

Assimilation of ground-based and airborne lidar data into the MM5 4D-Var system

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)

Fakultät Naturwissenschaften
Universität Hohenheim

Institut für Physik und Meteorologie (IPM)

vorgelegt von
Matthias Grzeschik

aus Elsterwerda, Deutschland

Dekan: Prof. Dr. Heinz Breer

Berichterstatter, 1. Prüfer: Prof. Dr. Volker Wulfmeyer

Mitberichterstatterin, 2. Prüferin: Prof. Dr. Evelyne Richard

3. Prüfer: Prof. Dr. Kurt Jetter

Eingereicht am: 12. Oktober 2009

Mündliche Prüfung am: 26. März 2010

Die vorliegende Arbeit wurde am 9. Februar 2010 von der Fakultät Naturwissenschaften der Universität Hohenheim als "Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften" angenommen.

Abstract

This work investigates the impact of assimilating water vapor Light Detection and Ranging (lidar) data into mesoscale Numerical Weather Prediction (NWP) models. Two cases from the field campaigns International H₂O Project 2002 (IHOP-2002) and International Lindenberg Campaign for Assessment of Humidity- and Cloud-Profiling Systems and its Impact on High-Resolution Modelling 2005 (LAUNCH-2005) are presented. In the first case, airborne water vapor Differential Absorption Lidar (DIAL) data are used for an assimilation for 24 May 2002, where convection occurred along an eastward moving dryline in western Texas and Oklahoma south of a triple point that formed in western Oklahoma. In the second case, a network of three ground based water vapor Raman lidars, operated behind a sharp frontal rain band with a northwesterly flow, are used.

The method employed, Four-Dimensional Variational Data Assimilation (4D-Var), is described in relation to other methods and the implementation is given in detail. The data assimilation results in a large modification of the initial fields. The assimilation into the pre-convective conditions changed not only the water vapor field but also the location of convergence lines, causing positive modification of Convective Initiation (CI). In the LAUNCH-2005 case a strong correction of the vertical structure and the absolute values of the initial water-vapor field of the order of 1 g/kg was found. This occurred mainly upstream of the lidar systems within an area that was comparable with the domain covered by the lidar systems. The correction of the water-vapor field was validated using independent Global Positioning System (GPS) sensors. Much better agreement with GPS zenith wet path delay was achieved with the initial water-vapor field after 4D-Var. Furthermore, the impact of the assimilation and its temporal evolution was investigated with introduced measures. The results demonstrate the high value of accurate vertically resolved mesoscale water vapor observations and advanced data assimilation systems for short-range weather forecasting.

Zusammenfassung

Diese Arbeit untersucht die Auswirkungen der Assimilation von Wasserdampf *Light Detection and Ranging* (Lidar)-Daten in mesoskalige numerische Wettervorhersagemodelle (NWP). Zwei Fallstudien aus den Feldkampagnen *International H₂O Project 2002* (IHOP_2002) und *International Lindenberg Campaign for Assessment of Humidity- and Cloud-Profiling Systems and its Impact on High-Resolution Modelling 2005* (LAUNCH-2005) werden vorgestellt. Im ersten Fall wurden flugzeuggestützte *Differential Absorption Lidar* (DIAL)-Wasserdampf-Daten für eine Assimilation am 24. Mai 2002 genutzt. An diesem Tag trat Konvektion entlang einer sich ostwärts bewegenden Feuchtefront im Westen Texas und in Oklahoma, südlich eines Tripelpunkt, auf. Im zweiten Fall wurde ein Netzwerk von drei bodengestützten Wasserdampf-Raman-Lidar-Systemen benutzt, welches nach Durchzug eines Regenbandes entlang einer Front und bei nordwestlichem Wind betrieben wurde.

Die verwendete Methode, *Four-Dimensional Variational Data Assimilation* (4D-Var), wird erläutert und mit anderen Datenassimilationsmethoden verglichen, wobei die hierfür notwendigen Implementationen detailliert beschrieben werden. Die Datenassimilation hat eine große Änderung der Anfangsbedingungen zur Folge. Die Assimilation in die prekonvektiven Bedingungen veränderte nicht nur die Verteilung des Wasserdampfes, sondern auch die Lage von Konvergenzlinien, was die Auslösung von Konvektion (CI) positiv modifizierte. Im LAUNCH-2005 Fall wurde eine starke Korrektur der vertikalen Struktur und der absoluten Menge des initialen Wasserdampfes in der Größenordnung von 1 g/kg festgestellt. Dies trifft hauptsächlich auf Gebiete im Lee der Lidar-Systeme zu. Die Korrektur des Wasserdampffeldes wurde mit unabhängigen *Global Positioning System* (GPS)-Messungen validiert. Durch die 4D-Var wurde eine bessere Übereinstimmung des Wasserdampf-Anfangsfeldes mit GPS *Zenit Wet Path Delay* (ZWD) erreicht. Außerdem wurde die Auswirkung der Assimilation und die weitere zeitlicher Veränderung mithilfe eingeführter Maße untersucht. Das Resultat zeigt den großen Wert akkurater und vertikal aufgelöster mesoskaliger Wasserdampfbeobachtungen sowie fortschrittlicher Datenassimilationsmethoden für die kurzfristige Wettervorhersage.

Contents

1. Acronyms	1
2. Introduction	5
3. Data assimilation schemes	7
3.1. Notation	8
3.1.1. Variable declaration	8
3.1.2. Statistical definitions	9
3.1.3. The model state vector	12
3.1.4. The observation operator	12
3.1.5. Time evolution	13
3.2. Optimal Interpolation (OI)	15
3.3. Three Dimensional Variational Analysis (3D-Var)	16
3.3.1. Equivalence to OI	18
3.3.2. Practical Implementation	19
3.4. Four Dimensional Variational Analysis (4D-Var)	20
3.4.1. Incremental formulation of Four-Dimensional Variational Data Assimilation (4D-Var)	21
3.5. Extended Kalman Filter (EKF)	22
3.5.1. Equivalence to the 4D-Var	23
3.6. Ensemble Kalman Filter (EnKF)	24
3.6.1. Stochastic Ensemble Kalman filter (EnKF)	25
3.6.2. Ensemble Square Root Filter (EnSRF)	25
3.6.3. Localization	26
3.7. Comparison of 4D-Var and EnKF	26
3.8. Optimal data assimilation scheme for lidar	27
4. MM5 4D-Var implementation	29
4.1. MM5 system	29
4.2. MM5 4D-Var system	32
4.2.1. The background error covariance matrix	32
4.2.2. The observation operator	33
4.2.3. The observation error covariance matrix	33
4.2.4. Calculation of the cost function value	33
4.2.5. Calculation of the gradient	34
4.3. Setup of the assimilation experiment	36
5. The IHOP_2002 experiment	39
5.1. The IHOP_2002 field experiment	39
5.2. Synoptic situation on 24 May 2002	40

5.3. LASE data	43
5.4. Observation operator	46
5.5. Minimization	49
5.6. The impact of 4D-Var on the initial fields	53
5.7. The impact on the forecast of CI	55
6. The LAUNCH-2005 experiment	63
6.1. LAUNCH-2005	63
6.2. Observational data	64
6.2.1. RAMSES	65
6.2.2. IFT multi wavelength lidar	65
6.2.3. Ziegendorf	65
6.2.4. Operational data analysis	66
6.2.5. System performance	66
6.2.6. Temporal evolution of water vapor profiles	69
6.3. Synoptic situation	69
6.4. Minimization	71
6.5. Impact on the fields	73
6.6. Quantitative description of the impact	78
6.7. Validation of the impact with GPS ZWD observations	80
7. Summary and conclusion	83
A. Process and verification studies using the D-PHASE ensemble	85
A.1. Implementation	85
A.2. Example plots from COPS IOP9c	86
Bibliography	93

1. Acronyms

3D-Var Three-Dimensional Variational Data Assimilation

4D-Var Four-Dimensional Variational Data Assimilation

4DVAR identifier for the simulation with assimilation

4D-EnKF Four-dimensional EnKF

ADJ Adjoint Model

AGL Above Ground Level

ARM Atmospheric Radiation Measurement

ASL Above Sea Level

BFGS Broyden-Fletcher-Goldfarb-Shanno algorithm

BLUE Best Linear Unbiased Estimate

CAP Working Group on Cloud and Aerosol Profiling

CAPE Convective Available Potential Energy

CART Cloud and Radiation Testbed

CARL Cloud and Radiation Testbed (CART) Raman lidar

CERA Climate and Environmental Retrieving and Archiving

CI Convective Initiation

CIN Convective Inhibition

CM-SAF European Climate Monitoring Satellite Application Facility

CMC Canadian Meteorological Centre

CMCGEM Canadian Meteorological Centre (CMC) Global Environmental Multiscale

CONTROL identifier for the simulation without assimilation

COPS Convective and Orographically-induced Precipitation Study

- COST** European Cooperation in the field of Scientific and Technical Research
- COSMO** Consortium for Small-scale Modeling
- CPU** Central processing unit
- D-PHASE** Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region
- DA** Data Assimilation
- DIAL** Differential Absorption Lidar
- DWD** German Meteorological Service
- ECMWF** European Centre for Medium-Range Weather Forecasts
- EISAR** Emission-Infrared Spectrometer for Atmospheric Research
- EnKF** Ensemble Kalman filter
- EnSRF** Ensemble Square Root Filter
- ESF** European Science Foundation
- EKF** Extended Kalman Filter
- FDP** Forecast Demonstration Project
- FTIR** Fourier Transform Infrared Spectroscopy
- GEWEX** Global Energy and Water Cycle Experiment
- GOES** Geostationary Operational Environmental Satellite
- GPS** Global Positioning System
- GRIB** GRIdded Binary
- GrADS** Grid Analysis and Display System
- GVaP** Global Water Vapor Project
- HIRLAM** High Resolution Limited Area Modelling
- IFT** Leibniz Institute for Tropospheric Research
- IHOP_2002** International H₂O Project 2002
- IOP** Intensive Observation Period
- IPM** Institut für Physik und Meteorologie

KF Kalman Filter

L-BFGS Limited Memory Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) method

LFM Limited Fine Mesh

LASE Lidar Atmospheric Sensing Experiment

LAUNCH-2005 International Lindenberg Campaign for Assessment of Humidity- and Cloud-Profiling Systems and its Impact on High-Resolution Modelling 2005

lidar Light Detection and Ranging

MAP Mesoscale Alpine Programme

Meteosat Meteorological Satellite

MCS Mesoscale Convective System

MM5 Fifth-generation Pennsylvania State University/National Center for Atmospheric Research Mesoscale Model

MOL Meteorological Richard-Aßmann-Observatory in Lindenberg

MRF Medium Range Forecast

MSG Meteorological Satellite (Meteosat) Second Generation

NASA National Aeronautics and Space Administration

NCEP National Center for Environmental Prediction

NEXRAD Next Generation Weather Radar

NWP Numerical Weather Prediction

OI Optimal Interpolation

OSE Observation System Experiment

OSSE Observation System Simulation Experiment

PDF Probability Density Function

PBL Planetary Boundary Layer

PNG Portable Network Graphics

POSIX Portable Operating System Interface

QPF Quantitative Precipitation Forecasting

RAFS Regional Analysis and Forecasting System
RAMSES Raman Lidar for Atmospheric Moisture Sensing
RDP Research and Development Project
RMS Root Mean Square
RUC Rapid Update Cycle
SGP Southern Great Plains
SNR Signal-to-Noise Ratio
SPC Storm Prediction Center
SSI Spectral Statistical Interpolation
STD Slant Total Delay
TAF Transformation of Algorithms in Fortran
TAMC Tangent linear and Adjoint Model Compiler
TLM Tangent Linear Model
US United States
WALES Water Vapour Lidar Experiment in Space
WDCC World Data Center for Climate
WMO World Meteorological Organization
WWRP World Weather Research Programme
ZTD Zenith Total Delay
ZWD Zenith Wet Delay

2. Introduction

Numerical Weather Prediction (NWP) is the main tool for forecasting future weather and giving society the possibility to prepare for natural disasters like storms and floods. Such extreme weather events cause heavy financial losses and will probably appear more frequently due to global warming.

The idea of NWP was summarized by Bjerknes (1911) as two parts:

1. The present state of the atmosphere must be characterized as accurately as possible.
2. The intrinsic laws, according to which the subsequent states develop out of the preceding ones, must be known.

The first experiment to solve this problem was performed by Richardson (1922). A grid having a horizontal resolution of 200 km and four layers, centered over Germany, was used to perform a numerical integration of the full primitive equations of motion. The first one-day weather forecast using a barotropic (one-layer) filtered model was made by Charney et al. (1950) and was possible thanks to one of the first electronic computers. In 1954, the first operational real time NWP was started in Sweden.

These early experiments used hand interpolations of available observations to the grid points. The first objective analysis was introduced by Panofsky (1949) and performed by a polynomial expansion; it was later refined by Bergthorsson and Döös (1955) and Cressman (1959). Since the available observations had a very non-uniform spatial distribution, it became obvious that additional information was needed. First, climatology was used (Gandin, 1963) but later a short-range forecast was also taken as a first guess and Data Assimilation (DA) was performed in a continuous analysis cycle. The first regional model (Limited Fine Mesh (LFM) model) was implemented by Howcroft (1971). The Regional Analysis and Forecasting System (RAFS) was developed as the main system for North America in 1982, including a more advanced model and a data assimilation system based on a regional Optimal Interpolation (OI) scheme.

Based on the idea of using stochastic forecasts instead of performing deterministic ones (Epstein, 1969; Leith, 1974), an operational ensemble forecasting system was implemented in December 1992 at both the National Center for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) (Tracton and Kalnay, 1993; Toth and Kalnay, 1993; Palmer et al., 1993; Molteni et al., 1996; Toth and Kalnay, 1997). Ensemble forecasts accomplished two main goals: first, they provided an ensemble average forecast which was more accurate since the forecast uncertainties tended to average out. Second, they provided an estimation of the forecasts' reliability. With this forecasting method, the upper limit of weather predictability, theoretically estimated by Lorenz (1965) as about two weeks, became attainable in practice.

The observations used for NWP can be divided into three classes:

1. In situ measurements, including surface measurements, observations by radiosondes and pilot balloons, and aircraft reports. The instruments used measure variables like temperature, humidity, and pressure.

2. Remote sampling of an area or a volume. The operational instruments are passive, receiving satellite data, or active, obtaining data from a wind profiler or radar. Also available, but not used operationally, are instruments like Light Detection and Ranging (lidar), which measure water vapor, temperature and winds.
3. Wind velocities calculated from Lagrangian trajectories using wind measurements from radiosondes, balloons or cloud elements tracked visually from geostationary satellites.

The number of class one observations available is strongly dependent on the time of day. Mainly, they are only available at synoptic times. They are also irregularly distributed in space, with a higher concentration over land and in rich countries, than over sea and in poorer countries.

The higher resolution of mesoscale NWP involves a greater under-determination of the initial state. Therefore a higher density of observation is necessary than is given by the synoptic observations used for DA on the global scale. This higher resolution of observation can be satisfied by a higher spatial resolution, as in observation networks, together with a higher temporal resolution, if the DA scheme can handle it.

Passive satellite observations are available continuously in time and space. But the information they provide about the atmosphere consists of values integrated through the whole atmosphere, like radiances in several spectral channels. Active instruments of class two, like radar and lidar, can provide spatially resolved observations with a high temporal resolution. While radar provides information such as reflectivity or radial wind velocity and is already exhaustively used, lidar provides information like temperature, humidity radial wind velocity, and concentrations of various gases such as ozone and CO_2 . Unfortunately lidar instruments are very expensive so they are not extensively used. They have mainly been used so far in field campaigns, which are testbeds to show the possible impact of these instruments on DA.

Studying the impact of lidar observations on DA is the aim of this work. At the beginning, there is a short overview of present DA schemes. The scheme used here, 4D-Var, and its implementation in the Fifth-generation Pennsylvania State University/National Center for Atmospheric Research Mesoscale Model (MM5) are described and two assimilation experiments are shown. The first one is an assimilation of water vapor Differential Absorption Lidar (DIAL) measurements from an airborne instrument during the International H₂O Project 2002 (IHOP_2002) field campaign and the second one is an experiment assimilating water vapor data from three ground based lidar systems during the International Lindenberg Campaign for Assessment of Humidity- and Cloud-Profiling Systems and its Impact on High-Resolution Modelling 2005 (LAUNCH-2005) campaign. Then, a plotting system is presented which was used during the Convective and Orographically-induced Precipitation Study (COPS) field campaign and the Forecast Demonstration Project (FDP) Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region (D-PHASE).

3. Data assimilation schemes

A numerical model like NWP and climate models are an approximation of the true state of the atmosphere with uncertain initial and boundary conditions. In his pioneering study, Lorenz (1963) showed the large sensitivity of NWP to initial conditions.

During the determination of the initial state, two major problems occur. First, finding the full initial state for a current model is an under determined problem, since there are not enough observations available. For example, a model grid that covers Central Europe (approx. 2500 km by 2500 km) with a horizontal resolution of 10 km and 40 levels consists of 2.5 million grid boxes. To provide a proper initial state and fill all these grid boxes with initial values of temperature, pressure, water vapor and horizontal wind, 12.5 million scalar values would have to be determined. The number of available observations is usually about two to three orders of magnitude less than this.

To solve this problem, not only observations are used for the initialization. In addition, a climatology or a previous forecast, usually called background, is used to fill gaps and derive the best estimate from observations and the background.

Second, observations are unevenly distributed in space and time and their density varies from region to region. The coverage over the continents is much higher than over the ocean. Even from country to country the density differs, depending on the financial means available.

The gaps in the observational network are filled using an analysis cycle. In this, a model is used to distribute information from areas with a high density of observations to regions with a lack of observations and another estimation is made, in a constant time interval, of the true state with an analysis and the new observations.

According to Talagrand (1997) “Assimilation of meteorological (or oceanographic) observations can be described as the process through which all the available information is used in order to estimate as accurately as possible the state of the atmospheric (or oceanic) flow. The available information essentially consists of the observations proper, and of the physical laws that govern the evolution of the flow. The latter are available in practice under the form of a numerical model.”

The computation of the probability density function of the model solution conditioned on the measurements defines the data assimilation problem considered in atmospheric modeling, namely of the atmosphere and ocean, and is given e.g. by Kalnay (2003) and Evensen (2006).

DA schemes have been successfully applied and extensively studied for global NWP. The applicability of the schemes in mesoscale NWP is currently disputed, since the high resolution induces problems, which do not occur on the global scale. The non-linearity of the model increases with the resolution, while DA schemes often assume linearity of the model. It also introduces processes like convection and discontinuities into the model, which are difficult to handle in DA.

3.1. Notation

3.1.1. Variable declaration

In the following sections the following variables and parameters are used.

τ : Time. If something is valid for a specific time, this time is marked as a superscript.

\vec{z} : State of the atmosphere

\vec{x} : Model state

\vec{x}_t : The model representation of the true atmosphere

\vec{x}_b : The background model state, obtained from a previous model forecast

\vec{x}_a : The analysis model state, obtained from a DA scheme

\vec{y}_o : An observation

L : The number of different observation times

\mathbb{Z} : The space of all possible states of the atmosphere

\mathbb{M} : The space of all possible states of the model

\mathbb{O} : The observations space

N : Dimension of the model space

K : Dimension of the observation space

$\mathcal{P}_{\mathbb{M}}$: Projection from \mathbb{Z} to \mathbb{M}

$\mathcal{P}_{\mathbb{O}}$: Projection from \mathbb{Z} to \mathbb{O}

$\varrho[\vec{x}]$: Marginal Probability Density Function (PDF)

$\varrho[\vec{x}|\vec{y}]$: Conditional PDF

$\langle \bullet \rangle$: Expected value

$\langle \bullet \rangle_{\mathbf{e}}$: Approximation of the expected value by an ensemble

$\vec{\varepsilon}$: The actual error of a state

\vec{b} : The bias of a state

σ_i : The variance of component i of a state

ρ_{ij} : The covariance between components i and j of a state

P: Error covariance matrix

P_e: Approximation of the error covariance matrix by an ensemble

B: Background error covariance matrix

R: Observation error covariance matrix

A: Analysis error covariance matrix

\vec{H} : Observation operator, or the projection from \mathbb{M} to \mathbb{O} .

$\mathbf{H}_{\vec{x}}$: Tangent Linear Model (TLM) of the observation operator

$\vec{\mathcal{M}}_{\tau \rightarrow \tau'}^Z$: Time propagator for the atmosphere from time τ to τ'

$\vec{\mathcal{M}}_{\tau \rightarrow \tau'}$: Representation of the time propagator in model space or the model

$\mathbf{M}_{\tau \rightarrow \tau', \vec{x}}$: The TLM

$\mathbf{Q}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'}$: The model error covariance

W: The optimal gain matrix

$\mathcal{J}[\vec{x}]$: The cost or penalty function

K: The Kalman gain

X: The ensemble matrix, describing all the deviations of ensemble members

3.1.2. Statistical definitions

Marginal Probability Density Function (PDF)

The PDF or marginal PDF ϱ gives the probability density that a random state \vec{x} is taken by a system. The PDF must satisfy the conditions

$$\begin{aligned} \varrho[\vec{x}] &\geq 0 \quad \forall \vec{x} \\ \int_{\forall \vec{x}} \varrho[\vec{x}] \, dV &= 1. \end{aligned} \tag{3.1.1}$$

Conditional Probability Density Function (PDF)

The conditional PDF is the probability density of some state given the occurrence of some other state. For example, we could define the conditional probability density $\varrho[\vec{x}|\vec{x}_t]$ of a state \vec{x} , forecasted by a model, when the system described by the model takes the state \vec{x}_t .

Bayes' Theorem

Bayes' Theorem associates a conditional PDF with the conditional PDF where the states are interchanged and reads:

$$\varrho[\vec{x}|\vec{x}_t] \varrho_A[\vec{x}_t] = \varrho[\vec{x}_t|\vec{x}] \varrho_M[\vec{x}], \quad (3.1.2)$$

where $\varrho[\vec{x}_t|\vec{x}]$ can be read as the probability density that the atmosphere takes the state \vec{x}_t , if the state \vec{x} is forecasted by the model and the marginal PDFs are defined as ϱ_A for the atmosphere and ϱ_M for the model.

In the case of a perfect model, model and atmosphere describe the same system and the marginal probability densities are identical, e.g.

$$\varrho_M[\vec{x}] = \varrho_A[\vec{x}]. \quad (3.1.3)$$

which yields, with this assumption on the identity of the conditional PDFs

$$\varrho[\vec{x}_t|\vec{x}] = \varrho[\vec{x}|\vec{x}_t] \frac{\varrho[\vec{x}_t]}{\varrho[\vec{x}]} \quad (3.1.4)$$

and as $\varrho[\vec{x}]$ is only a constant value for an actual measurement, we have

$$\varrho[\vec{x}_t|\vec{x}] \propto \varrho[\vec{x}|\vec{x}_t] \varrho[\vec{x}_t]. \quad (3.1.5)$$

Practically, equation (3.1.5) can express the probability of the atmospherical state \vec{x}_t , if \vec{x} was measured.

The basic idea of DA is, to find the state \vec{x}_t which maximizes the probability $\varrho[\vec{x}_t|\vec{x}]$ under the condition that \vec{x} was measured. This state \vec{x}_t is named the analysis and is labeled \vec{x}_a .

Expected Value

The expected value of a quantity q on condition that \vec{x}_t is taken by the atmosphere is defined as

$$\langle q \rangle = \int_{\forall \vec{x}} q \varrho[\vec{x}|\vec{x}_t] dV. \quad (3.1.6)$$

Bias

Comparing a model state \vec{x} with the true state of the atmosphere \vec{x}_t , the error of the model state is defined as

$$\vec{\varepsilon} = \vec{x} - \vec{x}_t \quad (3.1.7)$$

and the bias of the model as the expected value of this error

$$\vec{b} = \langle \vec{\varepsilon} \rangle = \langle \vec{x} \rangle - \vec{x}_t. \quad (3.1.8)$$

Error Covariance Matrix

The error covariance of a model is defined as

$$\mathbf{P} = \left\langle (\vec{x} - \langle \vec{x} \rangle) (\vec{x} - \langle \vec{x} \rangle)^T \right\rangle$$

and in the case of no bias ($\vec{b} = 0$)

$$\begin{aligned} &= \langle (\vec{x} - \vec{x}_t) (\vec{x} - \vec{x}_t)^T \rangle \\ &= \langle \vec{\varepsilon} \vec{\varepsilon}^T \rangle. \end{aligned} \quad (3.1.9)$$

The diagonal elements of this matrix represent the variances of each variable

$$\sigma_i^2 = \langle (x_i - \langle x_i \rangle)^2 \rangle$$

and in the case of no bias

$$= \langle \varepsilon_i^2 \rangle \quad (3.1.10)$$

The off-diagonal elements contain the covariances

$$\rho_{ij} = \frac{\langle (x_i - \langle x_i \rangle) (x_j - \langle x_j \rangle) \rangle}{\sigma_i \sigma_j}$$

and in the case of no bias

$$= \frac{\langle \varepsilon_i \varepsilon_j \rangle}{\sigma_i \sigma_j} \quad (3.1.11)$$

The error covariance matrix \mathbf{P} can be written as

$$\mathbf{P} = \begin{pmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 & \cdots & \rho_{1N} \sigma_1 \sigma_N \\ \rho_{21} \sigma_2 \sigma_1 & \sigma_2^2 & \cdots & \rho_{2N} \sigma_2 \sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} \sigma_N \sigma_1 & \rho_{N2} \sigma_N \sigma_2 & \cdots & \sigma_N^2 \end{pmatrix}. \quad (3.1.12)$$

Normal Distribution

The most common distribution is the Gaussian or normal distribution, which is fully defined by its mean $\langle \vec{x} \rangle$ and the error covariance matrix \mathbf{P} and has a bell shaped form. The normal distribution has the form

$$\begin{aligned} \varrho[\vec{x}|\vec{x}_t] &= \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \mathbf{P}}} \exp\left(-\frac{1}{2} (\vec{x} - \langle \vec{x} \rangle)^T \mathbf{P}^{-1} (\vec{x} - \langle \vec{x} \rangle)\right) \\ &= \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \mathbf{P}}} \exp\left(-\frac{1}{2} (\vec{x} - \vec{x}_t - \vec{b})^T \mathbf{P}^{-1} (\vec{x} - \vec{x}_t - \vec{b})\right) \end{aligned} \quad (3.1.13)$$

where N is the dimension of the state \vec{x} .

3.1.3. The model state vector

In a model, a set of variables describes the state of a system in a finite way. Therefore the state of a system $\vec{z} \in \mathbb{Z} = \mathbb{R}^\infty$ is brought to a grid with a finite number of grid points and a set of variables. A meteorological model has a set of variables $x_{vij k}$, where the index v denotes the state variables, i and j denote the horizontal grid box numbers, and k identifies the vertical grid box number. E.g. the variables temperature and pressure can be expressed as

$$\begin{aligned} T_{ijk} &= x_{0ijk} \\ p_{ijk} &= x_{1ijk} \end{aligned} \quad (3.1.14)$$

The total number of variables is N_v and the number of grid boxes in each direction is N_i , N_j , and N_k . Here the indices start with zero, i.e.

$$\begin{aligned} v &\in [0, N_v - 1] \\ i &\in [0, N_i - 1] \\ j &\in [0, N_j - 1] \\ k &\in [0, N_k - 1] \end{aligned} \quad (3.1.15)$$

To avoid matrices, the variables are mapped to vectors, e.g.

$$\vec{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix} \quad (3.1.16)$$

with

$$n \in [0, N - 1] \quad \text{and} \quad N = N_v N_i N_j N_k. \quad (3.1.17)$$

The resulting $\vec{x} \in \mathbb{M} = \mathbb{R}^N$ is called a model state and \mathbb{M} is called the model space. Furthermore, the method of how the system state $\vec{z} \in \mathbb{Z}$ is brought to model space, is denoted as $\mathcal{P}_{\mathbb{M}}$ with

$$\vec{x} = \mathcal{P}_{\mathbb{M}} [\vec{z}]. \quad (3.1.18)$$

For example, a model grid covering Central Europe ($\approx 2500 \text{ km} \times 2500 \text{ km}$) with a horizontal resolution of 10 km and 40 levels would have 2.5×10^6 grid boxes. Model variables, such as temperature, pressure, water vapor and horizontal wind, are associated with each grid box. This would lead to a dimension of the model space of $N \approx 12.5 \times 10^6$.

3.1.4. The observation operator

An observation is similar to equation (3.1.18) a projection from the state of a system $\vec{z} \in \mathbb{Z}$ on to a set of variables $\vec{y} \in \mathbb{O} = \mathbb{R}^M$ where \mathbb{O} is called the observation space

$$\vec{y} = \mathcal{P}_{\mathbb{O}} [\vec{z}]. \quad (3.1.19)$$

Such an observation is subjected to an uncertainty and therefore can be described by a PDF, i.e. if the observation is normally distributed, it is characterized by its bias and error covariance matrix.

The location, time, and often the variable of the observation do not agree with the model state variable. Therefore an observation operator, the link between model and observation, is defined by

$$\mathcal{P}_\circ [\vec{z}] = \vec{\mathcal{H}} [\mathcal{P}_\mathcal{M} [\vec{z}]]. \quad (3.1.20)$$

The observation operator is said to be linear if the following relation is valid.

$$\vec{\mathcal{H}} [\alpha \vec{x} + \vec{\varepsilon}] = \alpha \vec{\mathcal{H}} [\vec{x}] + \vec{\mathcal{H}} [\vec{\varepsilon}] \quad (3.1.21)$$

In the common DA methods, linearity of $\vec{\mathcal{H}}$ is assumed if the Taylor expansion

$$\vec{\mathcal{H}} [\vec{x} + \vec{\varepsilon}] = \vec{\mathcal{H}} [\vec{x}] + \mathbf{H}_{\vec{x}} \vec{\varepsilon} + \mathcal{O}(\|\vec{\varepsilon}\|^2) \quad (3.1.22)$$

can be cut off after the linear term, i.e. $\mathcal{O}(\|\vec{\varepsilon}\|^2) = 0$, which is approximately valid for small $\vec{\varepsilon}$. The $M \times N$ matrix $\mathbf{H}_{\vec{x}}$ is the TLM of the observation operator and is defined as the Jacobi matrix of $\vec{\mathcal{H}}$

$$\mathbf{H}_{\vec{x}} = \left(\begin{array}{cccc} \partial_{x_0} \mathcal{H}_0 & \partial_{x_1} \mathcal{H}_0 & \cdots & \partial_{x_{N-1}} \mathcal{H}_0 \\ \partial_{x_0} \mathcal{H}_1 & \partial_{x_1} \mathcal{H}_1 & \cdots & \partial_{x_{N-1}} \mathcal{H}_1 \\ \vdots & \vdots & & \vdots \\ \partial_{x_0} \mathcal{H}_{M-1} & \partial_{x_1} \mathcal{H}_{M-1} & \cdots & \partial_{x_{N-1}} \mathcal{H}_{M-1} \end{array} \right) \Bigg|_{\vec{x}} \quad (3.1.23)$$

The adjoint of the observation operator $\mathbf{H}_{\vec{x}}^T$ is the transpose of the linear tangent and an $N \times M$ matrix.

$$\mathbf{H}_{\vec{x}}^T = \left(\begin{array}{cccc} \partial_{x_0} \mathcal{H}_0 & \partial_{x_0} \mathcal{H}_1 & \cdots & \partial_{x_0} \mathcal{H}_{M-1} \\ \partial_{x_1} \mathcal{H}_0 & \partial_{x_1} \mathcal{H}_1 & \cdots & \partial_{x_1} \mathcal{H}_{M-1} \\ \vdots & \vdots & & \vdots \\ \partial_{x_{N-1}} \mathcal{H}_0 & \partial_{x_{N-1}} \mathcal{H}_1 & \cdots & \partial_{x_{N-1}} \mathcal{H}_{M-1} \end{array} \right) \Bigg|_{\vec{x}} \quad (3.1.24)$$

3.1.5. Time evolution

The time evolution of a system can be defined as a time propagator which transforms the state of the system at time τ to the state of the system a time τ' . This time propagator is denoted $\vec{\mathcal{M}}_{\tau \rightarrow \tau'}^Z$ for the system and $\vec{\mathcal{M}}_{\tau \rightarrow \tau'}$ for the model description of the same system and is defined by

$$\begin{aligned} \vec{z}^{\tau'} &= \vec{\mathcal{M}}_{\tau \rightarrow \tau'}^Z [\vec{z}^\tau] \\ \vec{x}^{\tau'} &= \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}^\tau] \end{aligned} \quad (3.1.25)$$

and, as a time propagator, $\vec{\mathcal{M}}$ has the properties

$$\vec{\mathcal{M}}_{\tau_1 \rightarrow \tau_3} [\vec{x}^{\tau_1}] = \vec{\mathcal{M}}_{\tau_2 \rightarrow \tau_3} \left[\vec{\mathcal{M}}_{\tau_1 \rightarrow \tau_2} [\vec{x}^{\tau_1}] \right] \quad \text{with} \quad \tau_1 \leq \tau_2 \leq \tau_3 \quad (3.1.26)$$

$$\vec{\mathcal{M}}_{\tau \rightarrow \tau} [\vec{x}^\tau] = \vec{x}^\tau. \quad (3.1.27)$$

The model $\vec{\mathcal{M}}$ is assumed to be linear if the Taylor expansion

$$\vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x} + \vec{\varepsilon}] = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}] + \mathbf{M}_{\tau \rightarrow \tau', \vec{x}} \vec{\varepsilon} + \mathcal{O}(\|\vec{\varepsilon}\|^2) \quad (3.1.28)$$

can be cut off after the linear term, i.e. $\mathcal{O}(\|\vec{\varepsilon}\|^2) = 0$. The $N \times N$ matrix $\mathbf{M}_{\tau \rightarrow \tau', \vec{x}} = \left(\nabla \vec{\mathcal{M}}_{\tau \rightarrow \tau'}^T \Big|_{\vec{x}} \right)^T$ is called the TLM and the transposed $\mathbf{M}_{\tau \rightarrow \tau', \vec{x}}^T$ is the Adjoint Model (ADJ).

