On the Benefits of a High-Resolution Analysis for Convective Data Assimilation of Radar Observations using a Local Ensemble Kalman Filter

Heiner Lange * and George C. Craig

Hans Ertel Centre for Weather Research, Data Assimilation Branch, Ludwig-Maximilians-Universität München, Munich,

* *Corresponding author address:* Heiner Lange, Meteorological Institute Munich, Theresienstrasse 37, 80333 Muenchen, Germany

E-mail: heiner.lange@lmu.de

ABSTRACT

The Deutscher Wetterdienst (DWD) is developing an implementation of the Local Ensemble Transform Kalman Filter (LETKF) for the cloud resolving COSMO model. This study shows in an idealized convective testbed that the LETKF is able to perform storm-scale Data Assimilation of simulated Doppler radar observations.

Perfect model experiments are used to investigate how the limited predictability of convective storms affects precipitation forecasts by comparing a fine scheme with low analysis error to a coarse scheme that allows variance regarding position, shape and occurence of storms in the ensemble. To get there, the coarse scheme uses averaged superobservations and a coarser evaluation of the analysis weights, a larger localization radius and a weaker Gaussian constraint on the analysis solution. Performing 3-hour forecasts of convective systems with typical lifetimes exceeding 6 hours, forecasts from the detailed analyses of the fine scheme are found to be advantageous to those of the coarse scheme during the first 1-2 hours, regarding the predicted storm positions. After 3 hours in the convective regime used here, the forecast quality of the different schemes appears indiscernible, judging by RMSE and verification methods for rain-fields and objects.

It is concluded that, for operational assimilation systems, the analysis might not necessarily need to be detailed on the grid scale of the model. Depending on the forecast lead time, and on the presence of orographic or synoptic forcings that enhance the predictability of storm occurences, analyses from a coarser scheme might suffice. As a positive side-effect, the computational cost of the Kalman Filter solution can be reduced strongly.

1. Introduction

In the last decade, Data Assimilation (DA) of Doppler radar observations using an Ensemble Kalman Filter (EnKF) (Evensen 1994; Houtekamer and Mitchell 1998) has been explored to be a feasible method to obtain suitable initial states of convective storms for very short-range ensemble forecasts. As Stensrud et al. (2009) note, a goal of convective DA is to be competitive to nowcast-warning systems that have generally been superior to model forecasts without data assimilation for at least the first three hours of leadtime (Kober et al. 2012).

Previous studies that used radar data in an EnKF focused on storm-analyses that apply relatively small scales of ~ 10 km for the covariance localization and converge the analysis ensemble closely towards the observations, pursuing the goal of obtaining initial states with low errors as the basis of their ensemble forecasts. This study tries to asses the performance of that converging approach, as it is costly and and may be of limited value due to the short predictability of convection in the chaotic atmospheric system where small errors grow rapidly (Lorenz 1969; Lilly 1990). The framework of the Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al. 2007) is used here in an idealized setup of the nonhydrostatic COSMO model (Baldauf et al. 2011) with severe and long lived convection.

In contrast to the converging approach, the alternative of a coarse assimilation scheme is proposed which is cheaper to compute and focuses less on the small-scale convergence of the analysis ensemble; variance and errors of the analysis should be allowed that are larger in scale *and* amplitude. For lead times of a few hours, after the strongly nonlinear growth of small errors has saturated (Zhang et al. 2003), this coarse scheme may not be subject to much penalty in forecast accuracy in comparison to a fine and converged scheme. Secondly, the coarse scheme with its larger variance may avoid dynamical imbalances caused by analysis increments than can arise when the member states of the localized analyses are rigorously converged (Greybush et al. 2011).

The sections of the introduction below are devoted to the problems of converging analyses and limited predictability and how they could be adressed using assimilation schemes of different precision. Finally, the setup for the numerical experiments is laid out.

a. Converged storm analyses in previous studies

The main observation types in convective EnKF are reflectivity and radial Doppler wind. All studies on this topic mentioned here used the EnSRF-algorithm (Houtekamer and Mitchell 2001) or a similar algorithm that localizes the background error covariance matrix $\mathbf{P}^b$, provided by a background ensemble of 50-100 members, to increase the effective sample size. In their first OSSE study on convective EnKF, Snyder and Zhang (2003) identified spurious storms in the members to be detrimental to the analysis solution. Tong and Xue (2005) were able to suppress spurious storms by assimilating observations of no-reflectivity outside the observed storms; their positive results were confirmed in the real data cases of Aksoy et al. (2009, 2010). As the ensemble mean is the solution that minimizes the cost function of the Kalman Filter, most studies focused on the improvement of this mean solution. The non-Gaussian and multimodal distribution of the observed dynamical systems (Dance 2004) forces the probability distribution of the resulting analyses, constrained to be Gaussian by the filter, to have a deficient amount of variance. A common solution for this is the inflation of the background error covariance, preferably adaptive to the respective situation (Anderson 2008). For real observations, Dowell and Wicker (2009) and Dawson et al. (2012) added noise during the assimilation inside the observed storms and Aksoy et al. (2009) perturbed the vertical wind profile. Both methods resulted in better analyses with a more consistent variance. Wang et al. (2012) applied the "Running in Place"-method of Kalnay and Yang (2010) on convective EnKF experiments, resulting in a better spin-up of the storms and a generally lower error level.

Using a Gaussian-like correlation function for the localization (Gaspari and Cohn 1999), all studies agreed on the usage of a small localization cutoff radius between 4 and 12 km, depending on the spatial stage of the storm development (Sobash and Stensrud 2013) and assumed an observation error of approximately 5 dBZ for (no-)reflectivity and 1 m s^{-1} for wind observations in the observation error covariance matrix $\mathbf{R}$. The resulting precise storm-analyses without spurious convection are hereby referred to as "converged" or "fine" analysis ensembles because all members essentially agree on position and shape of the observed storms.

b. Limited predictability of convective storms

Lilly (1990) and Skamarock (2004) estimated the predictability of mesoscale convective systems to be in the range of tens of minutes to 1 hour before the upscale error growth (Lorenz 1969) taints the forecast completely. Zhang et al. (2003), Hohenegger and Schär (2007) and Done et al. (2012) compared randomly perturbed forecasts of organized convection to unperturbed reference runs. They found small-scale perturbations to grow very quickly and nonlinearly, reaching a growth-saturation at 3-6 hours. The specific predictability limit in these studies depended on the presence of a large scale forcing that predetermined the occurence of convection spatially and modally. Craig et al. (2012) used perturbations from a Latent Heat Nudging assimilation scheme for convective storm forecasts and concluded a lower predictability for convection in regimes without synoptic forcing where the storms positions are mostly random.

c. Fine and coarse assimilation schemes

Taking the limited predictability of convection into account, and the fact that forecast products of local models are usually disseminated with lead times > 3 h, the benefits of the fine and converging approach of the previous studies on convective EnKF may not justify the cost: *A coarse and less converged analysis ensemble could provide equally good forecasts* because small perturbations grow quicker than large-scale disturbances (Lorenz 1969). As in Kuhl et al. (2007), the deviations of the analysis members from the reference run are regarded here as the perturbations which grow during the forecast.

In this paper, a fine and a coarse assimilation scheme are described. The aim of the fine scheme is to reproduce the converging ensemble behavior of the previous studies mentioned in Section 1a. The fine analysis localizes the covariances on a scale of a few kilometers and uses high resolution observations. It will therefore contain observed details on the smallest resolved scales of the model grid – which is smaller than the "effective resolution" of the model (Skamarock 2004) where the numerical solution of the equations can be considered accurate. The LETKF-algorithm used here localizes in the observation space, so it will be of interest if the results are comparable to those studies that localize the background covariance in model space. In constrast, the coarse scheme is configured to have perturbations that are larger with respect to scale, amplitude and horizontal positions of the updraft cores and are thus subject to a slower error growth (Lorenz 1969). It is based on the fine scheme and uses a scaling factor of 4 for most of the assimilation parameters: (i) a larger horizontal localization radius is applied, (ii) coarsened superobservations are used that provide information only on this larger scale and (iii) the local analysis weights of the LETKF are computed on a coarse grid and then spatially interpolated. The latter method deprives the analysis of some precision but allows the linear combination of the analysis members to vary more smoothly (Yang et al. 2009). The experiments in both the fine and the coarse scheme will show that it is also necessary to (iv) adjust $\mathbf{R}$ by inflating it in order to attenuate the rigorous convergence towards the observations, as the application of an untreated $\mathbf{R}$ results in noisy and imbalanced analysis member states that exhibit poor physical resemblance to the reference run.

Applying (i)-(iv) should allow states in the coarse analysis that are more consistent with the model physics because larger horizontal portions of background member storms are contained in the analysis member storms. As a technical benefit, the lower number of observations and the lower analysis resolution result in a significant computational acceleration of the analysis. Attributing the superobservations with less weight in the analysis using method (iv) will push the members less rigorously towards the observations by analysis increments (cf. Fig 1). This aims for a multimodal posterior distribution containing variance not only storm-internally but also with respect to the occurence of storms themselves, with the eventuality of spurious convection. These properties define the "non-converged", coarse assimilation scheme.

In addition to the convergence properties of the two experiments, other issues that are caused by the localized analysis of non-observed variables, such as a cold bias of the temperature variable, are explored.

d. Numerical experiments

The experiments are performed in a perfect model environment to focus solely on the predictability of the dynamics, excluding effects of model error. Simulated radar observations are drawn from a nature run that contains organized convective systems with lifetimes over six hours. A cycled assimilation covers a time-span of three hours, followed by three hours of ensemble forecast. For all experiments, the EnKF will be challenged by a convective ensemble that is spun up from random initial noise and therefore lacks any prior knowledge about the position of the observed storms in the nature run. This "bad background" makes the convergence of the ensemble challenging for the filter, but it is free from possibly beneficial influence e.g. of convective triggers that are introduced outside the EnKF algorithm at predetermined locations (Aksoy et al. 2009, 2010).

The 3-hour forecasts from both schemes are evaluated using the RMSE of the states together with object- and field-based forecast skill scores to evaluate how quickly the advantage of a fine analysis state is lost in the forecast with respect to the coarse analysis, due to the limited predictability.

2. Fine and coarse assimilation scheme

This section first describes the data assimilation setup, consisting of nature run, synthetic observations, convective ensemble and LETKF-algorithm, followed by the implementation of the fine and the coarse scheme. The Kilometer-scale ENsemble Data Assimilation (KENDA) system (Reich et al. 2011) is being developed at the Deutscher Wetterdienst (DWD). It couples an LETKF-implementation with an ensemble of the COSMO-DE model. The latter is used for operational forecasting over Germany (Baldauf et al. 2011).

