Development of a LETKF for km-scale Ensemble Data Assimilation for the COSMO model

Christoph Schraff, Hendrik Reich, **Andreas Rhodin**, Annika Schomburg

Deutscher Wetterdienst, D-63067 Offenbach, Germany

Motivation : Why develop Ensemble-Based Data Assimilation ?

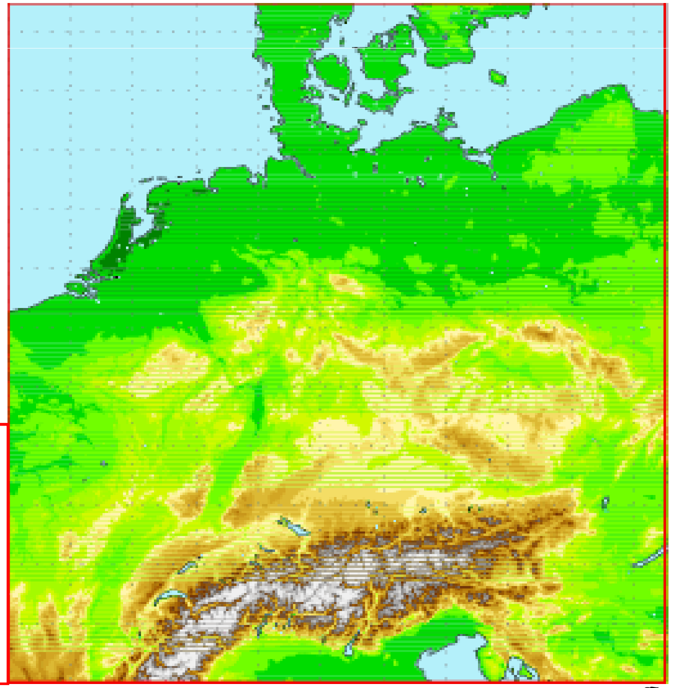
convection-permitting NWP:
after ‘few’ hours,
a forecast of convection is a long-term forecast

→ deliver probabilistic (pdf) rather than deterministic forecast

→ need ensemble forecast and data assimilation system

→ forecast component: COSMO-DE EPS → talk by Christoph Gebhardt

no data assimilation component yet



COSMO-DE: $\Delta x = 2.8$ km
(deep convection explicit,
shallow convection param.)

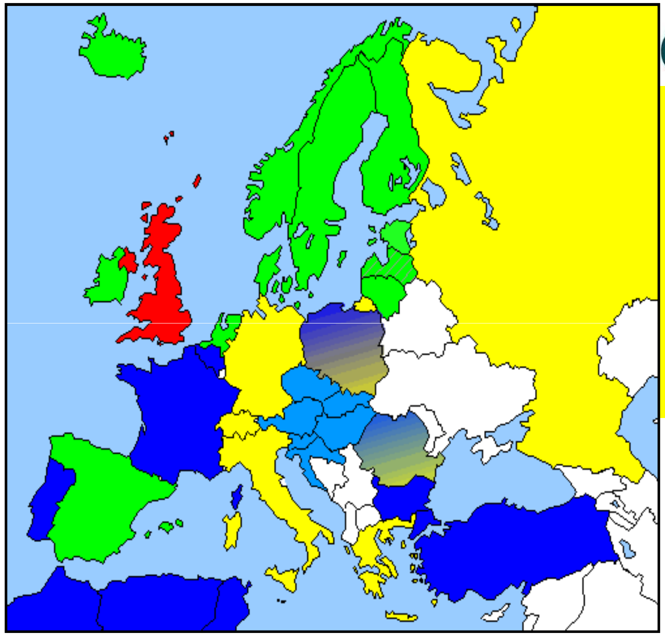
domain size : ~ 1250 x 1150 km

Motivation : Why develop Ensemble-Based Data Assimilation ?

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

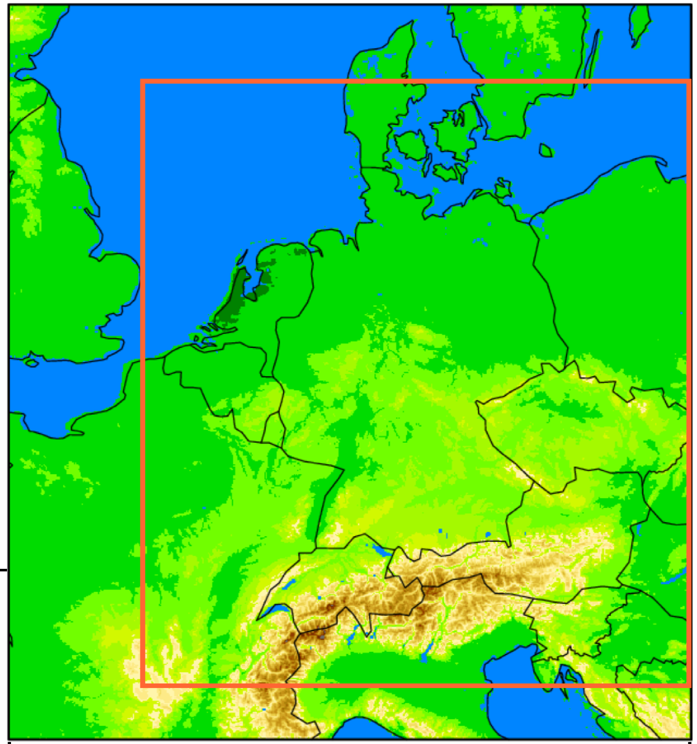
→ data assimilation: priority project within COSMO consortium

Km-scale ENsemble-based Data Assimilation (**KENDA**):



Germany
Greece
Italy
Poland
Romania
Russia
Switzerland

↓
similar configurations
$\Delta x = 1 – 3$ km



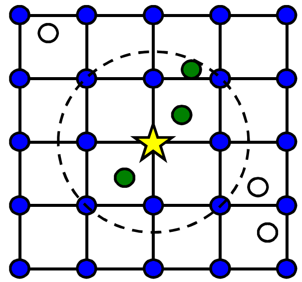
~ 2015 : $\Delta x \cong 2$ km , LETKF

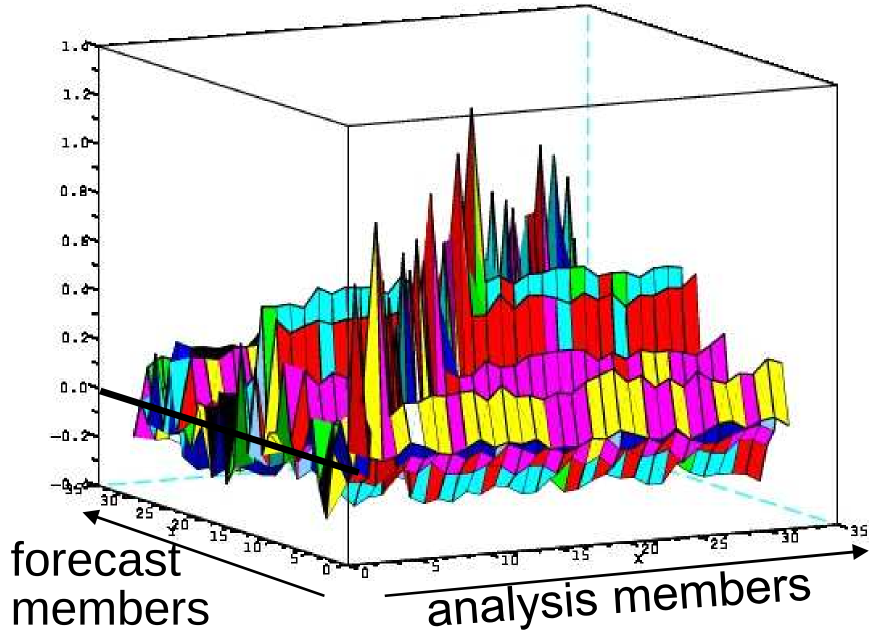
→ Local Ensemble Transform Kalman Filter (LETKF) ,
(because of its relatively low computational costs)

