

Some Future Perspectives for Assimilation

Olivier Talagrand
Laboratoire de Météorologie Dynamique, École Normale Supérieure
Paris, France

Workshop *Mathematical Advancement in Geophysical Data Assimilation*
Banff International Research Station
for Mathematical Innovation and Discovery
Banff, Canada
7 February 2008

Purpose of assimilation : reconstruct as accurately as possible the state of the atmosphere (the ocean, or whatever the system of interest is), using all available appropriate information. The latter essentially consists of

- The observations.
- The physical laws governing the system, available in practice in the form of a discretized, and necessarily approximate, numerical model.
- 'Asymptotic' properties of the flow, such as, e. g., geostrophic balance of middle latitudes. Although they basically are necessary consequences of the physical laws which govern the flow, these properties can usefully be explicitly introduced in the assimilation process.

Both observations and 'model' are affected with some uncertainty \Rightarrow uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don't know too well why, but it works).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (unambiguously defined if a prior probability distribution is defined; see Tarantola, 2005).

Ensemble Assimilation : the final product consists of a finite ensemble of points in state space, whose distribution is meant to sample the looked-for conditional probability distribution.

Ensemble Assimilation exists at present in two forms

- **Ensemble Kalman Filter (EnKF)**. Still linear and Gaussian as concerns updating phase.

- **Particle filters**. Dimension !

Ensemble elements may be 'equal and independent' (EnKF) or have (time-varying) weights w_i (particle filters)

Another approach for updating ensemble: 'acceptance-rejection' for generating sample of equal elements of posterior distribution (Miller *et al.*, 1999, *Tellus*).

(in Ensemble Prediction, there usually is a high-resolution 'control forecast', and a number of lower-resolution ensemble forecasts).

High cost, in particular for non-gaussian filters. Is the cost intrinsic to the problem, or could it be significantly reduced by new algorithmic developments ?

Evaluation of assimilation ensembles

Ensembles must be evaluated as descriptors of probability distributions (and not for instance on the basis of properties of individual elements). This implies, among others

- Validation of the expectation of the ensembles
- Validation of the spread (*spread-skill relationship*)

Reduced Centred Random Variable (RCRV, Candille *et al.*, 2006)

For some scalar variable x , ensemble has mean μ and standard deviation σ . Ratio

$$s = \frac{\xi - \mu}{\sigma}$$

where ξ is verifying observation. Over a large number of realizations

$$E(s) = 0 \quad , \quad E(s^2) = 1$$

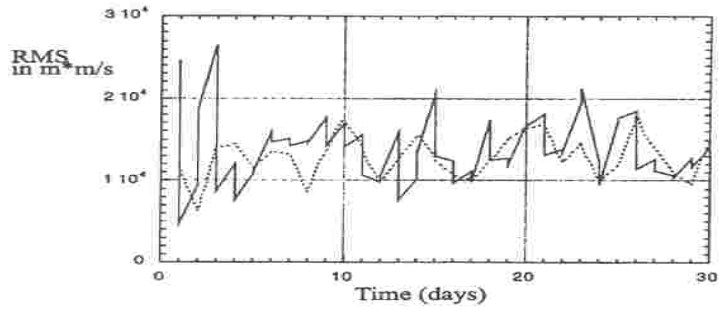


FIG. 12. Comparison of rms error ($m^2 s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

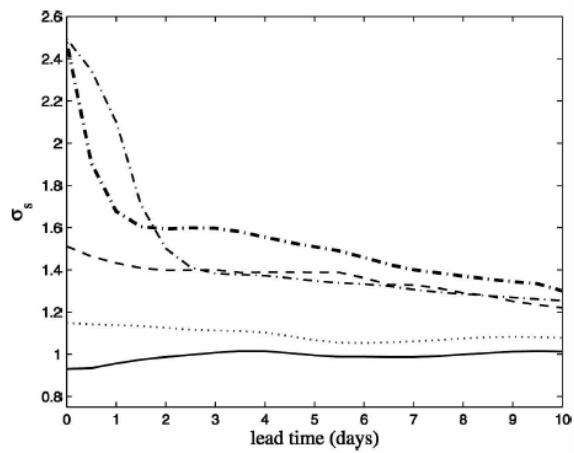


FIG. 4. Evolution of the std dev of the RCRV, as a function of lead time, for the four different methods: EnKF (solid line), ETKF (dotted line), BM (dashed line), and SV computed over a 24-h optimization period (heavy dashed-dotted line) and SV computed over a 48-h optimization period (thin dashed-dotted line).

