# Physical Initialisation of Precipitation in a Mesoscale Numerical Weather Forecast Model

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
**Marco Milan**
aus
S. Pietro in Gu

Bonn, 2009

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Referent: Prof. Dr. Clemens Simmer

2. Referent: Prof. Dr. Andreas Hense

Tag der Promotion: 02/02/2010

Erscheinungsjahr: 2010

# Abstract

Short term quantitative precipitation forecast (QPF) is an important task for numerical weather prediction (NWP) models, particularly in summer. The increase of model resolution requires the understanding of the initiation and evolution of convection. Initialisation schemes based on radar derived precipitation fields can reduce the model forecast error in convective cases. Any improvement of QPF denotes a correct forecast of the dynamics and the moisture content of the atmosphere, thus upgrading QPF generically improves NWP forecast.

The method, which we call *Physical initialisation Bonn* (PIB), uses as the most important input the precipitation estimation from the German weather service (DWD) radar network and assimilates the data into the operational non-hydrostatic COSMO model. During the assimilation window, PIB converts the input data (radar precipitation and cloud top height from satellite data) into prognostic COSMO variables, which are relevant for the development of rain events. PIB directly adjusts vertical wind, humidity, cloud water, and cloud ice in order to force the model state towards the measurements. The most distinctive feature of the algorithm is the adjustment of the vertical wind profile in the framework of a simple precipitation generation scheme.

In a first study we performed an identical twin experiment with three convective cases. The consistency of PIB with the physics of the NWP model is proved using qualitative comparisons and quantitative evaluations (e.g. objective skill scores).

The performance of PIB, using real data, is investigated by applying the scheme to the whole month of August 2007, with three simulations every day, at 00, 08 and 16 UTC. Every simulation consists of two hours of data assimilation followed by seven hours of free forecast. The comparison with the Control run and with Latent heat nudging, the operational radar data assimilation scheme from the DWD, is also made. PIB succeeds in improving QPF for up to six hours. Its results are comparable to the forecast by LHN. The sensitive of PIB to different assimilation windows is tested. An assimilation window of only 15 minutes is enough to provide the trigger for convection and to enhance the forecast quality. Thus PIB is much more time efficient than LHN and need much less observation values.

# Contents

# Chapter 1

# Introduction

Clouds and precipitations are essential components of the earth's water and energy cycle; they are important meteorological parameters of our habitat and influence life in various ways.

The accurate forecast of precipitation events, particularly for heavy precipitation, is of capital interest for the economy. For many applications a Quantitative Precipitation Forecast (QPF) is essential. In agriculture, precipitation can have different effects, sometimes positive, sometimes negative depending also on the quantity. In hydrology, precipitation is the input of a range of simulation models, e. g. for runoff prediction, flashflood warnings, and water use and quality management. By improving QPF severe weather warnings can be issued more precisely, and reduce flood damages and save lives.

The goal of this work is the improvement of QPF, especially for convective events in the midlatitudes. We are particularly interested in improving the short time range forecast, up to about 9 hours.

## 1.1 Motivation

A Numerical Weather Prediction (NWP) model integrates the governing equations of hydrodynamics using numerical methods subject to specified initial conditions. Numerical approximations are fundamental to almost all dynamical weather prediction schemes since the complexity and nonlinearity of the hydrodynamic equations do not allow exact solutions of the continuous equations. For this reason meteorology was one of the very first fields of physical science that had the opportunity and necessity to exploit high speed computers for the solution of multi-dimensional time-dependent non-linear problems (Mesinger and Arakawa, 1976 [71]).

In the early NWP experiments (the first unsuccessful one was from Richardson in 1922 [81]) the need for an automatic "objective" analysis became quickly apparent (Charney, 1951 [17]). The analysis is a procedure to estimate the atmospheric dependent variables on a regular two- or three-dimensional grid using the data available from irregularly spaced observation networks (Daley, 1991, [26]).

The determination of the initial condition for a NWP model and the subsequent assimilation of new observation data in the model run are fundamental for the quality of the forecast. The right determination of the model prognostic variables for the analysis provides the forecast model with the *starting point*, from which we can make the prognosis.

The most difficult weather element to predict correctly is rainfall ( Ebert et al. 2003, [34]) because it is not a continuous field in space and time, like wind and temperature for instance; it is rather a collection of solid or liquid particles developed within weather systems with a typical life time between an hour to few days. Furthermore its formation and evolution is highly complex and nonlinear and occurs on scales that are several orders of magnitude smaller than the size of a numerical grid box (in our case the grid box has a dimension of $2.8km \times 2.8km$, while rain develops in spatial scales smaller than $1cm$). Parameterisation schemes are therefore used to treat subgrid scale processes such as precipitation, but these had to be simplified for reasons of computational efficiency or the lack of knowledge about the true characteristics of those processes. We have to remark that normally the QPF skill also varies seasonally (Uccellini et al. 1999, [93], Weckwert et al. 2004 [96]) with the summer marked by lower forecast skill (we can clearly see this in terms of *Threat Score* in Fig. 1.1, for the definition of threat score see appendix A).

The problem of warm season QPF is the high unpredictability of the convective events (Lorenz, 1969 [68]). These events are particularly challenging to predict because of their small spatial scale, short lifetime and non-linear, chaotic behaviour. In the current NWP models convection is parameterized. It is believed that an improved representation of convection in forecast models is a necessary path through which major advances in QPF will be realized.

One necessary condition for a prediction of convective rainfall is a good forecast of where and when convection will initially develop; moreover the knowledge of the atmospheric state before the development of convection is crucial in order to initialise the NWP model. Currently the prediction and understanding of convection initiation processes, as well as the subsequent intensity, areal coverage and distribution of convective rainfall, are largely impeded by inaccurate and incomplete water vapour measurements (National Research Council, 1998 [22]).

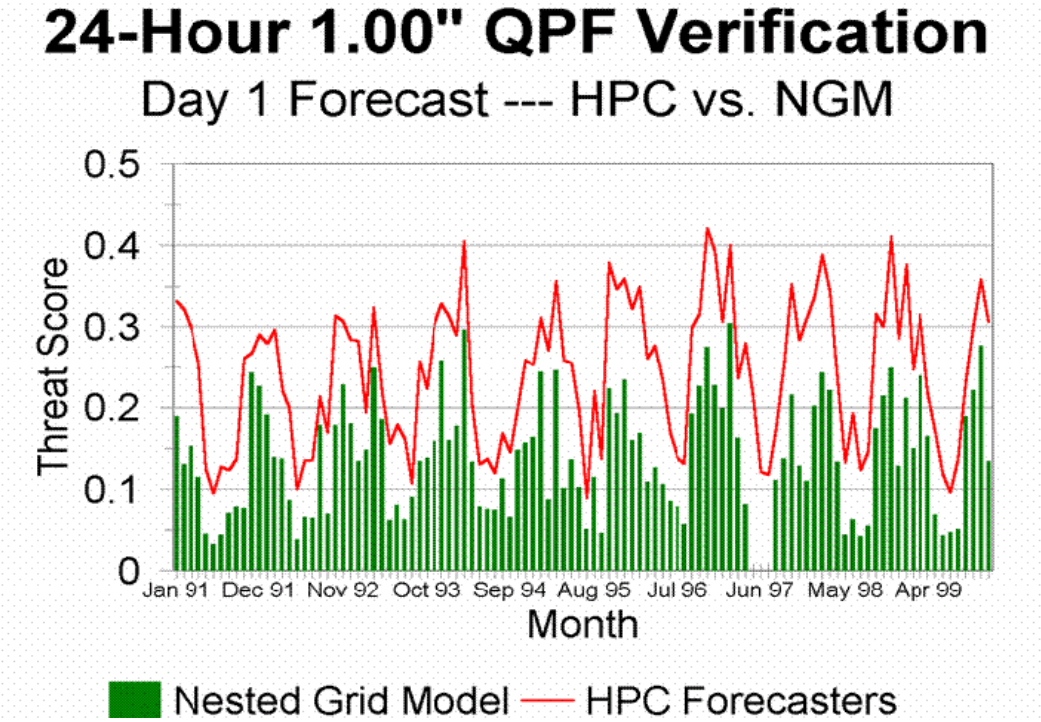


Figure 1.1: *Monthly threat scores for quantitative precipitation forecasts for the period form Jan 1991 to Dec 1999. Threat scores are plotted for the NGM (Nested Grid Model, green histogram) and for the HPC (Hydrometeorol. Prediction Center) forecaster predictions (red line) for the 24h period., picture from Wecwert at all., 2004 [96]*

The improvement of precipitation forecasts depends strongly on the coupling of humidity and wind fields (Zou and Xiao, 2000 [104]). An error in the initial state can amplify through the modelling process and result in drastic under or overpredictions of precipitation (Park, 1999 [77]). A model-consistent assimilation of precipitation and the corresponding fields of humidity can reduce the lack in the hydrological variables results in the so called spin-up (spin-down) problem (Krishnamurti, 1993 [57]), i.e. a hydrological imbalance between precipitation and evaporation (e.g. Fig. 1.2). This improvement is especially beneficial for nowcasting and short-range forecasting, with an improvement of the forecast for surface pressure and dynamics.

Chang and Holt (1994 [16]) used a theoretical experiment and showed that, for winter cases, the improvements in precipitation forecast due to data assimilation are still noticeable after 30 hours of forecast.

Another study (Anderson et al., 2000 [4]) has investigated the impact of observations on meso-scale model forecasts of three-hourly rainfall accumulations. They assimilated three-dimensional humidity fields from the *Moisture Observation Processing System* (MOPS). The main output from MOPS consists of a three-dimensional analysis of cloud fraction. This is converted into a set of relative humidity soundings at each model grid point, which are then assimilated in the same way as radiosondes (Wright, 1993 [99]). The MOPS leads to a significant increase in the forecasts quality. This is a proof of the importance of humidity assimilation in NWP (see also Macpherson et al., 1996 [69] ).

The reason for the model forecast improvement due to the humidity assimilation is that QPF is directly linked to the model's water cycle. The spin-up problem is due to the fact that NWP model simulations are typically initialised in so-called dry state, leaving the state variables for condensed water at zero. Usually several hours of simulated weather evolution will pass until the hydrological cycle is established in such a quasi-equilibrium.

Another parameter that influences the skill of QPF is orography; its forcing is generally badly reproduced by the model dynamics, the precipitation pattern (Gollvik, 1999 [40]) are especially poor for convective events in these areas. In our work, therefore, we started with simulations in an area where orography is not such an important factor (northern Germany) and then we applied the model using a larger domain covering Germany as a whole including the alps. In this way we tested the improvement of our scheme in both situations: with and without complex orography.

We have implemented an assimilation technique that tries to improve the forecast in terms of location, structure and movement within the area of activity. The special aim is to improve *nowcasting* (about two hours) and

*short range forecasts* (less then 24 hours). Specifically we restricted our self to predictions of only 9 hours, because the life cycle of the assimilated synoptic situation in the studied area was shorter than that. The benefit of data assimilation in high resolution models (in this case 2.8 km) for longer periods decreases because of the increase of unpredictability with the spatial resolution in the NWP models (Germann and Zawadzki, 2002 [38]) and the increasing influence of the boundaries.

For model output evaluation a comparison with radar observations is performed, by Objective Skill Scores (OBSS, defined in appendix A) and other statistical quantities.

## 1.2 Assimilation techniques

In order to perform a numerical weather forecast, it is necessary to have not only an appropriate numerical model but also a description of the corresponding initial conditions as accurate as possible (Talagrand, 1997 [90]).

In the early NWP model experiments (Richardson, 1922 [81], and Charney et al., 1950 [18]) the initial conditions were derived from an interpolation of the available observations to the grid, and these fields were manually digitized. It becomes immediately clear that a production of an automatic objective analysis was necessary also in order to economize time.

The available observations, however, do not provide the temporally synchronous and the spatially homogeneous description of the atmosphere, that is required by a numerical model. Moreover many observations are in quantities that are not prognostic model variables (for example in our case reflectivity from radar data has to be converted into precipitation rate). In this sense, assimilation is the process by which observations are combined together with a numerical model in order to produce, as accurate as possible, a description of the state of the atmosphere in terms of model variables.

The description of that atmospheric state is called analysis. It tries to reproduce the true state using observations.

In case the model state is overdetermined by the observations, the analysis reduces to an interpolation problem (Bouttier and Cortier, 1999 [13]). In most cases the analysis problem is under-determined because data is sparse especially in ocean regions. Moreover, in these regions most of the data are only indirectly related to the model variables (e. g. satellite data).

In order to make analysis a well-posed problem it is necessary to rely on some *background* (or first guess) information in the form of an *a priori* estimate of the model state. The background could be a climatological state or, in the

recent concepts, a previous forecast. Using the background state and the new observations a new state for the atmosphere is given from the assimilation techniques. This state will be propagated using the NWP model. For more detailed information about data assimilation concepts the reader is referred to Daley, 1991 [26] or Kalnay, 2003 [53] or Bouttier and Courtier, 1999 [13]. In this section we will describe briefly only some methods, in order to have a general idea about data assimilation.

We can subdivide the assimilation techniques basically in two categories: variational and non-variational.

### 1.2.1 Non variational approaches

#### Successive corrections method

We take a dependent variable $f(\vec{r})$ (Daley, 1991 [26] ), where $\vec{r}$ defines the spatial location. The $\vec{r}_i$ is the $i$th analysis gridpoint and $\vec{r}_k$ the $k$th observing location. The letter **O** stands for observation, **B** for background and **A** for the analysis, the error is $\epsilon$. For example the background error at an analysis grid point $i$ is $\epsilon_B(\vec{r}_i)$.

If we consider a single observation $\vec{r}_k$ we can estimate the analysis increment at the analysis gridpoint $i$:

$$f_A(\vec{r}_i) - f_B(\vec{r}_i) = W_{ik}[f_O(\vec{r}_k) - f_B(\vec{r}_k)] \tag{1.1}$$

where $[f_O(\vec{r}_k) - f_B(\vec{r}_k)]$ is the observation increment and $W_{ik}$ depends on the expected background and observation error variance. The observation increment is related to a weight depending on the distance $w(\vec{r}_k - \vec{r}_i)$ (in the original system from Cressman, 1959 [24] the weight depends only on the distance). For such a system it si problematic to define the weight function, $W_{ik}$.

Assume that the background error is homogeneous and the observation errors are spatially uncorrelated. Then equation 1.1 at the first iteration step can be written as

$$f_A^1(\vec{r}_i) = f_B(\vec{r}_i) + W_i^T[\vec{f_O} - \vec{f_B}] \tag{1.2}$$

where $\vec{f_O}$ and $\vec{f_B}$ are column vectors of lenght $K_i$ (the number of observations available in the range of influence) and $W_i$ is a column vector of length $K_i$ of weights.

At this point an iteration process is applied. The second step uses the $f_A^1$ as background and we have $f_A^2(\vec{r}_i) = f_A^1(\vec{r}_i) + W_i^T[\vec{f_O} - \vec{f_B}]$. Then we iterate the

process for $n$ times, until convergence.

### Optimal interpolation

Optimal interpolation, also called statistical interpolation (Daley, 1991 [26]), is a minimum variance method. The basic form is the same as the *successive correction method*. If we take $K$ observation points for the analysis influencing the $i$th point we get:

$$f_A(\vec{r}_i) = f_B(\vec{r}_i) + \sum_{k=1}^{K} W_{ik}[f_O(\vec{r}_k) - f_B(\vec{r}_k)] \tag{1.3}$$

or, with a more common notation (generally used in the variational approach): $A_i = B_i + \sum_{k=1}^{K} W_{ik}[O_k - B_k]$. The optimal interpolation uses a minimum variance estimation procedure for the definition of the weights $W_{ik}$. This assimilation scheme tries to minimize the expected analysis error variance as a function of the weights. The expected analysis error variance is $< (A_i - T_i)^2 >$, where $T_i$ is the true value at the point $i$. In this assimilation system both the background field and the observations are assumed to be unbiased.

### Latent heat nudging

The *Latent Heat Nudging (LHN)* scheme is based on the work of Manobianco et al. (1994, [70]) and on the successive application from Jones and Macpherson (1997, [50]). The principal idea is to correct the model's latent heating at each time step by an amount calculated from observed and modelled precipitation. They assumed that the vertically integrated latent heating is proportional to the surface rain rate (Leuenberger D., 2005 [65] ). Practically this scheme nudges (Anthes, 1974 [5] and Hoke and Anthes, 1976 [45]) the model temperature profile to the estimated temperature profile (using a saturation adjustment technique).

The extra heating is a source term in the prognostic model temperature equation and that changes the buoyancy and by ascent the precipitation. For a description of LHN and its use in recent years see Leuenberger, 2005 [65], Leuenberger and Rossa, 2004 [66] and Klink and Stephan, 2004 [54].

### The Physical Initialisation scheme

The idea of using precipitation data in NWP models was in the first experiments applied in the tropics (Krishnamurti et al. 1984 [58]). The use of an assimilation technique in such an area was necessary because of the

sparse presence of the classical data (for example radiosonde and synoptic station), especially over oceanic areas. In the tropics the sparsity of data also results in large errors in the divergent wind which in turn usually compounds the humidity errors through moisture convergence, resulting in a erroneous specification of the diabatic forcing (Krishnamurti et al., 1991 [59]). The spin up problem (described in Fig. 1.2) is inherent to any NWP models (Krishnamurti et al., 1988 [56]). The work of Krishnamurti tries to reduce it by assimilating the precipitation fields (a review of this work can be found in Krishnamurt et al., 1996, [55]).

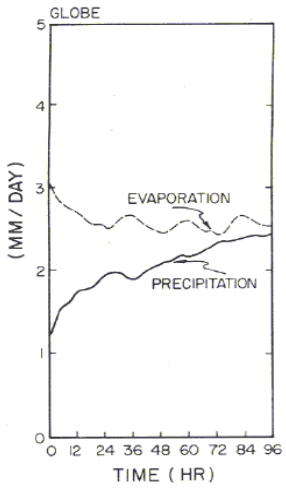


Figure 1.2: *Globally averaged value of evaporation and precipitation (units $mm \cdot day^{-1}$) plotted as a function of forecast period. Krishnamurti et al. 1988, [56]*

Krishnamurti used a nudging scheme (or *Newtonian relaxation* ) involving linear forcing to lead the evolution of variables towards certain predetermined space-time estimates. The general form of the equation for these processes is:

$$\frac{\partial A(x,y,t)}{\partial t} = F(A,x,y,t) + N(A,t) \cdot (A^0(x,y,t) - A(x,y.t)) \qquad (1.4)$$

Where $A$ is the predicted value of the variable at a time level during time integration and $A^0$ is the predetermined *true* estimate of the variable at that time (analysis); $F(A,x,y,t)$ is the time rate of change for the variable $A$ due to various dynamical and physical processes in the numerical model; $N(A,t)$ is the relaxation coefficient, the value that determines the degree with which the variable is forced towards its analysed value $A^0$.

When the model predicted value $A$ is equal to the analysed value the Newtonian term (the second on the right side in the eq. 1.4) does not operate. The value

of the relaxation coefficient can be selected experimentally; a large relaxation coefficient will force the variable too strongly, resulting in dynamic imbalance, while a too weak relaxation coefficient is not efficient.

The Krishnamurti scheme uses a reverse Kuo algorithm for the improvement of the humidity analysis. In the Kuo algorithm the moistening and heating by the cumulus cloud are made using the temperature difference between the cloud and the undisturbed environment and the large scale convergence of moisture as indicators (Kuo, 1974 [62]). Therefore Krishnamurti et al. tried to find an analysis of the humidity field consistent with an imposed precipitation field and a cumulus parameterisation algorithm. For each time step a reverse cumulus parameterisation algorithm provides a modified humidity field over the convective areas each time step. A forward prediction of the humidity variable with a direct Kuo algorithm nearly reproduces the imposed rain.
In the reverse Kuo scheme the specific humidity is calculated using the relation:

$$q_m(p) = \frac{R}{-\frac{1}{g}\int_{p_T}^{p_B} \omega \frac{\partial q}{\partial p} dp} q(p) + \frac{\frac{1}{g}\int_{p_T}^{p_B} q dp}{\frac{1}{g}\int_{p_T}^{p_B} dp} \times [1 - R/(-\frac{1}{g}\int_{p_T}^{p_B} \omega \frac{\partial q}{\partial p} dp)] \quad (1.5)$$

where $q(p)$ and $q_m(p)$ are the original and modified values of specific humidity at pressure level $p$, $\omega$ is the vertical velocity and $p_B$ and $p_T$ denote the bottom and top pressure of the cumulus column, respectively. The modification is applied only in the precipitation regions, with:

$$-\frac{1}{g}\int_{p_T}^{p_B} \omega \frac{\partial q}{\partial p} dp > 0 \quad (1.6)$$

and with the restriction $q_m(p) \leq q_s$. In other words the modified specific humidity has as upper limit the saturation value. This scheme was based on observations from satellites and raingauges.

Krishnamurti's scheme in the rain-free areas has a humidity adjustment based on the two estimates (model and satellite) of the outgoing longwave radiation above the planetary boundary layer (PBL). The scheme used a measure of the advective-radiative imbalance and tries to minimize it using a modification of $q$.

### 1.2.2 Variational approaches

Variational approaches consider the atmospheric state system as a whole and do not deal explicitly with the individual components of the system. It involves the determination of stationary points (Daley,1991 [26]) of integral expressions

known as *functionals*[1]. A stationary point is a point (in the domain of the functional) where the rate of change of the functional in every possible direction from that point is zero.

One disadvantage of this approach, as for Optimal Interpolation and all versions of the Kalman filter, is the assumption of Gaussian error distributions.

### 3D-VAR

We can write the optimal interpolation scheme (for the notation see table 1.1) as $\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - H[\mathbf{x}_b]$.

Table 1.1: *Notation used in this section with the dimensions of the matrixes*

| | **variable** | **dim** | | **variable** | **dim** |
|---|---|---|---|---|---|
| $\mathbf{x}_t$ | true model state | n | $\mathbf{x}_b$ | background model state | n |
| $\mathbf{x}_a$ | analysis model state | n | $\mathbf{y}$ | vector of observations | p |
| $H$ | observation operator | $n \rightarrow p$ | $\mathbf{B}$ | cov. matrix of the background error | $n \times n$ |
| $R$ | cov. matrix of the observation error | $p \times p$ | $\mathbf{A}$ | cov. matrix of the analyse error | $n \times n$ |
| $\mathbf{K}$ | weight matrix | $n \times n$ | | | |

The three dimensional variational assimilation principle tries to avoid the computation of the weight matrix $\mathbf{K}$ completely by looking for the analysis as an approximate solution of the equivalent minimization problem defined by a cost function $\mathbf{J}$, which could be seen as a measurement of the distance of the estimate state from the observations.

$$\mathbf{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}]) \tag{1.7}$$

The problem is now to find the minimum of $\mathbf{J}$ using the gradient $\nabla \mathbf{J}(\mathbf{x})$; usually a descent algorithm is used.

In 3D-Var the $\mathbf{B}$ matrix must be defined, i.e. the background error covariance for all pairs of model variables.

### 4D-VAR

4D-Var is a generalization of 3D-Var for observations that are distributed in time. We must include a forecast model operator $M$ and an index $i$ for the model time; $M_{0 \rightarrow i}$ is the forecast model operator from the initial time to $i$.

[1] functional: the principal meaning is a function whose domain is a set of functions

The cost function is now defined as:

$$\mathbf{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \sum_i^n (\mathbf{y}_i - H_i[\mathbf{x}_i])^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i[\mathbf{x}_i]) \quad (1.8)$$

In this equation we assume for every model variable $\mathbf{x}$ that we have $\mathbf{x}_i = M_{0 \to i}(\mathbf{x})$; in this case we have to solve a nonlinear constrained optimisation problem. Two assumptions are necessary:

- The forecast model can be expressed as the product of intermediate forecast steps.
- Tangent linear hypothesis: The M operator can be linearised.

One major problem in 4D-var is, that it requires the calculation of $M_i^T$, the so-called adjoint model. For complex models the computational work can be very high.

## 1.3 Data assimilation strategies

Assimilation methods can also be classified as *continuous* or *intermittent* in time.

In intermittent methods the observations (Fig. 1.3) are assembled in small batches. This method stops the model at regular intervals (i.e. at analysis times) during the numerical integration, and uses the model fields at those times as a first guess for a new objective analysis, which is used to restart the model for the next integration interval. Intermittent updating takes advantage of observational data that becomes available at standard time (e.g., 00:00 UTC and 12:00 UTC). Each restart incorporates fresh data to limit error growth but also can cause spin-up error due to imbalances in the reinitialised fields.

In continuous assimilation observation batches over long time periods are considered. The corrections to the analysed state are smooth in time. The observational data are incorporated in the forecasts at every time step between the previous analysis step and the observation time. Dynamic assimilation uses the nudging or Newtonian relaxation technique to relax the model state toward the observed state (at the next model time) by adding extra forcing terms to the governing equations. In this way the model fields gradually correct themselves, requiring no further dynamic balancing through initialisation.

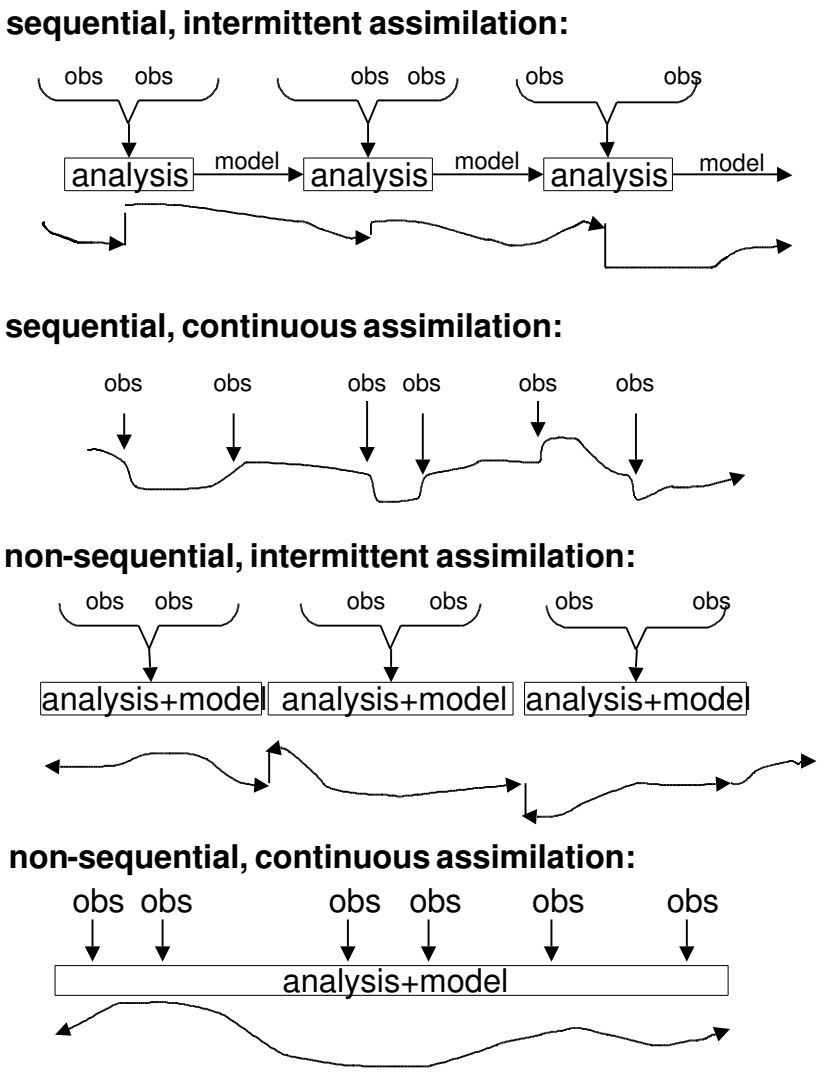


Figure 1.3: *Representation of four basic strategies for data assimilation as a function of time, figure from Data assimilation concept and methods Bouttier and Couttier, 1999 [13]*

# Chapter 2

# Model and data

Data assimilation combines observations with the results of a NWP model with the goal to produce a three dimensional picture of the atmosphere (analysis). The output of the data assimilation procedure is used to set the initial conditions for a forecast model. A description of data assimilation methods is given in Chapter 1.

This Chapter introduces the COSMO model (the NWP model , Consortium for Small-Scale MOdelling) and the observations (radar and satellite measurements), which are the elementary components of an assimilation system. Both the model and the data were provided from the DWD.

## 2.1 COSMO model

COSMO is a non-hydrostatic limited area atmospheric prediction model(Doms and Schättler, 2002 [29]). It is based on the primitive hydro-thermodynamical equations describing compressible flow in a moist atmosphere. This basic set of equations comprises the prognostic Eulerian equations for momentum, heat, total mass, mass of water substance and the equation of the state. The model equations are formulated in rotated geographical coordinates and a generalized terrain following height coordinate. These characteristics are useful because the model has been designed for both operational NWP and various scientific applications on the meso-$\beta$ and meso-$\gamma$ scale .

By employing 1 to 3 km grid spacing for operational forecasts over a large domain, it is expected that deep moist convection and the associated feedback mechanisms to the larger scales of motion can be explicitly resolved (Bryan et al., 2003 [14]). In case of large grid spacings, normally $7km$ or more, convection needs to be parameterised. In COSMO the Tiedtke parameterisation scheme (Tiedke, 1989 [92]) is operationally used, which uses moisture convergence and

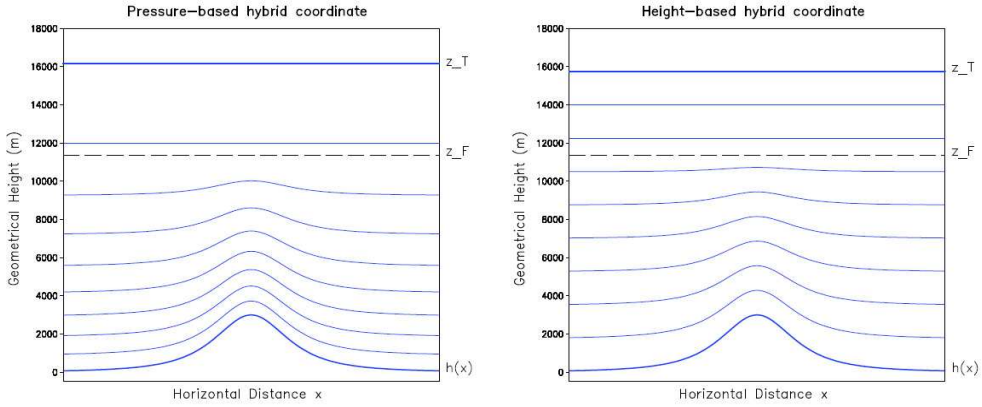


Figure 2.1: *Examples of terrain following levels, with pressure (on the right) and height (on the left) based hybrid coordinate, $z_T$ is the top of the model domain, $z_F$ is the height where the terrain following surface change to horizontal, figure from Doms et al., 2003 [30].*

boundary layer turbulence to determine the intensity and type of convection. Convection is separated in shallow, deep and midlevel. A Kain-Fritsch cumulus parameterisation scheme is included as an option, here the convergence is initiated taking into account the subcloud layer convergence and the intensity is connected to CAPE. (Kain and Fritsch, 1993 [52]).

For small-scale grid scales we assume that the model resolves convection by itself. Normally organized convective structures with a smaller grid spacing than 4 km could be resolved by the model; obviously convection could be activated also on smaller scales (1 km or less). We assume that with our resolution (2.8 km) we can resolve the basic aspects of convective processes using.

In case of complex topography high resolution can resolve explicitly the forcing and the enhancement of convection due to the presence of the mountainous regions.

The model variables are discretised on a staggered Arakawa-C/Lorenz grid with scalars defined at the centre of a grid box and the normal velocity components defined on the corresponding box faces (Fig. 2.2). In the vertical direction the COSMO atmosphere is divided in 50 layers. The centre of one layer is called *level* , the boundary of one layer is called *half level* . The principal half levels are defined from 0 at the top of the atmosphere to 50 at the soil interface (Fig. 2.1). The vertical velocity is defined at the half levels $k \pm 1/2$ (from Schulz and Schättler, 2005 [84]).

COSMO is used by DWD as the operational limited area forecast model since December 1998; its previous name was LM (Lokal-Modell). Since then, the

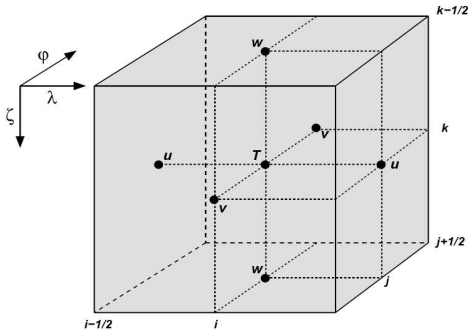


Figure 2.2: *dependent COSMO variables on the Arakawa-c Lorenz grid , figure from Doms et al., 2003 [30].*

model has been improved constantly and has undergone many modifications. In this work we have used COSMO-DE version 4.6. Until April 15, 2007, only a version with a horizontal resolution of 7 km was operational, This model was called LME (Lokal Modell Europa). Currently the new model, COSMO-DE, with a resolution for the meso-$\gamma$ events ($\Delta x \sim 2.8$ km) is also operational.

The default time integration scheme is a second order leapfrog HE-VI (horizontally explicit, vertically implicit), otherwise in this work as well as in the operational forecast the three step Runge-Kutta scheme is used.

The new model version with high resolution ($2.8km$ also called LMK) was tested in many research works (i.e. Baldauf et al., 2006 [8]). The boundary conditions are interpolated from the output of COSMO-EU ('COSMO Modell Europa', 7 km horizontal resolution). The further development of the COSMO model chain are embedded in an international project. In this project (see www.cosmo-model.org) the DWD, the German military geophysical advisory service and the national weather services of Italy, Greece, Poland and Switzerland are working together.

The most important differences between COSMO-EU and COSMO-DE are:

- The horizontal resolutions are different ($7km$ in COSMO-EU , $2.8km$ in COSMO-DE).
- The vertical resolution change from 40 in COSMO-EU to 50 levels in COSMO-DE.
- Deep convection is not parameterised any more in COSMO-DE. For shallow convection there is still a convection parameterisation active in COSMO.
- COSMO-DE takes graupel into account as an additional microphysical particle.

### 2.1.1 Data Assimilation in COSMO

Data assimilation is a requirement for NWP. Its value increases even more for nowcasting purposes (Laroche et al., 2005 [63]). In the operational COSMO model a scheme based on the observation nudging technique has been chosen for the purpose of data assimilation. The assimilation of the radar data is realised using a latent heat nudging scheme (see Section 1.2.1).

A detailed high-resolution analysis has to be produced frequently and efficiently. This requires a thorough use of synoptic and high-frequency observations. For this purpose an optimal interpolation (OI) or a 3DVAR are not optimal, since they do not allow to account for the exact observation time of synoptic data. 3DVAR and OI also neglect most of the high-frequent data unless the analysis scheme is applied very frequently, but in this case the computational costs are very expensive and the data density may become very low and inhomogeneous. Methods based on 4 dimensions offer potential advantages since they include the model dynamics in the assimilation process directly. The problem is that for the current available calculation power 4DVAR is too expensive for nowcasting.

In the observation nudging technique (using atmospheric fields and some of the surface and soil fields, Doms and Schättler, 2002 [29]) for COSMO the observation processing is performed only once for each observation at the timestep when the observation is available. Thus it is a continuously assimilation scheme (see Section 1.3 for explanations). In some cases the model values and observations are combined to complement missing pieces of information, e.g. model humidity is used to convert virtual temperature measured by RASS (Radio Acoustic Sounding System) into 'observed temperature'. The nudging scheme uses the traditional data (Synop, Pilot, ...) and also precipitation derived from radar reflectivity. An observation increment is applied, which is a function of the quality of the observation; quality weights are calculated as well.

The observational information contained in the observation increments and in their quality is spread explicitly to the model grid points. The assimilation scheme works for more than one grid point because if we correct the model only at one point the meteorological fields will not be smooth any more and we can create instability. In this scheme the modified model field is assumed to have an error prior to the correction. The smoothness of the meteorological fields then suggests that the corresponding model field is likely to contain a similar error in close vicinity. Hence we have to change the field near the observation taking into account the errors of the model and the errors of the field as well as their correlation.

The assimilation scheme nudges the information in a certain period of

time (assimilation window) and not only at the observation time. Thus is accomplished by using a temporal weight function.

COSMO-DE assimilates also precipitation from the composite of the DWD Radar network. The data has a temporal resolution of 5 minutes and a 1 km horizontal resolution. The Radar data are interpolated onto the COSMO-DE grid and then assimilated with a latent heat nudging scheme (Klink and Stephan, 2004 [54], Leuenberger and Rossa, 2004 [66] and Leuenberger, 2005 [65]).

### 2.1.2 Initial and boundary conditions

Regional models have a high horizontal resolution that cannot be afforded in a global model (Kalnay, 2003 [53]). Operational regional models have been embedded or nested into coarser resolution hemispheric or global models since the 1970s (e.g. Jones, 1973 [51]; Harrison and Elsberry 1972, [43];Chen and Miyakoda, 1974 [19]). The lower boundary conditions are physical but at the top and the sides they are usually artificial. For the lateral boundary condition a periodic one (for specific scientific applications) can be used, open or inflow-outflow boundary conditions can be used, which allow the atmosphere in the model interior domain to interact with the external environment.

The bottom boundary conditions play an important role since the transfer of physical properties as heat and moisture across this interface plays a fundamental role in most mesoscale circulation phenomena (Pielke R., 1984 [80]).

