Evaluation of assimilation algorithms
Part 5

Olivier Talagrand
School *Data Assimilation*
Nordic Institute for Theoretical Physics (NORDITA)
Stockholm, Sweden
30 April 2011

Best Linear Unbiased Estimate

State vector x, belonging to *state space* $\mathcal{S}$ ($\dim \mathcal{S} = n$), to be estimated.

Available data in the form of

- A '*background*' estimate, belonging to *state space*, with dimension n

  $x^b = x + \zeta^b$

- An additional set of data (*e. g.* observations), belonging to *observation space*, with dimension p

  $y = Hx + \varepsilon$

  H is known linear *observation operator*.

Best Linear Unbiased Estimate (continuation 2)

Assume $E(\boldsymbol{\zeta}^b) = 0$, $E(\boldsymbol{\varepsilon}) = 0$

Set $\boldsymbol{d} \equiv \boldsymbol{y} - H\boldsymbol{x}^b$ (innovation vector)

$$\boldsymbol{x}^a = \boldsymbol{x}^b - E(\boldsymbol{\zeta}^b\boldsymbol{d}^T)\,[E(\boldsymbol{d}\boldsymbol{d}^T)]^{-1}\,(\boldsymbol{y} - H\boldsymbol{x}^b)$$

$$\boldsymbol{P}^a = E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{bT}) - E(\boldsymbol{\zeta}^b\boldsymbol{d}^T)\,[E(\boldsymbol{d}\boldsymbol{d}^T)]^{-1}\,E(\boldsymbol{d}\boldsymbol{\zeta}^{bT})$$

Assume $E(\boldsymbol{\zeta}^b\boldsymbol{\varepsilon}^T) = 0$ (not restrictive). Set $E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{bT}) = \boldsymbol{P}^b$ (also often denoted $\boldsymbol{B}$), $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \boldsymbol{R}$

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b\boldsymbol{H}^T\,[\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1}\,(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$

$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b\boldsymbol{H}^T\,[\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1}\,\boldsymbol{H}\boldsymbol{P}^b$$

$\boldsymbol{x}^a$ is the *Best Linear Unbiased Estimate* (*BLUE*) of $\boldsymbol{x}$ from $\boldsymbol{x}^b$ and $\boldsymbol{y}$.

If probability distributions are *globally* gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

Determination of the *BLUE* requires (at least apparently) the *a priori* specification of the expectation and covariance matrix, *i. e.* the statistical moments of orders 1 and 2, of the errors. The expectation is required for unbiasing the data in the first place.

Questions

- Is it possible to objectively evaluate the quality of an assimilation system ?

- Is it possible to objectively evaluate the first- and second-order statistical moments of the data errors, whose specification is required for determining the *BLUE* ?

- Is it possible to objectively determine whether an assimilation system is optimal ?

- More generally, how to make the best of an assimilation system ?

Objective validation

Objective validation is possible only by comparison with unbiased *independent observations*, *i. e.* observations that have not been used in the asssimilation, and that are affected with errors that are statistically independent of the errors affecting the data used in the assimilation.

Amplitude of forecast error, if estimated against observations that are really independent of observations used in assimilation, is an objective measure of quality of assimilation.

$$x^b = x + \zeta^b$$
$$y = Hx + \varepsilon$$

The only combination of the data that is a function of only the error is the innovation vector

$$d = y - Hx^b = \varepsilon - H\zeta^b$$

Innovation is the only objective source of information on errors. Now innovation is a combination of background and observation errors, while determination of the *BLUE* requires explicit knowledge of the statistics of both observation and background errors.

$$\boldsymbol{x}^a = \boldsymbol{x}^b + P^b H^{\mathrm{T}} [HP^bH^{\mathrm{T}} + R]^{-1} (\boldsymbol{y} - H\boldsymbol{x}^b)$$

Innovation alone will never be sufficient to determine the required statistics.

With hypotheses made above

$$E(d) = 0 \quad ; \quad E(dd^{\mathrm{T}}) = HP^bH^{\mathrm{T}} + R$$

Possible to check statistical consistency between *a priori* assumed and *a posteriori* observed statistics of innovation.

Consider assimilation scheme of the form

$$\boldsymbol{x}^a = \boldsymbol{x}^b + K(\boldsymbol{y} - H\boldsymbol{x}^b) \qquad (1)$$

with any (*i. e.* not necessarily optimal) gain matrix K.

(1) $\Leftrightarrow$ if data are perfect, then so is the estimate $\boldsymbol{x}^a$.

Data-minus-Analysis (*DmA*) difference

$$\delta \equiv \begin{pmatrix} x^b - x^a \\ y - Hx^a \end{pmatrix} = \begin{pmatrix} -Kd \\ (I_p - HK)d \end{pmatrix}$$

For given gain matrix K, one-to-one correspondance $d \Leftrightarrow \delta$

It is exactly equivalent to compute statistics on either the innovation d or on the *DmA* difference δ.

