Meteorologisk institutt
met.no

Ensemble Prediction and Data Assimilation: an enriching partnership

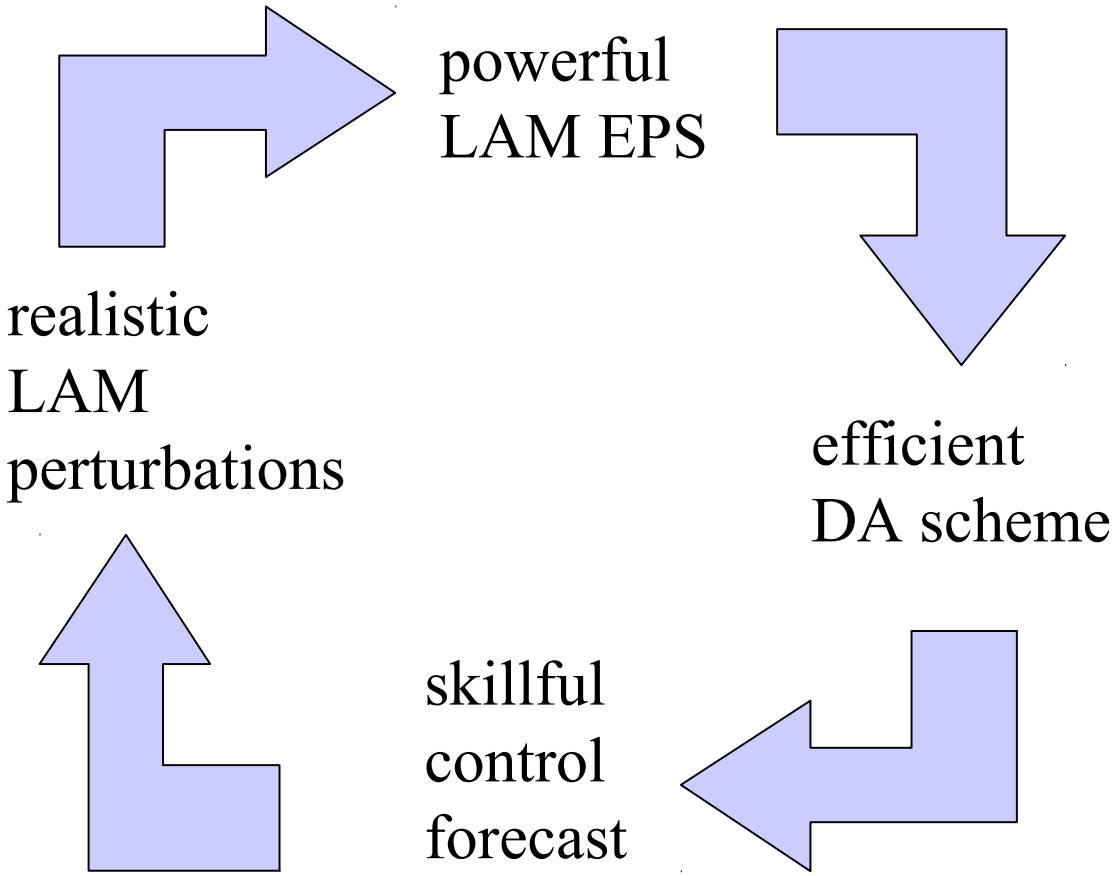
The HIRLAM ASM and ALADIN Workshop,
Reykjavik April 2013

Jelena Bojarova

Structure of the talk

I am going to talk about relations between ***Data Assimilation*** and ***Ensemble Prediction***. Here will be not so much results , mainly an invitation to discussion...

My Dream:
Consistent EPS-DA system on meso-scales



An enriching partnership is possible, isn't it ?!

EPS

A methodology to quantify uncertainty in the estimation of the model state

Mutual development

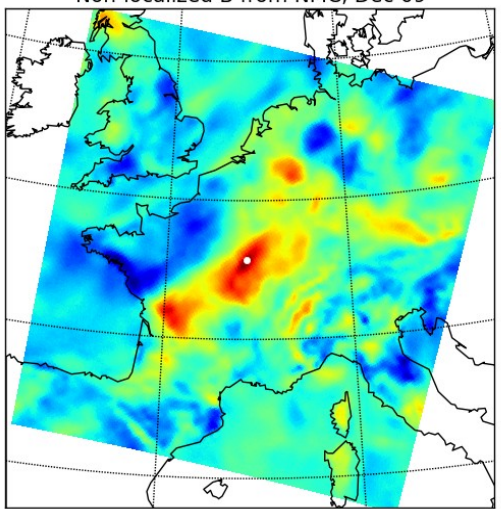
DA

An estimate of the initial model state + an equipment to generate the uncertainty in the initial conditions

