LETKF at CNMCA: current set-up and comparison with 3D-VAR

Lucio Torrisi, Francesca Marcucci, Massimo Bonavita[1]

and contributions of

Antonio Vocino, Marco Alemanno, Alessandro Cheloni

CNMCA, National Meteorological Center,Italy

[1] *Current affiliation: ECMWF, Reading (UK)*

***COSMO** General Meeting, **Moscow** 6-9 September 2010*

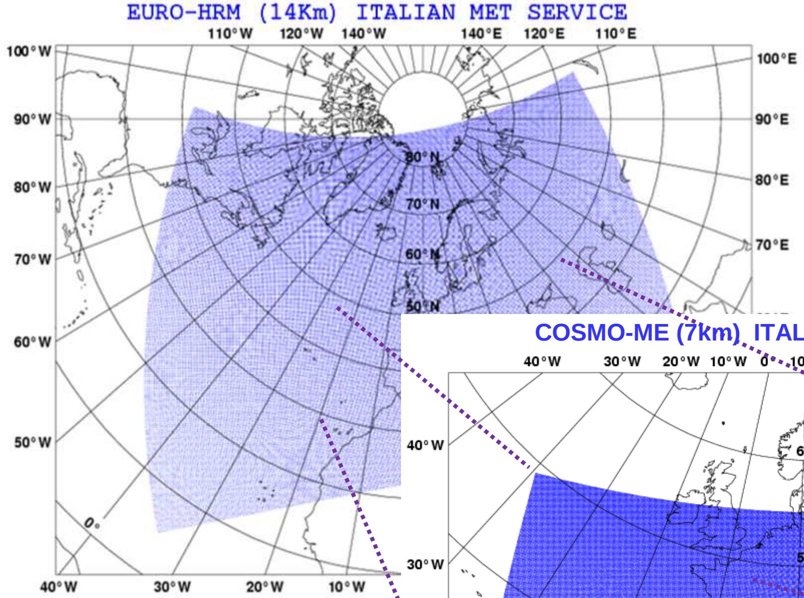
Outline

- CNMCA LETKF implementation
- Stochastic physics schemes
- Lateral boundary perturbation
- Radiances treatment
- Improvement of algorithm efficiency
- Handling non-linearities: outer loop
- Comparison with 3DVAR and IFS analysis
- Future developments

CNMCA NWP System

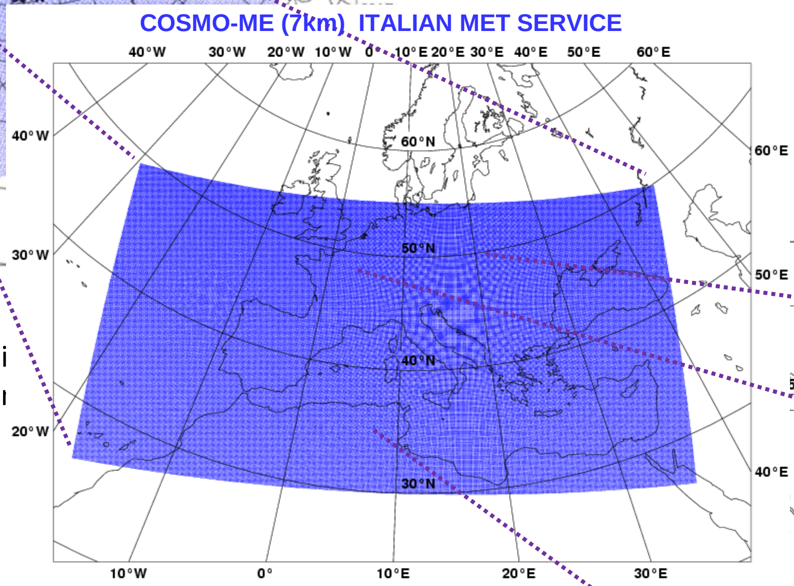
Data Assimilation:

3DVAR PSAS FGAT (T,u,v,qv,ps) every 3h using
TEMP, PILOT, SYNOP, SHIP, BUOY, Wind Profiler,
AMSUA rad., AMDAR-ACAR-AIREP, MSG/MODIS
AMV, METOP/QUIKSCAT/ERS2 scatt. winds
+ Land SAF snow mask,
IFS SST analysis once a day



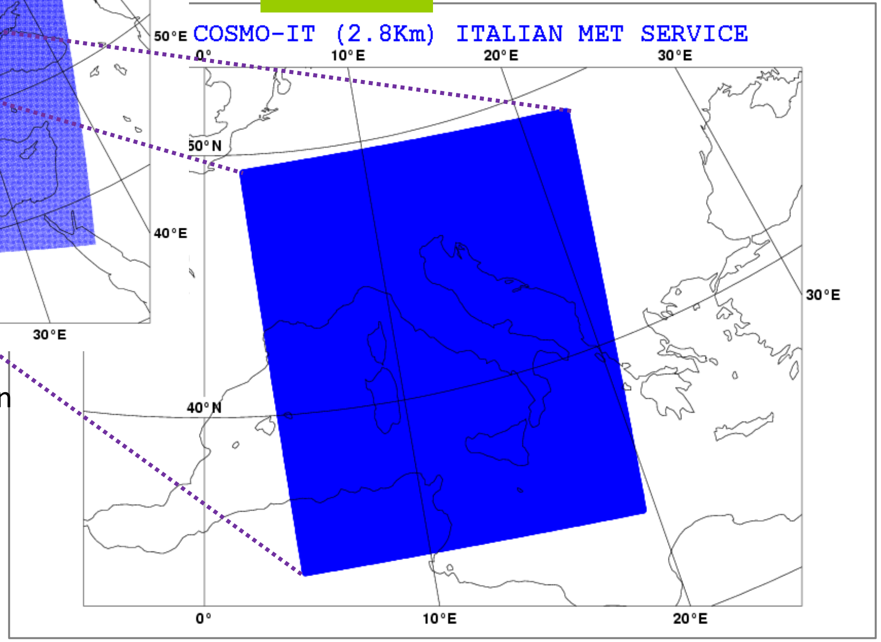
14 km
40 v.l.

- hydrostatic equati
- parameterized coı



7 km
40 v.l.

- compressible equations
- parameterized convection



2.8 km
50 v.l.

- compressible equations
- explicit convection

Local Area Modelling: COSMO

Ensemble Kalman Filter D.A.

Uses an ensemble of N system states to parametrize the distribution

$$P^b = \frac{1}{N-1} X^b X^{b^T} \qquad X^b = x^b - \overline{x}^b$$

It follows the time evolution of the mean and covariance (Gaussian assumption) by propagating the ensemble of states

LETKF FORMULATION (Hunt et al,2007)

$$\tilde{H}_n = H(x_n^b) - \overline{H}(x^b)$$

$$\tilde{P}^a = \left[(\tilde{H}^T R^{-1} \tilde{H} + (N-1)I\right]^{-1}$$

$$K = X^b \tilde{P}^a \tilde{H}^T R^{-1}$$

$$X^a = X^b W^a$$

$$W^a = \left[(N-1)\tilde{P}^a\right]^{1/2} \quad \text{(Square root filter)}$$

$$\overline{x}^a = \overline{x}^b + K(y^o - H(\overline{x}^b))$$

$$x^a = \overline{x}^a + X^a$$

The analysis ensemble mean is the linear combination of forecast ensemble states which best fits the observational dataset

CNMCA LETKF Implementation

(new Implementation)

- 40 member ensemble at 0.25° (~28Km) grid spacing (EU RO-HRM domain), 40 hybrid p-sigma vertical levels (top at 10 hPa)
- Initial ensemble from different EURO-HRM forecasts valid in ± 48h around start time and boundary conditions from IFS for all members (not perturbed)
- 6-hourly assimilation cycle run and (T,u,v,qv,ps) as a set of control variables
- Operational 3DVar cycle run in parallel at same spatial resolution
- Observations: RAOB, SYNOP(ps), SHIP(ps), BUOY(ps), AIREP, AMDAR, ACAR, AMV (MSG,MODIS), WPROF, SCAT(METOP,ERS2,QSCAT)
- Horizontal localization with 800 Km circular local patches (obs weight smoothly decay $\propto r^{-1}$)
- Vertical localization to layers whose depth increases from 0.2 scale heights at the lowest model levels to 2 scale heights at the model top
- 3D adaptive multiplicative (Li et al., 2009) + additive inflaction factor

More details in:

Bonavita M, Torrisi L, Marcucci F. 2008. The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations. *Q. J. R. Meteorol. Soc.* **134**: 1733-1744.

Bonavita M, Torrisi L, Marcucci F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q. J. R. Meteorol. Soc.* **136**: 132-145.

Stochastic Physics

- LETKF outperforms 3DVAR (Bonavita et al, 2010), but forecast runs initialized by LETKF have larger biases than those starting from 3DVAR → Necessity of bias correction
- Model uncertainty could be represented also with a stochastic physics scheme (Buizza et al, 1999; Palmer et al, 2009) implemented in the prognostic model
- This scheme perturbs physics tendencies by adding perturbations, which are proportional in amplitude to the unperturbed tendencies X_c:

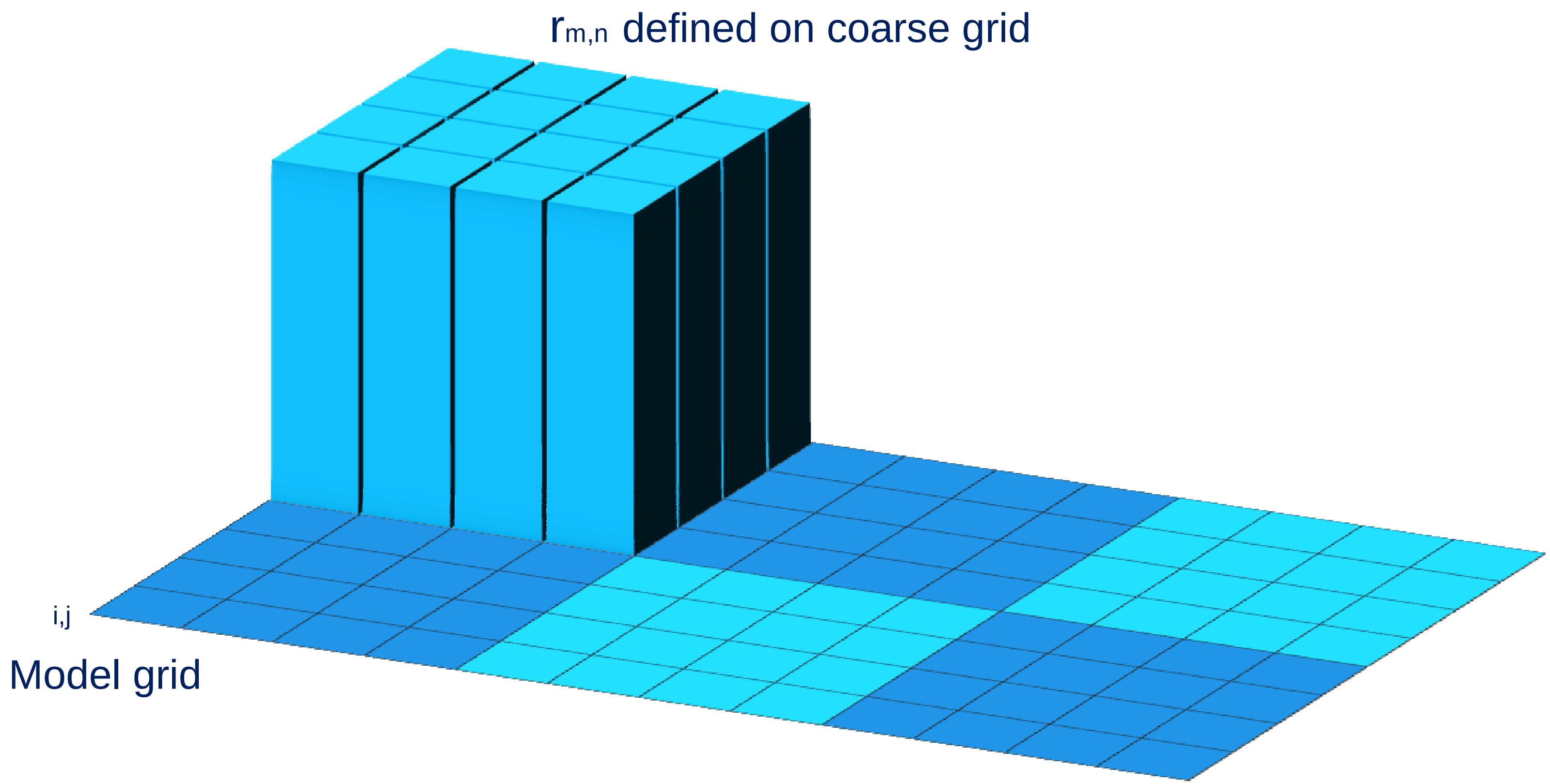
$$X_p=(1+r\mu)X_c$$

where r is a random number and μ is a tapering factor (μ=1 in Buizza et al, 1999)