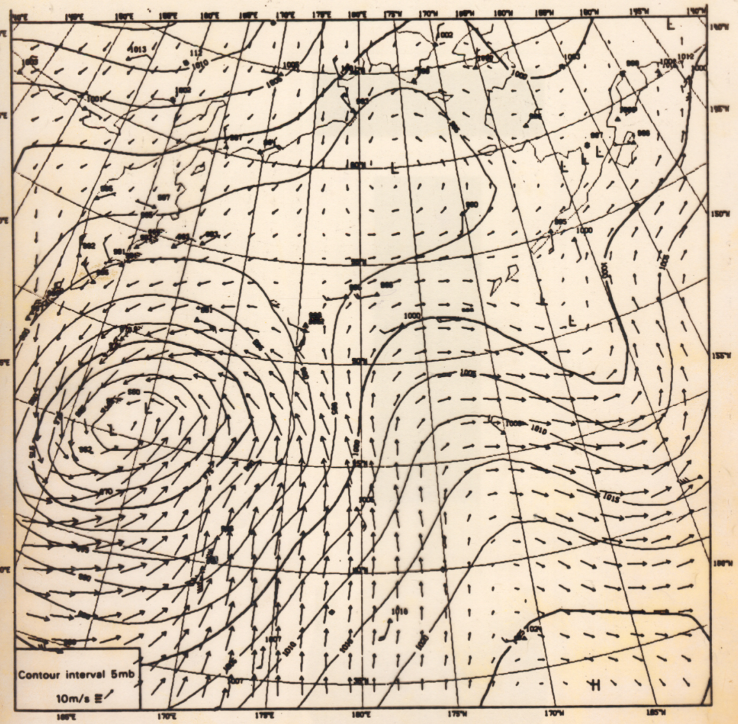


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

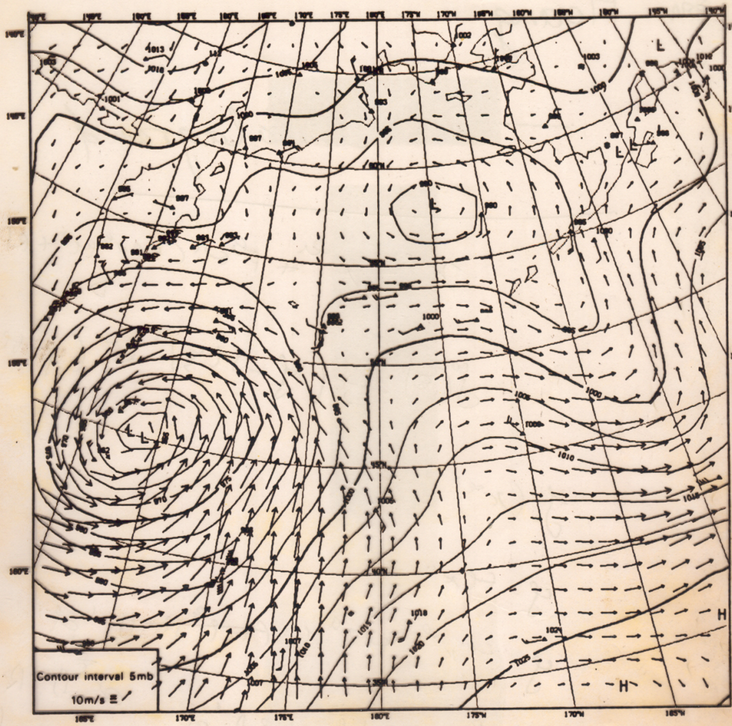


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

After A. Lorenc

For perfectly consistent system (*i. e.*, system that uses the exact error statistics):

$$E(\boldsymbol{d}) = 0 \ (\Leftrightarrow \ E(\boldsymbol{\delta}) = 0)$$

Any systematic bias in either the innovation vector or the DmA difference is the signature of an inappropriately taken into account bias in either the background or the observation (or both).

$$E[(\boldsymbol{x}^b-\boldsymbol{x}^a)(\boldsymbol{x}^b-\boldsymbol{x}^a)^{\mathrm{T}}] = \boldsymbol{P}^b - \boldsymbol{P}^a$$

$$E[(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^a)(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^a)^{\mathrm{T}}] = \boldsymbol{R} - \boldsymbol{H}\boldsymbol{P}^a\boldsymbol{H}^{\mathrm{T}}$$

A perfectly consistent analysis statistically fits the data to within their own accuracy.

If new data are added to (removed from) an optimal analysis system, *DmA* difference must increase (decrease).

Assume inconsistency has been found between *a priori* assumed and *a posteriori* observed statistics of innovation or DmA difference.

- What can be done ?

or, equivalently

- Which bounds does the knowledge of the statistics of innovation put on the error statistics whose knowledge is required by the *BLUE* ?

Data assumed to consist of a vector z, belonging to data space $\mathcal{D}$ ($\dim \mathcal{D} = m$), in the form

$$z = \Gamma x + \zeta$$

where Γ is a known (mxn)-matrix, and ζ an unknown 'error'

For instance

$$z = \begin{pmatrix} x^b = x + \zeta^b \\ y = Hx + \varepsilon \end{pmatrix}$$

which corresponds to

$$\Gamma = \begin{pmatrix} I_n \\ H \end{pmatrix} \qquad \zeta = \begin{pmatrix} \zeta^b \\ \varepsilon \end{pmatrix}$$

Look for estimated state vector x^a of the form

$$x^a = \alpha + Az$$

subject to

- invariance in change of origin in state space $\Rightarrow$ $A\Gamma = I_m$

- quadratic estimation error $E[(x^a_i - x_i)^2]$ minimum for any component x_i.

