}_{i-1/2}$

$$\sqrt{\mathbf{K}_{r,r}} = \mathrm{c}\,\frac{N}{\triangle t}\left[\max_s \left|\widehat{\mathbf{U}^{\mathrm{b}}}(s)\right|\right]^{-\frac{1}{l}} \left[\left|\widehat{\mathbf{U}^{\mathrm{b}}}(r)\right|\right]^{-\frac{l-1}{l}}, \quad l \geq 1, \tag{3.26}$$

enables an approximation of $\nabla\mathcal{J}$ which fulfils the three properties above:

$$\nabla\mathcal{J}(\mathbf{v},\mathbf{w}) \approx \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathrm{c}\left(\frac{\widehat{\mathbf{U}^{\mathrm{b}}}}{\max|\widehat{\mathbf{U}^{\mathrm{b}}}|}\right)^{\frac{1}{l}}\end{pmatrix} \sum_{i=0}^{N} \begin{bmatrix}(\mathbf{T}_{0,i})^{\mathrm{T}}\\ \sum_{j=1}^{i}(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\end{bmatrix} \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right], \tag{3.27}$$

see Appendix A.2 for a derivation of Eqs. (3.26) and (3.27). Equation (3.26) is implemented for each of the three fluxes considered, thus the constant c and $\max_s \left|\widehat{\mathbf{U}^{\mathrm{b}}}(s)\right|$ are with respect to anthropogenic emissions ($\mathrm{c}^{\mathrm{ant}}$), photosynthesis ($\mathrm{c}^{\mathrm{phot}}$), and respiration ($\mathrm{c}^{\mathrm{resp}}$) respectively.
Thus, for $l = 1$ (i.e. the diagonal of $\mathbf{K}$ is constant) the right hand side of Eq. (3.27) depends linearly on $\mathbf{U}^{\mathrm{b}}$. With increasing l the sensitivity of the system to smaller fluxes increases, while the robustness decreases.
After several tests $l = 4$ has been chosen for anthropogenic emissions and $l = 2$ for photosynthesis and respiration, which is caused by different spatial heterogeneity of the three fluxes. The off-diagonal elements of $\mathbf{K}$ are calculated with the diffusion approach analogue to the technique outlined for matrix $\mathbf{B}$. A correlation of 0.4 between photosynthesis and respiration has been chosen empirically after several identical twin experiments with the EURAD-IM (Sect. 5.1). No correlation is assumed between anthropogenic emissions and the two biogenic fluxes.

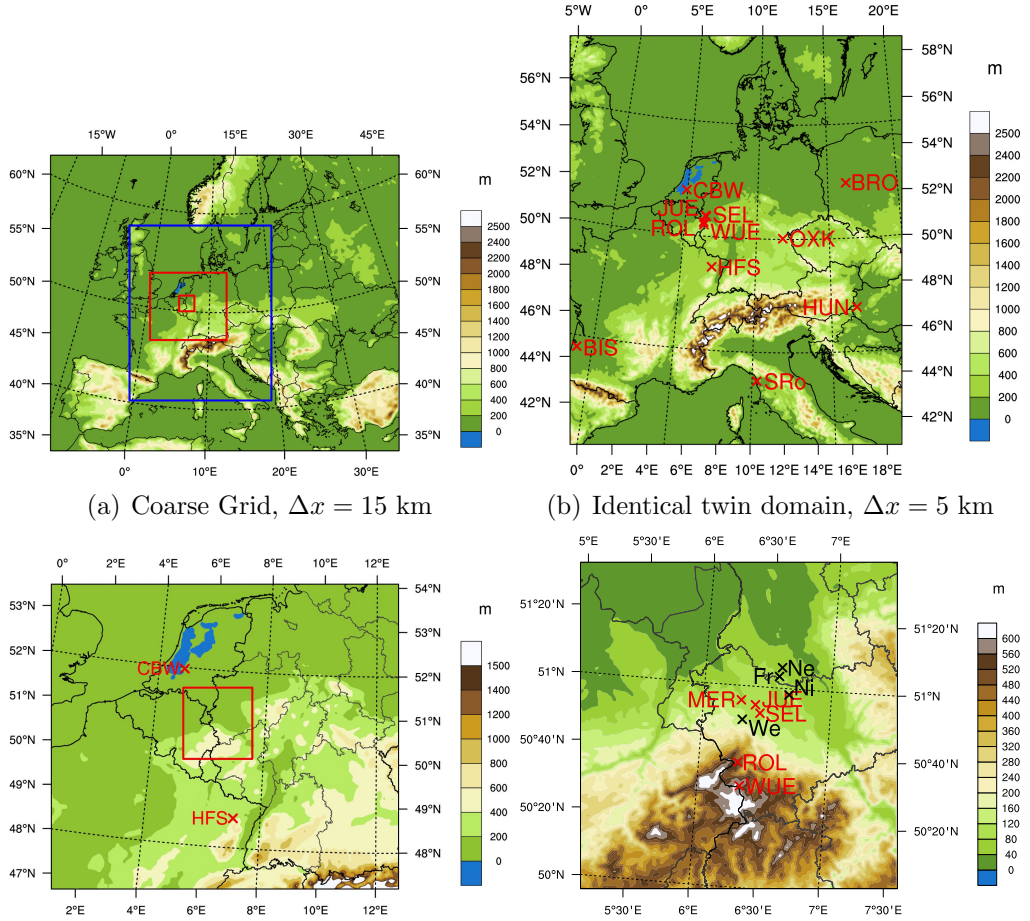
Chapter 4

Model set-up and observations

4.1 Model set-up

4.1.1 Model domain and nesting configuration

In this study EURAD-IM and WRF model use a Lambert conformal projection, where the centre of the coarse domain is located at 51° N latitude, 12.5° E longitude. The initial background concentration values of CO_2 for all domains are assumed to be constant 386 ppmV.
Two nesting configurations have been chosen in this study: For real-data cases a horizontal resolution sequence of 15×15 km^2, 5×5 km^2 and 1×1 km^2 is used. The mother domain (Fig. 4.1(a)) covers almost whole Europe. The sub-model domains are depicted in Fig. 4.1(c) (5 km) and 4.1(d) (1 km). For identical twin experiments the same mother domain as for the real-data case is used and only one sub-domain with 5 km horizontal resolution (Fig. 4.1(b)).
Boundary values for the mother domain are assumed to be constant for identical twin experiments (see Sect. 5.1). For real case studies (see Sect. 5.2) global atmospheric CO_2 values from "MACC-III greenhouse gases inversions" (*http://apps.ecmwf.int/datasets/data/macc-ghg-inversions/?version=v10*) is used to interpolate boundary values for the mother domain. This global data has 3.75° longitude and 2.0° latitude horizontal resolution, the vertical grid includes 22 layers up to the upper boundary of the EURAD-IM at 100 hPa, and the temporal resolution is three hours. Boundary values from each parent domain are written out every hour and are interpolated temporally for the next smaller daughter domain.



(a) Coarse Grid, $\Delta x = 15$ km

(b) Identical twin domain, $\Delta x = 5$ km

(c) Real case domain, $\Delta x = 5$ km

(d) Real case domain, $\Delta x = 1$ km

Figure 4.1: *Mother domain (a) with the daughter domain for identical twin experiments (blue rectangle shown in (b)) and the real case study (red rectangle shown in (c) and (d)). Red crosses indicate the location of the measurement stations according to Table 4.1. Black crosses show the location of the four biggest power plants Weisweiler, Niederaußem, Frimmersdorf, and Neurath, see Table 5.5.*

4.1.2 Different CO_2 tracers in EURAD-IM

To allow a detailed assessment of the model, six different tagged tracers of CO_2 are simulated. The first is the total CO_2 amount. The second is the atmospheric background concentration, influenced only from the global boundary values from MACC-III, but not from anthropogenic or biogenic fluxes of the simulated domains in this work.

Four additional CO_2 tracers are simulated driven by the surface-atmosphere fluxes to aid budget analysis: one comprises the CO_2 concentration removed by photosynthesis, and the latter three comprise the concentration introduced by anthropogenic emissions, leaf respiration, and soil respiration. These four tracers are initialised with 0 ppmV for all domains and the boundary values of the mother domain are constantly set to 0 ppmV as well. Figure 4.2 shows a detailed time series of observations at station Rollesbroich and the six tracers of the background run of the 5 km domain for the real case on 6 and 7 June 2012. The four tracers, which represent

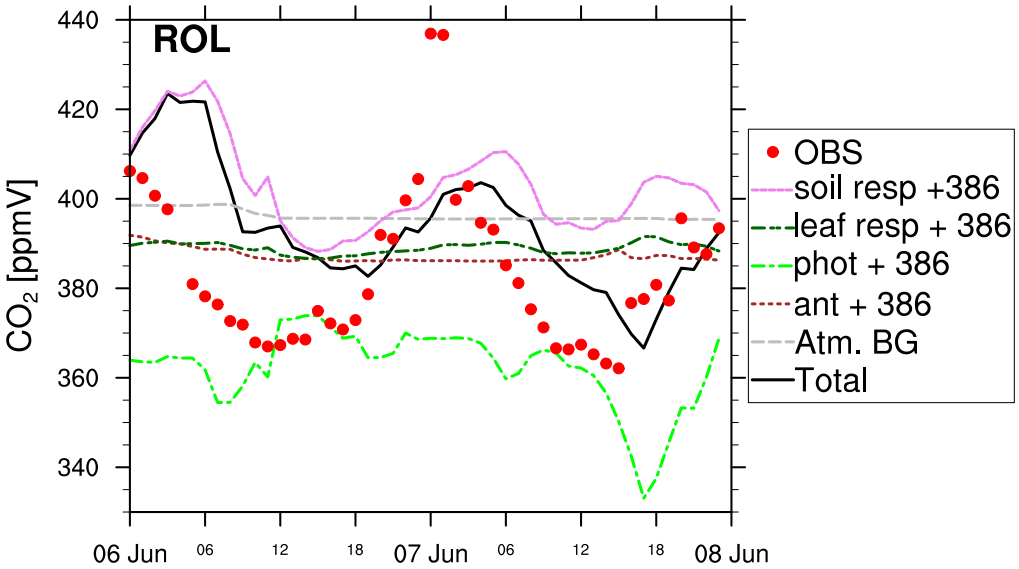


Figure 4.2: *Time series of six tagged CO_2 tracers in EURAD-IM and the observations at the station of Rollesbroich from 6 to 7 June 2012 of the real case domain with 5* km *horizontal resolution.*

the surface-atmosphere fluxes, are added to the initial CO_2 concentration (386 ppmV) to enable a comparison of these tracers with the total atmospheric CO_2 amount. Soil respiration and photosynthesis introduce large variations. Leaf respiration and anthropogenic emissions are much smaller with less amplitudes in this case.

4.2 Observations

4.2.1 Available measurements

This work uses in situ atmospheric concentrations measurements of seven ground stations and **n**et **e**cosystem **e**xchange (NEE) flux measurements of five stations for a real case study (Sect. 5.2). Concentration observations are compared with the total atmospheric CO_2 concentration, NEE with the sum of soil and leaf respiration subtracted by the photosynthesis. An overview of the used stations for real and synthetic experiments is given in Table 4.1.
The data of the stations Merzenhausen (MER) [*Graf et al.*, 2013], Rollesbroich (ROL) [*Post et al.*, 2015], and Selhausen (SEL) [*Langensiepen et al.*, 2012; *Mauder et al.*, 2013] is collected by the SFB-TR32 "Pattern in Soil-Vegetation-Atmosphere-Systems: Monitoring, Modelling, and Data Assimilation". Within the TERENO project (*http://teodoor.icg.kfa-juelich.de/overview-de*) the data at Wuestebach (WUE) site is collected (provided by Patrizia Ney, FZ Jülich, IBG-3). These stations have turbulence sensors to measure the CO_2 flux with the eddy covariance method, by measuring concentration values at a frequency of 20 Hz. Flux and concentration measurements are averaged over 30 minutes and are only used for evaluation or assimilation if the data has the highest quality flag as recommended (personal communication by M. Schmidt and A. Graf). Records of instruments at MER, ROL, SEL, and WUE are not primarily taken for their absolute accuracy. Instead, the instruments are very precise with regard to the high frequent change of CO_2 concentration, which is more relevant for optimisation of fluxes in this work. To ensure the high precision, calibration of the instruments deployed at those stations is performed with the same reference gas (personal communication by A. Graf).
The concentration measurements of the stations MER, ROL, SEL, and WUE are given in $\frac{\text{mmol}}{\text{m}^3}$ and are converted to ppmV with the observed temperature and pressure and the ideal gas equation:

$$\chi_{CO_2} = n_{CO_2}/n_{\text{Air}} = \frac{n_{CO_2}}{V} \cdot R_{\text{gas}} \cdot \frac{T}{P} \tag{4.1}$$
$$\left[\frac{10^{-3}\,\text{mol}}{\text{m}^3} \cdot \frac{\text{kg} \cdot \text{m}^2}{\text{s}^2 \cdot \text{K} \cdot \text{mol}} \cdot \frac{\text{K} \cdot \text{m} \cdot \text{s}^2}{10^2\,\text{kg}} = \frac{1}{10^5}\right].$$

Here χ_{CO_2} is the volume mixing ratio, n_{CO_2} the amount of CO_2 in mmol, V the volume in m^3, R_{gas} the universal gas constant in $\frac{\text{kg}\cdot\text{m}^2}{\text{s}^2\cdot\text{K}\cdot\text{mol}}$, T the temperature in K, and P the pressure in hPa $= \frac{10^2\,\text{kg}}{\text{K}\cdot\text{m}\cdot\text{s}^2}$.

Table 4.1: *Concentration measurement stations used in this study for real case study (upper part) and availability of flux measurements. Additional stations for identical twin experiments (lower part) are from FLUXNET (http://fluxnet.ornl.gov/). Only Merzenhausen station (indexed with *) is omitted for identical twin experiments, since it is very close to Selhausen. (ENF = Evergreen Needle leaf Forests, DBF = Deciduous Broad leaf Forests, W. Wheat = Winter Wheat).*

Exp.	Code	Lon, °	Lat, °	Elev., m a.s.l.	Measurement Height, m a.g.l.	Name	Veg.	NEE meas. available
stations for real case	CBW	4.93	51.97	5	20, 60, 120, 200	Cabauw, Ned	Grass	×
	HFS	7.06	48.67	304	22	Hesse F. S., Fra	DBF	✓
	JUE	6.41	50.91	101	100	Juelich, Deu	-	×
	MER*	6.29	50.93	93	3	Merzenhausen	W. Wheat	✓
	ROL	6.30	50.62	486	3	Rollesbroich, Deu	Grass	✓
	SEL	6.45	50.87	109	3	Selhausen, Deu	Potatoes	✓
	WUE	6.33	50.50	578	38	Wuestebach, Deu	ENF	✓
stations used additional for ident. twin exp.	BIS	-1.23	52.43	120	42	Biscarosse, Fra	ENF	×
	BRO	16.29	52.43	83	5	Brody, Pol	Crop	×
	JUE	6.41	50.91	101	20	Juelich, Deu	-	×
	HUN	16.65	46.96	234	10, 48, 82, 117	Hegyhatsal, Hun	Crop	×
	OXK	11.81	50.03	1020	23, 90, 163	Ochsenkopf, Deu	ENF	×
	SRo	10.29	43.73	6	23.5	San Rossore2, Ita	ENF	×

Measurements of the Jülich tower (JUE), operated by the FZ Jülich GmbH, (provided by Marc von Hobe, FZ Jülich, IBG-3) are used only in 100 m height, as confidence in measurements of lower altitudes is not high enough during June 2012 (personal communication by M. von Hobe). The concentration measurements are averaged 10 minute values and are further averaged to 30 minutes for assimilation. Concentration measurements from Cabauw (CBW) [*Vermeulen et al.*, 2011] are taken at 4 different heights in 20, 60, 120, and 200 m above ground layer (a.g.l.). Hesse Forest Sarrebourg (HFS) [*Longdoz et al.*, 2007] and CBW measurements are also averaged to 30 minutes. Both stations use an infrared gas analyser for CO_2 concentration measurements and additionally an ultrasonic anemometer to infer fluxes.

4.2.2 Observation error covariance matrices

The method to estimate $\mathbf{R}_i \in \mathbb{R}^{p_i \times p_i}$ ($i = 0, \ldots, N$, p_i= number of observations at t_i) presented here is similar to the method in *Elbern et al.* [2007]. The covariance matrices $\mathbf{R}_i$ are assumed to be diagonal, such that observation errors are not correlated. The total error is given as the sum of a measurement error and a representativeness error

$$\mathbf{R}_i(j,j) = r^{\text{meas}} + r_j^{\text{repr}}, \quad j = 1, \ldots, p_i. \tag{4.2}$$

The measurement error r^{meas} is assumed to have a constant value of 0.25 ppmV for all stations. Except from station HFS all measurements are provided with information about the standard deviation of averaged CO_2 concentration data r_j^{std}, $j = 1, \ldots, p_i$. The representativeness error is defined in dependence on the grid resolution Δx [km] and this standard deviation

$$r_j^{\text{repr}} := \sqrt{\Delta x}\, r_j^{\text{std}}, \quad j = 1, \ldots, p_i. \tag{4.3}$$

The standard deviation of the measurements at HFS are estimated with $\mathbf{y}_{i,j}/1000.0$, being on average close to the standard deviations of the other measurements.

Chapter 5

Simulation results

In this chapter ,the main results of the executed data assimilation model runs are presented. In the first part, results from identical twin experiments are shown. In the second part, a real case study spanning one month is presented.

5.1 Identical twin experiments

Identical twin experiments are used to evaluate the performance of the data assimilation system under simplifying assumptions. Contrary to real case experiments, they use synthetic observations which are generated by a forward model run. This run is commonly called nature run. The first guess (or background) run relies on disturbed initial values and flux rates. During the assimilation, the 4D-Var cost function measures the distance between the first guess and the synthetic observations generated by the nature run. The identical twin analysis aims to assess how well the 4D-Var system is able to reconstruct the perturbed parameters. The model run with optimised parameters is called the analysis run. The forward model used in identical twin experiments is presumed to be perfect (with respect to the synthetic observations, compare Eq. 2.5), which is a severe simplification of reality. *Daley* [1991] concludes that identical twin experiments "err on the optimistic side". However, identical twin experiments deliver valuable information about the potential performance of the assimilation system and they are widely used in CO_2 inversions [*Lauvaux et al.*, 2008; *Carouge et al.*, 2010; *Broquet et al.*, 2011]. The selection of observations can be chosen with realistic space and time distribution. The use of real meteorological data shows the ability of the adjoint model to reconstruct dispersion of atmo-

spheric CO_2 backwards in time (see Sect. 5.1.2). The effect of observation errors can be simulated by adding noise to the artificial observations.

5.1.1 Experiments with one-dimensional version of EURAD-IM

A spatially one-dimensional version of the EURAD-IM with 100 grid cells is used to address two aspects of the optimisation approach:

(i) influence of the variance of $\mathbf{K}$ to the analysis of the flux factors,
(ii) ill-posedness of the joint optimisation of initial values and flux factors.

Although a one-dimensional model is far from reality, both aspects can be illustrated in this framework to investigate limits of the optimisation approach. The ill-posedness is indeed the major difficulty for real case assimilation in this work. A one-dimensional model implies the usage of constant winds (here always from left to right), since the wind field has to be divergence free.

