The Impact of Data Assimilation Length Scales on Analysis and Prediction of Convective Storms

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

An idealized convective testbed for the LETKF is set up to perform storm-scale Data Assimilation of simulated Doppler radar observations. Convective systems with lifetimes exceeding six hours are triggered in a doubly periodic domain. Perfect model experiments are used to investigate the limited predictability in precipitation forecasts by comparing analysis schemes on different length scales. Starting from a high resolution reference scheme with 8 km covariance localization and observations with 2 km resolution on a 5 minute cycle, an experimental hierarchy is set up with a larger covariance localization radius of 32 km, observations that are horizontally coarse-grained by a factor of 4, a coarser resolution of the analysis weights, and a cycling interval of 20 minutes.

After 3 hours of assimilation, the high resolution analysis scheme is clearly superior to the configurations with coarser scales in terms of RMS error and field-oriented measures. The difference is associated with the observation resolution and a larger localization radius required for filter convergence with coarse observations. The high resolution analysis leads to better forecasts for the first hour, but after three hours, the forecast quality of the schemes is indistinguishable. The more rapid error growth in forecasts from the high resolution analysis appears to be associated with gravity wave noise and spurious convective cells, suggesting that the field is in some sense less balanced, or less consistent with the model dynamics, than in the coarser resolution analysis.

1. Introduction

Over the last decade, Data Assimilation (DA) with an Ensemble Kalman Filter (EnKF) (Evensen 1994; Houtekamer and Mitchell 1998) for Doppler radar observations has been demonstrated to be a feasible method to obtain suitable initial states of convective storms for very short-range ensemble forecasts in studies with both simulated (Snyder and Zhang 2003; Tong and Xue 2005) and real observations (Dowell et al. 2004; Dowell and Wicker 2009; Aksoy et al. 2009, 2010). 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 have focused on storm-scale analyses that apply relatively small covariance localization lengths in the range of ~ 10 km (Sobash and Stensrud 2013) and converge the analysis ensemble closely towards the observations, pursuing the goal of obtaining initial states with small errors as the basis of their ensemble forecasts.

This study tries to assess the benefits of such close convergence, because even a very precise analysis with small errors may be of limited value when used as an initial state for a forecast. The reason for this is the short predictability of convection in the chaotic atmospheric system where small scale errors grow rapidly (Lorenz 1969) because the small-scale error spectrum saturates faster than the larger-scale error spectrum. By adding random perturbations to convection in twin-experiments, Zhang et al. (2003, 2007) observed a saturation of small scale error growth within a few hours – a limitation that directly affects approaches of highly-resolved ensemble data assimilation on the convective scale.

Taking the limited predictability of convection into account, the benefits of a high-resolution approach in convective EnKF system may not justify the cost: an analysis ensemble not constrained to accurately reproduce the smallest scales in the observations might provide comparably good Quantitative Precipitation Forecasts (QPFs) for leadtimes where the small scale error growth has had enough time to saturate.

a. Limited predictability in convective scale data assimilation

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 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, saturating within 3-6 hours. The specific predictability limit in these studies depended on the presence of a large scale forcing that determined the type of convection and the spatial position of the cells. 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 with weak synoptic forcing in which the storm properties are not strongly constrained by the large scale forcing.

To assess the error growth processes in the framework of data assimilation (Kuhl et al. 2007), this study performs experiments with the Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al. 2007) coupled to an idealized setup of the nonhydrostatic COSMO model (Consortium for Small-scale Modelling) (Baldauf et al. 2011) containing severe and long lived convection. Observation System Simulation Experiments (OSSEs) of cycled DA are performed where synthetic observations radar observations are drawn from a nature run. A perfect model approach and a horizontally homogeneous environment without large-scale forcing are applied to focus on the intrinsic predictability of convective storms. The authors are aware that model error is usually the largest contributor to forecast errors of convection. This study uses a perfect model to be able to focus on the properties and results of the data assimilation cycling wherein it is favorable for the interpretation to know the true atmospheric state (viz. the nature run).

b. *Assimilation experiments with different length scales*

First, a reference experiment is devised in order to reproduce the results of the previous studies on convective EnKF DA. For this, fine-scale observations of radar data are drawn from the nature run at the full model resolution and be assimilated with a suitably small covariance localization length in order to converge the analysis ensemble closely towards the nature run.

Analysis schemes with different spatial and temporal resolution of the observations in combination with different covariance localization radii are then constructed to provide EnKF-generated perturbations with errors at different scales. Successively, (i) the horizontal localization radius of the covariances is increased (Sobash and Stensrud 2013), (ii) the scale of the observations are coarsened by averaging them into superobservations (Alpert and Kumar 2007), (iii) the computation of the LETKF analysis is done on a reduced horizontal grid (Yang et al. 2009), (iv) the temporal assimilation interval is extended to provide the analysis with observations less frequently.

The cycled assimilation covers a time-span of three hours, followed by three hours of ensemble forecast. Different combinations of the experiments given by (i)-(iv) are evaluated using the Root Mean Square Error (RMSE) of the states together with object- and field-based forecast skill scores to see how quickly the advantage of a fine analysis state is lost in the QPF due to changes in the precision and scale of the analysis errors. It is also discussed how a close convergence might cause problems for the dynamics of the forecast model.

The LETKF-experiments of this study try to focus on basic properties of such an assimilation system and do not make use of some recent innovations that can help to improve EnKF analyses, such as adaptive covariance inflation (Anderson 2008), additive inflation (Dowell and Wicker 2009), "Running-In-Place" (Kalnay and Yang 2010; Wang et al. 2012) or Gaussian anamorphosis of precipitation observations (Lien et al. 2013).

For all experiments, the EnKF will be initiated with a convective ensemble that is spun up from random initial white noise and therefore lacks any prior knowledge about the position of the observed storms in the nature run. This "bad background" leads to a longer period for the initial convergence of the ensemble, but avoids the possibly beneficial influence e.g. of arbitrary convective triggers that are "manually" introduced at predetermined locations (Tong and Xue 2005; Aksoy et al. 2009, 2010) to help the background ensemble resemble the observations even in the first assimilation step.

2. Model configuration and experimental design

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 scale-varying experiments. 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 Consortium for Small-scale Modeling (COSMO) model simulations in the domain over Germany (COSMO-DE) (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 PP from a stationary hydrostatic base state. The moist physics uses a single-moment bulk microphysics scheme with six state variables: water vapor QV, cloud water QC, cloud ice QI, rain QR, snow QS and graupel QG. A radiation scheme for long- and shortwave radiation is applied. Surface fluxes of latent and sensible heat are parametrized and constrained by a constant surface temperature and a constant surface specific humidity throughout the simulation.

a. Nature run

This study uses the testbed setup of COSMO with idealized initial state, periodic boundary conditions and a homogeneous flat landscape as the lower boundary. A convection-permitting horizontal resolution of 2 km and 50 vertical levels is set up in a domain of $396 \times 396 \times 22$ km extent. The vertical resolution ranges from 800 m at the model top to 100 m at the surface. The initial profile of all model runs is horizontally homogeneous and based on the sounding of Payerne (CH, Radiosonde 06610) at 12 UTC on July 30th 2007, a day with strong convective storms and mesoscale convective systems, 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 with amplitudes of 0.02 K and 0.02 m s^{-1}, respectively. The model runs start at 06 UTC and quickly develop a convective boundary layer. Instability increases prior to the outbreak of convection due to radiative cooling of the upper troposphere while the surface temperature is held constant.

Small showers initialize at random locations at 08 UTC, grow until 10 UTC and mostly die off by 12 UTC (Fig. 1). The surviving systems grow into intense storms and mesoscale convective systems by 14 UTC and propagate with the mean wind in a north-eastern direction through the domain with lifetimes $\geq$ 6 h (Fig. 2). The horizontally contiguous rain areas extend over distances from 30 to 150 km with reflectivity larger than 30 dBZ. Surface fluxes of sensible and latent heat lead to gradual decay of the cold pools in the wake of the storms. The periodic boundary conditions allow the storms to spin up in a way that is "natural" for the model physics in the given sounding despite the modest domain size. The time-window between 14 and 20 UTC is chosen for the experiments because the storm properties such as size and organization change little during this period (Fig. 1). For 3 hours, cycled assimilation is performed every 5 (or 20) minutes from 14 until 17 UTC, followed by ensemble forecasts with 3 hours leadtime from 17 until 20 UTC.

All experiments in this study are repeated five times. Each repetion uses a different seed for the random noise field to initialize its nature run (not shown). These five repetitions represent a variety of different storm positions and shapes which are possible given the initial sounding. When generating the nature runs, one random case was excluded in which the preliminary showers (Fig. 1 at 12 UTC) died out and no larger storm grew. An examination of the full set of ensemble members, which were also randomly initialized, showed that this was a rare event, occurring for $\approx \frac{1}{50}$ of the members, and therefore no such case was included in the five nature runs.

b. *Synthetic observations*

Synthetic radar observations of reflectivity (Z) and the component of horizontal wind in the x-direction (U) are generated from the nature runs. In order to mimic a region of good radar coverage, the observations are taken at every model gridpoint in the horizontal direction with the grid spacing of $\Delta x_{model} = \Delta x_{obs} = 2$ km. In the vertical, every third grid point between 500 m and 13 km above the model surface is observed.

Reflectivity Z is computed from mixing ratios of graupel (QG), rain (QR) and snow (QS) using the simple formula of Done et al. (2004): $Z_{fac,Q} = A_Q(\rho Q)^{1.75}$ where $A_{QR} = 3.63 \cdot 10^9$ for $Q = QR$, $A_{QS} = 2.19 \cdot 10^9$ and $A_{QG} = 1.03 \cdot 10^9$ for with the air density ρ. The reflectivity is given by $Z = 10 \log_{10}(Z_{fac,QR} + Z_{fac,QS} + Z_{fac,QG})$ in dBZ. These values were designed for the WRF model but also give reasonable values of dBZ with COSMO, mimicking the behaviour of real logarithmic reflectivity observations.

As in other OSSE studies (Tong and Xue 2005), Gaussian noise with a standard deviation of $\sigma_{refl} = 5$ dBZ is added to simulate measurement errors. The reflectivity observations Z are masked to regions where $Z > 5$ dBZ. Below this threshold, they are assimilated as observations of no-reflectivity (Tong and Xue 2005; Aksoy et al. 2009; Weygandt et al. 2008; Benjamin et al. 2012).

In reality, most elevation angles of a radar volume scan are shallow, so the radial wind mainly contains information about the horizontal wind. Observations of the horizontal wind component U, masked to $Z > 5$ dBZ and with an added error of $\sigma_U = 1$ m s^{-1}, are therefore used as a proxy for radial wind observations. As the storms move in a north-eastern direction, U contains information of both storm propagation and horizontal divergence patterns. The regular observation geometry and the usage of U ensures that the observed information 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 seed for 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. Snapshots of the reflectivity field at different times throughout one of the nature runs are shown in Fig. 2. Other runs appear similar, but with random displacement and shape variations of the intense storm systems. This initialization method was chosen to deprive the ensemble of any prior knowledge about the state of the nature run when the assimilation starts, as would 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 (described fully by Hunt et al. (2007)) determines the vector $\mathbf{w} = \bar{\mathbf{w}}^a$ that minimizes the cost function

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

where the k-dimensional vector $\mathbf{w}$ defines the optimal linear combination of ensemble member states that minimizes J^*. $\mathbf{y}^o$ is the vector of observations and $\mathbf{R}$ is the observation error covariance matrix which is treated as diagonal here. $\bar{\mathbf{y}}^b$ 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 $\bar{\mathbf{x}}^b$, *viz.*

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

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 minimum of the cost function (1) is computed locally for every analysis grid point to determine the best local linear combination of forecast members in the weighting vector $\mathbf{w}^{a(i)}$. However, these single analyses do not necessarily need to be computed at the full model resolution: the spatial field of the local $\mathbf{w}^{a(i)}$ is usually quite smooth in case of $\mathbf{R}$-localization (Janjić et al. 2011), so a coarser *analysis grid* can be chosen horizontally and vertically on which the local analysis weights $\mathbf{w}^{a(i)}$ are computed before being 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. Consistent with the doubly periodic lateral boundary conditions of the model used in this study, the synthetic observations are periodically replicated around the original domain and the filter algorithm of KENDA is configured to take also observations into account that are nominally outside of the domain while still within the horizontal localization radius of the outermost grid points. This leads to a fully periodic LETKF analysis.

For all experiments, the analysis weights $\mathbf{w}^{a(i)}$ are multiplied by a constant covariance inflation factor of $\rho = 1.05$ (Aksoy et al. 2009) in order to enhance the span of the analysis state space. All prognostic variables of the model are updated in the analysis computations (zonal wind U, meridional wind V, vertical wind W, temperature T, pressure perturbation

PP, water vapor mixing ratio QV, cloud water mixing ratio QC, cloud ice mixing ratio QI, rain water mixing ratio QR, snow mixing ratio QS, graupel mixing ratio QG).

e. *Reference configuration*

To reproduce the typical behaviour of an EnKF DA system with the assimilation of simulated radar observations, a reference setup with the name L8 was used and is described here. In L8, an assimilation interval of $\Delta t_{ass} = 5$ minutes represents the typical availability of volume observations from a scanning Doppler radar network (Lu and Xu 2009).

L8 uses a horizontal localization cutoff length of $r_{Loc,h} = 8$ km, so the ensemble covariances contain storm-internal structures, while the overall structure of the observed storms has to be recovered by assembling the overlapping local analyses from neighboring analysis grid points. In L8, the horizontal resolution of this analysis grid coincides with the full model grid ($\Delta x_{ana} = \Delta x_{model} = 2$ km). The analysis grid has 20 vertical levels with a spacing that varies with the logarithm of the reference pressure. This is similar to the vertical grid structure of the model, but with a vertical spacing varying from 1600 m at the model top to 250 m at the surface.

The vertical localization radius $r_{Loc,v}$ varies with height, so that observations close to the surface have a vertical influence of ~ 1 km, while observations taken a height of 12 km have a vertical influence of ~ 6 km. The combination of the vertical analysis grid structure, the vertical localization radius and the resolution of the synthetic observations at every third grid point vertically should provide a sufficient overlap for vertically consistent analysis computations. The authors recognize that the vertical covariance structures in deep convection can extend up to lengths of 10 km when sampled by an EnKF background ensemble (Tong and Xue 2005) – the shallower vertical localization was chosen here also as a compromise for computational efficiency.

In the L8 configuration, the positions of the storms in the analysis should closely coincide with those of 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 for a "converged" analysis ensemble as defined in the introduction.

f. *Configurations with different length scales*

All experiments are performed with the full model resolution of $\Delta x_{model} = 2$ km for both nature run and ensemble members. To produce analyses with varying scales, $r_{Loc,h}$, Δx_{obs}, Δx_{ana} and Δt_{ass} are varied to create the experimental hierarchy shown in Table 1

(i) *Horizontal localization (L8, L32)*: the horizontal localization is increased from of $r_{Loc,h} = 8$ km to 32 km. The vertical localization is not varied.