In a model (in our case COSMO-DE) the information about the variables at the lateral boundaries and their time evolution must be specified by an external data set. These external data can be obtained by interpolation from a forecast run of another model or from a coarser resolution run of COSMO, for example COSMO-EU (the boundary conditions are available every hour). At present, boundary data from the operational hydrostatic global model GME are supported for running COSMO-EU. Time dependent relaxation boundary conditions are used to force the solution at the lateral boundaries using the external data.

For COSMO-DE the initial conditions are not really well defined; because of the difference in the horizontal and vertical resolution, there is a spin-up time of 3-6 hours, during which the humidity fluxes are adjusted to the new topography.

### 2.1.3 Forecast and assimilation cycle

NWP is an initial-boundary value problem; given an estimate of the present state of the atmosphere (initial conditions) and appropriate surface and lateral boundary conditions, the model simulates (forecasts) the atmospheric evolution.

For the operational COSMO-DE forecast chain the four dimensional data assimilation cycle based on a nudging analysis scheme is installed. The initial and boundary conditions come from the COSMO-EU run, which initial and boundary conditions come from the global model GME.

In every grid point between two different boundary conditions a linear interpolation is made; in this way we can obtain the needed boundary conditions for every time step. For a simple model chain scheme see Fig. 2.3.

The assimilation cycle has a routine time of 3 hours. Every hour an analysis is written. Every 6 hours a new forecast simulation is performed; the forecast is computed for 18 hours.

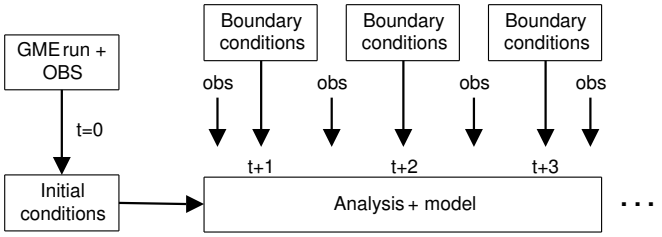


Figure 2.3: *Simple model chain scheme*

### 2.1.4 Grid scale precipitation

A cloud/precipitation model is a mathematical description (a set of equations) for the overall evolution of a cloud. These equations allow to predict numerically the water mass in each category of hydrometeors.

The grid scale precipitation scheme in COSMO model follows a one moment bulk formulation; the basic idea of this method is to assume as few categories of water as necessary and to predict the total mass fraction of water in each category.

In COSMO model different microphysical schemes are possible: from the simplest scheme with only cloud water content ($q_c$) and rain water content ($q_r$) to the most recent one which accounts for cloud snow ($q_s$), cloud ice ($q_i$) and the graupel phase ($q_g$).

Diagnostic and prognostic schemes are available for the rain processes (prognostic for liquid rain and snow). In prognostic mode the full budget equations for the precipitating hydrometeors are solved, whereas in diagnostic mode an approximated quasi-stationary version is utilised. For high resolution simulation, like in our case, the prognostic scheme is normally used.

The diagnostic precipitation formulation assumes a column equilibrium for precipitating constituents; this assumption is only as an acceptable approximation for coarse resolution. For this reason the prognostic precipitation was introduced. For deep convection the vertical advection of the precipitation phases must be taken into account since the air's vertical velocity is in the same magnitude as the fall velocity of rain and snow.

## 2.2 Radar data

The weather radar network of DWD consists of 16 operational Doppler RADAR (RAdio Detection And Ranging) systems in C-Band (wavelength of 4-8 cm and a frequency of 4-8 GHz), as shown in Fig. 2.4.

The DWD uses two scanning techniques:

- In the **volume scan** the antenna passes through 18 different angles of elevation from 37.0° to 0.5° every 15 minutes, thus covering the atmosphere up to an altitude of 12 km. The volume scan consists of two different measuring modes: the intensity mode covers the lower elevation angles from 0.5° to 4.5°, while the Doppler mode covers the elevation angles above. The horizontal range is 230 km for the intensity mode and 120 km for the Doppler mode.

- In the **precipitation scan** the lowest elevations are covered with a time resolution of five minutes The range is 128 km and the radar beam has an elevation between 0.5° and 1.8° depending on the orography. This scan is made because for hydrometeorology the use of the lowest positions (where the rain rate is estimated) has the greatest significance.

### 2.2.1 RY product

The RY product derived from the precipitation scan (see precedent Section), it contains the precipitation echoes measured every 5 minutes, making very short range forecasts of precipitation possible. This is the most important input for our assimilation scheme. The starting point of the PIB is the observed rain rate, which the RY product supplies based on the radar reflectivity. A

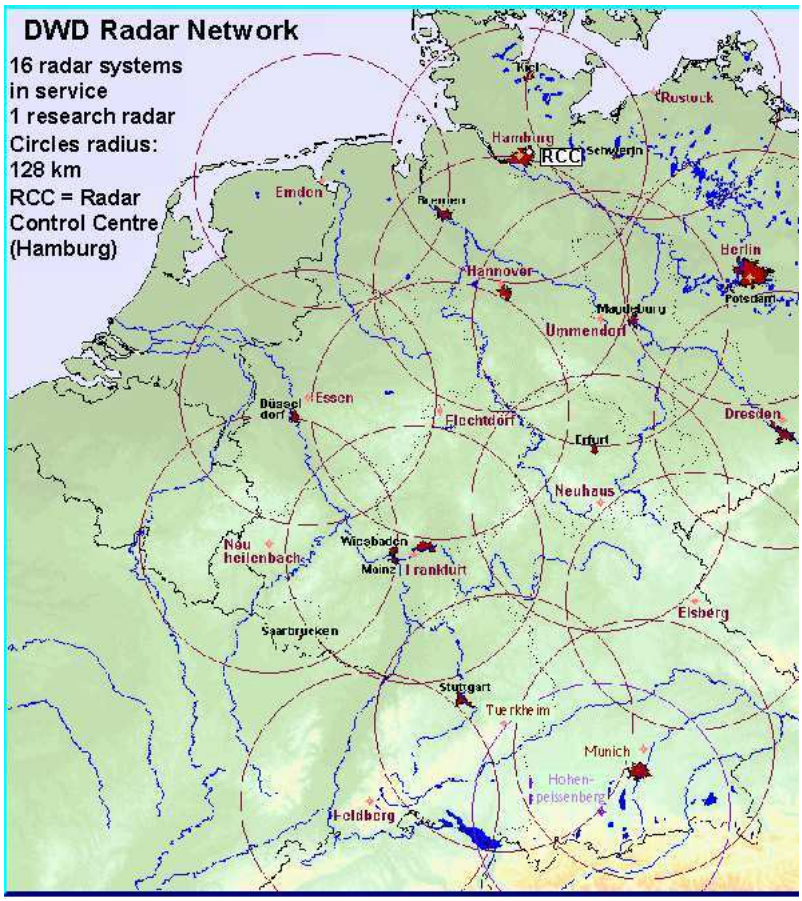


Figure 2.4: *DWD radar network. The circles around each radar site have a radius of 128km (Figure from www.dwd.de).*

quality flag is used in the generation of the precipitation fields; if the quality is insufficient in a pixel a negative value is attributed to this pixel.

Normally an empirical power-law relationship is used (R. E. Rinehart, 2004 [82]) to convert reflectivity into precipitation. This is, however, an approximation and a source of errors. But there are also numerous others error sources, which will be explained in the next Section.

### 2.2.2 Typical errors in radar precipitation estimates

Some typical problems of radar-based rainfall estimates are:

- **Anomalous propagation** (anaprop): In free space the path of a radar beam (a plane electromagnetic wave) propagates straight because dielectric permittivity $\varepsilon_0$ and magnetic permeability $\mu_0$ are constant.

The refractive index $n = (\mu_r \varepsilon_r)^{-1/2}$ is unity in vacuum; $\varepsilon_r$ is the relative dielectric permittivity and $\mu_r$ the relative magnetic permeability. If $\nu$ is the phase velocity of radiation of a specific frequency in a specific material, the refractive index is given by $n = \frac{c}{\nu}$ . The atmosphere's permittivity $\varepsilon = \varepsilon_0 \cdot \varepsilon_r$ is larger than $\varepsilon_0$ since the atmosphere is vertically stratified, microwaves propagate at speeds $\nu < c$. Even small variations in the refraction index can have remarkable effects in the electromagnetic wave propagation.

Water vapour particles have a permanent dipole momentum; therefore its refractive index depends on the wave frequency. For microwaves the relation can be parameterised by (following Bean and Dutton, 1968 [9]):

$$n = K_2 \frac{e}{T} + K_3 \frac{e}{T^2} \tag{2.1}$$

with $e$ the water vapour pressure in $hPa$, $K_2$ and $K_3$ are constants. If we take into account also the dry air, with pressure $P_d$, we can write

$$n = K_1 \frac{P_d}{T} + K_2 \frac{e}{T} + K_3 \frac{e}{T^2} \tag{2.2}$$

With $K_1$ a constant. With some approximations applied we have:

$$n = \frac{77.6}{T}(P + 4810\frac{e}{T}) \tag{2.3}$$

With P is the total pressure in $hPa$.

In the troposphere the temperature and pressure profiles decrease with the height. Since the pressure decrease is stronger than the temperature one, $n$ decreases with the height. Using the Snell' s law (Fig. 2.5), the wave front is beat downwards:

$$n_1 sin\theta_1 = n_2 sin\theta_2 \tag{2.4}$$

Only the first kilometres of the atmosphere are important for most radar meteorology applications, where the refractivity gradient is approximately $-40km^{-1}$ in standard conditions. In cases of temperature inversion and very humid air conditions, this value can be lower than $-157km^{-1}$, leading to anaprop (Bech et al., 1998 [10]).

- **Bright band**: Whether icy precipitation is from stratiform or convective clouds, it often falls into an environment where the temperature is above the freezing point. When snow melts into rain, radars typically observe

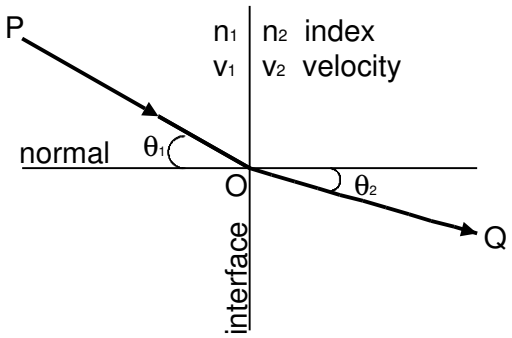


Figure 2.5: *Refraction of light at the interface between two media of different refractive indices, with $n_2 > n_1$. Since the velocity is lower in the second medium ($v_2 < v_1$), the angle of refraction $\theta_2$ is less than the angle of incidence $\theta_1$*

a narrow horizontal layer of stronger radar reflectivity that has been termed the *bright band* (Atlas, 1990 [7]).

The falling snowflakes are crossing the $0°C$ layer and melt from the outside. The reflectivity maximum is explained by the difference in the value of the dielectric factor of water and ice: when a water film begins to form on a melting snowflake, its radar reflectivity may increase by as much as 6.5 dB. The reflectivity decreases again below the melting level when the flakes collapse into raindrops; their fall velocities increase causing a decrease in the number of precipitation particles per unit volume. The size of the particles also becomes smaller in the melting process, as their density increases from that of the snow and melting snow to that of liquid water.

In case of strong convective currents, showers and thunderstorms tend to destroy the horizontal stratification essential for creating and sustaining the bright band. Normally the presence of the bright-band near the melting level is a signature that helps to distinguish convective mode from stratiform mode (Llsat et al., 2004 ).

- **Attenuation**: The intensity of electromagnetic radiation decreases due to attenuation, when passing through a medium (e.g. atmosphere, cloud, rain ...). This loss of radar energy is due to scattering and absorption. Because of attenuation, storms close to the radar are better sampled than storms far from the radar site. Small wavelength radar beams (i.e. X-band) attenuate more rapidly than long wavelength radar (i.e. S-band).

### 2.2.3 The composite

The composite products, e. g. by DWD, are more complicated then a simply assemblage of separate parts. In the overlapping areas of several radar sites the value with the best quality is selected. If the quality is the same then the one with the strongest signal is chosen.

The data of the national composite of DWD are exchanged with the data of the meteorological services of others European states (Belgium, Denmark, Austria, Switzerland, France, the Netherlands, Great Britain and the Czech Republic); in this way it is possible to build an international composite image. Unfortunately these data was not available for the current study.

## 2.3 Satellite data

In our work we need information about the cloud top height. For this reason we use products of the Satellite Application Facility on support to NoWCasting and Very Short-Range Forecasting (SAFNWC) (for further information see the SAFNWC manual, [72]). The SAFNWC products are based on the MSG SEVIRI data, generate by DWD and available with a temporal resolution of 15 min.

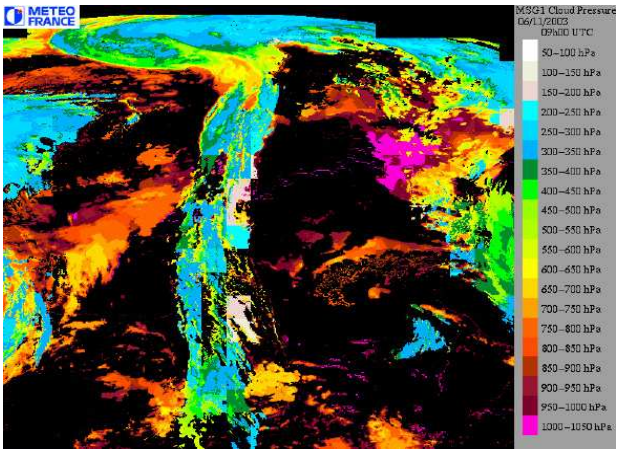


Figure 2.6: *CTTH data, image from http://nwcsaf.inm.es/*

The general objective of the SAFNWC is to provide operational services to ensure the optimum use of meteorological satellite data in Nowcasting and Very Short Range Forecasting to the targeted users. In our case we will use the data in order to define the cloud top height.

We use CTTH (Cloud Top Temperature and Height) and CT (Cloud Type).

### 2.3.1 Cloud top temperature and height

This product contains information about the cloud top temperature and height for all pixels identified as cloudy in the satellite scene.

For its determination the following process is applied:

- The RTTOV (Rapid Transmissions for *TOVS* [1]) radiative transfer model is implemented using NWP temperature and humidity vertical profile to simulate $6.2\mu m$, $7.3\mu m$, $10.8\mu m$, and $12.0\mu m$ cloud free and overcast (clouds successively on each RTTOV vertical pressure levels) radiances and brightness temperatures.
- Different techniques are employed for different cloud types in order to retrieve the best approximation for the cloud top pressure. The cloud top temperature is linear by interpolated using the temperature of the two nearest pressure levels in the vertical profile.
- General modules for pressure are used in order to define the cloud top height.

### 2.3.2 Cloud type

The CT algorithm provides a detailed cloud analysis; it may be used as input to an objective meso-scale analysis, as an intermediate product input to other products, or as a final image product. The CT product is also applied in the calculation of cloud top temperature and height. Different channels are used to determine CTTH depending on cloud type.

The CT classification algorithm is based on a sequence of thresholds tests; the cloud phase flag is not yet available.

The CT selects one of the following twenty-one categories :

1) not processed containing no data or corrupted data
2) cloud free land
3) cloud free sea
4) land contaminated by snow
5) sea contaminated by snow/ice
6) very low and cumuliform clouds
7) very low and stratiform clouds
8) low and cumuliform clouds
9) low and stratiform clouds

[1] TOVS instruments provide information on mainly vertical temperature and moisture distribution

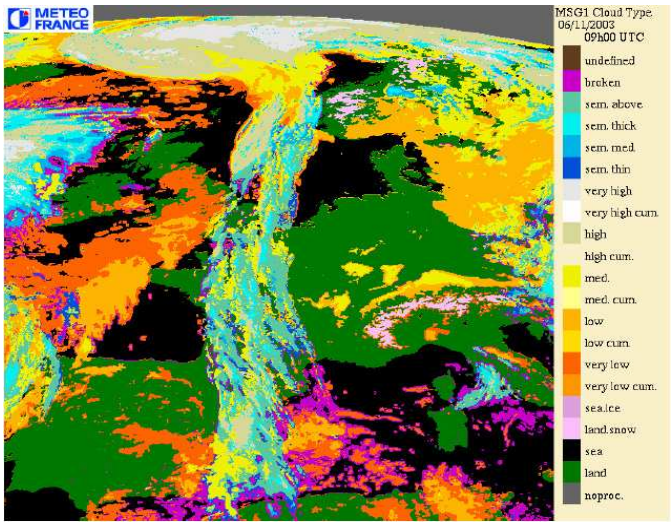


Figure 2.7: *CT data, image from http://nwcsaf.inm.es/*

10) medium and cumuliform clouds
11) medium and stratiform clouds
12) high opaque and cumuliform clouds
13) high opaque and stratiform clouds
14) very high opaque and cumuliform clouds
15) high opaque and stratiform clouds
16) high semitransparent thin clouds
17) high semitransparent thick clouds
18) high semitransparent above low or medium clouds
19) fractional clouds (sub-pixel water clouds)
20) undefined by cloud mask

The algorithm for the cloud type has some typical problems: e. g. low clouds surmounted by thin cirrus may be classified as medium clouds, very thin cirrus can be classified as fractional clouds and very low clouds (pressure larger than 800hPa) in a strong thermal inversion may be classified as medium clouds (pressure between 450hPa and 650hPa).

# Chapter 3

# The PIB algorithm

Horizontal humidity flux convergences in the lower part of a cloud tend to trigger updrafts and cumulus convection (Xi and Reuter, 1996 [102]). The result from many experimental studies (see Cotton and Anthes, 1989 [20]) was that the two quantities, which are best correlated with convective precipitation, are upward motion and moisture convergence. In his work, G. Haase (2002 [41]) asserts that in this context instability of atmospheric stratification is secondary. He investigated the relationship between stability indices (calculated from model variables) and precipitation; and did not find a direct connection.

The starting point of PIB (Physical Initialisation Bonn) is the analysis of the rain field using model precipitation and the rain rate product from the radar composite of the DWD (see Section 2.2.1). Satellite data (Section 2.3) are used in a simple cloud model; the PIB connects the observation space directly with model space.

A simplified precipitation scheme connects the observation space with model space. The vertical wind and the humidity are computed from this scheme.

The main difference between the original PIB developed by G. Haase and the current one is in the use of satellite data (SAFNWC see section 2.3) leading to a different definition of the cloud top and to the assimilation of non precipitating clouds. In addition we use a different definition of cloud base and of the analysed precipitation. We introduced a mixed phase and a more flexible definition for the conversion efficiency of saturated water vapour into rain. In the original version it was defined at the first time step for the whole simulated area.

In this Chapter we describe the PIB algorithm including the data processing from the analysed precipitation (Section 3.1) to the modification of COSMO variables (Section 3.4). In Sections 3.2 and 3.3 the assumed precipitation scheme and the cloud model are described. A scheme of the PIB algorithm is

shown in Fig. 3.5.

## 3.1 Analysed precipitation

For the determination of the analysed precipitation the first step is the measurement of the observed precipitation. After that the scheme attributes a time weight to the analysed precipitation.

For the formulation of the observed precipitation at every time step and for every grid point a linear interpolation between the adjacent measurements to the current model time step is made. The scheme takes into account only the radar data in a time range of 10 minutes (5 before and 5 after the current time step).

Upon the availability of radar observation directly after the current model time step and before it, the time weight ($\alpha$, eq. 3.1) is set to one. In other cases the time weight is reduced depending on the time distance of the observation to the current time step. In other words, if one measurement is missing either at the last or the following observation time, the available data is used but with a reduced $\alpha$ depending on the difference between the observation time and the current model time.

In COSMO model we use the prognostic precipitation scheme where precipitation can be horizontally advected when falling from the cloud base to the surface. Thus the position of the precipitation at the soil can be shifted relative to the cloud base. The position of the observed precipitation is defined at the height of the lowest radar beam which can be approximated by the cloud base. We avoid this problem, due to the use of different heights, by using the model precipitation at the cloud base.

For every time step and every grid point the analysed precipitation $R_{ana}$ is calculated using:

$$R_{ana} = \alpha R_{rad} + (1. - \alpha) R_{mod} \tag{3.1}$$

$R_{rad}$ is the radar precipitation flux and $R_{mod}$ the model precipitation flux.

The model rainfall rates are provided by the COSMO grid-scale cloud and precipitation scheme. If $\alpha$ is small (large), $R_{ana}$ is approximated by $R_{mod}$ ($R_{rad}$). Radar data and model field are mixed only if the temporal availability of observations is inadequate. In this case measurements can be merged partly with model values, because the latter are already affected by assimilation. Finally, the analysed precipitation rates in $[mm/h]$ are transformed to fluxes in $[kg \cdot (m^2/s)^{-1}]$ which are used by COSMO as diagnostic output. The $R_{ana}$ field can now be used as input to the one-dimensional cloud model (Section

3.3).

## 3.2 Precipitation scheme

The PIB uses a one-dimensional cloud model with the aim to generate COSMO-consistent vertical wind profiles. PIB consequently reproduces the precipitation rates available from our radar products, which are used also in the verification. In other words, the validation of the COSMO output field is made by a comparison with the same type of data that we assimilate. We want to check the ability of PIB to change the model atmosphere to a state similar to the radar state. We assume the radar data to be perfect.

The PIB cloud scheme takes into account both liquid water and ice, when calculating the water vapour saturation mixing between $-5°C$ and $-25°C$. In this temperature region the fraction of liquid water and ice is calculated as a function of temperature, and the saturated vapour pressure is determined as a weighted mean of the saturated vapour pressure over ice and water (see Kuell et al., 2007 [60]). Between cloud bottom and cloud top the specific humidity is set to its saturation level.

In this Chapter we describe the precipitation scheme using only the liquid phase for simplicity.

PIB starts using the partial densities of saturated water vapour ($\rho_v^*$) and rain water ($\rho_r$), both in $[kg/m^3]$. The latter is in the form of liquid-phase drops, which are large enough to have a non-negligible fall velocity.

The continuity equations for $\rho_v^*$ and $\rho_r$, taking into account the conversion rates, build the basic framework of the scheme and can be written as:

$$\frac{\partial \rho_v^*}{\partial t} = -\vec{\nabla}(\rho_v^* \vec{v}) - S_{con} \tag{3.2}$$

$$\frac{\partial \rho_r}{\partial t} = -\vec{\nabla}(\rho_r \vec{v_r}) - S_{ev} \tag{3.3}$$

where $\vec{v}$ and $\vec{v_r}$ are the three dimensional velocity vectors of water saturated air and rain water, respectively. The terms $S_{con}$ and $S_{ev}$ are the conversion rates due to condensation and evaporation, respectively. Assuming that the microphysical auto conversion process of cloud water into rain water has an efficiency of 100% additional water vapour is instantaneously converted into rain water. In case of saturation equilibrium within water clouds, both conversion rates ($S_{con}$ and $S_{ev}$) has the same absolute value. Another assumption is stationarity, which is also applied in the COSMO grid scale cloud and precipitation parameterization scheme. Adding equations 3.2 and

3.3 under these assumptions and splitting each term into a horizontal and a vertical part, leads to:

$$\vec{\nabla}_h(\rho_v^*\vec{v}_h) + \frac{\partial \rho_v^* \hat{w}}{\partial z} + \vec{\nabla}_h(\rho_r \vec{v}_{h,r}) + \frac{\partial \rho_r w_r}{\partial z} = 0 \tag{3.4}$$

where $(\rho_r w_r)$ is the precipitation flux, $R$ in $[kg/(m^2 \cdot s^{-1})]$. The sign of $R$ is the same as the one of the vertical velocity of rain water $(w_r)$ because $\rho_r$ is always positive. $R$ and $w_r$ are defined positive in downward direction. Neglecting horizontal transport of rain water (i.e. condensate falls out instantaneously and completely) we have:

$$\vec{\nabla}_h(\rho_v^*\vec{v}_h) + \frac{\partial \rho_v^* \hat{w}}{\partial z} = -\frac{\partial R}{\partial z} \tag{3.5}$$

Since the derivation of vertical wind is the main goal of the cloud model, a precipitation flux profile has to be assumed (the right part of eq. 3.5). Currently, no observational information is available about its actual vertical structure, because available radar data do not include the complete volume scans. Therefore, a synthetic first guess profile yet has to be assumed. It has two fixed points: at the cloud top $(z_{ct})$ the precipitation flux is zero and at the cloud base $(z_{cb})$ it is derived from the analysed precipitation flux ($R_{ana}$ at the cloud base, see Section 3.1). Inside the cloud, the precipitation flux is assumed to change linearly:

$$R(z) = \begin{cases} 0 & : \quad z = z_{ct} \\ R(z_{cb}) + \frac{R(z_{ct}) - R(z_{cb})}{z_{ct} - z_{cb}}(z - z_{cb}) & : \quad z_{ct} > z > z_{cb} \\ R_{ana}(z_{cb}) & : \quad z = z_{cb} \end{cases} \tag{3.6}$$

All microphysical interactions beneath the cloud base (e.g. evaporation of rain water) are neglected. The use of radar derived precipitation profiles instead of the above linear approximation has a serious drawback: the reflectivity structure of a rain shower, e.g. measured by a vertically pointing Doppler radar, varies extremely in time and space (Rogers and Yau, 1989 [83]). Therefore, profiles have to be filtered before they can be applied to the cloud model. Additionally, the measured reflectivity factor cannot be subdivided into components caused by different hydrometeor types unless the radar has polarimetric capabilities. Another problem is that the mid-latitude convective storms are dominated by the ice phase.

From the PhD thesis of G. Haase ([41]) we have some information about COSMO precipitation flux profiles of the one-category ice scheme (stationarity and homogeneity assumed). In his study parts of the horizontally averaged rain rate profiles show a linear behaviour, at least inside the cloud. Beneath the cloud (approximately 1500 m ASL) evaporation effects are visible.

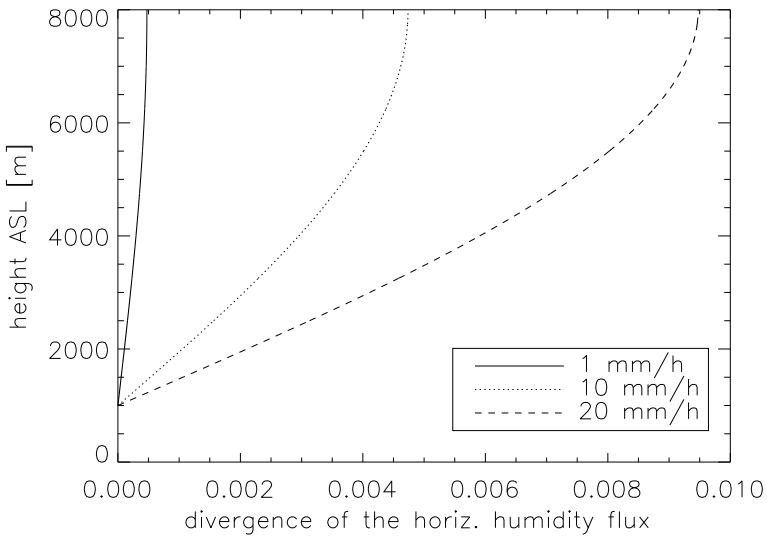


Figure 3.1: *Profiles of horizontal humidity flux divergences ($kg/(m^3 \cdot s)$) according to eq 3.7, using different rain rate at the cloud base. the conversion efficiency defined in eq. 3.11 is 0.9; cloud base and top heights are 1000m and 8000m, respectively.*

Summing up, the use of COSMO-consistent linear profiles is favourable, but as soon as three-dimensional polarimetric measurements from the DWD radar network are available, this approach should be reconsidered.

The vertical advection contributes to create the clouds. The cloud top is defined by the height where no further advection is present and at this level the vertical wind must vanish. In order to derive an equation for the vertical wind, the divergence of the horizontal flux of saturated water vapour in eq. 3.5 must satisfy this condition; the result follow on eq. 3.9; a sinusoidal profile meets all requirements:

$$\vec{\nabla}_h(\rho_v^* \vec{v}_h) = \tilde{c} \sin\left(\frac{\pi}{2} \frac{z - z_{cb}}{z_{ct} - z_{cb}}\right) \tag{3.7}$$

where

$$\tilde{c} = \frac{R(z_{cb})}{z_{ct} - z_{cb}} \frac{\pi}{2} \left(1 + \frac{\rho_v^*(z_{cb}) \hat{w}(z_{cb})}{R(z_{cb})}\right) \tag{3.8}$$

It has to be mentioned that a cosine approach would provide a similar profile characteristic as the sinusoidal one. The divergence of the horizontal flux of saturated water vapour vanishes at the cloud base and becomes maximum at cloud top (Fig. 3.1). The approximation in Eq. 3.7 is confirmed by simulations

with the non-hydrostatic *Mesoscale Model* (MM5) performed by Fehr (2000, [36]).

Inserting Eqs. 3.6 and 3.7 in Eq. 3.5 inside the cloud leads to:

$$\begin{aligned}\frac{\partial \rho_v^* \hat{w}}{\partial z} &= \frac{R(z_{cb})}{z_{ct}-z_{cb}} \\ &\left[1-\frac{\pi}{2}\left(1+\frac{\rho_v^*(z_{cb})\hat{w}(z_{cb})}{R(z_{cb})}\right) sin\left(\frac{\pi}{2}\frac{z-z_{cb}}{z_{ct}-z_{cb}}\right)\right] \end{aligned} \tag{3.9}$$

After the discretisation of 3.9 an expression for the vertical wind inside the cloud at model level $k$ can be derived:

$$\begin{aligned}\hat{w}_k &= (\rho_{v,k}^*)^{-1}\left\{\rho_{v,k-1}^*\hat{w}_{k-1}-(z_{k-1}-z_k)\frac{R(z_{cb})}{z_{ct}-z_{cb}}\right. \\ &\left.\left[1-\frac{\pi}{2}\left(1+\frac{1}{c}\right) sin\left(\frac{\pi}{2}\frac{z_{k-1/2}-z_{cb}}{z_{ct}-z_{cb}}\right)\right]\right\} \\ with& \\ c &= \left.\frac{R}{\rho_v^*\hat{w}}\right|_{z=z_{cb}} \qquad [0,1]\end{aligned} \tag{3.10}$$

The level $(k-1/2)$ is located between levels $k$ and $(k-1)$. The parameter $c$ is a measure for the *conversion efficiency of saturated water vapour into rain water at cloud base* . This is the closure assumption of the cloud model and characterizes the shape of the vertical wind profile. Assuming a constant rain rate, large (small) values of c mean high (low) conversion efficiencies resulting in moderate (strong) vertical updrafts.

The cloud model includes several assumptions and simplifications, the only tuning parameter is $c$. This parameter is set using a search algorithm in the nearby model points or dynamically adjusted (see par. 3.2.1). Some idealised vertical wind profiles for different rain rates and conversion efficiencies are shown in Fig. 3.2. These profiles strongly depend on the partial density of saturated water vapour profiles. According to this dependence we took $\rho_v^*$ from the model precipitation point in order to calculate $w$.

We can note that the maximum wind speed grows with increasing conversion efficiency and, from our calculation, also with cloud depth. For the explanation we take a deep convective (cumulus clouds) case, with estimated cloud top height of 8000 m. A strong vertical wind located in the upper third of the cloud is well represented. As clearly seen Fig. 3.2, the max vertical wind is near $20m/s$ for strong precipitation (20 mm/h) and $c = 0.1$.

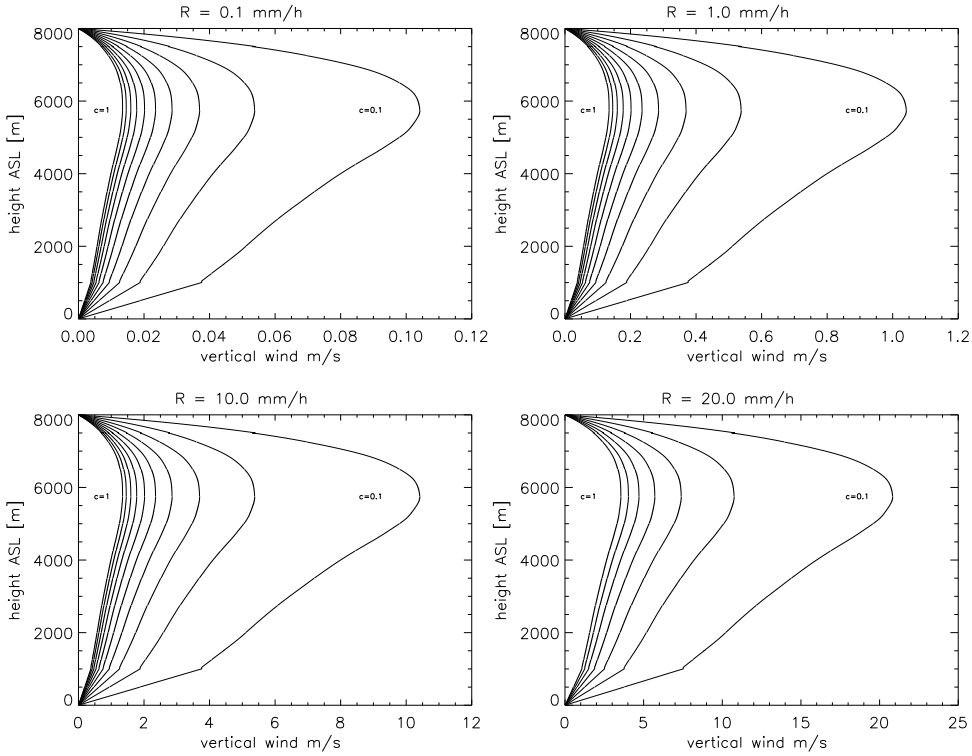


Figure 3.2: *Vertical wind profiles for different rain rates(from 0.1 mm/h to 20 mm/h) and conversion efficiencies (from 0.1 to 1 with 0.1 step size) from the simplified precipitation scheme. In this case we have artificial set the cloud base to 1000m and the cloud top to 8000m. The essential partial density of saturated water vapour is taken from a COSMO run.*

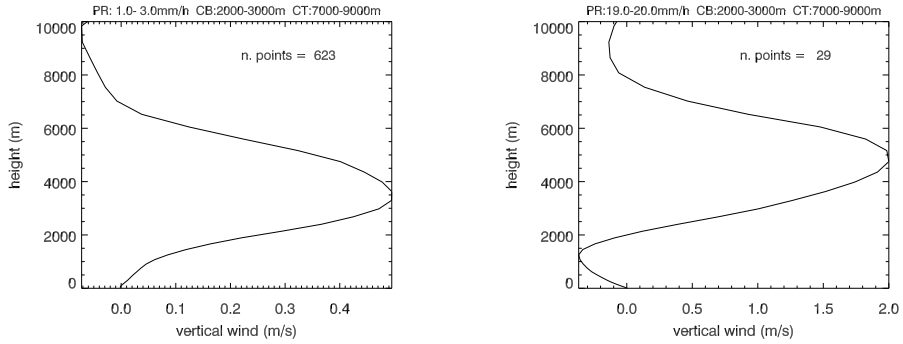


Figure 3.3: *COSMO reference run vertical wind profiles. Mean of the vertical wind for the points with a precipitation at the cloud base of about 0.1 mm/h on the left side and 20 mm/h on the right side.*

In Fig. 3.3 average vertical wind profiles simulated by the COSMO reference run without the use of convection parameterization are plotted. Averages have been made over profiles are from a run of 5 hours with an output every 5 minutes. We compare clouds and precipitation with similar characteristics as in Fig. 3.2. In the COSMO wind profiles (Fig. 3.3) we took the profiles with cloud base between $2000m$ and $3000m$ and cloud top between $7000m$ and $9000m$. Two rain rate intervals are taken into account, in the range of one and three millimetres per hour (plot on the left side) and between 19 and 20 mm/h (plot on the right side).

The curves on fig. 3.3 and Fig. 3.2 look similar, the location of the maximum wind speed is lower for the COSMO reference run especially for the higher rain rate, maybe due to the low number of points that we could find with strong precipitation. For strong rain rates a weak downdraft (maximal $-0.3m/s$) is present in the region below the cloud base.

Negative velocities are not present in the one-dimensional cloud model. We are not assimilating the downdraft but only the updraft for the precipitative clouds. We assumed that the downdraft is weaker than the updraft (Emanuel, 1994 [35]). This relationship and the consistency with the model vertical wind is proven in Chapter 5. In case of no precipitating clouds the vertical wind is set to 0.

The results of the simplified precipitation scheme demonstrate that the simple approach meets all requirements and provides COSMO-consistent vertical wind profiles similar to observations (Rogers and Yau 1989 [83], Shibagaki et al. 2000 [86] , Xu and Randall 2001 [103], Stull, 1999 [88]) for convection in continental areas (first statistics from Byers and Braham, 1949 [15]). However, all these studies were made with the motivation of improving cumulus parameterization and not an initialisation scheme.

### 3.2.1 Conversion efficiency determination

According to eq. 3.11, $c$ is important for the determination of the profile shape and the intensity of the vertical wind. Thus $c$ is adapted in a dynamical way using a search algorithm taking into account the quality of the forecast. The main idea is that columns close to each other with similar precipitation rate also have similar values of $c$.