5.1.1.1 Influence of modelling the variance of K to the flux factor optimisation

To quantify the influence of the variance of $\mathbf{K}$ on the flux factor analysis, the following scenario is investigated: Six emitting grid cells are defined, located at position 5, 20, 35 50, 65 and 80, each using the same diurnal cycle of industrial combustion split into 96 time steps. The total amount of emissions of each emitting grid cell is always the half of the left neighboured emitting grid cell (see Table 5.1). Synthetic observations are taken at every time step at six stations, such that each measurement station is influenced only by one emitting grid cell. The experiment setup is visualised in Fig. 5.1. Observation errors are 1/1000 of the measured concentration ($\sim$ 0.4 ppmV) and the observation error covariance matrices $\mathbf{R}_i$ are diagonal. The background run starts with an emission factor of 0.5 at each grid cell. $\sqrt{(\mathbf{K}_{r,r})}$ is modelled according to Eq. (3.26), once with $l = 1$ leading to a constant diagonal of $\mathbf{K}$ and once with $l = 4$ giving smaller emissions a smaller error compared to $l = 1$. The two cases will be called $\mathbf{K}^{l4}$ and $\mathbf{K}^{l1}$ in the following.
The analysed flux factor for the largest emitting grid cell is almost identical for $\mathbf{K}^{l1}$ and $\mathbf{K}^{l4}$ (Fig. 5.2(a)). For grid cells with smaller emissions, the analysed flux factor gets also smaller. The decrease of the analysed flux factor is much faster for $\mathbf{K}^{l1}$ compared to $\mathbf{K}^{l4}$, see Table 5.1 and Fig. 5.2(b). For

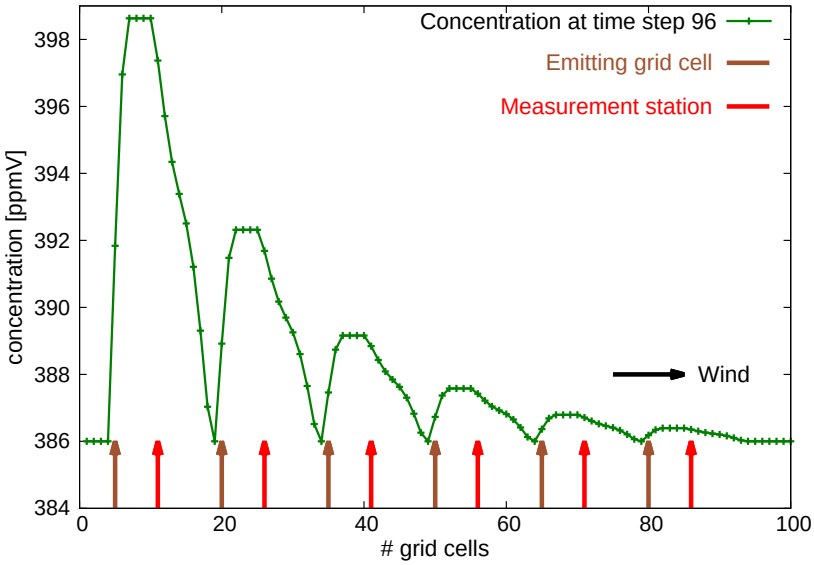
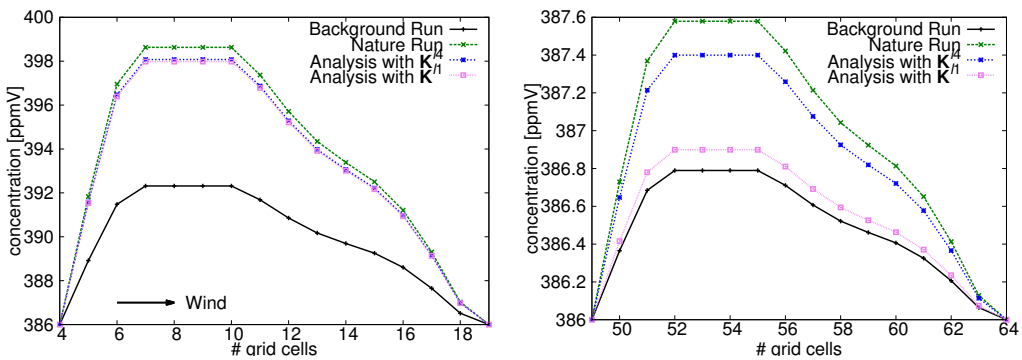


Figure 5.1: *Overview of the experiment setup of Sect. 5.1.1.1. The green line represents the CO_2 concentration after 96 time steps at the end of the assimilation window. Brown and red arrows indicate the location of the emissions and the measurement stations respectively. Each measurement station is influenced only by one emitting grid cell.*



(a) zoom for emissions introduced at grid cell 5

(b) zoom for emissions introduced at grid cell 50

Figure 5.2: *Identical twin experiment, showing the background run (black), nature run (green) and the analysis differing in $\sqrt{\boldsymbol{K}_{r,r}}$ modelled according to Eq.* (3.26) *once with $l = 4$ (blue) and once with $l = 1$ (pink). On the left panel a zoom of the emissions introduced by the largest source is shown, the emission strength in the right panel is 8 times smaller.*

both cases the properties 1.-3. of Section 3.4.4.2 are fulfilled due to the construction of the diagonal of $\mathbf{K}$, with $\mathbf{K}^{l4}$ being more sensitive for smaller

Table 5.1: *Analysed flux factors (ff) and variance of* $\boldsymbol{K}$ *(both without units) for* $\boldsymbol{K}^{l4}$ *and* $\boldsymbol{K}^{l1}$ *corresponding to Fig. 5.2.*

position of emitting grid cell	5	20	35	50	65	80
avg. emitted amount [μg/s]	1.0	1/2	1/4	1/8	1/16	1/32
background ff	0.5	0.5	0.5	0.5	0.5	0.5
analysed ff for $\mathbf{K}^{l1}$	**0.95**	**0.88**	**0.74**	**0.54**	**0.52**	**0.51**
analysed ff for $\mathbf{K}^{l4}$	**0.95**	**0.95**	**0.94**	**0.87**	**0.74**	**0.70**
$\sqrt{\mathbf{K}_{r,r}}$ for $\mathbf{K}^{l1}$	0.99	0.99	0.99	0.99	0.99	0.99
$\sqrt{\mathbf{K}_{r,r}}$ for $\mathbf{K}^{l4}$	0.99	1.67	2.80	4.72	7.93	13.3

emissions than $\mathbf{K}^{l1}$.

5.1.1.2 Ill-posedness of the problem of joint optimisation

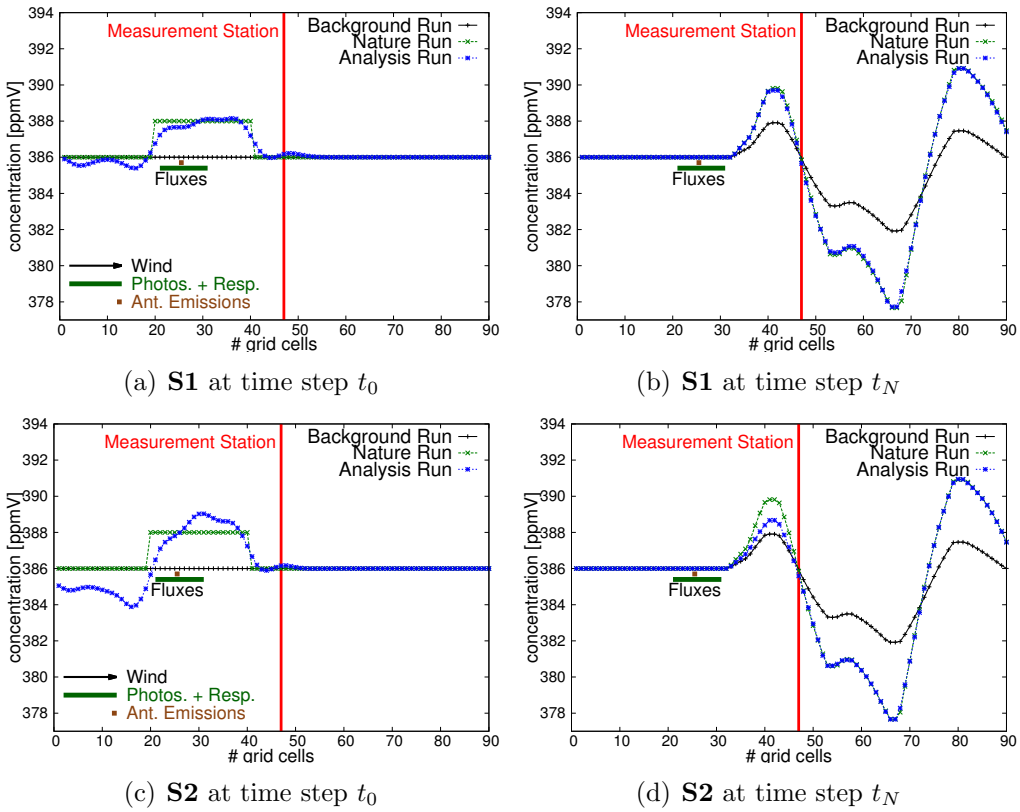
The following experiment assesses further limits of the assimilation system due to the ill-posedness of the problem. It is a step towards a more realistic set-up as it uses ten neighboured grid cells with biogenic fluxes and one with anthropogenic emissions, each with disturbed flux factor of 1/2. Additionally, initial values are disturbed by 2 ppmV at 20 grid cells. Measurements are taken only at one station, located downwind of the perturbed parameters, see Fig. 5.3. The analysis is performed twice (called **S1** and **S2** hereinafter), differing only in the BECM's for initial values $\mathbf{B}$ and flux factors $\mathbf{K}$:

$$\mathbf{B}^{\mathbf{S1}} = \frac{1}{2}\,\mathbf{B}^{\mathbf{S2}}, \quad \mathbf{K}^{\mathbf{S1}} = 2\,\mathbf{K}^{\mathbf{S2}}. \tag{5.1}$$

Table 5.2: *RMSE and bias of the flux factors (without units) of the analysis runs* **S1** *and* **S2** *and the background run with respect to the nature run.*

	(**S1**, Nature)	(**S2**, Nature)	(Background,Nature)
RMSE flux factors	0.13	0.38	0.5
bias flux factors	1.21	6.3	10.5

Contrary to the three dimensional case, the current set-up simplifies with respect to two aspects substantially: vertical diffusion is excluded and the complete time series of all flux cells can be observed, although there is only one measurement station. In combination with the knowledge of the location of the flux cells, an optimisation of flux factors only is not ill-posed



(a) **S1** at time step t_0

(b) **S1** at time step t_N

(c) **S2** at time step t_0

(d) **S2** at time step t_N

Figure 5.3: *Identical twin experiments* **S1** *(upper row) and* **S2** *(lower row) which differ by their background matrices according to Eq.* (5.1)*. Background run (black), nature run (green) and analysis run (blue) are shown at first time step (left column) and last time step (right column). At grid cells 21-30 photosynthesis and respiration (dark-green bulk) are introduced, anthropogenic emissions (brown small bulk) at grid cell 25. Initial values are disturbed for grid cells 20-39.*

in this set-up, if initial values are undisturbed. The good properties of the adjoint advection routine (see Section 3.4.2) enable a robust optimisation of the corresponding flux factors (not presented here).
The joint optimisation, however, allows several combinations of flux factors and initial values, which reproduce the synthetic observations. This can be seen by a comparison of Fig. 5.3(b) and 5.3(d). The initial values are able to adjust for perturbed flux factors in the case of **S2** as can be seen in Fig. 5.3(c). The cost function for both cases **S1** and **S2** decreases similarly. As a consequence, to evaluate the performance of the analysis of **S1** and **S2**, the RMSE and BIAS of the analysed flux factors are given in Table 5.2. Here, it can be seen that the analysis of the flux factors of **S1** is much better compared to **S2**. This illustrates that the inferred CO_2 surface-atmosphere

fluxes are highly sensitive with respect to **B** and **K**. In other words, the ratio of the BECM's is decisive for a successful analysis of CO_2 fluxes.

5.1.2 Experiments with EURAD-IM 5.8.1

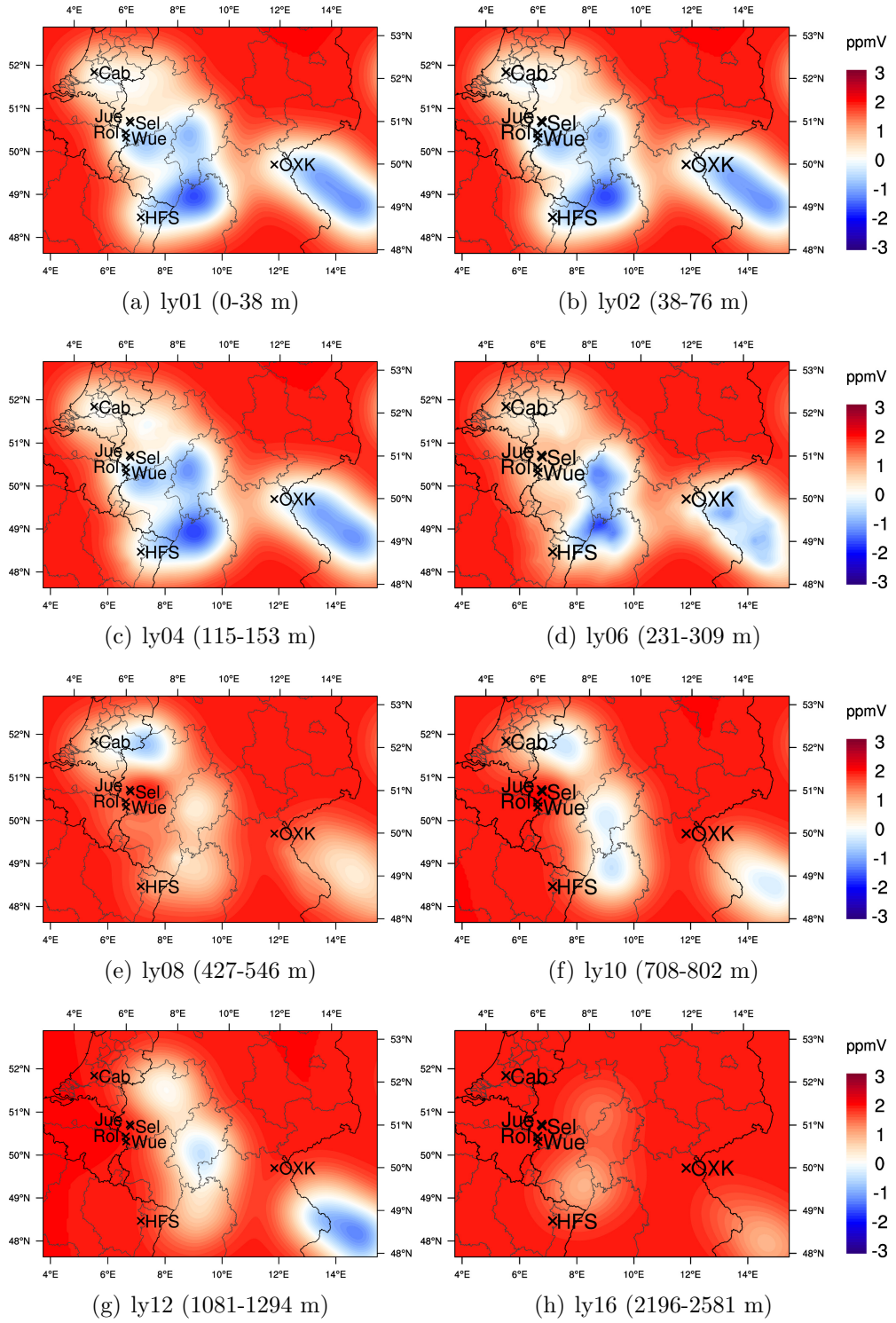
The following experiments are published in *Klimpt et al.* [2016] and use an input of real meteorology, anthropogenic emissions, photosynthesis, and respiration as described in Chapter 3. Numerical experiments are executed with four different background configurations, as listed in Table 5.3. For each configuration two experiments are executed. One optimising only flux factors for anthropogenic emissions, biogenic respiration, and photosynthesis and another optimising initial values and flux factors jointly for each grid cell. We call the first case hereinafter "only flux factor analysis" and the second case "joint analysis".

Table 5.3: *Configuration of background initial states and flux factors, applied for anthropogenic emissions, photosynthesis, and respiration.*

configuration	background initial state	background flux factor
1	nature run + 2 ppmV	0.8
2	nature run − 2 ppmV	0.8
3	nature run + 2 ppmV	1.2
4	nature run − 2 ppmV	1.2

The model domain encompasses central Europe with 5 km horizontal resolution (see Fig. 4.1(b)). After a model spinup time of 30 h spanning 23 July 00 UTC to 24 July 2012 06 UTC, a 12 h 4D-Var analysis run is initialised from 06-18 UTC. Synthetic measurements from the nature run are taken every 30 minutes (the model time step is 120 s) at eleven stations (see Fig. 4.1(b) and Table 4.1). The observation errors are constantly 2 ppmV and the error covariance matrices $\mathbf{R}_i$ are diagonal. In order to balance the analysis run with respect to background errors of initial values and flux factors, several values c^{iv} of Eq. (3.19) and c^{ant} (anthropogenic emissions), c^{phot} (photosynthesis), and c^{resp} (respiration) of Eq. (3.26) have been tested for all four configurations. Empirically $c^{\mathrm{iv}} = 0.0025$, $c^{\mathrm{ant}} = 1.4$, $c^{\mathrm{phot}} = 0.6$, and $c^{\mathrm{resp}} = 0.8$ have been chosen. The matrix **K** is modelled with $l = 4$ (see Sect. 3.4.4) for anthropogenic emissions and with $l = 2$ for biogenic fluxes. This is due to the fact, that the spatial distribution of flux variations is much smoother for biogenic fluxes compared to anthropogenic emissions. While the amount of biogenic fluxes at different grid cells has mostly the

same order of magnitude, the main contribution to anthropogenic emissions is caused by very few power plants, e.g. Weisweiler and Niederaußem in the Rur catchment area.
The off-diagonal entries of **B** and **K** are calculated using an influence radius of 30 km at the bottom layer, 45 km at the top of the modelled planetary boundary layer (PBL) and 60 km at the top of the model domain. The L-BFGS optimisation is carried out for at most 30 iterations, showing only minor changes of the cost function for the last iterations (Fig. 5.6).
Analysis configuration 4 is evaluated in the following. Figure 5.4 shows the difference of initial values for different vertical layers between the NmA (nature run minus joint analysis run). White areas indicate compliance of the analysis and nature run. The initial value correction is always towards the nature run, showing small overestimation easterly of Hesse Forest-Sarrebourg, the TR32 region, and Ochsenkopf at lower layers. Due to the set-up of the experiment with sparse observations, the meteorology is decisive for the distribution of corrections in the analysis run. Since westerly winds are predominant during the assimilation window, correction of the initial values occurs at all measurement stations and to the east of these. Correction of initial values can be seen up the top the vertical boundary layer at $\sim$ 2000 m (Fig. 5.4 (h)). Artificial overestimation can be seen in layer 12 (Fig. 5.4 (g)) south-east of station Ochsenkopf. Additional tests not presented here, show that a longer assimilation window increases this artificial deterioration of the analysis. Due to temporal high frequent measurements and the diffusive nature of atmospheric transport, the LBFGS amplifies small wiggles, primarily during later iterations, resulting in an artificial over- and underestimation.
A comparison of the assimilated flux factors for configuration 4, calculated once with joint analysis and once with only flux factor analysis, illustrates the benefits of the joint analysis. The results of the joint analysis in the left column of Fig. 5.5 are discussed first. Flux factors of the nature run are set to 1.
The flux factor analysis of anthropogenic emissions is dominated by a few but large point sources. In the Rur catchment area, due to westerly winds, the power plant Niederaußem is captured by the observation sites Jülich and Selhausen and is better reanalysed than Weisweiler, which is located a few kilometres upstream from both observation sites. The optimised flux factors of biogenic respiration (Fig. 5.5 (c)) and photosynthesis (Fig. 5.5 (e)) comply with the nature run very well. Although slight underestimation persists, the flux factors of the joint analysis are mainly between background (1.2) and nature run (1.0).
The analysis of the only flux factor optimisation has a stronger overestim-



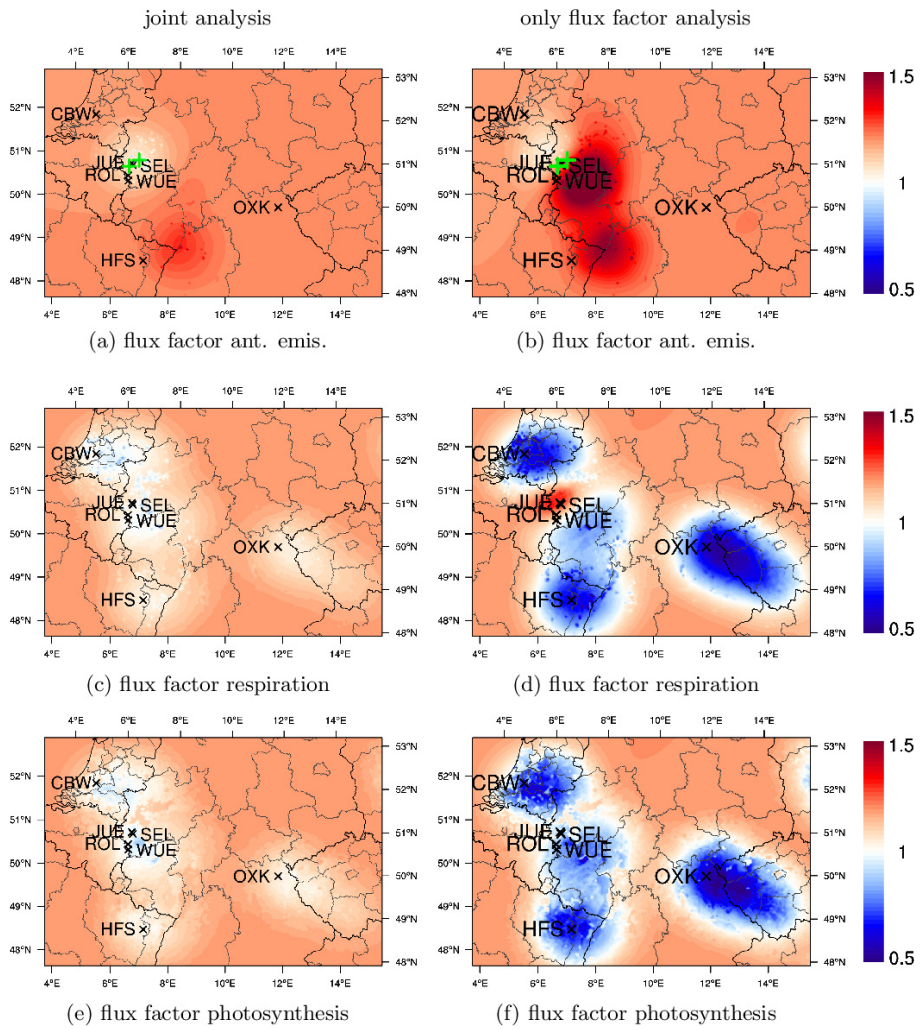
(a) ly01 (0-38 m) (b) ly02 (38-76 m)

(c) ly04 (115-153 m) (d) ly06 (231-309 m)

(e) ly08 (427-546 m) (f) ly10 (708-802 m)

(g) ly12 (1081-1294 m) (h) ly16 (2196-2581 m)

Figure 5.4: *Zoom of the 5* km *domain for the initial values NmA (nature run minus joint analysis run) of configuration 4 for different layers on 24 July, 06 UTC. Black crosses show synthetic observation sites. The given heights are valid for the U.S. standard atmosphere.*



(a) flux factor ant. emis.
(b) flux factor ant. emis.
(c) flux factor respiration
(d) flux factor respiration
(e) flux factor photosynthesis
(f) flux factor photosynthesis

Figure 5.5: *Analysed flux factors of anthropogenic emissions (first row), biogenic respiration (second row), and photosynthesis (third row). The analysis at the left column shows results of the joint initial value and flux factor assimilation while the right column shows an analysis which optimised only flux factors. The two biogenic fluxes are shown at surface level, while anthropogenic emissions are given at* ≈ 270 m *height, which is the layer with the highest impact due to power plants. Black crosses show synthetic observation sites. Green plus signs in the first row indicate the location of the two biggest power plants Niederaußem and Weisweiler.*

ation of anthropogenic emissions than the joint analysis (Fig. 5.5 (a), (b)), due to 2 ppmV higher initial values of the background run. Both biogenic fluxes (Fig. 5.5 (d), (f)) are underestimated in large areas, as observations at the beginning of the assimilation window result in a too strong forcing

of photosynthesis. The underestimation of photosynthesis causes higher concentration values than observed during the late afternoon, which forces a decrease of the atmospheric source biogenic respiration. In the area of the observation stations Jülich and Selhausen, biogenic respiration is also overestimated (Fig. 5.5 (d)), a clear deterioration of analysis performance compared to the joint analysis.