Stochastic Physics

According to Buizza et al, 1999

Spatial correlation is imposed using the same r in a whole column and drawing r for a coarse grid with spacing DL (boxes)



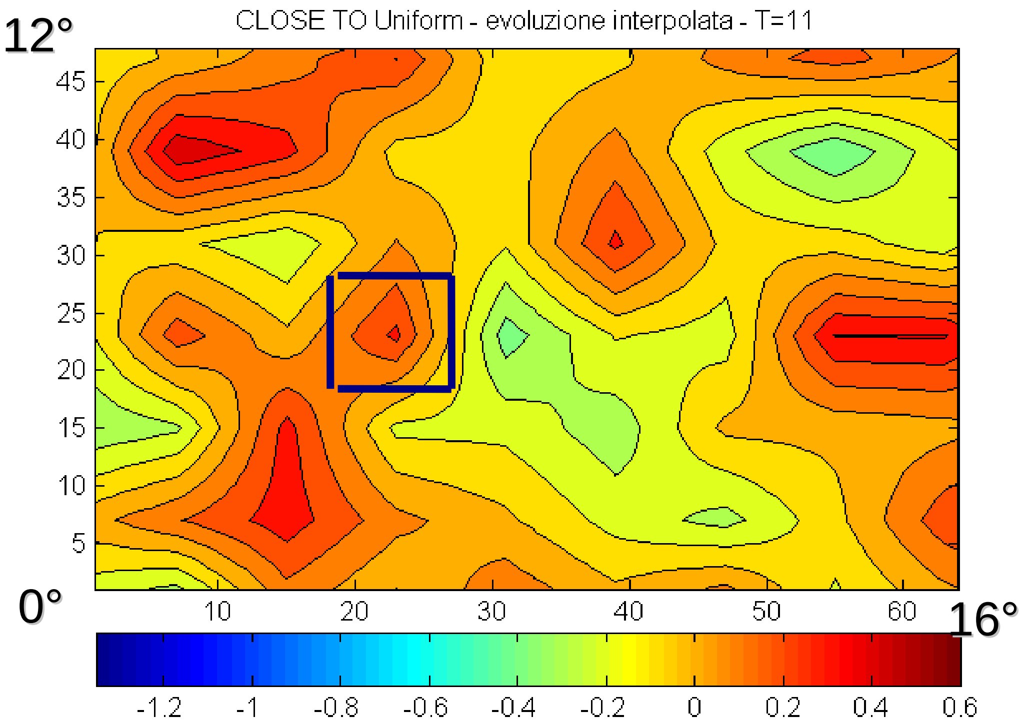
Modified to have a smoother pattern horizontally and to reduce the perturbation close to the surface and in the stratosphere

Stochastic Physics

Version 1

Toy model and plots by A. Cheloni

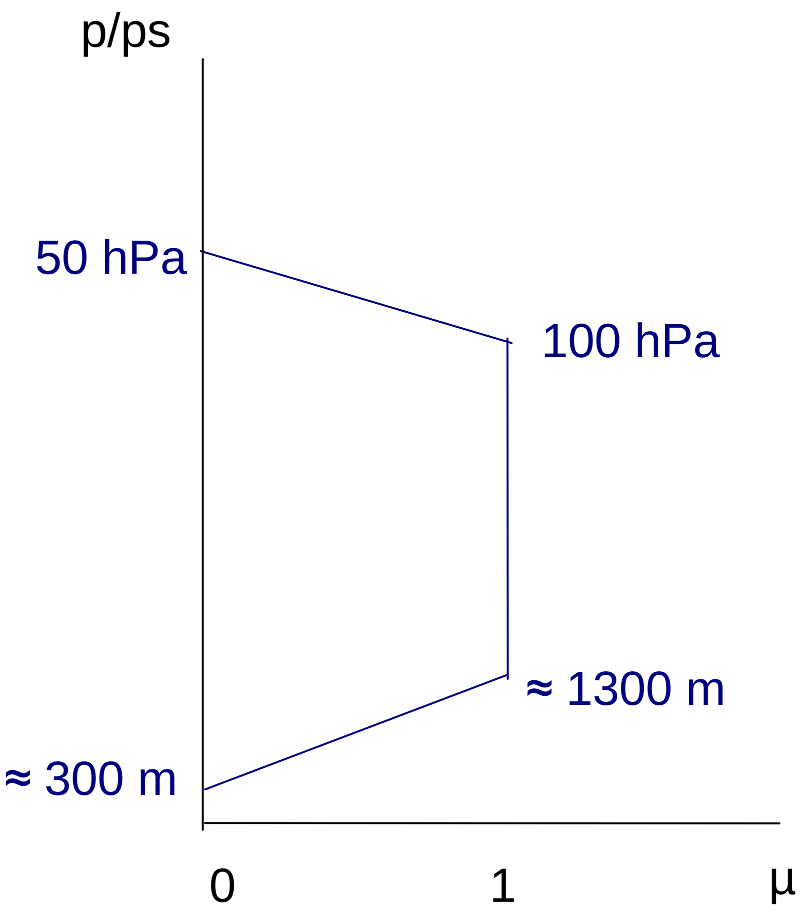
Model grid spacing: 0.25° (28 km)



2.5° coarse grid with bilin. interp.

Stochastic Physics

Version 1



According to Palmer et al, 2009:
"…. For reasons of numerical stability and physical realism, the perturbations have been tapered to zero in the lowermost atmosphere and in the stratosphere.

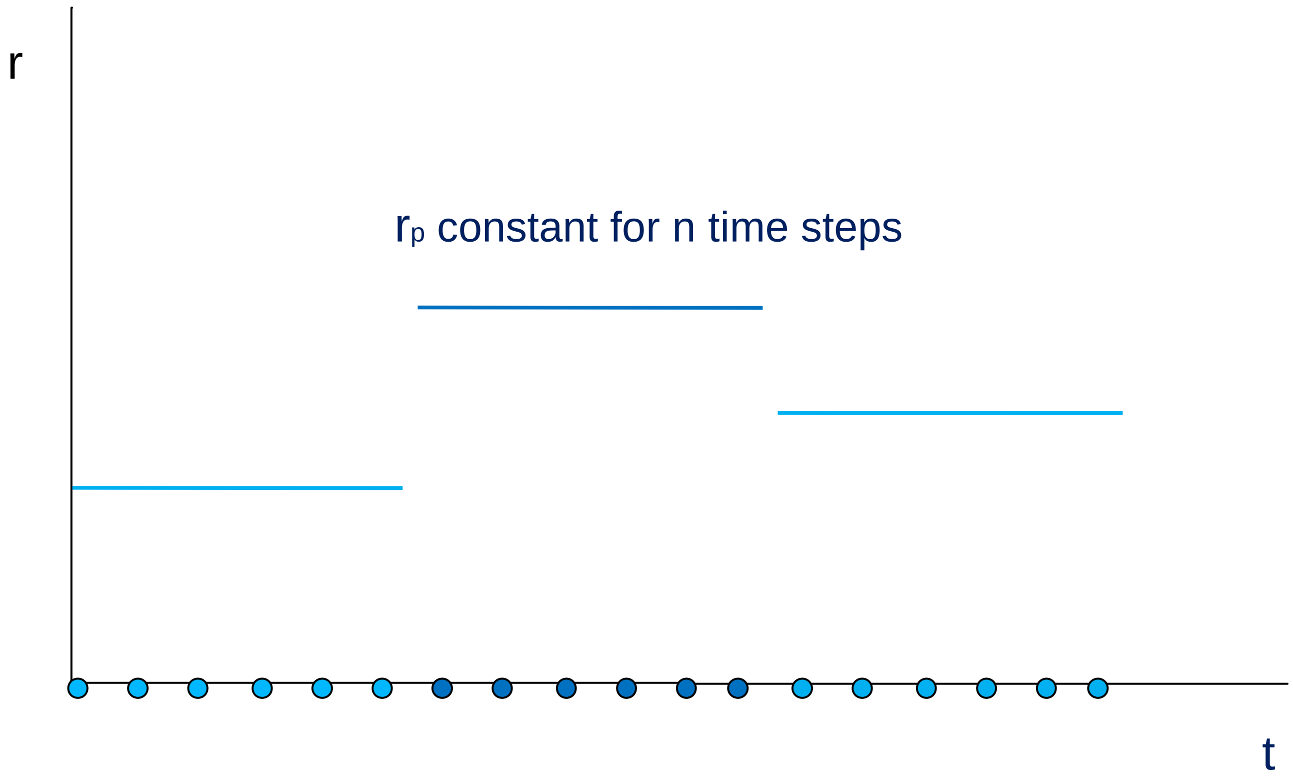
- In initial tests, tendencies were perturbed in the entire atmosphere. For standard deviations of 0.5, numerical instabilities were encountered. Further testing showed that the cause of the numerical instability are the perturbations in the lowermost part of the atmosphere.
- Radiative tendencies are expected to be relatively accurate in the stratosphere and with errors that are predominantly large scale, i.e. with correlation lengths far larger than 500 km….. "

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

Stochastic Physics

According to Buizza et al, 1999

Perturbations are multivariate (different r for u,v,T,qv). Temporal correlation is achieved by drawing r every n time steps (Dt)



Modified to have a univariate distribution (as in Palmer et al, 1999) and a smoother pattern in time

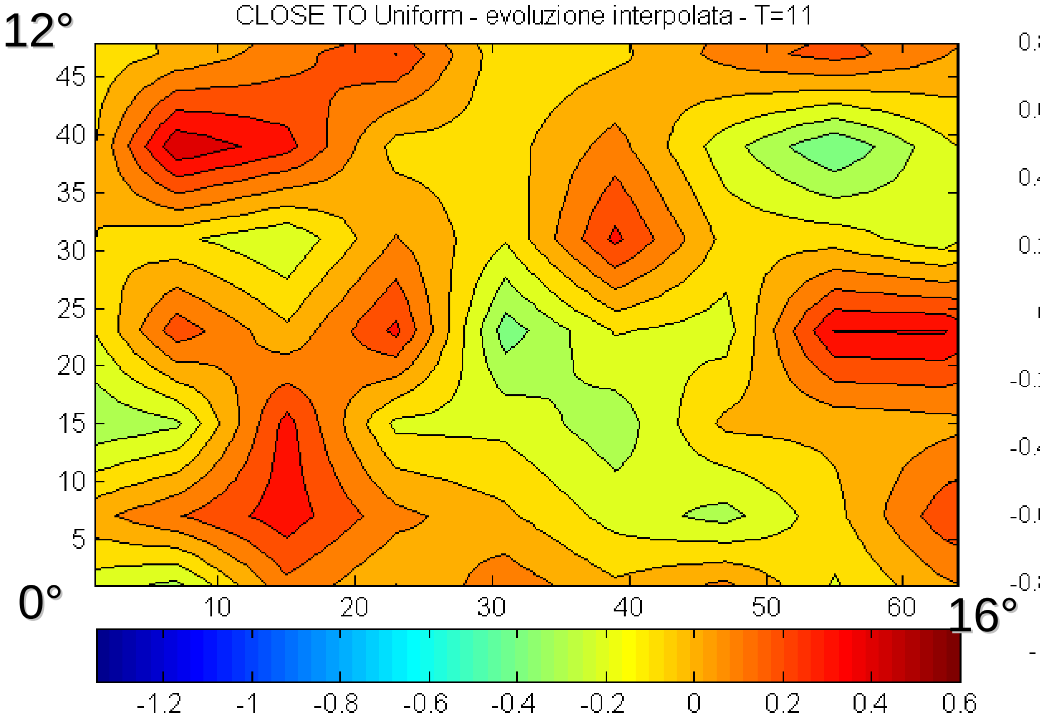
Stochastic Physics

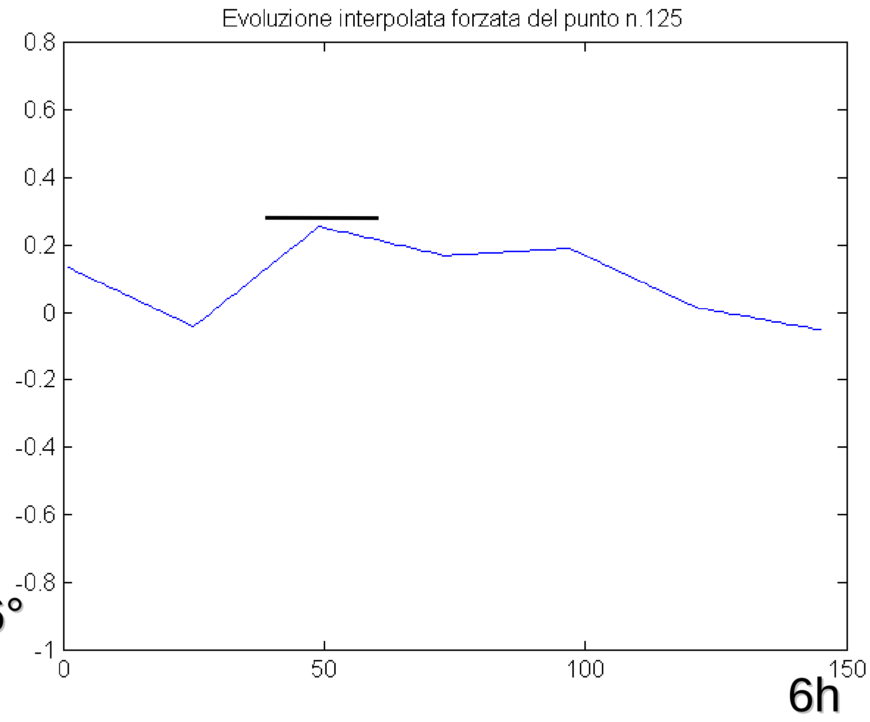
Version 1

Toy model and plots by A. Cheloni

Model grid spacing: 0.25° (28 km)

Time step: 150 s





2.5° coarse grid with bilin. interp.

