On the Use of Radar and Aircraft Data in Ensemble Data Assimilation of Convection for Non-hydrostatic Numerical Weather Prediction

Heiner Lange

München, 2016

On the Use of Radar and Aircraft Data in Ensemble Data Assimilation of Convection for Non-hydrostatic Numerical Weather Prediction

Heiner Lange

Dissertation
an der Faculty of Physics
der Ludwig–Maximilians–Universität
München

vorgelegt von
Heiner Lange
aus München

München, den 29.09.2016

Erstgutachter: Prof. Dr. George Craig und PD Tijana Janjić
Zweitgutachter: Prof. Dr. Bernhard Mayer
Tag der mündlichen Prüfung: 30. November 2016

Contents

Zusammenfassung

Anfangszustände (sog. Analysen) für die numerische Wettervorhersage von Gewittern und mesoskaligen konvektiven Systemen können mithilfe von Ensemble-Datenassimilation (DA) generiert werden. Reflektivitäts- und Radialwindbeobachtungen von Wetterradaren sind hierzu hilfreich, da sie die Gewitter zeitlich und räumlich hochaufgelöst beobachten. Gewitter besitzen eine komplexe innere Struktur und sind gleichzeitig eingebettet in die umgebende Atmosphäre, die durch das Vertikalprofil von Temperatur, Wind und Feuchte beschrieben wird. Die vorliegende Dissertation setzt es sich zum Ziel, Zusammenhänge zwischen beobachteter Dynamik und numerischer Modellvorhersage aufzudecken. Zudem wird die Frage bearbeitet, inwiefern die begrenzte praktische und intrinsische Vorhersagbarkeit den Anstrengungen der konvektiven DA entgegenwirkt.

Im chaotischen System der Atmosphäre wachsen Fehler der Anfangszustände während der Vorhersage rasch an und begrenzen die Vorhersagbarkeit. Kleinskalige Fehler und solche mit kleiner Amplitude intensivieren sich dabei relativ gesehen schneller als größere Fehler. Dementsprechend können Gewittervorhersagen, die von Analysen mit größeren Fehlern gestartet wurden, nach einiger Zeit dieselbe Qualität besitzen wie Vorhersagen von Analysen mit kleineren Fehlern. Um dies zu testen, wurden simulierte Radardaten mithilfe eines Ensemble Kalman Filters (EnKF) und unter der Annahme eines perfekten Modells in das konvektionsauflösende COSMO-Modell assimiliert, bei einer Modellauflösung von 2 km. Die räumliche und zeitliche Auflösung der Beobachtungsdaten wurde von 2 auf 8 km und von 5 auf 20 min variiert, zusammen mit einer Vergrößerung des räumlichen Einflussradius von 8 auf 32 km. Die Vorhersagequalität auf Basis solcher groben Analysen erwies sich nach drei Stunden als vergleichbar mit Vorhersagen auf Basis von Analysen, die mit hochaufgelösten Beobachtungen produziert wurden. Dieser Zeitraum wurde als obere Grenze für die Vorhersagbarkeit in dieser Situation gewertet. Im Falle der hochaufgelösten Beobachtungen waren die Analysen mit Störungen des vertikalen Windfelds (sog. *noise*) behaftet, die in Form von Schwerewellen die Dynamik der beobachteten Gewitter beeinflussten. Dies führte zum Auftreten von übermäßiger Konvektion an falschen Orten (sog. *spurious convection*). Zur Beschreibung dieses *noise* wurde eine Raum-Zeit-Korrelation von verschiedenen Modellfeldern berechnet. Hierbei zeigte sich eine Entkopplung der neu auftretenden *spurious convection* von den Böenfronten der beobachteten Gewitter, und das gleichzeitige Auftreten von übermäßiger Variabilität des vertikalen Windfelds im nahen und fernen Umfeld der beobachteten Konvektion.

Zur Verbesserung des atmosphärischen Vertikalprofils wurden reale Flugzeugbeobach-

tungen in das vom Deutschen Wetterdienst (DWD) entwickelte, prä-operationelle System KENDA (Kilometre-scale Ensemble Data Assimilation) assimiliert, welches das deutsche COSMO-DE-Modell mit 2.8 km Auflösung mit einem EnKF verbindet. Diese Flugzeugbeobachtungen stammen aus dem Mode-S-System der Flugsicherung und enthalten Beobachtungen von Wind und Temperatur, wobei die Mode-S-Daten eine zehnmal höhere Datendichte aufweisen als die bisher verwendeten Daten des Systems AMDAR (Aircraft Meteorological Data Relay). Nach Implementierung der Mode-S-Assimilation konnte eine Fehlerverringerung in dreistündigen Vorhersagen des Vertikalprofils erreicht werden, wodurch auch eine bessere Gewittervorhersage erwartbar ist. Um einer operationellen Nutzung der Mode-S-Daten den Weg zu bereiten, wurden die Auswirkungen der stark erhöhten Datenmenge auf Parameter des KENDA-Systems wie Lokalisierung und *ensemble spread* untersucht.

Der Aufwindstrom von Gewitterzellen bewirkt eine divergente und konvergente Horizontalzirkulation in den umgebenden Bereichen, die den Kern des Gewitters umgibt. Um den Einfluss der Radarbeobachtungstypen der Reflektivität und des Radialwinds auf Aufwindstrom und umgebende Zirkulation zu untersuchen, wurden diese in das COSMO-MUC-KENDA-System assimiliert. Dieses wurde im Rahmen dieser Dissertation als Untersystem zu COSMO-DE-KENDA implementiert. Ein konzeptuelles Modell wurde aufgestellt, welches die Einflussskalen der zwei Beobachtungstypen abschätzt. Durch Anwendung der zuvor entwickelten Maße für Modell-*noise* konnte gezeigt werden, dass eine Assimilation von Radialwinden die Störungen im Vertikalwindfeld dämpft. Reflektivitätsbeobachtungen hingegen verbesserten die Analyse der Gewitterpositionen. Zudem wurde gezeigt, dass die EnKF-Assimilation von Radardaten bessere Vorhersagen liefert und weniger Störungen verursacht als das zuvor genutzte Nudging-Verfahren. Die in dieser Dissertation entwickelte Kombination von Radar- und Flugzeugbeobachtungen stellt einen wichtigen Beitrag zur Verbesserung der Gewittervorhersage mithilfe des KENDA-Systems dar.

Abstract

The application of Ensemble Data Assimilation (DA) is a method to produce initial states (so-called *analyses*) for numerical forecasts of thunderstorms. By the means of radar observations of reflectivity and radial winds, the complex inner structure of a convective cell or a mesoscale convective system can be analyzed. At the same time, thunderstorms are embedded in their surrounding air mass and interact dynamically with this convective environment. This thesis aims to answer some questions which link the observational aspect of convective dynamics to the forecasting aspects of the numerical model, and to the practical and intrinsic predictability when applying convective scale DA.

Since the atmosphere is a chaotic system, errors in the initial states grow during the forecast and thus limit the predictability. Errors that are small in amplitude and spatial scale intensify with a faster rate than large errors, so that an initial state with relatively large errors may produce, after a few hours, an equivalently good thunderstorm forecast compared to an initial state with smaller errors. To test this hypothesis, simulated radar observations of wind and reflectivity were assimilated into the convection-permitting COSMO model with a resolution of 2 km, by the means of an Ensemble Kalman Filter (EnKF), and under the assumption of a perfect forecast model. The spatial and temporal resolution of the observations was coarsened from 2 to 8 km and from 5 to 20 minutes, and their spatial influence was increased from 8 to 32 km in order to produce analyses that contained less detailed information about the observed storms. After three hours, the resulting forecasts were comparable to such experiments in which the observations had been assimilated with the highest resolution and a close fit in the analysis had been reached. Thus, a recommendable setting for the assimilation of real radar data was found that respects the limited predictability of convection. Due to the input of high resolution data, noise was introduced into the model that disturbed the dynamics and triggered spurious convection in wrong locations. The study showed that this noise consisted of internal gravity waves which caused an over-abundance of variability in the vertical velocity field, both in the storm vicinity and in the environment that was at least 32 km distant to the storms. This noise disturbed the internal dynamics of the convective systems in such a way that the emergence of new updrafts became decoupled from the underlying pools of cold air. This was quantified by applying a spatio-temporal correlation method.

In addition to the DA-aspects of mainly storm-interior dynamics, aircraft observations in clear air regions were assimilated into the pre-operational Kilometre-scale Ensemble Data Assimilation (KENDA) system of the German Weather Service (DWD) that couples

an EnKF to the COSMO-DE model over Germany with a resolution of 2.8 km. These observations are collected by Mode-S air traffic control radars and contain measurements of wind and temperature along the flight tracks of commercial aircrafts. Their data density is ten times larger than the previously assimilated Aircraft Meteorological Data Relay (AMDAR) observations. Their assimilation was successfully implemented and it was shown that forecasts of three hours lead time benefit from the additional data by exhibiting a smaller error in the tropospheric wind and temperature profiles which, together with humidity, define the convective environment. Conclusions were drawn on how KENDA reacts to the large increase of observation data in terms of ensemble spread and other EnKF parameters, and recommendations on the operational use of Mode-S data were derived.

The point-like updraft of a convective storm is coupled to the environment by a divergent circulation that has a wider scale than the core. By assimilating real radar observations in the newly implemented framework of the nested COSMO-MUC-KENDA system, the influence of reflectivity and radial wind observations was investigated. A conceptual model was introduced that relates the scales that are influenced by the assimilation of these two observation types. By applying the noise measures that were developed earlier in the thesis, it was shown that the assimilation of radial wind observations has a damping effect on the noise that is caused by the DA. Reflectivity observations on the other hand are helpful to improve the storm positions for the first tens of minutes into the forecast. It was shown that the EnKF assimilation of radar data produces better forecasts and less noise than the previously used nudging scheme. Combining the assimilation Mode-S data and radar observations with the methods that were developed in this thesis is a promising way to improve analyses and forecasts of severe and damaging convection with the KENDA system.

Chapter 1

Introduction

The idea of Numerical Weather Prediction (NWP) was introduced by Vilhelm Bjerknes (Bjerknes, 1904) as a twofold problem in which the initial conditions of the atmosphere have to be determined first and then the governing equations of the system need to be solved, so that a forecast of the future atmospheric state can be produced. The determination of this initial state by the means of meteorological observations is known today as Data Assimilation (DA) (see Daley, 1991; Kalnay, 2003), in which a previous model forecast is used to fill the gaps of the necessarily imperfect observation system. Bjerknes already stated that the amount and density of observation data would be too small, since by that time, only observations from surface stations on land were available but none on the oceans and in upper air. The first real NWP forecast was attempted by Richardson (1922) who manually solved the *primitive equations* to produce a forecast of surface pressure tendency on a certain point over Europe. Despite the correctness of his forecasting method on principle, the initial state had been imbalanced, so that an unwanted oscillation of gravity waves (see Gill, 1982) was excited in the solution. The effect of this spurious gravity wave obscured the actual meteorological signal by more than an order of magnitude, as shown in Lynch (2006).

Since the first successful numerical forecast (Charney et al., 1950), atmospheric models have been steadily improved and the increase in complexity and resolution has significantly driven the evolution of high-performance computing systems in general (Bauer et al., 2015). An important factor in the computational effort has been the emergence of ensemble forecasting systems in the last two decades. Herein, multiple parallel forecasts are produced in a Monte-Carlo approach with slightly altered initial conditions that reflect the limited predictability of the chaotic system of the atmosphere (Lorenz, 1963, 1969).

Besides the ensemble technique, the most significant improvements in operational weather forecasting have been driven by the refinement of DA algorithms and the assimilation of previously unused observation data. The goal of DA is *to make use of meteorological observations in order to find the state of a forecast model that is closest to the real state of the atmosphere* by the means of observable errors and consistency in the dynamics. This so-called *analysis state* can then be used as the initial state for a numerical forecast. Together with the observations, a model forecast state (or *background*

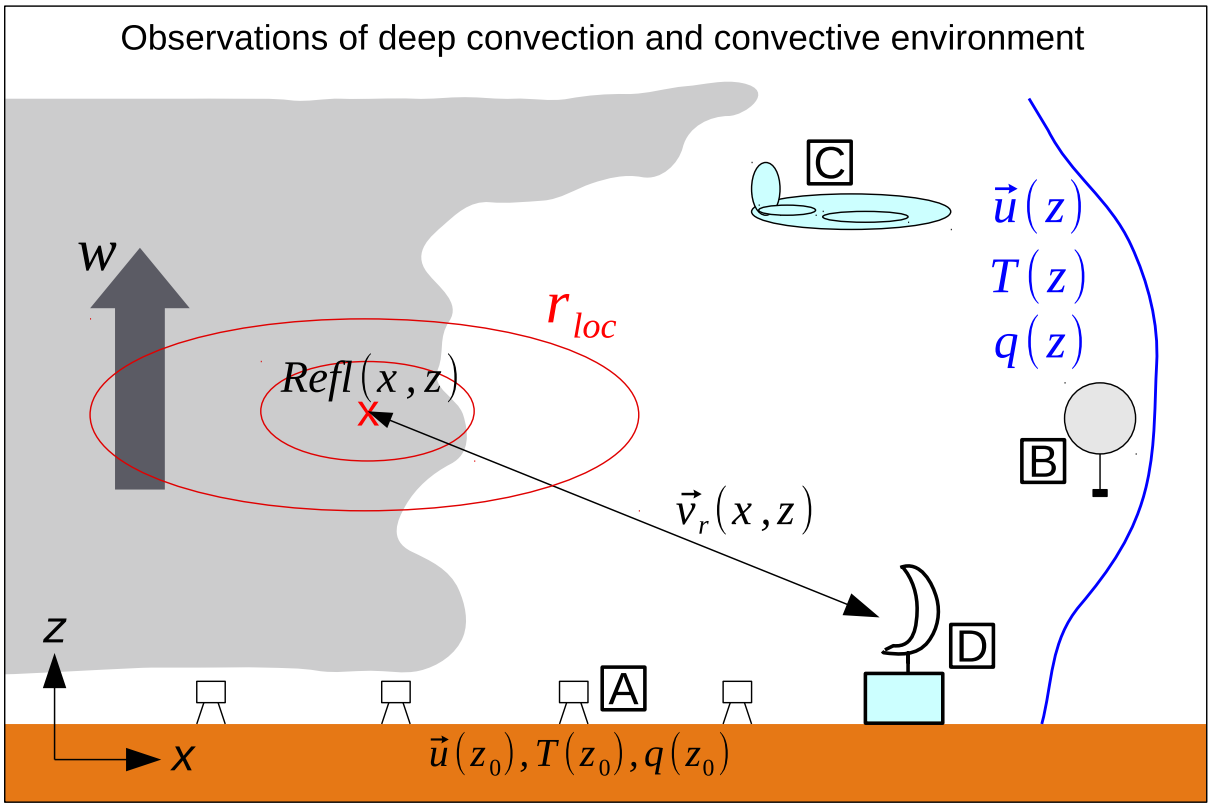

Figure 1.1: Meteorological observation systems that cover the volume inside deep convection (left side, gray cloud with vertical velocity w) and the surrounding environment (right side): Surface stations (A), radiosondes (B) and aircrafts (C) observe wind $\vec{u}$, temperature T and humidity q as a function of height z or at surface level z_0, and Doppler radars (D) that observe reflectivity and radial velocity $\vec{v}_r$ in the rainy volume. A variable localization radius r_{loc} for radar observations is indicated by red ellipses. For examples of real observation systems, see Figures in Sections 4.2 and 5.2.

state), which is an estimate of the weather situation at the time of the observations, is used as the basis for the DA computation. For the finite grid-point models that are used in this study, a vector $\mathbf{x}^b$ expresses the background state[1]. It has the length $m = v \cdot g$ and its entries are given by the values of the v prognostic model variables at all g grid points of the model. The formulation of the finite model therefore defines the basis vectors of the m-dimensional *model space* in which $\mathbf{x}^b$ is defined. The observed measurement values are collected in the vector $\mathbf{y}^o$ with the length n that is given by the number of single observations. For one observation time, the *observation space* that constitutes the basis vectors of $\mathbf{y}^o$ is defined by the locations, the instruments, and the observed variables (exemplified in Figure 1.1), such as a temperature measurement of an aircraft over Iceland at 800 hPa or a visible reflectance measurement pixel over Germany from a satellite imager. To make $\mathbf{x}^b$ and $\mathbf{y}^o$ comparable, $\mathbf{x}^b$ is mapped into the observation space as

$$\mathbf{y}^b = H(\mathbf{x}^b) \tag{1.1}$$

by a *forward observation operator* H. $\mathbf{y}^b$ is usually named the *model equivalent* or *first guess* value. In the example of an in-situ temperature observation in upper air, the temperature value of the nearest model grid point of $\mathbf{x}^b$ could be taken as the model equivalent. H is then a mostly linear map. In the example of a remote sensing observation of reflectance, H needs to simulate the radiative transfer through a slanted model column and the surrounding air, thereby taking into account the solar irradiance, the surface albedo, the three-dimensional cloud variables of $\mathbf{x}^b$, such as cloud water and cloud ice mixing ratio, and the scattering, reflecting and shadowing phenomena that depend on the cloud field. H, as exemplified by images of $\mathbf{y}^o$ and $\mathbf{y}^b$ in Figure 1.2, is then a highly non-linear map. The number of available observations is typically much smaller than the number of model variables ($n \ll m$), and the observations do not cover the full model space. Therefore, the problem of finding an analysis state is under-determined, and statistical methods have to be applied to allow an inverse solution from the observation space into the model space.

The error statistics of $\mathbf{x}^b$ (with respect to the *true* model state $\mathbf{x}^t$ that has no forecast errors) and the errors statistics of $\mathbf{y}^o$ (with respect to the *true* observations $\mathbf{y}^t$ that have no observation errors) are usually assumed to have multivariate Gaussian distributions based on the form

$$f_{\mathbf{x}}(x_1, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right) \tag{1.2}$$

with k-dimensional vectors $\mathbf{x}$ and mean μ, and $k \times k$ error covariance matrix Σ. Since this is assumed for both model and observation space, their respective probability functions of the form (1.2) can be multiplied to obtain the joint distribution in which the two exponents are added and the true state terms are eliminated. The resulting exponent (see Lorenc, 1986) is called the *cost function*

$$J = [\mathbf{x} - \mathbf{x}^b]^T (\mathbf{P}^b)^{-1} [\mathbf{x} - \mathbf{x}^b] + [\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x})] \tag{1.3}$$

[1] In the DA-formalism, superscript expressions like $\mathbf{x}^b$ are often used equivalently to subscripts in other physicals formalisms (see Ide et al., 1997).

with the background error covariance matrix $\mathbf{P}^b$ and the observation error covariance matrix $\mathbf{R}$. The *most likely* model state $\mathbf{x} = \mathbf{x}^a$ is sought (the so-called *analysis state*) by minimizing J such that the distance of $\mathbf{x}$ to the background forecast $\mathbf{x}^b$ is minimal in model space, weighted by the background error covariance matrix $\mathbf{P}^b$, and such that the distance of $H(\mathbf{x})$ to the observations $\mathbf{y}^o$ is minimal in observation space, weighted by the observation error covariance matrix $\mathbf{R}$. The observation error covariance $\mathbf{R}$ of size $n \times n$ contains the covariances of the observation error which is mostly the measurement error, but also errors of the operator and of the representativeness of a single measurement (e.g. Janjić and Cohn, 2006) for the phenomena that the model can simulate or resolve. An entry on the diagonal of $\mathbf{R}$ could be e.g. the square of an observation error standard deviation $\sigma_{obs} = 1$ K, for one temperature observation. The off-diagonal entries may be filled by correlated observation errors like error covariances between different channels of a multispectral satellite sounder. The background error covariance matrix $\mathbf{P}^b$ of size $m \times m$ contains information about the pointwise forecast uncertainty on the diagonal. The off-diagonal entries contain covariances and thus correlations between basis vectors of the model space. These correlations can be point-wise between variables such as wind, temperature, and pressure, and/or spatial across horizontal and vertical distances. For example, a forecast error of wind speed is correlated to the wind speed in distances of a few or thousands of kilometers, depending on the meteorological situation, and at the same time it is correlated to the pressure and the temperature fields through the relationships of hydrostatic, geostrophic, and thermal wind balance (see e.g. Holton, 2004). In the most general form, the cost function J (1.3) is minimized to find the vector $\mathbf{x}^a$ for all grid-points of the model simultaneously, which is called a *global solution*. As an intended consequence, the influence of the observation information is spread spatially, and across variables, according to the correlations in $\mathbf{P}^b$. The effect of information distribution, in case of an analytically prescribed $\mathbf{P}$-matrix, therefore acts also as a *balancing* operation on $\mathbf{x}^a$ such that the observation information is assimilated in accordance with the model dynamics. A naive assimilation of observation information on the other hand, e.g. by simply setting one grid point to the value of the nearest observation, usually results in *imbalances* of the initial state which may cause destructive artifacts such as spurious gravity waves – the same phenomenon which Richardson (1922) had encountered unknowingly.

Once a DA-analysis is produced, a short-term forecast is performed using the numerical forecast model M, which produces a background state for the next assimilation step as

$$\mathbf{x}^b_{t+\Delta t} = M_{t \to t+\Delta t}(\mathbf{x}^a_t) \ . \tag{1.4}$$

Using $\mathbf{x}^b_{t+\Delta t}$, a new analysis $\mathbf{x}^a_{t+\Delta t}$ is produced by the DA. This procedure is called *cycled* data assimilation, or *cycling*. The *forecast interval* or *assimilation interval* Δt that lies between two subsequent analyses can have the length of a few hours or a few minutes, depending on the dynamics of the observed phenomenon that is to be assimilated.

In terms of DA, a predominant operational method has been 3D-Var (see Kalnay, 2003), which still forms the conceptual basis of 4D-Var that is operational e.g. at the European Centre for Medium-Range Weather Forecasts (ECMWF). 3D-Var uses a static background error covariance matrix ($\mathbf{P}^b$) that is analytically constructed to spread the

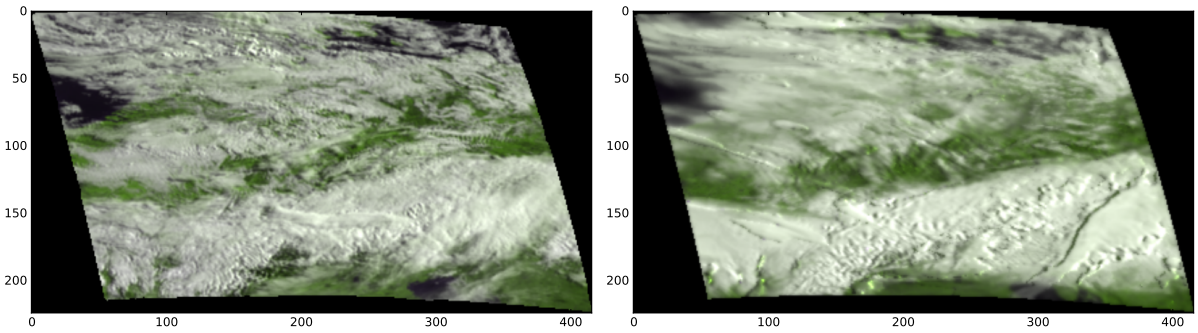

Figure 1.2: Spinning Enhanced Visible and Infrared Imager (SEVIRI) observation of the COSMO-DE region for 25 June 2012, 6 UTC ($\mathbf{y}^o$, left). Model equivalent ($\mathbf{y}^b$, right) generated from the 3h operational COSMO-DE model forecast using the forward operator developed by Scheck et al. (2016). To generate these RGB images, the 0.6 μm reflectance was used for the red channel, the 0.8 μm reflectance for the green channel and the mean value of the 0.6 μm and 0.8 μm reflectances for the blue channel.

correcting influence of observations into the forecast model in a manner that conserves important state properties such as hydrostatic balance, geostrophic balance and thermal wind balance. Evensen (1994) proposed the Ensemble Kalman Filter (EnKF) which uses an ensemble of k forecasts

$$\{\mathbf{x}^{b(i)}\} \ , \ i \in \{1, \ldots, k\} \tag{1.5}$$

with mean

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

to sample

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

The $\mathbf{P}^b$ of an EnKF contains the correlations that are covered by the ensemble of forecasts, which are by nature non-static and depend on the current atmospheric flow. The flow-dependency is the one unique feature which makes the EnKF suitable for weather phenomena such as convective storms where analytic relationships and spatial correlations within the moist and buoyancy driven dynamics can hardly be constructed analytically, as opposed to synoptic scale dynamics on which global models focus. The DA procedure of the EnKF solves (1.3) for the most likely state $\mathbf{x}^a$. Due to the Gaussian assumption, this $\mathbf{x}^a$ is regarded as the *analysis ensemble mean* $\bar{\mathbf{x}}^a$. In a resampling step, the EnKF produces an analysis ensemble

$$\{\mathbf{x}^{a(i)}\} \ , \ i \in \{1, \ldots, k\} \tag{1.8}$$

whose members then serve as initial states for the (ensemble) forecast described by (1.4) that produce the next background ensemble to sample (1.7).

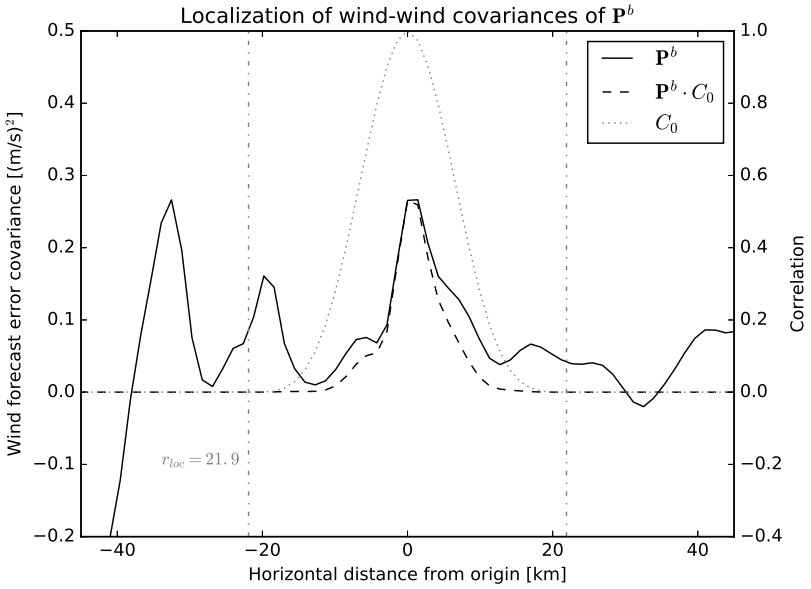

Figure 1.3: Covariances of wind-to-wind along a horizontal line through a thunderstorm simulation of Chapter 2, sampled by an ensemble of 50 members. Shown are the raw covariances (solid), the Gaussian-like correlation function (dotted, function from Gaspari and Cohn, 1999), and their product, the localized covariances (dashed). The correlation function in this example goes to zero at a distance of 21.9 km.

In an EnKF, the k ensemble members form the k basis vectors that span the *ensemble space* of possible analysis states. This means that any EnKF analysis is a *linear combination* of the background ensemble members. Due to the limitation of computational power, k is chosen in the order of tens or hundreds of members while m is in the range of millions, so that $k \ll m$. In the ensemble space are therefore much fewer degrees of freedom than in model space, and the sampled $\mathbf{P}^b$ is necessarily rank deficient. In order to compute sensible analyses, a spatial localization of the ensemble covariances has to be performed, so that a number of *local* analyses can be computed using k local basis vectors. As Ott et al. (2002) stated, the atmospheric dynamics often have a locally lower dimensionality than the number of degrees of freedom in the model would suggest, and therefore local analyses are possible if the local subdomains are small enough. The localization is done by multiplying $\mathbf{P}^b$ (or $\mathbf{R}$, as done in this thesis) by a distance-dependent correlation function that eliminates covariances which are assumed to be spurious due to their spatial distance, as exemplified by Figure 1.3. After their computation, the overlapping local analyses are merged to form the new global analysis, as sketched in Figure 1.4. If applied too rigorously, the localization may disrupt balance properties that would have been contained in the non-localized covariances. As a result, imbalances in the initial state may arise (Greybush et al., 2011), whose quantification are a central aspect of this thesis.

For global weather models, the *synoptic scales* of interest are on the order of 1000 km

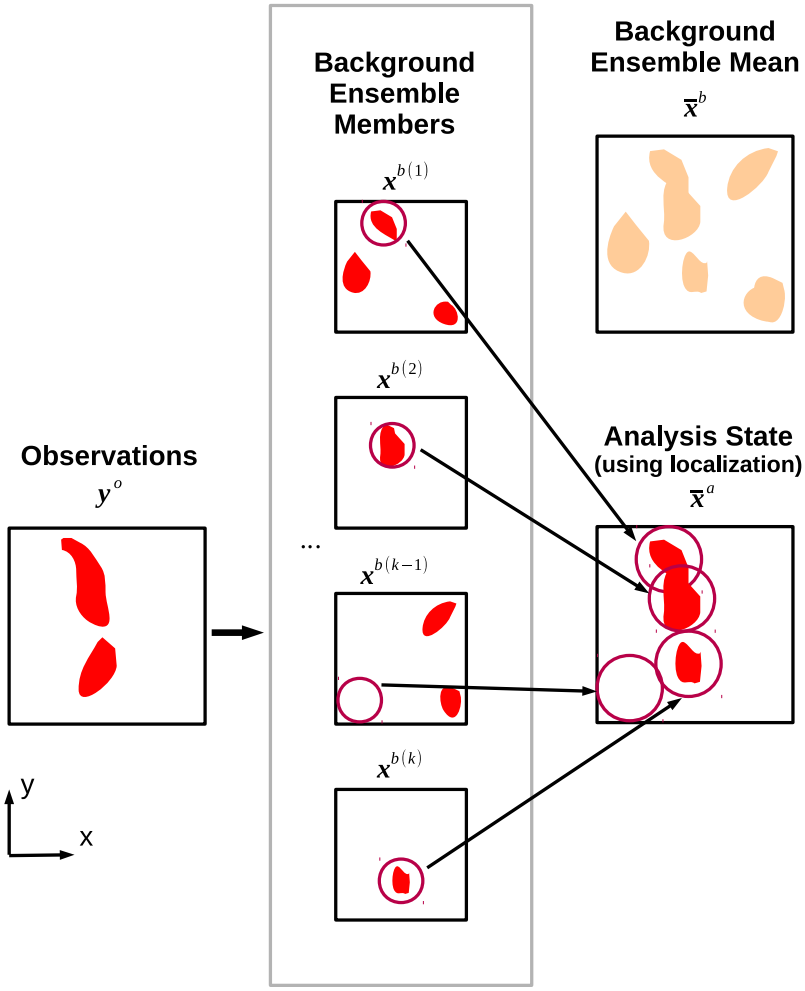

Figure 1.4: Schematic representation of how an EnKF analysis $\bar{\mathbf{x}}^a$ is assembled by overlapping analyses based on background ensembles members which partly resemble the observations (e.g. cloud observations from a satellite or a radar). In this sketch, the observation operator H can be regarded as a self-mapping from model to observation space. Note that also cloud-free observations are taken into account when constructing the analysis.

for low-pressure systems and on the order of a few 100 km for their associated fronts. For meteorological phenomena whose lateral extent is larger than the vertical extent – the latter usually being the height of the tropopause of roughly 10 km – the absence of explicit vertical motion is not a strong limitation (see Holton, 2004). Therefore, some global models even use the hydrostatic form of the atmospheric equations of motion, which means that they do not explicitly simulate motion of air along the vertical axis.

Today, all national weather centers run non-hydrostatic local area models (LAM) with horizontal grid-spacings around 1 km and domains with extents in the order of 1000 km (for a historical overview see Dudhia, 2014). The lateral boundary conditions are usually provided by hydrostatic global forecast models. Non-hydrostatic models explicitly forecast vertical motion which is meteorologicaly associated to buoyancy-driven cumulus convection, wherein single thermals or features of rising air have scales between 100 m and 10 km. These models permit deep convection, such as thunderstorms, that are driven by the release of the latent heat of water vapor from the boundary layer, and also dry inertia-gravity waves that are either excited by clouds or by air flowing over orographic obstacles such as mountains. Examples of non-hydrostatic models are the Weather Research and Forecasting (WRF) model (Done et al., 2004) and the Consortium for Small-scale Modelling (COSMO) model (Baldauf et al., 2011) that is operational at DWD and has been used for the simulations of this thesis.

Single convective storms with horizontal extents of order 10 km can, under the right circumstances, organize into a Mesoscale Convective System (MCS) that extends to scales of order 100 km. MCS, such as multi-cell thunderstorms or clusters of rotating supercells, are characterized by buoyancy-driven updrafts that are coupled to an underlying region of cold air which is caused by the precipitation of dying subcells (see Houze, 1993) – the so called *cold pool*. While the life cycle of a single updraft cell is on the order of 30 minutes, MCSs can have lifetimes of a few hours and may move across distances of several hundred kilometers. They can cause damaging winds and perilous amounts of rainfall in short time, as seen in Germany by the floodings of Braunsbach and Simbach in May and June 2016. Also, thunderstorms and MCS present a threat to civil aviation and may cause unnecessary delays at airports (Sauer and Gerz, 2014; Forster et al., 2014).

Figure 1.1 provides a sketch of the most relevant observation systems for the present study. In Europe and Northern America, surface stations with a spacing of roughly 50 km and radiosonde ascents with a spacing of 100 to 500 km provide an observational network that may resolve synoptic features in the atmospheric flow. Another important data source is civil aviation which provides soundings of wind and temperature close to airports and along the flight tracks. Over the oceans and in the southern hemisphere, satellite observations of temperature and humidity profiles and wind fields derived by cloud tracking are the most important source of large-scale observation data, but satellite observations will be disregarded in this study on convection, apart from their determination of the global boundary conditions for local models. An analysis that is computed by the means of these synoptic scale observations can be used as an initial state to start a forecast of synoptic features. Forecasts with lead times of multiple days are possible which are either skillful or at least correlated to the real synoptic scale flow.

In order to assimilate and forecast convective scale phenomena such as thunderstorms,

observations on the scale of the storms (of order 1 km) are necessary, which are provided by radar stations (see Figure 1.1). By measuring the reflectivity and frequency-shift of emitted microwaves, Doppler-radars gather information about the position, size, density and radial velocity of reflecting hydrometeors such as rain, snow or hail stones. First presented by Snyder and Zhang (2003) and further developed in various studies (for a more detailed list see Section 2.1), it has been shown that the assimilation of radar data in an EnKF framework is feasible and can lead to strongly improved forecast results of deep convection (e.g. Snook et al., 2015; Wheatley et al., 2015; Bick et al., 2016; Yussouf et al., 2016). The EnKF in its form of the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) is planned to be operational at the German Weather Service (Deutscher Wetterdienst, DWD) by 2017 in the Kilometre-scale Ensemble Data Assimilation (KENDA) system (Schraff et al., 2016), which provides the technical and algorithmical framework of this dissertation.

As Stensrud et al. (2009) states, it is an important task to further the development of convective scale DA in order to be able to create initial states for non-hydrostatic models that help to forecast threats posed by the occurrence of deep convection. The aim of the present dissertation is to tackle some of the possible shortcomings and limitations in this field. These different, yet connected aspects are shortly introduced in the following sections of the introduction, and are further elaborated in their dedicated chapters.

Convective DA in the face of limited predictability

In contrast to e.g. a mid-latitude cyclone that grows from a baroclinic instability within days, the growth of a thunderstorm is much more nonlinear and is within the order of minutes to hours. A small disturbance can trigger a cloud that, once initiated, may grow via latent heat release into a long-lived storm whose evolution can then be predicted (Lilly, 1990), or may die off randomly before it reaches a stable state, leaving the non-existence of a storm as the alternative. Sobash and Stensrud (2015) showed that a wrong initial state of mesoscale boundary layer features can lead to convective initiation in wrong locations. The predictability of organized convective systems with respect to their occurrence and their spatio-temporal position is therefore quite limited. Once a cloud is observed, e.g. by a radar system, a well-set DA framework should be able to lock onto that cloud (Craig and Würsch, 2013) and be able to forecast it (Craig et al., 2012) within the limits of the growth of the errors in the initial condition (Lorenz, 1969). By using twin experiments that only differ by very small and meteorologicaly insignificant perturbations in the fields of the initial model state, Selz and Craig (2015) found the error growth in convective scales to saturate within a few hours. In twin-experiments that perturbed mesoscale convective systems, Done et al. (2012) and Zhang et al. (2016) showed that the predictability of sub-cells of a mesoscale convective system is usually small, while the predictability of the mesoscale complex itself is much higher.

Taking this principle of very fast error growth, it is questionable whether forecasts that cover a range of a few hours are better if they start from an initial state in which every single observed cloud has the correct position and shape in every member (denoted as a *phase-locked* property in Neef et al., 2006, 2009), or whether such forecasts are better

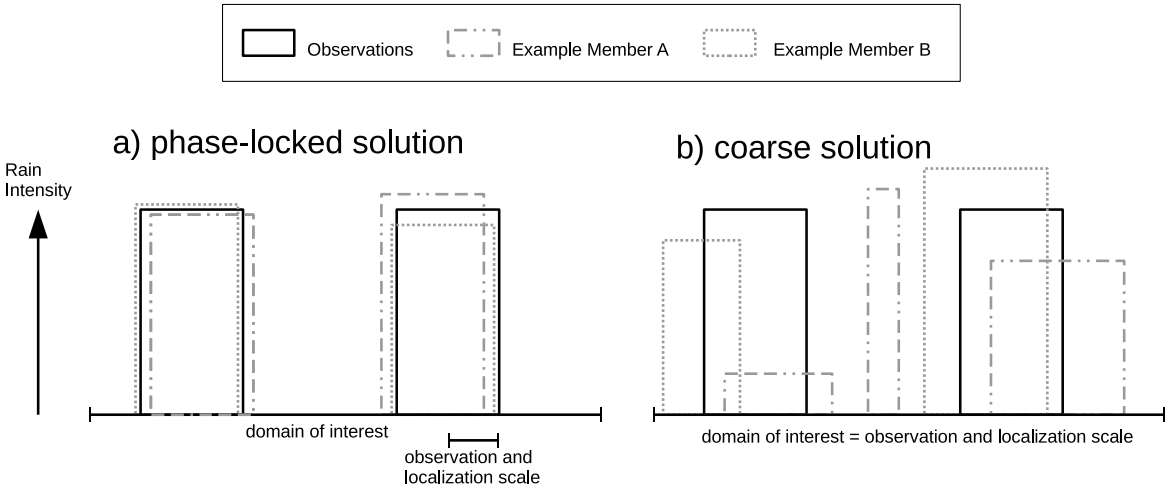

Figure 1.5: Solutions of analysis ensembles that are fitted to the observations a) in a small-scale phase lock with the observations and b) with a coarse-grained constraint.

that start from an initial state where the clouds are only correct on a coarser scale. Both possibilities are illustrated in Figure 1.5. Lange (2013) criticized that most convective EnKF studies tried to assimilate every detail of the observed storms into their analyses, while disregarding the limitations imposed by a finite predictability. By simultaneously increasing the localization length, the assimilation interval, the observation resolution and the grid-resolution of the analysis computation, Lange (2013) concluded that forecasts from initial states that were subject to a coarse-grained constraint (Figure 1.5) can have the same quality as forecasts that are started from initial states which were constrained with a small-scale phase lock. As an explanation, it was hypothesized that errors in coarse analyses are closer to a saturation of the displacement errors than fine analyses are, so that after a certain time the fast growing displacement errors in forecasts from fine analyses reach the same level as forecasts that are started from coarse analyses. Chapter 2, which has been published as Lange and Craig (2014), reassesses the same question, but with a refined methodology. The sensitivity on spatio-temporal parameters of the assimilation system is examined separately, so that those spatial and temporal assimilation parameters are identified which can lead to good forecasts for the first half hour, and those which are sufficient to produce equivalently good three-hour forecasts.

Spurious oscillations and noise caused by convective DA

If an air parcel is vertically lifted in an otherwise resting atmosphere, the related change of pressure and volume causes an adiabatic cooling of the air in the parcel. Since the parcel is now cooler than its surrounding, it is negatively buoyant and will return back to its original height, driven by the restoring force of the gravitational acceleration. The inertia of the returning motion will cause the parcel to descend below its original height. The adiabatic process is thereby reversed and an oscillation is initiated that spreads both vertically and horizontally. This atmospheric phenomenon is called Inertia-Gravity

Wave (IGW) (or also internal gravity wave, see Gill, 1982) and is abundant wherever vertical motion – such as convection and thunderstorms – is present in the atmosphere.

While the energy density of dry IGWs is usually much lower than for the latent heat-driven motion of deep convection, both are necessarily related to each other. Clark et al. (1986); Balaji and Clark (1988) presented idealized experiments in which deep convection was initiated by the coupling of different modes of IGW. Hauf and Clark (1989); Yang and Houze (1995); Tulich and Mapes (2008) gave further insight on the generation of IGW by deep convection, how an MCS can be described as a packet of vertically trapped IGW, and how convectively forced IGW spread and propagate horizontally. Lac et al. (2002); Stephan et al. (2016) analyzed observation data on the initiation of convection by IGW which had been emitted from previous and spatially separated convective systems, and were able to reproduce the observed interaction in simulations with non-hydrostatic models.

In Chapter 2, some convective cells are observed to arise in wrong locations that are distant to the observed MCS, and therefore uncoupled to the true cold pools. This spurious convection arises most severely in such experiments that appear to contain an excessive amount of vertical fluctuations and IGW. Chapter 3 of this dissertation investigates the decoupled property of this spurious convection and quantifies the noise that is caused by such assimilation schemes that do or do not force the analysis strongly onto the observations. The quantification of noise is done in a two-way approach by dynamical arguments about the disturbed coupling of new convection to the cold pools and the spatial statistics about IGW-related noise inside and outside of the storms. The chapter concludes with a recommended setting for DA of real radar observations which could prevent noise issues that are otherwise caused by the DA.

The role of the convective environment

An important property of deep convection is its coupling to the surrounding environmental air. The vertical sounding of temperature and humidity (as indicated in Figure 1.1) decides whether a thunderstorm may arise and how intense an updraft can become (see Houze, 1993). The maximum intensity is controlled by the Convectively Available Potential Energy (CAPE) of the temperature and humidity profile, and the negatively buoyant stratification that has to be overcome by rising air is called Convective Inhibition (CIN). Another important factor is the vertical shear of the horizontal wind which determines the form and longevity (the *mode*) of a convective storm (Weisman and Klemp, 1982). Without shear, an updraft destroys itself by its own precipitating downdraft and is therefore shortlived. With small amounts of shear, the precipitation is transported away from the updraft region and therefore more long-lived multicell storms may arise. With strong and helically turning environmental shear, rotating supercell storms may arise wherein a rotating updraft arises that is very long-lived and may even spawn a tornado at the surface.

By assimilationg observations that do not observe the convection itself, such as surface stations, aircraft data and radiosondes, the convective environment is determined. The data assimilation of the convective cells themselves therefore may fail if the environmental

sounding – a synoptic or mesoscale property – is wrong, as has been shown by Stensrud and Gao (2010). Aksoy et al. (2009, 2010) assimilated convective storms using radar data, and observed their forecasts to die off or misbehave with respect to the convective mode. They attributed this to errors in the environmental sounding which did not support the observed convective mode, e.g. a wrong shear or CAPE. Also, Dawson et al. (2012) and Cintineo and Stensrud (2013) saw a decline in forecast quality of supercell storms when they varied the low-level or the whole tropospheric sounding in their EnKF assimilation of radar data. These studies, with the addition of Aksoy et al. (2010) and Yussouf et al. (2013), mentioned that a certain degree of variability in the environmental sounding of their ensemble proved beneficial for the assimilation of the radar data itself. Still, they stress that the smallest possible error in the environment is always the main goal for an ensemble mean solution. Weisman et al. (2015) describes the Mesoscale Predictability Experiment (MPEX) in which a dense net of dropsonde measurements was deployed prior to a severe convective outbreak in the mid-west of the United States. Hitchcock et al. (2016) assimilated these additional observations in a convective EnKF framework and found the forecast of the convective initiation to improve strongly, since the initial states contained better information about the mesoscale environment.

These studies strongly suggest the use of a higher density of mesoscale observations in order to prepare a better convective environment. In Chapter 4, which has been published as Lange and Janjić (2016), experiments are performed which assimilate additional aircraft observations in the KENDA system that is spanned by an ensemble of COSMO-DE over Germany. These observations are of the type *Mode-S* which are collected by air traffic control and have a data density which is ten times larger than the previously dominant Aircraft Meteorological Data Relay (AMDAR) observations. As a part of this thesis, the preliminary KENDA-cycling system was modified to correctly ingest and assimilate this previously unused observation set. In Chapter 4, forecasts are performed to show that this increased data density leads to a sustainable improvement of the vertical sounding of temperature and wind, and thus of the convective environment.

Dynamical consistency and influence of observation types

Doppler radars observe both the three-dimensional wind and precipitation fields, but radial velocity and reflectivity are not model variables – the latter are defined by the three-dimensional wind vector and the mixing ratios of hydrometeors such as rain, snow, graupel and hail. The wind vector needs to be in accordance with a pressure perturbation field (Houze, 1993) which is not observed. In a series of studies that can be regarded as precursors to contemporary convective DA, Gal-Chen (1978), followed by Hane et al. (1981), Gal-Chen and Hane (1981) and Roux (1985), developed a method to retrieve the pressure perturbation field that is in accordance with the observed three-dimensional wind field, the latter given by a triangulation of radial wind observations from multiple radar sites. To perform this retrieval, two subsequent volume observations of wind are needed by which the local acceleration is computed. Using this field a source term, the anelastic form of the equations of motion can be solved like a Poisson equation, yielding the three-dimensional pressure perturbation field that is in balance with the wind field.

Both fields can then be used to initialize a cloud resolving numerical model to produce a forecast based on the observations.

In the context of the EnKF, forward operators are applied (see Chapter 2) to map the simulated model state into the vector space of the observations. The analysis update in model space is then performed only by the means of the background ensemble covariances and usually without the application of an analytic retrieval of model variables such as pressure. Since the analysis is used as an initial state, the analysis increments need to be consistent with the model dynamics. This means that the wind- and pressure fields have to be dynamically balanced in order to introduce or damp the vertical motion of the convective cells according to the observations, and to prevent the introduction of noise and excessive IGW (see Chapter 3). Potvin and Wicker (2013) observed waves of very fast modes like sound waves in forecasts that were started from EnKF analyses of radar assimilation. Using the method of Gal-Chen (1978) on subsequent wind analysis fields, they were able to modify the pressure perturbation field in such a way that spurious noise was suppressed. Ge et al. (2012) assimilated radar data in a variational framework and were able to reduce noise in the initial states by constraining the analysis state solution with the diagnostic pressure equation and the mass continuity equation which relate the wind and mass fields of the model to the pressure field.

In a convective storm, the precipitation is closely coupled to vertical motion inside the cells and divergent/convergent motion around them. The former is observed by reflectivity, the latter by radial wind. Tong and Xue (2005) tested the combined and isolated influence of reflectivity and radial velocity in idealized convective EnKF experiments. They found the background ensemble covariances to produce worse analyses when only reflectivity was assimilated, compared to the isolated assimilation of radial wind. This motivates a closer look at the effects of these two observation types in a real world setup.

In Chapter 5, real radar observations of reflectivity and radial wind of convective storms are assimilated in a setup similar to that of Chapter 4. The model variety of COSMO-MUC is used that is nested into COSMO-DE with a smaller domain that is centered around the Munich airport (MUC). As a part of this thesis, the ensemble version of COSMO-MUC and a coupled COSMO-MUC-KENDA sub-cycling was implemented in order to produce analyses of multiple parallel COSMO-MUC experiments which are nested into and synchronized with one cycled COSMO-DE-KENDA experiment. For the radar observations, the recommended parameters of Chapter 2 and Chapter 3 are applied. A conceptual relation is derived for the effects of introducing an observed thunderstorm into the model and how the associated circulation of the convective cell is coupled to the latent heat release. Thereby, it is estimated which influence the observations of rain and divergent winds have, depending on the scales that they observe. Using the methods of Chapter 3, it is shown that a combination of both observation types produces a more balanced solution than using only reflectivity, and that the assimilation of radial winds alone already leads to a damping of fluctuations that are induced by the assimilation of conventional observations. Additionally, the noise properties of the KENDA system are compared to the scheme of Latent Heat Nudging (LHN) of radar-derived precipitation observations, which is currently (2016) used in the operational deterministic analysis computation of the COSMO-DE model.

It is advisable to read Chapters 2, 3, 4, and 5 in a consecutive manner since the later Chapters rely on definitions and findings of the earlier Chapters. The final Chapter 6 summarizes the efforts of this dissertation and discusses how the assimilation of convective storms that are coupled to the surrounding environment was addressed. The possible limitations of convective DA that were exposed in the foregoing chapters are summarized, and it is discussed how they could be overcome. Finally, suggestions for further studies are provided that could address some of the new questions that arose in the course of this dissertation.

Chapter 2

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

2.1 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; Scheufele et al., 2014).

Previous studies that used radar data in an EnKF have focused on storm-scale analyses, applied relatively small covariance localization lengths in the range of $\sim$ 10 km (Sobash and Stensrud, 2013) and converged 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 lead-times where the small scale error growth has had enough time to saturate.

