Status Overview on PP KENDA
Km-scale ENsemble-based Data Assimilation

Lucio Torrisi - CNMCA-Italian Met. Centre
on behalf of Christoph Schraff – DWD (P.L.)

Contributions / input by:

Hendrik Reich, Andreas Rhodin, Roland Potthast, Uli Blahak, Klaus Stephan, Africa Perianez, Michael Bender (DWD)

Annika Schomburg (DWD / Eumetsat)

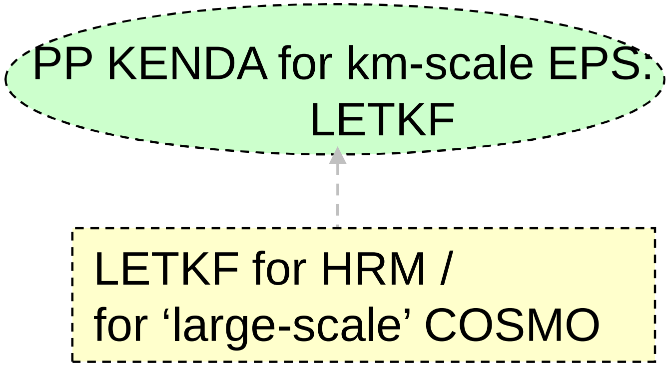
Yuefei Zeng, Dorit Epperlein (KIT Karlsruhe)

Daniel Leuenberger (MeteoSwiss)

Mikhail Tsyrulnikov, Vadim Gorin, Igor Mamay (HMC)

Lucio Torrisi (CNMCA)

Amalia Iriza (NMA)



PP KENDA for km-scale EPS:
LETKF

LETKF for HRM /
for ‘large-scale’ COSMO

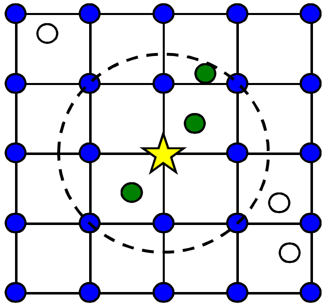
- Task 1: General issues in the convective scale (e.g. non-Gaussianity)
- Task 2: Technical implementation of an ensemble DA framework / LETKF
- Task 3: Tackling major scientific issues, tuning, comparison with nudging
- Task 4: Inclusion of additional observations in LETKF

Status Overview on PP KENDA
Km-scale ENsemble-based Data Assimilation

- Adaptive horizontal localization, preliminary results
- Stochastic physics implementation in COSMO model
- Use of cloud info: NWCSAF cloud products (SEVIRI/MSG)
- Work on radar observations and GNSS slant path delay

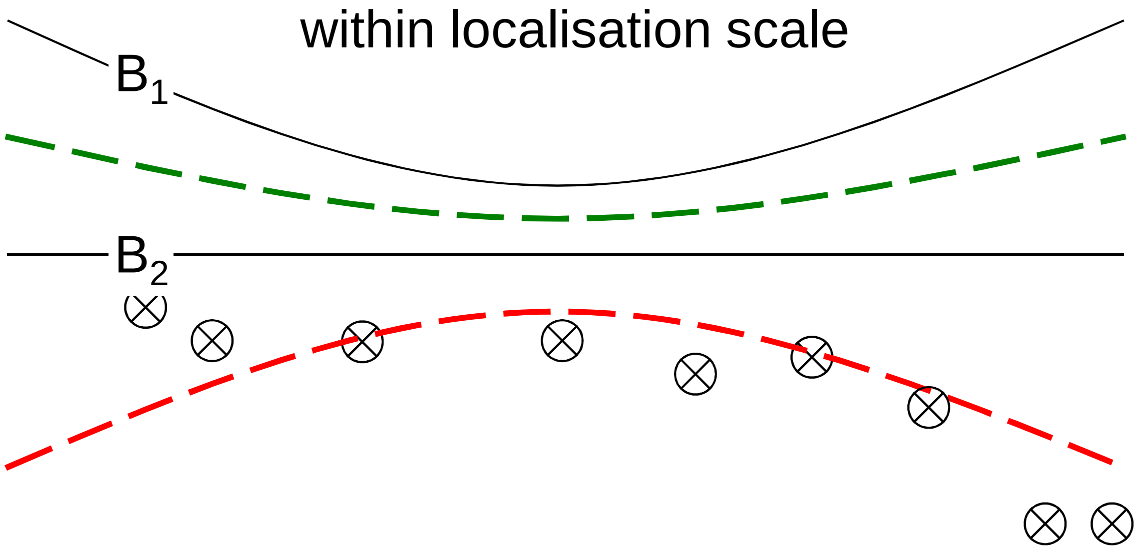
LETKF (COSMO) : method

- COSMO priority project KENDA (Km-scale ENsemble-based Data Assimilation)
- 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



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

however: large N_{obs} : adaptive increase of R indicates non-optimal use of obs



f.g. mean = $0.5\ B_1 + 0.5\ B_2$

$N_{obs} > N_{ens}$ → least square fit

→ localisation !

(or data selection / superobbing ?)

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

LETKF, preliminary results: horizontal localisation

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

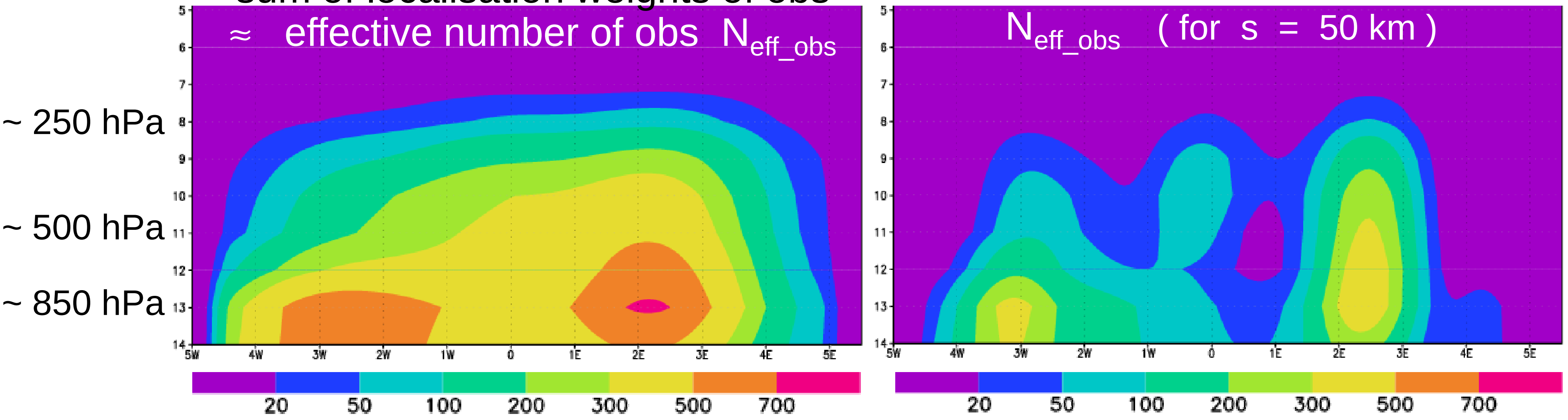
→ 0.4 at $r \cong (2)^{1/2} \cdot s \cong 141$ km

→ 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



≈ effective number of obs N_{eff_obs}

N_{eff_obs} (for $s = 50$ km)

→ $N_{eff\_obs} >> N_{ens}$

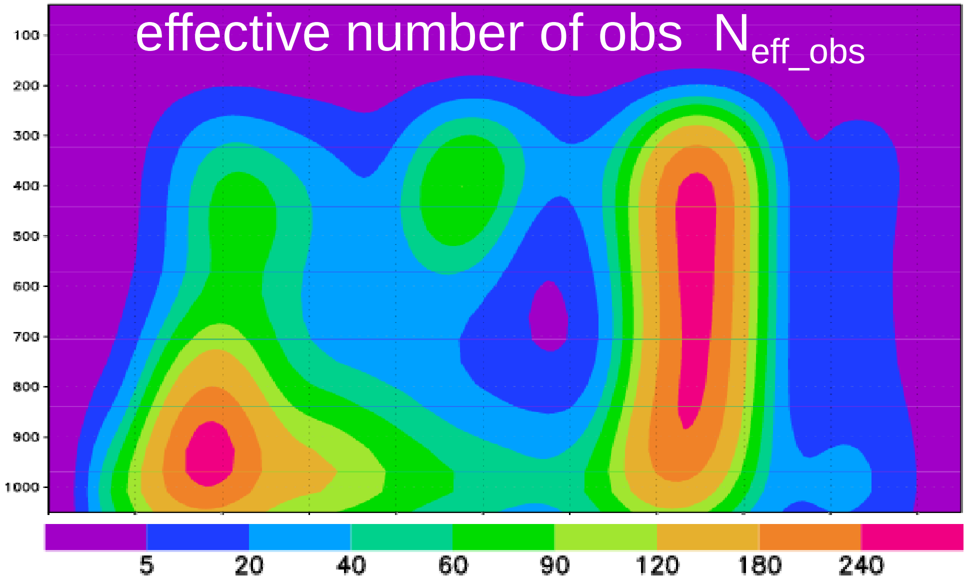
→ too few degrees of freedom in order to fit the observations

