Status Overview on PP KENDA

Km-scale ENsemble-based Data Assimilation

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

christoph.schraff@dwd.de
Deutscher Wetterdienst, D-63067 Offenbach, Germany

Contributions / input by:

Hendrik Reich, Andreas Rhodin, Roland Potthast, Uli Blahak, Klaus Stephan (DWD)
Annika Schomburg (DWD / Eumetat)
Yuefei Zeng, Dorit Epperlein (KIT Karlsruhe)
Daniel Leuenberger, Manuel Bischof (MeteoSwiss)
Mikhail Tsyrulnikov, Vadim Gorin (HMC)
Lucio Torrisi (CNMCA)
Amalia Iriza (NMA)

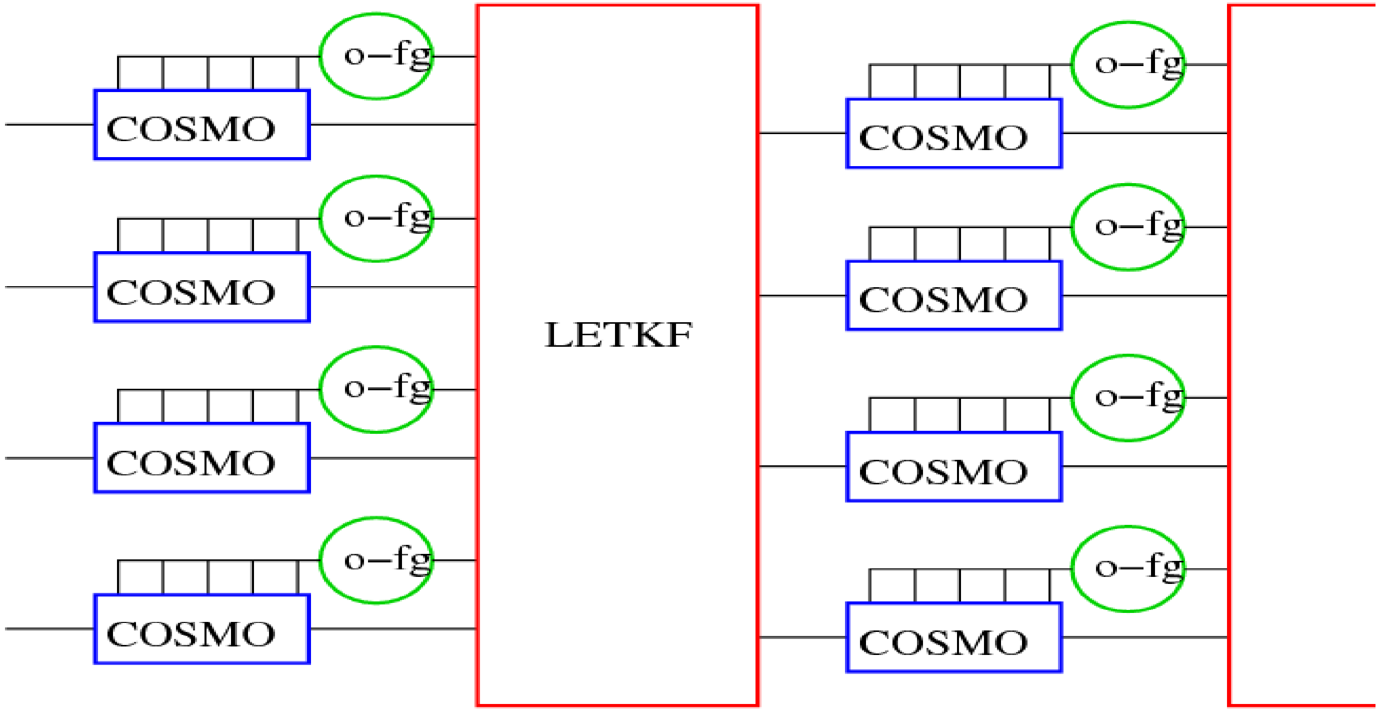
- Task 1: General issues in the convective scale (e.g. non-Gaussianity)
- Task 2: Technical implementation of an ensemble DA framework / LETKF
- Task 3: Tackling major scientific issues, tuning, comparison with nudging
- Task 4: Inclusion of additional observations in LETKF

Task 2: Technical implementation of an ensemble DA framework / LETKF

analysis step (LETKF) outside COSMO code

→ ensemble of independent COSMO runs up to next analysis time

→ separate analysis step code, LETKF included in 3DVAR code of DWD



collect obs from $]t_{ana}–3\text{h} , t_{ana}]$

old setup: obs from $]t_{ana}–1.5\text{h} , t_{ana}+1.5\text{h}]$ thinned

Task 2: Technical implementation of an ensemble DA framework / LETKF

- modifications in COSMO fully implemented, but not yet in official code

  (also in order to have a sub-hourly update frequency)

- first version LETKF (KENDA) implemented in NUMEX, to be tested

  $\rightarrow$ still use scripts to do a few stand-alone cycles with LETKF

- deterministic analysis implemented $\rightarrow$ talk of Hendrik

- LBC at analysis time : use (perturbed) analysis ensemble member itself instead of (temporally interpolated) LBC fields from steering ensemble system

- next step: LBC from GME LETKF ensemble
  interpolate GME ensemble perturbations and add to deterministic COSMO-DE LBC