2.1.1 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. While model error is usually the largest contributor to forecast errors of convection, this study uses a perfect model to investigate the properties and results of the data assimilation cycling alone.

2.1.2 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 are 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 due to changes in the precision and scale of the analysis errors. It is also discussed how a tight fit to the observations 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., 2013) 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.

2.2 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 throughout the simulation by a constant surface temperature and a constant surface specific humidity in both the nature run and the ensembles.

2.2.1 Nature run

This study uses the testbed setup of COSMO with idealized initial state, periodic boundary conditions and a homogeneous flat landscape at an elevation of 500 m 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). The convection inhibition (CIN) of the sounding is 40 J kg^{-1}, with can be regarded as a fairly strong "cap" that has to be overcome before a thermal can rise and trigger a deep convective cell.

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. 2.1). 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 ≥ 6 h (Fig. 2.2). Horizontally contiguous rain areas with a composite reflectivity > 30 dBZ extend over distances from 30 to 150 km. 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 for 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. 2.1). 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 (seen in all nature runs, e.g. in Fig. 2.1 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.

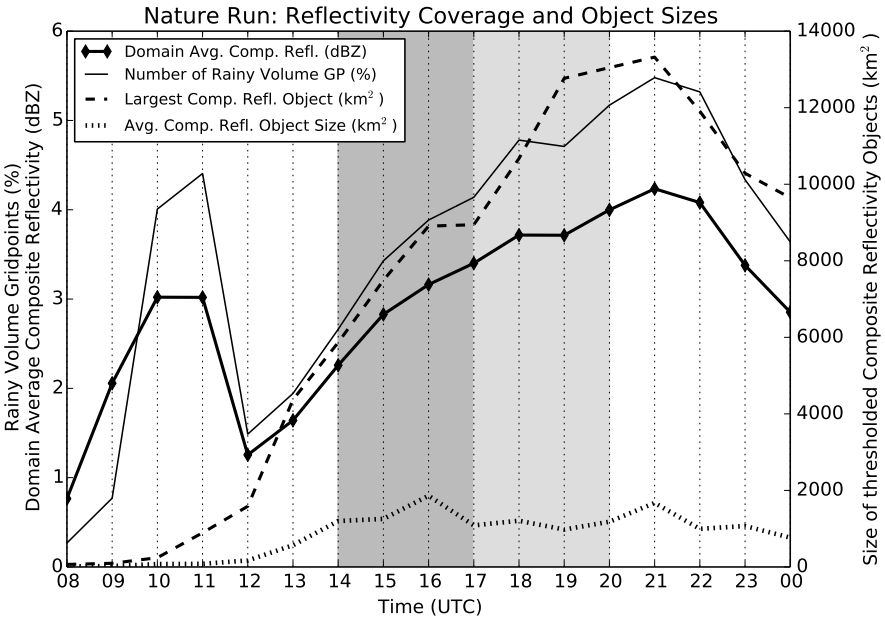

Figure 2.1: Time-series for the Nature Run (averaged over five realizations): Domain-average of the composite reflectivity in dBZ (solid with diamonds), average horizontal (dotted) and maximum (dashed) size of the rain-objects in the composite reflectivity thresholded to > 10 dBZ, percentage of rainy and therefore observed volume grid points with a volume reflectivity > 5 dBZ. Assimilation window is between 14 and 17 UTC (shaded dark gray), forecast window is between 17 and 20 UTC (shaded light gray). Variation of the object-sizes can be due to merger of anvils of separate convective systems.

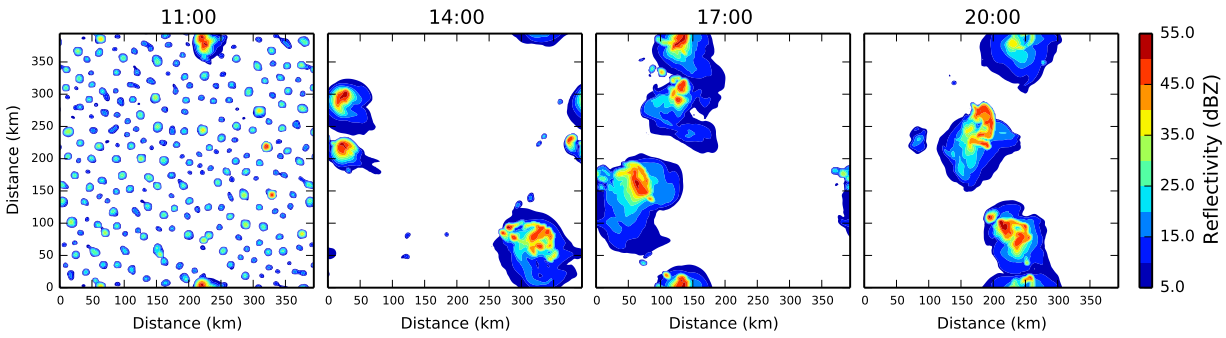

Figure 2.2: Snapshots of composite reflectivity from Nature Run 03.

2.2.2 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 grid point 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 $A_{QG} = 1.03 \cdot 10^9$ where ρ is the air density. 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.

2.2.3 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. 2.2. 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).

2.2.4 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

$$J^*(\mathbf{w}) = (k-1)\mathbf{w}^{\mathrm{T}}\mathbf{w} + [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b\mathbf{w}], \tag{2.1}$$

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.*

$$H(\bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}) \approx \bar{\mathbf{y}}^b + \mathbf{Y}^b\mathbf{w}, \tag{2.2}$$

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

$$\mathbf{Y}^b = H(\mathbf{x}^{b(i)}) - k^{-1}\sum_{i=1}^{k} H(\mathbf{x}^{b(i)}). \tag{2.3}$$

The minimum of the cost function (2.1) 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).

2.2.5 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's 50 layers, but with a vertical spacing twice as large, 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 $\sim$ 1 km, while observations taken a height of 12 km have a vertical influence of $\sim$ 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. Although 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), a shallow vertical localization appears to be sufficient here due to the good data coverage, and has an advantage for computational efficiency as less observations have to be taken into account for each local analysis computation.

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.

2.2.6 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 2.1

(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. If the original observations do not cover the full 4×4 block due to their reflectivity threshold, like at the edge of an anvil, the SO is the average of the available original observations in the partially observed

Table 2.1: Length scales used in the assimilation experiments, as described in Section 2.2. $r_{Loc,h}$ 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. $n_{obs,ana} = (r_{Loc,h}/\Delta x_{obs})^2$ is the number of horizontal observation points per one local analysis within a $r_{Loc,h}^2$ square. $n_{obs,total} = n_{obs,ana} \cdot 180/\Delta t_{ass}$ is the total number of assimilated local observations in the 3 hour assimilation period.

	$r_{Loc,h}$ (km)	Δx_{obs} (km)	Δx_{ana} (km)	Δt_{ass} (min)	$n_{obs,ana}$ (per ana)	$n_{obs,total}$ (in 3 h)
L8	8	2	2	5	16	576
L8SO	8	8	2	5	4	144
L8SOCG	8	8	8	5	4	144
L8SOCG20	8	8	8	20	4	36
L32	32	2	2	5	64	2304
L32SO	32	8	2	5	16	576
L32SOCG	32	8	8	5	16	576
L32SOCG20	32	8	8	20	16	144

block. This procedure 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; Xu, 2011). 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. A comparable amount of observations is provided at every intermittent analysis time for the analogous 5-minute and 20-minute experiments, as the generation of synthetic observations from the nature run is not changed but only performed less often. By increasing the cycling interval, the amount of assimilated information within the 3 hour assimilation period is therefore reduced and the model dynamics of the ensemble are given a longer time to adapt to the updated analysis states.

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 (see n_{obs} in Table 2.1).

The parameter combinations listed in Table 2.1 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.

2.2.7 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 $\bar{\mathbf{x}} = k^{-1} \sum_{i=1}^{k} \mathbf{x}^i$, computed in model space for the different model variables:

$$RMSE(\bar{\mathbf{x}}) = \sqrt{m^{-1} \sum_{l=1}^{m} (x_l^{nature} - \bar{x}_l)^2}, \tag{2.4}$$

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

$$\overline{spr} = \sqrt{m^{-1} \sum_{l=1}^{m} \left((k-1)^{-1} \sum_{i=1}^{k} (x_l^i - \bar{x}_l)^2 \right)}, \tag{2.5}$$

where k is the number of ensemble members. To fulfill the Gaussian assumption of the filter, the ensemble spread should represent the actual error of the analysis, so the consistency ratio CR

$$CR = \frac{\overline{spr}}{RMSE} \tag{2.6}$$

should be close to $CR = 1$. In addition to (2.4), the mean RMSE of the single members with respect to the nature run is computed for some comparisons:

$$\overline{RMSE_{mem}}(\mathbf{x}) = k^{-1} \sum_{i=1}^{k} \left(\sqrt{m^{-1} \sum_{l=1}^{m} (x_l^{nature} - x_l^i)^2} \right). \tag{2.7}$$

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 from a free-running control ensemble that has the same initial conditions as the background ensemble of the LETKF but does not undergo assimilation and runs parallely to the assimilation experiments.

In Section 2.4, additional feature-based scores are described and applied to analyses and forecasts.

2.3 Assimilation results

First the reference experiment L8 is evaluated during the assimilation window 14-17 UTC in Section 2.3.1, then the assimilation results of the scale-varying experiments L8S0-L32SOCG20 (Table 2.1) are compared to L8 in Section 2.3.2. The ensemble forecast parts of the experiments are discussed in Section 2.4.

An example of the assimilation results for one realization of the nature run is given in Figs. 2.3, 2.4 and 2.5, which compare snapshots of the nature run 03 to the analysis ensemble means at 17 UTC. The composite reflectivity is the vertical maximum of the 3D-reflectivity field (Fig. 2.3). T at the height of $z = 150$ m shows the cold pools (Fig. 2.4), W at $z = 3500$ m shows regions of up- and downdrafts (Fig. 2.5). Visually, the L8 analysis closely matches the nature run, particularly in reflectivity and vertical velocity.

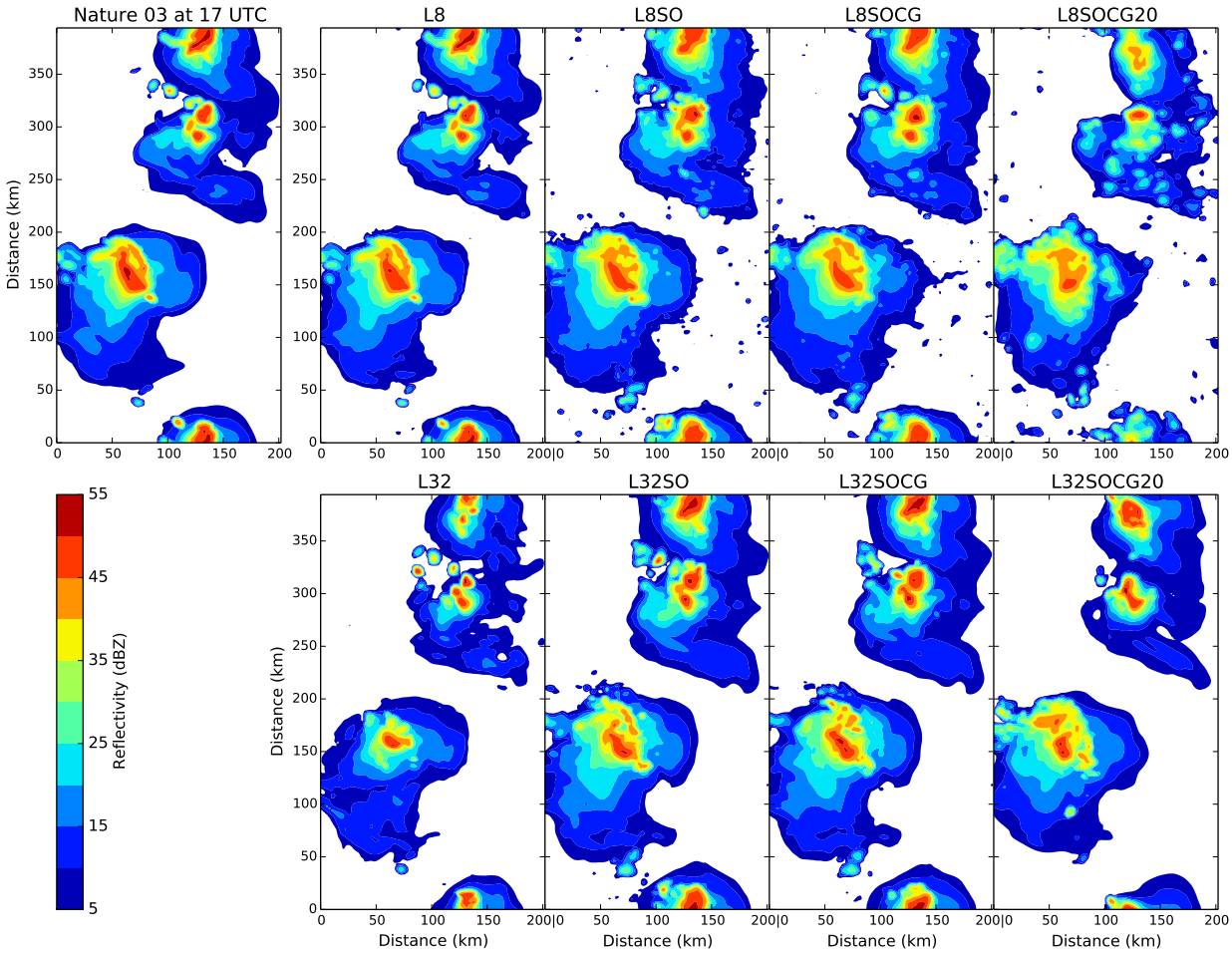

Figure 2.3: Composite Reflectivity of Nature Run 03 (cf. Fig. 2.2 at 17 UTC) and of the corresponding **Analysis Ensemble Means** of the different scale-experiments at the last analysis time 17 UTC.

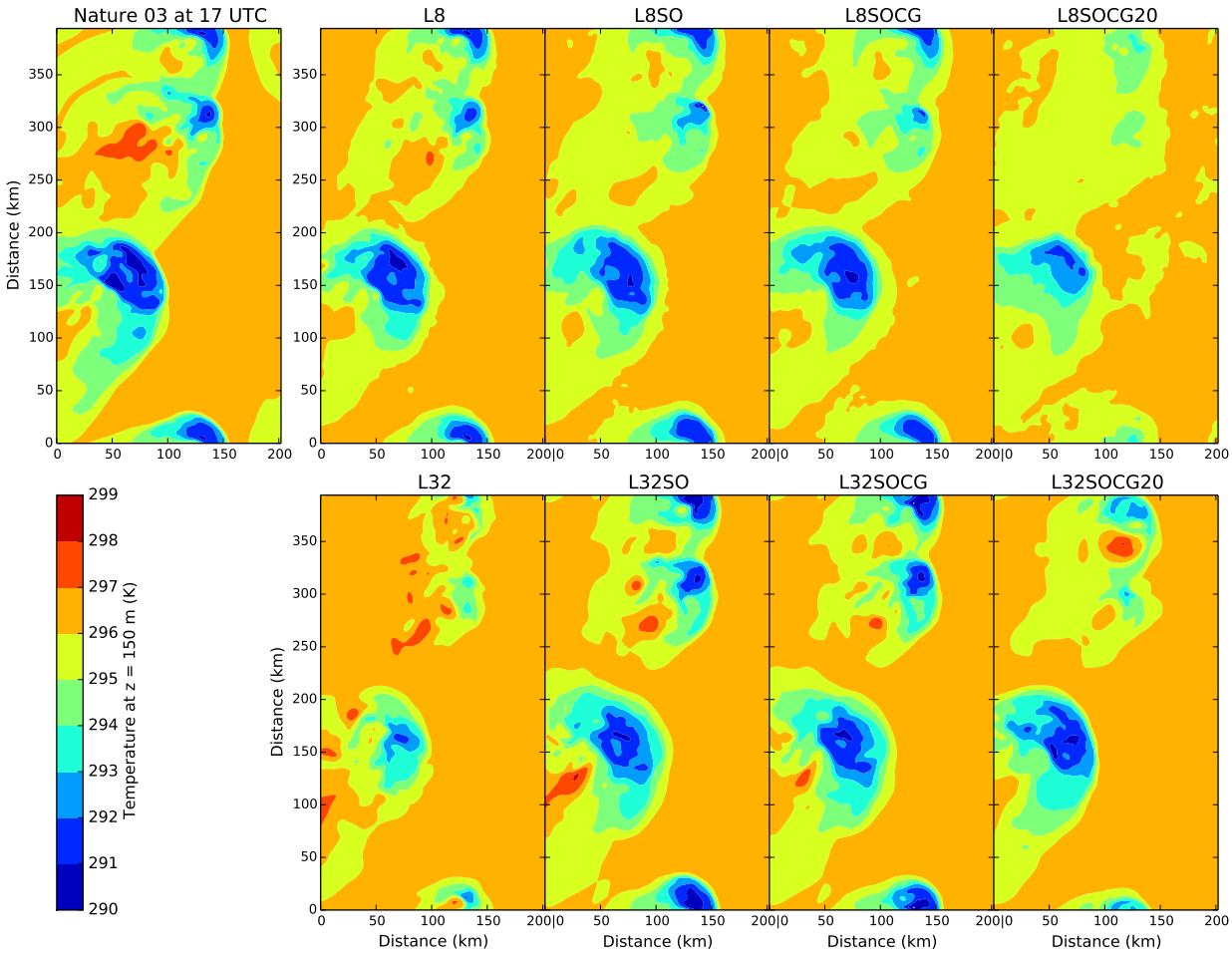

Figure 2.4: As Fig. 2.3, but showing the temperature T at $z = 150$ m.

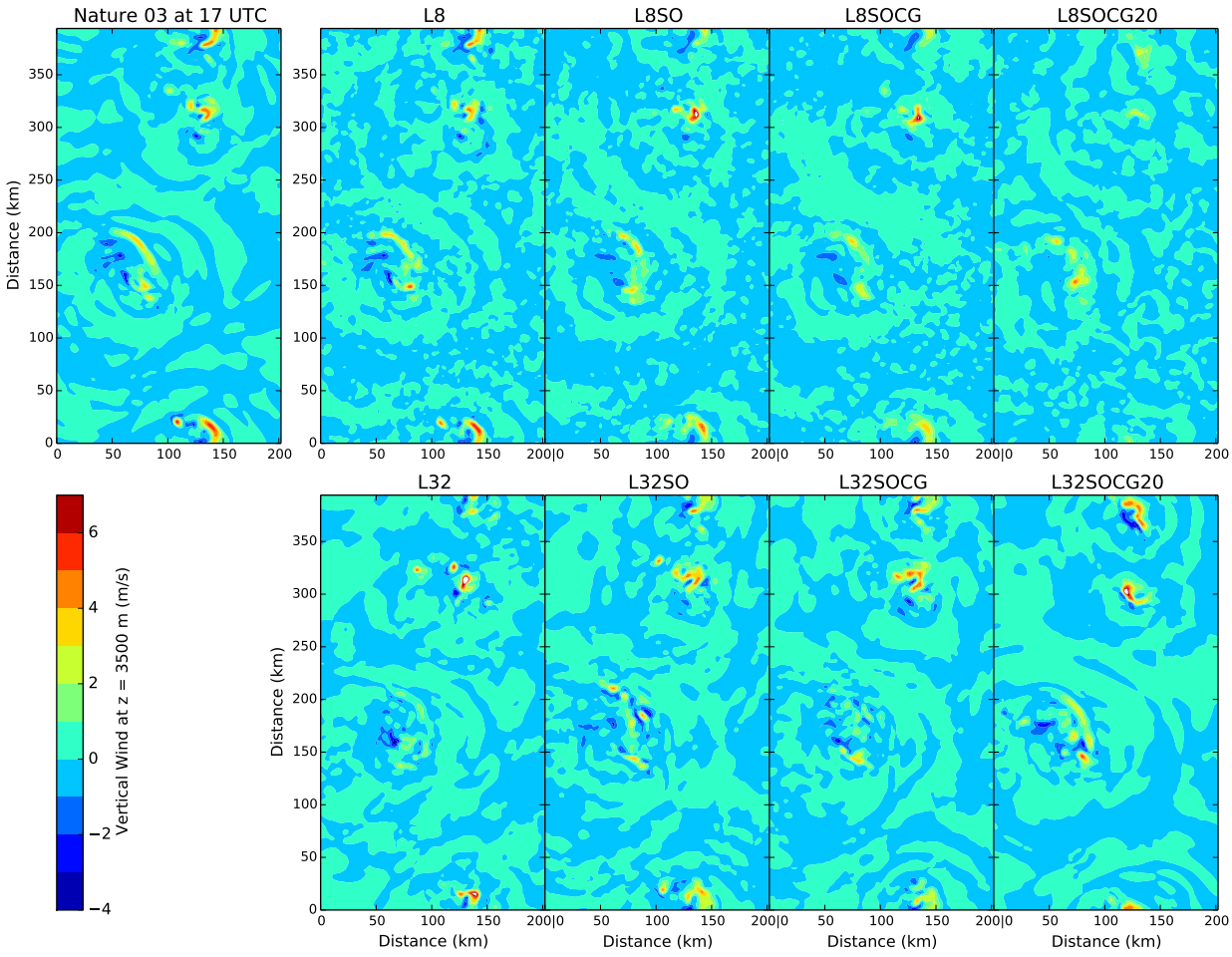

Figure 2.5: As Fig. 2.3, but showing the vertical velocity W at $z = 3500$ m.

2.3.1 Performance of the reference scheme L8

Fig. 2.6 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 grid points. 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 given by the control ensemble, which changes throughout the diurnal cycle due to an increase in domain-wide storm coverage (cf. Fig. 2.1). 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 south-westerly background wind.

The pressure field PP, the humidity QV and the temperature T (illustrated by cold pool structures in Fig. 2.4) 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. 2.6), 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. 2.5) 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. 2.7), 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-constrained by 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 probably happens due to the non-Gaussian and multimodal climatological distribution of the clouds prior to the assimilation which is converged towards a Gaussian solution in the analyses. The variance of a multimodal distribution with rain and no-rain present is larger than the variance of Gaussion distributions around rain and no-rain (Dance, 2004; Craig and Würsch, 2013). The localized Gaussian analysis solutions of the filter therefore exhibit a smaller spread than the climatology which is the free error level, displayed for example in Fig. 2.6. This convergence towards a Gaussian

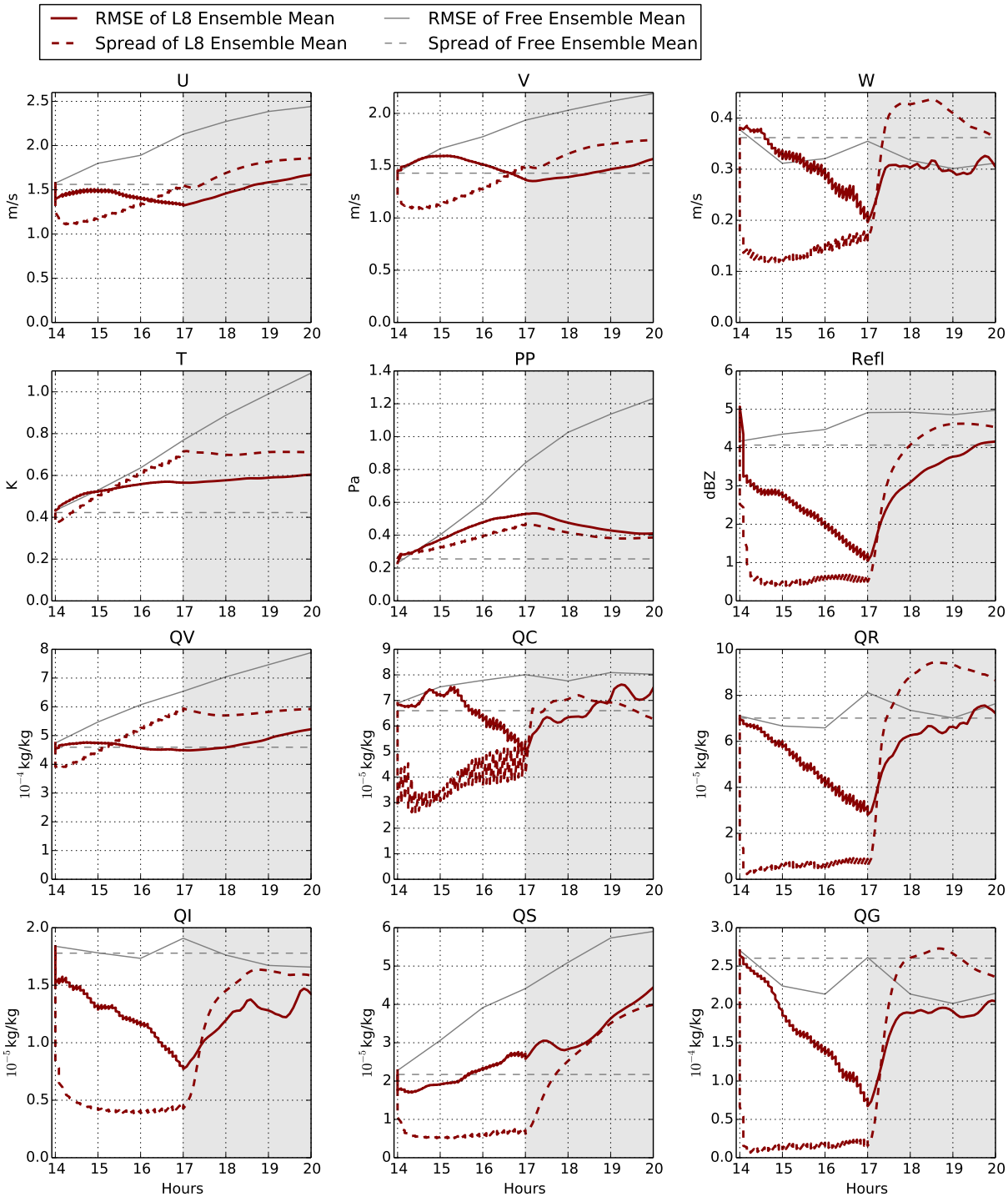

Figure 2.6: RMSE and spread of the ensemble mean of L8 through the assimilation (14-17 UTC) and forecast (17-20 UTC) phases of the experiment, with gray shading indicating the forecast phase. The gray lines show the RMSE and spread of the mean of the free control ensemble without any assimilation. All grid points are evaluated. The error values are averaged over five repetitions of the experiment.

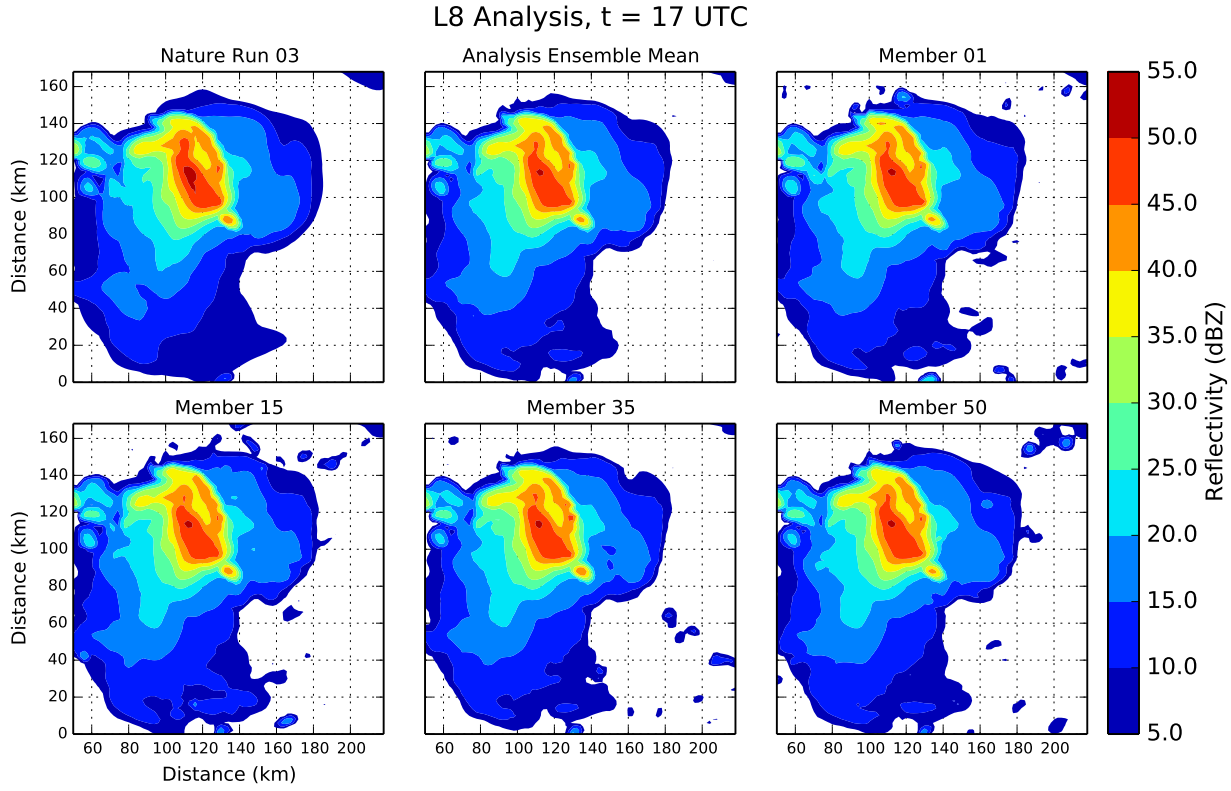

Figure 2.7: Composite reflectivity of Nature Run 03 (cutout of the domain in Fig. 2.2), **Analysis Ensemble Members** 01,15,35,50 and Analysis Ensemble Mean of L8 at the last assimilation time 17 UTC

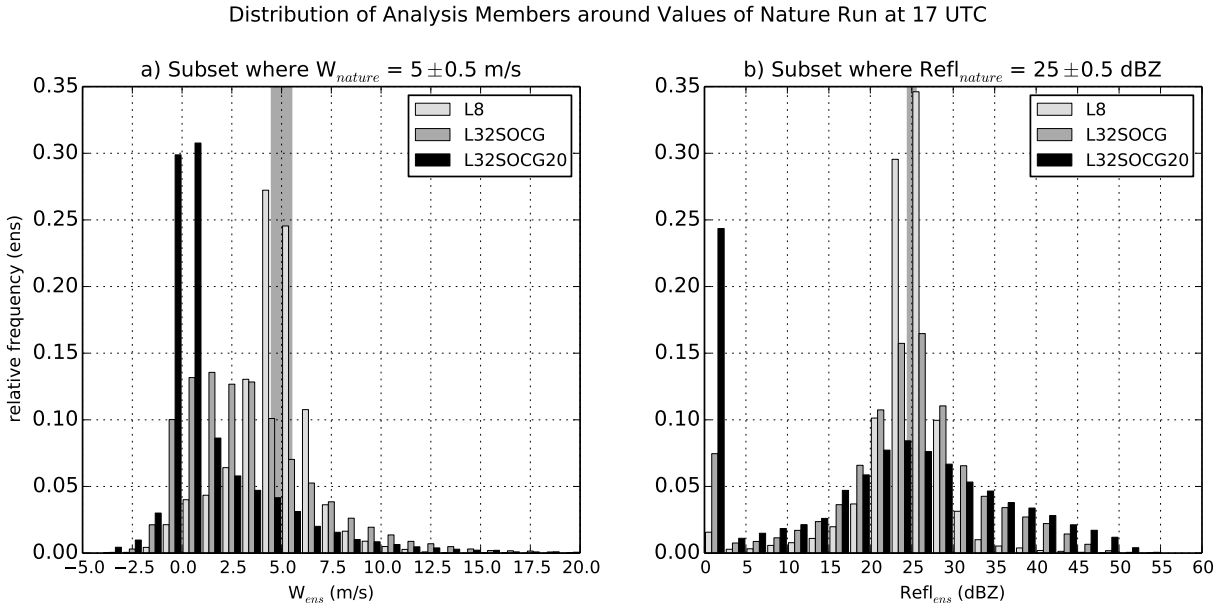

Figure 2.8: Relative frequencies of model values within the analysis ensemble-members of L8 (light gray), L32SOCG (gray) and L32SOCG20 (black) at the last analysis time of 17 UTC (cf. Figs. 2.3, 2.5, 2.7, 2.13), averaged over the five repetitions of the experiments. Distributions are computed for regions inside storms defined by a) grid points 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, both illustrated by a gray bar in the background. The bin width for the ensemble values is 3 times their bar width.

analysis solution is illustrated in Fig. 2.8 which shows histograms of W and $Refl$ of the L8 analysis ensemble, computed from the grid point 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. 2.9 shows the RMSE and spread of L8 as in Fig. 2.6, 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. 2.6 and Fig. 2.9 during the assimilation period from 14 to 17 UTC 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 grid points. During the forecast period from 17 to 20 UTC, the spread of such variables which are strongly associated with convective updrafts (W, $Refl$, QR, QG and QI) quickly exceeds the RMSE of the forecast ensemble mean if all grid points are considered (Fig. 2.6). This increase in variance is due to the development of spurious updrafts in the L8 forecast ensemble members. The grid points of these spurious cells are not contained in the thresholded subset of the observed storms (Fig. 2.9). They will be discussed in the forecast results Section 2.4.

As the relative differences between the two choices of grid point sets for RMSE-computations are small during the assimilation period, and the full-domain RMSEs of Fig. 2.6 contain information about the whole three-dimensional fields, further discussion of RMSEs in this paper will refer to the full domain. Furthermore, since the RMSE behaviour of many variables are similar to each other, only plots of four representative variables U, W, T and QR 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. 2.10).

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.

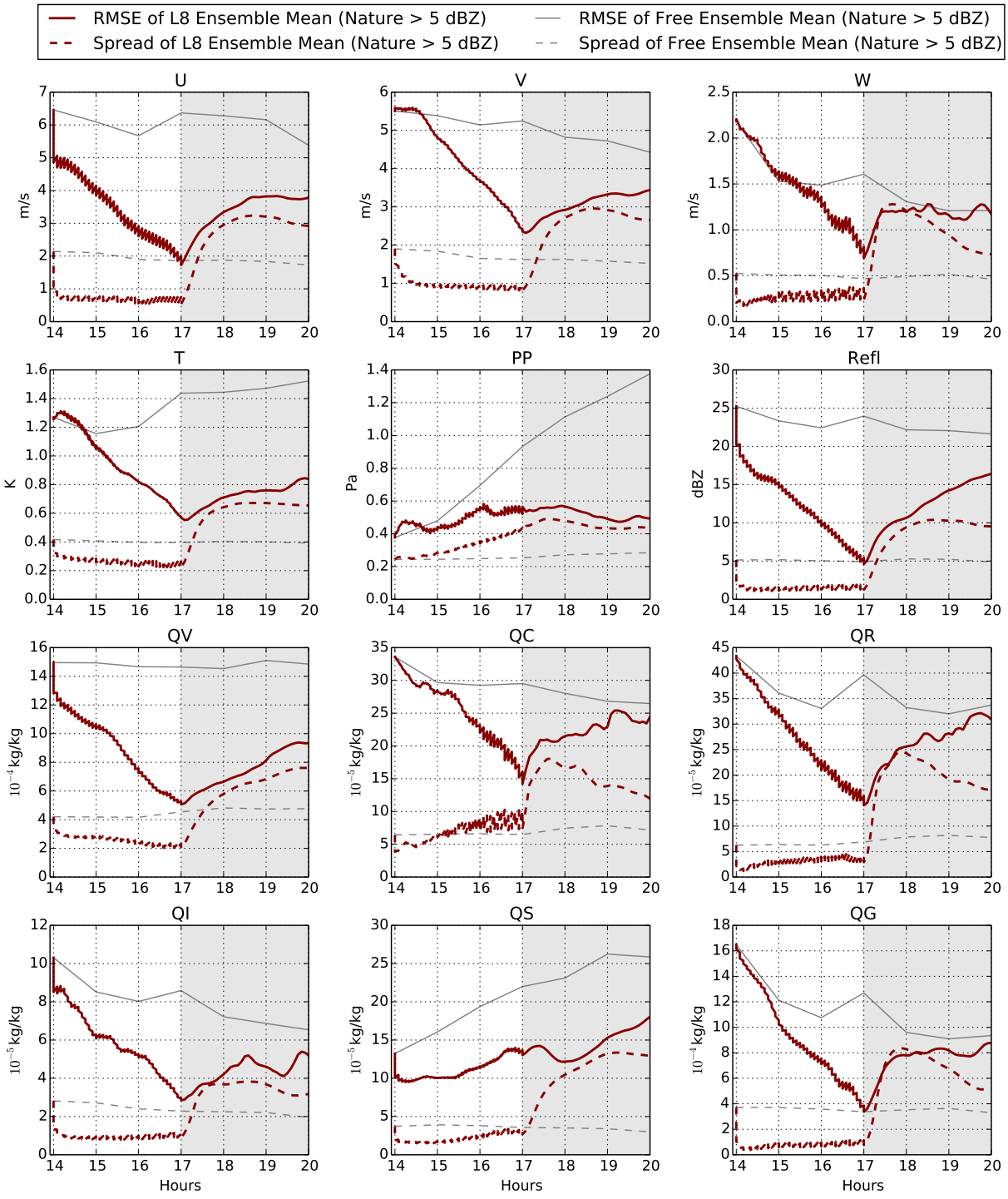

Figure 2.9: As Fig. 2.6, but only the subset of grid points is evaluated where the volume reflectivity of the Nature Run exceeds the observation threshold of 5 dBZ.

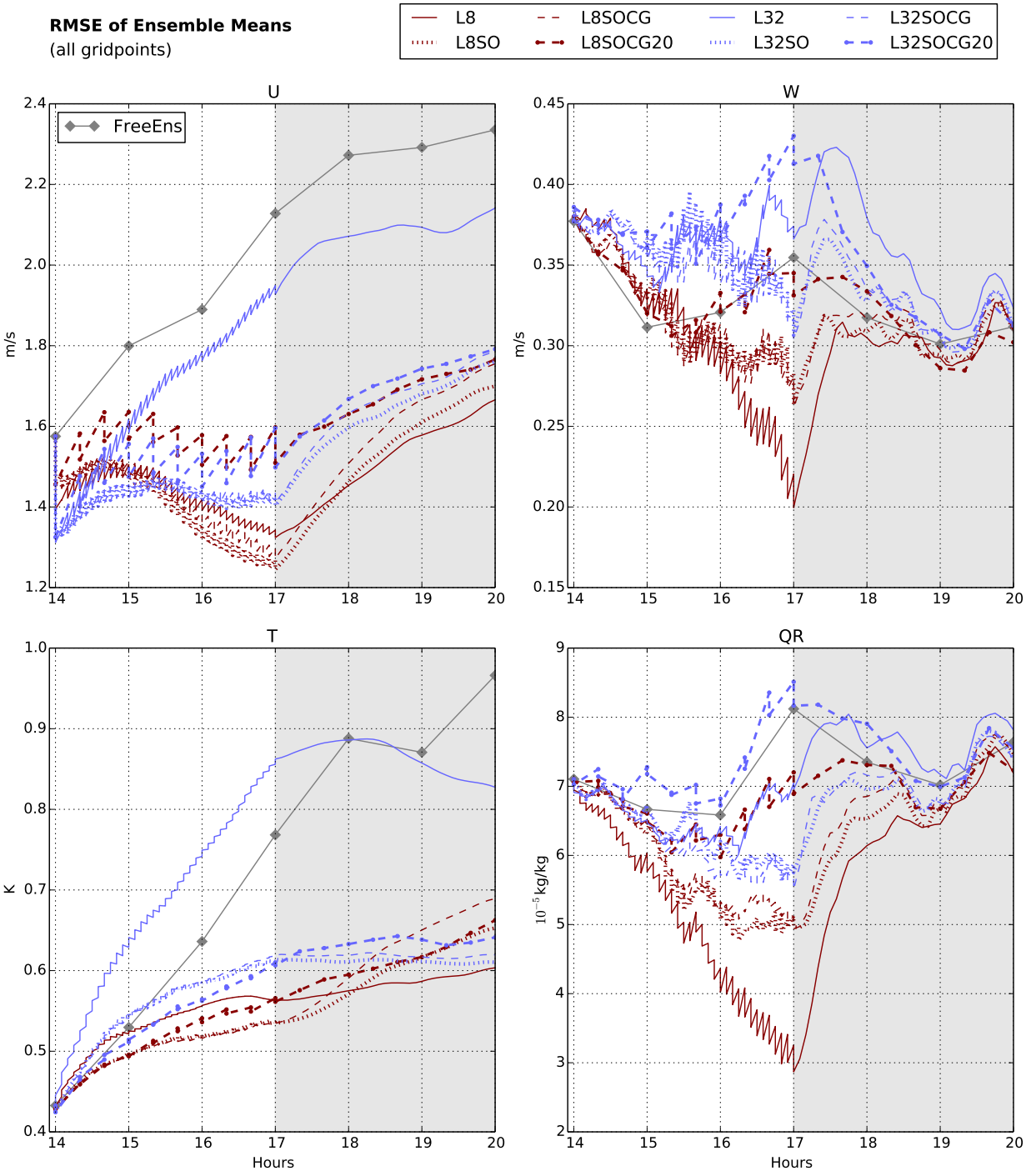

Figure 2.10: As Fig. 2.6, but now showing RMSE of the ensemble means for all experiments (spread not shown). All grid points are evaluated. The error values are averaged over five repetitions of the experiments.

2.3.2 Influence of length scales on assimilation results

In Section 2.2.6 and Table 2.1, 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 2.3.1).

First, the effect of increasing only the localization radius is considered (step i in the experimental hierarchy of Section 2.2.6), 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. 2.3, 2.4 and 2.5. The mean reflectivity field is much weaker than the nature run, as is the cold pool intensity. These results are typical for the five repetitions of the experiment although the fields differ in detail. The RMSE of the L32 ensemble mean (Fig. 2.10) 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. 2.3) than in L8. In single members of L8SO in Fig. 2.11, 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. 2.10). 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 cycling does not completely fail although there is very little horizontal overlap in the solution of adjacent local analyses when the observation resolution of $\Delta x_{obs} = 8$ km matches the localization radius of $r_{Loc,h} = 8$ km, resulting in a very non-smooth analysis solution.

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. 2.10), albeit with less precision in the small scales than L8. This can be seen comparing Fig. 2.3 and Fig. 2.12, 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

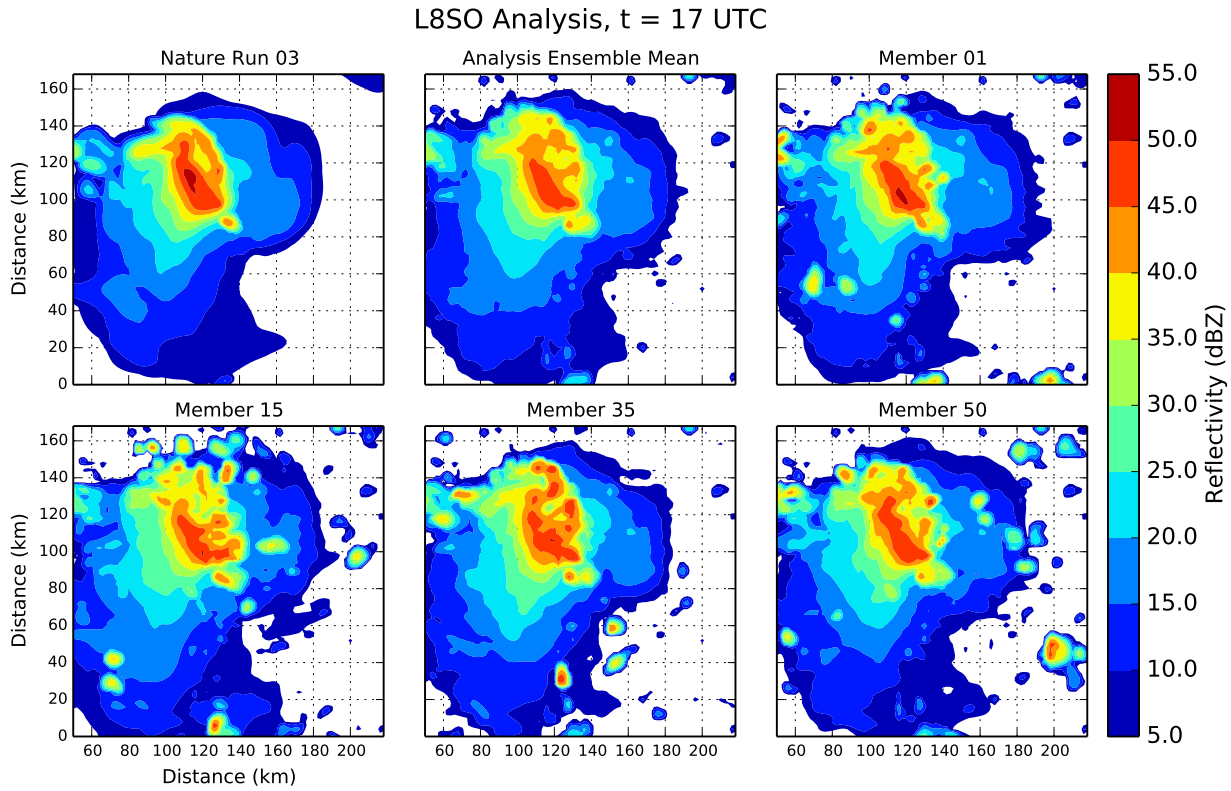

Figure 2.11: As Fig. 2.7, but for L8SO.

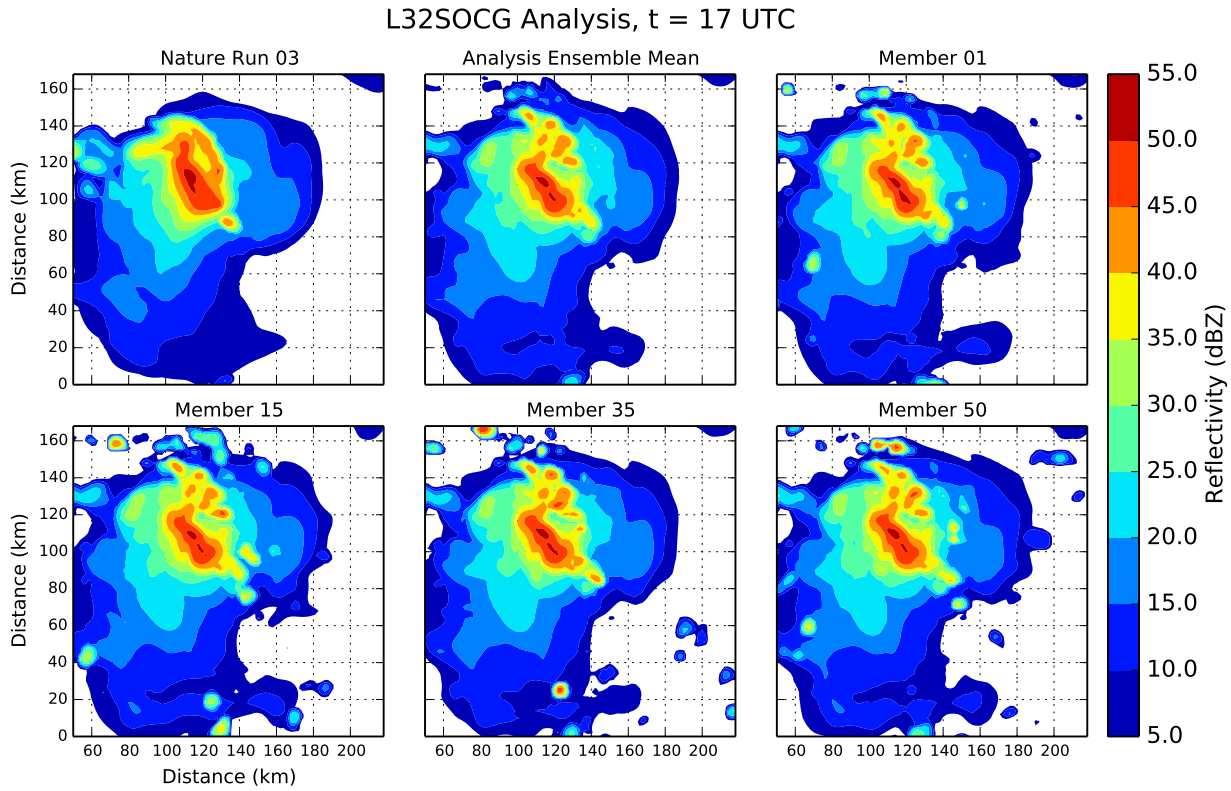

Figure 2.12: As Fig. 2.7, but for L32SOCG.