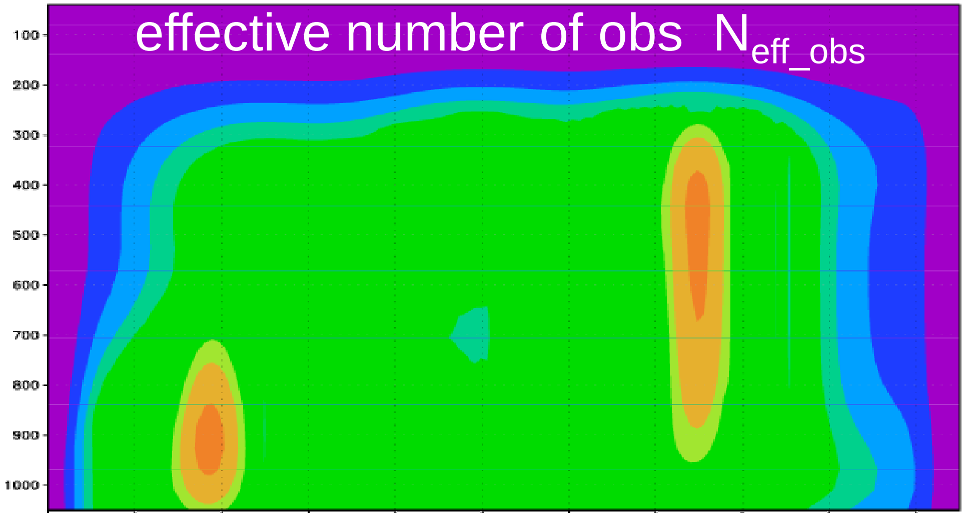
LETKF, preliminary results:
adaptive horizontal localisation

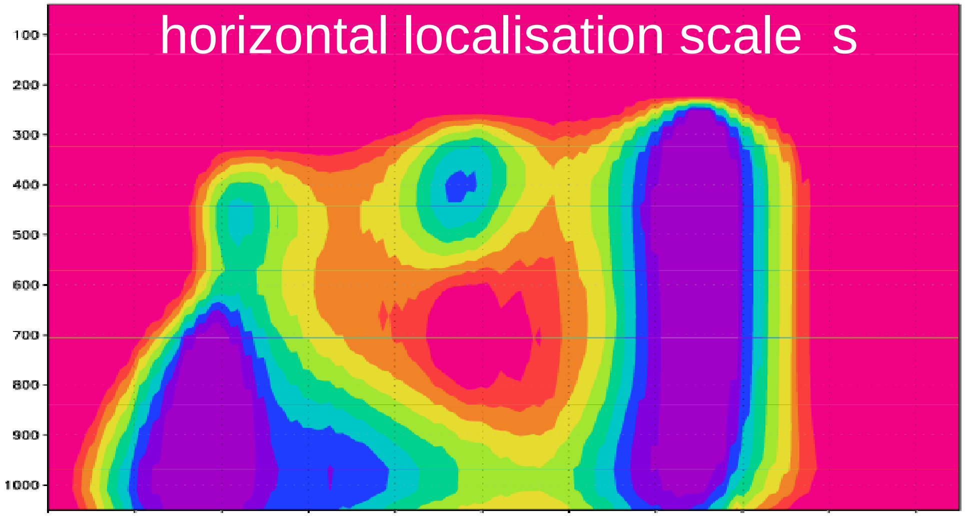
Caspari-Cohn function: scale s = 50 km → adaptive scale s (32 members):

adapt s such that $N_{eff\_obs} \cong 70$
and 30 km ≤ s ≤ 80 km

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


effective number of obs N_{eff_obs}


effective number of obs N_{eff_obs}

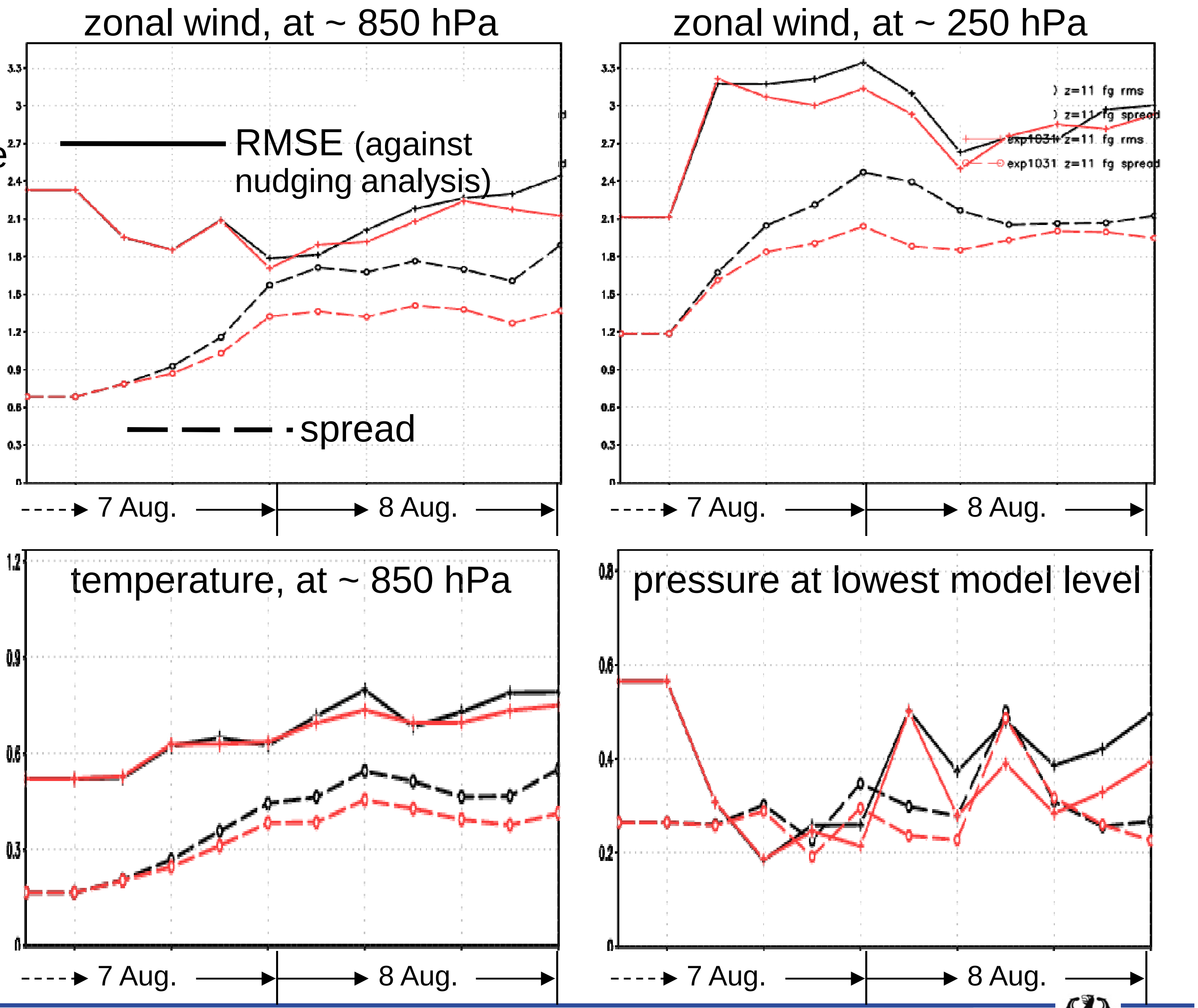

horizontal localisation scale s

scale s ·· $(10/3)^{½}$

LETKF, preliminary results: adaptive horizontal localisation

first guess mean
(inner domain average)
(variable vertical localisation, adaptive R and multiplicative covariance inflation ρ)

s = 50 km
adaptive s


zonal wind, at ~ 850 hPa
zonal wind, at ~ 250 hPa
temperature, at ~ 850 hPa
pressure at lowest model level

→ smaller spread

→ mostly smaller RMSE

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

LETKF:
account of model error / additive inflation

→ new task for a pattern generator (PG)

purely stochastic tool to generate 4-D pseudo-random fields with selectable scales / ampl.,
used to generate additive perturbations / for stochastic physics
(~ 0.4 FTE / y)

- stochastic physics: perturbing total physics tendency by a random factor
  at any given grid point (Palmer et al., 2009) (Torrisi)
  - modified Buizza version running
    → tuning required
  - perturb all physics tendencies in same way ?

→ 2013 Ekaterina Machulskaya from SFP for (more physically based) stochastic physics !
+ 1 N.N. (renewable energy project)

- additional additive inflation: - by scaled forecast differences (e.g. Bonavita et al.) ?
  - 3DVAR – B ?

