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

Année Universitaire 2013-2014

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 6

26 Mai 2014

Sequential Assimilation. *Kalman Filter*

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad\qquad k = 0, \ldots, K$$

$$E(\varepsilon_k) = 0 \quad ; \quad E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) = R_k\, \delta_{kj}$$

H_k linear

- Evolution equation

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

$$E(\eta_k) = 0 \quad ; \quad E(\eta_k \eta_j^{\mathrm{T}}) = Q_k\, \delta_{kj}$$

M_k linear

- $E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background $x^b{}_k$ and associated error covariance matrix $P^b{}_k$ known

- Analysis step

$$x^a{}_k = x^b{}_k + P^b{}_k H_k{}^T [H_k P^b{}_k H_k{}^T + R_k]^{-1} (y_k - H_k x^b{}_k)$$
$$P^a{}_k = P^b{}_k - P^b{}_k H_k{}^T [H_k P^b{}_k H_k{}^T + R_k]^{-1} H_k P^b{}_k$$

- Forecast step

$$x^b{}_{k+1} = M_k x^a{}_k$$
$$P^b{}_{k+1} = M_k P^a{}_k M_k{}^T + Q_k$$

Kalman filter (KF, Kalman, 1960)

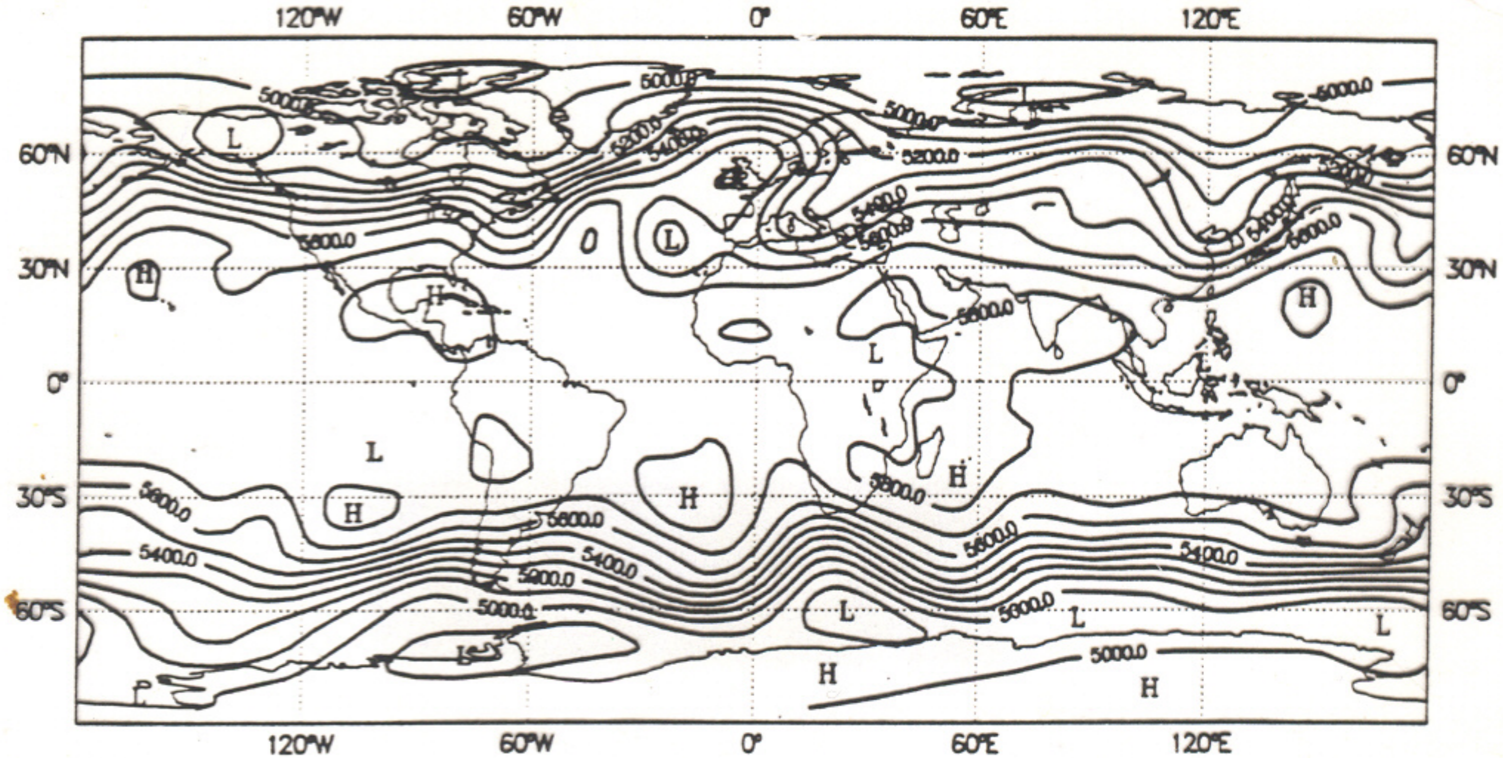
Must be started from some initial estimate $(x^b{}_0, P^b{}_0)$

