École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2011-2012

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation d'Observations

Olivier Talagrand

Cours 7

21 Mai 2012

Best Linear Unbiased Estimate

$\boldsymbol{x}^b = \boldsymbol{x} + \zeta^b$ (1)

$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{\varepsilon}$ (2)

with $E(\zeta^b) = 0$, $E(\varepsilon) = 0$, $E(\zeta^b \varepsilon^{\mathrm{T}}) = 0$ (not restrictive), $E(\zeta^b \zeta^{b\mathrm{T}}) = P^b$, $E(\varepsilon\varepsilon^{\mathrm{T}}) = R$

Best Linear Unbiased Estimate (*BLUE*)

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

Variational form. *BLUE* x^a minimizes following scalar *objective function*, defined on state space

$$\xi \in \mathcal{S} \rightarrow$$
$$\mathcal{J}(\xi) = (1/2)\,(x^b - \xi)^{\mathrm{T}} [P^b]^{-1} (x^b - \xi) + (1/2)\,(y - H\xi)^{\mathrm{T}} R^{-1} (y - H\xi)$$
$$= \mathcal{J}_b + \mathcal{J}_o$$

P^a is inverse of matrix of second derivatives (*hessian*) of function $\mathcal{J}(\xi)$

Variational approach can easily be extended to time dimension.

Suppose for instance available data consist of

- Background estimate at time 0

  $x_0^b = x_0 + \zeta_0^b$ $\qquad\qquad$ $E(\zeta_0^b \zeta_0^{bT}) = P_0^b$

- Observations at times $k = 0, \ldots, K$

  $y_k = H_k x_k + \varepsilon_k$ $\qquad\qquad$ $E(\varepsilon_k \varepsilon_j^T) = R_k \delta_{kj}$

- Model (supposed for the time being to be exact)

  $x_{k+1} = M_k x_k$ $\qquad\qquad$ $k = 0, \ldots, K\text{-}1$

Errors assumed to be unbiased and uncorrelated in time, H_k and M_k linear

Then objective function

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

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^T\,R_k^{-1}\,[y_k - H_k\xi_k]$$

subject to $\xi_{k+1} = M_k \xi_k$, $\qquad\qquad$ $k = 0, \ldots, K\text{-}1$

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^{\,b} - \xi_0)^{\mathrm{T}}\,[P_0^{\,b}]^{-1}\,(x_0^{\,b} - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^{\mathrm{T}}\,R_k^{-1}\,[y_k - H_k\xi_k]$$

Background is not necessary, if observations are in sufficient number to overdetermine the problem. Nor is strict linearity.

How to minimize objective function with respect to initial state $u = \xi_0$ (u is called the *control variable* of the problem) ?

Use iterative minimization algorithm, each step of which requires the explicit knowledge of the local gradient $\nabla_u \mathcal{J} \equiv (\partial \mathcal{J}/\partial u_i)$ of $\mathcal{J}$ with respect to u.

How to numerically compute the gradient $\nabla_u \mathcal{J}$?

Direct perturbation, in order to obtain partial derivatives $\partial \mathcal{J}/\partial u_i$ by finite differences ? That would require as many explicit computations of the objective function $\mathcal{J}$ as there are components in u. Practically impossible.

Gradient computed by *adjoint method*.

Adjoint Method

Input vector $\boldsymbol{u} = (u_i)$, $\dim \boldsymbol{u} = n$

Numerical process, implemented on computer ($e.\ g.$ integration of numerical model)

$$\boldsymbol{u} \rightarrow \boldsymbol{v} = \boldsymbol{G}(\boldsymbol{u})$$

$\boldsymbol{v} = (v_j)$ is *output vector* , $\dim \boldsymbol{v} = m$

Perturbation $\delta \boldsymbol{u} = (\delta u_i)$ of input. Resulting first-order perturbation on $\boldsymbol{v}$

$$\delta v_j = \Sigma_i (\partial v_j / \partial u_i)\ \delta u_i$$

or, in matrix form

$$\delta \boldsymbol{v} \ = \ \boldsymbol{G}' \delta \boldsymbol{u}$$

where $\boldsymbol{G}' \equiv (\partial v_j / \partial u_i)$ is local matrix of partial derivatives, or jacobian matrix, of $\boldsymbol{G}$.

Adjoint Method (continued 1)

$$\delta \boldsymbol{v} = \boldsymbol{G}' \delta \boldsymbol{u} \qquad \text{(D)}$$

Scalar function of output

$$\mathcal{J}(\boldsymbol{v}) = \mathcal{J}[\boldsymbol{G}(\boldsymbol{u})]$$

Gradient $\boldsymbol{\nabla}_{\boldsymbol{u}}\mathcal{J}$ of $\mathcal{J}$ with respect to input $\boldsymbol{u}$?