Solution

$\boldsymbol{x}^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1} \Gamma^{\mathrm{T}} S^{-1} [z - \boldsymbol{\mu}]$
$P^a \equiv E[(\boldsymbol{x}^a - \boldsymbol{x})(\boldsymbol{x}^a - \boldsymbol{x})^{\mathrm{T}}] = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1}$

where $\boldsymbol{\mu} \equiv E(\zeta)$, $S \equiv E(\zeta' \zeta'^{\mathrm{T}})$, $\zeta' \equiv \zeta - \boldsymbol{\mu}$

Requires (at least apparently) *a priori* explicit knowledge of $E(\zeta)$ and $E(\zeta' \zeta'^{\mathrm{T}})$

Unambiguously defined iff rank$\Gamma = n$. *Determinacy condition*. Requires $m \geq n$.
We shall set $m = n + p$.

Invariant in any invertible linear change of coordinates, either in data or state space.

In case ζ is gaussian, $\zeta = \mathcal{N}[\boldsymbol{\mu}, S]$, *BLUE* achieves bayesian estimation in the sense that $P(\boldsymbol{x} \mid z) = \mathcal{N}[\boldsymbol{x}^a, P^a]$

If determinacy condition is verified, it is always possible to decompose data vector $\boldsymbol{z}$ into

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b$$
$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

with $E(\boldsymbol{\zeta}^b) = 0 \quad ; E(\boldsymbol{\varepsilon}) = 0 \quad ; E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}}) = 0$

$\boldsymbol{x}^a$ is the same estimate (*BLUE*) as before, *viz.*,

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^a \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$
$$[\boldsymbol{P}^a]^{-1} = [\boldsymbol{P}^b]^{-1} + \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1}\boldsymbol{H}$$

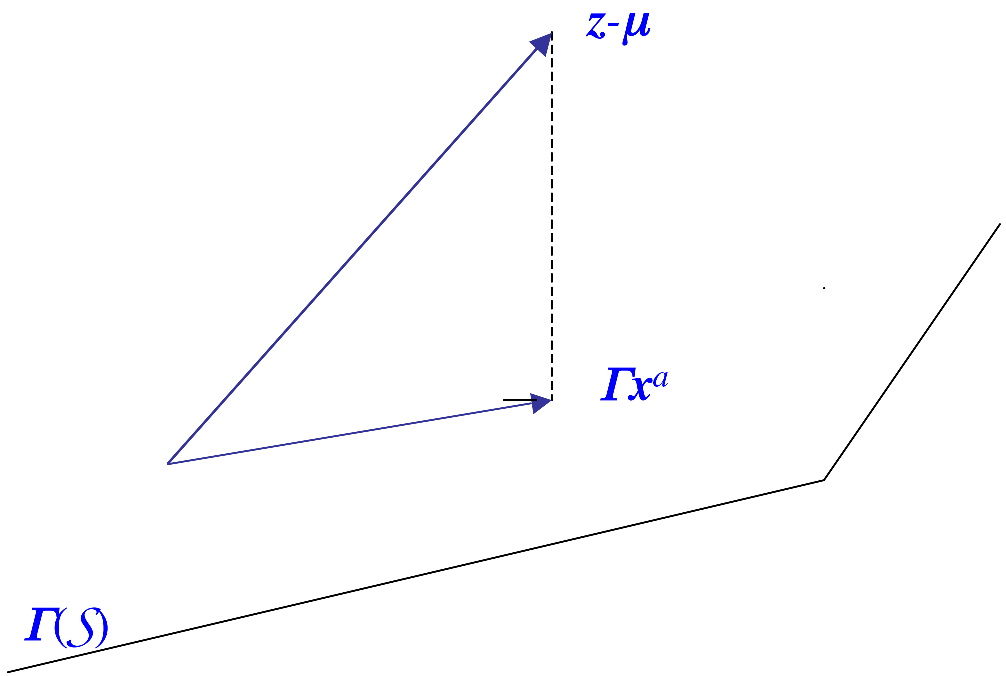
$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$
$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \boldsymbol{H}\boldsymbol{P}^b$$

Variational form.

$\boldsymbol{x}^a$ minimizes following scalar *objective function*, defined on state space $\mathcal{S}$

$$\mathcal{J}(\boldsymbol{\xi}) \equiv (1/2)\,[\boldsymbol{\Gamma\xi} - (\boldsymbol{z}-\boldsymbol{\mu})]^{\mathrm{T}}\, \boldsymbol{S}^{-1}\, [\boldsymbol{\Gamma\xi} - (\boldsymbol{z}-\boldsymbol{\mu})]$$

$$\mathcal{J}(\boldsymbol{\xi}) \equiv (1/2)\,[\boldsymbol{\Gamma}\boldsymbol{\xi} - (\boldsymbol{z}-\boldsymbol{\mu})]^{\mathrm{T}}\, S^{-1}\, [\boldsymbol{\Gamma}\boldsymbol{\xi} - (\boldsymbol{z}-\boldsymbol{\mu})]$$



Minimizing $\mathcal{J}(\xi)$ amounts to

- unbias z

- project orthogonally onto space $\Gamma(\mathcal{S})$ according to Mahalanobis S-metric

- take inverse through Γ (inverse unambiguously defined through determinacy condition)

Decompose data space $\mathcal{D}$ into image space $\Gamma(\mathcal{S})$ (index 1) and its $\mathcal{S}$-orthogonal space $\perp\Gamma(\mathcal{S})$ (index 2)

$$\Gamma = \begin{pmatrix} \Gamma_1 \\ 0 \end{pmatrix} \qquad \boldsymbol{\Gamma}_1 \text{ invertible} \qquad z = \begin{pmatrix} z_1 = \Gamma_1 x + \zeta_1 \\ z_2 = \zeta_2 \end{pmatrix}$$

Assume $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

Then

$$\boldsymbol{x}^a = \boldsymbol{\Gamma}_1^{-1} [\boldsymbol{z}_1 - \boldsymbol{\mu}_1]$$

$$\boldsymbol{x}^a = \boldsymbol{\Gamma}_1^{-1}\,[\boldsymbol{z}_1 - \boldsymbol{\mu}_1]$$

The probability distribution of the error

$$\boldsymbol{x}^a - \boldsymbol{x} = \boldsymbol{\Gamma}_1^{-1}\,[\boldsymbol{\zeta}_1 - \boldsymbol{\mu}_1]$$

depends on the probability distribution of $\boldsymbol{\zeta}_1$.