Costliest part of computation

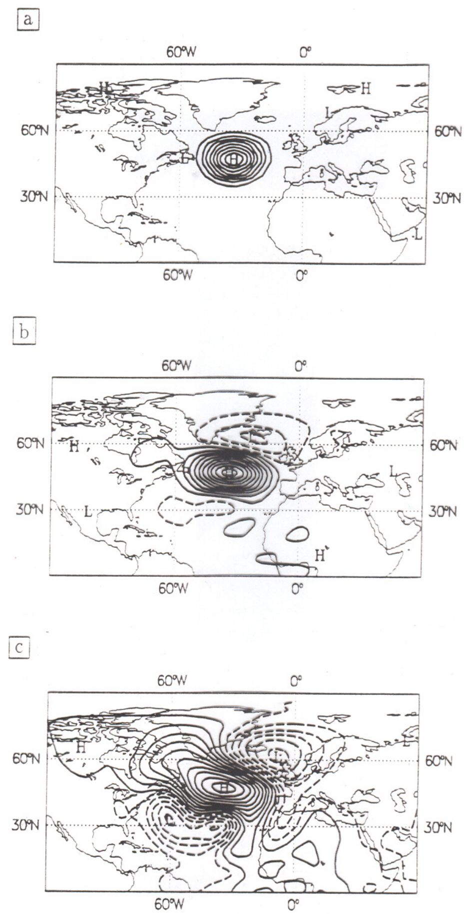
$$P^b{}_{k+1} = M_k P^a{}_k M_k{}^{\mathrm{T}} + Q_k$$

Multiplication by M_k = one integration of the model between times k and $k+1$.
Computation of $M_k P^a{}_k M_k{}^{\mathrm{T}} \approx 2n$ integrations of the model

Need for determining the temporal evolution of the uncertainty on the state of the system is the major difficulty in assimilation of meteorological and oceanographical observations



Analysis of 500-hPa geopotential for 1 December 1989, 00:00 UTC (ECMWF, spectral truncation T21, unit *m*. After F. Bouttier)



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.

Two solutions :

- *Low-rank filters*

  Use low-rank covariance matrix, restricted to modes in state space on which it is known, or at least assumed, that a large part of the uncertainty is concentrated (this requires the definition of a norm on state space).

Reduced Rank Square Root Filters (*RRSQRT*, Heemink)

Singular Evolutive Extended Kalman Filter (*SEEK*, Pham)

….

Second solution :

- *Ensemble filters*

  Uncertainty is represented, not by a covariance matrix, but by an ensemble of point estimates in state space that are meant to sample the conditional probability distribution for the state of the system (dimension $L \approx O(10\text{-}100)$).

  Ensemble is evolved in time through the full model, which eliminates any need for linear hypothesis as to the temporal evolution.

Ensemble Kalman Filter (*EnKF*, Evensen, Anderson, …)

How to update predicted ensemble with new observations ?

Predicted ensemble at time k : $\{x^b_l\}$, $\qquad l = 1, \ldots, L$

Observation vector at same time : $y = Hx + \varepsilon$

- Gaussian approach

Produce sample of probability distribution for real observed quantity Hx

$y_l = y - \varepsilon_l$

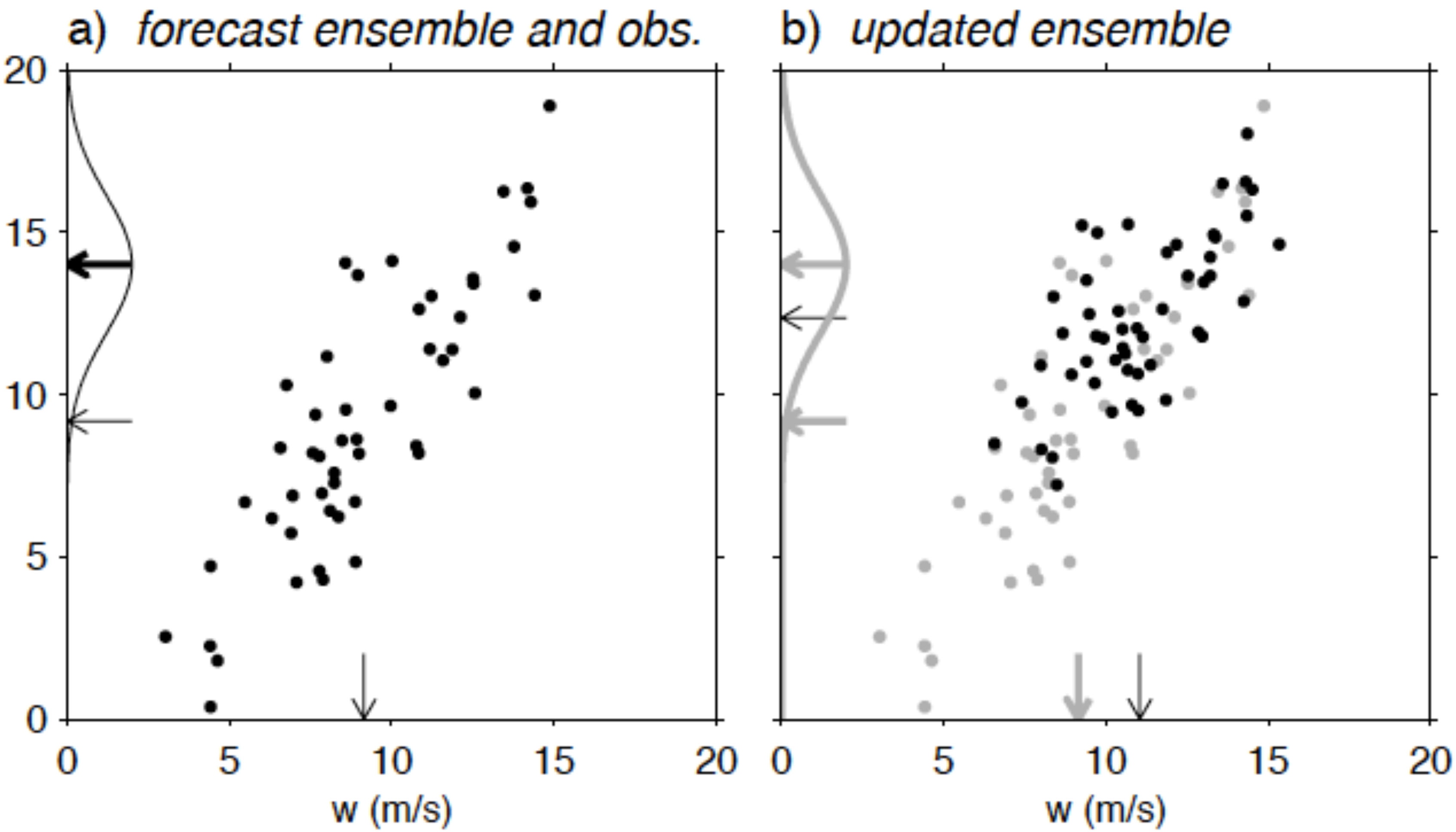
where ε_l is distributed according to probability distribution for observation error ε.

