École Doctorale des Sciences de l'Environnement d’Île-de-France
Année 2008-2009

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

Olivier Talagrand

Cours 7

22 Juin 2009

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 $k = 0, \ldots, K$

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

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

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

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.

Gradient computed by *adjoint method*.

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.

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 $\nabla_u \mathcal{J}$ 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

$$\nabla_u \mathcal{J} \;=\; \boldsymbol{G}'^{\mathrm{T}}\, \nabla_v \mathcal{J} \qquad \text{(A)}$$

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 gain 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

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

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

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

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

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

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

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 ?

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

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

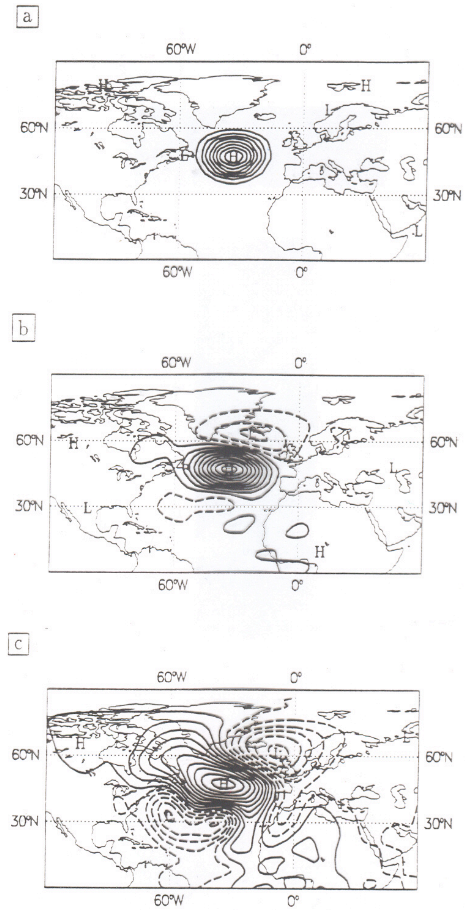
$$\lambda_K = \qquad H_K{}'^{\mathrm{T}}\,R_K^{-1}\,[H_K(\xi_K) - y_K]$$

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

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

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

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.

