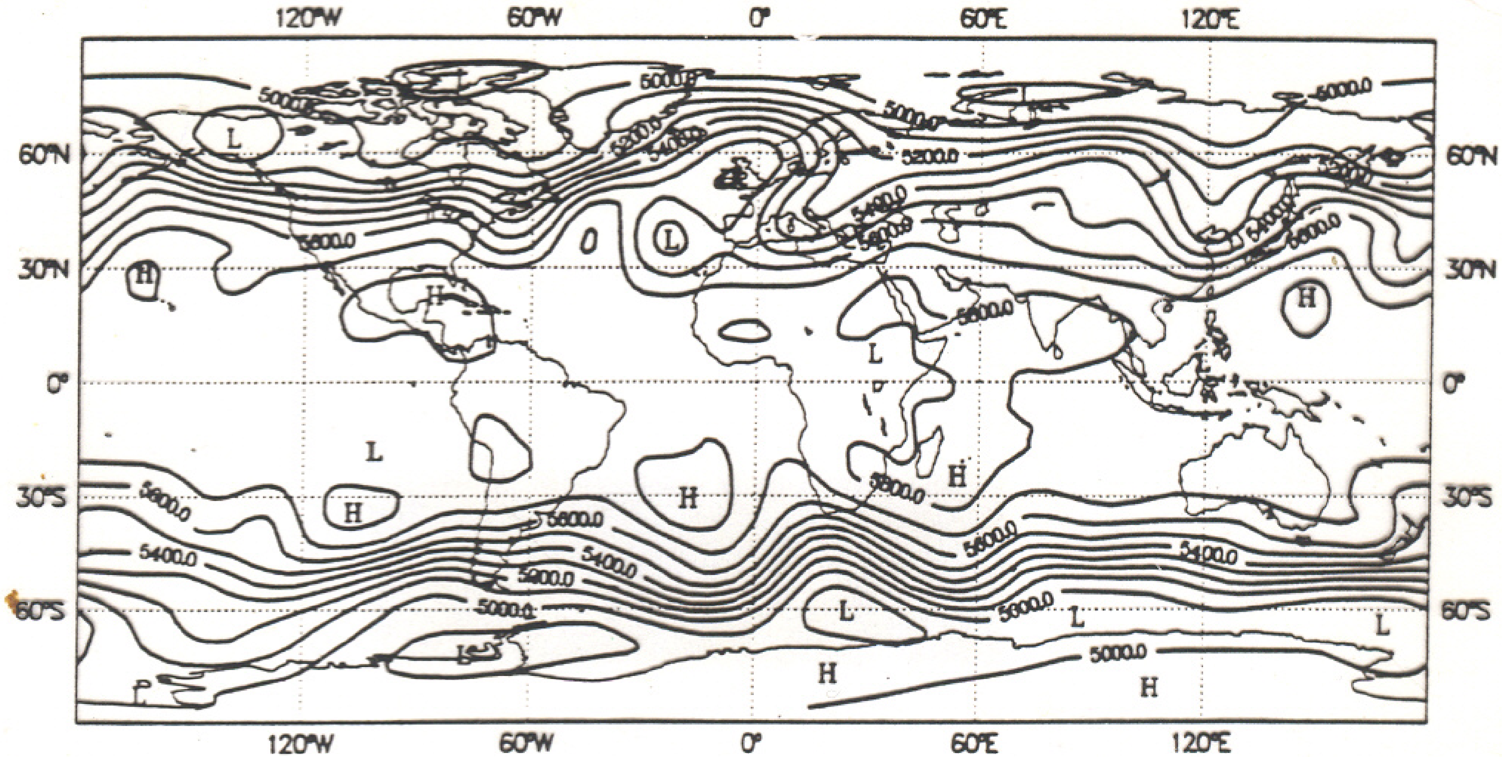
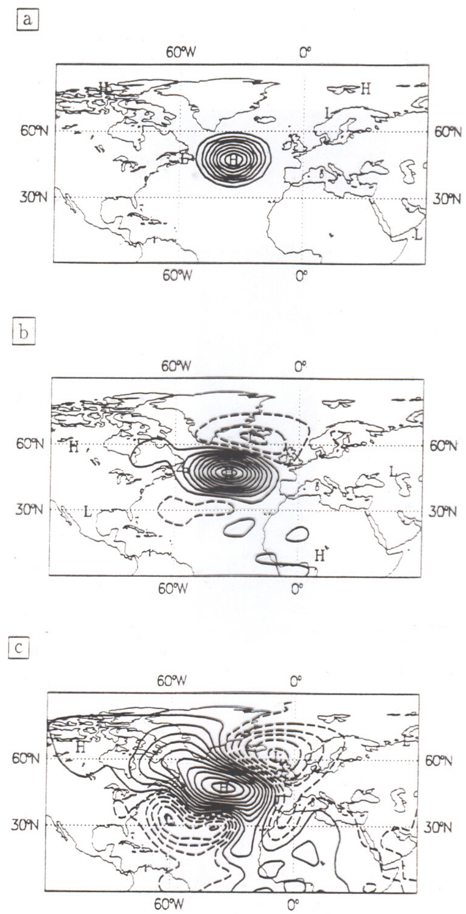
Advanced assimilation methods.
Variational assimilation. Adjoint equations
Part 4

Olivier Talagrand
School *Data Assimilation*
Nordic Institute for Theoretical Physics (NORDITA)
Stockholm, Sweden
28 April 2011

Costly part in assimilation of meteorological and oceanographical observations is to carry in time the uncertainty on the estimate if the state of the flow. Kalman filter (KF) does it explicitly, by evolving an error covariance matrix (standard KF) or an ensemble of points in state space (EnKF).



