Regional Cooperation for
Limited Area Modeling in Central Europe

LACE
nwp central europe

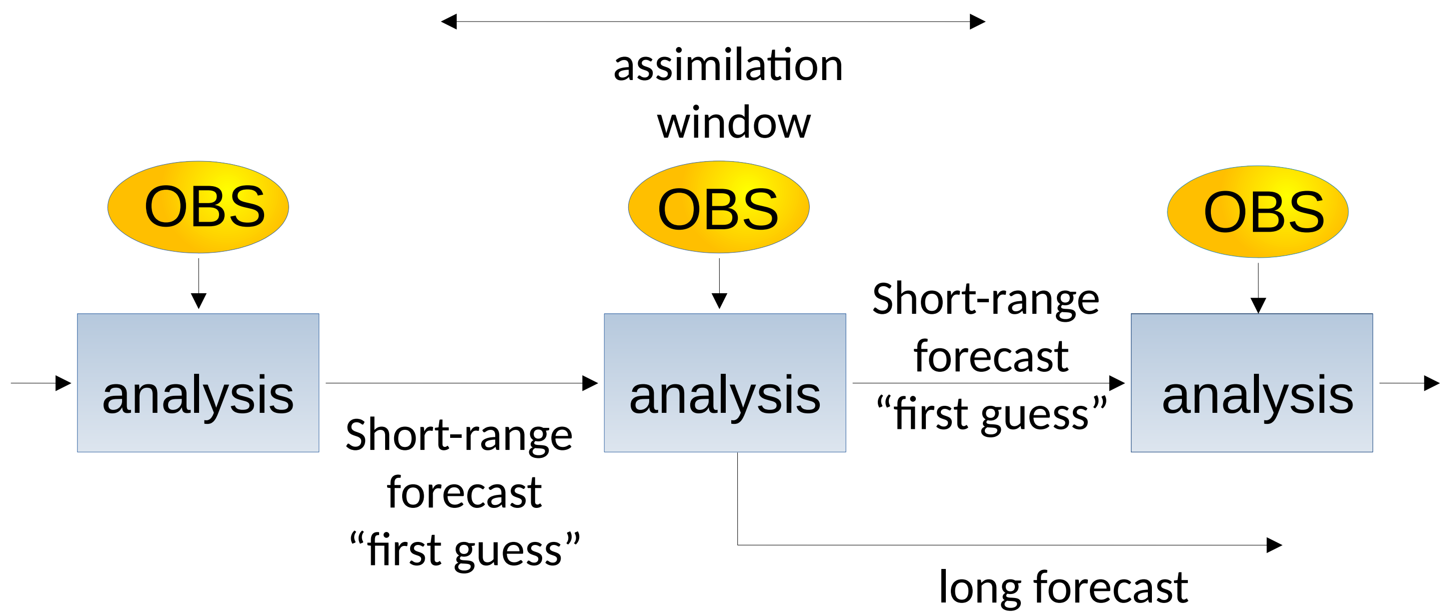
Diagnostics in data assimilation

Benedikt Strajnar

DHMZ ZAMG ROMANIA ANM

Contents

- 1) Analysis of DA residuals
- 2) Relative strength of obs. in analysis (DFS)
- 3) Relative impact of obs. on forecast (MTEN)
- 4) Covariances of residuals: tuning of DA



Why diagnostics?

- Data assimilation: combines “a priori” information with observations

$$J = (x - x_b)^T B^{-1} (x - x_b) + (y - H(x))^T R^{-1} (y - H(x))$$

- Ingredients: previous forecast, observations, data assimilation method, *statistical properties of errors*
- Aims of diagnostics in DA:
  - Validity of data assimilation method
  - Validity of assumptions about errors
  - Impact of (groups of) obs. on analysis and forecast
  - Statistical quality control of new observations

1) Analysis of DA residuals

- OMG (fg_depar): $y - H(x_b)$
  - Depends on quality of both first guess and observations
- OMA (an_depar): $y - H(x_a)$
  - Depends on DA method and its tuning, not a measure of quality
- OMA < OMB if assimilation converged
- Mean OMA and OMG are measures of systematic errors in the system (bias)