1h coarse time grid with lin. interp.

Stochastic Physics 1

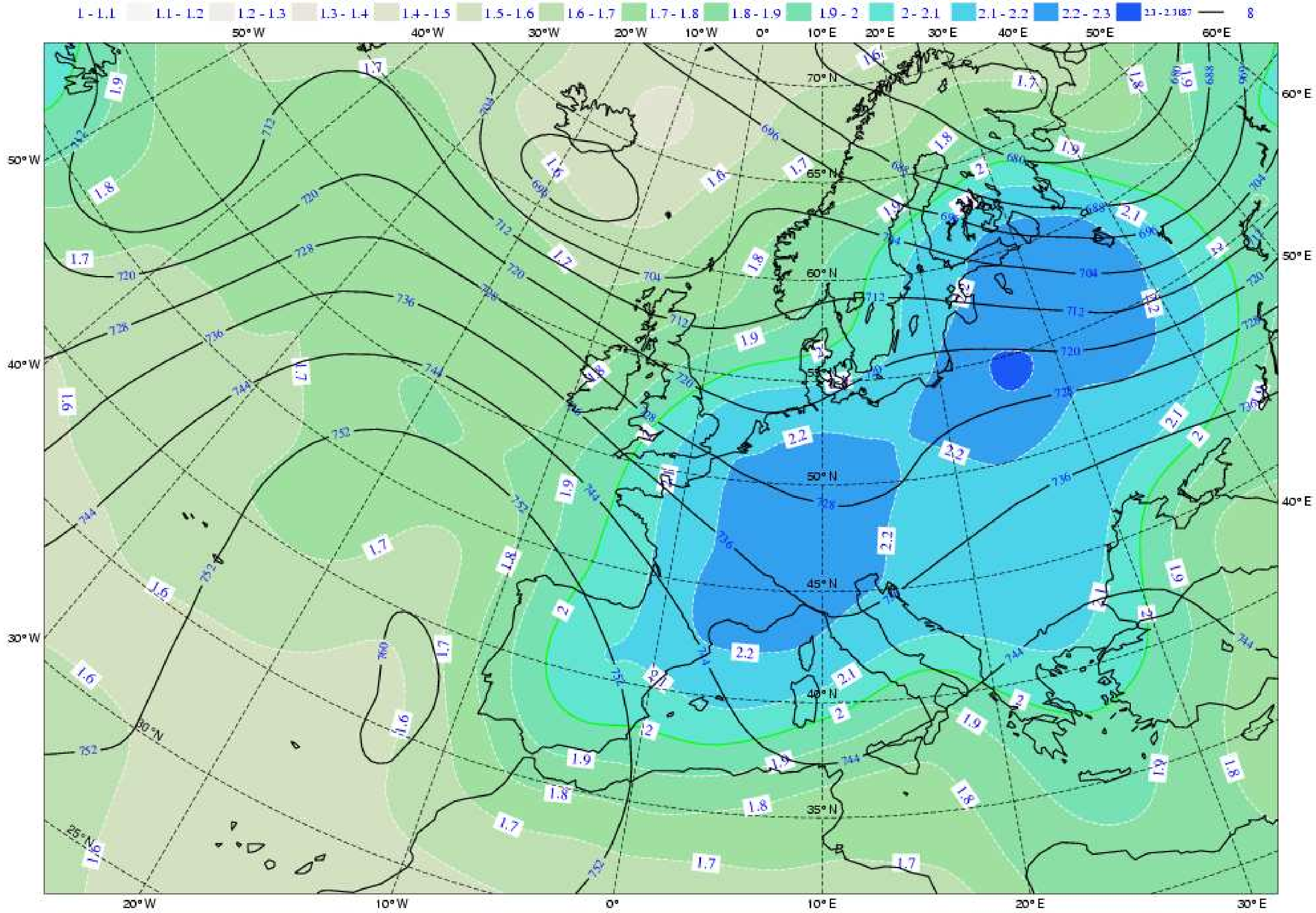
Version 1 (in blue, differences from Buizza et al, 1999)

- For all variables (u,v,T,qv), the random numbers r are drawn from a uniform distribution in the range [-0.5,0.5]
- 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 critical 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

No Stochastic Physics

Averaged from 20091009-18 to 20091011-06

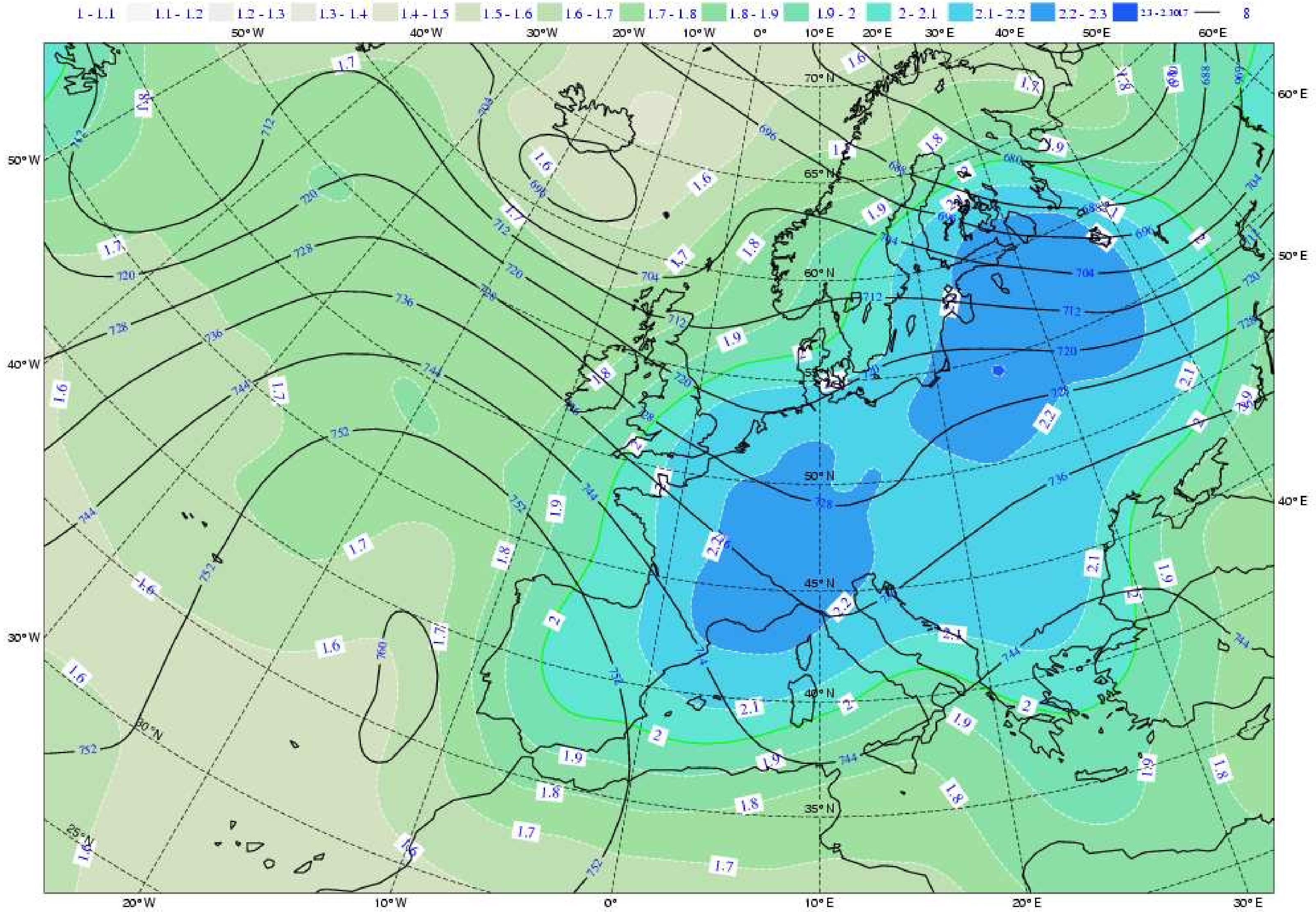
Multiplicative Inflaction Factor: Model level 14



Stochastic Physics

Version 1

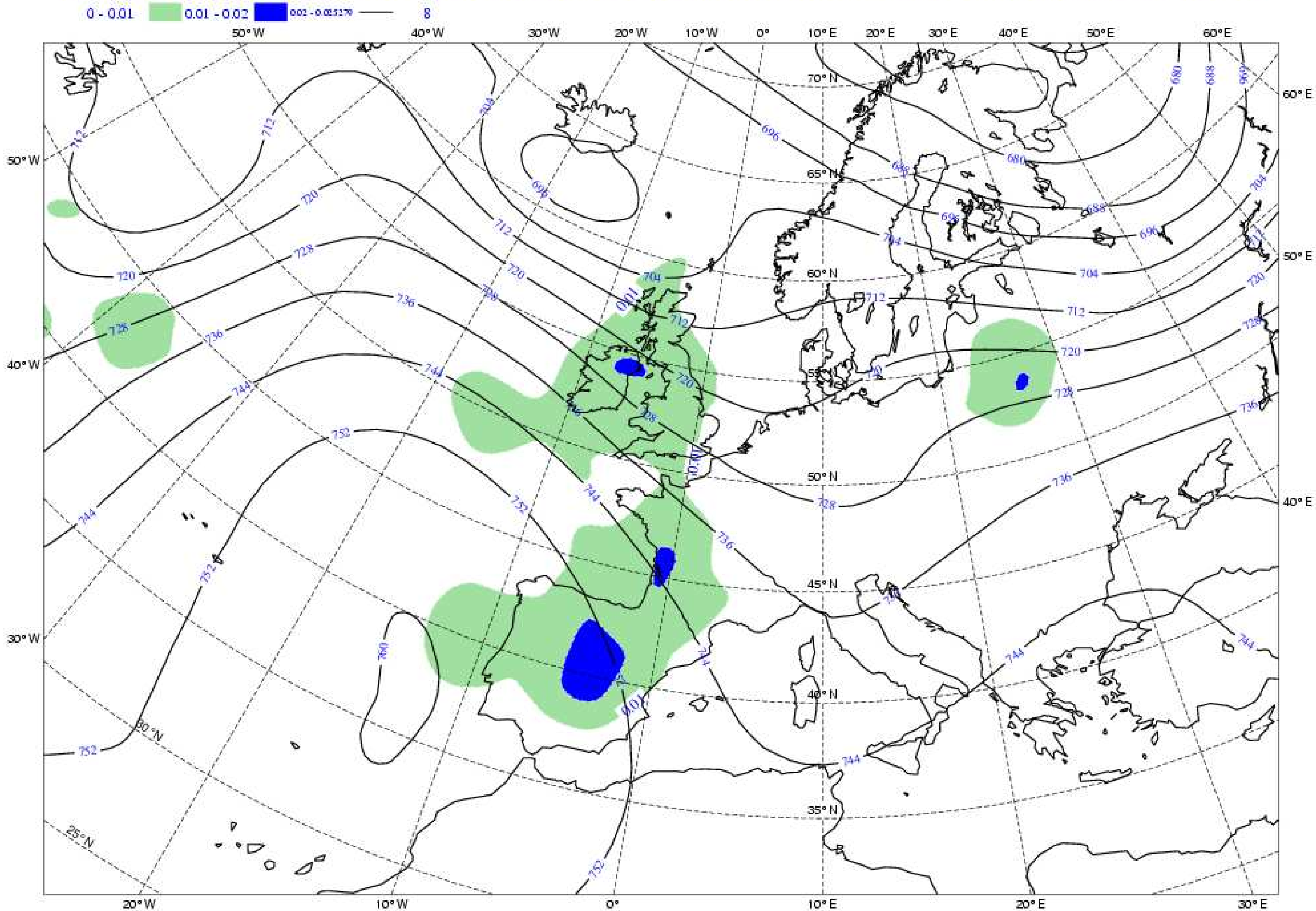
Multiplicative Inflaction Factor: Model level 14



