On-line observation covariance matrix tuning based on optimality diagnostic

Bernard Chapnik

A Few preliminary words

- More and more observations, and observation types.

- Need for fast and efficient procedures to prescribe correct weights (in DA).

- Also need monitoring tools to understand how the observations impact the analysis.

Outline

- Introduction.
- J min diagnostic.
  - The J min diagnostic.
  - Desroziers and Ivanov (2001) algorithm.
- Decorrelation of the analysis error and the innovation.
  -The diagnostic.
  -Desroziers et al. (2005) algorithm.
- Conclusion and future plans.

J min diagnostic

Some notations, Tarantola (1988)

$$z = \begin{pmatrix} x_b \\ y \end{pmatrix} = \begin{pmatrix} x_t + \epsilon_b \\ H x_t + \epsilon_o \end{pmatrix} = \Gamma x_t + \epsilon,$$

$$E(\epsilon) = 0,\ E(\epsilon \epsilon^T) = S = \begin{pmatrix} B & 0 \\ 0 & R \end{pmatrix}$$

Variational framework

$$J(x) = (z - \Gamma x)^T S^{-1} (z - \Gamma x)$$

For any subpart z_i of the data vector with size k_i

$$J_i(x) = (z_i - \Gamma_i x)^T S_i^{-1} (z_i - \Gamma_i x)$$

Talagrand (1999), if the DA system is consistent :

$$E(J_i(x_a)) = k_i - Trace\left(S_i^{\frac{-1}{2}} \Gamma_i P_a \Gamma_i^T S_i^{\frac{-1}{2}}\right)$$

If data set = obs. $E(J_o(x_a)) = p - Trace\left(R^{\frac{-1}{2}} H P_a H^T R^{\frac{-1}{2}}\right) = Trace(I_p - HK)$

If data set = guess $E(J_b(x_a)) = n - Trace\left(B^{\frac{-1}{2}} P_a B^{\frac{-1}{2}}\right) = Trace(KH)$

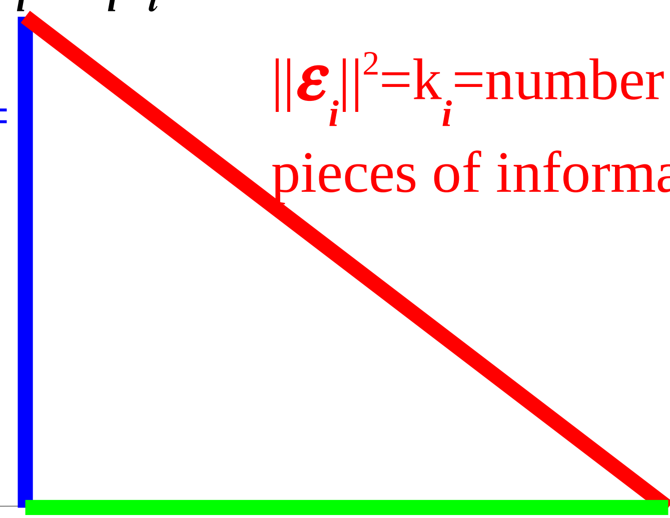
Whole dataset $E(J(x_a)) = p$

J min diagnostic

Geometrical interpretation

Scalar product : $<u,v>=E(u^T S_i^{-1} v)$

$z_i = \Gamma_i x_t + \varepsilon$



$\|z_i - \Gamma_i x_a\|^2 = E(J_i(x_a)) =$ $k_i - DFS_i$, not used by the analysis

$\|\varepsilon_i\|^2 = k_i$ = number of pieces of information

$\Gamma_i x$ plan

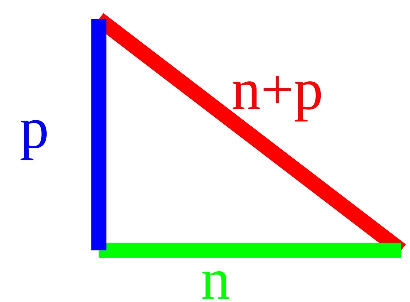
$\Gamma_i x_a$

$\Gamma_i x_t$

contribution to the analysis, Degrees of Freedom for Signal

$$\|\Gamma_i \epsilon_a\|^2 = Trace(S_i^{-\frac{1}{2}} \Gamma_i P_a \Gamma_i^T S_i^{-\frac{1}{2}}) = DFS_i$$

For the whole set of data



Desroziers and Ivanov (2001), the algorithm.

How to use this diagnostic in order to tune DA prescribed variances ?