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.

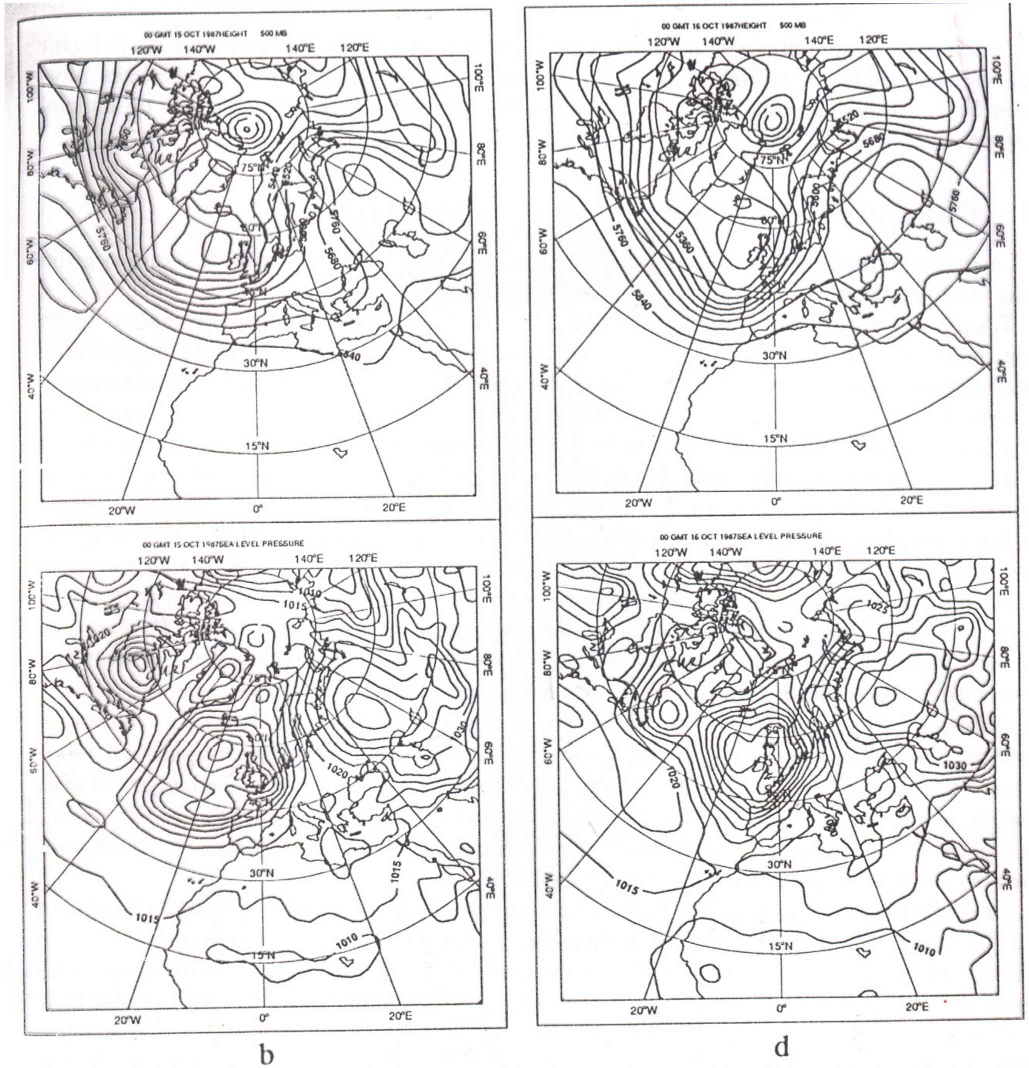
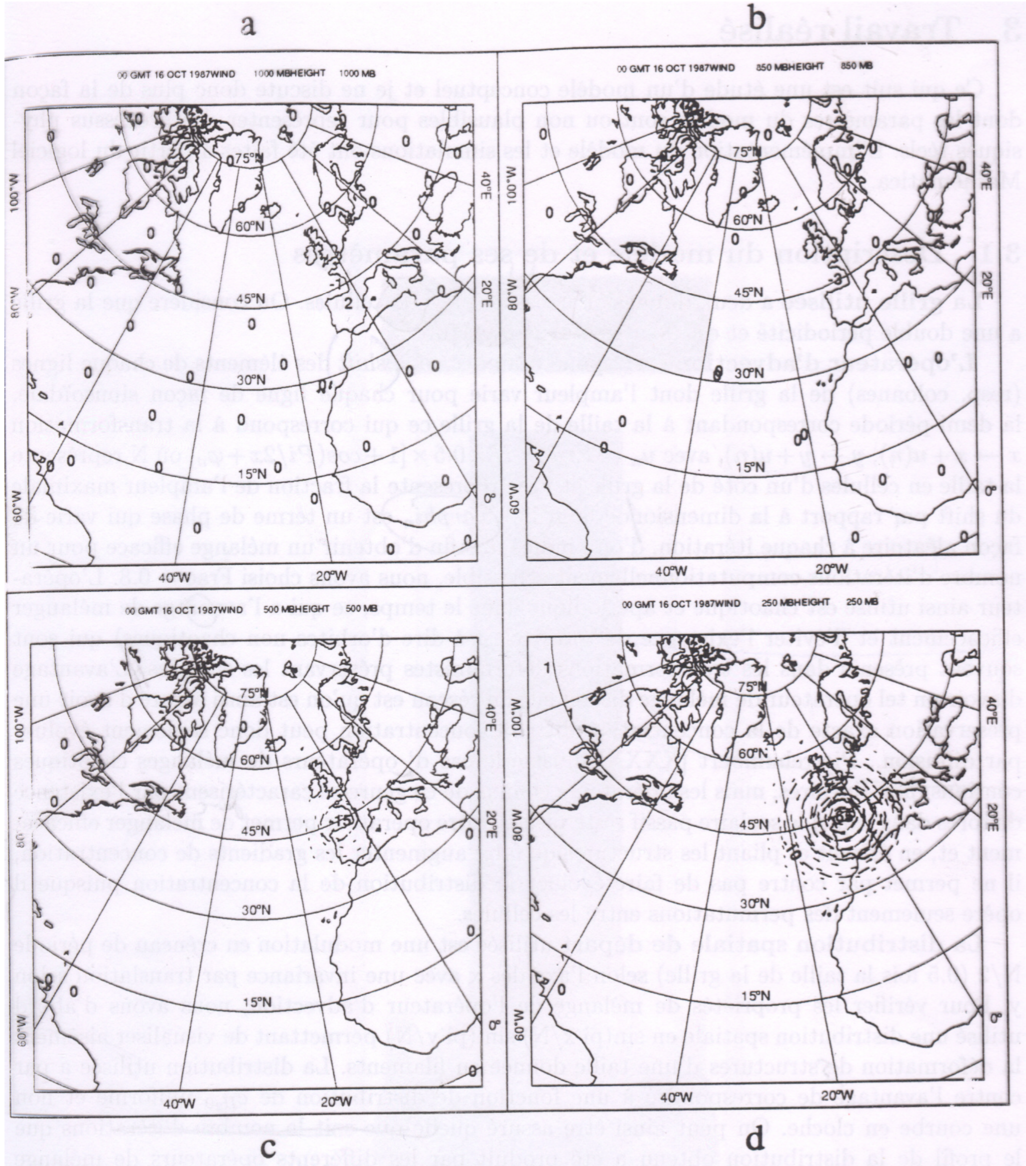


