A posteriori diagnostics

Gérald Desroziers
ISDA, Reading, 20 July 2016

Outline

Outline

1. Introduction
2. « J_{min} » diagnostics
3. Observation space diagnostics
4. Combination of diagnostics
5. Conclusion

Introduction

« A posteriori diagnostics »: diagnostics made after the analysis.

A noncomprehensive list of references on the subject!

- *Bennett 1992*: J_{min}, …
- *Talagrand 2000, 2010, 2015*: « a posteriori diagnostics », J_{min}, DFS, …
- *Desroziers and Ivanov 2001, Chapnik et al 2006, Desroziers et al 2009,*
- *Michel 2014*: « a posteriori diagnostics », J_{min}, DFS, σ^o, σ^b tuning …
- *Fisher 2003*: DFS, …
- *Desroziers et al 2005*: diagnostics in observation space …
- *Todling 2015, Howes 2015*: diagnostic of model error, ...
- *Waller et al 2015*: study of observation space diagnostics …
- *Ménard 2000, 2015a, 2015b*: study of « a posteriori diagnostics » …

- …

General framework

- Statistical linear estimation

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

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

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

- Also solution of the variational problem

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

Outline

1. Introduction
2. « J_{min} » diagnostics
3. Observation space diagnostics
4. Combination of diagnostics
5. Conclusion

J_{min} / χ^2

- Value of the cost function at its minimum

$$J_{min} = J^b(\delta\mathbf{x}^a) + J^o(\delta\mathbf{x}^a)$$
$$= (\mathbf{Kd})^T\,\mathbf{B}^{-1}\,(\mathbf{Kd}) + ((\mathbf{I}-\mathbf{HK})\mathbf{d})^T\,\mathbf{R}^{-1}\,((\mathbf{I}-\mathbf{HK})\mathbf{d})$$
$$= \mathbf{d}^T\,(\mathbf{HBH}^T+\mathbf{R})^{-T}\,\mathbf{HBH}^T\,(\mathbf{HBH}^T+\mathbf{R})^{-1}\,\mathbf{d} + \mathbf{d}^T\,(\mathbf{HBH}^T+\mathbf{R})^{-T}\,\mathbf{R}\,(\mathbf{HBH}^T+\mathbf{R})^{-1}\,\mathbf{d}$$
$$= \mathbf{d}^T\,(\mathbf{HBH}^T+\mathbf{R})^{-1}\,\mathbf{d}$$
$$= \mathbf{d}^T\,\mathbf{D}^{-1}\,\mathbf{d}.$$

- Statistical expectation of J_{min}

$$E(J_{min}) = E(\mathbf{d}^T\,\mathbf{D}^{-1}\,\mathbf{d})$$
$$= \mathrm{Tr}(\mathbf{D}^{-1}\,E(\mathbf{d}\,\mathbf{d}^T))$$
$$= p\text{, number of observations, if } \mathbf{D} = E(\mathbf{d}\,\mathbf{d}^T).$$

- Unnecessary: $\mathbf{HBH}^T = \alpha\,\mathbf{HB}^t\,\mathbf{H}^T$ and $\mathbf{R} = \alpha\,\mathbf{R}^t$, $\delta\mathbf{x}^a$ optimal but $E(J_{min}) = p / \alpha$!

- Statistical distribution of J_{min}

✓ Gaussian **d**: J_{min} follows a χ^2 distribution of order p, with mean p and var. 2p.

✓ Central limit theorem: J_{min} follows a $\mathcal{N}(p,2p)$ law, if p is large.

Expectation of subparts of J_{min}

- J^b part

$E(J^b_{min})$ = $Tr(\mathbf{HK})$
= total DFS (Degrees of freedom for Signal)
= number of degrees of freedom,
provided by observations, to deviate from background.

- J^o part

$E(J^o_{min})$ = $p - Tr(\mathbf{HK})$
= total DFN (Degrees of freedom for Noise)
= number of degrees of freedom ,
provided by background, to deviate from observations.

- J^o subparts

$E(J^o_{i,\,min}) = p_i - Tr(\Pi_i\, \mathbf{HK}\, \Pi_i^T)$,
Π_i selection of subset of observations i, with number p_i.

- $Tr(\Pi_i\, \mathbf{HK}\, \Pi_i^T)$ = DFS_i
= partial number of degrees of freedom,
provided by observations i, to deviate from background.

Computation of subparts of J_{min} in a variational assimilation

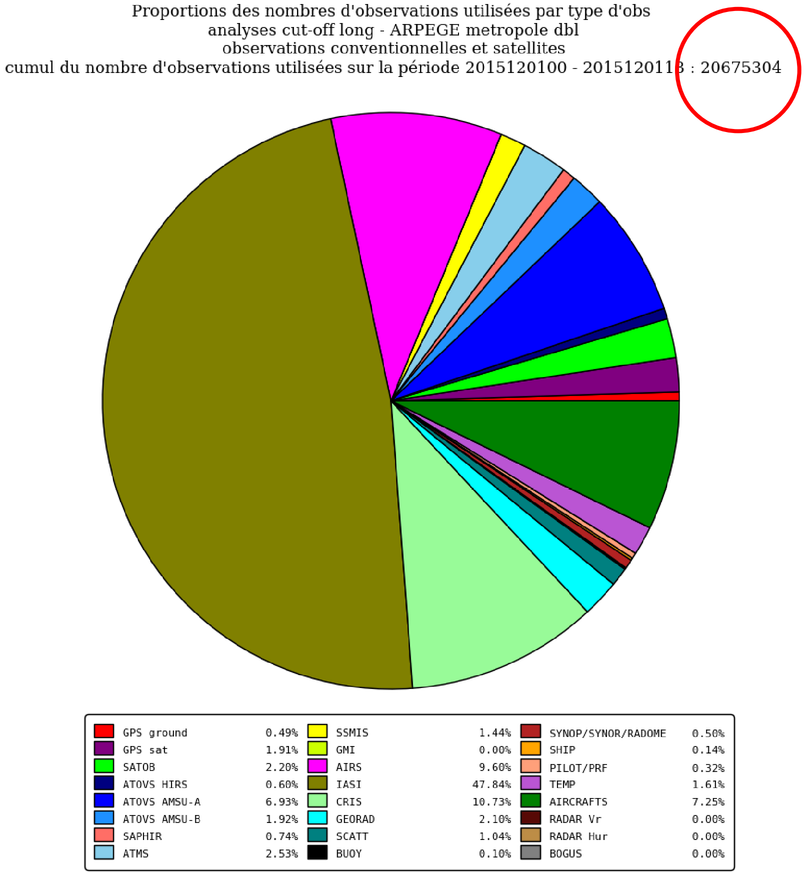
- Matrix **K** is not explicitely computed in a variational assimilation.
- Computation of $Tr(\Pi_i \mathbf{HK} \Pi_i^T)$ with a Monte Carlo procedure