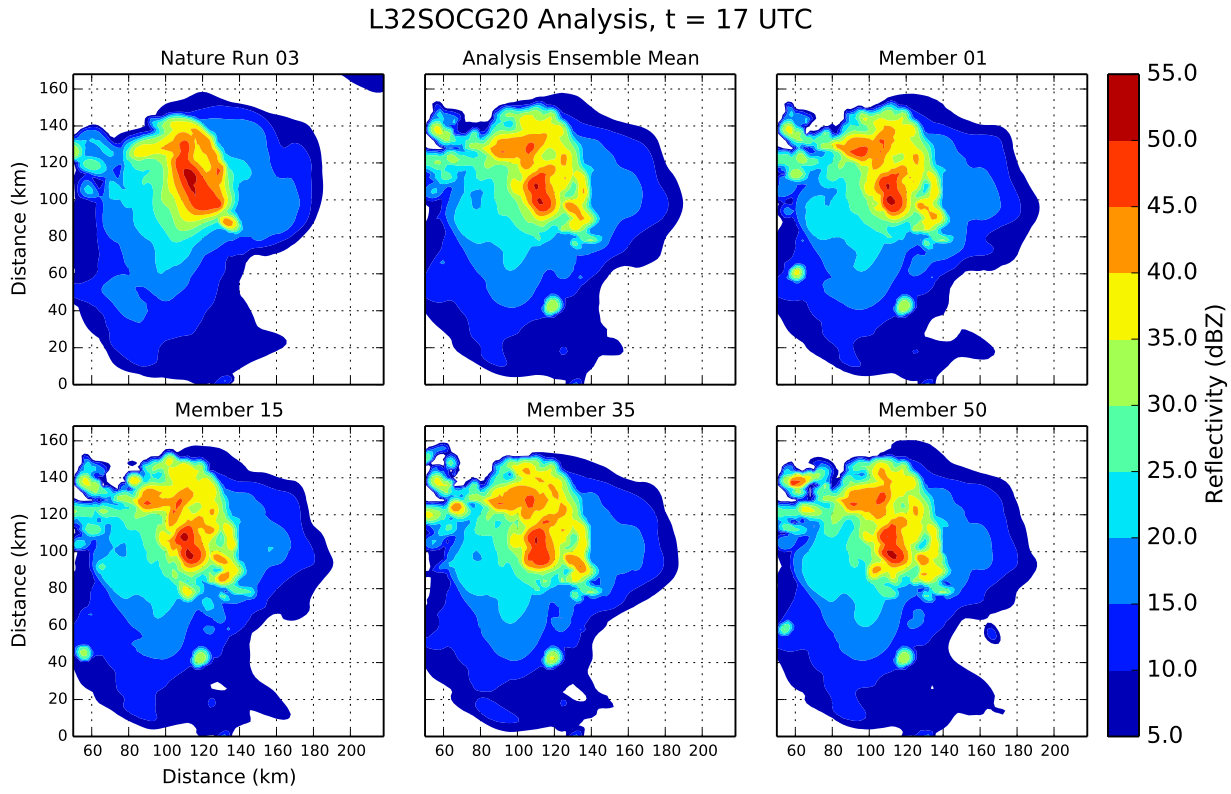

Figure 2.13: As Fig. 2.7, but for L32SOCG20.

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. 2.10 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. 2.3 and 2.4) and the RMSEs are significantly worse than for L8SOCG (Fig. 2.10).

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. 2.10). In Fig. 2.12, the analysis members of L32SOCG show a much larger spatial variability of the storm field than L8 (Fig. 2.7). This variability is even larger with L32SOCG20 (Fig. 2.13). Fig. 2.8 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. 2.8b 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. 2.8a where L32SOCG and L32SOCG20 show a more climatological distribution than the closely converged L8, where climatology is defined here by the frequency distribution of $Refl$ or W at the rainy grid points in the nature runs (cf. Fig. 2.2).

The RMSE of the L32SOCG20 analysis mean of W (Fig. 2.10) 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. 2.3, 2.13). 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. 2.10), 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. 2.14 at the mean RMSE of the individual ensemble members with respect to the nature run, as defined in (2.7). 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. 2.10. For U and T, this property indicates that horizontally smooth fields such as U and T are sensitive to the

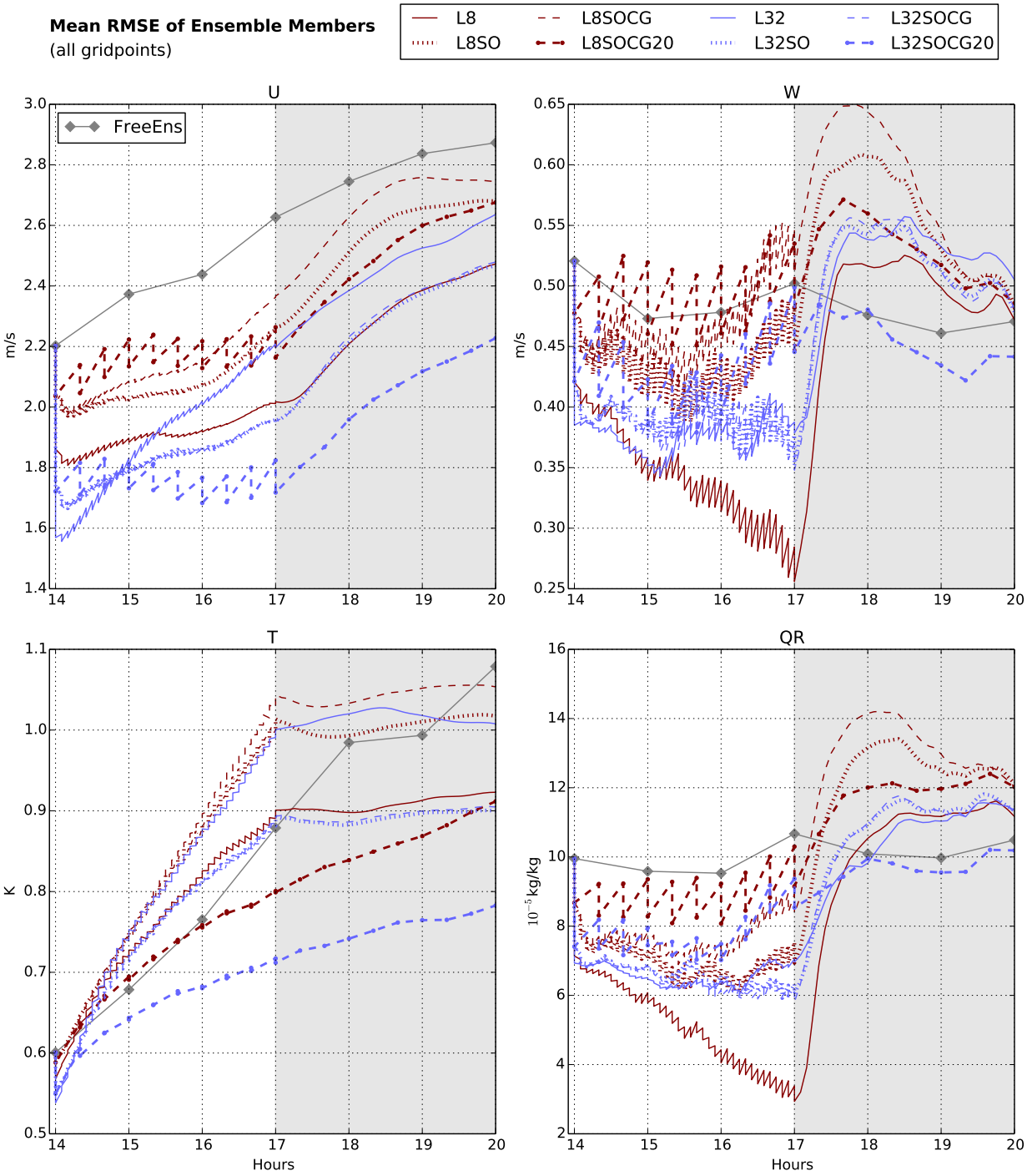

Figure 2.14: As Fig. 2.10, but showing the mean RMSE of the individual ensemble members of all experiments (see equation (2.7)).

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. 2.5 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.

2.4 Ensemble forecasts from analyses with different length scales

The main goal of this paper is to evaluate the influence of the spatial 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 leadtimes up to 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.

2.4.1 Forecast fields and RMS Error

For U and T, the RMSE of the forecast ensemble mean (Fig. 2.10) 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. 2.10). 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. 2.10 is consistent with W and QR which smoothe out in the forecast ensemble mean, becoming similar to the very smooth free ensemble mean fields of W and QR due to the random position of the storms in the free control ensemble. This smoothing can be seen directly in Fig. 2.15 for the mean forecast composite reflectivity, in Fig. 2.16 for T and in Fig. 2.17 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. 2.8) inside the nature run's storm locations, with the values for L8 narrowly distributed around the observation. However, after 3 hours (Fig. 2.18), the distributions of W and $Refl$ from

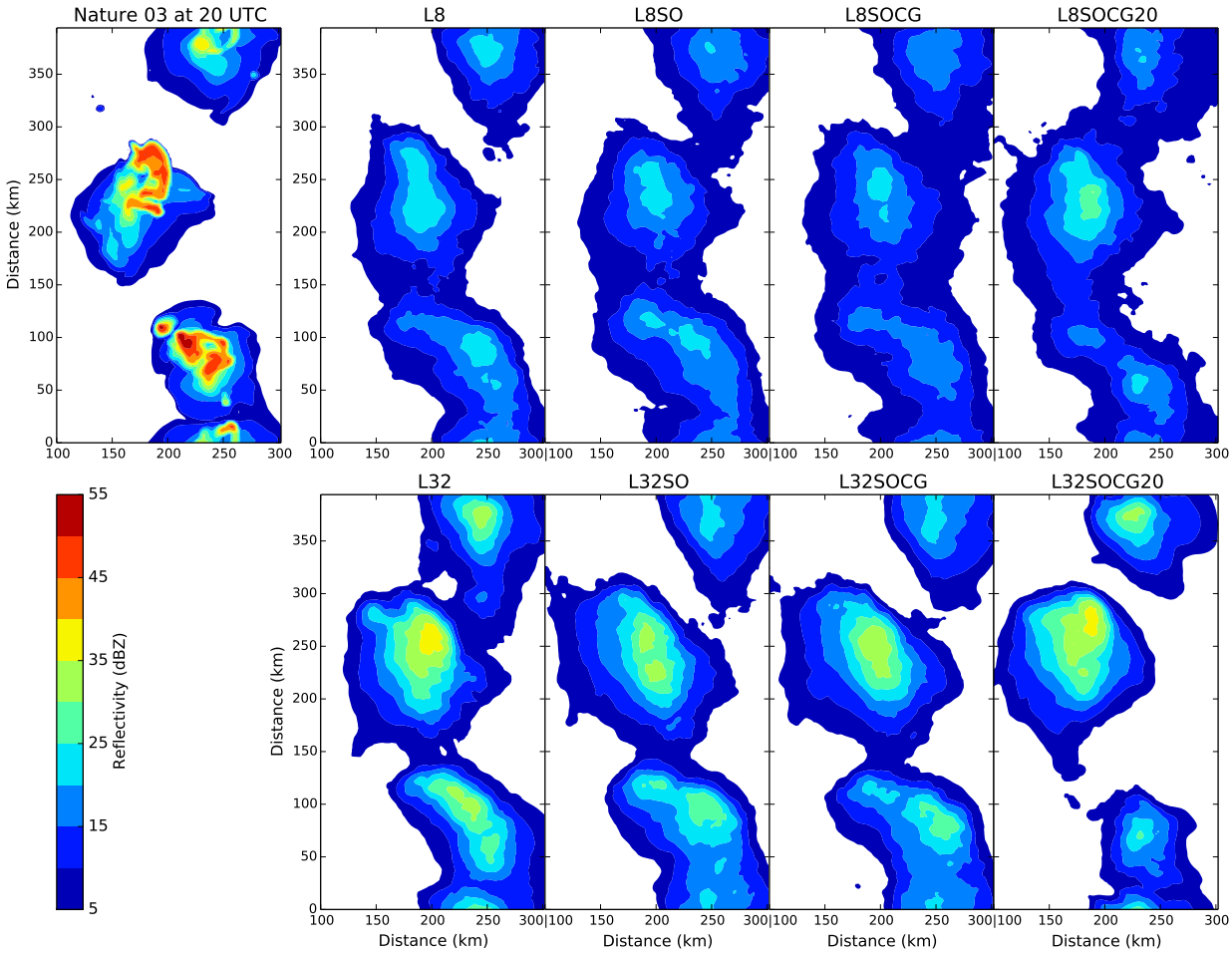

Figure 2.15: As Fig. 2.3, but showing Composite Reflectivity of Nature Run 03 (cf. Fig. 2.2 at 20 UTC), and the **Forecast Ensemble Means** of the different scale-experiments at 20 UTC, after 3 hours of ensemble forecast.

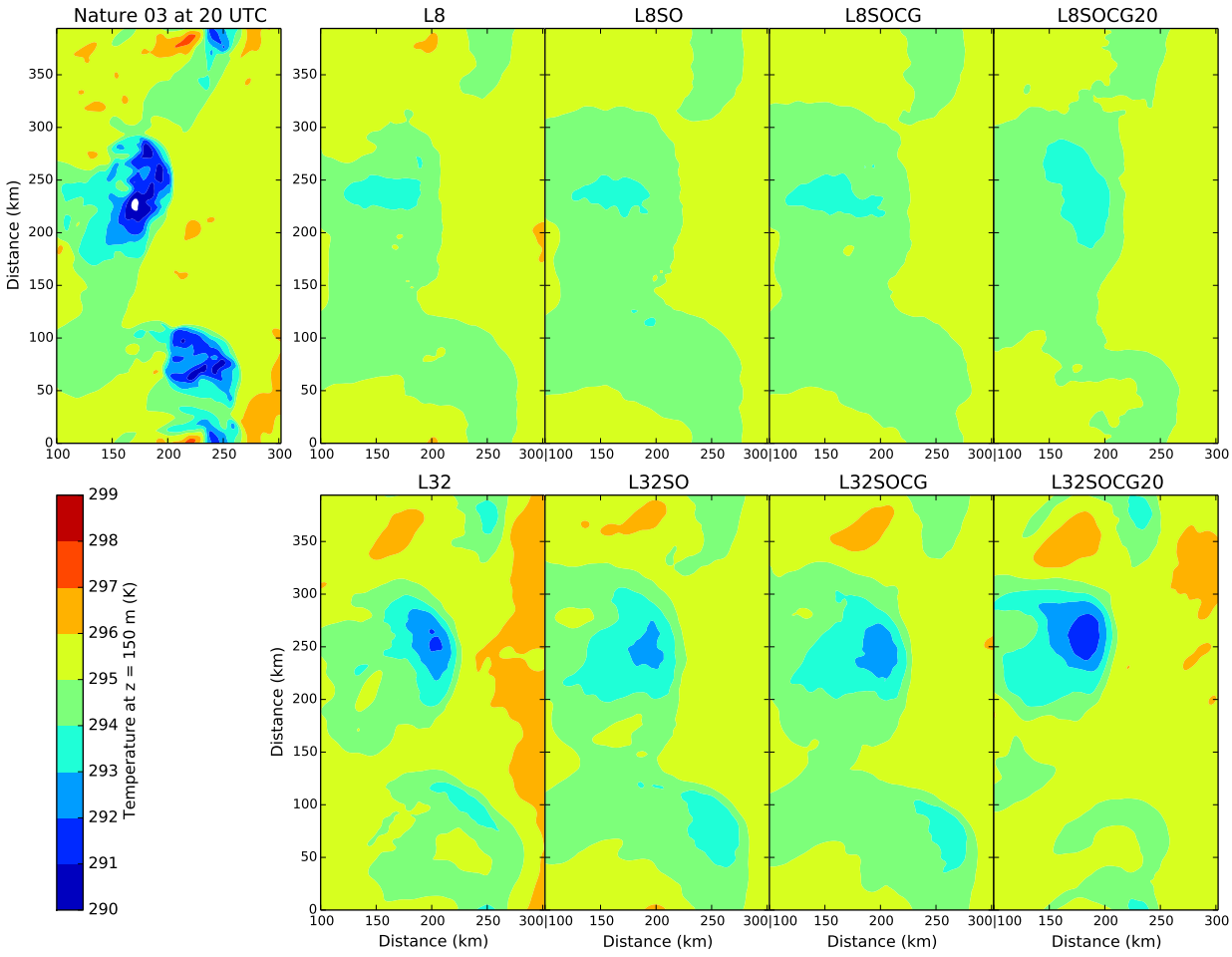

Figure 2.16: As Fig. 2.15, but showing the temperature T at $z = 150$ m.

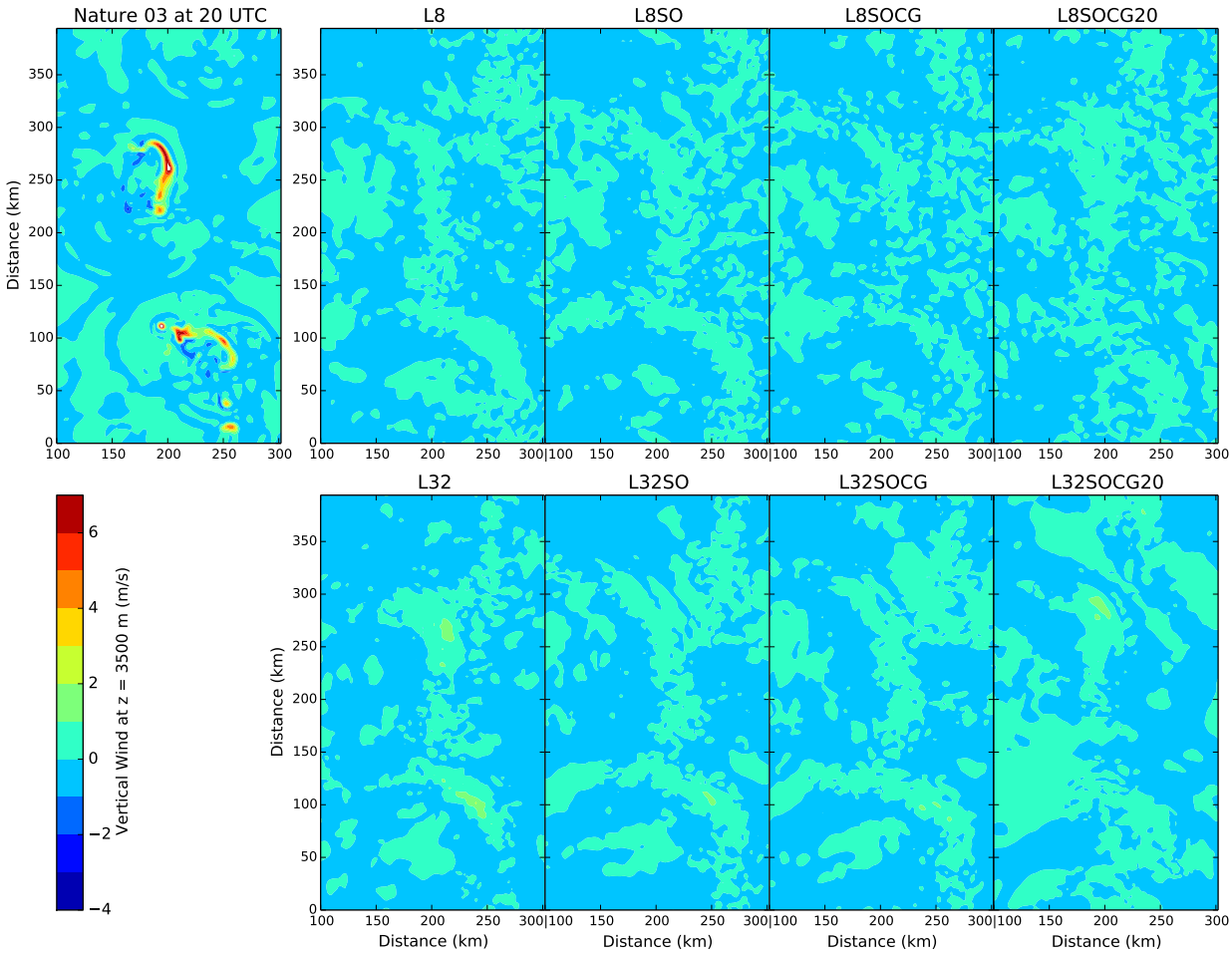

Figure 2.17: As Fig. 2.15, but showing the vertical velocity W at $z = 3500$ m.

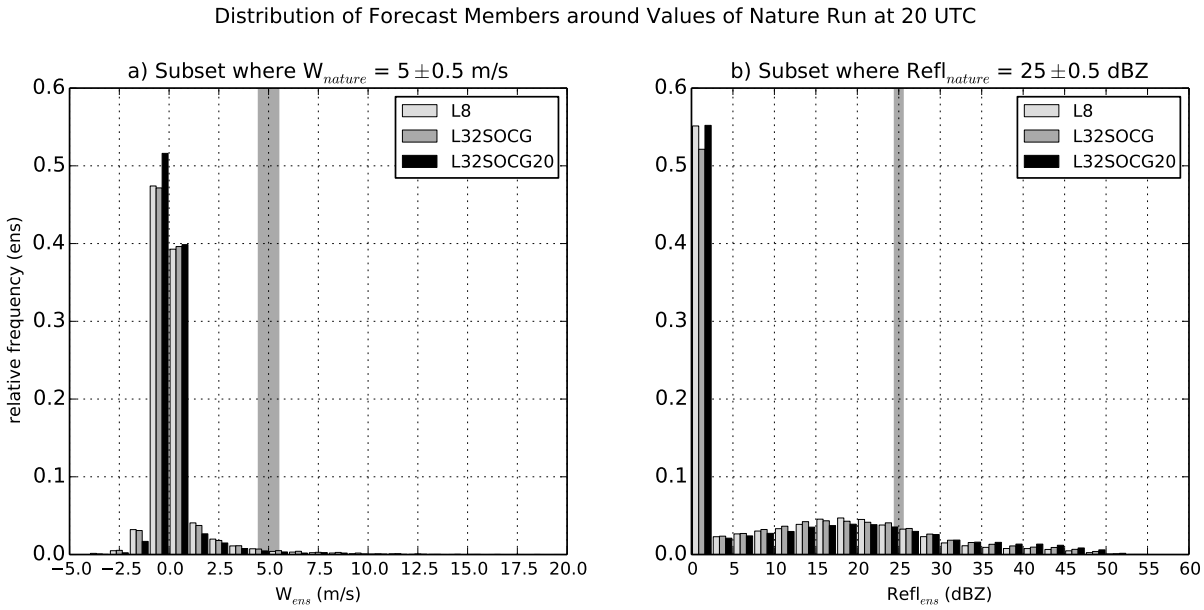

Figure 2.18: As Fig. 2.8, but for the forecast ensemble-members at 20 UTC after 3 hours of ensemble forecast (cf. Figs. 2.15 and 2.17)

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. 2.14 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. 2.10, is strongly penalized due to the random storm positions in the single free control 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. 2.19, 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 updrafts 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)). This is also evident through the rapid growth of spread for storm-core variables like W, $Refl$, QR, QG and QI in Fig. 2.6 after 17 UTC. In L32SOCG20 however, almost no spurious convection is seen in the 1 hour forecast outside of the organized convective systems (cf. Fig. 2.20) and the respective spread of W, $Refl$, QR, QG and QI grows more slowly and stays below the RMSE of the forecast mean (not shown), as opposed to L8. This indicates that the analysis states of L32SOCG20 are internally more 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.

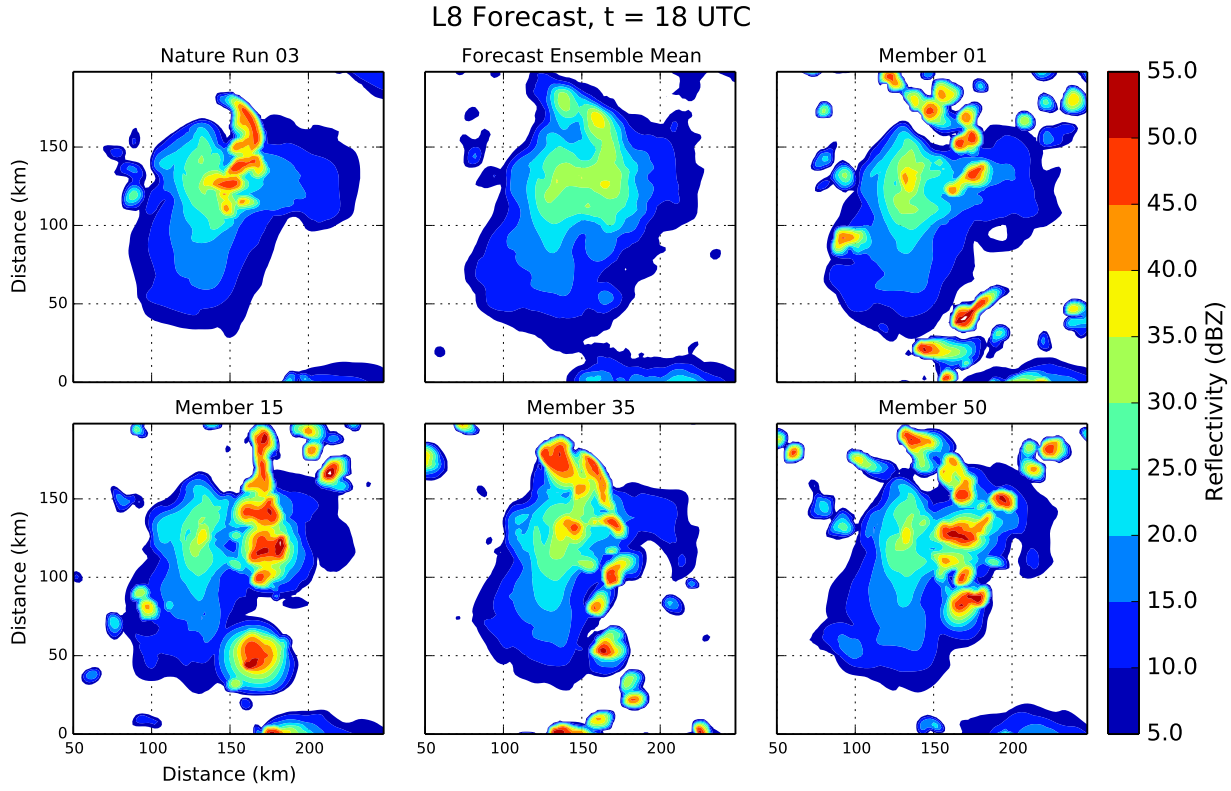

Figure 2.19: Composite reflectivity of Nature Run 03 (cutout of the domain), **Forecast Ensemble Members** 01,15,35,50 and Forecast Ensemble Mean of L8 at 18 after 1 hour of forecast.

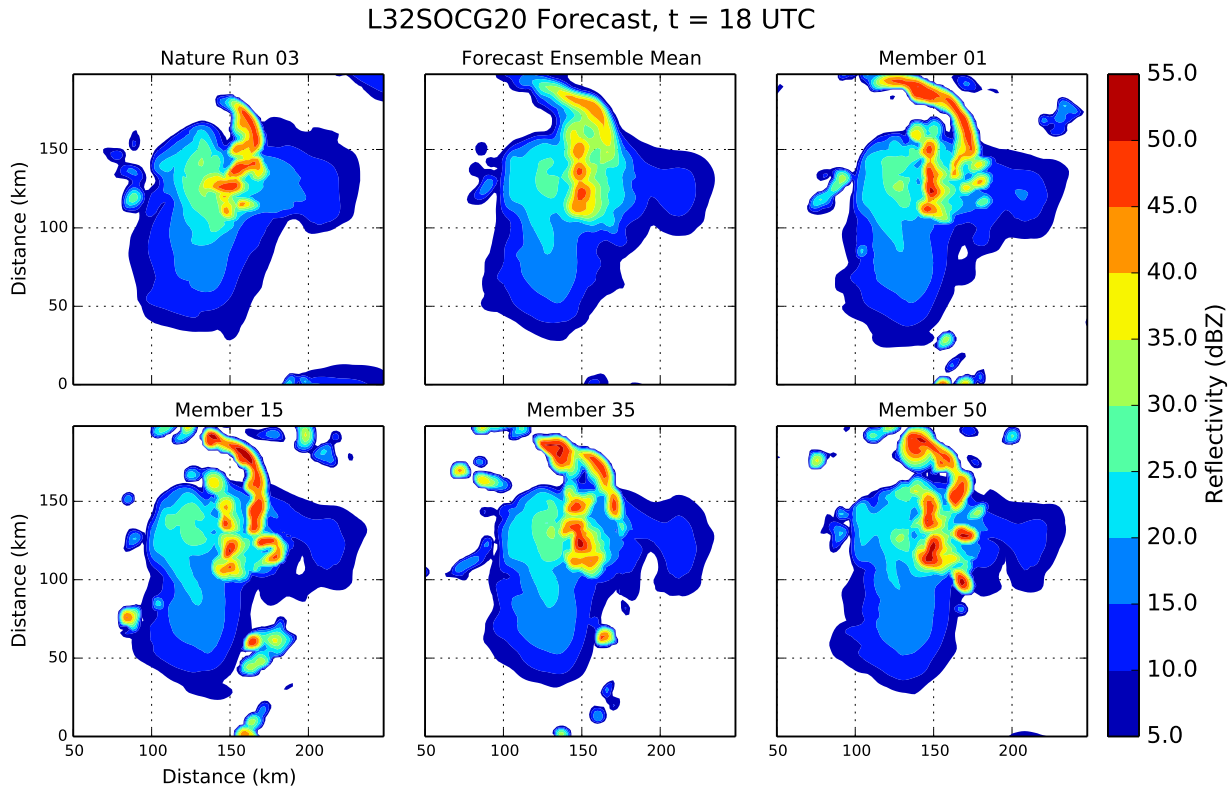

Figure 2.20: As Fig. 2.19, but for L32SOCG20.

2.4.2 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. For this purpose, the composite reflectivity fields of both the nature run and the forecasts are 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. 2.21.

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. 2.22a). The amplitude or SAL-A-component (Fig. 2.22b) compares the domain-wide total reflectivity (not thresholded) between forecast and observation and thus displays the overall bias. The SAL-L-component (Fig. 2.22c) 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.

2.4.3 Spatial forecast verification results

The DIS-score in Fig. 2.21 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. 2.14 which showed the influence of spurious storms in the forecast (cf. Figs. 2.19 and 2.20).

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

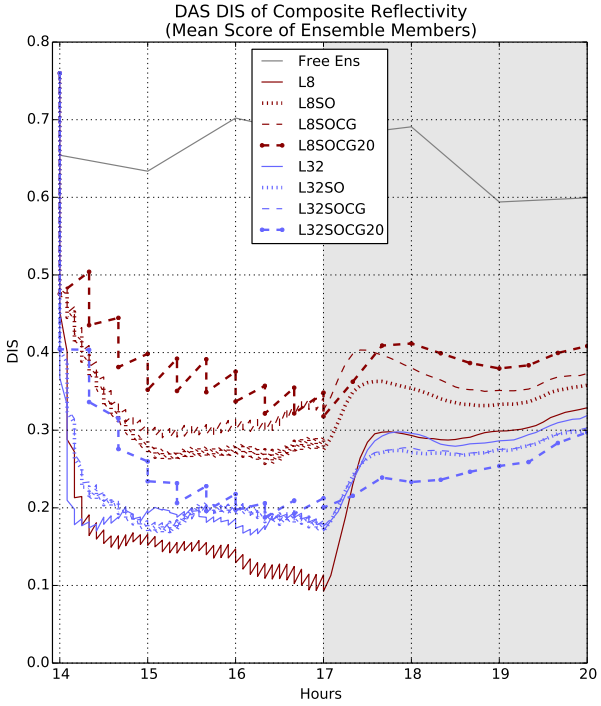

Figure 2.21: DIS-component of the DAS-score applied to the Composite Reflectivity field thresholded by 10 dBZ. Displayed is the mean DIS-score of the ensemble members of all experiments, with the Nature Run as reference. The score is averaged over the five random repetitions of all experiments. (A value of DIS = 0 is a perfect match.)

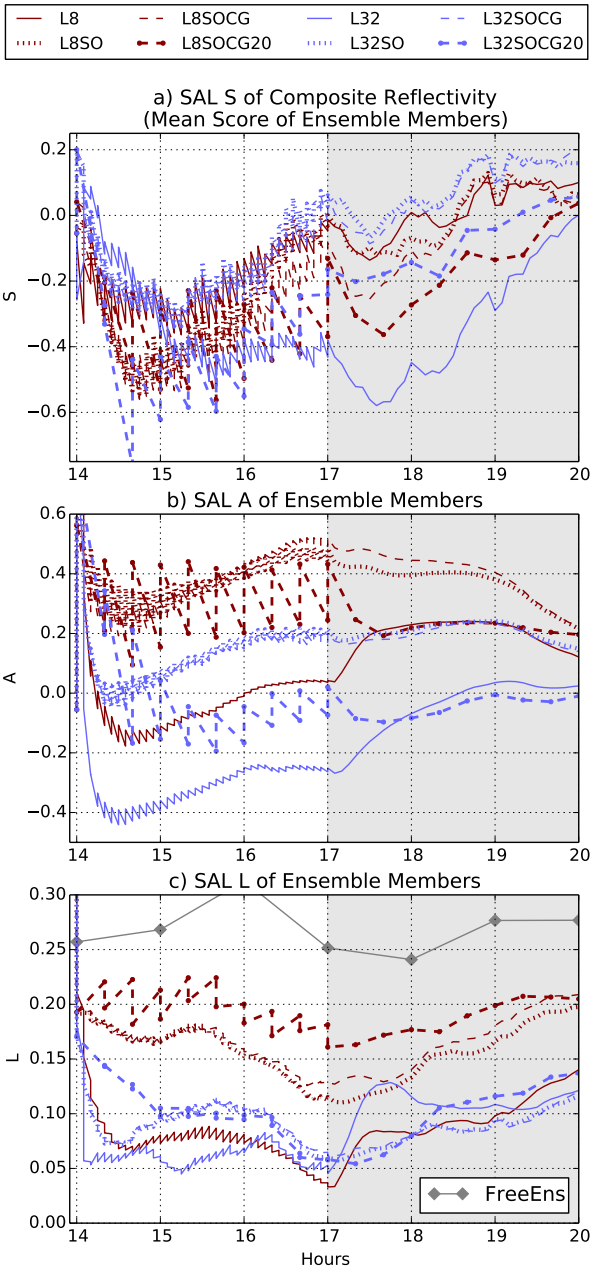

Figure 2.22: SAL score components S, A and L for all experiments, with the Nature Run as reference (see text). The scored field is the Composite Reflectivity, the scores are averaged over the five random repetitions of all experiments.

hour additional small and mislocated convective cells are triggered in L8: additional cells increase the bias SAL-A, small decrease the structure SAL-S, and mislocated cells 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.

2.4.4 "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. 2.5, 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. 2.23, 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.

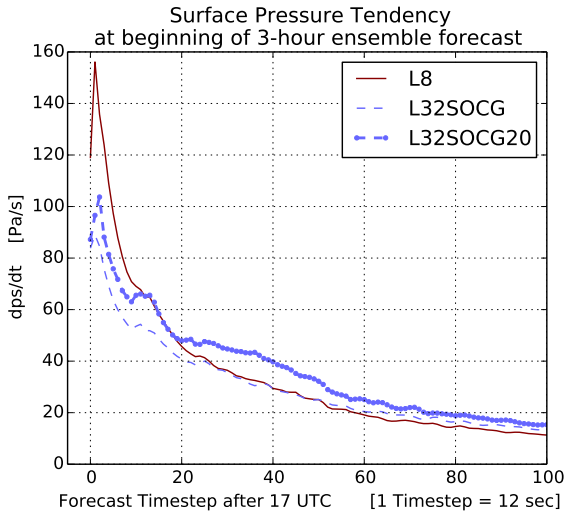

Figure 2.23: Surface pressure tendencies dp_s/dt (domainwide maximum) of the first member (averaged over five repetitions of the experiments) within the first 20 minutes after starting the 3 hour ensemble forecast. Shown are L8, L32SOCG and L32SOCG20.

2.5 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 computed with 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.

2.5.1 Assimilation schemes with different length scales

A reference experiment L8 was set up to reproduce the typical behavior of a convective EnKF DA system in which the exact observation error covariance matrix $\mathbf{R}$ was known and used by the LETKF. Assimilation schemes were devised in which the localization radius of $\mathbf{R}$ 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 under-

sampling due to the large localization radius in combination with full-resolution observations. This undersampling of the background error covariance in L32 resulted in an overconfidence of the filter that converged towards a wrong solution which missed some convective cells. 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.

2.5.2 Ensemble forecasts from analyses with different scales

In the ensemble quantitative precipitation forecasts (QPF), the displacement error of the storm's reflectivity field 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.

2.5.3 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 ensemble 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.

2.6 Discussion and Outlook

2.6.1 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. This means that the analysis states of the members are full scale for all experiments, regardless of the reduction of the physical observation resolution in the SO-experiments or the posteriorily downscaled coarse grid analysis in the CG-experiments. An alternative approach would be to spatially coarsen not only the input-parameters of the transforming LETKF-solver but also the final 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). Damping the analysis noise by the means of more diffusion on the small scales could be an instrument to keep the model state stable during a rapid update cycling. Incremental Analysis Update (IAU) (Bloom et al., 1996) is often used to prevent insertion shocks by distributing the increments of intermittent analyses throughout the forecast window. Also, a digital filter initalization (DFI) (Lynch and Huang, 1992) could be helpful to

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

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, so essential parts of the dynamic signal could be filtered out erroneously.

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.

2.6.2 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 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).

2.6.3 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., 2014) 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). In the present LETKF-framework with intermittent and instantaneous state updates consisting of linear combinations of the member states, any noise that is not dissipated between

subsequent analyses is always fed back into the system and can possibly amplify or resonate. Methods that accelerate the convergence of the filter, such as Running-In-Place, could lead to more consistent states by the end of analysis period, if their increments are smooth enough. 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 further issue is the difficulty 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, 2013). 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.

Chapter 3

Noise and spurious convection in idealized radar DA

In Chapter 2 it was shown that a high-resolution assimilation scheme such as L8 can produce analyses that strongly resemble the observed storms (Figure 2.7). In the forecast of L8, spurious convection arises (Figure 2.19) and taints the forecast skill (e.g. Figure 2.21). The experiment L32SOCG20 has less precise analyses (Figure 2.13), but also less spurious convection (Figure 2.20).

The two main differences of the opposing experiments L8 and L32SOCG20 are the spatial and temporal parameters of the assimilation. L8 assimilates observations with a spatial resolution of $\Delta x_{obs} = 2$ km, localized with $r_{Loc,h} = 8$ km, while L32SOCG20 has $\Delta x_{obs} = 8$ km, localized with $r_{Loc,h} = 32$ km (see Table 2.1). For L8, this is done with an interval of $\Delta t_{ass} = 5$ min, and L32SOCG20 uses $\Delta t_{ass} = 20$ min. The present chapter is devoted to the question whether the temporal or the spatial parameters during the assimilation cause a decline in forecast quality. Apart from the parameters that are active at every intermittent assimilation, other parameters such as the starting time of the background ensemble shall also be investigated.

In Sections 2.4.4 and 2.5.3 it was hypothesized that the development of spurious convective cells might be related to gravity wave noise. For this reason, surface pressure tendencies during the forecast were displayed in Figure 2.23. In radar data assimilation experiments with real observations, Bick et al. (2016) found that the surface pressure tendencies may not relax towards the level of a free model run during the cycling if analysis updates are performed with a 5 minute interval. It is therefore necessary to look at the properties of hydrostatic imbalance for the experiments of Chapter 2. For global data assimilation, this quantity has been used to describe the balance of analysis states with respect to the model dynamics (e.g. Greybush et al., 2011). Also for global models, DA algorithms can be devised that update not the wind variable, but instead the balanced stream function which then results in more balanced analyses (Kepert, 2009). Deep convection, as opposed to synoptic scale dynamics, mainly consists of divergent motion which does not follow simple constraints like geostrophic or thermal wind balance (Holton, 2004). It is therefore necessary for the present setup to find a measure that describes whether the abundance of vertical motion in the background ensemble members resembles

the truth of the nature run. For example, Gottwald (2014) proposes the mean and variance of model variables associated to gravity waves as a predictor for the dynamical imbalance of analyses and forecasts, when compared to a climatology or a free reference run. The model variable that is mainly associated to gravity waves in the present model is vertical velocity.

Organized and long-lived convection is coupled to the underlying cold pools (Houze, 1993). Therefore, it shall first be examined whether the dynamics of the spurious convection of the ensemble forecasts are cold-pool driven or if their triggering is spatially decoupled. If a decoupling is visible, it could be sensible to evaluate the properties of vertical motion inside and outside of the storms to learn about the influence of the assimilation on the dynamics of the model.

Chapter 2 showed that under certain circumstances the most rigorous assimilation schemes do not necessarily deliver the best results in the following forecasts. In their data assimilation experiments with varying spatio-temporal scales, Zeng and Janjić (2016) suggested a connection between noise in the analyses and forecast quality. This question is immanent to the present experiments, since the possible existence of a trade-off between precision and noise would have strong consequences on the operational use of convective DA.

Section 3.1 describes the cold pool coupling evaluation, leading to the vicinity-masking described in Section 3.3. Section 3.2 describes the measures for the abundance of vertical motion. The experiments for this Chapter are described in Section 3.4 and the results are presented in Section 3.5. Section 3.6 discusses the implications on further experiments with real observations which are presented in Chapter 5.

3.1 Cold Pool Coupling

In the spin-up of the nature run (Figures 2.2 and 2.1), some convective systems survive the stage of small precursor-cumuli and grow into large systems after 11 UTC. The development of new updraft cores is closely coupled to the gust front of their underlying cold pool. The spatial propagation of the gust fronts and the rain-maxima is therefore synchronous. The spurious convection in L8, as depicted in Figure 2.19, arises distant to the gust front of the cold pools that underly those analysis storms that matched the observations closely (Figure 2.7).

To quantify the coupled and the distant triggering, a spatio-temporal correlation $C(\vec{x}, t)$ is computed between cold pools and the emergence of new convective cells. The field of cold pool edges is defined as the absolute value of the temperature gradient $|\vec{\nabla} T(\vec{x}, t_0)|$ in the lowest model level at the starting time of the free forecast (t_0 = 17 UTC). The largest temperature gradient is located at the leading edges of the cold pools in the north-eastern direction (Figure 3.1). The field describing new convection is defined by the absolute column condensation rate $dCond/dt(\vec{x}, t)$, which has a dipole structure with positive values for new cells and negative values for decaying and precipitating cells behind them. The field that describes the gust front (which triggers new cells) is given by $|\vec{\nabla} T[\vec{x}(t), t_0]|$ (Figure 3.1) with a moving frame of reference

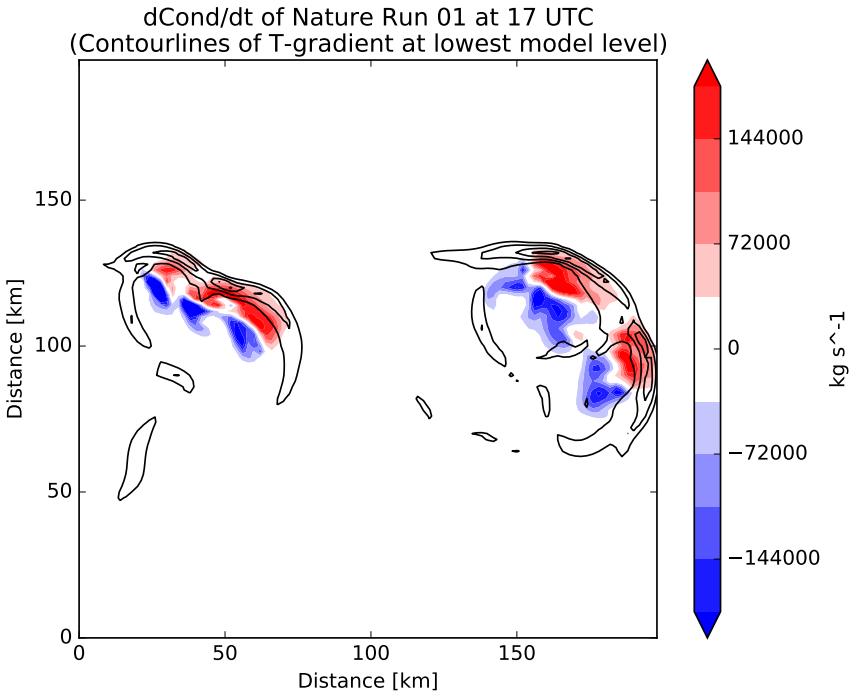

Figure 3.1: Column Condensation Rate (dCond/dt) of nature run 01 at 17 UTC. Black contour lines show the absolute value of the horizontal temperature gradient at the lowest level (lines every 0.0025 K m^{-1}). The storm systems move towards the north east of the domain.

$\vec{x}(t) = -\vec{v}(t) \cdot \Delta t$. The moving frame of reference follows the typical storm motion with $\vec{v} = (v_x, v_y) = (10, 8)$ ms^{-1} to the north east. It was deduced from a time-series like in Figure 2.2. Every 5 minutes from 17 to 18 UTC, the spatial correlation between $dCond/dt(\vec{x}, t)$ and $|\vec{\nabla} T[\vec{x}(t), t_0]|$ is computed, giving a non-isotropic pointwise two-dimensional correlation field $C(\vec{x}, t)$. The values of $C(\vec{x}, t)$ are evaluated along the direction that is oriented parallel to the storm motion in order to catch the signal of rising and decaying convection before and behind the triggering cold pool edge.

The result is shown in Figure 3.2. The nature run exhibits a clear streak of positive correlation which, due to the moving frame of reference, is almost stationary through time. This isolated streak of positive correlation is considered as the main reference feature of the *coldpool coupling*, and deviations from it are considered to be *uncoupled* features. In the direction behind the storm propagation (in the plot: to the left), a negative correlation is visible which means that the rate of precipitation and evaporation is largest behind the coldpool edges. Ahead of the propagating gust front (in the plot: to the right), some weak secondary triggering maxima are visible. For L8, a signal of triggering spurious cells can be seen by a broad pattern of positive correlation ahead of the original gust fronts, e.g. at the coordinates of (0-50 km, 5-20 min) in Figure 3.2b. Since this positive signal of spurious convection arises on a phase line in the plot that is faster than the cold pool propagation speed of 12 ms^{-1} (vertical lines), it is regarded as

decoupled from the coldpool edges. The experiment L32SOCG20 exhibits less spurious signal, supporting the previous evaluations.

The moving frame of reference was chosen by a manual evaluation of time-series plots of the nature run. This could be regarded as a limitation on the interpretation of the cold pool coupling results, because the cold pools of the L-experiments might move with a different speed than the ones in the nature run, e.g. due to a cold bias that causes a faster density current. A possible indication of this is visible in the different tilting of the main correlation streak in L8 and L32SOCG20. Still, it would not describe the broad pattern of decoupled convection in the frontal vicinity which can be regarded as a solid evidence to the decoupled emergence of spurious convection.

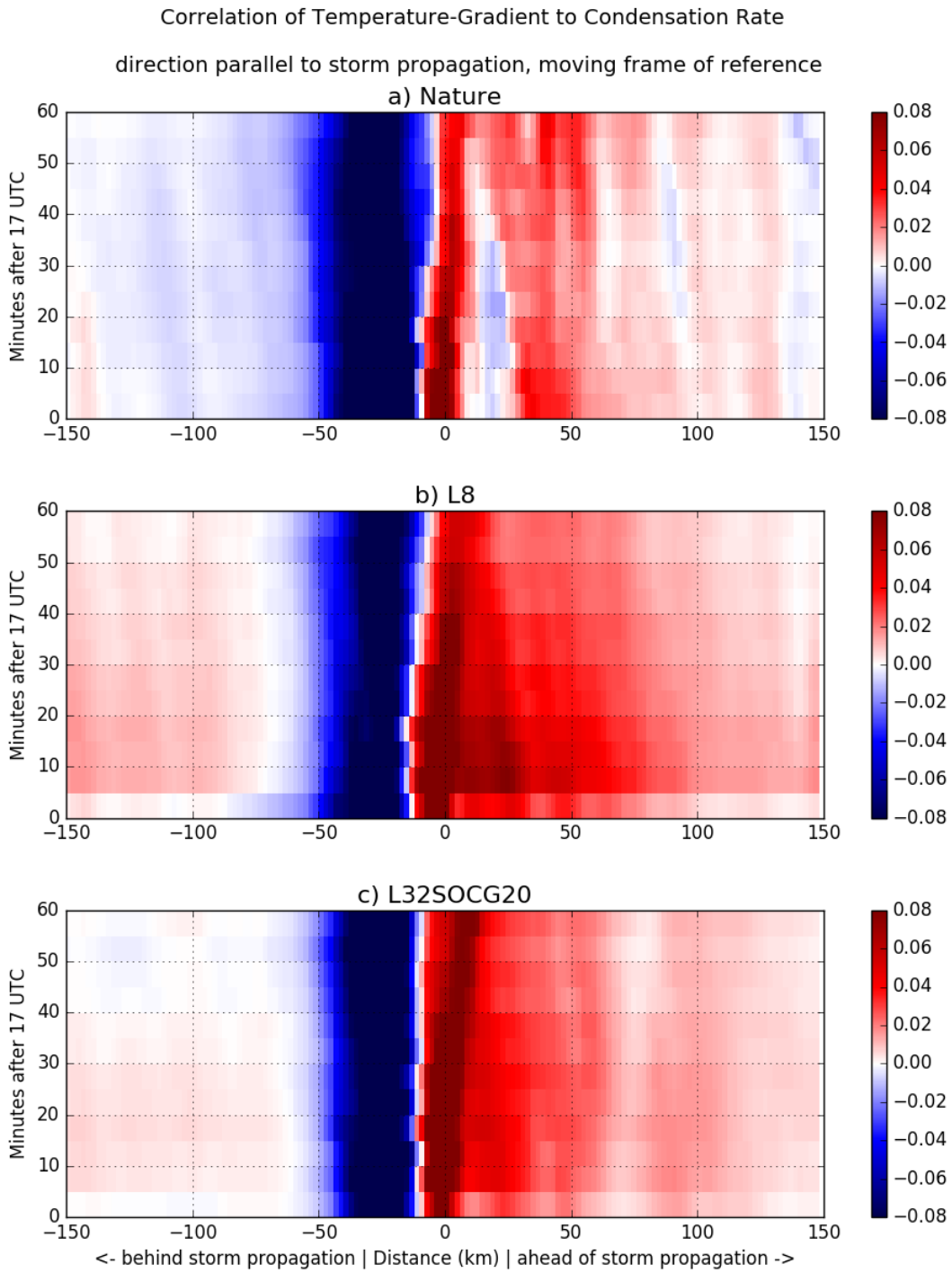

Figure 3.2: Coldpool coupling $C(\vec{x}, t)$ between $dCond/dt(\vec{x}, t)$ and $|\vec{\nabla} T(\vec{x}, t_0)|$ for nature run, L8 and L32SOCG20. The x-axis describes the correlation distance between the grid points, the y-axis describes the time. For the ensemble forecast, the correlation is computed separately for every member field. Due to the moving frame of reference, vertical lines are associated to a phase speed of roughly 12 m/s (see text).