On the other hand, the probability distribution of

$$\boldsymbol{\delta} = (\boldsymbol{z}-\boldsymbol{\mu}) - \boldsymbol{\Gamma}\boldsymbol{x}^a = \begin{pmatrix} 0 \\ \zeta_2 - \mu_2 \end{pmatrix}$$

depends only on the probability distribution of $\boldsymbol{\zeta}_2$.

DmA difference, *i. e.* $(\boldsymbol{z}-\boldsymbol{\mu}) - \boldsymbol{\Gamma x}^a$, is in effect 'rejected' by the assimilation. Its expectation and covariance are irrelevant for the assimilation.

Consequence. Any assimilation scheme (*i. e.*, *a priori* subtracted bias and gain matrix K) is compatible with any observed statistics of either *DmA* or innovation. Not only is not consistency between *a priori* assumed and *a posteriori* observed statistics of innovation (or *DmA*) sufficient for optimality of an assimilation scheme, it is not even necessary.

Example

$$z_1 = x + \zeta_1$$
$$z_2 = x + \zeta_2$$

Errors ζ_1 and ζ_2 assumed to be centred ($E(\zeta_1) = E(\zeta_2) = 0$), to have same variance s and to be mutually uncorrelated. Then

$$x^a = (1/2)\,(z_1 + z_2)$$

with expected quadratic estimation error

$$E[(x^a - x)^2] = s/2$$

Innovation is difference $z_1 - z_2$. With above hypotheses, one expects to observe

$$E(z_1 - z_2) = 0 \quad ; \quad E[(z_1 - z_2)^2] = 2s$$

Assume one observes

$$E(z_1 - z_2) = b \quad ; \quad E[(z_1 - z_2)^2] = b^2 + 2\gamma$$

Inconsistency if $b \neq 0$ and/or $\gamma \neq s$

Inconsistency can always be resolved by assuming that

$$E(\zeta_1) = -E(\zeta_2) = -b/2$$

$$E(\zeta'_1{}^2) = E(\zeta'_2{}^2) = (s+\gamma)/2$$
$$E(\zeta'_1\zeta'_2) = (s-\gamma)/2$$

This alters neither the *BLUE* x^a, nor the corresponding quadratic estimation error $E[(x^a-x)^2]$.

Explanation. It is not necessary to know explicitly the complete expectation μ and covariance matrix S in order to perform the assimilation. It is necessary to know the projection of $\boldsymbol{\mu}$ and S onto the subspace $\Gamma(\mathcal{S})$. As for the subspace that is S-orthogonal to $\Gamma(\mathcal{S})$, it suffices to know what it is, but it is not necessary to know the projection of $\boldsymbol{\mu}$ and S onto it. A number of degrees of freedom are therefore useless for the assimilation. The parameters determined by the statistics of $\boldsymbol{d}$ are equal in number to those useless degrees of freedom, to which any inconsistency between *a priori* and *a posteriori* statistics of the innovation can always mathematically be attributed.

However, it may be that resolving the inconsistency in that way requires conditions that are (independently) known to be very unlikely, if not simply impossible. For instance, in the above example, consistency when $\gamma \neq s$ requires the errors ζ_1 and ζ_2 to be mutually correlated, which may be known to be very unlikely.

That result, which is purely mathematical, means that the specification of the error statistics required by the assimilation must always be based, in the last resort, on external hypotheses, *i. e.* on hypotheses that cannot be validated on the basis of the innovation alone. Now, such knowledge always exists.

Problem. Identify hypotheses

- That will not be questioned (errors on observation perfomed a long distance apart by radiosondes made by different manufacturers are uncorrelated)

- That sound reasonable, but may be questioned (observation and background errors are uncorrelated)

- That are undoubtedly questionable (model errors are negligible)

Ideally, define a minimum set of hypotheses such that all remaining undetermined error statistics can be objectively determined from observed statistics of innovation.

Objective function

$$\mathcal{J}(\boldsymbol{\xi}) \equiv (1/2)\,[\boldsymbol{\Gamma\xi} - \boldsymbol{z}]^{\mathrm{T}}\, S^{-1}\, [\boldsymbol{\Gamma\xi} - \boldsymbol{z}]$$

$$\mathcal{J}_{min} \equiv \mathcal{J}(\boldsymbol{x}^a) = (1/2)\,[\boldsymbol{\Gamma x}^a - \boldsymbol{z}]^{\mathrm{T}}\, S^{-1}\, [\boldsymbol{\Gamma x}^a - \boldsymbol{z}]$$

$$= (1/2)\,\boldsymbol{d}^{\mathrm{T}}\,[E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}})]^{-1}\,\boldsymbol{d}$$

$$\Rightarrow \qquad E(\mathcal{J}_{min}) = p/2 \qquad (p = \dim \boldsymbol{y} = \dim \boldsymbol{d})$$

If p is large, a few realizations are sufficient for determining $E(\mathcal{J}_{min})$
Often called χ^2 criterion.