LETKF:
account of model error / additive inflation

Stochastic Physics

$X_p=(1+r\ \mu)X_c$

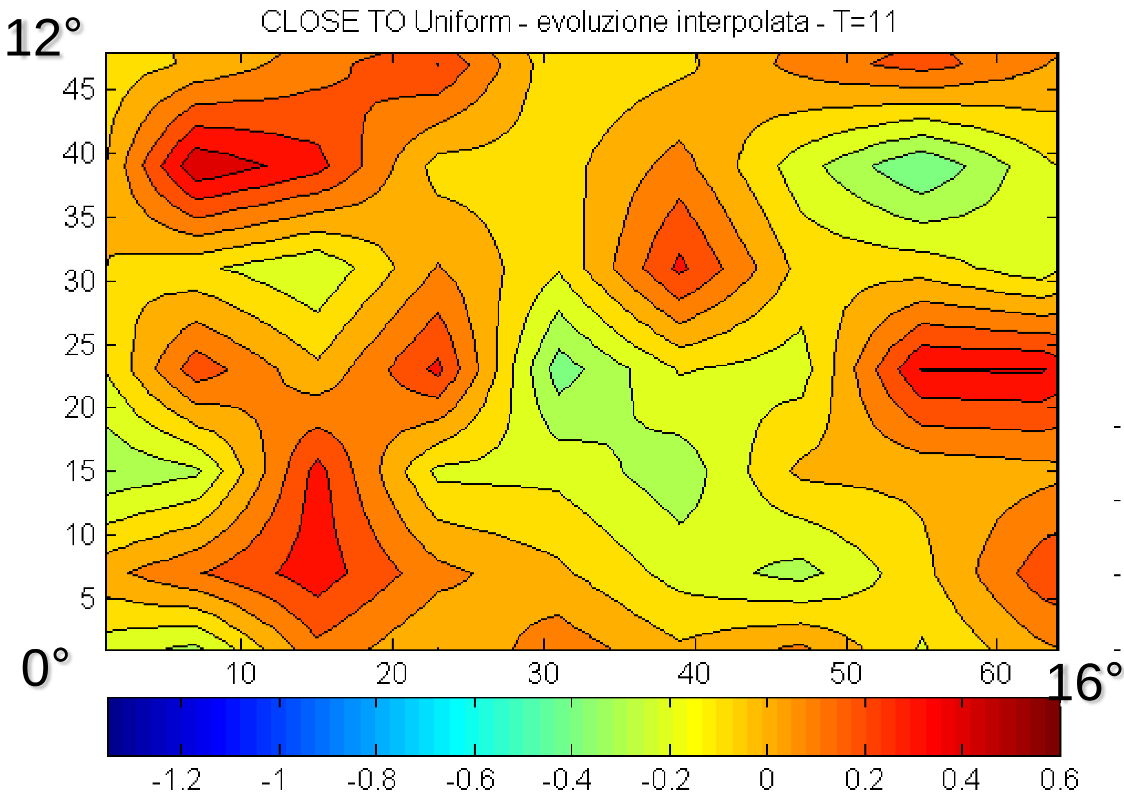
Modified Version (in blue, differences from Buizza et al, 1999)

- For all variables (u,v,T,qv), the random numbers r are drawn from a uniform distribution in a certain range [-0.5,0.5] or a gaussian distribution with stdv (0.1-0.5) bounded to a certain value (range= ± 2-3 stdv) on a coarse horizontal grid every n time-steps
- A tapering factor μ is used to reduce r close to the surface and in the stratosphere (Palmer et al, 2009)
- The perturbations of T and qv are not applied if they lead to particular humidity values (exceeding the saturation value or negative values)
- Spatial correlation is imposed using the same r in a whole column and drawing r for a coarse grid with spacing DL (boxes); then they are bilinearly interpolated on the finer grid to have a smooth pattern in space
- Temporal correlation is achieved by drawing r every n time steps (Dt); then they are linearly interpolated for the intermediate steps to have a smooth pattern in time
- Coarse grid SW corner is different for each member

LETKF:
account of model error / additive inflation

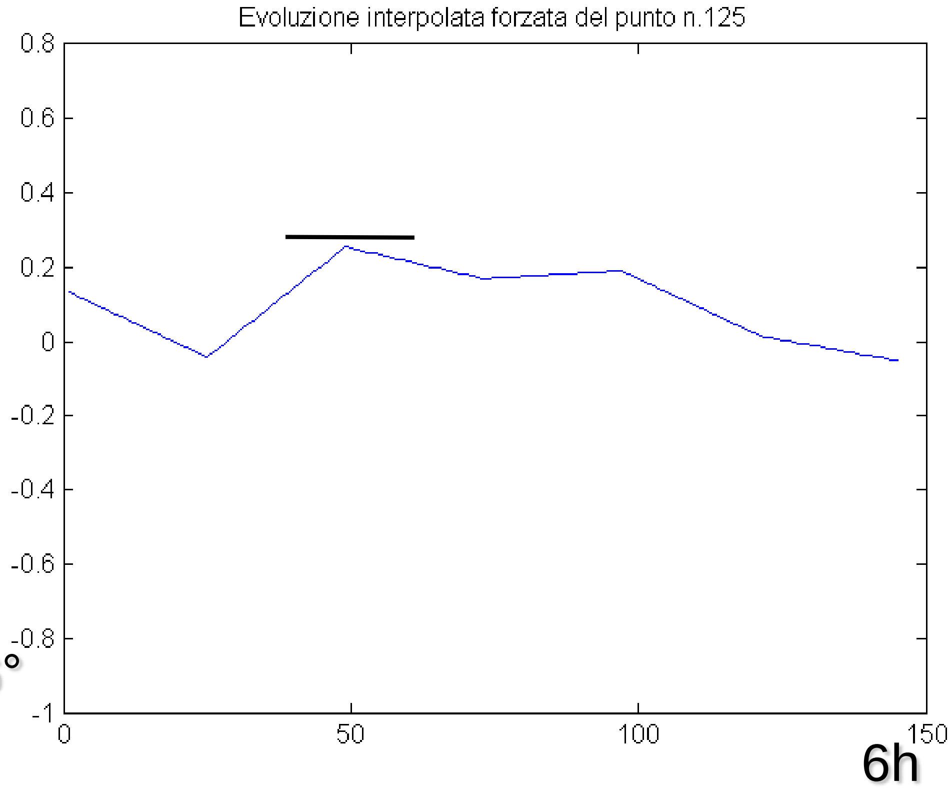
Modified Version of Stochastic Physics

Spatial pattern



Bilin. Interpolation from coarse grid

Temporal pattern



Lin. Interpolation from coarse grid

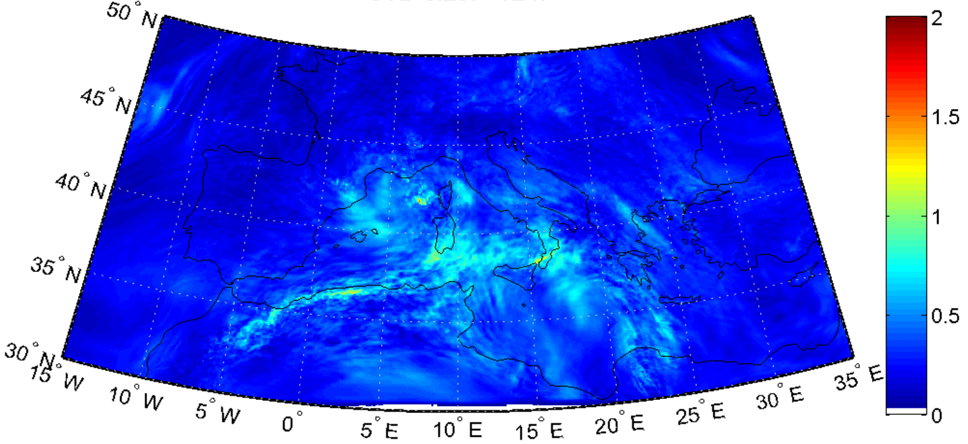
LETKF:
account of model error / additive inflation

