

DIPLOMARBEIT

Titel der Diplomarbeit

Improvement of short term precipitation forecasts in the
Alpine Region using WRF with 3DVAR RADAR
Reflectivity and SYNOP assimilation

Verfasser

Johannes Rausch

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat)

Wien, im August 2012

Studienkennzahl lt. Studienblatt: A 415

Studienrichtung lt. Studienblatt: Meteorologie

Betreuer: ao. Univ.-Prof. Mag. Dr. Leopold Haimberger

Contents

1	Introduction	4
2	Data Assimilation	5
2.1	The data Assimilation Problem	5
2.2	Variational Analysis - 3DVAR	7
3	WRFVAR-System	11
3.1	WRF Modelcore	11
3.2	WRF Physics	12
3.3	WRF Workflow	14
3.4	Data assimilation in WRF	16
3.4.1	WRFVAR System	16
3.4.2	Rapid Updated Cycle WRF	18
3.4.3	Background Error Generation with GEN_BE	19
4	Observation Preprocessing	22
4.1	Station Data Preprocessing	22
4.2	Radar analysis	23
5	Reflectivity Assimilation in WRFVAR	28
5.1	Forward Operator	28
5.2	Single OBS experiments	30
6	Case study 19.08.2011	36
6.1	Analysis increments found in the case study	36
6.2	WSM3 vs. WSM6 twin experiment	46
7	Verification August 2011	51
7.1	Skill Scores	51
7.2	Verification results	52
8	Conclusion and perspectives for further research	55

Abstract

The forecast of convective precipitation is still an unsolved issue of numerical weather prediction, because of the small scales of the involved meteorological processes and the never optimal model initialization of a global model. The goal of this work is to improve the short term forecast of convective precipitation by 3DVAR assimilation of radar reflectivities and station data within a mesoscale model (WRF - Weather and Research Forecasting). The author therefore implements a RUC (Rapid Update Cycling) system to generate hourly updated analysis, which are serving as new initial conditions for subsequent forecast runs. Due to the incompleteness of the available RADAR data in the research domain, the author develops a RADAR merging procedure where two dimensional RADAR data get expanded to three dimensional data by using artificial vertical profiles. Finally these data are merged with existing three dimensional radar over Austria and BOGUS radar data, which are created out of station observation precipitation.

The impact of the assimilated RADAR reflectivities and all available European SYNOP stations on the analysis and the computed forecast depends critically on an optimal configuration of the background error covariances, which are the determining factor how an observation is spread in space and time in the model. A case study shows the improved capability to predict severe weather developments in the Alpine region. The author shows an improvement of the short term precipitation forecast skill scores over Austria due to a cycled 3DVAR assimilation of radar and station data.

Abstract

Die Prognose konvektiver Niederschlagsereignisse stellt aufgrund der Kleinskaligkeit der meteorologischen Ereignisse und der niemals optimalen Modellinitialisierung globaler Modelle noch immer ein grosses Problem in der numerischen Wettervorhersage dar. Ziel dieser Arbeit ist die kurzfristige Vorhersage konvektiver Niederschlagsereignisse mittels 3DVAR Assimilation von RADAR-Reflektivitäten und Stationsdaten innerhalb eines mesoskaligen Modells (WRF - Weather and Research Forecasting) zu verbessern. Der Autor implementiert ein sog. RUC (Rapid Update Cycling) System um stündlich aktualisierte Analysen zu generieren, die als neue Startbedingungen des nächsten Vorhersagelaufes dienen. Da dem Autor kein vollständiges dreidimensionales Reflektivitätsfeld Mitteleuropas zur Verfügung steht wird ein Verfahren entwickelt, bei dem künstliche vertikale Radarprofile generiert werden, die im Anschluss mit dem vorhandenen 3D-RADAR aus Österreich und mit den aus Stationswerten generierten BOGUS RADAR-Reflektivitäten verschmolzen werden. Der Einfluss der Assimilation von RADAR-Reflektivitäten und aller verfügbarer SYNOP Stationsdaten auf die kurzfristige Prognose hängt vorwiegend von der Konfiguration und Analyse der Background Error Kovarianzen ab, die für die zeitliche und räumliche Ausbreitung der assimilierten Observationen ausschlaggebend sind. Anhand einer ausgewählten Fallstudie soll das Potenzial zur Prognose schadensträchtiger Unwetterereignisse im Alpenraum aufgezeigt werden. Der Autor zeigt, dass bei entsprechender Konfiguration des Assimilations- und des Modellsetups eine signifikante Verbesserung der Kurzfristvorhersage möglich ist.

1 Introduction

In Numerical Weather Prediction the quality of forecasts is highly dependent on the quality of the data assimilation process, which uses many different meteorological observation types to calculate an initial state from which a model can start a forecast. In recent years it became a major issue in meteorology to improve the short term forecast on smaller scales, to capture mesoscale processes like convection in space and time. Especially the (non)-occurrence of thunderstorms and their hazards in a convective atmospheric environment is a matter of public interest and it would be desirable to have access to a reliable nowcasting system. To achieve this it is common to use mesoscale models and feed them with high resolution observational data (wind profiler, RADAR¹, satellite) for a more realistic state vector. The objective of this work is to evaluate the impact of assimilated RADAR data on short term weather forecasts over Austria and Central Europe using the non-hydrostatic WRF (Weather Research and Forecasting) model (Skamarock et al. 2008) with the included variational analysis package WRFVAR. The implementation of a RUC (Rapid Update Cycling) system with one hour cycling seems to be an appropriate work-flow setup to take advantage of the high spatial and temporal resolution of the radar data. We will concentrate just on assimilation of the reflectivity factor, although it is already proven (Kuo et al. et al. 2005) that assimilation of Doppler velocities has a positive impact on the retrieval of the vertical wind component, which is of course crucial for any type of precipitation process. A key point of this study is the tuning of the background error covariances, which determine and spread the impact of the observations in space and time. To find the best configuration many experiments are verified with the MET (Meteorological Evaluation Tools) framework.

¹Even though RADAR is an acronym, it's lower case written in the following text.

2 Data Assimilation

2.1 The data Assimilation Problem

The numerical weather prediction is an initial value problem. For every grid point of the three dimensional model field, information of the prognostic variables (pressure, specific humidity, wind and temperature) are needed. Even if one takes all available observations one has access to on earth, like satellite radiances, sondes, surface station data, wind profilers, Doppler radars, ships and airplane reports, one will not be able to assign the corresponding value to every grid point. The meteorological equation system stays under determined and therefore not solvable. To get rid of this major issue one uses a previous forecast as background, in other words the model itself. Now the data assimilation process takes place and corrects the background (first-guess) in direction of the observation, wherever differences between observations and background are occurring. The data assimilation is the missing link between the background and the observations. For the first run (cold start) ever, the background consists of pure climatology data and is naturally far away from the true state of atmosphere.

To get a reliable forecast, one has to ensure, that the data assimilation process has done a good job (Simmons and Hollingsworth 2002). The principle aim is to adjust an existing model forecast closer to the true state of the atmosphere with as many observations as possible to generate an analysis from which a new forecast can be launched. In the beginnings of data assimilation comparatively simple methods were used, for example the Cressman Scheme (Bouttier and Courtier 1999), a simple interpolation between observations and background with distance weighting, where physical consistency was not guaranteed.

The Optimal Interpolation (Bouttier and Courtier 1999) or Statistical Interpolation, where for the first time in history of NWP error covariances of the background and the observation field have been used for spreading information more accurate in space and time was more sophisticated than the Cressman scheme. Finally the variational analysis was developed, namely 3DVAR and 4DVAR (Talagrand and Courtier 1987, Andersson et al. 1998). To find the most likely state of the current atmosphere a cost function gets minimized through an iterative method. One of the biggest benefits of this approach is the capability to assimilate observation data with a non-linear relationship (forward operator) to the background variables like satellite radiances or reflectivities. Furthermore the analysed meteorological fields are consistent with the governing equations of numerical weather prediction. 3DVAR is using all available observations at one given analysis time whereas 4DVAR includes the time dimension, hence high frequency observations over a longer time window (usually 6 hours) can be assimilated at appropriate times. Further

used in data assimilation, but not focus of this work is the Kalman Filter algorithm (Anderson, 2001). For this work, emphasis is on the 3D-VAR algorithm as implemented by WRF(Weather Research and Forecasting Model).

2.2 Variational Analysis - 3DVAR

There are two approaches to find the derivation of the 3DVAR algorithm, the least squares method and the maximum likelihood method (Faulwetter 2008, Lenz 2011) which will be the procedure of our choice. The unknown state of the atmosphere on the model grid is x . The background state x_b is a previous forecast. The background error χ is analogously defined as

$$\chi = x - x_b \quad (1)$$

The expectation value of all background errors assuming a Gaussian distribution is zero.

$$E(\chi) = 0 \quad (2)$$

$$Cov(\chi) = E(\chi\chi^T) = \mathbf{B} \quad (3)$$

Assuming the distribution of x is Gaussian and the background and its covariance matrix are given, we can estimate the probability of x with the cost function J_b

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{B})^{\frac{1}{2}}} e^{-J_b(x)} \quad (4)$$

$$J_b(x) = \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) \quad (5)$$

We say that there is a non-linear connection between the observations y and the model state x .

$$y = \mathbf{H}(x) \quad (6)$$

The Forward Operator $\mathbf{H}(x)$ is transforming data from model space to observation space. That could be a spatial interpolation from grid points to the location of the observation or the transformation from radar reflectivity to a corresponding rain water mixing ratio or from Doppler velocity to a 3D wind vector or from satellite radiances to temperature and relative humidity.

With the combined representivity and observation error ε this leads to

$$z = \mathbf{H}(x) + \varepsilon \quad (7)$$

Analogously to χ we define under the assumption of a Gaussian distribution and unbiased

observations

$$E(\boldsymbol{\varepsilon}) = 0 \quad (8)$$

$$Cov(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \mathbf{R} \quad (9)$$

where \mathbf{R} holds the covariance observation error data.

It is worth to mention that is not uncommon to deal with biased observations. In that case, the Bias has to be determined(Dee 2005). Furthermore the observation error and the background error are assumed to be uncorrelated.

$$E(\boldsymbol{\varepsilon}\boldsymbol{\chi}^T) = E(\boldsymbol{\chi}\boldsymbol{\varepsilon}^T) = 0 \quad (10)$$

The probability density function of an observation z for a given state x is written

$$p(z|x) = \frac{1}{(2\pi)^{\frac{m}{2}} \det(\mathbf{R})^{\frac{1}{2}}} e^{-J_o(x)} \quad (11)$$

$$(12)$$

with J_o as costfunction for the observations

$$J_o(x) = \frac{1}{2}(z - \mathbf{H}(x))^T \mathbf{R}^{-1}(z - \mathbf{H}(x)) \quad (13)$$

For us much more interesting is the probability of x for a given z . For that reason we have to introduce a formalism for conditional probabilities, the Bayes' theorem to modify our equation to get $p(x|z)$:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \quad (14)$$

where $P(B|A)$ is the probability of B given A

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A) \quad (15)$$

In our case

$$p(z \cap x) = p(z|x)p(x) = p(x|z)p(z) \quad (16)$$

which we can convert to

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} \quad (17)$$

where

$$p(z) = \sum_j p(z|x)p(x) \quad (18)$$

or

$$p(z) = \int_{R^n} p(z|x)p(x)dx \quad (19)$$

Since we want to vary x and $p(z)$ does not depend on x , we assume $p(z) = \text{constant}$

$$K = \frac{1}{p(z)} \quad (20)$$

Inserting equations $p(z|x)$ and $p(x)$ in the equation of $p(x|z)$ we get as result:

$$p(x|z) = \frac{K}{(2\pi)^{\frac{m+n}{2}} (\det(\mathbf{R})\det(\mathbf{B}))^{\frac{1}{2}}} e^{-(J_o(x)+J_b(x))} \quad (21)$$

The first term on the right side of the equation doesn't depend on x , hence we can examine the maximum of $p(x|z)$, if we try to find the absolute minimum of the exponent of the exponential function, which is our **3DVAR cost function $J(x)$** . The state x , where $J(x)$ is minimal is the analysis. The scalar $J(x)$ is the distance from x to the observation and the background weighted with the inverses of the error-covariances.

$$J(x) := J_o(x) + J_b(x) \quad (22)$$

$$= \frac{1}{2}(z - \mathbf{H}(x))^T \mathbf{R}^{-1}(z - \mathbf{H}(x)) + (x - x_b)^T \mathbf{B}^{-1}(x - x_b) \quad (23)$$

In case of a non linear observation operator, like satellite radiances for example, it is common to use an iterative approach to minimize the cost function. In the following the Newton method is explained, where the minimum is determined by incremental iterations. A Taylor series of 2nd order for $J(x)$ near to x_c is calculated.

$$Q(y) := J(x_c + y) \approx J(x_c) + (\nabla J(x_c))^T y + \frac{1}{2} y^T \nabla^2 J(x_c) y \quad (24)$$

$$(25)$$

The gradient is calculated with

$$\nabla Q(y) = \nabla J(x_c) + \nabla^2 J(x_c)y \quad (26)$$

In order of minimization, set $\nabla Q(y) = 0$

$$\nabla^2 J(x_c)y = -\nabla J(x_c) \quad (27)$$

hence, the result of this equation system, where the Jacobi-matrix $\nabla J(x_c)$ and the Hesse-matrix $\nabla^2 J(x_c)$ have to be known is

$$x_{new} = x_c + y_{min} \quad (28)$$

The minimum can be seen as an increment and creates the starting point x_{new} of the next iteration of minimization until the gradient reaches a convergence criterion, that is specified in WRFVAR.

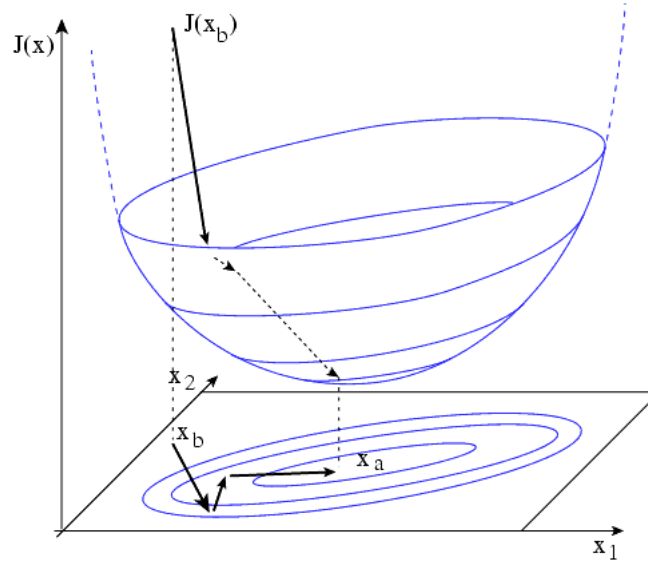


Figure 1: From Bouttier and Courtier (1999): Schematic representation of the variational cost function minimization (here in a two-variable model space): the quadratic cost function has the shape of a paraboloid, or bowl, with the minimum at the optimal analysis. The minimization works by performing several line-searches to move the control variable to areas where the cost function is smaller, usually by looking at the local slope (the gradient) of the cost function.

Next we describe a specific implementation of this algorithm in the WRFVAR framework.

3 WRFVAR-System

This chapter summarizes the key components of the WRFVAR-System. General information comes from the WRF User's Guide 2011, while more detailed technical and physical information can be found in the WRF Technical Note(Skamarock et al. 2008).

3.1 WRF Modelcore

The WRF (Weather Research and Forecasting) model is a fully compressible, Eulerian mesoscale model developed by many institutes worldwide, mainly NCAR and NCEP, and is available with two different cores

- **ARW** Advanced **R**esearch **W**Rf created from NCAR
- **NMM** Nonhydrostatic **M**esoscale **M**odel created from NCEP

Both model cores are using a 3rd order Runge Kutta time integration scheme and a 2nd to 6th order spatial discretization. The main difference between these two model cores is that ARW has more complex dynamic and physics settings than NMM. The WRF of this work was driven by the ARW core due to these reasons. WRF comes with public license and is therefore free for private use. WRF can be driven from user-created initial conditions for meteorological research(reanalysis) or from an external NWP product(real data case) in an operational forecasting setup.

There are two different downscaling methods available for WRF operation:

- **one way nesting** downscaling in the common sense, parallel computing of both domains, the output of the parent domain (coarse grid) works as lateral boundary condition for the child domain (fine grid).
- **two way nesting** works like one-way nesting, additionally there is information exchange (feedback) between the fine grid and the coarse grid.

ARW uses a staggered ARAKAWA C grid (Fig.2), where the vector values (u, v) are placed in the middle of two gridpoints and the scalar values (T, q, p) are placed in the center of an entire gridcell. In the next picture a nesting based on an ARAKAWA C grid with the ratio 1:3 like in our setup is illustrated. For the information feedback from the fine grid to the coarse grid for scalar values the mean of all nine points is fed back, whereas for the vector values only the mean of three values, the face of a grid is fed back.

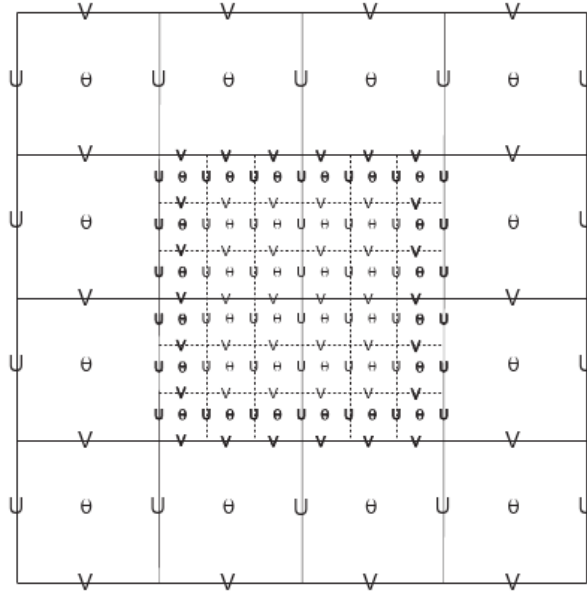


Figure 2: From ARW Technical Note Skamarock et al. 2008: Arakawa C grid

The choice of the integration timestep is limited by the Courant Number

$$\Delta t_{max} < \frac{C_R}{\sqrt{3}} \cdot \frac{\Delta x}{u_{max}}$$

where Δt_{max} is the timestep where the RK3 time integration is still stable, $C_R(1.08)$ is the Courant number taken from Wicker and Skamarock (2002), Δx is the gridsize of the domain and u_{max} the highest expected velocity in the entire model domain. As rough rule of thumb for the estimation of the ARW timestep it is common to use timesteps of 6 times the gridsize (km). For the precise calculation based on our model setup with gridsize of 12/4 km and a maximum velocity of 100ms^{-1} we get timesteps for domain one and two of 130s and 43s. In the WRF user's guide it is recommended to choose an approximately 25 percent smaller timestep, which leads to values of 90 and 30 seconds, respectively.

