SOME ASPECTS OF VALIDATION OF ASSIMILATION ALGORITHMS

Olivier TALAGRAND

Bernard CHAPNIK

Carlos PIRES

- Validation

  (How to define

  pdf of data error?)

- Optimality

- Gaussianity of error?

- Quasi-static

  Variational Assimilation

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

Data vector z, belonging to *data space* $\mathcal{D}$ ($\dim\mathcal{D} = m$), available.

$$F(z, x, \zeta) = 0 \qquad (1)$$

where ζ is a random element representing the uncertainty on the data (or, more precisely, on the link between the data and the unknown state vector).

For example

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

Bayesian estimation

Probability that $x = \xi$ for given ξ ?

$$x = \xi \quad \Leftrightarrow \quad F(z, \xi, \zeta) = 0$$

$$P(x = \xi \mid z) = P[F(z, \xi, \zeta) = 0] \,/\, \int_{\xi'} P[F(z, \xi', \zeta) = 0]$$

Unambiguously defined iff, for any ζ, there is at most one x such that (1) is verified.

$\Leftrightarrow$ data contain information, either directly or indirectly, on any component of x. *Determinacy* condition.

Bayesian estimation is however impossible in its general theoretical form in meteorological or oceanographical practice because of

- very large numerical dimension of state vector to be estimated.

- probability distribution of errors on data very poorly known (model errors in particular).

In practice one determines ≈ expectation + (a small number of components of) covariance matrix of conditional probability distribution, or (≈ equivalently) an ensemble of states (size a few hundred at most) meant to be a sample of conditional probability distribution

Data can contain 'equations'
(e.g. evolution equations)

$$\frac{x_{t+1} - x_t}{\Delta t} - F(x_t) = 0 + \zeta$$

ζ ↓ z

Least-variance unbiased linear estimation

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

Γ known (mxn)-matrix, ζ unknown 'error'

Look for estimated state vector x^a of the form

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

subject to

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

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

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

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

where $\mu = E(\zeta)$ (expectation) and $S = E\{[\zeta - \mu][\zeta - \mu]^T\}$ (covariance matrix)

Best Linear Unbiased Estimator (*BLUE*) of x from z.

Requires (at least apparently) *a priori* explicit knowledge of first- and second-order statistical moments of error ζ.

Determinacy condition $\Leftrightarrow$ rank$\Gamma = n$. $m = n + p \quad p \geq 0$

In case ζ is gaussian, $\zeta = \mathcal{N}[\mu, S]$, BLUE achieves bayesian estimation in the sense that

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

Variational form.

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

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

BLUE is invariant in any invertible linear change of coordinates, either in state or data space.

From now on, unless specified otherwise, data assumed to be unbiased ($\mu = 0$).

If determinacy condition is verified, it is always possible to decompose data into

A '*background*' estimate (*e. g.* forecast from the past), 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 $m - n = p$

$$y = Hx + \varepsilon$$

Then

$$x^a = x^b + P^b H^T (HP^b H^T + R)^{-1} (y - Hx^b)$$

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

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

$d \equiv y - Hx^b$ is called the *innovation vector*

Variational form

Objective function reads

$$\xi \rightarrow$$
$$\mathcal{J}(\xi) \equiv (1/2)(\xi - x^b)^T [P^b]^{-1} (\xi - x^b) + (1/2)(H\xi - y)^T R^{-1} (H\xi - y)$$

Two approaches

1. Least-variance unbiased linear estimation

- Optimal interpolation
- Kalman filter
- Kalman smoother
- Extended Kalman filter
- 3D Variational analysis (in both its primal and dual algorithmic implementations)
- 4D Variational assimilation (in both its strong- and weak-constraint formulations, and in both primal and dual algorithmic implementations).

Amounts in practice to determining ≈ expectation + (a small number of components of) covariance matrix of conditional probability distribution.

Valid in nonlinear cases if so-called tangent linear approximation is valid.

Difficulty : describe temporal evolution of uncertainty on the flow (determine the matrix '***B***').

2. Ensemble assimilation

Conditional probability distribution described by a finite sample of points in state space (size ≈ a few 10^2).

That approach avoids in principle any need for a linear hypothesis. Most algorithms are still however partially linear.

Determination of the *BLUE* requires (at least apparently) the *a priori* specification of the first- and second-order statistical moments of the errors affecting the data (general bayesian estimation requires the specification of the entire probability distribution of the errors).

