JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001546

Representation of Model Error in Convective-Scale Data Assimilation: Additive Noise Based on Model Truncation Error

Yuefei Zeng[1,2], Tijana Janjić[2], Matthias Sommer[1,2], Alberto de Lozar[3], Ulrich Blahak[3], and Axel Seifert[3]

[1]Meteorologisches Institut, Ludwig-Maximilians-Universität München, Munich, Germany, [2]Hans Ertel Centre for Weather Research, Deutscher Wetterdienst, Offenbach, Germany, [3]Deutscher Wetterdienst, Offenbach, Germany

Key Points:
- Additive noise based on model truncation error is used to represent small-scale model error
- The combination of large- and small-scale noise outperforms the application of large-scale noise only in short-term forecasts
- The outperformance is especially significant in weak forcing situations

Correspondence to:
Y. Zeng,
yuefei.zeng@lmu.de

Citation:
Zeng, Y., Janjic, T., Sommer, M., de Lozar, A., Blahak, U., & Seifert, A. (2019). Representation of model error in convective-scale data assimilation: Additive noise based on model truncation error. *Journal of Advances in Modeling Earth Systems*, *11*, 752–770. https://doi.org/10.1029/2018MS001546

Received 29 OCT 2018
Accepted 20 FEB 2019
Accepted article online 21 FEB 2019
Published online 20 MAR 2019

Abstract To account for model error on multiple scales in convective-scale data assimilation, we incorporate the small-scale additive noise based on random samples of model truncation error and combine it with the large-scale additive noise based on random samples from global climatological atmospheric background error covariance. A series of experiments have been executed in the framework of the operational Kilometre-scale ENsemble Data Assimilation system of the Deutscher Wetterdienst for a 2-week period with different types of synoptic forcing of convection (i.e., strong or weak forcing). It is shown that the combination of large- and small-scale additive noise is better than the application of large-scale noise only. The specific increase in the background ensemble spread during data assimilation enhances the quality of short-term 6-hr precipitation forecasts. The improvement is especially significant during the weak forcing period, since the small-scale additive noise increases the small-scale variability which may favor occurrence of convection. It is also shown that additional perturbation of vertical velocity can further advance the performance of combination.

1. Introduction

For ensemble Kalman filter (Evensen, 1994), explicit mechanisms are required to mitigate the underestimation of background error covariances that are limited by the ensemble size and model imperfections. One of the well-established techniques is additive covariance inflation, also called additive noise, which modifies the background/analysis error covariance by adding a covariance matrix or by adding a perturbation to each ensemble member. The perturbed ensemble may introduce an error-growing subspace that may be missed in the original ensemble (Hamill & Whitaker, 2005). To generate a perturbed ensemble, Houtekamer et al. (2005) drew random samples from the Three-Dimensional VARiational (3DVAR) background error covariance that represents the large-scale and barotropic structure. Whitaker et al. (2008) randomly selected samples based on differences between adjacent 6-hr analyses from a subset of National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis to emphasize growing baroclinic synoptic-scale structures in middle latitudes. Yang et al. (2015) used the leading ensemble singular vectors derived from ensemble forecasts of quasi-geostrophic multilayer channel model as additive inflation to correct the fastest-growing errors. Reviews of treatment of model error through covariance inflation can be found in Meng and Zhang (2011) and Houtekamer and Zhang (2016).

In the context of convective-scale data assimilation, Snyder and Zhang (2003) assimilated radar observations of radial wind in an idealized setup with 3-D wind and potential temperature of initial ensemble perturbed by grid point Gaussian noise either throughout the domain or around the known storm location, and they found the latter one produced less spurious convective cells. Rather than grid point noise, Dowell, Zhang, et al. (2004) initialized the ensemble by adding smooth ellipsoidal perturbations to horizontal wind, potential temperature, rainwater, and total water and achieved faster updraft and more spread throughout the assimilation than Snyder and Zhang (2003). Furthermore, Dowell, Wicker, and Stensrud (2004) added warm bubbles at random locations close to the storm in the initial ensemble to trigger convective cells. In contrast to those studies that relied on a priori knowledge about the locations of storms, Caya et al. (2005) used observed reflectivities as indicator for storms and added localized smooth noise to the initial ensemble. Furthermore, Dowell and Wicker (2009) introduced random smoothed noise to horizontal

wind, potential temperature, and humidity at places with observed high reflectivities (20 dBZ) throughout assimilation cycles and showed that this technique produced appropriately large spread within the convective cells and small spread in the environment during data assimilation. Later on, Sobash and Wicker (2015) argued that magnitudes of smoothed noise in Dowell and Wicker (2009) can vary significantly depending on the smoothing length scales. Noise of large magnitude is effective to spin-up convective systems while noise of small magnitude is preferable when convective systems become already maturely established in analyses. The repeated addition of large noise may introduce temperature and moisture biases to the surface cold pool and to the tropopause. Therefore, Sobash and Wicker (2015) suggested adding noise where the reflectivity innovation exceeded a threshold value. By using this sort of adaptive technique, they were able to reduce thermodynamic biases. Zeng et al. (2018) showed that large-scale additive noise (hereafter denoted by "LAN") that is based on random samples from climatological atmospheric background error covariance used by the global EnVar data assimilation system may partially represent large-scale error arising from the global driving model ICON (ICOsahedral Nonhydrostatic; Zängl et al., 2015). The LAN performs equally or even better than the relaxation methods RTPP (relaxation to prior perturbations; Zhang et al., 2004) and RTPS (relaxation to prior spread; Whitaker & Hamill, 2012) as well as combinations under strong forcing weather conditions, but the performance of the LAN degrades a bit under weak forcing conditions, assumedly due to being less representative for small-scale features.

A number of studies demonstrated that increasing model resolution could improve the performance of ensemble-based data assimilation and forecasting system (Buizza et al., 2005; Lei & Whitaker, 2017; Pellerin et al., 2003). The refinement of resolution may resolve not only new phenomena but also enable their nonlinear interactions with large-scale motions due to multiscale properties of atmospheric dynamics (Navarra et al., 2010). For the convective-scale precipitation forecasts, Lean et al. (2008) showed that a delay may occur in the convection initiation because the model fails to reproduce small initial convective plumes due to the large viscosity at a coarser resolution. Buzzi et al. (2014) argued that the finer resolution results in more accurate small-scale features such as low-level convergence and orographic lifting that in turn determines the onset of convection. Obviously, convective circulations that are permitted but not properly resolved by models are associated with large errors at the smallest scales allowed by the limited resolution. Hence, model truncation error is considered as one of the important sources of model error for convection. To account for model truncation error in the ensemble data assimilation scheme for a primitive equation global model, Hamill and Whitaker (2005) used samples from a time series of differences between model forecasts at different resolutions in the fashion of additive noise. They showed that this method is superior to additive noise based on samples of 24-hr forecast tendencies and of model state's anomaly from the model climatology. This additive noise method based on model truncation error is hereafter denoted by "SAN" (small-scale noise) since it is expected to somehow represent small-scale or unresolved model error. In this work, we incorporate the SAN into the data assimilation scheme of Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) in the operational Kilometre-scale ENsemble Data Assimilation (KENDA) system of the Deutscher Wetterdienst (DWD; Schraff et al., 2016). The operational numerical weather prediction (NWP) model is the convection permitting COnsortium for Small-scale MOdelling (COSMO, Baldauf et al., 2011) model. We will investigate the performance of the LAN, SAN, and combinations of both in the context of convective-scale data assimilation. Similar to Zeng et al. (2018), we choose a 2-week period in Germany with different types of synoptic forcing. The application of the SAN is expected to be especially useful for the weak forcing conditions.

The paper is organized as follows. Section 2 demonstrates how to create a set of samples for model truncation error of the COSMO model. We investigate as well their statistical properties and kinetic energy spectra. Section 3 shortly describes the experimental setup. Section 4 compares the performance of different experiments, and section 5 summarizes the obtained results.

2. Additive Noise Based on Model Truncation Error

2.1. Methodology

Similar to Hamill and Whitaker (2005), differences between forecasts that are valid at the same time but equipped with different resolutions are treated as samples for model truncation error:

$$\boldsymbol{\eta} = \mathcal{T}\{\mathcal{M}^H[\boldsymbol{x}^H(t_k - t)]\} - \mathcal{M}^L\{\mathcal{T}[\boldsymbol{x}^H(t_k - t)]\}, \quad (1)$$

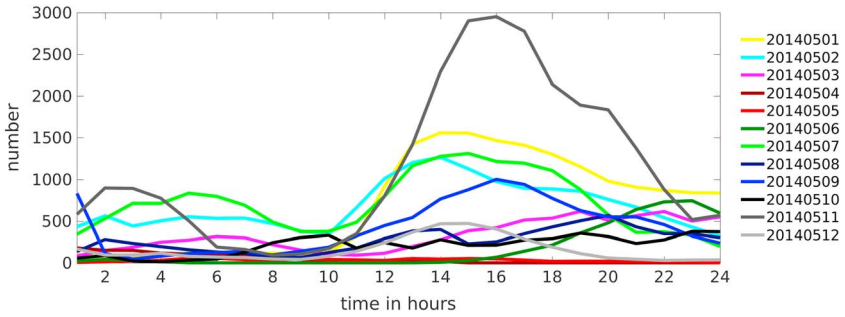

Figure 1. Hourly variations in numbers of model grid points with precipitation rate ≥ 5.0 mm/hr during the training period. All data are derived from the mean of ensemble forecasts of COSMO-DE.

where $\mathcal{T}$ is the interpolation operator from the high to the low resolution, x^H is model state at the high resolution, and $\mathcal{M}^H$ and $\mathcal{M}^L$ are the integration operators of model at high and low resolutions, respectively. The forecast time interval t is predefined, and the valid time t_k is randomly chosen. Therefore, the model truncation error is formed by differences between the truncated forecast from a high-resolution model using a high-resolution initial condition and a low-resolution forecast from a truncated initial condition. It may describe the (short-term) forecast error at resolved scales owing to the lack of ability to model activities at unresolved scales.

The operational horizontal resolution of the COSMO-DE model is 2.8 km. To create a set of samples for model truncation error, the COSMO-DE model is run with a high resolution of 1.4 km for a training period from 1 May 2014 at 00:00 UTC to 12 May 2014 at 00:00 UTC. During the period, a low-pressure cyclone affected a large area of southeastern and central Europe, causing severe thunderstorms with heavy rain as indicated by hourly variations of numbers of model grid points with precipitation rate ≥ 5.0 mm/hr for each day in Figure 1. The hourly outputs of forecasts at resolution of 1.4 km are then interpolated (via the program INT2LM Schättler, 2013) onto the coarse grid with the lower (operational) resolution of 2.8 km. These interpolated fields serve as initial conditions for 24-hr COSMO-DE forecast runs at a resolution of 2.8 km (see Figure 2). Lateral boundary conditions are provided by Integrated Forecast System (IFS) analysis every 6 hr. Therefore, for any chosen forecast time interval $t \in \{1, \ldots, 24\}$ hr, there is a corresponding set of samples for model truncation error. For instance, for $t = 1$ hr, the set is composed of 288 samples.

At the analysis step, for each ensemble member a sample $\eta^{(i)}$ is randomly drawn from the set and added to the analysis ensemble member $x^{a(i)}$ as follows:

$$\boldsymbol{x}^{a(i)} \leftarrow \boldsymbol{x}^{a(i)} + \alpha_S \boldsymbol{\eta}^{(i)}, \tag{2}$$

where α_S is a tunable parameter. For instance, Hamill and Whitaker (2005) chose $\alpha_S = 1.20$ by experimentation.

