

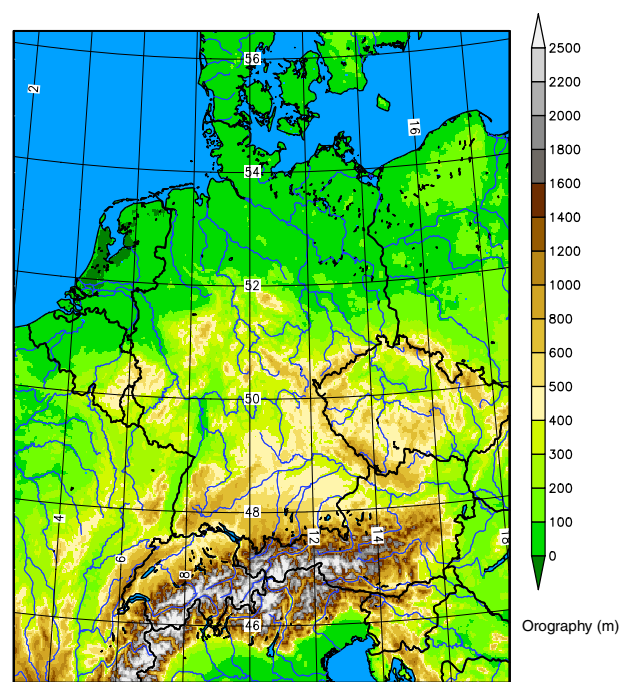


*et al.*, 2011; Dow and Macpherson, 2013). The most widely used schemes for this purpose are based on 3D-Var. For the convective scale, however, there are a number of arguments suggesting that the EnKF should be employed instead of conventional 3D-Var. The strongest argument is the explicit determination of the background-error covariance by the ensemble in the EnKF, which is highly flow-dependent at the convective scale. Its description should, for example, be different for cloudy and cloud-free areas. Moreover, generally all the variables simulated by a numerical weather prediction (NWP) model can be updated, whereas in variational schemes (VAR) for NWP there must be an explicit linear and adjoint relation to the control vector of prognostic variables. In principle, adding cloud variables, for example, to the control vector is possible in VAR; however, the definition of their background error and the coupling with temperature and moisture background errors is challenging. Some of these aspects are partly addressed in extensions of conventional 3D-Var. For example, some flow-dependence of the background errors is obtained implicitly in 4D-Var. Some explicit flow-dependence can be introduced using background-error variances (e.g. Berre and Desroziers, 2010) and covariances (Bonavita *et al.*, 2014) from an ensemble of perturbed variational assimilations, or by using hybrid EnKF–Var techniques (e.g. Buehner, 2005; Liu *et al.*, 2008; Wang *et al.*, 2008; Clayton *et al.*, 2013; Gustafsson and Bojarova, 2014, some of which do not require linear and adjoint operators). So far, however, these approaches have mostly been applied to larger scale NWP.

There are a rapidly growing number of studies with (up to now experimental) applications of the EnKF to the convective scale, including storm prediction. For example, Dowell *et al.* (2011) produced realistic storm-scale structures using an EnKF for a cloud model with 1 km horizontal grid spacing and obtained an improved fit to independent reflectivities when they assimilated radar radial velocities and/or radar reflectivities. Aksoy *et al.* (2009) employed the EnKF for the assimilation of radar observations at the convective scale with a 2 km resolution Weather Research and Forecasting (WRF) model version.

In the context of the COnsortium for Small-scale MOdelling (COSMO)\* priority project Kilometre-scale ENsemble Data Assimilation (KENDA), an LETKF system following Hunt *et al.* (2007) has been developed at the German Weather Service (Deutscher Wetterdienst, DWD) with a view to future operational use for convection-permitting configurations of the limited-area model COSMO. At the convective scale, a probabilistic approach is not only appropriate for data assimilation by means of EnKF, as pointed out above, but is also advisable for the forecasting problem, due to the stochastic nature of convection. At DWD, an ensemble prediction system (EPS), named COSMO-DE-EPS (Gebhardt *et al.*, 2011), already runs operationally, but without its own data assimilation cycle. However, a deterministic forecast is still required for various purposes and the first aim of KENDA, related to the focus of the present article, is to replace the current operational nudging-based data assimilation scheme of the COSMO model for the deterministic forecasts, using a similar observation set. The data types used operationally in the observation nudging scheme are conventional data (radiosonde, aircraft, wind profiler, surface station data) and radar-derived surface precipitation rates are assimilated with a latent heat nudging approach.

The goal of the present work is to describe the basic implementation of the KENDA system and its main components and some of their effects, as this forms the basis for a variety of further studies. We also aim to give an indication of the potential of the LETKF at the convective scale for operational NWP by comparing its performance for a limited test period with a nudging-based system that has been run operationally with



**Figure 1.** The orography of the operational COSMO-DE domain used for the numerical tests of the KENDA–LETKF ensemble data assimilation system with 2.8 km horizontal resolution. The domain size is about 1170 km × 1280 km.

success for many years. For this purpose, latent heat nudging (LHN) is included in the LETKF data assimilation cycle to allow for the use of radar-derived precipitation rates. To our knowledge, it is the first time that this combination has been tested and that an LETKF system has been developed for operational NWP at the convective scale. It should be noted that Lei *et al.* (2012) have already combined EnKF with nudging; however, in contrast to our approach, their scheme was a truly hybrid technique based on EnKF and observation nudging (rather than latent heat nudging), in which the two components were used to optimize the assimilation of only one data set (conventional *in situ* data).

The article is organized as follows. In section 2 we describe the basic design and set-up, as well as the algorithm for KENDA. This includes brief descriptions of the COSMO model and its current operational observation nudging and LHN assimilation techniques. The LETKF algorithm has been extended by adding a range of tools and components as described in section 3. Section 4 describes the basic experimental set-up and section 5 presents the results. Conclusions follow in section 6.

## 2. The forecast model and assimilation methods

### 2.1. The COSMO model

The COSMO model is a limited-area numerical weather prediction system, which is also deployed for regional climate simulations and used at many universities and research institutes in research mode. It is developed and maintained by the national weather services of the COSMO consortium. These countries run the model for daily operational weather forecasting in various configurations and at various resolutions.

Here, we describe the convective-scale configuration COSMO-DE (Baldauf *et al.*, 2011), which is run operationally at the German Weather Service (DWD) and used for all the experiments in the present study.

COSMO-DE has a 2.8 km horizontal grid spacing with a model domain covering Germany and parts of the neighbouring countries (Figure 1). It employs 50 hybrid vertical layers, which are terrain-following in the lower levels and horizontally flat in the upper levels. The model top is at a height of 22 km (i.e. about 40 hPa). In the operational set-up at DWD, hourly lateral boundary conditions are provided by a 7 km version of COSMO

\*COSMO stands for COnsortium for Small-scale MOdelling in Numerical Weather Prediction, a collaboration between the national weather services of Germany, Greece, Italy, Poland, Romania, Russia and Switzerland.

run at DWD (COSMO-EU), which in turn was driven by the global model GME (Majewski *et al.*, 2002) until January 2015 and by the new operational global model ICON (ICOSahedral Nonhydrostatic; Zängl *et al.*, 2015) after this date. In both COSMO configurations, the fields are also relaxed towards the respective driving model in the interior of the model domain from the model top down to a height of 11 km (about 235 hPa).

The atmospheric prognostic variables are the three-dimensional wind vector, temperature, pressure (deviation from a reference state), turbulent kinetic energy (TKE) and specific contents of water vapour, cloud water, cloud ice, rain, snow and graupel. The equations for the dynamic variables are solved using the time-splitting approach of Wicker and Skamarock (2002). Deep convection is computed explicitly, while shallow convection is parametrized based on a simplified version of the mass-flux convection scheme by Tiedtke (1989). A Lin-type cloud microphysics scheme is deployed for cloud formation and decay (Lin *et al.*, 1983). It contains a simplified version of the parametrization of Seifert and Beheng (2001) for autoconversion, accretion and self-collection, which has been reduced to a one-moment scheme assuming constant cloud droplet concentration and shape parameter. The turbulence parametrization is based on the prognostic TKE equation, i.e. a closure of order 2.5 (Raschendorfer, 2001). Heating rates due to radiative effects are provided by the radiation scheme of Ritter and Geleyn (1992), which is a  $\delta$ -two-stream scheme. This is updated every 15 min on an averaged  $2 \times 2$  grid points grid. A multi-layer soil model (TERRA: Doms *et al.*, 2011) provides the lower boundary condition at the ground.

Also included in the code of the COSMO model is the current operational nudging-based assimilation scheme combined with latent heat nudging of radar precipitation; see sections 2.2 and 2.3.

## 2.2. Observation nudging

In a *nudging* or *Newtonian relaxation* approach, the prognostic variables of a model are relaxed towards prescribed values within a predetermined time window. If the prescribed values are observations, the technique is called observation nudging. The method has a long history (see Anthes, 1974; Davies and Turner, 1977; Stauffer and Seaman, 1990, 1994; Schraff, 1997; Lei *et al.*, 2012) and has been successfully employed in operational convective-scale NWP models. A relaxation term is used in the model equations, such that the tendency for the prognostic variable  $\psi$  is given by

$$\frac{\partial}{\partial t} \psi(r, t) = F(\psi, r, t) + G_y \sum_j W_j(r, t) \cdot [y_j - \psi(r_j, t)]. \quad (1)$$

Here,  $F$  stands for the model dynamics and physical parametrization,  $y_j$  is the value of the  $j$ th observation influencing the grid point  $r$  at time  $t$ ,  $r_j$  is the observation location,  $G_y$  a constant denoted the *nudging coefficient* and  $W_j$  is an observation-dependent weight. Neglecting  $F$ , Eq. (1) describes an exponential relaxation of the prognostic variable state  $\psi(r, t)$  towards the observation, modulated by its observation weight, and  $1/G_y$  is the e-folding decay time, which is chosen to be typically about half an hour in COSMO. For more details we refer to Schraff and Hess (2012).

## 2.3. Latent heat nudging

Latent heat nudging (LHN) is deployed operationally for the assimilation of radar-derived precipitation rates. It is based on the assumption that the precipitation rate near the surface is approximately proportional to the release of latent heat at any

time. A temperature increment is calculated according to

$$\Delta T_{\text{LHN}} = (\alpha - 1) \Delta T_{\text{LH,mod}}, \quad (2)$$

with  $\alpha = RR_{\text{obs}}/RR_{\text{ref}}$ . This scaling of the modelled latent heat release  $\Delta T_{\text{LH,mod}}$  to obtain the temperature increment  $\Delta T_{\text{LHN}}$  is applied only at model levels where  $\Delta T_{\text{LH,mod}} > 0$ . The scaling factor depends on the ratio  $\alpha$  of the observed ( $RR_{\text{obs}}$ ) over the modelled precipitation rate ( $RR_{\text{ref}}$ ); for details we refer to Stephan *et al.* (2008). The increment is positive when the model underestimates the observed value and negative when the model overestimates precipitation. The increment changes the dynamical state of the model. Adding positive increments leads to updraught motion, which often yields enhanced condensation and subsequent precipitation.

In addition to the temperature increment, an increment in specific humidity is applied so as to maintain the relative humidity. Otherwise, adding a positive temperature increment would result immediately in a decrease of cloud liquid water because a saturation adjustment would convert cloud liquid water into water vapour in the case of undersaturation.

The frequency of the radar data used for LHN is 5 min. To apply the LHN at every time step (i.e. 25 s), the radar data are linearly interpolated in time, between two consecutive observations. Apart from a two-grid-point filter (a simple two-dimensional averaging operator) applied to the precipitation fields given on the model grid, the LHN is applied horizontally locally at the grid-point scale.

## 2.4. LETKF

The KENDA implementation follows the LETKF as described by Hunt *et al.* (2007). It is an efficient version of the EnKF and employs a square-root filter to calculate the analysis ensemble from the background ensemble.

In this method, the background-ensemble perturbations are used to estimate the background-error covariance  $\mathbf{P}^b$ :

$$\mathbf{P}^b = (k - 1) \mathbf{X}^b (\mathbf{X}^b)^T, \quad (3)$$

where the  $k$  columns of the matrix  $\mathbf{X}^b$  consist of the background-ensemble perturbations  $\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b$ , i.e. the deviations of the ensemble members  $\mathbf{x}^{b(i)}$  from the ensemble mean  $\bar{\mathbf{x}}^b$ , and  $k$  is the ensemble size. The well-known quadratic cost function in physical space can then be transformed into a cost function for a vector  $\mathbf{w}$  of dimension  $k$  in the ensemble space:

$$J(\mathbf{w}) = (k - 1)(\mathbf{w}^T \mathbf{w}) + [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b \mathbf{w}]^T \mathbf{R}^{-1} [\mathbf{y}^o - \bar{\mathbf{y}}^b - \mathbf{Y}^b \mathbf{w}]. \quad (4)$$

Here,  $\mathbf{R}$  is the observation-error covariance matrix,  $\mathbf{y}^o$  the observation vector and the linear approximation

$$H(\bar{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}) \approx \bar{\mathbf{y}}^b + \mathbf{Y}^b \mathbf{w} \quad (5)$$

is applied.  $\mathbf{Y}^b$  is the ensemble perturbation matrix in observation space, i.e. its  $k$  columns are the ensemble of deviations of the background observation vectors  $\mathbf{y}^{b(i)} = H(\mathbf{x}^{b(i)})$  from their mean  $\bar{\mathbf{y}}^b = \overline{H(\mathbf{x}^b)} = k^{-1} \sum_i H(\mathbf{x}^{b(i)})$ . Since the cost function (4) is formulated in the low-dimensional ensemble space, its minimum can be computed explicitly by

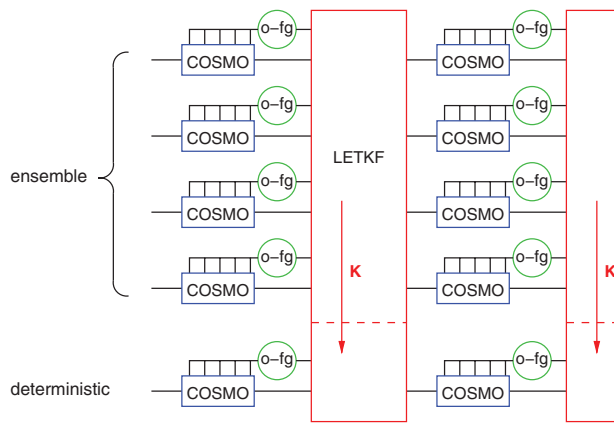
$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - \overline{H(\mathbf{x}^b)}), \quad (6)$$

$$\tilde{\mathbf{P}}^a = [(k - 1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1}, \quad (7)$$

where  $\bar{\mathbf{w}}^a$  is the analysis, i.e. analysis mean, in ensemble space, and  $\tilde{\mathbf{P}}^a$  the analysis-error estimation in ensemble space. The analysis ensemble perturbations are obtained from

$$\mathbf{W}^a = [(k - 1)\tilde{\mathbf{P}}^a]^{1/2}. \quad (8)$$





**Figure 2.** KENDA-LETKF system set-up; ‘o-fg’ denotes observation minus first guess, ‘K’ the Kalman gain for the analysis mean.

The background-ensemble perturbations matrix  $\mathbf{X}^b$  has been used to transform the analysis problem into ensemble space for Eq. (4) and is therefore used to transform the solution in ensemble space (Eqs (6), (8)) back to physical space. The analysis ensemble members are then given by

$$\mathbf{x}^{a(i)} = \bar{\mathbf{x}}^b + \mathbf{X}^b(\bar{\mathbf{w}}^a + \mathbf{W}^{a(i)}), \quad (9)$$

where  $\mathbf{W}^{a(i)}$  is the  $i$ th column of  $\mathbf{W}^a$ . Thus, each analysis ensemble member is given by a linear combination of the background ensemble members. The analysis mean in physical space does not need to be computed explicitly in the LETKF data assimilation cycle. It is given by Eq. (9) without the term  $\mathbf{W}^{a(i)}$ . Using this and Eq. (6), the so-called Kalman gain  $\mathbf{K}$  can be expressed for the ensemble mean by  $\mathbf{K} = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1}$ , where  $\mathbf{K}$  is defined to determine the analysis increments given the innovations ( $\mathbf{y}^o - \overline{H(\mathbf{x}^b)}$ ).

The solution in ensemble space given by Eqs (6) and (8) can be computed on a coarser resolution *analysis grid*. It is then interpolated to the model grid before being used to compute the solution in model space by Eq. (9). For more details of this implementation, we refer to Hunt *et al.* (2007) and Szunyogh *et al.* (2008).