3.2 WRF Physics

There are atmospheric mechanisms and exchange processes, which cannot be represented in discretized equations. These processes need to be parameterized. Within a forecast model, the parameterizations are referred to as "physics package". The WRF physics packages (Skamarock et al. 2008) are developed to incorporate these processes through the computation of tendencies for the velocity components, potential temperature and moisture fields based on diagnostic equations and parameterizations. The WRF physics package offers a huge range of different configuration possibilities dependent on the par-

ticular user application. In general the physics are made up of microphysics, cumulus parameterization, planetary boundary layer (PBL), land-surface model and radiation. Due to the countless number of available options, only a short overview is given here. Cumulus parameterization, planetary boundary layer (PBL), land-surface model, and radiation tendencies are computed in the first Runge Kutta step and stay fixed through the entire RK timestep integration. Microphysics are parameterizations of water vapor, cloud and precipitation processes. They are called at the end of every RK3 timestep as a condensation-adjustment to guarantee that the saturation conditions are fitting with the updated moisture and temperature from the RK3 integration. The microphysics are also estimating the concentration and number of ice hydrometeors using a function of temperature and are evaluating the fall speed and finally the amount of precipitation (mixing ratio x fall-speed), which reaches the ground. It is advantageous to use a more complex microphysics scheme like WSM6 (Hong and Lim 2006), if high resolution and cloud resolving grids are targeted. The WSM6 scheme is a good candidate for that, because it's a mixed-phase microphysics scheme, taking processes between ice and rain particles (graupel and hail) into account. Generally the exchange processes between the physics and the model state variables are very complex as illustrated in the next graphic.

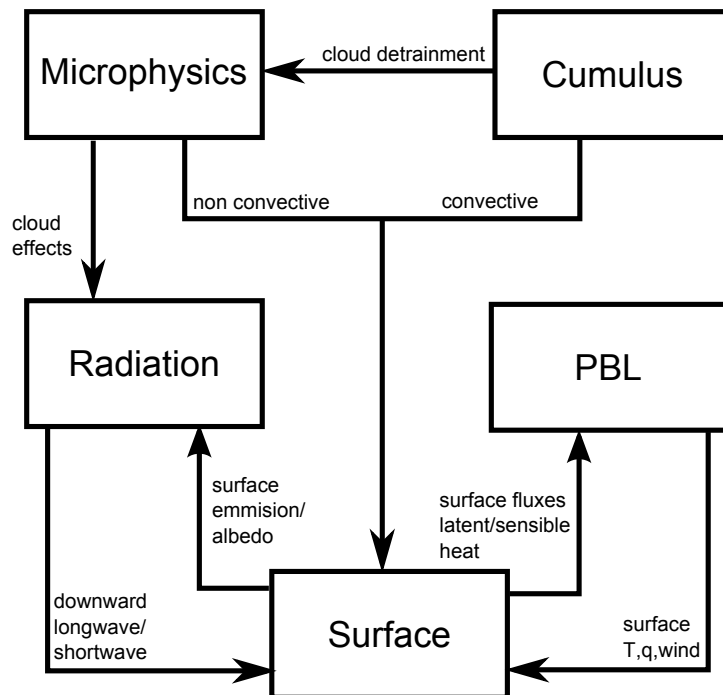


Figure 3: From Skamarock et al. (2008): Physical interactions

3.3 WRF Workflow

The main work-flow of real-data cases is split up into the execution of Fortran 90 programs `geogrid.exe(serial)`, `ungrib.exe(serial)`, `metgrid.exe(parallel)`, `real.exe(parallel)`, `wrf.exe(parallel)`. Parallel means that this program can be executed on parallel distributed-memory computer systems, serial just on one CPU. Geogrid, Ungrib and Metgrid are combined in the WRF Preprocessing System WPS, which prepares the input data for the later WRF execution.

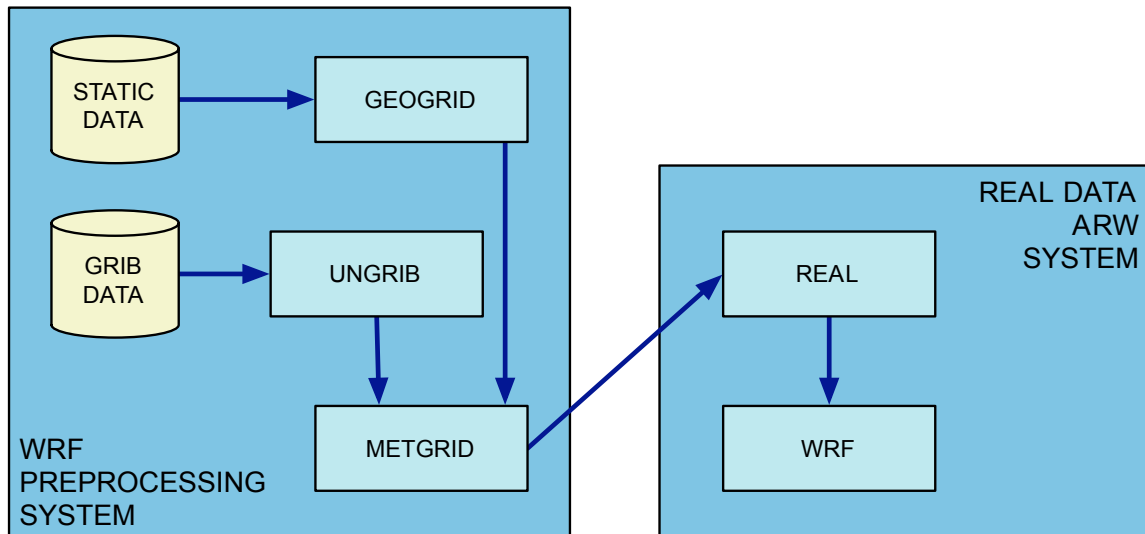


Figure 4: From WRF User's Guide: Basic WRF workflow

Geogrid creates a netcdf file of all static WRF-relevant data based on the user-configuration defining the projection type, number of grid points, nest locations, and grid distances. The content of the netcdf file is data like albedo, Coriolis parameters, terrain elevation, vegetation/land-use type, land/water mask, map scale factors, vegetation greenness fraction, annual mean temperature and latitude/longitude. It is possible to define a topography data resolution up to 30m (SRTM - GIS Data).

Ungrib extracts all relevant time varying meteorological fields from GRIB files for WRF initialization and saves them in an intermediate format for further Metgrid processing.

Metgrid horizontally interpolates the meteorological fields from the intermediate files to the model grid predefined by Geogrid. For every parameter the most suitable interpolation can be selected out of many. To achieve a successful Metgrid completion the following list of meteorological/geographical data from an external NWP product should be available:

- 3-dimensional fields of temperature (K), relative humidity, horizontal wind components u, v (m/s)
- 2-dimensional fields of surface pressure and sea-level pressure (Pa), layers of soil temperature (K) and soil moisture (kg/kg), snow depth (m), skin temperature (K), sea surface, temperature (K), and a sea ice flag

Real interpolates the Metgrid output (in netcdf format) vertically on the hydrostatic pressure coordinate

$$\eta = \frac{p - p_T}{p_S - p_T},$$

where p is the pressure at a given level, p_T is the pressure at the top of atmosphere and p_S is the pressure on the surface. η has values of 1 at the surface and 0 at the top of the atmosphere and can be seen as a terrain following pressure coordinate. In NWP terrain following coordinates are used with finer vertical spacing near the surface to resolve smaller scales of the boundary layer and coarser spacing at higher levels of the atmosphere to reduce computational costs.

Real splits the geopotential and the column pressure into a reference state(dry) and a perturbation(with moisture) form to avoid truncation errors from the discrete ARW solver. Further the module Real prepares soil fields dependent on the selected land surface scheme, checks, if the Metgrid interpolated parameters are consistent with each other and creates a lateral boundary condition netcdf file(wrfbdy_d0x) for WRF initialisation and finally a wrfinput-file with the initial conditions. The lateral boundary file consists of lateral boundary values for every forecast time of the Metgrid data and their tendency values to get to the lateral boundary values in the next time period.

3.4 Data assimilation in WRF

WRF-ARW comes with WRFVAR, a data assimilation package offering several techniques for data assimilation:

- basic 3DVAR, widely used computationally efficient assimilation method
- FGAT - First guess at appropriate time assimilation
- 4DVAR, best algorithm but computationally expensive
- Hybrid data assimilation with flow dependent background error covariances using ensemble methods

The differences between 3DVAR and 4DVAR are the longer assimilation window and the implicitly generated flow-dependent background errors. In reality the benefits of 4DVAR against 3DVAR are shrinking, if a frequently cycled 3DVAR forecasting system is used. Like WRF the WRFVAR-package consists of a bundle of program modules - writing routines for observation preprocessing, guaranteeing the availability of the needed input data and to produce the proper WRFVAR workflow is user's business.

3.4.1 WRFVAR System

The WRFVAR workflow is a mixture of already existing WRFDA modules (blue), interacting with user programmed routines (little_r obs-data generation, radar preprocessing or cloud analysis) as illustrated in the following flow diagram. As already discussed the

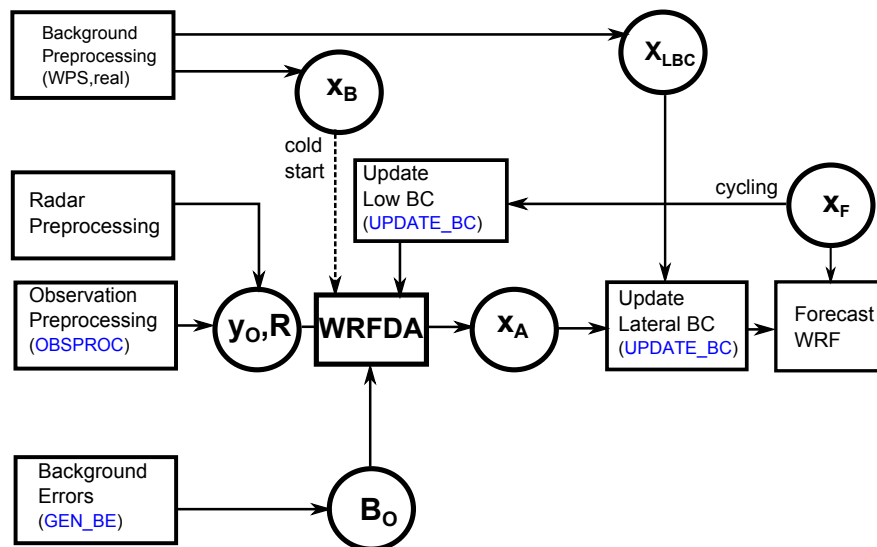


Figure 5: WRFVAR workflow (blue: WRFDA implemented modules)

main goal of WRFVAR is to minimize a cost function $J(x)$ in an iterative way to find an optimal estimate of the state x of the atmosphere:

$$J(x) := J_o(x) + J_b(x) \quad (29)$$

$$= \frac{1}{2}(y - y_o)^T \underbrace{(E + F)}_R^{-1} (y - y_o) + (x - x_b)^T B^{-1} (x - x_b) \quad (30)$$

where

- x is the analysis state we're trying to find
- x_b is the "first-guess", a previous forecast (warmstart-cycling) or an analysis of an external NWP product interpolated to the ARW grid (coldstart-mode)
- y_o is the observation
- y is the counterpart of y_o computed through the forward operator $\mathbf{H}(x)$ from analysis space
- B is the background error covariance matrix
- E is the instrumental observation error covariance matrix
- F is the representivity error covariance matrix
- R is the total observation error covariance matrix

The representivity error is describing the inaccuracy in the transformation from analysis space to observation space and interpolation from grid points to the observation location with the forward operator $\mathbf{H}(x)$.

There are approaches to tune the radar observation error covariance matrix, described in Desroziers and Ivanov (2001). In our configuration WRFVAR-default-observation errors are used.

For calculation of B GEN_BE is used. GEN_BE is a WRFDA-integrated tool, which calculates B out of a forecast series. This procedure will be discussed detailed in a later chapter of this work.

The parallel executable da_wrfvar.exe solves the cost function and generates important assimilation statistics, which are helpful for tuning and post-configuration of the WRFVAR-system. Outer loops (Skamarock et al. 2008) are used as a strategy for more efficient assimilation of non-linearities (observation operators and balance constraints) and for variational quality control. Therefore resulting analysis of the first assimilation procedure acts

as the background for the subsequent minimization. Observations, which were rejected in the first loop due their too large departures from the background will possibly be ingested in the next minimization, because the observation-background difference falls under the maximum-error in the internal quality control. This means that in areas with a high forecast error (probably due to poor assimilation in the global model), valuable observations are getting accepted in a later outer loop.

After successful completion of da_wrfvar.exe, the wrfvar_output file holds the new analysis data, the new initial conditions for a subsequent forecast. In the next step the WRFDA included utility update_bc updates the lateral boundary conditions, the tendencies for the first timestep of the wrfbdy_d01 file.

3.4.2 Rapid Updated Cycle WRF

A RUC system is implemented to take advantage of the high temporal and spatial resolution of the radar data. RUC systems are operated e.g by the U.S Storm Prediction Center(SPC). Start of the cycling-procedure is a cold start. GFS (Global Forecast System) 0.5 degree analyses are taken as background and observations are assimilated for the first time to create a new analysis, which is used as new initial condition for the cold-start. For the subsequent cycles (warm-starts) the 1-hour forecasts of the previous WRF are used as first guess. For the analysis the first guess is updated with fresh RADAR and observation data. This procedure is continued until the next GFS run is available.

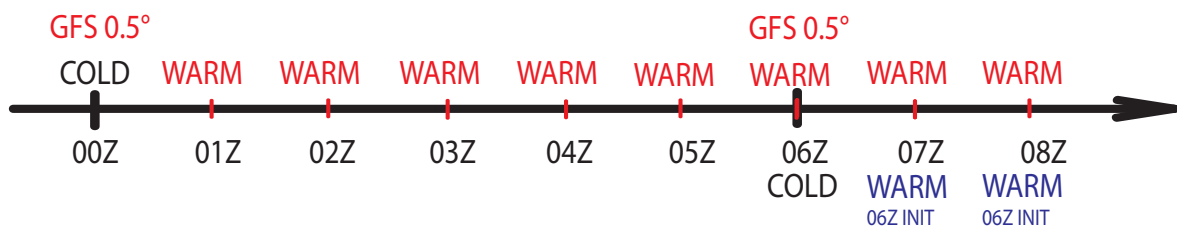


Figure 6: Rapid Update Cycling model procedure

3.4.3 Background Error Generation with GEN_BE

As already mentioned above, the estimation of the background errors is made with the WRFVAR tool `gen_be`. Due to the fact, that the background errors are unknown in reality, the NMC method (Parrish and Derber 1992) is used to create an approximation of the error covariances based on averaged forecast differences (24 hour - 12 hour) for the same forecast step.

$$B = \overline{(x^b - x^f)(x^b - x^f)^T} = \overline{\varepsilon_b \varepsilon_b^T} \approx \overline{(x^{T+24} - x^{T+12})(x^{T+24} - x^{T+12})^T} \quad (31)$$

where the overbar means an average over time and/or a geographical area. An alternative strategy proposed by Fisher (2003) uses ensemble forecast output, defining the ε vectors as ensemble perturbations (ensemble minus ensemble mean). It is recommended by the WRF user's guide to calculate the NMC B's based on a WRF forecast series over a month. The background errors used for cold-start should differ from the warm-start ones, because of the different resolution from the driving global model. Unfortunately GFS is only available in GRIB format, which cannot be read from `gen_be`. So slightly adapted warm-start background errors are used for the cold-starts.

The size of B is in the order of $10^7 \times 10^7$ (number of grid-points \times number of independent variables). The enormously high number comes from the cross-correlation of the physical variables of the model, namely u, v, T, p and q . The reduction of the cross-correlations can be achieved by introducing less correlated control variables v . Then the cross-correlations are neglected to block-diagonalize the background error covariance matrix, leaving only spatial correlations between individual analysis variables (Barker et al. 2003). Furthermore the computationally most efficient choice of the variables is essential. If e.g. vorticity and divergence are used as control variables, it is computationally more expensive (solution over poisson equation) to calculate u and v wind components, than to choose stream function and velocity potential as control variables. The variable transformation (Barker et al. 2003) is written

$$x' = Uv \quad (32)$$

where $x' = x - x_b$ and $B = UU^T$. The control variables chosen in WRFVAR are

- stream function ψ
- unbalanced part of velocity potential χ
- unbalanced part of temperature T
- relative or specific humidity q

- unbalanced part of surface pressure ps

According this assumption the 3DVAR equation is modified to

$$J(v) = J_b + J_o = \frac{1}{2} \mathbf{v}^T \mathbf{v} + \frac{1}{2} (\mathbf{y}^{o'} - \mathbf{H} \mathbf{U} \mathbf{v})^T (\mathbf{E} + \mathbf{F})^{-1} (\mathbf{y}^{o'} - \mathbf{H} \mathbf{U} \mathbf{v}) \quad (33)$$

where

$$\mathbf{y}^{o'} = \mathbf{y}^o - \mathbf{H}(\mathbf{x}_b) \quad (34)$$

is the innovation vector. The background error covariance matrix \mathbf{B} is now block-diagonalized and therefore much more efficient for computation. The number of required calculations is reduced from $O(10^7 \times 10^7)$ to 10^7 . A further advantage of preconditioning is the simplified use of adjoints of the cost function J during minimization process.

In practice the computation of \mathbf{B} is divided in three different non-separable operations

$$\mathbf{B} = U_h U_v U_p U_h^T U_v^T U_p^T \quad (35)$$

where

- U_h is the horizontal transform via recursive filters
- U_v is the vertical transform via EOF eigen-decomposition
- U_p is the physical transform from control variables to model variables with regression coefficients

The inverse transformations of above are calculated in `gen_be`, which consists of 5 operational stages(stage 0 - stage 4).

Stage 0: Converts u, v to the control variables ψ and χ and calculate forecast differences based on the NMC-method. Further the mean is calculated.