Then use Kalman formula to produce sample of 'analysed' states

$$x^a_l = x^b_l + P^b H^T [HP^bH^T + R]^{-1} (y_l - Hx^b_l) , \qquad l = 1, \ldots, L \qquad (2)$$

where P^b is covariance matrix of predicted ensemble $\{x^b_l\}$.

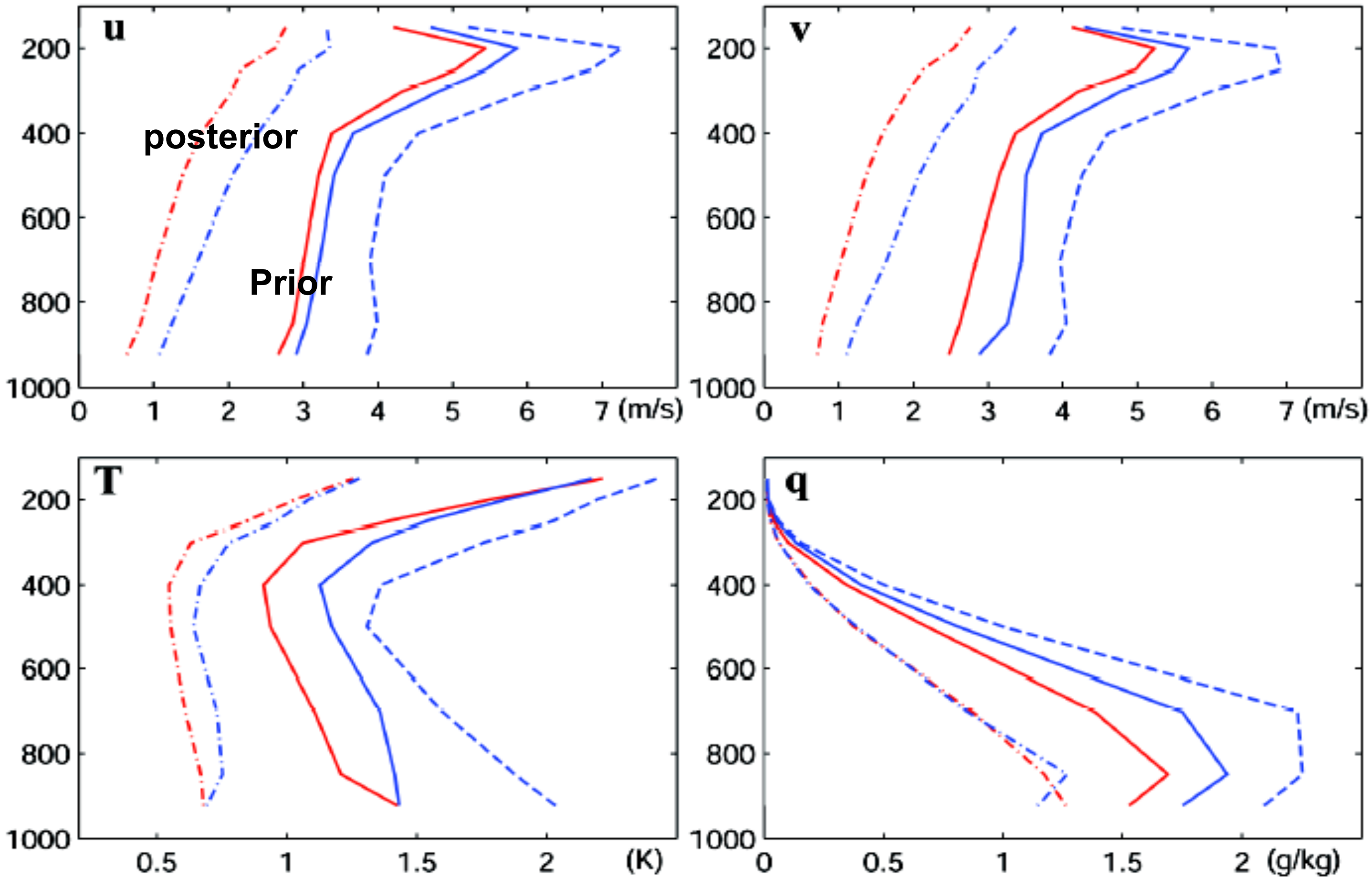
Remark. In case of Gaussian errors, if P^b was exact covariance matrix of background error, (2) would achieve Bayesian estimation, in the sense that $\{x^a_l\}$ would be a sample of conditional probability distribution for x, given all data up to time k.


a) forecast ensemble and obs.
b) updated ensemble
20
15
10
5
0
0
5
10
15
20
w (m/s)
0
5
10
15
20
w (m/s)

C. Snyder

Month-long Performance of EnKF vs. 3Dvar with WRF

— EnKF — 3DVar (prior, solid; posterior, dotted)



Better performance of EnKF than 3DVar also seen in both 12-h forecast and posterior analysis in terms of root-mean square difference averaged over the entire month

(Meng and Zhang 2007c, MWR, in review)

Nonlinear Observation Operators

The cases in which $\mathbf{y} = h(\mathbf{x}) + \epsilon$ or $\mathbf{x}$ is non-Gaussian, or both, can be handled easily.