Descamps and Talagrand, *Mon. Wea. Rev.*, 2007

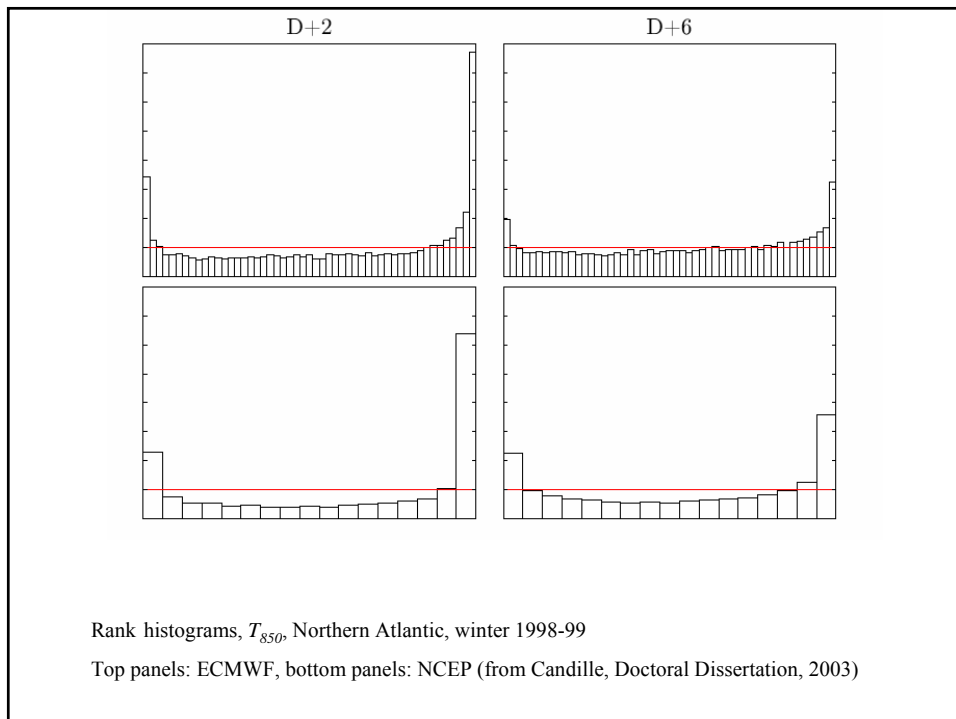
Rank Histograms

For some scalar variable x , N ensemble values, assumed to be N independent realizations of the same probability distribution, ranked in increasing order

$$x_1 < x_2 < \dots < x_N$$

Define $N+1$ intervals.

If verifying observation ξ is an $N+1$ st independent realization of the same probability distribution, it must be statistically undistinguishable from the x_i 's. In particular, must be uniformly distributed among the $N+1$ intervals defined by the x_i 's.



Two properties make the value of an ensemble estimation system (either for assimilation or for prediction)

Reliability is statistical consistency between estimated probability distributions and verifying observations. Is objectively and quantitatively measured by a number of standard diagnostics (among which Reduced Centred Random Variable and Rank Histograms, reliability component of Brier and Brier-like scores).

Resolution (semantic disagreement) is the property that reliably predicted probability distributions are useful (essentially have small spread). Also measured by a number of standard diagnostics (resolution component of Brier and Brier-like scores).

To-day's message. Evaluate assimilation ensembles in terms of reliability and resolution.

