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

Année Universitaire 2017-2018

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 6

24 Mai 2018

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.

Kalman smoother

Propagates information both forward and backward in time, as does 4DVar, but uses Kalman-type formulæ

Various possibilities

- Define new state vector $x^T \equiv (x_0^T, \ldots, x_K^T)$
  and use Kalman formula from a background x_b and associated covariance matrix Π_b.
  Can take into account temporal correlations

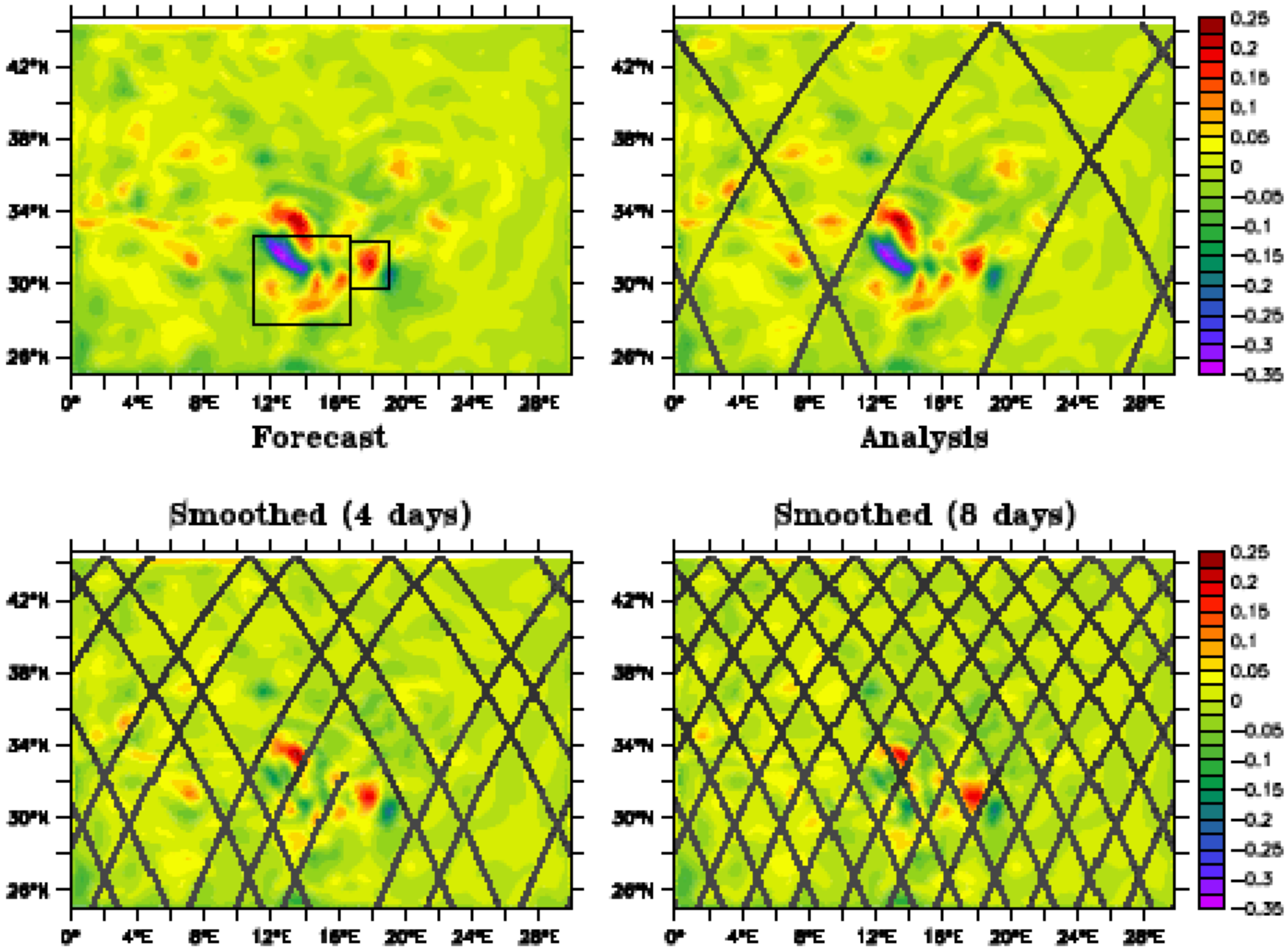
- Update sequentially vector $(x_0^T, \ldots, x_k^T)^T$ for increasing k
  Cannot take into account temporal correlations

Algorithms exist in ensemble form

E. Cosme (1995)

Ensemble smoother based on *Singular Evolutive Extended Kalman Filter* (*SEEK*)

Of second type above. Retropropagates corrections on fields backwards in time, but without modifying relative weights given to previous data, *i.e.* cannot be optimal in case of temporal dependence between errors.



E. Cosme, HDR, 2015, Lissage d'ensemble SEEK

Données synthétiques

FIGURE 3.6 – Evolution du champ d'erreur en SSH du jour 38, au cours des étapes d'analyse successives. En haut à gauche : prévision du filtre ; en haut à droite : analyse du filtre. Les observations utilisées pour cette analyse sont distribuées le long des traces grises. En bas à gauche : analyse du lisseur après introduction des observations des jours 40 et 42 ; En bas à droite : analyse du lisseur après introduction des observations des jours 40 à 46.