'Chain rule'

$$\partial\mathcal{J}/\partial u_i = \Sigma_j \, \partial\mathcal{J}/\partial v_j \, (\partial v_j/\partial u_i)$$

or

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\mathcal{J} = \boldsymbol{G}'^{\mathrm{T}} \, \boldsymbol{\nabla}_{\boldsymbol{v}}\mathcal{J} \qquad \text{(A)}$$

Adjoint Method (continued 2)

$\boldsymbol{G}$ is the composition of a number of successive steps

$$\boldsymbol{G} = \boldsymbol{G}_N \circ \ldots \circ \boldsymbol{G}_2 \circ \boldsymbol{G}_1$$

'Chain rule'

$$\boldsymbol{G}' = \boldsymbol{G}_N' \ldots \boldsymbol{G}_2' \, \boldsymbol{G}_1'$$

Transpose

$$\boldsymbol{G}'^{\mathrm{T}} = \boldsymbol{G}_1'^{\mathrm{T}} \, \boldsymbol{G}_2'^{\mathrm{T}} \ldots \boldsymbol{G}_N'^{\mathrm{T}}$$

Transpose, or *adjoint*, computations are performed in reversed order of direct computations.

If $\boldsymbol{G}$ is nonlinear, local jacobian $\boldsymbol{G}'$ depends on local value of input $\boldsymbol{u}$. Any quantity which is an argument of a nonlinear operation in the direct computation will be used again in the adjoint computation. It must be kept in memory from the direct computation (or else be recomputed again in the course of the adjoint computation).

If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

Adjoint Approach

$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^T\,R_k^{-1}\,[y_k - H_k\xi_k]$

subject to $\xi_{k+1} = M_k\xi_k\,, \qquad k = 0, \ldots, K\text{-}1$

Control variable $\qquad \xi_0 = \boldsymbol{u}$

Adjoint equation

$$\lambda_K = \qquad H_K^T\,R_K^{-1}\,[H_K\,\xi_K - y_K]$$

$$\lambda_k = M_k^T\lambda_{k+1} + H_k^T\,R_k^{-1}\,[H_k\,\xi_k - y_k] \qquad k = K\text{-}1, \ldots, 1$$

$$\lambda_0 = M_0^T\lambda_1 \quad + H_0^T\,R_0^{-1}\,[H_0\,\xi_0 - y_0] \;+\; [P_0^b]^{-1}\,(\xi_0 - x_0^b)$$

$$\nabla_u\mathcal{J} = \lambda_0$$

Result of direct integration (ξ_k), which appears in quadratic terms in expression of objective function, must be kept in memory from direct integration.

Adjoint Approach (continued 2)

Nonlinearities ?

$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k(\xi_k)]^T\,R_k^{-1}\,[y_k - H_k(\xi_k)]$

subject to $\xi_{k+1} = M_k(\xi_k)$, $\quad k = 0, \ldots, K\text{-}1$

Control variable $\quad \xi_0 = \boldsymbol{u}$

Adjoint equation

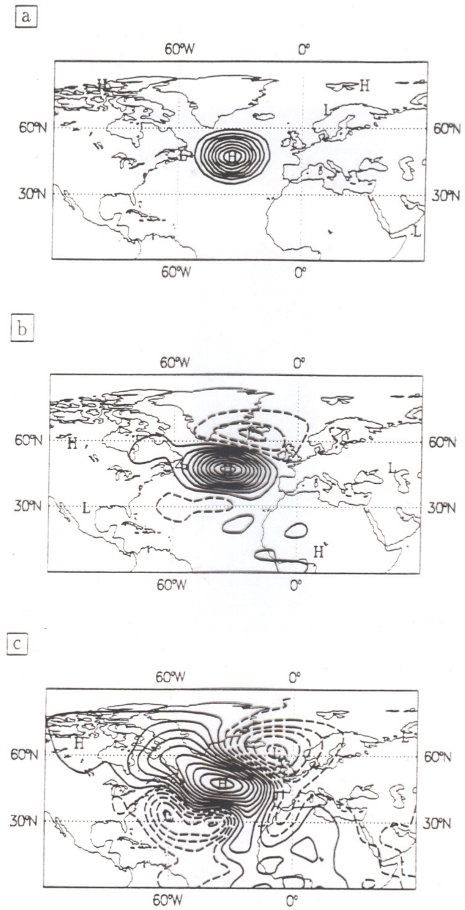
$\lambda_K = \quad H_K'^T R_K^{-1}\,[H_K(\xi_K) - y_K]$

$\lambda_k = M_k'^T \lambda_{k+1} + H_k'^T R_k^{-1}\,[H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$

$\lambda_0 = M_0'^T \lambda_1 + H_0'^T R_0^{-1}\,[H_0(\xi_0) - y_0] + [P_0^b]^{-1}\,(\xi_0 - x_0^b)$

$$\nabla_u \mathcal{J} = \lambda_0$$

Not heuristic (it gives the exact gradient $\nabla_u \mathcal{J}$), and really used as described here.



Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.

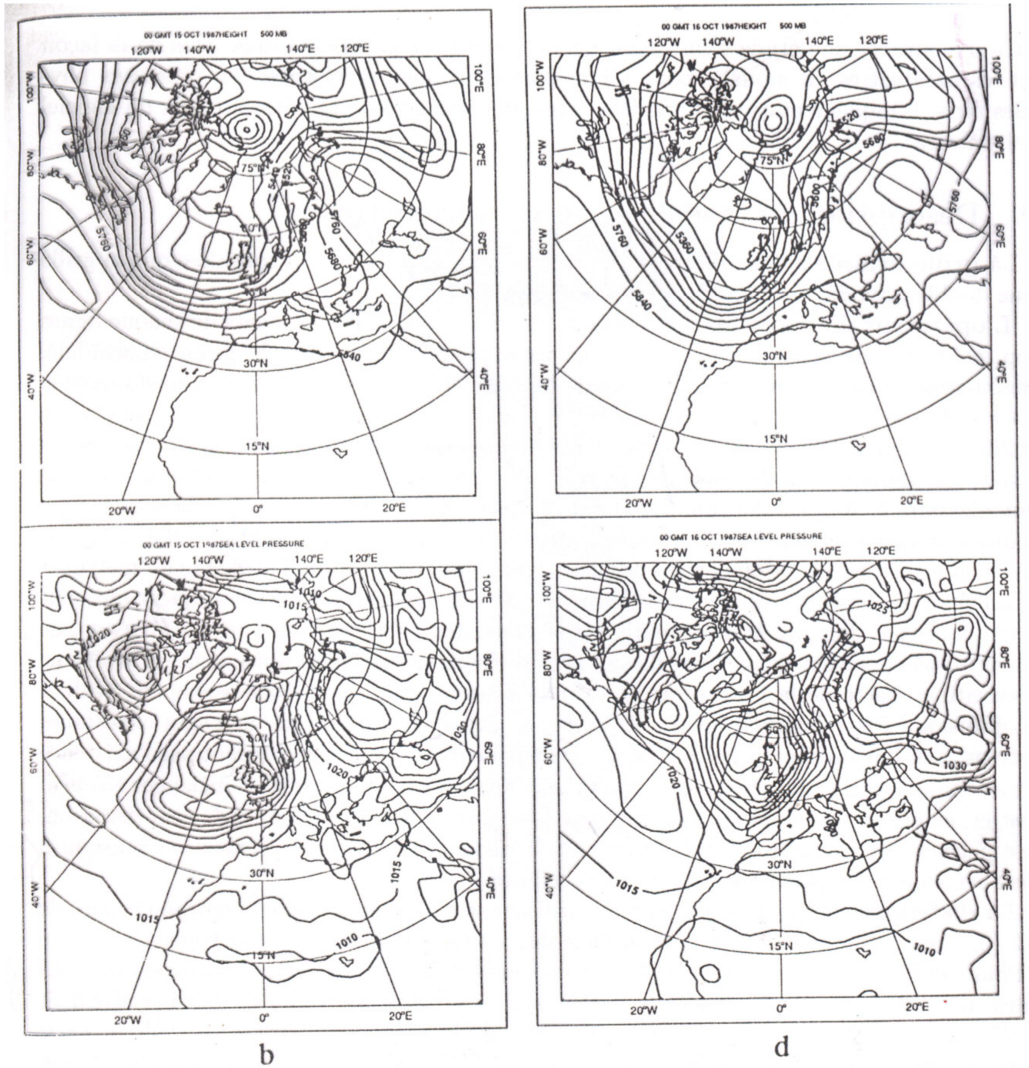
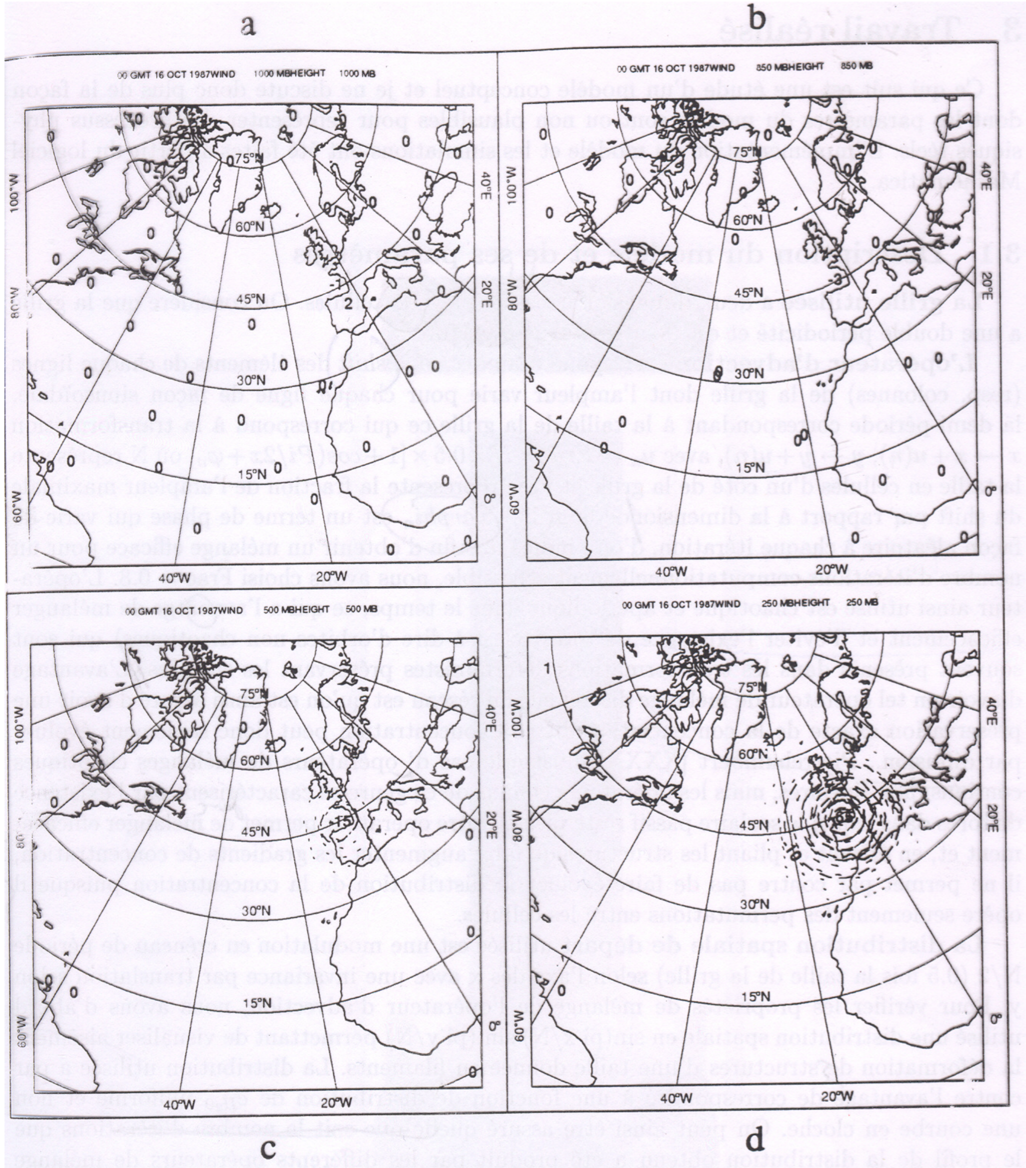