Remark. If in addition errors are gaussian $Var(\mathcal{J}_{min}) = p/2$

Results for ECMWF (January 2003, $n = 8\ 10^6$)

- Operations ($p = 1.4\ 10^6$, has almost doubled since then)

$$2\mathcal{J}_{min}/p = 0.40 - 0.45$$

Innovation is significantly smaller than implied by $\boldsymbol{P}^b$ and $\boldsymbol{R}$ (a residual bias in $\boldsymbol{d}$ would make $\mathcal{J}_{min}$ too large).

- Assimilation without satellite observations ($p = 2 - 3\ 10^5$)

$$2\mathcal{J}_{min}/p = 1. - 1.05$$

Similar results obtained at other NWP centres (C. Fischer, W. Sadiki with Aladin model, T. Payne at Meteorological Office, UK).

Probable explanation: error variance of satellite observations overestimated in order to compensate for ignored spatial correlation.

Informative content

Objective function

$$\mathcal{J}(\boldsymbol{\xi}) = \Sigma_k \mathcal{J}_k(\boldsymbol{\xi})$$

where

$$\mathcal{J}_k(\boldsymbol{\xi}) \equiv (1/2)\,(H_k\boldsymbol{\xi} - \boldsymbol{y}_k)^{\mathrm{T}}\, S_k^{-1}\,(H_k\boldsymbol{\xi} - \boldsymbol{y}_k)$$

with $\quad \dim \boldsymbol{y}_k = m_k$

Accuracy of analysis

$$P^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1}$$

$$[P^a]^{-1} = \Sigma_k\, H_k^{\mathrm{T}}\, S_k^{-1}\, H_k$$

$$\begin{aligned} 1 &= (1/n)\, \Sigma_k\, \mathrm{tr}(P^a H_k^{\mathrm{T}}\, S_k^{-1}\, H_k) \\ &= (1/n)\, \Sigma_k\, \mathrm{tr}(S_k^{-1/2}\, H_k\, P^a\, H_k^{\mathrm{T}}\, S_k^{-1/2}) \end{aligned}$$

Informative content (continuation 1)

$$(1/n)\ \Sigma_k\ \mathrm{tr}(S_k^{-1/2}\ H_k\ P^a\ H_k^{\mathrm{T}}\ S_k^{-1/2}) = 1$$

$I(\boldsymbol{y}_k) \equiv (1/n)\ \mathrm{tr}(S_k^{-1/2}\ H_k\ P^a\ H_k^{\mathrm{T}}\ S_k^{-1/2})$ is a measure of the relative contribution of subset of data $\boldsymbol{y}_k$ to overall accuracy of assimilation. Invariant in linear change of coordinates in data space ⇒ valid for *any* subset of data.

In particular

$$I(\boldsymbol{x}^b) = (1/n)\ \mathrm{tr}[P^a\,(P^b)^{-1}]\ = 1 - (1/n)\ \mathrm{tr}(KH)$$

$$I(\boldsymbol{y}) = (1/n)\ \mathrm{tr}(KH)$$

Rodgers, 2000, calls those quantities *Degrees of Freedom for Signal*, or *for Noise*, depending on whether considered subset belongs to 'observations' or 'background'.