→ apply multiplicative coefficients S_b and S_o to your B and R matrices (or S_{oi} and S_{bj} to any independent sub matrix R_i and B_j you can define) so that the J min criterion is fulfilled.

Let $\boldsymbol{x_a}(s_o, s_b)$ be the minimizer of $J_s = \frac{J_o}{s_o} + \frac{J_b}{s_b}$.

Supposing the optimality condition is achieved then

$$E\left(\frac{J_o(\boldsymbol{x}_a(s_o, s_b))}{s_o}\right) = Trace(\boldsymbol{I_p} - \boldsymbol{HK}(s_o, s_b)), \; E\left(\frac{J_b(\boldsymbol{x}_a(s_o, s_b))}{s_b}\right) = Trace(\boldsymbol{K}(s_o, s_b)\boldsymbol{H}).$$

Therefore $$s_o = \frac{J_o(\boldsymbol{x}_a(s_o, s_b))}{Trace(\boldsymbol{I_p} - \boldsymbol{HK}(s_o, s_b))}, \; s_b = \frac{J_b(\boldsymbol{x}_a(s_o, s_b))}{Trace(\boldsymbol{K}(s_o, s_b)\boldsymbol{H})}.$$

Thus defining the fixed point relation we use to compute
The tuning coefficients

METEO FRANCE
Toujours un temps d'avance

Desroziers and Ivanov, trace computation.

Problem : How do we compute the Trace(...) term ?
Three randomized methods...

1) Girard (1987)'s method : compute a perturbed analysis xa* with perturbed observations y*= y + a noise consistent with **R**, then it can be shown that

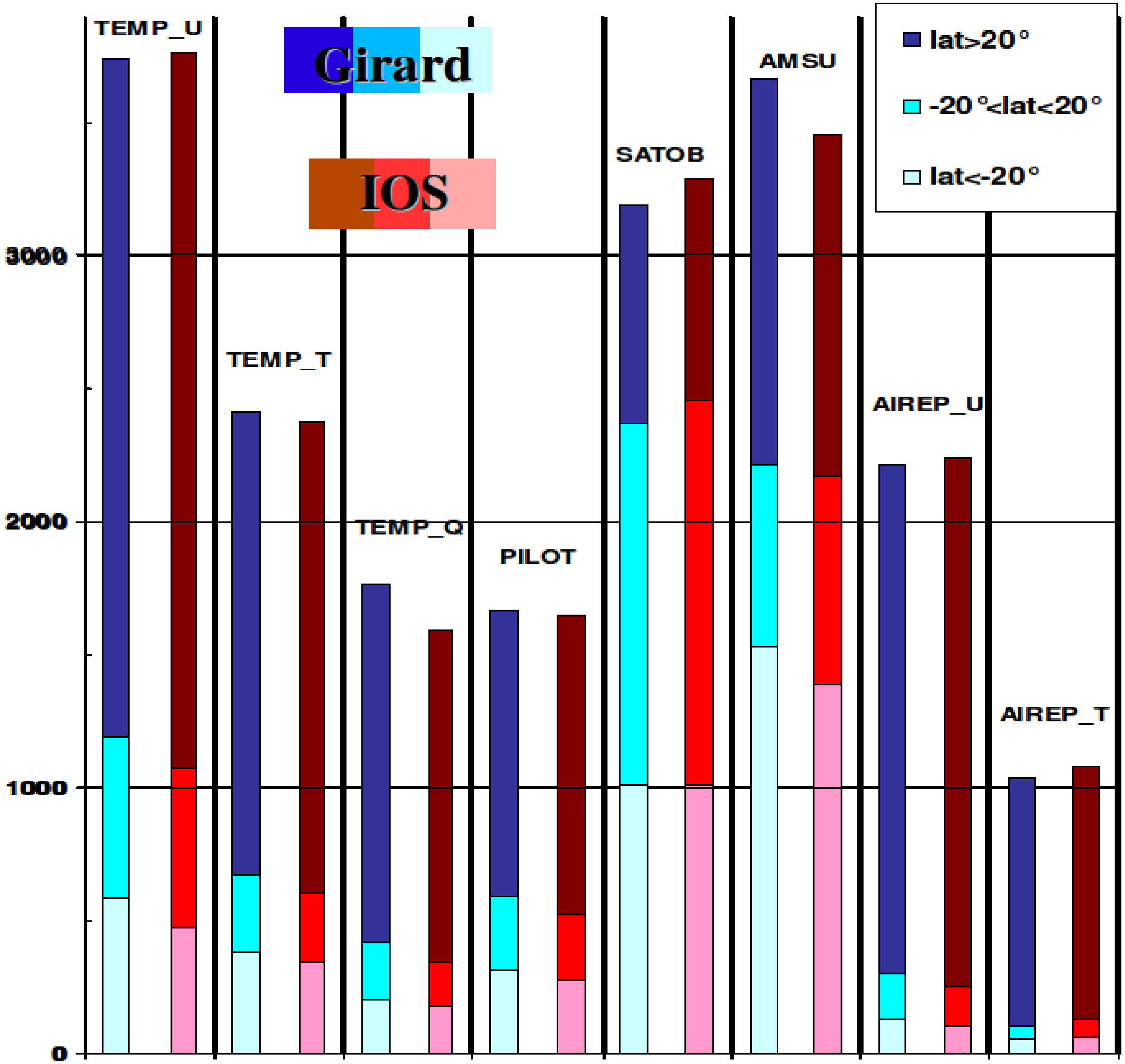
$$(\boldsymbol{y}^* - \boldsymbol{y})^T \boldsymbol{R}^{-1} (\boldsymbol{Hx}_a^* - \boldsymbol{Hx}_a) \simeq Trace(\boldsymbol{HK}).$$

