Diagnosing the optimality of data assimilation systems

Gérald Desroziers
Météo-France

Outline

Outline

1. General framework
2. Lagged innovation covariance
3. «Jmin» diagnostics
4. Observation space diagnostics
5. Ensemble variance diagnostics
6. Observation impact and optimality
7. Conclusion

General formalism

- *Statistical linear estimation* :

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x} = \mathbf{x}^b + \mathbf{K}\ \mathbf{d} = \mathbf{x}^b + \mathbf{BH}^T(\mathbf{HBH}^T\mathbf{+R})^{-1}\ \mathbf{d},$$

with $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$, *innovation*, **K**, *gain matrix*,

B et **R**, covariances of background and observation errors,

- Solution of the variational problem

$$J(\delta\mathbf{x}) = \delta\mathbf{x}^T\ \mathbf{B}^{-1}\ \delta\mathbf{x} + (\mathbf{d}\mathbf{-H}\ \delta\mathbf{x})^T\ \mathbf{R}^{-1}\ (\mathbf{d}\mathbf{-H}\ \delta\mathbf{x})$$

Non-linear formulation

- *Incremental formulation (Courtier et al, 1994)* : a strategy for minimizing the original non-linear cost-function:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathcal{H}(\mathbf{x}))$$

- Even in such a (slightly) non-linear problem, analysis, background and observation errors (or perturbations) are linked:

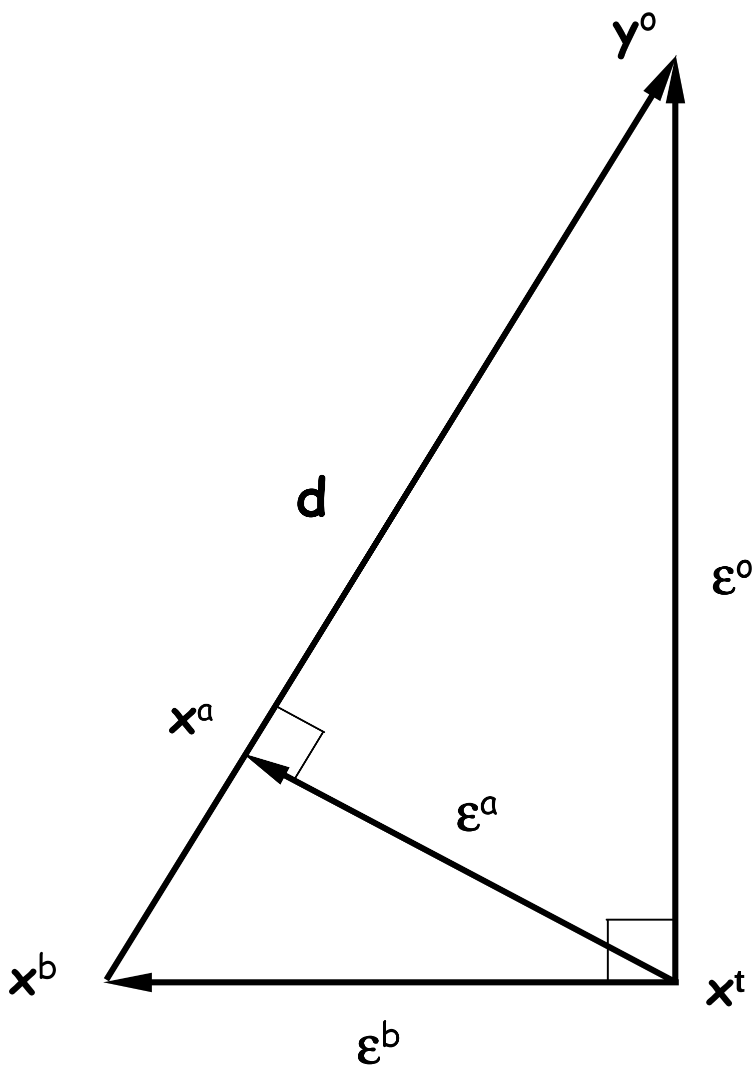
$$\varepsilon^a = (\mathbf{I} - \mathbf{KH})\, \varepsilon^b + \mathbf{K}\, \varepsilon^o \text{, with}$$

$$\varepsilon^a = \mathbf{x}^a - \mathbf{x}^t$$

$$\varepsilon^b = \mathbf{x}^b - \mathbf{x}^t$$

$$\varepsilon^o = \mathbf{y}^o - \mathcal{H}(\mathbf{x}^t)$$

Geometrical interpretation of analysis



$\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$

Scalar product:

$\langle \varepsilon, \varepsilon' \rangle = E[\varepsilon \ \varepsilon'^T]$

$\langle \varepsilon^a, \mathbf{d} \rangle = E[\varepsilon^a \mathbf{d}^T] = 0$:

- ✓ ε^a and **d** are orthogonal
- ✓ or, in other words, there is no projection of ε^a on **d**

Outline

1. General framework
2. Lagged innovation covariance
3. «Jmin» diagnostics
4. Observation space diagnostics
5. Ensemble variance diagnostics
6. Observation impact and optimality
7. Conclusion

Lagged innovation covariance

- *Kalman Filter sequence:*

$\mathbf{d}_n = \mathbf{y}^o_n - \mathcal{H}_n(\mathbf{x}^f_n)$
$\mathbf{x}^a_n = \mathbf{x}^f_n + \mathbf{K}_n \mathbf{d}_n$
$\mathbf{x}^f_{n+1} = \mathcal{M}_n(\mathbf{x}^a_n)$

$\mathbf{d}_{n+1} = \mathbf{y}^o_{n+1} - \mathcal{H}_{n+1}(\mathbf{x}^f_{n+1})$
...

- The lagged innovations $\mathbf{d}_n$ and $\mathbf{d}_{n+1}$ should be decorrelated.