500 hPa Temperature Spread for 10 members

stdv=0.25 range=0.75

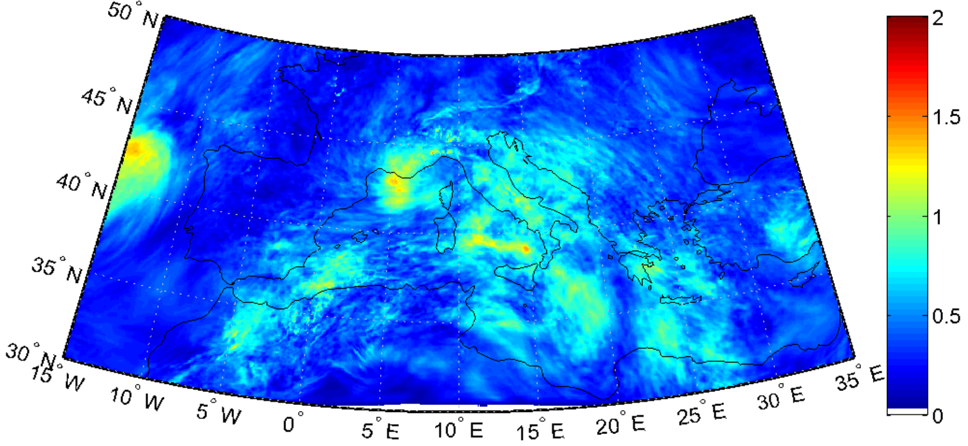
T+12h

STD 0.25: +12 h



T+24h

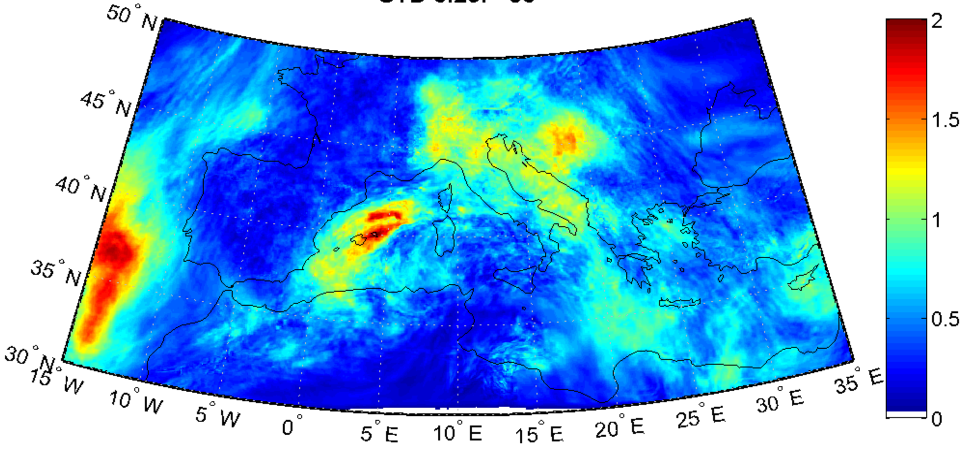
STD 0.25: +24 h



05 June 2011 00UTC

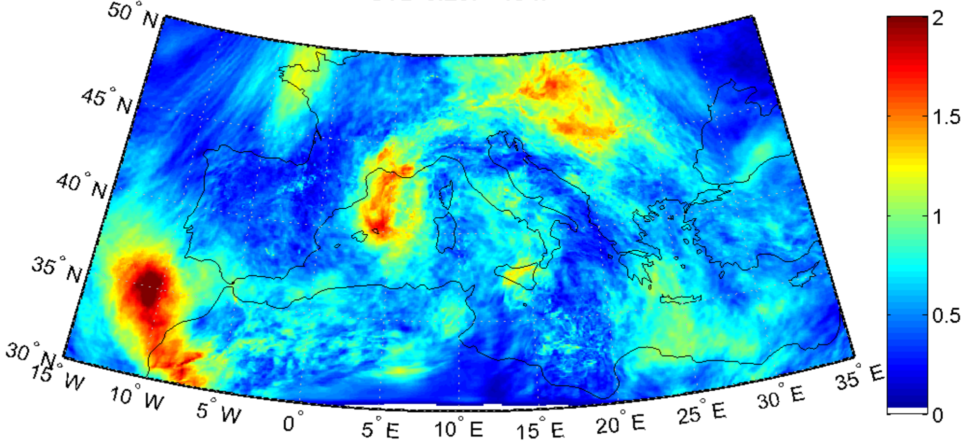
T+36h

STD 0.25: +36



T+48h

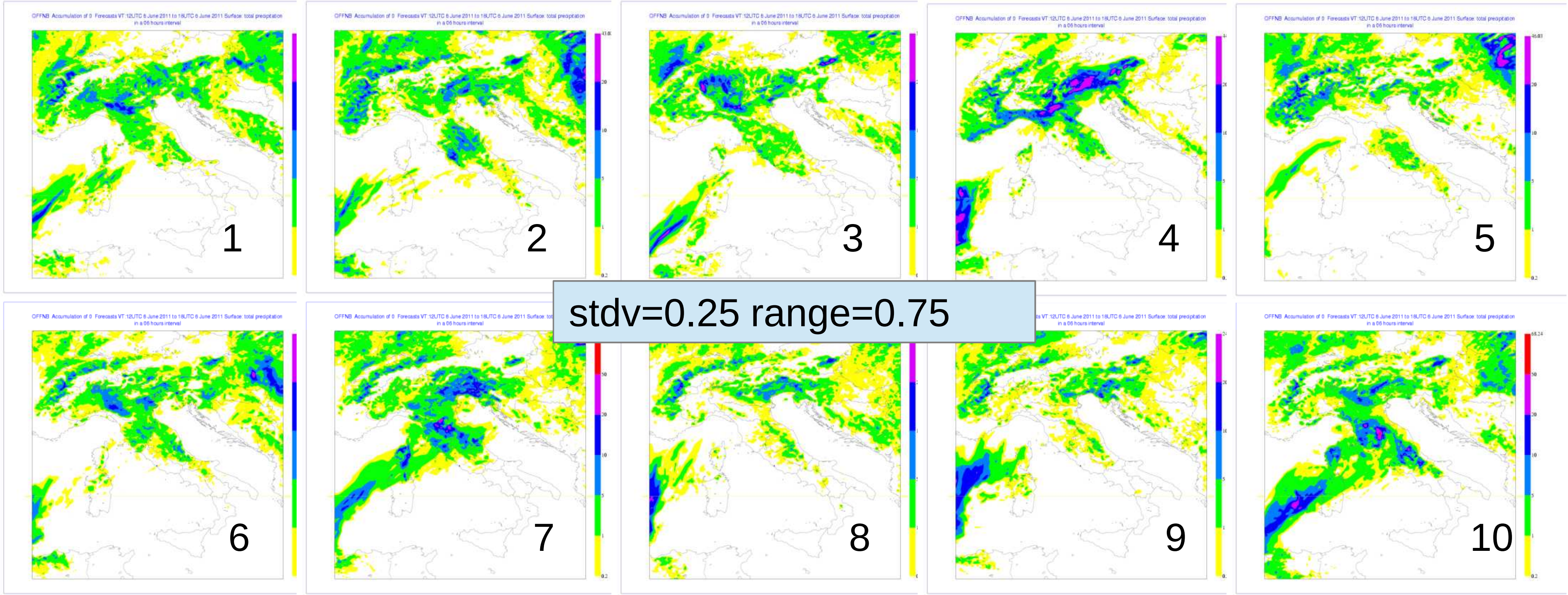
STD 0.25: +48 h



LETKF:
account of model error / additive inflation

05 June 2011 00UTC

6h Accumulated Precipitation (T+36 - T+42h)



stdv=0.25 range=0.75

Task 4.4: use of cloud info

- cloud information based on satellite and conventional data

1. derive incomplete analysis of cloud top + base height, using conventional obs (synop, radiosonde, ceilometer) and NWC-SAF cloud products from SEVIRI

   use cloud top height info in LETKF

1. (Annika Schomburg , DWD / Eumetsat)

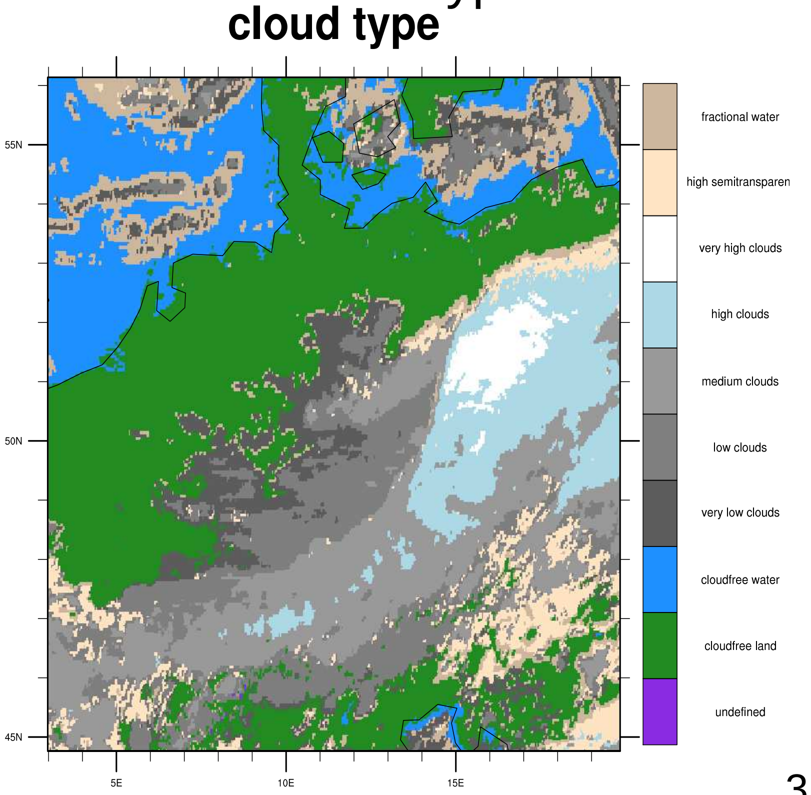
2. use SEVIRI brightness temperature directly in LETKF in cloudy (+ cloud-free) conditions, in view of improving the horizontal distribution of cloud and the height of its top (2013: Africa Perianez, Annika Schomburg)

→ compare approaches

Particular issues: non-linear observation operators,
non-Gaussian distribution of observation increments

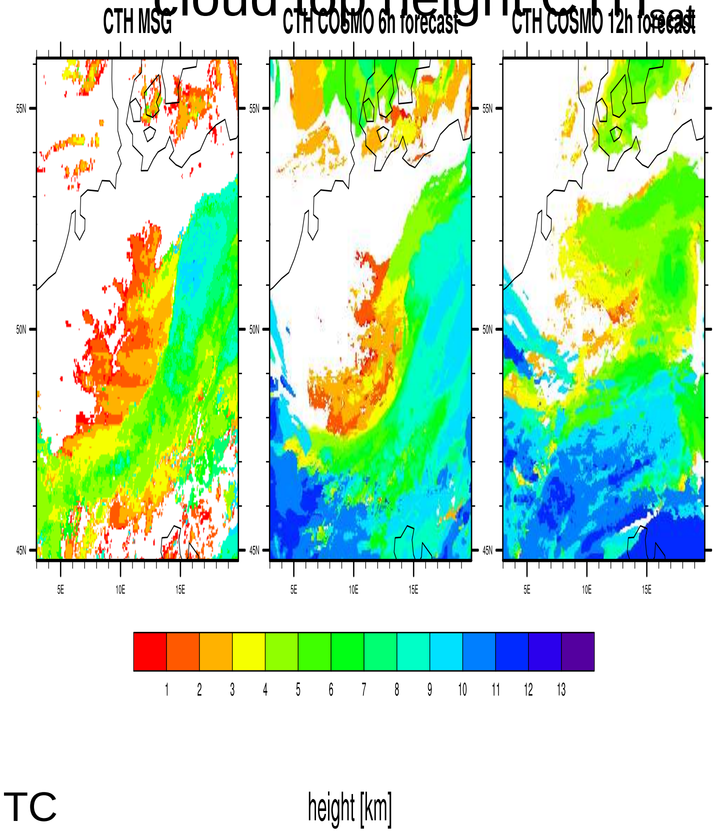
use of cloud info :
NWCSAF cloud products (SEVIRI/MSG)

cloud type



fractional
high semi-transparent
very high
high clouds
medium clouds
low clouds
very low clouds
cloudfree
cloudfree
undefined