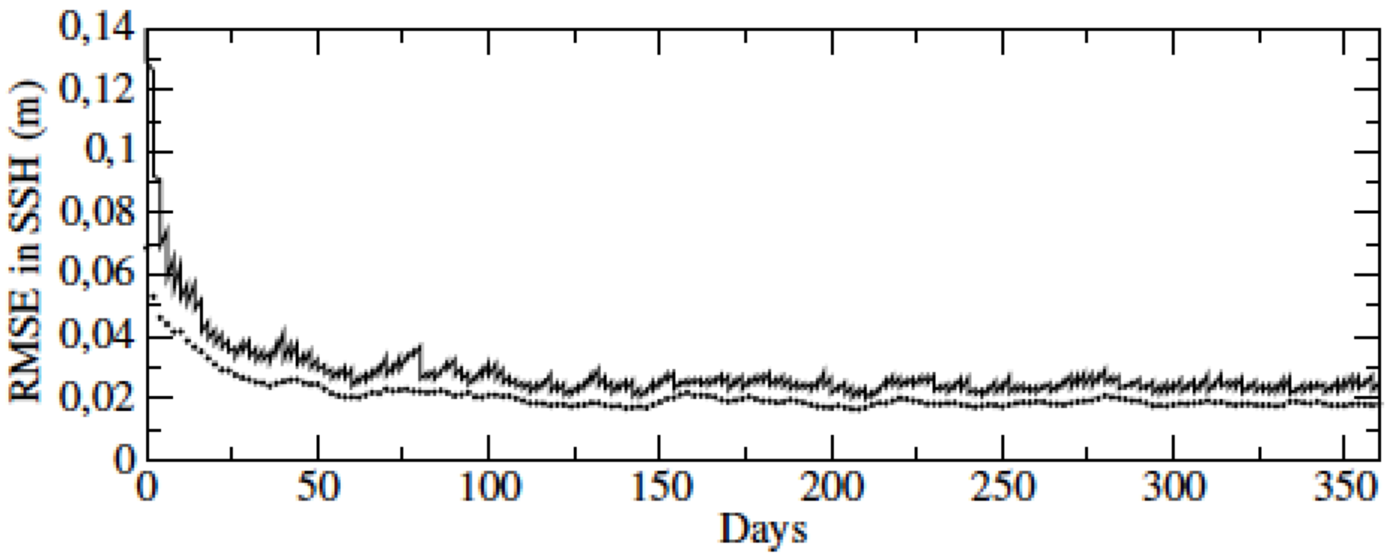


FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse) ; Points : lisseur à retard fixe de 8 jours.

E. Cosme, HDR, 2015, Lissage d'ensemble SEEK

Bayesian Estimation

Data of the form

$$z = \Gamma x + \zeta, \qquad\qquad \zeta \sim \mathcal{N}[0, S]$$

Known data vector z belongs to *data space* $\mathcal{D}$, $dim\mathcal{D} = m$,
Unknown state vector x belongs to *state space* $\mathcal{X}$, $dim\mathcal{X} = n$
Γ known (mxn)-matrix, ζ unknown 'error'

Probability that $x = \xi$ given ? $\quad x = \xi \Rightarrow \zeta = z - \Gamma\xi$

$$\mathrm{P}(\zeta = z - \Gamma\xi) \propto \exp[\,-(z - \Gamma\xi)^{\mathrm{T}} S^{-1} (z - \Gamma\xi)/2\,] \propto \exp[\,-(\xi - x^a)^{\mathrm{T}} (P^a)^{-1} (\xi - x^a)/2\,]$$

where

$$x^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1} \Gamma^{\mathrm{T}} S^{-1} z$$
$$P^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1}$$

Then conditional probability distribution is

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

Bayesian Estimation (continuation 1)

$$z = \Gamma x + \zeta, \qquad \zeta \sim \mathcal{N}[0, S]$$

Then

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

with

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

Determinacy condition : rank$\Gamma = n$. Data contain information, directly or indirectly, on every component of state vector x. Requires $m \geq n$.

Variational form

$$P(x \mid z) \propto \exp[\, -(z - \Gamma\xi)^T S^{-1} (z - \Gamma\xi)/2 \,] \propto \exp[\, -(\xi - x^a)^T (P^a)^{-1} (\xi - x^a)/2 \,]$$

Conditional expectation x^a minimizes following scalar *objective function*, defined on state space $\mathcal{X}$

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

$$P^a = [\partial^2 \mathcal{J} / \partial\xi^2]^{-1}$$

If data still of the form

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

but 'error' ζ , which still has expectation 0 and covariance S, is not Gaussian, expressions

$$x^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1} \Gamma^{\mathrm{T}} S^{-1} z$$
$$P^a = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1}$$

do not achieve Bayesian estimation, but define least-variance linear estimate of x from z (*Best Linear Unbiased Estimator, BLUE*), and associated estimation error covariance matrix.

From course 4

Best Linear Unbiased Estimate (continuation 1)

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

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

A probability distribution being known for the couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$, eqs (1-2) define probability distribution for the couple $(\boldsymbol{x}, \boldsymbol{y})$, with

$$E(\boldsymbol{x}) = \boldsymbol{x}^b \ , \ \ \boldsymbol{x}' = \boldsymbol{x} - E(\boldsymbol{x}) = - \boldsymbol{\zeta}^b$$

$$E(\boldsymbol{y}) = H\boldsymbol{x}^b \ , \ \ \boldsymbol{y}' = \boldsymbol{y} - E(\boldsymbol{y}) = \boldsymbol{y} - H\boldsymbol{x}^b = \boldsymbol{\varepsilon} - H\boldsymbol{\zeta}^b$$

$\boldsymbol{d} \equiv \boldsymbol{y} - H\boldsymbol{x}^b$ is called the *innovation vector*.

From course 4

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

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

x^a is the *Best Linear Unbiased Estimate* (*BLUE*) of x from x^b and y.

Equivalent set of formulæ

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

Vector $d \equiv \boldsymbol{y} - H\boldsymbol{x}^b$ is *innovation vector*

Matrix $K \equiv P^b H^{\mathrm{T}} [HP^bH^{\mathrm{T}} + R]^{-1} = P^a H^{\mathrm{T}} R^{-1}$ is *gain matrix*.

If probability distributions are *globally* gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

Variational Assimilation

Variational form of the *BLUE*

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

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

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

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

'*3D-Var*'

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

Used operationally in USA, Australia, China, …

Case of data that are distributed over time

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

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

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)

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

$\boldsymbol{v} = (v_j)$ 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 (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

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

....

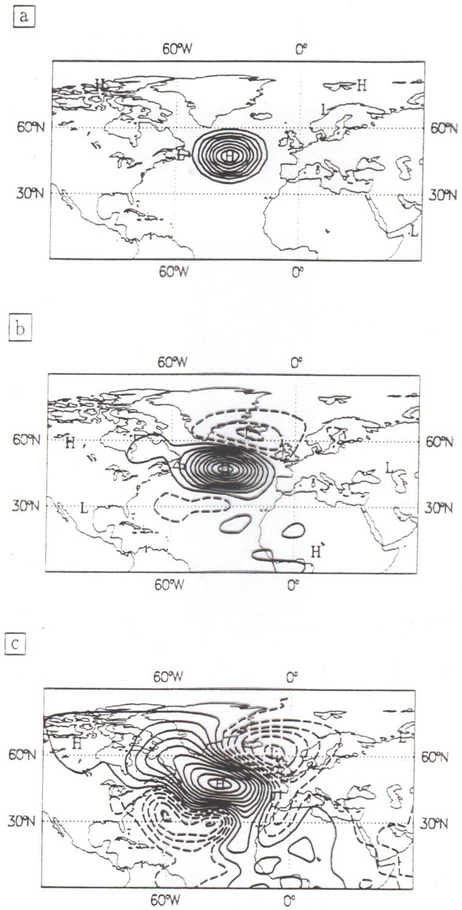
$$\lambda_k = M_k{}'^T \lambda_{k+1} + H_k{}'^T R_k^{-1} [H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$$

....

