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

Année Universitaire 2015-2016

# Modélisation Numérique de l’Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 6

18 Mai 2016

**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

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

- Forecast step

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

*Kalman filter* (KF, Kalman, 1960)

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

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, …)

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Collapse originates in the fact that gain matrix $P^b H^T [HP^bH^T + R]^{-1}$ is nonlinear wrt background error matrix $P^b$, resulting in a systematic sampling effect. Solution : empirical 'covariance inflation'.

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

- In formula

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

$P^b$, 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 $L \ll p$ , $L \ll n$. Localisation, in addition to eliminating spurious long-range correlations, increases the rank of the gain matrix.

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 $L$ x $L$ 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.*)

*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.

Situation still not entirely clear.

In any case, Kalman Filter propagates information only forward in time, and 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.

Two questions

- *How to propagate information backwards in time ?* (useful for reassimilation of past data)

- *How to take into account possible dependence in time ?*

Kalman Filter, whether in its standard linear form or in its Ensemble form, does neither.

## 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

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

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

# Conclusion

*Sequential assimilation, in which data are processed by batches, the data of one batch being discarded once that batch has been used, cannot be optimal if data in different batches are affected with correlated errors.* ***This is so even if one keeps trace of the correlations.***

## Solution

Process all correlated in the same batch (4DVar, some smoothers)

**Variational Assimilation**

Variational form of the *BLUE*

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

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

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

$$\equiv \qquad \mathcal{J}_b \qquad + \qquad \mathcal{J}_o$$

'*3D-Var*'

Can easily, and heuristically, be extended to the case of a nonlinear observation operator $H$.

Used operationally in USA, Australia, China, …

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

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

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

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

$\ldots.$

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

$\ldots.$

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

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

Result of direct integration $(\xi_k)$, 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

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

....

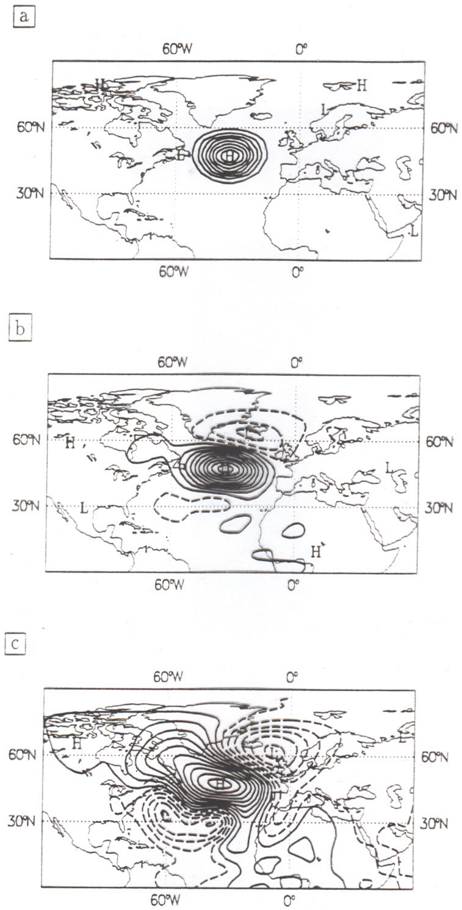
$$\lambda_k = M_k'^T \lambda_{k+1} + H_k'^T R_k^{-1}\,[H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$$

....

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

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

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.