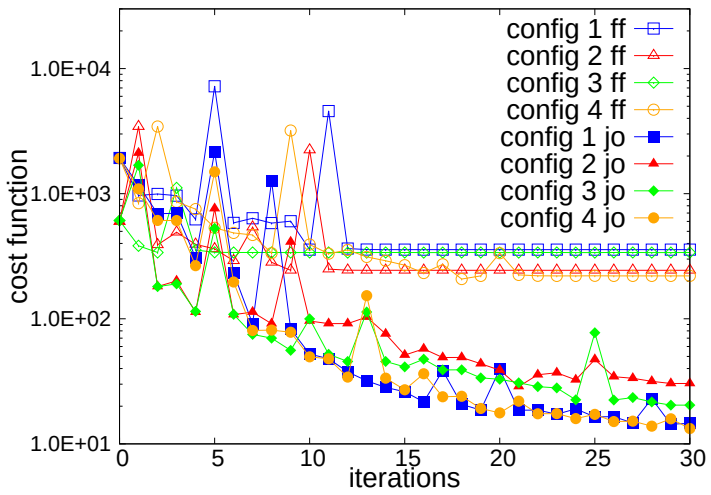
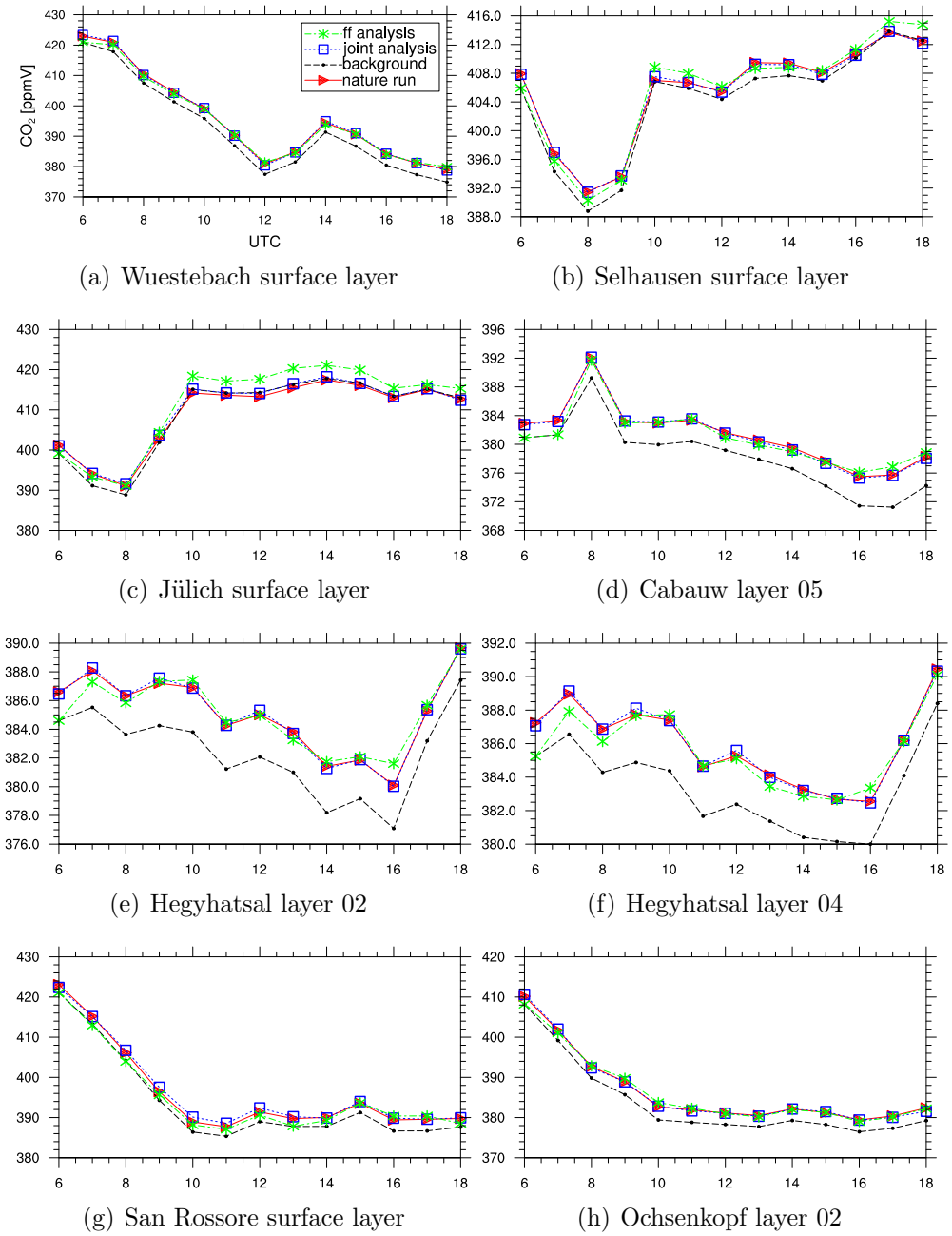


Figure 5.6: *Calculated cost function (without units) for joint (jo) and flux factor (ff) analysis of the four configurations listed in Table 5.3. Filled markers use joint optimisation. Only flux factor optimisation is indicated with empty markers, showing hardly any decrease of the cost function after 15 iterations.*

A comparison of the cost reduction for the four configurations for joint optimisation and only flux factor optimisation is shown in Fig. 5.6. For each configuration the cost function shows a stronger decrease (by a factor of $\approx$ 20) for the joint optimisation compared to the flux factor optimisation.
Time series of CO_2 concentrations for configuration 4 (Table 5.3) are depicted in Fig. 5.7. In general, a good compliance of the nature run and joint analysis at the observation sites is seen. The CO_2 concentration of the background run and the only flux factor analysis at the initial time is always 2 ppmV lower than the nature run. Surprisingly, also the only flux factor analysis is often in accordance with the nature run, except at Selhausen, Jülich, and Hegyhatsal. As can be seen from Fig. 5.5, the correction of the flux factors close to Wuestebach, Cabauw, and Ochsenkopf, by optimising only flux factors is poor. This shows clearly that solely an evaluation of the



(a) Wuestebach surface layer (b) Selhausen surface layer

(c) Jülich surface layer (d) Cabauw layer 05

(e) Hegyhatsal layer 02 (f) Hegyhatsal layer 04

(g) San Rossore surface layer (h) Ochsenkopf layer 02

Figure 5.7: *Time series of* CO_2 *concentration of configuration 4 at synthetic observation sites for nature run (red), background run (black), joint analysis (blue) and flux factor (ff) analysis (green).*

time series does not imply an adequate analysis.
The vertical concentration profiles at the last hour of the assimilation win-

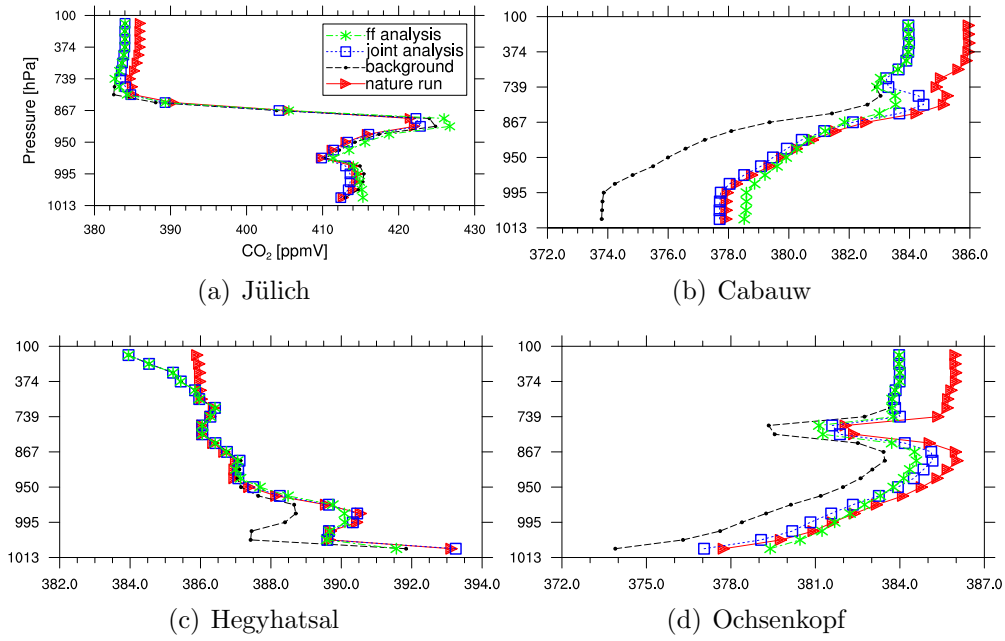


Figure 5.8: *Vertical profile of* CO_2 *concentration of configuration 4 at observation sites with measurements also in higher layers for the nature run (red), background run (black), joint analysis (blue) and flux factor (ff) analysis (green) at the end of the assimilation window on 24 July 2012 at 18 UTC.*

dow of the towers with measurements in higher layers are shown in Fig. 5.8. The vertical profiles of Cabauw, Ochsenkopf and Jülich display a slightly better concentration correction of the joint analysis compared to the flux factor analysis. Although the vertical profile of Hegyhatsal (Fig. 5.8(c)) is very similar for the joint and the flux factor analysis, the temporal evolution (Fig. 5.7(e) and 5.7(f)) towards the last hour of the assimilation interval is quite different, indicating again that the analysed flux factors are not sufficiently well estimated by the flux factor analysis.

The present study shows the beneficial effect of joint optimisation for the analysis of CO_2 fluxes. Accounting for errors from initial values has a smoothing effect on the inferred fluxes. The optimisation of only flux factors results in a strong forcing of fluxes due to observations especially at the beginning of the assimilation interval. Even a small deviation of background initial values of 2 ppmV has the potential to disturb the analysis of fluxes severely, if the initial value uncertainty is not taken into account.

5.2 Real case study for TR32 domain

The numerical experiments presented in Sect. 5.1.2 provide an upper boundary of the developed optimisation strategy. However, the experiments simplify substantially with respect to the forward model performance and important input parameters e.g. meteorology or the CO_2 fluxes themselves. To evaluate the performance of the presented 4D-Var system, a long lasting period is simulated to gather reliable statistical model data. A case study of 38 days lengths is presented, spanning the time from 24 May to 30 June 2012. EURAD-IM is applied with the nested model domain set-up, depicted in Figure 4.1(a), 4.1(c), and 4.1(d). The boundary values for the mother domain with 15 km horizontal resolution are taken from the global reanalysis MACC-III greenhouse gases inversions. The simulation of the mother domain was executed by a forward run of the EURAD model. Forward runs for the complete 38 days are also executed for the daughter nests with 5 km and 1 km horizontal resolution, using boundary values from the coarser domain. These runs are called hereinafter background runs.
Afterwards, 4D-Var assimilation runs are executed for the domains of 5 and 1 km horizontal resolution, which are called analysis runs. The assimilation window of each analysis run is 24 h from midnight to midnight. The analysis run of each day calculates first guess initial values and flux factors for the following day. The boundary values for the 5 km analysis run are the same as used by the forward run. The 1 km analysis run uses boundary values from the 5 km analysis run.
This study assimilates half hourly atmospheric CO_2 concentration measurements. The analysis results are validated with concentration measurements of CO_2, which were withheld from the assimilation procedure. Further CO_2 flux measurements of NEE, independent from the analysis, are used for validation. An overview of all observations used in this study is given in Table 4.1. To evaluate the analysis results only the last 30 days from 1 to 30 June are used, the first 8 days are required for the model spinup.
The meteorology for June 2012 is characterised by several low-pressure systems passing the model domain. The jet stream was often within the 1 km model domain, alternating frequently from north to south. Hence, the domain is situated on both sides of the jet stream characterised by cold and warm air, respectively. There is no long lasting dry phase in June 2012 for the model domain, such that plants do not suffer from water stress. The mean temperature of this month is slightly below long-time average for North Rhine-Westphalia, which implies good conditions for photosynthesis and respiration. Mowing of the meadow at ROL was done on 15 May, allow-

ing a recovery of the grassland for photosynthesis and respiration until the beginning of June, and on 5 July, such that no inconsistencies can be seen in the observed time series. At SEL sprouting of potatoes began around 15 May and a little earlier at MER (winter wheat). At neither station was harvested during June, such that the considered period represents a continuous growing season.

5.2.1 5 km domain

Three different setups are simulated and validated for the 5 km analysis run differing by the selection of measurement stations, either taken for assimilation or taken for evaluation. Additionally, the empirical parameters for the BECM's c^{iv}, c^{ant}, c^{phot}, and c^{resp} are modified slightly. The influence radii for the calculation of the non-diagonal entries for the initial values are 15 km at the surface layer, 25 km at the PBL, and 40 km at top of the model domain at 100 hPa. The influence radius for photosynthesis and respiration is 40 km at the surface layer in order to analyse many grid cells. For anthropogenic emissions, dominated by point sources of power plants, small influence radii of 5 km (surface layer), 10 km (PBL) and 20 km (at 100 hPa) are chosen. Table 5.4 gives an overview of the assimilated measurements of stations for the 5 km analysis run. The measurement data with more than 100 m a.g.l. (CBW and JUE) are assimilated to enlarge the area of optimised biogenic fluxes. The lower measurement data of CBW are for evaluation, since flux data at CBW was not available for this study. HFS was chosen for assimilation due to its location in the south of the model domain. Since SEL and MER are very close together, the almost complete time series data of MER is used for assimilation, while the measurements at SEL, having a gap between 07-16 June, are used for evaluation. ROL and WUE are also close to each other, but are located in different vegetation types (grassland and forest). Therefore the data of ROL and WUE are used for assimilation and evaluation in different configurations.

Table 5.4: *Configurations of stations whose data is used for assimilation (✓) or evaluation (×) for the 5* km *analysis runs.*

	CBW at height a.g.l.									
conf.	20	60	120	200	HFS	JUE	MER	ROL	SEL	WUE
1	×	×	✓	✓	✓	✓	✓	✓	×	×
2	×	×	✓	✓	✓	✓	✓	✓	×	✓
3	×	×	✓	✓	✓	✓	✓	×	×	✓

Concentration evaluation

To evaluate the analysis runs the mean differences (bias) and root mean square errors (RMSE) analysed concentrations are shown in Fig. 5.9. The

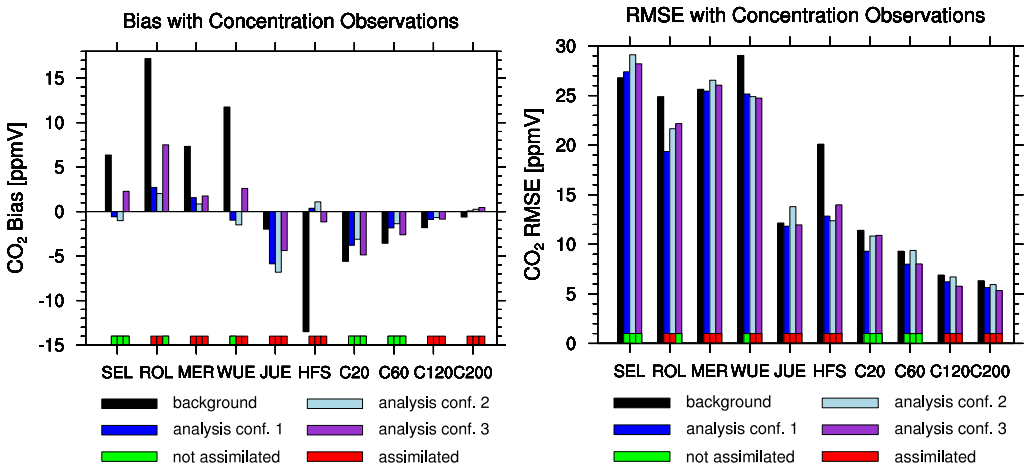
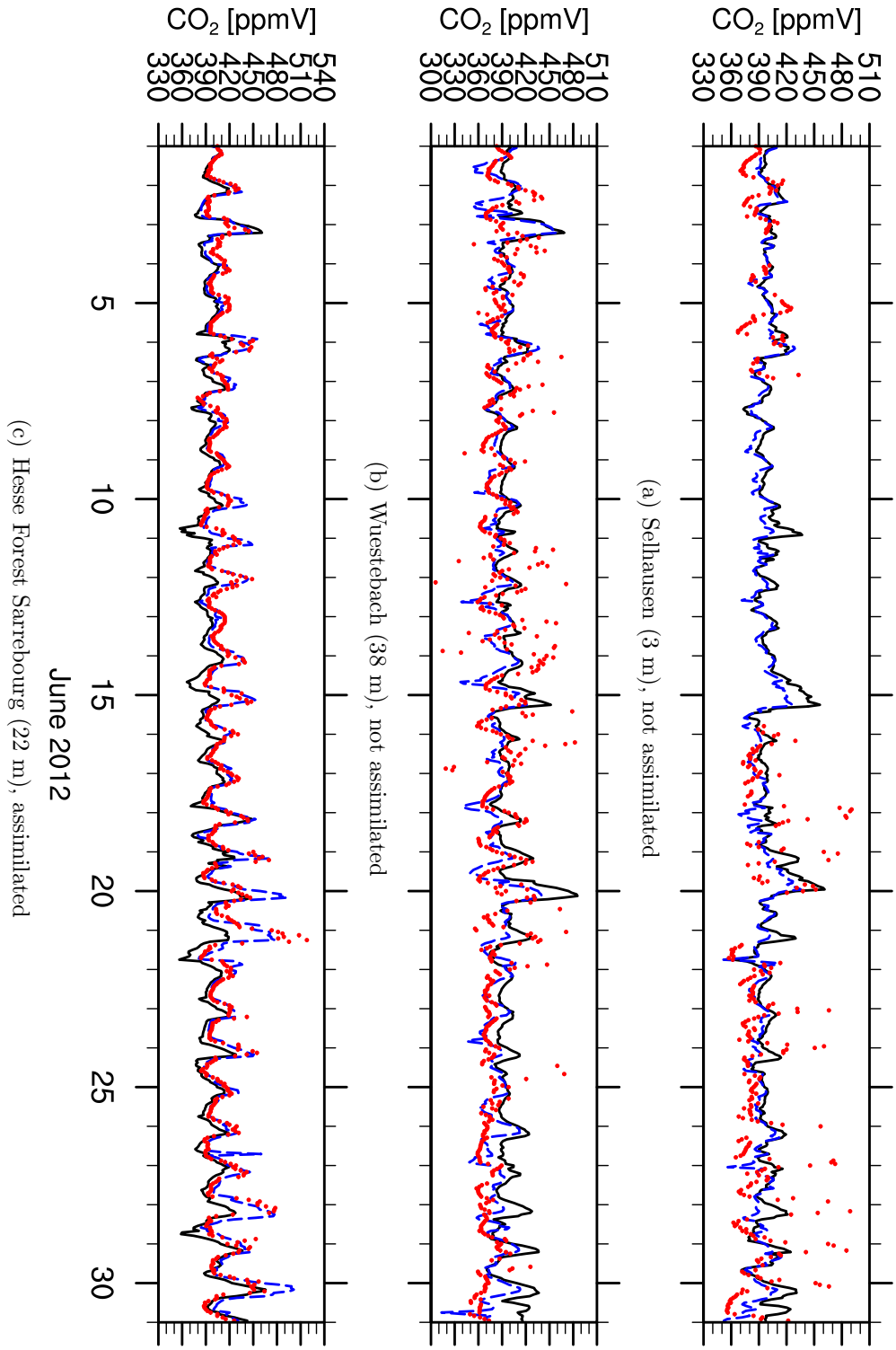


Figure 5.9: *Bias (left) and RMSE (right) between the concentration observations to the background run (black) and the analysis runs of configuration 1 (blue), 2 (light-blue), and 3 (violet) for June 2012. For each configuration the stations whose measurements are not assimilated are shown with a green bar, respectively with a red bar if the measurements are assimilated.*

absolute bias of the three analysis configurations reduces at all stations compared to the background bias, except at JUE. The RMSE shows a decrease for most of the stations of the analysis runs compared to the background run. For Selhausen, Merzenhausen, and Jülich the analysis RMSE increases for at least one configuration.