(ii) *Superobservations (SO)*: The observation interval Δx_{obs} is increased from 2 km to 8 km in experiments L8SO and L32SO, by horizontally averaging the values and positions of 4×4 blocks of the original observations into one central 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 entries of the observation error covariance matrix $\mathbf{R}$ are kept the same for one SO as for one original observation. The vertical resolution of the observations is not varied.

(iii) *Coarse Analysis Grid (CG)*: The analysis grid spacing Δx_{ana} is increased from 2 km to 8 km in experiments L8SOCG and L32SOCG by computing the local analysis weights $\mathbf{w}^{a(i)}$ at every fourth model grid point ($\Delta x_{model} = 2$) km and then linearly interpolating onto the full model grid (Yang et al. 2009). After interpolation, the transformation from ensemble space into model space is performed in the LETKF as a linear combination of background members. The vertical resolution of the analysis grid is not varied.

(iv) *Assimilation interval (20)*: The cycling interval Δt_{ass} is increased from 5 min to 20 min in experiments L8SOCG20 and L32SOCG20 to test the effect of less frequent introduction of observation information.

The factor 4 difference in all parameters between (i) and (iv) implies that all experiments with "L32SO" have the same number of observations per local analysis as the reference experiment L8, while the pure L32 has to fit the analysis ensemble to more local observations.

The parameter combinations listed in Table 1 were selected first (i) to use observation and background information about larger-scale correlations for the analysis, (ii) to reduce the horizontal resolution of the observation information, (iii) to let the filter compute only directly on this coarsened scale before the full scales are updated, and (iv) to further temporally reduce the amount of information provided to the analysis.

The analysis field of local linear ensemble member combinations $\mathbf{w}^{a(i)}$ in (i) is therefore computed with a larger horizontal overlap ($r_{Loc} = 32$ km), so the "assembly" of members in L32 is performed less locally than with the small localization of $r_{Loc} = 8$ km in L8. It might be dynamically favorable to have less variation in the linear combination of ensemble members especially in regions with large horizontal gradients such as up- or downdraft cores because a linear combination of nonlinear dynamics such as convection is not necessarily a dynamically consistent state for every member.

g. RMSE, spread and consistency ratio

The accuracy of analysis and forecast states $\mathbf{x}$ is measured by the Root Mean Square Error (RMSE) of the ensemble mean $\bar{\mathbf{x}} = k^{-1} \sum\limits_{i=1}^{k} \mathbf{x}^i$, computed in model space for the different model variables:

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

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 $\bar{\mathbf{x}}$, 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 close to $CR = 1$. In addition to (4), the mean RMSE of the single members with respect to the nature run is computed for some comparisons:

$$\overline{RMSE_{mem}}(\mathbf{x}) = k^{-1} \sum_{i=1}^{k} \left(\sqrt{m^{-1} \sum_{l=1}^{m} (x_l^{nature} - x_l^i)^2} \right). \tag{7}$$

As in previous studies, RMSE and spread are evaluated for rainy model grid points above the detection threshold of 5 dBZ where the up- and downdraft dynamics are most active and the error reduction most significant. Additionally, the present study evaluates RMSE and spread for all grid points of the domain. This results in a generally low error level because many small error values of clear air regions contribute to the mean error, but is of interest since temperature and wind errors can occur outside of precipitating regions, especially in the boundary layer. Analysis and forecast errors are compared to a reference error level computed for a free-running ensemble that evolves through the diurnal cycle from 14 to 20 UTC, but without any assimilation of observations.

In Section 4, additional feature-based scores are described and applied to analyses and forecasts.

3. Assimilation results

First the reference experiment L8 is evaluated during the assimilation window 14-17 UTC in Section 3a, then the assimilation results of the scale-varying experiments L8S0-L32SOCG20 (Table 1) are compared to L8 in Section 3b. The free forecast parts of the experiments are discussed in Section 4.

An example of the assimilation results for one realization of the nature run is given in Figs. 3, 4 and 5, which compare snapshots of the nature run 03 to the analysis ensemble means at 17 UTC. The composite reflectivity is the vertical maximum the 3D-reflectivity field (Fig. 3). T at the height of $z = 150$ m shows the cold pools (Fig. 4), W at $z = 3500$ m shows regions of up- and downdrafts (Fig. 5). Visually, the L8 analysis closely matches the nature run, particularly in reflectivity and vertical velocity.

a. *Performance of the reference scheme L8*

Fig. 6 shows the RMSE of the ensemble mean and the spread of L8 during the assimilation window and during the forecast window from 17 to 20 UTC for each of the prognostic model variables and the derived Reflectivity ($Refl$), computed at all model gridpoints. To illustrate the relative error reduction, the RMSE and spread of the free-running ensemble (which has not undergone any assimilation) are also depicted. An "error-reduction" is therefore a reduction of RMSE with respect to the free error level. The two observed variables are U and $Refl$, all other variables are updated only through covariances provided by the background ensemble.

The reduction of RMSE for all model variables shows the effectiveness of the LETKF-cycling. The error of the meriodional wind U decreases during the assimilation window. At the same time the spread adjusts towards a good consistency ratio by 17 UTC, showing that the chosen inflation factor is appropriate. The zonal wind V, although only updated through covariances, behaves similarly to U. This is probably due to the strong coupling of U and V in the domain sounding with a north-westerly background wind.

The pressure field PP, the humidity QV and the temperature T (illustrated by cold pool structures in Fig. 4) also benefit from the LETKF-cycling updates through ensemble covariances. Note that the filter-update of T is slightly detrimental at analysis times before 16 UTC, showing a "reversed" saw-tooth pattern (visible in a close examination of Fig. 6), while the error decreases during the 5 minutes of forecast intervals due to the dynamical convergence of the ensemble members towards a physical meaningful state. This indicates that the direct ensemble covariances of T with the observed U and $Refl$ are probably not well-sampled by the background ensemble. The deficient T-analysis in the EnKF-assimilation of radar data has been noticed in previous studies (Zhang et al. 2004; Dong et al. 2011).

Information about vertical motion of W (up- and downdrafts in Fig. 5) is provided by observations of $Refl$ (cf. Fig. 5 of Tong and Xue (2005)) because the observed reflectivity field is confined to vertically active regions and by horizontal convergence patterns of U that enclose the up- and downdrafts (cf. Fig. 8 of Snyder and Zhang (2003)).

The horizontally intermittent $Refl$ field is well-captured by the analysis ensemble (cf. Fig. 7), and the RMSEs of the precipitation variables QR, QS and QG are reduced. QR, QS and QG are used to compute observations and first guesses of $Refl$ in the observation operator and therefore are well-contained in the ensemble covariances. The unobserved cloud variables QC and QI also benefit from the LETKF-update. On the other hand, the spread of the precipitation variables is strongly reduced and does not recover during the assimilation window. This is probably caused by the non-Gaussian climatological distribution of the clouds which is converged towards a Gaussian solution of the filter (Dance 2004) which then has a smaller spread than the climatology. This is illustrated in Fig. 8 which shows histograms of W and $Refl$ of the L8 analysis ensemble, computed from the gridpoint values of all analysis members at locations inside the storm in the nature run (*i.e.* points satisfying $W_{nature} = 5 \pm 0.5$ m s^{-1} or $Refl_{nature} = 25 \pm 0.5$ dBZ in left and right panels, resp.). These show an approximately Gaussian distribution around the observed ($Refl$) or covariance-derived (W) values from the nature run.

Fig. 9 shows the RMSE and spread of L8 as in Fig. 6, but now computed only at those grid points where $Refl$ of the nature run exceeds the observation threshold of 5 dBZ. The overall variance of all variables is larger in these convective regions, as is the relative amount of error reduction. This is because observations of U and $Refl$ are available for the whole subset volume where RMSE and spread are evaluated. The main difference between Fig. 6 and Fig. 9 appears in U, V, T and QV with a stronger error reduction inside the thresholded subset, but also a smaller spread when compared to the evaluation including all gridpoints.

As the differences between the two choices of grid point subsets for RMSE-computations are small, further discussion of RMSEs in this paper will refer to the full-domain RMSEs of Fig. 6. Furthermore, since the RMSE behaviour of many variables are similar to each other, only plots of four representative variables will be considered. U will serve as a proxy for V, T as proxy for PP and QV, and QR as a proxy for QC, QI, QR, QS, QG and $Refl$ (Fig. 10).

Summarizing the results for the high-resolution assimilation, the L8-scheme appears to produce LETKF-analyses that are comparable in quality to the previous convective EnKF studies with radar data mentioned earlier. The mean of this strongly converged ensemble is representative of the best possible solution, with little variance among the analysis members inside the observed storms.

b. Influence of length scales on assimilation results

In Section 2f and Table 1, the different length scales employed in the schemes L8SO – L32SOCG20 are specified. Here, the assimilation results of the various experiments are evaluated in comparison to the reference experiment L8 (Section 3a).

First, the effect of increasing only the localization radius is considered (step i in the experimental hierarchy of Section 2f), by comparing L8 to L32, where the horizontal localization $r_{Loc,h}$ is changed from 8 km to 32 km. In L32, the storms in the analysis ensemble mean do not converge well onto the observations. This is clearly visible in the plots of the last analysis ensemble mean in Figs. 3, 4 and 5. The mean reflectivity field is much weaker than the nature run, as is the cold pool intensity. These results are typical of the five repetitions of the experiment although the fields differ in detail. The RMSE of the L32 ensemble mean (Fig. 10) shows very poor performance for U, W and QR. For T, the ensemble mean of L32 is strongly degraded by every analysis cycle and is even worse than the free error level. These deficiencies indicate that the number of 50 ensemble members is too small in L32: for every local analysis, 4 times the number of full-scale ($\Delta x_{obs} = 2$ km) observations must be fitted by the LETKF, in comparison to L8. As a result, the spread of L32 decreases, leading to a very poor consistency ratio for all variables (not shown).

Second, the change in horizontal observation resolution Δx_{obs} from 2 to 8 km (step ii: superobservations SO) is evaluated by comparing L8SO to L8, and L32SO to L32. In L8SO, the analysis field of reflectivity appears more spotty (Fig. 3) than in L8. In single members of L8SO in Fig. 11, the inner storm core is somewhat broken up and spurious convective cells exist in many locations. The superobservations are obtained by coarse-graining horizontally onto 4x4 blocks, so that fine-scale errors are obscured and are not penalized appropriately in the analysis. This is evident in the worse RMSE of W and QR of L8SO compared to L8 (Fig. 10). For U and T, L8SO does not perform much worse than L8, indicating that less horizontal information is needed for a reasonable analysis of the horizontally smoother variables U and T. It is perhaps surprising that in L8SO the filter does not diverge although there is very little overlap in the solution of adjacent local analyses.

In L32SO, the same horizontal number of observations per local analysis is available as in L8, but with a coarser observation resolution of 8 km. Indeed, the analysis ensemble of L32SO is now able to converge towards the superobservations, in contrast to L32 (Fig. 10), albeit with less precision in the small scales than L8. This can be seen comparing Fig. 3 and Fig. 12, which shows L32SOCG as a proxy for L32SO.

In the third set of experiments, the LETKF computations are performed on an analysis grid with a horizontal resolution decreased from $\Delta x_{ana} = 2$ km to 8 km (step iii: coarse analysis grid CG). The resulting field of local analysis weights $\mathbf{w}^{a(i)}$ is then interpolated onto the model grid. Using $\Delta x_{obs} = 8$ km (SO) in combination with $\Delta x_{ana} = 8$ km (SOCG), the analyses of the SOCG-experiments are computed at the same horizontal scale where the observations are available. The full-resolution covariances in model space sampled from the background ensemble are still used for the analysis ensemble. Comparing the ensemble mean field plots of L8SOCG to L8SO and L32SOCG to L32SO, the change from SO to SOCG appears insignificant. The RMSE shown in Fig. 10 shows little influence of the coarse grid method in both L32SO(CG) and L8SO(CG). This indicates that the fields of $\mathbf{w}^{a(i)}$ are smooth enough to be accurately represented on the coarse grid (although the setup of L8SO and L8SOCG could not be recommended for actual usage due to the lack of horizontal overlap between local analysis regions).

Finally, the interval between the cycling steps is increased from 5 to 20 minutes (step iv: 20 minutes). In L8SOCG20, this less frequent introduction of observation information leads to strong deterioration in the analysis mean solution of storm positions and cold pools (Figs. 3 and 4) and the RMSEs are significantly worse than for L8SOCG (Fig. 10).

In L32SOCG, the lowest spatial resolution of observations and analysis is reached, and L32SOCG20 additionally lowers the temporal resolution. As for the L8SOCG20 experiment the precision of the analysis mean is degraded (Fig. 10). In Fig. 12, the analysis members of L32SOCG show a much larger spatial variability of the storm field than L8 (Fig. 7). This variability is even larger with L32SOCG20 (Fig. 13). Fig. 8 compares the distributions of W and $Refl$ of the analysis ensembles of L32SOCG and L32SOCG20 to analysis ensemble L8, at subsets of points inside the nature run's storms. Fig. 8b shows that the analysis distribution of L32SOCG is broader than L8 and includes values of zero reflectivity. This is even more evident for L32SOCG20 where many non-precipitating points are present, and the values of precipitating points are distributed broadly around the observed value. This behavior is closely coupled to the non-observed W-updrafts in Fig. 8a where L32SOCG and L32SOCG20 show a more climatological distribution than the closely converged L8, where climatology is defined here by the frequency distribution of rainy grid points in the nature runs (cf. Fig. 2).

The RMSE of the L32SOCG20 analysis mean of W (Fig. 10) is larger than that of the free ensemble and it shows that the ensemble mean of L32SOCG20 has converged towards a solution where the mean updrafts cores are displaced with respect to the nature run (cf. Figs. 3, 13). This large RMSE of W is therefore associated with a doubly penalty, in contrast to the error of the free ensemble, where the randomly-placed convective updrafts contribute no strong features to the ensemble mean. Despite the larger RMSE, the consistency ratio of L32SOCG20 and L8 in W at 17 UTC is $CG \approx 0.2$ inside the storms (cf. W in Fig. 10), so both the finest and the coarsest experiment have converged to a comparable degree on the updrafts of the nature run.

Before the forecast results are presented in the next section, we look briefly in Fig. 14 at the mean RMSE of the individual ensemble members with respect to the nature run, as defined in (7). Although the single members are not themselves the solution of the LETKF (which is the ensemble mean and analysis perturbations from the mean), $\overline{RMSE_{mem}}$ can tell us about the deviations of the single members from the truth. It is remarkable that $\overline{RMSE_{mem}}$ for U and T is the lowest in L32SOCG20, and for W and QR it is well within the range of other experiments, compared to Fig. 10. For U and T, this property indicates that horizontally smooth fields such as U and T are sensitive to the introduction of noise by the filter-increments (Greybush et al. 2011; Holland and Wang 2013), especially for T, where the error of the "low-information" experiment L8SOCG20 is also smaller than for other experiments.

This introduction of noise is also evident in the vertical wind field in Fig. 5 in the form of spurious gravity waves, where the W-field of L32SOCG20 in locations of updrafts appears smoother than in all other experiments, and outside the storms is much less tainted by gravity wave noise (compared to the nature run), especially those with $\Delta t_{ass} = 5$ min.