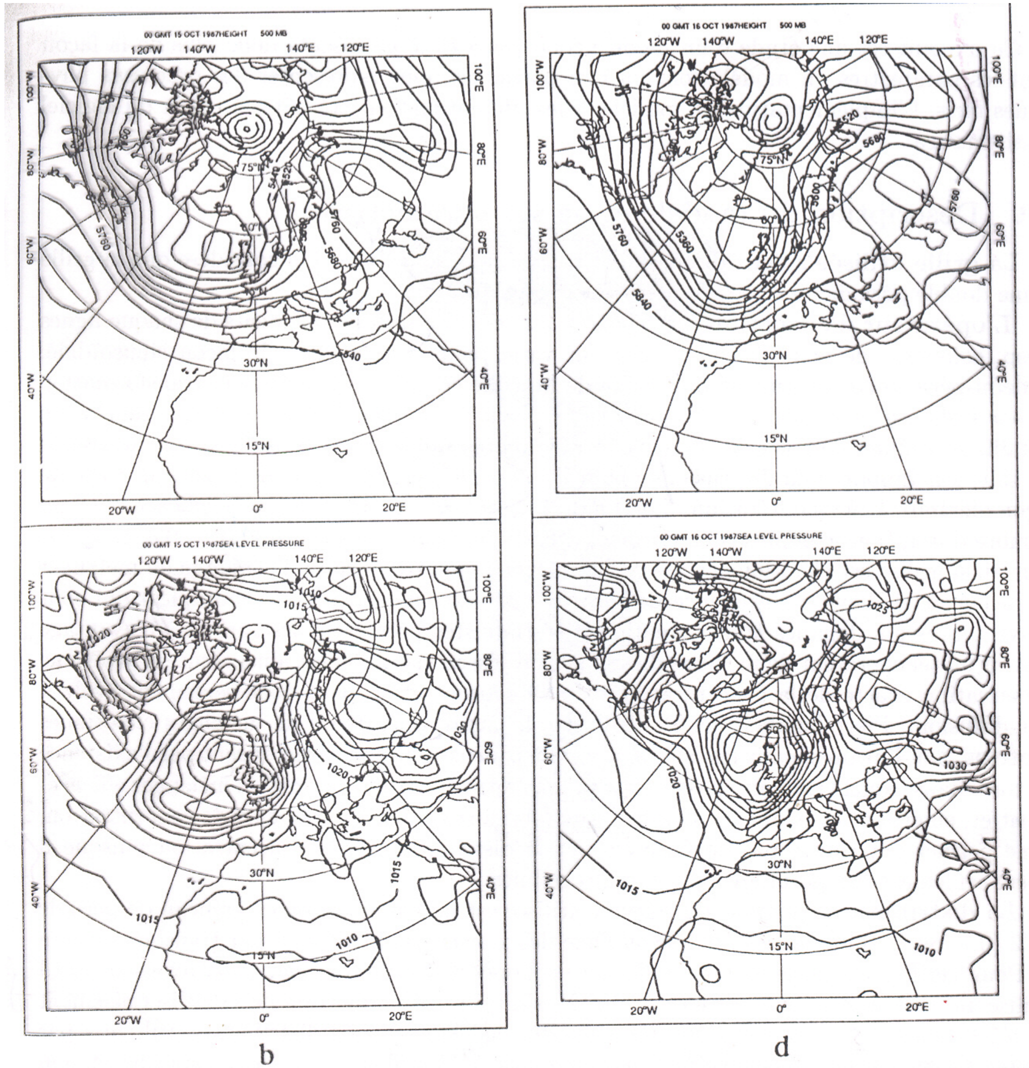
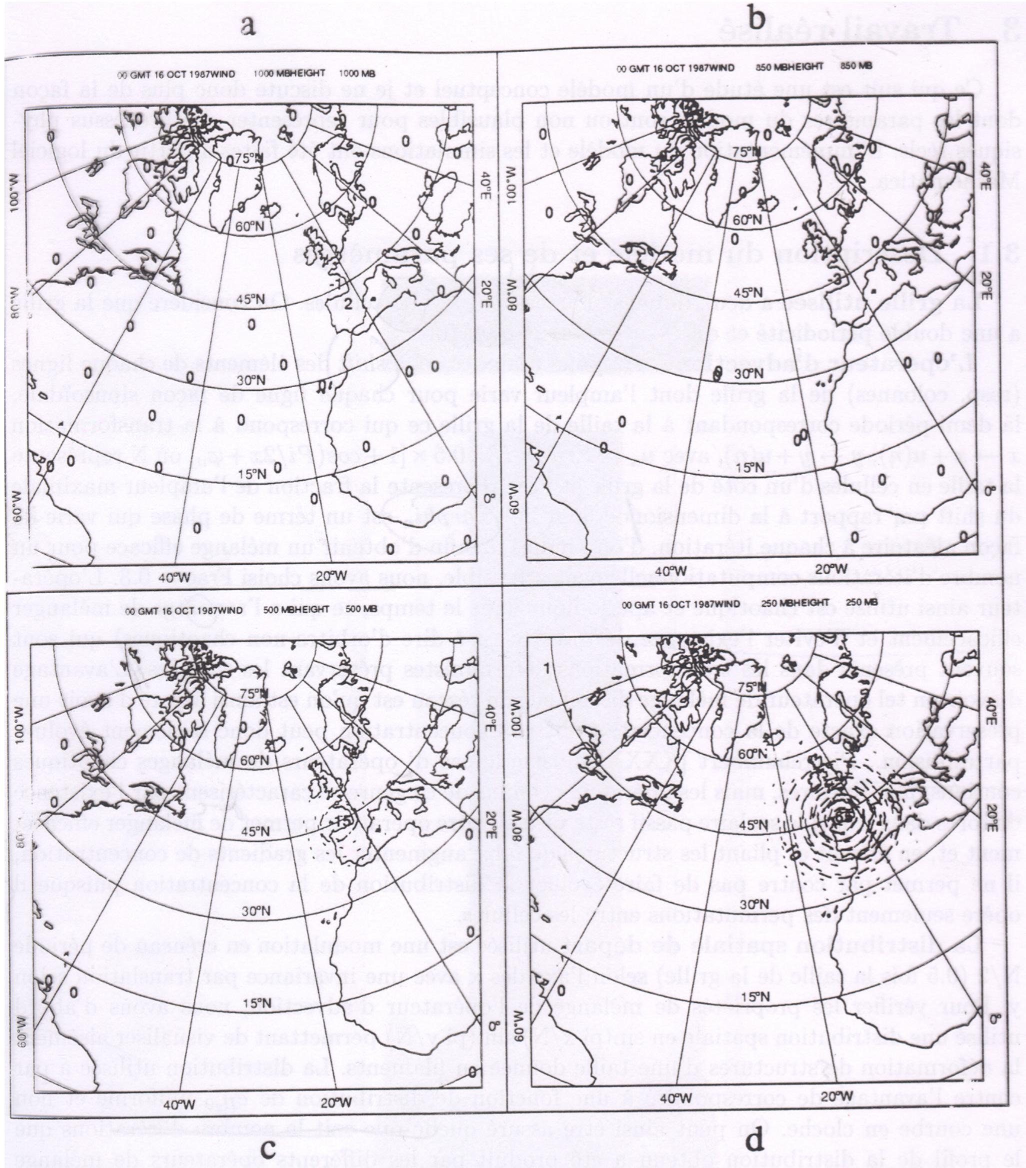


