Model evaluation and data assimilation impact studies in the framework of COPS

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

Fakultät Naturwissenschaften
Universität Hohenheim

Institut für Physik und Meteorologie (IPM)

vorgelegt von
Thomas Schwitalla

aus Ostfildern-Ruit
2012

Dekan: Prof. Dr. rer. nat. Heinz Breer

1. berichtende Person: Prof. Dr. rer. nat. Volker Wulfmeyer

2. berichtende Person: Prof. Dr. rer. nat. Clemens Simmer

Eingereicht am: 17.11.2011

Mündliche Prüfung am: 16.04.2012

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

Abstract

The goal of this thesis was the study of new approaches for improving and investigating quantitative precipitation forecasting (QPF), e.g., by optimizing model resolution, physics combination, and data assimilation.

A forecasting system based on the Mesoscale Model 5 (MM5) was compared against other operational numerical weather prediction models from Météo France, MeteoSwiss and the German Weather Service primarily with respect to daytime precipitation. First, a notable daytime dry bias was observed. It appears to be the result of a too small high-resolution domain and the switched-off convection parameterization from the second to the innermost domain. Even the application of a 4-dimensional variational data assimilation (4DVAR) with GPS slant total delays (STD) does not solve this problem due to inconsistent model physics between the 4DVAR and the forecasting model. Nevertheless, the MM5 is in good agreement with the shape of the observed diurnal cycle after the spin-up phase.

As the development of the MM5 was suspended, a transition to the new Weather Research and Forecasting (WRF) model system was made after the D-PHASE period (end of 2007). This system features state-of-the-art physics packages and also a variational data assimilation system. As a new observing system, GPS Zenith Total Delay (ZTD) data from Central Europe were incorporated into the 3-dimensional variational data assimilation (3DVAR) system to further improve the initial water vapor field. A first study with this system revealed an improvement of the integrated water vapor RMSE of about 15% and a small but positive impact on the spatial and quantitative precipitation forecast. Additionally, the importance of assimilating upper air observations and the necessity to select a large, convection permitting model domain emerged.

Finally a rapid update cycle (RUC) approach, comparable to operational forecast centers, has been developed for a convection-permitting configuration of the WRF model. The system is capable to assimilate radar observations from Germany and France, GPS-ZTD data and satellite radiances and can be applied even for near real-time applications. First experiments with this system show promising results in comparison to other operational models.

Zusammenfassung

Das Ziel dieser Arbeit war die Untersuchung von neuen Ansätzen zur Verbesserung und Evaluierung der quantitativen Niederschlagsvorhersage z.B. durch anpassen der Modellauflösung, der Kombination von verschiedenen Parametrisierungen sowie der Datenassimilation.

Ein Vorhersagesystem auf Basis des mesoskaligen Atmosphärenmodells MM5 wurde mit anderen operationellen Vorhersagesystemen von Météo France, MeteoSchweiz und dem Deutschen Wetterdienst hinsichtlich des Tagesniederschlags verglichen. Zu Beginn wurde eine deutliche Unterschätzung des Niederschlags festgestellt. Diese ist das Resultat eines zu kleinen hoch aufgelösten Modellgebiets sowie des Abschaltens der Konvektionsparametrisierung im innersten Modellgebiet. Der Einsatz einer 4-dimensionalen, variationellen Datenassimilation mit GPS slant total delays (STD) brachte auch keine wesentlich Verbesserung der Resultate. Dennoch konnte eine gute Übereinstimmung mit der Gestalt des beobachteten Tagesgangs nach Ablauf der spin-up Phase erzielt werden.

Nachdem die Entwicklung des MM5 zwischenzeitlich eingestellt worden war, wurde am Ende der D-PHASE Periode (November 2007) auf das Weather Research and Forecasting (WRF) Modellsystem gewechselt. Dieses beinhaltet dem aktuellen Stand der Forschung entsprechende Modellphysik sowie ein variationelles Datenassimilationssystem. Als neuartige Beobachtungen wurden GPS zenith total delay (ZTD) Messungen in die 3DVAR mit eingebunden, um das Wasserdampfanfangsfeld weiter zu verbessern.

Eine erste Fallstudie ergab eine Verbesserung des RMSE des integrierten Wasserdampfgehalts um 15%, und einen schwachen, aber positiven Einfluß auf die räumliche und quantitative Niederschlagsvorhersage. Außerdem ergaben sich Hinweise, dass es wichtig ist, Höheninformationen zu assimilieren und ein möglichst großes, konvektionserlaubendes Modellgebiet zu wählen.

Im letzten Teil wurde ein rapid update cycle (RUC) Ansatz, vergleichbar mit anderen Vorhersagezentren, entwickelt. Zusätzlich ist das 3DVAR-System nun so vorbereitet, dass es gleichzeitig Radardaten aus Deutschland und Frankreich, GPS-ZTDs sowie Satellitendaten benutzen und in nahezu Echtzeit verwendet werden kann. Ein erstes Experiment mit diesen System ergab Erfolg versprechende Resultate im Vergleich zu anderen operationellen Modellen.

Contents

1 Motivation

Numerical weather prediction (NWP) models are the basis for weather forecasting by simulating the temporal and spatial evolution of the atmospheric state. They consist of several differential equations describing the evolution of wind, temperature, humidity and pressure (so-called "prognostic variables") which are in based on the continuity equation and Euler-Lagrange equation. The first attempt to solve this set of differential equations was performed by Richardson in 1922. Without the aid of computers, he and his staff needed 6 weeks for a 6 h forecast. The forecast failed with a pressure change of 145 hPa due to the lack of accurate initial conditions and deficits in numerics (Lynch, 2006). Because of this experience, NWP was considered impossible and further research activities concerning this topic were abandoned for decades but as computers became available, NWP models have been developed.

In principle, NWP models can be divided into two categories. The first category are global models with current horizontal resolutions between 50 km (e.g. Global Forecast System (GFS) of NOAA[1]) and 12–15 km (ECMWF[2]). In global models, the model top often reaches 0.01 hPa ($\approx$ 100 km) as they should also represent variations in the ozone concentration and are often coupled to aerosol models. Global models are mostly applied for medium range forecasting up to 10 days and climate projections. The main focus is on the correct representation of synoptic scale features of the global circulation. Examples are traveling high and low pressure systems (e.g. Hadley cell or Monsoon in the tropics).

The second category are limited area (mesoscale) models (LAM) for short-range predictions. In the past, these models typically featured a horizontal resolution of 10 km with a model top of about 50–20 hPa corresponding to about 20–25 km altitude. An example is the COSMO-EU[3] model operated at the German Meteorological Service (DWD). Increasing computing performance allows to further increase the horizontal resolution down to a few kilometers like in COSMO-DE (Baldauf et al., 2009), AROME (Bouttier, 2007), Mesoscale Model 5 (MM5; Grell et al., 1995) and the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008). In contrast to global models, limited area models require boundary conditions which are obtained from coarser resolution LAMs or global models. They are necessary to conserve the mass due to advection in the model.

In todays NWP models the horizontal resolution is still far too coarse to represent physical processes except from gridscale advection and diffusion. It is therefore necessary to parameterize these processes. This is achieved through several physics packages for cloud microphysics, the planetary boundary layer (PBL), radiation, land-surface interaction and convection. The required complexity of the schemes is different for global and regional applications. E.g on the global scale it is not necessary to have a highly complex cloud microphysics scheme containing prognostic variables for graupel and hail because the model is not capable to simulate strong updrafts required for the formation of graupel due to the comparatively coarse horizontal resolution, . In current global models like the ECMWF model, the applied microphysics scheme only distinguishes between water vapor, liquid water and ice.

[1] National Oceanic and Atmospheric Administration
[2] European Centre for Medium range Weather Forecast
[3] Consortium for Small-scale Modeling, www.cosmo-model.org

To support the model to simulate (summertime) convective precipitation, a convection parameterization (also known as cumulus scheme) is applied at resolutions down to ≈ 4–8 km (Kain et al., 2008).

Basically two types of cumulus schemes exist. 1) A mass-flux based scheme where the strength of convection is controlled by entrainment and detrainment in clouds (Tiedtke, 1989) and 2) an approach with a CAPE (Convective Available Potential Energy) closure, where CAPE is removed within a convective time period (Kain, 2004). The disadvantage of these schemes are that they contain only a simple single column cloud model compared to cloud microphysics schemes which allow a horizontal mass transport to neighboring grid cells. Due to the former and in combination with the usually medium horizontal resolution, the models show a windward-lee precipitation pattern and also other systematic errors described e.g in Schwitalla et al., 2008.

If the horizontal resolution increases to a few kilometers, the model is assumed to be able to simulate convection explicitly and thus the cumulus scheme is often switched off. More sophisticated cloud microphysics schemes are necessary to include more microphysical physical processes relevant at higher resolutions like the formation of graupel and hail. The most commonly applied type of scheme are one-moment schemes containing hydrometeor mass mixing ratios as prognostic variables like the Reisner2 scheme available in the MM5 (Reisner et al., 1998). In contrast to cumulus schemes, they update the temperature tendency with an additional term coming from the latent heat release during the phase transition of different hydrometeors.

In recent years, 2-moment schemes have been developed which additionally predict hydrometeor number concentrations like the schemes of Milbrandt and Yau (2005), Seifert and Beheng (2006) and Morrison et al. (2009). With these schemes, it is possible to more realistically describe the variability of the size distributions of different hydrometeors due to the fact that the total number concentration and mass mixing ratio are not monotonically related (Milbrandt and Yau, 2005). The prognostic calculation of number concentrations is a key for deriving the microphysical process rates (e.g. autoconversion, accretion) which becomes especially important in mid-latitude mixed-phase clouds. 2-moment schemes are also required for the coupling of aerosol models because they require prognostic number concentrations of hydrometeors. Compared to the one-moment schemes available e.g. in the WRF model, the computing time is not an issue due to improved parallelization of the code.

When getting to higher horizontal resolutions of a few kilometers, the PBL processes are becoming more and more important. Thus it is necessary to increase not only the horizontal but also the vertical resolution to avoid violating the Courant-Friedrichs-Lewy (CFL) condition. The maximum CFL number depends on the applied time integration scheme (e.g. 1.43 for a 3^{rd}-order Runge-Kutta scheme with 5^{th}-order advection) and gives an upper limit for the model time step in combination with a given horizontal and vertical resolution. Typically, LAMs on the mesoscale consists of 45 levels up to altitudes of 20 to 25 km from which often 15 or more layers reside below 3 km above ground level (AGL) and typically require a time step [s] of ~ 6 times the horizontal resolution [km].

To represent boundary layer properties in an NWP model, two different approaches are commonly applied. 1) The Turbulent Kinetic Energy (TKE) closure approach based on the work of Mellor and Yamada (1974) implying that the exchange coefficients for heat and momentum are derived from TKE and 2) the countergradient approach used by e.g. Hong et al. (2006) where the exchange coefficients are calculated from prescribed profile functions depending on

the Prandtl number complemented by a countergradient term to take non-local mixing (mixing with e.g the second level from the current one) into account.

As the surface is the lower boundary for the PBL schemes, Land-Surface-Models (LSM, e.g. Ek et al., 2003) are necessary. They provide information of the soil properties like temperature, snow cover, moisture and surface fluxes which are then fed back to the PBL scheme. Additionally they calculate the transport of energy and water in the soil. As the resolution of mesoscale models is continuously increasing, it is important to properly initialize the soil variables especially when performing climate projections. Although soil moisture and temperature are prognostic variables in most of todays NWP models, each new forecast cycle needs a new soil analysis which is often not included in the 3/4-dimensional variational data assimilation (3DVAR/4DVAR) system. Thus, e.g. ECMWF performs an off-line soil analysis based on 2 m temperature, humidity and precipitation or snow coverage (Douville et al., 2001) due to the lack of operational soil measurements. As this can lead to erroneous soil water content, further improvements for the initialization of LSM variables are necessary which are e.g. addressed in the Water and Earth System Science (WESS) project[4].

A still remaining problem in this context is to get fine resolution data for surface and soil textures. As an example, the topography data for the WRF model comes from the United State Geological Survey (USGS) and for COSMO-DE from the GLOBE data set of NOAA with a resolution of 30" (1 km). Landuse information for WRF is e.g. provided by the International Geosphere-Biosphere Programme (IGBP) MODIS data set. It is based on a satellite retrieval made in 2001 available at a 1 km resolution containing 20 different land use types.

For the representation of the soil, currently data from the United Nations Food and Agricultural Organization (FAO) is used which is only available on a 10' resolution (20 km). This can lead to erroneous results especially in climate simulations because the soil type determines the energy and water budget and thus the surface fluxes. Figure 1 shows the land use and soil data currently available in the WRF model.

In addition to the model physics and numerics, the forecast quality also depends on the quality of the initial conditions. NWP with LAMs is an initial and boundary problem and their forecast quality depends, besides the model physics, on the quality of the driving model. Depending on the desired resolution of the selected model, this driving model can either be a global model or also a LAM covering the desired area.

One of the largest drawbacks of using global models as initial field is their relatively coarse resolution of about 30–50 km (except ECMWF) which is only able to represent large scale synoptic patterns. To reduce inaccuracies at initialization, NWP centers like Météo France, the DWD or the UK MetOffice developed their own model chain containing a global model, a mesoscale model and a convection-permitting model with similar physics packages to overcome major inconsistencies due to the different physics and numerics schemes applied in the models.

In summer, quantitative precipitation forecasting (QPF) in low mountain regions is still a very important research area as NWP models still have deficiencies in correctly forecasting the spatial and temporal evolution of precipitation.

[4] www.wess.info

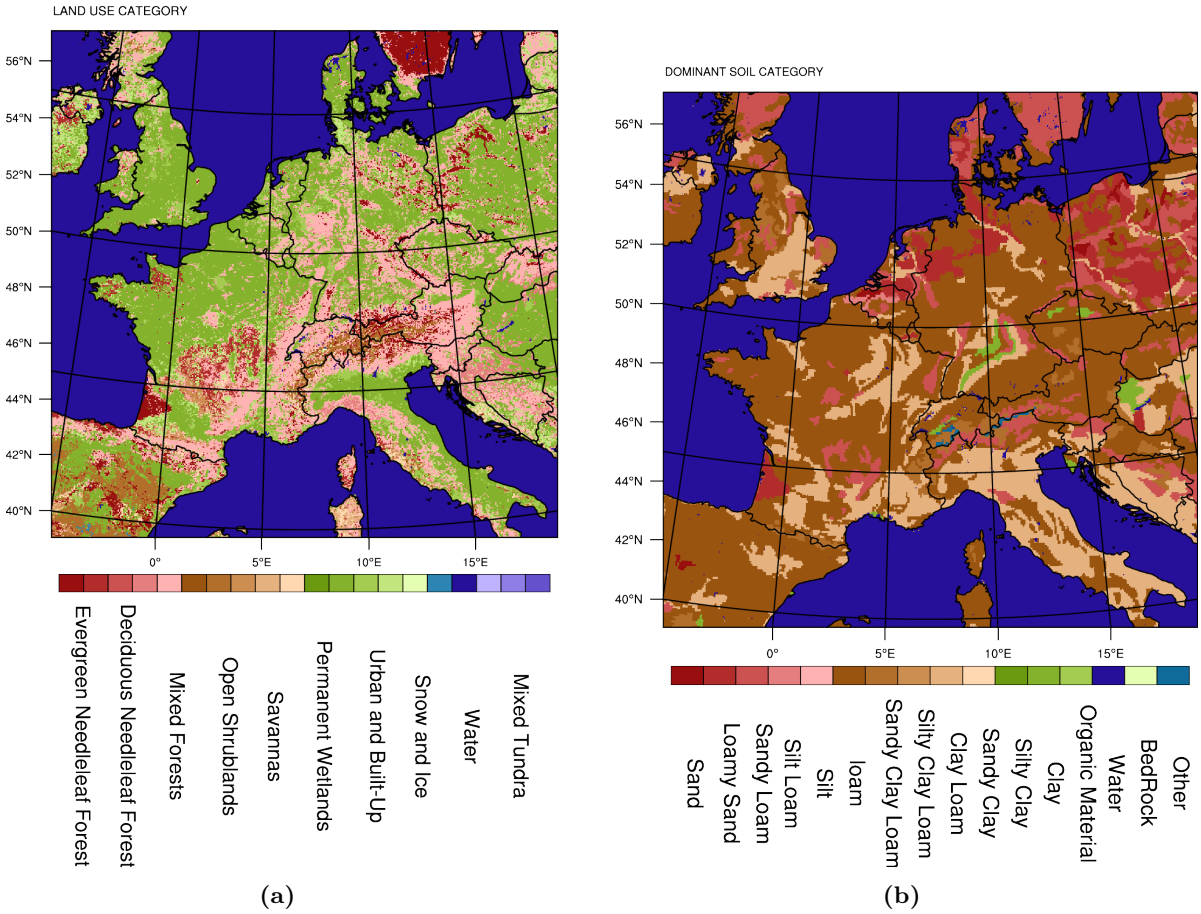


Figure 1: (a) shows the landuse data available from the MODIS satellite and (b) shows the FAO soil data set.

To improve the process understanding and quantitative precipitation forecast (QPF), the World Weather Research Program (WWRP) Research and Development Project (RDP) COPS[5] (Wulfmeyer et al., 2011) was carried out in summer 2007. The main purpose behind COPS was to understand the whole precipitation life cycle. This includes the investigation of processes leading to the development of convection, precipitation and the decay of convection in the Black Forest and the Vosges Mountains. Outcome of this campaign is a large dataset of observations including the exploitation of sensor synergies. For instance, on the Hornisgrinde mountain a Raman-LIDAR[6], a water vapor DIAL[7] , a precipitation radar, a cloud radar, radiosondes and occasionally a research aircraft were operated simultaneously during intensive observations periods (IOPs) where convective activity was expected. This allows the validation of different observing systems and the retrieval of additional parameters not measured by the different instruments directly.

COPS was coordinated with the WWRP Forecast Demonstration Project (FDP) D-PHASE[8]

[5] **C**onvective and **O**rographically-induced **P**recipitation **S**tudy
[6] **Li**ght **D**etection **A**nd **R**anging
[7] **Di**fferential **A**bsorption **LIDAR**
[8] **D**emonstration of **P**robabilistic **H**ydrological and **A**tmospheric **S**imulation of flood **E**vents in the Alpine

(Rotach et al., 2009) which took place from June to November 2007. During this time, 31 operational and research NWP models with different horizontal resolutions ranging from 60 km to 2 km were operated simultaneously. The aim was to validate these models over the same time period applying the rich data set collected during COPS.

The German part of COPS was partially funded by the the priority program 1167 "Quantitative Precipitation Forecast" which was established in 2004 by the German Research Foundation (DFG). Within this framework, the objectives of this thesis are:

- Select the most suitable physics combination in the atmospheric model MM5 for central Europe and especially the COPS region (Black Forest, Vosges Mountains and Swabian Jura)
- Evaluate the MM5 performance on 2 km resolution in comparison with other operational high-resolution models
- Perform data assimilation case studies at a convection-permitting resolution with GPS-ZTD data and conventional observations over Europe
- Developing a rapid update cycle (RUC) with GPS and radar data on a convection-permitting resolution

In the following chapter, an introduction into variational data assimilation will be given. Chapters 3 and 4 will give information about GPS and radar data and how they can be beneficial for data assimilation. Chapter 5 contains a summary of three useful publications followed by an outlook into possible future research activities.

region

2 Data assimilation

Due to the large number of model grid cells compared to the number of observations, NWP is an under-determined initial value problem. To overcome this handicap, a technique is required which gives a best estimate of the true atmosphere by taking the actual model field as background and observations into account: the so-called “data assimilation”.

The oldest data assimilation scheme in operation is the Newtonian Relaxation technique (Nudging) or Four-Dimensional Data Assimilation (FDDA) which is operationally applied in the COSMO models (Schraff, 1997) and is an option for MM5 and WRF (Stauffer and Seaman, 1994). Here, an extra relaxation term is added to the prognostic equations containing a value for the strength (how reliable is the observation) and a time weighting factor. The advantage of this technique is that the additional amount of computing time is negligible. A big drawback is that only observations of model prognostic variables can be assimilated.

Additional observations which do not measure the model's prognostic variables can, if at all, only be processed with additional error-prone assumptions. In COSMO-DE, radar latent heat nudging (LHN) is used because the nudging scheme does not have the possibility to assimilate radar reflectivities directly to support the development of convection. Another example is the assimilation of GPS-ZTD data in the COSMO model where an IWV content, calculated from the Zenith Wet Delay (ZWD), is utilized to adjust the vertical humidity profile.

As new indirect observation types arose in recent decades, new techniques had to be developed that also allow to assimilate non-prognostic variables like remote sensing observations. These are variational data assimilation techniques where a least square fit is performed by minimizing a cost function which contains the difference between the model state (background) and the observations. Currently two different deterministic variational approaches are applied both on the global scale and on the mesoscale: The 3DVAR and 4DVAR approach (Courtier et al., 1998; Barker et al., 2004; Huang et al., 2009) where the least square fit is performed on a deterministic analysis and the Ensemble Kalman filter (EnKF, Evensen, 1994, 2003) where an ensemble with O$\sim$ 30-100 members is applied to estimate the new analysis.

While both variational techniques require a forward operator to simulate the observation in the model world, the 4DVAR, in addition, requires the tangent linear and adjoint of the whole forecast model which, depending on the complexity of the model, can become very difficult or even impossible to derive. Concerning this matter, the Ensemble filter has the advantage that the adjoint is not necessary.

For both variational assimilation schemes, information about the background error and covariances between different model variables is needed which has to be derived from a climatological estimate or from a forecast ensemble. This estimation is crucial because the real model errors are unknown and it is assumed the model is free of a bias. As the number of model variables is O$\sim 10^7$, the covariance matrix has 10^{14} elements. As this matrix has to be inverted, techniques to reduce the matrix dimension have to be applied in addition.

Today's 3DVAR and 4DVAR systems make use of a so-called ”incremental approach“. Here, the full non-linear model forward operators are applied to calculate the initial value of the cost function (outer loop). Then, linear versions of forward and adjoint operators are applied to minimize the cost function (inner loop) saving computing time at the cost of accuracy and

allowing to apply optimized numerical algorithms to minimize the cost function. The inner loop is executed until the gradient becomes sufficiently small.

When performing a 4DVAR more often the inner loop is performed on a coarser resolution (e.g. at ECMWF) due to the enormous computing time consumption because of the application of especially the adjoint of the forecast model. Ongoing research of combining e.g. 3DVAR and EnKF (Hybrid approach) shows promising results but due to the large number of ensemble members it is also very time consuming. A future task may be the application of an ensemble of 3(4)DVAR schemes on a deterministic forecast. Further details about the mathematical aspects of variational data assimilation can be found below in section 2.1.

The most commonly assimilated variables are temperature, moisture, wind, and surface pressure (conventional observations). In recent years, more often remote sensing observations like satellite brightness temperatures (Hollinger, 1989; Mo and Liu, 2008) and Atmospheric Motion Vectors (AMV, EUMETSAT, 2009) are used. The assimilation of brightness temperatures measured from satellite sounders like IASI (Infrared Atmospheric Sounding Interferometer) or AMSU (Advanced Microwave Sounding Unit) requires a radiative transfer model and is currently performed only in a cloud free environment. The available radiative transfer models are RTTOV (Radiative Transfer for TOVS; Saunders et al., 1999) and the CRTM (Community Radiative Transfer Model; Chen et al., 2008) which are applied in todays NWP models. The former is applied e.g. by ECMWF or DWD either in a 4DVAR or for the generation of synthetic satellite images respectively.

The retrieval algorithm for satellite winds for MSG (Meteosat Second Generation; e.g. Singh et al., 2011) is based on a combination of brightness temperatures of the water vapor and infrared channels combined with the US standard atmosphere or the ECMWF model and has been developed by EUMETSAT. It is currently only reliable above 700 hPa because the assumption that clouds move with the main wind breaks down over orography (EUMETSAT, personal communication).

A still remaining problem is the distribution of the different types of observations. Synoptic stations (SYNOP) together with airport reports (METAR), measuring hourly 2 m temperature, 2 m humidity, 10 m wind and surface pressure, are commonly available over land. Depending on the time of the day, also large amounts of aircraft measurements (AMDAR, AIREP) are available but they mostly measure only temperature and wind.

A new TAMDAR (Tropospheric Airborne Meteorological Data Reporting) network, currently only available over the US, additionally measures humidity together with wind and temperature. These three variables are important for representing the convective environment and thus for the prediction of the precipitation development. Radiosonde ascents are, apart from some research vessels, only available over land and usually only available at 0000 UTC and 1200 UTC. In spite of their sparsity, they are still the major source for information about the 3-dimensional humidity field due to their accuracy.