2) Simulated Optimal Innovation method (Chapnik et al, 2006) : take a state vector you will consider as your truth, from this « truth » generate a background and observations by adding some noise consistent with B and R, then perform a (variational) analysis. The system being consistent, the subparts of the cost function at the minimum approximate their expectations (if the dataset is big enough).

3) Girard revisited in ensemble framework (Desroziers et al, 2009). If you have an assimilation ensemble based on observation perturbation, then a Girard-like computation can be performed at no significant extra-cost.

Moreover all these methods give access to DFS computation...

Desroziers and Ivanov, trace computation.



Upper air observations DFS computed by two methods on 02/04/2004 at 00Z.

Desroziers and Ivanov, properties of the method.

The method can be shown to achieve a maximum likelihood tuning of the tuning coefficients : i.e. considering a gaussian pdf for the innovation ($\boldsymbol{d}=\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_b$) with covariance $\boldsymbol{D}=s_o\boldsymbol{R}+s_b\boldsymbol{HBH}^T$, The coefficients provided by the method maximize

$$f(\boldsymbol{d}|s_o,s_b)=\frac{1}{\sqrt{(2\pi)^p det(\boldsymbol{D})}}\exp\left(\frac{-\boldsymbol{d}^T\boldsymbol{D}^{-1}\boldsymbol{d}}{2}\right).$$

This has some consequences :

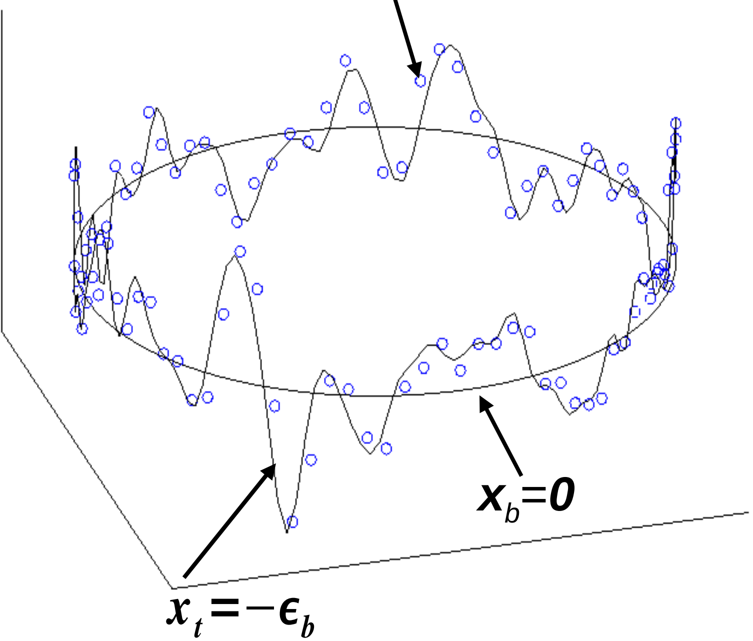
- Need of a large observation set (ok for a global tuning of a satellite sounder channel, not for RS at a certain level for a certain parameter in a limited geographical area);

- B and R must be very different in terms of correlation length scale, both for the actual covariances and the prescribed covariance model;