(Dee, 1983; Daley, 1992)

- Consequence of estimation error and innovation decorrelation.

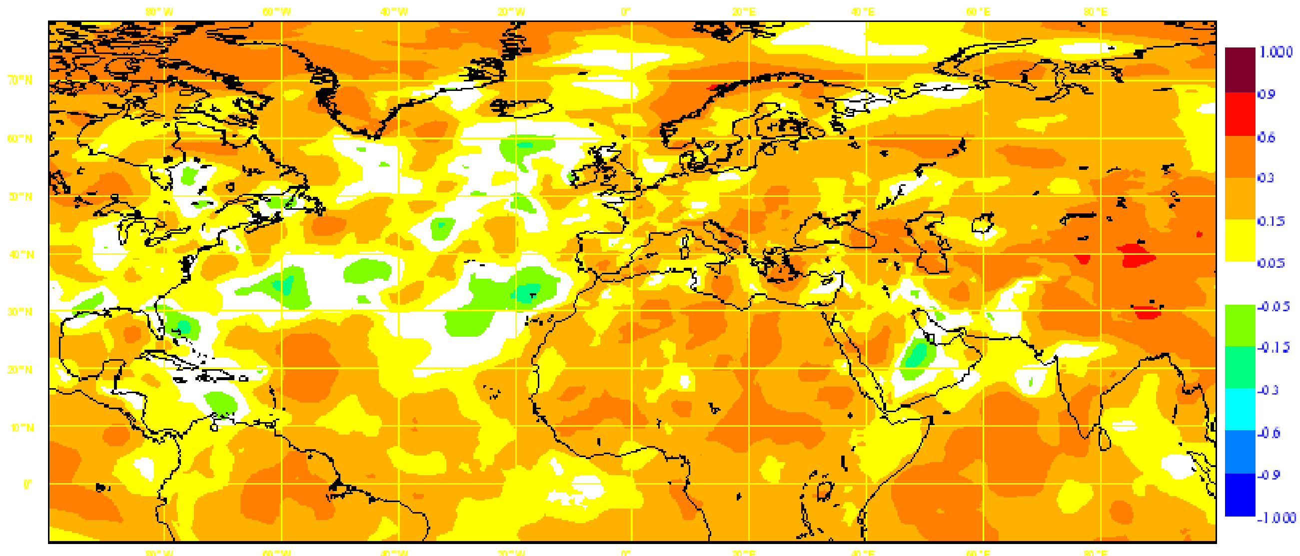
- Translates into $\delta\mathbf{x}_n (\delta\mathbf{x}_{n+1})^T = 0$.

(Chapnik, 2006)

Lagged increment covariance

Fig 4

Correlation of mean sea level pressure increments series.



(from Chapnik, 2006)

Outline

1. General framework
2. Lagged innovation covariance
3. «Jmin» diagnostics
4. Observation space diagnostics
5. Ensemble variance diagnostics
6. Observation impact and optimality
7. Conclusion

A posteriori « Jmin » diagnostics

- We should have

$$E[J(\mathbf{x}^a)] = p,$$

with p = total number of observations.

(Bennett et al, 1993)

- More precisely

$$E[J_i(\mathbf{x}^a)] = p_i - \mathrm{Tr}(\mathbf{S}_i^{-1/2}\,\Gamma_i\,\mathbf{A}\,\Gamma_i^T\,\mathbf{S}_i^{-1/2}),$$

p_i : number of pieces of information ($\mathbf{x}^b$ or $\mathbf{y}^o$) associated with J_i,

$\mathbf{S}_i$, Γ_i : associated error cov. matrix and « observation » operator.

(Talagrand, 1999)

Particular cases

- Complete background term:

$$\Gamma_i = \mathbf{I}_n,$$
$$\mathbf{S}_i = \mathbf{B},$$
$$E[J^b(\mathbf{x}^a)] = n - Tr(\mathbf{B}^{-1/2}\,\mathbf{I}_n\,\mathbf{A}\,\mathbf{I}_n^T\,\mathbf{B}^{-1/2})$$
$$= Tr(\mathbf{K}\;\mathbf{H})$$

- Complete observation term:

$$\Gamma_i = \mathbf{H},$$
$$\mathbf{S}_i = \mathbf{R},$$
$$E[J^o(\mathbf{x}^a)] = p - Tr(\mathbf{R}^{-1/2}\,\mathbf{H}\,\mathbf{A}\,\mathbf{H}^T\,\mathbf{R}^{-1/2})$$
$$= p - Tr(\mathbf{H}\;\mathbf{K})$$

- Subpart of obs. term:

$$\Gamma_i = \mathbf{H}_i,$$
$$\mathbf{S}_i = \mathbf{R}_i,$$
$$E[J^o_i(\mathbf{x}^a)] = p_i - Tr(\mathbf{R}_i^{-1/2}\,\mathbf{H}_i\,\mathbf{A}\,\mathbf{H}_i^T\,\mathbf{R}_i^{-1/2})$$
$$= p_i - Tr(\mathbf{H}_i\;\mathbf{K}_i),$$

with $\mathbf{H}_i$, $\mathbf{K}_i$ the restrictions of $\mathbf{H}$, $\mathbf{K}$ to subset i.