Residuals in DA

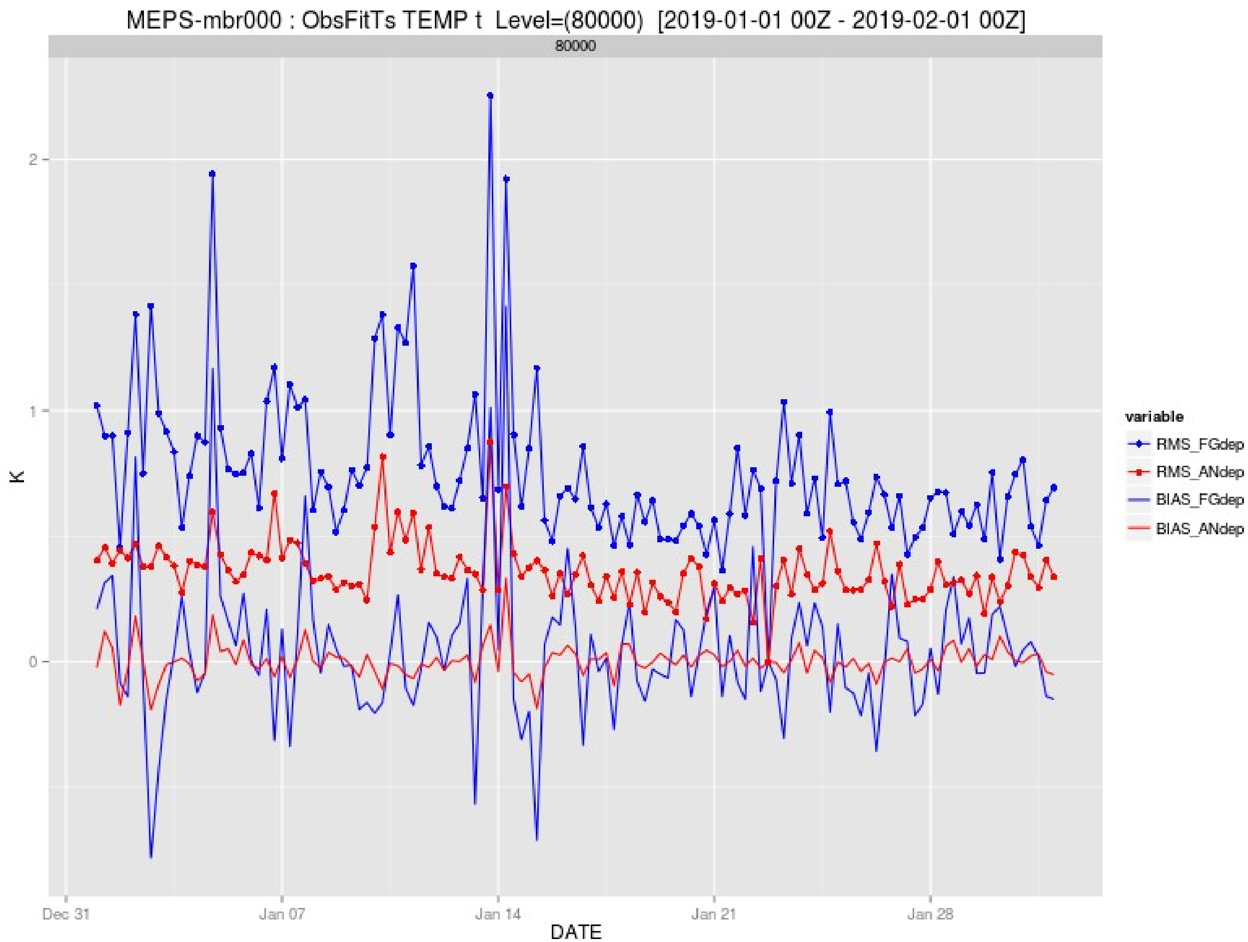
- Obtained by a simple ODB query

  odbsql -q 'select an_depar, fg_depar, obsvalue from hdr, body'

  from CCMA once minimization has finished
- ECMA: after screening, an_depar = fg_depar

Monitoring of active observations

- Does analysis reduce the OMG rms and bias?
- How close the analysis fits observations?

