Overview of the COSMO Priority Project **KENDA**
for Km-Scale Ensemble-Based Data Assimilation

Christoph Schraff,
Deutscher Wetterdienst, Offenbach, Germany

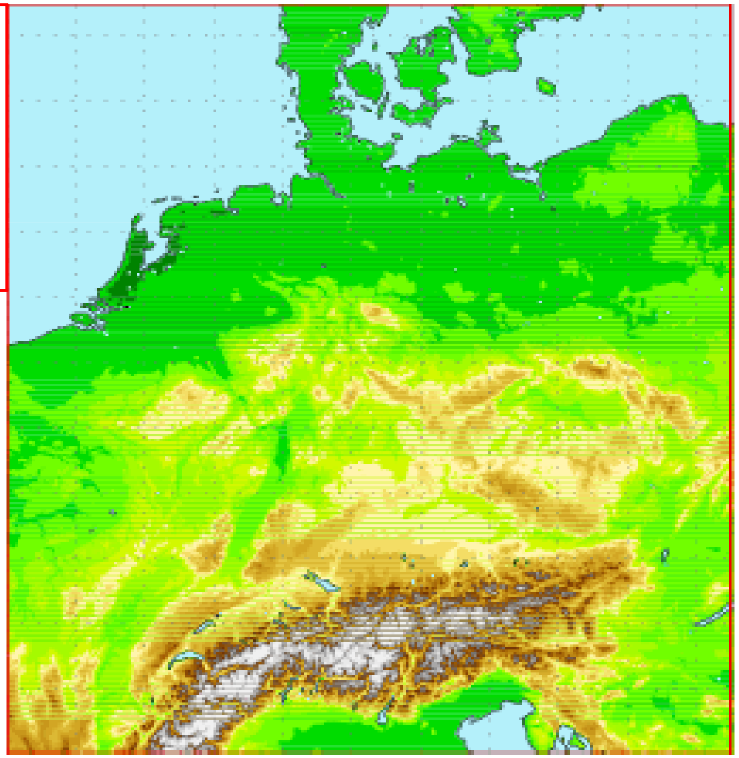
Hendrik Reich, Andreas Rhodin (DWD)

Roland Potthast, Ulrich Blahak, Klaus Stephan (DWD)
Yuefei Zeng, Dorit Epperlein (KIT, D)
Daniel Leuenberger, Tanja Weusthoff (MeteoSwiss)
Mikhail Tsyrulnikov, Vadim Gorin (RosHydromet, RU)
Amalia Iriza (NMA, RO)
Marek Lazanowicz (IMGW, PL)
Lucio Torrisi (CNMCA, I)

KENDA : Why develop Ensemble-Based Data Assimilation ?

COSMO-DE: Δx = 2.8 km
(deep convection explicit,
shallow convection param.)

domain size : ~ 1250 x 1150 km



convection-permitting NWP:
after ‘few’ hours,
a forecast of convection is a long-term forecast

→ deliver probabilistic (pdf) rather than deterministic forecast

→ need ensemble forecast and data assimilation system

→ forecast component: COSMO-DE EPS

KENDA : Why develop Ensemble-Based Data Assimilation ?

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

→ data assimilation: priority project within COSMO consortium
Km-scale ENsemble-based Data Assimilation (**KENDA**):

→ Local Ensemble Transform Kalman Filter (LETKF) ,
(because of its relatively low computational costs)