cloud top height CTH_{sat}



3 May 2011, 6:00 UTC

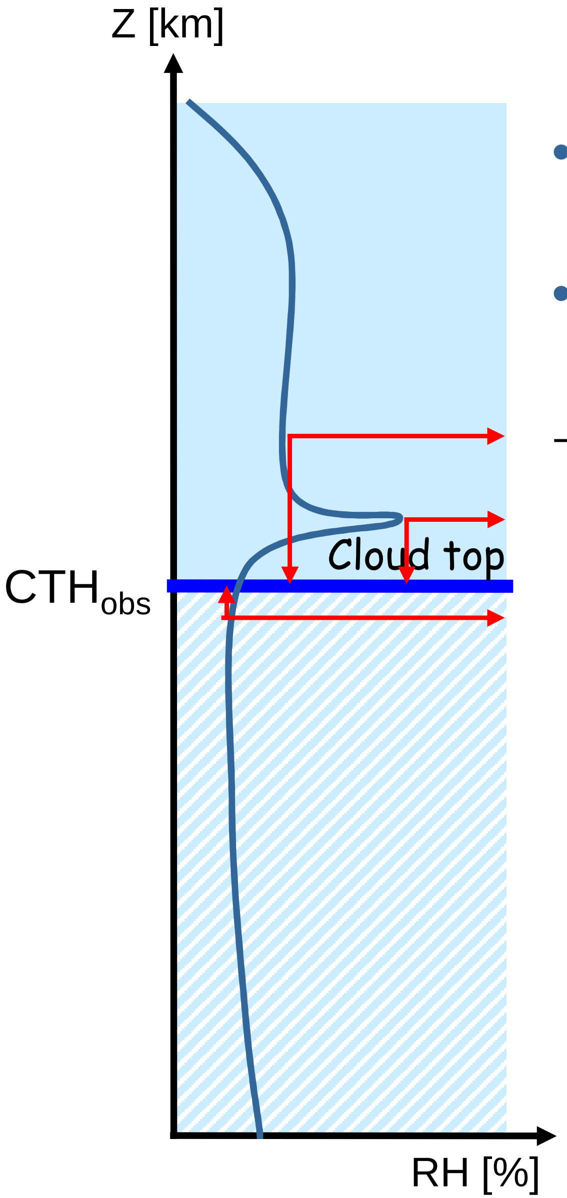
retrieval algorithm needs temperature and humidity profile from a NWP model

→ cloud top height CTH_{sat} wrong if temperature profile in NWP model wrong !

→ combine good horizontal resolution of satellite info
with good vertical resolution of radiosonde info :

use nearby radiosondes with same cloud type to correct CTH_{sat}

use of cloud info : assimilation of ‘cloud analysis’



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

- avoid too strong penalizing of members with high humidity but no cloud
- avoid strong penalizing of members which are dry at CTH_{obs} but have a cloud or even only high humidity close to CTH_{obs}

→ search in a vertical range Δh_{max} around CTH_{obs} for a ‘best fitting’ model level k, i.e. with minimum ‘distance’ d:

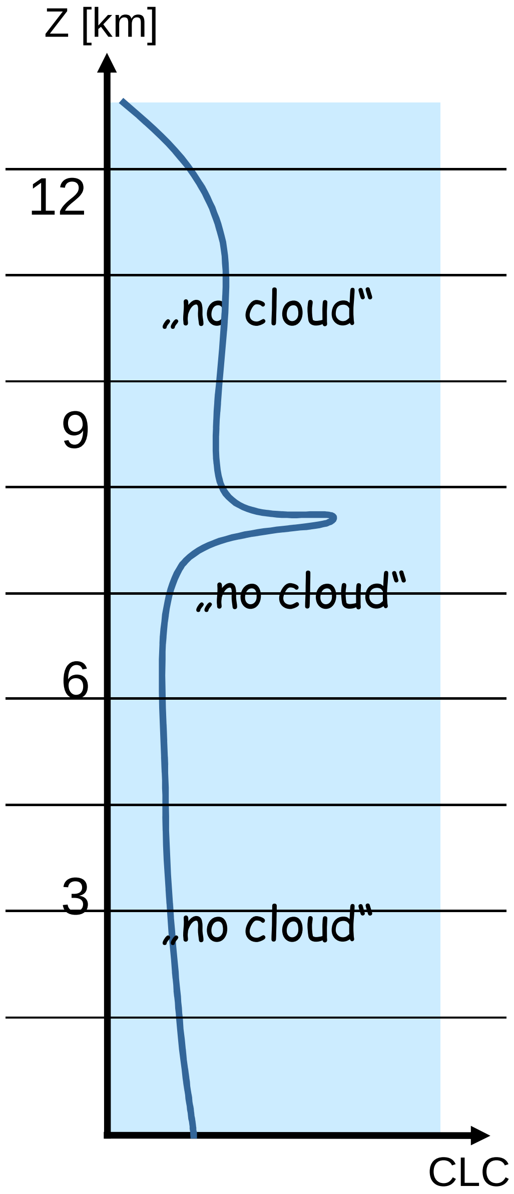
$$d = \min_k \sqrt{(f(RH_k) - f(RH_{obs}))^2 + \frac{1}{\Delta h_{max}}(h_k - CTH_{obs})^2}$$

function of relative humidity

= 1

height of model level k

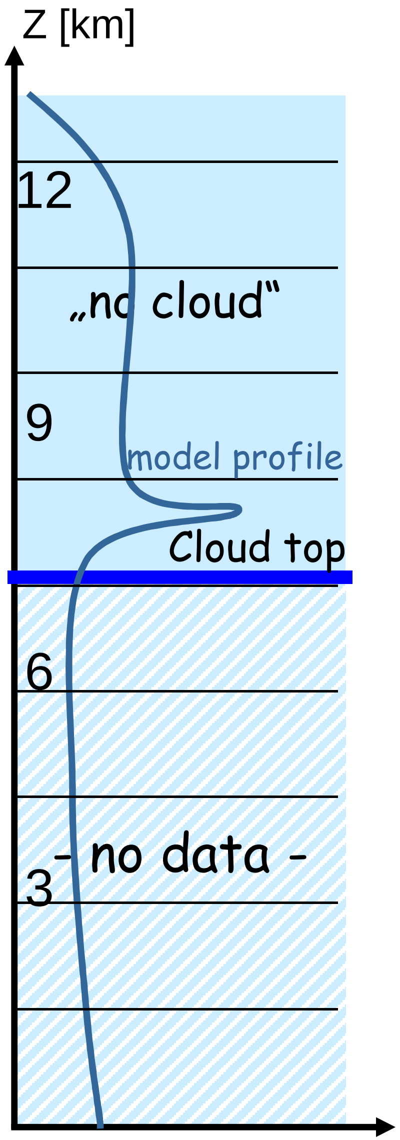
- use $f(RH_{obs}) - f(RH_k) = 1 - f(RH_k)$
  and $CTH_{obs} - h_k$
  as 2 separate obs increments in LETKF



type of obs increment , if no cloud observed ?

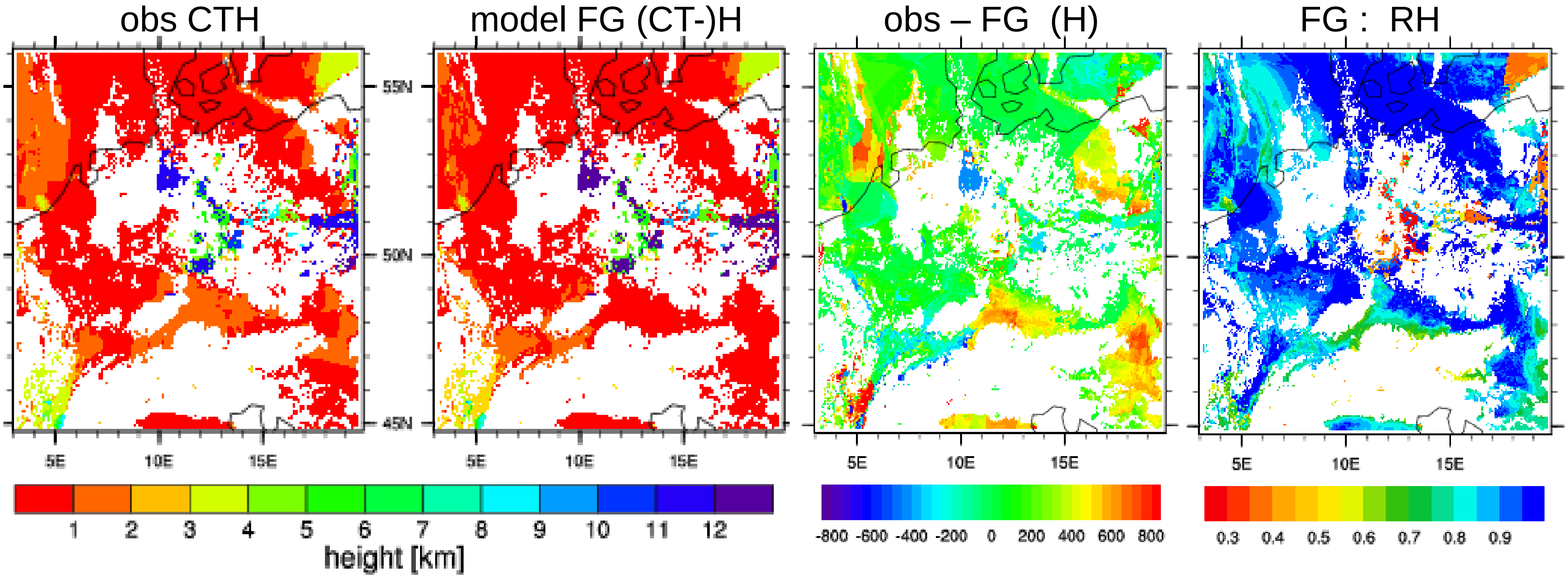
- assimilate CLC = 0 separately
  for high, medium, low clouds