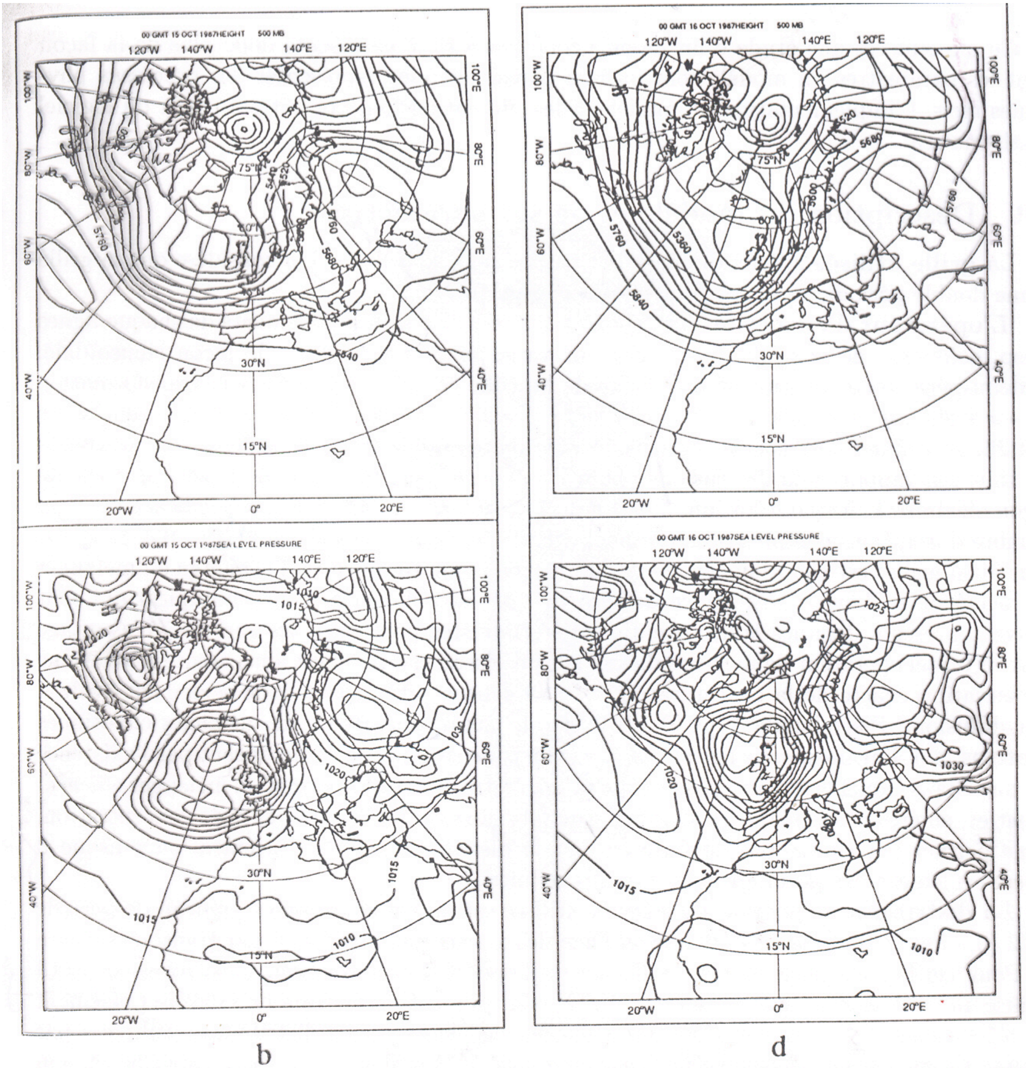
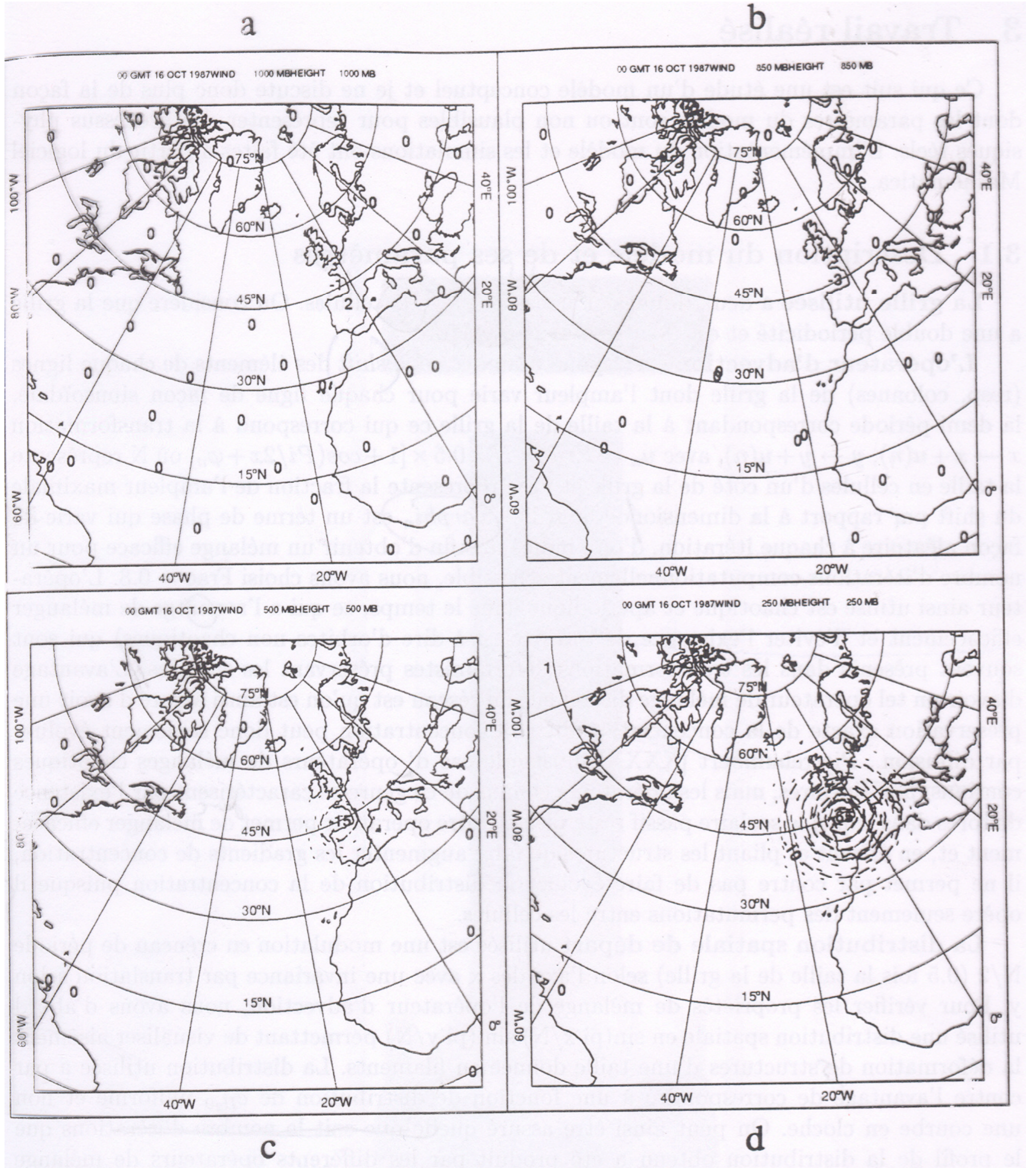


