Data assimilation of severe convection using a Local Ensemble Kalman Filter

Masterarbeit

Heiner Lange

Betreuer: Prof. Dr. George Craig

Studiengang: Master Meteorologie

Ludwig-Maximilians-Universität München

Wintersemester 2012/2013

Erklärung:

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, den 17.01.2013 Heiner Lange

Abstract

Data Assimilation (DA) of convective storms using observations of Doppler radars could help produce better forecasts in cloud-resolving weather models. The DWD is developing Kilometer-scale ENsemble Data Assimilation (KENDA), an implementation of the Local Ensemble Transform Kalman Filter (LETKF) for the cloud-resolving COSMO model. An OSSE-framework is developed here for KENDA to test the performance of the filter in a regime with severe and organized convection, using simulated radar observations drawn from idealized stochastic thunderstorms of a nature run. Storm analysis ensembles are produced and the LETKF is found to perform well, reproducing the results of previous studies by other authors about the usage of an Ensemble Kalman Filter (EnKF) in convective DA.

It is investigated how the limited predictability of the chaotic dynamical weather system affects the forecasts and what the implications are on the choice of an appropriate analysis. A fine analysis scheme is developed using detailed observations with small-scale covariance localization, opposing a coarse scheme using averaged superobservations with a larger localization scale. Their analysis-precision is evaluated using RMSE-statistics and it is investigated how strongly the storms in the background ensemble members of the fine scheme collapse onto the observed storms, controlled by an inflated observation error covariance matrix. The coarse scheme allows the background ensemble to not-collapse, keeping a greater variance in storm-position and intensity in the ensemble.

Short-range ensemble forecasts with a lead time of 3 hours are initiated from the differently detailed analyses and compared using object based skill-scores like the Displacement and Amplitude Score DAS or the SAL-score for errors of forecast Structure, Amplitude and Location. The forecast quality of the fine scheme is found to be better in the first hour of the forecast but subsequently converges to the error level of a forecast that used a coarse analysis as the initial state. Any advantage of the initially fine analysis state is lost by the end of the forecast period.

It is concluded that, due to the limited predictability of convective storm-systems, a very fine EnKF-analysis of thunderstorms, with an ensemble collapsed onto observed clouds, is only valuable for very short-range forecasts; a coarse and non-collapsed storm-analysis is considered advantageous for most purposes, in terms of dynamical consistence of the model states and the much lower computational effort it is gained by. It is discussed that these implications could be valid even for more severe convection than simulated here. Uncertainties of vertical sounding, mesoscale situation and model error in local forecast models for convective events are expected to diminish the possible advances of a fine analysis scheme even further.

Contents

1 Introduction

Convective storms and mesoscale convective systems like squall lines or multicell storms can cause dangerous winds and flash floodings. An early warning, based on a radar-based nowcast or a good short-term forecast, can help to warn of possible hazards and damages.

Quantitative Precipitation Forecasts (QPF) are an important product to predict the local amount of rainfall or to issue warnings as strong precipitation and hazardous winds are usually collocated in the case of thunderstorms.

For a time window of 6 hours into the future, nowcasts of precipitation predict a more skillful rain field than QPF produced by mesoscale or convective forecast models (Kober et al., 2012, Lin et al., 2005). This is due to the insufficient representation of the initial field of precipitation in the model and the imperfect prediction of related meteorological phenomena like convective storms by cloud-scale models. Data Assimilation (DA) (Kalnay, 2003) of convection, using small-scale observations like Doppler radar winds and reflectivity, is regarded with potential to fill the "gap" between short-term nowcasts with lead times of $\mathcal{O}(1\,\mathrm{h})$ and mid-term forecast by mesoscale models of $\mathcal{O}(12\,\mathrm{h})$. A promising technique seems to be the Ensemble Kalman Filter (EnKF) (Evensen, 1994) which does not only propagate the most likely state of the atmosphere explicitly, but also the background error covariances.

Previous storm-scale assimilation of radar data by EnKF (Section 1.2.4) typically focused on observations of single storms and how to represent these storms as exactly as possible in the analysis state of the forecast model. That approach is questioned by this study: Even with a fairly detailed representation of the storm in the model, the limited predictability of thunderstorms due to rapid growth of small errors in the initial state (Section 1.1.3) taint the forecasts of convective scale phenomena within hours or even tens of minutes. A coarser representation of the observed convective system could give a comparably good forecast for a typical forecast lead time of three hours.

In this study, DA of convective systems is performed in an idealized convective regime, using synthetic observations from a reference nature run. A fine analysis scheme uses all observations at full resolution and localizes the influence of the observations on small distances in order to reproduce the results of previous studies about assimilation of thunderstorms; hereby the potential of the highly parallelized EnKF-algorithm "Local Ensemble Transform Kalman Filter (LETKF)" (Hunt et al., 2007) for this particular purpose is tested for the first time. Opposed to the fine scheme, a coarse scheme uses averaged superobservations and a larger radius of influence for the observations. Short-range 3-hour forecasts from these analyses are conducted and compared to the reference in terms of their bias, location error and structure-agreement to see how perturbations of

different spatial scale grow in a regime of strongly limited predictability.

The thesis is structured as follows: Section 1.1 gives a short overview about the explicit forecast of convection via numeric models and about their limitations. Section 1.2 reviews the properties, development and current state of convective DA by radar data. The algorithm of the LETKF is reviewed and discussed in terms of the intended study whose aims are specified in Section 1.3. In Chapter 2 the methodology is presented: Section 2.1 describes the fine and coarse assimilation experiments of this study and their technical setup while Section 2.2 shows the error measures and scores by which they are evaluated. Results of assimilation and ensemble forecasts are presented in Chapter 3. Chapter 4 summarizes the work, discusses the results and gives implications on objectives for further investigation.

1.1 Prediction and predictability of convective storms

In order to explicitly predict the initialization and evolution of deep convection, most weather services are running (local) forecast models with a horizontal gridspacing of $\mathcal{O}(1\,\mathrm{km})$ like COSMO-model (Baldauf et al., 2011b) or the WRF-model (Done et al., 2004), usually referring to them as "local models" or "cloud(-scale) models". These models represent the atmospheric flow by solving the three-dimensional Navier-Stokes equation to predict the future state of the contained variables (wind speed, pressure, temperature) together with moist atmospheric processes (evaporation, condensation, precipitation), discretized on a mesh grid. They are not constrained by hydrostatic balance as global models usually are, meaning they explicitly resolve nonhydrostatic vertical motion and can simulate the evolutionary stages of a thunderstorm (Figure 1, for a detailed description see Houze (1993)):

- initial ascent of air, e.g. thermal, orographic, frontally forced or triggered by the gust front of a cold pool
- condensation of water vapor and release of latent heat
- development of updraft core, anvil, overshooting top and attributed gravity waves
- spatial distribution of cloud water, snow, rain, graupel
- downdrafts induced by precipitation
- development of surface cold pools and gust fronts
- upscale growth via clustering of convection into mesoscale convective systems

A high value of Convectively Available Potential Energy (CAPE) in the environmental sounding of the atmosphere's vertical profile is an important condition to allow deep convection with strong updrafts. A vertical shear of the horizontal wind favors mesoscale organization (Houze, 1993). When CAPE-values over roughly $1500\,\mathrm{J/kg}$ are present, the occurence of intense thunderstorms is possible. A large value of CAPE is a sign

for "conditional instability": When a warm air parcel rises and latent heat is released by condensation of the contained water vapor, this additional heat source enhances the buoyancy of the air parcel. A large buoyancy difference causes a greater vertical acceleration of the air parcel – an updraft core evolves. The larger the vertical cooling gradient of the environmental temperature, the greater the instability and the more CAPE is available for the vertical acceleration of the parcel. Hydrometeors like rain and graupel grow in the updraft cores; the faster the updraft wind speeds, the larger hydrometeors can grow by condensation or grow interactively when they collide.

A highly unstable atmosphere is a precondition for this study and all studies it is based on. Whenever terms like "convection", "convective systems", "convective storms", "storms" or "thunderstorms" are mentioned in this thesis, they all refer to the atmospheric phenomena just described.

1.1.1 Prediction

The Deutscher Wetterdienst (DWD) runs the nonhydrostatic Consortium for SMall-scale Modelling (COSMO)-model COSMO-DE (Baldauf et al., 2011a) with 50 vertical levels and a horizontal resolution of 2.8 km, which is also used in this study with a higher resolution of 2 km. The term "resolution" refers to the "horizontal resolution" of a model from here on, in contrast to the "vertical resolution".

A resolution of 4 km or less appears sufficient to initiate mesoscale convective motion (Bryan et al., 2003, Weisman et al., 1997). Recent studies (Bryan and Morisson, 2012) show that a further increase of resolution down to 250 m enhances the spectrum of resolved phenomena with small pointed updrafts and vertical turbulence, possibly entering the inertial subrange. Nonetheless, the overall features of a mesoscale convective system like upper level outflow, mid level inflow, low level rear-to-front flow and downmixing of midlevel air into the cold pool[1] were also represented when using a resolution of 4 km.

Skamarock (2004) notes that the "effective resolution" of such cloud models is substantially lower than the nominal resolution of the grid-spacing Δx. Due to the downscale energy cascade there needs to be dissipation at the smallest resolved scales, either implicitly within the numerical scheme (mostly finite differences) or by an explicit closure. Comparing the missing energy at the smallest scales of cloud models to observed spectra, Skamarock estimates the effective resolution to be in the order of roughly $7\Delta x$, specifically to the Weather Research and Forecasting (WRF)-model (Done et al., 2004), but in comparable magnitude with other models. He concludes that an explicit forecast of a thunderstorm with cloud resolving models might be limited to a range of tens of minutes, even if the initial state on the grid-resolved scale is correct. Nonetheless, the up-scale influence of convection on the larger field could still be correctly described by the model.

1 For a detailed description, again see Houze (1993)

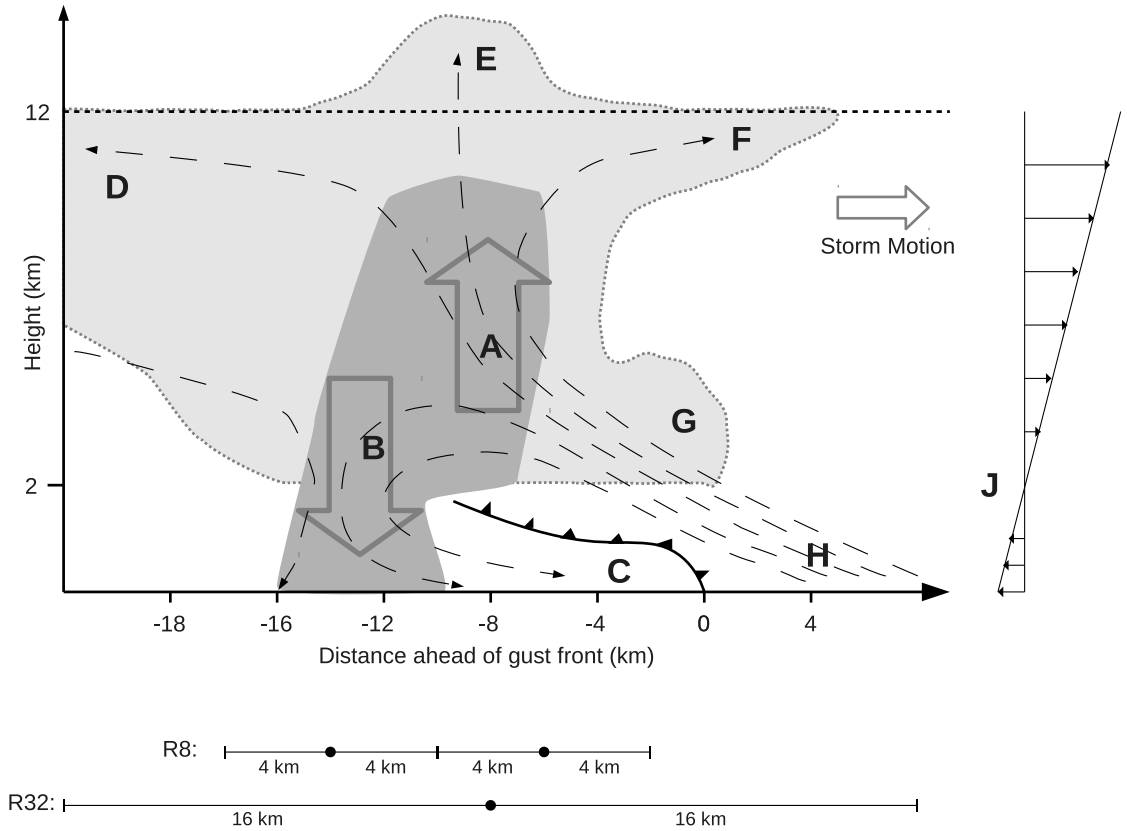

Figure 1: Schematic slice along the propagation direction of a convective storm system. Light gray shading shows the cloud outlines containing low and medium amounts of hydrometeors (mostly cloud water, ice, snow), strong gray shading shows the cell core with high amounts of precipitation (rain, graupel, hail). Some main storm features are: A) Updraft core of active cell, B) downdraft core with precipitation that reaches the surface, C) surface cold pool with gust front, D) backsheared anvil, E) overshooting top, F) forward-sheared anvil, G) newly developing subcell of the system, H) moist air ahead of storm together with possible trajectories, J) vertical shear of the environmental horizontal wind (relative to storm motion).
Below the horizontal axis the line R8 depicts two possible observations (circles) with a horizontal radius of $8/2 = 4\,\text{km}$ around each. The line R32 shows one possible observation with a radius of $32/2 = 16\,\text{km}$.
This sketch is inspired by the sketch of the Raymer Hail storm in Houze (1993).

1.1.2 Parameterizations and model error

The goal of a weather model is to predict the future state of the atmosphere as well as possible. Thus, all processes involved in real weather need to be represented in the model. The Navier-Stokes equation of the flow, filtered or in primitive form, is finitely discretized on the model grid (Kalnay, 2003). So, in a cloud model with 2 km resolution as used here, there are unresolved processes like real three-dimensional turbulence at scales of $\mathcal{O}(100\,\mathrm{m})$ and smaller, or diffusion at very small scales of $\mathcal{O}(1\,\mathrm{cm})$. These processes need to be parameterized, together with exchange processes like fluxes of sensible or latent heat from the surface or energy transfer by solar and thermal radiation. Also, the evolution and distribution of single hydrometeor objects (particles of cloud water, cloud ice, rain, snow, graupel, hail) is usually too expensive to be computed explicitly, so it is parameterized in bulk microphysics schemes. These parameterizations are simplified representations and only work as a limited substitute for an explicit simulation of the respective processes. For more information on parameterizations, see Kalnay (2003).

These technical limitations of the forecast model (together with the chaotic limitations of the atmospheric system described in Section 1.1.3) will render any forecast subsequently defective, even if the initial states of all the model's discretized variables are correct. This technical defectiveness is called "model error".

As an example, model error caused by unresolved subgrid processes shows up as a positive bias in precipitation amounts if mesoscale convective systems are simulated with a resolution of 2 km instead of 100 m (Bryan and Morisson, 2012). A 100 m simulation shows lower and more realistic precipitation amounts as it resolves entrainment of dry air into updrafts by vertical turbulence whose eddies extend on a much smaller scale than 2 km. Defective parameterizations like bulk microphysics will obviously give an imperfect forecast due to their model error, being unable to resolve all processes of real hydrometeors.

This study excludes model error by making the assumption of a "perfect model": An idealized "nature run" is used as the source of synthetic observations for the DA scheme that uses the same model as the nature run for its analyses and forecasts. These are then evaluated against the nature run, so a perfect forecast would be possible if the initial state was correct. This approach is known as a perfect-model Observations System Simulation Experiment (OSSE).

In such a perfect model experiment, the defectiveness of forecasts from an imperfect initial state is only subject to the chaotic nature of the atmosphere that limits the predictability.

1.1.3 Predictability of convection

Weather forecasting is an inital value problem: The future development of the flow is determined by integrating the model's equations forward in time, starting at the initial state. If this integration is strictly deterministic, a perfect inital state gives a perfect forecast. If the initial state is perturbed by an error, this perturbation grows nonlinearly

in amplitude and size. The maximum size and amplitude are determined by the system's dynamics. The growth speed of the perturbation depends on whether unstable modes are excited, and is attributed to the Lyapunov vectors in perturbation space (Kalnay, 2003). The fact that even tiny perturbations will inevitably destroy the forecast's resemblance to the truth is known as "chaotic behavior".

The chaotic nature of the atmosphere imposes an intrinsic limit on predictability for forecast models (Lorenz, 1963). Lorenz (1969) proposed the concept of an upscale growth of perturbations via turbulent dynamics: The error of one scale contaminates the next larger scale within one "turnover time" of an eddy of the respective scale. Regarding the spectrum of kinetic energy of atmospheric scales, he estimated an error doubling time of roughly 1 hour in the convective scale of $\mathcal{O}(10\,\mathrm{km})$ and 3 hours in the mescoscale of $\mathcal{O}(100\,\mathrm{km})$.

Studies on convective predictability

Various studies have been conducted following Lorenz' concept in order to find a limiting time of the atmospheric predictability. A selection of those that focused on convective systems is presented here, reviewing their methods and findings. They all conduct perfect model experiment, excluding effects of model error.

Investigating the possibility of operational forecasts of thunderstorms, Lilly (1990) presumed a higher predictability of long-lived organized convective systems in atmospheric conditions with helical hodographs, i.e. a turning of the horizontal wind direction with height, contrary to short-lived storms that extinguish themselves quickly by cutting themselves off suppliant air due to their precipitation. He noted that the intermittent nature of Mescoscale Convective Systems (MCS) could speak against regarding them as part of a continuous spectrum of turbulent scales, so they might be subject to a smaller error growth rate.

More recent studies (Hohenegger et al., 2006, Walser et al., 2004, Zhang et al., 2003) compared randomly perturbed model runs to unperturbed runs, whereas Craig et al. (2012) and Aksoy et al. (2009) used "perturbations" caused by increments of DA schemes.

Zhang et al. (2003) recognized moist convection as the cause of rapid growth of error amplitude and scale in a mesoscale convective area influencing a winterly cyclogenesis over North America. In their study, they used a convection-permitting model with 3.3 km resolution. They perturbed the temperature field within the nested domain of their forecast model and drew a comparison to a forecast with unperturbed initial conditions. The Boundary Conditions (BC), provided by a surrounding mesoscale model, were left identical in order to study domain-internal growth of convective/mesoscale error only, excluding possible forecast errors of the synoptic scale. They observed small perturbation amplitudes to grow faster in the first three hours and then to reach an approximate saturation of growth rate (Figure 2), pointing at the strong nonlinearity of the involved processes at the smallest scales resolved by their model (i.e. moist convection).

Walser et al. (2004) compared the predictability of convection in different synoptic

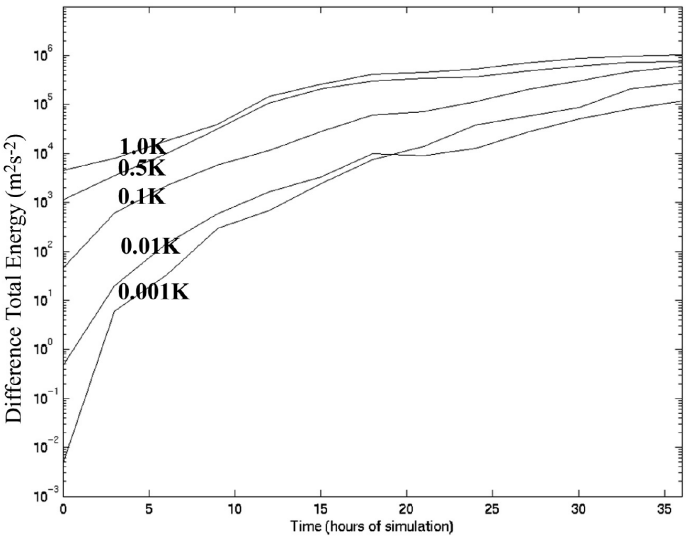

FIG. 4. Evolution of DTE (in $m^2\ s^{-2}$) in experiments with idealized perturbations of different magnitudes in the initial temperature field. Curves are labeled with the values of T_0 used in (1).

Figure 2: Figure 4 of Zhang et al. (2003). The Difference Total Energy (DTE) measures the difference of the perturbed runs compared to the unperturbed run. T_0 is the amplitude of the initial temperature perturbations.

regimes over the Alps by also using a nested domain with identical BC for their ensemble[2], perturbed by lagged initiation time of the forecast. They observed higher predictability of the distribution of convectively induced precipitation when the convection was forced by a cold front moving through the domain. In contrast, the predictability in a weakly forced regime appeared lower, owing to the more stochastic behaviour of convective storm initiation in the absence of a synoptic forcing pattern[3]. They also noted orography to have an increasing impact on predictability, meaning that thunderstorms are likely to initiate over the same orographic features (e.g. mountains) even in slightly perturbed atmospheric conditions.

Hohenegger et al. (2006) conducted a similar study where they observed the mean flow to have the largest influence on the predictability of convection. They hypothesized that the propagation of perturbations through the domain takes place via gravity waves. They reason that the gravity waves which are caused by one thunderstorm in one location influence the initiation of another storm in a different and potentially distant location of the domain. They name the windspeed of the background flow as the constraint on this effect: If the mean flow is strong and exceeds the group speed of gravity waves (as

2 A definition of ensemble forecasting is provided in Section 1.1.4
3 A cold front, in that study

for example in a cold front case), the upstream propagation of disturbances by gravity waves is impossible. On the other hand, in a slow mean flow of a weakly forced synoptic regime the disturbances may travel upstream. Hohenegger and Schär (2007) drew similar conclusions, using also domain-internal perturbation fields like temperature noise.

Craig et al. (2012) conducted sensitivity-experiments with a Latent Heat Nudging scheme for the assimilation of radar reflectivity observations into a local model. The positive impact of the observations decayed with different forecast times, depending on the convective equilibrium characteristics of the cases, measured by the convective time-scale (Done et al., 2006). In weakly forced situations where the location of the cells was not determined by synoptic features, the impact of the observations was strong and lasted relatively long (roughly 3 hours), hinting to the triggering of long lasting convective cells which did not initiate in the control forecast. In a frontally forced situation, a shorter decay time was observed. They suggest that in the latter case the locations of single convective cells were already predetermined by the location of the cold front that was provided by the boundary and initial conditions.

Aksoy et al. (2009, 2010) assimilated radar observations of different types of organized convection. They observed the predictability of single convective systems to depend on their degree of internal organization. A squall-line with its mesoscale organization structure and triggering of new cells along the well defined gust front seemed to have a slower forecast error growth than a multicell thunderstorm in which the new subcells emerged in a more random manner –which would agree with the aforementioned presumption of Lilly (1990). They estimated the error-doubling time to be in the order of 10 minutes, using a 2 km resolving instance of the WRF-model.

Implications on this study

This study focuses on the intrinsic predictability limits of forecasts of convection. Anticipating Section 2.1 shortly: Enhanced predictability by orographic forcing will be disabled by using a flat domain. Enhanced predictability by synoptic or frontal forcing will be disabled by using a horizontally homogeneous initial sounding. The initial perturbations for the conducted ensemble forecasts are contained in the imperfect analysis states of the DA system.

Using perturbations of an ensemble in an EnKF-framework to assess atmospheric predictability was shown to be a reasonable approach by Kuhl et al. (2007), using a global model. They showed how the local low dimensionality of the ensemble vanished rapidly in the beginning of the forecast period. Using a perfect-model approach, they warned that in the real atmospheric system the local low dimensionality might be less present, therefore the predictability will generally be minor in the presence of model error. This warning should be kept in mind for this study where predictability of thunderstorms is assessed using a perfect-model approach in a cloud model.

1.1.4 Ensemble Forecasts

To account for the limited atmospheric predictability, Epstein (1969) and Leith (1974) suggested to represent the Probability Density Function (PDF) of the atmospheric state $P(\mathbf{x})$ by a limited sample, consisting of an ensemble of k members at initial time $j-1$, with mean $\bar{\mathbf{x}}$ and standard deviation σ:

$$P(\mathbf{x}_{j-1}) = \sum_{i=1}^{k} \delta(\mathbf{x}_{j-1} - \mathbf{x}_{j-1}^{i}) \tag{1.1}$$

[4] where δ is the Dirac-function with

$$\delta(\mathbf{x}_j - \mathbf{x}_j^i) = \begin{cases} 1, & \text{if } \mathbf{x}_j = \mathbf{x}_j^i \\ 0, & \text{else} \end{cases} \tag{1.2}$$

The members are, like single deterministic forecast, integrated from the initial time $j-1$ to the time j by the model $M_{j-1,j}$:

$$\mathbf{x}_j = M_{j-1,j}(\mathbf{x}_{j-1}) \tag{1.3}$$

Thus, also the PDF is integrated forward in time:

$$P(\mathbf{x}_j) = \sum_{i=1}^{k} \delta[\mathbf{x}_j - M_{j-1,j}(\mathbf{x}_{j-1}^i)] \tag{1.4}$$

$P(\mathbf{x}_j)$ then represents the PDF of the future state in which the initial perturbations contained in $P(\mathbf{x}_{j-1})$ have either decayed or evolved into larger disturbances. Due to the nonlinearity of the forecast model M, $P(\mathbf{x}_j)$ will not have the same distribution as $P(\mathbf{x}_{j-1})$ (Bocquet et al., 2010). Even if the states of the ensemble have a simple Gaussian distribution at time j, they diverge in the case of a (nonlinear) convective forecast. The resulting distribution may be multimodal (Figure 3) so that the mean may not even represent a distinct mode anymore. The sequential Kalman Filter (Kalman, 1960) which is involved in this study assumes a normal distribution $N(\bar{\mathbf{x}}^b, \mathbf{P}^b)$ (here often called "Gaussian") of the background ensemble and the contained variables at all times, together with a linear forecast model. As the background states are not normally distributed and the weather forecast model is linear only very approximately, the filter will either give a bad solution or at least collapse to single modes in certain locations, i.e. with updraft cells, when the ensemble states are pushed into the direction of the observation.

4 (1.1) is similar to eq. 12 in Bocquet et al. (2010)

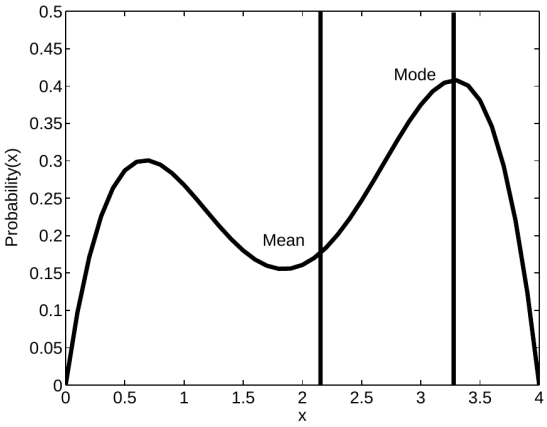

Fig. 1. Schematic of a bimodal probability density function (pdf), with the mean and mode indicated using vertical lines.

Figure 3: Figure 1 of Dance (2004).

1.2 Convective Data Assimilation

The fundamental idea of DA is to determine the most likely state of the atmosphere and its uncertainty by combining state and uncertainty of a model forecast and state and uncertainty of real world observations. State (first moment) and uncertainty (second moment) in DA are regarded as the mean and the variance of the random variables "atmospheric state" and "observation". To determine the mean and the variance, a cost function is minimized that has the same structure as the argument of a multidimensional Gaussian distribution. This "Gaussian assumption" is a useful approximation that keeps the equations simple and computational costs comparably low, but it may fail when the actual distributions are non-Gaussian (Dance, 2004).

DA at synoptic scales, e.g. via the 3DVAR-scheme or Optimal Interpolation (Kalnay, 2003), traditionally uses a prefabricated background error covariance matrix to minimize the respective cost function for the most likely state called "analysis state" or simply "analysis". The entries of this matrix represent correlations that are constrained by the atmospheric system such as the Rossby radius of deformation in terms of distance correlations or the geostrophic and thermal wind balance in terms of cross-correlations between variables (i.e. wind, temperature and pressure). Through these correlations, the information of point-observations like measurements from radiosondes is widely distributed in the model space in a dynamically consistent manner.

The 4DVAR-scheme initially uses a similar matrix as 3DVAR, but additionally takes the flow dependence of the covariances into account by minimizing not only for the most likely model state but also for the most likely trajectory which led to that state (Dance, 2004).

At convective scales, the mentioned dynamic balances do not hold anymore – the flow

will even be nonhydrostatic if convection is present. Also, the background error covariances are strongly flow dependent and should ideally be constructed for every single analysis. This approach is executed in Ensemble Kalman Filtering (Evensen, 1994). An ensemble of the forecast model is used to finitely sample the PDF of the atmospheric state and to propagate it in time, providing background error covariances for analyses. These are produced sequentially: An analysis is followed by an ensemble forecast, followed again by an analysis that now benefits from the previous analysis; in other words, a Bayesian update of the most likely state and its probability is performed, conditioned on the probability of the forecast ensemble and the observations (Kalnay, 2003).

The analysis of intermittent phenomena like convective systems can strongly benefit from an EnKF-scheme. For example, assume the state space (provided by an ensemble forecast) in a certain region spans the possibility of either a weak single storm or a strong organized convective system. Then the DA algorithm can relate this to observations of either low reflectivity or high radar reflectivity and construct an analysis state that hopefully resembles reality by containing the observed feature. On the other hand, the DA-scheme will fail if all ensemble members do not predict a storm but one is observed. Approaches to solve this particular problem of storm initialization by previous convective EnKF studies are reviewed in Section 1.2.4.

The following subsections give an overview about the Ensemble Kalman Filtering of convection with respect to the used observations, the filter algorithm, previous studies and the problems that arise from this topic.

1.2.1 Observations

At convective scales, the most widely available observation types are reflectivity (in dBZ) and radial wind (in m/s) from a scanning Doppler radar. Figure 4 shows an example of such radar data. In the US, the National Oceanic and Atmospheric Administration (NOAA) operates the NEXRAD network of roughly 150 WSR-88D S-band Doppler weather radars (Crum and Alberty, 1993) which covers the continental US almost completely (Figure 5). Similar radar networks are also installed in European countries[5].

Scanning at 14 horizontal angles from 0.48 ° to 19.5 °, the single stations provide one volume scan per 5 minutes (Lu and Xu, 2009). The azimuthal resolution is 1° and the radial resolution ranges from 250 m nearby to 1 km at the maximum distance. Reflectivity is observed up to a radius of 345 km, radial wind up to 230 km away from the radar site. The beam of the lowest elevation angle goes up to 5 km at the maximum radius of 230 km, leaving increasing parts of the lower atmosphere unobserved. If two radar sites are close together, this effect is diminished as their scanning swaths already cross at lower elevations. If an area is observed by more than one radar, the absolute wind field can be reconstructed by triangulating the single radial wind speeds into a mutual frame of reference (Potvin

5 http://www.knmi.nl/opera/

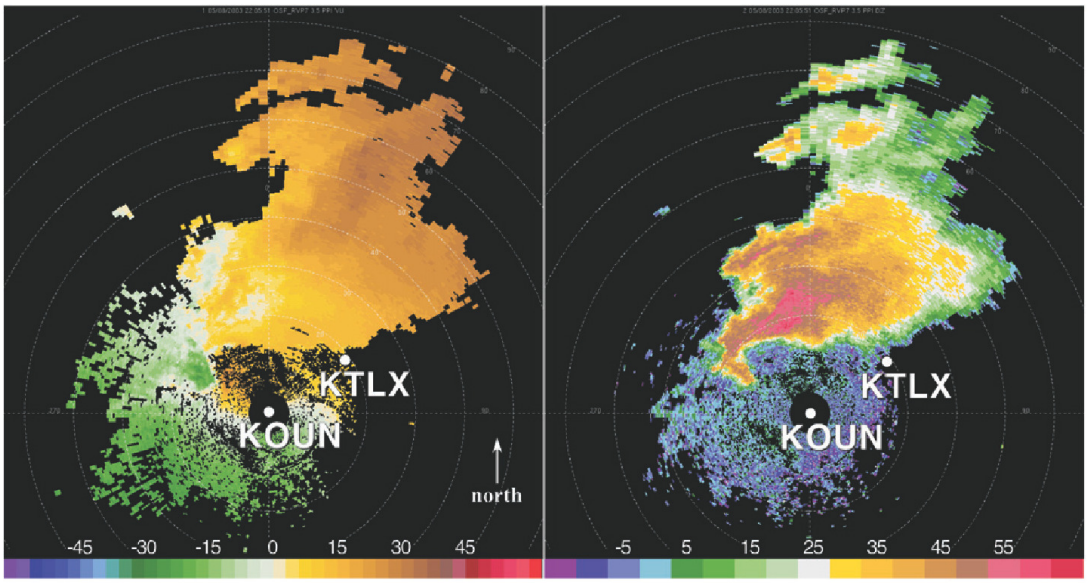

FIG. 1. (left) Doppler velocity (m s^{-1}) and (right) effective reflectivity factor (dBZ) from the KOUN radar at 3.5° elevation angle at 2205 UTC 8 May 2003. Range rings are every 10 km. The locations of the KOUN and KTLX radars are shown.

Figure 4: Fig. 1 from Dowell and Wicker (2009). Radial Wind (left, in m/s) and Reflectivity (right, in dBZ) of a supercell thunderstorm in Oklahoma

and Wicker, 2012).

To assimilate such observations, *observation operators* need to map the model fields onto the observations. For radial velocity, the three model wind components (u, v, w) are combined with the elevation angle α and the azimuth angle β to obtain the radial wind V_r:

$$V_r = u \cos\alpha \sin\beta + v \cos\alpha \cos\beta + w \sin\alpha \tag{1.5}$$

[6] For reflectivity, the mixing ratios of reflective hydrometeors (rain, snow, graupel) and their microphysical properties (particle size, density, intercept parameter, dielectric factor) need to be taken into account to get the logarithm of the reflectivity Z in dBZ. Typically observed values of Z are ranging from 10 dBZ for drizzle to 70 dBZ for large hail. Values below 5 dBZ can be treated as "no rain"-observations which can help suppressing spurious convection (Aksoy et al., 2009, Tong and Xue, 2005).

To weigh the observations against the model background in the DA scheme, a variance value for the observation error covariance matrix is needed. This is typically assumed

6 (1.5) from Tong and Xue (2005)

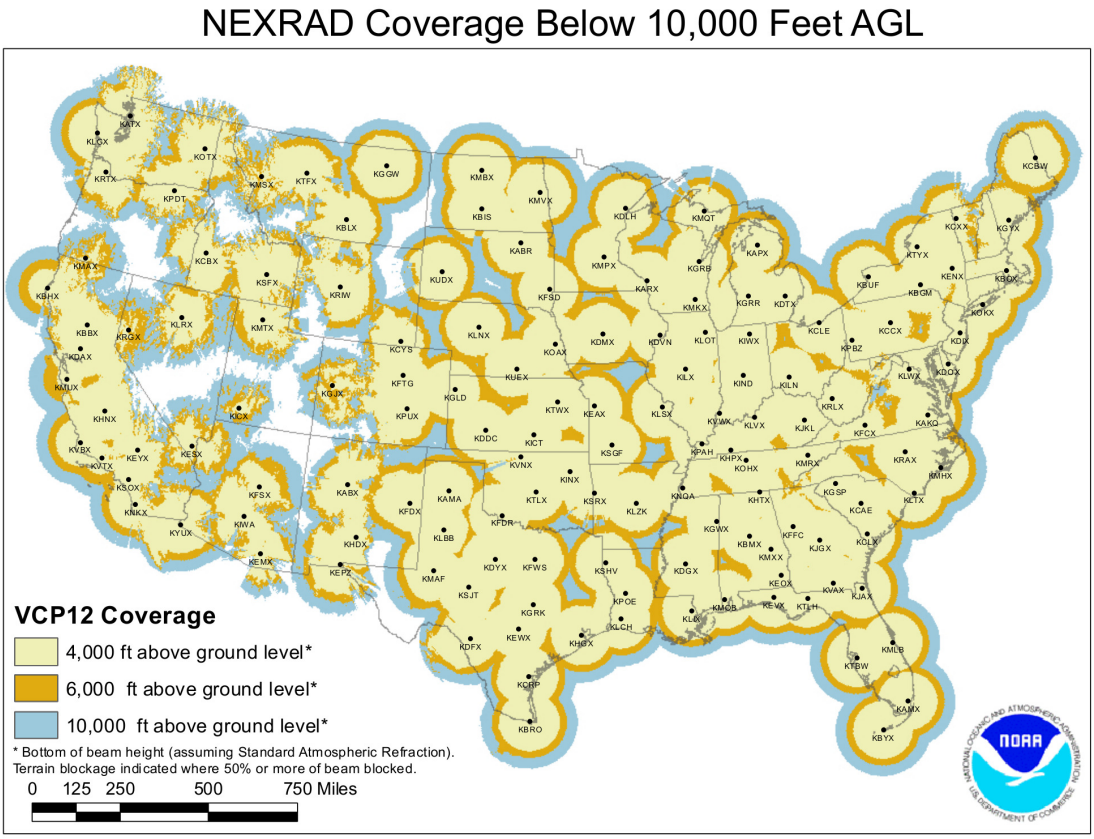

Figure 5: Spatial coverage of the NEXRAD Radar Network. Different shadings depict the height of the lowest beam per radar. Image source: http://www.roc.noaa.gov/

to be σ^2 of the standard deviations $\sigma = 1 - 2\,\mathrm{m/s}$ for wind and roughly $\sigma = 5\,\mathrm{dBZ}$ for reflectivity, already including measurement error and estimated representativity error. The representativity error is defined as the error that arises when scales smaller than the observational resolution influence the measurement. Observation averaging is an approach to diminish that effect (Salonen et al., 2009).