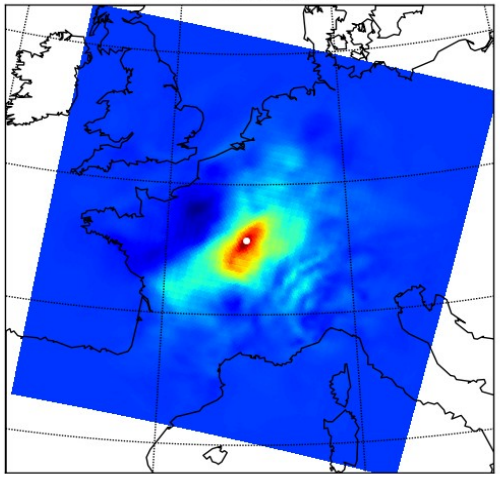
EPS & DA : advanced spatialisation tool

Long time series climatological ensemble is able to represent non-homogeneity and anisotropy induced by orography and land-sea mask

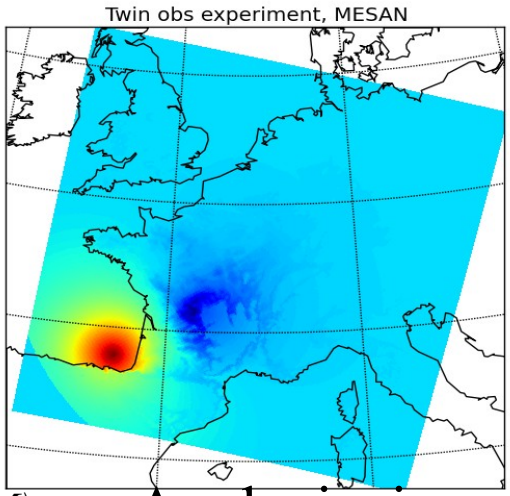
Non-localized B from NMC



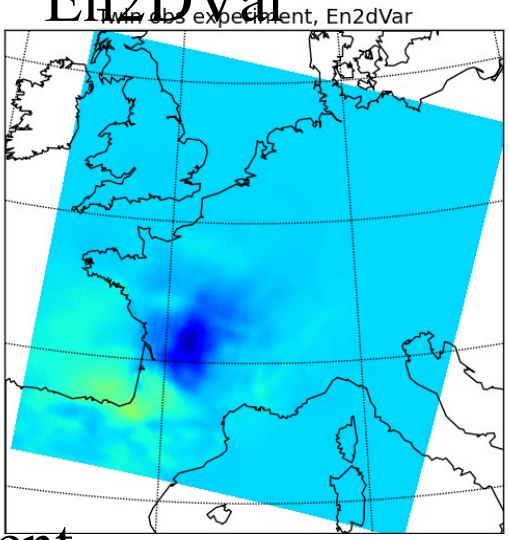
Localized B from NMC (Schur product)



MESCAN



En2DVar



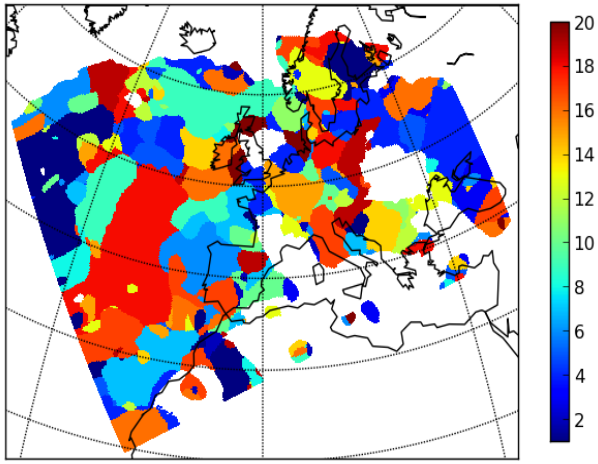
Analysis increment

By Tomas Landelius, SMHI
(cooperation with MF within EURO4M)