COSMO solves the full non-hydrostatic and compressible Navier-Stokes equations using a time-splitting Runge Kutta approach for fast and slow tendencies in the prognostic wind variables U, V and W and the deviations of temperature T' and pressure p' from a stationary hydrostatic base state. The moist physics use six state variables of water vapor, cloud water, cloud ice, rain, snow and graupel with a first order bulk microphysics scheme. A radiation scheme for long- and shortwave radiation is applied. Surface fluxes of latent and sensible heat are parametrized and constrained by constant surface temperature and surface specific humidity throughout the simulation.

a. Nature run

This study uses the testbed setup of COSMO with idealized initial state, periodic boundary conditions (BCs) and a homogeneous flat landscape as the lower boundary. A convection-permitting horizontal resolution of 2 km and 50 vertical levels are used in a domain of $396 \times 396 \times 20$ km extent. The initial profile of all model runs is horizontally homogeneous and based on the sounding of Payerne (CH, Radiosonde 06610) at 12 UTC on Juli 30th 2007, a day with strong convective storms and mesoscale convective systems (MCS), favored by a high CAPE value of 2200 J kg^{-1} together with a vertical wind shear that allows organized convection with heavy precipitation and propagating gust fronts (Bischof 2011).

Instead of initializing convection with predefined warm bubbles (Aksoy et al. 2009) or targeted noise (Dowell et al. 2004; Tong and Xue 2005) with amplitudes that directly trigger thermals, uncorrelated grid point noise is added at the inital time t_0 to the temperature field T and the vertical wind speed W in the boundary layer (BL) with amplitudes of 0.02 K and 0.02 m s^{-1}. The model runs start at 06 UTC and rapidly develops a convective boundary layer. Triggering instabilities are generated through the effect of the radiation scheme on the tropospheric temperature.

First preliminary showers evolve in random locations at 08 UTC which grow until 10 UTC and mostly die off by 12 UTC (Fig. 2). The surviving systems grow into intense storms and MCS by 14 UTC and propagate with the mean wind in a 45°-direction through the domain with lifetimes $\geq$ 6 h (Fig. 3). The horizontally contiguous rain areas extend between 30 and 200 km with a reflectivity larger than 30 dBZ. Surface fluxes of sensible and latent heat allow the cold pools to relax in the wake of the storms. The periodic BC allow the storms to grow in a way that is "natural" for the model physics in the given sounding. The time-window between 14 and 20 UTC is chosen for 3 hours of cycled assimilation until 17 UTC, followed by ensemble forecasts with 3 hours leadtime until 20 UTC.

Five random realizations of the initial noise in the BL are used as nature runs to cover the possible storm characteristics and positions.

b. Synthetic observations

Synthetic radar observations of reflectivity and radial windspeed are generated for each realization. The observations are computed on a regular grid with a horizontal spacing of 2 km and a vertical spacing of 1 km, covering the heights between 500 m and 12500 m, in order to mimic a region of good radar coverage. Reflectivity Z is computed from mixing ratios of graupel (QG), rain (QR) and snow (QS) using the simple implementation of Done et al. (2004). As in other OSSE studies (Tong and Xue 2005), Gaussian noise with a standard deviation of $\sigma_{refl} = 5$ dBZ is added to simulate errors of measurement and representativity. The reflectivity observations Z are masked to regions where $Z > 5$ dBZ.

Below this threshold, they are regarded as observations of no-reflectivity (Tong and Xue 2005; Aksoy et al. 2009) with a nominal value of 0 dBZ and $\sigma_{no-refl} = 2.5$ dBZ, assuming a good post-processing that can identify rain-free areas. The horizontal wind component U, with an added error of $\sigma_U = 1$ m s^{-1}, is used as an analogue for radial wind observations. This would be an accurate approximation if the storms are far away from the radar site. The regular observation geometry and the usage of U ensures that the data coverage is uniform and all storms are equally well observed.

c. *Initial ensemble*

The synthetic radar observations are assimilated using an ensemble of $k = 50$ members which differ from the nature run and among themselves only in the random distribution of the initial noise. The spin-up time between 06 and 14 UTC enables the members to contain storms with similar characteristics but completely uncorrelated horizontal positions. This approach was chosen to deprive the ensemble of any prior knowledge about the reference when the assimilation starts, as it could have been provided by a "manual" positioning of warm bubbles in the members or a confinement of the initial noise to regions of observed reflectivity (Tong and Xue 2005; Dowell and Wicker 2009).

d. *Implementation of the LETKF*

To produce an analysis ensemble, the LETKF-algorithm (for a full description see the original paper) determines the $\mathbf{w} = \bar{\mathbf{w}}^a$ that minimizes the cost function

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

where the k-dimensional vector $\mathbf{w}$ defines the optimal linear combination of ensemble member states that minimizes J^*. $\mathbf{y}^o$ contains the observations and $\mathbf{R}$ is the observation error covariance matrix. $\bar{\mathbf{y}}^b$ and $\mathbf{Y}^b\mathbf{w}$ are given by approximating the observation operator H to be linear about the m-dimensional background ensemble mean state $\bar{\mathbf{x}}^b$

$$H(\bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}) \approx \bar{\mathbf{y}}^b + \mathbf{Y}^b\mathbf{w} \tag{2}$$

where $\mathbf{X}^b$ is a $m \times k$ matrix whose columns are given by the deviations of the single forecast members from their mean $\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b$ and

$$\mathbf{Y}^b = H(\mathbf{x}^{b(i)}) - k^{-1}\sum_{i=1}^{k} H(\mathbf{x}^{b(i)}) \tag{3}$$

The analysis of (1) are computed locally for every analysis grid point. However, they do not necessarily need to be taken out on the full model resolution: the spatial field of the local $\mathbf{w}^{a(i)}$ is usually quite smooth, so a coarser "analysis grid" can be chosen horizontally and vertically on which the local analysis weights $\mathbf{w}^{a(i)}$ are first determined and then interpolated onto the model grid (Yang et al. 2009).

For the local analysis, only nearby observations are taken into account by localizing the observation error covariance matrix $\mathbf{R}$ with a Gaussian-like correlation function (Gaspari and Cohn 1999) that is zero where the distance r of the single observations is larger than the "cutoff-length" r_{Loc} of the localization radius. The dimension of the local $\mathbf{R}$ corresponds then to the number of local observations.

Since periodic boundary conditions have not been implemented in the LETKF, observations closer to the border of the domain than r_{Loc} are discarded. The untreated $\mathbf{R}$ contains the σ^2-variances mentioned in Section 2b and is diagonal in the present case because the added errors are uncorrelated.

$\mathbf{R}$ here is the only manipulable factor in (1) that determines how closely the analysis members are drawn towards the observations by the filter. As the localized assimilation of observations produced by nonlinear operators can lead to imbalances (Greybush et al. 2011), it appears sensible to inflate $\mathbf{R}$ to lessen the impact of the observations in (1), as will be shown later.

e. *Fine scheme R4*

The fine analysis scheme R4 (Table 1) is based on previous convective-scale data assimilation studies (Section 1a). It is devised as a control run to reproduce previous results and to serve as a benchmark for the performance of the coarse experiment later.

A horizontal localization length scale $l_{Loc,h} = 4$ km is chosen that corresponds to a Gaussian correlation function with $G(4\text{ km}) = \exp(-1/2) \approx 0.61$. A Gaussian-like correlation function C is used that goes to zero within the finite distance of the *localization radius* $r_{Loc,h} = 2l_{Loc,h}\sqrt{10/3} \approx 14.6$ km (Gaspari and Cohn 1999; Hamill et al. 2001). This $l_{Loc,h}$ is chosen for all observation types, as previous studies found it to be an effective compromise (Sobash and Stensrud 2013). The resolution of the synthetic observations is $\Delta x_{obs} = 2$ km, same as the model resolution. The vertical localization length scale, $l_{Loc,v}$, ranges from 3 km near the surface to 5 km at the model top. An assimilation interval Δt_{ass} of 5 minutes represents the typical availability of volume observations from a scanning Doppler radar (Lu and Xu 2009). The analysis grid here has the full horizontal resolution of $\Delta x_{ana} = 2$ km. The vertical resolution, provided by the 20 levels n_{lev}^{ana} of the analysis grid, is lower than the model resolution of 50 levels but should suffice, given the strong overlap due to the vertical $l_{Loc,v}$. The weights $\mathbf{w}^{a(i)}$ are multiplied by a constant covariance inflation factor of $\rho = 1.05$ in order to enhance the span of the analysis state space.

In this R4-setup, the positions of the analysis storms should coincide with the observed storms. Within the storm-cores, the analysis states are expected to be detailed with a low error and small variance, while spurious convection is suppressed outside of them by assimilating volume-observations of no-reflectivity. These are the requirements on a converged and detailed analysis ensemble formulated in the introduction.

As it turned out in preliminary experiments, providing the filter with the untreated $\mathbf{R}$ can result in imbalanced increments. These taint the solution by introducing gravity wave noise that arises from dynamically inconsistent model states when the members are drawn too closely to the observations. It was therefore chosen to inflate the entries of $\mathbf{R}$ from

($\sigma_U = 1$ m/s)^2 to ($\sigma_U = 5$ m/s)^2, from ($\sigma_{refl} = 5$ dBZ)^2 to ($\sigma_{refl} = 20$ dBZ)^2 and from ($\sigma_{no-refl} = 2.5$ dBZ)^2 to ($\sigma_{no-refl} = 20$ dBZ)^2 (cf. Table 1). Lessening the influence of the observations in (1) by this method resulted in more consistent and stable analyses as with the untreated $\mathbf{R}$.

The deterioriating effect appeared to be most critical with the reflectivity observations that are computed by a very nonlinear operator whose increments in model space are nonetheless approximated to be linear by (2). To test this hypothesis, R4_forced was devised as a sensitivity experiment wherein $\mathbf{R}$ contains the reflectivity error values with which the synthetic observations were originally perturbed: ($\sigma_{refl} = 5$ dBZ)^2 and ($\sigma_{no-refl} = 2.5$ dBZ)^2 (cf. Table 1).

f. Coarse scheme R16

In the coarse scheme R16, more horizontal variance should be allowed, so a larger $l_{Loc,h} =$ 16 km is chosen and the observational resolution Δx_{obs} is decreased from 2 km to 8 km. At the same time, the resolution of the analysis grid Δx_{ana} is decreased from 2 km to 8 km, so there are 4^2 times fewer local analysis vectors $\mathbf{w}^{a(i)}$ to be computed by the local cost functions (1), reducing the computational effort and smoothing the analysis field further. Consequently, for every local analysis the same number of observations is used in R16 as in – but now on a coarser scale.

The observations are coarse-grained by horizontally averaging the values and positions of 4×4 blocks of the original observations into one central superobservation (SO); this is preferable to data-thinning and reduces the information to the desired coarse scale (Alpert and Kumar 2007; Salonen et al. 2009; Seko et al. 2004).

The analysis increments should therefore consist of larger parts of storms that are more dynamically consistent internally than in R4, as the analysis weights of the different members will vary less between adjacent model grid points and the influence radius of the observations is larger. In one single analysis, a whole storm-core with up- and downdraft could be drawn from the background ensemble in R16, whereas in R4 the updraft-region of an analysis storm might originate from a different background member than the downdraft region.

The fact that the coarse and large-scale scheme R16 has the same number of observations per local analysis as the fine scheme R4 suggests that the absolute influence of the observations on the local solution of (1) should also be the same. As this is regulated by the magnitude of the local $\mathbf{R}^{-1}$-entries, R16 uses an inflated $\mathbf{R}$-matrix that contains the error same values as in R4 (cf. Table 1), but now for the SOs.