Computation of $\mathrm{Tr}(\mathbf{H}_i \mathbf{K}_i)$ in a variational scheme

- **K** unknown, but relation between errors (or perturbations) still holds:

$$\varepsilon^a = (\mathbf{I} - \mathbf{KH})\, \varepsilon^b + \mathbf{K}\, \varepsilon^o$$

- For observation subset i:

$$\mathbf{H}_i\, \varepsilon^a = \mathbf{H}_i (\mathbf{I} - \mathbf{KH})\, \varepsilon^b + \mathbf{H}_i \mathbf{K}\, \varepsilon^o$$
$$= \mathbf{H}_i (\mathbf{I} - \mathbf{KH})\, \varepsilon^b + \mathbf{H}_i \Sigma_j \mathbf{K}_j\, \varepsilon^o_j$$

- Linear regression:

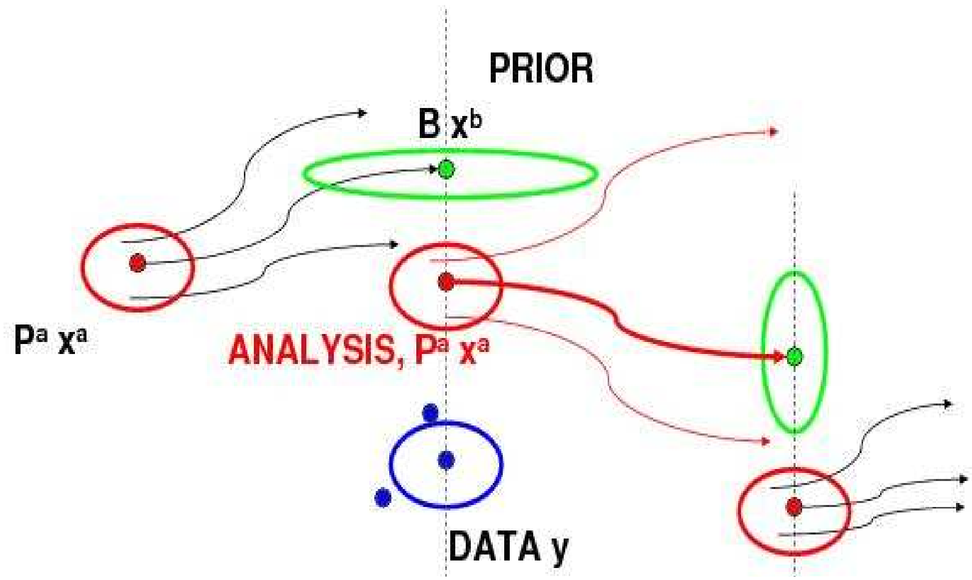
$$\mathbf{H}_i \mathbf{K}_i = \mathrm{cov}(\mathbf{H}_i\, \varepsilon^a, \varepsilon^o_i) / \mathrm{cov}(\varepsilon^o_i, \varepsilon^o_i)$$
$$= \mathrm{cov}(\mathbf{H}_i\, \varepsilon^a, \varepsilon^o_i) / \mathbf{R}_i$$

- Or:

$$\mathrm{Tr}(\mathbf{H}_i \mathbf{K}_i) = {\varepsilon^o_i}^T \mathbf{R}_i^{-1} \mathbf{H}_i\, \varepsilon^a$$

(Desroziers and Ivanov, 2001)

Computation from an ensemble of perturbed assimilations



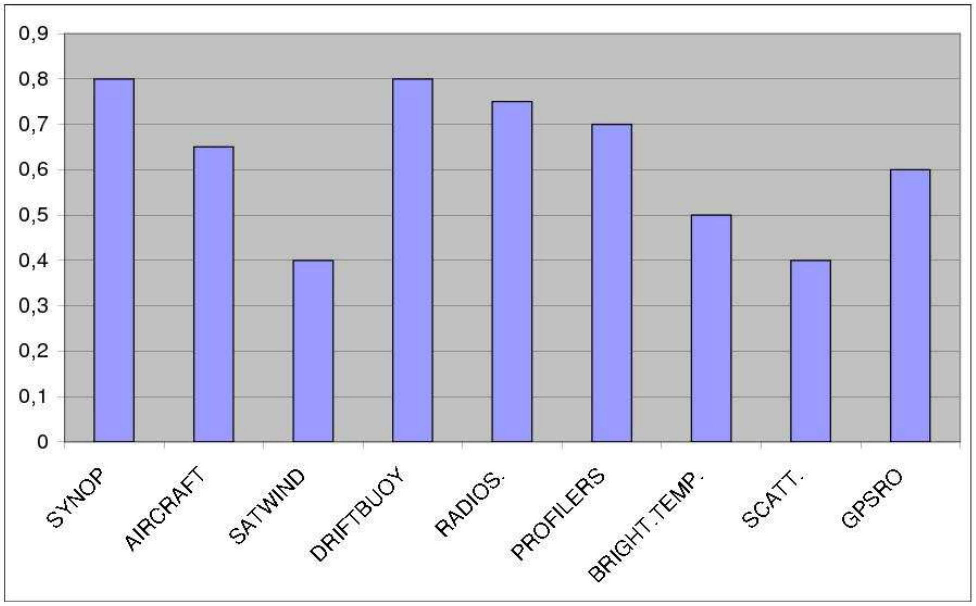
(from Ehrendorfer, 2006)

- Ensemble assimilation : simulation of the joint evolution of analysis, background and observation errors.

- $E[J^o_i(\mathbf{x}^a)] = Tr(\mathbf{H}_i\,\mathbf{K}_i)$ are sub-products of an ensemble of perturbed analyses.

(Desroziers et al, 2009)

Application : optimization of R



Normalization of $\mathbf{R}_i$:

$s^o_i \, \mathbf{R}_i$

Coef. s^o_i diagnosed with

$s^o_i = E[J^o_i(\mathbf{x}^a)]/(E[J^o_i(\mathbf{x}^a)])^{opt}$.

Normalization coefficients of σ^o_i in the French Arpège 4D-Var

(Chapnik, et al, 2004; Buehner, 2005; Desroziers et al, 2009)

Application : normalization of $\mathbf{B}$

- Normalization of $\mathbf{B}$: $s^b\,\mathbf{B}$.

- Coefficient s^b diagnosed with $s^b = E[J^b(\mathbf{x}^a)]/(E[J^b(\mathbf{x}^a)])^{opt}$.