For the estimation of $c$, a constant background field (with a value of 0.4) is chosen; this value matches best the mean observed vertical wind profile for raining clouds (Haase et al., 2000 [42]). Starting from this value, $c$ is changed in order to move the model state into the direction of the observation state. The first step is the definition of a *good forecasted point* using the criterion:

$$R_{ana} - 20\%R_{ana} < R_{mod} < R_{ana} + 20\%R_{ana} \tag{3.11}$$

Here $R_{mod}$ is the model precipitation at lowest radar height for the considered grid point and $R_{ana}$ the analysed precipitation.

In the model columns, where eq. 3.11 is fulfilled, $c$ is not changed. In the other columns we search in the surrounding area (within a radius of 5 points) for the column with the best matching rain rate. In this case $c$ is calculated using the model variables (precipitation and wind both at the cloud base), the resulting model-consistent parameter is then applied.

If no columns satisfy eq. 3.11 we apply a dynamical determination of the conversion efficiency. The adjustment is made using the difference between analysed precipitation and the model precipitation (normalized with the sum of the two precipitation rates), and the value of the parameter at the previous time step in order to enhance or reduce the vertical wind (and then the precipitation) to meet the requirements in eq. 3.11.

The conversion efficiency is adapted using:

$$c(n) = c(n-1)\left[1 + 0.25 \cdot sin\left(\frac{\pi}{2}\frac{R_{ana} - R_{mod}}{R_{ana} + R_{mod}}\right)\right] \tag{3.12}$$

Here $c(n)$ is the conversion efficiency at the time step $n$. A sinusoidal profile is used. The conversion efficiency is confined to the interval:

$$0.25 < c(n) < 1 \tag{3.13}$$

## 3.3 Cloud Analysis

In this Section we describe how the lower and upper boundaries for the clouds are set. Normally the cloud base and cloud top heights derived from the COSMO itself are not in agreement with the observation. On the other hand, the use of observed clouds with their properties is often thermodynamically inconsistent with the model state. The intention of the PIB is to assimilate clouds in accordance with the observations but without losing the model balance.

### 3.3.1 Cloud base height

The first assumption is a well mixed atmospheric boundary layer. Under this condition lifting condensation level (LCL) and convective condensation level

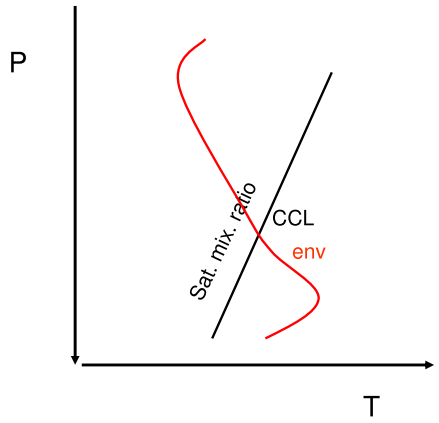


Figure 3.4: *Convective condensation level*

(CCL) are identical (Rogers and Yau, 1989 [83]). The LCL is defined as the level to which a parcel of moist air can be lifted adiabatically until it becomes saturated (Wallace and Hobbs, 1977 [95]). The CCL on a thermodynamic diagram is defined as the point of intersection of a sounding curve (representing the vertical distribution of temperature in an atmospheric column, red line in Fig. 3.4) with the saturation mixing ratio line corresponding to the average mixing ratio in the surface layer (black line). In both cases the air parcel is displaced to the level where it will become saturated as a result of its adiabatic cooling, for the LCL the displacement depend on mechanical lifting while for the CCL is the results of turbulent circulations forced by surface heating (Holton et al., 2003 [46]).

PIB calculates the LCL using the model variables and identifies it with the cloud base. For Craven et al. (2002, [23]) a clearly wrong calculation for the LCL is possible if the surface temperature and/or the dewpoint is not representative of the thermodynamic profile in the boundary layer. Thus, PIB uses the mean temperature and the mean dewpoint temperature calculated from the model variables in the lowest 100-hPa layer; the choice of 100-hPa is recommended from the results of Craven et al..

The calculation of the temperature at the LCL ($T_L$) uses the temperature ($T$) and the dew point temperature ($T_D$) of the mixed layer, consequently (Bolton, 1980 [12]):

$$T_L = \frac{1}{\frac{1}{T_D - 56} + \frac{ln(T/T_D)}{800}} + 56 \tag{3.14}$$

all temperature values are in Kelvin.

From the temperature of the LCL the cloud base model level is searched.

Another approximation of the cloud base is derived from the calculation of

the CCL. An air parcel with environmental properties (plus a small excess temperature) is lifted dry-adiabatically from the lowest model level. The convective condensation level is then defined to be the (full) level where the parcel becomes saturated and its buoyancy is greater or equal zero. Also in this case a mixed layer of 100-hPa is used.

Both estimates of the cloud base are nearly independent so we assume that the average of both has a smaller error than the individual estimates.

The determination of the cloud base height is a very sensible parameter for the PIB; an error in its definition leads to a different CAPE, since not only the vertical wind is changed but also saturation occurs at different levels, for the levels from cloud base to cloud top; therefore the dynamics of the atmosphere is strongly influenced.

Forsythe et al. (2000, [37]) proposed a combination of satellite cloud classification with surface cloud reports, which results in a superior estimate of cloud base height (especially for broken cloudiness) compared with an estimate using only interpolated surface data.

### 3.3.2 Cloud top height

The cloud top height is taken from the SAFNWC product from the DWD (see Section 2.3). The assimilation is made every 15 minutes and the cloud top height is left constant in the time between two assimilations.

An algorithm for the definition of a cloud top height depends on the cloud type (for the cloud type definition see Section 2.3.2). The satellite data are interpolated to the model grid and the algorithm assigns the following values to each grid point:

- -1 for missing data.
- -2 for no cloud type or cloud type with a low water content.
- -3 in case high clouds obscure the top of the precipitating clouds.
- the cloud top height in the other cases.

An important part in the PIB algorithm (Fig. 3.5) is the assimilation of the precipitating grid points. These are the points where the analysed precipitation is greater than 0.1 mm/h. In areas where the radar data indicate precipitation and where the cloud top height is set to a value less than 0, an optimal interpolation algorithm is used to fill these gaps (Ament, 2001 [3]). The interpolation algorithm follows the general theory in Daley, 1991 [26] (see Chapter 1.2.1). The background cloud top height is set to $8000m$, which

is approximately the average height for a convective cloud (Lopez, 1977 [67], Sherwood et al., 2004 [85]).

An optimal interpolation algorithm aims to minimize the excepted analysis error variance as a function of the weights (from eq. 1.3 ), $i$ and $k$ are the analysis and the observation grid points, respectively:

$$\frac{\partial(A_i - T_i)^2}{\partial W_{ik}} = 0 \tag{3.15}$$

where $A_i = B_i + \sum_{k=1}^{K} W_{ik}[O_k - B_k]$ (for notation see Chapter 1.2.1). Subtracting the true values leads to:

$$A_i - T_i = B_i - T_i + \sum_{k=1}^{K} W_{ik}[O_k - B_k] \tag{3.16}$$

We assume that both the background field and the observations are unbiased, $< \cdots >$ is the expected value operator:

$$< B_i - T_i >=< B_k - T_k >=< O_k - T_k >= 0 \tag{3.17}$$

Thus $< O_k - B_k >=< (O_k - T_k) - (B_k - T_k) >$. This implies from 3.16 that the analysed value $A_i$ is also unbiased. With these assumptions the analysis error at grid point $i$, $E_A^2$ is given by:

$$\begin{aligned} E_A^2 &= < (A_i - T_i)^2 >=< (B_i - T_i)^2 > +2 \cdot \sum_{k=1}^{K} W_{ik} < (O_k - B_k)(B_i - T_i) > \\ &+ \sum_{k=1}^{K}\sum_{l=1}^{K} W_{ik}W_{il} < (O_k - B_k)(O_l - B_l) > \end{aligned} \tag{3.18}$$

$< (O_k - B_k)(O_l - B_l) >$ is the covariance between observation increments at observation location $r_k$ and $r_l$. $< (O_k - B_k)(B_i - T_i) >$ is the covariance between background error at $r_i$ and the observation increment at $r_k$. Using eq. 3.15, we obtain:

$$\sum_{l=1}^{N} W_{il} < (O_k - B_k)(O_l - B_l) >= - < (O_k - B_k)(B_i - T_i) > \tag{3.19}$$

The term at the right side of the eq. 3.19 is determined using :

$$< (O_k - B_k)(B_i - T_i) >= C_f|\vec{r}_k - \vec{r}_l| \tag{3.20}$$

The value for $C_f$ is set to:

$$C_f = exp(-0.001841 \cdot |\vec{r}_k - \vec{r}_l|) \tag{3.21}$$

The parameterisation of $C_f$ is based on an experiment performed in August 1998 using the analysis of the cloud data for the whole month, in an analysis by F. Ament, 2001.

At this point an analysis of the cloud top height is available. The model cloud top height derived from the cloud water content profile is nudged to the analysed cloud top height, with a nudging coefficient of 0.5:

$$ctlev = \text{INT}\left(\frac{ctlev_{mod} + ctlev}{2}\right) + 1 \tag{3.22}$$

where $ctlev_{mod}$ is the model level closest to the cloud top, and $ctlev$ is the cloud top from the analysis.

### 3.3.3 Corrections

A difference between shallow and deep convection is made using the results from Bechtold et al. (2001, [11]), who give some parameters for the minimum cloud depth in case of deep convection.

Since the cloud top can be reasonably retrieved from satellite measurements, we assume that the definition of the cloud base can lead to the bigger errors. Since cloud depth is a very sensitive parameter in PIB, we use a correction for cases when the calculated cloud base and top height (see cap. 3.3.1 and 3.3.2) are inconsistent with the model dynamics.

For every cloudy grid point we check whether shallow or deep convection is expected. The deep convection grid points are defined, in the first approximation, as the points where precipitation greater than 5 mm/h occurs. For this decision we apply the criteria of the convection parameterisation by Tiedtke (1989, [92]), i.e. we assume that deep convection takes place in connection with strong low-level convergent flows. Thus we use a parameterisation based on the three dimensional humidity convergence ($Q$) and the boundary layer turbulence diffusion ($\Omega$). Deep convection in a grid point occurs in case of:

$$\frac{g}{R_d}\int_0^H \frac{Q \cdot p}{T} dz > -1.1 \cdot \Omega \cdot g \tag{3.23}$$

where $R_d$ is the gas constant for dry air, $g$ the acceleration of gravity, $H$ the height of the model atmosphere, $p$ the pressure, $T$ the temperature, $z$ the height.

For deep convection the cloud depth should be at least 250 hPa (Bechtold et al., 2001 [11]). If the differences between cloud base and cloud top is less, the cloud base is lowered in order to match this thickness. Similarly, if shallow convection is analysed and the difference between the cloud boundaries is too large, the cloud base is raised until the cloud thickness is less than 250 hPa.

## 3.4 Modification of model profiles

After introducing the simplified precipitation scheme, including its input parameters (par. 3.2), the model is now applied to assimilate the vertical wind profiles into COSMO. Additionally, the fields of specific water vapour, cloud water content and cloud ice content are adjusted (see Tab. 3.1). PIB is applied only when the difference between model precipitation and analysed precipitation exceeds 20%. This threshold is a rough guess of the minimal uncertainty of the radar-based precipitation estimate.

The applied modification depends on the intensity of the analysed precipitation rate (eq. 3.6) and on the vertical location of the model grid point. When PIB is used a threshold for the precipitation rates ($R_{th}$) of 0.1 mm/h is assumed. Therefore, at grid points where radar data are available with an analysed rainfall rate of 0.1 mm/h or less, model precipitation has to be suppressed taking into account the presence of clouds from the satellite data. PIB (as described in table 3.1) divides the 1D vertical atmosphere in three Sections: a subcloud layer ($z < z_{cb}$), a cloud layer ($z_{ct} \leq z \leq z_{cb}$) and an above-cloud layer ($z > z_{ct}$).

The PIB scheme changes the model variables considering:

- Changes made to the COSMO fields should be as small as possible.
- The model fields should be drawn to an atmospheric state comparable with the observations.

### 3.4.1 Forcing of precipitation

Using the analysed rainfall rate, grid points with precipitation greater than 0.1 mm/h are forced to the analysed precipitation. To achieve this purpose the vertical wind in the entire column is changed. Inside the cloud (from $z_{cb}$ to $z_{ct}$) $w$ is replaced using the artificial profile described in Section 3.2, eq. 3.11. Since the upper boundary of the cloud is defined by the zero-buoyancy level, vertical wind above the cloud top height is set to zero. Between cloud

Table 3.1: *Modification of prognostic COSMO variables $w$, $q_v$ and $q_c$ depending on the intensity of the analysed surface precipitation rate , see the calculation of $\hat{w}$ eq. 3.11, and the vertical location of the model grid point, $z_{srf}$ is the height of the surface above the sea level. $R_{ana}$ is the analysed precipitation, $R_{th}$ a threshold to determine the rain area 0.1 mm/h, $q_v^*$ specific humidity at the saturation point, qcmin is the threshold for stratocumulus $0.24g \cdot kg^{-1}$. Outside of the radar area the variables are not modified.*

| **height** | $R_{ana} > R_{th}$ | $R_{ana} < R_{th}$ | |
|---|---|---|---|
| | | cloud area | no cloud area |
| $z > z_{ct}$ | $w = 0$<br>$rh = min(rh, 99\%)$<br>$q_c = 0$<br>$q_i = 0$ | $w = 0$<br>$rh = min(rh, 99\%)$<br>$q_c = 0$<br>$q_i = 0$ | $w$ : not modified<br>$rh = min(rh, 99\%)$<br>$q_c = 0$<br>$q_i = 0$ |
| $z_{ct} \geq z \geq z_{cb}$ | $w = \hat{w}(z)$<br>$q_v = q_v^*$<br>$q_c$ : not modified<br>$q_i$ : not modified | $w = 0$<br>$q_v = q_v^*$<br>$q_c = min(qc, qcmin)$<br>$q_i = 0$ | $w$ : not modified<br>$rh = min(rh, 99\%)$<br>$q_c = 0$<br>$q_i = 0$ |
| $z < z_{cb}$ | $w = \hat{w}(z_{cb})\frac{z - z_{srf}}{z_{cb} - z_{srf}}$<br>$q_v = q_v^*(z_{cb})$<br>$q_c = 0$<br>$q_i = 0$ | $w = 0$<br>$q_v = q_v^*(z_{cb})$<br>$q_c = 0$<br>$q_i = 0$ | $w$ : not modified<br>$rh = min(rh, 99\%)$<br>$q_c = 0$<br>$q_i = 0$ |
| | | | |

base and surface the vertical wind is replaced by a linearly decreasing updraft (from the value at the cloud base to 0 at the surface):

$$w(z) = \hat{w}(z_{cb})\frac{z - z_{srf}}{z_{cb} - z_{srf}} \mid_{z<z_{cb}} \tag{3.24}$$

with $z_{srf}$ the height of the surface above sea level (Fig. 3.2 for the vertical wind profiles). The change of the vertical wind in the model could create instabilities in the model dynamics. In order to minimize the formation of gravity waves Newtonian nudging is used.

The maximum relative humidity above the cloud top is fixed to 99%. Therefore the maximum of the specific water vapour content above the cloud top is calculated by:

$$q_v = \frac{0.622 \cdot 0.99e^*}{p - 0.378 \cdot 0.99e^*} \tag{3.25}$$

Inside the cloud the air is assumed to be saturated ($q_v = q_v^*$, where $q_v^*$ is the

saturation specific humidity). Below the cloud base the specific humidity is equal to the saturation specific humidity at the cloud base: $q_v = q_v^*(z_{cb})$. It is assumed that the specific water vapour content is a conservative quantity in the PBL (Stull, 1999 [88]).

The specific cloud water and cloud ice content above the cloud top and below the cloud base are set to zero. Inside the assimilated cloud, water and ice contents are not modified. Therefore the COSMO is forced to generate clouds according to the prescribed specific humidity distribution.

### 3.4.2 Suppression of precipitation

In the case of no radar data (or missing data) the PIB is not applied, and all profiles remain unchanged. No radar data are grid points where the composite has no information, while missing data are points where the radar data are not available for a limited time.

For grid points where the analysed surface precipitation rate is equal or less than 0.1 mm/h, PIB suppresses precipitation taking into account the presence of clouds (derived from satellite data).

At grid points where no cloud is observed the model vertical wind field is not changed, but the maximum relative humidity is reduced to 99%. The cloud water and ice content is set to 0 in order to remove the source of precipitation formation consistent with the previous request.

For the cloud area the vertical wind is set to 0 in the entire column, and in the cloud the specific cloud water content has a maximum of $0.24 g/kg$. This value reflects a usual property of stratocumulus (Miles et al., 2000 [74]). Generally PIB tries to build broad regions of gentle downdraft, or vertical wind set to 0, surrounding smaller regions of strong updraft, as described in Stull, 1999 [88].

By this assumptions the model precipitation is suppressed at grid points with an analysed surface rainfall rate of 0.1 mm/h or less without influencing the nearby assimilated clouds.

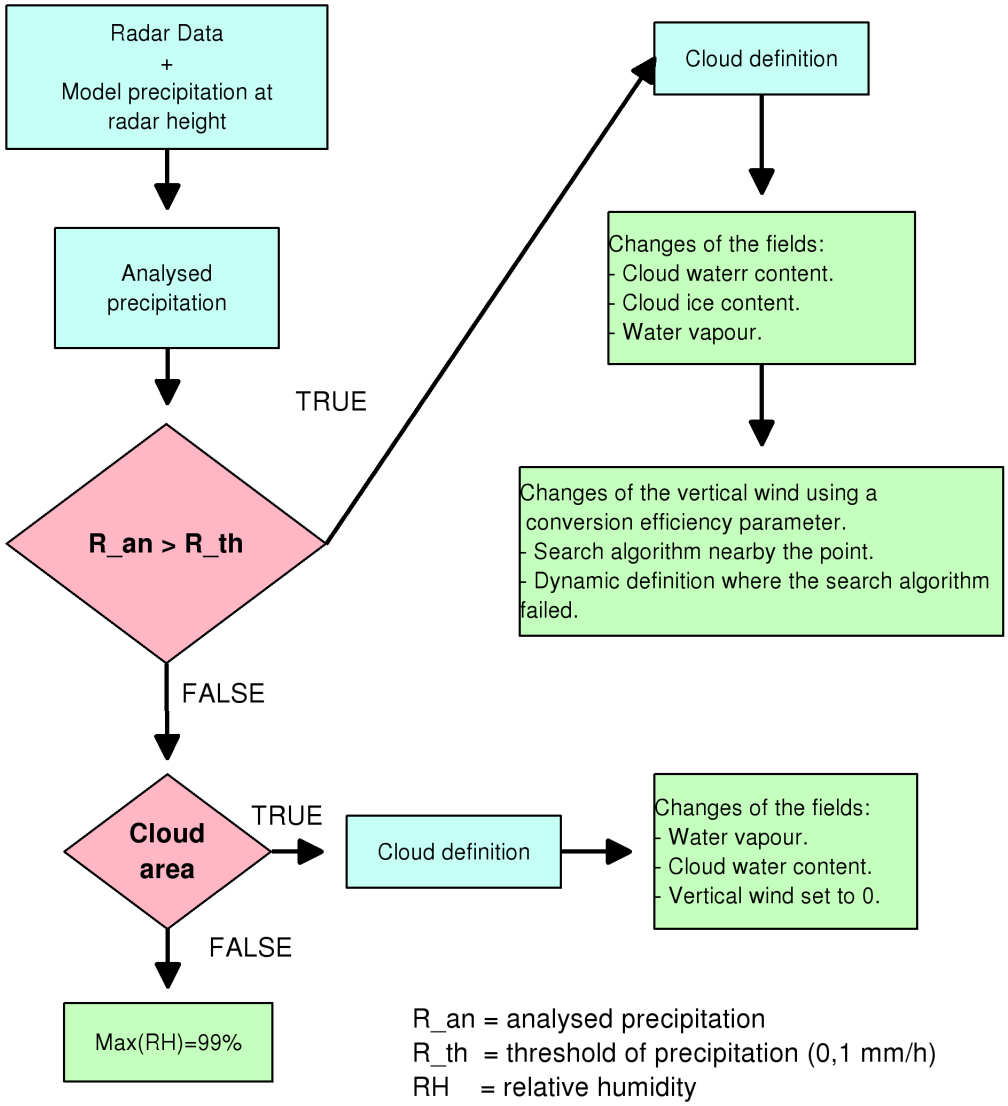


Figure 3.5: *PIB scheme*

# Chapter 4

# Identical twin

In this Chapter an identical twin experiment is performed with the aim to check the possible existence of inconsistencies between the assimilation scheme and the NWP model in generating precipitation.

In an identical twin experiment (a broad literature to this issue exists, e. g. Elbern et al. 1997 [33], De Lima Nascimento and Drögemeier 2006 [27], Orlandi et al. 2001, [76], Jiang et al. 2002 [48] ) the model is first run without any assimilation (*control run* , CTL). The output from CTL is used to generate a set of pseudo observations (the reference atmosphere), in our case the fields of rain rate and cloud top height. In the second step the model is run while assimilating these pseudo observations (*identical twin run*, IDTWIN). For a good description of the identical twin experiment theory see Arnold and Dey (1986, [6]) where the development of this technique through 30 years is described.

In such experiments both the model and observation error are added to the pseudo observations before assimilation in order to judge how well the analyses fit the reference atmosphere. In our case we only check the existence of imbalances and the consistency of PIB-modified model variables with the reference atmosphere. PIB is therefore applied without adding errors. A perfect assimilation algorithm must be consistent with the NWP model, it must not change the model field during this type of experiment. However a real assimilation algorithm changes the model fields and its physics. The IDTWIN run differs from the CTL because the pseudo-observations are generated only from a restricted CTL output to mimic the availability of the observations in space and time, i.e. 5 minutes for the radar data and 15 minutes for the satellite data.

A comparison of the control and the identical twin runs allows to investigate the properties of the assimilation algorithms, setting aside the observation errors. During the initialisation window, the improvements from the PIB procedure

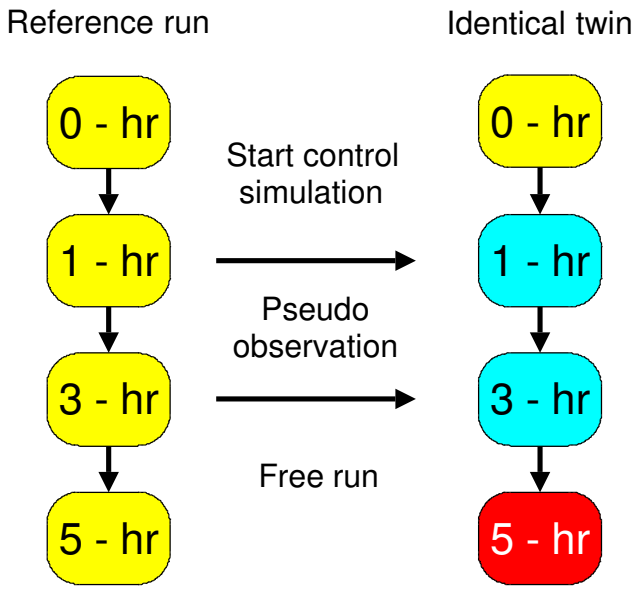


Figure 4.1: Experiment methodology

are evaluated by analysing the time evolution of the differences between the state of the atmosphere simulated in the two runs. We require for our scheme that the differences between the forecasted fields remain small (something like a near approximation to a perfect algorithm), also at the end of the forecast period. The model fields for the two runs will have always differences at the end of the initialisation period because the assimilation scheme is not perfect consistent with the model. The two free forecasts start from two different states.

In our case the numerical model (COSMO model 4.6) is first integrated from the initial conditions over 5 hours for the reference run. Then the pseudo observations are assimilated into a run that has an hour of control run, followed by a two hours assimilation window, and then by another two hours of free forecast (as elucidated in the Fig. 4.1). Normally PIB modifies only those columns for which the difference between analysed and modelled precipitation is larger than 20%; this constraint is not applied for the identical twin experiment, because in that case PIB would not apply any action; CTL and IDTWIN would be identical. We change the model variables using PIB for every model time step during the assimilation window. We do not apply PIB for those points where both CTL and IDTWIN show no precipitation.

The model domain is the operational used by COSMO-DE (see Fig. 4.2), the operational horizontal resolution ($2.8km \times 2.8km$) is used as well.

First we describe the three test cases (Section 4.1) then the results are discussed

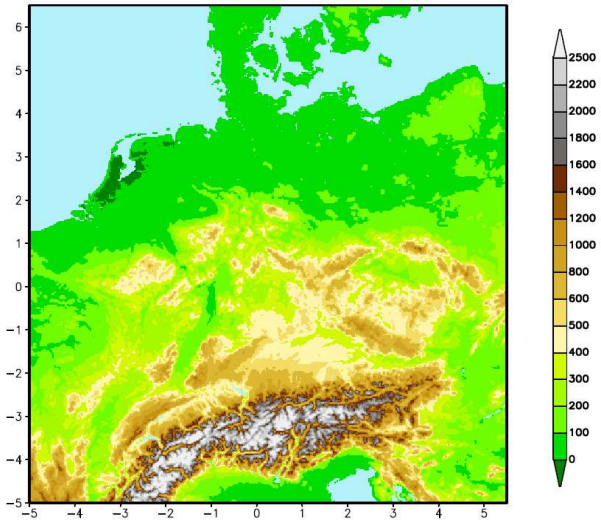


Figure 4.2: orography of the COSMO integration area, the coordinate are for the rotated grid model

(Section 4.2).

The position of the cloud base (Section 4.3) is revealed in this work as an interesting parameter in the PIB precipitation forecast success, because of the sensitivity of CAPE to this parameter. The simulated convection is sensitive to low level humidity: even small changes can lead to different storm developments, particularly in environments classified at the transition between different storm regimes (Park, 1999 [78], Weisman and Klemp, 1982 [97]). Generally the low-level moisture controls the timing and persistence of clouds (Hong et al., 1995 [47]) and variations in moisture content can alter convective rainfall significantly (Cotton and Tripoli, 1981 [21]). Moreover, Crook (1996, [25]) showed that the storm strength is more sensitive to the variability in moisture than in temperature.

The comparison between the horizontal wind (direction, absolute value and divergence) in CTL and IDTWIN is interesting. We demonstrate that the change of the vertical wind adjusts also the horizontal wind field (from the mass conservation equation). The convergence near the cloud base as well as the divergence near the cloud top will be analysed too (Section 4.4).

Finally results without the assimilation of the vertical wind and without the humidity adjustment will be evaluated (Section 4.5), in order to determine the impact of this parts of the scheme.

## 4.1 Case studies

We investigate the performance of the PIB scheme with three convective cases in Germany:

- CASE 1: system characterised by a convective line moving toward central Germany.
- CASE 2: cold front leading to many small convective systems.
- CASE 3: convective precipitation that moves across southern and south eastern Germany.

The synoptic descriptions are based on the radar composite output from DWD (Figs. 4.3, 4.4 and 4.5) and the *Berliner Wetterkarte* (Figs. D.1, D.2 and D.3). The description of the weather was performed by the *Freie Universität Berlin* .

### 4.1.1 Case 1 : June 29, 2005

An upper-air low pressure zone (the surface pressure chart is in Fig. D.1) is present near the west part of Brittany in the early hours of the day. Its cold front moves eastward. A strong upper-level ridge, whose axis extends from the North Sea above West Germany and North Italy to Tunisia, advects warm air from North Africa to middle Europe.

Southwest Germany reaches high temperatures, for example in Freiburg the maximum temperature was $35°C$ and in Karlsruhe $34°C$. In the afternoon several storms developed along the Danube and near the Black Forest, which are pulled southward with the upper air flow. The studied event, well visible in the radar data (Fig. 4.3), is a convective line moving across Southern and central Germany during the night. It starts at about 00 UTC and finishes at 11 UTC. As measured by DWD the precipitation amount for this event is high, e.g. 48 mm for a six hours accumulation at the Cologne-Bonn airport.

The distinctive feature of this case is a poor forecast: the operational DWD forecast between 00 and 06 UTC predicted no rain or only small amounts in the area of interest. The cold front is not caught from this forecast.

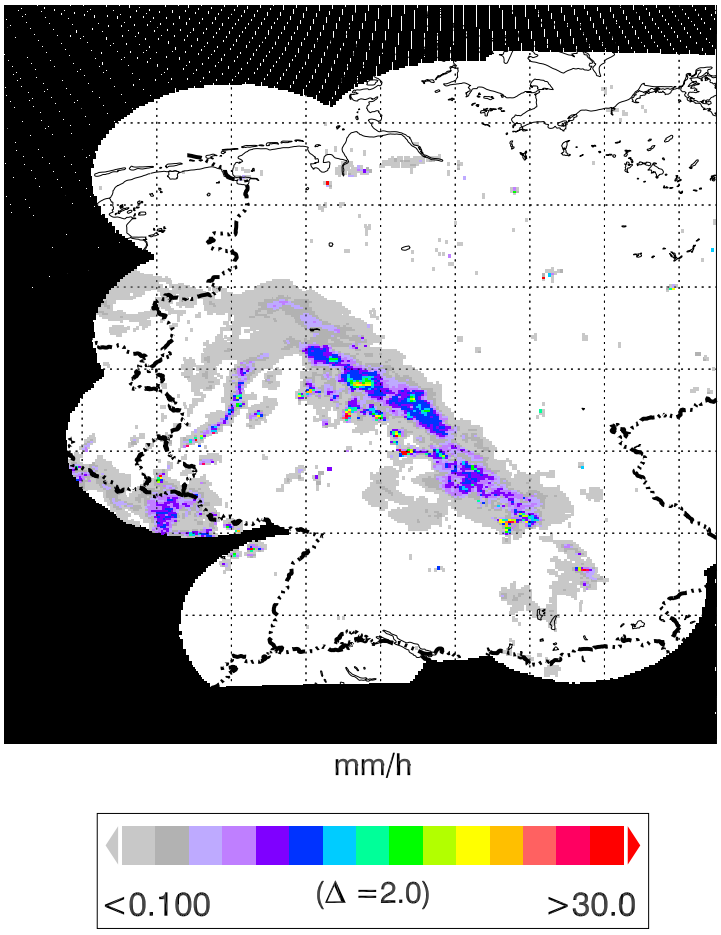


Figure 4.3: Radar composite image for June 29, 2005, 06:00 UTC (CASE1).

### 4.1.2 Case 2 : August 19, 2005

The influence of ex-hurricane *Irene* is important for this event. This leads to advection of tropical sea air in the middle latitudes, (Fig. D.2). Near the Azores hurricane *Harvey* dissipated some days before, leaving an upper level trough extended until Madeira and moving to South-West Europe by the westerly flow. At the same time the anticyclone EDDA moves to the Baltic region and a cold front evolves on the west part of Germany. Convective precipitation started around 12 UTC and covered the area for the rest of the day until about 22 UTC.

This day is interesting for our study because it is a possibility to test the behaviour of PIB in the presence of multicells (Fig. 4.4).

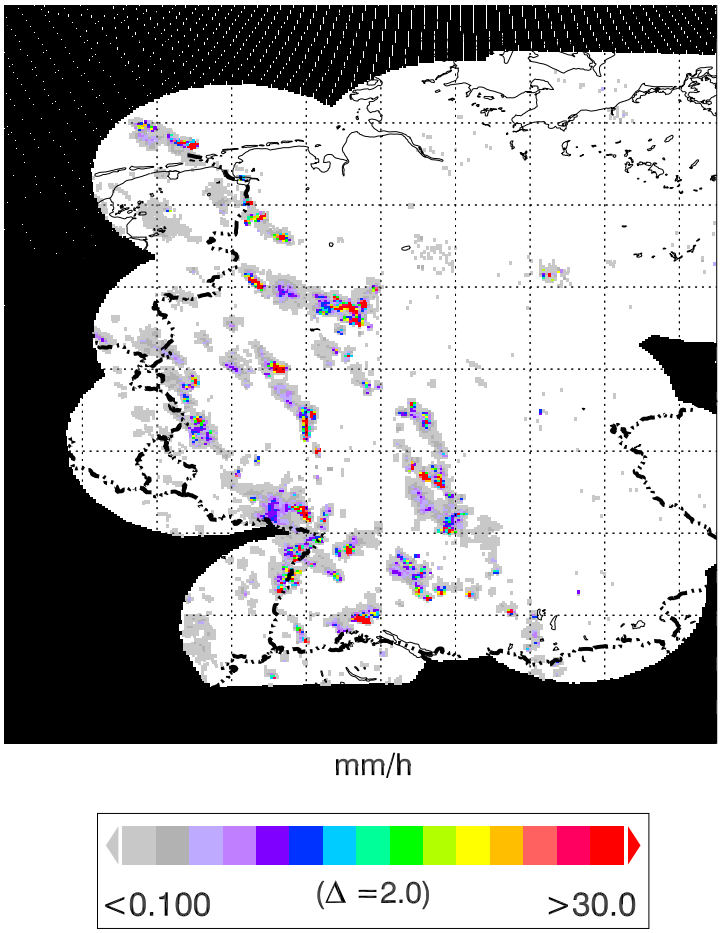


Figure 4.4: Radar composite image for August 19, 2005, 16:00 UTC (CASE2).

### 4.1.3 Case 3 : June 28, 2006

In northern Germany the influence of the low pressure region QUEENY leads to low temperatures; in southern Germany a high pressure ridge advects warm air from northern Africa. Therefore a high temperature gradient is present. The maximal temperature rises from 16°$C$ in List/Sylt to 30°$C$ in Bavaria, the minimum temperature rises from 6°$C$ in Lower Saxony to 18°$C$ in Baden-Wüttemberg.

In Southern Germany a convective region, activated by a weak trough, moves eastward during the first hours of the day (Fig. 4.5).

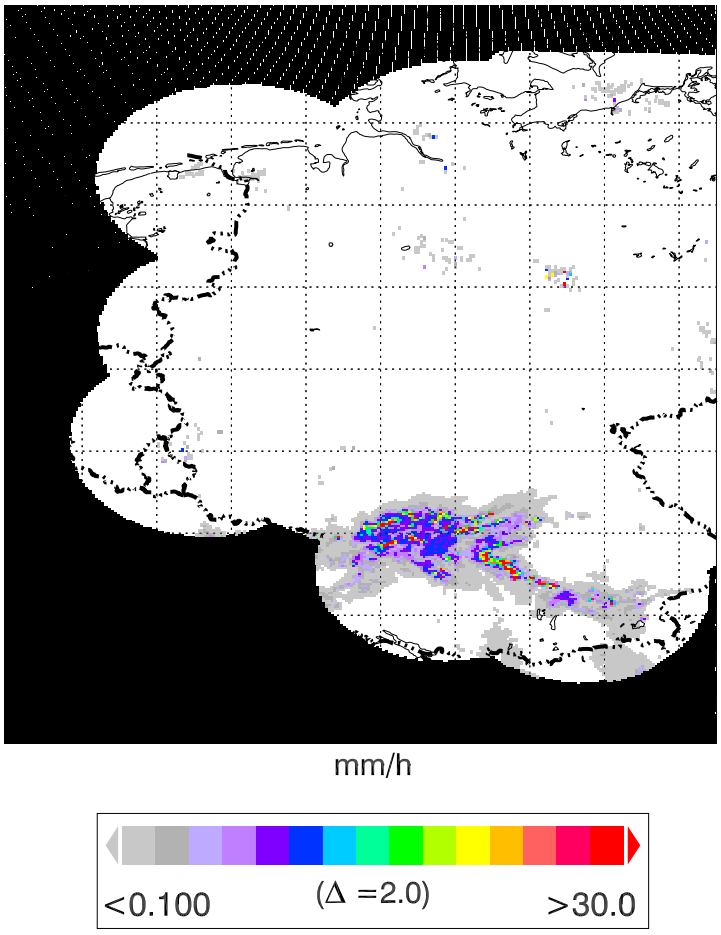


Figure 4.5: Radar composite image for June 28, 2006, 05:00 UTC (CASE3).

## 4.2 Precipitation and CAPE field

In this Section the precipitation field is analysed, at the end of the two hours assimilation window and after one hour of free run. It is important to note that the behaviour of the model during the free run depends on the general predictability of the situation. The first hours of free run are useful to analyse how the model reacts to the adjustments due to the PIB without any other constriction.

A first comparison between CTL and IDTWIN is made, looking at the precipitation fields for every case study (Figs. 4.6, 4.8 and 4.10). The comparison reveals the general ability of PIB to capture the main features of the storm evolution but also the general tendency to underestimate the precipitation amount at the end of the assimilation window. Otherwise the

position of convective cells is well represented. A relationship with the positive CAPE field (defined in appendix C) is also discussed.

One hour after the end of the assimilation window (that means one hour free run at the beginning, two hours of assimilation and then one hour free run; as described in Fig. 4.1), the precipitation of IDTWIN has the common feature to remain in good approximation with CTL, this is a prove of the PIB's consistency with the model. In other words the PIB does not upset the model physic.

In CASE1 IDTWIN underestimates CAPE, and consequently precipitation, during the assimilation window. As visible in Figs. 4.6 and 4.7 for the eastern part of France CAPE presents in CTL is reduced during the assimilation window in IDTWIN. In the same region after one hour of free run the two CAPE fields are very similar but in the precipitation fields some differences are present, in the amount as well as in the path. In this case PIB assimilates the weather situation quite well but it suppresses also some convective cells in east France, moreover PIB anticipate the end of convection. A possible explanation for this behaviour of PIB is present in the next sections.