LETKF for km-scale COSMO : method

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- 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 and scale $\mathbf{R}^{-1}$)
  - analysis ensemble members are locally linear combinations of first guess ensemble members

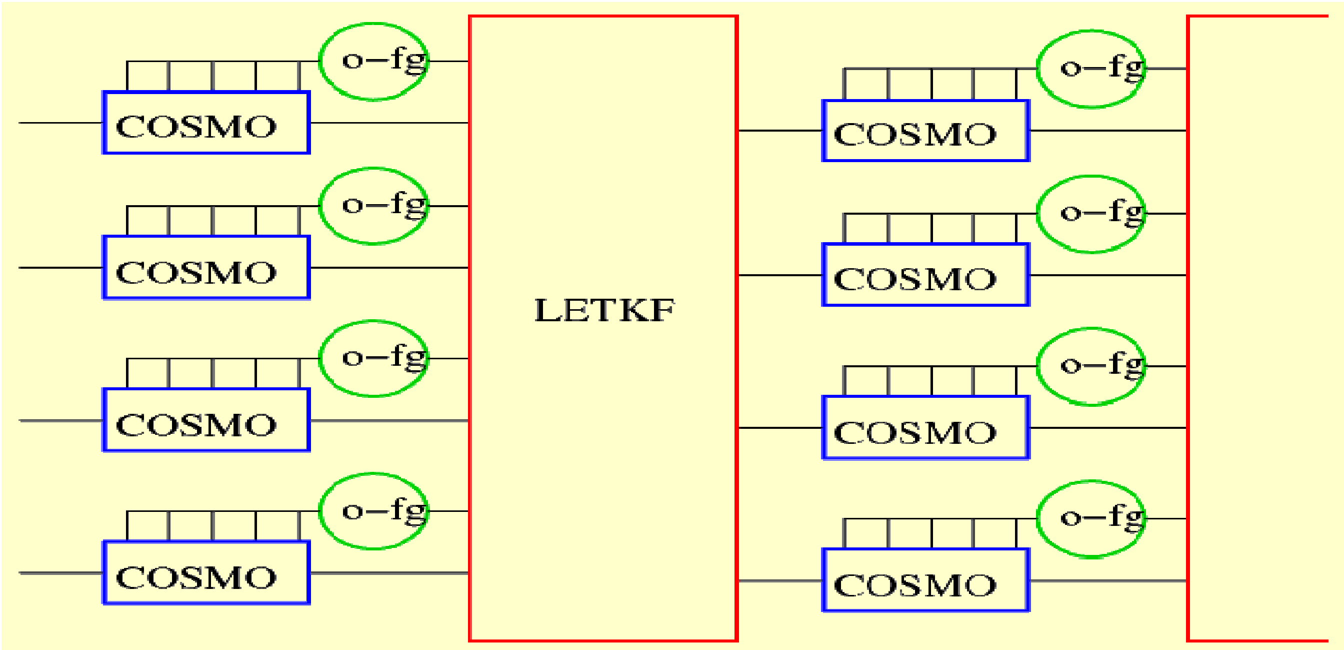




LETKF for km-scale COSMO : implementation

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

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

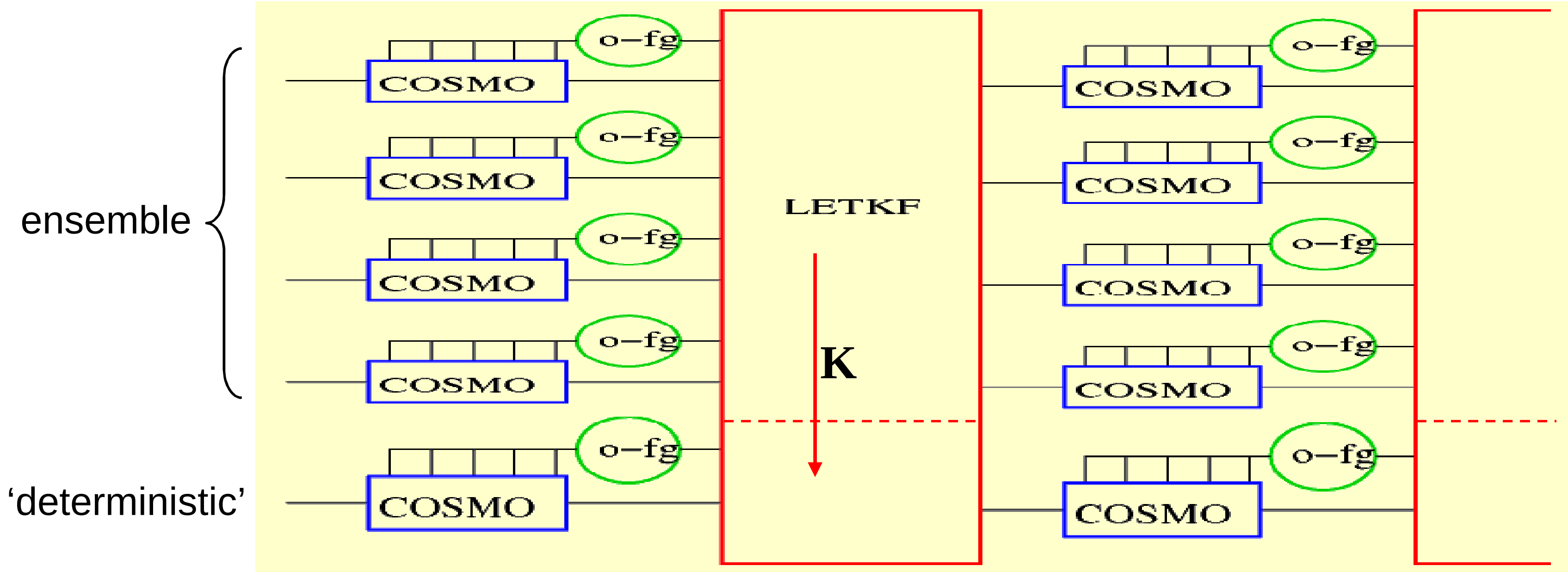


- COSMO obs operators incl. quality control being implemented in 3DVAR / LETKF code
  - → 1st step for hybrid 3DVAR-EnKF approaches in COSMO as potential alternative

LETKF for km-scale COSMO : implementation

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

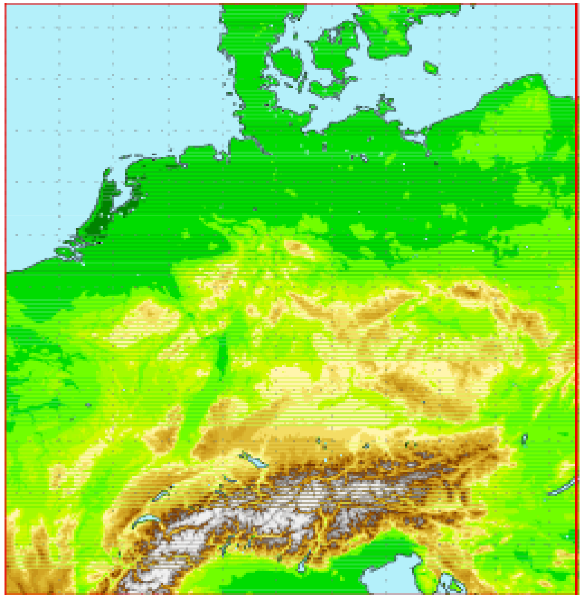
- analysis for a deterministic forecast run: apply Kalman Gain $\mathbf{K}$ of LETKF for analysis mean to obs increments from additional control run



- deterministic run must use same set of observations as the ensemble system !
- Kalman gain / analysis increments not optimal, if deterministic background $\mathbf{x}^B$ (strongly) deviates from ensemble mean background

preliminary experiments: adaptive methods

- only preliminary LETKF experiments with COSMO-DE so far:
  - use **conventional obs** (TEMP, AIREP, SYNOP, Wind Profiler)
  - up to 2 days (7 – 8 Aug. 2009: quiet + convective day)
    $\rightarrow$ **3-hourly** LETKF cycle
  - **32** ensemble members
  - perturbed LBC: COSMO-SREPS, 3 * 4 members