In the following configuration 1 will be evaluated in more detail. The time series of atmospheric concentration is presented in Fig. 5.10 at SEL and WUE which are not assimilated and at HFS and CBW which are not situated in the 1 km domain.

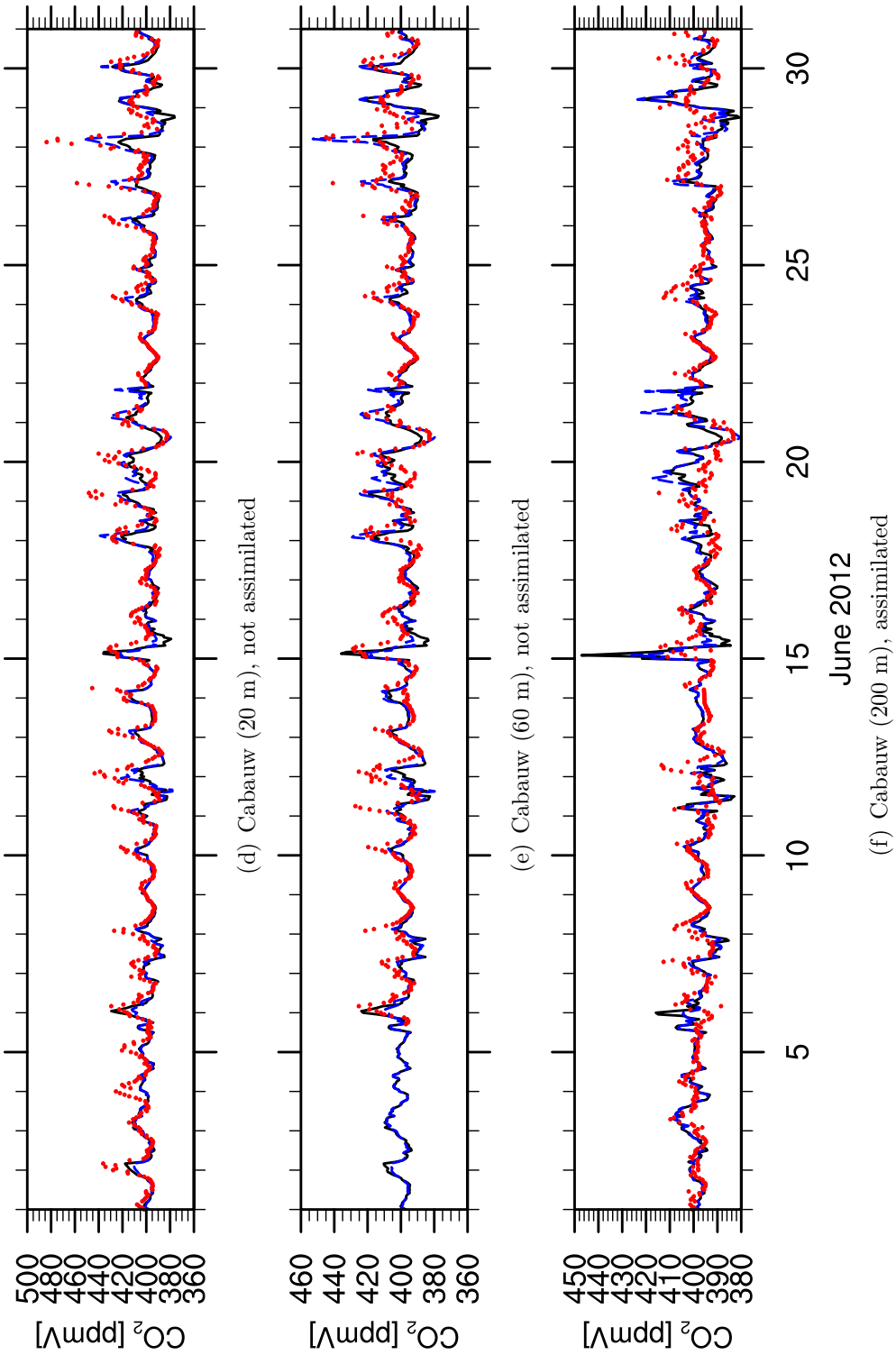
At SEL (Fig. 5.10(a)), during the nights of 18, 19, 23, 24, and 26-30 June 2012 high concentration peaks are observed which are not simulated by the background run, resulting in a model-observation mismatch up to 90 ppmV. These high concentration peaks from observations cannot be reproduced by the analysis run although similar observations of the nearby station MER (not plotted) are assimilated. One reason is that the surface layer of EURAD-IM ($\sim 0 - 38$ m vertical extension) is too coarse for the quite low measurement height of 3 m above ground level (a.g.l.) of SEL and MER during these nights with extremely low PBL height. Thus, the



(a) Selhausen (3 m), not assimilated

(b) Wuestebach (38 m), not assimilated

(c) Hesse Forest Sarrebourg (22 m), assimilated



(d) Cabauw (20 m), not assimilated

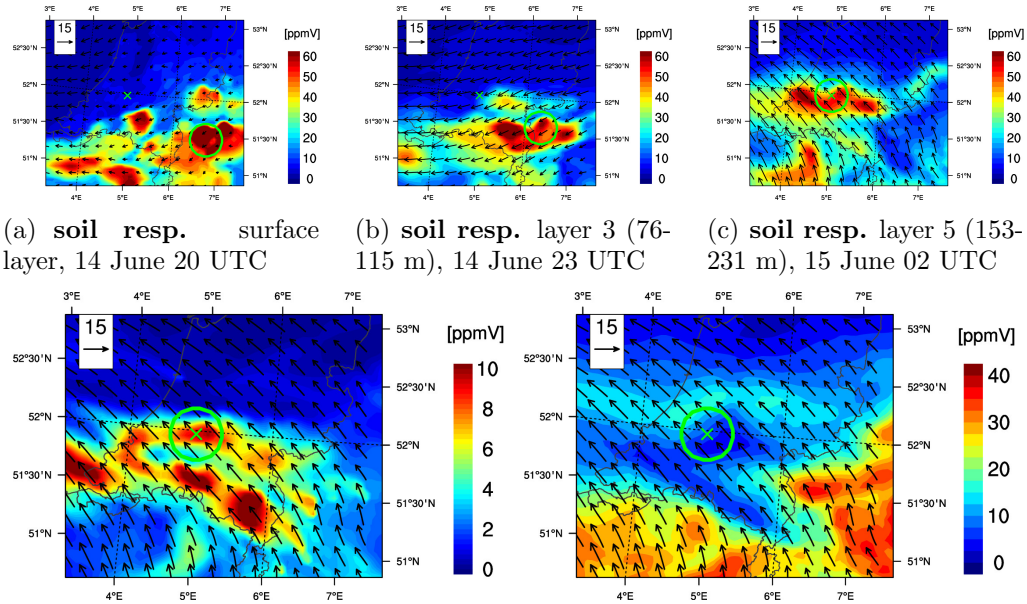
(e) Cabauw (60 m), not assimilated

(f) Cabauw (200 m), assimilated

Figure 5.10: *Concentration time series of observations (red dots), background (black solid line), and analysis run of configuration 1 (blue dashed line) for 5* km *domain for June 2012.*

measurements are not representative for the model during these nights.
In the case of WUE (Fig. 5.10(b)) we see in general a closer fit of the analysis run to the observations compared with the background run. The remarkable distribution of the background run between 19 and 21 June is also visible at ROL (not plotted). At ROL, there are also strikingly similar observations compared to WUE during this time period, explaining the good fit of the analysis run to the observations between 19 and 21 June 2012.
The best fit of the analysis run to the observation time series compared to the background run can be seen at HFS (Fig. 5.10(c)), which has the highest reduction of the RMSE (Fig. 5.9). The measurement height of 22 m a.g.l. is representative for the EURAD-IM surface layer also during nights with low PBL on 21 and 28 June 2012. The observed high concentration peaks, that appeared due to an accumulation of respired CO_2, are well reproduced by the analysis during these two nights.
Similarly, the observed high concentration peaks of CBW at 20 (Fig. 5.10(d)) and 60 m a.g.l. (Fig. 5.10(e)) on 27 and 28 June 2012 are well reproduced by the analysis, although these observations are not assimilated. The assimilated observations of CBW at 120 (not plotted) and 200 m a.g.l. (Fig. 5.10(f)) do not show such high concentration values during the nights of 27 and 28 June 2012, indicating a good performance of the analysis for this period due to independent observations.

A detailed evaluation of different CO_2 tracers at CBW at 200 m a.g.l (Fig. 5.10(f)) reveals a surprisingly good analysis on 15 June 2012. Hereby the background time series shows much higher concentration values than the observations. Figure 5.11 shows that the high background values (concentration peak on 15 July 02 UTC) are due to a combination of three factors: concentration values up to 60 ppmV originating from soil respiration, are emitted already at the previous day and are transported to CBW in model layer 5 (153-231 m) on 15 June 02 UTC (Fig. 5.11 (a)-(c)). At the same time, additional 10 ppmV from anthropogenic emissions are transported to CBW (Fig. 5.11(d)) and total atmospheric CO_2 is hardly reduced by photosynthesis (Fig. 5.11(e)), resulting in the background peak on 15 June at 02 UTC.
A comparison between the background and analysis of the tagged CO_2 tracers shows significant smaller analysed soil respiration (Fig. 5.12(b)) and slightly increased anthropogenic emissions (Fig. 5.12(a)). The analysis performance for this period is remarkably well due to two reasons. First, the fluxes which influence the CO_2 concentration at CBW in model layer 5 (153-231 m are emitted during the previous



(a) **soil resp.** surface layer, 14 June 20 UTC

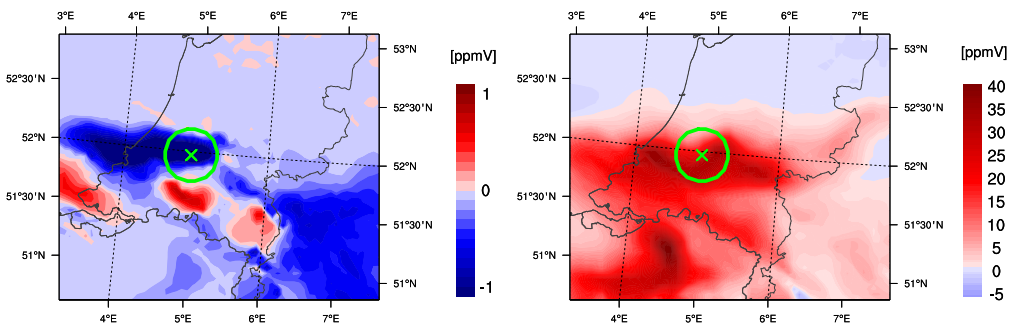
(b) **soil resp.** layer 3 (76-115 m), 14 June 23 UTC

(c) **soil resp.** layer 5 (153-231 m), 15 June 02 UTC

(d) **ant. emis.** layer 5 (153-231 m), 15 June 02 UTC

(e) **phot.** layer 5 (153-231 m), 15 June 02 UTC

Figure 5.11: *Zoom of the 5* km *domain of different tagged* CO_2 *tracers around the Cabauw tower, indicated by the green cross. The green circle shows the location of the concentration reaching Cabauw tower on 15 June, 02 UTC. Tagged* CO_2 *tracer for soil respiration (a-c) at three different times. Tagged* CO_2 *tracers for anthropogenic emissions (d) and photosynthesis (e) on 15 June 2012, 02 UTC.*



(a) **ant. emis.**

(b) **soil resp.**

Figure 5.12: *Background run minus analysis run (BmA) at model layer 5 (153-231* m*) of anthropogenic emissions (left) and soil respiration (right) is shown on 15 June 2012, 02 UTC. Same area as plotted in Figure 5.11 (green cross indicates CBW).*

assimilation window, thus the correction shown in Fig. 5.12(b) is not

due to the observations on 15 June 2012. Second, it can also be seen from Fig. 5.12 that the corrections of analysed anthropogenic emissions and soil respiration are opposed to each other regarding the background. This indicates also that the correction of both fluxes is independent of the observations on 15 June 2012 and shows that the benefit of 4D-Var persists longer the assimilation window.

Flux evaluation

Flux observations are available at 5 different stations (HFS, MER, SEL, ROL, WUE) and are always independent from the analysis runs. The RMSE and the bias are presented in Fig. 5.13. An improvement of the

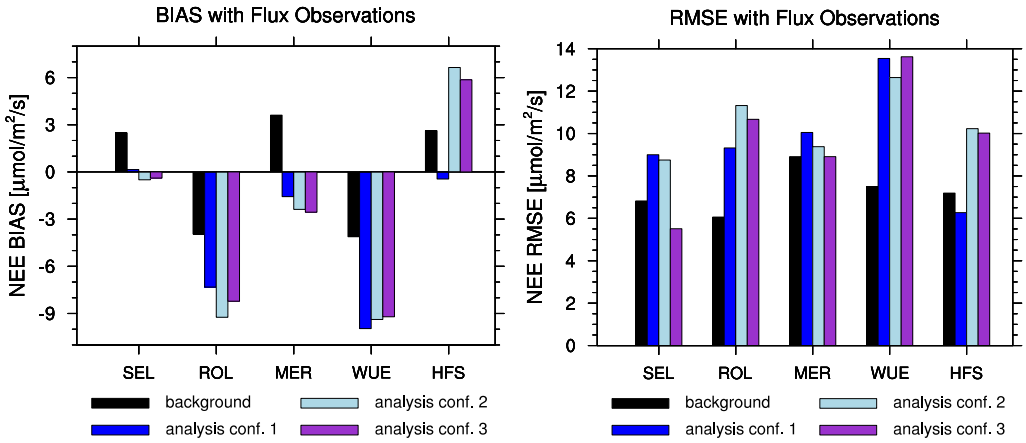
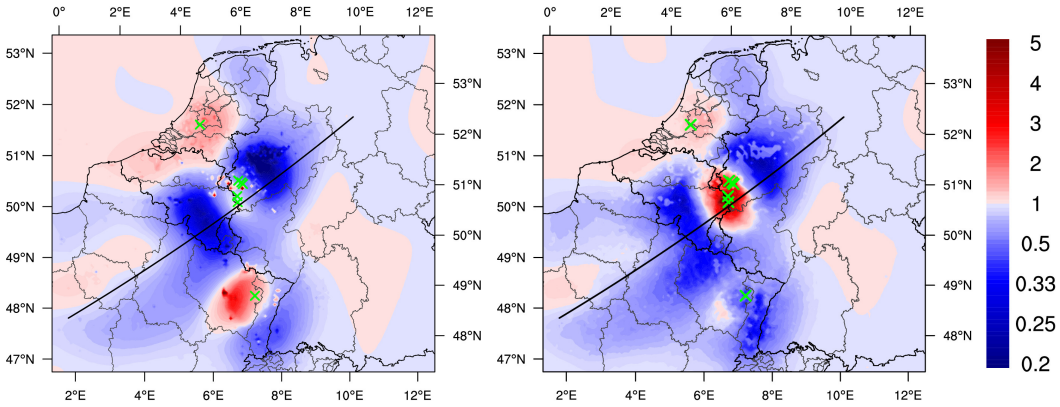


Figure 5.13: *Bias (left) and RMSE (right) of the not assimilated flux observations to the background run (black) and the analysis runs of configuration 1 (blue), 2 (light-blue), and 3 (violet) for June 2012.*

bias of the analysis run with configuration 1 can be seen at SEL, MER, and HFS. The RMSE is only improved for the analysis at station HFS, whereas the analysis RMSE deteriorates for SEL, MER, ROL, and WUE compared to the background RMSE. One reason for this is the heterogeneous spatial distribution of analysed flux factors, which are shown in Fig. 5.14 for the entire 5 km domain for 14 June 2012. In the TR32 region (measurement stations are shown as green crosses) a strong alteration of photosynthesis can be seen, whereas the analysis leaves respiration quite stable in this region for this day. A reduction of respiration and photosynthesis can be seen approximately 20 km north-east and 30 km south-west of the TR32 measurement stations. A similar structure can be seen at HFS, but only with increased respiration and quite constant photosynthesis at the region close to the station.

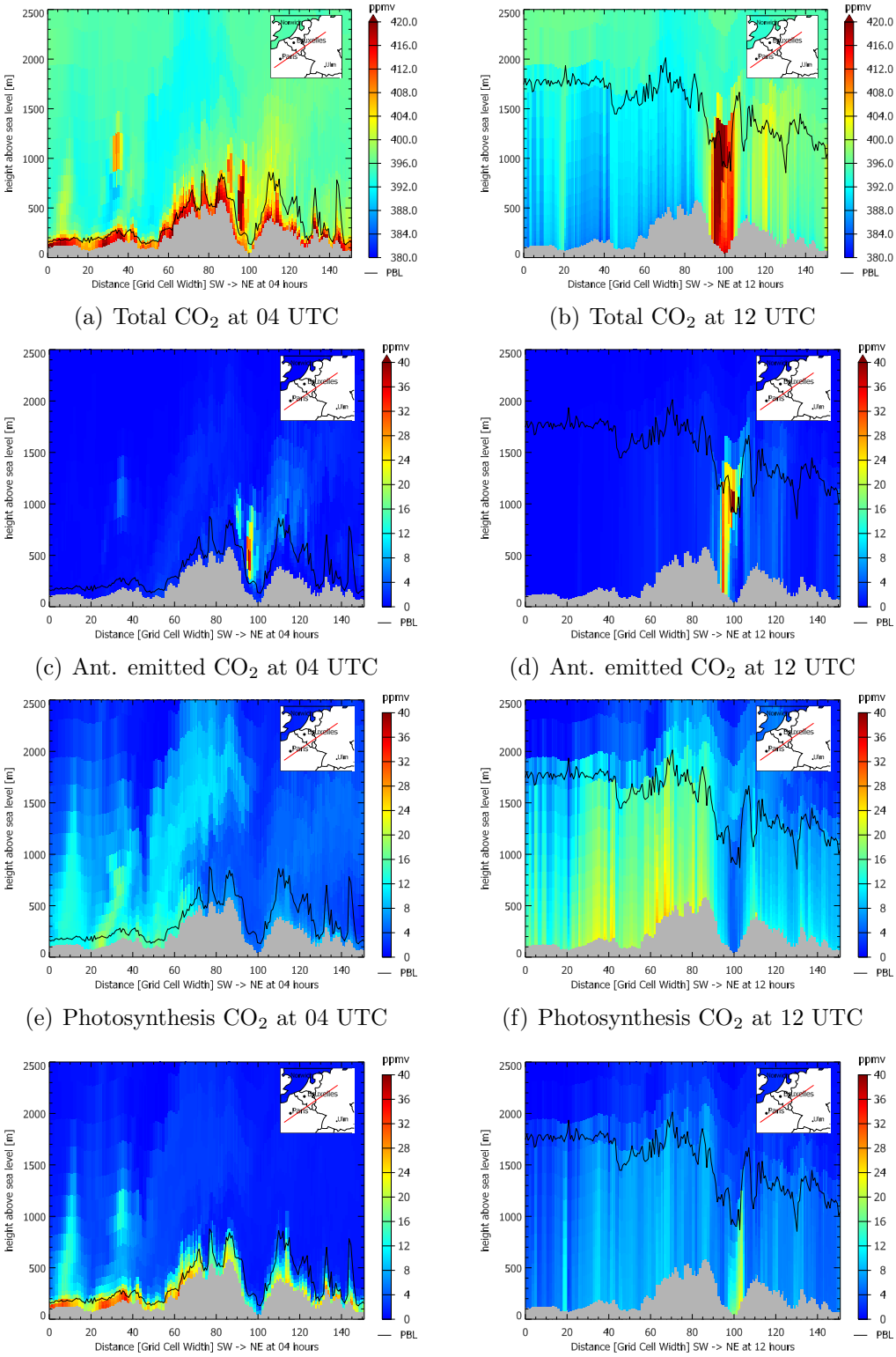
The correction of biogenic flux factors influenced by the measurements at



(a) ff of biogenic respiration (b) ff of photosynthesis

Figure 5.14: *Analysed flux factor (ff) at 5* km *domain for 14 June 2012. Measurement stations are plotted as green crosses. The black solid line defines the location for the vertical cross section in Fig. 5.15 and 5.16.*

