Estimation and representation of background error covariances

L. Berre

Météo-France/CNRS

How can we estimate error covariances ?

- Definition of background errors : $\mathbf{e_b = x_b - x_t}$
- The true atmospheric state $\mathbf{x_t}$ is never exactly known.
- Use observation-minus-background departures :

  $\Rightarrow$ space- and time-averaged variances and correlations,
  using assumptions on spatial structures of errors.

- Use ensemble to simulate the error evolution and
  to estimate complex background error structures.

Outlook

- Simulation of error cycling in DA
- Innovation-based estimates
- Filtering of variances and correlations

METEO FRANCE

Error cycling in DA (linear or weakly nl)

- Equations of analysis state and errors :

$$\mathbf{x_a = (I-KH)\, x_b + K\, y_o}$$

$$\mathbf{e_a = (I-KH)\, e_b + K\, e_o}$$

- Equations of forecast state and errors :

$$\mathbf{x_f = M\, x_a}$$

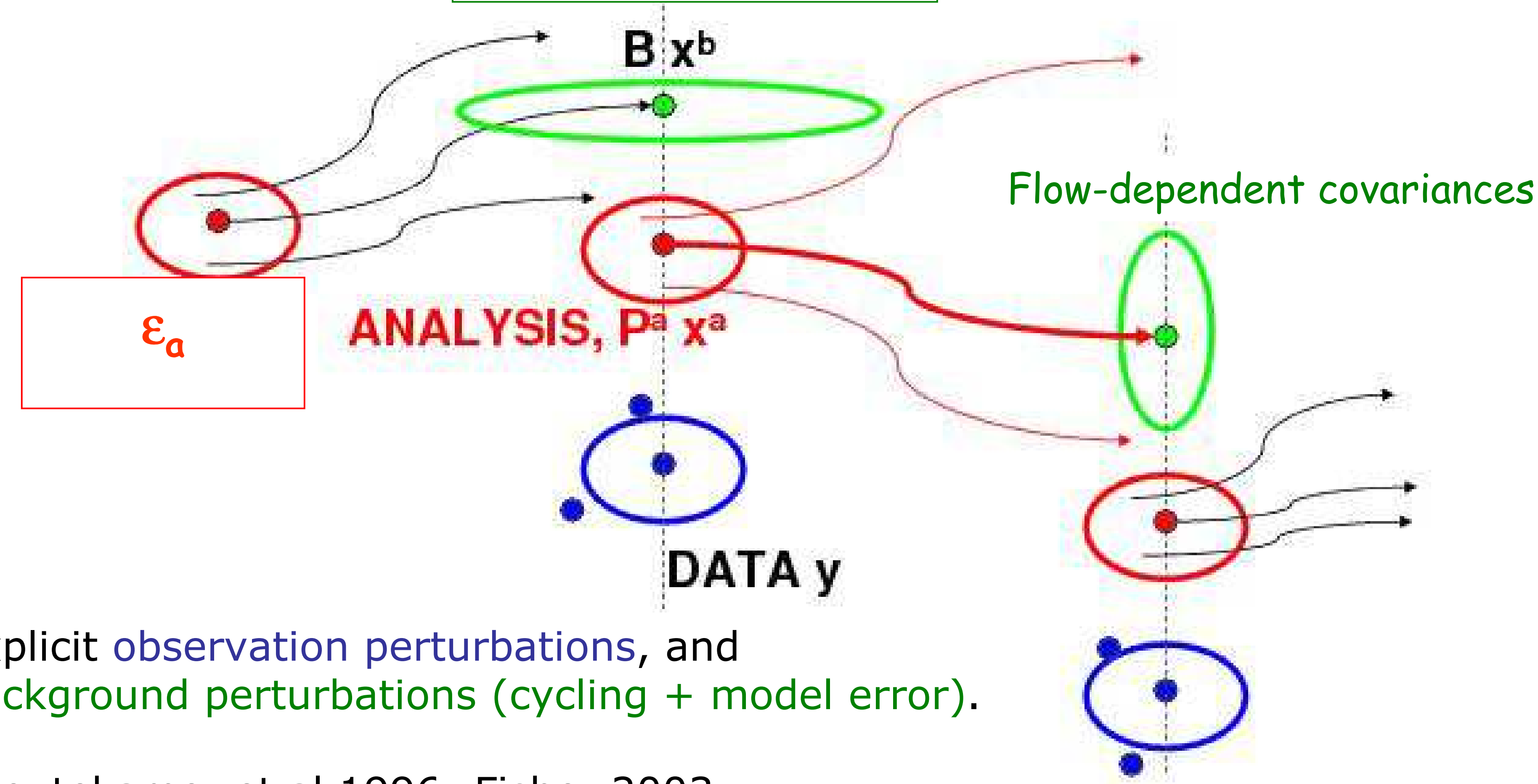
$$\mathbf{e_f = M\, e_a + e_m}$$

$\Rightarrow$ Ensemble DA : simulate this error cycling, via evolution of (cycled) perturbations of observations (ε_o) and of model (ε_m).

METEO FRANCE

Ensemble DA : simulation of error cycling

$$\varepsilon_f = M\,\varepsilon_a + \varepsilon_m$$



Explicit observation perturbations, and background perturbations (cycling + model error).

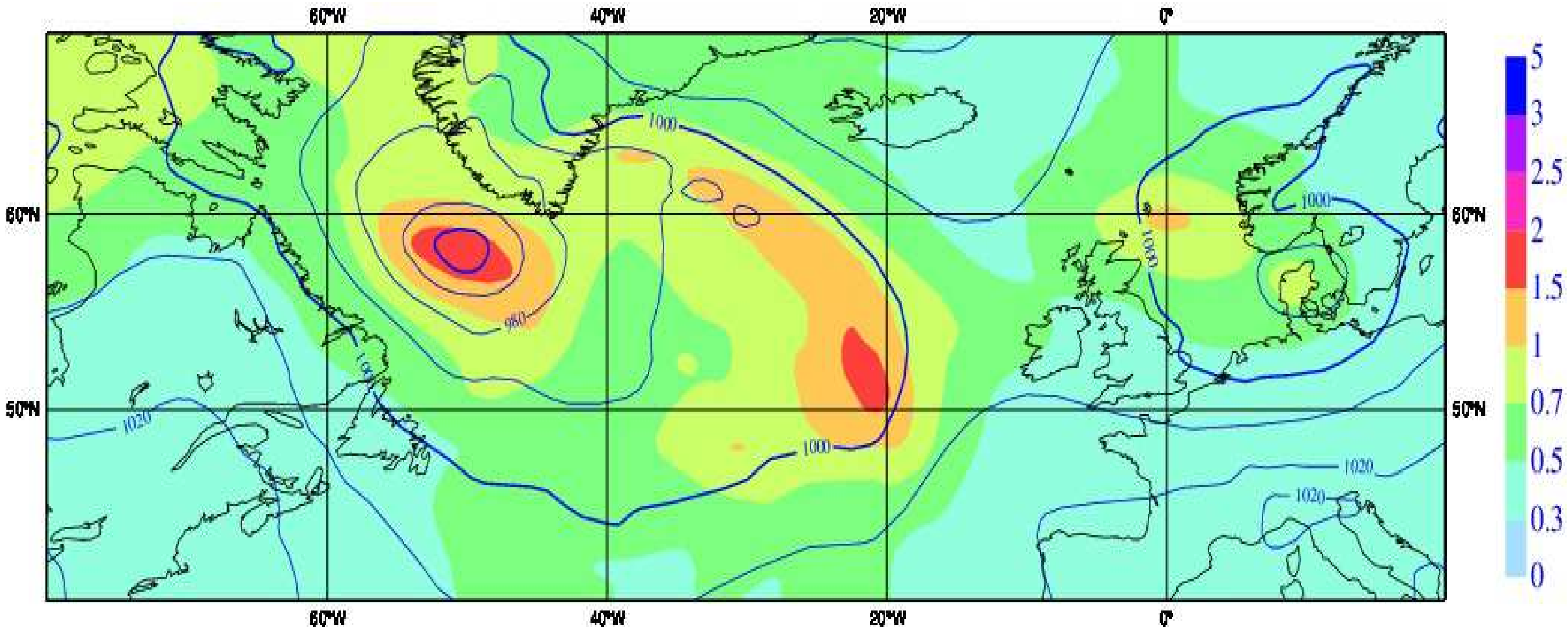
(Houtekamer et al 1996; Fisher 2003 ; Ehrendorfer 2006 ; Berre et al 2006)

METEO FRANCE

Ensemble 4D-Var at Météo-France (global)

- 25 members with 4D-Var, T479 (40 km) L105, Arpege (minim T149).
- Perturbations of 4D-Var analyses :
  obs perturbs. (drawn from **R**) and
  background perturbs (cycling of analysis perturbs and model perturbs).
- Inflation (cycled) of forecast perturbations (model errors),
  based on innovation estimates.
- Spatial filtering of flow-dependent variances,
  for minim. and obs. quality control.
- Wavelet filtering of flow-dependent 3D correlations,
  including sliding temporal average over 1 day 1/2
  (25 members x 6 cycles = total sample of 150 perturbations).