Time-correlated Errors

Example of time-correlated observation errors

$$z_1 = x + \zeta_1$$

$$z_2 = x + \zeta_2$$

$$E(\zeta_1) = E(\zeta_2) = 0 \quad ; \quad E(\zeta_1^2) = E(\zeta_2^2) = s \quad ; \quad E(\zeta_1 \zeta_2) = 0$$

BLUE of x from z_1 and z_2 gives equal weights to z_1 and z_2 .

Additional observation then becomes available

$$z_3 = x + \zeta_3$$

$$E(\zeta_3) = 0 \quad ; \quad E(\zeta_3^2) = s \quad ; \quad E(\zeta_1 \zeta_3) = cs \quad ; \quad E(\zeta_2 \zeta_3) = 0$$

BLUE of x from (z_1, z_2, z_3) has weights in the proportion $(1, 1+c, 1)$

Time-correlated Errors (continuation 1)

Example of time-correlated model errors

Evolution equation

$$x_{k+1} = x_k + \eta_k \quad E(\eta_k^2) = q$$

Observations

$$y_k = x_k + \varepsilon_k, \quad k = 0, 1, 2 \quad E(\varepsilon_k^2) = r, \text{ errors uncorrelated in time}$$

Sequential assimilation. Weights given to y_0 and y_1 in analysis at time 1 are in the ratio $r/(r+q)$. That ratio will be conserved in sequential assimilation. All right if model errors are uncorrelated in time.

Assume $E(\eta_0 \eta_1) = cq$

Weights given to y_0 and y_1 in estimation of x_2 are in the ratio

$$\rho = \frac{r - qc}{r + q + qc}$$

Time-correlated Errors (continuation 2)

Moral. If data errors are correlated in time, it is not possible to discard observations as they are used while preserving optimality of the estimation process. In particular, if model error is correlated in time, all observations are liable to be reweighted as assimilation proceeds.

Variational assimilation can take time-correlated errors into account.

Example of time-correlated observation errors. Global covariance matrix

$$\mathcal{R} = (R_{kk'} = E(\varepsilon_k \varepsilon_{k'}^T))$$

Objective function

$$\xi_0 \in \mathcal{S} \rightarrow$$

$$J(\xi_0) = (1/2) (x_0^b - \xi_0)^T [P_0^b]^{-1} (x_0^b - \xi_0) + (1/2) \sum_{kk'} [y_k - H_k \xi_k]^T [\mathcal{R}^{-1}]_{kk'} [y_{k'} - H_{k'} \xi_{k'}]$$

where $[\mathcal{R}^{-1}]_{kk'}$ is the kk' -sub-block of global inverse matrix \mathcal{R}^{-1} .

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 3)

Time correlation of observational error has been introduced by ECMWF (Järvinen *et al.*, 1999) in variational assimilation of high-frequency surface pressure observations (correlation originates in that case in representativeness error).

Identification and quantification of temporal correlation of errors, especially model errors ?

Q. Is it possible to develop fully bayesian algorithms for systems with dimensions encountered in meteorology and oceanography ? Would that require totally new algorithmic developments ?

Q. Is it possible to have at the same time the advantages of both ensemble estimation and variational assimilation (propagation of information both forward and backward in time, and, more importantly, possibility to take temporal dependence into account) ?

Observability

What must one observe to know what ?

Dynamical 'downscaling'

Q. Is it possible to determine the small scales of the motion from the observed history of the large scales ?

Least-variance linear estimation, on which a large fraction of assimilation algorithms are still based, determines the *Best Linear Unbiased Estimate (BLUE)* of the state of the system from the available data. It achieves bayesian estimation if the errors affecting the data are globally gaussian.

It requires the *a priori* knowledge of the first- and second-order statistical moments of the errors affecting the data.

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 assimilation, 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.

$$\begin{aligned}x^b &= x + \zeta^b \\y &= Hx + \varepsilon\end{aligned}$$

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.