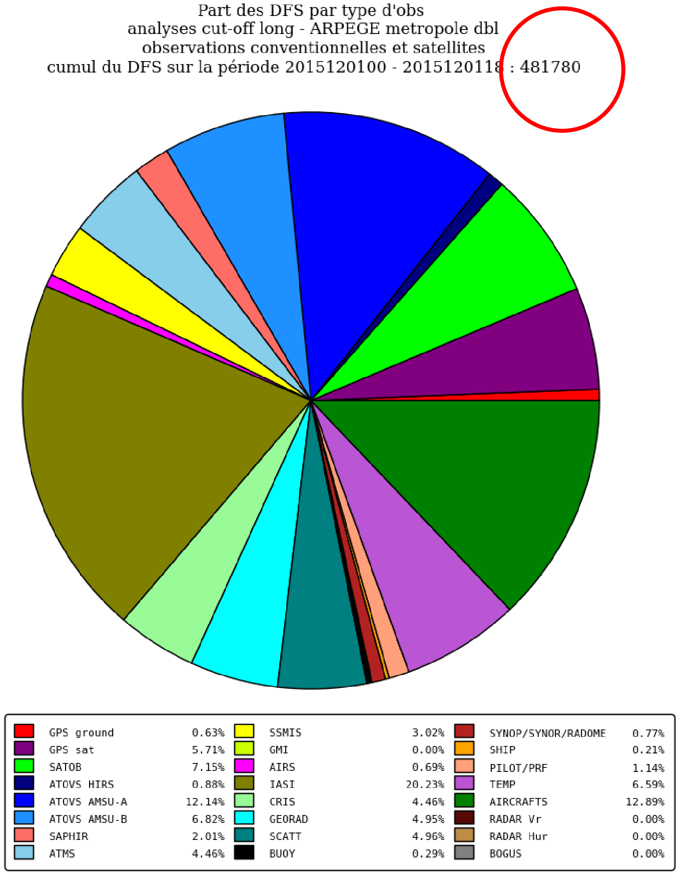
$$
\begin{aligned}
Tr(\Pi_i \mathbf{HK} \Pi_i^T) &= Tr(\Pi_i^T \Pi_i \mathbf{HK}) \\
&= E(\delta \mathbf{y}^{o\,T}_i \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{K} \delta \mathbf{y}^o) \\
&= E(\delta \mathbf{y}^{o\,T}_i \mathbf{R}_i^{-1} \mathbf{H}_i \delta \mathbf{x}^a(\delta \mathbf{y}^o)),
\end{aligned}
$$

with $\delta \mathbf{y}^o = \mathcal{N}(\mathbf{0}, \mathbf{R})$.

(Desroziers and Ivanov 2001, Desroziers et al 2009, Michel 2014)

DFS_i computation





Tuning of error variances

- Normalization of $\mathbf{R}_i$: $s^o{}_i{}^2\ \mathbf{R}_i$

  Coef. $s^o{}_i{}^2$ diagnosed with $s^o{}_i{}^2 = E[J^o{}_i(\mathbf{x}^a)] \,/\, (\, E[J^o{}_i(\mathbf{x}^a)]\,)^{opt}$
  $= E[J^o{}_i(\mathbf{x}^a)] \,/\, (\, p_i - Tr(\mathbf{P}_i\ \mathbf{HK}\ \mathbf{P}_i{}^T)\,)$.

- Normalization of $\mathbf{B}$: $s^{b2}\ \mathbf{B}$

  Coef. s^{b2} diagnosed with $s^{b2} = E[J^b(\mathbf{x}^a)] \,/\, (\, E[J^b(\mathbf{x}^a)]\,)^{opt}$
  $= E[J^b(\mathbf{x}^a)] \,/\, Tr(\mathbf{HK})$.

  (Desroziers and Ivanov 2001, Chapnik et al, 2004, Desroziers et al 2009)

- Equivalent to Maximum-likelihood estimation (Chapnik et al 2006, Menard 2015)

  $f(\mathbf{d}|\mathbf{s}) = 1 \,/\, ((2p)^p \det(\mathbf{D}(\mathbf{s}))^{1/2} \exp(-1/2\ \mathbf{d}^T\ \mathbf{D}(\mathbf{s})^{-1}\ \mathbf{d})$,
  where $\mathbf{D}(\mathbf{s})$ is the covariance matrix of innovations with parameters **s.**