This reasoning is tested by the sensitivity-experiment R16_forced which uses the error values for the SOs which have been effectively lowered by the averanging: The original observations of R4 are a normally distributed random variable (cf. Section 2b), so the block-averaging in R16 should reduce the error standard deviation of the SO by a factor of $\sqrt{1/4^2} = 1/4$ to multipy $\mathbf{R}$ with. If this is done, the observational part of the local cost function (1) becomes larger, because the local analysis has the same number of observations but now with a larger $\mathbf{R}^{-1}$. This rigorous formulation is tested in R16_forced whose $\mathbf{R}$-entries are derived from the control experiment R4. The entries of $\mathbf{R}_{R16\_forced}$ with SOs are $(\sigma_{U,SO} = 1.25\text{m/s})^2$, $(\sigma_{refl,SO} = 5\ \text{dBZ})^2$ and $(\sigma_{no_refl,SO} = 5\ \text{dBZ})^2$ (cf. Table 1).

The the inflated $\mathbf{R}$-matrix of R16 is intended to allow a not-converged ensemble wherein even the occurence of an observed storm, given by observations of U and reflectivity, should have a variance; also, spurious convection should be permitted by lessening the suppressive influence of no-reflectivity observations. These are the requirements on a non-converged and coarse analysis ensemble formulated in Section 1.

The observations in R16 are assimilated using a longer cycling interval Δt_{ass} of 20 minutes between subsequent analyses. In sensitivity experiments for R16 with $\Delta t_{ass} = 5$ minutes, it was found that the inevitable noise of the increments was introduced too frequently and deteriorated the analysis. Also, significant large-scale differences in the members storm structures could not spin up during $\Delta t_{ass} = 5$, rendering the analyses worse than with $\Delta t_{ass} = 20$ min. The resulting noisy structures were similar to the disturbances of R16_forced that are shown later, albeit with lower amplitude.

Taking advantage of the reduced memory requirements of the coarser horizontal grid, $n_{lev}^{ana} = 25$ had been chosen for R16, giving a modest improvement over $n_{lev}^{ana} = 20$. The slight advantage over R4 with respect to the additional vertical levels is assumed to be offset by the larger number of discarded observations in the wider border regions of $l_{Loc,h} = 16$ km, so R16 and R4 should generally be comparable.

g. RMSE, spread and consistency ratio

The accuracy of analysis and forecast states $\mathbf{x}$ is measured by the Root Mean Square Error (RMSE), computed in model space for the different model variables:

$$RMSE(\mathbf{x}) = \sqrt{m^{-1}\sum_{l=1}^{m}(x_l^{nature} - x_l)^2} \tag{4}$$

where m is the number of grid points. The corresponding variance is the spread $\overline{spr}$ of the ensemble $\mathbf{x}^i$ around its mean $\bar{\mathbf{x}} = k^{-1}\sum_{i=1}^{k}\mathbf{x}^i$, given by

$$\overline{spr} = m^{-1}\sum_{l=1}^{m}\sqrt{(k-1)^{-1}\sum_{i=1}^{k}(x_l^i - \bar{x}_l)^2} \tag{5}$$

To fulfill the Gaussian assumption of the filter, the ensemble spread should represent the actual error of the analysis, so the consistency ratio CR

$$CR = \frac{\overline{spr}}{RMSE} \tag{6}$$

should be $CR = 1$.

Unlike studies that only regarded convectively active regions above a certain reflectivity threshold, the present study evaluates RMSE and spread for all grid points of the domain. This results a) in a generally low error level because many small error values of clear air regions contribute to the mean error and b) in analysis errors which should be regarded relatively to the free error level of an ensemble running in parallel without assimilation, as the overall variance of the convective situation may change during the diurnal cycle.

3. Assimilation results at fine and coarse resolution

First the fine experiment R4 is evaluated during the assimilation window 14-17 UTC in Section 3a, then the results of R16 are presented for comparison in Section 3b. An example of the results for one realization of the nature run is given in Fig. 4, which compares snapshots of nature run 01 to the analysis means of R4 and R16 at 17 UTC. The maximum column reflectivity is the vertical 2D-projection of the 3D-reflectivity field. T at the height of $z = 150$ m shows the cold pools, W at $z = 3500$ m shows regions of up- and downdrafts. Fig. 5 shows the RMSE and spread of the ensemble means during the assimilation window and during the forecast window from 17 to 20 UTC for the model variables U, W, T and QR. Analogously, Fig. 6 shows the mean RMSE of the single members. The dash-dotted lines in the RMSE-Figures provide an error level by showing the error of a free-running ensemble that has not undergone any assimilation.

a. Performance of R4

The fine scheme R4 is able to converge the ensemble mean onto the observed storms for the directly observed precipitation variables contained in the reflectivity observations, so the reflectivity field of the R4-mean closely resembles the nature run by 17 UTC (Fig. 4). The RMSE of the rain mixing ratio QR is strongly reduced in the analysis mean of R4 (Fig. 5) and exhibits a very low spread: the graupel mixing ratio (QG) behaves analogously (not shown). This points at a strong convergence of the ensemble members onto the observed clouds. Fig. 7 compares the reflectivity of some representative analysis-members of R4 to the nature run. All member-storms look similar to the nature run and spurious convection is largely suppressed by observations of no-reflectivity. In Fig. 9, the analysis ensemble distributions are evaluated at selected points of updrafts where the nature run shows a vertical velocity of +10 m s^{-1} (a) or a reflectivity of 40 dBZ (b) at 17 UTC. These are locations where observations are favorably present: the reflectivity is observed directly, the vertical velocity of the analysis is provided through background covariances where reflectivity observations and horizontal convergence observations (U) indicate an updraft. The analysis member states of R4 have converged closely around these values with a distribution that exhibits a Gaussian shape.

Judging by Figs. 4, 7 and 9, the cycled analyses produce up- and downdrafts – although only provided by background error covariances – in the right positions, shapes and amplitudes (Fig. 4); the RMSE of W is clearly reduced for the R4-mean relative to the free-running ensemble, and a good consistency ratio is reached by maintaining ensemble spread (Fig. 5).

The RMSE of the horizontally smooth T-field is not reduced as much relative to the free error level as for the horizontally intermittent variables W and QR. The analysis increments even appear to be largely detrimental in the case of R4 (showing as a reversed sawtooth pattern), especially when the single members are considered (Fig. 6). Fig. 4 also shows that the shapes and positions of the analysis cold pools are in accordance with the nature run but exhibit a cold bias. This bias is assumed to be caused by the covariance localization: the mean temperature is not conserved when increments add local linear combinations of members, (Janjić et al. 2012, 2013) possibly resulting in convective systems that are generally too wet and therefore develop too strong cold pools due to evaporative cooling. The deficient T-analysis resulting from EnKF-assimilation of radar data has been noticed in previous studies (Zhang et al. 2004; Dong et al. 2011). The previous studies found that temperature observations from a dense surface mesonet can reduce the problem of too strong cold pools, but this method is not considered here.

Observations of the horizontally smooth U-field are available where reflectivity is present. The filter acts beneficially on U for R4 and reduces the RMSE relative to the free error level and also absolutely. As with T, the consistency ratio of U exceeds 1 by 17 UTC, indicating that the chosen covariance inflation factor $\rho = 1.05$ might be too large for these variables. The V-field is closely correlated to U by the dynamics of the storms and the 45° background wind field and behaves analogously to U in terms of errors and spread.

Fig. 10 compares R4 to the sensitivity experiment R4_forced which has an uninflated $\mathbf{R}$-matrix for (no-)reflectivity observations. For all variables, the errors are larger in R4_forced, and the detrimental effect on T shows even more clearly. In (2) it was assumed that a deviation from the mean in the observation space is directly proportional to the attributed deviation in the model space. The inflated $\mathbf{R}$ in R4 allows a greater weight for the background ensemble in (1) and thus does not follow the assumption of (2) as rigorously which then acts beneficially on the quality of the analysis.

Summarizing these results, the R4-scheme appears to produce analyses that are comparable to previous convective EnKF studies with radar data. The covariance localization in observation space of the LETKF-algorithm appears to work well, compared to the localization in model space in the EnSRF-filters of other studies. The storms in the analysis ensemble resemble the storms of the nature run closely in terms of position, amplitude and dynamics. *The mean of the strongly converged ensemble is representative for the best solution, the single analysis members show a strong agreement with little variance inside the observed storms.* These were the requirements for a fine and converged scheme formulated in the introduction. R4 is regarded to fulfill them.

b. Performance of R16

Now the assimilation results of the coarse scheme R16 are evaluated. With the horizontally smooth variable U, Fig. 5 exhibits a similar error for the R16-mean as for the R4-mean; the slight advantage of R4 for U appears even less significant if the mean of the member-RMSEs in Fig. 6 is considered where no difference between the two schemes is apparent. The analysis quality of T seems to be better for R16 – the analysis increments are always beneficial here (Fig. 6). R16 assembles the analysis storms from horizontally larger portions of member-storms, apparently reducing the dynamical inconsistencies in smooth variables U and T that are introduced by the localization. Nevertheless, also R16 exhibits a cold bias in the ensemble mean, especially due to additional cold pools of spurious storms.

For the horizontally intermittent rain-variables and thus for the reflectivity, the analysis quality of R16 is inferior to R4. In Fig. 4, the counterparts of the observed storms in the R16-mean have roughly the right position and shape but are blurred, caused by variability of the updraft core positions within the ensemble (Fig. 8). This was expected due to the combination of the SOs that provide information only on a coarser scale, a larger $l_{Loc,h}$ and coarser analysis grid that facilitate horizontally smoother increments, and the inflated $\mathbf{R}$ that allows a higher level of variance. The resulting R16-ensemble is not as converged as the R4-ensemble: its distribution is broader, less Gaussian and even multimodal in locations of updrafts and strong reflectivity in the nature run (Fig. 9). Spurious storms exist largely and show up in the reflectivity fields of the ensemble members (Fig. 8) and the mean (Fig. 4). The RMSE-values of the variables related to vertical motion (W and QR) are larger for R16, lying just slightly below the free error level (Fig. 6), but now possess significant spread in QR (Fig. 5).

The introduction of the coarse scheme required a larger variance for the storm positions than in the fine scheme, together with larger analysis error and more spread, allowing also spurious convection. These requirements are regarded to be fulfilled by the coarse scheme R16.

Fig. 11 compares a typical storm system of one nature run to the corresponding analysis means of R16 and the sensitivity experiment R16_forced, the latter with an uninflated $\mathbf{R}$-matrix for (no-)reflectivity observations. In contrast to R16, the mean W-field of R16_forced reveals single updraft cores with large amplitudes where all members agree on their positions, recognizable also in the larger amplitude of the mean reflectivity field: *R16_forced corresponds to a stringently converged version of R16, but unlike to R4, the converged mean does not resemble the physical properties of the nature run.* The strong and imbalanced increments of R16_forced cause gravity waves that show in the W-field where noise is present in the analysis mean. Strong updrafts lie close to downdrafts in an alternating pattern which appears to be dynamically inconsistent with the nature run.

The effective sample size in the R16-experiments is lower than in the R4-experiments due to the larger localization radius (Hunt et al. 2007). Producing a converged ensemble with horizontally broad increments as in R16_forced would possibly need a larger ensemble size – with the given 50 members, the non-converged ensemble of R16 turned out to work better.