3.2 Measures for abundance and variability of vertical motion

In Sections 2.4.4 and 2.5.3 it was hypothesized that the development of spurious convective cells might be related to gravity wave noise, and surface pressure tendencies during the forecast were displayed in Figure 2.23. Figure 3.3 shows horizontal and vertical slices of the vertical velocity W. In the left part of the vertical slices, a convective system is depicted and it is apparent that the LETKF-generated analyses are able to capture position and intensity of the true convective activity. Yet in clear air, only L32SOCG20 resembles the nature run with respect to the amount and intensity of dry vertical motion. In L8 the clear air regions appear very noisy. In all cases, the vertical motion directly above the surface does not appear to reflect the dynamical intensity of the overlying layers, which means that the surface pressure tendencies might give only limited insight on how balanced the members are. Therefore it is advisable to look at the amount of absolute vertical motion W_{abs} on a pressure level p which is a standard diagnostic output of COSMO given by

$$W_{\text{abs}} = \overline{|W(p)|}. \tag{3.1}$$

Due to its definition, large-scale subsidence or rising motion contribute to W_{abs}. Therefore, the mean $\overline{W}$ and variance $\text{var}(W)$ of the vertical velocity W is split here additionally as

$$\overline{W} = \frac{1}{n}\sum_{i}^{n} W_i \ , \ i \in M \tag{3.2}$$

and

$$\text{var}(W) = \frac{1}{n}\sum_{i}^{n}(W_i - \overline{W})^2 \ , \ i \in M \tag{3.3}$$

where M denotes the set of all model grid points or a specific subset of n points. When an ensemble experiment is referred to, the expression $\text{var}(W)$ shall describe the same quantity, but averaged over the members $[1, \ldots, k]$, given by

$$\text{var}(W) = \frac{1}{k}\sum_{l}^{k}\left[\frac{1}{n}\sum_{i}^{n}(W_i - \overline{W})^2\right] \ , \ i \in M \ , \ l \in [1, \ldots, k] \tag{3.4}$$

When comparing nature run and ensemble members of an experiment, the definitions of (3.3) and (3.4) shall be used equivalently. Note that in the absence of large scale vertical motion, the values of (3.1) and (3.3) describe essentially the same phenomenon: the amount of vertical motion that deviates from the (zero) mean state. In case of (3.3), extreme values are given a higher weight due to the quadratic norm that uses the square of the differences from the mean.

The variance of W along the ensemble axis can be regarded to see how similar the ensemble members are among each other. This shall be done by the means of

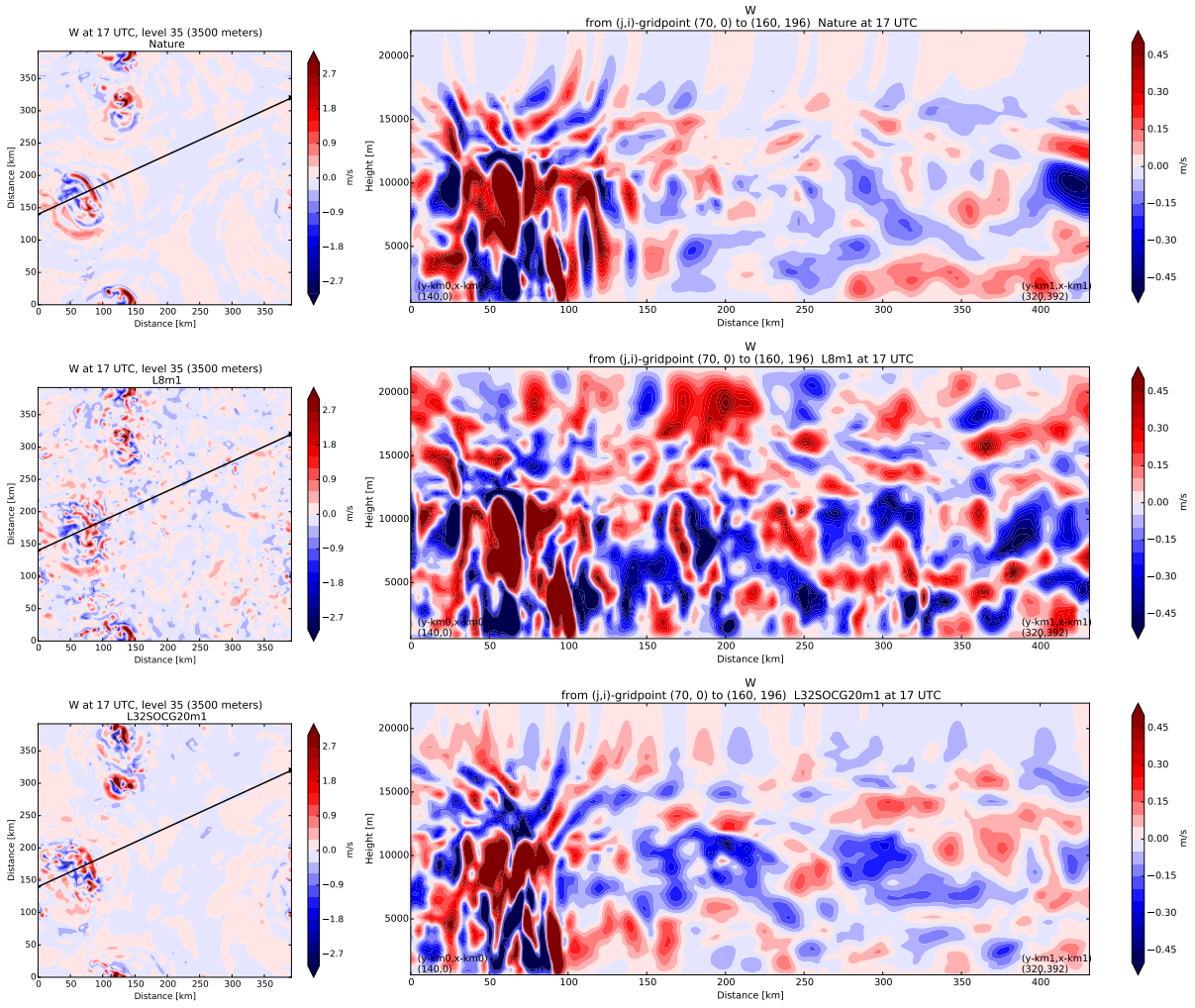

Figure 3.3: Vertical velocity W of Nature Run (top, random repetition 3 of 5) and the corresponding analysis member 1 of experiments L8 (middle) and L32SOCG20 (bottom) at the last analysis time of 17 UTC. Left are horizontal slices through the domain, the black line demarks the plane on which the vertical slices (right) are displayed.

$$\mathrm{var}_{\mathrm{ens}}(W) = \frac{1}{n}\sum_{i}^{n}\left[\frac{1}{k}\sum_{l}^{k}(W_l - \frac{1}{k}\sum_{l}^{k}W_l)^2\right] \ , \ i \in M \ , \ l \in [1,\ldots,k] \tag{3.5}$$

which is the square of the *spread* quantity defined by (2.5).

3.3 Masking of vicinities

The coldpool coupling results of Section 3.1 and the impression of Figure 3.3 motivate to look at the error and statistical properties of the model fields with a distinction of the spaces into three categories: (i) At grid points *within* in the storms, (ii) in their *vicinity*, and (iii) in the *outer environment* between storms. The vicinity is defined in a geometrical sense with horizontal distance. In Figure 3.4, the masking algorithm is exemplified:

The *inner* grid points are simply given by the three-dimensional set of points in the nature run where the volume reflectivity exceeds 5 dBZ. For the *vicinity mask*, the two-dimensional vertical maximum of the nature run's reflectivity field is taken as the source field, thresholded to 5 dBZ, and the mask is dilated horizontally with a circle of a predefined diameter. From Figure 3.2, a radius of 16 km appears reasonable, since a relatively wide region distant to the storm cores is already covered by the spreaded anvils. This means that rain-free regions up to roughly 50 km in front of the cold pool edges (cf. Figure 3.2) are covered by the *vicinity mask* which does not contain the grid-points inside of the storms. All points that are not in of the dilated vicinity-mask are categorized as *outer environment* and the mask is applied on all vertical levels to form a three-dimensional volume mask. Figure 3.5 shows how many points the different masks comprise. To gain a full impression, the reader is advised to look again at Figures 2.1 and 2.2 to see how number, size and structure of the storm systems influence the masks over the forecast time. For example, at 10 and 11 UTC the vicinity masks covers almost the full domain, since the small precursor storms are numerous and evenly placed with only small distances between them.

Note that the masks are defined by the nature runs and then applied to the nature runs themselves *and* at the same time to the members of forecast and analysis ensembles. Since the displacement errors of the analysis storms are relatively small (Figure 2.3), the *inner* masking points of the truth also contain storms in the ensemble members, at least during the cycling. Still, the point-set of the nature run's "dry" outer storm environment may also contain "wet" spurious convective cells when applied to the ensemble members, which is especially the case in the forecast. The growth of spurious convection in the forecast ensemble members *within* the outer region is expected to be distinguishable from slow displacement of storms that had been correctly positioned in the last analysis, since spurious growth is observed to be very quick (Figure 3.2), while the displacement of good storms is rather slow (Figure 2.20).

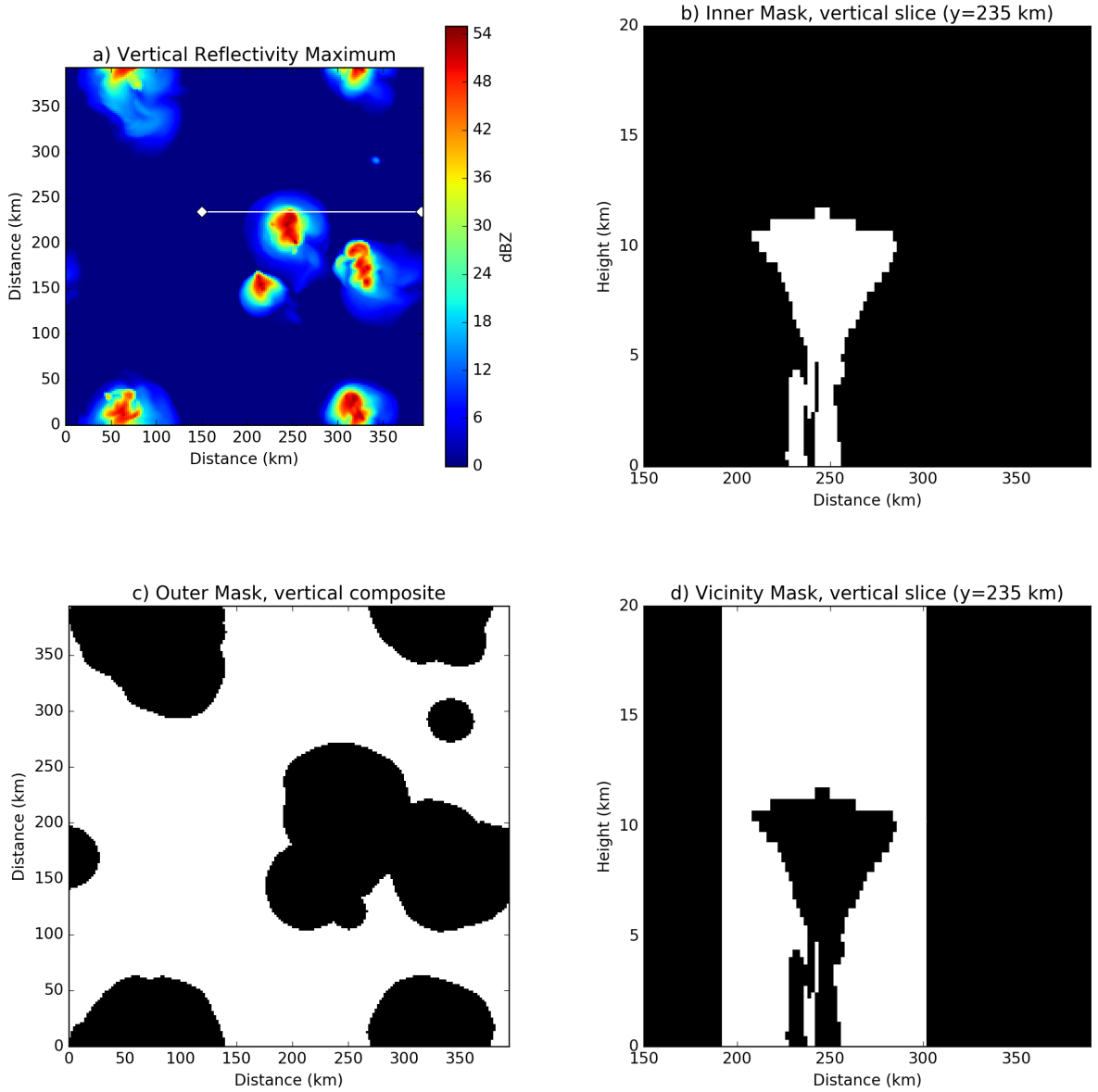

Figure 3.4: Vicinity masking for one state of a nature run. The reflectivity (a) is the source field, the white line in (a) marks the vertical slice that is depicted for the inner mask (b) of grid points with Refl $>$ 5 dBZ and the vicinity mask (d) with a horizontal dilation of 16 km. The remaining grid points in (c) form the outer mask. White color in the mask-plots demarks valid grid points.

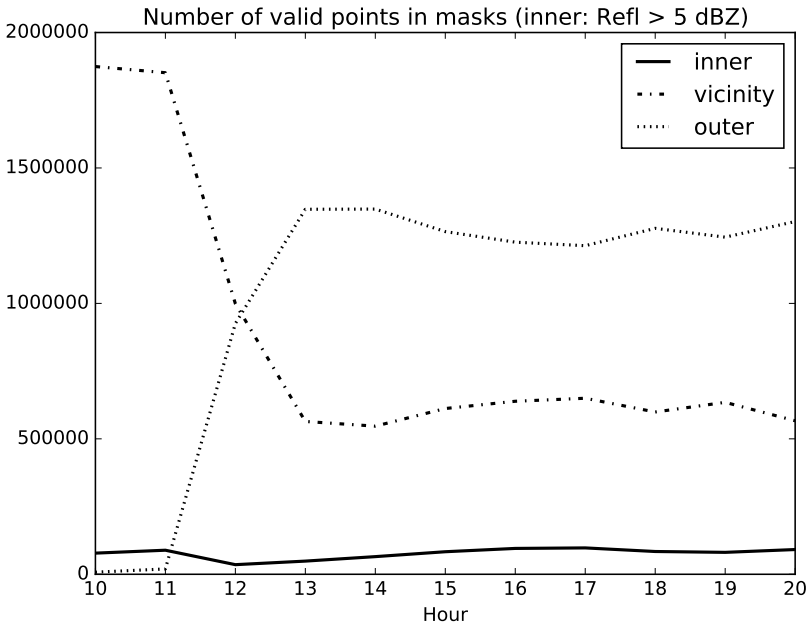

Figure 3.5: Time series of number of points in the masks (Figure 3.4) with respect to the nature runs.

3.4 Experiments on spatial and temporal scales of the analysis

L8 and L32SOCG20 (Table 2.1) are now regarded as reference experiments. The experiment L32SOCG (also Table 2.1) has $\Delta t_{ass} = 5$ min and is otherwise equivalent to L32SOCG20. An additional sensitivity experiment L820 is introduced with $\Delta t_{ass} = 20$ min which is otherwise equivalent to L8. The sensitivity of the assimilation system to multiples of the temporal parameter $\Delta t_{ass} = 5$ min can now be investigated by the pairs of (L8, L820) and (L32SOCG, L32SOCG20). The sensitivity to multiples of the spatial parameter combination $4\Delta x_{obs} = 4\Delta x_{ass} = r_{Loc,h}$ can be investigated by the pairs of (L8, L32SOCG) and (L820, L32SOCG20). An additional sensitivity experiment L8T10 is performed that is started at 10 UTC and is otherwise identical to L8, which will be discussed in Section 3.5.4.

3.5 Results

3.5.1 Non-hydrostatic properties

Figure 3.6 shows the surface pressure tendencies (cf. Figure 2.23) throughout the whole experimental time. The two groups of sensitivity experiments can be distinguished by two features: (i) (L820, L32SOCG20) relax towards relatively low levels of noise by the end

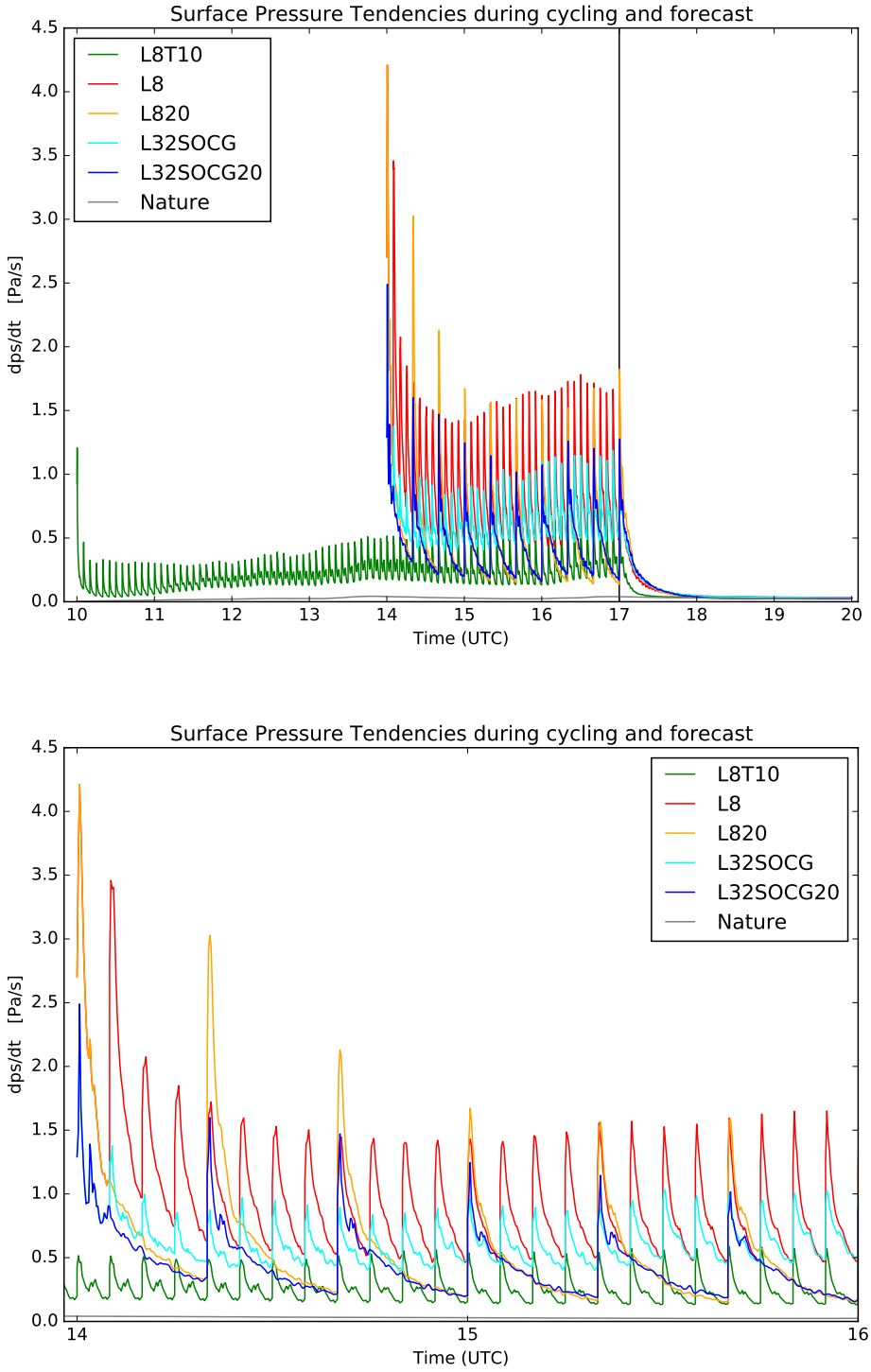

Figure 3.6: Top: Surface pressure tendencies of one random realization of nature run and experiments (L8T10, L8, L820, L32SOCG, L32SOCG20) (cf. Figure 2.23), during cycling (10 or 14 to 17 UTC) and 3 hour forecasts (17 to 20 UTC). Bottom: Detail of top picture.

of their forecast intervals when compared to (L8, L32SOCG). (ii) (L8, L820) experience stronger initial noise after starting from the intermittent analyses than (L32SOCG, L32SOCG20). These two features mean that, during the cycling, analyses of (L32SOCG, L32SOCG20) are more balanced due to their coarser spatial parameters and forecasts of (L820, L32SOCG20) are more balanced due to their longer time interval wherein noise is being damped by the model dynamics.

3.5.2 Abundance of vertical motion

A time-series of the unmasked W_{abs} from (3.1) is depicted in Figure 3.7 for different pressure levels. It is apparent that the members of (L8, L32SOCG) overestimate the amount of vertical motion by the end of the cycling period and throughout the forecast. Due to the absence of masking this can either hint at storms that are too strong or too large in number and therefore spurious. The members of (L820, L32SOCG20) mostly underestimate W_{abs} by the end of the cycling, but converge to the nature run after 3 hours, meaning that their amount of vertical motion is correct then.

Using var(W) from (3.3) and (3.4) as a measure of the abundance of vertical motion in the ensemble members and the nature run, results for the *inner*, *outer* and *vicinity* spaces are shown in Figure 3.8. The large difference between nature run and the free reference ensemble without data assimilation is caused by the application of the nature's mask to the free ensemble members in which the positions of the convective cells are completely uncorrelated to the nature run. In the clear air outer regions (left panel), the Δt_{ass}-grouping of experiments with DA (that was indicated by Figure 3.7) is visible again: The experiments with $\Delta t_{ass} = 20$ min converge to the same abundance of vertical motion by the end of the cycling. In their forecast, var(W) starts to increase only after 40 minutes which is caused by storm cores in the members that displace themselves into clear air regions with respect to the nature run. The experiments with $\Delta t_{ass} = 5$ min grow to a larger abundance of vertical motion in the observed outer space during the cycling, and a strong overgrowth is visible in the forecast, which is a sign of spurious convection. While the cycling interval of L8 and L32SOCG is the same, the latter exhibits slightly smaller values of var(W) in the outer space and slightly larger values in the vicinity, which can only be attributed to the coarser spatial parameters of the experiments. In the vicinity region (middle panel), a grouping with respect to abundance of var(W) is not discernible during the cycling period. In all experiments, the DA-increments are acting as suppressors of spurious vertical motion. Since with $\Delta t_{ass} = 20$ min there is more convective growth during the forecast intervals, the suppressing increments of (L820, L32SOCG20) are also larger. In the forecast of the vicinity mask, the Δt-grouping is visible again by the feature of a strong spurious convective overgrowth in (L8, L32SOCG) as indicated in the cold pool decoupling (Figure 3.2). In the inner regions (right panel), no clear grouping is visible both during the cycling and the forecast window.

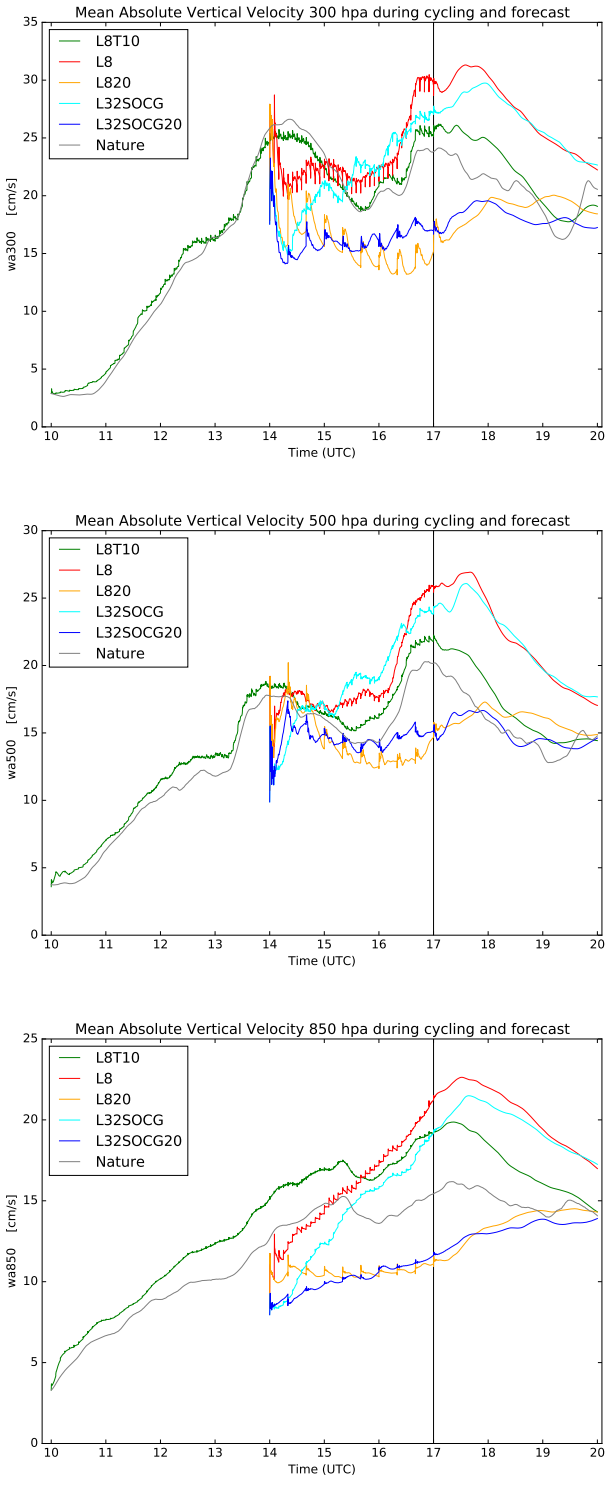

Figure 3.7: W_{abs} for nature run, and experiment (L8T10, L8, L32SOCG, L820, L32SOCG20) for one random realization, during cycling (10 or 14 to 17 UTC) and 3 hour forecasts (17 to 20 UTC), for pressure levels of 300 hPa (top), 500 hPa (middle), 850 hPa (bottom).

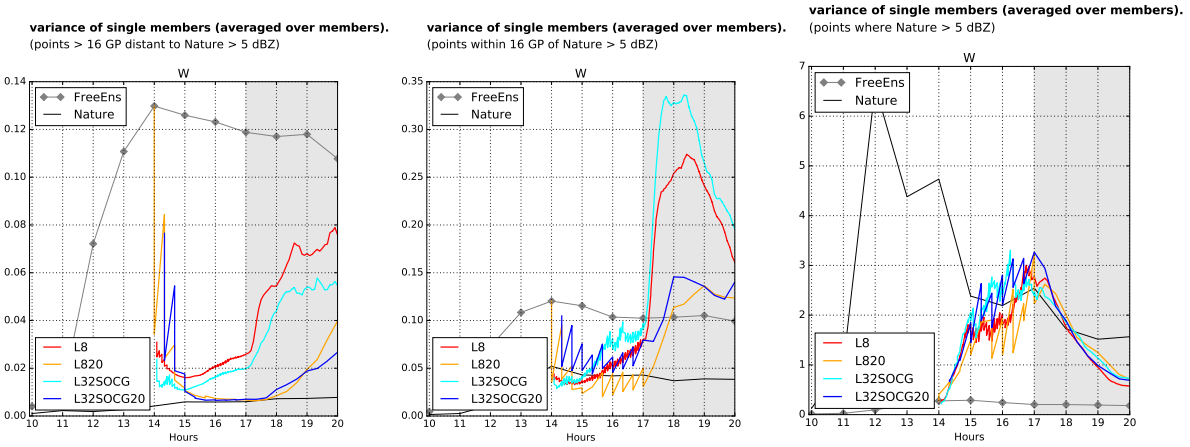

Figure 3.8: Variance of W (averaged over all ensemble members) for (L8, L32SOCG, L820, L32SOCG20) during cycling and forecast, for outer (left), vicinity (middle) and inner space (right) with respect to the reflectivity of true storms in the nature run (five random repetitions). Note that for some lines the first assimilation and cycling step is missing due to lost data – their starting point is always at 14 UTC at the value of the free ensemble. The free ensemble is depicted as a grey line, the nature run is the black line.

3.5.3 Ensemble Spread

As depicted in Figure 3.9, $\mathrm{var}_{\mathrm{ens}}(W)$ behaves in the same way as $\mathrm{var}(W)$ (Figure 3.8) for the spaces outside of the storms and in their vicinity (left and middle panel) during the cycling and forecast windows. During the cycling window and inside the storms (right panel of Figure 3.8), $\mathrm{var}_{\mathrm{ens}}(W)$ of (L8, L820) is forced onto small levels at every analysis due to the smaller spatial parameter of their LETKF-scheme. L32SOCG20 has the largest analysis and background ensemble variances, while L32SOCG exhibits an intermediate behavior with an analysis ensemble variance smaller than L32SOCG20 and background ensemble variances that do not grow as large as in L32SOCG20 within the shorter cycling interval of $\Delta t_{ass} = 5$ min. In the first 40 minutes of the forecast window in the storm-inner space, $\mathrm{var}_{\mathrm{ens}}(W)$ of (L8, L32SOCG) strongly outgrows the general level that was reached by their background ensembles every 5 minutes during the cycling. This means that all information on small-scale structure that was forced onto the L8-members (with a benefit in skill, cf. 2.10) during the cycling is very quickly lost in the forecast. This feature of variance-overgrowth is less apparent for L820 and even less so for L32SOCG20.

3.5.4 Recycling of initial noise

As Figures 2.9 and 2.10 have shown, the very first intermittent analysis step produces the most dominant increments with respect to the model states of the ensemble members. This happens because the storms in the members have fully developed by that time (cf. Figure 2.2) and are randomly positioned, thus the ensemble spread is already very large. A large ensemble spread in the background error covariance matrix of equation (2.1) gives

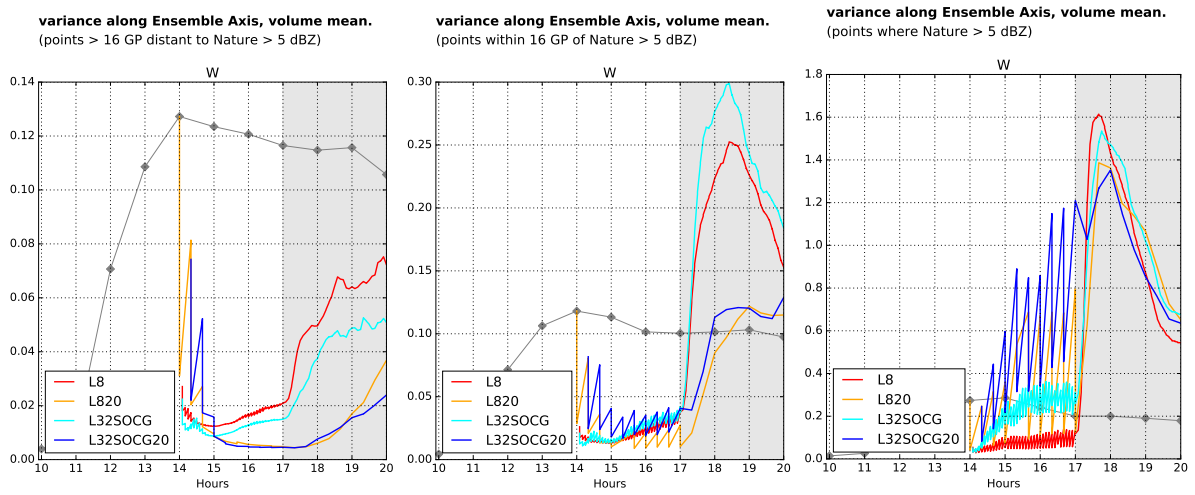

Figure 3.9: Mean pointwise ensemble variance of W for (L8, L32SOCG, L820, L32SOCG20) during cycling and forecast, for outer (left), vicinity (middle) and inner space (right) with respect to the reflectivity of true storms in the nature run. The free ensemble is depicted as a grey line.

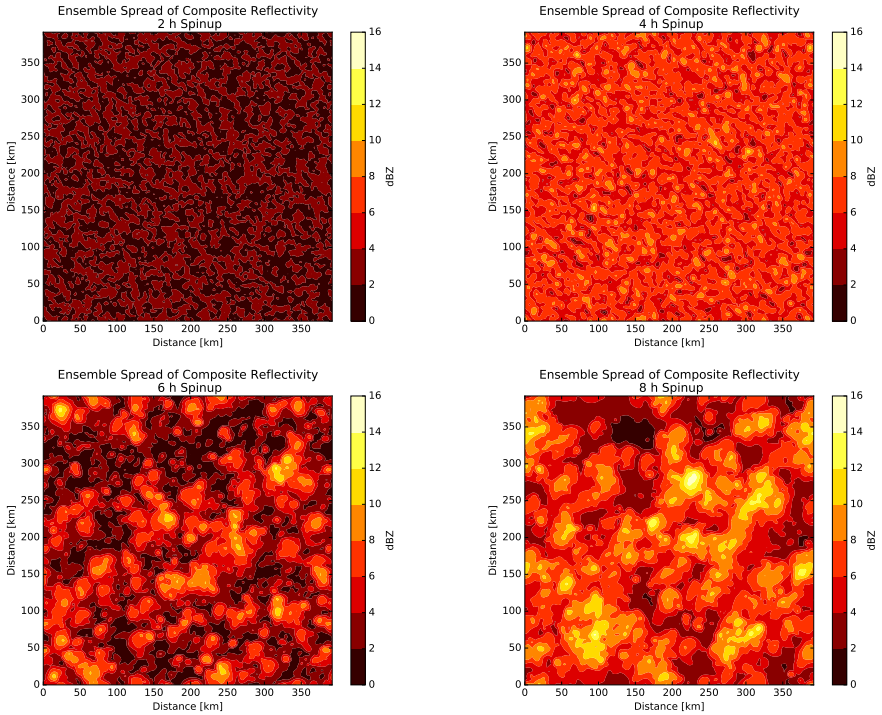

Figure 3.10: Spread of the composite reflectivity field of 50 free ensemble members (one random repetition) after 2,4,6 and 8 hours of spin-up, starting time was 6 UTC (cf. Figure 2.2, 14 UTC equals 8 h spin-up).

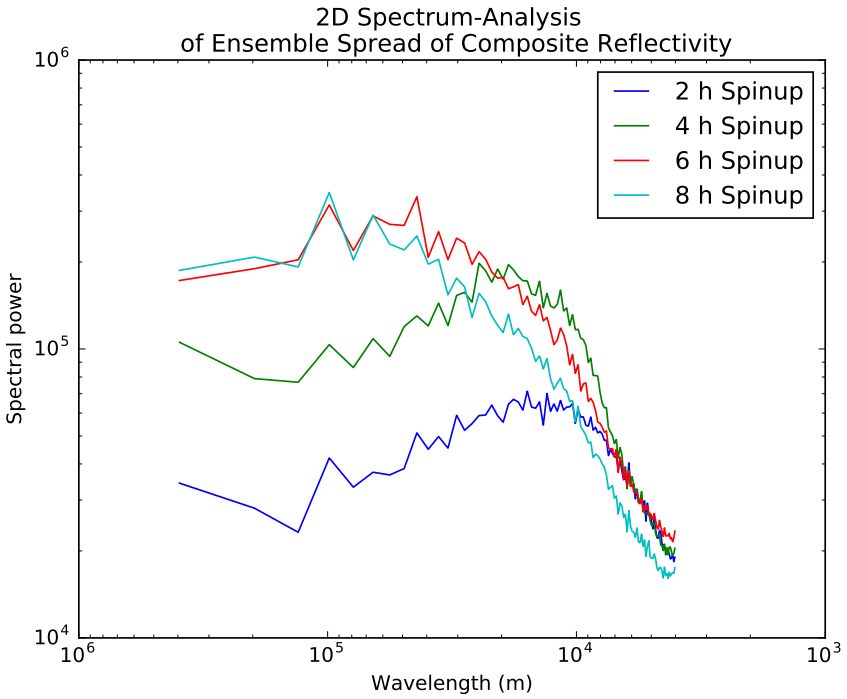

Figure 3.11: Power Spectrum of the spread fields of composite reflectivity for all five repetitions, exemplified in Figure 3.10.

the observations, whose error covariance matrix is fixed here, a large weight.

Figures 3.10 and 3.11 show the temporal evolution of the ensemble spread and its spectral decomposition (cf. Section 4.7) during the spin-up. It is apparent that at early times such as 10 UTC (4 h spin-up) the spread is small in scale and amplitude and relatively homogeneous, as the distribution of small showers in Figure 2.2 had already indicated. The large number of points in the storm-vicinity at 10 UTC (Figure 3.5) illustrates that an experiment that starts early will produce localized increments of the observed small showers that influence much more adjacent grid points than at 14 UTC.

At the later times of 12 and 14 UTC, the spread structure is dominated by the signal of large-scale storms. These have a smoother shape than before and thus more energy in the large scales and less energy in the small scales (Figure 2.2). Since the storms are always randomly positioned and the ensemble size is very limited, the spread distribution is not as homogeneous as before (e.g. at 10 UTC) and gaps exist in the spread field where the members randomly agree on the state of "no storm forecast". Due to the random initialization (cf. Section 2.2.1), there is no rule that a certain location should be storm-free in all members at a certain point in time – which will be problematic if a storm is observed exactly on that particular location in the nature run. Therefore, the ensemble does not cover all possible states, so the background error covariance matrix is rank deficient, even when localized to a single grid point.

The experiment L8T10 investigates the influence of an earlier start of L8 at 10 UTC. Figure 3.12 shows that the first increment of L8T10 is much smaller than for L8 and the

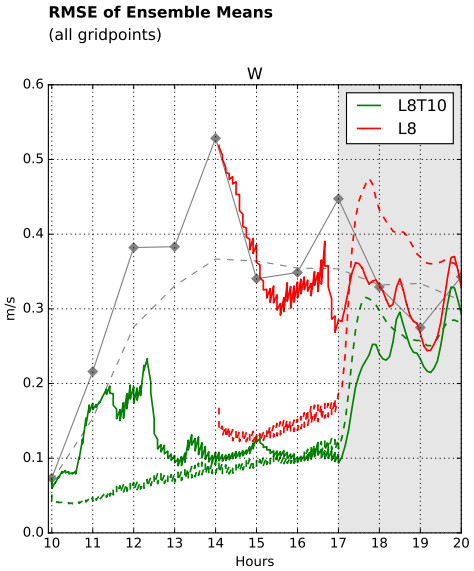

Figure 3.12: RMSE (solid) and spread (dashed) of the ensemble mean of L8 and L8T10 through the assimilation (10 or 14-17 UTC) and forecast (17-20 UTC) phases of the experiment, with gray shading indicating the forecast phase. The gray lines show the RMSE and spread of the free control ensemble without any assimilation. All grid points are evaluated. The error values are shown for one random repetition of the experiment.

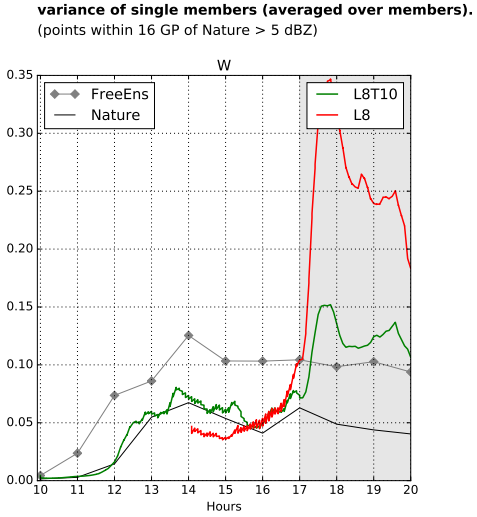

Figure 3.13: Variance of W (averaged over all ensemble members) for one random repetition of L8T10 and L8 during cycling and forecast, for the vicinity region with respect to the reflectivity of true storms in the nature run. The free ensemble is depicted as a grey line, the nature run is the black line.

RMSE of all variables is kept smaller throughout the cycling period, and in the forecast. Figures 3.7 and 3.13 show that the noise properties of L8T10 are favorable to L8, and spurious convection does not grow as explosively. The result can be interpreted such that the noise of the very first increment in L8T10 is small enough to be damped away by the model dynamics and is therefore not fed back as strongly as in L8 during the cycling.

3.6 Summary and discussion

Motivated by Chapter 2, additional experiments have been performed that vary the assimilation interval but keep the spatial properties constant, and vice versa. Two mechanism have been identified that cause noise in the analyses and spurious convection in the forecast. On the one hand, small-scale localization of 8 km combined with grid-scale observations of 2 km resolution cause strong analysis increments that are less hydrostatically balanced compared to the choice of larger spatial parameters. Together with this, an over-abundance of vertical motion is seen in the noisy members, which then grows into spurious convection. This spurious convection is both dislocated with respect to the truth and decoupled from the triggering cold pool mechanisms. On the other hand, an assimilation interval of 5 minutes appears to cause a feedback-growth of noise in the analysis steps, while with an interval of 20 minutes this noise is damped before the next analysis update, leading to less spurious abundance of vertical motion.

Different measures were used to quantify the abundance of vertical motion. *Surface pressure tendencies* mainly gave an indication on gross imbalances that were introduced by the intermittent analysis updates and how quickly they were dissipated. The *amount of absolute vertical motion* provided a first overall impression whether there was too much or too little convection during the cycling and during the forecasts. The *cold pool coupling* method motivated to look at the *mean and variance of vertical velocity* within and outside of the observed storms. By this method, spurious overgrowth could be identified in the vicinity and distance of the true storms, while all ensemble experiments showed a comparably good match to the nature run inside the well-observed inner parts of the storms.

A sensitivity experiment L8T10 was started at a four hours earlier time of the spin-up where scales and amplitude of the background ensemble spread were smaller and had a more homogeneous structure. As a consequence, the previously unfavorable parameter choices of the noisy L8-experiment did not exhibit the same behavior of noise-recycling, since their initial analysis increments were smaller and caused less noise in the initial update step. The forecasts from this experiment were very precise and exhibited no serious spurious convection, just as L32SOCG20. Still, L8T10 must be regarded as an extreme choice of parameters, since such a homogeneous distribution of clouds in the initial ensemble is usually never observed in the reality of mid-latitude convective systems, since they are not triggered in the same way as the very idealized ensemble that is used here. An expected behavior would rather be the observation of a growing thunderstorm that is only contained in few or no background ensemble members. This scenario resembles the conditions of the presented experiments which started at 14 UTC after 8 hours of

spin-up, so the conclusions drawn from experiments which start at that time are still valid for further applications.

With the present DA-setup, the combinations of the experiment L32SOCG20 exhibit a low noise level and do not overly force the analysis ensemble members into accordance with respect to the inner-storm structure. Following the suggestions of Zeng and Janjić (2016), this is a sign that the parameter combination of the latter experiment is well-chosen for forecasts of three hours and should be used as a first orientation for setups with real observations.

Chapter 4

Assimilation of Mode-S EHS Aircraft Observations in COSMO-KENDA

4.1 Introduction

For more than two decades, operational Numerical Weather Prediction (NWP) has benefited from aircraft observations. Typically transmitted by the Aircraft Meteorological Data Relay (AMDAR, WMO, 2003), the aircraft observations provide the Data Assimilation (DA) systems with in-situ measurements of the important dynamical variables of temperature and wind (Daley, 1991). AMDAR data contain vertical profiles near airports and upper-air observations at flight level with cruising altitudes around 10 km above sea level (ASL), corresponding to a layer between 200 and 300 hPa.

A new aircraft data source has been introduced in de Haan (2011). It uses airport radars in the surveillance *Mode-S* to collect traffic control reports of commercial aircrafts from which observations of wind and temperature can be derived. de Haan (2011) has described this data set of Mode-S EnHanced Surveillance (Mode-S EHS)[1] and has found a quality comparable to AMDAR by comparing collocated measurements from the same aircrafts. de Haan and Stoffelen (2012) have assimilated Mode-S EHS observations from Schiphol Airport into the High Resolution Local Area Model (HIRLAM) using the 3dVar algorithm. In the study, HIRLAM is run with a resolution of 10 km and Mode-S observations are assimilated in 3-hourly and 1-hourly intervals. The assimilation of Mode-S data has shown a reduction in wind forecast errors by approximately 5 percent (de Haan and Stoffelen, 2012). Strajnar et al. (2015) also found benefits when assimilating Mode-S observations from a different data source in a local model over Slovenia.

The present study aims to assimilate the Mode-S EHS data with the Kilometer-scale ENsemble Data Assimilation (KENDA) system (Schraff et al., 2016) which is being developed at Deutscher Wetterdienst (DWD). KENDA couples a Local Ensemble Transform Kalman Filter (LETKF, as in Bishop et al., 2001; Hunt et al., 2007) with a 40 member

[1] https://www.eurocontrol.int/articles/mode-s-operational-overview

ensemble of the Consortium for Small-scale Modeling (COSMO) (Baldauf et al., 2011) model in the domain over Germany (COSMO-DE). COSMO-DE is a non-hydrostatic, convection permitting model that runs operationally with a horizontal resolution of 2.8 km and 50 vertical levels. COSMO-KENDA provides a framework for convective-scale ensemble data assimilation of different observation types and aims to replace the previously used nudging scheme for observed model variables and the latent heat nudging scheme (Stephan et al., 2008).

The present study aims at using the additional aircraft Mode-S EHS data to improve in particular the estimates of temperature and wind, in order to provide better background forecasts for the assimilation of radar data. Past studies have assimilated observations from Doppler radars with an Ensemble Kalman Filter (EnKF) in order to forecast convective systems (for example Stensrud et al., 2009). Aksoy et al. (2010) and Stensrud and Gao (2010) stress the important dependence of convective forecasts on correct environmental profiles of wind and temperature which are usually provided by radiosondes and aircraft observations. In the present study, NWP experiments with and without Mode-S EHS are performed, i.e. by supplementing the previously used AMDAR data or by replacing it. For verification purposes, radiosondes are not assimilated and used as fully independent observations.

Since the large amount of additional aircraft observations is expected to reduce the ensemble spread, it is tested how robust the alleviating mechanisms of background covariance inflation (Zhang et al., 2004; Li et al., 2009) and adaptive covariance localization (Perianez et al., 2014) are. Further, estimates of the observation error are made following Desroziers et al. (2005). Thinning experiments are performed that gradually increase the amount of Mode-S EHS data to see whether the assimilation system reaches a point of saturation where no more additional benefit is taken from the observations. Forecasts with lead times up to three hours are computed to see how the error reduction of the Mode-S EHS data persists.

Since any error reduction is related to a change of the model state, a quantification of those model state modifications is of interest in this study, particularly with respect to the assimilated data amount. Using OSSEs of a regional EnKF system that assimilate datasets comparable to the present study, Zhang et al. (2006) and Meng and Zhang (2007) have investigated how the difference total energy grows for selected variables with and without data assimilation. They found that for vertical velocity and cloud variables, the growth of difference energy is mainly in the small scales, while for wind and temperature, the growth in difference is dominated by the larger scales. This was done using a localization radius of 900 km. In the present study, adaptive localization length scales of order 100 km are applied. The true state is not known, so a control experiment without data assimilation will be used as the reference state. It will be evaluated which scales of the horizontal kinetic energy are affected by the data assimilation experiments and how quickly the different scales will relax towards the control experiment during the three hour forecasts.