CBW is much more homogeneous and does not show strong gradients along the main wind axis from south-west to north-east. The higher position of the assimilated measurements at 120 and 200 m a.g.l. enables a smoother analysis field of flux factors.
One reason for the spatial high gradients of analysed flux factors is the ill-posedness of the flux optimisation problem in general, which is intensified by optimising three flux factors and initial values for each grid cell separately in this work. Atmospheric transport modelling acts as a smoother for the optimisation of flux factors or initial values. Therefore, a tendency of the 4D-Var system can be seen to overestimate photosynthesis close to the assimilated measurements from the surface stations MER, ROL, and HFS. In contrast, the observations at CBW, are at higher altitudes and thus in higher model layers. The adjoint model simulates therefore smoother gradients with respect to the biogenic flux factors, resulting in a spatially more homogeneous analysed flux field around this station (Fig. 5.14).
For the examination of the composition of the total atmospheric concentration, a vertical cross section of differently tagged CO_2 tracers is presented in Fig 5.15, crossing the TR32 region from south-west to north-east. During the morning hours, soil respiration increases CO_2 concentrations up to 40 ppmV in low layers, as vertical transport is quite low due to a low PBL (Fig. 5.15(a) and 5.15(g)). After sunrise photosynthesis (juxtaposed to the other tracers) is becoming more and more dominant compared to biogenic respiration. Until 12 UTC the increase of the PBL in vertical extent leads to lower concentrations between the surface and up to 2000 m due to pho-



(a) Total CO_2 at 04 UTC

(b) Total CO_2 at 12 UTC

(c) Ant. emitted CO_2 at 04 UTC

(d) Ant. emitted CO_2 at 12 UTC

(e) Photosynthesis CO_2 at 04 UTC

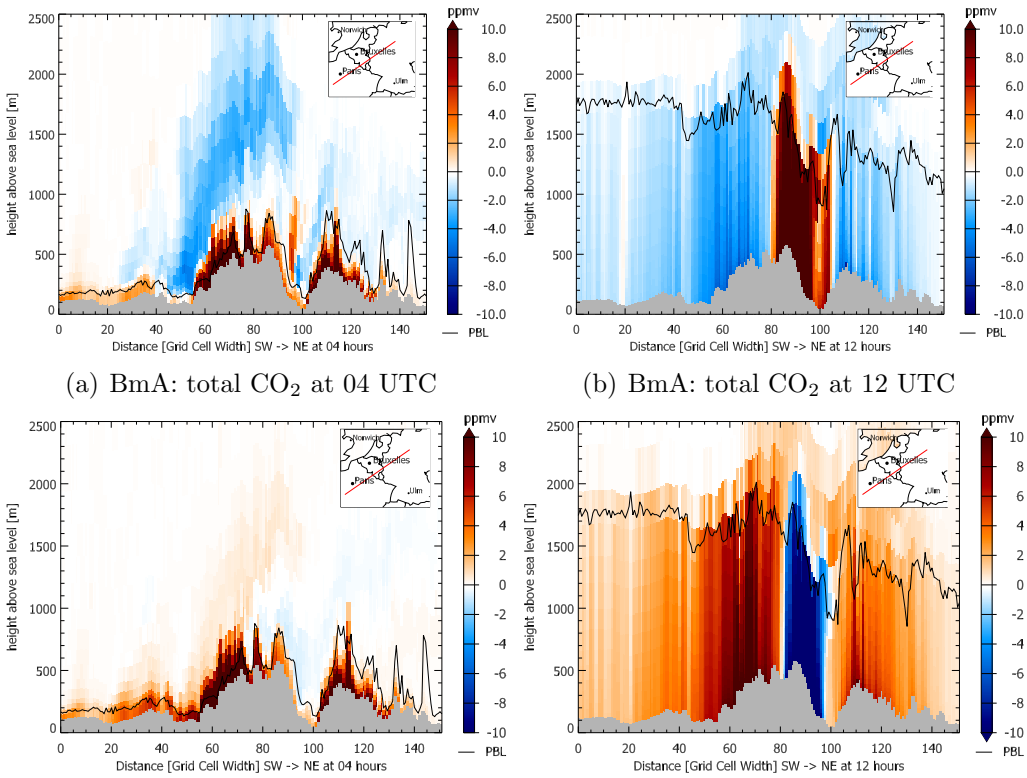
(f) Photosynthesis CO_2 at 12 UTC

(g) Soil respiration CO_2 at 04 UTC

(h) Soil respiration CO_2 at 12 UTC

Figure 5.15: *Vertical cross section of different tagged* CO_2 *tracers of the background run along the black solid line in Fig. 5.14 from south-west to north-east. Left column at 04 UTC 14 June, right column at 12 UTC 14 June 2012. Photosynthesis (third row) is opposed to the other tracers.*

tosynthesis (Fig. 5.15(b) and 5.15(f)). Anthropogenic emissions, leading to concentration increase of the same magnitude as caused by biogenic fluxes, are observed only locally due to large power plants, as can be seen in Fig. 5.15(d), induced by Weisweiler. Increased total concentration at 12 UTC between grid cell 90 and 105 is due to a combination of high anthropogenic emissions and soil respiration, respectively low photosynthesis in this region (Fig 5.15(d), 5.15(f) and 5.15(h)).
The difference of the BmA along the same cross section shown in Fig. 5.16 reveals the problem of spatially inhomogeneous optimised flux factors. The



(a) BmA: total CO_2 at 04 UTC (b) BmA: total CO_2 at 12 UTC

(c) BmA: soil respiration CO_2 at 04 UTC (d) BmA: photosynthesis CO_2 at 12 UTC

Figure 5.16: *Vertical cross section of background minus analysis run (BmA) of different tagged* CO_2 *tracers on 14 June 2012 at 04 UTC (left column) and 12 UTC (right column). Photosynthesis (d) is opposed to the other tracers.*

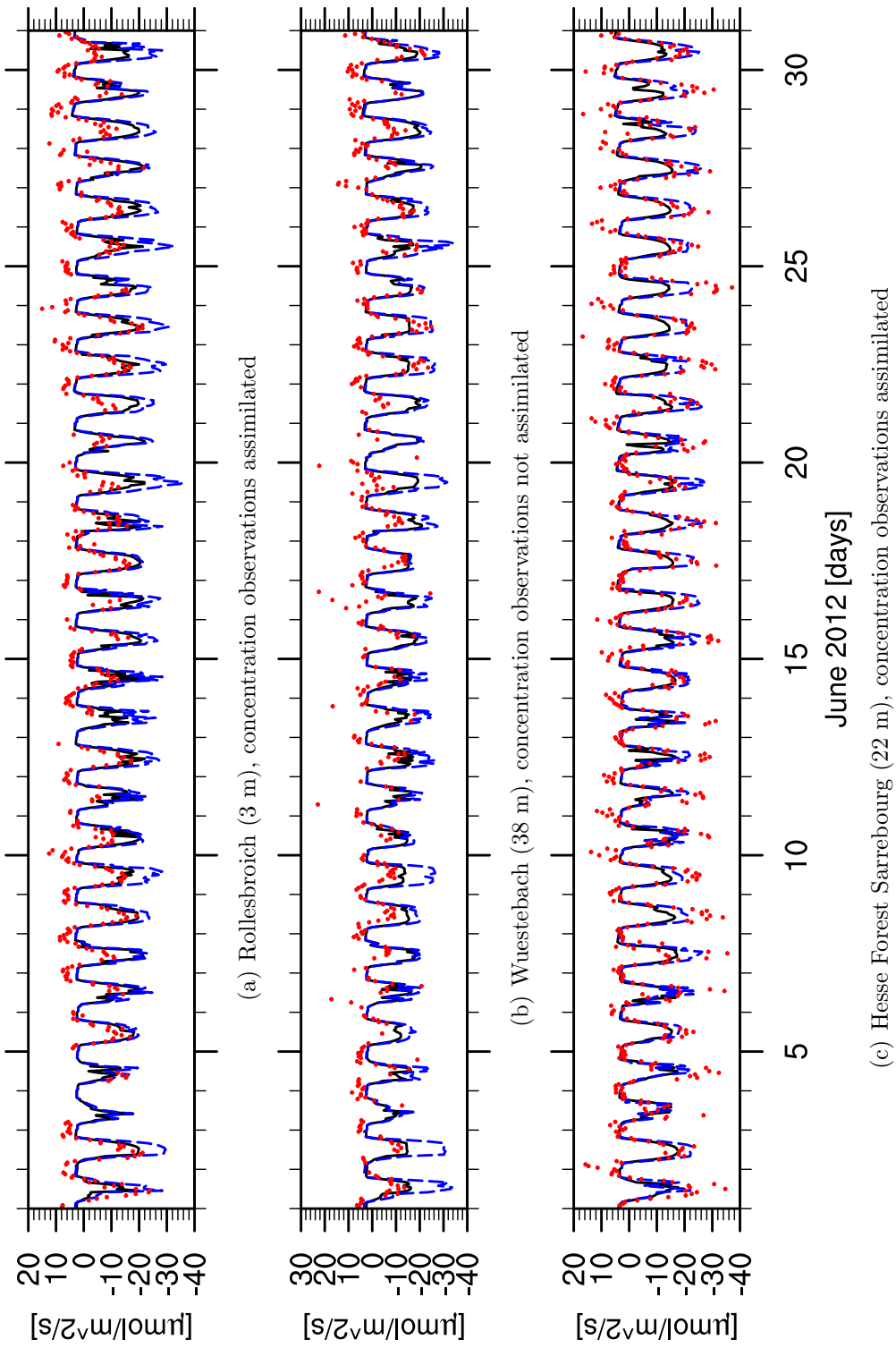
TR32 region lies approximately between grid cells 80 and 115 in Fig. 5.16. The analysis of flux factors for photosynthesis is overestimated in this region (Fig. 5.16(d)), leading to a decrease of atmospheric CO_2 compared to the background run in the TR32 region (Fig. 5.16(b)). South-west and north-east of the TR32 region the photosynthesis has lower analysed flux

factors compared to the background, leading to higher concentration values outside the TR32 region. The analysis of soil respiration reveals increased atmospheric CO_2 values compared to the background run south-west and north-east of the TR32, which is mainly visible in Fig. 5.16(a) and 5.16(c) at 04 UTC before sunrise.

Finally, the NEE time series is shown in Fig. 5.17 at ROL, WUE, and HFS. The fairly low observed photosynthesis at WUE (needle leaf forest, Fig. 5.17(b)) compared to ROL (grassland, Fig. 5.17(a)) is surprising. An overestimation of photosynthesis, which has its maximum in the early afternoon hours, can be seen in comparison to the observed NEE at several days at ROL and WUE. Respiration during night is often underestimated by the background run compared to the observations and is not sufficiently increased by the analysis.
The NEE time series at HFS (Fig. 5.17(c)) shows a better compliance of the analysis run with the observations compared to the background run. NEE fluxes during afternoon reveal the best fit of the analysis to the observations. The underestimation of the analysis run compared to observed NEE time series during night, is not confirmed by the concentration time series. Analysed concentration time series at HFS show barely lower values compared to the observations in Fig. 5.10(c).
The temporal mean of the biogenic fluxes of June 2012 is shown in Fig. 5.18, showing a very similar structure as the analysed flux factors on 14 June (Fig. 5.14). The analysis photosynthesis rate is estimated to be larger in the TR32 region compared to the background. North-east and south-west of the TR32 region analysed photosynthesis and respiration rates are lower. At HFS analysed respiration is larger and photosynthesis lower, compared to the background. A slight increase of all biogenic fluxes is seen in the area of CBW, resulting in a slightly lower NEE for this region. Therefore it is very likely, that simulated background NEE is slightly overestimated close to CBW tower for June 2012.

5.2.2 1 km domain

The analysis runs of the 1 km domain assimilate the measurements of the TR32 region, which are also used in configuration 1 of the 5 km analysis run: JUE, MER, and ROL. The stations SEL and WUE are used for independent evaluation. Three analysis runs are executed, differing slightly by parameters c^{iv}, c^{ant}, c^{phot}, and c^{resp} to model the BECM's **B** and **K**.
The RMSE and the bias of the background run and analysis configurations for concentration and flux observations are presented in Fig. 5.19 in order

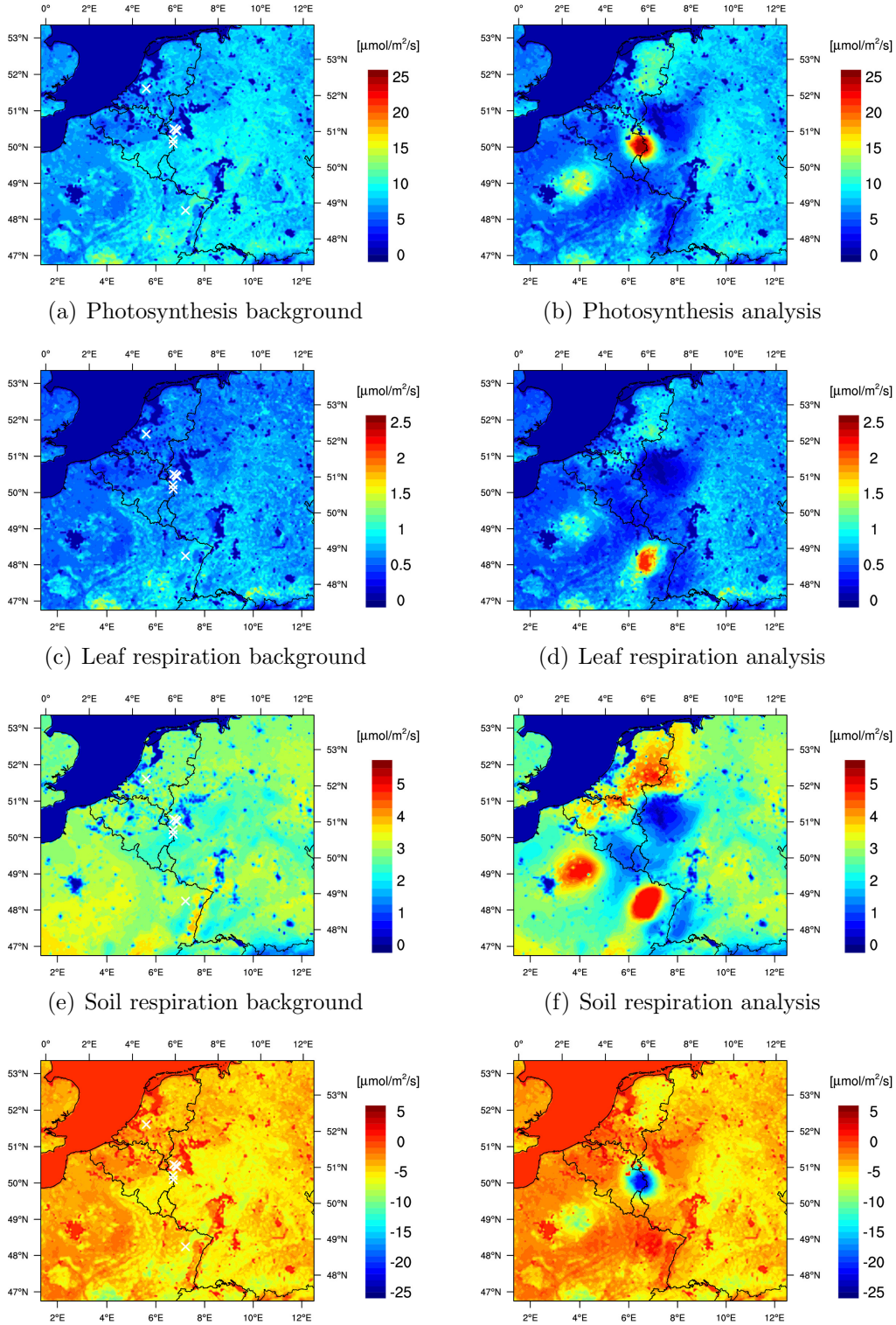


(a) Rollesbroich (3 m), concentration observations assimilated

(b) Wuestebach (38 m), concentration observations not assimilated

(c) Hesse Forest Sarrebourg (22 m), concentration observations assimilated

Figure 5.17: *NEE series of observations (red dots), background run (black solid line) and analysis run of configuration 1 (blue dashed line) with 5* km *horizontal resolution for June 2012.*



(a) Photosynthesis background

(b) Photosynthesis analysis

(c) Leaf respiration background

(d) Leaf respiration analysis

(e) Soil respiration background

(f) Soil respiration analysis

(g) NEE background

(h) NEE analysis

Figure 5.18: *Mean biogenic* CO_2 *fluxes of the background (left column) and analysis run (right column) for June 2012. Photosynthesis (first row) is opposed to the other fluxes. Measurement stations are plotted as white crosses.*

to evaluate the analysis performance. Regarding the concentrations, the

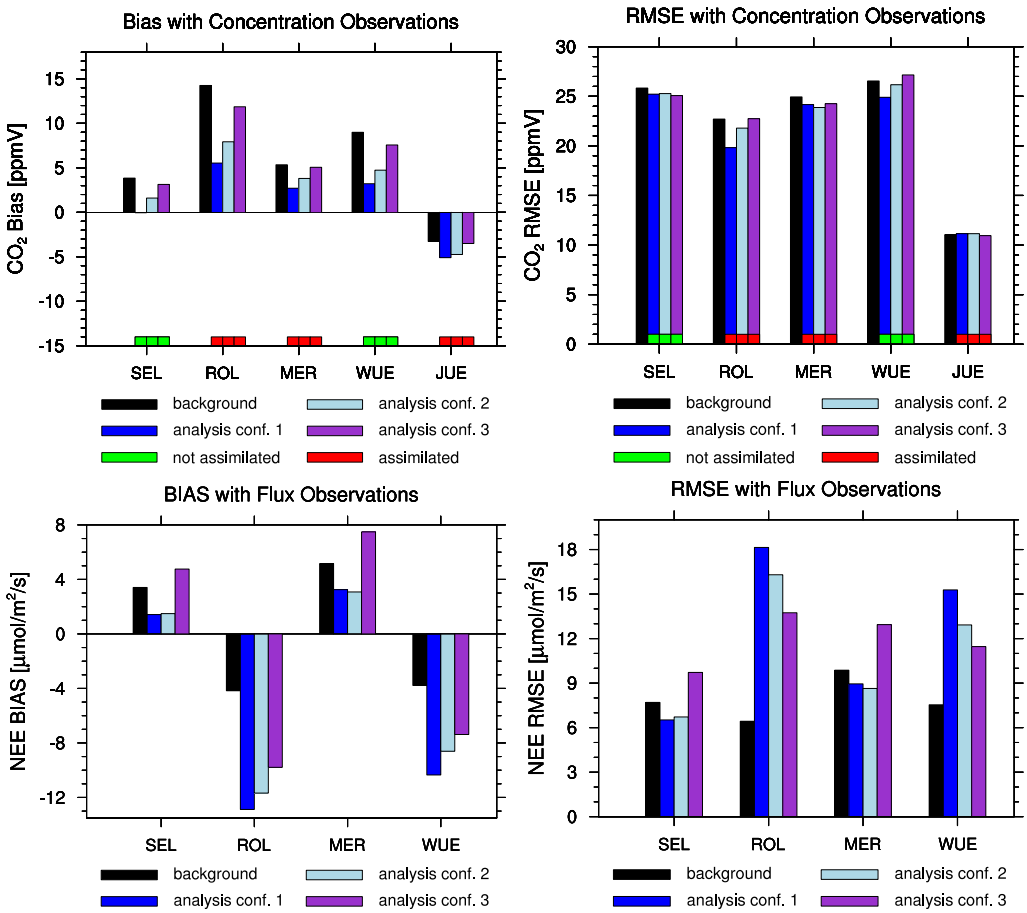


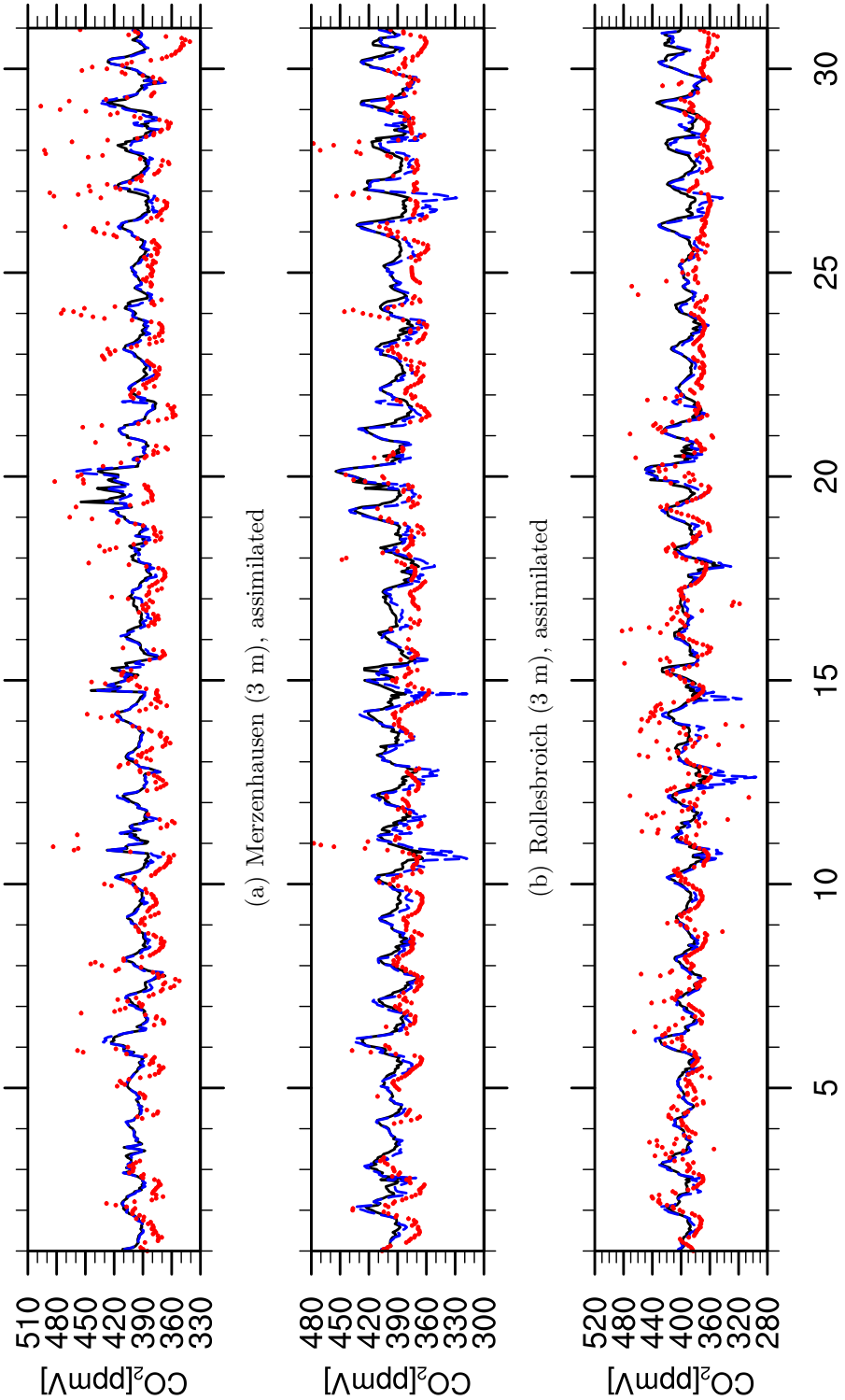
Figure 5.19: *Bias (left) and RMSE (right) of the concentration observations (upper row) and flux observations (lower row) to the background run (black) and the analysis runs of configuration 1 (blue), 2 (light-blue), and 3 (violet) with 1* km *horizontal resolution for June 2012. The concentration measurements at SEL and WUE are not assimilated for all analysis configurations.*

bias for all analysis configurations is improved compared to the background run for all stations except JUE (Fig. 5.19 upper left panel). The RMSE of all analysis configurations shows a slight reduction compared to the background run, except from configuration 3 at WUE and configuration 1 and 2 at JUE (Fig. 5.19 upper right panel).
To evaluate the analysis performance with respect to the flux observations, analysis configurations 1 and 2 are assessed first. At stations SEL and MER the bias shows a good improvement (Fig. 5.19 lower left panel), the RMSE a slight improvement (Fig. 5.19 lower right panel). This is opposed to ROL

and WUE, where a clear deterioration of the analysis performance for the bias and RMSE is seen. This deterioration is due to an overestimation of photosynthesis and will be discussed later.
Analysis configuration 3 is not able to improve either the bias nor the RMSE at the four stations. The reason is an underestimation of respiration and photosynthesis at SEL and MER and an overestimation of the two biogenic fluxes in the area of ROL and WUE, which is not further evaluated here.