Let $\hat{h} = (N_e)^{-1} \sum h(\mathbf{x}^i)$. Define $\mathbf{X}$ as before and

$$\mathbf{Y} \quad = \quad \text{matrix with columns } (N_e - 1)^{-1/2} \left(h(\mathbf{x}^i) + \epsilon^i - \hat{h} - \hat{\epsilon} \right),$$

EnKF update is as before, but with nonlinear $h(\mathbf{x}^i)$:

$$\xi^i = \mathbf{x}^i + \hat{\mathbf{K}} \left(\mathbf{y}^o - (h(\mathbf{x}^i) + \epsilon^i) \right)$$

$$\hat{\mathbf{K}} = \mathbf{X}\mathbf{Y}^T \left(\mathbf{Y}\mathbf{Y}^T \right)^{-1}$$

Nonlinear Observation Operators (cont.)

Note relation to the BLUE:

- $\mathbf{X}\mathbf{Y}^T$ is a sample estimate of $\mathbf{P}_{xy} = \mathrm{cov}(\mathbf{x}, \mathbf{y})$
- $\mathbf{Y}\mathbf{Y}^T$ is a sample estimate of $\mathbf{P}_{yy} = \mathrm{cov}(\mathbf{y})$
- $\hat{\mathbf{K}} \rightarrow \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}$ for large N_e

Thus, sample mean converges to the BLUE.

$$\begin{aligned} \hat{\mathbf{x}}^a &= \hat{\mathbf{x}}^f + \hat{\mathbf{K}}\left(\mathbf{y}^o - (\hat{h} + \hat{\epsilon})\right) \\ &\rightarrow E(\mathbf{x}) + \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}(\mathbf{y}^o - E(\mathbf{y})) \end{aligned}$$

EnKF for nonlinear observation operators approximates the BLUE.

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Empirical 'covariance inflation'

- Spurious correlations appear at large geographical distances. Empirical 'localization' (see Gaspari and Cohn, 1999, *Q. J. R. Meteorol. Soc.*)

- In formula

$$x^a{}_l = x^b{}_l + P^b H^T [HP^bH^T + R]^{-1} (y_l - Hx^b{}_l) , \qquad l = 1, \dots, L$$

P^b, which is covariance matrix of an L-size ensemble, has rank L-1 at most. This means that corrections made on ensemble elements are contained in a subspace with dimension L-1. Obviously very restrictive if $L \ll p$, $L \ll n$.

Houtekamer and Mitchell (1998) use two ensembles, the elements of each of which are updated with covariance matrix of other ensemble.

There exist many variants of Ensemble Kalman Filter

Ensemble Transform Kalman Filter (***ETKF***, Bishop *et al.*, *Mon. Wea. Rev.*, 2001)

Requires a prior 'control' analysis x_c^a, emanating from a background x_c^b. An ensemble is evolved about that control without explicit use of the observations (and without feedback to control)

More precisely, define $L \times L$ matrix T such that, given $P^b = ZZ^T$, then $P^a = ZTT^TZ^T$ (not trivial, but possible). Then the background deviations $x^b{}_l - x_c{}^b$ are transformed through $Z \rightarrow ZT$ into an ensemble of analysis deviations $x^a{}_l - x_c{}^a$.

(does not avoid collapse of ensembles)

Local Ensemble Transform Kalman Filter (***LETKF***, Hunt *et al.*, *Physica D*, 2007)

Each gridpoint is corrected only through the use of neighbouring observations.

Other variants of Ensemble Kalman Filter

‘*Unscented*’ *Kalman Filter* (Wan and van der Merve, 2001, Wiley Publishing)

Weighted Kalman Filter (Papadakis *et al*., 2010, *Tellus A*)

Inflation-free Ensemble Kalman Filters (Bocquet and Sakov, 2012, *Nonlin. Processes Geophys.*)

Situation still not entirely clear.

In any case, optimality always requires errors to be independent in time. In order to relax that constraint, it is necessarily to augment the state vector in the temporal dimension.

Bayesian properties of Ensemble Kalman Filter ?

Very little is known.

Le Gland *et al*. (2011). In the linear and gaussian case, the discrete pdf defined by the filter, in the limit of infinite sample size L, tends to the bayesian gaussian pdf.

No result for finite size (note that ensemble elements are not mutually independent)

In the nonlinear case, the discrete pdf tends to a limit which is in general not the bayesian pdf.

Variational Assimilation

Variational approach can easily be extended to time dimension.

Suppose for instance available data consist of

- Background estimate at time 0

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

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

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

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

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

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

Then objective function

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

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

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

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

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.

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 $\nabla_{\boldsymbol{u}}\mathcal{J}$ 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

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

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

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

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

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

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

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

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

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 ?

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

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

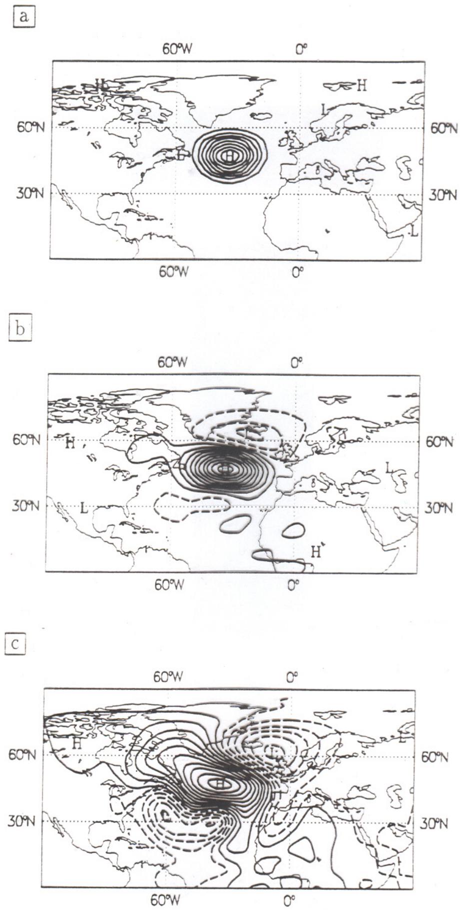
$$\lambda_K = H_K'^T R_K^{-1} [H_K(\xi_K) - y_K]$$