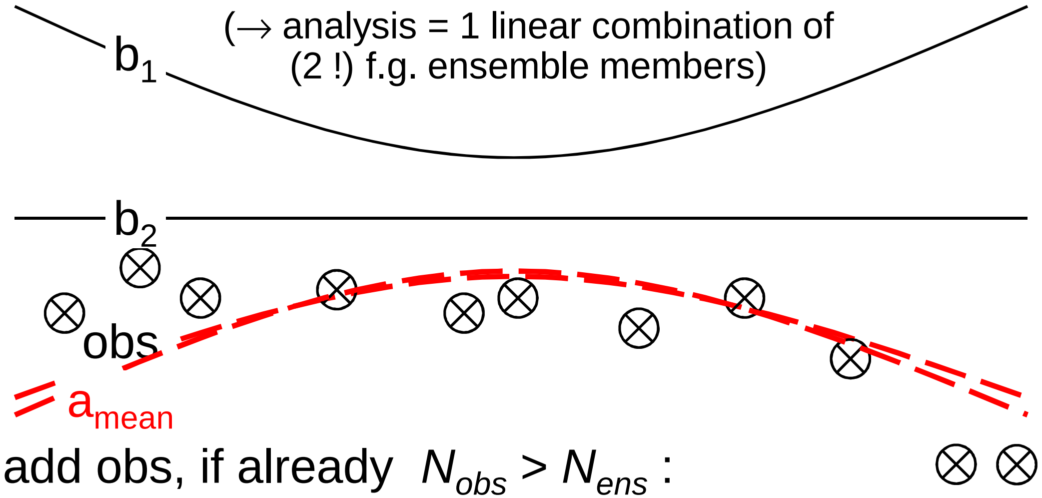


- adaptive methods (important when N_{obs} large) :
  - $\rightarrow$ multiplicative covariance inflation $X_b \rightarrow \rho \cdot X_b$
    different adaptive approaches
    (Desroziers et al., QJRMS 2005; Whitaker et al., 2010)
    positive impact found (not shown)
  - $\rightarrow$ adaptive estimation of obs error covariance $\mathbf{R}$ (Li et al., QJRMS 2009)

preliminary experiments: adaptive methods

- adaptive estimation of obs error covariance $\mathbf{R}$ (Li et al., QJRMS 2009)
  but our implementation: in ensemble space !

within 'localisation scale'
(→ analysis = 1 linear combination of (2 !) f.g. ensemble members)



$N_{obs} > N_{ens} - 1$ (perfect) obs
→ only least square fit

add obs, if already $N_{obs} > N_{ens}$:

- cannot be fitted well, improves analysis only slightly
- but decreases analysis error / spread !
  (filter blind to rank deficiency)
  (non-perfect obs: problem mitigated, but still exists)

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

→ adaptive $\mathbf{R}$ in ensemble space takes that into account and increases $\mathbf{R}$

adaptive localisation: motivation

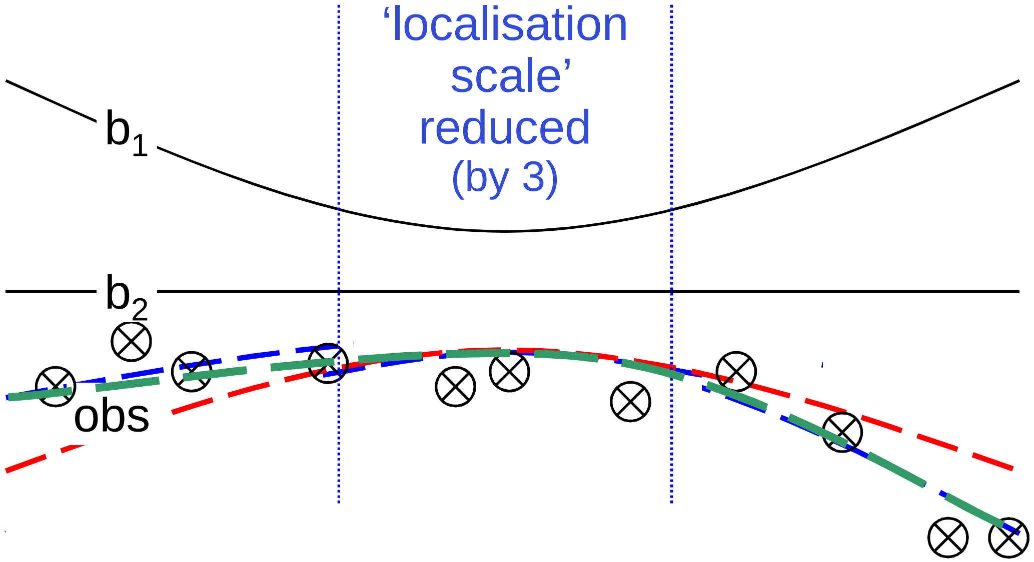
→ positive impact

nevertheless: large N_{obs} : adaptive increase of $\mathbf{R}$ indicates non-optimal use of obs

→ localisation ! → talk by Africa Perianez

(or data selection / superobbing ?)

→ basic idea for adaptive localisation : keep N_{obs} constant (> N_{ens} , not >> N_{ens}) !



Caspari-Cohn function: scale s = 100 km → reduce scale : s = 50 km

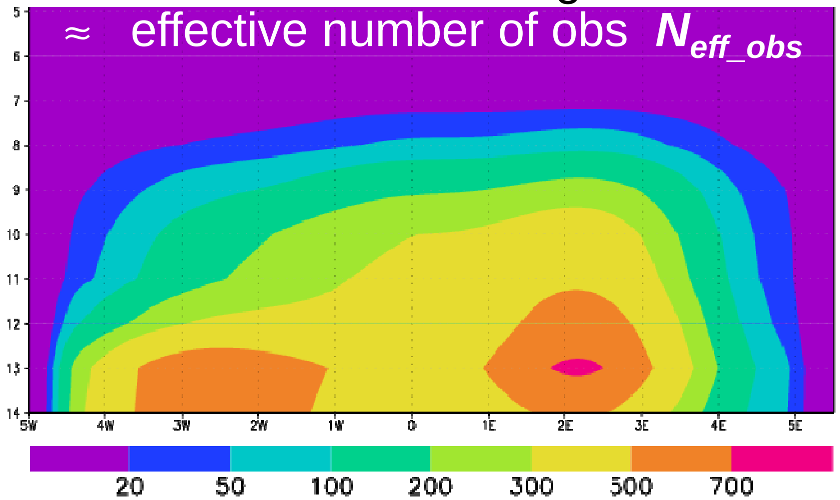
→ 0.4 at $r \cong (2)^{1/2} \cdot s \cong$ 141 km

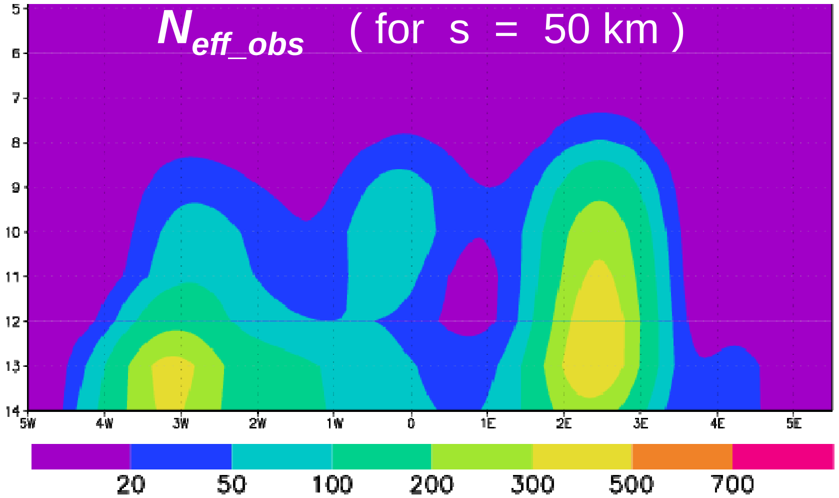
→ 0 at $r = 2 \cdot (10/3)^{1/2} \cdot s \cong$ 365 km