Analysis configuration 2, showing on average the best improvement, is evaluated in more detail in the following. Time series of June 2012 are plotted for concentrations in Fig. 5.20 and for NEE in Fig. 5.21 at MER, ROL, and WUE. MER is presented instead of SEL since the time series at both stations show similar structures, but an observation gap at SEL between 6 and 15 June hampers evaluation. A comparison of background and observation NEE (Fig. 5.21(a)) reveals an underestimation of photosynthesis, which is improved slightly for several days by the analysis. Consequently, concentration time series at MER show lower observations compared to the model during afternoon hours of several days (Fig. 5.20(a)).The observed concentration peaks during nights e.g. on 26-29 June are due to a combination of underestimated respiration and the low measurement height of 3 m a.g.l., which is discussed in the case of the 5 km domain.
The concentration time series at ROL (Fig. 5.20(b)) shows the strongest bias of observations and background model (bias=14.3 ppmV), which is one reason for the overestimation of analysed photosynthesis seen in Fig. 5.21(b). The concentration time series of the analysis also overestimates the observations (bias=3.8 ppmV), except on 10, 14 and 26 June. Together with the strong negative bias of analysed NEE at ROL (bias=-11.7 $\frac{\mu\mathrm{mol}}{\mathrm{m^2s}}$), this indicates also a too coarse vertical model resolution for the measurement height of 3 m a.g.l.
At WUE the non-assimilated concentration observations and the analysis agree well until 22 June 2012 (Fig. 5.20(c)). The observed concentrations of the following days show a weak daily amplitude compared to earlier days. The observed fluxes in contrast (Fig. 5.21(c)), do not show a weaker daily amplitude for NEE between 22 and 30 June. This can be explained by an increase of observed friction velocity u^* on 22 June, which indicates enhanced mixing between vertical layers. This increase of turbulent mixing can be seen as well in the EURAD-IM. The strong diurnal amplitude of the fluxes is reflected by the modelled concentrations of the 5 lowest model layers ($0 - 230$ m), leaving the question open why model and observation time series differ between 22 and 30 June at 38 m a.g.l..

The overestimation of photosynthesis at ROL and WUE is mainly caused

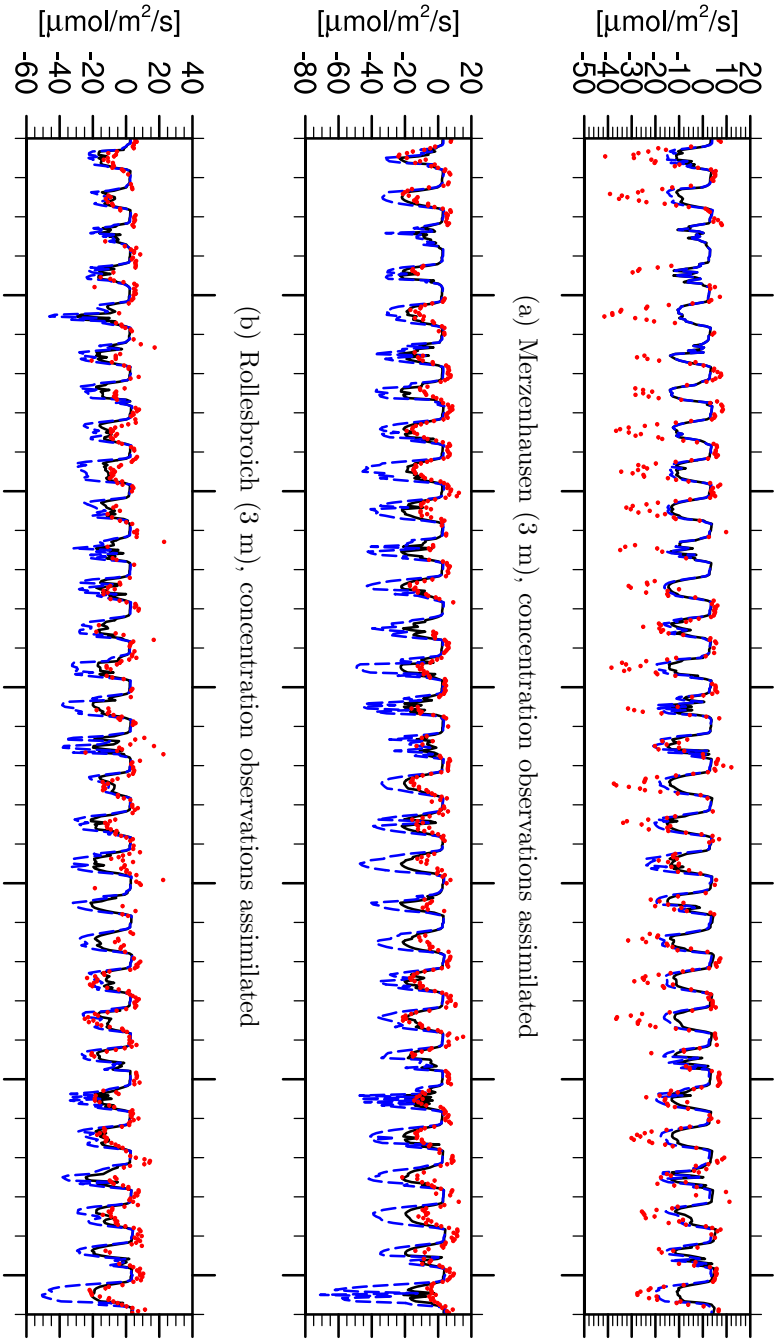


(a) Merzenhausen (3 m), assimilated

(b) Rollesbroich (3 m), assimilated

(c) Wuestebach (38 m), not assimilated

Figure 5.20: *Concentration time series of observations (red dots), background (black solid line), and analysis run of configuration 2 (blue dashed line) for the 1* km *domain for June 2012.*

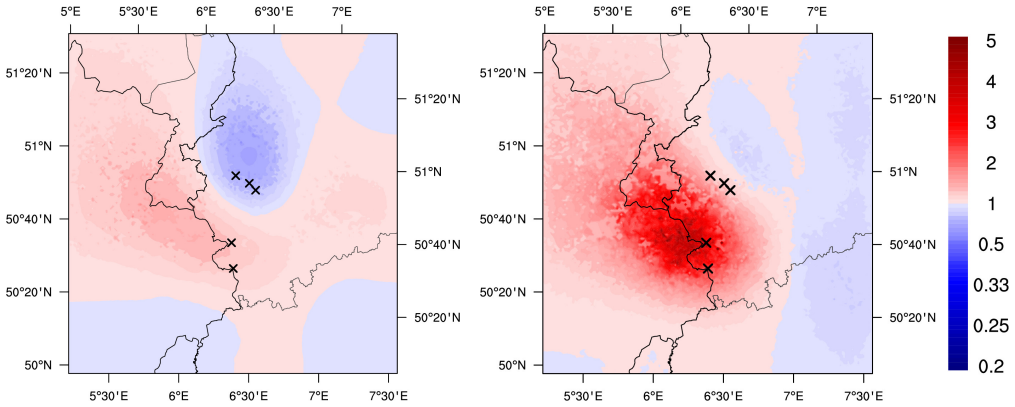


(a) Merzenhausen (3 m), concentration observations assimilated

(b) Rollesbroich (3 m), concentration observations assimilated

(c) Wuestebach (38 m), concentration observations not assimilated

Figure 5.21: *NEE time series of observations (red dots), background run (black solid line) and analysis run of configuration 2 (blue dashed line) with 1* km *horizontal resolution for June 2012.*



(a) ff of biogenic respiration (b) ff of photosynthesis

Figure 5.22: *Analysed flux factor (ff) of the 1* km *domain for biogenic respiration (left panel) and photosynthesis (right panel) on 26 June 2012. Measurement stations are plotted as black crosses.*

by the boundary values of the 5 km analysis run. A much shorter assimilation window or the optimisation of boundary values is required to improve the analysis. The 5 km analysis run estimated low photosynthesis and respiration rate outside the 1 km domain (Fig. 5.14). Inside the 1 km domain, photosynthesis was analysed to be higher compared to the background run. Due to the small spatial extent, the atmospheric CO_2 concentration of the 1 km is strongly determined by its boundary values. Therefore, the analysis of the flux factors for photosynthesis and respiration of the 1 km domain (Fig. 5.22) is similar to that of the inner 5 km domain, which holds also true for the mean biogenic fluxes for June 2012.

The analysis of the four biggest power plants of the 1 km domain, Weisweiler, Niederaußem, Frimmersdorf, and Neurath, is presented in Fig. 5.23. Since these four point sources accumulate to more than 25 % of all anthropogenic CO_2 emissions in Northrine-Westfalia, a large part of GHG emissions of the model domain is analysed. Emissions of all four sources are corrected downwards. The gradient of the fluxes reveals during which days the emissions are observed by one of the measurement stations. Using the temporal mean of the flux factors, the total amount of CO_2, induced by these power plants for June 2012, is presented in Table 5.5.

Since the 1 km domain is heavily influenced by the boundary values of the simulations at the 5 km domain, the analysed biogenic fluxes with 1 km resolution reveal a similar spatial pattern as the analysed fluxes at the 5 km domain. Therefore photosynthesis is overestimated in the TR32 region and

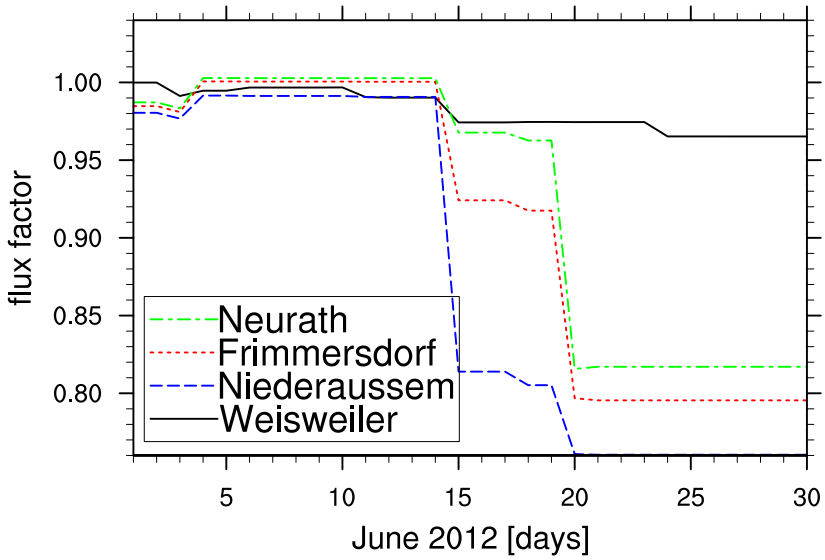


Figure 5.23: *Analysed flux factors for Weisweiler, Niederaußem, Frimmersdorf, and Neurath for June 2012.*

Table 5.5: CO_2 *emissions of Weisweiler, Niederaußem, Frimmersdorf, and Neurath for June 2012 in tonnes per month.*

Power plant	Background Emissions CO_2 [t/m]	Analysis Emissions CO_2 [t/m]
Weisweiler	1.87 E+06	1.83 E+06
Niederaußem	2.63 E+06	2.31 E+06
Frimmersdorf	1.34 E+06	1.22 E+06
Neurath	1.72 E+06	1.60 E+06

underestimated in areas north-east and south-west of this region. Respiration is underestimated outside of the TR32 region as well. The reasons for the spatial gradients of analysed biogenic were already discussed in detail in Sect. 5.2.1, with regard to the vertical resolution of the model, the low measurement height within the TR32 region, and the missing smoothing properties of the atmospheric transport.

Chapter 6

Summary and Outlook

6.1 Summary

This study investigated the possibilities and limits of CO_2 flux estimation by assimilating concentration time series with the 4D-Var system EURAD-IM. CO_2 surface-atmosphere fluxes anthropogenic emissions, photosynthesis, and biogenic respiration are optimised jointly with the initial atmospheric concentration at high spatio-temporal resolution for each grid cell.
This approach involves a high degree of freedom of the optimisation parameters. On the one hand, this provides new opportunities to infer surface-atmosphere CO_2 fluxes from concentration observations. On the other hand, the increased ill-posedness of the CO_2 inversion problem must be controlled.

To enable CO_2 inversion with the EURAD-IM by optimising four parameters jointly, the following developments and modifications were implemented. Biogenic fluxes are simulated with the land surface model CLM 4.0, whose diurnal cycle is optimised in EURAD-IM. Anthropogenic CO_2 emissions of the TNO inventory are scaled down with additional land use information to 5 km and 1 km horizontal resolution. To interpret the quite weak signal of surface-atmosphere fluxes in atmospheric CO_2 concentration time series the absolute monotone Walcek [*Walcek*, 2000] advection scheme has been implemented. The adjoint advection routine is implemented such that the properties of the forward advection scheme are preserved. This allows also a precise processing of the gradient of the optimisation parameters backwards in time. The ratio of the different gradients with respect to the initial values and flux factors is derived. Based on this derivation the variance of the background error covariance matrix for flux factors $\mathbf{K}$ is modelled in a new manner.

Identical twin experiments reveal a higher sensitivity for the optimisation of small fluxes by the new modelling compared to earlier implementations of **K**. Further experiments with a spatial one-dimensional version of the EURAD-IM show the impact of the choice of the background error covariance matrices **B** and **K** to the joint analysis result. It is possible to reproduce measured time series by several combinations of initial values and flux factors. The error covariance matrices determine to which extent initial values and flux factors for photosynthesis, respiration, and anthropogenic emissions are adjusted by the analysis.
Identical twin experiments with the EURAD-IM present the influence of disturbed initial values for the analysis of CO_2 fluxes. Concentration time series at the location of the observations may still show a sufficient compliance between synthetic measurements and the analysis of optimising flux factors only. Nevertheless, the optimised flux factors show severe deterioration, if the uncertainty of initial values is not taken into account. Joint optimisation leads to a substantially better analysis of flux factor compared to that of optimising flux factors only. The minimum of the cost function is 20 times smaller on average for the joint analysis compared to the flux factor analysis.

An experiment to analyse surface-atmosphere CO_2 fluxes for the TR32 region during June 2012 was executed. 4D-Var data assimilation is performed for the domain with 5 km and 1 km horizontal resolution. The evaluation of concentration time series at Selhausen demonstrates that the vertical model resolution has to be refined for a better analysis of measurements positioned 3 m a.g.l.. Observed concentration peaks, by accumulated respiration during night, are difficult to analyse due to the coarse model surface layer. Time series at Hesse Forest Sarrebourg and Cabauw show a good compliance of the analysis to the observation. The analysis run shows a significant improvement compared to the background run for high observed concentration peaks during night at these stations.
The analysed flux factors for photosynthesis and respiration show strong spatial gradients in the area of the measurement stations, which are located in the surface layer. Fluxes close to these stations are over- or underestimated by the 4D-Var system. The main cause is the ill-posed nature of the CO_2 inversion problem, which is intensified by the chosen optimisation strategy. A better analysis of fluxes is seen close to the Cabauw tower, where measurements are assimilated at 120 m and 200 m a.g.l.. The smoothing properties of atmospheric transport result in smooth analysed flux factors, where less over- or underestimation can be seen.
An example of the long term impact of 4D-Var is investigated in detail. A

high peak of background concentrations on 15 June 02 UTC at Cabauw tower, is reduced by the analysis in compliance with the observations. It is shown that the high background peak is due to fluxes released in the previous assimilation window, which are corrected by the analysis. It could be shown that elevated concentrations due to soil respiration are traced back and corrected by the EURAD-IM.

The 4D-Var System EURAD-IM demonstrated the ability to infer CO_2 fluxes from temporal high resolved concentration time series, with a large degree of freedom for the optimisation space. The main obstacle for the chosen optimisation strategy of this work is the ill-posedness of the problem. High requirements have to be fulfilled to analyse surface-atmosphere fluxes successfully regarding the modelling of atmospheric transport, the a priori fluxes themselves, the initial CO_2 concentration, and the correct specification of error covariance matrices.

6.2 Outlook

To facilitate the solution of the ill-posed top-down CO_2 inverse problem the degree of freedom of the solution space has to be reduced or further constraints, i.e. more observations, are needed. Measurements from radon and measurements for the closure of the energy budget were used by *Tolk et al.* [2009] to account for boundary layer characteristics. The bottom-up approach, which uses direct measurement of NEE, can be applied to obtain more reliable a priori fluxes for EURAD-IM. This allows for smaller background errors for biogenic fluxes, which reduces the solution space.
A further improvement of the analysed fluxes could be achieved by an enlargement of the assimilation window. Measurements at high model layers contain often signals from surface fluxes, which are released more than 24 hours ago. Using longer assimilation windows requires also the treatment of transport errors, as it is feasible with weak constraint 4D-Var.