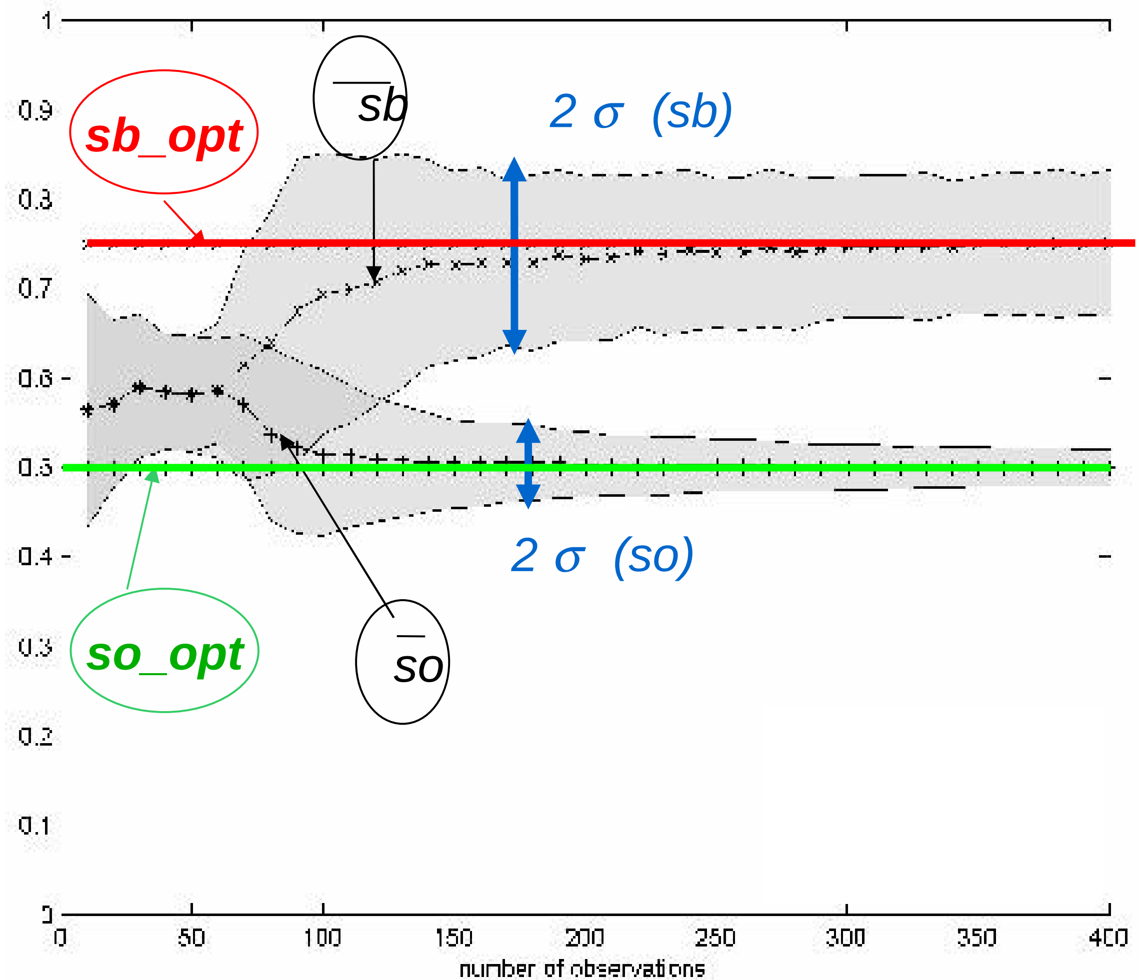
- so and sb can be tuned independently.

Independently of ML equivalence : the algorithm converges very quickly (good estimate at the first step).