vertical cross section
(at rot lat = 2°, 8 Aug 2009, 12 UTC)

sum of localisation weights of obs

~ 250 hPa
~ 500 hPa
~ 850 hPa



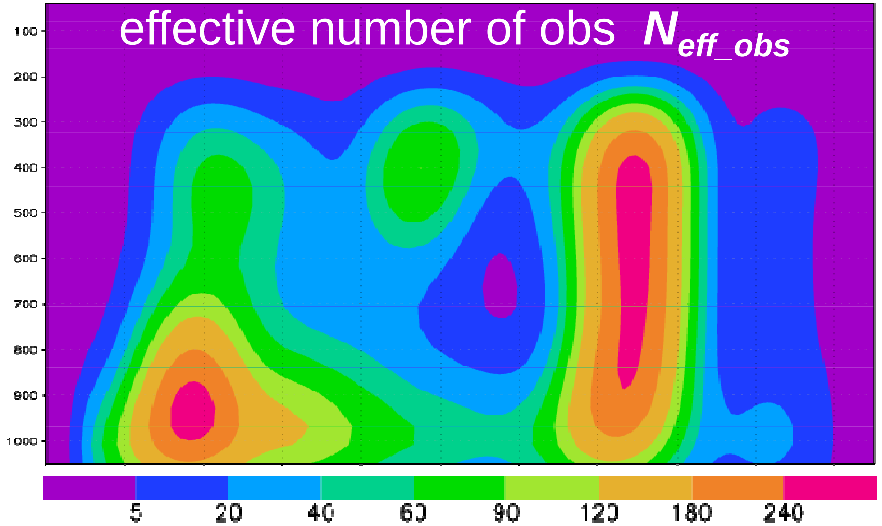


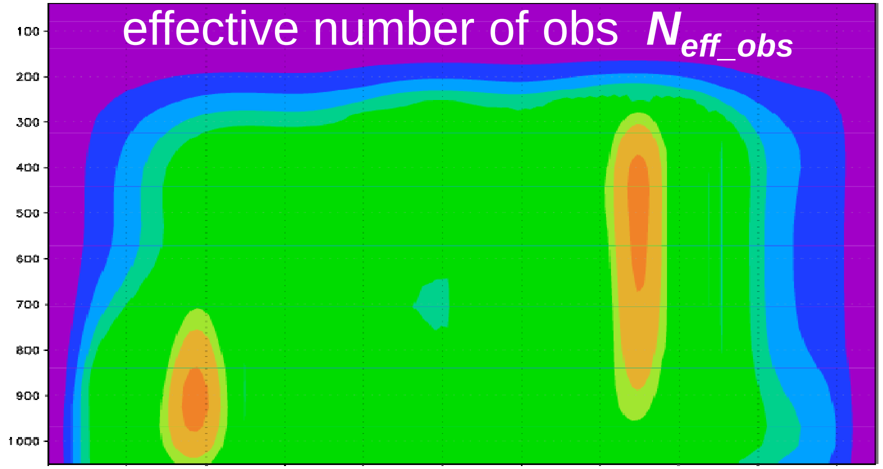
→ N_{eff_obs} >> N_{ens}
→ too few degrees of freedom in order to fit the observations

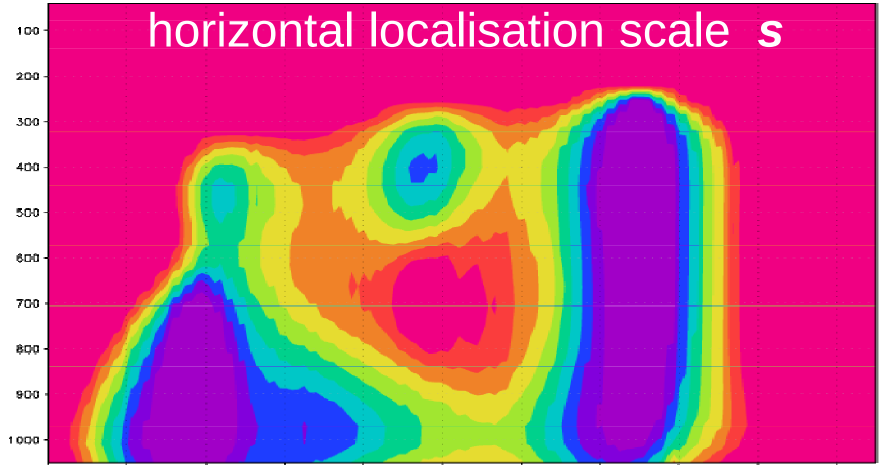
→ in cycled LETKF experiment, increases f.g. spread strongly decreases f.g. rmse slightly (not shown)

Caspari-Cohn function: scale s = 50 km → adaptive scale s :
adapt s such that $N_{eff\_obs} \cong 70$
and 30 km $\leq s \leq$ 80 km

vertical cross section
(at rot lat = 2°, 8 Aug 2009 , 12 UTC)







scale s $\cdot$ $(10/3)^{½}$

adaptive localisation: preliminary results

first guess mean
(inner domain average)

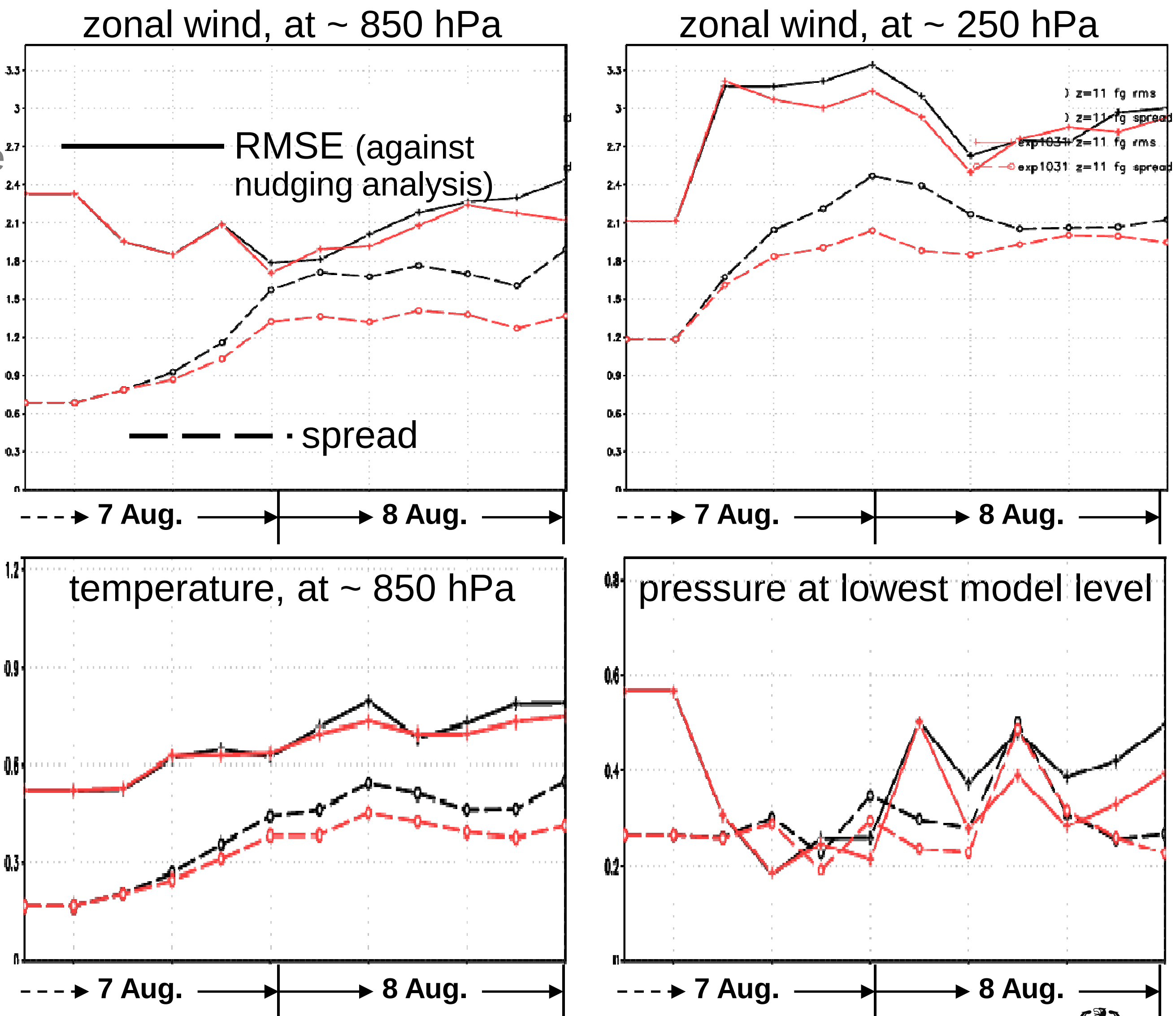
(variable vertical localisation, adaptive **R** and multiplicative covariance inflation ρ)