The model error, which is produced from τ to τ' is defined by the comparison of $\vec{\mathcal{M}}$ with $\vec{\mathcal{M}}^{\mathbb{Z}}$:

$$\vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'} = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\mathcal{P}_{\mathbb{M}} [\vec{z}^\tau]] - \mathcal{P}_{\mathbb{M}} \left[\vec{\mathcal{M}}_{\tau \rightarrow \tau'}^{\mathbb{Z}} [\vec{z}^\tau] \right] \quad (3.1.29)$$

A true state of the atmosphere is propagated according to equation (3.1.25) with

$$\vec{x}_t^{\tau'} = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau] - \vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'} \quad (3.1.30)$$

and this true state can also be expressed by

$$\vec{x}_t^{\tau'} = \vec{x}^{\tau'} - \vec{\varepsilon}^{\tau'} = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}^\tau] - \vec{\varepsilon}^{\tau'}. \quad (3.1.31)$$

Combining equation (3.1.30) and 3.1.31 yields the error of the background state \vec{x}^τ at time τ' of

$$\vec{\varepsilon}^{\tau'} = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}^\tau] - \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau] + \vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'}$$

and with the error at time τ

$$\vec{\varepsilon}^{\tau'} = \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau + \vec{\varepsilon}^\tau] - \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau] + \vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'}$$

and with a linear model

$$\begin{aligned} \vec{\varepsilon}^{\tau'} &= \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau] + \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^\tau} \vec{\varepsilon}^\tau - \vec{\mathcal{M}}_{\tau \rightarrow \tau'} [\vec{x}_t^\tau] + \vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'} \\ \vec{\varepsilon}^{\tau'} &= \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^\tau} \vec{\varepsilon}^\tau + \vec{\varepsilon}_{\vec{\mathcal{M}}, \tau \rightarrow \tau'} \end{aligned} \quad (3.1.32)$$

Therefore, the covariance matrix at time τ' is

$$\begin{aligned}
 \mathbf{P}^{\tau'} &= \left\langle \vec{\varepsilon}^{\tau'} \vec{\varepsilon}^{\tau'T} \right\rangle \\
 &= \left\langle \left(\mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \vec{\varepsilon}^{\tau} + \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \right) \left(\mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \vec{\varepsilon}^{\tau} + \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \right)^T \right\rangle \\
 &= \left\langle \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \vec{\varepsilon}^{\tau} \vec{\varepsilon}^{\tau T} \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}}^T + \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \vec{\varepsilon}^{\tau} \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'}^T \right. \\
 &\quad \left. + \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \vec{\varepsilon}^{\tau T} \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}}^T + \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'}^T \right\rangle \\
 &= \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \left\langle \vec{\varepsilon}^{\tau} \vec{\varepsilon}^{\tau T} \right\rangle \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}}^T + \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \left\langle \vec{\varepsilon}^{\tau} \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'}^T \right\rangle \\
 &\quad + \left\langle \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \vec{\varepsilon}^{\tau T} \right\rangle \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}}^T + \left\langle \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'} \vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'}^T \right\rangle.
 \end{aligned}$$

The mixed expected values would be zero if the model error $\vec{\varepsilon}_{\mathcal{M}, \tau \rightarrow \tau'}$ and the state error $\vec{\varepsilon}^{\tau}$ were uncorrelated. This assumption is hard to fulfill since the model state is created by model forecasts. Nevertheless, these terms are neglected for simplicity (Bouttier and Courtier, 1999). The remaining expected values are the error covariance matrix \mathbf{P}^{τ} of time τ and the model error covariance $\mathbf{Q}_{\mathcal{M}, \tau \rightarrow \tau'}$ from time τ to τ' .

$$\mathbf{P}^{\tau'} = \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}} \mathbf{P}^{\tau} \mathbf{M}_{\tau \rightarrow \tau', \vec{x}_t^{\tau}}^T + \mathbf{Q}_{\mathcal{M}, \tau \rightarrow \tau'} \quad (3.1.33)$$

3.2. Optimal Interpolation (OI)

The OI¹ searches for an analysis state \vec{x}_a . It is assumed that the analysis can be written as the analysis equation

$$\vec{x}_a = \vec{x}_b + \mathbf{W} \left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_b] \right) \quad (3.2.1)$$

with the unknown $N \times M$ matrix \mathbf{W} .

The analysis error is

$$\begin{aligned}
 \vec{\varepsilon}_a &= \vec{x}_a - \vec{x}_t \\
 &= \vec{x}_b + \mathbf{W} \left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_b] \right) - \vec{x}_t \\
 &= \vec{x}_b - \vec{x}_t + \mathbf{W} \left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t + \vec{x}_b - \vec{x}_t] \right)
 \end{aligned}$$

and if the observation operator is linear

$$\begin{aligned}
 &= \vec{\varepsilon}_b + \mathbf{W} \left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t] + \mathbf{H}(\vec{x}_b - \vec{x}_t) \right) \\
 &= \vec{\varepsilon}_b + \mathbf{W} (\vec{y}_o - \vec{y}_t + \mathbf{H}\vec{\varepsilon}_b) \\
 &= \vec{\varepsilon}_b + \mathbf{W} (\vec{\varepsilon}_o + \mathbf{H}\vec{\varepsilon}_b) \quad (3.2.2)
 \end{aligned}$$

¹In mathematical terms the OI is not an interpolation rather it is a fit.

The optimal gain matrix \mathbf{W} is determined in OI by minimizing the sum of variances, i.e.

$$0 \stackrel{!}{=} \frac{\partial}{\partial w_{ij}} \left\langle \vec{\varepsilon}_a^T \vec{\varepsilon}_a \right\rangle \quad \forall i, j \quad (3.2.3)$$

with w_{ij} , the entries of the matrix \mathbf{W} . Assuming that background and observation are uncorrelated, i.e. $\langle \vec{\varepsilon}_b \vec{\varepsilon}_o^T \rangle = 0$ equation (3.2.3) leads to

$$0 \stackrel{!}{=} \mathbf{W}\mathbf{R} + \mathbf{W}\mathbf{H}\mathbf{B}\mathbf{H}^T - \mathbf{B}\mathbf{H}^T \quad (3.2.4)$$

where \mathbf{B} and \mathbf{R} are the error covariance matrices of the background and the observation. I.e. in the case of minimal variances of the analysis \mathbf{W} calculates to

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \right)^{-1}. \quad (3.2.5)$$

The analysis error covariance matrix is

$$\begin{aligned} \mathbf{A} &= \left\langle \vec{\varepsilon}_a \vec{\varepsilon}_a^T \right\rangle \\ &= \left\langle (\vec{\varepsilon}_b + \mathbf{W}(\vec{\varepsilon}_o - \mathbf{H}\vec{\varepsilon}_b)) (\vec{\varepsilon}_b + \mathbf{W}(\vec{\varepsilon}_o - \mathbf{H}\vec{\varepsilon}_b))^T \right\rangle \\ &= \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_b^T \right\rangle + \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_o^T \right\rangle \mathbf{W}^T - \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_b^T \right\rangle \mathbf{H}^T \mathbf{W}^T \\ &\quad + \mathbf{W} \left\langle \vec{\varepsilon}_o \vec{\varepsilon}_b^T \right\rangle + \mathbf{W} \left\langle \vec{\varepsilon}_o \vec{\varepsilon}_o^T \right\rangle \mathbf{W}^T - \mathbf{W} \left\langle \vec{\varepsilon}_o \vec{\varepsilon}_b^T \right\rangle \mathbf{H}^T \mathbf{W}^T \\ &\quad - \mathbf{W}\mathbf{H} \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_b^T \right\rangle - \mathbf{W}\mathbf{H} \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_o^T \right\rangle \mathbf{W}^T + \mathbf{W}\mathbf{H} \left\langle \vec{\varepsilon}_b \vec{\varepsilon}_b^T \right\rangle \mathbf{H}^T \mathbf{W}^T \end{aligned}$$

In the case of $\langle \vec{\varepsilon}_b \vec{\varepsilon}_o^T \rangle = 0$ and $\mathbb{1}$ the unity matrix

$$\begin{aligned} \mathbf{A} &= (\mathbb{1} - \mathbf{W}\mathbf{H})\mathbf{B} - \mathbf{B}\mathbf{H}^T \mathbf{W}^T + \mathbf{W}\mathbf{R}\mathbf{W}^T + \mathbf{W}\mathbf{H}\mathbf{B}\mathbf{H}^T \mathbf{W}^T \\ &= (\mathbb{1} - \mathbf{W}\mathbf{H})\mathbf{B} - \mathbf{B}\mathbf{H}^T \mathbf{W}^T + \mathbf{W} \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \right) \mathbf{W}^T \end{aligned}$$

and with equation (3.2.5)

$$\begin{aligned} \mathbf{A} &= (\mathbb{1} - \mathbf{W}\mathbf{H})\mathbf{B} - \mathbf{B}\mathbf{H}^T \mathbf{W}^T + \mathbf{B}\mathbf{H}^T \mathbf{W}^T \\ &= (\mathbb{1} - \mathbf{W}\mathbf{H})\mathbf{B}. \end{aligned} \quad (3.2.6)$$

3.3. Three Dimensional Variational Analysis (3D-Var)

3D-Var also combines a background state \vec{x}_b with an observation \vec{y}_o to obtain an analysis state \vec{x}_a which is the optimal estimation of the true state \vec{x}_t . It assumes unbiased model simulation and observation with a Gaussian distribution of errors, the background error covariance matrix \mathbf{B} and the observation error covariance matrix \mathbf{R} .

From equation (3.1.5), it follows that the probability of the true state, when \vec{y}_o is measured,

is

$$\varrho[\vec{x}_t|\vec{y}_o] = \varrho\left[\vec{\mathcal{H}}[\vec{x}_t]|\vec{y}_o\right] \propto \varrho\left[\vec{\mathcal{H}}[\vec{x}_t]|\vec{y}_o\right] \varrho\left[\vec{\mathcal{H}}[\vec{x}_t]\right] \quad (3.3.1)$$

and with $\varrho\left[\vec{\mathcal{H}}[\vec{x}_t]\right] = \varrho[\vec{x}_t]$

$$\varrho[\vec{x}_t|\vec{y}_o] \propto \varrho\left[\vec{\mathcal{H}}[\vec{x}_t]|\vec{y}_o\right] \varrho[\vec{x}_t]. \quad (3.3.2)$$

Since the observation is assumed to be normally distributed, the conditional PDF is

$$\varrho\left[\vec{\mathcal{H}}[\vec{x}_t]|\vec{y}_o\right] = \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det \mathbf{R}}} \exp\left(-\frac{1}{2}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t]\right)^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t]\right)\right) \quad (3.3.3)$$

and the marginal PDF is assumed to be a normal distribution with the background as the mean:

$$\varrho[\vec{x}_t] = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \mathbf{B}}} \exp\left(-\frac{1}{2}\left(\vec{x}_t - \vec{x}_b\right)^T \mathbf{B}^{-1}\left(\vec{x}_t - \vec{x}_b\right)\right) \quad (3.3.4)$$

The idea of 3D-Var is to determine the state with the maximum probability, which will be the analysis \vec{x}_a . The requirement

$$\varrho[\vec{x}_t|\vec{y}_o] \stackrel{!}{=} \max \quad (3.3.5)$$

can be simplified to the minimization of a cost function

$$\mathcal{J}[\vec{x}] = \frac{1}{2}\left(\vec{x}_b - \vec{x}\right)^T \mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}\right) + \frac{1}{2}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}]\right)^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}]\right) \stackrel{!}{=} \min. \quad (3.3.6)$$

The minimum of $\mathcal{J}[\vec{x}]$ is attained for $\vec{x} = \vec{x}_a$ and the solution of

$$\begin{aligned} 0 &= \nabla \mathcal{J}|_{\vec{x}_a} \\ &= -\mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}_a\right) - \mathbf{H}_{\vec{x}}^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_a]\right) \end{aligned} \quad (3.3.7)$$

gives the analysis.

The analysis error covariance matrix \mathbf{A} can be derived from equation (3.3.7) including the true state \vec{x}_t .

$$0 = -\mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}_t + \vec{x}_t - \vec{x}_a\right) - \mathbf{H}^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t + \vec{x}_a - \vec{x}_t]\right)$$

and, assuming a linear observation operator,

$$\begin{aligned} &= \mathbf{B}^{-1}\left(\vec{x}_a - \vec{x}_t\right) - \mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}_t\right) - \mathbf{H}^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t] - \mathbf{H}\left(\vec{x}_a - \vec{x}_t\right)\right) \\ &= \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\right)\left(\vec{x}_a - \vec{x}_t\right) - \mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}_t\right) - \mathbf{H}^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t]\right) \end{aligned} \quad (3.3.8)$$

which yields the following identity:

$$\left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\right)\left(\vec{x}_a - \vec{x}_t\right) = \mathbf{B}^{-1}\left(\vec{x}_b - \vec{x}_t\right) + \mathbf{H}^T \mathbf{R}^{-1}\left(\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_t]\right) \quad (3.3.9)$$

Multiplying equation (3.3.9) by its transpose from the right and using the background error $\vec{\varepsilon}_b = \vec{x}_b - \vec{x}_t$, the analysis error $\vec{\varepsilon}_a = \vec{x}_a - \vec{x}_t$, and the observation error $\vec{\varepsilon}_o = \vec{y} - \vec{\mathcal{H}}[\vec{x}_t]$ gives

$$\begin{aligned} & \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \vec{\varepsilon}_a \vec{\varepsilon}_a^T \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) = \\ & \mathbf{B}^{-1} \vec{\varepsilon}_b \vec{\varepsilon}_b^T \mathbf{B}^{-1} + \mathbf{B}^{-1} \vec{\varepsilon}_b \vec{\varepsilon}_o^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{H}^T \mathbf{R}^{-1} \vec{\varepsilon}_o \vec{\varepsilon}_b^T \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \vec{\varepsilon}_o \vec{\varepsilon}_o^T \mathbf{R}^{-1} \mathbf{H} \end{aligned} \quad (3.3.10)$$

Applying the expected value to this equation and using the assumption of no correlation of observations with the background finally leads to

$$\begin{aligned} & \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \mathbf{A} \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) = \mathbf{B}^{-1} \mathbf{B} \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{H} \\ & \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \mathbf{A} \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \\ & \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \mathbf{A} = \mathbf{1} \\ & \mathbf{A} = \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \end{aligned} \quad (3.3.11)$$

which is the same as the inverse of the Hessian matrix of the cost function.

3.3.1. Equivalence to OI

In the case of 3D-Var, an analysis state \vec{x}_a is found which fulfills equation (3.3.7) and the following equation can be taken as valid:

$$0 = -\mathbf{B}^{-1} (\vec{x}_b - \vec{x}_a) - \mathbf{H}_{\vec{x}_a} \mathbf{R}^{-1} (\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_a]). \quad (3.3.12)$$

In the case of a linear observation operator $\vec{\mathcal{H}}$, this can be transformed into

$$\vec{x}_a = \vec{x}_b + \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_b]). \quad (3.3.13)$$

With the Sherman-Morrison-Woodbury formula (Hager, 1989), it can be shown that

$$\left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} = \mathbf{B} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T \right)^{-1} = \mathbf{W} \quad (3.3.14)$$

with \mathbf{W} from equation (3.2.5).

Further, the analysis error covariance matrix of the 3D-Var can be transformed by the

Sherman-Morrison-Woodbury formula to

$$\begin{aligned}
 \mathbf{A} &= \left(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \\
 &= \mathbf{B} - \mathbf{B} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T \right) \mathbf{H} \mathbf{B} \\
 &= \left(\mathbb{1} - \mathbf{B} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T \right)^{-1} \mathbf{H} \right) \mathbf{B} \\
 &= (\mathbb{1} - \mathbf{W} \mathbf{H}) \mathbf{B}
 \end{aligned} \tag{3.3.15}$$

In the case of a linear observation operator $\vec{\mathcal{H}}$, the 3D-Var is therefore equivalent to the OI.

3.3.2. Practical Implementation

A line search algorithm, also called steepest ascent, is implemented to minimize the cost function $\mathcal{J}[\vec{x}]$. The algorithm applies the following iterative method:

1. Starting from a first guess \vec{x}_g the minimum of \mathcal{J} is sought in the direction of the gradient.

$$\vec{g} = \nabla \mathcal{J}|_{\vec{x}_g} \tag{3.3.16}$$

The minimum in direction of the gradient is located at

$$\vec{x} = \vec{x}_g + \alpha \vec{g} \tag{3.3.17}$$

2. α is determined by solving

$$\mathcal{J}[\vec{x}_g + \alpha \vec{g}] \stackrel{!}{=} \min \tag{3.3.18}$$

and

$$\frac{d}{d\alpha} \mathcal{J}[\vec{x}_g + \alpha \vec{g}] \stackrel{!}{=} 0 \tag{3.3.19}$$

The function $\mathcal{J}[\vec{x}_g + \alpha \vec{g}]$ is approximated by a Taylor series expansion up to second order at $\alpha = 0$.

$$\mathcal{J}[\vec{x}_g + \alpha \vec{g}] \approx \mathcal{J}[\vec{x}_g] + \alpha \vec{g}^T \vec{g} + \frac{1}{2} \alpha^2 \vec{g}^T \mathbb{H} \vec{g} \tag{3.3.20}$$

with the Hessian matrix \mathbb{H} of \mathcal{J} .

The Hessian matrix \mathbb{H} is obtained by a series expansion of the gradient up to first order at the point \vec{x}_g :

$$\nabla \mathcal{J}|_{\vec{x}} \approx -\mathbf{B}^{-1} (\vec{x}_b - \vec{x}) - \mathbf{H}^T \mathbf{R}^{-1} (\vec{y}_o - \vec{\mathcal{H}}[\vec{x}_g]) + \mathbf{H} \mathbf{R}^{-1} \mathbf{H} (\vec{x} - \vec{x}_g) \tag{3.3.21}$$

$$\mathbb{H} \approx \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \tag{3.3.22}$$

3. Differentiation and root finding of equation (3.3.20) results in

$$\alpha = -\frac{\vec{g}^T \vec{g}}{\vec{g}^T \mathbb{H} \vec{g}} \quad (3.3.23)$$

4. The minimum of this line search

$$\vec{x} = \vec{x}_g - \frac{\vec{g}^T \vec{g}}{\vec{g}^T \mathbb{H} \vec{g}} \vec{g} \quad (3.3.24)$$

is applied as the first guess in step 1 until the length of the gradient drops below a given limit.

The algorithm is an example of possible algorithms with many possibilities for improving the convergence.

For the BFGS method, only an estimate of the Hessian matrix is needed for the first step. Usually, the identity matrix is taken for the first guess. Afterwards the Hessian matrix is updated and evolves during the iterations. In addition to this capability, to work without calculating an explicit Hessian matrix, the Limited Memory BFGS method (L-BFGS) (Liu and Nocedal, 1989) is able to handle large dimension problems and is very often used in DA

3.4. Four Dimensional Variational Analysis (4D-Var)

4D-Var is an extension of 3D-Var to allow for the inclusion of observations distributed in time. The probability of the true state $\vec{x}_t^{\tau_0}$ at analysis time τ_0 when the observations $\vec{y}_o^{\tau_i}$ at time τ_i with $i \in [1, L]$ are given by Bayes' theorem with:

$$\varrho [\vec{x}_t^{\tau_0} | \vec{y}_o^{\tau_1}, \vec{y}_o^{\tau_2}, \dots, \vec{y}_o^{\tau_L}] \propto \varrho [\vec{y}_o^{\tau_1} | \vec{x}_t^{\tau_1}] \varrho [\vec{y}_o^{\tau_2} | \vec{x}_t^{\tau_2}] \dots \varrho [\vec{y}_o^{\tau_L} | \vec{x}_t^{\tau_L}] \varrho [\vec{x}_t^{\tau_0}] \quad (3.4.1)$$

where the model states at time τ_i are linked with that at time τ_0 by the model:

$$\begin{aligned} &\varrho [\vec{x}_t^{\tau_0} | \vec{y}_o^{\tau_1}, \vec{y}_o^{\tau_2}, \dots, \vec{y}_o^{\tau_L}] \propto \\ &\varrho [\vec{y}_o^{\tau_1} | \vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_1} [\vec{x}_t^{\tau_0}]] \varrho [\vec{y}_o^{\tau_2} | \vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_2} [\vec{x}_t^{\tau_0}]] \dots \varrho [\vec{y}_o^{\tau_L} | \vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_L} [\vec{x}_t^{\tau_0}]] \varrho [\vec{x}_t^{\tau_0}]. \end{aligned} \quad (3.4.2)$$

The conditional PDFs are again assumed to have a Gaussian distribution and are

$$\begin{aligned} \varrho [\vec{y}_o^{\tau_i} | \vec{x}_t^{\tau_0}] &= \frac{1}{(2\pi)^{\frac{M}{2}} \sqrt{\det \mathbf{R}_i}} \cdot \\ &\exp \left(-\frac{1}{2} \left(\vec{y}_o^{\tau_i} - \vec{\mathcal{H}}_i [\vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_i} [\vec{x}_t^{\tau_0}]] \right)^T \mathbf{R}_i^{-1} \left(\vec{y}_o^{\tau_i} - \vec{\mathcal{H}}_i [\vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_i} [\vec{x}_t^{\tau_0}]] \right) \right) \end{aligned} \quad (3.4.3)$$

and the marginal PDF is again as in equation (3.3.4).

This probability leads to the cost function

$$\begin{aligned} \mathcal{J}[\bar{x}^{\tau_0}] &= \frac{1}{2} (\bar{x}_b^{\tau_0} - \bar{x}^{\tau_0})^T \mathbf{B}^{-1} (\bar{x}_b^{\tau_0} - \bar{x}^{\tau_0}) \\ &\quad + \frac{1}{2} \sum_{i=1}^L \left(\bar{y}_o^{\tau_i} - \bar{\mathcal{H}}_i [\bar{\mathcal{M}}_{\tau_0 \rightarrow \tau_i} [\bar{x}^{\tau_0}]] \right)^T \mathbf{R}_i^{-1} \left(\bar{y}_o^{\tau_i} - \bar{\mathcal{H}}_i [\bar{\mathcal{M}}_{\tau_0 \rightarrow \tau_i} [\bar{x}^{\tau_0}]] \right) \end{aligned} \quad (3.4.4)$$

which has to be minimized to find the optimal probability for an analysis state $\bar{x}_a^{\tau_0}$ to be the truth.

The root of the gradient of the cost function is sought to minimize the cost function.

$$\begin{aligned} \nabla \mathcal{J}|_{\bar{x}^{\tau_0}} &= -\mathbf{B}^{-1} (\bar{x}_b^{\tau_0} - \bar{x}^{\tau_0}) \\ &\quad - \sum_{i=1}^L \mathbf{M}_{\tau_0 \rightarrow \tau_i, \bar{x}^{\tau_0}}^T \mathbf{H}_{i, \bar{x}^{\tau_i}}^T \mathbf{R}_i^{-1} \left(\bar{y}_o^{\tau_i} - \bar{\mathcal{H}}_i [\bar{\mathcal{M}}_{\tau_0 \rightarrow \tau_i} [\bar{x}^{\tau_0}]] \right) \end{aligned} \quad (3.4.5)$$

As for 3D-Var, the minimum could be found by a line search algorithm as described in section 3.3.2. The method to calculate cost function and gradient is described in more detail in section 4.2 for the MM5 4D-Var system.

In a way similar to the derivation of the analysis error covariance matrix of 3D-Var in equation (3.3.11), this matrix can also be written for 4D-Var as the Hessian matrix of the cost function.

$$\mathbf{A} = \left(\mathbf{B}^{-1} + \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{M} \right)^{-1} \quad (3.4.6)$$

3.4.1. Incremental formulation of 4D-Var

In contrast to the 4D-Var described so far and used in this work, there is also the so called incremental formulation of the 4D-Var. Thus the increment between the background and the actual initial state is defined as:

$$\delta \bar{x}^{\tau_0} = \bar{x}^{\tau_0} - \bar{x}_b^{\tau_0}. \quad (3.4.7)$$

With this increment, equation (3.4.4) reads as:

$$\begin{aligned} \mathcal{J}[\delta \bar{x}^{\tau_0}] &= \frac{1}{2} \delta \bar{x}^{\tau_0 T} \mathbf{B}^{-1} \delta \bar{x}^{\tau_0} \\ &\quad + \frac{1}{2} \sum_{i=1}^L \left(\bar{y}_o^{\tau_i} - \bar{\mathcal{H}}_i [\bar{\mathcal{M}}_{\tau_i} [\bar{x}_b^{\tau_0} + \delta \bar{x}^{\tau_0}]] \right)^T \mathbf{R}_i^{-1} \left(\bar{y}_o^{\tau_i} - \bar{\mathcal{H}}_i [\bar{\mathcal{M}}_{\tau_i} [\bar{x}_b^{\tau_0} + \delta \bar{x}^{\tau_0}]] \right) \end{aligned} \quad (3.4.8)$$

and with a linear model and observation operator:

$$\begin{aligned} \mathcal{J}[\delta\vec{x}^{\tau_0}] &= \frac{1}{2} \delta\vec{x}^{\tau_0 T} \mathbf{B}^{-1} \delta\vec{x}^{\tau_0} \\ &+ \frac{1}{2} \sum_{i=1}^L \left(\vec{y}_o^{\tau_i} - \vec{\mathcal{H}}_i \left[\vec{\mathcal{M}}_{\tau_i} \left[\vec{x}_b^{\tau_0} \right] \right] + \mathbf{H}_i \mathbf{M}_{\tau_i, \vec{x}} \delta\vec{x}^{\tau_0} \right)^T \\ &\cdot \mathbf{R}_i^{-1} \left(\vec{y}_o^{\tau_i} - \vec{\mathcal{H}}_i \left[\vec{\mathcal{M}}_{\tau_i} \left[\vec{x}_b^{\tau_0} \right] \right] + \mathbf{H}_i \mathbf{M}_{\tau_i, \vec{x}} \delta\vec{x}^{\tau_0} \right) \end{aligned} \quad (3.4.9)$$

and with the observational increment:

$$\delta\vec{y}_o^{\tau_i} = \vec{y}_o^{\tau_i} - \vec{\mathcal{H}}_i \left[\vec{\mathcal{M}}_{\tau_i} \left[\vec{x}_b^{\tau_0} \right] \right] \quad (3.4.10)$$

the cost function of the incremental formulation of 4D-Var reads:

$$\begin{aligned} \mathcal{J}[\delta\vec{x}^{\tau_0}] &= \frac{1}{2} \delta\vec{x}^{\tau_0 T} \mathbf{B}^{-1} \delta\vec{x}^{\tau_0} \\ &+ \frac{1}{2} \sum_{i=1}^L \left(\delta\vec{y}_o^{\tau_i} - \mathbf{H}_i \mathbf{M}_{\tau_i, \vec{x}} \delta\vec{x}^{\tau_0} \right)^T \mathbf{R}_i^{-1} \left(\delta\vec{y}_o^{\tau_i} - \mathbf{H}_i \mathbf{M}_{\tau_i, \vec{x}} \delta\vec{x}^{\tau_0} \right) \end{aligned} \quad (3.4.11)$$

and the appropriate gradient is

$$\nabla \mathcal{J} |_{\delta\vec{x}^{\tau_0}} = \mathbf{B}^{-1} \delta\vec{x}^{\tau_0} - \sum_{i=1}^L \mathbf{M}_{\tau_i, \vec{x}}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \left(\delta\vec{y}_o^{\tau_i} - \mathbf{H}_i \mathbf{M}_{\tau_i, \vec{x}} \delta\vec{x}^{\tau_0} \right). \quad (3.4.12)$$

This formulation of the 4D-Var has the advantage that the full forward model is only needed to calculate the observational increments $\delta\vec{y}_o^{\tau_i}$. For the calculation of the cost function and the gradient, only the linear version and adjoint of the model are needed, which leads to much faster calculations. Furthermore, the cost function in equation (3.4.11) is quadratic in the increments $\delta\vec{x}^{\tau_0}$ and therefore there exists only one minimum of the cost function, which makes the minimization much easier and faster.

The disadvantage of this formulation is the assumption of a linear observation operator and model, which is not needed for the full 4D-Var described in section 3.4. This means that this approach gives correct results if, in the time window concerned, the model behavior is nearly linear or the increment $\delta\vec{x}^{\tau_0}$ stays small enough, i.e. the correctness of this formulation depends on the stability of the weather condition, on the size of the assimilation window and on the quality of the background $\vec{x}_b^{\tau_0}$.

3.5. Extended Kalman Filter (EKF)

The Kalman Filter assumes that the initial state of the system is normally distributed and forms a Markov Chain, i.e. the state can be described by a mean and a covariance. The Extended Kalman Filter (EKF) is an extension of the Kalman Filter (KF). While the latter only allows for linear models, the EKF may be applied to non-linear models. EKF and 3D-Var are theoretically equivalent since they both solve equation (3.3.7). However, 3D-Var applies

an iterative algorithm to obtain the analysis while the EKF calculates the error covariance matrices of the background and the analysis and allows the model to propagate then in time to obtain the analysis. Since the numerical method differs between EKF and 3D-Var, a different solution of equation (3.3.7) is obtained. The advantage of EKF is that it gives the Best Linear Unbiased Estimate (BLUE) of the atmosphere and its covariance (Kalnay, 2003) but it has the disadvantages of expensive computing and the possibility that, in cases of unstable systems where observations are too scarce, the EKF might become unstable and not lead to a solution.

The KF is a cycle of OIs concatenated by the forecast of the model state and the error covariance matrix. Since the KF assumes a linear model, this procedure applied to a nonlinear model is called EKF. The assimilation is performed in the following cycle:

1. A previous analysis $\vec{x}_a^{\tau_i}$ with its analysis covariance matrix \mathbf{A}^{τ_i} at time τ_i is propagated to time τ_{i+1} and taken as the background.

$$\begin{aligned}\vec{x}_b^{\tau_{i+1}} &= \vec{\mathcal{M}}_{\tau_i \rightarrow \tau_{i+1}} [\vec{x}_a^{\tau_i}] \\ \mathbf{B}^{\tau_{i+1}} &= \mathbf{M}_{\tau_i \rightarrow \tau_{i+1}} \mathbf{A}^{\tau_i} \mathbf{M}_{\tau_i \rightarrow \tau_{i+1}}^T + \mathbf{Q}_{\vec{\mathcal{M}}, \tau_i \rightarrow \tau_{i+1}}\end{aligned}\quad (3.5.1)$$

2. According to the optimal gain matrix in equation (3.2.5), the Kalman gain matrix is defined by

$$\mathbf{K}^{\tau_{i+1}} = \mathbf{B}^{\tau_{i+1}} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{B}^{\tau_{i+1}} \mathbf{H}^T \right)^{-1}. \quad (3.5.2)$$

and the analysis at time τ_{i+1} is determined according to equation (3.2.1) and 3.2.6.

$$\begin{aligned}\vec{x}_a^{\tau_{i+1}} &= \vec{x}_b^{\tau_{i+1}} + \mathbf{K}^{\tau_{i+1}} \left(\vec{y}_o^{\tau_{i+1}} - \vec{\mathcal{H}} [\vec{x}_b^{\tau_{i+1}}] \right) \\ \mathbf{A}^{\tau_{i+1}} &= (\mathbb{1} - \mathbf{K}^{\tau_{i+1}} \mathbf{H}) \mathbf{B}^{\tau_{i+1}}\end{aligned}\quad (3.5.3)$$

The analysis and analysis covariance matrix of time τ_{i+1} is then used in step 1 to propagate to the background at time τ_{i+2} .

3.5.1. Equivalence to the 4D-Var

Assuming an unbiased linear model and observations at only one time, it is possible to show the equivalence of 4D-Var and the KF. For the sake of simplicity, it is assumed that $\mathbf{M} = \mathbf{M}_{\tau_0 \rightarrow \tau_1, \vec{x}^{\tau_0}}$, $\vec{\mathcal{M}} = \vec{\mathcal{M}}_{\tau_0 \rightarrow \tau_1}$, and $\mathbf{H} = \mathbf{H}_{\vec{x}^{\tau_1}}$.