4. Ensemble forecasts from analyses with different length scales

The main goal of this paper is to evaluate the influence of the spacial scale of ensemble perturbations on the quality of convective ensemble forecasts, and to assess how the benefits of a high-resolution analysis are limited by the predictability of the atmospheric system.

As described before, after the 3 hours of cycled LETKF-assimilation, ensemble forecasts with a leadtime of 3 hours are run from 17 to 20 UTC for all experiments. This section will focus on the forecast results of L8, L32SOCG and L32SOCG20, which exemplify the differences between fine and coarse resolution LETKF analyses.

a. Forecast fields and RMS Error

For U and T, the RMSE of the forecast ensemble mean (Fig. 10) grows similarly slowly for all experiments, aside from the badly converged L32. The RMSE of the forecast ensemble mean is generally larger for the L32-experiments than for the L8-experiments, but appears to converge towards the end of the forecast window.

For W and QR, the RMSE of the L8 forecast ensemble mean increases from a low error to the free error level within the 3 hours from 17 to 20 UTC (Fig. 10). The RMSEs of L32SOCG and L32SOCG20 also converge towards the free error level, but from a larger (and for L32SOCG20 doubly penalized) inital error at 17 UTC. As large values of W and QR are present mainly within convective cell cores, the convergence to the uncorrelated free ensemble error level in Fig. 10 indicates that W and QR become smoother in the ensemble mean, becoming similar to the very smooth free ensemble mean fields of W and QR due to the random position of the free storms. This smoothing can be seen directly in Fig. 15 for the mean forecast composite reflectivity, in Fig. 16 for T and in Fig. 17 for W. The forecast ensemble mean reflectivity field, in particular, resembles a smoothed probability map of the nature run's storm position. For W and $Refl$, the analysis ensembles of L8, L32SOCG and L32SOCG20 had different initial distributions of values (Fig. 8) inside the nature run's storm locations, with the values for L8 narrowly distributed around the observation. However, after 3 hours (Fig. 18), the distributions of W and $Refl$ from the three forecast ensembles are indistinguishable and similar to the climatology of the simulated convective regime.

The $\overline{RMSE_{mem}}$ values of the members in Fig. 14 grow similarly slowly for U and T to the RMSE of the ensemble mean for all experiments. However, the $\overline{RMSE_{mem}}$ of W and QR overshoots the free error level for all experiments except L32SOCG20. The free error level here, in contrast to error of the ensemble mean in Fig. 10, is strongly penalized due to the random storm positions in the single free members. The overshooting $\overline{RMSE_{mem}}$ of the forecasts of L8 compared to L32SOCG indicates a strong double penalty for the forecast updrafts. In Fig. 19, a snapshot of a 1 hour forecast of L8 is displayed. It can be seen that spurious convection arises in the L8 forecast members outside of the true storm position of the nature run. Within the region of stratiform precipitation, the positions of the active parts of the updraft have diverged very strongly. Such spurious development was found in almost all experiments (L8SO – L32SOCG, not shown) and has been observed in previous studies (cf. Fig 3 of Aksoy et al. (2010)). In L32SOCG20 however, almost no spurious convection is seen in the 1 hour forecast outside of the organized convective system (Fig. 20). This indicates that the the analysis states of L32SOCG20 are internally consistent although the members had not converged closely to the observations, whereas the strongly converged analysis of L8 is not well handled by the model dynamics and is probably unbalanced in some sense.

b. Spatial forecast verification methods

To supplement the visible and the RMSE forecast evaluation of the volume fields, two spatial verification measures for QPFs are chosen to compare the forecast rain fields to the nature run. The composite reflectivity field is used here because high reflectivity is usually accompanied with strong precipitation and winds – the essential threats to be predicted by a convective storm forecast. Where needed, the observation field of the nature run is masked to values above a threshold of 10 dBZ to separate the storms and overlapping anvils for object identification.

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, displayed in Fig. 21.

The SAL-score (Wernli et al. 2008) compares statistical properties of thresholded rain-objects. The structure or S-component indicates whether 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 rather than matching individual objects, so SAL-S is independent of location errors and biases (Fig. 22a). The amplitude or SAL-A-component (Fig. 22b) compares the domain-wide total reflectivity (not thresholded) between forecast and observation and thus displays the overall bias. The SAL-L-component (Fig. 22c) determines the location error by measuring the horizontal deviation of the centroid of the forecast reflectivity field from the centroid of the observations. Without matching objects, the curves of the location error SAL-L are therefore less continuous than the DIS-score. This is a consequence of the small sample of five random cases which is used in this study.

c. Spatial forecast verification results

The DIS-score in Fig. 21 shows that after the 3 hours of assimilation, the reference experiment L8 has the lowest position error for the reflectivity field. The L8 experiments with superobservations suffer from the low SO-observation resolution and the small overlap, so the L8SOx experiments are disregarded here. The L32-experiments however appear comparable. In the first 30 minutes of the ensemble forecast, the DIS displacement error of the L8-members grows rapidly and exceeds L32SOCG20, then saturates. The DIS of the L32SOCG20-members grows slowly and converges with L8, L32SO and L32SOCG towards the end of the forecast window. This finding corroborates the results of the the $\overline{RMSE_{mem}}$ of QR and W in Fig. 14 which showed the influence of spurious storms in the forecast (cf. Figs. 19 and 20).

The rapid displacement error growth of the L8 members is also displayed by the SAL-score: In the first 30 minutes of the ensemble forecast, the bias SAL-A of L8 increases, the location error SAL-L of L8 grows strongly, and the structure SAL-S of L8 decreases to negative values associated with spikier objects. This means that during this first half hour additional small and mislocated convective cells are triggered in L8: additional because the bias SAL-A increases, small because SAL-S decreases, and mislocated because they increase the location error SAL-L (DIS). In contrast, none of this appears to happen in the experiment L32SOCG20 which has the coarsest analysis properties.

d. "Balance" of initial states

The development of spurious updraft cores in L8 – L32SOCG suggests that the states of the analysis ensemble members are in some way inconsistent with the forecast model dynamics. In Fig. 5, it is apparent that all analyses except that of L32SOCG20 have fine-scale gravity wave noise present in the W variable that is not found in the nature run. While it is not expected that convective forecasts are close to geostrophic balance, the apparent excess of free gravity waves suggests the initial states for the ensemble forecasts may suffer from a similar "dynamical imbalance", caused by noisy analysis increments of the LETKF. In Fig. 23, the surface pressure tendencies of L8, L32SOCG and L32SOCG20 during the first 20 minutes after 17 UTC are plotted as an indicator of possible imbalance (Greybush et al. 2011). The initial pressure tendencies in L8 are substantially larger than for the large-scale analyses until timestep 25 (= 5 minutes), which is also the cycling interval of all experiments except L8SOCG20 and L32SOCG20.

5. Summary of results

The aim of this study on convective scale data assimilation was to assess how errors grow in 3-hour ensemble forecasts from analysis ensembles with precision on differing length scales. Data assimilation of long-lived and organized convective systems was performed under the assumption of a perfect model in order to focus on the intrinsic predictability of the storm systems. An LETKF system with 50 members was used in an idealized OSSE testbed with simulated Doppler radar observations of reflectivity and radial wind. The reference and ensemble storms were triggered randomly in the convection-permitting COSMO-model with $\Delta x_{model} = 2$ km, using radiative forcing and initial small-amplitude random noise in a horizontally homogeneous environment with periodic boundary conditions. The nature run and the ensemble all used the same model with the same horizontal resolution and the same initial sounding.

a. Assimilation schemes with different length scales

A reference experiment L8 was set up to reproduce the typical behavior of a convective EnKF DA system. Assimilation schemes were devised in which the localization radius of the observation error covariance matrix $\mathbf{R}$ was increased from 8 to 32 km (L8 and L32), the observation resolution was coarsened from 2 to 8 km as superobservations (SO), the analysis grid resolution of the LETKF was coarsened (CG) from 2 to 8 km, and the assimilation interval was increased from 5 to 20 minutes.

The background and analysis ensembles of each of the experiments succeeded in converging towards the observations, with the exception of L32 which suffered from undersampling due to the large localization radius in combination with full-resolution observations. The analyses of L8 had the lowest RMSE and displacement errors after 3 hours of cycling. The introduction of superobservations made the analyses of the L8SO-experiments less precise than L8 while enabling the L32SO-experiments to converge properly. Using a coarser analysis grid (CG experiments) in the SO-experiments made no significant difference, because the analysis computation was merely performed on the same coarse horizontal scale as the observations. An increased cycling interval of 20 minutes degraded the precision of the analyses, but resulted in less spurious gravity wave noise in the members through less frequent introduction of analysis increments. It was found that the spatially and temporally coarsened observation information in L32SOCG and L32SOCG20 rendered their filter solutions much less Gaussian than L8. The point-value distributions of these solutions were closer to the climatology of the simulated convective regime.

b. Ensemble forecasts from analyses with different scales

In the ensemble quantitative precipitation forecasts (QPF), the displacement error of the storms was measured by the object-based and field-based scores DAS and SAL. During the first hour, forecasts from the reference experiment L8 were clearly superior to those of the coarser schemes in terms of storm positions and internal structure. This advantage was lost in as little as half an hour of forecast lead time through the rapid error growth of small perturbations and the emergence of spurious convective cells. The coarsest scheme L32SOCG20 with large perturbations had much slower error growth and fewer spurious cells.

c. Imbalance and limited predictability

Inspection of the forecast fields suggested that rapid growth of spurious convective cells occurred in LETKF configurations where the analysis showed small-scale gravity wave noise. Surface pressure tendencies at the start of the free forecast were therefore evaluated for indications of dynamical imbalance introduced by the analysis increments. The smaller initial pressure tendencies in the L32SOCG20 experiment in comparison with L8 suggest less imbalance, although the analysis errors are larger in the coarse resolution experiment. The L32SOCG20 experiment also featured slower initial growth rates of RMSE. This was not caused by a need to spin up small scale motions since the analysis ensemble was always constructed from the full resolution background ensemble. Rather, the rapid error growth in L8 was associated with the appearance of small-scale spurious convective cells which may have been triggered by spurious gravity wave noise from the high resolution analysis increments as noted previously. As this noise was also present in clear air regions around the observed storms, the possiblity of triggering spurious convection by interfering gravity waves

(Hohenegger and Schär 2007) cannot be ruled out. The lower noise level of the experiments with a 20 minute cycling interval agrees with experiences in operational convective data assimilation (Seity et al. 2010). A hypothetical better data assimilation scheme that did not introduce spurious imbalance might have slower initial error growth than the L8 results here, but it is unlikely that the results would would be better than those of a coarser resolution analysis after 1-2 hours since some rapid development of new convective cores is seen in all configurations and the perturbations of the L32SOCG20 updraft positions from the analysis are already larger than coarse observation scale of 8 km within the first forecast hour.

6. Discussion and Outlook

a. On the methods used in this study

The forecast results displayed the limited predictability of the dynamics in large convective systems. In previous twin-experiment studies on error growth (Zhang et al. 2003, 2007; Hohenegger and Schär 2007; Done et al. 2012), uncorrelated noise with very small amplitude was used, which does not disrupt the model dynamics as the LETKF ensemble perturbations do. This study should therefore be seen in the tradition of predictability studies that used perturbations created by data assimilation systems (Kuhl et al. 2007; Aksoy et al. 2010; Craig et al. 2012).

One feature of the experiments of this paper is that the linear combinations of ensemble members $\mathbf{w}^{a(i)}$ are computed by the LETKF on different scales only in the transformed ensemble perturbation space, but the final analysis increments are added in the physical space and therefore contain also the smallest scales, as they are sampled from the full-resolution background ensemble members. An alternative approach would be coarsen not only the input-parameters of the LETKF-system are spatially coarsened but also the physical solution of the algorithm, in order to update larger physical scales only and leave the fine scales untouched. For example, Gao and Xue (2008) computed ensemble covariances from a background ensemble with 4 km horizontal resolution in an Ensemble Square Root Filter to update an analysis state of a 1 km resolution model. To mimic this for the present LETKF study, one could i) coarse-grain the background ensemble for the LETKF, ii) compute the analysis weights on this coarse scale and iii) apply the analysis update in physical space on the coarse scales only[1]. This might also reduce the imbalances that are introduced by the filter, and take into account that the effective resolution of numerical models is much lower than the grid-spacing implies (Skamarock 2004). A digital filter initalization (DFI) (Lynch and Huang 1992) could be helpful to reduce the initial gravity wave noise of the ensemble forecasts (Whitaker et al. 2008), although DFI might be problematic for a non-hydrostatic model as used here wherein the analysis states contain gravity wave motion explicitly.

As a general remark, the spatial and temporal predictability depends 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. These become addressable with horizontal resolutions of 250 m and less (Bryan and Morisson 2012; Craig and Dörnbrack 2008). 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 addition, using a different environmental sounding could result in different convective modes, such as mesocyclones, multicell storms and linear squall lines, and the predictability limit will probably be influenced by the degree of storm-internal organization (Lilly 1990; Aksoy et al. 2010). The quantitative time limit found in this study is specific to the atmospheric model configuration and the data assimilation algorithm used here and should not be generalized.

b. Idealized vs. operational convective data assimilation

The present study adressed the limitation of *intrinsic* convective predictability, assuming a perfect model. In the current convective DA systems (e.g. KENDA), model error may well

[1]This idea was discussed in personal communication with Chris Snyder.

be the largest factor limiting the *practical* predictability in convective QPF. In real world experiments, it may also happen that the radar DA tries to converge the analysis members towards convective modes that are not supported by the model physics and the predicted sounding (Stensrud and Gao 2010; Aksoy et al. 2010). On the other hand, in an operational model the predictability of convective system may be enhanced by effects of synoptic and orographic forcing (Hohenegger and Schär 2007; Craig et al. 2012).

c. Outlook

The results of this study showed that the impact of high resolution information in convective-scale data assimilation is limited by the intrinsic predictability of the flow to a time interval of a few hours at most, and probably also by the balance and noise properties of the initial states generated by the LETKF. These could potentially be mitigated through a priori constraints (Janjić et al. 2013) or post-processing like DFI. The observation error covariance matrix (Desroziers et al. 2005) is also likely to affect convergence and balance properties of the solution – choosing a larger observation error than actually added to the synthetic observations may result in less convergence but also in more compatible model states, as preliminary experiments (not shown) indicated. Another related issue is the initial ensemble generation for the filter. The random storms used here lead to large analysis increments in the first couple of cycles, which is likely to have detrimental effects on the dynamics (Lien et al. 2013; Kalnay and Yang 2010). Methods that accelerate the convergence of the filter, such as Running-In-Place, could also lead to more consistent states by the end of analysis period. However, it is not clear how successful these methods would be for convective-scale data assimilation, where the notions of "balanced" or "consistent" states cannot yet be precisely defined or measured.

A final issue is the difficultly of assimilating observations of precipitation due to their non-Gaussian climatological distribution, 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 anamorphosis of precipitation observations could help here to fulfill the Gaussian assumptions of the EnKF (Simon and Bertino 2009; Bocquet et al. 2010; Lien et al. 2013), also on the convective scales.