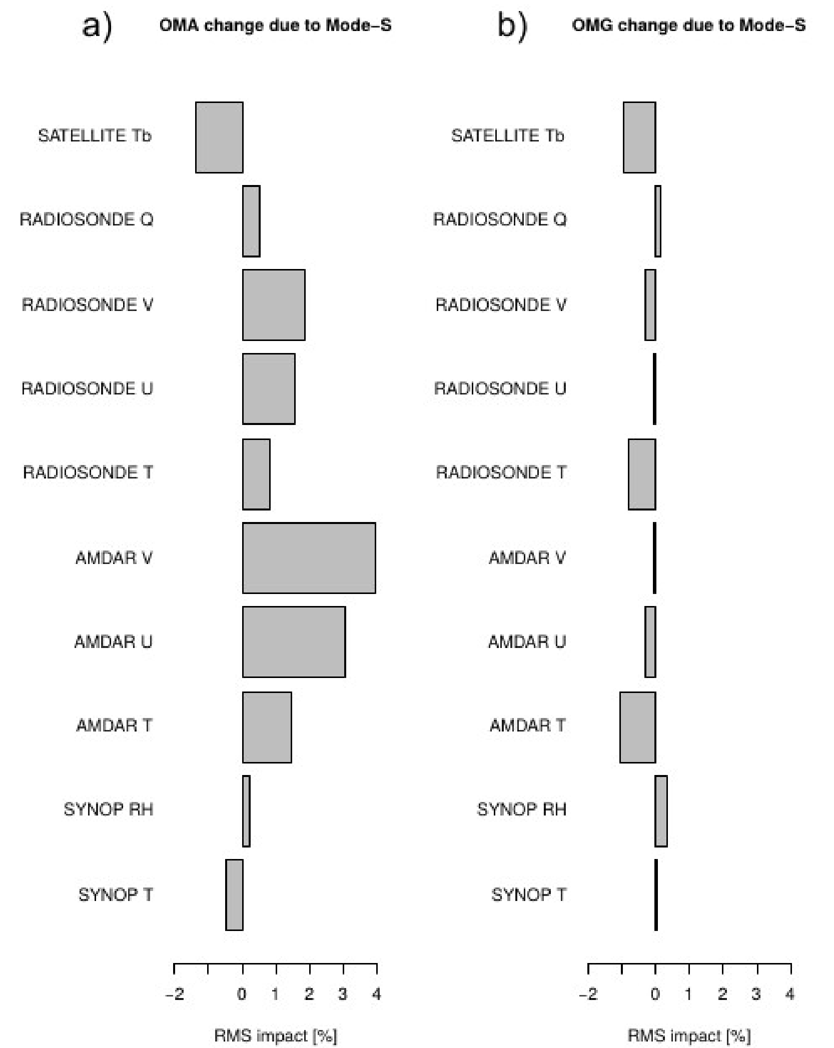


Residuals in active DA

- Relative change in OMA and OMG for a new observation set tells us about its impact.

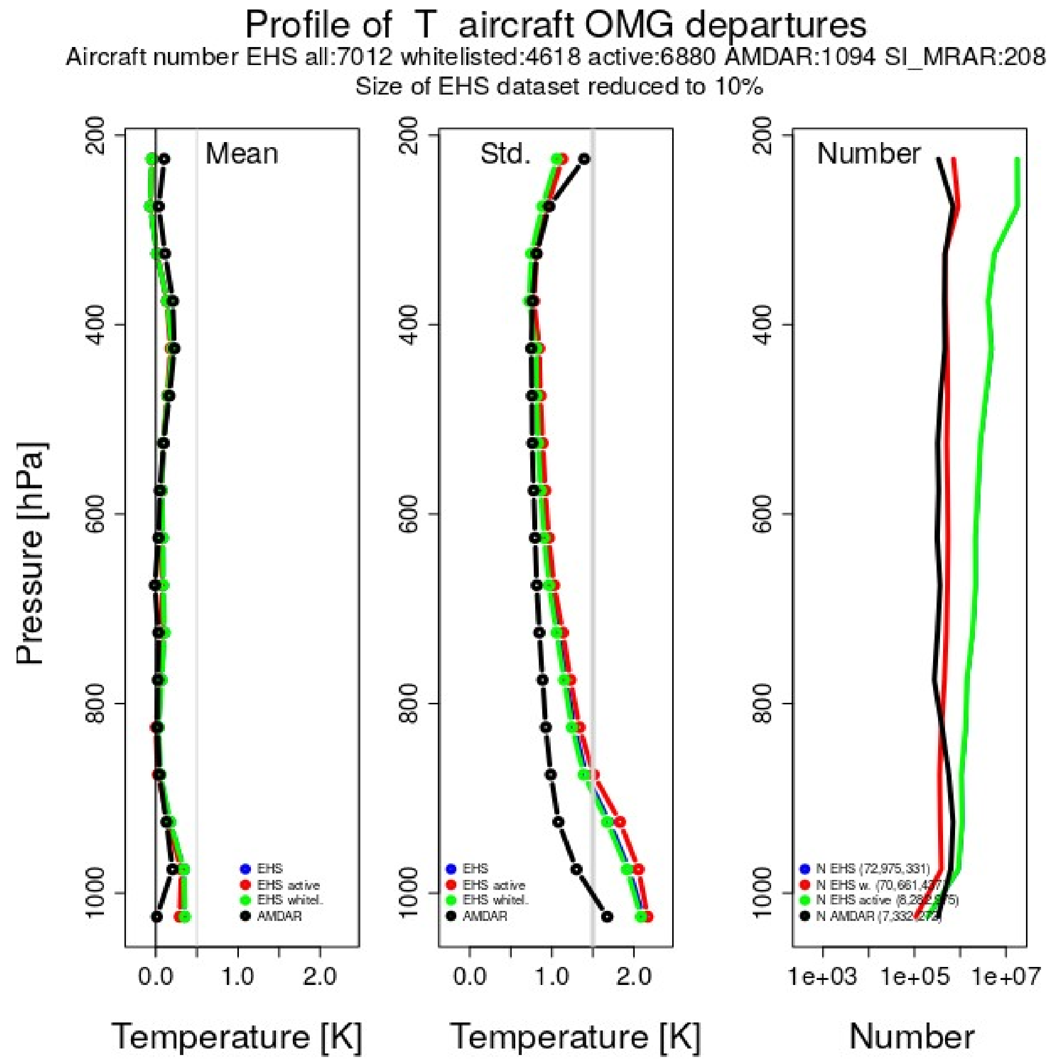
Left: Relative change of OMA because of assimilated new observations