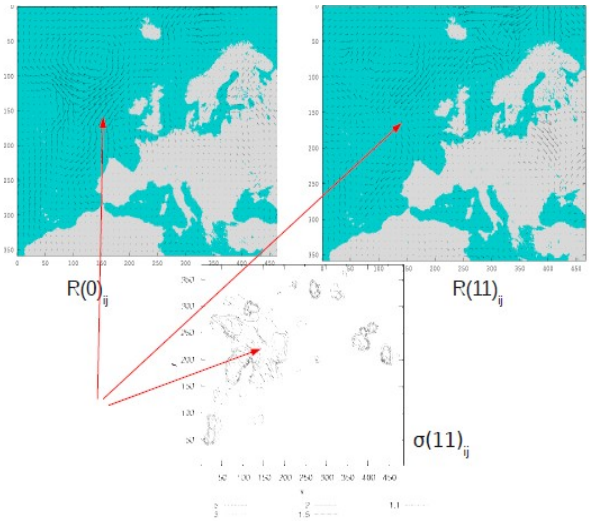
EPS&DA: phase-error correction

The ETKF based ensemble of perturbations allows the assimilation of structures by combining image warping and pseudo-observation techniques within the hybrid 3D-VAR

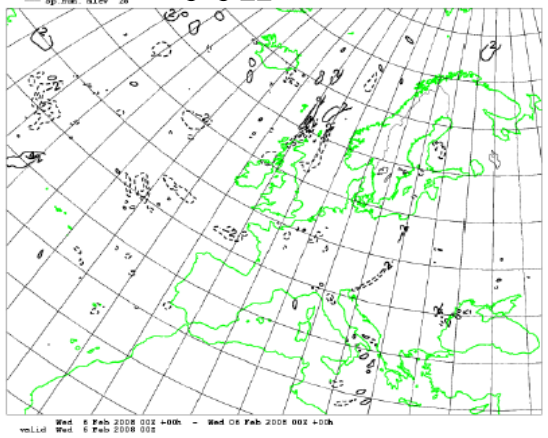
Locally "best" ensemble member



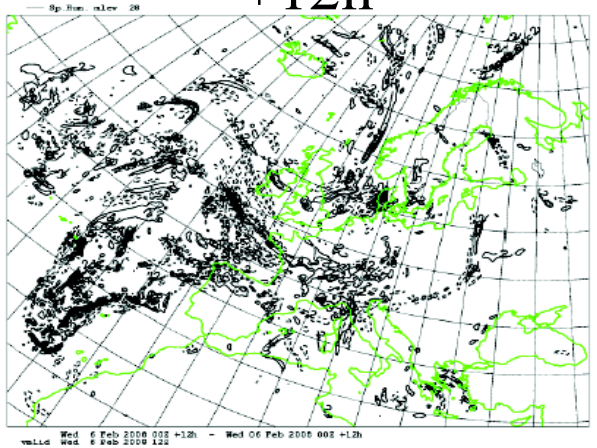
Locally "best" member selection



+00h



+12h



Impact of the phase-error corrections

By Nils Gustafsson et al, SMHI (see poster for details)

DA & EPS : performance of the prediction system

It is important how we sample ensemble

Leith (1974) showed that averaging the ensemble forecast yields a mean forecast superior (in mean square error score) to the control forecast, provided the ensemble perturbations are representative of the initial probability distribution of the basic flow around the control forecast.

E. Kalnay and Z.Toth (1997)

It is more important what we sample the ensemble around and what processes we resolve