s = 50 km
adaptive *s*

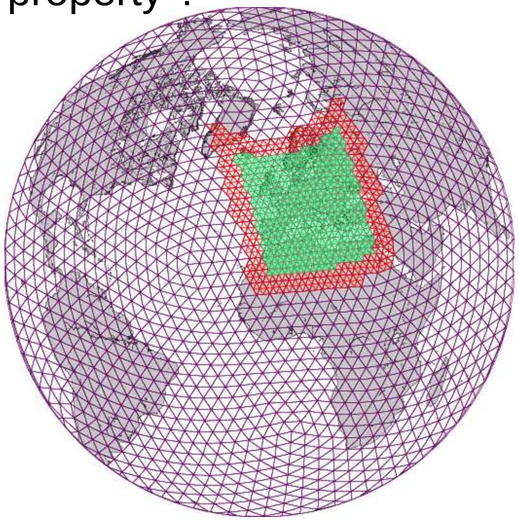
→ smaller spread

→ mostly smaller RMSE

(mixed results in verif vs. upper-air obs (T pos., wind neg.))



- ensemble size N_{ens} = 32 $\rightarrow$ 40
- covariance inflation (adaptive multiplicative, additive: e.g. stochastic physics)
- **localisation** (multi-scale data assimilation,
  successive LETKF steps with different obs / localisation ?
  adaptive , dep. on obs density ?)
- **update frequency** $\Delta_a t$? 3 hr $\rightarrow$ RUC 1 hr $\geq \Delta_a t \geq$ 15 min ! (high-res. obs)
  non-linearity vs. noise / lack of spread / 4D property ?
- perturbed lateral BC : (at DWD) **hybrid VAR-EnKF** / EPS based on ICON (non-hydrostatic global 40 km, regional grid refinement 10 km)



- non-linear aspects, **convection initiation**
  (examine LETKF extensions :
  outer loop like approaches, latent heat nudging ?)

LETKF for km-scale COSMO : use of high-resolution obs : focus on

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

- **ground-based GPS slant path delay**

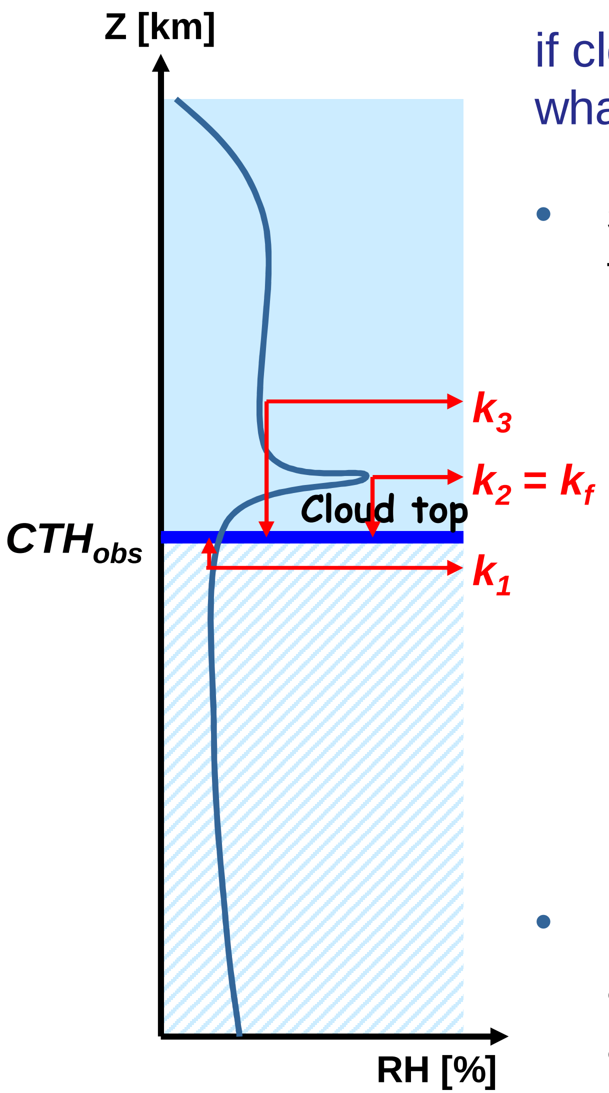
  (direct use (of vertically & horizontally non-local obs) in LETKF , or tomography)

- **cloud information based on satellite and conventional data**
  - use SEVIRI radiances directly, or
  - derive incomplete analysis of cloud top height, based on NWC-SAF cloud products from SEVIRI with correction derived from radiosonde obs

    use cloud top height info in LETKF

  $\rightarrow$ compare / combine approaches

use of cloud top height info in LETKF: method



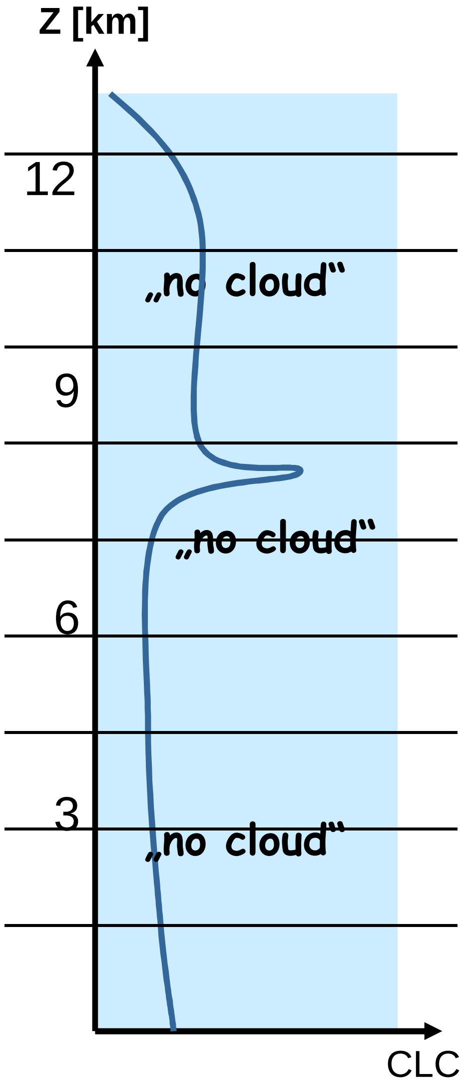
if cloud is observed with cloud top height $\boldsymbol{CTH_{obs}}$, what is the appropriate type of obs increment ?