To give some numbers, within an assimilation time window of two hours, $\sim$ 3000 aircraft measurements, $\sim$ 1000 surface measurements, 10 wind profiler, $\sim$ 80 ship measurements and $\sim$ 50 radiosonde ascents are available from the ECMWF data archive (MARS) over central Europe.

To further improve the 3-dimensional humidity field, Global Positioning System (GPS) derived water vapor information are assimilated at e.g. Météo France to complement other observations.

Here, the GPS signal delay between the line of sight through the neutral atmosphere and the real ray path through the ionosphere and earth atmosphere is considered. After performing extensive ionospheric corrections through the application of dual-frequency receivers and applying a complex retrieval algorithm, the atmospheric refractivity N is obtained (Gendt et al., 2004). Integration of N along the ray path through the atmosphere and subsequent multiplication with a constant factor of 10^{-6} gives the slant total delay (STD) and zenith total delay (ZTD). In this study, GPS data is provided by the Helmholtz Centre Potsdam German Research Centre for Geosciences (GFZ).

Although both STD and ZTD are integrated quantities, they can be used to adjust the vertical (and horizontal in case of STD) distribution of water vapor in a variational assimilation scheme. A critical point is the estimation of the receiver altitude. It has to be very accurate because a few meters difference in the altitude assignment of the receiver can cause a difference of 1 cm in the value of the wet delay which can become critical during hot and dry weather conditions.

In recent years, the most commonly used parameter derived from GPS data for assimilation purposes is the Integrated Water Vapor (IWV) content. In a variational scheme, the IWV content is converted to water vapor and air density, allowing to adjust the water vapor distribution. The disadvantage is that the calculation of GPS-observed IWV requires the assumption of a mean temperature and moisture measurement at the GPS receiver location. Therefore, the NWP community switches more and more to the assimilation of ZTD measurements (Vedel and Huang, 2004) which are only dependent on surface temperature and pressure. When assimilating ZTD data in a variational scheme, it is transferred to pressure, temperature and humidity via the adjoint operator and thus has a higher information content compared to IWV (see section 3.1).

The next step in complexity and realism is the application of GPS STDs (e.g. Zus et al., 2008; Bauer et al., 2011b) for the assimilation where the real ray path between satellite and receiver is considered. This allows for fully exploiting the water vapor information provided by GPS. This means that due to the large spatial coverage information about the vertical water vapor distribution especially in the boundary layer can be obtained which can have a significant impact on summertime convection. Further examples for the application of GPS data beside data assimilation is the 3-dimensional reconstruction of the water vapor field (tomography) applying STD data (Bender et al., 2009) and the validation of other remote sensing instruments and models (Bender et al., 2008; Bauer et al., 2011a).

As the assimilation is an under-determined problem, it is desired to have as many as possible observations on the one hand side. On the other side, the quality of the observations is a critical point during the assimilation. In general, a variational data assimilation approach minimizes a cost function which roughly contains differences between the model background and the observations (O-B) weighted by the measurement quality. If the distance between the model and the observations is large, the cost function has a high value and thus the minimization takes more and more iterations to converge. It is possible, that the model is forced to a new analysis resulting in a large model imbalance as e.g. hydrometeors are not updated during a 3DVAR except when applying radar reflectivities.

As shown later in section 2.1, the observations are weighted by $1/\sigma^2$ (with σ being the observation error) during the assimilation process. This implies that observations with a large error are degraded and thus the influence in the assimilation process and the new analysis becomes smaller. However, users wish to have low observation errors but this can reduce the number of

observations significantly due to the O-B check. This is usually performed in the outer loop of an incremental variational assimilation (section 2.4) where the full non-linear observation operator is applied to simulate the observation. This rejection can e.g happen for GPS measurements due to an inaccurate water vapor background field or for radar reflectivities in the case that the model did not trigger convection.

The ECMWF and the WRF assimilation systems e.g. apply statistically derived errors for standard variables like wind, temperature and humidity. For remote sensing observations like satellite radiances or radar radial velocities, the user has to apply a bias correction (for satellite data) and also a thinning by averaging values in a model grid box due to the high resolution (more than one observation can exist in one model grid box) and correlations between neighboring pixels in the measurements. E.g in the French AROME Model, data are thinned to a 15*15 km grid box (Montmerle and Faccani, 2009) and satellite brightness temperatures are usually thinned to a distance of 50–100 km.

It can also occur that the minimization algorithm, often done with the conjugate gradient method, is only able to find a relative minimum of the cost function instead of a global minimum degrading the quality of the new analysis. Therefore the observations have to be screened in a way that the difference between the model and the observations is not too large (so-called O-B check). However this can be dangerous if the background field has a bad quality, because one may reject observations which have a good quality but are too far away from the analysis. This happened e.g. at DWD for the winter storm Lothar in December 1999 (Wergen and Buchhold, 2002). Here a sounding over Newfoundland was rejected in the data assimilation of the global model GME (Majewski et al., 2009). This happened due to unreasonable (but really observed) low pressure values as the balloon of the scheduled radiosonde burst early and a second radiosonde had to be launched later which was assigned at the original time in the former data assimilation scheme.

The following section will give further details about the mathematical background applied in variational data assimilation schemes with a focus on the WRF model.

2.1 Mathematical background of variational data assimilation

Data assimilation is a suitable technique to improve the model initial state. Variational data assimilation schemes, independent on whether it is an deterministic or ensemble-base method, are based on the maximum likelihood and least-square method and the Bayes-Theorem (Lorenc, 1986; Bouttier and Courtier, 1999). The model state $\mathbf{x}_b$ with the dimension of the model domain in West-East direction i, South-North direction j, z-direction k and the number of prognostic variables v can be described by a vector

$$\mathbf{x}_b = \begin{pmatrix} x_1 \\ x_2 \\ .... \\ .... \\ x_n \end{pmatrix} \tag{2.1}$$

with the dimension of n= i·j·k·v in the range of 10^7.

The Bayes's Theorem describes the joint probability P for the occurrence of two events $\mathbf{x}_b$ (model background) and $\mathbf{y}$ (observations):

$$P(\mathbf{y}) \cdot P(\mathbf{x}_b \mid \mathbf{y}) = P(\mathbf{y} \mid \mathbf{x}_b) \cdot P(\mathbf{x}_b) \tag{2.2}$$

The first term on the l.h.s. denotes the a priori pdf of an observation $\mathbf{y}$, the second term denotes the conditional probability to find $\mathbf{x}_b$ for given observation $\mathbf{y}$. The first term on the r.h.s. denotes the probability for the observation with a given background $\mathbf{x}_b$ and the second term denotes the a priori pdf for a model background $\mathbf{x}_b$. As the probability P($\mathbf{y}$) for a measurement is equal to one, the probability for finding $\mathbf{x}$ with the observation $\mathbf{y}$ is

$$P(\mathbf{x} \mid \mathbf{y}) = P(\mathbf{y} \mid \mathbf{x}) \cdot P(\mathbf{x}) \tag{2.3}$$

which is the basis for all variational data assimilation schemes. Under the assumption of Gaussian distribution, the probability for true observations $\mathbf{y}_t$ when observations $\mathbf{y}$ are given is

$$P(\mathbf{y_t} \mid \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\mathbf{O}}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{y}_t)^T\mathbf{O}^{-1}(\mathbf{y}-\mathbf{y}_t)} \tag{2.4}$$

$\mathbf{O}$ denotes the observation error covariance matrix describing error correlations between different observation types and variables and n is the dimension of the observation vector $\mathbf{y}$. In case of no bias it is defined as

$$\begin{aligned} \mathbf{O} &= < (\mathbf{y}-\mathbf{y}_t)(\mathbf{y}-\mathbf{y}_t)^T > \\ &= < \epsilon\epsilon^T > \end{aligned} \tag{2.5}$$

where <> means the expectation value.

The same mathematical formalism can be applied for the probability of a true model state $\mathbf{x}_t$ when a given model background $\mathbf{x}_b$ is present:

$$P(\mathbf{x_t} \mid \mathbf{x}_b) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\mathbf{B}}} e^{-\frac{1}{2}(\mathbf{x}_b-\mathbf{x}_t)^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}_t)} \tag{2.6}$$

Similar to equation (2.4), $\mathbf{B}$ denotes the model background error covariance matrix. In general it describes correlations of different model variables on the model grid, e.g. the correlation between water vapor and horizontal wind. In case only one observation of e.g. water vapor is available, $\mathbf{B}$ spatially distributes the new information to other variables if correlations exist.

To find the most likely values for the new model analysis state $\mathbf{x}_a$ for $\mathbf{x}_t$ and $\mathbf{y}_a$ for $\mathbf{y}_t$, when both observation and model background are given, the probability P has to be maximized:

$$P(\mathbf{x}_a \mid \mathbf{x}_b)P(\mathbf{y}_a \mid \mathbf{y}) = P(\mathbf{x}_a, \mathbf{y}_a) \stackrel{!}{=} max \tag{2.7}$$

$$P(\mathbf{x}_a \mid \mathbf{x}_b)P(\mathbf{y}_a \mid \mathbf{y}) \propto e^{\frac{1}{2}(\mathbf{x}_b-\mathbf{x}_a)^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}_a)} \cdot e^{\frac{1}{2}(\mathbf{y}-\mathbf{y}_a)^T\mathbf{O}^{-1}(\mathbf{y}-\mathbf{y}_a)} \tag{2.8}$$

To obtain the maximum value for P($\mathbf{x}_a$,$\mathbf{y}_a$), it is equivalent to minimize -ln (P($\mathbf{x}_a$,$\mathbf{y}_a$)) which is referred to as cost function J:

$$J(\mathbf{x}_a, \mathbf{y}_a) = \frac{1}{2}((\mathbf{x}_b-\mathbf{x}_a)^T\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}_a) + (\mathbf{y}-\mathbf{y}_a)^T\mathbf{O}^{-1}(\mathbf{y}-\mathbf{y}_a)) \stackrel{!}{=} min \tag{2.9}$$

As $\mathbf{y}_a$ is not directly available from the model, a (not necessarily linear) "interpolation" operator (forward operator) $\overrightarrow{H}$ has to be developed for each type of observations which generates the observation from the model state $\mathbf{x}_a$:

$$\mathbf{y}_a = \overrightarrow{H}(\mathbf{x}_a) \tag{2.10}$$

Due to the application of $\overrightarrow{H}$, the observation error covariance matrix $\mathbf{O}$ has to be replaced by the matrix $\mathbf{R}$

$$\mathbf{R} =< (\mathbf{y} - \overrightarrow{H}(\mathbf{x}_a))(\mathbf{y} - \overrightarrow{H}(\mathbf{x}_a))^T > \tag{2.11}$$

which includes the covariances (observation and representativeness errors) between $\mathbf{y}$ and $\overrightarrow{H}(\mathbf{x}_a)$. Therefore equation 2.9 for a single observing system reads now:

$$J(\mathbf{x}_a) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x}_a)^T\mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_a) + \frac{1}{2}[(\mathbf{y} - \overrightarrow{H}(\mathbf{x}_a))^T\mathbf{R}^{-1}(\mathbf{y} - \overrightarrow{H}(\mathbf{x}_a))] \stackrel{!}{=} min \tag{2.12}$$

In most of the current variational data assimilation schemes, $\mathbf{R}$ is a block-diagonal matrix assuming no correlation of errors between the single observations and thus most off-diagonal elements are zero. This assumption can be violated especially for satellite radiances or radar radial velocities. Due to their high spatial resolution, neighboring pixels are correlated with each other.

To obtain the minimum of $\mathrm{J}(\mathbf{x}_a)$, the gradient has to be zero with respect to $\mathbf{x}_a$. This requires the solution of

$$\nabla J(\mathbf{x}_a) = -\mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_a) - \mathbf{H}^*\mathbf{R}^{-1}(\mathbf{y} - \overrightarrow{H}(\mathbf{x}_a)) = 0. \tag{2.13}$$

$\mathbf{H}^*$ is the adjoint of the forward operator $\overrightarrow{H}$. Assuming $\mathbf{x} = \mathbf{x}_b + \delta\mathbf{x}$ with an infinitesimal perturbation $\delta\mathbf{x}$, the forward operator $\overrightarrow{H}(\mathbf{x})$ can be developed into a Taylor expansion

$$\overrightarrow{H}(\mathbf{x}_b + \delta x) = \overrightarrow{H}(\mathbf{x}_b) + \frac{\partial\overrightarrow{H}}{\partial\mathbf{x}}\delta x + \frac{1}{2}\frac{\partial^2\overrightarrow{H}}{\partial\mathbf{x}^2}\delta x^2 + O[(\delta x)^3]. \tag{2.14}$$

In most of today's data assimilation schemes, $\overrightarrow{H}(\mathbf{x})$ is assumed to be linear. Thus higher order terms except the linear in the Taylor expansion are neglected. The first derivative $\frac{\partial\overrightarrow{H}}{\partial\mathbf{x}}\delta x$ is called the Tangent Linear Model (TLM) in data assimilation and is the Jacobi Matrix $\mathbf{H}$ of $\overrightarrow{H}(\mathbf{x})$:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial\overrightarrow{H}_1}{\partial\mathbf{x}_1} & .... & \frac{\partial\overrightarrow{H}_1}{\partial\mathbf{x}_n} \\ .... & .... & .... \\ \frac{\partial\overrightarrow{H}_m}{\partial\mathbf{x}_1} & .... & \frac{\partial\overrightarrow{H}_m}{\partial\mathbf{x}_n} \end{pmatrix} \tag{2.15}$$

m denotes the model space dimension, more precisely the number of model grid points and n the number grid boxes times the number of model state variables which are pressure p, temperature T, specific humidity q, zonal and meridional wind u and v.

The adjoint $\mathbf{H}^*$ of the forward operator $\overrightarrow{H}$ is the transpose of $\mathbf{H}$ and thus reads

$$\mathbf{H}^* = \mathbf{H}^T = \begin{pmatrix} \frac{\partial\overrightarrow{H}_1}{\partial\mathbf{x}_1} & .... & \frac{\partial\overrightarrow{H}_m}{\partial\mathbf{x}_1} \\ .... & .... & .... \\ \frac{\partial\overrightarrow{H}_1}{\partial\mathbf{x}_n} & .... & \frac{\partial\overrightarrow{H}_m}{\partial\mathbf{x}_n} \end{pmatrix} \tag{2.16}$$

An example for the derivation of $\mathbf{H}$ and $\mathbf{H}^T$ for GPS-ZTD data is given later in section 3.1.

2.2 FGAT/4DVAR

The details given in the previous section apply for the 3DVAR, where the observations within a user specified assimilation window around the analysis time are assumed to be exactly at the analysis time. This is, at the same time, its biggest advantage and the biggest disadvantage. A disadvantage because the observations are packed together independent of the observation time and no model dynamics is included in the new analysis. An advantage as the observations are fixed to the analysis time and thus no NWP model is necessary to propagate the model forward and backward in time which is significantly less computationally demanding.

To partially resolve this problem, the 3DVAR-FGAT (First Guess at Appropriate Time) can be applied. Here the observations are arranged into several time slices around the desired analysis time. E.g. if the analysis time is 00 UTC, then observations would be arranged into three parts with observations ±30 minutes around the analysis time, 23 UTC and 01 UTC the next day. Then for each date a separate innovation vector and cost function are calculated. Finally the cost functions from each time slots are added and a 3DVAR is performed.

To include model dynamics and assimilate observations within the assimilation window at their measurement times, the 4DVAR approach is applied. In contrast to the 3DVAR, the 4DVAR does not only need an adjoint of the observation operators but also requires an adjoint of the NWP model M.

Applying the model to a model state $\mathbf{x}_{i-k}$ with k being the assimilation time step in the assimilation window gives

$$\mathbf{x}_{a_i} = M(\mathbf{x}_{a_{i-k}}) \tag{2.17}$$

with $\mathbf{x}_{a_i}$ being the model state at time i. This leads to a redefined cost function

$$J(\mathbf{x}_a) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x}_a)^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_a) + \sum_{i=1}^{n}(\mathbf{y}_i - \overrightarrow{H}_i(\mathbf{x}_{a_i}))^T \mathbf{R}^{-1}(\mathbf{y}_i - \overrightarrow{H}_i(\mathbf{x}_{a_i})) \tag{2.18}$$

(2.18) is only valid if the model operator M can be linearized. As this is usually not the case, simplified physics have to be applied during the 4DVAR. E.g in the latest WRF 4DVAR release, a simple water vapor condensation scheme, a simple convection scheme (both do not consider frozen hydrometeors) and a surface drag scheme are available.

2.3 Properties and derivation of the background error covariance matrix B

The background error covariance matrix **B** contains the spatial information about background (model) errors. To derive **B**, two different approaches are used: 1) The National Meteorological Center (NMC) method developed by Parrish and Derber (1992), where **B** is estimated from forecast differences and 2) the ensemble method described by Fisher (2003) where **B** is calculated from an ensemble using ensemble member perturbations. As it is not easy to find a suitable ensemble, the current mostly common variant is the NMC method which is briefly addressed here.

The basic concept of the NMC method is the calculation of **B** from forecast differences over a longer time period. Usually at least a one month time period or a season is chosen. For global

models like ECMWF the forecast differences $\mathbf{x}'$ between the 48 h and 24 h forecast of different initial dates are calculated for a target date. For regional models like WRF, usually differences between a 24 h and a 12 h forecast are chosen:

$$\begin{aligned}\mathbf{B} &= \overline{\mathbf{x}'\mathbf{x}'^{T}} \\ &= \overline{(\mathbf{x}_{t+24h} - \mathbf{x}_{t+12h})\,(\mathbf{x}_{t+24h} - \mathbf{x}_{t+12h})^{T}}\end{aligned} \tag{2.19}$$

which is similar to the definition of the observation error covariance matrix $\mathbf{O}$ shown in equation 2.5. As the reader may have noticed, this is not a real background error as the name promises. It is a climatological estimate of the model variances. The reason not to chose differences between the model and observations at certain time is the lack of observations in this case.

Each model grid point requires an observation which is not possible and thus would also introduce interpolation errors which are not useful for model error characteristics. The reason not to choose forecast differences between a 6 h forecast and the analysis are twofold. First this would include the model spin-up when starting from an analysis and secondly, the error of the analysis would be taken into account.

Another problem is the representation of the model error covariances in $\mathbf{B}$. If assuming 550x550x50 model grid boxes, they contain about $\sim 10^7$ prognostic variables. The representation of errors in the $\mathbf{B}$-matrix would lead to $N_e = 0.5 \cdot (10^7)^2 = 0.5 \cdot 10^{14}$ elements in $\mathbf{B}$. As the data has to be stored with double precision, this would lead to a file size of ∼364 Petabyte, which is impossible to store even on a high-performance computer as e.g the German Climate Computing Center which has about 20 Petabyte memory. Therefore, a technique is required, which reduces the size of $\mathbf{B}$ without loosing too much information. This can be done with a transformation of $\mathbf{B}$ into an almost diagonal matrix.

Starting point for this transformation are equations 2.14 and 2.22. First, a control variable transformation $\delta\mathbf{x} = \mathbf{U}\mathbf{v}$ is introduced where $\mathbf{U}$ has to fulfill the relation $\mathbf{B} = \mathbf{U}\mathbf{U}^T$ and $\mathbf{v}$ is a control variable. This transformation ensures that covariances between the new control variables $\mathbf{v}$ are minimized. For the WRF model, the control variable $\mathbf{v}$ consists of five variables to meet the 5 prognostic variables:

- Stream function Ψ, derived from vorticity by solving the Poisson equation
- Velocity potential derived from divergence by solving the Poisson equation
- Temperature
- Pseudo relative humidity
- Surface pressure

Pseudo relative humidity is defined as ratio between the current water vapor mixing ratio and the saturation water vapor mixing ratio of the background field.

Furthermore, the transformation matrix $\mathbf{U}$ is separated into 3 matrices so that $\mathbf{U} = \mathbf{U}_h\mathbf{U}_v\mathbf{U}_p$. $\mathbf{U}_h$ is the horizontal correlation matrix which holds horizontal correlations and $\mathbf{U}_v$ denotes the vertical correlation matrix. $\mathbf{U}_p$ finally contains the the transformation of the control variables

back to model prognostic variables. This decomposition reduces the elements from N_e to $\sim \sqrt{N_e}$ with very small off-diagonal terms. Further details about the transformation are given in Lorenc et al. (2000), Barker et al. (2003) and Barker et al. (2004).

As an example for the properties of **B**, Figures 2 and 3 show how the information of a single temperature increment of 10 K on model level 10 in the middle of the domain is spread out horizontally and vertically by the background matrix **B**.

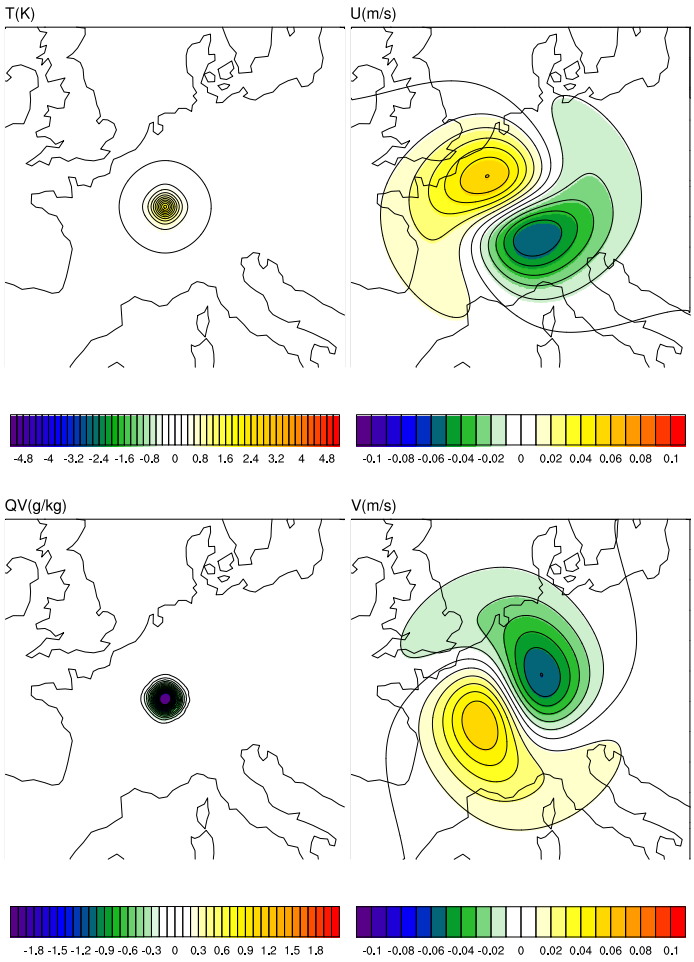


Figure 2: Single observation test with a temperature increment of 10 K at model level 10. (a) Shown are analysis increments for temperature T (upper left panel), U-wind component (upper right panel), water vapor mixing ratio Q (lower left panel) and V-wind component (lower right panel).

Due to horizontal and vertical correlations between the variable itself and other prognostic variables, not only T is influenced but also U and V. This indicates that **B** is not a fully diagonal matrix. In a newer version of the WRF assimilation system, **B** is designed such that temperature or wind observations also influence the water vapor mixing ratio and vice versa.

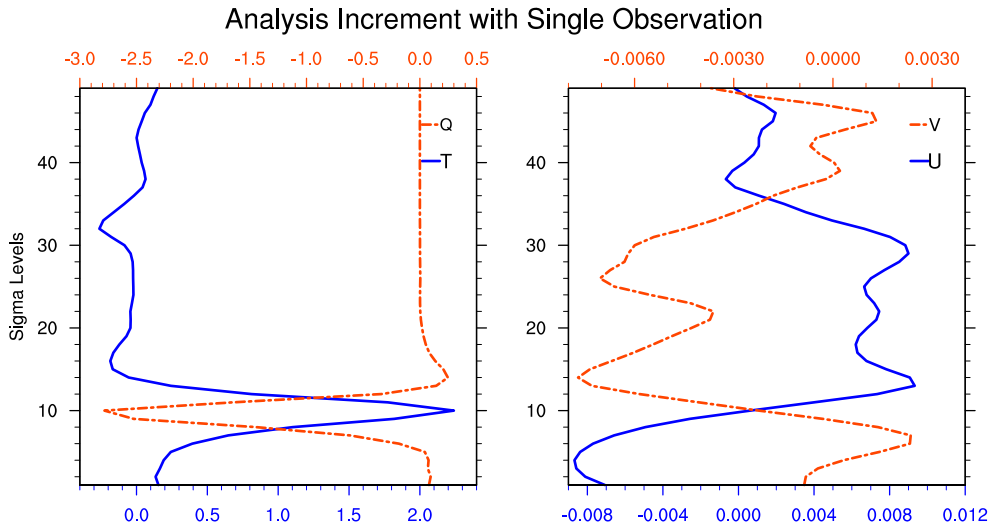


Figure 3: Vertical distribution of T, u, v and Q increments at the increment position.