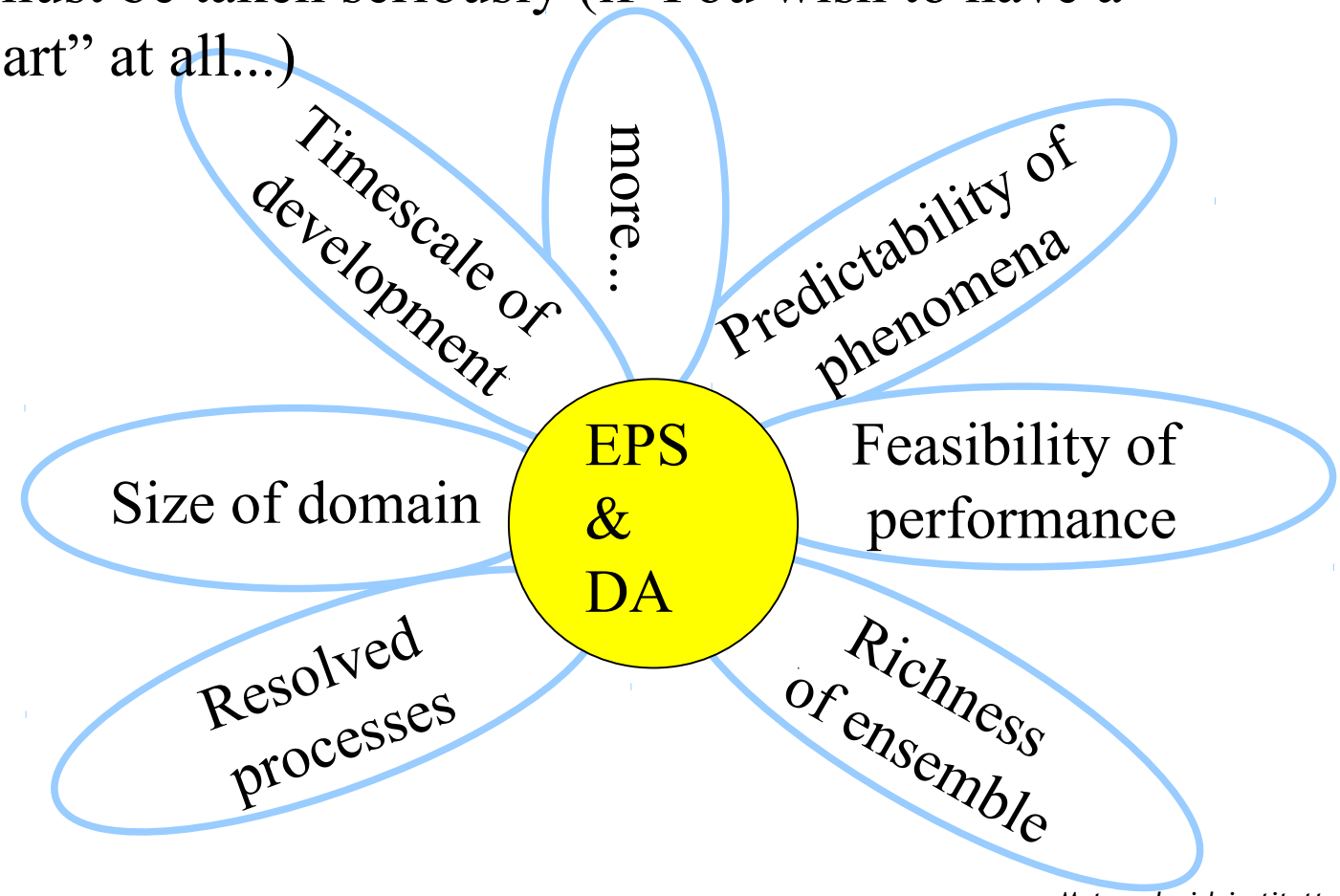
The performance of an ensemble prediction system strongly depends on the quality of the data-assimilation system used to create the control forecast and the numerical model used to generate the forecast perturbations

R. Buizza, M. Leutbecher and L. Isaksen (2008)

DA & EPS: Constraints

DA and ***EPS*** are grown up enough to understand that a perfect world exists … but only in their imagination !

In reality there are constraints imposed by the "counterpart" and they must be taken seriously (if You wish to have a "counterpart" at all...)



Factors influencing the choice of the model domain

Domain should be **large enough** to allow LAM perturbations to develop: a signal/perturbation inserted in the LAM domain should **have time to breed within the area of domain**.

group velocity of disturbance propagation (Norman Phillips, 1990: 850km <=> 6h)

freedom developing the position and the intensity for the phenomena of interest

the age of lateral boundary conditions;

Other factors

ability of lateral boundary conditions to propagate information through lateral boundaries;