It is important to note the following properties of the above formulation of the LETKF. Firstly, if observations are available and used not only at analysis time but over a certain period up to the time of the next analysis step, then the LETKF is a truly four-dimensional data assimilation scheme. This applies not just in the sense that the innovations ( $\mathbf{y}^o - \overline{H(\mathbf{x}^b, t^o)}$ ) can be computed at appropriate time (i.e. observation time  $t^o$ ). Rather, the 4D-LETKF is based on the four-dimensional background-error covariances evaluated in observation space and projected on to the ensemble subspace,  $(k-1)^{1/2} \mathbf{Y}^b (\mathbf{Y}^b)^T$ . ( $\mathbf{Y}^b$  takes a similar role to  $\mathbf{HMB}^{1/2}$  in 4D-Var, where  $\mathbf{H}$  is the linearized observation operator,  $\mathbf{M}$  the tangent-linear forward model operator and  $\mathbf{B}$  the background-error covariance at analysis time.) In other words, the ensemble forecast trajectories over the assimilation window provide an estimate of the 4D background-error covariances at the times and locations of the observations. In the KENDA implementation, the 4D capabilities are obtained by applying the observation operators  $H(\mathbf{x}^{b(i)}, t^o)$  within the COSMO model during its forward integration (see Figure 2). This allows for the operators to be evaluated at the exact observation times  $t^o$ . For the data types used commonly in the LETKF and the observation nudging, the same operators are deployed for quality control and also for the computation of those innovations that require the forward operator. (In the nudging scheme, a retrieval operator is applied for some observation types and levels, e.g. surface-level data, for which the observation increment is required at the model orography rather than at the station height.)

Secondly, as seen in Eqs (6)–(9), the formulation makes explicit use of only the full nonlinear observation operator; the linearized

and adjoint operators are not required. This allows enhanced freedom in the formulation of the observation operators. For instance, the operator used by Schomburg *et al.* (2015) for the assimilation of cloud-top height would be difficult to invert. (It has to be seen how well such strongly nonlinear operators work in practice, considering that an implicit linearization is imposed by Eq. (5).)

Thirdly, as pointed out by Hunt *et al.* (2007), the choice to derive the analysis ensemble perturbations by taking the square root of the analysis-error covariance in Eq. (8) is the solution that minimizes the (mean-square) distance between  $\mathbf{W}^a$  and the identity. Thus, in this metric, the analysis ensemble perturbations are as close as possible to the background-ensemble perturbations subject to the constraint on their sample covariance. It also guarantees that  $\mathbf{W}^a$  depends continuously on  $\tilde{\mathbf{P}}^a$ . This is crucial when localization (see Houtekamer and Mitchell, 1998; Hamill *et al.*, 2001) is applied, because it ensures that at neighbouring grid points with slightly different matrices  $\tilde{\mathbf{P}}^a$ , similar linear combinations ( $\bar{\mathbf{w}}^a + \mathbf{W}^{a(i)}$ ) are obtained to compute the analysis ensemble members. Otherwise, the analysis members could be very unbalanced after localization.

Localization is achieved by calculating the weight matrix  $\mathbf{w}^a$  in Eq. (6) independently at each grid point of the *analysis grid*. This grid can be the model grid itself, or a coarser grid from which  $\mathbf{w}^a$  is interpolated to the model grid afterwards. At each analysis grid point, the inverse observation-error covariance matrix  $\mathbf{R}^{-1}$  is scaled according to the distance of the observations from the grid point. The scaling makes use of a *Gaspari–Cohn* function (which is similar to a Gaussian but has compact support) with specified horizontal and vertical localization length-scale (Gaspari and Cohn, 1999). Prior to the scaling of  $\mathbf{R}^{-1}$ , observations outside the area given by the compact support of the function are discarded. In the experiments of this study, the vertical localization scale increases with increasing height in the range 0.075–0.5 in terms of the logarithm of pressure, while the horizontal scale is determined adaptively (see section 3.5).

Finally, as an initial condition for a deterministic forecast, the analysis ensemble mean could be deployed. However, in case of non-Gaussian distributions of the ensemble members, their mean will in general not be in balance. Cloud-related variables, which are of particular interest for very short-range forecasts in the convective scale, often have smoothed fields in the ensemble mean. Therefore, a different approach is adopted in KENDA (see Figure 2). The analysis  $\mathbf{x}^a$  for a deterministic data assimilation and forecast cycle is determined by applying the Kalman gain matrix for the ensemble mean  $\mathbf{K} = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1}$  to the innovations of the unperturbed deterministic (or control) run:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{L} \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - H(\mathbf{x}^b)). \quad (10)$$

It is clear from Eq. (10) that the deterministic run must use exactly the same set of observations as the ensemble in the LETKF. Optionally, the grid resolution can be higher than that of the ensemble and, in this case, the analysis increments on the ensemble grid finally have to be interpolated to the fine grid of the deterministic run by an interpolation  $L$ .

The rationale of using the gain of the ensemble mean is that both the ensemble mean and the deterministic analysis aim to provide an unperturbed ‘best’ estimate of the true state. This allows the deterministic analysis to take full advantage of the flow-dependent ensemble background covariances. The deterministic analysis will not be optimal if its background deviates significantly from the background ensemble mean. This is because the background-ensemble perturbations of the LETKF would not reflect the background errors of the deterministic run in such a case.

Initializing the deterministic forecast with the analysis ensemble mean would be an alternative approach (we call this ‘forecast on mean’ hereafter). However, preliminary tests over two weeks showed increased spin-up effects on precipitation during the first 2 h of these forecasts. Thereafter, their forecast quality in terms

Table 1. Standard deviation of observation errors  $\mathbf{R}_k$  used in the experiments for all upper-air observations, based on the Desroziers statistics (Eq. (11)) from a period in June 2011. In the brackets, the first number indicates the original averaged observation errors  $\mathbf{R}_n$  used operationally in the nudging; the second number denotes the errors  $\mathbf{R}_e$  as estimated by Eq. (11) from the output of the standard LETKF experiment for the main period (in 2014) used in this study. The values are for horizontal wind components in  $\text{m s}^{-1}$ , temperature  $T$  in K and relative humidity  $RH$  in % as a function of vertical pressure level in hPa.

Level	Wind	Temperature	Rel. humidity
300	2.38 (3.85; 1.93)	0.56 (0.57, .60)	14 (19, 15)
400	2.08 (3.71; 1.71)	0.54 (0.50, .56)	13 (13, 17)
500	1.92 (3.40; 1.59)	0.55 (0.51, .62)	13 (13, 17)
700	1.89 (2.63; 1.55)	0.68 (0.73, .74)	12 (11, 15)
850	2.00 (2.45; 1.55)	0.85 (0.98, .69)	13 (10, 14)
1000	1.95 (2.46; 1.69)	1.07 (1.18, .90)	9 (10, 10)

of verification scores was overall similar to that of the forecasts started from the deterministic analysis. The area-averaged absolute values of surface pressure tendency and of vertical velocity at different pressure levels are also indicators of the balance. While they were found to be rather similar for the deterministic run and the ensemble members, the forecast on mean increased the pressure tendencies somewhat in the first hour. Remarkably, its vertical velocities are lower, by 40% initially and by 5–10% even after 12 h. Thus, the different balance of the forecast on mean remains for a long time into the forecast. It may be worth mentioning that the fairly severe vertical localization is not found to be the main source of imbalance. To counter the imbalance problems, some form of initialization could be applied, e.g. the physically based method of Hamrud *et al.* (2014) aimed at adjusting the horizontal wind divergence. However, the impact of such an initialization applied to the convective scale would need to be evaluated carefully and we have not carried out such tests yet. We note finally that using the ensemble mean analysis for a higher-resolution deterministic forecast would imply the need to interpolate the full analysis fields rather than the analysis increments.

### 3. KENDA suite components

In the EnKF, the uncertainty of the background is sampled by the ensemble perturbations, which are used to estimate the background covariance. It has emerged that, without any extra actions, the ensemble variances of the background and analysis are typically underestimated (e.g. Anderson, 2009). This effect is due partly to sampling errors caused by an ensemble size too small for high-dimensional applications and partly to inappropriate representation of model and observational errors. It results in too much confidence being attributed to the background and too little weight given to the observations in the analysis and it can easily lead to filter divergence.

The KENDA suite contains a range of components addressing these issues. They deal with the estimation of observation errors, consideration of model and background errors and adaptive localization. These tools are described briefly in the following subsections.

#### 3.1. Determination of observation error and its adaptation

Our estimation of the error covariance  $\mathbf{R}$  follows Desroziers *et al.* 2005a 2005b and Li *et al.* (2009). It is based on the relationship

$$\langle \mathbf{d}_{o-a} \mathbf{d}_{o-b}^T \rangle = \mathbf{R}, \quad (11)$$

where the brackets  $\langle \cdot \rangle$  denote statistical expectation,  $\mathbf{d}_{o-a}$  are the differences between observations and analysis in observation space and  $\mathbf{d}_{o-b}$  are the corresponding differences between observations and background. This tool is applied offline to estimate the observation-error variances after the analysis cycle is completed. In a second iteration, the estimated variances are

then used as pre-specified variance values in the analysis cycle and re-estimated thereafter from the new cycle. It is found that, after two iterations, the variance estimations do not change much any more and these values can then be used as pre-specified observation errors in all the subsequent experiments.

Table 1 indicates the observation error values  $\mathbf{R}_k$  for the upper-air data that were used for the LETKF experiments in this study. Surface-pressure errors are about 0.5 hPa. These values are based on the Desroziers statistics (Eq. (11)) from an LETKF experiment in another period (during June 2011) different from those used in the present study, in order to avoid an incestuous offline tuning that could not be applied in an operational set-up.

For comparison, Table 1 also shows the average values  $\mathbf{R}_n$  used operationally in the nudging scheme and the values  $\mathbf{R}_e$  as estimated by Eq. (11) from the output of the standard LETKF experiment for the main period (i.e. in 2014) used in this study. These values are an estimate of the true average observation errors occurring in this period and may give an indication of the optimal error values for the LETKF assimilation. Table 1 reveals that, for wind and for low-level temperature, the errors  $\mathbf{R}_k$  used in the experiments are much smaller than the operational values  $\mathbf{R}_n$  (by up to 40%) and the estimated true and possibly optimal errors  $\mathbf{R}_e$  are even smaller (by up to 20%). For humidity, the opposite tends to be true.

#### 3.2. Multiplicative covariance inflation

A *multiplicative covariance inflation* is implemented following Anderson and Anderson (1999). Inflating the analysis-error covariance  $\mathbf{P}^a$  by a factor  $\rho$  greater than one is achieved by multiplying the analysis perturbation matrix  $\mathbf{W}^a$  (see Eq. (8)) by  $\sqrt{\rho}$ .

The covariance inflation factor  $\rho$  can be pre-specified as a constant value, or it can be determined adaptively following Houtekamer *et al.* (2005), based on the ideas of Dee (1995) and Wang and Bishop (2003) according to

$$\rho = \frac{\mathbf{d}_{o-b}^T \mathbf{d}_{o-b} - \text{trace}(\mathbf{R})}{\text{trace}(\mathbf{H} \mathbf{P}^b \mathbf{H}^T)}, \quad (12)$$

where  $\text{trace}()$  denotes the trace of a matrix. The idea is to compare the observation- and background-error estimation in the LETKF with the actual errors as obtained by the observation minus background statistics. This is done at each analysis grid point separately, i.e. the adaptive inflation factor  $\rho$  varies in space and time. To mitigate sampling errors, an upper and lower limit are imposed for  $\rho$  at first; in the present study, these limits are set to 3.0 and 0.5 respectively. If only few observations are available to compute Eq. (12),  $\rho$  is relaxed towards 1. Thereafter, a temporal filter is applied by means of a weighted average of this bounded value of  $\rho$  (with weight 0.1) and the value used at the previous analysis step (with weight 0.9). The resulting inflation factors vary smoothly in time and a final lower limit of 1.0 is imposed to avoid covariance deflation.

These final values of  $\rho$  are found to have local maxima as high as 2, but, averaged over the horizontal domain, they are typically in the range of 1.1–1.4 at low levels and decrease with height. At 600 hPa, they are usually less than 1.1 on average.

#### 3.3. Relaxation to prior perturbations (RTPP)

Another kind of covariance inflation has been suggested by Zhang *et al.* (2004), where the posterior analysis ensemble perturbations  $\mathbf{X}^a$  are relaxed towards the first guess or background-ensemble perturbations  $\mathbf{X}^b$ . This relaxation to prior perturbations (RTPP) is carried out at each analysis grid point in an efficient manner by replacing the analysis perturbation matrix in ensemble space  $\mathbf{W}^a$  by

$$(1 - \alpha_p) \mathbf{W}^a + \alpha_p \mathbf{I}. \quad (13)$$

In most experiments of the current study, RTPP with  $\alpha_p = 0.75$  (following Whitaker and Hamill, 2012; Harnisch and Keil, 2015) and adaptive multiplicative covariance inflation as described above are applied together at each LETKF analysis step. This has been motivated by previous tests done at DWD with the global ICON–LETKF system. Introducing RTPP (with  $\alpha_p = 0.75$ ) had a clear benefit, but adding the adaptive multiplicative inflation on top of it further improved the spread-skill ratio. It increased the spread and had an at worst neutral impact on the errors. With KENDA, adaptive multiplicative inflation has been deployed in all experiments and no tests have been done with RTPP only. Adding RTPP on top of adaptive multiplicative inflation will be discussed in section 5.

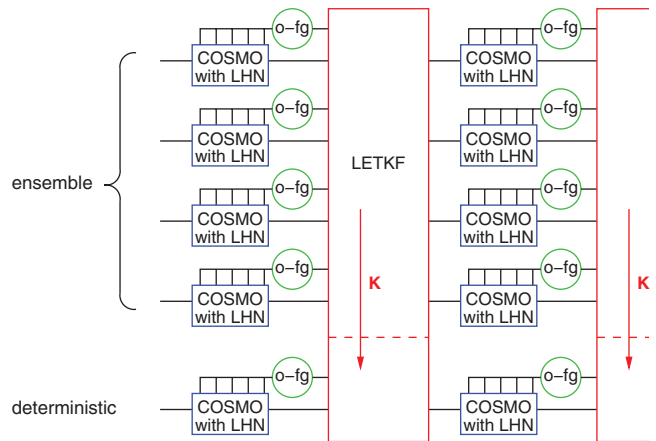
### 3.4. Perturbations of boundary conditions

While the two previous sections address the uncertainty in the interior of the model domain, it is also important to represent the uncertainty at the boundaries of the domain adequately. At the lateral and upper boundary, random perturbations result from the ensemble boundary conditions provided by the global ICON ensemble driven by an LETKF assimilation cycle (A. Cress, 2015; personal communication).

In order to account for uncertain lower boundary conditions, correlated random perturbations can also be applied to the constant-in-time sea-surface temperature (SST) and the prognostic soil moisture. Without such perturbations, the ensemble spread remains much too small in the boundary layer. Currently, these perturbations are created based on an elementary approach. The implemented pattern generator derives uncorrelated, normally distributed pseudo-random numbers on a coarse three-dimensional horizontal–temporal grid and accounts for the correlations by applying the square root of the desired covariance function. In practice, this is done by convolution with Gaspari–Cohn functions (Gaspari and Cohn, 1999) with appropriate length- and time-scales. For SST, the 2D perturbations computed in this way at an LETKF analysis time are added directly to the SST ensemble mean to obtain the perturbed SST. For prognostic soil moisture, a fraction  $\Delta t_a / \Delta t_{\text{corl}}$  of the 2D perturbation fields is added to the present soil moisture of each ensemble member independently from the other members. Here,  $\Delta t_a$  is the LETKF analysis update frequency and  $\Delta t_{\text{corl}}$  the pre-specified temporal correlation scale. Since the soil moisture perturbations are always added to the already existing perturbations, the total perturbations start with small values and tend to grow beyond the time-scale  $\Delta t_{\text{corl}}$  until saturation is nearly attained after about  $2 \Delta t_{\text{corl}}$ . In the vertical, the perturbations are decreased with increasing depth of the soil layers, also according to a Gaspari–Cohn function. Multiple random patterns with different scales may be superposed for multi-scale correlations. In the experiments of the current study, the scales for soil moisture were specified to be 100 and 10 km in the horizontal, 1 m in the vertical and 1 day in time. More details on the soil moisture perturbations, their size and impact is provided in section 5.1. For SST, the scales were 100 km in the horizontal and 1 day in time and the amplitude was 1 K.

### 3.5. Adaptive localization

Localization techniques are key ingredients of ensemble data assimilation systems based on the ensemble Kalman filter (Houtekamer and Mitchell, 1998, 2001; Anderson, 2001; Hamill *et al.*, 2001; Ott *et al.*, 2004; Janjić *et al.*, 2011). Localization means that computations are performed in a local region instead of a global space (Szunyogh *et al.*, 2005, 2008), to mitigate the limitations of an EnKF due to the limited ensemble size. Spurious correlations between spatially or temporally separated events are removed by localization of the background-error covariance matrix. Moreover, the rank of the covariance matrix is increased strongly by localization, so that the effective number of degrees



**Figure 3.** KENDA system set-up when LETKF is combined with LHN; ‘o–fg’ denotes observation minus first guess, ‘K’ the Kalman gain for the analysis mean and ‘COSMO with LHN’ means that the COSMO model is integrated with the LHN switched on.

of freedom of the data assimilation system is increased by a large factor. Adaptive localization methods have been studied by Anderson (2007, 2012), Bishop and Hodyss (2007, 2009a, 2009b), Zupanski *et al.* (2007), Migliorini (2013), Kirchgessner *et al.* (2014) and Periañez *et al.* (2014).