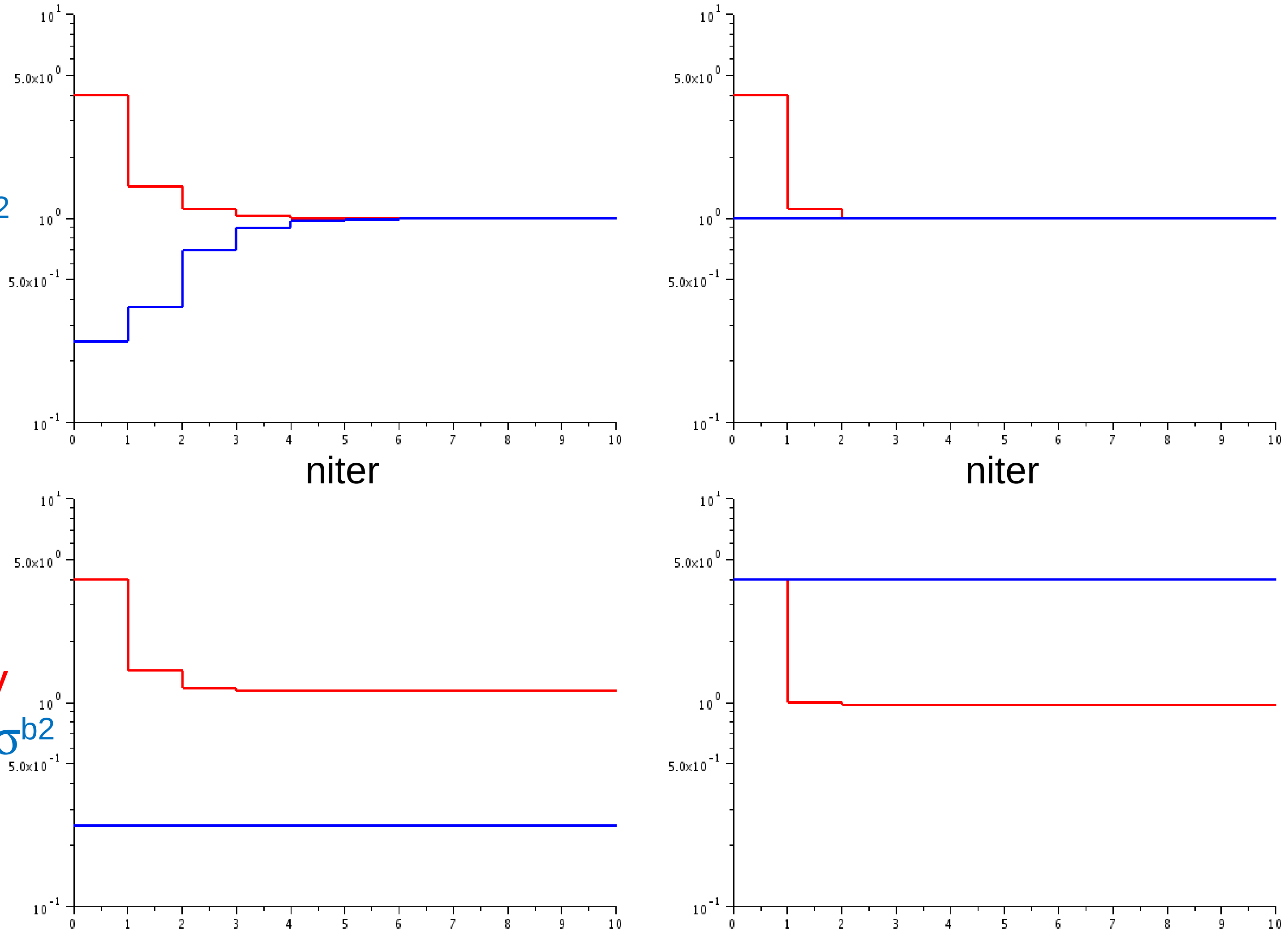
  Optimal parameters **s** are those that minimize the Log-likelihood
  $L(s) = -\log(\, f(\mathbf{d}|\mathbf{s})\,)$.

J_{min} diagnostics: mean σ^o / σ^b

$\sigma^{ot} = 1$ ($L^{ot} = 0$ km) / $\sigma^{bt} = 1$ ($L^{bt} = 500$ km)

σ^{o2} / σ^{b2}

σ^{o2} only



σ^{o2} only
wrong σ^{b2}

σ^{o2} only
wrong σ^{b2}

Outline

1. Introduction
2. « J_{min} » diagnostics
3. Observation space diagnostics
4. Combination of diagnostics
5. Conclusion

Diagnostics in observation space

- Observations / background departures

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

- Observations / analysis departures

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

$\sim (\mathbf{I}\text{-}\mathbf{HK})\,\mathbf{d}$

$\sim \mathbf{R}\,(\mathbf{HBH}^T + \mathbf{R})^{-1}\,\mathbf{d}.$

- Background / analysis departures in observation space

$\mathbf{d}^{ba} = H(\mathbf{x}^b) - H(\mathbf{x}^a) = H(\mathbf{x}^b) - H(\mathbf{x}^b + \mathbf{K}\,\mathbf{d})$