Right: Relative change of OMG because of assimilated new observations



Passive monitoring in DA – example 1

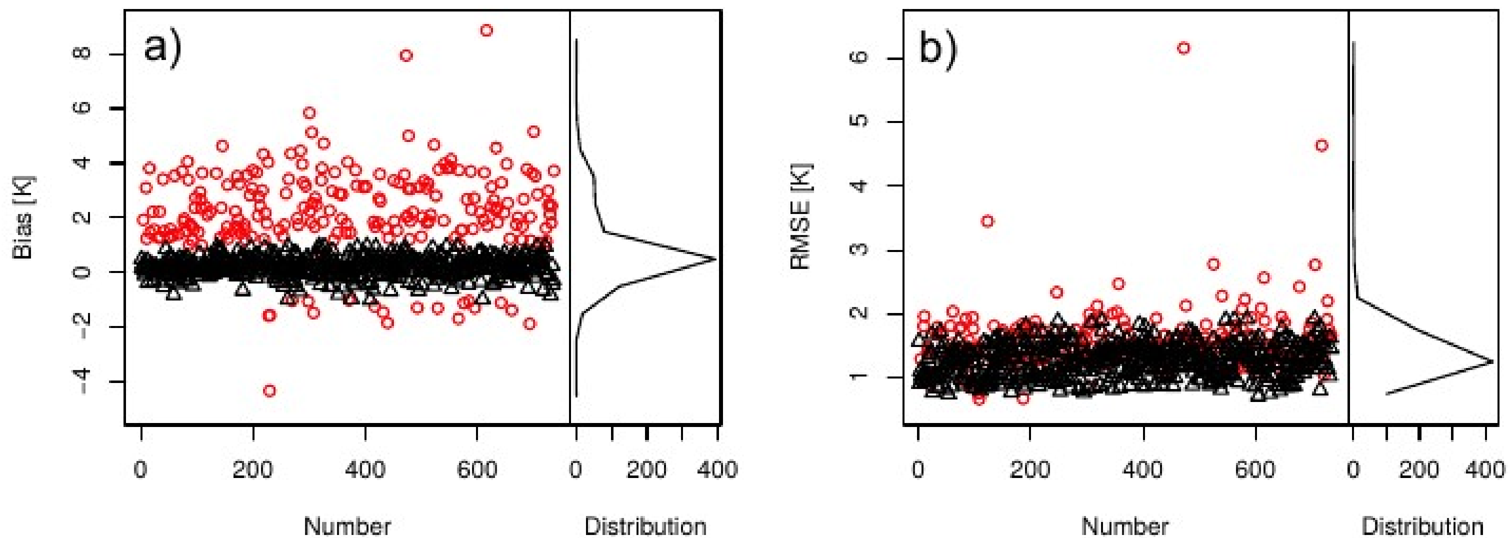
- If OMG is accumulated over long time for different observations, its relative quality can be estimated



Mean and std. of passive OMG for different aircraft observation groups

Passive monitoring in DA – example 2

- Long term OMG can be used to detect erroneous observations (instrument errors) – beware of other errors contained in R!



Mean and std. of passive temperature **OMG** for different airplanes.

Degrees of freedom for signal (DFS)

- Question: What is the strength of i-th obs. group in analysis:

$$DFS_i = \sum_{y \in obs} \left(\frac{\partial H_i x_a}{\partial y_i} \right)$$

$$DFS_i = Tr(\mathbf{KH})_i$$

- As there is no explicit K in the variational assimilation, we apply „Monte Carlo“ approach

$$\partial y'^T \mathbf{HK} \partial y' = Tr(\mathbf{HK} y'^T \partial y') = Tr(\mathbf{HK})$$

Cardinali et al. (2004), Chapnik et al. (2006)

Degrees of freedom for signal (DFS)

- If one sets

$$y' = y + R^{1/2}\partial y'$$

the trace can be computed by two analyses x_a, x_a' using y, y' observation sets.