2.4 Incremental 3/4DVAR

As an example, a model configuration of grid 550x550x50 grid boxes contains about $\sim 10^7$ prognostic variables which are described by the vector $\mathbf{x}_b$ in equations (2.12) and (2.18). This implies that the minimization algorithm has to handle 10^7 derivatives in each iteration which is quite time-consuming. Therefore, in todays 3DVAR and 4DVAR algorithms, an incremental approach is applied. At the beginning, the full non-linear observation forward operator $\overrightarrow{H}(\mathbf{x}_b)$ and, in case of a 4DVAR, also the full non-linear forward model $\overrightarrow{M}(\mathbf{x}_{b_{i-k}})$is applied to calculate the initial cost function (outer loop).

With the definition of the linearized observation operator

$$\overrightarrow{H}(\mathbf{x}) = \overrightarrow{H}(\mathbf{x}_b + \delta\mathbf{x}) = \overrightarrow{H}(\mathbf{x}_b) + \mathbf{H}(\delta\mathbf{x}) \tag{2.20}$$

and the original departures (observation minus background)

$$\mathbf{d}_i = \mathbf{y} - \overrightarrow{H}(\mathbf{x}_b) \tag{2.21}$$

a new "reduced" cost function

$$\widetilde{J}(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x}\mathbf{B}^{-1}\delta\mathbf{x}^T) + \frac{1}{2}[(\mathbf{d}_i - \mathbf{H}(\delta\mathbf{x}))^T\mathbf{R}^{-1}((\mathbf{d}_i - \mathbf{H}(\delta\mathbf{x}))] \tag{2.22}$$

can be defined which is solved by calculating the gradient with respect to $\delta\mathbf{x}$. The advantage of this technique is, that the degrees of freedom is significantly reduced. For example when

4000 observations are available, the number of elements of $\delta\mathbf{x}$ is $\sim 10^4$ allowing a much faster computation of the gradient compared to the original formulation 2.12. Further details can be found in Courtier (1997), Barker et al. (2003) and Huang et al. (2009).

2.5 Minimization Algorithm

The general purpose of a variational data assimilation scheme is the minimization of the cost function $\widetilde{J}$ (2.22) describing the sum of differences between model background and modeled observations. To minimize $\widetilde{J}$, it is required to set $\nabla\widetilde{J} = 0$. As the dimension can be in the range of $n^2 \sim 10^{14}$ it is very inefficient or even impossible to solve equation 2.13 directly. Therefore, in todays assimilation schemes two different minimization algorithms are applied: 1) the Lanczos algorithm described by Golub and Van Loan (1996) and the conjugate gradient (CG) method (Shewchuk, 1994) which is explained in more detail using the example implemented into the WRF model.

The conjugate gradient method is an iterative method which can be applied to systems of linear equations as shown in equation 2.22. A precondition for this method is that the matrix $\mathbf{R}$ has to be a symmetric matrix and positive definite. The latter relation is fulfilled, if

$$[\mathbf{d}_i - \mathbf{H}(\delta\mathbf{x})]^T\mathbf{R}^{-1}[\mathbf{d}_i - \mathbf{H}(\delta\mathbf{x})] \quad > \quad 0 \tag{2.23}$$

with $\mathbf{d}_i$ defined in equation 2.21.

In a first guess, a residual vector

$$\mathbf{g}_0 = \mathbf{y} - \overrightarrow{H}(\mathbf{x}_b) - \overrightarrow{H}(\delta\mathbf{x}_0) \tag{2.24}$$

describing the negative gradient of a quadratic function $f(\delta\mathbf{x}_0)$ and a conjugate vector

$$\mathbf{p}_0 = \mathbf{g}_0 \tag{2.25}$$

as first guess are calculated. The second term on the right hand side of equation 2.24 is zero as no minimization took place so far. In an incremental approach, this is done in the outer loop where the full observation operator is applied to calculate the observed quantity. From this starting point, the inner iteration process starts.

In the first inner iteration k (starting at k=0), a scalar

$$\begin{aligned} \alpha_k &= \frac{\mid \mathbf{g}_k \mid^2}{\mathbf{p}_k^T(\nabla^2\widetilde{J})\mathbf{p}_k} \\ &= \frac{\mid \mathbf{g}_k \mid^2}{\mathbf{p}_k^T(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{p}_k} \end{aligned} \tag{2.26}$$

is calculated leading to an updated guess

$$\delta\mathbf{x}_{k+1} = \delta\mathbf{x}_k + \alpha_k \cdot \mathbf{p}_k \tag{2.27}$$

and gradient (residual)

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \cdot (\nabla^2\widetilde{J}) \cdot \mathbf{p}_k \tag{2.28}$$

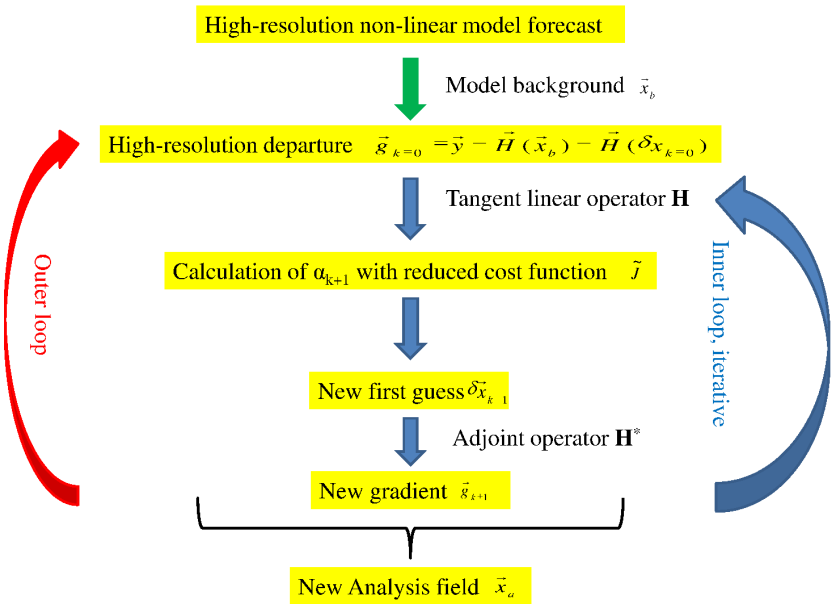


Figure 4: Schematic flow diagram of an incremental variational assimilation scheme.

where $\mathbf{H}$ is the linearized observation operator and $\mathbf{H}^T$ is the adjoint observation operator. Still in the inner loop k, a scalar

$$\beta = \frac{|\ \mathbf{g}_{k+1}\ |^2}{|\ \mathbf{g}_k\ |^2} \tag{2.29}$$

is introduced which gives a new estimate for the conjugate vector

$$\mathbf{p}_{k+1} = \mathbf{g}_{k+1} + \beta \cdot \mathbf{p}_k. \tag{2.30}$$

with

$$\delta\mathbf{x}_{k+1}^T(\nabla^2\widetilde{J})\mathbf{p}_{k+1} = 0 \tag{2.31}$$

This procedure is executed in a loop until $\mathbf{g}_{k+1}$ is sufficiently small to end the minimization process here. A peculiarity in the WRF model is the fact, that the minimization procedure is performed in the control variable space described in section 2.3.

Figure 4 shows a schematic illustration of the minimization procedure. Depending on the quality of the observations and the model background, it can happen that this method takes either a lot of inner iterations (~ 100), that the gradient bounces at higher iterations or that no minimum is achieved. Figure 5 shows an example for the behavior of the cost function $\widetilde{J}$ and its gradient during a minimization process.

2.6 Brief outlook to ensemble data assimilation

For an incremental 3DVAR algorithm, the minimum of equation 2.22 with respect to $\delta\mathbf{x}$ similar to equation 2.13 is calculated. It reads as follows:

$$\nabla\widetilde{J} = \mathbf{B}^{-1}\delta\mathbf{x} - \mathbf{H}^*\mathbf{R}^{-1}(\mathbf{d}_i - \mathbf{H}(\delta\mathbf{x})) = 0. \tag{2.32}$$

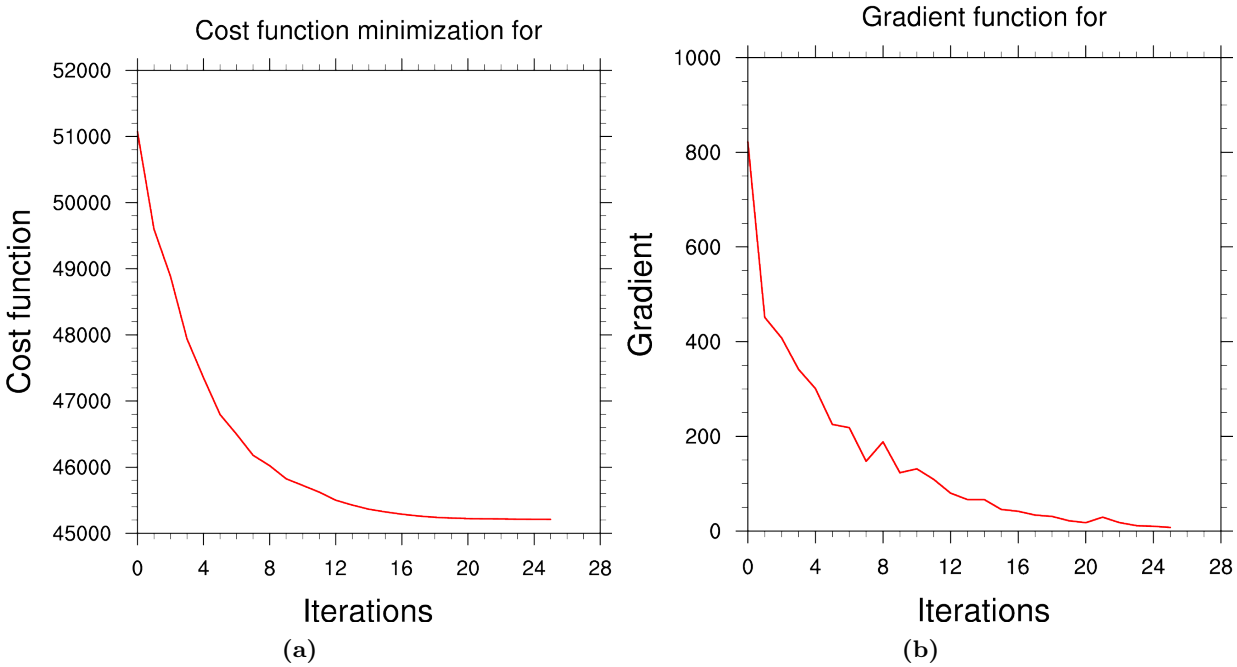


Figure 5: Example for the behavior of the cost function $\widetilde{J}$ (a) and its gradient (b) during the inner loop.

This equation can be rewritten as

$$\delta\mathbf{x} = (\mathbf{B}^{-1} + \mathbf{H}^*\mathbf{R}^{-1}\mathbf{H})^{-1}[\mathbf{H}^*\mathbf{R}^{-1}\mathbf{d}_i]. \tag{2.33}$$

Applying the Woodbury matrix identity (Hager, 1989), this leads to

$$\delta\mathbf{x} = \underbrace{\mathbf{B}\mathbf{H}^*(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^*)^{-1}}_{\mathbf{K}}\mathbf{d}_i. \tag{2.34}$$

which is then added to $\mathbf{x}_b$ to obtain the new analysis. The matrix product $\mathbf{B}\mathbf{H}^*(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^*)^{-1}$ is the "Kalman gain matrix" $\mathbf{K}$ for a deterministic forecast.

In the previous section, the creation of the static matrix $\mathbf{B}$ describing model errors and covariances of the different model prognostic variables was performed by the NMC method of Parrish and Derber (1992) describing errors and covariances by a model climatology.

Recently, operational forecast centers like Météo France tend to derive $\mathbf{B}$ from an ensemble of forecasts from the selected NWP model (Berre, 2000; Fisher, 2003) applying a Kalman Filter (Evensen, 2003). Here the errors and covariances are calculated from ensemble perturbations with respect to the ensemble mean. Thus, the matrix $\mathbf{B}$, now renamed as $\mathbf{P}_e$, is the ensemble derived error covariance matrix and can be rewritten as

$$\mathbf{P}_e = \frac{1}{K-1}\sum_{i=1}^{K}(\mathbf{x}_f^i - \mathbf{x}_e)(\mathbf{x}_f^i - \mathbf{x}_e)^T = \mathbf{x}_p\mathbf{x}_p^T \tag{2.35}$$

with $\mathbf{x}_e$ denoting the ensemble mean of the K members and $\mathbf{x}_f^i$ being the forecast of the i-th ensemble member. In a matrix notation this can be rewritten as

$$\mathbf{P}_e = \frac{1}{K-1}\sum_{i=1}^{K}\mathbf{x}_p\mathbf{x}_p^T = \frac{1}{K-1}\mathbf{X}_p\mathbf{X}_p^T \tag{2.36}$$

with $\mathbf{X}_p$ denoting the matrix of the ensemble perturbation vectors $\mathbf{x}_p$. With the notation of the ensemble Kalman Gain Matrix $\mathbf{K}_e$, the new analysis mean $\mathbf{x}_e^a$ is obtained from

$$\mathbf{x}_e^a = \mathbf{x}_e + \mathbf{P}_e\mathbf{H}^*(\mathbf{R} + \mathbf{H}\mathbf{P}_e\mathbf{H}^*)^{-1}[\mathbf{y}_0 - \mathbf{H}(\mathbf{x}_e)] = \mathbf{x}_e + \mathbf{K}_e[\mathbf{y}_0 - \mathbf{H}(\mathbf{x}_e)] \tag{2.37}$$

The analysis error covariance matrix $\mathbf{P}_e^a$ is obtained by (Evensen, 1994)

$$\mathbf{P}_e^a = (\mathbf{1} - \mathbf{K}_e\mathbf{H})\mathbf{P}_e. \tag{2.38}$$

The analysis error covariances for each member are added to the ensemble mean and the next ensemble forecast is performed. The advantage of applying a Kalman Filter is that a situation dependent error covariance matrix is obtained but the quality critically depends on the ensemble quality.

To realize a good ensemble spread, today often a multi-physics ensemble with perturbed boundary conditions of the applied NWP model is selected (e.g. Zhang et al., 2011). To update the individual ensemble members, a transformation matrix $\mathbf{T}$ is needed, which contains an "inflation factor" to give more weight on the ensemble perturbations (e.g. Wang et al., 2008). As the number of ensembles is limited by the available computing power, hybrid data assimilation schemes come into focus which combine an ensemble prediction system with a 3DVAR on the ensemble mean. During the 3DVAR update, not only $\mathbf{B}$ but also the ensemble error covariance $\mathbf{P}_e$ is considered in the following way (Wang et al., 2007):

$$\mathbf{B}_e = (1-\alpha)\mathbf{P}_e + \alpha\mathbf{B} \tag{2.39}$$

α is a weighting factor determining the influence of the ensemble on the 3DVAR analysis. In the next step, the analysis error covariances $\mathbf{P}_e^a$ are added to the updated ensemble mean to get an updated ensemble for the next forecast step. The method described above for calculating $\mathbf{P}_e^a$ is the Ensemble Transform Kalman Filter (ETKF; Bishop et al., 2001). The main difference to the EnKF is that only the analysis covariances are calculated and the ensemble mean is not updated. The latter is done by the 3DVAR in this case.

2.7 Rapid Update Cycle (RUC) system with the WRF model

As mentioned in the section 2.2, applying the 4DVAR can be an enormous challenge on the km-scale both with respect to computing resources and model physics. Therefore, the 3DVAR-RUC approach is in operation at many forecast centers around the world.

In general, due to the coarser resolution and different physics configurations of the driving model compared to the LAM, inconsistencies can occur during the first few hours because the model has to find its own balance. To overcome this difficulties, the model is initialized once with the coarser driving model at the beginning of the desired RUC. After the initialization, optional with

a digital filter initialization (DFI) to remove noise coming from the interpolation, the model is run for a 3 h forecast.

In the next step, the lower boundaries have to be updated. This includes fields like skin temperature, snow-free albedo and snow coverage. Thereafter the 3DVAR is performed using a previous forecast as background field $\mathbf{x}_b$ to obtain the new analysis $\mathbf{x}_a$. As the new analysis is not consistent with the input $\mathbf{x}_b$, the lateral boundaries have to be updated for the next forecast cycle. Following, if favored, a DFI can be applied again to remove noise introduced by the variational analysis. Preliminary tests with radar data assimilation revealed that the usage of DFI slightly weakens the reflectivity information and thus needs further investigation.

After the next short term forecast is finished, the procedure with updating the lower boundaries, 3DVAR and updating the lateral boundaries is performed until the final forecast is required by the end-user (see Figure 6). Typically the cycle frequency varies between 3 h and 6 h with an

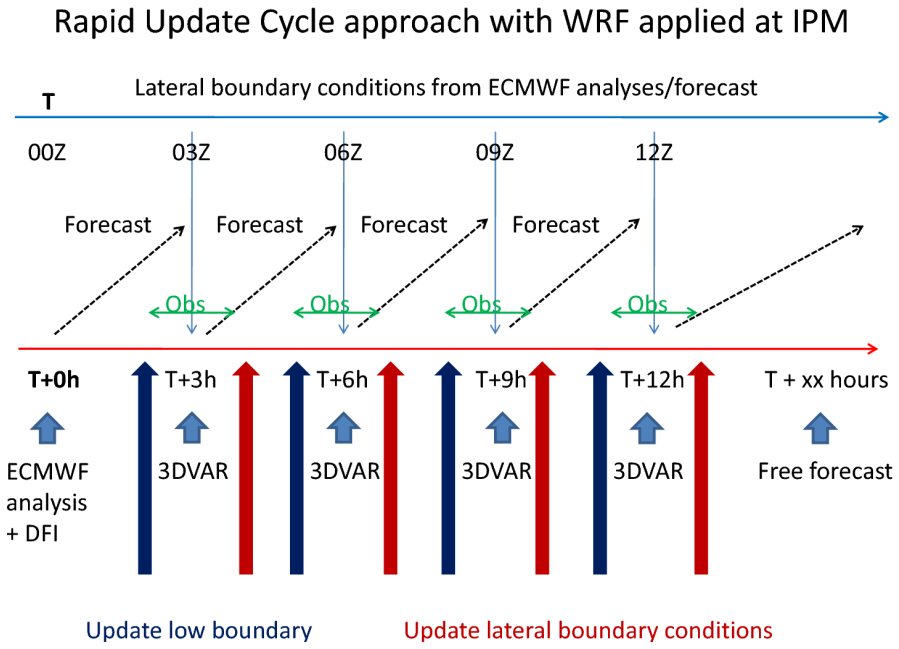


Figure 6: Schematic diagram of a Rapid Update Cycle.

assimilation window of typically ± 1hr. NCEP successfully operates the convection permitting High Resolution Rapid Refresh (HRRR, http://ruc.noaa.gov/hrrr) model based on WRF-ARW with a 1 h RUC frequency but tests with a 1 h cycle frequency revealed a detrimental performance probably because the model does not have enough time to develop its own "climate" (see later in section 4.3).

The advantage of performing a RUC instead of a cold start with 3DVAR is that the model is kept in its own climate and spin-up problems, regardless of which reason, are significantly reduced. In general, this RUC approach can be arbitrary modified with a 4DVAR or hybrid data assimilation scheme as it is often done by the global models from ECMWF, DWD, Météo France and UK MetOffice.

Another important point is the size of the LAM model domain where the assimilation is performed especially in case of a 3DVAR. During COPS and D-PHASE, the 4DVAR for the MM5 was performed on 18 km grid with a domain size of 64x70 grid points ($\sim$ 1100x1200 km; Zus et al., 2008) in order to minimize computational costs. Afterwards, a forecast with a 2-way nested configuration was performed with a resolution of 18-6-2 km. In retrospect this resulted in 2 deficiencies (see Publication II in chapter 5): The application of a convection parameterization and the obviously too small model domain for the inner nest.

To avoid nesting and the application of a convection parameterization, it is the best way to design a single domain for both the assimilation and the subsequent free forecast. E.g. the ALADIN model of Météo France covers an area of $\sim$ 3000x3000 km, and the high resolution AROME model covers an area of $\sim$ 1900x1800 km. A particular point for Central Europe is that at least the British Channel is covered by the convection-permitting domain. The importance of the model domain is shown later in publication III in chapter 5.

As the desired model domain is set up, the question arises how to use already available and new (remote sensing) observations. For conventional observations data simply have to be reformatted into the required data structure. For observations like GPS ZTD or STD, radar data and satellite radiances, several things have to be taken into account.

E.g. GPS-ZTD data are available every 15 min with about 400 stations. They are only assimilated in the nearest time windows around the assimilation date as the Zenith Wet Delay (ZWD) field is highly variable in time. The same applies for STD data, but the data density usually is far too high so that data thinning is necessary also to avoid that more than one STD ray goes though one model grid box. As similar procedure has to be performed for assimilating radar radial velocities. This is described in more detail in chapter 4.

For satellite radiances it is a bit more difficult as they are not available on a regular schedule, except geostationary satellite data like MSG. Therefore they will be assimilated within a $\pm$ 1 h time window around the analysis time.

To give the reader an impression of the RUC work flow in detail, appendix 6 shows an example how a RUC for the COPS period was recently set up including GPS-ZTD, radar and satellite radiance data for near real time applications.

3 Water vapor information from Global Positioning System

In this section, a brief introduction into the GPS and the retrieval of water vapor information will be given. GPS was originally introduced to provide accurate position information under all weather conditions that is to say like cloudy conditions, rain or snow everywhere on earth. The only requirement is that a direct line of sight exists to at least four of the satellites of the GPS network which currently consists of 32 satellites at an altitude of $\sim$ 20000 km above the earth surface.

GPS satellites transmit two different radio frequencies of 1.57542 GHz (f_1) and 1.2276 GHz (f_2) corresponding to wavelengths of 0.19 m (λ_1) and 0.244 m (λ_2). To determine the position of a antenna (receiver), it is necessary to have at least four satellites i available. The unknown parameters for the determination of position are the coordinates x,y,z and time difference between the satellite and receiver time. To obtain the position, one has to know the vector

$$\mathbf{x}_i = \begin{pmatrix} x_i \\ y_i \\ z_i \\ t_i \end{pmatrix} \tag{3.1}$$

of each of the four satellites plus a clock bias t_b of the receiver. From this point, a pseudo range Θ describing a phase difference in length units for each satellite can be defined:

$$\Theta = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} - t_b c \tag{3.2}$$

with

$$\Theta = (t_r - t_i)c \tag{3.3}$$

where t_r denotes the receiver time, t_i denotes the satellite time and c is the speed of light in the vacuum. With equations 3.2 and 3.3 the receiver location coordinates x,y and z can be determined with an accuracy of a few centimeters.

If the exact location of the GPS receiver is known, additional information can be obtained. Due to the earth's troposphere (with water vapor) and the ionosphere, the received signal is delayed compared to the vacuum which is known as tropospheric or slant path total and ionospheric delay. Following Hofmann-Wellenhof et al. (1997), the pseudorange can be also estimated by

$$\Theta = d + c(dt_s - dt_r) + \lambda_i n_i + STD + I \tag{3.4}$$

where d is the distance between the receiver and satellite, dt_s and dt_r denote the offset of satellite and receiver time compared to GPS reference time, λ_i is the carrier wave length, n_i denotes an integer of wavelengths λ_i between satellite and receiver due to an unknown phase and STD and I represent the slant path and ionospheric delay respectively. If dual frequency (f_1 and f_2) receivers are available, the ionospheric delay can be removed from equation 3.4 and thus the accuracy of Θ increases.

The slant total delay STD caused by the neutral atmosphere is given by (Bevis et al., 1992)

$$STD = 10^{-6} \int_S N ds \tag{3.5}$$

where N is the atmospheric refractivity which depends on pressure, temperature and water vapor and ds denotes the signal path from the GPS satellite to the receiver.