Questions

- Is it possible to objectively evaluate those first- and second-order statistical moments of the data errors ?

- How sensitive is the quality of the analysis to misspecification of those statistical moments ?

And also :

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

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

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

and

- Is assimilation worth all the concern we give to it ?

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

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

Innovation vector is the only objective source of information on data error. Implementing assimilation requires knowing, at least to some extent, how the background on the one hand, the observations on the other, contribute to the innovation. That cannot be obtained from the innovation alone. Consistency between *a priori* assumed and observed statistics of d is not sufficient for ensuring optimality of assimilation.

We will consider assimilation schemes of the form

$$x^a = x^b + K\,d \qquad (2)$$

where K is the *gain matrix* (not necessarily optimal).

(2) $\Leftrightarrow$ if data are exact, then analysis is exact too ($x^a = x$).

Difference between data and assimilated fields

$$\delta \equiv \begin{pmatrix} x^b - x^a \\ y - \boldsymbol{H}x^a \end{pmatrix} = z - \Gamma x^a$$

$$\delta = \begin{pmatrix} -\boldsymbol{K}d \\ (I_p - \boldsymbol{HK})d \end{pmatrix}$$

For given gain matrix K, one-to-one transformation between d and δ. Exactly equivalent to perform diagnostics on either innovation or *data-minus-analysis* (*DmA*) difference.

Objective validation of quality of an assimilation system can be made only by comparison with unbiased independent data, *i. e.*, data affected by errors that are statistically independent of errors affecting the data used in the assimilation (but the required independence cannot be objectively validated, and has to be assumed *a priori*).

If errors are uncorrelated in time, magnitude of innovation vector is an objective measure of the quality of the assimilation.

Other possible diagnostic : check if observed statistics of innovation vector are consistent with *a priori* assumed statistics, *i. e.* check that

$$E(d) = 0$$

Any systematic bias in d (or equivalently in δ) is the signature of an improperly taken account bias in the data.

With hypotheses made above, check that

$$E(dd^T) = HP^bH^T + R$$

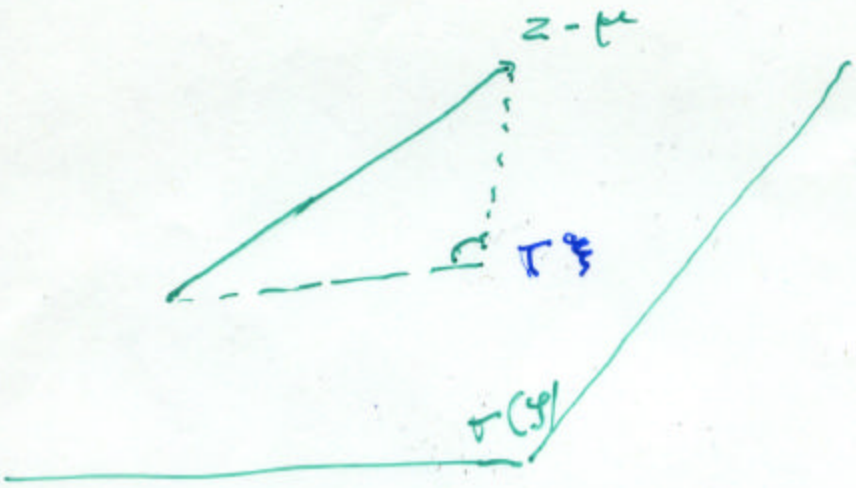
In particular

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

Assimilated fields must fit data to within assumed accuracy of the latter (system is *efficient*, as defined by Hollingsworth and Lönnberg, 1989). If they do not, inconsistency between *a priori* assumed and *a posteriori* observed statistics of d (or δ).

$$f(\xi) = \frac{1}{2}\left[\Gamma\xi - (z-\mu)\right]^t S^{-1}\left[\Gamma\xi - (z-\mu)\right]$$

$\Gamma\xi$ is projection of $z-\mu$ onto image space $\Gamma(\mathcal{G})$ according to S-Mahalanobis scalar product.