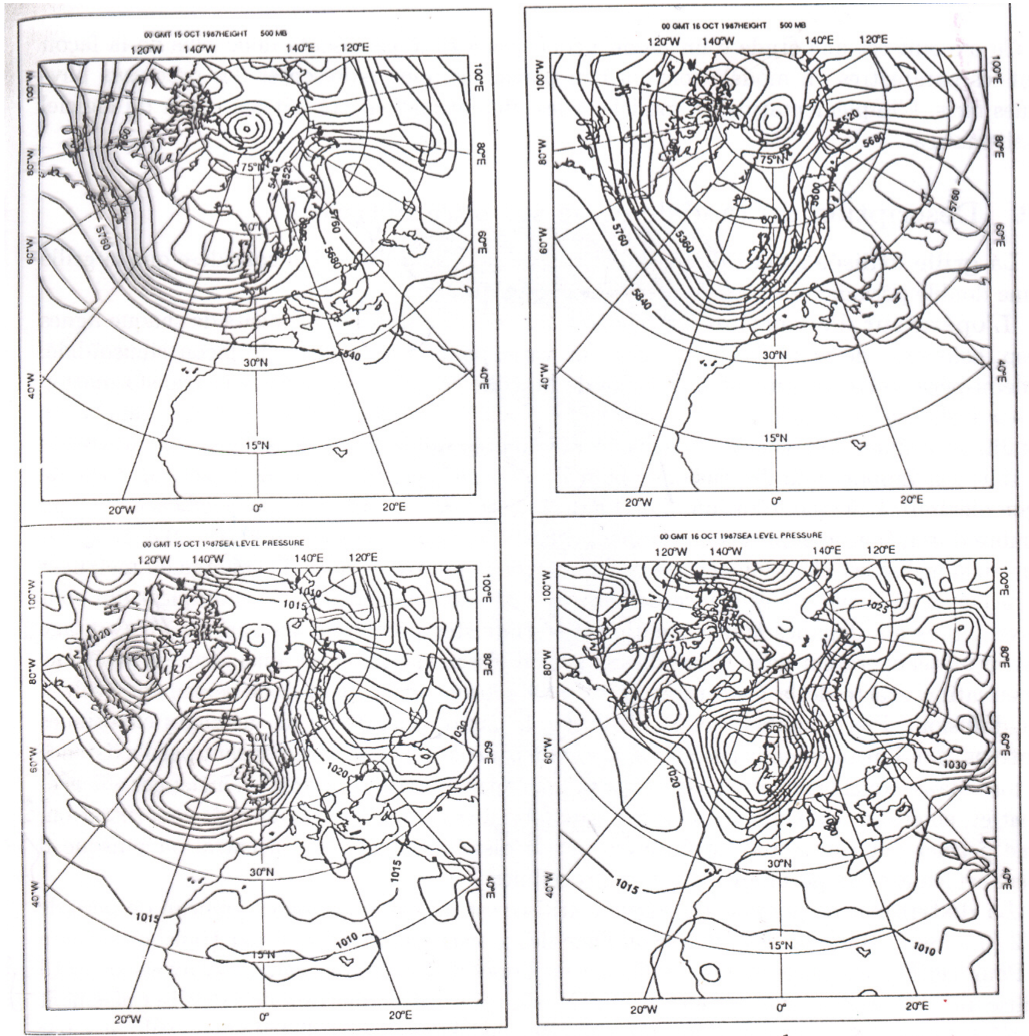
$$\lambda_0 = M_0{}'^T \lambda_1 + H_0{}'^T R_0^{-1} [H_0(\xi_0) - y_0] + [P_0^b]^{-1} (\xi_0 - x_0^b)$$