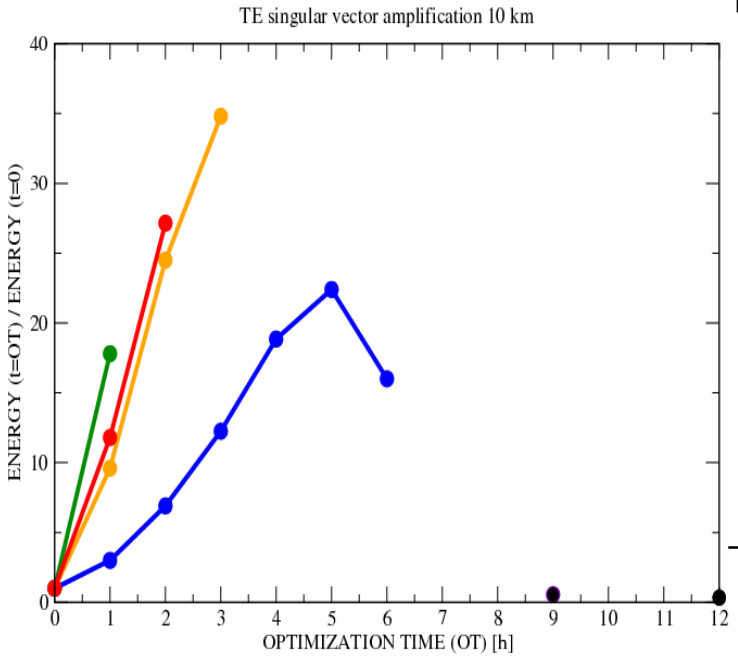
deformation of information through lateral boundaries due to differences in model resolution, physics, orography;

intrinsic limitations of LAM DA schemes

practical limitation of the “small institute” groups

Constraint : size of domain

Domain size:
1000km x 1000km

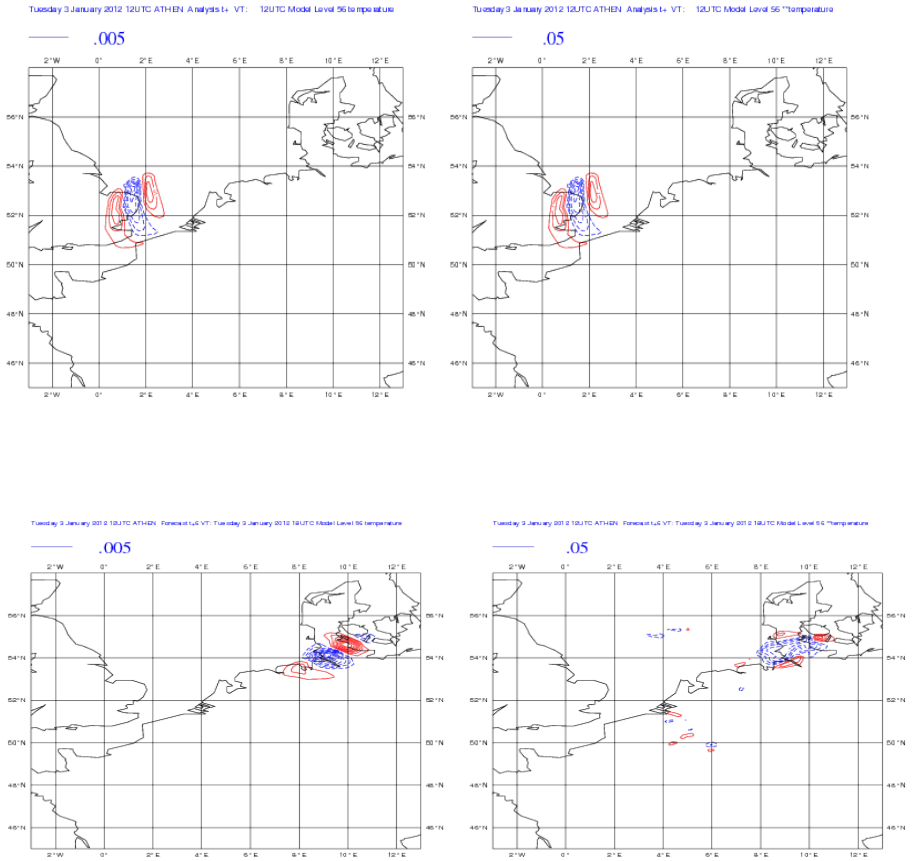


+00

+06

SV

IC + 10*SV -IC



Amplification of
total dry energy

(from Jan Barkmeijer)

Constraint : size of domain

Feasibility to manage by a "small institute"