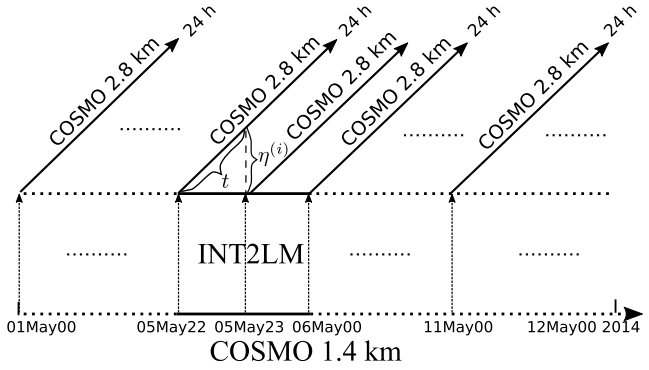

Figure 2. The COSMO-DE model with horizontal resolution of 1.4 km is run from 1 May 2014, 00:00 a.m. to 12 May 2014, 00:00 a.m. with hourly outputs. The hourly outputs are then interpolated (via the program INT2LM) onto the coarser grid with horizontal resolution of 2.8 km. These interpolated fields serve as initial conditions for 24-hr COSMO-DE forecasts.

2.2. Spectra and Statistics of Model Truncation Error

In the following, several properties of model truncation error, including kinetic energy spectrum and horizontal correlation, are computed, and their physical plausibility is elaborated. Since our goal is to represent model error at convective scale and since we will be doing data assimilation in hourly intervals, we choose samples with $t = 1$ hr in (1). Some important features of these samples are shown in the following.

Similar to Dowell and Wicker (2009) and Sobash and Wicker (2015), model variables of zonal velocity u, meridional velocity v, temperature T, and relative humidity q_v are perturbed. Perturbations of u and v can generate horizontal mass convergence, which is favorable for convection initiation (Banacos & Schulz, 2005). However, if only perturbing u and v without perturbing vertical velocity w, conservation of momentum may be violated due to mass continuity equation, and gravity waves can be evoked, which could be an additional source of imbalance. Therefore, the

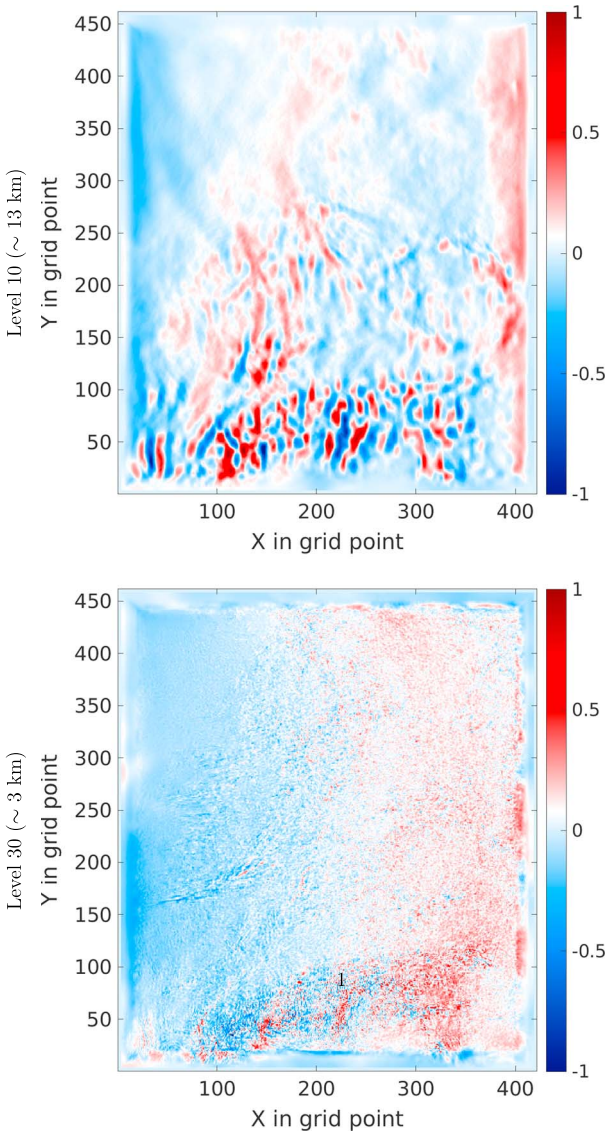

Figure 3. Mean model error of u at model levels 10 (~ 13 km, upper) and 30 (~ 3 km, lower), averaged over 40 random samples in case of $t = 1$ hr.

effects of perturbing w will be examined in this work. The other microphysical variables such as cloud droplets q_c, cloud ice q_i, rain q_r, snow q_s, and graupel q_g are currently not perturbed.

First of all, Figure 3 exemplifies mean model truncation error of u (averaged over 40 random samples) at model levels 10 (~ 13 km, lower stratosphere) and 30 (~ 3 km, lower troposphere). At model level 10, small-scale structures of model error can be particularly seen in the south (Alpine Region), probably due to vertical propagation of gravity waves triggered by differently resolved orography. The northwestern part of the domain over the sea is relatively smooth. Additionally, it can be seen in Figure 4 that the kinetic energy spectrum ranges from 10^0 to 10^4 m^3s^{-2} and the power declines strongly when the wavelength is smaller than 100 km. At model level 30, model truncation error exhibits even smaller features. Accordingly, the kinetic energy spectrum varies within a very narrow range (from 10^2 to 10^4 m^3s^{-2}) and the amplitudes of power are up to 3 orders of magnitude larger than at model level 10 for wavelength $\leq$ 100 km. Smaller scales of model truncation error are visible at lower atmosphere (e.g., 3 km) due to more convection being triggered with higher resolution model, resulting in more variability, than in the upper atmosphere (e.g., 13 km). It should be also pointed out that 40 random samples are used in Figure 3 since this is the ensemble size of data assimilation system. If the model truncation error is averaged over all samples (288), the absolute magnitude of

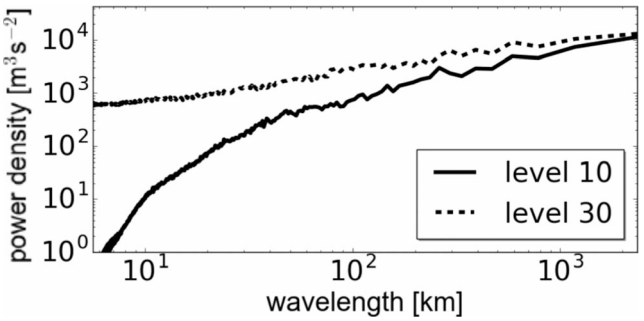

Figure 4. Power spectra of mean model error of u at model levels 10 ($\sim$ 13 km, solid) and 30 ($\sim$ 3 km, dashed), averaged over 40 random samples. The spectrum is calculated over a square domain by removing 20 grid points in latitude on each side of the COSMO-DE domain. One-dimensional spectrum results from 2-D Fourier transformation into linearly detrended field and subsequent summation of the Fourier coefficients over annuli in wave number space.

mean error decreases due to larger amount of samples, but the mean sample error is subtracted from each sample while applying the SAN.

Figure 5 illustrates the horizontal correlation of model truncation error for perturbed variables u (v is omitted for brevity), w, T, and q_v for distance 0 to 1,400 km at model levels 10 and 30. At model level 10, the correlation for u decreases slowly with distance and small correlation ($\approx$ 0.1) still exists up to 1,400 km; the correlation for w decreases rapidly to 0 within 400 km; the correlation for T decreases slowly and ends up with 0 at 1,400 km; the correlation for q_v decreases faster than for u and T but much more slowly than w and approaches 0 at around 800 km. At model level 30, the correlation for u decreases sharply to around 0.1 within tens of kilometers and completely disappears at about 800 km. Similar behavior of correlation can also be seen for T and q_v. For w little or no correlation can be recognized. To sum up, the horizontal correlation length scale reduces with decreasing height, probably due to the fact that model variables generally exhibit much more spatial variability in the lower troposphere than in the lower stratosphere; it also corresponds to what has been seen in Figure 3. We have also computed horizontal correlations of model truncation error resulting from $t > 1$ hr; the correlation length scale increases with larger t as expected (not shown).

Figure 6 shows the histograms of model truncation error for u, w, T, and q_v at model levels 10 and 30. First of all, it can be seen that the distributions of error of all variables at both levels have mean (denoted by μ) of 0. At model level 10, the distribution of error of u has standard deviation σ of 0.353 m/s and kurtosis γ of 2.162. The nonzero γ indicates that the distribution has slightly heavier tails and a slightly higher peak than the corresponding Gaussian distribution (note that $\gamma = 0.0$ for the Gaussian distribution). The distribution of error of w has $\sigma = 0.032$ m/s and $\gamma = 33.139$, and it is much more strongly tailed and peaked than the Gaussian distribution. The distribution of error of T has $\sigma = 0.157$ K and $\gamma = 2.162$. The distribution of model error of q_v has $\sigma = 2.5 \times 10^{-8}$ and $\gamma = 37.293$. At model level 30, the distribution of error of u has $\sigma = 0.425$ m/s and $\gamma = 28.63$, which is much greater than at level 10. The distribution of error of w has $\sigma = 0.02$ m/s which is 1 order of magnitude higher than that at level 10, probably due to occurrence of convection at the lower troposphere. The kurtosis of $\gamma = 329.122$ is significantly greater than at level 10. The distribution of error of T has $\sigma = 0.153$ K and $\gamma = 18.602$ which is also much greater than at

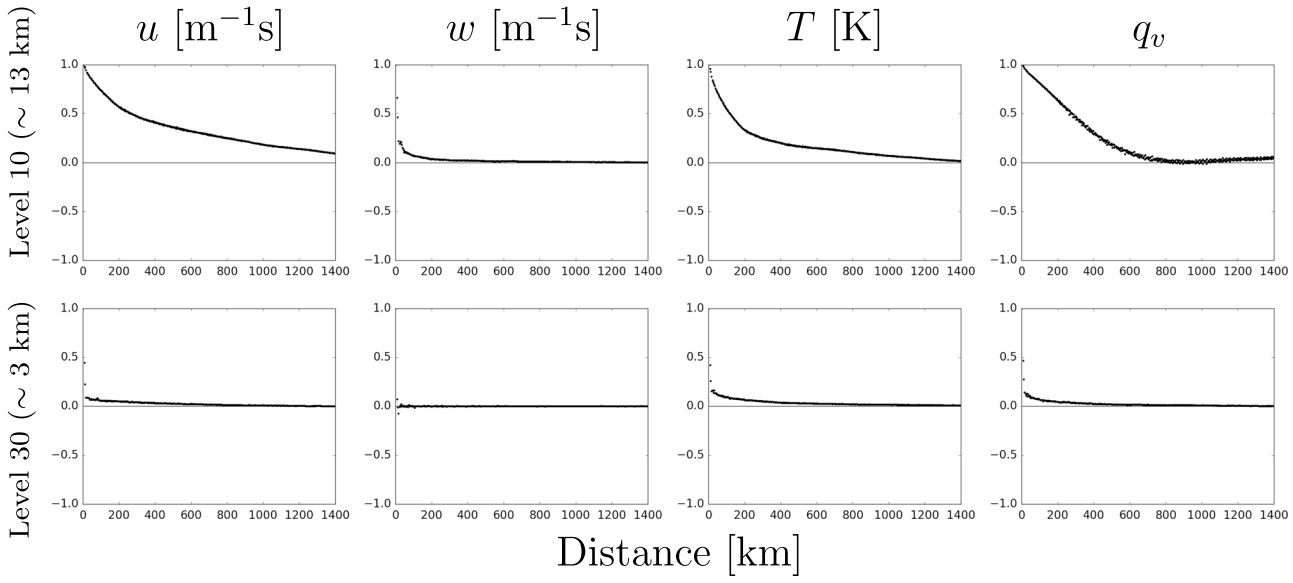

Figure 5. Horizontal correlation of model truncation error for model variables u, w, T, and q_v at model levels 10 ($\sim$ 13 km) and 30 ($\sim$ 3 km) in case of $t = 1$ hr. The correlations are averaged over all the horizontal grids given a vertical level.