Section 4.2 describes the conventional observations in COSMO-KENDA and the properties of the Mode-S EHS aircraft observations. Section 4.3 briefly introduces the forecast model COSMO-DE, describes how the LETKF and the adaptive algorithms of the DA

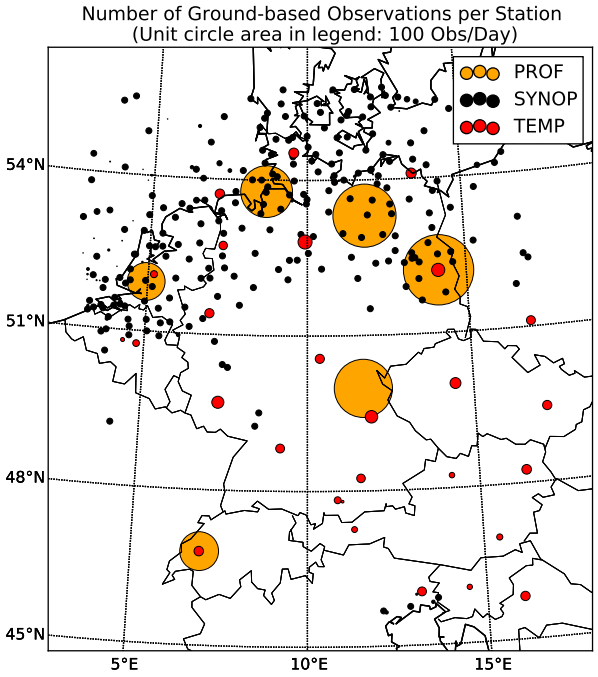

Figure 4.1: Domain extent of COSMO-DE with conventional observations. The circles indicate positions of the surface based stations. The areas of the circles correspond to the average daily number of single observations of the wind variable per station. The average number of observations per day are: 11851 PROF, 5813 SYNOP, 1571 TEMP.

system are set up and how the observation errors are estimated. Section 4.4 describes the experimental period and the different configurations for the assimilation experiments. In Section 4.5, the DA results during the continuous cycling with an interval of one hour and of three hour forecasts are evaluated. Section 4.6 discusses the observation error estimates and how the adaptive KENDA settings act when the Mode-S EHS data density is varied. Section 4.7 discusses the impact of Mode-S EHS data on the kinetic energy spectrum of the forecasts. Section 4.8 summarizes the results and discusses their implications on forecasts with COSMO-KENDA.

4.2 Observations

The conventional observations which are currently assimilated in the COSMO-KENDA system consist of surface pressure and 10-m wind observations from surface stations (SYNOP), upper air observations of wind, temperature and humidity from radiosondes (TEMP) and aircrafts, and wind profiler data (PROF). Fig. 4.1 illustrates the amount and spatial distribution of the conventional data set that is used in the experimental period from 7th till 12th of May 2014. Most of the 30 radiosondes provide vertical soundings of wind, temperature and relative humidity only at 0 and 12 UTC, some also at 6 and 18 UTC. The six wind profilers continuously sample with intervals between 5 and 30

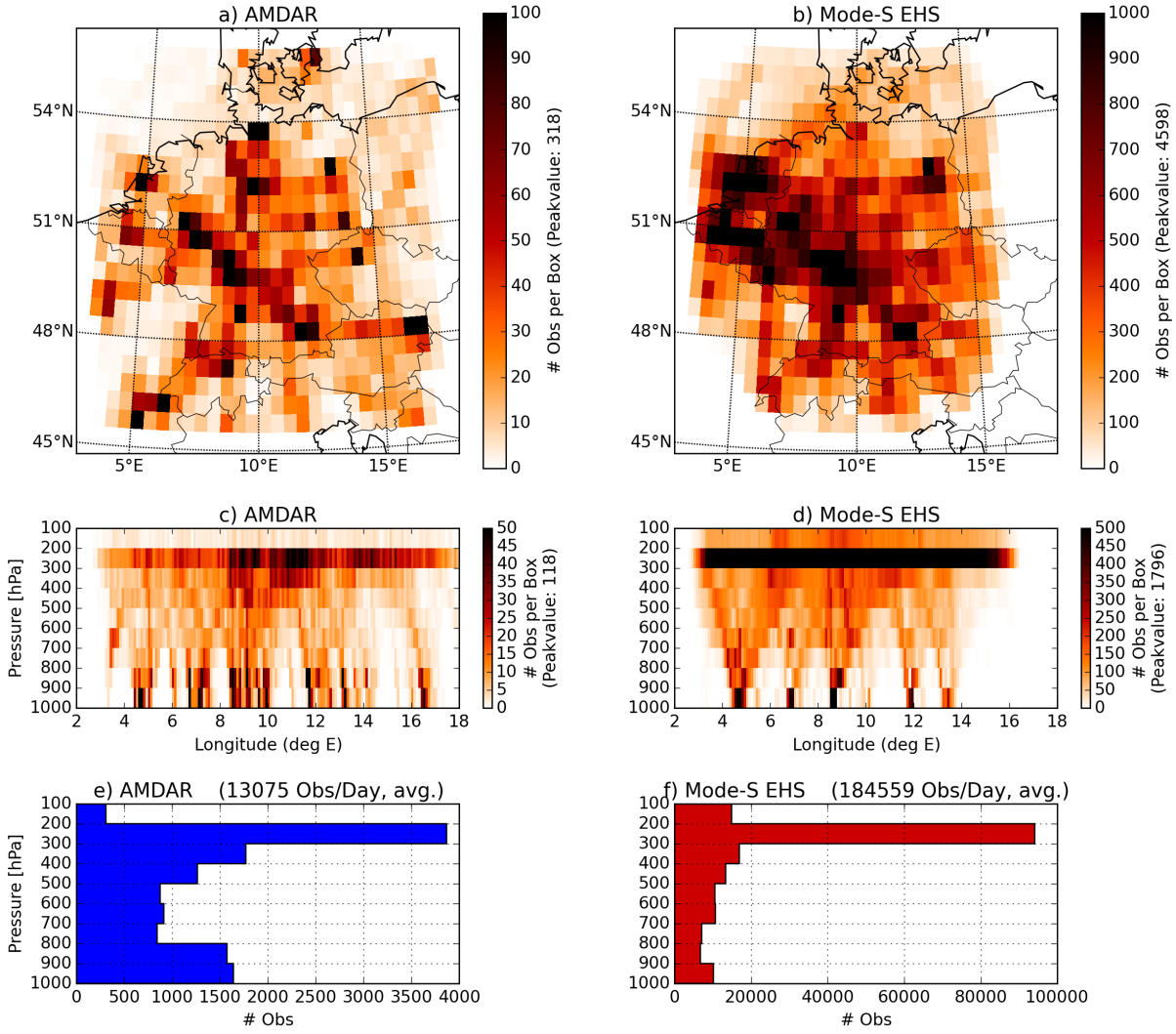

Figure 4.2: Panels a) and b) show the horizontal density of assimilated aircraft observations of AMDAR and the full Mode-S EHS set per average day. The color scale in b) is enhanced by a factor of 10 with respect to a). Panels c) and d) show the vertical density against longitude. Differences in the low level distributions are due to highly frequented airports and are overly exaggerated by the color scales of c) and d). Panels e) and f) show the overall vertical density.

minutes. Hourly assimilated surface data from SYNOP stations are located only in such in land areas with a surface height below 100 m ASL and on the sea surface, due to a rigorous quality control (Schraff et al., 2016). The amount of aircraft measurements of wind and temperature collected in the AMDAR format are illustrated in Fig. 4.2. While the bulk number of single observations in AMDAR is similar to the profiler category, the horizontal distribution is broader.

de Haan (2011) has first presented Mode-S EHS aircraft data as a meteorological observation source. In contrast to the actively submitted AMDAR reports from specially equipped planes, flight control messages of Mode-S EHS are collected by airport radars from all commercial aircrafts within a radius of 270 km every 4-20 s (depending on the purpose of the scanning radars) in the region of Mode-S EHS over Europe and are distributed in a postprocessed quality by the Royal Netherlands Meteorological Institute (KNMI)[2]. Therefore, this experimental data set is limited to the range of the collecting airport radars (Fig. 4.2b). The horizontal extent in the COSMO-DE domain is therefore slightly smaller than for AMDAR, but the number of observations is roughly 15 times larger (Fig. 4.2ef). The single Mode-S EHS report messages contain information on latitude/longitude, flight level, ground speed, airspeed, track angle, heading angle and Mach number. As with AMDAR, the horizontal wind observations are derived from the vectors of airspeed with heading angle and ground speed with track angle. The reported Mach number has been computed by the board computer of the aircraft from the in-situ temperature and airspeed. Reverting this calculation using the ideal gas law, the observed temperature is regained and bias-corrected (see de Haan, 2011; de Haan and Stoffelen, 2012, for details), whereas AMDAR reports contain the direct temperature measurement from the airplane thermometer. Relative humidity is not contained in the Mode-S EHS reports. The number of AMDAR reports which contain relative humidity are negliglibe in the current setup.

de Haan (2011) further describes the post-processing that is applied to the Mode-S EHS data, e.g. exclusion of observations with a large roll angle, magnetic heading offset corrections and bias corrections for different aircraft types. In collocated comparisons with AMDAR-data of the same airplane de Haan (2011) finds observation error standard deviations of comparable magnitude for wind and a slightly larger magnitude for the temperature in case of Mode-S EHS.

Along the flight-tracks of single planes, the raw Mode-S EHS observations have been averaged in order to meet the criteria of AMDAR (WMO, 2003; de Haan and Stoffelen, 2012): During the ascending flight phase, the first nine observations are vertically 10 hPa apart and after that they are 50 hPa apart. On the flight level, consecutive observations are 60 s apart, resulting in a distance of approximately 15 km for a typical flight speed of 250 m s^{-1}. In the descending flight phase, observations with $p < 900$ hPa are 50 hPa apart, for $p > 900$ hPa they are 10 hPa apart. The resulting horizontal distribution of Mode-S EHS within the observed region (Fig. 4.2b) is more homogenous than AMDAR (Fig. 4.2a), while the latter also contains profiles from airports outside of the Mode-S EHS coverage (Fig. 4.2cd). In both AMDAR and Mode-S EHS, the cruising flight level

[2]http://mode-s.knmi.nl

between 200 hPa and 300 hPa is relatively well-observed. In middle levels, patterns of ascending and descending aircrafts become visible which converge towards the locations of the airports, leaving horizontal data gaps in the low levels between 700 and 1000 hPa (Fig. 4.2). The ratio of flight level data compared to data at p > 300 hPa is larger for Mode-S EHS than for AMDAR (Fig. 4.2ef). This is because Mode-S EHS also observes a lot of aircrafts which are passing over the COSMO-DE model domain in flight level without ascent or descent. The number of available observations at different times of day depends on the flight activity which is highest during the daytime between 6 UTC and 21 UTC and lowest at night (de Haan and Stoffelen, 2012).

4.3 Data Assimilation with COSMO-KENDA

COSMO-DE (Baldauf et al., 2011) is a non-hydrostatic model, which is run operationally at DWD with a model domain of 461 by 421 grid points horizontally, covering Germany and parts of neighbouring countries (Fig. 4.1). It has a 2.8 km horizontal grid resolution and 50 vertical layers. The data assimilation framework of KENDA has been developed at the DWD for operational use (Schraff et al., 2016). It is an ensemble based data assimilation system that uses the LETKF algorithm (Hunt et al., 2007). This section describes the settings of the KENDA system for the assimilation experiments with COSMO-DE of this study.

4.3.1 Local Ensemble Transform Kalman Filter

COSMO-KENDA uses a 40 member ensemble to sample the background error covariance, whose each member i with state $\mathbf{x}_i^b$ consists of the prognostic variables of wind, vertical velocity, temperature, pressure perturbation, specific humidity, and cloud water and ice. The prognostic variables of turbulent kinetic energy, rain, snow and graupel are excluded from the analysis update (Schraff et al., 2016). In this study, an analysis ensemble $\mathbf{x}_i^a$ is computed every full hour using the observations $\mathbf{y}^o$ from -1 h to 0 h of the analysis time. Observations closer than 0.67 degree to the lateral boundaries are excluded so that the analysis update is only performed in the inner domain. All observations above 100 hPa are excluded and no analysis update is performed above 100 hPa in order to keep the state close to the model top (20 km, approximately 40 hPa) unchanged between background and analysis.

The LETKF calculations are done on a coarse analysis grid and the analysis weights are interpolated onto the model grid before the ensemble forecasts are started (see Yang et al., 2009, for details). The coarse analysis grid is horizontally reduced by a factor of 3 with respect to the COSMO-DE model, resulting in a spacing of roughly 8 km (COSMO-DE: 2.8 km), and it has 30 vertical levels (COSMO-DE: 50 levels).

The boundary conditions (BCs) are provided by an LETKF-ensemble of the ICON (Icosahedral non-hydrostatic) general circulation model (Zängl et al., 2014). This global LETKF system uses a 3 hourly cycle in which observations from surface stations, radiosondes, aircrafts, and satellites are assimilated (Fernandez et al., 2015; Schraff et al.,

2016), using a localization radius of 1000 km. The ICON forecasts between the analyses of 0, 3, 6, ... UTC are saved every hour and provide the BCs for the COSMO-DE-ensemble. Since the BCs of 0, 3, 6, ... UTC are represented by global analyses, they contain information also from observations inside the COSMO-DE domain. Due to the lack of longer global forecasts, the local ensemble forecasts of this study have been limited to the 3 hours of leadtime between the global analyses to study only the internal error evolution within COSMO-DE.

4.3.2 Observation error statistics

The computation of analyses in the LETKF depends on a proper specification of the observation error statistics. Following Desroziers et al. (2005) and Li et al. (2009), the observation error covariance $\mathbf{R}$ can be estimated from the differences between observations and background in the observation space $\mathbf{d}_{o-b} = \mathbf{y}^o - H(\mathbf{x}^b)$ and between observations and analysis in observation space $\mathbf{d}_{o-a} = \mathbf{y}^o - H(\mathbf{x}^a)$ by using $\mathbf{R}_{\text{est}} = E[\mathbf{d}_{o-a}{\mathbf{d}_{o-b}}^T]$, where E denotes the expectation operator, H is the observation operator, and $\mathbf{x}^a$, $\mathbf{x}^b$ are the ensemble means of analysis and background, respectively. Before the estimation, the means of $\mathbf{d}_{o-b}$ and $\mathbf{d}_{o-a}$ are subtracted to make the estimation unbiased. In COSMO-KENDA, only the diagonal of the estimated $\mathbf{R}_{\text{est}}$ is used for the observation error covariance matrix. The cross covariances between different observations are neglected. The variance is estimated offline after the analysis cycling, using the bulk $\mathbf{d}_{o-b}$ and $\mathbf{d}_{o-a}$ of the five day period. The observation error standard deviation of the wind observations U and V is estimated from the windspeed $\sqrt{U^2 + V^2}$. $\mathbf{R}_{\text{est}}(p)$ is evaluated for 100 hPa tall bins of pressure height p. For SYNOP, PROF, and AMDAR, the estimates of the observational error statistics have been made in a previous study by Schraff et al. (2016) using the same approach and are denoted as $\mathbf{R}_0$.

The estimates obtained for Mode-S EHS should indicate the quality of the data set compared to AMDAR. It will be tested if the estimated observation error standard deviations $\sqrt{\mathbf{R}_{\text{est}}}$ are the same or similar to the previously estimated values $\sqrt{\mathbf{R}_0}$ for AMDAR. Note however that this technique is based on the innovation statistics of the DA system and can not replace the external comparison already done in de Haan (2011) for the estimation of the measurement error for Mode-S EHS and AMDAR.

4.3.3 Methods of accounting for model and sampling error

The background error covariance derived from the ensemble does not adequately represent the uncertainty of the background field due to sampling and model error. In KENDA, the methods of background covariance inflation, adaptive covariance localization, relaxation-to-prior-perturbation (RTPP) and random surface perturbations are implemented. These methods are briefly introduced here and their effects are presented in Section 4.6.

The RTPP scheme of Zhang et al. (2004) relaxes the analysis ensemble perturbations of the LETKF towards the background ensemble perturbations at every grid point of the model. In KENDA, the relaxation is performed with a factor of $\alpha_{\text{r}} = 0.75$, thus limiting the maximum influence of the analysis to 25% of the maximum change. This helps to

keep the analysis ensemble spread alive by acting as an inflation with both multiplicative and additive aspects, albeit in the space spanned by the ensemble only (Zhang et al., 2004; Whitaker and Hamill, 2012). Harnisch and Keil (2015) showed its beneficial effects on ensemble forecasts with the COSMO-KENDA system.

Following Li et al. (2009), an adaptive multiplicative covariance inflation with factor ρ is applied separately for every coarse analysis grid point (Schraff et al., 2016) and relaxed in time by mixing the current ρ estimate with the estimate of the previous cycle. ρ is larger than 1 where the background ensemble spread in observation space $Tr(\mathbf{HP}^b\mathbf{H})$ is small in proportion to how much the magnitude of $\mathbf{d}_{o-b}$ deviates from the specified observation error $\mathbf{R}$, expressed by the difference $[{\mathbf{d}_{o-b}}^T\mathbf{d}_{o-b} - Tr(\mathbf{R})]$.

In the present KENDA setup based on Schraff et al. (2016), the horizontal localization is adaptive to the number of observations in the vicinity of an analysis grid point (Perianez et al., 2014) so that the spatial influence is distributed to the appropriate scales (Janjić et al., 2011, 2012). The horizontal localization length scale l_{loc} is defined here as the radius where the Gaussian-like covariance function (Eq. 5.1 from Gaspari and Cohn, 1999) goes down to $\exp(-0.5)$. With lower and upper bounds of 50 and 100 km, l_{loc} is adaptively determined to contain 100 observations which is 2.5 times the ensemble size of 40. The vertical localization length scale is fixed and ranges from 0.075 $\ln(p)$ at the surface to 0.5 $\ln(p)$ at the model top.

As described by Schraff et al. (2016), the sea surface temperature (SST) and the soil moisture are randomly perturbed in every analysis by adding temporally and spatially correlated perturbation fields in order to enhance the spread near the model surface and in the boundary layer.

4.4 Experiments with varying aircraft datasets

The experiments of this study were performed from 7th till 12th of May 2014. The meteorological situation was a west-wind period with passing cyclones. A warm front passage on 8th of May caused stratiform precipitation and a cold front passage on 10th of May caused predominantly convective precipitation with embedded thunderstorms and showers. The period was chosen because synoptically and locally forced convection was present and a convection permitting model such as COSMO-DE is expected to produce more realistic weather systems than its driving global model.

The initial COSMO-DE ensemble, consisting of the downscaled global ICON-LETKF ensemble, is started at 0 UTC on the 6th of May. One day of cycled data assimilation is performed to spin up COSMO-DE and the time-relaxed adaptive covariance inflation. The resulting analysis ensemble and the covariance inflation field ρ at 0 UTC on the 7th of May are then used as initial conditions for all other experiments.

Table 4.1 summarizes the performed experiments for the chosen test period. The *monitored observations* (mon.) are not assimilated, therefore their model equivalents are only influenced by the cross correlations calculated by the LETKF by the means of the *assimilated observations* (denoted by •). The control experiment NoDA is performed without data assimilation. All DA experiments assimilate the selected data sets in an

	Mode-S EHS	AMDAR	SYNOP	PROF	TEMP
NoDA	mon.	mon.	mon.	mon.	mon.
Aconv	mon.	•	•	•	mon.
Mconv	•	mon.	•	•	mon.
MAconv	•	•	•	•	mon.
MAconvTh10	• 10%	•	•	•	mon.
MAconvTh50	• 50%	•	•	•	mon.

Table 4.1: Experiments (rows) performed with COSMO-KENDA. Observations with bullet points • are assimilated, observations with 'mon.' are *monitored*. In case of '• x %', only a random subset of x percent is assimilated.

hourly interval. The basic experiment Aconv assimilates conventional (SYNOP, PROF) and AMDAR observations. From Aconv, the Mconv experiment is derived by setting AMDAR to monitoring and assimilating Mode-S EHS. The experiment MAconv assimilates both AMDAR and Mode-S EHS data sets. In MAconvTh10 and MAconvTh50 the Mode-S EHS data is randomly thinned to 10 percent and 50 percent of the full density that is used in MAconv. The random thinning makes sure that the spatial and temporal distribution characteristics of the Mode-S EHS observation space are preserved, only with a lower density. The observation error standart deviation $\sqrt{\mathbf{R}_0}$ which was preliminarily estimated for AMDAR is used identically for the assimilation of Mode-S EHS. The observation operator is applied during the model integration of COSMO at the appropriate time of the respective observation. Therefore the $\mathbf{d}_{o-b}$ of a 1-hour forecast is a collection of innovations along the temporal trajectory. In case of the 3 hour forecasts, $\mathbf{d}_{o-b}$ refers to the observations within the third hour of the forecast window.

The experimental results are evaluated separately in the observation spaces of each upper air system (TEMP, AMDAR, Mode-S EHS) for wind speed, direction and temperature. Where the results of the observation systems are qualitatively similar, only the radiosonde space is shown. The Root Mean Square (RMS), the mean and the standard deviation (STD) of $\mathbf{d}_{o-b}$ are computed at every forecast step for every ensemble member and are averaged over the 5 day period. To test whether the ensemble is underdispersive, for every pressure level the consistency ratio (CR) is computed which also takes the observation error variance $\mathbf{R}$ into account, as in Schraff et al. (2016):

$$\mathrm{CR} = \frac{\mathrm{STD}}{\sqrt{\mathrm{RMS}^2 - Tr(\mathbf{R})n_{\mathrm{obs}}^{-1}}} \tag{4.1}$$

The CR should ideally be one. For $\mathbf{R}$, the previously defined $\mathbf{R}_{\text{est}}$ was used posterior to the experiments.

The profiles of RMS and mean of $\mathbf{d}_{o-b}$ are tested for significant differences among the experiments. Using the Student's t-test, the nullhypothesis that the results were identical shall be rejected on a confidence level of 95 percent. Since the experiments share the same model and boundary conditions and have a mutual subset of assimilated observations, the forecast errors are correlated among the experiments. Therefore, the t-test is performed as a paired difference test in which the number of degrees of freedom n is reduced by half. In case of the 1-hourly cycling, the test is performed using the hourly profiles of RMS and mean of $\mathbf{d}_{o-b}$ of the single members as the testing sample with $n = (5 \text{ days} \cdot 24 \text{ cycles} \cdot 40 \text{ members }) \cdot 2^{-1} = 2400$. In case of the 3 hour forecasts, two ensemble sets of profiles per day are tested between the experiments with $n = (5 \cdot 2 \cdot 40) \cdot 2^{-1} = 200$.

4.5 Results

The RMS, mean and STD of $\mathbf{d}_{o-b}$ in the control experiment NoDA are shown in Figs. 4.3, 4.4, and 4.5, respectively. The wind speed STD of NoDA is largest at 300 hPa height (Fig. 4.5adg), the largest STD in wind direction is close to the surface (Fig. 4.5beh) and the largest temperature STD in NoDA is at 250 hPa (Fig. 4.5cfi). In the radiosonde space, a warm bias of the background exists in NoDA above 400 hPa and a cold bias around 700 hPa (Fig. 4.4c).

As a first diagnostic for the observation error, the RMS of NoDA is examined, under the assumption that the model equivalents in the upper air observation spaces of all three systems are subject to the same forecast error characteristics. For wind speed, the observation systems TEMP, AMDAR and Mode-S EHS behave similarly (Fig. 4.3adg) which hints at comparable observation errors. For temperature (Fig. 4.3cfi), the NoDA RMS of Mode-S EHS in lower levels (Fig. 4.3i) is larger than for the other systems. For wind direction (Fig. 4.3beh), the NoDA RMS of Mode-S EHS is much smaller. This difference in directional observation error could be a consequence of the heading correction in the post-processing of Mode-S EHS (de Haan, 2011) which is calibrated for different commercial aircraft models. According to WMO (2003), such a correction is not performed for AMDAR.

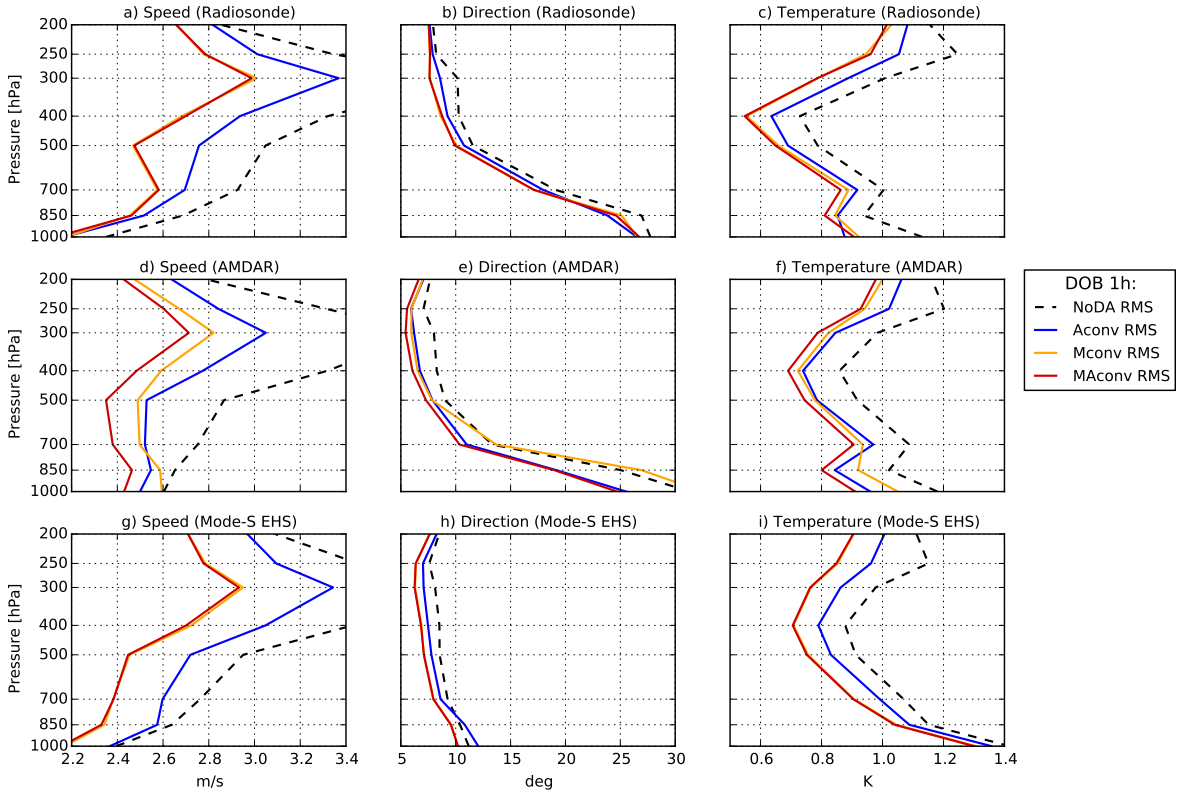

Figure 4.3: Mean RMS ($\mathbf{d}_{o-b}$) of the ensemble members for upper air observations of wind speed (left), wind direction (middle) and temperature (right) throughout the 1-hourly cycling, for experiments NoDA, Aconv, Mconv and MAconv.

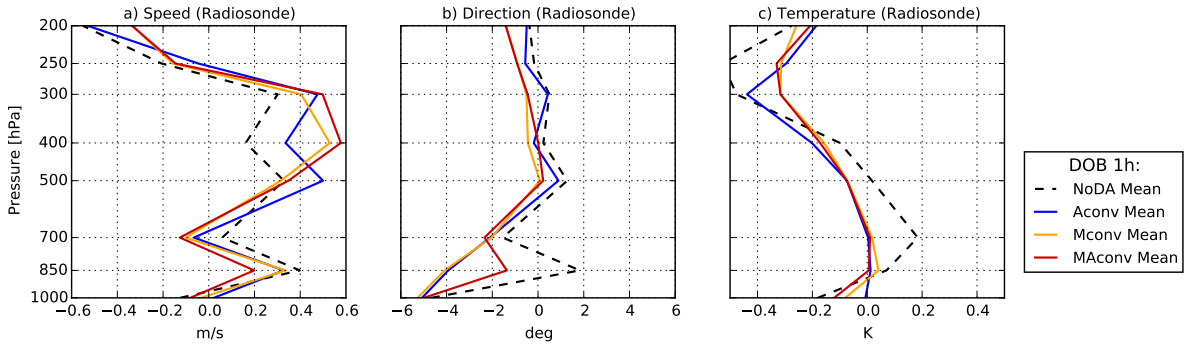

Figure 4.4: Mean ($\mathbf{d}_{o-b}$) of the ensemble members for upper air observations of wind speed (left), wind direction (middle) and temperature (right) throughout the 1-hourly cycling, for experiments NoDA, Aconv, Mconv and MAconv.

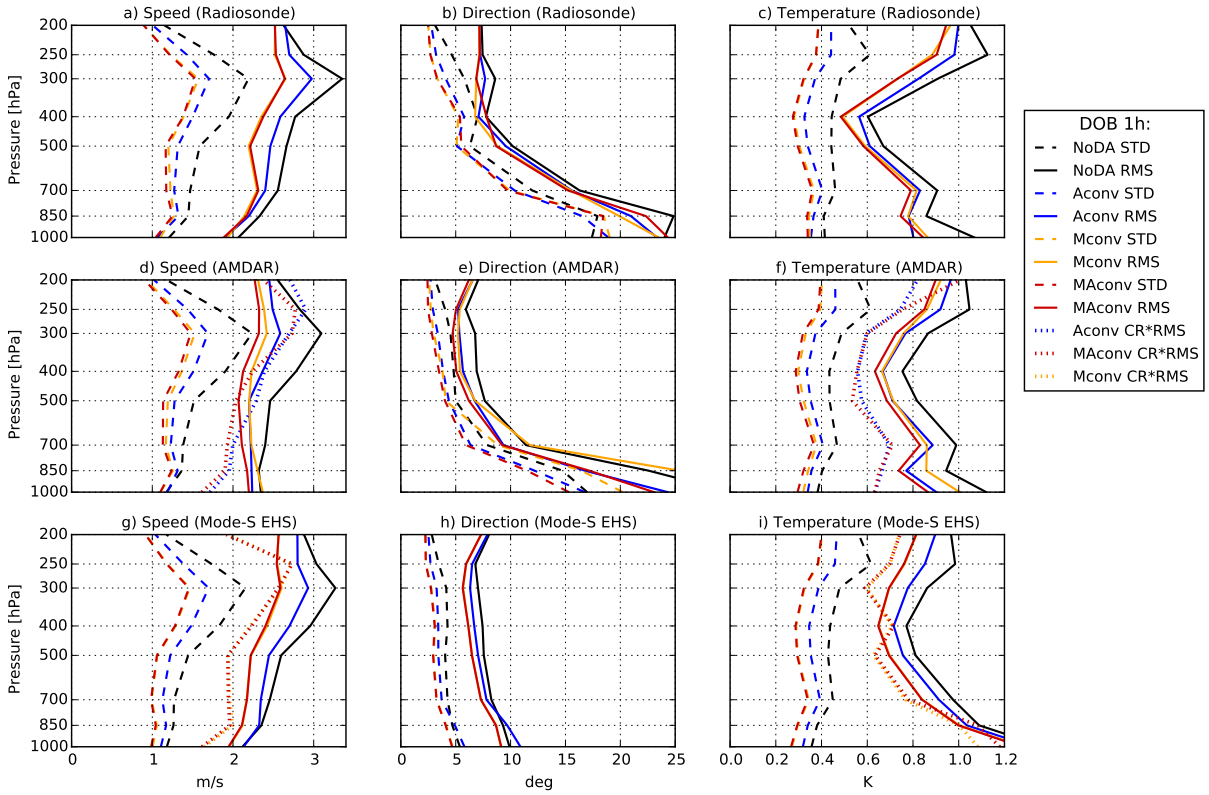

Figure 4.5: STD of the ensemble $\mathbf{d}_{o-b}(\mathbf{x}_i^b)$ and RMS of the ensemble mean $\mathbf{d}_{o-b}(\overline{\mathbf{x}^b})$ for upper air observations of wind speed (left), wind direction (middle) and temperature (right) throughout the 1-hourly cycling, for experiments NoDA, Aconv, Mconv and MAconv. Where data is actively assimilated, the consistency ratio CR is computed using the estimated values of Fig. 4.9, and is scaled by the respective RMS to fit into the x-axes of the plots.

4.5.1 Results with AMDAR and Mode-S EHS observations

The main experiments Aconv, Mconv and MAconv compare the usage of AMDAR and Mode-S EHS as the only assimilated aircraft observation types (Aconv and Mconv) and the combination of assimilating both (MAconv). In Fig. 4.3, the significance symbols are omitted because where the RMSs of DOB are reduced by adding additional data, the reductions are significant in all cases.

The RMS in the radiosonde space is continuously reduced with respect to NoDA for wind speed and temperature (Fig. 4.3ac) and less so for wind direction (Fig. 4.3b) when AMDAR and additionally Mode-S EHS data are assimilated in the experiments Aconv and MAconv. Also, the STD is continuously diminished (Fig. 4.5abc).

The RMS values of Mconv in AMDAR space and of Aconv in Mode-S EHS space show the influence of the data sets across the two aircraft observation spaces. The RMSs of wind speed (Fig. 4.3dg) and temperature (Fig. 4.3fi) are reduced across the spaces for Aconv and Mconv, while the cross-space RMSs in wind direction (Fig. 4.3eh) are increased for levels below 700 hPa. This might again be a consequence of the heading correction of Mode-S EHS.

In Fig. 4.5, the ratio of the RMS and the STD give information about the consistency of background error and spread, assuming that the biases in DOB are small enough. The CRs of windspeed and temperature are close to 0.9 or 1 for all experiments, except close to the surface where values around 0.75 are reached.

4.5.2 Results with varying Mode-S EHS data coverage

As there are far more Mode-S EHS data than AMDAR data, it is tested how the forecasts depend on the amount of Mode-S EHS observations. In the experiments MAconvTh50 and MAconvTh10, the assimilated portion of the Mode-S EHS observations is randomly thinned to roughly 50 and 10 percent of the original set.

During the cycling, the wind speed RMS in the radiosondes (Fig. 4.6a) is significantly reduced going from 10 percent to 50 percent for all levels except 700 and 1000 hPa, while an increase from 50 percent to 100 percent does not yield benefits. For the wind direction (Fig. 4.6b), there is no significant RMS reduction when increasing the data amount. For temperature (Fig. 4.6c), the RMS is decreased going from 10 percent to 50 percent and from 10 to 100 percent for all levels except the surface. At 400 and 300 hPa, an improvement is also visible when going from 50 percent to 100 percent of Mode-S EHS. The wind and temperature STD in the radiosonde space is continuously reduced when adding more data (Fig. 4.7, AMDAR and Mode-S are analogous and not shown), but not in a linear dependence on the data amount.

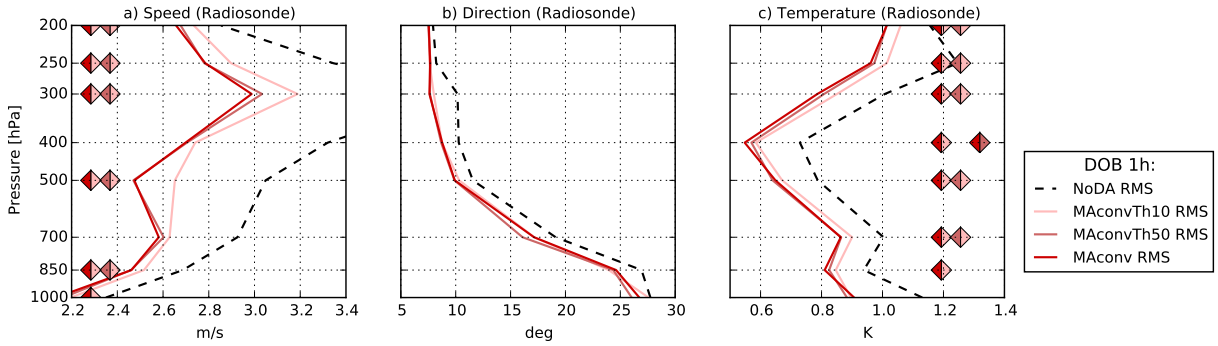

Figure 4.6: As Fig. 4.3, for experiments NoDA, MAconvTh10, MAconvTh50, MAconv. A two-colored diamond on a level means that the RMSs of the left experiment are significantly smaller than of the right experiment.

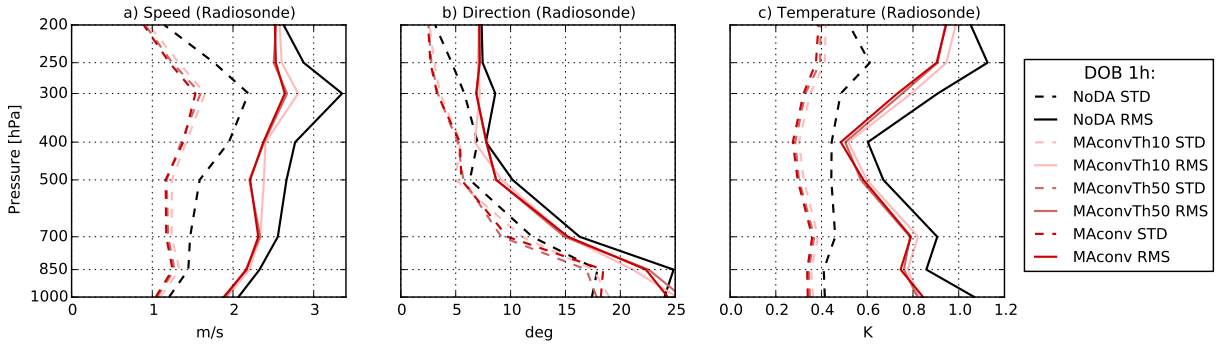

Figure 4.7: As Fig. 4.5, for experiments NoDA, MAconvTh10, MAconvTh50, MAconv.

4.5.3 Results of the 3 hour ensemble forecasts

The daily analysis ensembles at 9 and 21 UTC were saved and ensemble forecasts of three hours were initialized from all experiments, so that also the radiosondes launched at 12 and 0 UTC could be used for the verification. The forecast RMS of $\mathbf{d}_{o-b}$ (third hour of 3 h window) is shown in Fig. 4.8.

For wind speed, significant reductions in forecast RMS between Aconv and MAconv take place at all levels except the surface. The clearest reduction is visible when using Mode-S EHS for the verification (Fig. 4.8g). The radiosonde space (Fig. 4.8a) is positively affected mainly in levels above 500 hPa. For AMDAR, additional Mode-S EHS data of MAconvTh10 and MAconvTh50 is beneficial at 700 hPa (Fig. 4.8d), which could be due to the shared observation spaces of the aircraft systems (Fig. 4.2). For wind direction RMS, the radiosonde space verification shows no significant improvement due to Mode-S EHS (Fig. 4.8b). Still, in AMDAR the wind direction forecast is improved on the flight levels (Fig. 4.8e), and in the Mode-S EHS space also the 700 hPa level is improved (Fig. 4.8h). For temperature RMS, an improvement due to additional Mode-S EHS data is visible for all three systems (Fig. 4.8cfi) from 700 to 200 hPa.

The 3-hour forecast RMSs are, in some cases, improved by increasing the amount of assimilated Mode-S EHS. The wind speed RMS of MAconv is smaller than both Aconv and MAconvTh10 at 300 hPa (Fig. 4.8a), 500 and 700 hPa (Fig. 4.8g). The 700 hPa wind direction is improved within Mode-S EHS going from MAconvTh10 to MAconv (Fig. 4.8h). For temperature, an improvement from MAconvTh10 to MAconv is visible from 700 to 200 hPa (Fig. 4.8cfi). An improvement from MAconvTh50 to MAconv appears only in the 300 hPa radiosonde wind speed (Fig. 4.8a). A deterioration of forecast RMS at 500 hPa is apparent in some cases when increasing the data amount from MAconvTh50 to MAconv (Fig. 4.8dec).

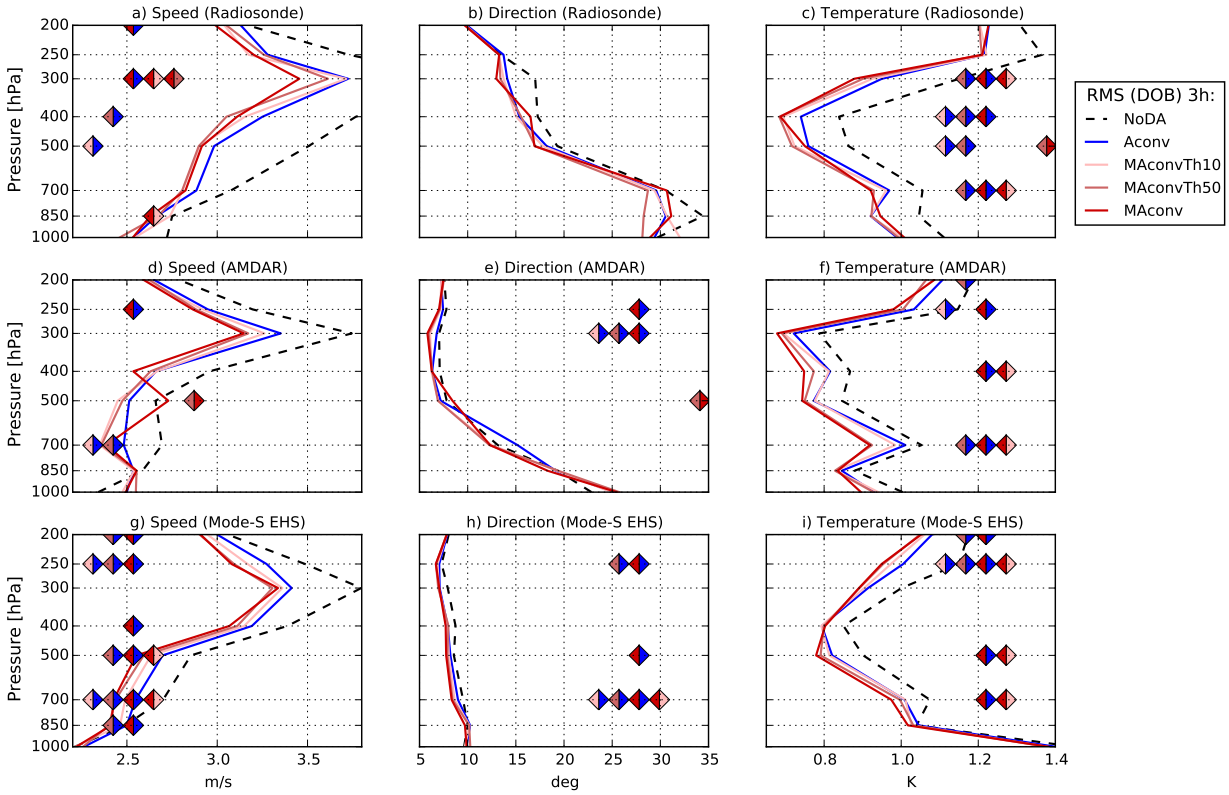

Figure 4.8: Forecast RMS of the Mode-S EHS thinning experiments and Aconv, evaluated for the last hour of the 3-hour forecast windows from 9 to 12 UTC and 21 to 0 UTC. A two-colored diamond on a level means that the RMSs of the left experiment are significantly smaller than of the right experiment. At levels where the RMS of an experiment is smaller than the RMS of NoDA, the difference is always significant and the diamonds are not plotted.

4.6 Effects of adaptive KENDA settings

The settings used in KENDA affect the representations of both background and analysis error covariances. The specified values of the observation error are discussed first in Section 4.6.1. Secondly, the influence of the methods that account for model and sampling error in KENDA are examined in Section 4.6.2.

4.6.1 Results of the observation error estimation

The estimated error standard deviation $\sqrt{\mathbf{R}_{\text{est}}}$ of AMDAR and Mode-S EHS are depicted in Fig. 4.9 together with the initial error estimate $\sqrt{\mathbf{R}_0}$ of Schraff et al. (2016) that was used for the assimilation. $\sqrt{\mathbf{R}_{\text{est}}}$ for windspeed in AMDAR is lower than $\sqrt{\mathbf{R}_0}$ (Fig. 4.9a), which can be attributed to a different experimental period and a difference in the driving global ensemble.

The values are similar among the experiments that actively assimilate AMDAR (Aconv, MAconvTh10, MAconv). A lower estimate is reached (Fig. 4.9a) when Mode-S EHS observations are used which reduce the estimators $\mathbf{d}_{o-b}$ and $\mathbf{d}_{o-a}$ also across the AMDAR observation space. $\sqrt{\mathbf{R}_{\text{est}}}$ for temperature in AMDAR (Fig. 4.9b) is close to $\sqrt{\mathbf{R}_0}$ between 700 and 400 hPa and much smaller in the lower levels.

For the windspeed of Mode-S EHS (Fig. 4.9c), $\sqrt{\mathbf{R}_{\text{est}}}$ is, within a range of 10 percent difference, quantitatively and qualitatively similar to the windspeed of AMDAR (Fig. 4.9a) in case of the joint data experiment MAconv. For the Mode-S EHS temperature (Fig. 4.9d), $\sqrt{\mathbf{R}_{\text{est}}}$ follows the shape of $\sqrt{\mathbf{R}_0}$ on all levels and is larger for levels below 700 hPa where it reaches 1.2 K for Mode-S EHS in MAconv, but only 0.8 K for AMDAR in MAconv (Fig. 4.9b). This difference in low level temperature errors is qualitatively consistent with the findings of de Haan (2011) and de Haan and Stoffelen (2012), although they diagnosed larger absolute error standard deviations.

The hypothesis that a feedback $\mathbf{R}$-input of the values $\sqrt{\mathbf{R}_{\text{est}}}$ estimated by MAconv (Fig. 4.9) would give better results during the cycling was tested an iterated experiment MAconvR1 (not shown) which did not exhibit significant improvements with respect to MAconv.

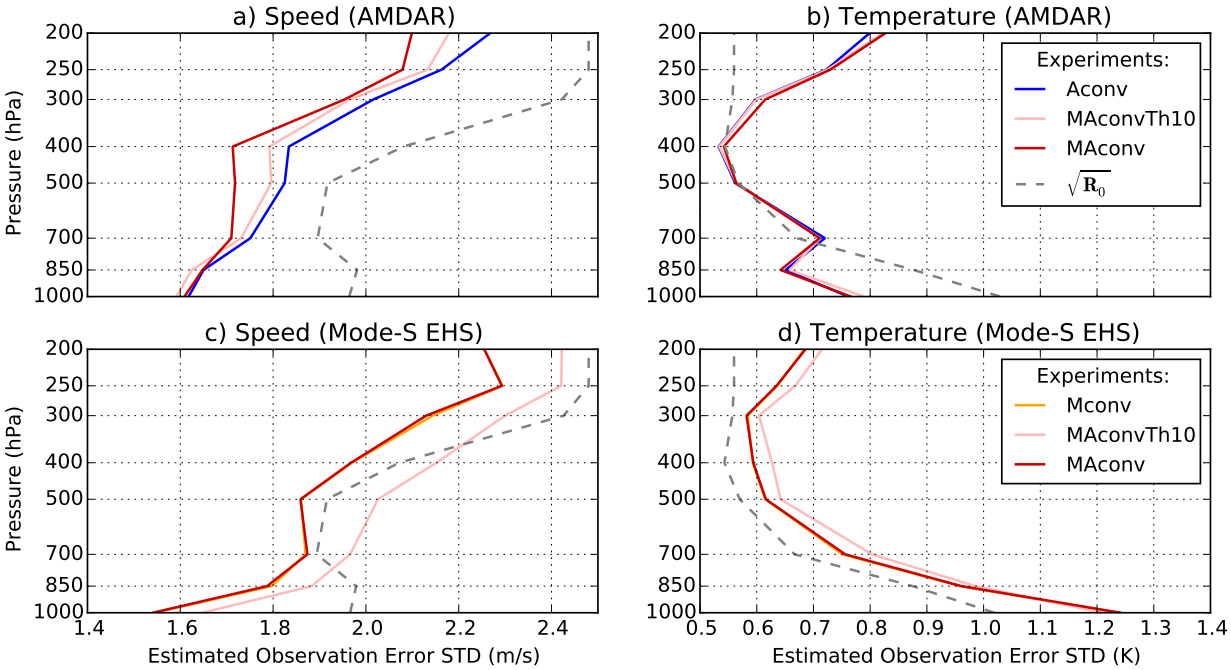

Figure 4.9: Estimated observation error STD $\sqrt{\mathbf{R}_{\text{est}}}$ for AMDAR and Mode-S EHS observation spaces. $\sqrt{\mathbf{R}_0}$ is the initial error standard deviation value used as input for the LETKF for both AMDAR and Mode-S EHS for all experiments.

4.6.2 Accounting for model and sampling error

Fig. 4.10a shows the adaptive covariance inflation factor. It is almost continuously larger for experiments with more observations. One maximum lies between 300 and 200 hPa where the high data density of the flight level is situated (Fig. 4.2). Another maximum is between 700 hPa and the surface. This is caused by a spread deficiency in the lower levels (see CR in Fig. 4.5). While the adaptive inflation is set to counteract this deficiency, the RTPP factor of 0.75 limits the overall reduction and inflation of spread by the analysis update. As seen in Fig. 4.10b, the adaptive localization directly depends on the number of observations per level. In the horizontal snapshot of Fig. 4.10c, highly frequented airports (Amsterdam, Frankfurt, Berlin) become visible.