$\sim \mathbf{HK}\,\mathbf{d}$

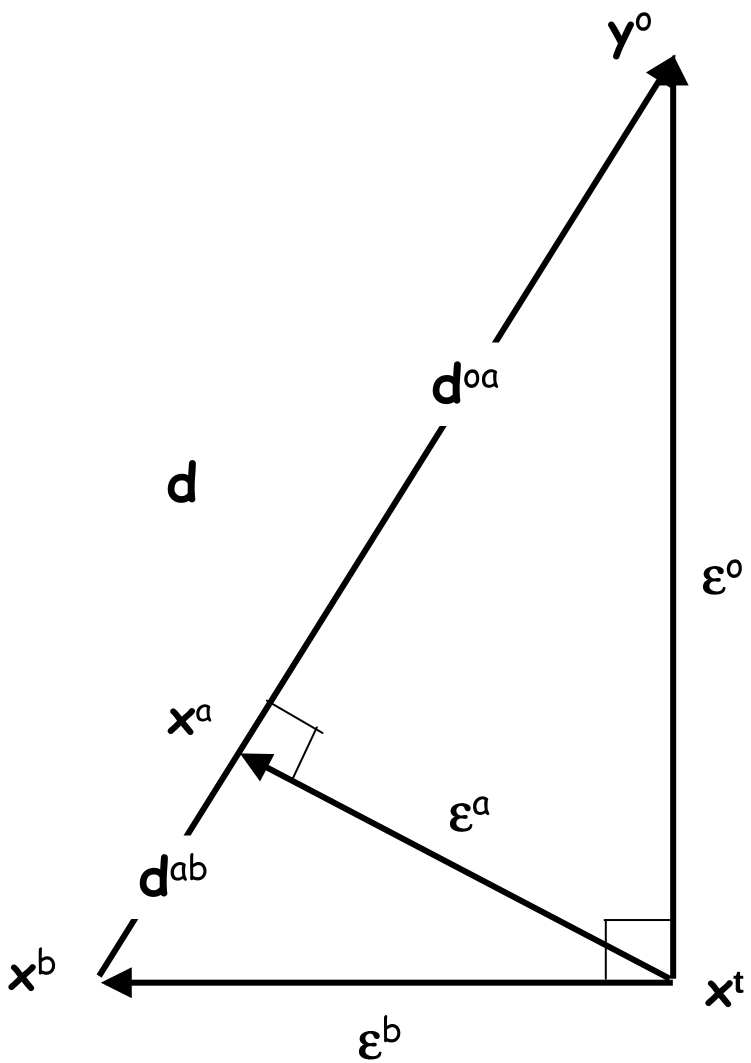
$\sim \mathbf{HBH}^T\,(\mathbf{HBH}^T + \mathbf{R})^{-1}\,\mathbf{d}.$

- A posteriori diagnostics

$E[\mathbf{d}^{oa}\,\mathbf{d}^T] = \mathbf{R}\,(\mathbf{HBH}^T + \mathbf{R})^{-1}\,E[\mathbf{d}\,\mathbf{d}^T] = \mathbf{R}\,(\boldsymbol{HBH}^T + \boldsymbol{R})^{-1}\,(\boldsymbol{HB}^t\boldsymbol{H}^T + \boldsymbol{R}^t) = \mathbf{R}^t$

$E[\mathbf{d}^{ba}\,\mathbf{d}^T] = \mathbf{HBH}^T\,(\mathbf{HBH}^T + \mathbf{R})^{-1}\,E[\mathbf{d}\,\mathbf{d}^T] = \mathbf{HBH}^T\,(\boldsymbol{HBH}^T + \boldsymbol{R})^{-1}\,(\boldsymbol{HB}^t\boldsymbol{H}^T + \boldsymbol{R}^t) = \mathbf{HB}^t\mathbf{H}^T.$

Diagnostics in observation space



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

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

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

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

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

$$E[\mathbf{d}^{ba}\,\mathbf{d}^{oaT}] = \mathbf{HAH}^T$$

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

(Desroziers et al, 2005)

Diagnostics in observation space: practical implementation

- Difficulty

No multiple statistical realizations of **d**, $\mathbf{d}^{oa}$, $\mathbf{d}^{ba}$ for a given analysis!

- An ergodicity hypothesis is necessary

Average over time, space, subset of observations…

- Practical computation

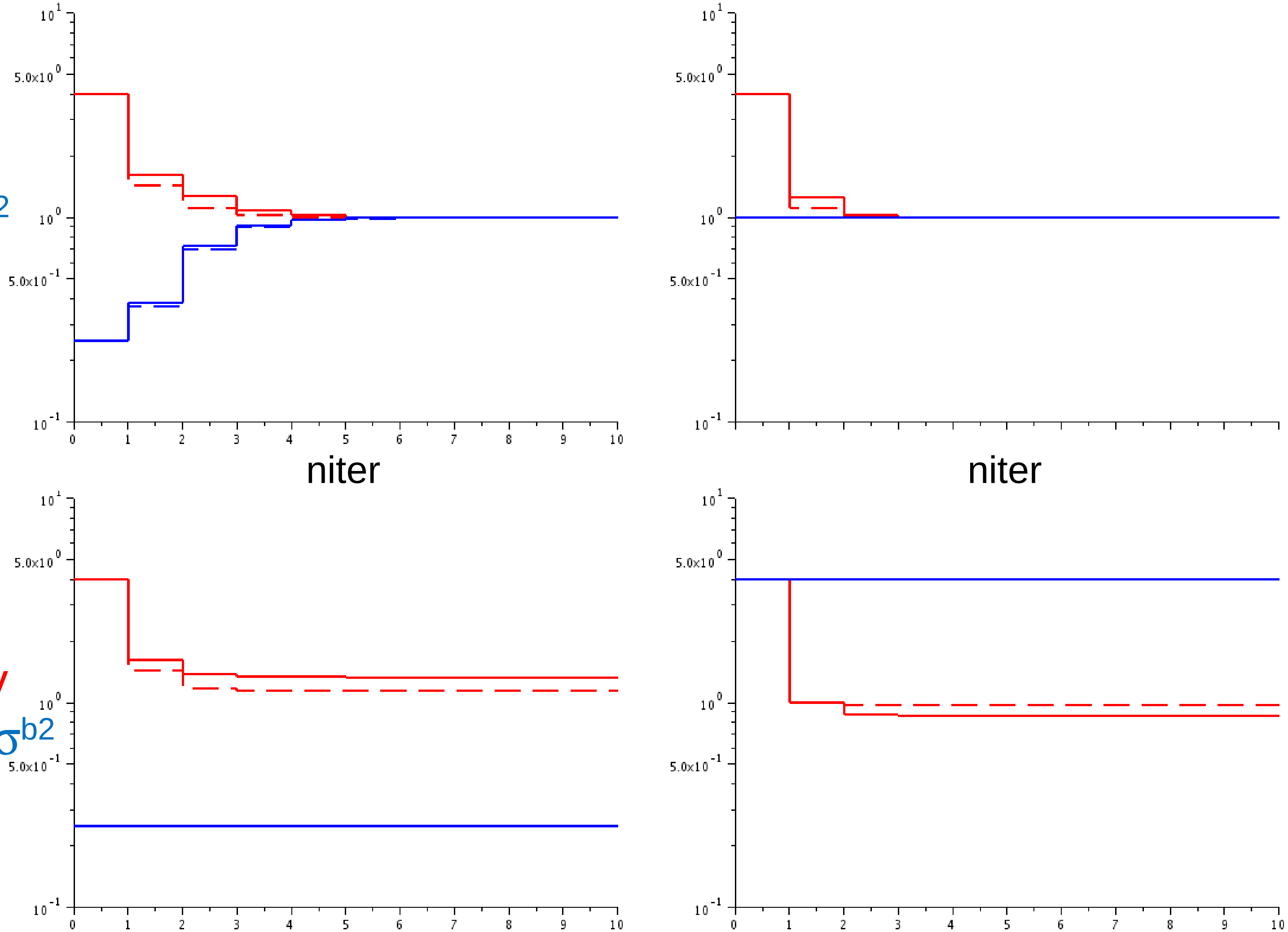
$(\sigma^{oi})^2 = 1 / p_i \, \Sigma_{k=1,pi} \, (y^{oi}_k - \boldsymbol{H}^i_k(\mathbf{x}^a)) \, (y^{oi}_k - \boldsymbol{H}^i_k(\mathbf{x}^b))_,$