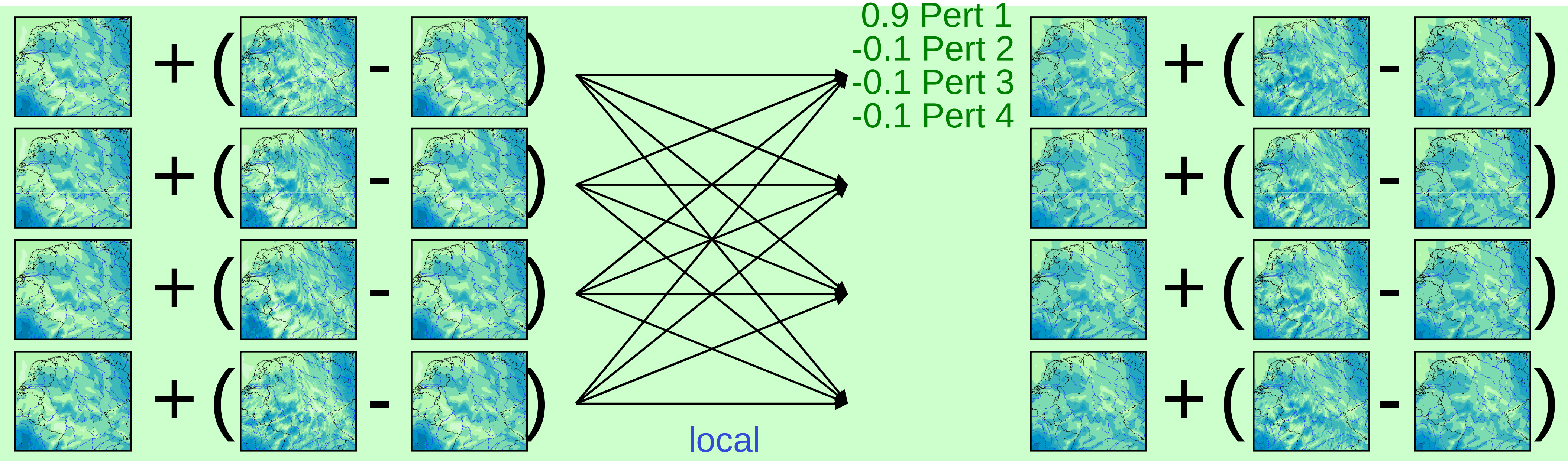
This talk: - method and implementation
- some scientific issues

talk by Hendrik Reich : preliminary experiments at DWD

Assume: Gaussian errors

$$J(\mathbf{x}) \;=\; \left(\mathbf{x}-\overline{\mathbf{x}}^{b}\right)^{T}\mathbf{P}^{b^{-1}}\left(\mathbf{x}-\overline{\mathbf{x}}^{b}\right) + \left[\mathbf{y}^{\mathbf{O}}-H(\mathbf{x})\right]^{T}\mathbf{R}^{-1}\left[\mathbf{y}^{\mathbf{O}}-H(\mathbf{x})\right]$$

for $\mathbf{P}^{b} = (k-1)\,\mathbf{X}^{b}(\mathbf{X}^{b})^{T}$, $J(\mathbf{x})$ is well-defined in sub-space S spanned by $\mathbf{X}^{b}$



ensemble mean forecast

k perturbed forecasts

ensemble mean fcst.

forecast perturbations $\mathbf{X}^b$

↓

flow-dep background error cov.

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

↓

transform matrix $\mathbf{w}$

analysis mean $\overline{\mathbf{w}}^a$

analysis perturbations $\mathbf{W}^{a(i)}$

analysis error covariance (computed only in ensemble space)

$$\mathbf{w}^{a(i)} = \overline{\mathbf{w}}^a + \mathbf{W}^{a(i)}$$

perturbed analyses

↑

in the (k-1) -dimensional (!) sub-space S spanned by background perturbations :