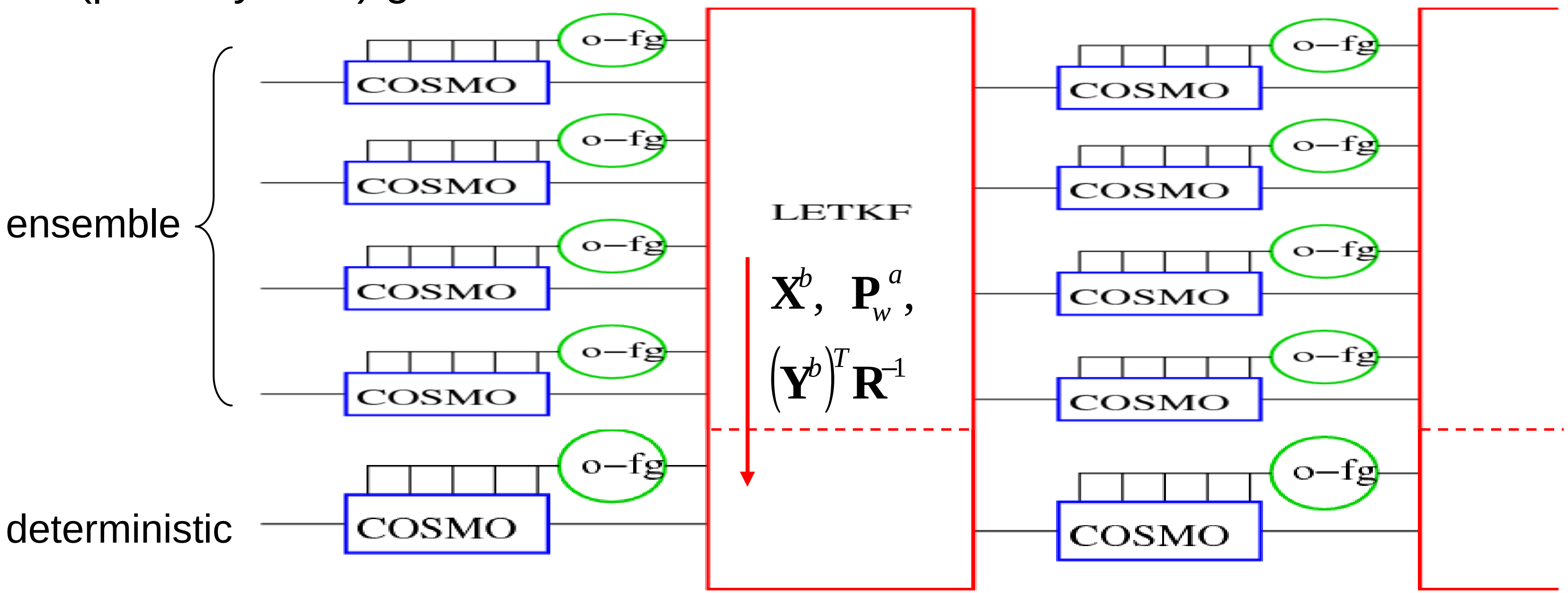
Analysis for a deterministic forecast run : use Kalman Gain $\mathbf{K}$ of analysis mean

$$\mathbf{x}^A = \mathbf{x}^B + \mathbf{K}\left[\mathbf{y}^o - H(\mathbf{x}^B)\right] \quad , \quad \mathbf{K} = L\ \mathbf{X}^b \left(\mathbf{P}_w^{\ a}\right)^{-1} \left(\mathbf{Y}^b\right)^T \mathbf{R}^{-1}$$

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

deterministic analysis recently implemented

L : interpolation of analysis increments from grid of LETKF ensemble to (possibly finer) grid of deterministic 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

Task 2: Technical implementation: verification for KENDA

‘stat’-utility:
compute model (forecast) – obs for complete NEFF (NetCDF feedback files) :

→ adapt verification mode of 3DVar/LETKF package

→ need to implement COSMO observation operators with QC (‘library 2’) in 3DVar/LETKF package
(→ hybrid 3DVar–EnKF approaches in principle applicable to COSMO)

→ build COSMO modules with clean interfaces & introduce them in LETKF
- no further progress
- still plan to first implement upper-air obs (needed as input for VERSUS)

NEFFprove (Amalia Iriza, NMA) :
tool to compute and plot verification scores based on NetCDF feedback files (NEFF)

- uses NEFF files from 3DVAR and COSMO environments
- computes statistical scores for different runs ('experiments'),

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

  → sort observations in a unique order (lon / level / time / lat / variable / code ..)

  → select obs according to namelist values (area, obs quality, status, …)

- compute scores (for each experiment, variable, forecast time, vertical level, … , upper-air obs, screen-level obs, RR, cloud, etc.)
  - deterministic continuous: BIAS, RMSE, MSE
  - dichotomous: accuracy, Heidke Skill Score, Hanssen and Kuiper discriminant
  - ensemble scores (to be done): reliability, ROC, Brier Skill Score, (continuous) Ranked Probability Score
- write scores to ASCII files (1 file for each variable, with all scores for each exp.)
- plot the scores, using GNUplot (work is in progress)

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 ?

$\rightarrow$ multiplicative $X_b \rightarrow \rho \cdot X_b$ (tuning, or adaptive ($y - H(x) \sim R + H^T P_b H$))

$\rightarrow$ additive ((e.g. statistical 3DVAR-B), stochastic physics (Torrisi))

– localisation (multi-scale data assimilation, 2 successive LETKF steps with different obs / localisation ?)