In case of an elevation of 90° (see Figure 10), the STD is called Zenith Total Delay (ZTD). Both can be split into a wet delay depending on the 3-dimensional temperature T, water vapor mixing ratio q, pressure p and a dry delay only depending on p_s. For a ZTD, the wet delay ZWD can be determined by integrating refractivity profiles i starting at the surface:

$$ZWD = \sum_{i=1}^{n} k_1 \frac{p_i \; q_i}{0.622 \; T_i} + k_2 \frac{p_i \; q_i}{0.622 \; T_i^2} \tag{3.6}$$

with constants $k_1 = 2.21 \cdot 10^{-7}$ K/Pa and $k_2 = 3.73 \cdot 10^{-3}$ K^2/Pa (Bevis et al., 1994). The dry hydrostatic delay ZHD is that suggested by Saastamoinen (1972):

$$ZHD = \frac{k_4 \; p_s}{1 - 0.00266 \cos(2\phi) - 0.00000028h} \tag{3.7}$$

with $k_4 = 0.000022768$ m Pa^{-1}, ϕ being the receiver latitude [deg], h the receiver altitude above the surface, p_s the surface pressure [Pa]. Thus the ZTD [m] reads now as

$$ZTD \;\; = \;\; ZWD + ZHD \tag{3.8}$$

The magnitude of the total delay depends on the elevation angle to the receiver with respect to a tangent at the earth surface. It can reach up to 2.5 m for ZTD and 10 m for low elevation STD measurements. Deriving STDs from an NWP model can be challenging as the refractivity profiles N_i have to be derived according to the line of sight to the satellite and the ray bending near the surface have to be taken into account. Further details about this issue can be found in the PhD thesis of Zus (2010).

Although STD and ZTD are integrated quantities, the application of a variational assimilation scheme gives information about the 3-dimensional structure of the atmosphere. With the application of an adjoint operator it is possible to adjust the 3-dimensional humidity, temperature and pressure fields in the model either above the receiver station in case of a ZTD or along the ray path through the atmosphere in case of STDs.

Application of STD instead of ZTD has a beneficial effect in a data assimilation scheme. As STDs can be measured with all elevations greater than zero, the tilted STD can give additional information about the horizontal water vapor distribution in the boundary layer which is especially important for summertime convection. The disadvantage or the critical point is that when STDs are obtained at low elevations $< 20°$(ϵ in Figure 10), the ray bending due to gravity has to be taken into account (Ereesma and Järvinen, 2006; Zus, 2010) in a similar way as it is considered for radar measurements (see later in section 4.3). An example for a difference modeled minus observed STD from the WRF model is given in Figure 7. A simple STD operator was implemented into the 3DVAR system to obtain an O-B statistics but cannot be used for assimilation purposes.

An example for the derivation of an adjoint model for ZTD is given in section 2.1 and Figure 8 shows the distribution of the GPS receiver network currently re-processed by the GFZ[9] EPOS software (Gendt et al., 2004) for 2007. Figure 9 shows an example for the coverage of GPS-STD data and Figure 10 illustrates how the STD is estimated.

[9]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences

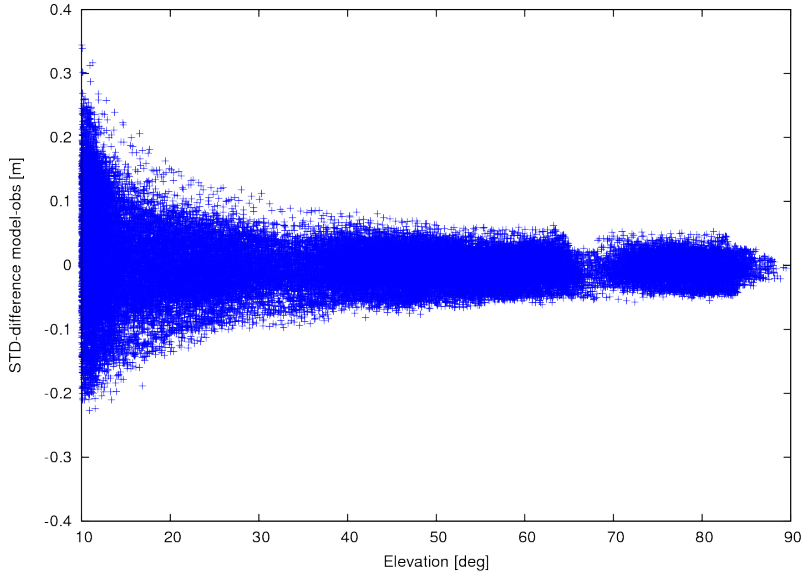


Figure 7: Example for a STD difference between model and observation depending on the elevation angle.

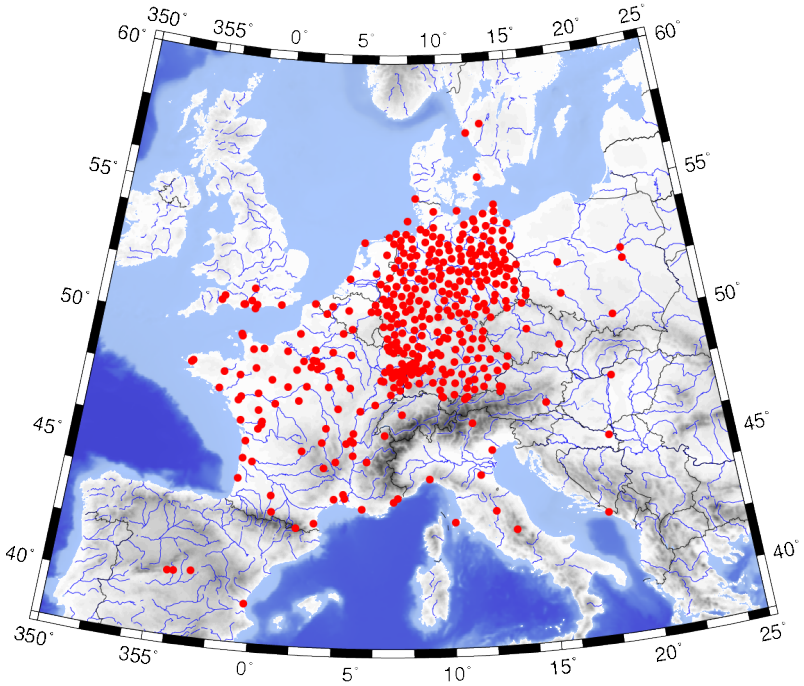


Figure 8: Overview about the GPS stations currently processed by the GFZ for 2007. The number of stations is about 400.

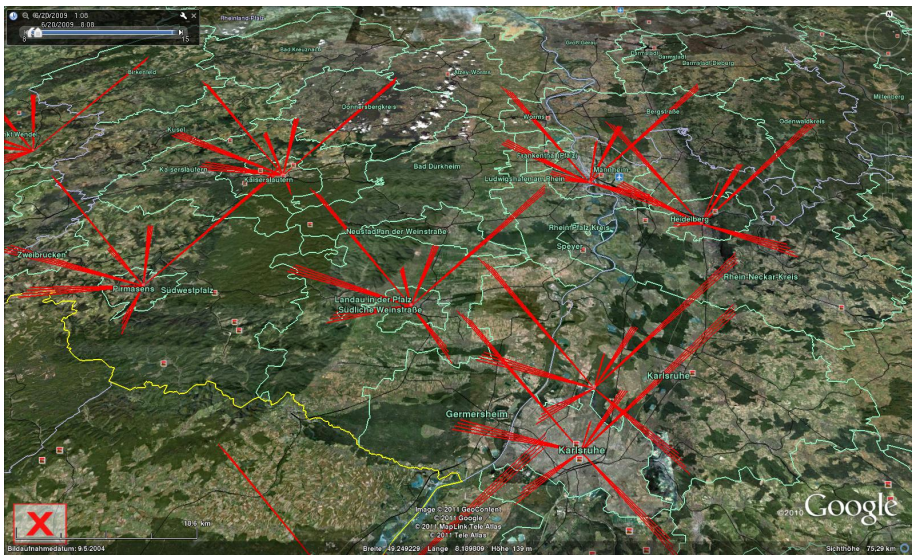


Figure 9: Example for the areal coverage of GPS-STD measurements. The red lines show the ray path which ends at the satellite. This image was taken from Google Earth with data provided by M. Bender, GFZ Potsdam.

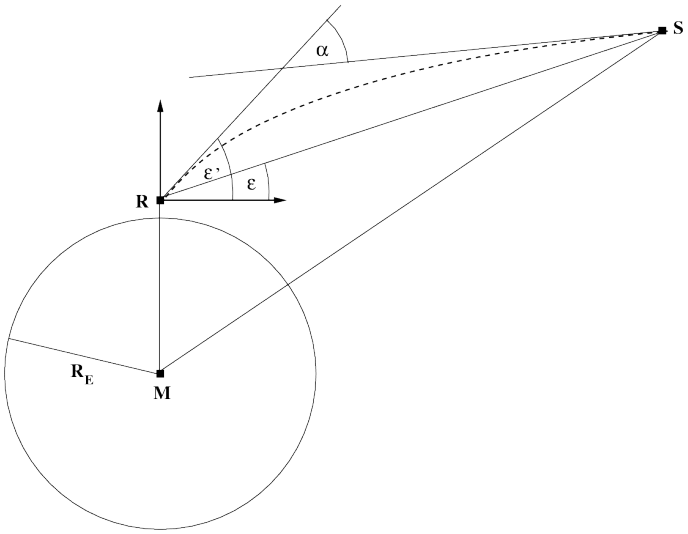


Figure 10: Schematic illustration how STDs are estimated. The continuous line is the ideal (vacuum) ray path from the satellite while the curved dashed line shows the real ray path through the atmosphere. ϵ denotes the elevation angle with respect to the surface. (Courtesy of F. Zus)

3.1 Example for the generation of an adjoint for the GPS-ZTD forward operator in the WRF model

As already mentioned, ZTD is split into a hydrostatic delay (ZHD) depending on pressure and temperature and a part that depends on the water vapor content and temperature (Zenith Wet Delay, ZWD):

$$ZTD = \int_0^s [(\underbrace{k_1 \frac{p\ q}{0.622\ T} + k_2 \frac{p\ q}{0.622\ T^2}}_{ZWD}) + ZHD] ds \tag{3.9}$$

Here p is the 3-dimensional air pressure, T the 3-dimensional temperature, q the 3-dimensional water vapor mixing ratio and e the 3-dimensional water vapor pressure. The empirical constants are given by $k_1 = 2.21 \cdot 10^{-7}$ K/Pa and $k_2 = 3.73 \cdot 10^{-3}$ K^2/Pa (Bevis et al., 1994). For the sake of clarity, vector quantities are neglected in the following calculations.

For simplicity, the new variables

$$a = \frac{\Delta h}{0.622}, \quad b = \frac{p\ q}{T}, \quad c = a\ b\ k_1, \quad d = \frac{a\ b\ k_2}{T}, \quad ZHD = \frac{z_1\ p_s}{z_f} \tag{3.10}$$

are introduced. Δh is the height difference between two model layers, p_s is the surface pressure and z_1 and z_f are constants for the calculation the hydrostatic delay above the model top (Saastamoinen, 1972). With these abbreviations, the vector $\mathbf{x}$ for the calculation of the ZTD field can be written as:

$$\mathbf{x} = \begin{pmatrix} q \\ p \\ T \\ p_s \\ b \\ c \\ d \\ ZHD \\ ZWD \\ ZTD \end{pmatrix} \tag{3.11}$$

The tangent linear model (TLM) requires the Jacobian of $\overrightarrow{H}$ as shown in equation (2.15) and can be written as a matrix-vector product:

$$\underbrace{\begin{pmatrix} \delta q \\ \delta p \\ \delta T \\ \delta p_s \\ \delta b \\ \delta c \\ \delta d \\ \delta ZHD \\ \delta ZWD \\ \delta ZTD \end{pmatrix}}_{\delta \mathbf{x}_{TLM}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{p}{T} & \frac{q}{T} & \frac{-pq}{T^2} & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & ak_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{abk_2}{T} & 0 & \frac{ak_2}{T} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{z_1}{z_f} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 100 & 100 & 0 \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} \delta q \\ \delta p \\ \delta T \\ \delta p_s \\ \delta b \\ \delta c \\ \delta d \\ \delta ZHD \\ \delta ZWD \\ \delta ZTD \end{pmatrix}}_{\delta \mathbf{x}_{TLM}} \tag{3.12}$$

The adjoint model (ADJ) is then calculated from (3.12) by the transpose of $\mathbf{H}$:

$$\begin{aligned}
\delta q^* &= \delta q^* + \frac{p}{T}\delta b \\
\delta p^* &= \delta p^* + \frac{q}{T}\delta b \\
\delta T^* &= \delta T^* - \frac{pq}{T^2}\delta b - \frac{abk_2}{T}\delta d \\
\delta p_s^* &= \delta p_s^* + \frac{z_1}{z_f}\delta ZHD \\
\delta b^* &= ak_1\delta c^* + \frac{ak_2}{T}\delta d \\
\delta c^* &= \delta ZWD \\
\delta d^* &= \delta ZWD \\
\delta ZHD^* &= 100\delta ZTD \\
\delta ZWD^* &= \delta WZD^* + 100\delta ZTD \\
\delta ZTD^* &= 0
\end{aligned} \tag{3.13}$$

With this adjoint model, increments of water vapor δq^*, Temperature δT^* and pressure δp^* are added to the background field $\mathbf{x}_b$ as new analysis $\mathbf{x}_a$.

4 Radar data assimilation

In the previous chapter, the focus was set on the application of (active remote sensing) GPS STD and ZTD data to improve the spatially and temporally highly variable water vapor distribution. As for example in an area which is highly favorable for convection or convective initiation, it can be inadequate to assimilate only tropospheric humidity in this conditions. An option to obtain information about dynamics are active remote sensing systems. Recently, Doppler LIDAR and Doppler precipitation RADAR[10] are employed to obtain information on the dynamic structure.

A Doppler LIDAR employs a LASER beam (mostly in the ultraviolet) which is send out to the atmosphere through a telescope. Due to the aerosols in the atmosphere, the scattered beam exhibits a frequency change which allows the detection of wind speed along the ray path (Line of sight, LOS). A LIDAR system typically has a LOS resolution of 50–100 m and can measure up to altitudes of 15 km. As large as the advantage of the high resolution is, due to the scattering properties of particles a LIDAR measurement is restricted to cloud free areas. Due to the applied wavelength, the beam does not penetrate clouds and thus a strong backscatter signal is detected of which no doppler shift can be extracted. Further details about the basics and recent developments of Doppler LIDAR wind measurements can be found e.g. in Reitebuch et al. (2009).

In contrast to a LIDAR System, radar systems have an larger spatial coverage but they need tracers for the detection of Doppler radial velocities. Additionally, reflectivity measurements describing the intensity ratio between the emitted and scattered/reflected microwave beam are in use to allow meteorologists drawing conclusions of the strength of a convective system, e.g if hail can be expected on the ground. This makes it currently a favorable measurement system in areas where convection already took place. Radar systems use microwaves (so called “C-Band”, “X-Band“ or ”S-Band“ radars, depending on the frequency) and have a range of $\approx$ 200 km.

As a supplement for getting information about dynamics (wind), passive remote sensing instrument retrievals from satellites (e.g. MSG from EUMETSAT; EUMETSAT (2009)) are applied in current NWP data assimilation schemes. The advantage is that a large area is covered by the satellite. Depending on the satellite orbit, the temporal resolution can be very coarse, e.g polar-orbiting satellites overpasses one location twice a day whereas satellite derived winds from MSG are available every hour. Another disadvantage is the fact, that clouds have to be present, since the wind information is retrieved by tracking cloudy pixels in the satellite image.

As the application of radar data in combination with NWP is a relatively new research area, the next section will give further details about radar radial velocity and reflectivity measurements and their application in current NWP assimilation schemes.

4.1 Measurement principle

Doppler precipitation radars emit electromagnetic pulses which are backscattered from hydrometeors like rain droplets or hail particles as well as from insects (”clear air echoes”). Precipitation radars in general are usually classified into 3 categories depending on the emitted wavelength:

[10] **R**adio **D**etection **A**nd **R**anging

- S-Band Radar with frequencies of ≈ 3 GHz
- C-Band Radar with frequencies of ≈ 6 GHz
- X-Band Radar with frequencies of ≈ 10 GHz

In Europe mostly C-Band and X-Band radar systems are in operation. They have a typical range of 200–250 km with a horizontal resolution of 30–250 m which is often averaged to about 500–1000 m in the horizontal to reduce the amount of raw data. The radar antenna can be rotated so that different antenna azimuths and elevations can be scanned. Typical azimuth increments are 1°and the elevation angle increments increase with the antenna elevation from 1°–10°.

To benefit from the Doppler effect, precipitation radars operate with a pulse repetition rate (PRT) of 200–1500 Hz meaning that e.g. every 5 μs an electromagnetic pulse is emitted. Within this time, the hydrometeors can move in the 3 directions u,v, and w. Due to the motion, the scattered electromagnetic wave has a different phase compared to the emitted wave. From this behavior, the radial velocity v_r of a target can be derived following e.g. Doviak and Zrnić (1993):

$$v_r = \frac{\lambda \cdot PRT}{4} = \frac{\lambda \Delta \Phi}{4\pi \Delta t} \tag{4.1}$$

with λ being the wavelength of the pulse, $\Delta\Phi$ the phase difference of the scattered wave and Δt the time between two radar pulses.

Equation 4.1 shows, that Doppler radars do not really measure a Doppler shift rather than a phase shift between two electromagnetic waves from which the radial velocity is derived. As the number of particles in the scanned volume, which increases with the distance from the antenna, is assumed to be large, a continuous velocity distribution of the particles is assumed. The velocity spectrum is then weighted by the power to obtain the mean radial velocity (see e.g. Doviak and Zrnić, 1993).

Unfortunately, the so-called "Doppler dilemma" complicates the derivation of radial velocities (Doviak and Zrnić, 1993). It allows only one unambiguous radial velocity v^u_{max} (Nyquist velocity, see e.g. Holleman and Beekhuis, 2003)

$$v^u_{max} = \frac{1}{R^u_{max}} \frac{\lambda \cdot c}{8} \tag{4.2}$$

dependent on the wavelength λ with the speed of light c and the unambiguous maximum range R^u_{max}. The latter is defined by

$$R^u_{max} = \frac{c}{2 \cdot PRF} \tag{4.3}$$

If a scatterer is at a distance R farther than R^u_{max}, it appears that the scatter has a distance of $\mathrm{R}' = R - (M-1) \cdot R^u_{max}$ where M denotes the number of R^u_{max} intervals. PRF describes the pulse repetition frequency.

As an example, for the DWD C-Band radar λ is 5.3 cm and the pulse repetition frequency is 600 Hz. This would lead to R^u_{max} of 250 km and to an unambiguous radial velocity of ±3.31 m/s. Here the range is acceptable but the unambiguous velocity is too small as particles can move

up to 80 m/s when measuring in the jet stream area. Therefore, the dual PRF described in Holleman and Beekhuis (2003) is often used which decreases R^u_{max} but significantly increases v^u_{max}.

To keep the DWD example, the low PRF is 800 Hz and the high PRF is 1200 Hz leading to a maximum unambiguous range of 125 km and velocity of ±32 m/s. Recently Météo France employs a triple PRF scheme allowing unambiguous velocities of about 60 m/s and a maximum range of 250 km (Tahanout et al., 2009). The data availability window for reflectivity and radial velocity in Europe is usually 15 min because of the antenna rotation speed and other scan procedures (e.g. PPI scan at a constant elevation with a comparatively low PRF).

Another important feature for short range forecasting and nowcasting is the intensity of the scattered radar beam. This is called the reflectivity Z which depends on the particle diameter D and the assumed drop size distribution N(D) and is related to the number of droplets per unit volume and the sixth power of the droplet diameter (mm^6/m^3). Assuming e.g. a Marshall-Palmer distribution N(D) for hydrometeors (Marshall and Palmer, 1948), the reflectivity Z reads

$$Z = \int_0^D N(D)dD = \int_0^D N_0 e^{-\Gamma D} D^6 dD \tag{4.4}$$

N_0 describes a droplet concentration at 0 radius. For radar reflectivity the unit dbZ is used which is 10 times the logarithm of Z. From this kind of reflectivity, one can detect whether a convective system contains only rain or also hail which is more likely when the reflectivity exceeds 55 dbZ. It is also interesting to note that aircrafts are advised to avoid areas with reflectivities greater than 37 dbZ as this can lead to substantial damage to the aircraft and loss of control due to assumed strong updrafts.

4.2 Possible error sources

Due to the synoptic situation and the ambient conditions, several error sources are possible. The most prominent error source is ground clutter. This can occur when objects like buildings are present within a radius of a few kilometers around the radar site. They can produce strong echoes up to 30–40 dbZ and erroneous radial velocity field. Recently, clutter removal techniques have been developed (e.g. Seltmann, 2000) applying a high-pass filter or FFT (Fast Fourier Transformation) to use near radar radial velocities. Figure 11 shows an example of a low elevation scan from the the Feldberg radar for near distance ground clutter effects. A recently arising problem for obtaining reliable radar data are wind turbines. The rotating rotor blades also reflect the radar beam but due to the rotation large radial velocities can occur. Therefore the corresponding range bins are blacklisted for the DWD radar sites (Helmert et al., 2008; Hengstenbeck et al., 2010). An example for this negative spike is shown in Figure 12 from the Feldberg radar. Positive spikes can e.g occur due to interference of different radars or wireless networks as they use the same wave length (band) as precipitation radars (Hengstenbeck et al., 2010).

In strong convective systems, that is to say large rain drops or hail occur, the radar beam can be completely reflected by the hydrometeors. As the radar beam usually penetrates the cloud, this absorption can lead to a strong attenuation and thus to too small reflectivities at more distant

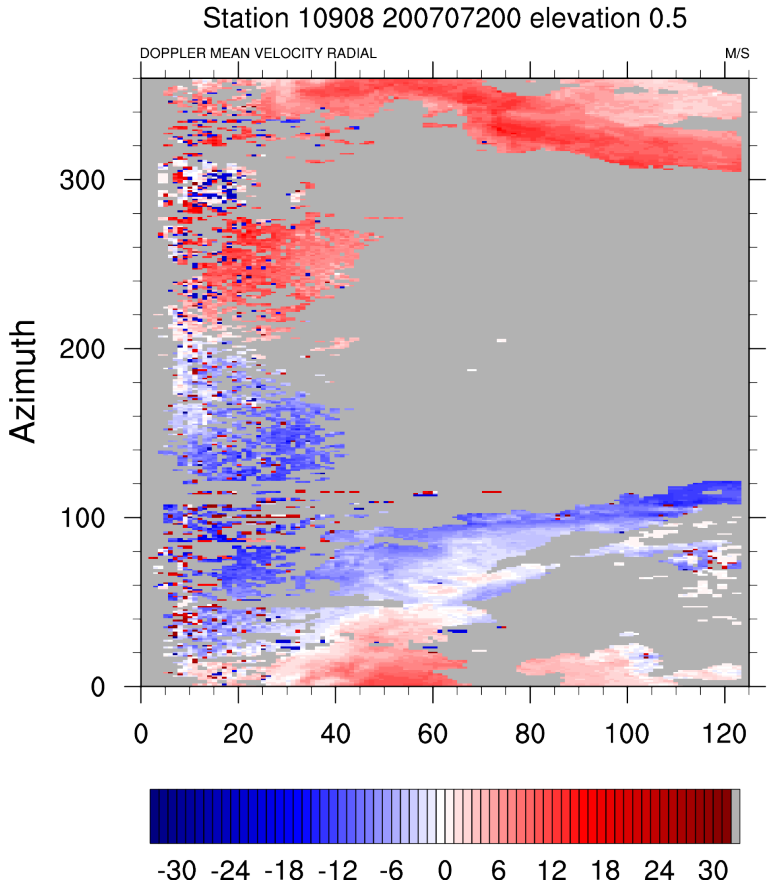


Figure 11: Example for near distance clutter echoes for radial velocities from radar Feldberg at an elevation of 0.5°.