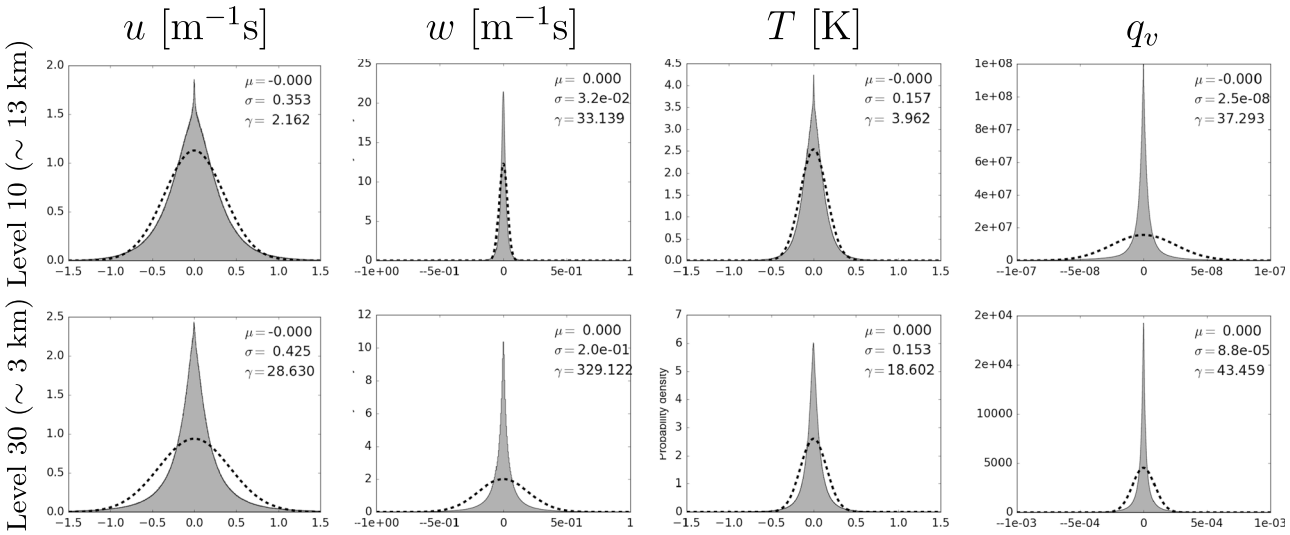

Figure 6. Histogram of model error samples for model variables u, w, T, and q_v at model levels 10 ($\sim$ 13 km) and 30 ($\sim$ 3 km) in case of $t = 1$. The variable μ is the mean of the distribution, σ is the standard deviation, and γ is the kurtosis. Dash lines depict the corresponding Gaussian distributions.

level 10. The distribution of error of q_v has $\sigma = 8.8 \times 10^{-5}$, 3 orders of magnitude greater than at level 10 since the lower troposphere is much wetter, while $\gamma = 43.459$ is only slightly greater than at level 10. In general, it can be said that the distributions of model truncation error of u and T are much more Gaussian in the lower stratosphere than in the low troposphere, and the distributions of model error of w and q_v are strongly non-Gaussian. The application of the SAN may introduce some extra non-Gaussianity into the ensemble, which is undesired by the ensemble Kalman filter. However, since the perturbations are added to the analysis ensemble as (2), the side effect of non-Gaussianity may be secondary compared to the nonlinear model integration.

Finally, it is worth noting that we have also explored the sensitivity of model truncation error to the length of the training period (not shown). It is found that 6 days is required to achieve approximate statistics (e.g., correlation length scale) as in the whole training period (12 days).

3. Brief Summary of Experimental Setup

Similar to Zeng et al. (2018), we choose a 2-week period from 27 May to 9 June 2016 with extraordinarily many severe convective storms in Germany. In the first week (from 27 May to 2 June) under strong large-scale forcing weather conditions, the convective activity was characterized by larger-scale precipitation patterns caused by frontal ascent, whereas much more scattered convective cells triggered by local mechanisms prevailed in the second week (from 3 June to 9 June) under weak forcing conditions. It has been shown in Zeng et al. (2018) that the LAN, based on random samples from climatological atmospheric background error covariance used by the global EnVar data assimilation system, mimics large-scale uncertainties arising from the global driving model and performs equally or even better than relaxation methods as well as

Table 1
Experimental Setup

		SAN ($\alpha_S = 1.25$)	
Experiment	LAN ($\alpha_L = 0.1$)	w unperturbed	w perturbed
E_LAN0.10	✓	×	×
E_SAN1.25	×	×	✓
E_LAN0.10SAN1.25NW	✓	✓	×
E_LAN0.10SAN1.25	✓	×	✓

Note. In each experiment either the LAN with $\alpha_L = 0.1$ or the SAN with $\alpha_S = 1.25$ or both can be applied. For the SAN it can be chosen if w is perturbed. ✓ means "on" and × means "off." LAN = large-scale noise; SAN = small-scale noise.

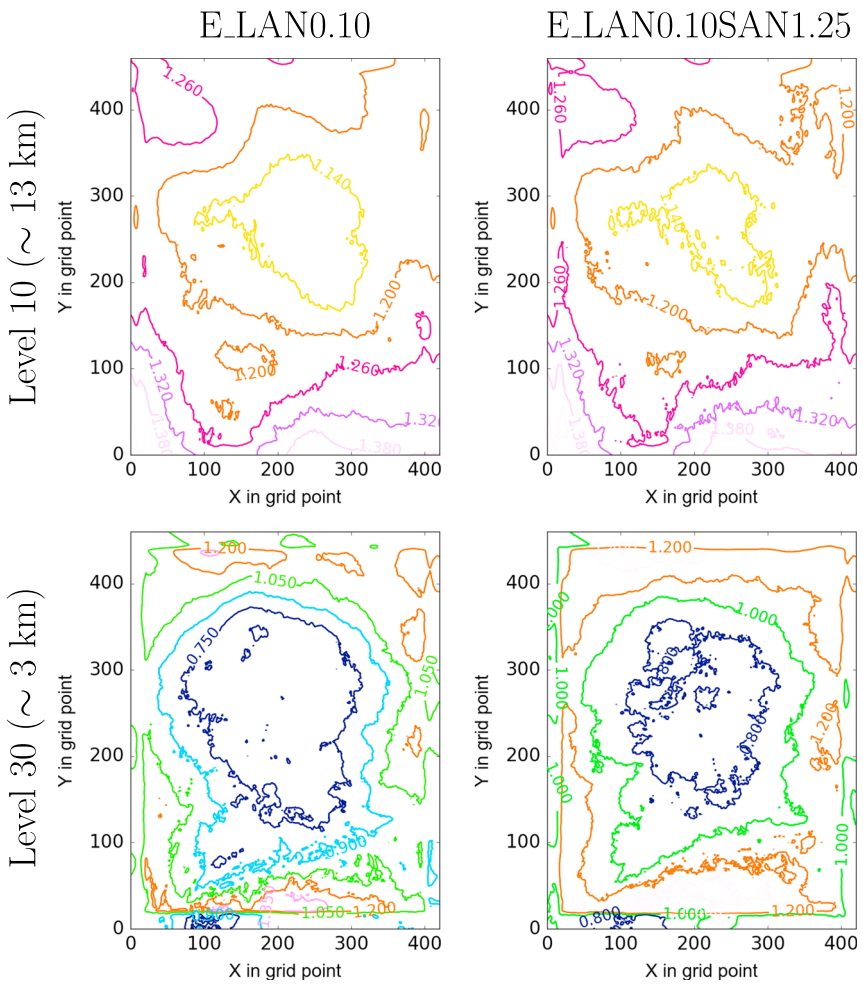

Figure 7. Mean spread of the background ensemble for variable u at model levels 10 (~ 13 km, upper) and 30 (~ 3 km, lower) for E_LAN01.10 and E_LAN0.10SAN1.25. The spread is averaged over all assimilation cycles.

combinations under strong forcing weather conditions. Its performance degrades a bit under weak forcing conditions, assumedly due to being less representative for small-scale features. Since the SAN may remedy model states with information on small scales, it may be especially useful to enhance the performance in case of weak forcing. Therefore, we first conduct a series of experiments for the second week in Study 1 (weak forcing) and then repeat the same experiments for the first week in the Study 2 (strong forcing). Note that the SAN randomly draws samples from the year 2014, which was governed by different synoptic conditions than the year 2016 of experiments. There was no attempt made to have samples with similar convective activities so that the results can be generalized outside of the training manifold. The experimental setup is given in Table 1, including combinations of the LAN and SAN, that is,

$$\boldsymbol{x}^{a(i)} \leftarrow \boldsymbol{x}^{a(i)} + \alpha_L \boldsymbol{\eta}_L^{(i)} + \alpha_S \boldsymbol{\eta}_S^{(i)}, \tag{3}$$

where $\boldsymbol{\eta}_L^{(i)}$ and $\boldsymbol{\eta}_S^{(i)}$ are random large- and small-scale samples, respectively. The tunable parameter α_L for the LAN has been tuned in Zeng et al. (2018) and set to 0.1. With the LAN, u, v, q_v, T, and p are perturbed. For the SAN α_S has been also tuned. It is found out that $\alpha_S = 1.25$ has the best performance both in assimilation cycling and 6-hr ensemble forecasts for the SAN alone as well as for combinations with the LAN (not shown). Note that it is very close to $\alpha_S = 1.20$ chosen by Hamill and Whitaker (2005) for their application.

Experiments are run by using the KENDA system at the DWD. Conventional observations from radiosondes (TEMP), wind profilers (PROF), aircraft reports (AIREP), and synoptic surface stations (SYNOP) are assimilated by the LETKF. In addition, radar reflectivity and no-precipitation observations (i.e., observations of reflectivity ≤ 5 dBZ) are also directly assimilated via the application of the radar forward operator EMVO-

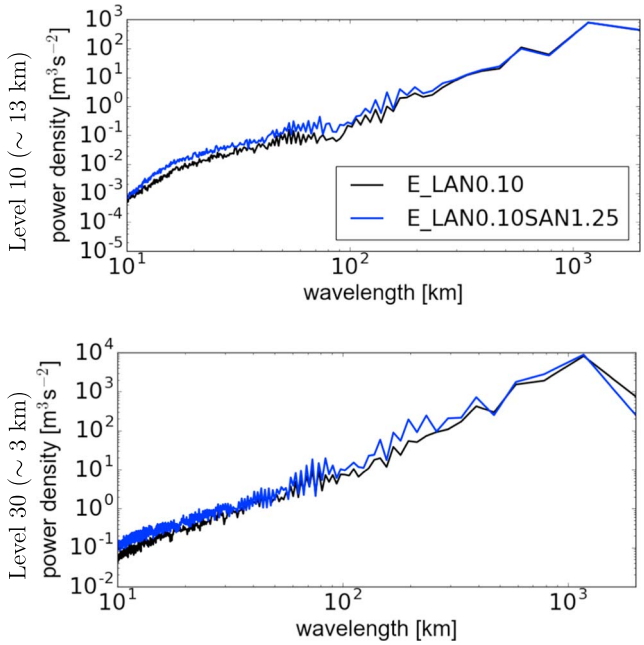

Figure 8. The power spectra of spread of *u* in Figure 7 at model levels 10 (- 13 km, upper) and 30 (- 3 km, lower) for E_LAN01.10 and E_LAN0.10SAN1.25.

RADO (Zeng et al., 2014, 2016). The assimilation window is 1 hr, and the ensemble size is 40. In addition, ensemble forecasts (of 20 members) are run daily at 10:00, 11:00, … , 17:00 and 18:00 UTC.