Dynamics of error variances
using Ensemble 4D-Var

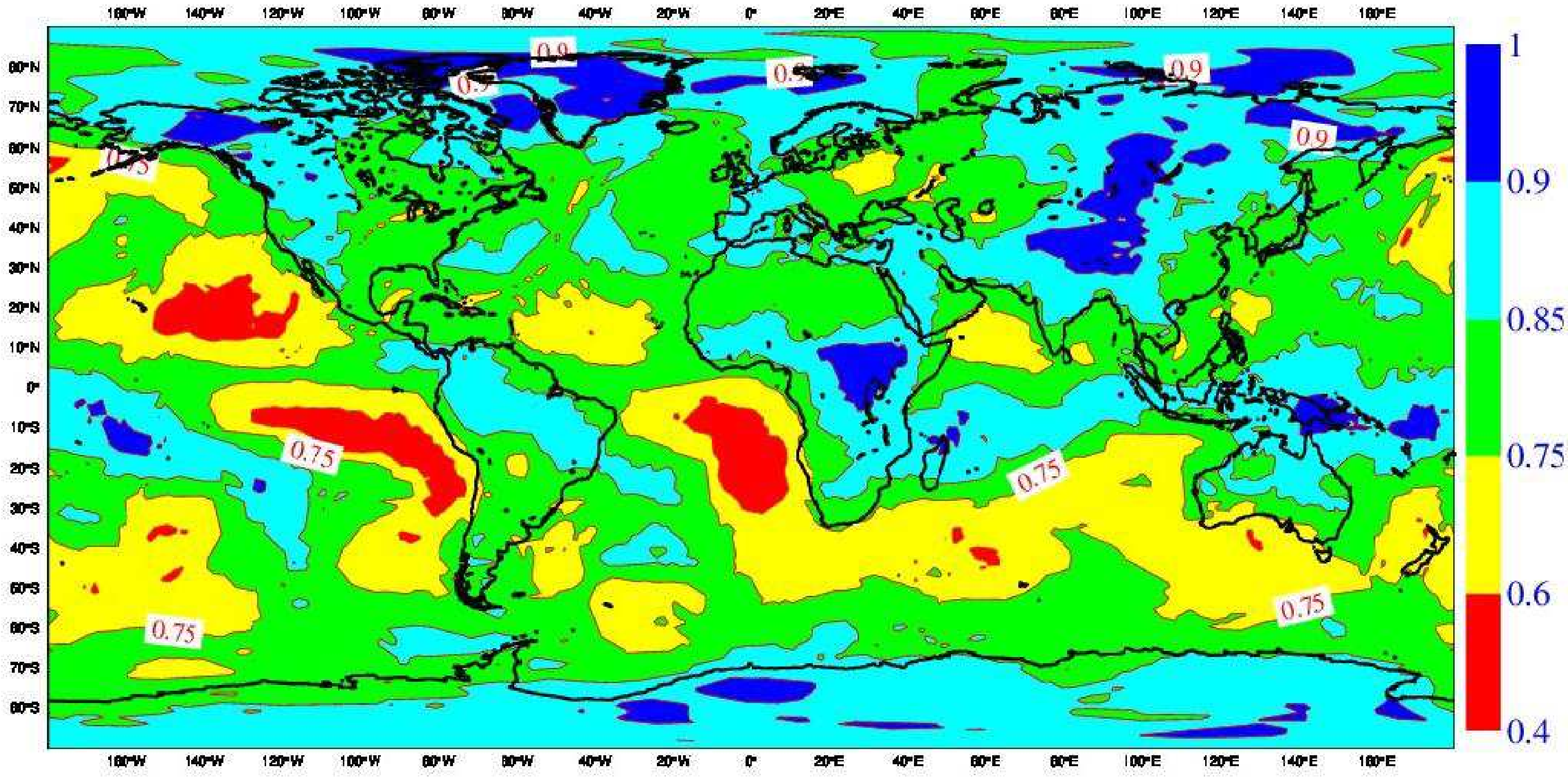


Error standard deviations of vorticity near 500 hPa (2/2/2010), spatially filtered (superimposed with geopotential).

(Berre et al 2007, Raynaud et al 2012)

METEO FRANCE

Dynamics of vertical correlations
using Ensemble 4D-Var and wavelet filtering



Vertical correlations of temperature between 850 & 870 hPa (28/2/2010)

(Berre et al 2015)

METEO FRANCE

Error simulation : some features and issues

- Requires plausible estimates of **R** and **Q** (often innovation-based ; ensemble not self-sufficient).

- Sampling noise : filtering of both variances and correlations.

- Consistence with the deterministic 4D-Var (ensemble of 4D-Var’s vs EnKF/ETKF), choice of horizontal resolution, …

- Representation of 4D-Var weakly non linear effects (outer loops, Desroziers et al 2009).

Outlook

- Simulation of error cycling in DA
- Innovation-based estimates
- Filtering of variances and correlations

Covariances of innovations

- Innovation vector :

$$\mathbf{y_o - H\,x_b = (y_o - H\,x_t) + (H\,x_t - H\,x_b)}$$
$$\mathbf{= \quad e_o - H\,e_b}$$

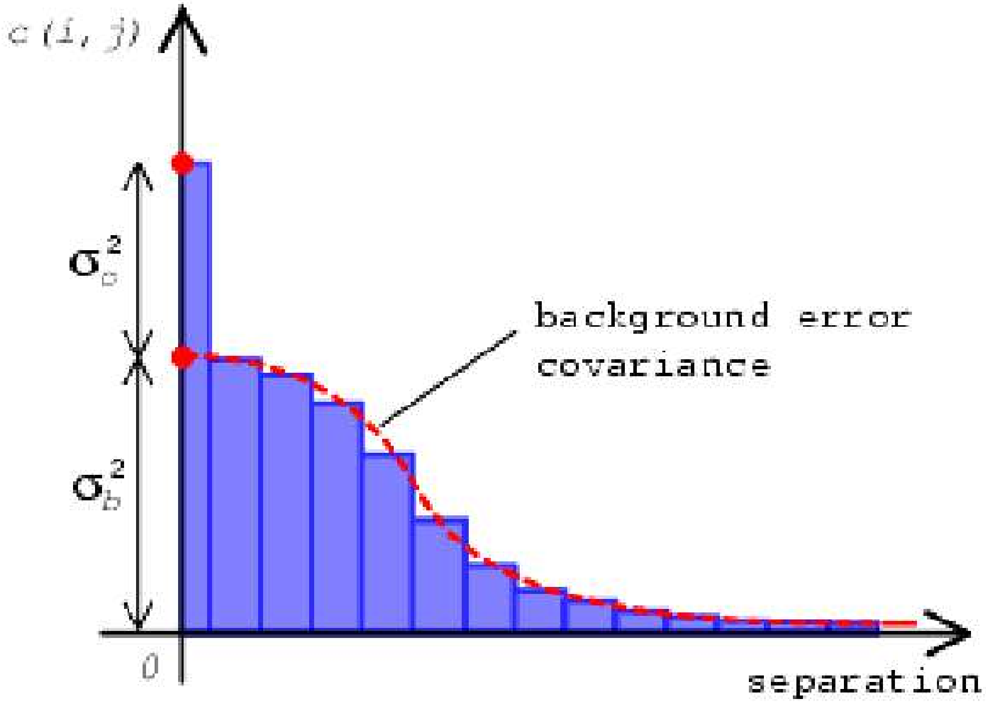
- Innovation covariances :

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

assuming that $\mathbf{E[(e_o)(He_b)^T]=0}$.