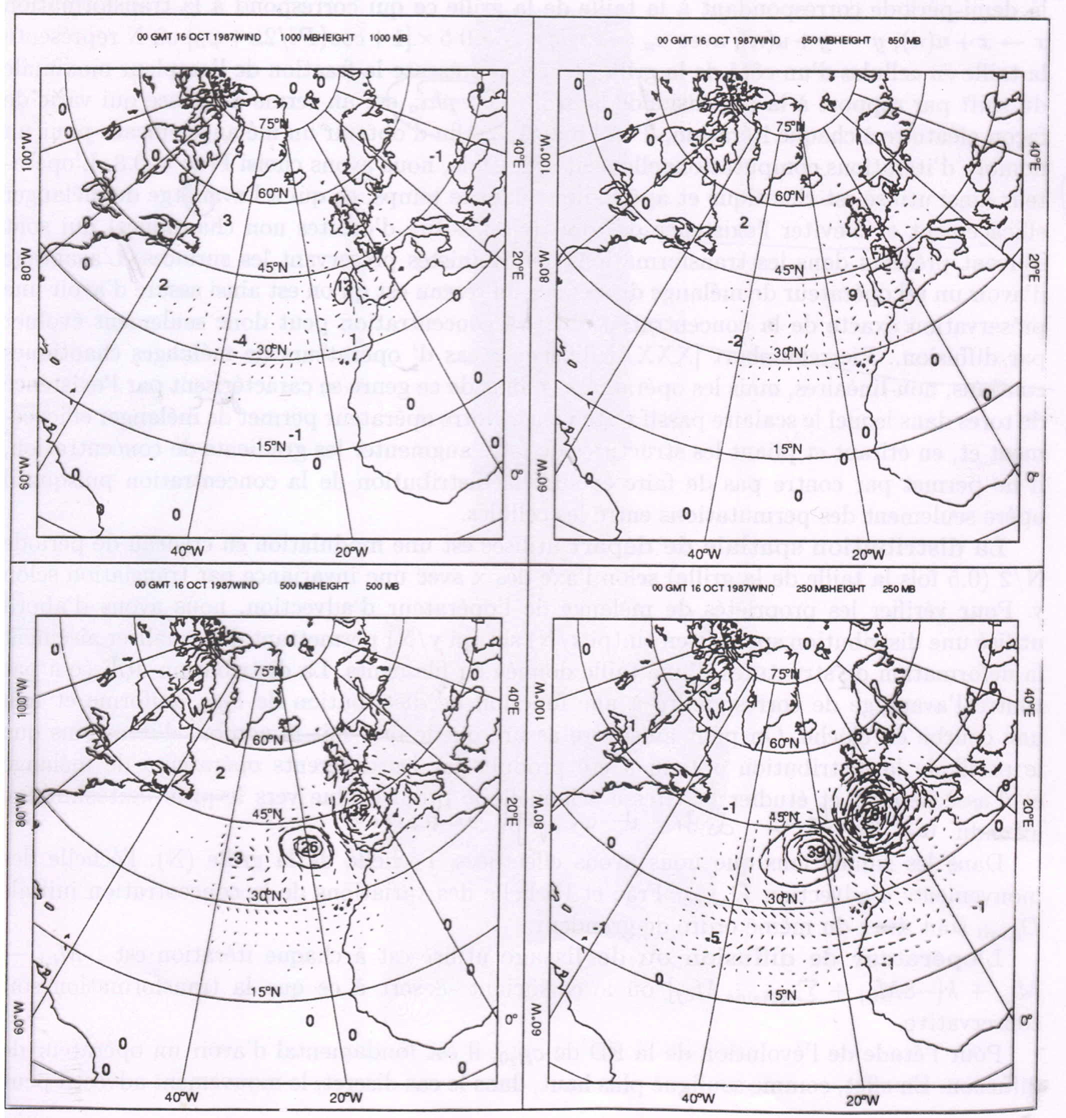
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



