École Doctorale des Sciences de l'Environnement d'Île-de-France

Année Universitaire 2019-2020

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 8

26 Mai 2020

- Particle filters. Principle. Variants (*Proposal Densities*). A few results.

- Validation of assimilation algorithms

- Conclusions

Exact bayesian estimation ?

Particle filters

Predicted ensemble at time t : $\{x^b{}_l,\ l = 1, \ldots, L\}$, each element with its own weight (probability) $P(x^b{}_l)$

Observation vector at same time : $y = H(x) + \varepsilon$

Bayes' formula

$$P(x^b{}_l|y) \sim P(y|x^b{}_l)\ P(x^b{}_l)$$

Defines updating of weights

Bayes' formula

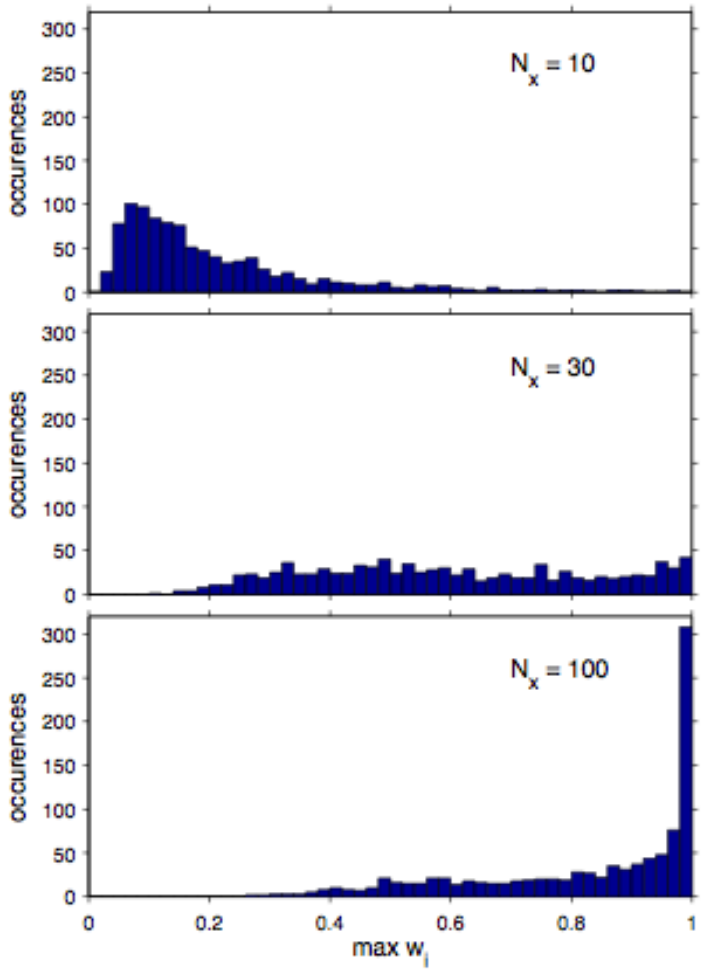
$$P(x^b{}_l|y) \sim P(y|x^b{}_l)\, P(x^b{}_l)$$

Defines updating of weights; particles are not modified. Asymptotically converges to bayesian pdf. Very easy to implement.

Observed fact. For large state dimension, ensemble tends to collapse.

Behavior of $\max w^i$

- $N_e = 10^3$; $N_x = 10, 30, 100$; 10^3 realizations



average squared error of posterior mean $= 5.5$

$\ldots = 25$

$\ldots = 127$

C. Snyder, http://www.cawcr.gov.au/staff/pxs/wmoda5/Oral/Snyder.pdf

Problem originates in the 'curse of dimensionality'. Large dimension pdf's are very diffuse, so that very few particles (if any) are present in areas where conditional probability ('*likelihood*') $P(y|x)$ is large.

Curse of dimensionality

Standard one-dimensional gaussian random variable X

$$P[\,|X| < \sigma\,] \approx 0.84$$

In dimension $n = 100$, $0.84^{100} = 3.10^{-8}$

.

Bengtsson *et al*. (2008) and Snyder *et al*. (2008) evaluate that stability of filter requires the size of ensembles to increase exponentially with space dimension.

Alternative possibilities (review in van Leeuwen, 2017, *Annales de la faculté des sciences de Toulouse Mathématiques*, **26** (4), 1051-1085)

Resampling. Define new ensemble.

Simplest way. Draw new ensemble according to probability distribution defined by the updated weights. Give same weight to all particles. Particles are not modified, but particles with low weights are likely to be eliminated, while particles with large weights are likely to be drawn repeatedly. For multiple particles, add noise, either from the start, or in the form of 'model noise' in ensuing temporal integration.

Random character of the sampling introduces noise. Alternatives exist, such as *residual sampling* (Lui and Chen, 1998, van Leeuwen, 2003). Updated weights w_l are multiplied by ensemble dimension L. Then p copies of each particle l are taken, where p is the integer part of Lw_l. Remaining particles, if needed, are taken randomly from the resulting distribution.

However, resampling is not sufficient to avoid degeneracy of filters.

Idea :

Use a *proposal density* that is closer to the new observations than the density defined by the predicted particles (for instance the density defined by EnKF, after the latter has used the new observations).

We are now to discuss a very interesting property of particle filters that has received little attention in the geophysical community. We start from Bayes:

$$p(x^{0:n}|y^{0:n}) = \frac{p(y^n|x^n)p(x^n|x^{n-1})}{p(y^n)} p(x^{0:n-1}|y^{1:n-1}). \quad (5.1)$$

To simplify the analysis, and since we concentrate on a filter here, let us first integrate out the past, to get:

$$p(x^n|y^{0:n}) = \frac{p(y^n|x^n)}{p(y^n)} \int p(x^n|x^{n-1})p(x^{n-1}|y^{1:n-1})\, \mathrm{d}x^{n-1}. \quad (5.2)$$

This expression does not change when we multiply and divide by a so-called proposal transition density $q(x^n|x^{n-1}, y^n)$, so:

$$p(x^n|y^{0:n})$$
$$= \frac{p(y^n|x^n)}{p(y^n)} \int \frac{p(x^n|x^{n-1})}{q(x^n|x^{n-1}, y^n)} q(x^n|x^{n-1}, y^n)p(x^{n-1}|y^{1:n-1})\, \mathrm{d}x^{n-1}. \quad (5.3)$$

van Leeuwen, 2017, *Annales de la faculté des sciences de Toulouse Mathématiques*, **26** (4), 1051-1085

As long as the support of $q(x^n|x^{n-1}, y^n)$ is equal to or larger than that of $p(x^n|x^{n-1})$ we can always do this. This last condition makes sure we don't divide by zero. Let us now assume that we have an equal-weight ensemble of particles from the previous analysis at time $n-1$, so

$$p(x^{n-1}|y^{0:n-1}) = \sum_{i=1}^{N} \frac{1}{N} \delta_{x_i^{n-1}}. \tag{5.4}$$

Using this in the equation above gives:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} \frac{1}{N} \frac{p(y^n|x^n)}{p(y^n)} \frac{p(x^n|x_i^{n-1})}{q(x^n|x_i^{n-1}, y^n)} q(x^n|x_i^{n-1}, y^n). \tag{5.5}$$

As a last step, we run the particles from time $n-1$ to n, i.e. we sample from the transition density. However, instead of drawing from $p(x^n|x_i^{n-1})$, so running the original model, we sample from $q(x^n|x_i^{n-1}, y^n)$, so from a modified model. Let us write this modified model as

$$x^n = g(x^{n-1}, y^n) + \hat{\beta}^n \tag{5.6}$$

so that we can write for the transition density, assuming $\hat{\beta}^n$ is Gaussian distributed with covariance $\hat{Q}$:

$$q(x^n|x^{n-1}, y^n) = N(g(x^{n-1}, y^n), \hat{Q}). \tag{5.7}$$

van Leeuwen, 2017, *ibid.*

Drawing from this density leads to:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} \frac{1}{N} \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)} \delta(x^n - x_i^n) \tag{5.8}$$

so the posterior pdf at time n can be written as:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} w_i \delta_{x_i^n} \tag{5.9}$$

with weights w_i given by:

$$w_i = \frac{1}{N} \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}. \tag{5.10}$$

We recognise the first factor in this expression as the likelihood, and the second as a factor related to using the proposal transition density instead of the original transition density to propagate from time $n-1$ to n, so it is related to the use of the proposed model instead of the original model. Note that because the factor $1/N$ and $p(y^n)$ are the same for each particle and we are only interested in relative weights, we will drop them from now on, so

$$w_i = p(y^n|x_i^n) \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}. \tag{5.11}$$