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

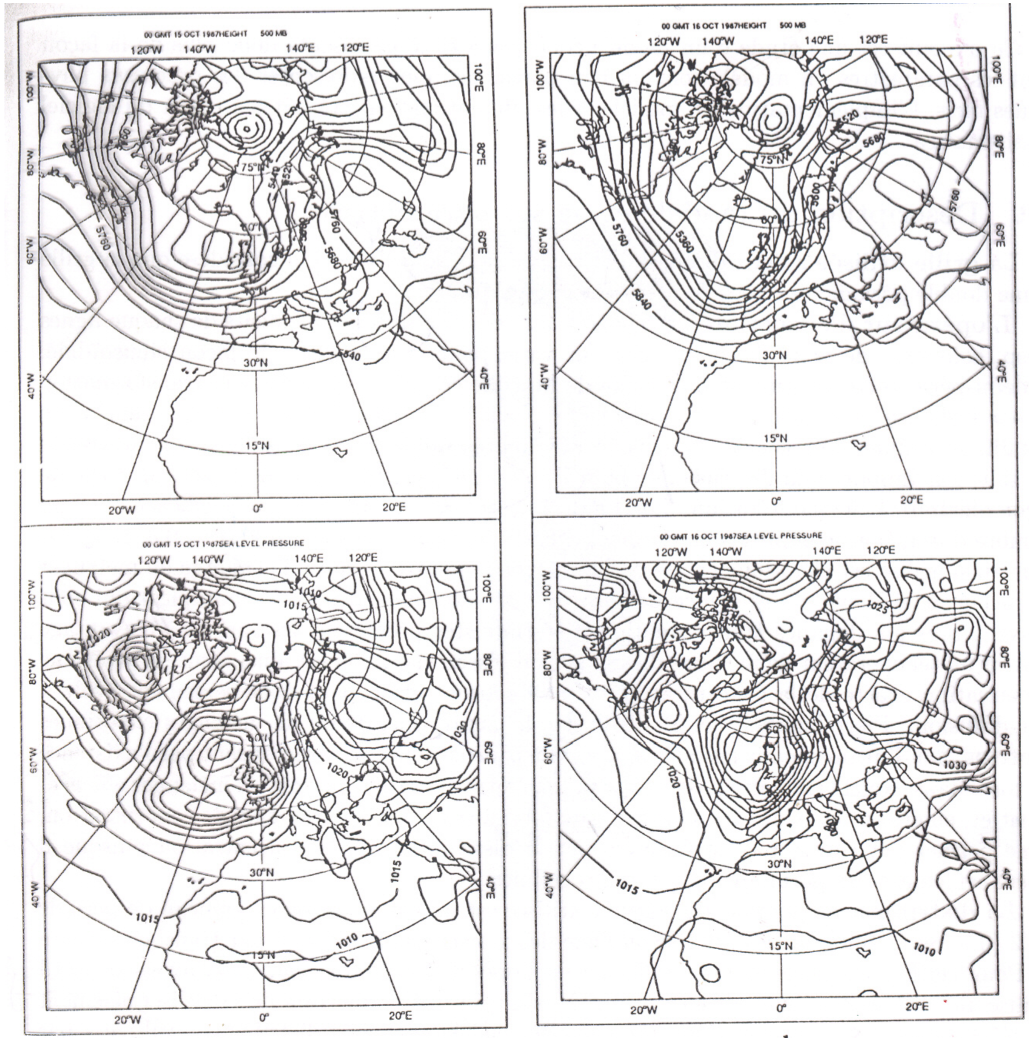
$$\lambda_0 = M_0'^T \lambda_1 + H_0'^T R_0^{-1} [H_0(\xi_0) - y_0] + [P_0^b]^{-1} (\xi_0 - x_0^b)$$

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

Not approximate (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.