For CASE2 the path of CTL CAPE is very similar to the IDTWIN ones, also the approximation of the quantity is good (Fig. 4.9), at the end of the assimilation window as well as after one hour of free run. Only small differences are present, consequently the precipitation is well reproduced with respect to position and size of the convective cells (Fig. 4.8). In this case PIB is able to reproduce the same dynamic present in the model.

The IDTWIN in CASE3 underestimates the CAPE in the precipitating regions (Fig. 4.11). In the CTL the CAPE in the precipitating regions is very low (about 350 J/kg). The precipitation in the IDTWIN is underestimated (Fig. 4.10) although the path is very well approximated.

The good CAPE approximation of IDTWIN leads to the good quantitative approximation in precipitation for these regions, as we expected. The initialisation of the model variables leads to a CAPE initialisation and furthermore improves the forecast during the free run. Generally PIB has the characteristic to underestimate CAPE, this feature will be discussed also in the next Sections.

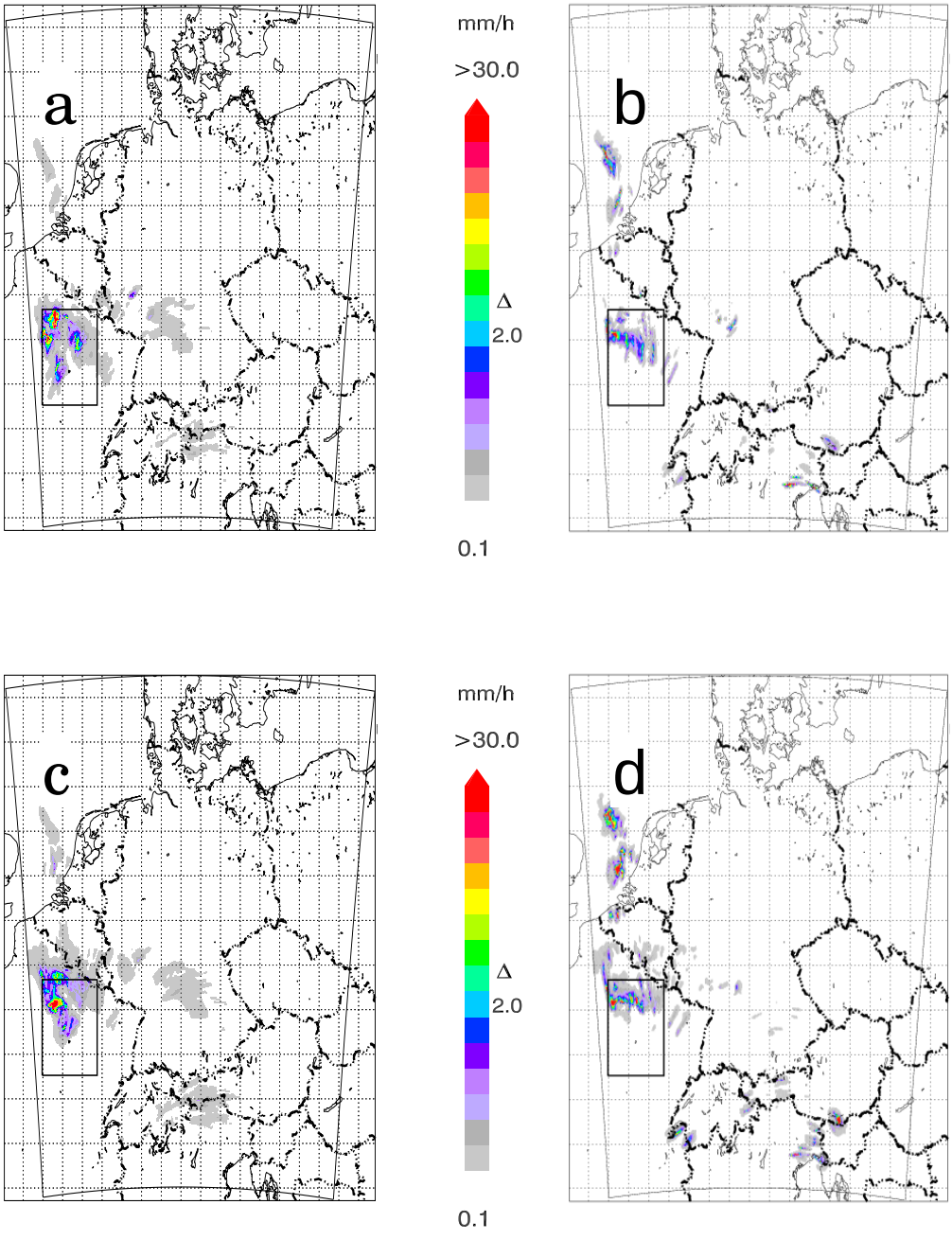


Figure 4.6: *Selected results of the hourly accumulated precipitation field for 29/06/2005. After two hours of assimilation. a) control run, b) identical twin. After an hour of free run. c) control run, d) identical twin.*

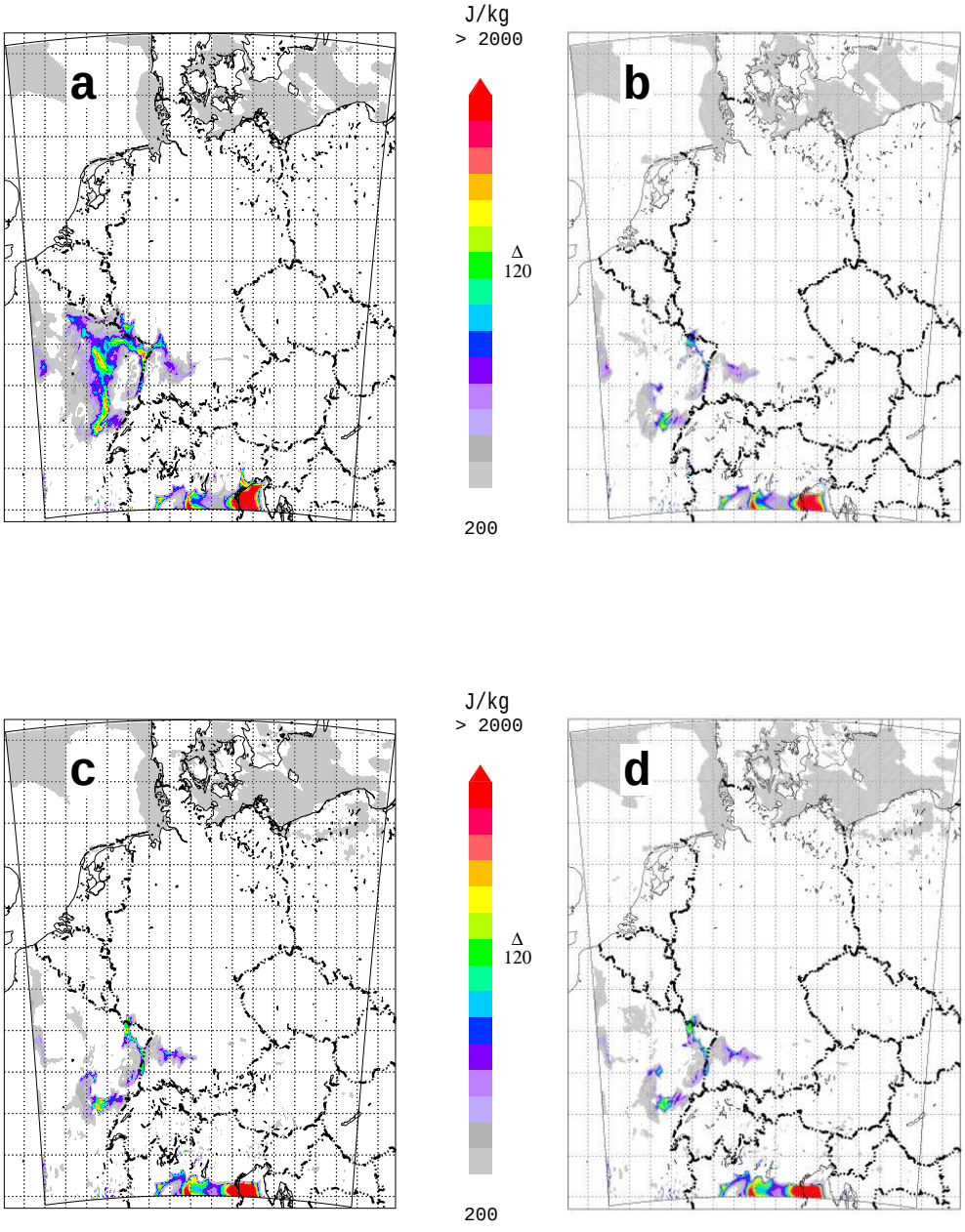


Figure 4.7: *The same as in Fig. 4.6 but for the CAPE field.*

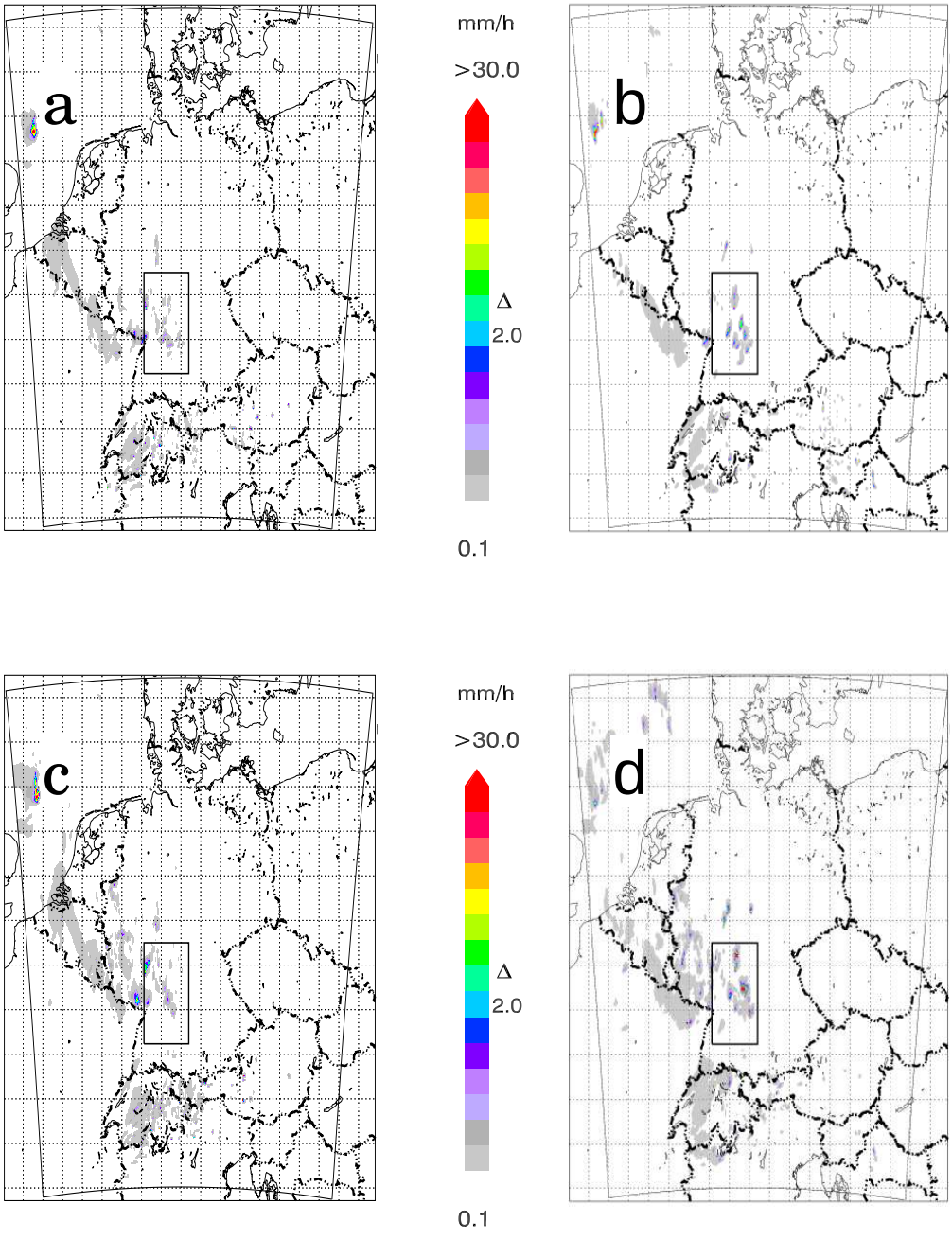


Figure 4.8: *Selected results of the hourly accumulated precipitation field for 19/08/2005. After two hours of assimilation. a) control run, b) identical twin. After an hour of free run. c) control run, d) identical twin.*

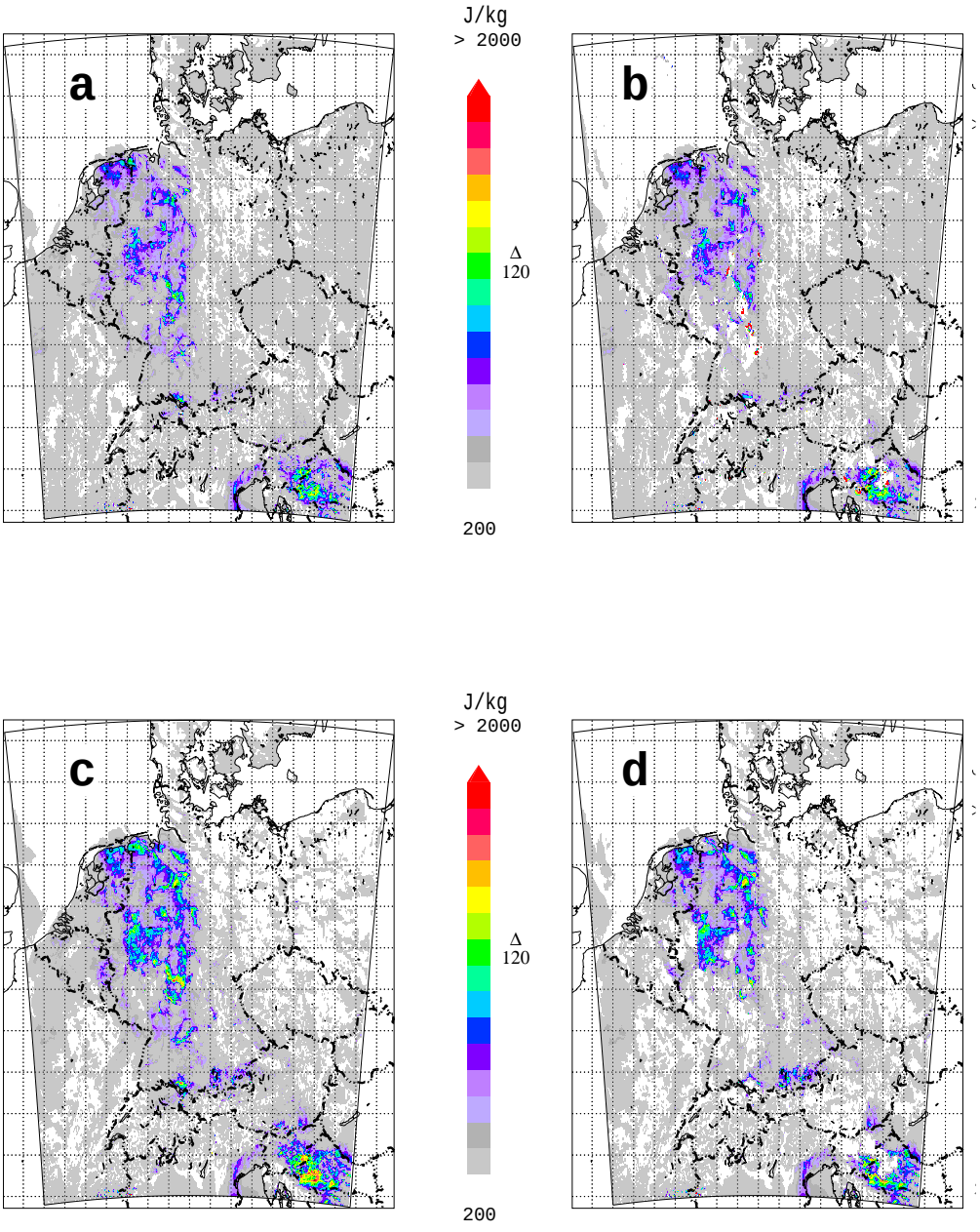


Figure 4.9: *The same as Fig. 4.8 but for the CAPE field.*

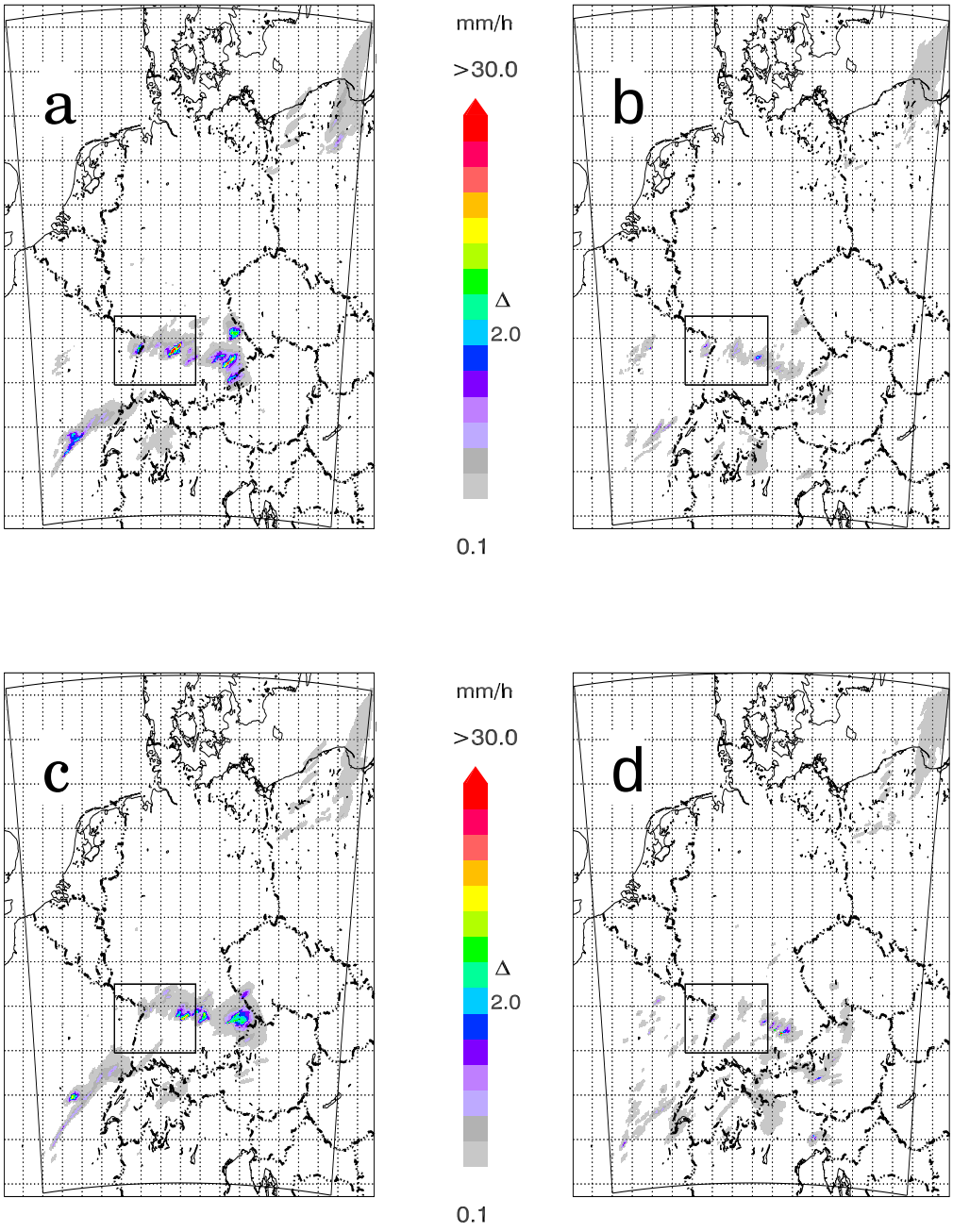


Figure 4.10: *Selected results of the hourly accumulated precipitation field for 28/06/2006. After two hours of assimilation. a) control run, b) identical twin. After an hour of free run. c) control run, d) identical twin.*

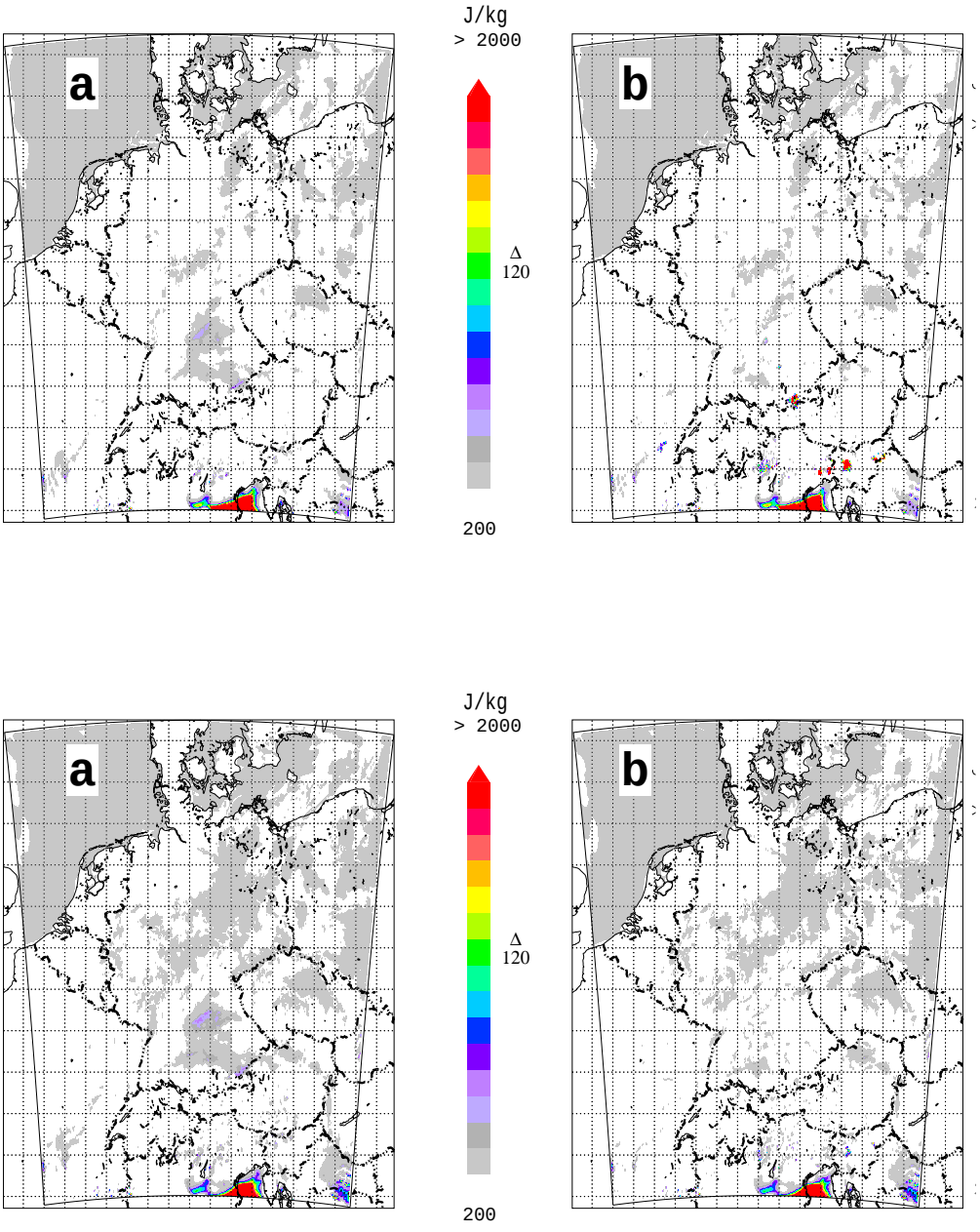


Figure 4.11: *The same as Fig. 4.10 but for the CAPE field.*

Figs. 4.12, 4.13 and 4.14 show the pdf of the ratio of precipitation generated by the CTL to precipitation generated by the IDTWIN, and the frequency bias of IDTWIN. In the pdf distributions only grid points, where both simulations produced precipitation between 1 and 60 mm/h were compared. Note the change of axis increments in the pdf plots, from ratios lower than 1 to higher than one.

A value of one would indicate a perfect reproduction by IDTWIN of the precipitation field produced by CTL. Values less (more) than 1 indicates overestimation (underestimation) by IDTWIN. In order to simplify the comparison between different pdfs the integral between 1/3 and 3 is added. This quantity denotes the probability that the IDTWIN precipitation is between 1/3 and 3 times the CTL precipitation.

The frequency bias (see Appendix A.1) with a threshold of 0.1 mm/h indicates an overestimation (values bigger than 1) or an underestimation (less than 1) of the precipitative area, in this case the frequency bias completes the information given from the PDF, because it takes into account also the points where the simulations have no precipitation.

During the assimilation window the pdf of CTL/IDTWIN is always shifted to values larger than one, with the maximum placed between 1 and 2. This means that the amount is underestimated. The integral values underlines that IDTWIN approximates CTL worst in the second hour of assimilation compared to the first one, especially in CASE1. During the free run the pdf is closer to one, moreover the integral values increases or remain quite constant. The IDTWIN tends to return to the CTL state. The model dynamics and physics are not upset proving the PIB consistency with the model.

The frequency bias stays close to one in CASE2 and CASE3, but for CASE1 the precipitation areas is clearly underestimated (frequency bias always negative with values of also 0.5).

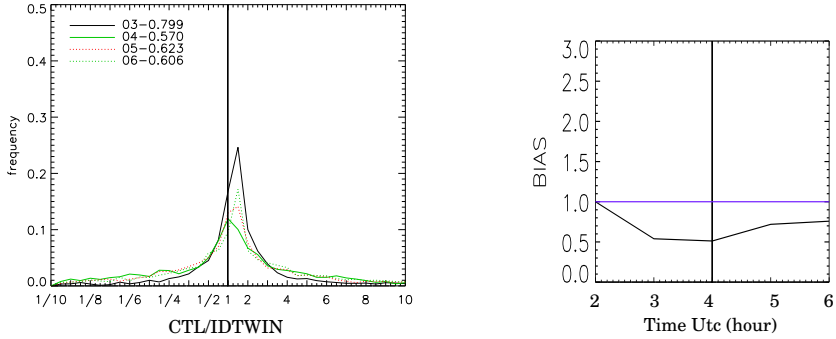


Figure 4.12: *PDF of CTL/IDTWIN for CASE1. Continuous line during the assimilation window, dashed line during the free run. In the legend beside the hour, the integral between* 1/3 *and* 3 *is written. On the right side Bias frequency for the IDTWIN.*

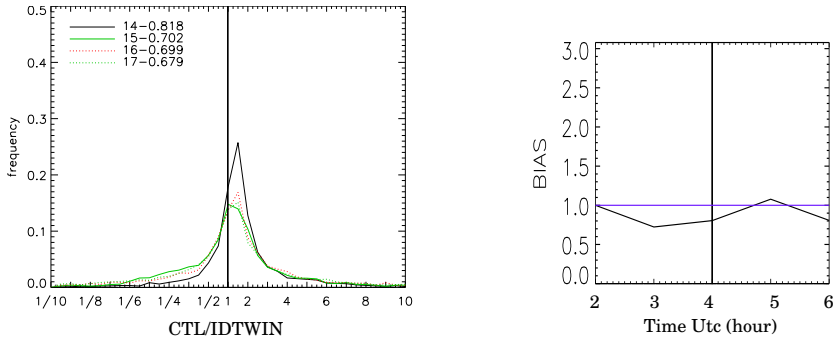


Figure 4.13: *The same as Fig. 4.13 but for CASE2.*

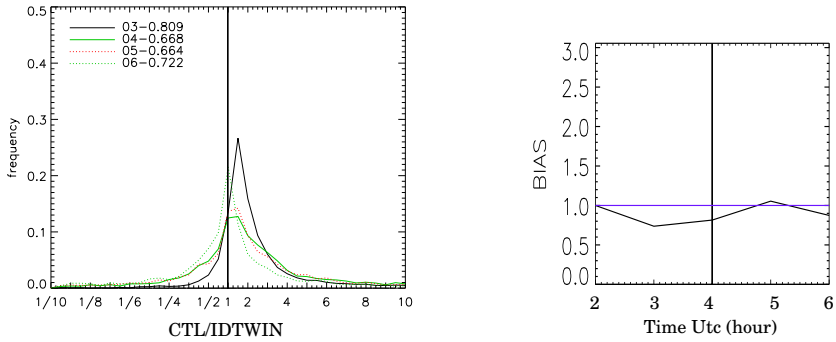


Figure 4.14: *The same as Fig. 4.13 but for CASE3.*

## 4.3 Cloud base in the convective regions

In the following Section we analyse in particular the convective subareas marked as black boxes in Figs. 4.6, 4.8 and 4.10.

A strong relation exists between cloud base height and CAPE, when cloud base is estimated from the level of free convection. In a conditionally unstable atmosphere the parcel becomes warmer than its surroundings when passing that level. This also means that the modelled precipitation increments are sensitive to the cloud base determination. The PIB is also sensible to this determination because above the cloud base the air is assumed to be saturated, which changes CAPE accordingly.

In Figs. 4.15, 4.16 and 4.17 we compare precipitation (upper pictures) and cloud base (lower pictures) in the convective areas, in order to visualize the behaviour of IDTWIN in the convective regions. The cloud base height is given as model level (level 50 is closest to the soil, level 1 is the top of the atmosphere). The figures show the situations at the end of the assimilation window. IDTWIN assimilates the precipitation of CTL, thus the cloud base in IDTWIN must be associated with the precipitation in CTL.

In this work the prognostic precipitation is used (for explanation see Section 2.1.4). The precipitation is advected on its way from the cloud base to the ground. The cloud base can be displaced somewhat from the precipitating area.

In CASE1 (Fig. 4.15) such a displacement is present. The convective cells in IDTWIN has both position and intensity errors. The large area with a low cloud base in CTL (red colour) is not present in IDTWIN, consequently no precipitation is formed in this area. Also looking to precipitation in CTL and cloud base in IDTWIN the large precipitating area at the box bottom has no correspondence in the IDTWIN cloud, for the other convective cells clouds are present. A connection between low cloud base height and the precipitation field is however well visible. Where IDTWIN succeeds into form low cloud base also precipitation is present. In this case precipitation in IDTWIN is underestimated and connected with the overestimation of cloud base.

In CASE2 (Fig. 4.16) a positioning error is present. Consequently a displacement is also present in the precipitation field. The cloud base is better approximated concerning its height than in CASE1, but an underestimation is visible. Accordingly to that the precipitation estimation is also overestimated and the convective cells are more intense. In particular a cloud area at the bottom-right part for IDTWIN produces strong convective cells not presents in CTL.

In CASE3 (Fig. 4.17) the cloud area is underestimated but its structure is

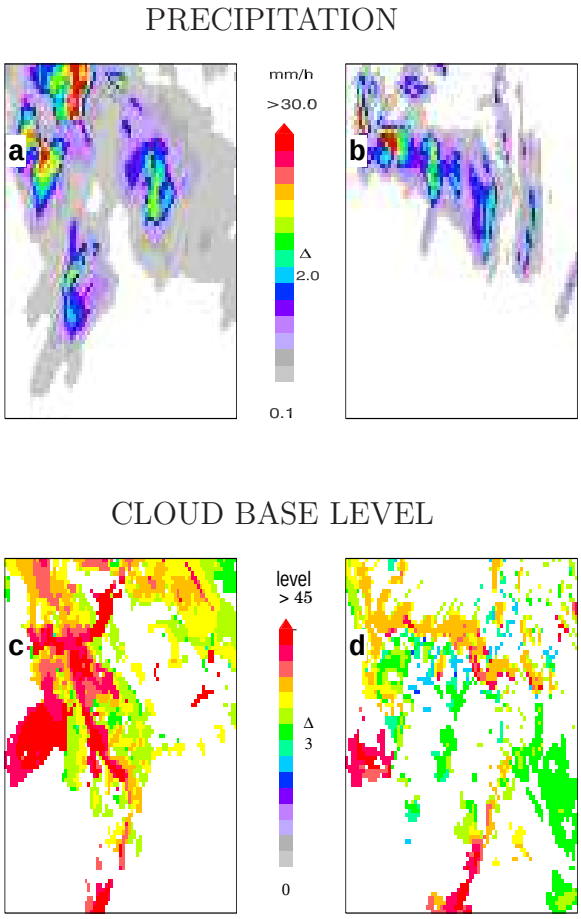


Figure 4.15: *Convective area for the day of 29/06/2005 (area defined in Fig. 4.6). Hourly accumulated precipitation at the end of the assimilation window, a: for the control run, b: for the identical twin run. c and d: the corresponding cloud base distribution.*

well approximated. Also for this event the relationship between convective cells and cloud base is evident.

In presence of convective cells PIB succeeds quite well into assimilate the corresponding cloud base, however in CASE1 a precipitating area in CTL is not reproduced in the IDTWIN clouds.

From this more qualitative approach we can assert that the height of the cloud base is higher in IDTWIN than in CTL. But PIB maintains the relationship between the presence of precipitation and the height of the cloud base. As already seen in Section 4.2, the computation of the cloud base influences CAPE and thus precipitation.

This bias of PIB in the determination of the cloud base is also evident in the

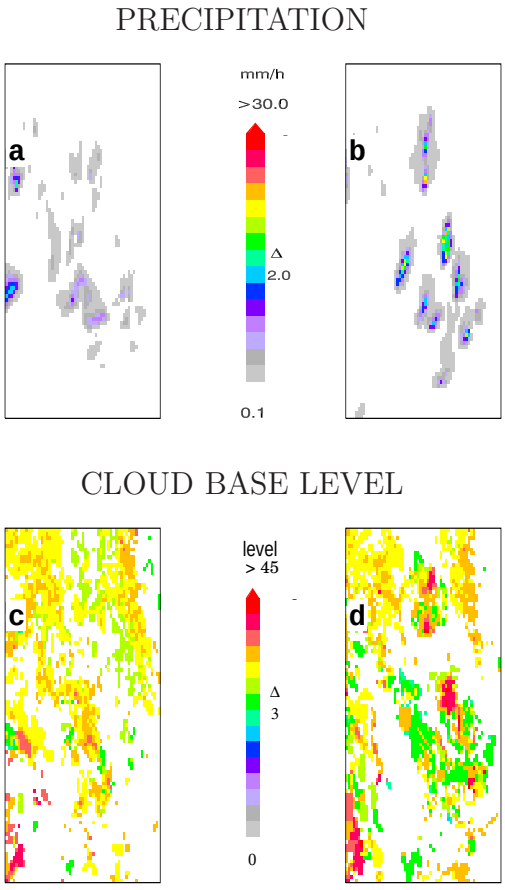


Figure 4.16: *Convective area for the day of 19/08/2005 (area defined in Fig. 4.8). Hourly accumulated precipitation at the end of the assimilation window, a: for the control run, b: for the identical twin run. c and d: the corresponding cloud base distribution.*

relative frequency of the cloud base levels, for the whole simulated area at the end of the assimilation window, described in the histogram plot in Fig. 4.18. Generally IDTWIN overestimates the determination of the cloud base level.

In CASE1 the CTL has a larger area with a cloud base close to $700hPa$, in other words IDTWIN underestimates the second peak and we have an overestimation of the frequency of higher and lower clouds. Also in CASE2 and CASE3 the frequency peak close to $700hPa$ is underestimated but the overestimation interests only the higher clouds (more than $600hPa$).

Commonly PIB underestimates the medium height cloud bases and overestimates the upper level cloud bases, the lower cloud bases are in good agreement with CTL, thus the frequency of clouds more sensible to CAPE is well assimilated.