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 BMVI (Federal Ministry of Transport and Digital Infrastructure).

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.

Benjamin, S., S. Weygandt, and G. Manikin, 2012: Overview of the rapid refresh rap. URL `http://ruc.noaa.gov/pdf/RAPbrief.NCEP-Dir-20Mar2012.pdf`.

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.

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

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. and A. Dörnbrack, 2008: Entrainment in Cumulus Clouds: What Resolution is Cloud-Resolving? *J. Atmos. Sci.*, **65 (12)**, 3978–3988, doi:10.1175/2008JAS2613.1.

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.

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.

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

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.

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

Holland, B. and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **139 (672)**, 758–770, doi:10.1002/qj.2006.

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

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. *Mon. Wea. Rev.*, **142 (2)**, 755–773, doi:10.1175/MWR-D-13-00056.1.

Janjić, T., L. Nerger, A. Albertella, J. Schrter, and S. Skachko, 2011: On domain localization in ensemble-based kalman filter algorithms. *Mon. Wea. Rev.*, **139 (7)**, 2046–2060, doi:10.1175/2011MWR3552.1.

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.

Lynch, P. and X.-Y. Huang, 1992: Initialization of the HIRLAM Model Using a Digital Filter. *Mon. Wea. Rev.*, **120 (6)**, 1019–1034, doi:10.1175/1520-0493(1992)120⟨1019:IOTHMU⟩2.0.CO;2.

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.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bnard, F. Bouttier, C. Lac, and V. Masson, 2010: The arome-france convective-scale operational model. *Mon. Wea. Rev.*, **139 (3)**, 976–991, doi:10.1175/2010MWR3425.1.

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.

Simon, E. and L. Bertino, 2009: Application of the gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the enkf: a twin experiment. *Ocean Science*, **5 (4)**, 495–510, doi:10.5194/os-5-495-2009.

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.

Weygandt, S. S., S. Benjamin, T. G. Smirnova, and J. M. Brown, 2008: Assimilation of radar reflectivity data using a diabatic digital filter within the rapid update cycle. *12th Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, New Orleans, LA, AMS+, URL `https://ams.confex.com/ams/88Annual/techprogram/paper\_134081.htm`.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the ncep global forecast system. *Mon. Wea. Rev.*, **136 (2)**, 463–482, doi:10.1175/2007MWR2018.1.

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., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64 (10)**, 3579–3594, doi:10.1175/JAS4028.1.

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

Table 1. Length scales used in the assimilation experiments, as described in Section 2. $r_{Loc,h}$ is the cutoff length of the horizontal covariance localization function, Δx_{obs} is the horizontal resolution of observations, Δx_{ana} is the horizontal resolution of the analysis grid, Δt_{ass} is the assimilation interval between two subsequent analyses.

	$r_{Loc,h}$ (km)	Δx_{obs} (km)	Δx_{ana} (km)	Δt_{ass} (min)
L8	8	2	2	5
L8SO	8	8	2	5
L8SOCG	8	8	8	5
L8SOCG20	8	8	8	20
L32	32	2	2	5
L32SO	32	8	2	5
L32SOCG	32	8	8	5
L32SOCG20	32	8	8	20

List of Figures

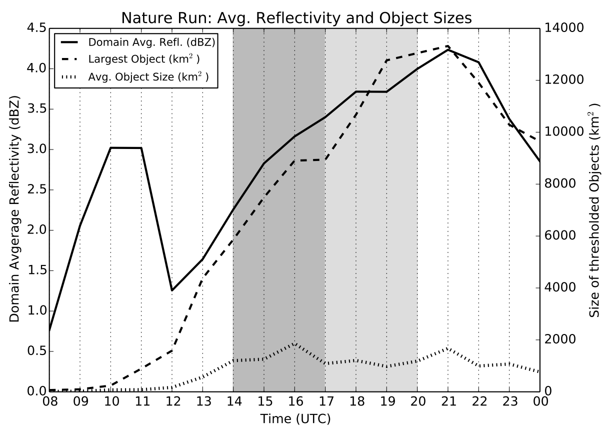

Fig. 1. Time-series for the Nature Run (averaged over five realizations): Domain-average of the composite reflectivity in dBZ (solid), together with average horizontal (dash-dotted) and maximum (dashed) size of the rain-objects, thresholded to > 10 dBZ. Assimilation window is between 14 and 17 UTC (shaded dark gray), forecast window is between 17 and 20 UTC (shaded light grey). Variation of the object-sizes can be due to merger of anvils of separate convective systems.

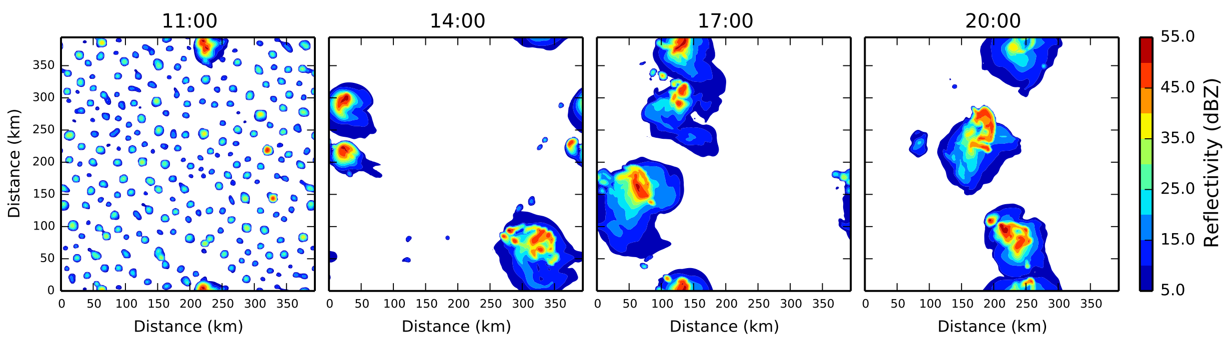

Fig. 2. Snapshots of composite reflectivity from Nature Run 03.

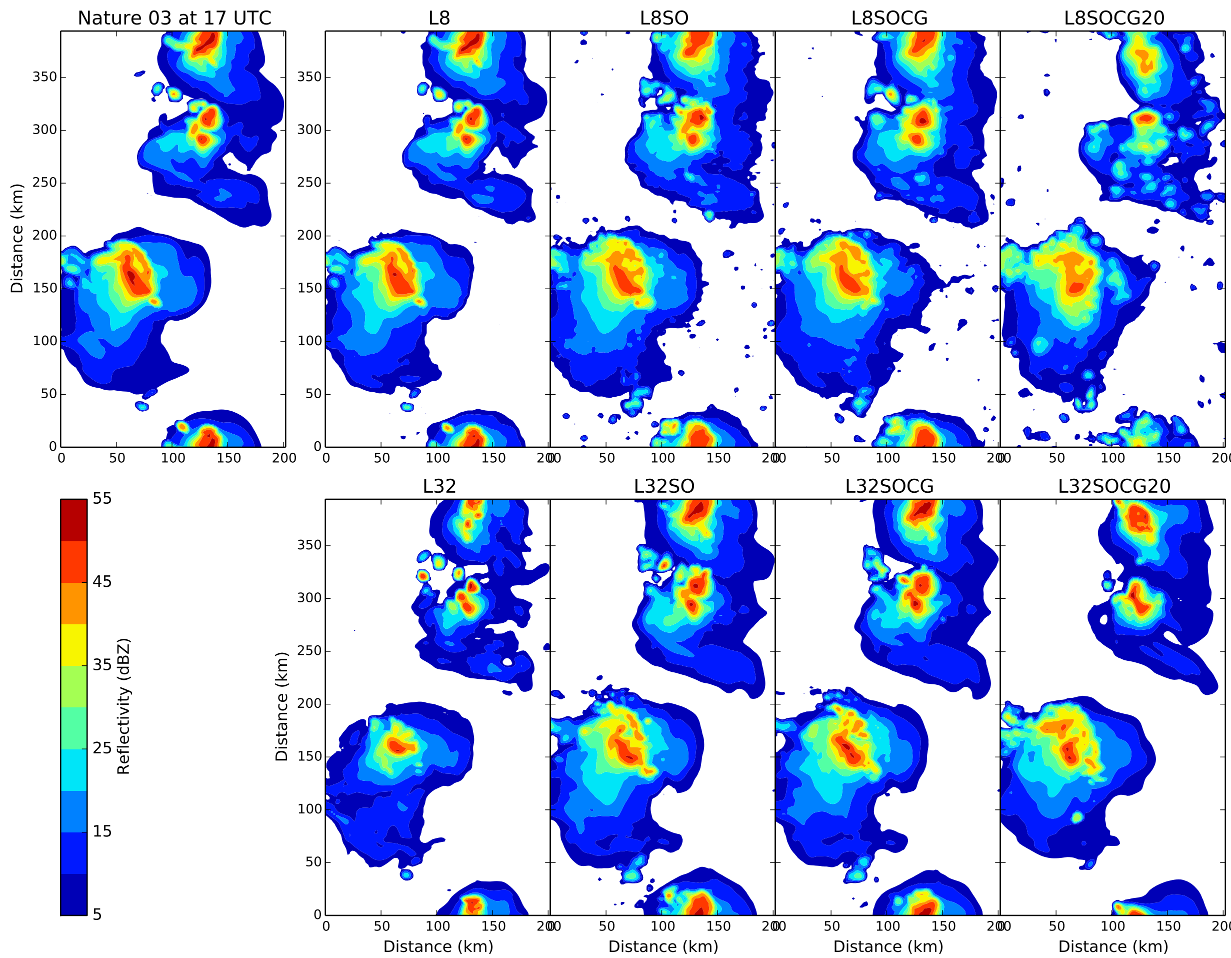

FIG. 3. Composite Reflectivity of Nature Run 03 (cf. Fig. 2 at 17 UTC) and of the corresponding **Analysis Ensemble Means** of the different scale-experiments at the last analysis time 17 UTC.

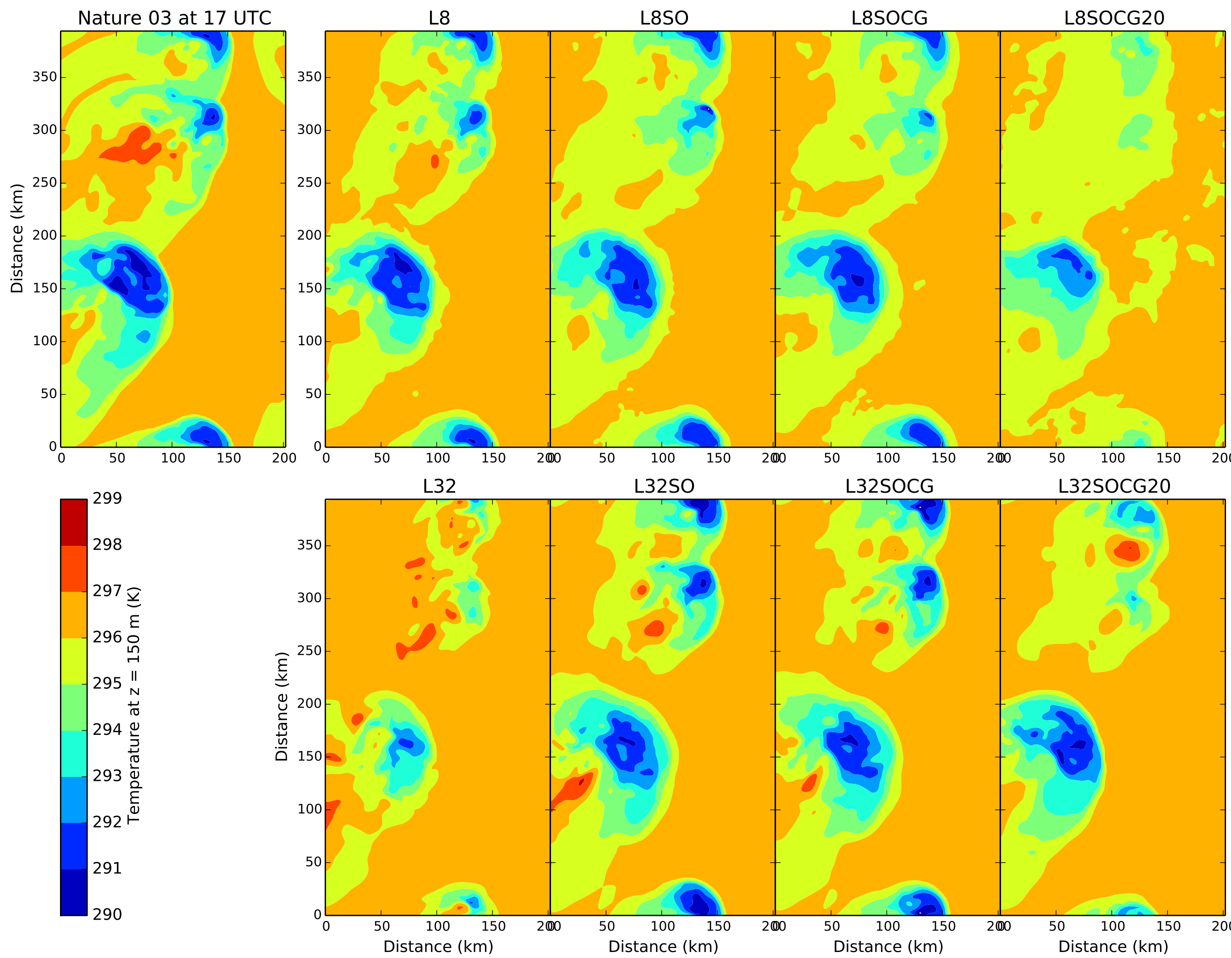

FIG. 4. As Fig. 3, but showing the temperature T at $z = 150$ m.