Stage 1: Calculate perturbations $\psi', \chi', T', q', ps'$ by removing the mean of stage 0 of the control variable fields

Stage 2: The initial conditions have to fulfill the wind - mass balance to prevent/decrease the model spin-up and permit an optimal data assimilation. The balanced/correlated parts of the variables $\chi_b(k), T_b(k), p_{sb}$ are calculated via statistical regression analysis with the

stream-function ψ as predictor-field

$$\chi_b(k) = c(k)\psi(k) \quad (36)$$

$$T_b(k) = \sum_{k1} G(k1, k)\psi(k1) \quad (37)$$

$$P_{sb} = \sum_{k1} W(k1)\psi(k1) \quad (38)$$

where c, G, W are the regression coefficients. The regression is allowed, because the relations between the stream function and the other meteorological variables are linear. The filtering matrix W is calculated from the correlation between actual pressure forecast differences (used in the NMC-method) and balanced pressure increments derived from the wind forecast difference data (Barker et al. 2003). The factor W is chosen to vary with height and latitude to represent the fact that geostrophic balance is not appropriate in either the tropics or the planetary boundary layer. Studies (Wu et al. 2002) have shown that e.g. the balanced part of the increment of stream-function integrated over all levels accounts for 86% of the variance of surface pressure increments.

For the error covariances the balanced/correlated components are not essential and therefore neglected/ subtracted from covariance calculation.

$$\chi_u(k) = \chi'(k) - c(k)\psi(k) \quad (39)$$

$$T_u(k) = T'(k) - \sum_{k1} G(k1, k)\psi(k1) \quad (40)$$

$$P_{su} = P_s' - \sum_{k1} W(k1)\psi(k1) \quad (41)$$

Hence the unbalanced part of the variables χ_u, T_u, P_s is the difference between full and balanced increment of the control variables. These results are input for the spatial covariance calculation of the next gen_be stage.

Stage 3: removal of the mean of χ_u, T_u, P_s . For each control variable a computation of global (domain-averaged) and local eigenvectors and eigenvalues via eigen decomposition $B = \epsilon\epsilon^T = E \Lambda E^T$ is made. Then a projection from the entire three dimensional control variable field onto the EOFs is performed. That yields a significant CPU-saving during minimization, because of the projection onto the uncorrelated eigenvectors.

Stage 4: horizontal correlations are computed between grid-points of each 2D field, binned as a function of distance. A Gaussian curve is then fitted to the data as described in Barker et al. 2004 to provide correlation length scales for use in the recursive filter algorithm (Skamarock et al. 2008). The influence is local and limited to its neighboring grid points (depending on the length scale of the filter).

4 Observation Preprocessing

Before observations can be ingested by WRFVAR, preprocessing and quality control is performed for station data as well as for radar data to ensure that the best possible input data is available for data assimilation.

4.1 Station Data Preprocessing

400 weather stations all over Europe are available for hourly ingestion by WRFVAR in this work. Relevant meteorological parameters are pressure, wind, temperature, relative humidity and height. The observations must be converted into a proprietary *Ascii* data format and are read subsequently from *Obsproc*, a WRFDA tool for observation preprocessing. *Obsproc* calculates momentum components u and v out of wind and wind direction and applies quality control e.g. stations getting rejected if their altitude differs more than 100 meters from the model orography. If both pressure and height are missing, the whole data is discarded. A further important capability of *Obsproc* is the assignment of observation errors to the observations, which are not available at the time this work was created. This originates due the complexity of the calculation of the observations errors and will be a future issue of research. The output of *Obsproc* deals as input for *da_wrfvar*.

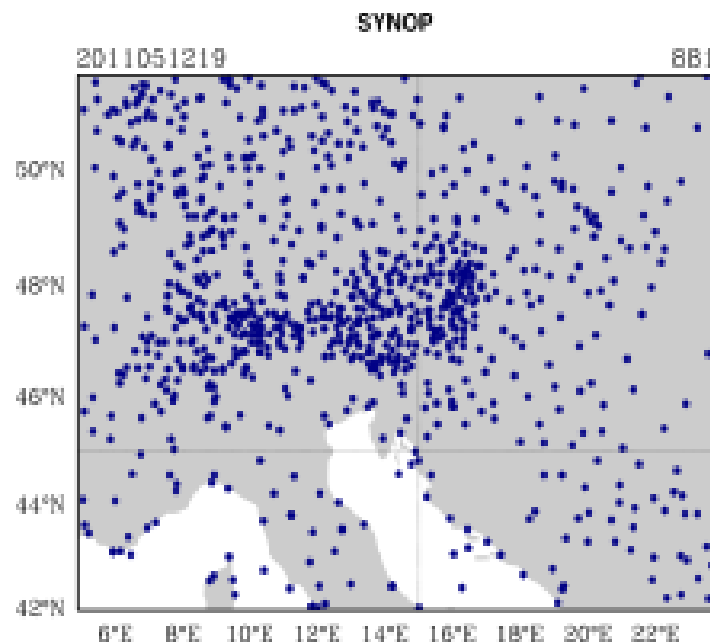


Figure 7: Station data distribution over Europe

4.2 Radar analysis

The data used for radar data assimilation consists of

- MAX-CAPPI CERAD (Central European Radar Network) reflectivity data, with a spatial resolution of 2km and a temporal resolution of 15min (Fig. 8)
- Austrian ACG (AustroControlGroup) radar reflectivity data located at Patscherkofel (Tyrol), Zirbitzkogel (Styria), Mattighofen (Salzburg) and Schwechat(Lower Austria), with a resolution of 1 km and 16 vertical levels, temporal resolution of 5min
- 400 weather stations all over Central Europe, that are used to generate BOGUS-RADAR data, temporal resolution of 10min
- Reflectivity of model background for 2D to 3D expansion of CERAD

The benefit of radar data compared to weather stations is the much higher spatial(1 km) and temporal resolution(5 min). In fact a lot of preprocessing is necessary to create a homogeneous and roughly complete reflectivity field. Reasons of errors and inconsistency in the data are

- Shadowing effects of the Alps
- Ground Clutter and other noise
- Missing elevations in raw data
- Radar device unavailability

To receive an optimal data preparation of measured radar reflectivity a C++ merging routine is developed, where CERAD data, ACG-AT radar and BOGUS data are combined to one final radar analysis for ingestion by WRFVAR.

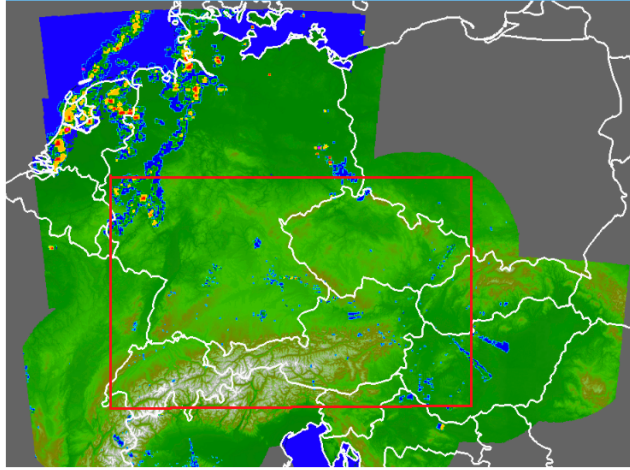


Figure 8: CERAD domain, red rectangle: domain of interest for radar assimilation

The raw data from ACG come in binary format in 16 quantised steps in units mm/hour. The Z-R equation

$$Z = 10 * \log(200 * R^{1.6}) \quad (42)$$

provided by AustroControl, is used to convert the precipitation variable R to a reflectivity Z . Due to the different resolutions of the radar data, data thinning is applied to fit the data with the model mesh (12 and 4km). Erroneous radar regions (shadowing effects over the Alps or missing radar stations) are getting corrected with BOGUS radar data derived from 10 minutes interval station-observations of precipitation by an empirically found Z-R equation,

$$Z = 2 + 10 * \log(300 * R^1), \quad (43)$$

which is based on the fact, that a converted reflectivity of measured 0.1mm-0.3mm precipitation would lead to too low reflectivities calculated with Marshall Palmer Z/R. If instead of ten minutes station data, one hour station data is used for BOGUS data creation, fast moving systems would generate large areas of reflectivity, falsifying the current image of precipitation. Next a linear interpolation(inverse distance rating) is used to create areas of BOGUS reflectivity as shown in Fig.10 and 11.

Another major issue of preprocessing are the missing vertical level data of the ACG Radar. A simple interpolation is used in order to fill the missing values with interpolated values from the neighboring upper and lower levels.

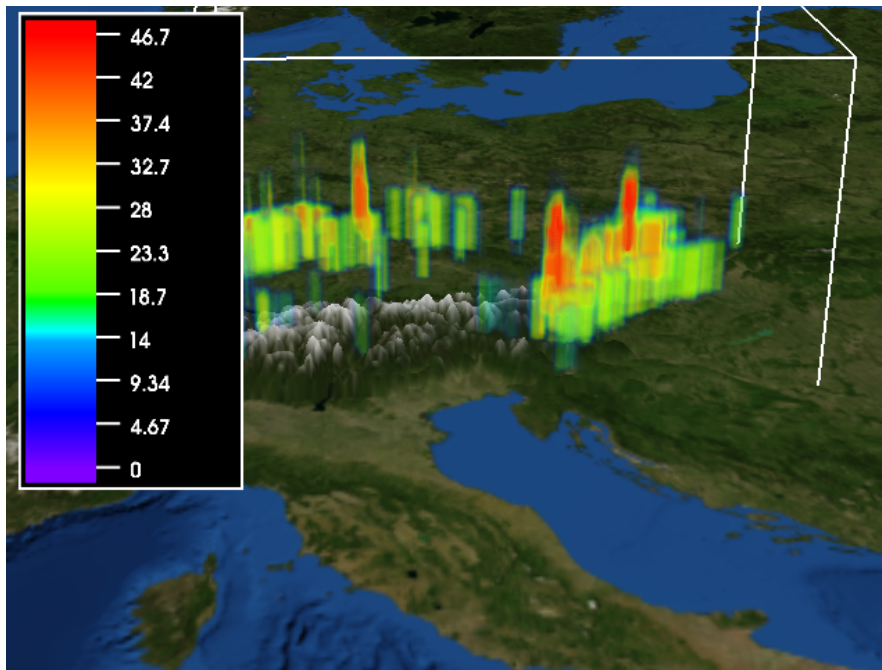


Figure 9: 2011.08.19 12Z: Three-Dimensional Reflectivity[dbZ] over 4km x 4km domain after Preprocessing; Viewing angle: from Italy to the Alps; plotted with NCAR's VA-POR(Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers) Engine

As next step artificial radar vertical profiles are created in order to expand CERAD from 2 dimensions to 3 dimensions. In the beginning of this work the profiles were created on base of averaged vertical reflectivity distributions at different latitudes (Troxel and Engholm 1990). Later on an analysis of a model reflectivity field from a different WRF without data assimilation was applied, averaging the echo tops over an user-defined grid area(10x10 grid points) and use this data as vertical profiles. This method is showing more reliable results. If no reflectivity is found in the background field the artificial profiles are used. In the end a noise correction is performed deleting all single reflectivity data points. In the following graphics the merging procedure is illustrated for the coarse and the fine domain. One can clearly see, how the missing CERAD data over lower Austria get replaced by BOGUS and Austrian radar data.

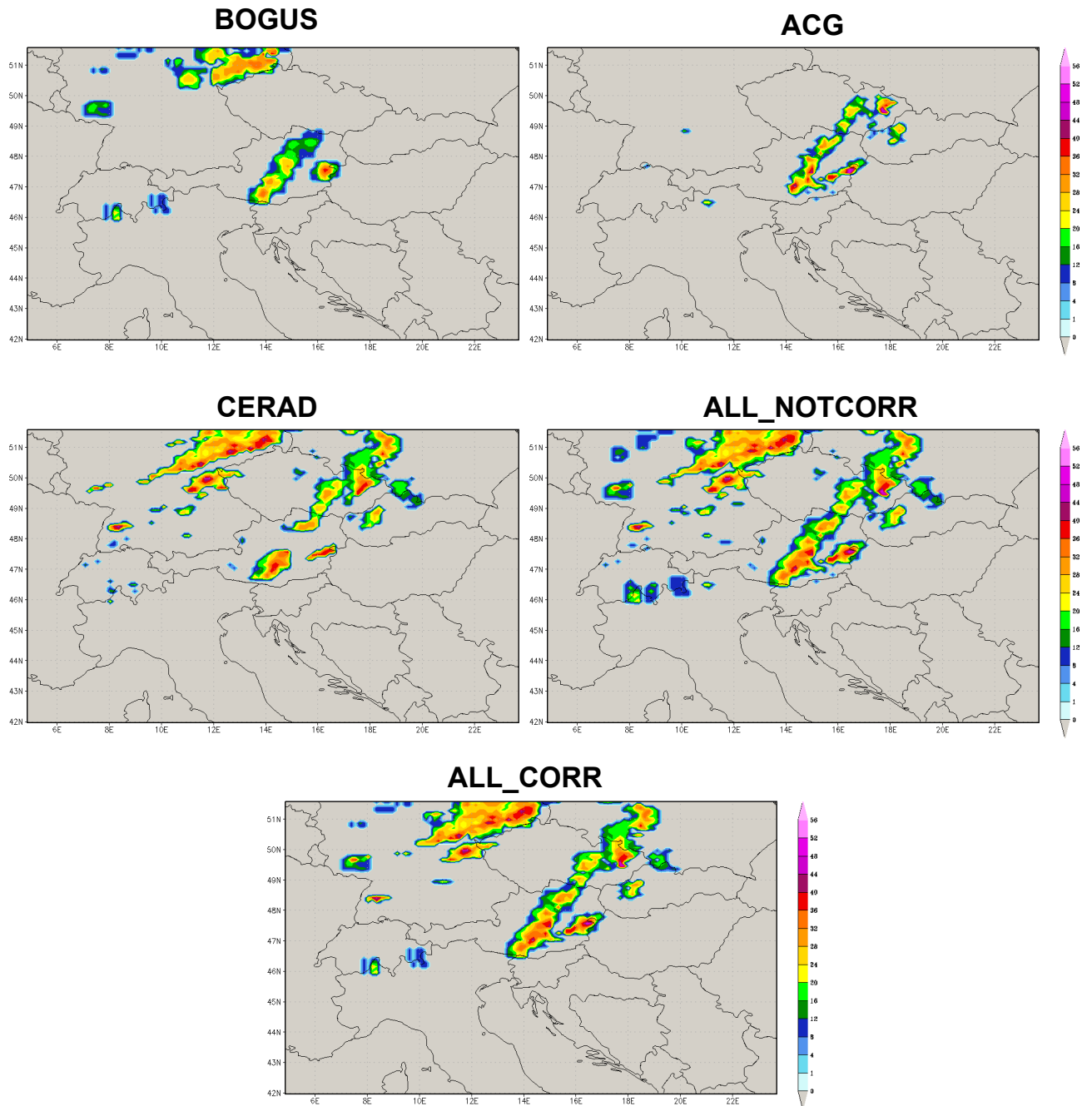


Figure 10: Radar analysis merging procedure with BOGUS+ACG+CERAD uncorrected (ALL_NOTCORR) and final correction of noise echoes(ALL_CORR) on the 12km x 12km domain, date: 2011.08.19 12Z, scale is reflectivity (dBZ)

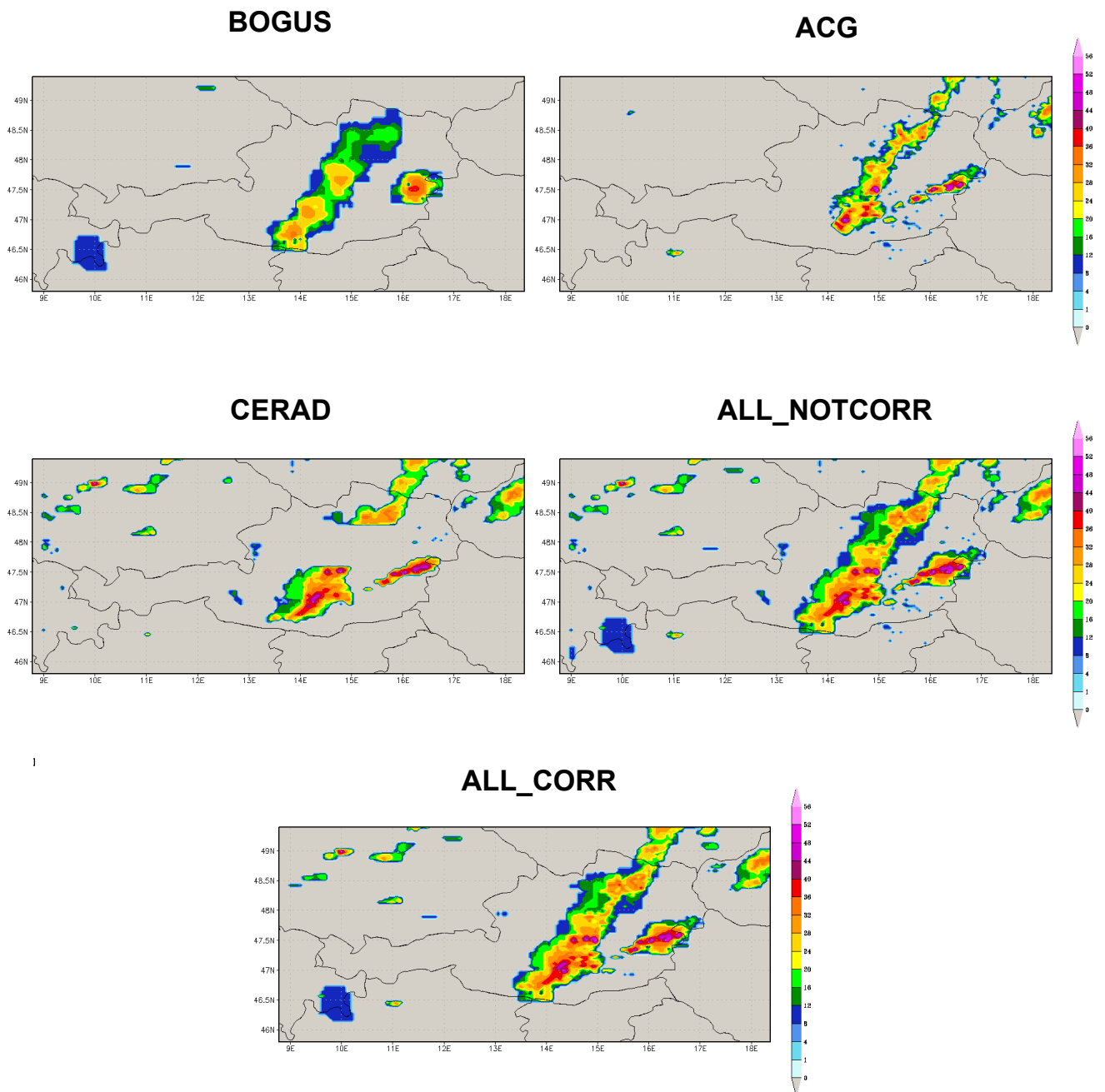


Figure 11: Radar analysis merging procedure with BOGUS+ACG+CERAD uncorrected (ALL_NOTCORR) and final correction of noise echoes(ALL_CORR) on the 4km x 4km domain, date: 2011.08.19 12Z, scale is reflectivity (dBZ)

5 Reflectivity Assimilation in WRFVAR

The aim of reflectivity factor assimilation is to retrieve the temperature (buoyancy) field through partitioning total water increment through microphysical processes. The on or off switch of processes in microphysics is critically important and is triggered through a warm rain scheme.