1.2.2 LETKF

Now the building blocks for the DA are defined and the LETKF-scheme that is used here can be described. The formulation by Hunt et al. (2007) is recapitulated – for a full description of the algorithm see the original paper.

LETKF-Algorithm

The m-dimensional "state space" is defined by the discretized model variables contained in the state vector $\mathbf{x}$. The n-dimensional "observation space" is defined by the entries of the observation vector $\mathbf{y}^o$. The number of ensemble members is given by k. Usually $m \gg n \gg k$ holds.

$\mathbf{x}$ *:* m-dimensional state vector (of forecast model)

$\bar{\mathbf{x}}^b$ *:* Background state estimate (forecast ensemble mean)

$\mathbf{P}^b$ *:* Background error covariance matrix

$\bar{\mathbf{x}}^a$ *:* Analysis state estimate (analysis ensemble mean)

$\mathbf{P}^a$ *:* Analysis error covariance matrix

$\mathbf{y}^o$ *:* n-dimensional observation vector in observation space

$\mathbf{R}$ *:* Observation error covariance matrix (diagonal here)

H *:* Observation operator, mapping from state space to observation space

$H(\mathbf{x})$ *:* first guess (model state $\mathbf{x}$ mapped onto observations $\mathbf{y}^o$)

The background state estimate is given by the ensemble mean

$$\bar{\mathbf{x}}^b = k^{-1} \sum_{i=1}^{k} \mathbf{x}^{b(i)} \tag{1.6}$$

where k is the number of ensemble members and $\mathbf{x}^{b(i)}$ is the i-th member.

The background error covariance matrix $\mathbf{P}^b$ is given by multiplying the deviations of the single members from their mean by their transpose:

$$\mathbf{P}^b = (k-1)^{-1} \sum_{i=1}^{k} [\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b][\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b]^T = (k-1)^{-1} \mathbf{X}^b (\mathbf{X}^b)^T \tag{1.7}$$

where $\mathbf{X}^b$ is a $m \times k$ matrix with i columns of $\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b$. $\mathbf{P}^b$ and $\mathbf{X}^b$ are thus at most of rank $k-1$.

The observation error covariance matrix $\mathbf{R}$ is filled by the variances and covariances of the observational error. The cost function to be minimized for the analysis state $\mathbf{x} = \bar{\mathbf{x}}^a$ is

$$J(\mathbf{x}) = [\mathbf{x} - \bar{\mathbf{x}}^b]^T (\mathbf{P}^b)^{-1} [\mathbf{x} - \bar{\mathbf{x}}^b] + [\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x})] \tag{1.8}$$

$[\mathbf{x} - \bar{\mathbf{x}}^b]$ denotes the deviation of $\bar{\mathbf{x}}^b$ from the minimizing state $\mathbf{x}$ and is weighted by $\mathbf{P}^b$. $[\mathbf{y}^o - H(\mathbf{x})]$ is the deviation of $\mathbf{x}$ mapped onto the observations from the observations $\mathbf{y}^o$, weighted by $\mathbf{R}$. When J is minimized, the one $\mathbf{x}$ is determined that i) either lies closer to the background ensemble mean $\bar{\mathbf{x}}^b$ if $\mathbf{P}^b$ is small and thus dominant or ii) whose $H(\mathbf{x})$ lies closer to $\mathbf{y}^o$ if $\mathbf{R}$ is small and dominant.

Instead of minimizing J in the m-dimensional state space (i.e. for every single model variable and grid point at once with an explicitly computed $\mathbf{P}^b$), the minimization is taken out in the $(k-1)$-dimensional subspace S of the ensemble perturbations spanned by the columns of $\mathbf{X}^b$.

$\mathbf{X}^b$ is regarded as a linear transformation from a k-dimensional space $\tilde{S}$ onto S. The analysis is performed in $\tilde{S}$ and solves for the vector $\mathbf{w}$ that contains the k factors for the linear combination of the ensemble perturbations that represent the most likely state estimate: $\mathbf{X}^b\mathbf{w}$ then belongs to the space S and $\mathbf{x} = \bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}$ minimizes (1.8) in state space, by then called "analysis" state $\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b\bar{\mathbf{w}}^a$

The transformed cost function $\tilde{J}$ to be minimized for $\mathbf{w} = \bar{\mathbf{w}}^a$ in $\tilde{S}$ is

$$\tilde{J}(\mathbf{w}) = (k-1)\mathbf{w}^T\mathbf{w} + [\mathbf{y}^o - H(\bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})]^T\mathbf{R}^{-1}[\mathbf{y}^o - H(\bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})] \tag{1.9}$$

If H is too nonlinear, $\tilde{J}(\mathbf{w})$ could have multiple minima, so H is linearized about the background ensemble mean state $\bar{\mathbf{x}}^b$. The background observation vectors are defined as

$$\mathbf{y}^{b(i)} = H(\mathbf{x}^{b(i)}) \tag{1.10}$$

with the mean $\bar{\mathbf{y}}^b$ and an $n \times k$ matrix $\mathbf{Y}^b$ whose ith column is $\mathbf{y}^{b(i)} - \bar{\mathbf{y}}^b$. The linearizing of H about $\bar{\mathbf{x}}^b$ is given by

$$H(\bar{\mathbf{x}} + \mathbf{X}^b\mathbf{w}) \approx \bar{\mathbf{y}}^b + \mathbf{Y}^b\mathbf{w} \tag{1.11}$$

This yields the Kalman Filter cost function

$$J^*(\mathbf{w}) = (k-1)\mathbf{w}^T\mathbf{w} + [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}]^T\mathbf{R}^{-1}[\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}] \tag{1.12}$$

using $\bar{\mathbf{w}}^b = \mathbf{0}$ and $\tilde{\mathbf{P}}^b = (k-1)^{-1}\mathbf{I}$ (which means that if the analysis mean is equal to the forecast mean, zero perturbations are added). The matrix $\mathbf{Y}^b$ plays the role of the observation operator H as defined in (1.11). The minimization of $\mathbf{J}^*$ for $\mathbf{w} = \bar{\mathbf{w}}^a$ approximately minimizes the cost-function (1.9) in $\tilde{S}$ and yields

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}(\mathbf{y}^o - \bar{\mathbf{y}}^b) \tag{1.13}$$

with

$$\tilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\mathbf{Y}^b]^{-1} = (k-1)^{-1}\mathbf{W}^a(\mathbf{W}^a)^T \tag{1.14}$$

The analysis perturbation weight matrix $\mathbf{W}^a$ chooses the analysis ensemble through $\mathbf{X}^a = \mathbf{X}^b\mathbf{W}^a$ and is given by the analysis member perturbation weights $\mathbf{w}^{a(i)}$ via

$$\mathbf{W}^a = [(k-1)\tilde{\mathbf{P}}^a]^{1/2} \tag{1.15}$$

where $\mathbf{W}^a$ has i entries of $\mathbf{w}^{a(i)} - \bar{\mathbf{w}}^a$. The $1/2$ power of the symmetric matrix $\tilde{\mathbf{P}}^a$ specifies the symmetric square root. From (1.15) the single analysis perturbation weight vectors $\mathbf{w}^{a(i)}$ are gained to finally compute the analysis ensemble members in state space in the "filter update":

$$\mathbf{x}^{a(i)} = \bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}^{a(i)} \tag{1.16}$$

where the terms $\mathbf{X}^b\mathbf{w}^{a(i)}$ are called the "increments of the analysis" respectively "of the update".

The single members of $\mathbf{x}^{a(i)}$ are then used as inital conditions for the forecast model and integrated forward during the assimilation interval Δt_{ass} from the analysis time $j-1$ to the next assimilation time j where they represent the new background ensemble $\mathbf{x}^{b(i)}$:

$$\mathbf{x}_j^{b(i)} = M(\mathbf{x}_{j-1}^{a(i)}) \tag{1.17}$$

The cycling interval Δt_{ass} is a characteristic parameter of the EnKF and should be chosen in a length where the background PDF has time to explore new dynamical developments of the system without diverging from the mean state too strongly. Δt_{ass} is also referred to as "analysis interval" because at the end an analysis takes place, as "forecast interval" because the forecast model propagates the ensemble through that time-interval, or simply as the "assimilation interval" of the cycling.

The symmetric square root formulation enables the LETKF, like the Ensemble Square Root Filter (EnSRF), to make use of the untreated observations. In the classical EnKF, also an ensemble of observations $\mathbf{y}^{o(i)}$ needs to be generated by adding random noise so that the analysis error covariance is not underestimated (Whitaker and Hamill, 2002).

Notes on the filtering scheme

The analysis ensemble members $\mathbf{x}^{a(i)}$ consist of the background ensemble mean $\bar{\mathbf{x}}^b$ with a weighted linear combination ($\mathbf{w}^{a(i)}$) of the background ensemble members' perturbations ($\mathbf{X}^b$) added (1.16), so $\mathbf{x}^{a(i)}$ can only be constructed from states that were already forecasted by the background forecast ensemble $\mathbf{x}^{b(i)}$. By constructing a linear combination of a finite number of members, the analysis is expected to be a consistent and stable model state.

This assumption does not necessarily hold for linear combinations of all possible phenomena: Assume the updraft core of a thunderstorm is observed but only one member comes close to the observations by containing the updraft of a small shower. The filter will then exaggerate this shower by giving the respective member a higher weight; this might result in an imbalanced state that causes noise and degrades the analysis (Greybush

et al., 2011).

Localization

Since typically there are much more state variables than ensemble members, the $m \times k$ matrix $\mathbf{X}^b$ is rank deficient. Correlations that are indicated by the limited sample of the ensemble members are thus not necessarily trustworthy, especially with long distances: For example, a single thunderstorm over Hamburg does not necessarily coincide with a single shower over Munich, even if the finite sample of the ensemble were accidentally to suggest so. In order to get rid of the influence of such spurious correlations in $\mathbf{P}^b$, the spatial influence of the observations needs to be limited.

The LETKF performs the analysis for every single model gridpoint, taking into account only nearby observations. The influence of more and more distant observations on the local analysis is first diminished and then set to zero by multiplying their entries in the inverse observation error covariance matrix $\mathbf{R}^{-1}$ with a correlation function that goes to zero within the finite distance of the *localization radius* r_{Loc}. Here a polynomial function is used whose shape is very similar to a Gaussian bell but is zero for $r > r_{Loc}$ (Gaspari and Cohn, 1999). Outside r_{Loc}, the entries of the localized $\mathbf{R}^{-1}$ are nullified, so distant observations do not have an influence on the local linear combinations determined by (1.14). If r_{Loc} is short, it is possible that even within compact phenomena like convective systems different members are chosen for different parts of the storm if they fit the local observations better (cf. R8 in Figure 1).

Localization in observation space as in the LETKF enhances the effective ensemble size: Without localization, k members provide $k - 1$ degrees of freedom for the linear combination in the analysis of the whole model domain. If the domain is decomposed into l subdomains separated by r_{Loc}, $k \cdot l$ degrees of freedom are available, reducing the sampling noise (Berre and Desroziers, 2010).

The cutoff distance r_{Loc} should be chosen larger than the grid resolution of the model so a single observation influences several grid points and the solution is smooth. If r_{Loc} is too small or too large, the solution can be dynamically imbalanced (Greybush et al., 2011).

If the analysis is taken out for the whole domain at once (like in the EnSRF), one can also use localization in state space by Schur-multiplying $(\mathbf{P}^b)^{-1}$ with a matrix that contains a similar spatial correlation pattern with zero-entries outside the predefined localization radius (Caya et al., 2005).

Patil et al. (2001) showed with an ensemble of 5 members of a global model that the local state probability density may span over a number of degrees of freedom which is lower than the number of dimensions of the model itself. The convective EnKF studies summarized in Section 1.2.4 show that an ensemble of at least 50 members seems to be sufficient for spanning the space of possible states when assimilating convective events.

Inflation of $\mathbf{R}$

In the LETKF of the KENDA-system that is used here (Section 2.1), the matrix $\mathbf{R}$ is assumed to be diagonal for computational reasons. Thus, no covariances between single observations are taken into account in the solution. If the observations are correlated and thus their errors actually possess non-zero covariances, the diagonal (and thus too small) $\mathbf{R}$ leads to a systematic overestimation of the contribution of the observations $\mathbf{y}^o$ in the computation of the analysis ensemble weights $\mathbf{w}^{a(i)}$ in (1.15) via (1.13). This can be accounted for by inflating the variance-entries in the diagonal $\mathbf{R}$. Setting $\tilde{\mathbf{P}}^b = (k-1)^{-1}\mathbf{I}$ into (1.14) yields

$$\tilde{\mathbf{P}}^a = [(\tilde{\mathbf{P}}^b)^{-1} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\mathbf{Y}^b]^{-1} \tag{1.18}$$

so the magnitude of $\mathbf{R}^{-1}$ determines how strongly the observations $\mathbf{y}^o$ and their deviations from the first guess mean $\bar{\mathbf{y}}^b$ affect the computation of the analysis weights in (1.13). Inflating $\mathbf{R}$ (meaning: multiplying it by a factor > 1) and thus deflating $\mathbf{R}^{-1}$ is therefore regarded as an instrument to correct the possibly wrong magnitude of the observation influence.

1.2.3 RMSE, Spread, Consistency Ratio, Inflation

It is common to measure the accuracy of the analysis state with the Root Mean Square Error (RMSE). In cases of real observations, it can only be computed in observation space. As this study uses a nature run as the reference, the RMSEs of analyses and forecasts are computed separately for different model variables (contained in $\mathbf{x}$) in the state space:

$$RMSE(\mathbf{x}) = \sqrt{m^{-1}\sum_{i=1}^{m}(x_i^{true} - x_i)^2} \tag{1.19}$$

The resulting RMSE-value is a scalar value. To measure the width of the PDF spanned by the analysis or forecast ensemble of k members $(\mathbf{x}^1, ..., \mathbf{x}^k)$, their mean standard deviation **spr** (the so called "spread") from the ensemble mean $\bar{\mathbf{x}}$ is computed for different model variables:

$$SPREAD(\mathbf{x}^1, ..., \mathbf{x}^k) = \sqrt{(k-1)^{-1}\sum_{l=1}^{k}(\mathbf{x}^l - \bar{\mathbf{x}})^2} = \mathbf{spr} \tag{1.20}$$

The resulting **spr** is a vector in state space with m entries representing the standard deviation of single variables at single gridpoints. To compare the spread with the RMSE, the mean is computed gridpoint-wise separately for different variables:

$$\bar{spr} = m^{-1}\sum_{i=1}^{m} spr_i \tag{1.21}$$

From here on, the word "spread" usually describes $s\bar{p}r$. If the analysis is to be consistent with the Gaussian assumptions of the filter, the ensemble spread should represent the actual error of the analysis. The "consistency ratio" CR is defined to test this requirement:

$$CR = \frac{s\bar{p}r}{RMSE} \tag{1.22}$$

If $CR = 1$, the filter works accordingly to the Gaussian assumptions. Several reasons exist why this is hardly reached in the case of convective EnKF:

- The physical states of thunderstorms and their precipitation fields are horizontally distributed in a highly non-Gaussian manner (QR and W in Figure 6) which may force the analysis ensemble to "collapse" horizontally onto the observed cells with no cells elsewhere.
- The limited ensemble size k does not cover the whole space of possible states and perturbations ($k \ll m$). Fast evolving disturbances may be missed (Kalnay, 2003, chap. 6).
- With real observations, model error will cause the members to miss developments that are contained in the observations.

Thus the spread will generically be too low, forcing the consistency ratio to be too small with $CR < 1$. This is the case of an "underdispersive ensemble".

The simplest and most common solution to this is "covariance inflation": The analysis error covariance (represented by $\mathbf{P}^a$) is multiplied at every cycle by a scalar inflation factor ρ before the analysis members are resampled by taking the square root of $\mathbf{P}^a$. Typical values are slightly above 1 (e.g. $\rho = 1.05$). This enhancement of the subspace covered by the ensemble is beneficial for the spread, but as it overestimates the background error covariance systematically, it can cause unphysical perturbations like spurious cells or other noise that is detrimental for the overall solution.

So, one needs to be careful when choosing a fixed value for ρ as is done here – a tuning process of this parameter appears inevitable when held constant. Anderson (2008) presented a method for "adaptive inflation" for EnKF that compensates the loss of variance by inflating locally more where there are many observations present and less where few observations are present, following a Bayesian approach and gaining overall more consistent analyses and lower errors at the same time.

1.2.4 Previous EnKF-studies on convection

Before the specific aims of this study will be presented in the next section, this section reviews previous studies on convective DA using an EnKF.

In the last decade, various studies about the use of local EnKF on convection have been conducted. Some are theoretical experiments with synthetic observations while some studies

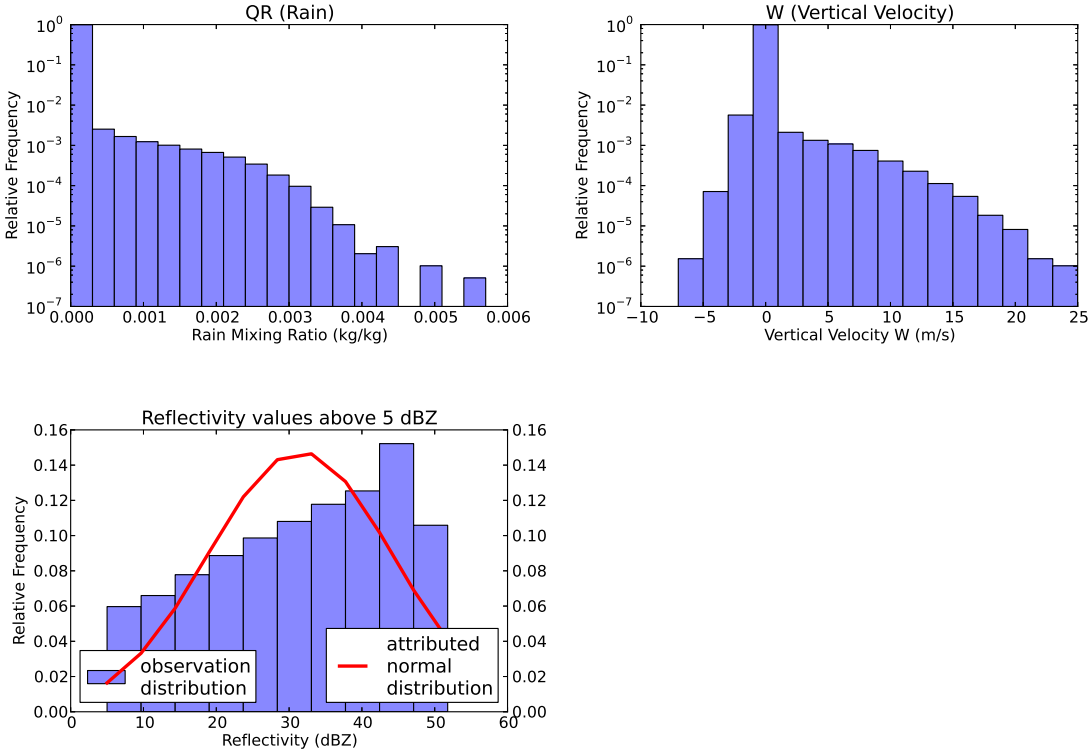

Figure 6: Relative frequencies of distinct values for QR (rain mixing ratio), W (vertical velocity) and reflectivity (retrieved value, no model variable), computed at 14:00 on model level 30 ($z \approx 3500\,m$) as an average distribution of the free forecast ensemble members depicted in Figure 16 that have not undergone any assimilation. The linear axis of the reflectivity distribution allows to show also the attributed normal distribution for the empirical mean and σ of the reflectivity.

assimilate real Doppler radar observations to reproduce the observed storm structures. The following studies all use a localized EnKF, gaining their analysis ensemble either through the "perturbed observations"-method, an EnSRF or another EnKF that localizes the background error covariance matrix $\mathbf{P}^b$. The LETKF, localizing the observation error matrix $\mathbf{R}$, appears to have not been applied so far on storm scale DA using radar data.

Primal questions to be dealed within convective DA are i) the choice of observation types, error and resolution, ii) the ensemble initialization, iii) the treatment of an underdispersive ensemble, iv) the cycling interval, v) the choice of the right covariance localization and vi) the quality of the ensemble forecasts.

Different approaches to these issues are reviewed here. Table 1 provides a supportive overview whether the studies are idealized or with real data, over their methods of initializing the ensemble and whether they use their observations multiply.

Study:	synth / real Obs	Storm Initialization in Members			Multiple Usage of Observations
		Warm Bubble(s)	Random Perturbations		
			targeted	non-targeted	
Snyder and Zhang (2003)	s		•	•	•
Zhang et al. (2004)	s	• / s-copy			•
Dowell et al. (2004)	R	•	•	•	•
Caya et al. (2005)	s		•		•
Tong and Xue (2005)	s			•	
Gao and Xue (2008)	s			•	
Xu et al. (2008)	s			•	
Lu and Xu (2009)	s			•	
Xue et al. (2006)	s			•	
Tong and Xue (2008a)	s			•	
Jung et al. (2008)	s			•	
Dowell and Wicker (2009)	R	•	•		Additive Noise
Aksoy et al. (2009)	R	•			(only bub pos)
Yussouf and Stensrud (2010)	s	•			(only bub pos)
Dong et al. (2011)	s		•		•
Dowell et al. (2011)	R		•		Additive Noise
Sobash and Stensrud (2012)	s	•			Additive Noise
Dawson et al. (2012)	R	•	•		Additive Noise
Potvin and Wicker (2012)	s	•			Additive Noise
Wang et al. (2012)	s		•		RIP

Table 1: Properties of studies mentioned in Section 1.2.4, per column: Synthetic (s) or real observations (R), approaches on the initialization of storms in the ensemble members (warm bubbles vs. random perturbations, targeted or non-targeted), multiple usage of observations (yes/no, Running in Place RIP, Additive Noise or Bubble Noise). "s-copy" refers to a complete copy of the reference storm, only dislocated in the members.

i) Observation types, error and resolution

The basic observation types are reflectivity and radial wind from a real or simulated Doppler radar. An error of 3-5 dBZ and 1-2 m/s respectively is usually chosen. Tong and Xue (2005) found the best horizontal resolution of their synthetic observations to be coinciding with the resolution of their ensemble forecast model (2 km in their case). Further studies about the optimal observational resolution were conducted by Gao and Xue (2008), Lu and Xu (2009), Xue et al. (2006), Yussouf and Stensrud (2010). They agreed that an observational resolution higher than that of the forecast model does not give any benefit to the analysis state and rather deteriorates the forecasts by forcing the model into "unphysical" (i.e. unbalanced) analysis states.

Zhang et al. (2004) found low radar observations crucially important for the cold pool analysis. An additional mesonet of surface stations that observed wind and temperature proved helpful for this purpose. The theoretical benefit of such a surface mesonet with a horizontal spacing of 20-6 km is supported by Dong et al. (2011) who found surface wind observations to be most useful for determining cold pool structures, followed by

temperature and tailed by pressure observations, regarding their magnitude of impact.

Tong and Xue (2005) were able to suppress spurious convection by assimilating simulated reflectivity observations of "clear air", instead of simply discarding them as "absent observations". The usability of assimilating such "zero-reflectivity" (or "no-rain") observations is supported by Aksoy et al. (2009) who found their analysis strongly improved and spurious convection suppressed when using "zero-reflectivity" in real-data cases.

In order to make better use of radar observations, Tong and Xue (2008a,b) estimated microphysical properties of their model with a newly developed radar operator. Jung et al. (2008) assimilated observations from a polarimetric weather radar to make use of the different polarization ratios of different hydrometeors. Potvin and Wicker (2012) showed that covering a storm with two observing Doppler radars gives a better analysis due to the better information of the 3D-windfield; they also pointed out that using observations from only one radar gives sensible results for convective EnKF, as it is done in most studies (including this one).

ii) Ensemble initialization

Unless further mentioned, all studies use a flat domain with one horizontally homogeneous sounding as the initial state of the ensemble members. To introduce variance, perturbations are added to the members' states.

In the first convective EnKF study by Snyder and Zhang (2003), these perturbations consisted merely of uncorrelated gridpoint noise in the wind and temperature field. Zhang et al. (2004) used warm bubbles/thermals to trigger storms in the members, randomly positioned around the storm of their nature run. They also experimented with perturbed copies of the reference storm.

Tong and Xue (2005) used gridpoint noise all over the domain that was smoothed with a correlation length of the expected system size. Dowell et al. (2004) targeted this noise to regions where reflectivity was observed which resulted in a better analysis with less spurious clouds.

While these studies used a horizontally homogeneous sounding from just one radiosonde for their whole domain, Stensrud and Gao (2010) compared 3DVAR analyses of strongly convective days to ensemble forecasts of the same situations, pointing out how important a realistic horizontal inhomogeneous initial state is for a good forecast of convection.

Aksoy et al. (2009), initializing the members with warm bubbles, added sinusoidal perturbations to the vertical wind profiles of the single members. This enabled the members to keep more spread throughout the assimilation window, maintaining a higher consistency ratio.

iii) Underdispersive ensemble

Most studies use cycled covariance inflation with typical values near $\rho = 1.05$ to maintain the spread. The consequent overestimation of the covariances can cause spurious convection

as observed by Snyder and Zhang (2003).

Dowell and Wicker (2009) introduced the concept of "additive noise": Their ensemble members are not only perturbed initially in locations of high reflectivity, but also throughout the cycling of the EnKF – outside the actual minimization algorithm. They observed a more accurate analysis and better spread. They argue that the additive noise helps the imperfect model to explore possible states it would not have gone into otherwise, i.e. by starting from the analytical perturbations after the resampling of the filter. This additive noise was used again by Dawson et al. (2012), Dowell et al. (2011), Sobash and Stensrud (2012). It appeared to fulfill its purpose, but showed strong dependence on the subjectively chosen variance values in the last mentioned study.

Xu et al. (2008) used time-lagged forecast states of the ensemble to construct additional virtual members with states unaccounted for in the usual simultaneous sample of members at analysis times. They used the model states lagged by 1/2 analysis-interval into the past and future from the analysis time, thus enhancing the PDF of background states to compensate e.g. for timing-errors in the positioning and propagation of storms in the model. The method appears to be computationally quite cheap as it uses states that were forecasted anyway.

Kalnay and Yang (2010) proposed the "Running In Place (RIP)"-method to accelerate spin-up and filter convergence in highly nonlinear situations such as the initiation and growth of thunderstorms. To give the ensemble the possibility to explore more of the state space suggested by the observations, the analysis weights of their LETKF are re-applied at the previous analysis time to construct a new analysis ensemble. Then the "new" ensemble is re-integrated for the same analysis interval, then re-analyzed and re-cycled again until its mean and spread satisfy the observations properly. Thus the observations are used more than once, but in an objective manner which is consistent with the Bayesian approach of the filter, rather than the subjective ad-hoc solution of "additive noise". Wang et al. (2012) tested the RIP method by assimilating simulated radar observations from a supercell and found the storm-spinup to benefit strongly from the repeatedly updated ensemble. They used an EnSRF-filter for this, not the LETKF that RIP was originally designed for.

iv) Cycling interval

All studies agree that the cycling interval between consecutive assimilation steps should be of $\mathcal{O}(5\,\text{min})$ to track the storm developments, which is also the typical time-interval of a volume-scan from a rotating Doppler radar. Aksoy et al. (2009) succesfully used an interval of 2 minutes. A more frequent update seems to be too restrictive, as the insertion of too many observations diminishes the spread strongly and the model dynamics with their limited resolution cannot benefit from observations that are too detailed.

Reference	Ensemble Size	Analysis grid	Radar obs spacing	Localization Cutoff (r)
Snyder and Zhang (2003)	50	2 km	Model grid	H: 4 km V: 4 km
Dowell et al. (2004a)	50	2 km	Radar scans	H: 6 km V: 6 km
Tong and Xue (2005)	100	2 km	Model grid	H: 8 km V: 8 km
Caya et al. 2005	100	2 km	Radar scans	H: 7.3 km V: 7.3 km
Aksoy et al. (2009a)	50	2 km	Radar scans	H: 5 km V: 4 km
Dowell and Wicker 2009	50	1 km	Radar scans	H: 6 km V: 6 km
Dowell et al. 2011	50	1 km	Radar scans	H: 6 km V: 6 km
Dong et al. 2011	50	2 km	Radar scans	H: 6 km V: 6 km
Dawson et al. 2012	30	1 km	Radar scans	H: 12 km V: 6 km

Table 1: Summary of covariance localization choices in a selection of previous convective-scale enKF studies.

Table 2: Copy of Table 1 of Sobash and Stensrud (2012)

v) Covariance localization

To gain knowledge about the optimal strategy for covariance localization, Sobash and Stensrud (2012) reviewed the choices of localization radii of recent convective EnKF studies (Table 2). They tested different combinations of horizontal (6/12/18 km) and vertical localization radii (referring to the cut-off distance of the Gaspari-Cohn correlation function). They assimilated severe convection in a typical unstable sounding with high CAPE through simulated radar observations. To initialize their nature run and ensemble members, they used warm bubbles as triggers. The warm bubbles of the members were distributed around the reference warm bubbles of the nature run[7]. They found a horizontal cut-off distance of 12 km together with 3 km in the vertical to give the best analyses. They observed a growth of (prior) correlation lengths as their initially single cells merged into a mesoscale convective system. The RMSE of their analyses only showed significant differences for smaller or larger radii within the first 100 minutes after initialization. After that, the RMSE of the different radii-experiments was regarded as comparable during the assimilation. They attribute this phenomenon to the evolution of convection, arguing that

7 a prior knowledge which will not be given in this study!

the optimal localization strategy might differ, depending on the stage of the convective evolution and thus the spatial scale of the system and its internal correlations.

vi) Quality of the ensemble forecasts

Most studies conducted ensemble forecasts from their last analysis step with lead times of 1-2 hours. They all observed rapid error growth in the storms with a horizontal divergence of the members' storms, although in most studies the members had all collapsed onto the observed storms and showed little internal variance. Aksoy et al. (2010) observed a better forecast quality for convective systems with strong internal organization like supercells and gustfronts compared to a less organized multicell storm. Using such storm analyses as the basis of very short-range forecasts of one hour is generally judged to be sensible, although the technical implementation is difficult: The process of operationally gathering the data in real-time time, processing and assimilating it, issuing ensemble forecasts from the analyses and distributing the forecast products (Stensrud et al., 2009) is critically limited by the computational power available.

1.3 Aims of this study

Properties and problems of Data Assimilation of radar observations in Ensemble Kalman Filter frameworks have been reviewed. Overall, the reviewed frameworks appear to succeed in assimilating severe storms like supercells and squall lines by producing good analyses. Their common goal appears to be a representation of the observed convective cells as accurate as possible in their ensemble. They all apply covariance localization of $\mathbf{P}^b$; a horizontal localization cutoff-radius of $r_{Loc} \leq 12\,\text{km}$ is identified to give the best results by limiting the spatial covariances to a scale smaller than the horizontal extent $\geq 30\,\text{km}$ of the observed convective systems.

These studies aimed at analysis means to be as precise and detailed as possible, having a low RMSE and a hopefully good consistency ratio. Spurious convection in the members was identified to have detrimental influence on the analyses and therefore suppressed by assimilating "no-rain" observations. This results in a "collapsed" ensemble whose members contain only the observed storms at their respective locations. Therein, the uncertainty and thus the spread only span internal variations. The model states of the members are drawn very closely towards the observations by the filter; inconsistent model states as a consequence of this seem to be possible but are difficult to quantify.

Taking the limited predictability of convection into account, and the fact that forecast products of local models are usually disseminated with lead times $> 3\,\text{h}$, the "detailed and collapsed" approach is questioned here: A coarse and non-collapsed analysis of the storms could provide short-range forecasts that are as good or as bad as forecasts from a fine analysis. The coarse analysis would need much less computational power and could prohibit unbalanced states by drawing the ensemble storms less rigidly towards the observations than in a collapsed ensemble.

1.3.1 Fine and coarse analysis using different observational localization

Such a coarse representation is developed here: Therein the variance between the members should not only be storm-internal but should also span uncertainties of storm position, shape and intensity. This can be reached by using observations that contain different amounts of detail, by relaxing the filter constraint using an inflated $\mathbf{R}$ and by localizing the observation influence on different scales. First, the different localization is illustrated:

A "fine" storm-analysis is defined here by the correct representation of location and intensity of the observed updrafts in the analysis. The error-distribution within the fine analysis ensemble should only span details of the specific storms like updraft speeds or coldpool intensity. As an LETKF-analysis is the ensemble mean plus a local linear combination of the members' deviations from the mean, a "fine" analysis state of one convective system is allowed to be an assembly of different members: Using a small r_{Loc} for $\mathbf{R}$ with a cutoff-length of 8 km, the frontal region of the storm may be mainly taken from one member while in the downdraft region a different and better fitting member might get a greater weight (R8 in Figure 1). The smoothness of the localization function hopefully produces dynamically consistent transitions between these building blocks.

In the "coarse" storm-analysis, the storms should be allowed to vary also in position within the ensemble. Errors in extent and intensity of the single clouds are then contained anyway because the single clouds in the members do not need to be exactly collocated as in the fine analysis. The reduction of the error of the ensemble mean is not the first priority of the coarse analysis which aims rather on internally consistent members around the most likely mean than on a perfect mean state. Using a larger $r_{Loc} = 32$ km, the member-weights should vary less within single storms (R8 in Figure 1), so the storms in the single members are rather whole one-members-storms than assembled many-member-storms, hopefully resulting in model states that are more consistent with the model dynamics.

1.3.2 Collapsed and non-collapsed analysis

To recreate the previous studies in Section 1.2.4, the ensemble of the posterior "fine" analysis should collapse onto the observed clouds, meaning all members should contain clouds at the observed position and no spurious clouds. The ensemble of the posterior "coarse" analysis will be allowed to *not-collapse* in order to see whether spurious storms act detrimentally on the ensemble forecasts of the assimilated storms. This non-collapse means that the rain-distribution in the coarse ensemble may be multimodal even if rain is observed with a high certainty.

Figure 7 illustrates the idea of (a) collapsing and (b) non-collapsing for a relatively good prior background ensemble with a few outliers. Assume the analysis is taken out locally for one point in the domain where a value of $Z^{obs} = 50$ dBZ is observed. The background ensemble mean predicts $\bar{Z}^b = 46$ dBZ so the predicted Z-values of the background ensemble members are already distributed around the observation, but in the outliers no storm is present and $Z^{b(i)} = 0$ dBZ.

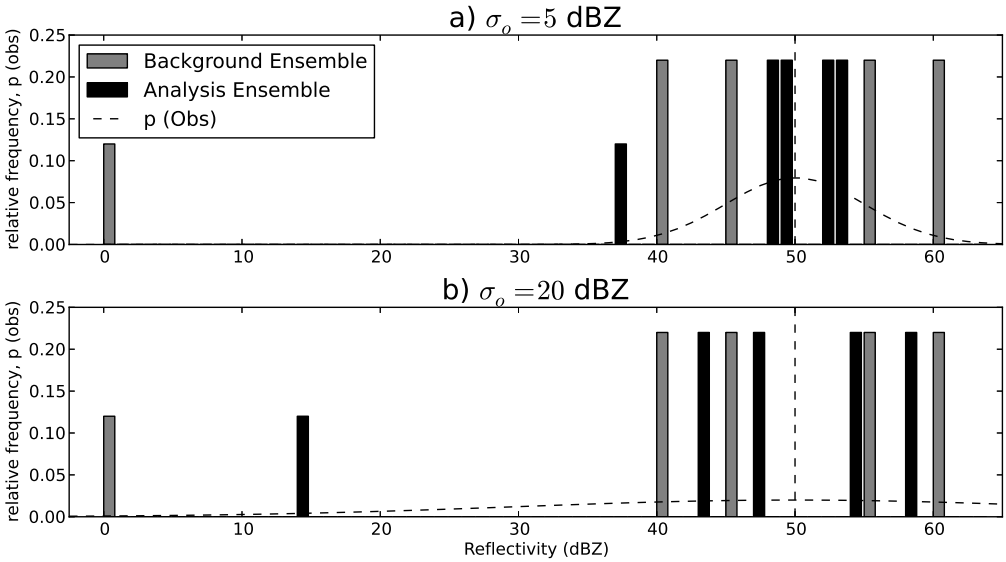

Figure 7: Relative frequencies of predicted reflectivities for a relatively good background ensemble (black) and an analysis-ensemble after determining the analysis weights of the LETKF once. The observed value is $Z^{obs} = 50\,\text{dBZ}$. In a) a small observations error of $\sigma_o = 5\,\text{dBZ}$ is reported to the filter, in b) an inflated value of $\sigma_o = 20\,\text{dBZ}$ is reported. The dashed Gaussian curve depicts shows the probability distribution of the measurement.