$\rightarrow$ talk by Hendrik Reich

– model bias
– update frequency : up to now only 3-hourly
– noise control : new: use IC as initial LBC
– non-linear aspects, convection initiation (latent heat nudging ?)
– technical aspects: efficiency, system robustness

(Nov. 2012: regular ‘pre-operational’ LETKF suite)

LETKF scientific issues / refinements: model error estimation

- **stochastic physics (Palmer et al., 2009)** : by Lucio Torrisi, available Oct. 2011

  → talk by Lucio Torrisi

- **objective estimation and modelling of model (tendency) errors**

  → talk by Michael Tsyrulnikov

- **investigate LETKF in Observing System Simulation Experiments (OSSE)**
  - apply LETKF to idealized convective weather systems,
    tune LETKF settings (localization, covariance inflation)
  - quantify non-Gaussianity and spin-up time in the evolving LETKF ensemble during the assimilation of a convective storm
  - develop ways to improve Gaussianity of errors and reduce spin-up time of the LETKF → test "running in place" (Kalnay and Yang, 2010)
    and "outer loop" (Yang and Kalnay, 2010) algorithms

MeteoSwiss: plan for 2-year project not accepted

- however master thesis (5 months): Manuel Bischof, MeteoSwiss
  perturbation for ensemble generation:
  - which quantities, describing the storm environment, suitable to perturb ?
  - meaningful perturbation amplitudes ?

→ talk by Daniel Leuenberger

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

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

  – Implement full, sophisticated observation operators
    (by end 2011)
  – then develop different approximations and test their validity in order to obtain sufficiently accurate and efficient operators
    – by looking at the simulated obs (by March 2012)
    – doing assimilation experiments : OSSE setup

Particular issues for use in LETKF: obs error variances and correlations, superobbing, thinning, localisation

→ talk by Uli Blahak

Task 4.3: use of GNSS slant path delay

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- **ground-based GNSS slant path delay** (early 2012, N.N.)
  - implement non-local obs operator in parallel model environment
  - test and possibly compare with using GNSS data in the form of IWV or of tomographic refractivity profiles

Particular issue: localisation for (vertically and horizontally) non-local obs

Task 4.4: use of cloud info

- **cloud information based on satellite and conventional data**

  Annika Schomburg (DWD / Eumetsat)

  - 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
  - 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
  - compare approaches

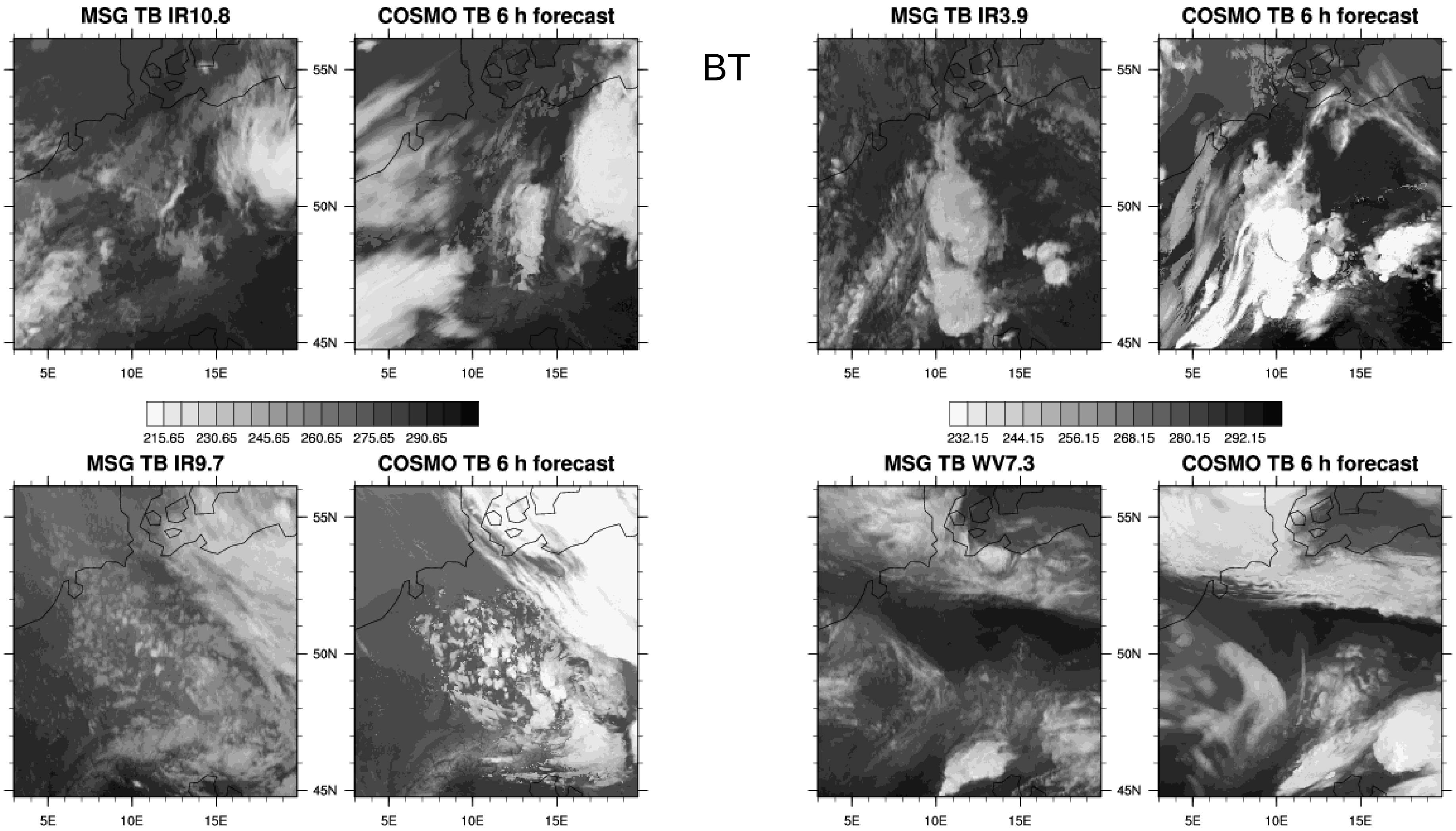
Particular issues: non-linear observation operators,
non-Gaussian distribution of observation increments