The gradient can be written as

$$\nabla \mathcal{J}|_{\vec{x}^{\tau_0}} = -\mathbf{B}^{-1} (\vec{x}_b^{\tau_0} - \vec{x}^{\tau_0}) - \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \left(\vec{y}_o^{\tau_1} - \vec{\mathcal{H}}_1 [\vec{x}^{\tau_1}] \right). \quad (3.5.4)$$

In the case of a linear model, this is

$$\vec{x}_b^{\tau_0} - \vec{x}^{\tau_0} = \mathbf{M}^{-1} (\vec{x}_b^{\tau_1} - \vec{x}^{\tau_1}) \quad (3.5.5)$$

and equation (3.5.4) becomes

$$\nabla \mathcal{J}|_{\vec{x}^{\tau_0}} = -\mathbf{B}^{-1} \mathbf{M}^{-1} \left(\vec{\mathcal{M}} [\vec{x}_b^{\tau_0}] - \vec{x}^{\tau_1} \right) - \mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \left(\vec{y}_o^{\tau_1} - \vec{\mathcal{H}}_1 [\vec{x}^{\tau_1}] \right) \quad (3.5.6)$$

or

$$\begin{aligned} \nabla \mathcal{J}|_{\vec{x} \tau_1} = (\mathbf{M}^T)^{-1} \nabla \mathcal{J}|_{\vec{x} \tau_0} = & - (\mathbf{M}^T)^{-1} \mathbf{B}^{-1} \mathbf{M}^{-1} \left(\mathcal{M} [\vec{x}_b^{\tau_0}] - \vec{x}^{\tau_1} \right) \\ & - \mathbf{H}^T \mathbf{R}^{-1} \left(\vec{y}_o^{\tau_1} - \vec{H}_1 [\vec{x}^{\tau_1}] \right) \end{aligned} \quad (3.5.7)$$

i.e.

$$\nabla \mathcal{J}|_{\vec{x} \tau_1} = (\mathbf{M} \mathbf{B} \mathbf{M}^T)^{-1} \left(\mathcal{M} [\vec{x}_b^{\tau_0}] - \vec{x}^{\tau_1} \right) - \mathbf{H}^T \mathbf{R}^{-1} \left(\vec{y}_o^{\tau_1} - \vec{H}_1 [\vec{x}^{\tau_1}] \right). \quad (3.5.8)$$

Equation (3.5.8) is equivalent to a 3D-Var at time τ_1 with a background error covariance matrix $\mathbf{B}' = \mathbf{M} \mathbf{B} \mathbf{M}^T$, i.e. the background error covariance matrix was propagated as described for the Kalman Filter. Since the equivalence of 3D-Var and OI has already been shown, this means that the 4D-Var is equivalent to a KF in this case.

3.6. Ensemble Kalman Filter (EnKF)

The EnKF was first introduced by Evensen (1994) and is described in detail in Evensen (2003). The EnKF applies the so-called Markov Chain Monte Carlo method to solve the Fokker-Planck equation, which describes the time evolution of the probability density function of the model state.

In contrast to 3D-Var and EKF, a PDF is not characterized by its mean and its covariance matrix, but is described by a set of samples, where each sample is called a member and the whole set is called an ensemble. K is the number of members and the i -th member is \vec{x}_i . Therefore the PDF is not longer restricted to normal distributions.

The expected value from equation (3.1.6) is approximated from the ensemble by

$$\langle q \rangle \approx \langle q \rangle_e = \frac{1}{K} \sum_{i=1}^K q_i. \quad (3.6.1)$$

which yields to the mean of

$$\langle \vec{x} \rangle_e = \frac{1}{K} \sum_{i=1}^K \vec{x}_i. \quad (3.6.2)$$

The covariance matrix \mathbf{P} is approximated by

$$\mathbf{P} \approx \mathbf{P}_e = \frac{1}{K-1} \sum_{i=1}^K (\vec{x}_i - \langle \vec{x} \rangle_e) (\vec{x}_i - \langle \vec{x} \rangle_e)^T \quad (3.6.3)$$

and with the definition of the $N \times K$ ensemble matrix \mathbf{X} consisting of the member deviations, which is the difference of the member and the mean, as columns

$$\mathbf{X} = (\vec{x}_1 - \langle \vec{x} \rangle_e, \vec{x}_2 - \langle \vec{x} \rangle_e, \dots, \vec{x}_K - \langle \vec{x} \rangle_e) \quad (3.6.4)$$

equation (3.6.3) can be written as

$$\mathbf{P}_e = \frac{1}{K-1} \mathbf{X}\mathbf{X}^T. \quad (3.6.5)$$

If \mathbf{X}_b is the ensemble matrix of the background, the Kalman gain of the EKF (equation (3.5.2)) reads then approximately as

$$\mathbf{K}_e = \frac{1}{K-1} \mathbf{X}_b \mathbf{X}_b^T \mathbf{H}^T \left(\mathbf{R} + \frac{1}{K-1} \mathbf{H} \mathbf{X}_b \mathbf{X}_b^T \mathbf{H}^T \right)^{-1}. \quad (3.6.6)$$

and the time propagation of EKF (see equation (3.5.1)) is calculated in the ensemble formulation by

$$\vec{x}_i^{\tau_{j+1}} = \vec{\mathcal{M}}_{\tau_j \rightarrow \tau_{j+1}} \left[\vec{x}_i^{\tau_j} \right] \quad \forall i. \quad (3.6.7)$$

3.6.1. Stochastic EnKF

The stochastic EnKF is performing the analysis step by creating also from the observations an ensemble according \mathbf{R} with \vec{y}_{oi} the K member of this observation ensemble. The analysis equation of the EKF (see equation (3.5.3)) is applied to each member of the ensembles.

$$\vec{x}_{ai} = \vec{x}_{bi} + \mathbf{K}_e \left(\vec{y}_{oi} - \vec{\mathcal{H}}[\vec{x}_{bi}] \right) \quad (3.6.8)$$

This approach gives directly an analysis ensemble, on which equation (3.6.7) can be applied for the time propagation to the next observation time in a cycle.

3.6.2. Ensemble Square Root Filter (EnSRF)

In the square root filter approach, the analysis is performed for the mean

$$\langle \vec{x}_a \rangle_e = \langle \vec{x}_b \rangle_e + \mathbf{K}_e \left(\vec{y}_o - \vec{\mathcal{H}}[\langle \vec{x}_b \rangle_e] \right) \quad (3.6.9)$$

and the analysis covariance matrix is calculated according to equation (3.5.3)

$$\mathbf{A}_e = (\mathbb{1} - \mathbf{K}_e \mathbf{H}) \mathbf{X}_b \mathbf{X}_b^T. \quad (3.6.10)$$

For the further time propagation, an ensemble has to be chosen which fulfills the calculated mean from equation (3.6.9) and the analysis covariance matrix from equation (3.6.10). Therefore an \mathbf{X}_a is searched, which fulfills

$$\mathbf{A}_e = \frac{1}{K-1} \mathbf{X}_a \mathbf{X}_a^T \quad (3.6.11)$$

This decomposition of \mathbf{A}_e can be done with an eigenvalue/eigenvector decomposition. If the matrix \mathbf{U} is the matrix with the eigenvectors of \mathbf{A}_e as columns and \mathbf{D} is the diagonal matrix with the eigenvalues as diagonal elements, it applies:

$$\mathbf{A}_e = \mathbf{U} \mathbf{D} \mathbf{U}^T \quad (3.6.12)$$

The matrix $\sqrt{K-1}\mathbf{U}\sqrt{\mathbf{D}}$, where $\sqrt{\mathbf{D}}$ is the diagonal matrix where all the elements are the square root of the elements of \mathbf{D} , is already a candidate for the matrix \mathbf{X}_a , which fulfills (3.6.11)

$$\begin{aligned} \frac{1}{K-1}\mathbf{X}_a\mathbf{X}_a^T &= \frac{1}{K-1}\left(\sqrt{k-1}\mathbf{U}\sqrt{\mathbf{D}}\right)\left(\sqrt{k-1}\mathbf{U}\sqrt{\mathbf{D}}\right)^T \\ &= \mathbf{U}\sqrt{\mathbf{D}}\sqrt{\mathbf{D}}^T\mathbf{U}^T \\ &= \mathbf{U}\mathbf{D}\mathbf{U}^T \\ &= \mathbf{A}_e, \end{aligned} \tag{3.6.13}$$

but represents an ensemble of size N . If instead we define the matrix \mathbf{D}^K as the diagonal matrix with the K largest eigenvalues and the matrix \mathbf{U}^K is defined as the $N \times K$ -matrix with the associated eigenvectors, \mathbf{X}_a can be chosen as:

$$\mathbf{X}_a = \sqrt{K-1}\mathbf{U}^K\sqrt{\mathbf{D}^K} \tag{3.6.14}$$

for which (3.6.11) is approximately true.

With \mathbf{X}_a and $\langle \bar{x}_a \rangle_e$ the analysis ensemble is completely described and a time propagation can be performed.

3.6.3. Localization

An ensemble with finite size represents only approximately the PDF. Especially variables, which are weak or even not correlated, which is represented by zeros or small numbers in \mathbf{P} , can produce noise in \mathbf{P}_e if the ensemble size is small. Localization is a technique to reduce the noise by adding a priori knowledge of not correlated variables. Technically this localization is performed by a $N \times N$ localization matrix \mathbf{C} which consists of zeros and ones. The diagonal elements c_{ii} of the matrix \mathbf{C} are set to one and the off diagonal elements c_{ij} are set to a value between zero and one according to a-priori knowledge, how strong the correlation is between the variables x_i and x_j . The estimator of the background error \mathbf{P}_e is then replaced by a localized version:

$$\mathbf{P}_e^{\text{loc}} = \mathbf{C} \circ \mathbf{P}_e \tag{3.6.15}$$

which is the Schur-product of the localization matrix and the background error covariance matrix estimator. The Schur-product means elementwise multiplication:

$$\mathbf{A} = \mathbf{B} \circ \mathbf{C} \quad \Leftrightarrow \quad a_{ij} = b_{ij}c_{ij} \quad \forall_{ij} \tag{3.6.16}$$

Note, that in (3.6.16) the sum convention of Einstein is not applied.

Only the technique of localization allows the application of the EnKF to problems with a dimension as high as in weather forecast with a small enough member size of the ensembles that is computationally affordable.

3.7. Comparison of 4D-Var and EnKF

If the model and observation operator are linear and PDFs are Gaussian, the analyzed values produced by 4D-Var at the end of an assimilation window are the same as those produced by the

full KF filter applied to the same time window (e.g. Lorenc (1986)). The incremental 4D-Var is regarded as a practical approximation to the EKF. However, the length of the assimilation window of 4D-Var is limited for non-linear systems because the tangent linear approximation in the incremental 4D-Var becomes less valid with a longer time window. Since 4D-Var does not explicitly calculate the evolution of the background error covariance, it evolves the background error covariance only within a short time window starting from a statistically derived initial value. In this respect, 4D-Var is suboptimal compared to Kalman filtering.

EnKF is an approximation of Kalman filtering and is becoming a viable choice for the operational NWP as a method for data assimilation and the generation of ensemble perturbations. EnKF has advantages over 4D-Var in that covariances are evolved indefinitely. EnKF also provides the analysis errors while other data assimilation schemes, including variational methods, cannot. Moreover, a major advantage of EnKF is its simple implementation and model independence.

It should be noted that EnKF assimilates observations at synoptic times only. In practice, many asynoptic observations are reported, and it is important to treat them adequately. 4D-Var provides the solution, since it uses the dynamical model for the temporal propagation within the data assimilation system. 4D-Var treats asynoptic observations simultaneously to generate the analysis field at any given time. Although the treatment of asynoptic observations looks like a disadvantage of EnKF, Hunt et al. (2004) proposed the Four-dimensional EnKF (4D-EnKF) approach. 4D-EnKF enables adequate treatment of asynoptic observations under the same approximation as 4D-Var but without an adjoint model. Thus, the treatment of asynoptic observations is no longer disadvantageous to EnKF.

Lorenc (2003) and Kalnay et al. (2007a) discussed theoretical comparisons between 4D-Var and EnKF. Although EnKF has several advantages over the variational methods, the major problem with EnKF is associated with a limited ensemble size. The ability to fit observations is limited by the ensemble size, although localization techniques could reduce this problem. It is not clear how well EnKF simultaneously estimates the atmospheric state in a wide range of scales compared to 4D-Var. Since the 4D-Var system is generally complicated due to the presence of TLM and ADJ, the comparative research between 4D-Var and EnKF is rather limited. Once EnKF has been adopted in experimental systems at operational NWP centers, the comparison with the operational 4D-Var system should provide important research outcomes.

Theoretically, 4D-Var and EnKF for linear systems are equivalent to the KF, if 4D-Var has no limitation of assimilation window length and EnKF employs infinite ensemble size. Since those conditions cannot be achieved, they are inherent limitations. In this respect, a hybrid approach may be attractive. Fisher (1998) tried to combine 4D-Var with a reduced-rank KF to indefinitely evolve the background error covariance used in 4D-Var. A hybrid approach of 4D-Var and EnKF is proposed by Lorenc (2003). Since larger-scale fields tend to have longer predictability, a procedure for evolving large-scale background errors indefinitely is strongly desirable.

3.8. Optimal data assimilation scheme for lidar

The most promising candidates for improving DA in the mesoscale are 4D-Var (Rabier et al., 2000) and EnKF (Evensen, 1994), see the discussion by Kalnay et al. (2007a); Gustafsson (2007); Kalnay et al. (2007b).

The 4D-Var needs derivatives of the model and therefore has problems with discontinuities, while the EnKF is independent of derivatives and therefore also handles discontinuities without

difficulty and, for the same reason, the EnKF is much easier to implement. So, normally, simplifications are made for the adjoints of a model for the 4D-Var. Although, for example, the High Resolution Limited Area Modelling (HIRLAM) 4D-Var (Huang et al., 2002), the nonlinear model, is based on a finite difference representation while the tangent linear and the adjoint models are based on a spectral representation, these model differences do not seem to affect the good results of HIRLAM 4D-Var (Gustafsson, 2007). Also automatic compilers (e.g. Tangent linear and Adjoint Model Compiler (TAMC) or Transformation of Algorithms in Fortran (TAF)) are available, which makes the implementation of the TLM and the ADJ much easier. Nevertheless, a lot of manual work is still needed to find and correct errors.

4D-Var has the advantage of being able to use asynchronous observations as they are provided by today's observing systems, like satellites, in a model-consistent way, while EnKF includes discontinuities during the assimilation window. Hunt et al. (2004) extended the EnKF to four dimensions, allowing the assimilation of asynchronous observations.

Finally, for 4D-Var, a background error covariance matrix has to be assumed for the initial state, and it is common to start with a 3D-Var background covariance matrix. Only with a sufficiently long assimilation window does the lack of the errors of the current day become negligible (Kalnay et al., 2006).

As things stand at the moment, 4D-Var and EnKF are strong contenders for operational utilization, but 4D-Var has the advantage of having been in use for a long time at the synoptic scale, e.g. by ECMWF and Meteo-France and also on the mesoscale with HIRLAM. Such a full-scale test of EnKF is not available yet.

For this work, the 4D-Var was chosen since it supported the very high temporal resolution of lidar data and it was already available for use.

4. MM5 4D-Var implementation

4.1. MM5 system

As described in section 3.4 the 4D-Var works with a model to close the gap between analysis time and observation times. The assumption of a perfect model is approximated by an existing model, in this case MM5. It is then also used to generate a forecast from the optimized state in order to investigate the impact of the assimilation.

MM5 is a limited-area, non-hydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale and regional-scale atmospheric circulation. It is the latest in a series developed from a mesoscale model used by Anthes at Penn State in the early 70's that was later documented by Anthes and Warner (1978). Since that time, it has undergone many changes designed to broaden its usage (Grell et al., 1995). These include

- a multiple-nesting capability (1-way, 2-way, moving nests),
- non-hydrostatic dynamics, which allows the model to be used at kilometer scales,
- multitasking capability on shared- and distributed-memory machines,
- various physics options of different complexity.

In the horizontal plane, MM5 is able to use the Lambert conformal (for mid latitudes), Polar stereographic (for polar regions), and Mercator (for regions near to the equator) map projections. A detailed description of the projections can be found in appendix 3 of Grell et al. (1995). For this work, the Lambert conformal projection was used since all model domains were located in mid latitudes. Figure 4.1.1 shows, for example, lines of constant longitude and latitude in the model space (left panel), and lines of constant range in the x and y -directions in the longitudes and latitudes space (right panel).

The vertical discretization is performed in σ -coordinates, given by

$$\sigma = \frac{p - p_t}{p^*} \quad (4.1.1)$$

with p_t the constant pressure of the model top and $p^* = p_s - p_t$, where p_s is the reference pressure at the surface, obtained from the orography, the hydrostatic equation and a constant temperature lapse rate with time. The i^{th} vertical full level is defined by the constant $\sigma_i \in (0, 1)$, while the half levels are centered between two full levels.

The model variables are located in the grid according to Arakawa-B, i.e. the scalar variables are located in the center of a horizontal grid box and on a half level, the horizontal wind is located at the edge of the horizontal grid box and on a half level, and the vertical wind is located in the middle of the horizontal grid box and on full levels.

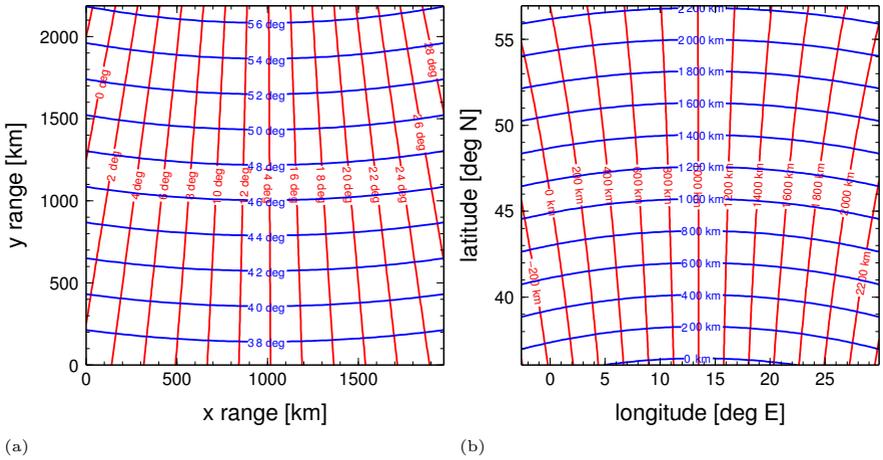


Figure 4.1.1.: Horizontal map projection of the MM5. Shown are lines of constant longitudes and latitudes in model space (a) and lines of constant range in x and y-direction in the space of longitudes and latitudes (b).

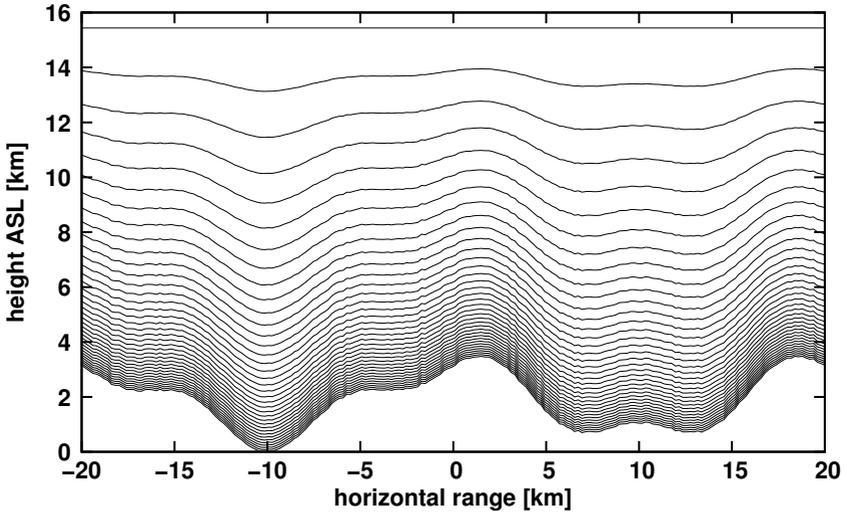


Figure 4.1.2.: Sketch of the terrain following sigma level discretization for an arbitrary orography and 36 vertical levels.

The core of MM5 is composed of the differential equations for pressure

$$\begin{aligned} \partial_t [p^* p'] = & -m^2 \left(\partial_x \left[\frac{p^* u p'}{m} \right] + \partial_y \left[\frac{p^* v p'}{m} \right] \right) - \partial_\sigma (p^* p' \dot{\sigma}) + p' DIV \\ & - m^2 p^* \gamma p \left(\partial_x \left[\frac{u}{m} \right] - \frac{\sigma}{m p^*} \partial_x p^* \partial_\sigma u + \partial_y \left[\frac{v}{m} \right] - \frac{\sigma}{m p^*} \partial_y p^* \partial_\sigma v \right) \\ & + \varrho_0 g \gamma p \partial_\sigma w + p^* \varrho_0 g w, \end{aligned} \quad (4.1.2)$$

momentum in the x -direction

$$\begin{aligned} \partial_t [p^* u] = & -m^2 \left(\partial_x \left[\frac{p^* u u}{m} \right] + \partial_y \left[\frac{p^* v u}{m} \right] \right) - \partial_\sigma [p^* u \dot{\sigma}] + u DIV \\ & - \frac{m p^*}{\varrho} \left(\partial_x p' - \frac{\sigma}{p^*} \partial_x p^* \partial_\sigma p' \right) + p^* f v + D_u, \end{aligned} \quad (4.1.3)$$

momentum in the y -direction

$$\begin{aligned} \partial_t [p^* v] = & -m^2 \left(\partial_x \left[\frac{p^* u v}{m} \right] + \partial_y \left[\frac{p^* v v}{m} \right] \right) - \partial_\sigma [p^* v \dot{\sigma}] + v DIV \\ & - \frac{m p^*}{\varrho} \left(\partial_y p' - \frac{\sigma}{p^*} \partial_y p^* \partial_\sigma p' \right) - p^* f u + D_v, \end{aligned} \quad (4.1.4)$$

momentum in the z -direction

$$\begin{aligned} \partial_t [p^* w] = & -m^2 \left(\partial_x \left[\frac{p^* u w}{m} \right] + \partial_y \left[\frac{p^* v w}{m} \right] \right) - \partial_\sigma [p^* w \dot{\sigma}] + w DIV \\ & + p^* g \frac{\varrho_0}{\varrho} \left(\frac{1}{p^*} \partial_\sigma p' + \frac{T'_v}{T} - \frac{T_0 p^*}{T p_0} \right) - p^* g (q_c + q_r) + D_w, \end{aligned} \quad (4.1.5)$$

and the temperature

$$\begin{aligned} \partial_t [p^* T] = & -m^2 \left(\partial_x \left[\frac{p^* u T}{m} \right] + \partial_y \left[\frac{p^* v T}{m} \right] \right) - \partial_\sigma [p^* T \dot{\sigma}] + T DIV \\ & + \frac{1}{\varrho c_p} \left(p^* \frac{D p'}{D t} - \varrho_0 g p^* w - D_{p'} \right) + D_T \end{aligned} \quad (4.1.6)$$

with

$$DIV = m^2 \left(\partial_x \left[\frac{p^* u}{m} \right] + \partial_y \left[\frac{p^* v}{m} \right] \right) + \partial_\sigma [p^* \dot{\sigma}] \quad (4.1.7)$$

and

$$\dot{\sigma} = -\frac{\varrho_0 g}{p^*} w - \frac{m \sigma}{p^*} \partial_x p^* u - \frac{m \sigma}{p^*} \partial_y p^* v. \quad (4.1.8)$$

Furthermore, the quantities of temperature, pressure and density are calculated as reference

values and perturbations

$$\begin{aligned}
 T(x, y, z, t) &= T_0(z) + T'(x, y, z, t) \\
 p(x, y, z, t) &= p^* \sigma_i + p_t + p' \\
 \varrho(x, y, z, t) &= \varrho_0(z) + \varrho'(x, y, z, t).
 \end{aligned}
 \tag{4.1.9}$$

The differential equations are solved with a semi-implicit scheme developed by Klemp and Wilhelmson (1978). The subscale processes of convection, turbulence microphysics, radiation, and soil interaction are calculated by various parametrizations¹.

Since MM5 is a regional model, it requires an initial condition as well as lateral boundary conditions to run. For a model run, gridded data covering the entire time period are needed. For our experiments, the ECMWF operational analysis provided the necessary data.

The 4D-Var requires adjoint versions of the parametrizations. Since these versions are only available for simple parametrizations, the assimilation and the subsequent free forecasts use a different physical set-up.

4.2. MM5 4D-Var system

As shown in section 3.4, for the implementation of the 4D-Var using this formulation, the calculation of the cost function value and its gradient has to be carried out for an arbitrary initial state. Furthermore, the implementation of an optimization algorithm, in our case the L-BFGS method (see section 3.3.2), is needed. The error covariance matrices for the background \mathbf{B} and the observations \mathbf{R} are needed or have to be based on sufficiently good assumptions. These components are briefly described in the following subsections.

4.2.1. The background error covariance matrix

Assuming that the probability of an analysis for a true state can be described with a Gaussian distribution, the statistical description of the analysis error is the error covariance matrix.

The background-error covariance matrix is the most critical component of every data assimilation system, since it describes the spatial and multivariate structure of the analysis increment. Ideally, \mathbf{B} should express the covariance statistics of the “true” forecast errors.

In its current release, the MM5 4D-Var system only offers diagonal background error covariance matrices. This approximation has, however, proved to work well for most studies conducted with the system (Zou et al., 1995; Xiao et al., 2000). This can be explained by the ability of 4D-Var systems to generate physically consistent structure functions during the model integration. Therefore, the correct specification of \mathbf{B} at the initial time may not be necessary. For each control variable, the background error variances (diagonal elements of the matrix) are specified by the differences between a short term forecast and the initial values at each grid point. At each vertical level, the maximum value of this difference is assigned to all grid points of that level. This creates a vertical profile of forecast errors valid everywhere in the model domain. These forecast errors, squared, are the diagonal elements of the background error covariance matrix. With this method, the background matrix covers the variability of the modeled quantities for each model level.

¹An overview of these parameterisations is given in <http://www.mmm.ucar.edu/mm5/documents/mm5-desc-pdf/sec5.pdf>.

Another option in the MM5 4D-Var system, not used in this thesis, is to prescribe the background error variances from tables published in Parrish and Derber (1992). They contain the observational errors used for the assimilation of radio-soundings in the NCEP Spectral Statistical Interpolation (SSI) global analysis system at the time of the publication.

4.2.2. The observation operator

The observation operator, as introduced in section 3.1.4, maps the state of the atmosphere as represented in the model (model space) to a set of measured values (observational space) unevenly distributed in the model domain. This process takes place in two steps.

The first step considers the calculations that do not depend on the actual state of the model \vec{x} , like the spatial discretization and the mapping of real world coordinates to grid box numbers. All these calculations are executed before the minimization of the cost function is started. Observations are combined into so called super observations by averaging real observations in order to filter information that cannot be represented by the model. For example, it is not possible to reproduce the highly variable water vapor field, as measured with high-resolution observing systems like lidar, with a model using a horizontal resolution of the order of 20–30 km applying a vertical discretization from 1000–100 hPa with 36 levels. At the end of these tasks, the super observations, the indices of the grid boxes concerned and eventually interpolation weights are written in a data format accessible by the assimilation system.

This first part is often called the observation decoder as it rewrites the observation files into a form usable by the assimilation system and avoids multiple calculations in the assimilation system.

The second step is performed during the assimilation, where the modeled state is mapped to observational space. For example, for an integrated observation like Global Positioning System (GPS) slant path delay, the integral has to be calculated from the forecasted values of the model along the path precalculated by the decoder.

4.2.3. The observation error covariance matrix

In the version of the MM5 4D-Var used here, only diagonal observation error covariance matrices are implemented, or rather the errors on each single point observation are given to the system and used internally to calculate the diagonal terms of the error covariance matrix, while the off-diagonal terms are neglected.

Since the observation operator is defined by the description of the observation \mathcal{P}_0 and the description of the model \mathcal{P}_M , see equation (3.1.20), the errors on the observation given to the assimilation system also consist of two parts. First, the error on the observation described by the measurement system and second, the error from the discretization of the model.

4.2.4. Calculation of the cost function value

The cost function is calculated according to section 3.4, applied with the simplification of diagonal background and observation error covariance matrices. Equation (3.4.4) then reads:

$$\begin{aligned} \mathcal{J}[\vec{x}_{t_0}] &= \frac{1}{2} \sum_i \frac{1}{B_{ii}} (x_{b,i} - x_{t_0,i})^2 \\ &+ \frac{1}{2} \sum_{i=1}^N \sum_j \frac{1}{R_{i,jj}} \left(y_{i,j} - \left(\vec{\mathcal{H}}_i \left[\vec{\mathcal{M}}_{t_0 \rightarrow t_i} [\vec{x}_{t_0}] \right] \right)_j \right)^2. \end{aligned} \quad (4.2.1)$$

At the beginning, the background part is calculated from the current state at analysis time, the background state and the background error covariance matrix. Since the background error covariance matrix is diagonal in our case, this part simplifies to

$$\frac{1}{2} \sum_i (\mathbf{B}^{-1})_{ii} (x_{bi} - x_{t_0i})^2,$$

where the index i runs over all grid boxes in the three spatial directions and all model variables.

To calculate the observational part of the cost function, the observations \vec{y}_i are sorted with respect to time, so that $t_i \leq t_{i+1}$ is valid for all \vec{y}_i . Beginning with the first observation time t_1 , the state \vec{x}_{t_1} can be calculated with a model forecast from t_0 to t_1 , $\vec{\mathcal{M}}_{t_0 \rightarrow t_1}$. The observation operator of the corresponding observation is applied to this state in order to map from model space to observational space and calculate the difference between modeled observation and real observation $\vec{y}_1 - \vec{\mathcal{H}}_1[\vec{x}_{t_1}]$. Since the observation error covariance matrix is also diagonal in our case, the observational part of the cost function simplifies to

$$\frac{1}{2} \sum_j (\mathbf{R}_1^{-1})_{jj} \left(y_{1j} - \left(\vec{\mathcal{H}}_1[\vec{x}_{t_1}] \right)_j \right)^2.$$

Afterwards, the model is applied again to the new current state \vec{x}_{t_1} to calculate the state for the next observation time $\vec{x}_{t_2} = \vec{\mathcal{M}}_{t_1 \rightarrow t_2}[\vec{x}_{t_1}]$ and the next summand of the observational part of the cost function can be calculated.

4.2.5. Calculation of the gradient

For the calculation of the gradient of the cost function, the property of the model $\vec{\mathcal{M}}$ of being a time propagator is used.

$$\vec{\mathcal{M}}_{t_0 \rightarrow t_2}[\vec{x}] = \vec{\mathcal{M}}_{t_1 \rightarrow t_2} \left[\vec{\mathcal{M}}_{t_0 \rightarrow t_1}[\vec{x}] \right] \quad \text{when} \quad t_0 \leq t_1 \leq t_2 \quad (4.2.2)$$

From equation (4.2.2) it follows by differentiation

$$\mathbf{M}_{t_0 \rightarrow t_2, \vec{x}_{t_0}}^T = \mathbf{M}_{t_0 \rightarrow t_1, \vec{x}_{t_0}}^T \mathbf{M}_{t_1 \rightarrow t_2, \vec{x}_{t_1}}^T \quad (4.2.3)$$

as a property of the adjoint. Therefore the gradient of the observational part of the cost function can be calculated with the following steps:

1. Calculate the gradient for the last observation time with respect to its time step.

$$\vec{g}_N = \mathbf{H}_{N, \vec{x}_N}^T \mathbf{R}_N^T \left(\vec{y}_N - \vec{\mathcal{H}}_N [\vec{x}_{t_N}] \right) \quad (4.2.4)$$

2. Calculate all gradients from the $(N - 1)^{\text{th}}$ to the first observation for the appropriate time step.

$$\vec{g}_i = \mathbf{M}_{t_i \rightarrow t_{i+1}, \vec{x}_{t_i}}^T \vec{g}_{i+1} + \mathbf{H}_{i, \vec{x}_i}^T \mathbf{R}_i^T \left(\vec{y}_i - \vec{\mathcal{H}}_i [\vec{x}_{t_i}] \right) \quad (4.2.5)$$

3. Calculate the final gradient of the observational part of the cost function.

$$\vec{g}_0 = \mathbf{M}_{t_0 \rightarrow t_1, \vec{x}_{t_0}}^T \vec{g}_1 \quad (4.2.6)$$

4. Finally add the background part of the gradient.

$$\nabla \mathcal{J}|_{\vec{x}_{t_0}} = \mathbf{B}^{-1} (x_{b,i} - x_{t_0,i}) + \vec{g}_0 \quad (4.2.7)$$

Thus, the adjoint of the model $\mathbf{M}_{t_1 \rightarrow t_2, \vec{x}_{t_1}}$ is composed of operators per time step of the model

$$\mathbf{M}_{t \rightarrow T, \vec{x}_t}^T = \prod_{i=0}^{N-1} \mathbf{M}_{t_i \rightarrow t_{i+1}, \vec{x}_{t_i}}^T \quad (4.2.8)$$

with the equidistant decomposition of the time interval using the model time step size Δt

$$\begin{aligned} t_{i+1} &= t_i + \Delta t \\ t_0 &= t \\ t_N &= T. \end{aligned} \quad (4.2.9)$$

Adjoint coding

For the calculation of the gradient, the necessary adjoint operators $\mathbf{H}_{i, \vec{x}}^T$ and $\mathbf{M}_{t \rightarrow T, \vec{x}_t}^T$ are never calculated as matrices with all their elements. They are written as subroutines, as an instruction for how to calculate the result, if the operator is applied to a vector.