- Project $z-\mu$ onto $\Gamma(\mathcal{G})$
- Take inverse of projection through Γ^{-1} (unambiguously defined since Γ is of full rank)

Choose first u basis vectors in image space $\Gamma(\varphi)$, remaining $\rho = m - u$ vectors in S-orthogonal space $\perp\Gamma(\varphi)$

$$\Gamma = \begin{pmatrix} \Gamma_1 \\ 0 \end{pmatrix} \quad \Gamma_1 \text{ invertible.}$$

$$\zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad z = \begin{pmatrix} z_1 = \Gamma_1 x + \zeta_1 \\ z_2 = \zeta_2 \end{pmatrix}$$

$$x^a = \Gamma_1^{-1}(z_1 - \mu_1) = x + \Gamma_1^{-1}(\zeta_1 - \mu_1)$$

$$P^a = E\left[(x^a - x)(x^a - x)^T\right] = \Gamma_1^{-1} S_1 \Gamma_1^{-T}$$

Performing analysis and determining corresponding covariance matrix P^a requires knowledge of only μ_1 and S_1, i.e probability distribution of ζ_1

DmA difference

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

whose probability distribution depends only on distribution of ζ_2

Question. An inconsistency having been observed, what can be done in order to, *e. g.*, improve *a priori* statistics of data errors ?

Variational formulation of the *BLUE*. Analysis x^a minimizes objective function :

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

Shows that Γx^a is orthogonal projection of data vector z onto image space $\Gamma(\mathcal{S})$ (in the sense of Mahalanobis norm associated with covariance matrix S). The DmA difference $\delta = z - \Gamma x^a$ is the part of the data that has been rejected in the estimation. Its statistical properties are entirely independent of the statistical properties of the analysis error.

Consequence. A possible inconsistency between observed and expected statistics of the data-minus-analysis difference δ (or, equivalently, of the innovation d) can always be 'explained out' without changing either the analysis x^a or the estimated analysis error covariance matrix P^a. Answer to question above is : nothing can be done.

As already said, consistency between *a priori* assumed and observed statistics of d (or equivalently of δ) is not sufficient for ensuring optimality of assimilation. It is not even necessary.

True of bayesian estimation in general ? I think yes.

Independent hypotheses, which cannot be objectively validated (at least on the basis of the innovation) will always be necessary.

Problem. Find minimum set of hypotheses leaving only parameters that can be determined from statistics of the innovation (or of the DmA difference).

Variational assimilation (either 3-D or 4-D)

$$\mathcal{J}(\xi) \equiv (1/2)\,(\xi - x^b)^T\,[P^b]^{-1}\,(\xi - x^b) + (1/2)\,(H\xi - y)^T\,R^{-1}\,(H\xi - y)$$

Minimum reached for $\xi = x^a$ For a perfectly consistent system

$$\mathcal{J}(x^a) = (1/2)\,d^T.[E(dd^T)]^{-1}\,d$$

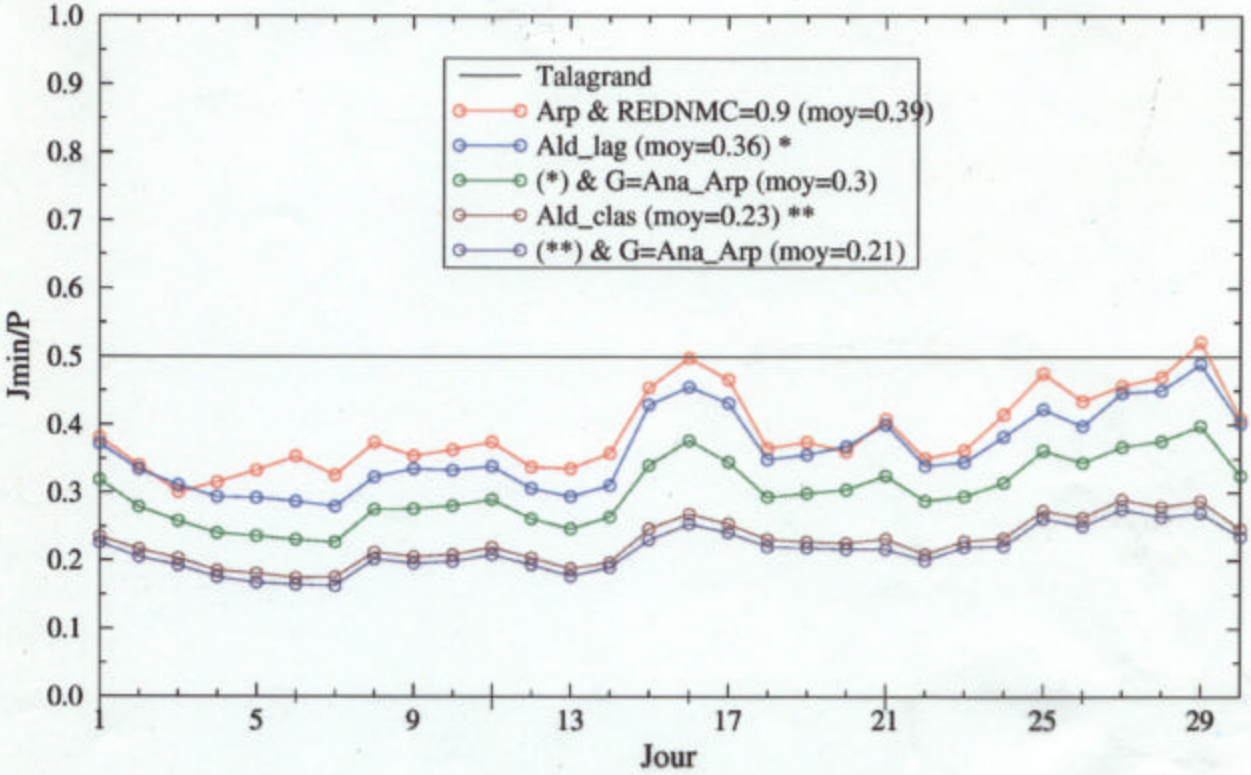
Minimum of objective function is norm of innovation with respect to its own Mahalanobis scalar product. On expectation

