LETKF for the nonhydrostatic regional model COSMO-DE

Hendrik Reich, Andreas Rhodin, Christoph Schraff

Deutscher Wetterdienst, Frankfurter Strasse 135, 63067 Offenbach, Germany

1. Introduction

Data assimilation for numerical weather prediction (NWP) at the convective scale meets with a number of challenges. They include: strongly flow dependent and unknown spatial balances between the different model variables, importance of nonlinear processes, non-Gaussian error statistics and large forecast errors in 'weather'-parameters due to imperfections in the physics, in particular in the cloud and boundary layer formulations.

The Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al., 2007) offers some very attractive features: it is a simple algorithm, no tangent linear and adjoint versions of the prognostic model are required, and the forecast error covariance matrix is cycled and thus flow-dependent.

At Deutscher Wetterdienst (DWD) it is planned to use the LETKF on the global scale (GME/ICON, in a hybrid approach together with 3dVar) as well as on the local scale (COSMO-DE). COSMO-DE is a nonhydrostatic COSMO-version with a horizontal resolution of 2.8 km, covering Germany and parts of its neighbouring countries (Baldauf et al., 2010). The LETKF analysis ensemble will also serve as initial conditions for COSMO-DE EPS, a convection permitting EPS system under development at DWD.

The outline is as follows: we will give a short overview on the LETKF in section 2. In section 3 we present the results of our LETKF experiments and we conclude in section 4.

2. LETKF theory

Our Implementation follows (Hunt et al., 2007). The basic idea of the LETKF is to do the analysis in the space of the *ensemble perturbations*. This is computationally efficient, but also restricts corrections to the subspace spanned by the ensemble. An explicit localization is necessary to confine the ensemble size; this means to compute a separate analysis at every grid point, where only certain observations are selected. Thus, the analysis ensemble members are a locally linear combination of first guess ensemble members.

The analysis mean $\bar{\mathbf{x}}^a$ is given by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{H}\mathbf{X}^b)^T \mathbf{R}^{-1} (\mathbf{y} - \bar{\mathbf{y}}^b) \tag{1}$$

where $\bar{\mathbf{x}}^b$ is the first guess mean; $\mathbf{H}$ is the (linearized) observation operator and $\mathbf{X}^b$ are the first guess ensemble perturbations. The analysis ensemble is obtained via

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

Here, k is the number of ensemble members and $\tilde{\mathbf{P}}^a$ is the analysis error covariance matrix which is (in the ensemble space) given by

$$\tilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{H}\mathbf{X}^b)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{X}^b]^{-1}. \tag{3}$$

3. Experiments and results

We performed several preliminary experiments with successive LETKF assimilation cycles. In all experiments, 32 ensemble members were used. The initial ensemble members where drawn from the 3dVar B-Matrix of the global model GME. Conventional observations from the global network were assimilated. We have run a 3-hourly cycle up to 2 days (7-8 Aug. 2009: 1 quiet + 1 convective day) and used lateral boundary conditions (BC) from COSMO-SREPS (3 * 4 members) or deterministic BC.

In our first experiments we concentrate on general topics, such as the rms/spread ratio of the ensemble, the noise (as measured by dps/dt) and the general behaviour of LETKF (analysis increments, spread structures). The effect of parameter variation (e.g. localization length scales) was tested, but fine tuning is left to be done with longer running experiments.

The set of analysed variables is given by $u, v, w, T, pp, qv, qcl, qci$ (wind components including vertical velocity w, temperature, pressure pertubation, humidity, cloud water and cloud ice content). Here, 'analysed' means that linear combination is applied to these variables; other variables are taken from first guess ensemble members or ensemble mean.

We verify the LETKF results (i.e. the analysis *mean*) against the nudging analysis and observations. When comparing with the nudging analysis one has to take into account that the nudging scheme uses a larger set of observations. A verification tool (deterministic/ensemble scores) is currently under development within the COSMO consortium.