- search in a vertical range Δh_{max} (≅ 2500 m) around $\boldsymbol{CTH_{obs}}$ for a 'best fitting' model level k_f :
  - if cloud fraction > 70 % in some levels k_{cld} at or above $\boldsymbol{CTH_{obs}}$ then set k_f to uppermost of these levels k_{cld}
  - otherwise find 'best fitting humid level nearby', i.e. with minimum

$$d = \min_k \sqrt{(f(RH_k) - f(RH_{obs}))^2 + \frac{1}{\Delta h_{\max}}(h_k - CTH_{obs})^2}$$

function of relative humidity ($f(RH_k)$); $f(RH_{obs})$ = 1; height of model level k (h_k)

- use $f(RH_{obs}) - f(RH_{k_f}) = 1 - f(RH_{k_f})$
  and $CTH_{obs} - h_{k_f}$
  as 2 separate obs increments in LETKF

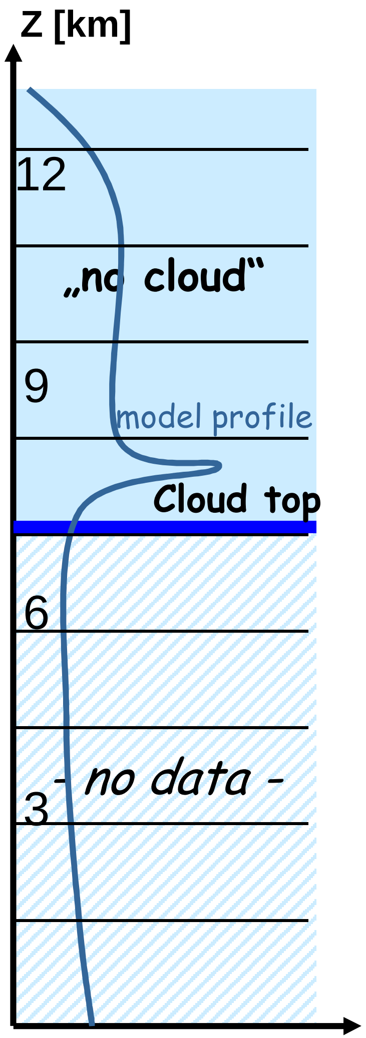


type of obs increment ,
if **no cloud** observed ?

- assimilate cloud fraction = 0
  separately for
  high, medium, low clouds

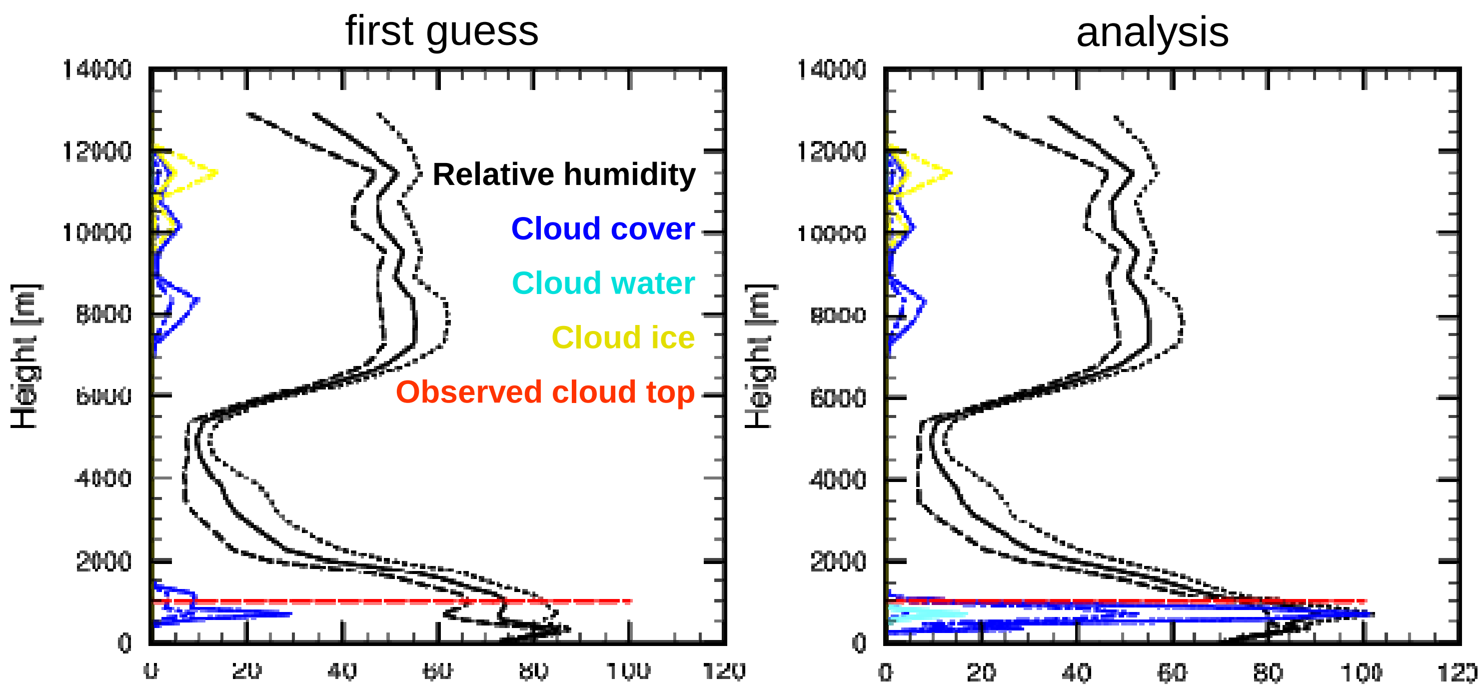
- model equivalent:

  maximum cloud fraction within
  vertical range



- single analysis step, 17 Nov. 2011, 6:00 UTC (low stratus case)
  perturbed LBC: only 4 different members based on 4 global models

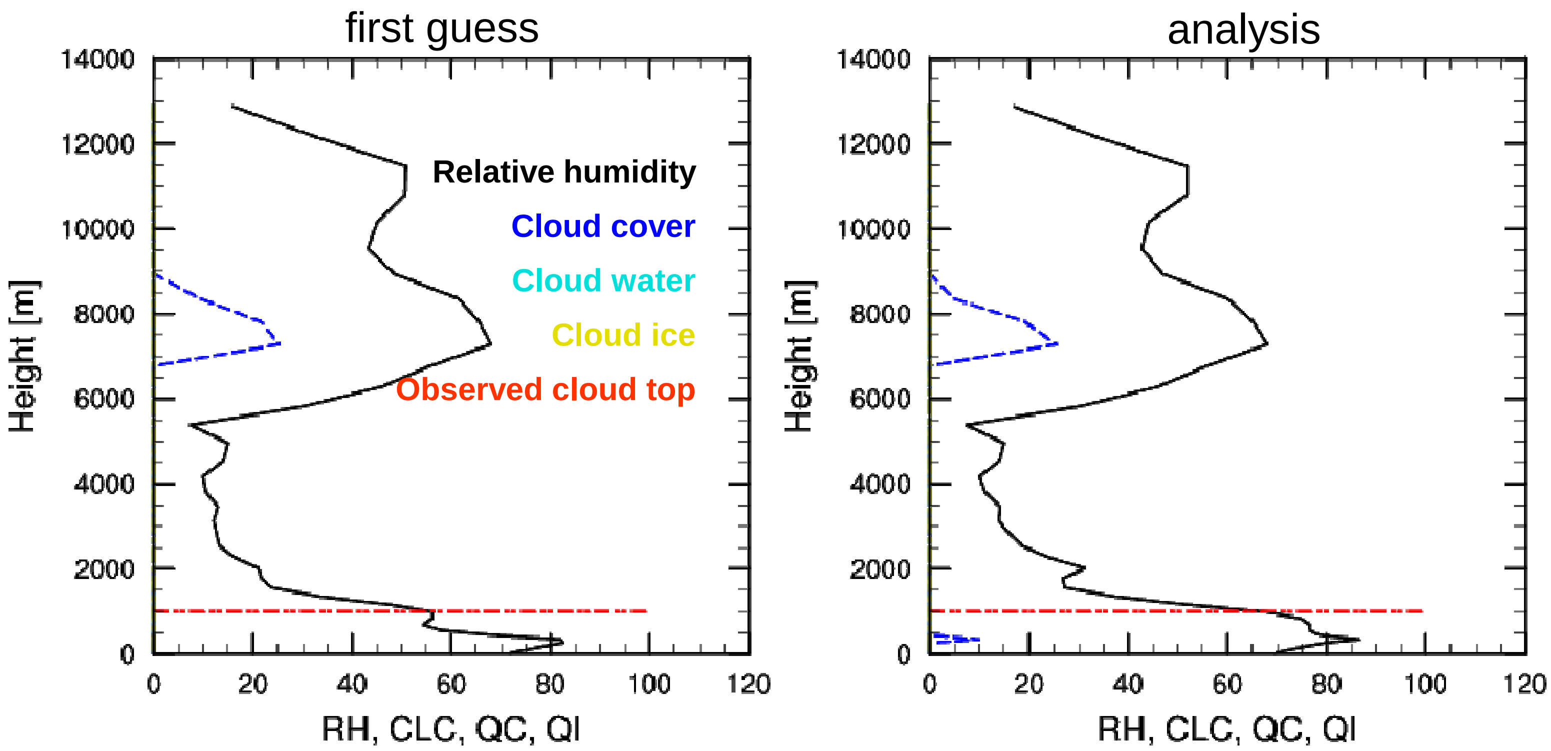
- example 1 : observed low cloud, too little cloud in model → use CTH, RH incr.