In the framework of the LETKF, localization is performed in observation space by carrying out the analysis independently at each *analysis grid* point, taking into account only observations in a neighbourhood of this grid point and applying localization weights to the data (see section 2.4).

While the vertical localization scale is fixed in KENDA (see section 2.4), a very simple scheme for adaptive horizontal localization has been developed. Here, the effective number of observations used for a local analysis is kept constant, i.e. the localization radius  $r_{\text{loc}}$  is adapted such that the sum over all localization-weighted observations is equal to a tuning constant  $N_{\text{loc}}^{\text{Oef}}$ . The rationale for this data-density-dependent localization scale is the following: Within the range of the characteristic localization radius, the number of degrees of freedom of the assimilation is roughly given by the ensemble size  $k$ . Adding more data to an already sufficiently large data set would then have little potential to improve the (mean) analysis itself but would make the LETKF (more) overconfident by further decreasing the estimated analysis error via Eq. (7). For perfect observations,  $N_{\text{loc}}^{\text{Oef}} \approx k$ , but for imperfect data,  $N_{\text{loc}}^{\text{Oef}}$  should be chosen somewhat larger.

To prevent the analysis increments being projected on too small scales or being contaminated too much by sampling noise of the background-error estimate, lower and upper limits,  $r_{\text{loc}}^{\text{min}}$  and  $r_{\text{loc}}^{\text{max}}$  respectively, are imposed on  $r_{\text{loc}}$ . For the experiments in this study, these limits are set to  $r_{\text{loc}}^{\text{min}} = 50$  km and  $r_{\text{loc}}^{\text{max}} = 100$  km and the target number of observations is  $N_{\text{loc}}^{\text{Oef}} = 100$  for an ensemble size of 40.

### 3.6. LETKF combined with LHN

In the KENDA system, the LETKF can be combined with LHN in a very simple way. During the forward integration from one LETKF analysis step to the next, the LHN is applied to each COSMO ensemble member and to the deterministic run (see Figure 3). This means that the COSMO integrations are not free forecast runs any more, since LHN increments are added to the model fields at each model time step. For observations valid between two analysis steps, such increments are hence applied before and after applying the observation operator for that observation.

Algorithmically, the LETKF and the LHN are separate. However, the effects of the two schemes influence each other within the data assimilation cycle via the model state. The LHN modifies the model trajectories from which the innovations for the LETKF are computed during the forward integration. It thereby



Table 2. Number of actively used conventional observations for analysis times with relatively few data (0200 UTC) and many data (1100 UTC, 22 July 2012). 'Wind' denotes the horizontal wind vector. The TEMP data at 1100 UTC are from 13 stations.

Obs. type and variable	0200 UTC	1100 UTC
SYNOP pressure	567	563
SYNOP 10 m wind	243	243
AMDAR wind	0	311
AMDAR temperature	0	305
TEMP wind	0	250
TEMP temperature	0	265
TEMP humidity	0	248
Wind profiler	471	419

also affects the four-dimensional background-error covariances evaluated in observation space in the LETKF. By modifying the first-guess ensemble perturbations, it has a further impact on the LETKF analysis increments as well as on the effect of schemes like RTPP. Finally (and quite trivially), it modifies the first-guess fields to which the LETKF analysis increments are added. Conversely, the LETKF increments change the model state and hence the precipitation rates produced by the model after the LETKF step and this (also trivially) has an effect on the LHN increments.

#### 4. Basic experimental set-up

The first goal of the KENDA development is to replace the deterministic operational nudging-based data assimilation with the KENDA deterministic analysis. Therefore, our main aim here is to compare the two schemes by means of real-data cycling data assimilation and deterministic forecast experiments. Care is taken to make the comparison as fair as possible. In particular, we use the same lateral boundary conditions and the same initial and lower boundary condition at the beginning of the experiment period and provide the schemes with the same set of observations. On the other hand, care is also taken to make use of observations and choice of nudging parameters in the control experiment as similar to the operational COSMO-DE set-up as possible.

The control experiment has the following differences compared with the operational set-up:

- (very!) different lateral boundary conditions (see section 4.3);
- no use of 2 m humidity data (as discussed in section 4.1);
- no data cut-off time in the assimilation cycle;
- no nudging of data valid after analysis time in the forecast run (the operational forecast run can nudge towards available data valid up to about 30 min after the analysis time, given a cut-off time of 40 min); and
- a different model version compared with the one used during the experimental periods, no six-hourly analysis of snow depth.

The large positive impact of the LHN on the operational precipitation forecasts for short lead times (Stephan *et al.*, 2008; Stephan, 2012; private communication) suggests that KENDA cannot replace nudging operationally in the future, unless it uses information related to the radar reflectivity data. As experimentation (for the longer-term goal) using 3D radar reflectivity (and radial velocity) in the LETKF directly is still in a relatively early phase, we test combination of the LETKF with LHN here, with a view to its operational application.

This section outlines the experimentation environment and describes the observation sets and basic set-up of the experiments.

##### 4.1. Observation data sets used

The observation data set used for the experiments was selected to allow for a fair comparison between LETKF and nudging; at

the same time, it is similar to the data set used operationally for COSMO-DE. For the upper air (see Table 2), this comprises AMDAR aircraft wind and temperature data without thinning in the vertical, approximately half-hourly wind profiler data from six stations (four German, one Swiss, one Dutch; other stations are blacklisted) and radiosonde (TEMP) data from up to 21 stations (standard and significant level data). As noted in section 2.1, the model fields are drawn towards the boundary fields of the driving model within an upper boundary Rayleigh relaxation zone reaching down to about 235 hPa. To limit the impact of background-error covariances distorted by this relaxation, data above 300 hPa were generally discarded in the LETKF experiments. In the nudging experiments shown here, these data were used as in the operational set-up, but tests have shown their impact to be negligible in the test periods of this study, due to Rayleigh relaxation. Thus, this difference in the datasets used in the nudging and LETKF experiments is irrelevant for the results of this study.

From the surface stations (and ships), hourly surface pressure is assimilated. 10 m wind is used from a smaller subset of stations over low terrain (mainly the plains in the northern half of the domain). 2 m temperature is not used. 2 m humidity is used operationally in the nudging, but is given a very limited vertical influence in a way that cannot be mimicked in an LETKF. Therefore, 2 m humidity is not used in the experiments, to allow for a fair comparison between LETKF and nudging. However, additional control experiments have revealed that nudging of 2 m humidity with the operational settings has negligible forecast impact in the test periods.

Table 2 provides typical numbers of observations used in an LETKF analysis step (for an hourly update frequency) for a time with relatively few data and another time with many data.

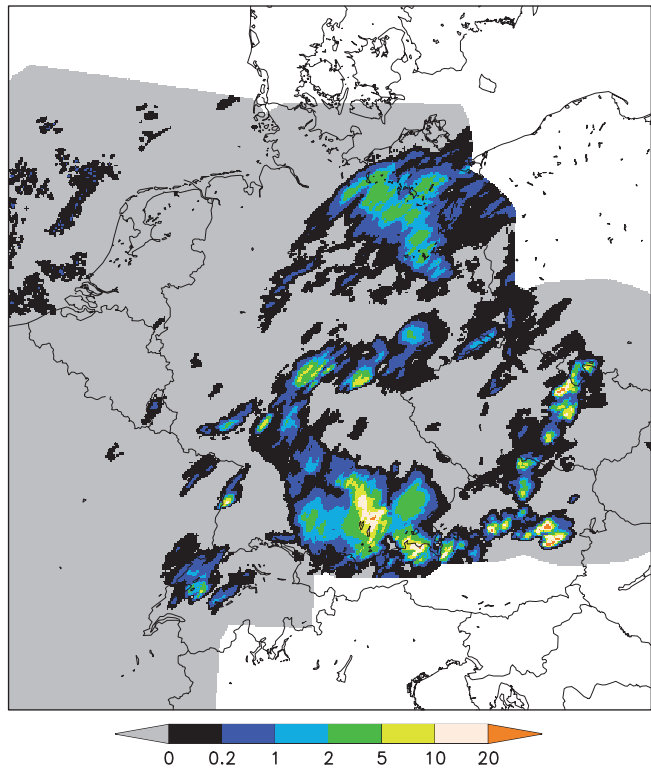
The pre-specified observation errors  $R_k$  used for the LETKF experiments are based on the Desroziers statistics (Eq. (11)), as described in section 3.1. Table 1 specifies these values and the operational values  $R_n$  used in the nudging experiments. For wind and low-level temperature, the operational values  $R_n$  are much larger. We therefore emphasize that, in the nudging scheme, observation-error values have little influence except in the quality control. (Different weighting for different observation types is imposed via different values of the nudging coefficient  $G_\gamma$  in Eq. (1)). Thus, replacing  $R_n$  by  $R_k$  in the nudging experiments would have had little impact on the results and the comparison between LETKF with  $R_k$  and nudging with  $R_n$  remains fair.

In addition to these standard conventional data, radar-derived precipitation rates are assimilated operationally by LHN (see section 2.3). Germany employs 17 radar stations, which provide reflectivity data at high spatial resolution every 5 min. Reflectivity from the lowest elevation angle is converted into a near-surface precipitation rate using a multi-state  $Z-R$  relation. The data are combined with data from 16 radars from neighbouring countries to obtain the final composite, which covers a large part of the COSMO-DE domain (Figure 4).

A careful operational quality control is carried out for frequently occurring errors due to non-rain echoes from sources like ground clutter, anomalous propagation, positive/negative strokes and rings (Hengstebeck *et al.*, 2010). A blacklist of data points in the composite is also deployed. Finally, a bright-band detection uses radar beam height maps and the freezing level temperature zone as simulated by the COSMO-DE run that applies LHN. Within LHN, only data with the best available quality are processed.

##### 4.2. Basic cycling (BACY) testing environment

A data assimilation and forecast basic cycling environment (BACY) has been developed at DWD for rapid experimentation, particularly in view of ensemble data assimilation (EnDA). It also allows for comparison with the current nudging-based



**Figure 4.** Snapshot (for 20 June 2012, 1600 UTC) of the radar composite from the German radar station network and neighbouring countries such as the Netherlands, Belgium, France, Switzerland, Poland and the Czech Republic; the shaded areas show the coverage of the precipitation data used for assimilation and verification. The values in the colour bar are in mm h<sup>−1</sup>.

operational assimilation scheme and has been used for this study. EPS capabilities are also included.

Compared with DWD’s standard experimentation system, which is also deployed for the operational suite, BACY has several advantages. Firstly, BACY experiments, and in particular series of experiments for the same period, run at a much faster speed, since they cycle from hard drive instead of using a database and archiving system based on tapes. This speed-up is crucial for EnDA, particularly with rapid update cycling, due to its huge need for data access. Secondly, BACY is fully portable to standard LINUX-based computing environments. This greatly improves the technical foundation for collaboration, for example with universities. Thirdly, it is rather flexible and easy to modify, so that changes and extensions of the pure LETKF system can be tested fairly readily.

Finally, we mention that the computational cost of providing a deterministic analysis using an LETKF with 40 ensemble members is about 20–40 times higher compared with that of nudging. For operational purposes, however, the cost of KENDA will still be moderate (i.e. of the order of 15%) compared with the cost of the complete COSMO-DE system, as this includes three-hourly ensemble forecasts of 27 h or longer in the COSMO-DE-EPS system.

4.3. Study set-up and description of test periods

The main difference between the operational set-up and the experimental setup adopted for this study is the different lateral boundary conditions (LBC). Operationally, they are derived from COSMO-EU, a 7 km resolution regional COSMO configuration, while, for the current study, ensemble LBC are required for the LETKF. These were obtained from an LETKF experiment with the new global model ICON (Zängl et al., 2015), which has been operational at DWD since January 2015. The ICON–LETKF experiment has a horizontal grid spacing of 80 km for the ensemble and 40 km for the deterministic run. Obviously, this low resolution does not provide LBC with optimal quality, but

Table 3. General settings for the LETKF used in the experiments.

Variable/feature	Value
Ensemble size $k$	40
Deterministic run	1
Horiz. resolution ens. + det. run	2.8 km
Forecast frequency / length	6 h 24 h
Analysis update frequency	1 h
Vert. localiz. length-scale (ln $p$ )	0.075–0.5
Horizontal localization	Adaptive
→ Target weighted no. obs. $N_{loc}^{oef}$	100
→ Min. local. length-scale $r_{loc}^{min}$	50 km
→ Max. local. length-scale $r_{loc}^{max}$	100 km
Multiplicative covariance inflation	Adaptive
RTTP relaxation weight $\alpha_p$	0.75

the relevant point is that it allows a fair comparison between KENDA–LETKF and nudging, as it is applied to both.

Similar considerations hold for downscaling the initial state from the ICON–LETKF experiment at the beginning of the whole experiment period. For the atmospheric fields, this provides reasonable large-scale balance and ensemble perturbations to initialize the KENDA–LETKF. The initial atmospheric ensemble mean has a rather short memory in the assimilation cycle. The downscaled soil moisture and temperature fields, however, have a large long-term impact throughout the experiment period. Therefore, all experiments except those in section 5.1 were performed by replacing the ICON–LETKF soil moisture and temperature at the initial time of the experiments with that of the operational COSMO-DE, in order to have a set-up more similar to the operational one. The impact of this replacement is addressed in section 5.1.

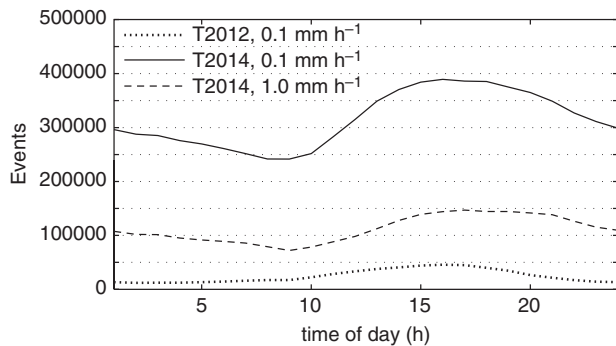
Table 3 provides an overview of the general settings of the LETKF cycle in the experiments. Besides the direct comparison between LETKF and nudging, further experiments have been performed to evaluate certain aspects of the system. The topics are as follows.

- *Study 1.* We assess the sensitivity of the LETKF to *soil moisture perturbations* and *soil moisture conditions* with regard to low-level ensemble spread and errors.
- *Study 2.* We explore how the impact of RTTP is influenced by the spread of the lateral boundary conditions on a model domain as small as that of COSMO-DE.
- *Study 3.* We show the impact on the deterministic forecast from adding *latent heat nudging* (LHN) to the LETKF assimilation cycle, as described in section 3.6.
- *Study 4.* We compare the *deterministic* forecast based on *nudging* data assimilation for *conventional* data with the deterministic forecast based on the *LETKF* deterministic analysis.

Due to the computational costs of comparing many different settings, studies 1 and 2 were carried out over the relatively short period from 19–25 July 2012. The evaluation of the experiments for these two studies focuses on upper-air verification scores for 20–25 July. The period is characterized by rather dry anticyclonic conditions and convective and some larger-scale rain is mostly limited to the southern half of the domain from noon to evening during three days. It is named T2012 hereafter.

To extend the statistical basis, in particular with respect to convective precipitation, for the assessment of the impact of LHN and for the main comparison between LETKF and nudging, studies 3 and 4 were computed additionally for a second, longer period from 17 May 2014–15 June 2014, 0000 UTC. The verification starts on 18 May and extends over 28 days. This period, named T2014, is characterized by mostly rather weak advection and contains a variety of weather situations: a slow passage of a weak depression through the model domain, episodes with isolated strong convection, periods with more widespread convective and frontal precipitation and dry episodes.





**Figure 5.** Number of observed events (grid points) with hourly precipitation exceeding a given threshold. Dotted line  $0.1 \text{ mm h}^{-1}$  for period T2012, solid line  $0.1 \text{ mm h}^{-1}$  for period T2014 and dashed line  $1.0 \text{ mm h}^{-1}$  for period T2014.

Due to the shortness of the six-day period T2012 and the relatively few precipitation events (see Figure 5) relative to the 28 day period T2014, we will present here only the results from latter period for studies 3 and 4. Even though the results in the two periods differ in some details, the general results and conclusions from them are similar.

In order to obtain measures for the statistical confidence of the main results in studies 3 and 4, bootstrapping was carried out for the period T2014. For the surface and upper-air verification, a dataset was collected for each day containing the 0000, 0600, 1200 and 1800 UTC forecast runs and corresponding observations. For the precipitation verification, which is done separately for 0000, 0600, 1200 and 1800 UTC runs, each dataset contained the data from a single forecast run. Thus, 28 datasets with (assumed) uncorrelated errors between them were obtained for the 28 days for each verification. Each of the 10 000 bootstrap realizations was then based on 28 random draws with replacement from these daily datasets. Percentage improvement (PI) was calculated for each realization, so that a positive value indicates an improvement of the experiment over the reference.