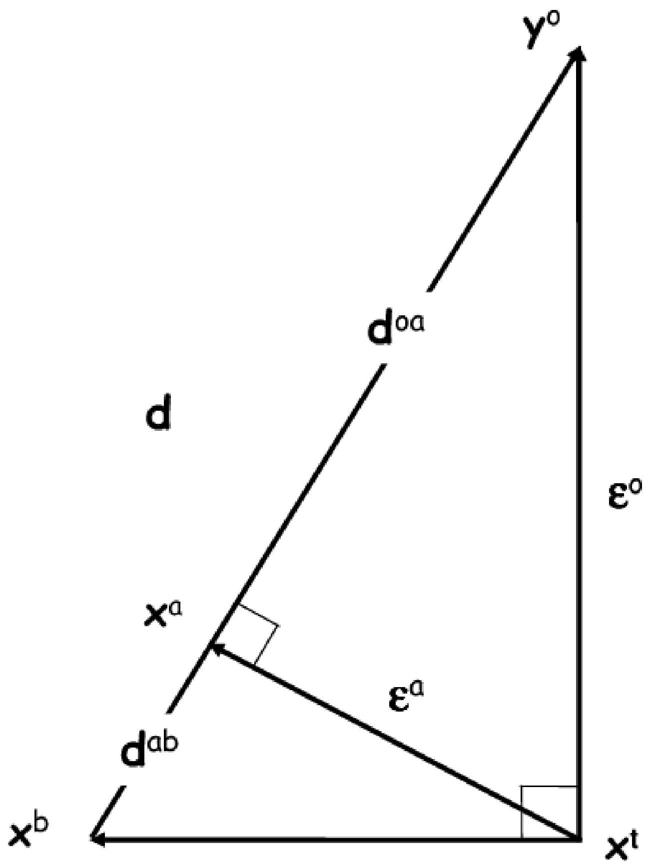
- Assumptions on spatial structures of errors in order to distinguish **R** and **B** contributions (e.g. Hollingsworth and Lönnberg 1986).

Hollingsworth and Lönnberg method



(From Bouttier and Courtier, ECMWF)

Diagnostics in observation space



(Desroziers et al, 2005)

- $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$
- $\mathbf{d}^{oa} = \mathbf{y}^o - H(\mathbf{x}^a)$
- $\mathbf{d}^{ab} = H(\mathbf{x}^a) - H(\mathbf{x}^b)$
- $E[\mathbf{d}^{oa}\ \mathbf{d}^T] = \mathbf{R}$
- $E[\mathbf{d}^{ab}\ \mathbf{d}^T] = \mathbf{HBH}^T$
- $E[\mathbf{d}^{ab}\ \mathbf{d}^{oaT}] = \mathbf{HAH}^T$
- $<\varepsilon, \varepsilon'> = E[\varepsilon\ \ \varepsilon'^T]$

Sample size issues for innovation methods

Ex: $\mathbf{cov(\ H\ dx\ ,\ dy\ )\ \sim\ E[H\ dx\ (dy)^T]\ \sim\ H\ B\ H^T}$

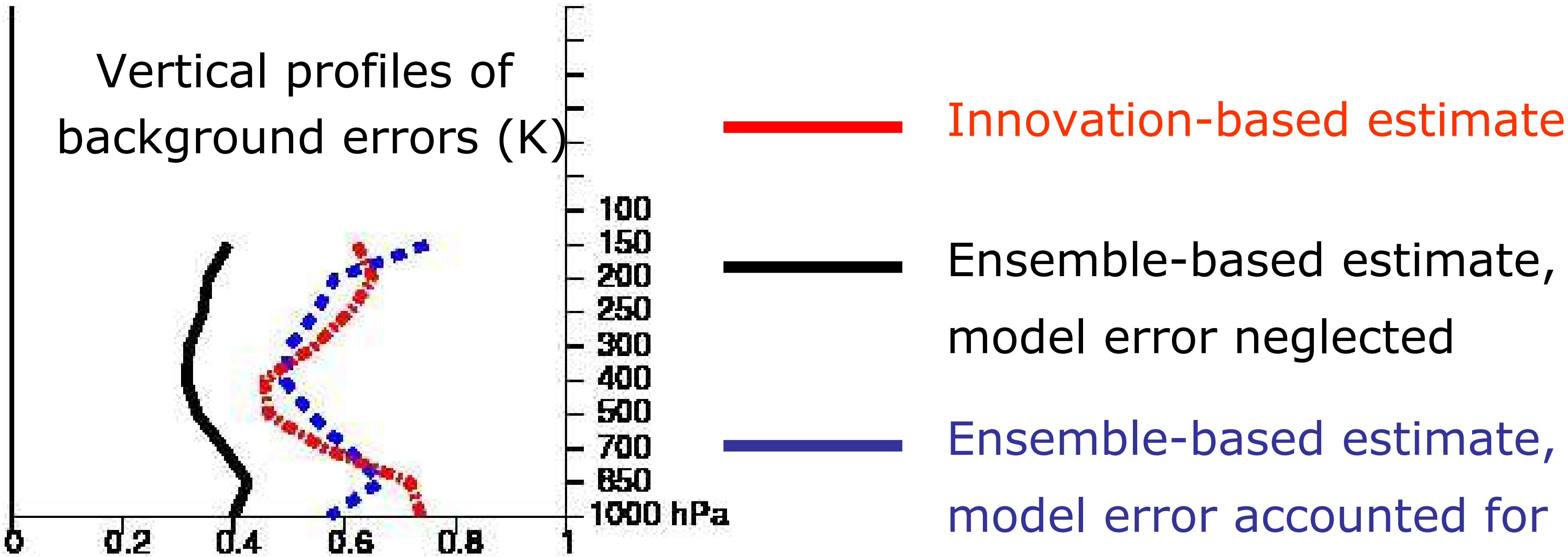
E[.] (statistical average) is « purely theoretical » :
only one realization of the observation vector is available (for a given date).

$\Rightarrow$ a single error realization, at a given time and location.

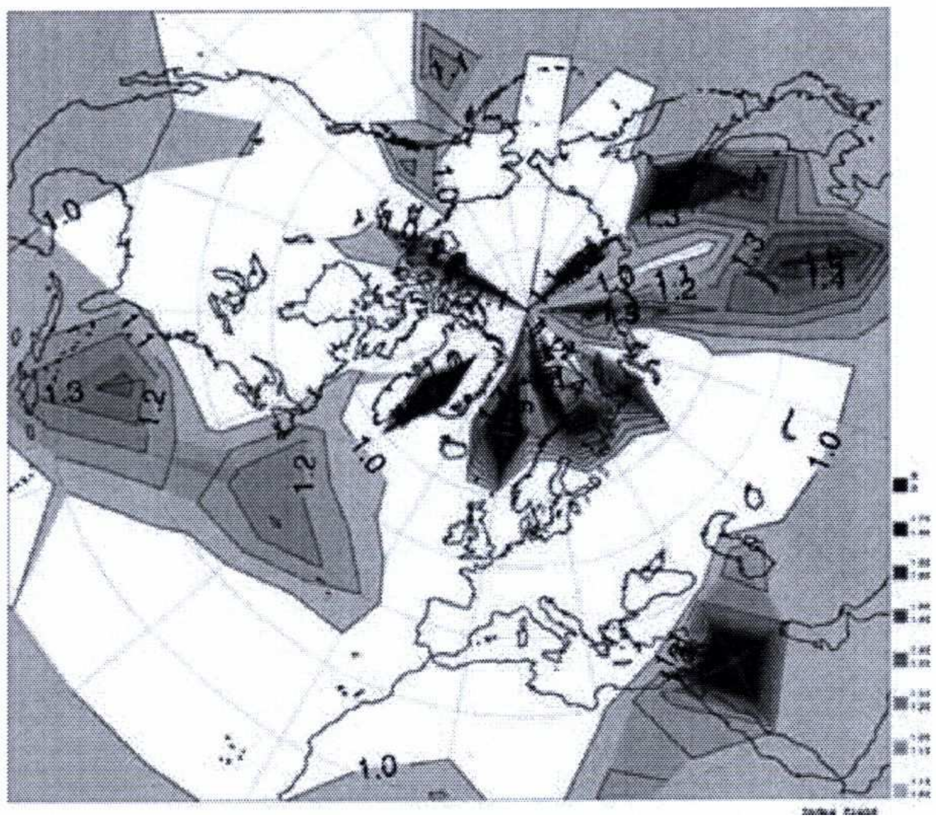
E[.] is usually replaced by spatial and/or temporal averages,
using ergodicity assumptions.