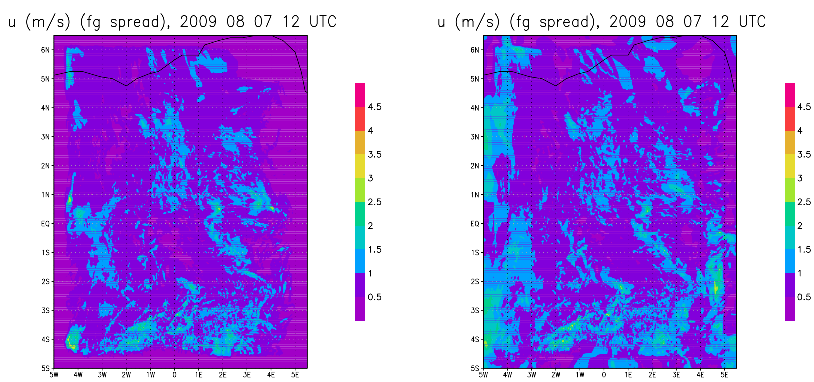

Figure 1: spread (wind component u in m/s) of first guess on 7 Aug. 2009 at 12 UTC (after 4 analysis cycles) for deterministic BC (left) and ensemble BC (right)

Fig.1 shows the spread of the u-wind component of the first guess ensemble, obtained with deterministic and ensemble BC, respectively. In the case of deterministic BC we observe a lack of spread at the lateral boundaries, whereas "new" spread is coming in from the west with ensemble BC. This demonstrates the need to use ensemble BC's, and it can be seen that a large amount of the whole domain is influenced by the spread stemming from the BC. As we will see later, the use of ensemble BC leads to some difficulties.

In order to test the implementation of the LETKF and it's capability of making use of observations we compare the analysis produced by the LETKF with a free forecast which uses the same BC but no observations. Fig.2 shows the temporal development of the first guess and free forecast rmse of the u-component of wind velocity (as measured with respect to the nudging analysis) on the 500 hPa level. One can see that the LETKF performs better than the free forecast on all levels.

Next we verify the LETKF analysis against observations; the results are shown in Fig. **??**.

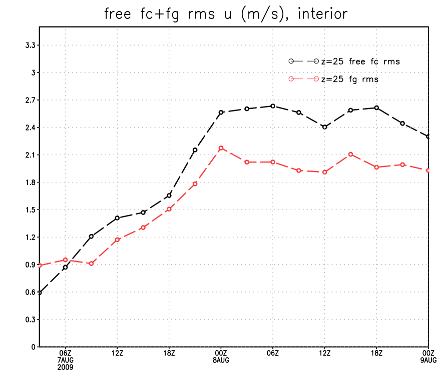

Figure 2: rms of u,(m/s) (interior) of first guess and free forecast; results for det BC.

The reduction of spread between first guess and analysis indicates that the LETKF makes use of the observations. As also the rmse of the analysis is smaller than that of the first guess we conclude that the LETKF is able to use the information contained in the observations. The spread of first guess and analysis is much smaller than the corresponding rmse; this means that the ensemble is underdispersive.

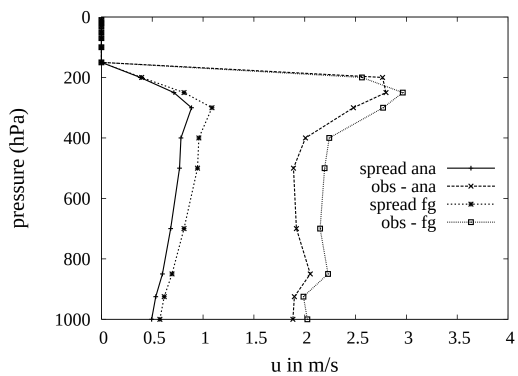

Figure 3: time average (20090807 15 UTC - 20090809 00 UTC)of obs-fg and spread of u,(m/s) (whole area), AIREP

The lack of spread is (partly) due to model error which is not accounted for so far. One (simple) method to increase spread is multiplicative covariance inflation:

$$\mathbf{X}_b \rightarrow \rho \mathbf{X}_b \tag{4}$$

with $\mathbf{X}_b$ being the ensemble perturbations and $\rho > 1$. The tuning of the inflation factor ρ takes much time, and it is expected that the optimal value will change in time, depending e.g. on observation density. For this reason, an adaptive procedure is preferable. (Li et al., 2009) propose an online estimation of the inflation factor. The idea is to compare the “observed” obs minus first guess, given by $(\mathbf{y} - H(\mathbf{x}_b))$ with the “predicted” one, given by $(\mathbf{R} + \mathbf{H}\mathbf{P}_b\mathbf{H}^t)$. This method was applied in a LETKF environment by Bonavita et al. (2010), where ρ was time and space dependent. Here, in a first step, we tested a version with a space independent ρ.

It is also assumed that the observation errors and thus the $\mathbf{R}$-matrix are specified incorrectly, and a correction is desirable. This can be achieved by comparing the observed observation covariance with the assumed one (in ensemble space) and correcting $\mathbf{R}$ automatically if necessary. Both methods (est. of inflation factor / $\mathbf{R}$ matrix) have been tested with reasonable numerical cost and success with a toy model, and have been implemented in the COSMO LETKF. For deterministic BC, a positive effect of the adaptive ρ inflation is visible, which is shown in Fig.4. In the case with ensemble BC the method was not succesful. Currently, an improved version with a space dependent ρ and doing the computation in ensemble space is tested.