Several variants of proposal densities have been defined and studied : perform an EnKF up to observation time, and then use the obtained ensemble as proposal density, *nudge* the model integration between times n-1 and n towards the observations at time n, perform a 4D-Var on each particle, *optimal proposal density*, …

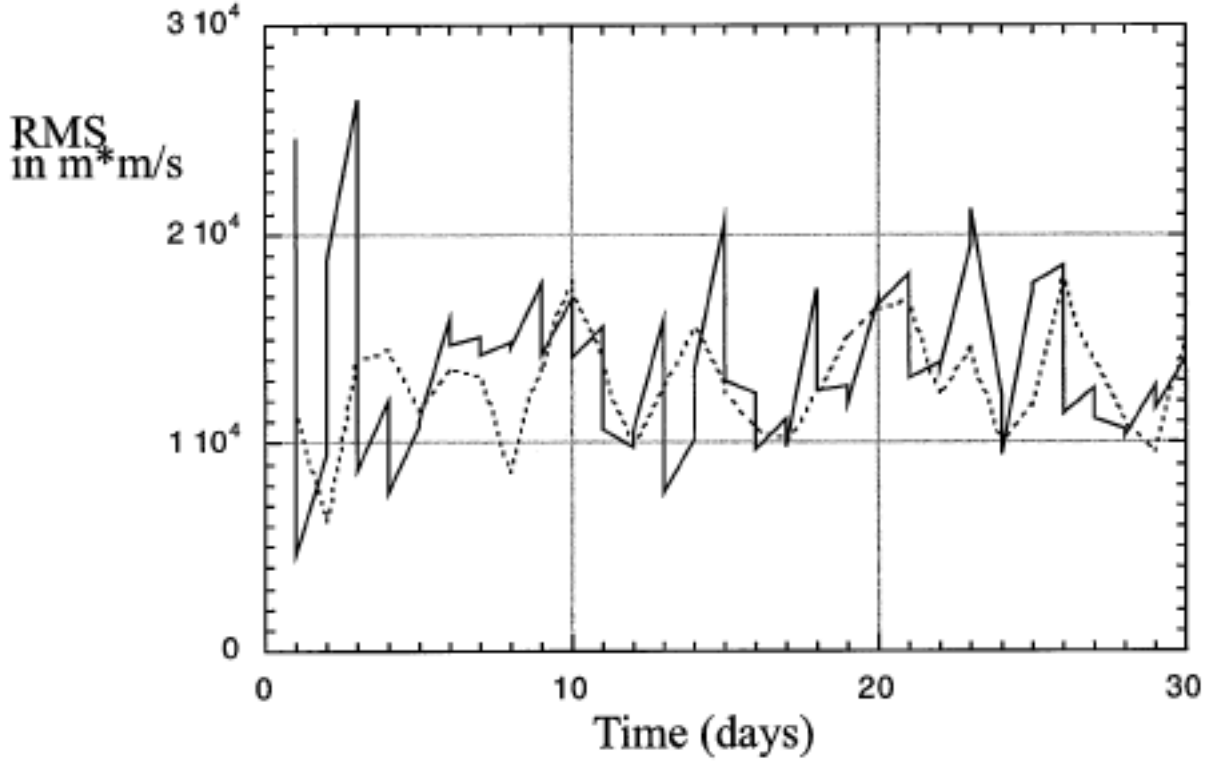


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

van Leeuwen, 2003, *Mon. Wea. Rev.*, **131**, 2071-2084

The Equivalent-Weights Particle Filter (Ades and van Leeuwen, *QJRMS*, 2013).

Make the proposal density depend on the whole ensemble at time n-1, and not only on x_i^{n-1}, use density of the form $q(x^n \mid x^{n-1}{}_{1,\mathrm{L}}, y^n)$, where 1,L denotes all ensemble indices, rather than of the more restrictive form $q(x^n \mid x_i^{n-1}, y^n)$. This gives many degrees of freedom which can be exploited for obtaining at time n an ensemble with almost equal weights.

Example Vorticity equation model with random error.

State-vector dimension $\approx 65{,}000$

Decorrelation time: 25 timesteps

One complete noisy model field observed every 50 timesteps

24 particles

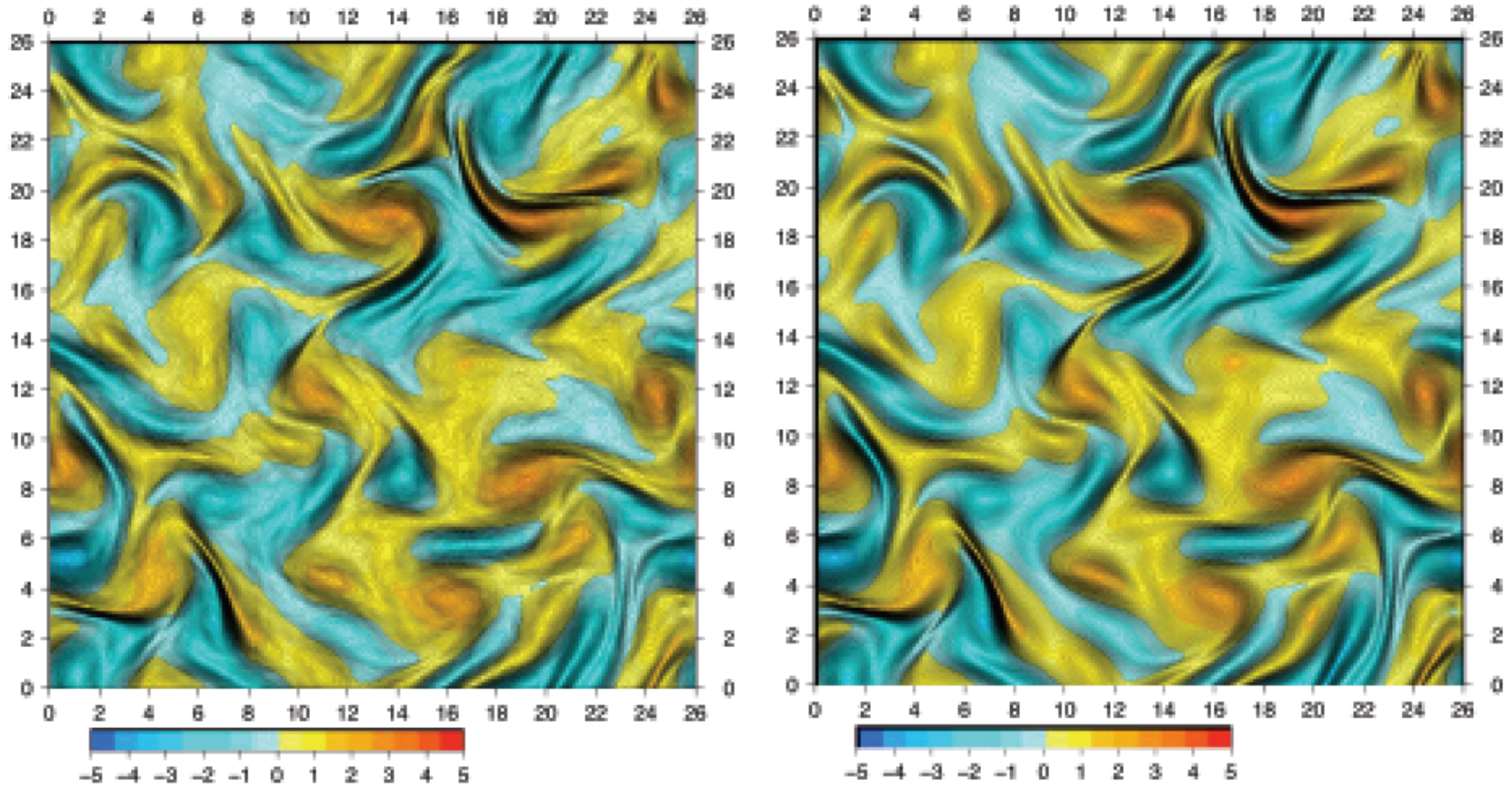


Figure 5.3. Snap shot of the vorticity field of the truth (right) and the particle filter mean (left) at time 25. Note the highly chaotic state of the fields, and the close to perfect tracking.