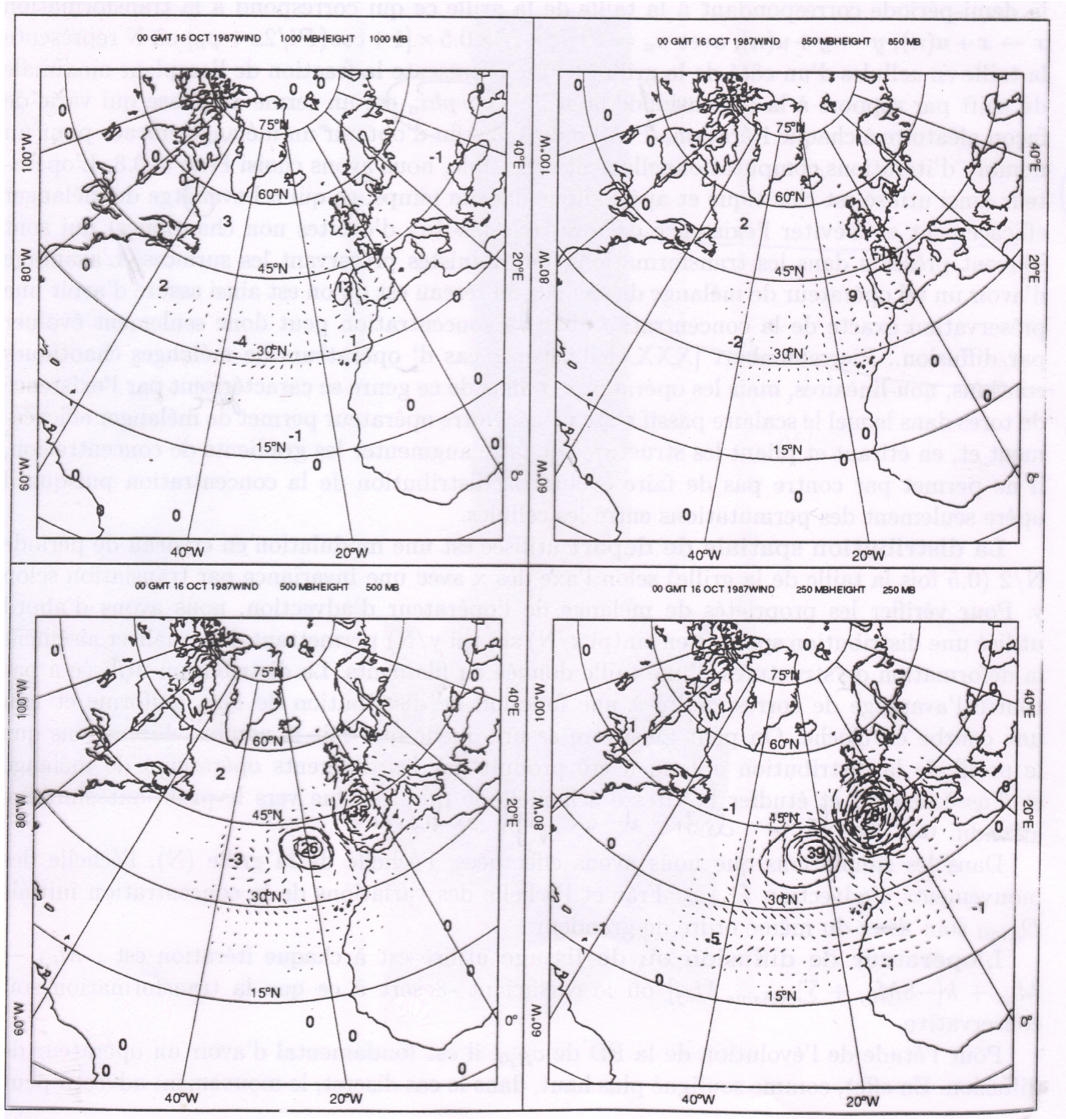
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October are from the initial estimate of the initial conditions for the 4DVAR minimization. The fields for 16 October are from the 24-h T63 adiabatic model forecast from the initial conditions. Contour intervals are 80 m and 5 hPa.