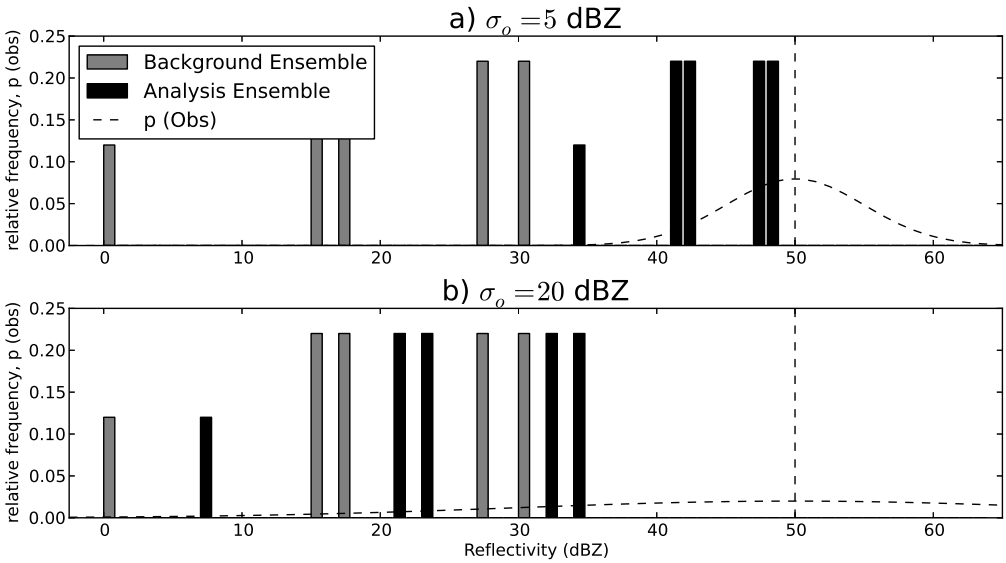

Figure 8: Like Figure 7, but for a bad background ensemble

In the collapsing case a), the observation error for $\mathbf{R}$ is $\sigma_o = 5\,\text{dBZ}$, so the analysis weights (determined as defined in Section 1.2.2) pull the good members very close to the observed value with a posterior analysis mean of $\bar{Z}^a = 49.6\,\text{dBZ}$. For the outliers, a large analysis increment of $\Delta Z \approx 37\,\text{dBZ}$ is added; this could cause an insertion shock and impose a dynamic imbalance on the model dynamics, e.g. by triggering propagating gravity waves which taint the solution – but only for the outlying members.

In the non-collapsing case b) with $\sigma_o = 20\,\text{dBZ}$ chosen, those members that were good already are not affected much and the bad outliers are only drawn gently towards the observed value with an increment of $\Delta Z \approx 15\,\text{dBZ}$. The posterior mean value of $\bar{Z}^a = 46.7\,\text{dBZ}$ is slightly improved with respect to the prior, although not as precisely as in a).

In Figure 8, the same value of $Z^{obs} = 50\,\text{dBZ}$ is observed but the prior background ensemble is bad and far away with $\bar{Z}^b = 19.8\,\text{dBZ}$. For a), all members are pushed strongly to collapse onto the observed value which might cause imbalances in all members. For b), the increments are again rather gentle, resulting in a worse but probably more stable analysis.

In both Figure 7 and 8, the spread of the collapsed ensemble a) is more diminished than for b). This is a consequence of the Gaussian constraint of the filter that tries to resample the analysis ensemble to be normally distributed by $N(\bar{x}^a, \sigma^a)$ (one-dimensional in the example); the analysis mean $\bar{x}^a$ is a weighted average of $\bar{x}^b$ and the observation y^o. If the background ensemble $x^{b(i)}$ is widely distributed with a large σ^b, a small σ^o (case a) pushes all members closely towards $\bar{x}^a$ so all members adopt the convective mode of the observation with low internal spread. If a "no-cloud"-observation is assimilated, the filter will try to suppress storms in the whole analysis ensemble, adopting the local convective mode of "no-convection".

In the non-collapsed ensemble (case b) a variety of modes is possible, including also weak or non-existent storms. The precision of the analysis is sacrificed for that, but the weaker Gaussian constraint of the filter may allow the ensemble to persist with a more "natural" distribution. The sequential update cycling of the filter should gradually bring also the non-collapsing ensemble close to the observations. A tradeoff between a) and b) could be possible – here the extreme cases are tested to investigate their properties.

Previous studies on storm-assimilation followed the collapsing approach, so the aforementioned "fine" analysis shall consist of a collapsed ensemble to recreate their results. Opposing this, the "coarse" analysis will be allowed to not-collapse.

Craig and Würsch (2012) noted that such a non-collapsing background ensemble can turn up in the case of convective data assimilation. They applied the LETKF on a simple stochastic cloud model with random birth-death-processes. Although the analyses were able to lock onto most observed clouds, spurious clouds persisted in the background ensemble and could not be compensated effectively because increments of "negative"

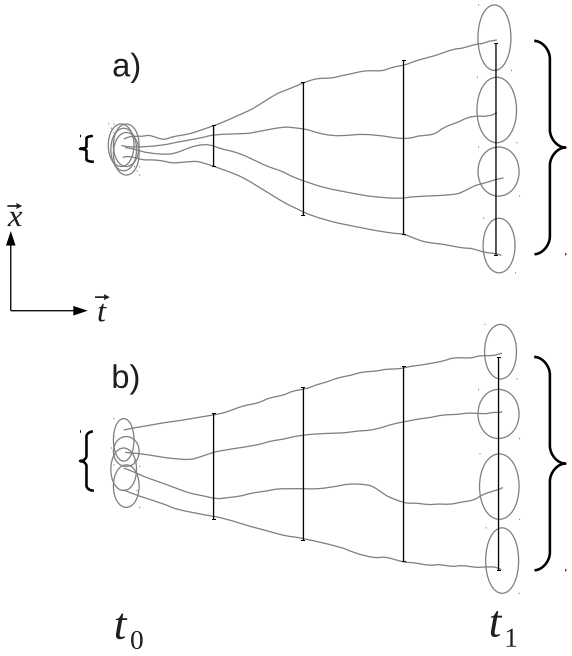

Figure 9: Schematic divergence in state space $\vec{x}$ through time $\vec{t}$ of the states of a fine collapsed analysis ensemble of one observed storm (a) and a coarse un-collapsed analysis ensemble (b). The distance of the members' storm-circles along $\vec{x}$ depicts both accuracy and spread.

clouds are physically impossible[8]; the favoring of cloudless members by the filter appeared to be insufficient to destroy the spurious clouds. The clouds in their model were only binary events, so in this study it will be seen how a model with continuous dynamics and correlations between them behaves in a collapsing or non-collapsing DA scheme.

1.3.3 Detailed and non-detailed analyses and forecasts

To assess how the predictability limit of the convective systems challenges the forecasts from the fine analysis, the fine and coarse scheme defined in Section 1.3.1 need to have different perturbation scales. To reach this, the fine system will be supplied by detailed synthetic observations with the maximum resolution of 2 km that can be gained from the nature run, so the fine analysis ensemble has to represent observed details of the storms. These detailed observations will be reported to the filter with small σ-values attributed in order to collapse the analysis onto the observed storms and their details.

The coarse scheme will be supplied with less and coarser superobservation (SO) (cf.

8 Which is not the case e.g. for variables like wind or temperature; these do not have a phase transition such as water vapor and cloud water.

Section 2.1.2) with a horizontal resolution of 8 km wherein storm-details have been averaged out, so the internal structure of the analysis-storms is allowed to vary between the members. By reporting an inflated **R** to the filter, the analysis storms of the coarse scheme will not need to collapse completely onto the observations.

With these three instruments of different localization, observation averaging and **R**-inflation, it should be possible to gain a fine and a coarse analysis ensemble by which the 3-hour ensemble forecasts can be initiated. It can then be compared how fast the perturbations of the fine analysis grow in a forecast, relative to the perturbations of the coarse analysis. It is expected that an ensemble forecast from the fine analysis will be better than a forecast from the coarse analysis in the beginning and then saturate to a comparable error level, as sketched in Figure 9, due to the nonlinear growth of small perturbations in the chaotic system of the atmosphere.

Testing categories for analyses and forecasts are the correct representation of convective modes, storm positions and fields of wind, temperature and precipitation. RMSE and spread of the model variables are evaluated and object based verification methods are applied on the Quantitative Precipitation Forecasts.

Contrary to most previous studies, the convection in the members is initiated in a "stochastic" manner comparable to the simple stochastic convection model in Craig and Würsch (2012), i.e. without "manual" determination of the storm positions outside of the actual EnKF scheme. Such a manual approach is for example conducted by Caya et al. (2005), where the initial noise in the members is targeted to locations of observed reflectivity, or by Aksoy et al. (2009), Sobash and Stensrud (2012) where the storms in the ensemble are triggered by warm bubbles placed around the assumed origin of the observed storm, giving the initial background ensemble a prior knowledge about the true storm position. It is suspected here that a part of the strong initial error reduction in those studies could be a consequence of that prior knowledge.

To avoid such a benefitial but manual manipulation of the assimilation system, this study uses initial members with fully developed convective systems that are randomly distributed throughout the model domain. This is supposed to deprive the initial ensemble and the analyses descending from it of a prior knowledge about the reference in order to make the results statistically more robust.

2 Methods

Section 2.1 contains a description of the experimental setup and Section 2.2 describes how the experiments are evaluated.

2.1 Experimental setup

To test the questions formulated in Section 1.3, numerical experiments are conducted with the COSMO-KENDA system which couples the COSMO-DE model with a LETKF[1].

In the operational KENDA, COSMO-DE is fixed onto the orography of Germany; boundary conditions are provided by the nonhydrostatic COSMO-EU model that surrounds the domain.

This study uses the testbed setup of COSMO's "artifical mode" with idealized initial state, periodic boundary conditions and a homogeneous flat landscape as the lower model boundary. Synthetic observations are drawn from a nature run which also functions as the reference. The framework that was developed for this study is named COSMO-KENDA-SOFIA[2] (Synthetic Observations From Idealized nAture-run).

2.1.1 Nature run as reference

The nature run is based on the 2 km single cell nature run contained in Bischof (2011). The initial atmospheric profile is horizontally homogeneous with the sounding from Payerne (CH, Radiosonde 06610) at 12 UTC on Juli 30th 2007, a day with severe convection over Switzerland, favored by a high CAPE value of 2200 J/kg.

The model domain has a horizontal extent of 396 km in the x- and y-direction with a horizontal resolution of 2 km and is 20 km deep. The vertical resolution varies from 150 m near the surface to 800 m at the model top.

Instead of initializing convection with predefined "warm bubbles" or targeted noise, uncorrelated grid point noise is added at the initial time to the temperature field T and the vertical wind speed W (which is zero in the basic state) in the boundary layer with amplitudes of 0.02 K and 0.02 m/s. The random distribution of the developing convective cells only depends on this initial random T- and W-noise.

1 developed by Hendrik Reich and Andreas Rhodin at DWD
2 A technical description is provided in Appendix A

To obtain results that are statistically robust, five realizations of the nature run are computed with different random seeds for the initial noise of T and W. The resulting storms cover a spectrum of small intense storms up to MCS (Figure 12).

The radiation scheme of the model is active in order to initialize convection by thermal forcing from the surface. Surface fluxes of latent and sensible heat are enabled to relax the cold pools realistically. Figure 10 shows a time series of the convective activity in the domain, while Figure 11 shows the domain-averaged properties of precipitation and storm size over the integration time of one day.

The run starts at 06:00 in the morning. The surface is heated by solar irradiation. Thermals with a typical spacing of 20 km evolve and start to rise. First preliminary showers develop around 8:00 which grow until 10:00 and die off mostly by 12:00. The remaining convective cells evolve into full single storms from 12:00 on. These intensify rapidly, develop cold pools and thus trigger new cells ahead which organize into small MCS by 14:00. The maximum activity with developing and dying cells is between 17:00 and 22:00. After that, the systems have run out of supplying air from the boundary and rain out mostly with few new cells being triggered.

As the overall activity is strong between 14:00 and 20:00 and the characteristics of the convective cells are on a high level with cells still developing and growing, this time window is used for 3 hours of assimilation between 14:00 and 17:00, followed by a 3 hour forecast.

The convective systems propagate in a north-eastern direction with a subjectively determined speed of roughly 15 m/s, steered and driven by the background wind and the density flow of the cold pools' gust fronts (cf. Figure 1). Their horizontal extent ranges from 20 - 180 km, updraft speeds reach 30 m/s. A lifetime longer than 6 hours is a typical value for single MCS in this regime (cf. Figure 12).

Periodic boundary conditions are used to let the thermals evolve properly in the mean flow, as they need more than one pass through the domain to amplify into full convective systems. Open boundaries would sweep the information out of the domain within one pass of the mean flow.

2.1.2 Synthetic observations

Doppler radar observations of radial wind and reflectivity are simulated (Figure 13) by applying the same observation operator on the nature run that is used for computing the first guesses during the assimilation.

As the observational resolution of a single radar site decreases with distance (Section 1.2.1), storms nearby the radar are observed superiorly to more distant storms in such a single-radar observing system. The positions of the reference storms in the experiments are arbitrary, therefore a homogeneous coverage of the observational system is necessary if no particular observed storm is to be favored: Observations are drawn from the nature run at a horizontal resolution of 2 km throughout the domain and a vertical resolution of 1 km, ranging from a height of 500 m up to 12.5 km. The observation points are shifted 1 km horizontally in the x/y-direction relative to the nature run's grid; the observation values

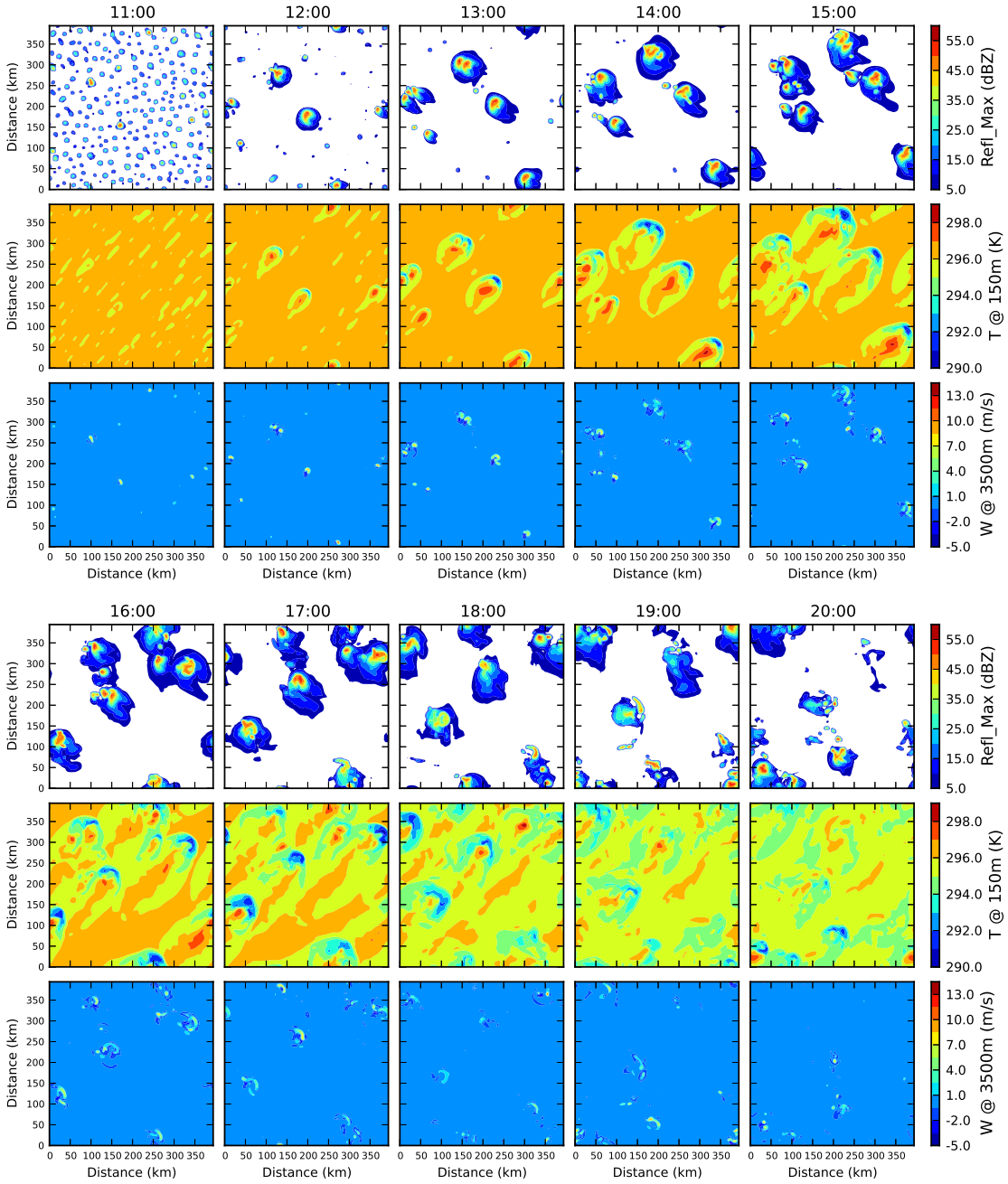

Figure 10: Time series of the nature run (realization 01). Top rows: Maximum reflectivity of column. Middle rows: The temperature at $z = 150m$. Bottom rows: The vertical velocity W at $z = 3500m$.

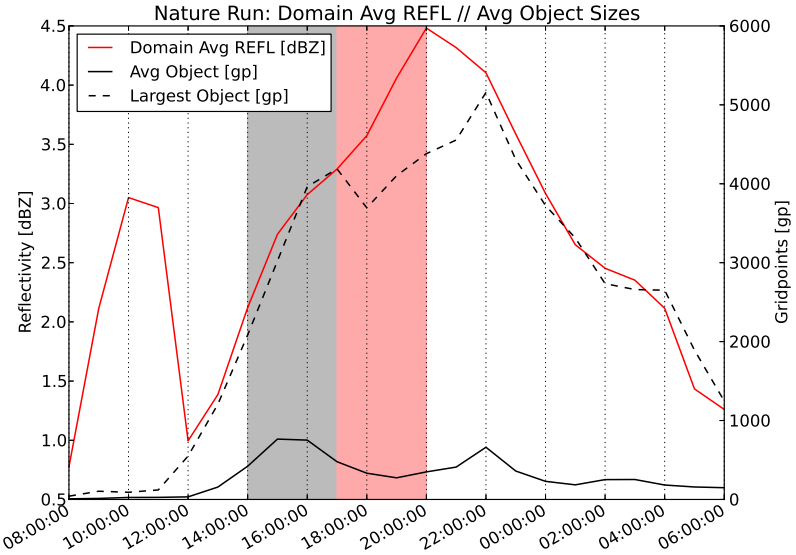

Figure 11: Time-series of the nature run (averaged over 5 realizations): Domain-average of the maximum column reflectivity in dBZ (red), together with average size (solid black) and maximum size (dashed black) of the rain-objects, thresholded to > 5 dBZ, in units of gridpoints. Assimilation window is between 14:00 and 17:00 (shaded gray), forecast window is between 17:00 and 20:00 (shaded pink). Peaks of the object-sizes can be due to merger of anvils of separate convective systems.

are computed by trilinear interpolation from the model grid. This observation geometry imitates a region covered by the crossing beams of more than one radar site. Letting the observational coverage start at a height of 500 m is quite optimistic compared to a real observation network but will provide the DA scheme with important information about the low level convergence structures (Zhang et al., 2004).

Wind observations

In a real observation system, the radar beams with the longest range are those with a very small inclination angle. These beams are almost parallel to the horizontal plane, so the radial wind they observe is mainly composed of horizontal wind components.

Following this consideration, the U-wind component is used as a substitute for a "real" radial wind here. This is equivalent to one radar site infinitely far away in the east.

Observations of U are masked to points with a reflectivity value higher than 5 dBZ. To simulate the error of measurement and representativity, Gaussian noise with a standard deviation of 1 m/s is added to the original observations.

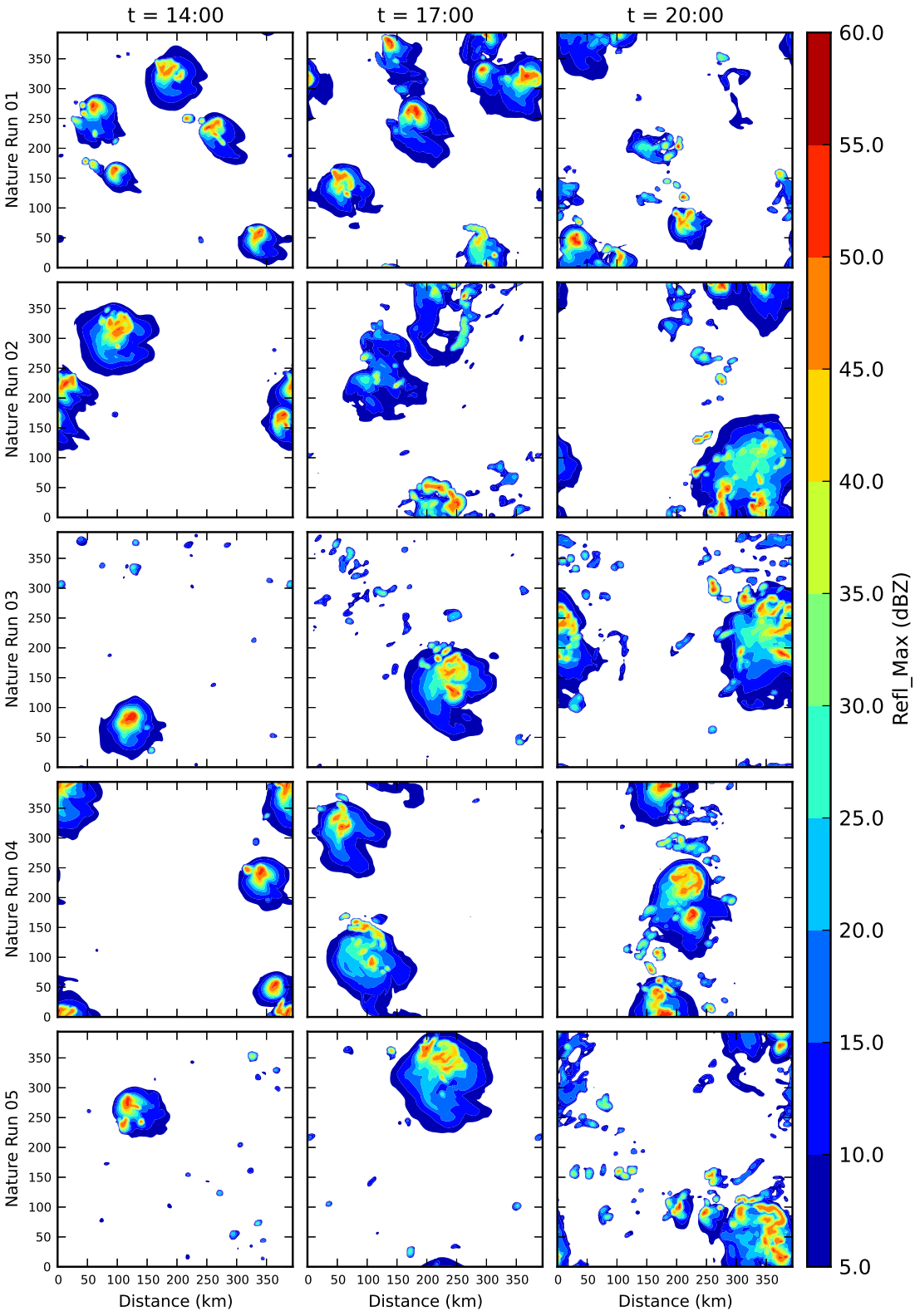

Figure 12: Five realizations (rows) of the nature run. The three columns show the time evolutions of the maximum column reflectivity. The convective systems are moving with the 45 ° background wind.

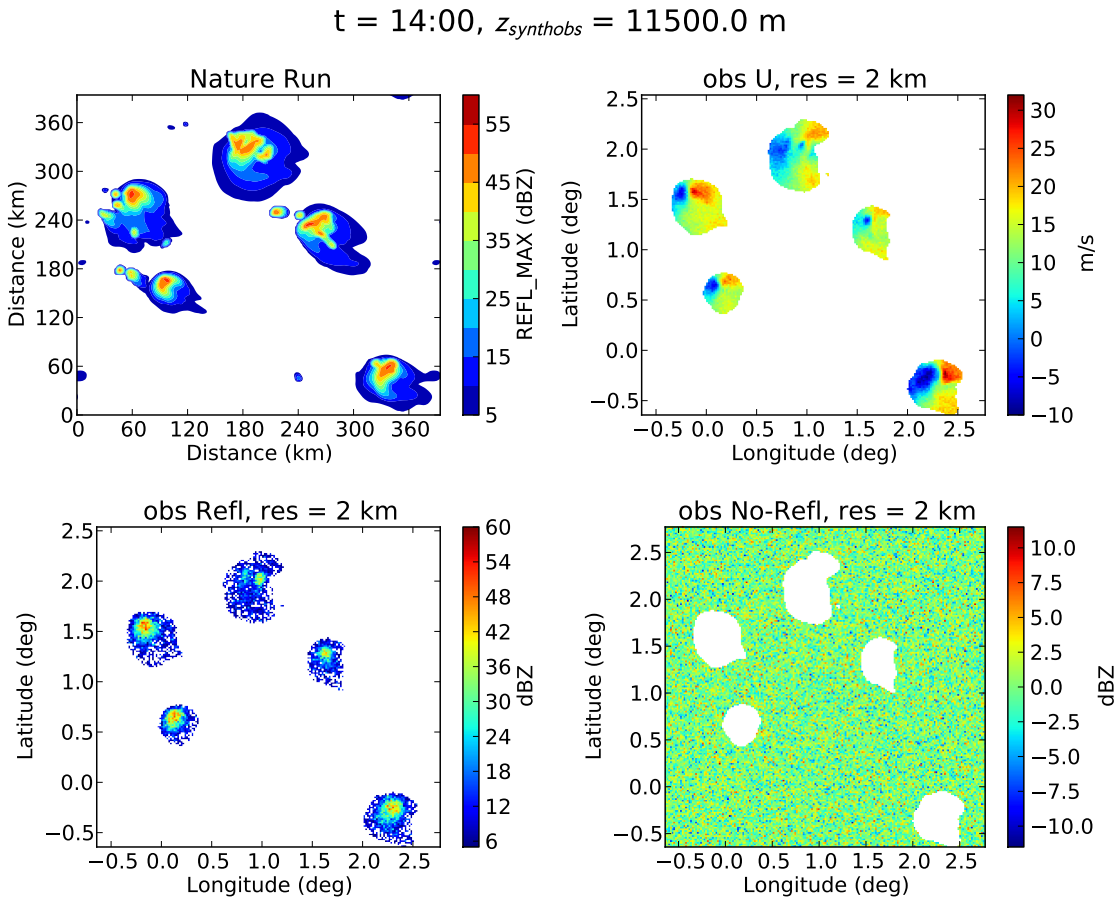

Figure 13: *REFL_MAX* of nature run 01 (upper left) and synthetic observations of U (upper right), *reflectivity* (lower left) and *no-reflectivity* (lower right) at an anvil-containing height of 11500 m. The horizontal resolution of the observations is 2 km

Reflectivity and no-reflectivity

Reflectivity observations (sometimes written as a value Z) are computed using the model variables of rain, snow and graupel following the formulation of Done et al. (2004), simulating how strongly a radar beam would be reflected by the hydrometeors. The observations are only available above the threshold of $Z = 5$ dBZ; $\sigma_Z = 5$ dBZ is also the standard deviation of the added noise. If an observed value is below the threshold of 5 dBZ, it is set to a nominal value of 0 dBZ and regarded as an observation of *no-reflectivity* (Aksoy et al., 2009, Tong and Xue, 2005) with an error of 2.5 dBZ added, assuming a good post-processing of the radar measurements that can identify regions without rain properly.

Reflectivity and *no-reflectivity* are treated as different observation types so they can be assigned with different errors to influence their respective impact in the assimilation.

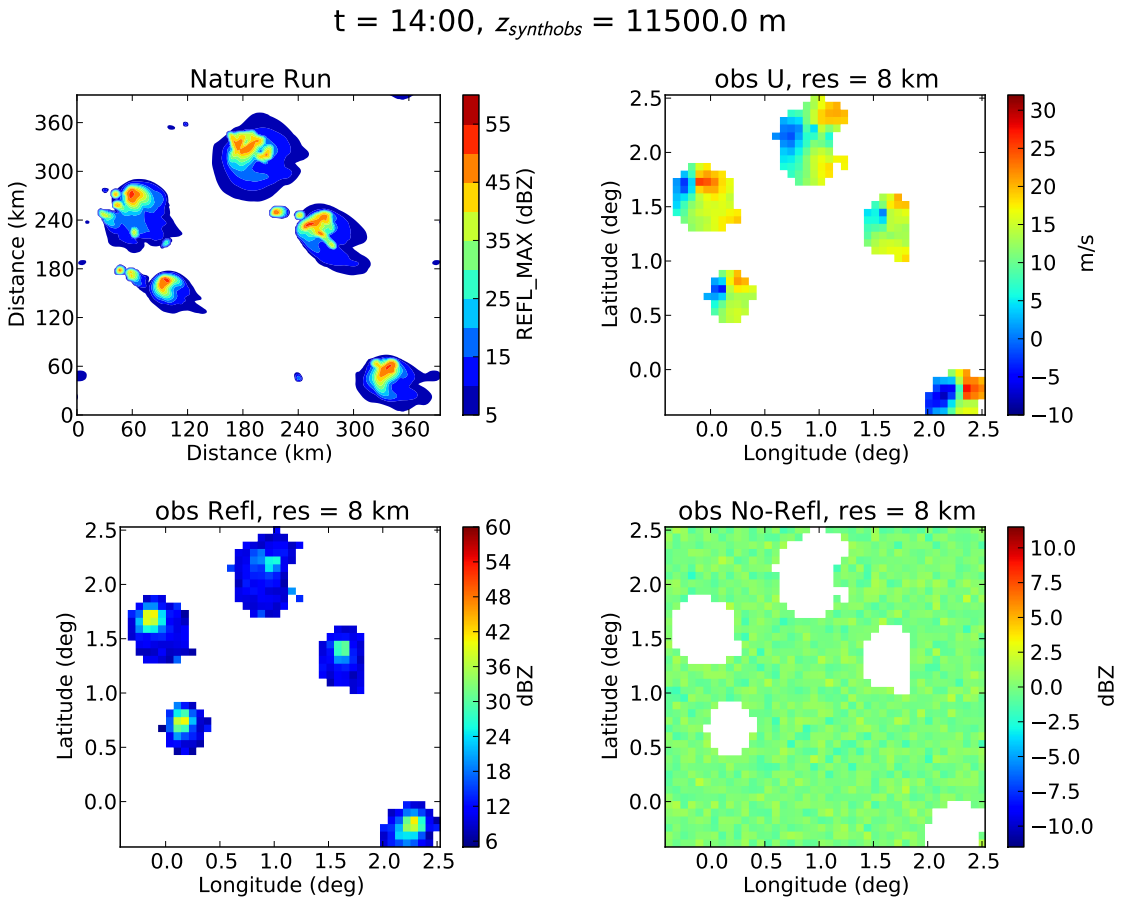

Figure 14: Like Figure 13, but coarsened superobservations with a horizontal resolution of 8 km.

Observation errors and inflation of $\mathbf{R}$

As said, Gaussian noise is added to the observations with a prescribed standard deviation σ. The observation error covariance matrix $\mathbf{R}$ that is used by the LETKF here is given by the diagonal matrix $\mathbf{R} = \sigma^2 \cdot \mathbf{I}$ with σ^2 being the true variance of the observational errors. As explained in Section 1.2.2, providing the filter only this raw matrix is not sensible as it does not contain error covariances caused by correlations of neighboring observations; the undersized magnitude of $\mathbf{R}$ would lead to an exaggerated influence of the observations in the choice of the analysis ensemble weights. This may result in unbalanced analysis increments that destroy the solution, especially with increments induced by observations of reflectivity (shown later by experiment R8_forced).

To compensate this defect, an inflated $\mathbf{R}_{infl}$ instead of the original $\mathbf{R}$ is reported to the filter. $\mathbf{R}_{infl}$ appears helpful for maintaining a good consistency ratio of the analysis

solution by keeping the spread alive. Thus, the magnitude of $\mathbf{R}_{infl}$ is regarded as a necessary tuning factor here. Observations of U have a reported standard deviation of 5 m/s instead of the added 1 m/s when they are used with their full resolution of 2 km. *Reflectivity* and *no-reflectivity* are assigned an error of 20 dBZ instead of the error of 5 dBZ.

Observation Coarsening / Superobservations

For the experiments with a coarse analysis (Section 1.3.3), the observation field is horizontally coarsened by a factor $f_c = 4$ (Figure 14) to get rid of the details of the original observations. The average value of a horizontal block of f_c^2 original observations y_j^o is the new superobservation (SO) value y^{SO}:

$$y^{SO} = f_c^{-2} \sum_{j=1}^{f_c^2} y_j^o \tag{2.1}$$

The new SO position lies in the middle of the block (cf. Figure 15). The same coarsing that is used to generate the synthetic SO is applied on the first guesses $H(\mathbf{x}^{b(i)})$ for dynamical consistence. In this block-averaging, every original observation is only used for one SO, so the observed field is coarsened but not smeared. In the averaging, every original observation has the same weight of $1/f_c^2$ – which is intended: The SO here are supposed to be spatially coarse and represent only larger scales of the observed storms without small details.

The horizontal block-averaging may seem crude, but an absolute loss of information by a distance-weighting of the f_c^2 original observations that go into one SO (Seko et al., 2004) is not the goal here; the data here should only be coarsened, not be thinned. Salonen et al. (2009), Seko et al. (2004) averaged along the radial beams of the radar and the azimutal scan angles which is not done here due to the simpler grid-shaped observation geometry.

In previous studies that used SO of Doppler radar data, the aims were a) to get rid of random errors of measurement and representativity and b) to reduce the number of observations and their resolution to an appropriate scale for the model they are assimilated in (Alpert and Kumar, 2007, Salonen et al., 2009, Seko et al., 2004). Because the averaging of the SO helps for aspect a), SO are preferable over "thinned"[3] raw radar observation data which still contains the raw random errors.

3 "thinned" meaning: a major fraction of the observation data points is simply discarded

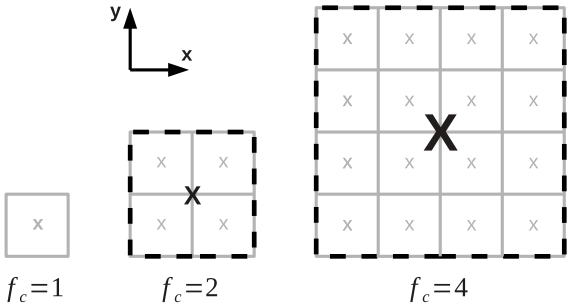

Figure 15: Observation coarsing with $f_c = 1$, 2 and 4. Gray (black) lines and crosses show the original observations (SO) with positions and extent of representativity.