Task 4.4: use of cloud info

- make observations available
  - aggregation / interpolation to COSMO-DE / COSMO-EU grids
  - regular archiving:
    - NWC-SAF cloud products available since May 2011:
      (CT: cloud type, CTT: cloud top temperature, CTH: cloud top height)
    - BT available since June 2011: all 12 SEVIRI channels
- assumption in LETKF: no bias

  $\rightarrow$ look at systematic differences in BT / cloud products

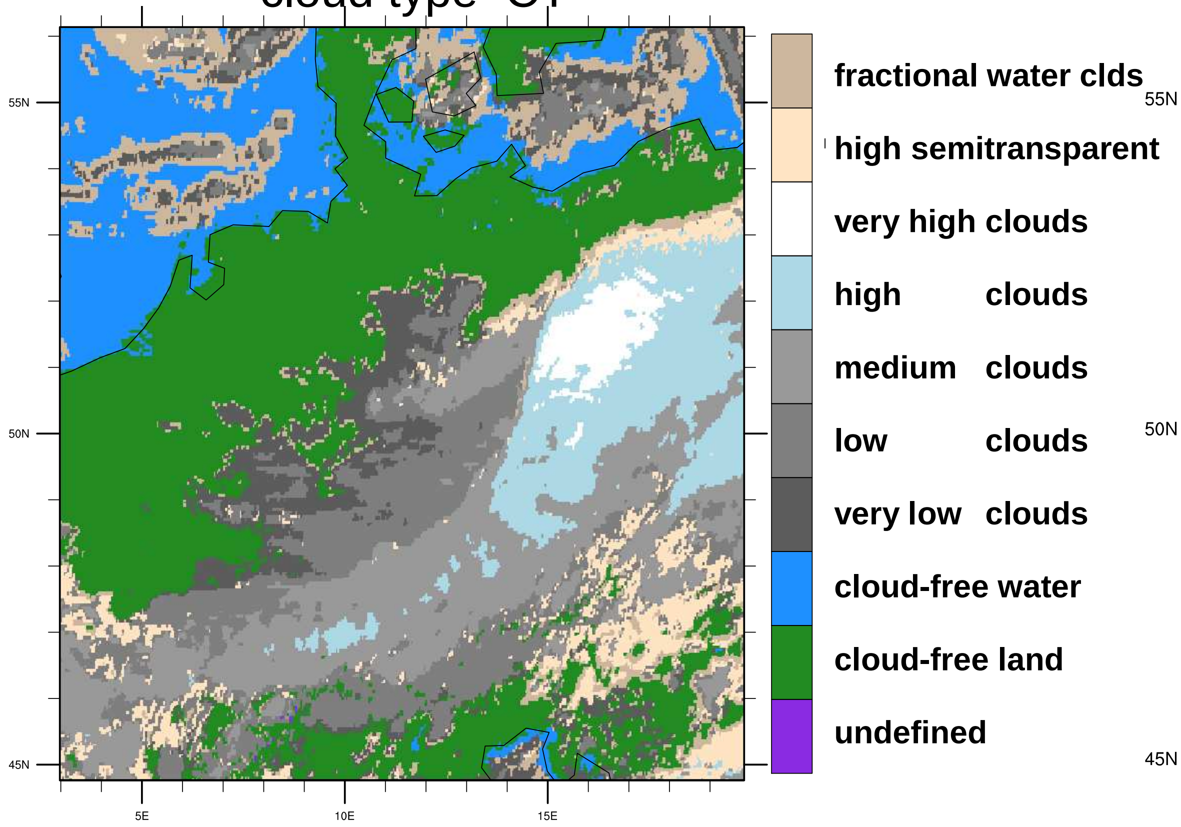
Task 4.4: use of cloud info



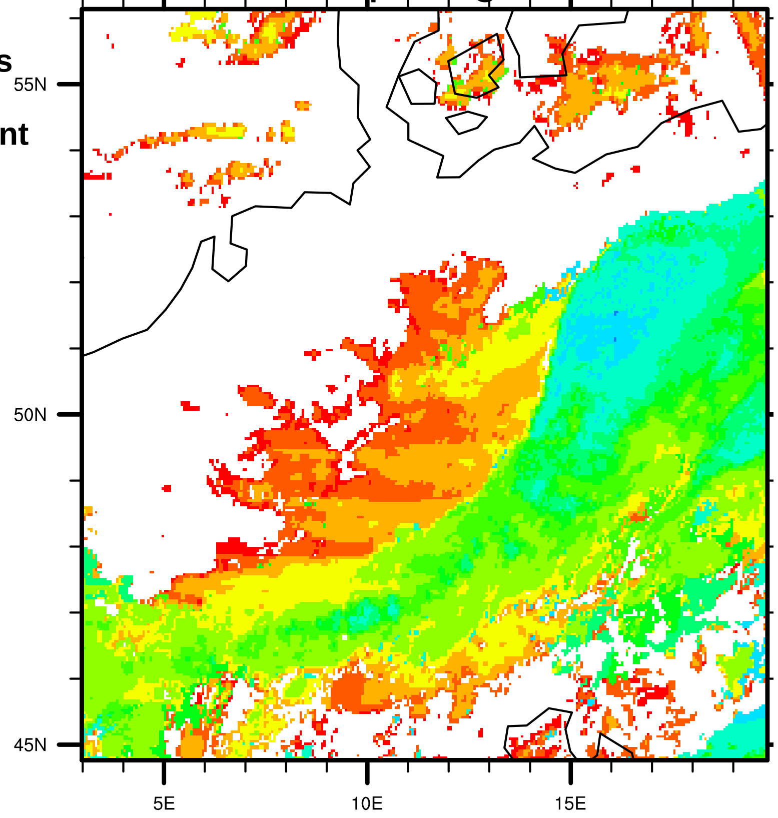
high clouds: cold bias of model-BT, (partly ?) due to known bias in RTTOV-9