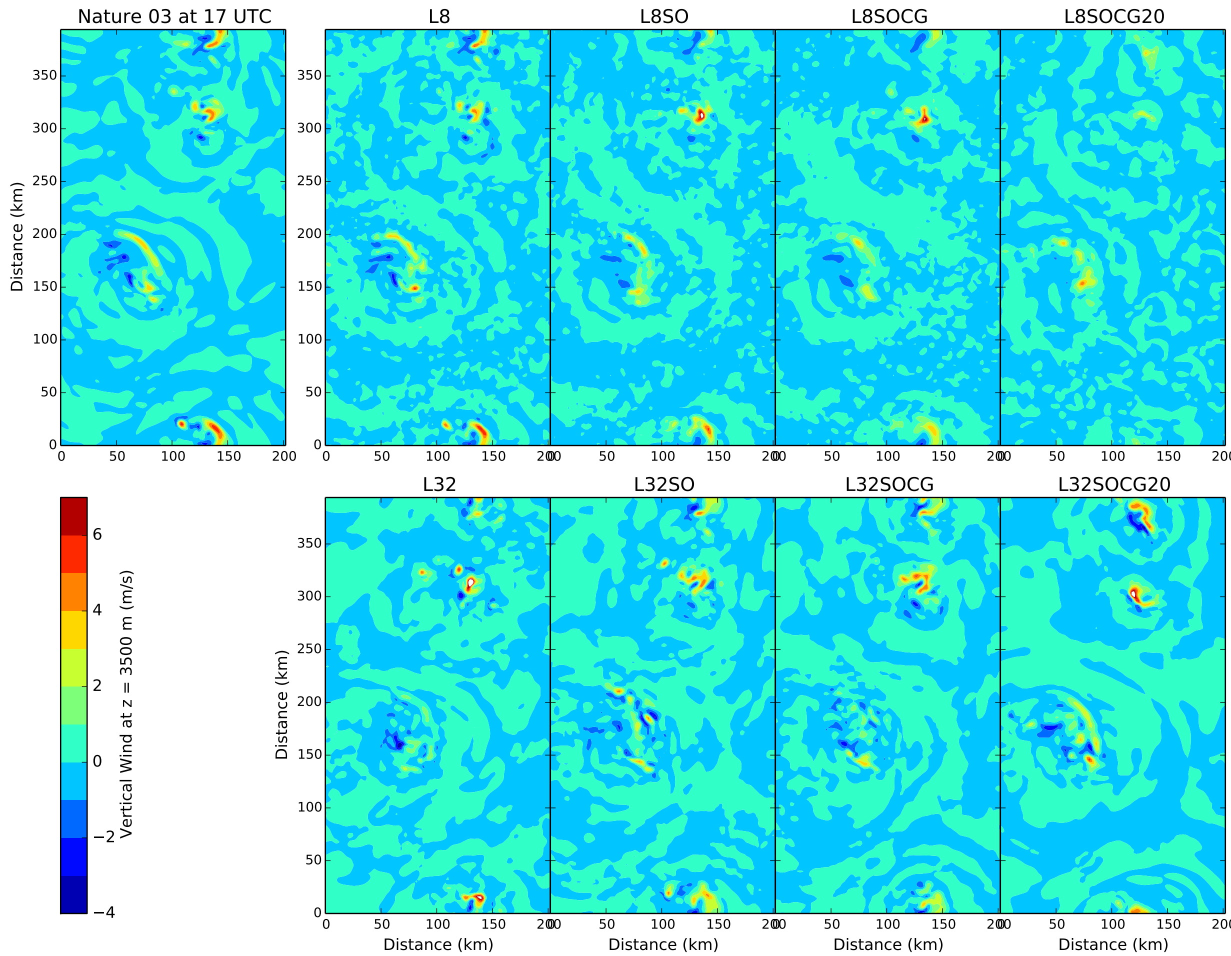

FIG. 5. As Fig. 3, but showing the vertical velocity W at $z = 3500$ m.

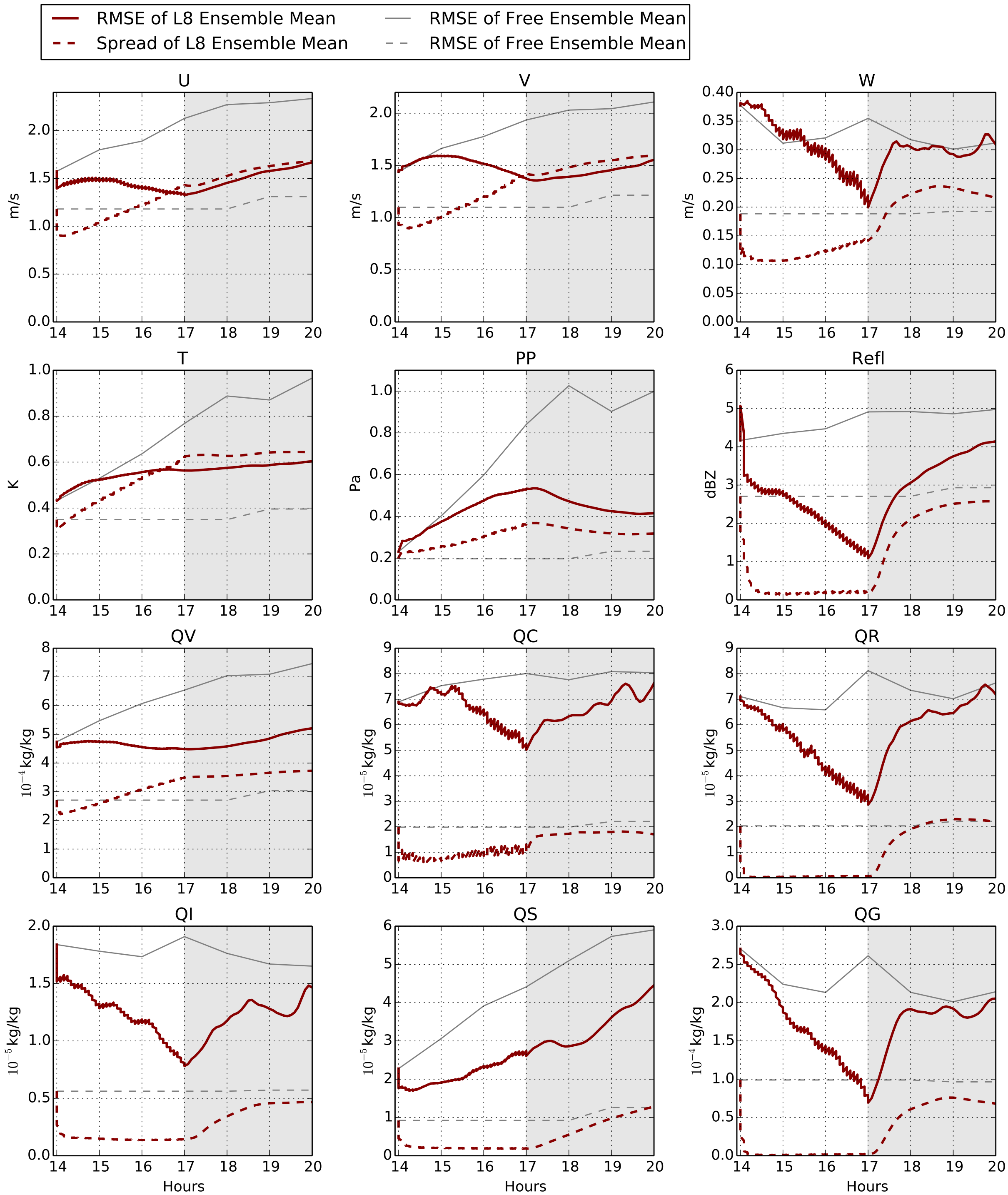

FIG. 6. RMSE and spread of the ensemble mean of L8 through the assimilation (14-17 UTC) and forecast (17-20 UTC) phases of the experiment, with gray shading indicating the forecast phase. The gray lines show the RMSE and spread of the mean of the free ensemble without any assimilation. All gridpoints are evaluated. The error values are averaged over five repetitions of the experiment.

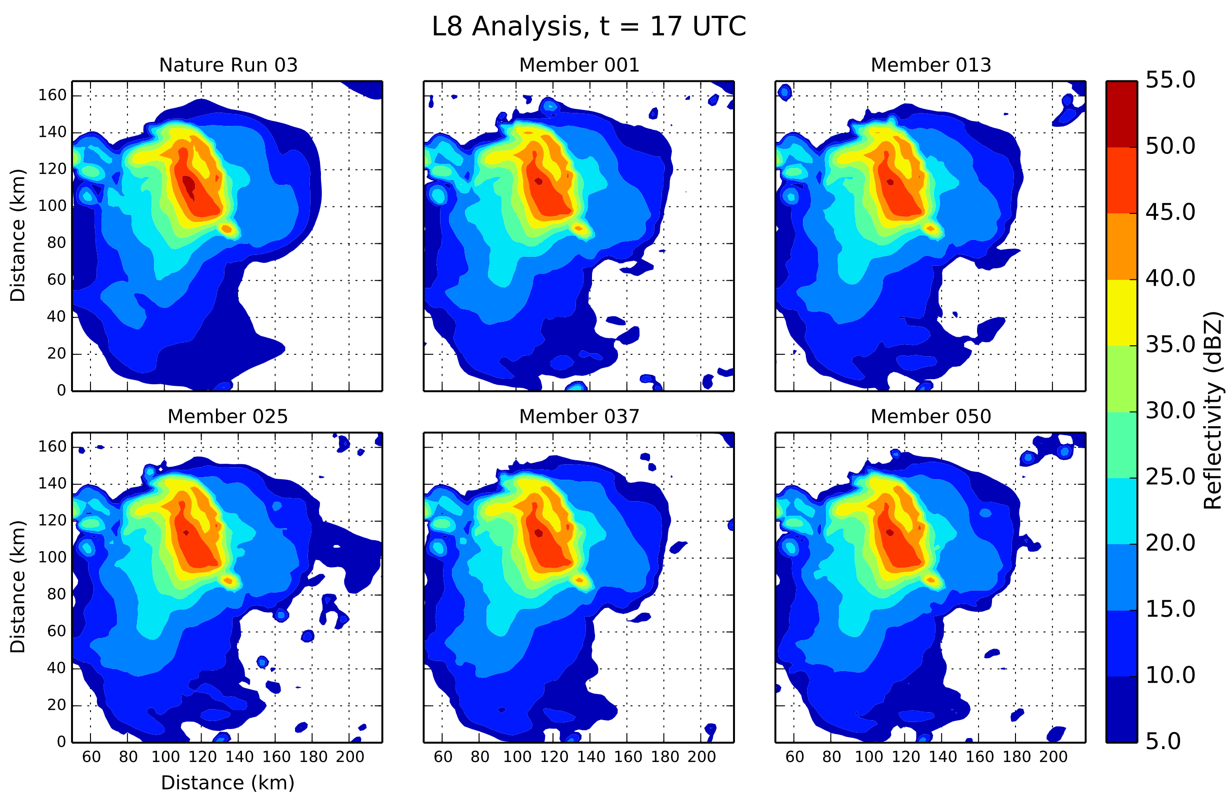

Fig. 7. Composite reflectivity of Nature Run 03 (cutout of the domain in Fig. 2) and **Analysis Ensemble Members** 1,13,25,37,50 of L8 at the last assimilation time 17 UTC

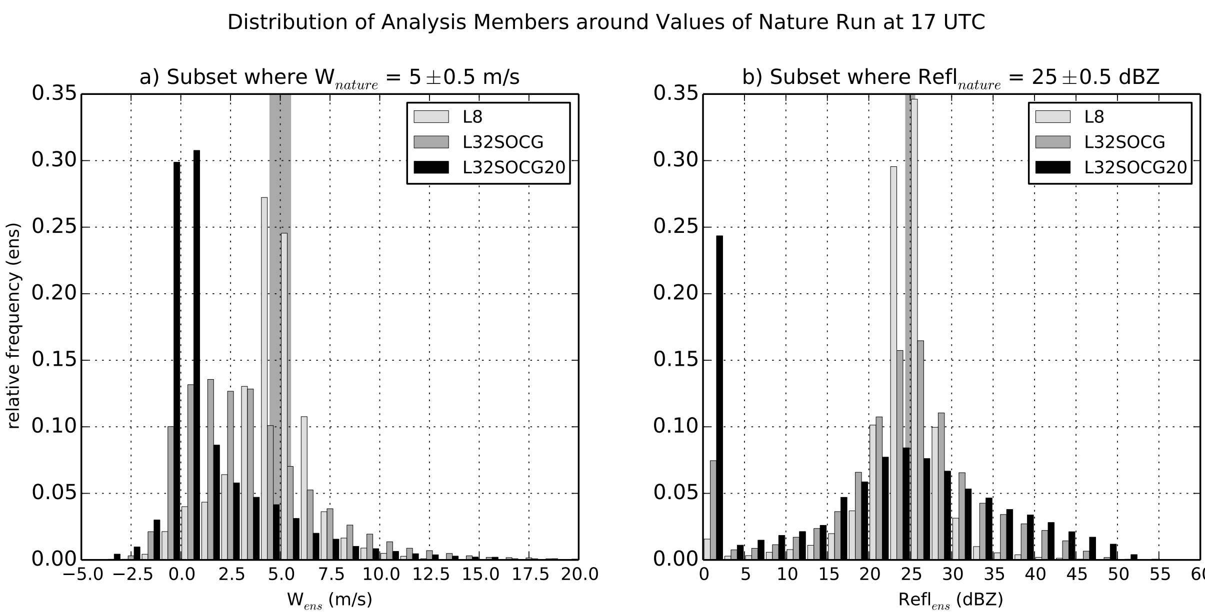

FIG. 8. Relative frequencies of model values within the analysis ensemble-members of L8 (light grey), L32SOCG (gray) and L32SOCG20 (black) at the last analysis time of 17 UTC (cf. Figs. 3, 5, 7, 13), averaged over the five repetitions of the experiments. Distributions are computed for regions inside storms defined by a) gridpoints with updrafts where W of the Nature Run is 5 ± 0.5 m s^{-1} and b) where $Refl_{nature} = 25 \pm 0.5$ dBZ. The bar width is 1/3 of the binwidth.

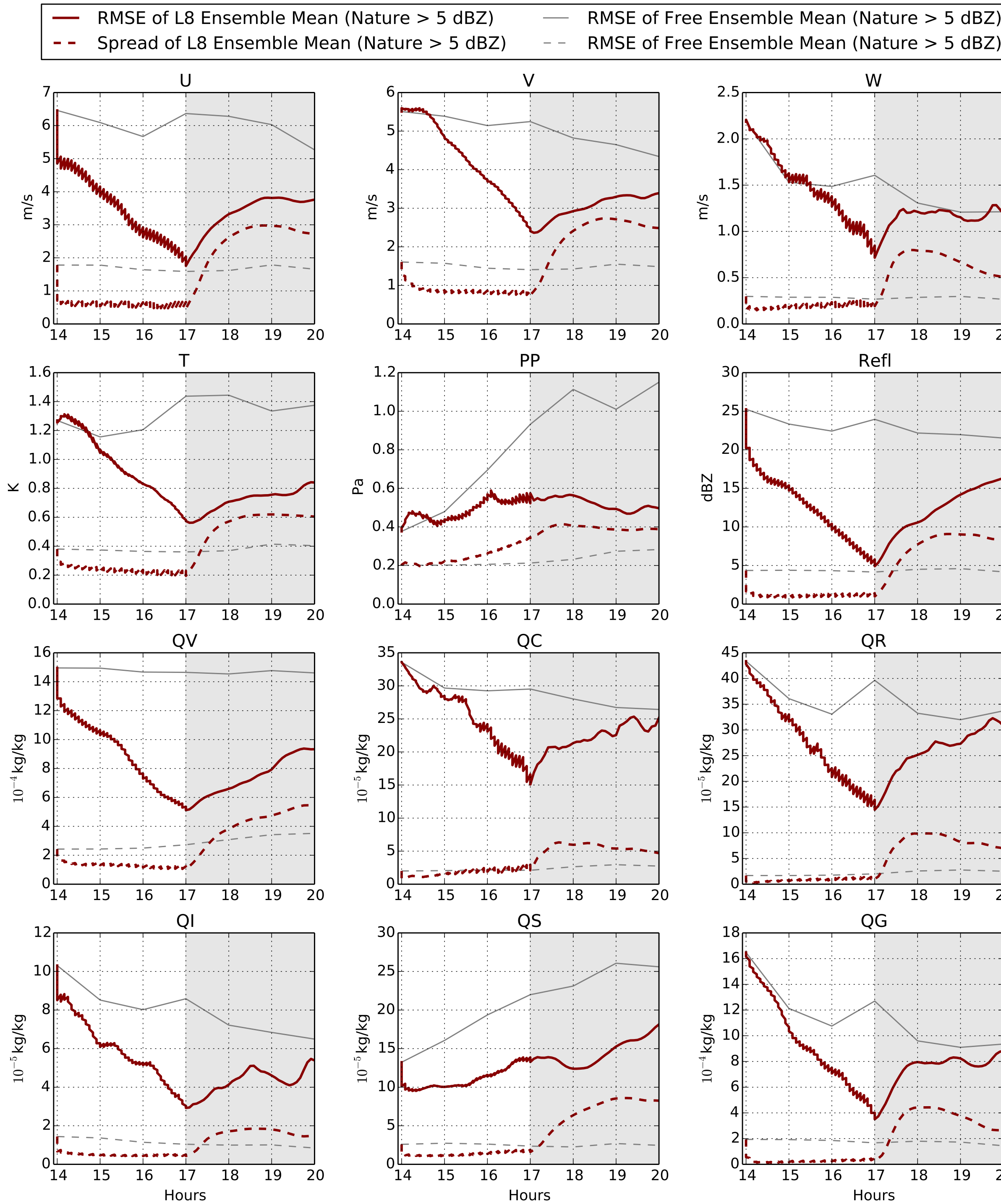

FIG. 9. As Fig. 6, but only the subset of gridpoints is evaluated where the volume reflectivity of the Nature Run exceededs the observation threshold of 5 dBZ.