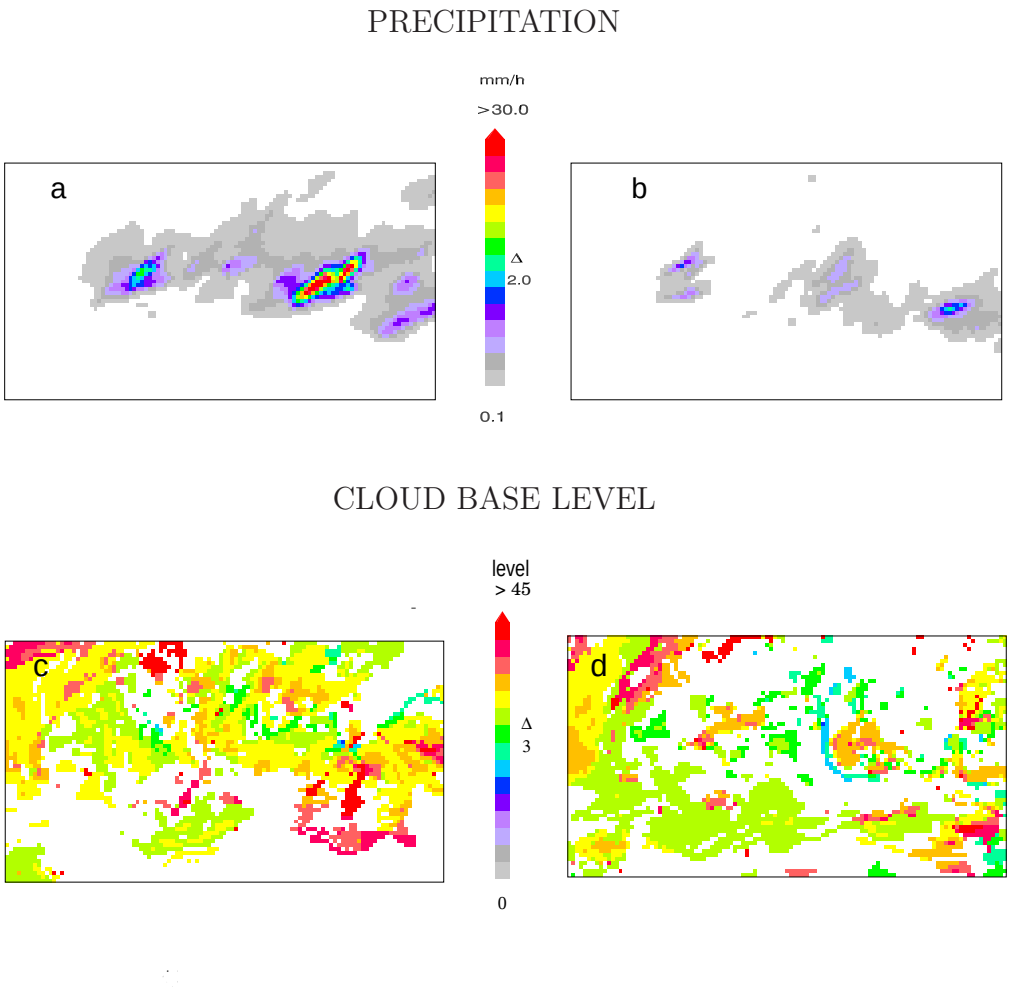


Figure 4.17: *Convective area for the day of 28/06/2005 (area defined in Fig. 4.10). Hourly accumulated precipitation at the end of the assimilation window, a: for the control run, b: for the identical twin run. c and d: the corresponding cloud base distribution.*

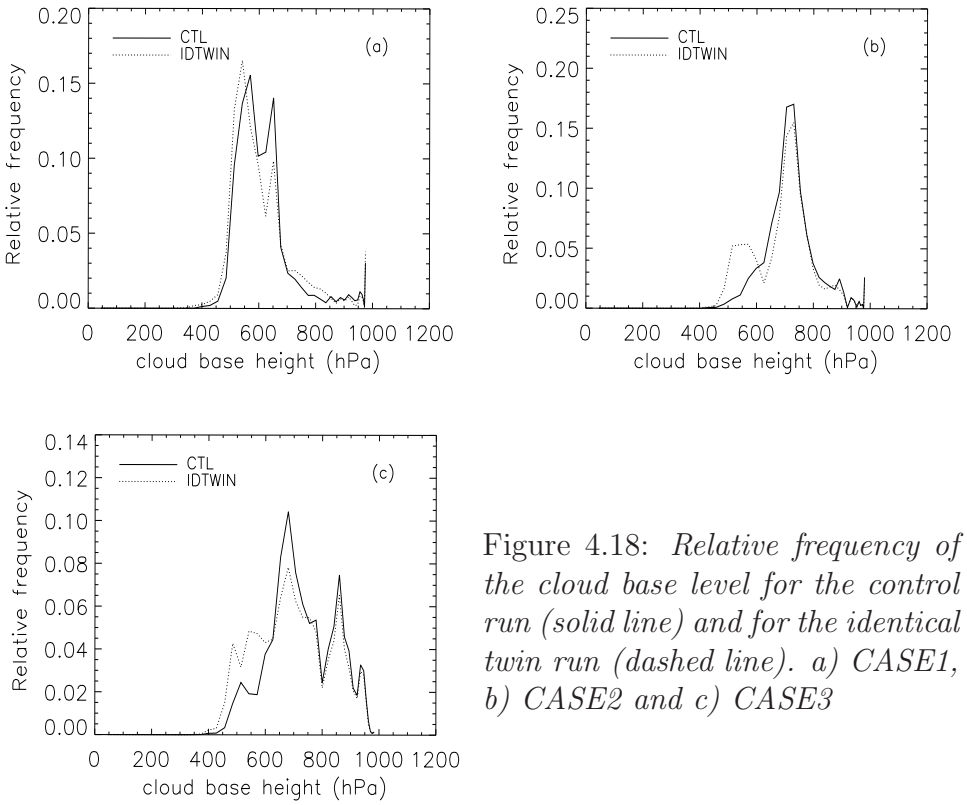


Figure 4.18: *Relative frequency of the cloud base level for the control run (solid line) and for the identical twin run (dashed line). a) CASE1, b) CASE2 and c) CASE3*

## 4.4 Mass flux divergence

PIB changes the vertical wind (Section 3.2), thus the horizontal wind should be adjusted to this change via the mass conservation equation. This assumption is verified in this Section.

In a convective boundary layer (Stull, 1999 [88]) horizontal convergence region under the cloud base should be connected to the updraft. Above the level of neutral buoyancy divergence should prevail. Both are connected with the position of the convective cells shown in Figs. 4.15, 4.16 and 4.17.

We analyse the convergence for the model level 31 (corresponding to a pressure of about $700hPa$), which should be near the cloud base, according to the results of Section 4.3. In this Section we focus on the fields at the end of the two hours assimilation window.

From the results shown in Section 3.2 we expect the maximum of the vertical wind lower for CTL than for IDTWIN, as well as for the corresponding maximum of the divergence field. Thus the divergence near the cloud top is analysed at model level 15 (corresponding to a pressure of about $250hPa$) for

IDTWIN and at model level 20 (corresponding to a pressure of about $400hPa$) for CTL.

The horizontal wind field (Figs. 4.19, 4.20 and 4.21, c and d) near the cloud base is well approximated in the Identical Twin run. Both the distributions of the divergence field and the direction of the wind field are similar to CTL, for all cases. The magnitude of divergence and convergence in the convective region is properly reproduced. Convergence (negative divergence) is consistent with the different positions and intensities of the convective cells.

At the cloud top (Figs. 4.19, 4.20 and 4.21, a and b) the wind field and the correspondent maximum of mass flux divergence are quite well approximated in CASE2 as well as in CASE3. In CASE1 a position error is present, as in the precipitation (Fig. 4.15). In all cases the correspondence between convective cells and divergence at high levels is clear. Moreover, the error in the precipitation is low near the cloud top, because the most part of the humidity is situated in the lower level. Therefore, the influence on the precipitation field is low.

The precipitation fields (Figs. 4.15, 4.16 and 4.17) for the IDTWIN and CTL simulations contain some differences; the convective cells have not the same position, therefore the corresponding horizontal wind divergence fields are not expected to be correlated.

For this reason the total horizontal wind divergence field is analysed in more detail using its statistic. We subdivide divergence into classes with regular intervals of $10^{-4}\,s^{-1}$ (Figs. 4.22 and 4.23). As a measure of the differences between the IDTWIN and CTL distributions we use as metric the Kullback-Leibler divergence, also named relative entropy, of $Z_{IDT}/Z_{CTL}$ (Kullback, 1968 [61]):

$$rel.entropy = \sum_{i}^{n}(Z_{CTL}\log\frac{Z_{CTL}}{Z_{IDT}}) \tag{4.1}$$

where n is the number of bins used, $Z_{CTL}$ is the frequency of values in a given bin of CTL, $Z_{IDT}$ is the frequency of values in a given bin of IDTWIN. A perfect simulation will have a relative entropy equal to 0. In the relative entropy larger values denote a model output with a poorer approximation to the reference data. With this analysis we are able to compare IDTWIN and CTL in terms of relative frequency.

This analysis is performed for all cases, for cloud top and for cloud bottom. In every plot the distributions are very similar with a low relative entropy (less than 0.02, Figs. 4.22 and 4.23 ).

For a better analysis of the convective regions only the grid points with precipitation above 2 mm are taken into account. With regard to the cloud

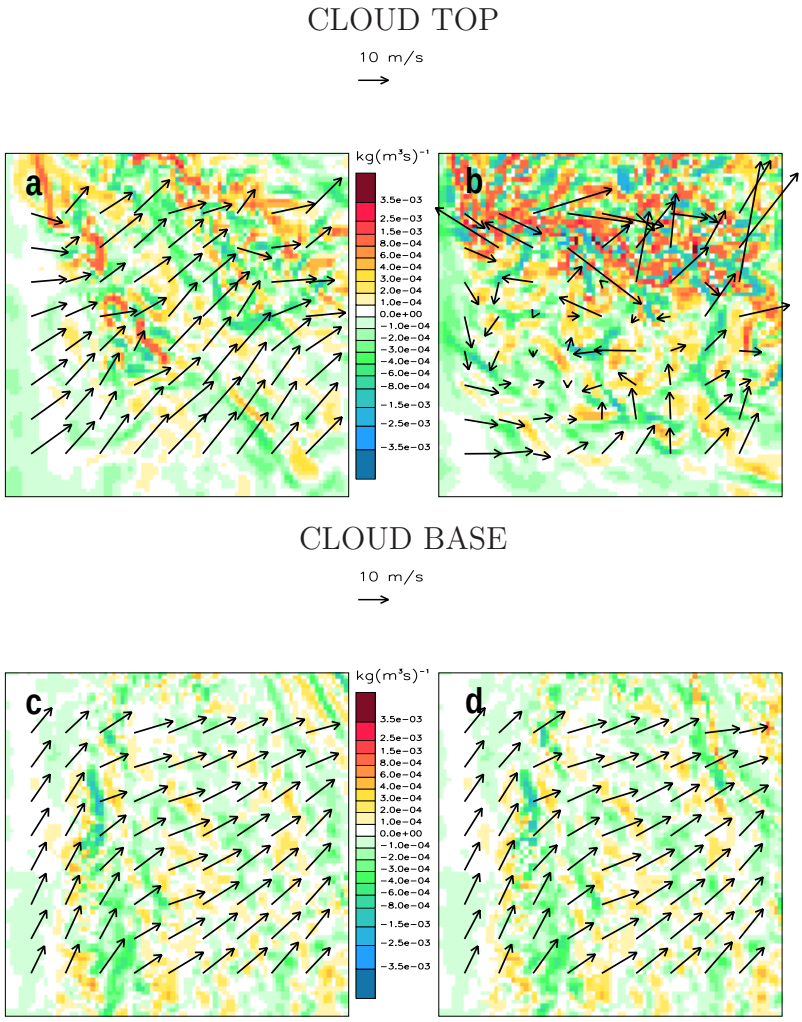


Figure 4.19: *29/06/2005. Horizontal wind divergence (colour field) and wind (black arrows) for the area in figure 4.15. The upper panel refers to the cloud top, the lower panel to the cloud base. a and c: IDTWIN; b and d: CTL.*

base (Fig. 4.24) a good approximation for all cases is still evident, for CASE2 and CASE3 (Fig. 4.24 b and c) a slight overestimation of the positive values is present. The relative entropy increases respect to the entire field but it remains less than 0.1.

Close to cloud top (Fig. 4.25) the approximation between IDTWIN and CTL is worst than near the cloud base. In all cases the mass flux divergence in IDTWIN has a flatter distribution; the extreme values of divergence/convergence are overestimated. Nevertheless, the central parts of the two distributions are fitting well and the relative entropy remains less than 0.25.

In spite of the PIB changes of the vertical wind, the horizontal wind near the cloud base seems to be adjusted in a model consistent way, that from the analysis of its divergence. A local problem is present near the cloud top for the grib points with strong precipitation. Anyway the divergence distributions are similar and the relative entropy remains limited.

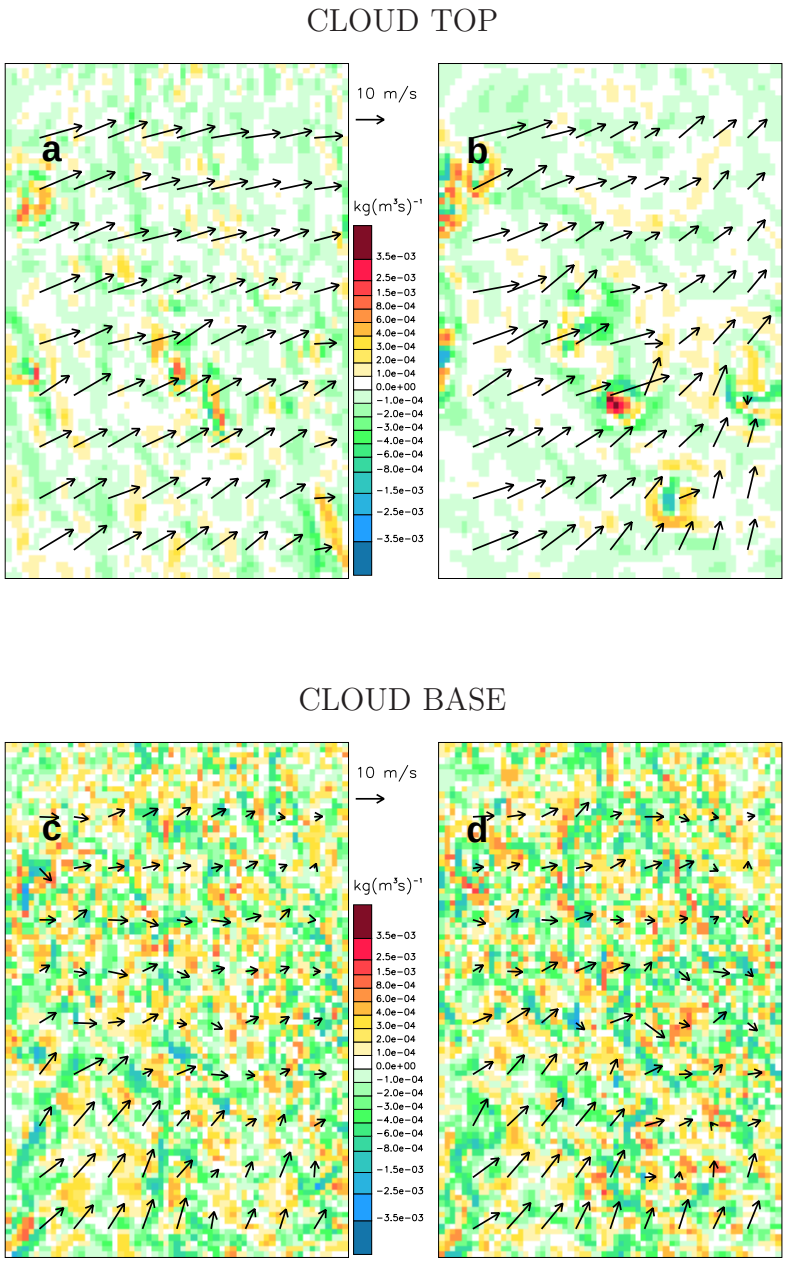


Figure 4.20: *The same as the Fig. 4.19 but for CASE2.*

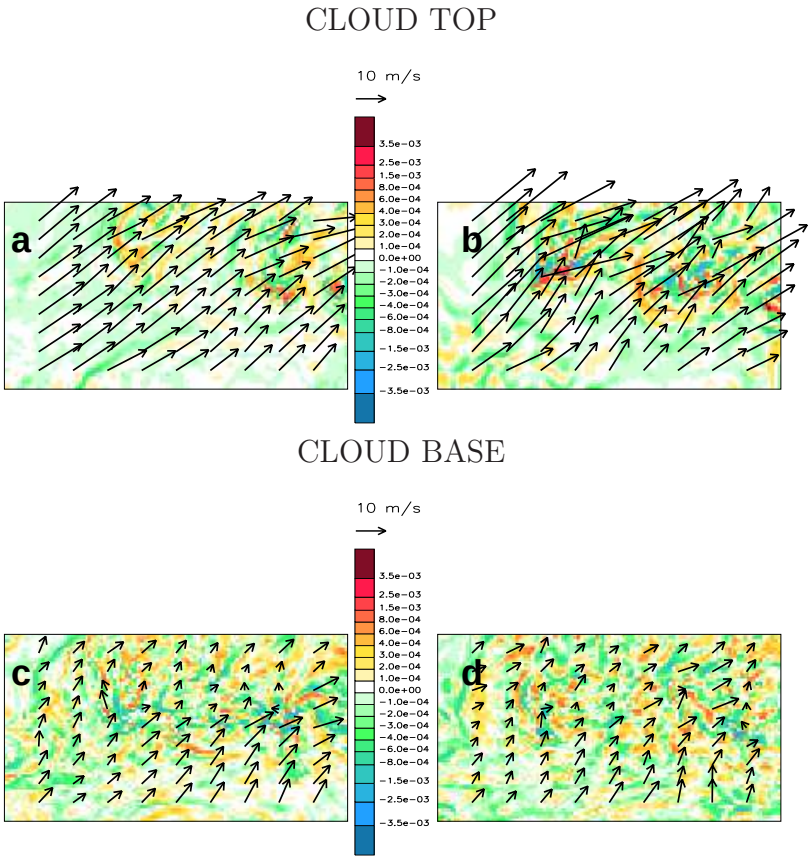


Figure 4.21: *The same as the Fig. 4.19 but for CASE3.*

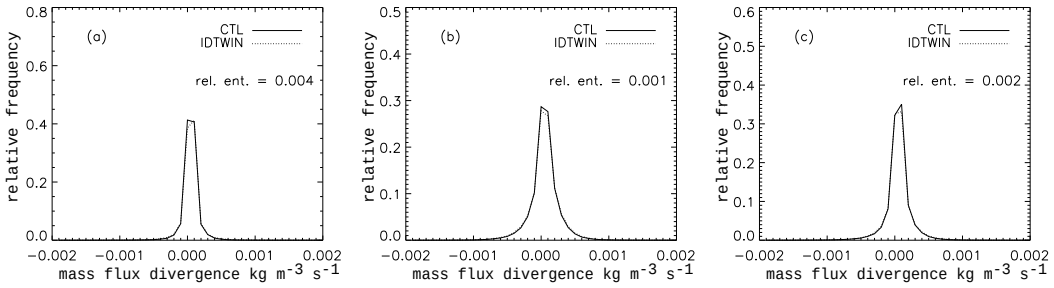


Figure 4.22: *Relative frequency of the mass flux convergence at the cloud base subdivided in classes. a) CASE1, b) CASE2, c) CASE3*

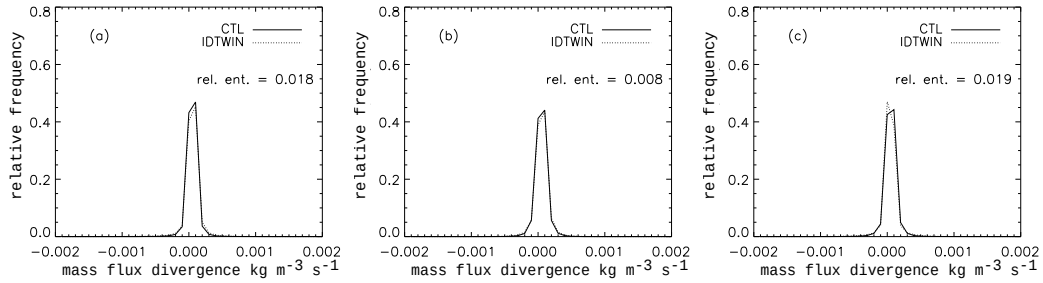


Figure 4.23: *The same as the Fig. 4.22 but for the cloud top*

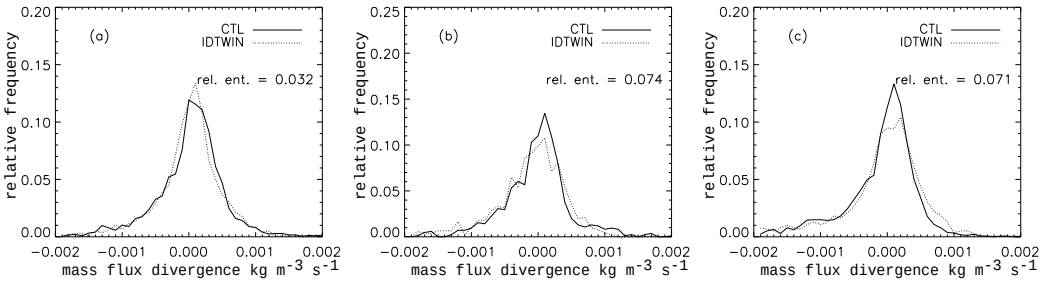


Figure 4.24: *Relative frequency of the mass flux divergence at the cloud base subdivided in classes, for the points where the precipitation is greater than 2 mm. a) CASE1, b) CASE2, c) CASE3*

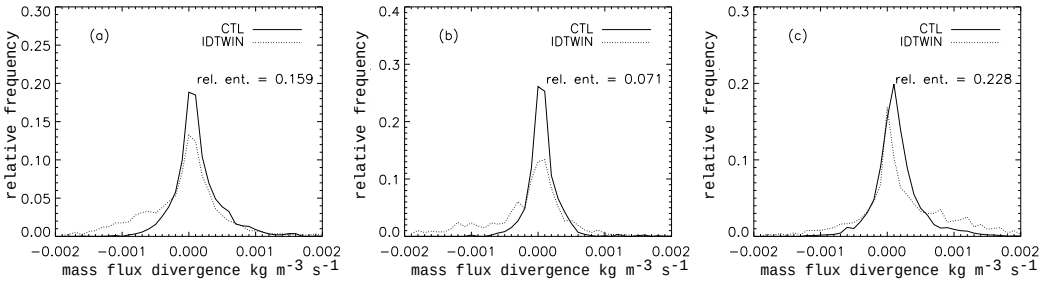


Figure 4.25: *The same as the figure 4.24 but for the cloud top*

## 4.5 PIB without vertical wind or without humidity assimilation

An identical twin run without the vertical wind (PIB nowind) and one without the humidity assimilation (PIB noqv) are made in order to evaluate their respective impact in the assimilation scheme.

As measure of the error in the vertical wind field we use the Root Mean Square Difference (RMSD, see appendix A) between CTL and IDTWIN wind fields, the comparison is made between outputs with time resolution of five minutes.

### 4.5.1 Precipitation

In Figs. 4.26, 4.27 and 4.28 the precipitation distribution of the two hours of assimilation and the two hours of free run is shown. These hours relate to the time between 04 UTC and 06 UTC for CASE1 and CASE3, and between 15 UTC and 17 UTC for CASE2. As in Section 4.2 the integral of the pdf between 1/3 and 3 is added.

For CASE1 from the second hour of assimilation, the precipitation from PIB nowind and from PIB noqv is more strongly overestimated (CTL/IDTWIN less than 1 has a higher probability). The complete PIB scheme is generally better distributed around 1. Only in the first hour of assimilation the distribution is better in PIB noqv. The probability that the IDTWIN precipitation is between 1/3 and 3 times the CTL is, during the assimilation window, better for PIB noqv. During the free run the integral values are comparable.

In CASE2 PIB noqv shows a better behaviour than the complete PIB during the assimilation window, but not during the free forecast. Moreover, the integral value for PIB noqv is better than the one of the complete PIB during the assimilation window but not during the free run. Thus the assimilation with the complete PIB seem to disturb the model dynamic less than the assimilation with the PIB noqv. The PIB nowind is the worst option.

The behaviour of CASE3 is very similar to the one of CASE1.

The complete PIB system has a pdf of the precipitation similar to the CTL's pdf, during the assimilation window and also during the free forecast. The PIB without the humidity adjustment has generally better results during the assimilation window, the only change of the vertical wind field maintain the PIB noqv state near the CTL state. On the contrary the quality of the complete PIB does not decrease during the free run, thus we can conclude that it has demonstrated a major model consistency. Both the assimilation of humidity and vertical wind are necessary in order to maintain model consistency.

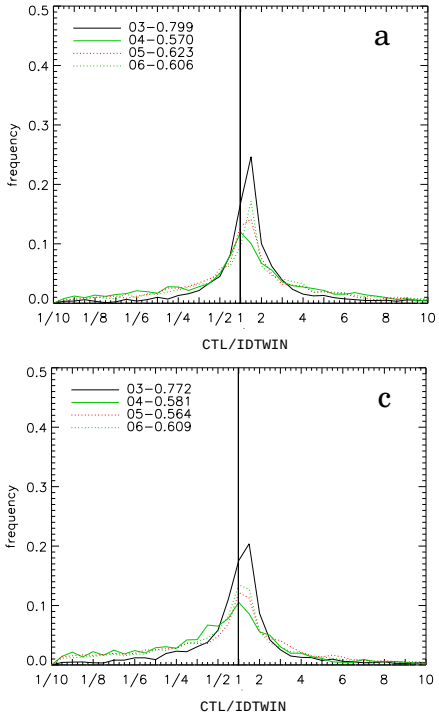


Figure 4.26: *PDF of CTL/IDTWIN for CASE1: PIB (a), PIB noqv (b) and PIB nowind (c). Continue line during the assimilation window, dashed line during the free run. In the legend beside the hour, the integral between* 1/3 *and* 3 *is wrote.*

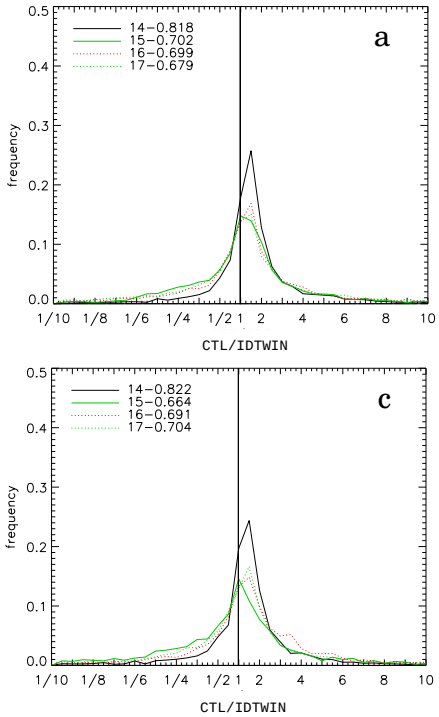


Figure 4.27: *The same as Fig. 4.26 but for CASE2*

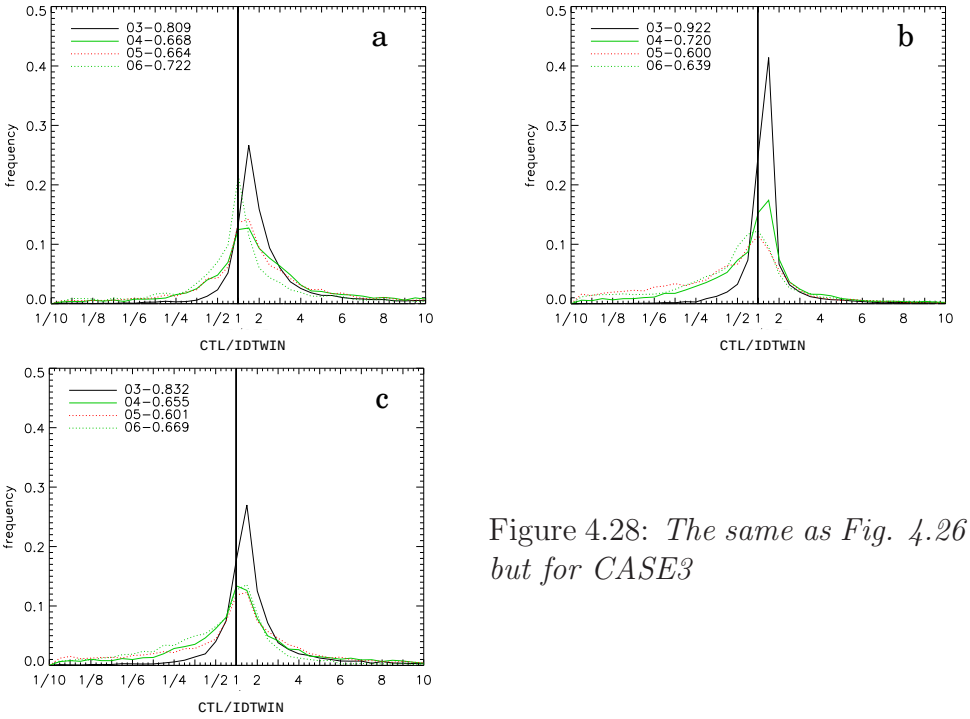


Figure 4.28: *The same as Fig. 4.26 but for CASE3*

The PDF is again calculated only from those grid points where in both IDTWIN and CTL precipitation is greater than 0.1 mm/h. This type of verification doesn't take into account overestimations or underestimations of the precipitative area in the IDTWIN in comparison to the CTL. The frequency BIAS with threshold 0.1 mm/h is given in Fig. 4.29 in order to give a complete information. In all cases all schemes have good scores, but particularly for CASE3 PIB nowind and PIB noqv have the tendency to slightly overestimate precipitative areas during the free run (frequency BIAS bigger than 1.5).

We can conclude that both the assimilation of vertical wind and humidity are useful and necessary.

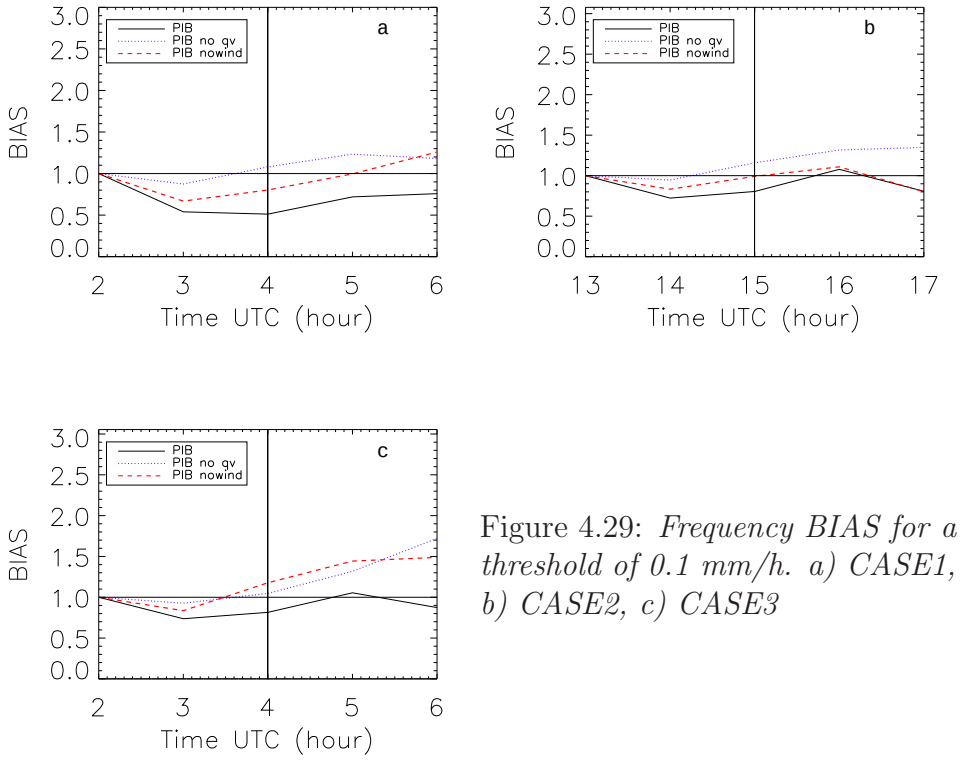


Figure 4.29: *Frequency BIAS for a threshold of 0.1 mm/h. a) CASE1, b) CASE2, c) CASE3*

### 4.5.2 Vertical wind

The RMSD between IDTWIN and CTL of the three dimensional vertical wind field is calculated (Fig. 4.30). The error is similar for all cases. For the complete PIB the error at the beginning of the assimilation window is very high (a mean in the whole field between 0.5 and 0.7 m/s). That is exactly what we expect, since we are changing the vertical wind directly. The error decreases, however, during the assimilation window. The two phase states become even more similar when PIB is applied for a longer assimilation window. The opposite behaviour is visible for PIB nowind and PIB noqv, their application lead to an error that increases with the time. Thus we can gather that the state of the complete PIB tends to return to the CTL state, but that not happens for PIB nowind and PIB noqv.

At the beginning of the free run the four different model simulations (CTL and three PIB versions) have the same boundary conditions but different initial condition. After enough time they will converge to the same phase state, for this reason the error decrease with the time but generally the PIB has a better score (lower RMSD) because it modifies the model dynamics less than

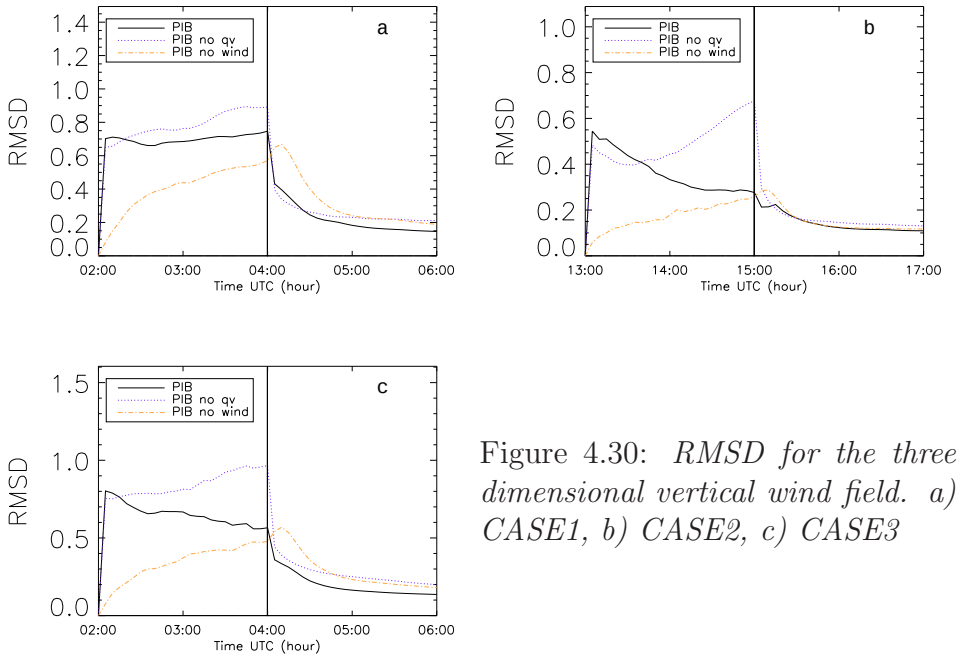


Figure 4.30: *RMSD for the three dimensional vertical wind field. a) CASE1, b) CASE2, c) CASE3*

PIBnowind and PIB noqv.

This effect could be due to the *feedback* effect of the vertical wind assimilation and humidity assimilation. If we assimilate only the humidity (PIB nowind) we overestimate the vertical wind due to false CAPE production. If we assimilate only the vertical wind we enhance the vertical wind in the precipitation field in order to recreate the precipitation. But without the humidity adjustment we need more vertical wind to obtain the same result.

COSMO's response to the assimilation of the merely water component (cloud water content, cloud ice content, water vapour, the PIB nowind) is a strong enhancement of the vertical wind. The maximum values of the vertical wind are plotted in Fig. 4.31. In all PIB versions for all cases the maximum of the vertical wind is overestimated during the assimilation window, the CTL remains between 10 and 20 m/s while the PIB simulations have values of also 50 m/s. The behaviour of PIB and PIB noqv are very similar, that means that the overestimation is originated from the direct wind adjustment. The value increases at the beginning of the assimilation window and remains quite constant during the initialisation time. PIB nowind needs some time to reach these high values and the maximum of the vertical wind fluctuates sharply. This variable also converges to CTL during the free run. Generally the convergence of the complete PIB is faster than the other two.

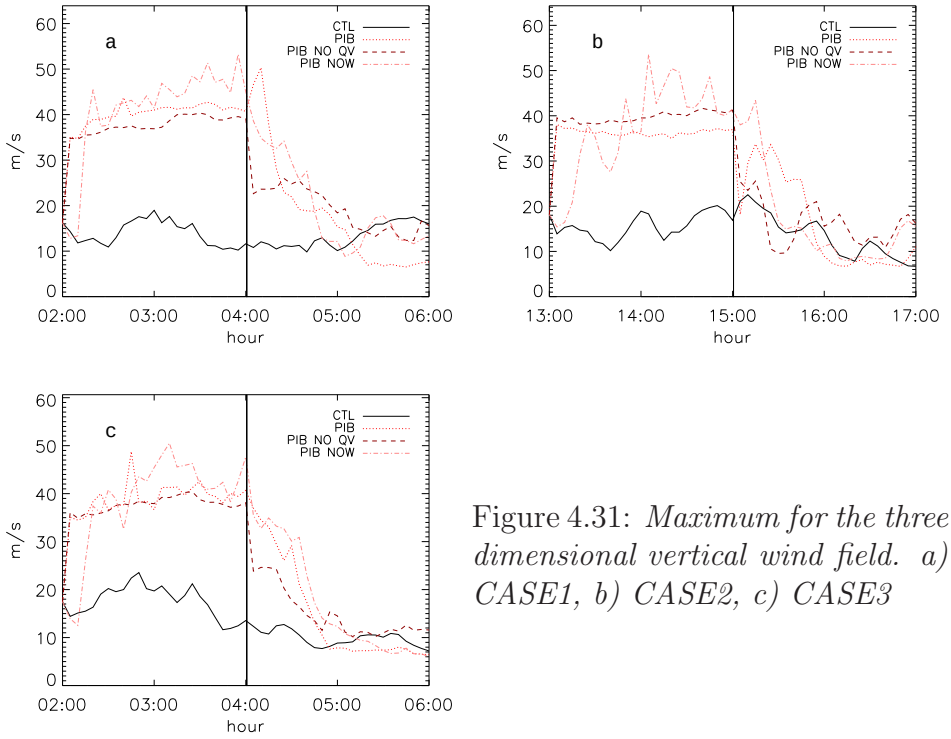


Figure 4.31: *Maximum for the three dimensional vertical wind field. a) CASE1, b) CASE2, c) CASE3*

## 4.6 Summary of the Identical twin experiment

We made an identical twin experiment for three convective cases. A convective line drawn from a synoptic situation, a quasi-chaotic convective event, and a large convective region on mountainous terrain.