Stochastic Physics

Version 1

Multiplicative Inflaction Factor Difference (NoStochP-StochP): Model level 14



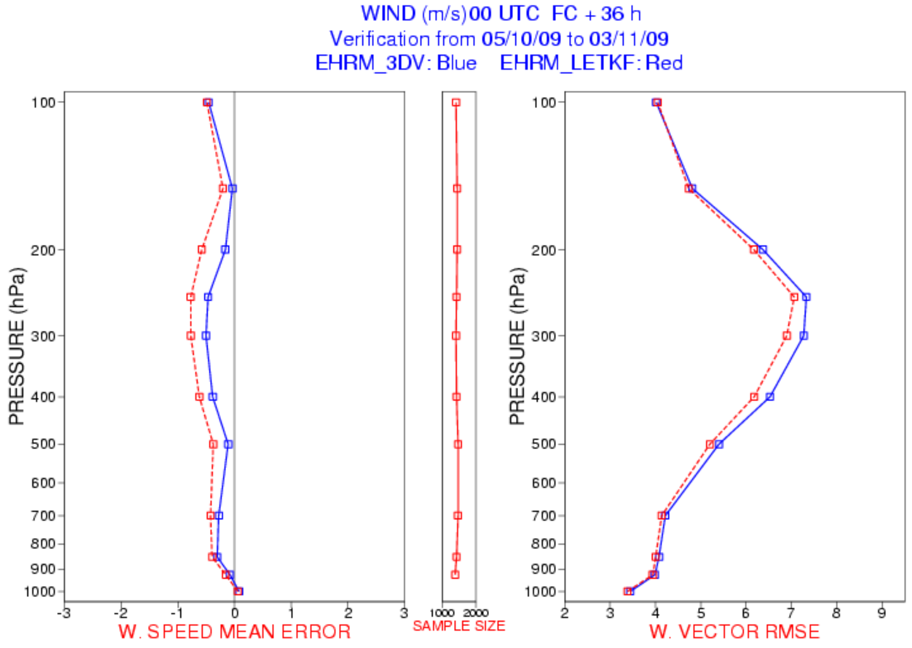
to 0.2.

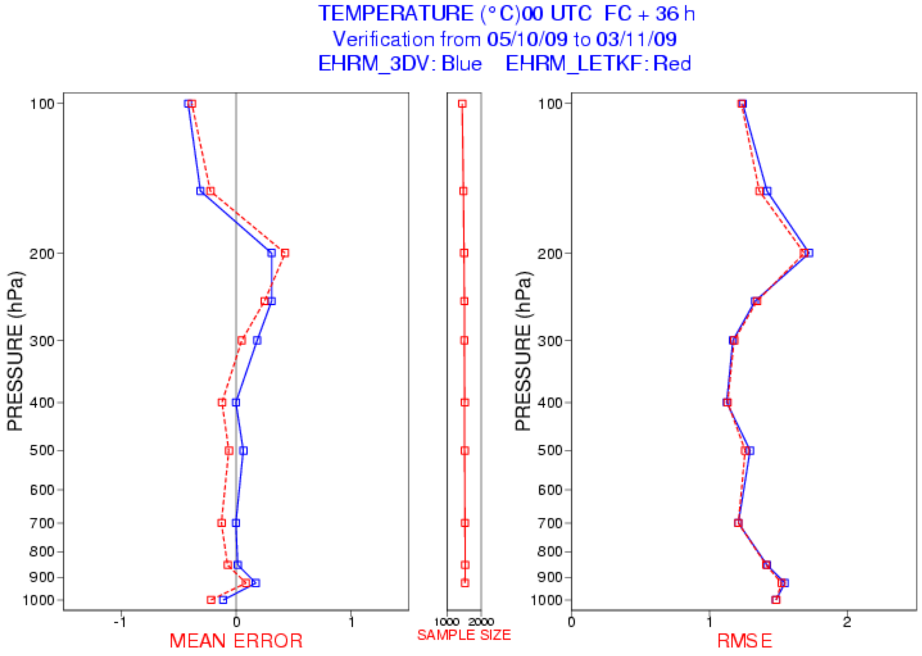
Modified CNMCA LETKF Setup

- The stochastic physics scheme (version 1) was included in the HRM model and some experiments were performed to tune the length and the time scale of random number pattern. 10°/5°/2.5° and 1/3/6 hours were tested. A very small improvement (especially in wind after T+24h) was found using the stochastic physics
- For a one month experiment 2.5° (5° similar) and 1h (3h similar) were chosen (5 Oct-3Nov 2009)
- The additive inflaction factor was reduced (scal.f. 0.5 to 0.2)
- SST perturbations derived from randomly selected, scaled consecutive analysis differences were daily inserted
- A daily blending of the mean upper level analysis with the IFS one was also introduced to compensate the limited satellite data usage of the system at that levels

3D-VAR / LETKF Comparison

Objective verification against
European radiounding stations





Stochastic Physics 2

Version 2 (similar to Palmer et al, 2009)

- For all variables (u,v,T,qv), the random numbers r are drawn for each time step from a distribution close to a gaussian one with stdv=0.5 (bounded to the range ± 3 stdv)
- r are described with auto-regressive processes of first order (AR1) with a decorrelation time scale τ forced by gaussian random numbers η with zero mean and unit variance

$$\mathbf{r(t+Dt)=\Phi\, r(t) + \sigma\, \eta(t)}$$

where $\Phi=\exp(-Dt/\tau)$ and $\sigma=stdv\,(1-\Phi^2)^{0.5}$

- The perturbations of T and qv are not applied if they lead to particular humidity values (exceeding the critical saturation value or negative)
- Spatial correlation is imposed using the same r in a whole column and drawing r for a coarse grid with spacing DL (similarly to the boxes of Buizza et al, 1999); then they are *bilinearly interpolated* on the finer grid to have a smooth pattern in space

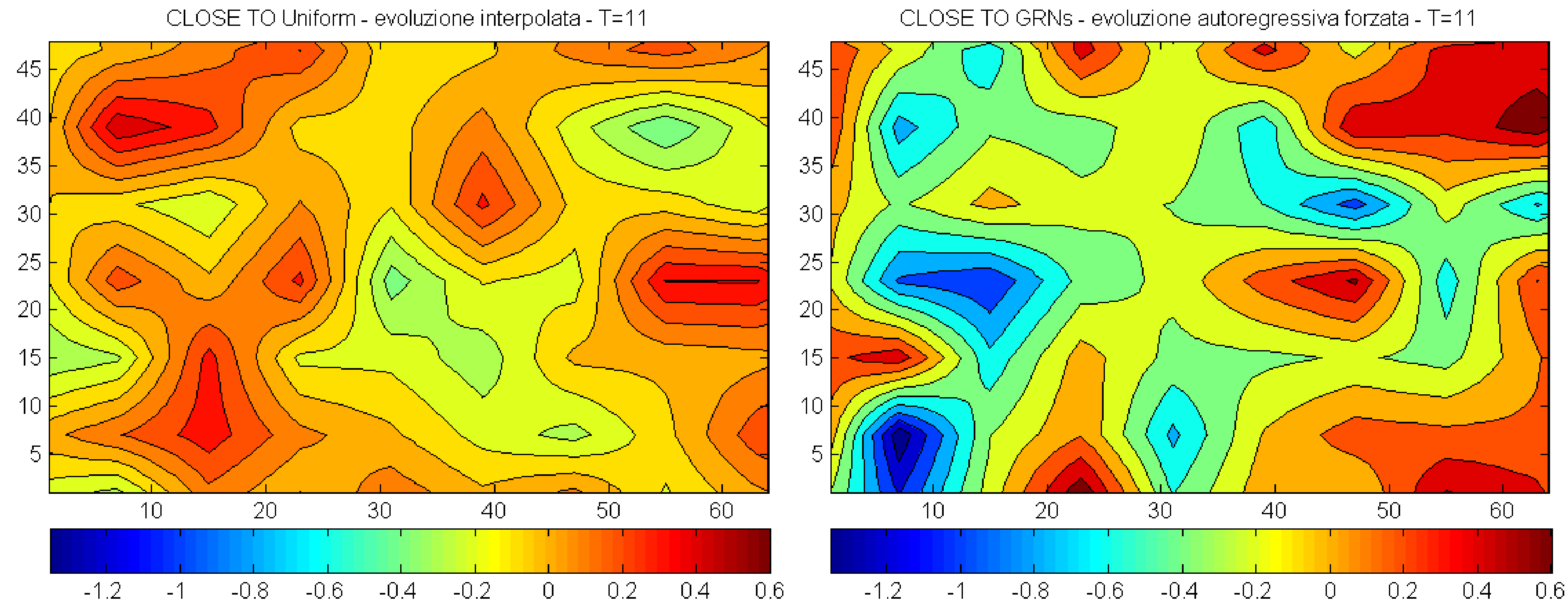
Stochastic Physics

Version 2

Toy model and plots by A. Cheloni

2.5° coarse grid with bilin. interp.

Version 1: 1h coarse time grid with lin. interp.   Version 2: AR1 with 1h decorr. length



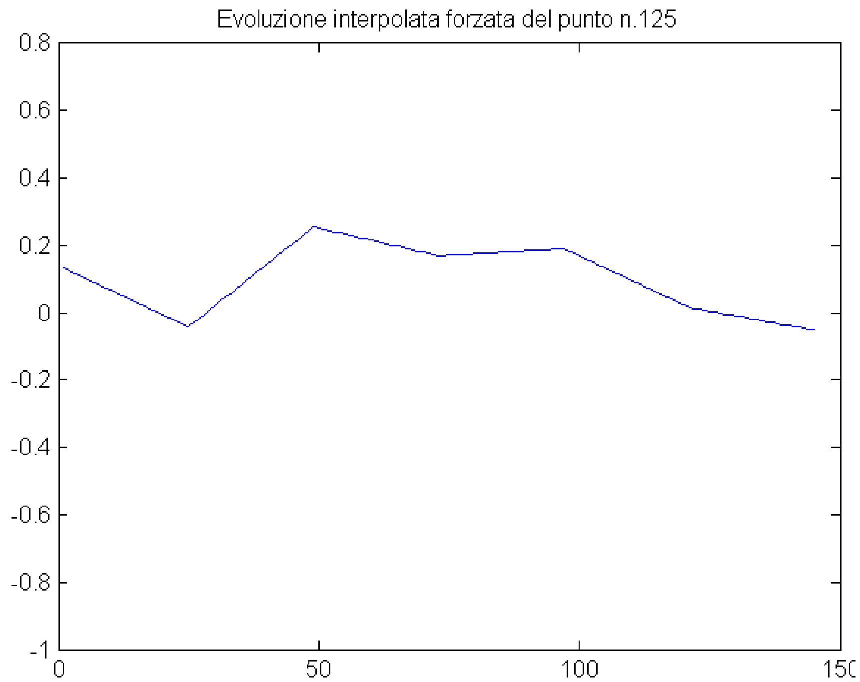
Version 2 can have larger perturbations than version 1

Stochastic Physics

Version 2

Toy model and plots by A. Cheloni

1h coarse time grid with lin. interp.