for each colour (except red), 3 lines indicate 'ensemble mean' and 'mean ± spread'

use of cloud top height info in LETKF: single observation experiments

- example 1 : deterministic run → humidity at cloud layer is also increased



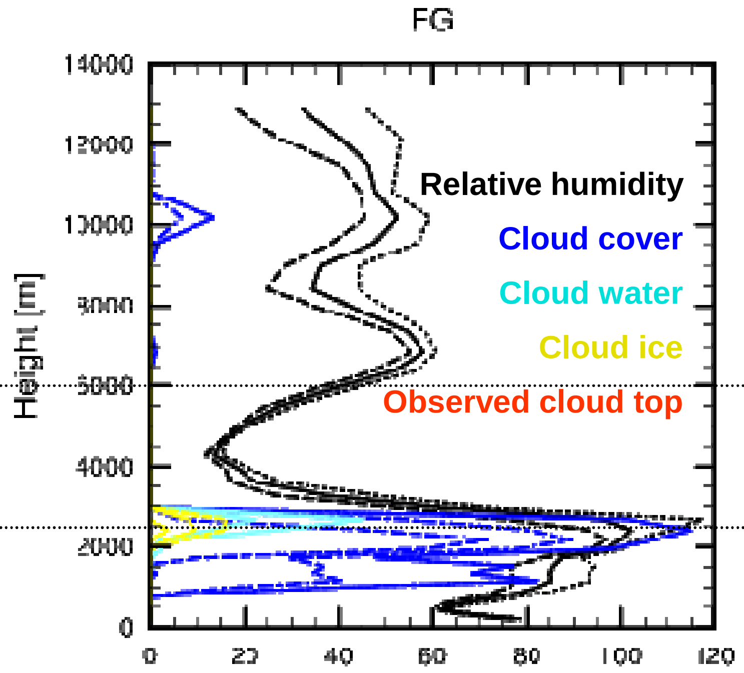
- example 2 : ‘false alarm cloud’ in cloud-free case → use cloud fraction info

cloud levels:

high

mid-level

low

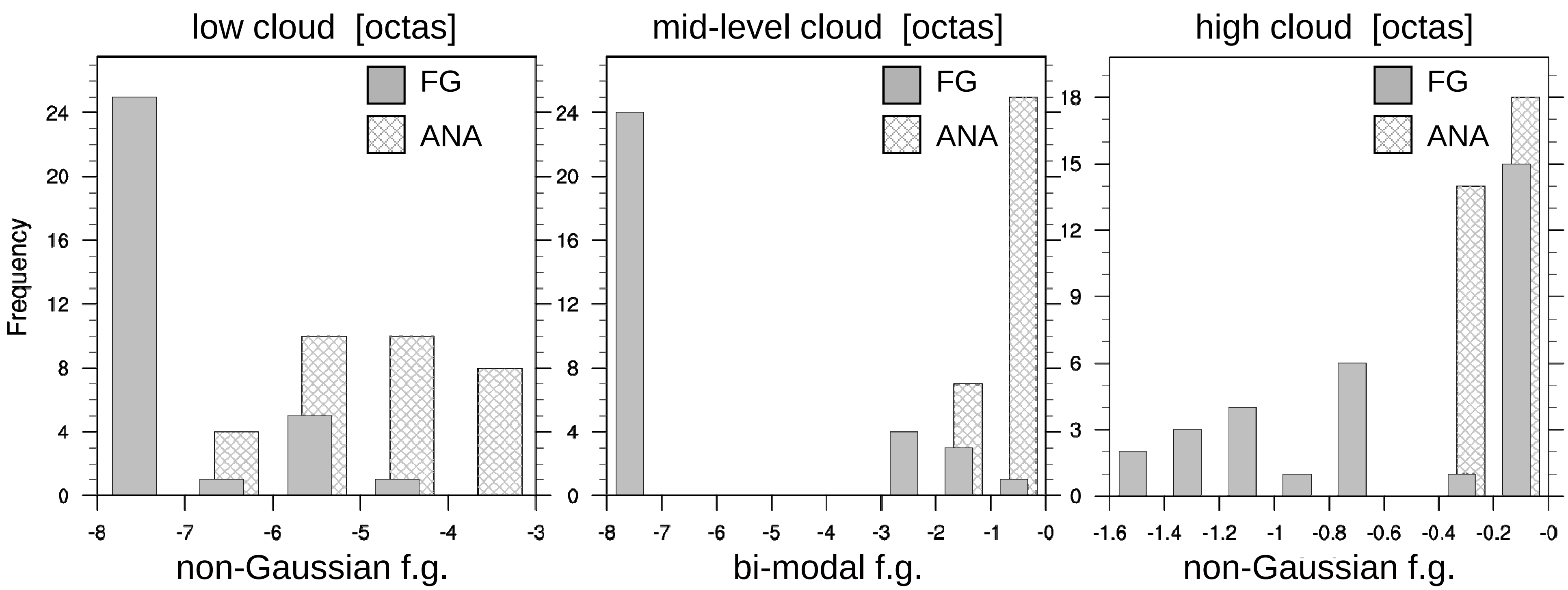


use of cloud top height info in LETKF: single observation experiments

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- example 2 : 'false alarm cloud' in cloud-free case → use cloud fraction info

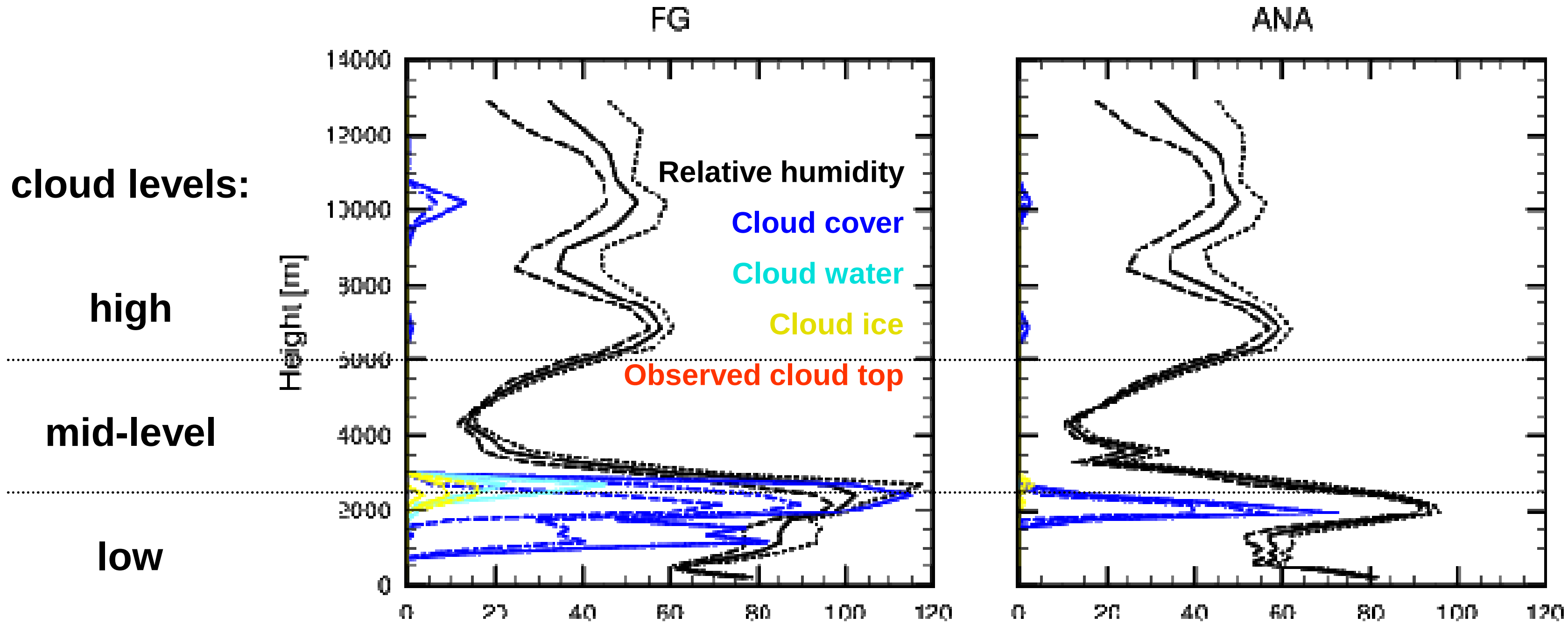
obs minus model deviation histograms (cloud fraction)