## 5. Results of numerical tests in an operational environment

### 5.1. Soil moisture perturbations

To study the impact of soil moisture perturbations, four experiments are compared for period T2012.

- (1) SMNO: LETKF cycle without any soil moisture perturbations, i.e. at each analysis step all the ensemble members are initialized with the same soil moisture content.
- (2) SMCYCL: soil moisture is cycled in the ensemble members independently of each other, i.e. the analyzed soil moisture is set equal to the first guess at each LETKF step. As a result, the soil moisture evolves differently in each member. We call this *implicit soil moisture perturbations*.
- (3) SMPERT: in addition to the implicit perturbations generated by cycling, explicit soil moisture perturbations are added as described in Section 3.4. The temporal correlation scale  $\Delta t_{\text{corl}}$  is 1 day. Perturbations with two horizontal length-scales, 100 km and 10 km, are superposed. For each length-scale, the standard deviation of the random perturbations added incrementally in the top soil layer is 0.05 SMI (soil moisture index) per day, where SMI varies between zero at the air dryness point and one at the pore volume. The vertical length-scale is 1 m, so that, in soil layer 5 with depth 27–81 cm, the amplitude in terms of SMI is 77% of that in the top soil layer with 0–1 cm depth. As a result, the perturbations in terms of absolute amount of soil water are largest in layer 5, because it is much thicker than the layers above.

It is noted that using different perturbation time-scales for the different soil layers would decouple the perturbations in the vertical and this would compromise their impact. On

the other hand, a time-scale of 1 day for the perturbations in all layers is much smaller than the typical time-scale of soil moisture variations in the lower soil layers. However, we do not consider this a problem, provided that only atmospheric data are assimilated. In this case, the (lower limit for the) time-scale of the soil moisture perturbations should be related to the memory of the atmospheric state with respect to the soil conditions. For example, the near-surface temperature at noon on a sunny day in summer will depend on the surface flux of latent heat and hence the soil moisture throughout the morning of the same day. Conversely, it will depend rather little on the soil conditions of the previous day. Thus, the atmospheric state of a particular ensemble member depends mainly on its soil conditions during the same day (or same night). Therefore, the soil moisture perturbation should not vary much during that day, but it should not matter much if it is different on the previous day.

- (4) SMPERT2: as SMPERT, but with doubled amplitude for the explicit perturbations of soil moisture (and SST).

Figure 6 indicates that the soil moisture perturbations in SMPERT are typically about an order of magnitude larger than in SMCYCL during the experiment period. Note that in a quasi-operational setting with continued data assimilation cycling over whole seasons, the implicit soil moisture perturbations would grow to larger amplitudes than within the short six-day period of experiment SMCYCL.

As shown in the upper-air verification against observations in Figure 7, the weak implicit soil moisture perturbations in SMCYCL reduce the temperature and humidity errors slightly below 850 hPa, but have only a very small impact on the spread. Below 700 hPa, the spread decreases with decreasing height for all variables, while the root-mean-square error (RMSE) increases for temperature and remains nearly constant in height for wind. Thus, the ratio between spread and RMSE is lower close to the ground.

A major purpose of an ensemble of forecasts  $f$  is to estimate the forecast uncertainty by means of the ensemble spread  $\sigma_f - \bar{f}$  (root-mean-square (RMS) ensemble deviation from the mean  $\bar{f}$ ). The dispersion of ensembles can be judged by comparing the spread with the *forecast skill*. Ideally, they should be equally large. Often, the RMS difference  $\sigma_{f-o}$  between the forecast and observations  $o$  (or a verifying analysis) is used to measure the skill and is commonly denoted as the RMSE of the forecast. However, as pointed out e.g. by Szunyogh *et al.* (2008), the true *forecast skill*  $T$  is given by

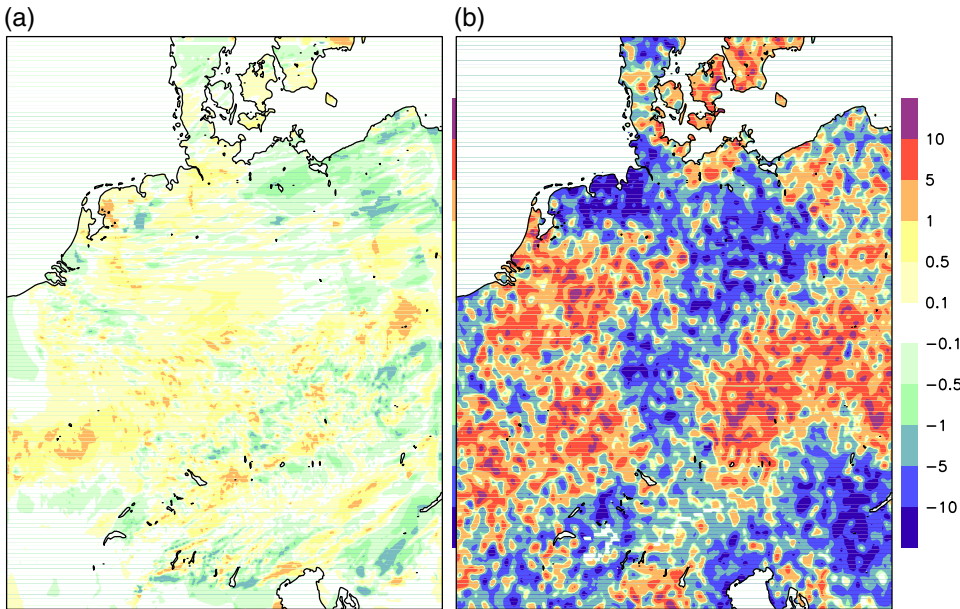
$$T^2 = \sigma_{f-o}^2 - \sigma_o^2 + 2C, \quad (14)$$

where  $\sigma_o^2$  is the observation error variance and  $C$  the error covariance between the forecast and the observations. Assuming zero covariance and using the deterministic forecast  $f_d$  to measure the skill, the spread-skill ratio  $r_s$  is given by

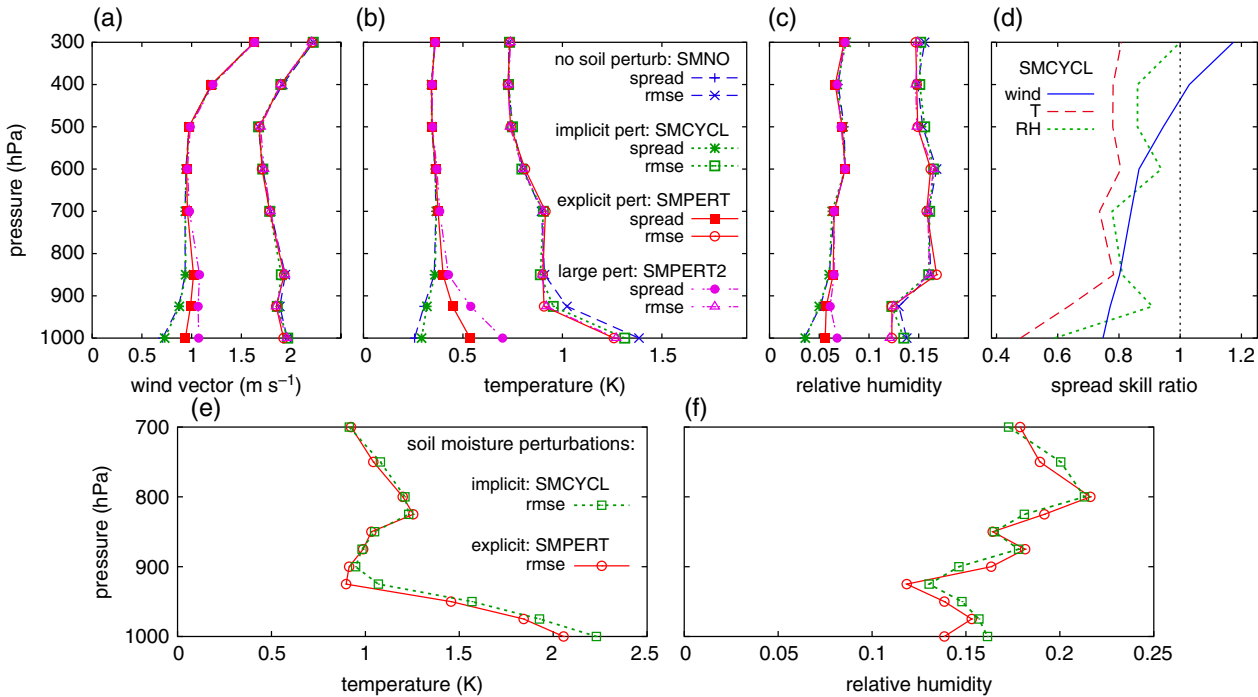
$$r_s = \frac{\sigma_f - \bar{f}}{\sqrt{\sigma_{f_d-o}^2 - \sigma_o^2}}. \quad (15)$$

Neglecting  $\sigma_o^2$ , as is often done, is not a good approximation for short-range forecasts, where  $\sigma_{f-o}$  is not much larger than  $\sigma_o$ . However,  $\sigma_o$  is not known exactly. Here, we use the estimates  $R_e$  obtained from Eq. (11) (see Table 1).

As shown in the upper right panel of Figure 7, the spread-skill ratio of the first guess of experiment SMCYCL is mostly smaller than one and generally tends to decrease with decreasing height. The smallest values are found at 1000 hPa and vary between 0.75 for wind and less than 0.5 for temperature. Obviously, the first-guess ensemble is more underdispersive towards the surface than further aloft. This indicates that the background uncertainty



**Figure 6.** Difference of soil moisture (WSO in  $\text{kg m}^{-2}$ ) in layer 3 (i.e. 3–9 cm depth) between two ensemble members (members 1 and 2) after 5 days of cycling (on 24 July 2012, 0000 UTC): (a) for experiment SMCYCL with implicit soil moisture perturbations, (b) for SMPERT with explicit soil moisture perturbations.



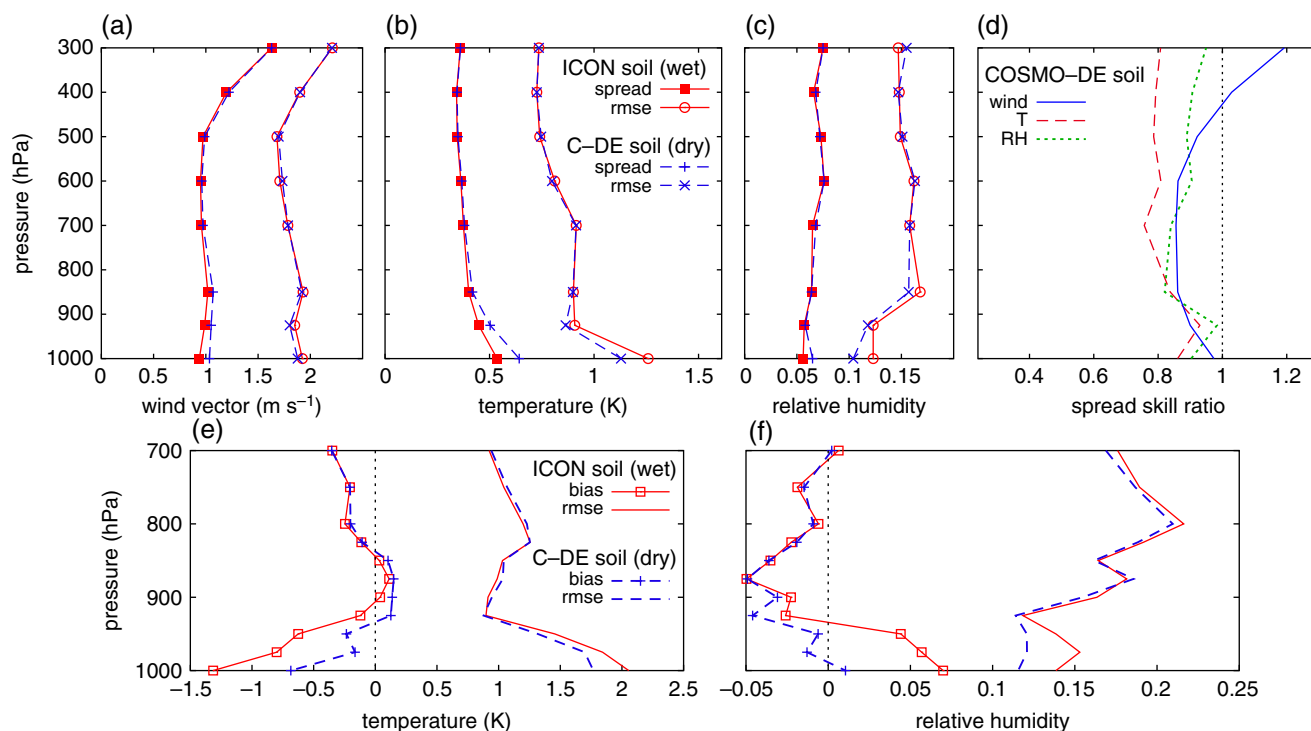
**Figure 7.** (a)–(c) Vertical profiles of first-guess ensemble spread and root-mean-square error (RMSE) of the deterministic first guess (i.e. 1 h forecast) against upper-air observations (from radiosondes, aircraft and wind profilers) of horizontal wind vector, temperature and relative humidity. Each panel shows experiments SMNO without soil moisture perturbations (dashed lines), SMCYCL with implicit perturbations (solid), SMPERT with explicit perturbations (dotted) and SMPERT2 with doubled explicit perturbations (thin dash–dotted lines) for period T2012 from 20–25 July 2012. (d) Spread-skill ratio as given by Eq. (14) for horizontal wind vector (solid), temperature (dotted) and relative humidity (dashed) for experiment SMCYCL. (e) and (f) Vertical profiles (up to 700 hPa) of 6 h forecast RMSE against radiosonde observations of (e) temperature and (f) relative humidity for SMCYCL (dotted) and SMPERT (solid line).

is underrepresented at the surface and the planetary boundary layer.

This is greatly improved by adding the explicit soil moisture perturbations in SMPERT. The spread is increased significantly up to 850 hPa for all variables and the errors are decreased slightly. A small error reduction also prevails during the forecast for low-tropospheric temperature and humidity (see Figure 7 for 6 h forecasts), as well as for 2 m temperature, 2 m humidity and cloud fraction (not shown). Doubling the amplitude of the perturbations in SMPERT2 does not reduce the errors further but increases the spread, so that the spread-skill ratio becomes larger near the surface than aloft. This indicates that background errors at the surface are over-represented, i.e. that the doubled explicit perturbations are too strong. Therefore, all further experiments

in this study are performed using our standard values for the explicit soil perturbations, as in experiment SMPERT.

For the above series of experiments including SMPERT, the initial state of the soil moisture and temperature at the beginning of the experiment period (19 July 2012, 0000 UTC for the period T2012) was obtained by interpolation from the ICON–LETKF experiment, which also provides the lateral boundary conditions. That experiment was started one week earlier, from the soil conditions present at that time in the operational assimilation cycle for GME. Hence, since soil moisture and temperature have a long memory, particularly in the thick lower layers, the average soil moisture content in SMPERT roughly reflects that of GME. The operational GME system deployed a variational soil moisture analysis (SMA), which by design aims at improving daytime



**Figure 8.** (a)–(c) Profiles of first-guess ensemble spread and RMSE of the deterministic first guess, as in the corresponding panels of Figure 7, but for experiments SMPERT with ICON–LETKF soil conditions (solid lines) and KENDA with COSMO-DE soil conditions (dashed lines) for period T2012. (d) Spread-skill ratio based on Eq. (14) as in Figure 7(d), but for experiment KENDA. (e) and (f) Vertical profiles (up to 700 hPa) of 6 h forecast bias and RMSE against radiosonde observations of (e) temperature and (f) relative humidity for SMPERT (solid) and KENDA (dashed lines).

2 m temperature forecasts. No data related to precipitation were assimilated. The operational COSMO-DE suite, on the other hand, uses latent heat nudging to assimilate precipitation data but does not apply SMA.

We thus assess the influence of the different soil conditions by comparing two experiments.

1. SMPERT: see above; the initial soil moisture and temperature values reflect the ICON–LETKF and ultimately the operational GME soil conditions (we name this ‘ICON–LETKF soil conditions’).
2. KENDA: identical to SMPERT, except that at the beginning of the six-day experiment period (19 July 2012, 0000 UTC) the soil temperature and soil moisture index (SMI) values were replaced by the operational COSMO-DE values valid on that date (‘COSMO-DE soil conditions’). This experiment uses our standard KENDA–LETKF settings as described in sections 2.4, 3 and 4 (without LHN).