- model equivalent:
  maximum CLC within vertical range



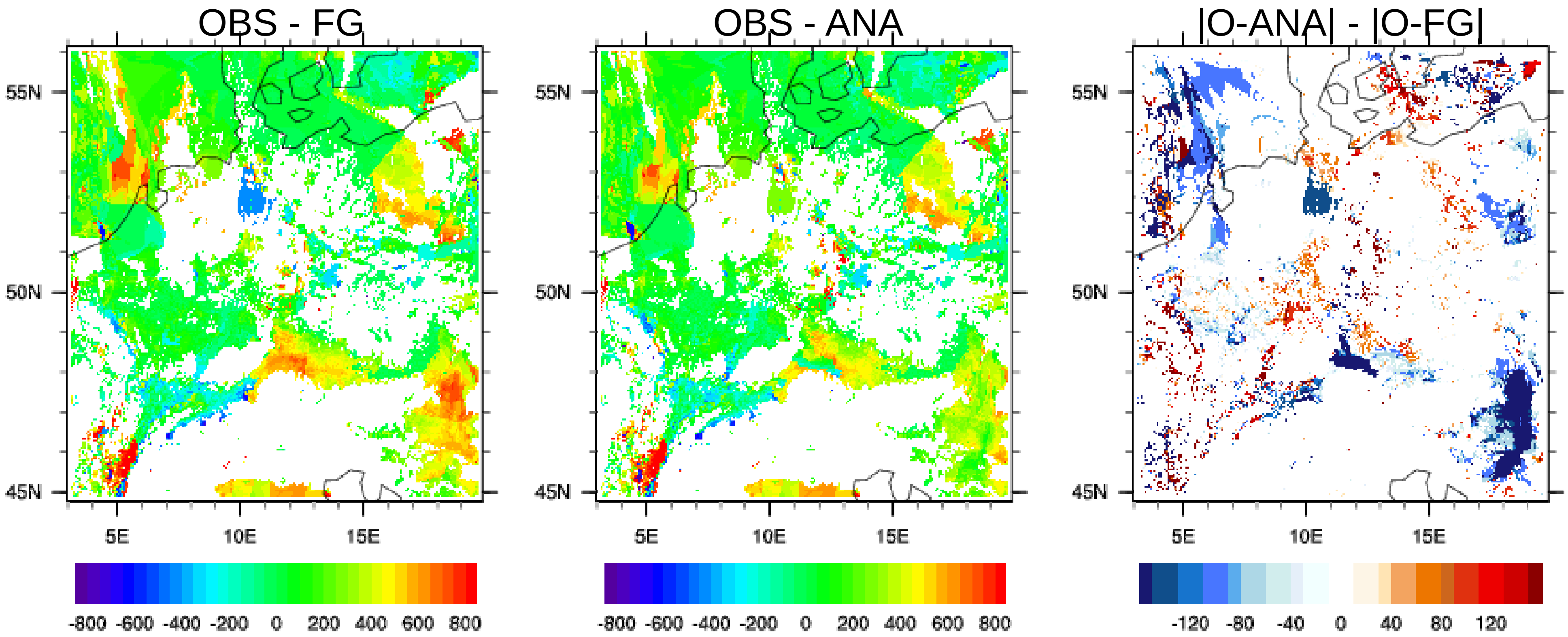
use of cloud info :
assimilation of ‘cloud analysis’ : example

17 Nov 2011, 6 UTC (low stratus case)
pixels where observation has clouds
(output from feedback files)



use of cloud info :
first assimilation experiment

‘cloud‘ top height



blue means ANA is better

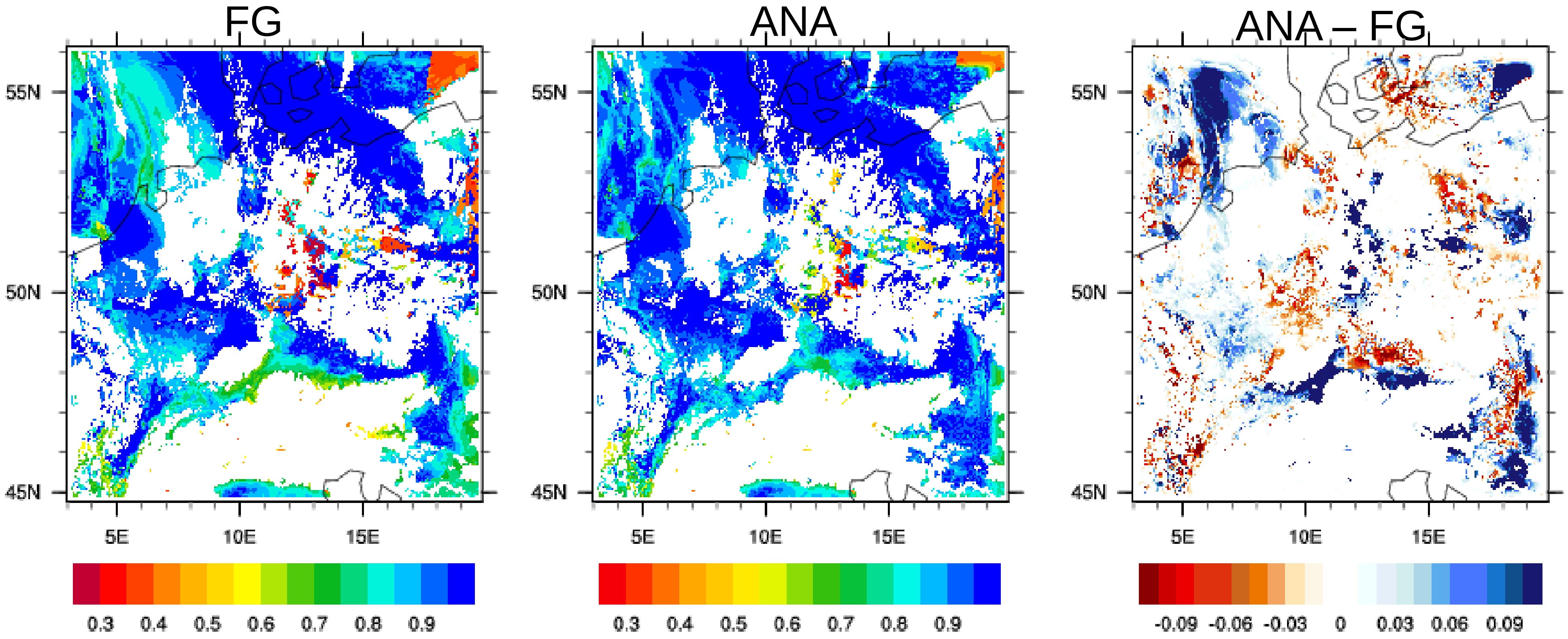
Here: results of deterministic run in LETKF framework

(Kalman gain matrix applied to standard (unperturbed) model integration)

use of cloud info :
first assimilation experiment

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Relative humidity at ‘cloud‘ level

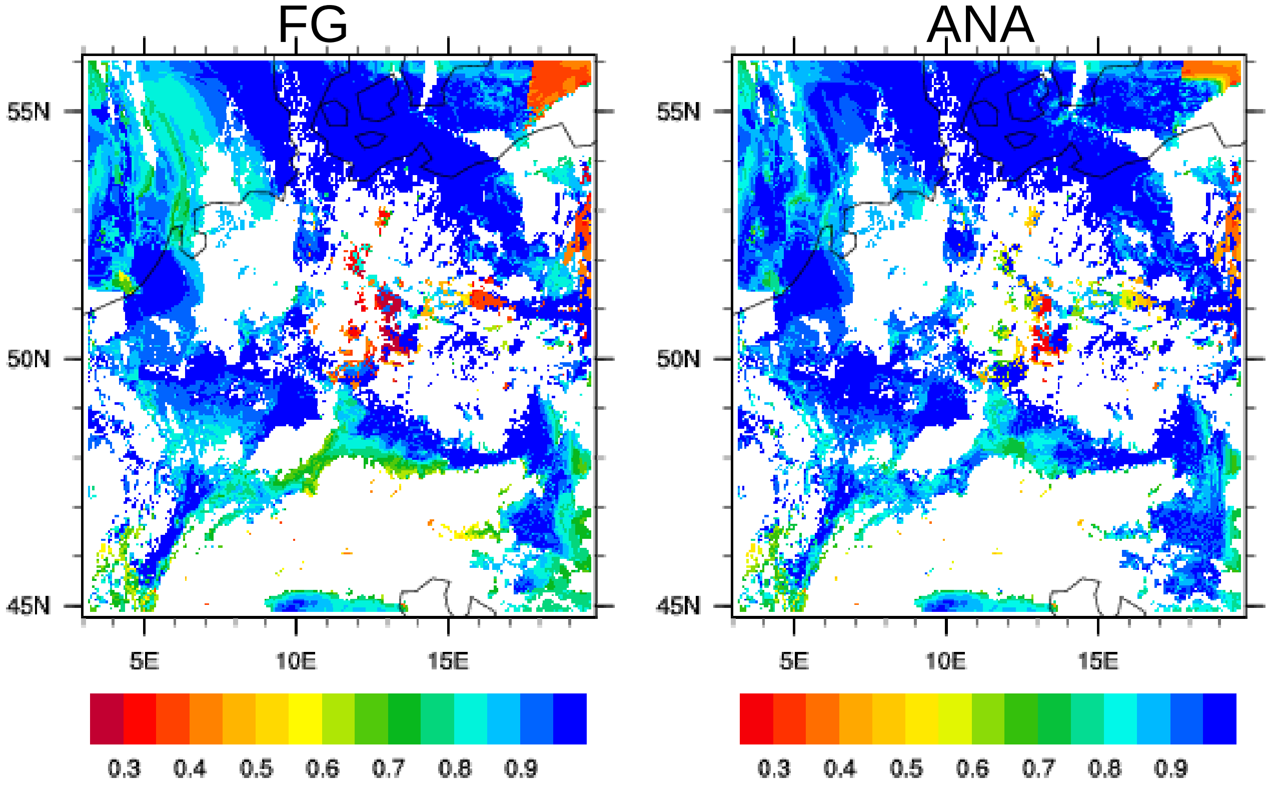


Here: results of deterministic run in LETKF framework

(Kalman gain matrix applied to standard (unperturbed) model integration)