Amount of inferred information should be limited
(e.g. large scale variance structures in space and/or in time).

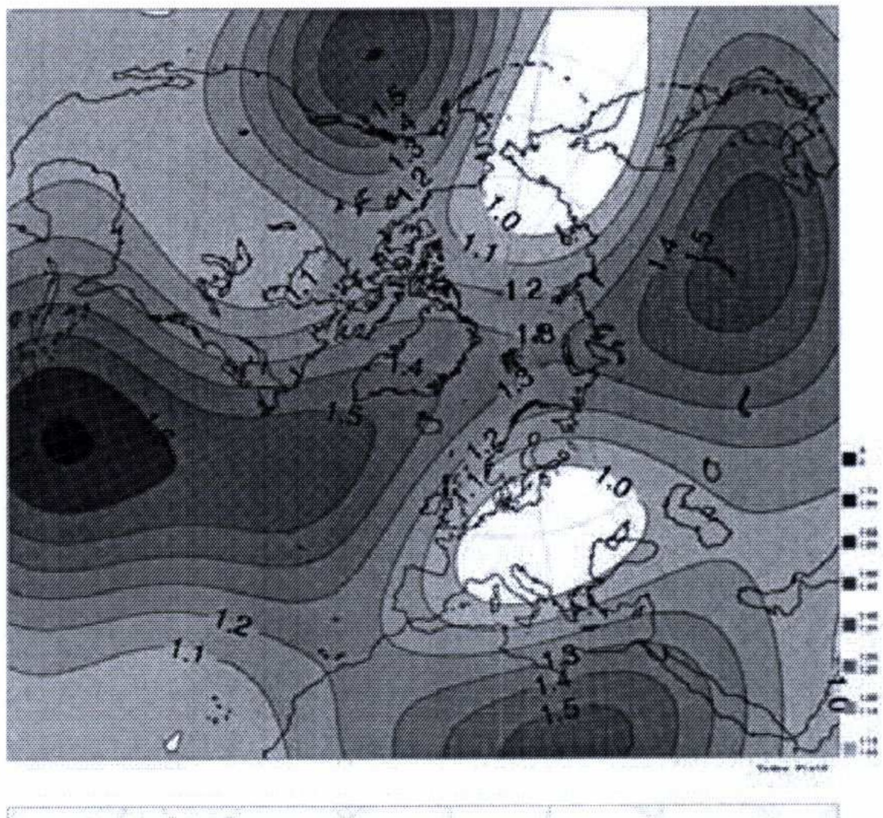
Use of innovations to estimate accumulated model error contributions (Raynaud et al 2012)



Spatial variations of innovation-based estimates
(Lindskog et al 2006 ; 1-year of surface pressure data)



« Raw » innovation-based variances

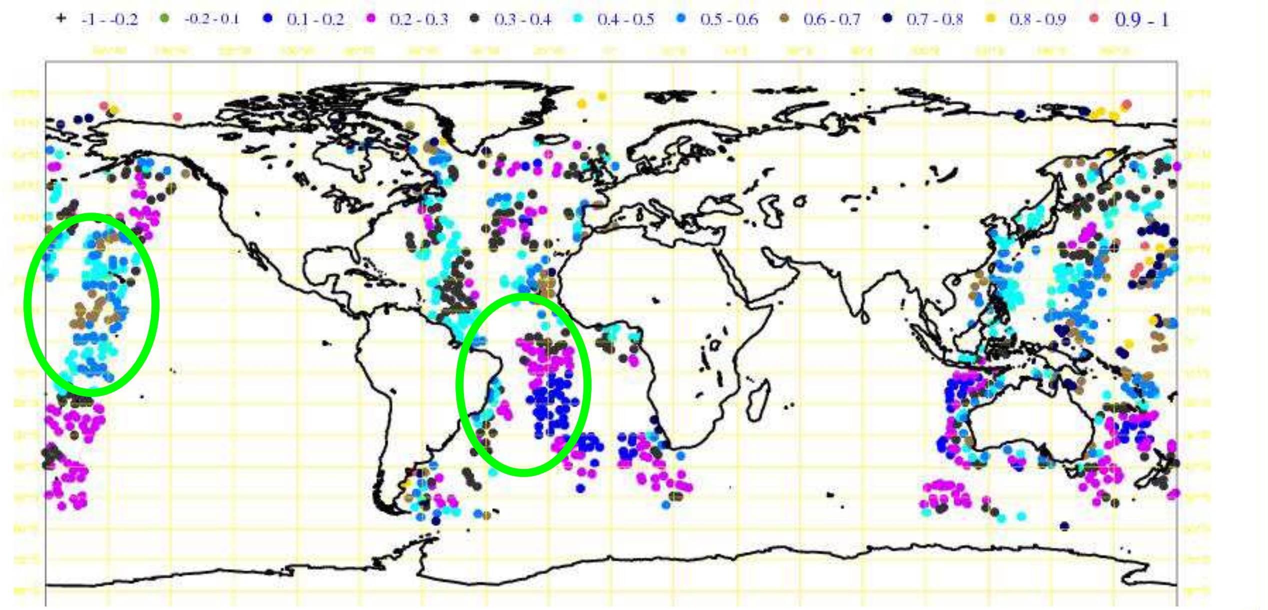


« Filtered » innovation-based variances

⇒ Some relevant geographical variations (e.g. data density effects), after spatial filtering ?

Background error estimates (sigmab's) in HIRS 7 space (28/8/2006 00h ; local averages over 500 km)