Thépaut *et al*., 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



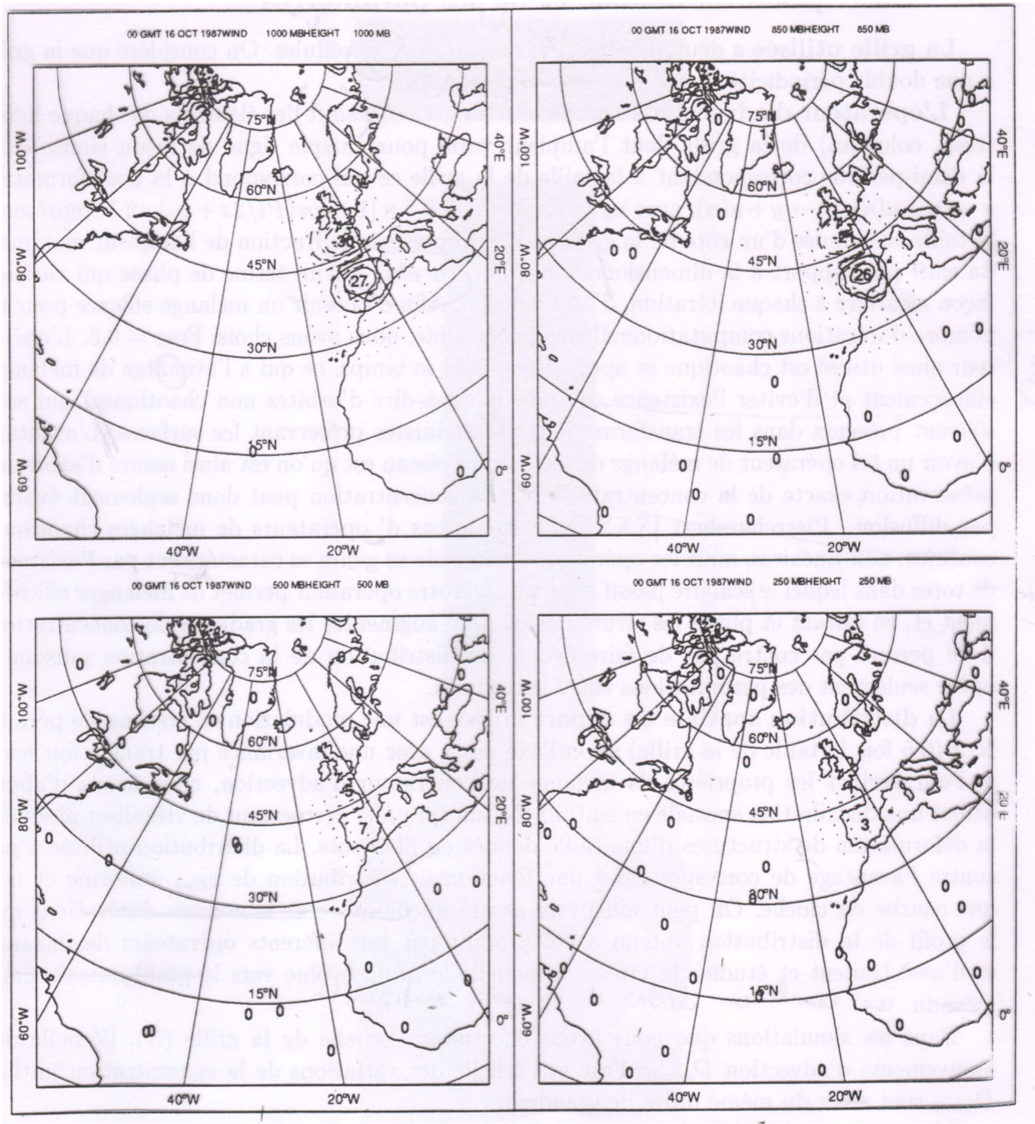
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