As Figure 8 shows, using the COSMO-DE soil conditions decreases the RMSE and particularly the biases of temperature and humidity in the lowest 100 hPa of the atmosphere quite strongly. Moreover, it is found to increase the ensemble spread of temperature, humidity and wind near the surface. The reason is as follows. While the two soil conditions are found to be similar in the top layers, *COSMO-DE soil conditions* are much drier in layer 5, with mostly 0.1–0.5 SMI, compared with *ICON–LETKF soil conditions* with 0.4–0.7 SMI. Soil layer 5 extends from 27–81 cm and it is the layer with the largest impact on plant evapotranspiration, due to its thickness and penetration by the roots. Evapotranspiration is clearly more sensitive to soil moisture variations in dry soil than in wet soil. Therefore, with dry soil, explicit soil moisture perturbations of the same size result in more spread in the temperature of the top soil layer and in the near-surface atmosphere.

As a consequence of the enhanced spread and reduced errors at low levels in experiment KENDA, the spread-skill ratio (see Figure 8(d)) of the first guess is no longer lower close to the ground. The vertical profiles are relatively uniform, with values in the range between 0.8 and 1.0 for all variables (except for the wind at 300 hPa).

It can be concluded that the covariance inflation methods applied in KENDA–LETKF work well in the sense that they lead to a reasonable ensemble spread in the data assimilation cycle. This does not imply that the spatio-temporal correlations of the ensemble perturbations are optimal.

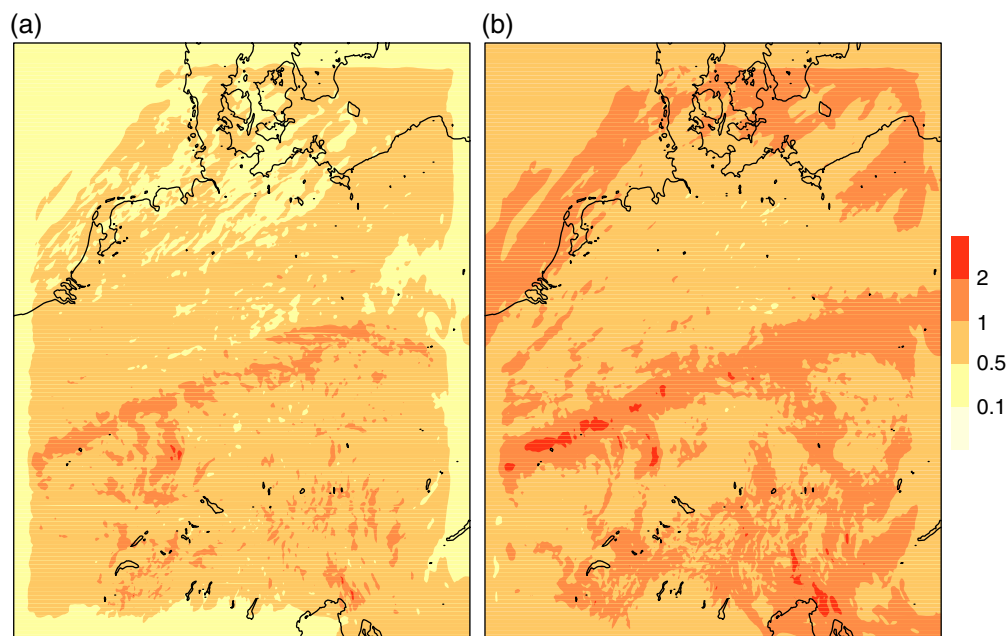
## 5.2. Influence of the spread in the lateral boundary conditions on RTPP

The impact of RTPP (see section 3.3) was found to be smaller in some of our preliminary tests with KENDA compared with the results of Harnisch and Keil (2015). The experimental set-ups were dissimilar in several respects, one of them being the different lateral boundary conditions (LBC). This motivated an investigation into the influence of the LBC spread on RTPP by means of the following four experiments, which were carried out for the 6 day period T2012. For practical reasons, latent heat nudging (LHN) was added to the LETKF cycle in each of them; this aspect is addressed at the end of the present section and a detailed study on the influence of LHN in the KENDA system is given in the subsequent section.

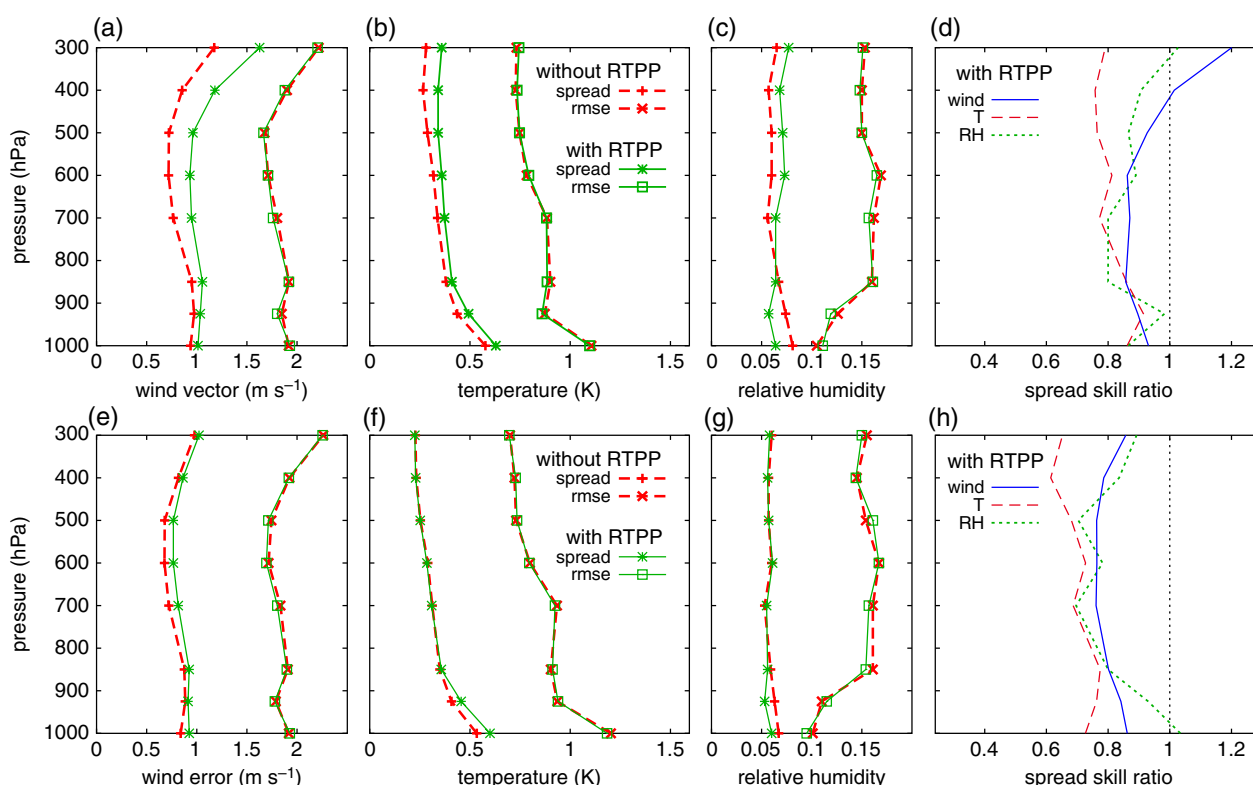
- (1) KENDA–LHN: standard LETKF set-up as in experiment KENDA (see section 5.1), but with LHN. This set-up applies RTPP (in addition to adaptive multiplicative covariance inflation, see section 3) and uses ensemble LBC derived from an ICON–LETKF experiment with ‘standard’ settings (which also include the use of RTPP); these LBC are subsequently called *standard LBC*.
- (2) KLNOR: as KENDA–LHN, but without RTPP.
- (3) BCRTTP: as KENDA–LHN (i.e. with RTPP), but with LBC from an ICON–LETKF experiment without RTPP that had a lower ensemble spread; these LBC are denoted as *low-spread LBC*.
- (4) BCNOR: as BCRTTP, but without RTPP.

Figure 9 illustrates that, in experiment BCRTTP, the ensemble spread is much smaller in the boundary zone than in the inner model domain. In KENDA–LHN, this difference is much less pronounced.





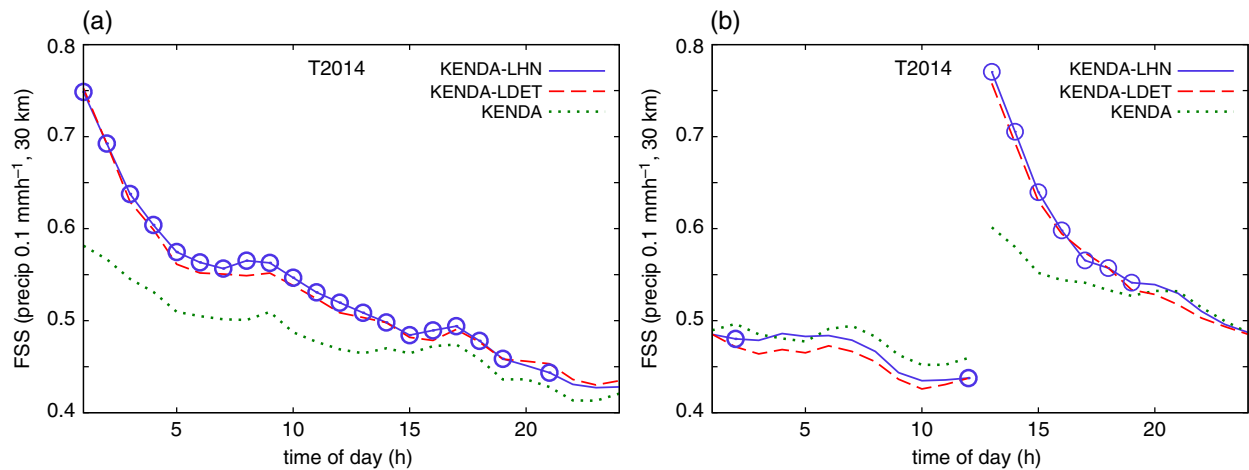
**Figure 9.** Analysis ensemble spread for the zonal wind component at model level 25 ( $\approx 500$  hPa) for 24 July 2012, 0000 UTC: (a) for experiment BCRTTP with lateral boundary conditions (LBC) from an ICON-LETKF experiment with low spread; (b) for experiment KENDA-LHN with LBC from the standard ICON-LETKF experiment. The values on the colour bar are in  $\text{m s}^{-1}$ .



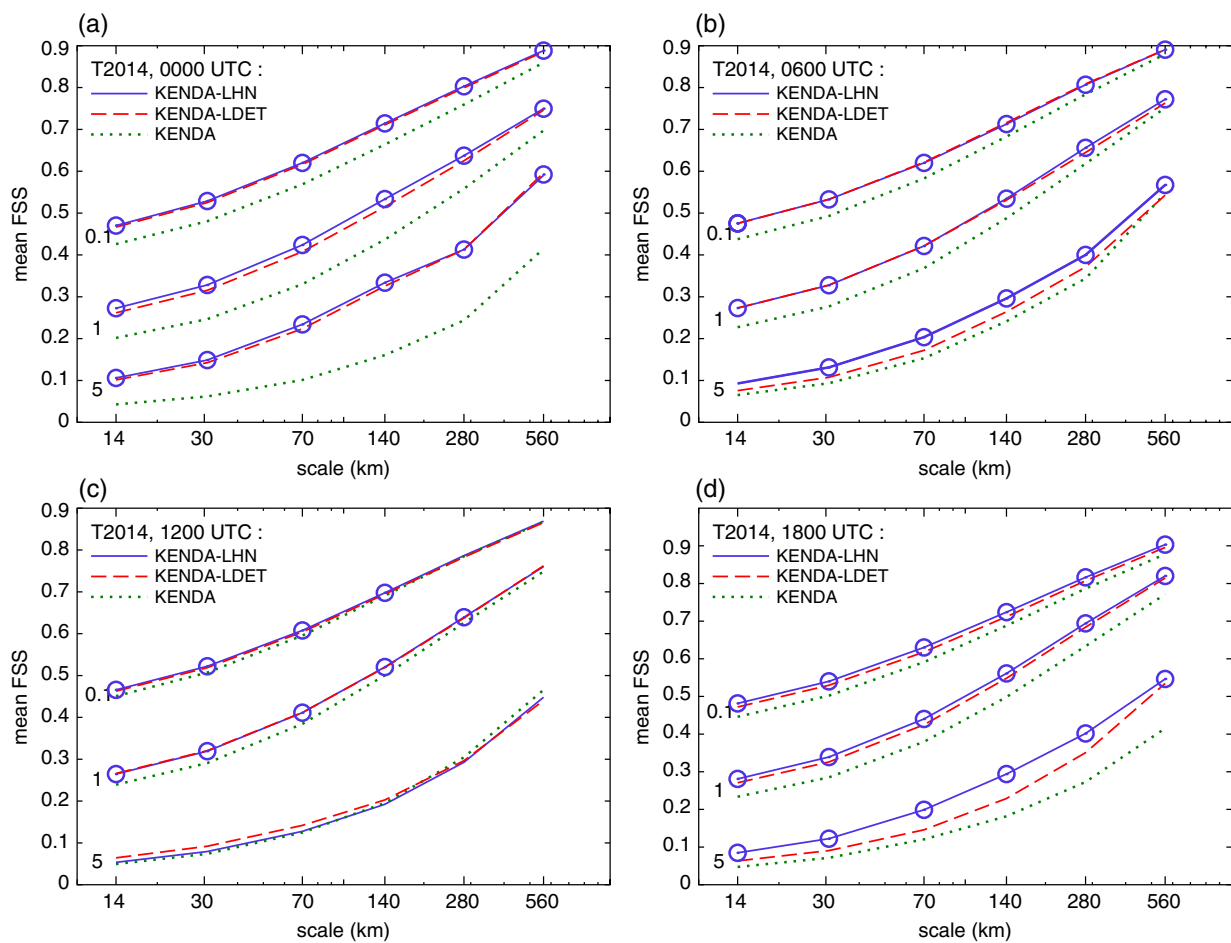
**Figure 10.** (a)–(c) Profiles of first-guess ensemble spread and RMSE of the deterministic first guess as in the corresponding panels of Figure 7, but for experiments KENDA-LHN with RTPP (solid lines) and KLNOR without RTPP (dashed lines), both driven by the *standard* LBC. (d) Spread-skill ratio based on Eq. (14) as in Figure 7(d), but for experiment KENDA-LHN. (e)–(h) As in the upper row (a)–(d), but for the corresponding experiments BCRTTP with RTPP (solid lines) and BCNOR without RTPP (dashed lines), both driven by the *low-spread* LBC. All results relate to the 6 day period T2012.

The influence of the LBC spread on RTPP is evident in the spread and errors of the 1 h forecasts, as shown in Figure 10. With the *standard* LBC, RTPP reduces the errors marginally, increases the spread substantially and improves the mean spread-skill ratio to values between 0.8 and 0.9 at almost all levels for wind, temperature and humidity. In contrast, if RTPP is applied with the *low-spread* LBC, it increases the spread only slightly for wind and is not able to increase the spread at all for temperature and humidity at most levels. The effect on the errors is small. As a result, the spread-skill ratio remains between 0.6 and 0.8 for temperature and around 0.8 for wind and humidity.

The reason why RTPP does not work well with strongly underdispersive LBC (on a small model domain such as that of COSMO-DE) is easy to understand. If too little spread enters the model domain from the boundaries during the forecast between two analysis steps and certain measures (e.g. adaptive multiplicative covariance inflation based on innovation statistics) are applied to keep a certain amount of spread in the analysis, the first-guess ensemble spread can get consistently smaller than the analysis ensemble spread. In this case, RTPP tends to be a covariance deflation towards the small spread of the LBC, at least near the inflow boundary.



**Figure 11.** Verification of one-hourly precipitation against radar-derived precipitation for the 28 day period T2014 from 18 May 2014–14 June 2014: fraction skill score (FSS) with a scale of 11 grid points ( $\approx 30$  km) for a threshold of  $0.1 \text{ mm h}^{-1}$  as a function of time of day, for (a) 0000 UTC forecast runs and (b) 1200 UTC forecast runs, for experiments KENDA without LHN (dotted line), KENDA-LDET with LHN only in the deterministic run (dashed) and KENDA-LHN with LHN additionally in all the LETKF ensemble members (solid). The circles on the solid line indicate forecast ranges where the confidence for an FSS improvement of KENDA-LHN over KENDA exceeds 95% according to bootstrapping.