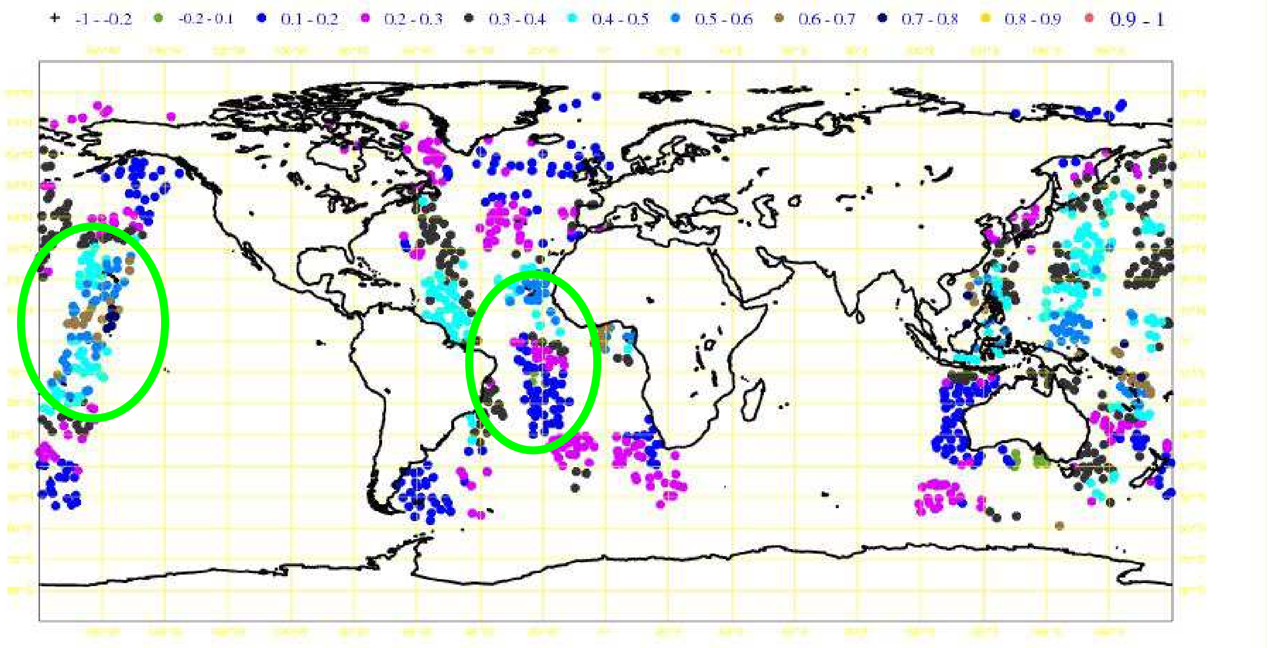
Ensemble spread



« Observed » std-dev's

cov(H dx , dy) ~ H B H^T

(Desroziers et al 2005)



(Berre et al 2007,2010)

Innovation method(s) : properties

- Provides estimates in observation space only.
- A good quality data dense network is needed.
- Assumptions on spatial scales (differences), in order to separate obs and bkd error contributions.
- Issues of sample size (time and/or space averaging).
- Source of information on **B** and **R**.

Outlook

- Simulation of error cycling in DA
- Innovation-based estimates
- Filtering of variances and correlations

Challenges for representing **B**

- Size of **B** is far too large (10^9 x 10^9 = 10^{18}).

- Can't be computed explicitly
  (nor stored in memory).

$\Rightarrow$ Model **B** as product of sparse operators.

- Finite ensemble size : sampling noise.

$\Rightarrow$ Apply filtering techniques.

Modelling of covariances in ensemble space

$$\mathbf{B}_{\mathbf{raw}} = \varepsilon_b \, \varepsilon_b^T / (N-1)$$

$$\mathbf{B}_{\mathbf{raw}}^{1/2} = \varepsilon_b / \sqrt{N-1}$$

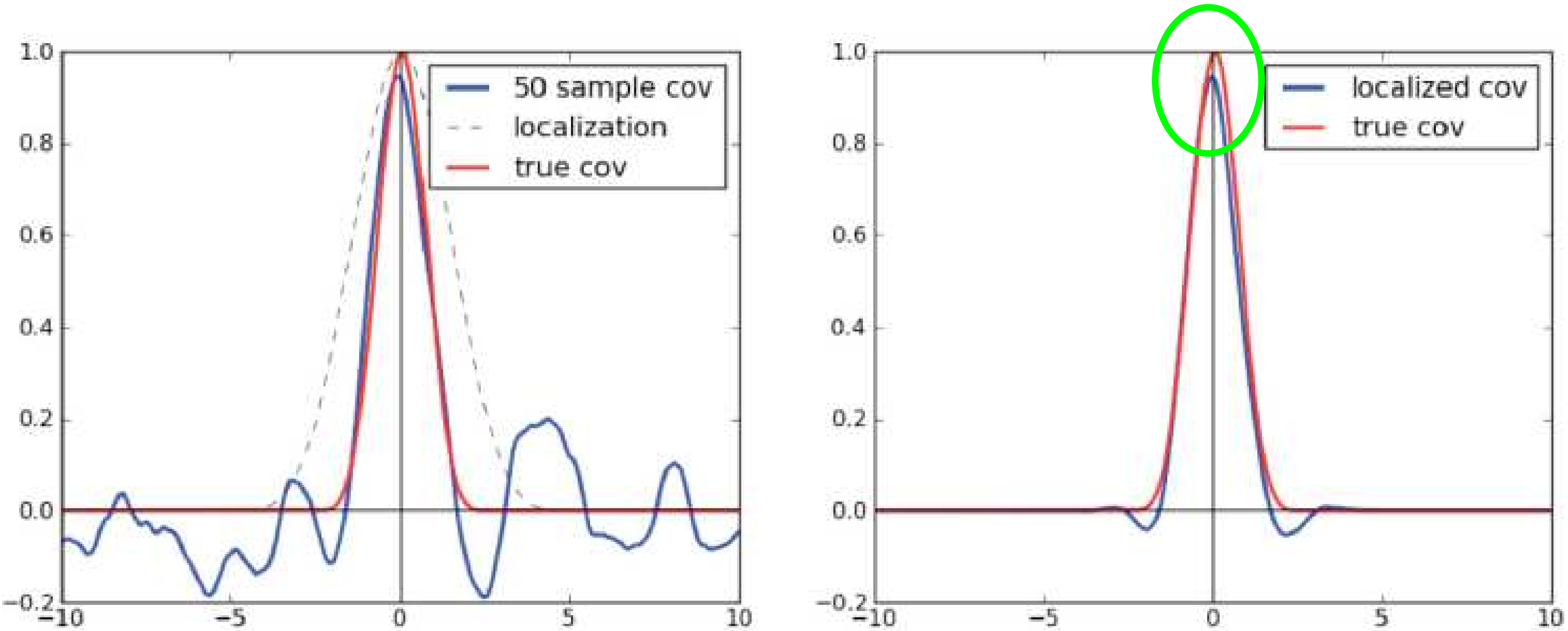
where perturbation matrix ε_b contains each member as a column (Lorenc 2003).

Schur filtering for long-distance correlations :

$$\mathbf{B} = \mathbf{B}_{\mathbf{raw}} \circ \mathbf{C}_{\mathbf{L}}$$

where $\mathbf{C}_{\mathbf{L}}$ is a localisation matrix (~ correlation model).

Schur filtering of covariances



(Figure Whitaker (2011), N=50)

METEO FRANCE

Schur filtering : some features and issues

- Attenuates noise in long-distance correlations.

- Does not account for noise in variances.

- Usual Schur filter is local (gridpoint by gridpoint) :
  could it be generalized to account for spatial coherences ?
  (e.g. Buehner 2012)

- Is all this related to the « practice » of
  a linear combination with modelled covariances ?

$$\mathbf{B} = \beta\, \mathbf{B}_{\mathbf{raw}} \circ \mathbf{C}_{\mathbf{L}} + (1-\beta)\, \mathbf{B}_{\mathbf{mod}}$$

Modelling and filtering of spatial covariances estimated from an ensemble

Decomposition of the covariance matrix :

$$\mathbf{B} = \mathbf{L}\,\Sigma\,\mathbf{C}\,\Sigma^T\,\mathbf{L}^T$$

with :

$\mathbf{L}$ ~ mass/wind cross-covariances,

including flow dependent nl balances (Fisher 2003).