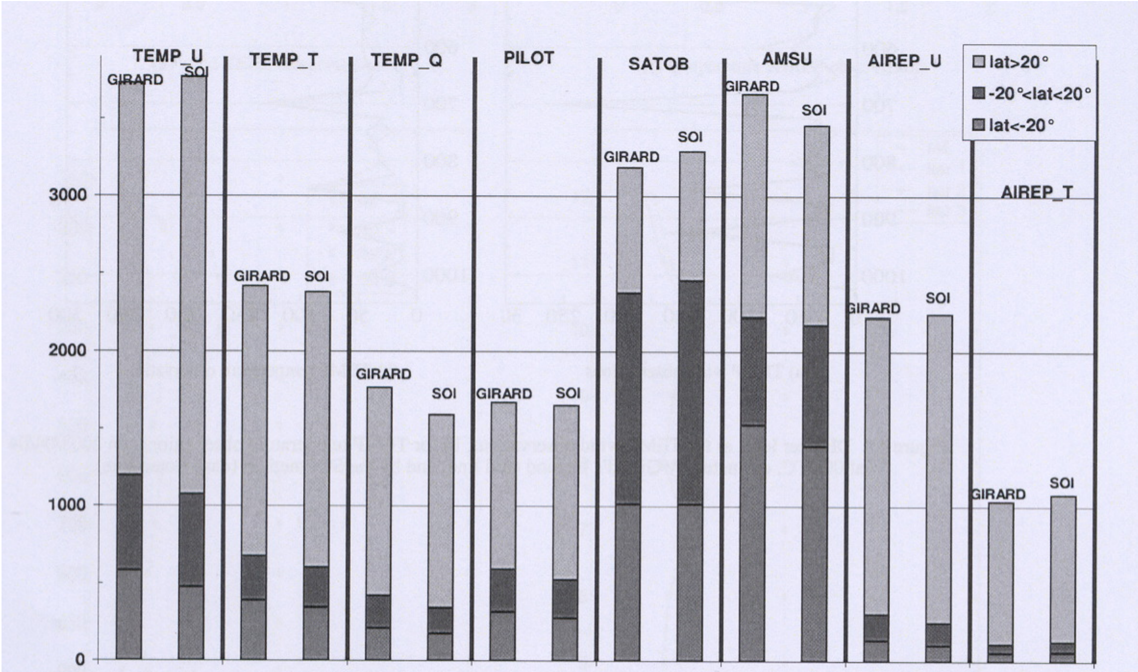
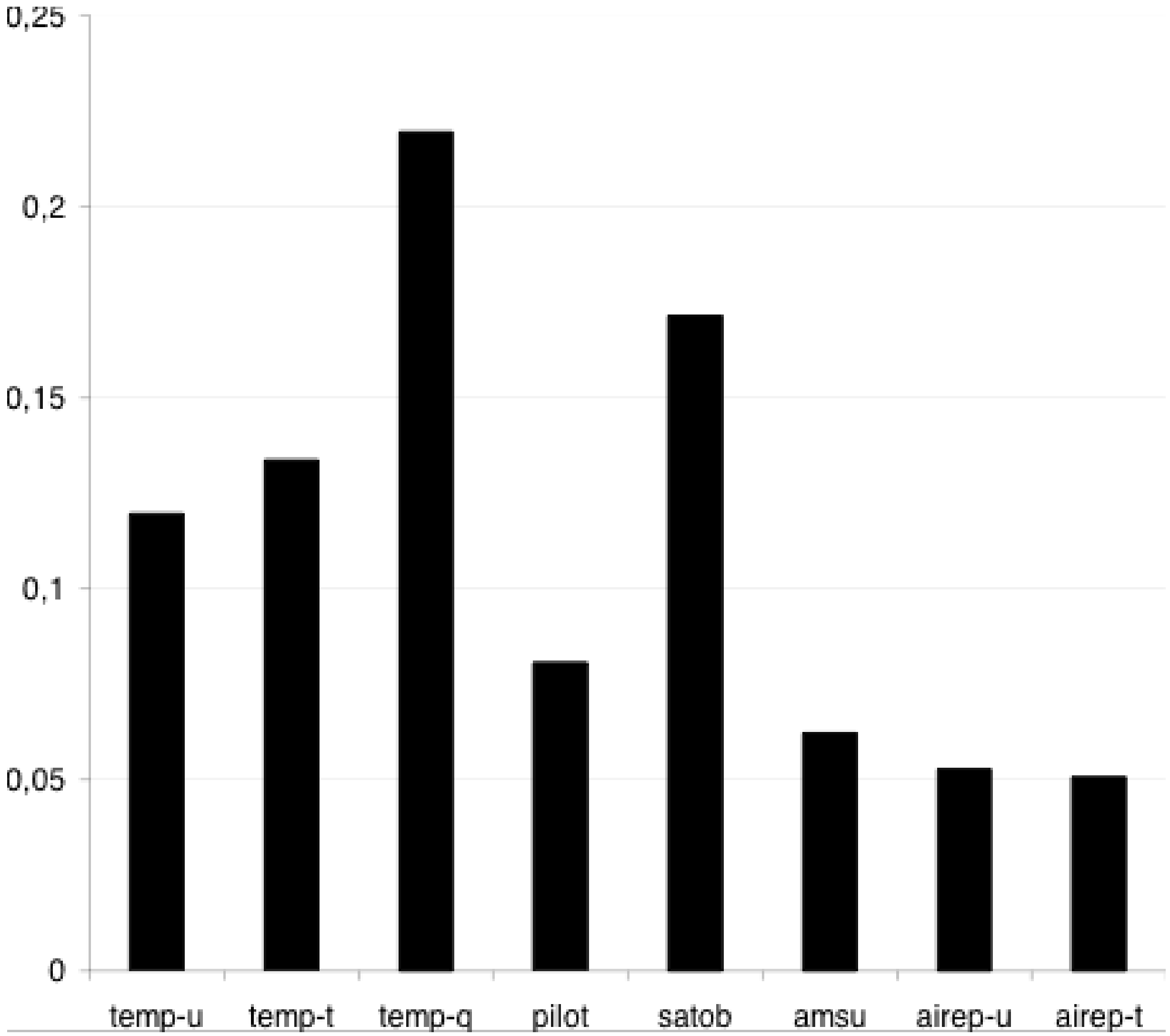


Figure 3. DFS for several subsets of observations. The subsets are separated by thick black lines. For each subset the left hand side bar was computed with Girard's method and the right hand side bar was computed with the SOI method. Each bar is divided into three parts, showing the contribution from the north hemisphere observations, the tropical observations and the southern observations. TEMP U and AIREP U types account for both zonal and

Informative content of subsets of observations (Arpège Assimilation System, Météo-France)

Chapnik *et al.*, 2006, *QJRMS*, **132**, 543-565



Informative content per individual (scalar) observation (courtesy B. Chapnik)

Objective function

$$\mathcal{J}(\xi) = \Sigma_k \mathcal{J}_k(\xi)$$

where

$$\mathcal{J}_k(\xi) \equiv (1/2)\,(H_k\xi - \boldsymbol{y}_k)^{\mathrm{T}}\, S_k^{-1}\,(H_k\xi - \boldsymbol{y}_k)$$

with dim$\boldsymbol{y}_k = m_k$

For a perfectly consistent system

$$E[\mathcal{J}_k(x^a)] = (1/2)\,[m_k - \mathrm{tr}(S_k^{-1/2}\, H_k\, P^a\, H_k^{\mathrm{T}}\, S_k^{-1/2})]$$

(in particular, $E(\mathcal{J}_{min}) = p/2$)

For same vector dimension m_k, more informative data subsets lead at the minimum to smaller terms in the objective function.