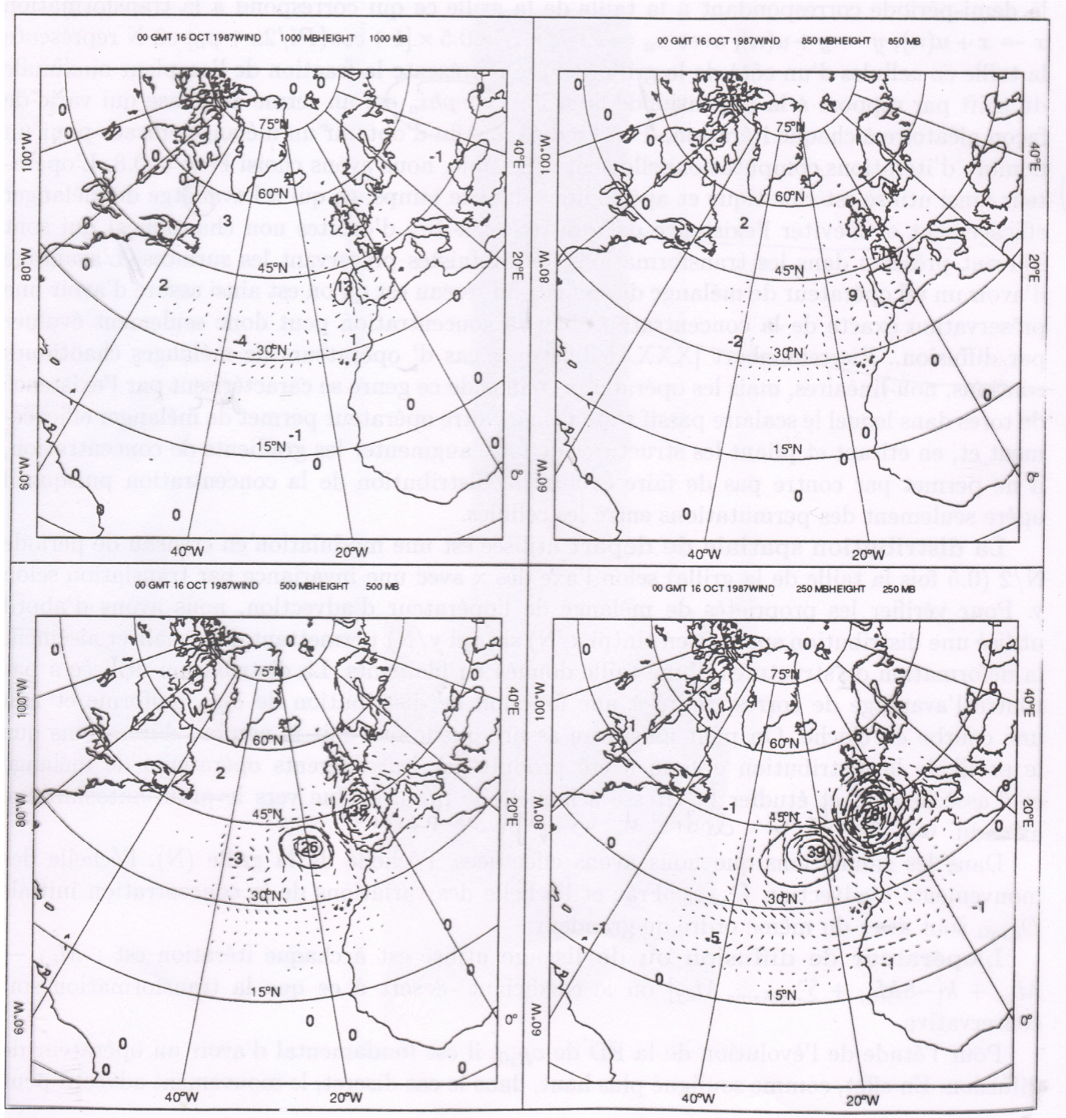
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October 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