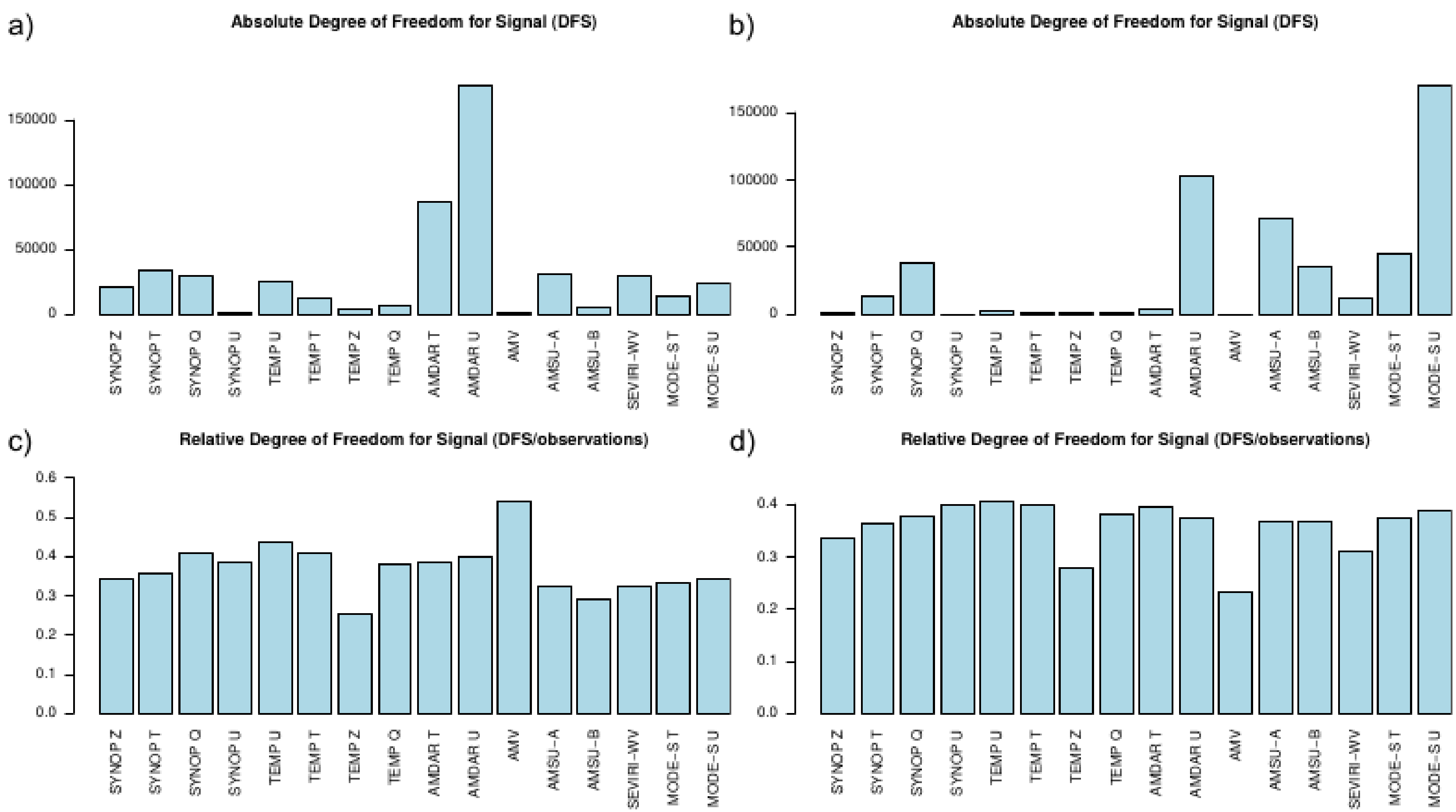
$$\mathrm{Tr}(\mathbf{KH}) = (y' - y)\, \mathbf{R}^{-1}\mathbf{H}(x_a' - x_a)$$

- By applying to what is available from the ODB we get:

$$\mathrm{DFS} = (\mathrm{OMG'} - \mathrm{OMG})^T \mathbf{R}^{-1}(OMA' - OMA)$$

- DFS can then be computed by observation groups.

DFS - interpretation



DFS- how to perturb observations?

- Program PERTCMA:
  - Copy/duplicate the existing unperturbed CCMA

```
${mpiexec} ./PERTCMA -N CCMA
```

  - N is seed for random number generator (int)
  - Adds perturbations with zero mean and sigma = sqrt(**R**) to all observations
  - Can also be executed inline in screening
    - NAMSCC: LPERTURB=T

DFS - computation?

- Predefined query to ODB:

```
odbsql -q 'select obstype, codetype, statid, varno, vertco_reference_1, degrees(lat), degrees(lon), an_depar, tdiff(date,time,andate,antime)/60, fg_depar, obsvalue from hdr,desc,body where varno/=91 and an_depar/="NULL" and obstype/=7' | grep -v obsvalue
```

- Program dfscomp (by A. Storto)

```
./dfstot.x pert_CCMA unpert_CCMA
```

- Can be then aggregated over the period, variable, observation group etc.

2. Moist total energy norm (MTEN)

- Measures strenght of influence of observation group on forecast

$$\mathrm{J}^i(t) =< x_t^i - x_t^{ctrl}, x_t^i - x_t^{ctrl} >$$

Where J is an arbitrary norm.