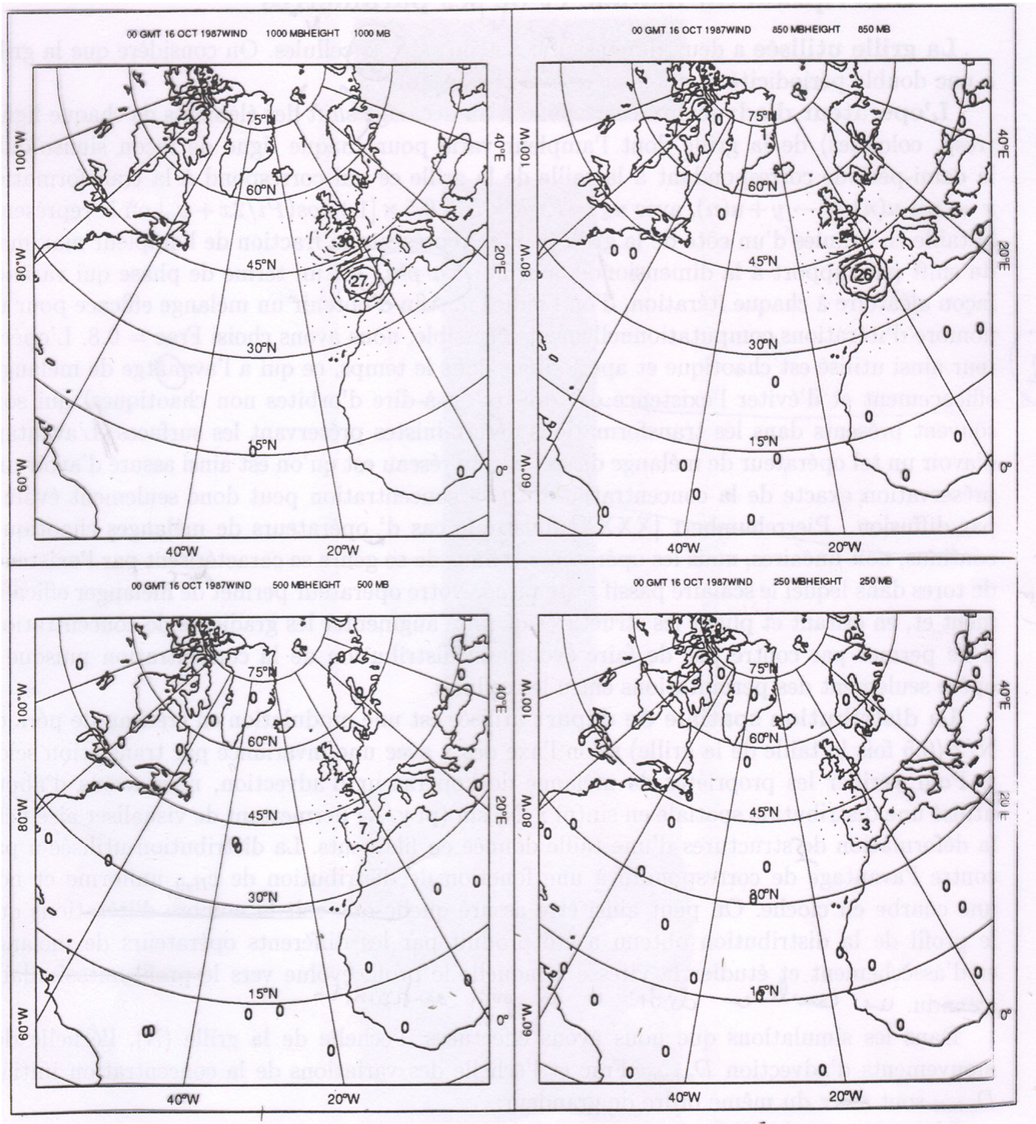
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