As an example, let us look at a non-linear functional $\vec{\mathcal{F}}$ and how to implement the subroutines representing the application of this functional, its linearization and the adjoint.

$$\vec{\mathcal{F}} \left[\begin{pmatrix} a \\ b \\ c \end{pmatrix} \right] = \begin{pmatrix} \sin(a) + \cos(b) \\ ab + 5ac \\ c \end{pmatrix} \quad (4.2.10)$$

This would be coded as:

```
f( a, b, c ) {
  ap = sin(a) + cos(b);
  bp = a*b + 5*a*c;
  a = ap;
  b = bp;
```

}

The linearization would be:

$$\mathbf{F}_{a,b,c} = \begin{pmatrix} \frac{\partial \mathcal{F}_1}{\partial a}(a, b, c) & \frac{\partial \mathcal{F}_1}{\partial b}(a, b, c) & \frac{\partial \mathcal{F}_1}{\partial c}(a, b, c) \\ \frac{\partial \mathcal{F}_2}{\partial a}(a, b, c) & \frac{\partial \mathcal{F}_2}{\partial b}(a, b, c) & \frac{\partial \mathcal{F}_2}{\partial c}(a, b, c) \\ \frac{\partial \mathcal{F}_3}{\partial a}(a, b, c) & \frac{\partial \mathcal{F}_3}{\partial b}(a, b, c) & \frac{\partial \mathcal{F}_3}{\partial c}(a, b, c) \end{pmatrix} \quad (4.2.11)$$

$$= \begin{pmatrix} \cos(a) & -\sin(b) & 0 \\ b + 5c & a & 5a \\ 0 & 0 & 1 \end{pmatrix}$$

and its implementation:

```
l_f( a, da, b, db, c, dc ) {
  dap = cos(a)*da - sin(b)*db;
  dbp = (b + 5*c)*da + a*db + 5*a*dc;
  da = dap;
  db = dbp;
}
```

And the adjoint would be:

$$\mathbf{F}_{a,b,c}^T = \begin{pmatrix} \cos(a) & b + 5c & 0 \\ -\sin(b) & a & 0 \\ 0 & 5a & 1 \end{pmatrix} \quad (4.2.12)$$

with its coding as:

```
a_f( a, da, b, db, c, dc ) {
  dap = cos(a)*da + (5 + b)*db;
  dbp = -sin(b)*da + a*db;
  dcp = 5*a*db + dc;
  da = dap;
  db = dbp;
  dc = dcp;
}
```

Of course, this is a simple example, but this method is also applicable for more complicated routines when the functions are differentiable (Giering and Kaminski, 1998; Nehr Korn et al., 2001).

4.3. Setup of the assimilation experiment

The assimilation experiments are divided into two major steps. First, the assimilation is performed using a coarse horizontal resolution and simplified physics (shown in table 4.3.1). The

restriction to a simplified physical package is necessary since adjoint versions of the parametrization are required for the 4D-Var, which to date are only available for the simplest ones. Therefore, the state vectors in the cost function only include wind, temperature, and water vapor fields on model levels but no surface variables.

The assimilation is performed in coarser resolution since it is much more computationally demanding than a free forecast. For each iteration of the assimilation, the value of the cost function is calculated with a model run and the gradient for each observation is backward propagated with the adjoint model. The number of iterations in the examples presented was of the order of 20 for an observation time window of three hours. To illustrate the computing time needed, it is comparable with the time needed for a free forecast with two additional nests in the domain with a three-times- and nine-times-larger horizontal resolution for 24 hours.

Table 4.3.1.: Parametrization schemes used for the assimilation run.

Physical process	Parametrization used
Convection	Kuo
Boundary Layer	Medium Range Forecast (MRF) (Hong and Pan, 1996)
Cloud micro physics	simple
Land surface	N/A

For the configuration of the MM5 model used to produce free forecasts from both the original ECMWF initial state and an improved initial state generated by the MM5 4D-Var, more sophisticated physical schemes (see table 4.3.2) were selected (Schwitalla, 2007; Schwitalla et al., 2008). The parameterizations are described in more detail in Grell et al. (1995) and references therein.

Table 4.3.2.: Parametrization schemes selected for free forecasts. For the resolution of 1km, the convection parametrization was switched off; otherwise Kain Fritsch 2 was operated.

Physical process	Parametrization used
Radiation	RRTM LW (Mlawer et al., 1997) / Dudhia SW (Dudhia, 1989)
Convection	Kain Fritsch 2 (Kain and Fritsch, 1990, 1993; Kain, 2004) or no parametrization
Boundary Layer	MRF (Hong and Pan, 1996)
Cloud micro physics	Reisner 2 (Reisner et al., 1998)
Land surface	5-layer soil model

5. The IHOP_2002 experiment

5.1. The IHOP_2002 field experiment

The IHOP_2002 field experiment took place from 13 May to 25 June 2002 in the Southern Great Plains (SGP) (see figure 5.1.1) of the United States (US). Its main objective was to determine the relation between the four-dimensional distribution of water vapor and Quantitative Precipitation Forecasting (QPF) skill in a warm season.

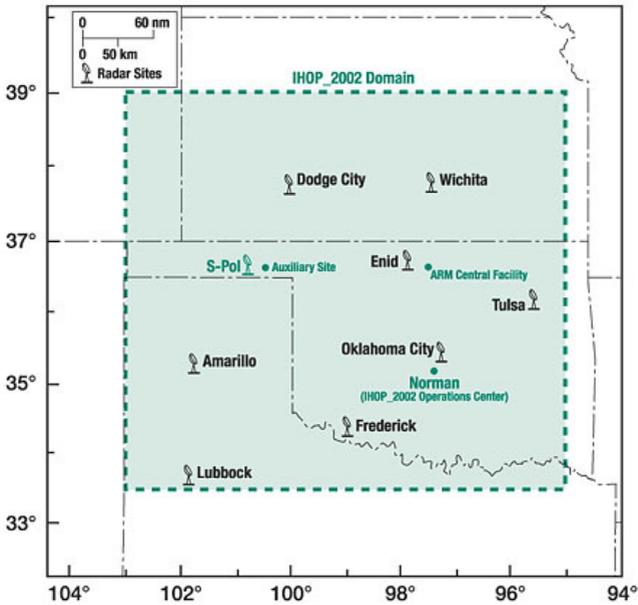


Figure 5.1.1.: Region in the SGP defined as the IHOP_2002 domain.

The scientists associated with IHOP_2002 addressed these general issues through four coordinated research components:

- the QPF research component sought to determine the degree of improvement in forecast skill that occurred through improved characterization of the water vapor field,

- the Convective Initiation (CI) research component sought to better understand and predict the processes that determined where and when convection first formed,
- the Planetary Boundary Layer (PBL) processes research component sought to improve understanding of the relationship between atmospheric water vapor and surface and boundary layer processes and their impact on CI, and
- the instrumentation research component sought to determine the future optimal mix of operational water vapor measurement strategies to better predict warm-season rainfall. This group is also working toward better quantification of measurement accuracy, precision, and performance limitations.

The SGP was chosen for IHOP_2002 for several reasons:

- the extensive array of preexisting operational and experimental instruments,
- both convection initiation and active convection occurring over the region at this time of the year,
- large horizontal gradients in water vapor, and
- a strong east-west gradient in rainfall over this area.

A detailed list of the instrumentation and a description of the aims of IHOP_2002 can be found in Weckwerth et al. (2004).

5.2. Synoptic situation on 24 May 2002

The upper-level analysis showed a short-wave trough moving from west to east over the central Great Plains. This was captured by the 300 hPa velocity field of the 1800 UTC ECMWF operational analysis (see figure 5.2.1(a)). It was associated with a well-defined near-surface short-wave trough visible in the 850 hPa temperature field (figure 5.2.1(b)). It was forecasted to move east-northeastward across the central high plains through 0100 UTC, increasing the gradients of geopotential height, temperature, and deep-layer shear in the IHOP_2002 domain. The associated large-scale ascent also helped to maintain already steep midlevel temperature lapse rates of 8 K/km to 9 K/km. They were superimposed by surface dewpoints of the order of 20 °C in the moist sector. This favored high values of Convective Available Potential Energy (CAPE).

A significant dryline developed during the day resulting in a sharp gradient in the relative humidity over western Texas and eastern New Mexico. This was supported by northward moisture transport from the Gulf of Mexico, which continued during the night by a broad low-level jet. The situation became more complex due to the presence of a cold front, which pushed southward during the day. The dryline and the cold front merged in a triple point, enhancing the probability of CI. These large-scale processes were fairly well forecasted by the different operational forecast systems operated for mission planning, for example, the Eta and the Rapid Update Cycle (RUC) models. Predictions of the locations of the cold front and the dryline together with probabilities for CI were performed by the Storm Prediction Center (SPC) (available from the IHOP_2002 field catalogue; see <http://catalog.eol.ucar.edu/ihop/>). The region of high probability of convection developed around 2000 UTC at the southwestern corner

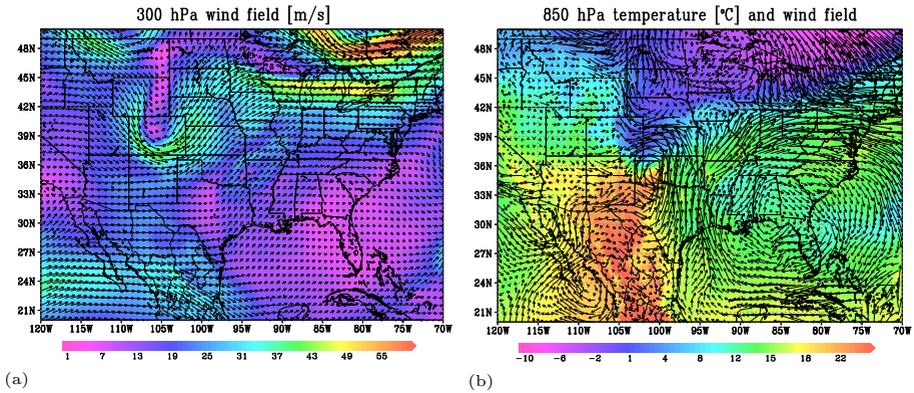


Figure 5.2.1.: ECMWF operational analysis at 1800 UTC on 24 May 2002 of (a) the horizontal wind velocity (m/s) at 300 hPa (colors) overlaid by wind vectors and (b) 850 hPa temperature (°C) (colors) and wind vectors.

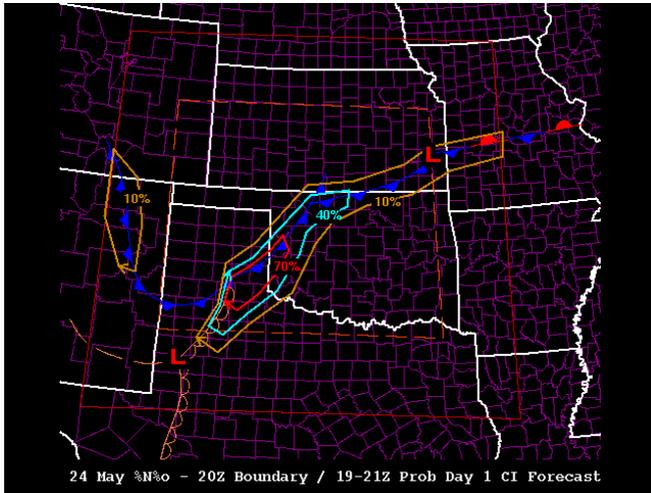


Figure 5.2.2.: Forecasted probability of convection for 2000 UTC, initialized 1700 UTC, by the Storm Prediction Center.

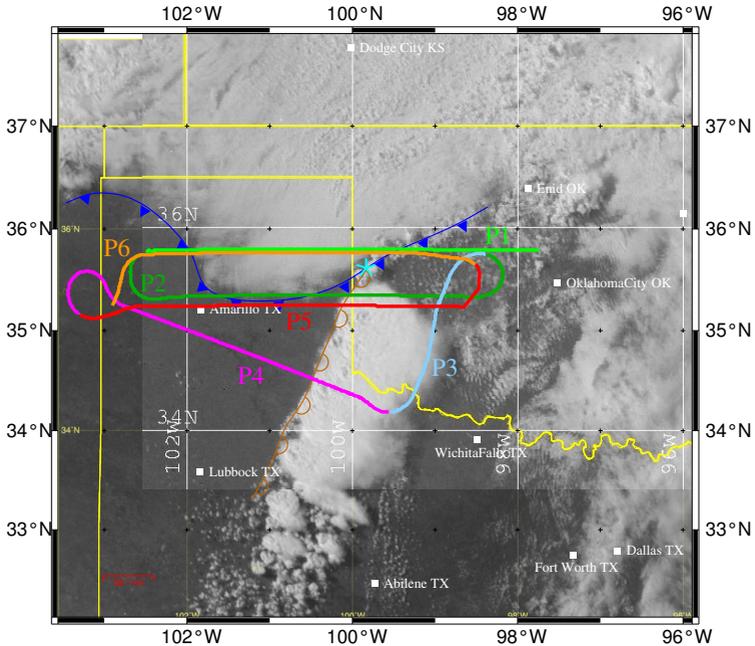


Figure 5.2.3.: The flight path of the NASA DC-8 on 24 May 2002. The path is marked by the labels P1 to P6 to show the temporal order of the segments. It covers the dry, moist and cold regions. Also marked are the dryline and the cold front with the triple point as a blue star. Underlying is the visible GOES-8 image from 2125 UTC.

of Oklahoma (see figure 5.2.2), extended slowly southeastward, and reached its maximum extent around 2300 UTC. For this region, the development of an Mesoscale Convective System (MCS) was predicted during the night.

The development and penetration of the cold front was influenced by outflow boundaries of the previous day’s convective developments in northeastern Texas and Colorado. At 1900 UTC, the cold front was slowly moving southward through the Texas panhandle and northwestern Oklahoma, where high values of CAPE and low values of Convective Inhibition (CIN) were present. According to the mesoscale discussion at the SPC, convection was forecasted to trigger near the triple point at about 2000 UTC (see figure 5.2.2). The estimated locations of the dryline, the cold front, and the triple point at 2125 UTC, based on satellite images and the Lidar Atmospheric Sensing Experiment (LASE) data, described in section 5.3, are indicated in figure 5.2.3.

The convection was initiated along a southwest-northeast band from central Texas to the southwestern corner of Oklahoma. This activity along and east of the dryline occurred in a deeply mixed area and extended rapidly north-northeastward. The development was well captured by the visible channel of GOES-8 (see figure 5.2.3). The east-west-oriented cold front

continued to surge south and undercut the storm near Erick, Oklahoma.

In retrospect, the timing of convection was well predicted by the operational regional models. However, the location of the dryline and the triple point were predicted too far west. CI did not take place at the triple point but about 200 km southwest along the dryline. The intensity of convection was overestimated. Furthermore, details such as the location of outflow boundaries and the fine structure of the dryline were not well predicted. Further discussions and the evolution of convection during this case can be found in Geerts et al. (2006); Martin and Xue (2006); Wakimoto et al. (2006); Xue and Martin (2006a,b); Wulfmeyer et al. (2006).

5.3. LASE data

The DIAL methodology has been introduced in several publications (Browell et al., 1979; Ismail and Browell, 1989; Bösenberg, 1998; Wulfmeyer and Bösenberg, 1998). Further details of advanced DIAL processing schemes can be found e.g. in Bauer et al. (2004); Di Girolamo et al. (2004).

LASE is the first autonomous DIAL system (Browell and Ismail, 1995) that has been operated on various aircraft (Browell et al., 1998). During IHOP_2002, it was deployed on the NASA DC-8. The flight speed was typically 230 m/s and the flight altitude was about 8 km. The retrieved LASE water vapor profiles covered the entire troposphere from the ground to about 6.5 km Above Sea Level (ASL). Further details about the LASE system can be found in Moore Jr. et al. (1997).

During IHOP_2002, LASE water vapor number density profiles were available at a horizontal resolution of about 14 km and a vertical resolution of about 330 m. The horizontal error weighting function of the data was unity for a distance of less than 14 km. For distances of more than 14 km, the errors can be considered uncorrelated. In the vertical direction, the error weighting function W was a triangle function with a side length of 330 m (Wulfmeyer et al., 2004).

Using this retrieval technique, the water vapor profile started at 330 m above the surface. During the postprocessing, water vapor retrievals close to the ground were filled in by utilizing on- and offline ground return signals from the low-gain channel (Browell et al., 1997). Averaged (210 m vertically and 600 m horizontally) atmospheric signals centered at 120 m above ground level were calculated and used in conjunction with ground return signals to obtain the average water vapor number density at 60 m Above Ground Level (AGL). Afterwards, 14 km horizontal averages were calculated. This resulted in a vertical coverage of the water vapor profiles from 60 m above ground up to 6.5 km. In this height range, full overlap between the laser transmitter beam and the receiver was achieved, so associated systematic errors could be neglected.

The power of the DIAL technology is not only the high horizontal and high range resolution of its measurements but also the capability to specify system errors very accurately. In the following discussions, systematic errors are called bias. These errors cannot be removed by signal averaging. Another part of the error is due to system noise. It is assumed that the different noise sources (detector and amplifier noise, Poisson statistics, etc.) are uncorrelated in each range bin. In the following discussions, this kind of error is called noise error.

Systematic errors can be assessed using analytical error propagation (Ismail and Browell, 1989; Bösenberg, 1998; Wulfmeyer and Bösenberg, 1998; Wulfmeyer and Walther, 2001a,b), end-to-end simulators (Bauer et al., 2004; Di Girolamo et al., 2004), and comparisons with other instruments. LASE has been extensively characterized in validation campaigns and the

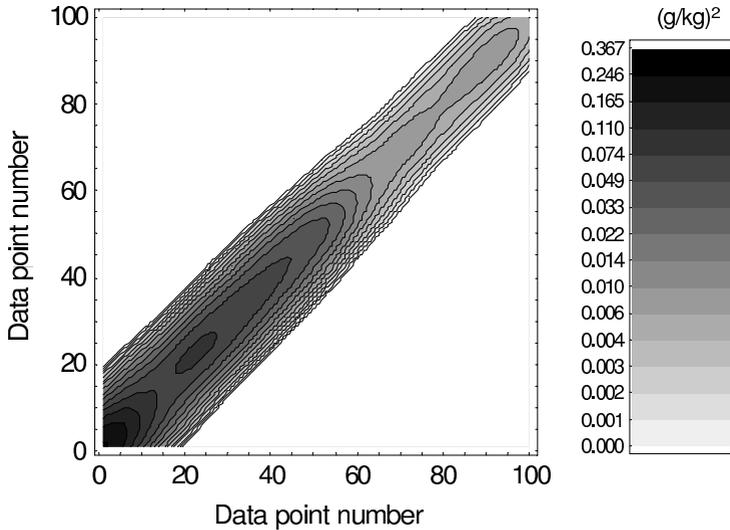


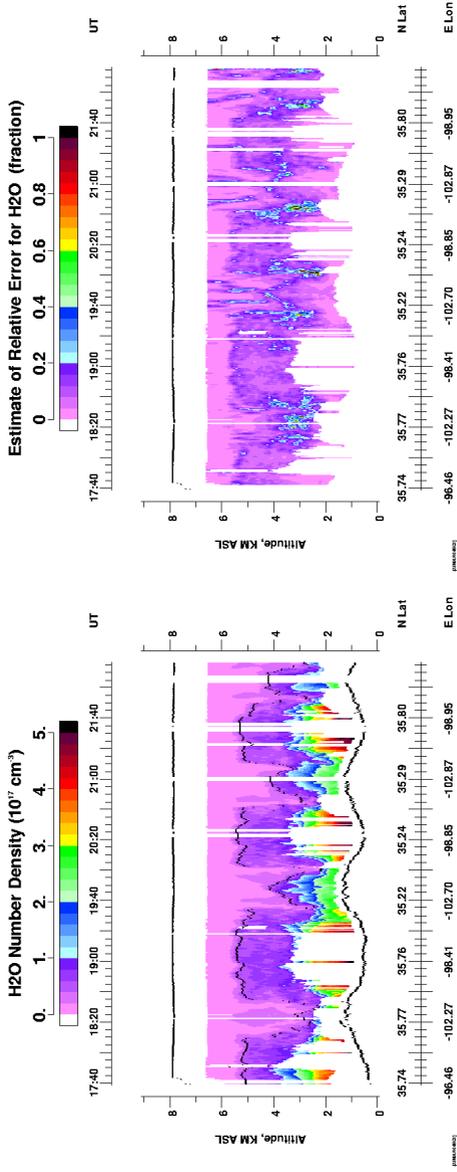
Figure 5.3.1.: Vertical observation error correlation matrix \mathbf{R} of LASE water vapor mixing ratio for 3000 m ASL. The units are in (g^2/kg^2) . The matrix indices between 0 and 100 correspond to a height from 1477 m to 4477 m with a resolution of 30 m

overall Root Mean Square (RMS) accuracy has been determined to be better than 6% over the entire measurement range (Browell et al., 1997).

For data assimilation, it is important to separate the overall error into uncorrelated errors due to system noise and a bias due to remaining uncertainties in the retrieval. The majority of the systematic errors are due to laser performance and a limited knowledge of water vapor spectroscopy. Extensive previous systematic error analyses and comparisons suggest an overall systematic error of LASE water vapor profiles of about 3%.

The Signal-to-Noise Ratio (SNR) of the backscattered signal depends on the atmospheric backscatter properties and the efficiency of the receiver system. An elegant technique exists which can be used to estimate noise errors in each single water vapor profile. This realistic and direct noise error analysis not relying on any other information or instruments is another significant advantage of the DIAL technique. The corresponding analysis has been presented e.g. in Wulfmeyer (1999); Lenschow et al. (2000), so only a short introduction is given here.

First of all, a high-resolution water vapor time series was calculated using a resolution of 1.4 km, which corresponded to an averaging time of 6 s. The autocovariance function of this time series was determined at each height. The difference between the zero lag and the first lag was a good approximation of the noise variance at this horizontal resolution. The noise contribution in the first lag was nearly eliminated, as noise errors were uncorrelated between samples. Noise errors at a different horizontal resolution were estimated by taking the square root of the variance and dividing by the square root of the horizontal resolutions ratio between



(a)

(b)

Figure 5.3.2.: Time-height cross section of water vapor number density measured by LASE (a) and the relative noise error of LASE water vapor number density (b).

the high-resolution time series and the horizontal resolution of interest (Wulfmeyer, 1999). In some cases, water vapor estimates were not available at all heights in the 6 s profiles due to statistical errors being too large. Then, the statistics were calculated from a 2 min running window, thereby filling in midprofile gaps. Within the 2 min window, an error estimate was reported if at least 3 of the 20 water vapor values were valid. This procedure provided some error estimates where there were no water vapor values available. The final step was to eliminate these by masking the error estimates on the 1 min water vapor field.

Finally, the error covariance matrix \mathbf{R} was constructed for each noise error profile ε_n . It reads

$$R_{mk} = \varepsilon_{n,m}^2 C_{mk}. \quad (5.3.1)$$

Here, $\varepsilon_{n,m}$ is the noise variance at model level m . C_{mk} is the autocorrelation function of W between the model levels m and k , hence $C_{mk} = C(z_m - z_k)$ where z_m and z_k are the heights of the model levels. Figure 5.3.1 shows an example of \mathbf{R} for mixing ratio m , calculated by means of equation (5.4.4) using forecasted temperature and pressure profiles from MM5 at the location of the DIAL measurement. The \mathbf{R} shown is nearly diagonal with a strong decay of the off-diagonal elements and supports the assumption of a diagonal observation error covariance matrix in section 4.2.3.

The water vapor number density field measured by LASE is presented in figure 5.3.2(a) the corresponding noise error field is depicted in figure 5.3.2(b) White areas indicate clouds which could not be penetrated by the LASE transmitter. The black line at ground level indicates the height of the surface ASL. The water vapor and error profiles start at 60 m above the ground. Observations were taken from all three regimes present on this day (see section 5.2 for details), along the flight track shown in figure 5.2.3.

Figure 5.3.2 illustrates the high-resolution and good vertical coverage of LASE. Over the entire domain, water vapor profiles could be measured from near the ground up to 6.5 km. The black line, at a height of 5 km, indicates the region where the DIAL retrieval was switched between different absorption line strengths. The height level of switching decreased considerably in the region of the dryline, as the drying also affected the lower troposphere up to a height of 5 km.

Figure 5.3.2(b) demonstrates that the horizontal and vertical variability of noise errors should be taken into account in data assimilation efforts. The vertical variability of the noise is mainly due to changes in the humidity structure. The errors are largest close to regions of frequency switching, since minimal SNRs occur for both online signals. Errors of up to 80 % can appear in limited areas. The errors are typically smaller than 20 %, which is a promising performance for data assimilation.

The dryline was easily detected by the LASE system. Strong gradients of water vapor were observed across the front with dry conditions on the west side of the dryline and very moist conditions to the east. At the dryline, the humidity across the entire troposphere decreased by several g/kg and the cloud coverage decreased.

5.4. Observation operator

As described in section 4.2.2 the observation operator maps from model space to observational space. For the LASE data, this means mapping from the modeled variable water vapor mixing ratio q , temperature T and pressure p to the number density of water vapor molecules N_W , as described in equation (5.4.1), where N_L is Loschmidt's number, M_W is the molecular weight of

water vapor, and R_L is the gas constant of dry air. Equation (5.4.1) follows by simple textbook transformations using the ideal gas law, the definition of the number density, $N_W = \varrho_W \frac{N_L}{M_W}$, and mixing ratio, $q = \frac{\varrho_W}{\varrho_L}$.

$$N_W = \mathcal{H}[p, T, q] = \frac{N_L}{M_W R_L} \frac{p}{T} \frac{q}{1 + 1.608 q} \quad (5.4.1)$$

$$\mathbf{H}_{p,T,q}^T = \begin{pmatrix} \frac{N_L}{M_W R_L} \frac{1}{T} \frac{q}{1 + 1.608 q} \\ -\frac{N_L}{M_W R_L} \frac{p}{T^2} \frac{q}{1 + 1.608 q} \\ \frac{N_L}{M_W R_L} \frac{p}{T} \frac{1}{(1 + 1.608 q)^2} \end{pmatrix} \quad (5.4.2)$$

To examine the meaning of a measurement of water vapor number density, a system is used which is completely described by the triplet of pressure p , temperature T , and water vapor mixing ratio q . The state of the system is therefore:

$$\vec{z} = \begin{pmatrix} p \\ T \\ q \end{pmatrix} \in \mathbb{Z}. \quad (5.4.3)$$

The model which describes this system has the same space and the projection \mathcal{P}_M is unity. Therefore, the observation operator $\vec{\mathcal{H}}$ and the Projection to observation space \mathcal{P}_O are also the same as equation (5.4.1). For this example, the system is assumed to be in the state with $q = 30 \text{ g/kg}$, $T = 20^\circ\text{C}$, and $p = 1024 \text{ hPa}$. The measurement is given by $\mathcal{P}_O[\vec{z}]$ and evaluates to a water vapor number density of $N_W = 1.066 \times 10^{-24} \text{ m}^{-3}$. This measurement does not describe the system completely; instead it describes a subspace of \mathbb{Z} or \mathbb{M} where the quantities of p , T , and q describe the same water vapor number density.

Figure 5.4.1 shows the cost function as a shaded plot and the gradient as vectors assuming an observation error of 6% with a fixed pressure of 1024 hPa. The gradient points mainly in the direction of the water vapor mixing ratio. This means that the measured water vapor density is, as expected, strongly correlated with the water mixing ratio and only weakly coupled with the temperature.

For this reason, the coupling of the water vapor number density to temperature and pressure was skipped by calculating the water vapor mixing ratio in the observation decoder with the temperature and pressure from the first guess using the following equation:

$$q = \frac{N_W}{\frac{N_L}{M_W R_L} \frac{p}{T} - 1.608 N_W}. \quad (5.4.4)$$

This mixing ratio was provided to the assimilation system. The errors of the LASE water vapor number density data were taken from the data providers. In cases where the error was larger than the observation, the observation was refused and not used. Otherwise, if the error was smaller than 6%, a fixed value of 6% of the observation was assumed and was transformed to a water vapor number density error by the following equation.

$$\Delta q = \frac{M_W R_L}{N_L} \frac{T}{p} \left(1 - \frac{M_W R_L}{N_L} \frac{T}{p} 1.608 N_W \right)^{-2} \Delta N_W \quad (5.4.5)$$

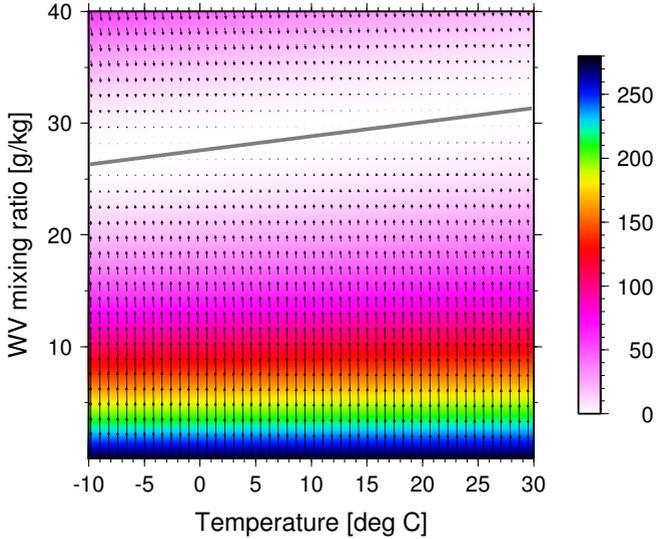


Figure 5.4.1.: Example for the observational part of the cost function (shaded plot) and its gradient (vector plot). The settings of the observed system were $T = 20^\circ\text{C}$, $q = 30\text{ g/kg}$, and $p = 1024\text{ hPa}$. The gray line shows the region where q , T and the fixed value of $p = 1024\text{ hPa}$ represents a water vapor number density of $N_W = 1.066 \times 10^{-24}\text{ m}^{-3}$. The cost function and its gradient were calculated with the assumption of a 6% error in the observation.

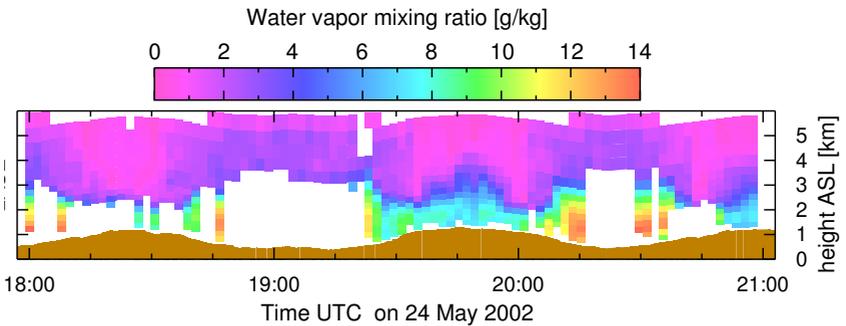


Figure 5.4.2.: Resulting superobservations after mapping the LASE observations to the MM5 model grid.

All observations in one grid box were collected and one super observation per grid box was created as the average of all water vapor mixing ratio values.

The observations were assumed uncorrelated in time and space. This means that the observational error covariance matrix is also diagonal. The diagonal elements of this matrix, or the error for each individual observation, were calculated from the error propagation given by the data providers, unless they were too small.

The diagonal matrix elements are the square of the error values. For example, figure 5.3.1 shows the observation error covariance matrix for one vertical profile of a DIAL measurement. The covariance matrix is nearly diagonal and the decay of the off-diagonal elements is fast.

5.5. Minimization

The assimilation was performed with a horizontal resolution of 30 km in a model domain of 40×40 grid points and 20 vertical layers. The layer thickness is shown in figure 5.5.2 and the domain configuration is given in figure 5.5.1. Figure 5.5.3 shows the vertical error profile as described in section 4.2.1 applied for the simplified background error covariance matrix.