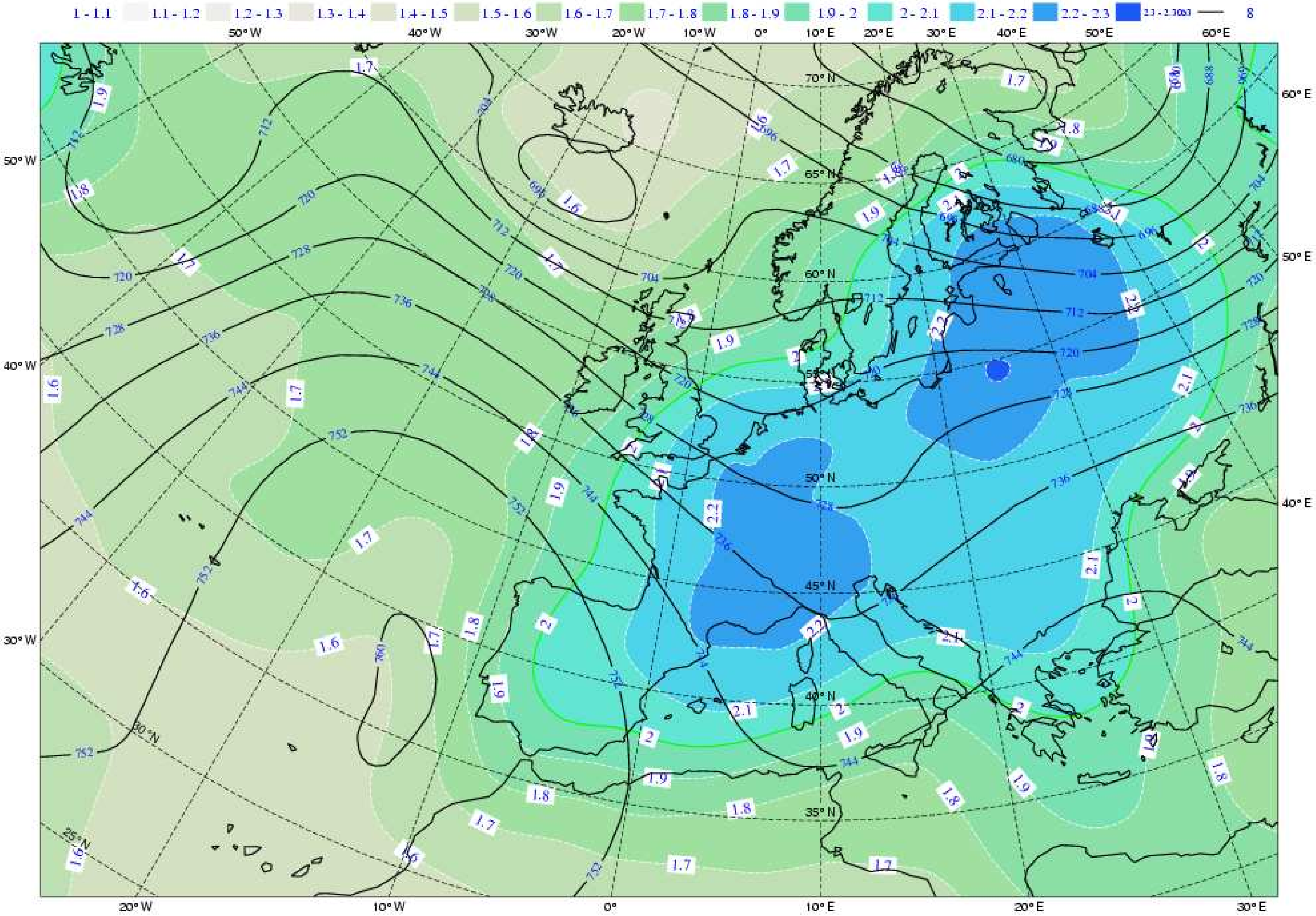


AR1 with 1h decorr. length

Stochastic Physics

Version 2

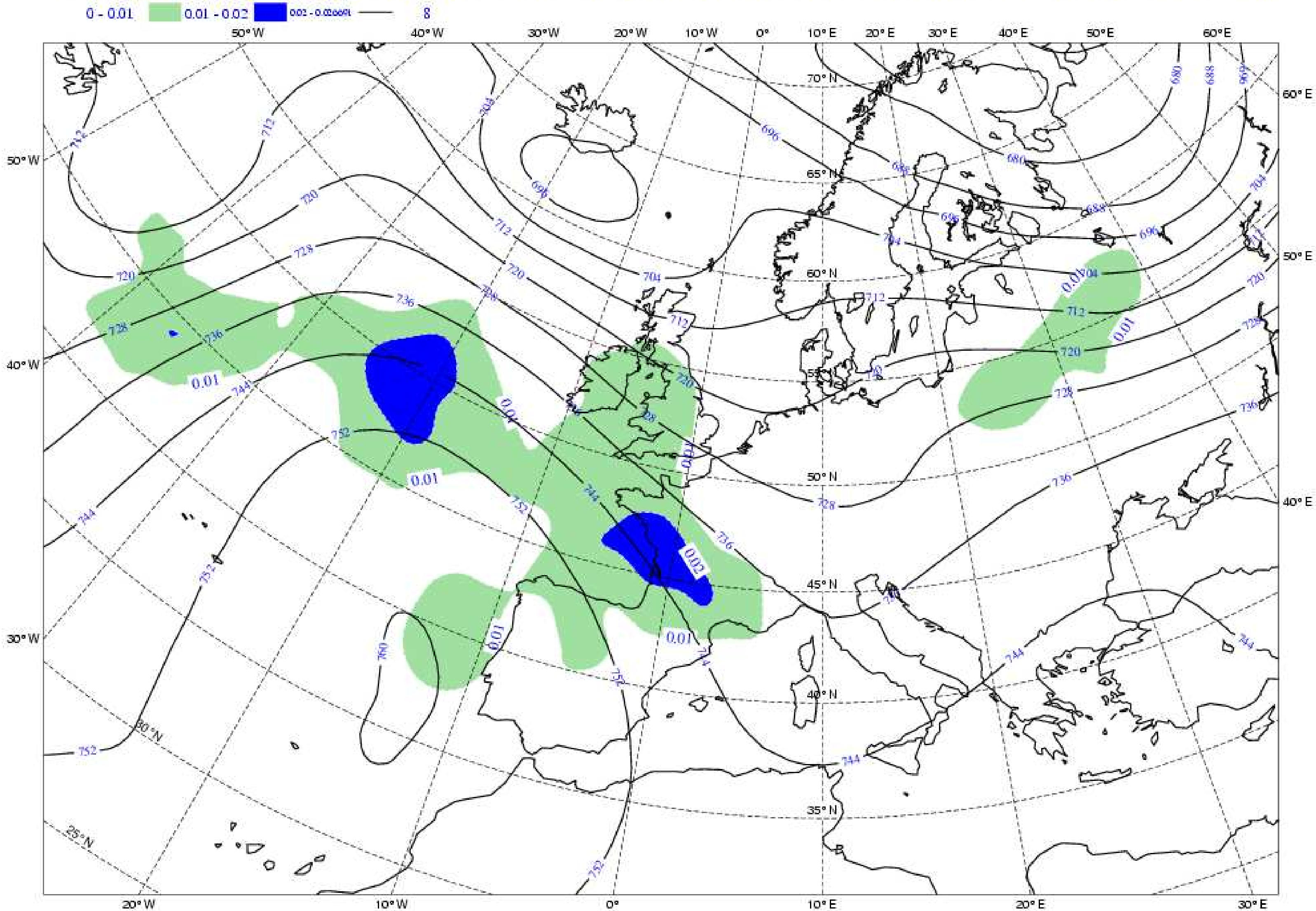
Multiplicative Inflaction Factor: Model level 14



Stochastic Physics

Version 2

Multiplicative Inflaction Factor Difference (NoStochP-StochP): Model level 14

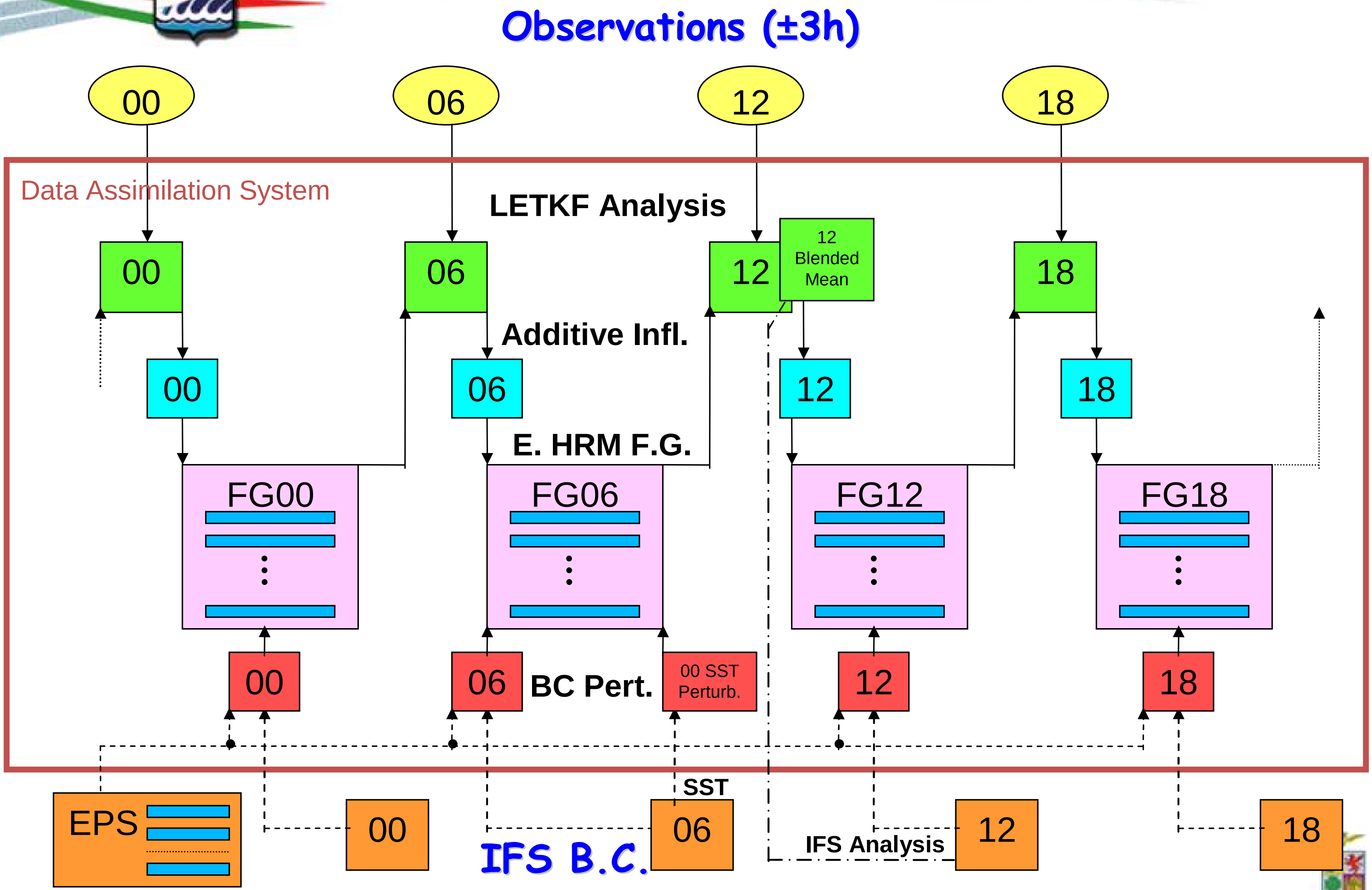


WHAT'S NEW ?

Stochastic Physics

- In the preliminary versions the turbulence parameterization is not perturbed
- This implementation of the stochastic physics has a small/negligible impact on the system (up to now, perhaps more impact in summer periods)
- Version1 works also including the perturbation of the turbulence parameterization; version 2 in certain cases is unstable
- The impact of this new implementation has to be investigated

EXPERIMENTAL LETKF SYSTEM



Lat. Boundary Perturbation

In a limited area ENKF the specification of a suitable ensemble of lateral boundary condition should be addressed (a climatological variance infl. has been used at CNMCA).

According *R.D.Torn et al. (2005)* "… The most natural method for ensemble lateral boundary conditions is one that uses values from a global-model ensemble. … This method involves pairing each limited-area ensemble member with a global ensemble member. … The main advantage of this method is that the covariance relationships are derived from a state-dependent estimate, …". It is called Global Ensemble Sampling (GES).

A variant of this GES scheme using ECMWF-EPS data was implemented.