Equality

$$E[\mathcal{J}_k(x^a)] = (1/2)\,[m_k - \mathrm{tr}(S_k^{-1/2} H_k P^a H_k^{\mathrm{T}} S_k^{-1/2})] \qquad (1)$$

can be objectively checked.

Chapnik *et al*. (2004, 2005). Multiply each observation error covariance matrix S_k by a coefficient α_k such that (1) is verified simultaneously for all observation types.

System of equations fot the α_k's solved iteratively.

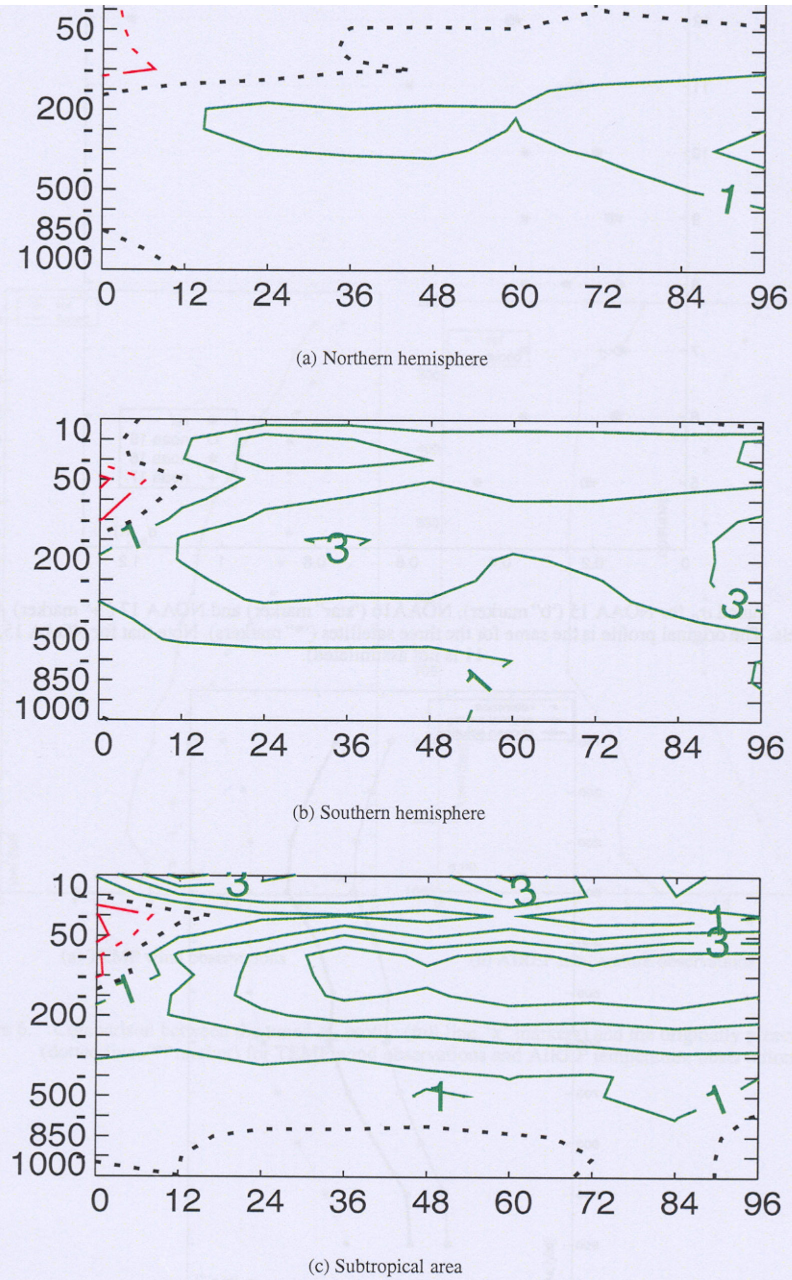


Figure 9. Difference between tunned rms (tuned geopotential forecasts - geopotential TEMP obserrvations) and the operational rms computed over 21 siutations. the x axis is the forecast term and the y axis is the vertical pressure level. Dashed lines mean that the tuned forecast is further from the observations than the operational one (degradation), on the contrary the solid lines mean that the tuned forecast is better than the operational. the difference between two colored line is 1 m. Subpanel a is for the northern hemisphere, subpanel b for the southern

Chapnik *et al*., 2006, *QJRMS*, **132**, 543-565

Informative content (continuation 2)

$$I(\boldsymbol{y}_k) \equiv (1/n)\ \mathrm{tr}(S_k^{-1/2}\, H_k\, P^a\, H_k^{\mathrm{T}}\, S_k^{-1/2})$$

Two subsets of data $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$

If errors affecting $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are uncorrelated, then $I(\boldsymbol{z}_1 \cup \boldsymbol{z}_2) = I(\boldsymbol{z}_1) + I(\boldsymbol{z}_2)$

If errors are correlated $I(\boldsymbol{z}_1 \cup \boldsymbol{z}_2) \neq I(\boldsymbol{z}_1) + I(\boldsymbol{z}_2)$

Informative content (continuation 3)

Example 1

$z_1 = x + \zeta_1$
$z_2 = x + \zeta_2$

Errors ζ_1 and ζ_2 assumed to centred, to have same variance and correlation coefficient c.

$$I(z_1) \;=\; I(z_2) \;=\; (1/2)\,(1 + c)$$

Example 2

State vector $\boldsymbol{x}$ evolving in time according to

$$\boldsymbol{x}_2 \;=\; \alpha\, \boldsymbol{x}_1$$

Observations are performed at times 1 and 2. Observation errors are assumed centred, uncorrelated and with same variance. Information contents are then in ratio $(1/\alpha\,,\,\alpha)$. For an unstable system ($\alpha > 1$), later observation contains more information (and the opposite for a stable system).