Desroziers and Ivanov, properties of the method



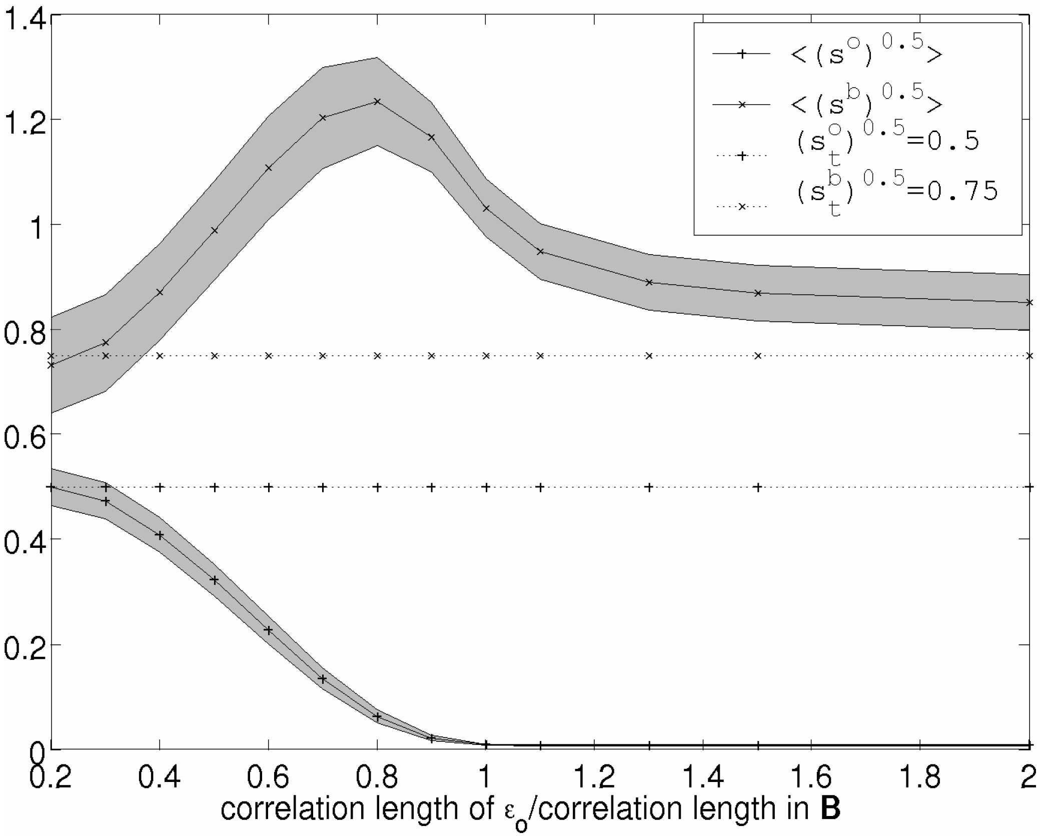
Experimental framework



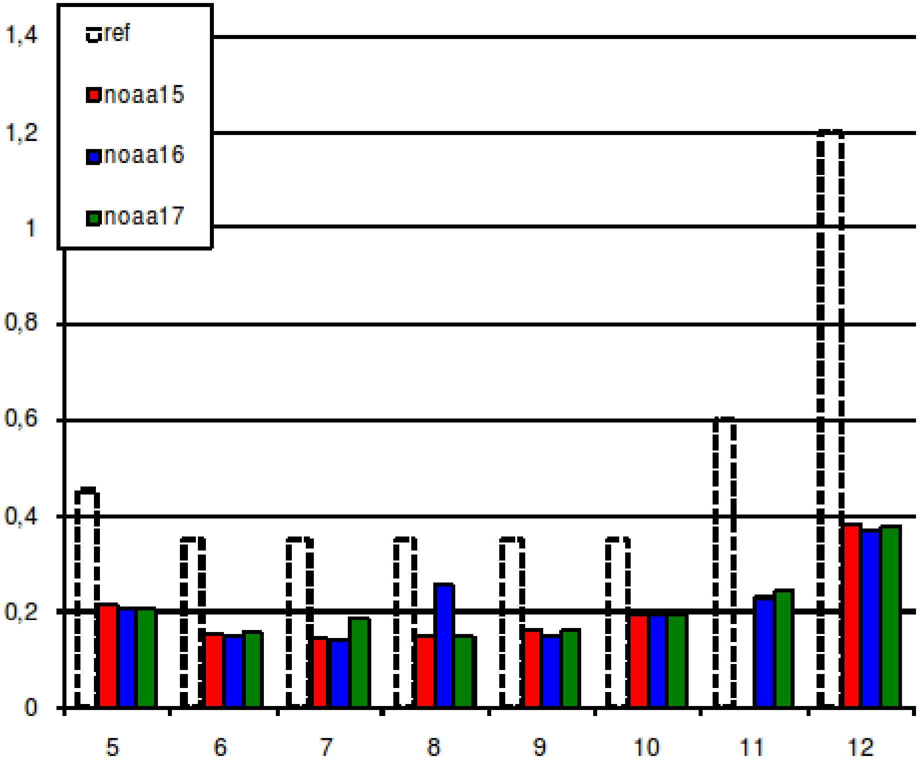
P<50, not enough observations,
HBH^T=α R.
s_o precision increases with p.