5.1 Forward Operator

The observation operator $H(x)$ for reflectivity Z (Sun and Crook 1997) is assuming a Marshall Palmer drop size distribution and is written

$$Z = 43.1 + 17.5 \log(\rho q_r) \quad (44)$$

where Z is only function of ρ the air density (kgm^{-3}) and q_r the rainwater mixing ratio (gkg^{-1}), which means that no ice phases are contributing to Z . $H(x)$ is used to transform a radar observation Z from observation space to model space q_r and vice versa. The moisture control variable for the radar assimilation is total water mixing ratio q_t instead of q_v in the normal WRFVAR control space. Therefore a physical transformation U_p is introduced for partitioning q_t in water vapor mixing ratio q_v , rain water mixing ratio q_r and q_c cloud water mixing ratio. This partitioning process is performed by a simple warm rain microphysics scheme (no ice phases included - Sun and Crook 1997). The interactions between the mixing ratios are shown in the next image.

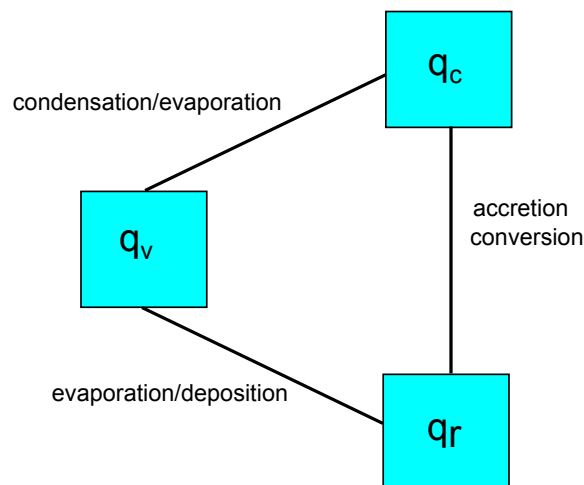


Figure 12: WRFVAR microphysics processes

The warm rain scheme parameterizes the following microphysical processes:

- evaporation of raindrops in subsaturated air
- autoconversion of cloud to rain
- accretion of cloud by rain
- sedimentation of rain

The work flow of the radar data assimilation in WRFVAR is pictured on a further image.

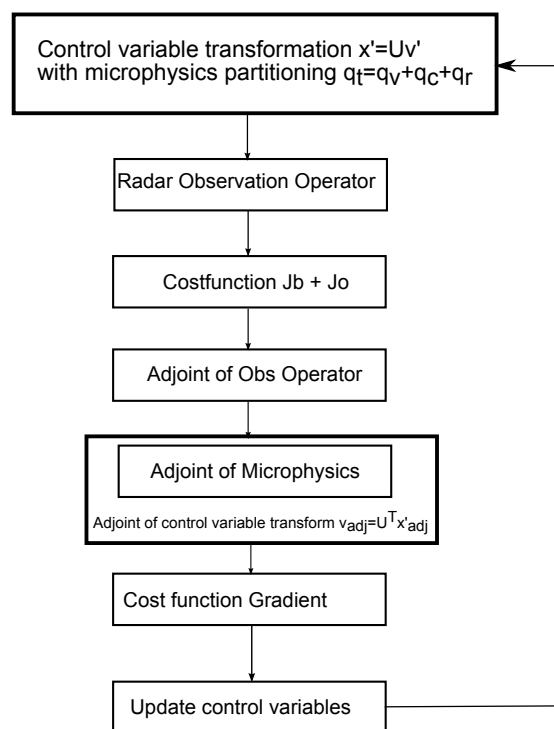


Figure 13: Radarassimilation flowchart

The warm rain scheme works like a one-dimensional column-model (Xiao et al. 2007). During the minimization iteration procedure reflectivity information gets distributed to the other corresponding variables (q_c, q_v and T). At the beginning of the minimization the q_r and q_c increments are zero. If supersaturation is identified in the column cloudwater and rainwater is created. This leads to moisture and temperature change because of the condensation and evaporation processes included in the warm rain scheme. It is noteworthy that zero derivatives of the cost function are inducing computational problems, when q_r gets very small and the gradient becomes large. A minimum threshold for radar reflectivity data was implemented in preprocessing to avoid this problem.

It is obvious that this approach is not a perfect solution, because of the missing ice phases in the warm rain process. This weak point could be solved in future by implementation of the more complex WSM6 scheme in WRFVAR.

Sugimoto et al. (2009) point out, that it is advantageous to perform a cloud analysis before the WRFVAR reflectivity assimilation is processed. The WRFVAR assimilation is only able to insert convection when reflectivity data exists in the area. It cannot reduce or correct a model state if convection is in the background, but was not measured in reality. In the cloud analysis the model q_r gets blended with the observational q_r . If the observational q_r is under a distinct threshold, the model q_r is set to zero. This is important because wrong evaporation effects are taken out, if q_r is zero. The model stays in a consistent physical state, because the internal thermodynamics are not changed. The use of a simple cloud analysis leads to a more realistic atmospheric model state and therefore should improve the score of the model. The cloud analysis is a promising method, but was not implemented at the time the work was written.

5.2 Single OBS experiments

Single observations experiments are realized to determine the spatial and quantitative impact of observations controlled by the error covariances on the analysis. One assimilates an artificial observation variable to get an insight, how the different model variables are interacting with each other. After the WRFVAR minimization procedure has created an analysis, the increments are plotted in multiple ways (horizontal and vertical visualization). The background error covariances generated with the NMC-method can be tuned with variance scaling and/or length scaling. Each control variable has different spatial and quantitative impacts on the analysis, which makes clear how important and time consuming efficient background error tuning can be. Through variance scaling one can increase the influence of the observation by decreasing the influence of the background and vice versa. This results in a higher magnitude of the analysis increments. Length scaling is used to control how the analysis increments are spread horizontally. Dozens of single observation experiments and real cases with different variance scalings and length scalings were tested to find the optimal WRFVAR setup. An appropriate configuration for an operational WRFVAR forecasting system seems to be variance scaling of 0.5 and a length scaling of 0.1. That indicates that the calculated NMC length scales are too large for accurate reflectivity assimilation and tuning becomes necessary. Similar results are found in Sugimoto et al. (2009), even though they recommend a length scaling of 0.3 and a variance scaling of 0.9, which is not in best agreement with our test cases.

As single observation variable water vapor q is chosen, because it's the most affected variable from the radar reflectivity forward operator during assimilation. To create reasonable

analysis increments from a q innovation a value of 0.01 kg/kg is considered for assimilation. In Fig. 14 and 15 different length scalings (0.1 and 0.5) are shown. Especially the vertical cross sections in Fig. 16-18 are showing explicitly, that the default NMC length scales wouldn't be appropriate for radar assimilation, because the increment is spread too wide in horizontal direction. The increments of the vertical wind component w in Fig. 18 are indicating areas of lifting and sinking air masses and are spread over a larger part of the air column due to mass continuity reasons.

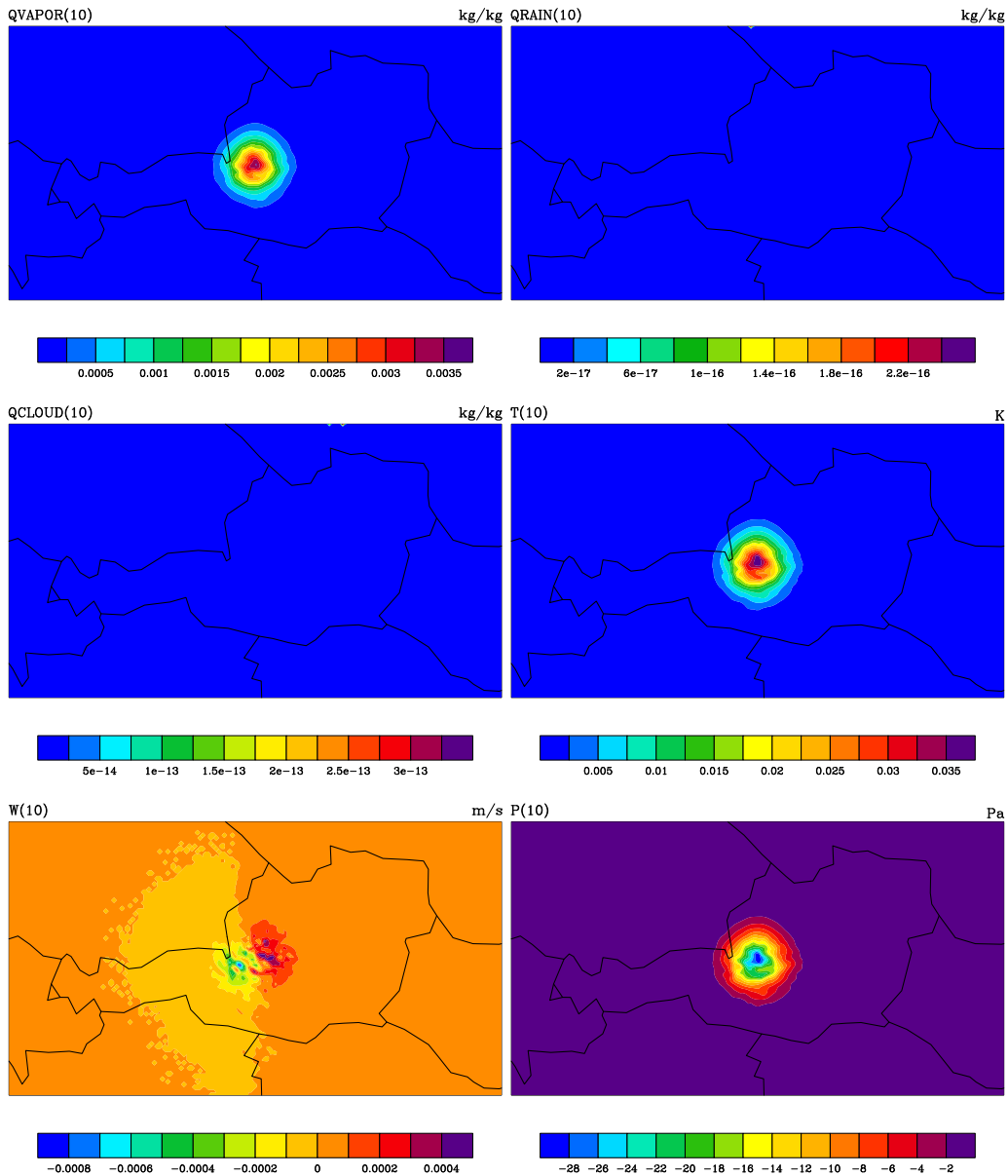


Figure 14: Single Obs experiment with len_scaling=0.5 and var_scaling=0.5, innovation variable= q , innovation value=0.01 kg/kg, inserted at eta-level=10(700hPa)

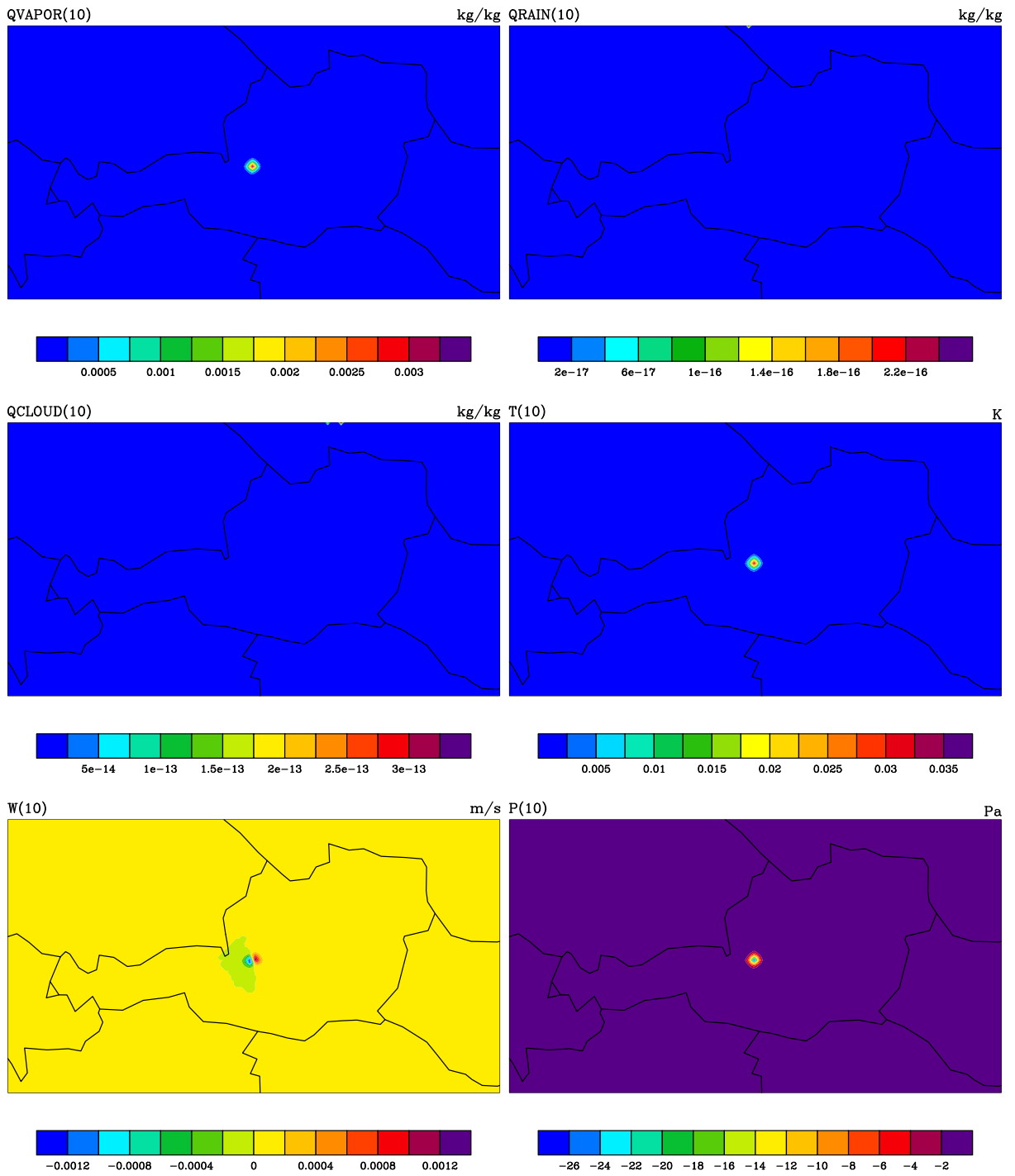


Figure 15: Single Obs experiment with len_scaling=0.1 and var_scaling=0.5, assimilated variable=q, innovation value=0.01 kg/kg, inserted at eta-level=10(700hPa)

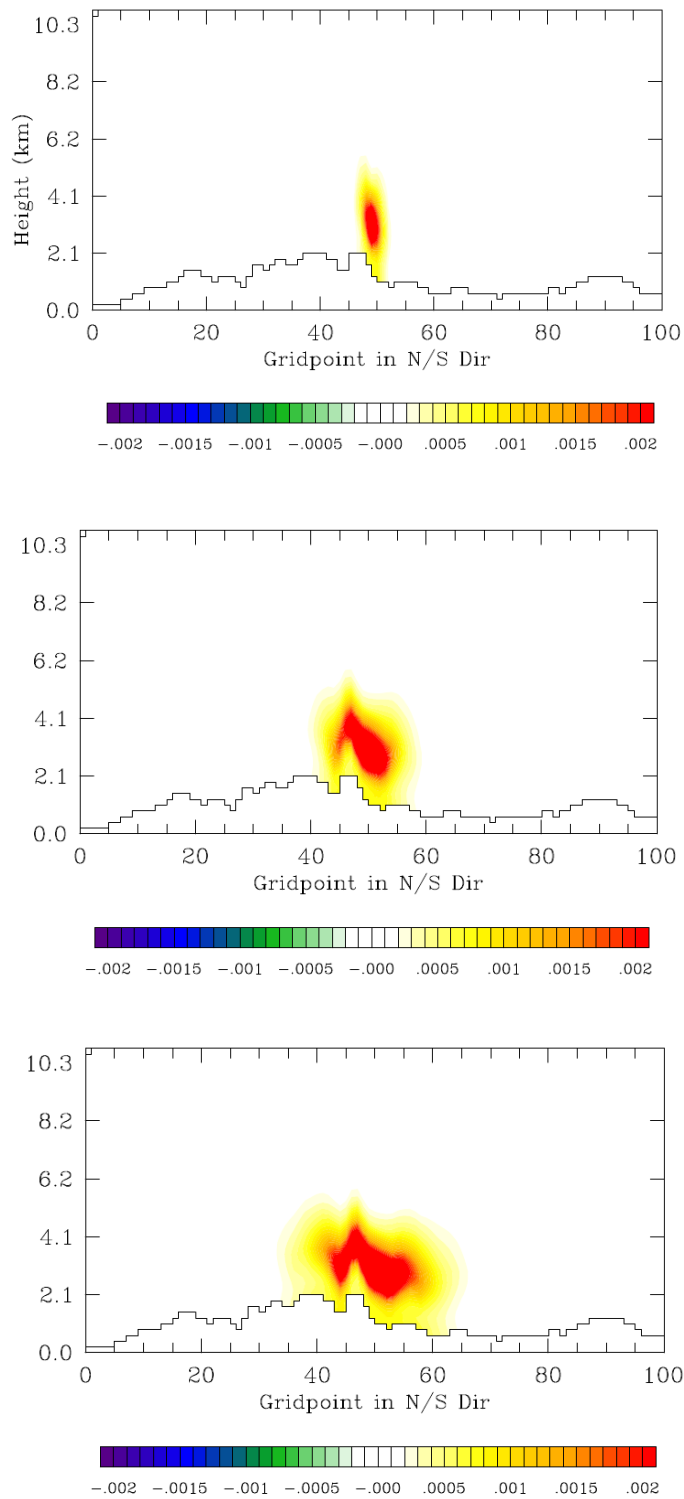


Figure 16: Vertical cross section through fine domain: longitude=constant(domain center), latitude=from south to north
 Analysis increments of water vapor q in kg/kg with $len_scaling=0.1$ (upper), $len_scaling=0.3$ (middle) and $len_scaling=0.5$ (lower)