- $(E[J^b(\mathbf{x}^a)])^{opt}$ given by $(E[J^b(\mathbf{x}^a)])^{opt} = Tr(\mathbf{HK}) = \Sigma_i\, Tr(\mathbf{H}_i\,\mathbf{K}_i)$.

- Allows the global inflation of background error variances given by an ensemble of perturbed assimilations.

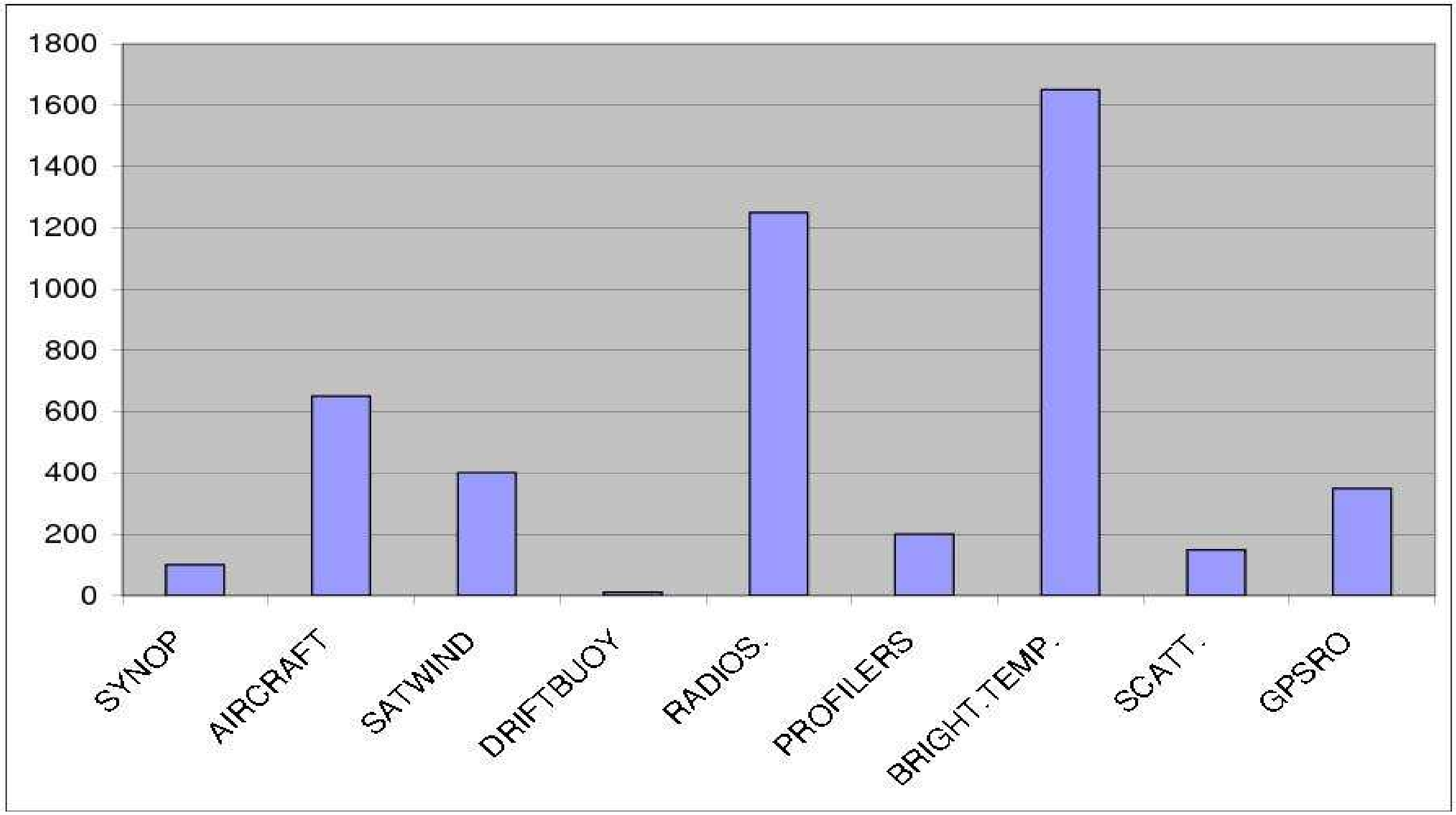
Link with different measures of the impact of independent observations

- $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \Sigma_i \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i$ — $\mathbf{A}^{-1}$ = background « precision » + obs. « precisions »
- $\mathbf{I}_n = \mathbf{A}\ \mathbf{B}^{-1} + \Sigma_i \mathbf{A}\ \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i = (\mathbf{I}_n - \mathbf{KH}) + \Sigma_i \mathbf{K}_i \mathbf{H}_i$ — $\mathbf{I}_n$ = background ponderation + obs. ponderations
- $n = \mathrm{Tr}(\mathbf{I}_n - \mathbf{KH}) + \Sigma_i \mathrm{Tr}(\mathbf{K}_i \mathbf{H}_i)$ — n = DFS background + DFS observations
- $\mathbf{B} = \mathbf{A} + \Sigma_i \mathbf{A}\ \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{B} = \mathbf{A} + \Sigma_i \mathbf{K}_i \mathbf{H}_i \mathbf{B}$ — bg error cov. = res. error cov. + explained error cov.

DFS: Degrees of Freedom for Signal : Information content.