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

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

With hypotheses made above

$$E(d) = 0 \quad ; \quad E(dd^T) = HP^b H^T + R$$

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

Consider assimilation scheme of the form

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

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

(1) \Leftrightarrow if data are perfect, then so is the estimate \mathbf{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 .

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

$$E(d) = 0 \quad (\Leftrightarrow \quad E(\delta) = 0)$$

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).

$$E[(x^b - x^a)(x^b - x^a)^T] = P^b - P^a$$

$$E[(y - Hx^a)(y - Hx^a)^T] = R - HP^aH^T$$

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 $(m \times n)$ -matrix, and ζ an unknown 'error'

For instance

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

which corresponds to

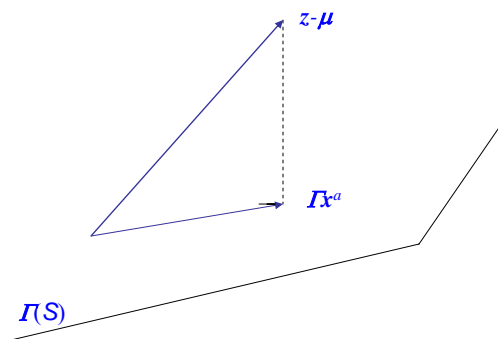
$$\Gamma = \begin{pmatrix} I_n \\ H \end{pmatrix} \quad \zeta = \begin{pmatrix} \zeta^b \\ \varepsilon \end{pmatrix}$$

Variational form.

x^a minimizes following scalar *objective function*, defined on state space S

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

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



DmA difference, *i. e.* $(z - \mu) - T\bar{x}^a$, 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

$$\begin{aligned} z_1 &= x + \zeta_1 \\ z_2 &= x + \zeta_2 \end{aligned}$$

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 μ and S onto the subspace $I(S)$. As for the subspace that is S -orthogonal to $I(S)$, it suffices to know what it is, but it is not necessary to know the projection of μ and S onto it. A number of degrees of freedom are therefore useless for the assimilation. The parameters determined by the statistics of 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 performed 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.

Informative content

Objective function

$$J(\xi) = \sum_k J_k(\xi)$$

where

$$J_k(\xi) \equiv (1/2) (H_k \xi - y_k)^T S_k^{-1} (H_k \xi - y_k)$$

with $\dim y_k = m_k$

Accuracy of analysis

$$P^a = (I^T S^{-1} I)^{-1}$$

$$[P^a]^{-1} = \sum_k H_k^T S_k^{-1} H_k$$

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

Informative content (continuation 1)

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

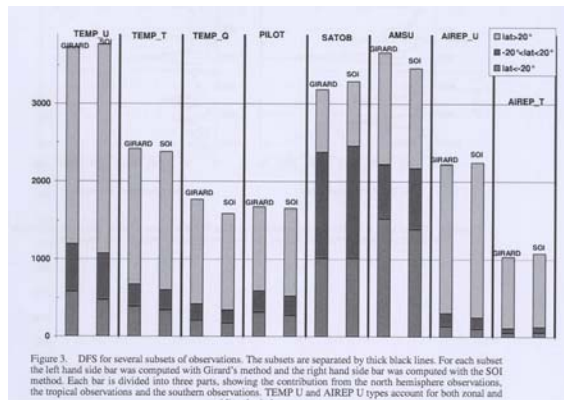
$I(y_k) \equiv (1/n) \text{tr}(S_k^{-1/2} H_k P^a H_k^T S_k^{-1/2})$ is a measure of the relative contribution of subset of data y_k to overall accuracy of assimilation. Invariant in linear change of coordinates in data space \Rightarrow valid for *any* subset of data.

In particular

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

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

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



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

Chapnik *et al.*, 2006, *QJRM*S, **132**, 543-565

QuickTime™ et un
décompresseur TIFF (LZW)
sont requis pour visionner cette image.