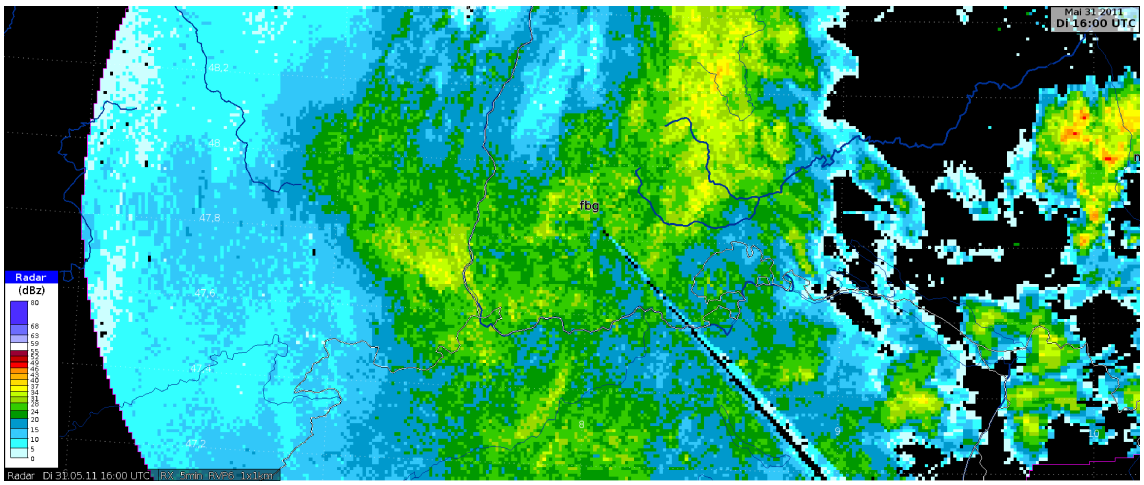


Figure 12: Example for a negative spike from Feldberg radar (marked with fbg) which can occur due to erected cranes or wind turbines. The range bins are blacklisted. This image was taken from the IPM-Ninjo system.

ranges. This can partially be overcome by an overlapping area between two radar sites but is technically not easy to implement.

When working with radars, it is assumed that the backscattered beam has the same direction as the emitted beam but this is not fully correct. The wave is scattered in all directions and especially when hail is present, a non-negligible part is scattered towards the surface and scattered back to the hydrometeor and is recorded as an increased intensity also leading to an position error of the signal. This phenomena is referred to as "Hail spike" and is described in more detail by Lemon (1998).

Other phenomena are the anomalous beam propagation due to super refractivity and the so called "bright band". The former especially occurs in the morning when a strong temperature inversion can occur which prevents the radar beam penetrating the warm air above the inversion layer (Bech et al., 2003). The bright-band effect occurs when frozen particles penetrate the melting layer. They are covered by a thin layer of water which better reflects the radar beam compared to snow leading to increased reflectivities. Usually this zone has a depth of few hundred meters. Further details about possible detection algorithms for bright band effects can be found e.g. in Gourley and Calvert (2003).

4.3 Application of radar data in NWP models

The direct assimilation of radar reflectivities, polarizations and radial velocities are only possible within a variational or ensemble data assimilation framework as observation operators are required. For NWP models where the analysis is updated with a simpler nudging scheme, a different method has to be applied. E.g. the DWD applies the LHN procedure for COSMO-DE which should force convection in the short term time-scale. LHN is based on a radar derived precipitation rate R_r which is obtained the following relation:

$$R_r = aZ^b \tag{4.5}$$

where Z is the 3-dimensional radar reflectivity and a,b are empirically derived constants. It is assumed that the falling hydrometeors and their phase transitions release a certain amount of latent heat leading to a warming ΔT of the ambient air masses. This temperature increment is then related to the model derived precipitation R_{model}, the latent heat release and observed rain rate in the following way (see e.g. Stephan et al., 2008 for the application in COSMO-DE):

$$\Delta T = (\alpha - 1)\frac{1}{C_p}\Delta LH \tag{4.6}$$

where α is defined as $\frac{R_r}{R_{model}}$, C_p is the specific heat capacity for dry air at constant pressure. ΔLH is the latent heat released by condensation which is usually calculated in the cloud microphysics scheme. The profile temperature increment is finally added to the model background with an additional weighting term.

A weak point is the addition of only the temperature increment to the model background state. This can lead to strong precipitation in a short-term forecast as the often positive increment leads to a potential instability (increase of CAPE).

A similar method compared to LHN is the Physical Initialisation Bonn (PIB, developed by Haase, 2002) for improving the very short term precipitation forecast. The aim of this approach is to change the vertical velocity field in order to initiate and support convection. PIB uses the same radar derived and model rain rates (R_r and R_{model}) to obtain an updated rain rate R_{update}. Then, a new rain rate R_{new} is derived from R_{update} based whether a cloud exists in the model or not. Based on R_{update}, the vertical velocity inside the clouds is calculated. Compared to the LHN, this method also changes the vertical structure of cloud water and cloud ice. The disadvantage of this approach is that a synthetic 3-dimensional radar derived rain rate has to be assumed because the radar does not scan the whole hemisphere. Further details and results can be found in Milan et al. (2008).

Especially in summer, convergence lines play a major role for the initiation of convection. To better represent (near) surface wind, to date mostly synoptic and airport measurements are incorporated into data assimilation schemes. Figure 13 shows an example for surface wind data obtained from the Global Telecommunication System (GTS) and for the COPS year 2007. It can be seen from Figure 13a, that the station density obtained from the GTS is very coarse with an average distance of 25–30 km compared to the Joint D-PHASE COPS data set data set (Dorninger et al. (2009), Figure 13b). In the summer season, this can be too coarse to represent convergence lines so that this may be not very useful for correctly predicting the triggering of convection with the aid of an assimilation system. Therefore, the assimilation of high resolution radar radial velocities and reflectivities becomes important.

Unfortunately radial velocities are only available in the presence of scattering targets, that is insects or hydrometeors. One of the first attempts using radar data for assimilation in NWP models was performed by Sun and Crook (1997) who developed a 3DVAR containing observation operators for radar radial velocities and reflectivities for the application of a convective storm in Florida. Further attempts were made e.g. by Xiao et al. (2005) and Xiao and Sun (2007) for the U.S. applying the WRF model and by Lindskog et al. (2004) and Montmerle and Faccani (2009) applying the HIRLAM and AROME models over Europe.

To assimilate radar radial velocities in a variational assimilation system, a model forward operator is needed. The three model wind components u, v, and w have to be projected to the radar beam position. In the WRF model, this is done in the following way (Xiao and Sun, 2007):

$$v_r = \frac{x - x_i}{r_i}u + \frac{y - y_i}{r_i}v + \frac{z - z_i}{r_i}(w - v_T) \tag{4.7}$$

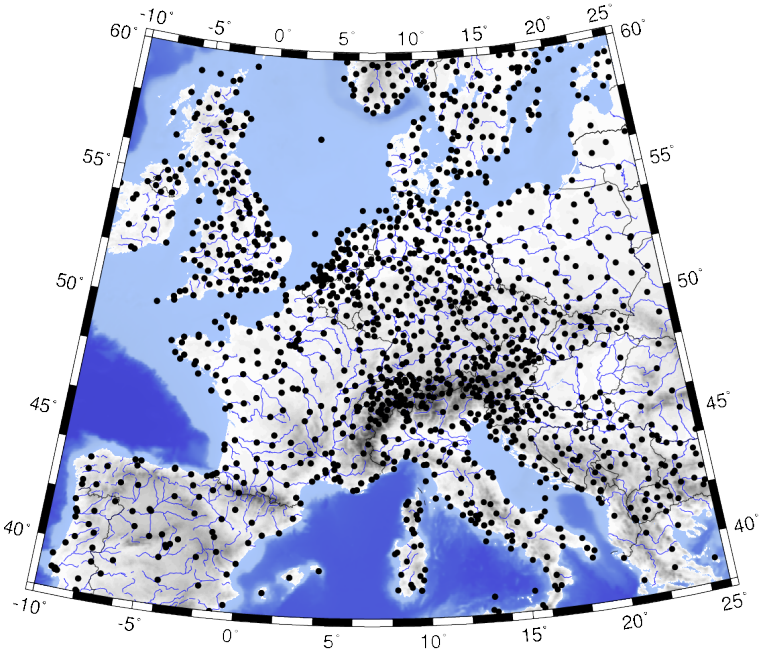
Here x,y, and z denotes the radar location, x_i,y_i and z_i are the location of the radar observation and r_i is the distance between the radar location and the observation relative to the center of the earth (see Figure 14). v_T is the so called "terminal velocity" describing the fallspeed of rain particles in the model and is defined by

$$v_T = a\ b\ q_r^{0.125} \tag{4.8}$$

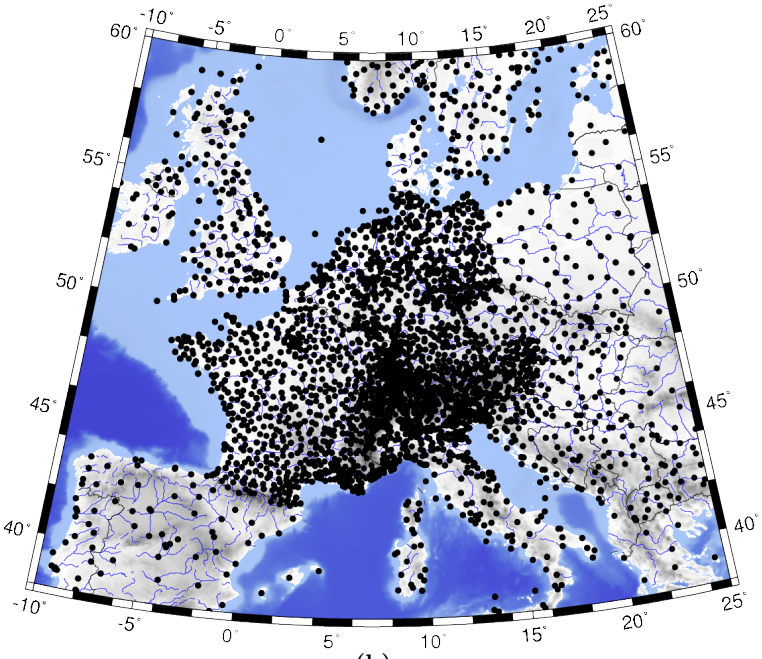
where q_r is the rain water mixing ratio and b is estimated as

$$b = \frac{p_s}{p_b} \tag{4.9}$$

p_s is the surface pressure and p_b denotes the base state model pressure. a has a value of 5.4 m/s and is derived under the assumption that the Reynolds Number is $\sim$ 300 and with a drag coefficient of 1. Further details can be found in Pruppacher and Klett (1997).



(a)



(b)

Figure 13: Spatial distribution of surface wind measurements. (a) shows data available from GTS, (b) shows data available for 2007 including the JDC data set. Note that a lot of stations in the mountains are rejected in the assimilation system due to too large height differences between model and reality.

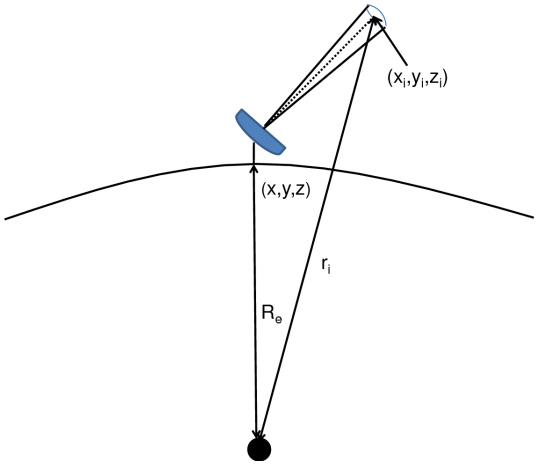


Figure 14: Schematic diagram for the radar radial velocity operator. R_e is the radius of the earth.

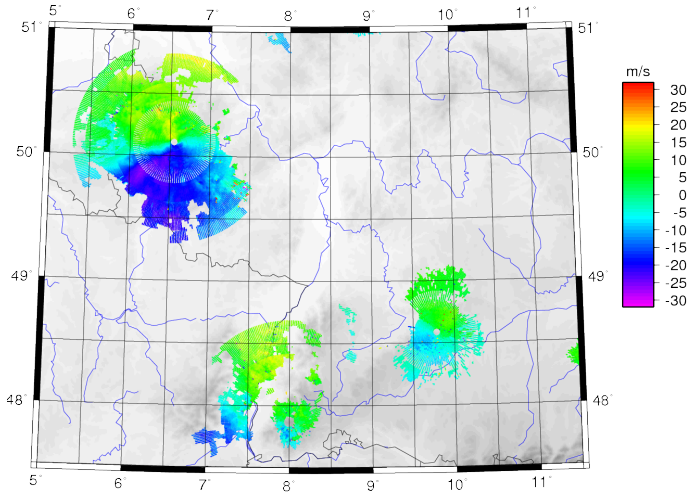
As mentioned at the beginning of this section, radar data are available with a horizontal resolution of $\sim$ 1 km which should contain information about small scale features. Figure 15a shows the observed radial velocity for altitudes below 4000 m above ground for four DWD radar stations at 10 UTC July 20, 2007. It is seen that there are noise, spikes and circles in the observation. Before going into an assimilation, this has to be removed with the help of quality flags available.

A result of the filtering developed in this thesis is shown Figure 15b where noise is reduced. As the number of observations is still too large compared to the model grid points after filtering, a data thinning has to be performed in addition. An example for filtered and thinned radial velocities is shown in Figure 15c.

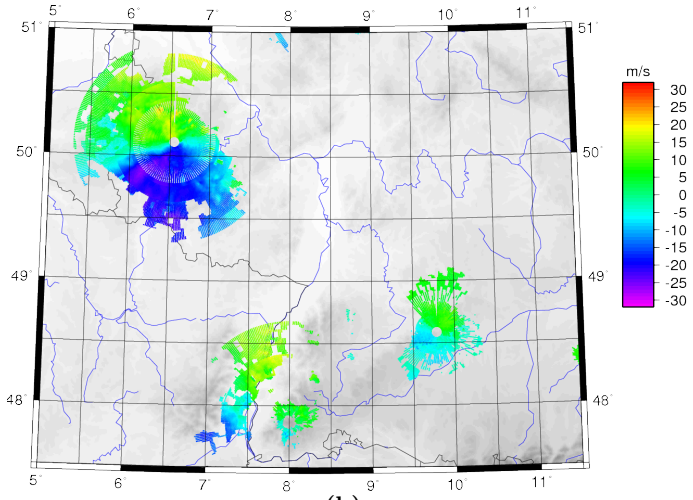
The process of data thinning is still very preliminary for the DWD radar data and has to be refined to have them thinned on a regular latitude-longitude grid with approximately four to five times the horizontal model resolution (Olivier Caumont, personal communication). What has to be taken into account in addition is the ray bending of the radar beam due to the curvature of the earth. With the assumption that the change of atmospheric refractivity is very small and the altitude is limited to 20 km, the 4/3 earth model of Doviak and Zrnić (1993) can be applied.

Currently, the German radar data are filtered similar to an approach of Xiao et al. (2008) so that only observations with a variance of the surrounding observations smaller than 60 m^2/s^2 are accepted. Otherwise they are set to missing values. Data thinning and filtering for French radar data is done with a procedure described by Montmerle and Faccani (2009) and is currently performed on a $\sim$ 10x10 km grid. First simulations with the additional usage of German and French radar data show promising results when applying a 3 h RUC whereas a 1 h RUC shows degraded results probably due to spin-up effects after the assimilation.

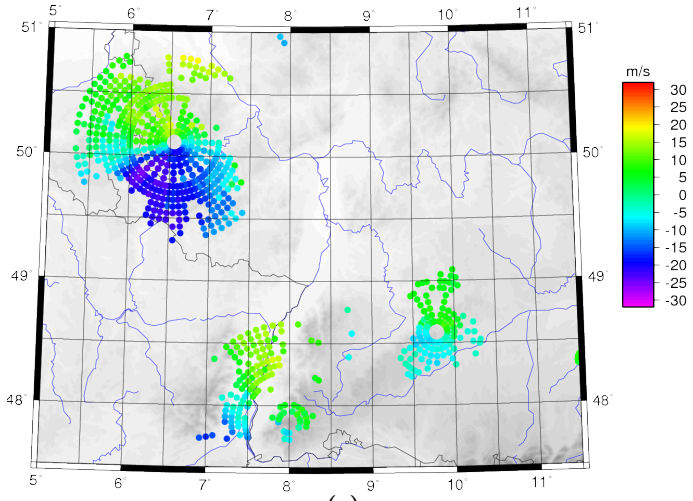
Figures 16 and 17 show an example for the different results during a RUC applying radar data for the wind field at the surface and around 5 km altitude. The observations fed in the assimilation system are the same, the only difference is the interval between two assimilation steps. It can



(a)



(b)



(c)

Figure 15: Observed radar radial velocities below 4000 m for 20 July 2007. From top to bottom: Raw data, filtered observations using quality flags, thinned data used for assimilation.

be clearly seen, that the 1 h RUC analysis shows a much higher variability compared to the 3 h RUC interval which degrades the forecast quality.

As radial velocities can be obtained also in case of insects as scatterers (with the assumption that they move with the main wind direction), the assimilation of radar reflectivity requires hydrometeors to be present in the atmosphere. It is obtained with the assumption from equation 4.4. Assimilating radar reflectivity allows, depending on the applied forward operator, an adjustment of rain and cloud water content or in addition of solid hydrometeors. For example, in the WRF model the reflectivity forward operator (Sun and Crook, 1997) considers rain water mixing ratio and is defined by

$$Z = 43.1 + 17.5 \ \log\left(\frac{\rho_{air}\ q_r}{1\ kg\ m^{-3}}\right)[dbZ]\,. \tag{4.10}$$

ρ_{air} is the dry air density and q_r describes the rain water mixing ratio at the corresponding observation location. Relation 4.10 is derived from a Marshall-Palmer distribution with a number density N_0 of $8{\cdot}10^6$ m^{-4}.

When assimilating radar reflectivity data, the control variable pseudo relative humidity is not only based on the water vapor mixing ratio q (see section 2.3) but on the total liquid water mixing ratio q_t which contains mixing ratios of water vapor, cloud water and rain water.

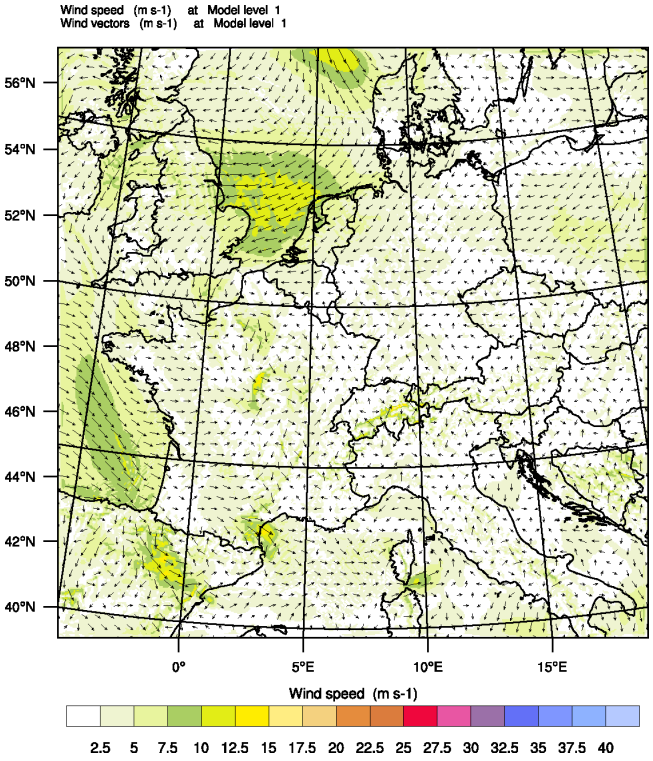
The total liquid water content is transferred to a warm rain scheme which includes condensation of water vapor into cloud water, the accretion of cloud water by rain and evaporation of rain water to cloud water (Xiao and Sun, 2007; Xiao et al., 2007). Note that this partitioning takes place in the inner loops of the assimilation process (section 2.4). In the adjoint process, increments of temperature (due to evaporation), water vapor, cloud water and rain water are added to the model background field.

One of the biggest drawbacks of this method is the fact, that in case of strong radar reflectivities, the assumption of the observation operator that only rain water is present can be violated. This is considered by the fact that observed reflectivities > 55 dbZ at altitudes larger than 5000 m are rejected in our experiments before going into the assimilation.

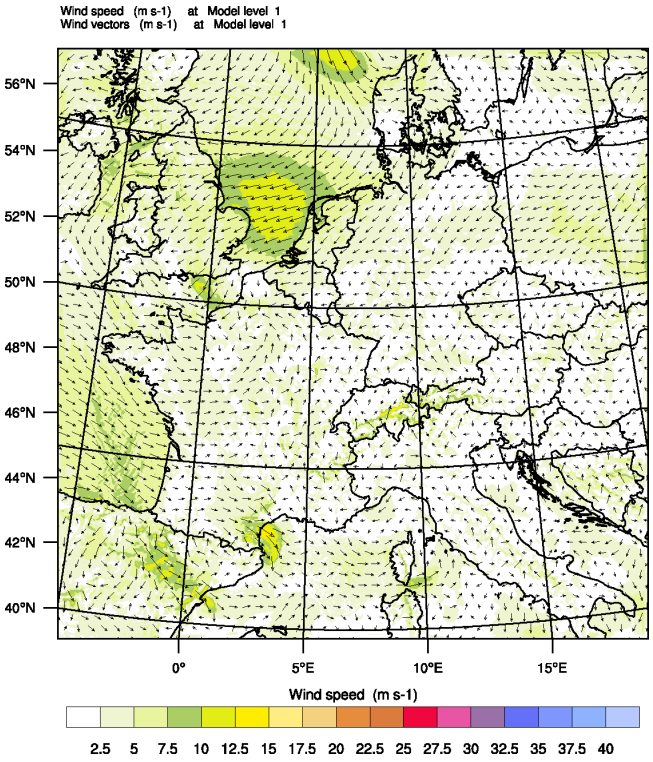
The data thinning follows the same procedure as described for radial velocities except that for reflectivities data are rejected if the variance of 150 dbZ^2 is exceeded (Xiao et al., 2008). Figure 18 shows an example for reflectivity data quality control.

So far, only reflectivities of spherical particles have been considered. As hydrometeors usually do not have a circular shape, and thus showing different scattering properties, great efforts have been made on the development of dual-polarization Doppler precipitation radars. These radars emit two microwaves with different polarizations. From the returning signal, several quantities like differential reflectivity (which gives information about the drop size) can be derived to get information of the hydrometeors present in the cloud (e.g if rain or snow is present).

Jung et al. (2008) recently developed a forward operator to calculate polarized reflectivities depending on rain water, snow and hail content. Gallus and Pfeiffer (2008) compared reflectivities calculated from several microphysics schemes in the WRF model based on the synthetic polarimetric radar (SynPolRad, Pfeiffer, 2007). They both suppose that when two-moment microphysics schemes become available, the results will get even better because the number of assumptions for the drop size distribution N(D) in the polarization operator decreases.

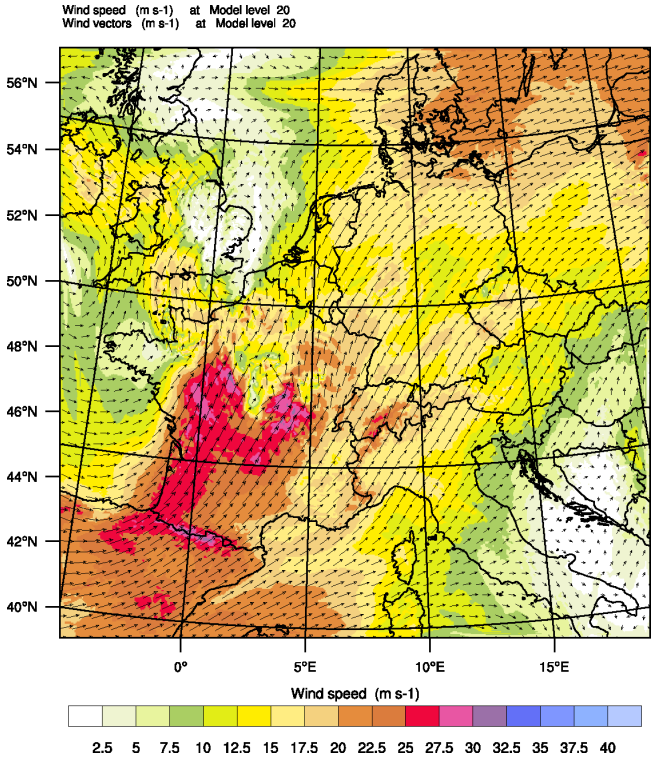


(a)

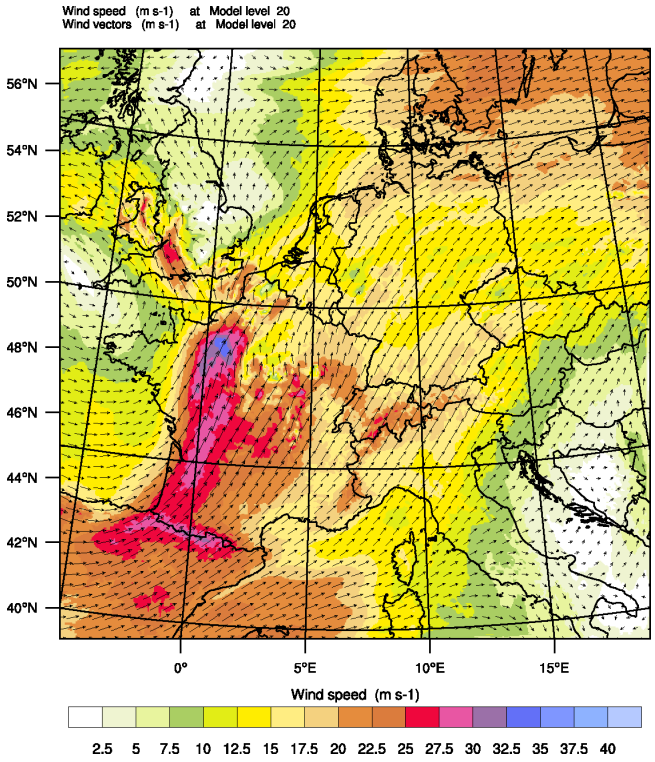


(b)

Figure 16: Wind field at the lowest model level for different RUC intervals. (a) shows the analysis for a 1 h interval, (b) shows the analysis for a 3 h interval.