(Cardinali, 2004)

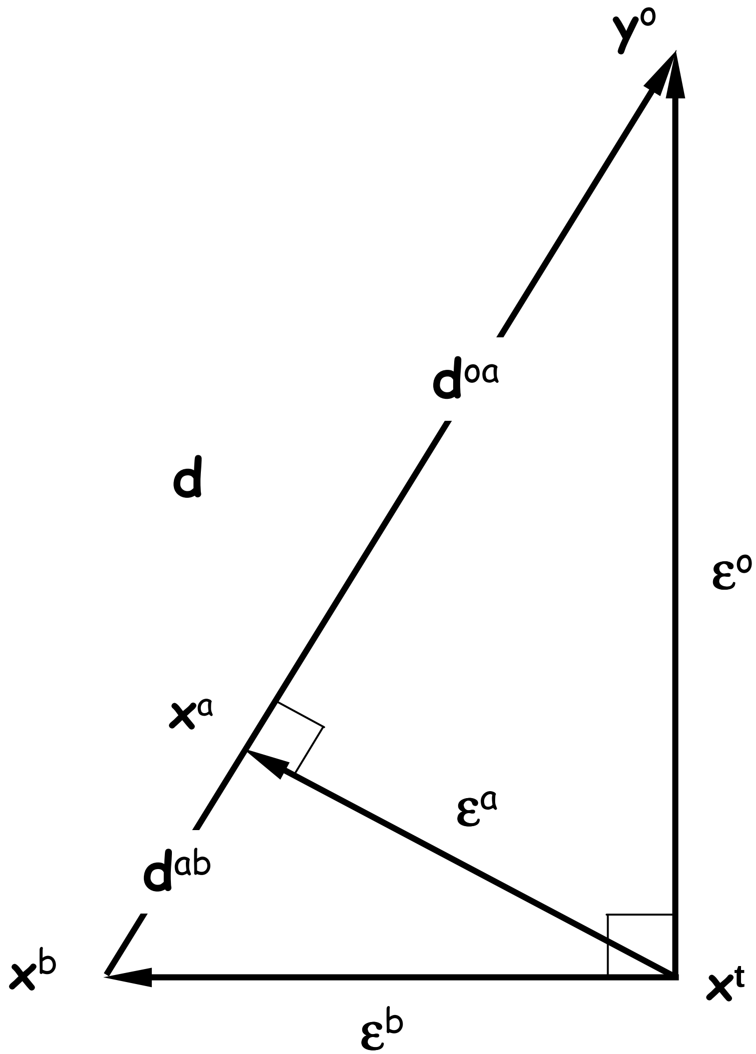
Degrees of Freedom for Signal



Information content of observations in the French Arpège 4D-Var

Outline

Diagnostics in observation space



(Desroziers et al, 2005)

$\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$

$\mathbf{d}^{oa} = \mathbf{y}^o - H(\mathbf{x}^a)$

$\mathbf{d}^{ab} = H(\mathbf{x}^a) - H(\mathbf{x}^b)$

$E[\mathbf{d}^{oa}\ \mathbf{d}^T] = \mathbf{R}$

$E[\mathbf{d}^{ab}\ \mathbf{d}^T] = \mathbf{HBH}^T$

$\langle \varepsilon, \varepsilon' \rangle = E[\varepsilon\ \varepsilon'^T]$

Practical implementation

- For any subset i with p_i observations, simply compute

$$(\sigma^o_i)^2 = \Sigma_{j=1,pi} (y^o_j - \mathcal{H}_j(\mathbf{x}^a)) (y^o_j - \mathcal{H}_j(\mathbf{x}^b)) / p_i$$