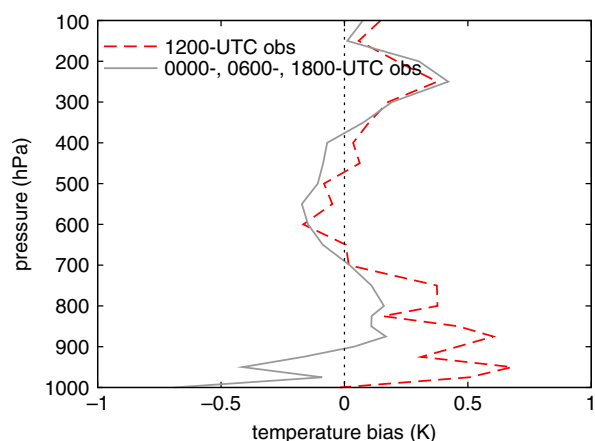
**Figure 12.** Verification against radar-derived precipitation for the 28 day period T2014: fraction skill scores (FSS) for one-hourly precipitation as a function of scale, where each of the values shown is an average over the 24 FSS values valid for the forecast lead times from 1–24 h. The labels for the abscissa denote the (approximate) scales for which the FSS is computed. In each panel, the upper, middle and lower triplet of lines are for thresholds of 0.1, 1.0 and  $5.0 \text{ mm h}^{-1}$ , respectively. Each triplet of lines shows the scores for the experiments KENDA without LHN (dotted line), KENDA-LDET with LHN only in the deterministic run (dashed) and KENDA-LHN with LHN additionally in all the LETKF ensemble (solid). The circles on the solid lines indicate scales where the confidence for an FSS improvement of KENDA-LHN over KENDA exceeds 95%. The four panels are for (a) 0000, (b) 0600, (c) 1200 and (d) 1800 UTC forecast runs, respectively.

Even though we did not carry out an analogous set of experiments without LHN, we have no indication that the above findings are influenced by LHN. The comparison between the experiments KENDA (see Figure 8(a)–(d)) and KENDA-LHN (Figure 10(e)–(h)) suggests that (at least for the period T2012) LHN has a limited impact on the ensemble spread and the errors (on average), so that the influence on the spread-skill ratio is much smaller than that of RTPP or the LBC.

### 5.3. Latent heat nudging and LETKF

The motivation for combining the LETKF with latent heat nudging (LHN) has been outlined at the beginning of section 4. We compare three experiments for the 28 day period T2014 (from 18 May 2014–14 June 2014).

- (1) KENDA: LETKF with the standard set-up and without LHN.



**Figure 13.** Vertical profiles of the temperature bias of the 6 h forecasts against radiosonde observations from 1200 UTC (dashed line) and from 0000, 0600 and 1800 UTC (solid line), respectively. Results are for experiment KENDA–LHN during period T2014.

- (2) KENDA–LDET: LETKF combined with LHN in the deterministic run only. This leaves the LETKF assimilation cycle unaffected by LHN. However, the system's states and hence the background-error covariances of the ensemble and the deterministic run might be more likely to diverge. As a result, the Kalman gain from the LETKF may potentially become inappropriate to compute the analysis increments for the deterministic run from its innovations.
- (3) KENDA–LHN: LETKF combined with LHN both in the deterministic run and in all the ensemble members. This avoids the enhanced potential of divergence between the two, but the LHN affects the background-error covariances used in the LETKF directly in an unknown way. For instance, the covariances and analysis increments could possibly be projected on to smaller scales, as LHN operates on the grid-point scale.

The focus of the evaluation is on the forecast of one-hourly precipitation, which is verified against radar-derived precipitation using e.g. the fraction skill score (FSS: Roberts and Lean, 2008; Mittermaier and Roberts, 2010). This score introduces a fuzziness of pre-specified scale  $s_f$  into the verification, in order to account for the double penalty problem at scales smaller than  $s_f$ . Higher values of the score indicate better performance.

Figure 11 depicts FSS as a function of time of day for a scale  $s_f \approx 30$  km and a low threshold of precipitation intensity. It shows that the application of LHN has a large positive impact on precipitation at the beginning of the forecast. In the 0000 UTC forecast runs, this improvement decreases rather slowly and prevails throughout the forecast range of 24 h. Bootstrapping, as

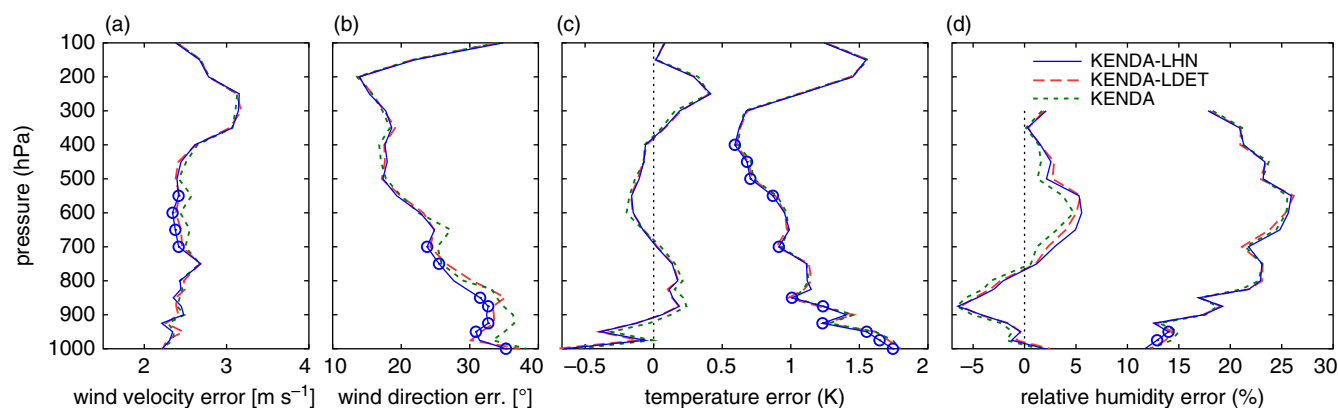
described in section 4.3, indicates that the statistical confidence for an improvement is greater than 95% up to a forecast lead time of 19 h. In the 0600 and 1800 UTC runs, the benefit also lasts for 24 h, even though it tends to be smaller overall (not shown). In the 1200 UTC forecast runs, the positive impact is limited to the first 6 h.

In Figure 12, the sample noise is reduced by averaging the hourly FSS values over the forecast range of 24 h. These averaged FSS scores are displayed as a function of scale  $s_f$  for low, middle and high precipitation thresholds. Overall, the benefit from LHN is similarly large for all scales and thresholds. With respect to the initial time of the forecast runs, the benefit is rather large for the 0600 and 1800 UTC runs, largest for the 0000 UTC runs and small for the 1200 UTC runs. These improvements are statistically significant (in terms of >95% confidence), except for the large precipitation threshold ( $5 \text{ mm h}^{-1}$ ) of the 1200 UTC runs.

The reason for this more limited impact of LHN on the 1200 UTC forecast runs might be related to the fact that COSMO-DE has a warm bias around noon in the summer season (see Figure 13). With its limited horizontal resolution and without parametrization of deep convection, the model requires excessively unstable conditions to produce realistic air-mass convection. In such an environment, assimilating unbiased temperature profile data tends to suppress convection. In spite of this, LHN is able to generate convective precipitation, albeit without destabilizing the convective environment. As a result, the model tends to dissolve the convective cells in the subsequent free forecast and the impact of LHN is comparatively short-lived.

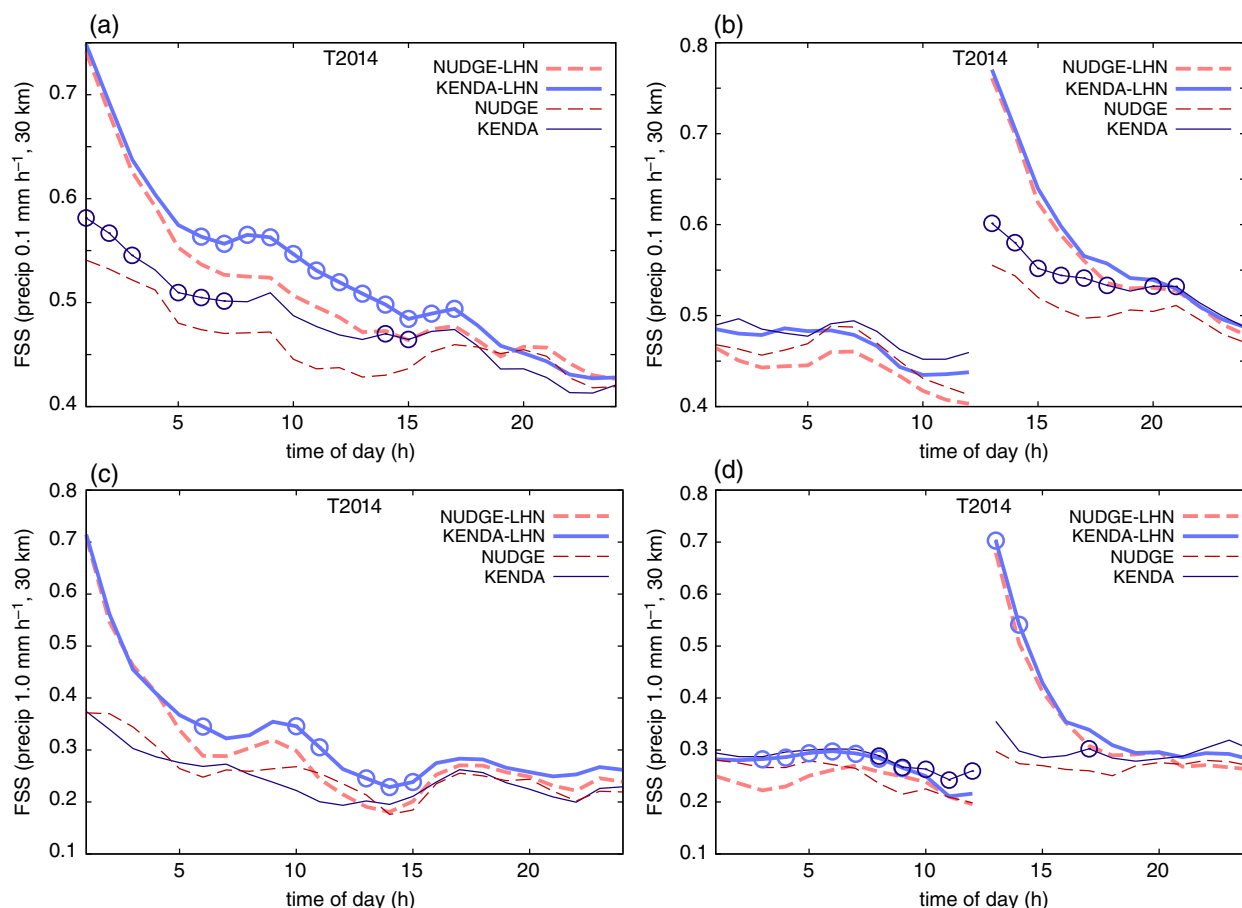
On average, LHN also reduces the upper-air wind (RMS) errors, but has little impact on the humidity errors (Figure 14). The reduction of the upper-air temperature errors is about 3% (percentage improvement of RMSE) and is statistically significant at most vertical levels up to 400 hPa. Small error reductions with a confidence of more than 95% are also obtained for 2 m temperature, 2 m dew-point and surface pressure (see Table 4). It may be noted that the upper-air wind direction errors are mainly decreased at low levels, where speeds are lower. This might reflect improved low-level convergence fields as a result of a more accurate simulation of convective heating and updrafts. The corresponding improved upper-level divergence would then affect wind speed more than direction (as seen in Figure 14), because of the stronger winds aloft.

Compared with KENDA–LHN, KENDA–LDET has a slightly inferior forecast quality overall for precipitation (see Figures 11, 12) and for upper-air wind (Figure 14). The degradation does not attain statistical significance for most thresholds, forecast initial times and scales, but it exceeds 95% confidence for  $5 \text{ mm h}^{-1}$  precipitation in the 1800 UTC runs at all scales. Thus, it is advantageous to apply LHN not only in the deterministic run but also in the ensemble. Regarding the deterministic analysis, the difference between the two experiments is that the Kalman gain is



**Figure 14.** Vertical profiles of 6 h forecast errors against radiosonde observations for the experiments KENDA without LHN (dotted line), KENDA–LDET with LHN in the deterministic run only (dashed line) and KENDA–LHN with LHN additionally in all ensemble members (solid line). The circles on the solid line indicate levels where the confidence for an RMSE reduction of KENDA–LHN over KENDA exceeds 95% according to bootstrapping. Panels from left to right: (a) wind speed (only RMSE), (b) wind direction (only RMSE), (c) temperature (bias and RMSE) and (d) relative humidity (bias and RMSE) for the 28 day period T2014.





**Figure 15.** Verification of one-hourly precipitation against radar-derived precipitation: FSS with a scale of 11 grid points ( $\approx 30$  km) for experiments KENDA-LHN (LETKF + LHN; thick solid line), NUDGE-LHN (nudging + LHN; thick dashed line), KENDA (LETKF without LHN; thin solid line) and NUDGE (nudging without LHN; thin dashed line) for period T2014. The circles on the solid lines indicate forecast ranges where the confidence for an FSS improvement of KENDA-LHN over KENDA (thick line) (resp. KENDA-LHN over KENDA, thin line) exceeds 95% according to bootstrapping. The upper row of panels (a,b) is for a threshold of  $0.1 \text{ mm h}^{-1}$ ; the lower row (c,d) for  $1.0 \text{ mm h}^{-1}$ . The left panels (a,c) are for 0000 UTC forecast runs; the right panels (b,d) for 1200 UTC forecast runs.

influenced by LHN in KENDA-LHN only. Thus, the influence of LHN on the background-error covariances in the LETKF appears to be slightly beneficial (rather than disadvantageous as could have been suspected), at least on average.

#### 5.4. Comparing LETKF with nudging

Here, we compare the overall performance of the LETKF versus nudging. We also compare the benefit latent heat nudging (LHN) has if combined with either of the two methods. For this purpose, results are shown from the following experiments.

- (1) KENDA-LHN: LETKF combined with LHN both in the deterministic run and in all the ensemble members. The LETKF uses our standard settings as described in sections 2.4, 3 and 4.
- (2) NUDGE-LHN: nudging combined with LHN.
- (3) KENDA: as KENDA-LHN, but without LHN.
- (4) NUDGE: as NUDGE-LHN, but without LHN.

Since the current operational data assimilation system for COSMO-DE uses nudging plus latent heat nudging (LHN), this is the benchmark with which we are most interested to compare the LETKF. Therefore, the emphasis of our discussion will be on the comparison between KENDA-LHN and NUDGE-LHN.

Figure 15 reveals that KENDA-LHN provides better precipitation forecasts than NUDGE-LHN almost throughout the forecast range, except for the first 4 h. Here, the immediate impact of the LHN reflected by the high scores at the start of the forecast appears to be dominant and similar for both experiments, so that more subtle analysis differences influence the precipitation scores only later in the forecast. For the  $0.1 \text{ mm h}^{-1}$  threshold

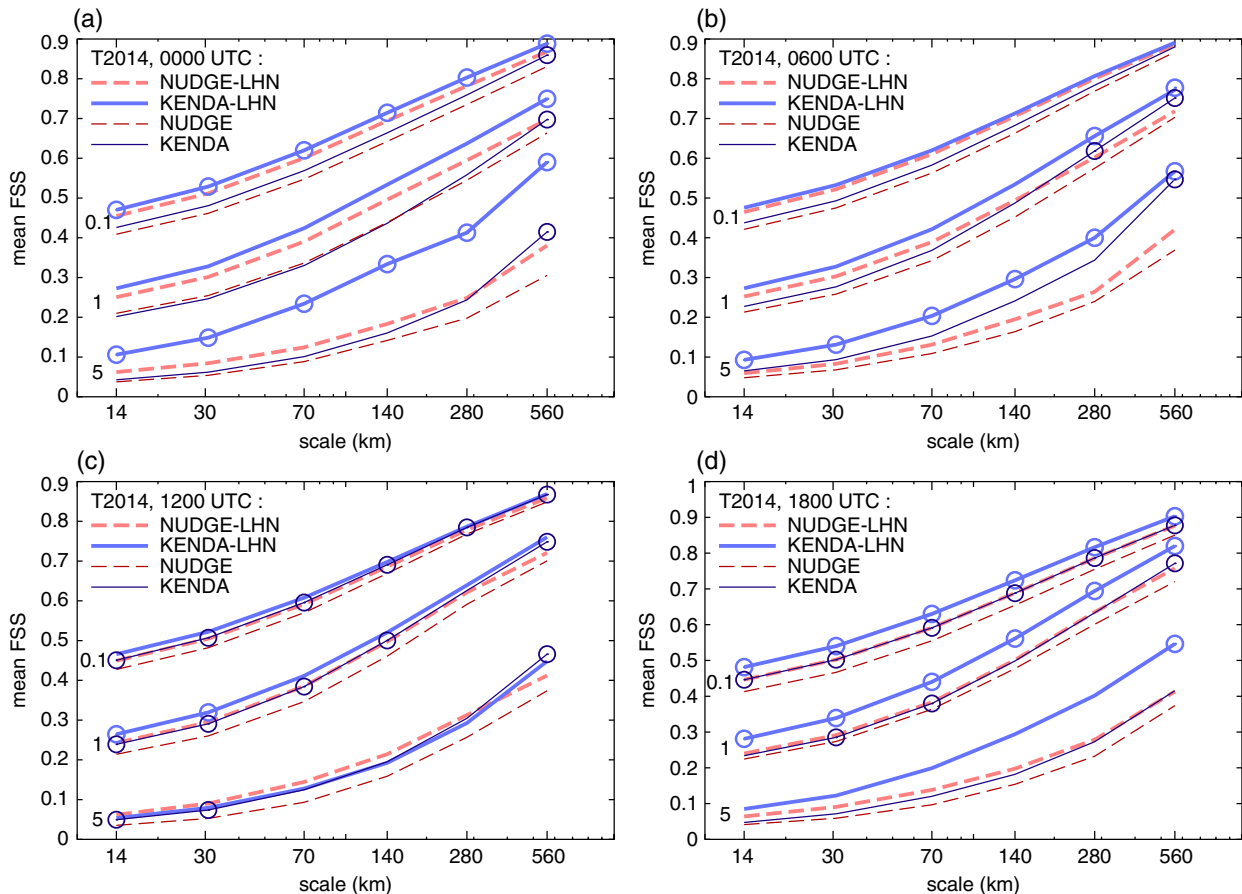
of 0000 UTC runs, for example, the improvement in the FSS is statistically significant (in terms of confidence  $> 95\%$ ) for lead times from +6 to +17 h and the percentage improvement (PI) of FSS amounts to about 8% from +8 to +13 h.