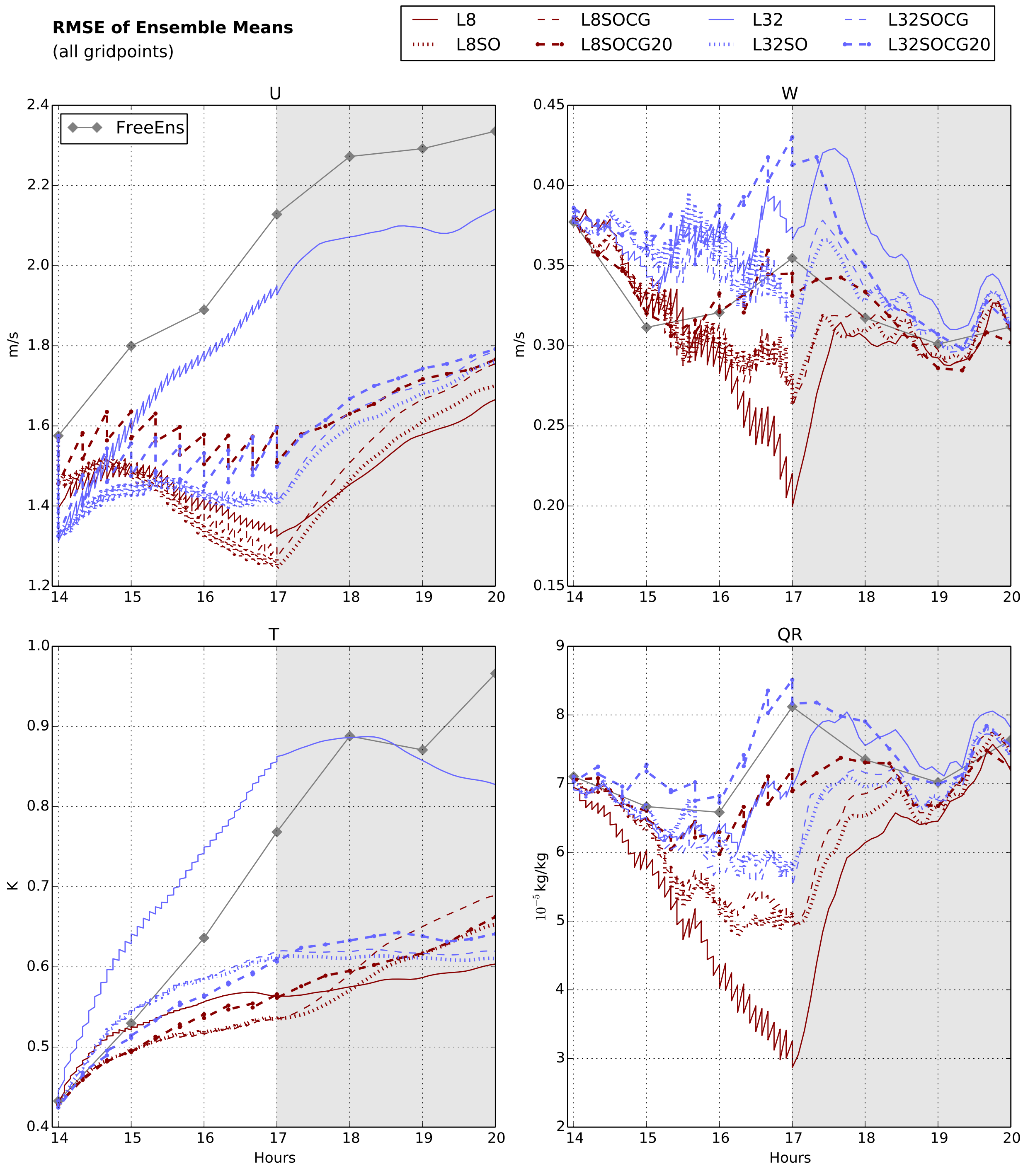

FIG. 10. As Fig. 6, but now showing RMSE of the ensemble means for all experiments (spread not shown). All gridpoints are evaluated. The error values are averaged over five repetitions of the experiments.

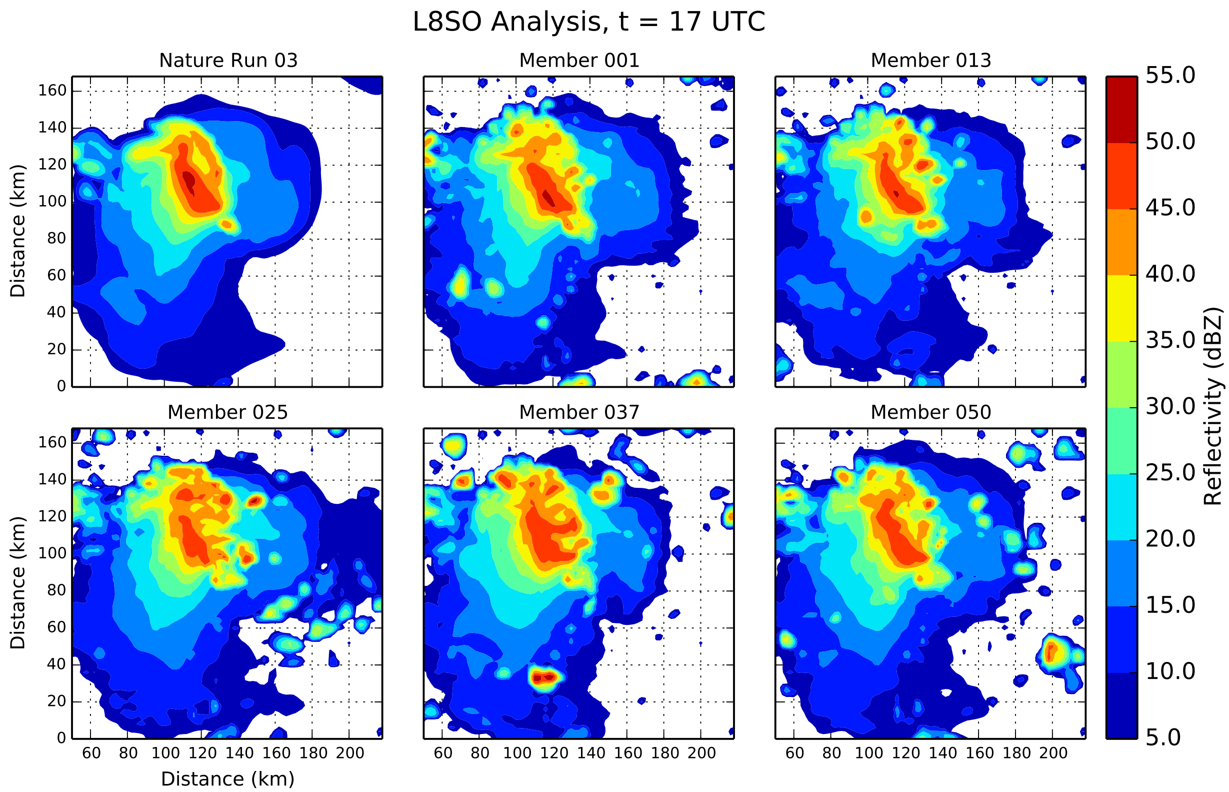

FIG. 11. As Fig. 7, but for L8SO.

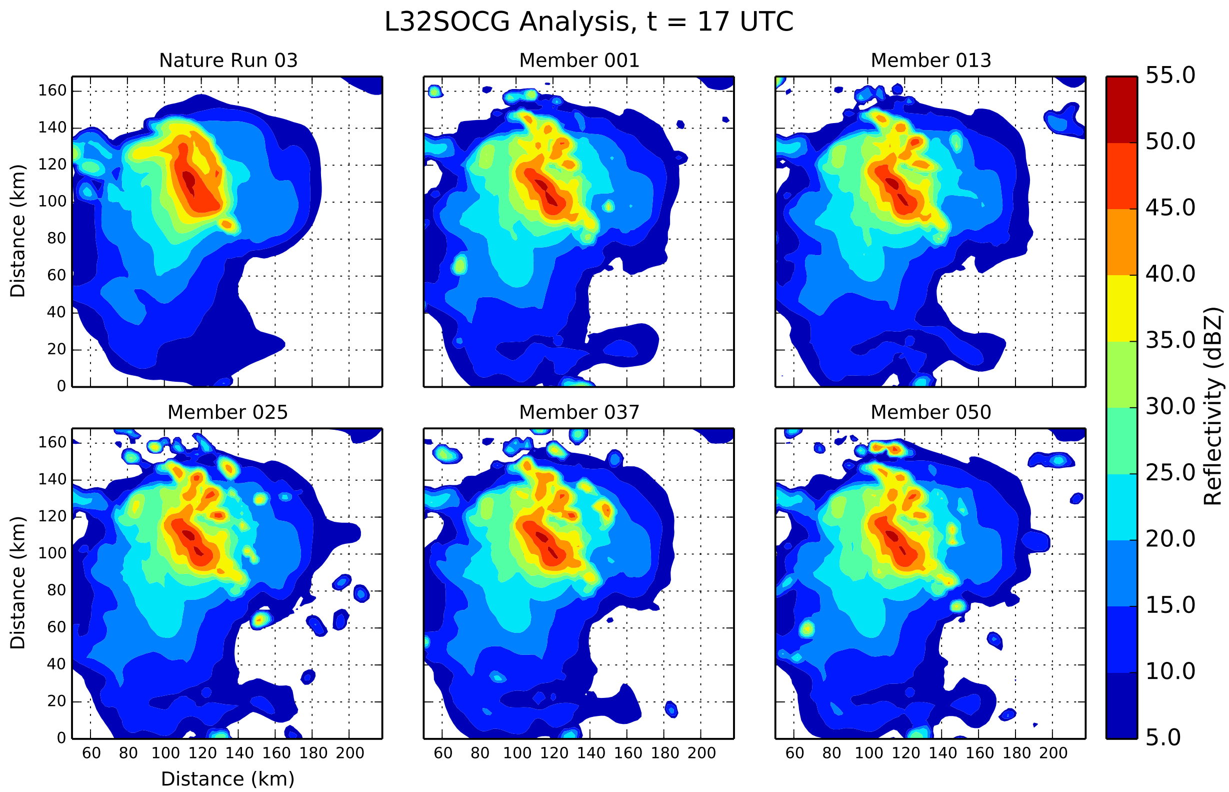

Fig. 12. As Fig. 7, but for L32SOCG.

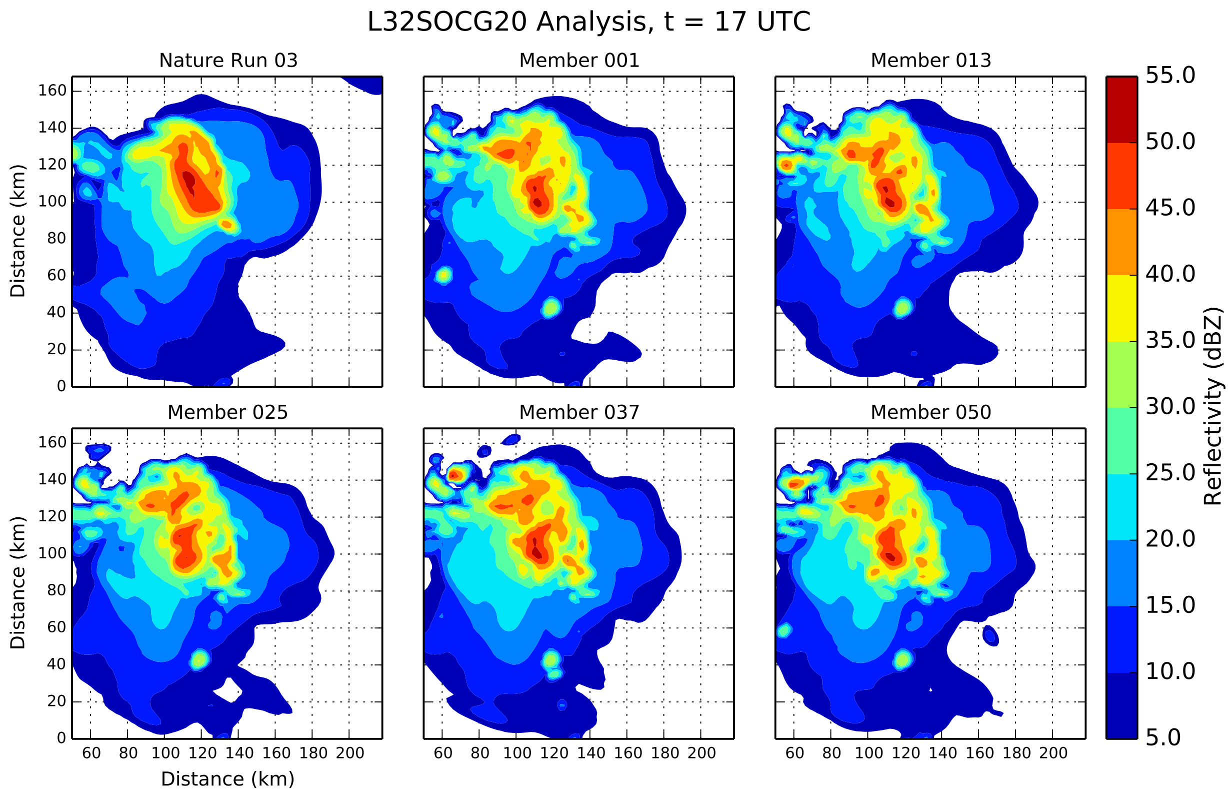

FIG. 13. As Fig. 7, but for L32SOCG20.