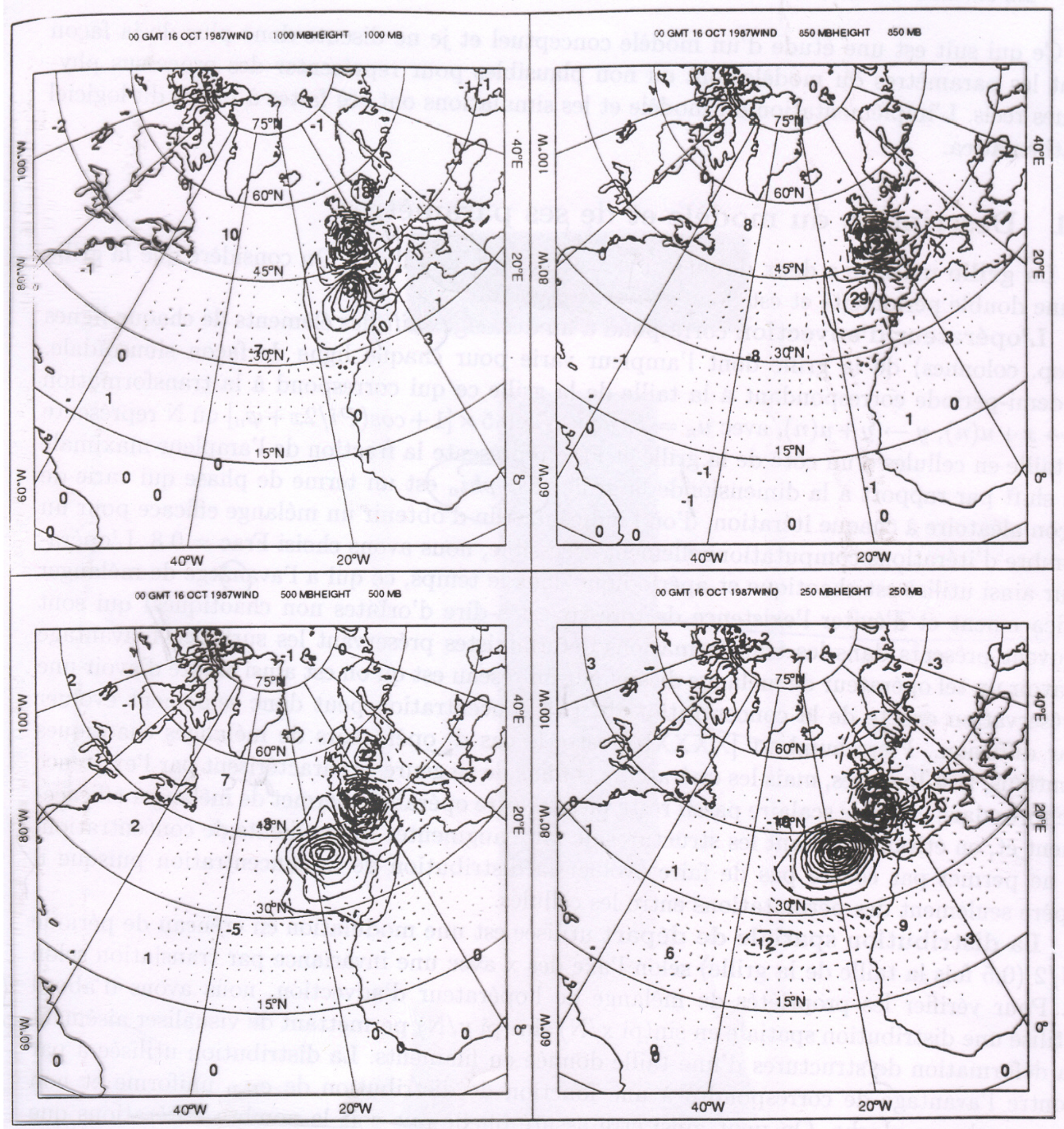
Same as before, but at the end of a 24-hr 4D-Var

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



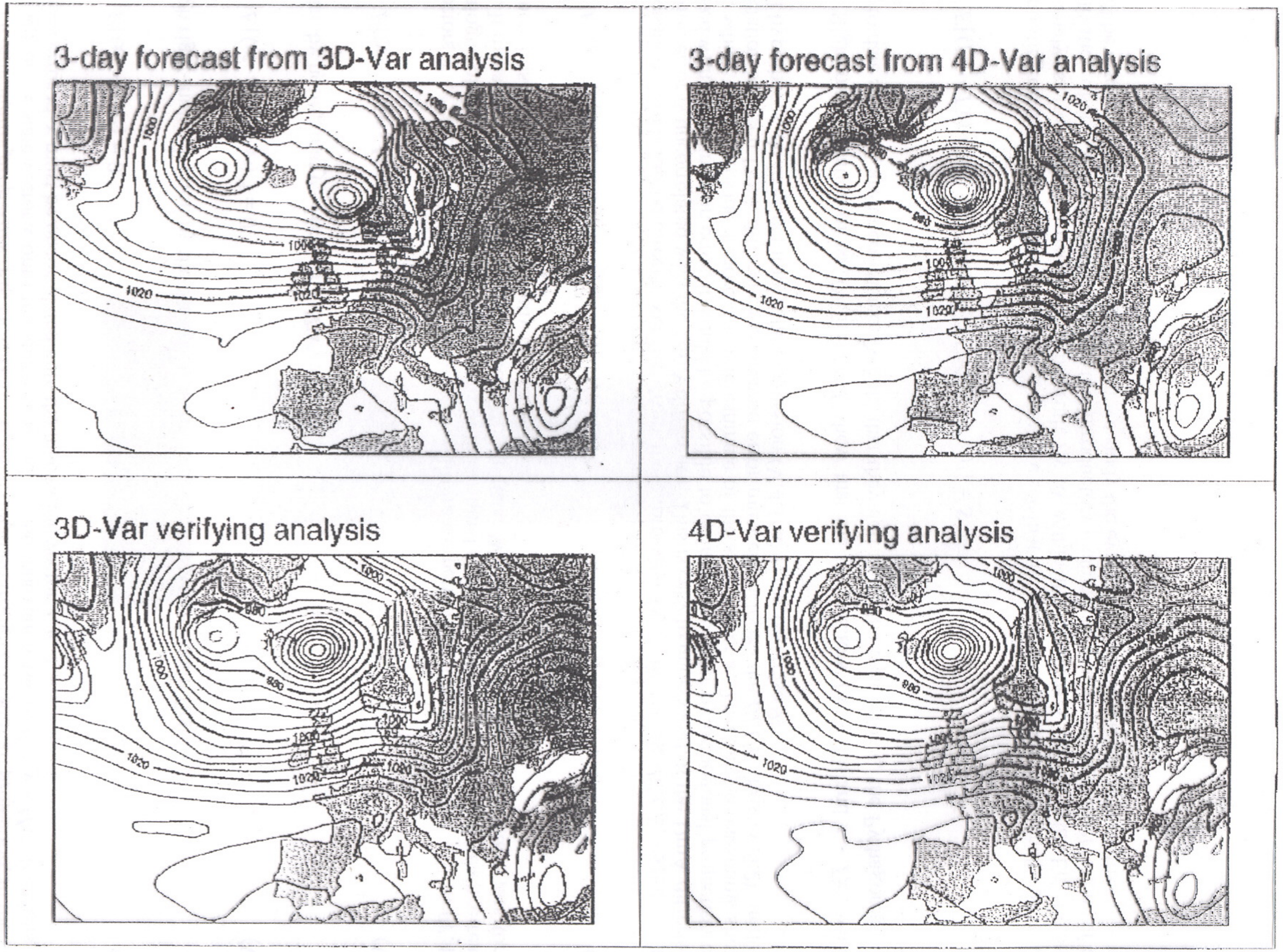
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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)