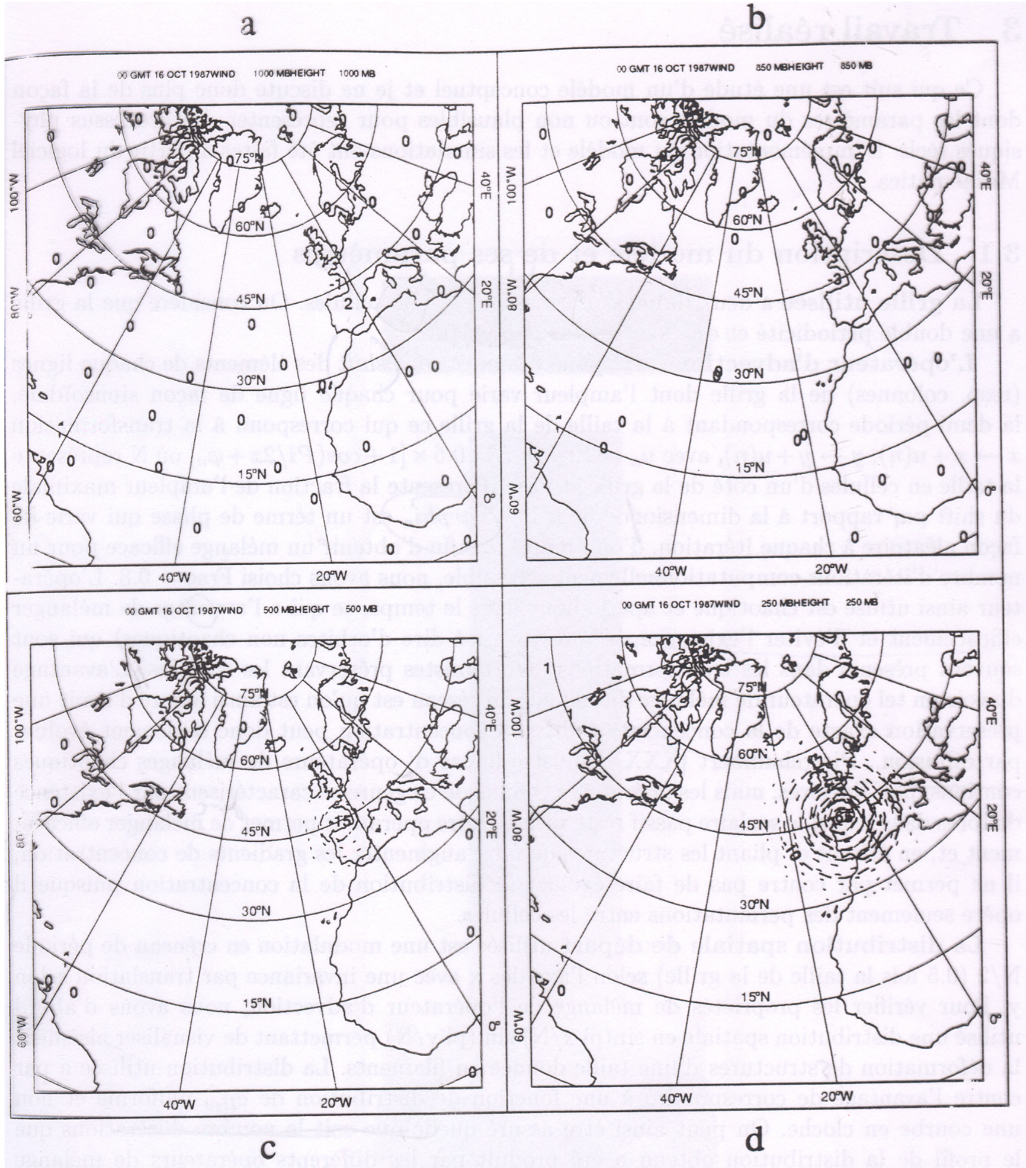


b

d

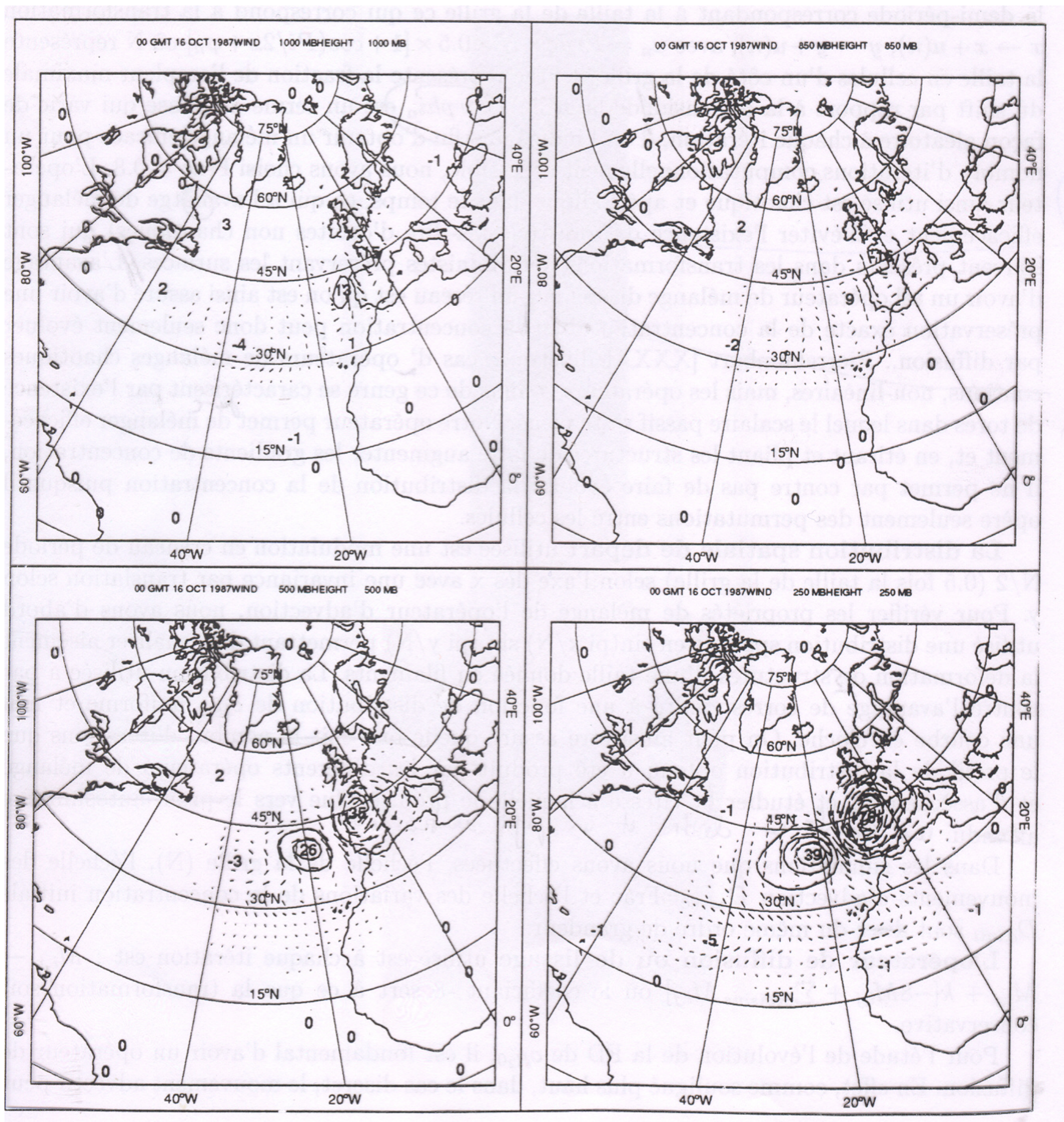
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