Random errors of superobservation and their contribution to $\mathbf{R}$

Regarding y_j^o as a sample of a normally distributed random variable with known standard deviation σ^o, the standard deviation σ^{SO} of the SO is given by

$$\sigma^{SO} = \sqrt{\sigma_o^{-2} f_c^2}^{(-1)} = \frac{\sigma^o}{f_c} \tag{2.2}$$

This consideration as a random variable can also be regarded from the perspective of the Gaussian assumption of the filter. Reducing σ^{SO} as in (2.2) will give one SO the same weight in the analysis as the bulk of the original observatios. This is now shown shortly:

To give one SO the same weight in the analysis, its contribution to the cost function needs to be the same as for the bulk of the original observations. One can simplify the observational part of the cost function (equation 1.8) to

$$J_{obs} = [\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x})] = \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d} \tag{2.3}$$

$\mathbf{R}^{-1}$ is diagonal:

$$\mathbf{R}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sigma_n^2} \end{pmatrix} \tag{2.4}$$

σ_i^2 denotes the variance assigned to the i-th observation. So one can simplify J_{obs} further to get

$$J_{obs} = \sum_{i=1}^{n} \frac{d_i^2}{\sigma_i^2} = \sum_{i=1}^{n} \left(\frac{d_i}{\sigma_i} \right)^2 \tag{2.5}$$

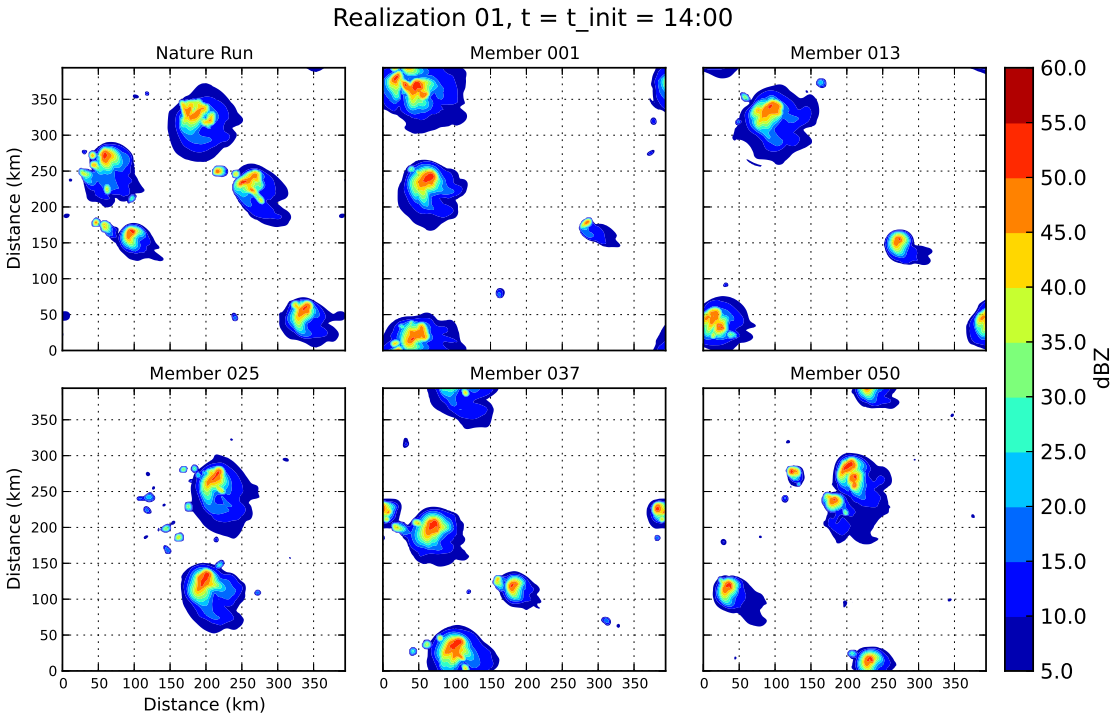

Figure 16: Maximum reflectivity of nature run and forecast ensemble members 1,13,25,37,50 at 14:00 (realization 01), before the first assimilation step

The length l of the new SO vector $\mathbf{y}^{SO}$ is given by $l = \frac{n}{f_c^2}$. The respective part of the cost function with SO is

$$J_{SO} = \sum_{j=1}^{l} \left(\frac{d_j}{\sigma_j} \right)^2 \tag{2.6}$$

If the SO should to have the same quantitative contribution to the cost function (and thus the solution) as the original observations, the equality (2.5) = (2.6) must hold. Assuming a high number of independent observations, the consequential linearity of the sum yields

$$\sum_{i=1}^{n} \left(\frac{d_i}{\sigma_i} \right)^2 = f_c^2 \sum_{j=1}^{l=\frac{n}{f_c^2}} \left(\frac{d_j}{\sigma_j} \right)^2 = \sum_{j=1}^{l} \left(\frac{d_j}{\frac{\sigma_j}{f_c}} \right)^2 \tag{2.7}$$

which is equivalent to (2.2).

2.1.3 Convective Ensemble

An ensemble of 50 members is used to sample the background error covariance in the LETKF. The members differ from the nature run only in the random realization of their initial T and W-noise. The convective systems in the members evolve analogously to the nature run with similar characteristics in terms of onset time, updraft speed, extent and organization. Only their spatial distribution is completely uncorrelated between the members (Figure 16).

The assimilation experiments R8 and R32 are repeated five times with different initial perturbations for the ensemble and for the nature run (Figure 12), so observations from every nature run are assimilated with an ensemble independent from the other ensembles. The results are then averaged.

2.1.4 Assimilation Setup

The cycled assimilation starts at 14:00 when the convective systems have fully developed (Section 2.1.1). It is assumed that at this time the internal correlation lengths of the systems have finished their upscale growth, so the coarse experiment is not additionally challenged by small correlation lengths when only intensifying systems are present (Sobash and Stensrud, 2012). The 3-hour period of assimilation from 14:00 to 17:00 is followed by a 3-hour ensemble forecast until 20:00.

Fine vs. Coarse Experiments

As devised in Section 1.3, a "fine" experiment R8 is set up to yield a collapsing analysis ensemble; opposed to R8, a "coarse" experiment R32 is set up to allow the ensemble to not collapse (Table 3).

R8 has a localization radius of $r_{Loc} = 8\,\text{km}$ and a fine observational resolution of $\Delta x^{obs} = 2\,\text{km}$. $\sigma_U^{obs} = 5\,\text{m/s}$ and $\sigma_{refl,no\_refl}^{obs} = 20\,\text{dBZ}$ are inflated values (for $\mathbf{R}$) of the originally added noise (cf. Section 2.1.2) that were found by tuning to prohibit imbalanced increments which arise when the analysis state is drawn too closely to the observations. To justify this, in the experiment R8_forced no $\mathbf{R}$-inflation is applied for the reflectivity observations by using the uninflated error values $\sigma_{refl,no\_refl}^{obs} = (5, 2.5)\,\text{dBZ}$. A cycling interval of $\Delta t_{ass} = 5\,\text{min}$ is chosen to track detailed developments in the observations.

R32 has a localization radius of $r_{Loc} = 32\,\text{km}$ and coarsened SO with a resolution of $\Delta x^{SO} = 8\,\text{km}$. R32 should produce "coarse" analyses with an ensemble that has convective cells of the roughly right intensity in the roughly right locations without collapsing solely onto the observed storms. The spatially less detailed SO will not force the members to look too similar. The larger localization of 32 km should let the filter choose linear combinations of the members that vary smoother horizontally; more ample structures of single members are therefore preserved in the analyses members, opposed to the analyses of R8 that can be assembled from many different members. A slower updating interval of $\Delta t_{ass} = 20\,\text{min}$ is used in R32 to let the members develop significant changes in the scale

	R8	R32	R8_forced	R32_forced
$r_{Loc,hor}$ (km)	8	32	8	32
$\Delta x_{obs/ana}(km)$	2	8	2	8
n_{lev}^{ana}	20	25	20	25
Δt_{ass} (min)	5	20	5	20
σ_U(m/s)	5	5	5	1.25
σ_{refl} (dBZ)	20	20	5	5
$\sigma_{no-refl}$ (dBZ)	20	20	2.5	5

Table 3: $r_{Loc,hor}$ is the cutoff length of the horizontal covariance localization function, $\Delta x_{obs/ana}$ is the horizontal resolution of observations and of the analysis grid, n_{lev}^{ana} is the number of vertical levels of the analysis grid, Δt_{ass} is time-intervall between the two consequent analyses, σ_U, σ_{refl} and $\sigma_{no-refl}$ are the standard deviations of the observations (for R32: superobservations) contained in $\mathbf{R}$ for the filter, inflated for R8 and for R32.

of the SO during the longer forecast interval. This also means that in R32 less information per time is provided than in R8 which contributes to the "coarse" aspect of R32[4].

To allow a real non-collapse in R32, the same σ-values are chosen for the superobservations as for the original observations in R8: $\sigma_U^{SO} = 5\,\mathrm{m/s}$ and $\sigma_{refl,no\_refl}^{SO} = 20\,\mathrm{dBZ}$. This is equivalent to an additional $\mathbf{R}$-inflation in R32: The number of SOs coarsened with $f_c = 4$ is $N_{SO} = 1/16 N_{obs}$ of the number of original observations, so $\mathbf{R}_{R32}$ has 1/16th of the dimensions of $\mathbf{R}_{R8}$. If the magnitude of $\mathbf{R}_{R32}^{-1}$ should be the same as the magnitude of $\mathbf{R}_{R8}^{-1}$ in equation (2.3), σ^{SO} would need to be multiplied by 1/4 (cf. equations (2.2) and (2.7)).

In R32, the influence of the superobservations on the choice of the linear combinations of analysis members is less than of the original observations in R8. This contributes to the coarse and non-collapsing aspect of R32. This consideration is particulary tested by the experiment R32_forced with $\sigma_{refl,no\_refl}^{SO} = 5\,\mathrm{dBZ}$ and $\sigma_U^{SO} = 1.25\,\mathrm{m/s}$ for $\mathbf{R}_{R32\_forced}$; these σ-values were determined referring to the σ-values of R8, following equation (2.7).

Filter settings

The analysis weights are computed by the LETKF on an analysis grid with a horizontal resolution Δx_{ana} equal to the horizontal resolution of the observations Δx_{obs}. When the

4 An interval of 5 minutes for R32 did not give a better analysis but rather a worse one due to the frequent introduction of noise

analysis grid is coarser than the model grid, the analysis weights are first evaluated on the coarse grid and then interpolated onto the model grid. For R8, the observations and the analysis grid have the full horizontal resolution of the model (2 km). For R32, the analysis grid has a horizontal resolution of 8 km; at model points that lie between observation points horizontally, the analysis weights vary linearly. This causes the analysis state of R32 to contain more structures from single members that are horizontally more coherent than in R8, supporting the fine/coarse approach in Section 1.3. This is also illustrated in Figure 1.

The analysis grid has $n_{lev}^{ana} = 25$ vertical levels for R32 and $n_{lev}^{ana} = 20$ for R8, the latter because of memory-limitations. Common to both experiments is a vertical localization cutoff radius of 3 km near the surface and 5 km at the model top, and a constant covariance inflation factor of $\rho = 1.05$. This ρ-value appears slightly too large for R8, overinflating the domainwide spread with time, but gives reasonable results.

All model variables are updated in the analysis with the exception of the snow mixing ratio (QS). A snow-update seemed to deteriorate the solution, possibly due to the ambiguous projection of rain, graupel and snow in the *reflectivity* operator.

The solution of the analysis is slightly smoothed, prohibiting unphysical negative moisture values and strongly nonhydrostatic pressure increments that would taint the analysis by their noise.

The LETKF of KENDA is designed for the purpose of operational DA, therefore its analyses are non-periodic, in contrast to the present model BC. To make the analysis state stable and noise-free, the analysis increments at the boundaries need to be zero, so all observations within the distance r_{Loc} to the lateral model boundaries are discarded. This is evident when comparing the observational coverage of R8 in Figure 13 to the coverage of R32 in Figure 14. As a consequence, the quality of the analyses in the bordering regions is inferior to the inner regions (Figure 17). This static deficiency at the borders is partly compensated by the filter information which is dynamically advected from the inner region by the 45° mean flow. Due to this direction of the flow, the regions with the least influence of the filter are the upper-left and lower-right corner. If a convective system leaves the domain near to the upper-right corner, it has to pass this particular "region" due to the periodicity, but it will re-enter the well-observed inner domain afterwards.

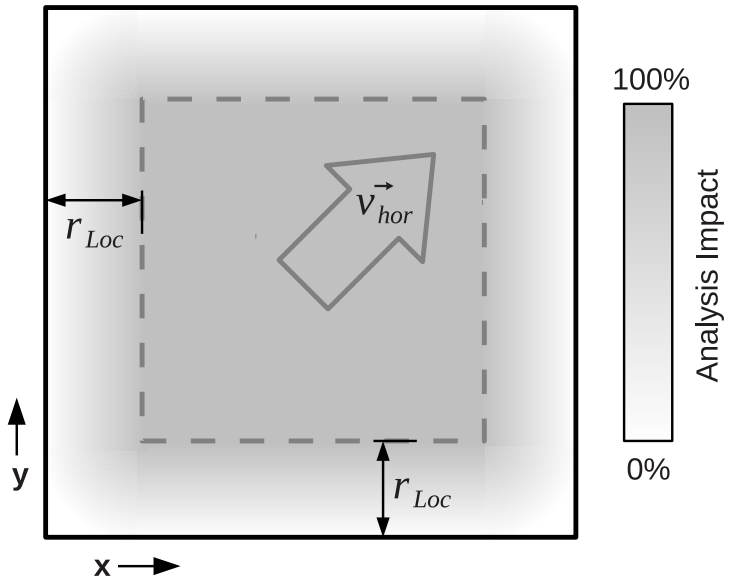

Figure 17: Model domain from above. The shading shows the analysis impact due to observations present in the inner domain (dashed box) and due to discarded observations within r_{Loc} to the domain boundaries. $\vec{v_{hor}}$ denotes the horizontal mean flow.

2.2 Evaluation: Errors and Skill-Scores

The error of analysis or forecast states is assessed here with different measures. Quantitative measures like the RMSE are combined with object-based skill scores. The ensemble forecast skill is interpreted using the Brier Score.

During the cycling of the filter from 14:00 to 17:00, errors and scores of forecast and analysis are plotted simultaneously, resulting in a "saw-tooth" pattern of the time series: The error is expected to grow from the initial state $\bar{\mathbf{x}}^a_{t_0}$ to the forecast state $\bar{\mathbf{x}}^b_{t_1}$ during the forecast intervals. After the filter update at t_1, the error of $\bar{\mathbf{x}}^a_{t_1}$ should be lower than the error of $\bar{\mathbf{x}}^b_{t_1}$, if the filter works correctly. The sharper the spikes are, the larger the filter increments from forecast state to analysis are. This pattern also holds for the spread which is usually reduced in the analysis step.

In the free forecast time between 17:00 and 20:00, only the forecast states are evaluated, so the error and spread appear as continuous lines without a saw-tooth pattern superimposed.

2.2.1 Quantitative Evaluation: RMS Error

RMSE, spread and consistency ratio are defined in Section 1.2.3. To evaluate the quality of a state, the RMSE of the following model variables are computed in model space with the nature run as the reference state: U, V, W and T show the wind and temperature

field while QR, QS, QG (rain, snow and graupel mixing ratio) show the field of actual precipitation. Additionally, the error of the observed variable *reflectivity* ($REFL$) is evaluated on the model grid to see the impact of the respective observations. To estimate the consistency ratio, the spread of the variable is plotted together with the error values.

The RMSE is computed on the set of all gridpoints and then averaged. In most studies presented in Section 1.2.4, RMSE and spread are evaluated only on those points with a reflectivity larger than e.g. $Z = 10\,\mathrm{dBZ}$. That method gives detailed information about the filter activity because wind- and reflectivity observations of a radar are usually only available above a certain reflectivity threshold. The displayed error values are relatively high due to the strong convective activity on the reflecting subset of points and the error curve tends to go down distinctly.

If the error is evaluated on all grid points as it is done here, many inactive regions with low error values are averaged in, causing the average error value to be relatively low. The activity of the filter is still recognizable in these sawtooth plots but may seem less distinct because only a subset of the points is affected by the filter updates. Nonetheless, the influence of the filter on the *whole* field is evaluated this way, and as model forecasts are initiated by full physical fields, this is a challenge the EnKF has to face.

2.2.2 Object-based Evaluation

For quantitative precipitation forecasts the 2D rain-field is represented here by the maximum reflectivity of the overlying atmospheric column ($REFL\_MAX$). This field is thresholded above $Z = 10\,\mathrm{dBZ}$ and looks similar to operational radar products. Convective systems with intense cores and stratiform anvils can be identified well.

Categories for the evaluation of forecasts (and of analyses, during the cycling) are:

a) the right position of the rain-objects (correct or shifted)

b) the right amplitude of the rain-field (over- or underestimation)

c) the right mode of precipitation (strong and pointed rain-features or weak but stratiform rain-areas)

Two skill scores are chosen to check these requirements in analyses and forecasts. The SAL-score (Wernli et al., 2008) gives a quantitative measure for categories a)-c) with three separate scalar score values. The DAS-Score (Keil and Craig, 2009) is used here for b) to check the displacement of the rain fields. A short description of both follows.

Skill scores that are designed to take missing observations into account such as the Equitable Threat Score (ETS) are not used here because in the present perfect-model experiment the whole field is observed.

Structure, Amplitude and Location Skill Score (SAL)

A schematic SAL-scoring of simplified objects is shown in Figure 18.

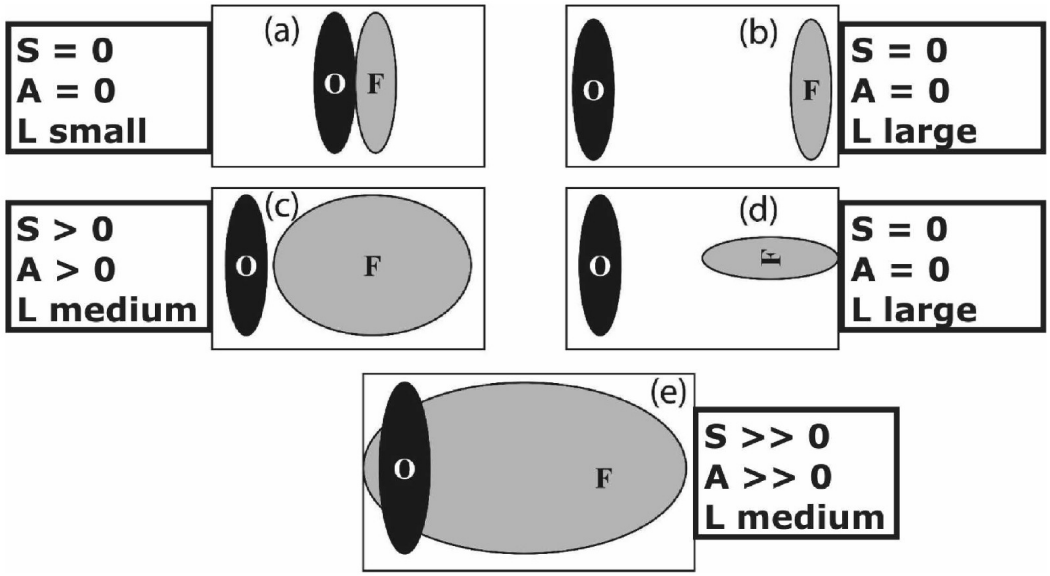

FIG. 1. A schematic example of various forecast and observation combinations, modified from Davis et al. (2006a). For the qualitative application of SAL, it was assumed that precipitation rates are uniform and the same in all objects.

Figure 18: Schematic overview on SAL-properties (original Fig. 1 from Wernli et al. (2008)

To check for the right rain-amount of the overall forecast (category b), the average rain-amount of the raw observation field is divided by the average rain-amount of the forecast field, without any thresholding. This yields the component SAL-A which can also be regarded as the overall bias of the forecast. A positive value of SAL-A represents an overestimation in the forecast, a negative value an underestimation.

The right location of the forecast rain (category a) is checked by SAL-L. Observation and forecast of *REFL_MAX* are first thresholded, then the respective centroids of the contained rain-objects are computed. SAL-L consists of two parts L1 and L2. For L1, the position of the mean centroid in the forecast field is compared to the position of the mean centroid in the observation field. The spatial difference of these is normalized by the diagonal extent of the domain; e.g. a location difference of 50 km of the two centroids on a $400 \times 400\,\mathrm{km}^2$-domain yields L1 $= 50/(400 \cdot \sqrt{2}) = 0.09$. L1 has one handicap: Assume one storm is observed but two distinct storms are forecasted which happen to have their common centroid at the same location as the centroid of the one observed storm: L1 is zero in that case. This is compensated for by L2: The mean radial distance of the single objects' centroids from the common centroid can distinguish between "one centered storm" and "two storms around that position", so if this property is not equal in observation and forecast, L2 is > 0. The sum SAL-L = L1 + L2 is thus unambiguous as it measures the

displacement of the mean centroids of forecast and observations via L1 and compares the respective distances of the rain-objects from the mean-centroid via L2. SAL-L is always positive; a value of 0 is a perfect match.

An important category for the consistency of the forecast is c): The question if the right *type* of rain-event is forecasted. To check this, SAL-S divides the volume (the integral over the product of extent and rain-content) of the single objects by their respective maximum value, then computes the average value of these "normalized volumes" for forecast rain objects and observation rain objects, respectively. Spiked objects e.g. like small thunderstorms of $30 \times 30\,\mathrm{km}^2$ with intense rain-cores of 50 dBZ can be distinguished from blurry stratiform rain regions extending $100 \times 100\,\mathrm{km}^2$ with a core of 30 dBZ. A positive value of SAL-S means "forecast rain objects too blurry" while a negative value is attributed to "forecast rain objects too spiked".

The fact that SAL only compares the mean properties of rain-objects means that there is no need to match specific objects. This kind of evaluation is useful in a regime where rain-objects appear rather stochastically (as in the present setup) and their properties (the convective modes) are mainly predefined by the environmental sounding.

Another advantage of SAL is that the assessment of forecast location by SAL-L is independent of the amplitude SAL-A, and the assessment of the forecast structure by SAL-S is independent of the location SAL-L.

Displacement and Amplitude Score (DAS)

If there are specific thresholded objects to be compared between observation and forecast, the DAS-score matches them. DAS consists of the two components AMP, measuring the amplitude error, and DIS, measuring the displacement error of the forecast. The AMP component is not used here as the overall bias of the forecast field is already measured by SAL-A. To compare observed and forecast field, the DAS-score uses a pyramidal matching algorithm where distinct objects are first matched on a coarsened field and finally on the full resolution. The field is coarsened here by a maximum exponent of 3, so one gridpoint of the most coarse field is constructed from 2^3 gridpoints in the full field, per x/y-direction. A search-radius of ± 2 gridpoints per x/y-direction is used for the object-matching going down the pyramidally coarsened fields. The maximum horizontal search radius on the present model grid is thus

$$R_{DAS}^{max} = \pm 2 \; \Delta x_{model} \; 2^3 \; \sqrt{2} = \pm 2 \cdot 2\,\mathrm{km} \; 8 \; \sqrt{2} \approx 45\,\mathrm{km} \tag{2.8}$$

where the diagonal factor $\sqrt{2}$ indicates that the matching search is performed on both x- and y-axis. A maximum search radius of 45 km is considered to be a fair limit for the displacement of an assimilated thunderstorm system that is forecasted over 3 hours.

The matching algorithm uses an optical flow technique and computes two vector fields that morph the observation field onto the forecast and vice versa (cf. Figure 19). The average magnitude of these displacement vector fields is measured by the DIS-component

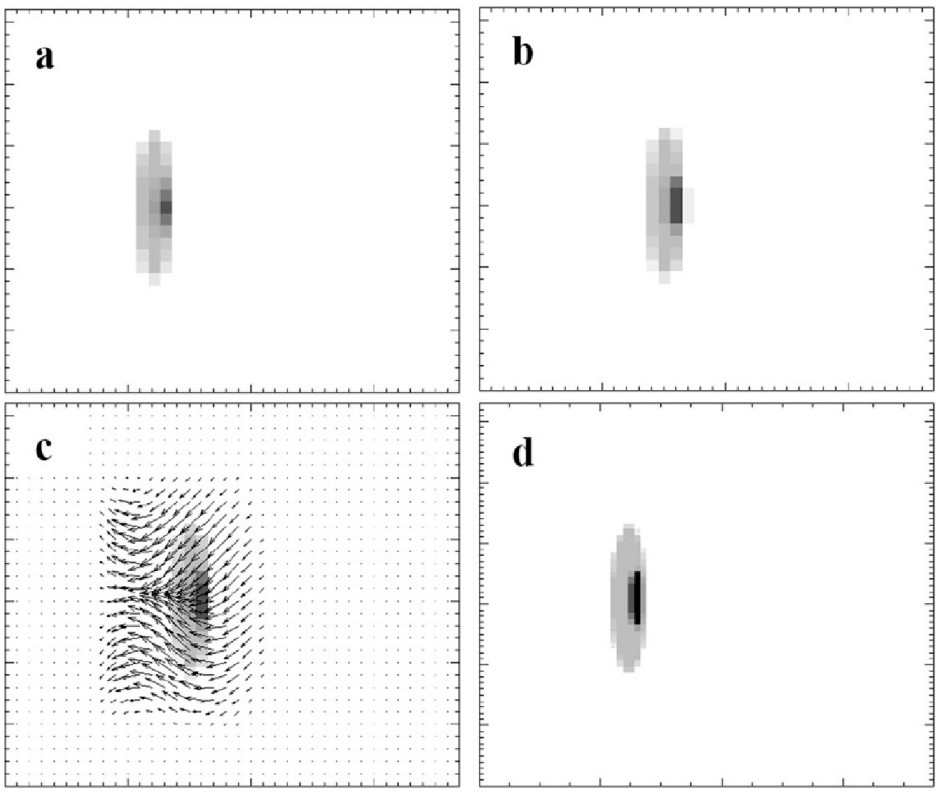

Figure 19: Scheme of the matching in DAS: (a) observed field. (b) forecast field which is similar to the observed field, but displaced to the right. (c) vector field that morphs (b) onto (a), resulting in the morphed forecast field (d). (original Fig. 2 from Keil and Craig (2009))

of the DAS-score. A positive DIS-value is a location error; a DIS-value of 0 is either a complete miss (of the forecast field with respect to the observed field) or a perfect match – neither of which appear in the relatively precise forecasts of this study, so the evaluation using DIS is regarded as safe and unambiguous.

To perform the pyramidal matching, a bordering region that is $\pm 2 \cdot 2\,\mathrm{km}\ 8 = 22.5\,\mathrm{km}$ wide needs to be discarded – a necessary evil that should not spoil the statement of the score.

The DAS-score is applied on the single members of a forecast or analysis and averaged over the members. This is more sensible than comparing the ensemble mean rain-field to the nature run as the storms in the R32-mean are not collapsed, and in both R8/32 the members' storms diverge during the forecast and therefore blur the mean precipitation field.

Comparison of DIS and SAL-L

The fact that DIS measures the displacement error by the mean vector-magnitude of directional object-displacements makes it comparable to SAL-L. SAL-L does not bother with matching distinct objects, but the measured location-differences are effectively similar to the directional displacement vectors of DIS: the displacement of the mean forecast centroid from the observation is given by SAL-L1, and if SAL-L1 fails, SAL-L2 compensates by applying a penalty if the distances of the rain-objects in the radial direction from the mean centroid of the forecast are differing from these distances in the observation.

This equality of DIS and SAL-L is independent of the domain-size: Assume a large domain with many small storms (like the situation at 11:00 in Figure 10) and a perfect forecast, resulting in DIS = SAL-L = 0. Now the forecast field is disturbed by shifting one small storm slightly to the side. The mean vector-magnitude of DIS will indicate the displacement of one matched object; DIS will be > 0 and very small because the rest of the displacement vector field is zero. SAL-L will recognize the shift also because the mean centroid is shifted; SAL-L will be > 0 and very small because the shifted centroid distance is weighted by the diagonal extent of the domain, similar to the weighting by the domain-area in the mean vector displacement of DIS where a lot of no-displacement-zeros are averaged in if the rest of the forecast field is perfect.

Now assume all forecasted small storms are shifted randomly: The one half slightly to the right, the other half slightly to the left. DIS will be > 0 but relatively large because all objects are displaced. SAL-L will still be zero because the distances of the mean centroids (L1) and the mean-distances of the objects to the common centroids (L2) do not change. In cases where convective cells pop up stochastically in random places but homogeneously distributed on a larger scale (e.g. with a good forecast of a weakly forced convective regime), SAL-L does not regard this as problematic – in contrary to DIS which will penalize the random displacements.

The setup in this study contains stochastic cloud fields (like 11:00 in Figure 10) before the assimilation window, then a distinct storm field during the assimilation window, and then, in the ensemble forecast window, a field of storms that move apart from the positions they were given by the filter. This variety of field-properties lets the usage of both DIS and SAL-L appear reasonable.

2.2.3 Ensemble Forecast Evaluation

To accompany the RMSE and object-based scores, the probabilities of the ensemble forecasts are evaluated using the Brier Score (Brier, 1950) und the method of Relative Operating Characteristics (ROC).

A convective ensemble forecast produces a probability distribution of the future rain field. The binary event of “occurence of an intense thunderstorm” is here defined as a gridpoint-value of $REFL\_MAX$ above a threshold of $30dBZ$. The nature-run gives an observation field with values of 0 (no storm) and 1 (storm). The forecast ensemble yields a

		Event Observed? Yes	No
Event Forecast?	Yes	a	b
	No	c	d

Table 4: Contingency table for binary event (Yes/No)

probability distribution field with values ranging from 0 (no storm in any forecast member) over 0.5 (storm in 50% of the members) to 1 (storm in all members), discretized in steps of $\Delta p = 1/N_{members}$. The Brier Score is a verification measure for probabilistic forecasts:

$$BS := \frac{1}{n}(f_k - o_k)^2 \tag{2.9}$$

where k denotes a set of n forecast-event pairs, f_k the forecast probability and o_k the collocated observation. The Brier Score is negatively oriented and represents the mean squared error of the forecast probabilities, so $BS = 0$ is desirable. For a detailed description of the Brier Score and its decomposition see Wilks (2006).

For a chosen spectrum of forecast probabilities between 0 and 1 of a binary storm event, contingency tables (Table 4) can be constructed to compare the agreement of the forecasts with the observations. Hit rate H and false alarm rate F are defined as

$$H = \frac{a}{a+c}, \qquad F = \frac{b}{b+d} \tag{2.10}$$

$H(F)$ for different forecast probabilities can be plotted in a Relative Operating Characteristics (ROC)-diagram (Kober, 2010, Stanski et al., 1989) like shown in Figure 20. F should be zero for all forecast probabilities, and the probability should be high that an observed binary event lies within in a smeared forecast field, meaning H should increase going towards lower forecast probabilities. The area below the ROC-curve should therefor be 1 for a perfect probabilistic forecast. This "ROC-area" is used as an additional probabilistic measure.

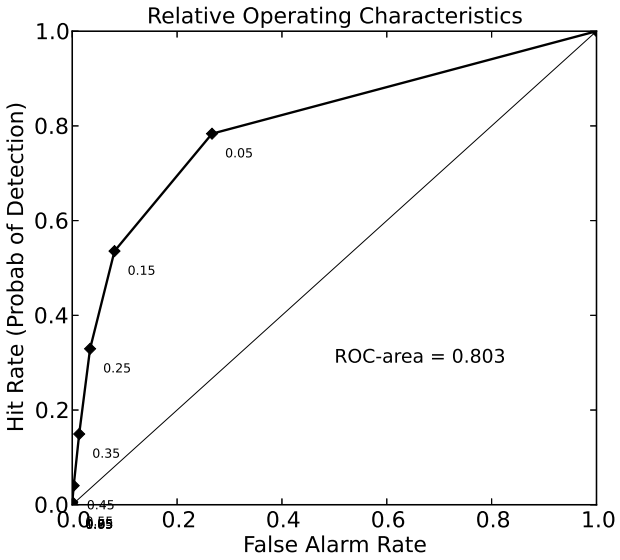

Figure 20: ROC-diagram, constructed from the experiment R8 (later shown) for a 3 hour forecast of events where $Z > 30\,\mathrm{dBZ}$. Small numbers next to the data points denote the forecast probabilities.

3 Results

Section 3.1 evaluates the general performance of the LETKF in the assimilation of the nature run's thunderstorms by a convective ensemble, using the experiment. Therein, results of the Ensemble Kalman filtering are evaluated subjectively and quantitatively by RMSE and spread of the model fields. The focus lies on R8 in the assimilation window between 14:00 and 17:00 to see if the LETKF is able to create a comparably good storm-analysis like the EnSRF-storm-analyses reviewed in Section 1.2.4.

Section 3.2 compares the fine experiment R8 and the coarse experiment R32 in terms of analysis quality and in terms of the collapse of the ensemble onto the observed clouds. R8/32 are evaluated in the forecast window from 17:00 to 20:00 subjectively and quantitatively by the object based skill scores SAL and DAS.

3.1 Performance of the LETKF

Basic properties of the assimilation process are evaluated qualitative in Section 3.1.1 and quantitatively in Section 3.1.2 on the basis of the R8-experiment.

3.1.1 Qualitative examination of R8

Figure 21 and 22, 23, 24 show the analysis ensemble means of the R8 and R32-series versus the nature run from which the synthetic observations were drawn.

The three rows show the most important dynamical features of thunderstorms: The maximum reflectivity $REFL\_MAX$ of the vertical atmospheric column gives an idea about the vertical content of hydrometeors; the temperature field T at $z = 150\,\mathrm{m}$ shows the surface cold pools that trigger new cells along their gust-front, the vertical velocity W at $z = 3500\,\mathrm{m}$ shows the actual up- and downdrafts of the convective systems.

Regarding only the column of R8 in these figures, one can see that the reflectivity patterns are reproduced well by 16:00 after 2 hours of cycled assimilation. This shows a positive influence of the observations on the rain-, graupel- and snowfield. Different processes are possible during the assimilation: Using observations of U, the horizontal divergence structure is determined and thus the positions of the updrafts. The updrafts themselves are collocated with high reflectivity, so reflectivity-observations enable the filter to choose members that have an updraft there. This updraft does not need to be fully evolved – a mere triggering can be enough to let it spin up during the forecast window, converging to the convective mode the observations were caused by. Also, direct analysis increments of hydrometeors caused by observations of reflectivity are possible.

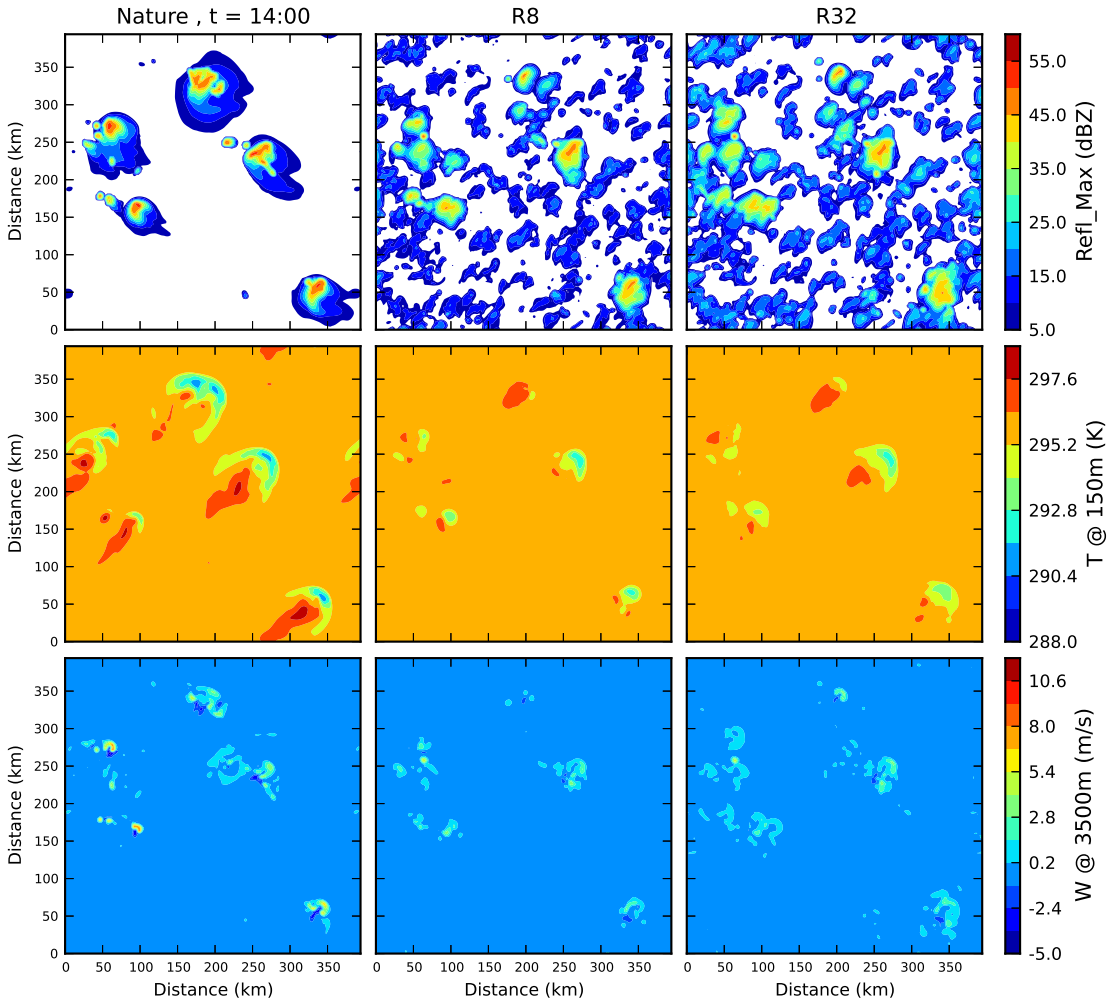

Figure 21: Nature Run and Analysis Means of R8, and R32 (both of random realization r01) at t=14:00. Top rows: Maximum reflectivity. Middle rows: Temperature T at $z = 150m$. Bottom rows: Vertical velocity W at $z = 3500m$.