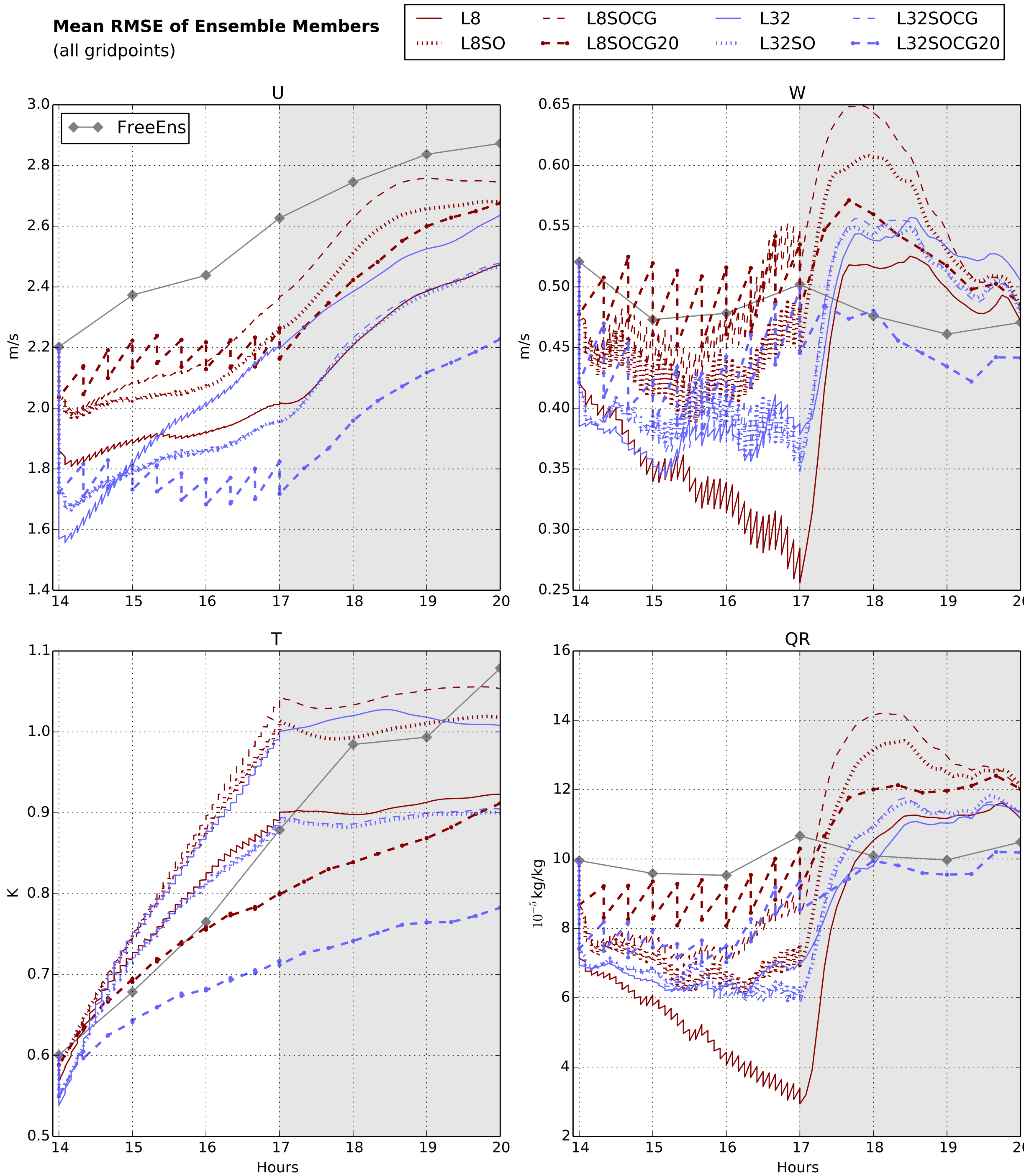

FIG. 14. As Fig. 10, but showing the mean RMSE of the individual ensemble members of all experiments (see equation (7)).

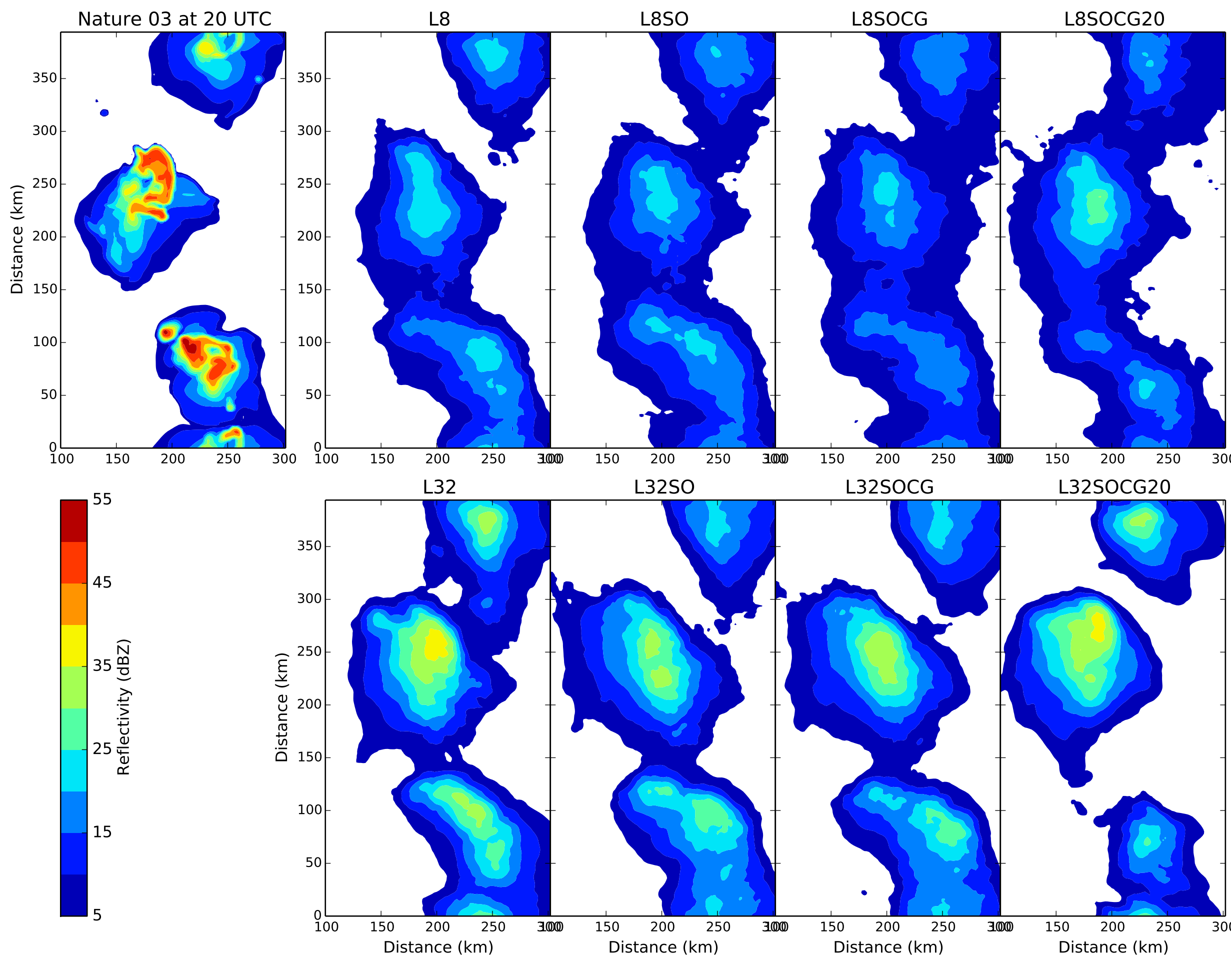

FIG. 15. As Fig. 3, but showing Composite Reflectivity of Nature Run 03 (cf. Fig. 2 at 20 UTC), and the **Forecast Ensemble Means** of the different scale-experiments at 20 UTC, after 3 hours of ensemble forecast.

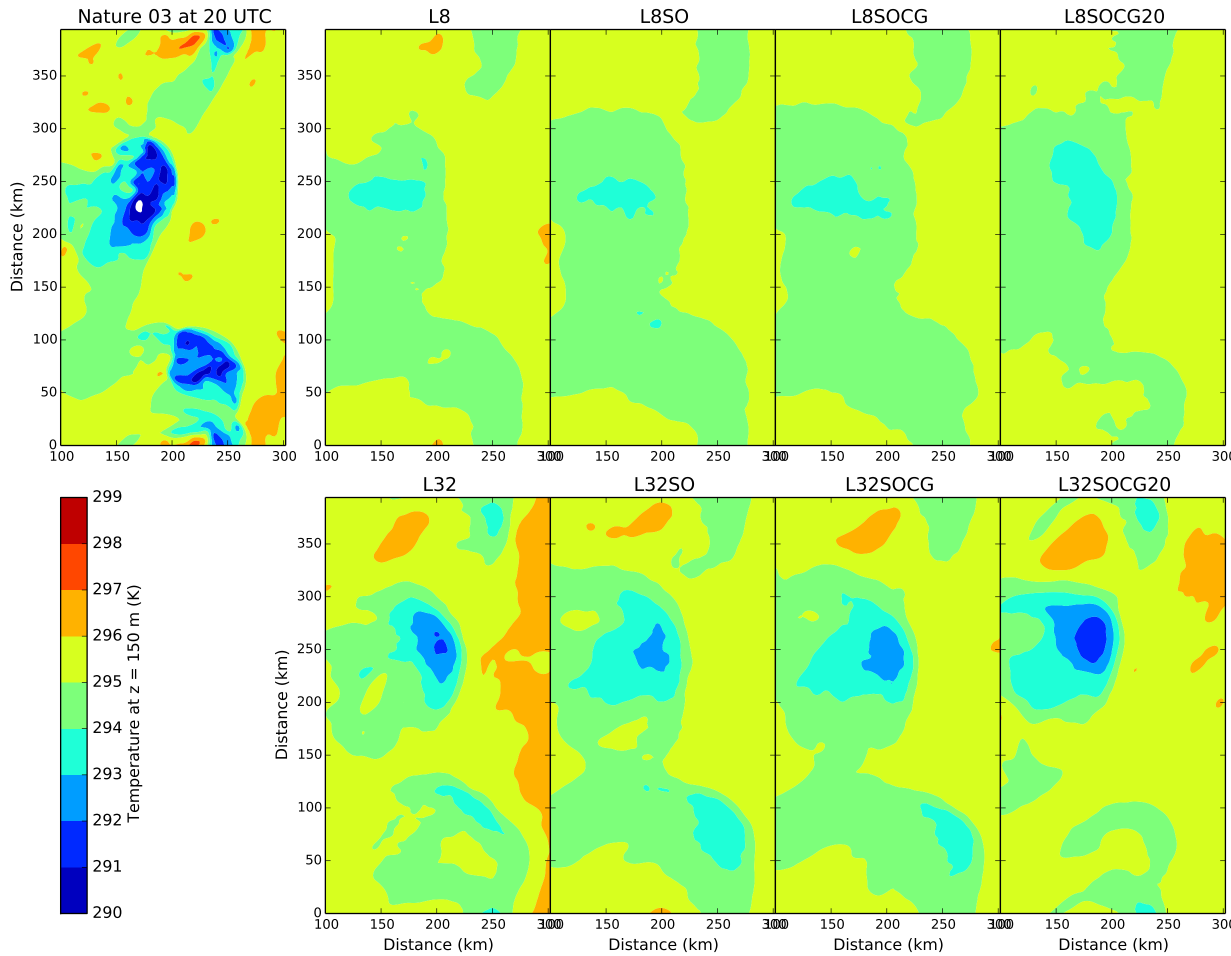

FIG. 16. As Fig. 15, but showing the temperature T at $z = 150$ m.

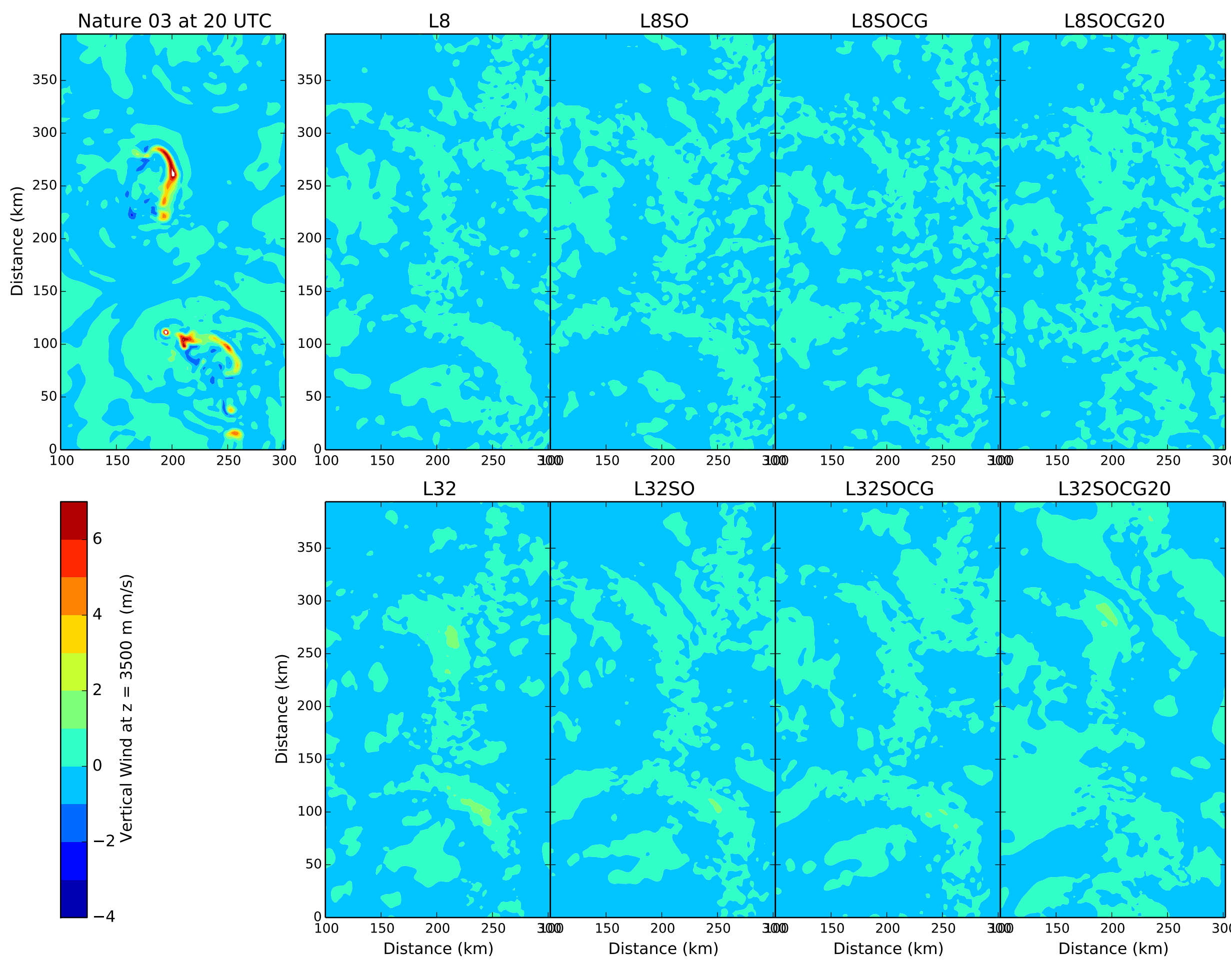

FIG. 17. As Fig. 15, but showing the vertical velocity W at $z = 3500$ m.