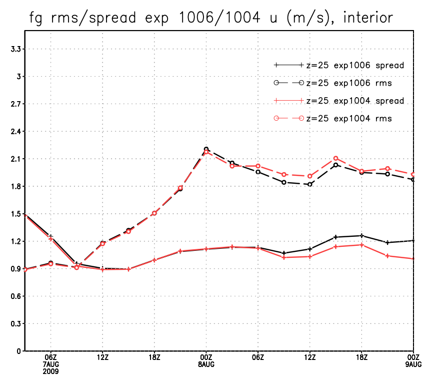

Figure 4: intercomparison of first guess rms and spread of u,(m/s) (interior); results for det BC and constant inflation factor ρ (exp1004) and adaptive covariance inflation (exp1006)

The LETKF produces an analysis ensemble as a (local) linear combination of the first guess ensemble. The analysis fields obtained are not necessarily balanced, and noise (e.g. external gravity waves, measured by dps/dt) might be present when starting the integration. Indeed we find an increased level of noise (as compared with the nudging scheme). Fig.5 (left plot) shows that noise is present in the whole domain.

We observed that the diagonal elements of weight matrices W are larger than the off diagonal elements; this means that the analysis ensemble member k gets the largest contribution from first guess ensemble member k plus (smaller) corrections from members $i \neq k$. Thus, the difference between analysis and first guess ensemble member k (the analysis increment) is small compared to the full fields, and hydrostatic balancing can be applied to this increment: this leaves the full fields nonhydrostatic as it should be in a nonhydrostatic model.

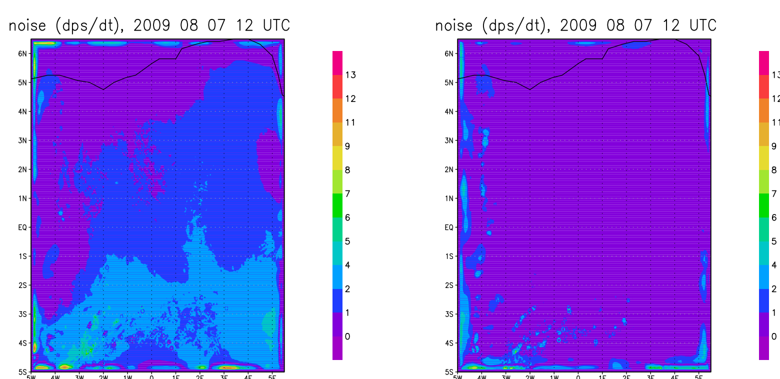

Figure 5: area plots of noise (dPs/dt) at the first time step; ens BC without hydrostatic balancing (left) and with hydrostatic balancing applied (right).

Applying this method reduces the noise in the interior of the domain (Fig.5, right plot); at the boundaries, there is still noise present.

4. Conclusions

We have tested the LETKF in preliminary, short assimilation cycles. The LETKF demonstrated its capability of assimilating conventional observations correctly. Problems such as a lack of spread and noise introduced by the ensemble BC were identified. The latter could be alleviated by applying hydrostatic balancing to the analysis increments. The effect of this balancing on the rms/spread ratio will have to be investigated. Furthermore, we will study the effect of the remaining noise on e.g. precipitation at the beginning of the integration.

The adaptive covariance inflation, which was tested in a simple version, was successfully applied in the case of deterministic BC. For ensemble BC a more sophisticated version is currently tested. Within the COSMO consortium, alternative methods to account for model errors are developed and will also be implemented in the LETKF.

In the future we will run experiments over a period of weeks or months. In this experiments, we will use more observations (radar data in particular) and do the analysis more frequently ($\approx$ all 15 min). With this more realistic setup, parameters as the localization length scales will have to be tuned.

References

[1] Baldauf M., Seifert A., Förstner J., Majewski D., Raschendorfer M., Reinhardt T., 2010: Operational convective-scale numerical weather prediction with the COSMO model. *submitted to MWR*

[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

[3] Hunt B.R., Kalnay E., Kostelich E.J., Szunyogh I., 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **230**: 112-126

[4] 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

[5] Stephan K., S. Klink, C. Schraff, 2008: Assimilation of radar derived rain rates into the convective scale model COSMO-DE at DWD. *Q.J.R.Meteorol.Soc.* **134**: 1315-1326