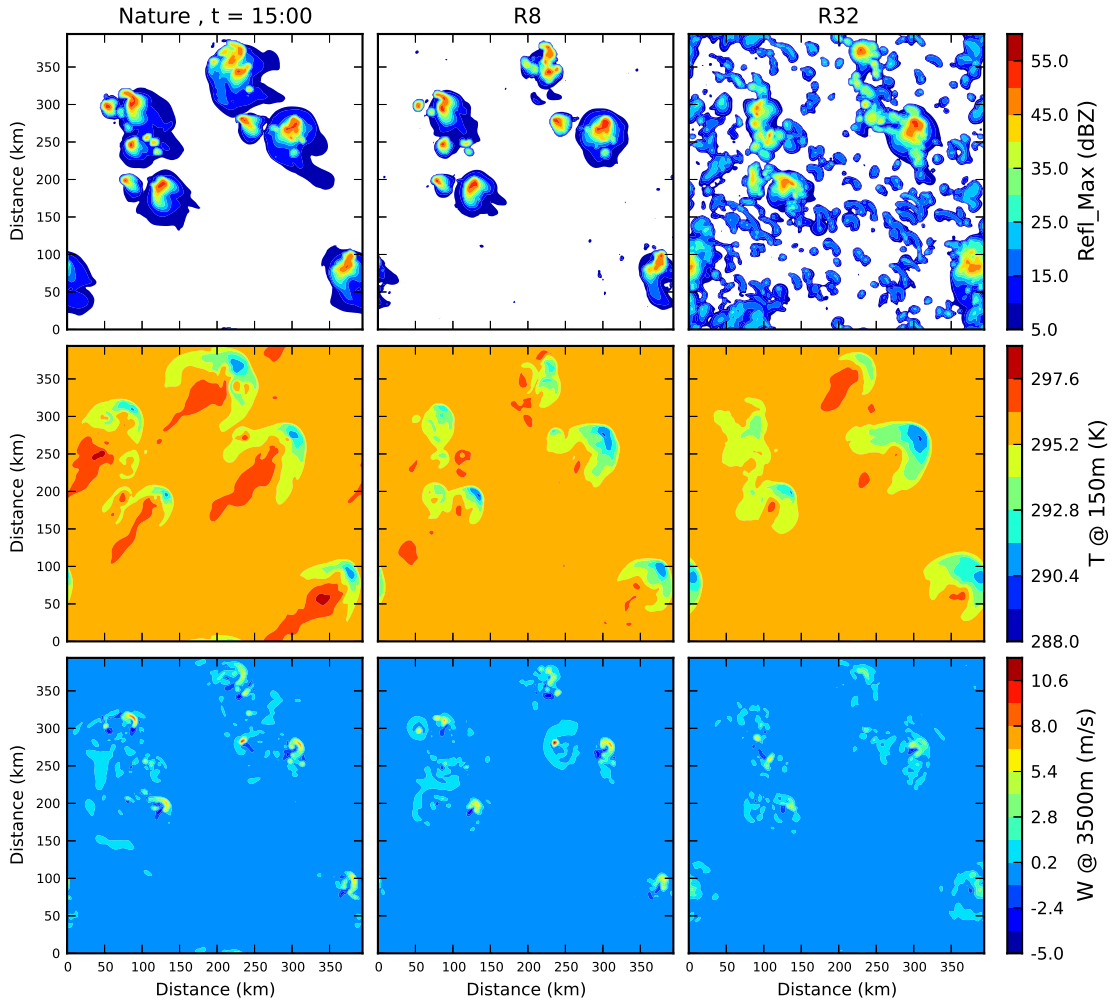

Figure 22: Like Figure 21. t=15:00

Sensitivity experiments without observations of reflectivity struggled with the reproduction of the actual precipitation field (not shown)[1]. The analyses of the dynamic fields like wind and temperature therein were also worse.

The vertical velocity field of the ensemble mean of R8 looks very similar to the nature run. Single updraft cells and downdraft regions are reproduced by the background error covariances between the observed U-wind and W, together with the covariances of the observed precipitation-fields and W. Whenever background error covariances are mentioned here, they refer to the implicit background error covariances that are a consequence of

1 But can be easily tested using the SOFIA-framework!

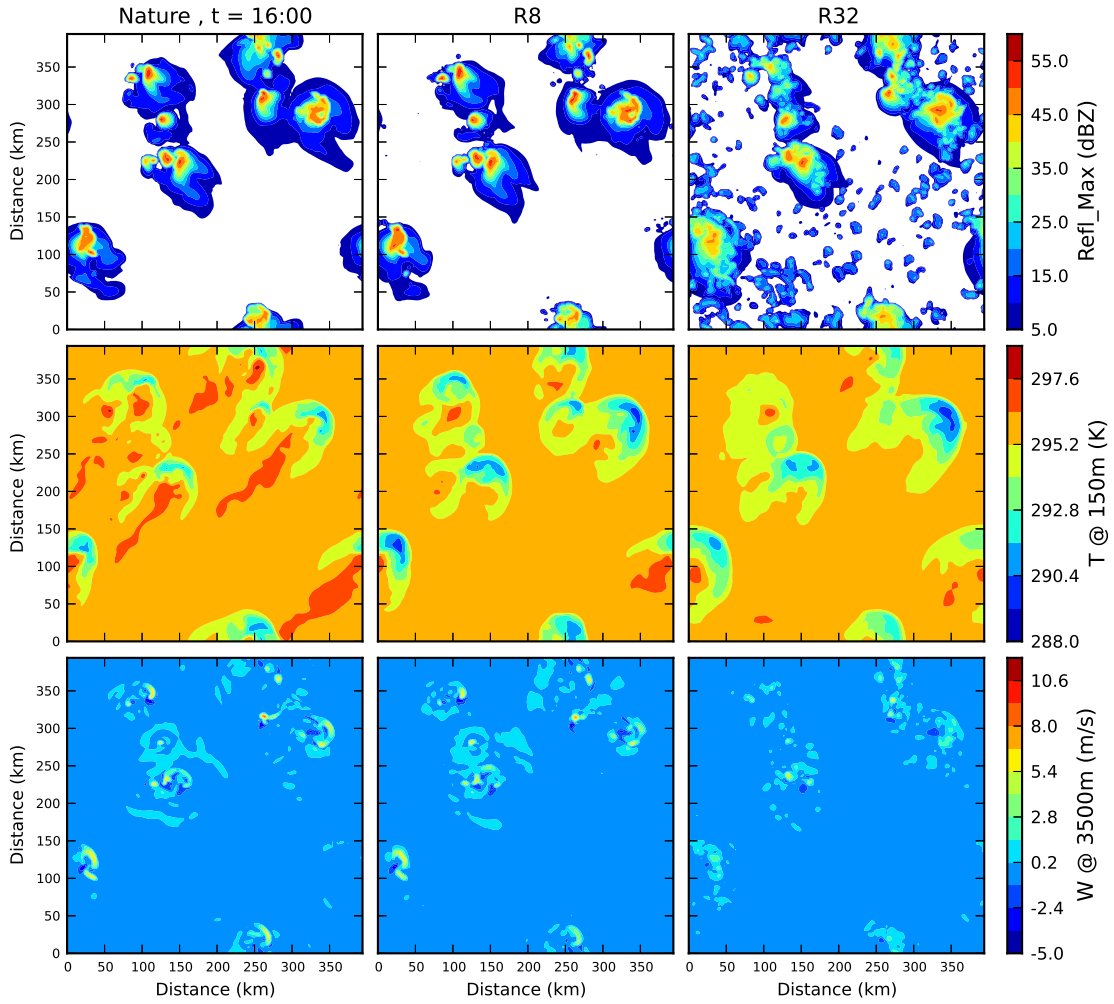

Figure 23: Like Figure 21. t=16:00

the choice of the linear members-combinations in the LETKF-analyses; the LETKF itself does not compute $\mathbf{P}^b$ explicitly.

The cold pools in the temperature field of the analysis resemble those of the nature run, even though the temperature field is completely unobserved. It is only reconstructed through covariances from other observed variables and through the dynamics: Storm cells initiated by the assimilation scheme produce precipitating downdrafts and thus surface cold pools.

This dynamical spin up of the cold pools can be seen from 14:00 to 17:00 for R8 in Figure 21, 22, 23, 24: In the beginning the cold pools appear too weak as only a few members provide good increments, while in the end they are too strong, hinting at an

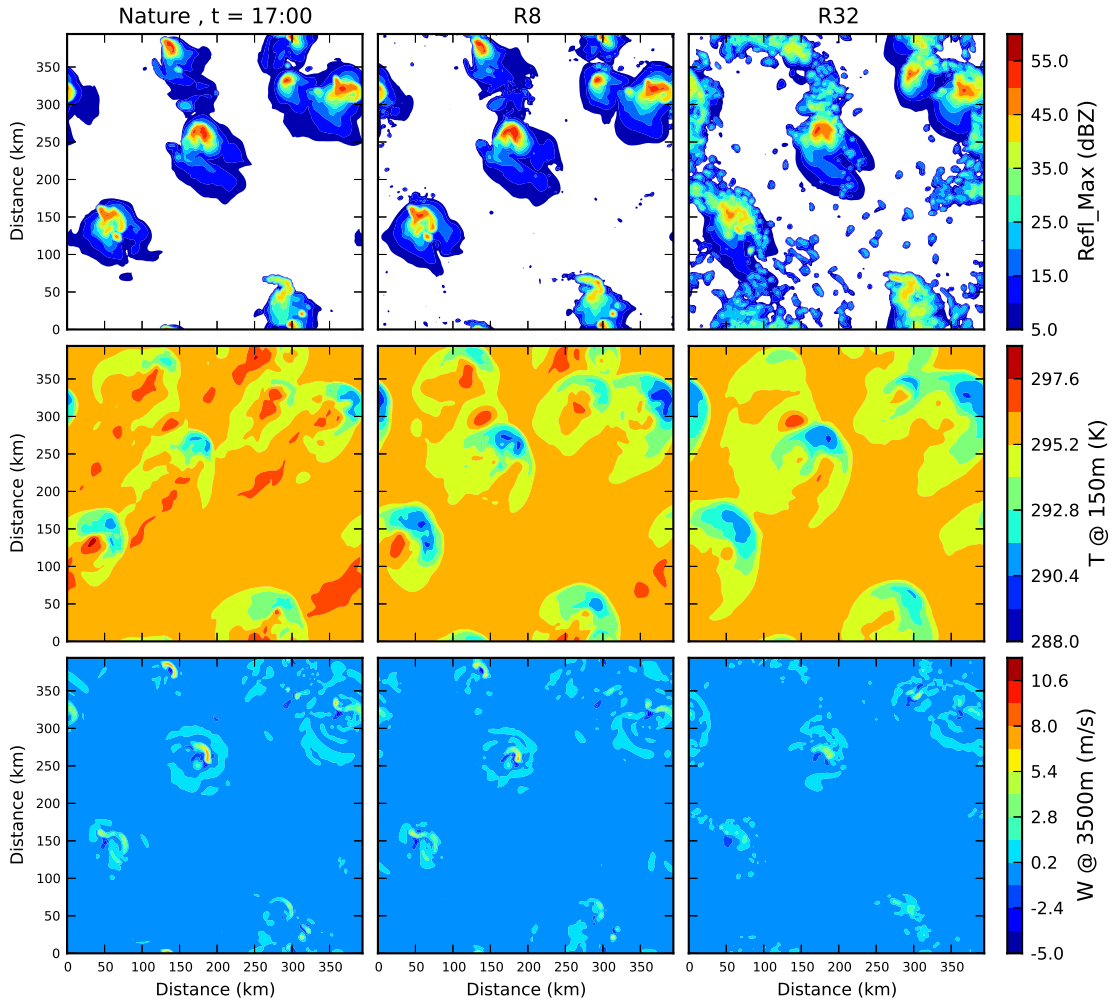

Figure 24: Like Figure 21. t=17:00

overcompensation through the dynamics of the initiated storms. This strong cold pool issue is addressed later in Section 3.4.1.

3.1.2 Quantitative examination of R8

Figure 25 shows the RMSE and spread of R8 and R32 for the wind variables U, V and W, for the temperature T, for the water vapor mixing ratio QV, for the hydrometeor mixing ratios of the precipitating variables rain, graupel and snow (QR, QG and QS) and the reflectivity $REFL$ in model space. Mixing ratios of cloud-water and cloud-ice are omitted here because the focus lies on the precipitating parts of the storms.

RMSE and spread of the free ensemble without any assimilation (and thus: without

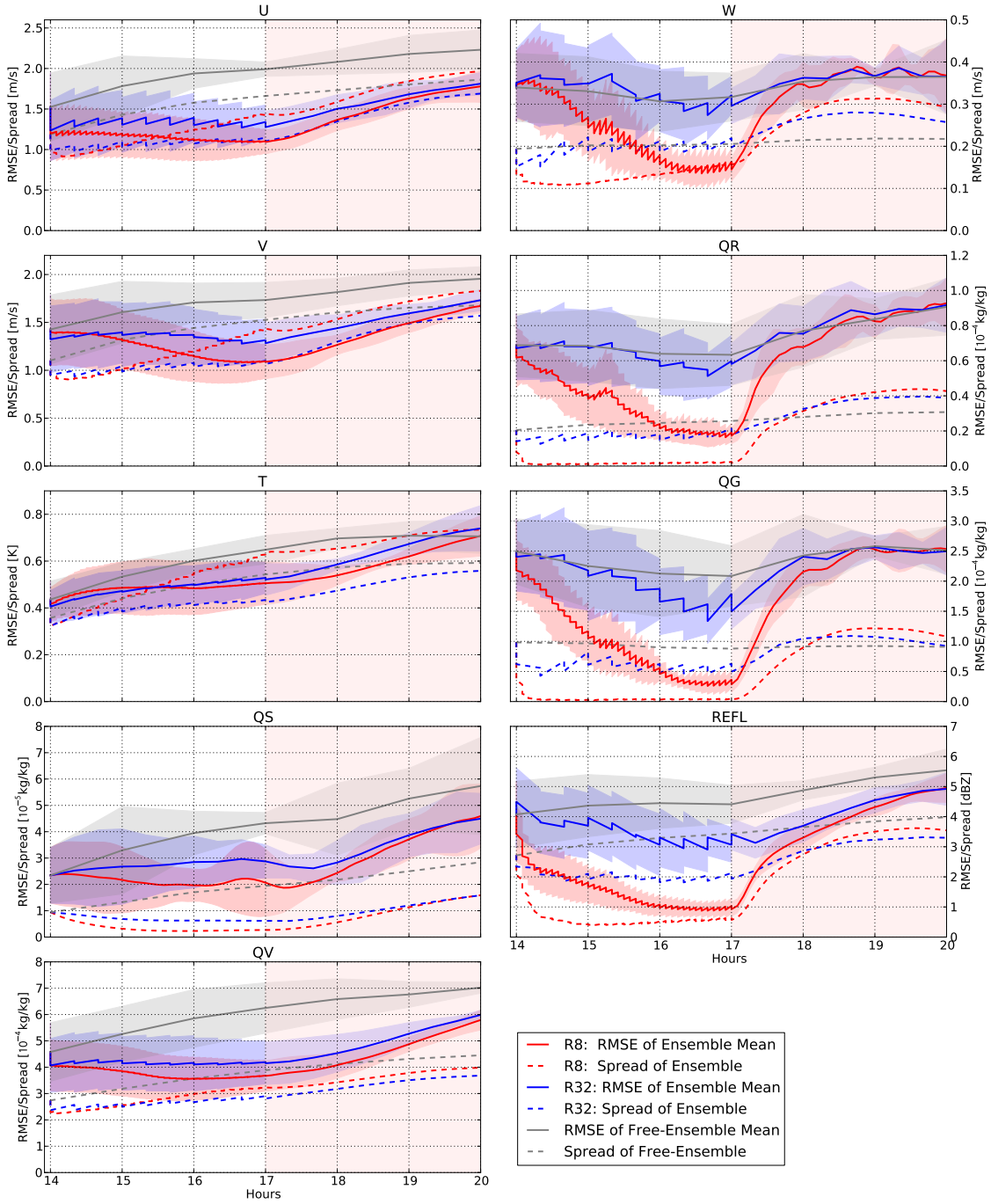

Figure 25: RMSE and spread of the ensemble means of R8 (red) and R32 (blue) during assimilation (14-17 h, forecast ensemble mean and analysis ensemble mean as one consecutive line) and ensemble forecast (17-20 h, only forecast ensemble mean). The gray lines show the RMSE and spread of the mean of the free ensemble without any assimilation. All gridpoints are evaluated. The error values are averaged over five realizations of the experiment. An envelope of $\pm\sigma$ (one empirical standard deviation) is depicted by the shaded areas.

correlation to the nature run) are depicted to provide a general error level. The error of R8/32 should be below this "free error level" by the means of the assimilation cycling to be regarded as a successful assimilation. When an ensemble of R8/32 becomes as uncorrelated to the nature run as the free ensemble, the free error level should again be reached. An anticorrelation of the R8/R32 analysis is indicated by an error above the free error level. RMSE and spread are evaluated at all gridpoints of the domain (Section 2.2.1) and then averaged. The experiment was repeated five times.

Regarding only the RMSE-curves of the ensemble mean of R8 and the spread of R8 (R32 is evaluated later), the filter seems to act beneficially on the ensemble:

The variable U is updated almost directly as U itself is an observed variable. The strongest reduction in the RMSE of R8 takes place in the first assimilation step. After that, the error of the ensemble mean is subsequently reduced relative to the free error level and also absolutely, showing the convergence of the scheme. At the same time, the spread of R8 grows due to the cycled covariance inflation, resulting in a consistency ratio > 1 at 17:00. The analysis increments are always beneficial, meaning they reduce the error of U of the ensemble mean in every analysis step.

Error and spread of the unobserved variable V behave similarly to U. This is expectable as the storms move in a 45 °-angle to both V and U, so information about storm-movement and divergence structures in the U-observations help the filter to choose members that are also good in the strongly correlated V-variable. Reflectivity-observations help both U and V of the ensemble mean to get closer to the nature run because they precisely indicate storm positions.

W here is only updated by the impact of observations of U and reflectivity. For the reference experiment R8, the filter quickly finds the right positions of the up- and downdrafts (Figure 24), resulting in an error level of half the free error at 17:00. This is regarded here as a good quality, concerning that the whole model field is evaluated. The consistency ratio of W reaches an almost perfect 1 by the end of the assimilation period.

QR and QG are mostly present in the updraft cores and downdrafts of the convective systems, thus closely coupled to W. For R8, the RMSE of both variables is reduced strongly but the spread becomes very small quickly and does not recover at all until 17:00. The consistency ratio is thus very low. This is a sign of the intended ensemble collapse in R8: Variability of rain and graupel are only present within the members' representation of the observed clouds while outside of them the ensemble agrees on the observed *no-reflectivity*.

Although QS is mapped onto the observations of $reflectivity$, it is excluded from the filter update, so the state of QS can only be improved passively by spin-up processes during the forecast interval. Sensitivity experiments revealed a detrimental influence of a QS-update by the filter (not shown), possibly due to erroneous increments interfering with spin-up processes of the members. The error of QS is below the free error level in the fine experiment R8 during the assimilation period, indicating a successful passive spin-up.

$REFL$ is a logarithmic mapping of QR, QG and QS. The RMSE behaves similarly to the error of QR and QG while exhibiting a larger spread value. This larger spread is

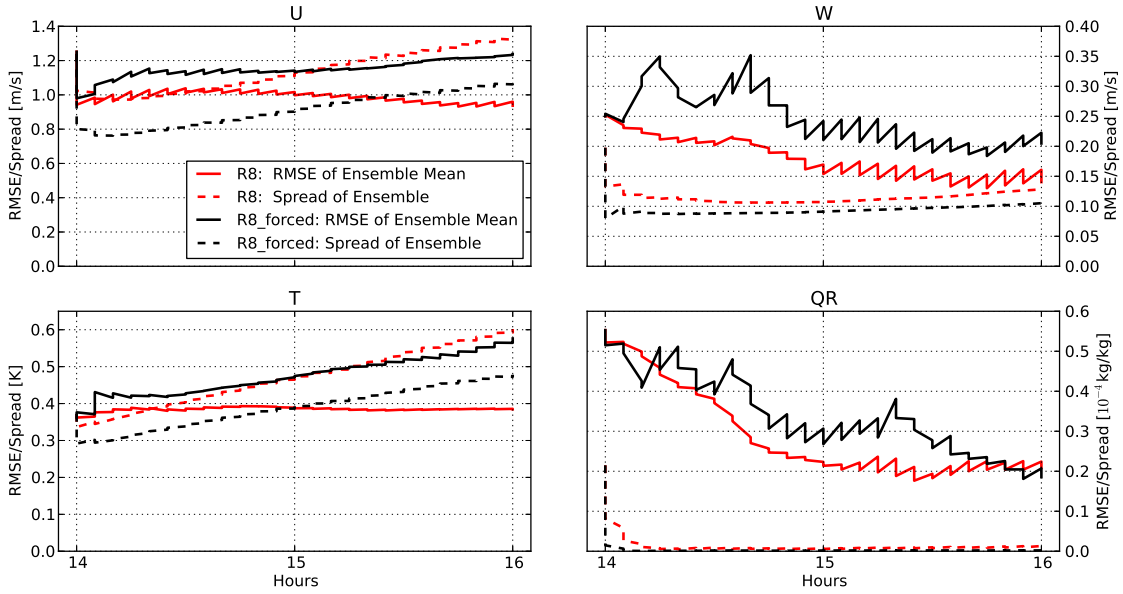

Figure 26: Like Figure 25, but comparing R8 (red) to R8_forced (black) for the primary variables U, W, T and QR in the first two hours of assimilation. (both experiments: only Realization 03, no repititions plotted here)

probably allowed due to the fact that the logarithmic mapping of the hydrometeors makes the distribution of $REFL$ more Gaussian (QR and $REFL$ in Figure 6).

T and QV are not mapped onto observations and therefore only updated by increments from members that fit the other observations well. With QV, the error in R8 stays below the free error level and is reduced also absolutely.

For T, the updates are slightly detrimental in the beginning of the assimilation window, resulting in a "backwards saw tooth" pattern where the error of an analysis state is higher than that of the forecast state it is based on. During the forecast intervals, the error is reduced away from the free error level but is pushed higher again by the analysis updates. This detrimental behaviour appears to hold only for the first hour – after that the analyses in R8 have almost no influence the mean T field. Thus, the error is not reduced by the analysis, but the spread is inflated. A nonoptimal consistency ratio larger than 1 is reached by 17:00 in R8.

R8 compared to R8_forced

R8 uses an **R**-matrix with inflated values for the errors of U- and (no-)reflectivity-observations (Table 3). In R8_forced, the diagonal **R**-matrix contains the original variance-values of the noise that was added to the synthetic observations before the assimilation (which was done in all experiments, simulating errors of measurement and representativity). Figure 26 compares the errors of the analysis means of R8 and R8_forced. Remember

that $\bar{\mathbf{x}}^a$ is the minimizer of the cost function and should thus be optimal. R8_forced shows higher error levels than R8, although the same observations were used. The analysis increments are often detrimental for U and T. The spread of R8_forced is below the error of its mean which might appear as a better consistency ratio than for the overspreaded R8, but this mainly expresses that the model's analysis states were forced very strongly towards the observations. Dynamical inconsistencies in R8_forced show strongly in the first assimilation hour for W. The error of QR is worse or the same as for R8, showing that the restrictive filtering in R8_forced is detrimental. R8_forced is thus not used in the rest of the study and is regarded only as a sensitivity-experiment[2].

Successful application of the LETKF

Together with the qualitative examination of R8 in Section 3.1.1, the behavior of this fine LETKF-scheme closely resembles the convective EnKF experiments of the previous studies by other authors (Section 1.2.4) and is regarded as a successful recreation. This shows that in convective data assimilation the LETKF-approach of combining perturbations of ensemble members linearly in separate local analyses with a localized $\mathbf{R}$ can keep up with the EnSRF-approach where $\mathbf{P}^b$ is localized.

3.2 Assimilation results for fine and coarse experiment

The fine experiment R8 is now compared to the coarse experiment R32 during the assimilation window in Section 3.2 and during the forecast window in Section 3.3, showing how they fulfill their "fine"/"coarse" purpose formulated in Section 2.1.4. First the different degree of ensemble collapse is evaluated and then it is shown how the advantages of R8 are limited with respect to R32 due to the limited predictability of the convection.

Ensemble collapse and non-collapse during the assimilation

Looking at the reflectivity patterns of R32 in Figure 21 and 22, 23, 24 of the assimilation window 14:00 to 17:00, the convective systems in the coarse experiment R32 appear to be roughly at the right horizontal position like in R8. However, the internal structure of the storms in the analysis ensemble mean of R32 is in less accordance with the nature run than the structures in R8. The amplitude of the storms in the analysis mean of R32 is smaller than in the nature run which is not the case in R8.

The reason for this is that the positions of the updrafts with high reflectivity have an actual spatial uncertainty within the ensemble members of R32 (Figure 28) so they appear blurred in the ensemble mean, while the updrafts are at the exact observed positions in

2 Field-plots of R8 and R8_forced look very similar and exhibit no visible difference by eye, thus they are not shown. This is different in the case of R32 compared to R32_forced as shown later.

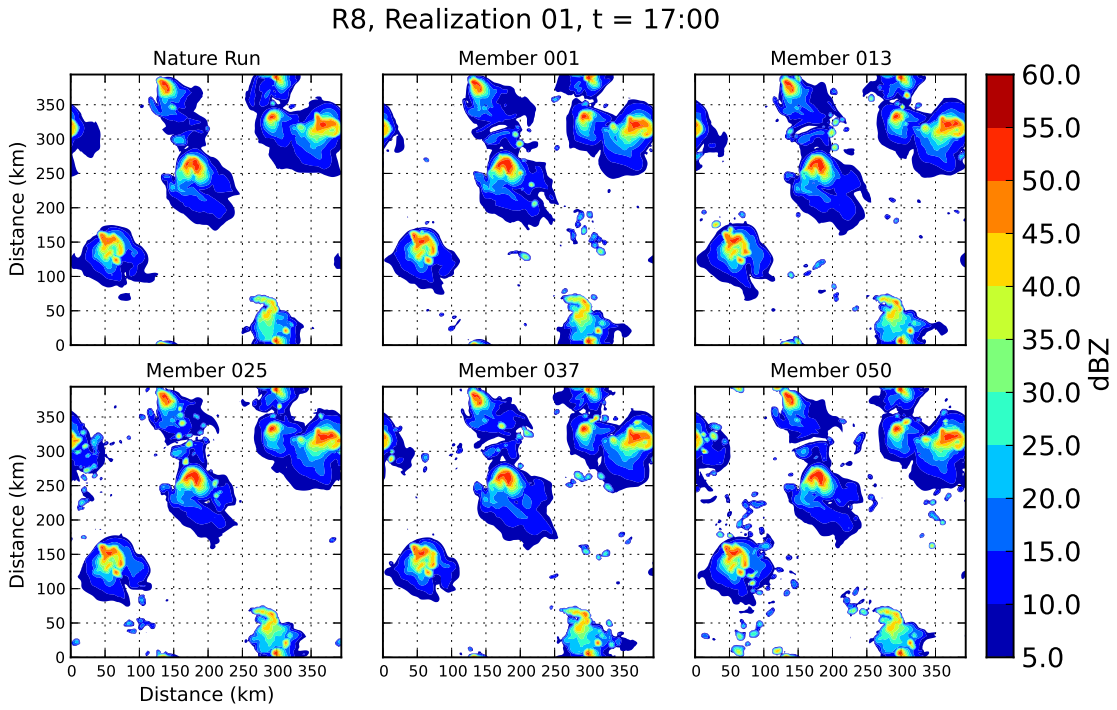

Figure 27: Maximum reflectivity ($REFL\_MAX$) of nature run and analysis ensemble members 1,13,25,37,50 of R8 (Realization 01) at the last assimilation time 17:00.

the ensemble members of R8 (Figure 27). This blurring of the R32-mean also takes place for the ensemble mean of the cold pools and the updrafts.

Contrary to R8, R32 has actual spurious storms in its members (e.g. Member 13 at $(x,y) = (200, 50)$ km in Figure 28) which also show up in the ensemble mean as a patchy carpet in regions without convection of the nature run (Figure 24). The collapsed ensemble of R8 has a very low amount of spurious convection which is also weaker.

For the last analysis time at 17:00, Figure 29 b) examplifies how the R8 ensemble has collapsed distinctly onto single observed storms where $REFL_{nature} = 40\,\text{dBZ}$. The reflectivity-distribution of R8, centered on the observed value, resembles a normal distribution which illustrates the strong Gaussian constraints of the filter (cf. Section 1.3.2). R32 has a wider distribution, bell shaped around a maximum at $\approx 35\,\text{dBZ}$ but also containing members without any precipitation. Here the Gaussian filter-constraint does not destroy the multimodality of the prior distribution.

At points with an updraft speed of $W_{nature} = 10\,\text{m/s}$, the analysis ensemble-distribution of R8 is again bell-shaped around that value with a slight negative bias, while R32 contains also weak parts of storms and downdrafts. For the gust front windspeed $U48_{nature} = 15\,\text{ms/s}$, R8 again looks Gaussian and R32 multimodal; in the coldpool-temperature $T48_{nature} = 292\,\text{K}$, R32 also shows a much broader spectrum than R8 due to the non-

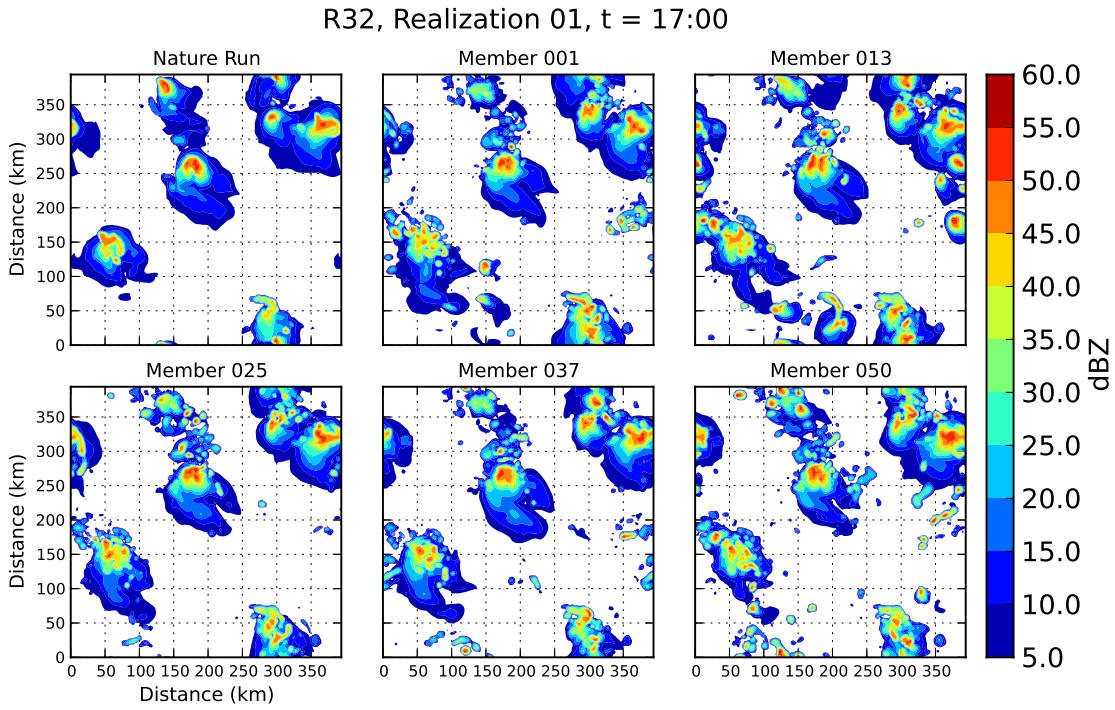

Figure 28: Like Figure 27, for R32

collapsed ensemble. A negative bias of the cold pool temperature is apparent for both R8/32 and will be addressed later.

In regions outside storms and cold pools, the value distribution of R32 presents more and stronger spurious updrafts and cells with high reflectivity than R8 (cf. Figure 30 a and b), together with slightly stronger spurious gust fronts. Judging spurious cold pools by the distribution of $T48$ in Figure 30 d) is difficult as there is no distinct temperature outside the pool; an "outside"-value of $T48_{nature} = 296\,\mathrm{K}$ was chosen subjectively, but is seldomly met due to the quite continuous $T48$-field (Figure 10). Similarly, $U48_{nature} = 1\,\mathrm{m/s}$ as an ambient wind speed does not give sharp results.

RMSE evaluation of the collapse

The collapse is recognizable by regarding RMSE of the ensemble means of R8/R32 and their spreads in Figure 25. For QR and QG, R8 shows a small error with a very small spread, resulting in a very bad consistency ratio of ≈ 0.1. R32 shows a larger error and a consistency ratio of ≈ 0.3 due to the non-collapse. Still, the RMSE of R32 in QR, QG and $REFL$ is below the free error level, supporting the qualitative assessment that the clouds in R32 are roughly correct.

For W, the RMSE of the ensemble mean R32 is reduced to a level just closely below the

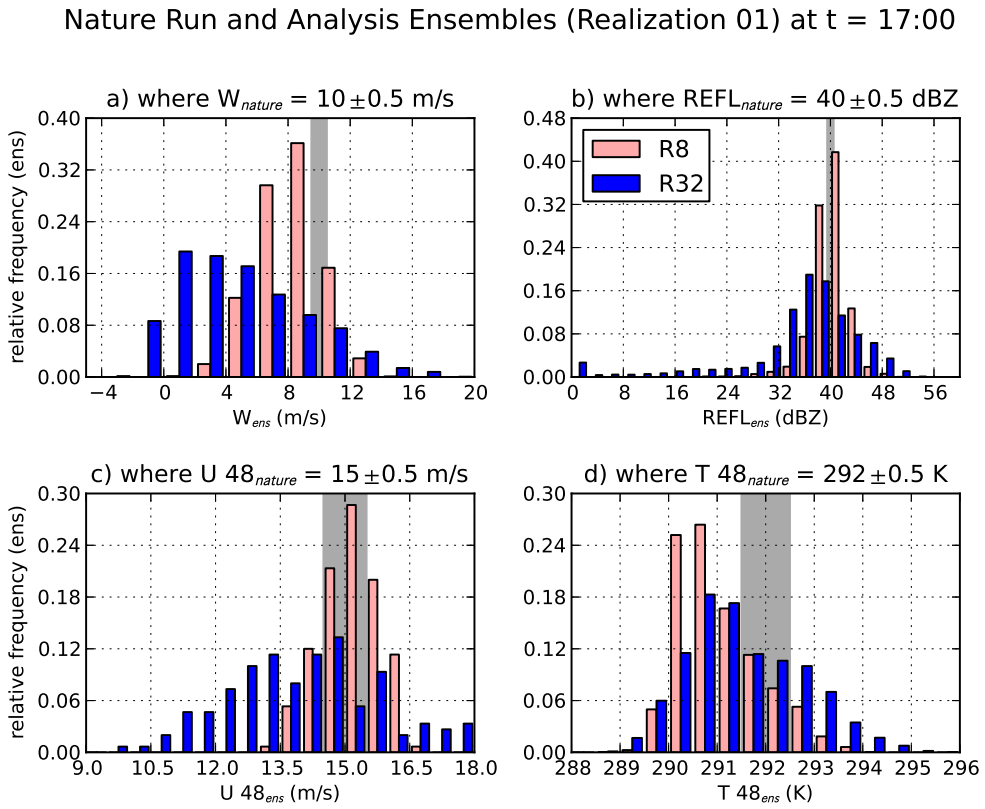

Figure 29: Relative frequencies of model values within the ensemble-members of R8 (pink bars) and R32 (blue bars). The bar width is 1/2 of the binwidth for the histograms. Regions inside storms are given by a) and b) and regions inside coldpools by c) and d): a) at locations of the domain where W of the nature run is $10 \pm 0.5\,\mathrm{m/s}$, b) where $REFL_{nature} = 40 \pm 0.5\,\mathrm{dBZ}$, c) where gust fronts with $U48_{nature} = 15 \pm 0.5\,\mathrm{m/s}$ are present and d) where $T48_{nature} = 292 \pm 0.5\,\mathrm{K}$ is a typical coldpool temperature. a) and b) are evaluated in the whole column, c) and d) are limited on the model level 48 ($z \approx 150\,\mathrm{m}$) for the cold pools. Regions closer than 64 km to the lateral boundaries are excluded so the histograms are computed only where the filter is fully active in R8/32 (Figure 17).

free error level by 17:00. An error above the free level would mean that updrafts in the mean are in the wrong places or even coince with downdraft. This seems to be the case for the first $1\frac{1}{2}$ hours of the R32-assimilation but equalizes until 17:00 where the blurrier field and the lower amplitude of the R32-W-mean result in a bad RMSE, although the updrafts are roughly in the right places (cf. Figure 24).