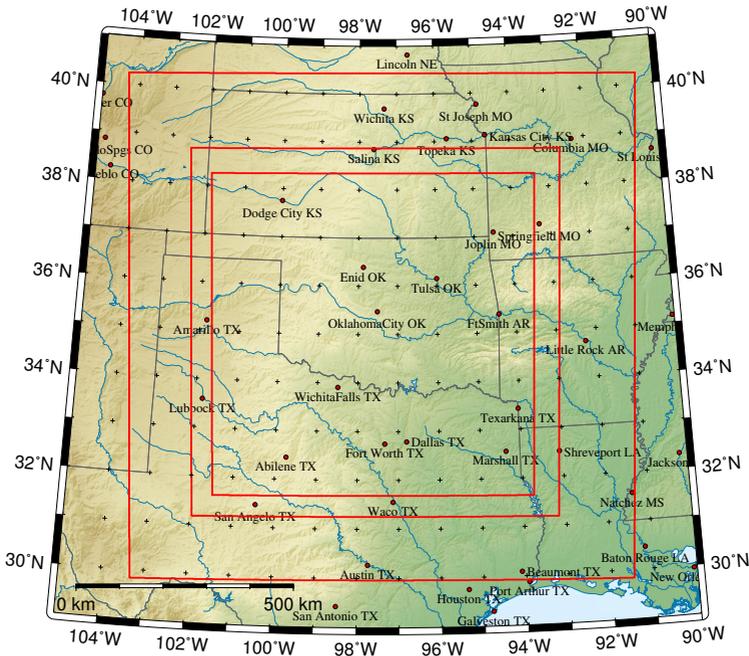


Figure 5.5.1.: The location of the three model domains. The assimilation was performed in the outer domain with 30 km horizontal resolution. For the free forecasts, all domains were used with one-way nesting.

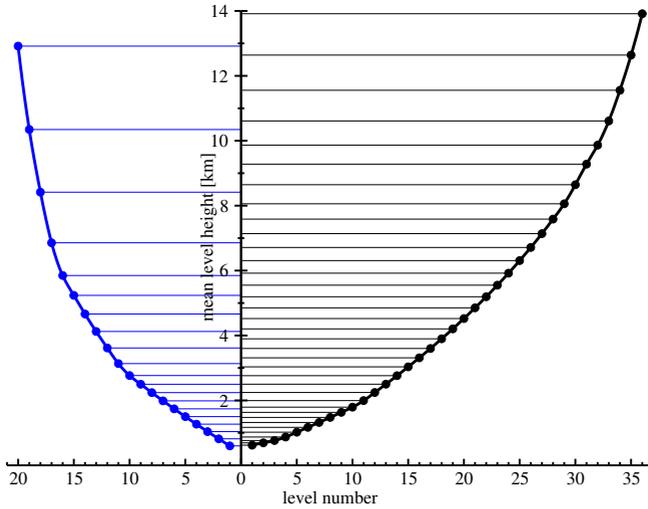


Figure 5.5.2.: Domain averaged heights of the half levels in the configurations with 20 levels (left) and with 36 levels (right).

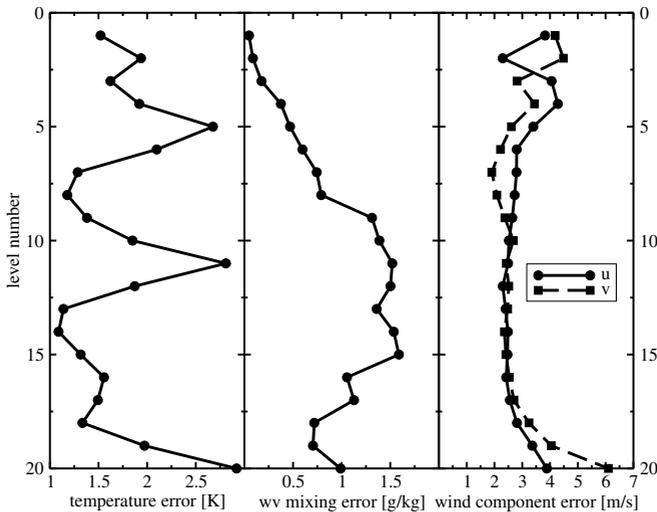


Figure 5.5.3.: Background error estimates for the LASE assimilation experiment in each model level for the three variables temperature, water vapor mixing ratio, and the horizontal wind directions.

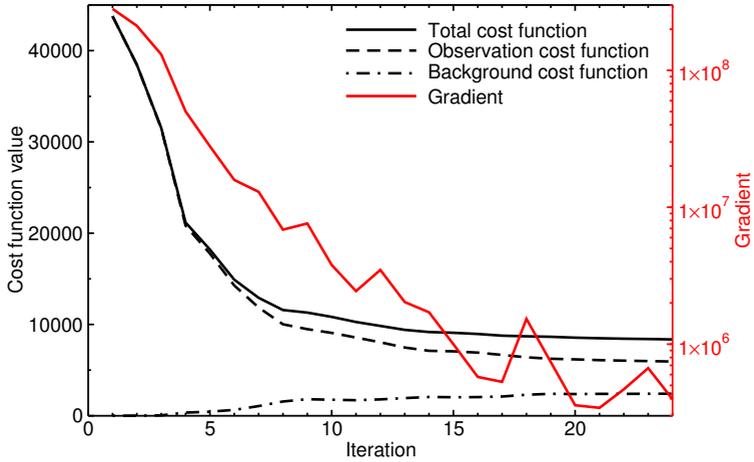


Figure 5.5.4.: Development of the value of the cost function and its components (black) with increasing iterations. Also shown is the decrease of the absolute value of the gradient (red).

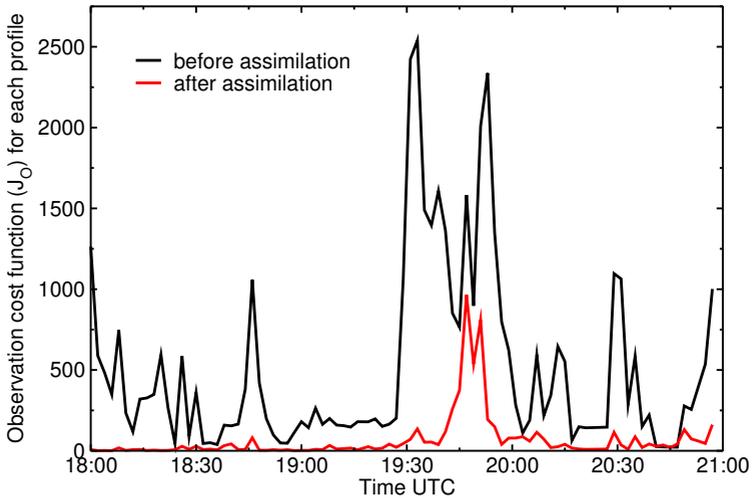


Figure 5.5.5.: Temporal evolution of the observational part of the cost function before and after the minimization.

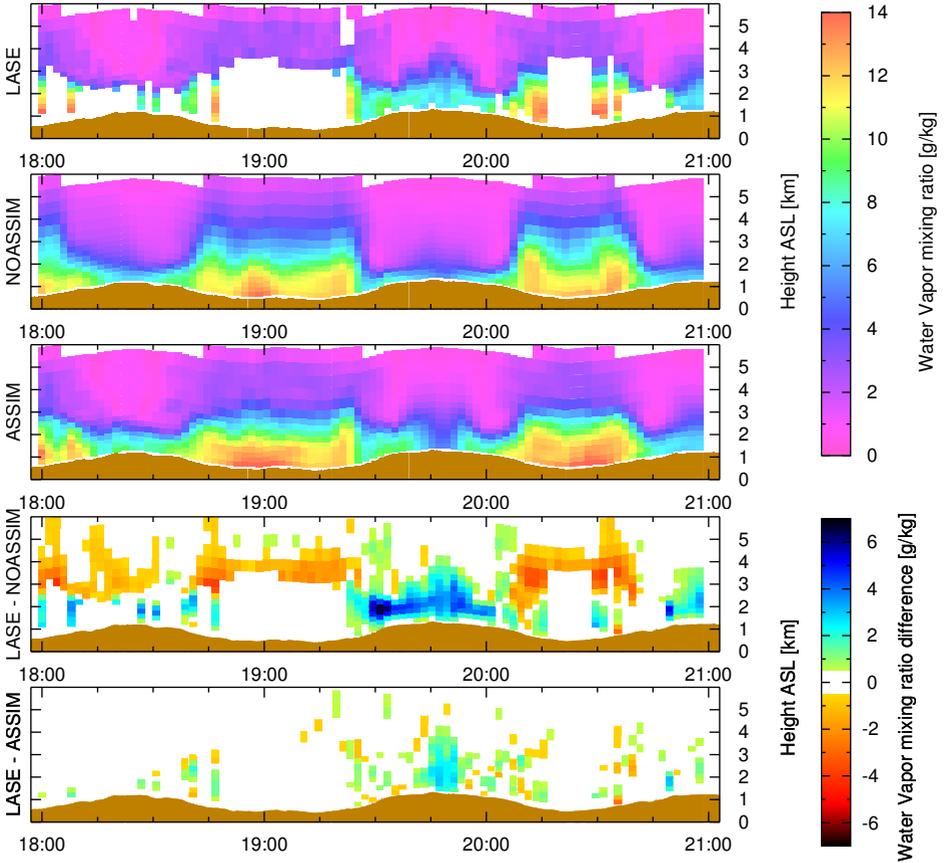


Figure 5.5.6.: Impact of the assimilation on the water vapor field along the flight track. Shown from top to bottom are: the assimilated superobservations as in figure 5.4.2 (LASE), the water vapor field from the model run without assimilation (NOASSIM), the water vapor field from the model run started from the initial state changed by the assimilation (ASSIM), the difference between the observation and the modeled water vapor field without assimilation (LASE-NOASSIM), and the difference between the observation and the modeled water vapor field with assimilation (LASE-ASSIM) (Wulfmeyer et al., 2006).

After the data assimilation run, its impact was studied. Figure 5.5.4 shows the decrease of the cost function during the iterations of the minimization process. The different parts of the cost function, namely the observational and background parts, and the total cost function are shown separately. The observational part decreased rapidly, while the background part increased slightly, which means that the analyzed state left the background state and converged on the observation. The total cost function, the sum of the observational and the background parts, also decreased rapidly. The gradient shown rapidly approached zero, which means that a local minimum of the cost function was reached. After 24 iterations, the minimization was aborted manually.

Figure 5.5.5 shows the observational part of the cost function for each assimilated profile for the first guess and after 24 iterations. For nearly every profile this value was much smaller for the 24th iteration. Only profiles observed between 1945 and 2000 UTC could not be reproduced by the assimilation. Since the gradient in figure 5.5.4 was nearly zero and the total cost function seemed to converge, the assimilation was probably not able to find an initial state, from which the model could reproduce these observations.

In addition to figure 5.5.5, figure 5.5.6 illustrates the vertical impact of the assimilation. Shown in the five panels, from top to bottom, are the assimilated superobservation (first), the observation reproduced from the model with the first guess (second) and the analysis as initial state (third), and the differences between model and observation before (fourth) and after the assimilation (fifth). This illustrates the correction of the model forecast by reducing the amount of water vapor in the upper troposphere and increasing the water vapor content in the boundary layer. The lower two panels of figure 5.5.6, which show the differences between the forecasts with and without assimilation and the observation, also demonstrate this decrease of the cost function.

From the modified initial state at 1800 UTC, a free forecast with the domains shown in table 5.5.1 and figure 5.5.1 was started. Its setup is described in section 4.3.

Table 5.5.1.: MM5 domain configuration applied for the free forecasts. (see also figure 5.5.1)

Resolution	Size
30 km	$40 \times 40 \times 20$
10 km	$85 \times 85 \times 36$
3.33 km	$220 \times 220 \times 36$

5.6. The impact of 4D-Var on the initial fields

The assimilation did not only change the initial fields at the location of the observation, since the gradient of the observational part of the cost function was spread in the model domain by the adjoint of the model. This information spreading of the observation also took place for other variables. In the MM5 4D-Var, information spreading to temperature, wind, water vapor, and pressure is possible. Other parts of the gradient are set to zero by implementation.

Figure 5.6.1 shows an example for the impact of the assimilation on the fields influenced by the 4D-Var at initial time. In figure 5.6.1(b) a decrease of moisture around 35.8°N , 99.8°W and an increase of moisture around 35.8°N , 100.5°W and 35.7°N , 99.2°W at 850 hPa are visible. Also the temperature decreases at 850 hPa around location 35.0°N , 100.2°W . The

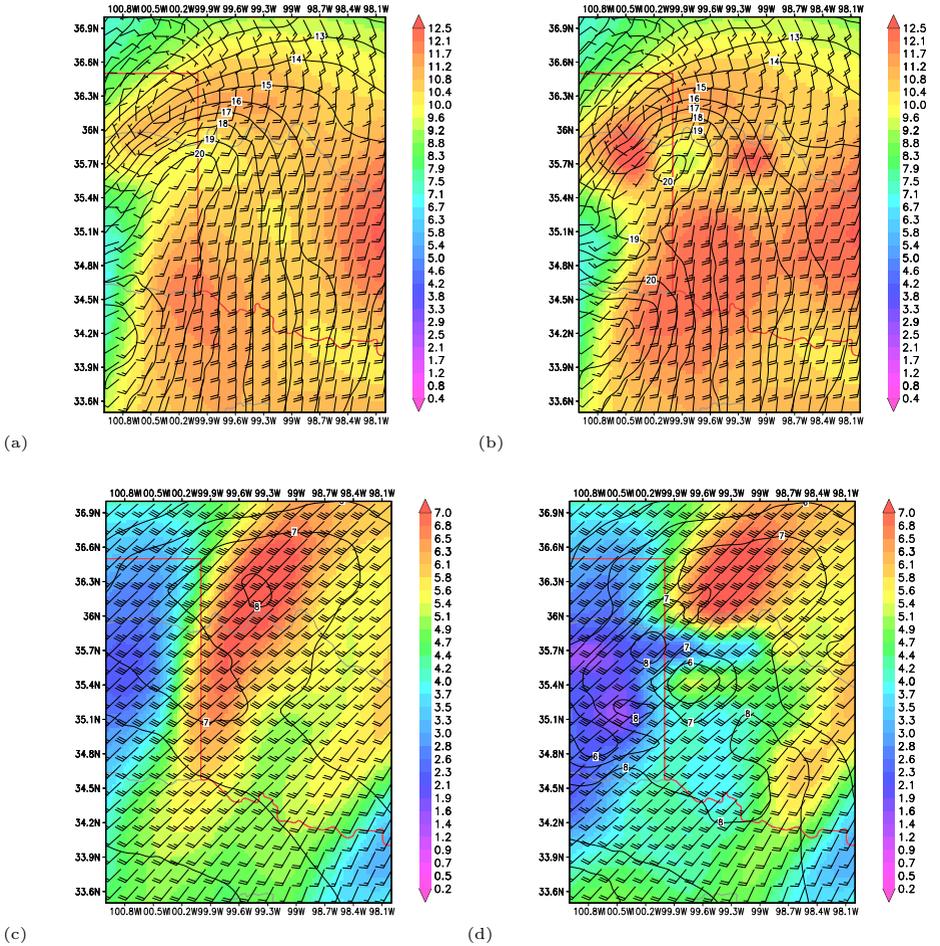


Figure 5.6.1.: Influence of the assimilation on temperature ($^{\circ}\text{C}$) (contours), water vapor (g/kg) (shaded), and wind (barbs) at 850 hPa (upper two pictures) and at 700 hPa (lower two pictures) for the analysis time (1800 UTC). The left column shows the driving ECMWF analysis interpolated into the MM5 grid, whereas the right column shows the initial field, created with the 4D-Var system.

wind field also changes. A vortex, which is visible without assimilation at 36.6°N , 100.4°W , is moved to 36.0°N , 100.4°W .

Figure 5.6.1(d) shows that the change of the water vapor field is more pronounced at 700 hPa. South of 36.0°N , the specific humidity drops from about 6 g/kg to values sometimes lower than 4 g/kg. The temperature is only slightly changed and no change in the wind field is visible.

5.7. The impact on the forecast of convective initiation

In the Mission Scientist Summary for the day in question, convective initiation was reported at 2100 UTC. Figure 5.7.1 is a composite of the Next Generation Weather Radar (NEXRAD) stations Amarillo, Dodge City, and Oklahoma City made for the purposes of the present work and shows the line of developing storms at 2032 UTC from the Texas-Oklahoma border at 34.5°N , 100.0°W directed towards the southwest along the dryline. On the other hand, the CI predicted by SPC to the North of the triple point did not occur.

In the simulation without assimilation (CONTROL) convection also developed along the dryline at 2030 UTC. However, the line was extended northeastward to the triple point (see figure 5.7.2(a)).

On the other hand, in the simulation with assimilation (4DVAR) the convection along the northern part of the dryline was suppressed (see figure 5.7.2(b)) although some precipitation still occurred near the triple point. Furthermore, the southern precipitating area was slightly shifted to the south.

To detect the differences in the model fields, the time of convective initiation was analyzed. Figure 5.7.3 shows the water vapor mixing ratio, temperature, and wind field of the lowest model layer, which is approximately 35 m deep. The dryline is clearly visible as a strong gradient of water vapor. Also obvious is the cold front as a temperature gradient from about 16°C in northwestern Oklahoma, to 25°C and more to the south of the front.

On the plots from 2000 UTC, a clear gradient in the temperature field along the dryline is visible. It is the result of the low-level cold air outflow from the regions of developing convection. Also a drop in the water vapor mixing ratio is visible and the wind field illustrates the outflow from this region with a low-level divergence.

A first indication of this footprint is already visible at 1930 UTC at 35.2°N , 99.9°W in the CONTROL simulation and at 34.9°N , 100.0°W in the 4DVAR simulation.

This means that the initiation of convection was slightly shifted to the south due to the assimilation and was simulated between 1930 UTC and 2000 UTC. Comparing panels (c) and (d) of figure 5.7.3 it appears that the evolution of convection started earlier in the CONTROL simulation as compared to the 4DVAR simulation, since the reduction of water vapor is already further advanced.

Figure 5.7.4 shows vertical soundings from the model fields. They were generated for 35.7°N , 100.0°W in the region where convection was suppressed by the assimilation. The soundings were taken at 1930 UTC, just before the convection was initiated.

No significant differences are visible. The CIN is a little smaller with assimilation. This would mean that it is slightly easier to initiate the convection. So the reason why no convection was triggered along the northern part of the dryline has to be sought elsewhere.

Figure 5.7.5 compares the lowest model layer at 1900 UTC, before the convection was triggered. In the 4DVAR simulation, the northern part of the dryline is not as sharp as in the CONTROL simulation. The temperature gradient is weaker since, in the region around 35.7°N , 100.5°W , an additional region of moist air was included by 4D-Var.

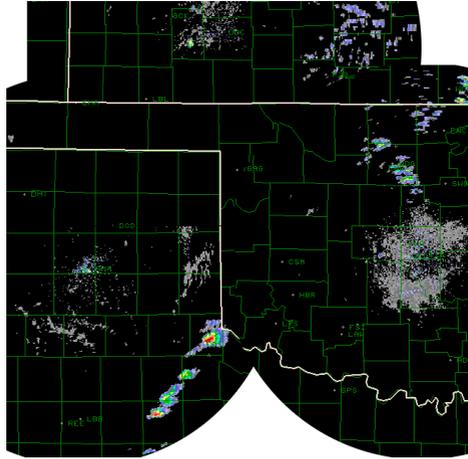


Figure 5.7.1.: Manual composite of the radar reflectivity from the NEXRAD stations at Amarillo, Dodge City, and Oklahoma City at 2032 UTC.

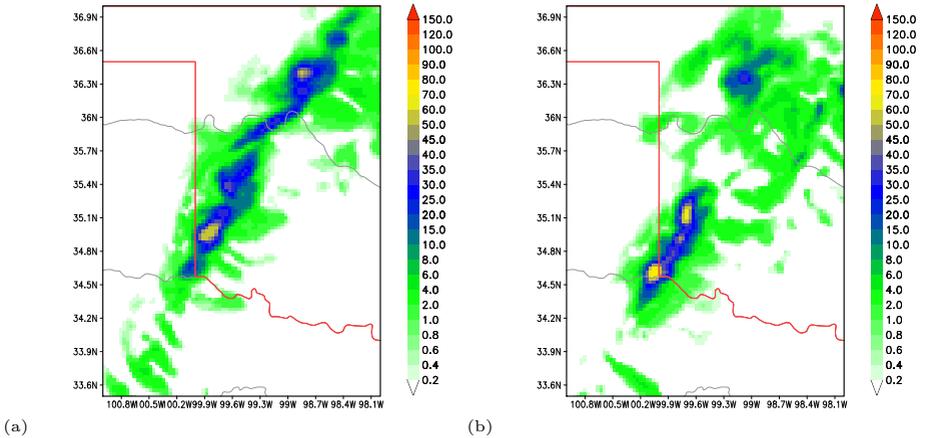


Figure 5.7.2.: 30 min accumulated total precipitation in mm from 2000 UTC to 2030 UTC as simulated by the free MM5 forecast, initialized by ECMWF (a), and 4D-Var (b).

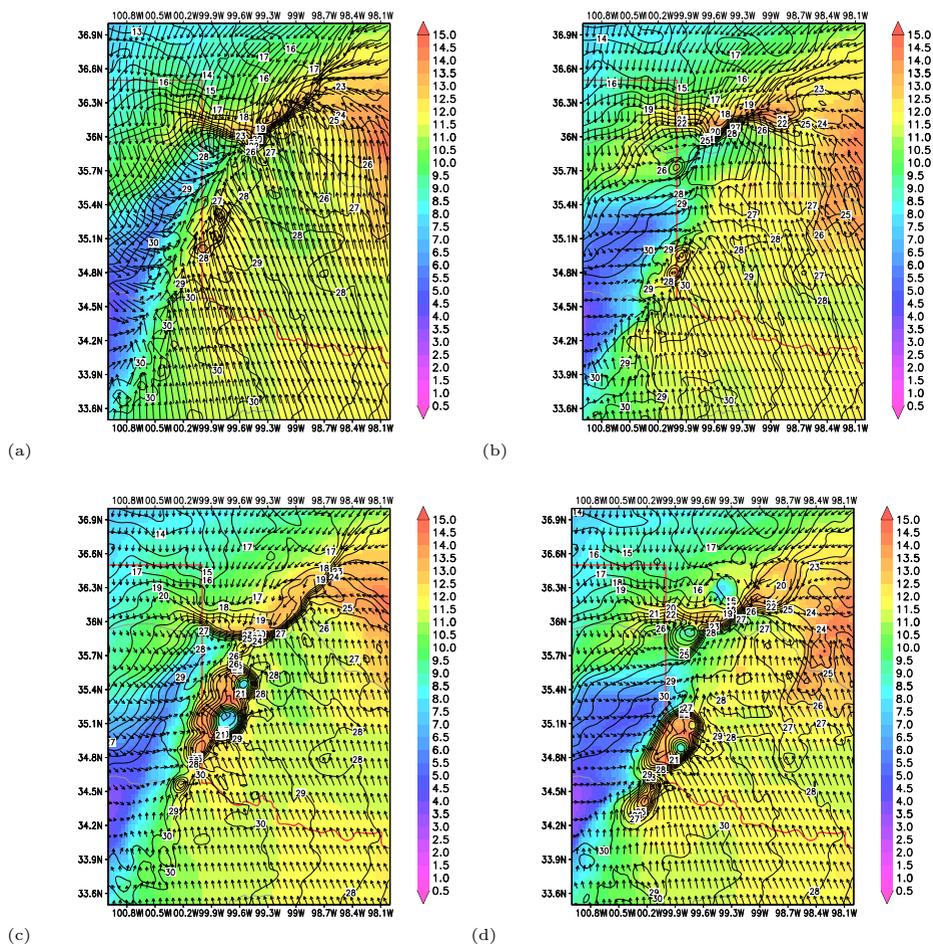


Figure 5.7.3.: Simulated specific humidity, temperature, and wind fields of the lowest model layer (approximately 35 m thick) at 1930 UTC (upper row) and at 2000 UTC (lower row). On the left hand side the CONTROL simulation is shown. The right column shows the result of the 4DVAR simulation. Color shading shows the specific humidity (g/kg), the contours are the temperature ($^{\circ}\text{C}$) and the arrows represent the wind field.

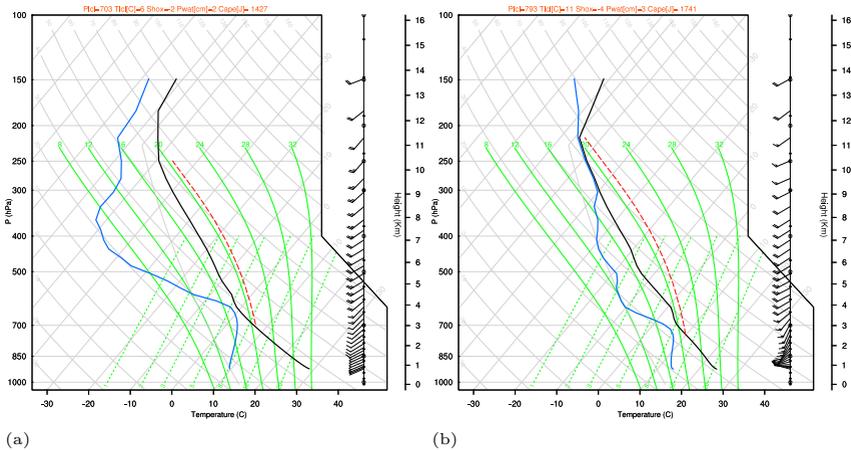


Figure 5.7.4.: Vertical sounding generated from the CONTROL (a) and the 4DVAR simulation (b) at 35.7°N, 100.0°W and 1930 UTC. The black curve is the temperature (°C) and the blue curve is the dewpoint temperature (°C). Also marked are the areas that represent CAPE (red striped) and CIN (green striped).

A stronger change included by the assimilation is visible in the wind field (see figure 5.7.5(c) and (d)). Without assimilation, the wind field to the west of the dryline is bent from southerly flow to northeasterly flow to a convergence zone along the dryline. With assimilation, the wind field is already much further to the north, directed to a separate convergence line which is located about 50 km to the northwest of the previous convergence line.

A reasonable scenario is that the triggering of the convection along the convergence zone along the dryline occurs where enough moist air is available in the lower atmosphere, indicated by high values of CAPE. With the assimilation, the northern part of the convergence line was moved away from the dryline into the drier region with a more stable atmosphere, which explains the suppression of convection.

Also, the vertical cross sections in figure 5.7.6 and 5.7.7 show the modification of the wind and water vapor fields due to the assimilation. An additional divergence line is introduced. The moisture field is modified, as is clearly visible in figure 5.7.7, where the dryline, the region with a strong gradient in the moisture field, is shifted from about 98.4°W to 99.6°W. A second divergence line appears, due to the assimilation in the region of the dryline, and weakens the convergence line. Without assimilation, the dryline is correlated with a region of calm but without divergence or convergence. Nevertheless, the convergence line is associated with a moister area to the west of the still-existing dryline. At this convergence line, convection is initiated, as seen in figure 5.7.8, in the narrow region of reduced temperature and moisture while no CI occurs with assimilation. Although the region near to the convergence line is moister with assimilation, no CI appears, since the convergence is weakened and therefore the CIN is not negotiated.

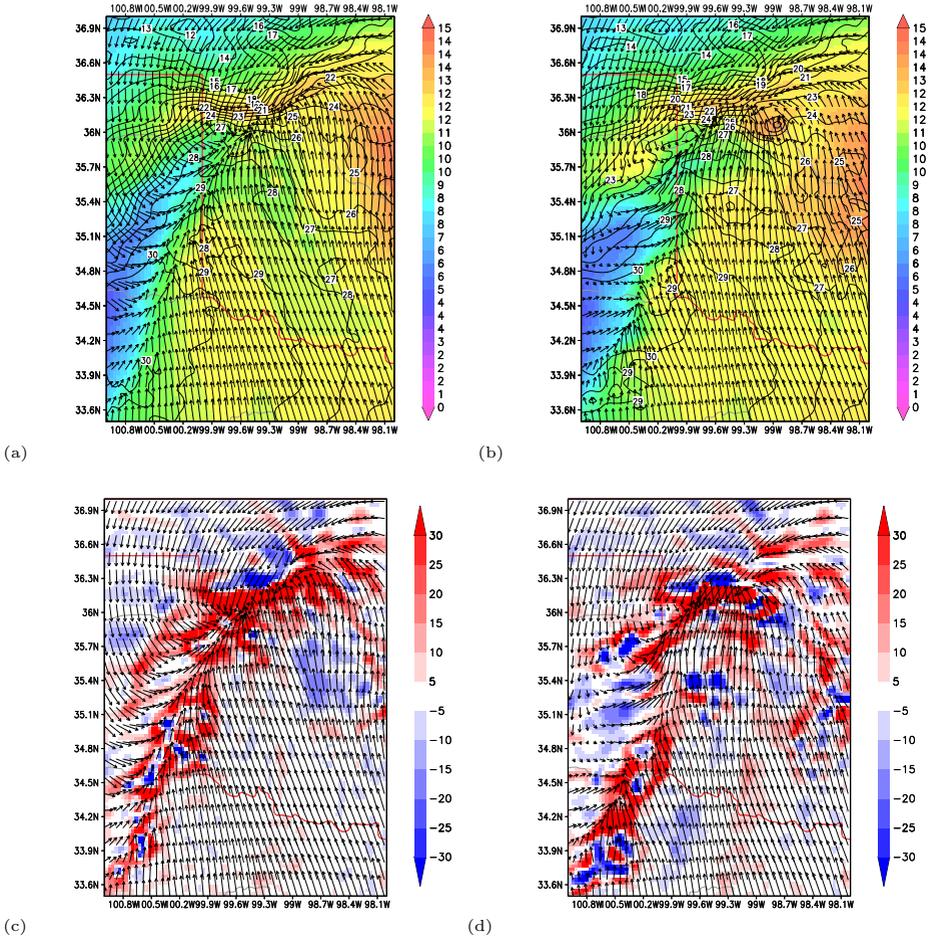
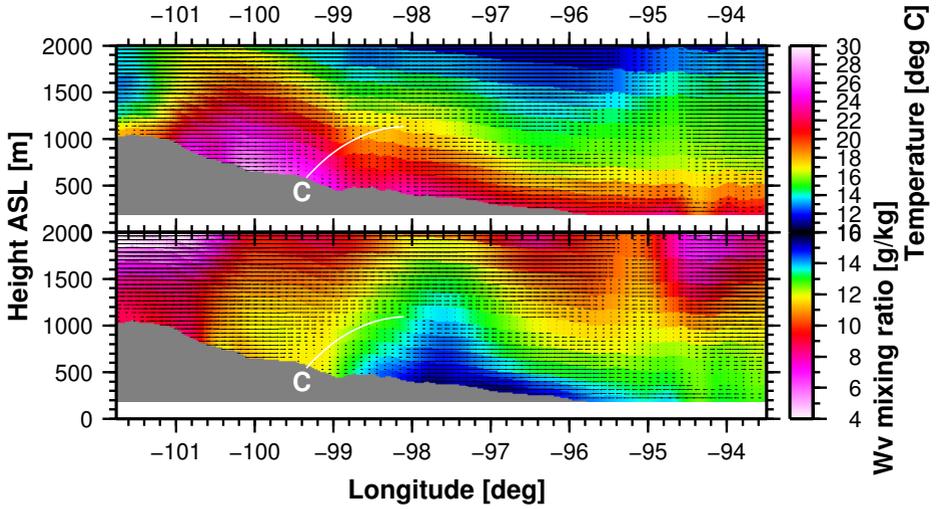
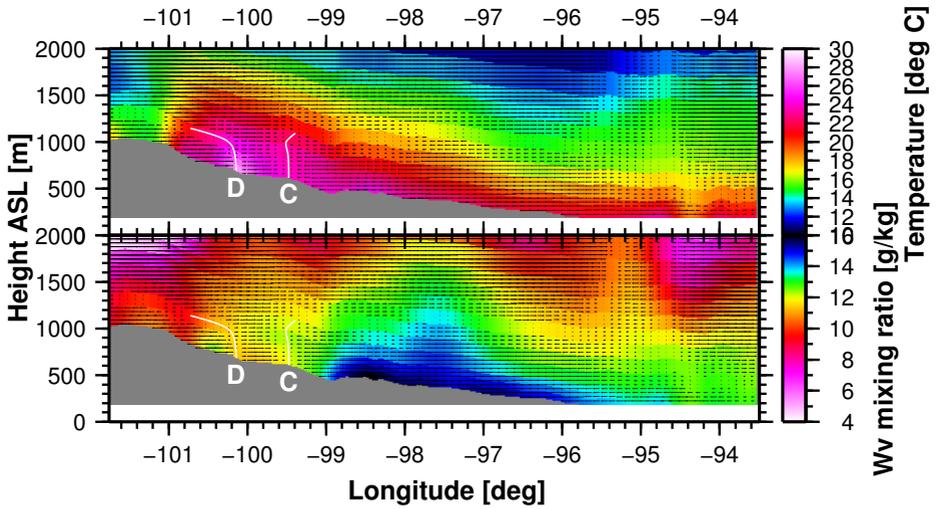


Figure 5.7.5.: The atmosphere state of the lowest model layer (approximately 35 m) at 1900 UTC. The upper two images show the specific humidity (g/kg), the temperature (°C), and the wind field as in figure 5.7.3. In the lower two images, the same wind field is combined with the the moisture convergence (g/(kg s)). Again, on the left side, the state without and, on the right side, the state with assimilation are shown.