The PIB has shown the ability to maintain the main features of the storm evolution, during the assimilation window as well as during the free forecast. The determination of the cloud base in PIB is connected to the amount of CAPE available for the atmosphere. Furthermore the presence of CAPE is of crucial importance for the development of the convection. The link between CAPE and precipitation is obvious. In CASE1 PIB has the tendency to overestimate the cloud base height and consequently to underestimate the precipitation. In CASE2 and CASE3 the coincidence between CTL and IDTWIN is better. This is also proved from the pdf of CTL/IDTWIN (Figs. 4.12,4.13 and 4.14), where the probability of CTL/IDTWIN to lie between 1/3 and 3 at the end of the assimilation window is 0.570 for CASE1 and about 0.7 for CASE2 and CASE3. The probability remains quite constant during the free forecast. The PIB changes the model variables but during the free run the model states do not further diverge.

At the end of the assimilation window the variation of the horizontal wind

is analysed. PIB succeeds in reproducing the mass flux convergence near the cloud base. The relative entropy of the IDTWIN pdf of divergence compare to the CTL pdf of divergence is bigger close to cloud top than near cloud base. However such a problem seems not to have negative consequence for the precipitation field.

PIB mainly consists of the adjustment of the humidity and the adjustment of the vertical wind. In order to test the impact of the two components, two additional simulations are made, with either one part switched off (PIB nowind and PIB noqv). For PIB nowind and PIB noqv the error during the free run increases, for the complete PIB it remains quite constant. Both, humidity and wind, assimilation are necessary in order to avoid the presence of imbalances.

# Chapter 5

# Real data assimilation

In this Chapter the performance of the PIB scheme is investigated with meso-$\gamma$ scale simulation for August 2007 in the COSMO-DE area (Fig. 4.2). The simulations are initialised every eight hours (at 00:00, 08:00 and 16:00 UTC) and consist of forecasts of 9 hours. We want to quantify the improvement achieved using PIB for QPF. The evaluation is made, based on the statistical distribution of the precipitation to take the stochastic nature of precipitation into account, this method is useful in order to understand the main features of the forecast.

The control run (CTL) is created from a nine hours forecast from the COSMO-DE 4.6, the same model as in the previous Chapter. It will be compared with the run using PIB and LHN, both with an assimilation window of two hours followed by 7 hours of free runs.

The model's ability to reproduce the precipitation and the response of the model dynamics will be investigated during the step by step assimilation, where the model state is forced to the *observation state* (Section 5.2), as well as during seven hours of free forecast (Section 5.1).

The simulations are validated with radar-estimated precipitation data (RY-product, see Section 2.2.1), the same data used for the assimilation. Statistiscal scores are used to quantify the quality of the forecasts. An improvement due to the radar data assimilation is well visible, for both assimilation schemes. The suitability of PIB for nowcasting purposes is very similar to the LHN one. However, LHN has the tendency to overestimate the number of grid points with stronger precipitation during the time periods dominated by diurnal convection (simulations initialised at 08:00 and 16:00 UTC).

Furthermore, the sensitivity of PIB to the width of the assimilation window is tested during the three case studies described in Chapter 4. The results show a high performance of PIB using short initialisation period, also 15 minutes. On

the contrary LHN shows poor forecast quality with short assimilation windows.

## 5.1 Evaluation of the free forecast

The precipitation fields of all forecasts for August 2007 are evaluated using statistical quantities. For PIB and LHN we analyse the seven hours of free forecast, for the CTL we take the same time range (that means the seven hours of forecast between the third and the ninth hour).

### 5.1.1 Precipitation PDF and PDF of RADAR/MODEL

The pdf of the ratio of estimated precipitation (RADAR) to the forecasted precipitation (MODEL) is computed (see Section 4.2 for explanations), the computation is based on the accumulated precipitation over August, using hourly data. Together with the pdf of the accumulated precipitation over the seven hours of free forecast (Figs. 5.1, 5.2 and 5.3), it is used to test the ability of the control run and the assimilation methods to simulate the distribution of the precipitation. In order to estimate how well the model pdf approximates the radar pdf, we use as metric the relative entropy of the model probability compared to the radar probability, described in Section 4.4, Eq. 4.1.

The uncertainty of the model pdf is defined using a bootstrapping method (Wilks, 2006 [98]). The used bootstrap method resamples the sequence of days with repetition, in this way we do not disturb the spatial correlation of the pdf.

The pdf of seven hours accumulated precipitation from the three simulations indicates general underestimation of heavy rainfall. The frequency of precipitation rates less than 4 mm is overestimated with respect to the radar data, while stronger rain values are underestimated. This behaviour is particularly strong for CTL. Both assimilation methods are able to reduce these tendencies significantly.

For the simulations initialised at 00 UTC, the behaviour of the three model types are similar. They well approximate precipitation between 0.1 and 12 mm and the relative entropies are very similar, they are within the uncertainty limits. All simulations show underestimation of the precipitation greater than 12 mm. Also for the simulations initialised at 08 UTC, CTL and LHN show very similar relative entropies. PIB shows a lower relative entropy (better skill), due to the better approximation of the precipitation between higher than 4 mm. CTL has the larger underestimation of precipitation greater than 8 mm. PIB and LHN have similar pdf for the simulations initialised at 16 UTC. They underestimate less than CTL the strong precipitation (more than

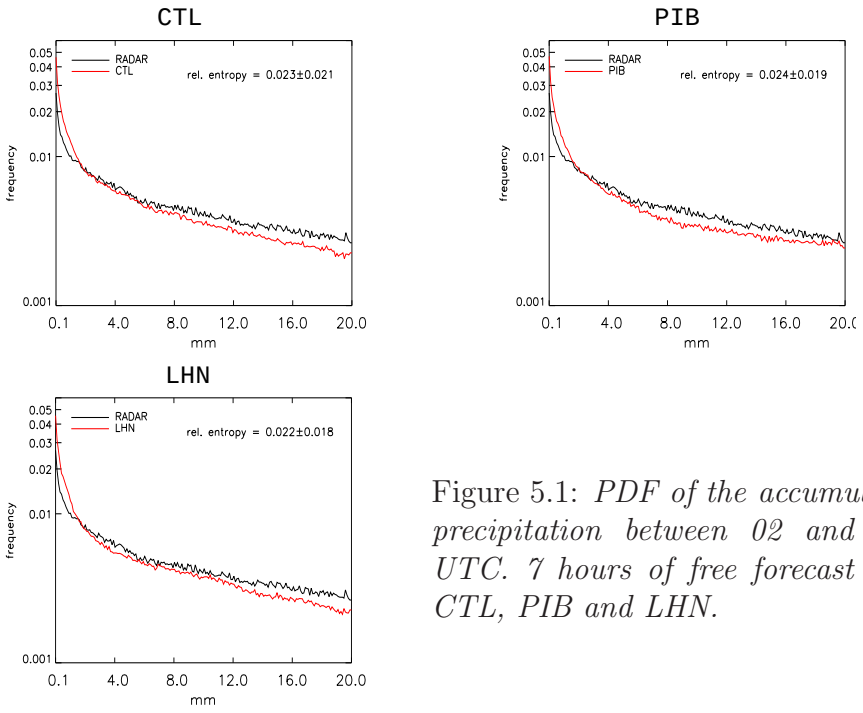


Figure 5.1: *PDF of the accumulate precipitation between 02 and 09 UTC. 7 hours of free forecast for CTL, PIB and LHN.*

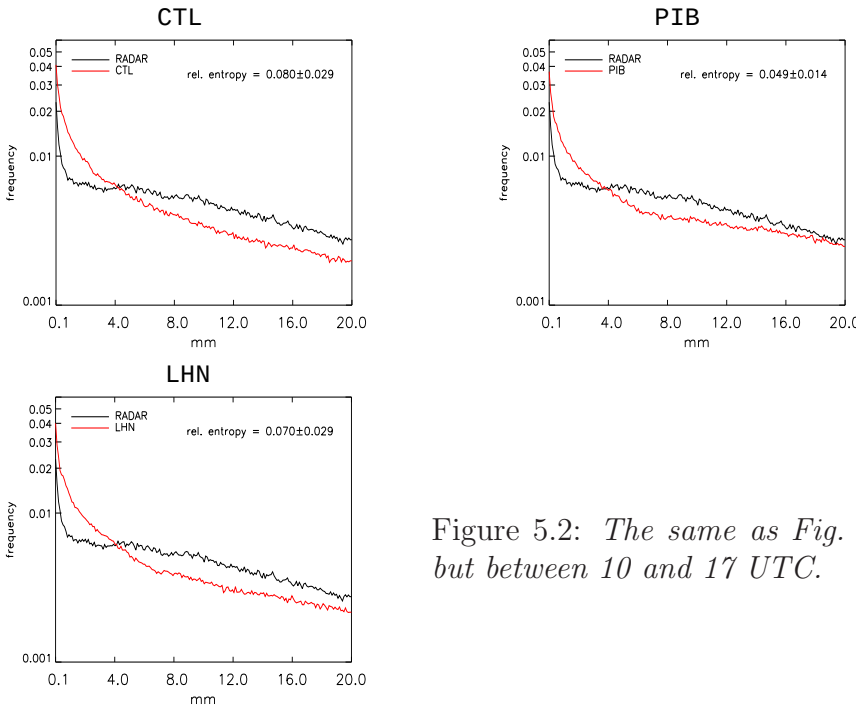


Figure 5.2: *The same as Fig. 5.1 but between 10 and 17 UTC.*

12 mm). In this case the relative entropy of the two assimilation schemes are very similar, while the one of CTL is higher.

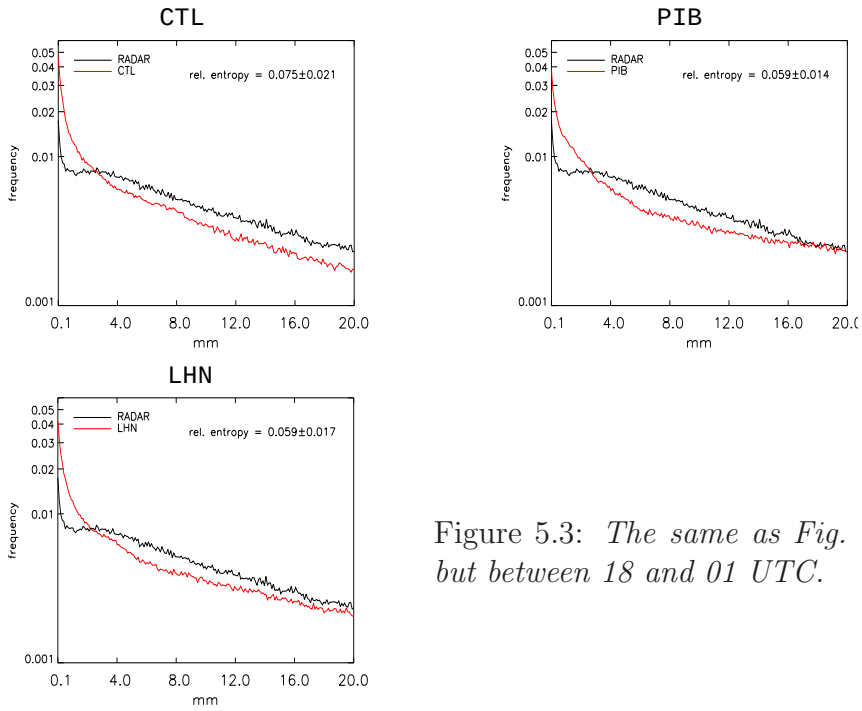


Figure 5.3: *The same as Fig. 5.1 but between 18 and 01 UTC.*

The relationship between the pixel sum of precipitation from the radar data and the model data is described by the pdf of RADAR/MODEL (see Chapter 4 for explanation). For this quantity we take into account the precipitation sum over the entire month during the seven hours of free forecast, for each initialisation time. As in Chapter 4, in order to simplify the comparison between different pdfs the integral between 1/3 and 3 is added.

For the cases initialised at 00 UTC PIB exhibits a slightly lower quality (the reference integral is less than the one of the CTL, 0.818 for PIB and 0.856 for CTL). In the other cases PIB leads to an improvement. The two assimilation methods show comparable results.

CTL tends to underestimate the precipitation (the right tail of the pdf is always bigger then the left). This problem is reduced by both assimilation methods.

The improvement due to the initialisation schemes is clear, the distribution is always closer to one compared to CTL. For the runs initialised at 00 UTC, PIB leads to a larger overestimation than LHN (the curve is higher on the left side). The opposite happens for the simulations initialised at 08 and 16 UTC where PIB overestimates less than LHN, but underestimate.

Both assimilation schemes (especially PIB) improve the forecast when the initialisation is during daytime. Both PIB and LHN employ precipitation/cloud models which better represent convective rainfall.

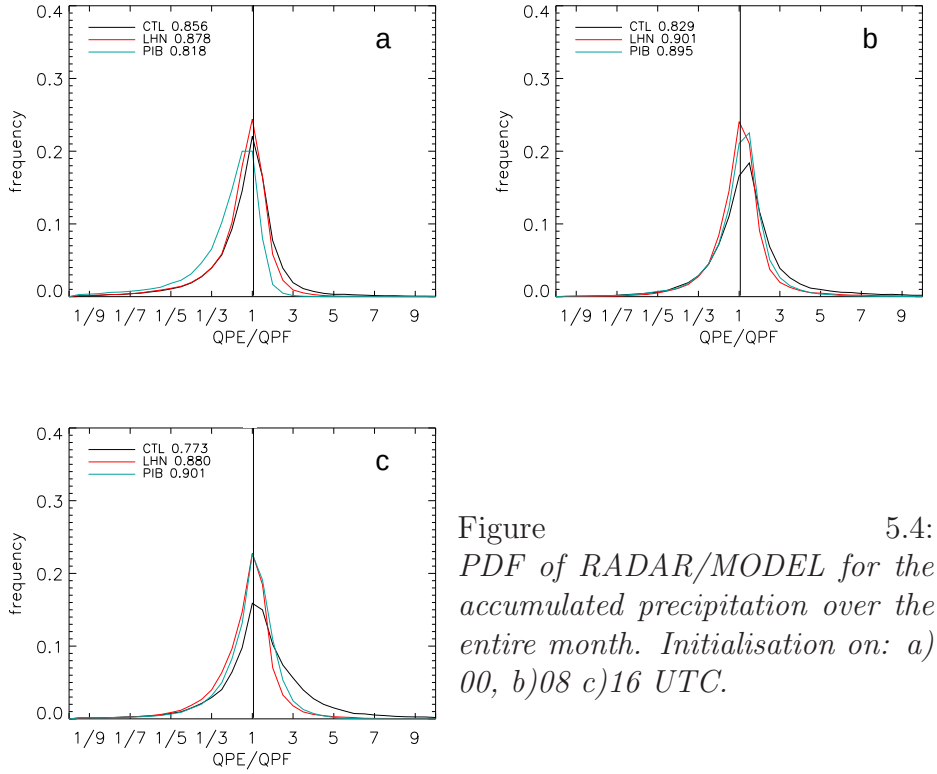


Figure 5.4: *PDF of RADAR/MODEL for the accumulated precipitation over the entire month. Initialisation on: a) 00, b)08 c)16 UTC.*

### 5.1.2 Relative precipitation

We use the quantity total relative precipitation (in per cent) to measure the reproduction of precipitation by the three types of simulations:

$$\frac{R_{rad} - R_{mod}}{R_{rad} + R_{mod}} \cdot 100 \tag{5.1}$$

Where $R_{rad}$ is the radar precipitation and $R_{mod}$ the model precipitation. The precipitation is referred to the precipitation sum over the entire month during the seven hours of free forecast, for each initialisation time.

The quantity is computed for every pixel in the radar area, thus we can identify local bias which may be caused by orography. The value of the total relative precipitation is between $-100$ (the model has precipitation but not the radar) and 100 (precipitation absent in the model but present in the radar). A value of 0 means perfect forecast by the model, values greater (lower) than 0 indicate underestimation (overestimation).

Figs. 5.5, 5.6 and 5.7 display the spatial behaviour of the error of the model precipitation for the three different initialisation times. Nearby the frequency

of the total relative precipitation is shown using bins of 5%. The areas mean the integral on the left side of 0 (probability of overestimation) and on the right side of 0 (probability of underestimation).

Generally CTL exhibits more large errors, both in overestimation and underestimation leading to a flatter distribution. Large areas with strong underestimation/overestimation are visible especially for the runs initialised at 08 and 16 UTC.

For the simulations initialised at 00 UTC all three scheme show higher probability of overestimation, especially PIB. The CTL shows a local large area with high underestimation on the northern side of Germany, this feature is diminished in LHN and completely disappeared in PIB.

For the simulations initialised at 08 UTC, PIB exhibits a better quality in the east side parts (the underestimation has values with less intensity) of the radar composite compared to CTL and LHN. But LHN is better in the centre of the area. CTL has large areas with overestimation and underestimation, as visible from its highly flat distribution (maximum of the frequency about 0.04 and quite constant between $-40\%$ and 40%). Both the assimilation schemes succeed into improve the precipitation frequency, the distributions have both the maximum nearby 0, with a value about 0.06.

For the simulations initialised at 16 UTC a large area of underestimation is present mainly in the centre and in the northern part of Germany for the CTL simulations. This underestimation has a minor intensity in PIB and the improvement is even better for LHN. Also in this case the pdfs of the two assimilation schemes are well distributed around 0, while the CTL distribution is flatter. As in the previous Section we can conclude that both assimilation schemes improve the forecast when the initialisation is during daytime. For both assimilation schemes, errors induce by the orography or local bias could not be determined.

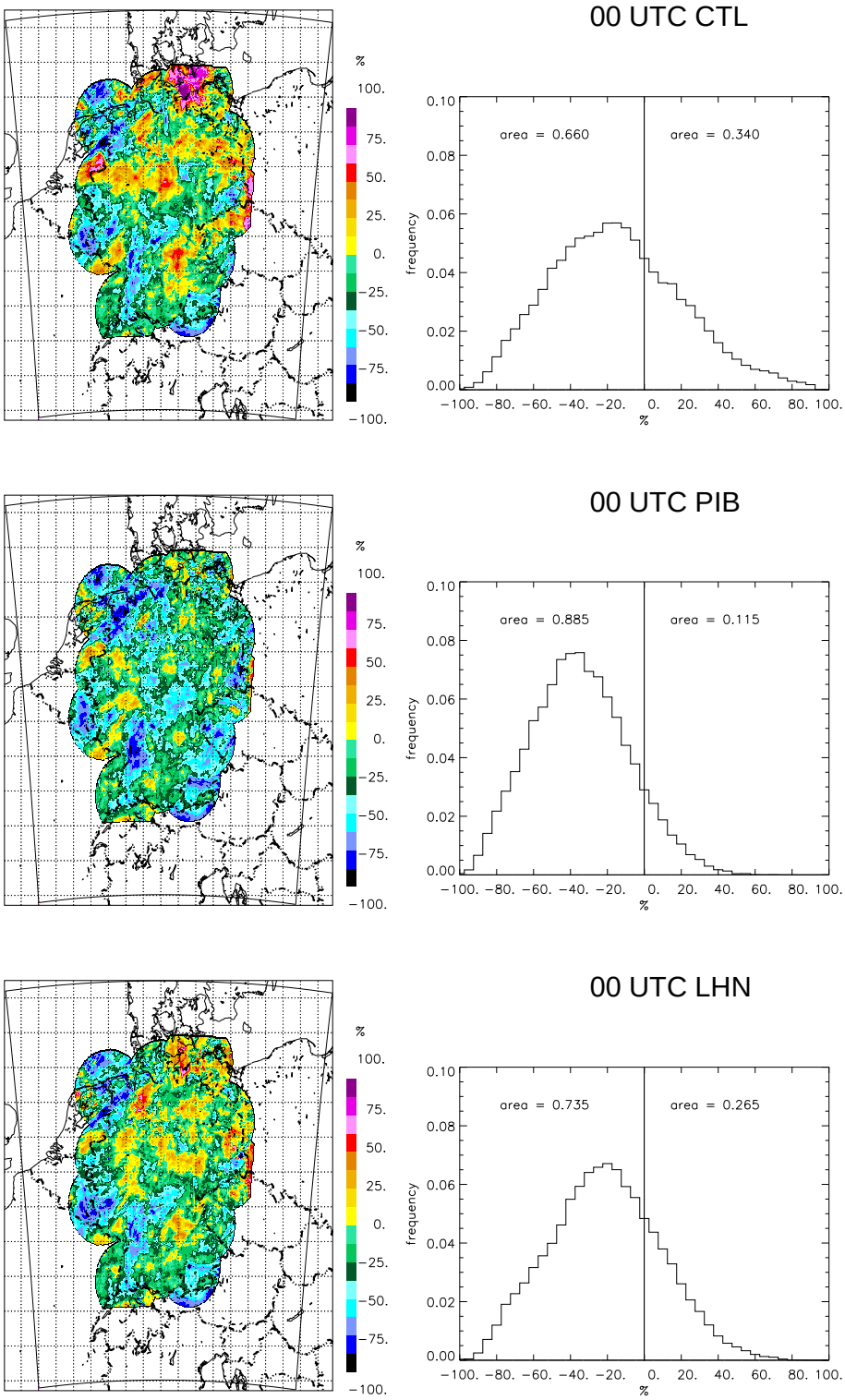


Figure 5.5: *Relative precipitation during the free run, between 02 and 09 UTC. Computation using August 2007. Green and blue colours denote overestimation, yellow and red colours underestimation. Frequency computes using bins of* 5%.

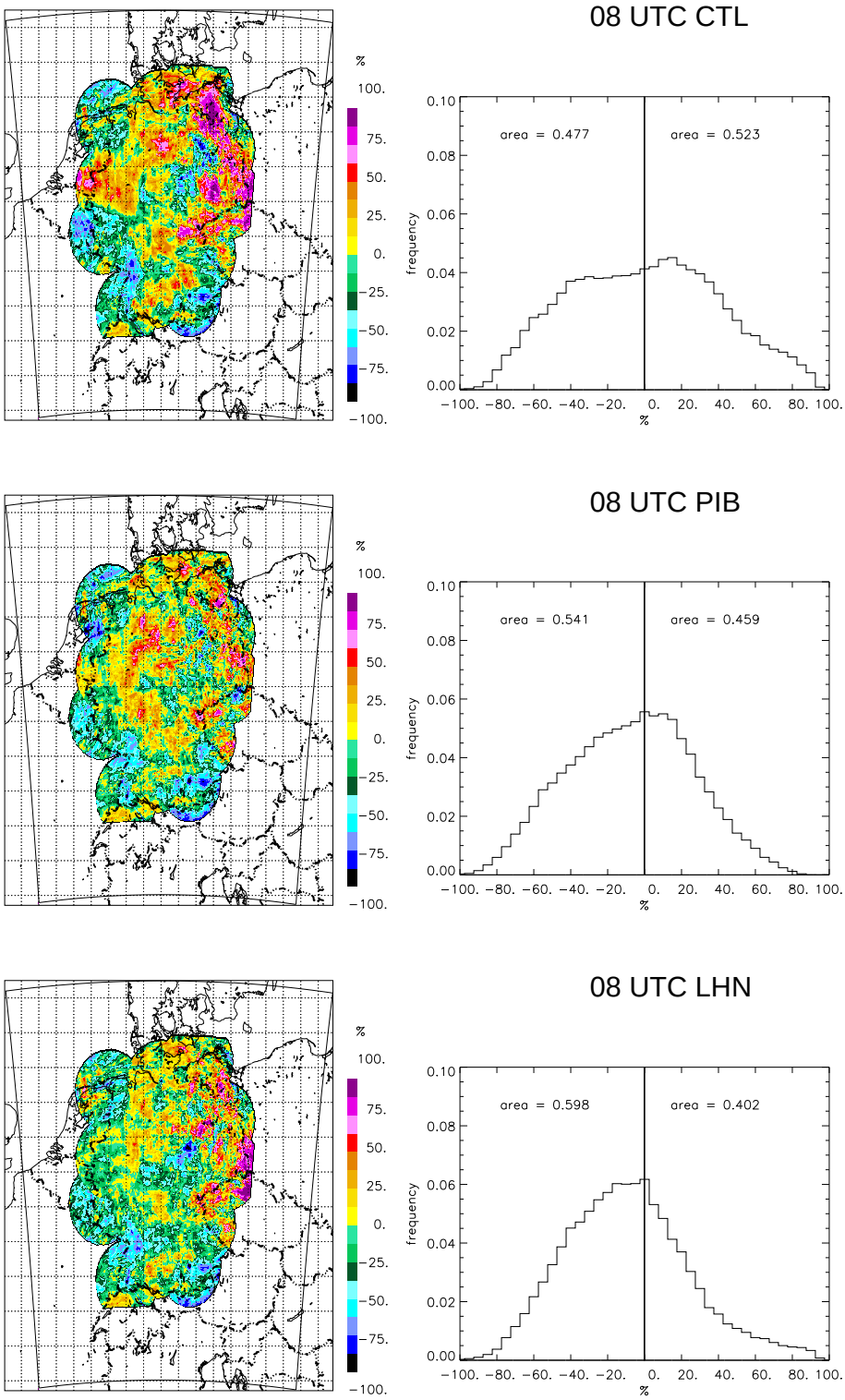


Figure 5.6: *The same as Fig. 5.5 but for the time range between 10 and 17 UTC.*

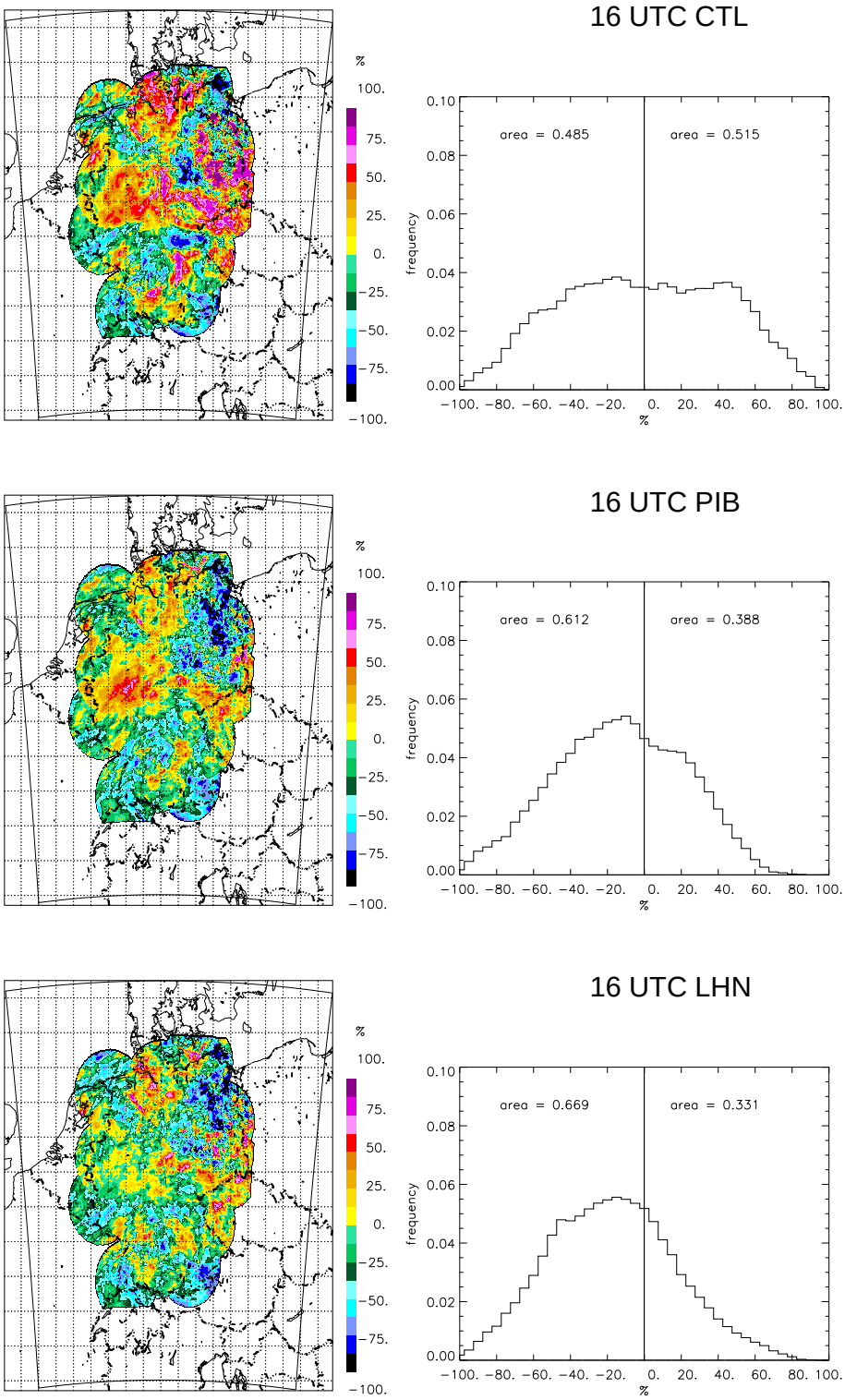


Figure 5.7: *The same as Fig. 5.5 but for the time range between 18 and 01 UTC.*

### 5.1.3 Objective skill scores

In the previous Sections the quality of the forecast is estimated from a statistical and a qualitative point of view. In this Section we want to use objective skill scores based on contingency tables. The ability of the models to reproduce the radar data can be quantify by the ROC (Relative operating characteristic, see appendix A), with the use of the ROC area as additional measure for the comparison between the models. ROC is, however, not sensitive to biases, for this reason the frequency bias is also computed. The ROC shows the behaviour of the forecast in terms of FAR (false alarm rate , see appendix A) and HR (hit rate, see appendix A), the ROC area has a range between 0 and 1 with a perfect score of 1.

The same time periods and forecast ranges as in the previous Section are analysed. The contingency table (Tab. A.1) is computed based on the cumulative precipitation during the forecast time. The skills are then related to the sum of the marginal distributions over the whole month. The quantities are computed for different thresholds (in mm for every 7 hours): 0.1, 0.5, 1.0, 2.0, 5.0, 10.0 .

The behaviour of the three models is shown in Figs. 5.8, 5.9, 5.10.

The improvement of the skills in terms of ROC by using the initialisation schemes is clear, as the ROC area shows an enhancement between 0.05 and 0.1, for both PIB as well as LHN. For all simulations, both curves of PIB and LHN are over the curve of CTL. In every simulation PIB and LHN have lower FAR and higher HR. In all cases LHN shows the lowest FAR for all thresholds, that means the lowest per cent of forecasted yes events that did not occur in the observation, on the contrary PIB shows higher HR for smaller threshold (the fraction of observed yes events correctly forecasted). The ROC areas of PIB and LHN are similar for the simulations initialised at 00 UTC and 16 UTC, for the simulations initialised at 08 UTC the ROC area of LHN is bigger than the PIB area (about 0.06).

The BIAS frequency is influenced by the amount of precipitation. Normally the assimilation schemes leads the model to a state with more precipitation, as shown in Section 5.1.1 from pdf of the ratio RADAR/MODEL and from the total relative precipitation. In the time range between 02 and 09 UTC the CTL has a better skill. The two initialisation schemes have larger BIAS frequencies, the worst one is from PIB. For the other two time ranges (between 10 and 17 UTC and between 18 and 01 UTC) the initialisation schemes improve also the bias frequency. CTL underestimates the precipitation area, while LHN strongly overestimates especially for the higher thresholds. The overestimation by PIB is moderate compared to LHN; this is in agreement with the results show by the ROC. Using a verification based on objective skill scores PIB and

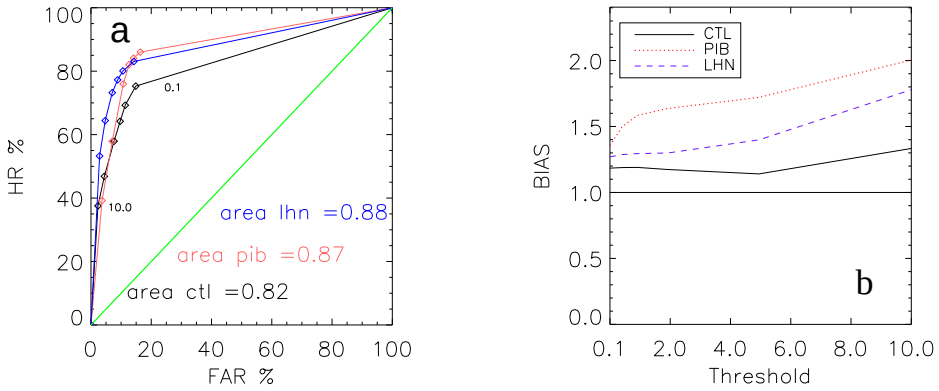


Figure 5.8: *Objective skill scores for the accumulated precipitation during the free forecast between 02 UTC and 09 UTC, for CTL, PIB and LHN. a) ROC, b) frequency BIAS.*

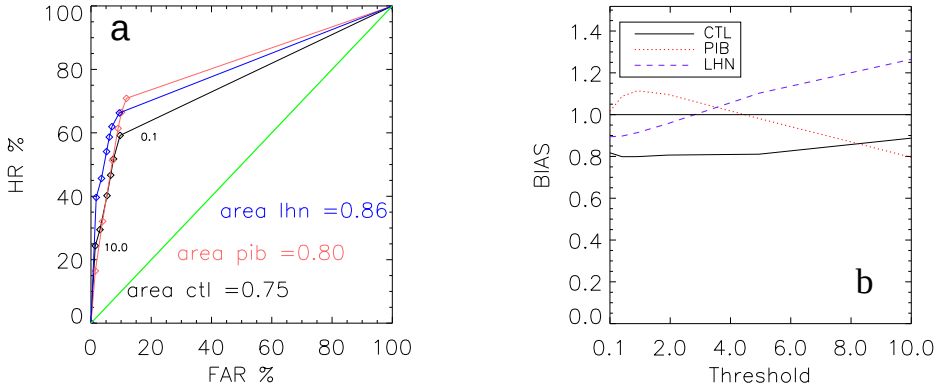


Figure 5.9: *The same as Fig. 5.8 but for the hours between 10 and 17 UTC.*

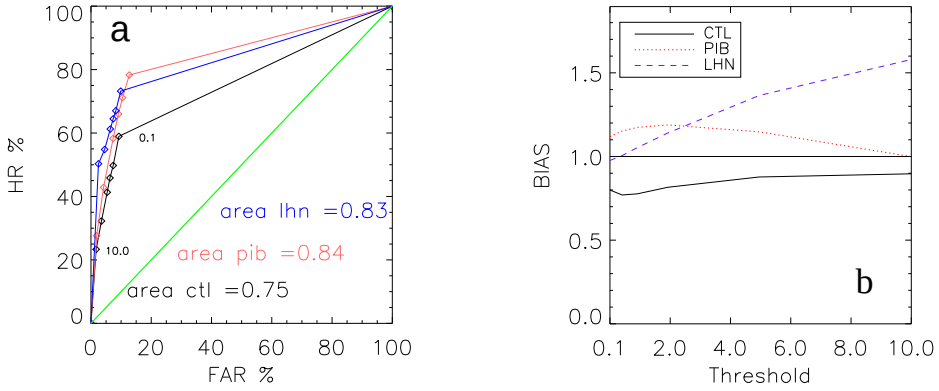


Figure 5.10: *The same as Fig. 5.8 but for the hours between 18 and 01 UTC.*

LHN enhance the quality of the forecast. From these results LHN has the tendency to overestimate the stronger precipitation rates and both assimilation schemes overestimates the precipitation when the assimilation is during the night (from 00 to 02 UTC).

## 5.2 Duration of the assimilation window

The PIB assimilates directly the model variables connected with the convection: specific water vapour content, cloud water and ice content, and vertical wind. Thus we expect the scheme to react faster compared to LHN, which very indirectly forces convection. In this Section we want to quantify this assumption by reducing the assimilation window for PIB and LHN. The shorter the necessary window the more efficient PIB.

The comparison between simulations with different assimilation windows with radar data is made in terms of ETS and BIAS score. A threshold of 0.1 mm/h is used in order to apply the statistics only to the precipitative area. A quantitative evaluation using a fuzzy logic system based on the work of Ebert, 2008 [32] is also applied. A upscaling method is used, forecast and observation are averaged to consecutively larger scale and compared using ETS for a range of thresholds. The basic idea is that even for high resolution forecasts the verification should also take into account the large area mean and not only a grid point value. This accounts in some way for the stochastic nature of convective precipitation. In a consistent forecast the ETS score should decrease with increasing threshold value and it should increase when enlarging the spatial scale. For a better readability the values are colour coded (Figs. 5.13, 5.14, 5.15).

The calculations are based on hourly accumulated precipitation. Different durations of the assimilation window are chosen, from 15 minutes (the time range between two satellite observations) to two hours. We change the starting time of PIB, while the time step when the free run starts is maintained constant. The output is compared to that of LHN. For this study we only analysed the three cases evaluated in the IDTWIN (Chapter 4).