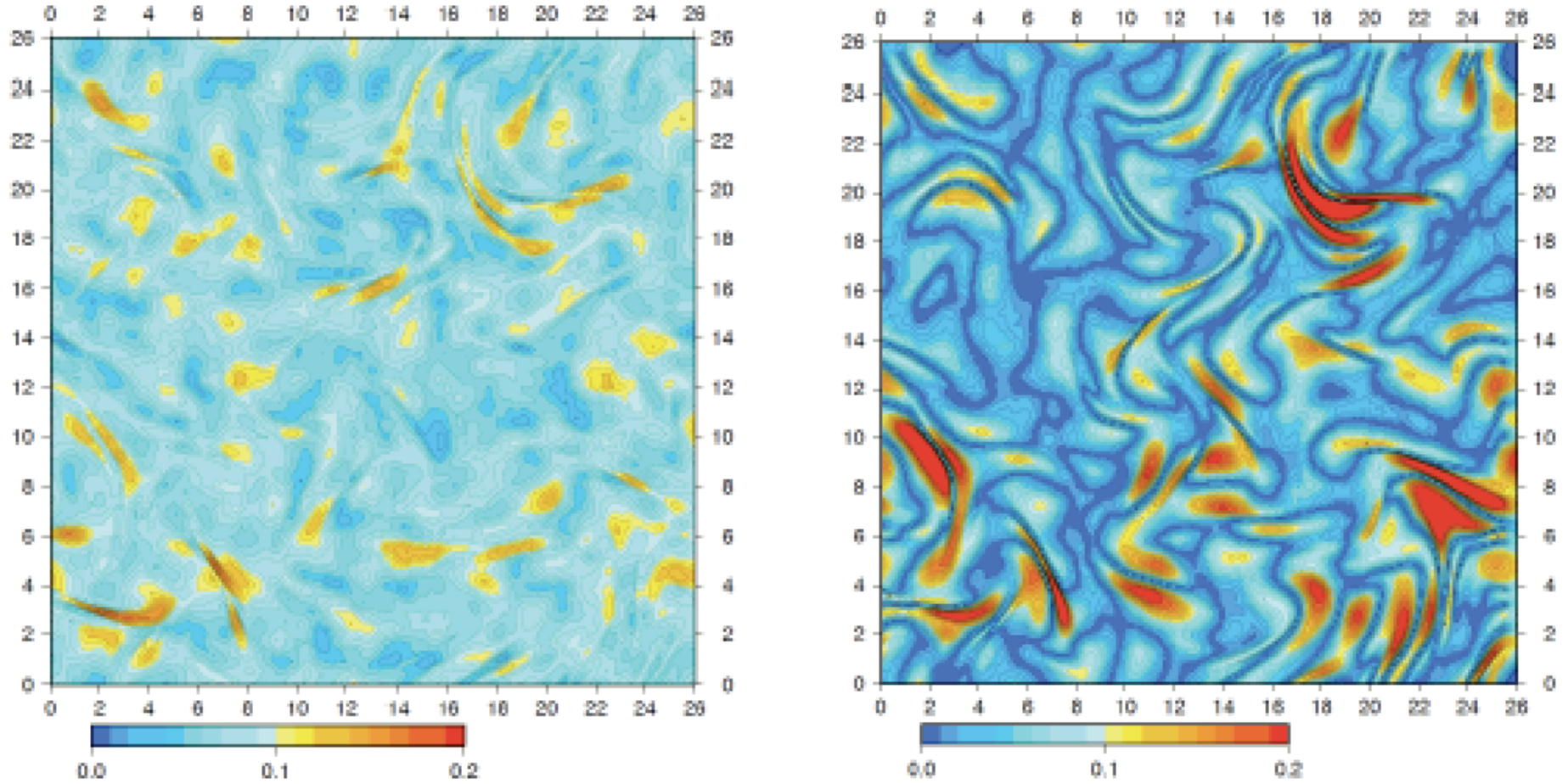


Figure 5.4. Snap shot of the absolute value of the mean-truth misfit and the standard deviation in the ensemble. The ensemble underestimates the spread at several locations, but averaged over the field it is slightly higher, 0.074 versus 0.056.

Bayesianity : experts say all these filters are bayesian (in the limit of infinite ensemble size)

Possible difficulties : numerical implementation, numerical cost

Particle filters are actively studied (van Leeuwen, Morzfeld, …)

- Validation of assimilation algorithms

Unknown $\boldsymbol{x}$ to be determined. Belongs to *state space* $\mathcal{S}$, with dimension n
Data, belonging to *data space* $\mathcal{D}$, with dimension m, available in the form

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta}$$

where $\boldsymbol{\Gamma}$ is a known $(m \times n)$-matrix, rank$\boldsymbol{\Gamma} = n$ and $\boldsymbol{\zeta}$ is 'error'

Best Linear Unbiased Estimate (BLUE)

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

with $\boldsymbol{\mu} = E(\boldsymbol{\zeta})$ and $\boldsymbol{S} = E[(\boldsymbol{\zeta} - E(\boldsymbol{\zeta}) (\boldsymbol{\zeta} - E(\boldsymbol{\zeta})^{\mathrm{T}}]$.

$E(\boldsymbol{x}^a - \boldsymbol{x}) = 0$ $\qquad$ $E[(\boldsymbol{x}^a - \boldsymbol{x}) (\boldsymbol{x}^a - \boldsymbol{x})^{\mathrm{T}}] \equiv \boldsymbol{P}^a = (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1}$

Determinacy condition : rank$\boldsymbol{\Gamma} = n$. Data contain information, directly or indirectly, on every component of state vector x. Requires $m \geq n$.

BLUE is invariant in any change of origin, or in any invertible linear transformation, in either data or state space. In particular, it is independent of the choice of a scalar product or norm in either of those spaces. *BLUE* minimizes the quadratic estimation error on any component of $\boldsymbol{x}$.

If error ζ is gaussian, $\zeta \sim \mathcal{N}[\mu, S]$, *BLUE* achieves bayesian estimation in the sense that

$$P(x \mid z) = \mathcal{N}[x^a, P^a]$$

Signification of x^a and P^a is however different. In particular, in the gaussian gase, P^a is covariance matrix of conditional probability distribution of x for *any* data set z, while it is only, in the general *BLUE* case, the covariance of the estimation error $\boldsymbol{x}^a - \boldsymbol{x}$, taken over all realizations of the error ζ.

The *BLUE* can be obtianed by minimization of the following scalar objective function, defined on state space $\mathcal{X}$

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

$$P^a = [\partial^2 \mathcal{J} / \partial \xi^2]^{-1}$$

$\mathcal{J}(\xi)$ is squared *Mahalanobis norm* of difference $\Gamma\xi - (z-\mu)$. That norm, which is associated with covariance matrix S. is defined on data space $\mathcal{D}$.

Prasanta Chandra Mahalanobis (1893 -1972)

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

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 error. The expectation is required for debiasing the data in the first place.

If determinacy condition is verified, it is always possible to decompose data vector z, through change of origin and invertible linear change of coordinates in data space, into

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

$dim x^b = n$, $dim y = p$, H known linear *observation operator.*

and $E(\zeta^b) = 0$, $E(\varepsilon) = 0$, $E(\zeta^b \varepsilon^T) = 0$
Set $E(\zeta^b \zeta^{bT}) = P^b$ (also often denoted B), $E(\varepsilon\varepsilon^T) = R$

Then

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

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

In those formulations too, determination of the *BLUE* (apparently) requires the *a priori* specification of the expectation and covariance matrix of the errors.

Questions

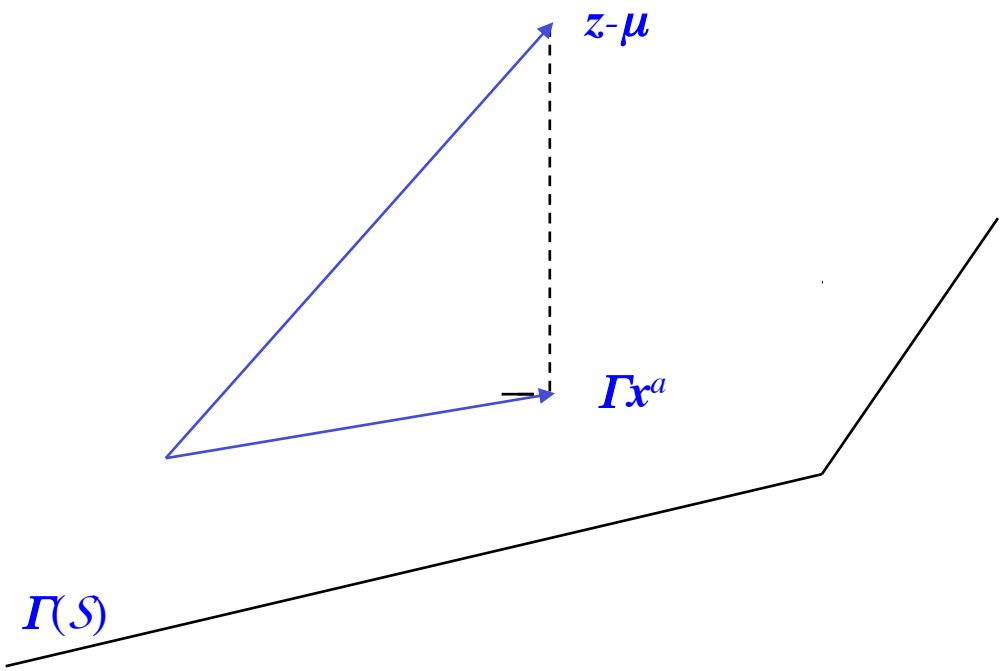
- Is it possible to objectively evaluate the quality of an assimilation system ?
- Is it possible to objectively evaluate the first- and second-order statistical moments of the data errors, whose specification is required for determining the *BLUE* ?
- Is it possible to objectively determine whether 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, and everything else being the same, is an objective measure of quality of assimilation.