- As the IFS deterministic run has better skill in the short range than the EPS mean, the EPS ensemble is used only to generate an ensemble of zero-mean perturbations to be added to the IFS run and obtain the desired ensemble of lateral boundary conditions

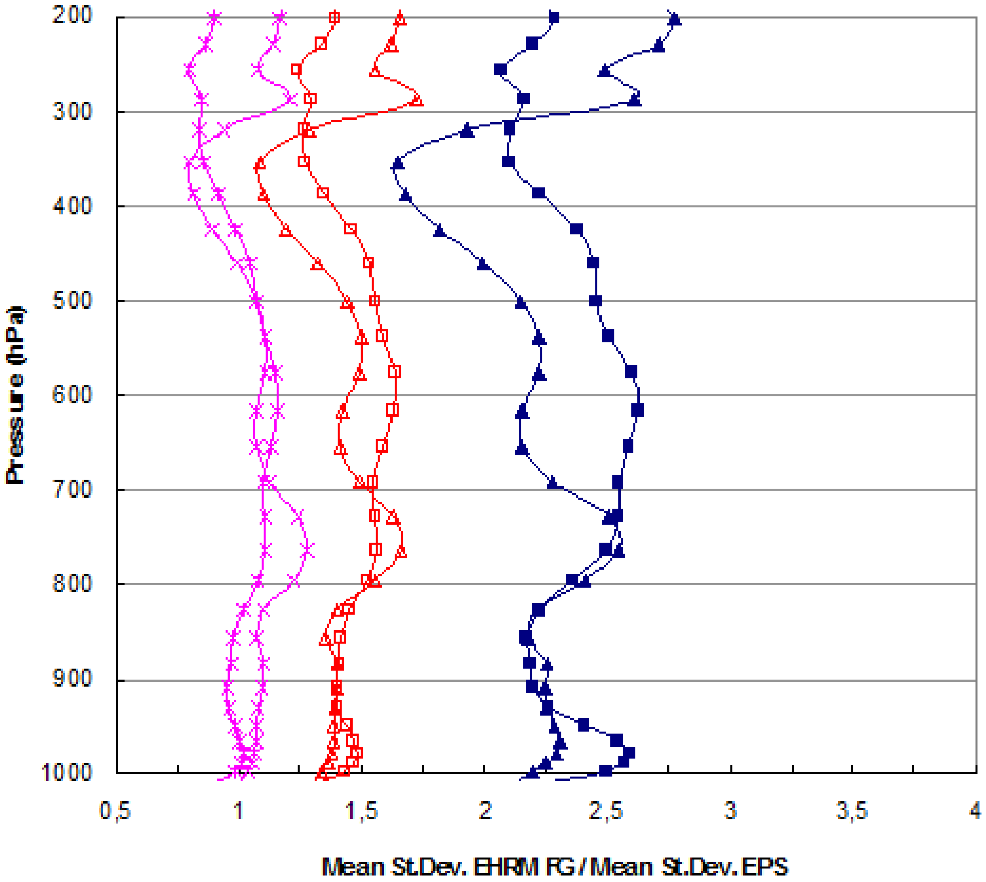
- The IFS run is the closest to the analysis time (i.e. 6 h early). The EPS ensemble perturbations (only two runs a day) are generally "older" than IFS run, also in order to match as much as possible the Euro-HRM spread

- 40 over 50 EPS members are randomly selected and interpolated over the EuroHRM grid

Lat. Boundary Perturbation

Spread comparison: EHRM / EPS

EHRM: 9oct 18 UTC +3h

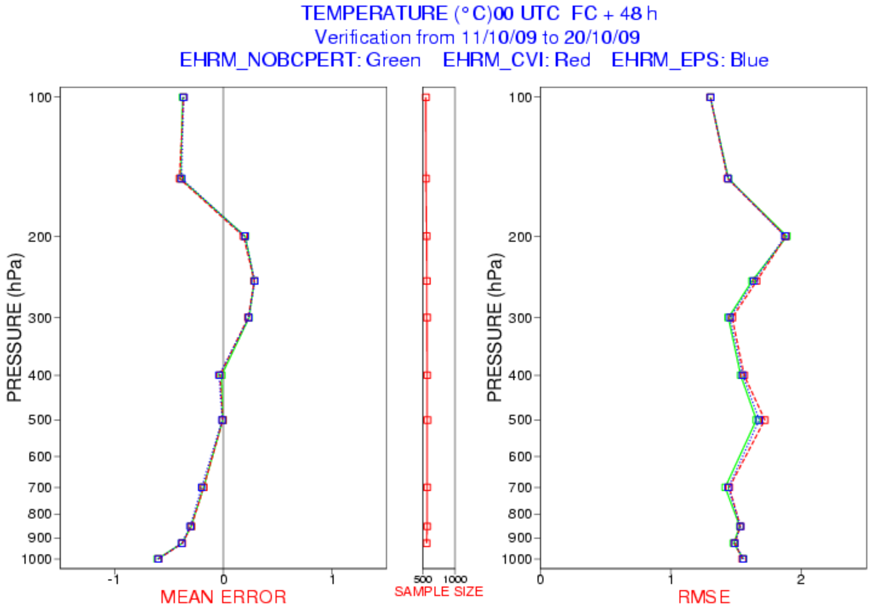


u 09/10/2009 12+9h T 09/10/2009 12+9h	Closest (avail.) EPS run to EHRM one: 12UTC **EPS: 9oct 12 UTC +9h**
u 09/10/2009 00+21h T 09/10/2009 00+21h	EPS run started 12h before : 00 UTC **EPS: 9oct 00 UTC +21h**
u 08/10/2009 12+33h T 08/10/2009 12+33h	EPS run started 24h before: 12 UTC day before **EPS: 8oct 12 UTC +33h**

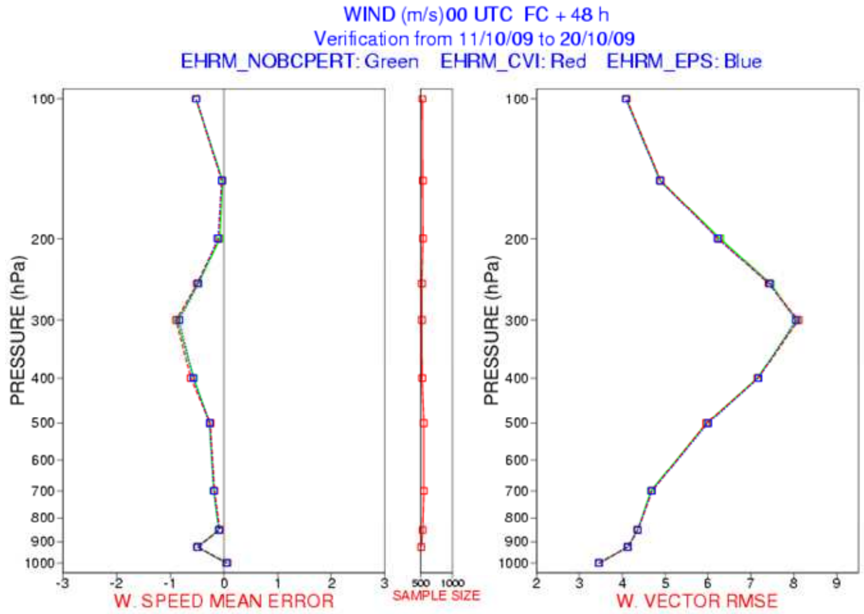
Lat. Boundary Perturbation

Verification against observations

Comparison of two LBC perturbation methods: Climatological Variance Inflaction (CVI) and the Global Ensemble Sampling of EPS perturbation are compared with runs without any BC perturbation



The EPS run started 12h before to the closest one is used to have a much larger spread

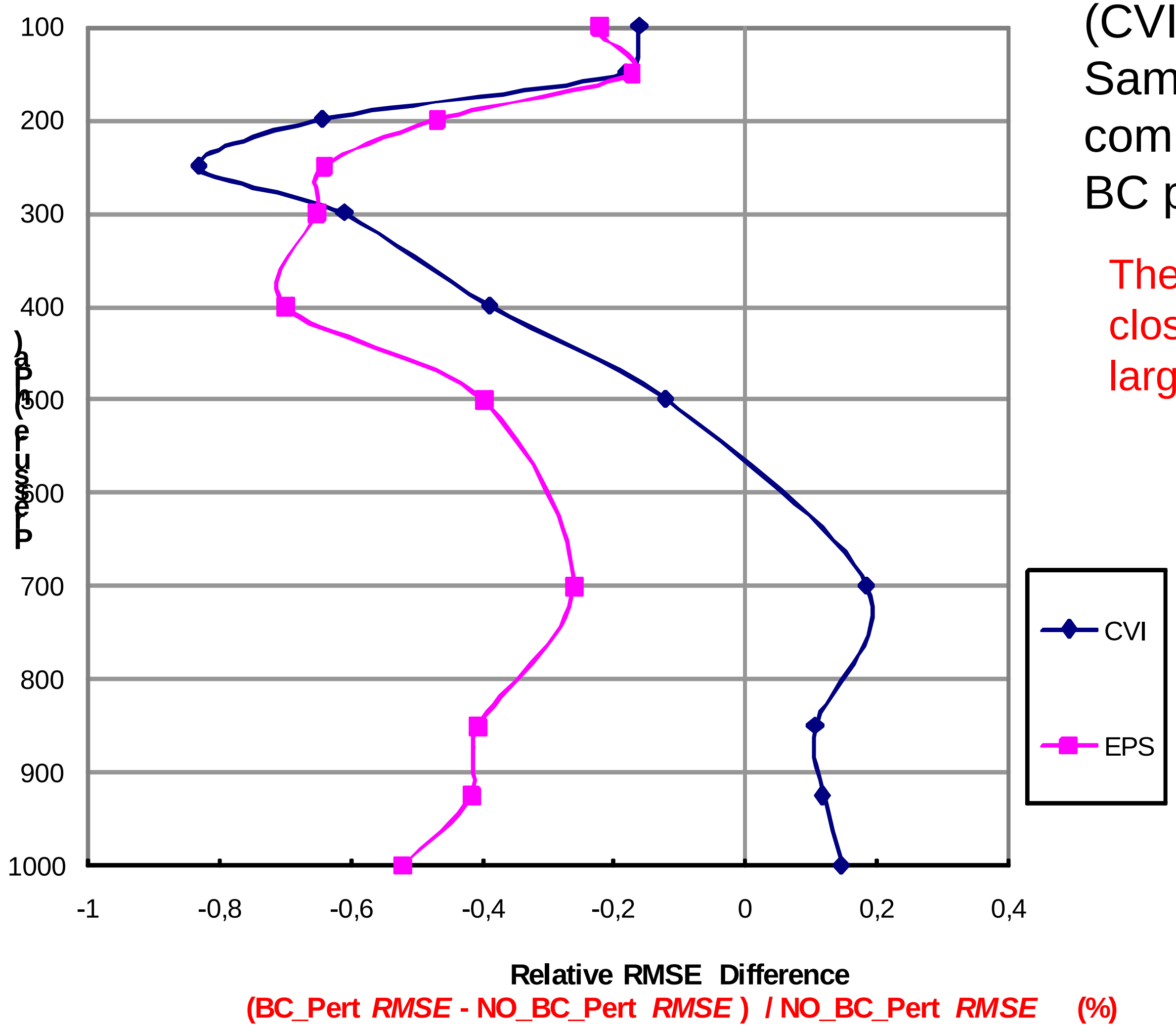


Lat. Boundary Perturbation

Verification against IFS analysis

Comparison of two LBC perturbation methods: Climatological Variance Inflaction (CVI) and the Global Ensemble Sampling of EPS perturbation are compared with runs without any BC perturbation

The EPS run started 12h before to the closest one is used to have a much larger spread



The method using EPS perturbation has generally a positive impact.

A much older EPS run has to be tested (24h).

Radiances Treatment

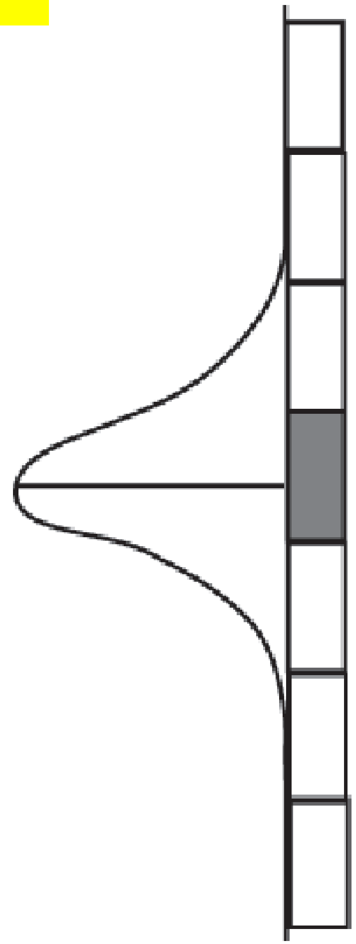
AMSU-A Assimilation

Selecting which satellite radiances to assimilate is complicated by the fact that they not have a single well-defined vertical location

The weighting function at a particular model point indicates the sensitivity of that observation to the state at that model grid point

→"**MAXIMUM-BASED SELECTION**" **METHOD** (Fertig et al. 2007)

- **Assign radiance observations to the model level for which the magnitude of the weighting function is largest**
- **Use this location to select observations within the local region in model space**
- **AMSU-A are treated as "single-level" observations**