Appendix A

A.1 The adjoint model $\mathbf{M}^{\mathrm{T}}$

To derive the adjoint model we combine Eq. (3.8) and (3.9) of the forward run and introduce the following notation (remember that the transport operator (see Eq. (3.15)) $\mathbf{T}$ merges the advection and diffusion operator considering the operator splitting of the EURAD-IM: $\mathbf{T}_{i,i+1/2} = \mathbf{D}_{i,i+1/2}\mathbf{A}^1_{i,i+1/2}$, $\mathbf{T}_{i+1/2,i+1} = \mathbf{A}^2_{i+1/2,i+1}\mathbf{D}_{i+1/2,i+1}$)

$$\widetilde{\mathbf{T}}_{i,i+1/2} := \begin{pmatrix} \mathbf{T}_{i,i+1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad \widetilde{\mathbf{T}}_{i+1/2,i+1} := \begin{pmatrix} \mathbf{T}_{i+1/2,i+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \tag{A.1}$$

$$\widetilde{\mathbf{F}}_{i+1/2} := \begin{pmatrix} \mathbf{I} & \triangle t\, \mathbf{U}^{\mathrm{b}}_{i+1/2} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{A.2}$$

We can now write in accordance with Eq. (3.8) and (3.9)

$$\begin{pmatrix} \mathbf{x}_{i+1} \\ \mathbf{f} \end{pmatrix} = \underbrace{\widetilde{\mathbf{T}}_{i+1/2,i+1}\, \widetilde{\mathbf{F}}_{i+1/2}\, \widetilde{\mathbf{T}}_{i,i+1/2}}_{=\mathbf{M}_{i,i+1}} \begin{pmatrix} \mathbf{x}_{i} \\ \mathbf{f} \end{pmatrix}, \tag{A.3}$$

as $\widetilde{\mathbf{F}}$ is equivalent to Eq. (3.9). Using Eq. (A.3), the adjoint model can be derived

$$(\mathbf{M}_{i,i+1})^{\mathrm{T}} \begin{pmatrix} \mathbf{x}^*_{i+1} \\ \mathbf{f}^*_{i+1} \end{pmatrix} = \left(\widetilde{\mathbf{T}}_{i,i+1/2}\right)^{\mathrm{T}} \left(\widetilde{\mathbf{F}}_{i+1/2}\right)^{\mathrm{T}} \left(\widetilde{\mathbf{T}}_{i+1/2,i+1}\right)^{\mathrm{T}} \begin{pmatrix} \mathbf{x}^*_{i+1} \\ \mathbf{f}^*_{i+1} \end{pmatrix} \tag{A.4}$$

$$= \begin{pmatrix} (\mathbf{T}_{i,i+1/2})^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \triangle t\, \mathbf{U}^{\mathrm{b}}_{i+1/2} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\mathbf{T}_{i+1/2,i+1})^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}^*_{i+1} \\ \mathbf{f}^*_{i+1} \end{pmatrix} \tag{A.5}$$

$$= \begin{pmatrix} (\mathbf{T}_{i,i+1/2})^{\mathrm{T}} (\mathbf{T}_{i+1/2,i+1})^{\mathrm{T}} \mathbf{x}^*_{i+1} \\ \triangle t\, \mathbf{U}^{\mathrm{b}}_{i+1/2}\, (\mathbf{T}_{i+1/2,i+1})^{\mathrm{T}} \mathbf{x}^*_{i+1} + \mathbf{f}^*_{i+1} \end{pmatrix} \tag{A.6}$$

which is Eq. (3.14). Equation (A.6) can now be used iteratively for the calculation of $(\mathbf{M}_{0,i})^{\mathrm{T}}$. For the sake of shorter notation we write $\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\left[\mathbf{d}_i - \mathbf{H}_i\mathbf{M}_{0,i}\begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix}\right] =: \begin{pmatrix}\mathbf{x}_i^*\\ \mathbf{0}\end{pmatrix}$,

$$(\mathbf{M}_{0,i})^{\mathrm{T}}\begin{pmatrix}\mathbf{x}_i^*\\ \mathbf{0}\end{pmatrix} = (\mathbf{M}_{0,i-1})^{\mathrm{T}}\begin{pmatrix}(\mathbf{T}_{i-1,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \triangle t\,\mathbf{U}^{\mathrm{b}}_{i-1/2}\,(\mathbf{T}_{i-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix} \tag{A.7}$$

$$= (\mathbf{M}_{0,i-2})^{\mathrm{T}}\begin{pmatrix}(\mathbf{T}_{i-2,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \triangle t\,\mathbf{U}^{\mathrm{b}}_{i-1-1/2}\,(\mathbf{T}_{i-1-1/2,i})^{\mathrm{T}}\mathbf{x}_i^* + \triangle t\,\mathbf{U}^{\mathrm{b}}_{i-1/2}\,(\mathbf{T}_{i-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix} \tag{A.8}$$

$$= \begin{pmatrix}(\mathbf{T}_{0,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \sum_{j=1}^{i}\triangle t\,\mathbf{U}^{\mathrm{b}}_{j-1/2}\,(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix}, \tag{A.9}$$

which is Eq. (3.25).

A.2 Derivation of the diagonal of K

In this Sect. the specific construction of the diagonal of $\mathbf{K}$ (Eq.(3.26)) and an approximation of the gradient of the cost function (Eq. (3.27)) is shown. First an approximation for the flux factor part of $\sum_{i=0}^{N}(\mathbf{M}_{0,i})^{\mathrm{T}}$ (Eq. (3.18)) is derived with the crude approximation

$$\mathbf{U}^{\mathrm{b}}_{j-1/2} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{U}^{\mathrm{b}}_{i-1/2} =: \frac{1}{N}\widehat{\mathbf{U}^{\mathrm{b}}} \tag{A.10}$$

Using Eq. (A.9) and the notation $[\cdot]^{\mathrm{FF}}$ as the flux factor part we can write

$$\left[\sum_{i=0}^{N}(\mathbf{M}_{0,i})^{\mathrm{T}}\begin{pmatrix}\mathbf{x}_i^*\\ \mathbf{0}\end{pmatrix}\right]^{\mathrm{FF}} \overset{(\mathrm{A.9})}{=} \triangle t\sum_{i=0}^{N}\left(\sum_{j=1}^{i}\mathbf{U}^{\mathrm{b}}_{j-1/2}(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\right)\mathbf{x}_i^* \tag{A.11}$$

$$\overset{(\mathrm{A.10})}{\approx} \frac{\triangle t}{N}\widehat{\mathbf{U}^{\mathrm{b}}}\sum_{i=0}^{N}\left(\sum_{j=1}^{i}(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\right)\mathbf{x}_i^*. \tag{A.12}$$

Using additionally the construction of diag($\mathbf{K}$) (c designates a constant number)

$$\sqrt{K_{r,r}} = \mathrm{c}\,\frac{N}{\triangle t}\left[\max_s\left|\widehat{\mathbf{U}^{\mathrm{b}}}(s)\right|\right]^{-\frac{1}{l}}\left[\left|\widehat{\mathbf{U}^{\mathrm{b}}}(r)\right|\right]^{-\frac{l-1}{l}}, \tag{A.13}$$

the gradient of the preconditioned cost function can be approximated

$$\nabla\mathcal{J}(\mathbf{v},\mathbf{w}) \overset{(3.18)}{=} \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathbf{K}^{\mathrm{T}/2}\end{pmatrix} \sum_{i=0}^{N} (\mathbf{M}_{0,i})^{\mathrm{T}} \begin{pmatrix}\mathbf{x}_i^*\\ \mathbf{0}\end{pmatrix} \tag{A.14}$$

$$\overset{(A.9)}{=} \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathbf{K}^{\mathrm{T}/2}\end{pmatrix} \sum_{i=0}^{N} \begin{pmatrix}(\mathbf{T}_{0,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \sum_{j=1}^{i} \triangle t\, \mathbf{U}^{\mathrm{b}}_{j-1/2}\, (\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix} \tag{A.15}$$

$$\overset{(A.12)}{\approx} \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \mathbf{K}^{\mathrm{T}/2}\left(\mathrm{c}\,\frac{\triangle t}{N}\widehat{|\mathbf{U}^{\mathrm{b}}|}\right)\end{pmatrix} \sum_{i=0}^{N} \begin{pmatrix}(\mathbf{T}_{0,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \sum_{j=1}^{i}(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix} \tag{A.16}$$

$$\overset{(A.13)}{\approx} \begin{pmatrix}\mathbf{v}\\ \mathbf{w}\end{pmatrix} - \begin{pmatrix}\mathbf{B}^{\mathrm{T}/2} & 0\\ 0 & \left(\mathrm{c}\,\frac{\widehat{\mathbf{U}}^{\mathrm{b}}}{\max_s |\widehat{\mathbf{U}^{\mathrm{b}}}(s)|}\right)^{\frac{1}{l}}\end{pmatrix} \sum_{i=0}^{N} \begin{pmatrix}(\mathbf{T}_{0,i})^{\mathrm{T}}\mathbf{x}_i^*\\ \sum_{j=1}^{i}(\mathbf{T}_{j-1/2,i})^{\mathrm{T}}\mathbf{x}_i^*\end{pmatrix}, \tag{A.17}$$

such that we get Eq. (3.27). As already mentioned in Sect. 3.4.2 and 3.4.4.2 the notation presented here is straightforward to apply also for three fluxes but is avoided here for the sake of clarity.

Bibliography

Alfieri, J. G., W. P. Kustas, J. H. Prueger, L. E. Hipps, J. L. Chávez, A. N. French and S. R. Evett, Intercomparison of nine micrometeorological stations during the BEAREX08 field campaign, *Journal of Atmospheric and Oceanic Technology*, *28*, (11), 1390–1406, 2011.

Andres, R. J., T. A. Boden, F.-M. Bréon, P. Ciais, S. Davis, D. Erickson, J. S. Gregg, A. Jacobson, G. Marland, J. Miller et al., A synthesis of carbon dioxide emissions from fossil-fuel combustion, *Biogeosciences*, *9*, 2012.

Aubinet, M., A. Grelle, A. Ibrom, Ü. Rannik, J. Moncrieff, T. Foken, A. Kowalski, P. Martin, P. Berbigier, C. Bernhofer et al., *Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology.*, vol 30, 2000.

Aumann, H. H., M. T. Chahine, C. Gautier, M. D. Goldberg, E. Kalnay, L. M. McMillin, H. Revercomb, P. W. Rosenkranz, W. L. Smith, D. H. Staelin et al., AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems, *IEEE Transactions on Geoscience and Remote Sensing*, *41*, (2), 253–264, 2003.

Austin, J., Toward the four dimensional assimilation of stratospheric chemical constituents, *Journal of Geophysical Research: Atmospheres*, *97*, (D2), 2569–2588, 1992.

Baker, I., L. Prihodko, A. Denning, M. Goulden, S. Miller and H. Da Rocha, Seasonal drought stress in the amazon: Reconciling models and observations, *Journal of Geophysical Research: Biogeosciences*, *113*, (G1), 2008.

Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, M. Jung, N. Carvalhais, C. Rödenbeck, M. A. Arain, D. Baldocchi, G. B.

Bonan et al., Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate, *Science*, *329*, (5993), 834–838, 2010.

Bennett, A. F., *Inverse methods in physical oceanography*, Cambridge university press, 1992.

Bernacchi, C. J., J. E. Bagley, S. P. Serbin, U. M. Ruiz-Vera, D. M. Rosenthal and A. VanLoocke, Modelling C3 photosynthesis from the chloroplast to the ecosystem, *Plant, Cell & Environment*, *36*, (9), 1641–1657, 2013.

Bouttier, F. and P. Courtier, Data assimilation concepts and methods March 1999, *Meteorological training course lecture series. ECMWF*, 2002.

Broquet, G., F. Chevallier, P. Rayner, C. Aulagnier, I. Pison, M. Ramonet, M. Schmidt, A. T. Vermeulen and P. Ciais, A European summertime CO2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *J. Geophys. Res.: Atmospheres (1984–2012)*, *116*, (D23), 2011.

Broquet, G., F. Chevallier, F.-M. Bréon, N. Kadygrov, M. Alemanno, F. Apadula, S. Hammer, L. Haszpra, F. Meinhardt, J. Morguí et al., Regional inversion of CO2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, *Atmospheric Chemistry and Physics*, *13*, (17), 9039–9056, 2013.

Bruhwiler, L., A. Michalak, W. Peters, D. Baker and P. Tans, An improved Kalman Smoother for atmospheric inversions, *Atmospheric Chemistry and Physics*, *5*, (10), 2691–2702, 2005.

Calvet, J., J. Noilhan and P. Bessemoulin, Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements, *Journal of Applied Meteorology*, *37*, (4), 371–386, 1998.

Canadell, J. G., C. Le Quéré, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R. Houghton and G. Marland, Contributions to accelerating atmospheric co2 growth from economic activity, carbon intensity, and efficiency of natural sinks, *Proceedings of the National Academy of Sciences*, *104*, (47), 18866–18870, 2007.

Cardinali, C., S. Pezzulli and E. Andersson, Influence-matrix diagnostic of a data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *130*, (603), 2767–2786, 2004.

Carouge, C., P. Bousquet, P. Peylin, P. Rayner and P. Ciais, What can we learn from european continuous atmospheric CO2 measurements to quantify regional fluxes–Part 1: Potential of the 2001 network, *Atmos. Chem. Phys*, *10*, (6), 3107–3117, 2010.

Chevallier, F., M. Fisher, P. Peylin, S. Serrar, P. Bousquet, F.-M. Bréon, A. Chédin and P. Ciais, Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research: Atmospheres*, *110*, (D24), 2005.

Ciais, P., P. P. Tans, J. W. White, M. Trolier, R. J. Francey, J. A. Berry, D. R. Randall, P. J. Sellers, J. G. Collatz and D. S. Schimel, Partitioning of ocean and land uptake of CO2 as inferred by δ13C measurements from the NOAA Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network, *Journal of Geophysical Research: Atmospheres*, *100*, (D3), 5051–5070, 1995.

Collatz, G., J. Ball, C. Grivet and J. A. Berry, Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, (24), 107 – 136, 1991.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, *120*, (519), 1367–1387, 1994.

Crevoisier, C., C. Sweeney, M. Gloor, J. L. Sarmiento and P. P. Tans, Regional US carbon sinks from three-dimensional atmospheric CO2 sampling, *Proceedings of the National Academy of Sciences*, *107*, (43), 18348–18353, 2010.

Daley, R., Atmospheric data analysis, Cambridge atmospheric and space science series, *Cambridge University Press*, *6966*, 25, 1991.

Daley, R. and E. Barker, NAVDAS: Formulation and diagnostics, *Monthly Weather Review*, *129*, (4), 869–883, 2001.

Deacon, E., The measurement of turbulent transfer in the lower atmosphere, *Advances in Geophysics*, *6*, 211–228, 1959.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer et al., The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, *137*, (656), 553–597, 2011.

Denning, A. S., M. Holzer, K. R. Gurney, M. Heimann, R. M. Law, P. J. Rayner, I. Y. Fung, S.-M. FAN, S. Taguchi, P. Friedlingstein et al., Three-dimensional transport and concentration of SF6, *Tellus B*, *51*, (2), 266–297, 1999*a*.

Denning, A. S., T. Takahashi and P. Friedlingstein, Can a strong atmospheric CO2 rectifier effect be reconciled with a reasonable carbon budget?, *Tellus B*, *51*, (2), 249–253, 1999*b*.

Desai, A. R., Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape, *Journal of Geophysical Research: Biogeosciences*, *115*, (G3), 2010.

Desai, A. R., D. J. Moore, W. K. Ahue, P. T. Wilkes, S. F. De Wekker, B. G. Brooks, T. L. Campos, B. B. Stephens, R. K. Monson, S. P. Burns et al., Seasonal pattern of regional carbon balance in the central Rocky Mountains from surface and airborne measurements, *Journal of Geophysical Research: Biogeosciences*, *116*, (G4), 2011.

Dietze, M. C., R. Vargas, A. D. Richardson, P. C. Stoy, A. G. Barr, R. S. Anderson, M. A. Arain, I. T. Baker, T. A. Black, J. M. Chen et al., Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis, *Journal of Geophysical Research: Biogeosciences*, *116*, (G4), 2011.

Dowd, J. E. and D. S. Riggs, A comparison of estimates of michaelis-menten kinetic constants from various linear transformations, *J. biol. Chem*, *240*, (2), 863–869, 1965.

Elbern, H. and H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.: Atmospheres (1984–2012)*, *106*, (D4), 3569–3590, 2001.

Elbern, H., H. Schmidt and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *Journal of Geophysical Research: Atmospheres*, *102*, (D13), 15967–15985, 1997.

Elbern, H., H. Schmidt, O. Talagrand and A. Ebel, 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Ecol. Stat.*, *15*, (67), 539 – 548, 2000.

Elbern, H., A. Strunk, H. Schmidt and O. Talagrand, Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, *7*, (14), 3749–3769, 2007.

EMEP, E., EEA air pollutant emission inventory guidebook 2013, *European Environment Agency, Copenhagen*, 2013.

Engelen, R. J. and A. P. McNally, Estimating atmospheric CO2 from advanced infrared satellite radiances within an operational four-dimensional variational (4D-Var) data assimilation system: Results and validation, *Journal of Geophysical Research: Atmospheres*, *110*, (D18), 2005.

Engelen, R. J., A. S. Denning and K. R. Gurney, On error estimation in atmospheric CO2 inversions, *Journal of Geophysical Research: Atmospheres*, *107*, (D22), 2002.

Enting, I. and J. Mansbridge, Seasonal sources and sinks of atmospheric CO2 direct inversion of filtered data, *Tellus B*, *41*, (2), 1989.

Errico, R. M., What is an adjoint model?, *Bulletin of the American Meteorological Society*, *78*, (11), 2577–2591, 1997.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, *99*, (C5), 10143–10162, 1994.

Evensen, G., *Data assimilation: the ensemble Kalman filter*, Springer Science & Business Media, 2009.

Fan, S., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi and P. Tans, A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models, *Science*, *282*, (5388), 442–446, 1998.

Farquhar, G., S. v. von Caemmerer and J. A. Berry, A biochemical model of photosynthetic co2 assimilation in leaves of c3 species, *Planta*, *149*, (1), 78–90, 1980.

Fisher, M. and D. J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Quarterly Journal of the Royal Meteorological Society*, *121*, (527), 1681–1704, 1995.

Fletcher, S., Mixed Gaussian-lognormal four-dimensional data assimilation, *Tellus A*, *62*, (3), 266–287, 2010.

Fletcher, S. J. and M. Zupanski, A hybrid multivariate normal and lognormal distribution for data assimilation, *Atmos. Sci. Lett.*, *7*, (2), 43–46, 2006.

Foken, T. and B. Wichura, Tools for quality assessment of surface-based flux measurements, *Agricultural and forest meteorology*, *78*, (1), 83–105, 1996.

Gelb, A., *Applied optimal estimation*, MIT press, 1974.

Gerbig, C., J. Lin, S. Wofsy, B. Daube, A. Andrews, B. Stephens, P. Bakwin and C. Grainger, Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *Journal of Geophysical Research: Atmospheres*, *108*, (D24), 2003*a*.

Gerbig, C., J. Lin, S. Wofsy, B. Daube, A. Andrews, B. Stephens, P. Bakwin and C. Grainger, Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *Journal of Geophysical Research: Atmospheres*, *108*, (D24), 2003*b*.

Goris, N. and H. Elbern, Singular vector based targeted observations of chemical constituents: description and first application of the EURAD-IM-SVA, *Geoscientific Model Development Discussions*, *8*, (8), 6267–6307, 2015.

Gou, T. and A. Sandu, Continuous versus Discrete Advection Adjoints in Chemical Data Assimilation with CMAQ, *Atmos. Env.*, *45*, 4868–4881, 2011.

Goulden, M. L., J. W. Munger, S.-M. Fan, B. C. Daube and S. C. Wofsy, Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability, *Science*, *271*, (5255), 1996.

Gourdji, S. M., North American CO2 exchange: Inter comparison of modeled estimates with results from a fine scale atmospheric inversion, *Biogeosciences*, 2013.

Gourdji, S., A. Hirsch, K. Mueller, V. Yadav, A. Andrews and A. Michalak, Regional-scale geostatistical inverse modeling of North American CO2 fluxes: a synthetic data study, *Atmospheric Chemistry and Physics*, *10*, (13), 6151–6167, 2010.

Graf, A., J. Werner, M. Langensiepen, A. van de Boer, M. Schmidt, M. Kupisch and H. Vereecken, Validation of a minimum microclimate disturbance chamber for net ecosystem flux measurements, *Agricultural and Forest Meteorology*, *174*, 1–14, 2013.

Gurney, K. R., R. M. Law, A. S. Denning, P. J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y.-H. Chen, P. Ciais, S. Fan et al., Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, *Nature*, *415*, (6872), 626–630, 2002.

Hamazaki, T., A. Kuze and K. Kondo, Sensor system for greenhouse gas observing satellite (GOSAT), in *Optical Science and Technology, the SPIE 49th Annual Meeting*, 275–282, International Society for Optics and Photonics, 2004.

Hascoet, L. and V. Pascual, The Tapenade Automatic Differentiation tool: principles, model, and specification, *ACM Transactions on Mathematical Software (TOMS)*, *39*, (3), 20, 2013.

Hollinger, D., J. Aber, B. Dail, E. Davidson, S. Goltz, H. Hughes, M. Leclerc, J. Lee, A. Richardson, C. Rodrigues et al., Spatial and temporal variability in forest-atmosphere CO2 exchange, *Global Change Biology*, *10*, (10), 1689–1706, 2004.

Houghton, R. A., Counting terrestrial sources and sinks of carbon, *Climatic Change*, *48*, (4), 525–534, 2001.

Houghton, R. A., Balancing the global carbon budget, *Annu. Rev. Earth Planet. Sci.*, *35*, 313–347, 2007.

Houghton, R. A., How well do we know the flux of CO2 from land-use change?, *Tellus B*, *62*, (5), 337–351, 2010.

Huntzinger, D. N., W. M. Post, Y. Wei, A. Michalak, T. O. West, A. Jacobson, I. Baker, J. M. Chen, K. Davis, D. Hayes et al., North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, *Ecological Modelling*, *232*, 144–157, 2012.

IPCC, Climate change 2014: Synthesis report, 2014.

Jung, M., M. Reichstein, H. A. Margolis, A. Cescatti, A. D. Richardson, M. A. Arain, A. Arneth, C. Bernhofer, D. Bonal, J. Chen et al., Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.: Biogeosciences (2005–2012)*, *116*, (G3), 2011.

Kalman, R. E., A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, *82*, (1), 35–45, 1960.

Kalman, R. E. and R. S. Bucy, New results in linear filtering and prediction theory, *Journal of Basic Engineering*, *83*, (1), 95–108, 1961.

Kaminski, T., P. J. Rayner, M. Heimann and I. G. Enting, On aggregation errors in atmospheric transport inversions, *Journal of Geophysical Research: Atmospheres*, *106*, (D5), 4703–4715, 2001.