4. Ensemble forecasts from fine and coarse resolution analyses

From the last analysis ensembles of the experiments at 17 UTC, ensemble forecasts are performed until 20 UTC.

a. RMSE forecast error

For W and QR, the RMSE in R4 converges to the error level of R16 within 1-2 hours, both for the ensemble mean (Fig. 5) and the members (Fig. 6). Within this period, R4 quickly regains spread for QR, caused by the precipitating cores of the member storms that are moving apart from the well-determined position of the converged analysis. At 20 UTC, the degree of horizontal blurring in R4 equals that of R16 (Fig. 12).

For U and T, the RMSE of the forecasts grows more slowly with no significant difference between R4 and R16. Qualitative differences of the analysis perturbations in R4 and R16 with respect to the reference run are thus contained mostly in the horizontally intermittent variables such as W and QR.

b. Field-based evaluation by the DAS-score

Anticipating that much of the error in the R4 and R16 ensemble mean forecast fields (Fig. 12), two spatial verification measures are chosen to compare the forecast fields to the nature run. Fields of maximum column reflectivity are chosen for this evaluation because high reflectivity is usually accompanied with strong precipitation and winds – these two are the essential threats to be predicted by a convective storm forecast. The observation field of the nature run here is thresholded above 10 dBZ to separate the storms and eventually overlapping anvils.

The Displacement and Amplitude Score DAS (Keil and Craig 2009) uses a pyramidal matching algorithm to compare two fields by an optical flow technique. A vector field is computed that morphs the forecast onto the reference field and vice versa, using a maximum search radius of 45 km here. The average magnitude of the displacement vector field, normalized by maximum search radius, defines the DIS-component of the DAS-score, and is displayed in Fig. 13. During the assimilation window, R4 is clearly superior to R16 due to the converged ensemble. This advantage is lost during the forecast window within 1 hour – after that, the forecast quality of storm positions is indistinguishable between the fine and coarse schemes.

c. *Object-based evaluation by the SAL-score*

The SAL-score (Wernli et al. 2008) compares statistical properties of thresholded rain-objects. The structure or S-component indicates if the forecast objects are smoother and broader than the observations ($S > 0$) or spikier ($S < 0$), with $S = 0$ indicating the correct structure. Only average object properties are compared, without matching, so S is independent of location errors and biases. In Fig. 14a, R4 and R16 appear qualitatively equal with $S \approx 0$ at 17 UTC, meaning that the right type of objects are present in the members after the assimilation cycling; the observed convective modes are reproduced, as far as the projection of the column is concerned. This property holds during the forecast for the members (dashed lines), but for the ensemble mean (solid lines), the blurring of forecast features is apparent with an increasing $S > 0$.

The amplitude or A-component (Fig. 14b) compares the domain-wide total reflectivity between forecast and observation and thus displays the overall bias. During the assimilation window, R4 and R16 behave differently. In the first hour of R4, most spurious storms are subsequently suppressed while the inserted storms have not yet fully developed, resulting in a negative bias with $A < 0$, that then overcompensates to a positive bias with $A > 0$ by 17 UTC. In R16, the analysis increments always increase the bias by inserting observed storm-increments without complete suppression of spurious storms, counteracting the relaxing tendency of the dynamics during the forecast intervals. During the first 90 minutes of the forecast window, the A-bias of R16 decreases, while small and incompletely suppressed spurious storms in R4 grow quickly and enhance the R4-bias, but eventually converge with R16. After 2 hours of forecast, R4 and R16 are indistinguishable.

The L-component (Fig. 14c) determines the location error by measuring the horizontal deviation of the centroid of the forecast reflectivity field from the centroid of the observations. The results here are similar to those of the DIS-score: R4 has an advantage over R16 by the end of the analysis period, but this is lost within 1-2 hours of forecast time.

The curves of the location error L are less continuous than the DIS-score and sometimes even decrease; this is a consequence of the small sample five random realizations that are evaluated here. Comparing the L-error between the experiments and the free error level here is more meaningful than the absolute L-values.

The results of DAS-DIS and the SAL score were verified by evaluating the Brier Score (Wilks 2006) of the forecast probabilities of R4 and R16 with respect to convective events exceeding a threshold of 10 dBZ (Fig. 15). A similar convergence in skill can be seen, and at 20 UTC there is no significant difference of skill between R4 and R16 in the sample of the five repetitions of the experiments.

5. Summary and Discussion

The aim of this study on convective scale data assimilation was to assess the benefits of high-resolution analysis schemes in the provision of initial conditions for 3-hour precipitation forecasts, taking into account the limited predictability of convective systems due to scale-dependent error growth.

Data Assimilation of long-lived and organized convective systems was performed under the assumption of a perfect model, using a LETKF in an idealized testbed with simulated Doppler radar observations of reflectivity and radial wind. The storms were triggered randomly in the convection-permitting COSMO-model using radiative forcing and initial small-amplitude random noise in a horizontally homogeneous environment with periodic boundary conditions.

Updraft positions, storm intensities and cold pool structures were determined well by the cycled analyses of the high-resolution assimilation scheme R4; this appeared as a succesful reproduction of previous studies on this topic. The usability of the LETKF-algorithm, which localizes in observation space, for detailed storm-analyses with radar data is therefore demonstrated. Reducing the influence of observations on the analysis increments by an inflated error covariance matrix $\mathbf{R}$ proved useful for obtaining dynamically consistent model states.

The coarse scheme R16 was devised to produce less precise analyses than R4 by increasing the horizontal localization length scale from 4 to 16 km, by reducing the resolution of observations and of the analysis grid from 2 to 8 km and by inflating $\mathbf{R}$ further; by these means, variance was introduced into the ensemble with respect to the position and occurence of updrafts. Spurious convection was permitted in multimodal posterior distributions.

Both schemes showed a cold bias due to the introduction of localized analysis increments that were too wet and caused excessive cooling in low levels.

The 3-hour period of cycled assimilation of the two schemes was followed by ensemble forecasts with a lead time of 3 hours. During the first hour, precipitation forecasts from analyses of the fine scheme were clearly advantageous to analyses of the coarse scheme in terms of the storm positions and internal structure, measured by RMSE and the object based scores DAS and SAL. This advantage of the fine analyses was lost within the first 1-2 hours of forecast by the rapid error growth of the small perturbations.

a. *Conclusions*

These results suggest that, due to the limited predictability, convective forecasts with typical leadtimes of 3 hours might not benefit from storm analyses that are detailed on the scale of the model grid; a coarser representation of storm positions and occurence as in the coarse scheme might be sufficient. The computational acceleration hereby is notable.

Nonetheless, a fine analysis with low errors is superior for the first hour and could be useful for forecasts of phenomena smaller than the storm system such as tornados and gust fronts.

b. *Discussion*

Looking back at the methods, the coarse scheme R16 is based on the fine scheme R4 which reproduced previous results from other studies. The means by which the coarsening of the analysis was achieved in R16, namely the larger localization together with a less detailed analysis computation and coarsened observations, were only useful in combination with an inflated **R**-matrix to relax the Gaussian constraint of the cost function (shown by R16_forced compared to R16). The **R**-inflation was chosen empirically to let the system produce reasonable analyses. More sophisticated methods exist for the estimation of appropriate observation errors for **R** (Desroziers et al. 2005) which are also implemented in the KENDA-system but were not applied here. Evaluating the effects of the coarsening methods of R16 separately could give much insight but would be beyond the scope of this paper.

The cold bias of the analyses hints at the substantial problem that mass and energy is not conserved in analyses produced by a localized EnKF (Janjić et al. 2012, 2013). Covariance localization in observation space as done here shows no substantial differences to localization in model space.

The forecast results displayed the limited predictability of the dynamics in large convective systems. In an operational model, the predictability will be enhanced by effects of synoptic and orographic forcing (Hohenegger and Schär 2007) that might overlay the efforts of a fine analysis; in that case, the larger analysis errors of a coarse scheme might be even less disadavantageous than in the idealized situation presented in this study. Effects of model error could further diminish the benefits of the fine scheme, especially if it tries to converge the model states into convective modes that are not supported by the model physics and the predicted sounding (Stensrud and Gao 2010) – here the gentler approach of the coarse scheme could be profitable for dynamic consistence.

The spatial and temporal quantities of dynamics and predictability that are discussed here depend on the model resolution, the model physics and the type of long-lived storm that is simulated. A resolution of 2 km is not sufficient to simulate storm-internal variance on the scale of single plumes which come addressable with resolutions of 250 m and less (Bryan and Morisson 2012). Using such a model that is able to resolve three-dimensional turbulence, an even finer assimilation scheme could be applied to further investigate the limits of predictability.

In the case of a different sounding and the resulting different convective modes such as mesocyclones, multicell storms and linear squall lines, the predictability limit will be influenced by the degree of storm-internal organization (Aksoy et al. 2010). The quantitative time limit found in this study can thus not be generalized for all kinds of convection.

The lagged detection of storm cells by radar observations remains a problem. Further studies could investigate if an ensemble that contains spurious storms might be advantageous when an assimilation system has to catch storms that develop in locations where clear air was observed and assimilated before: it may be harmful to the dynamics to select single precipitating forecast members that have to form large increments for the analysis mean (Lien et al. 2013).

Assimilating reflectivity observations is still difficult due to their non-Gaussian distributions, making the suppression of spurious convection insufficient because (a) the background distribution of convection is typically intermittent and therefore non-Gaussian and (b) there is no “negative rain” to assimilate in unaltered observations (Craig and Würsch 2012). A Gaussian transformation of precipitation observations could help here to fulfill the Gaussian assumptions of the EnKF (Lien et al. 2013).

Acknowledgments.

The authors thank Daniel Leuenberger for providing the template of the COSMO setup, Ulrich Blahak for the support with the idealized COSMO code and Hendrik Reich and Andreas Rhodin for providing code and support of the KENDA-LETKF.

This study was carried out in the Hans Ertel Centre for Weather Research. This research network of universities, research institutes, and the Deutscher Wetterdienst is funded by the BMVBS (Federal Ministry of Transport, Building and Urban Development).

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 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.

Aksoy, A., D. C. Dowell, and C. Snyder, 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.

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

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

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

Bischof, M., 2011: Ensemble Simulations of Convective Storms. M.S. thesis, Institute for Atmospheric and Climatic Sciences (IACETH) at Swiss Federal Institute of Technology Zürich.

Bryan, G. H. and H. Morisson, 2012: Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics. *Monthly Weather Review*, **140**, 202–225.

Craig, G. C., C. Keil, and D. Leuenberger, 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.

Craig, G. C. and M. Würsch, 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*.

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

Dawson, D. T., L. J. Wicker, and E. R. Mansell, 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.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3385–3396.

Done, J. M., G. C. Craig, S. Gray, and P. Clark, 2012: Case-to-case variability of predictability of deep convection in a mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, **138**, 638–648.

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

Dong, J., M. Xue, and K. K. Droegemeier, 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.

Dowell, D. C. and L. J. Wicker, 2009: Additive Noise for Storm-Scale Ensemble Data Assimilation. *Monthly Weather Review*, **132**, 1982–2005.

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

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**, 10 143–10 162.