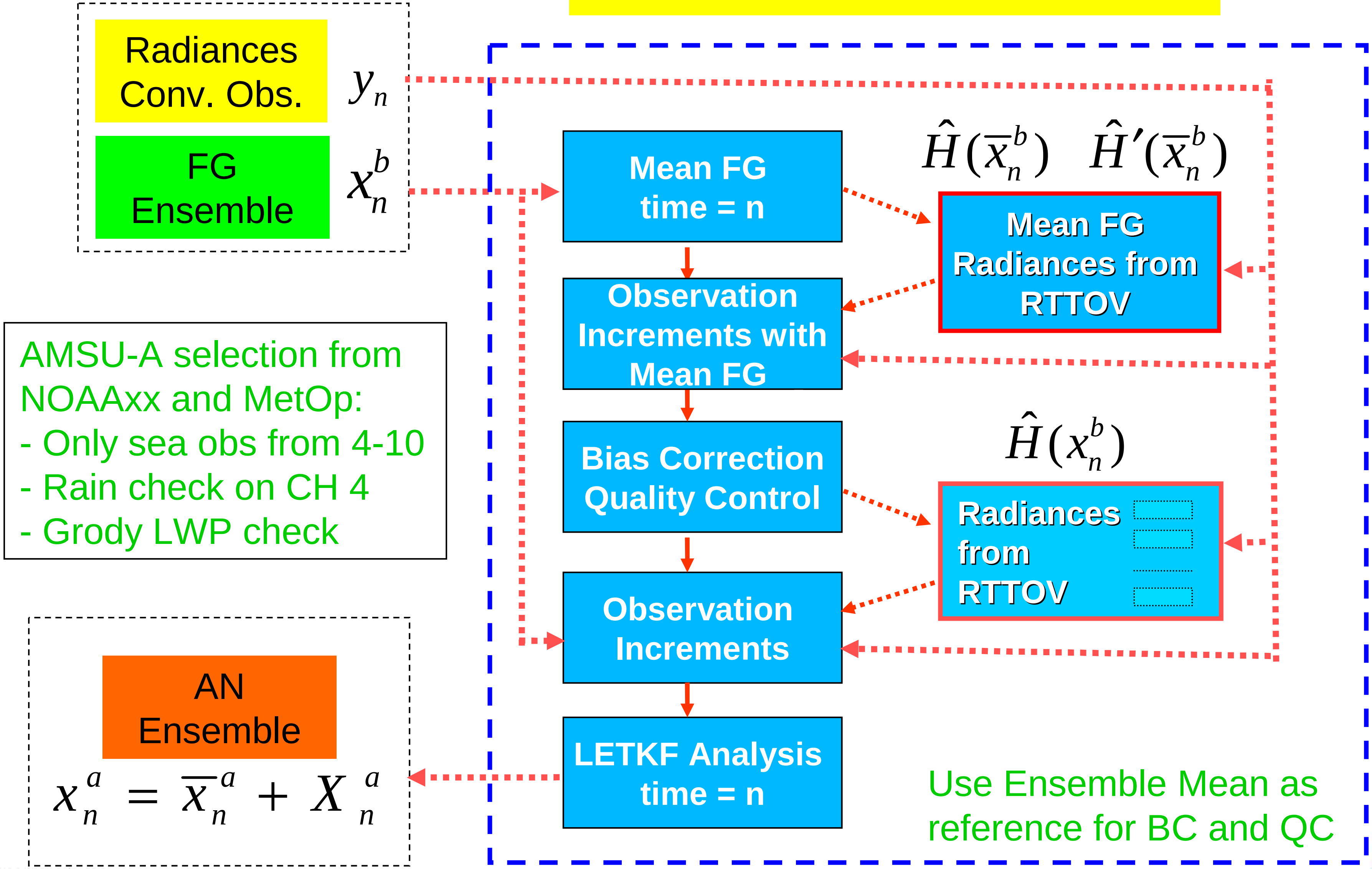


(a) Maximum-Based Selection

An experiment with AMSUA has been recently started

Radiances Treatment

AMSU-A Assimilation



Improvement of algorithm efficiency

AIM :

Reduce the computation time, when the number of observations increase (ex. no thinning in obs pre-processing, radiances assimilation, etc)

HOW DOES IT WORKS?

Whitaker et al. 2008

- **Estimate how much each observation inside the local patch would reduce the ensemble variance for the state element if it were assimilated in isolation**

- **Select observations according to this estimate, so that only the observations with the largest (>= 1%) expected variance reduction are treated**

Improvement of algorithm efficiency

EXAMPLE: 14 oct 2009 12 UTC

Assimilated OBS: 18672 + 2272 (SFC)

Computation time	
OPE	NEW
4785 s	1286 s

Scores: **No skill deteriorations**

It would be effective when a large number of observation will be assimilated
(ex. AMSU + NO SPATIAL THINNING in PRE-PROCESSING)

OUTER-LOOP (Yang et al., 2010)

- The outer-loop, widely used in incremental 3D- and 4D-variational methods, is adapted into EnKF to improve its ability to handle the nonlinearity of the evolving dynamics that can take place in long assimilation windows.

- The outer-loop is constructed within LETKF with the purpose of re-centering the ensemble perturbations around a more accurate ensemble mean

OUTER LOOP: scheme

$$\underbrace{x_n^a = \overline{x}_n^b + X_n^b \overline{w}_n^a}_{\overline{x}_n^a} + X_n^a$$

$X_n^{a,0}$; $\overline{w}_n^{a,0}$

LETKF
time = n

i = 0

i = i+1

Mean analysis time = n-1

$$\overline{x}_{n-1}^{a,i+1} = \overline{x}_{n-1}^{a,i} + X_{n-1}^{a,0}\overline{w}_n^{a,i}$$

Model integration time n-1 to n

$$\overline{x}_n^{b,i+1} = M(\overline{x}_{n-1}^{a,i+1})$$

Observation increments

Mean analysis time = n

$$\overline{x}_n^{a,i+1} = \overline{x}_n^{b,i+1} + X_n^{b,i+1}\overline{w}_n^{a,i+1}$$

$$x_n^{a,i+1} = \overline{x}_n^{a,i+1} + X_n^{a,0}$$

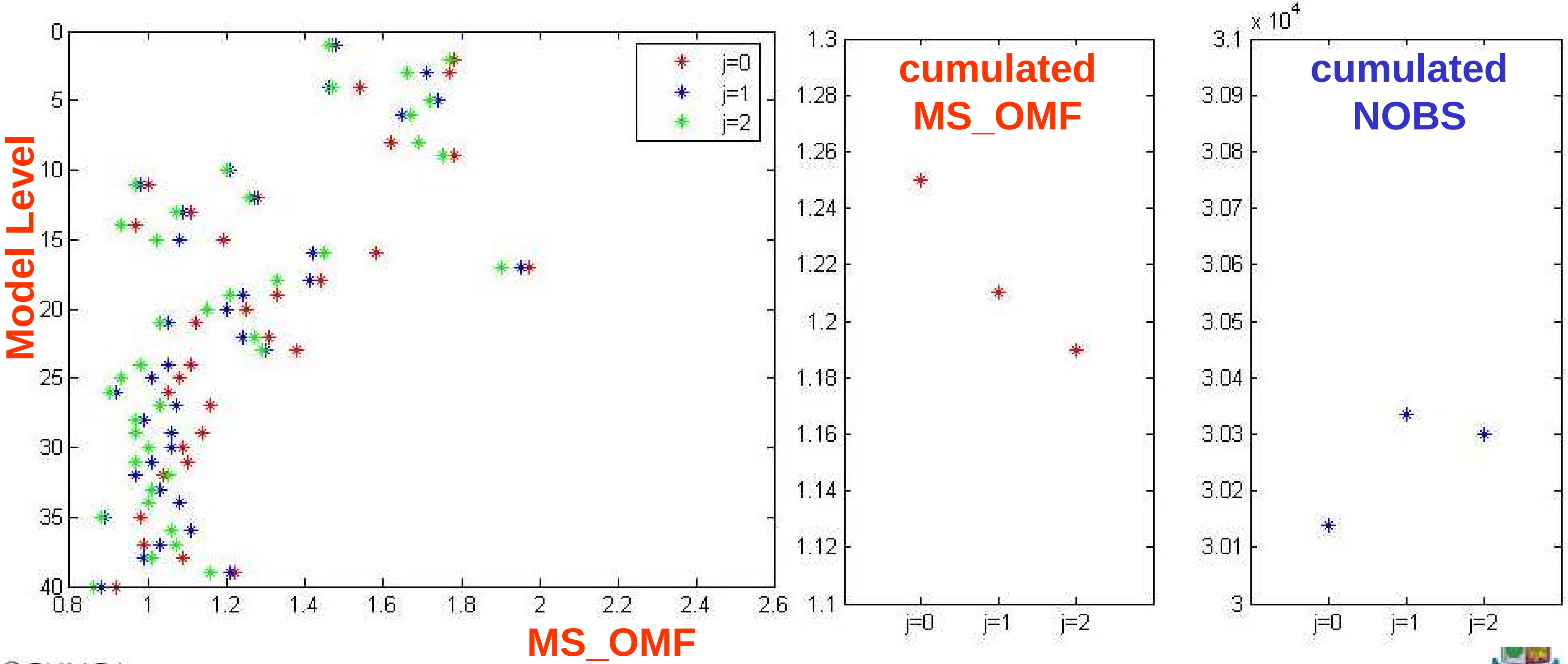
$\overline{w}_n^{a,i+1}$; $X_n^{a,0}$

OUTER LOOP: results

The effectiveness of the algorithm is evaluated looking at the reduction of mean observation increment (OMF), because the best definition of mean analysis

$$\frac{\mathrm{RMS\_OMF}_{i-1} - \mathrm{RMS\_OMF}_{i}}{\mathrm{RMS\_OMF}_{i}}$$

For one case:



Objective Verification

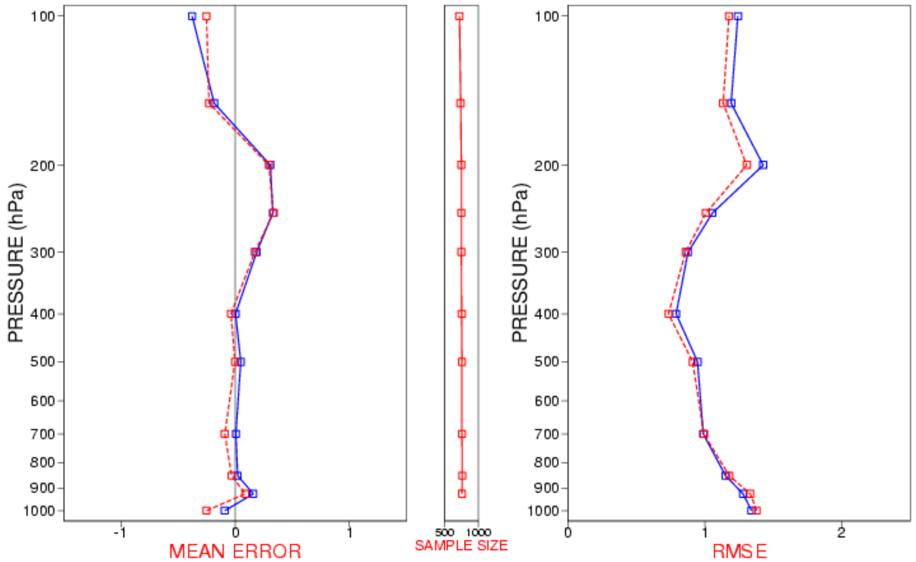
Two experiments:

- 3DV: The deterministic Euro-HRM runs, initialized by the 00 UTC mean LETKF analysis and by the 3DVAR system and driven by IFS boundary conditions, are objectively verified against the European radiosounding observations.

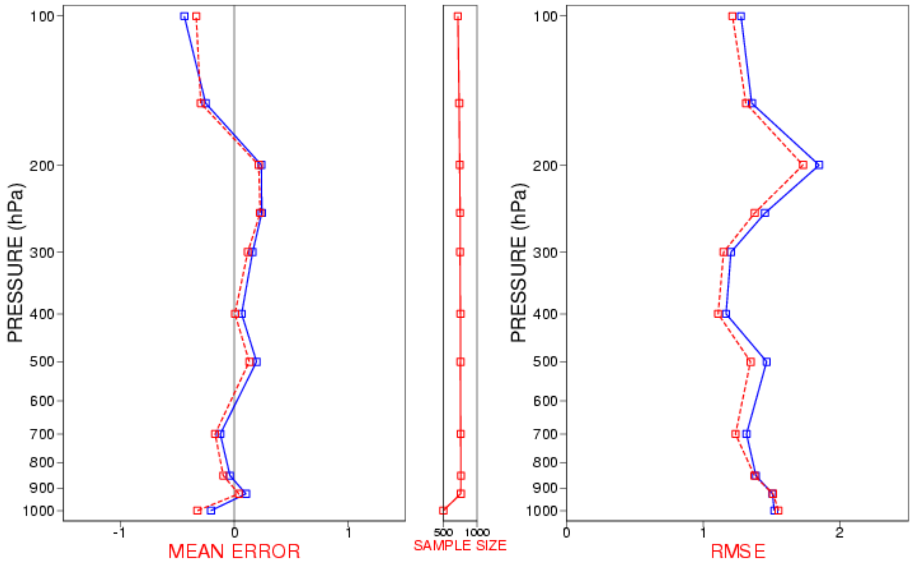
- IFS: Deterministic Euro-HRM integrations starting from the IFS analysis are also performed to assess the skill of the mean LETKF analysis compared to the global IFS 4DVAR system. In this evaluation you have to take into account the reduced number of observation types (no radiances in this experiment) used in the CNMCA LETKF system.