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

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



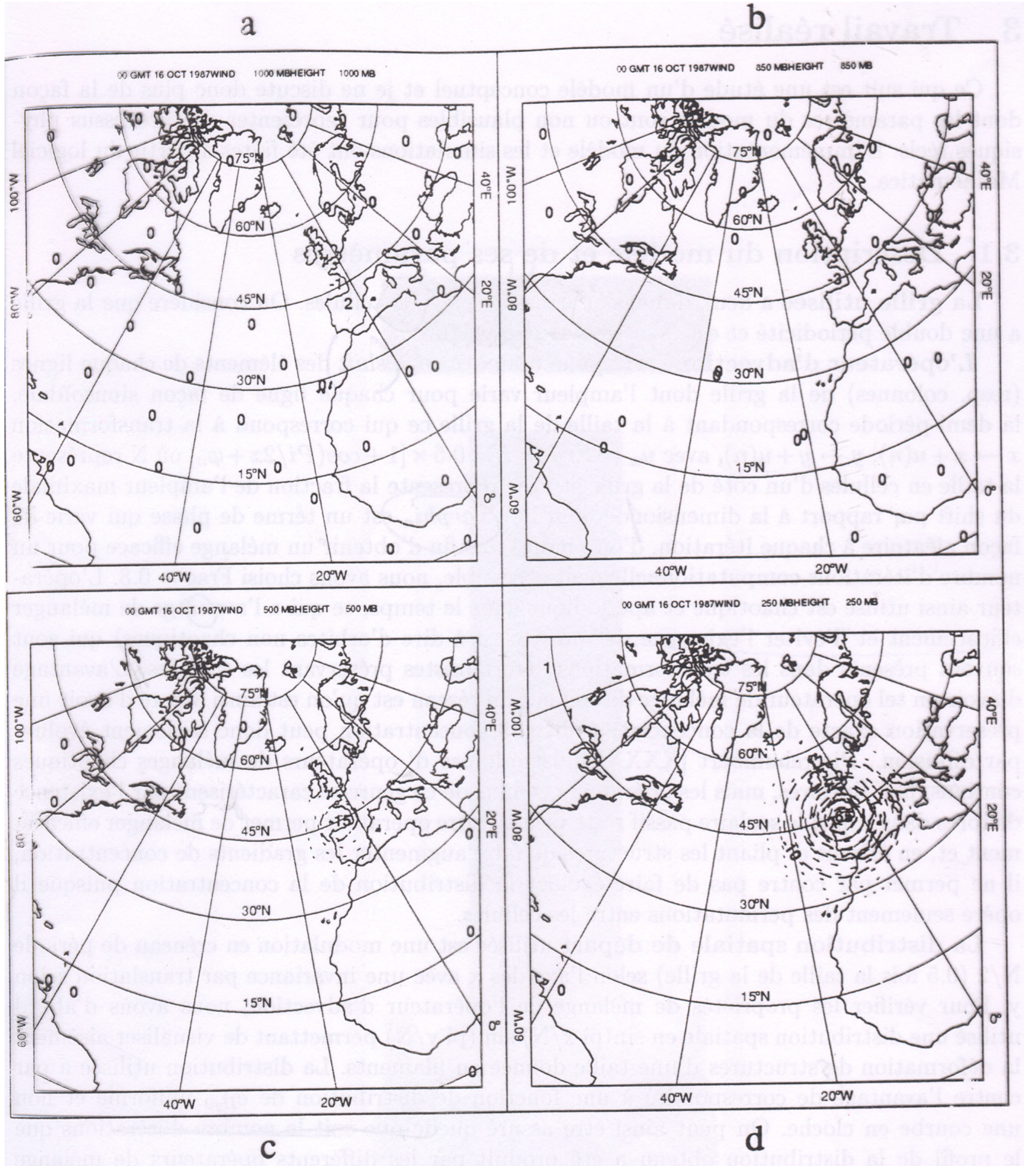
Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.



b d

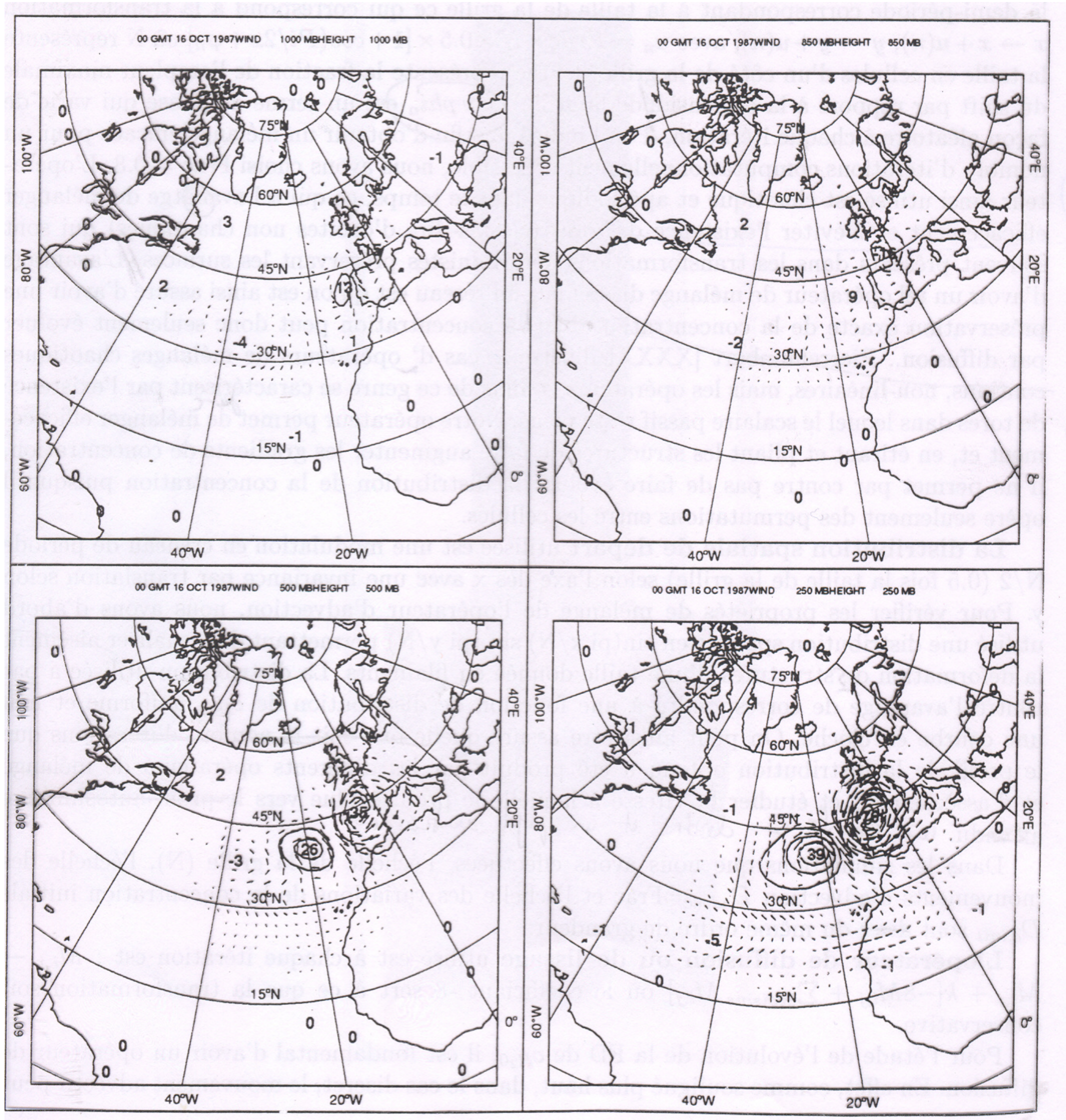
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