The NWP model is the operational convective-permitting COSMO model, which is fully compressible and nonhydrostatic (Baldauf et al., 2011; Doms et al., 2011; Doms & Baldauf, 2015). The model size is 421 × 461 × 50 with horizontal resolution of 2.8 km. The one-moment microphysical scheme based on Lin et al. (1983) and Reinhardt and Seifert (2006) is utilized. Lateral boundary conditions for the ensemble are provided by the EPS (Ensemble Prediction System) of the operational global model ICON.

A general description of the KENDA system and the LETKF can be found in Schraff et al. (2016). More details about weather situation (e.g., strong/weak forcing) and treatment of observations assimilated (e.g., superobbing, localization, and specification of observation error) are given in Zeng et al. (2018).

4. Experiment Results

As mentioned above, a series of experiments including the application of the LAN only (E_LAN0.10), application of the SAN only (E_SAN1.25), and the combinations of both (E_LAN0.10SAN1.25NW and E_LAN0.10SAN1.25, the former one is without perturbing *w*) are first conducted for the weak forcing conditions (from 27 May to 2 June) in Study 1 and then restarted for strong forcing conditions (from 3 June to 9 June) in Study 2. In the following, we will begin with looking into the background ensemble spread and kinetic energy spectrum of analysis during the cycling and then we will mainly focus on the performance during 6-hr ensemble forecasts.

To account for uncertainties in verification scores, such as fractions skill score (FSS, Robert & Lean, 2008) and false alarm rate (FAR), the bootstrap method (Efron & Tibshirani, 1993) is used. E_LAN0.10 is taken as the reference run, the relative differences compared to E_LAN0.10 are calculated, and then 10,000 bootstrap resampling is carried out to examine the statistical significance at 95% confidence intervals.

4.1. Study 1

Figure 7 depicts the background spread of *u* at different levels for E_LAN0.10 and E_LAN0.10SAN1.25. At level 10, E_LAN0.10 results in a spread that decreases from 1.32 m/s at the boundary to 1.140 m/s in the inner domain since more observations are available and assimilated in the inner domain. In comparison, E_LAN0.10SAN1.25 generates a larger spread than E_LAN0.10 in the inner domain, and more importantly, the spread is larger for smaller scales as indicated by the power spectra in Figure 8. At level 30, the spread

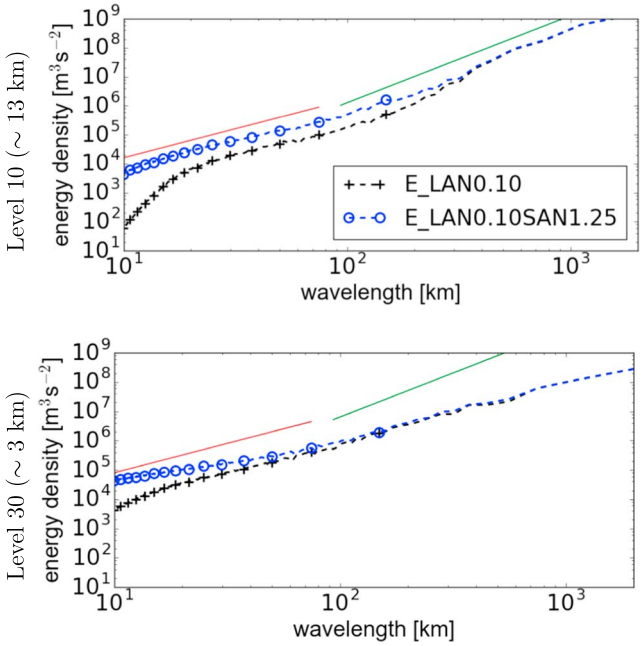

Figure 9. Mean kinetic energy spectra of the analysis ensemble at model levels 10 (~ 13 km, upper) and 30 (~ 3 km, lower) for E_LAN01.10 and E_LAN0.10SAN1.25. The kinetic energy spectra are calculated for each ensemble member and then averaged over the ensemble size. This is done for 00:00 UTC of each day during the period of Study 1, and then the average is calculated. Mean kinetic energy spectra of E_SAN1.25 and E_LAN0.10SAN1.25NW are very comparable to those of E_LAN0.10SAN1.25 (not shown). The green and red lines show the reference lines for −3 and −5/3 power laws, respectively.

of E_LAN0.10 has a sharper decrease (from 1.2 m/s at the boundary to 0.75 m/s in the inner domain) than at level 10, due to denser observations in the lower troposphere. Again, E_LAN0.10SAN1.25 results in a larger spread especially at smaller scales. It is likely that small-scale structures in background ensemble spread allow small-scale updates and create more small-scale variability in model states, which favors strong horizontal convergence and consequently onset of convection. Moreover, the mean kinetic energy spectra of analysis ensemble are shown in Figure 9. It can be seen that the slope of kinetic energy spectrum of E_LAN0.10SAN1.25 at model level 10 (~ 13 km) is almost parallel to −5/3 and much shallower than that of E_LAN0.10 for wavelengths shorter than 100 km. Selz et al. (2018) showed that the kinetic energy on scales smaller than 300 km is strongly positively correlated with convective precipitation. Therefore, E_LAN0.10SAN1.25 is prone to triggering more convective precipitation than E_LAN0.10, however, it may have some corresponding side effects. Figure 10 shows that E_LAN0.10SAN1.25 has much higher surface pressure tendency S_t than E_LAN0.10 at analysis steps, which indicates more imbalanced model states in E_LAN0.10SAN1.25. One would usually expect that imbalanced model states might cause spurious

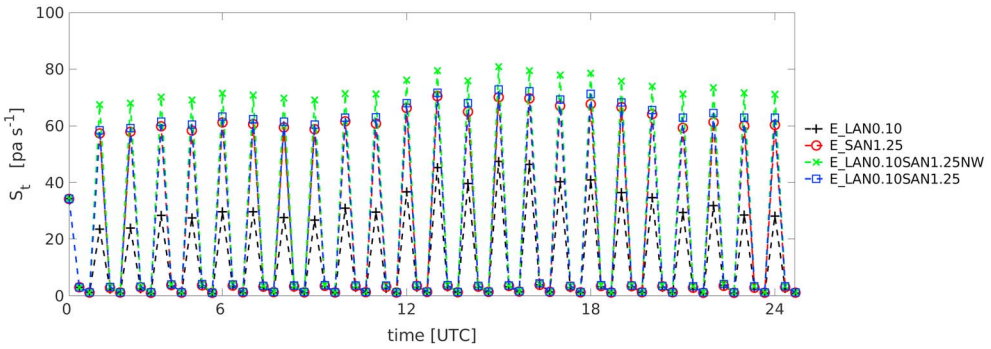

Figure 10. Half-hourly evolution of surface pressure tendency S_t for E_LAN0.10, E_SAN1.25, E_LAN0.10SAN1.25NW, and E_LAN0.10SAN1.25, averaged over all ensemble members for the period 00:00 UTC 3 June to 00:00 UTC 4 June. Recall that the assimilation window is 1 hr.

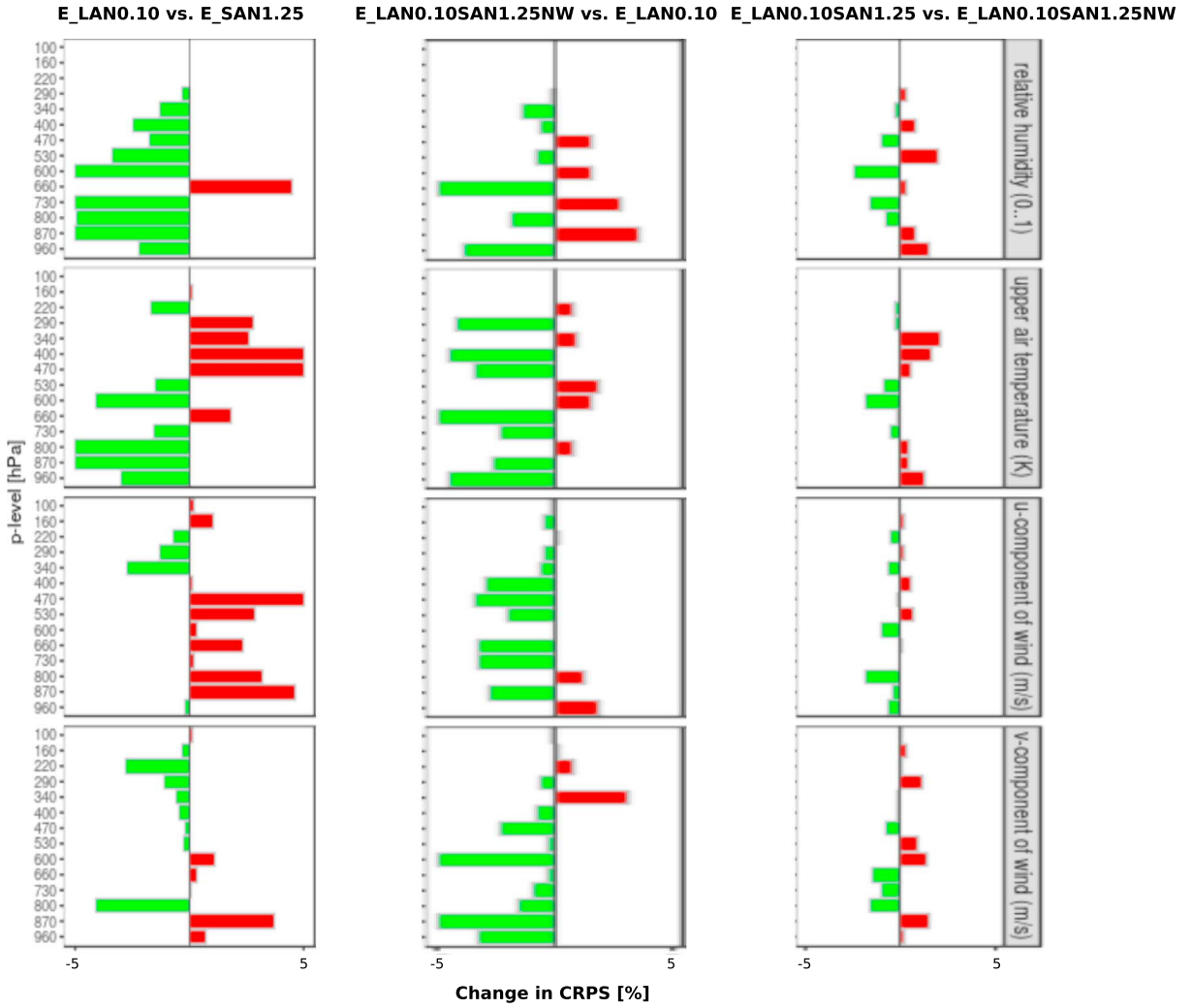

Figure 11. Pairwise comparison (exp1 vs. exp2) of relative differences (in percentage [%]) of the CRPS for relative humidity, temperature, and u and v components of horizontal wind (from top to bottom) in Study 1. The relative difference is calculated by the formula $\frac{(\mathrm{CRPS(exp1)-CRPS(exp2)})}{(\mathrm{CRPS(exp1)+CRPS(exp2)})/2} * 100$, where CRPS(exp1) and CRPS(exp2) denote the CRPS values of the first and second experiments, respectively, and then averaged over all forecast lead times. The observations used for verification are upper air observations, and the scores are aggregated over all initial times. From left to right are the comparisons between E_LAN0.10 and E_SAN1.25, between E_LAN0.10SAN1.25NW and E_LAN0.10, and between E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, respectively. The color of green means that exp1 is better, and red means that exp2 is better. CRPS = continuous ranked probability score.

convection and lead to poor quality of forecasts, but in this case it seems that "noisy" model states not only provide more convection-friendly conditions but also surprisingly reduce spurious convection as shown below. It can be also seen that E_LAN0.10SAN1.25NW is more imbalanced than E_LAN0.10SAN1.25 as expected.