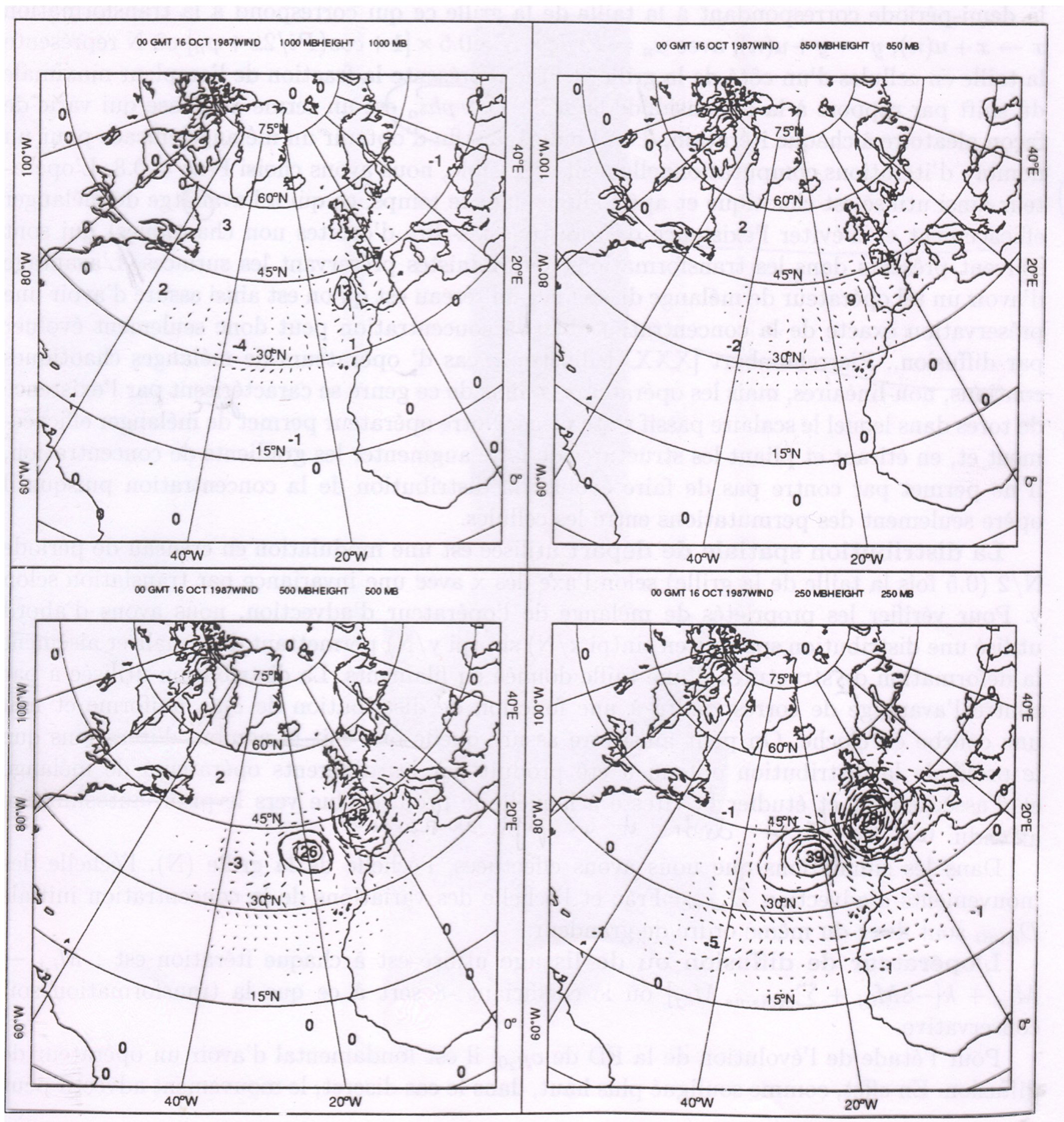

FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



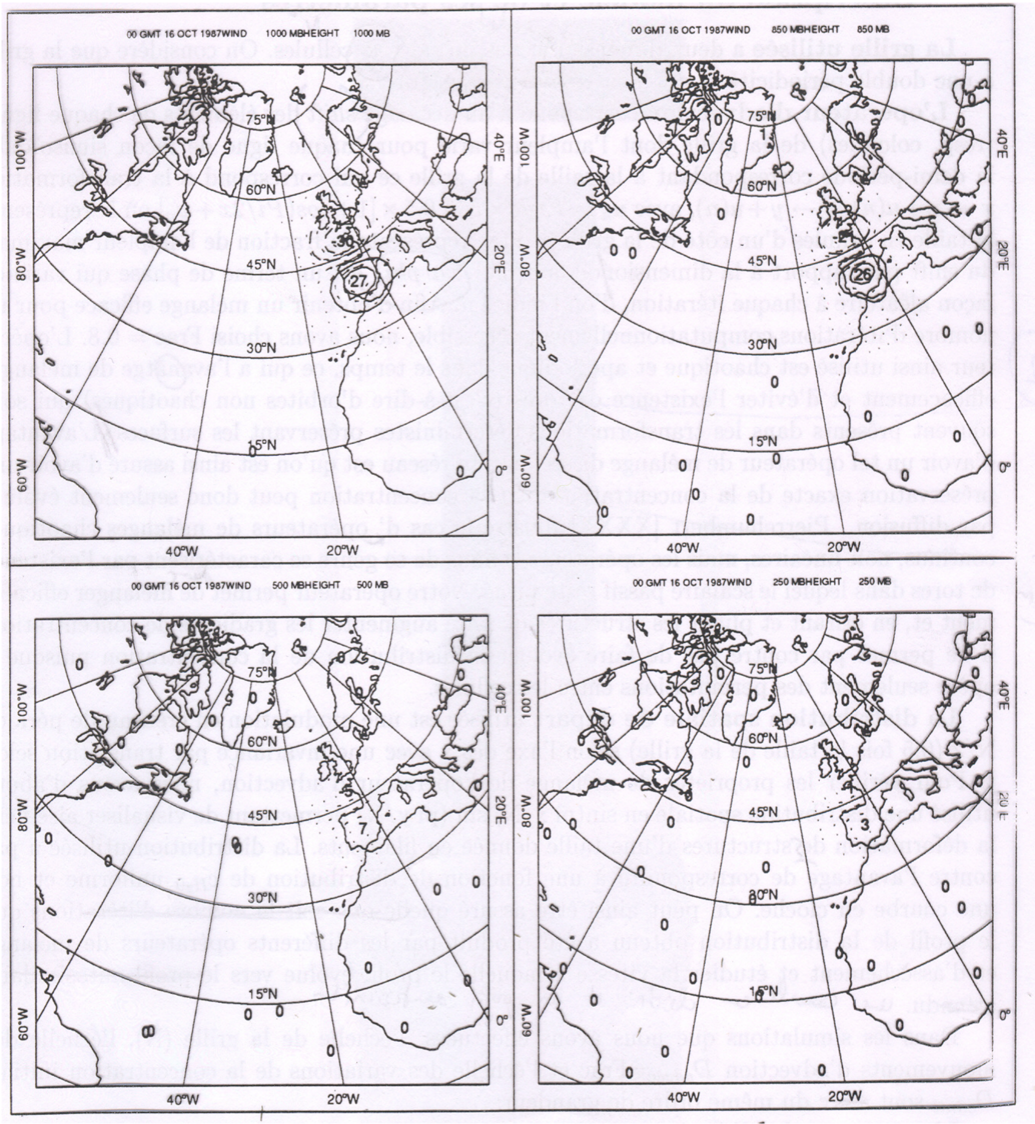

Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



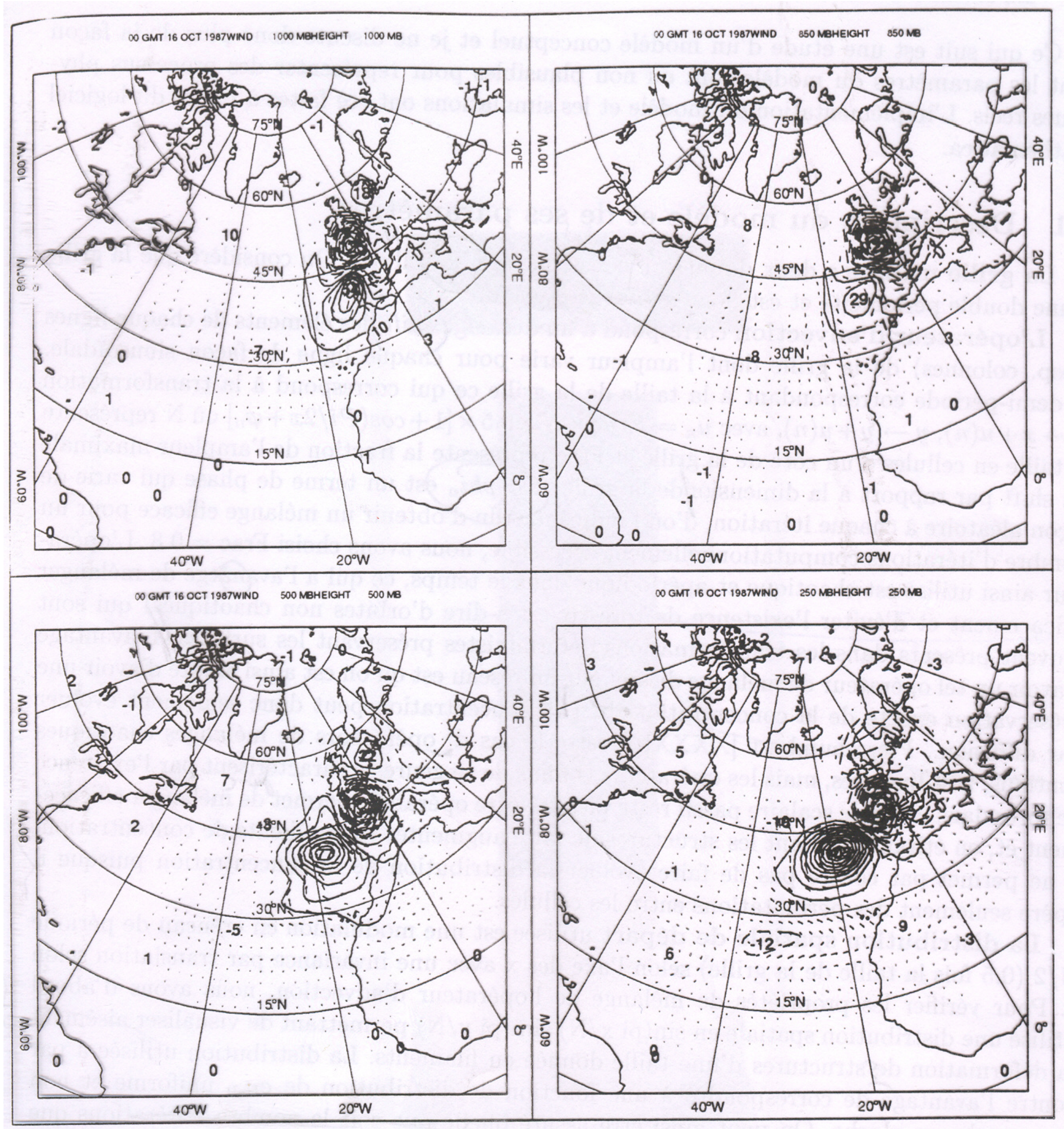

Same as before, but at the end of a 24-hr 4D-Var

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