NWC-SAF SEVIRI cloud products: example

cloud type CT



cloud top height CTH



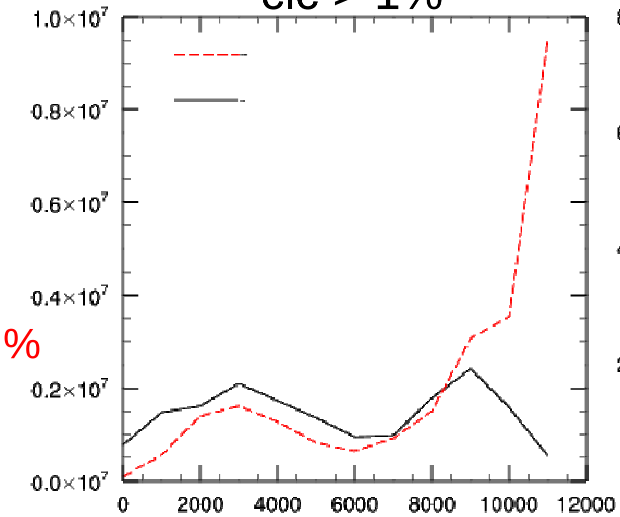
COSMO: cloud water $q_c > 0$, or cloud ice $q_i > 5 \cdot 10^{-5}$ kg/kg $\rightarrow$ clc = 100 %
subgrid-scale clouds $\rightarrow$ clc = f(RH; shallow convection; q_i , $q_{i,sgs}$) < 100 %

Task 4.4: use of cloud info

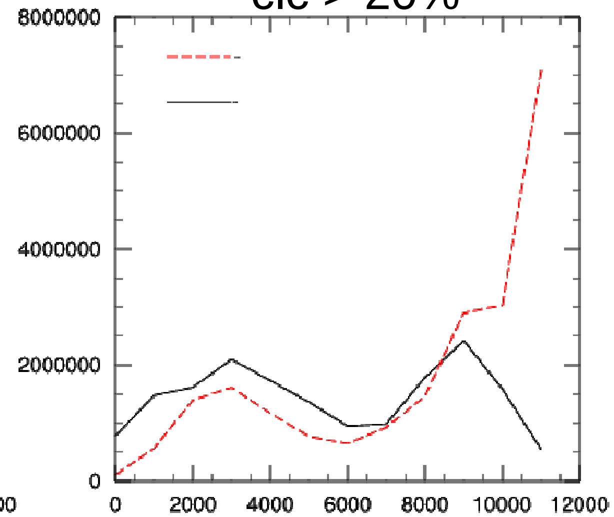
cloud top height CTH distribution
May – July 2011