To demonstrate the performance of 6-hr ensemble forecasts, the continuous ranked probability score (CRPS, Hersbach, 2000) is computed. The CRPS is defined as the integrated squared difference between the cumulative density function (CDF) of forecasts and observations:

$$\mathrm{CRPS} = \int_{+\infty}^{+\infty} [F^f(x) - F^o(x)]^2 \mathrm{d}x, \tag{4}$$

where F^f and F^o are CDFs for forecasts and observations, respectively. The CRPS is regarded as a measurement of error for probabilistic prediction, it ranges from 0 to infinity with lower values representing a better score and the value of 0 indicates a perfect forecast accuracy. Figure 11 compares pairs of experiments and illustrates the vertical profiles of relative differences (in percentage [%]) of the CRPS for relative humidity, temperature, and u and v components of horizontal wind. The CRPS is computed for all 6-hr

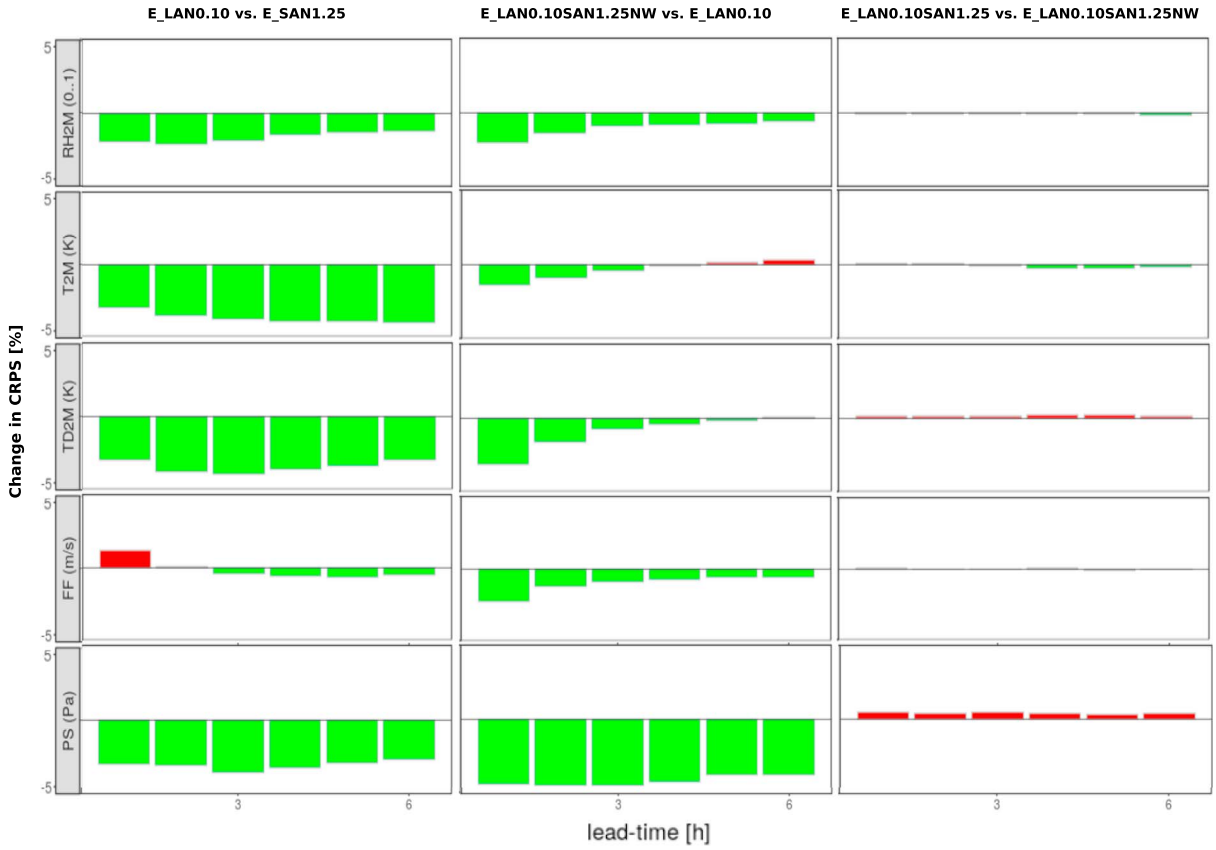

Figure 12. Pairwise comparison (exp1 vs. exp2) of vertical profiles of relative differences (in percentage [%]) of the CRPS for 2-m relative humidity, 2-m temperature and 2-m dew point temperature, and 10-m horizontal wind and pressure (from top to bottom) in Study 1. The relative difference is calculated by the formula $\frac{(\mathrm{CRPS(exp1)}-\mathrm{CRPS(exp2)})}{(\mathrm{CRPS(exp1)}+\mathrm{CRPS(exp2)})/2} * 100$, where CRPS(exp1) and CRPS(exp2) denote the CRPS values of the first and second experiments, respectively. The observations used for verification are SYNOP observations, and the scores are aggregated over all initial times. From left to right are the comparisons between E_LAN0.10 and E_SAN1.25, between E_LAN0.10SAN1.25NW and E_LAN0.10, and between E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, respectively. The color of green means that exp1 is better, and red means that exp2 is better. CRPS = continuous ranked probability score; SYNOP = synoptic surface stations.

ensemble forecasts (note that forecasts are daily initiated at 10:00, 11:00, … , 17:00 and 18:00 UTC), verified against upper air observations. For the pair of E_LAN0.10 and E_SAN1.25, E_LAN0.10 is better than E_SAN1.25 at most levels for relative humidity; E_LAN0.10 is better (worse) than E_SAN1.25 in the lower (upper) atmosphere for temperature, and it is almost the opposite for horizontal wind. For the pair of E_LAN0.10SAN1.25NW and E_LAN0.10, E_LAN0.10SAN1.25NW is better than E_LAN0.10 at most levels for all variables except relative humidity. For the pair of E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, no experiment is certainly better than the other. Figure 12 also compares relative differences of the CRPS values as a function of forecast lead time, the variables (2-m relative humidity, 2-m temperture, 2-m temperture dew point, 10-m horizontal wind, and pressure) are verified against surface SYNOP observations. For the pair of E_LAN0.10 and E_SAN1.25, E_LAN0.10 is considerably better throughout the forecast lead time for all variables except the first hour for horizontal wind. For the pair of E_LAN0.10SAN1.25NW and E_LAN0.10, E_LAN0.10SAN1.25NW is advantageous for all variables, especially for pressure (almost 5.0%), and the advantage seems to decreases gradually with the forecast lead time. For the pair of E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, no significant differences can be recognized for all variables. In total, based on CRPS values it can be stated that E_LAN0.10 is better than E_SAN1.25 for surface variables, E_LAN0.10SAN1.25NW is better than E_LAN0.10 for both surface and upper air variables, and E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW are very comparable.

Figure 13 compares 6-hr ensemble forecasts of E_LAN0.10 and E_LAN0.10SAN1.15NW by means of radar reflectivity composite of elevation 0.5°. The simulated reflectivities of the ensemble are represented by probability, defined as ratio of the number of ensemble members that exceed the given threshold value (here 20 dBZ) divided by the ensemble size. The ensemble forecasts are initialized at 12:00 UTC 6 June 2016. In

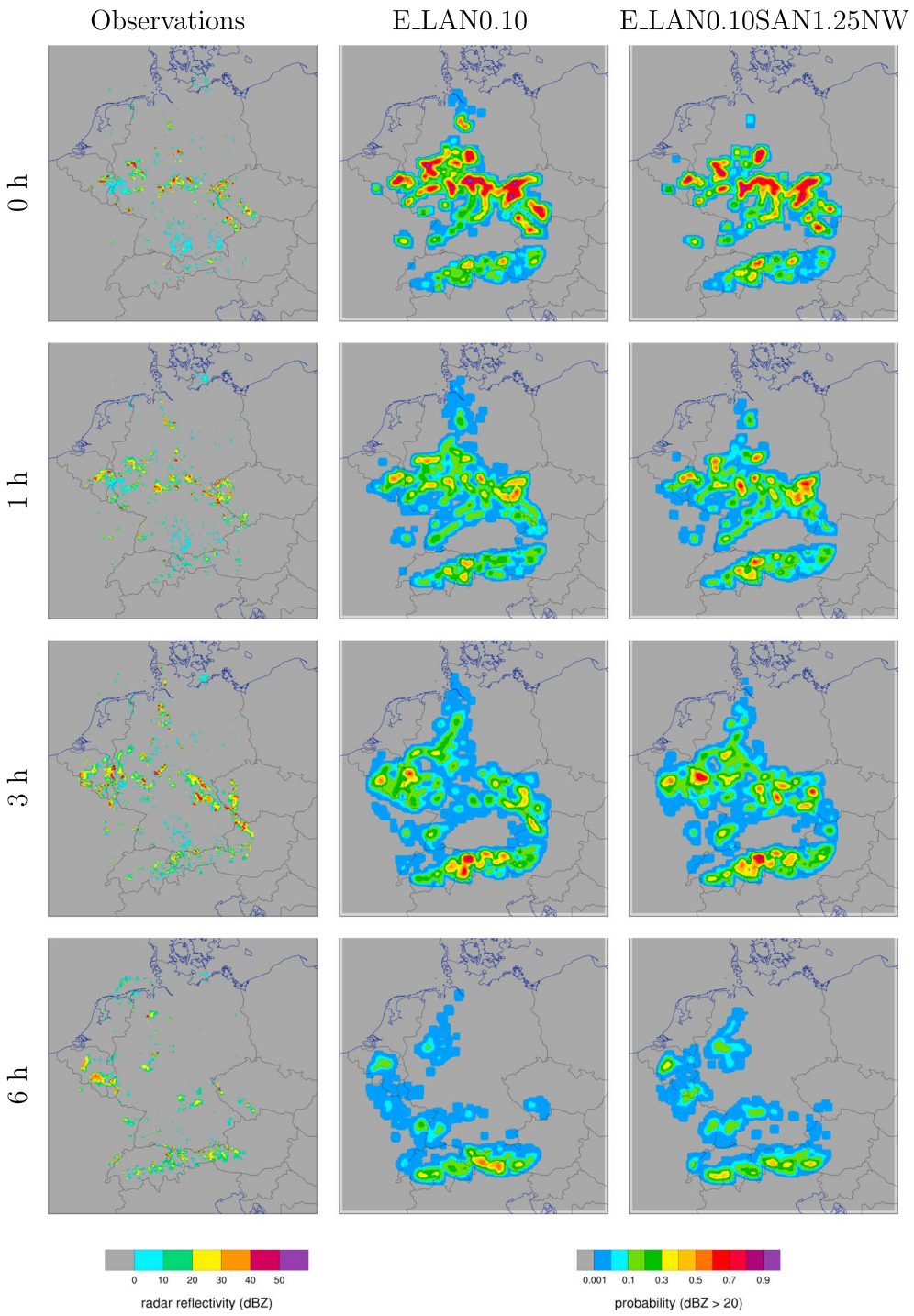

Figure 13. Six-hour evolution of reflectivities from the lowest elevation 0.5°, starting at 12:00 on 6 June. (first column) Observations; (second and third columns) probabilities of E_LAN0.10 and E_LAN0.10SAN1.25, defined as the number of ensemble members exceeding the threshold value 20 dBZ (~ 0.6 mm/hr) divided by ensemble size.