Small domain

Large domain

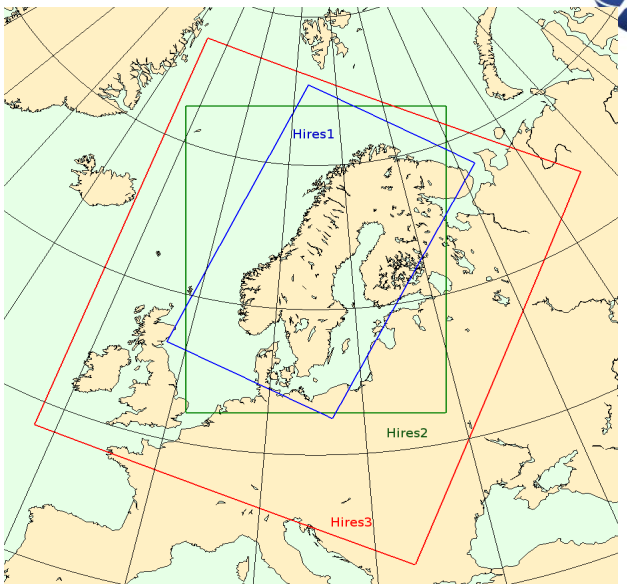
Relevant to fewer

Relevant several

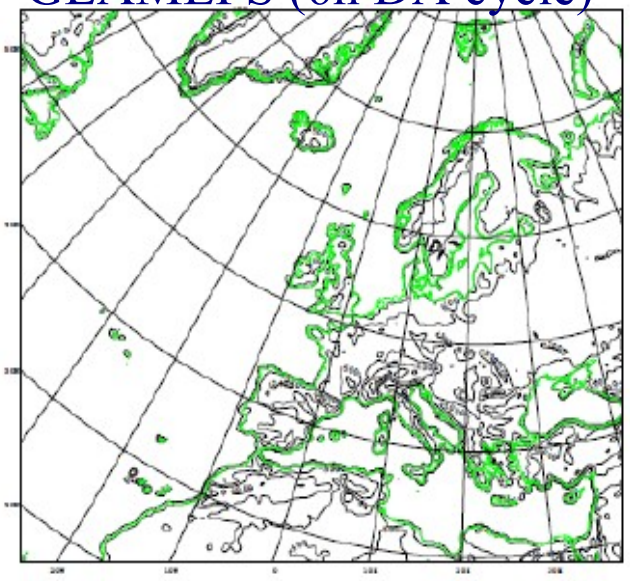
Small cost per institute

Even smaller cost per institute

HarmonEPS (3h DA cycle)



GLAMEPS (6h DA cycle)



Constraint: perturbation generation

Hybrid Data Assimilation
(50% EPS; 50% Climatological
By Ricardo Todling & Amal El Akkraoui)

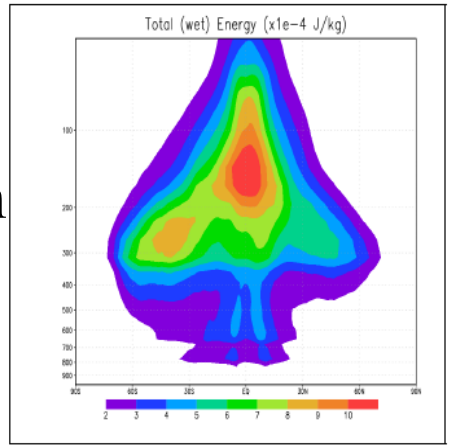
32 EnKF perturbations+
Additive inflation

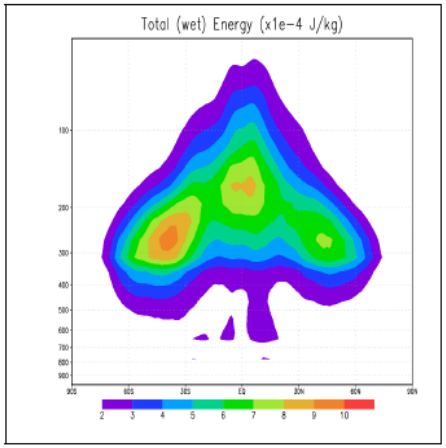
Filter Free:
Additive perturbations downscaled from NMC-like 48-24 hr forecasts

