LETKF for the COSMO-DE

H. Reich, A. Rhodin, C. Schraff

SRNWP workshop Bad Orb 2011

16-18. May 2011

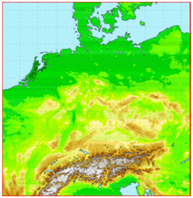
LETKF / KENDA at DWD

- global scale (GME/ICON): hybrid system (3dvar/LETKF) planned
- convective scale: *KENDA* project

  Kilometerscale **En**semble **D**ata **A**ssimilation

  - priority project within COSMO consortium
  - at DWD: Hendrik Reich, Andreas Rhodin, Christoph Schraff
  - will be applied to nonhydrostatic COSMO-DE model

COSMO-DE domain



LETKF basics

- Implementation following *Hunt et al., 2007*
- basic idea: do analysis in the space of the ensemble perturbations
  - computational efficient, but also restricts corrections to subspace spanned by the ensemble
  - explicit localization (doing separate analysis at every grid point, select only certain obs)
  - analysis ensemble members are locally linear combination of first guess ensemble members

LETKF experiments

- technical implementation of experiments (up to now):
  - stand-alone LETKF script environment to run COSMO-DE LETKF + diagnostics / plotting
  - toy model (Lorenz-96,40 grid points) to test LETKF components
- preliminary experiments with successive LETKF assimilation cycles (32 ensemble members, drawn from 3dVar B-Matrix)
  - 3-hourly cycles, up to 2 days (7-8 Aug. 2009: quiet + convective day)
  - lateral boundary conditions (LBC) from COSMO-SREPS (3 * 4 members) or deterministic LBC
  - old experiments: use obs from GME NetCDF feedback files (sparse density)
  - new experiments: use obs from NetCDF files written by COSMO-model during integration (same obs set as nudging); QC of obs has to be adapted
  - option for *deterministic analysis* has been implemented

LETKF experiments

- results are *preliminary* because of:
  - 3h update (later $\approx$ 15 min)
  - sparse observation density (old exp.)
  - only 2 day period
- $\rightarrow$ concentrate on general topics:
  - rms/spread of ensemble
  - noise (dps/dt and $wa500$)
  - general behaviour of LETKF (analysis increments etc.)
- test effect of parameter variation, but no fine tuning
- some (adaptive) methods to increase spread/reduce noise have been tested with toy model/COSMO-LETKF

LETKF experiments

- analysed variables are $u, v, w, T, pp, qv, qcl, qci$
- analysed means that linear combination is applied to these variables (other variables taken from first guess ensemble / ensemble mean)
- verify LETKF *mean* / *det run* against
  - nudging analysis (u, v, T, pp, qv)
  - observations (u, v, T)
- verification tool (deterministic/ensemble scores) is currently under development

spread (ens BC)

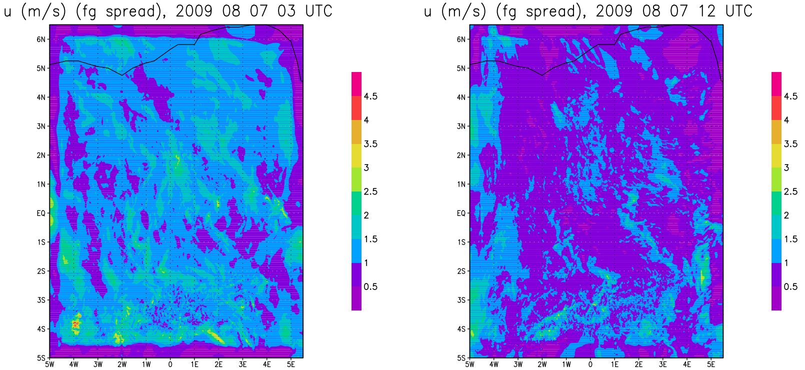


Fig.1: spread (wind component u in m/s) of first guess on 7 Aug. 2009 at 03 UTC (after 1 LETKF analysis with 3DVAR-B) (left) and at 12 UTC (after 4 analysis cycles) (right)

The large scale spread decreases and “new” spread comes in from the west due to the lateral boundary fields.

spread (det BC)