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



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

- debias z
- project orthogonally onto space $\Gamma(S)$ according to Mahalanobis S-metric
- take inverse through Γ (inverse unambiguously defined through determinacy condition)

Computation of the *BLUE* is a generalized (Moore-Penrose) inverse. $(\Gamma^T S^{-1} \Gamma)^{-1} \Gamma^T S^{-1}$ is a left-inverse of Γ. Conversely, any left-inverse of Γ is of the form $(\Gamma^T \Sigma^{-1} \Gamma)^{-1} \Gamma^T \Sigma^{-1}$, with Σ a (non-uniquely defined) symmetric positive definite mxm matrix.

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

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

Set $$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$$

Then

$$\boldsymbol{x^a = \Gamma_1^{-1} [z_1 - \mu_1]}$$

$$\boldsymbol{P^a = (\Gamma_1^{\mathrm{T}} S_1^{-1} \Gamma_1)^{-1}}$$

$$x^a = \Gamma_1^{-1} [z_1 - \mu_1]$$

The probability distribution of the error

$$x^a - x = \Gamma_1^{-1} [\zeta_1 - \mu_1]$$

depends on the probability distribution of ζ_1.

On the other hand, the probability distribution of

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

depends only on the probability distribution of ζ_2.

Contrary to what equations suggest, complete specification of expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{S}$ is not necessary for determining $\boldsymbol{x}^a$ and $\boldsymbol{P}^a$. It suffices to specify the subspace $\perp\boldsymbol{\Gamma}(S)$ which is S-orthogonal to the image subspace $\boldsymbol{\Gamma}(S)$ in data space, and the respective components $\boldsymbol{\mu}_1$ and $\boldsymbol{S}_1$ of $\boldsymbol{\mu}$ and $\boldsymbol{S}$ along $\boldsymbol{\Gamma}(S)$.

Practical implications ? Actually, not many. Data space $\mathcal{D}$ varies every day with observing system, and the above decomposition varies accordingly. It is only in the case of a stationary observing system (*i.e.*, a system in which $\mathcal{D}$, Γ, $\boldsymbol{\mu}$ and $\boldsymbol{S}$ did not vary) that the above decomposition would be practically useful. Even if some components of the observing system are permanent (*e.g.*, observation operators and/or variances of associated errors), one can think it will in general be preferable to introduce those permanent components as such in a general estimation algorithm, rather than modifying the algorithm as such.

Evaluation of first- and second-order moments of error statistics ?

Systematic search among all possible $\boldsymbol{\mu}$ and $\boldsymbol{S}$, *i. e.* performing assimilations for each possible couple $(\boldsymbol{\mu}, \boldsymbol{S})$, and then evaluating results against independent observations ? Forget it.

Cross-validation (Wahba and others). For given instrument, search among possible values for error variance through validation against independent observations. Possible, may not have been sufficiently considered.

$(\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1}$ is left-inverse of $\boldsymbol{\Gamma}$ for any $\boldsymbol{S}$ $\Rightarrow$ Estimation schemes of the form

$$\boldsymbol{x}^a = (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{z} \qquad (1)$$

will not spoil exact data.

In background-observation $(\boldsymbol{x}^b, \boldsymbol{y})$ format, same property holds for schemes of the form

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{K} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b) \qquad (2)$$

where 'gain matrix' $\boldsymbol{K}$ is any nxp matrix (and holds only for those schemes).

We will consider a scheme of form (1-2), built on *a priori* assumed (but not necessarily correct) error statistics, and try and determine whether a possible misspecification of those statistics can be detected, and then corrected.

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

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

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

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.

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

Innovation alone will never be sufficient to entirely determine the required statistics, but it may impose constraints, in particular in the form of bounds, on those statistics.

A priori hypotheses made on error statistics define statistics of innovation

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

E. g., standard hypotheses and definitions

$$E(\boldsymbol{\zeta}^b) = 0, E(\boldsymbol{\varepsilon}) = 0, E(\boldsymbol{\zeta}^b \boldsymbol{\varepsilon}^{\mathrm{T}}) = 0$$
$$E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{b\mathrm{T}}) = \boldsymbol{P}^b, E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = R$$

imply

$$E(\boldsymbol{d}) = 0 \quad ; \quad E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}) = \boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}$$

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

Data-minus-Analysis (*DmA*) difference $\delta = z - \Gamma x^a$

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

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.

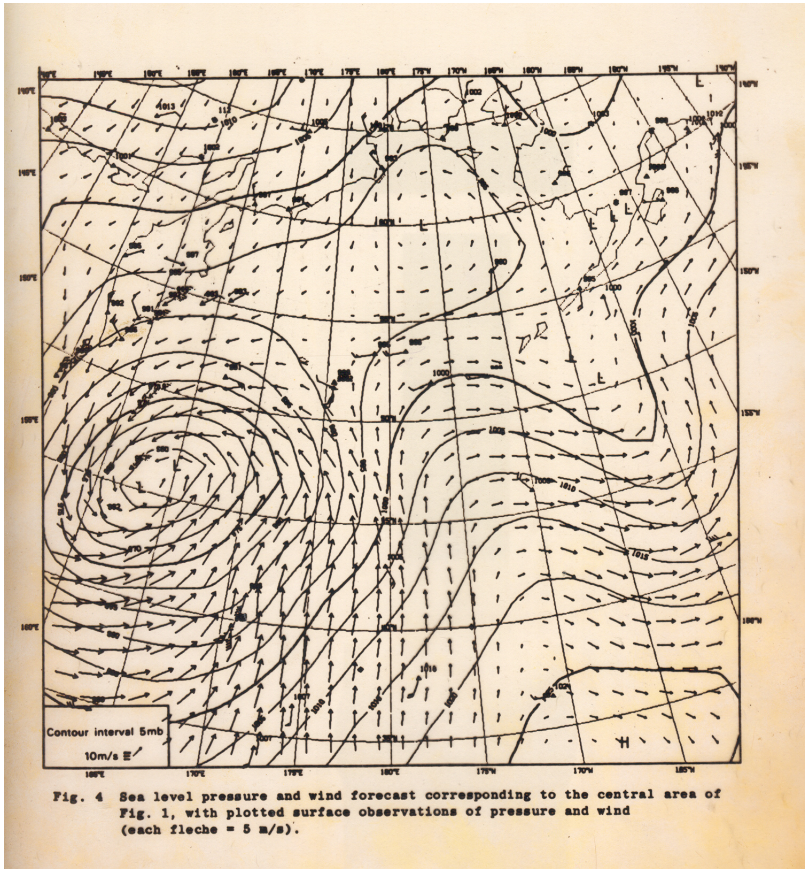


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

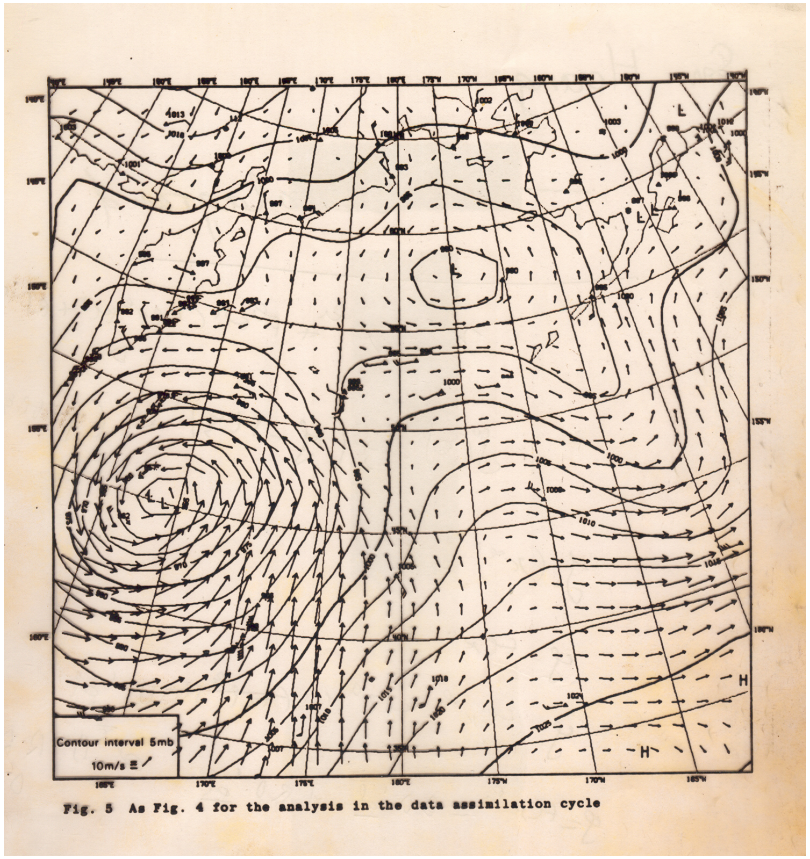


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

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).