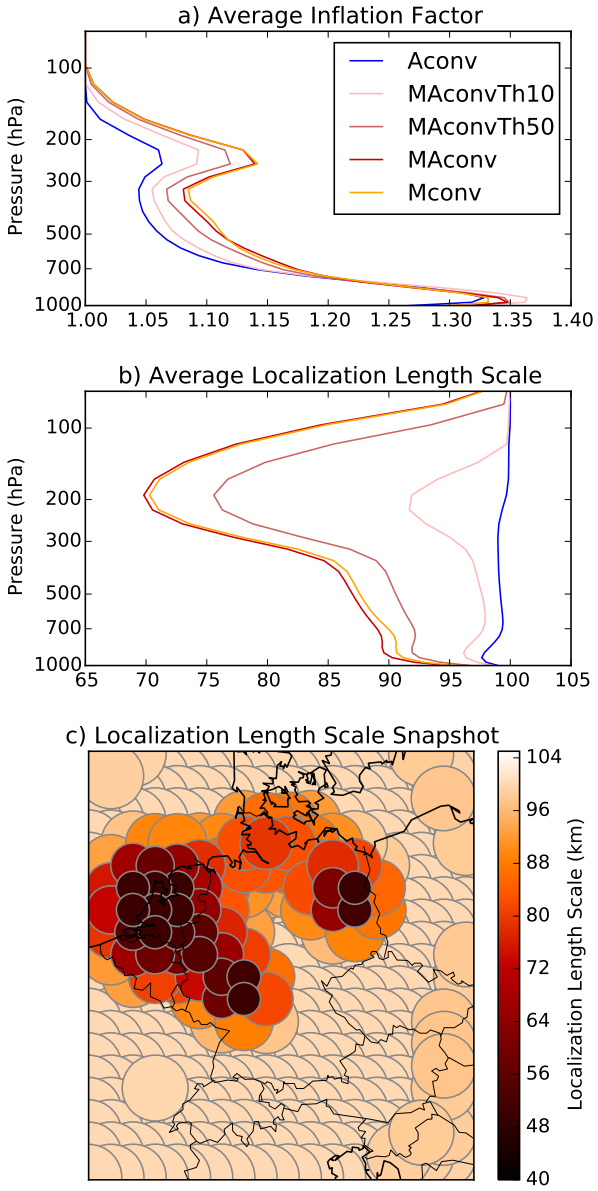

Figure 4.10: a) Adaptive covariance inflation factor, averaged over the experimental period on the pressure levels of the analysis grid. b) Adaptive localization length scale. c) Snapshot of the localization length scale at the surface level at 8th of May 2014, 12 UTC. The plotted circles are centered on every 8th point of the coarse analysis grid, so the overlap of the local LETKF solutions is denser than in this illustration.

4.7 Kinetic energy spectra of three hour forecasts

One-dimensional kinetic energy spectra are derived from the analysis and forecast ensemble members as in Bierdel et al. (2012). Due to the predominant westerly flow, isotropy is not assumed and the energy spectra are calculated separately in longitudinal and latitudinal directions on levels with a constant height ASL. Since on the flight levels there are significant amounts of AMDAR and Mode-S EHS observations, an average over three heights of 10, 11, and 12 km height ASL is chosen for the investigation. The wind field of the ensemble was saved every 15 minutes throughout the previously mentioned 3 hour forecasts and the corresponding spectra were calculated. Fig. 4.11 shows the difference kinetic energy spectrum of Aconv, MAconvTh10, MAconvTh50, and MAconv with respect to the NoDA experiment. In order to enhance the relative differences, the energy of the difference wind field from the NoDA experiment is scaled by the absolute kinetic energy spectrum of NoDA. The evaluation presumes that during the three forecast hours the BCs are not dominant and the main part of the dynamical evolution is happening within the COSMO-DE domain itself.

Starting from the analyses (forecast minute 0), the largest relative differences are present on the scales below the effective resolution of $5\Delta x$ found for COSMO-DE (Bierdel et al., 2012). Between the wavelengths of 10 and 40 km, there is a level of relatively large difference energy, and above 40 km there is an almost steady drop. Among the series of Fig. 4.11a-d, the analysis states are subsequently further away from NoDA when more data is assimilated. With the ten percent of Mode-S EHS data in MAconvTh10, the characteristics are similar to Aconv. For MAconvTh50, a significant difference to Aconv becomes visible on all scales.

Throughout the three forecast hours, the relative difference energy on all scales is gradually decreasing, where a value of zero would express a wind field identical to NoDA. In all experiments, the spectrum of the relative forecast difference energy remains similar to the analysis of the respective experiment for up to one hour, especially on scales smaller than 200 km. After one hour, a drop in relative difference energy with respect to the analysis can be noted on all scales. As a consequence of pulling the states further away from the attractor of COSMO-DE, the rate of convergence towards NoDA is higher when more data is assimilated. Although the experiments MAconvTh50 and MAconv are almost indistinguishable after three hours of forecast with respect to the forecast RMS (Fig. 4.8), the higher level of the relative spectra at 180 minutes in MAconv still indicates differences in the wind field. It can be concluded that the usage of 100 percent Mode-S data causes excessive modifications, possibly in the form of gravity wave noise (Houtekamer and Mitchell, 2005), which are more extensive than necessary to actually improve the forecast.

4.8 Summary and Discussion

In order to evaluate the benefit of additional aircraft in-situ observations of wind and temperature that are collected by airport radars (Mode-S EHS), experiments were per-

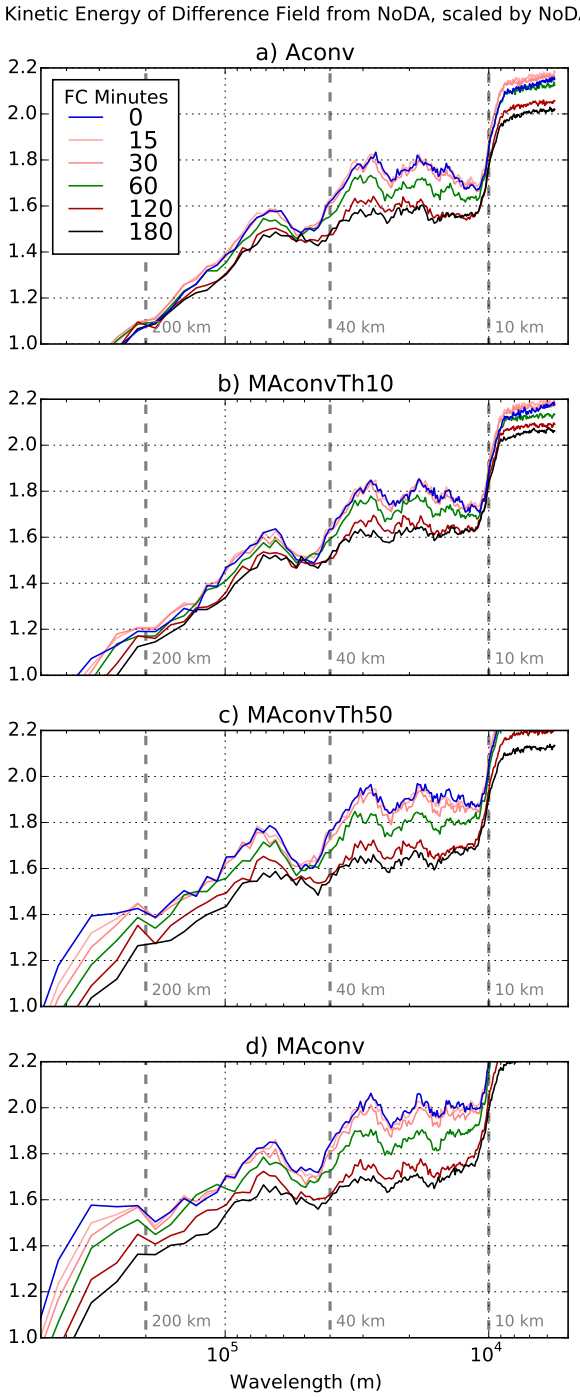

Figure 4.11: Kinetic energy of the difference wind fields from NoDA (along latitudes between 10 and 12 km ASL), scaled by absolute kinetic energy of the NoDA wind field. The upper air spectra are shown for the experiments Aconv, MAconvTh10, MAconvTh50 and MAconv (top to bottom).

formed with the COSMO-KENDA system, which couples an LETKF to an ensemble of the nonhydrostatic model COSMO-DE. The Mode-S EHS observations are available within most of the domain of the COSMO-DE model over Germany. The amount of Mode-S EHS observations is 15 times larger than of the traditionally used AMDAR, with a similar observation density distribution. Cycled ensemble data assimilation was performed for a period of 5 days with a cycling interval of 1 hour. Mode-S EHS data were assimilated in addition to AMDAR, assuming the same observation error. Using 100 percent of the Mode-S data improved the upper air innovation RMSs during the 1-hour cycling and in the verification of 3 hour ensemble forecasts, in comparison to assimilating only AMDAR aircraft data. A comparable improvement was visible when assimilating 50 percent of Mode-S EHS instead of none. The 3-hour forecasts that were started from the experiment with 100 percent Mode-S EHS data had often significantly lower errors than from the experiment with 10 percent. Still, using 100 percent Mode-S EHS instead of 50 percent degraded the 3-hour forecasts in some cases. It is therefore concluded that a saturation of benefit exists between 50 percent and using 100 percent of Mode-S EHS.

By using innovation statistics in observation space, the observation error standard deviation of Mode-S EHS was estimated to be comparable to AMDAR in the wind variable. For temperature, the observation error of Mode-S EHS was diagnosed to be 50 percent larger for levels below 700 hPa. It must be noted that observation error standard deviations obtained by this method depend on the background ensemble spread of the cycling system, so the estimated observation errors can only be regarded as inherently compatible with the present DA system. Still, the current KENDA setting appeared mostly insensitive to slightly wrong prescribed observation errors.

The ensemble spread was generally consistent with the forecast errors, except near the surface due to a deficiency of small scale error growth (Schraff et al., 2016). KENDA contains mechanisms that maintain the ensemble spread, namely the RTPP scheme, adaptive covariance inflation, and adaptive localization. The effects of these methods were evaluated and it was concluded that they enable the assimilation system to use the larger number of aircraft observations without a collapse of the ensemble spread. The spread reduction exhibited a non-proportional dependence on the increasing number of observations. In the future, unified inflation methods such as presented in Ying and Zhang (2015) could be helpful in further dealing with this issue.

The ability to benefit from a high observation density strongly depends on the RTPP relaxation parameter. With less relaxation towards the background ensemble than the currently used factor of 0.75, the resulting analysis ensemble could contain smaller errors because the observations are given more weight. In future experiments, also a more sophisticated thinning could be applied to reduce the observation density only where it is redundant or where saturation effects are apparent. The high temporal density of Mode-S EHS observations could be exploited in a 15 minute cycling interval. An observation system that fills the low level data gaps between airports could be useful for airport forecasts, especially for warnings of small scale phenomena such as rapidly evolving gustfronts and downbursts. A wider usage of surface stations than in the present study could supplement the high density of upper air observations, if provided with proper vertical localization.

Spectra of the difference kinetic energy with respect to a control experiment showed that the LETKF data assimilation affects all model scales. The amount of observation data in the experiments was reflected by the degree of modification of analysis and forecast states. A stronger modification caused a faster relaxation towards the attractor of the control experiment. In further studies, different regimes (dry/wet) or seasons (winter/summer) could be chosen to compare the influence of dense observation sets when error growth is dominated by either synoptic or local processes. The spectral difference energy of temperature and cloud fields could give insight on scale dependent error growths (Zhang et al., 2007; Selz and Craig, 2015; Zhang et al., 2016). Forecasts with longer lead times than three hours should be performed to evaluate if the Mode-S EHS benefit is persistent, along with a testperiod longer than 5 days to improve the significance of the results. A rigorous survey of the imbalances that depend on the Mode-S EHS assmilation should be performed, in order to identify effects of a possible overfitting and overly small localization lengths (see Chapter 2). In Chapter 5, a version of the present DA-setup is used to additionally assimilate observation sets with even higher resolutions, such as radial winds and reflectivity of convective systems from Doppler radars.

Chapter 5

Influence of Observation Types in Radar Data Assimilation

In Chapter 3 it was investigated how temporal and spatial parameters of radar assimilation influence the dynamical properties of analyses and forecasts in an idealized convective situation. The present Chapter is devoted to the question which influence the observation types of wind and reflectivity have on the dynamical properties of EnKF-produced analyses.

5.1 Introduction and motivation

The question how a DA-analysis reacts to different observation types has been raised by Talagrand (1981) who discusses how the assimilation of different variables influence the dynamics of a Shallow Water Model (SWM) (Gill, 1982). The SWM equations mimic the behavior of global atmospheric dynamics whose rotational components are geostrophically balanced with the surface height field and whose divergent components are expressed by fast-moving inertia-gravity waves of the surface height.

Talagrand (1981) identifies observations of the surface height variable, which expresses the underlying mass, and of the wind variable, which is related to rotational and divergent motions, as the two main observation types. He concludes that by using only mass observations, the first time derivative of divergence is modified, but only the second time derivative of vorticity. Using only wind observations, gravity wave motion is reconstructed directly. Since this reconstruction may introduce noisy increments, geostrophic adjustment to the mass field has to take place, which is a slow process that may be overridden by further increments due to wind observations. For increments of mass observations (as was done already by Charney et al., 1969, using temperature observations), the slow geostrophic adjustment is the only process that may rectify the vorticity. Kalnay (2003) further elaborates on this topic in their Chapter 5.7, and derives an equation that predicts which influence wind observations $\mathbf{v}_{ob}$ with an error $|\delta\mathbf{v}_{ob}|$ and height observations Φ_{ob} with an error $|\delta\Phi_{ob}|$ have on the error of the reconstructed rotational part of the stream

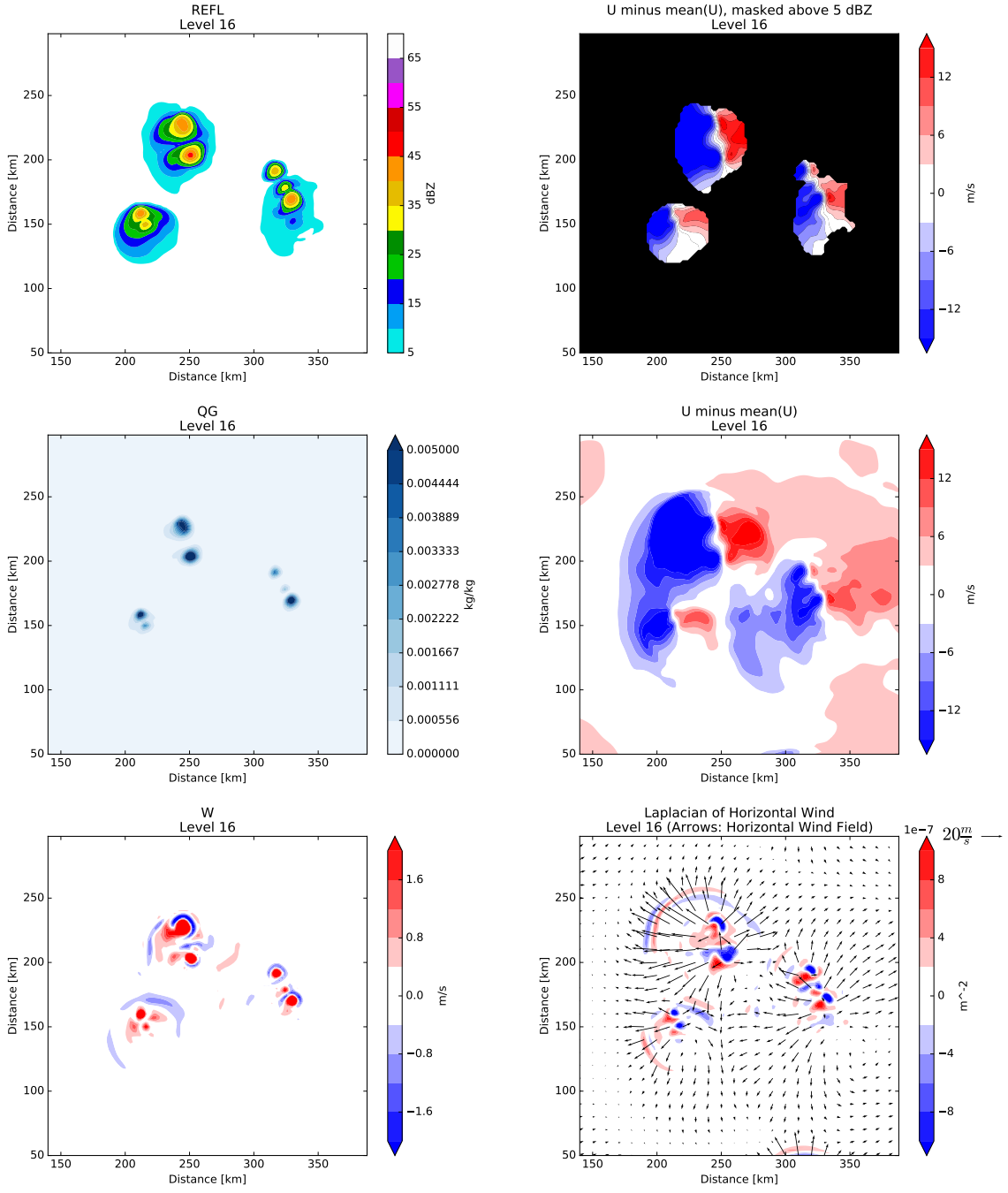

Figure 5.1: Left column, top to bottom: Observation fields of reflectivity (REFL), graupel mixing ratio (QG), snow mixing ratio (QS), vertical velocity (W). Right column, top to bottom: horizontal wind perturbation (U minus mean of U, masked by a threshold of 5 dBZ reflectivity, and unmasked), Laplacian of wind field with overlaid wind field arrows. All at model level 16 (12 km above ground) from nature run 01 of 2 at 14 UTC.

function $|\delta\psi|$ when assimilated simultaneously:

$$\frac{1}{|\delta\psi|^2} = \frac{f_0^2}{g^2|\delta\Phi_{ob}|^2} + \frac{n^2}{|\delta\mathbf{v}_{ob}|^2} \tag{5.1}$$

Here f_0 is the coriolis parameter, which has its maximum value at the poles and is zero at the equator, g is the gravitational acceleration, and n is the wavenumber of the observed motion. Kalnay (2003) concludes that short waves (large n) in the tropics (small f_0) are well-reconstructed using wind observations, while large waves (small n) outside of the tropics (large f_0) benefit most from mass-observations of the surface height.

Zeng and Janjić (2016) performed Local Ensemble Transform Kalman Filter (LETKF)-assimilation experiments with a SWM and assimilated combinations of surface height and wind observations. They observed a reduction in the noise of the analysis fields when adding wind observations to the assimilation of height observations. In the same process, the generation of spurious enstrophy (and therefore vorticity) is bounded by the wind observations. These findings only apply for global scales where geostrophically balanced rotational components have the largest scale (bounded by the Rossby radius of deformation) and divergent motion like fronts and large hydrostatically balanced gravity waves have a relatively small scale. As an example for a practical implementation of noise reduction in global models, Hamrud et al. (2015) present an EnKF setup for the hydrostatic Integrated Forecasting System (IFS) in which the surface pressure tendency (cf. Chapter 3) is analyzed as a control variable (Lorenc, 2003; Kepert, 2009). From this surface analysis, the overlying mass divergence is derived and distributed in the members, which causes a damping of noise and thereby a reduction of forecast errors (as in Zeng and Janjić, 2016) for surface based observations.

For convective scales, it is necessary to look at the small scale phenomena that are given by the non-hydrostatic cumulus dynamics of deep convection such as thunderstorms. When viewed from above at the height of the tropopause, thunderstorms are characterized by a point-like updraft core (or overshooting top) and a divergent wind field that is centered on the core and spreads into the surrounding (Figure 5.1). The model variable that is related to strong up- and downdrafts is vertical velocity which gains its driving buoyancy by the means of diabatic heating due to condensation. The resulting hydrometeors are a good tracer of vertical motion and are observed e.g. by the means of radar reflectivity. In case of COSMO, the dominant hydrometeor variable of the updraft cores is graupel (QG in Figure 5.1). The divergent motion is, depending on the number of overlapping beams from different radar sites, fully or partially observed by the radial wind observations which contain mostly horizontal wind components (such as U in Section 2.2.2). Figure 5.2 provides a simplified picture of the associated dynamics.

A background ensemble which contains a convective updraft also contains correlations to the corresponding divergent flow pattern through its ensemble covariances (Snyder and Zhang, 2003; Tong and Xue, 2005). Therefore, without localization and with an assumed absence of sampling noise, the assimilation of reflectivity observations will cause an analysis increment that also contains a wind pattern which modifies both the stream function and the divergence.

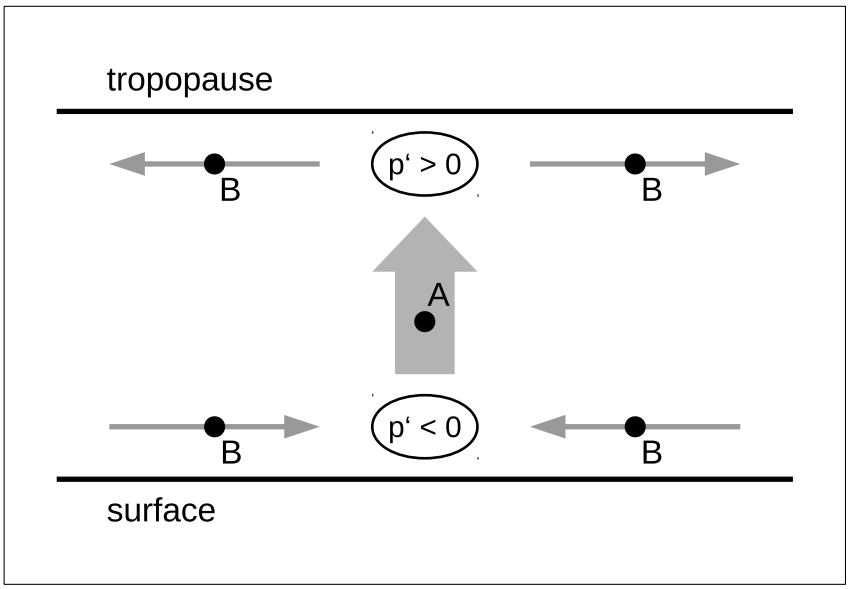

Figure 5.2: Conceptual illustration of the convective dynamics in Figure 5.1 (see also Figure 1.1). Point (A) is inside the updraft, points with (B) are in the divergent wind pattern. p' is the perturbation pressure that is associated to the updraft (cf. Figure 7.1 of Houze, 1993).

It therefore makes sense to formulate an equation similar to (5.1) for the divergence D. As a basis, the thermodynamic equation

$$\frac{\partial \Theta}{\partial t} + \mathbf{v}_h \cdot \nabla_h \Theta + w \frac{\partial \bar{\Theta}}{\partial z} = Q \tag{5.2}$$

relates the potential temperature Θ (Holton, 2004) with the horizontal wind $\mathbf{v}_h$ and the vertical velocity w to a heat-source term Q. Assuming a quasi-steady state of an updraft circulation (as in Figure 5.2), the temporal derivative $\partial\Theta/\partial t$ is canceled. The advection term is assumed to be negligible since the horizontal gradient of potential temperature is much smaller than the vertical gradient (and evened out by gravity waves very fast), thus $\mathbf{v}_h \cdot \nabla_h \Theta = 0$. The remaining term on the left side is $w\partial\bar{\Theta}/\partial z$ where $\bar{\Theta}(z)$ is the vertical sounding of the environmental temperature, with the vertical derivative as the stability parameter. In a quasi-steady state of an updraft, the rising motion is therefore in balance with the local heating rate Q which is renamed as $\dot{\Theta}$:

$$w \frac{\partial \bar{\Theta}}{\partial z} = \dot{\Theta} \tag{5.3}$$

The divergence is given by $D = \nabla \cdot \mathbf{v}_{3d}$ and can, via scale analysis, be related to the vertical velocity scale W over a scale height H as

$$D \propto \frac{W}{H} \tag{5.4}$$

and to the horizontal wind V over a horizontal length scale L as

$$D \propto \frac{V}{L} \ . \tag{5.5}$$

Substituting the W of (5.4) into (5.3) provides an estimate of how D is related to $\dot{\Theta}$ and the stability $\partial\bar{\Theta}/\partial z$:

$$DH\frac{\partial\bar{\Theta}}{\partial z} \propto \dot{\Theta} \ . \tag{5.6}$$

The estimation equation for D can then, analogous to Equ. (5.1), be formulated as

$$\frac{1}{|\delta D|^2} = \frac{(H\frac{\partial\bar{\Theta}}{\partial z})^2}{|\delta\dot{\Theta}_{ob}|^2} + \frac{L^2}{|\delta\mathbf{v}_{ob}|^2} \ . \tag{5.7}$$

$\delta\dot{\Theta}_{ob}$ is the observed change in heating rate in the column and can be estimated by

$$\delta\dot{\Theta}_{ob} = \frac{L_v\delta r_{ob}}{c_p H\rho} \ , \tag{5.8}$$

where δr_{ob} is the observed change of rain rate in the column (in kg m^{-2} h^{-1}), H is the scale height, ρ is a reference air density, $L_v = 2.6\ 10^6$ J kg^{-1} is the latent heat of condensation, and $c_p = 1004$ J K^{-1} kg^{-1} is the specific heat of dry air at constant pressure. For H, a typical value is the height of the tropopause at 10 km, and a typical air density is $\rho = 1$ kg m^{-3}.

The question *which relative influence a rain-observation and a wind-observation has on the divergence in the model state* may now be reformulated as a question of scale: *Below which scale is the influence of rain-observations dominant, and above which scale is the influence of horizontal wind observations dominant?* To answer this, one may equate the right hand side terms of Equ. (5.7), with the substitute (5.8), to get

$$\frac{(H\frac{\partial\bar{\Theta}}{\partial z})^2}{|\frac{L_v\delta r_{ob}}{c_p H\rho}|^2} = \frac{L^2}{|\delta\mathbf{v}_{ob}|^2} \ . \tag{5.9}$$

Using $\partial\bar{\Theta}/\partial z = 3$ K km^{-1} as a typical value of stability (computed from the Payerne sounding in Chapter 2), an observed change in rain rate of $\delta r_{ob} = 10$ kg m^{-2} h^{-1} and an observed change in horizontal wind of $|\delta\mathbf{v}_{ob}| = 10$ m s^{-1}, a horizontal scale of approximately $L \approx 40$ km is estimated. This scale concurs with the depiction in Figure 5.1 and the vicinity size around convective storms that was derived in Chapter 3.

When performing an analysis update with an EnKF that assimilated rain- and wind-observations, the increments are by definition tainted by errors such as sampling noise in the background ensemble covariances. These errors can, as was shown in Chapter 3, disrupt the dynamical balance between model variables and thus be harmful to the forecasts. Since there is an interplay between the effects of hydrometeor observations (reflectivity) and wind observations (radial velocity), the present chapter is devoted to the question *which influence the different radar-based observations types exert on the dynamical balance*

of the forecast model. If possible, the question *whether a small noise level is related to small forecast errors* (Hamrud et al., 2015; Zeng and Janjić, 2016) shall also be asked. As a research tool, a self-nested version of COSMO is used, in which radar data assimilation is performed similar to the experiments of Chapter 2. Two versions of assimilating rain-observation will be evaluated. On the one hand, the statistically based EnKF method is used to update the model according to radar observations. On the other hand, the physically based Latent Heat Nudging (LHN) method (Stephan et al., 2008) is tested.

Section 5.2 describes the model and the DA-setup, together with the observation system. Section 5.3 describes the convective case and the evaluation measures. In Section 5.4, the results during cycling and 3-hour forecasts are presented. Section 5.5 provides a summary and gives an outlook on possible further studies based on the presented setup and research question.

5.2 Setup of the COSMO-MUC Model and the DA

5.2.1 COSMO-MUC model

Similar to Hagelin et al. (2014), the model of COSMO-MUC (Soelch, 2015) is being developed as a nested version of COSMO-DE (see Chapter 4), centered around the Munich Airport MUC as shown in Figure 5.3. The Initial Conditions (IC) of COSMO-MUC are cut-out fields from COSMO-DE. The Boundary Conditions (BC) for COSMO-MUC are provided by COSMO-DE with an interval of 15 minutes. In between these times, the BC are interpolated linearly in time. COSMO-MUC has the same horizontal resolution of 2.8 km as COSMO-DE and the same vertical structure with 50 levels.

5.2.2 Data Assimilation with COSMO-MUC-KENDA

As in Chapter 4, DA is performed with the LETKF of the Kilometre-scale Ensemble Data Assimilation (KENDA) for both COSMO-DE and COSMO-MUC. COSMO-DE-KENDA employs 40 ensemble members and 1 deterministic run. As in Schraff et al. (2016) and Chapter 4, an LETKF-ensemble of ICON with 3-hourly global analyses (Fernandez et al., 2015) provides the hourly BCs. Every member of COSMO-DE then provides the BCs for one member of the COSMO-MUC ensemble. As part of this study, the additional nesting-layer was introduced into the cycling-system to run COSMO-MUC as an inner ensemble system that is driven by the outer ensemble system of COSMO-DE. This included the development and testing of newly written software that drives the COSMO-MUC ensemble experiments with partly different DA-parameters, correctly transfers the boundary conditions, and archives the very high number of output files in a consistent data structure.

5.2.3 Observation System

As in Chapter 4, observations of surface stations (SYNOP) are used, but now with a less strict filtering: All observations of surface temperature (T2M), wind (U10M, V10M) and

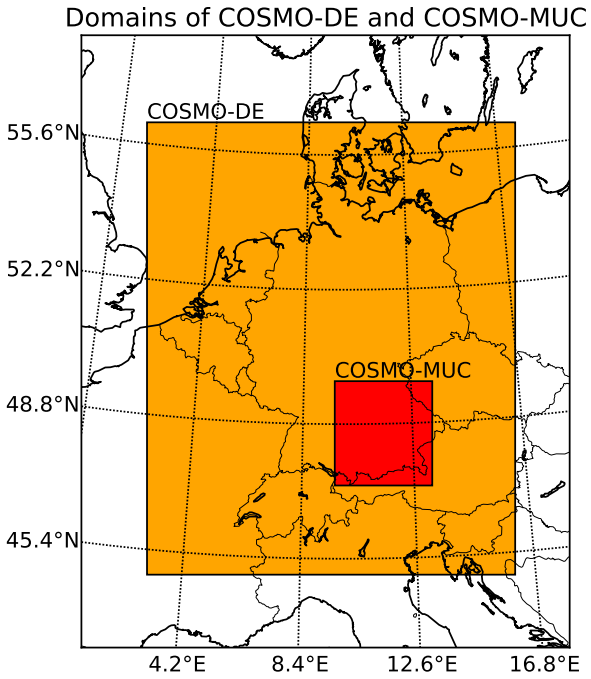

Figure 5.3: Domain extents of COSMO-DE (orange) and COSMO-MUC (red). The global driving model is ICON (see text).

relative humidity (RH2M) are assimilated in COSMO-DE-KENDA and COSMO-MUC-KENDA, as long as the true station height does not differ more than 150 meters from the station height that is represented by the model (for wind: 100 meters). Aircraft observations of AMDAR and Mode-S are assimilated in all experiments without thinning, as in the experiment MAconv described in Table 4.1. The COSMO-MUC domain contains three of the C-Band radar stations of the DWD-network: Türkheim (TUR), Memmingen (MEM), and Isen (ISN). The model equivalents of radar observations are computed in COSMO using the Efficient Modular VOlume scanning RADar Operator (EMVORADO, Zeng et al., 2016). An example of their horizontal coverage is given in Figure 5.4. The grid-like structure in the horizontal is caused by averaging raw radar observations into superobservations (see Section 2.3.2) with a horizontal bin-resolution of 10 km. Similar to (Aksoy et al., 2009; Bick et al., 2016) and Chapter 2, a value of 5 dBZ was assigned to the reflectivity observations and the model equivalents below 5 dBZ, in order to use them to suppress spurious convection in the analysis members (Aksoy et al., 2009). The observation limits of the single stations are apparent: TUR covers the western half of the MUC domain, MEM a circle up to Regensburg, ISN almost the full domain. Note that due to the upward ray propagation of the lowest scan angle, the height of the lowest observed level increases with distance from the radar stations. Still, the radar observations, if provided with enough hydrometeors, can cover a larger volume than the aircraft observations, so these two observation spaces are not congruent (illustrated by Figure 5.5). Figure 5.6 illustrates how radar observations provide information about po-

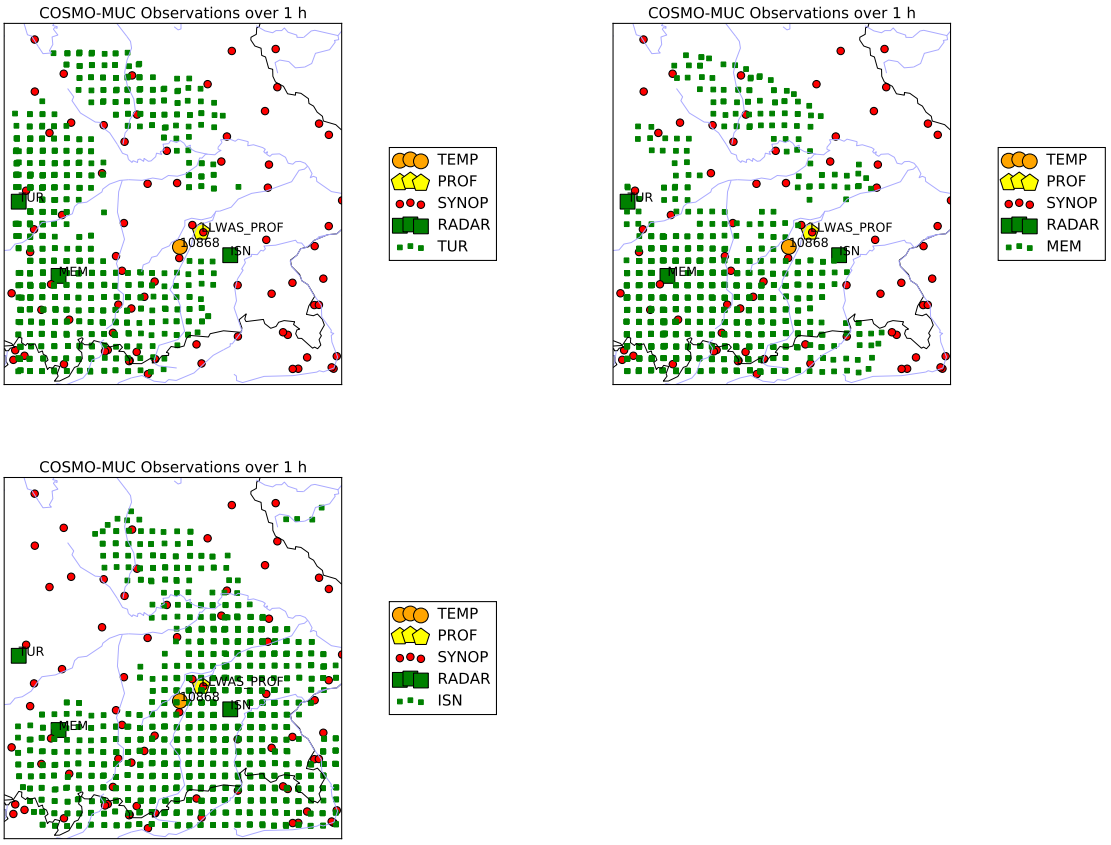

Figure 5.4: COSMO-MUC observation coverage at analysis time of 26.5.2014, 12 UTC. Shown are SYNOP stations, LLWAS profiler at airport MUC, radiosonde 10868 (Oberschleissheim). The different panels show the radial velocity observation coverage of the radar stations TUR, MEM, ISN in that particular hour.

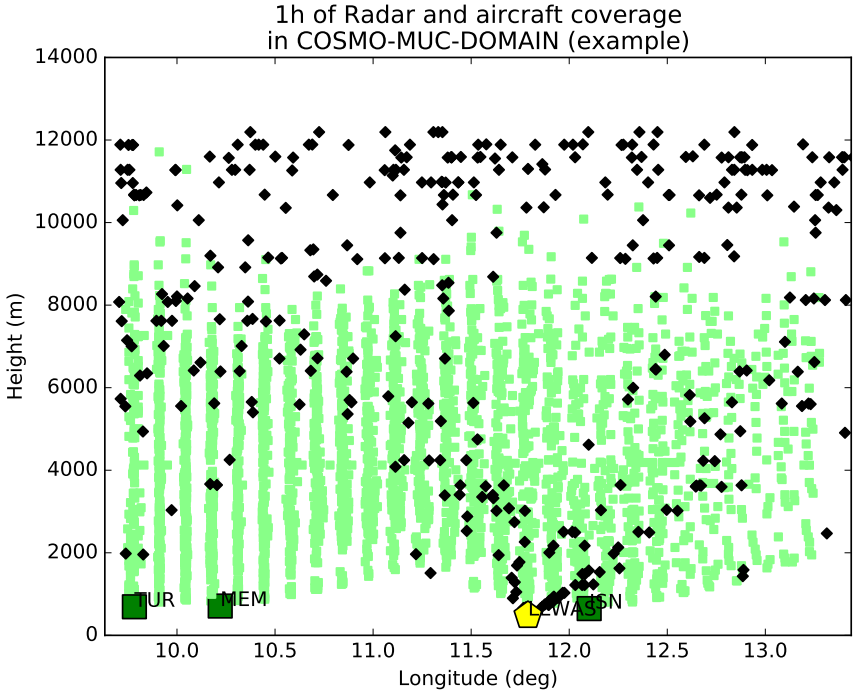

Figure 5.5: Vertical coverage of radar radial velocity (green) from Figure 5.4 and aircraft wind data (black) as depicted in Figure 4.2.

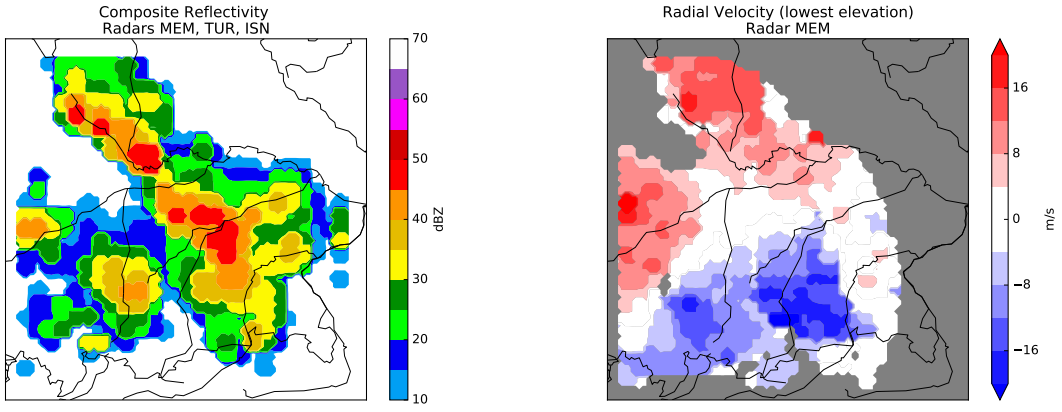

Figure 5.6: Composite reflectivity of all three radars in the MUC domain (left) and corresponding radial velocity of the MEM-station (cf. Figure 5.4) at 15 UTC, 26.5.2016

sition and intensity of convection through reflectivity and about the underlying dynamics through radial velocity.

5.2.4 DA Setup

Following the recommendations of Chapters 2 and 3, the DA parameters for radar are chosen similar to the experiment L32SOCG20 (as was done in Bick et al., 2016). The horizontal localization length scale (see Section 2.2.5) for radar reflectivity and radial wind was chosen as $l_{hor,rad} = 16$ km, corresponding to a cutoff length of roughly 60 km. For conventional observations, the horizontal localization was chosen adaptively (cf. Section 4.3.3) with a lower limit of $l_{hor,min} = 25$ km and an upper limit of $l_{hor,min} = 50$ km. The forecast window between analyses was chosen with a length of $\Delta t = 15$ min. As a conservative choice, the observation error standard deviation was assumed to be 10 dBZ for reflectivity and 5 m s^{-1} for radial velocity as in (Bick et al., 2016).

5.2.5 Latent Heat Nudging Setup

Instead of updating the model state with ensemble covariances produced by an LETKF that assimilates rain, one can also directly force the model to produce or suppress rain when the forecast deviates from the observations. This method is called Latent Heat Nudging (LHN) and is operational for COSMO (Leuenberger, 2005; Stephan et al., 2008). In LHN, a surface rain rate $RR_{obs}(x, y)$ (in mm/h) is heuristically derived from radar observations of reflectivity. $RR_{obs}(x, y)$ is compared to the surface rain rate $RR_{mod}(x, y)$ that is predicted by the model, resulting in a local mismatch factor

$$f(x, y) = RR_{obs}(x, y)/RR_{mod}(x, y). \tag{5.10}$$

In the model, the latent heat (LH) release is predicted as $\Delta T_{mod,LH}$. If a mismatch of forecast and observation of rain is found (i.e. $f \neq 1$), a heating increment

$$\Delta T_{LHN} = (f - 1)\ \Delta T_{mod,LH} \tag{5.11}$$

is applied in the model column. For $f(x, y) > 1$, $\Delta T_{LHN} > 0$ is a heating increment that increases the local buoyancy and thus amplifies vertical motion, subsequently producing more rain in the model. For $f(x, y) < 1$, $\Delta T_{LHN} < 0$ is a cooling increment that decreases the local buoyancy and thus suppresses vertical motion, subsequently producing less rain in the model. ΔT_{LHN} is applied incrementally as a forcing term in the prognostic equations and is therefore a smoother method than the intermittent analysis updates of the LETKF system. While this can be regarded as an advantage with respect to the model dynamics, the wind circulation associated to rain-producing convection is only reproduced as a divergent secondary effect of LHN since the radial velocity observations are not nudged. This means that the influence on the divergence field, as estimated in Equ. (5.7), is performed only indirectly through the response of the model dynamics. Therefore, differences in the model response are expected when the assimilation of hydrometeor observations between LHN and LETKF is compared inside the MUC domain, especially when radial velocity is added in the LETKF.

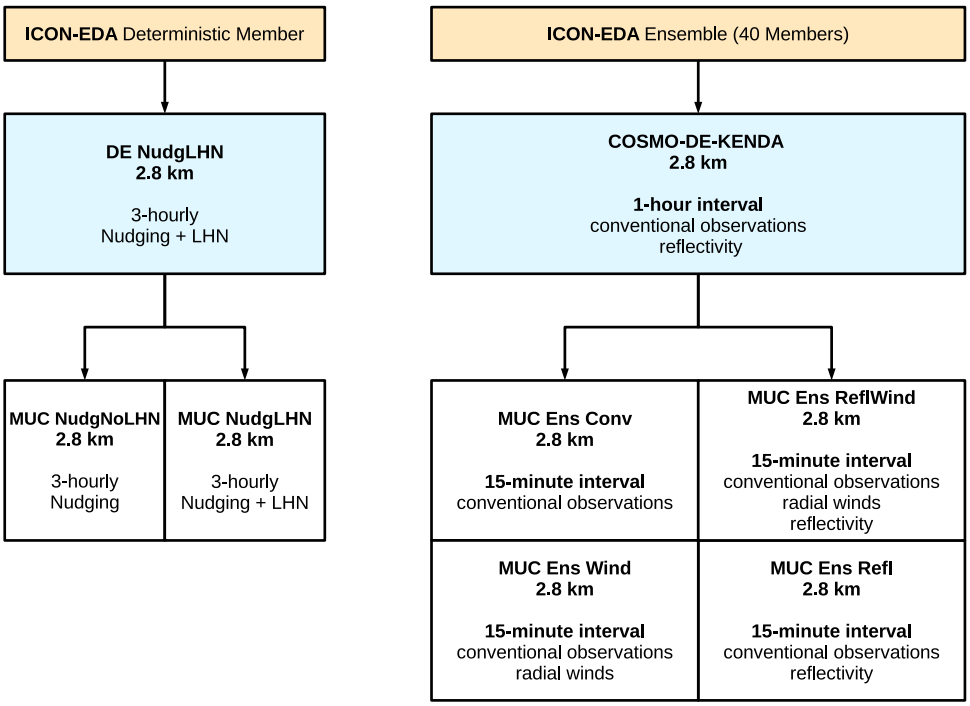

Figure 5.7: Flowchart of the experiments described in Section 5.3

5.3 Experiments on the influence of observation types

In order to investigate the influence that radar observation types have on the model dynamics, a basic ensemble data assimilation experiment (i, *MUC Ens Conv*) is performed in which only conventional observations are assimilated (Figure 5.7). In parallel, experiments are performed where reflectivity (ii, *MUC Ens Refl*), radial velocity (iii, *MUC Ens Wind*) and a combination of both are assimilated (iv, *MUC Ens ReflWind*) in addition to the conventional observations. The boundary conditions for these MUC-KENDA experiments with 40 members are provided by one COSMO-DE-KENDA experiment (*DE Ens Refl*) with 40 members which assimilates conventional observations and radar reflectivity in an hourly cycle. *DE Ens Refl* does not assimilate radial wind in order to keep the input of observation information similar to the nudging experiment. Two deterministic experiments with nudging are performed, one with LHN (*MUC NudgLHN*) and one without LHN (*MUC NudgNoLHN*) to determine the influence of the LHN scheme within the MUC domain. Both are driven by a deterministic COSMO-DE experiment (*DE NudgLHN*) that has LHN activated and is therefore comparable (with respect to the data input) to the *DE Ens Refl* boundary experiment. *DE NudgLHN* has the same assimilation settings as *MUC NudgLHN* and is driven by the boundary conditions of the deterministic run of the ICON-Ensemble. Every 3 hours, a forecast with a lead time of 3 hours is started for both deterministic and probabilistic COSMO-MUC experiments and for the respective boundary experiments of COSMO-DE. The forecasts of COSMO-DE and COSMO-MUC are therefore synchronized with the 3-hourly cycle of the global ICON BCs (cf. Section 4.3.1). As in Section 4.4, one day of cycled data assimilation without

assimilating radar data was performed beforehand to spin up the COSMO fields and the adaptive covariance inflation.

5.3.1 Description of the convective case

The experimental period comprises one day (26th of May 2014) where a convergence line was located inside the MUC domain, diagonally from the north-west to south-east (Figure 5.8). After the decay of some previous convection by 6 to 9 UTC, new convective cells were triggered along the convergence line and merged into a larger system by 18 UTC that filled a great part of the MUC domain (Figures 5.9 and 5.10) and continued to rain until the morning of the next day. Figure 5.11 shows the number of radar-observations per type throughout the day, wherein *no-reflectivity* observations are reflectivity observations below 5 dBZ. Due to this temporal development, the evaluation focuses on the times between 9 UTC and 0 UTC to cover the growth phase of the convective systems and the mature and decaying phase.

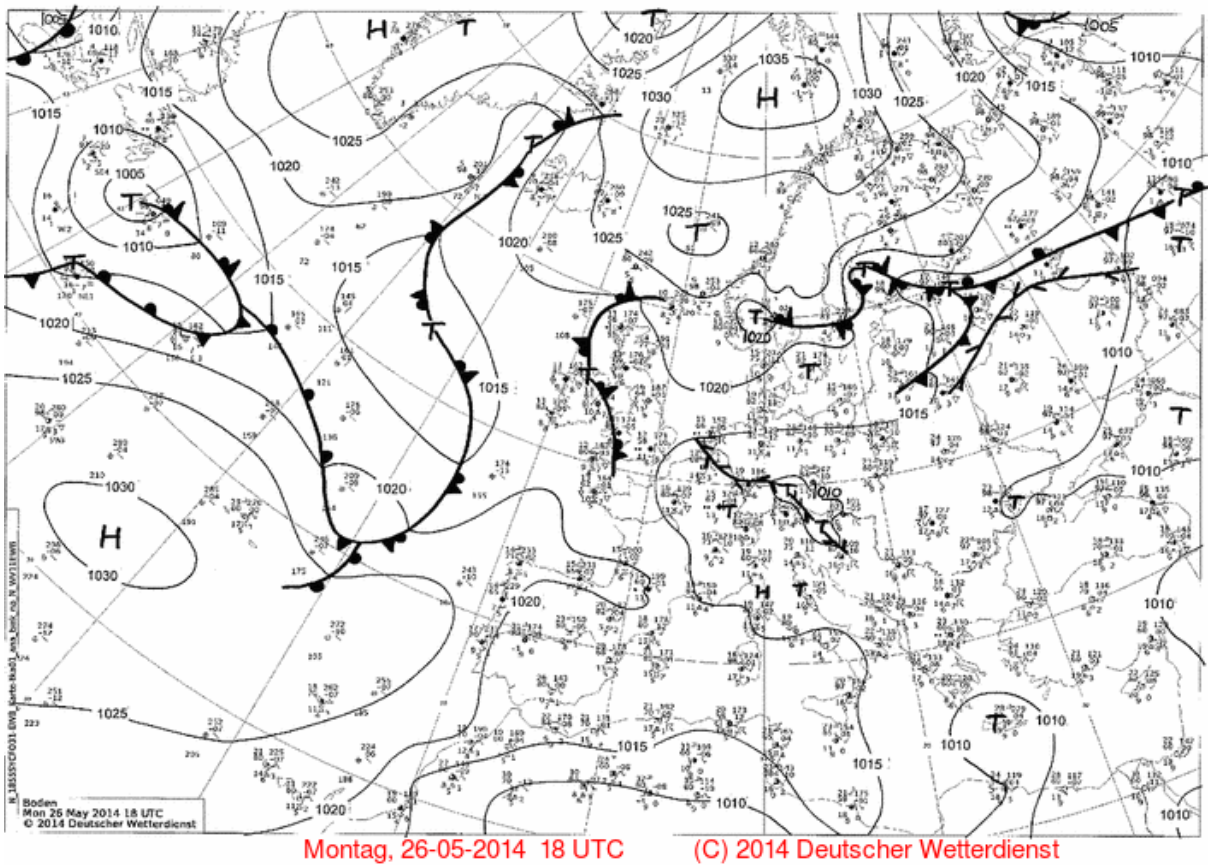

Figure 5.8: Synoptical situation on the 26th of May 2014 at 18 UTC.