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

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

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, **129**, 2776–2790.

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

Houtekamer, P. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811.

Houtekamer, P. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **129**, 123–137.

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

Janjić, T., D. McLaughlin, S. E. Cohn, and M.Verlaan, 2013: Conservation of mass and preservation of positivity with ensemble-type kalman filter algorithms. *Monthly Weather Review*, accepted.

Janjić, T., D. B. McLaughlin, and S. E. Cohn, 2012: Preservation of physical properties with ensemble based kalman filter algorithms. *Oberwolfach Reports*, **9**, 3414–3471.

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

Keil, C. and G. C. Craig, 2009: A Displacement and Amplitude Score Employing an Optical Flow Technique. *Weather and Forecasting*, **24**, 1297–1308.

Kober, K., G. C. Craig, C. Keil, and A. Dörnbrack, 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.

Kuhl, D., et al., 2007: Assessing Predictability with a Local Ensemble Kalman Filter. *Journal of the Atmospheric Sciences*, **64**, 1116–1140.

Lien, G.-Y., E. Kalnay, and T. Miyoshi, 2013: Effictive assimilation of global precipitation: Simulation experiments. *Tellus A*, doi:10.3402/tellusa.v65i0.19915.

Lilly, D. K., 1990: Numerical prediction of thunderstorms – has its time come? *Monthly Weather Review*, **116**, 779–798.

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

Lu, H. and Q. Xu, 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.

Reich, H., A. Rhodin, and C. Schraff, 2011: LETKF for the nonhydrostatic regional model COSMO-DE. Newsletter 11, Consortium for Small-scale Modelling, 27–31 pp. URL `http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter11`.

Salonen, K., H. Järvinen, G. Haase, S. Niemelä, and R. Eresma, 2009: Doppler radar radial winds in HIRLAM. Part II: optimizing the super-observation processing. *Tellus A*, **61**, 288–295.

Seko, H., T. Kawabata, and T. Tsuyuki, 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.

Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review*, **132**, 3019–3032.

Snyder, C. and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. *Monthly Weather Review*, **131**, 1663–1677.

Sobash, R. A. and D. J. Stensrud, 2013: The impact of covariance localization for radar data on enkf analyses of a developing mcs: Observing system simulation experiments. *Mon. Wea. Rev.*, –, doi:10.1175/MWR-D-12-00203.1.

Stensrud, D. J. and J. Gao, 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.

Stensrud, D. J., et al., 2009: Convective-Scale Warn-on-Forecast System - A Vision for 2020. *Bulletin of the American Meteorological Society*, **90**, 1487–1499.

Tong, M. and M. Xue, 2005: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Monthly Weather Review*, **133**, 1789–1807.

Wang, S., M. Xue, A. D. Schenkman, and J. Min, 2012: An Iterative Ensemble Square Root Filter and Tests with Simulated Radar Data for Storm Scale Data Assimilation, submitted to QJR.

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

Wilks, D., 2006: *Statistical methods in the Atmospheric Sciences*. Academic Press, San Diego, London.

Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the local ensemble transform kalman filter. *Q.J.R. Meteorol. Soc.*, **135 (638)**, 251–262, doi:10.1002/qj.353.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of Moist Convection on Mesoscale Predictability. *Journal of the Atmospheric Sciencies*, **60**, 1173–1185.

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

List of Tables

1 Experiments as described in Section 2. R4_f and R16_f are R4_forced and R16_forced. $l_{Loc,h}$ is the horizontal localization lengthscale, $r_{Loc,h}$ 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-interval between the two consequent analyses, σ_U, σ_{refl} and $\sigma_{no-refl}$ are the standard deviations of the observations (for R16 and R16_f: SOs) that are contained in the **R** for the respective experiment. Some **R**-entries are inflated, see text. 33

Table 1. Experiments as described in Section 2. R4_f and R16_f are R4_forced and R16_forced. $l_{Loc,h}$ is the horizontal localization lengthscale, $r_{Loc,h}$ 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-interval between the two consequent analyses, σ_U, σ_{refl} and $\sigma_{no-refl}$ are the standard deviations of the observations (for R16 and R16_f: SOs) that are contained in the **R** for the respective experiment. Some **R**-entries are inflated, see text.

	R4	**R16**	**R4_f**	**R16_f**	
$l_{Loc,h}$	4	16	4	16	km
$r_{Loc,h}$	14.6	58	14.6	58	km
$\Delta x_{obs/ana}$	2	8	2	8	km
n_{lev}^{ana}	20	25	20	25	
Δt_{ass}	5	20	5	20	min
σ_U	5	5	5	1.25	m s^{-1}
σ_{refl}	20	20	5	5	dBZ
$\sigma_{no-refl}$	20	20	2.5	5	dBZ

List of Figures

1 Relative frequencies of predicted reflectivities for a relatively good background ensemble (gray) and two analysis-ensembles a) and b) (black), computed by by the LETKF-algorithm. The observed value is 50 dBZ. In a) a small observation error of $\sigma = 5$ dBZ is reported to the filter via $\mathbf{R}$, in b) an inflated value of $\sigma = 20$ dBZ. The dashed curve is the assumed normal distribution around the observation. 36

2 Time-series of the nature run (averaged over five realizations): Domain-average of the maximum column reflectivity in dBZ (solid), together with average horizontal (dash-dotted) and maximum (dashed) size of the rain-objects, thresholded to > 5 dBZ. Assimilation window is between 14 and 17 UTC (shaded gray), forecast window is between 17 and 20 UTC (shaded pink). Peaks of the object-sizes can be due to merger of anvils of separate convective systems. 37

3 Time series of maximum column reflectivity of nature run 01. Note that the decline in average reflectivity from 17 to 20 UTC is not representative for the five repitions (cf. Fig. 2) 38

4 Nature Run and Analysis Means of R4, and R16 (both of repetition R01) at 17 UTC. Top rows: Maximum reflectivity. Middle rows: Temperature T at $z = 150$ m. Bottom rows: Vertical velocity W at $z = 3500$ m. 39

5 RMSE and spread of the ensemble means of R4 (black) and R16 (gray) during assimilation (14-17 UTC, forecast ensemble mean and analysis ensemble mean as one consecutive line) and ensemble forecast (17-20 UTC, 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 one empirical standard deviation is depicted by the shaded areas. 40

6 Like Fig. 5, but for the mean of the RMSE of the single ensemble members of R4 (black) and R16 (gray) instead of the RMSE of the ensemble-mean states. 41

7 Maximum reflectivity ($REFL\_MAX$) of nature run and analysis ensemble members 1,13,25,37,50 of R4 (Realization R01) at the last assimilation time 17:00. 42

8 Like Fig. 7, for R16. 43

9 Relative frequencies of model values within the ensemble-members of R4 (black bars) and R16 (gray bars). Regions inside storms are given by a) and b): a) at gridpoints where W of the nature run is 10 ± 0.5 m s^{-1}, b) where $REFL_{nature} = 40 \pm 0.5$ dBZ. Regions closer than 64 km to the lateral boundaries are excluded so the histograms are computed only where the filter is fully active in R4 and R16. The bar width is 1/2 of the binwidth. 44

10 Like Fig. 5, but comparing R4 (black) to R4_forced (gray) for the primary variables U, W, T and QR in the first two hours of assimilation. 45

11 Like Fig. 4, but comparing the analysis means of R16 and R16_forced to the nature run at 16 UTC during the assimilation (Realization 03). 46

12 Like Fig. 4, but forecast ensemble means (Realization 01) at 20 UTC. 47

13 DIS-component of the DAS-score of the maximum column reflectivity $REFL\_MAX$ $>$ 10 dBZ. Displayed is the mean DIS-score of the ensemble members of R4 (black), R16 (gray). The error lines are the average over five experimental repetitions. An envelope of the empirical standard deviation is given by the shaded areas. A value of DIS = 0 is a perfect match. 48

14 SAL score components S, A and L (see text) of $REFL\_MAX$ $>$ 10 dBZ, averaged over the five repetitions. 49

15 Brier Score of forecast ensembles R4 and R16 at locations where $REFL\_MAX$ $>$ 10 dBZ in the nature run. Errorbars show the empirical standard deviation of the five experimental repetitions. 50

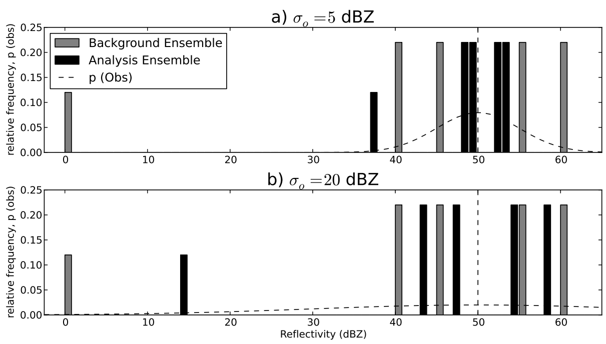

FIG. 1. Relative frequencies of predicted reflectivities for a relatively good background ensemble (gray) and two analysis-ensembles a) and b) (black), computed by by the LETKF-algorithm. The observed value is 50 dBZ. In a) a small observation error of $\sigma = 5$ dBZ is reported to the filter via $\mathbf{R}$, in b) an inflated value of $\sigma = 20$ dBZ. The dashed curve is the assumed normal distribution around the observation.

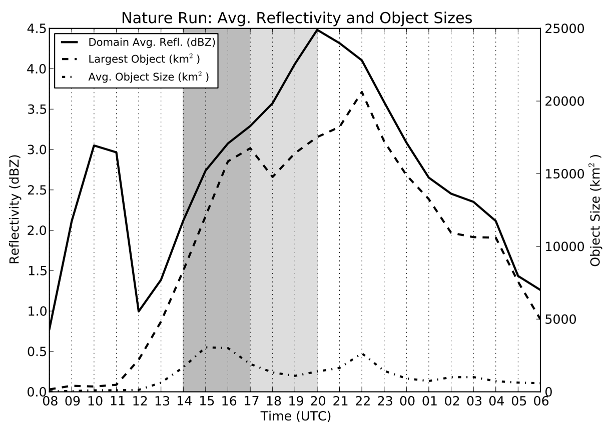

FIG. 2. Time-series of the nature run (averaged over five realizations): Domain-average of the maximum column reflectivity in dBZ (solid), together with average horizontal (dash-dotted) and maximum (dashed) size of the rain-objects, thresholded to > 5 dBZ. Assimilation window is between 14 and 17 UTC (shaded gray), forecast window is between 17 and 20 UTC (shaded pink). Peaks of the object-sizes can be due to merger of anvils of separate convective systems.