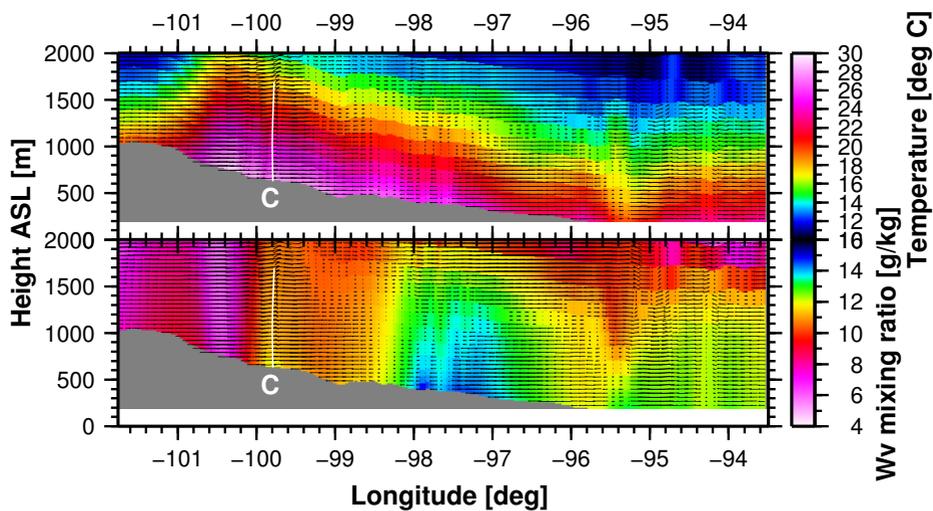


(a)

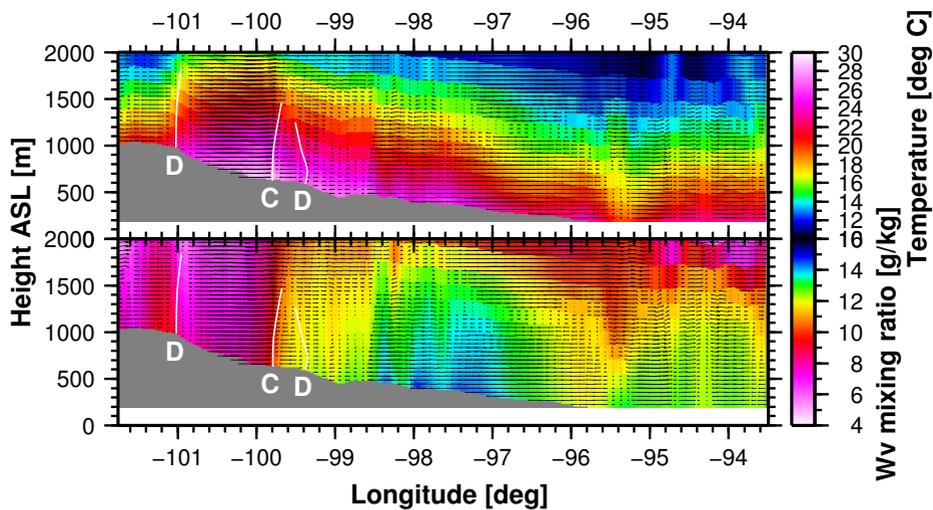


(b)

Figure 5.7.6.: Vertical cross section at 35.4°N through the initial field at 1800 UTC without (a) and with assimilation (b). The upper panel shows the temperature and the lower panel the water vapor mixing ratio. The wind field is overlaid as vectors. Marked convergence lines (C) and divergence lines (D) can be seen.

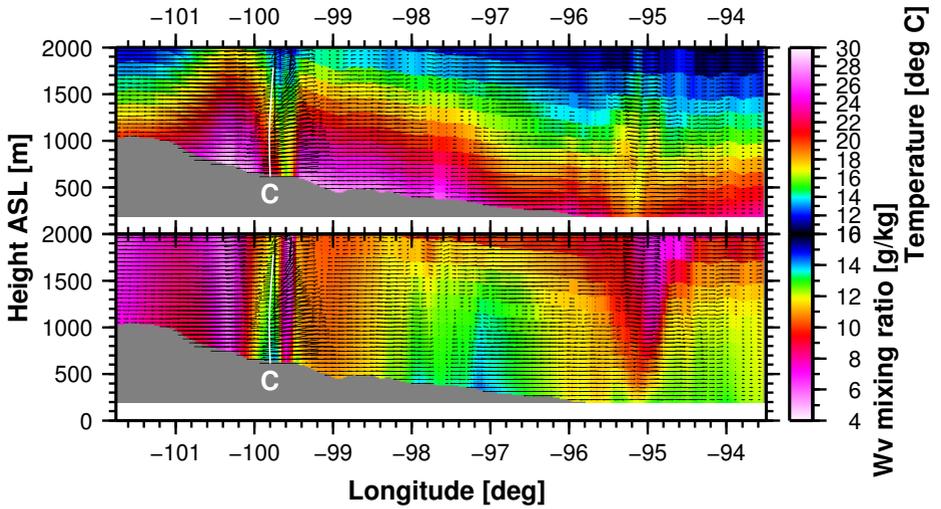


(a)

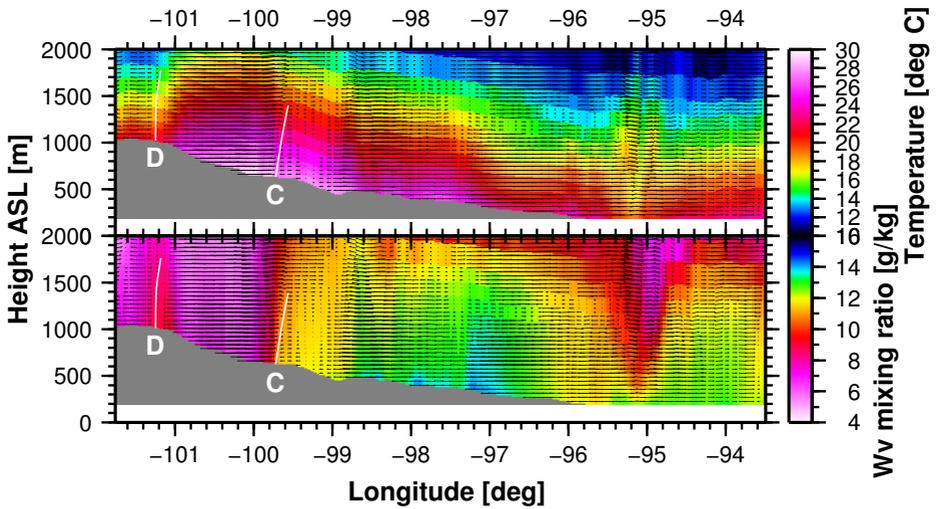


(b)

Figure 5.7.7.: Same as figure 5.7.6 but for 1930 UTC.



(a)



(b)

Figure 5.7.8.: Same as figure 5.7.6 but for 2000 UTC.

6. The LAUNCH-2005 experiment

6.1. LAUNCH-2005

In the framework of the European Cooperation in the field of Scientific and Technical Research (COST)-720, “Integrated Ground-based Remote-Sensing Stations for Atmospheric Profiling”, of the European Science Foundation (ESF) (see also <http://www.cost.esf.org>) and in connection with the World Meteorological Organization (WMO) Global Energy and Water Cycle Experiment (GEWEX) Working Group on Cloud and Aerosol Profiling (CAP), Meteorological Richard-Aßmann-Observatory in Lindenberg (MOL), operated by the German Meteorological Service (DWD), organized the international campaign LAUNCH-2005 (Engelbart and Haas, 2006) in late summer/early fall 2005. It was designed to accomplish four major scientific objectives:

1. Assessment of new or improved profiling systems like water vapor lidars, cloud-radar systems, various microwave profiler systems, a Doppler wind lidar, a new single-photon counting high-range ceilometer, and the redesigned Fourier Transform Infrared Spectroscopy (FTIR) spectrometer, Emission-Infrared Spectrometer for Atmospheric Research (EISAR)
2. Assessment of various algorithms, combining different techniques for profiling of cloud parameters
3. Provision of a data set designed for validation and comparisons between measurements and NWP output
4. Provision of a data set for data assimilation experiments using high-resolution water vapor profiling systems in regional NWP modelling

It was planned to concentrate on approaching precipitation systems and to investigate their correct representation in numerical models. This required a large coverage of the observations. For a first Observation System Experiment (OSE) using high-resolution water vapor profiling systems, a network of water vapor Raman lidar systems was installed at a number of locations in Germany, the United Kingdom, and Italy. This network consisted of 13 Raman lidars and two water vapor DIAL systems intended to provide insight into the usefulness of lidar data assimilation for operational NWP forecasts.

The part of the LAUNCH-2005 network that is important for this study is depicted in figure 6.1.1. Three Raman lidar systems were located in Lindenberg (52.21°N, 14.12°E), Ziegendorf (53.31°N, 11.84°E), and Leipzig (51.35°N, 12.43°E). The sides of this triangle had lengths of 150 km (Leipzig - Lindenberg), 200 km (Lindenberg - Ziegendorf), and 220 km (Leipzig - Ziegendorf). Major questions to be investigated included:

- Do data of water vapor lidar systems have a detectable impact on NWP forecasts?
- How long does this potential impact survive after the end of the data assimilation window?

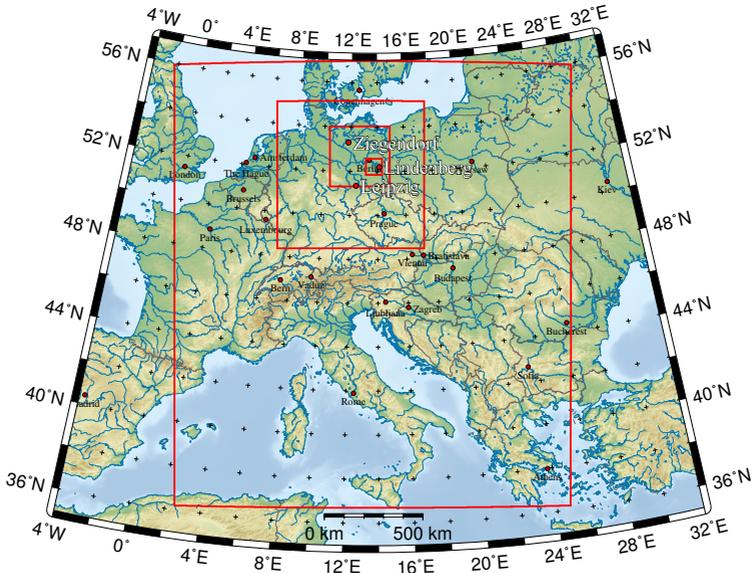


Figure 6.1.1.: The locations of the Raman lidar systems in Lindenberg, Leipzig, and Ziegenhof. The locations of the model domains used are shown as red boxes.

During LAUNCH-2005, data were collected in 7 Intensive Observation Periods (IOPs), each having a duration of 2–3 days. The Raman lidar systems mainly made measurements during nighttime, which resulted in 10 nights of data collected in Germany.

Prior to the supply of all network lidar data to the assimilation centers at the universities of L’Aquila/Italy and Hohenheim/Germany, all water vapor profiles were prepared using the same signal processing software and data quality was ensured with collocated radiosonde launches at each of the lidar sites. This procedure ensured high precision of the water vapor measurements as well as their comparability.

6.2. Observational data

Since the first demonstration of Raman lidar measurements, this technique has become an operational active water vapor remote-sensing technique. This is due to the fact that huge technological advances have been made with respect to the development of high-power laser transmitters operating at suitable wavelengths, optical receiver technology, and detector technology. Details of the set up and water vapor retrievals of state-of-the-art Raman lidar systems can be found e.g. in Whiteman (2003) and Wandinger (2005). In particular, the CART Raman lidar (CARL) located at the Atmospheric Radiation Measurement (ARM) site in Oklahoma, has been demonstrating since the 1990s that routine measurements of water vapor profiles with the Raman lidar technique are possible during daytime and nighttime (Turner et al., 2002).

6.2.1. RAMSES

Since the summer of 2005, the DWD has been running Raman Lidar for Atmospheric Moisture Sensing (RAMSES) at MOL in Lindenberg, Germany. RAMSES supplies quasi-operational, high-precision water vapor profiles both for climate monitoring of the troposphere (e.g. in the frame of the Global Water Vapor Project (GVaP) (WMO, 1999)) and as a reference for the assessment and validation of new ground-based and/or space-borne sensors (e.g. for the European Climate Monitoring Satellite Application Facility (CM-SAF) (DWD, 2000)). Since these main objectives are long-term activities and require continuous operation of the lidar, it was specifically designed for unattended and independent operation. During its test phase, RAMSES was restricted to nighttime measurements. It will later be upgraded with the capability to also measure water vapor during daytime. RAMSES is the second fully operational lidar system in the world after CARL and is equipped with a fully autonomous data retrieval system (Mattis and Jaenisch, 2006).

RAMSES is based on an injection-seeded frequency-tripled Nd:YAG laser with a total pulse energy of up to 1.6 J. Only third-harmonic radiation at 354.7 nm is emitted into the atmosphere. The typical pulse energy at this wavelength, chosen for operational conditions, is 300 mJ. The pulse repetition rate is 30 Hz. The laser beam is expanded tenfold and directed onto the axis of the far-field telescope with three beam-folding mirrors.

The receiving optics is presently optimized for nighttime water vapor measurements throughout the troposphere. RAMSES is operated with two receiver telescopes, with a diameter of 0.2 m for near-range and a diameter of 0.8 m for far-range observations, simultaneously. Two nearly identical receiver boxes for the far-field and the near-field channels are deployed. After beam collimation, dichroic beamsplitters and interference filters separate the elastically backscattered light at 355 nm and the vibrational-rotational Raman signals of water vapor at 408 nm and of nitrogen at 387 nm. The ratio of the corresponding backscattered signals at these wavelengths is already proportional to the water vapor mixing ratio. RAMSES is supplied with a combined analog and photon-counting data acquisition system. In the framework of this study, only signals based on the photon-counting technique were evaluated.

6.2.2. IFT multi wavelength lidar

RAMSES benefited strongly from the in-depth and long-term experience on Raman lidar developments at the Leibniz Institute for Tropospheric Research (IfT) (Mattis et al., 2002; Wandinger, 2005). In Leipzig, the stationary three-wavelength Raman lidar of the IfT contributed to the LAUNCH-2005 campaign. This powerful instrument applies a Nd:YAG laser and emits radiation at 355, 532, and 1064 nm with an overall pulse energy of 1.6 J. The backscattered light is collected with a 1 m telescope. Ten independent return signals are evaluated. Water vapor measurements are performed on the basis of the vibration-rotation Raman signals of nitrogen and water vapor at 387 and 408 nm, respectively, for the primary wavelength of 355 nm. The signals are acquired in the photon-counting mode. Details of the system are given by Mattis et al. (2002).

6.2.3. Ziegendorf

The IfT multiwavelength Raman lidar was set up, together with the IfT Doppler wind lidar, next to the DWD wind profiler site in Ziegendorf. The system has been described in detail by Althausen et al. (2000). Two Nd:YAG lasers and two Ti:Sapphire lasers generate radiation

at 355, 400, 532, 710, 800, and 1064 nm. At the receiver end, the light collected by a 0.56 m telescope is divided into 15 detection channels. The water vapor mixing ratio is determined from the vibration-rotation Raman signals of water vapor and nitrogen at 660 and 607 nm, respectively, which are generated by laser radiation of 532 nm wavelength. The photon-counting technique is applied for the Raman signals.

6.2.4. Operational data analysis

A common data evaluation algorithm developed at IFT (Mattis and Jaenisch, 2006) was used to analyze all data measured in Lindenberg, Leipzig, and Ziegendorf during LAUNCH-2005. The water-vapor-to-dry-air mixing ratio is calculated from the ratio of the water vapor Raman signal to the nitrogen Raman signal (Wandinger, 2005). The calibration factor needed to convert the relative to an absolute measurement is derived by comparing the lidar profile for a selected time and height range with a local radiosonde measurement performed parallel to the lidar observation. At Lindenberg, the on-site operational soundings launched at 0000, 0600, 1200, and 1800 UTC were used. Both IFT lidar systems were equipped with radiosonde stations, so individual profiles for calibration were also available in Leipzig and Ziegendorf. Consequently, the overall accuracy of the Raman lidar water vapor measurements was limited by the accuracy of the radiosondes used for calibration. Extended comparisons demonstrated that the systematic calibration error was of the order of 5 % (Wandinger, 2005).

The statistical measurement error results from the signal noise in the Raman signals and depends on height as well as on spatial and temporal averaging. These noise errors were derived for each profile using the analytical error propagation of photon statistics in the water vapor Raman lidar equation. For details, see Whiteman (2003); Wandinger (2005). Typically, the errors are smaller than 5 % in the lower troposphere and of the order of 10 % to 20 % in the upper troposphere for averaging periods of 15 min to 30 min.

6.2.5. System performance

Figure 6.2.1 shows a typical comparison of Raman lidar and radiosonde water vapor profiles performed on 30 October 2005 using resolutions of 10 min and 67.5 m to 307.05 m, respectively.

The agreement is excellent and no indication was found of systematic errors larger than 0.1 g/kg between 500 m and 7500 m.

These data also demonstrate the large vertical variability of water vapor, which was observed during all IOPs. Several dry layers were found, e.g. at about 3500 m AGL, which corresponded to about 10 % relative humidity, whereas at 6000 m the relative humidity increased to 60 %. These structures changed continuously, confirming the importance of high-resolution observations for model evaluation and water vapor data assimilation.

In figure 6.2.1, the thin red lines around the lidar profiles correspond to the noise error estimates. The error profiles agree with the estimates made above. The errors are about 5 % or 0.1 g/kg, whichever is lower, over the entire range, which can be transferred to other vertical and temporal resolutions using noise error propagation. Based on these results, the following conclusions for the data assimilation efforts could be drawn: No bias correction of the Raman lidar data was considered necessary in this study, as the systematic errors were considerably less than the background errors of the model. The noise error analyses were stable and reasonable and could be ingested in the observation error covariance matrix.

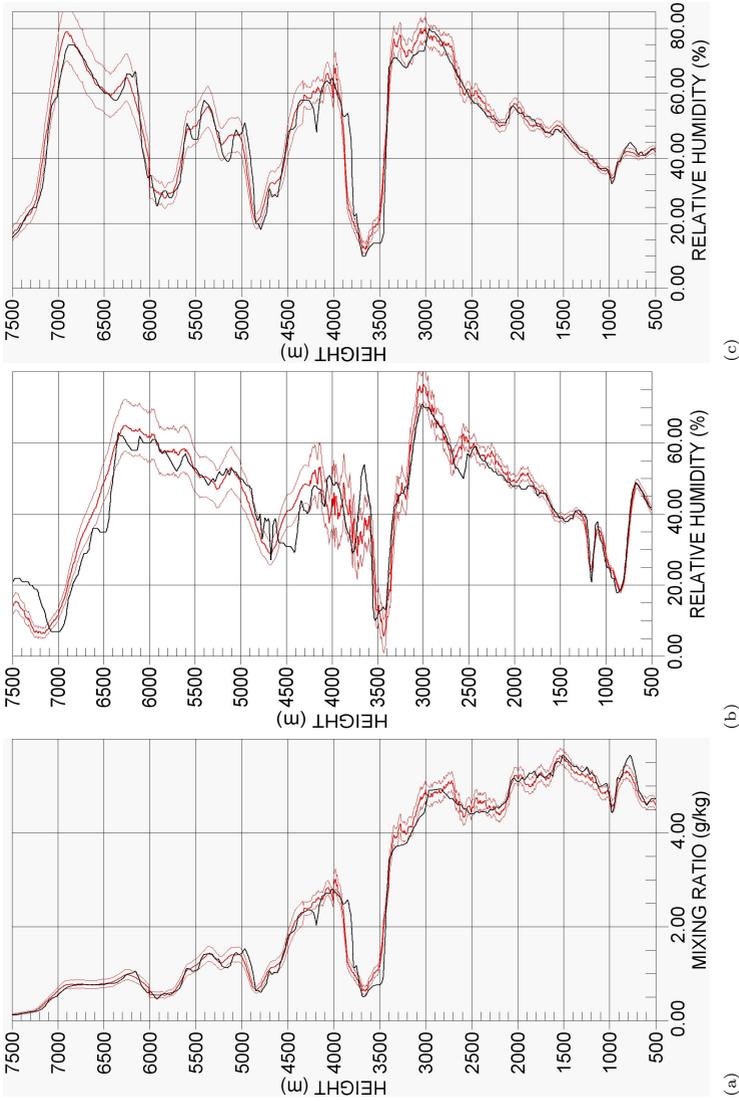


Figure 6.2.1.: Profiles of water vapor mixing ratio (a) and relative humidity (b), (c), observed over Lindenberg, Germany, between 2313 and 2323 UTC on 30 October 2005 (a), (b), and between 0503 and 0513 UTC on 31 October 2005 (c). Bold (red) lines correspond to the lidar profiles, thin (red) lines indicate the measurement uncertainties. Dotted (black) lines show the profiles from the 0000 UTC radiosonde (a), (b) and from the 0600 UTC radiosonde (right). The vertical resolution ranges from 67.5 m to 307.5 m.

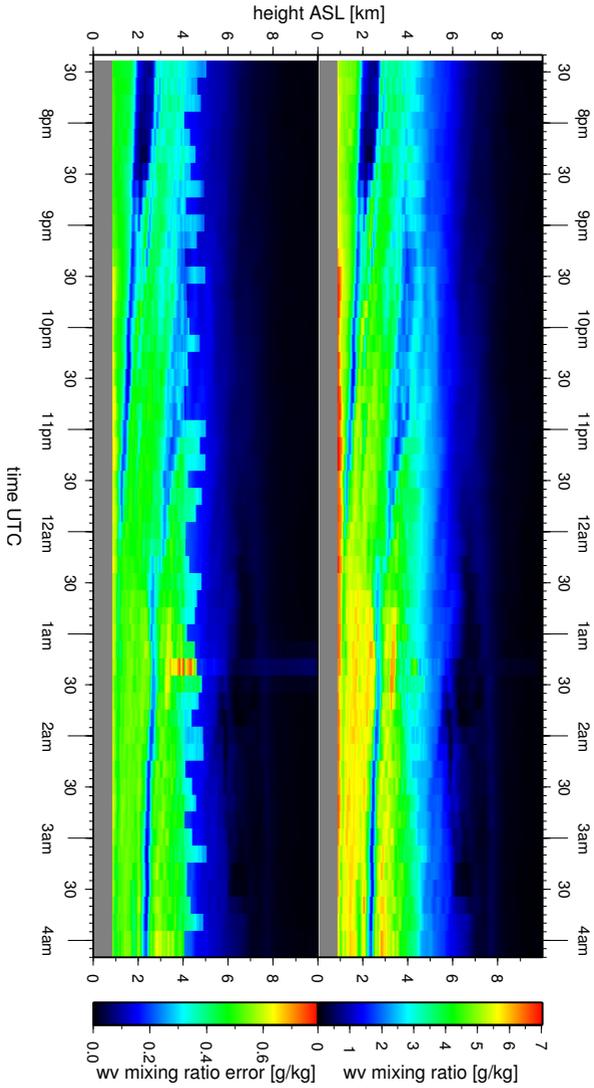


Figure 6.2.2: Time height cross section of the water vapor mixing ratio observed by RAMS/S between 1900 UTC on 26th October 2005 and 0500 UTC on 27th October 2005. The upper panel shows the observed water vapor mixing ratio in a time resolution of 10 min. The lower panel shows the estimated absolute error of the observations.

6.2.6. Temporal evolution of water vapor profiles

All Raman lidars were able to provide high-quality water-vapor profiles up to the middle troposphere. RAMSES had the best SNR with a time resolution of 10 min due to the high average power of the laser transmitter and high efficiency of the receiver. However, the RAMSES measurements were limited to heights greater than 500 m. To obtain similar noise errors, the time resolution of the IFT and Ziegendorf Raman lidar water vapor profiles were reduced to 30 min but they provided water vapor observations down to the ground.

Figure 6.2.2 presents the temporal development of the moisture profiles from sunset to sunrise on 26–27 October 2005 as observed with RAMSES in Lindenberg. Similar vertical resolutions were used as in figure 6.2.1. Noise error profiles are also shown and demonstrate that the noise errors were typically below 0.5 g/kg over the entire range, which sets strong constraints for data assimilation.

The high vertical resolution of the Raman lidar water vapor measurements allowed the detection of thin dry layers up to a height of 8 km alternating with moist layers with relative humidities up to 85 %. During the observation period, subsidence of the dry layers by about 2 km was observed. This was probably due to the effect of the approaching warm front, which glided up to the cold sector of the low over Russia. Furthermore, the approach of the warm front appeared as an overall increase of humidity by about 1 g/kg to 2 g/kg in the lower troposphere, starting at about 0100 UTC, as expected by the weather analysis in section 6.3.

6.3. Synoptic situation

For this study, IOP-7 (26-27 October 2005) was chosen because all high priority systems of the network in eastern Germany were in operation and the synoptic conditions ahead of a sharp frontal rain band approaching from the west were favorable in terms of the campaign aims. Figure 6.1(a) shows the 0000 UTC surface analysis by DWD providing an overview of the synoptic situation. Two low pressure systems were located over the eastern North Atlantic and northwestern Russia. In between the two systems, a northwesterly flow occurred along a boundary that connected the warm front of the western with the cold front of the eastern low. The surface front separated warm and moist air to the west and southwest from drier and cooler air over northern and eastern Germany. During the IOP, the front moved slowly eastwards and diluted during 27 October.

The NCEP analysis of the 500 hPa geopotential height (see figure 6.1(b)) showed that westerly to northwesterly flow was present in the middle troposphere. However, eastern Germany was already located beneath a northward extending ridge that moved slowly to the east. In the region of the ridge, warm air was transported northwards.

The Meteosat Second Generation (MSG) channel 6 image (water vapor, 7.6 μm) shown in figure 6.3.2(a) provides information on the water vapor content at around 500 hPa. Over the southern part of central Europe and over the Alps, south of the warm front, a region of very dry air was observed where very warm air from the Mediterranean and northern Africa was transported northwards into the western flank of the warm sector. Above the Raman lidar sites, the complex structure of the warm front was observed with indications of cirrus clouds and increased humidity. Since the surface front was present over eastern Germany, this suggests that the warm and moist air in the warm sector of the low pressure system approaching from the west was already gliding up onto the cooler and drier air to the east of the surface front. The MSG infrared image (shown in figure 6.3.2(b)) for the same time step showed that large parts of

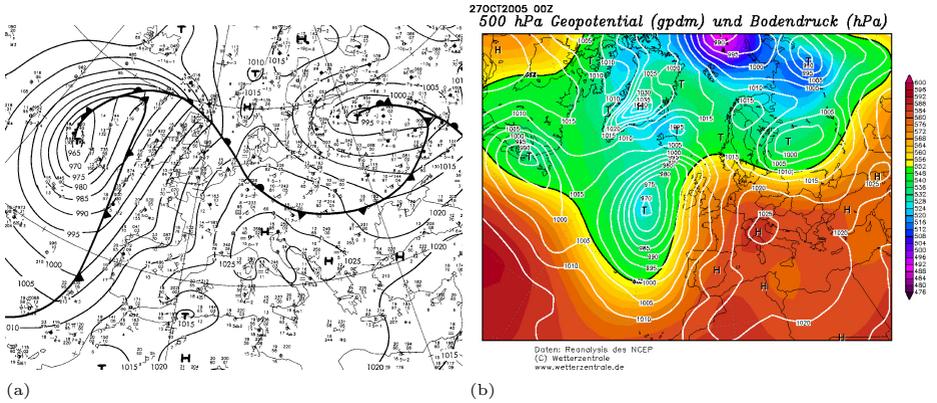


Figure 6.3.1.: Surface analysis at 0000 UTC, 27 October 2005, provided by DWD (a), and 500 hPa geopotential height and surface pressure of the NCEP reanalysis at the same time (b).

central Europe were cloud free or only covered by low-level clouds. In the broad northwesterly flow along the surface front optically thicker clouds with lower cloud-top temperatures were transported into the region of the lidar sites.

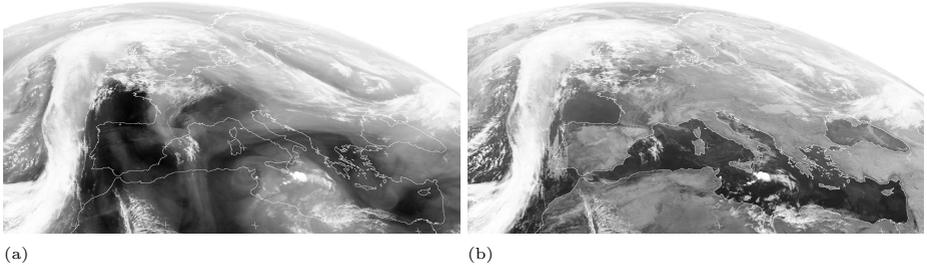


Figure 6.3.2.: MSG water vapor channel (a) and infrared channel (b) at 0000 UTC, 27th October 2005

In such a situation, high values of humidity are expected to occur on the cold side of the front at the surface with a sharp humidity gradient above the boundary layer. With the approach of the warm air and its up-riding on the colder near-surface air, the moist region is expected to rise and thicken with time.

6.4. Minimization

For the assimilation run, a horizontal resolution of 27 km was used in a model domain with 82×74 grid points and 36 vertical layers with thicknesses ranging from 70 m near the surface to about 1 km at the top of the model domain (see figure 5.5.2). The result of the assimilation was an optimal initial condition at the beginning of the assimilation window. This approach was considered valid as the major uncertainties in the initial fields were expected and were largely due to initial conditions and less due to the model physics. In our study, a data assimilation window from 2300 UTC on 26 October to 0200 UTC on 27 October 2005 was chosen.

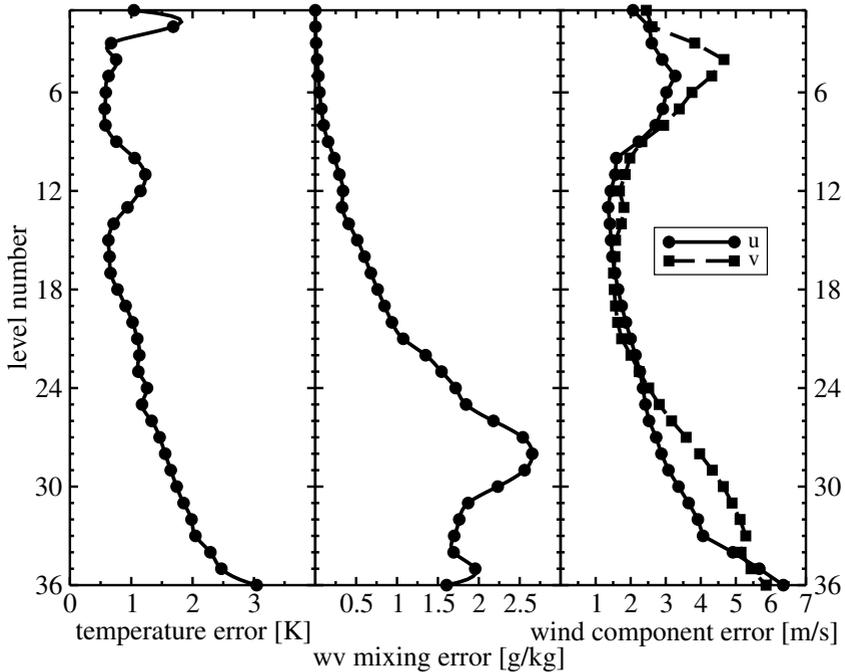


Figure 6.4.1.: Background error estimation for the LAUNCH-2005 assimilation experiment

Figure 6.4.1 shows the vertical error profile used as described in section 4.2.1 and used for the simplified background error covariance matrix.

The observation operator for water vapor Raman lidar measurements is very similar to that described in section 5.4. Since the variable measured by this lidar is the water vapor mixing ratio, which is the same as in the MM5 model, no conversion of water vapor number density to mixing ratio is necessary.

After the assimilation, two forecasts were performed. One was made with the modified initial state from the assimilation and the other with the original initial state provided by the

ECMWF analysis. For the free forecasts, the most sophisticated parameterizations available (see table 4.3.2) and four 2-way interactive nested domains were chosen. All domains had the same vertical structure as the assimilation run. The location of the domains is shown in figure 6.1.1 and their size and resolutions are summarized in table 6.4.1.

Table 6.4.1.: MM5 domain configurations used for the assimilation of ground based water vapor Raman lidar data of the LAUNCH-2005 campaign and subsequent free forecasts.

Resolution	Size
27 km	$82 \times 74 \times 36$
9 km	$82 \times 82 \times 36$
3 km	$100 \times 100 \times 36$
1 km	$82 \times 82 \times 36$

The observation operator \mathcal{H} was the same as for the DIAL case in section 5.4. Since the Raman lidar already delivers the observable of the cost function, namely the water vapor mixing ratio, the conversion from absolute humidity to mixing ratio, necessary during the DIAL data assimilation, was not needed.