With regard to the forecast quality averaged over the forecast range of 24 h (see Figure 16), KENDA-LHN clearly outperforms NUDGE-LHN for all thresholds, scales, periods and initial times of the forecast runs, except for the 1200 UTC runs, where it is only slightly better for the two lower thresholds and slightly worse for  $5.0 \text{ mm h}^{-1}$ . These results are mostly significant, except for the 1200 UTC runs. Figure 16 also gives an idea of the PI of FSS; for  $5.0 \text{ mm h}^{-1}$  in 0000 UTC runs, the PI varies between 55% and 89%.

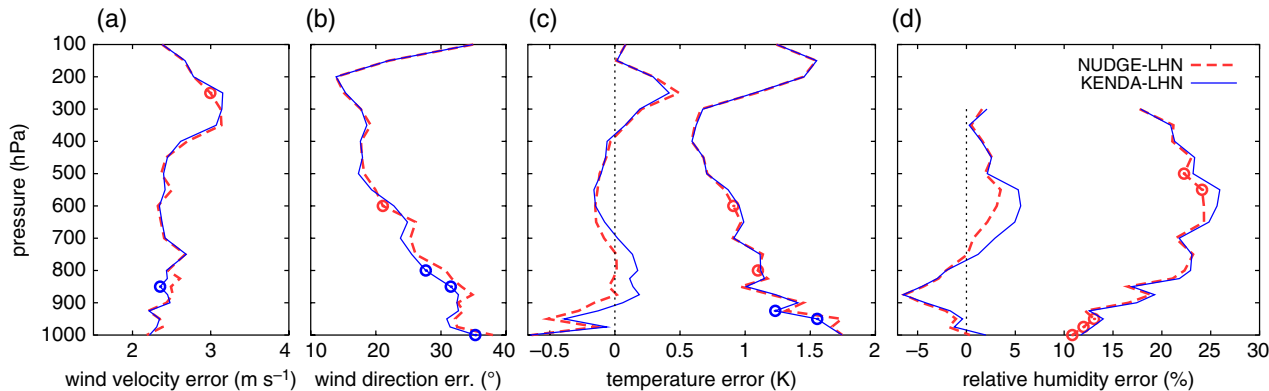
Without LHN, the LETKF experiment KENDA is also clearly better than the nudging experiment NUDGE. However, the advantage of the LETKF is smaller than in the comparison with LHN, except for the 1200 UTC runs, which are affected by the temperature bias issue discussed in section 5.3.

From this, it can also be concluded that LHN has greater positive impact on precipitation forecasts if combined with the LETKF than with nudging. (This is obvious in Figures 15, particularly for the  $1 \text{ mm h}^{-1}$  threshold in 0000 UTC runs, and 16, by comparing the differences between the two solid lines with the differences between the two dashed lines.) This confirms the indication obtained in section 5.3 that the influence of LHN on the background-error covariances in the LETKF tends to be beneficial on average.

In the upper-air verification of the 6 h forecasts against radiosondes (see Figure 17), the biases of KENDA-LHN are similar to those of NUDGE-LHN, except for upper-level humidity and lower-level temperature. A *free forecast* experiment, which cycles without assimilating any observations, has a wet



**Figure 16.** FSS for one-hourly precipitation verified against radar-derived precipitation as a function of scale, where each of the values shown is an average over the 24 FSS values valid for forecast lead times from 1–24 h. The labels for the abscissa denote the (approximate) scales for which the FSS is computed. In each panel, the upper, middle and lower quadruplet of lines are for thresholds of 0.1, 1.0 and 5.0 mm h<sup>−1</sup> respectively. Each quadruplet of lines shows the scores for the experiments KENDA–LHN (LETKF + LHN; thick solid line), NUDGE–LHN (nudging + LHN; thick dashed line), KENDA (LETKF without LHN; thin solid line) and NUDGE (nudging without LHN; thin dashed line). The circles on the solid lines indicate scales where the confidence for an FSS improvement of KENDA–LHN over KENDA (thick line) (resp. KENDA–LHN over KENDA, thin line) exceeds 95%. The panels are for (a) 0000, (b) 0600, (c) 1200 and (d) 1800 UTC forecast runs of the 28 day period T2014 respectively.

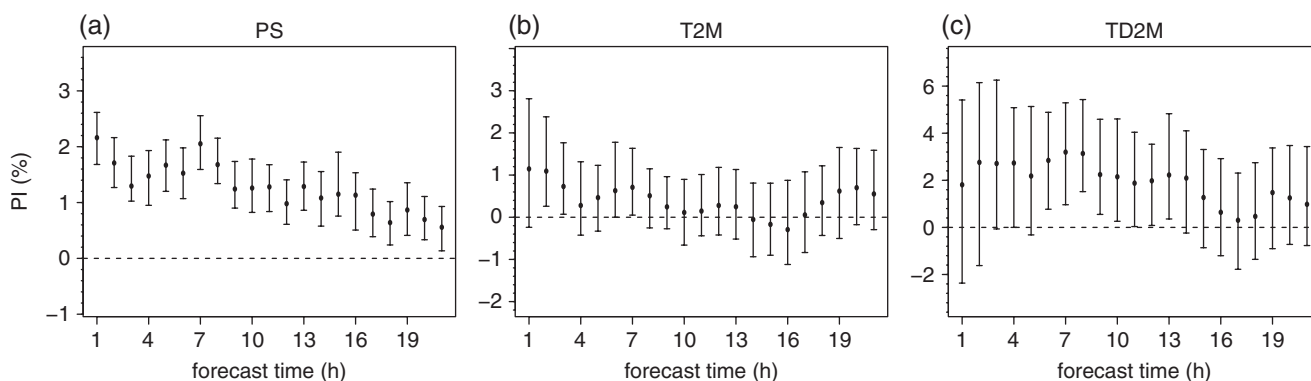


**Figure 17.** Vertical profiles of 6 h forecast errors against radiosonde observations for the 28 day period T2014, as in Figure 14, but for the experiments KENDA–LHN (solid line) and NUDGE–LHN (dashed line). The circles on the solid (dashed) line indicate levels where the confidence for an RMSE reduction (increase) of KENDA–LHN over KENDA exceeds 95%.

bias around 600 hPa in period T2014 (not shown) that is not corrected by the LETKF in KENDA–LHN, but is corrected by nudging in NUDGE–LHN. On the other hand, most of the warm temperature bias in the *free forecast* experiment is corrected by the LETKF. Nudging tends to overdo the correction, so that NUDGE–LHN has a cold bias against aircraft observations, unlike KENDA–LHN (not shown). Apparently, the aircraft data have a warm bias against the radiosonde data on average. It is also interesting to note that radiosonde profiles, as well as most aircraft ascent and descent data, are assimilated as multi-level profile data in the nudging, while LETKF assimilates all data as single-level data for which the vertical influence on the analysis increments can overlap strongly. Since the aircraft data often have a higher

vertical resolution than the radiosonde data in the lower and middle troposphere, the influence of the aircraft data relative to that of the radiosonde data is larger in the LETKF than in the nudging. This may explain the different temperature biases in the two experiments (KENDA–LHN has a smaller temperature bias against aircraft data, NUDGE–LHN against radiosonde data).

In terms of RMSE against radiosonde data, KENDA–LHN is better than NUDGE–LHN for wind direction (percentage improvement PI = 2.1% averaged over all levels up to 300 hPa, i.e. where data are assimilated in the LETKF), similar for wind speed (PI = 0.5%) and temperature (PI = −0.2%) and worse for relative humidity (PI = −3.4%). However, these differences are



**Figure 18.** Forecast verification against observations from surface stations: percentage improvement (PI) of RMSE of experiment KENDA–LHN versus NUDGE–LHN for (a) surface pressure (denoted ‘PS’), (b) 2 m temperature (‘T2M’) and (c) 2 m dew-point depression (‘TD2M’) for the 28 day period T2014. The dots indicate the median of the PI and the error bars the 5th and 95th percentiles from bootstrapping.

Table 4. Forecast verification against observations from surface stations: percentage improvement (PI) of root-mean-square errors (RMSE) of the ‘experiment’ versus the RMSE of the ‘reference’, for surface pressure (denoted ‘PS’), 2 m temperature (‘T2M’) and 2 m dew-point depression (‘TD2M’) for the 28 day period T2014. Each PI value shown is the median from bootstrapping and each RMSE value is an average over the 21 RMSE values valid for the forecast lead times from 1–21 h. For values without brackets, the PI is positive (i.e. indicating an error reduction for the experiment compared with the reference) with more than 95% confidence, whereas the values in brackets (narrowly) miss this criterion.

Experiment	Reference	PS	T2M	TD2M
KENDA–LHN	NUDGE–LHN	1.2	(0.4)	(1.9)
KENDA	NUDGE	1.1	1.1	2.4
KENDA–LHN	KENDA	0.3	1.1	1.6

statistically significant only at a few vertical levels. For larger forecast lead times, the differences between the two experiments are qualitatively similar but somewhat smaller, due to the impact of the identical lateral boundary conditions.

The degradation for humidity is partly related to the larger bias. Further possible reasons include non-Gaussianity of relative humidity or first-guess error cross-correlations with temperature and wind, which are not compliant with the innovations of the humidity, temperature and wind observations. In the latter case, the information from humidity observations will not be assimilated well, given that substantially more temperature and wind data are used than humidity data.

Table 4 summarizes the main results of the verification against surface synoptic stations and Figure 18 details the most important differences between KENDA–LHN and NUDGE–LHN. Experiment KENDA–LHN verifies slightly better for surface pressure and 2 m dew-point. The improvement for 2 m temperature is marginal and the differences in the scores for cloud cover are very small overall (not shown). While the statistical significance of the improvement is attained only at a few forecast hours for dew-point and temperature, it prevails throughout the forecast for surface pressure.

As already mentioned in section 4.3, the results are qualitatively similar in period T2012. As an exception, the surface pressure of KENDA–LHN verifies slightly worse in that period. Particularly in T2012, the LETKF ensemble is found to be underdispersive for surface pressure and this is likely related to underdispersiveness of the lateral boundary conditions from the ICON–LETKF with regard to pressure.

Both in the upper-air and surface verification, similar differences between the LETKF and nudging are found when LHN is not applied to both, except that the improvements of 2 m temperature and 2 m dew-point are larger and significant if averaged over the forecast lead time (see Table 4). In the upper-air verification, the PI averaged over all levels up to 300 hPa is 1.9% for wind direction, 1.4% for wind speed, –1.6% for temperature and –2.7% for relative humidity.

## 6. Conclusions

This study documents the development of the KENDA data assimilation system, which is based on LETKF following Hunt *et al.* (2007) and is envisaged to become the operational data assimilation system at DWD for the regional convective-scale model COSMO-DE in the near future. The KENDA system comprises an advanced suite of various tools and configurations, which can be set as required. These include adaptive observation-error estimation, adaptive localization, adaptive multiplicative covariance inflation, relaxation to prior perturbations and relaxation to prior spread. In addition, the highly flexible cycling environment BACY, which offers a fast and user-friendly test environment, has been introduced. Its portability to any Linux system makes it a useful tool for research at universities.

In this study, various configurations have been tested and deterministic forecasts started from deterministic analyses of the LETKF system have been compared with forecasts based on the currently operational observation nudging scheme at DWD. In both schemes, only conventional observations were assimilated. It is shown that sufficient spread in both the lateral and lower boundary conditions is important for the performance of the LETKF system. Explicit soil moisture perturbations have a beneficial impact on the spread and forecast scores in the boundary layer.

Since a direct assimilation of radar reflectivity is not yet ready for operational use, a novel approach has been adopted in this study: in order to assimilate radar precipitation, latent heat nudging (LHN) has been incorporated into the LETKF cycle. One of the core results is that LHN works well in combination with the LETKF, in particular if it is applied to all ensemble members. With this set-up, LHN even has slightly more benefit in combination with LETKF than with observation nudging. This indicates that the influence of LHN on the background-error covariances in LETKF tends to be beneficial on average.

Compared with the standard nudging scheme, the KENDA–LETKF system yields superior deterministic forecasts for almost all variables and for precipitation in particular. This is true without the use of LHN, but when LHN is added to both, the advantage of KENDA–LETKF tends to be even larger. Remaining problems with slightly higher errors are found for surface pressure in one of the test periods and for upper-air humidity. The former might be related to the global driving ensemble being underdispersive for pressure. Work on the global LETKF system is ongoing and improvements in that system will be beneficial to the regional system. The problems with upper-air humidity are partly related to the finding that LETKF is less able than nudging to correct humidity biases. They may also be linked to errors in the estimation of the first-guess error cross-correlations between humidity and other observed variables such as temperature and wind in the LETKF. Sampling errors due to limited ensemble size and model errors may contribute to this.



The experiments shown here cover two summer periods of 6 and 28 days continuous cycling. The KENDA system is also found to work well in some other, shorter periods with different synoptic situations (not shown here). A continuous, pre-operational online cycling is scheduled to start in March or April 2016. Before this, it is also planned to test additive covariance inflation, since this has often been found to have benefits in other EnKF systems (e.g. Whitaker *et al.*, 2008; Li *et al.*, 2009; Bonavita *et al.*, 2010).

A further goal is to enhance the use of observations in KENDA, in particular high-resolution remote sensing data providing information on the planetary boundary layer, the humidity field and weather parameters such as cloud and precipitation. A number of research projects and studies based on the KENDA–LETKF system are ongoing in view of this, both at DWD and also at universities, for example within the Hans Ertel Centre for Weather Research<sup>†</sup> (HERZ: Weissmann *et al.*, 2014). Forward operators have been developed at DWD for three-dimensional radar radial velocity (Zeng, 2014) and reflectivity (Jerger, 2014), ground-based Global Navigation Satellite System (GNSS) slant path delay data, Spinning Enhanced Visible and InfraRed Imager (SEVIRI) derived cloud-top height (Schomburg *et al.*, 2015) and SEVIRI VIS/NIR radiances (at HERZ: Kostka *et al.*, 2014). Testing and optimization of the use of these data is ongoing. The same holds for the use of Mode-S aircraft data (at HERZ: Lange and Janjic, 2015), as well as the direct assimilation of SEVIRI IR radiances (using the Radiative Transfer for TIROS Operational Vertical sounders (RTTOV) operator) in view of assimilating cloud information. This also makes use of the experience of Otkin (2012), who assimilated satellite radiances in an EnKF framework in an Observing System Simulation Experiment (OSSE) study. Another study is the development and test of the forecast sensitivity to observations (FSO) in the KENDA–LETKF framework (within HERZ: Sommer and Weissmann, 2014).

When incorporating high-resolution remote sensing data in the future, the analysis update interval of 1 h may be decreased to 30 min or even less, as these data are available at high temporal resolution. Moreover, when the updates are applied more frequently they become more linear. The noise generated in the analysis levels off in less than 30 min (not shown), hence at least at this cycling frequency our system works well without any digital filtering or incremental analysis updating. It is noted that most of these high-resolution data are indirect observations and LETKF has the additional advantage over nudging of offering much better prospects for their future operational use. Firstly, it avoids the need to compute retrievals. Secondly, it provides and uses situation-dependent multivariate covariances and this has particular potential for the assimilation of data related to cloud, precipitation, humidity and the planetary boundary layer.

Furthermore, it is planned to combine the KENDA ensemble data assimilation with the convective-scale ensemble prediction system COSMO-DE-EPS (Gebhardt *et al.*, 2011; Peralta *et al.*, 2012) by using the KENDA analysis ensemble as initial condition for the ensemble forecasts (Harnisch and Keil, 2015), likely in combination with additional stochastic physics perturbations.