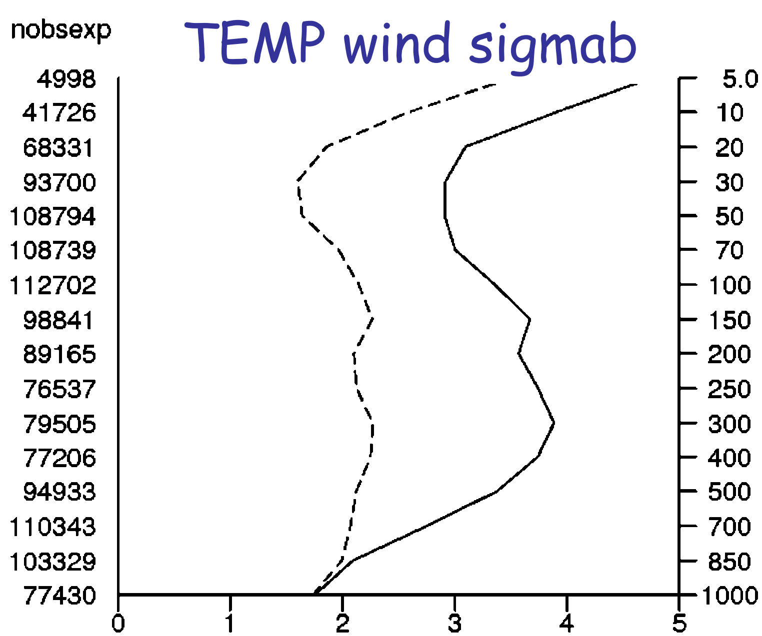
and

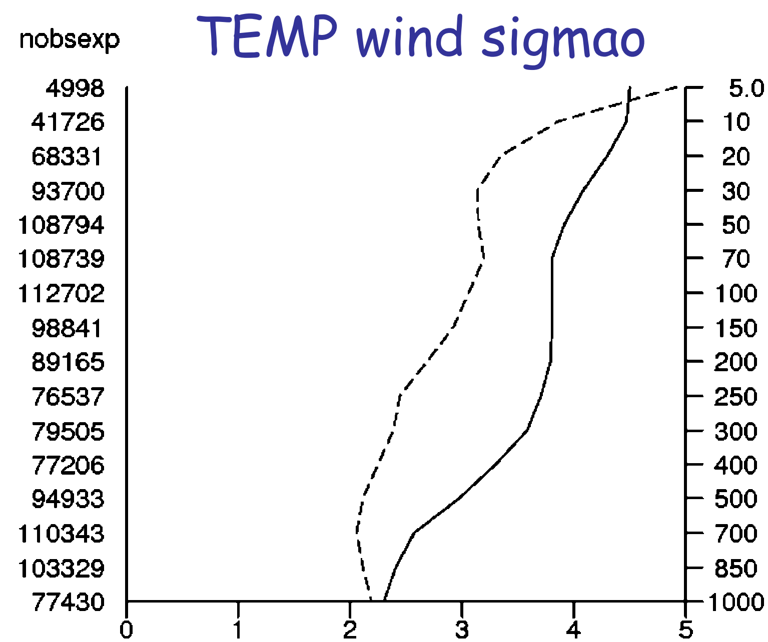
$$(\sigma^b_i)^2 = \Sigma_{j=1,pi} (\mathcal{H}_j(\mathbf{x}^b) - \mathcal{H}_j(\mathbf{x}^a)) (y^o_j - \mathcal{H}_j(\mathbf{x}^b)) / p_i$$

- This is nearly cost-free and can be computed

✓ a posteriori,
✓ over one or several analyses,
✓ in any data assimilation scheme (including 4D-Var).

Practical implementation



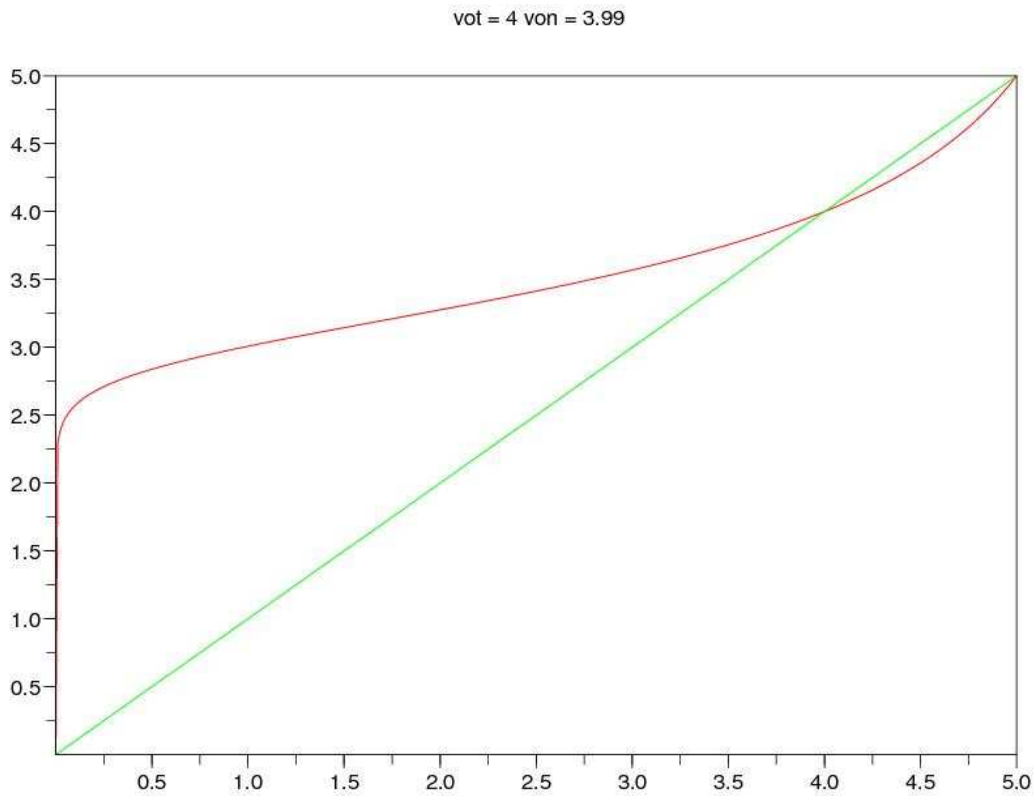


___ specified in Arpège 4D-Var

--- diagnosed in observation space

(20081127 00H – 20081228 18H)