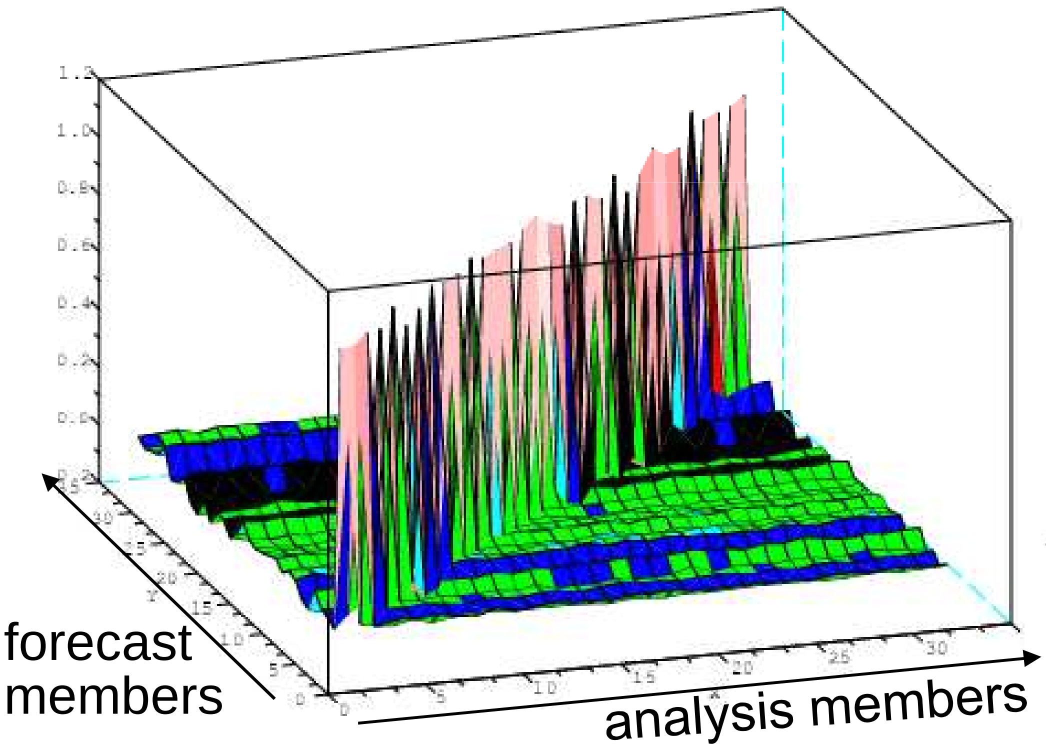
$$\mathbf{x} = \overline{\mathbf{x}}^b + \mathbf{X}^b \mathbf{w}$$

set up cost function $J(\mathbf{w})$ in ensemble space,
explicit solution for minimisation (Hunt et al., 2007)

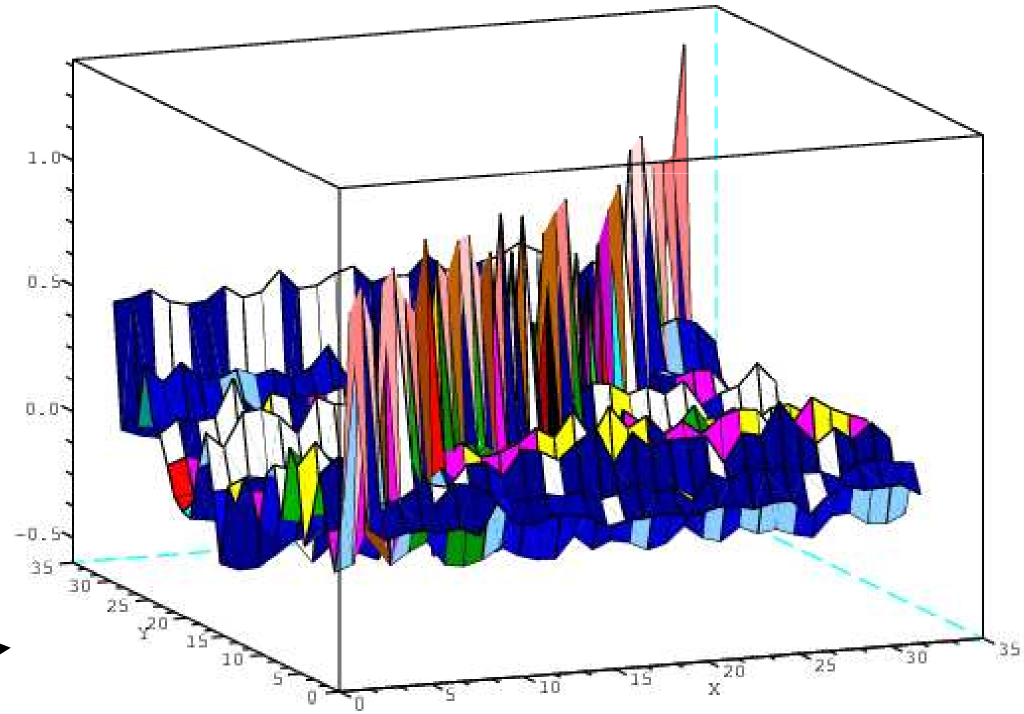
LETKF for COSMO : summary of method

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

weight matrices (transform matrices) : determine linear combination



normal values for obs errors **R**
(for a grid point influenced by
> 200 conventional obs)