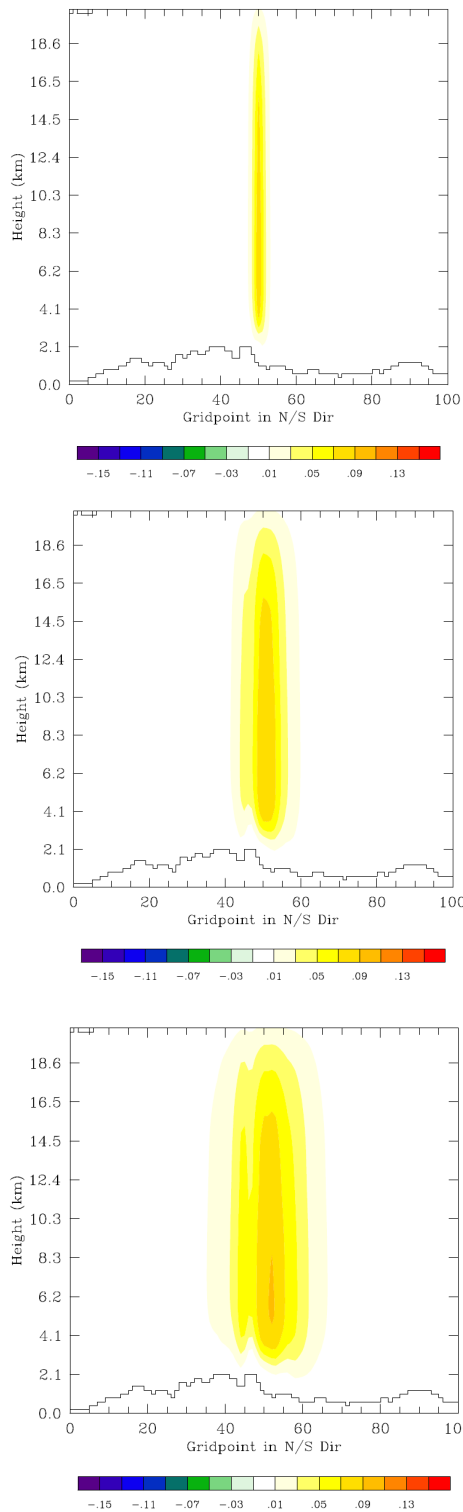


Figure 17: Vertical cross section through fine domain: longitude=constant(domain center), latitude=from south to north

Analysis increments of temperature T in K with $len_scaling=0.1$ (upper), $len_scaling=0.3$ (middle) and $len_scaling=0.5$ (lower)

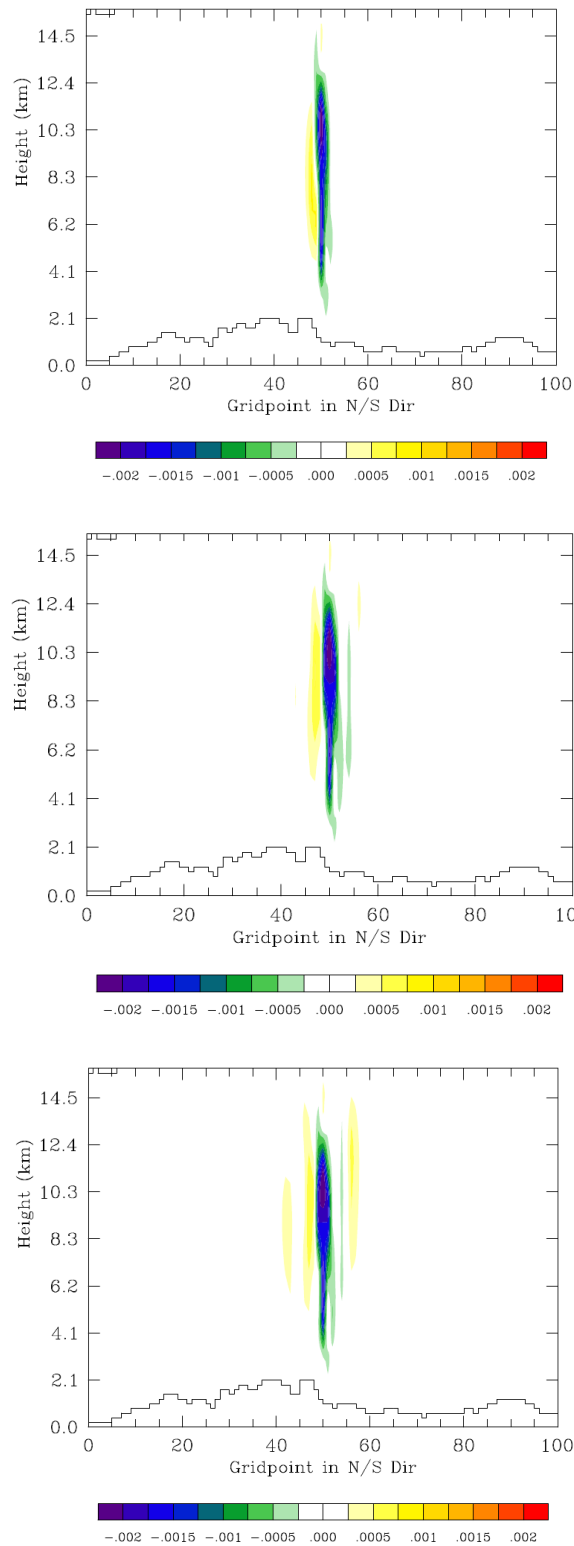


Figure 18: Vertical cross section through fine domain: longitude=constant(domain center), latitude=from south to north
 Analysis increments of the vertical wind component w in m/s with $len_scaling=0.1$ (upper), $len_scaling=0.3$ (middle) and $len_scaling=0.5$ (lower)

6 Case study 19.08.2011

On 19th August 2011 a low pressure system moved from Northern Germany to South Scandinavia. The corresponding coldfront moved eastward over Austria, producing an organized line of in the beginning weak thunderstorms. As the front approached the south-east region of Austria, labilisation took place and widespread severe thunderstorms developed. The maximum precipitation amount of 31.6 mm/h was measured at the TAWES station Eisenkappel in Carinthia. In the regions where the highest radar reflectivity was measured, hail was reported.

6.1 Analysis increments found in the case study

The increments of the assimilated surface station data can be seen in Fig. 19,20 and 21. A big problem is the rejection of station data due to the difference of the station altitude to the model altitude. It is the reason why in broad areas of the alpine region the analysis can not be improved. That of course is not ideal for a consistent analysis. Nevertheless the station data have a significant impact on analysis. The temperature increments are vertically spread only to a limited extent as show in the vertical cross section in Fig. 21. The increments from moisture assimilation of water vapor q_v are in the order of 10^{-4} kg/kg, hence too small to affect the analysis efficiently. q_c, q_r are not affected. The statistically computed NMC length scales of ingested pressure observations are large compared to others.

The retrieval of the microphysical fields from reflectivity assimilation is shown in Fig. 22,23,24 for the fine domain and in Fig. 25,26 and 27 for the coarse domain. The positive q_v increment in the analysis is a result of an underpredicted q_v in the background, which correlates directly with a positive temperature increment coming from latent heat release through condensation. If q_r is partitioned during the minimization process evaporational cooling is included by the WRFVAR warm rain scheme. The water vapor spreads broader than the rainwater since the length scale of the water vapor is larger than the rainwater. Through the wind and temperature statistical relation in the background covariance, the wind response is obtained(no Figure of u and v available). The response of w (updrafts and downdrafts) appears plausible. The pressure increments are small compared to other variables.

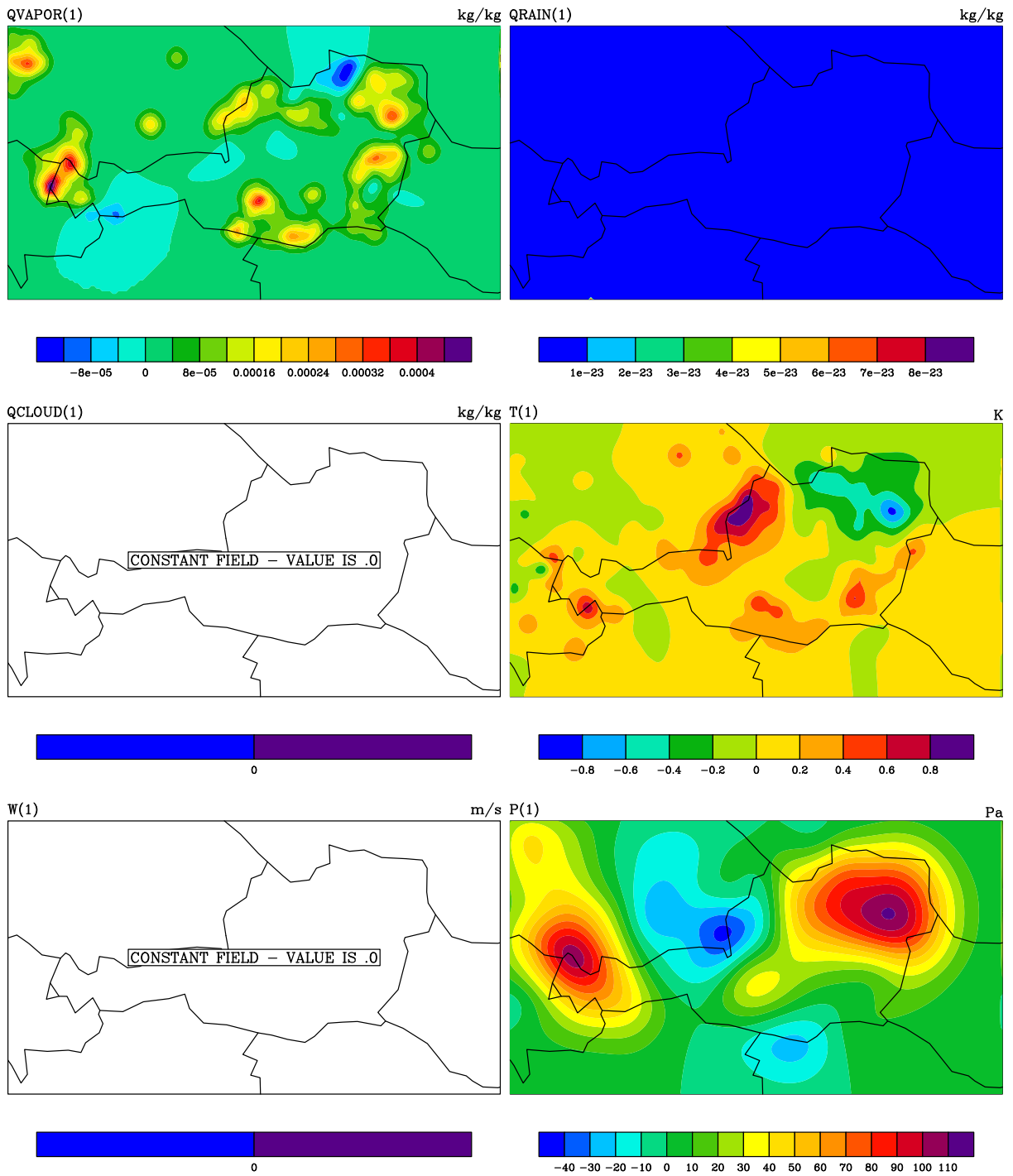


Figure 19: Analysis increments of multiple variables after SYNOP data (t,u,v,rh,p) assimilation on eta level 1(ground) on fine domain with len_scaling=0.3

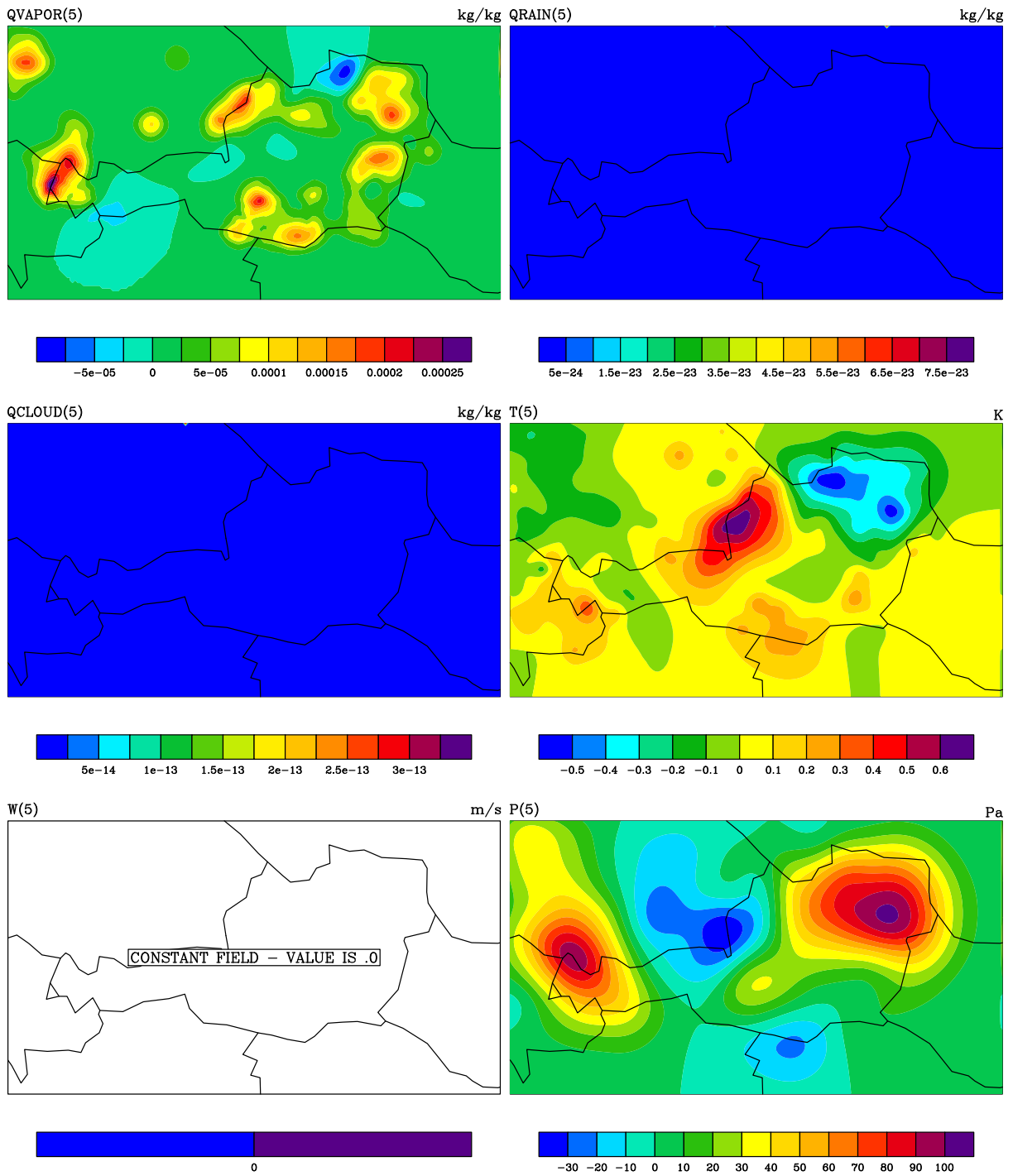


Figure 20: Analysis increments of multiple variables after SYNOP data (t,u,v,rh,p) assimilation on eta level 5(850hPa) on fine domain with len_scaling=0.3

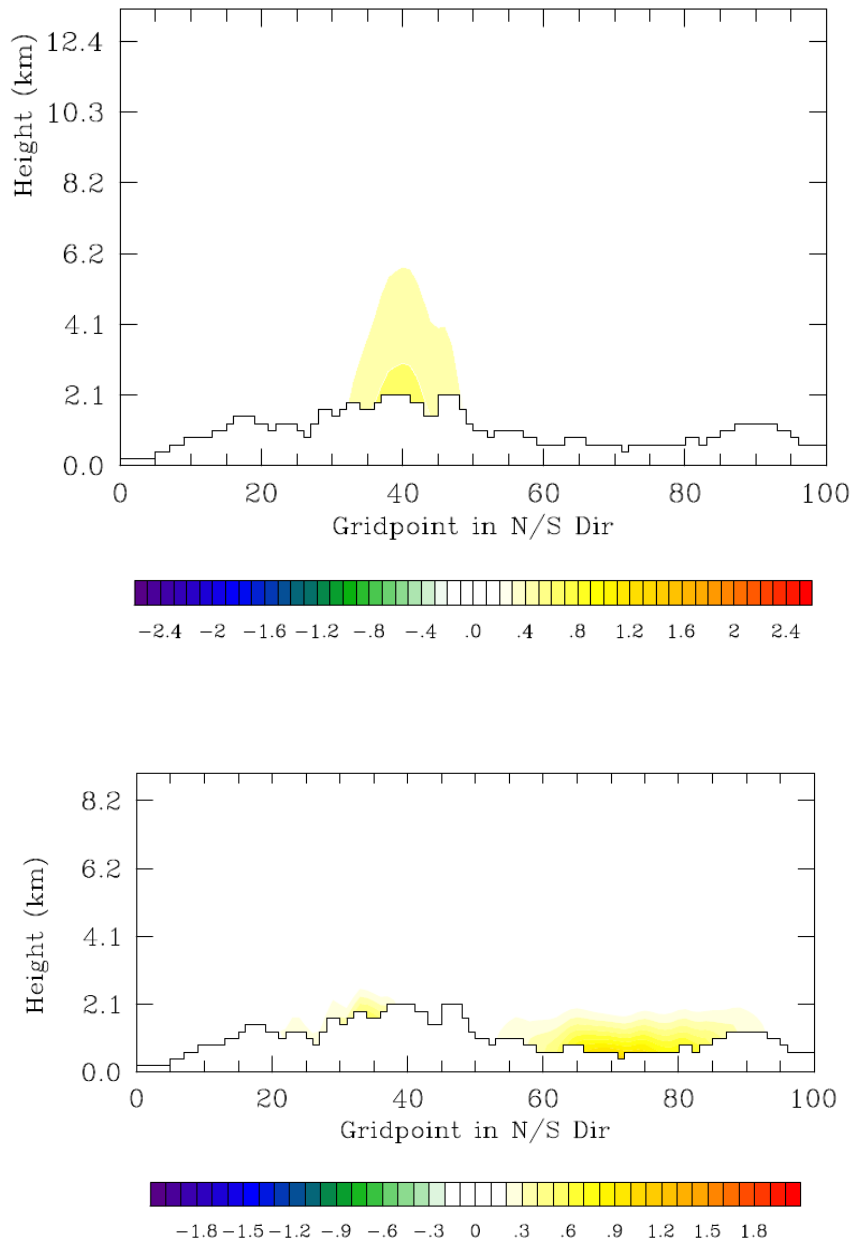


Figure 21: Vertical cross section through fine domain after SYNOP assimilation: longitude=constant(domain center), latitude=from south to north, Analysis increments of Pressure[hPa](upper) and Temperature[K](lower)