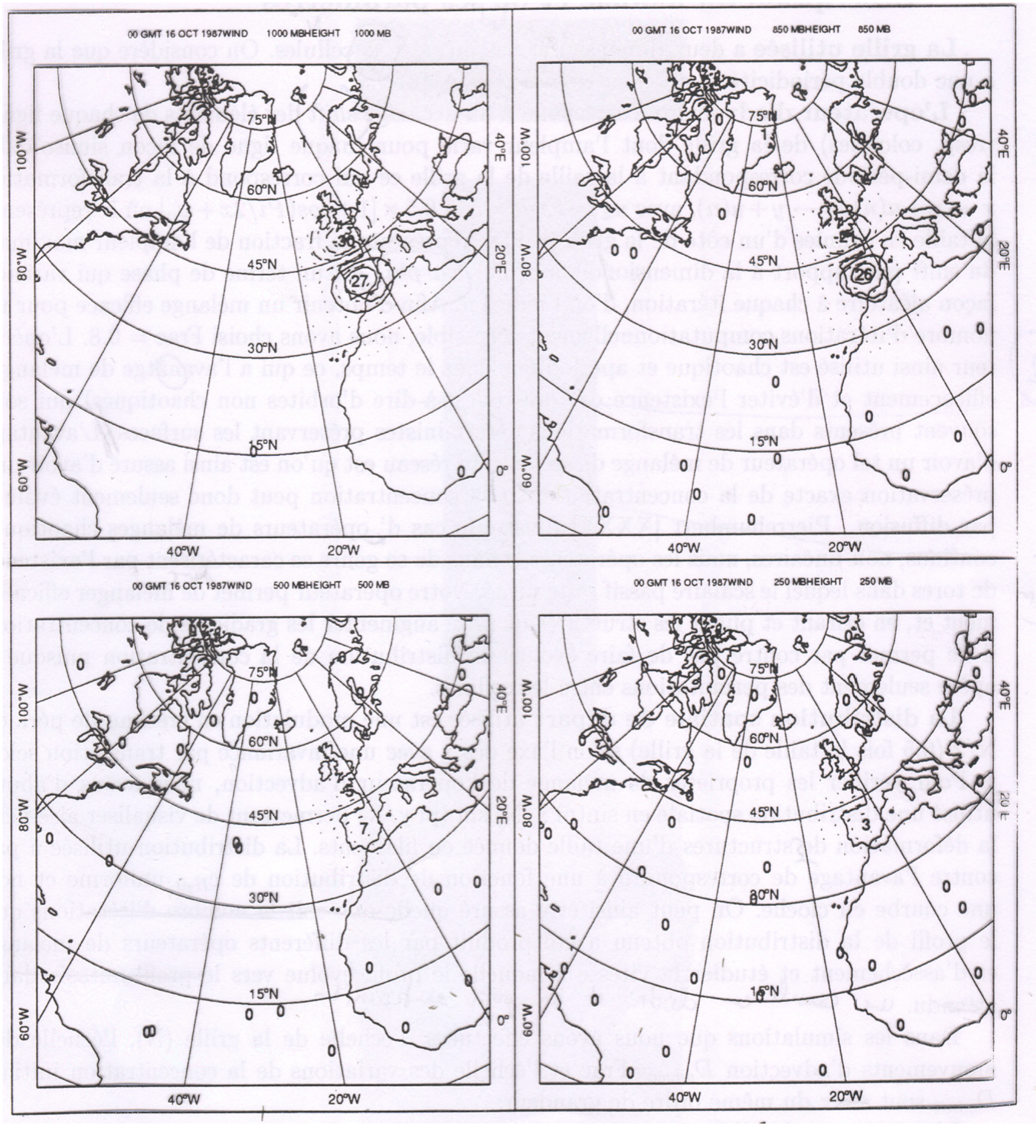
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



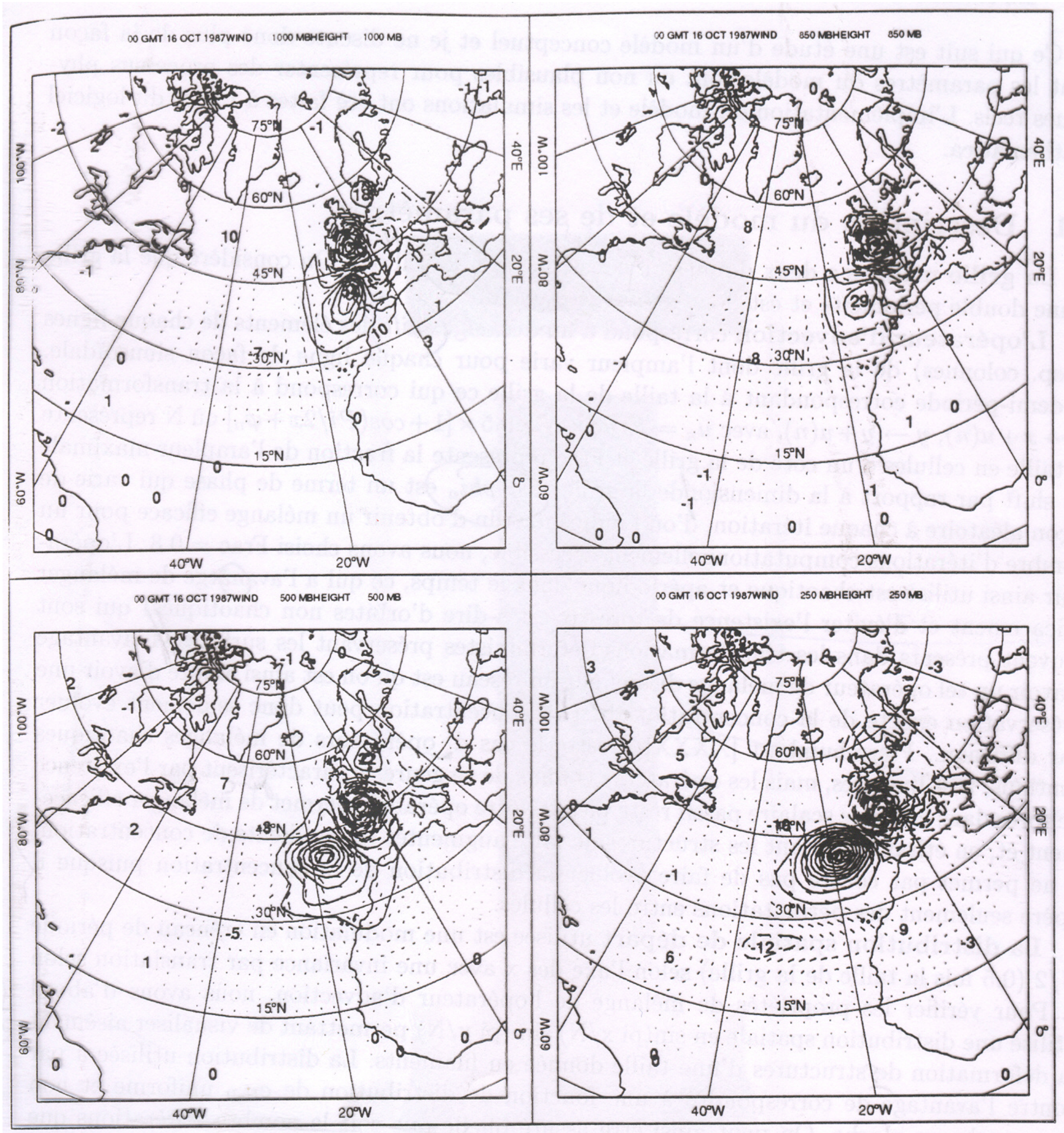
Same as before, but at the end of a 24-hr 4D-Var