$$E[\mathcal{J}(x^a)] = p/2$$

Often called χ^2-criterion (but, unless data errors are gaussian, $\mathcal{J}(x^a)$ will not in general follow a χ^2-probability distribution).

If errors are gaussian

$$Var\left[\mathcal{J}(x^a)\right] = \frac{p}{2}$$



W. Sadiki (2005)

Results for ECMWF (January 2003, $n \approx 8\text{x}10^6$)

- Operations ($p \approx 1.4\text{x}10^6$)

$$2\,\mathcal{J}(x^a) / p \approx 0.40 - 0.45$$

Magnitude of innovation d largely overestimated by P^b and R (an unaccounted for bias in d would make $\mathcal{J}(x^a)$ too large).

- Assimilation without satellite observations ($p \approx 2 - 3\text{x}10^5$)

$$2\,\mathcal{J}(x^a) / p \approx 1. - 1.05$$

Similar results obtained at other NWP centres (C. Fischer, W. Sadiki with Aladin model, Météo-France, T. Payne, UKMO).

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

Objective function

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

where

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

with $\dim y_k = m_k$

Accuracy of analysis

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

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

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

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.

Can be numerically computed (Wahba, Fisher, Desroziers and Ivanov).

Can be extended to measure of relative contribution of any subset of data to accuracy of any subjet of analysed fields (but practical computation ?).

Relative informative contents can be related to entropy!

Informative contents of uncorrelated sets of data are additive

$$I(Y_1 \cup Y_2) = I(Y_1) + I(Y_2)$$

If mutual correlation

$$I(Y_1 \cup Y_2) \geq I(Y_1) + I(Y_2)$$

In particular, relative informative content of background x^b

$$\frac{1}{n} Tr\left[P^a (P^b)^{-1}\right] = 1 - \frac{1}{n} Tr(KH)$$

(Rodgers, 2000)

Relative informative content of observation vector y (if $E(\xi^a \varepsilon^T) = 0$)