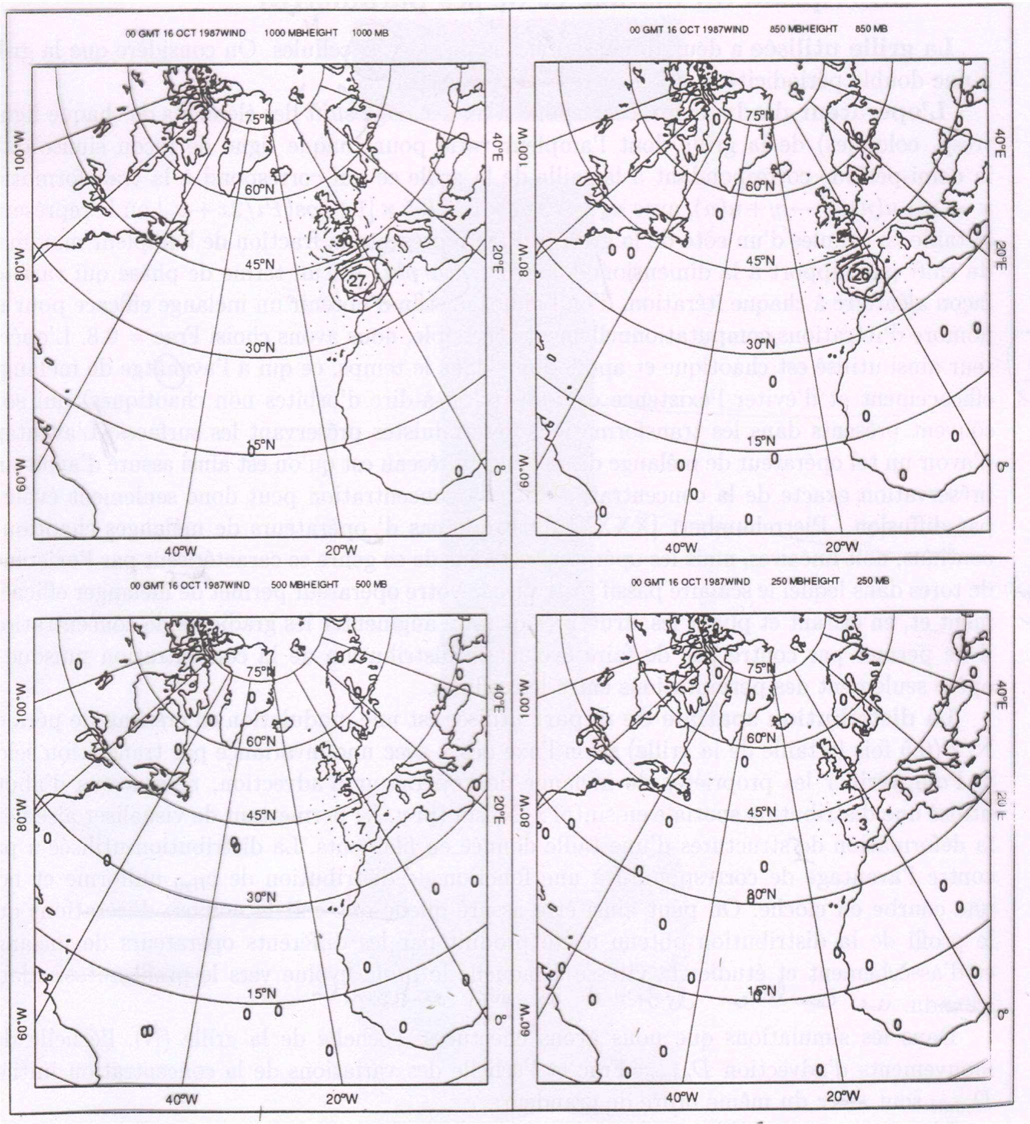
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