where p_i is the number of observations in subset i.

- Easy to implement, especially to try to diagnose observation error covariances!

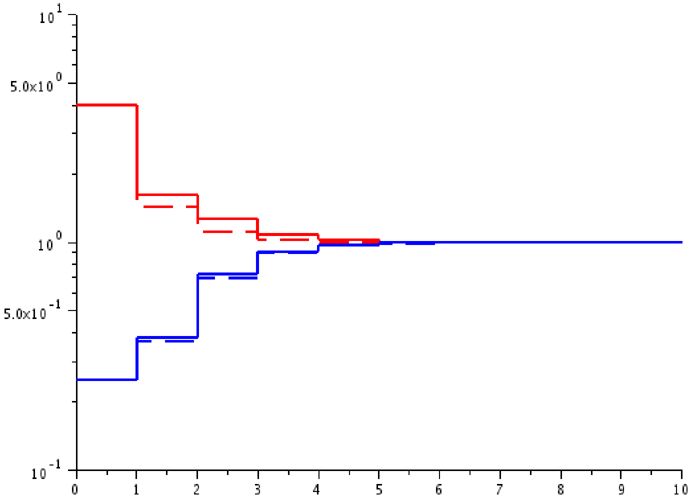
Diagnostics in observation space: mean σ^o / σ^b

σ^{ot} = 1 (L^{ot} = 0 km) / σ^{bt} = 1 (L^{bt} = 500 km)



Convergence of σ^o, σ^b

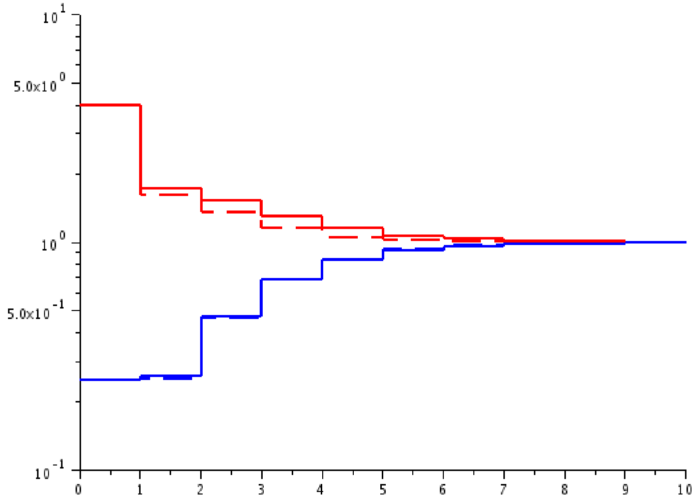
$\sigma^{o2} / \sigma^{b2}$



niter

L^b = 500 km

L^o = 0 km



niter

L^b = 250 km

L^o = 0 km

Spectral interpretation

Λ^{o} specified spectral variances (obs. error)

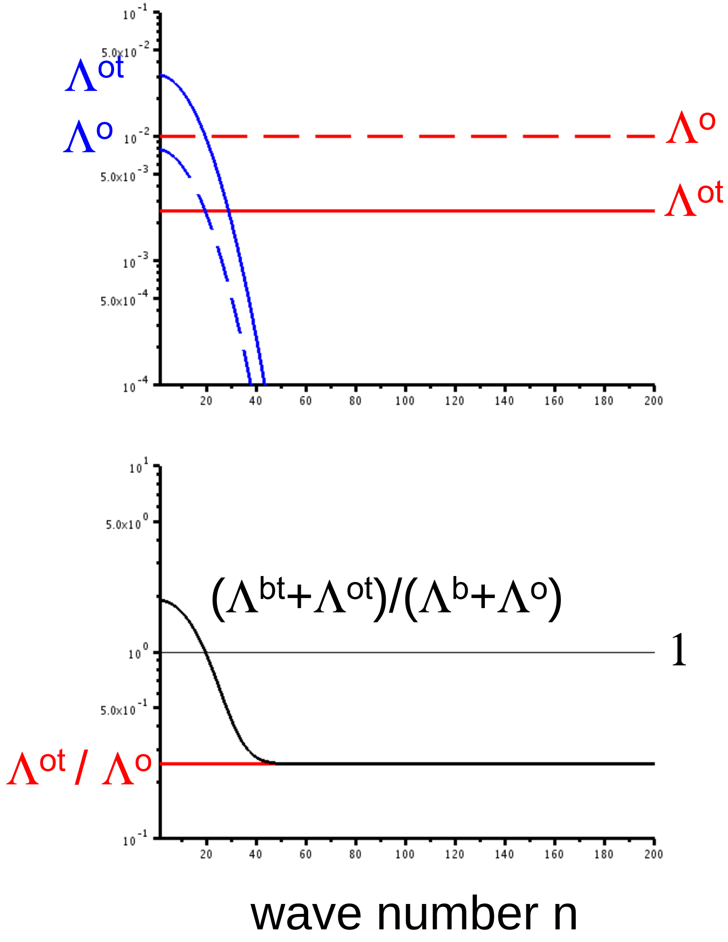
Λ^{ot} true spectral variances

Diagnosed spectral variances:

$$\Lambda^{o'} = \Lambda^{o} \, (\Lambda^{bt} + \Lambda^{ot}) / (\Lambda^{b} + \Lambda^{o})$$
$$\simeq \Lambda^{o} \quad \Lambda^{ot} / \Lambda^{o}$$
$$\simeq \Lambda^{ot}.$$

Spatial mean of variance:

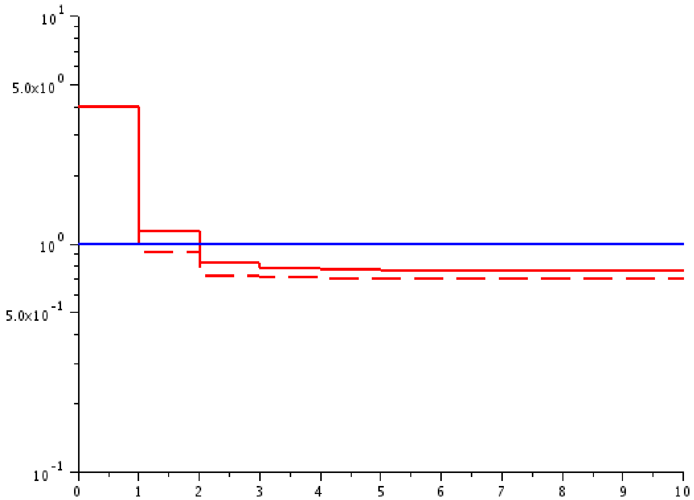
$$v^{o'} = \sum_{n} \Lambda^{o'}(n) \simeq v^{ot}.$$



Mis-specification of correlation in R (L^o = 0 km)

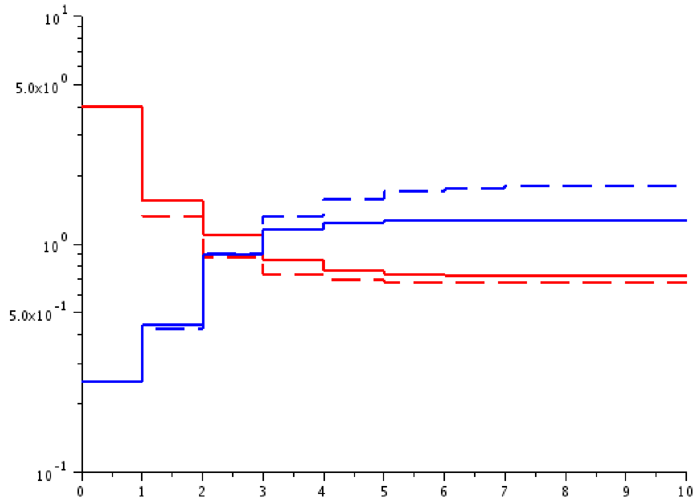
σ^{ot} =1 (L^{ot} =100km) / σ^{bt} =1 (L^{bt} =500km)