Thépaut *et al*., 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



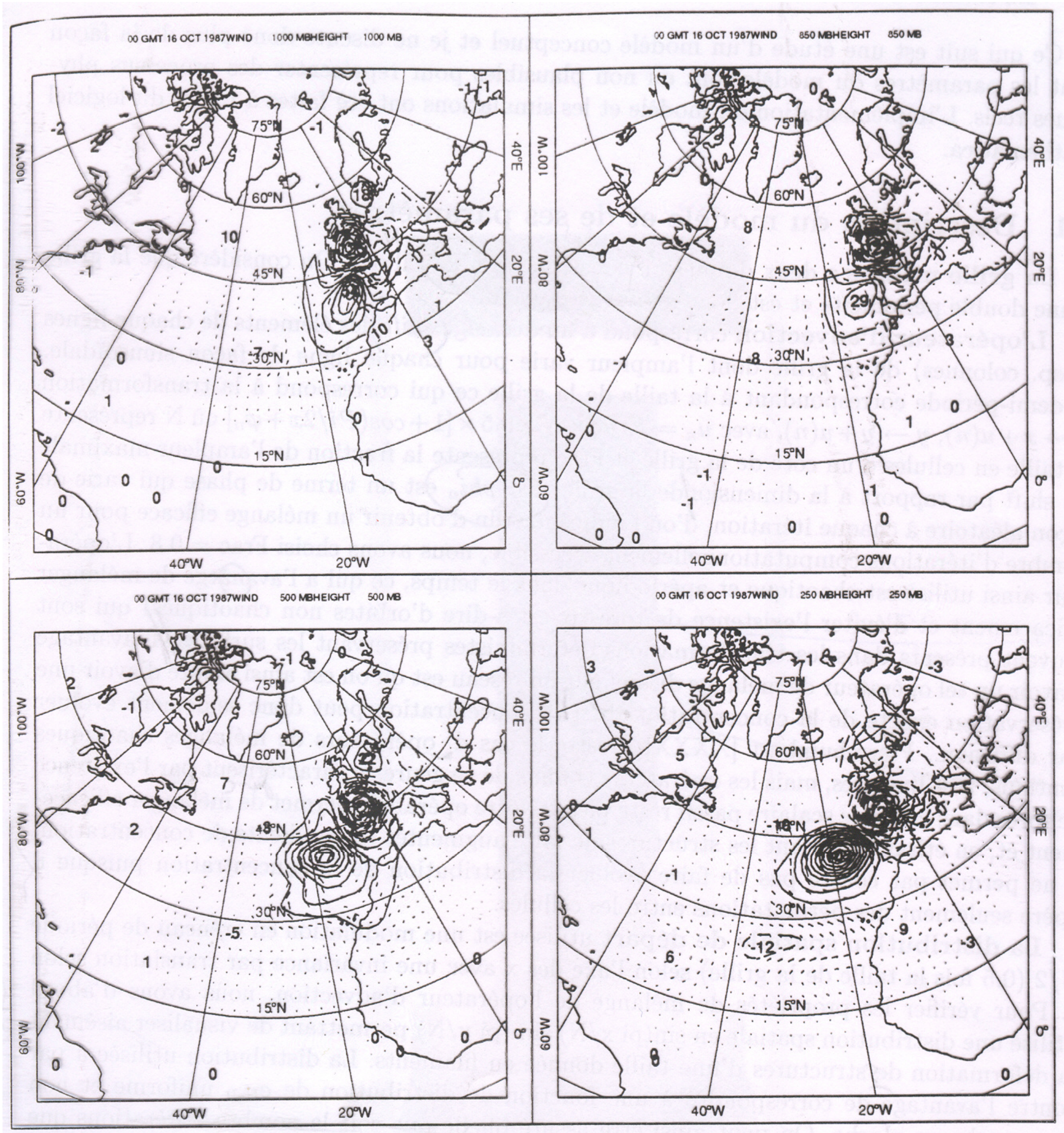
Same as before, but at the end of a 24-hr 4D-Var