- MTEN norm combines all main meteorological variables:

$$\mathrm{J}_i = \int_{p_0}^{p_1} \int_D (u^2 + v^2 + \frac{c_p}{T_r} T^2 + \frac{R T_r}{p_r^2} p^2 + \frac{L^2}{c_p T_r} q^2) \frac{\partial p_r}{\partial p} dp dD$$

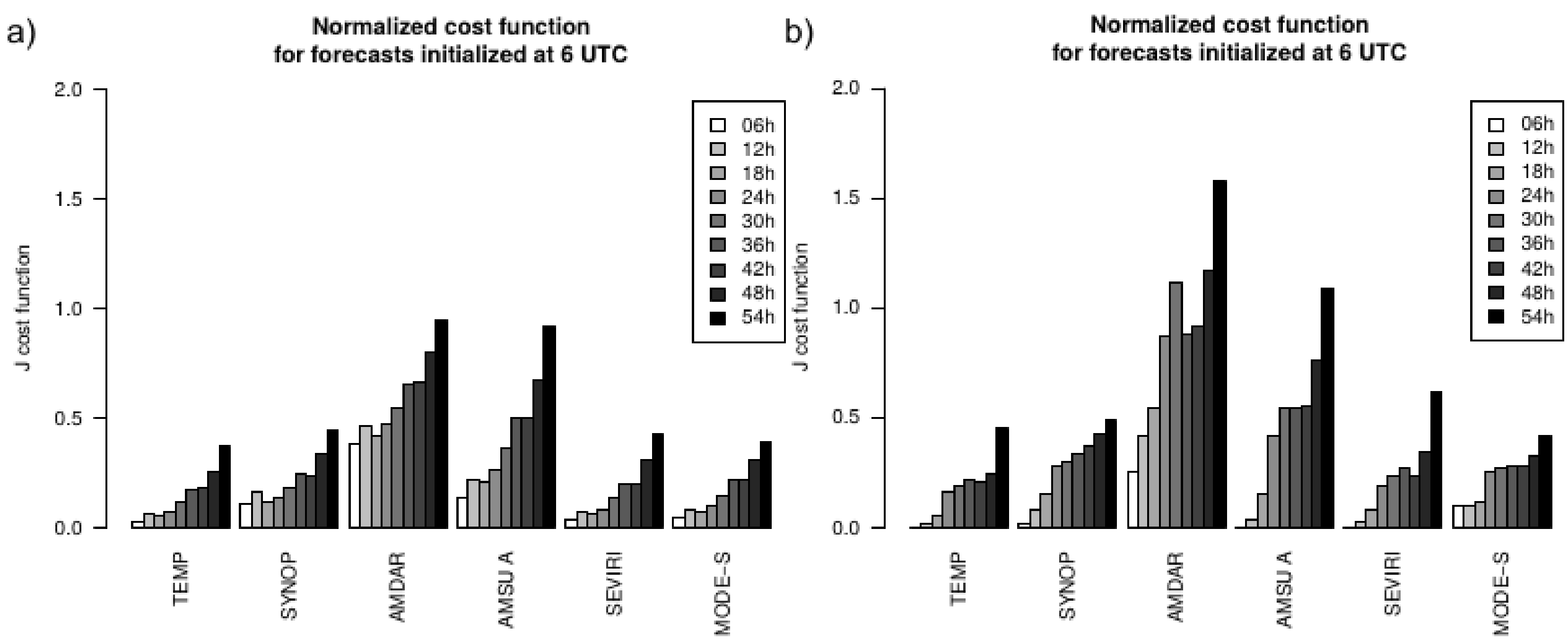
A. Storto, R. Randriamampianina (2010)

MTEN - computation

- Package on the LACE forum: http://www.rclace.eu/forum/viewtopic.php?f=21&t=375&p=1564&hilit=mten#p1564
- A special executable needs to be compiled by replacing some of the model source routines and entry point
- The analysis step is repeated several times, each time one observation is omitted
- Additional sets of forecasts are generated, history files are saved
- MTEN uses history files (control and the one with omitted observations) to compute the norms.