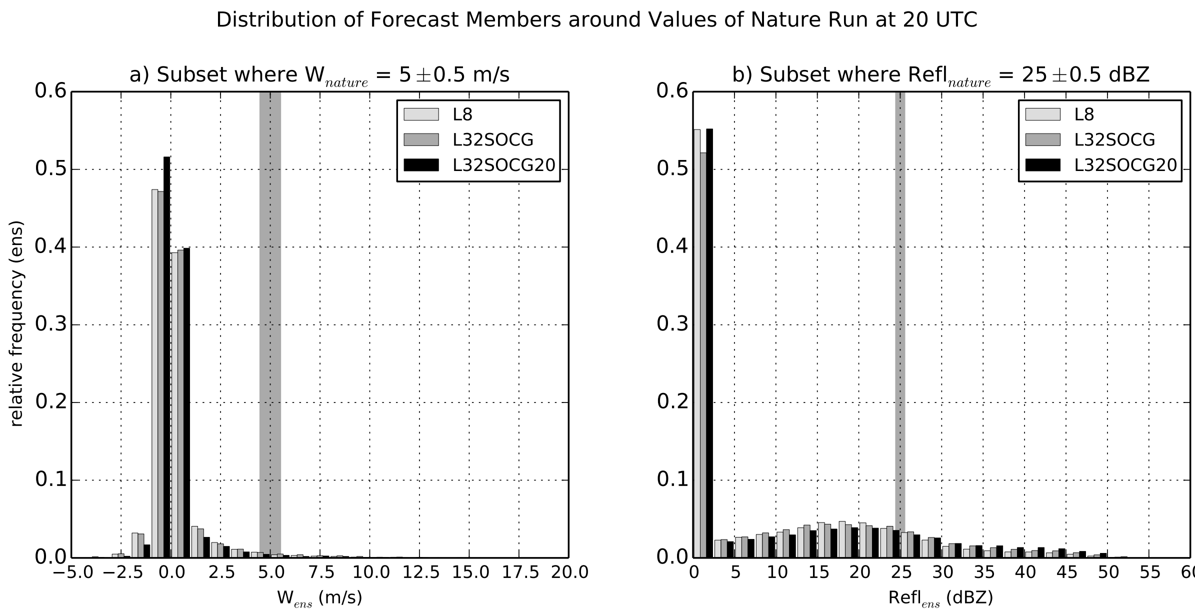

FIG. 18. As Fig. 8, but for the forecast ensemble-members at 20 UTC after 3 hours of ensemble forecast (cf. Figs. 15 and 17)

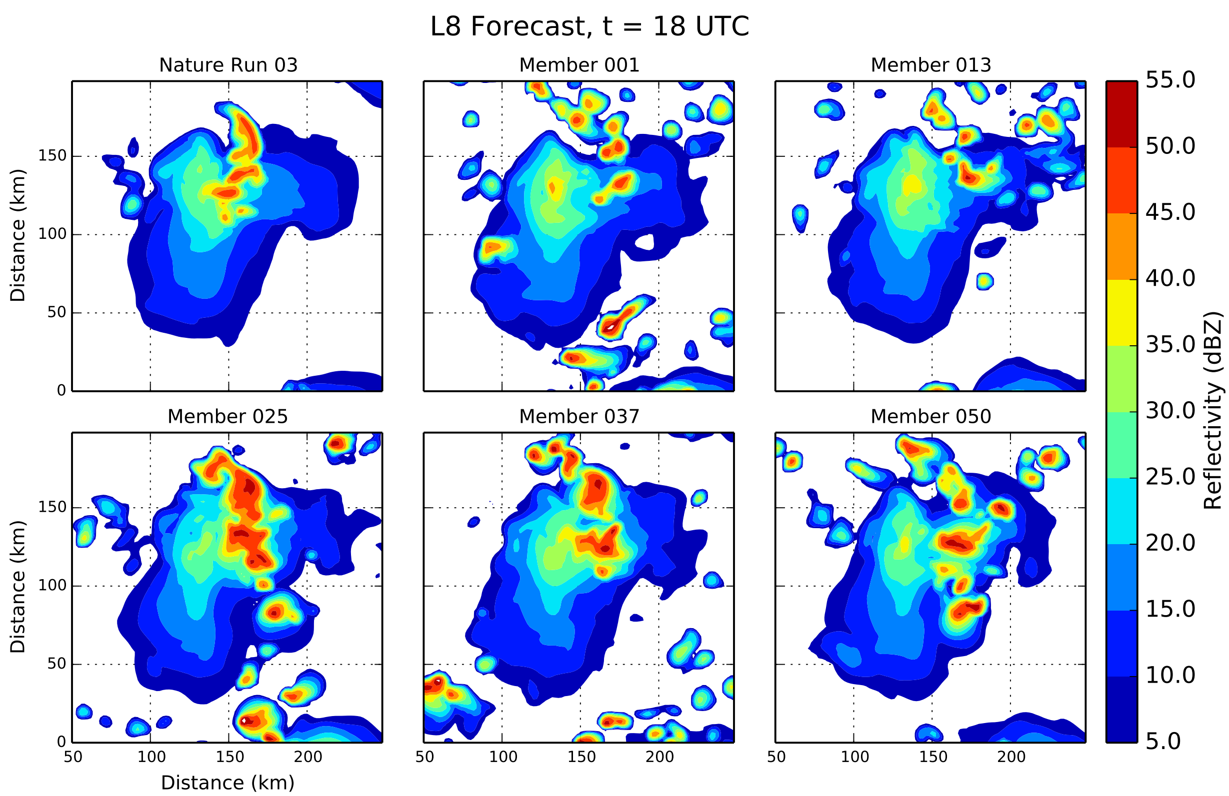

Fig. 19. Composite reflectivity of Nature Run 03 (cutout of the domain) and **Forecast Ensemble Members** 1,13,25,37,50 of L8 at 18 after 1 hour of forecast.

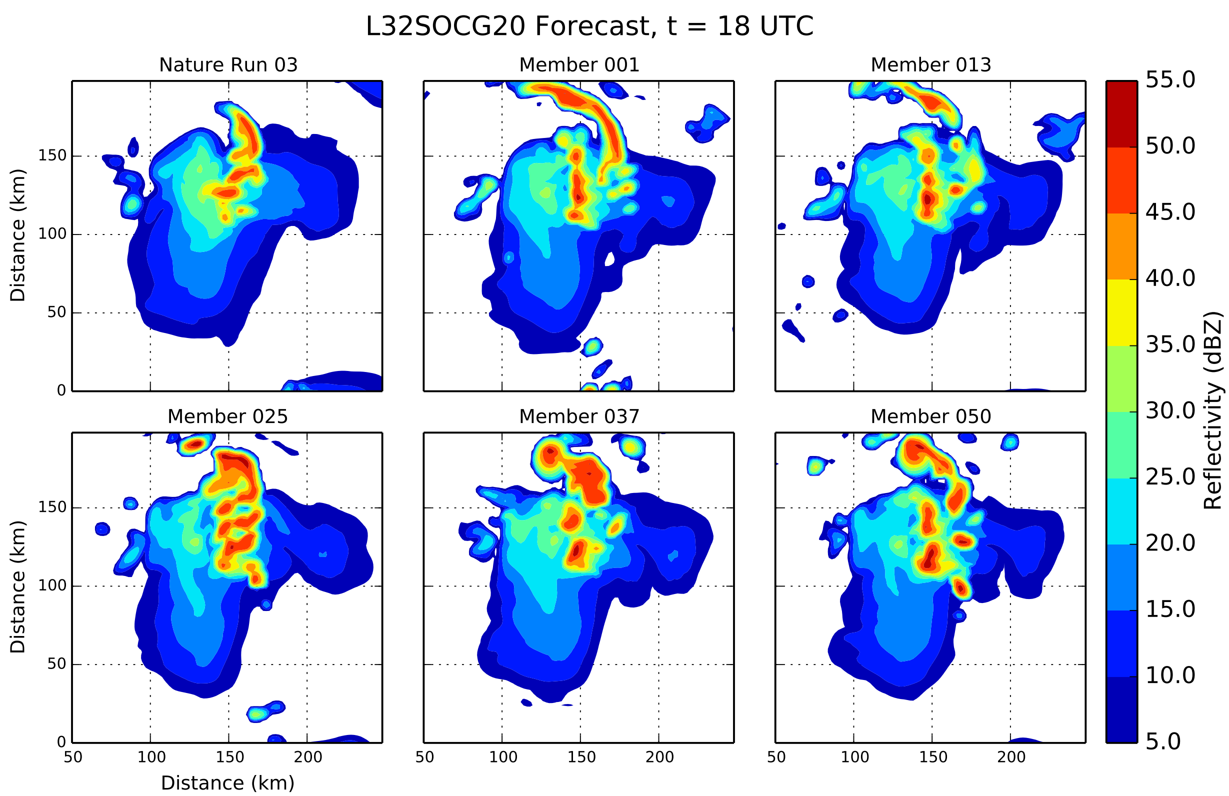

FIG. 20. As Fig. 19, but for L32SOCG20.

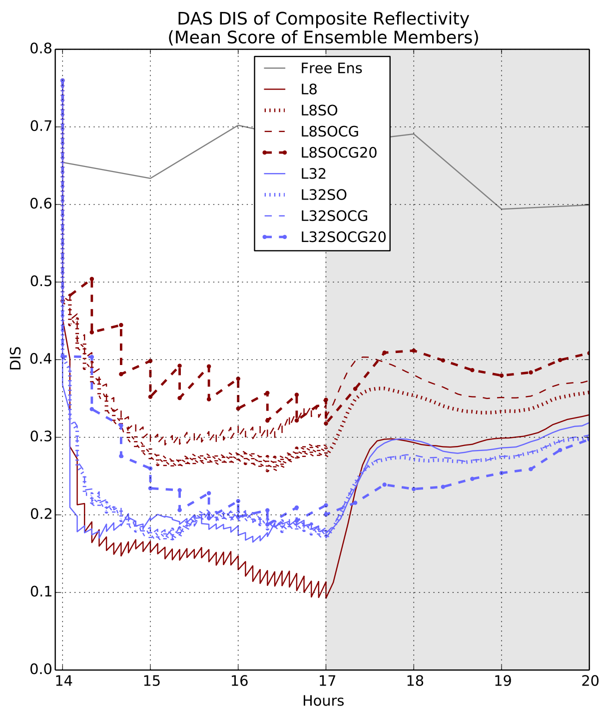

FIG. 21. DIS-component of the DAS-score applied to the Composite Reflectivity field thresholded by 10 dBZ. Displayed is the mean DIS-score of the ensemble members of all experiments, with the Nature Run as reference. The score is averaged over the five random repetitions of all experiments. (A value of DIS $= 0$ is a perfect match.)

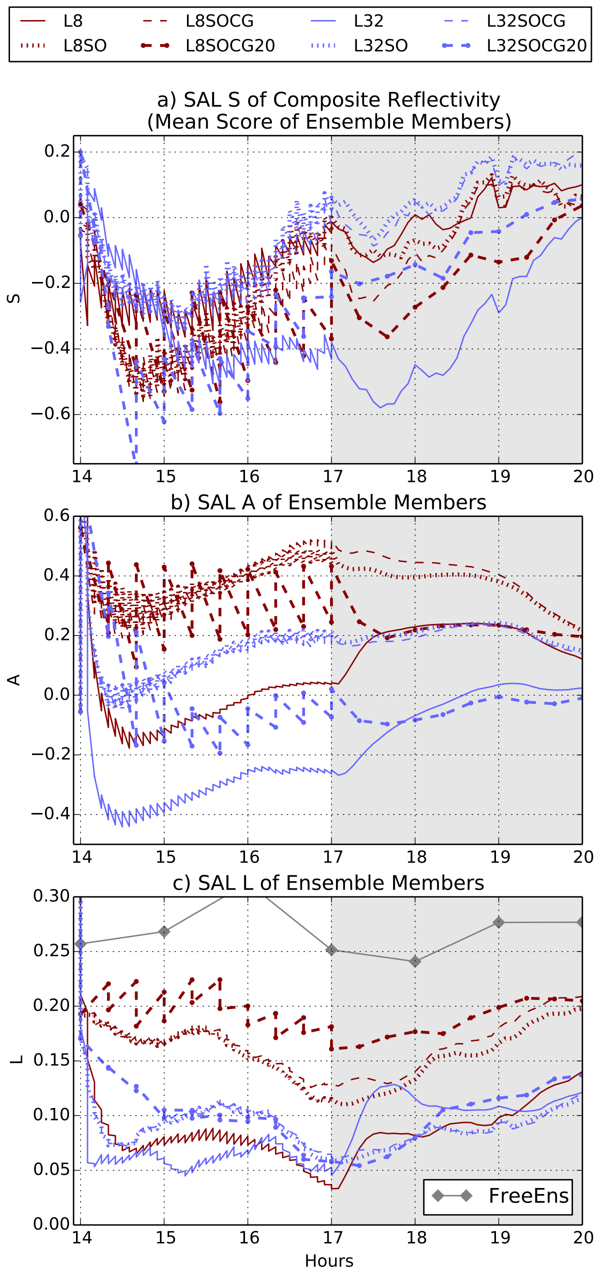

Fig. 22. SAL score components S, A and L for all experiments, with the Nature Run as reference (see text). The scored field is the Composite Reflectivity, the scores are averaged over the five random repetitions of all experiments.

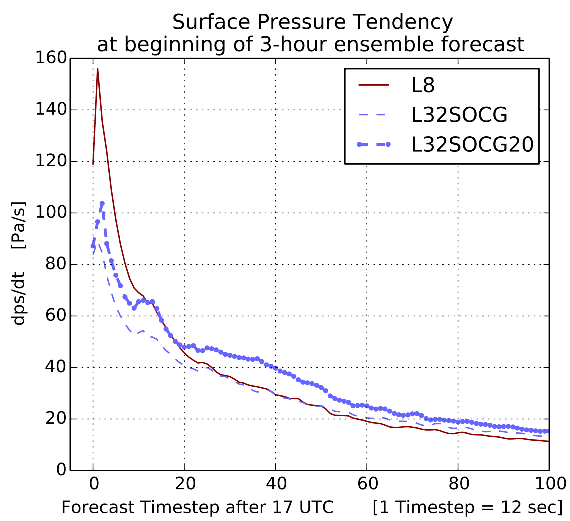

FIG. 23. Surface pressure tendencies dp_s/dt (domainwide maximum) of the first member (averaged over five repetitions of the experiments) within the first 20 minutes after starting the 3 hour ensemble forecast. Shown are L8, L32SOCG and L32SOCG20.