R divided by 100
(simulating many more obs)

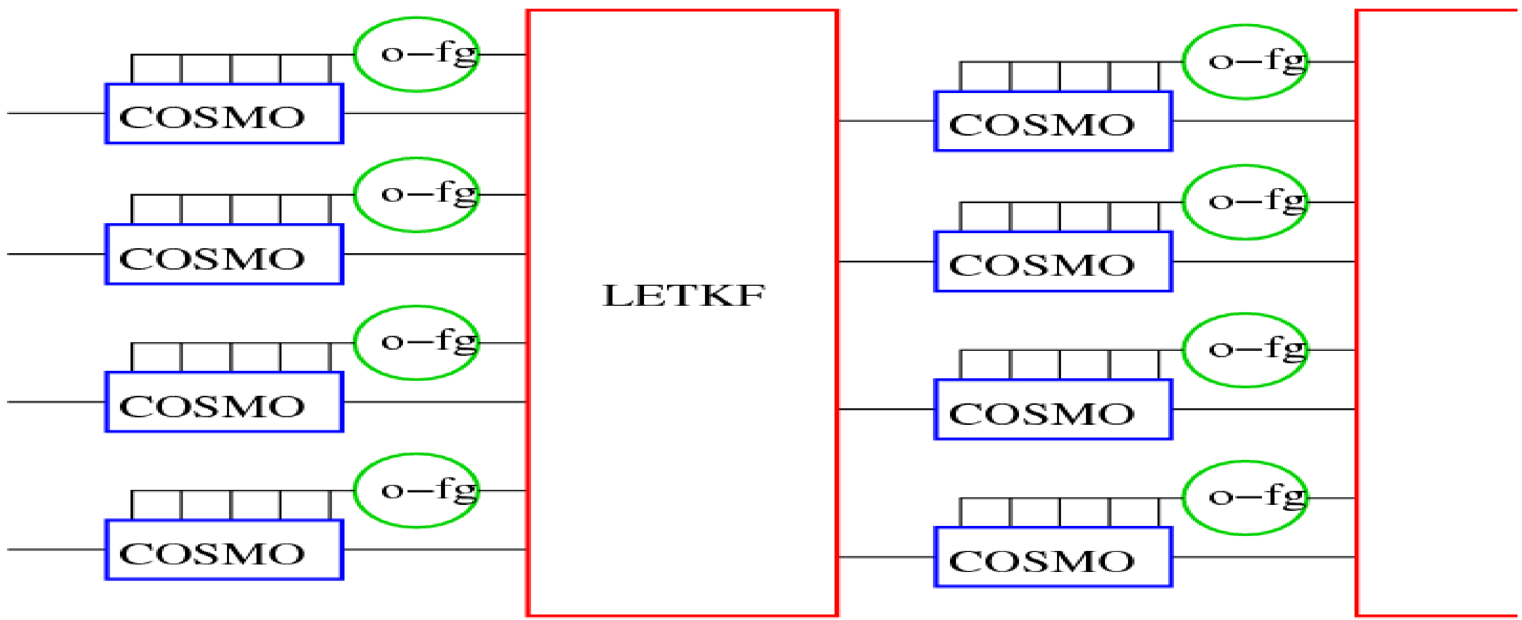
→ diagonal elements >> off-diagonal elements (→ analysis increments 'small')