Verification Results: 3DV/LETKF

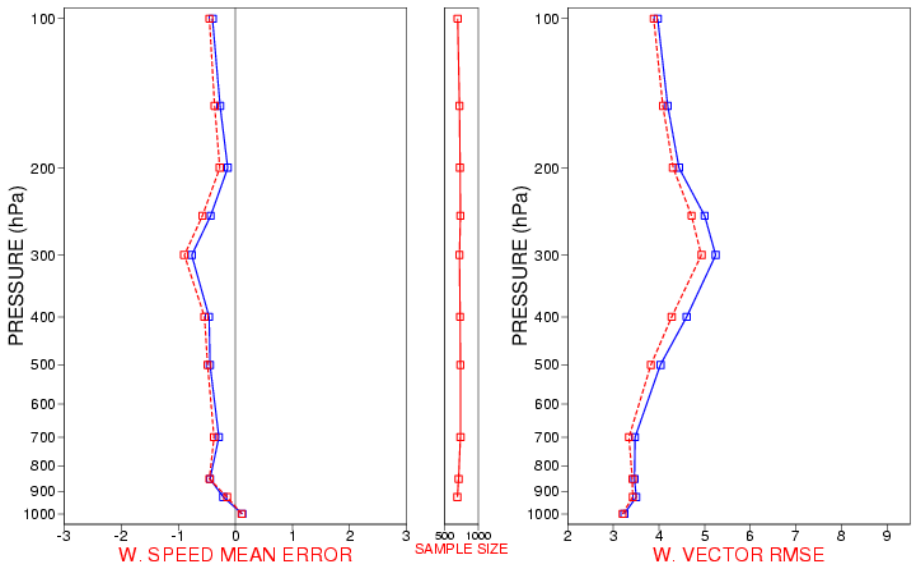
TEMPERATURE (°C)00 UTC FC + 12 h
Verification from 11/10/09 to 25/10/09
EHRM_3DV: Blue EHRM_LETKF: Red



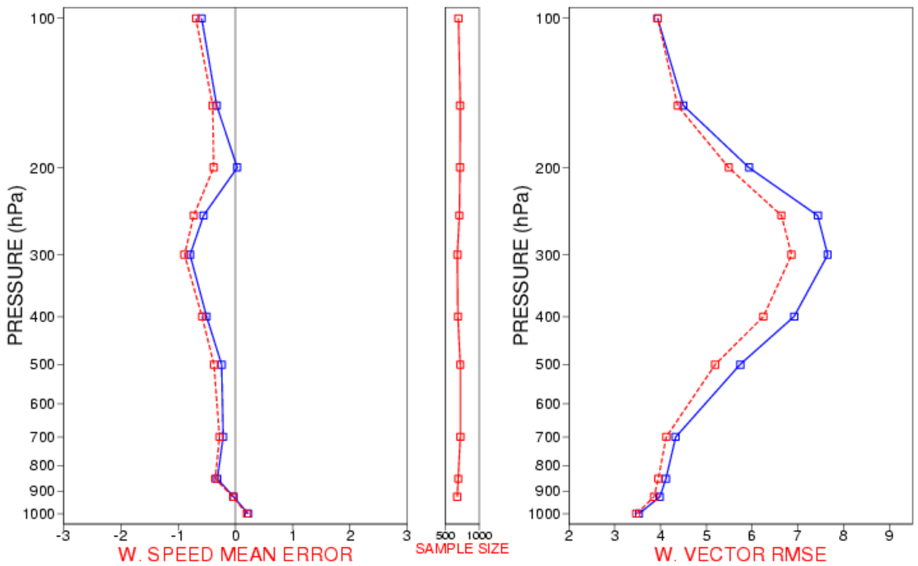
TEMPERATURE (°C)00 UTC FC + 36 h
Verification from 11/10/09 to 25/10/09
EHRM_3DV: Blue EHRM_LETKF: Red



WIND (m/s)00 UTC FC + 12 h
Verification from 11/10/09 to 25/10/09
EHRM_3DV: Blue EHRM_LETKF: Red

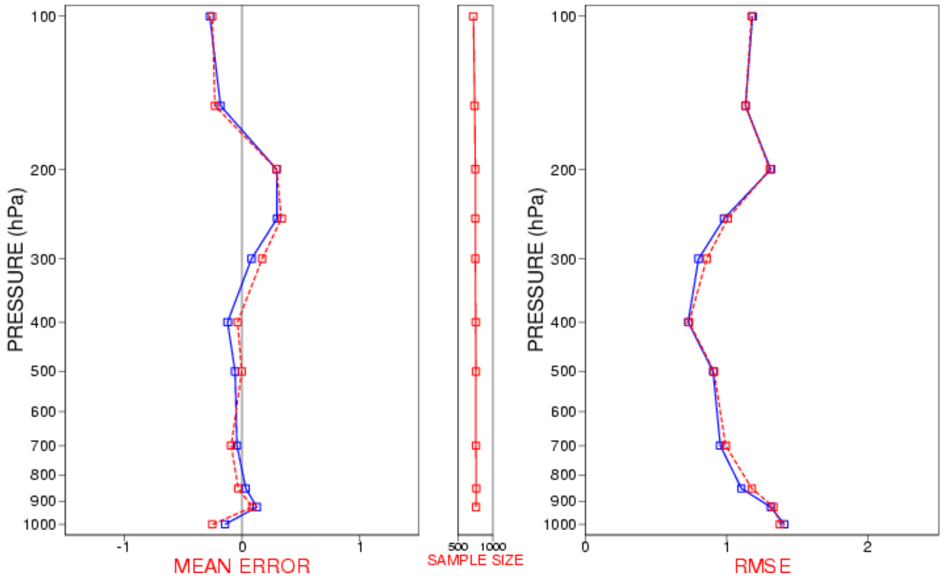


WIND (m/s)00 UTC FC + 36 h
Verification from 11/10/09 to 25/10/09
EHRM_3DV: Blue EHRM_LETKF: Red

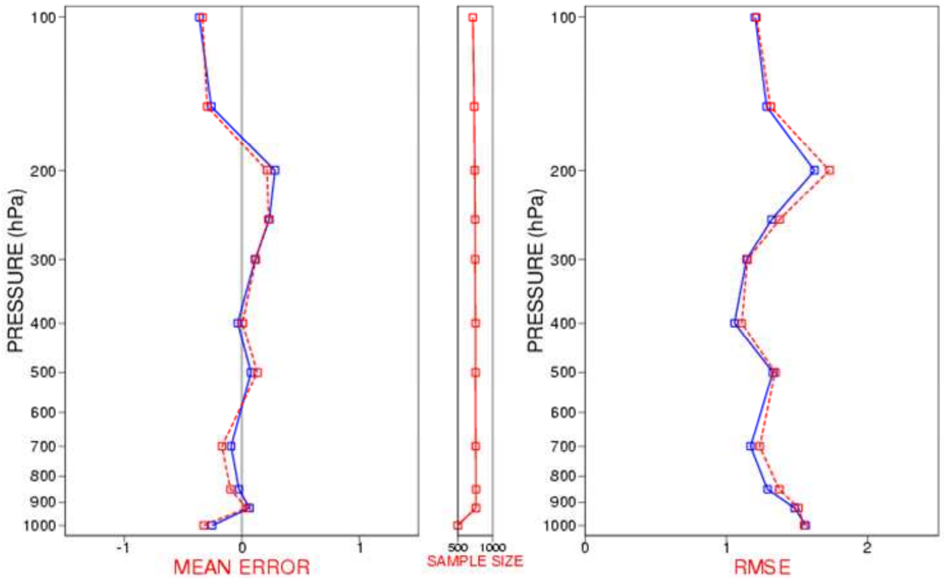


Verification Results: IFS/LETKF

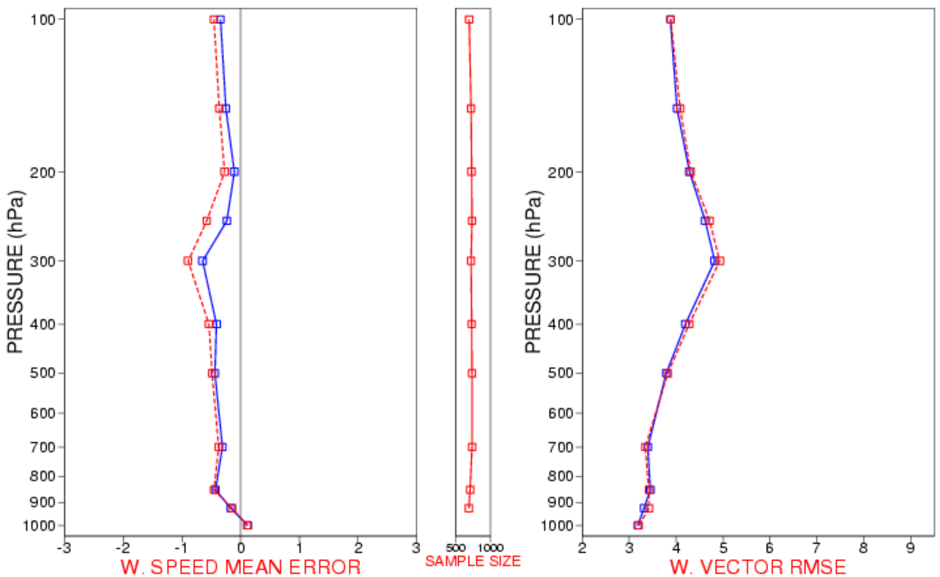
TEMPERATURE (°C)00 UTC FC + 12 h
Verification from 11/10/09 to 25/10/09
EHRM_IFS: Blue EHRM_LETKF: Red



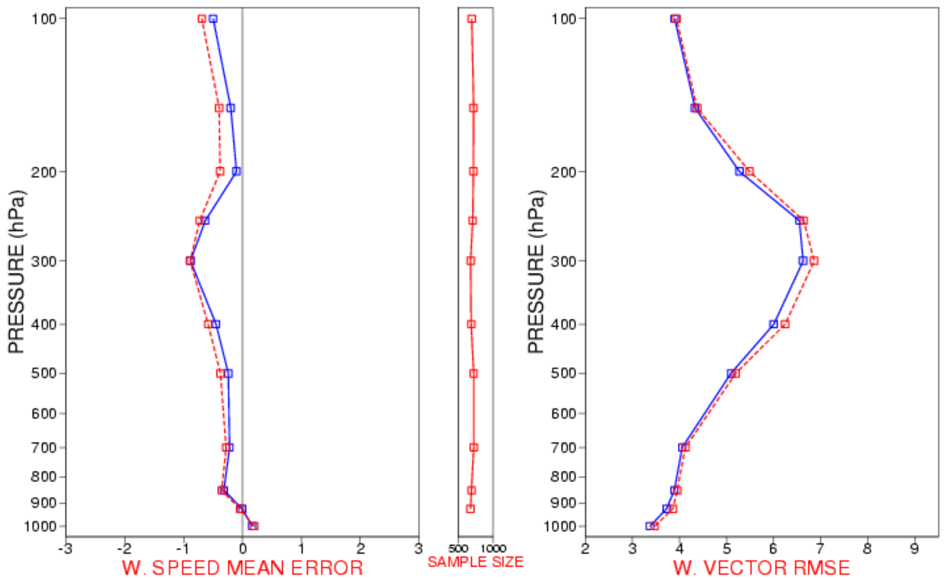
TEMPERATURE (°C)00 UTC FC + 36 h
Verification from 11/10/09 to 25/10/09
EHRM_IFS: Blue EHRM_LETKF: Red



WIND (m/s)00 UTC FC + 12 h
Verification from 11/10/09 to 25/10/09
EHRM_IFS: Blue EHRM_LETKF: Red



WIND (m/s)00 UTC FC + 36 h
Verification from 11/10/09 to 25/10/09
EHRM_IFS: Blue EHRM_LETKF: Red



FUTURE DEVELOPMENTS

- Test AMSU radiances treatment and outer loop
- Further investigation on perturbation of lateral boundary conditions (saturation check, member selection)
- Further tuning of model error representation (stochastic physics version 2, bias correction)
- Increase the number of assimilated observations removing spatial thinning in obs pre-processing
- Use weight interpolation for finer spatial resolution runs
- Test methods of adaptive covariance localization

Thanks for the attention!
Any questions?