Adjoint Method. Practical implementation

Code

Input variables $u_1, u_2, \ldots, u_n$

$*$

$*$

$*$

$a = b \, x \, c$

$*$

$*$

$*$

$J = x_1^2 + x_2^2 + x_3^2$

Purpose. Determine partial derivatives of J with respect to $u_1, u_2, \ldots, u_n$

Adjoint Method. Practical implementation (2)

Input variables $u_1, u_2, \ldots, u_n$

$*$

$*$

$*$

$a = b \, x \, c$

$*$

$*$

$*$

$J = x_1^2 + x_2^2 + x_3^2$

Last instruction

$\partial J/\partial x_1 = 2x_1$, $\partial J/\partial x_2 = 2x_2$, $\partial J/\partial x_3 = 2x_3$

And then proceed backwards

Adjoint Method. Practical implementation (3)

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

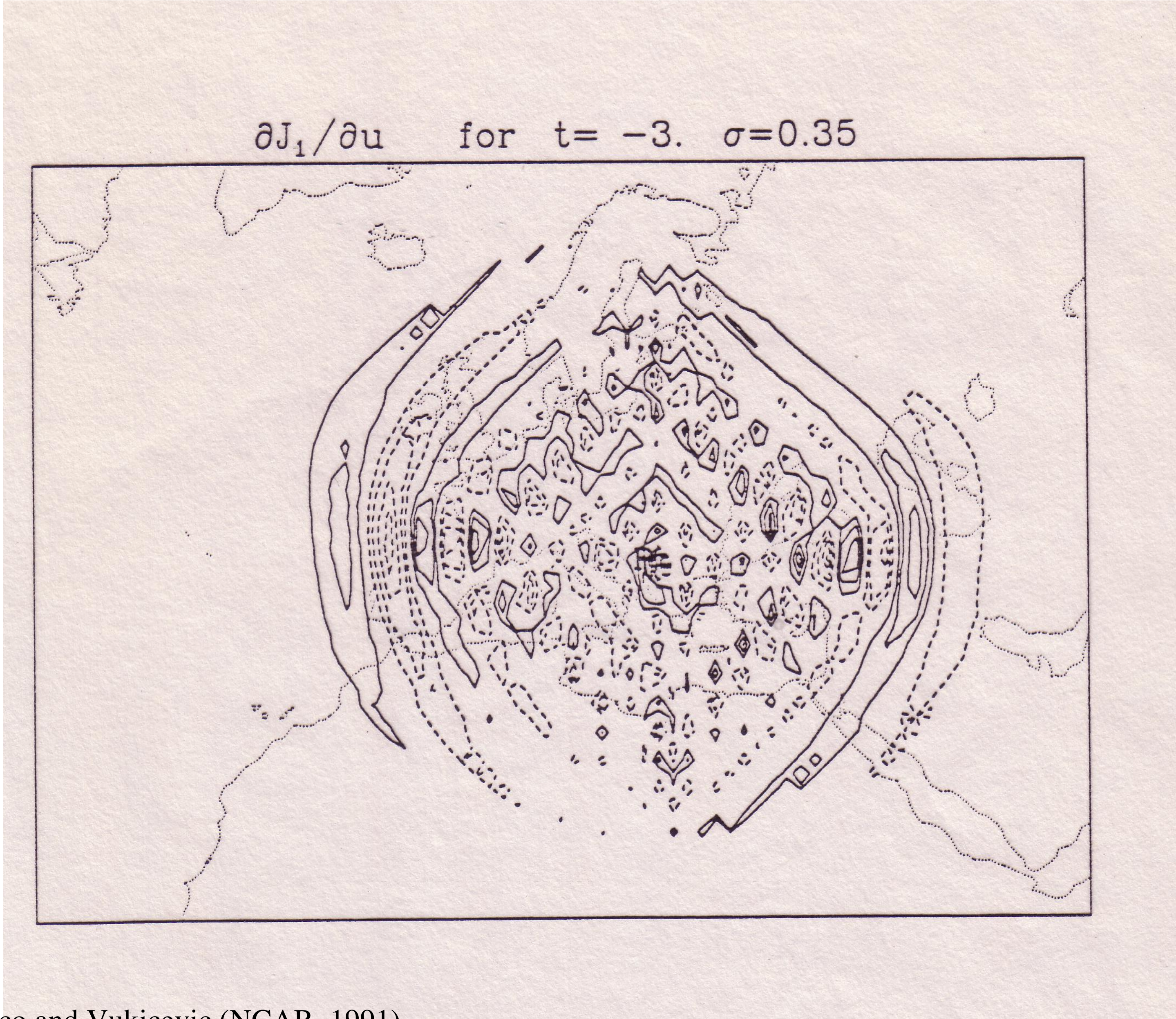
$$\partial J/\partial c = (\partial J/\partial a)\, b + \partial J/\partial c'$$

Adjoint Method is very powerful (actually, there is no competitor as far as computing gradients with respect to a large number of variables is concerned).