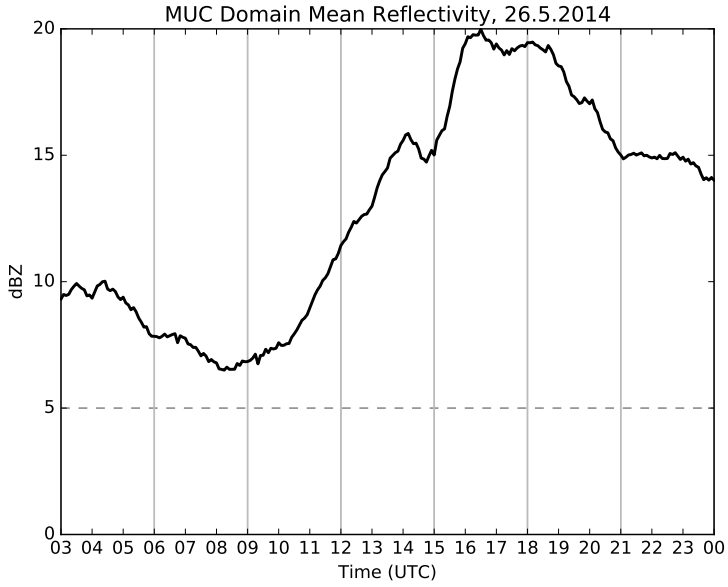

Figure 5.9: Observed Mean Reflectivity in the MUC domain on the 26th of May 2014 (cf. Figure 5.10). The lowest observed value is the lower threshold of 5 dBZ.

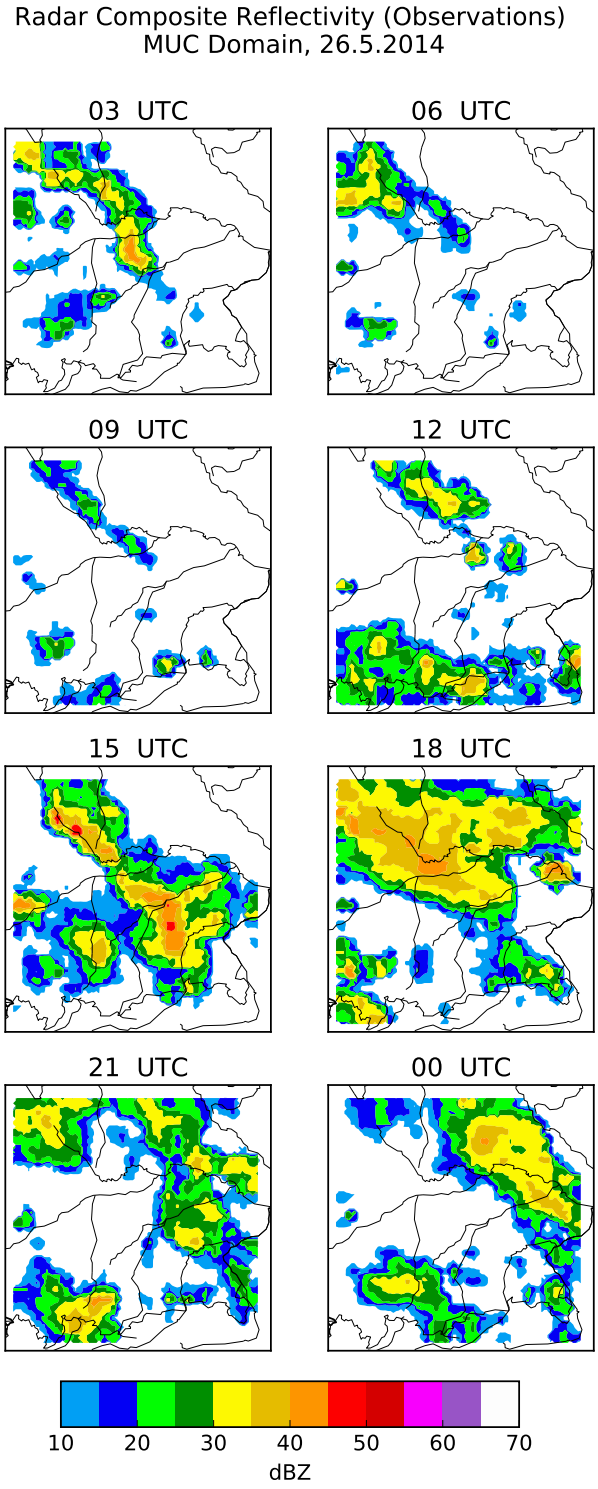

Figure 5.10: Observed reflectivity in the MUC domain on the 26th of May 2014 for every 3 hours (cf. Figure 5.9)

5.3.2 Evaluation Measures

As in Chapter 2, the performance of convection forecasts with lead times of 3 hours will be evaluated by the means of the composite reflectivity field. This field is created for this study by a self-developed software that maps the unstructured radar observation points (cf. Figure 5.6) onto the model grid. A brute-force algorithm is employed in which all observations within a horizontal radius of 7.5 km around a single grid point are averaged. The vertical maximum projects the interpolated three-dimensional volume reflectivity onto the two horizontal dimensions. Figure 5.10 shows a typical product of this algorithm – the coarse structures are a result of the prior 10 km averaging of the superobservations.

The Displacement and Amplitude Score (DAS) and the Structure, Amplitude, Location (SAL) score (Section 2.4.2) are computed on the composite reflectivity fields that are thresholded above 20 dBZ. For DAS, both the displacement error component (DAS-DIS, cf. Section 2.4.2) and the amplitude error component (DAS-AMP) are shown. DAS-AMP (Keil and Craig, 2009) is the sum of AMP_{obs} and AMP_{fct}. AMP_{obs} is the RMS difference between the observed rain field and the forecast rain field that is morphed onto the observed field, and is therefore sensitive to rain objects that are missed by the forecast. For AMP_{fct}, the morphing is the other way around, and it is therefore sensitive to rain objects that are falsely predicted by the forecast. DAS-AMP can be regarded as a mixture of the categories *miss rate* and *false alarm rate* (see Wilks, 2006). Additionally, the Fraction Skill Score (FSS) is applied. Like the DAS, the FSS (Roberts and Lean, 2008) takes into account that there are necessarily some location errors in rain forecasts that could be counted as double penalties, even when the displacement of the forecast feature is only one grid point away from the observed location. To prohibit this, the observed and forecast rain-fields are thresholded (here the 20 dBZ contour is used) and the agreement of the fractional coverage inside a spatial neighborhood (here 9 grid points are used, resulting in a box size of around $(25\ \mathrm{km})^2$) is compared. An FSS-value of 1 denotes a perfect forecast with respect to the chosen threshold and neighborhood size.

The noise-measures of Chapter 3 are evaluated during the cycling for all experiments to see whether a connection between analysis noise and forecast error can be made as in Zeng and Janjić (2016). The masking algorithm defined in Section 3.3 is applied to the three-dimensional reflectivity observation field on the model grid, with a vicinity distance of 45 km around grid points that have an observed reflectivity larger than 5 dBZ. The number of points in the masks is shown in Figure 5.12. As a traditional measure, the forecast RMSE, spread and bias (cf. Section 2.2.7 and Section 4.4) of reflectivity and radial velocity observations are shown.

5.4 Results

5.4.1 Quantitative Precipitation Forecast Results

Here the results of the Quantitative Precipitation Forecast (QPF) scores are presented. For ensemble experiments, the scores are computed for every member separately and then

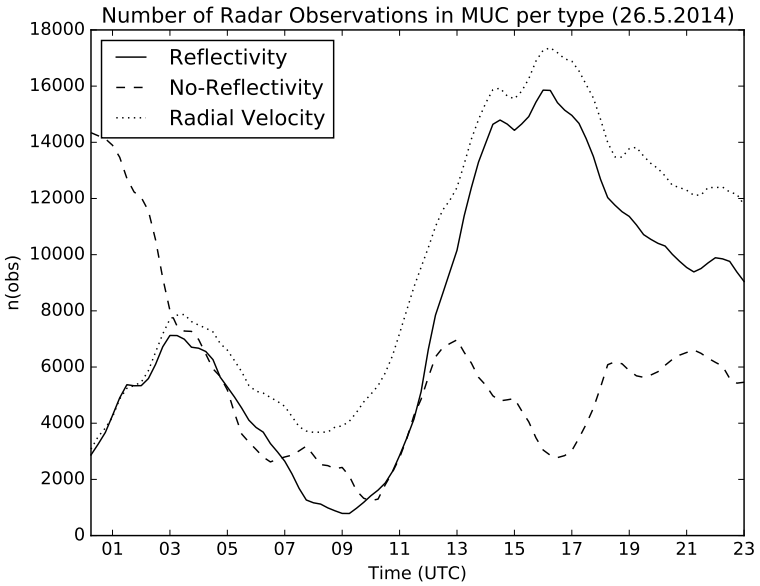

Figure 5.11: Number of radar observations in the MUC domain on the 26.5.2014 by type.

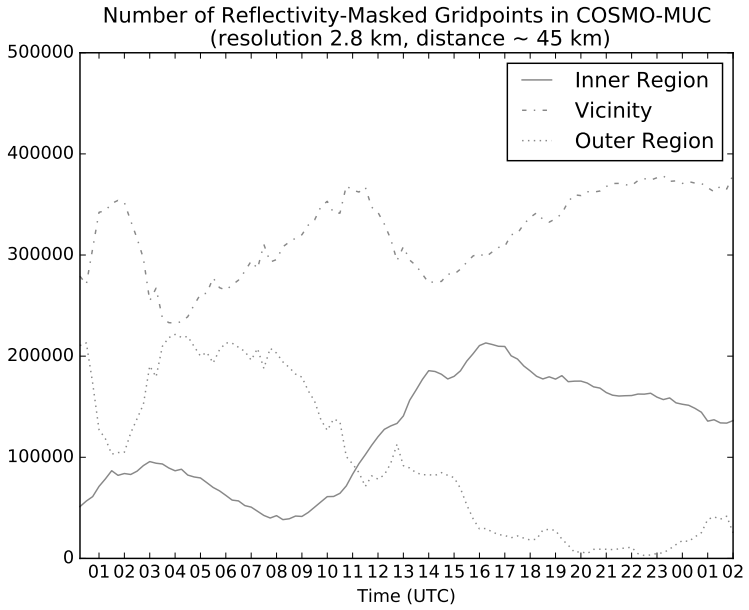

Figure 5.12: Timeseries of number of points in the masks with respect to the observed reflectivity field (cf. Figures 5.9, 5.10, and 5.11).

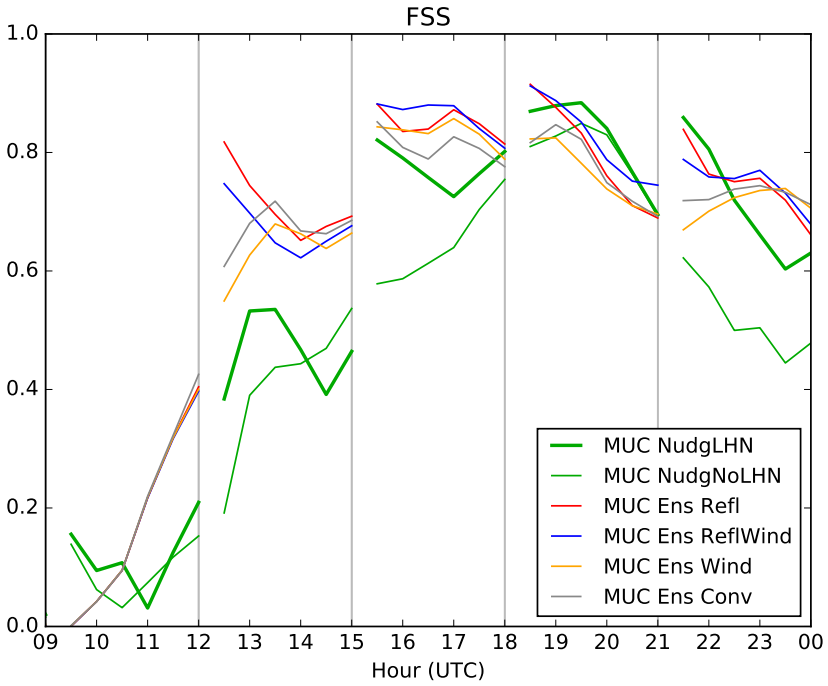

Figure 5.13: FSS of the COSMO-MUC experiments (threshold of 20 dBZ, neighborhood size of 25 km). Vertical lines represent the start-times of 3-hour forecasts.

averaged.

Influence of LHN

First, *MUC NudgNoLHN* is compared to *MUC NudgLHN*. In the FSS (Figure 5.13), it is apparent that short lead times (up to 1 or 2 hours) of the 3-hour forecasts benefit from the prior influence of LHN. Within 3 hours this advantage of *MUC NudgLHN* may be lost, indicated by a convergence with the FSS-value of *MUC NudgNoLHN* (e.g. at 14, 18, 20 UTC). The bias (Figure 5.14) shows that this behavior is not due to a continuous systematic difference in convection amount between the two experiment, since sometimes *MUC NudgNoLHN* has more convection present, and sometimes *MUC NudgLHN*. DAS-DIS (Figure 5.15) shows that LHN improves the positioning of the convection in short forecast lead times, but within most 3 hour forecast windows, this benefit is lost. For the mature and late phase of the convection between 15 and 0 UTC, DAS-AMP (Figure 5.16) indicates a beneficial influence of LHN on misses and false alarms that holds throughout the 3 hours of forecast lead time.

Influence of radar assimilation in ensemble experiments

As with the comparison of deterministic LHN against NoLHN, the influence of radar observations in the ensemble experiments are evaluated now. In the FSS (Figure 5.13), the experiments with assimilation of reflectivity (*MUC Ens Refl* and *MUC Ens Re-*

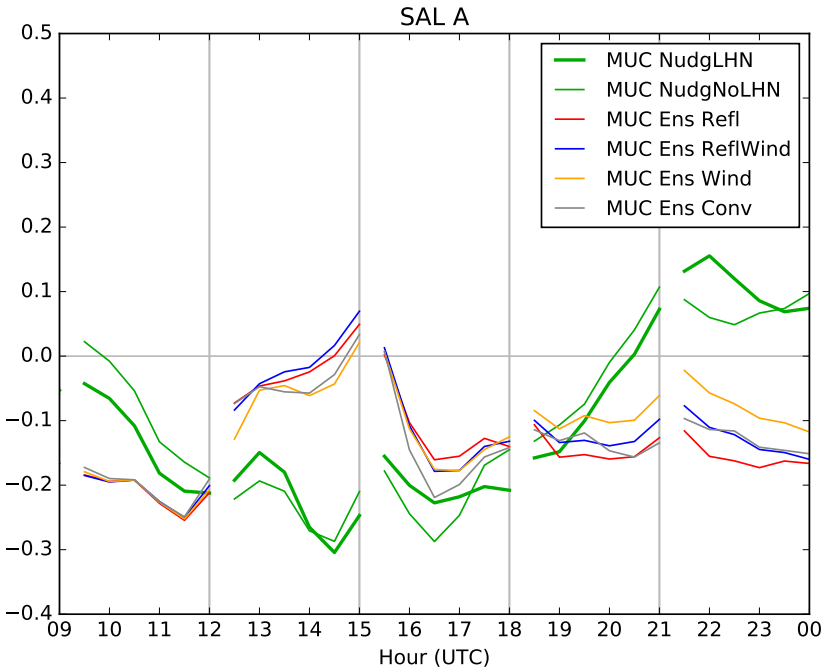

Figure 5.14: Forecast bias (SAL-A, cf. 2) of the composite reflectivity field.

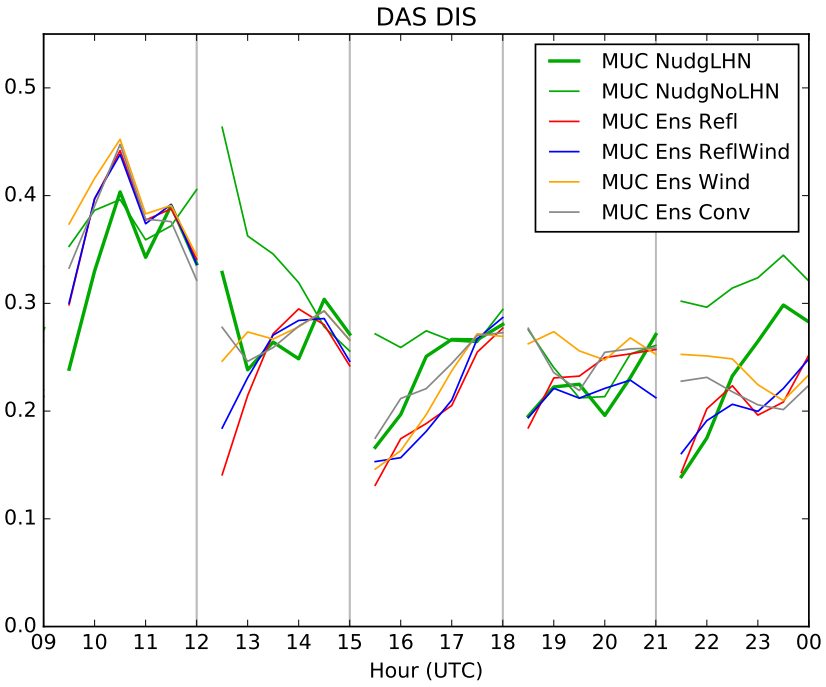

Figure 5.15: As Figure 5.13, for the DAS-DIS component (threshold of 20 dBZ, search radius of 45 km).

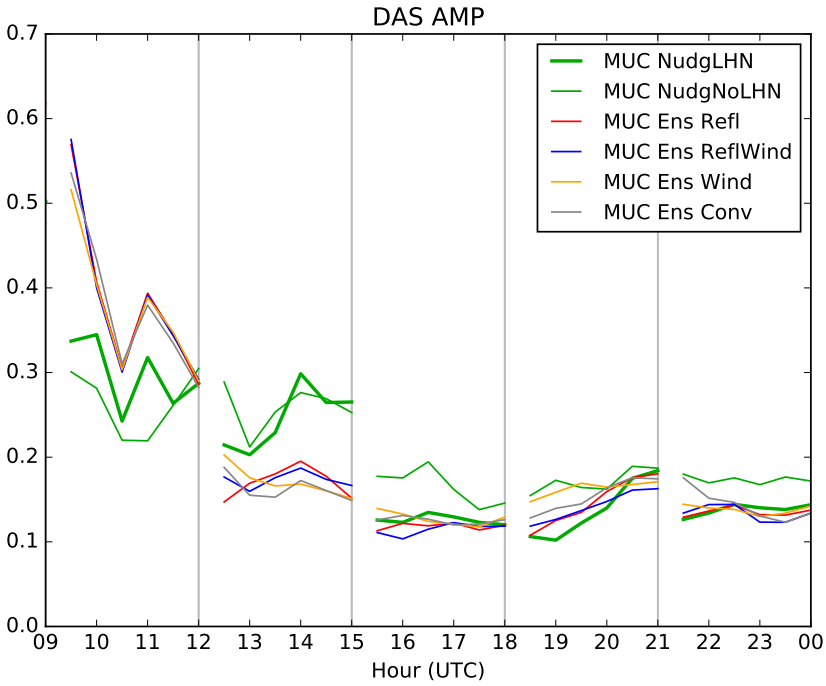

Figure 5.16: As Figure 5.13, for the DAS-AMP component (threshold of 20 dBZ, search radius of 45 km).

flWind) exhibit a better performance than experiments without assimilation of reflectivity (*MUC Ens Conv* and *MUC Ens Wind*), especially for short forecast lead times. This is expectable, since the FSS is computed using the composite derivative of the assimilated volume reflectivity field. At the end of the 3 hour forecast windows, the advantage is lost and the FSS scores of the ensemble experiments have mostly converged. SAL-A (Figure 5.14) indicates that between 18 and 00 UTC, there is less negative bias in the amount of convection if only radial velocity is assimilated and reflectivity is not (*MUC Ens Wind* compared to *MUC Ens ReflWind*). This concurs with the findings of Bick et al. (2016) who diagnosed a lack of rain in forecasts that were initiated after the thresholded reflectivity had been assimilated. As a possible reason, it was suggested that the suppression of spurious convection due to assimilating no-reflectivity observations could damp the evolution of the truly observed storms in the forecast members. DAS-DIS (Figure 5.15) shows even clearer than the FSS that short lead times benefit from the assimilation of reflectivity with respect to the positioning of the storms. A convergence towards the ensemble experiments without radar data after 3 hours is again apparent. When *MUC Ens ReflWind* is compared to *MUC Ens Refl*, it is apparent that the addition of radial velocity to the assimilation of reflectivity does not yield significant benefits in the 3 hour forecasts of this convective case, with respect to the presented measures. DAS-AMP (Figure 5.16) does not show significant differences between the ensemble experiments.

Deterministic nudging compared to ensemble data assimilation

MUC NudgNoLHN and *MUC Ens Conv* use the same observation set where radar data is excluded in the MUC domain. It can be seen that the positioning of the convection is better in the *MUC Ens Conv* members than in *MUC NudgNoLHN* run, especially in the phase where convection triggers locally inside the MUC domain (FSS and DAS-DIS between 9 and 18 UTC). SAL-A shows that this goes in hand with a larger negative bias of *MUC NudgNoLHN* between 12 and 18 UTC. In the decaying phase between 18 and 0 UTC, this bias behavior is reversed. Overall, the ensemble experiment *MUC Ens Conv* appears to make better usage of the conventional data for the assimilation of convection than the *MUC NudgNoLHN* experiment (see also DAS-AMP in Figure 5.16). As a second possibility, the boundary data from *DE Ens Refl* could be more accurate than to the boundary data from *DE NudgLHN*, but only if convection or convective triggers are advected from the boundaries after they have been assimilated in the surrounding COSMO-DE domain. Since this is not really the case for the stationary convergence line of this case, it can be regarded as a negligible effect for this particular day.

LHN compared to radar assimilation in ensemble experiments

Now it is tested how beneficial LHN is when compared to the ensemble assimilation of radar data. Comparing *MUC NudgLHN* to *MUC Ens Refl* and *MUC Ens ReflWind*, the forecast performance of convection location (FSS in Figure 5.13 and DAS-DIS in Figure 5.15) is generally better in the ensemble data assimilation system. This is especially true for the rapid growth phase between 12 and 15 UTC where there is a negative bias in *MUC NudgLHN* (Figure 5.14) and its FSS, DAS-DIS and DAS-AMP (Figure 5.16) scores are comparatively bad.

5.4.2 Verification using radar volume observations

At the start of the 3-hour forecasts, the ensemble experiments which assimilate reflectivity (*MUC Ens Refl* and *ReflWind*) have the lowest RMSE and spread for volume reflectivity observations (Figure 5.17). By the end of the forecasts, their error has converged with *MUC Ens Conv*, while the spread in the reflectivity space is still smaller. *MUC Ens Wind* is the ensemble experiment with the largest error in the reflectivity observations. The nudging experiments are worse for reflectivity than the ensemble experiments, wherein *MUC NudgLHN* has a lower RMSE than *NudgNoLHN* throughout the whole forecast. The fact that *MUC NudgLHN* still has a much larger error than the ensemble experiments is probably due to the fact that LHN only assimilates a two-dimensional rain field (cf. Figure 5.10) while the ensemble experiments assimilate the full three-dimensional volume data of the radars (cf. Figure 5.5). All experiments exhibit a negative bias in reflectivity which is least severe for *MUC NudgNoLHN* at the end of the 3-hour forecast and for *MUC Ens Wind* at the beginning. This indicates that both LHN and ensemble assimilation of reflectivity have an overall suppressing influence on the hydrometeor density in the model, assuming that the model and the observation operator are mostly unbiased.

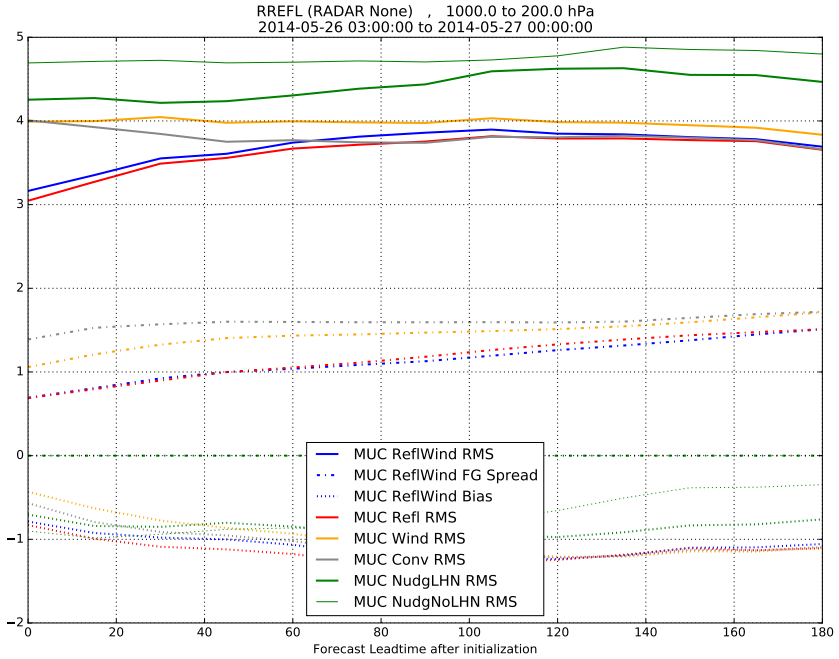

Figure 5.17: Forecast RMSE, spread and bias against observations of radar reflectivity (RREFL, in dBZ) of the MUC experiments. Results are averaged over all 3 hour forecasts.

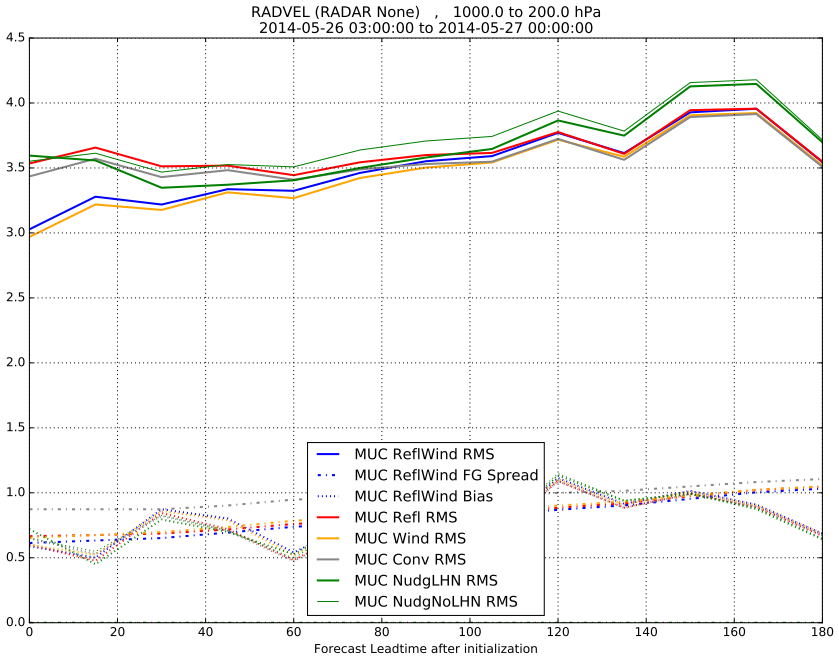

Figure 5.18: As Figure 5.17, for observations of radial velocity (RADVEL, in m s^{-1}).

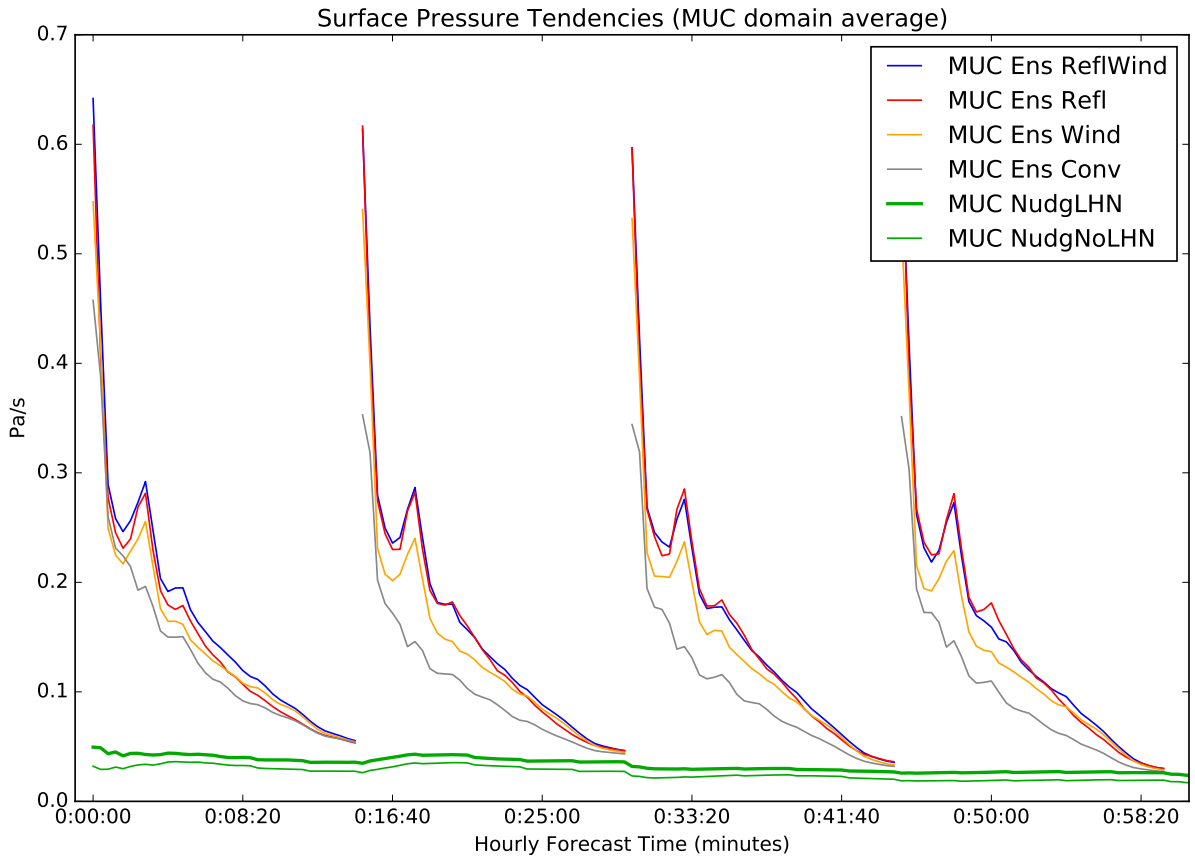

Figure 5.19: Surface Pressure Tendencies for MUC Experiments. The ensemble (*MUC Ens*) experiments are averaged with respect to their 15-minute assimilation windows within the hourly analysis update of their boundary experiments (*DE Ens Refl*). The nudging (Nudg) experiments are averaged for every 3-hour window since they continuously assimilate observations (only the first hour is shown).

At the start of 3-hour forecasts, the ensemble experiments which assimilate radial velocity (*MUC Ens Wind* and *ReflWind*) have the lowest RMSE for radial velocity (Figure 5.18). Any addition of reflectivity observations increased the initial error in the radial wind space. By the end of the forecasts, the error has converged with *MUC Ens Refl* and *Conv*. The performance of *MUC NudgLHN* and *NudgNoLHN* is initially similar to Muc Ens Conv and Refl, but becomes worse than all ensemble experiments at the end of the forecast windows. Despite the fact that LHN makes no use of radial velocity observations, the error of *MUC NudgLHN* is generally smaller than of *MUC NudgNoLHN*. This could be related to circulation patterns induced by the LHN-forced convection that are closer to the truth than in the absence of LHN-forcing.

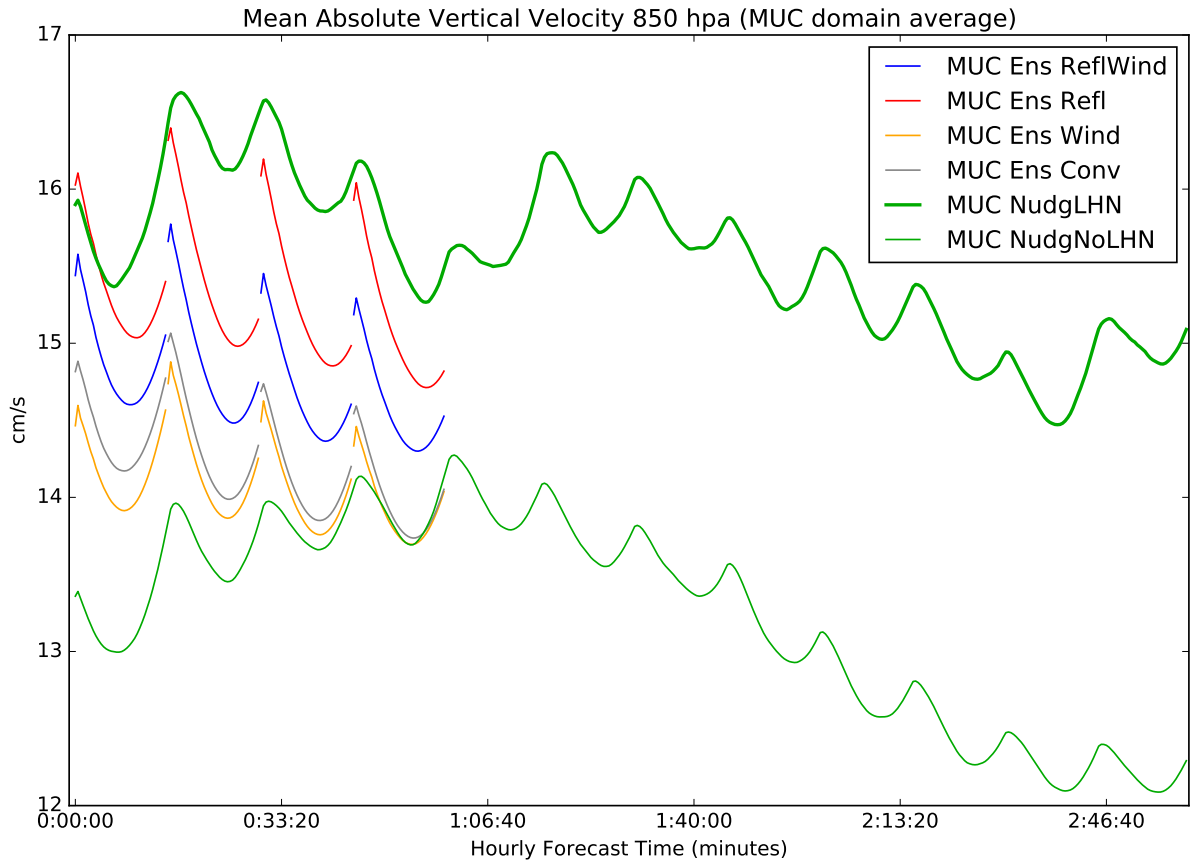

Figure 5.20: W_{abs} for MUC Experiments on the pressure level of 850 hPa. The ensemble (*MUC Ens*) experiments are averaged with respect to their 15-minute assimilation windows within the hourly analysis update of their boundary experiments (*DE Ens Refl*). The nudging (Nudg) experiments are averaged for every 3-hour window since they continuously assimilate observations.

5.4.3 Abundance of vertical motion

Figure 5.19 shows the surface pressure tendencies for the MUC experiments during the assimilation cycling. The nudging experiments exhibit the smallest deviations from hydrostatic balance, and the activation of LHN does not increase the signal strongly. For the ensemble experiments, the surface pressure tendencies increase when radial velocity or reflectivity is added to *MUC Ens Conv* It is remarkable that dp_s/dt of *MUC Ens Wind* and *MUC Ens Reflwind* are very similar, which indicates that the addition of wind observations to the reflectivity observations does not increase the imbalance in this particular diagnostic. After the 15 minute forecast window, the noise signal of all ensemble experiments has decayed towards the nudging experiments. This means that the assimilation interval was chosen in good accordance with the results of L32SOCG20 in Chapter 3.

Absolute amount of vertical motion

Figure 5.20 shows W_{abs} from Equ. (3.1) of Section 3.2 for all six MUC experiments. The 15-minute oscillation pattern is caused by the BCs which are updated every 15 minutes. The reference experiment is *MUC NudgNoLHN* which does not use any radar data inside the MUC domain. It has the lowest values of W_{abs}. Since dry noisy waves and deep convection are both contained in W_{abs}, the low values in *MUC NudgNoLHN* can be attributed solely to the absence of noisy forcing by LHN, since the amount of convection in *MUC NudgNoLHN* is very close to *MUC NudgLHN* on average (cf. Figure 5.14). The W_{abs} values of *MUC NudgLHN* are the largest of all experiments. In the intercomparison of the *MUC Ens* experiments, *MUC Ens Refl* has the largest noise level. When radial velocity is added (*MUC Ens ReflWind*), this noise is significantly damped. Likewise, the addition of radial velocity observations (*MUC Ens Wind*) to the assimilation of conventional data (*MUC Ens Conv*) has again a damping effect. This, together with the behavior in the surface pressure tendencies, partly answers the question posed in the introduction: Wind observations appear to cause less imbalance than rain observations.

In terms of the scaling relationship (5.9), the wind observations appear to have a correcting influence on the associated circulation. They are assimilated with a localization length scale of 16 km, which has a cut-off length of 30 km that is in the range of the estimated 40 km in the scaling relationship, but overlapping analyses may contain wind-patterns of even larger scale, opposed to the point-like updrafts that are observed by reflectivity. The mere forcing of small-scale updrafts due to LHN, whose scales are on the order of a few Δx of the model (Stephan et al., 2008), has a strong influence on the excitation of divergent patterns that necessarily underly the field of W_{abs}.

Vertical Motion in Masked Regions

The masked noise measures of Chapter 3 are evaluated here to further tackle the question which influence the observation types have on the dynamical balance, as posed in Section 5.1.

In the masked regions of the MUC domain, the variance of different model variables (horizontal wind U, temperature T, vertical velocity W, reflectivity $REFL$, horizontal divergence HOR_DIV, rain mixing ratio QR) is computed as for W in Equs. (3.4) and (3.5). Figure 5.21 shows $\text{var}(X)$ for different variables X in the storm-inner region during the cycling, and Figure 5.22 shows $\text{var}_{\text{ens}}(X)$.

$\text{var}(U)$ (Figure 5.21) is generally smaller when radial velocity is assimilated, which is expectable, since U is a direct component of the radial velocities for most directions. $\text{var}_{\text{ens}}(U)$ and $\text{var}_{\text{ens}}(T)$ (Figure 5.22) are largest when no radar data is assimilated, and smallest when all radar types are assimilated, which can be attributed to the number of observations that decrease the ensemble spread. $\text{var}(T)$ is similar for all experiments.

The results for $REFL$ are qualitatively similar to QR. When reflectivity is assimilated, $\text{var}(REFL)$ is largest and $\text{var}_{\text{ens}}(REFL)$ is smallest. Assimilating radial velocity in *MUC Ens Wind* decreases the spread $\text{var}_{\text{ens}}(REFL)$ with respect to *MUC Ens Conv*, which is caused by increments due to ensemble covariances between wind and hydrometeors (cf. Figure 5.1).

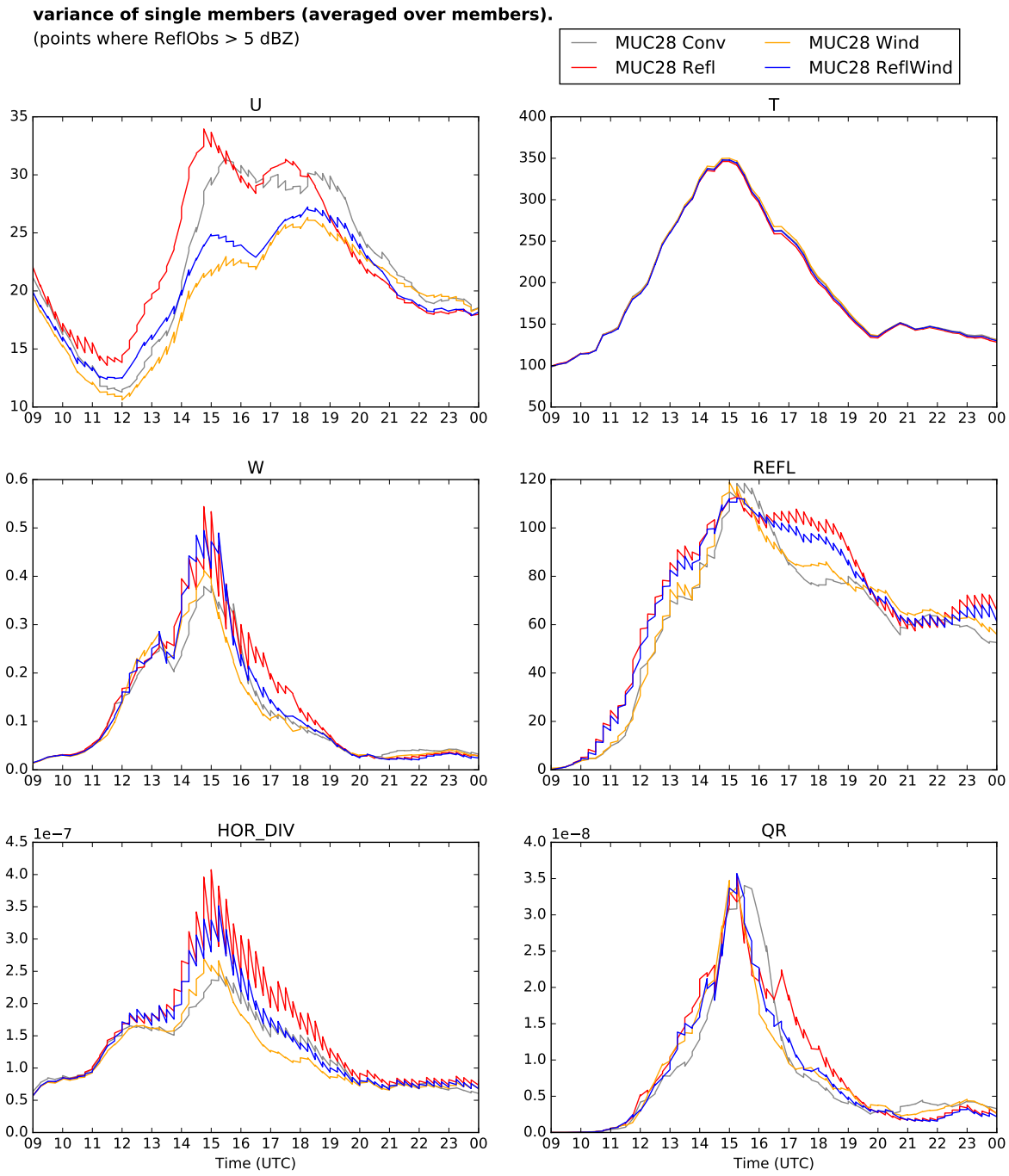

Figure 5.21: Variance of U, T, W, $REFL$, HOR_DIV, QR, for grid points inside an observed reflectivity larger than 5 dBZ for COSMO-MUC ensemble experiments (averaged over all ensemble members) during the cycling on 26.5.2014.

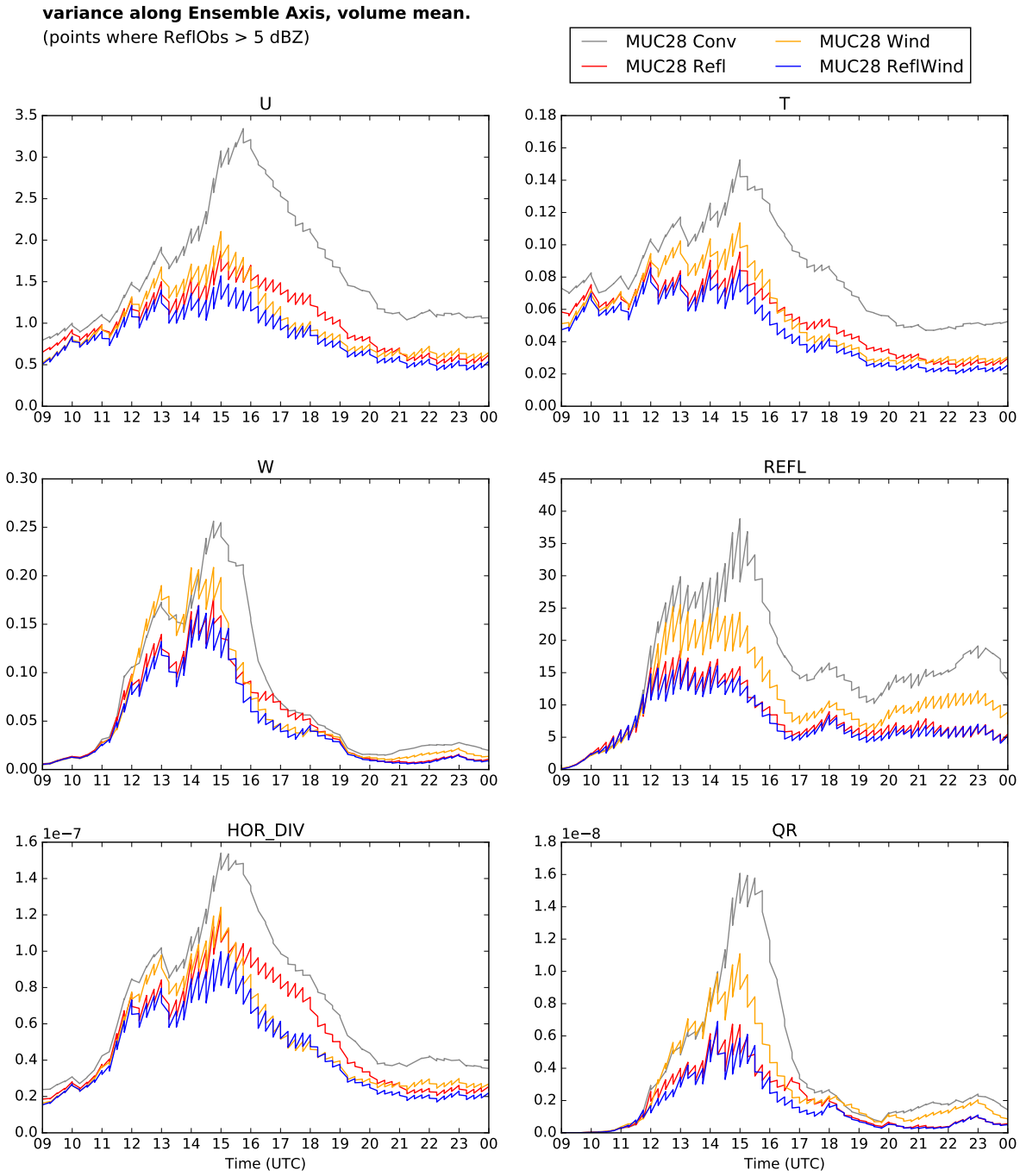

Figure 5.22: As Figure 5.21, now for mean pointwise ensemble variance.

The results for W are qualitatively similar to HOR_DIV. var(W) is largest and increased at every analysis step when reflectivity is assimilated. These increments are caused by ensemble covariances between hydrometeors and vertical velocity (cf. Figure 5.1). For *MUC Ens Wind*, var(HOR_DIV) is smallest which hints at the damping effect of radial winds that was already indicated in Figure 5.20. The increments of var(HOR_DIV) in *MUC Ens Wind* are smaller than in *MUC Ens Conv*, which is another sign that assimilating radial velocity has either a damping effect or does not overly excite vertical and divergent motion. This conclusion is supported by $\mathrm{var}_{\mathrm{ens}}(HOR\_DIV)$ which is smallest when radial velocity is added to reflectivity in *MUC Ens ReflWind*. For the vicinity space (not shown), the variance results of the ensemble experiments are similar in the same manner as for the inner space (Figures 5.21 and 5.22). For the space distant to the convection, no significant differences were visible between the experiments (not shown).

Differences between nudging and ensemble data assimilation

When comparing the nudging experiments with the ensemble DA experiments, it makes sense to first look at the unmasked, domain-wide mean values of model variables, as shown in Figure 5.23. U and T are not solely associated with convection. The mean value of U in the nudging experiments is larger than in the ensemble experiments, while the mean value of T is smaller. This already indicates that due to the differences in boundary conditions and data assimilation, the nudging experiments assume a slightly different meteorological situation than the ensemble experiments. Regarding $REFL$ and QR in Figure 5.23, the onset of convection is delayed with respect to the ensemble experiments and happens only slightly earlier when LHN is active. The analysis increments for $REFL$ and QR are generally positive in the ensemble experiments, especially between 9 and 17 UTC and even for *MUC Ens Conv*. This means that the LETKF is more effective in generating additional convection than LHN and nudging are.