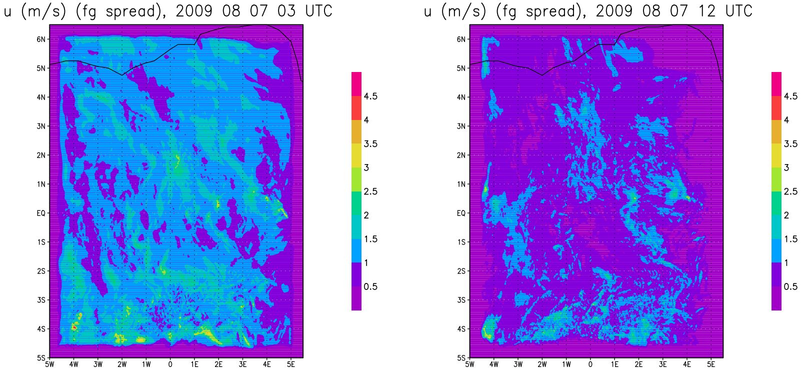


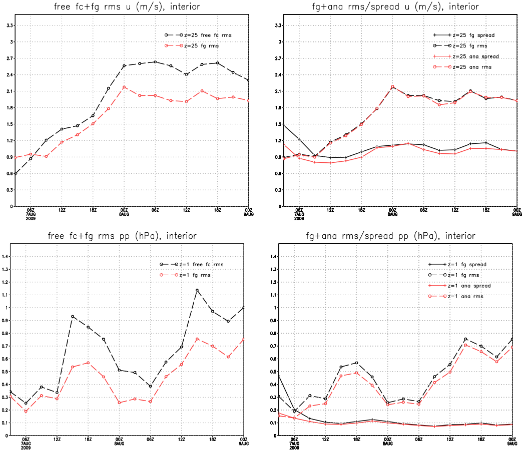
Fig.2: same as Fig.1 but with *deterministic* boundary conditions

The large scale spread decreases faster as no “new” spread comes in from the lateral boundary fields.

free fc and rms/spread (interior), det BC

free fc rmse

first guess rmse



first guess

rmse+spread

analysis

rmse+spread

Fig.3: upper row: *u* (m/s) at 500 hPa; lower row: pressure perturbation *pp* (hPa).

LETKF performs better than free forecast (→ obs info is assimilated), but small differences between rms of analysis/forecast (esp. for *u*)

free fc and rms/spread (interior), ens BC

free fc rmse

first guess rmse

first guess

rmse+spread

analysis

rmse+spread

Fig.4: upper row: *u* (m/s) at 500 hPa; lower row: pressure perturbation *pp* (hPa).

larger differences between rms of first guess/analysis (esp. for *pp*)

u obs-fg/spread (time average, whole area), ens BC (exp 1008)

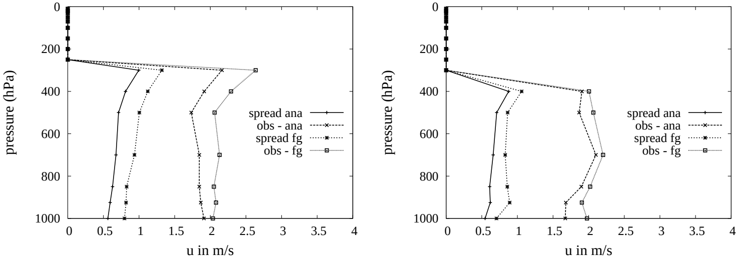


Fig.5: time average (20090807 15 UTC - 20090809 00 UTC)of obs-fg and spread of u (m/s) (whole area), AIREP (left) and TEMP (right); results for ens BC and active vertical localization (exp1008)

larger differences between analysis and first guess at observation locations, but LETKF is underdispersive

u rms/spread (interior), det BC (effect of vertical localization)

no vert loc

vert loc

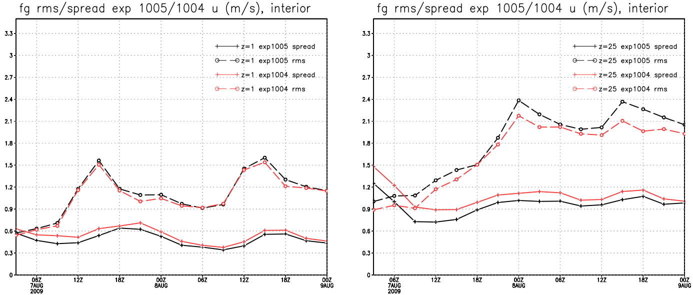


Fig.6: intercomparison of first guess rms and spread of *u*,(m/s) (interior); results for det BC and active vertical localization (exp1004) and no vertical localization (exp1005)

turning off the vertical localization increases rms and decreases spread; but effect on noise → later!