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

Keeling, R. F., S. C. Piper, M. Heimann et al., Global and hemispheric CO2 sinks deduced from changes in atmospheric O2 concentration, *Nature*, *381*, (6579), 218–221, 1996.

Keeling, C. D., T. P. Whorf et al., Atmospheric CO2 records from sites in the SIO air sampling network, *Trends: a compendium of data on global change*, 16–26, 2005.

Keenan, T., I. Baker, A. Barr, P. Ciais, K. Davis, M. Dietze, D. Dragoni, C. M. Gough, R. Grant, D. Hollinger et al., Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange, *Global Change Biology*, *18*, (6), 1971–1987, 2012.

Klimpt, J., E. Friese and H. Elbern, Joint CO2 state and flux estimation with the 4D-Var system EURAD-IM, *Geoscientific Model Development Discussions*, *2016*, 1–24, 2016.

Klinker, E., F. Rabier, G. Kelly and J.-F. Mahfouf, The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration,

Quarterly Journal of the Royal Meteorological Society, *126*, (564), 1191–1215, 2000.

Knorr, W. and M. Heimann, Uncertainties in global terrestrial biosphere modeling: 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, *Global Biogeochemical Cycles*, *15*, (1), 207–225, 2001.

Kretschmer, R., C. Gerbig, U. Karstens, G. Biavati, A. Vermeulen, F. Vogel, S. Hammer and K. Totsche, Impact of optimized mixing heights on simulated regional atmospheric transport of CO2, *Atmos. Chem. Phys.*, *14*, (14), 7149–7172, 2014.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch and I. C. Prentice, A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles*, *19*, (1), 2005.

Kuenen, J., A. Visschedijk, M. Jozwicka and H. Denier van der Gon, TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, *Atmos. Chem. Phys.*, *14*, (20), 10963–10976, 2014.

Lahoz, W., B. Khattatov and R. Menard, *Data assimilation: making sense of observations*, Springer Science & Business Media, 2010.

Langensiepen, M., M. Kupisch, M. T. van Wijk and F. Ewert, Analyzing transient closed chamber effects on canopy gas exchange for optimizing flux calculation timing, *Agricultural and forest meteorology*, *164*, 61–70, 2012.

Langland, R. H. and N. L. Baker, Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system, *Tellus A*, *56*, (3), 189–201, 2004.

Laprise, R., The Euler equations of motion with hydrostatic pressure as an independent variable, *Monthly weather review*, *120*, (1), 197–207, 1992.

Lauvaux, T., M. Uliasz, C. Sarrat, F. Chevallier, P. Bousquet, C. Lac, K. J. Davis, P. Ciais, A. S. Denning and P. J. Rayner, Mesoscale inversion: first results from the CERES campaign with synthetic data, *Atmos. Chem. Phys.*, *8*, (13), 3459–3471, 2008.

Lauvaux, T., B. Gioli, C. Sarrat, P. Rayner, P. Ciais, F. Chevallier, J. Noilhan, F. Miglietta, Y. Brunet, E. Ceschia et al., Bridging the gap between atmospheric concentrations and local ecosystem measurements, *Geophysical Research Letters*, *36*, (19), 2009*a*.

Lauvaux, T., O. Pannekoucke, C. Sarrat, F. Chevallier, P. Ciais, J. Noilhan and P. Rayner, Structure of the transport uncertainty in mesoscale inversions of CO2 sources and sinks using ensemble model simulations, *Biogeosciences*, *6*, (6), 1089–1102, 2009*b*.

Law, R., P. Rayner, A. Denning, S. Taguchi and J. Taylor, Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO2 inversions, *Global Biogeochemical Cycles*, *10*, (4), 783–796, 1996.

Le Quéré, C., R. Moriarty, R. M. Andrew, J. G. Canadell, S. Sitch, J. I. Korsbakken, P. Friedlingstein, G. P. Peters, R. J. Andres, T. Boden et al., Global carbon budget 2015, *Earth System Science Data*, *7*, (2), 349–396, 2015.

Levin, I. and U. Karstens, Inferring high-resolution fossil fuel CO2 records at continental sites from combined 14 CO2 and CO observations, *Tellus B*, *59*, (2), 245–250, 2007.

Liu, D. C. and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical programming*, *45*, (1-3), 503–528, 1989.

Lloyd, J. and J. Taylor, On the temperature dependence of soil respiration, *Funct. Ecol.*, 315–323, 1994.

Longdoz, B., P. Gross and A. Granier, Multiple quality tests for analysing CO2 fluxes in a beech temperate forest, *Biogeosciences Discussions*, *4*, (6), 4197–4228, 2007.

Lorenc, A. C., Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, (474), 1177–1194, 1986.

Mahadevan, P., S. C. Wofsy, D. M. Matross, X. Xiao, A. L. Dunn, J. C. Lin, C. Gerbig, J. W. Munger, V. Y. Chow and E. W. Gottlieb, A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochemical Cycles*, *22*, (2), 2008.

Matross, D. M., A. Andrews, M. Pathmathevan, C. Gerbig, J. C. Lin, S. C. Wofsy, B. C. Daube, E. W. Gottlieb, V. Y. Chow, J. T. Lee et al., Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus B*, *58*, (5), 344–358, 2006.

Mauder, M., M. Cuntz, C. Drüe, A. Graf, C. Rebmann, H. P. Schmid, M. Schmidt and R. Steinbrecher, A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, *Agricultural and Forest Meteorology*, *169*, 122–135, 2013.

McKinley, D. C., M. G. Ryan, R. A. Birdsey, C. P. Giardina, M. E. Harmon, L. S. Heath, R. A. Houghton, R. B. Jackson, J. F. Morrison, B. C. Murray et al., A synthesis of current knowledge on forests and carbon storage in the United States, *Ecological Applications*, *21*, (6), 1902–1924, 2011.

Memmesheimer, M., H. Hass, J. Tippke and A. Ebel, Modeling of episodic emission data for Europe with the EURAD emission model EEM, *Regional Photochemical Measurement and Modeling Studies*, *2*, 495–499, 1995.

Ménard, R. and R. Daley, The application of Kalman smoother theory to the estimation of 4DVAR error statistics, *Tellus A*, *48*, (2), 221–237, 1996.

Niwa, Y., T. Machida, Y. Sawa, H. Matsueda, T. J. Schuck, C. A. Brenninkmeijer, R. Imasu and M. Satoh, Imposing strong constraints on tropical terrestrial CO2 fluxes using passenger aircraft based measurements, *Journal of Geophysical Research: Atmospheres*, *117*, (D11), 2012.

Nocedal, J., Updating quasi-Newton matrices with limited storage, *Math. Comput.*, *35*, (151), 773–782, 1980.

Oleson, K. W., D. M. Lawrence, B. Gordon, M. G. Flanner, E. Kluzek, J. Peter, S. Levis, S. C. Swenson, E. Thornton, J. Feddema et al., Technical description of version 4.0 of the Community Land Model (CLM), *NCAR Technical Note STR*, 2010.

Papale, D. and R. Valentini, A new assessment of european forests carbon exchanges by eddy fluxes and artificial neural network spatialization, *Global Change Biology*, *9*, (4), 525–535, 2003.

Peters, W., J. Miller, J. Whitaker, A. Denning, A. Hirsch, M. Krol, D. Zupanski, L. Bruhwiler and P. Tans, An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, *Journal of Geophysical Research: Atmospheres*, *110*, (D24), 2005.

Peters, W., M. Krol, G. Van Der Werf, S. Houweling, C. Jones, J. Hughes, K. Schaefer, K. Masarie, A. Jacobson, J. Miller et al., Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Global Change Biology*, *16*, (4), 1317–1337, 2010.

Peylin, P., P. Rayner, P. Bousquet, C. Carouge, F. Hourdin, P. Heinrich, P. Ciais et al., Daily CO2 flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology, *Atmos. Chem. Phys.*, *5*, (12), 3173–3186, 2005.

Peylin, P., C. Bacour, N. MacBean, S. Leonard, P. J. Rayner, S. Kuppel, E. N. Koffi, A. Kane, F. Maignan, F. Chevallier, P. Ciais and P. Prunet, A new step-wise Carbon Cycle Data Assimilation System using multiple data streams to constrain the simulated land surface carbon cycle, *Geoscientific Model Development Discussions*, *2016*, 1–52, 2016.

Post, H., H.-J. Hendricks Franssen, A. Graf, M. Schmidt and H. Vereecken, Uncertainty analysis of eddy covariance CO2 flux measurements for different EC tower distances using an extended two-tower approach, *Biogeosciences*, *12*, (4), 1205–1221, 2015.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney and S. A. Klooster, Terrestrial ecosystem production: a process model based on global satellite and surface data, *Global Biogeochemical Cycles*, *7*, (4), 811–841, 1993.

Pouliot, G., T. Pierce, H. D. van der Gon, M. Schaap, M. Moran and U. Nopmongcol, Comparing emission inventories and model-ready emission datasets between Europe and North America for the AQMEII project, *Atmos. Env.*, *53*, 4–14, 2012.

Rayner, P. J., The current state of carbon-cycle data assimilation, *Current Opinion in Environmental Sustainability*, *2*, (4), 289–296, 2010.

Rödenbeck, C., Estimating CO2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, 2005.

Rödenbeck, C., C. Gerbig, K. Trusilova and M. Heimann, A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, *Atmospheric Chemistry and Physics*, *9*, (14), 5331–5342, 2009.

Rodgers, C. D., *Inverse methods for atmospheric sounding: theory and practice*, vol 2, World scientific, 2000.

Ruimy, A., L. Kergoat, C. B. Field and B. Saugier, The use of CO2 flux measurements in models of the global terrestrial carbon budget, *Global Change Biology*, *2*, (3), 287–296, 1996.

Ryan, M. G. and B. E. Law, Interpreting, measuring, and modeling soil respiration, *Biogeochemistry*, *73*, (1), 3–27, 2005.

Sarmiento, J., M. Gloor, N. Gruber, C. Beaulieu, A. Jacobson, S. Mikaloff Fletcher, S. Pacala and K. Rodgers, Trends and regional distributions of land and ocean carbon sinks, *Biogeosciences*, *7*, (8), 2351–2367, 2010.

Sasaki, Y., Numerical variational analysis formulated under the constraints as determined by longwave equations and a low-pass filter, *Mon. Wea. Rev*, *98*, (12), 884–898, 1970*a*.

Sasaki, Y., Numerical variational analysis with weak constraint and application to surface analysis of severe storm gust, *Mon. Wea. Rev*, *98*, (12), 899–910, 1970*b*.

Schwalm, C. R., C. A. Williams, K. Schaefer, R. Anderson, M. A. Arain, I. Baker, A. Barr, T. A. Black, G. Chen, J. M. Chen et al., A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis, *Journal of Geophysical Research: Biogeosciences*, *115*, (G3), 2010.

Sharkey, T. D., Photosynthesis in intact leaves of C3 plants: physics, physiology and rate limitations, *The Botanical Review*, *51*, (1), 53–105, 1985.

Sitch, S., P. Friedlingstein, N. Gruber, S. Jones, G. Murray-Tortarolo, A. Ahlström, S. C. Doney, H. Graven, C. Heinze,

C. Huntingford et al., Trends and drivers of regional sources and sinks of carbon dioxide over the past two decades, *Biogeosciences Discuss.*, *10*, 20113–20177, 2013.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, A Description of the Advanced Research WRF Version 3 , *NCAR technical note, NCAR, Boulder CO*, 2008.

Stephens, B. B., K. R. Gurney, P. P. Tans, C. Sweeney, W. Peters, L. Bruhwiler, P. Ciais, M. Ramonet, P. Bousquet, T. Nakazawa et al., Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2, *Science*, *316*, (5832), 1732–1735, 2007.

Stoy, P. C., A. D. Richardson, D. D. Baldocchi, G. G. Katul, J. Stanovick, M. D. Mahecha, M. Reichstein, M. Detto, B. E. Law, G. Wohlfahrt et al., Biosphere-atmosphere exchange of CO2 in relation to climate: a cross-biome analysis across multiple time scales, *Biogeosciences*, *6*, (10), 2297–2312, 2009.

Talagrand, O., Variational Assimilation, in *Data Assimilation, Making sense of observations*, 41–66, Springer, 2010.

Talagrand, O. and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quarterly Journal of the Royal Meteorological Society*, *113*, (478), 1311–1328, 1987.

Tans, P. P., T. J. Conway and T. Nakazawa, Latitudinal distribution of the sources and sinks of atmospheric carbon dioxide derived from surface observations and an atmospheric transport model, *Journal of Geophysical Research: Atmospheres*, *94*, (D4), 5151–5172, 1989.

Tans, P. P., I. Y. Fung and T. Takahashi, Observational contrains on the global atmospheric CO2 budget, *Science*, *247*, (4949), 1431–1438, 1990.

Thacker, W. C. and R. B. Long, Fitting dynamics to data, *Journal of Geophysical Research: Oceans*, *93*, (C2), 1227–1240, 1988.

Thepaut, J.-N. and P. Courtier, Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model, *Quarterly Journal of the Royal Meteorological Society*, *117*, (502), 1225–1254, 1991.

Thompson, R. L., P. K. Patra, F. Chevallier, S. Maksyutov, R. M. Law, T. Ziehn, I. T. van der Laan-Luijkx, W. Peters, A. Ganshin, R. Zhuravlev et al., Top-down assessment of the Asian carbon budget since the mid 1990s, *Nature communications*, *7*, 2016.

Tolk, L. F., W. Peters, A. G. Meesters, M. Groenendijk, A. T. Vermeulen, G. J. Steeneveld and A. J. Dolman, Modelling regional scale surface fluxes, meteorology and CO2 mixing ratios for the Cabauw tower in the Netherlands, *bg*, *6*, (10), 2265–2280, 2009.

Tolk, L. F., A. J. Dolman, A. G. Meesters and W. Peters, A comparison of different inverse carbon flux estimation approaches for application on a regional domain, *Atmos. Chem. Phys.*, *11*, (20), 10349–10365, 2011.

Trémolet, Y., Model-error estimation in 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, *133*, (626), 1267–1280, 2007.

Tromp, J., C. Tape and Q. Liu, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, *160*, (1), 195–216, 2005.

Turner, D. P., M. Göckede, B. E. Law, W. D. Ritts, W. B. Cohen, Z. Yang, T. Hudiburg, R. Kennedy and M. Duane, Multiple constraint analysis of regional land–surface carbon flux, *Tellus B*, *63*, (2), 207–221, 2011.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky and J. W. Munger, Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest, *Journal of Geophysical Research: Biogeosciences*, *112*, (G2), 2007.

Van der Molen, M. K. and A. J. Dolman, Regional carbon fluxes and the effect of topography on the variability of atmospheric CO2, *Journal of Geophysical Research: Atmospheres*, *112*, (D1), 2007.

Van Leer, B., Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, *J. Comp. Phys*, *23*, (3), 276–299, 1977.

Vermeulen, A. T., A. Hensen, M. E. Popa, W. C. M. Van Den Bulk and P. A. C. Jongejan, Greenhouse gas observations from Cabauw tall tower (1992–2010), *Atmos. Meas. Tech*, *4*, (3), 617–644, 2011.

Vukicevic, T., Nonlinear and linear evolution of initial forecast errors, *Mon. Weather Rev*, *119*, (7), 1602–1611, 1991.

Walcek, C., Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection, *J. Geophys. Res.*, *105*, (D7), 9335–9348, 2000.

Wang, Y.-P., C. M. Trudinger and I. G. Enting, A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales, *Agricultural and Forest Meteorology*, *149*, (11), 1829–1842, 2009.

Wanninkhof, R., G.-H. Park, T. Takahashi, C. Sweeney, R. A. Feely, Y. Nojiri, N. Gruber, S. C. Doney, G. A. McKinley, A. Lenton et al., Global ocean carbon uptake: magnitude, variability and trends, *Biogeosciences*, *10*, (3), 1983–2000, 2013.

Weaver, A. and P. Courtier, Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. R. Meteorol. Soc.*, *127*, (575), 1815–1846, 2001.

Yanenko, N. N., *The method of fractional steps*, Springer, 1971.

Zhu, Q., Q. Zhuang, D. Henze, K. Bowman, M. Chen, L. Yaling, Y. He, H. Matsueda, T. Machida, Y. Sawa et al., Constraining terrestrial ecosystem CO2 fluxes by integrating models of biogeochemistry and atmospheric transport and data of surface carbon fluxes and atmospheric CO_2 concentrations, *Atmos. Chem. Phys. Discuss.*, *14*, 22587–22638, 2014.

Danksagung

Zuerst möchte ich mich bei PD Dr. Hendrik Elbern für seine Unterstützung in den letzten Jahren bedanken. Die vorliegende Arbeit wurde durch ihn überhaupt erst möglich gemacht und durch unzählige Diskussionen und Denkanstöße vorangetrieben. Ich bedanke mich außerdem für den Blick über den Tellerrand zur Herangehensweise an unterschiedlichste Probleme und für die vielen Verbesserungen und Anregungen zu dieser Arbeit.

Weiterhin möchte ich mich bei Prof. Dr. Susanne Crewell für die vielen Diskussionen in den letzten Jahren und die Korrektur dieser Arbeit bedanken.

Diese Arbeit wurde teilweise durch das TR32 Projekt (http://tr32new.uni-koeln.de/), Sonderforschungsbereich der DFG, finanziert. Über die finanzielle Unterstützung hinaus, wurden wertvolle Beobachtungsdaten von vier Messstationen zur Verfügung gestellt. Dafür, und für die geduldigen Erklärungen von Marius Schmidt und Alexander Graf zu den Daten, möchte ich hiermit meinen Dank aussprechen.

Ich bedanke mich für die bereitgestellte Rechenzeit der durchgeführten numerischen Experimente vom "John von Neumann Institute for Computing" (NIC), durchgeführt auf dem Supercomputer JURECA am "Jülich Supercomputing Center" (JSC).

Großer Dank gilt all denjenigen Personen und Institutionen, die diese Arbeit durch das Bereitstellen weiterer wertvoller Daten unterstützt haben und einen sinnvollen Vergleich der entstandenen Simulationen ermöglichten. Hugo Denier van der Gon stellte die aktuellen Daten des "TNO emission inventory" für die anthropogenen Emissionen zur Verfügung. Die CO_2-Messwerte des Jülicher Turms wurden von Marc von Hobe bereitgestellt. Bernard Longdoz danke ich für die Messungen von Hesse-Forest Sarrebourg in Frankreich und Arjan Hensen für die Beobachtungsdaten des Cabauw Turms in den Niederlanden. Patrizia Ney danke ich für die Messreihen von Wüstebach und ihren Erläuterungen zur Eddy-Kovarianzmethode. Markus Übel danke ich für das Bereitstellen seiner Implementation der Blattatmung in das CLM und für die Diskussionen zur CO_2 Modellierung.

Schließlich ist es mir ein besonderes Anliegen, mich bei meinen Kollegen am RIU zu bedanken. Ohne die Hilfe von Elmar Friese wäre ich vermutlich in den Tiefen des EURAD-Codes auf der Suche nach kleinen Käfern verloren gegangen. Weiterhin konnte ich mich immer auf ein offenes Ohr und einen guten Rat von Elmar bei komplizierten Problemen verlassen. Michael Memmersheimer danke ich für seine Erklärungen zur Meteorologie, zu Emissionen und zur Modellierung allgemein. Anne, Isabel, Keti, Luise, Voula, Zoe, Jonas, Kai und Philipp danke ich für die vielen Diskussionen zur Arbeit, durch die ich vieles erst richtig verstanden habe und für die anstrengende Arbeit des Korrekturlesens. Viel wichtiger als die Beiträge zur wissenschaftlichen Arbeit, möchte ich mich bei allen oben Genannten für die gemeinsam verbrachte Zeit bedanken, für die unzähligen kleinen Momente des Glücks im Büroalltag, für die interessanten Diskussionen aller Themen abseits der Arbeit.

Zu guter Letzt bedanke ich mich für die Unterstützung meiner Eltern und meiner Schwester. Danke an meinen Sohn Ole, der es schafft, mir in jeder Situation ein Lächeln zu entlocken. Meiner Frau danke ich für ihre Begleitung durch schwere und leichte Zeiten, ohne dich wäre das Selbe niemals das Gleiche.

Erklärung

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

Teilpublikationen

Klimpt, J., E. Friese and H.Elbern, Joint CO_2 state and flux estimation with the 4D-Var system EURAD-IM, in *Geoscientific Model Development Discussions, 2016*, 1-24, 2016.

Curriculum Vitae

Johannes Klimpt

Address

Dauner Straße 18
50937 Cologne
Germany
Phone: +49-176-23508428
Email: jklimpt@uni-koeln.de

Personal Details

Gender: Male
Date of birth: 9th of March, 1985
Place of birth: Eisenach, Germany
Present citizenship: German
Family status: married, one child

Education

since 06/2012	Ph.D. student at the Rhenish Institute for Environments at the University of Cologne, Project title: *CO_2 state and flux estimation by four dimensional variational assimilation*; Supervisor: PD Dr. Hendrik Elbern
2008–2012	Continuation of the academic studies at the University of Bonn; degree: Diplom Mathematician; focal point: stochastic analysis
2007–2008	Academic studies at the Universitá degli Studi di Trieste
2004–2007	Begin of studies at the University of Bonn; major fields of study: mathematics, minor field of study: economics
1995–2003	Secondary school at Ernst Abbe Gymnasium in Eisenach; A levels