As for the assimilation of LASE data, a first impression of the result of the minimization was given by the development of the cost function and the gradient with the iterations. Figure 6.4.2 shows the decrease of the value of the cost function. The behavior expected of a successful reduction of the cost function is visible. The observational part generally decreases, so that the changed initial fields produced a forecast for the assimilation window which fitted the observations better. Only the cost function part of Ziegendorf developed differently. The cost value was already small for the first guess, so it only changed slightly with the number of iterations. The background part of the cost function increased, indicating a change of the forecast with respect to the first guess.

In Figure 6.4.3, 6.4.4, and 6.4.5 the water vapor fields of the control run and the run after data assimilation are compared with the observation itself. The coarse structure of the water vapor field, a humid layer up to about 4 km in Lindenberg and Ziegendorf and 2 km in Leipzig, was present and the model also showed the moistening of the lower troposphere due to the approaching warm front. However, in all cases, the control simulation was not able to reproduce fine structures in the vertical variability of the water vapor field and their strong vertical gradients. The dry subsiding layers, which were particularly visible in Lindenberg and Ziegendorf, were not simulated and the humidity gradients above the humid layer were largely underestimated. These deviations may have been caused by too coarse a vertical resolution of the model.

An indication of biases in the model field was found in the middle troposphere between 2 and 6 km. In all cases, the middle troposphere was slightly too dry. In Leipzig, the model PBL was too humid whereas it was too dry in Ziegendorf. In any case, the data indicated that both biases in the model field were significant and the vertical distribution of water vapor could deviate strongly from reality.

The improvement in the initial fields after data assimilation was substantial. In all cases, the vertical structure and gradients in the water vapor fields were corrected. Furthermore, the above mentioned biases in the middle troposphere and deviations in the PBL were strongly reduced.

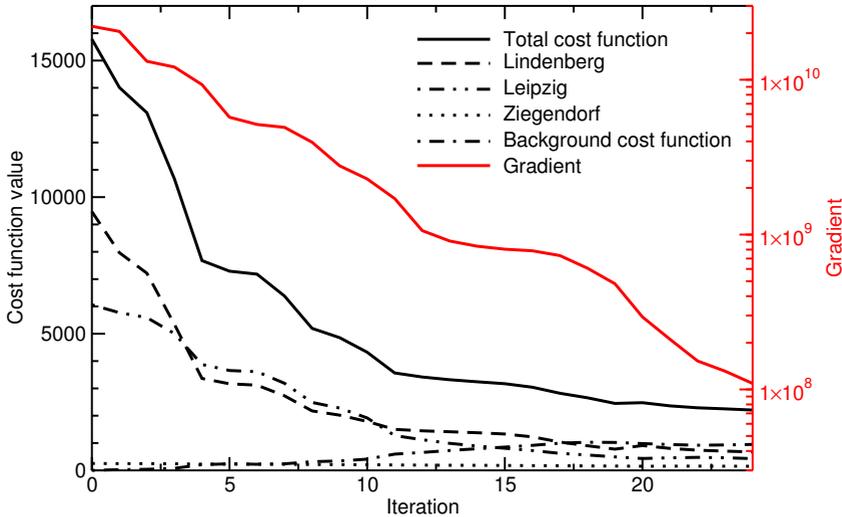


Figure 6.4.2.: Decrease of the cost value with increasing iteration number (black). Also shown is the decrease of the absolute value of the gradient (red).

The small impact in the region of Ziegendorf can be explained by figure 6.4.2. The cost value at iteration step zero was already small for Ziegendorf and, during the 4D-Var, this value changed only slightly. Therefore, the first guess from the ECMWF analysis was close to the observation and the 4D-Var had no reason to change the initial field of the model for this region.

6.5. Impact on the fields

Figure 6.5.1 shows horizontal differences of the mixing ratio, temperature, and wind for 850 hPa and 500 hPa heights in the initial state at the beginning of the assimilation window at 2300 UTC, 26th October 2005. This is the time step that was changed by the 4D-Var run. In the region of Lindenberg, the amount of moisture increased at 850 hPa and decreased slightly at 500 hPa, whereas the moisture in Ziegendorf and Leipzig were hardly influenced.

A strong impact was induced on the water vapor field in a region even larger than the Raman lidar triangle. From the lower to the middle troposphere, a redistribution of moisture of up to 1 g/kg occurred. The vertical impact also reached the 500 hPa level.

The shape of the impact region mainly indicates that the water vapor content upstream of the Leipzig and Lindenberg Raman lidar system was reduced. This is consistent with the prevailing wind direction from the northwest. Obviously, the water vapor field was adjusted by the modification of the water vapor, which was advected to the observation sites during the assimilation window. The extent of the impact region is promising, as it shows that a coarse grid of three Raman lidars can produce a nearly homogeneous impact over an area of 200 km \times 200 km.

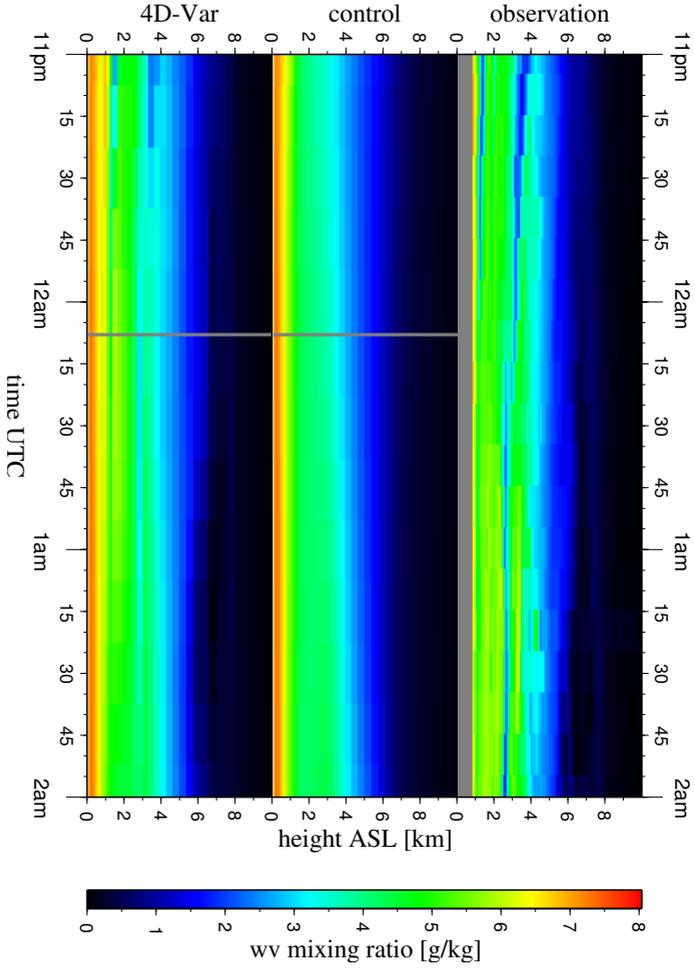


Figure 6.4.3.: Comparison of the assimilated observations with forecasts with 4D-Var. Upper panel: Observations of RANSES, middle panel: forecast initialized with ECMWF analysis, lower panel: forecast from the initial state modified by 4D-Var.

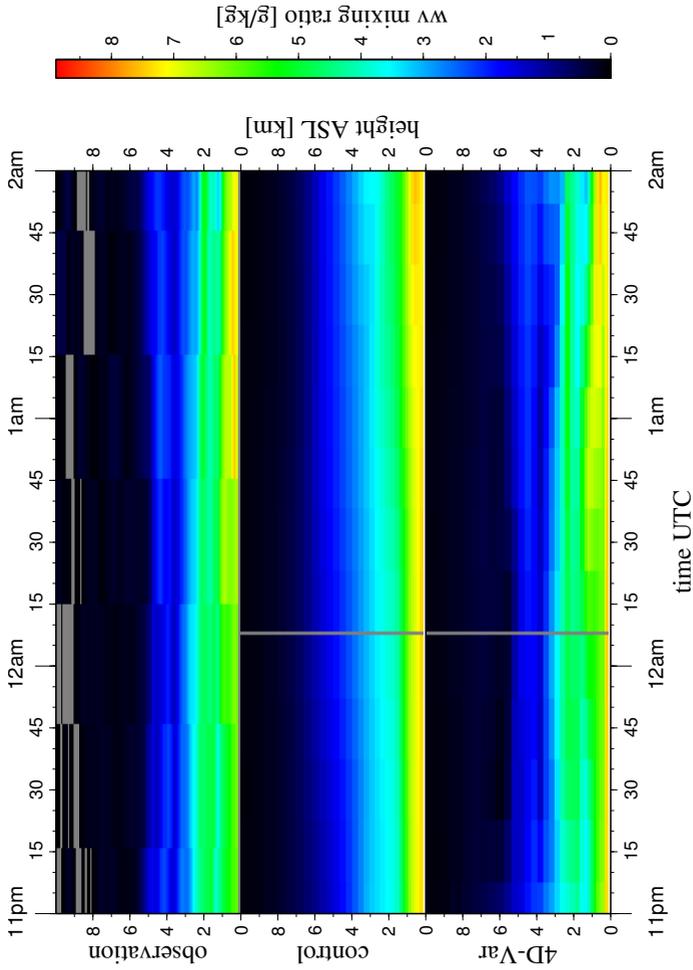


Figure 6.4.4.: Same as figure 6.4.3 but for the Leipzig Raman lidar.

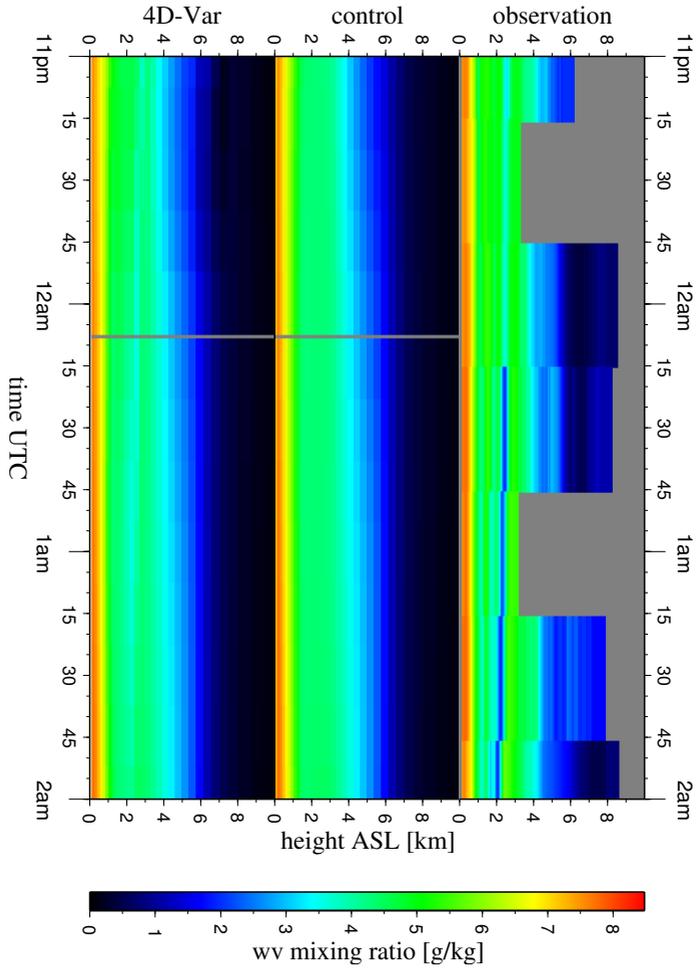


Figure 6.4.5.: Same as figure 6.4.3 but for the Ziegendorf Raman lidar.

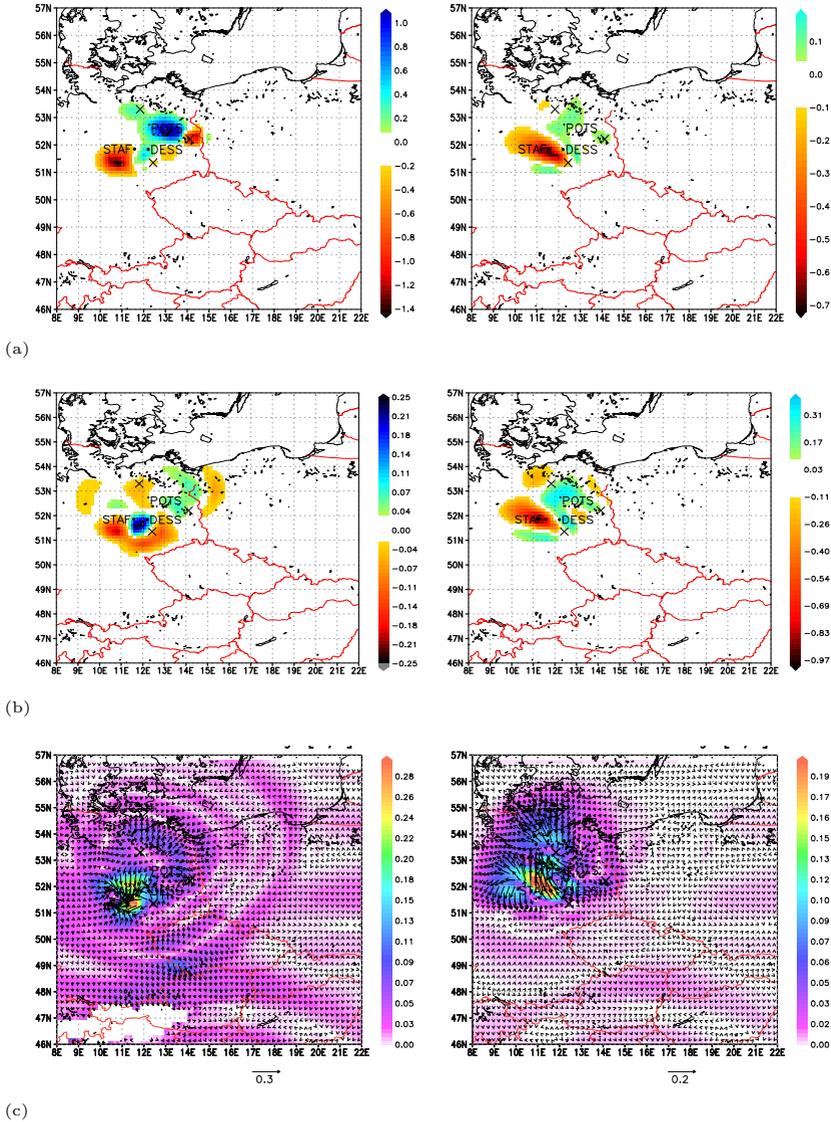


Figure 6.5.1.: Difference of water vapor mixing ratio (g/kg) (a), temperature (K) (b), and wind velocity (m/s) (c) (4D-Var – CONTROL) for different levels of the troposphere at the beginning of the assimilation window. The three lidar stations are marked with crosses. The left column shows the difference at 850 hPa, while the right column represents the middle troposphere at 500 hPa. The GPS stations Potsdam (POTS), Dessau (DESS), and Stauffurt (STAF) used for validation are also indicated.

Although only water vapor observations were assimilated, the temperature was also influenced. Figure 6.5.1(b) shows temperature changes of -1 K to 0.5 K in both 850 hPa and 500 hPa heights. This illustrates the dynamical coupling of the humidity and temperature fields in the assimilation system.

Also studied was the change in the wind field due to the assimilation (see figure 6.5.1(c)) The wind field was even more strongly correlated in space. Although the maximal magnitude of the impact in the wind field was small, with 0.3 m/s at 850 hPa height and 0.2 m/s at 500 hPa height, the area affected was much larger.

6.6. Quantitative description of the impact

To describe the impact of the assimilation more quantitatively, an impact density of a model variable γ is introduced by

$$\varrho_\gamma(\vec{x}) = \frac{\Delta\gamma(\vec{x})^2}{\int \Delta\gamma(\vec{x})^2 dV} \quad (6.6.1)$$

where $\Delta\gamma$ is the difference between the cases with and without assimilation of the variable γ , i.e. $\Delta\gamma = \gamma_{\text{4D-Var}} - \gamma_{\text{Control}}$, and \vec{x} denotes the location in the three dimensional space. The difference $\Delta\gamma$ is squared, to make the density definitely positive, and the denominator is the integral over the whole volume, to fulfill the property of a density $\int \varrho_\gamma(\vec{x}) dV = 1$. This density describes the impact density at a location \vec{x} . Furthermore, the expected value according to this impact density is defined by

$$\langle r \rangle_\gamma = \int r \varrho_\gamma(\vec{x}) dV \quad (6.6.2)$$

With this expected value, various measures can be defined to describe the impact of the assimilation:

1. The expected value of the location describes the center of the impact.

$$\text{center} = \langle \vec{x} \rangle_\gamma \quad (6.6.3)$$

2. The spatial volume expansion of the impact region can be measured by:

$$\text{width} = \sqrt{\langle \vec{x}^T \vec{x} \rangle_\gamma - \langle \vec{x}^T \rangle_\gamma \langle \vec{x} \rangle_\gamma} \quad (6.6.4)$$

3. And a measure of the impact strength is:

$$\text{strength} = \langle \gamma \rangle \quad (6.6.5)$$

The theoretical maximum width would be a constant impact in the whole domain, i.e. the impact density function would also be constant, but zero outside the domain. In this case, the width can be calculated analytically as $\frac{L}{2\sqrt{3}}$, where L is the edge length of the model domain. For the coarsest domain, the theoretical maximum impact width is then ≈ 620 km.

In the lowest left panel of figure 6.6.1, the width of the impact area of different model variables is shown for the coarsest domain. The impact width for water vapor starts from about 100 km at analysis time and grows slowly during the first 10 h of the forecast. Later,

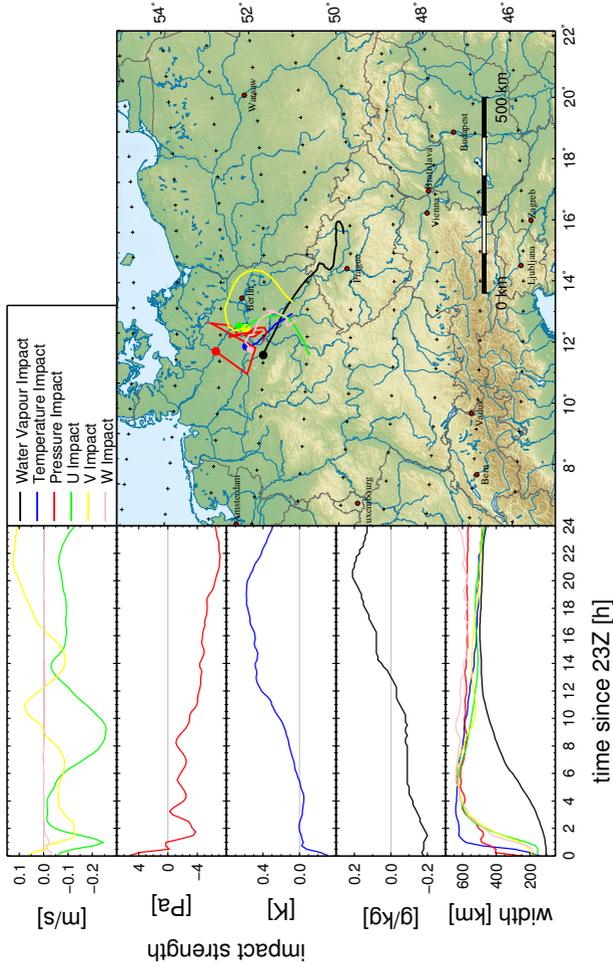


Figure 6.6.1.: Shown are the measures, defined by equation (6.6.3), (6.6.4) and (6.6.5) for the outermost domain with a resolution of 27 km. In the lowest left panel, the widths of the impact region for the model variables water vapor mixing ratio, temperature, pressure and wind are shown. The upper four panels show the impact strength of those variables. In the right panel, the path of the impact center is shown for the different variables. The paths of the water vapor impact are shown for the first 14 hours of model integration, while the paths of the pressure and wind impact are shown for the first four hours and the path of the temperature impact is shown for the first hour. The starting point for each line is marked with a dot.

the impact width remains nearly constant with values of around 400 km. The impact width of the wind components starts at about 200 km and quickly reaches the theoretical maximum after about four hours. The pressure impact starts with a larger width of about 300 km and increases as fast as for the wind. The temperature impact broadens most quickly, to the whole model domain within about two hours.

In the upper four panels on the left of figure 6.6.1, the impact strength for the chosen variables is shown. The time of broadening to the maximum width changes rapidly. This is clearly seen for the temperature and the pressure, while the impact strength is zero for the temperature and the pressure has a negative impact strength of about 4 Pa after the broadening of the impact region. After about 8 hours for the pressure and about 4 hours for the temperature, the impact strength changes again.

Since the width of the temperature and pressure impact are already maximal, a constant impact density has to be assumed and therefore also a constant $\Delta\gamma(\vec{x})$. In this case, equation (6.6.5) suggests that $\langle\gamma\rangle_\gamma$ is equal to $\Delta\gamma$. If a maximum possible impact width is assumed at 20 hours after analysis time, the temperature is globally increased by 1 K, the pressure is decreased by 10 Pa and the water vapor mixing ratio is increased by 0.3 g/kg.

The center of the impact is of course only meaningful as long as the impact width has not reached its maximum. In the right panel of figure 6.6.1, the center of the impact region from analysis time to the time of maximal broadening is shown. The movement of the impact in the water vapor field is clearly visible with advection during 14 hours.

It is possible that the long-term impact, visible in the impact strength, is related to the independence of the impact with respect to the initial perturbation, as shown by Hohenegger and Schär (2007). In this study, an impact on the temperature, with a magnitude of about 1 K, was found after an integration time of 11 hours, and was independent of the initial perturbation. Here, a global impact of about 1 K was found after an integration time of 20 hours. The difference of time scale can be explained by the much coarser resolution of 27 km in this study as compared to 2.2 km in the experiments of Hohenegger and Schär.

6.7. Validation of the impact with GPS ZWD observations

If high-accuracy observations are available, they strongly constrain the resulting water vapor field after data assimilation. The 4D-Var system minimizes the cost function by combined changes of the water vapor, wind, and temperature fields. It can be expected that a minimization of the cost function will always be possible if the observations do not deviate too much from the control fields.

However, the question arises as to whether the 4DVAR results are in better agreement with reality than the CONTROL simulation. This is not necessarily the case considering the limitations of the 4D-Var system, namely diagonal error covariance matrices, coarse horizontal resolution, and simplified physical parametrizations and a change of the model physics from the minimization to the free forecasts. Furthermore, the impact area and number of observations were limited. Therefore, it is reasonable to validate the 4D-Var run using an independent water vapor data set having good accuracy and coverage. To do so, data from the German GPS network was used, which is well-suited to this purpose (Gendt et al., 2004).

The GPS stations determine the Zenith Wet Delay (ZWD) according to the equation

$$ZWD = 3.37 \times 10^{-1} \frac{\text{K}^2}{\text{hPa}} \int_{\text{ground}}^{\text{top of model}} \frac{e}{T^2} dh, \quad (6.7.1)$$

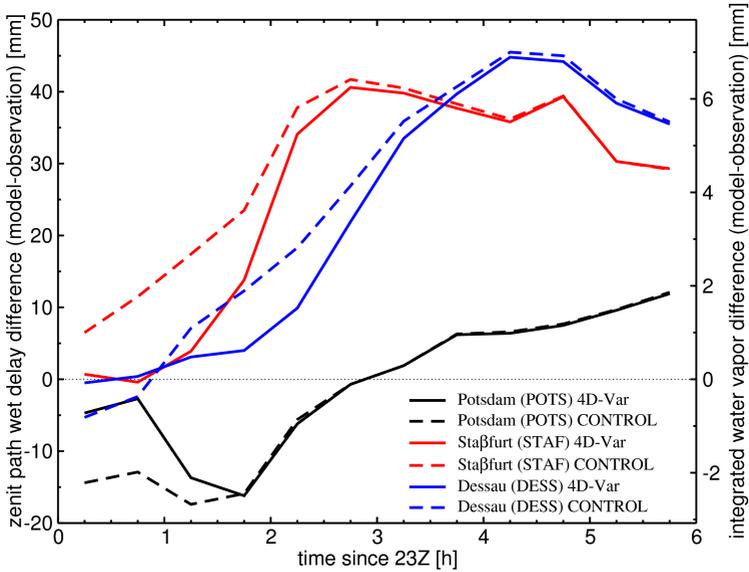


Figure 6.7.1.: Differences between observed and modeled GPS ZTD for the three GPS stations of Potsdam (POTS), Dessau (DESS), and Staßfurt (STAF). Dashed lines are the differences between the observations and the forecast from the ECMWF first guess (CONTROL). Solid lines are the differences with the forecast initialized by the 4D-Var.

It is thus proportional to the integrated water vapor in the column. In this equation, e is water vapor pressure, T is temperature, and h is height. Consequently, if strong changes in water vapor profiles occur during the assimilation process, this is reflected in ZWD . The observed zenith wet delay ZWD_{obs} was determined at some GPS stations close to the impact region, namely at Potsdam (POTS), Dessau (DESS), and Staßfurt (STAF), which are also indicated in figure 6.5.1. Then we calculated the differences between the observed ZWDs and ZWDs derived from the model fields using a forward operator developed for the MM5 4D-Var system (Zus et al., 2008). For this purpose, we used equation (6.7.1) and evaluated the integral via the trapezium rule from the model levels. This forward operator for GPS observation was also used in a generalized form for the operational assimilation of GPS Slant Total Delay (STD) during COPS/D-PHASE.

Figure 6.7.1 presents the results. It clearly shows better agreement of the 4DVAR simulation with the observed wet delay during the time when the region of influence of the assimilation covers the location of the GPS measurements. This impact was reduced after several hours because the impact region was advected away from the GPS stations with the ambient flow. This demonstrated not only an impact of Raman lidar data assimilation but also the fact that the results agreed much better with reality. Unfortunately no GPS stations were available for Poland and the Czech Republic to validate the impact of the assimilation later in the forecast. (Grzeschik et al., 2008)

7. Summary and conclusion

The aim of this work was to study the impact of lidar observations on data assimilation and possible improvements in NWP. Two cases were selected to investigate the influence of an airborne as well as a ground-based network of lidars. For the data assimilation studies, the 4D-Var was chosen as the assimilation scheme since it allowed the assimilation of asynchronous time series of observations while information from these observations were propagated not only to the future, as in sequential assimilation schemes like the Kalman filter and ensemble Kalman filter, but also to the past of the model state. This allowed a forecast to be consistent with the model physics even in the assimilation window.

The implementation of the 4D-Var scheme in the framework of the mesoscale NWP model MM5 has been described, including details of the calculation of the cost function, its gradient and the necessary generation of the adjoint of the model. Also, the steps necessary for the estimation and simplification of the error covariance matrices have been shown. Although the background error covariance matrix was limited by implementation to a diagonal matrix, the time propagation of the 4D-Var permits the use of such a simplified background error covariance matrix. Last but not least, the implementation of the observation operator for lidar has been described

To investigate the influence of lidar data on the performance of NWP, two case studies were investigated in detail. The first case studied the impact of the assimilation of airborne water vapor dial data for one IOP of the IHOP_2002 campaign. During this case, the downward pointing LASE water vapor DIAL mapped the water vapor field in a sensitive region of an upcoming convective event. Apart from the result that the 4D-Var was working successfully, the modification of the initial state by the assimilation system was investigated. Although only water vapor data were assimilated, the initial state of the forecast also changed the temperature and wind fields. The impact on model variables, not directly assimilated, is explained by the backward propagation in time of the observed information, which was performed in a way consistent with the model physics in the 4D-Var.

It was shown that the initiation of convection was clearly modified by the assimilation. The wind field was changed in such a way that a convergence line was shifted. This convergence line was moved from an area with high CAPE to a region with a much smaller potential to initiate convection. The resulting suppressed initiation of convection was in agreement with the observed radar reflectivity.

The second assimilation study investigated an IOP of the LAUNCH-2005 campaign, where three ground based, vertically pointing water vapor Raman lidar systems were measuring simultaneously in northeastern Germany. The three lidar systems were arranged in a triangle with a side length of about 170 km. The capability of the 4D-Var to include a layer structure in the water vapor field with alternating dry and moist layers, which was missing in the control simulation, was shown.

To describe the impact of the assimilation and its time evolution in the following free forecast, measures were introduced to characterize the location, the extent, and the strength of the impact. Investigations on the time evolution of the impact measures showed a variable spatial

broadening of the impact for the different model variables. The temperature impact reached maximal possible broadening of the model domain within two hours, pressure and wind variables broadened in four hours to the maximal width, and the water vapor impact was distributed over the whole model domain within ten hours. The impact center of water vapor was advected from northwest to southeast with the large-scale flow. Due to the fast broadening of the temperature, pressure, and wind impacts, their impact center could not be successfully tracked.

To investigate whether the assimilation of the Raman lidar water vapor profiles improved the forecast, a comparison was carried out with GPS ZWD data as independent observations. It showed better agreement of the integrated water vapor of model and reality after assimilation.

Both experiments showed a positive impact on the representation of the water vapor field in numerical weather prediction when data from state-of-the-art lidar systems were assimilated. Also the modification of other model variables, which were not observed, could be shown. The modification of the wind field, although the water vapor mixing ratio was assimilated, was the main trigger for improving the forecast of CI in the IHOP_2002 experiment.

In this work, the impact of lidar observation in comparison to observations already in operational use was not discussed. A conceivable continuation of this work would be an assimilation study investigating the possible information content of lidar observations for assimilation in comparison to satellite observations, GPS, and conventional in-situ measurements.

Based on the three ground based lidar systems from the LAUNCH-2005 experiment, Observation System Simulation Experiment (OSSE) experiments with a synthetic network of lidar systems are interesting to find optimal configuration for assimilation of a possible network. Also OSSE experiments could investigate the impact of the Water Vapour Lidar Experiment in Space (WALES). Furthermore investigations should be made into the extent to which 4D-Var is necessary or assimilation schemes with specific analyses times, where observations are accepted, are sufficient. Further, an extension from water vapor to wind observations, like radial wind from doppler radar, and their combination would be very interesting.

A. Process and verification studies using the D-PHASE ensemble

A.1. Implementation

For the World Weather Research Programme (WWRP) Research and Development Project (RDP) COPS field campaign (Wulfmeyer et al., 2008) and the WWRP FDP Mesoscale Alpine Programme (MAP) D-PHASE (Rotach et al., 2008), a plotting system was developed that provided a common way to generate visualizations from a large number of models. 24 deterministic models and 7 probabilistic models participated in the D-PHASE FDP. The tool to provide the same plots from all of these models was impelled by the idea of comparing all the models. For ease of comparison, it was considered necessary to define the same verification domains, variables, and color tables. A major step was already taken by Andrea Montani (Arpagaus, 2007) in defining a common GRIdded Binary (GRIB) format for all D-PHASE models. With this prerequisite, it was possible to set up a chain of programs which were used by each D-PHASE group on their computing environments. The selected implementation of the whole system was, as far as possible, compatible with Portable Operating System Interface (POSIX) to enable easy installation on the common computing systems.

The main program used was the Grid Analysis and Display System (GrADS) Version 1.9b4. GrADS is a combination of plotting software and a scripting language to control the plotting software. GrADS has been implemented worldwide on a variety of commonly used operating systems and is freely distributed over the Internet.

The GRIB files were parsed by the freely available programs *wgrib* and *grib2ctl* and an index file was created, which allowed GrADS to directly access the GRIB files without conversion to another data format.

For each of the necessary plot series documented in the D-PHASE implementation plan, a program, written in the GrADS scripting language, was developed. These programs created horizontal plots of different variables, precipitation plots with different accumulation times, plots of the probability of exceeding critical thresholds in precipitation for ensembles, meteograms at 87 locations, and 32 cross-sections.

The GrADS software permits different map projections of the imported data. Internally, the data are interpolated on to an equally spaced, orthogonal lat/lon grid and a set of possible input projections are available. Some of the D-PHASE models used a rotated lat/lon grid, which was originally not supported by GrADS. This projection was implemented in GrADS. Furthermore, this change was provided to the developers of GrADS and is now implemented in the current 2.0.a3 version.

Furthermore a possibility was added of starting the system several times on several Central processing units (CPUs), not recreating already processed plots, to reach a possibility of accelerating the execution of the plotting package with this simple but, for this case, sufficient kind of parallelization.

The system was developed during the preparation phase of D-PHASE and the future users

were able to test it in an early state. During this testing, it was easy to adjust the system to the users' needs and also to find problems in the data format, file name conventions or data organization for each group. The D-PHASE groups were supported by email in the installation and execution of the system.