adaptive methods

- 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:
  - $X_b \to \rho X_b$ with $\rho > 1$
  - more advanced methods to account for model error (esp. in limited-area models) need to be developed
- problem: tuning inflation factor ρ takes much time, adaptive procedure preferable
  - (*Li et al.*:) online estimation of inflation factor
  - compare "observed" (obs - f.g.) : $(y - H(x))$
    with "predicted" (obs - f.g.) : $(R + HP_bH^t)$
  - Following *Bonavita et al.* ρ is computed at every gridpoint

adaptive methods

- obs errors / R-matrix probably assumed incorrectly, correction desirable
  - compare observed obs covariance with assumed one and correct R automatically if necessary
  - this is done in *ensemble space*
- both methods (est. of inflation factor / R matrix) have been tested with reasonable numerical cost and success within the toy model, and have been implemented in the LETKF (COSMO and GME)
- adaptive ρ inflation works quite well (see next slides), adaptive R matrix correction does not show positive effects right now

u/T rms/spread (interior), ens BC (effect of adaptive cov. inflation)

ρ adap

ρ const

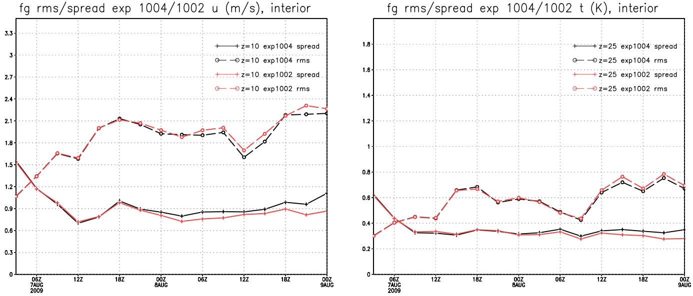
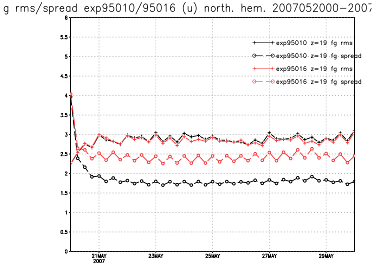


Fig.7: intercomparison of fg rms / spread with adapt. cov. inflation (exp1004) and $\rho = \text{const.}$ (exp1002); results for ens BC (u in m/s at 850 hPa (left), T in K at 500 hPa (right).

adaptive covariance inflation increases spread and (on average) decreases rms, but effect is relatively small

u rms/spread (GME), (effect of adaptive cov. inflation)



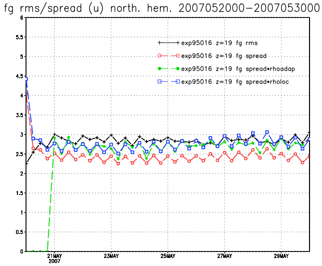


Fig.8: left: intercomparison of fg rms and spread with adaptive covariance inflation (exp95010) and $\rho =$ const. (exp95016); results from GME (u in m/s at 500 hPa, northern hemisphere) right: fg rms and spread mulitplied with adaptively obtained inflation factor.

larger effect on spread, effect on rms still small

u rms/spread (COSMO), (effect of adaptive cov. inflation)

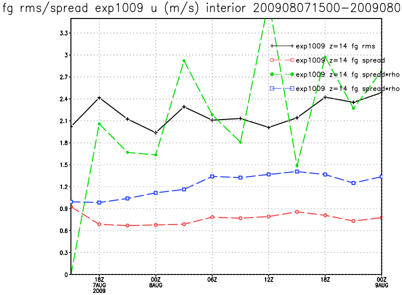


Fig.9: same as Fig. 8 right, but results from COSMO (u in m/s at 850 hPa, interior)

effect of vertical localization on noise

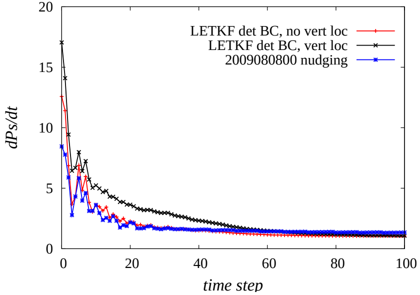


Fig.10: noise (*dPs*/*dt* in *Pa*/*s*, area mean) of one ensemble member at 20090808 00 UTC (left) and 09 UTC (right) for det BC and vertical localization switched on/off (time step: 25 s)