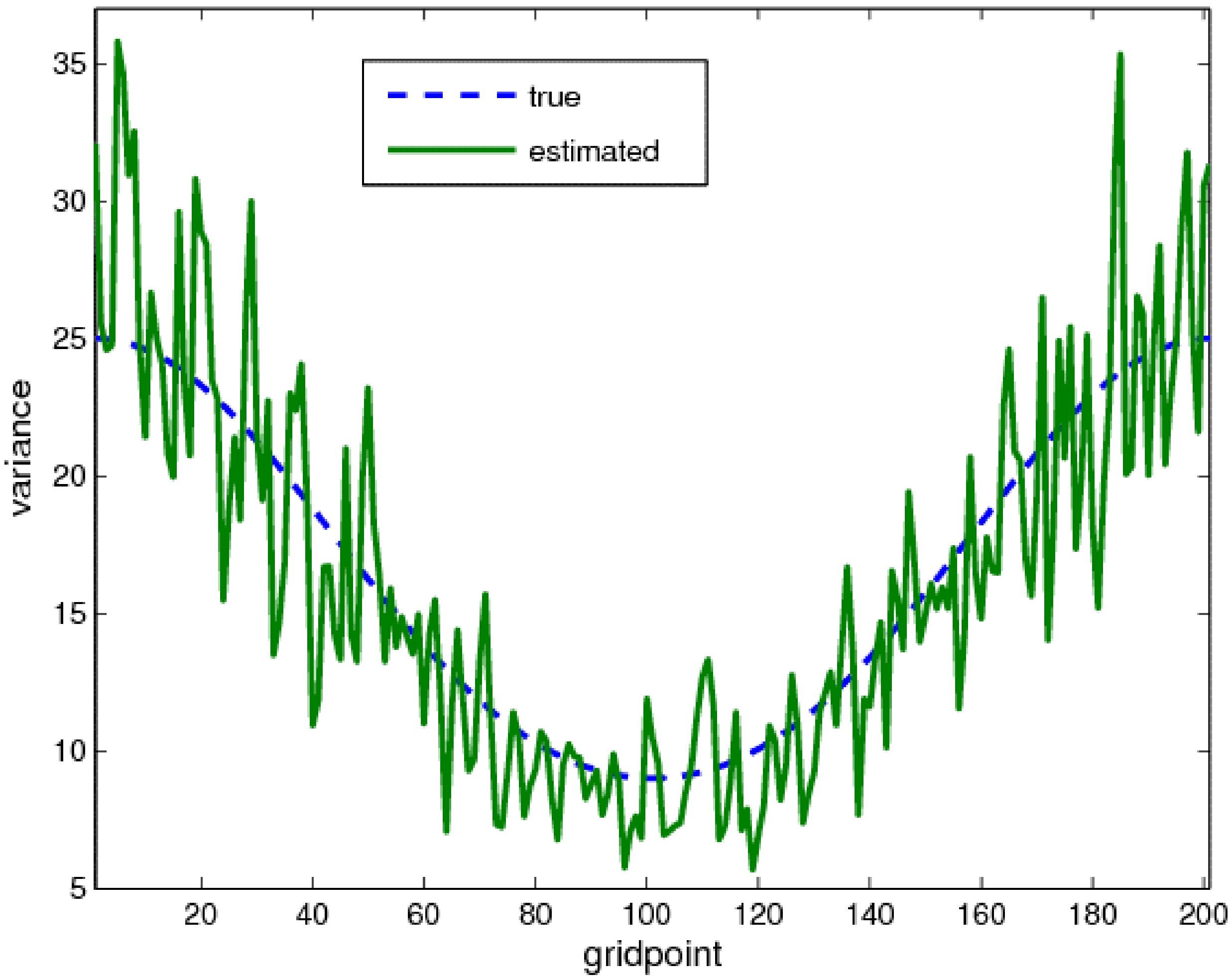
Σ error standard deviations,

spatially filtered (e.g. Berre et Desroziers 2010).

$\mathbf{C}$ matrix of 3D spatial correlations,

can be filtered in different ways.

Spatial structure of sampling noise for variances (Raynaud et al 2008)



$\varepsilon_b = B^{1/2}\,\eta$

$\eta \sim \mathcal{N}(0,1)$

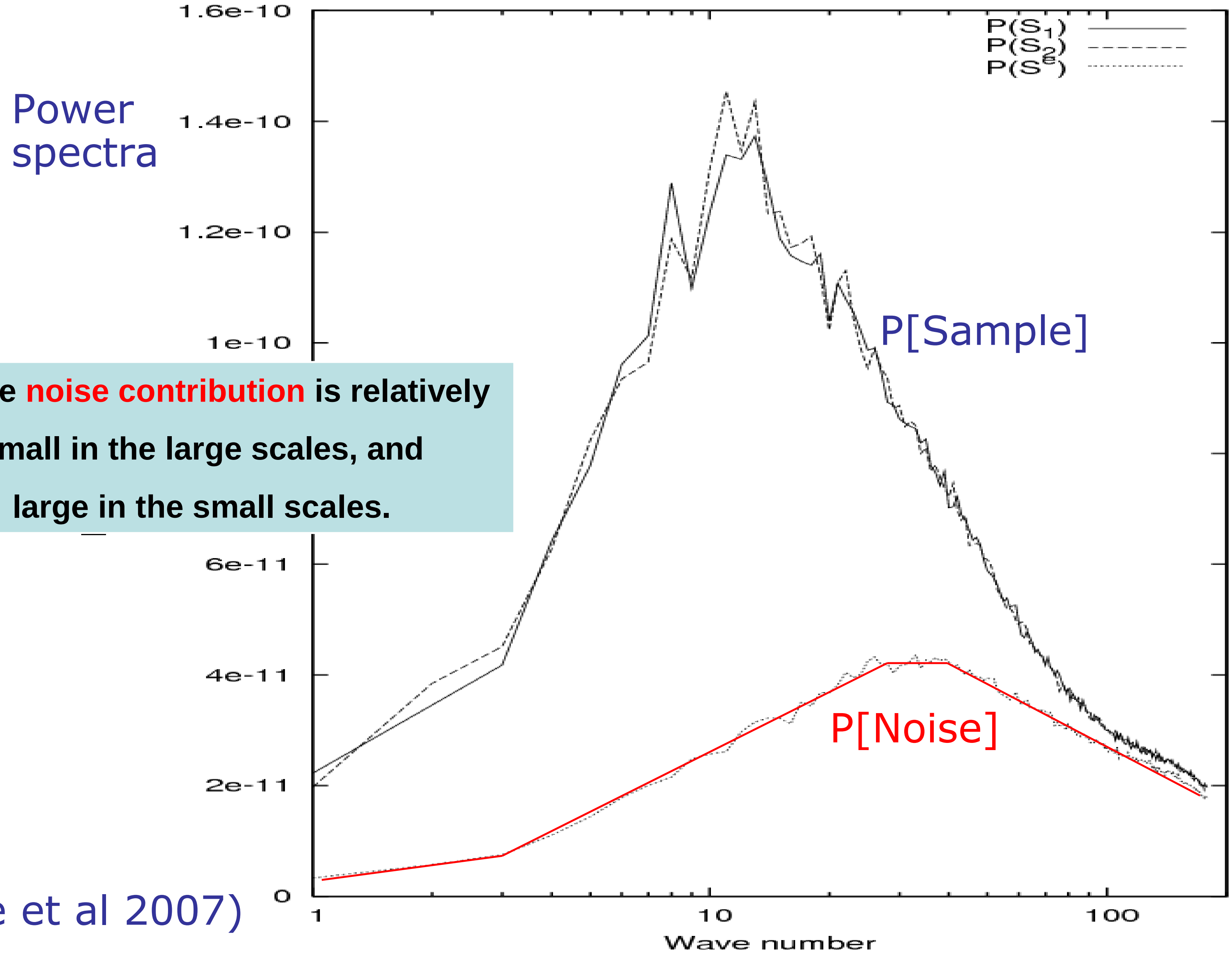
N = 50 members

L(ε_b) = 200 km

$\Rightarrow$ While the signal of interest is large scale, the sampling noise is small scale.

METEO FRANCE

CONTRIBUTION OF SIGNAL AND NOISE TO POWER SPECTRA OF RAW VARIANCE FIELDS



=> The noise contribution is relatively small in the large scales, and large in the small scales.