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

Objective function

$$J(\xi) = \sum_k J_k(\xi)$$

where

$$J_k(\xi) \equiv (1/2) (H_k \xi - y_k)^T S_k^{-1} (H_k \xi - y_k)$$

with $\dim y_k = m_k$

For a perfectly consistent system

$$E[J_k(x^a)] = (1/2) [m_k - \text{tr}(S_k^{-1/2} H_k P^a H_k^T S_k^{-1/2})]$$

(in particular, $E(J_{min}) = p/2$)

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

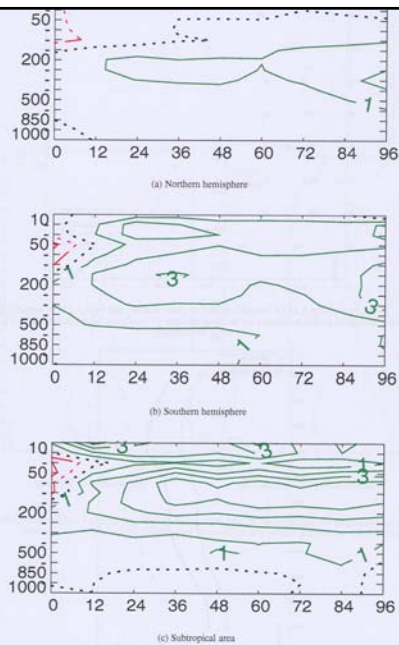
Equality

$$E[J_k(x^a)] = (1/2) [m_k - \text{tr}(S_k^{-1/2} H_k P^a H_k^T S_k^{-1/2})] \quad (1)$$

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 for the α_k 's solved iteratively.



Chapnik *et al.*, 2006,
*QJRM*S, **132**,
 543-565

Figure 9. Difference between tuned rms (tuned geopotential forecasts - geopotential TEMP observations) and the operational rms computed over 21 situations. 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

Informative content (continuation 2)

$$I(y_k) \equiv (1/n) \text{tr}(S_k^{-1/2} H_k P^\alpha H_k^T S_k^{-1/2})$$

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 $I(z_1 \cup z_2) \neq I(z_1) + I(z_2)$

Informative content (continuation 3)

Example 1

$$\begin{aligned} z_1 &= x + \zeta_1 \\ z_2 &= x + \zeta_2 \end{aligned}$$

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 x evolving in time according to

$$x_2 = \alpha x_1$$

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 u_1 of analyzed fields, $dim u_1 = n_1$. Define relative contribution of subset y_k of data to accuracy of u_1 ?

u_2 : component of x orthogonal to u_1 with respect to Mahalanobis norm associated with P^a (analysis errors on u_1 and u_2 are uncorrelated).

$x = (u_1^T, u_2^T)^T$. In basis (u_1, 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

$$[P^a_1]^{-1} = \sum_k H_{k1}^T S_k^{-1} H_{k1}$$

$$[P^a_2]^{-1} = \sum_k H_{k2}^T S_k^{-1} H_{k2}$$

Same development as before shows that the quantity

$$(1/n_1) \text{tr}(S_k^{-1/2} H_{k1} P^a_1 H_{k1}^T S_k^{-1/2})$$

is a measure of the relative contribution of subset y_k of data to analysis of subset u_1 of state vector.

But can it be computed in practice for large dimension systems (requires the explicit decomposition $x = (u_1^T, u_2^T)^T$)?

Informative content (continuation 6)

Q. Can those notions be extended to general nonlinear case ?

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

For a consistent system

$$E[\mathbf{H}(\mathbf{x}^a - \mathbf{x}^b)(\mathbf{y} - \mathbf{H}\mathbf{x}^b)^\top] = E[\mathbf{H}(\mathbf{x}^a - \mathbf{x}^b)\mathbf{d}^\top] = \mathbf{H}\mathbf{P}^b\mathbf{H}^\top$$

$$E[(\mathbf{y} - \mathbf{H}\mathbf{x}^a)(\mathbf{y} - \mathbf{H}\mathbf{x}^b)^\top] = E[(\mathbf{y} - \mathbf{H}\mathbf{x}^a)\mathbf{d}^\top] = \mathbf{R}$$