My conclusion:
We are far away from being optimal generating perturbations

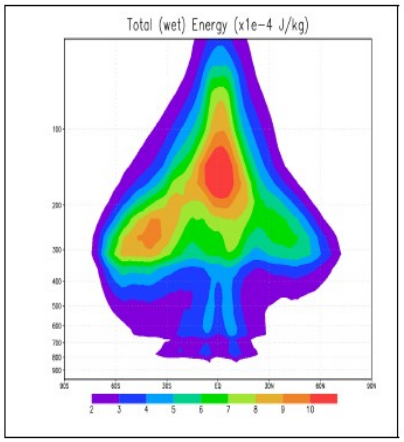
Background spread
Total wet energy norm

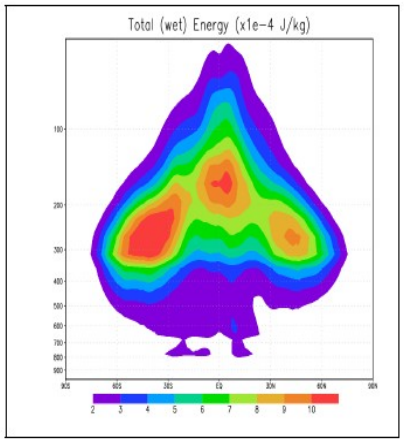
+03h





+09h





EnKF based

Filter-Free

Perturbation generation : Who is the winner ...?

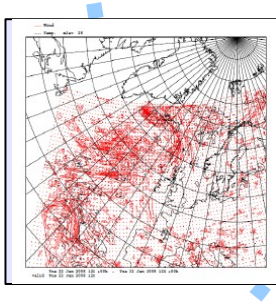
Randomization of climatological B

NMC method 48h -24h

Hessian based perturbations

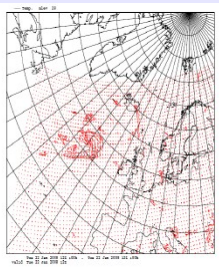
$$P_a^{1/2} \approx [I + \tilde{V}(\tilde{\Lambda}^{-1/2} - I)\tilde{V}^T]P_f^{1/2}$$

Competition or synthesis



EnsDA

ETKF



? ...

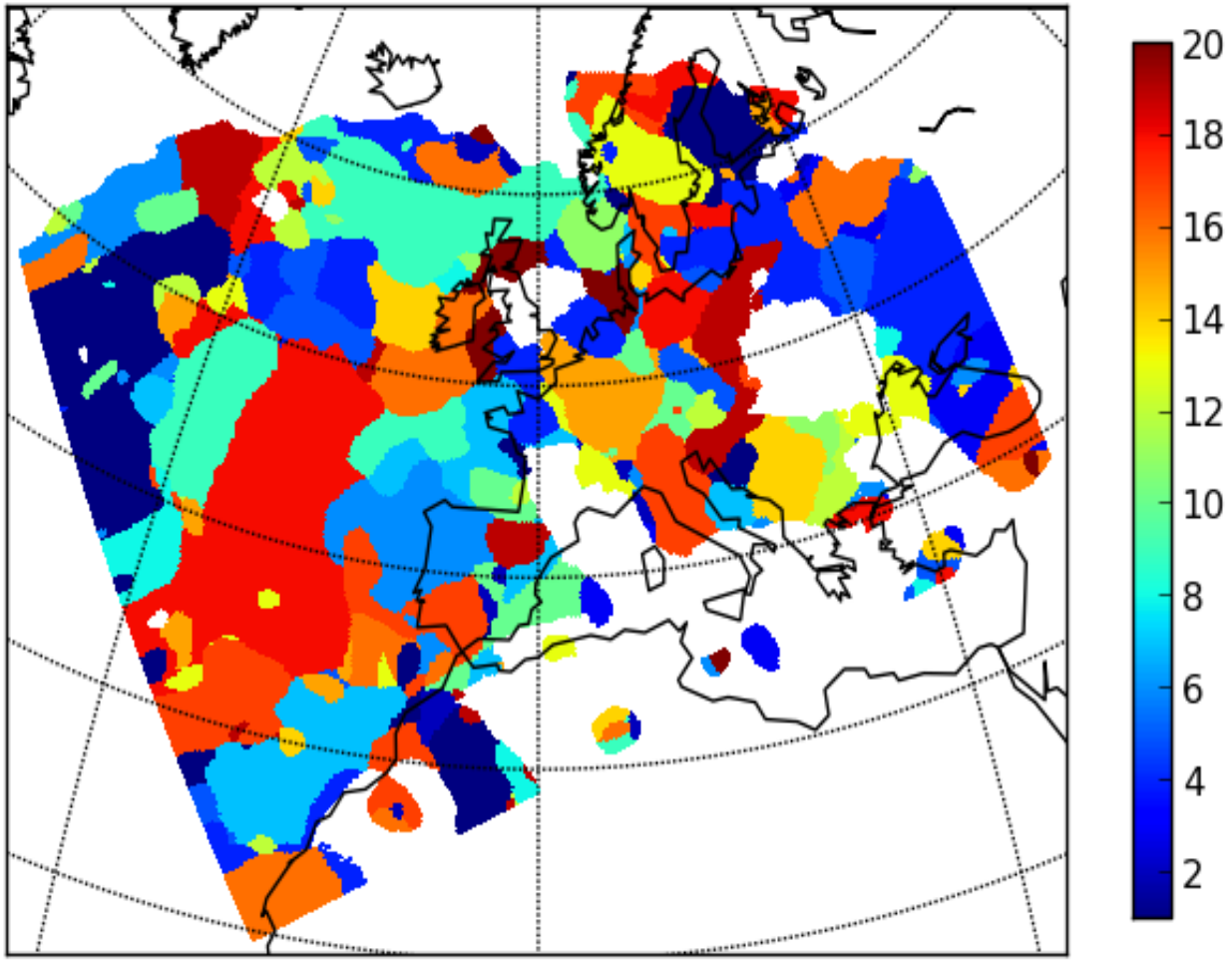
If competition is selected as a way to move forward, the quality issue should not be put aside...