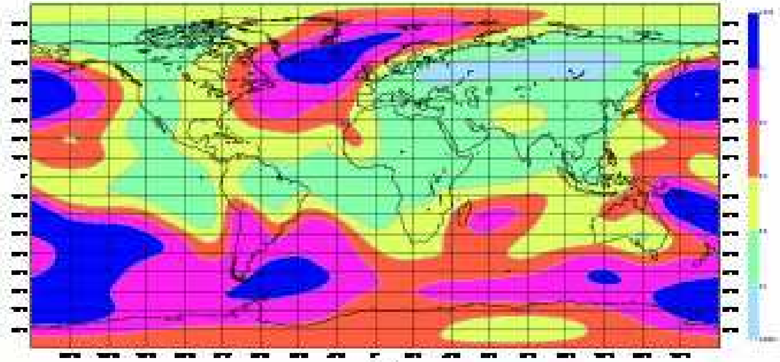
(Berre et al 2007)

"OPTIMIZED" SPATIAL FILTERING

OF THE VARIANCE FIELD

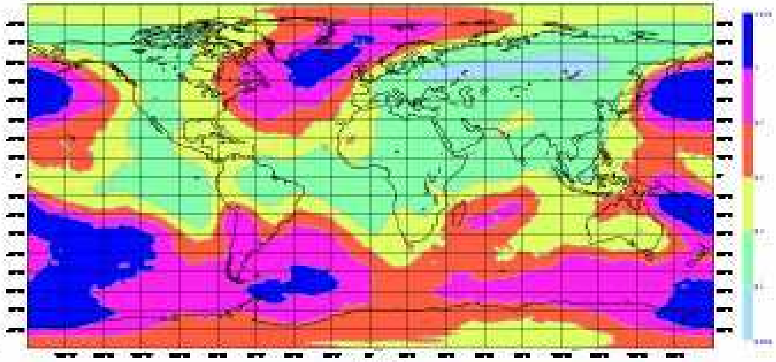
(Berre et al 2007,2010,
Raynaud et al 2008,2009,
Ménétrier et al 2015a,b)

« TRUE » VARIANCES

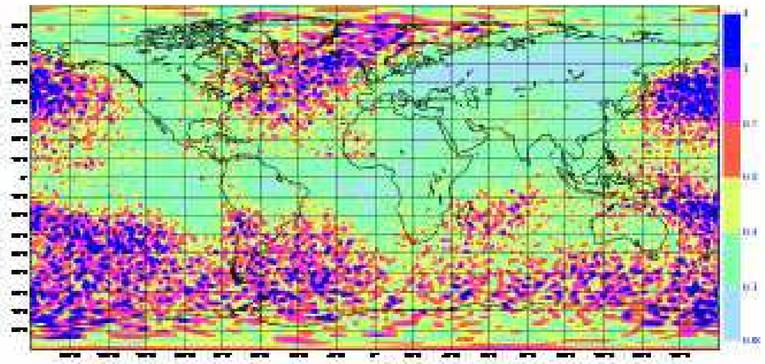


(a)

FILTERED VARIANCES (N = 6)



(b)



(c)

RAW VARIANCES (N = 6)

$V_b^* \sim F\ V_b$; min° of RMSE (spectral):

$F = 1 / (1 + E[noise^2]/signal^2)$

with $E[sample^2] = signal^2 + E[noise^2]$

Modelling and filtering of spatial correlations

- Spectral modelling :

  homogeneous correlations, but robust

  (global spatial averaging).

- Raw ensemble :

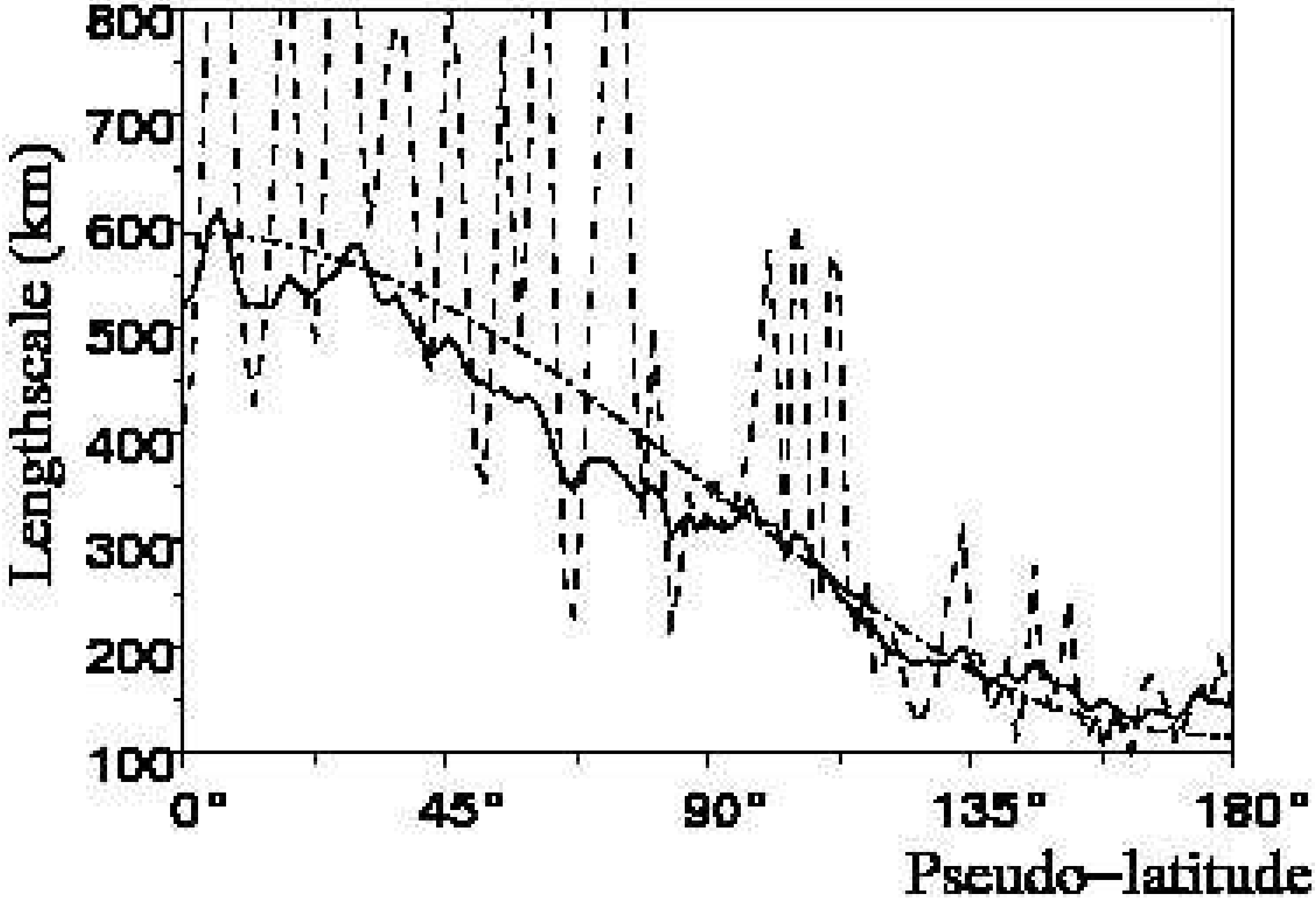
  very heterogeneous correlations, but noisy.

- Wavelet modelling :

  heterogeneous correlations, and robust

  (local spatial averaging).

Wavelet filtering of correlations
(1D illustration)



……… Exact correlations

- - - Raw correlations

—— Wavelet correlations

Ensemble size :
N = 10

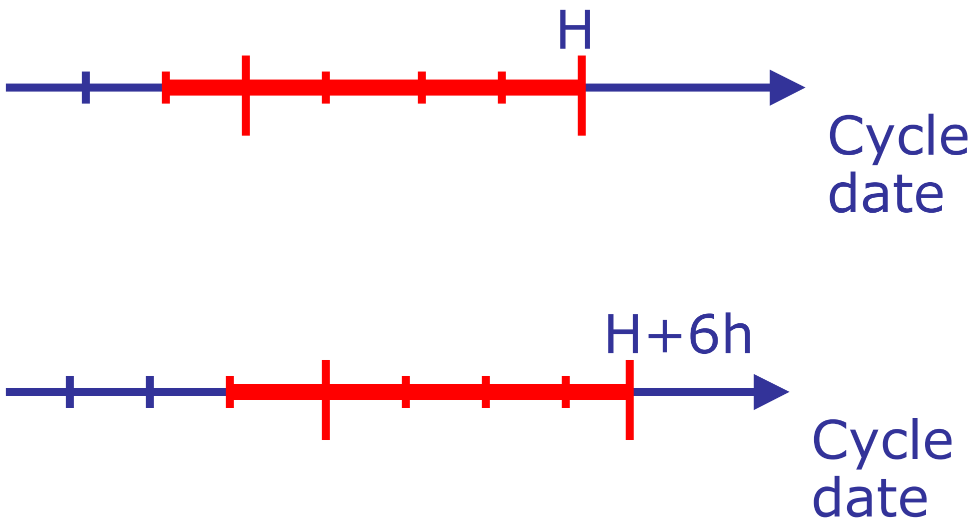
Spatial variations of correlation length-scales