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



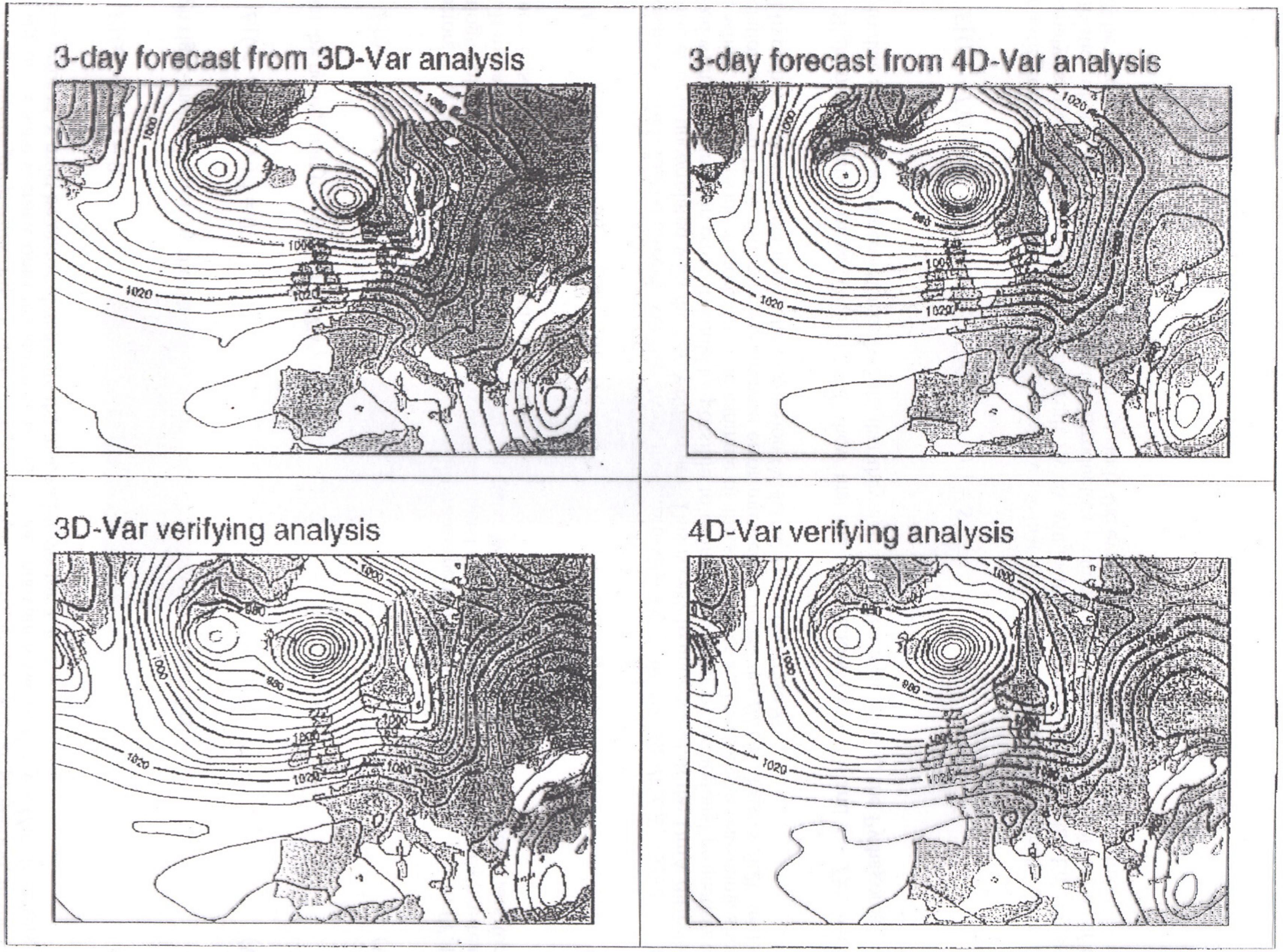
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



Same as before, but at the end of a 24-hr 4D-Var

Thépaut *et al*., 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



ECMWF, Results on one FASTEX case (1997)

Strong Constraint 4D-Var is now used operationally at several meteorological centres (Météo-France, UK Meteorological Office, Canadian Meteorological Centre, Japan Meteorological Agency, …) and, until recently, at ECMWF. The latter now has a 'weak constraint' component in its operational system.