→ 'good' forecast members get larger weight in all analysis members

LETKF for COSMO : technical implementation

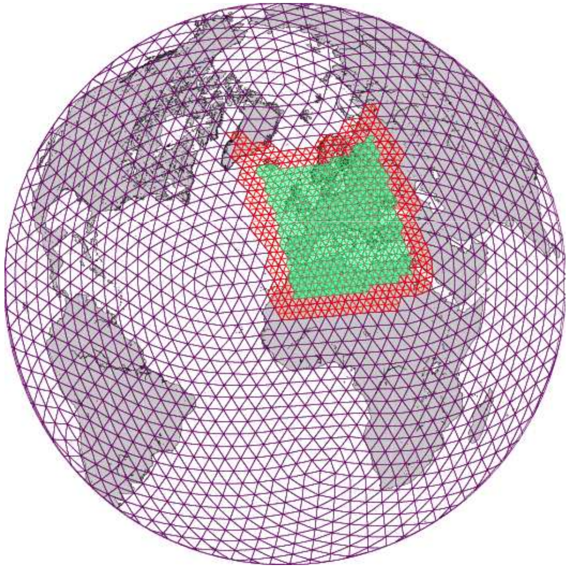
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- analysis step (LETKF) outside COSMO code
  - → ensemble of independent COSMO runs up to next analysis time (collecting obs – f.g. → 4D -LETKF)
  - → separate analysis step code, LETKF included in 3DVAR package of DWD



- basically for verification purposes, COSMO obs operators incl. quality control will be implemented in 3DVAR / LETKF environment
  - → future: hybrid 3DVAR-EnKF approaches in principle applicable to COSMO

LETKF (km-scale COSMO) : implementation



- lateral BC :
  - future: from global EnKF / EPS based on
    ICON (non-hydrostatic,
    with regional grid refinement)
  - currently: COSMO-SREPS (or deterministic)
- standard experimentation system not yet adapted to perform LETKF (but soon)
  $\rightarrow$ stand-alone scripts, only preliminary LETKF experiments up to now

$\rightarrow$ talk by Hendrik Reich

LETKF (km-scale COSMO) : scientific issues / refinement

(forecast / analysis) ensemble spread 'characterises' (forecast / analysis) error, but

- ensemble size is limited, ensemble can only sample but not fully represent errors
- model error is not accounted for by algorithm

→ lack of spread: (partly) due to model error and limited ensemble size
which is not accounted directly by the algorithm

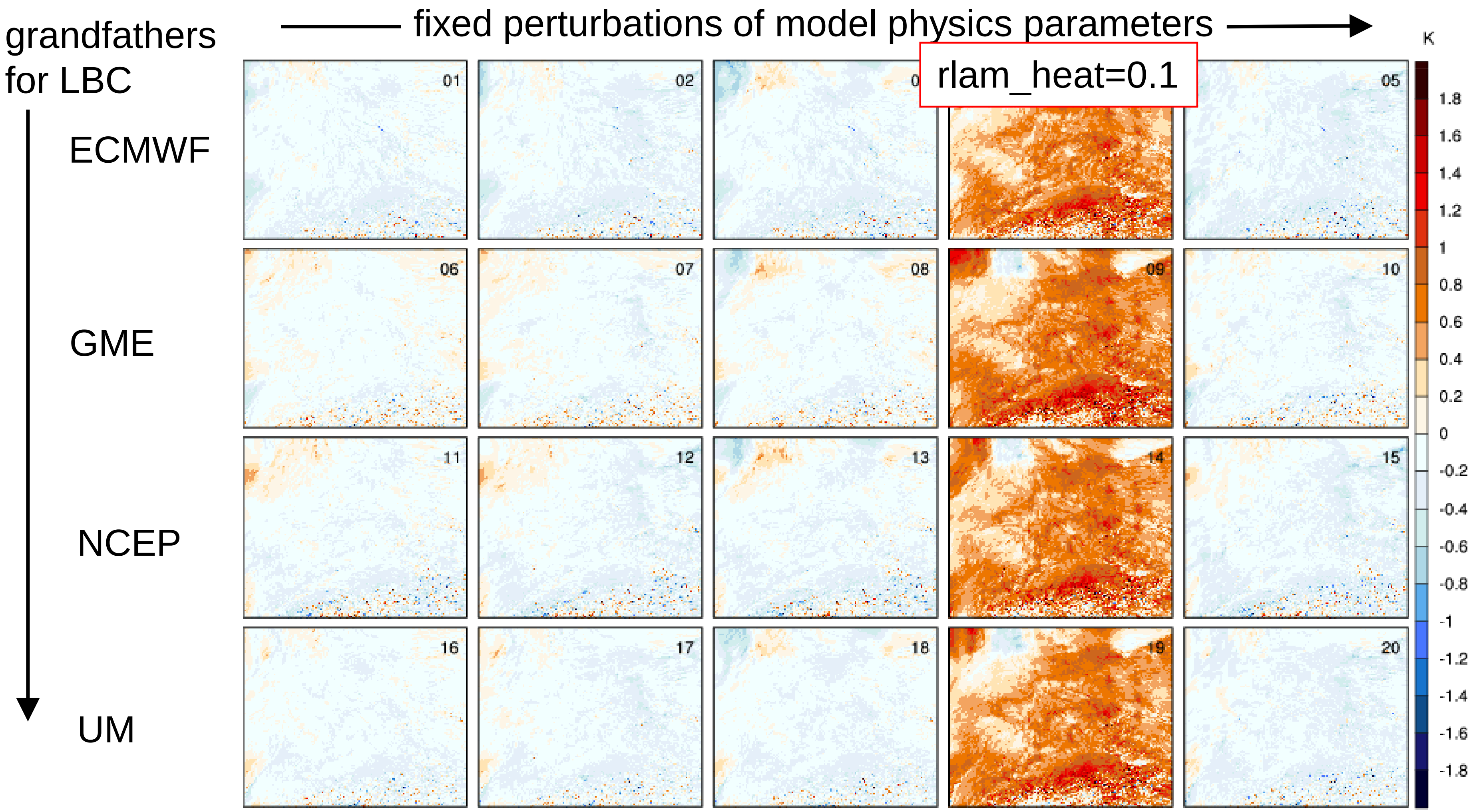
to account for it: covariance inflation, what is needed ?