Noise decreases for vertical localization switched off

weight matrices

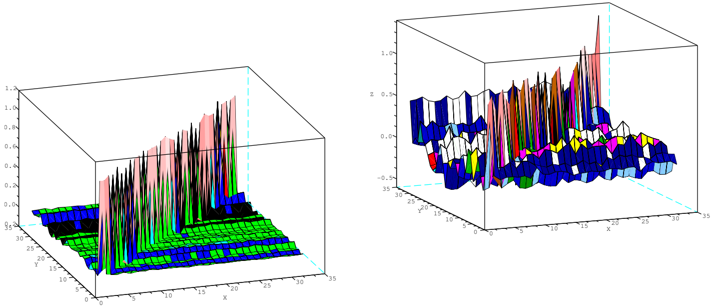


Fig.11: weight matrices (the matrix the first guess ensemble is multiplied with), for a case with "normal" number of observations (left) and with many observations (or small obs. errors; right).

off diagonal elements even for large number of obs ≤ 0.5 and diagonal elements > 0.5

hydrostatic balancing

- diagonal elements of weight matrix are larger than off diagonal elements
- $\rightarrow$ analysis ensemble k gets 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
- apply hydrostatic balancing to this increment; this leaves the full fields nonhydrostatic as it should be in a nonhydrostatic model

effect of hydrostatic balancing on noise

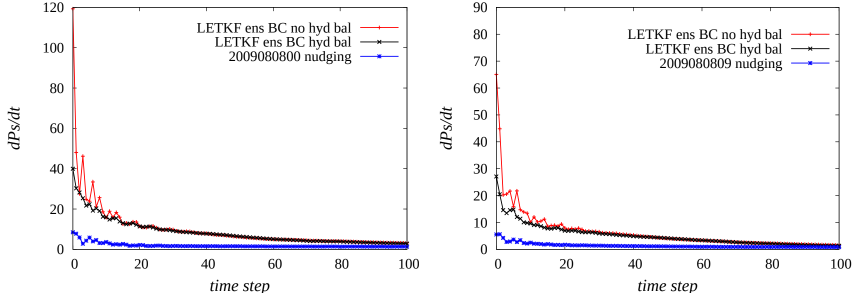


Fig.12: noise (dPs/dt in Pa/s, area mean) of one ensemble member at 20090808 00 UTC (left) and 09 UTC (right) for ens BC with hydrostatic balancing switched on/off

Noise is reduced by applying hydrostatic balancing

noise: area plots

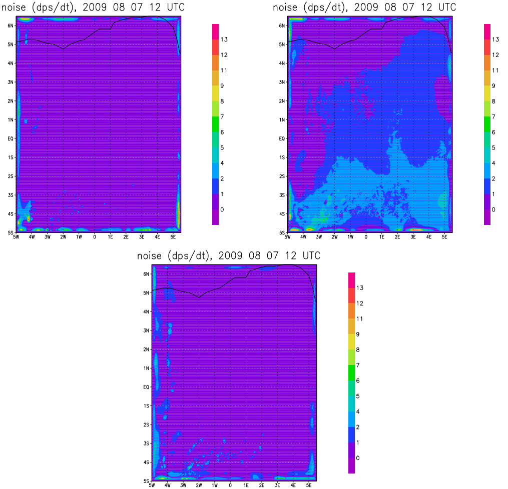


Fig.13: area plots of dPs/dt, 1st. time step; *analysis* with det. BC first guess, integration with ens BC; ens. BC first guess and ens BC integration; ens. BC first guess and ens BC integration, but hydrostatic balancing applied.