Informative content (continuation 4)

Subset $\boldsymbol{u}_1$ of analyzed fields, $dim\boldsymbol{u}_1 = n_1$. Define relative contribution of subset $\boldsymbol{y}_k$ of data to accuracy of $\boldsymbol{u}_1$?

$\boldsymbol{u}_2$: component of $\boldsymbol{x}$ orthogonal to $\boldsymbol{u}_1$ with respect to Mahalanobis norm associated with P^a (analysis errors on $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are uncorrelated).

$\boldsymbol{x} = (\boldsymbol{u}_1{}^{\mathrm{T}}, \boldsymbol{u}_2{}^{\mathrm{T}})^{\mathrm{T}}$. In basis $(\boldsymbol{u}_1, \boldsymbol{u}_2)$

$$P^a = \begin{pmatrix} P^a{}_1 & 0 \\ 0 & P^a{}_2 \end{pmatrix}$$

Informative content (continuation 5)

Observation operator H_k decomposes into

$$H_k = (H_{k1}, H_{k2})$$

and expression of estimation error covariance matrix into

$$[P^a{}_1]^{-1} = \Sigma_k \; H_{k1}{}^{\mathrm{T}} \, S_k{}^{-1} \, H_{k1}$$

$$[P^a{}_2]^{-1} = \Sigma_k \; H_{k2}{}^{\mathrm{T}} \, S_k{}^{-1} \, H_{k2}$$

Same development as before shows that the quantity

$$(1/n_1) \; \mathrm{tr}(S_k{}^{-1/2} \, H_{k1} \, P^a{}_1 \, H_{k1}{}^{\mathrm{T}} \, S_k{}^{-1/2})$$

is a measure of the relative contribution of subset $\boldsymbol{y}_k$ of data to analysis of subset $\boldsymbol{u}_1$ of state vector.

But can it be computed in practice for large dimension systems (requires the explicit decomposition $\boldsymbol{x} = (\boldsymbol{u}_1{}^{\mathrm{T}}, \boldsymbol{u}_2{}^{\mathrm{T}})^{\mathrm{T}}$) ?

Other possible diagnostics (Desroziers *et al*., 2006)

For a consistent system, with uncorrelation between background and observation errors

$$E[\boldsymbol{H}(\boldsymbol{x}^a\text{-}\boldsymbol{x}^b)(\boldsymbol{y}\text{-}\boldsymbol{H}\boldsymbol{x}^b)^{\mathrm{T}}] = E[\boldsymbol{H}(\boldsymbol{x}^a\text{-}\boldsymbol{x}^b)\boldsymbol{d}^{\mathrm{T}}] = \boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

$$E[(\boldsymbol{y}\text{-}\boldsymbol{H}\boldsymbol{x}^a)(\boldsymbol{y}\text{-}\boldsymbol{H}\boldsymbol{x}^b)^{\mathrm{T}}] = E[(\boldsymbol{y}\text{-}\boldsymbol{H}\boldsymbol{x}^a)\boldsymbol{d}^{\mathrm{T}}] = \boldsymbol{R}$$

Optimality

Equation

$$\boldsymbol{x}^a = \boldsymbol{x}^b - E(\boldsymbol{\zeta}^b \boldsymbol{d}^{\mathrm{T}})\ [E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}})]^{-1}\ (\boldsymbol{y} - H\boldsymbol{x}^b)$$

means that estimation error $\boldsymbol{x} - \boldsymbol{x}^a$ is uncorrelated with innovation $\boldsymbol{y} - H\boldsymbol{x}^b$ (if it was not, it would be possible to improve on $\boldsymbol{x}^a$ by statistical linear estimation).

Independent unbiased observation

$$\boldsymbol{v} = C\boldsymbol{x} + \boldsymbol{\gamma}$$

Fit to analysis $\quad \boldsymbol{v} - C\boldsymbol{x}^a = C(\boldsymbol{x} - \boldsymbol{x}^a) + \boldsymbol{\gamma}$

$$E[(\boldsymbol{v} - C\boldsymbol{x}^a)\ \boldsymbol{d}^{\mathrm{T}}] = CE[(\boldsymbol{x} - \boldsymbol{x}^a)\ \boldsymbol{d}^{\mathrm{T}}] + E(\boldsymbol{\gamma}\boldsymbol{d}^{\mathrm{T}})$$

First term is 0 if analysis is optimal, second is 0 if observation $\boldsymbol{v}$ is independent from previous data.

Daley (1992)

Conclusions

- **Absolute evaluation of analysis schemes, and comparison between different schemes**

  Can be evaluated only against independent unbiased data (independence and unbiasedness cannot be objectively checked). Fundamental, but not much to say.

- **Determination of required statistics**

  Impossible to achieve in a purely objective way. Will always require physical knowledge, educated guess, interaction with instrumentalists and modelers, and the like.

  Inconsistencies in specification of statistics can be objectively diagnosed, and can help in improving assimilation.

  For given error statistics, possible to quantify relative contribution of each subset of data to analysis of each subset of state vector.

  (and also Generalized Cross-Validation, Adaptive Filtering)

- **Optimality of analysis schemes**

  Optimality in the sense of least error variance can be objectively checked against independent unbiased data.