observations, a number of scattered convective cells can be seen in the middle of domain initially. The whole precipitation system weakens a bit in the first hour and then strongly intensifies at the third hour before it rapidly decays at the sixth hour. Within the 6 hr, the precipitation system stays fairly stationary. In comparison with observations at the initial time, except for the overestimation of the precipitation in the southern part, vicinities of all convective cells with observed reflectivity ≥ 20 dBZ are covered by high probabilities (≥ 50 %) of ensemble members of E_LAN0.10, which means an appropriate representation of the precipitation

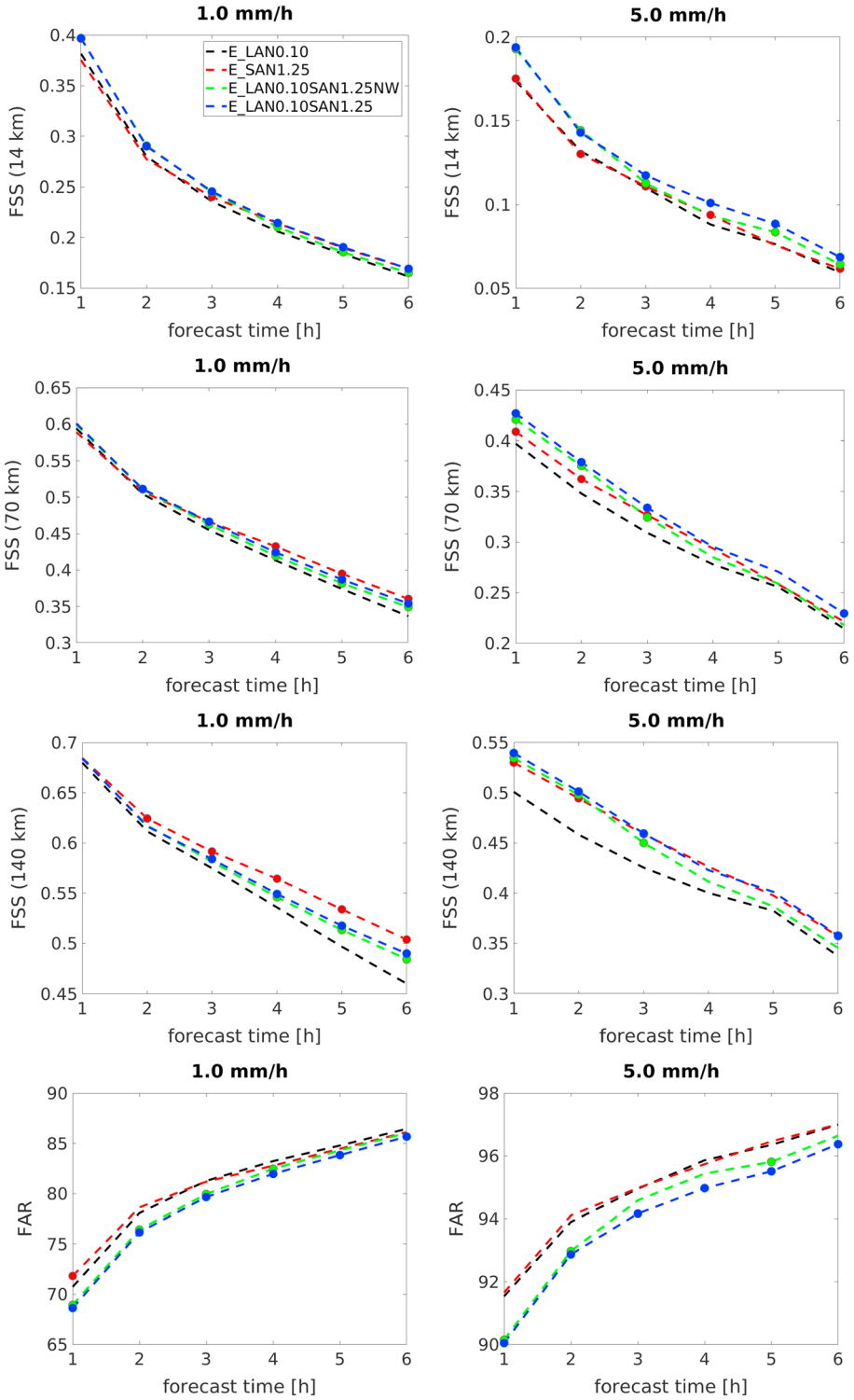

Figure 14. Verification of 6-hr ensemble forecasts against radar-derived precipitation rate for Study 1. (left column) The first to third panels illustrate the FSS values of experiments for the threshold value of 1.0 mm/hr as a function of forecast lead time for different scales of 14, 70, and 140 km, respectively. The fourth panel illustrates the FAR values for the threshold value of 1.0 mm/hr. (right column) the same as the left-hand side but for threshold value of 5.0 mm/hr. Each FSS (FAR) value is computed as an average over all 63 forecast runs (the study period contains 7 days, and each day has nine forecast runs). The lines are marked as filled dots at the forecast lead times where the differences compared to E_LAN0.10 are statistically significant at 95% confidence intervals after 10,000 bootstrap resamplings based on 63 difference samples. FSS = fractions skill score; FAR = false alarm rate.

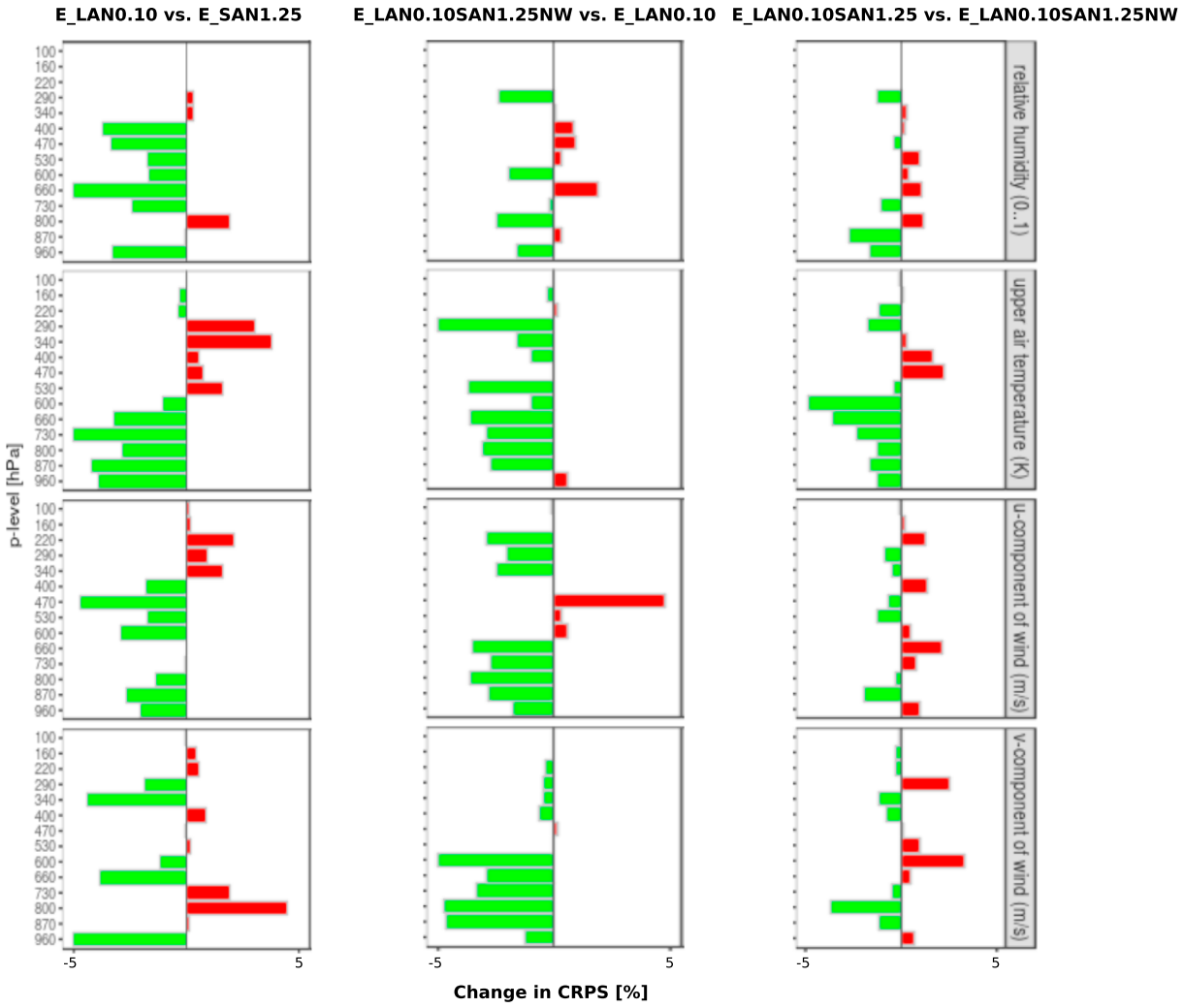

Figure 15. The same as Figure 11 but for Study 2.

system by the ensemble. Similar can be seen also for E_LAN0.10SAN1.15NW. For 1-hr forecast, it is obvious that the area with high probabilities (≥ 50 %) has significantly reduced, (i.e., the lifetime of individual convective cores is short and less than 1 hr for most cores and the practical predictability is very limited for individual convective cores), most of convective cells are represented by the ensemble of E_LAN0.10 with probabilities ranging from 30% to 50 %. Comparatively, convective cells are reproduced by a higher fraction of ensemble members (probabilities ranging from 40% to 60 %) in E_LAN0.10SAN1.15NW. This advantage of E_LAN0.10SAN1.15NW is also evident for 3-hr forecasts. For 6-hr forecasts, the predictability is already low and it is difficult to tell which one is better. This visual comparison indicates a better representation of convective cells by the ensemble of E_LAN0.10SAN1.15NW than that of E_LAN0.10 in ensemble forecasts, at least up to 3 hr. In the following, objective forecast verification scores are used to compare the performance of experiments in ensemble forecasts.

Figure 14 shows the verification of all 6-hr ensemble forecasts against radar-derived precipitation rate, using the FSS values, as function of forecast lead time for different scales of 14, 70, and 140 km and threshold values 1.0 (light rain) and 5.0 mm/hr (moderate rain). For 1.0 mm/hr and 14 km, E_LAN0.10SAN1.25 is considerably better than E_LAN0.10 with statistical significance throughout 6 hr. E_LAN0.10SAN1.25 is also better than E_SAN1.25 before they approach after 4 hr, while E_LAN0.10SAN1.25 is identical to E_LAN0.10SAN1.25NW until 3 hr and then becomes slightly superior. For 70 km, E_SAN1.25 is the best, followed sequentially by E_LAN0.10SAN1.25, E_LAN0.10SAN1.25NW, and E_LAN0.10. Similar pattern can also be seen for 140 km although the differences are much larger. For 5 mm/hr and 14 km,