→ multiplicative $X_b \rightarrow \rho \cdot X_b$ (tuning, or adaptive ($y - \mathbf{H}(x) \sim \mathbf{R} + \mathbf{H}^T \mathbf{P}_b \mathbf{H}$))

→ additive : perturbing the NWP model
  – fixed perturbations of model physics parameters

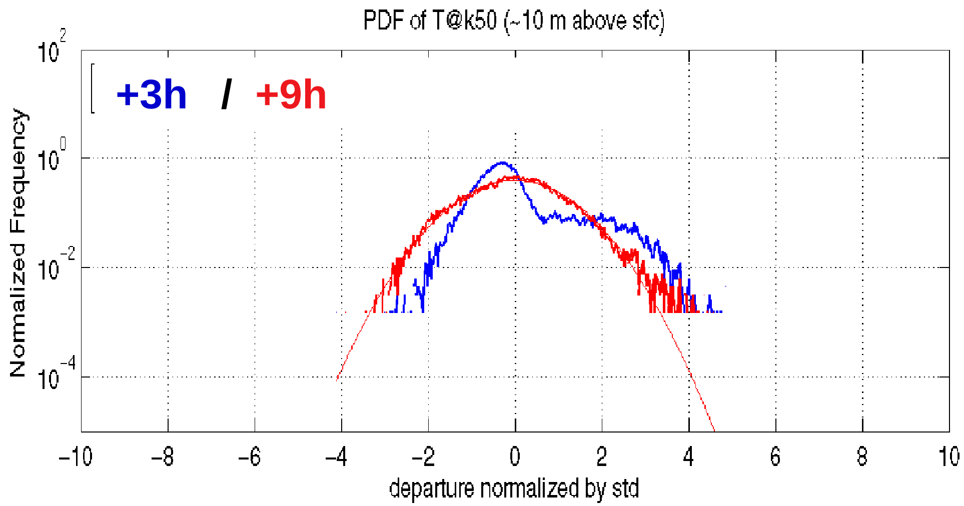
additive covariance inflation:
model perturbations

ensemble forecast perturbations from COSMO-DE-EPS
here: a result from an old version without initial perturbations



temperature at ~10m, +3h (03 UTC)

additive covariance inflation: model perturbations



ensemble forecast perturbations, statistics over 9 days
$T_{10\ m}$ over Northern Germany

→ ≤ +3h : bi-modal pdf
but EnKF optimal
only for Gaussian pdf

→ physics perturbations that may be appropriate for the forecast component of an EPS need not be appropriate for DA (EnKF)

- additive covariance inflation: perturbing the NWP model
  - fixed perturbations of model physics parameters : no
  - stochastic physics (will be implemented)
  - statistical 3DVAR-**B** → hybrid schemes !
  - additive inflation which reflects model error as estimated by statistics (comparing forecast tendencies with observed tendencies, Gorin & Tsyrulnikov)

LETKF (km-scale COSMO) : scientific issues / refinement

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- localisation (multi-scale data assimilation,
  successive LETKF steps with different obs / localisation ?)