hydrostatic balancing reduces noise in the interior, no effect at the boundaries

effect of hydrostatic balancing on rms/spread

no hyd bal

hyd bal

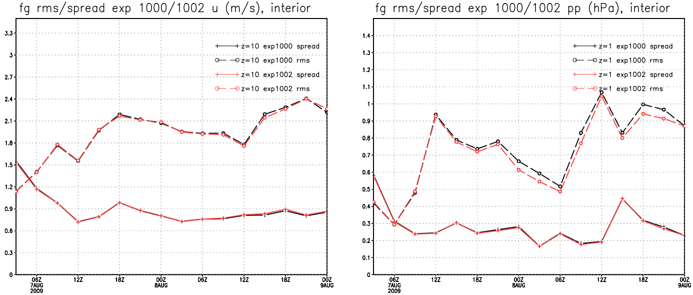


Fig.14: effect of hydrostatic balancing on rms/spread (*u* in *m*/*s*, 850 *hPa* (left); *pp* in *hPa*, surface level (right))

effect on rms/spread is neutral for most levels/variables, positive effect for rms of pressure deviation in low levels

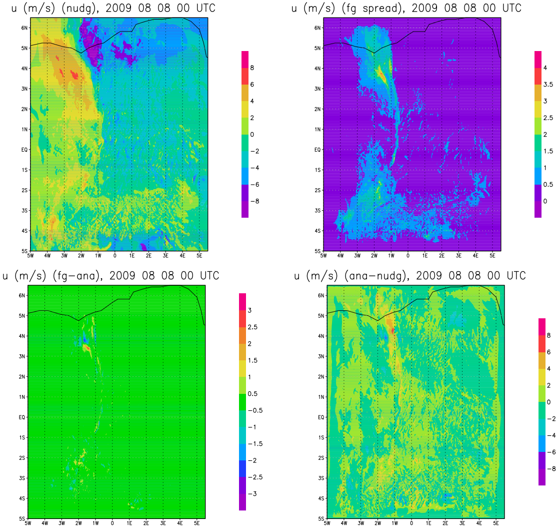


Fig.15: nudging at 2009080800 UTC, model level 1, spread (u in m/s) (det. BC) of first guess, fg minus LETKF analysis and LETKF analysis minus nudging

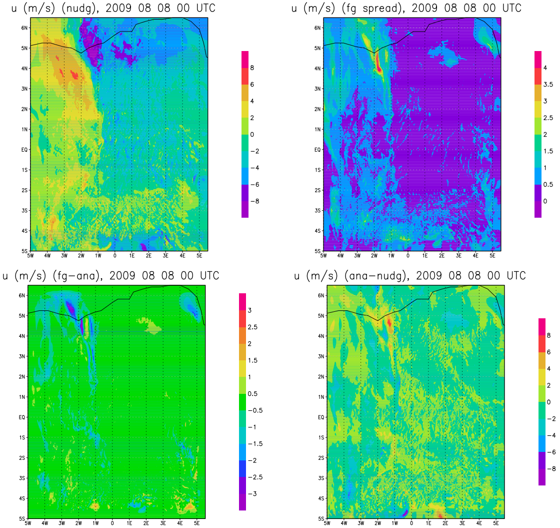


Fig.16: nudging at 2009080800 UTC, model level 1, spread (u in m/s) (ens. BC) of first guess, fg minus LETKF analysis and LETKF analysis minus nudging

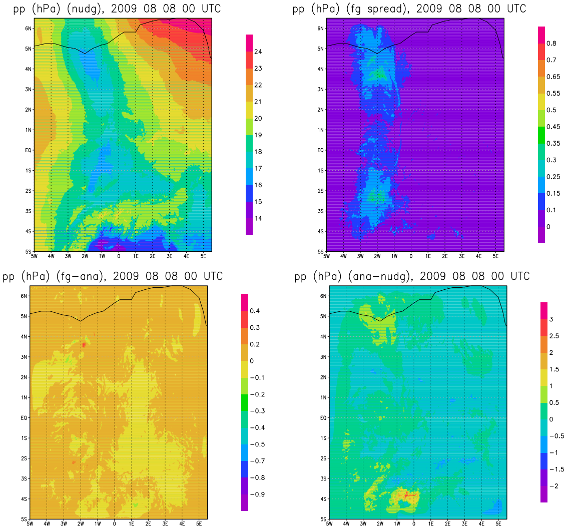


Fig.17: nudging at 2009080800 UTC, model level 1, spread (pp in hPa) (det. BC) of first guess, fg minus LETKF analysis and LETKF analysis minus nudging

