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

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

Contributions / input by:
Hendrik Reich, Andreas Rhodin, Roland Potthast, Uli Blahak, Klaus Stephan, Africa Perianez, Michael Bender (DWD)
Annika Schomburg (DWD / Eumetat)
Yuefei Zeng, Dorit Epperlein (KIT Karlsruhe)
Daniel Leuenberger (MeteoSwiss)
Mikhail Tsyrlunikov, Vadim Gorin, Igor Mamay (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
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
• 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, … )
    – problems with observation selection solved
    – implementing ensemble scores (reliability, ROC, Brier Skill Score, (continuous) Ranked Probability Score )
    – main part of documentation written
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
    \[
    \langle (y - H(x_b))(y - H(x_b))^T \rangle = R + \rho HP_b H^T
    \]
    → pre-specified \( R \) is used for adaptive \( \rho \):
    → need for careful specification / tuning of obs errors

- observation error covariance \( R \):
  also estimate adaptively (Li, Kalnay, Miyoshi, QJRMS 2009)
  \[
  \langle (y - H(x_a))(y - H(x_b))^T \rangle = R
  \]
Task 3: scientific issues & refinement of LETKF

- adaptive observation errors

\[
\left\langle (y - H(x_a))(y - H(x_b))^T \right\rangle = R
\]

(in observation space)

- adaptive $R$ in ensemble space:
  
  adjusts total weight, not relative weight of obs

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

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

within localisation scale

\[ \text{f.g. mean} = 0.5 \, B_1 + 0.5 \, B_2 \]

\[ \text{ana} \approx \text{f.g.mean} + (B_2 - B_1) \]
\[ = 1.5 \, B_2 - 0.5 \, B_1 \]  
(1 perfect obs)

2 obs $\rightarrow$ least square fit

add obs, if already $N_{\text{obs}} > N_{\text{ens}}$:
- cannot be fitted well, improve analysis only slightly
- decrease analysis error!

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

$\rightarrow$ adaptive $\mathbf{R}$ takes that into account and increases $\mathbf{R}$
however: large $N_{obs}$: adaptive increase of $R$ indicates non-optimal use of obs

\[ f.g. \text{ mean } = 0.5 B_1 + 0.5 B_2 \]

\[ N_{obs} > N_{ens} \rightarrow \text{ least square fit} \]

$\implies$ localisation ! $\rightarrow$ see also Hendrik’s talk !

(or data selection / superobbing ?)

$\implies$ basic idea for adaptive localisation: keep $N_{obs}$ constant ($> N_{ens}$, not $>> N_{ens}$) !
LETKF, preliminary results: horizontal localisation

Caspari-Cohn function: scale $s = 100 \text{ km}$

$\rightarrow 0.4 \text{ at } r \approx (2)^{\frac{1}{2}} \cdot s \approx 141 \text{ km}$

$\rightarrow 0 \text{ at } r = 2 \cdot (10/3)^{\frac{1}{2}} \cdot s \approx 365 \text{ km}$

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

sum of localisation weights of obs

$\approx$ effective number of obs $N_{\text{eff,obs}}$

$\rightarrow N_{\text{eff,obs}} >> N_{\text{ens}}$

$\rightarrow$ too few degrees of freedom in order to fit the observations
LETKF, preliminary results: adaptive horizontal localisation

Caspari-Cohn function: scale $s = 50$ km

vertical cross section
(at rot lat = $2^\circ$, 8 Aug 2009, 12 UTC)

$N_{\text{eff,obs}} \rightarrow$ adaptive scale $s$:
appt $s$ such that $N_{\text{eff,obs}} \approx 70$
and $30$ km $\leq s \leq 80$ km

effective number of obs $N_{\text{eff,obs}}$

horizontal localisation scale $s$
LETKF, preliminary results: adaptive horizontal localisation

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

$s = 50$ km adaptive $s$

$\rightarrow$ smaller spread
$\rightarrow$ mostly smaller RMSE
(mixed results in verif vs. upper-air obs (T pos., wind neg.) )
LETKF: account of model error / additive inflation

- **parameterisation of model error using statistics** (Tsyrlunikov, Gorin):
  - parameterisation: \( e = \mu \cdot F_{phys}(x) + e_{add} \)
  - 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: 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)
  – basic Buizza version running, occas. crashed if microphysics tendencies perturbed
    → 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  ?
• **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
     (e.g. 4/3 earth model for beam propagation)

2. assimilation experiments
   - technical work (feedback files)
   - 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 (Michael Bender, Erdem Altuntac)
  - 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)
      - adjoint (sensitivities needed for tomography)
      - implement complex obs operator with ray tracer
      - monitoring, test e.g. impact of straight line approximation
    - then implement obs operator in COSMO (in 2013)
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
   (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)

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 $CTH_{\text{obs}}$, 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 $CTH_{\text{obs}}$ but have a cloud or even only high humidity close to $CTH_{\text{obs}}$

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

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

- use $f(RH_{\text{obs}}) - f(RH_k) = 1 - f(RH_k)$
and $CTH_{\text{obs}} - h_k$
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

Here: results of \textit{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

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

→ LETKF draws model cloud tops closer to obs

next:
- detailed evaluation (cross section, profiles...)
- single observation experiments
- tuning of observation error, thinning, localization
thank you for your attention