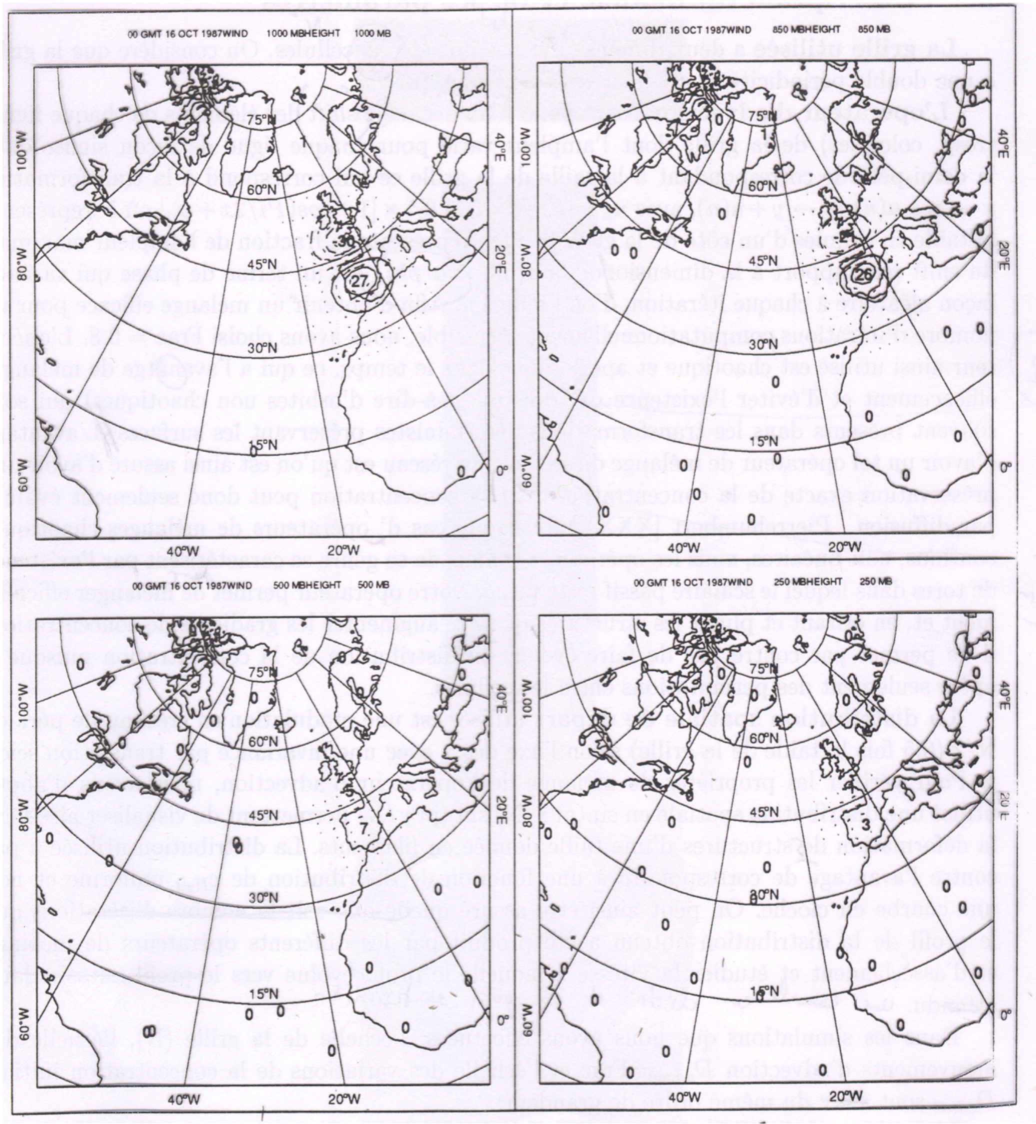
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



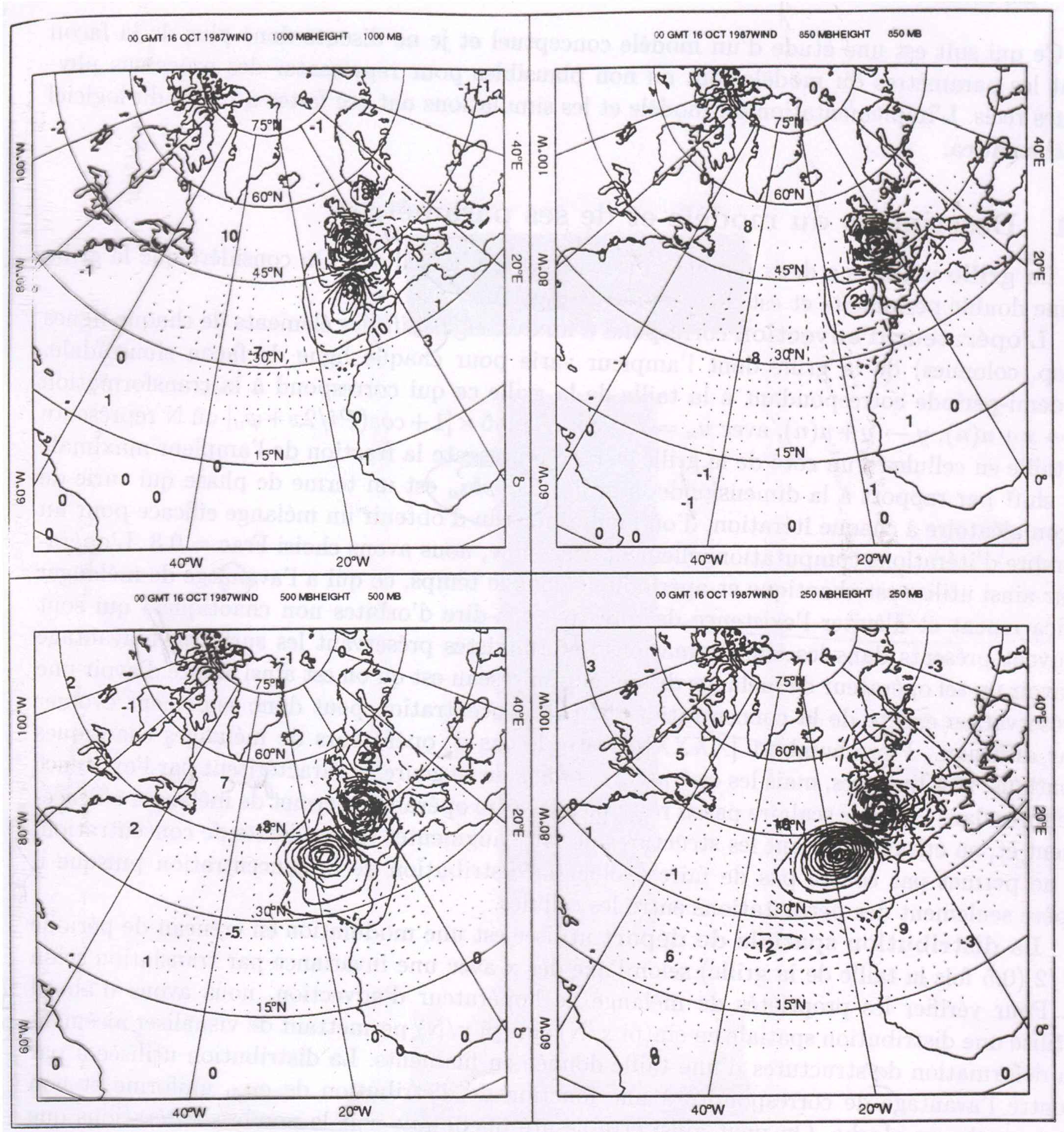
Same as before, but at the end of a 24-hr 4D-Var