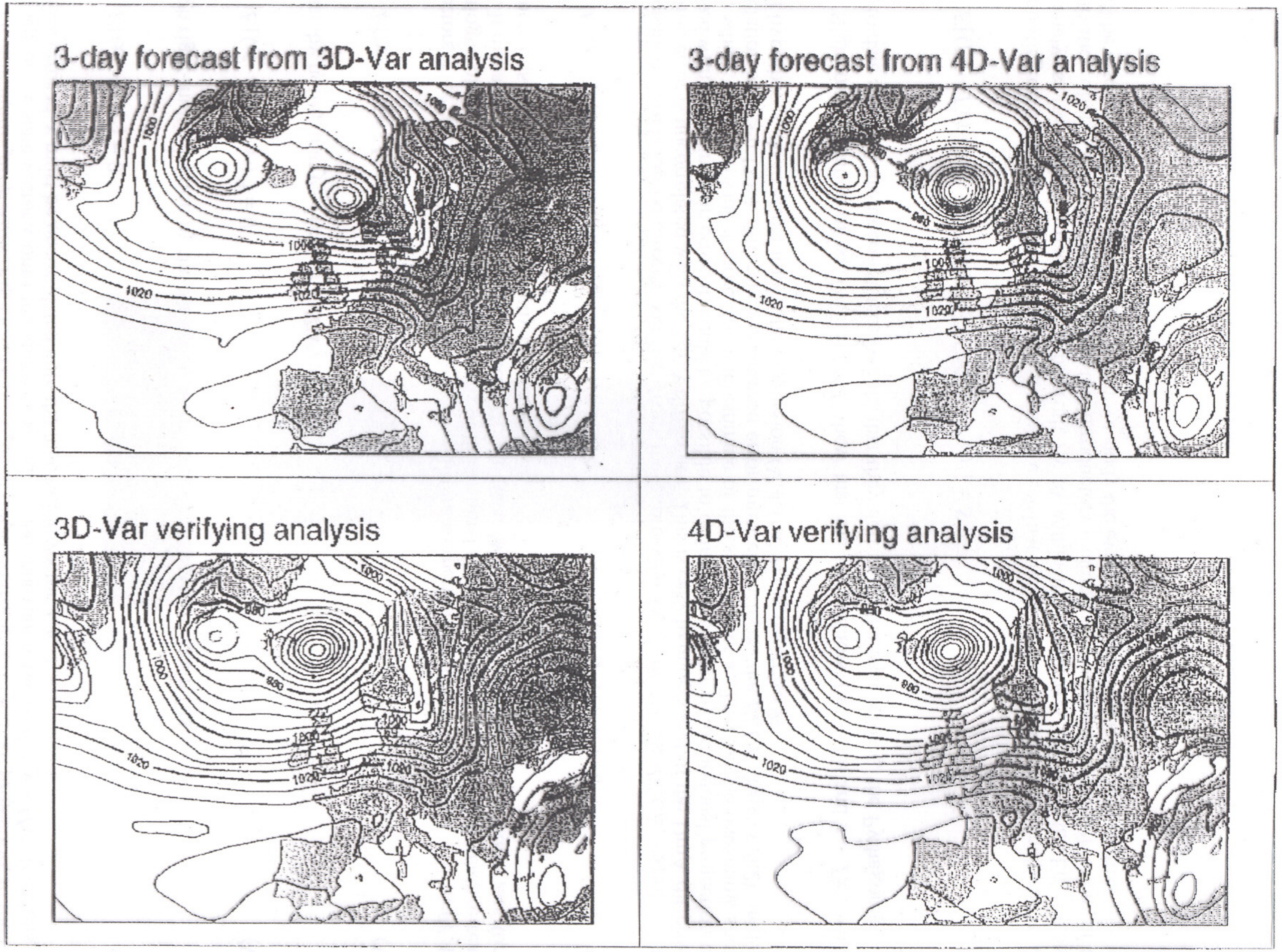

Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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.