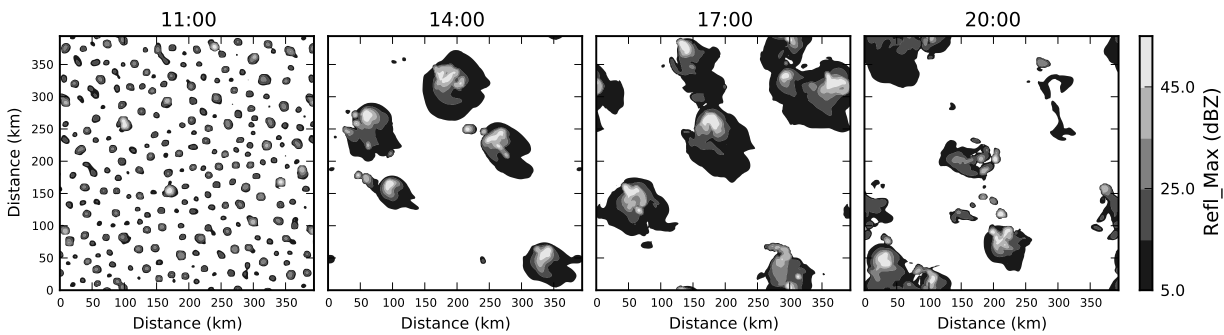

FIG. 3. Time series of maximum column reflectivity of nature run 01. Note that the decline in average reflectivity from 17 to 20 UTC is not representative for the five repitions (cf. Fig. 2)

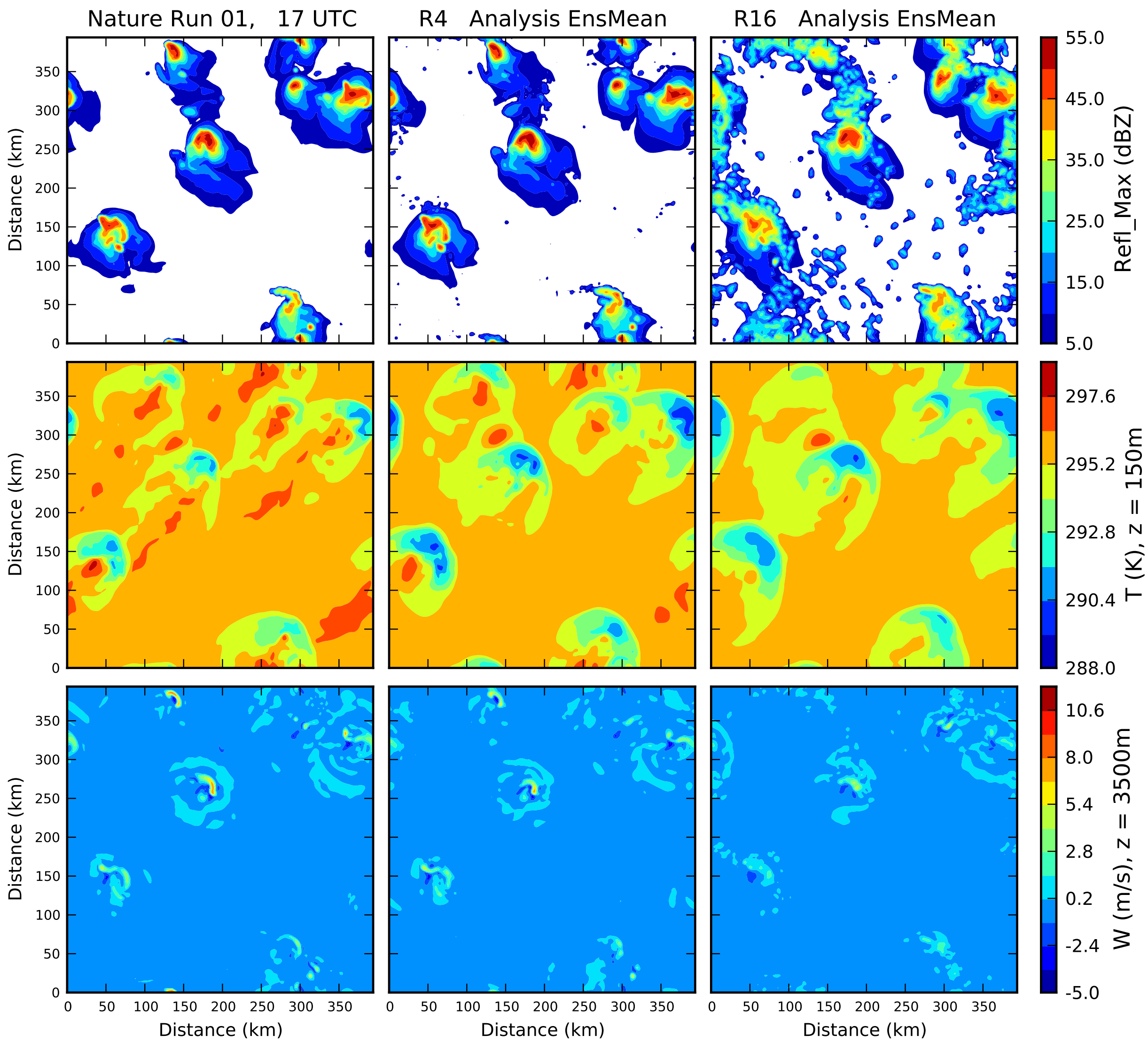

FIG. 4. Nature Run and Analysis Means of R4, and R16 (both of repetition R01) at 17 UTC. Top rows: Maximum reflectivity. Middle rows: Temperature T at $z = 150$ m. Bottom rows: Vertical velocity W at $z = 3500$ m.

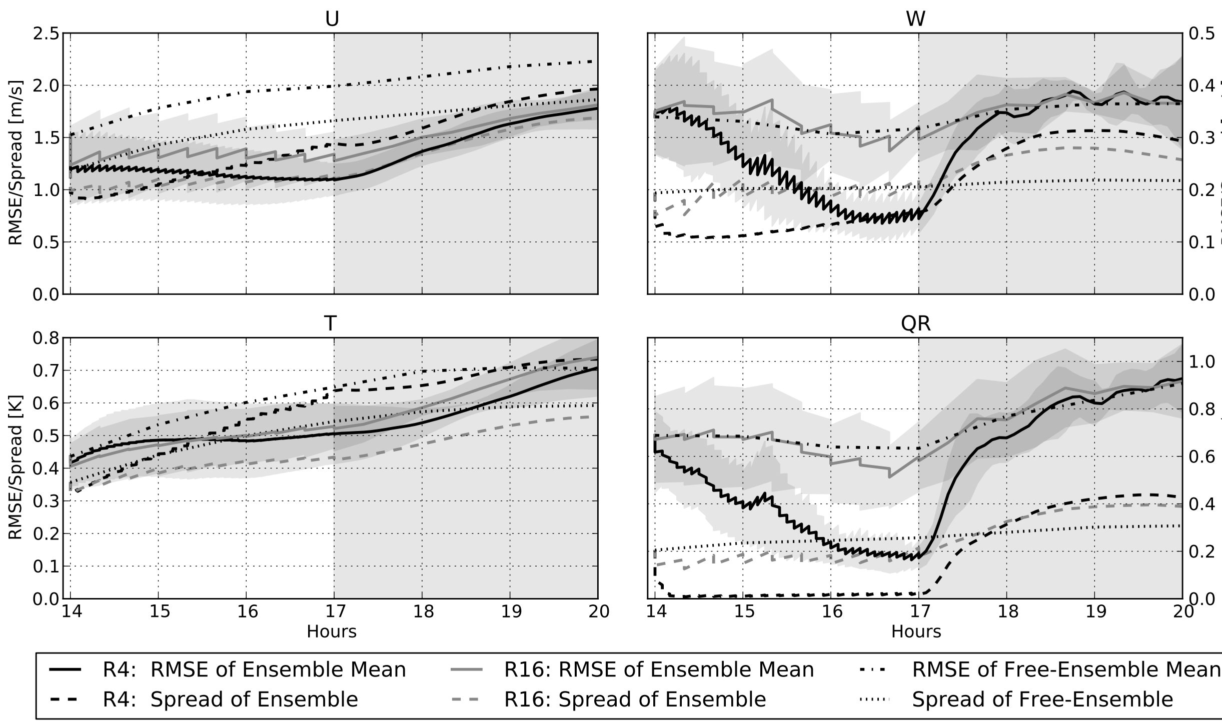

Fig. 5. RMSE and spread of the ensemble means of R4 (black) and R16 (gray) during assimilation (14-17 UTC, forecast ensemble mean and analysis ensemble mean as one consecutive line) and ensemble forecast (17-20 UTC, only forecast ensemble mean). The gray lines show the RMSE and spread of the mean of the free ensemble without any assimilation. All gridpoints are evaluated. The error values are averaged over five realizations of the experiment; an envelope of one empirical standard deviation is depicted by the shaded areas.

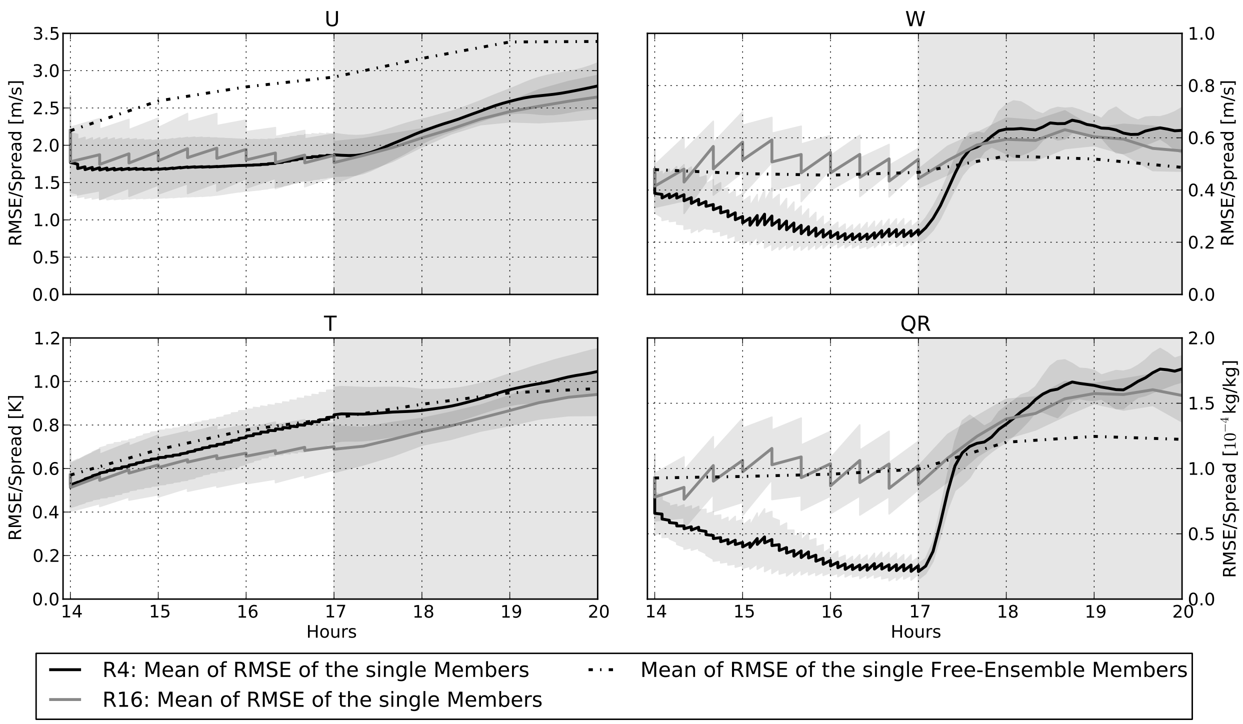

FIG. 6. Like Fig. 5, but for the mean of the RMSE of the single ensemble members of R4 (black) and R16 (gray) instead of the RMSE of the ensemble-mean states.