- update frequency $\Delta_a t$? 1 hr $\geq \Delta_a t \geq$ 15 min
  non-linearity vs. noise / lack of spread / 4D property ?

- perturbed lateral BC , how to deal with it ?
  ($\rightarrow$ source of noise)
  (distort implicit error covariances in filter $\rightarrow$ limit use of obs ?)

  note: shorter data cut-off & higher analysis frequency
  for COSMO-DE than for driving global system ICON

- non-linear aspects, convection initiation (outer loop , (latent heat nudging) ?)

LETKF (km-scale COSMO) : current work : new observation types

- **radar : radial velocity and (3-D) reflectivity**

- **ground-based GPS slant path delay** (direct use in LETKF , or tomography)

- **cloud information based on satellite and conventional data**
  - derive incomplete analysis of cloud top + cloud base, using conventional obs (synop, radiosonde, ceilometer) and NWC-SAF cloud products from SEVIRI, use obs increments of cloud or cloud top / base height or derived humidity
  - or use SEVIRI radiances directly

(Issues in LETKF: **non-Gaussian** distribution of obs increments, non-linear obs opr, non-local obs, obs error correlations / thinning …)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

thank you for your attention

$$J(\mathbf{x}) \;=\; \left(\mathbf{x}-\overline{\mathbf{x}}^b\right)^T \mathbf{P}^{b^{-1}} \left(\mathbf{x}-\overline{\mathbf{x}}^b\right) + \left[\mathbf{y}^{\mathbf{O}} - H(\mathbf{x})\right]^T \mathbf{R}^{-1} \left[\mathbf{y}^{\mathbf{O}} - H(\mathbf{x})\right]$$

for $\mathbf{P}^b = (k-1)\,\mathbf{X}^b\,(\mathbf{X}^b)^T$, $J(\mathbf{x})$ is well-defined in sub-space S spanned by $\mathbf{X}^b$

- if $\mathbf{w}$ is Gaussian random vector with mean 0 and covariance $(k-1)\,\mathbf{I}$,
  then $\mathbf{x} = \overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w}$ is Gaussian with mean $\overline{\mathbf{x}}^b$ and cov. $(k-1)\,\mathbf{X}^b\,(\mathbf{X}^b)^T$

  → set up cost function in (low-dimensional !) ensemble space

  $$J(\mathbf{w}) \;=\; (k-1)\mathbf{w}^T\mathbf{w} + \left[\mathbf{y}^{\mathbf{O}} - H(\overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})\right]^T \mathbf{R}^{-1} \left[\mathbf{y}^{\mathbf{O}} - H(\overline{\mathbf{x}}^b + \mathbf{X}^b\mathbf{w})\right]$$

  → apply nonlinear H to all forecast members
  and linearize around ensemble mean in observation space $\overline{H\left(\mathbf{x}^{b(i)}\right)}$

  → normal KF eq. in low-dim ensemble space, solve explicitly

  → analysis error cov. $\mathbf{P}_w^{\,a} = \left[(k-1)\,\mathbf{I} + \left(\mathbf{Y}^b\right)^T \mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1}$

  and analysis mean $\overline{\mathbf{w}}^a = \mathbf{P}_w^{\,a}\left(\mathbf{Y}^b\right)^T \mathbf{R}^{-1}\left[\mathbf{y}^{\mathbf{O}} - \overline{\mathbf{y}}^b\right]$

- need analysis ensemble members, with ensemble spread $(k-1)\,\mathbf{X}^a\,(\mathbf{X}^a)^T = \mathbf{P}^a$,
  choose: $\mathbf{X}^a = \mathbf{X}^b\mathbf{W}^a$, where $\mathbf{W}^a = \left[(k-1)\mathbf{P}_w^{\,a}\right]^{1/2}$