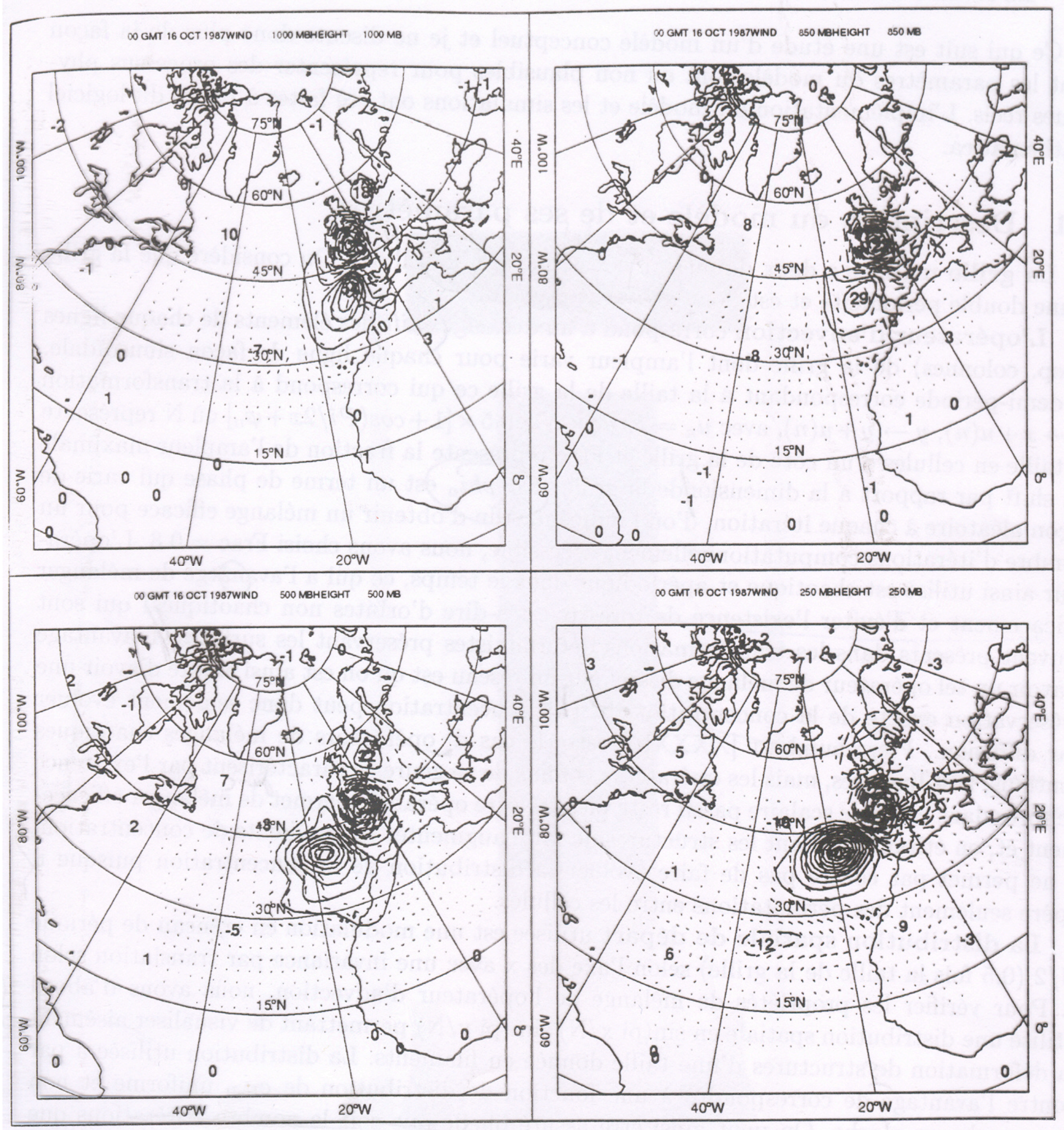
Same as before, but at the end of a 24-hr 4D-Var