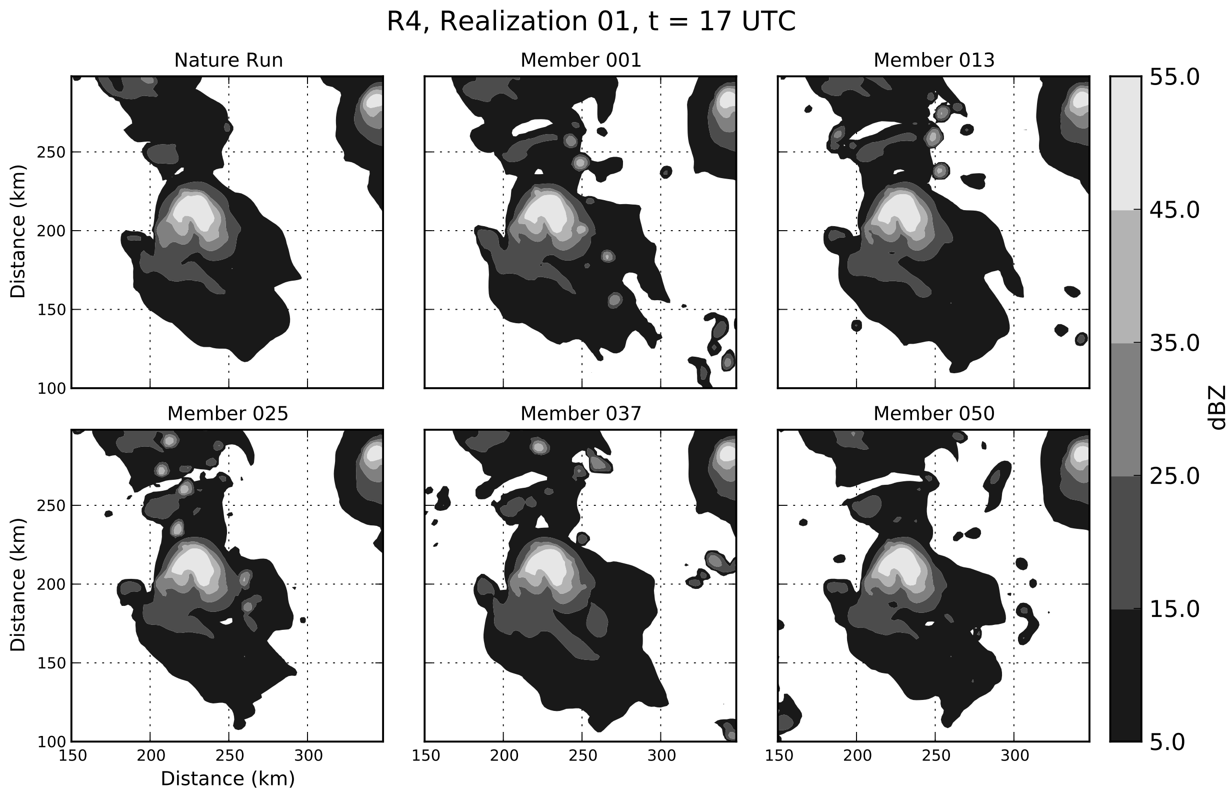

FIG. 7. Maximum reflectivity ($REFL\_MAX$) of nature run and analysis ensemble members 1,13,25,37,50 of R4 (Realization R01) at the last assimilation time 17:00.

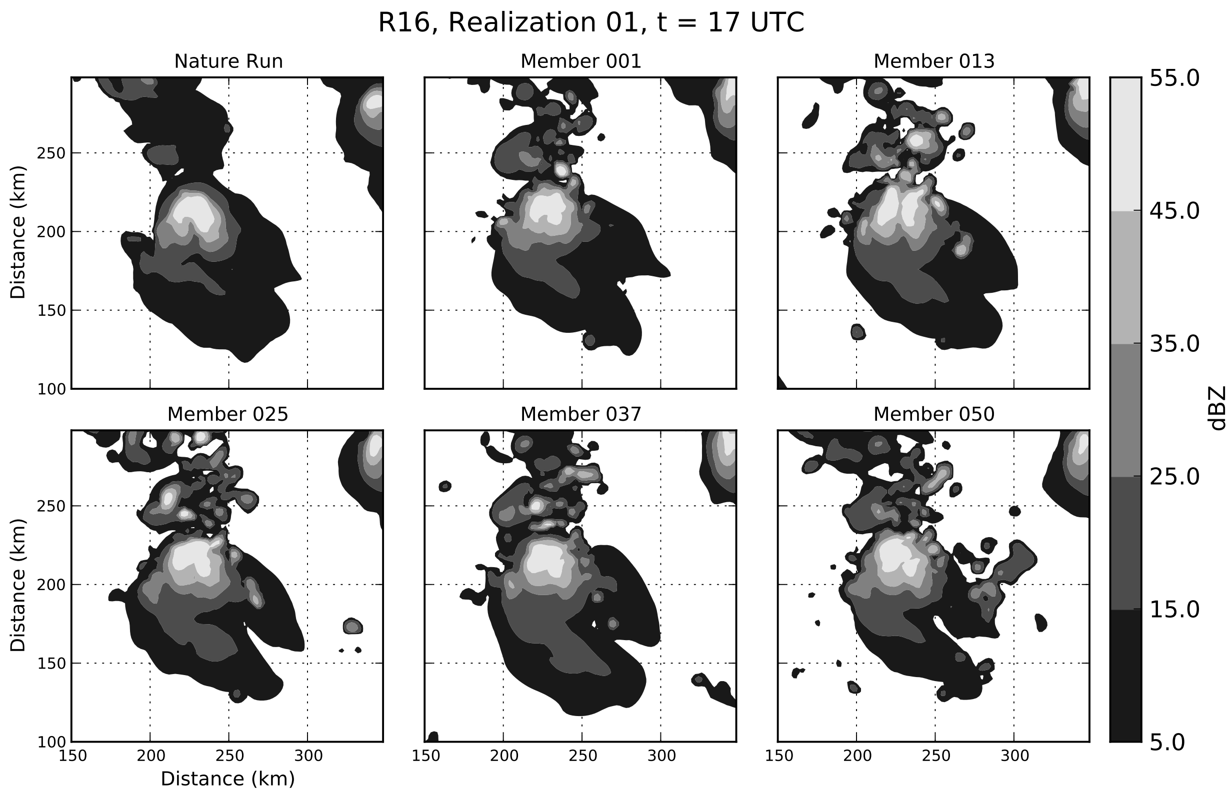

Fig. 8. Like Fig. 7, for R16.

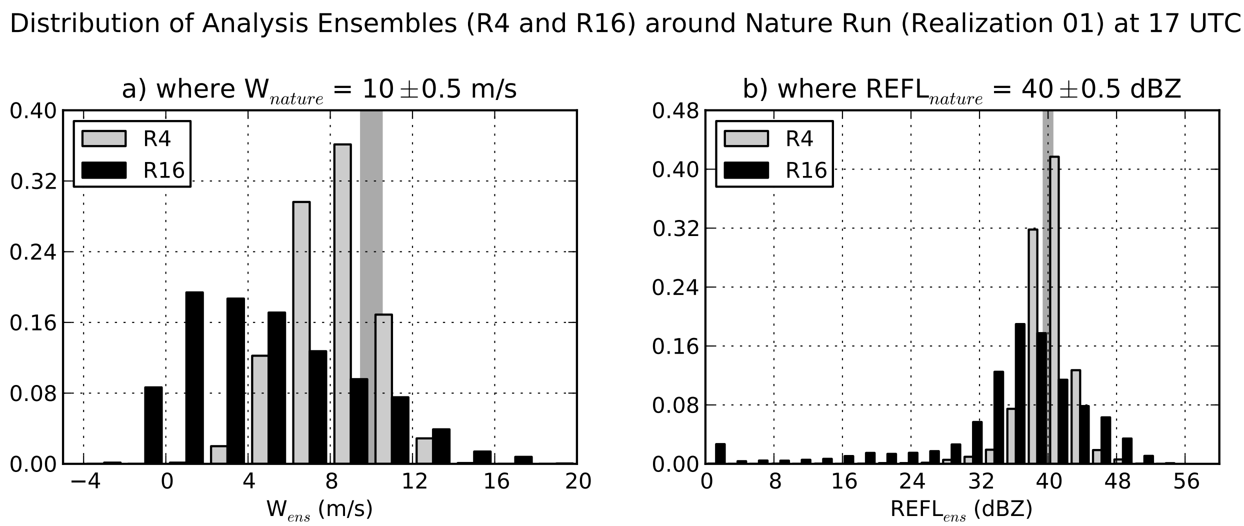

FIG. 9. Relative frequencies of model values within the ensemble-members of R4 (black bars) and R16 (gray bars). Regions inside storms are given by a) and b): a) at gridpoints where W of the nature run is 10 ± 0.5 m s^{-1}, b) where $REFL_{nature} = 40 \pm 0.5$ dBZ. Regions closer than 64 km to the lateral boundaries are excluded so the histograms are computed only where the filter is fully active in R4 and R16. The bar width is 1/2 of the binwidth.

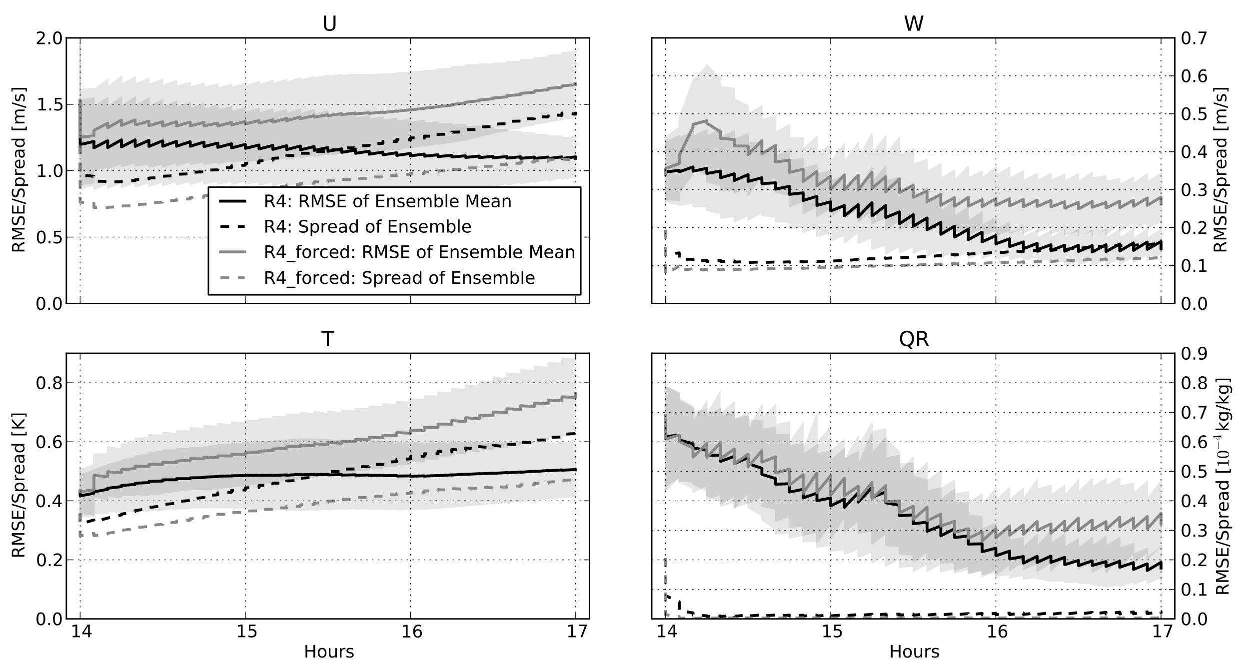

FIG. 10. Like Fig. 5, but comparing R4 (black) to R4_forced (gray) for the primary variables U, W, T and QR in the first two hours of assimilation.