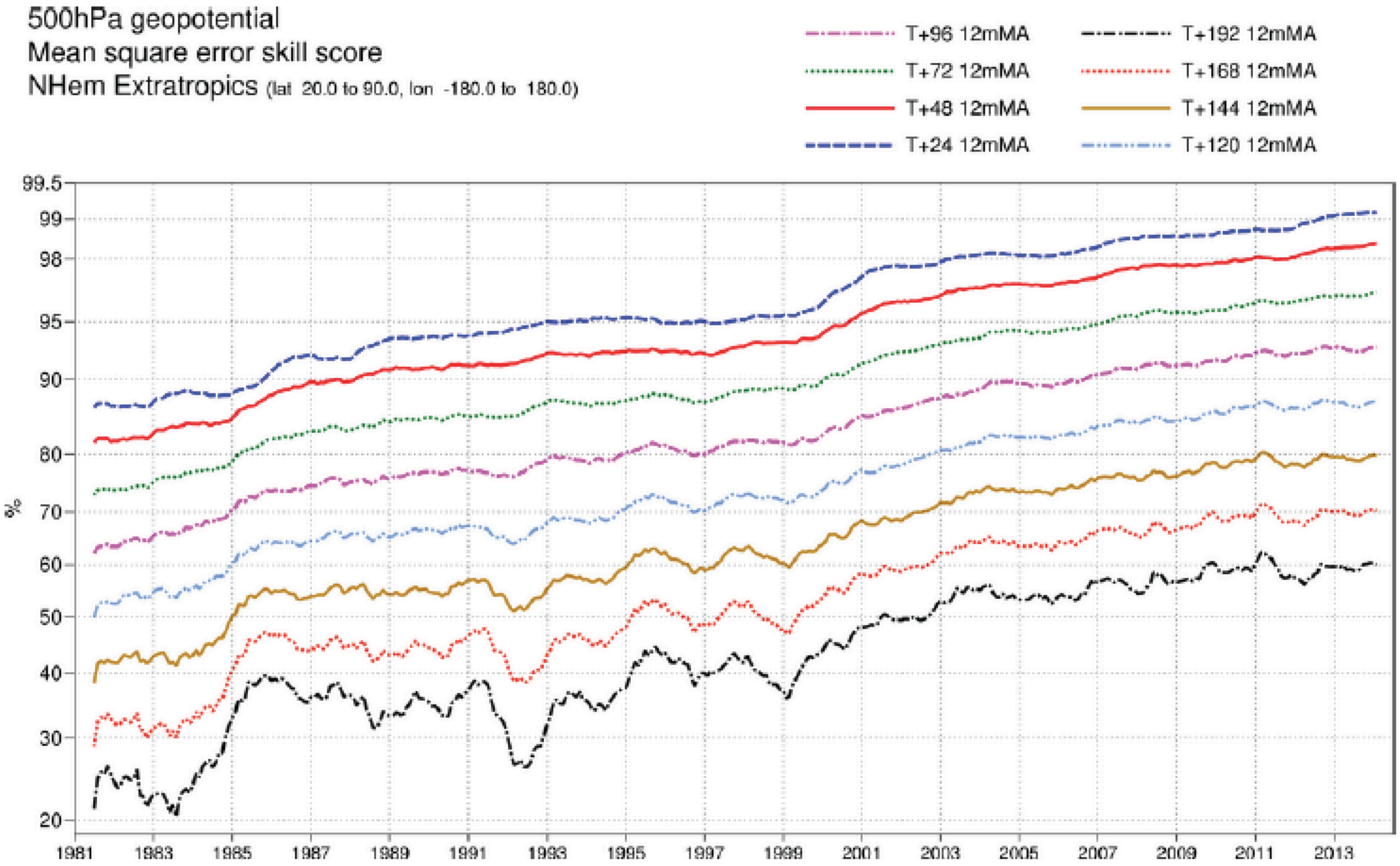