Desroziers and Ivanov, properties of the method

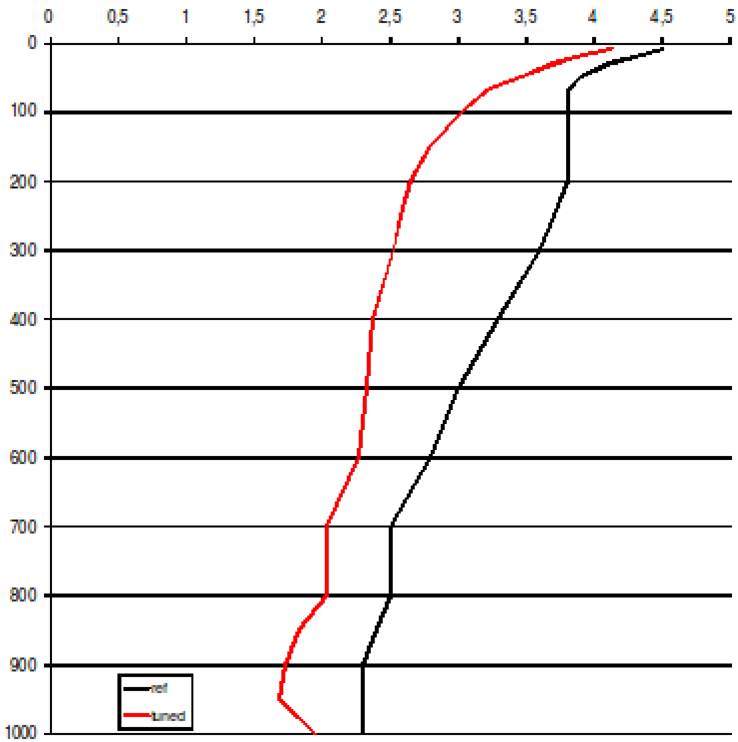
If obs. errors have spatial correlation...



Desroziers and Ivanov, examples of tuned σ_o .



AMSU-A



TEMP-UV

Decorrelation of the analysis error and the innovation.

The innovation vector and the analysis error should not be correlated in an optimal system.

$$E(\boldsymbol{\epsilon}_a \boldsymbol{d}^T) = \mathbf{0}$$

Already used by Daley (1992) under a slightly different form : if the system is optimal then the innovations computed at different steps of the Kalman filter are decorrelated, departure from optimality is measured by a “lagged innovation covariance”.

Talagrand (2003) suggested to assess the optimality of a DA system using this criterion, sampling the analysis error by its difference with unassimilated, independent verifying observations.

$$E(\boldsymbol{y}_v - \boldsymbol{H}_v \boldsymbol{x}_a)\boldsymbol{d}^T = E((\boldsymbol{\epsilon}_v - \boldsymbol{H}_v \boldsymbol{\epsilon}_a)\boldsymbol{d}^T) = \mathbf{0}$$

Using an equivalent “lagged analysis increment covariance”, Chapnik (2006) produced sub-optimality maps (see Gerald Desroziers's presentation).

Desroziers et al. (2005) algorithm

Based on the fact that the innovation vector and the analysis error should not be correlated in an optimal system.

$$E(\boldsymbol{\epsilon}_a \mathbf{d}^T) = \mathbf{0}$$

$$\Rightarrow \left\{ \begin{array}{c} E((\mathbf{y} - \mathbf{H}\mathbf{x}_a)\mathbf{d}^T) = (\boldsymbol{\epsilon}_o - \mathbf{H}\,\boldsymbol{\epsilon}_a)\mathbf{d}^T = \boldsymbol{\epsilon}_o \mathbf{d}^T = \mathbf{R} \\ E((\mathbf{x}_a - \mathbf{x}_b)\mathbf{d}^T) = E((\boldsymbol{\epsilon}_a - \boldsymbol{\epsilon}_b)\mathbf{d}^T) = E(\boldsymbol{\epsilon}_a \mathbf{d}^T) = \mathbf{B}\mathbf{H}^T \end{array} \right\}$$

Dropping the « E » term (with large enough an observation set), only considering these relations in observation space and considering only the trace of the matrices lead to the following relations :

$$\left\{ \begin{array}{c} \frac{1}{p}(\mathbf{y} - \mathbf{H}\mathbf{x}_a)^T \mathbf{d} = \sigma_o^2 \\ \frac{1}{p}(\mathbf{H}\,\boldsymbol{\delta}\,\mathbf{x}_a)^T \mathbf{d} = \sigma_b^2 \ \ in\, obs.\, space \end{array} \right\}$$