## Acknowledgements

We thank Alexander Cress for performing additional ICON–LETKF experiments, which were dedicated to providing the lateral boundary conditions for some of the KENDA experiments performed for this study. We also thank Liselotte

Bach for quickly providing us with a verification against surface stations that included bootstrapping. Finally, we acknowledge Richard Keane and Liselotte Bach for proof-reading.

## References

- Aksoy A, Dowell D, Snyder C. 2009. A multicaser comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Weather Rev.* **137**: 1805–1824, doi: 10.1175/2008MWR2691.1.
- Anderson J. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**: 2884–2903, doi: 10.1175/1520-0493(2001)129<2884:aeakff>2.0.CO;2.
- Anderson JL. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A* **59**: 210–224, doi: 10.1111/j.1600-0870.2006.00216.x.
- Anderson JL. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A* **61**: 72–83, doi: 10.1111/j.1600-0870.2008.00361.x.
- Anderson JL. 2012. Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Weather Rev.* **140**: 2359–2371, doi: 10.1175/MWR-D-11-00013.1.
- Anderson J, Anderson S. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**: 2741–2758, doi: 10.1175/1520-0493(1999)127<2741:amciot>2.0.CO;2.
- Anthes R. 1974. Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.* **31**: 702–719, doi: 10.1175/1520-0469(1974)031<0702:daaioh>2.0.CO;2.
- Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Weather Rev.* **139**: 3887–3905, doi: 10.1175/MWR-D-10-05013.1.
- Berre L, Desroziers G. 2010. Filtering of background error variances and correlations by local spatial averaging: A review. *Mon. Weather Rev.* **138**: 3693–3720, doi: 10.1175/2010MWR3111.1.
- Bishop CH, Hodyss D. 2007. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 2029–2044, doi: 10.1002/qj.169.
- Bishop CH, Hodyss D. 2009a. Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus A* **61**: 84–96, doi: 10.1111/j.1600-0870.2008.00371.x.
- Bishop CH, Hodyss D. 2009b. Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus A* **61**: 97–111, doi: 10.1111/j.1600-0870.2008.00372.x.
- Bishop C, Etherton B, Majumdar S. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather Rev.* **129**: 420–436, doi: 10.1175/1520-0493(2001)129<0420:aswtet>2.0.CO;2.
- Bonavita M, Torrisi L, Marcucci F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q. J. R. Meteorol. Soc.* **136**: 132–145, doi: 10.1002/qj.533.
- Bonavita M, Holm E, Isaksen I, Fisher M. 2014. *The Evolution of the ECMWF Hybrid Data Assimilation System, Technical Memorandum 743*. ECMWF: Reading, England. [www.ecmwf.int/en/library/8271-evolution-ecmwf-hybrid-data-assimilation-system](http://www.ecmwf.int/en/library/8271-evolution-ecmwf-hybrid-data-assimilation-system) (accessed 29 February 2016).
- Buehner M. 2005. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**: 1013–1043, doi: 10.1256/qj.04.15.
- Buehner M, Houtekamer P, Charette C, Mitchell H, He B. 2010. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Weather Rev.* **138**: 1567–1586, doi: 10.1175/2009MWR3158.1.
- Burgers G, van Leeuwen P, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**: 1719–1724, doi: 10.1175/1520-0493(1998)126<1719:asitek>2.0.CO;2.
- Clayton A, Lorenc A, Barker D. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**: 1445–1461, doi: 10.1002/qj.2054.
- Davies H, Turner R. 1977. Updating prediction models by dynamical relaxation – examination of technique. *Q. J. R. Meteorol. Soc.* **103**: 225–245, doi: 10.1002/qj.49710343602.
- Dee D. 1995. Online estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.* **123**: 1128–1145, doi: 10.1175/1520-0493(1995)123<1128:oleoc>2.0.CO;2.
- Desroziers G, Brousseau P, Chapnik B. 2005a. Use of randomization to diagnose the impact of observations on analyses and forecasts. *Q. J. R. Meteorol. Soc.* **131**: 2821–2837, doi: 10.1256/qj.04.151.
- Desroziers G, Berre L, Chapnik B, Poli P. 2005b. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396, doi: 10.1256/qj.05.108.
- Doms G, Förstner J, Heise E, Herzog HJ, Mironov D, Raschendorfer M, Reinhardt T, Ritter B, Schröder R, Schulz JP, Vogel G. 2011. ‘A description of the nonhydrostatic regional COSMO model, Part II: Physical parameterization’.

<sup>†</sup>The Hans Ertel Centre for Weather Research (HERZ) is a virtual centre including several German universities and the Max-Planck-Institute for Meteorology. It is funded since 2011 through DWD. The HERZ branch on Data Assimilation is located at LMU Munich and uses the KENDA system as its main research platform.

- www.cosmo-model.org/content/model/documentation/core (accessed 29 February 2016).
- Dow G, Macpherson B. 2013. 'Benefit of convective-scale data assimilation and observing systems in the UK models', Forecasting Research Technical Report 585. Met Office. [www.metoffice.gov.uk/media/pdf/t/l/FRT585.pdf](http://www.metoffice.gov.uk/media/pdf/t/l/FRT585.pdf) (accessed 29 February 2016).
- Dowell D, Wicker L, Snyder C. 2011. Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analysis. *Mon. Weather Rev.* **139**: 272–294, doi: 10.1175/2010MWR3438.1.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162, doi: 10.1029/94JC00572.
- Evensen G, van Leeuwen P. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**: 1852–1867, doi: 10.1175/1520-0493(2000)128<1852:aeksf>2.0.CO;2.
- Gaspari G, Cohn S. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757, doi: 10.1002/qj.4971255417.
- Gebhardt C, Theis S, Paulat M, Ben Boualleue Z. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.* **100**: 168–177, doi: 10.1016/j.atmosres.2010.12.008.
- Gustafsson N, Bojarova J. 2014. Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Proc. Geophys.* **21**: 745–762, doi: 10.5194/npg-21-745-2014.
- Hamill T, Whitaker J, Snyder C. 2001. Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790, doi: 10.1175/1520-0493(2001)129<2776:ddfobe>2.0.CO;2.
- Hamrud M, Bonavita M, Isaksen L. 2014. *EnKF and Hybrid Gain Ensemble Data Assimilation. Technical Memorandum 733*. ECMWF: Reading, England. [www.ecmwf.int/sites/default/files/elibrary/2014/9766-enkf-and-hybrid-gain-ensemble-data-assimilation.pdf](http://www.ecmwf.int/sites/default/files/elibrary/2014/9766-enkf-and-hybrid-gain-ensemble-data-assimilation.pdf) (accessed 29 February 2016).
- Harnisch F, Keil C. 2015. Initial conditions for convective-scale ensemble forecasting provided by ensemble data assimilation. *Mon. Weather Rev.* **143**: 1583–1600, doi: 10.1175/MWR-D-14-00209.1.
- Hengstebek T, Helmert K, Seltmann J. 2010. 'RadarQS – a standard quality control software for radar data at DWD'. In *Proceedings of the 6th European Conference on Radar in Meteorology and Hydrology*, Sibiu, Romania.
- Houtekamer P, Mitchell H. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**: 796–811, doi: 10.1175/1520-0493(1998)126<0796:dauaek>2.0.CO;2.
- Houtekamer P, Mitchell H. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137, doi: 10.1175/1520-0493(2001)129<0123:asekff>2.0.CO;2.
- Houtekamer P, Mitchell H, Pellerin G, Buehner M, Charron M, Spacek L, Hansen M. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Weather Rev.* **133**: 604–620, doi: 10.1175/MWR-2864.1.
- Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230**: 112–126, doi: 10.1016/j.physd.2006.11.008.
- Janjić T, Nerger L, Albertella A, Schroeter J, Skachko S. 2011. On domain localization in ensemble-based Kalman filter algorithms. *Mon. Weather Rev.* **139**: 2046–2060, doi: 10.1175/2011MWR3552.1.
- Jerger D. 2014. *Radar Forward Operator for Verification of Cloud Resolving Simulations within the COSMO model*. KIT Scientific Publishing: Karlsruhe, Germany, doi: 10.5445/KSP/1000038411.
- Kirchgessner P, Nerger L, Bunse-Gerstner A. 2014. On the choice of an optimal localization radius in ensemble Kalman filter methods. *Mon. Weather Rev.* **142**: 2165–2175, doi: 10.1175/MWR-D-13-00246.1.
- Kostka PM, Weissmann M, Buras R, Mayer B, Stiller O. 2014. Observation operator for visible and near-infrared satellite reflectances. *J. Atmos. Oceanic Technol.* **31**: 1216–1233, doi: 10.1175/JTECH-D-13-00116.1.
- Lange H, Janjić T. 2015. Assimilation of Mode-S aircraft observations in COSMO-KENDA. *Mon. Weather Rev.*, doi: 10.1175/MWR-D-15-0112.1.
- Lei L, Stauffer D, Deng A. 2012. A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. *Q. J. R. Meteorol. Soc.* **138**: 2066–2078, doi: 10.1002/qj.1939.
- Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 523–533, doi: 10.1002/qj.371.
- Lin YL, Farley R, Orville H. 1983. Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Meteorol.* **22**: 1065–1092, doi: 10.1175/1520-0450(1983)022<1065:bpotsf>2.0.CO;2.
- Liu C, Xiao Q, Wang B. 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Weather Rev.* **136**: 3363–3373, doi: 10.1175/2008MWR2312.1.
- Majewski D, Liermann D, Prohl P, Ritter B, Buchhold M, Hanisch T, Paul G, Wergen W. 2002. The operational global icosahedral hexagonal gridpoint model GME: Description and high-resolution tests. *Mon. Weather Rev.* **130**: 319–338, doi: 10.1175/1520-0493(2002)130<0319:TOGIHG>2.0.CO;2.
- Meng Z, Zhang F. 2008. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Weather Rev.* **136**: 522–540, doi: 10.1175/2007MWR2106.1.
- Meng Z, Zhang F. 2011. Limited-area ensemble-based data assimilation. *Mon. Weather Rev.* **139**: 2025–2045, doi: 10.1175/2011MWR3418.1.
- Migliorini S. 2013. Information-based data selection for ensemble data assimilation. *Q. J. R. Meteorol. Soc.* **139**: 2033–2054, doi: 10.1002/qj.2104.
- Mittermaier M, Roberts N. 2010. Inter-comparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Weather and Forecasting* **25**: 343–354, doi: 10.1175/2009WAF222260.1.
- Otkin J. 2012. Assimilation of water vapor sensitive infrared brightness temperature observations during a high impact weather event. *J. Geophys. Res.* **117**: D19203, doi: 10.1029/2012JD017568.
- Ott E, Hunt B, Szunyogh I, Zimin A, Kostelich E, Corazza M, Kalnay E, Patil D, Yorke J. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* **56**: 415–428, doi: 10.1111/j.1600-0870.2004.00076.x.
- Peralta C, Ben Boualleue Z, Theis S, Gebhardt C, Buchhold M. 2012. Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.* **117**: D07108, doi: 10.1029/2011JD016581.
- Periáñez A, Reich H, Potthast R. 2014. Optimal localization for ensemble Kalman filter. *J. Meteorol. Soc. Jpn.* **92**: 585–597, doi: 10.2151/jmsj.2014-605.
- Raschendorfer M. 2001. The new turbulence parametrization of LM. *COSMO Newsl.* **1**: 89–97.
- Ritter B, Geleyn JF. 1992. A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Weather Rev.* **120**: 303–325, doi: 10.1175/1520-0493(1992)120<0303:acrsfn>2.0.CO;2.
- Roberts N, Lean H. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Weather Rev.* **136**: 78–96, doi: 10.1175/2007MWR2123.1.
- Schomburg A, Schraff C, Potthast R. 2015. A concept for the assimilation of satellite cloud information in an ensemble Kalman filter: Single-observation experiments. *Q. J. R. Meteorol. Soc.* **141**: 893–908, doi: 10.1002/qj.2407.
- Schraff C. 1997. Mesoscale data assimilation and prediction of flow stratus in the Alpine region. *Meteorol. Atmos. Phys.* **64**: 21–50, doi: 10.1007/BF01044128.
- Schraff C, Hess R. 2012. 'A description of the nonhydrostatic regional COSMO-model, Part III: Data assimilation'. [www.cosmo-model.org/content/model/documentation/core](http://www.cosmo-model.org/content/model/documentation/core) (accessed 29 February 2016).
- Seifert A, Beheng K. 2001. A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmos. Res.* **59**: 265–281, doi: 10.1016/S0169-8095(01)00126-0.
- Seity Y, Brousseau P, Malardel S, Hello G, Benard P, Bouttier F, Lac C, Masson V. 2011. The AROME-France convective-scale operational model. *Mon. Weather Rev.* **139**: 976–991, doi: 10.1175/2010MWR3425.1.
- Sommer M, Weissmann M. 2014. Observation impact in a convective-scale localized ensemble transform Kalman filter. *Q. J. R. Meteorol. Soc.* **140**: 2672–2679, doi: 10.1002/qj.2343.
- Stauffer D, Seaman N. 1990. Use of 4-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Weather Rev.* **118**: 1250–1277, doi: 10.1175/1520-0493(1990)118<1250:uofdda>2.0.CO;2.
- Stauffer D, Seaman N. 1994. Multiscale 4-dimensional data assimilation. *J. Appl. Meteorol.* **33**: 416–434, doi: 10.1175/1520-0450(1994)033<0416:mfdada>2.0.CO;2.
- Stephan K, Klink S, Schraff C. 2008. Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Q. J. R. Meteorol. Soc.* **134**: 1315–1326, doi: 10.1002/qj.269.
- Szunyogh I, Kostelich E, Gyarmati G, Patil D, Hunt B, Kalnay E, Ott E, Yorke J. 2005. Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus A* **57**: 528–545, doi: 10.1111/j.1600-0870.2005.00136.x.
- Szunyogh I, Kostelich EJ, Gyarmati G, Kalnay E, Hunt BR, Ott E, Satterfield E, Yorke JA. 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A* **60**: 113–130, doi: 10.1111/j.2007.00274.x.
- Tang Y, Lean H, Bornemann J. 2013. The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorol. Appl.* **20**: 417–426, doi: 10.1002/met.1300.
- Tiedtke M. 1989. A comprehensive mass flux scheme for cumulus parametrization in large-scale models. *Mon. Weather Rev.* **117**: 1779–1799, doi: 10.1175/1520-0493(1989)117<1779:acmfsf>2.0.CO;2.
- Torn RD, Hakim GJ. 2008. Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Weather Rev.* **136**: 3947–3963, doi: 10.1175/2008MWR2443.1.
- Wang X, Bishop C. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**: 1140–1158, doi: 10.1175/1520-0469(2003)060<1140:acobaes>2.0.CO;2.
- Wang X, Barker D, Snyder C, Hamill T. 2008. A hybrid ETKF-3DVar data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Weather Rev.* **136**: 5132–5147, doi: 10.1175/2008MWR2445.1.
- Weissmann M, Göber M, Hohenegger C, Janjić T, Keller J, Ohlwein C, Seifert A, Trömel S, Ulbrich T, Wapler K, Bollmeyer C, Deneke H. 2014. Initial phase of the Hans-Ertel Centre for Weather Research: A virtual centre at the interface of basic and applied weather and climate research. *Meteorol. Z.* **23**: 193–208, doi: 10.1127/0941-2948/2014/0558.

- Whitaker J, Hamill T. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**: 1913–1924, doi: 10.1175/1520-0493(2002)130<1913:edawpo>2.0.CO;2.
- Whitaker JS, Hamill TM. 2012. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev.* **140**: 3078–3089, doi: 10.1175/MWR-D-11-00276.1.
- Whitaker J, Hamill T, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Weather Rev.* **136**: 463–482, doi: 10.1175/2007MWR2018.1.
- Wicker L, Skamarock WC. 2002. Time-splitting methods for elastic models using forward time schemes. *Mon. Weather Rev.* **130**: 2088–2097, doi: 10.1175/1520-0493(2002)130<2088:tsmfem>2.0.CO;2.
- Zängl G, Reinert D, Ripodas P, Baldauf M. 2015. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Q. J. R. Meteorol. Soc.* **141**: 563–579, doi: 10.1002/qj.2378.
- Zeng Y. 2014. *Efficient Radar Forward Operator for Operational Data Assimilation within the COSMO Model*. KIT Scientific Publishing: Karlsruhe, Germany, doi: 10.5445/KSP/1000036921.
- Zhang F, Snyder C, Sun J. 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **132**: 1238–1253, doi: 10.1175/1520-0493(2004)132<1238:ioieao>2.0.CO;2.
- Zhu K, Pan Y, Xue M, Wang X, Whitaker J, Benjamin S, Weygandt S, Hu M. 2013. A regional GSI-based ensemble Kalman filter data assimilation system for the rapid refresh configuration: Testing at reduced resolution. *Mon. Weather Rev.* **141**: 4118–4139, doi: 10.1175/MWR-D-13-00039.1.
- Zupanski D, Hou AY, Zhang SQ, Zupanski M, Kummerow CD, Cheung SH. 2007. Applications of information theory in ensemble data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 1533–1545, doi: 10.1002/qj.123.