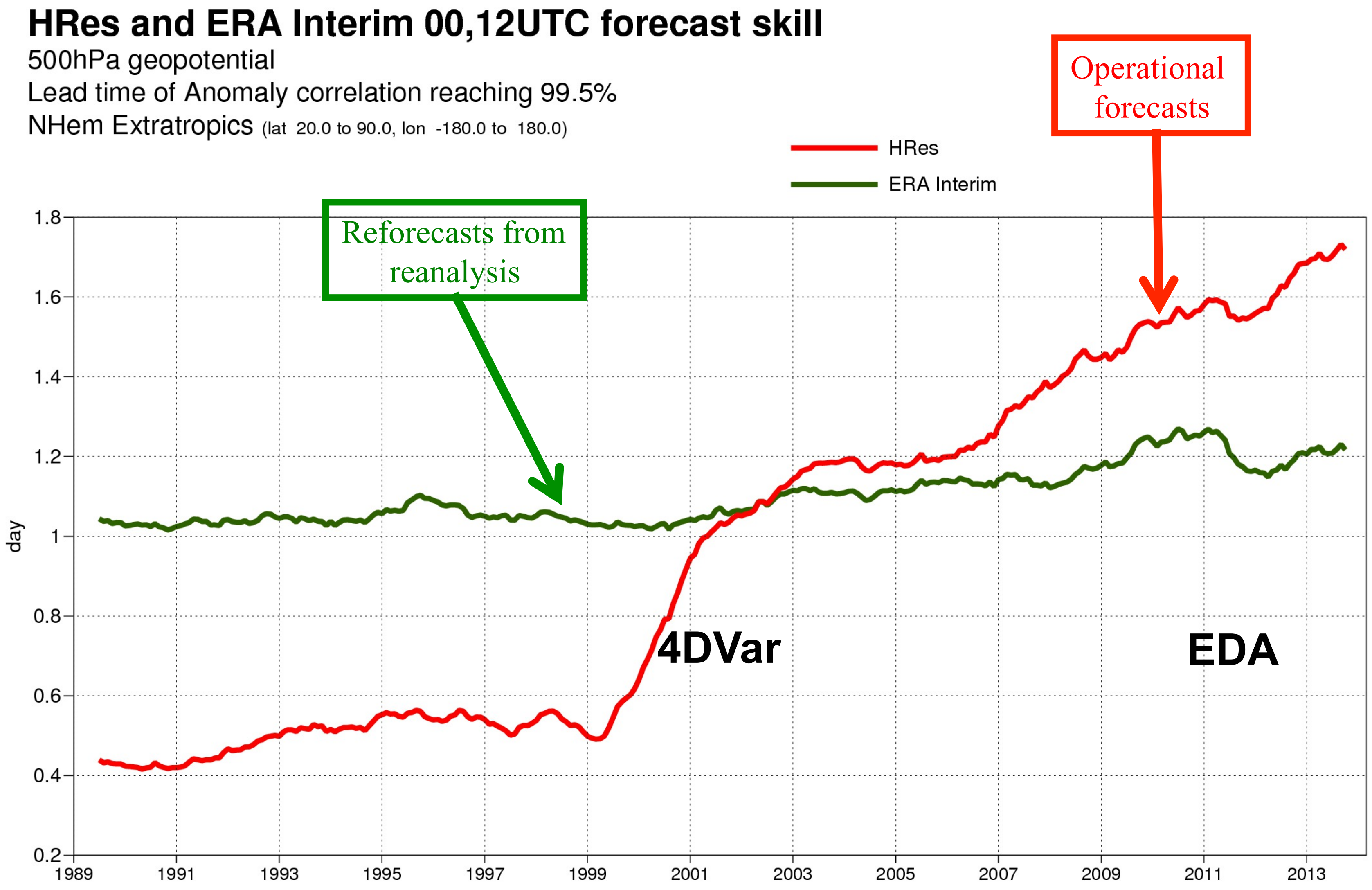