To start the package, the GRIB files are required to follow a pre-defined naming convention. Then the whole system is started with the following command:

```
usage: run.sh -d<datestring> -m<model> [-e<ensemble>] [-C] [-D] \
      [-R | -E] [-t] [-p] -g<grib_directory> [-a]
```

```
-d<datestring>      : initial date of the forecast
                    (form: 'YYYYMMDDHH')
-m<model>           : the model identifier string
-e<ensemble>       : the ensemble member
-g<grib_directory> : directory of grib files
-C                 : CDOM grib file is present
-D                 : DDOM grib file is present
-R                 : use special plot creation list for
                    driving models
-E                 : use special plot creation list for
                    ensembles
-a                 : use analyzed orography field instead of
                    fixed
-t                 : do not remove temporary files
-p                 : the script is executed as part of a
                    parallel run
```

A set of parameters has to be provided, such as the initial date of the forecast, the name of the model, and, optionally, the ensemble member number and the directory of the GRIB files. The list of plots to be generated is controlled by the options -C to create COPS plots, -D for the D-PHASE high-resolution model plots, -R for the D-PHASE driving models, and -E for the plot series of probabilistic models. With the -a switch, the orography data are taken from a different file, as was necessary for the Canadian CMC GEMH model.

A.2. Example plots from COPS IOP9c

As an example, some plots from the COPS IOP9c (2007-07-20) are shown here. This IOP covered a frontal zone oriented from southwest to northeast over the COPS region. As it passed over the COPS region, the frontal zone weakened. While ahead of the zone, east of the Black Forest and Swabian Alps CI took place (see figure A.2.1) and the whole system developed to a squall line again, ranging from the Netherlands down to central Germany.

During IOP9c, MCS related to a vorticity maximum at the east side of a jet, initiated over middle eastern France at about 0600 UTC, propagated northeastwards and reached the COPS region at 0845 UTC. After the passage of the MCS over the COPS region, which resembled that of IOP9b (slowdown over Rhine valley, transformation over COPS region), this time, severe CI took place over the southeastern Black Forest and Swabian Alps, merging with the northern part of the MCS. The whole system developed again in a bow-like structure, which could be

called a squall line, ranging from the Netherlands down to central Germany. It is interesting to note that additional cells were triggered in Bavaria at around 1400 UTC. These cells prevailed until 2200 UTC whereas the other part of the MCS had already moved to eastern Europe.

A surface low-pressure system is expected to move from central France over the Benelux towards the North Sea. Its cold front should cross the COPS area between 1500 and 1800 UTC. Ahead of this system, warm humid air is present.

Low clouds or fog in the valleys will disappear in the morning hours as surface heating increases, followed by quite fair conditions. From around noon, cumulus cloud development is expected eventually leading to showers/thunderstorms during the afternoon. Given CAPE and wind shear, they could be severe in places. (see science director summary¹ and weather summary²)



Figure A.2.1.: DWD radar reflectivity composite from 1041 UTC.

In figure A.2.2 the forecasts of four D-PHASE models are shown. The models are the CM-

¹http://www.cops2007.de/html/daily_reports/sd/sd20070720.pdf

²http://www.cops2007.de/html/daily_reports/ws/ws20070720.pdf

CGEMH, a high resolution version of the CMC Global Environmental Multiscale (CMCGEM) model; MESONH2, a 2 km resolution configuration of the Meso-nh model of the French research community; MM5_2_CT, the MM5 model at 2 km resolution appropriated by IPM, and COSMOCH2, an adapted version of the Consortium for Small-scale Modeling (COSMO) model of MeteoSwiss at a resolution of 2.2 km. For each model, the time step where CI appeared is shown. Table A.2.1 lists the times and locations of CI for each model.

Table A.2.1.: Location and time of CI forecast by the different models.

Model	Time	Longitude	Latitude
CMCGEMH	1215 UTC	8.4 °E	48.4 °N
MESONH2	1230 UTC	8.5 °E	48.9 °N
MM5_2_CT	1245 UTC	8.6 °E	49.1 °N
COSMOCH2	1330 UTC	9.4 °E	48.6 °N

CI is visible in all the four variables shown in the plots. The upper row shows the 2 m temperature where the location of the cold front as well as cold air outflow at the position of the CI are visible. The next row shows the 10 m wind as barbs, and gusts in colors. At the position of the CI, the strongest gust and a discontinuity in the wind field marks CI. In the third row, the high level cloud fraction is shown, with clouds that have grown at the location of CI. The 15 min accumulated precipitation is shown in the last row, where convective precipitation appeared at the location of CI.

Although several CI events are visible in the radar image (see figure A.2.1) over the central Black Forest at about 1030 UTC, CMCGEMH simulates a single event in this region, CI in MESONH2 is located in the northern Black Forest, MM5_2_CT shifts CI to the north of the Black Forest, and COSMOCH2 develops convection over the Swabian Alps. All the models have a time delay of about two hours.

Figure A.2.3 and A.2.4 show vertical cross sections of the forecasts at the same times as above. Upward transport of moisture is visible for all the models at the location determined for CI. For MM5_2_CT and COSMOCH2, high CAPE values are forecasted, while CIN is very low at this location. All the models except the MESONH2 show precipitation in the region of interest. The cross section for this model probably missed the exact location.

By using these tools, it was possible to generate the plots for the whole period of D-PHASE and COPS and all the D-PHASE models. Plots of all the variables, every 15 min for the COPS area and every 1 h for the D-PHASE area, were created by the data providers of D-PHASE and stored in the Climate and Environmental Retrieving and Archiving (CERA) database at the World Data Center for Climate (WDCC) together with the data. In total, about 4000 Portable Network Graphics (PNG) images were produced for each day and model.

A detailed comparison of the different models is possible with this database of images. All forecasts are visualized with the same plotting domain, the same cartographic projection, and the same color scales. So, this tool makes an important contribution to the fulfilment of the COPS and D-PHASE goals by supporting the discovery of weaknesses in numerical weather forecast models and improving prediction of convection and precipitation.

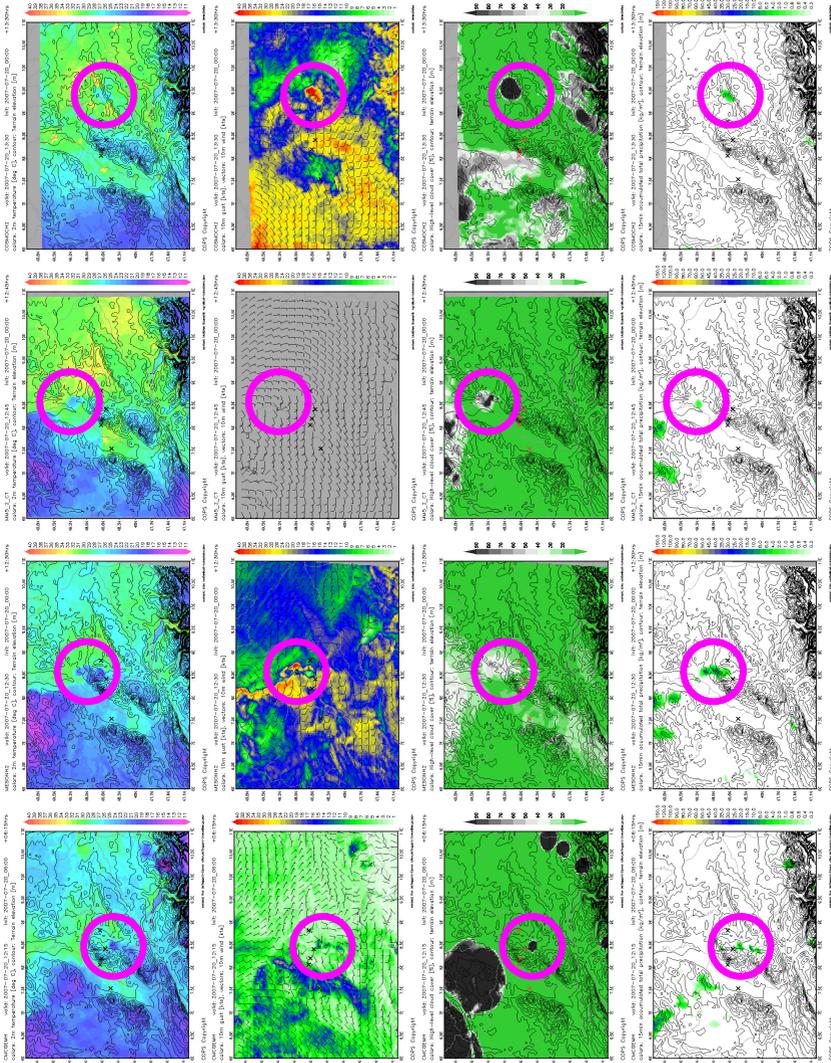


Figure A.2.2.: Forecast of the models CMC GEMH (1215 UTC), MESONH2 (1230 UTC), MM5_2-CT (12:45 UTC), and COSMOCH2 (13:30 UTC) in the rows from top to bottom the 2 m temperature, 10 m wind and gust as colors, cloud fraction of high level clouds, and the accumulated precipitation of the last 15 min are shown. The gust for the MM5_2-CT model is missing, since this variable was not provided. Marked are the locations of CI ahead of the cold front.

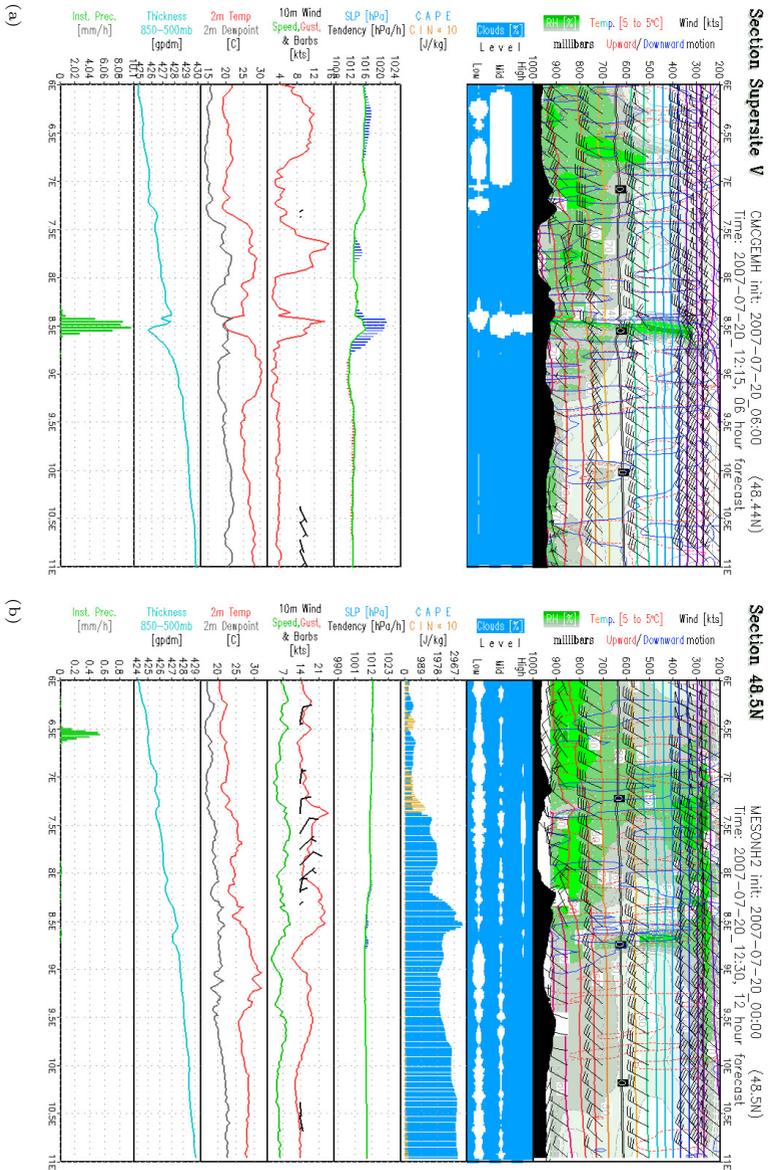
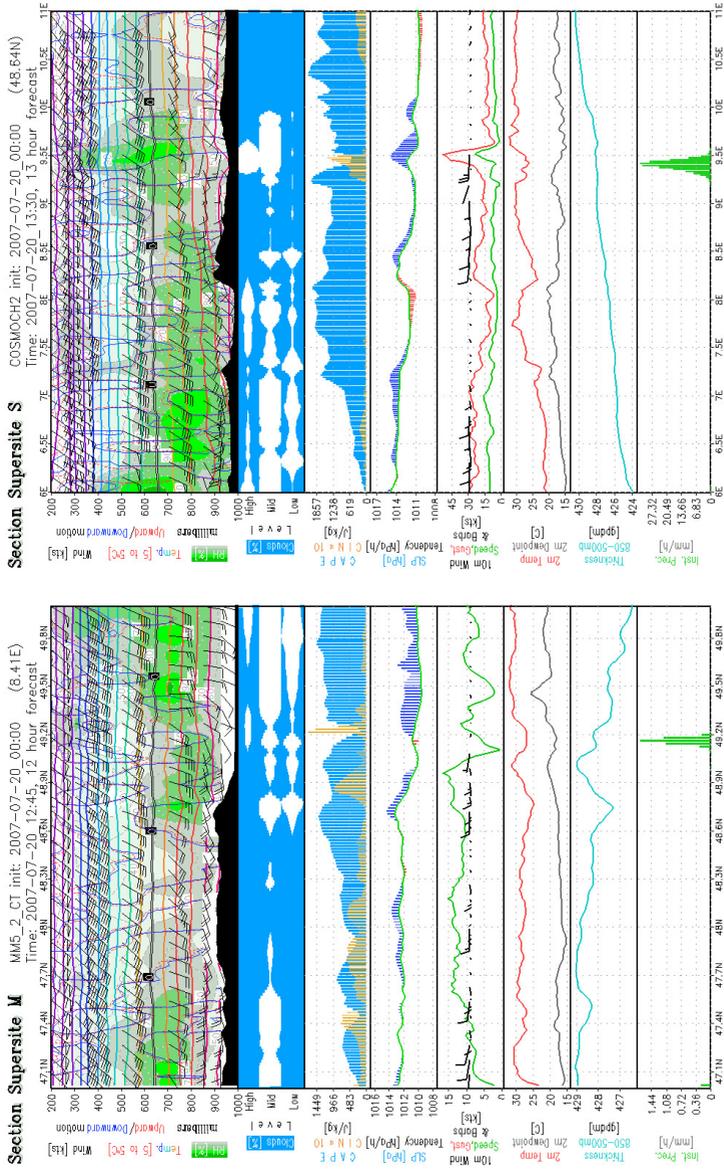


Figure A.2.3.: Vertical cross sections from west to east of the models CMCGEMH (a) and MESONH2 (b). The top panel shows a vertical cross-section of the humidity, wind, temperature and up- and downward motion. The other panels show, from top to bottom, the cloud cover at different heights, forecasted values for CAPE and CIN, sea level pressure, 10m wind speed, 2m temperature and dew point, the thickness of the layer between 850 and 500 hPa, and the 15 min accumulated precipitation. The CAPE/CIN panel is missing for CMCGEMH since these values were not provided by this model.



(a) (b)
 Figure A.2.4.: Same as figure A.2.3 but for the models MM5_2-CT and COSMOCH2. For MM5_2-CT, a cross section from south to north is shown, cutting through the location of CI.

Bibliography

- Althausen, D., D. Müller, A. Ansmann, U. Wandinger, H. Hube, E. Clauder and S. Zörner, 2000: Scanning 6-wavelength 11-channel aerosol lidar. *J. Atmos. Oceanic Technol.*, **17**, 1469–1482.
- Anthes, R. A. and T. T. Warner, 1978: Development of hydrodynamic models suitable for air pollution and other mesometeorological studies. *Mon. Wea. Rev.*, **106**, 1045–1078.
- Arpagaus, M., 2007: MAP D-PHASE Implementation Plan. Technical report.
- Bauer, H., H.-S. Bauer, V. Wulfmeyer, M. Wirth, B. Mayer, G. Ehret, D. Summa and P. Di Girolamo, 2004: End-to-end simulation of the performance of WALES: Forward module. In: *Proceedings of the 22nd International Laser Radar Conference*.
- Bergthorsson, P. and B. Döös, 1955: Numerical weather map analysis. *Tellus*, **7**, 329–340.
- Bjerknes, V., 1911: *Dynamic meteorology and hydrography. Part II. Kinematics*. Carnegie Institute, Gibson Bros., New York.
- Bösenberg, J., 1998: Ground-based differential absorption lidar for water vapour profiling. *Methodology. Appl. Opt.*, **37**, 3845–3860.
- Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods. Technical report, Meteorological Training Course Lecture Series, ECMWF.
- Browell, E. V. and S. Ismail, 1995: First lidar measurements of water vapor and aerosols from a high-altitude aircraft. Technical report, Proc. OSA Optical Remote Sensing of the Atmosphere, Salt Lake City, Utah. OSA Tech. Digest 2, 212.
- Browell, E. V., S. Ismail and W. B. Grant, 1998: Differential absorption lidar (DIAL) measurements from air and space. *Applied Physics B: Lasers and Optics*, **67**, 399–410.
- Browell, E. V., S. Ismail, W. M. Hall, A. S. Moore Jr., S. A. Kooi, V. G. Bracket, M. B. Clayton, J. D. W. Barrick, F. J. Schmidlin, N. S. Higdon, H. S. Melfi and D. N. Whiteman, 1997: LASE validation experiment. In: *Advances in Atmospheric Remote Sensing with lidar*, A. Ansmann, R. Neuber, R. Rairoux and U. Wandinger, eds., pp. 289–295. Springer Verlag, New York, Berlin, Heidelberg.
- Browell, E. V., T. D. Wilkerson and T. J. McIlrath, 1979: Water vapor differential absorption lidar development and evaluation. *Appl. Opt.*, **18**, 3474–3483.
- Charney, J. G., R. Fjørtoft and J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254.

- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.
- Di Girolamo, P., D. Summa, H. Bauer, V. Wulfmeyer, A. Behrendt, G. Ehret, B. Mayer, M. Wirth and C. Kiemle, 2004: Simulation of the performance of WALES based on an end-to-end model. In: *Proceedings of the 22nd International Laser Radar Conference*, pp. 957–960.
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.
- DWD, 2000: Satellite application facility on climate monitoring: Science Plan. Technical Report SAF/CM/DWD/SCI3.0, Deutscher Wetterdienst (DWD), 63004 Offenbach, Germany.
- Engelbart, D. A. M. and E. Haas, 2006: LAUNCH-2005 – International Lindenberg campaign for assessment of humidity and cloud profiling systems and its impact on high-resolution modelling. In: *TECO-2006 - WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation*. Geneva, Switzerland.
- Epstein, E. S., 1969: Stochastic-dynamic prediction. *Tellus*, **21**, 739–759.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.
- Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Fisher, M., 1998: Development of a Simplified Kalman Filter. Technical Memorandum 260, ECMWF.
- Gandin, L. S., 1963: *Objective analysis of meteorological fields*. Gidrometeor. Izd. Leningrad. (English Translation by Israel Program for Scientific Translation, Jerusalem, 1965).
- Geerts, B., R. Damiani and S. Haimov, 2006: Finescale vertical structure of a cold front as revealed by an airborne Doppler radar. *Mon. Wea. Rev.*, **134**, 251–271.
- Gendt, G., G. Dick, C. Reigber, M. Tomassini, Y. Lui and M. Ramatschi, 2004: Near real time GPS water vapor monitoring for numerical weather prediction in Germany. *J. Meteorol. Soc. Japan*, **82**, 1B, 361–370.
- Giering, R. and T. Kaminski, 1998: Recipes for Adjoint Code Construction. *ACM Trans. Math. Software*, **24**, 4, 437–474.
- Grell, G. A., J. Dudhia and R. R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Technical Note TN-398+STR, NCAR, P.O. Box 3000, Boulder, CO, 80307.

- Grzeschik, M., H.-S. Bauer, V. Wulfmeyer, D. Engelbart, U. Wandinger, I. Mattis, D. Althausen, R. Engelmann, M. Tesche and A. Riede, 2008: Four-Dimensional Variational Data Analysis of Water Vapor Raman Lidar Data and Their Impact on Mesoscale Forecasts. *J. Atmos. Oceanic Technol.*, **25**, 1437–1453.
- Gustafsson, N., 2007: Discussion on ‘4D-Var or EnKF?’. *Tellus A*, **59**, 5, 774–777.
- Hager, W. W., 1989: Updating the Inverse of a Matrix. *SIAM Review*, **31**, 2, 221–239.
- Hohenegger, C. and C. Schär, 2007: Predictability and Error Growth Dynamics on Cloud-Resolving Models. *J. Atmos. Sci.*, **64**, 4467–4478.
- Hong, S.-Y. and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **10**, 2322–2339.
- Howcroft, J. G., 1971: Local forecast model: present status and preliminary verification. NMC Office note 50, National Weather Service, NOAA, US Dept of Commerce. Available from the National Centers for Environmental Prediction, 5200 Auth Rd, Rm 100, Camp Springs, MD 20746.
- Huang, X.-Y., X. Yang, N. Gustafsson, K. Mogensen and M. Lindskog, 2002: Four-dimensional variational data assimilation for a limited area. HIRLAM Technical Report 57. Available from SMHI.
- Hunt, B. R., E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke and A. V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277.
- Ismail, S. and E. V. Browell, 1989: Airborne and spaceborne lidar measurements of water vapor profiles: A sensitivity analysis. *Appl. Opt.*, **17**, 3603–3615.
- Kain, J. S., 2004: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteorol.*, **43**, 170–181.
- Kain, J. S. and J. M. Fritsch, 1990: A one-dimensional entraining / detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.
- Kain, J. S. and J. M. Fritsch, 1993: Convective parameterization of mesoscale models: The Kain-Fritsch scheme. In: *The Representation of Cumulus Convection in Numerical Models*, volume 24 of *Meteorol. Monographs*, pp. 165–170. American Meteorological Society, Boston.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.
- Kalnay, E., B. Hunt, E. Ott and I. Szunyogh, 2006: Ensemble forecasting and data assimilation: two problems with the same solution? In: *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, eds., chapter 7, p. 157. Cambridge University Press.
- Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang and J. Ballabrera-Poy, 2007a: 4-D-Var or ensemble Kalman filter? *Tellus A*, **59**, 5, 758–773.

- Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang and J. Ballabrera-Poy, 2007b: Response to the discussion on “4-D-Var or EnKF?” by Nils Gustafsson. *Tellus A*, **59**, 5, 778–780.
- Klemp, J. B. and R. B. Wilhelmson, 1978: Simulations of three dimensional convective storm dynamics. *J. Atmos. Sci.*, **35**, 1070–1096.
- Leike, A., 2002: Demonstration of the exponential decay law using beer froth. *Eur. J. Phys.*, **23**, 21–26.
- Leith, C. E., 1974: Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev.*, **102**, 6, 409–418.
- Lenschow, D., V. Wulfmeyer and C. Senff, 2000: Measuring second– through forth–order moments in noisy data. *J. Atmos. Oceanic Technol.*, **17**, 1330–1347.
- Liu, D. C. and J. Nocedal, 1989: On the Limited Memory Method for Large Scale Optimization. *Mathematical Programming B*, **45**, 3, 503–528.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteorol. Soc.*, **112**, 1177–1194.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP - A comparison with 4D-VAR. *Quart. J. Roy. Meteorol. Soc.*, **129**, 3183–3203.
- Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, **20**, 130–141.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333.
- Martin, W. J. and M. Xue, 2006: Sensitivity analysis of convection of the 24 May 2002 IHOP case using very large ensembles. *Mon. Wea. Rev.*, **134**, 192–207.
- Mattis, I., A. Ansmann, D. Althausen, V. Jaenisch, U. Wandinger, D. Müller, Y. F. Arshinov, S. M. Bobrovnikov and I. B. Serikov, 2002: Relative-humidity profiling in the troposphere with a Raman lidar. *Appl. Opt.*, **41**, 6451–6462.
- Mattis, I. and V. Jaenisch, 2006: Automated Lidar Data Analyzer (ALDA) for RAMSES – the autonomously operating German Meteorological Service Raman Lidar for Atmospheric Moisture Sensing. In: *Reviewed and Revised Papers Presented at the 23rd International Laser Radar Conference*, C. Nagasawa and N. Sugimoto, eds., pp. 215–218. Nara, Japan.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. Iacono and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682.
- Molteni, F., R. Buizza, T. N. Palmer and T. Petroliaigis, 1996: The new ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteorol. Soc.*, **122**, 73–119.

- Moore Jr., A. S., K. E. Brown, W. M. Hall, J. C. Barnes, W. C. Edwards, L. B. Petway, A. D. Little, W. S. Luck Jr., I. W. Jones, C. W. Antill Jr., E. V. Browell and S. Ismail, 1997: Development of the Lidar Atmospheric Sensing Experiment (LASE) – An advanced airborne DIAL instrument. In: *Advances in Atmospheric Remote Sensing with Lidar*, A. Ansmann, R. Neuber, P. Rairoux and U. Wandinger, eds., pp. 281–288. Springer Verlag, New York.
- Nehrkorn, T., G. D. Modica, M. Cemiglia, F. H. Ruggiero, J. G. Michalakes and X. Zou, 2001: MM5 adjoint development using TAMC: Experiences with an automatic code generator. In: *Proceedings of 14th Conference on Numerical Weather Prediction*, pp. 481–484.
- Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet and J. Tribbia, 1993: Ensemble prediction. In: *ECMWF Seminar Proceedings on Validation of models over Europe*, volume 1. ECMWF, Shinfield Park, Reading, UK.
- Panofsky, H. A., 1949: Objective weather-map analysis. *J. Appl. Meteorol.*, **6**, 386–392.
- Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Centre’s spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteorol. Soc.*, **126**, 1143–1170.
- Reisner, J., J. Rasmussen and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteorol. Soc.*, **124B**, 1071–1107.
- Richardson, L. F., 1922: *Weather prediction by numerical process*. University Press, Cambridge.
- Rotach, M. W., M. Arpagaus, M. Dorninger, C. Hegg, A. Montani, R. Ranzi, F. Bouttier, A. Buzzi, G. Frustaci, K. Mylne, E. Richard, A. Rossa, C. Schär, M. Staudinger, H. Volkert, V. Wulfmeyer, P. Ambrosetti, F. Ament, C. Appenzeller, H.-S. Bauer, S. Davolio, M. Denhard, L. Fontannaz, J. Frick, F. Fundel, U. Germann, A. Hering, C. Keil, M. Liniger, C. Marsigli, Y. Seity, M. Stoll, A. Walser and M. Zappa, 2008: MAP D-PHASE: Real-time Demonstration of Weather Forecast Quality in the Alpine Region. *Bull. Amer. Meteorol. Soc.* submitted.
- Schwitalla, T., 2007: Kritische Prozessdarstellungen in Wettervorhersagemodellen zur quantitativen Niederschlagsvorhersage. Diplomarbeit. Institut für Physik und Meteorologie, Universität Hohenheim/Universität Stuttgart.
- Schwitalla, T., H.-S. Bauer, V. Wulfmeyer and G. Zängl, 2008: Systematic errors of QPF in low mountain regions. *Meteorol. Zeitschrift*. in press.
- Talagrand, O., 1997: Assimilation of Observation, an Introduction. *J. Meteorol. Soc. Japan*, **75**, 1B, 191–209.
- Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteorol. Soc.*, **74**, 2317–2330.

- Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Tracton, M. S. and E. Kalnay, 1993: Ensemble forecasting at NMC: Practical aspects. *Wea. Forecasting*, **8**, 379–398.
- Turner, D. D., R. A. Ferrare, L. A. Heilman Brasseur, W. F. Feltz and T. P. Tooman, 2002: Automated retrievals of water vapor and aerosol profiles from an operational Raman lidar. *J. Atmos. Oceanic Technol.*, **19**, 1, 37–50.
- Wakimoto, R. M., H. V. Murphey, E. V. Browell and S. Ismail, 2006: The “triple point” on 24 May 2002 during IHOP. Part I: Airborne Doppler and LASE analyses of the frontal boundaries and convection initiation. *Mon. Wea. Rev.*, **134**, 231–250.
- Wandinger, U., 2005: Raman lidar. In: *Lidar. Range-resolved optical remote sensing of the atmosphere*, C. Weitkamp, ed., pp. 241–271. Springer.
- Weckwerth, T. M., D. B. Parsons, S. E. Koch, J. A. Moore, M. A. LeMone, B. B. Demoz, C. Flamant, B. Geerts, J. Wang and W. F. Feltz, 2004: An Overview of the International H2O Project (IHOP_2002) and Some Preliminary Highlights. *Bull. Amer. Meteorol. Soc.*, **85**, 253–277.
- Whiteman, D. N., 2003: Examination of the traditional Raman lidar technique. II. Evaluating the ratios for water vapor and aerosols. *Appl. Opt.*, **42**, 2593–2608.
- WMO, 1999: The WCRP/GEWEX Global Water Vapor Project (GVaP): Science Plan. publication 27, International GEWEX Project Office, 1010 Wayne Ave., Silver Spring, MD 20910, USA.
- Wulfmeyer, V., 1999: Investigation of turbulent processes in the lower troposphere with water-vapor DIAL and Radar–RASS. *J. Atmos. Sci.*, **56**, 1055–1076.
- Wulfmeyer, V., H.-S. Bauer, A. Behrendt, F. Vandenberghe and E. V. Browell, 2004: Assimilation of DIAL data in mesoscale models: An impact study during IHOP_2002. In: *Proceedings of the 22nd International Laser Radar Conference*, pp. 957–960.
- Wulfmeyer, V., H.-S. Bauer, M. Grzeschik, A. Behrendt, F. Vandenberghe, E. V. Browell, S. Ismail and R. Ferrare, 2006: Four-Dimensional Variational Assimilation of Water Vapor Differential Absorption Lidar Data: The First Case Study within IHOP_2002. *Mon. Wea. Rev.*, **134**, 1, 209–230.
- Wulfmeyer, V., A. Behrendt, H.-S. Bauer, C. Kottmeier, U. Corsmeier, A. Blyth, G. Craig, U. Schumann, M. Hagen, S. Crewell, P. Di Girolamo, C. Flamant, M. Miller, A. Montani, S. Mobbs, E. Richard, M. W. Rotach, M. Arpagaus, H. Russchenberg, P. Schlüssel, M. König, V. Gärtner, R. Steinacker, M. Dorninger, D. D. Turner, T. Weckwerth, A. Hense and C. Simmer, 2008: RESEARCH CAMPAIGN: The Convective and Orographically Induced Precipitation Study. *Bull. Amer. Meteorol. Soc.*, **89**, 1477–1486.

- Wulfmeyer, V. and J. Bösenberg, 1998: Ground-based differential absorption lidar for water-vapor profiling: Assessment of accuracy, resolution, and meteorological applications. *Appl. Opt.*, **37**, 3825–3844.
- Wulfmeyer, V. and C. Walther, 2001a: Future performance of ground-based and airborne water vapor differential absorption lidar. I: Overview and theory. *Appl. Opt.*, **40**, 5304–5320.
- Wulfmeyer, V. and C. Walther, 2001b: Future performance of ground-based and airborne water vapor differential absorption lidar. II: Simulations of the precision of a near-infrared, high-power system. *Appl. Opt.*, **40**, 5321–5336.
- Xiao, Q., X. Zou and B. Wang, 2000: Initialization and simulation of a land-falling hurricane using a variational bogus data assimilation scheme. *Mon. Wea. Rev.*, **218**, 2252–2269.
- Xue, M. and W. J. Martin, 2006a: A high-resolution modeling study of the 24 May 2002 case during IHOP. Part I: Numerical simulation and general evolution of the dryline and convection. *Mon. Wea. Rev.*, **134**, 1, 149–171.
- Xue, M. and W. J. Martin, 2006b: A high-resolution modeling study of the 24 May 2002 case during IHOP. Part II: Horizontal convective rolls and convective initiation. *Mon. Wea. Rev.*, **134**, 1, 172–191.
- Zou, X., Y.-H. Kuo and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic adjoint model. *Mon. Wea. Rev.*, **123**, 2229–2249.
- Zus, F., M. Grzeschik, H.-S. Bauer, V. Wulfmeyer, G. Dick and M. Bender, 2008: Development and optimization of the IPM MM5 GPS slant path 4DVAR system. *Meteorologische Zeitschrift*, **17**, 867–885.

Acknowledgements

I would like to express my profound respect and gratitude to my principal supervisor Prof. Dr. Volker Wulfmeyer, Institute of Physics and Meteorology, University of Hohenheim, for his help, support and patience.

I take this opportunity to thank my co-advisor Dr. Hans-Stefan Bauer, Institute of Physics and Meteorology, University of Hohenheim, for continuous, invaluable and extensive support throughout the last years.

I would like to thank Francois Vandenberghe for his assistance with the MM5 4D-Var software package.

I thank Ed Browell, Ulla Wandinger, Dirk Engelbert for providing observational data for the assimilation and Galina Dick for GPS data for the validation.

I am thankful to the Deutsches Klimarechenzentrum GmbH (DKRZ) in Hamburg for the computing facilities and the Deutsche Forschungsgemeinschaft (DFG) for the funding in the framework of the priority program SPP1176 "Quantitative Precipitation Forecast".

I'm thankful to my colleagues Florian Zus for the fruitful discussions about data assimilation and Marcus Radlach about meteorology.

A special thanks go to Kirsten Warach-Sagi, Sandip Pal, Thomas Schwitalla for proof-reading. And again I would like to thank my colleges Marcus Radlach, Sandip Pal, Maxim Schiller, Florian Zus for the nice evenings mostly with Indian food.

Finally I thank my parents for their continuous support.