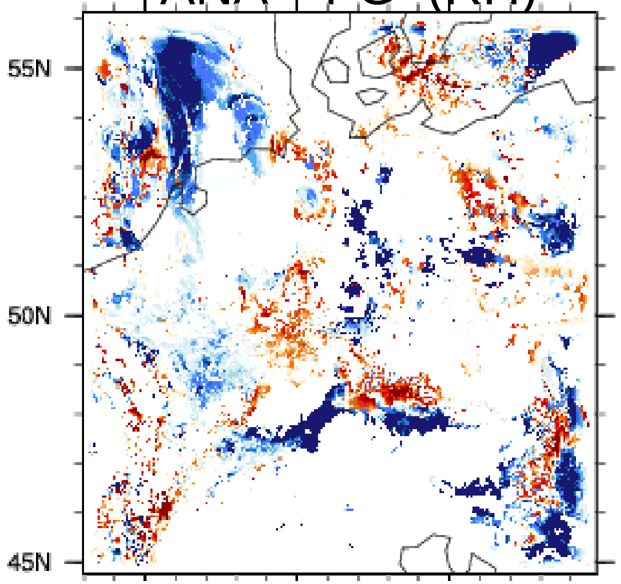
use of cloud info :
first assimilation experiment

Relative humidity at cloud level

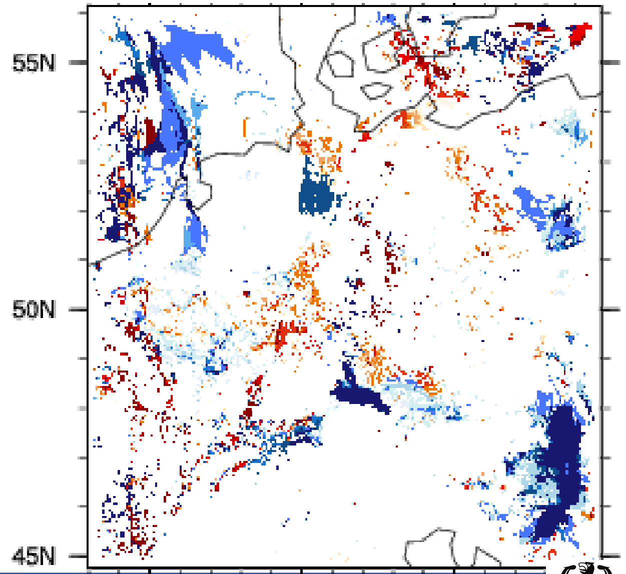

FG

ANA

ANA – FG (RH)



|O-ANA| - |O-FG| (‘CTH’)



→ LETKF draws model cloud tops closer to obs

next :
- detailed evaluation (cross section, profiles…)
- single observation experiments
- tuning of obs. error, thinning, localization

KENDA: use of additional observations
use of radar obs

- radar : assimilate 3-D radial velocity and 3-D reflectivity directly

1. observation operators implemented

   (Uli Blahak (DWD), Yuefei Zeng, Dorit Epperlein (PhD, KIT))

   – full, sophisticated

   – efficient (e.g. lookup tables for Mie scattering)

   – tested for sufficiently accurate and efficient approximations for DA

2. assimilation experiments

   – technical work finished this week

   – 1 - 2 assimilation case studies (Zeng)

   – 2013: Klaus Stephan : test periods, tuning …

Task 4.3: use of GNSS slant path delay

- ground-based GNSS slant path delay SPD
  - produce & use tomographic refractivity profiles (Erdem Altunac, PhD)
  - implement non-local SPD obs operator & use SPD (Michael Bender)

    - first implement SPD obs operator in 3DVAR package
      (environment for work on tomography)
      - implement simple operator (refractivity along straight line)
      - implement complex obs operator with ray tracer
      - monitoring, test e.g. impact of straight line approximation
    - then implement obs operator in COSMO (in 2013)

The 5th EnKF Workshop, Rensselaerville (New York), May 21-24, 2012

Session 5: Operational Implementations

Session chair: Lucio Torrisi (torrisi@meteoam.it)

Center	Scheme	Use	Operations	Presentation
CMC (Canada)	EnKF (Stochastic)	- EPS initialization	January 2005	Oral
NCEP (USA)	Hybrid (3DVAR/EnSRF)	- Deterministic forecast	May 2012	Oral (sess.2)
UKMO	Hybrid (4DVAR/ Local ETKF)	- Deterministic forecast - EPS Perturbations	July 2011 June 2006	Oral
MF (France)	EDA (Ensemble of 4DVAR)	- Initial B variances in 4DVAR - EPS initial perturbations together with SV	July 2008 December 2009	-
ECMWF	EDA (Ensemble of 4DVAR)	- Initial B variances in 4DVAR - EPS initial perturbations together with SV	May 2011 June 2010	Oral
CNMCA (Italy)	Regional EnKF (LETKF)	- Deterministic forecast	June 2011	Poster

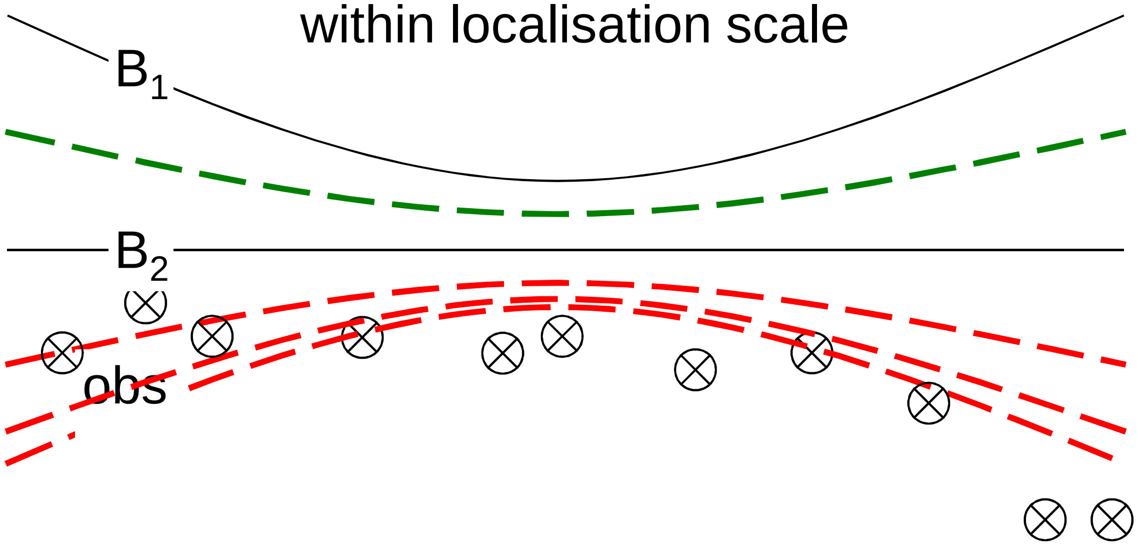
Overview KENDA

thank you for your attention

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

- adaptive estimation of obs error covariance R

  (Li, Kalnay, Miyoshi, QJRMS 2009) , but our implementation: in ensemble space !



f.g. mean = $0.5\ B_1 + 0.5\ B_2$

ana $\cong$ f.g.mean + $(B_2 - B_1)$
$= 1.5\ B_2 - 0.5\ B_1$ (1 perfect obs)

2 obs → least square fit

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

- cannot be fitted well, improve analysis only slightly
- decrease analysis error ! $\mathbf{P}_w^{\ a} = \left[(k-1)\,\mathbf{I} + \left(\mathbf{Y}^b\right)^T \mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1}$

→ adaptive R takes that into account and increases R

LETKF:
account of model error / additive inflation

- parameterisation of model error using statistics (Tsyrulnikov, Gorin) :
  - parameterisation: $\underline{e \;=\; \mu * F_{phys}(x) \;+\; e_{add}}$
  - estimate parameters by fitting to statistics from forecast tendency and observation tendency data (using a maximum likelihood based method)

failed in OSSE setup with simulated ME for finite-time 1 – 6 hr tendencies !!!

main methodological cause of failure : instantaneous ME is contaminated
in finite-time tendencies
by other tendency errors :

- trajectory drift as a result of ME themselves
- initial errors (plus the trajectory drift due to initial errors)

→ conclusion: observation accuracy and spatio-temporal coverage
far from being sufficient to reliably estimate ME !

→ task is stopped !