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



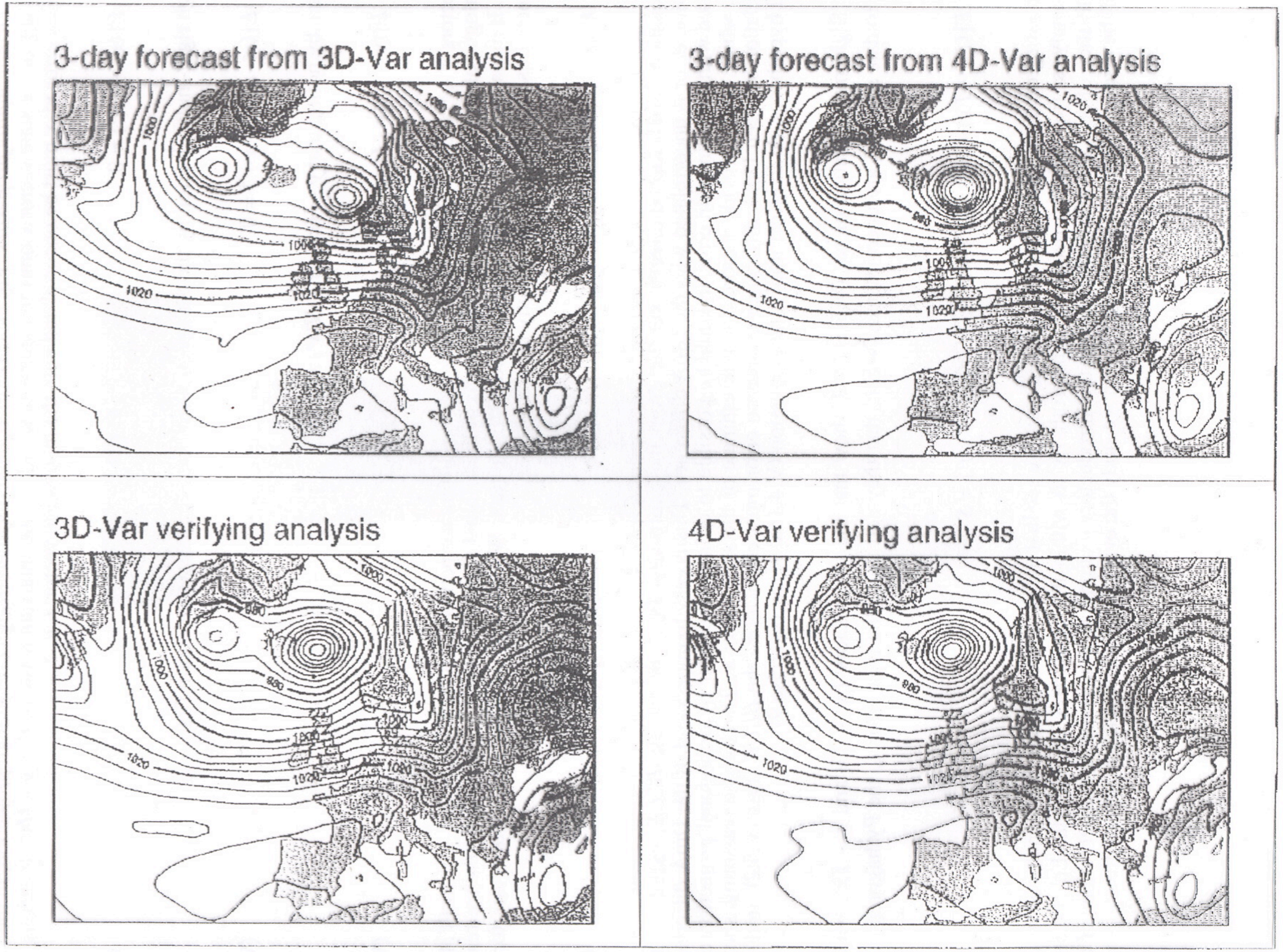
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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)

4D-Var is now used operationally at ECMWF, Météo-France, Meteorological Office (UK), Canadian Meteorological Service (together with an ensemble assimilation system), Japan Meteorological Agency

Model error is ignored

Strong Constraint Variational Assimilation