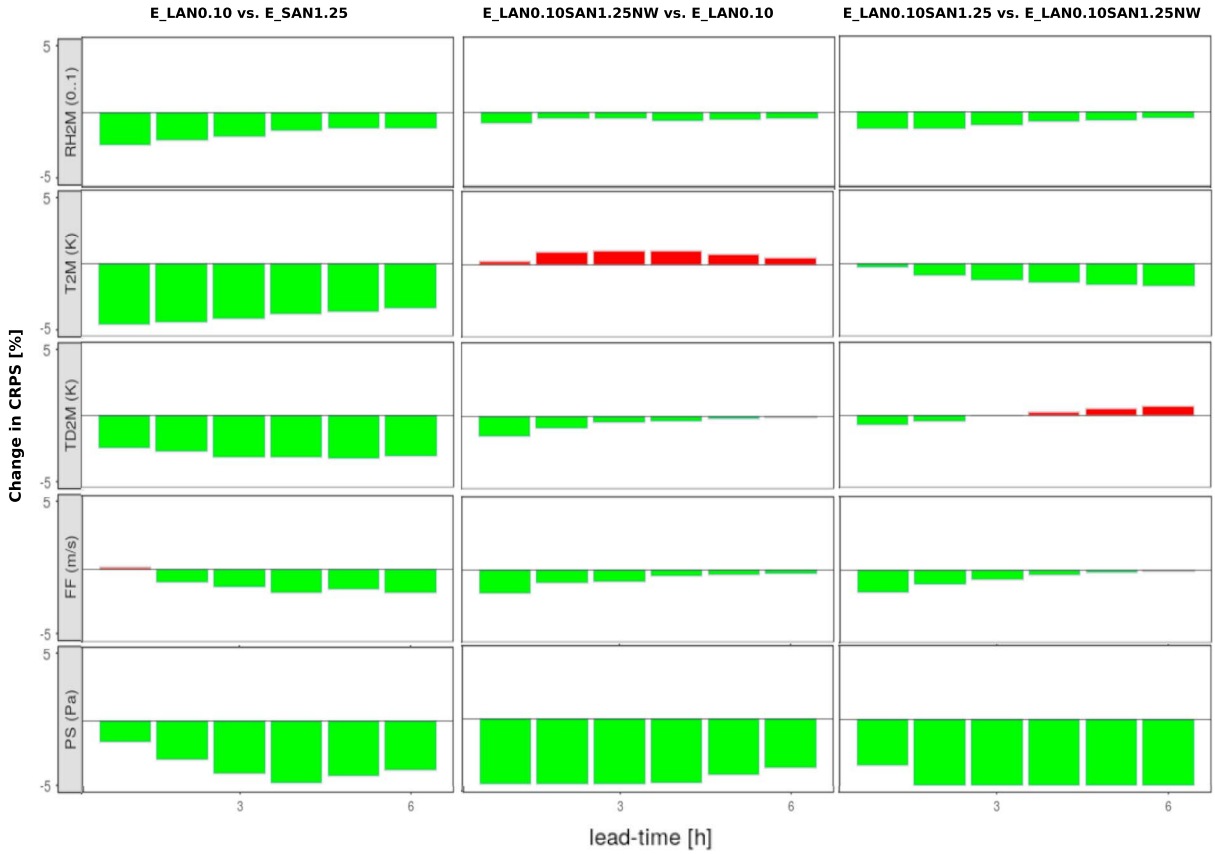

Figure 16. The same as Figure 12 but for Study 2.

E_LAN0.10SAN1.25 is slightly better than E_LAN0.10SAN1.25NW and clearly better than E_SAN1.25 and E_LAN0.10 with statistical significance throughout 6 hr. For both 70 and 140 km, E_LAN0.10SAN1.25 is slightly better than E_LAN0.10SAN1.25NW and E_SAN1.25 and much better than E_LAN0.10.

Figure 14 also compares the FAR values of precipitation for threshold values 1.0 and 5.0 mm/hr. The FAR is the number of false alarms divided by the total number of events forecast. Rather than spatial comparisons between forecasts and observations as the FSS does, the FAR makes point comparisons. It ranges from 0 to 1 and the perfect score is 0. For 1.0 mm/hr, E_LAN0.10 is slightly better than E_SAN1.25 in the first 3 hr and then becomes slightly worse afterward. E_LAN0.10SAN1.25NW is considerably better than E_SAN1.25 and E_LAN0.10 at the beginning and gets close to them with increasing time. E_LAN0.10SAN1.25 is evidently the best, which is identical to E_LAN0.10SAN1.25NW in 3 hr and becomes slightly better afterward. For 5 mm/hr, E_LAN0.10 and E_SAN1.25 are quite comparable. E_LAN0.10SAN1.25NW is considerably better than E_LAN0.10 and E_SAN1.25. E_LAN0.10SAN1.25 is close to E_LAN0.10SAN1.25NW up to 2 hr and then becomes slightly better. Since a larger FAR value indicates more spurious convection, it can be concluded from Figure 14 that both E_LAN0.10 and E_SAN1.25 results in much more spurious convection based on the FAR values of ensemble forecasts. Based on the FSS values, E_LAN0.10 may be slightly better than E_SAN1.25 for 1.0 mm/hr, while E_SAN1.25 is much better than E_LAN0.10 for 5.0 mm/hr. Overall, E_LAN0.10SAN1.25 produces the best ensemble forecasts based on the FSS values for all precipitation rates and scales accounted for. In addition, E_LAN0.10SAN1.25 also results in the least spurious convection. E_LAN0.10SAN1.25NW is comparable to E_LAN0.10SAN1.25 in the first few hours and degrades considerably later on.

Last but not least, it should be mentioned that we have also performed experiments in which we adjusted the SAN by perturbing model variables only at places with higher reflectivity (Dowell & Wicker, 2009) or larger reflectivity innovation (Sobash & Wicker, 2015) as done for smoothed random Gaussian noise. However, no satisfactory results compared to E_LAN0.10SAN1.25 could be achieved (not shown). Although the reason

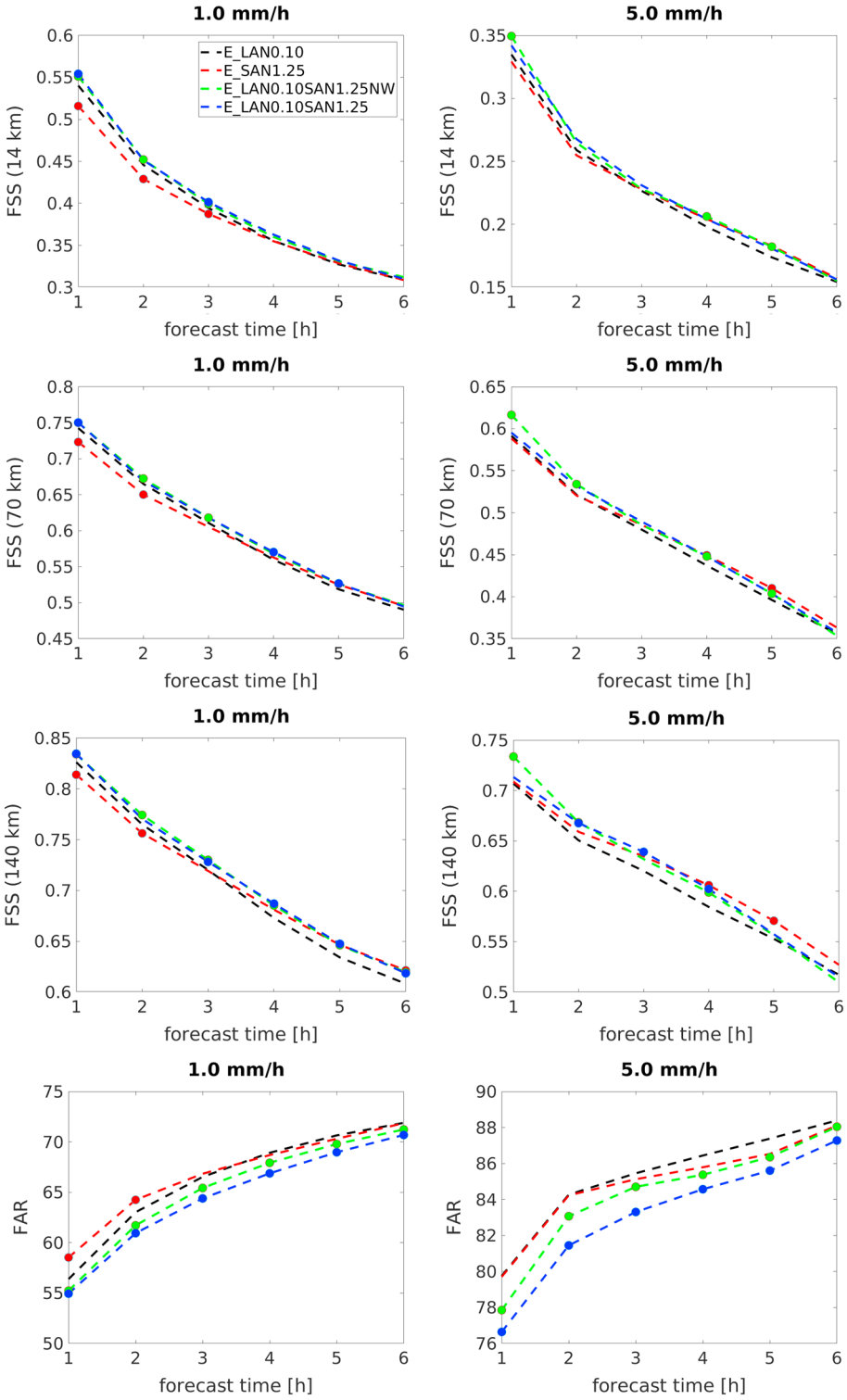

Figure 17. The same as Figure 14 but for Study 2.

has not been explored elaborately, it may be related to the limitation of perturbed area which may cause strong inconsistency of model states between perturbed and unperturbed areas.

4.2. Study 2

Similar to Figure 11, Figure 15 compares vertical profiles of the CRPS of experiments in pair (verified against upper air observations) but now for Study 2. For the pair of E_LAN0.10 and E_SAN1.25, E_LAN0.10 is better than E_SAN1.25 at most levels for relative humidity; E_LAN0.10 is better (worse) than E_SAN1.25 in the lower (upper) atmosphere for temperature and u component, and no experiment is evidently better for v component. For the pair of E_LAN0.10SAN1.25NW and E_LAN0.10, E_LAN0.10SAN1.25NW is better than E_LAN0.10 at most levels for all variables except relative humidity. For the pair of E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, no experiment is certainly better than the other. Similar to Figure 12, Figure 16 shows comparisons of 6-hr ensemble forecasts of experiments in pair by means of the CRPS (verified against SYNOP observations). For the pair of E_LAN0.10 and E_SAN1.25, E_LAN0.10 is much better than E_SAN1.25 for all variables. For the pair of E_LAN0.10SAN1.25NW and E_LAN0.10, E_LAN0.10SAN1.25NW is much better than E_LAN0.10 for pressure; slightly better for 2-m relative humidity, dew point and 10-m horizontal wind; and slightly worse for 2-m temperature. For the pair of E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW, E_LAN0.10SAN1.25 is much better than E_LAN0.10SAN1.25NW for pressure and slightly better for the other variables except 2-m temperature after 3 hr. Compared to Figure 12, the positive impacts of w perturbations could be explained as follows: the perturbations of u, v, and w are physically consistent (approximately divergence free). This means that if u, v, and w are perturbed together, the effect on w will be more strong and persistent during the early stage of the forecasts as compared to leaving out w perturbations. Under weak forcing conditions, convection is mainly caused by local thermal instabilities and characterized by the presence/absence of local temperature inversions and/or near-surface moisture contrasts. For strong forcing conditions, instabilities are more homogeneous over larger areas (less influenced by local conditions and less local extremes) and convection lives relatively longer, which may amplify the effects of w perturbations. Generally speaking, it can be concluded that E_LAN0.10 is better than E_SAN1.25 for surface variables, E_LAN0.10SAN1.25NW is better than E_LAN0.10 for both surface and upper air variables, and E_LAN0.10SAN1.25 is better than E_LAN0.10SAN1.25NW for surface variables.