LETKF :
implementation of verification

- tool for production of ‘full’ NetCDF feedback files
  - make clean interfaces to observation operators / QC in COSMO : done
  - … integrate them into 3DVAR package : in progress
  - and extend flow control (read correct (hourly) Grib files etc.) : to be done

  should be ready by end of 2012 (for VERSUS)

- ensemble-related diagnostic + verification tool, using feedback files : (Iriza, NMA)

  → computes statistical scores for different runs (‘experiments’),

  → focus: use exactly the same observation set in each experiment !

  → select obs according to namelist values (area, quality + status of obs, …)
  - implementing ensemble scores (reliability, ROC, Brier Skill Score, (continuous) Ranked Probability Score)
  - main part of documentation written

SMC: LETKF: implementation

- modifications in COSMO in official code (V4_24)
  (e.g. in order to have a sub-hourly update frequency)

- COSMO-DE LETKF implemented in NUMEX and tested
  (e.g. stand-alone 2-day experiment reproduced)

- GME-LETKF & ensemble INT2LM for DA cycle implemented in NUMEX, being tested, should be available end of Sept.

  → in Oct., start first KENDA experiments in NUMEX over several days/weeks
  but: - direct interpolation from 60 km to 2.8 km !
  - deterministic analysis not yet implemented in NUMEX

- ensemble LBC 2013 – 2014 :
  ensemble perturbations of interpolated ensemble GME fields,
  added to deterministic COSMO-DE LBC

LETKF :
implementation / activities

- in past year, still only preliminary LETKF experiments possible, using Hendrik's scripts:
  - up to 2 days (7 – 8 Aug. 2009: quiet + convective day)
    → 3-hourly (15-min) cycles
  - 32 ensemble members
  - perturbed LBC: COSMO-SREPS, 3 * 4 members

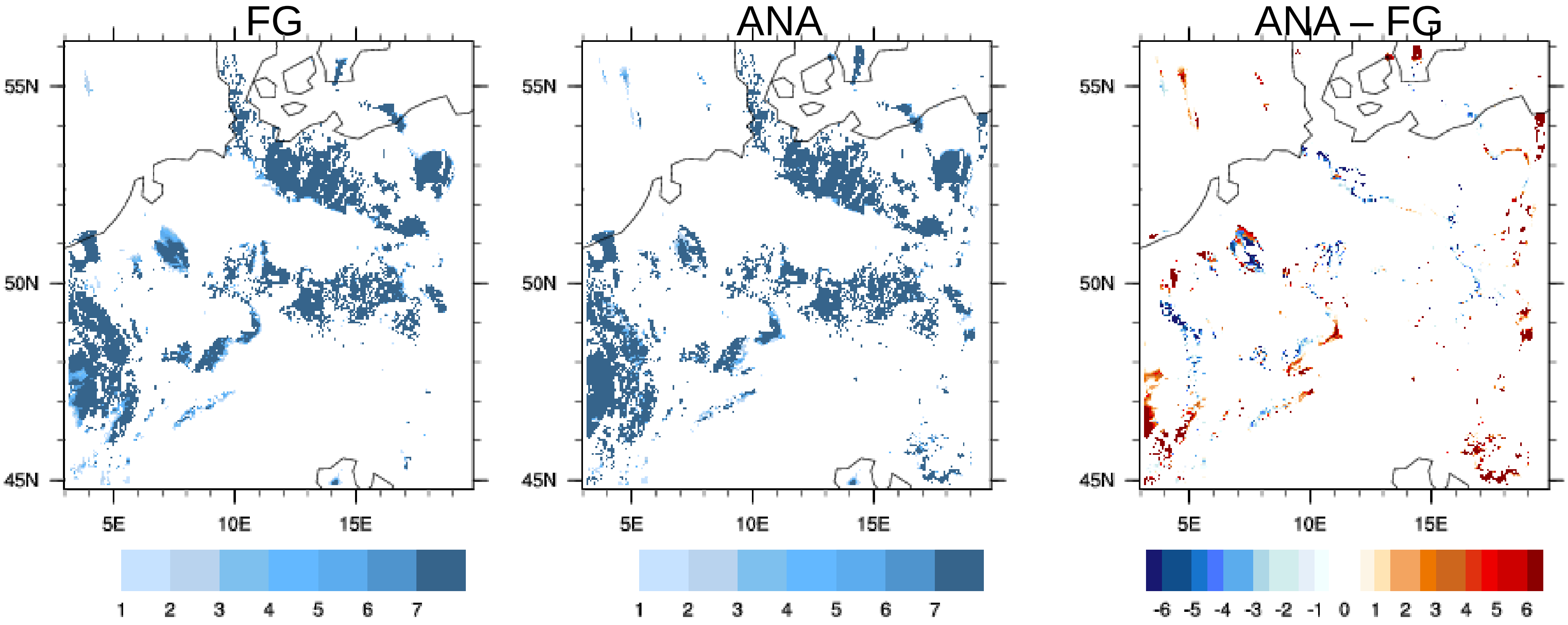
→ therefore

- theoretical studies, toy model experiments related to adaptive localisation
  → talk by Hendrik Reich
- benchmark , winter school on DA , support for HErZ centre , testing (e.g. NUMEX) …
- only few COSMO-DE experiments related to adaptive localisation

use of cloud info :
first assimilation experiment

pixels where model has clouds and observation not

<u>Low clouds</u>



→ from “cloudfree“ obs no positive impact so far

next steps :
- detailed evaluation (cross section, profiles…)
- single observation experiments
- tuning of observation error, thinning, localization

Task 3: scientific issues & refinement of LETKF

– lack of spread: (partly ?) due to model error and limited ensemble size
which is not accounted for so far

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

→ additive (see later)

→ multiplicative $X_b \rightarrow \rho \cdot X_b$

(by tuning, or) adaptive $$\left\langle (y - H(x_b))(y - H(x_b))^T \right\rangle = \mathbf{R} + \rho \mathbf{HP}_b\mathbf{H}^T$$

→ pre-specifed R is used for adaptive ρ :

→ need for careful specification / tuning of obs errors

– observation error covariance R :

also estimate adaptively (Li, Kalnay, Miyoshi, QJRMS 2009)

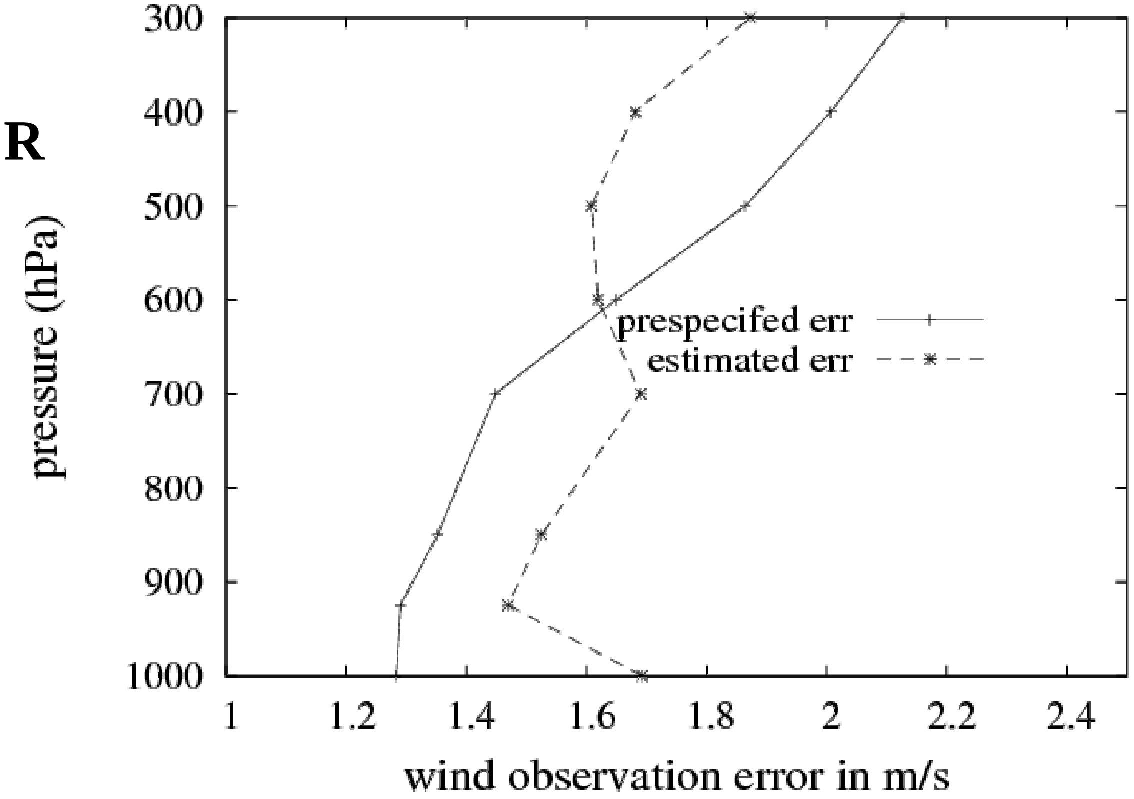
$$\left\langle (y - H(x_a))(y - H(x_b))^T \right\rangle = \mathbf{R}$$

Task 3: scientific issues & refinement of LETKF

- adaptive observation errors

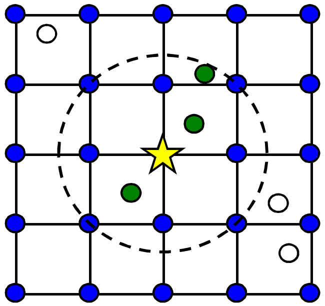
$$\left\langle \left(y - H(x_a)\right)\left(y - H(x_b)\right)^T \right\rangle \quad = \quad \mathbf{R}$$

(in observation space)



- adaptive R in ensemble space:

  adjusts total weight, not relative weight of obs



- localisation