(a)



(b)

Figure 17: Same as Figure 16 but for an altitude of 5000 m.

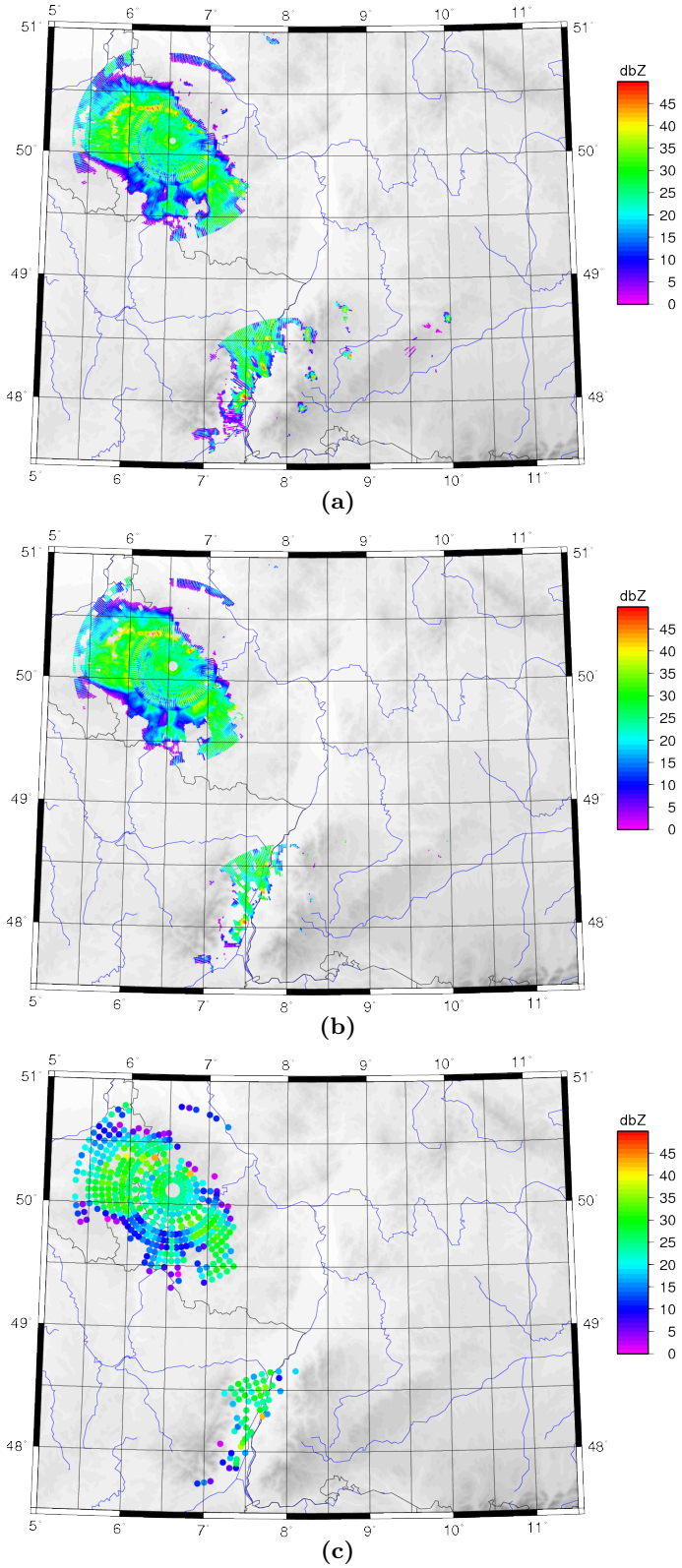


Figure 18: Observed reflectivities up to 4000 m altitude for 20 July 2007. From top to bottom: Raw data, filtered observations using quality flags, thinned data used for assimilation.

To partially overcome the problem of designing a reflectivity forward operator, Caumont et al. (2010) recently developed a 1DVAR scheme for radar reflectivity assimilation in the AROME model. The assimilation of reflectivity is split into two parts. First the reflectivity measurements at different elevations are split into vertical columns at one location (x,y). Afterwards, a 1DVAR retrieval is performed where the difference between modeled and observed reflectivity profile is applied as a weighting factor to the background relative humidity. The relative humidity profile is then assimilated as conventional observations.

This method is similar to latent heat nudging (e.g Schraff et al., 2006) but takes the hydrometeor background into account. As this method adjusts the humidity distribution, information can be lost if the lifting condensation level (LCL) is relatively high or the moisture imbalance after the assimilation can be very large. Also if the relative humidity is set to 100% below the LCL, precipitation may be not able to evaporate and thus additional water vapor is missing for possible further convective developments which is similar to the the PIB scheme of Milan (2009).

5 Summary of publications

In this chapter, brief summaries of three publications covering physics sensitivity tests with MM5 (I), the evaluation of the MM5 system operational during COPS (II), and first assimilation results with the new WRF model (III) are given in this section.

Publication I:
T. Schwitalla, H.-S. Bauer, V. Wulfmeyer und G. Zängl, 2008: Systematic errors of QPF in low-mountain regions as revealed by MM5 simulations, Meteorol. Z., **17**, 903 - 919.IF: 1.4

Publication II:
H.-S. Bauer, T. Weusthoff, M. Dorninger, V. Wulfmeyer, **T. Schwitalla**, T. Gorgas, M. Arpagaus, K. Warrach-Sagi, 2010: Predictive Skill of a Subset of models that participated during D-PHASE in the COPS Region. Q. J. R. Meteorol. Soc., **137**, S1, 287-305. IF: 2.98

Publication III:
T. Schwitalla, H.-S. Bauer, V. Wulfmeyer und F. Aoshima, 2010: High-resolution simulation over central Europe: Assimilation experiments during COPS IOP9c. Q. J. R. Meteorol. Soc., **137**, S1, 156-175. IF: 2.98

5.1 Schwitalla et al. (2008)

In publication I, high-resolution numerical simulations were performed using the mesoscale model MM5. The investigated area was the COPS region (Vosges Mountain, Black Forest and Swabian Jura) and the focus was set on convective-induced summertime precipitation.

A representative set of 13 precipitation events for this region was chosen and simulations with a horizontal resolution of 7 km and 1 km were performed. The major finding is that the 7 km simulations show a strong windward-lee effect for precipitation, implying an overestimation on the windward side with hardly any precipitation on the lee side.

Moreover precipitation occurred too early due to an incorrect simulation of the flow which is seen in averaged 850 hPa vertical velocity fields.

The high-resolution simulation with 1 km horizontal resolution, in contrast, shows a temporal delay of $\sim$ 2 h in the precipitation development while the shape of the diurnal cycle is fairly well represented.

To select an applicable physics configuration for an operational setup for COPS, a subset of 4 of these 13 cases was selected. Three different boundary layer and two different land-surface models were applied. In general, all configurations have problems simulating the surface humidity correctly showing a strong decrease during sunset probably due to a too strong mixing and have cold 2 m temperature bias during daytime.

A configuration combining a countergradient, non-local PBL scheme (that is mixing can occur not only between adjacent model layers in one time step) with the 5-layer soil model was found to be most successful and was finally applied for the operational model setup during COPS. In contrast, a turbulence closure scheme (ETA-scheme) produced unrealistic results with a moist and cold bias due to an inaccurate simulation of sensible and latent heat fluxes.

Finally a comparison with GPS IWV for the 1 km domain indicated that the model overestimated the IWV content by about 15–20% throughout the simulation indicating a too moist boundary layer. A possible error source seems to be the NOAH-LSM which, in combination with the MRF boundary layer scheme, predicts comparatively high latent heat fluxes and low sensible heat fluxes.

5.2 Bauer et al. (2011a)

In publication II, a subset of the D-PHASE model ensemble has been evaluated over the COPS period with the main focus set on daytime precipitation in the COPS area. This model subset also includes the MM5 with the physics package selected in publication I.

The selected models include the operational models of DWD and the pre-operational models of Meteo Swiss and Météo France and two MM5 forecasts. For the verification of the daytime precipitation, the gridded VERA data (Steinacker et al., 2006) and for the validation of the diurnal cycle of precipitation, surface wind and 2 m temperature from the JDC-plus data set (Dorninger et al., 2009) were employed which consists of up to 10000 surface observations over central Europe.

These observations had to be translated into a homogeneous format as every meteorological service has its own data format and quality information. The second step was the reformatting to feed this data into the MET package (DTC, 2009) while doublets had to be filtered or removed to avoid falsifying the evaluation.

In general, both COSMO models with convection parameterization overestimate the daytime precipitation and show the well known windward-lee effect. Both MM5 models (initialized with and without a 4DVAR) show, in contrast to all other models a dry bias. It is assumed, that the position of the inner model domain (which seems too small) and the turning-off of the cumulus scheme is the main reason for that. The 4DVAR initialized simulation slightly improves the situation but not as much as expected.

Regarding the diurnal cycle of precipitation, both MM5 simulations show a good performance compared to the operational models. Interestingly the 4DVAR driven MM5 shows a significant overestimation of precipitation in the first simulation hours whereas the CONTROL experiment shows the usual cold start spin-up. The former is probably the effect of different model and assimilation packages (3.7 vs. 3.4.3). Although the assimilation is performed on 18 km resolution, the innermost domain is also affected during the forecast because of the applied 2-way nesting option. In a further improved, unofficial version of the 4DVAR, a more complex cumulus scheme (Grell, 1993) and its adjoint was available attenuating the spin-up effect (Zus et al., 2008).

Even though applying a 4DVAR, it appears not sufficient to assimilate humidity data but to use as much observations as possible to constrain the model as much as possible. It appears also evident to switch to convection-resolving resolutions to avoid the necessity of a cumulus scheme for which it can be difficult to derive an adjoint code. The first attempt has been made in publication III with the WRF model.

The diurnal cycle of temperature is well simulated by all models but the 10 m wind speed is overestimated with a positive bias of 0.5 m/s which is, in case of the MM5, caused by the MRF PBL scheme and is also (but weakly) observed in the successor scheme YSU.

5.3 Schwitalla et al. (2011)

In publication III, the impact on the forecast quality of the assimilation of conventional and GPS-ZTD observations was investigated with the WRF model for COPS IOP 9c (July 20, 2007). For this purpose different experiments were carried out to investigate the influence of assimilating different types of observations.

A high-resolution, convection resolving model set-up with a horizontal resolution of 3600 m covering whole central Europe was designed (see the dashed rectangle in Figure 8).

The major reason for this large domain was that during D-PHASE several nested models had problems to simulate frontal or pre-frontal precipitation due to the fact that the front is disrupted at the boundary between the coarse outer and finer inner domain.

In such a large domain, the whole synoptic situation can completely develop inwardly of the domain avoiding inconsistencies between different physics selections responsible for destroying the synoptic features. Another reason was the observed dry bias in the precipitation patterns of

the high-resolution D-PHASE models due to the switch-off of the cumulus scheme in the nested domain (Bauer et al., 2011a).

it is also evident that as many high quality observations as possible should be assimilated simultaneously (independent whether 3DVAR or 4DVAR methodology is applied) to provide an optimal constraint of the model during the assimilation process. Due to a noticeable bias, GPS data from France were reprocessed at the GFZ Potsdam with their EPOS software so that in total about 330–360 GPS measurements were available every 15 minutes. A critical point was the estimation of the observation error for ZTD measurements. As there is hardly any expertise, we set the error to 2–3 mm which corresponds to about 1.5 mm IWV (Zus, 2010).

The first conclusion which can be drawn is that it is necessary to assimilate upper air (humidity) observations as otherwise the surface is uncoupled from the 3D atmosphere. When assimilating GPS-ZTD observations in addition to conventional observations, small improvements in the precipitation scores and its diurnal cycle were observed.

Also with data assimilation the simulated precipitation amounts are too high (bias $\sim$20%) which also does not change when applying the Thompson one-moment microphysics scheme. Both microphysics schemes show a large amount of integrated graupel which immediately melts into rain when the ambient temperature is greater than 0 °C.

We also discovered a too strong mixing in the boundary layer resulting in a too large water vapor content before/after sunrise/sunset similar to the behavior in publication I apparently caused by too large boundary layer heights and too strong mixing during night, respectively. In the meantime, this problem was confirmed by the developer of the boundary layer scheme (Hong, personal communication) and a bug fix will be tested soon.

Compared to other operational D-PHASE models, WRF was the only one predicting the redevelopment of the front in the lee of the black forest while e.g. COSMO-EU only produced light widespread precipitation. This appears to be the result of inaccurate initial conditions as WRF was driven by ECMWF data. COSMO-EU was driven by GME and applies a FDDA scheme to improve the initial conditions which seems to be not as good as the ECMWF analysis.

6 Summary and outlook

In this thesis, several different topics have been addressed. In the first part, a high-resolution setup of the mesoscale model MM5 was investigated to determine a suitable physics configuration to run an operational forecast system during COPS in 2007 and to compare a 7 km resolution MM5 model with the operational version of the Lokal Modell (LM, predecessor of the COSMO-EU) of DWD.

Here, 13 representative convective cases of 2005 have been selected and divided into three different categories depending on the synoptic situation. Independent of the synoptic situation, both 7 km resolution models show a systematic positive precipitation bias on the windward side of the Vosges Mountains and the Black Forest. The precipitation was initiated too early likely due to the application of a convection parameterization and comparatively strong upward motion at 850 hpa. For the high resolution setup, it was found that the shape of the diurnal cycle of precipitation follows the observation except that the precipitation maximum was simulated too late.

Additionally, a poor man's physics ensemble was created for four of the 13 convective cases. Here, the PBL schemes and the land-surface model have been exchanged. It was found that the 5-layer soil model produces more realistic latent and sensible heat fluxes compared to the NOAH-LSM leading to a better representation of the diurnal cycle of the temperature where the simulations with the NOAH-LSM show a notable cold bias.

Comparison with soundings and GPS integrated water vapor revealed a positive bias of the model with the different physics configurations indicating a too moist boundary layer. As this was state of the art at this time, the combination of the 5-layer soil model in combination with the MRF-PBL scheme was chosen for the operational forecasting system for COPS and D-PHASE with a slightly different nesting setup compared to the original study domain described in Schwitalla et al. (2008).

During D-PHASE, an operational forecasting system, providing two forecasts per day, was set up with one forecast driven by the ECMWF operational model level analysis only. For the second forecast, an additional 4DVAR applying GPS STD data (Zus et al., 2008) was performed to especially improve the water vapor distribution in the outermost domain. The investigation for the COPS period revealed an improvement of the water vapor field during the first six hours of the forecast. With the assimilation, the precipitation spin-up is reduced compared to the control experiment and the mean diurnal cycle fits better to the observation.

It appears that the inconsistent model physics (especially the poor convection parameterization) led to this "improvement" of the precipitation spin-up. In general, both forecasts tend to underestimate precipitation during the day probably due to the design of the domains. The innermost domain with a horizontal resolution of 2 km appears to be too small to capture precipitation events which are moving inside the domain. This behavior was also seen in several other models which were operated during D-PHASE.

As the development of the MM5 model was suspended before the COPS experiment, the WRF model system was introduced as a successor of MM5 including the preparation of the corresponding infrastructure for case study experiments. It incorporates a new dynamical core based on the Runge-Kutta scheme and more enhanced physics options. The NOAH-LSM is now a unified

version which is also applied at the National Centers for Environmental Prediction (NCEP) in its operational forecast system. A major step was the introduction of 2-moment cloud microphysics schemes allowing a better representation of microphysical processes.

With the WRF model version 3.1 an assimilation experiment for the COPS IOP 9c was carried out. The model set up was chosen in a way that only a single domain with a horizontal resolution of 3.6 km covering whole central Europe was applied. When simulating such a large computational demanding area, normally a nest is placed inside the coarser domain covering the region of interest. As this would imply that the convection parameterization is switched off in the inner nest, undesirable effects like destruction of frontal systems can occur (see e.g. Bauer et al., 2011a). Therefore this large domain with 550∗550 horizontal grid points was chosen.

Four different experiments have been conducted with this high resolution to determine the influence of assimilating different observations including GPS ZTD data on the analysis and forecast. This is unique in combination with the WRF model over Europe. Even the experiment with the assimilation of surface only data shows an improvement compared to the simulation without data assimilation. Its performance clearly improves with the assimilation of additional upper air observations, especially soundings. When assimilating additional GPS ZTD data, the vertical distribution of water vapor is clearly improved during the first 12 h of the forecast, but the impact on QPF is not clearly visible. One reason can be that from $\sim$ 320 observations only $\sim$ 100 stations are left for the assimilation partly due to the assumption of a too small observation error. This error determines the weight of the observations during the assimilation process and will be adjusted for future experiments.

What still remains is a significant overestimation of precipitation although the IWV content is on average only slightly higher as compared to GPS observations. The 2 m dew point shows a strong increase/decrease on the morning/evening due to a too strong mixing in the PBL during this time. In the meantime, an improvement of the YSU scheme is available which should prevent this unrealistic behavior in future case study experiments. In addition, a closer look into the energy partitioning in the LSM has to be made. Unfortunately there are currently no area-wide flux measurements available for the verification of the NOAH-LSM.

For the operational 4DVAR system in publication II, the 4DVAR was performed at the initial time step with a crude matrix $\mathbf{B}$ which was basically the ECMWF analysis on a similar resolution. As this can result in an additional model spin-up (publication II), the next step is to perform the assimilation a few hours after the initialization time so that the model is balanced (which is usually determined by the 3 h or 6 h surface pressure tendency). This was done in publication III in a simple way where the assimilation was performed 6 h after the initialization time so that the model has enough time to adjust the dynamics and moisture fields as well as hydrometeor fields.

As this may be not enough, a recently developed Rapid Update Cycle (RUC) will be applied in future experiments with an assimilation cycle of 3 h. It is expected that if the RUC is performed for 12–24 h, this will lead to a better dynamical balance despite the distortion due to the assimilation. The possible imbalance in the new initial field can be partially reduced by the application of a digital filter removing noise with a frequency higher than a certain cutoff frequency. This can become especially important for very short-range forecasts.

It is also evident from publications II and III, that one has to select a model domain which covers most of the prevailing synoptic features of the region of interest. Ideally, if nested into a

global model, the selected domain has a resolution which does not require the application of a convection scheme.

As stated earlier, the 3DVAR does not incorporate the model physics and can be regarded as a very sophisticated interpolation method. In contrast to a 4DVAR, a 3DVAR, although performed in a RUC, suffers from a static background matrix **B** which is only a climatological estimate of background errors independent of which method is applied to calculate the variances and covariances for **B**.

This can be resolved by the application of a 4DVAR which also starts from a static **B** but due to the incorporated model physics, the assimilation and the new analysis become flow dependent. The 4DVAR additionally requires the adjoint of the (full) forecast model which is, depending on the parameterization, impossible and thus simplified physics have to be used. The additional amount of computing time is not negligible especially for large grid sizes as for each minimization of the cost function the full forward and adjoint of the forecast model have to be run during the assimilation window (typically 6 h). Another possibility would be the derivation of **B** from a WRF ensemble which is based on the ECMWF ensemble members with the method described by Fisher (2003).

To overcome this 4DVAR specific problem, more and more Ensemble Kalman Filters are applied. They do not require the adjoint of a forecast model but the choice of the ensemble or the ensemble initialization, respectively, can be critical. Todays global ensemble forecast systems consist of 20–50 members (e.g. GFS and ECMWF). This is sufficient on the global scale but can be not enough to represent the synoptic situation on a finer grid. For a finer scale, $\sim$ 100 members could be necessary, but this would be prohibitively expensive in terms of computing power. It would however, have the advantage of providing situation dependent error covariances ("error of the day") from the ensemble perturbations.

Therefore the hybrid approach with e.g an ETKF is more and more considered as a compromise between the application of a 4DVAR and an EnKF. Here $\sim$ 20 members could be sufficient and the update of the ensemble perturbations is performed by applying the 3DVAR on the ensemble mean. This combines the advantages of reduced computational costs for a 3DVAR and a flow dependent background error matrix **B** and is subject for future research of high-resolution NWP.

As ensemble forecasting is subject for future research, the focus was set on the development of a RUC with a convection-permitting resolution of the WRF model based on the 3DVAR. The important point is, that this system includes GPS-ZTD data as well as radar radial velocities, reflectivities and satellite radiances. This system can be used in near-real time and is, to my knowledge, unique in Europe.

Another future task will be the application of polarization radar data, which allows the use of more hydrometeor distribution information, e.g. whether rain or hail is present, and thus allows a comparison with the applied microphysics scheme. If polarization data is employed in the assimilation process (independent whether 3/4DVAR or Ensemble Kalman Filter), it is beneficial if a 2-moment cloud microphysics scheme is selected because less assumptions for the forward operator are necessary.