Primary diagnostic to perform on analysis system

In z-form, DmA difference reads

$$\delta = (S - \Gamma P^a \Gamma^T)\, S^{-1}\, z$$

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

And, for a perfectly consistent system

$$E(\delta\delta^T) = S - \Gamma P^a \Gamma^T$$

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).

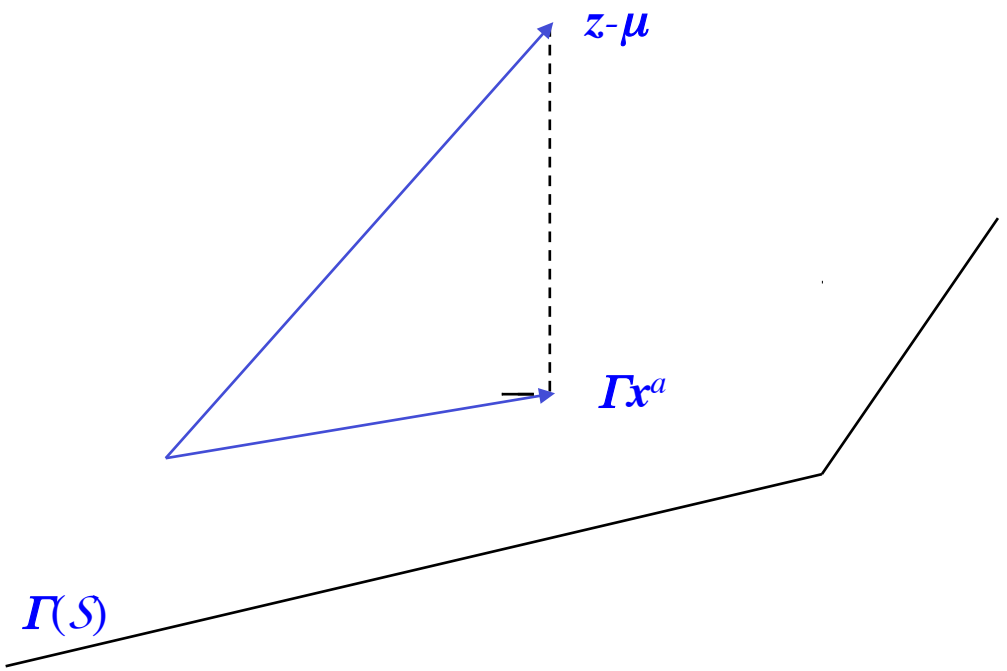
Assume inconsistency has been found between *a priori* assumed and *a posteriori* observed statistics of innovation or DmA difference.

- 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* ?

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



DmA difference, *i*. *e*. $(\boldsymbol{z}\text{-}\boldsymbol{\mu})$ - $\boldsymbol{\Gamma x}^{a}$, 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

$$E[(x^a\text{-}x)^2] = s/2$$

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

$$E(\xi_1) = -E(\xi_2) = -b/2$$

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

That 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, and to determine the associated estimation error covariance matrix. A number of degrees of freedom are therefore useless for the assimilation, and can therefore be used, in infinitely many ways, to resolve any observed inconsistency between *a priori* and *a posteriori* observed statistics of the innovation $\boldsymbol{d}$. The parameters determined by the statistics of $\boldsymbol{d}$ are equal in number, for both expectation and covariance, to those useless degrees of freedom. As a consequence, among the infinitely many possibilities for resolving the inconsistency, there is one in which neither the analysis nor its associated error covariance matrix is modified.

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.

Now, a resolution of the inconsistency that would change the orthogonal subspace $\perp\Gamma(S)$ would modify the analysis.

That result, which is purely mathematical, means that the specification of the error statistics required by the assimilation must always be entirely based, in the last resort, on external hypotheses, *i*. *e*. on hypotheses that cannot be objectively validated on the basis of the innovation alone. Given an inconsistency between *a priori* assumed and *a posteriori* observed innovation statistics, there is no mathematically fool-proof method for identifying the origin of the inconsistency.

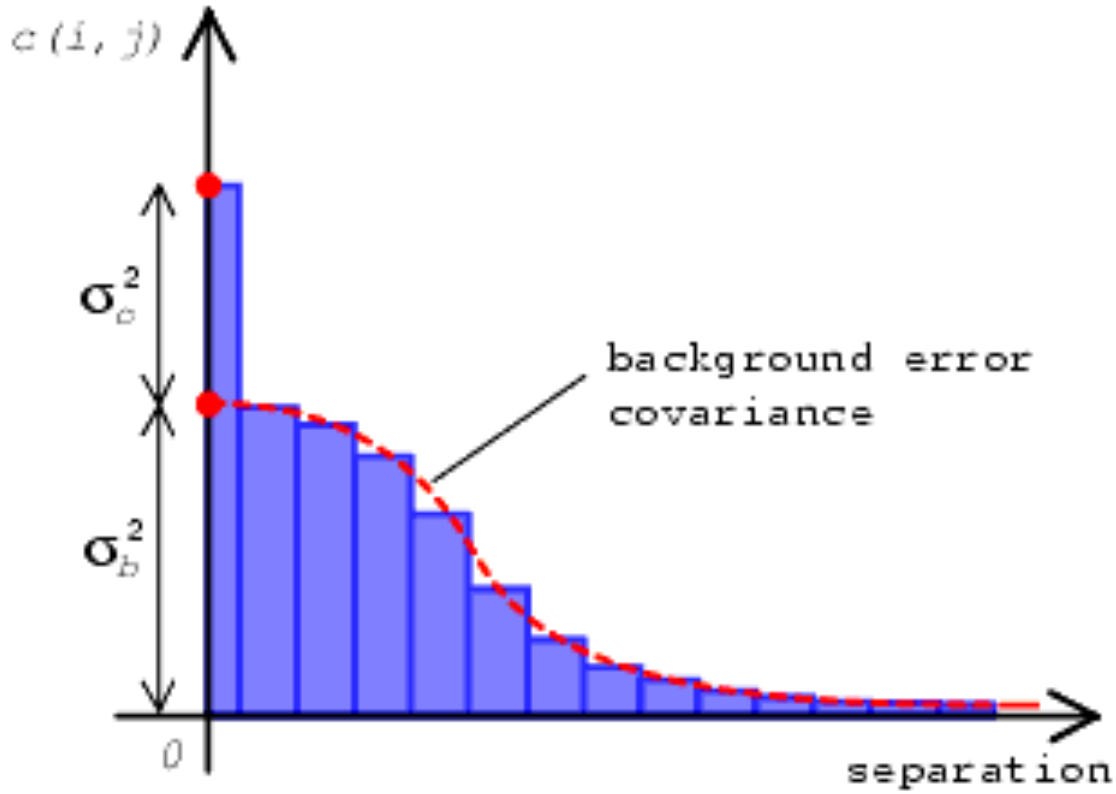
Question. Does this result hold true in a general nonlinear case ? I don't know. If anyone knows, tell me …

Problem. Identify hypotheses

- That will not be questioned (errors on observation performed a long distance apart by radiosondes made by different manufacturers are uncorrelated)
- That sound reasonable, but may be questioned (background errors are uncorrelated at scales of, say, 5000 km)
- That certainly look questionable (background and observation errors are mutually uncorrelated)
- That are undoubtedly questionable (model errors are negligible, or are uncorrelated in time)

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

Hollingsworth and Lönnberg method



(From Bouttier and Courtier, ECMWF)