$$\frac{1}{n} Tr(KH)$$

Chapnik et al., 2005

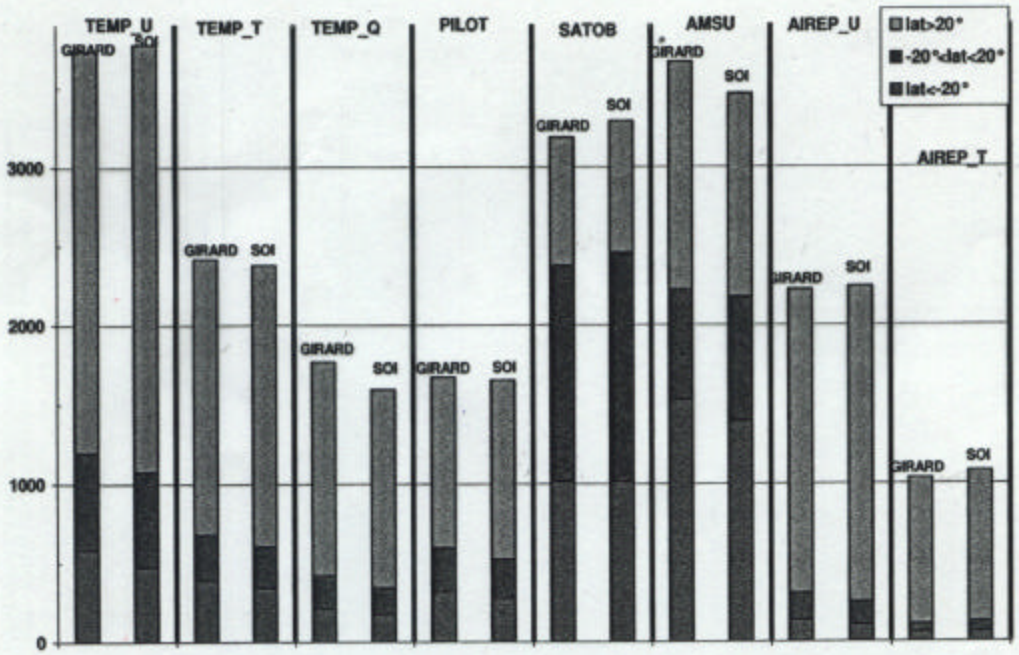


FIG. B.3 – DFS for several subsets of observations. The subsets are separated by thick black lines. For each subset the left hand side bar was computed with Girard's method and the right hand side bar was computed with the SOI method. Each bar is divided into three parts, showing the contribution from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

DFS = Degrees of Freedom for Signal

$$= \mathrm{Tr}\left(S_k^{-\frac{1}{2}} H_k P^a H_k^T S_k^{-\frac{1}{2}}\right)$$

Consider now subset x_1 of analyzed fields, $dim\, x_1 = n_1$.

x_2 : component of x orthogonal to x_1 with respect to Mahalanobis norm associated with P^a ($\Leftrightarrow$ analysis errors on x_1 and x_2 are uncorrelated).

In basis (x_1, x_2)

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

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

Then

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

and

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

measures relative contribution of data subset y_k to accuracy of analysis of state subset x_1.

But can it be numerically computed at an acceptable cost (requires decomposition of x into components x_1 and x_2) ?

For a perfectly consistent system

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

It is possible to compare $E[\mathcal{J}_k(x^a)]$, as determined operationally, and $(1/2)\,[m_k - \mathrm{tr}(S_k^{-1/2} H_k P^a H_k^T S_k^{-1/2})]$, as computed directly, thus providing a check of the consistency of the assimilation system (Chapnik *et al.*, 2003).

Also, $E[\mathcal{J}_k(x^a)]$ must be less than $m_k/2$ $\Leftrightarrow$ every piece of data must fit the analysis to within its assumed accuracy.

In particular, relative informative content of background x^b

$$\frac{1}{n} T_2(P^a P^{b-1}) = 1 - \frac{1}{n} T_2(KH)$$

(Rodgers, 2000)

Relative informative content of observations (if $E(\varepsilon \xi^{eT}) = 0$)

$$\frac{1}{n} T_2(KH)$$

Desroziers and Ivanov (2001)
Chapnik et al. (2004, 2005)

Rescale variances of ~~inde~~ independent blocks of data such that the condition

$$E\left[\mathcal{J}_\varepsilon(x^a)\right] = \frac{1}{2}\left[m_\varepsilon - \mathrm{tr}\left(S_\varepsilon^{-\frac{1}{2}} H_\varepsilon P^a H_\varepsilon^T S_\varepsilon^{-\frac{1}{2}}\right)\right]$$

is verified for all k's (correlations are not modified)

$\mathrm{tr}\left(S_\varepsilon^{-\frac{1}{2}} H_\varepsilon P^a H_\varepsilon^T S_\varepsilon^{-\frac{1}{2}}\right)$ computed by simulation

Fixed point algorithm.

Converges very rapidly

[6] D. Girard. A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. Technical Report 637-M, IMAG, Grenoble, France, 1987.

[7] A. Hollingsworth and P. Lönnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. part I: The wind field. *Tellus*, 38:111–136, 1986.