—— NWC-SAF CTH 'obs'
- - - - COSMO CTH for clc > x %

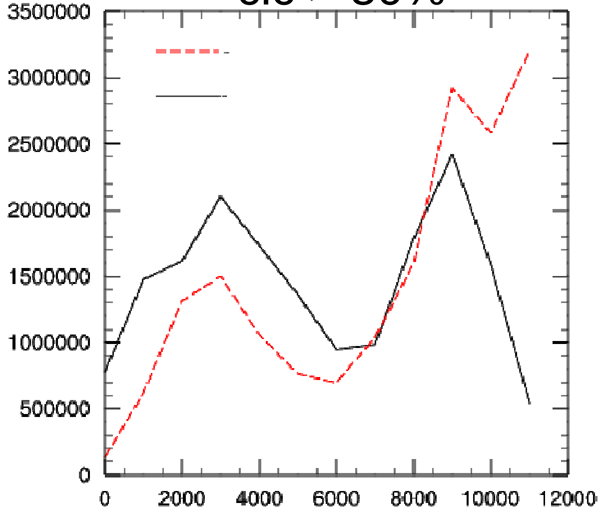
clc > 1%



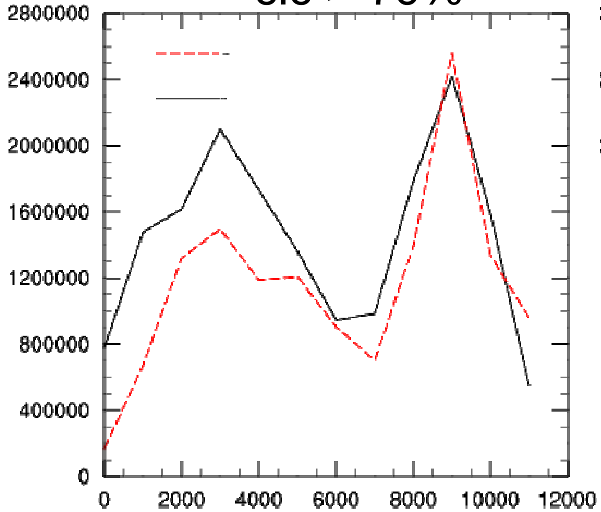
clc > 20%



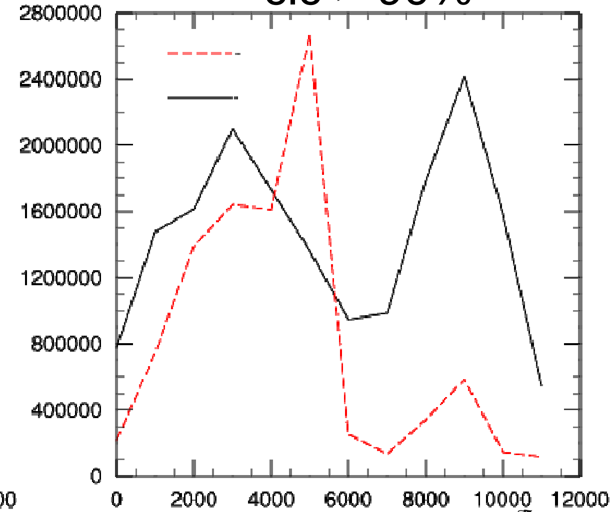
clc > 50%



clc > 70%



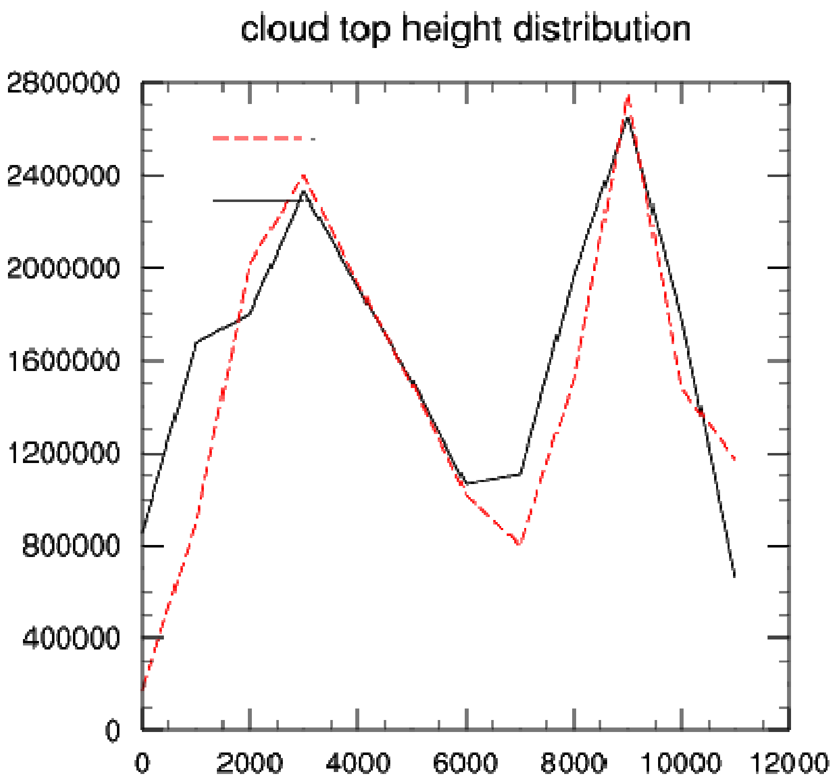
clc > 90%



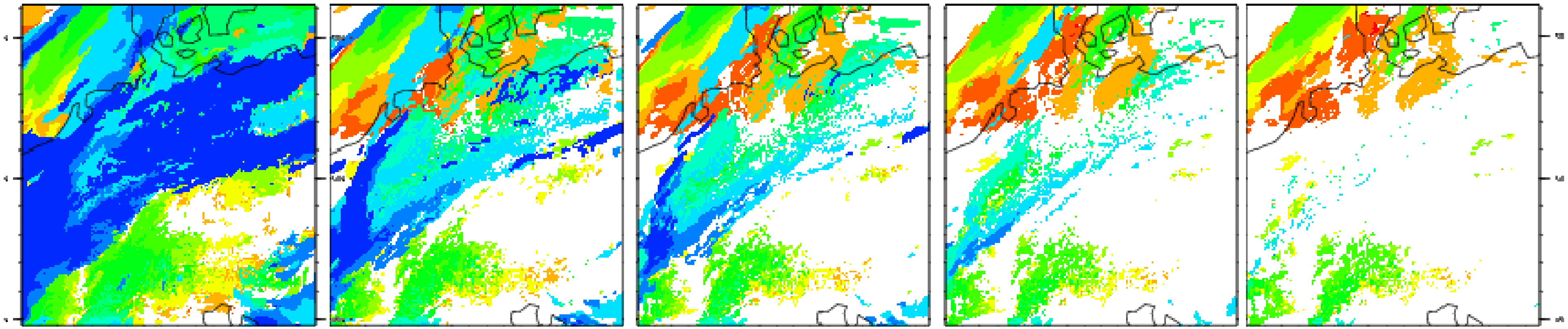
cloud top height CTH distribution May – July 2011

NWC-SAF CTH 'obs'

COSMO CTH with optimal threshold: for clc > 70 % for levels above 5000 m , for clc > 40 % for levels below 5000 m