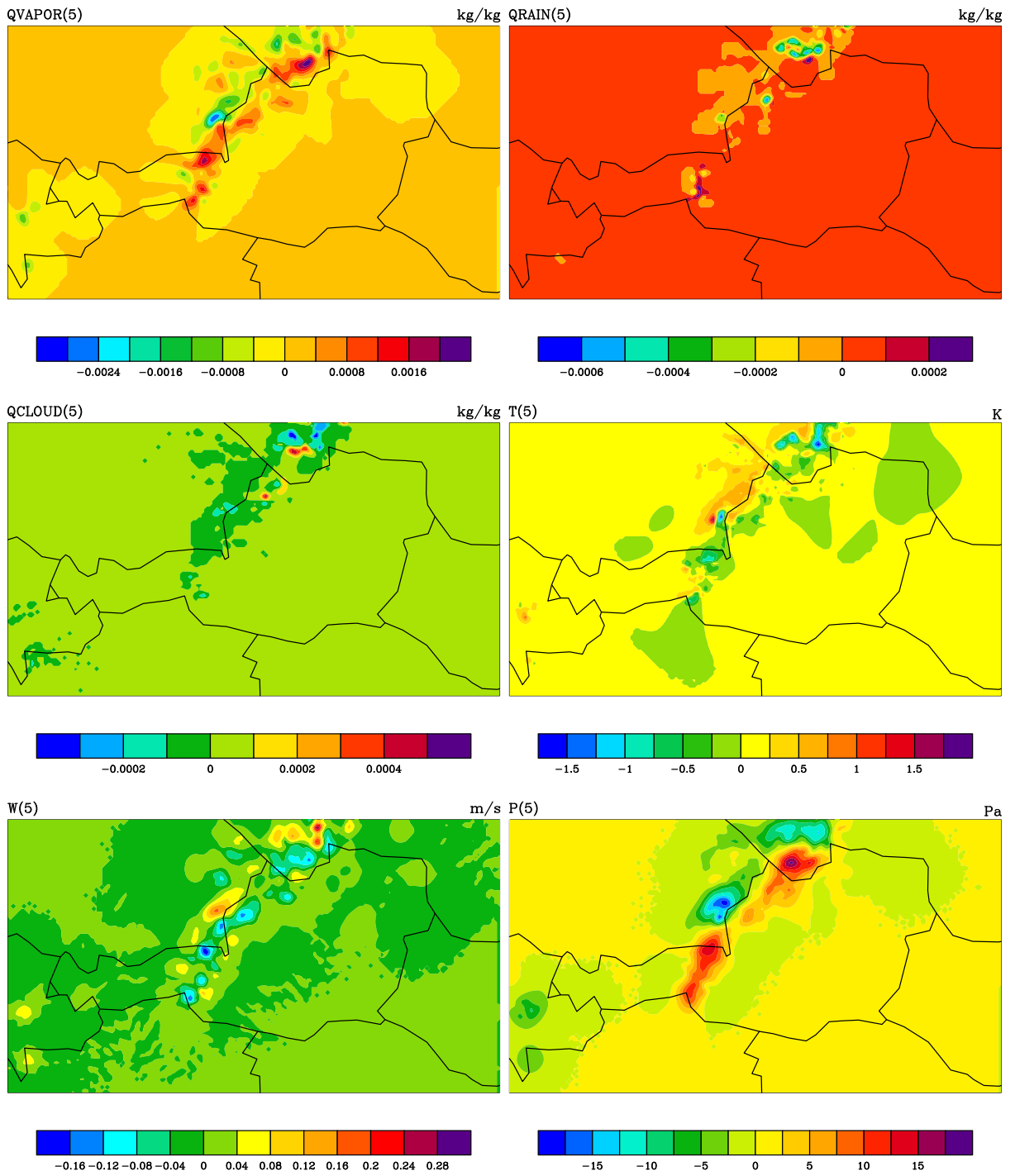


Figure 22: Analysis increments of multiple variables after radar assimilation on eta level 5(850hPa) on fine domain with len_scaling=0.1

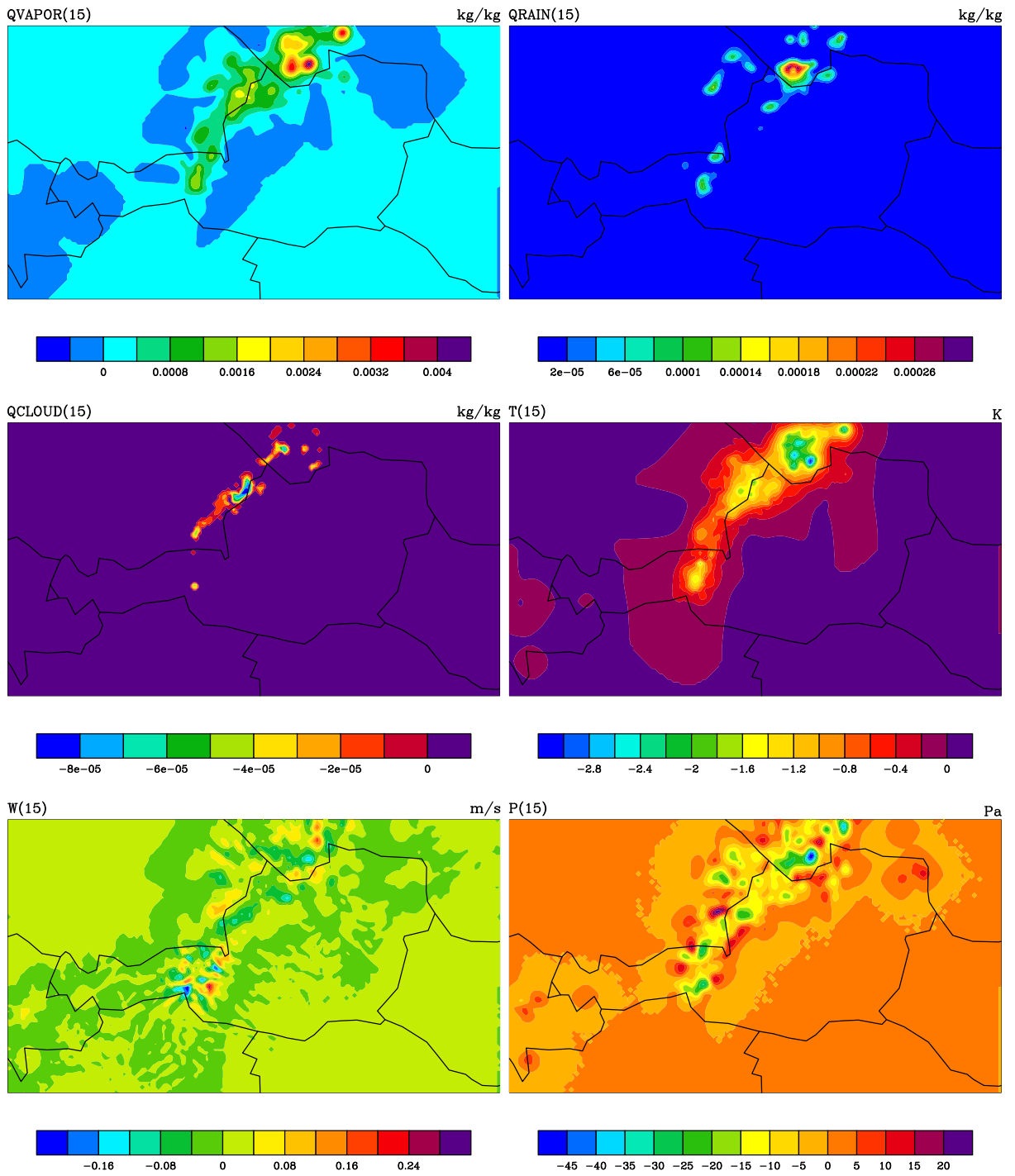


Figure 23: Analysis increments of multiple variables after radar assimilation on eta level 15(500hPa) on fine domain with len_scaling=0.1

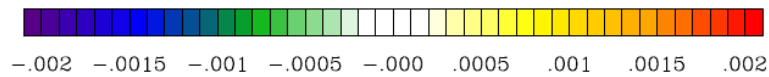
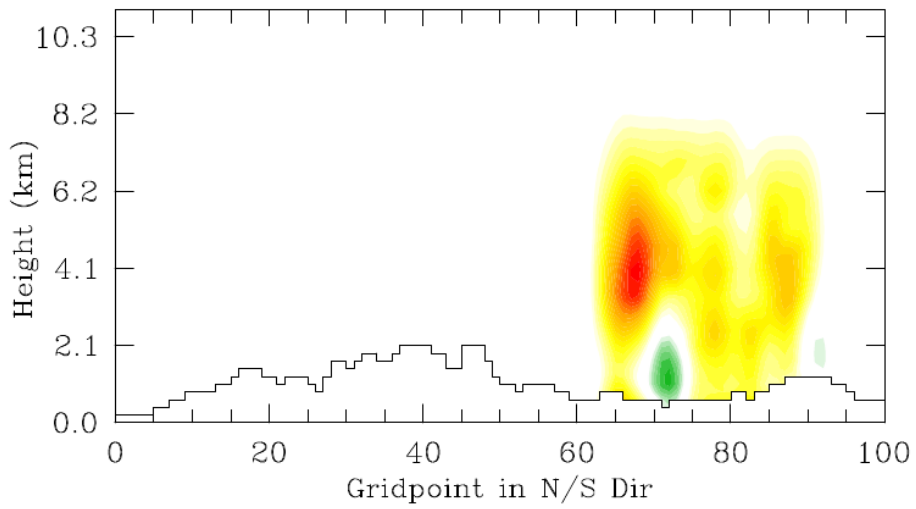
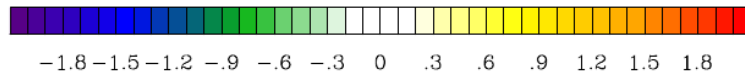
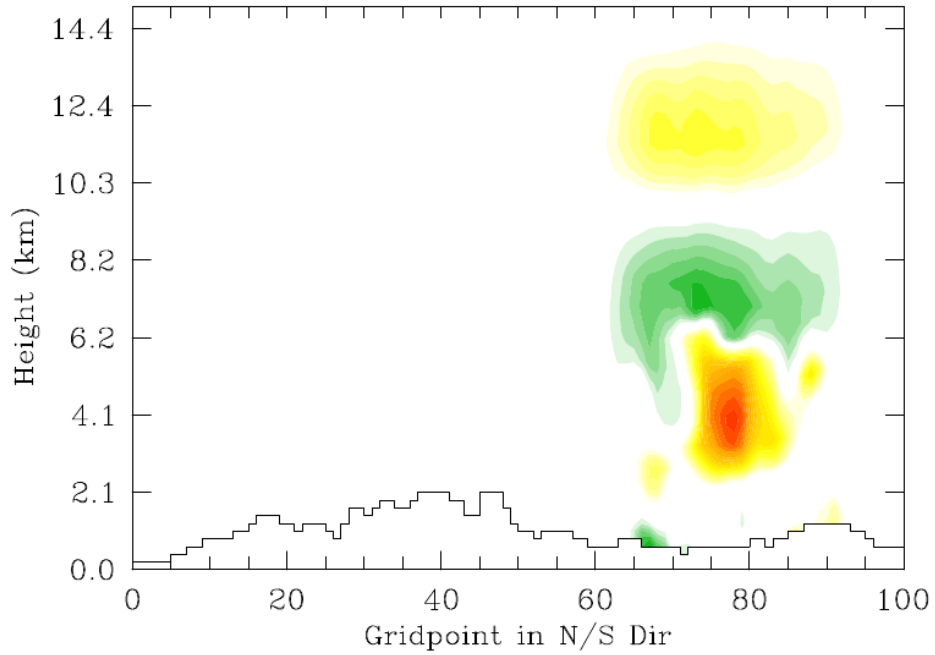


Figure 24: Vertical cross section through fine domain after radar assimilation: longitude=constant(domain center), latitude=from south to north, Analysis increments of Temperature[K](upper) and water vapor[kg/kg](lower)

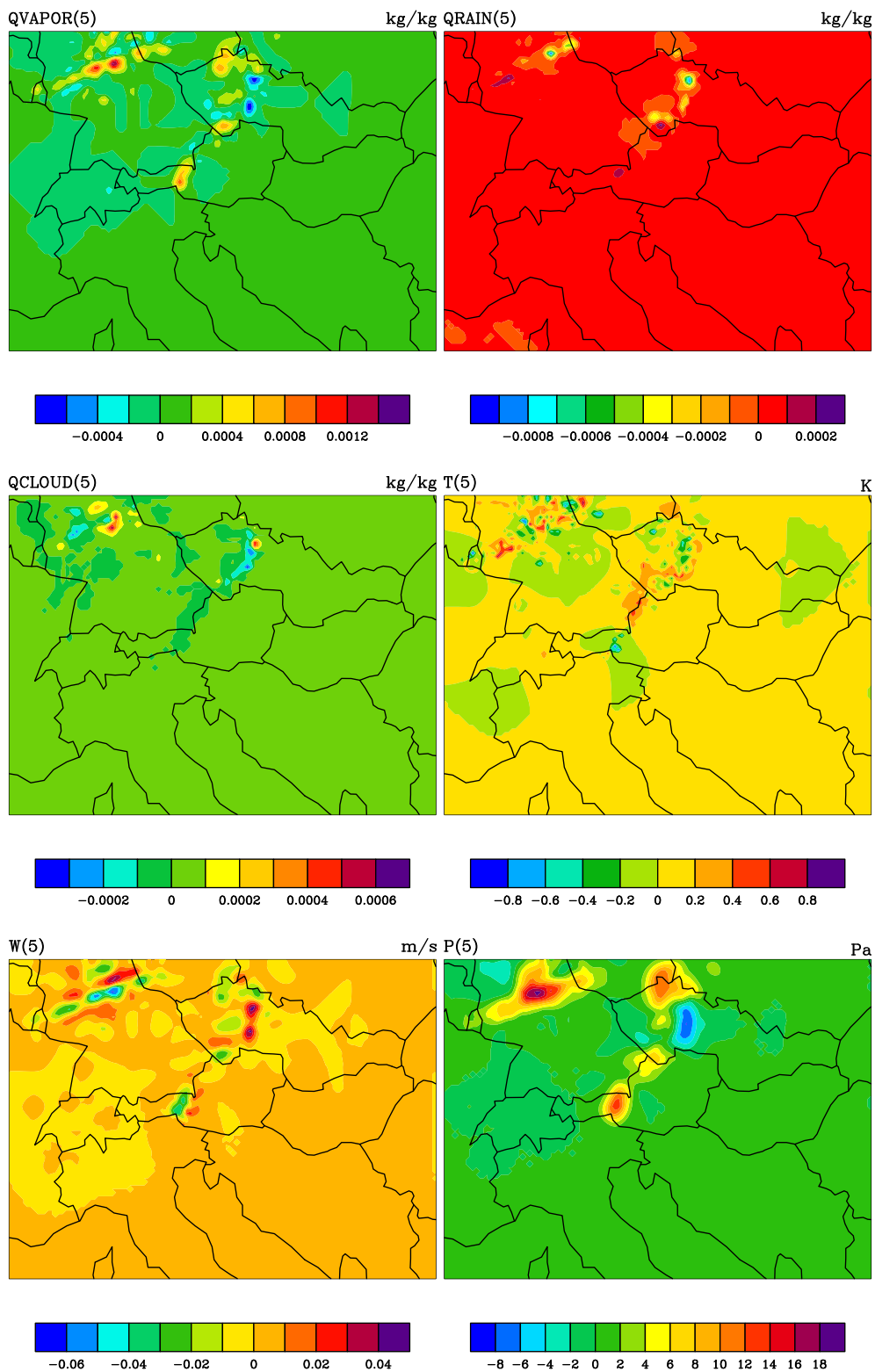


Figure 25: Analysis increments of multiple variables after radar assimilation on eta level 5(850hPa) on coarse domain with len_scaling=0.1

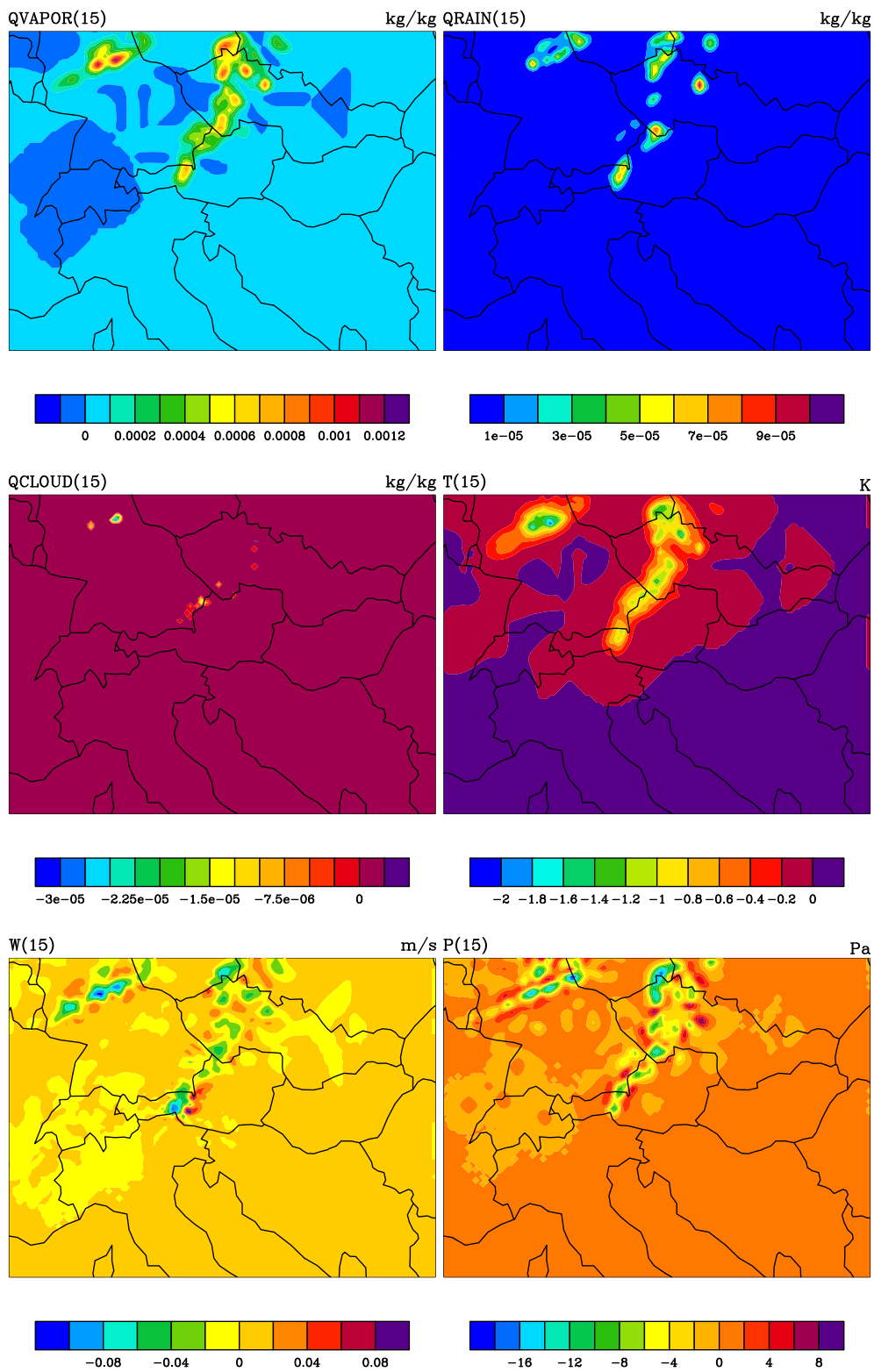


Figure 26: Analysis increments of multiple variables after radar assimilation on eta level 15(500hPa) on coarse domain with len_scaling=0.1