**Figure 3:** 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

Persistence = 0 ; climatology = 50 at long range

# Initial state error reduction




Credit E. Källén, ECMWF

## 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

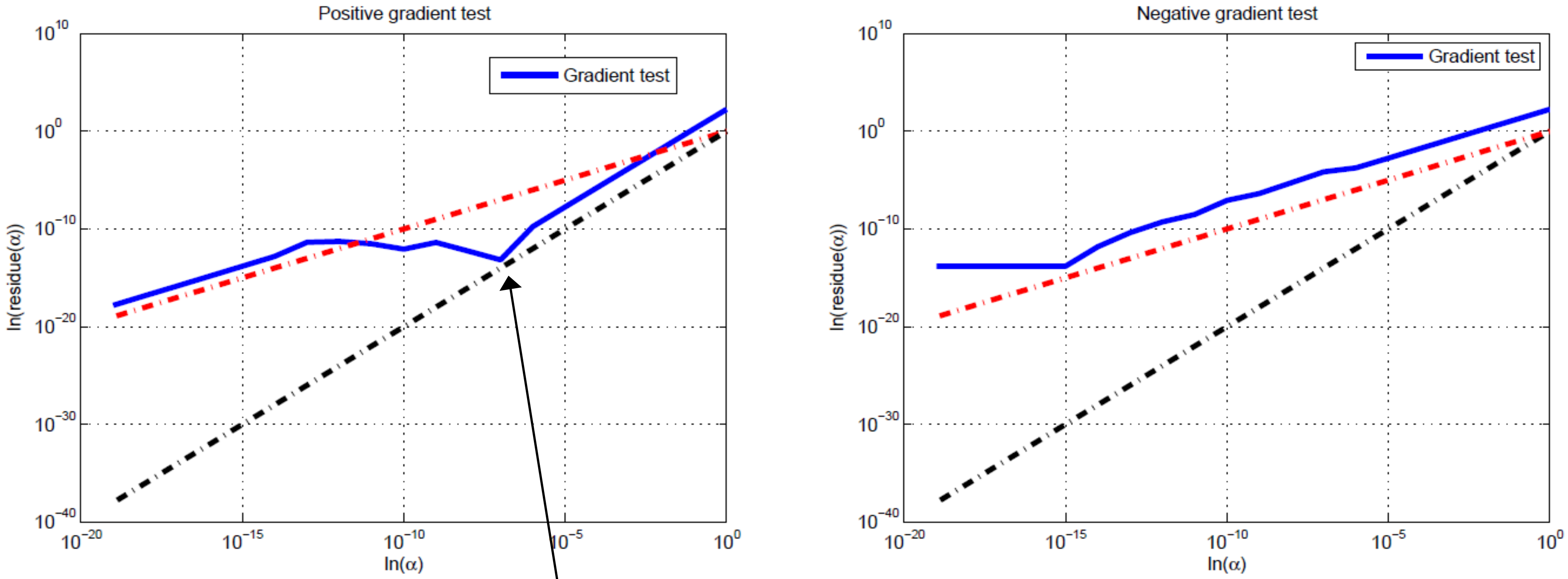
$$\partial J/\partial b = (\partial J/\partial a)\underset{c}{(\partial a/\partial b)} + (\partial J/\partial b')\underset{1}{(\partial b'/\partial b)} + (\partial J/\partial c')\underset{0}{(\partial c'/\partial b)}$$

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

Similarly

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

# Gradient test




$\epsilon \cdot \mathfrak{J}(\text{optimal control variable})$

$\epsilon = 2^{-53}$ zero machine

$$residue(\alpha) = (\mathfrak{J}(x + \alpha dx) - \mathfrak{J}(x)) - \alpha \nabla \mathfrak{J}(x) dx$$

M. Jardak