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



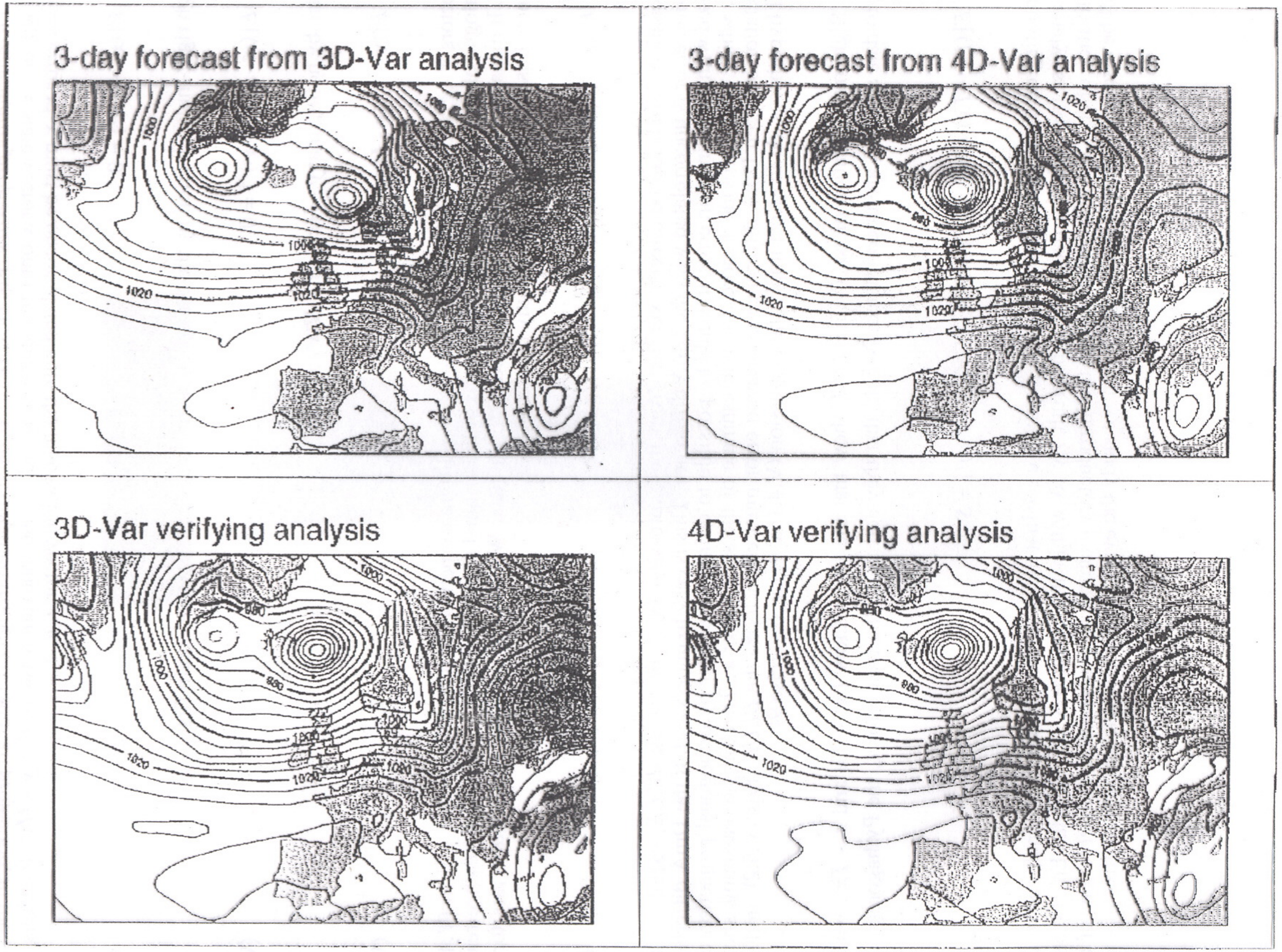
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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.