Desroziers et al. (2005)
Properties of the method

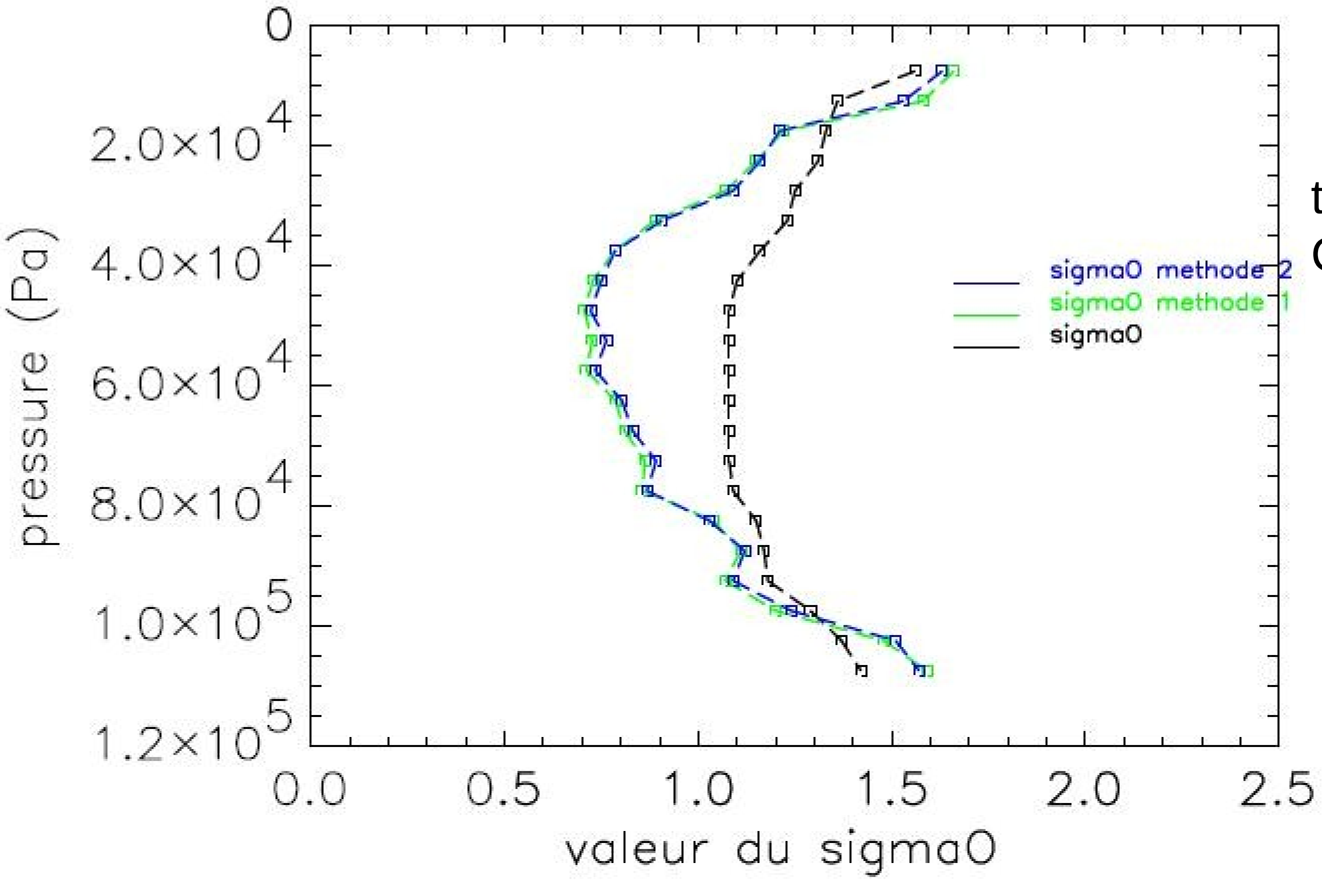
Unlike the first method no formal equivalence was found with another well documented method.
But both methods perform very much alike

For this method too
-need of a large number of observations
- All spatial correlation terms should be in B while there should be none in R.
-The result after a first iteration is a very good guess of the final result...

Moreover, performed over large enough a number of observations the results for both methods are strikingly consistent...

Desroziers et al.(2005)

Both methods provide similar (almost identical) results
but no formal proof of this (yet).



σ_o profiles for RS temperature obs. From Chabot (2008)

Tuning of inter channel covariances

Using the algorithm, inter channel covariances terms can be computed using :

$$\frac{1}{p}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_{\boldsymbol{a}})_i^T \boldsymbol{d}_{\boldsymbol{j}} = C_{ij}.$$