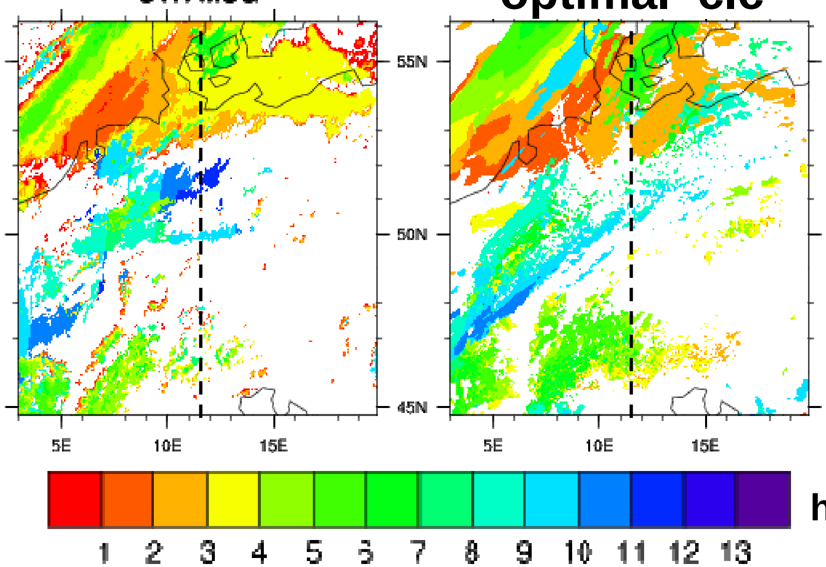


CTH COSMO 6-h forecast, for 18 May 2011, 18 UTC

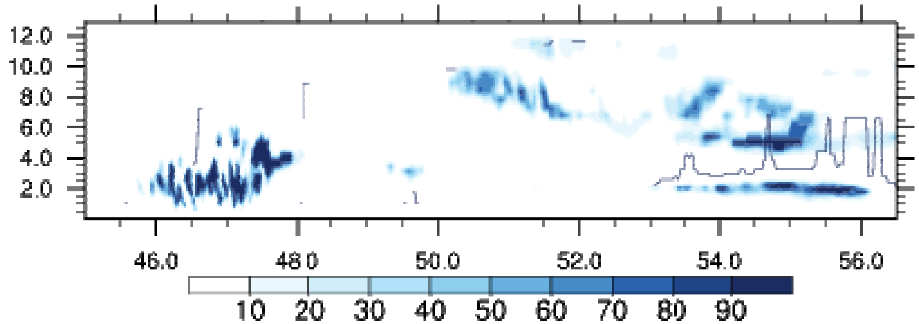


clc > 0.1% **clc > 30%** **clc > 50%** **clc > 70%** **clc > 90%**

CTH MSG **optimal clc**



height [km]



cloud cover [%] , shading: model
isoline: CTH MSG

Technical implementation of verification for KENDA

Proposition of WG3b (Jean-Marie Bettems) for other WGs :

WG1 : Assimilation of land surface data

To overcome some climatological assumptions, evaluate usage of LandSAF products

MSG/SEVIRI based products

- Wild Fires (Fire Radiative Power, Fire Risk Map, Fire Detection and Monitoring)
- Vegetation Parameters (Fraction of Vegetation Cover, Leaf Area Index, Fraction of Absorbed Photosynthetic Active Radiation)
- Snow Cover (daily, 15 mins)
- Other : Bi-Directional Reflectance Factor, Land Surface Emissivity
- Surface Albedo
- **Land Surface Temperature** (15 mins)
- Down-welling Surface Fluxes (short-wave, long-wave)
- **Evapotranspiration** (30 mins, daily) → additional control variable in var SMA ?

MetOp/AVHRR based products

- Land Surface Temperature
- Down-welling Surface Fluxes (long-wave)

christoph.schraff@dwd.de

Technical implementation of verification for KENDA

NEFFPROVE verification tool

- Post-processing routine to prepare statistics for ensemble runs
- Verification based on netcdf feedback files
- Uses files from the 3DVAR and COSMO environments
- Computes statistical scores for different runs, based on exactly the same observations for each of the used experiments

christoph.schraff@dwd.de

Technical implementation of verification for KENDA

- namelist reading: specifications on models + observation types used,
  selection criteria for the obs (lat/lon, quality, state, …) , etc.

- select obs for each 3-hrly time box :
  - sort observations to get a unique order which allows the identification of the observations in all experiments (reduces computing time for the next step):
    - determine maximum number of observations
    - sort observations according to three criteria:
      - longitude, vertical level
      - time, latitude, observed variable
      - code type, sub-centre, status
  - select the observations (dep. on namelist):
    - eliminate the observations that are not used in all the experiments, or do not correspond to at least one of the criteria specified in the namelist
    - select according to verification times

Technical implementation of verification for KENDA

- update contingency tables:
  - compute intermediate scores for upper air temperature, relative humidity, surface pressure, wind components, vertical velocity;
  - update contingency tables for dichotomous scores (for precipitation amount and total cloud amount).

- compute scores (for each experiment, each variable, forecast time, vertical level, …)
  - deterministic continuous: BIAS, RMSE, MSE
  - dichotomous: accuracy, Heidke Skill Score, Hanssen and Kuiper discriminant
  - ensemble score (to be done): reliability, ROC, Brier Skill Score, (continuous) Ranked Probability Score

- write scores to ASCII files (1 file for each variable, containing all computed scores for each experiment.

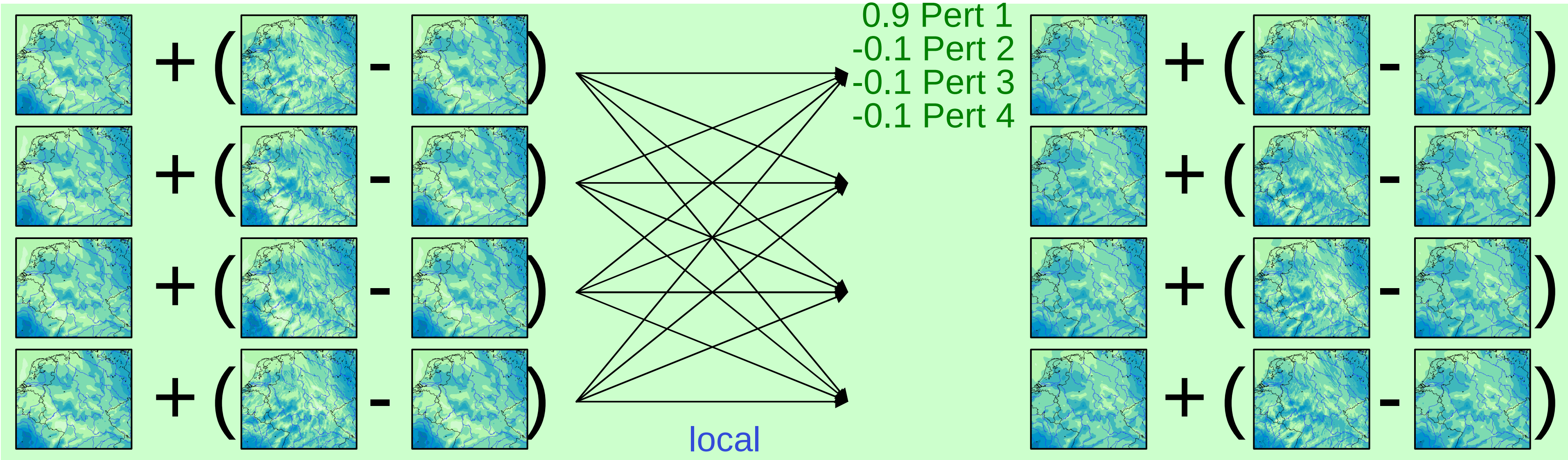
- plot the scores, using GNUplot (work is in progress)