Convergence : $v^o_{diag}(v^o)$

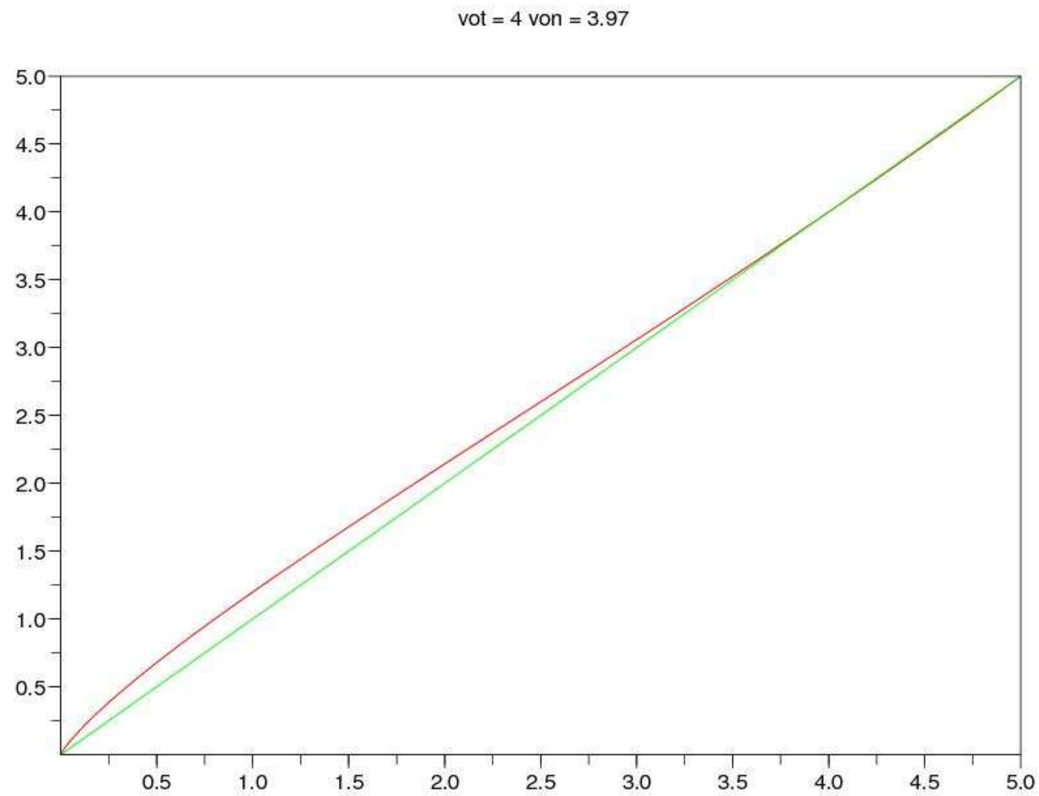


Idealized case: analysis on an equatorial circle (40 000km).

$v^o_{true} = 4.$

L^b = 300 km / L^0= 0 km.

Convergence : $v^o_{diag}(v^o)$



Idealized case: analysis on an equatorial circle (40 000km).

$v^o_{true} = 4$.

L^b = 300 km / L^0= 200 km.

Outline

1. General framework
2. Lagged innovation covariance
3. «Jmin» diagnostics
4. Observation space diagnostics
5. Ensemble variance diagnostics
6. Observation impact and optimality
7. Conclusion

Ensemble/diagnosed variances in observation space

- Ensemble variances can be computed at observation locations i with

$$(\sigma^{be}_i)^2 = \Sigma_{j=1,ne} (\mathbf{h}_j \, \varepsilon^b)^2 / ne,$$

where ne is the ensemble size.

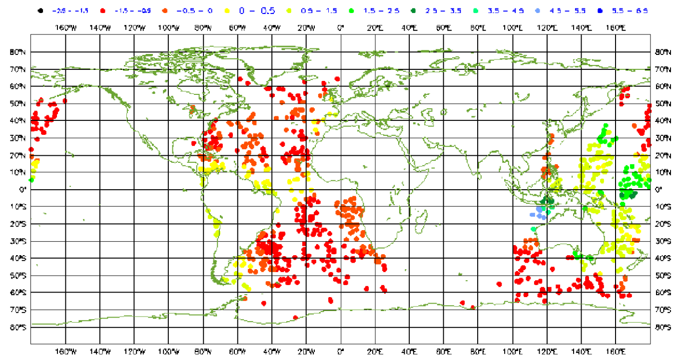
- Can be compared to diagnosed errors

$$(\sigma^{bd}_i)^2 = \Sigma_{j=1,pa} (h_j (\mathbf{x}^b) - h_j (\mathbf{x}^a)) (y^o_j - h_j (\mathbf{x}^b)) / pa,$$

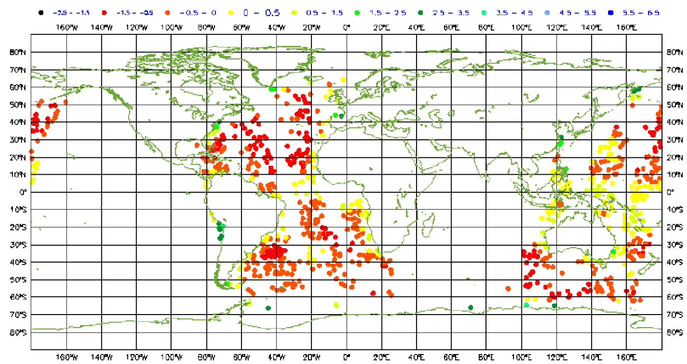
where pa is the number of obs. taken around each obs. location i.

- pa is optimized to maximize the correlation between $(\sigma^{be}_i)^2$ and $(\sigma^{bd}_i)^2$

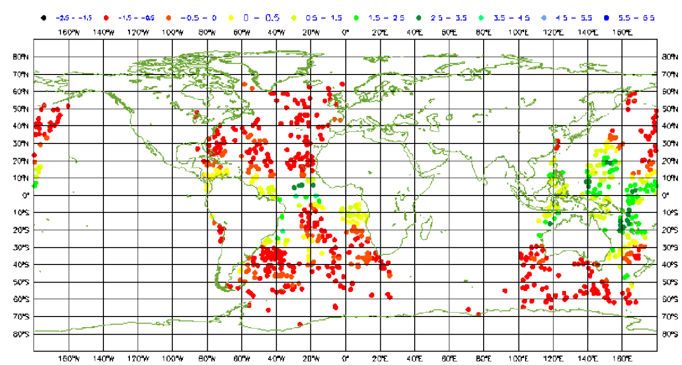
Ensemble / diagnosed background errors in HIRS-7 space