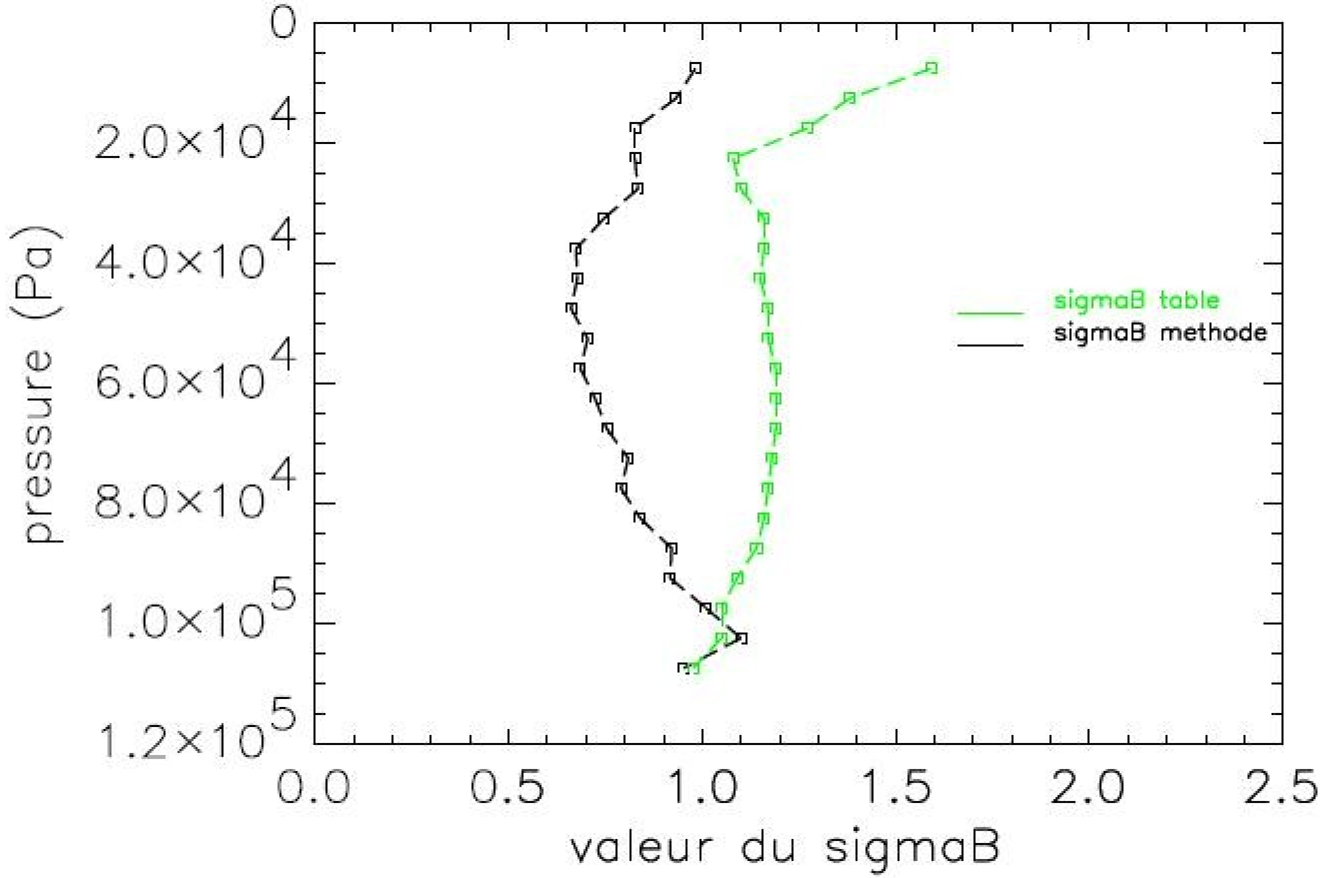
Or to make it more symmetric:

$$\frac{1}{2p}\left((\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_{\boldsymbol{a}})_i^T \boldsymbol{d}_{\boldsymbol{j}} + \boldsymbol{d}_{\boldsymbol{i}}^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_{\boldsymbol{a}})_j\right) = C_{ij}.$$

	true / diag.	true / diag.	true / diag.
obs. 1	1.00 / 1.00	1.00 / 1.01	0.10 / 0.10
obs. 2	1.00 / 1.01	4.00 / 3.73	-0.20 / -0.18
obs. 3	0.10 / 0.10	-0.20 / -0.18	0.25 / 0.25

True and diagnosed obs. error covariances in a simple « toy » experimental set, Desroziers et al. (2005)

Computing σ_b.



σ_b profile, sampled by RS temperature obs. From Chabot (2008).

Conclusions, future plans.

- Diagnostics of optimality have been a source of inspiration : 2 algorithms to tune error statistics, Fast computations of observation impact, almost costless in a perturbed observation ensemble context.
- Still to be seen, how do we deal with spatially correlated observation errors ?
- Operational use of the tuned statistics (need to remove bias first!).
- Assess and/or calibrate ensemble B.
- Evaluation of model error...