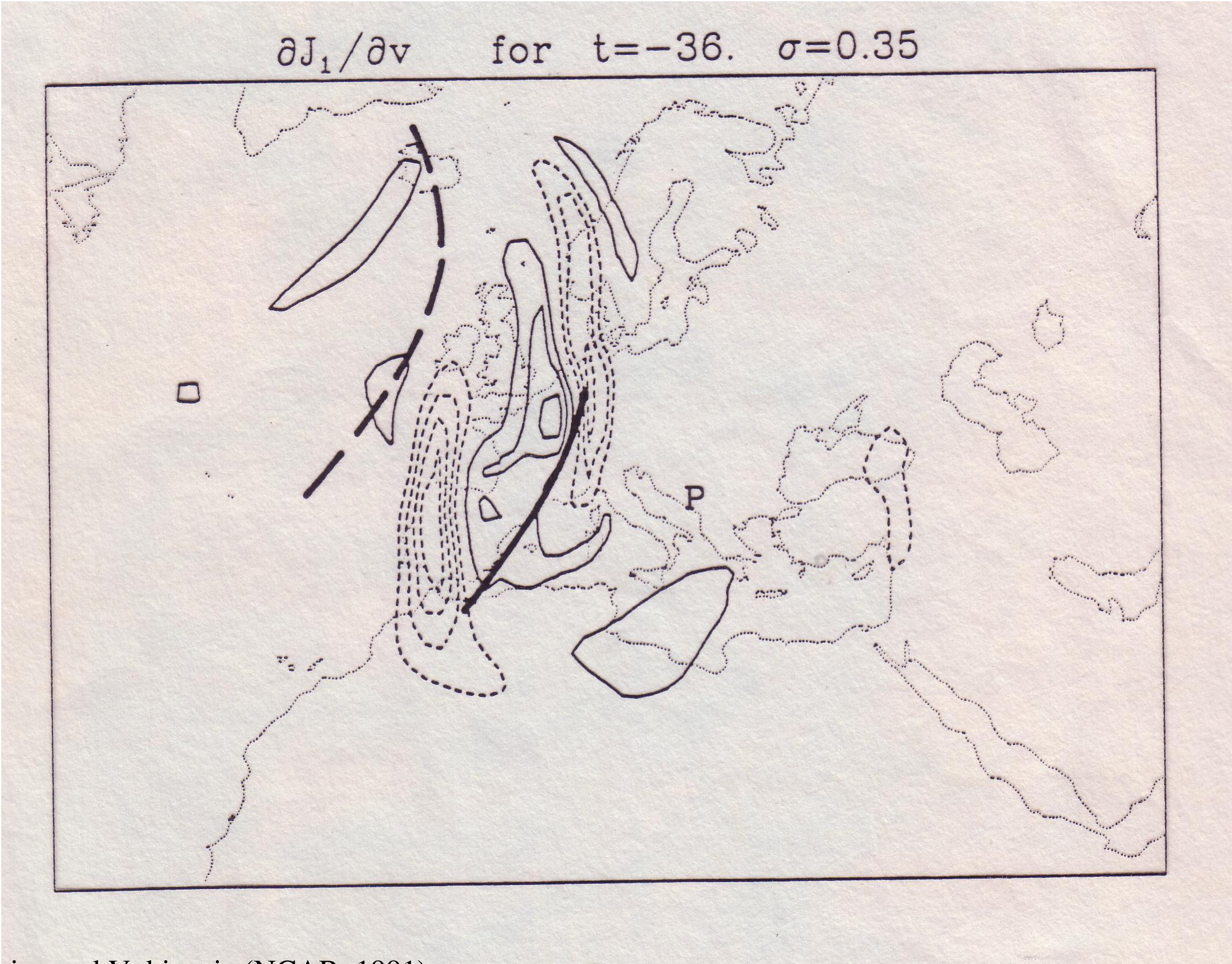
But there is a price to pay for that efficiency. Larger memory requirements and, especially necessity of developing, validating (and maintaining …) adjoint codes.

Quasi-mechanical rules exist for writing adjoint codes. Nevertheless, writing and validating an adjoint code can be a lengthy and tedious task. The best is often to write the direct and adjoint codes in parallel.

Adjoint compilers exist, that automatically derive the adjoint of a given code. But adjoints produced by those compilers are in general less efficient than adjoints written by hand, and require careful validation. Adjoint compilers nevertheless very useful for developing adjoints of very big codes, such as NWP models.



Errico and Vukicevic (NCAR, 1991)



Errico and Vukicevic (NCAR, 1991)

Varitaional assimilation minimizes objective function

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

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

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

$(\boldsymbol{x}_1, \boldsymbol{x}_2) \rightarrow \boldsymbol{x}_1^{\mathrm{T}}\,[\boldsymbol{P}^b]^{-1}\,\boldsymbol{x}_2$ and $(\boldsymbol{y}_1, \boldsymbol{y}_2) \rightarrow \boldsymbol{y}_1^{\mathrm{T}}\,\boldsymbol{R}^{-1}\,\boldsymbol{y}_2$ define intrinsic (*i.e.*, coordinate independent) scalar products in state and observation spaces respectively, called the *Mahalanobis scalar products* associated with the covariance matrices (tensors) $\boldsymbol{P}^b$ and $\boldsymbol{R}$ respectively.

Tangent Linear Approximation

Data in the form of background +additional 'observations'

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

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

where operator H is nonlinear (includes model in the case of 4D-Var).