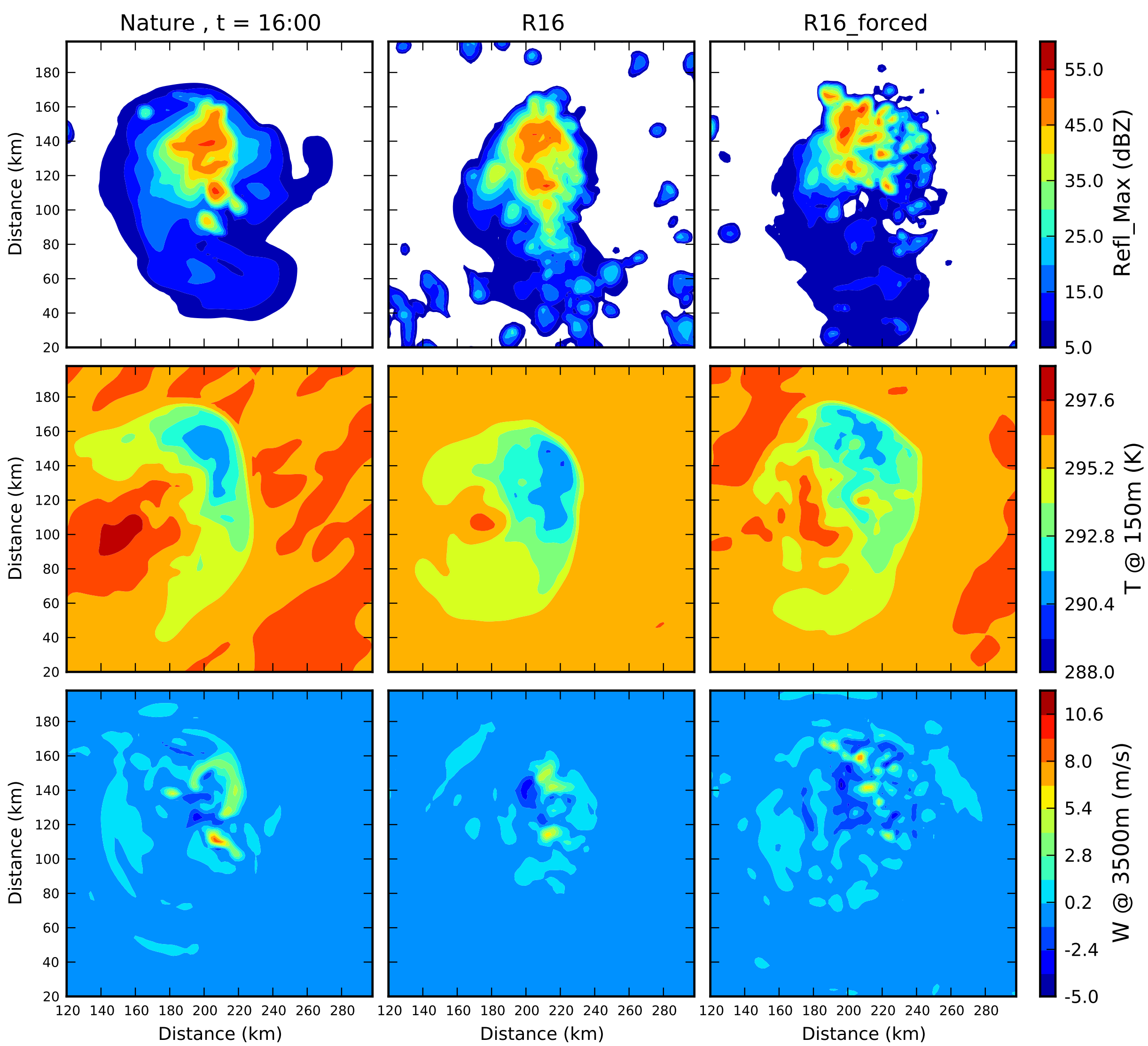

FIG. 11. Like Fig. 4, but comparing the analysis means of R16 and R16_forced to the nature run at 16 UTC during the assimilation (Realization 03).

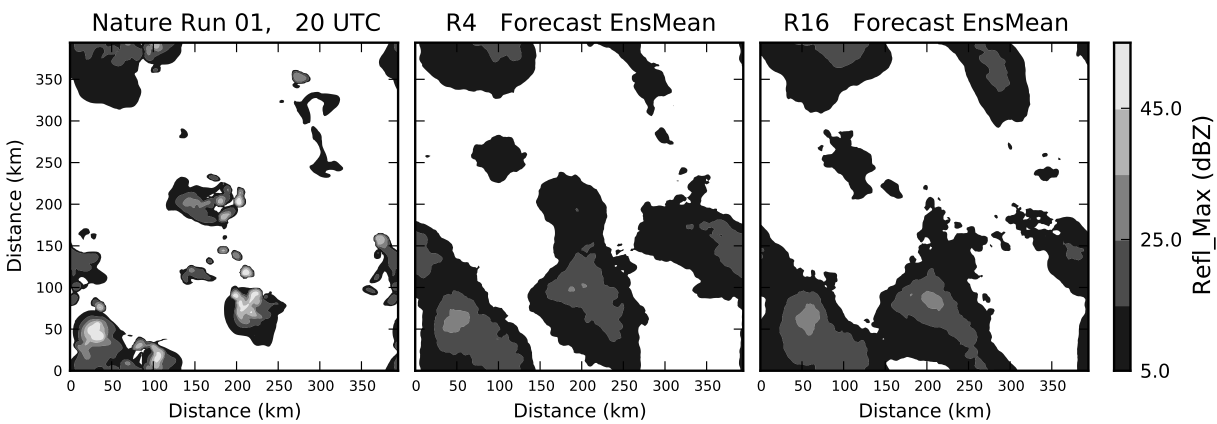

Fig. 12. Like Fig. 4, but forecast ensemble means (Realization 01) at 20 UTC.

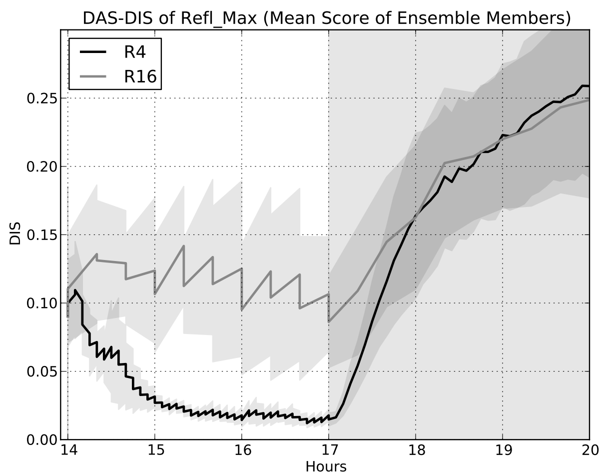

FIG. 13. DIS-component of the DAS-score of the maximum column reflectivity *REFL_MAX* > 10 dBZ. Displayed is the mean DIS-score of the ensemble members of R4 (black), R16 (gray). The error lines are the average over five experimental repetitions. An envelope of the empirical standard deviation is given by the shaded areas. A value of DIS = 0 is a perfect match.

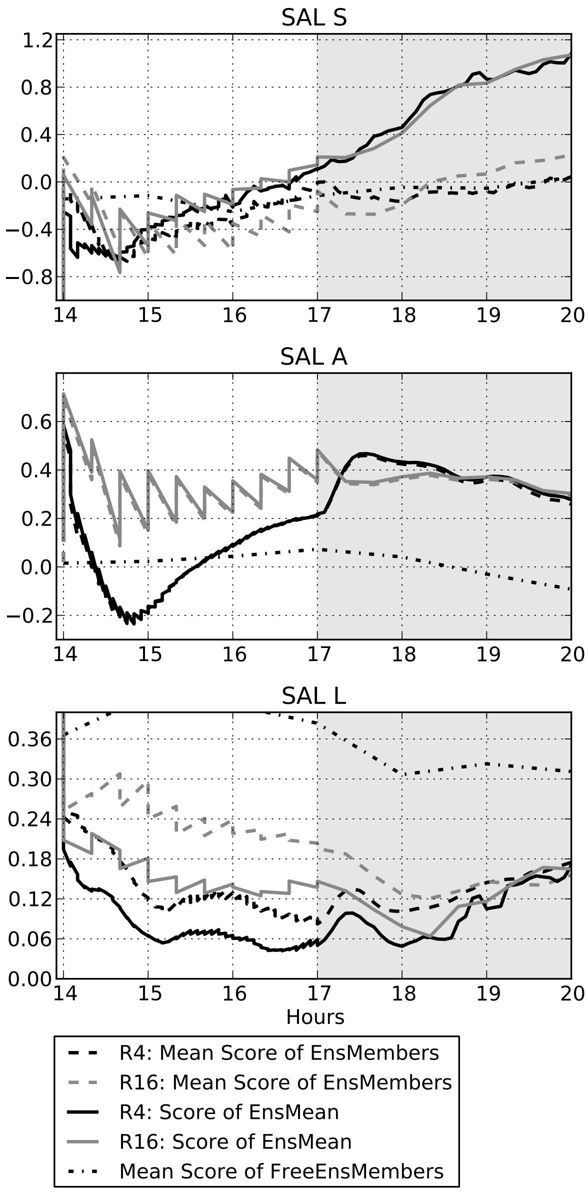

FIG. 14. SAL score components S, A and L (see text) of *REFL_MAX* > 10 dBZ, averaged over the five repetitions.

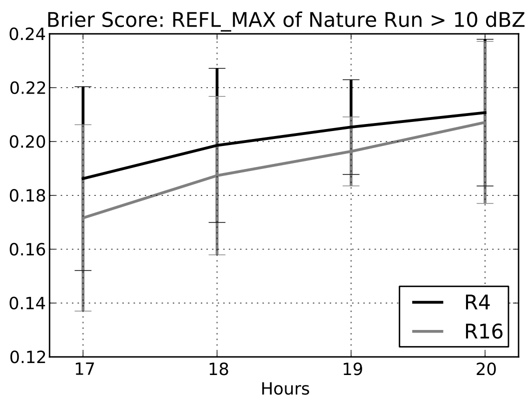

Fig. 15. Brier Score of forecast ensembles R4 and R16 at locations where *REFL_MAX* > 10 dBZ in the nature run. Errorbars show the empirical standard deviation of the five experimental repetitions.