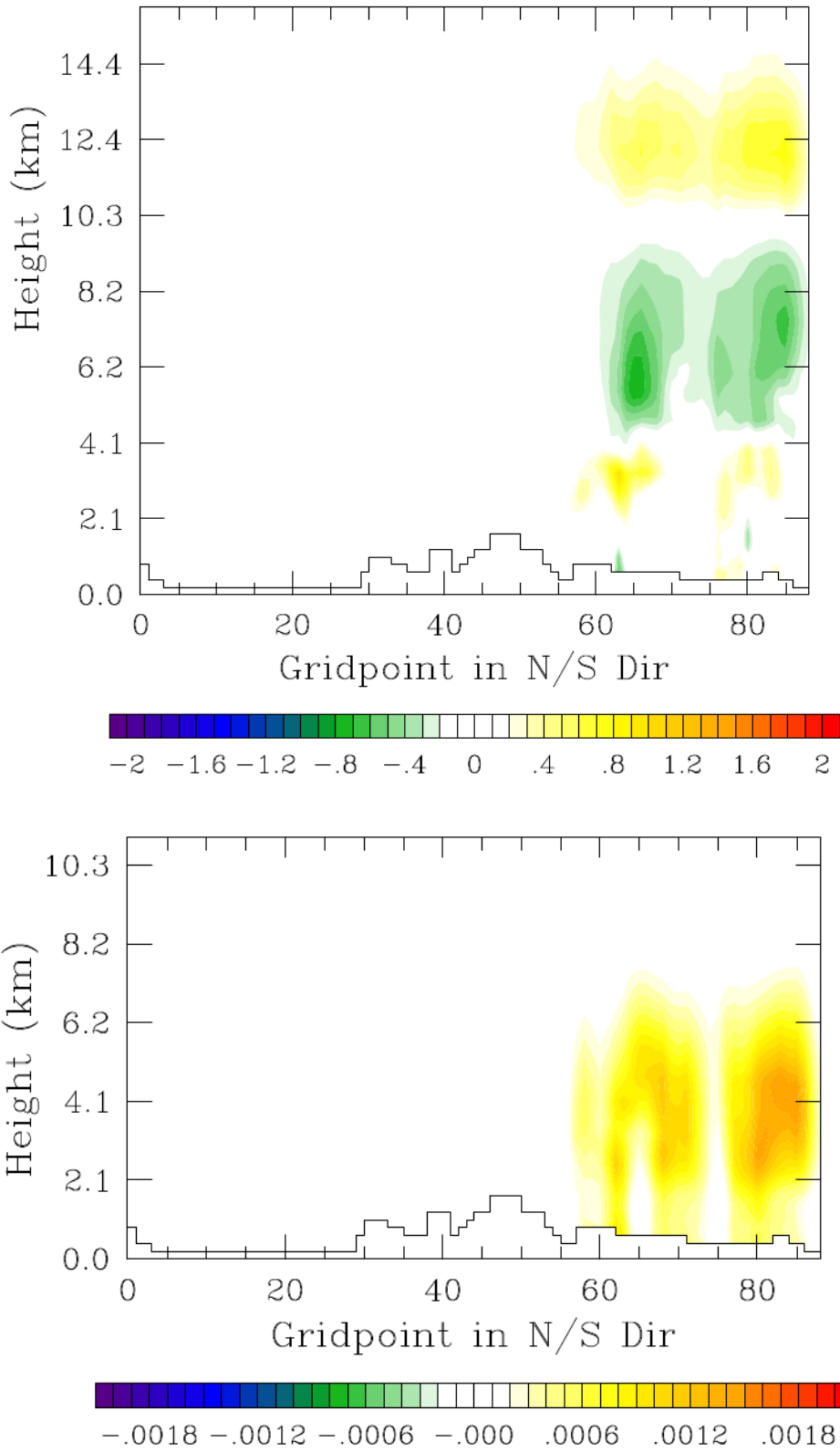


Figure 27: Vertical cross section through coarse domain after radar assimilation: longitude=constant(domain center), latitude=from south to north, Analysis increments of Temperature[K](upper) and water vapor[kg/kg](lower)

6.2 WSM3 vs. WSM6 twin experiment

As shown in the last chapters the configuration of the error covariances is essential for the quality of the forecast. Furthermore the choice of appropriate microphysics, decides if the analysis leads to the optimal forecast (Baker et al. 2007).

In the twin experiment two different WRF forecasts are generated for the same forecast date trying to distinguish the forecast skills of WSM3 and WSM6 microphysics compared to a regular WRF forecast without data assimilation and the "true" radar reflectivity field. The meteorological conditions on the 19th of August 2011 are ideal for this twin experiment. Two identical background error configurations are used with `var_scaling(0.5)` and `len_scaling(0.1)` for radar assimilation and `len_scaling(0.3)` for SYNOP assimilation.

In the Figures 28-31 a qualitative (not quantitative) comparison between "true" radar reflectivity and hourly model precipitation is shown. The regular WRF(without data assimilation) underestimates the amount and localisation of precipitation from the beginning, whereas the cycled 3DVAR WRF assimilates the radar reflectivity properly and captures intensity, shape and movement of the front for the first 2 to 3 hours from INIT almost correctly. The unsatisfying forecast of the regular WRF comes probably from a suboptimal GFS model initialization due to the much coarser resolution of the global model. In Fig. 28 the simple WSM3 microphysics scheme is used and shows a decreasing forecast skill within the third hour (only slight precipitation in the southeast of Austria), whereas the WSM6 in Fig. 29 shows a significant development of thunderstorms. With the more complex WSM6 scheme the severe thunderstorm development in the southeast is captured from WARMSTART INIT 09Z on, whereas WSM3 shows dissipation of the storms after 12Z.

The WARMSTART INIT 12Z(Fig. 30 and 31) shows the ability to create valuable short-range precipitation forecasts with rapid updated cycled WRF using radar assimilation with both microphysics schemes, even though the structure and intensity with WSM6 looks more reliable again.

During testing it turned out, that the ground station data has a big impact on the results of this case study, especially the temperature increments are producing a significant CAPE increase in the south-eastern parts of Austria. At the moment this work is written the WRFVAR quality control rejects the data of many higher situated stations, because the station altitude differs too much from model orography. The revision of this behaviour will be an issue of future research.

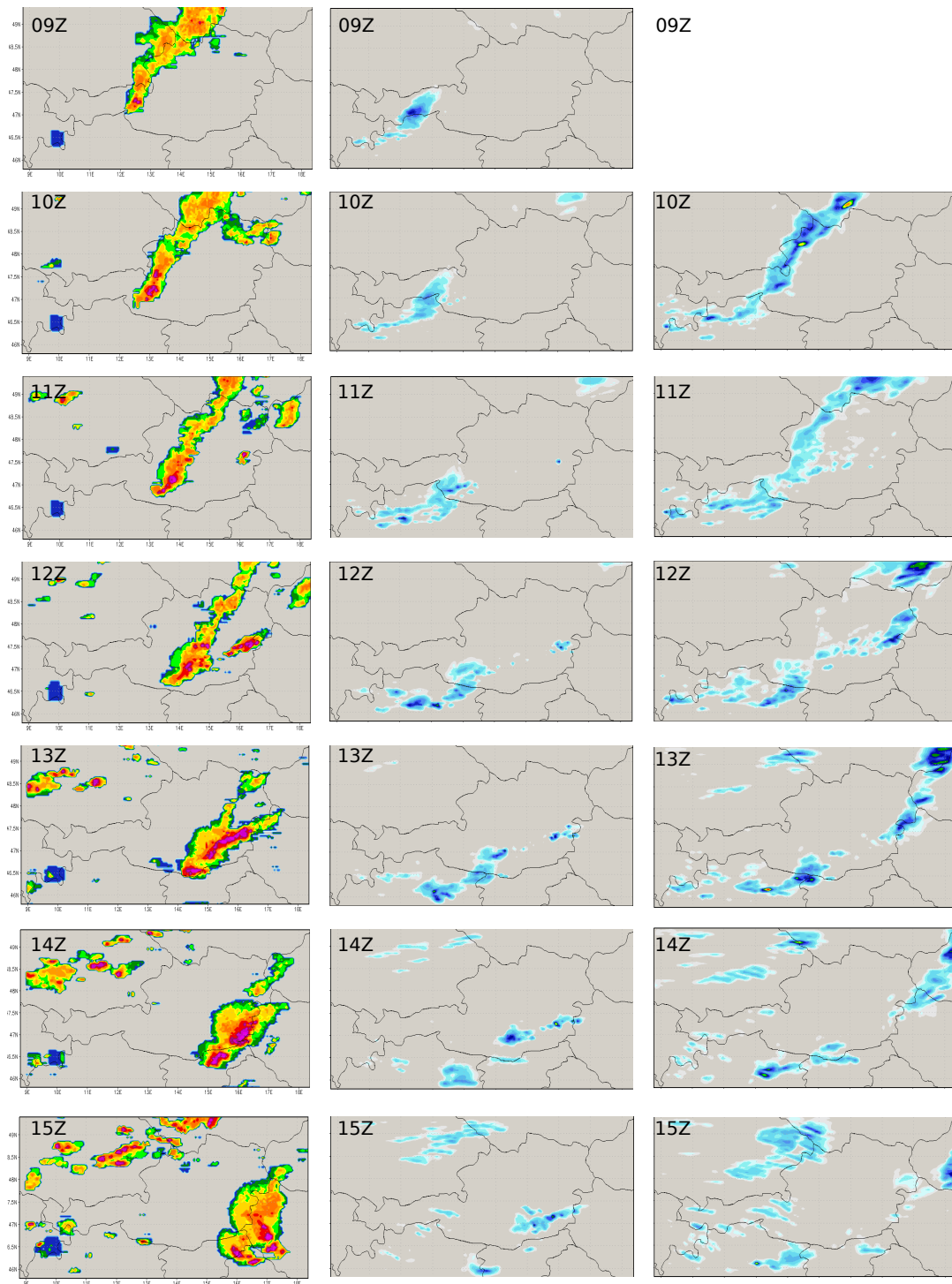


Figure 28: left: "true" radar reflectivity field, center: WRF forecast without assimilation 1 hour accumulated precipitation, right: 09Z INIT WRFVAR with WSM3 microphysics 1 hour accumulated precipitation

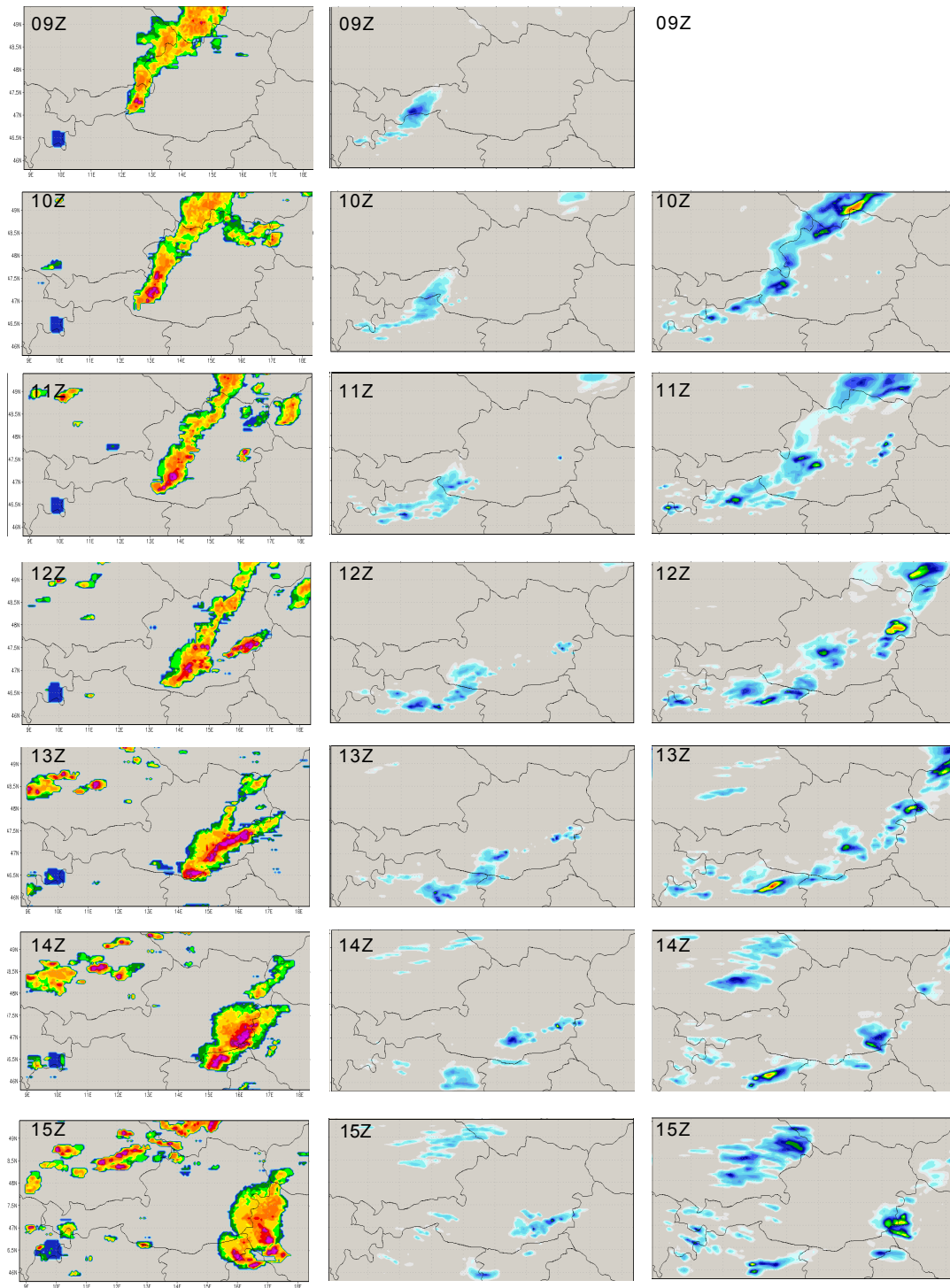


Figure 29: left: "true" radar reflectivity field, center: WRF forecast without assimilation 1 hour accumulated precipitation, right: 09Z INIT WRFVAR with WSM6 microphysics 1 hour accumulated precipitation

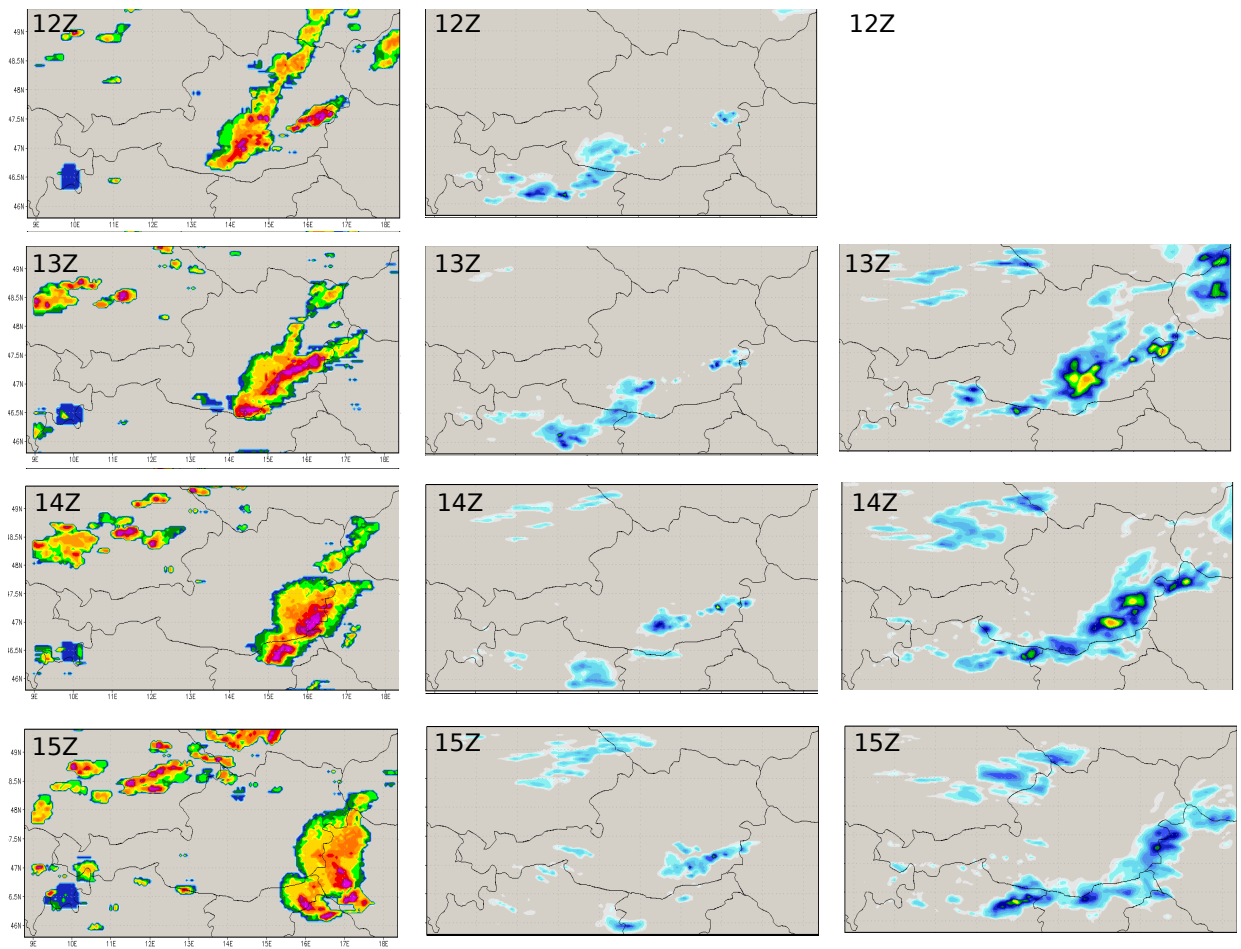


Figure 30: left: "true" radar reflectivity field, center: WRF forecast without assimilation
 1 hour accumulated precipitation, right: 12Z INIT WRFVAR with WSM3 microphysics
 1 hour accumulated precipitation