Use of innovations to estimate model errors (Q)

$$B = MAM^T + Q$$

- Use ensemble assimilation to estimate « MAM^T ».
- Use innovation diagnostics to estimate « B » (or at least HBH^T).
- Estimate Q by difference :

$$Q \sim B - MAM^T$$

(e.g. Daley 1992)

After L. Berre

Model error in M.F. ensemble 4D-Var (Raynaud et al 2012, QJRMS)

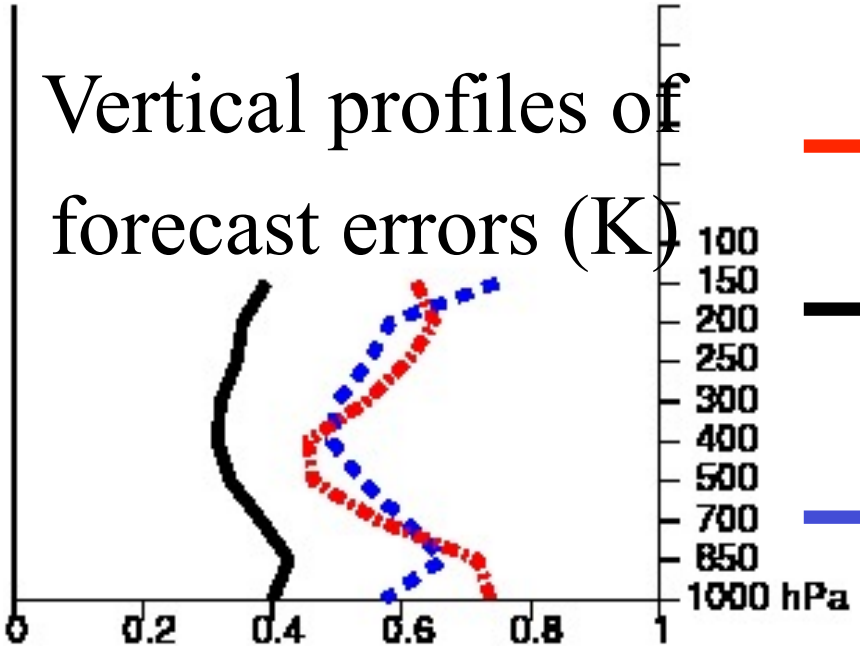
Methodology :

1. « Total » forecast error variances $V[M e^a + e^m]$ from innovations (Jb_min).

2. Compare / ensemble-based variances $V[M e^a]$ => inflation factor α.

3. Inflation of forecast perturbations (by $\alpha > 1$).

Model error in M.F. ensemble 4D-Var (Raynaud et al 2012, QJRMS)



Observation-based estimate

Ensemble-based estimate, model error neglected

Ensemble-based estimate, model error accounted for

Model error in M.F. ensemble 4D-Var

- Inflation of forecast perturbations by 15% every 6h.

- Much more realistic initial spread (by a factor 2-3)

  for ensemble prediction.

- A vertical and latitudinal dependence is needed

  w.r.t. high level tropical winds.

- Neutral impact of new variances on the forecast quality.

Objective function

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

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

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

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

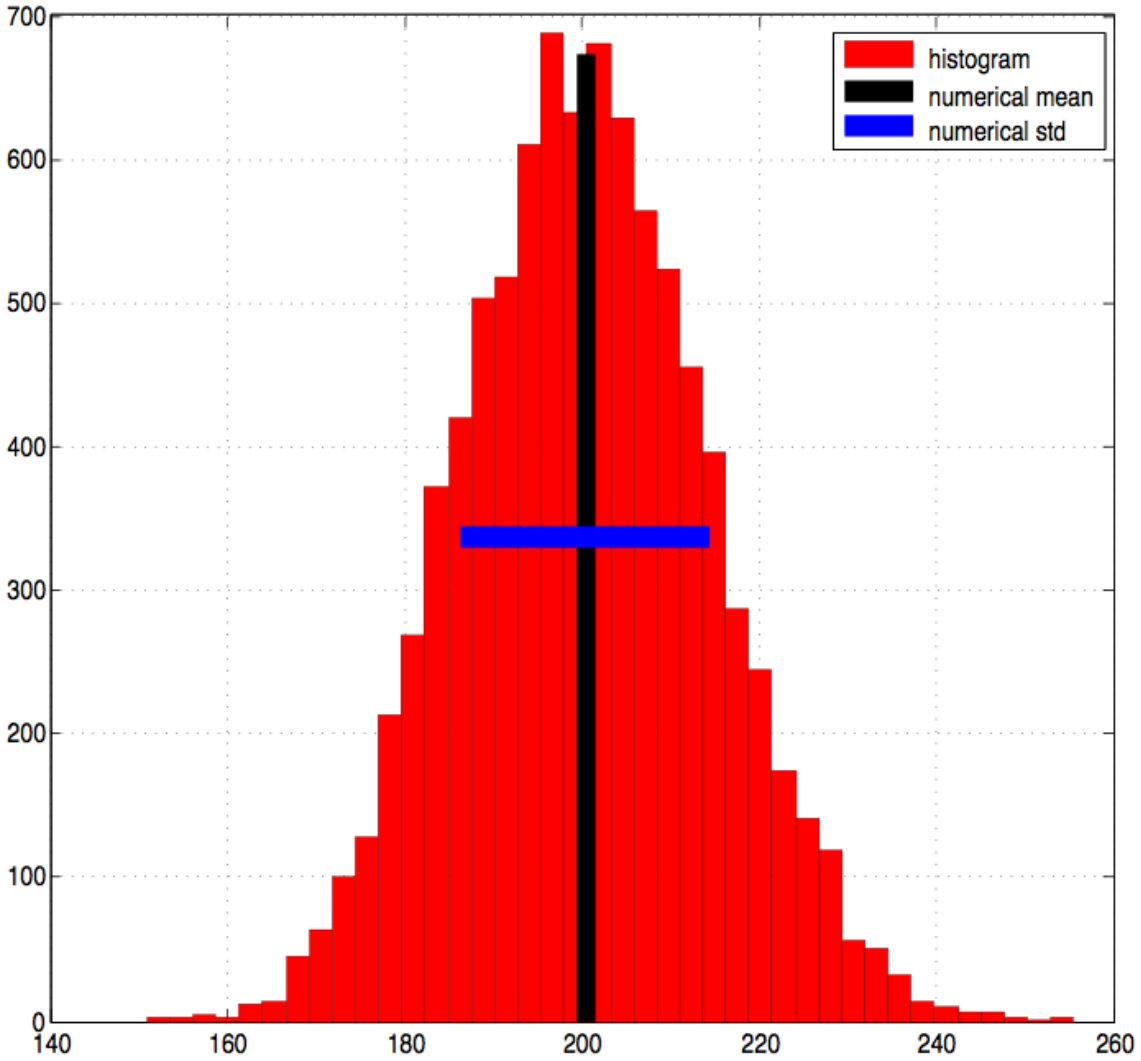
If p is large, a few realizations are sufficient for determining $E(\mathcal{J}_{min})$

If observed $E(\mathcal{J}_{min}) > p/2$, amplitude of innovation is *a priori* underestimated, and overestimated if $E(\mathcal{J}_{min}) < p/2$

Often called χ^2 criterion.

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

From course 7



Linearized Lorenz'96. 5 days. Histogram of $\mathcal{J}_{min}$
$E(\mathcal{J}_{min}) = p/2$ (=200) ; $\sigma(\mathcal{J}_{min}) = \sqrt{(p/2)}$ (≈14.14)

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

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

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

Accuracy of analysis

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

$$[\boldsymbol{P}^a]^{-1} \;=\; \Sigma_k \; \boldsymbol{H}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{H}_k$$

$$\begin{aligned} 1 &= \; (1/n)\, \Sigma_k \,\mathrm{tr}(\boldsymbol{P}^a\, \boldsymbol{H}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{H}_k) \\ &= \; (1/n)\, \Sigma_k \,\mathrm{tr}(\boldsymbol{S}_k^{-1/2}\, \boldsymbol{H}_k\, \boldsymbol{P}^a\, \boldsymbol{H}_k^{\mathrm{T}}\, \boldsymbol{S}_k^{-1/2}) \end{aligned}$$

Informative content (continuation 1)