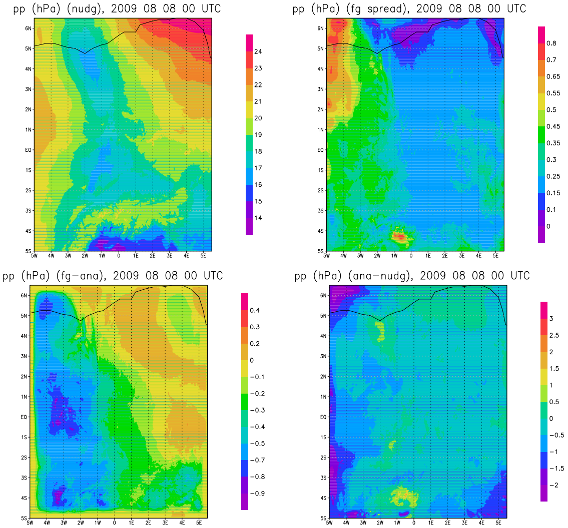


Fig.18: nudging at 2009080800 UTC, model level 1, spread (pp in hPa) (ens. BC) of first guess, fg minus LETKF analysis and LETKF analysis minus nudging

Conclusions / open questions

- noise:
  - no vertical localization: rms/spread ratio gets worse, noise slightly reduced; $\rightarrow$ use vertical localization
  - hydrostatic balancing of analysis increments reduces noise (at the beginning of integration) - effect on rms/spread is neutral (slightly positive for press. pert.)
- spread:
  - structures (small at large scales, high values at small scales) seem to be appropiate (in cases studied so far), but amplitude too small
  - adaptive covariance inflation increases spread, effect on rms smaller
  - further study adaptive R-matrix correction (increase spread in *analysis*)
- boundary conditions:
  - ens BC are essential, but also cause some problems

Outlook / next steps

next steps:

- for runs with obs from COSMO feedback files: tune QC of obs
- increase update frequency, use NUMEX
- compare performance of deterministic run with mean
- runs with BC from global LETKF; look at spread of BC

Outlook:

- further examine and combine adaptive methods
- tuning of parameters , e.g. localization length scales
- model error (model perturbations): 2 projects within COSMO to account for model error; (stochastic) physics perturbations
- additional observations: radar data (radial winds, reflectivity), GPS, ...

LETKF Theory

- let $\mathbf{w}$ denote gaussian vector in k-dimensional ensemble space with mean 0 and covariance $\mathbf{I}/(k-1)$
- let $\mathbf{X}^b$ denote the (background) ensemble perturbations
- then $\mathbf{x} = \bar{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}$ is the corresponding model state with mean $\bar{\mathbf{x}}^b$ and covariance $\mathbf{P}^b = (k-1)^{-1}\mathbf{X}^b(\mathbf{X}^b)^T$
- let $\mathbf{Y}^b$ denote the ensemble perturbations in observation space and $\mathbf{R}$ the observation error covariance matrix

LETKF Theory

- do analysis in the k-dimensional ensemble space

$$\bar{\mathbf{w}}^a = \tilde{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y} - \bar{\mathbf{y}}^b)$$
$$\tilde{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b]^{-1}$$

- in model space we have

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b \bar{\mathbf{w}}^a$$
$$\mathbf{P}^a = \mathbf{X}^b \tilde{\mathbf{P}}^a (\mathbf{X}^b)^T$$

- Now the analysis ensemble perturbations - with $\mathbf{P}^a$ given above - are obtained via

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}^a,$$

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

LETKF Theory

- it's possible to obtain a *deterministic run* via

$$\mathbf{x}_a^{det} = \mathbf{x}_b^{det} + \mathbf{K}\left[\mathbf{y} - H(\mathbf{x}_b^{det})\right]$$

  with the *Kalman gain* **K**:

$$\mathbf{K} = \mathbf{X}_b \left[(k-1)\mathbf{I} + \mathbf{Y}_b^T \mathbf{R}^{-1} \mathbf{Y}_b\right]^{-1} \mathbf{Y}_b^T \mathbf{R}^{-1}$$

- the deterministic analysis is obtained on the same grid as the ensemble is running on; the *analysis increments* can be interpolated to a higher resolution