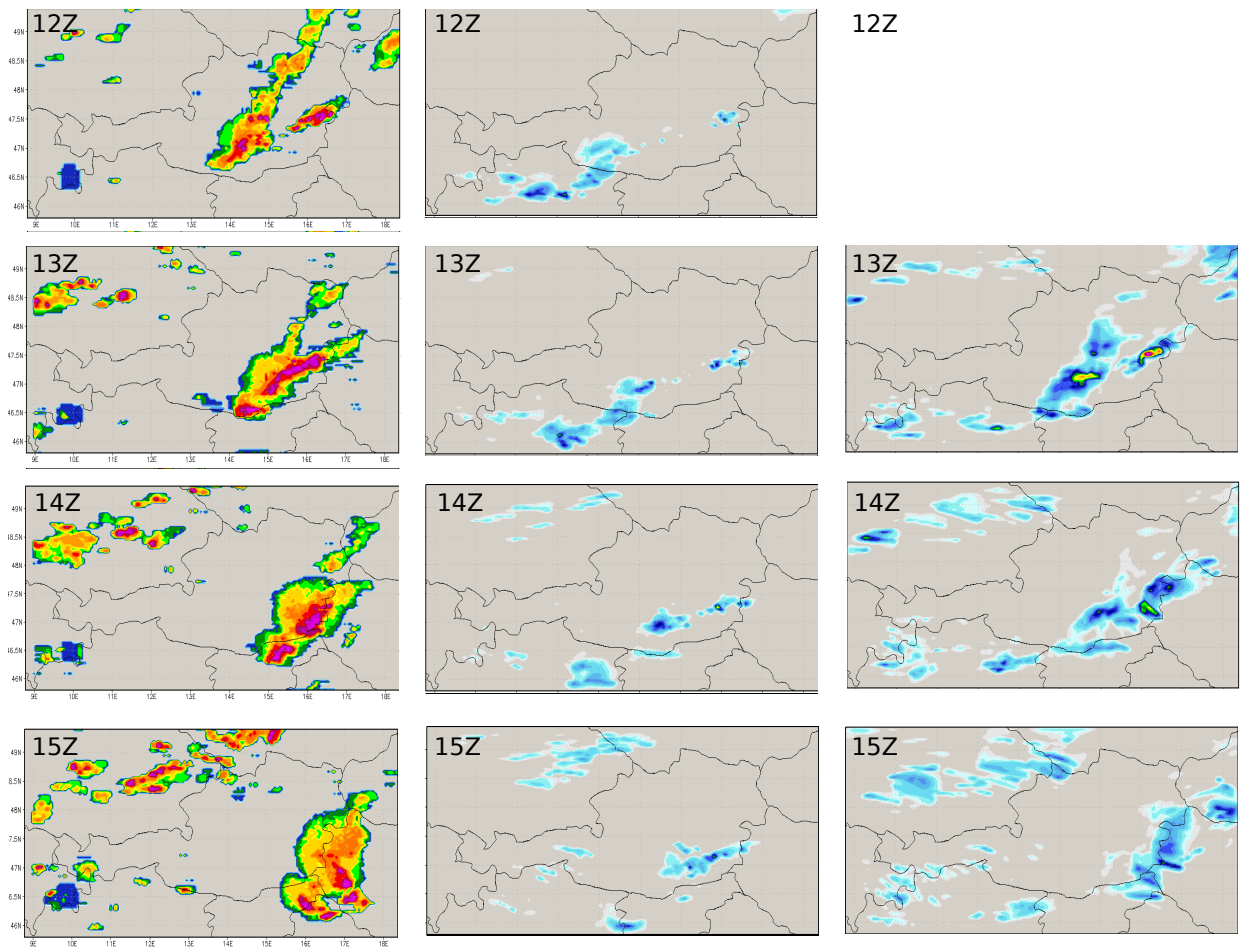


Figure 31: left: "true" radar reflectivity field, center: WRF forecast without assimilation
 1 hour accumulated precipitation, right: 12Z INIT WRFVAR with WSM6 microphysics
 1 hour accumulated precipitation

7 Verification August 2011

The verification includes 20 runs with occurring precipitation between 3rd of August 2011 and 19th of September 2011. For operational verification MET (Model Evaluation Tools www.dtcenter.org/met/users/), a free software package developed by the National Center for Atmospheric Research (NCAR) is used. At this particular time only point-to-point verification is performed, which means 1 hour accumulated model precipitation of the fine domain is verified against 1 hour precipitation rates of 200 Austrian TAWES stations. Higher sophisticated spatial verification of model reflectivity against measured reflectivity will be an important issue of future research.

7.1 Skill Scores

For point to point precipitation verification following discrete predictands(skill scores) are chosen for use:

- Critical Success Index(**CSI**)(also called Threat Score)
- Probabilty of Detection YES(**PODY**)
- Probability of Detection NO(**PODN**)

These skill scores are calculated based on a contingency table for binary events(detected or not detected) in the shape of

Forecasted	YES	NO	TOTAL
Obsrvated			
YES	hits(a)	misses(b)	a+b
NO	false alarms(c)	correct negative(d)	c+d
TOTAL	a+c	b+d	a+b+c+d

Table 1: Contingency table

The Critical Success Index is defined as

$$CSI = \frac{a}{a+b+c} \quad (45)$$

$$CSI = \frac{hits}{hits + false\ alarms + misses} \quad (46)$$

Its range is 0 to 1, with a value of 1 indicating a perfect forecast. Unlike the False Alarm Rate (FAR) and the Probability of Detection (POD) the Critical Success Index takes into account missed events (Observed YES, Forecasted NO) and false alarms (Observed NO, Forecasted YES).

To verify the accuracy of the forecast the Probability of detection YES (PODY)

$$PODY = \frac{a}{a+b} \quad (47)$$

$$PODY = \frac{hits}{hits + misses} \quad (48)$$

and Probability of detection NO(PODN)

$$PODN = \frac{d}{c+d} \quad (49)$$

$$PODN = \frac{correct\ negative}{correct\ negative + false\ alarms} \quad (50)$$

are calculated. PODY gives the percentage, how often an event is forecasted and detected. Values are ranging from 0 to 1, 1 is indicating the perfect forecast.

7.2 Verification results

The verification is performed for three precipitation thresholds: 0.1 mm, 1 mm, 5 mm precipitation per hour. Rain rates ≥ 5 mm per hour can be seen as heavy precipitation as produced in thunderstorms. For illustration only a few contingency tables for the 0.1 mm threshold are shown in the following. The Critical Success Index (CSI) indicates improvement for the forecast hours 0 - 9 for all thresholds. Probability of Detection YES (Fig. 33) is increasing in the mean at about 10 percent, with the best skill at forecast hour 7 and reaches values up to 80 percent, whereas Probability of Detection NO decreases only slightly, remaining on a high level. In Fig. 32 the positive impact of reflectivity assimilation on the WRFVAR CSI scores decreases from forecast hour 1 to forecast hour 3 after INIT (the displayed WRFVAR runs are initialized at forecast hour 3 and 6). This can be seen on the slightly downward directed bars. Like Sugimoto et al. 2009 pointed out, the reason therefore is the inconsistency between moisture and the unretrieved 3D wind field. The answer to this problem might be additional assimilation of Doppler velocities.

Forecasted	YES	NO	TOTAL
Obsvated			
YES	1036	390	1426
NO	303	1783	2086
TOTAL	1339	2173	3512

Table 2: Contingency table: hourly precipitation of WRFonly at forecast hour 7 for 0.1mm threshold

Forecasted	YES	NO	TOTAL
Obsvated			
YES	882	440	1322
NO	271	1911	2182
TOTAL	1153	2351	3504

Table 3: Contingency table: hourly precipitation of WRFonly at forecast hour 9 for 0.1mm threshold

Forecasted	YES	NO	TOTAL
Obsvated			
YES	1150	257	1407
NO	219	1884	2103
TOTAL	1369	2141	3510

Table 4: Contingency table: hourly precipitation of WRFVAR at forecast hour 7 for 0.1mm threshold

Forecasted	YES	NO	TOTAL
Obsvated			
YES	995	327	1322
NO	347	1835	2182
TOTAL	1342	2162	3504

Table 5: Contingency table: hourly precipitation of WRFVAR at forecast hour 9 for 0.1mm threshold

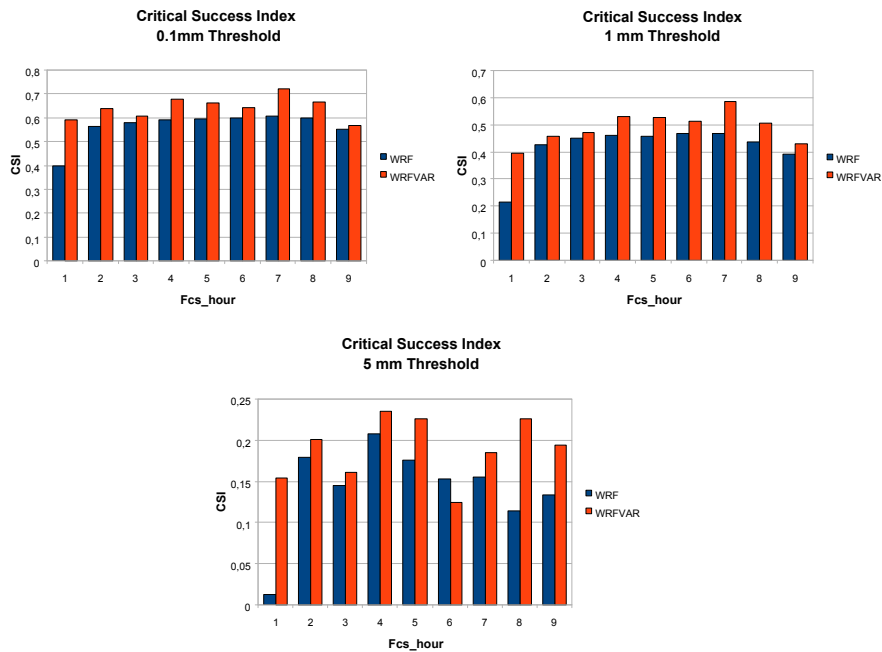


Figure 32: Critical Success Index for 0.1, 1.0 and 5.0 mm threshold WRFonly vs. WRFVAR

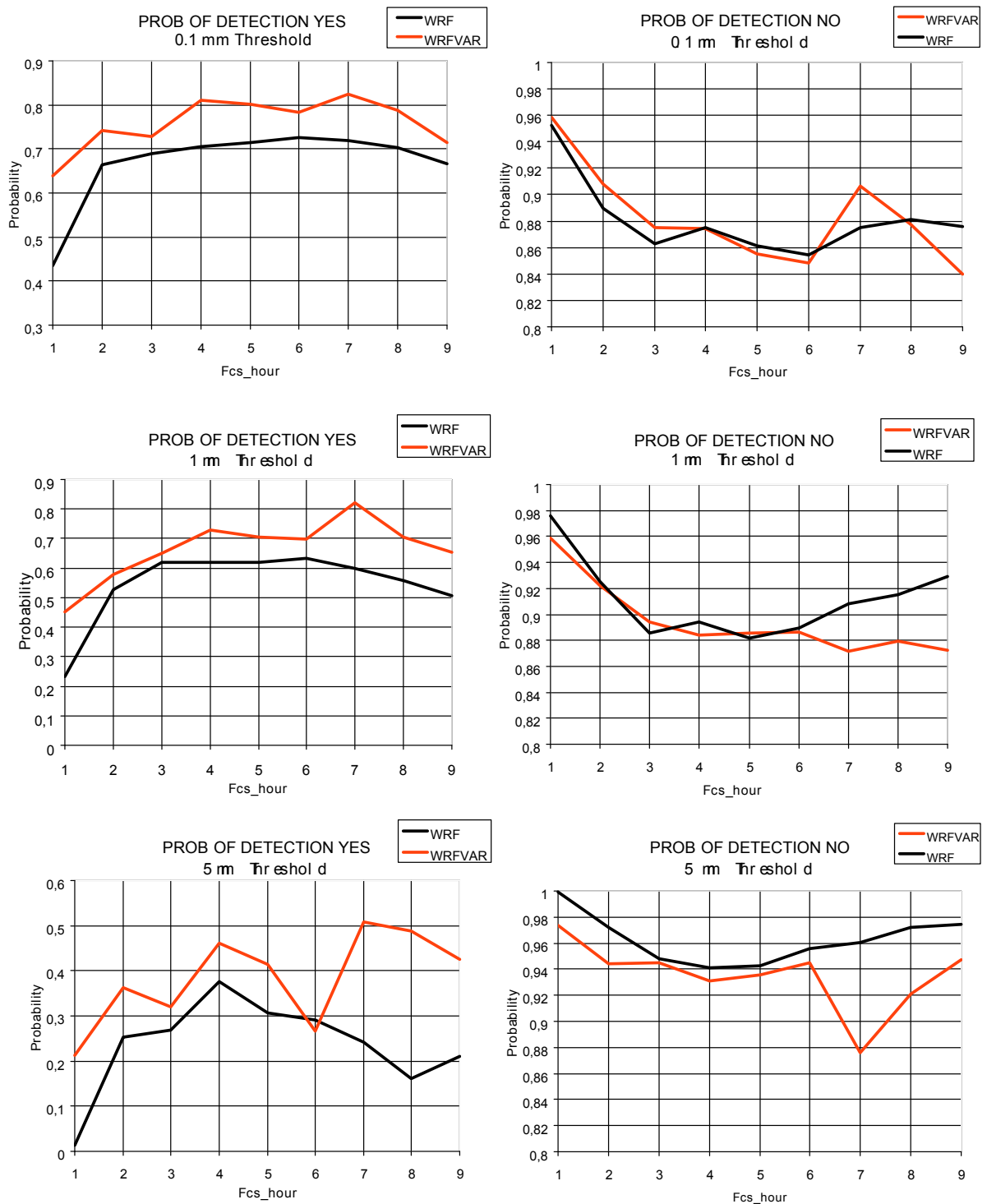


Figure 33: left: POD YES, right: POD NO for 0.1, 1.0 and 5.0 mm threshold WRFonly vs. WRFVAR

8 Conclusion and perspectives for further research

In this work the author could show the ability of improving the short term forecast (up to three hours) using radar assimilation in an hourly cycled WRF setup in combination with higher class microphysics scheme like WSM6. The tuning of the background error variance and length scaling was absolutely essential, determining if an updated analysis is leading to an usable forecast. Furthermore, the more data is available for assimilation, the better the quality of the analysis gets. That means that great efforts will be made in assimilating Doppler velocities and satellite radiances to retrieve the three dimensional wind-field and cloud motion (over radiances). Like Sugimoto et al. 2009 showed, a background preparation (*qr*-blending as described in 5.1) before assimilation is performed would be advantageous. The problem of station-rejection due to altitude differences between model terrain height and station elevation is also a future issue, increasing the number of available ground stations for assimilation.

Acknowledgements

I thank Prof. Leopold Haimberger for his continuous support working on this thesis. I thank Soichiro Sugimoto giving me helpful input for my research. I thank UBIMET for using their HPC infrastructure and financial support on this project.

References

- Anderson, J., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Q.J.R Meteorol. Soc.*, **129**, 2884–2903.
- Andersson, E., et al., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). part iii: Experimental results. *Q.J.R Meteorol. Soc.*, **550**, 1831–1860.
- Baker, D., T. Downs, M. Ku, W. Hao, G. Sistla, M. Kiss, M. Johnson, and D. Brown, 2007: Sensitivity testing of WRF physics parameterizations for meteorological modeling and protocol in support of regional SIP air quality modeling in the OTR. *Ozone Transport Commission Modelling Committee*.
- Barker, D. M., W. Huang, Y.-R. Guo, and A. Bourgeois, 2003: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. NCAR Tech. Note NCAR/TN-4531STR, NCAR Mesoscale & Microscale Meteorology Division, 68 pp., Boulder, CO.
- Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods. *Meteorological Training Course Lecture Series ECMWF*.
- Dee, D. P., 2005: Bias and data assimilation. *Q.J.R Meteorol. Soc.*, **131**, 3323–3343.
- Desroziers, G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q.J.R Meteorol. Soc.*, **127**, 1433–1452.
- Faulwetter, R., 2008: Datenassimilation und numerische Wettervorhersage. *Institute for Meteorology, University of Leipzig*.
- Fisher, M., 2003: Background error covariance modeling. *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean, ECMWF*, 45–63.
- Hong, S.-Y. and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Atmos. Sci.*, **42**, 129–151.
- Kuo, Y.-H., 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. Meteor.*, **44**, 768–788.
- Lenz, F., 2011: Variationsmethoden in der Datenassimilation zur Wetter- und Preisvorhersage. Ph.D. thesis, University of Muenster Germany.

- Parrish, D. F. and J. C. Derber, 1992: The national meteorological center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q.J.R Meteorol. Soc.*, **128**, 647–677.
- Skamarock, W., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, NCAR Mesoscale & Microscale Meteorology Division, 125 pp., Boulder, CO.
- Sugimoto, S., N. A. Crook, J. Sun, Q. Xiao, and D. M. Barker, 2009: An Examination of WRF 3DVAR Radar Data Assimilation on Its Capability in Retrieving Unobserved Variables and Forecasting Precipitation through Observing System Simulation Experiments. *Mon. Wea. Rev.*, **137**, 4011–4029.
- Sun, J. and N. 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**, 1642–1661.
- Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. *Q.J.R Meteorol. Soc.*, **113**, 1311–1328.
- Troxel, S. W. and C. D. Engholm, 1990: Vertical Reflectivity Profiles: Averaged Storm Structures and applications to Fan-Beam Radar Weather Detection in th U.S. *Amer. Meteor. Soc.*
- Wang, W., et al., 2011: WRF Modeling System User's Guide. Tech. rep., NCAR Mesoscale & Microscale Meteorology Division, Boulder, CO.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.
- Xiao, Q., Y.-H. Kuo, J. Sun, W.-C. Lee, D. M. Barker, and E. Lim, 2007: An approach of radar reflectivity data assimilation and its assessment with the inland QPF of Typhoon Rusa (2002) at landfall. *J. Appl. Meteor. Climatol.*, **46**, 14–22.

Curriculum vitae

Personal

Name	Johannes Rausch
Birth date	05.02.1985
Birth place	Vienna, Austria

Education

1999–2004	HTL Donaustadt, Technical Informatics, Vienna, Austria.
2005–2012	Diploma study in Meteorology, University of Vienna, Austria.

Employment

2003–2005	Traineeship Raiffeisen Zentral Bank, Software Department, Vienna, Austria.
since 2006	UBIMET GmbH 2006-2012, IT and NWP development, Vienna, Austria.