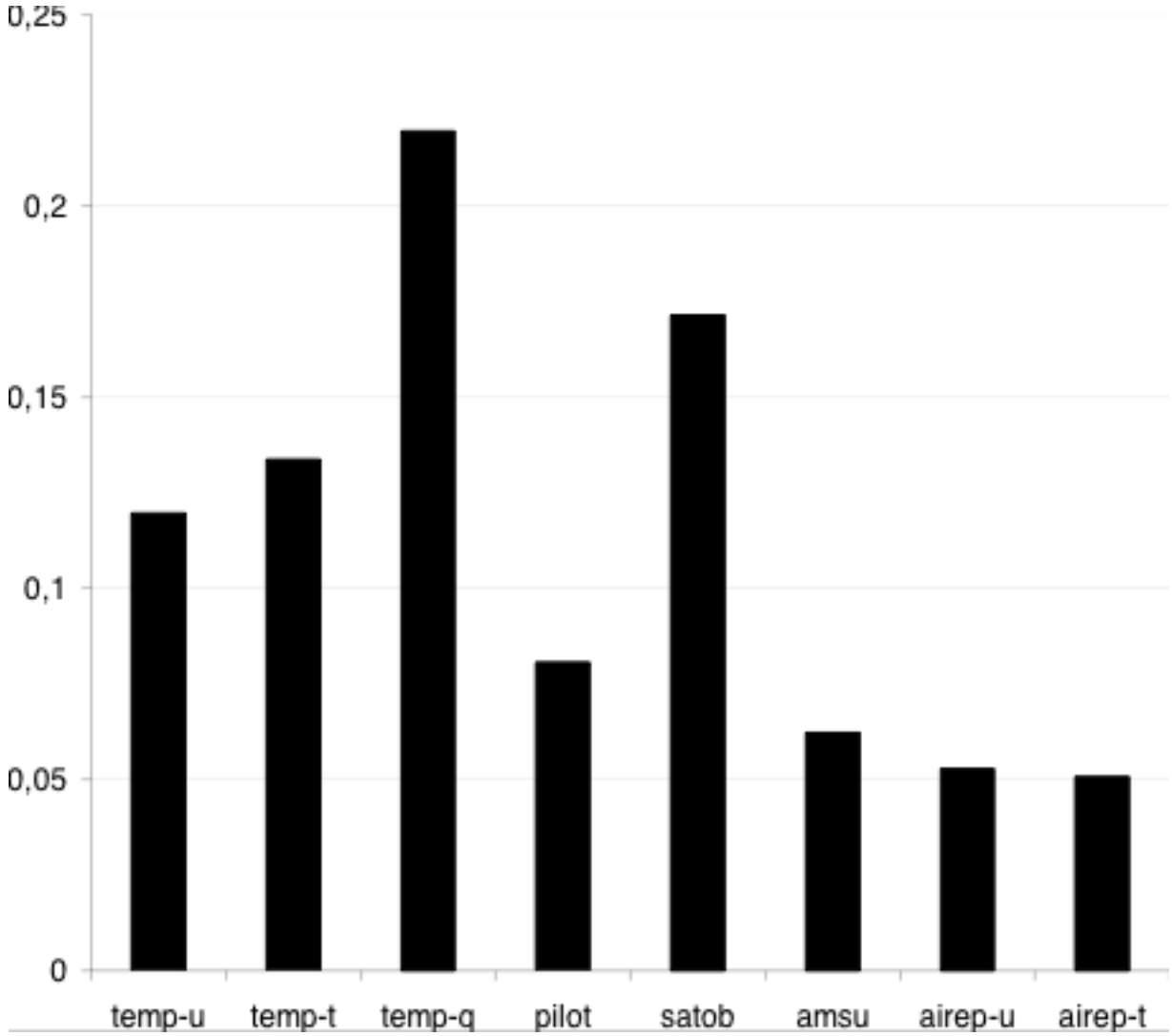
$$(1/n)\,\Sigma_k\,\mathrm{tr}(\boldsymbol{S}_k^{-1/2}\,\boldsymbol{H}_k\,\boldsymbol{P}^a\,\boldsymbol{H}_k^{\mathrm{T}}\,\boldsymbol{S}_k^{-1/2}) = 1$$

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

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

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



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

Objective function

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

where

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

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

For a perfectly consistent system

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

(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.

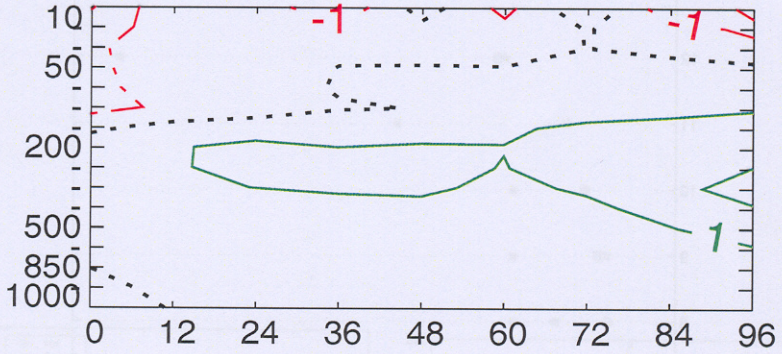
Equality

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

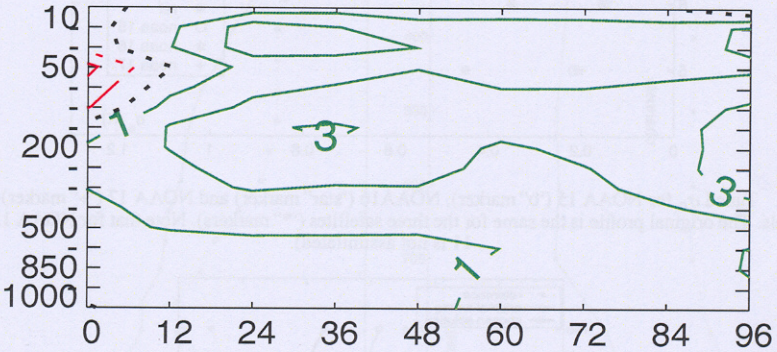
can be objectively checked (the required trace can be computed by implementing a variational assimilation on synthetic data).

Chapnik *et al*. (2004, 2005). Multiply each observation error covariance matrix $\boldsymbol{S}_k$ by a coefficient α_k such that (1) is verified simultaneously for all observation types.

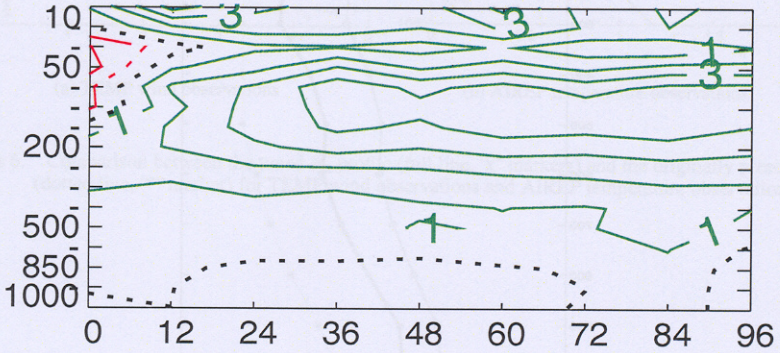
System of equations fot the α_k's solved iteratively.



(a) Northern hemisphere



(b) Southern hemisphere



(c) Subtropical area

Figure 9. Difference between tunned rms (tuned geopotential forecasts - geopotential TEMP observvations) 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 hemisphere and subpanel c for the equatorial regions.

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

Informative content (continuation 2)

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

Two subsets of data z_1 and z_2

If errors affecting z_1 and z_2 are uncorrelated, then $I(z_1 \cup z_2) = I(z_1) + I(z_2)$

If errors are correlated $I(z_1 \cup z_2) \neq I(z_1) + I(z_2)$

If $I(z_1 \cup z_2) < I(z_1) + I(z_2)$, subsets z_1 and z_2 can be said to be positively correlated, and negatively correlated if $I(z_1 \cup z_2) > I(z_1) + I(z_2)$

Informative content (continuation 3)

Example 1

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

Errors ζ_1 and ζ_2 assumed to be 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 $\boldsymbol{H}_k$ decomposes into

$$\boldsymbol{H}_k = (\boldsymbol{H}_{k1}\, \boldsymbol{H}_{k2})$$

and expression of estimation error covariance matrix into

$$[\boldsymbol{P}^a{}_1]^{-1} = \Sigma_k\, \boldsymbol{H}_{k1}{}^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{H}_{k1}$$

$$[\boldsymbol{P}^a{}_2]^{-1} = \Sigma_k\, \boldsymbol{H}_{k2}{}^{\mathrm{T}}\, \boldsymbol{S}_k^{-1}\, \boldsymbol{H}_{k2}$$

Same development as before shows that the quantity

$$(1/n_1)\ \mathrm{tr}(\boldsymbol{S}_k^{-1/2}\, \boldsymbol{H}_{k1}\, \boldsymbol{P}^a{}_1\, \boldsymbol{H}_{k1}{}^{\mathrm{T}}\, \boldsymbol{S}_k^{-1/2})$$

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)

If background and observation errors are assumed to be unbiased and mutually uncorrelated, then

$$E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}) = \boldsymbol{H}\boldsymbol{P}^{b}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}$$

If $\boldsymbol{H}\boldsymbol{P}^{b}\boldsymbol{H}^{\mathrm{T}}$ invertible, this is equivalent to

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

And, if $\boldsymbol{R}$ invertible, to

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

Optimality

Equation

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

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

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

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.