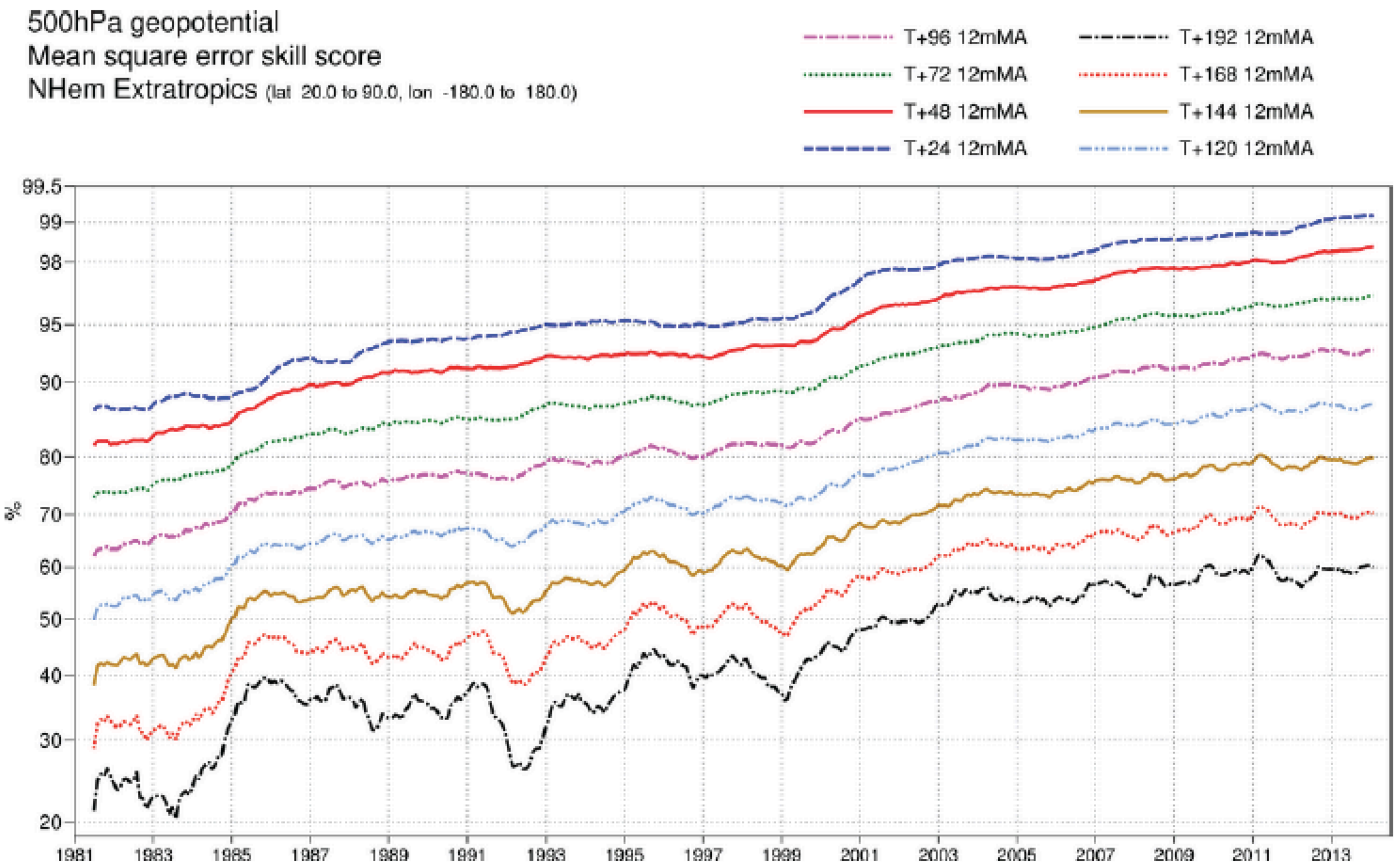
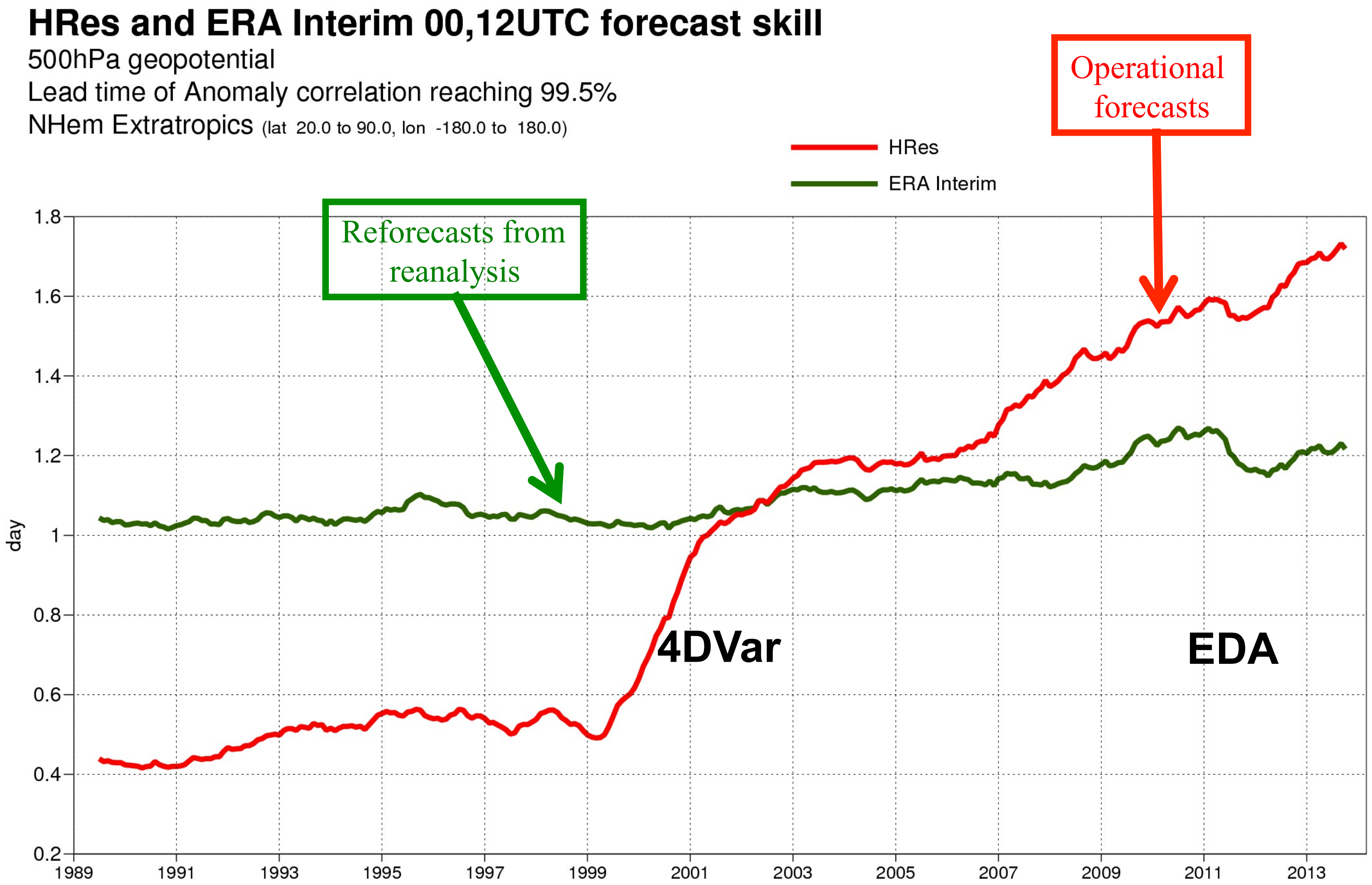


Figure 3: 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

Persistence = 0 ; climatology = 50 at long range

Initial state error reduction



Credit E. Källén, ECMWF

Time-correlated Errors (continuation 3)

Moral. If data errors are correlated in time, it is not possible to discard observations as they are used. In particular, if model error is correlated in time, all observations are liable to be reweighted as assimilation proceeds.