σ^{o2} / σ^{b2}



niter

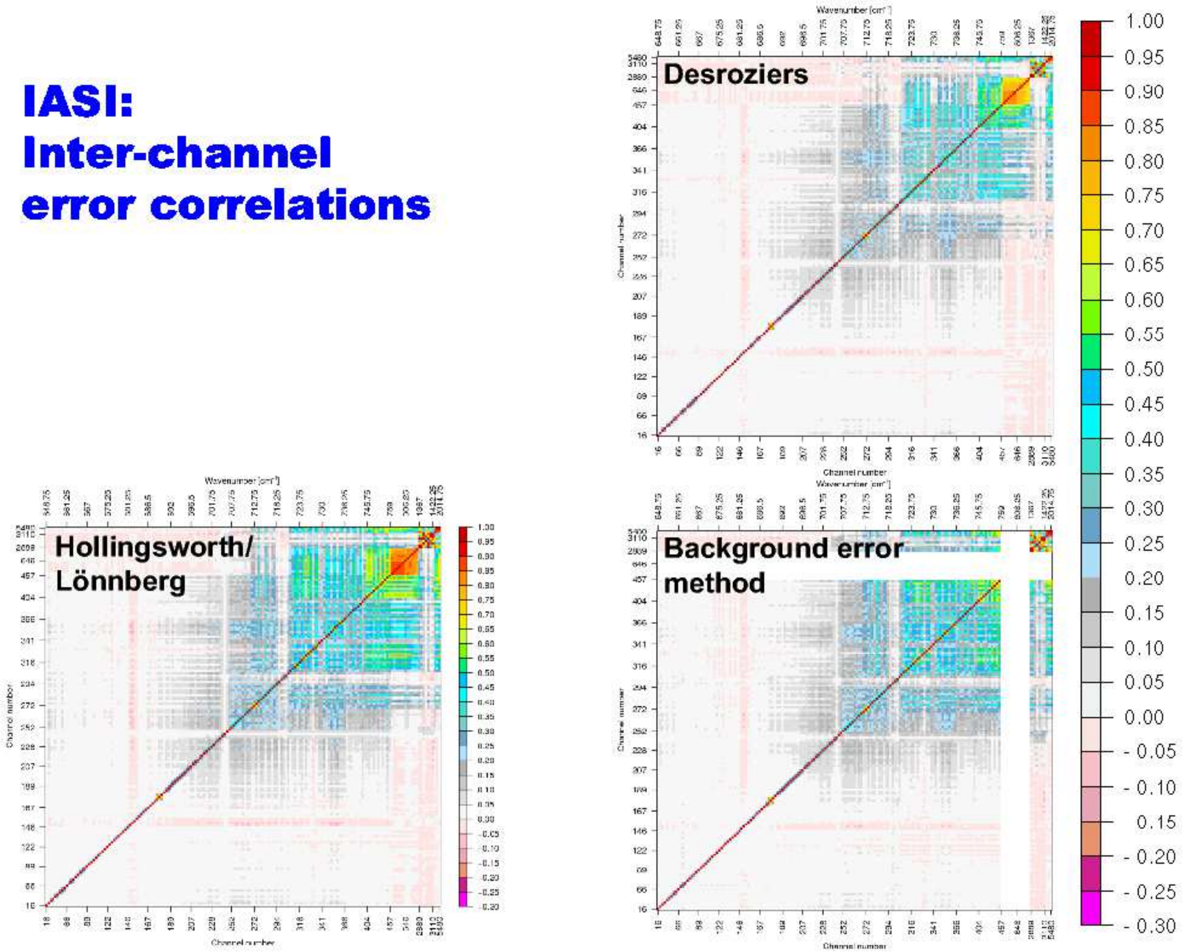
σ^o diagnosis only



niter

σ^o / σ^b diagnosis

IASI inter-channel error correlations



(Bormann et al, ECMWF, 2010)

Diagnostic of model error in 4D-Var $[t_{k-1}, t_k]$

- Observations / background departures at time t_k

$$\mathbf{d}_k = \mathbf{y}^o_k - H_k(M_k(\mathbf{x}^b_{k-1}))$$
$$\sim \mathbf{y}^o_k - H_k(M_k(\mathbf{x}^t_{k-1})) - \mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^b_{k-1}$$
$$\sim \mathbf{y}^o_k - H_k(\mathbf{x}^t_k + \boldsymbol{\varepsilon}^m_k)) - \mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^b_{k-1}$$
$$\sim \boldsymbol{\varepsilon}^o_k - \mathbf{H}_k \boldsymbol{\varepsilon}^m_k - \mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^b_{k-1}.$$

- Observations / analysis departures at t_k with analysis at t_{k-1} but using $\mathbf{y}^o_k$

$$\mathbf{d}^{oa}_k = \mathbf{y}^o_k - H_k(M_k(\mathbf{x}^a_{k-1}))$$
$$\sim \boldsymbol{\varepsilon}^o_k - \mathbf{H}_k \boldsymbol{\varepsilon}^m_k - \mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^a_{k-1}.$$

- A posteriori diagnostics

$$\mathrm{E}[\mathbf{d}^{oa}_k \, \mathbf{d}_k^T] = \mathrm{E}[(\boldsymbol{\varepsilon}^o_k - \mathbf{H}_k \boldsymbol{\varepsilon}^m_k) \, \mathbf{d}_k^T] - \mathrm{E}[(\mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^a_{k-1}) \, \mathbf{d}_k^T]$$
$$= \mathbf{R}_k + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^T - \mathrm{E}[(\mathbf{H}_k \mathbf{M}_k \boldsymbol{\varepsilon}^a_{k-1}) \, \mathbf{d}_k^T]$$
$$= \mathbf{R}_k + \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^T, \text{ if optimal } (\mathrm{E}[\boldsymbol{\varepsilon}^a_{k-1} \, \mathbf{d}_k^T] = \mathbf{0}).$$

(also see Todling 2015)

Outline

1. Introduction
2. « J_{min} » diagnostics
3. Observation space diagnostics
4. Combination of diagnostics
5. Conclusion

Lag-innovation correlation as a diagnostic of optimality

- Innovations used by analysis k-1

$\mathbf{d}_{k-1} \sim \varepsilon^{o}_{k-1} - \mathbf{H}_{k-1}\, \varepsilon^{b}_{k-1}.$

- Innovations used by analysis k

$$\begin{aligned}
\mathbf{d}_{k} &= \mathbf{y}^{o}_{k} - H_{k} (M_{k} (\mathbf{x}^{b}_{k-1} + \mathbf{K}_{k-1}\, \mathbf{d}_{k-1})) \\
&\sim \mathbf{y}^{o}_{k} - H_{k} (M_{k} (\mathbf{x}^{t}_{k-1}) + \mathbf{M}_{k} (\varepsilon^{b}_{k-1} + \mathbf{K}_{k-1}\, \mathbf{d}_{k-1})) \\
&\sim \mathbf{y}^{o}_{k} - H_{k} (\mathbf{x}^{t}_{k} + \varepsilon^{m}_{k} + \mathbf{M}_{k} (\varepsilon^{b}_{k-1} + \mathbf{K}_{k-1}\, \mathbf{d}_{k-1})) \\
&\sim \varepsilon^{o}_{k} - \mathbf{H}_{k}\, \mathbf{e}^{m}_{k} - \mathbf{H}_{k}\, \mathbf{M}_{k} (\varepsilon^{b}_{k-1} + \mathbf{K}_{k-1}\, \mathbf{d}_{k-1}).
\end{aligned}$$