Unknown $\boldsymbol{x}$ to be determined. Belongs to *state space* $\mathcal{S}$, with dimension n
Data, belonging to *data space* $\mathcal{D}$, with dimension m, available in the form

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta}$$

where $\boldsymbol{\Gamma}$ is a known $(m \times n)$-matrix, rank$\boldsymbol{\Gamma} = n$ and $\boldsymbol{\zeta}$ is 'error'

Best Linear Unbiased Estimate (BLUE)

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

with $\boldsymbol{\mu} = E(\boldsymbol{\zeta})$ and $\boldsymbol{S} = E[(\boldsymbol{\zeta} - E(\boldsymbol{\zeta}) (\boldsymbol{\zeta} - E(\boldsymbol{\zeta})^{\mathrm{T}}]$.

$E(\boldsymbol{x}^a - \boldsymbol{x}) = 0$ $\qquad$ $E[(\boldsymbol{x}^a - \boldsymbol{x}) (\boldsymbol{x}^a - \boldsymbol{x})^{\mathrm{T}}] \equiv \boldsymbol{P}^a = (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1}$

Determinacy condition : rank$\boldsymbol{\Gamma} = n$. Data contain information, directly or indirectly, on every component of state vector x. Requires $m \geq n$.

BLUE is invariant in any change of origin, or in any invertible linear transformation, in either data or state space. In particular, it is independent of the choice of a scalar product or norm in either of those spaces. *BLUE* minimizes the quadratic estimation error on any component of $\boldsymbol{x}$.

If error ζ is gaussian, $\zeta \sim \mathcal{N}[\boldsymbol{\mu}, S]$, *BLUE* achieves bayesian estimation in the sense that

$$P(x \mid z) = \mathcal{N}[x^a, P^a]$$

Any assumed probability distribution $P(\zeta)$ defines a conditional probability distribution $P(x \mid z)$ for x. In case the distribution $P(\zeta)$ is known only through its expectation $\boldsymbol{\mu}$ and covariance matrix S, the gaussian distribution $\mathcal{N}[\boldsymbol{\mu}, S]$ leads for $\boldsymbol{x}$ to the conditional probability distribution $P(x \mid z)$ with the largest entropy. The gaussian choice is in that sense the 'least-committing' choice.

BLUE is the simplest of non-simplicist algorithms.

The *BLUE* can be obtained by minimization of the following scalar objective function, defined on state space $\mathcal{X}$

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

And in case of nonlinearity ?

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

Variational approach can be heuristically implemented

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

It works !

If data are of the form (after possibly an appropriate transformation)

$$x^b = x + \zeta^b$$
$$y = H(x) + \varepsilon$$

Transformation

$$x^b = x + \zeta^b$$
$$y - H(x^b) = H(x) - H(x^b) + \varepsilon \approx H'(x - x^b) + \varepsilon$$

where H' is jacobian of H, makes the estimation problem linear in the deviation $x - x^b$
(*tangent linear approximation*)

All algorithms that have been presented in the course, with the exception of particle filters, are empirical heuristic extensions of the BLUE approach to approximate nonlinear and non-gaussian situations.

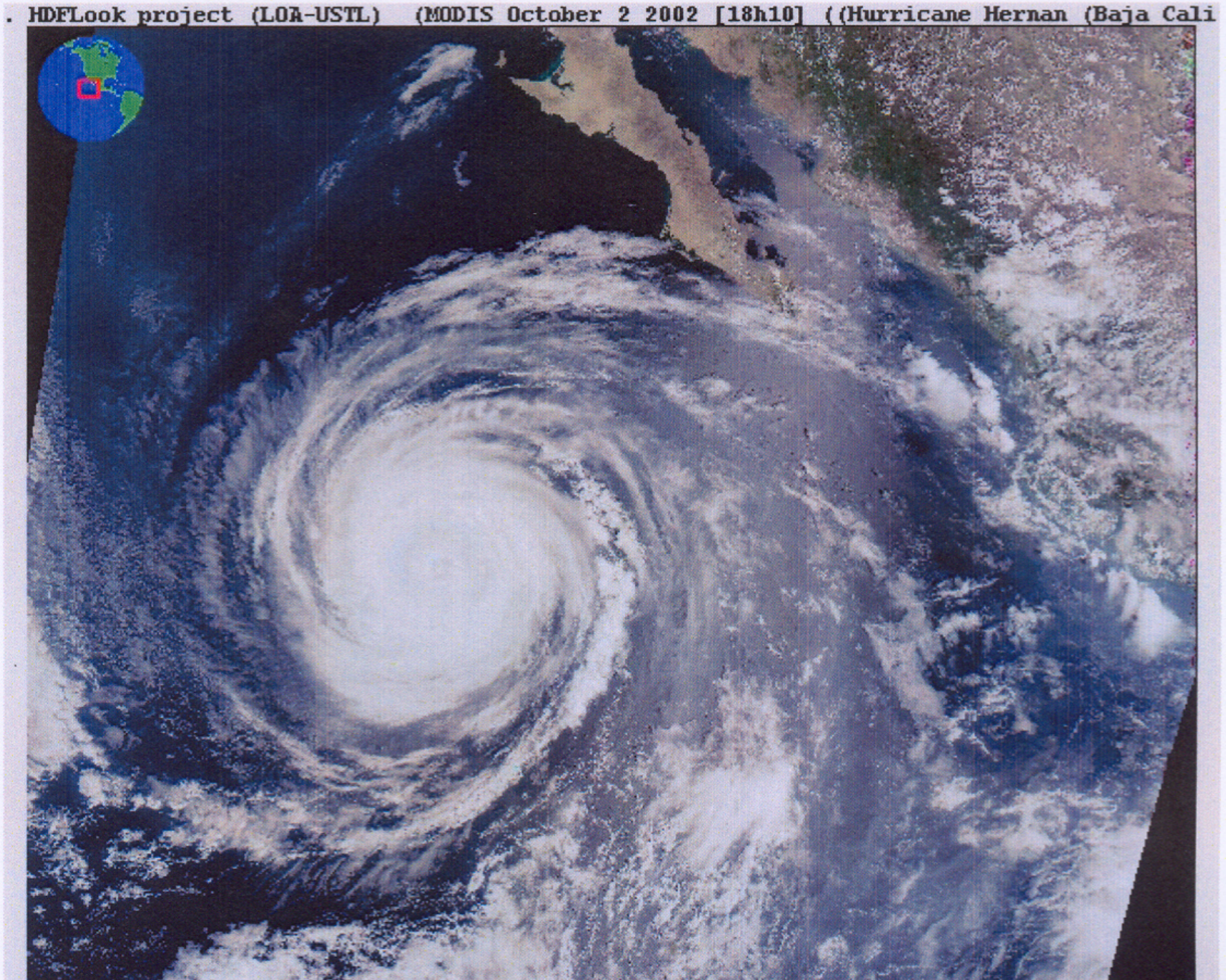
Assimilation, which originated from the need of defining initial conditions for numerical weather forecasts, has gradually extended to many diverse applications

- Oceanography
- Palaoclimatology
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciology
- Magnetism (both planetary and stellar)
- Plate tectonics
- Planetary atmospheres (Mars, …)
- Reassimilation of past observations (mostly for climatological purposes, ECMWF, NCEP/NCAR)
- Identification of source of tracers
- Parameter identification
- *A priori* evaluation of anticipated new instruments
- Definition of observing systems *(Observing Systems Simulation Experiments)*
- Validation of models
- Sensitivity studies (adjoints)
- Mathematical studies, independently of direct real life applications
- …

It has now become a major tool of numerical environmental science, and a subject of mathematical study in its own right.

A few of the (many) remaining problems :

- Observability (what to observe in order to know what we want to know ? Data are noisy, system is chaotic !)
- More accurate identification and quantification of errors affecting data particularly the assimilating model (will always require independent hypotheses)
- Assimilation of images
- …


. HDFLook project (LOA-USTL) (MODIS October 2 2002 [18h10] ((Hurricane Hernan (Baja Cali

La Fin du Cours …