A Practical implementation of a RUC system for the WRF model

```
##!/bin/ksh

###############################################################################
# This script automatizes the whole process necessary to perform forecasts with
# the mesoscale numerical weather prediction model WRF from the preparation of
# the model domain and initial fields to data assimilation, model forecast, and
# last but not least to post processing, visualization and data transfer.
###############################################################################

#############################################################
#
# no use without permission!!
#
#############################################################

# Author: Thomas Schwitalla
#         Institute of Physics and Meteorology
#         University of Hohenheim
#         Garbenstrasse 30
#         70599 Stuttgart
#         Germany
#         Email: thomas.schwitalla@uni-hohenheim.de

#############################################################################
# 1. Enable/disable things to be done
#############################################################################

MARS='false' #  get forcing data
GEOGDATA='false' # get geographical static data, necessary for creating domain
GEOGRID='false' # Run geogrid, creates domain
UNGRIB='false' # run ungrib
METGRID='false' # run metgrid
FORECAST='true' # you want to make a forecast

###############
#first reset everything
###############

WRF_REAL='false'
RUN_WRF_FREE='false'
VAR='false'
CYCLING='false'
RUN_WRF='false'
ARWPOST='false'
TRANSPORT='false'
DELETE='false'
READ_WRF='false'
NCL_VAR='false'
NCL_STAT_VAR='false'
OBSPROC='false'
JDC='false'
TRANSPORT_statistics='false'
```

```
TRANSPORT_model='false'
PLOT_RADIANCE='false'
USE_RADIANCE='false'
UPDATE_BDY_VAR='false'

####################
#now define what you want
####################

if [ $FORECAST = 'true' ]
then

fi

#############################################################################
#Define EXPERIMENT name which is attached to the date string as directory name
#############################################################################

#############################################################################
# 2. Define variables used for the call of the MARS retrieval scripts
#############################################################################

# Start and end of forecast. By setting the values separately it is later possible
# to get them automatically set with the MSJ system

DATA_SOURCE='forecast'  ### if you need boundaries from analysis or forecast data
FORECAST_LENGTH=            # Length of the forecast (hours)
LAST_DAY=                 # Last day of forecast if not a monthly statistics

############################
# adjust cycle length as you need it!!!
############################

######################
#special for JDC data for 2007
# if JDC is used for other years, the script will stop
######################

####################
#something may go wrong
####################

#############################################################################

#############################################################################
#
# create start and stop dates for the WRF forecast and assimilation
#
#############################################################################

### Now loop over days to automatise the work######

while [ ${START_DAY} -le ${LAST_DAY} ] ;
do

# Initial date and initial time needed by the MARS retrieval scripts
```

```
# Calculate STOP DATE. This was necessary, when analysis and not forecast data
    should
# be used to force WRF

# Stop date needed when analysis data shall be used for the forcing

###########################################################################
#
# end of calculation of date strings
#
###########################################################################
# 3. Define variables used to fill the namelist file "namelist.wps"
#    used by the WRF pre-processing system.
#
#******************* READ THIS CAREFULLY BEFORE YOU START!!!!!*****************
#
###########################################################################

MAX_DOM=1                              # Number of WRF model domains

###############################################
#different intervals for forecast and analysis data
###############################################

if [ ${DATA_SOURCE} = 'forecast' ]
then
INTERVAL_SECONDS=10800                 # Interval between two forecast timesteps
else
INTERVAL_SECONDS=21600                 # Interval between two forecast timesteps
fi

#########################################################
#domain settings for namelist.wps
#########################################################

PARENT_ID='1, 1, 2, 3'                 # ID of parent domain
PARENT_GRID_RATIO='1,  3,  3,  3' # grid ratio between domains
I_PARENT_START='1, 135, 41, 172 ' # X-coordinate of the lower left corner
J_PARENT_START='1, 82, 141, 132'  # Y-coordinate of the lower left corner
S_WE='1, 1, 1, 1'                      # Set to '1' for each domain
E_WE='550, 250, 367, 160'              # Number of east-west grid points
S_SN='1, 1, 1, 1'                      # Set to '1' for each domain
E_SN='550, 271, 271, 160'              # Number of south-north grid points

GEOG_DATA_RES='"modis_30s+30s", "modis_30s+30s", "modis_30s+30s", "modis_30s+30s"'
DX=3600                               # Grid resolution (m) of the outermost domain
DY=3600                               # Grid resolution (m) of the outermost domain
MAP_PROJ=lambert                      # Map projection
REF_LON=7.                            # Center point latitude of the outermost domain
REF_LAT=48.9                          # Center point longitude of the outermost domain
TRUELAT1=30.0                         # Southern latitude for Lambert projection
TRUELAT2=60.0                         # Northern latitude for Lambert projection
STAND_LON=7.                          # Longitude parallel to Y-axis

###########################################################################
```

```
# 4.) Define variables to fill the file "namelist.input" which controls
#the execution of the  programms "real.exe" and "wrf.exe".
#     IMPORTANT: The variables already defined for WPS are recycled
###############################################################################

if [ ${DATA_SOURCE} = 'forecast' ]
then
ANALYSIS_INTERVAL=3
else
ANALYSIS_INTERVAL=6
fi

#################################################################
# Time control options for the namelist "namelist.input"
#################################################################

#########################################################
# Domain options for the namelist "namelist.input" (not defined above for wps)
#########################################################

#########################################################
# Physics options (selection parameterizations schemes) for the "namelist.input"
#########################################################

#############################################
# for I/O quilting, useful for large domains
#############################################

#####################################################################
# DEFINE OBSERVATIONS you want to use in 3DVAR
#####################################################################

#####################################################################
#
# Warning!!!!!
# set correct location for GPS data file before using GPS-ZTD data!
# max_error_gpsxx was set to 5.0!
# set max_errors for other variables later!
#
#####################################################################

USE_GPSPWOBS='false' ### cannot used together with ZTD!
USE_GPSZTDOBS='true'
USE_BUOYOBS='true'
USE_SYNOPOBS='true'
USE_METAROBS='true'
USE_SOUNDOBS='true'
USE_AIREPOBS='true'
USE_SHIPSOBS='true'
USE_PROFILEROBS='true'
USE_RADAROBS1='true' ## scroll down to adjust first time when assimilating
USE_RADAR_RV='true'
USE_RADAR_RF='true'
CALC_STD='false'
USE_QSCATOBS='true'
USE_GEOAMVOBS='true'
```

```
USE_SSMITBOBS='false' ### data are available, but not tested

###################################
#for radiances
###################################

USE_TWO_RADIANCE_FILES='true'
if [ ${USE_RADIANCE} = 'true' ]
then
 USE_HIRS3OBS='true'
 USE_HIRS4OBS='false'
 USE_MHSOBS='true'
 USE_AMSUAOBS='true'
 USE_AMSUBOBS='true'
 USE_AIRSOBS='true'
 USE_EOS_AMSUAOBS='true'
 USE_HSBOBS='false'
 USE_SSMISOBS='false'
fi
#############################################################################
######################## From here no changes are necessary !! ##############
#############################################################################
#############################################################################
# 5. Create working directory and copy necessary files from the ecfs
#############################################################################

#################
# Create data directory and  working directory
#################

# Test whether the directories exist. Create if not the case. This will be changed
    in
# a way, that an existing directory is automatically deleted and recreated when
# everything is working.

cd ${WRKDIR}
#############################################################################
# 6. Include start date, end date into the scripts that retreive observations for
#the data assiimilation system and analysis fields from the MARS archive.
#############################################################################

if [ ${MARS} = 'true' ]
then

fi ## finished with MARS

#############################################################################
# 7. Pre-processing. Here, the three pre-processing steps to create the initial
# fields for "real.exe" are done.
#############################################################################

# Set up namelist "namelist.wps" with the information provided above.

cat > namelist.wps <<EOF

EOF
```

```
#############################################################################
# 8.) Create the model domains (using information from "namelist.wps").
#This step is only necessary when a new domain shall be created.
#############################################################################

if [ ${GEOGRID} = 'true' ]
then

fi ## finished geogrid

### check whether geogrid file is already available

#############################################################################
# 9.) Decode analysis grib files (using information from "namelist.wps")
#############################################################################

if [ ${UNGRIB} = 'true' ]
then

fi ## finished ungrib

#############################################################################
# 10.) Create meteorological initial fields for WRF
#############################################################################

if [ ${METGRID} = 'true' ]
then

fi ## finished metgrid

#############################################################################
# 12.) Create namelist "namelist.input" used by the final preprocessing
#step and the WRF forecast
#############################################################################

## Change to working directory
cd ${WRKDIR}

## Create namelist
cat > namelist.input <<EOF

EOF

#############################################################################
#################### WRF run real once at the beginning ####################
#############################################################################

cd ${WRKDIR}

if [ ${WRF_REAL} = 'true' ]
then

fi ### finished REAL

#############################################################################
```

```
###################END of preprocessing#############################################
##################################################################################

#######################################
#start free forecast before assimilation
#######################################

if [ ${RUN_WRF_FREE} = 'true' ]
then

fi ### WRF_RUN_FREE

#########################################################################
###### multiple cycling loop#############################################
#########################################################################

I=00
while [ ${I} -lt ${MULTIPLE} ]
do

#######################################
#### to prevent the model blowing up with radar data
#### check when radar data assimilation starts!!
#######################################

 if [ $I -le 0 ] # 0 to not assimilate radar data at the first cycle step
 then
 USE_RADAROBS='false'
 echo "Do not want to use radarobs"
 echo "Radar-Obs= " ${USE_RADAROBS}
 else
 USE_RADAROBS=${USE_RADAROBS1}
 echo "you want radarobs"
 echo "Radar-Obs= " ${USE_RADAROBS}
 fi
## fi blowing prevent

# some counters
J=`expr $I + 1`
K=`expr $J + 1`

##############################
#calculation of datestring for RUC
##############################

echo CYCLE_YEAR $CYCLE_YEAR
echo CYCLE_MONTH $CYCLE_MONTH
echo CYCLE_DAY $CYCLE_DAY
echo CYCLE_HOUR  $CYCLE_HOUR

#################################################################################
# export date for assimilation statistics later on
#################################################

NCL_VAR_DATE=${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}${CYCLE_HOUR}
```

```
#############################################################################
# OBSPROC FOR 3DVAR data assimilation
#############################################################################

#############################################################################
# create namelist.3dvar_obs with variables declared above
#############################################################################

if [ $OBSPROC = 'true' ]
then

### set timesteps necessary for backward times

##########################################
#
#Set correct directory for GPS DATA!!!!!
#
##########################################

##########################################
#
# for conventional data
#
##########################################
 if test -e "${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}_bufr.asc"
  then
  echo "conventional obs data already there"
  else
  ecp -o ec:OBS_DATA/${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}_bufr.asc.gz .
  gunzip -f ${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}_bufr.asc.gz
 fi

 OBSFILE=${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}${CYCLE_HOUR}_bufr.asc

############################
# copying obs files finished
############################

### delete files that are not necessary any more

### copy all necessary binary files for OBSPROC from ECFS

### calculate resolution in km from DX

### Create namelist for obsproc####

cat > namelist.obsproc  <<EOF

EOF

# execute obsproc
  ./obsproc.exe
fi # obsproc

#############################################################################
#############################################################################
```

```
# BEGIN OF 3DVAR
###############################################################################
###############################################################################

if [ ${VAR} = 'true' ]  # you want to run a 3DVAR first
then

## copy or link the necessary files to 3DVARDIR

################################
#necessary for version earlier than 3.3
################################
 cd $VARDIR

### copy or link the necessary files to 3DVARDIR once

### link and copy the necessary files for 3DVAR

   cp $WRKDIR/wrf_3dvar_input_d01_${CYCLE_YEAR}-${CYCLE_MONTH}-${CYCLE_DAY}_${
       CYCLE_HOUR}:00:00   fg
   ln -sf $VARDIR/obs_gts_${CYCLE_YEAR}-${CYCLE_MONTH}-${CYCLE_DAY}_${CYCLE_HOUR
       }:00:00.3DVAR ob.ascii

# use of SSMI/T data

## copy background matrix from ECFS
   ecp -o ec:be_3600m_550_550_50.dat be.dat

####################################
# create namelist.input for 3DVAR
####################################

## create namelist for wrfvar
 cat > namelist.input <<EOF

EOF

########################################
#need a new wrfbdy file for 3DVAR and forecast afterwards
#update low boundary before 3DVAR otherwise DFI will not work after assimilation
########################################
if [ ${UPDATE_BDY_VAR} = 'true' ]
then

cat > namelist.input_inside_ruc_for_bdy <<EOF

EOF

### start real.exe at first

#copy new wrfbdy to 3DVAR directory
cp   wrfbdy_d01 $VARDIR

fi ## update bdy before 3DVAR

#####################################################
```

```
# create parame.in for updating boundary conditions before 3DVAR
#####################################################

cat > parame.in <<EOF

EOF
# run da_update_bc

# end of update low BC

## for safety remove some radiance files, otherwise plots can be strange...
rm -rf 01_oma* 01_inv* ${VARDIR}/diag/${NCL_VAR_DATE}

# for radar data
if [ ${USE_RADAROBS} = 'true' ]
then

if [ $? -ne 0 ]
then
echo "copying of radar data failed for date " ${NCL_VAR_DATE}
exit 1
fi
ln -sf ob.radar_all_${NCL_VAR_DATE}.asc ob.radar
fi
###### fi radar data

########################################
# some preparation steps for radiances
########################################

if [ ${USE_RADIANCE} = 'true' ]
then
 if test -d "rttov10"
 then
  echo "RTTOV already there"
 else
  echo "RTTOV has to be copied at first"
 fi
fi

# copy necessary atatic files once
if [[ ${USE_RADIANCE} = 'true' && $I -le 0 ]]
then

fi

if [ ${USE_RADIANCE} = 'true' ]
then

mkdir -p ${VARDIR}/diag/${NCL_VAR_DATE}

# check if radiation data is already available
 if  test -e "${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}1bamua.tar"
 then
  echo "AMSUA already there"
 else
```

```
  ecp -o ec:OBS_DATA/RADIANCES/1bamua.${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.tar.
      gz  .
  gunzip -f 1bamua.${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.tar.gz
  tar -xvf 1bamua.${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.tar
 fi

## remove file links first to prevent double observations in assimilation!!
 rm -rf *01.bufr *02.bufr

 if [ ${CYCLE_HOUR} = 00 ]
 then
  ln -sf ${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.1bamua/*00z* amsua01.bufr
  ln -sf ${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.1bamub/*00z* amsub01.bufr
  ln -sf ${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.1bhrs3/*00z* hirs301.bufr
  ln -sf ${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}.1bmhs/*00z* mhs01.bufr
  ln -sf airsobs.${CYCLE_YEAR}${CYCLE_MONTH}${CYCLE_DAY}/*00z*airsev* airs01.bufr

# test if files are there for the next day
   if  test -e "gdasairsobs.${CYCLE_YEAR2}${CYCLE_MONTH2}${CYCLE_DAY2}.tar"
   then
    echo "AIRS for day 2 already there"
   else
    ecp -o ec:OBS_DATA/RADIANCES/gdasairsobs.${CYCLE_YEAR2}${CYCLE_MONTH2}${
        CYCLE_DAY2}.tar.gz  .
    gunzip -f gdasairsobs.${CYCLE_YEAR2}${CYCLE_MONTH2}${CYCLE_DAY2}.tar.gz
    tar -xvf gdasairsobs.${CYCLE_YEAR2}${CYCLE_MONTH2}${CYCLE_DAY2}.tar
   fi
 fi

fi
###### fi radiances

#############################################
# now run WRF 3DVAR inside the RUC
#############################################

## remove certain files
 rm -f cost_fn grad_fn fort.21

# search for failure
 grep -i completed rsl.out.0000
 if [ $? != 0 ]
 then
 echo "3DVAR failed" > mail_message
 exit 1
 fi
### fi run 3DVAR

##############################
#variational bias correction for radiances
##############################
 if [ ${USE_RADIANCE} = 'true' ]
 then
 cp VARBC.in VARBC.in_${NCL_VAR_DATE}
 cp VARBC.out VARBC.out_${NCL_VAR_DATE}
 mv VARBC.out VARBC.in
```

```
fi

#backup files
cp radar_omb_oma_01 radar_omb_oma_01_${NCL_VAR_DATE}
cp gts_omb_oma_01 gts_omb_oma_01_${NCL_VAR_DATE}
cp filtered_obs_01 filtered_obs_01_${NCL_VAR_DATE}
cp wrfvar_output wrfvar_output_${NCL_VAR_DATE}.nc

##############################################
# update lateral boundary conditions after 3DVAR
##########################################

### create parame.in for updating lateral boundary conditions

cp -f  $WRKDIR/wrfinput_d01 .

cat > parame.in <<EOF

EOF

# run update of BC conditions
 grep successfully update_lat_bdy_${NCL_VAR_DATE}.txt
 if [ $? -ne  0 ]
 then
 echo "update bc after 3DVAR failed, exiting"
 echo "update bc failed" > mail_message
 /usr/bin/mail -s "update bc after" thomas.schwitalla@uni-hohenheim.de <
     mail_message
 exit 1
 fi

fi # end of 3DVAR

#############################################################################
#############################################################################
# END OF 3DVAR
#############################################################################
#############################################################################

#####################################
### make assimilation statistic plots using default NCL scripts.....
#############################################

if [ ${NCL_VAR} = 'true' ]
then

#######################
 if [ ${NCL_STAT_VAR} = 'true' ]
 then

# make some horizontal plots

 if [ ${PLOT_HORIZONTAL} = 'true' ]
 then
   echo "finished horizontal plots at"
```

```
   date
 fi

# for plotting radiance statistics
 if [[ ${USE_RADIANCE} = 'true' && ${PLOT_RADIANCE} = 'true' ]]
 then

 fi

#########################################
# make an archive
#########################################

#transport statistics file to UHOH
 if [ $TRANSPORT_statistics = 'true' ]
 then

 fi #### transport

fi # finished statistics
#####################################
### end of NCL assimilation statistics
#####################################

## change back to WRKDIR
 cd ${WRKDIR}

###########################################
## after the 3dvar is finished start forecast from a wrfinput file
###########################################

###############################################
## Create namelist for warm start after a assimilataion
###############################################

cat > namelist.input_warm <<EOF

EOF

#################################################################
#
# starts forecast after assimilation in RUC mode
#
#################################################################

if [[ ${RUN_WRF} = 'true' &&  ${I} -lt $(expr ${MULTIPLE} - 1)  ]]
then

echo CYCLE_NUMBER  $I "+1"

#### finished WRF within RUC

fi # finished WRF forecast inside RUC

I=`expr $I + 1`
```

```
done ################ end multiple cycling loop#################################
#########################################################################################
#########################################################################################

#################################################################
# run free forecast after assimilation
#################################################################

if [[ $RUN_WRF_FINAL = true && $CYCLING = true ]]
then

 cd $WRKDIR

# Create namelist.input

cat > namelist.input <<EOF

EOF

### start real.exe at first for updating wrfbdy

 test -s wrfbdy_d01

### create parame.in for updating lateral boundary conditions
cat > parame.in <<EOF

EOF

# run update of BC conditions


#########################
# submit WRF
#########################

  #check if the job is still running

fi # finished WRF final

#################################
# Change to next day
#################################

  START_DAY=`expr $START_DAY + 1`

  if [ ${START_DAY} -lt 10 ]
    then
    START_DAY=0${START_DAY}
  fi

done # loop over days
###### end of script
```

References

Baldauf, M., J. Förstner, S. Klink, T. Reinhardt, C. Schraff, A. Seifert and K. Stephan, 2009: Kurze Beschreibung des Lokal-Modells Kürzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD. Technical report, DWD, Offenbach, Germany.

Barker, D. M., Y.-R. Guo, W. Huang and A. Bourgeois, 2003: The three-dimensional variational (3DVAR) data assimilation system for the use with MM5. Technical Report TN-453+STR, NCAR, Boulder/CO. available at http://www.mmm.ucar.edu/mm53dvar/docs/3DVARTechDoc.pdf.

Barker, D. M., W. Huang, G. Y.-R. and Q. N. Xiao, 2004: A three-dimensional variational (3DVAR) data assimilation system for MM5. Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.

Bauer, H.-S., T. Weusthoff, M. Dorninger, V. Wulfmeyer, T. Schwitalla, T. Gorgas, M. Arpagaus and K. Warrach-Sagi, 2011a: Predictive skill of a subset of models participating in D-PHASE in the COPS region. *Q. J. R. Meteorol. Soc.*, **137**, S1, 287–305.

Bauer, H.-S., F. Zus, V. Wulfmeyer, T. Schwitalla, M. Grzeschik, G. Dick and M. Bender, 2011b: Operational assimilation of GPS slant path delay measurements into the MM5 4DVAR system. *Tellus A*, **63**, 263–282.

Bech, J., B. Codina, J. Lorente and D. Bebbington, 2003: The Sensitivity of Single Polarization Weather Radar Beam Blockage Correction to Variability in the Vertical Refractivity Gradient. *J. Atmos. Oceanic Technol.*, **20**, 845–855.

Bender, M., G. Dick, T. Schmidt, S. Song, G. Gendt, M. Ge and M. Rothacher, 2008: Validation of GPS slant delays using water vapour radiometers and weather models. *Meterol. Z.*, **17**, 807–812.

Bender, M., G. Dick, J. Wickert, M. Ramatschi, M. Ge, G. Gendt, A. Rothacher, M. Raabe and G. Tetzlaff, 2009: Estimates of the information provided by GPS slant data observed in Germany regarding tomographic applications. *J. Geophys. Res.*, **114**, D06303, 11pp.

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Wea. Rev.*, **128**, 644–667.

Bevis, M., S. Businger, T. A. Herring, R. A. Anthes, C. Rocken and R. H. Ware, 1994: GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. *J. Appl. Meteor.*, **33**, 379–386.

Bevis, M., S. Businger, T. A. Herring, R. A. Anthes and R. H. Ware, 1992: GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System. *J. Geophys. Res.*, **97**, 15787–15801.

Bishop, C., E. B. J. and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Bouttier, F., 2007: AROME, avenir de la prévision régionale. *La Météorologie*, **58**, 12–20.

Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods. Technical report, Meteorological Training Course Lecture Series, ECMWF, European Center for Medium Range Weather Forecasting, Reading, UK. available online on: `http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/index.html`.

Caumont, O., V. Ducroq, E. Wattrelot, G. Jaubert and S. Pradier-Vabre, 2010: 1D+3DVar assimilation of radar reflectivity data: a proof of concept. *Tellus A*, **62**, 173–187.

Chen, Y., F. Weng, Y. Han and Q. Liu, 2008: Validation of the community radiative transfer model (CRTM) by using CloudSat Data. *J. Geophys.Res.*, **113**, 15pp.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, **123**, 2449–2461.

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, A. Hamruth, A. Hollingsworth, F. Rabier and M. Fisher, 1998: The ECMWF implementation of three dimensional variational assimilation (3DVAR). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124**, 1783–1807.

Dorninger, M., T. Gorgas, T. Schwitalla, M. Arpagaus, M. Rotach and V. Wulfmeyer, 2009: Joint D-PHASE-COPS data set (JDC data set). Technical report, University of Vienna/University of Hohenheim/Meteo Swiss.

Douville, H., P. Viterbo, J. F. Mahfouf and A. C. M. Beljars, 2001: Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733–1756.

Doviak, R. and D. Zrnić, 1993: *Doppler Radar and Weather Observations.* Academic Press,, 2nd edition.

DTC, 2009: MET: Version 2.0 Model Evaluation Tools User's Guide. `http://www.dtcenter.org/met/users/docs/overview.php`.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grummann., V. Koren, G. Gayno and J. D. Tarpley, 2003: Implementation of NOAH land surface model advances in the National Centers for Environmental Prediction operational Mesoscale Eta Model. *J. Geophys. Res.*, **108**, D22.

Ereesma, R. and H. Järvinen, 2006: An observation operator for ground-based GPS slant delays. *Tellus*, **58A**, 131–140.

EUMETSAT, 2009: MSG Meteorological Products Extraction Facility Algorithm Specification Document. Technical report, EUMETSAT, Darmstadt, Germany.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Fisher, M., 2003: Background error covariance modeling. In: *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pp. 45–63. ECMWF.

Gallus, W. A. J. and M. Pfeiffer, 2008: Intercomparison of simulations using 5 WRF microphysical schemes with dual-Polarization data for a German squall line. *Adv. Geosc.*, **16**, 109–116.

Gendt, G., G. Dick, C. Reigber, M. Tomassini, Y. Liu and M. Ramatschi, 2004: Near Real Time GPS Water Vapor Monitoring for Numerical Weather Prediction in Germany. *J. Meteor. Soc. Jpn.*, **82**, 1B, 361–370.

Golub, G. H. and C. F. Van Loan, 1996: *Matrix Computations.* Johns Hopkins University Press, 3rd edition.

Gourley, J. J. and C. Calvert, 2003: Automated Detection of the Bright Band Using WSR-88D Data. *Wea. Forecasting*, **18**, 585–599.

Grell, G., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764–787.

Grell, G. A., J. Dudhia and R. R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR technical Note TN-398+STR, NCAR, Boulder/CO. 122pp.

Haase, G., 2002: A physical initialization algorithm for non-hydrostatic weather prediction models using radar derived rain rates. Ph.D. thesis, University of Bonn.

Hager, W. W., 1989: Updating the inverse of a matrix. *SIAM Review*, **31**, 221–239.

Helmert, K., T. Hengstebeck and J. Seltmann, 2008: DWD s operational tool to enhance radar data quality. Extended abstract. ERAD 2008, Helsinki. http://erad2008.fmi.fi/proceedings/extended/erad2008-0083-extended.pdf.

Hengstenbeck, T., K. Helmert and J. Seltmann, 2010: RadarQS - a standard quality control software for radar data at DWD. Extended abstract. ERAD 2010, Sibiu. `http://www.erad2010.org/pdf/POSTER/03_Quality/05_ERAD2010_0318_modificat.pdf`.

Hofmann-Wellenhof, B., H. Lichtenegger and J. Collins, 1997: *GPS: Theory and Practice.* Springer, 4th edition.

Holleman, I. and H. Beekhuis, 2003: Analysis and Correction of Dual PRF Velocity Data. *J. Atmos. Oceanic Technol.*, **20**, 443–453.

Hollinger, J., 1989: DMSP Special Sensor Microwave/Imager Calibration/Validation. Vol. 1. Final rep., Naval Research Laboratory. 153pp.

Hong, S.-Y., Y. Noh and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,*, **134**, 2318–2341.

Huang, X. Y., Q. Xiao, D. M. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson, J. Brey, Y. Chen, Z. Ma, J. Dudhia, Y. Guo, X. Zhang, D. J. Won, H. C. Lin and Y. H. Kuo, 2009: Four-Dimensional Variational Data Assimilation for WRF: Formulation and Preliminary Results. *Mon. Wea. Rev.*, **137**, 299–314.

Jung, Y., G. Zhang and M. Xue, 2008: Assimilation of Simulated Polarimetric Radar Data for a Convective Storm Using the Ensemble Kalman Filter. Part I: Observation Operators for Reflectivity and Polarimetric Variables. *Mon. Wea. Rev.*, **136**, 2228–2245.

Kain, J. S., 2004: The Kain-Fritsch Convective Parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181.

Kain, J. S., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber and K. W. Thomas, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952.

Lemon, L. R., 1998: The Radar "Three-Body Scatter Spike": An Operational Large-Hail Signature. *Wea. Forecasting*, **13**, 327–340.