For the variables that are more horizontally continuous (U, V, T, QV), the mean of the non-collapsing ensemble of R32 gives RMSEs that are only slightly greater than the errors of R8. This advantage of R8 appears even less significant when the mean RMSEs of

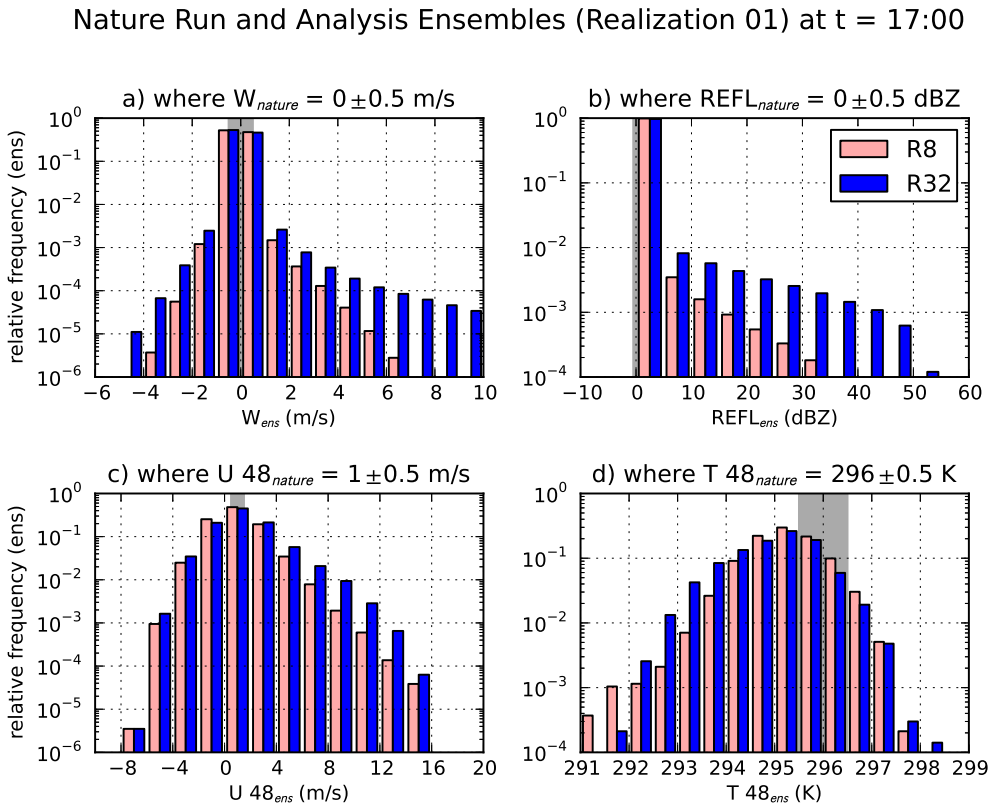

Figure 30: Like Figure 29 but for regions outside storms (a and b) and outside of cold-pools (c and d): a) at locations where W of the nature run is $0 \pm 0.5\,\mathrm{m/s}$, b) where $REFL_{nature} = 0 \pm 0.5\,\mathrm{dBZ}$, c) where the ambient wind speed of $U48_{nature} = 1 \pm 0.5\,\mathrm{m/s}$ is met and d) where $T48_{nature} = 296 \pm 0.5\,\mathrm{K}$ indicates the absence of a cold pool. A logarithmic scale was chosen here to show the differences for infrequent values.

the members are regarded in Figure 41: Here for U and V there is no difference in the assimilation quality between R8 and R32 – for T the R8-quality even appears worse (cf. Section 3.4.2).

Back to the RMSE of the mean states: Spread and consistency of U, V, T, QV show a similar behavior in R8 and R32 during the assimilation window. These horizontally continuous variables suffer less than the intermittent variables from the coarsened observations of R32 because the prior correlation lengths of the model fields U, V, T, QV appear to be larger than those of W, QR, QG. For example, the horizontal divergence patterns of U and V around single updrafts (Figure 13 and 14) are much smoother than the updrafts cores they are caused by.

Although R32 has less detailed observations, the larger-scale weighting of the analysis

members introduces more coherent storm structures that still are able to represent the nature run by the observations. This concurs with Sobash and Stensrud (2012) who observed a growth in horizontal correlation lengths during the growth and mesoscale organization process of their idealized convective systems.

Collapse of the horizontal precipitation field

Another measure for the collapse of the R8-ensemble are the SAL-scores of the horizontal *REFL_MAX* precipitation-field (Figure 31). For the location error SAL-L at 17:00, the mean score of the R8-ensemble members (red dashed line) and the score of the R8-ensemble mean (red solid line) are both lower than for R32. The more precise experiment is thus the fine R8.

The number of precipitating objects *NObjects* differs more strongly between members and mean in R8 than in R32: In R8 the spurious storms are weak and infrequent (Figure 27) so they only show up in the member-fields of *REFL_MAX* as tiny countable spots, but not in the thresholded mean-field (Figure 24). In R32 the stronger spurious storms (Figure 28) show up in the mean (Figure 24) and contribute to the mean *NObjects*-score of the R32-members – which thus does not differ strongly from the *NObjects*-score of the R32-mean.

In the structure error of the precipitation objects SAL-S, both R8/32 reach a level close to zero by 17:00, so in both experiments the assimilated convective modes mapped by *REFL_MAX* resemble those of the observed nature run, despite the collapse or non-collapse.

The overall amount of precipitation, given by SAL-A, is higher in R32 due to the spurious storms. But, in R32 there is no initial underestimation of the overall amount of precipitation as in R8 between 14:30 and 15:30. This phenomenon in R8 is caused by the incremental introduction of undeveloped storms at observed storm locations, together with the suppression of spurious storms at locations where no-reflectivity is observed.

It must be noted that in the object-space of the SAL-scored reflectivity, the filter seems to act largely detrimental for R32 if only SAL-A and *NObjects* would be regarded; here the R32-increments increase the error level with every analysis. This is due to the design of the non-collapsing R32: On the one hand, the filter adds storms at the observed locations, distributing the *positive* rain-amount of good members into other members by the analysis weights. At the same time, spurious convection can not be eliminated by adding *negative* rain-amounts – the filter can only choose members with observed zero-reflectivity as a compensation in the local analysis, resulting in an overestimation of rain in the uncollapsed R32. This disadvantage of a convective LETKF-assimilation with a non-collapsing ensemble was predicted by Craig and Würsch (2012). Using a simple stochastic cloud model in an LETKF testbed, their analysis mean caught the observed storms but was tainted by spurious convection in the uncollapsed background ensemble.

The curves of SAL-S and SAL-L, which are independent of this overestimation-bias, show that the analysis incremenents usually reduce the error for both R8/32, so the cloud-

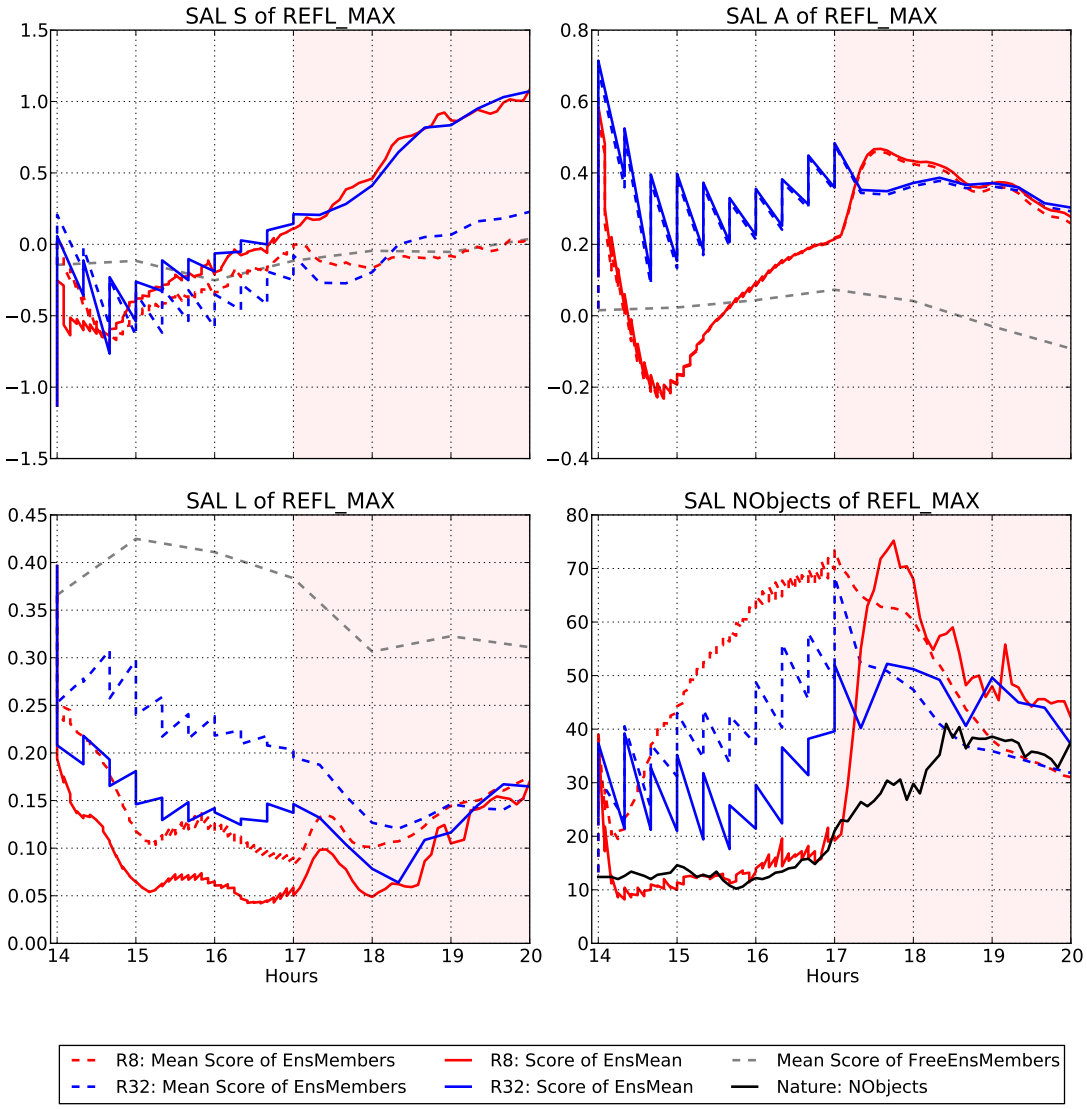

Figure 31: SAL score components S, A and L of the precipitation field *REFL_MAX*> 10 dBZ, together with SAL-*NObjects*.

internal assimilation is regarded as successful for both the collapsed and the uncollapsed regime. This is not the case for the sensitivity-experiment R32_forced, now shown briefly:

Dynamical advantage of an uncollapsed ensemble: R32 vs R32_forced

The experiment R32_forced, wherein the superobservations have a higher contribution to the filter-solution by a $\mathbf{R}^{-1}$-matrix with greater magnitude than in R32 (cf. Section 2.1.2), was devised in order to show what a collapsed coarse analysis would look like. Figure 32 shows an analysis mean state of R32_forced against R32 and the nature run. The W-field of R32_forced reveals stronger single updrafts than R32 which is only possible if all members of R32_forced agree on these particular updrafts. At the same time, the W field of R32_forced is noisier than in the nature run and tainted by gravity waves; strong spurious downdrafts lie next to strong updrafts, both in small patches. Comparing this to the W-field of the nature run, it is obvious that the member model states in R32_forced are dynamically inconsistent, whereas the members of R32 are good by that category. R32_forced is thus not used in the rest of the study and is regarded only as a sensitivity-experiment.

Summarizing the evaluations during the assimilation period, R8/R32 appear to fulfill their purpose of a "fine"/"coarse" representation of the observed storms in the analysis ensembles: With R8, the analysis storms are very close to the observed storms in terms of intensity, shape and horizontal position. In R32, these attributes of the analysis-storms vary around the observed storms.

The next section compares ensemble forecasts from the fine and coarse analysis to see how or if the more precise fine analysis is therein advantageous.

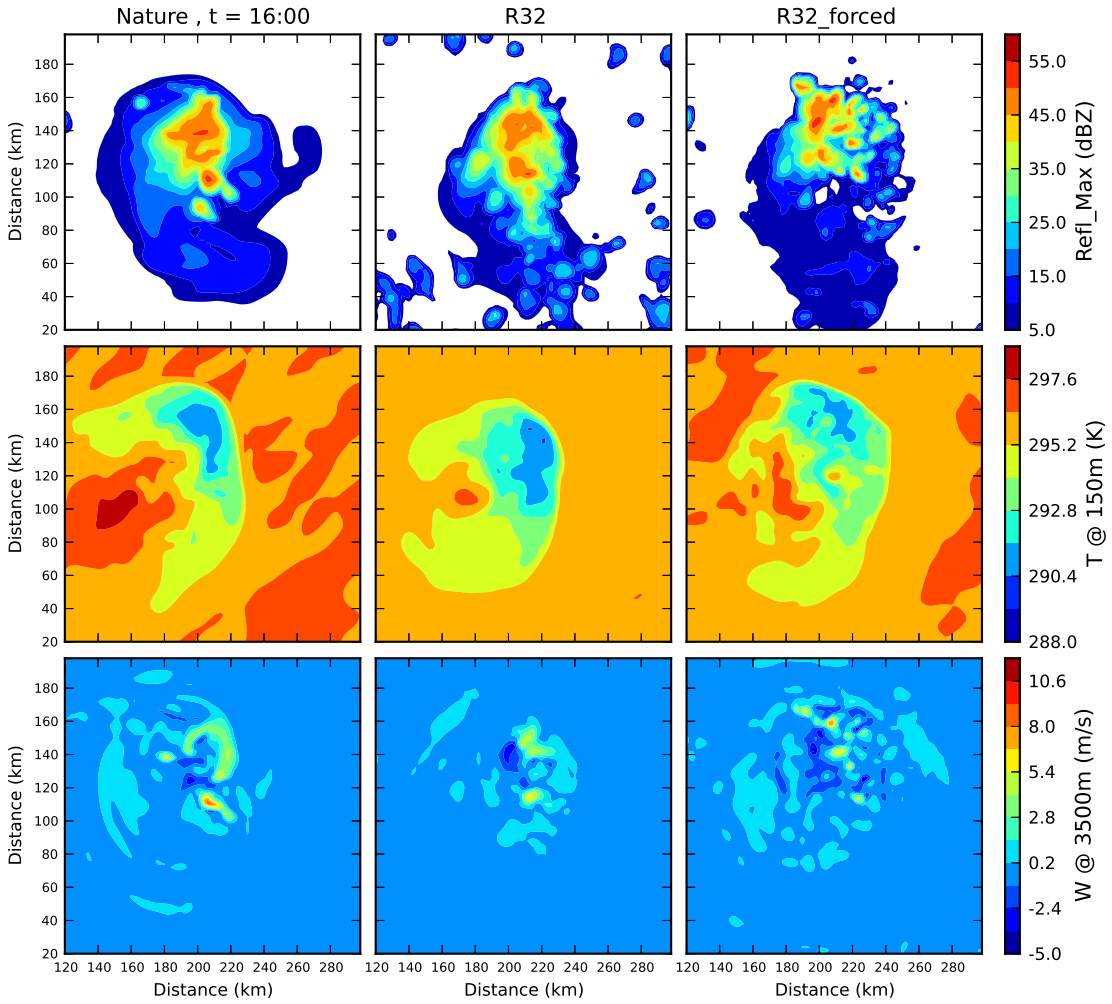

Figure 32: Like Figure 21, but comparing the analysis means of R32 and R32_forced to the nature run at 16:00 during the assimilation (Realization 03).

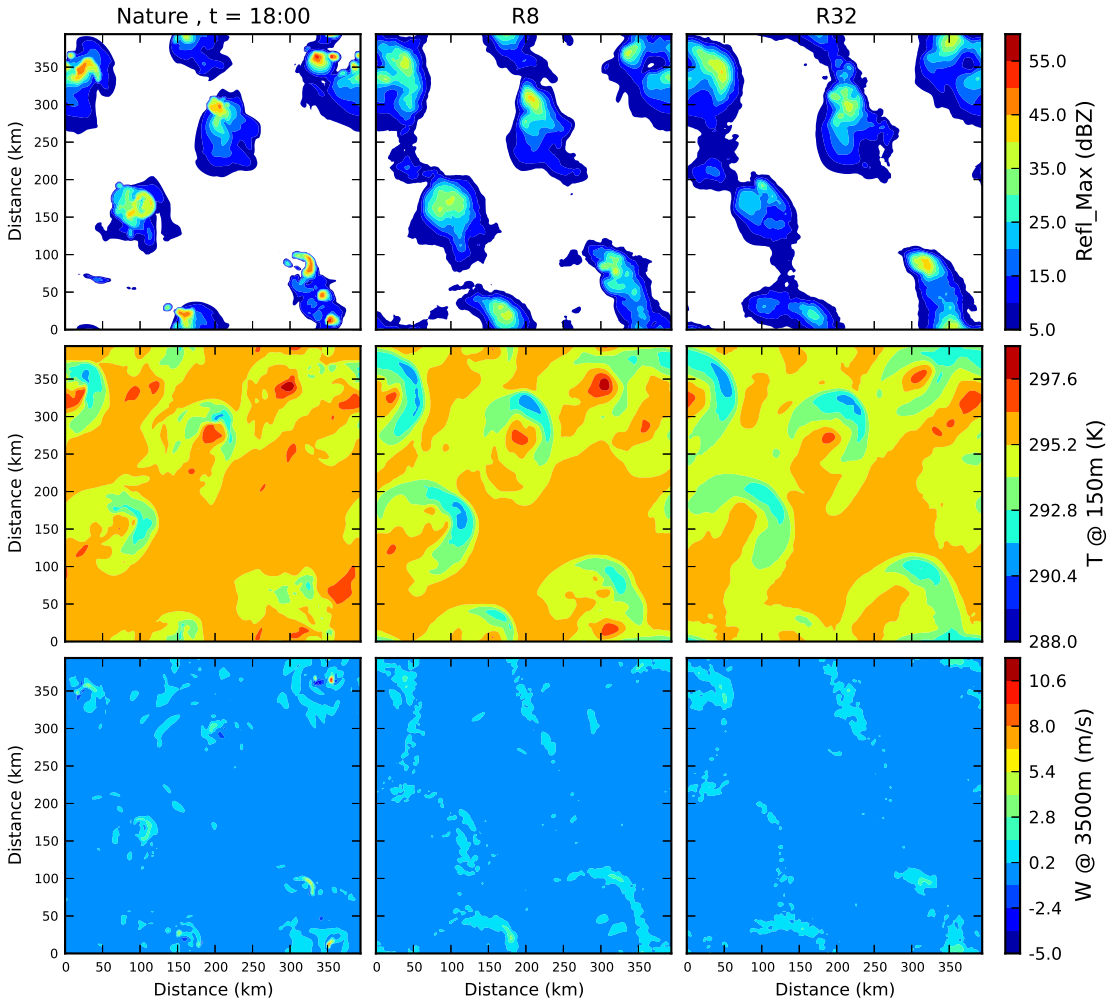

Figure 33: Like Figure 21, but forecast ensemble means (Realization 01) at t=18:00.

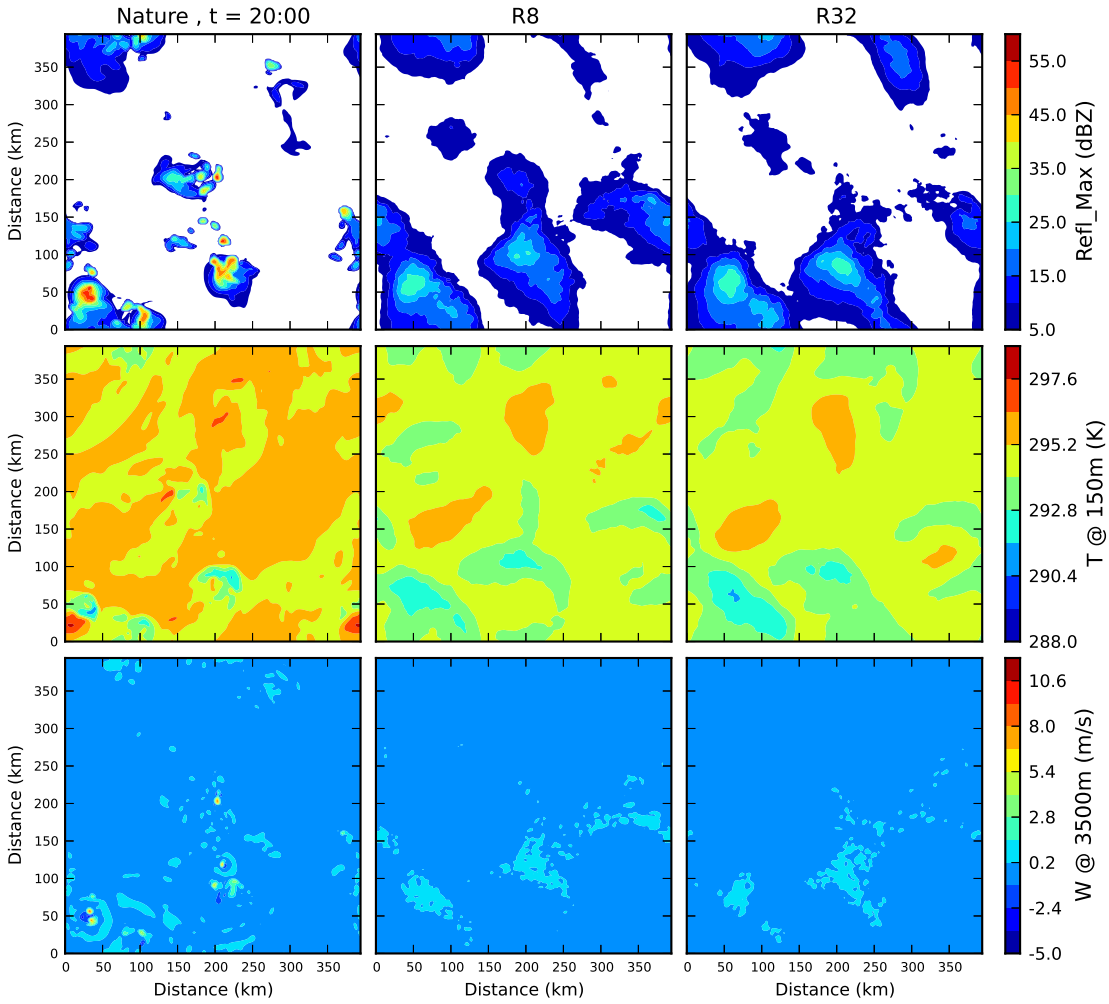

Figure 34: Like Figure 33. t=20:00.

3.3 Ensemble forecasts from fine and coarse experiment

It is evaluated now how the quality of an ensemble forecast from the fine analysis ensemble of R8 differs from the coarse analysis ensemble of R32. Figure 33 and 34 exemplarily show the forecast ensemble mean fields of R8 and R32 from 18:00 to 20:00. A blurring of the forecast mean fields due to diverging storms in the ensemble is apparent. It is hard to judge by eye which one of R8/32 gives a more accurate forecast, so objective methods have to be applied.

Object-based forecast evaluation

A good precipitation forecast should contain storms with the right intensity at the right locations. The intensity and thus the convective modes of the forecast storms can be assessed by combining SAL-A and SAL-S (cf. Figure 31).

At the end of the assimilation window, the ensembles of both R8/32 contain storms with the right structure and convective modes, shown by Figure 27 and 28 and by the mean SAL-S score of the R8/32 ensemble members which is almost zero at 17:00. During the forecast window, this property of SAL-S ≈ 0 is almost completely preserved, so one can assume that the analysis storms were dynamically consistent with the convective modes of the model. At the last forecast time 20:00, the ensembles of both R8/32 contain storms with the apparently right structures (Figure 35 and 36) – in R32 these are slightly blurrier, but not much.

The positive bias of SAL-A > 0 combined with SAL-S ≈ 0 could suggest that a) the storms have the right structure but there are too many of them or b) the forecast storms are larger with intenser cores, normalizing SAL-S to the same level. The average $NObjects$ of the R8 and R32 ensemble members decreases towards $NObjects_{nature}$ during the forecast-window, so b) is more likely: Larger forecast storms with intenser cores, given by more updraft-cores for larger anvil regions in the forecast members than in the nature run. This positive bias is regarded as a minor defect of the forecast system, and as it affects both R8 and R32, the comparison of both does not suffer.

At 17:00, R32 has a larger SAL-A than R8, but within 30 minutes of forecast time, the amount of precipitation in R8 grows quickly to the level of R32 and overshoots it, while the SAL-A of R32 relaxes a bit in the direction of zero. This suggests that the weak spurious cells over 10 dBZ in the R8-members (depicted by the dashed red line in $NObjects$; they do not show in the mean of the 10 dBZ-thresholded field which is the solid red line) are intensifying quickly – cells which might have been partly suppressed by the filter until 17:00 and which can then evolve to full strength in the forecast window. The already strong spurious cells in R32 at 17:00 seem to rather die off than to intensify, indicated by a decreasing SAL-A and $NObjects$ of the R32-members (blue dashed line) in the first forecast hour.

Now that the question of intensity has been explored, the important location property is checked. Here the mean error level of the free ensemble members (dashed gray line)

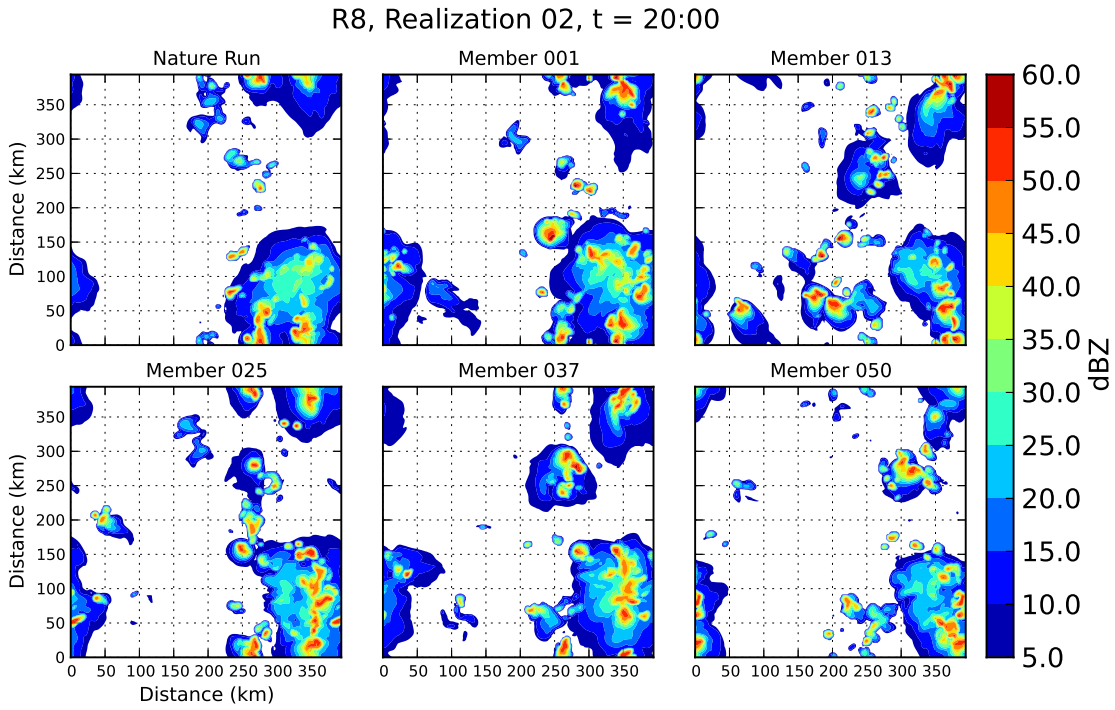

Figure 35: Maximum reflectivity (*REFL_MAX*) of nature run and forecast ensemble members 1,13,25,37,50 of R8 (realization 02) at the last forecast time 20:00

gives a good orientation of the relative goodness of the forecast locations. Mean- and memberlines of R8/32 converge within 2 hours; after that time, the SAL-L error of R8 and R32 is indistinguishable. This timespan can be regarded as the predictability limit of forecast storm location from the fine analysis, for this particular convective sounding. After 3 hours of forecast, SAL-L scores both R8 and R32 appear to be still quite skillful.

The location error seems to be caused mainly by the forecast storms moving apart from the filter-determined locations. SAL-S of the ensemble means of R8/32 grows positive during the forecast, indicating the blurring of the mean precipitation field (cf. Figure 33 and 34). The positive blurring gradient of the field-mean SAL-S (solid red and blue line) looks similar for both experiments. It could be expected that the gradient is steeper for R8 in the first forecast hour due to faster error growth, but SAL-S is apparently not precise enough to judge this. The RMSE-statistics of the forecasts in the next section are better for that purpose.

As an addition to the rather coarse comparison of average mass points in SAL-L, the DIS-component of the DAS-score compares individually matched precipitation objects between ensemble members and nature run and assignes an error > 1 to their displacement (Figure 37). With the chosen maximal search radius for the object-matching of 45 km, the DIS-error of R8 converges to the level of R32 already within 1 hour. This is shorter than

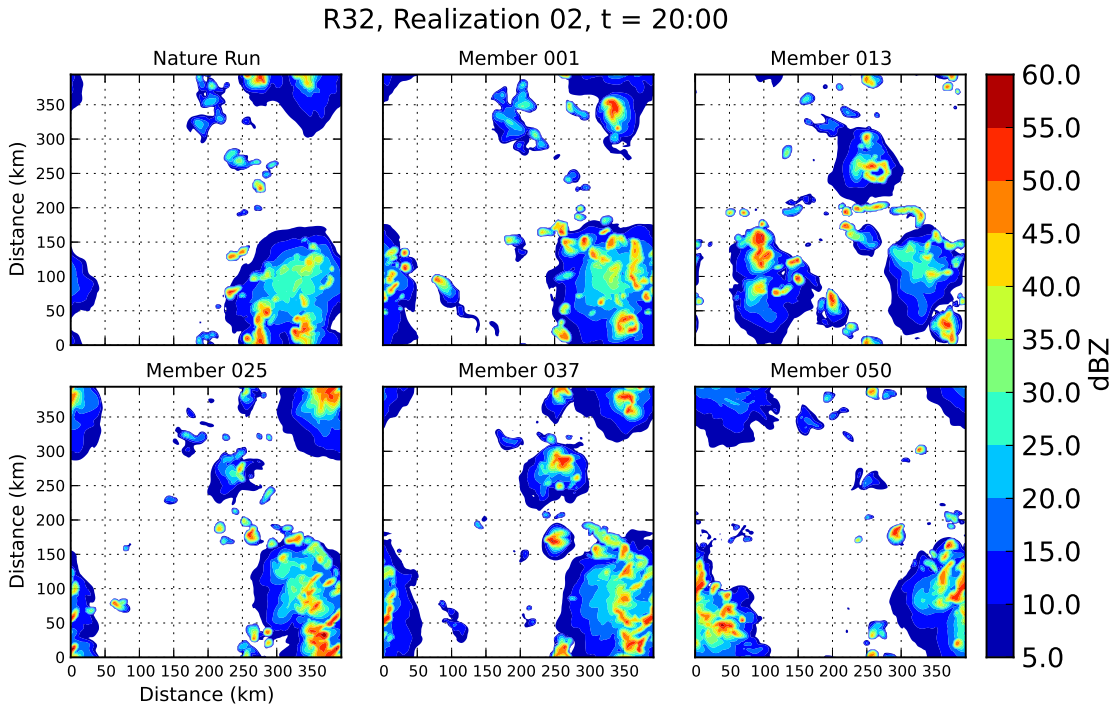

Figure 36: Like Figure 35, for R32

the 2 hours of convergence time in the non-matching SAL-L because DIS is sharper. The ability of SAL-L to distinguish R8 and R32 longer speaks for a greater "fairness" of SAL-L in such a regime of stochastic convection. Nonetheless, DIS endorses the statement of SAL-L and is helpful due to its sharper distinction between the precise R8 and the coarse R32 at 17:00.

RMSE and spread-based forecast evaluation

The fast growth of fine perturbations is evident in the RMSE and spread of the horizontally quite intermittent vertical wind speed W (Figure 25), evaluating whether the positions of the assimilated up- and downdrafts concur with the nature run.

For R8, in the first hour of forecast after 17:00 the RMSE of W grows to the level of R32, as does the spread which even overshoots the spread of R32. The even more intermittent precipitation variables QR and QG show this behaviour even more drastic: A very steep error growth is accompanied by the quick recovery of the almost non-existent spread of 17:00. With $REFL$, this behaviour is also present but slightly mollified, possibly due to the inclusion of the less intermittent snow QS that composes the horizontally extendent anvils of the convective systems.

Until the end of the forecast period, the error of these horizontally intermittent variables

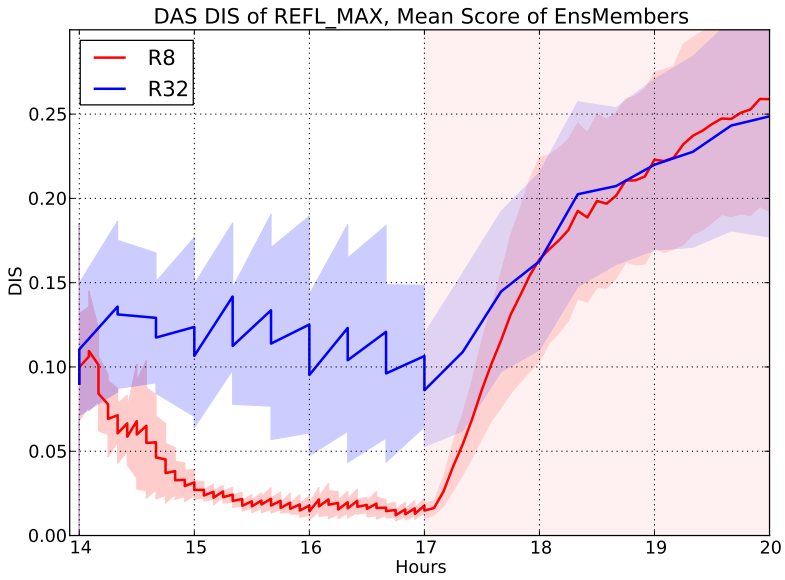

Figure 37: DIS-component of the DAS-score of *REFL_MAX*. Displayed is the mean DIS-score of the ensemble members of R8 (red), R32 (blue). The error lines are the average over five experimental repetitions. An envelope of the empirical $\pm\sigma_{DIS}$ is given by the shaded areas around the curves. Scoring the free ensemble members with DIS is not sensible (cf. Section 2.2.2) so their error is not displayed.

relaxes towards the error level of the free uncorrelated ensemble for both R8/32, the relaxation being faster again for R8. For both R8/32, the ensemble means of the forecast fields W, QR, QG at 20:00 do not agree more with the nature run than an uncorrelated ensemble; the $REFL$-forecasts seem to be better than uncorrelated, again due to the snow-component, but indistinguishable between R8 and R32.

Contrary to these horizontally intermittent variables, the mean state of R8 seems to keep a slight advantage throughout the forecast window for the horizontally continuous variables U, V, T and QV. Here the growth of the ensemble-mean error and spread is similar between R8 and R32. This advantage of R8 appears less significant when the mean errors of the members are regarded in Figure 41: Here the forecast errors of U, V, T do not seem to differ significantly between R8 and R32.

As shown, the horizontally intermittent variables are subject to a more non-linear error growth than the horizontally continuous variables. This nonlinearity is to be expected in a chaotic system (Section 1.1.3) of non-linear differential equations that the forecast model consists of.