Moist total energy norm (MTEN)

- Shows relative importance of observations over time



4) Covariances of residuals

- Linear estimation theory: $\mathrm{K} = \mathrm{BH}^T(HBH^T + R)^{-1}$
- From the definition of background departure:

$$\mathrm{OMG} = \mathrm{y} - \mathrm{H}(\mathrm{x}_b) = y - H(x_t) + H(x_t) - H(x_b) \simeq \epsilon_o - \mathbf{H}\epsilon_b$$

- Taking the expectation operator and assuming that **R** and **B** are not mutually correlated:

$$\begin{aligned} E[OMG(OMG^T)] &= E[\epsilon_o\epsilon_o^T] - E[\epsilon_o\epsilon_b^T]\mathbf{H}^T + \mathbf{H}E[\epsilon_b\epsilon_o^T] + \mathbf{H}E[\epsilon_b\epsilon_b^T]\mathbf{H}^T \\ &= E[\epsilon_o\epsilon_o^T] + \mathbf{H}E[\epsilon_b\epsilon_b^T]\mathbf{H}^T \\ &= \mathbf{R} + \mathbf{HBH}^T \end{aligned}$$

Desroziers et al., 2005

4) Covariances of residuals

- Similar derivation, using the residuals

$$OMG = y - H(x)$$

$$AMG = H(x_a) - H(x_b) = \mathbf{H}\partial x_a$$

$$OMA = y - H(x_a)$$

leads to further relations:

$$E[OMG(OMG)^T] = \mathbf{HBH}^T + \mathbf{R}$$

$$E[AMG(OMG)^T] = \mathbf{HBH}^T$$

$$E[OMA(OMG)^T] = \mathbf{R}$$

$$E[AMG(OMA)^T] = \mathbf{HAH}^T$$

Hollingsworth & Lonnberg method

- Based on covariances of OMG departures, binned by distance between observations
- **B** and **R** are separated by assumming no spatial correlation in R and by extrapolating OMG covariance to 0.