Variational assimilation can take time-correlated errors into account.

Example of time-correlated observation errors. Global covariance matrix

$$\mathcal{R} = (R_{kk'} = E(\varepsilon_k \varepsilon_{k'}^{\mathrm{T}}))$$

Objective function

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

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^{\mathrm{T}}\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_{kk'}[y_k - H_k\xi_k]^{\mathrm{T}}\,[\mathcal{R}^{-1}]_{kk'}\,[y_{k'} - H_{k'}\xi_{k'}]$$

where $[\mathcal{R}^{-1}]_{kk'}$ is the kk'-sub-block of global inverse matrix $\mathcal{R}^{-1}$.

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 4)

Temporal correlation of observational error has been introduced by ECMWF (Järvinen *et al.*, 1999) in variational assimilation of high-frequency surface pressure observations (correlation originates in that case in representativeness error).

Identification and quantification of time correlation of errors, especially model errors ?

In the linear case, and if errors are uncorrelated in time, Kalman Smoother and Variational Assimilation are algorithmically equivalent. They produce the *BLUE* of the state of the system from all available data, over the whole assimilation window (Kalman Filter produces the *BLUE* only at the end of the final time of the window). If in addition errors are Gaussian, both algorithms achieve Bayesian estimation.

If errors are correlated in time, one some Kalman Smoothers are equivalent with Variational Assimilation.

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.

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

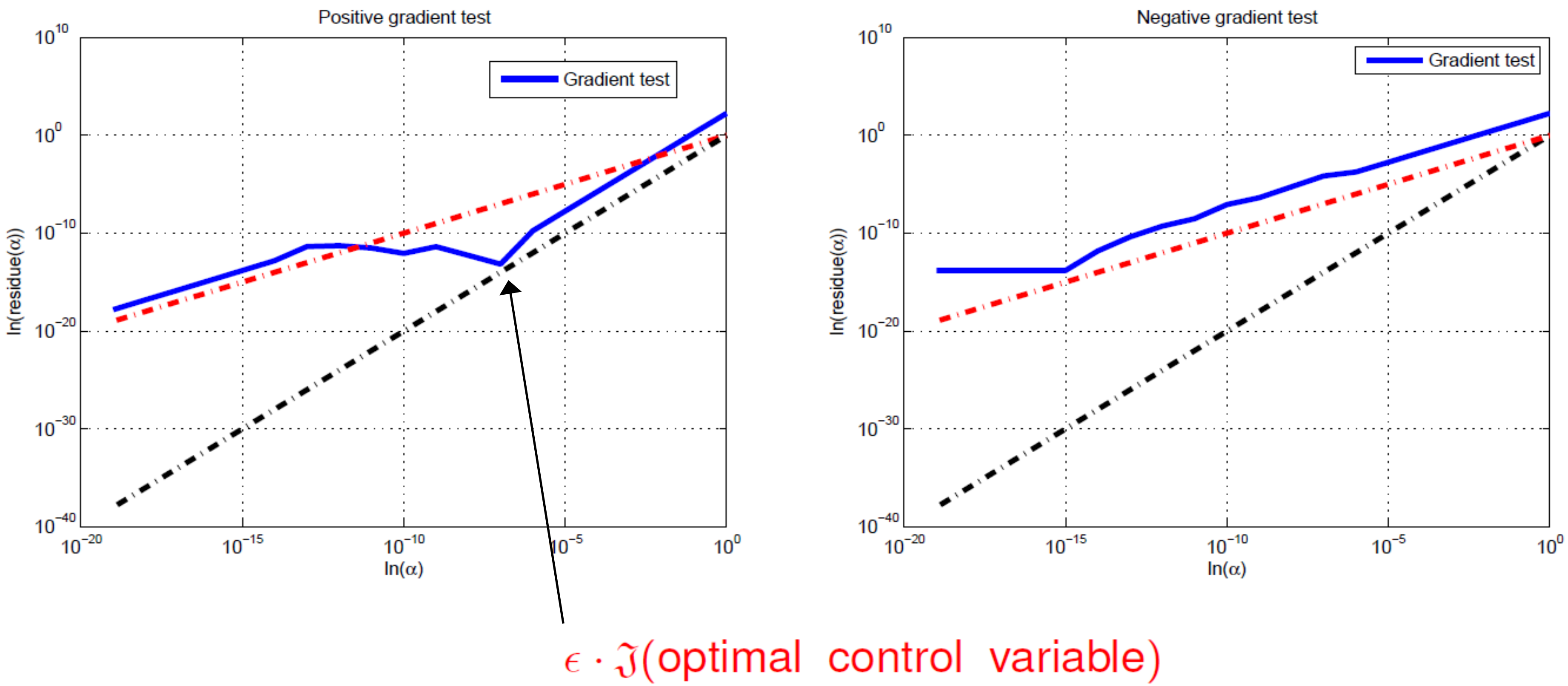
$$\partial J/\partial b = (\partial J/\partial a)\underbrace{(\partial a/\partial b)}_{c} + (\partial J/\partial b')\underbrace{(\partial b'/\partial b)}_{1} + (\partial J/\partial c')\underbrace{(\partial c'/\partial b)}_{0}$$

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

Similarly

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

Gradient test



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

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

M. Jardak

How to take model error into account in variational assimilation ?

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 $k = 0, \ldots, K$

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

  - Model

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

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

Then objective function

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

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

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

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

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

Can include nonlinear M_k and/or H_k.

Implemented operationally at ECMWF for the assimilation in the stratosphere.

Becomes singular in the strong constraint limit $Q_k \rightarrow 0$

Cours à venir

~~Jeudi 19 avril~~

~~Jeudi 26 avril~~

~~Jeudi 3 mai~~

~~Lundi 14 mai~~

~~Jeudi 17 mai~~

~~Jeudi 24 mai~~

Jeudi 7 juin

Jeudi 14 juin

De 10h00 à 12h30, Salle E314, 3ième étage, Département de Géosciences, École Normale Supérieure, 24, rue Lhomond, Paris 5