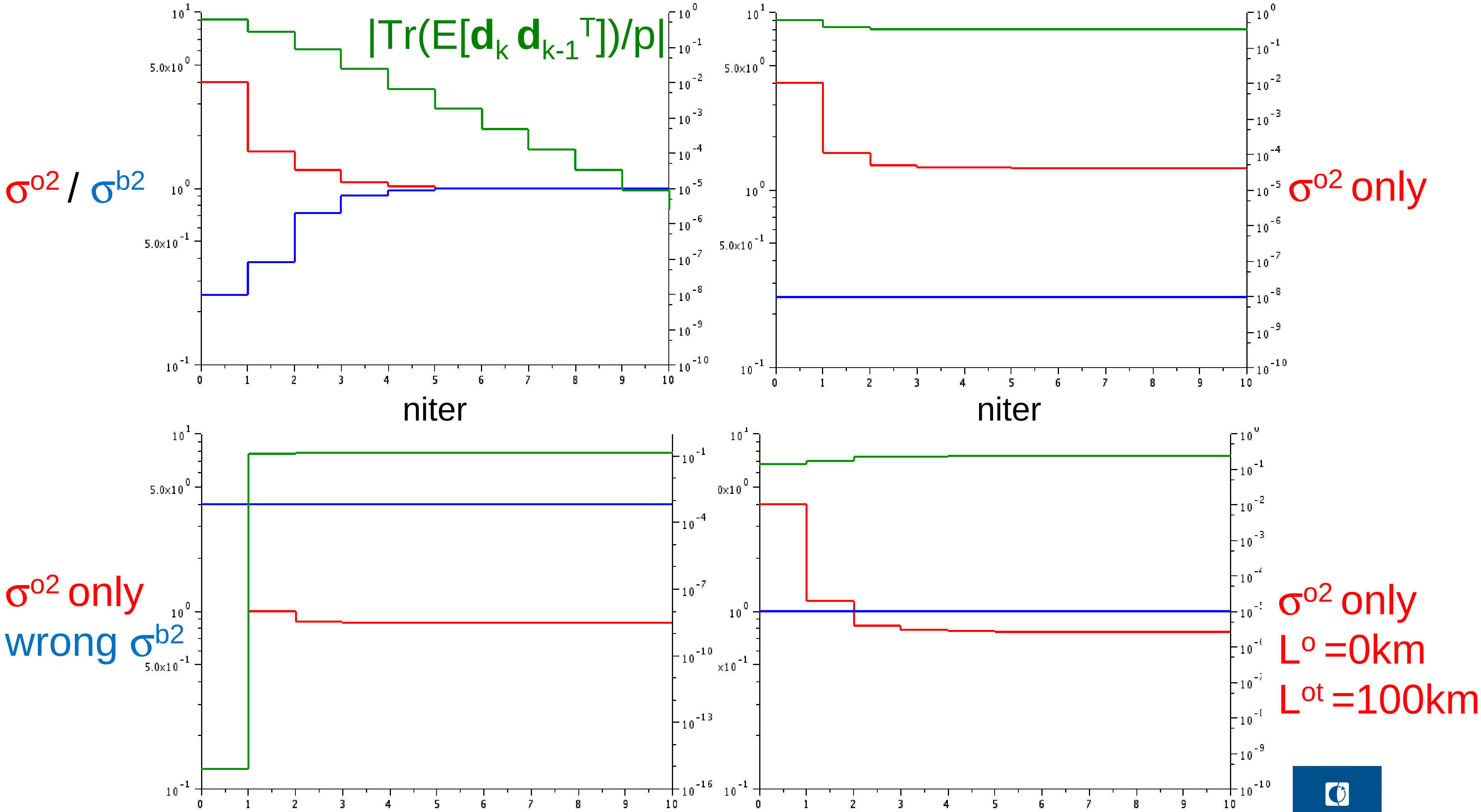
- Covariance of innovations

$$\begin{aligned}
\mathrm{E}[\mathbf{d}_{k}\, \mathbf{d}_{k-1}^{T}] &\sim \mathbf{H}_{k}\, \mathbf{M}_{k} (\mathbf{B}^{t}_{k-1}\, \mathbf{H}_{k-1}^{T} - \mathbf{K}_{k-1}\, \mathrm{E}[\mathbf{d}_{k-1}\, \mathbf{d}_{k-1}^{T}]) \\
&\sim \mathbf{H}_{k}\, \mathbf{M}_{k} (\mathbf{B}^{t}_{k-1}\, \mathbf{H}_{k-1}^{T}\, \mathrm{E}[\mathbf{d}_{k-1}\, \mathbf{d}_{k-1}^{T}]^{-1}\, \mathrm{E}[\mathbf{d}_{k-1}\, \mathbf{d}_{k-1}^{T}] - \mathbf{K}_{k-1})\, \mathrm{E}[\mathbf{d}_{k-1}\, \mathbf{d}_{k-1}^{T}] \\
&\sim \mathbf{H}_{k}\, \mathbf{M}_{k} (\mathbf{K}^{t}_{k-1} - \mathbf{K}_{k-1})\, \mathrm{E}[\mathbf{d}_{k-1}\, \mathbf{d}_{k-1}^{T}].
\end{aligned}$$

- $\mathrm{E}[\mathbf{d}_{k}\, \mathbf{d}_{k-1}^{T}] = 0$

Additional diagnostic of optimality (Ménard 2015)?

Diagnostics in observation space: mean σ^o / σ^b
σ^{ot} = 1 (L^{ot} = 0 km) / σ^{bt} = 1 (L^{bt} = 500 km)



Outline

1. Introduction
2. « J_{min} » diagnostics
3. Observation space diagnostics
4. Combination of diagnostics
5. Conclusion

Conclusion

- J_{min} diagnostics linked with basic foundations of BLUE estimation
- ✓ Give contributions of observations to the correction of background.
- ✓ Can be used to tune weights of **R** and **B** matrices.

- Observation space diagnostics
- ✓ Rather easy to implement.
- ✓ Well adapted to try to estimate **R**.
- ✓ Give apparently reasonable results in many cases.

- We need something else to validate tunings
- ✓ Impact on forecasts (compared to future independent observations).
- ✓ Compare tuned analysis to independent observations.
- ✓ Use (adapt) lag innovation correlation diagnostic.