For CASE 1 looking at the skills (BIAS score and ETS) during the free forecast (Figs. 5.11 a and 5.12 a; and Fig. 5.13 for the upscaling) after one hour of forecast (at 04:00 UTC) all assimilation simulations using PIB have a very similar quality, even with a 15 minutes assimilation window. This behaviour persists until the end of the simulation time. The upscaling ETS also shows no large discrepancies between PIB 2h and PIB 15min. Large differences can be found in the BIAS score, where for the run using 2 hours of assimilation window an overestimation during the first three hours of free forecast is present. This

overestimation is decreasing with the decreasing of the assimilation duration, with 15 minutes the BIAS score is everywhere less than one, showing an underestimation of the precipitation.

For this case the simulation with 2 hours of assimilation using LHN exhibits a similar behaviour to the PIB simulations. Looking at ETS and BIAS score, with regard to the upscaling ETS, LHN has a slightly better quality at the beginning of the free forecast. This could not be observed after 3 or 6 hours of forecast. The poor quality of LHN when using an assimilation window of only 15 minutes is well visible. The ETS remains always less than 20 % and the BIAS score denotes an underestimation of the precipitative area. The quality of this forecast is comparable to the quality of the CTL.

For CASE 2 (Figures 5.11 b and 5.12 b) the results using PIB with only 15 minutes of assimilation are reasonable compared with the other simulations. Like in the previous case the PIB overestimates the precipitation area for the longer assimilation windows at the beginning of the free run. However in this case the ETS and the BIAS score at the end of the free run are quite different with better results for the simulations with longer assimilation windows, as well visible also in the upscaling ETS (Fig. 5.14). The LHN shows the same behaviour than in CASE1, with better skills for 2h of assimilation windows and poorer skills for the shorter one, where the ETS is very similar to that of the CTL.

The CTL skills for CASE3 are better than in the other cases, as clear described from the upscaling ETS (Fig. 5.15). ETS and BIAS score for the two assimilation schemes are very similar to that of the first two cases. In this case for PIB the simulation using 15 minutes of assimilation is even better than the one using two hours of assimilation. For LHN the short assimilation run exhibit skills comparable to the PIB, this performance is maybe due to the already good skills of CTL. In this case the LHN with 2h of assimilation shows better skills than PIB.

The result of this test is that PIB proves to be a very fast initialisation scheme, an assimilation window of 15 minutes is sufficient to change the direction of the state of the model to the state of the atmosphere. On the contrary LHN seems to need longer assimilation windows in order to produce comparable results.

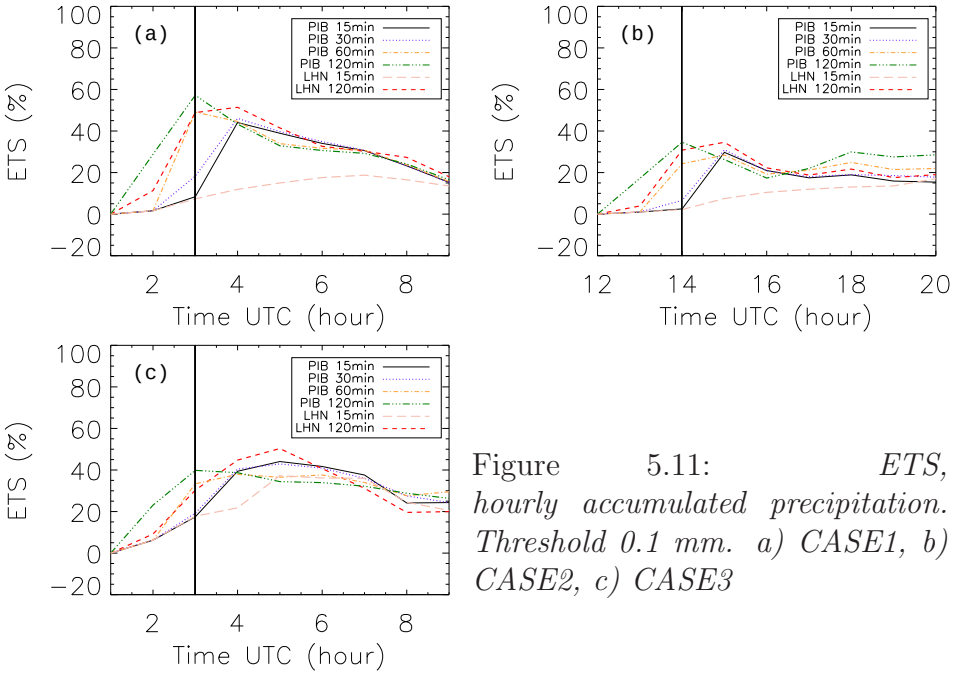


Figure 5.11: *ETS, hourly accumulated precipitation. Threshold 0.1 mm. a) CASE1, b) CASE2, c) CASE3*

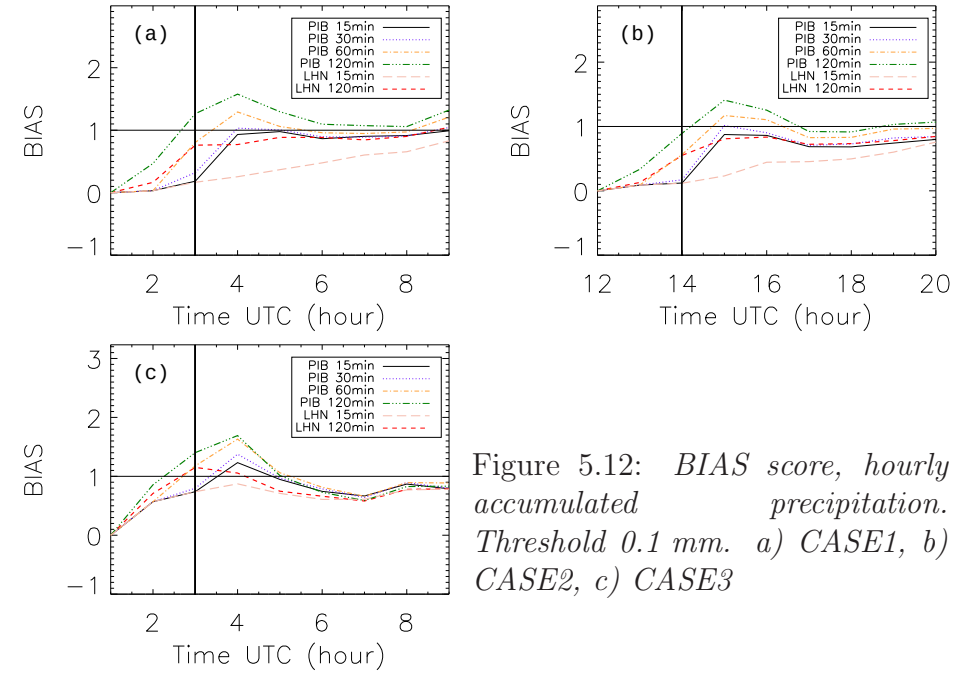


Figure 5.12: *BIAS score, hourly accumulated precipitation. Threshold 0.1 mm. a) CASE1, b) CASE2, c) CASE3*

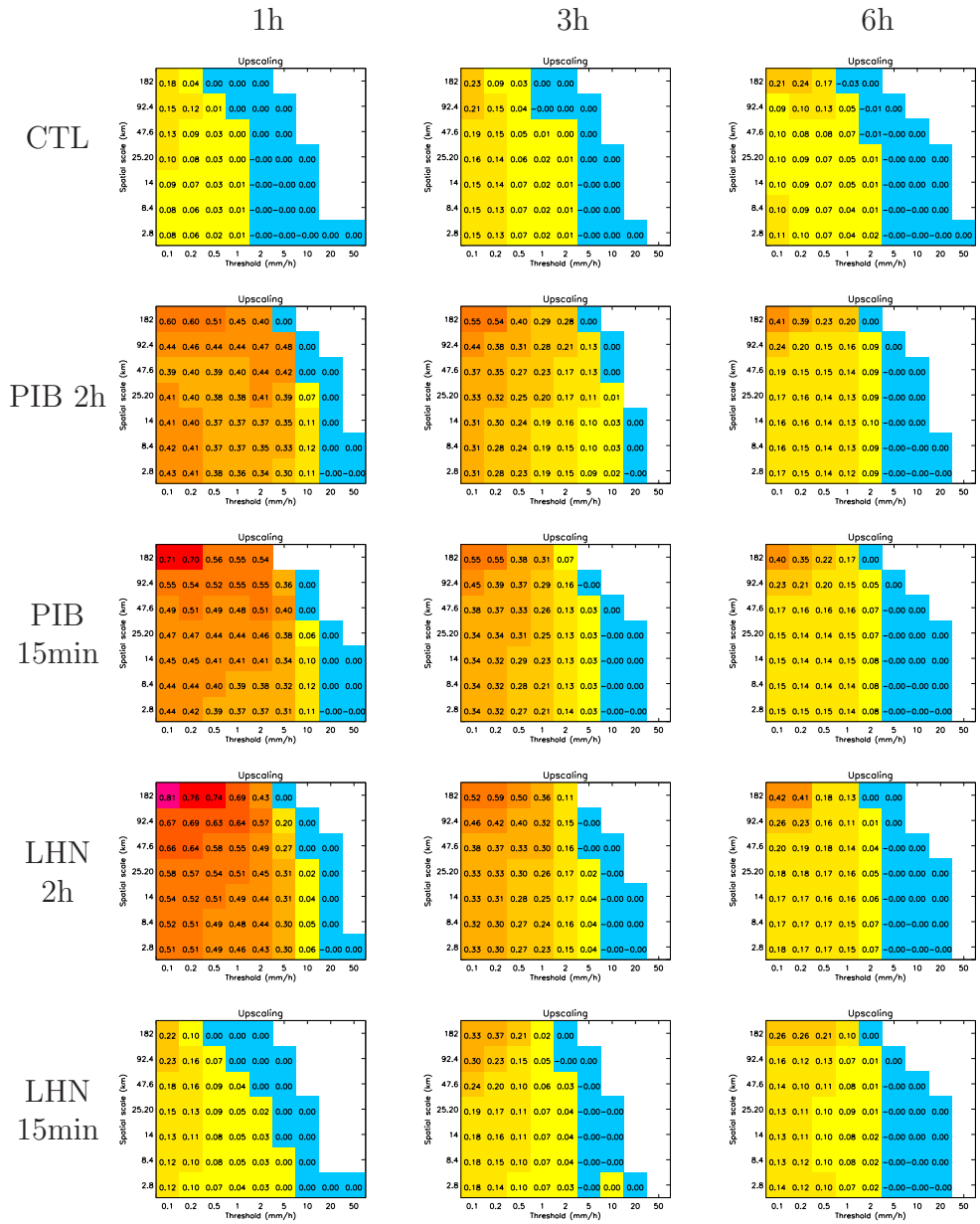


Figure 5.13: CASE1 : ETS for hourly accumulated precipitation. Results after an hour, first column, after three hours, second column, and after six hours of free forecast, third column.

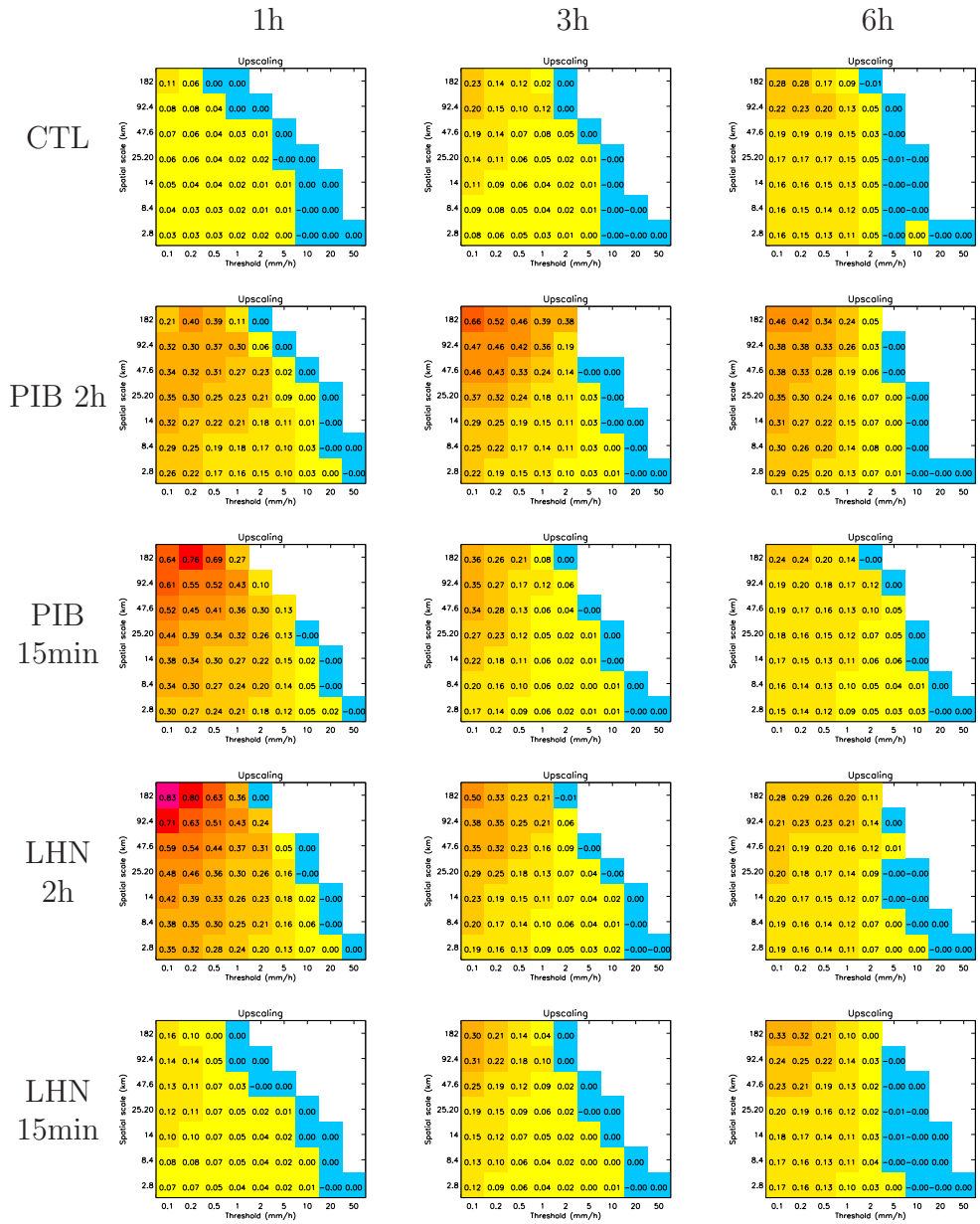


Figure 5.14: CASE2 : ETS for hourly accumulated precipitation. Results after an hour, first column, after three hours, second column, and after six hours of free forecast, third column.

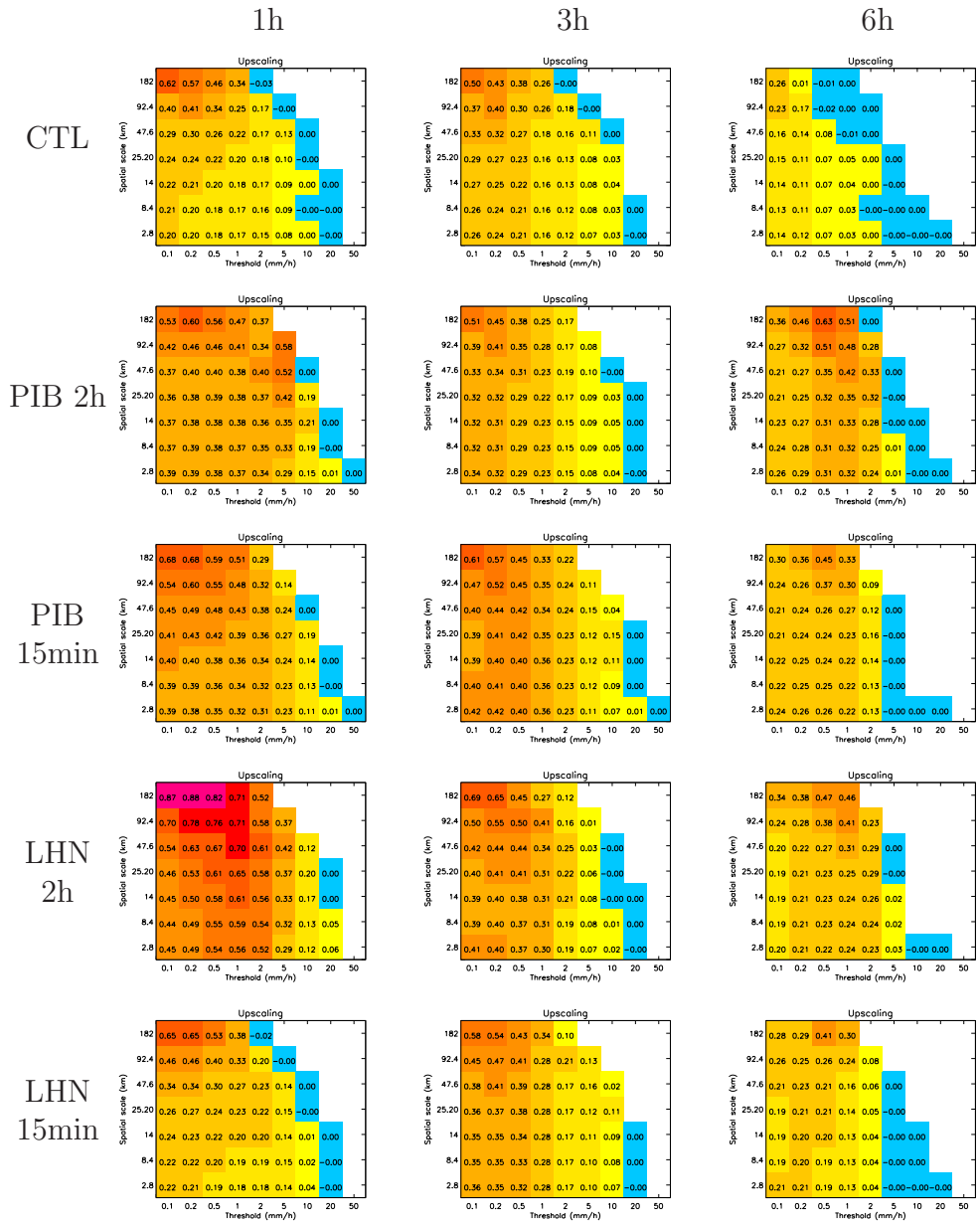


Figure 5.15: CASE3 : ETS for hourly accumulated precipitation. Results aften an hour, first column, after three hours, second column, and after six hours of free forecast, third column.

## 5.3 Summary of the real data experiment

In this work the whole august 2007 is simulated using CTL, PIB and LHN. For every day and every model type three different initial times are used: 00, 08 and 16 UTC. With those simulations we can check the quality of the assimilation schemes in relationship to different initial time. The comparison uses radar-estimated precipitation data as reference.

During the considered month COSMO generally underestimates precipitation, especially for stronger rain rates. LHN and PIB succeed in reducing this error, mainly for the simulations started at 08 and 16 UTC. That is in the case when convection happens.

No local biases or orography influences are detected in this experiments, with and without assimilation.

The quality of PIB for different assimilation windows is tested, for the three cases described in Section 4.1. PIB has demonstrated high quality in the free forecast also with an assimilation window of 15 minutes, especially in the first hours (3-4) of forecast. In this time range the enhancement in the ETS for PIB with short assimilation window is comparable to the one of PIB with two hours of assimilation window, for every threshold and spatial scale.

# Chapter 6

# Conclusions and future research

## 6.1 Synthesis of the results

Radar data can be used to describe mesoscale and convective-scale states of the atmosphere, for this reason they are useful in order to initialise numerical models with a resolution able to describe convective processes. The assimilation of such type of data improves precipitation forecast (Sun, 2005 [89]).

In this work the original *Physical initialisation algorithm developed at the University of Bonn, PIB*, is extended by an improved cloud characterisation and an adaptive efficiency factor. The performances of PIB are analysed by identical twin experiments and forecasts using observations. In the latter one, PIB is applied to assimilate precipitation estimates from weather radar network into operational non-hydrostatic NWP model at midlatitudes.

PIB includes a one-dimensional cloud model (see Section 3), which takes into account two cloud processes (condensation and evaporation) over ice and water, the existence of a mixed phase is also included. The definition of the cloud top is derived from the SAFNWC data, while the cloud bottom is approximated from the model variables. Both cloud bottom and cloud top are very sensitive parameters within PIB, because they determine the total water vapour content (influencing the CAPE) and the vertical wind profile inside the assimilated cloud. Originally the cloud top was defined as fix all over the assimilation window. With the current PIB, the cloud model is connected to the observations.

Besides the cloud variables, the other important assimilated model variable, is the vertical wind (table 3.1). The assumed vertical wind inside the precipitating cloud is derived from the original simplified precipitation model from Haase G., 2002 [41]. In the original PIB the conversion efficiency of

saturated water vapour into rain at the cloud base was general defined for all weather situations, while in the current scheme it is changed using a search algorithm and a dynamic approach.

The new PIB uses the advantage of the cloud analysis also in the assimilation outside the precipitating area. Such an approach leads to an analysed atmospheric state, which is more consistent with the reality. The original scheme has "dried" the columns where the analysed precipitation was zero. For a comparison between the original and the current scheme see Milan et al., 2008 [73].

The use of the prognostic precipitation decreases the positional errors of the rainfall estimates. Several sensitivity studies demonstrate that positional errors of rainfall lead to significantly reduced forecast quality, comparable to intensity errors (Manobianco et al., 1994 [70]). We used a definition of analysed precipitation at the cloud base (see Section 3.1) in order to avoid the errors due to the different height of model and radar precipitation. In this way we apply PIB at places where we actually have precipitation and therefore initialise a convective state.

The PIB nudges the model state to the observed one and it is applied to the variables which are directly related to development and evolution of convection, in particular humidity (e. g. Drucrocq et al., 2002 [31]; Lascaux et al., 2004 [64]). The relationship between PIB assimilation and CAPE is also proved in Section 4.2. The vertical wind is changed in order to give the model a convective profile related to the analysed precipitation.

The change of the model variables could lead the model state to instable condition. We checked the effects of inconsistencies between PIB and the model using an identical twin experiment (Section 4). Using this type of test, the monitoring of the behaviour of all model variables (e. g. the vertical wind), is possible and not only for precipitation.

Here, the main results from this study are assembled:

- A consistency between PIB and COSMO is found using the identical twin test. The dynamic of the model is not mainly influenced, the position of the convective cells remains very close to the original ones. The horizontal wind convergence at cloud base is connected to the convective cells position.

- In the Identical Twin experiment, PIB has the tendency to overestimate the cloud base (see Section 4.3), which leads to a local underestimation of CAPE and also of precipitation.

- The use of the prognostic precipitation leads to a more consistent

assimilation scheme with the observed data. The definition of the analysed precipitation at the cloud base, helps to reach this purpose.

- Both, assimilation of vertical wind and humidity, are useful in order to maintain the model consistency and the quality of the forecast.

- The assimilation of the non precipitative clouds avoid the creation of dry area outside the radar precipitative domains in presence of clouds.

- Assimilation experiments over August 2007 with observed radar data have demonstrated that COSMO simulations with PIB and LHN (Latent Heat Nudging, the operational radar data assimilation scheme in use by the German weather service) provide more realistic precipitation forecast than CTL. Assimilation of radar data has a positive impact on nowcasting of rainfall. From our results the quality of PIB is close to the one of LHN, moreover the LHN has the tendency to enhance the overestimation of the points with strong precipitation (rain rate greater 10 mm in 7 hours).

- PIB and LHN conserve the good impact of the assimilated precipitation for at least three hours of free forecast. Positive results are visible also after 6 hours (see Section 5.2). Both assimilation schemes reduce spin-up and position errors of precipitation forecasts.

- PIB has demonstrated the ability to assimilate the development of the convection when it is triggered from a synoptic situation (as CASE 1), as well as in a quasi-chaotic case (CASE 2).

- An assimilation window of 15 minutes is enough for nudge the model state to the observational state, which is a consequence of changing the variables directly related to convection.

- The quality of the rainfall observation is important. PIB is sensitive to the precipitation amplitude of the target precipitation in that it produces convective events with strong vertical wind in presence of rainfall observations with positive BIAS.

- Non-rain echoes derived from radar clutter can be significantly amplified, because PIB assimilates "stationary convective cells" leading to a strong local precipitation. At the end of the assimilation window these "fake cells" will be advected. With a negative effect in the quality of the forecast.

## 6.2 Future work

PIB and COSMO vertical wind profiles are analysed in Section 3.2, a difference in the position of the maximum is found. In order to enhance the model consistency of PIB the vertical wind profile should be revised.

In this thesis the quality of PIB was verified by using radar data products (RY-data), the same data as used in the assimilation. An evaluation based on other sources of data could be made by using the COPS (Convectively and Orographically induced Precipitation Study, Wulfmeyer et al.,2008 [100] ) data base.

PIB is suitable for nowcasting, the coupling with other nowcasting methods could leads to interesting results, where the enhancement of the QPF reaches longer time ranges.

The radar precipitation patterns can be analysed and the motion field of precipitation can be advected. The lifetime and the state of the convection can be also approximated. The idea is apply PIB to such an algorithm, in this case the output of the "radar forecast" will substitute the radar observation used in PIB.

An ensemble approach to the forecast is useful in order to generate a representative sample of the possible future states. The first crucial problem is therefore to construct a representative ensemble, based on the numerous sources of uncertainty in the forecast (Stensrud et al., 2000 [87]).

Both, LHN and PIB can be used to generate ensemble spread and to enhance ensemble density close to the observations. This goal can be reached using a SIRF (Sequential Importance Resampling Filter). For more information see Van Leeuwen, 2003 [94].

# Appendix A

# Forecast verification methods

This appendix describes the verification methods used in this work. The forecast verification has the aim to assign a quality to the forecast using the available observations (Wilks, 2006 [98], Jolliffe, 2004 [49] and the web site of WWRP/WGNE Joint Working Group on Verification, [101] ).

The quality of the forecast can be described from nine aspects or "attributes" (Murphy, 1993 [75]):,

- **Bias** the correspondence between mean forecast and mean observation.
- **Association** the strength of the linear relationship between forecasts and observations (for example the correlation coefficient).
- **Accuracy** the level of agreement between forecast and truth (the observations). The difference between forecast and observation is the error. The lower the errors, the greater the accuracy.
- **Skill** the relative accuracy of the forecast. The reference forecast is generally an unskilled forecast such as random chance, persistence (defined as the most recent set of observations, *persistence* implies no change in condition) or climatology. Skill refers to the increase in accuracy due purely to the quality of the forecast system. Weather forecasts may be more accurate simply because the weather is easier to forecast, skill takes this into account.
- **Reliability** the average agreement between forecast values and observed values. If all forecasts are considered together, then the *overall reliability* is the same as the *bias*.
- **Resolution** the ability of the forecast to sort or resolve the set of events into subsets with different frequency distributions. This means that

the distribution of outcomes when A is forecasted is different from the distribution of outcomes when B is forecasted. Even if the forecasts are wrong, the forecast system has resolution if it can successfully separate one type of outcome from another.

- **Sharpness** the tendency of the forecast to predict extreme values. To use a counter-example, a forecast of *climatology* has no sharpness. Sharpness is a property of the forecast only.
- **Discrimination** ability of the forecast to discriminate among observations, that is, to have a higher prediction frequency for an outcome whenever that outcome occurs.
- **Uncertainty** the variability of the observations. The greater the uncertainty, the more difficult the forecast will tend to be.

## A.1 Objective skill scores

A variety of Objective skill scores (OSS) can be computed from a yes/no contingency table (see table A.1).

Table A.1: *the contingency table*

| | | **Observation** | | |
|---|---|---|---|---|
| | | yes | no | Total |
| **Forecast** | yes | hits | false alarms | forecast yes |
| | no | misses | correct negatives | forecast no |
| Total | | observed yes | observed no | total |

The contingency table is a useful way to see what types of errors are being made. A perfect forecast system would produce only *hits* and *correct negatives* but neither *misses* or *false alarms* . The elements of the table A.1 are named *marginal distributions*.

In this section we describe some usefull OSS.

**Bias score**

$$BIAS = \frac{hits + false \quad alarms}{hits + misses} \tag{A.1}$$

The *bias score* or *frequency bias* is a number between 0 and infinity with a perfect score of 1. This OSS wants to answer to the question: How did the

forecast frequency of yes events compare to the observed frequency of yes events?

It measures the ratio of the frequency of forecast events to the frequency of observed events. A value less than 1 (more than 1) suggest that the system has the tendency to underforecast (overforecast) events. It is only a measure for the relative frequency not for how well the forecast corresponds to the observation.

**Hit rate**

$$HR = \frac{hits}{hits + misses} \tag{A.2}$$

The *Hit rate* has a value between 0 and 1 with a perfect score of 1. This OSS wants to answer to the question: What fraction of the observed "yes" events were correctly forecasted?

It is sensitive to hits, but ignores false alarms.

**False alarme ratio**

$$FAR = \frac{false \quad alarms}{hits + false \quad alarms} \tag{A.3}$$

The *False alar ratio* has a value between 0 and 1 with a perfect score of 0. This OSS wants to answer to the question: What fraction of the predicted *yes* events actually did not occur?

It is sensitive to false alarms, but ignores misses.

**Relative Operating Characteristic**

The *Relative operating characteristic* (ROC), plots hit rate (HR) vs false alarm rate (FAR), using a set of increasing probability thresholds (for example for Rain, 0.1, 0.5, 1.0, etc.). The area under the ROC curve is used as score. An example is visible on Figure A.1.

The ROC area has a range between 0 and 1 with a perfect score of 1. While 0.5 indicates no skill.

A forecast completely non-discriminatory is represented in the ROC by the diagonal HIT=FAR.

ROC is no sensitive to bias in the forecast, a biased forecast may still have good resolution and produce a good ROC curve.

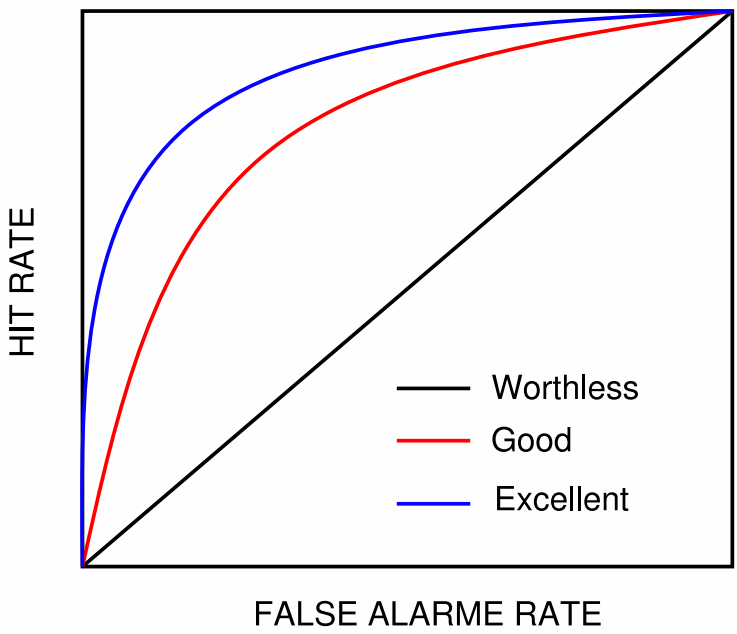


Figure A.1: *example of ROC*

**Threat score**

$$TS = \frac{hits}{hits + misses + false \quad alarms} \tag{A.4}$$

The *Threat score* has a value between 0 and 1 with a perfect score of 1. This OSS wants to answer to the question: How well did the forecast *yes* events correspond to the observed *yes* ?

It measures the fraction of observed and/or forecast events that were correctly predicted. It doesn't take into account the correct negatives and it is sensitive to hits, penalizes both misses and false alarms. It has not a filter for the random chance.

**Equitable threat score or Gilbert skill score**

$$\begin{aligned} ETS &= \frac{hits - hits_{random}}{hits + misses + false \quad alarms - hits_{random}} \\ hits_{random} &= \frac{(hits + misses)(hits + false \quad alarms)}{total} \end{aligned} \tag{A.5}$$

The *Equitable threat score* has a value between $-1/3$ and 1 with a perfect score of 1, a value of 0 indicates no skill. This OSS wants to answer to the question:

How well did the forecast *yes* events correspond to the observed *yes* events (accounting for hits due to chance)?

This OSS is very similar to TS (see above) but it is adjusted for hits associated wit random chance. It is sensitive to hits. Because it penalises both misses and false alarms in the same way, it does not distinguish the source of forecast error.

#### The Brier score

The Brier Score (Wilks, 2006 [98] ) is essentially the mean squared error of the probability forecast, using a binary approach. Considering that the observation $o_k$ can have a value of 1 if an event occurs (for example precipitation more than 0.1 mm/h) or a value of 0 if not. The Brier score (for n points) is defined as:

$$BS = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2 \tag{A.6}$$

Where $y_k$ is the binary value for the forecast. The BS has a value between 0 and 1, 0 for a perfect forecast. The score averages the square differences between points of forecast probability and the subsequent binary observations.

A consequent skill score of the BS is the Brier Skill Score:

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \tag{A.7}$$

This score is conventionally used for a comparison with a reference forecast (for example the climatology). The BSS has a range value between minus infinity to 1. Negative values means that the detected forecast is worst than the reference one.

## A.2 Root mean square difference

The root mean square difference (RMSD, Giorgi and Yunqiang, 2000 [39]) is a variable to measure the deviation between run and the baseline experiment for a given field.

Talking about a general variable $\phi$ the RMSD between a reference run *ref* and another run *ctl* at a certain model level and simulation time, is given by:

$$RMSD = \sqrt{\sum_i \frac{(\phi_i^{ref} - \phi_i^{ctl})^2}{N}} \tag{A.8}$$

the summation is carried out over the number of grid points N. A corresponding bias for the difference is defined from:

$$BIAS_D = \sum_i \frac{(\phi_i^{ref} - \phi_i^{ctl})}{N} \tag{A.9}$$

both quantities are measures of deviation: the RMSD compounds all the magnitudes of deviations at individual points regardless of their sign, while the $BIAS_D$ measures predominant or systematic deviations in given directions.

# Appendix B

# Differences between parallel and sequential simulations in COSMO

At the beginning of this work we run the COSMO model (the former Lokal Modell) in sequential mode, that means with only one processor and without using any parallel subroutine, that was possible because our first *run-domain* was not so big, only 200 points times 200 points (with 35 vertical level).

When we started with the operational COSMO-DE area (421 points times 461 points) we run in parallel mode also the *little domain* with 16 processors and we noted some differences in the output. At the beginning we hadn't explanation for that but we found that LM has different behaviour in case of sequential or parallel runs, due to the different use of a Newtonian approximation.

In this appendix we want to introduce the Newtonian approximation method, and than explain how it is implemented in LM and where is our *source of discrepancies*.

## B.1 The Newtonian approximation method

In many computational cases, general in fields like Physics, Mathematics, Chemistry ... we need to solve equations in form of :

$$f(x) = 0 \tag{B.1}$$

We describe here the *Newton's methods* for the resolution (Pascal and Xavier, 2001 [79]).

### B.1.1 First Newton's method

In the method developed by Newton there are no use of derivative but only application of polynomial equations. Around 1669 Newton should solve the polynomial equation $y^3 + 2y - 5 = 0$. The method starts with the guess of the root, in this case $y \simeq 2$. Then just takes $y = 2 + p$ and after a substitution:

$$p^3 + 6p^2 + 10p - 1 = 0 \tag{B.2}$$

we suppose that $p$ is small $p \ll 1$, for this reason we can take into account only the part $10p - 1 = 0$, we have $p \simeq 0.1$. We have now a better approximation for our root, $y \leq 2.1$. At this point we can iterate this process using $p = 0,1 + q$ and with a substitution we have:

$$q^3 + 6.3q^2 + 11.23q + 0.061 = 0 \tag{B.3}$$

using the same method of the B.2 we can say $q \leq -0.061/11.23 = -0.0054...$ and the new approximation for the root is $y = 2.0946....$ We can iterate this process how many times that we want.

### B.1.2 Newton's method

Nowadays the method is generalize for $C^2$ functions on a given interval, we can use a Taylor's expansion for a variable $h$ in a given interval near $x$, where $f(x) = 0$:

$$f(x + h) = f(x) + hf'(x) + O(h^2) \tag{B.4}$$

If we stop at the first order (linearization of the equation) we are looking for a small $h$ such as:

$$f(x + h) = 0 \leq f(x) + hf'(x) \tag{B.5}$$

giving

$$h = -\frac{f(x)}{f'(x)} \tag{B.6}$$

$$x + h = x - \frac{f(x)}{f'(x)} \tag{B.7}$$

The Newton iteration start with an initial guess of the root $x_0$, then find the limit of recurrence:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{B.8}$$

## B.2 Newton's method in COSMO for the saturation adjustment

The saturation adjustment in COSMO is called in many modules but at the end of the dynamical time stepping (*leapfrog* or *two time levels* ) there are some differences in the case of we are running in sequential mode or in parallel.

The Newton's iteration method is used for solved the equation for the temperature.

If we defined the enthalpy related temperature $T_h$ ( Doms et al., 2005 [28]) as:

$$T_h = T + \frac{L_v}{c_{pd}} q^v \tag{B.9}$$

where $q^v$ is the specific humidity, $L_v$ the latent heat of vaporisation and $c_{pd}$ the specific heat of dry air at constant pressure.

If we general defined $\tilde{x}$ the provisional values of $x$, in the case of the saturated cloud we have, for the total water content ( $q^T = q^v + q^c$, where $q^c$ is the liquid water content) $\tilde{q}^T \leq q^{v*}_{sw}$, with $q^{v*}_{sw}$ the saturation specific humidity.