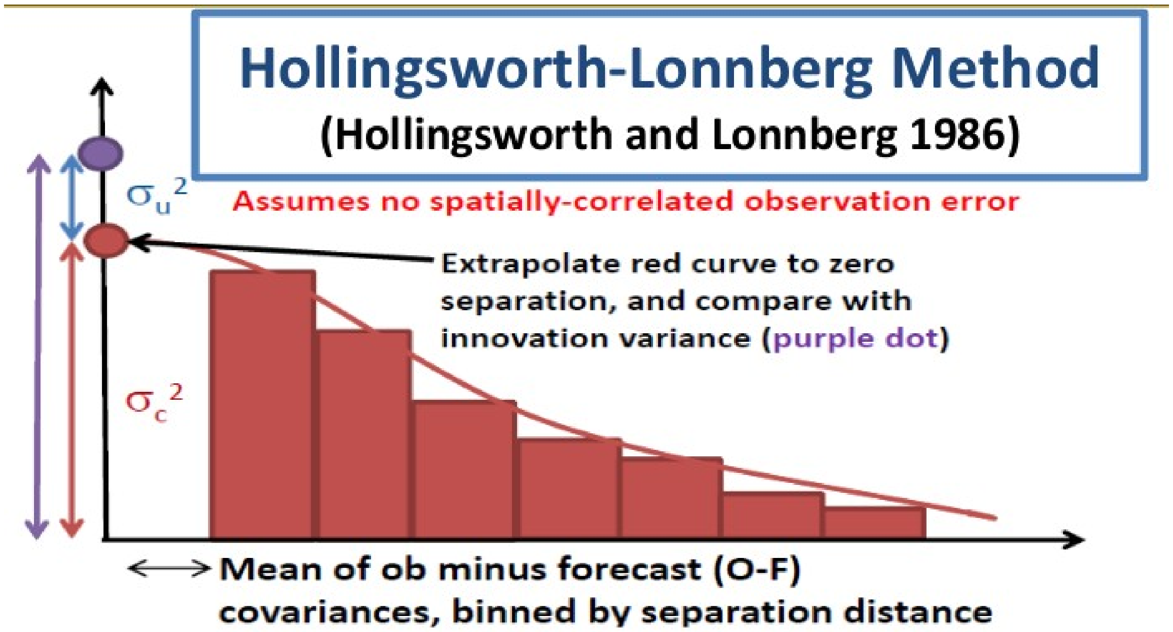


Figure by Campbell and Satterfield, NRL

A-posteriori tuning of **R** and **B**

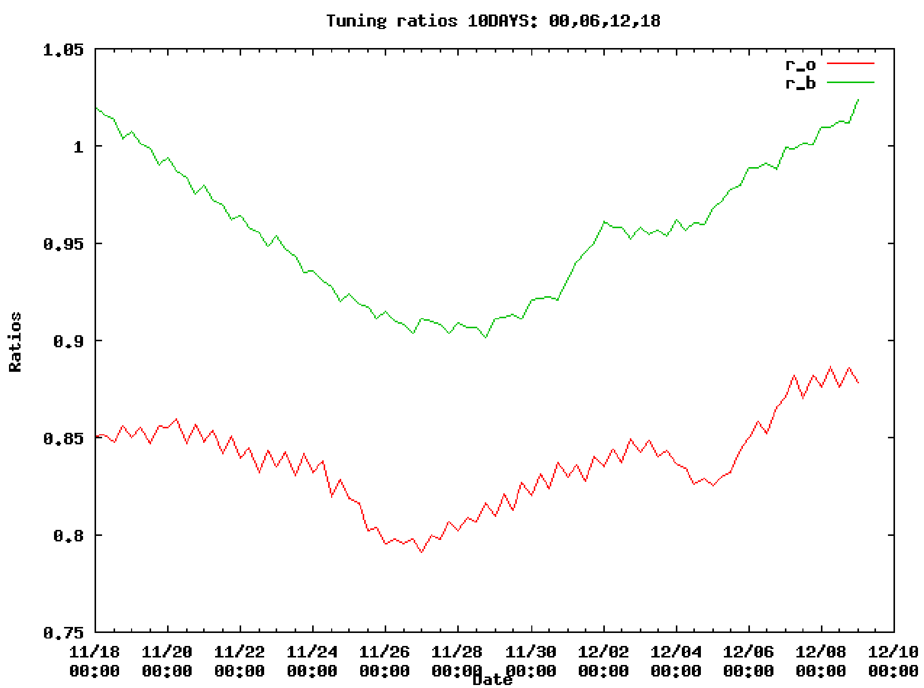
- The diagnosis is possible for any observation subset.
- If the diagnosed R and B are not in agreement with the predefined ones, we can use this method to tune them by *REDNMC* and *SIGMAO_COEFF*

$$B' = s\ B$$

- Covariances/correlations can be diagnosed (ObsTool in practicals).
- Tuning is an iterative process.

A-posteriori tuning of **R** and **B**

- Based on the ratio between predefined and diagnosed observation and background errors.
- Check impact on forecast as well!



Software for DA tuning

- Tune BR (G. Boloni)
  - http://www.rclace.eu/forum/viewtopic.php?f=30&t=248&hilit=TuneBR
  - Determines tuning factors for σ_o, σ_b
- OBSTool (P. Benacek)
  - Focuses on correlations in **B** (Desroziers and Hollingsworth&Lonnberg method) and **R**

Conclusions

- Diagnostic methods provide useful insights into assimilation regularity
- Enables tuning to reach more realistic balance between observations and first guess
- Typically influenced by many factors and depend on conditions which are not always met.
- Each diagnostic methods only focuses on a particular property of obs. impact.

Exercises

Exercise 1: DFS

- Run PERTOBS on the provided experiments.
- Preparation of perturbed CCMA ODB database

1.1) Run the script scr/prep_perturbed_ccma.sh

```
qsub scr/prep_perturbed_ccma.sh
```

This perturbs the original ODB database for 4 different cases.

Exercise 1: DFS

- Check the perturbed ODB and compare it with the unperturbed one

```
module load odb
odbsql -q "select obsvalue from hdr,body where obstype=1 and varno=39"  -o dump_perturbed.dat
```

- Exercise: Check that the perturbations really follow the Gaussian distribution with sigma = sqrt(R).

Exercise 1: DFS

- Run the DFS computation from the precomputed perturbed and unperturbed analysis departures.

  ./run.sh

- Check the outputs in results subdirectory and run the plotting script

  Rscript visualize_dfstot.R

  Rscript time_evolution

- Check the plots by using eog and display.

Exercise 1: DFS

- What is the main difference between the two experiments?
- Which obs. types and variables have the greatest impact on the analysis?
- Which obs. types are the most valuable for the analysis?

Exercise 2: ObsTool

- * Modify the DFS computation program in src/dfscomp.F90 in order to compute the total DFS for conventional (SYNOP,AMDAR,TEMP) and remote-sensed observations (AMSU,IASI,SCATT,AMV,..). Which observations have bigger impact and which are more valuable in the analysis?

Run the DFS computation from the precomputed perturbed and unperturbed analysis departures.

```
cd scr
./run.sh
```

2.1) Check the outputs in results subdirectory and run the plotting script

```
module load R
Rscript visualize_dfstot.R
Rscript time_evolution.R
```

Check the plots dfs_plot_total.ps, dfs_evolution_${EXPNAME}.png.

Exercise 2: ObsTool

- Run the obs. tool on experiment aos, period 2018080100 - 2018081006, analysis every 3 hour.
- You don't need to extract ODBs as this is done in advance.
- Examine the generated plots for SYNOP, AMDAR and TEMP.
- Determine the optimal thinning distance for AMDAR and SYNOP which needs to be applied if R is considered as spatially uncorrelated (usually we are safe at distance where correlation falls below 0.2).

Exercise 1: DFS

- Run PERTOBS on the files
- Use odbselect to read aircraft data and check that the observations were properly perturbed.
- 1. Preparation of perturbed CCMA ODB database
-
- 1.1) Run the script scr/prep_perturbed_ccma.sh
-
- qsub scr/prep_perturbed_ccma.sh
-

- This perturbs the original ODB database for 4

Exercise 3: ObsTool

- Run PERTOBS on the files
- Use odbselect to read aircraft data and check that the observations were properly perturbed.