Brier Score and ROC-area

Probabilistic three-hour forecasts of binary storm events in R8 and R32 compared to the observed field in Figure 38. The qualitative shapes and observation-accordances of the probability-fields are very similar between R8 and R32; the stronger positive bias of R32 is apparent. The evolution of Brier Score and ROC-area for the forecast period is depicted in Figure 39. A similar convergence of quality as in SAL-L and DIS can be identified for both probabilistic scores, using a low treshold of 10 dBZ and a rigorous threshold of 30 dBZ – neither the fine nor the coarse experiment show significant advantage over the other after 3 hours of forecast. This supports the quality-assessments that were already stated.

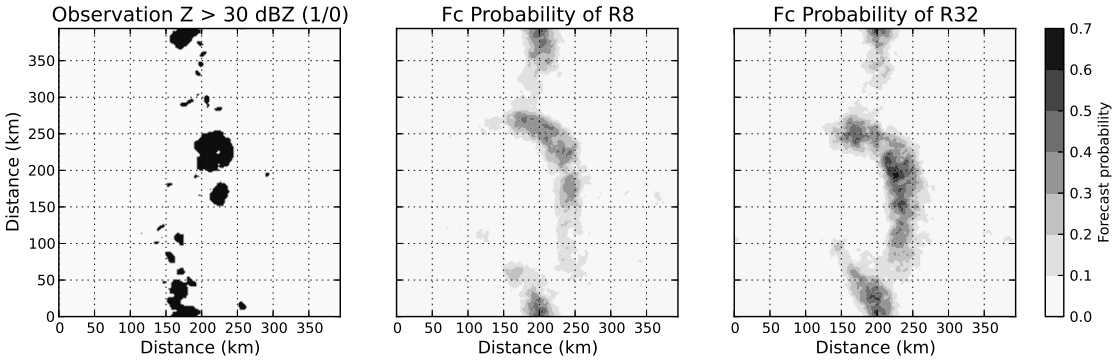

Figure 38: Observation field of $REFL\_MAX > 30\,\mathrm{dBZ}$ (left), forecast probabilities of R8 (middle) and R32 (right) at 20:00 after 3 hours of forecast. The observation is binary (black: observed); the colorbar from $p = 0$ to 0.7 refers only to the forecast probability fields. (Realization 04)

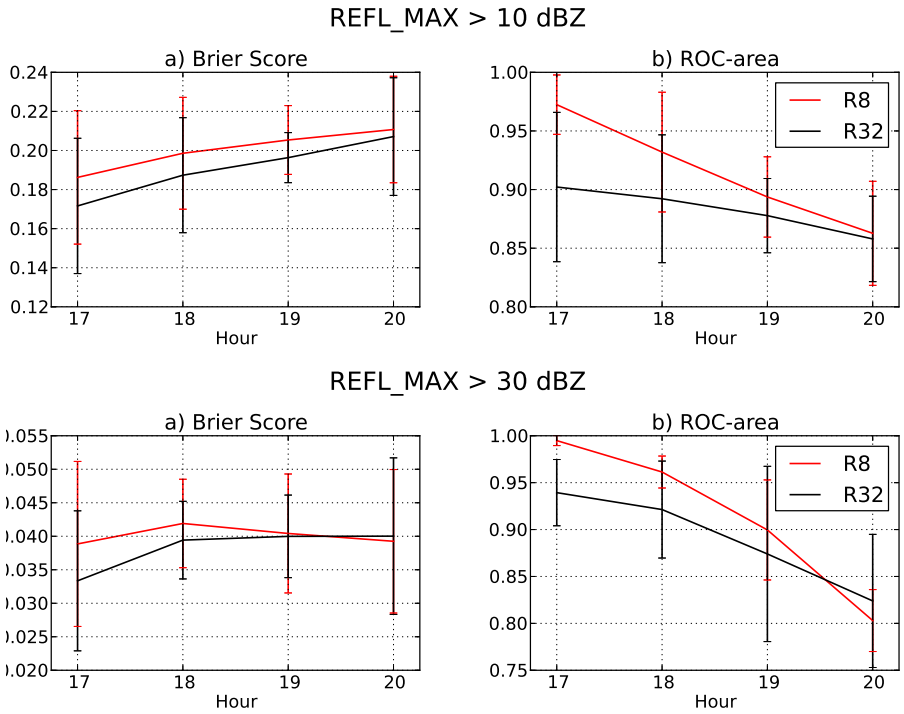

Figure 39: Brier Score and ROC-area for R8 and R32 during the forecast window 17:00 to 20:00, using thresholds of 10 and 30 dBZ for the reflectivity field. Errorbars with the empirical standard deviation of five experimental repetitions are added.

3.4 Cold pool bias and inconsistent temperature assimilation

This short section is dedicated to the assimilation quality of the cold pools and how the temperature field suffers from the indirect assimilation using radar data.

3.4.1 Cold pool bias

Surface cold pool objects are defined as those horizontal patches in model level 48 ($z \approx 150\,\mathrm{m}$) that are colder than the mean temperature of the nature run in this vertical layer. The negative deviation (usually between $+0.5$ and $+5\,\mathrm{K}$) is then inverted into a positive value and all other values around it are set to zero. The result is a patchy field similar and almost collocated to the patches of *REFL_MAX* but only dependent on the temperature field.

Figure 40 shows the development of the cold pools: In the beginning of the assimilation window they are too weak (SAL-A < 0) and too pointed (SAL-S < 0), then grow too strong and wide by the last assimilation step. This is more apparent for R32 than for R8 due to actual spurious cold pools; the non-collapse shows in the largely different SAL-L between R32-members and R32-mean. The overgrowth continues throughout the forecast window, concurring with the findings about the positive precipitation bias in Section 3.2.

3.4.2 Inconsistent temperature assimilation

Apart from the locally limited cold pools, the general properties of the temperature field during the assimilation are interesting. Figure 41 shows the mean-RMSE of the R8/32 ensemble-members (instead of the RMSE of the mean in Figure 25). For most variables the curves look similar, but for T the mean-errors of the R32 members appear to be smaller than for R8, during the assimilation and the forecast.

The analysis increments of T in R32 are always beneficial, contrary to R8 where the analyses subsequently worsen the T-errors of the R8-ensemble. This disparity is possibly caused by the fact that the analysis states of T are provided by the best-fitting members for wind and reflectivity observations; T itself is unobserved. $R_{Loc} = 8\,\mathrm{km}$ appears to be inferior to $R_{Loc} = 32\,\mathrm{km}$ because in the latter case, more complete and ample temperature structures are used (R32 in Figure 1) to construct the analysis members which can then be dynamically consistent. The greater T-error of the R8-members is then a sign that the common “craze for the best mean” in the application of convective EnKF can have negative influence on the dynamics of the members that construct the mean.

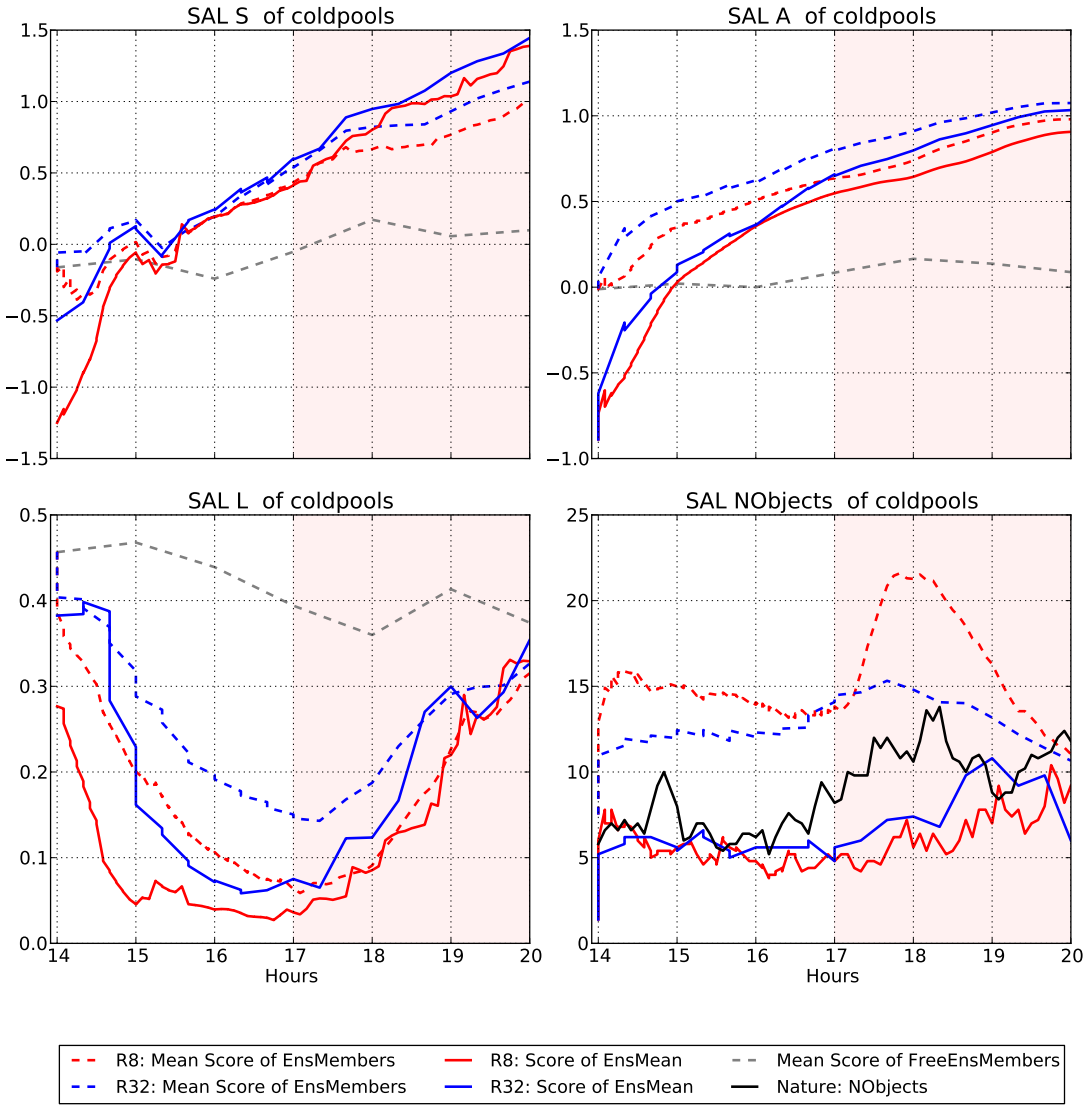

Figure 40: Like Figure 31, but here: SAL score components of the cold pools. A surface cold pool is defined in the lowest model layer where $T48_{nature} < \bar{T}48_{Nature}$.

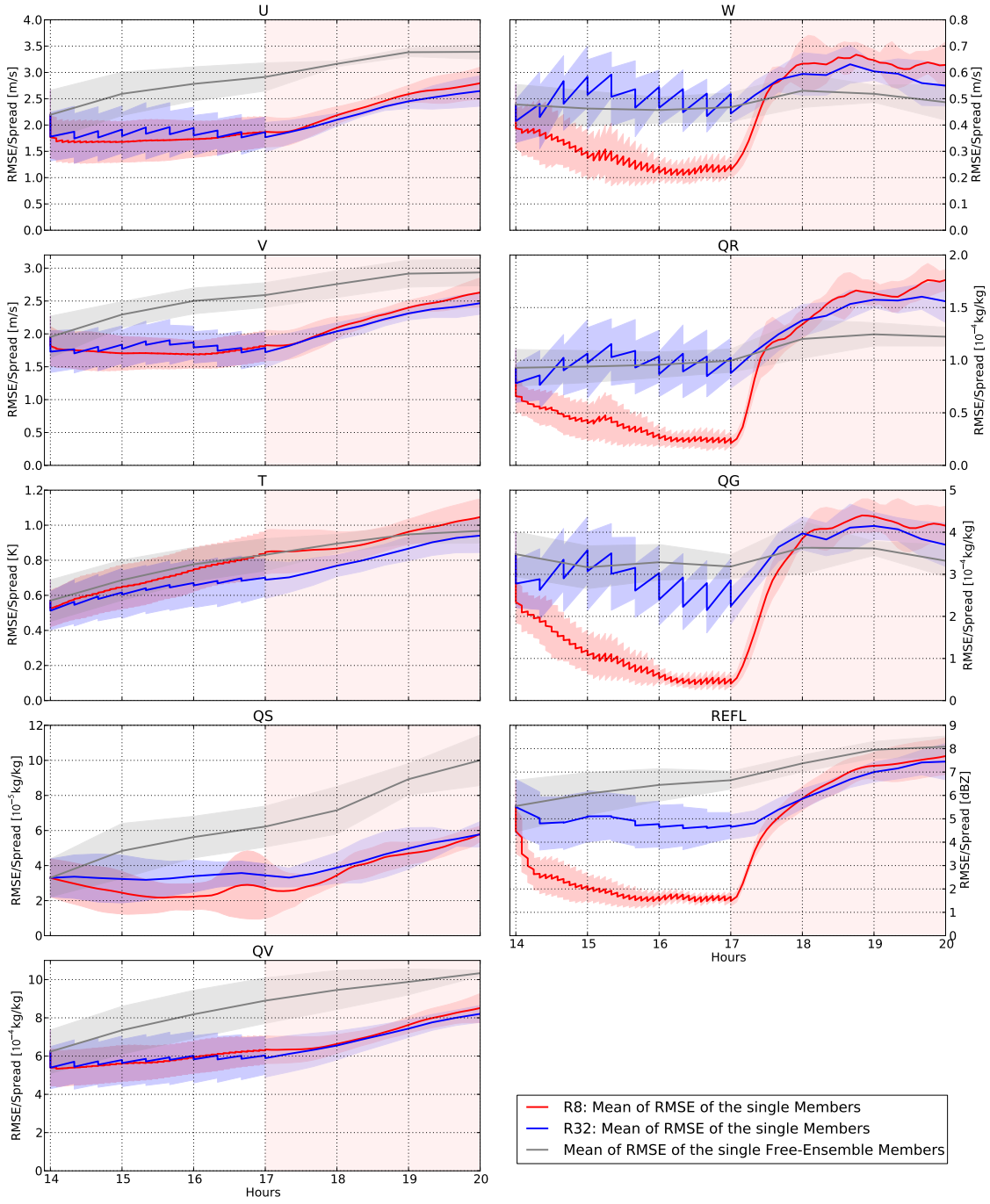

Figure 41: Like Figure 25, but for the mean of the RMSE of the single ensemble members of R8 (red) and R32 (blue) instead of the RMSE of the ensemble-mean states. All grid-points are evaluated. The error values are averaged over five repetitions of the experiment. An envelope of $\pm\sigma$ (one empirical standard deviation) is depicted by the shaded areas.

4 Summary, Discussion and Outlook

This chapter first summarizes the assimilation and forecast methods that were applied, and their results. A discussion of the results and what they implicate follows. Conclusively, the thesis generalizes the findings and takes an outlook on possible further topics.

4.1 Summary of assimilation and forecast methods

This study performed convective Data Assimilation with a Local Ensemble Transform Kalman Filter (LETKF) in an idealized testbed with simulated Doppler radar observations from a nature run, assuming a perfect model. Convective storms were triggered stochastically in the convection-permitting COSMO-model with 2 km horizontal resolution, using thermal forcing and initial small-amplitude random noise in a flat domain with periodic Boundary Conditions. A high value of CAPE in the environmental sounding enabled the occurence of strong single cells and Mescoscale Convective Systems with lifetimes > 6 h. The storms' positions in nature run and ensemble members were completely uncorrelated before the first assimilation step, challenging the filter to converge the background ensemble towards the real storms without any prior knowledge.

Using information about updraft positions from observations of reflectivity and information about the divergent structures from observations of the horizontal wind, the filter successfully lead the mean state of the ensemble to converge onto the observed storms during a 3-hour period of assimilation cycling. Updraft positions, storm intensities and cold pool structures were well-determined by the cycled analyses of the fine analysis scheme R8 that was devised to reproduce studies on convective data assimilation by other authors (cf. Section 1.2.4).

The LETKF of Hunt et al. (2007) in the DWD-implementation of KENDA was used the first time for this special purpose by this study and the results are regarded as a successful application of the filtering scheme. The highly parallelized LETKF, localizing in $\mathbf{R}$, should hereby be able to compete with EnKF-algorithms which localize $\mathbf{P}^b$, like the EnSRF does.

The assimilation period of three hours was followed by a 3-hour window of ensemble forecasts. Five repetitions of the experiments were performed to give the results more statistical robustness than for a case study.

4.1.1 Functionality of fine and coarse assimilation scheme

Two different approaches to represent the observed storms in the ensemble were devised to assess the usefulness and possible advantage of a fine analysis over a coarse analysis regarding the quality of ensemble forecasts from their analyses. These approaches were compared with respect to RMSE and spread of the model variables and object-based properties of the precipitation field.

a) A fine analysis scheme (R8) was used to recreate results of previous studies on convective DA. The full observational resolution of 2 km was used with an inflated observation error covariance matrix; the inflation was shown to be necessary by the non-inflated experiment R8_forced that produced worse analysis means. A short horizontal radius of 8 km was used for the observation error covariance localization, and a rapid update cycle of 5 minutes. These properties caused the storms in the R8-ensemble to collapse onto the observed storms with very similar updraft cores in the members, causing a low error in the mean but an ensemble with little spread regarding storm positions. Single analysis-storms were likely constructed internally from different ensemble members due to a localization radius smaller than the observed storms. Spurious storms were almost completely suppressed by assimilating observations of no-reflectivity outside the observed storms.

b) A coarse analysis scheme (R32) used horizontally averaged superobservations with a resolution of 8 km and an (effectively) even more inflated error covariance matrix; this additional inflation was shown to be necessary by the experiment R32_forced which had the same observational weighting as R8 and produced a collapsed and noisy coarse ensemble with unphysical member states. A wider horizontal localization of 32 km was used for R32, and a slower update cycle of 20 minutes. The mean state of this coarse scheme did not collapse completely onto the observed clouds, resulting in a blurrier mean field for the updraft cores, and the R32-ensemble retained full spurious storms in the members; this showed that the Gaussian filter-constraint on the analysis ensemble was weaker in R32, allowing multi-modal posteriors that were impossible in R8. The mean state of R32 therefore had a larger error compared to the nature run, and more spread was present for horizontally intermittent storm-positions than in the fine scheme – yielding a higher consistency ratio for W and the precipitation-variables.

The horizontally continuous variables of the horizontal winds U and V and the temperature T were assimilated comparably well in both coarse and fine analysis schemes; therein no scheme showed a significant advance over the other. The larger horizontal localization radius in R32 enabled the filter to construct ensemble members containing storms that were dissimilar between the members and presumably more consistent with the model dynamics than in R8, due to the larger-scale structures determined by the smoother varying analysis weights in R32. It must be noted that this coarse scheme worked although it was technically challenged by the absence of observations in the domain borders, due to the periodic flow and the non-periodic solution of the filter.

Analysis states of both R8 and R32 were negatively affected by a positive bias of

precipitation-content and, consequently, by surface cold pools that were cooled too strongly by too much precipitation. These biases were more severe in the non-collapsed experiment R32 wherein spurious storms together with spurious cold pools occured. Generally, analyses of the unobserved temperature variable were imperfect because the T-field was only provided by members that had great analysis weights determined by observations of wind and reflectivity; these two variables observed by Doppler radar are apparently not always well-correlated with T.

It was found that R8 and R32 generally fulfill the purposes of a fine and a coarse analysis, formulated in Section 1.3, with the mean of both centered on the observations, and perturbations of smaller scale contained in R8.

4.1.2 Forecasts from fine and coarse analyses

Three-hour Quantitative Precipitation Forecasts (QPF) of the assimilated thunderstorms in R8 and R32 were then conducted. The error growth of the small-scale perturbations in R8 was faster and more non-linear than the growth of the large-scale perturbations in R32. Distinct updraft cores that were caught by the R8-ensemble moved apart quickly in the forecast; with them, the RMSE of the horizontally intermittent variables like W, rain and graupel in R8 grew to the level of R32 within one hour. Although R8 had more precise storm positions in the last analysis, the positioning error of the storms in the forecast window became indistinguishable between R8 and R32 after two hours, evaluated by the field-matching and object-based scores DAS-DIS and SAL-L. Judging the quality of storm forecast probabilities by the Brier Score and ROC-area showed a similar convergence between R8 and R32. Forecasts for the horizontally continuous fields showed a qualitative behavior that was very similar between R8 and R32 – none of them appeared to forecast a significantly better horizontal field of wind or temperature.

In Section 1.3 it was presumed that forecasts based on the fine analysis scheme might only be significantly better and more skillful during the first one or two hours due to the limited predictability; this presumption is considered as confirmed.

4.2 Discussion of the assimilation and forecast results

All results here were gained in a testbed with a perfect model approach, which may mitigate their findings somewhat because the important problem of model error influence is disregarded. It is highly improbable that thunderstorm forecasts initiated from real-data-analyses will reach the same absolute quality as the forecasts that were presented here; various deficiencies of weather models concerning discretization and the very important microphysics will render forecasts to be worse in real situations. Nonetheless, information has been gained about the dynamical processes that limit the atmospherical predictability (cf. Section 1.1.3), showing that forecasts from a fine analysis will be superior to forecasts from a coarse analysis only for a time-interval of $\mathcal{O}(1\,\mathrm{h})$ wherein the rapid error growth of small perturbations has not destroyed the state's precision.

The cold-and-wet-bias of both experiments does not affect this finding, as it was more severe in the coarse analysis and forecasts from the coarse analysis were not worse than from the fine analysis in terms of the well-forecasted thunderstorm positions and convective modes.

A fine analysis that resolves dynamical details of thunderstorms could be useful for very short-range and local forecasts with lead times < 1 hour, e.g. for a probability forecast of a tornadic supercell or a downburst-warning for an airport. For operational forecasting with local models like COSMO-DE, where lead times of three hours are typical, a coarse analysis with large localization, coarse superobservations and non-collapsed ensemble might be a sufficient choice given the limited predictability even in a highly convective regime as the one presented here. The larger dynamic consistency of a coarse analysis is a strong argument when real observations are assimilated; these could be slightly or strongly incompatible with the model dynamics of those ensemble members that happen to provide a forecast at that specific time. Orographic triggering and synoptic forcing on scales that are larger than single convective systems (both effects were excluded in this study) will enhance the predictability in operational weather models; diminishing the advantage of a fine analysis for short-range forecasts even further as their benefits concerning analysis state quality might be overlayed by orographical and synoptic effects. Additionally, computing the analysis weights of the LETKF on a coarse analysis grid using few but significant superobservations (as in the R32 experiment) dramatically reduces the computational cost of one analysis step.

4.2.1 Generalization of the findings

It can be argued that these findings might not be applicable to other meteorological situations with lower or higher CAPE and smaller or larger convective systems.

In the case of smaller storms, this can be disproved quickly: In a more stable sounding with less vertical wind shear, forecasts of convective cells less intense and less organized in the mesoscale will generally be worse because the rapid error growth of the small scales affects the smaller convective cells as a whole more rapidly. The fine scheme might be the only one that gives reasonable analyses in that case at all, but the predictability would be low anyhow, so the case of smaller convection appears not so important here.

Let us consider storms in regimes with a higher CAPE than the 2000 J/kg that were used here, covering the dynamical spectrum of multicell storms, mesocyclonic supercells and squall lines. As mentioned in Section 1.1.1, Bryan and Morisson (2012) showed that simulations of MCS with a low resolution of 4 km were able to represent the coarse mesoscale structure of the storm (although not as well-detailed as very high resolving runs with 250 m resolution), apparently good enough to use the low resolution model to forecast Mescoscale Convective Systems (MCS). For such a low-resolving model, a coarse analysis might suffice, as the model could not make use of the fine scheme's additional information.

For higher resolving models, and given the long lifetime of the convective systems in this study, the findings should be applicable for intense storms that are even more

organized internally like supercells or squall lines with linear gust fronts where the degree of mesoscale organization enhances their mesoscale predictability (Aksoy et al., 2009). The coarse scheme could provide the right position and structure sufficiently and dynamically consistent. Given the fact that the storm-internal error-growth rate in three-dimensional turbulence depends mainly on the spatial scale of the perturbations and that small perturbations grow faster (Lorenz, 1969), the well-defined internal structure of ensemble storms generated by the fine-scheme would deteriorate also in such storms; the resulting storms would be comparable in quality to those from the coarse scheme again. Furthermore, a small initial error-amplitude only provides an advantage over a very short time due to the non-linear error-growth rate as shown by Zhang et al. (2003).

The convective modes in local cloud-models strongly depend on a good vertical sounding forecast of a surrounding global or mesoscale model that provides the boundary conditions. A bad forecast of the conditional instability and vertical shear of the atmosphere will produce the wrong kind of storms in the model, e.g. multicell storms instead of squall lines. Stensrud and Gao (2010) pointed out how important this is to forecast the right convection location and type; Aksoy et al. (2009) emulated the mesoscale uncertainty in their real-data cases by perturbing the vertical sounding of their members.

If real radar data of a storm is assimilated and the models' sounding is wrong, the assimilated storm might be inconsistent with the available modes of the environmental sounding it is embedded in if forced too strongly by a collapsing fine scheme. For example, Aksoy et al. (2010) found an assimilated convective line to be too shallow because the sounding for that case did not support long-lived convection. Ensemble forecasts initiated with such an inconsistently assimilated analysis storm are expected to converge to the preferred storm mode of the ensemble models, as indicated in the present (perfect model-) study by quite constant structure SAL-S and the converging number of storm-objects in Section 3.3. A coarse and non-collapsing scheme for real observations could allow at least the consistent specific type of convection within the model, disregarding initial conditions that are only precise storm-internally and therefore not useful in the forecast.

A capability of the filter to adjust a synoptically wrong sounding forecast by locally assimilating correct storms appears unlikely as the amount of energy introduced or extracted by local filter increments of the storms will probably not suffice for changing sounding properties in the whole domain of the local model. If that were so, the forceful fine scheme with its collapsed storm representations could have a slight advantage. Instead, launching radiosondes frequently to measure the current profile and giving their sounding-observations a large and ample influence appears more appropriate.

Another problem is posed by the late detection of storms using only radar observations, as radars can measure only precipitation – which is not present until roughly 30 minutes after the initial development of a thunderstorm (Houze, 1993). Dawson et al. (2012) tried to solve this by adding warm bubbles into the members during the cycling at locations where reflectivity was observed but not present in the ensemble, in order to encourage the development of updrafts. This appears appropriate for their collapsed ensemble, similar to the one presented in the fine experiment R8. The uncollapsed coarse ensemble of

R32 might be able to handle such a situation without such "help from outside the filter" because therein spurious storms are widely distributed over the whole domain and can be amplified by the filter if necessary, simply by giving the fitting member a high weight in the analysis.

4.3 Outlook

Considering this variety of eventualities just presented, testing the fine and coarse scheme in case-studies or test-periods with an operational model and real data could gain more insight. It will be especially interesting how the two schemes perform when synoptic forcings like cold-fronts or convergence lines are present, or, generally, in regimes with different convective time scales as in Craig et al. (2012). The KENDA system for COSMO-DE is an appropriate tool for such experiments.

It should be tested which influence the cold-and-wet bias of both schemes has in a real data setting or how it could be diminished. Using a simulated surface mesonet for observations of pressure, temperature and wind like in Dong et al. (2011) could be tested to improve the cold pool and gust front analyses. Spurious storms and cold pools in the coarse analysis scheme could be reduced by deflating the $\mathbf{R}$ entries for the superobservation of *no-reflectivity*, suppressing spurious convection also in R32 and leaving only representations of the observed clouds, still with storm-internal non-collapsed variance.

It appeared as a problem that the application of "zero-rain"-increments might not be enough to suppress spurious convection properly in an ensemble that has not collapsed onto the observed "positive-rain"-observations (Craig and Würsch, 2012). It was proposed by Eugenia Kalnay[1] that one could transform observations and first guesses of radar-reflectivity into a new observation variable that has a Gaussian distribution. In this transformed variable, "zero-reflectivity" corresponds to a negative value and could be assimilated in better accordance with the prior Gaussian assumptions of the Kalman filter.

The experiment R8_forced exhibited strong insertion shocks due to the rigorous filter increments caused by the large $\mathbf{R}^{-1}$. These shocks were present also in the other experiments when increments are added in the linear combination of the analyses, but the lower amplitude of these increments allowed the dynamics to adapt smoothly. If a rigorous forcing is inevitable, mollifying the filter-influence with hybrid-schemes like in Lei et al. (2012) could be a good approach.

Specifically using the framework developed here, other observations can be simulated and tested easily. Generating and assimilating simulated satellite images to detect storms in $\mathcal{O}(10\,\text{min})$ earlier before they are measured by the radar could help the initiating process of the convective systems. The Running In Place (RIP)-method by Kalnay and Yang

1 in a presentation at the International Symposium On Data Assimilation 2012 at DWD in Offenbach and during personal communication

(2010) could be investigated and how it could help the ensemble to spin-up quickly in a mathematically rigorous manner, similar to the study of Wang et al. (2012).

The author will be pleased to see his studies carried on and to see the COSMO-KENDA-SOFIA framework made use of.

Acronyms used in the text

BC Boundary Conditions

MCS Mescoscale Convective Systems

OSSE Observations System Simulation Experiment

DTE Difference Total Energy

NOAA National Oceanic and Atmospheric Administration

EnKF Ensemble Kalman Filter

LETKF Local Ensemble Transform Kalman Filter

EnSRF Ensemble Square Root Filter

DA Data Assimilation

KENDA Kilometer-scale ENsemble Data Assimilation

CAPE Convectively Available Potential Energy

WRF Weather Research and Forecasting

COSMO Consortium for SMall-scale Modelling

DWD Deutscher Wetterdienst

RMSE Root Mean Square Error

ETS Equitable Threat Score

QPF Quantitative Precipitation Forecasts

SOFIA Synthetic Observations From Idealized nAture-run

SO superobservation

PDF Probability Density Function

RIP Running In Place

ROC Relative Operating Characteristics

Bibliography

Aksoy, A., Dowell, D. C., and Snyder, C. (2009). A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Monthly Weather Review*, 137:1805–1824. (Cited on pages 6, 8, 12, 21, 22, 23, 30, 36, and 84)

Aksoy, A., Dowell, D. C., and Snyder, C. (2010). A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part II: Short-Range Ensemble Forecasts. *Monthly Weather Review*, 138:1273–1292. (Cited on pages 8, 25, and 84)

Alpert, J. C. and Kumar, V. K. (2007). Radial Wind Super-Obs from the WSR-88D Radars in the NCEP Operational Assimilation System. *Monthly Weather Review*, 135:1090–1109. (Cited on page 38)

Anderson, J. L. (2008). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61:72–83. (Cited on page 19)

Baldauf, M., Förstner, J., Kling, S., Reinhardt, T., Schraff, C., Seifert, A., and Stephan, K. (2011a). Kurze Beschreibung des Lokal-Modells Kurzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD. Technical report, DWD Deutscher Wetterdienst, Offenbach. (Cited on page 3)

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., and Raschendorfer, M. (2011b). Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Journal of the Atmospheric Sciences*, 12:3887–3905. (Cited on page 2)

Berre, L. and Desroziers, G. (2010). Filtering of Background Error Variances and Correlations by Local Spatial Averaging: A Review. *Monthly Weather Review*, 138:3693–3720. (Cited on page 17)

Bischof, M. (2011). Ensemble Simulations of Convective Storms, school = Institute for Atmospheric and Climatic Sciences (IACETH) at Swiss Federal Institute of Technology Zürich. Master's thesis. (Cited on page 31)

Bocquet, M., Pires, C. A., and Wu, L. (2010). Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Monthly Weather Review*, 138:2997–3023. (Cited on page 9)

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3. (Cited on page 49)

Bryan, G. H. and Morisson, H. (2012). Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics. *Monthly Weather Review*, 140:202–225. (Cited on pages 3, 5, and 83)

Bryan, G. H., Wyngaard, J. C., and Fritsch, J. M. (2003). Resolution Requirements for the Simulation of Deep Moist Convection. *Monthly Weather Review*, 131:2394–2416. (Cited on page 3)

Caya, A., Sun, J., and Snyder, C. (2005). A Comparison between the 4DVAR and the Ensemble Kalman Filter Techniques for Radar Data Assimilation. *Monthly Weather Review*, 133:3081–3094. (Cited on pages 17, 21, and 30)

Craig, G. C., Keil, C., and Leuenberger, D. (2012). Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 138:340–352. (Cited on pages 6, 8, and 85)

Craig, G. C. and Würsch, M. (2012). The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model. *Quarterly Journal of the Royal Meteorological Society*. (Cited on pages 28, 30, 65, and 85)

Crum, T. D. and Alberty, R. L. (1993). The WSR-88D and the WSR-88D Operational Support Facility. *Bulletin of the American Meteorological Society*, 74:1669–1687. (Cited on page 11)

Dance, S. (2004). Issues in high resolution limited area data assimilation for quantitative precipitation forecasting. *Physica D*, 196:1–27. (Cited on page 10)

Dawson, D. T., Wicker, L. J., and Mansell, E. R. (2012). Impact of the Environmental Low-Level Wind Profile on Ensemble Forecasts of the 4 May 2007 Greensburg, Kansas, Tornadic Storm and Associated Mesocyclones. *Monthly Weather Review*, 140:696–716. (Cited on pages 21, 23, and 84)

Done, J. M., Craig, G. C., Gray, S. L., Clark, P. A., and Gray, M. E. B. (2006). Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, 132:737–756. (Cited on page 8)

Done, J. M., Davis, C. A., and Weisman, M. L. (2004). The next generation of NWP: explicit forecastsmof convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, 5:110–117. (Cited on pages 2, 3, and 36)

Dong, J., Xue, M., and Droegemeier, K. K. (2011). The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. *Meteorology and Atmospheric Sciences*, 112:41–61. (Cited on pages 21 and 85)

Dowell, D. C. and Wicker, L. J. (2009). Additive Noise for Storm-Scale Ensemble Data Assimilation. *Monthly Weather Review*, 132:1982–2005. (Cited on pages 12, 21, and 23)

Dowell, D. C., Wicker, L. J., and Snyder, C. (2011). Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell: Influences of Reflectivity Observations on Storm-Scale Analyses. *Monthly Weather Review*, 139:272–294. (Cited on pages 21 and 23)

Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C., and Cook, N. A. (2004). Wind and Temperature Retrievals in the 17 May 1981 Arcadia, Oklahoma, Supercell: Ensemble Kalman Filter Experiments. *Monthly Weather Review*, 132:1982–2005. (Cited on pages 21 and 22)

Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus*, 21:739–759. (Cited on page 9)

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162. (Cited on pages 1 and 11)

Gao, J. and Xue, M. (2008). An Efficient Dual-Resolution Approach for Ensemble Data Assimilation and Tests with Simulated Doppler Radar Data. *Monthly Weather Review*, 136:945–963. (Cited on page 21)

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757. (Cited on page 17)

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R. (2011). Balance and Ensemble Kalman Filter Localization Techniques. *Monthly Weather Review*, 139:511–522. (Cited on pages 16 and 17)

Hohenegger, C., Lüthi, D., and Schär, C. (2006). Predictability Mysteries in Cloud-Resolving Models. *Monthly Weather Review*, 134:2095–2107. (Cited on pages 6 and 7)

Hohenegger, C. and Schär, C. (2007). Predictability and Error Growth Dynamics in Cloud-Resolving Models. *Journal of the Atmospheric Sciencies*, 64:4467–4478. (Cited on page 8)

Houze, Jr, R. A. (1993). *Cloud Dynamics*. Academic Press. (Cited on pages 2, 3, 4, and 84)

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 203:112–126. (Cited on pages 1, 13, and 80)

Jung, Y., Xue, M., Zhang, G., and Straka, J. M. (2008). Assimilation of Simulated Polarimetric Radar Data for a Convective Storm Using the Ensemble Kalman Filter. Part II: Impact of Polarimetric Data on Storm Analysis. *Monthly Weather Review*, 136:2246–2260. (Cited on pages 21 and 22)

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering*, 82:35–45. (Cited on page 9)

Kalnay, E. (2003). *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press. (Cited on pages 1, 5, 6, 10, 11, and 19)

Kalnay, E. and Yang, S.-C. (2010). Notes and Correspondence Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, 136:1644–1651. (Cited on pages 23 and 85)

Keil, C. and Craig, G. C. (2009). A Displacement and Amplitude Score Employing an Optical Flow Technique. *Weather and Forecasting*, 24:1297–1308. (Cited on pages 45 and 48)

Kober, K. (2010). *Probabilistic forecasting of convective precipitation by combining a nowcasting method with severel interpretations of a high resolution ensemble*. PhD thesis, Fakultät für Physik, Ludwig-Maximilians-Universtität München. (Cited on page 50)

Kober, K., Craig, G. C., Keil, C., and Dörnbrack, A. (2012). Blending a probabilistic nowcasting method with a high-resolution numerical weather prediction ensemble for convective precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138:755–768. (Cited on page 1)