We have for the temperature:

$$T = \tilde{T}_h - \frac{L_v}{c_{pd}} q^v_{sw}(T, p) \tag{B.10}$$

Where $p$ is the pressure. The B.10 is solved with the Newton's iteration method:

$$T^\nu = \frac{\tilde{T}_h - \frac{L_v}{c_{pd}} \left\{ q^v_{sw}(T^{\nu-1}, p) - T^{\nu-1} \left( \frac{\partial q^v_{sw}}{\partial T} \right)_{T^{\nu-1}} \right\}}{1 + \frac{L_v}{c_{pd}} \left( \frac{\partial q^v_{sw}}{\partial T} \right)_{T^{\nu-1}}} \tag{B.11}$$

where $\nu$, a natural number, denotes the iteration index.

The temperature at time level $n$ is used as an initial estimate for the iteration, i.e. $T^{\nu=0} = T^n$. Following the last iteration step, the saturation specific humidity is not calculated from $T^N$ but extrapolated from its value for $T^{N-1}$ using a linearly truncated Taylor series expansion. This is done to satisfy thermal energy conservation for the approximate iterative solution. The final values of the variables at time level $n + 1$ are given by:

$$\begin{aligned} T &= T^N \\ q^v &= q^v_{sw}(T^{N-1}, p) + (T^N - T^{N-1}) \left( \frac{\partial q^v_{sw}}{\partial T} \right)_{T^{N-1}} \\ q^c &= \tilde{q}^T - q^v \end{aligned} \tag{B.12}$$

By default, the number of iterations $N$ is set to one, it changes in relation of the maximal vertical wind of the considered level:

$$\begin{aligned} w < 2m/s &\implies N = 1 \\ 2m/s \leq w < 10m/s &\implies N = 2 \\ w \geq 10m/s &\implies N = 3 \end{aligned} \tag{B.13}$$

In the sequential case the maximal is calculated between the all point in the level for the simulation, instead for the parallel only between the point in the processor.

# Appendix C

# CAPE

CAPE (Convective Available Potential Energy) is defined as the amount of the available energy for the convection of a particular parcel (Emanuel 1994, [35]).

On the diagram in figure C.1 CAPE is the positive area between the lifted parcel profile (curve with $\theta_{e,p}$ constant) and the environmental sounding, from the parcel's level of free convection (LFC) to its level of neutral buoyancy (LNB). It may be defined as:

$$CAPE_i = \int_{p_n}^{p_i} (\alpha_p - \alpha) dp \tag{C.1}$$

Here $\alpha$ is the environmental specific volume profile, $\alpha_p$ is the specific volume of a air parcel moving upward moist-adiabatically from the level of free convection, $p_i$ is the pressure at the level of free convection, and $p_n$ is the pressure at the level of neutral buoyancy. All is integrated in $dp$ (differential of pressure). The value depends on whether the moist-adiabatic process is considered reversible or irreversible (conventionally irreversible) and whether the latent heat of freezing is considered (conventionally not).

SPC (Storm Prediction Centre) forecaster at the USA national weather service often refer to "weak instability" (CAPE less than 1000 Jkg-1), "moderate instability" (CAPE from 1000-2500 Jkg-1), "strong instability" (CAPE from 2500-4000 Jkg-1), and "extreme instability" (CAPE greater than 4000 Jkg-1); from the web side of the NOAA (National Oceanic & Atmospheric Administration), Thompson [91] .

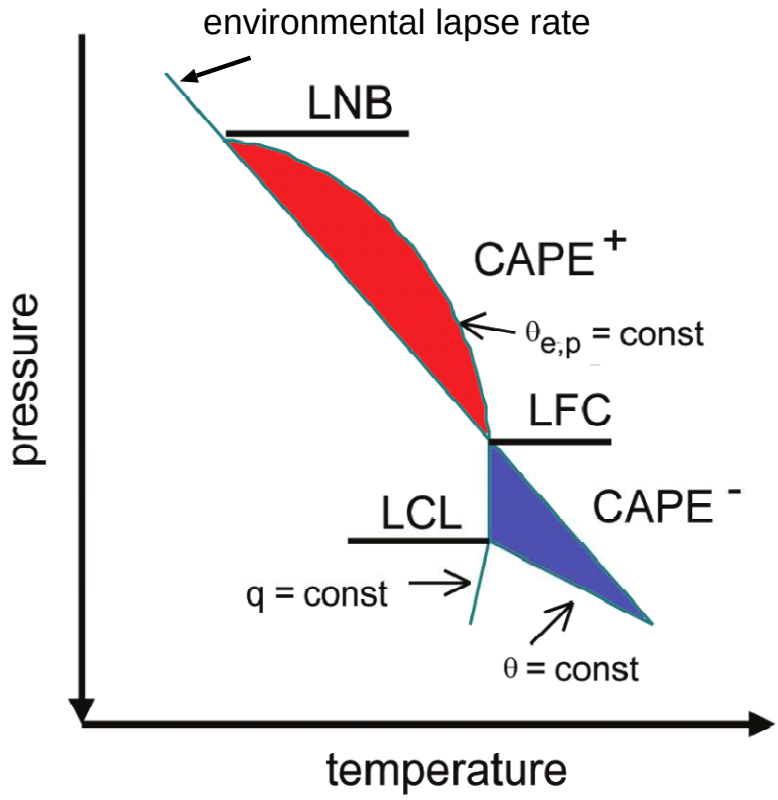


Figure C.1: *The CAPE, in our calculations only the positive part is considered, is defined as the integral between the level of free convection and the level of neutral buoyancy. Here q is the specific humidity, θ the potential temperature and $\theta_{e,p}$ equivalent potential temperature, picture from Hense, 2002 [44].*

# Appendix D

# Additional figures of chapter 4

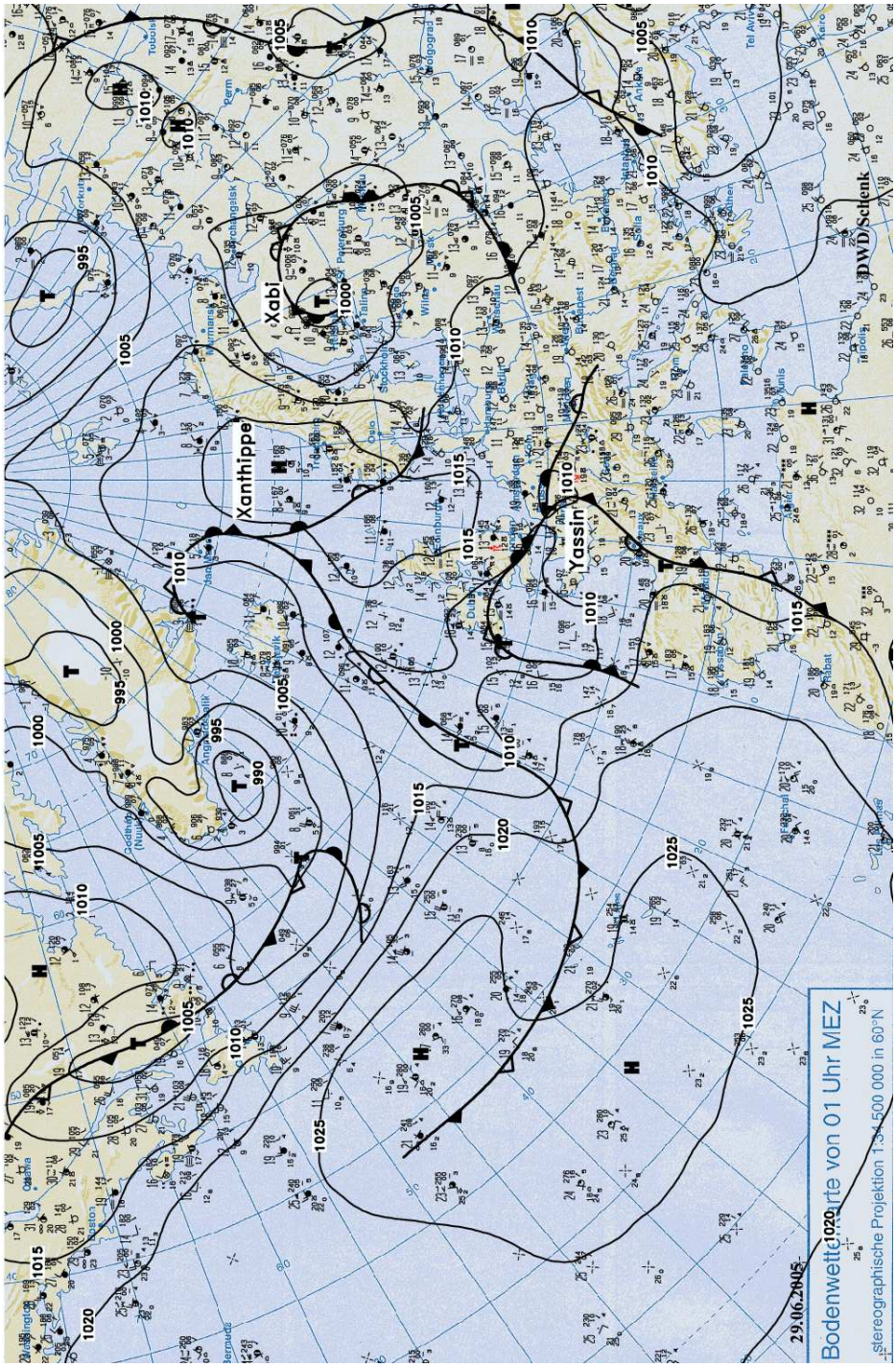


Figure D.1: Berliner wetterkarte: surface pressure with front systems, 29/06/2005 at 00:00 UTC

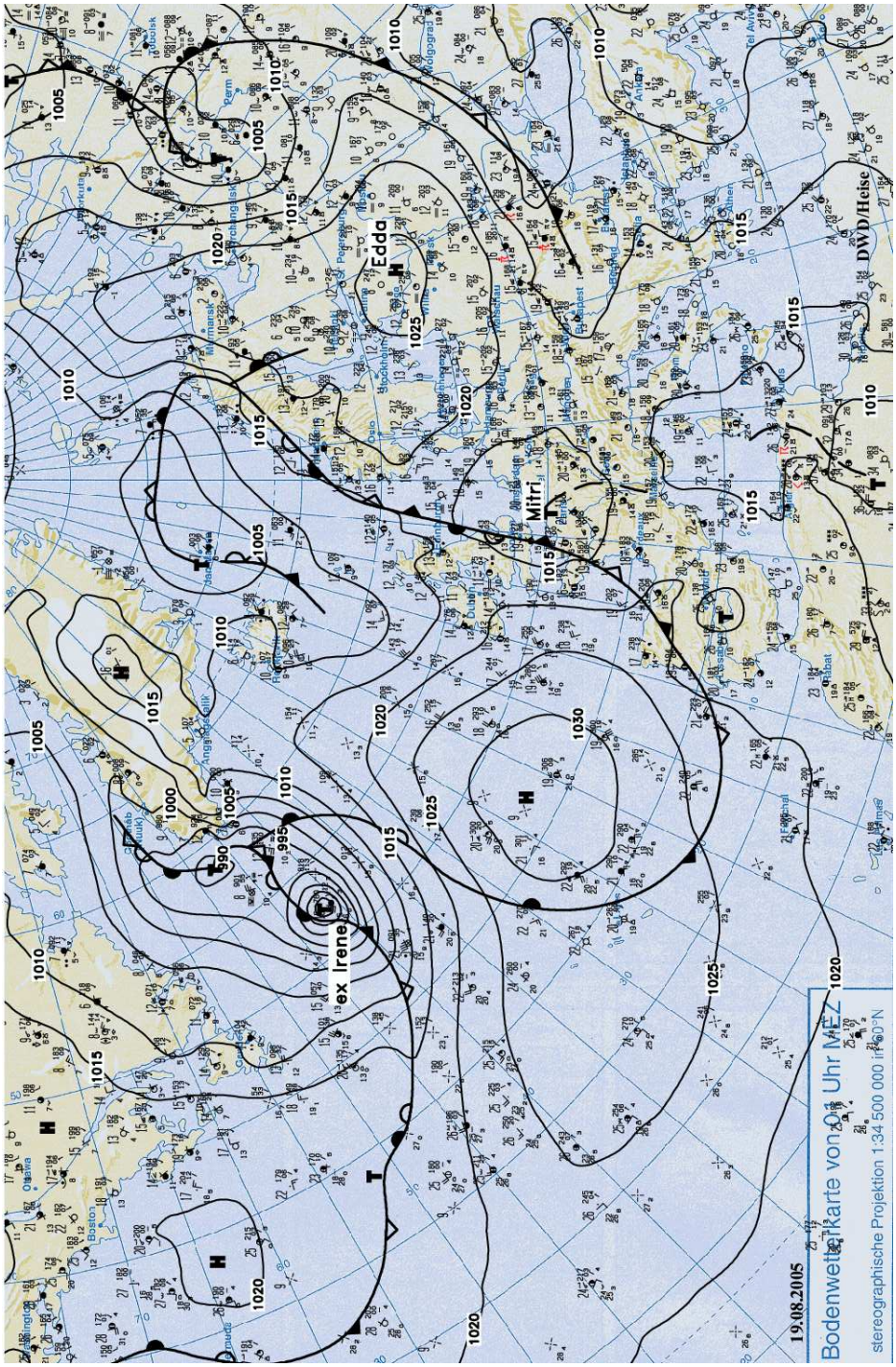


Figure D.2: Berliner wetterkarte: surface pressure with front systems, 19/08/2005 at 00:00 UTC

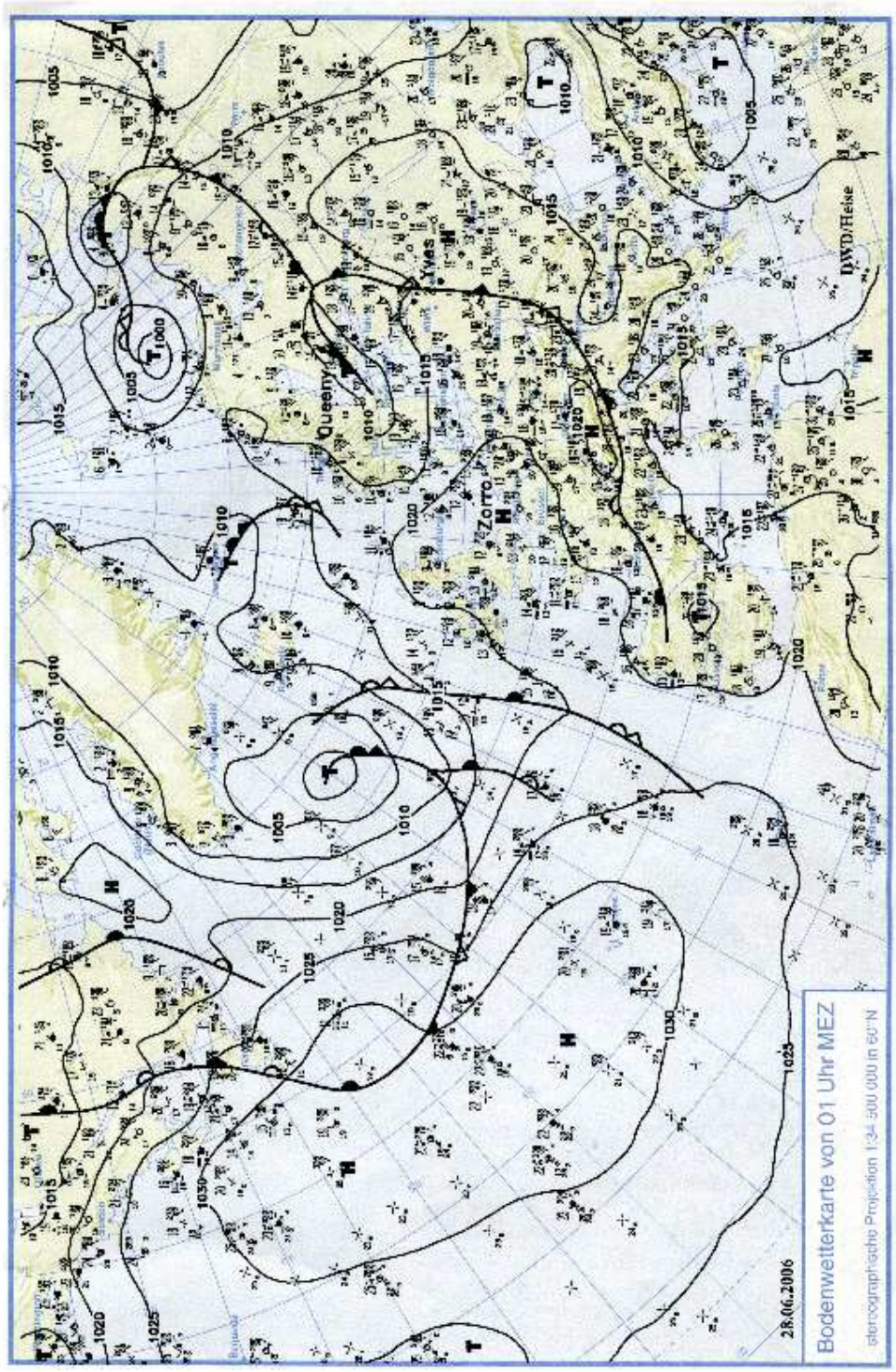


Figure D.3: Berliner wetterkarte: surface pressure with front systems, 28/06/2006 at 00:00 UTC

# Bibliography


[1] S. C. Albers. *The LAPS Wind Analysis*, 1995. `http://citeseer.ist.psu.edu/688.html`.

[2] S. C. Albers, John A. McGinley, Daniel L. Birkenheuer, and John R. Smart. *The Local Analysis and Prediction System (LAPS): Analyses of Clouds, Precipitation, and Temperature.* `http://citeseer.ist.psu.edu/67944.html`.

[3] F. Ament. Initialisierung von Wolken im Lokal-Modell aus Meteosat-Messungen. Diploma thesis, University of Bonn, Germany, 2001.

[4] S. R. Anderson, R.J Graham, and M.J. Bader. The impact of observations on mesoscale model forecasts of three-hourly rainfall accumulations. *Meteorological Applications*, 7:193–203, 2000.

[5] R. A. Anthes. Data assimilation of hurricane prediction models. *Journal of the Atmospheric Sciences*, 31:702–719, 1974.

[6] C. P. Arnold and H. C. Dey. Observing-Systems Simulation Experiments: past, present, and future. *Bulletin of the American Meteorological Society*, 67:687–695, 1986.

[7] D. Atlas. *Radar in meteorology: Battam memorial and 40th anniversary radar meteorology conference.* American Meteorological Society, 1990.

[8] M. Baldauf, J. Förstner, S. Klink, T. Reinhardt, C. Schraff, A. Seifert, and K. Stephan. *Kurze Beschreibung des Lokal-Modells Kürzestfrist LMK und seiner Datenbanken auf dem Datenserver des DWD.* German Weather Service (DWD), Research Department, P.O. 100465, D-63004 Offenbach, 2006.

[9] B. R. Bean and E. J. Dutton. *Radio meteorology.* Dover Publications, 1968.

[10] J. Bech, D. Bebbington, B. Codina, A. Sairouni, and J. Lorente. *Evaluation of atmospheric anomalous propagation conditions: an application for weather radars*, 199. EUROPTO Conference on Remote Sensing. Barcelona, Spain.

[11] P. Bechthold, E. Bazile, F. Guichard, P. Mascart, and E. Richard. A mass-flux convection scheme for regional and global models. *Quarterly Journal of the Royal Meteorological Society*, 127:869–886, 2001.

[12] D. Bolton. The computation of equivalent potential temperature. *Monthly Weather Review*, 108:1046–1053, 1980.

[13] F. Bouttier and P. Courtier. Data assimilation concepts and methods. Training course notes, ECMWF, 1999.

[14] G. H. Bryan, Wyngaard J.C., and Fritsch J.M. Resolution Requirements for the Simulation of Deep Moist Convection . *American Meteorological Society*, 131:2394–2416, 2003.

[15] H. R. Byers and R. R. Braham. The thunderstorm project. NTIS PB234515, US Wheather Bureau, 1949.

[16] S.W. Chang and T.R Holt. Impact of assimilating ssm/i rainfall rates on numerical winter cyclones. *Monthly Weather Review*, 122:151–164, 1994.

[17] J. G. Charney. *Dynamical forecasting by numerical process, Compendium of meteorology*. American Meteorological Society, Boston, MA, 1951.

[18] J. G. Charney, R. Fj$\phi$rtoft, and J. Von Neuman. Numerical integration of barotropic vorticity equation. *Tellus*, 2:237–254, 1950.

[19] J.H. Chen and K. Miyakoda. A nested grid computation for barotropic free surface atmosphere. *Monthly Weather Review*, 102:181–190, 1974.

[20] W. R. Cotton and R. A Anthes. *Storm and cloud dynamics*. Academic Press, 1989.

[21] W. R. Cotton and G. J. Tripoli. On the predictability of individual deep convective clouds. Conference on Numerical weather and prediction, 5th, Monterey, CA; United States; 2-6 Nov., 1981.

[22] National Research Council. *The Atmospherich Sciences:Entering the Twenty-first Century*. National Academy Press, 1998.

[23] J. P. Craven, R. E. Jewell, and H. E. Brooks. Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Wheather and Forecasting*, 17:885–890, 2002.

[24] G. Cressman. An operational objective analysis system. *Monthly Weather Review*, 87:367–374, 1959.

[25] N. A. Crook. Sensitivity of convection initiation to low-lewel thermodynamics fields. *Monthly Weather Review*, 124:1767–1785, 1996.

[26] R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, first edition, 1991.

[27] E. De Lima Nascimento and K. K. Droegemeier. Dynamic adjustment in a numerically simulated mesoscale convective system: Impacto of the velocity field. *Journal of the Atmospheric sciences*, 63:2246–2268, 2006.

[28] G. Doms, J. Förstner, E. Heise, H. J. Herzog, M. Raschendorfer, R. Schrodin, T. Reinhardt, and G. Vogel. *A description of the nonhydrostatic limited-area model LM (Lokal-Modell). Part II: Physical Parameterization* . German Weather Service (DWD), Research Department, P.O. 100465, D-63004 Offenbach, 2005.

[29] G. Doms and U. Schättler. *A description of the nonhydrostatic limited-area model LM (Lokal-Modell). Part I: Dynamisc and Numerics*. German Weather Service (DWD), Research Department, P.O. 100465, D-63004 Offenbach, 2002.

[30] G. Doms, U. Schättler, and J.P. Schulz. *Kurze Beschreibung des Lokal-Modells LM und seiner Datenbanken auf dem Datenserver (DAS) des DWD*. German Weather Service (DWD), Research Department, P.O. 100465, D-63004 Offenbach, 2003.

[31] V. Ducrocq, D. Ricard, J.-P. Lafore, and F. Orain. Storm-scale numerical rainfall prediction for five precipitating events over France: On the importance of the initial humidity field . *Weather and Forecasting*, 17:1236–1256, 2002.

[32] E. E. Ebert. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorol. Appl.*, 15:51–64, 2008.

[33] H. Elbern, H. Schmidt, and A. Ebel. Variational data assimilation for tropospheric chemistry modeling. *Journal of Geophysical Research*, 102:967–985, 1997.

[34] E. E. Elizabeth, U. Damrath, W. Wergen, , and M. E. Baldwin. The WGNE assessment of short term quantitative precipitation forecasts. *Bulletin of the American Meteorological Society*, 84:481–492, 2003.

[35] K. A. Emanuel. *Atmospheric convection*. Oxford University Press, 1994. 580 pp.

[36] T. Fehr. *Mesoskalige Modellierung der Produktion und des dreidimensionalen Transports von Stickoxiden durch Gewitter*. PhD-Thesis, Deutsches Zentrum für Luft- und Raumfahrt e.V., institut für Physik der Atmosphäre, Oberpfaffenhofen, 2000.

[37] J. M. Forsythe, T. H. V. Haar, and D. L. Reinke. Cloud-base height estimates using a combination of meteorological satellite imagery and surface reports. *Journal of Applied Meteorology*, 39:2336–2347, 2000.

[38] U Germann and I. Zawadzki. Scale-dependence of the predictability from continental Radar images. Part I: Description of the metodology . *Monthly Weather Review*, 130:2859–2873, 2002.

[39] F. Giorgi and B. Xunqiang. A study of internal variability of a regional climate model. *Journal of Geophysical Research*, 105:29,503–29,521, 2000.

[40] S Gollvik. On the effects of horizontal diffusion, resolution and orography on precipitation forecasting in a limited area model. *Meteorological Applications*, 6:49–59, 1999.

[41] G. Haase. *A physical initialization algorithm for non-hydrostatic weather prediction models using radar derived rain rates*. PhD-Thesis, Bonn University, 2002.

[42] G. Haase, S. Crewell, C. Simmer, and W. Wergen. Assimilation of radar data in mesoscale models: Physical initialization and latent heat nudging. *Physics and Chemestry of the Earth*, 25:1237–1242, 2000.

[43] E.J. Harrison and R.L. Elsberry. A method for incorporating nested finte grids in the solution of systems of geophysical equations. *Journal of the Atmospheric Sciences*, 29:1235–1245, 1972.

[44] A. Hense. Theoretische meteorologie 2. Lecture notes, 2002.

[45] J. E. Hoke and R. A. Anthes. The initialization of numerical models by a dynamic-initialization technique . *Monthly Weather Review*, 104:1551–1556, 1976.

[46] J. A. Curry Holton, J. R. and and J. A. Pile. *Encyclopedia of Atmospheric sciences*. Academic Press, first edition, 2003.

[47] X. Hong, M. Leach, and S. Raman. A sensitivity study of convective cloud formation by vegetation forcing with different atmospheric conditions. *Journal of Geophysical Research*, 34:2008–2028, 1995.

[48] Q. Jiang, T. Awaji, N. Sugiura, S. Masuda, K. Takeuchi, and N. Shikama. An Identical Twin Experiment for the Development of A 4D-VAR Data Assimilation System for the ARGO Data. *AGU Fall Meeting Abstracts*, pages C268+, December 2002.

[49] I. T. Jolliffe and D. B. Stephenson. *Forecast Verification*. Wiley, 2004.

[50] C. D. Jones and B. Macpherson. A latent heat nuging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Applied*, 4:269–277, 1997.

[51] R.W. Jones. *A nested grid method for a three dimensional model of a tropical cyclone* , 1997. Rept. National Hurricane resarch Lab., Miami, Fla.

[52] J.S. Kain and J. M. Fritsch. Convective parameterisation in mesoscale models: The Kain-Fritsch scheme. *Meteorological Monographs*, 24:165–170, 1993.

[53] E. Kalnay. *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

[54] S.. Klink and K. Stephan. Assimilation of radar data in the lm at dwd. *COSMO newsletter*, 4:143–150, 2004.

[55] T. N. Krishnamurti and H.S. Bedi. A brief review of physical initialization. *Meteorology and Atmospheric Physics*, 60:137–142, 1996.

[56] T. N. Krishnamurti, H.S. Bedi, William Heckley, and Kevin Ingles. Reduction of spinup time for evaporation and precipitation spectral model. *American Meteorological Society*, 116:907–920, 1988.

[57] T. N. Krishnamurti, H.S. Bedi, and K. Ingles. Physical initialization using SSM/I rain rates. *Tellus*, 45:247–269, 1993.

[58] T. N. Krishnamurti, K. Ingles, S. Cocke, and T. Kitade. Details of low latitude medium range numerical weather prediction using a global spectral model, part ii: effects of orography and physical initialization. *Journal of the Meteorological Society of Japan*, 62:613–649, 1984.

[59] T. N. Krishnamurti, J. Xue, H.S. Bedi, K. Ingles, and D. Oosterhof. Physical initialization for numerical weather prediction over the tropics. *Tellus*, 43:53–81, 1991.

[60] V. Kuell, A. Gassman, and A. Bott. Towards a new hybrid cumulus parametrization scheme for use in non-hydrostatic weather prediction models. *Quarterly Journal of the Royal Meteorological Society*, 133:479–490, 2007.

[61] S. Kullback. *Information theory and statistics*. Dover Publications Inc., second edition, 1968.

[62] H. L. Kuo. Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *Journal of the Atmospheric Sciences*, 31:1232–1240, 1974.

[63] S. Laroche, W. Szyrmer, and I. Zawadzki. A microphysical bulk formulation based on scaling normalization of the particle size distribution. Part II: data assimilation into physical processes . *Journal of the Atmospheric Sciences*, 62:4222–4237, 2005.

[64] F. Lascaux, E. Richard, C. Keil, and O. Bock. Impact of the MAP reanalysis on the numerical simulation of the MAP-IOP2a convective sistem . *Meteorol. Z.*, 13:49–54, 2004.

[65] D. Leuenberger. *High-resolution radar rainfall assimilation: Exploratory studies with latent heat nudging*. PhD-Thesis, Meteoschweiz, 2005.

[66] D. Leuenberger and A. Rossa. *Revisiting the latent heat nudging scheme for the rainfall assimilation in convective system*, 2004. ERAD 2004.

[67] R. E. Lopez. The lognormal distribution and cumulus cloud populations. *Monthly weather review*, 105:865–872, 1977.

[68] E.N. Lorenz. The predictability of a flow which possesses many scales of motion. *Tellus*, 21:289–307, 1969.

[69] B. Macpherson, B. J. Wright, W. H. Hand, and A. J. Maycock. The impact of MOPS moisture data in the U.K. meteorological office mesoscale data assimilation scheme. *Monthly Weather Review*, 124:1746–1766, 1996.

[70] J Manobianco, S. Koch, V.M. Karyampudi, and A.J. Negri. The impact of assimilating satellite-derived precipitation rates on numerical simulations of the ERICA IOP 4 cyclone. *Monthly Weather Review*, 122:341–365, 1994.

[71] F Mesinger and A. Arakawa. *Numerical metheods used in atmospheric models*. GARP publications, 1976.

[72] Meteo France. *User Manual for the PGE01-02-03 of the SAFNWC/MSG: Scientific part*, 2004.

[73] M. Milan, Venema V., Schüttemeyer D., and Simmer C. Assimilation of radar and satellite data in mesoscale models: A physical initialization scheme. *Meteorol. Z.*, 17:887–902, 2008.

[74] N. L. Miles, J. Verlinde, and E. E. Clothiaux. Cloud droplet size distributions in low-level stratiform clouds . *Journal of the Atmospheric Sciences*, 57:295–311, 2000.

[75] A. H. Murphy. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8:281–293, 1993.

[76] A. Orlandi, F. Meneguzzo, G. Messeri, A. Ortolani, M. Pasqui, M. Rossi, and A. Terzo. *Satellite rainfall assimilation to improve the quantitative precipitation forecasting*. IATA Institute for Agrometeorology and Environmental Analysis - National Research Council , Via G. Caproni 8, I-50144 Firenze, Italy, 2001.

[77] S. K. Park. Nonlinearity and predictability of convective rainfall associated with water vapor perturbations in a numerically simulated storm. *Journal of Geophysical Research*, 104:575–587, 1999.

[78] S. K. Park. Nonlinearity and predictability of convective rainfall associated with water vapour perturbations in a numerically simulated storm. *Journal applied meteorology*, 104:31:575–587, 1999.

[79] S. Pascal and G. Xavier. *Newton's method and high order iterations*, 2001. `http://numbers.computation.free.fr/Constants/Algorithms/newton.html`.

[80] R. A. Pielke. *Mesoscale Meteorological Modeling*. Academic Press, 1984.

[81] L. F. Richardson. *Weather prediction by numerical processes*. Cambridge University Press, Cambridge. Reprinted Dover 1965, newYork, 1922.

[82] E. R. Rinehart. *Radar for Meteorologists*. Rinehart publications, 2004.

[83] R. R. Rogers and M. K. Yau. *A short course in cloud physics*. Pergamon Press, 1989.

[84] J. Schulz and U. Schättler. *Kurze Beschreibung des Lokal-Modells LME und seiner Datenbanken auf dem Datenserver des DWD*. German Weather Service (DWD), Research Department, P.O. 100465, D-63004 Offenbach, 2005.

[85] S. C. Sherwood, P. Minnis, and M. McGill. Deep convective cloud-top heights and their thermodynamic control during CRYSTAL-FACe. *J. Geophys. res.*, 109:doi10.1029/2004JD00811, 2004.

[86] Y. Shibagaki, M. D. Yamanaka, S. Shimizu, H. Uyeda, A. Watanabe, Y. Maekawa, and S. Fukao. Meso-$\beta$ to-$\gamma$-scale wind circulations associated with precipitating clouds near baiu front observed by the mu and meteorological radars. *Journal of the Meteorological Society of Japan*, 78:69–91, 2000.

[87] D.J Stensrud, J.W. Bao, and T.T. Warner. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective system . *Monthly Weather Review*, 128:2077–2107, 2000.

[88] R. B. Stull. *Boundary layer meteorology*. Kluwer Academic Publishers, 1999.

[89] J. Sun. Convective-scale assimilation of radar data: Progress and cahllenges. *Quarterly Journal of the Royal Meteorological Society*, 131:3439–3463, 2005.

[90] O. Talagrand. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan Special Issue*, 75:191–209, 1997.

[91] R. Thompson. *Explanation of SPC severe weather parameters*. `http://www.spc.noaa.gov/exper/mesoanalysis/help/begin.html`.

[92] M. Tiedtke. A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Monthly Weather Review*, 117:1779–1800, 1989.

[93] L. W. Uccellini, P. J. Kocin, and J. M. Sienkiewicz. Advances in forecasting extratropical cyclogenesis at the national meteorological center. In M.A. Shapiro and S. Gronas, editors, *The live of Extratropical Cyclones*, pages 317–336. AMS, 1999.

[94] P.J. Van Leeuwen. A Variance minimizing filter for large scale applications . *Monthly Weather Review*, 131:2071–2083, 2003.

[95] J. M. Wallace and P. V. Hobbs. *Atmospheric science an introductory survey*. Academic Press, 1977.

[96] T. M. Weckwerth, D. B, S. E. Koch Parsons, J. A. Moore, M. A. LeMone, B. B. Demoz, C. Flamant, B. Geerts, J. Wang, and W. F. Feltz. An overview of the international $H_2O$ project (IHOP 2002) and some preliminary highlights. *Bulletin of the American Meteorological Society*, 85:253–277, 2004.

[97] S. K. Weisman and J. B. Klemp. The dependence of numerically simulated convective storms on vertical wind shear and buayancy. *Monthly Weather Review*, 110:504–520, 1982.

[98] D. S. Wilks. *Statistical methods in the atmospheric sciences*. Academic Press, 2006.

[99] B. J. Wright. *The moisture observation preprocessing system (MOPS).* , 1993. U.K. Meteorological office forecasting research division tech. rep. 38, 23pp.

[100] V. Wulfmeyer, Behrendt A., Bauer H. S., Kottmeier C., Corsmeier U., Blyth A., Craig G., Schumann U., Hagen M., Crewell S., Di Girolamo P., Flamant C., Miller M., Montani A., Mobbs S., Richard E., Rotach M. W., Arpagaus M., Russchenberg H., Schlussel P., Konig M., Gartner V., Steinacker R., Dorninger M., Turner D. D., Weckwerth T., Hense A., and Simmer C. THE CONVECTIVE AND OROGRAPHICALLY INDUCED PRECIPITATION STUDY. A Research and Development Project of the World Weather Research Program for Improving Quantitative Precipitation Forecasting in Low-Mountain Regions . *Bullettin of the American Meteorological Society*, 89:1477–1486, 2008.

[101] WWRP/WGNE. *Joint Working Group on Verification*. `http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html`.

[102] L. Xin and G. W. Reuter. Numerical simulation of the effects of mescale convergence on convectinve rain showers. *Monthly Weather Review*, 124:2828–2842, 1996.

[103] K. M. Xu and D. A. Randall. Updraft and downdraft statistics of simulated tropical and midlatitude cumulus convection. *Journal of the Atmospheric sciences*, 58:1630–1649, 2001.

[104] X. Zou and Q. Xiao. Studies on the initialization and simulation of a mature hurricane using a variational bogus data assimilation scheme. *Journal of the Atmospheric Sciences*, 57:836–860, 2000.

# Acknowledgment

I would like to gratefully acknowledge the supervision of Prof. Dr. Clemens Simmer for his scientific support, relevant discussions and his logical way of thinking during this work. His patience and personal advice were essential for the successful completion of this thesis. I thank Prof. Dr. Andreas Hense for his helpful suggestions and for reading this thesis as second reviewer. Furthermore I thank Prof. Dr. Bernd Diekkrüger and Prof. Dr. Armin Skowronek for their reviews.

This study was supported by Deutschen Forschungsgemeinschaft (DFG) under grant SI60617-1/2. Thanks to the Deutscher Wetterdienst (DWD) for providing the COSMO model and the data.

I am grateful to Dr. Dirk Schüttemeyer and Dr. Victor Venema for their guidance and inspiration during the development of this work. I have appreciated the many theoretical discussion with Volker Kuell and Henning Wilker. In addition, I would like to thank all the people who have improved this work with their ideas and experience, particularly all DAQUA members.

I am grateful to current and former members of the Meteorological Institute of the University of Bonn (MIUB) for their friendship and kind help, especially the proof-readers of this thesis. For the friendly atmosphere I will always remember the colleagues of the room 215: René Grasselt, Henning Wilker, Daniel Simonis, Mohammed Reza Marami, Ralf Schmitz and Linda Smoydzin.

Finally, I am forever indebted to Steffi for her understanding, endless patience and encouragement when it was most required.