[8] K. Ide, P. Courtier, M. Ghil, and A.C. Lorenc. Unified notation for data assimilation: operational, sequential and variational. *J. Met. Soc. Japan*, 75:181–189, 1997.

[9] A. Joly and et al. Overview of the field phase of the Fronts and Atlantic Storm-Track Experiment (FASTEX) project. *Quart. J. Roy. Meteor. Soc.*, 125, 1999. Submitted to Quart. J. Roy. Meteor. Soc., under revision.

[10] O. Talagrand. A posteriori verification of analysis and assimilation algorithms. In *Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems*, pages 17–28, Reading, 1999. 2-4 November.

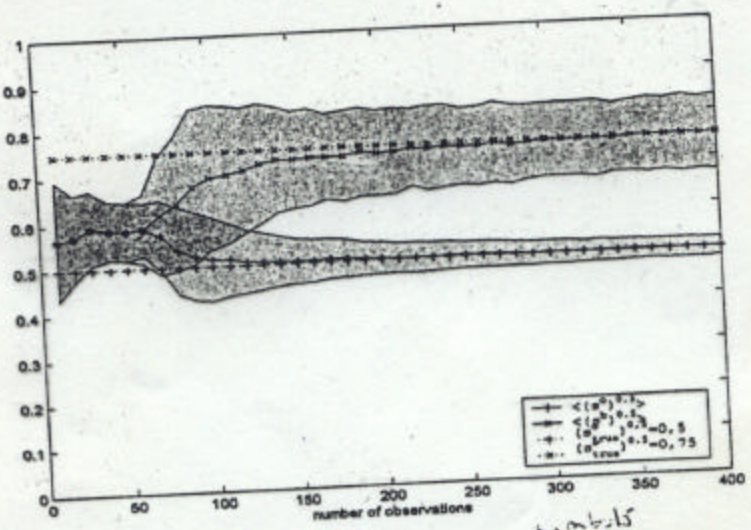


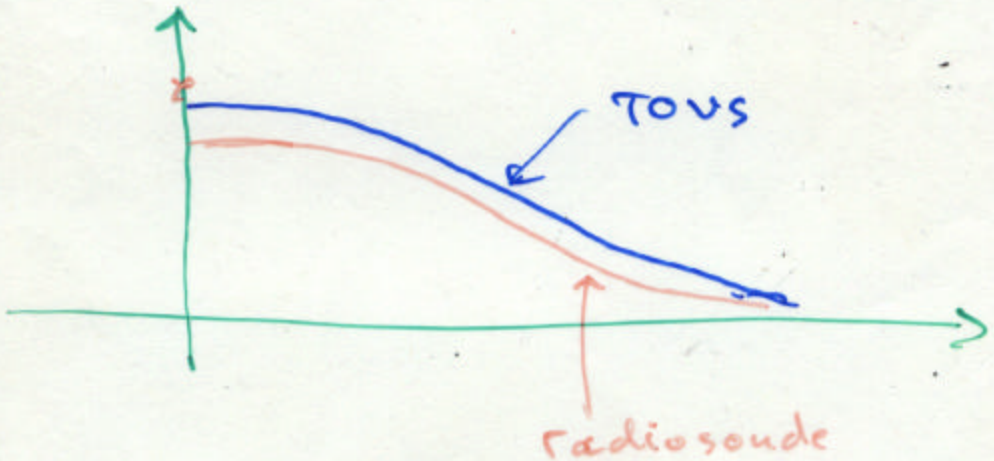
Figure 1. Influence of the number of observations on the tuning coefficient. The x-coordinate is the number of observations The full lines with "+" and "x" markers are respectively the means of the square roots of s^o and s^b computed over 200 experiments, while the dotted lines, with the same markers respectively indicate the true value for the square roots of s^o and s^b. The grey areas are the region within a range of one standard deviation (computed over the same experiments) from the mean.

Wahba 1985.