Kuhl, D., Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Oczkowski, M., Hunt, B. R., Kalnay, E., Ott, E., and Yorke, J. A. (2007). Assessing Predictability with a Local Ensemble Kalman Filter. *Journal of the Atmospheric Sciences*, 64:1116–1140. (Cited on page 8)

Lei, L., Stauffer, D. R., and Deng, A. (2012). A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. *Quarterly Journal of the Royal Meteorological Society*, 138:2066–2078. (Cited on page 85)

Leith, C. E. (1974). Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Review*, 102:409–418. (Cited on page 9)

Lilly, D. K. (1990). Numerical prediction of thunderstorms – has its time come? *Monthly Weather Review*, 116:779–798. (Cited on pages 6 and 8)

Lin, C., Vasic, S., Kilambi, A., Turner, B., and Zawadzki, I. (2005). Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophysical Research Letters*, 32:L14801. (Cited on page 1)

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciencies*, 20:130–141. (Cited on page 6)

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, 21:289–307. (Cited on pages 6 and 84)

Lu, H. and Xu, Q. (2009). Trade-Offs between Measurement Accuracy and Resolutions in Configuring Phased-Array Radar Velocity Scans for Ensemble-Based Storm-Scale Data Assimilation. *Journal of Applied Meteorology and Climatology*, 48:1230–1244. (Cited on pages 11 and 21)

Patil, D., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E. (2001). Local Low Dimensionality of Atmospheric Dynamics. *Physical Review Letters*, 86:5878–5881. (Cited on page 17)

Potvin, C. K. and Wicker, L. J. (2012). Comparison between Dual-Doppler and EnKF Storm-Scale Wind Analyses: Observing System Simulation Experiments with a Supercell Thunderstorm. *Monthly Weather Review*, 140:3972–3991. (Cited on pages 11, 21, and 22)

Salonen, K., Jäervinen, H., Haase, G., Niemelä, S., and Eresma, R. (2009). Doppler radar radial winds in HIRLAM. Part II: optimizing the super-observation processing. *Tellus A*, 61:288–295. (Cited on pages 13 and 38)

Seko, H., Kawabata, T., and Tsuyuki, T. (2004). Impacts of GPS-derived Water Vapor and Radial Wind Measured by Doppler Radar on Numerical Prediction of Precipitation. *Journal of the Meteorological Society of Japan. Ser. II*, 82:473–489. (Cited on page 38)

Skamarock, W. C. (2004). Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review*, 132:3019–3032. (Cited on page 3)

Snyder, C. and Zhang, F. (2003). Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. *Monthly Weather Review*, 131:1663–1677. (Cited on pages 21, 22, and 23)

Sobash, R. A. and Stensrud, D. J. (2012). The impact of covariance localization for radar data on enKF analyses of a developing MCS: Observing system simulation experiments. submitted to Monthly Weather Review. (Cited on pages 21, 23, 24, 30, 41, and 65)

Stanski, H., Wilson, L., and Burrows, W. (1989). Survey of common Verification Methods in Meteorology. Technical report, WMO, Geneva. (Cited on page 50)

Stensrud, D. J. and Gao, J. (2010). Importance of Horizontally Inhomogeneous Environmental Initial Conditions to Ensemble Storm-Scale Radar Data Assimilation and Very Short-Range Forecasts. *Monthly Weather Review*, 138:1250–1272. (Cited on pages 22 and 84)

Stensrud, D. J., Wicker, L. J., Kelleher, K. E., Xue, M., Foster, M. P., Schaefer, J. T., Schneider, R. S., Benjamin, S. G., Weygandt, S. S., Ferree, J. T., and Tuell, J. P. (2009). Convective-Scale Warn-on-Forecast System - A Vision for 2020. *Bulletin of the American Meteorological Society*, 90:1487–1499. (Cited on page 25)

Tong, M. and Xue, M. (2005). Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Monthly Weather Review*, 133:1789–1807. (Cited on pages 12, 21, 22, and 36)

Tong, M. and Xue, M. (2008a). Simultaneous Estimation of Microphysical Parameters and Atmospheric State with Simulated Radar Data and Ensemble Square Root Kalman Filter. Part I: Sensitivity Analysis and Parameter Identifiability. *Monthly Weather Review*, 136:1630–1648. (Cited on pages 21 and 22)

Tong, M. and Xue, M. (2008b). Simultaneous Estimation of Microphysical Parameters and Atmospheric State with Simulated Radar Data and Ensemble Square Root Kalman Filter. Part II: Parameter Estimation Experiments. *Monthly Weather Review*, 136:1649–1668. (Cited on page 22)

Walser, A., Lüthi, D., and Schär, C. (2004). Predictability of Precipitation in a Cloud-Resolving Model. *Monthly Weather Review*, 132:560–577. (Cited on page 6)

Wang, S., Xue, M., Schenkman, A. D., and Min, J. (2012). An Iterative Ensemble Square Root Filter and Tests with Simulated Radar Data for Storm Scale Data Assimilation. submitted to QJR. (Cited on pages 21, 23, and 86)

Weisman, M. L., Skamarock, W. C., and Klemp, J. B. (1997). The Resolution Dependence of Explicitly Modeled Convective Systems. *Monthly Weather Review*, 125:527–548. (Cited on page 3)

Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). SAL - A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136:4470–4487. (Cited on pages 45 and 46)

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130:1913–1924. (Cited on page 16)

Wilks, D. (2006). *Statistical methods in the Atmospheric Sciences*. Academic Press, San Diego, London. (Cited on page 50)

Xu, Q., Lu, H., Gao, S., Xue, M., and Tong, M. (2008). Time-Expanded Sampling for Ensemble Kalman Filter: Assimilation Experiments with Simulated Radar Observations. *Monthly Weather Review*, 136:2651–2667. (Cited on pages 21 and 23)

Xue, M., Tong, M., and Droegemeier, K. K. (2006). An OSSE Framework Based on the Ensemble Square Root Kalman Filter for Evaluating the Impact of Data from Radar

Networks on Thunderstorm Analysis and Forecasting. *Journal of Atmospheric and Oceanic Technology*, 23:46–66. (Cited on page 21)

Yussouf, N. and Stensrud, D. J. (2010). Impact of Phased-Array Radar Observations over a Short Assimilation Period: Observing System Simulation Experiments Using an Ensemble Kalman Filter. *Monthly Weather Review*, 138:517–538. (Cited on page 21)

Zhang, F., Snyder, C., and Rotunno, R. (2003). Effects of Moist Convection on Mesoscale Predictability. *Journal of the Atmospheric Sciencies*, 60:1173–1185. (Cited on pages 6, 7, and 84)

Zhang, F., Snyder, C., and Sun, J. (2004). Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, 132:1238–1253. (Cited on pages 21, 22, and 34)

The odd dream of the data assimilating scientist

I was completely lost in time and phase-space and felt like I was drunk. Walking in a linear direction appeared difficult. Small changes in my steps grew to large strides quickly. My perception of myself and the world around me was somehow rather discrete. Pixelized. Bulky. I was walking home, but I didn't know the way.

An LETKF walked by casually, wearing a bell shaped hat. He looked quite foreign – I assumed he was Gaussian. He said "Hey there, I heard you were lost. Somebody told me you were here, walking, and somebody observed that turning into a different direction could probably be a good idea for you. But he wasn't quite sure. Neither am I, seeing you around here. But, hey, better than being lost, right? So, turn roughly into that direction and walk on. I shall return."

Every five minutes, the LETKF turned up and said "Ah, there you are! You should probably rather turn that way." Whenever the LETKF showed again, he turned my direction less than before, although I was still stumbling and tripping, but apparently in the right direction. My senses slowly converged.

The last time the LETKF appeared and advised me where to go, he added:

"Go forth now. You're on your own."

Acknowledgements

I would like to thank:

Prof. Dr. George Craig for challenging me with the topic of convective data assimilation and for supporting me in nailing down the actual scientific question. Our vivid and sometimes passionate discussions helped me to understand existing problems and to find approaches to unsolved questions. To him, I am grateful for inflaming me with scientific enthusiasm and insatiable hunger for knowledge.

Michael Würsch for providing the preparatory work on the topic and for being a pleasant companion to discuss with and share an office.

Dr. Daniel Leuenberger for providing the basic COSMO setup, extensive support to get the model running and for the invitation to MeteoSchweiz.

Dr. Matthias Sommer for giving me technical support during the (long) phase of setting up the KENDA-SOFIA-framework.

Dr. Hendrik Reich for answering *all* the questions I had on the LETKF of KENDA and Dr. Andreas Rhodin for helping with the feedback file creation.

Dr. Ulrich Blahak for giving me insight into the COSMO-code and for fixing bugs in COSMO that stood in the way of a stochastic initiation of convection in the model.

Dr. Tobias Selz for his aid whenever COSMO decided to have a bad day.

Dr. Tijana Janjic-Pfander for kindly pointing out and discussing the ever-undeniable mathematical foundations of the problem. This helped to take a step back from the grinding technical part and to think about the actual questions to be answered.

Dr. Martin Weissmann for inexhaustibly providing impulse and for giving the opportunity to present my work in Offenbach.

Dr. Christian Keil for simply being the one to consult.

Heinz Lösslein for patiently taking care of my excessive hunger for computational power.

Dr. Philipp Kostka for pointing out the human side of science.

Theresa Bick and Frauke Alexander for a rapid proof-reading. I know now that the Latin ‘exspectare’ and the English ‘to expect’ differ in one ‘s’ and that ‘therefor’ is something else than ’therefore’. Therefor, I exspect to keep myself from making these mistakes. Now.

“PauL – Poesie aus Leidenschaft” and “Die Rationalversammlung” for allowing me to balance arts and science, especially Dr. Bumillo and Magister Scharri for being insightful scientific artists and trusted friends.

A Technical Appendix

Here the framework COSMO-KENDA-SOFIA (Synthetic Observations From Idealized nAture-run) and its modules are described. One should first read and understand the KENDA workflow before approaching the specialized SOFIA.

- Section A.1 explains KENDA.
- Section A.2 explains KENDA-SOFIA.
- Section A.3 explains some specific settings used for experiments of the thesis.

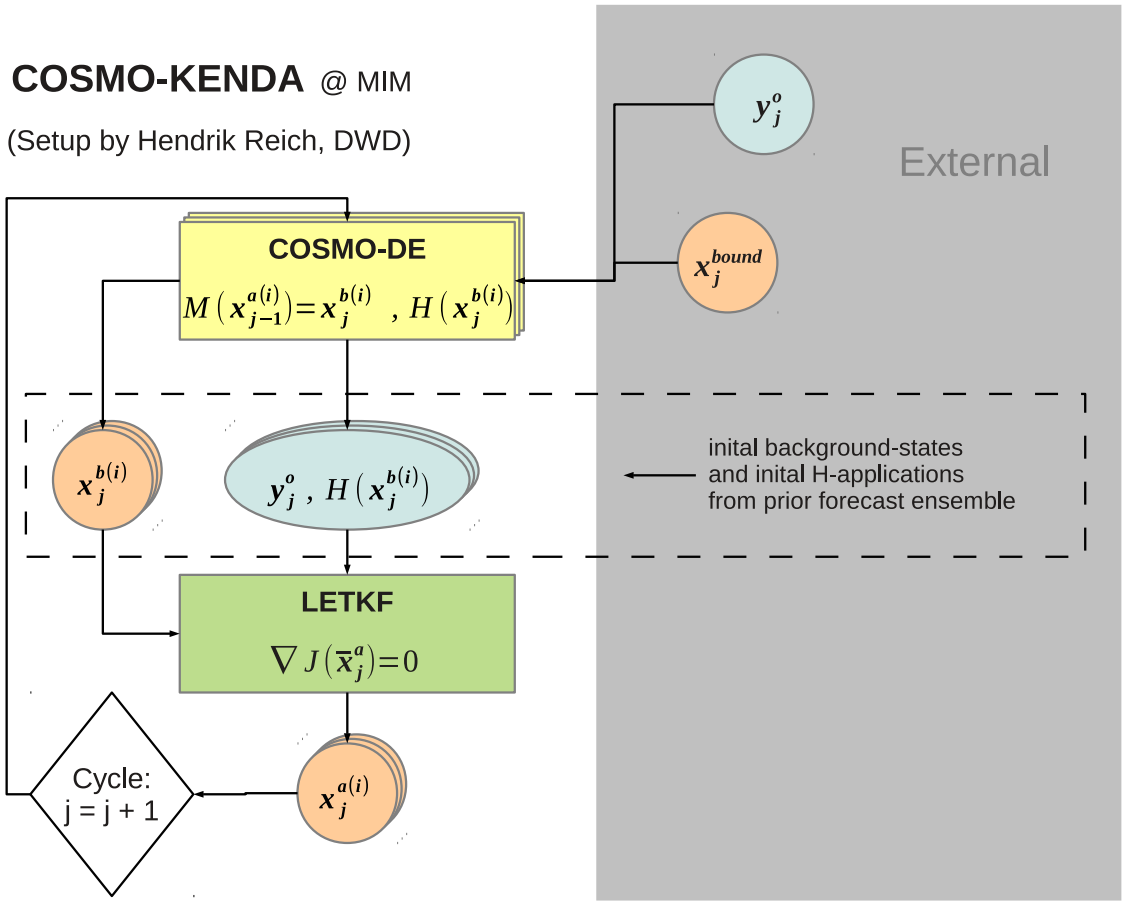

Figure 42: Flowchart of COSMO-KENDA as it is implemented at MIM.

A.1 COSMO-KENDA

The cycling flow of the quasi-operational KENDA (Figure 42) starts with the analysis step, followed by the forecast steps. Expressions in brackets are fileprefixes like `lff, laf, fof` and fileformats like GRIB, NetCDF, ASCII or NPY[1]. COSMO-KENDA uses the operational or quasi-operational COSMO-DE model.

Workflow

1. **Analysis Step (time-index $j = j$)**
   - A file-set of an initial background ensemble $\vec{x}^{b(i)}$ (`lff`, GRIB) is provided together with an ensemble of initial *feedback files* (`fof`, NetCDF) which contain the observations $\mathbf{y}^o$ and the first guesses of the members $H(\mathbf{x}^{b(i)})$. In the 0-th cycle, these need to be provided externally. Note that $\mathbf{y}^o$ is redundantly contained in every feedback-file.
   - $\vec{x}^{b(i)}$ and $[\mathbf{y}^o, H(\mathbf{x}^{b(i)})]$ are read in by the LETKF (which is the program `3dvar` in the LETKF-mode) which computes and saves the analysis-ensemble $\vec{x}^{a(i)}$ (`laf`, GRIB).
2. **Forecast Step (cycled time-index $j = j + 1$)**
   - An ensemble of COSMO-DE reads in the recently produced $\vec{x}^{a(i)}$ together with the externally provided boundary conditions x^{bound} (`lbf`, GRIB) and the observations $\mathbf{x}^o$ (NetCDF).
   - COSMO computes an ensemble of background forecasts $\mathbf{x}^{b(i)}$ for the next analysis step. The observation operator H is contained in COSMO and maps the observations onto the forecast (per member), producing the feedback files.
3. Next analysis time. Back to Analysis Step 1.

1 NPY-files with the suffix ".npy" are used to save and reimport arrays from the Python-module numpy

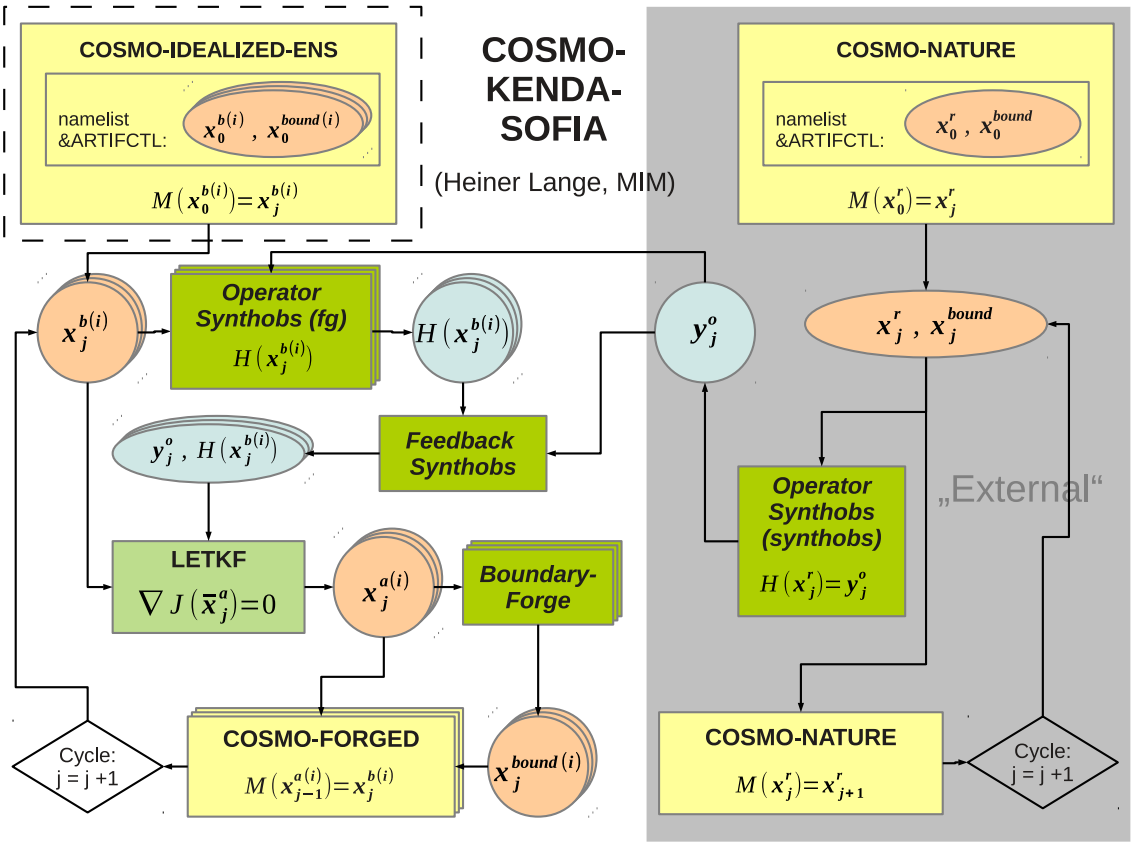

Figure 43: Flowchart of COSMO-KENDA-SOFIA.

A.2 COSMO-KENDA-SOFIA

To perform OSSE-experiments with SOFIA, additional steps have to be performed. It is recommended to keep these steps in the same order. SOFIA uses idealized instances of COSMO in the version `COSMO 4.22 UB/HL` (described later). The following work-flow is depicted in Figure 43. The number of ensemble members is k. Some modules are modified in peculiar but necessary ways – to understand the cycling completely, reading both the workflow and module-descriptions is mandatory. A complete and linear explanation would be to long for this description, so reading more than once might be necessary.

A.2.1 Workflow

1. **Preparation of Ensemble and Nature Run**

   - COSMO-NATURE produces the forecast files $\mathbf{x}^r$ (`lfff`) for the nature run. The superscript "r" here depicts that these are the "reference states". They need to be written in an interval that fits the forecast-intervals of the LETKF-cycling, e.g. 5 min in between. It is recommended to link the `lfffddhhmmss`-forecast-files as `lffyyyymmddhhmmss` (with the exact datetime of their validity!) to make the generation of the synthetic observations easy at all analysis times.
   - Files with initial $\mathbf{x}_0^r$ and boundary-conditions $\mathbf{x}_0^{bound}$ are not needed in the idealized COSMO, as boundary conditions are either open boundaries or periodic boundary conditions. The "cycling" part on the right of Figure 43 is just a graphical expression for "one COSMO instance, doing a forecast over a prescribed time-period and producing forecast files".
   - COSMO-IDEALIZED-ENS produces the initial ensemble $\vec{x}^{b(i)}$ (`lfff` linked as `lff` with full date). Every member typically has different initial conditions, more or less close to those of the nature run. It is recommended to let COSMO-IDEALIZED-ENS run for the same full forecast period as the nature run to have a free uncorrelated ensemble at all times for reasons of comparison, or to be able to choose an ensemble at a different time to start the cycling.

2. **Cycling in KENDA-SOFIA**

   a) First the synthetic observations are generated. The program `operator_synthobs` in mode "synthobs" reads the nature runs forecast $\mathbf{x}^r$ which is valid at that analysis time and applies the observation operator $H(\mathbf{x}^r) = \mathbf{y}^o$. The result is the synthetic observation vector $\mathbf{y}^o$ (`synthobs`, ASCII). Coordinate-informations of the observations are also exported (NPY).

   b) After that, `operator_synthobs` is called k times in mode "fg". Every i-th instance reads in $\mathbf{y}^o$ with coordinates and one member of $\vec{x}^{b(i)}$, then applies the operator to produce the first guess $H(\mathbf{x}^{b(i)})$ (`fg`, ASCII).

c) The program `feedback_synthobs` reads $\mathbf{y}^o$ and all $\vec{x}^{b(i)}$ and produces k *feedback files* containing $[\mathbf{y}^o, H(\mathbf{x}^{b(i)})]$.

d) As in operational KENDA, $\vec{x}^{b(i)}$ and $[\mathbf{y}^o, H(\mathbf{x}^{b(i)})]$ are read by the program `3dvar` in the LETKF-mode which computes and saves the analysis-ensemble $\vec{x}^{a(i)}$ (`laf`, GRIB).

e) The program `boundary_forge` is called k times and reads $\vec{x}^{a(i)}$, producing an ensemble of forged boundary conditions $\mathbf{x}^{bound(i)}$ (`laf`, GRIB).

f) An ensemble of k COSMO-FORGED instances reads the recently produced $\vec{x}^{a(i)}$ together with boundary conditions $x^{bound(i)}$ and from there computes the background ensemble for the next analysis time: $M(\mathbf{x}_{j-1}^{a(i)}) = \mathbf{x}_j^{b(i)}$.

g) Next analysis time at time $j = j + 1$. Back to step a)

A.2.2 Modules and Modifications

Here a list of scripts and modules that were created or modified by Heiner Lange for KENDA-SOFIA. For a detailed description of the modules see the scripts' codes and the COSMO-Userguide.

▪ LETKF runscript

`letkf_runscript.py` is a Python script that organizes the SOFIA-workflow shown in Section A.2.1. From here the modules are called, jobs are submitted into the cluster queue and their output is moved and archived.

Input: `letkf_runscript_config.py`, containing settings for

- experiment ID
- archive- and work-directory
- directory with initial ensemble (has to be precomputed!)
- directory with the respective nature run (has to be precomputed!)
- start and stopdate of assimilation
- forecast interval time between analysis steps
- number k of ensemble members
- a possible parent-experiment the current experiment is a derivative from

- a possible time-interval with a free ensemble forecast
- possible time-intervals for deterministic forecasts from analysis means
- binaries for COSMO and `var3d`
- namelists for the `var3d` and COSMO-FORGED (it is vitally important that the namelists for COSMO-FORGED have the same or compatible dynamical settings as the namelists of COSMO-IDEALIZED-ENS!)

Useful additional features besides organizing the workflow:

- collects and presents possible error-messages from modules (except LETKF and COSMO, they have own error outputs)
- sends the user an Email if the experiment has to be aborted because a module failed in the cycling (e.g. due to a non-responding file-system)
- also sends an email if an experiment has finished succesfully
- tolerates possible non-responding off-times of the queue-master by waiting

---

▪ Operator Synthobs

`operator_synthobs` is a Python-script. In the main-mode "synthobs" it computes synthetic observations, in the main-mode "fg" the first guesses:

▪ Operator Synthobs in mode "synthobs"

Call: Once for the synthetic observations

Input:

- `operator_synthobs_config.py` with settings for:
  - observed variables (COSMO-variables or combinations of them like *reflectivity*)
  - switch for reflectivity masks for different variables
  - error values that are added to different observation type
  - the error values that are reported for the filter's **R**-matrix
  - the observation geometry
  - settings for superobservations
- 1 nature run file (GRIB)

Computations: $\mathbf{y}^o = H(\mathbf{x}^\mathbf{r})$ and adding prescribed noise

Output:

- 1 synthobs-file (ASCII), containing observation values, locations and errors and types
- synthetic observations per obs-type, observation geometry and reflectivity masks (NPY)
- namelist file `namelist_feedback_synthobs` for `feedback_synthobs`

▪ Operator Synthobs in mode "fg"

Call: k-times for the k first guesses (running parallelized)

Input:

- same `operator_synthobs_config.py` as in mode "synthobs" (if perfect model experiment)
- 1 forecast member file (GRIB)

Computations: $H(\mathbf{x}^{\mathbf{b(i)}})$

Output: 1 first guess-file (ASCII), containing only first guess values

---

▪ Feedback Synthobs

`feedback_synthobs` is a utility program that uses routines of the `3dvar`-package it is compiled with.

Call: Once for the whole ensemble

Input:

- 1 synthetic observation file (ASCII)
- k first guess files (ASCII)
- namelist file `namelist_feedback_synthobs` generated by `operator_synthobs`. For a description, see the code there that generates the namelist.

Computations: None, only file-conversion

Output: k feedback files (NetCDF), each containing the i-th first guess and the observations

---

▪ **LETKF** via var3d

The LETKF-computations are performed by the program `var3d` (developed at DWD by Andreas Rhodin et al.) in the LETKF-mode (developed by Hendrik Reich et al.) [2]. Filter-Updates of the hydrometeor-mixing ratios of rain, graupel and snow (QR, QG, QS) were disabled by default and enabled for this thesis by modifying `mo_letkf.f90`. They can be enabled in the namelist of `var3d`. Also in `mo_letkf.f90`, observation analysis statistics (some parts around the routine `derive_obs_ana`) were disabled due to a problem with the parallelized computations (possibly due to the regular observation grid geometry).

Call: Once for the whole ensemble

Input:

- k ensemble forecast files
- k feedback files
- namelist file exported from `letkf_runscript_config.py`, containing essential LETKF-filter settings like:
  1. localizaton (horizontal and vertical)
  2. covariance inflation
  3. analysis grid coarsing
  4. model variables to be updated or passed through by the filter

Computations: $\nabla J(\bar{\mathbf{x}}^a) = 0$

2 Sometimes the program is called 3dvar, but the binary of it is named var3d. One should not get mixed up by this and the "3dvar" name of it: It IS an LETKF.

Output:

- k analysis members $\mathbf{x}^{a(i)}$ (`lff`, GRIB, filesuffix `.001` ... `.k`)
- 1 file containing the forecast mean $\bar{\mathbf{x}}^b$ with the same variables as the forecast members (`lff`, GRIB, `.mean`)
- 1 file containing the forecast spread fields gridpointwise with the same structure as the model grid (`lff`, GRIB, `.spread`)
- 1 file containing the analysis mean $\bar{\mathbf{x}}^a$ (`laf`, GRIB, `.mean`)
- 1 file containing the analysis spread fields gridpointwise with the same structure as the model grid (`laf`, GRIB, `.spread`)

The LETKF-implementation provides evaluation-tools that are not described here. The version of the LETKF was fixed somewhere around June 2012 for SOFIA – an update to the latest version is recommended but needs testing!

Possible bug in that version:

The analysis mean- and spread field of V seems to be mistaken with a wrong variable. Maybe somewhere the variable names get mixed up. Computing means and spreads externally using `cosmo_diag` is therefore recommended for safe evaluations.

▪ COSMO 4.22 UB/HL

COSMO-NATURE, COSMO-IDEALIZED-ENS and COSMO-FORGED are instances of `COSMO 4.22 UB/HL`. This version was bugfixed by Uli Blahak (UB) from DWD and modified by Heiner Lange. Modifications in the source code were noted by commenting "`lange`" at the respective lines. The COSMO-version contains the following features:

- Bugfix by Uli Blahak: When COSMO is run in idealized mode (`lartif_data=.true.`) with periodic boundary conditions and active radiation scheme, a discontinuity in radiative forcing showed (due to the earth curvature) when the flow passed the periodic boundaries. Now, when the radiation scheme is called, suntime and sun zenith angle are the same for the whole domain as with the coordinate-origin `(rlat,rlon)=(0,0)`.
- To allow a predefined random seed for the boundary layer noise of T and W, the parameters `iseed_noise_t` and `iseed_noise_w` are introduced in the namelist `&ARTIFCL` and worked into `src_artifdata.f90`. This enables different but predefined and reproducible realizations of nature run and ensemble members with convection evolving in locations uncorrelated amongst each other.

- It must be possible to start COSMO at times other than full hours. The newest COSMO-version should already suppport this – here it was manually, lengthening the time-variable `ydate_ini` to support a defined starting minute and second. This was changed in the following source-files:
  - `src_input`
  - `src_setup`
  - `data_runcontrol`
  - `utilites`
  - `data_io`
  - `src_gridpoints`
  - `src_meanvalues`
  - `src_radiation`
  - `src_output`
- To enable the combination of periodic boundary conditions together with starting COSMO using an analysis state file (`laf`), the logical switch `lartif_ana` was introduced in the namelist `&RUNCTL` and worked into `organize_data.f90`.

  `lartif_ana` can only be `.true.` if `lartif_data=.true.`. In standard COSMO, `lartif_data=.true.` causes the routines `gen_ini_data` and `gen_bound_data` in `organize_data.f90` to create an artificial initial model field and boundary conditions instead of reading them in from `laf` and `lbf`-files as COSMO-DE would where `lartif_data=.false.`.

  Now, want to have periodic boundaries that are only possible if `lartif_data=.true.`, but also we want the initial state to be the analysis produced by the LETKF. So, if `lartif_ana=.true.`, the routines `gen_ini_data` and `gen_bound_data` are skipped (although `lartif_data=.true.` !) and the analysis-`laf`-file is read in like in operational COSMO, together with the forged boundary conditions produced by the SOFIA-module `boundary_forge`. The periodic COSMO needs these forged boundary-conditions only to initialize the fields correctly – during the model's integration, the boundary conditions are periodic again due to `lpbc=.true.` enabled by `lartif_data=.true.`.[3]

3 This is a rather "quick and dirty" solution, but it fulfills its purposes. COSMO had not been designed for such purposes, admittedly, but they became necessary by the experimental design of this thesis.

▪ **Boundary Forge**

COSMO needs a least two boundary conditions for t_0 and for $t_1 = t_0 + 1\,\mathrm{h}$ when it is started. These can be operational (COSMO-DE) or fake (idealized COSMO with `lartif_ana=.true.`). COSMO is able to use analysis-files (`laf`) as boundary files by the switch `lbdana` in the namelist `&GRIBIN`. This feature is exploited here by the Python-script `boundary_forge.py`:

Call: k-times for k analysis ensemble members

Input: 1 `laf` file (GRIB)

Computations:

1. Read original `laf`-file (the original file stays unchanged!) and create a copy in the memory (binary). This copy is then manipulated:
2. The time-codes of all the contained GRIB-records (ca 500 records per file[4], containing the model fields etc.) are set to be valid one hour into the future.

Output: 1 manipulated copy of the input-file as new `laf`-file with manipulated time-codes inside and time-advanced filename (yyyymmddhhmmss + 1hh) outside

---

▪ **cosmo_diag**

A collection of functions for analyzing the data, plotting and scoring is provided in the collection `cosmo_diag.py`. This makes heavy use of the `Nio`-module (also called PyNio, created at NCAR), that is able to read GRIB-files and extract their fields.

A.3 Specific settings for experiments of this study

Here some specific settings for the different COSMO-instances are described. All of them use `COSMO 4.22 UB/HL`. Not every specific namelist parameter can be explained here – some training in the use of COSMO might be necessary.

---

4 Exact number uncertain but unimportant here.

▪ COSMO-NATURE

The basic settings of COSMO-NATURE are described in Manuel Bischofs MSc-thesis. Here mainly the modifications are named:

In the idealized COSMO (`lartif_data=.true.`), convection can be triggered by warm bubbles with predifined position and shape. With the used Payerne-Sounding with $CAPE = 2000 J/kg$, strong storms evolve. Storm triggered by warm bubbles cause strong gravity waves rush through the domain. Here a smoother approach is followed.

To gain the setup of this thesis' Chapter 2, the random seed for the T- and W-perturbations is given by `iseed_noise_t` and `iseed_noise_w` (integer). The radiative forcing scheme is enabled. Not much more has to be done – the storms evolve then due to the radiative forcing and are randomly distributed. `lpbc=.true.` enables periodic lateral boundaries.

The forecast files of COSMO-NATURE have to contain records a) for all variables that are needed by `operator_synthobs` and b) for all variables that are to be compared with the forecasts and analysis ensembles.

▪ COSMO-IDEALIZED-ENS

In COSMO-IDEALIZED-ENS, ALL settings of namelist variables are the same as in COSMO-NATURE. Only `iseed_noise_t` and `iseed_noise_w` are chosen differently for every member, taking also care that no member has the same random seeds as the nature-run. The position of resulting storms is then completely uncorrelated.

To make the ensemble member files compatible for the LETKF, they need to contain a variety of dynamical variables which are then either passed through by the filter or updated. The following collection of variables to be set in `&GRIBOUT` is regarded as safe:

```
yvarml=
'U       ','V       ','W       ','T       ','QV      ','QC      ',
'QI      ','QR      ','QS      ','PP      ','P       ','PS      ',
'T_SNOW  ','W_SNOW  ','QV_S    ','W_I     ','QG       ',
'T_SO    ','W_SO    ','W_SO_ICE',           'RUNOFF_S','RUNOFF_G',
'RAIN_GSP','SNOW_GSP','RAIN_CON','SNOW_CON','TOT_PREC',
'U_10M   ','V_10M   ','T_2M    ','TD_2M   ','RELHUM_2M','T_G     ',
'TMIN_2M ','TMAX_2M ','VMAX_10M','TCM     ','TCH     ',
'CLCT    ','CLCH    ','CLCM    ','CLCL    ',
'ALB_RAD ','ASOB_S  ','ATHB_S  ','ASOB_T  ','ATHB_T  ','APAB_S  ',
'AUMFL_S ','AVMFL_S ','ASHFL_S ','ALHFL_S ',
'BAS_CON ','TOP_CON ','HBAS_CON','HTOP_CON','HTOP_DC ','TKE     ',
'QCVG_CON','MFLX_CON','CAPE_CON',
```

```
'TQC     ','TQI     ','TQV     ','TWATER  ','TDIV_HUM',
'CLCT_MOD','CLC     ','HZEROCL ','SNOWLMT ','CLDEPTH ','AEVAP_S ',
'PMSL    ','DPSDT   ','Z0      ','ZTD     ','ZHD     ','ZWD     ',
'PLCOV   ','ROOTDP  ','FR_LAND ','FIS     ','RLAT    ','RLON    ',
'FRESHSNW','RHO_SNOW','W_ICE   ','H_SNOW  ',           'LAI     ',
'FOR_E   ','FOR_D   ',
'HHL     ','T_S     ','VIO3    ','HMO3    ','SOILTYP ','HSURF   ',
```

Also important: The reference atmosphere of the model-instances has to be compatible with the LETKF! If the LETKF assumes a different reference atmoshpere, in restarted members of COSMO-FORGE the complete airmass is imbalanced.

Possible bug:

Giving the background ensemble a long spin-up-time before the first assimilation appeared technically problematic, as the GRIB-records in the `lfff`-forecast-files appeared to contain wrong time-codes when read by the LETKF.

This problem (or bug?) is gone around here in the following crude way: The initial ensemble COSMO-IDEALIZED-ENS writes pseudo-analysis `laf`-files every hour. At the first planned assimilation time, an ensemble of COSMO-FORGED is initialized by these `laf`s. In this ensemble each member runs for zero seconds, only producing one `lfff`-GRIB-file wherein the time-codes then seem to be correct and readable by the LETKF.

▪ COSMO-FORGED

The namelist-template for COSMO-FORGED is contained in `letkf_runscript_config.py` and has to coincide completely with the setting of COSMO-IDEALIZED-ENS, except for the following settings that now in COSMO-FORGED:

- `lartif_ana=.true.` — this causes the read-in of the `laf`-ensemble and the forged boundary-files that need to be produced by `boundary_forge` for each member before. This way, the computations of $M(\mathbf{x}_{j-1}^{a(i)}) = \mathbf{x}_j^{b(i)}$ can be done.
- The namelist-settings for starting-time and end-time need to be replaced by the current analysis-time and the next analysis time. This is done when `letkf_runscript` modifies the COSMO-FORGED namelist-templates for each member.

As said in the description of `COSMO 4.22 UB/HL`, COSMO-FORGED does not initialize artificial fields by using the settings of `&ARTIFCL` anymore. Instead, `lartif_ana=.true.`

causes the read-in of the analysis state. COSMO-FORGED is thus a chimera of a quasi-operational COSMO-DE with the idealized set-up of COSMO.

When setting up new experiments, it should be closely checked if COSMO-FORGED works as demanded. Plotting initial and intermediate forecast member fields is helpful to check for dynamical consistency.

To the possible reader and user of this technical appendix: Good luck! And sometimes: Patience.

Checking things twice or going back the possible error-cascade is often inevitable.