Lindskog, M., K. Salonen, H. Järvinen and D. B. Michelson, 2004: Doppler Radar Wind Data Assimilation with HIRLAM 3DVAR. *Mon. Wea. Rev.*, **132**, 1081–1092.

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

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne and F. W. Saunders, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **126**, 2991–3012.

Lynch, P., 2006: *The emerge of numerical weather prediction: Richardson's Dream.* Cambridge University Press.

Majewski, D., H. Frank, D. Liermann and B. Ritter, 2009: Kurze Beschreibung des Global-Modells GME (40 km/L40) und seiner Datenbanken auf dem Datenserver des DWD. Technical report, DWD, Offenbach, Germany.

Marshall, J. S. and W. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165–166.

Mellor, G. L. and T. Yamada, 1974: A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers. *J. Atmos. Sci.*, **31**, 1791–1806.

Milan, M., 2009: Physical Initialisation of Precipitation in a Mesoscale Numerical Weather Forecast Model. Ph.D. thesis, University of Bonn.

Milan, M., V. Venema, D. Schüttemeyer and C. Simmer, 2008: Assimilation of radar and satellite data in mesoscale models: A physical initialization scheme. *Meteorol. Z.*, **17**, 6, 887–902.

Milbrandt, J. A. and M. K. Yau, 2005: A Multimoment Bulk Microphysics Parameterization. Part II: A Proposed Three-Moment Closure and Scheme Description. *J. Atmos. Sci.*, **62**, 3065–3081.

Mo, T. and Q. Liu, 2008: A study of AMSU-A measurement of brightness temperatures over the ocean. *J. Geophys. Res.*, **113**.

Montmerle, T. and C. Faccani, 2009: Mesoscale Assimilation of Radial Velocities from Doppler Radars in a Preoperational Framework. *Mon. Wea. Rev.*, **137**, 1939–1953.

Morrison, H., G. Thompson and V. Tatarskii, 2009: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Mon. Wea. Rev.*, **137**, 991–1007.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Centre's spectral statistical-interpolation analysis system. *Mon Wea. Rev.*, **120**, 1747–1763.

Pfeiffer, M., 2007: Evaluation of precipitation forecasts by polarimetric radar. Ph.D. thesis, LMU, Munich. 134pp. http://edoc.ub.uni-muenchen.de/7253/.

Pruppacher, H. R. and J. D. Klett, 1997: *Microphysics of clouds and precipitation.* Springer, Netherlands.

Reisner, J., R. Rasmussen and R. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Q. J. R. Meteor. Soc.*, **124**, 1071–1107.

Reitebuch, O., C. Lemmerz, E. Nagel, U. Paffrath, Y. Durand, M. Endemann, F. Fabre and M. Chaloupy, 2009: The Airborne Demonstrator for the Direct-Detection Doppler Wind Lidar ALADIN on ADM-Aeolus. Part I: Instrument Design and Comparison to Satellite Instrument. *J. Atmos. Oceanic Technol.*, **26**, 12, 2501–2515.

Richardson, L. F., 1922: *Atmospheric diffusion shown on a distance-neighbour graph.* Cambridge University Press.

Rotach, M. W., M. Arpagaus, M. Dorninger, C. Hegg, A. Montani, R. Ranzi, F. Bouttier, A. Buzzi, G. Frustaci, K. Mylne, E. Richard, A. Rossa, C. Schär, M. Staudinger, H. Volkert, V. Wulfmeyer, P. Ambrosetti, F. Ament, C. Appenzeller, H.-S. Bauer, S. Davolio, M. Denhard, L. Fontannaz, J. Frick, F. Fundel, U. Germann, A. Hering, C. Keil, M. Liniger, C. Marsigli, Y. Seity, M. Stoll, A. Walser and M. Zappa, 2009: MAP D-PHASE: Real-time Demonstration of Weather Forecast Quality in the Alpine Region. *Bull. Amer. Meteorol. Soc.*, **90**, 1321–1336.

Saastamoinen, J., 1972: Contributions to the theory of atmospheric refraction. *Bull. Geodesique*, **V105-107**, 279–298,383–397,13–34.

Saunders, R. W., M. Matricardi and P. Brunel, 1999: An Improved Fast Radiative Transfer Model for Assimilation of Satellite Radiance Observations. *Q. J. R. Meteorol. Soc.*, **125**, 1407–1425.

Schraff, C., 1997: Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteorol. Atmos. Phys.*, **64**, 21–50.

Schraff, C., K. Stephan and S. Klink, 2006: Revised Latent Heat Nudging to cope with Prognostic Precipitation. *COSMO Newsletter*, , 6, 31–37.

Schwitalla, T., H.-S. Bauer, V. Wulfmeyer and F. Aoshima, 2011: High-resolution simulation over central Europe: assimilation experiments during COPS IOP 9c. *Q. J. R. Meteorol. Soc.*, **137**, S1, 156–175.

Schwitalla, T., H.-S. Bauer, V. Wulfmeyer and G. Zängl, 2008: Systematic errors of QPF in low-mountain regions as revealed by MM5 simulations. *Meterol. Z.*, **17**, 903–919.

Seifert, A. and K. D. Beheng, 2006: A two-moment cloud microphysics parameterization for mixedphase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, **92**, 45–66.

Seltmann, J., 2000: Clutter versus radar winds. *Phys. Chem. Earth*, **25**, 1173–1178.

Shewchuk, J. R., 1994: An introduction to the conjugate gradient method without the agonizing pain. Technical report, School of Computer Science, Pittsburgh, PA.

Singh, R., C. M. Kishtawal, P. K. Pal and P. C. Joshi, 2011: Assimilation of the multisatellite data into the WRF model for track and intensity simulation of the Indian Ocean tropical cyclones. *Meteorol. Atmos. Phys.*, **111**, 103–119.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. Gill, D. O. Barker, M. G. Duda, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note TN-475+STR, NCAR, Boulder/CO. available at `http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf`.

Stauffer, D. R. and N. L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor.*, **33**, 416–434.

Steinacker, R., M. Ratheiser, B. Bica, B. Chimani, M. Dorninger, W. Gepp, C. Lotteraner, S. Schneider and S. Tschannett, 2006: A mesoscale data analysis and downscaling method over complex terrain. *Mon. Wea. Rev.*, **134**, 2758–2771.

Stephan, K., S. Klink and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Q. J. R. Meteorol. Soc.*, **134**, 1315–1326.

Sun, J. and N. A. Crook, 1997: Dynamical and microphysical retrieval from doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.*, **54**, 1624–1661.

Tahanout, M., J. Parent Du Chatelet and C. Augros, 2009: New multiple-PRT scheme for Météo-France Doppler radar network to improve spectral moment estimation of weather radar signal and ground clutter filtering. 34th AMS Conference on Radar Meteorology.

Tiedtke, M., 1989: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Vedel, H. and X.-Y. Huang, 2004: Impact of ground based GPS Zenith Troposheric Delay data on precipitation forecast in Mediterranean France and Spain. *J. Meteorol. Soc. Japan*, **82**, 1B, 459–472.

Wang, X., D. M. Barker, C. Snyder and T. M. Hamill, 2008: A Hybrid ETKF-3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Mon. Wea. Rev.*, **136**, 5116–5131.

Wang, X., T. M. Hamill, J. S. Whitaker and C. H. Bishop, 2007: A comparison of hybrid ensemble transform kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Monthly Weather Review*, **135**, 1055–1076.

Wergen, W. and M. Buchhold, 2002: Datenassimilation fr das Globalmodell GME. *PROMET*, **27**, 150–155.

Wulfmeyer, V., A. Behrendt, C. Kottmeier, U. Corsmeier, C. Barthlott, G. C. Craig, M. Hagen, D. Althausen, F. Aoshima, M. Arpagaus, H.-S. Bauer, L. Bennett, A. Blyth, C. Brandau, C. Champollion, S. Crewell, G. Dick, P. DiGirolamo, M. Dorninger, Y. Dufournet, R. Eigenmann, R. Engelmann, C. Flamant, T. Foken, T. Gorgas, M. Grzeschik, J. Handwerker, C. Hauck, H. Höller, W. Junkermann, N. Kalthoff, C. Kiemle, S. Klink, M. König, L. Krauss, C. N. Long, F. Madonna, S. Mobbs, B. Neininger, S. Pal, G. Peters, G. Pigeon, E. Richard, M. W. Rotach, H. Russchenberg, T. Schwitalla, V. Smith, R. Steinacker, J. Trentmann, D. D. Turner, J. van Baelen, S. Vogt, H. Volker, T. Weckwerth, H. Wernli, A. Wieser and M. Wirth, 2011: The Convective and Orographically Induced Precipitation Study (COPS): The Scientific Strategy, the Field Phase, and research Highlights. *Q. J. R. Meteorol. Soc.*, **137**, S1, 3–30.

Xiao, Q., Y. H. Kuo, J. Sun, W.-C. Lee and D. Barker, 2007: An Approach of Radar Reflectivity Data Assimilation and Its Assessment with the Inland QPF of Typhoon Rusa (2002) at Landfall. *J. Appl. Met.*, **46**, 14–22.

Xiao, Q., Y.-H. Kuo, J. Sun, W.-C. Lee, E. Lim, Y.-R. Guo and D. M. Barker, 2005: Assimilation of Doppler Radar Observations with a Regional 3DVAR System: Impact of Doppler Velocities on Forecasts of a Heavy Rainfall Case. *J. Appl. Met.*, **44**, 768–788.

Xiao, Q., E. Lim, X. Zhang, J. Sun and Z. Liu, 2008: Doppler Radar Data Assimilation with WRF 3D-Var: IHOP Retrospective Studies. 9th WRF users workshop, Boulder/CO. http://www.mmm.ucar.edu/wrf/users/workshops/WS2008/abstracts/P5-05.pdf.

Xiao, Q. and J. Sun, 2007: Multiple-Radar Data Assimilation and Short-Range Quantitative Precipitation Forecasting of a Squall Line Observed during IHOP_2002. *Mon. Wea. Rev.*, **135**, 3381–3403.

Zhang, M., F. Zhang, X.-Y. Huang and X. Zhang, 2011: Intercomparison of an Ensemble Kalman Filter with Three- and Four-Dimensional Variational Data Assimilation Methods in a Limited-Area Model over the Month of June 2003. *Mon. Wea. Rev.*, **139**, 566–572.

Zus, F., 2010: Application of Global Positioning System slant path delay data for mesoscale modelverification and four-dimensional variational assimilation. Ph.D. thesis, University of Hohenheim. http://opus.ub.uni-hohenheim.de/volltexte/2011/516/pdf/Dissertation_FZUS.pdf.

Zus, F., M. Grzeschik, H.-S. Bauer, V. Wulfmeyer, G. Dick and M. Bender, 2008: Development and optimization of the IPM MM5 GPS slant path 4DVAR system. *Meterol. Z.*, **17**, 867–885.

Acknowledgements

First I would thank Prof. Dr. Volker Wulfmeyer for the opportunity to prepare my Diploma and my PhD thesis in his institute and his supervision over the last 3 years. If he did not gave my the opportunity in his institute, I probably would stay in another research area. I also want to thank Prof. Dr. Clemens Simmer from the University of Bonn and Prof. Dr. Kurt Jetter from the University of Hohenheim for reviewing my dissertation.

Thanks goes to the head of the modeling group, Dr. Hans-Stefan Bauer, for his supervision and never ending patience with answering all my questions about meteorology.

Thanks goes also to Dr. Kirsten Warrach-Sagi for giving insights to ensemble data assimilation and to Dr. Matthias Grzeschik and Dr. Florian Zus for the fruitful discussions about modeling and the introduction to shell-scripting and help with Fortran.

I highly acknowledge the German Research Foundation for the establishment of the Priority Program 1167 in which this thesis was possible and also the Climate Service Center, BW-GRID for providing large amounts of computing time.

ECMWF is acknowledged for the provision of meteorological data, the large amount of computing time and their technical support to run WRF on their HPC system.

Thanks goes to Dr. Galina Dick and Dr. Michael Bender from GFZ for the provision and support of GPS data, and to Dr. Klaus Stephan, Dr. Kathleen Helmert and Karolin Eichler from DWD for support and provision of German radar data.

Special thanks goes to all members of the IPM for the very friendly atmosphere and discussions apart from meteorology.

I am most thankful to my parents Veronika and Joachim and my sister Tina who encouraged me to study physics and their never ending patience and support. Although my father will not see or hear me anymore, he will by proud of me.

Curriculum Vitae

Personnel data

Name: Thomas Schwitalla
Address: Herschelstrasse 12 A, 70565 Stuttgart
Date of birth: 12.05.1981
Place of birth: Ostfildern/Ruit
Citizenship: German
Email: Thomas.Schwitalla@web.de

Education

08/1987–07/1991 Primary school: Schönbuchschule, Stuttgart

09-1991–06/2000 Secondary school: Fanny-Leicht Gymnasium, Stuttgart

10/2001–02/2007 Studies of Physics, University of Stuttgart Elective subjects: Surfaces and interfaces, geophysics , atmospheric physics

02/2006–02/2007 Diploma thesis, University of Hohenheim

Subject: *Kritische Prozessdarstellungen in Wettervorhersagemodellen zur quantitativen Niederschlagsvorhersage*

Supervisor: Prof. Dr. V. Wulfmeyer

03/2007–04/2008 Graduate assistant, University of Hohenheim

05/2008–10/2011 PhD student, University of Hohenheim

Civilian service

07/2000–05/2001 Hans-Rehn Stift - Haus Rohrer Höhe, Stuttgart

Supervision of Diploma thesis

04/2010–04/2011 *Ramon Walter*: Wolkenmikrophysikalische Parametrisierungsätze im Wettervorhersage-Modell WRF

Stuttgart, 11. Juni 2012

Publications

Peer reviewed publications

F. Zus, H.-S. Bauer, **T. Schwitalla**, V. Wulfmeyer, J. Wickert, 2011: Experiments of GPS slant path data assimilation with an advanced MM5 4DVAR system. Met. Z., **20**, 173-184.

T. Schwitalla, H.-S. Bauer, V. Wulfmeyer und F. Aoshima, 2010: High-resolution simulation over central Europe: Assimilation experiments during COPS IOP9c. Q. J. R. Meteorol. Soc., **137**, S1, 156-175.

H.-S. Bauer, V. Wulfmeyer, **T. Schwitalla**, F. Zus, M. Grzeschik, 2010: Operational assimilation of GPS slant path delay measurements into the MM5 4DVAR system. Tellus A, **63**, 263-282.

C. Barthlott, R. Burton, D. Kirshbaum, K. Hanley, E. Richard, J.-P. Chaboureau, J. Trentmann, B. Kern, H.-S. Bauer, **T. Schwitalla**, C. Keil, Y. Seity, A. Gadian, A. Blyth, S. Mobbs,C. Flamant, J. Handwerker, 2010: Initiation of deep convection at marginal instability in an ensemble of mesoscale models: A case study from COPS,Q. J. R. Meteorol. Soc., **137**, S1, 118-136.

H.-S. Bauer, T. Weusthoff, M. Dorninger, V. Wulfmeyer, **T. Schwitalla**, T. Gorgas, M. Arpagaus, K. Warrach-Sagi, 2010: Predictive Skill of a Subset of models that participated during D-PHASE in the COPS Region. Q. J. R. Meteorol. Soc., **137**, S1, 287-305.

V. Wulfmeyer, A. Behrendt, C. Kottmeier, U. Corsmeier, C.Barthlott, G. C. Craig, M. Hagen, D. Althausen, F. Aoshima, M. Arpagaus, H.-S. Bauer, L. Bennett, A. Blyth, C. Brandau, C. Champollion, S. Crewell, G. Dick, P. Di Girolamo, M. Dorninger, Y. Dufournet, R. Eigenmann, R. Engelmann, C. Flamant, T.Foken, T. Gorgas, M. Grzeschik, J. Handwerker, C. Hauck, H. Höller, W. Junkermann, N. Kalthoff, C. Kiemle, S. Klink, M. König, L. Krauss, C. N. Long, F. Madonna, S. Mobbs, B. Neininger, S. Pal, G. Peters, G. Pigeon, E. Richard, M. W. Rotach, H. Russchenberg, **T. Schwitalla**, V. Smith, R. Steinacker, J. Trentmann, D. D. Turner, J. van Baelen, S. Vogt, H. Volkert, T. Weckwerth, H. Wernli, A. Wieser, M. Wirth, 2011: The Convective and Orographically Induced Precipitation Study (COPS): The Scientific Strategy, the Field Phase, and Research Highlights. Q. J. R. Meteorol. Soc., **137**, S1,3-30.

J. Trentmann, C. Keil, M. Salzmann, C. Barthlott, H.-S. Bauer, **T. Schwitalla**, M. G. Lawrence, D. Leuenberger, V. Wulfmeyer, U. Corsmeier, C. Kottmeier, H Wernli, 2009: Multimodel simulations of a convective situation in low-mountain terrain in central Europe. Met. Atmos. Phys., Vol. 103, 95-103.

T. Schwitalla, H.-S. Bauer, V. Wulfmeyer und G. Zängl, 2008: Systematic errors of QPF in low-mountain regions as revealed by MM5 simulations, Meteorol. Z., **17**, 903 - 919.

Monographs

T. Schwitalla, 2007: Kritische Prozessdarstellungen in Wettervorhersagemodellen zur quantitativen Niederschlagsvorhersage, Diploma thesis. Stuttgart University library, Signature Y 3398.

Non-reviewed publications

M. Dorninger, T. Gorgas, **T. Schwitalla**, M Arpagaus, M. Rotach, V. Wulfmeyer, 2009: Joint D-PHASE - COPS data set (JDC data set). Technical report. Available at `http://cera-www.dkrz.de/WDCC/ui/BrowseExperiments.jsp?proj=JDC`

H.-S. Bauer, M. Grzeschik, F. Zus, **T. Schwitalla**, V. Wulfmeyer, 2009: dphase_mm5_2_ct: MM5 model forecast with 2 km horizontal resolution driven by ECMWF operational forecast run by UHOH for the MAP D-PHASE project. World Data Center for Climate. DOI: 10.1594/WDCC/dphase_mm5_2_ct.

H.-S. Bauer, M. Grzeschik, F. Zus, **T. Schwitalla**, V. Wulfmeyer, 2009: dphase_mm5_2_ct: MM5 model forecast with 2 km horizontal resolution based on 4DVAR assimilation of GPS data run by UHOH for the MAP D-PHASE project. World Data Center for Climate. DOI: 10.1594/WDCC/dphase_mm5_2_4d.

T. Schwitalla, G.Zängl, H.-S. Bauer und V. Wulfmeyer: Convective initiation in the Black Forest region in high-resolution MM5 simulations. *Extended Abstract*, 29^{th} Int. Conference on Alpine Meteorology, Chambéry, Frankreich, 04.-08. Juni 2007

V. Wulfmeyer, A Behrendt, **T. Schwitalla**, H.-S. Bauer, M. Rotach und C. Kottmeier: The WWRP RDP COPS and the FDP D-PHASE. A unique opportunity for studying the impact of small scale/large-scale interaction on medium-range QPF. Second THORPEX International Science Symposium (STISS), Landshut, Germany, 04.-08. Dezember 2006

Presentations

M. Grzeschik, **T. Schwitalla**, A. Behrendt,H.-S. Bauer, V. Wulfmeyer: Assimilation of lidar and GPS data in the frame of COPS and beyond. COST 702 Meeting, Königswinter, November 2010.

T. Schwitalla, H.-S. Bauer, V. Wulfmeyer: High-resolution simulation over central Europe: Assimilation experiments during COPS IOP9c. 11^{th} WRF Workshop, Boulder/CO, USA, 22. June 2010.

T. Schwitalla, H.-S. Bauer, V. Wulfmeyer: Overview of the application of WRF at IPM UHOH. Ertel coordination meeting, Offenbach, June 2010.

T. Schwitalla, H.-S. Bauer, V. Wulfmeyer, F. Zus, M. Grzeschik, G. Dick: Overview over SRNWP/COPS-GRID in the framework of SPP 1167 at UHOH. SPP 1167 final meeting, Bonn,

April 2010.

T. Schwitalla: "Performance des UHOH WRF systems". COPS-GRID project meeting, Potsdam, March 2010.

H.-S. Bauer, V. Wulfmeyer, F. Zus, **T. Schwitalla**, G. Dick, M. Bender, J. Wickert, G. Gendt: The SPP1167 Project COPS-GRID and results of first studies using GPS and radar data. 30^{th} Int. Conference on Alpine Meteorology, Rastatt, May 2009.

T. Schwitalla: Assimilation mit dem WRF-Var System. COPS-GRID project meeting, Stuttgart, 2009.

T. Schwitalla, G.Zängl, H.-S. Bauer und V. Wulfmeyer: Convective initiation in the Black Forest region in high-resolution MM5 simulations. 29^{th} Int. Conference on Alpine Meteorology, Chambéry, France, June 2007.

Poster

V. Wulfmeyer, H.-S. Bauer, **T. Schwitalla**, K. Warrach-Sagi, 2010: Forecasting in Ethiopia, one of the most challenging and vulnerable regions with respect to weather and climate. EMS 2010, Zurich.

H.-S. Bauer, **T. Schwitalla**, G. Dick, V. Wulfmeyer, M. Bender, Z. Deng, J. Wickert, G. Gendt, K. Stephan, K. Helmert, K. Eichler, M. Dorninger, T. Gorgas, R. Steinacker, M. Arpagaus, T. Weusthoff, C. Kottmeier, U. Blahak, C. Wunram: Current status and first results of the SPP1167 projects COPS-GRID and D-PHASE Verification. 8^{th} COPS Workshop, Cambridge, UK, October 2009.

T. Schwitalla, H.-S. Bauer, F. Zus, V. Wulfmeyer: The WRF modeling system and first results of its application within the COPS period. 30^{th} Int. Conference on Alpine Meteorology, Rastatt, May 2009.

T. Schwitalla, H.-S. Bauer, F. Zus, V. Wulfmeyer: High-Resolution WRF simulations for selected IOPs during the field experiment COPS. 9^{th} WRF Workshop, Boulder/CO, USA, June 2008.

Proposals and project reports

Bauer, H.-S., V. Wulfmeyer and **T. Schwitalla**, 2011: Validation and improvement of high-impact weather process understanding with the aid of high-resolution WRF simulations and sophisticated data assimilation (VALPUDA), Proposal for a special project at the European Centre for Medium Range Weather Forecasting (ECMWF),7pp.

Bauer, H.-S., V. Wulfmeyer, **T. Schwitalla** and K. Warrach-Sagi 2011: High-Resolution Climate Predictions and Short-Range Forecasts to Improve the Process Understanding and the Representation of Land-Surface Interactions in the WRF Model in Southwest Germany (WRFCLIM), Progress report for the project WRFCLIM.

Bauer, H.-S., V. Wulfmeyer, K. Warrach-Sagi, **T. Schwitalla**, M. Grzeschik, S. Reich and J. Wickert, 2010: Research on Consistent Land-Surface-Atmosphere Data Assimilation on the Convection-Permitting Scale with the Hans-Ertel Center Data Assimilation Testbed. Proposal for the Hans-Ertel Research Center of DWD, 38pp.

Bauer, H.-S., V. Wulfmeyer and **T. Schwitalla**, 2010: Short-Range QPF by Combining 4D Observations of Water Vapor and Wind with Innovative Methods in Data Assimilation. Progress Report for computing time at the German Climate Computing Center (DKRZ), 4pp.

Bauer, H.-S., V. Wulfmeyer, **T. Schwitalla**, K. Warrach-Sagi and J. Chu, 2010: High-Resolution Climate Predictions and Short-Range Forecasts to Improve the Process Understanding and the Representation of Land-Surface Interactions in the WRF Model in Southwest Germany (WRFCLIM), Progress report for the project WRFCLIM, 10pp.

Bauer, H.-S., V. Wulfmeyer, **T. Schwitalla**, K. Warrach-Sagi and J. Chu, 2009: High-Resolution Climate Predictions and Short-Range Forecasts to Improve the Process Understanding and the Representation of Land-Surface Interactions in the WRF Model in Southwest Germany (WRFCLIM), Proposal for computing time at the High Performance Computing Center Stuttgart, 8pp.

Bauer, H.-S., V. Wulfmeyer and **T. Schwitalla**, 2009: Short-Range QPF by Combining 4D Observations of Water Vapor and Wind with Innovative Methods in Data Assimilation. Proposal for computing time at the German Climate Computing Center (DKRZ), 4pp.

Stuttgart, June 11, 2012

Erklärung:

Hiermit erkläre ich, dass ich die vorgelegte Dissertation selbständig angefertigt und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder inhaltlich übernommene Stellen sind als solche gekennzeichnet.

Stuttgart, den 17.07.2012

Thomas Schwitalla