(Fisher 2003, Pannekoucke et al 2007, Berre and Desroziers 2010)

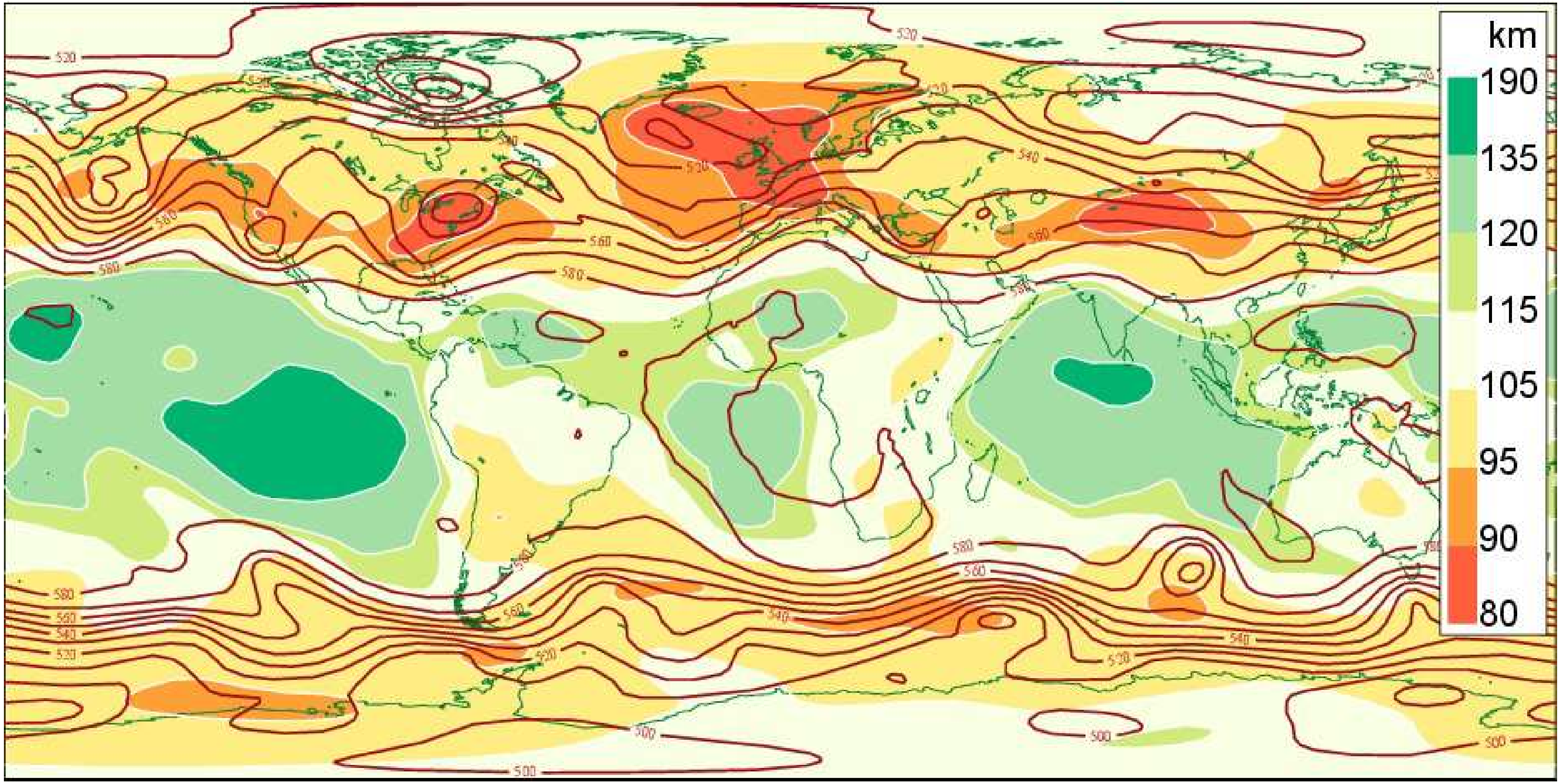
METEO FRANCE

Sliding temporal average of correlations over 6 cycles (1 day 1/2) x 25 members

25 members x 6 cycles= 150 perturbed forecasts



Horizontal correlation length-scales (diagnostics) using Ensemble 4D-Var and wavelets (28/2/2010)



Length-scales (in km) for wind near 500 hPa, superimposed with geopotential.

(Berre et al 2015)

METEO FRANCE

Conclusions

- Ensemble DA : flow-dependent covariances ;
  requires plausible estimates of **R,Q** ; consistence with 4D-Var.

- Innovation-based information for estimates of **R,Q** ;
  obs/bkd error separation ; space/time averaging.

- Filtering of both variances and correlations ;
  should account for spatial coherences.

- 4D aspects of ensemble covariances (4DEnVar).

References

Berre, L., S.E., Stefanescu, and M. Belo Pereira, 2006:
The representation of the analysis effect in three error simulation techniques.
Tellus, 58A, 196-209.

Berre, L., O. Pannekoucke, G. Desroziers, S.E. Stefanescu, B. Chapnik and L. Raynaud, 2007:
A variational assimilation ensemble and the spatial filtering of its error covariances:
increase of sample size by local spatial averaging. Proceedings of the ECMWF workshop
on flow-dependent aspects of data assimilation, 11-13 June 2007, 151-168.
(available on line at: http://www.ecmwf.int/publications/library/do/references/list/14092007)

Berre and Desroziers 2010, MWR.

Berre et al 2015, QJRMS.

Bonavita et al 2011,2012 QJRMS.

Buehner 2012, MWR.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005 :
Diagnosis of observation, background, and analysis error statistics in observation space,
Quart. Jour. Roy. Meteor. Soc., 131, pp. 3385-3396.

Desroziers et al 2009, MWR.

Ehrendorfer, M., 2006 :
Review of issues concerning Ensemble-Based data assimilation techniques.
Oral presentation at the Seventh Adjoint Workshop, Obergurgl, Austria.

Fisher, M., 2003 :
Background error covariance modelling.
Proceedings of the ECMWF seminar on recent developments in data assimilation
for atmosphere and ocean, 45-63.

References

Hamill, T., 2008 :
Chapter 6 of "*Predictability of Weather and Climate*".
See oral presentation at WMO Buenos Aires workshop in 2008.

Hollingsworth and Lönnberg 1986 :
The statistical structure of short-range forecast errors as determined from radiosonde data.
Part I: The wind field. Tellus A, 38A: 111-136.

Houtekamer, P., , L. LeFaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996:
A system simulation approach to ensemble prediction. Mon. Wea. Rev., 124, 1225–1242.

Lindskog et al 2006 :
Representation of background error standard deviations in a limited area model data assimilation system.
Tellus A, 58: 430–444..

Lorenc 2003, QJRMS.

Ménétrier et al 2015 (parts I and II), MWR.

Pannekoucke, O, L. Berre and G. Desroziers, 2007 :
Filtering properties of wavelets for the local background error correlations.
Quart. Jour. Roy. Meteor.Soc.133, 363-379

Raynaud L., L. Berre et G. Desroziers, 2008 :
Spatial averaging of ensemble-based background error variances.
Q. J. R. Meteorol. Soc., 134, 1003-1014.

Raynaud L., L. Berre et G. Desroziers, 2009 :
Objective filtering of ensemble-based background error variances.
Q. J. R. Meteorol. Soc.

Raynaud et al 2012 QJRMS.

Thank you

for your attention

METEO FRANCE