The unmasked variances are shown in Figure 5.24. The main difference between nudging and ensemble DA is again the delayed onset of rain ($REFL$ and QR). The maximum variance of W in *MUC NudgLHN* is twice as large as for *MUC Ens ReflWind* and is delayed by 3 hours. This underlines the impression of Figure 5.20 that LHN introduces more vertical fluctuations in the dynamics – at least with the presently chosen DA settings.

5.5 Summary and conclusions

The aim of this Chapter was to investigate the influence which the assimilation of radar observations of reflectivity and radial wind have on the model state and the performance of QPF.

An ensemble version of nested local model COSMO-MUC was developed for this purpose, as well as an augmented cycling environment for the KENDA system. Experiments were performed with ensemble data assimilation of conventional observations and radar

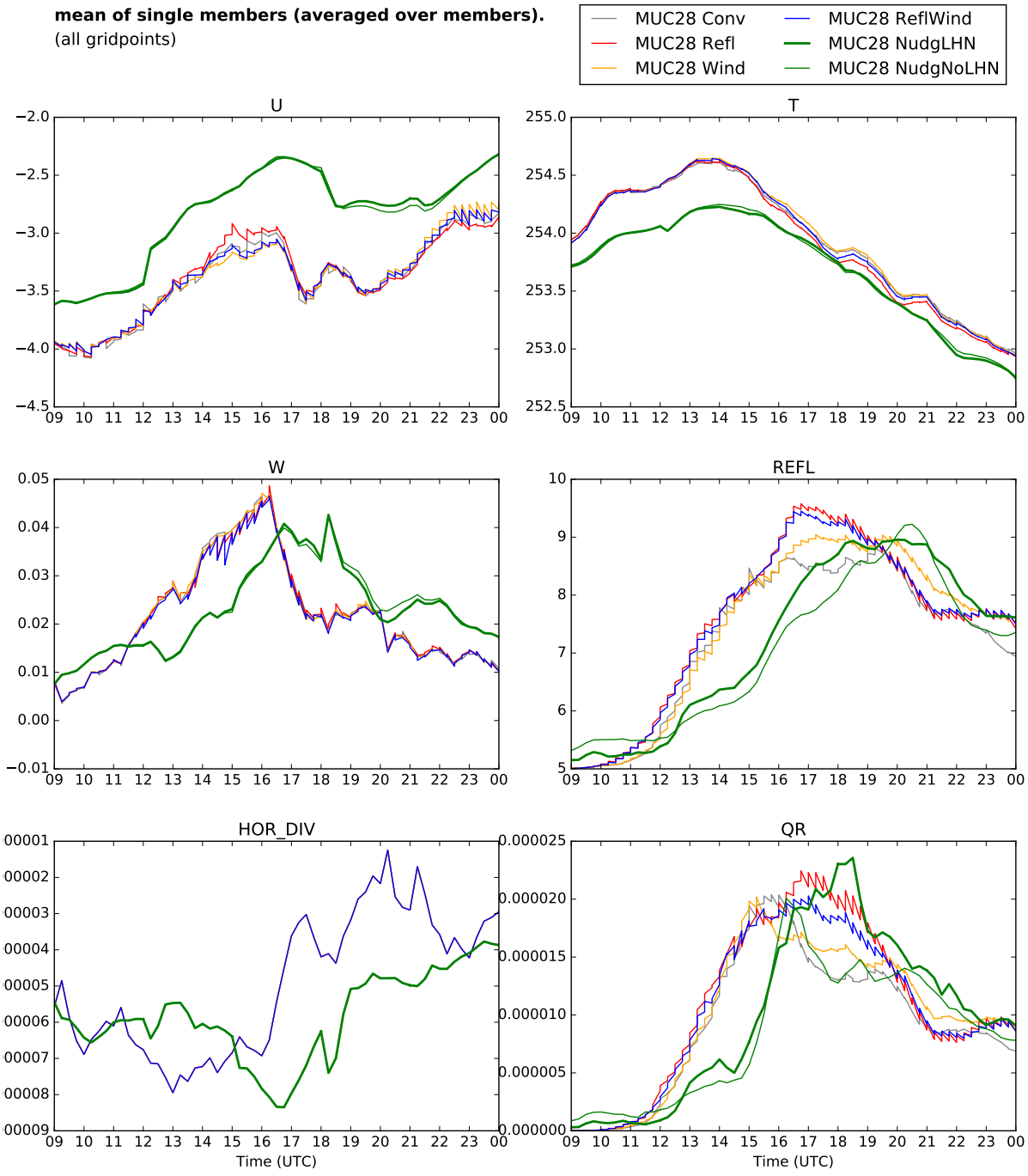

Figure 5.23: Domain mean values of U, T, W, $REFL$, HOR_DIV, QR for COSMO-MUC ensemble experiments (averaged over all ensemble members) and nudging experiments during the cycling on 26.5.2014.

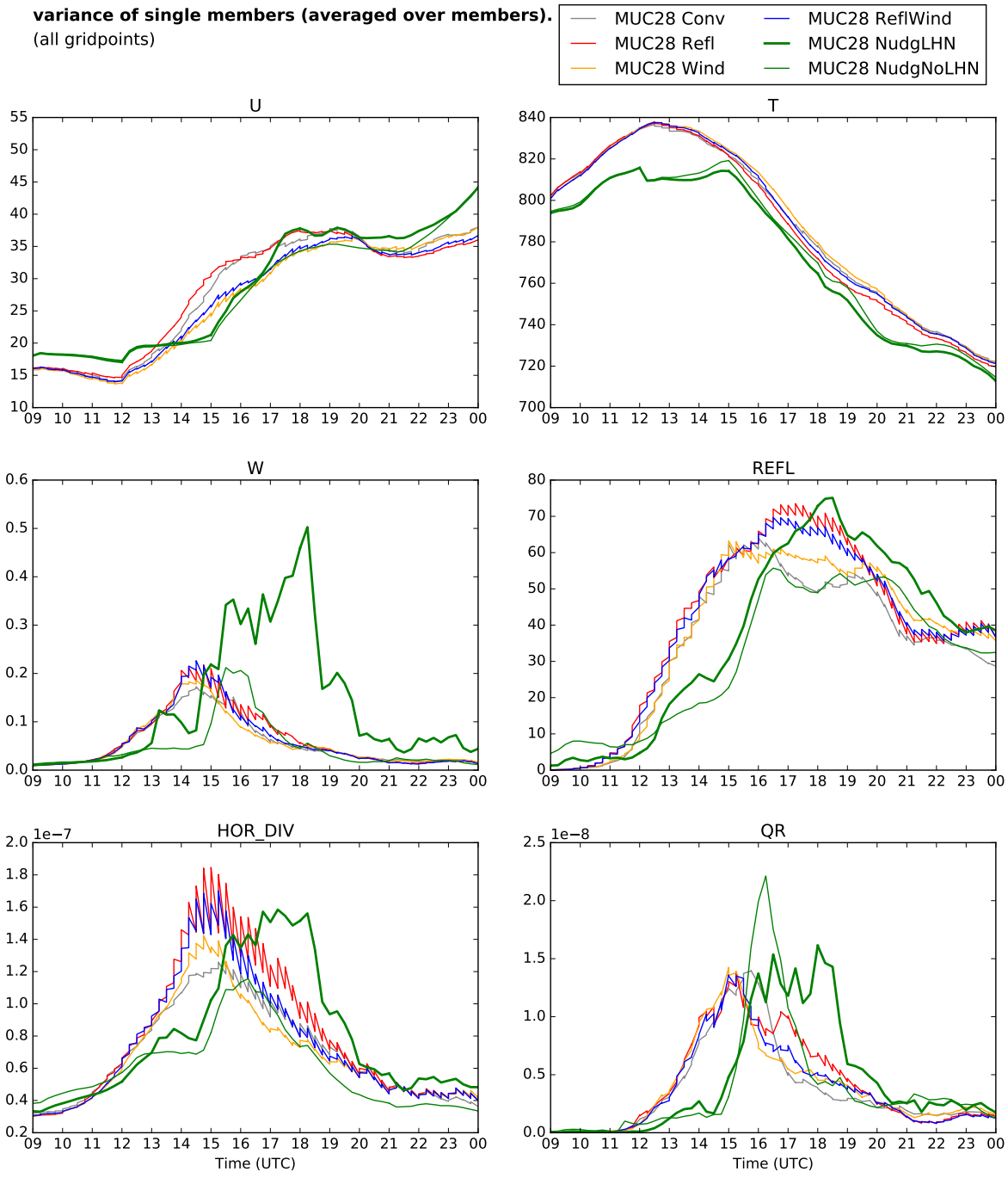

Figure 5.24: Variance of U,T,W,$REFL$,HOR_DIV, QR, for all grid points, for COSMO-MUC ensemble experiments (averaged over all ensemble members) and nudging experiments during the cycling on 26.5.2014.

data and with the deterministic nudging scheme together with LHN of radar-derived precipitation observations. The ensemble data assimilation was set up as a combination of the KENDA settings from Chapters 2 and 4. During the cycled data assimilation, the noise measures developed in Chapter 3 were applied and the reaction of the model on the different DA-schemes was assessed. One day in May 2014 was chosen as a multi-case study throughout the diurnal cycle of a stationary convective line along a low level convergence line. Five forecasts with a lead time of 3 hours were performed and assessed in terms of quantitative precipitation forecast skill and in terms of observation space statistics for reflectivity and radial wind.

The forecast error of the ensemble experiments was mostly smaller than the forecast error of the deterministic nudging experiments with LHN. This was especially dominant in a phase between 12 and 15 UTC where the convection in the domain was triggered and grew rapidly. Adding radar observations to the conventional observations proved helpful for the first 1 or 2 hours of forecast lead time in the ensemble experiments, which is in accordance to the findings of Chapter 2. In the same way, the prior application of LHN proved helpful for the forecast skill in the nudging experiments. The LHN caused oscillations of the vertical velocity and horizontal divergence that exceeded even the ensemble experiments that assimilated all radar data. These oscillations can be regarded as spurious gravity waves with dry *and* moist dynamics, since the absolute amount of convection was comparable to the nudging experiment that did not apply LHN. By applying EnKF-increments, the ensemble experiments experienced an introduction of non-hydrostatic noise which appeared as non-damaging to the model state consistency, as predicted by the experiment L32SOCG20 in Chapter 3. In the ensemble experiments, the additional assimilation of radial velocity observations had a damping effect on the intensity of vertical oscillations, compared to the assimilation of reflectivity only. This supports the conceptual model of observation type influence (cf. Section 5.1) wherein a precipitating storm core is coupled to a circulation of horizontal winds and vice versa. Still, the absolute quality of the recovered associated circulation could not be assessed by the means of this real data case in which the true state was unknown. Idealized experiments with convective cells that are quasi-stationary in space and time such as in Bretherton and Smolarkiewicz (1989) could bring insight on how effectively ensemble covariances of precipitating cores can recover the associated circulation, in comparison to the model response on a LHN forcing. An independent variation of the localization radius for the two observation types could bring insight on the scales of their spacial influence and correlation structure (Sobash and Stensrud, 2013).

With respect to the influence of the model noise on forecast errors, no conclusions could be drawn due to two reasons. On the one hand, the number of cases was too small for proper statistics. One the other hand, the true amount of convection and vertical oscillations was not known since it was not observed directly. Additional Observation System Simulation Experiment (OSSE) in an idealized framework or with a nature run for real cases as the present one could bring clarity.

A further limitation on the interpretation of the results is the unknown influence of the boundary conditions on the forecast skill and overall model state in the inner domain. Statistics of the mean model states showed that the deterministic nudging experiments

converged towards a slightly different meteorological situation than the ensemble experiments. This could already be a reason for the different behavior, e.g. the time of the onset of the main convection and its dynamical development, or the propagation direction. It could be helpful to rerun similar experiments with instances of COSMO-DE and ICON boundaries which use the ensemble mean as a BC for the nudging experiments, in order to have the deterministic state centered within the ensemble members. In the quantitative precipitation forecast, only an inner part of the domain should be evaluated that is far away enough from the model boundaries. For example, a frame of 150 km could be cut away in 3-hour forecasts for a typical advective wind speed of 50 km/h.

Chapter 6

Summary and Discussion

The aim of this thesis is to expose some of the difficulties that come with ensemble data assimilation of convection. These include the problem of convective predictability in dependency of the spatial scales that are updated by the DA, the quantification of gravity waves and spurious convective cells caused by the application of rigorous analysis schemes, the role of the convective environment and how it can be corrected with additional aircraft observation data, and the question which influence the Doppler radar observation types of wind and reflectivity have on the dynamical balance of the forecast model.

In Chapter 2, analysis schemes were devised that used a different amount of observation information when performing analysis updates of idealized convective cells. Observations were drawn from a reference nature run either on the model resolution of 2 km or on a coarse-grained resolution of 8 km, and they were used with a localization length scale that limited the spatial influence to distances of 8 km and 32 km, in temporal assimilation intervals of 5 and 20 minutes. It was hypothesized that within a lead time of a few hours, the quality of forecasts from very precise analyses with small errors would converge to the quality of forecasts from rather coarse analysis schemes with larger errors. The expected reason for this convergence was that small-scale errors grow faster than large-scale errors, until both saturate at a certain level. This expectation was met, and forecasts from a fine-scale analysis scheme only had a better performance for roughly 40 minutes of forecast, when compared to the coarse-scale analysis scheme. This means that there is indeed a limit on how precise an analysis needs to be, depending on the range of forecast lead time that is aimed at. In the simplest consequence, this means that the computational effort can be limited to reach only a certain accuracy, and the development of an operational DA system can focus on problems that are more pressing, such as model error and deficiencies of the spread of operational ensembles.

The main reason for the rapid decay in forecast quality described in Chapter 2 was identified to be the appearance of spurious convection in the forecast from the fine-scale analysis. The occurrence of spurious convective cells in an environment with considerable convective inhibition such as the Payerne sounding is a binary event and a major forecast error as such, compared e.g. to a forecast error that merely consists of a change in shape or position of a previously observed and analyzed storm. Following this consideration, Melhauser and Zhang (2012) developed a distinction between *practical* and

intrinsic predictability of convection. The practical predictability is related to errors in the initial conditions which can lead to a wrong forecast of a stable convective mode, but if these errors are reduced, the forecast catches the correct convective mode. This error could e.g. be a convergent wind pattern in the analysis state which has a scale of order 100 km and leads to the occurrence of a Mesoscale Convective System in the forecast – but neither the wind pattern occurred in reality, nor did the resulting storm. The intrinsic predictability is defined by the growth of errors that are at least a magnitude smaller than those related to the practical predictability and which may still lead to a bifurcation in the forecast (Lorenz, 1969). These errors could consist e.g. of a pattern of temperature-variability on the scale of the grid resolution and with an amplitude on the order of 0.1 K, which through chaotic error growth leads to the occurrence of a convective storm – a storm that does not appear in another experiment with a similar temperature perturbation field which only differs by its spacial pattern. Following this concept, the perturbations in the rigorous fine-scale scheme L8 that grew into spurious cells should be regarded as an artifact of the parameter choices in the data assimilation, thereby mainly affecting the *practical* predictability of the convective situation instead of the *intrinsic* predictability that had been targeted by the Chapter's hypothesis. In the presented convective situation, the intrinsic predictability should be assessed rather with perturbations of very small amplitude as was done in Selz and Craig (2015) and Zhang et al. (2016).

The DA settings of Chapter 2 could be criticized as too rigorous when compared to less strict spatio-temporal parameter sets of (pre-)operational EnKF systems that pursue to produce ensemble forecasts of severe convection by assimilating real observation data (e.g. the Warn-on-Forecast system described in Wheatley et al., 2015; Jones et al., 2016). The resulting noise which was characterized in this study could therefore be a phenomenon that rarely occurs in well-tuned operational systems and is otherwise seldom documented due to a generally success-driven tendency in scientific publishing. Still, it is possible that even well-tuned assimilation schemes disturb the forecast model in previously unaccounted meteorological situations, and it is therefore necessary to meticulously explore all corners of the parameter space in convective DA. Thus, the aim of Chapter 3 was to quantify the noise and dynamical imbalances caused by a rigorous assimilation scheme. As a first measure in the previous chapter, surface pressure tendencies indicated that a larger abundance of non-hydrostatic oscillations was present in the fine-scale scheme. In the next step, the coupling of cold pools and newly triggered convection was quantified by a newly developed spatio-temporal correlation algorithm, and it was shown that this particular coupling was disrupted by the occurrence of spurious convective cells that were uncoupled to the analyzed cold pools. While the coupling-properties will vary for different convective modes and the cold pools may not be as well-observed as the precipitation of new cells, the method of cold pool-coupling may still provide a quantitative guide line on whether the assimilation scheme produces dynamically reasonable model states of Mesoscale Convective Systems. As a second method, the abundance of vertical motion was quantified inside and outside of the storms by computing the variability and ensemble spread, thereby also providing a quantitative measure for the balance in the model (as done for an ocean model in Waters et al., 2016).

In possible follow-up studies, methods and DA-algorithms that may alleviate the

noise-production in convective DA should be assessed by the means of the quantifications that were successfully developed in Chapter 3. The following methods and ideas can be grouped into different categories. In the first group, considerations are made on how the dynamical properties of the forecast models can be respected when computing or post-processing the analysis. The algorithm of hydrostatic balancing of analysis increments (Rhodin et al., 2013) might reduce the noise that projects into surface pressure tendencies. The development of a transformation of variables into a hypothetical balanced space of convective updrafts (as done for hydrostatic global models by Kepert, 2009; Hamrud et al., 2015) could be a possible approach. General dynamical recipes against initialization noise are offered by the Digital Filter Initialization (DFI) scheme developed by Lynch et al. (1997), which reduces the effect of insertion shocks by filtering dynamical oscillations below a certain frequency, and the Incremental Analysis Update (IAU) technique (Bloom et al., 1996), which tries to avoid insertion shocks by distributing the analysis increments over a time-window, in contrast to the intermittent update approach of this study. Both techniques have proven successful for global models and also for mesoscale QPF forecasts (e.g. Peng and Zou, 2010), but it is not certain whether this approach is applicable for convective DA. In the second group of methods, statistical approaches are developed that intend to cure the problem of dynamical imbalance as a side-effect of the properties of the background and observation error distributions into account more consistently. It is possible that the noise level is sensitive to the degree of relaxation towards the background forecast in the analysis step (RTPP, Zhang et al., 2004), and a trade-off between precision and noise content of the analysis may exist. A combination of relaxation methods with the multi-pass Running-In-Place algorithm (Kalnay and Yang, 2010; Wang et al., 2013) could then increase the precision of the analyses while keeping the noise-level of individual update steps low. An inflated observation error covariance matrix, as done in Lange (2013), may alleviate detrimental effects in the analysis update that are caused by the non-Gaussianity of hydrometeor observations. A sophisticated observation error model, as presented in Harnisch et al. (2016) for satellite radiances, may provide an even better solution to account for the non-normal distribution of innovation statistics. Besides approaches that transform non-Gaussian probability density functions into a Gaussian form during the update step, the direct consideration of higher statistical moments (Hodyss, 2012) such as skewness and kurtosis could prove useful especially in the case of the non-Gaussian nature of clouds and convection. In the third group of methods, dynamical considerations are mixed with a statistical approach to constrain the DA solution to properties of the model state and dynamics. For example, the spatial structure of ensemble spread presented in Chapter 3 indicates that the ensemble contains spurious correlations on multiple scales. Since the imbalanced noise of the analysis updates is related to the covariance localization, it could be worth to investigate the influence of adaptive localization (as applied in Chapter 4) or more sophisticated methods like the empirical localization functions developed by Anderson and Lei (2013). Scale-dependent localization as proposed by Buehner and Shlyaeva (2015) may account for the different dynamical scales that are associated to the convection itself and to the surrounding environment. A simple increase in ensemble size will help to reduce sampling noise on all scales, or a simple spectral truncation could be applied to the background error covari-

ance matrix as a first step. Other dynamical-statistical approaches are to constrain the EnKF-solution with a variance limit (Gottwald, 2014), or to apply other algorithms that intend to conserve quantities like energy and enstrophy (Zeng and Janjić, 2016).

In Chapter 4, ensemble DA experiments were performed with a pre-operational version of KENDA for the COSMO-DE domain over Germany. It was discussed how the DA parameters steer and limit the influence of the additional data, and experience was gathered and documented that can help to better understand the effects of adaptive inflation and of adaptive localization. It was verified that the additional aircraft observations of the traffic control Mode-S are beneficial in the correction of the tropospheric profiles of wind and temperature. Together with humidity, they constitute the convective environment, and therefore an improvement of convective forecasts is expectable when new data sources are implemented analogously to the new developments of Chapter 4 (compare e.g. the effect of a large number of targeted dropsondes in Hitchcock et al., 2016). In further studies, a cross-validation of error reduction in the aircraft observation spaces and a resulting error reduction in the QPF skill of thunderstorms should be performed, either by data denial experiments or by using the observation impact method of Sommer and Weissmann (2014).

The coupling of model state properties inside convective storms and in the surrounding environment was investigated in Chapter 3 in terms of noisy oscillations. Chapter 5 expanded this view by investigating the different influence of radar reflectivity and radial wind observations on EnKF analyses and ensemble forecasts. A conceptual model was proposed which relates the influence of hydrometeor observations and the divergent circulation driven by the associated latent heating in terms of these two observations types. The setup used real observation data, as in Chapter 4, in the model domain of COSMO-MUC which is centered around the Munich airport and nested into COSMO-DE. The sub-cycling of multiple parallel experiments was implemented into KENDA for this purpose. Radar data was assimilated in a dynamically consistent way, based on the experiences gathered in Chapters 2 and 3, and 3-hour QPF were evaluated. During the assimilation, a lower noise level in the storms and their environment was visible when radial velocity was assimilated. It was concluded that the recovery of the convective circulation was dynamically more consistent due to the observations of the actual wind, as opposed to the recovery of the circulation by the means of ensemble covariances, and opposed to the dynamical response of the model on the heating source in the updrafts. The ensemble DA of radar data generally proved beneficial for the QPF skill and showed some advantage over the previously used deterministic LHN scheme. In future studies, more cases with different types of convective forcing (as in Craig et al., 2012) should be evaluated to obtain robust statistics about the benefit of radar data and the different subtypes of observations. Since the COSMO-MUC model is nested into COSMO-DE with a smaller domain, it could provide a good testing ground for the increase of model resolution. Horizontal resolutions down to the Large Eddy Simulation (LES) regime of order 100 m could be used to study the difference in model error and, especially important for DA purposes, the differences in model response to analysis increments from an EnKF system. The noise-measures of Chapter 3 could be applied to find proper parameters for radar data assimilation, such as localization lengths, covariance inflation and relaxation.

At the time of this publication, a master's thesis is being co-advised by the author, in which the dynamical properties of the spurious convection is investigated to further characterize the noise that is caused by the DA (see Chapter 3). Motivated by the work of Torri et al. (2015), dynamical properties of the convective triggering in the nature run are traced along the trajectories of Lagrangian particles. The same is done for the ensemble forecasts of the DA experiments, in order to characterize the conditions and mechanisms, such as perturbations of vertical velocity, pressure, and buoyancy, which lead to the arise of erroneous and spurious convection. A distinction between dynamically consistent and inconsistent triggering is sought, so that further measures can be developed to quantify the dynamical consistency of analyses, in dependency of the algorithm and the parameters of the DA. Further, an ongoing PhD project uses the technical framework of the idealized setting of Chapter 2 that was developed by the author, which consists of the idealized COSMO-setup with externally coupled observation operators and a proprietary cycling environment. In this project, the distinction of practical and intrinsic predictability is investigated in terms of the scales and amplitudes of DA-induced perturbations and other error sources. For these different perturbations, error doubling times and the resulting predictability limits are quantified, and situations of convective equilibrium are taken into consideration. Also, the influence of external constraints on the predictability is investigated, such as synoptic scale forcing or the influence of model orography, in order to assess the prospects of convective scale DA for the forecasting of severe and potentially hazardous convective phenomena.

Appendix A

List of Abbreviations

AMDAR	Aircraft Meteorological Data Relay
BC	Boundary Conditions
CAPE	Convectively Available Potential Energy
CIN	Convective Inhibition
COSMO	Consortium for Small-scale Modelling
DA	Data Assimilation
DAS	Displacement and Amplitude Score
DFI	Digital Filter Initialization
ECMWF	European Centre for Medium-Range Weather Forecasts
EnKF	Ensemble Kalman Filter
FSS	Fraction Skill Score
IAU	Incremental Analysis Update
IC	Initial Conditions
IFS	Integrated Forecasting System
IGW	Inertia-Gravity Wave
KENDA	Kilometre-scale Ensemble Data Assimilation
LETKF	Local Ensemble Transform Kalman Filter
LES	Large Eddy Simulation
LHN	Latent Heat Nudging

MCS	Mesoscale Convective System
NWP	Numerical Weather Prediction
OI	Optimal Interpolation
OSSE	Observation System Simulation Experiment
QPF	Quantitative Precipitation Forecast
SAL	Structure, Amplitude, Location
SWM	Shallow Water Model
WRF	Weather Research and Forecasting

Bibliography

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., and L. Lei, 2013: Empirical Localization of Observation Impact in Ensemble Kalman Filters. *Mon. Wea. Rev.*, **141**, 4140–4153, URL http://dx.doi.org/10.1175/MWR-D-12-00330.1.

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

Balaji, V., and T. L. Clark, 1988: Scale Selection in Locally Forced Convective Fields and the Initiation of Deep Cumulus. *J. Atmos. Sci.*, **45 (21)**, 3188–3211, URL http://dx.doi.org/10.1175/1520-0469(1988)045⟨3188:SSILFC⟩2.0.CO;2.

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. *Monthly Weather Review*, **139**, 3887–3905.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, URL http://dx.doi.org/10.1038/nature14956.

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.

Bick, T., and Coauthors, 2016: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quarterly Journal of the Royal Meteorological Society*, **142 (696)**, 1490–1504, URL http://dx.doi.org/10.1002/qj.2751.

Bierdel, L., P. Friederichs, and S. Bentzien, 2012: Spatial kinetic energy spectra in the convection-permitting limited-area NWP model COSMO-DE. *MetZ*, **21 (3)**, 245–258, URL http://dx.doi.org/10.1127/0941-2948/2012/0319.

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.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129**, 420–436.

Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorologische Zeitschrift*, **21**, 1–7.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Mon. Wea. Rev.*, **124 (6)**, 1256–1271, URL http://dx.doi.org/10.1175/1520-0493(1996)124⟨1256:DAUIAU⟩2.0.CO;2.

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

Bretherton, C. S., and P. K. Smolarkiewicz, 1989: Gravity waves, compensating subsidence and detrainment around cumulus clouds. *Journal of the Atmospheric Sciences*, **46**, 740–759.

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.

Buehner, M., and A. Shlyaeva, 2015: Scale-dependent background-error covariance localisation. *Tellus A*, **67 (0)**, URL http://www.tellusa.net/index.php/tellusa/article/view/28027.

Charney, J., M. Halem, and R. Jastrow, 1969: Use of Incomplete Historical Data to Infer the Present State of the Atmosphere. *Journal of the Atmospheric Sciences*, **26 (5)**, 1160–1163, URL http://dx.doi.org/10.1175/1520-0469(1969)026⟨1160:UOIHDT⟩2.0.CO;2.

Charney, J. G., R. Fjoertoft, and J. v. Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254.

Cintineo, R. M., and D. J. Stensrud, 2013: On the Predictability of Supercell Thunderstorm Evolution. *J. Atmos. Sci.*, **70 (7)**, 1993–2011, URL http://dx.doi.org/10.1175/JAS-D-12-0166.1.

Clark, T. L., T. Hauf, and J. P. Kuettner, 1986: Convectively forced internal gravity waves: Results from two-dimensional numerical experiments. *Quarterly Journal of the Royal Meteorological Society*, **112 (474)**, 899–925, URL http://dx.doi.org/10.1002/qj.49711247402.

Craig, G. C., and A. Dörnbrack, 2008: Entrainment in Cumulus Clouds: What Resolution is Cloud-Resolving? *J. Atmos. Sci.*, **65 (12)**, 3978–3988, URL http://dx.doi.org/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, URL http://dx.doi.org/10.1002/qj.929.

Craig, G. C., and M. Würsch, 2013: The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model. *Quarterly Journal of the Royal Meteorological Society*, **139**, 515–523.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press.

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.

de Haan, S., 2011: High-resolution wind and temperature observations from aircraft tracked by Mode-S air traffic control radar. *Journal of Geophysical Research: Atmospheres*, **116 (D10)**, URL http://dx.doi.org/10.1029/2010JD015264.

de Haan, S., and A. Stoffelen, 2012: Assimilation of High-Resolution Mode-S Wind and Temperature Observations in a Regional NWP Model for Nowcasting Applications. *Wea. Forecasting*, **27 (4)**, 918–937, URL http://dx.doi.org/10.1175/WAF-D-11-00088.1.

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 forecasts of 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.

Dudhia, J., 2014: A history of mesoscale model development. *Asia-Pacific Journal of Atmospheric Sciences*, **50 (1)**, 121–131, URL http://dx.doi.org/10.1007/s13143-014-0031-8.

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.

Fernandez, A., and Coauthors, 2015: An Ensemble Kalman Filter for global NWP based on the ICON model. *Monthly Weather Review*, **in preparation**.

Forster, C., A. Tafferner, and H.-D. Saffran, 2014: Gewittervorhersage. *Promet*, **39**, 45–54.

Gal-Chen, T., 1978: A method for the initialization of the anelastic equations: Implications for matching models with observations. *Mon. Wea. Rev.*, **106 (5)**, 587–606, URL http://dx.doi.org/10.1175/1520-0493(1978)106⟨0587:AMFTIO⟩2.0.CO;2.

Gal-Chen, T., and C. E. Hane, 1981: Retrieving buoyancy and pressure fluctuations from Doppler radar observations: A status report. *Atmospheric Technology*, **13**, 98–104, URL http://nldr.library.ucar.edu/repository/collections/ATD-000-000-000-034.

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, URL http://dx.doi.org/10.1002/qj.49712555417.

Ge, G., J. Gao, and M. Xue, 2012: Diagnostic pressure equation as a weak constraint in a storm-scale three-dimensional variational radar data assimilation system. *J. Atmos. Oceanic Technol.*, **29 (8)**, 1075–1092, URL http://dx.doi.org/10.1175/JTECH-D-11-00201.1.

Gill, A. E., 1982: *Atmosphere-ocean dynamics*. Academic Press (London).

Gottwald, G. A., 2014: Controlling balance in an ensemble Kalman filter. *Nonlinear Processes in Geophysics*, **21 (2)**, 417–426, URL http://www.nonlin-processes-geophys.net/21/417/2014/.

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, doi:10.1175/2010MWR3328.1.

Hagelin, S., L. Auger, P. Brovelli, and O. Dupont, 2014: Nowcasting with the AROME Model: First Results from the High-Resolution AROME Airport. *Wea. Forecasting*, **29 (4)**, 773–787, URL http://dx.doi.org/10.1175/WAF-D-13-00083.1.

Hamrud, M., M. Bonavita, and L. Isaksen, 2015: EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation. *Monthly Weather Review*, **143 (12)**, 4847–4864, URL http://dx.doi.org/10.1175/MWR-D-14-00333.1.

Hane, C. E., R. B. Wilhelmson, and T. Gal-Chen, 1981: Retrieval of thermodynamic variables within deep convective clouds: Experiments in three dimensions. *Mon. Wea. Rev.*, **109 (3)**, 564–576, URL http://dx.doi.org/10.1175/1520-0493(1981)109⟨0564:ROTVWD⟩2.0.CO;2.

Harnisch, F., and C. Keil, 2015: Initial conditions for convective-scale ensemble forecasting provided by ensemble data assimilation. *Mon. Wea. Rev.*, **143 (5)**, 1583–1600, URL http://dx.doi.org/10.1175/MWR-D-14-00209.1.

Harnisch, F., M. Weissmann, and . Periez, 2016: Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **142 (697)**, 1797–1808, URL http://dx.doi.org/10.1002/qj.2776.

Hauf, T., and T. L. Clark, 1989: Three-dimensional numerical experiments on convectively forced internal gravity waves. *Quarterly Journal of the Royal Meteorological Society*, **115 (486)**, 309–333, URL http://dx.doi.org/10.1002/qj.49711548606.

Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of MPEX Upsonde Observations on Ensemble Analyses and Forecasts of the 31 May 2013 Convective Event over Oklahoma. *Monthly Weather Review*, **144 (8)**, 2889–2913, URL http://dx.doi.org/10.1175/MWR-D-15-0344.1.

Hodyss, D., 2012: Accounting for skewness in ensemble data assimilation. *Monthly Weather Review*, **140 (7)**, 2346–2358, URL http://dx.doi.org/10.1175/MWR-D-11-00198.1.

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, URL http://dx.doi.org/10.1002/qj.2006.

Holton, J. R., 2004: *Dynamic Meteorology*. Elsevier Academic Press.

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

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3269–3289, URL http://dx.doi.org/10.1256/qj.05.135.

Houze, R. A., Jr, 1993: *Cloud Dynamics*. Academic Press.

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, URL http://dx.doi.org/10.1016/j.physd.2006.11.008.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan. Ser. II*, **75 (1B)**, 181–189.

Janjić, T., and S. E. Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, **134 (10)**, 2900–2915, URL http://dx.doi.org/10.1175/MWR3229.1.

Janjić, T., D. McLaughlin, S. E. Cohn, and M. Verlaan, 2014: Conservation of Mass and Preservation of Positivity with Ensemble-Type Kalman Filter Algorithms. *Monthly Weather Review*, **142 (2)**, 755–773, URL http://dx.doi.org/10.1175/MWR-D-13-00056.1, http://dx.doi.org/10.1175/MWR-D-13-00056.1.

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

Janjić, T., J. Schröter, R. Savcenko, W. Bosch, A. Albertella, R. Rummel, and O. Klatt, 2012: Impact of combining GRACE and GOCE gravity data on ocean circulation estimates. *Ocean Sci.*, **8**, 65–79, doi:10.5194/os-8-65-2012.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined Radar and Satellite Data Experiments. *Weather and Forecasting*, **31 (1)**, 297–327, URL http://dx.doi.org/10.1175/WAF-D-15-0107.1.

Kalnay, E., 2003: *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press.

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.

Kepert, J. D., 2009: Covariance localisation and balance in an Ensemble Kalman Filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (642)**, 1157–1176, URL http://dx.doi.org/10.1002/qj.443.

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., and Coauthors, 2007: Assessing Predictability with a Local Ensemble Kalman Filter. *Journal of the Atmospheric Sciences*, **64**, 1116–1140.

Lac, C., J.-P. Lafore, and J.-L. Redelsperger, 2002: Role of Gravity Waves in Triggering Deep Convection during TOGA COARE. *J. Atmos. Sci.*, **59 (8)**, 1293–1316, URL http://dx.doi.org/10.1175/1520-0469(2002)059⟨1293:ROGWIT⟩2.0.CO;2.

Lange, H., 2013: Data assimilation of severe convection using a Local Ensemble Kalman Filter. M.S. thesis, Ludwig-Maximilians-Universität München (LMU).

Lange, H., and G. C. Craig, 2014: The Impact of Data Assimilation Length Scales on Analysis and Prediction of Convective Storms. *Mon. Wea. Rev.*, **142 (10)**, 3781–3808, URL http://dx.doi.org/10.1175/MWR-D-13-00304.1.

Lange, H., and T. Janjić, 2016: Assimilation of Mode-S EHS Aircraft Observations in COSMO-KENDA. *Monthly Weather Review*, **144 (5)**, 1697–1711, URL http://dx.doi.org/10.1175/MWR-D-15-0112.1.

Leuenberger, D., 2005: High-resolution radar rainfall assimilation: Exploratory studies with latent heat nudging. Ph.D. thesis, Diss. ETH No. 15884.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (639)**, 523–533, URL http://dx.doi.org/10.1002/qj.371.

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

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

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112 (474)**, 1177–1194, URL http://dx.doi.org/10.1002/qj.49711247414.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Q.J.R. Meteorol. Soc.*, **129 (595)**, 3183–3203, URL http://dx.doi.org/10.1256/qj.02.132.

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciencies*, **20**, 130–141.

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., 2006: *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge Univ. Press.

Lynch, P., D. Giard, and V. Ivanovici, 1997: Improving the Efficiency of a Digital Filtering Scheme for Diabatic Initialization. *Mon. Wea. Rev.*, **125 (8)**, 1976–1982, URL http://dx.doi.org/10.1175/1520-0493(1997)125⟨1976:ITEOAD⟩2.0.CO;2.

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

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69 (11)**, 3350–3371, URL http://dx.doi.org/10.1175/JAS-D-11-0315.1.

Meng, Z., and F. Zhang, 2007: Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part ii: Imperfect model experiments. *Mon. Wea. Rev.*, **135 (4)**, 1403–1423, URL http://dx.doi.org/10.1175/MWR3352.1.

Neef, L. J., S. M. Polavarapu, and T. G. Shepherd, 2006: Four-Dimensional Data Assimilation and Balanced Dynamics. *J. Atmos. Sci.*, **63 (7)**, 1840–1858, URL http://dx.doi.org/10.1175/JAS3714.1.

Neef, L. J., S. M. Polavarapu, and T. G. Shepherd, 2009: A Low-Order Model Investigation of the Analysis of Gravity Waves in the Ensemble Kalman Filter. *J. Atmos. Sci.*, **66 (6)**, 1717–1734, URL http://dx.doi.org/10.1175/2008JAS2585.1.

Ott, E., and Coauthors, 2002: Exploiting Local Low Dimensionality of the Atmospheric Dynamics for Efficient Ensemble Kalman Filtering. *http://arxiv.org/abs/physics/0203058v3*, URL http://arxiv.org/abs/physics/0203058v3.

Peng, S., and X. Zou, 2010: Impact on quantitative precipitation forecasts of 4d-var rainfall data assimilation with a modified digital filter in favor of mesoscale gravity waves: A case study. *Journal of Geophysical Research: Atmospheres*, **115 (D23)**, URL http://dx.doi.org/10.1029/2010JD013993.

Perianez, A., H. Reich, and R. Potthast, 2014: Optimal localization for ensemble kalman filter systems. *Journal of the Meteorological Society of Japan*, **92**, 585–597.

Potvin, C. K., and L. J. Wicker, 2013: Correcting fast-mode pressure errors in storm-scale ensemble kalman filter analyses. *Advances in Meteorology*, **2013**, 14, URL http://dx.doi.org/10.1155/2013/624931.

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.

Rhodin, A., H. Lange, R. Potthast, and T. Janjić, 2013: *Documentation of the DWD Data Assimilation System*. Offenbach, Deutscher Wetterdienst.

Richardson, L. F., 1922: *Weather Prediction by Numerical Process*. Cambridge Univ. Press.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136 (1)**, 78–97, URL http://dx.doi.org/10.1175/2007MWR2123.1.

Roux, F., 1985: Retrieval of Thermodynamic Fields from Multiple-Doppler Radar Data Using the Equations of Motion and the Thermodynamic Equation. *Mon. Wea. Rev.*, **113 (12)**, 2142–2157, URL http://dx.doi.org/10.1175/1520-0493(1985)113⟨2142:ROTFFM⟩2.0.CO;2.

Salonen, K., H. Jäervinen, 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.

Sauer, M., and T. Gerz, 2014: Wettereinfluss auf Sicherheit und Effizienz im Luftverkehr - Analysen und Minimierungskonzepte. *Promet*, **38**, 157–165.

Scheck, L., T. Necker, M. Weissmann, and B. Mayer, 2016: An improved forward operator for visible satellite images. *Journal of Atmospheric and Oceanic Technology*, in preparation.

Scheufele, K., K. Kober, G. C. Craig, and C. Keil, 2014: Combining probabilistic precipitation forecasts from a nowcasting technique with a time-lagged ensemble. *Meteorological Applications*, **21 (2)**, 230–240, URL http://dx.doi.org/10.1002/met.1381.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, **142 (696)**, 1453–1472, URL http://dx.doi.org/10.1002/qj.2748.

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.

Selz, T., and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a summertime weather event over europe. *Monthly Weather Review*, **143 (3)**, 813–827, URL http://dx.doi.org/10.1175/MWR-D-14-00140.1.

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.

Snook, N., M. Xue, and Y. Jung, 2015: Multiscale EnKF Assimilation of Radar and Conventional Observations and Ensemble Forecasting for a Tornadic Mesoscale Convective System. *Monthly Weather Review*, **143 (4)**, 1035–1057, URL http://dx.doi.org/10.1175/MWR-D-13-00262.1.

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. *Monthly Weather Review*, **141 (11)**, 3691–3709, URL http://dx.doi.org/10.1175/MWR-D-12-00203.1.

Sobash, R. A., and D. J. Stensrud, 2015: Assimilating Surface Mesonet Observations with the EnKF to Improve Ensemble Forecasts of Convection Initiation on 29 May 2012. *Monthly Weather Review*, **143 (9)**, 3700–3725, URL http://dx.doi.org/10.1175/MWR-D-14-00126.1.

Soelch, I., 2015: Abschlussbericht COSMO-MUC Teilprojekt I. Tech. rep., DLR-PA.

Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **140 (685)**, 2672–2679, URL http://dx.doi.org/10.1002/qj.2343.

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, URL http://dx.doi.org/10.1038/310036a0.

Stensrud, D. J., and Coauthors, 2009: Convective-Scale Warn-on-Forecast System - A Vision for 2020. *Bulletin of the American Meteorological Society*, **90**, 1487–1499, URL http://dx.doi.org/10.1175/2009BAMS2795.1.

Stephan, C. C., M. J. Alexander, M. Hedlin, C. D. de Groot-Hedlin, and L. Hoffmann, 2016: A case study on the far-field properties of propagating tropospheric gravity waves. *Monthly Weather Review*, **144 (8)**, 2947–2961, URL http://dx.doi.org/10.1175/MWR-D-16-0054.1.

Stephan, K., S. Klink, and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model cosmo-de at dwd. *Quarterly Journal of the Royal Meteorological Society*, **134 (634)**, 1315–1326, URL http://dx.doi.org/10.1002/qj.269.

Strajnar, B., N. Zagar, and L. Berre, 2015: Impact of new aircraft observations Mode-S MRAR in a mesoscale NWP model. *Journal of Geophysical Research: Atmospheres*, **120 (9)**, 3920–3938, URL http://dx.doi.org/10.1002/2014JD022654, 2014JD022654.

Talagrand, O., 1981: Convergence of assimilation procedures. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Källen, Eds., Springer, 225–262.

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.

Torri, G., Z. Kuang, and Y. Tian, 2015: Mechanisms for convection triggering by cold pools. *Geophysical Research Letters*, **42**, 1943–1950, URL http://dx.doi.org/10.1002/2015GL063227.

Tulich, S. N., and B. E. Mapes, 2008: Multiscale convective wave disturbances in the tropics: Insights from a two-dimensional cloud-resolving model. *J. Atmos. Sci.*, **65 (1)**, 140–155, URL http://dx.doi.org/10.1175/2007JAS2353.1.

Wang, S., M. Xue, A. D. Schenkman, and J. Min, 2013: An iterative ensemble square root filter and tests with simulated radar data for storm-scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **139 (676)**, 1888–1903, URL http://dx.doi.org/10.1002/qj.2077.

Waters, J., M. J. Bell, M. J Martin, and D. J. Lea, 2016: Reducing ocean model imbalances in the equatorial region caused by data assimilation. *Quarterly Journal of the Royal Meteorological Society*, URL http://dx.doi.org/10.1002/qj.2912.

Weisman, M. L., and J. B. Klemp, 1982: The Dependence of Numerically Simulated Convective Storms on Vertical Wind Shear and Buoyancy. *Monthly Weather Review*, **110**, 504–520.

Weisman, M. L., and Coauthors, 2015: The Mesoscale Predictability Experiment (MPEX). *Bulletin of the American Meteorological Society*, **96 (12)**, 2127–2149, URL http://dx.doi.org/10.1175/BAMS-D-13-00281.1.

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.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar Data Experiments. *Weather and Forecasting*, **30 (6)**, 1795–1817, URL http://dx.doi.org/10.1175/WAF-D-15-0043.1.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, **140**, 3078–3089, doi: 10.1175/MWR-D-11-00276.1.

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, URL http://dx.doi.org/10.1175/2007MWR2018.1.

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

WMO, 2003: Aircraft Meteorological Data Relay (AMDAR) Reference Manual, WMO-958, 84 pp. Tech. rep., WMO. URL http://www.wmo.int/pages/prog/www/GOS/ABO/AMDAR/publications/AMDAR_Reference_Manual_2003.pdf.

Xu, Q., 2011: Measuring information content from observations for data assimilation: spectral formulations and their implications to observational data compression. *Tellus A*, **63 (4)**, 793–804, URL http://dx.doi.org/10.1111/j.1600-0870.2011.00524.x.

Yang, M.-J., and R. A. Houze, 1995: Multicell squall-line structure as a manifestation of vertically trapped gravity waves. *Mon. Wea. Rev.*, **123 (3)**, 641–661, URL http://dx.doi.org/10.1175/1520-0493(1995)123⟨0641:MSLSAA⟩2.0.CO;2.

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, URL http://dx.doi.org/10.1002/qj.353.

Ying, Y., and F. Zhang, 2015: An adaptive covariance relaxation method for ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **692**, 2898–2906, URL http://dx.doi.org/10.1002/qj.2576.

Yussouf, N., J. Gao, D. J. Stensrud, and G. Ge, 2013: The impact of mesoscale environmental uncertainty on the prediction of a tornadic supercell storm using ensemble data assimilation approach. *Advances in Meteorology*, **2013**, 15, URL http://dx.doi.org/10.1155/2013/624931.

Yussouf, N., J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 may 2013 oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Weather and Forecasting*, **31 (3)**, 957–983, URL http://dx.doi.org/10.1175/WAF-D-15-0160.1.

Zängl, G., D. Reinert, P. Rpodas, and M. Baldauf, 2014: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, **687**, 563–579, URL http://dx.doi.org/10.1002/qj.2378.

Zeng, Y., U. Blahak, and D. Jerger, 2016: An Efficient Modular Volume Scanning Radar Forward Operator for NWP-models: Description and coupling to the COSMO-model. *Quarterly Journal of the Royal Meteorological Society*, URL http://dx.doi.org/10.1002/qj.2904, qJ-16-0007.R2.

Zeng, Y., and T. Janjić, 2016: Study of conservation laws with the local ensemble transform kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **699**, 2359–2372, URL http://dx.doi.org/10.1002/qj.2829.

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, URL http://dx.doi.org/10.1175/JAS4028.1.

Zhang, F., Z. Meng, and A. Aksoy, 2006: Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part i: Perfect model experiments. *Mon. Wea. Rev.*, **134 (2)**, 722–736, URL http://dx.doi.org/10.1175/MWR3101.1.

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.

Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2016: Intrinsic predictability of the 20 may 2013 tornadic thunderstorm event in oklahoma at storm scales. *Monthly Weather Review*, **144 (4)**, 1273–1298, URL http://dx.doi.org/10.1175/MWR-D-15-0105.1.

Bedecke deinen Himmel, Zeus,
mit Wolkendunst!

Goethe, Prometheus

Acknowlegdments

First, I'd like to thank my supervisor George Craig, who has led me through my Bachelor thesis, my Master's thesis, and now also by supervising this Dissertation. Luckily, he never got tired of teaching (and preaching) his credo that the formulation of a clear hypothesis often leads to the creation of a comprehensible method that may sometimes even provide reproducible results. For me, who generally refrains from any idolization, working with him has illustrated what could be achieved by the glorious splendor of a focussed mind – if I had one. In this sense, I would like to thank George for sharing his place at the fountain of optimism and for keeping my mind on track when it was needed.

I'd like to deeply thank my co-advisor Tijana Janjić-Pfander for supporting me mathematically, methodically, and emotionally, and for providing an insider's perspective on the vast and multifaceted habitat that is academia. Thanks to Bernhard Mayer for co-examining this thesis – and also for provoking me into becoming a tolerable part-time runner.

I'd like to thank the employees and students of the Meteorological Institute for continuously creating an atmosphere of pleasant social envolvement, a productive handling of conflicts, and a relaxing background noise of dark humor.

Special thanks go to Yvonne Ruckstuhl for the fast and meticulous proofreading.

I'd like to thank my companions of *Die Rationalversammlung* for supplying the necessary artistic balance on the second rail of my life's interests – especially Philipp *Scharri* Scharrenberg and Dr. Christian *Bumillo* Bumeder for giving me the opportunity to deeply commit myself to science during the phase that led to this Dissertation.

I would like to thank Andreas Rhodin, Christoph Schraff, Ulrich Blahak and Hendrik Reich from DWD Offenbach for the technical support of the COSMO-Model, the LETKF and the BaCy-Environment of KENDA. I'd like to thank Ana Fernandez at DWD for the ICON Boundary conditions, and Siebren de Haan at KNMI for the Mode-S EHS aircraft dataset. Thanks to Ingo Soelch from DLR Oberpfaffenhofen for the basic setup of the deterministic COSMO-MUC. Thanks to Julia Mack at MIM for the code of the autocorrelation method and to Tobias Selz for the code of the spectral analysis.

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). The author was funded by the BMVI Forschungsvorhaben Nr. 50.0357-2013-L1.