(a) HIRS7-diagnostic



(b) HIRS7-63HE-HR



(c) HIRS7-63QH

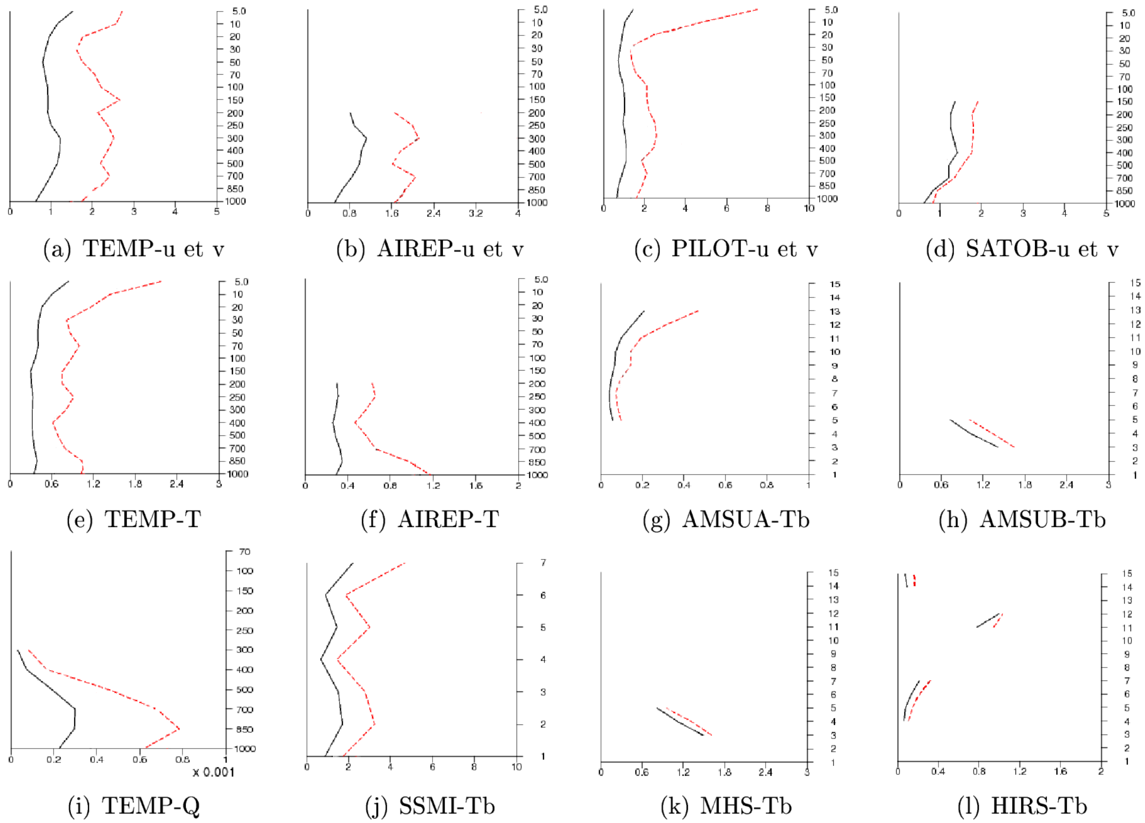
(from Gibier, 2009)

✓Diagnosed 4D-Var background errors

✓3D-Var FGAT ensemble

✓4D-Var ensemble

Ensemble / diagnosed background errors



__ ensemble

--- diagnosed

(from Gibier, 2009)

Outline

1. General framework
2. Lagged innovation covariance
3. «Jmin» diagnostics
4. Observation space diagnostics
5. Ensemble variance diagnostics
6. Observation impact and optimality
7. Conclusion

Impact of observations on forecasts

- Measure of the quality of a forecast $\mathbf{x}^f = M(\mathbf{x})$:

$$J(\mathbf{x}) = (M(\mathbf{x})-\mathbf{x}^v)^T \boldsymbol{C} (M(\mathbf{x})-\mathbf{x}^v),$$

with, for example, $\boldsymbol{C}$ = energy norm.
and $\mathbf{x}^v$ the verifying analysis at final time t^f.

(Langland et Baker, 2004; Gelaro and Zhu, 2009)

- Expression in terms of initial error:

$$J(\varepsilon) = (\mathbf{M}\ \varepsilon)^T \boldsymbol{C}\ (\mathbf{M}\ \varepsilon),$$

with $\varepsilon = \mathbf{x}-\mathbf{x}^t$ the error at initial time t^i.

Impact of observations on forecasts / optimality

- Taylor expansion at ε^a:

$$J(\varepsilon^b) = J(\varepsilon^a) + (\varepsilon^b - \varepsilon^a)^T J'(\varepsilon^a) + \tfrac{1}{2} (\varepsilon^b - \varepsilon^a)^T J''(\varepsilon^a) (\varepsilon^b - \varepsilon^a)$$
$$= J(\varepsilon^a) + 2 (\varepsilon^b - \varepsilon^a)^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, \varepsilon^a + (\varepsilon^b - \varepsilon^a)^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, (\varepsilon^b - \varepsilon^a)$$
$$= J(\varepsilon^a) + 2\, \mathbf{d}^T \mathbf{K}^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, \varepsilon^a + \mathbf{d}^T \mathbf{K}^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, (\varepsilon^b - \varepsilon^a)$$

First order term = 0 in an optimal system ($<\mathbf{d}, \varepsilon^a> = 0$)! (Cardinali, 2008)

- Taylor expansion at ε^b:

$$J(\varepsilon^a) = J(\varepsilon^b) + 2 (\varepsilon^a - \varepsilon^b)^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, \varepsilon^b + (\varepsilon^a - \varepsilon^b)^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, (\varepsilon^a - \varepsilon^b)$$
$$= J(\varepsilon^b) + 2\, \mathbf{d}^T \mathbf{K}^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, \varepsilon^b + \mathbf{d}^T \mathbf{K}^T \mathbf{M}^T \mathbf{C}\, \mathbf{M}\, (\varepsilon^a - \varepsilon^b)$$

First order term = $- 2\, \mathrm{Tr}(\mathbf{M}^T \mathbf{C}\, \mathbf{M}\, \mathbf{K}\, \mathbf{H}\, \mathbf{B})$ = twice the optimal value of error reduction by observations!

- 2nd order expansion required. (Errico, 2007)

Conclusion

- Wide range of diagnostics, linked with Extended KF formalism.

- Useful to keep in mind.

- Applicable to a slightly non-linear scheme such as incremental 4D-Var.

- *A posteriori* diagnostics are quite useful
  - ✓ to diagnose and tune background and observation error variances,
  - ✓ to measure information content of observations.

- Might be also useful to diagnose model error.