I myself believe more in synthesis ....

Perturbation generation: coherent structures

For DA variability captured by ensemble is very important !

HIRLAM ETKF Rescaling scheme (20 members)



The magic numbers

Domain size +
Perturbation generation +
Size of ensemble +
Computer power resources

Model resolution

What lies behind "2.5" ?

Pragmatic solutions

Pragmatic solutions should be taken in order to move forward, but ....

Small children feel themselves bad when they see what happens in the world around. But I know one can do!. The problem is that grown-up people accept the first solution what is working at the moment without thinking if it will work well in continuation. It would be better if they would question themselves more often whether it is the best solution what they have found now...

Maksim Olof Gustafsson

Estimating the affordable resolution

Grid size computation
Åke Johansson, SMHI

Large LAM Domain - C22 Domain - North Atlantic and Europe

The area of the C22 domain is approximately 3 10^{13} m^2 which imply that

Area size LAM

$$\alpha = \frac{a}{A} = \frac{3 \cdot 10^{13}}{5.1 \cdot 10^{14}} = 0.0588$$

Area size GL

The LAM is integrated out to 48 hours compared to the 240 hours of the GM which give that

Forecast length LAM

$$\gamma = \frac{t}{T} = \frac{48}{240} = 0.2$$

Forecast length GL

Equation (12) then gives that

$$\beta = \left(\frac{1}{\alpha\gamma}\right)^{\frac{1}{3}} = \left(\frac{1}{0.0588 \cdot 0.2}\right)^{\frac{1}{3}} = 4.4$$

which implies that the grid size could be

$$\Delta x = \Delta y = \frac{\Delta X}{\beta} = \frac{\Delta Y}{\beta} = \frac{15.436\ \text{km}}{4.4} = 3.5\ \text{km}$$

=> Δx=3.5

Predictability on meso-scales

(Question by Åke Johansson: *we know predictability on what scales we would like to describe, we know predictability on what scales we can describe, but do we know predictability on what scales should we describe* ?)

Three types of the error growth

Inverse cascade	of errors from smaller to larger scales is due to non-linear inertial term in the hydrodynamical equation and depends on a saturation of error energy
Baroclinic instability	The error energy is extracted from the background flow and is not transferred from smaller scales
Advection	The errors in the large scales will quickly create errors in the small scales due to erroneous large scale advection

What is the purpose of the EPS on meso-scales

(following Åke Johansson)

- Errors are initially present in **all scales**
- Errors in small scales saturate **very quickly**
- The predictable part of the flow in the 6h-48h window **is mostly on synoptic scales** which are due to baroclinic instabilities
- EPS is needed **to filter out** the increasing amount of small scales **with increasingly more limited predictability** due to **increased resolution**
- EPS is necessary to quantify the probability **of the small scale events** occuring which are driven **by large scale flow**

An invitation to discussion

Following Fabry & Sun

José A. Garcia-Moya
AEMET

HIRLAM ASM 2013

Working group on transversal predictability-data assimilation issues
Thursday the 18th of April 14:30-16:00