use of cloud top height info in LETKF: single observation experiments

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

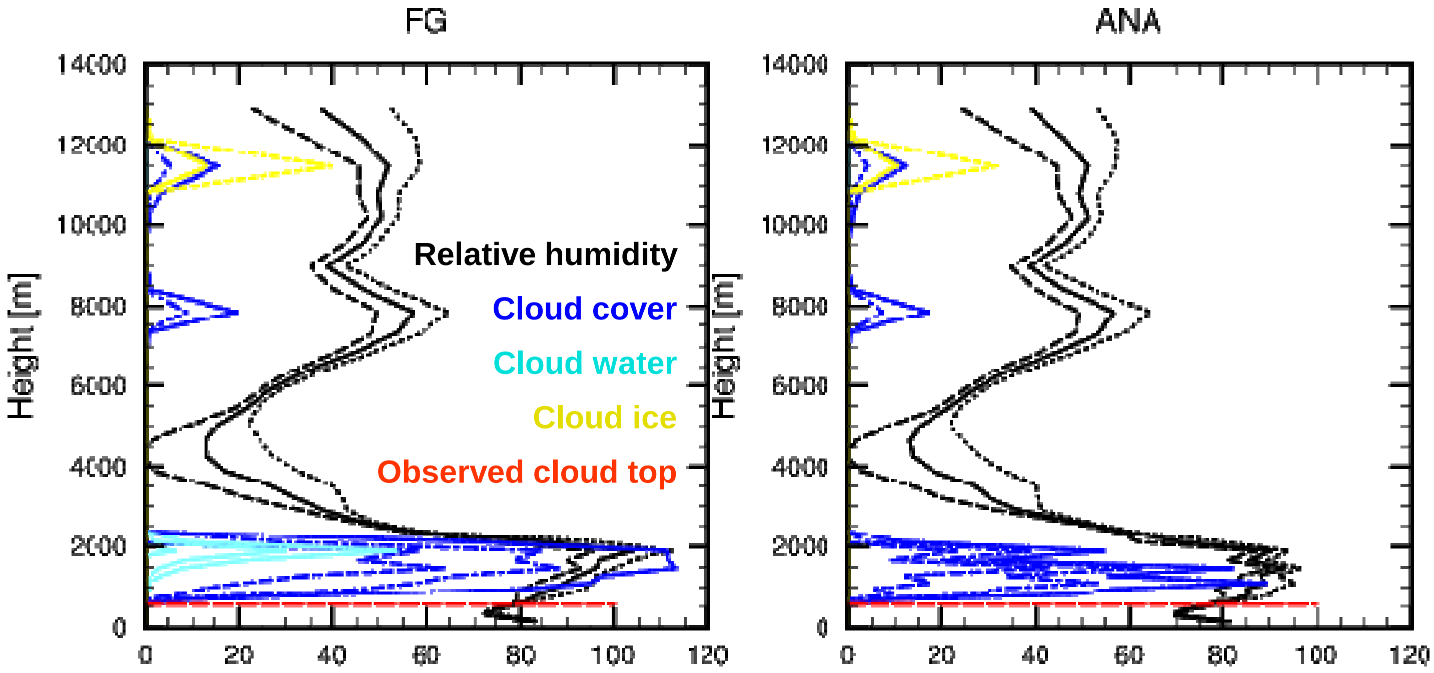
- example 2 : 'false alarm cloud' in cloud-free case → use cloud fraction info



- LETKF reduces erroneous cloud at all levels (low, medium, high), but spread becomes low sometimes
- overall preliminary results rather promising, but many open issues, e.g. specifying $\mathbf{R}$, thinning, maintaining spread in cycling LETKF, ...

thank you for your attention

- example 3 : cloud top height of low cloud is too high $\rightarrow$ use CTH, RH obs incr.



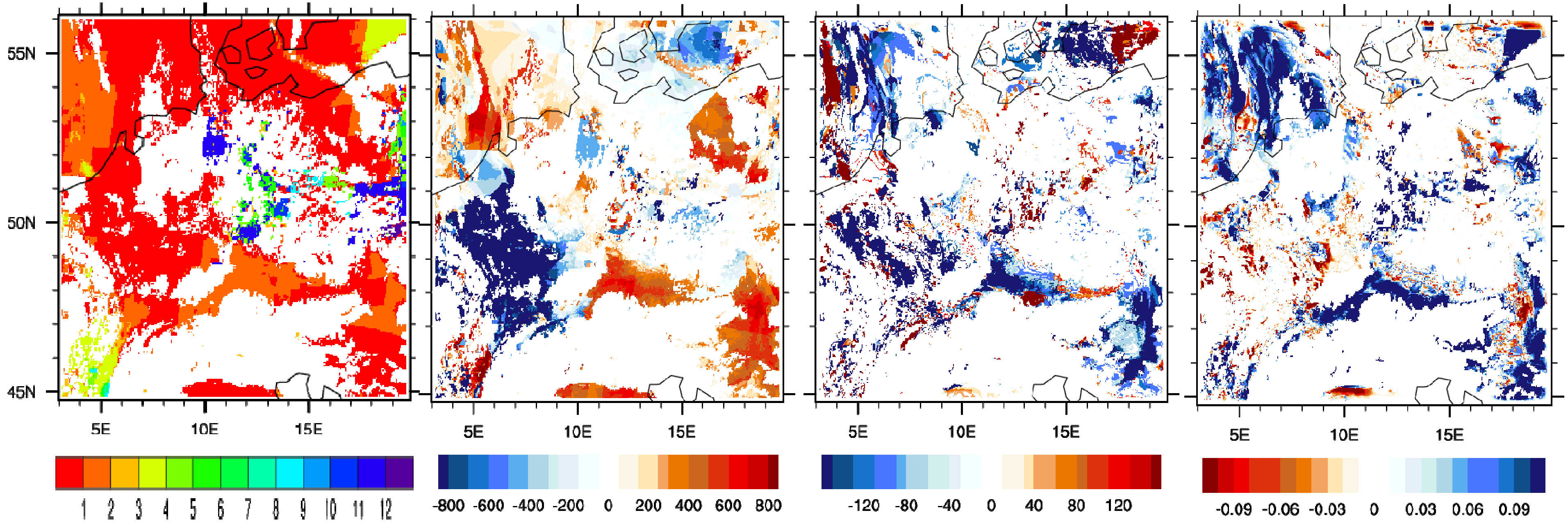
for each colour (except red), 3 lines indicate 'ensemble mean' and 'mean ± spread'

use of cloud top height info in LETKF: 1 LETKF analysis using all cloud obs

- use obs at all grid points (no thinning)

colour shading: grid points with cloud top height observed (i.e. cloud exists)

CTH_{obs} | $CTH_{obs} - H(k)_{FG}$ | $|O-ANA| - |O-FG|$ | $RH_{ana} - RH_{FG}$



blue means: ana better than f.g.

blue means: ana better than f.g.