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



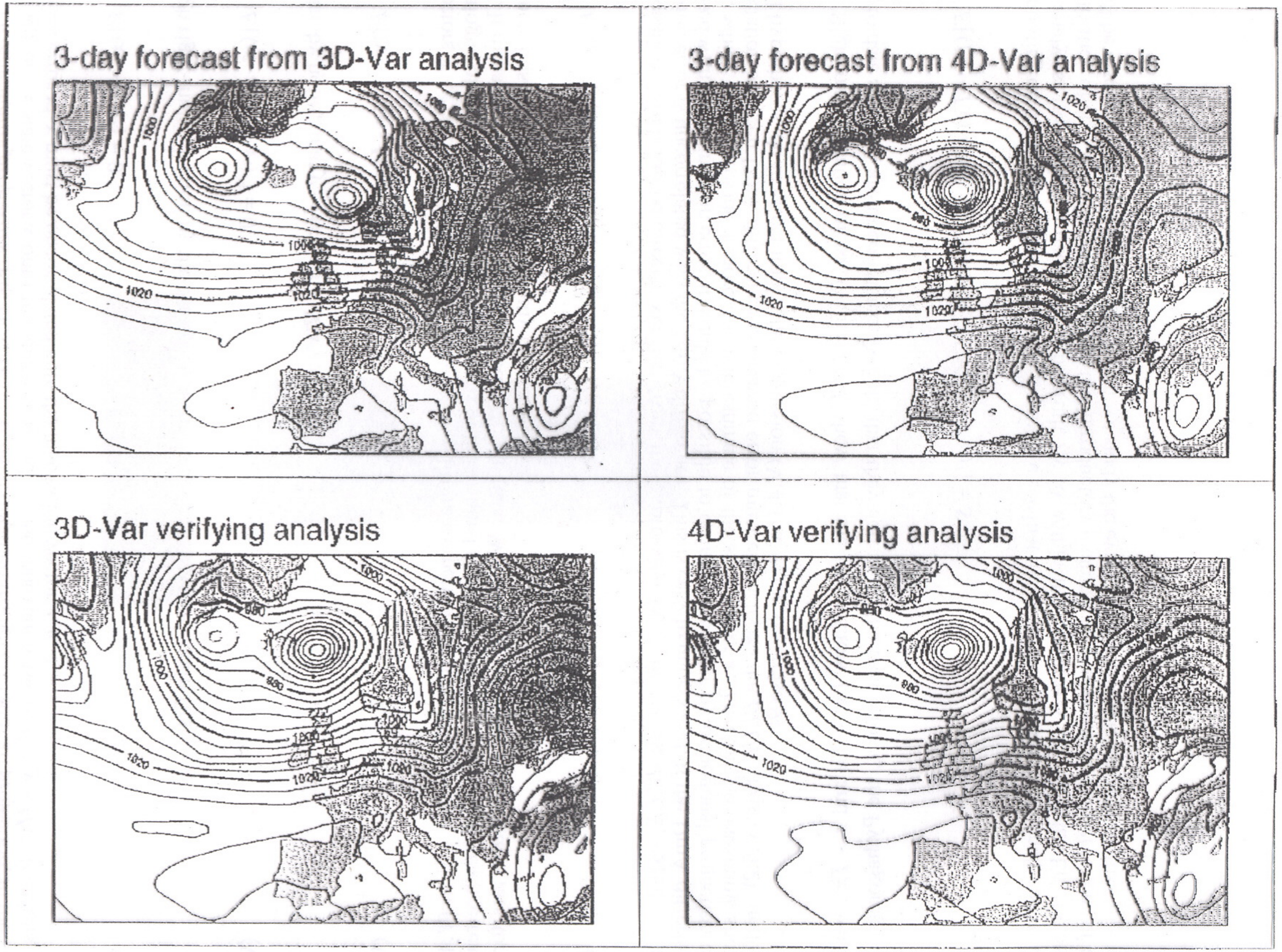
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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.