Regarding the FSS values of 6-hr ensemble forecasts for precipitation rate, it is shown in Figure 17 that, for 1.0 mm/hr, E_LAN0.10SAN1.25NW and E_LAN0.10SAN1.25 are almost the same for scale of 14 km. Both are slightly better than E_LAN0.10 which is considerably better than E_SAN1.25; however, differences among experiments become statistically insignificant with the increasing time. Similar behavior is also visible for scale 70 and 140 km except that for 140 km E_LAN0.10 loses more skills in the last 3 hr. For 5.0 mm/hr, E_LAN0.10SAN1.25NW is slightly better than E_LAN0.10SAN1.25 at the beginning and then both are strongly overlapped. The both are slightly better than E_LAN0.10 throughout the forecast lead time. Almost the same behavior can be also seen for 70 and 140 km, although E_LAN0.10SAN1.25NW seems to be more advantageous at the beginning. For all scales, E_SAN1.25 is worse than E_LAN0.10SAN1.25NW and E_LAN0.10SAN1.25 in the first few hours, but it approaches and becomes even better with the increasing time. With respect to the FAR for 1.0 mm/hr, E_LAN0.10SAN1.25 is better than E_LAN0.10SAN1.25NW, followed by E_LAN0.10 and E_SAN1.25, while E_LAN0.10 is considerably better than E_SAN1.25 in the first 3 hr before they approach. For 5.0 mm/hr, E_LAN0.10SAN1.25 is the best as well, followed by E_LAN0.10SAN1.25NW. E_LAN0.10 and E_SAN1.25 are the same in the first 2 hr, and then the latter one becomes better. To conclude, this is similar to what has been seen in Figure 14. E_LAN0.10 is better than E_SAN1.25 for 1.0 mm/hr, but E_SAN1.25 is better for 5.0 mm/hr based on the FSS and FAR values of ensemble forecasts. Overall, E_LAN0.10SAN1.25 and E_LAN0.10SAN1.25NW have better performance. Both are comparable based on the FSS values for all precipitation rates and scales considered, while the latter one may be associated with slightly more spurious convection. However, it can be seen that the advantage of E_LAN0.10SAN1.25 over E_LAN0.10 is not as significant as in Study 1 due to strong forcing conditions.

5. Conclusion and Outlook

In this work, we incorporate the small-scale additive noise based on random samples of model truncation error, combining it with the large-scale additive noise based on random samples from global climatological atmospheric background error covariance (Zeng et al., 2018), to account for model error on multiple scales in convective-scale data assimilation. A series of experiments have been executed in the framework of the

operational KENDA system at the DWD for a 2-week period in Germany with different types of synoptic forcing of convection (i.e., strong or weak forcing). It is shown that the combination of the large and small-scale noise results in a larger background ensemble spread than the application of the large-scale noise only and the analysis ensemble of the combination exhibits more small-scale variability and is more energetic at small scales. In terms of the quality of short-term 6-hr forecasts, the combination produces more accurate ensemble forecasts than the large-scale noise only based on conventional surface and upper air observations. It also produces better precipitation forecasts with less spurious convection under both weak and strong forcing situations, while the improvement is especially significant in the weak forcing situation. This may be due to the fact that extra energy is added to smaller scales and more small-scale variability is available, which favors creation of strong local convergence and thus occurrence of convection, and this effect is especially important for the weak forcing situation. It is also shown that additionally perturbing vertical velocity in the small-scale noise part of the combination can further improve the balance of model states and the quality of precipitation forecasts. Moreover, it is found that the application of small-scale noise only may produce better forecasts of precipitation than the large-scale noise, but it results in less accurate ensemble forecasts if verified against conventional surface observations. In total, it can be concluded that the combination has the best performance in short-term ensemble forecast under all analyzed synoptic forcing conditions due to its multiscale representation of model error.

Currently, the training period of model truncation error is chosen from convective days in summer. The properties of model truncation error may differ in a different season, for example a training period in winter may be needed for case studies of winter storms. There are some new approaches as the Adaptive Background Error Inflation (Minamide & Zhang, 2019), which attempts to treat model error and non-Gaussian sampling error adaptively, as well as a new approach that allows for using more ensemble members in the treatment of model error than forecasted with additive noise (Sommer & Janjić, 2018). These could be explored for convective scale data assimilation in the future together with the Adaptive Observation Error Inflation (Minamide & Zhang, 2017). We plan to supplement or compare the additive noise with other approaches that account for subgrid-scale model error, such as physically based stochastic perturbation scheme for turbulence (Kober & Craig, 2016; Rasp et al., 2018), which is flow dependent, and an advanced warm bubble technique which can automatically detect and trigger missing convective cells. The results will be presented in another article.

Acknowledgments
Data of Figure 1 were provided by Christian Keil from LMU. Thanks are also given to George Craig, Mijram Hirt, and Tobias Necker from LMU and Felix Fundel, Axel Hutt, Roland Potthast, Hendrik Reich, and Klaus Stephan from DWD for technical support and fruitful discussions. The study was carried out in the Hans Ertel Centre for Weather Research (Simmer et al., 2016; Weissmann et al., 2014). This research network of universities, research institutes, and the Deutscher Wetterdienst is funded by the BMVI (Federal Ministry of Transport and Digital Infrastructure). The assimilated data used were obtained from the DWD. The processed data and plotting scripts are freely accessible under the website (http://doi.org/10.5281/(zenodo.1473600).

References

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., & Raschendorfer, M. (2011). Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, *139*, 3887–3905.

Banacos, P., & Schulz, D. M. (2005). The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives. *Weather and Forecasting*, *20*, 351–366.

Buizza, R., Houtekamer, P., Toth, Z., Pellerin, G., & Wei, M. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, *133*, 1076–1097.

Buzzi, A., Davolio, S., Malguzzi, P., Drofa, O., & Mastrangelo, D. (2014). Heavy rainfall episodes over Liguria in autumn 2011: Numerical forecasting experiments. *Natural Hazards and Earth System Sciences*, *14*, 1325–1340.

Caya, A., Sun, J., & Snyder, C. (2005). A comparison between the 4D-VAR and the ensemble Kalman filter techniques for radar data assimilation. *Monthly Weather Review*, *133*, 3081–3094.

Doms, G., & Baldauf, M. (2015). A description of the nonhydrostatic regional cosmo-model LM. Part I: Dynamics and numerics. Germany: Consortium for Smallscale Modeling (COSMO). Retrieved from http://www.cosmo-model.org/content/model/documentation/core/cosmoDyncsNumcs.pdf

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., et al. (2011). A description of the nonhydrostatic regional cosmo-model. Part II: Physical parameterization. Germany: Consortium for Smallscale Modeling (COSMO). Retrieved from http://www.cosmo-model.org/content/model/documentation/core/cosmoPhysParamtr.pdf

Dowell, D. C., & Wicker, L. J. (2009). Additive noise for storm-scale ensemble data assimilation. *Monthly Weather Review*, *26*, 911–927.

Dowell, D. C., Wicker, L. J., & Stensrud, D. J. (2004). High resolution analyses of the 8 May 2003 Oklahoma City Storm. Part II: EnKF data assimilation and forecast experiments, *Preprints, 22d, Conf. on Severe Local Storms* (Vol. 12.5, pp. 1–6). Hyannis, MA: Amer. Meteor. Soc. Retrieved from https://ams.confex.com/ams/11aram22sls/techprogram/paper_81393.htm

Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C., & Crook, N. A. (2004). Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Monthly Weather Review*, *132*, 1982–2005.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Abingdon, UK: Chapman & Hall/CRC.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(10), 143–162.

Hamill, T. M., & Whitaker, J. S. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation. *Monthly Weather Review*, *133*, 3132–3147.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*, 559–570.

Houtekamer, P. L., Mitchell, H. L., Buehner, M., Charron, M., Spacek, L., & Hansen, M. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, *133*, 604–620.

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*, 4489–4532.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*, 112–126.

Kober, K., & Craig, G. C. (2016). Physically-based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, *73*, 2893–2911.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., & Halliwell, C. (2008). Characteristics of high-resolution versions of the Met Office unified model for forecasting convection over the United Kingdom. *Monthly Weather Review*, *136*(9), 3408–3424.

Lei, L., & Whitaker, S. (2017). Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *Journal of Advances in Modeling Earth Systems*, *9*, 781–789. https://doi.org/10.1002/2016MS000864

Lin, Y., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, *22*, 1065–1092.

Meng, Z., & Zhang, F. (2011). Limited-area ensemble-based data assimilation. *Monthly Weather Review*, *139*(7), 2025–2045.

Minamide, M., & Zhang, F. (2017). Adaptive observation error inflation for assimilating all-sky satellite radiance. *Monthly Weather Review*, *145*(3), 1063–1081.

Minamide, M., & Zhang, F. (2019). An adaptive background error inflation method for assimilating all-sky radiances. *Quarterly Journal of the Royal Meteorological Society*, 1–19. https://doi.org/10.1002/qj.3466

Navarra, A., Kinter, IIIJ. L., & Tribbia, J. (2010). Crucial experiments in climate science. *Bulletin of the American Meteorological Society*, *91*, 343–352.

Pellerin, G., Lefaivre, L., Houtekamer, P., & Girard, C. (2003). Increasing the horizontal resolution of ensemble forecasts at CMC. *Nonlinear Processes Geophysics*, *10*, 463–468.

Rasp, S., Selz, T., & Craig, G. C. (2018). Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *Journal of the Atmospheric Sciences*, *75*, 691–706.

Reinhardt, T., & Seifert, A. (2006). A three-category ice scheme for LMK. *COSMO News Letter*, *6*, 115–120.

Robert, N., & Lean, H. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*, 78–96.

Schättler, U. (2013). A description of the nonhydrostatic regional COSMO-model. Part V: Initial and boundary data for the COSMO-model. Germany: Consortium for Smallscale Modeling (COSMO). Retrieved from http://www.cosmo-model.org/content/model/documentation/core/cosmoInt2lm.pdf

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

Selz, T., Bierdel, L., & Craig, G. C. (2018). Estimation of the variability of mesoscale energy spectra with three years of COSMO-DE analyses. *Journal of the Atmospheric Sciences*, *76*, 627–637. https://doi.org/10.1175/JAS-D-18-0155.1

Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M., Hohenegger, C., et al. (2016). HErZ—the German Hans-Ertel Centre for Weather Research. *Bulletin of the American Meteorological Society*, *97*, 1057–1068. https://doi.org/10.1175/BAMS-D-13-00227.1

Snyder, C., & Zhang, F. (2003). Assimilation of simulated doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, *131*, 1663–1677.

Sobash, R., & Wicker, L. (2015). On the impact of additive noise in storm-scale EnKF experiments. *Monthly Weather Review*, *143*, 3067–3086.

Sommer, M., & Janjić, T. (2018). A flexible additive inflation scheme for treating model error in ensemble Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, *144*, 2026–2037.

Weissmann, M., Göber, M., Hohenegger, C., Janjić, T., Keller, J., Ohlwein, C., et al. (2014). The Hans-Ertel Centre for Weather Research—Research objectives and highlights from its first three years. *Meteorologische Zeitschrift*, *23*(3), 193–208.

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. (2008). Ensemble data assimilation with the NCEP global forecasting system. *Monthly Weather Review*, *136*, 463–482.

Yang, S. C., Kalnay, E., & Enomoto, T. (2015). Ensemble singular vectors and their use as additive inflation in enKF. *Tellus A*, *67*, 781–789.

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The ICON (ICO sahedral non-hydrostatic) modelling framework of DWD and MPI-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*, 563–579.

Zeng, Y., Blahak, U., & Jerger, D. (2016). An efficient modular volume-scanning radar forward operator for NWP models: Description and coupling to the COSMO model. *Quarterly Journal of the Royal Meteorological Society*, *142*, 3234–3256.

Zeng, Y., Blahak, U., Neuper, M., & Jerger, D. (2014). Radar beam tracing methods based on atmospheric refractive index. *Journal of Atmospheric and Oceanic Technology*, *31*, 2650–2670.

Zeng, Y., Janjić, T., de Lozar, A., Blahak, U., Reich, H., Keil, C., & Seifert, A. (2018). Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods and combinations. *Journal of Advances in Modeling Earth Systems*, *10*, 2889–2911. https://doi.org/10.1029/2018MS001375

Zhang, F., Snyder, C., & Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*(5), 1238–1253.