Chapnik, 2003

Tuning of background and observation error variances (simulated data)

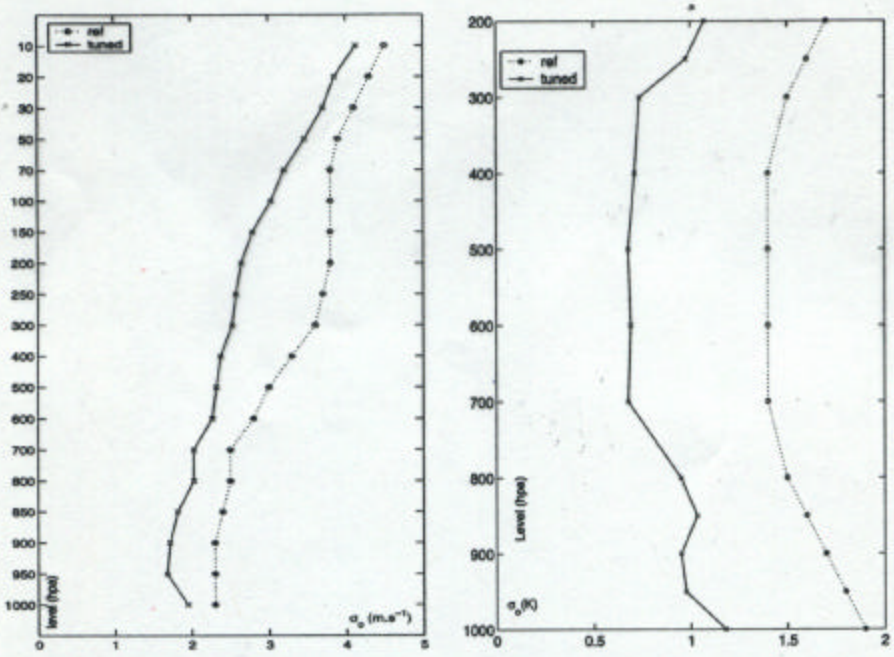
Spatial correlation ? (Lönnberg, Hollingsworth, Daley)



If difference detectable, spatial correlation can be estimated

Adjustment as done by Chapnik et al ?

Chapnik et al., 2005



(a) TEMP wind observations

(b) AIREP temperature observations

FIG. B.6 – Comparison between the tuned σ_o profile (full line, "x" markers) and the originally prescribed profile (dotted line, "*" marker) for TEMP wind observations and AIREP temperature observations.

Chapnick et al., 2005

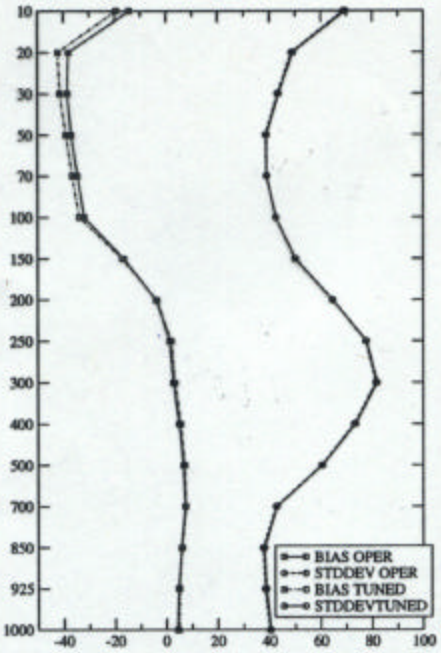


FIG. B.9 – Statistics of TEMP geopotential observations - 96H forecasts, for operational and tuned forecasts. The bias (square markers) and the standard deviation (circle markers) are shown for the operational forecast (dashed line) and for the tuned forecast (full line).

Impact on quality of forecasts negligible, or at best very weak

- Specification of variances already optimal?

- Specification not optimal, but improvements are to be made elsewhere (background error covariances to be made situation dependent, ..)

Independent, unbiased observation

$$u = Cx + \beta \qquad E(\beta) = 0 \qquad E(\beta \zeta^T) = 0$$

$$u - Cx^a = \underbrace{u - Cx}_{\beta} + C(x - x^a)$$

$$E(\beta d^T) = 0$$

Optimality of estimate Cx^a equivalent to

$$E(u - Cx^a) = 0$$

$$E[(u - Cx^a) d^T] = 0$$

Can be objectively checked

short (3–12 h) forecasts are followed by an analysis procedure such as statistical interpolation, employ (2.1) with a nonlinear rather than a linear model, together with an analysis equation of the form (2.4). The gain matrix or analysis weights are obtained from approximations to (2.3), (2.7), and (2.8), with the crudest approximations being made to (2.3).