christoph.schraff@dwd.de

Technical implementation of verification for KENDA

Local Ensemble Transform Kalman Filter
LETKF (Hunt et al., 2007)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand



ensemble mean forecast

k perturbed forecasts

ensemble mean fcst.

forecast perturbations $\mathbf{X}^b$

↓

flow-dep background error cov.

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

↓

local 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

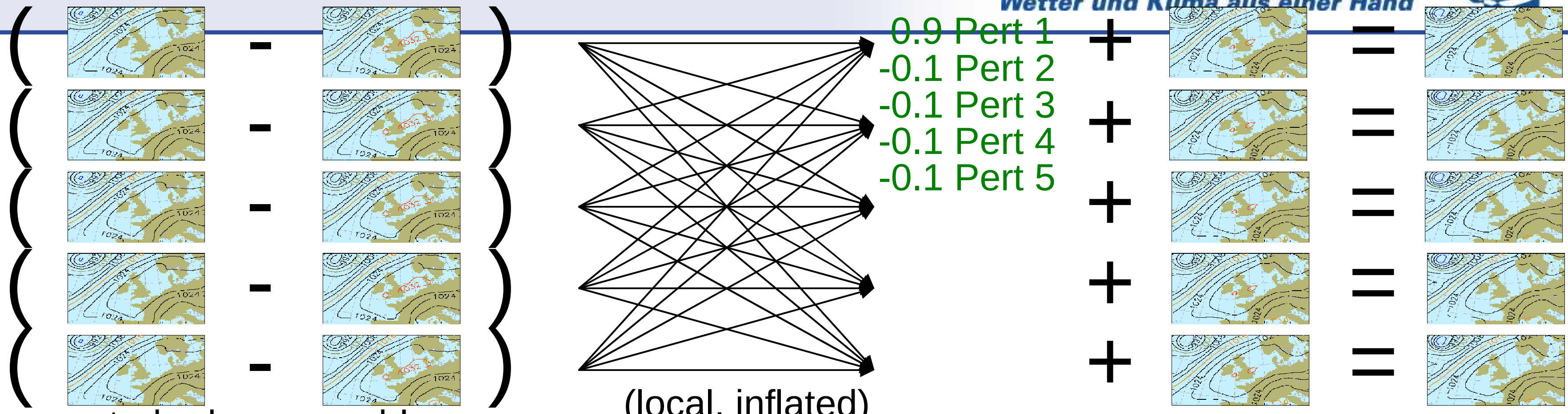
in the (*k*-1) -dimensional (!) sub-space S spanned by background perturbations :

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

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

Local Ensemble Transform Kalman Filter (LETKF)

(chart based on a slide of Neil Bowler, UK MetOffice)



perturbed forecasts – ensemble mean forecast

= background perturbations X^b

→ flow-dep. backgr. error covar.

(local, inflated) transform matrix $\mathbf{W^a}$

analysis error covariance

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

perturbed analyses

in the (5-dimensional) sub-space S spanned by background perturbations :

(linearise obs operator around the mean of ensemble of simulations of an obs) $\overline{\mathbf{y}}^b \equiv \overline{\mathbf{y}^{b(i)}} \equiv \overline{H(\mathbf{x}^{b(i)})}$

explicit solution for minimisation of cost function (Hunt et al., 2007)

→ analysis (mean) and analysis error covariance : columns of $\mathbf{Y}^b$: $\mathbf{y}^{b(i)} - \overline{\mathbf{y}}^b$

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

and in model space : $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{X}^b \overline{\mathbf{w}}^a$, $\mathbf{P}^a = \mathbf{X}^b \mathbf{P}_w{}^a \left(\mathbf{X}^b\right)^T$

→ analysis ensemble member : $\mathbf{x}^{a(i)} = \overline{\mathbf{x}}^a + \mathbf{X}^{a(i)} = \overline{\mathbf{x}}^b + \mathbf{X}^b\left(\overline{\mathbf{w}}^a + \mathbf{W}^{a(i)}\right)$ $\quad \mathbf{W}^a = \left[(k-1)\mathbf{P}_w{}^a\right]^{1/2}$

LETKF (COSMO) : method to deal with non-Gaussianity

- „No cost smoother“ (Kalnay and Yang, QJRS, submitted)
  - Apply weights from time t+1 at time t
- „Running in place“
  - Use observations more than once by iterating several times over same assimilation window using the no cost smoother until „convergence“
- „Outer loop“ in LETKF (Yang and Kalnay, MWR, submitted)
  - Bring analysis mean closer to observations by iterative integration of ensemble mean from time n-1 to n (re-centering)
  - Advance to next assimilation time by integration of whole ensemble to time n+1
- Have proven to improve LETKF in presence of nonlinearity / non-Gaussianity in Lorenz model