Buehner *et al*. (*Mon. Wea. Rev*., 2010)

For the same numerical cost, and in meteorologically realistic situations, Ensemble Kalman Filter and Variational Assimilation produce results of similar quality.

Weak constraint variational assimilation allows for errors in the assimilating model

Data

- Background estimate at time 0

$$x_0^b = x_0 + \zeta_0^b \qquad E(\zeta_0^b \zeta_0^{bT}) = P_0^b$$

- Observations at times $k = 0, \ldots, K$

$$y_k = H_k x_k + \varepsilon_k \qquad E(\varepsilon_k \varepsilon_k^T) = R_k$$

- Evolution equation

$$x_{k+1} = M_k x_k + \eta_k \qquad E(\eta_k \eta_k^T) = Q_k \qquad k = 0, \ldots, K\text{-}1$$

Errors assumed to be unbiased and uncorrelated in time, H_k and M_k linear

Then objective function

$$(\xi_0, \xi_1, \ldots, \xi_K) \rightarrow$$

$$\mathcal{J}(\xi_0, \xi_1, \ldots, \xi_K)$$

$$= (1/2)\ (x_0^{\,b} - \xi_0)^{\mathrm{T}}\ [P_0^{\,b}]^{-1}\ (x_0^{\,b} - \xi_0)$$

$$+ (1/2)\ \Sigma_{k=0,\ldots,K}[y_k - H_k\xi_k]^{\mathrm{T}}\ R_k^{-1}\ [y_k - H_k\xi_k]$$

$$+ (1/2)\ \Sigma_{k=0,\ldots,K-1}[\xi_{k+1} - M_k\xi_k]^{\mathrm{T}}\ Q_k^{-1}\ [\xi_{k+1} - M_k\xi_k]$$

Can include nonlinear M_k and/or H_k.

Time-correlated Errors

Example of time-correlated observation errors

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

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

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

Additional observation then becomes available

$$z_3 = x + \zeta_3$$

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

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

`

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

Observations

$y_k = x_k + \varepsilon_k\,,$ $\qquad k = 0, 1, 2$ $\qquad E(\varepsilon_k{}^2) = r$, 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

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

Variational assimilation has been extended to non Gaussian probability distributions (lognormal distributions), the unknown being the mode of the conditional distribution (M. Zupanski, Fletcher).

Bayesian character of variational assimilation ?

- If everything is linear and gaussian, ready recipe for obtaining bayesian sample

Perturb data (background, observations and model) according to their error probability distributions, do variational assimilation, and repeat process

Sample of system orbits thus obtained is bayesian

- If not, very little can be said at present

Conclusion on Sequential Assimilation

Pros

‘Natural’, and well adapted to many practical situations

Provides, at least relatively easily, explicit estimate of estimation error

Cons

Carries information only forward in time (of no importance if one is interested only in doing forecast)

In present form, optimality is possible only if errors are independent in time

Conclusion on Variational Assimilation

Pros

Carries information both forward and backward in time (important for reassimilation of past data).

Can easily take into account temporal statistical dependence (Järvinen *et al.*)

Does not require explicit computation of temporal evolution of estimation error

Very well adapted to some specific problems (*e. g.*, identification of tracer sources)

Cons

Does not readily provide estimate of estimation error

Requires development and maintenance of adjoint codes. But the latter can have other uses (sensitivity studies).

- Dual approach seems most promising. But still needs further development for application in non exactly linear cases.

- Is ensemble variational assimilation possible ? Probably yes. But also needs development.

How to write the adjoint of a code ?

Operation $a = b \times c$

Input b, c Output a but also b, c

For clarity, we write

$$a = b \times c$$
$$b' = b$$
$$c' = c$$

$\partial J/\partial a$, $\partial J/\partial b'$, $\partial J/\partial c'$ available. We want to determine $\partial J/\partial b$, $\partial J/\partial c$

Chain rule

$$\partial J/\partial b = (\partial J/\partial a)\underbrace{(\partial a/\partial b)}_{c} + (\partial J/\partial b')\underbrace{(\partial b'/\partial b)}_{1} + (\partial J/\partial c')\underbrace{(\partial c'/\partial b)}_{0}$$

$$\partial J/\partial b = (\partial J/\partial a)\, c + \partial J/\partial b'$$

Similarly

$$\partial J/\partial c = (\partial J/\partial a)\, b + \partial J/\partial c'$$