If present-day data assimilation systems are viewed in this light, it makes sense to estimate the lagged innovation covariances, using the definition (2.10). In particular, the nonzero lag innovation covariance estimation for real systems, together with the characteristic misspecification signatures of sections 4 and 5, might provide insight into the shortcomings of the system. The legitimacy of this approach, in light of the simplicity of the models considered above, will be discussed in section 7.

With this goal in mind, we attempted to estimate the lagged innovation correlations for the Canadian Meteorological Centre (CMC) four-dimensional data assimilation system. This system has been recently described by Mitchell et al. (1990). Briefly, the forecasts are made by a multilevel primitive equation hemispheric spectral model. The analysis stage employs a variant of statistical interpolation.

Innovation sequences from the CMC system were obtained for the North American radiosonde network for the period 1 February–31 March 1989. As discussed by Mitchell et al. (1990), a major change had been made to the specified forecast error statistics in January 1989, but the system was stable during the period of interest. The dataset consisted of 0000 and 1200 UTC 500-mb height and wind innovations for 118 radiosonde stations in the region 25°–90°N,20°–178°W.

There were a maximum of 118 reports for each station during the period. No station with less than 80 reports was used, but there were several days in the second half of March when there were very few reports.

The 60-day time mean was removed from the innovation sequence at each station. This innovation sequence was then normalized by dividing by the 60-day rms value from that station. Estimates of the lagged innovation correlation could then be produced simply by multiplying together the normalized innovations at any two stations, for a particular lag, and then averaging over the 60-day period.

Many operational centers have developed sophisticated software for estimating lag-zero innovation covariances. Rather than attempting to adapt this software to the nonzero lag case, a simpler approach was taken. For most cases, the correlations were calculated with respect to a key station in the center of the continent. This key station was chosen to be Omaha, Nebraska, and is marked with a solid dot in Fig. 11. These results were then substantiated by using two adjacent stations as key stations: Ogallala, Nebraska and Topeka, Kansas. However, only calculations for Omaha will be shown here. Omaha was chosen as the key station as it is in the center of the continent and is surrounded by a relatively dense, reasonably homogeneous network of radiosonde stations; thus, a complete two-dimensional pattern could be obtained.

The h/h, u/u, v/v, h/v, v/h, h/u, and u/h innovation correlations were determined for Omaha (and the other two key stations) for 12-h lags of -3, -2, -1, 0, $+1$, $+2$, $+3$. However, most attention was paid to the -1 and 0 lags.

Nonzero lag innovation correlations have apparently not been previously calculated. However, Buell (1972) has calculated lagged correlations for the 500-mb wind field itself. Taking wind observations from the North American radiosonde network, he subtracted off the

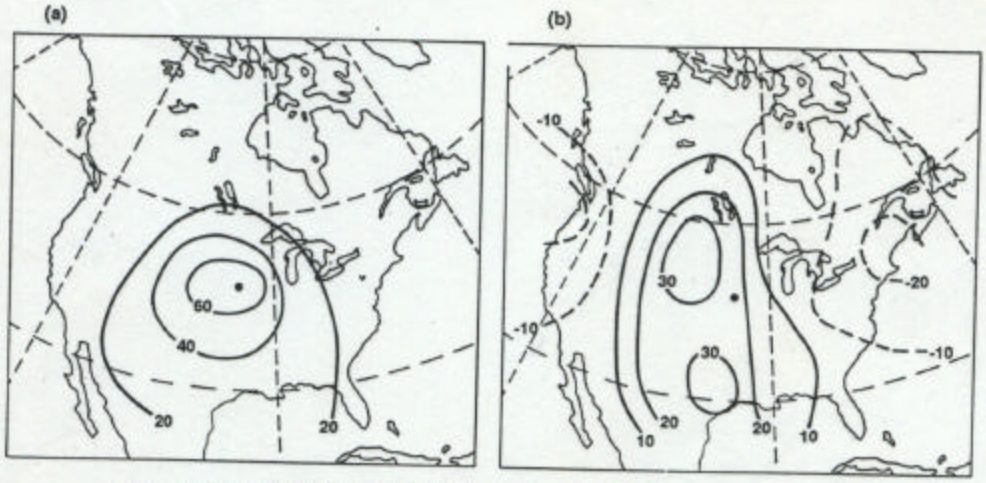


FIG. 11. Correlations c^0_{AA} (a) and c^{-1}_{AA} (b) for Omaha. Correlations are multiplied by 100.