Data (1-2) are equivalent to

$$\boldsymbol{x}^b = \boldsymbol{x} + \zeta^b \qquad (3)$$

$$\boldsymbol{d} \equiv \boldsymbol{y} - H(\boldsymbol{x}^b) + \boldsymbol{\varepsilon} \qquad (4) \text{ (innovation)}$$

If background $\boldsymbol{x}^b$ is sufficiently accurate such that

$$\boldsymbol{d} = H(\boldsymbol{x}) - H(\boldsymbol{x}^b) + \boldsymbol{\varepsilon} \approx \boldsymbol{H}'(\mathbf{x}-\boldsymbol{x}^b) + \boldsymbol{\varepsilon}$$

where $\boldsymbol{H}'$ is Jacobian of H at point $\boldsymbol{x}^b$, then estimation problem is linear in terms of deviation $\mathbf{x}-\boldsymbol{x}^b$ from background.

$$H(\boldsymbol{x}) - H(\boldsymbol{x}^b) \approx \boldsymbol{H}'(\mathbf{x}-\boldsymbol{x}^b)$$

Tangent Linear Approximation, valid for large scale meteorology (scales > 100 km) up to ranges 24-48 hours.

Explicitly implemented in Extended Kalman Filter

'Incremental' approach to Variational Assimilation

If you want to implement 4D-Var, do you have to to first develop the adjoint of the whole model, and then take your chance ? Is there not a way to proceed more progressively ?

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

First-order perturbation to solution $\{\xi_k\}$

$\delta\xi_{k+1} = M_k'\ \delta\xi_k$, where is Jacobian of M_k at point ξ_k

Adjoint equation

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

…

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

…

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

‘Incremental’ approach to Variational Assimilation (continuation 1)

Simplify adjoint ? That is *not* the solution. Gradient will be erroneous, and minimization will fail.

Solution : simplify simultaneously, and *consistently*, both the direct perturbation dynamics and the associated adjoint.

Basic dynamics $\xi_{k+1} = M_k(\xi_k)$, $k = 0, \dots, K\text{-}1$

Perturbation dynamics

$\delta\xi_{k+1} = G_k(\delta\xi_k)$, where G_k is appropriately simplified form of M_k'

Adjoint equation

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

…

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

…

$$\lambda_0 = G_0'^{\mathrm{T}} \lambda_1 + H_0'^{\mathrm{T}} R_0^{-1} [H_0(\xi_0) - y_0] + [P_0^b]^{-1} (\xi_0 - x_0^b)$$

‘Incremental’ approach to Variational Assimilation (continuation 2)

Incremental approach generally (systematically ?) implemented in weather centres where 4D-Var is used for operational assimilation (ECMWF, Météo-France, UK Met Office, …).

Can be made iterative, with increasing complexity in G_k (inner and outer loops).

Largely heuristic.

Strong nonlinearities can result in existence of multiple minima for objective function.

Quasi-Static Variational Assimilation (Swanson *et al*., Luong, Järvinen *et al*.)

Progressively extend length of assimilation window, each new minimization being started from the result of the previous one, so as to keep track of the absolute minimum.

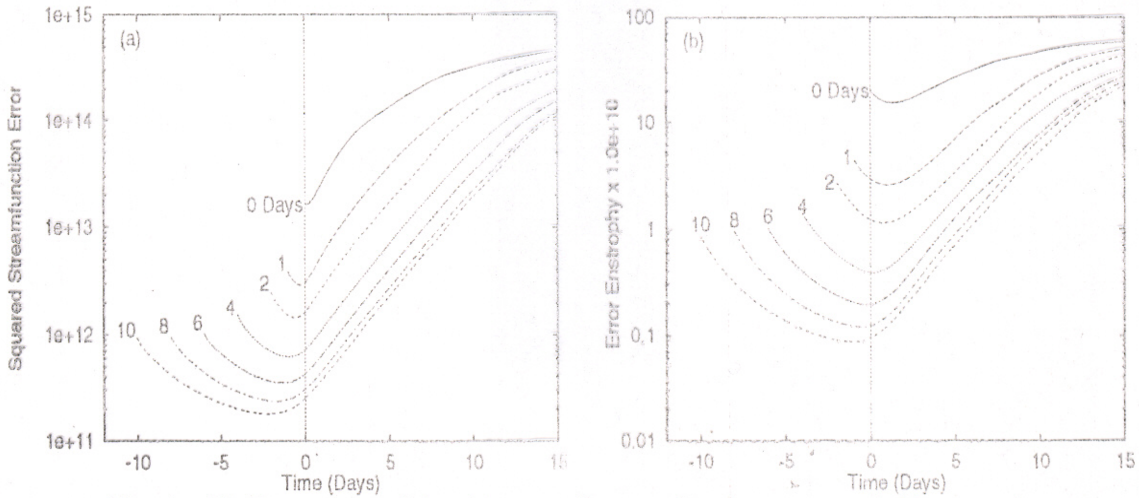


Fig. 4 Median values of the (a) squared streamfunction error and (b) enstrophy error for the 200 forecast set as a function of forecast time and of the assimilation time interval T_a.

Swanson, Vautard and Pires, 1998, *Tellus*, **50A**, 369-390

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

`

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

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

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

Time-correlated Errors (continuation 2)

Moral. If data errors are correlated in time, it is not possible to discard observations as they are used while preserving optimality of the estimation process. 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'-subblock of global inverse matrix $\mathcal{R}^{-1}$.

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 3)

Time 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 temporal correlation of errors, especially model errors ?

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

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.

Dual Algorithm for Variational Assimilation (aka *Physical Space Analysis System*, *PSAS,* pronounced *'peezaz',* developed at Data Assimilation Office, NASA, Greenbelt)

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

$$x^a = x^b + P^b H^T \Lambda^{-1} d = x^b + P^b H^T m$$

where $\Lambda \equiv HP^bH^T + R$, $d \equiv y - Hx^b$ and $m \equiv \Lambda^{-1} d$ maximises

$$\mu \rightarrow \mathcal{K}(\mu) = -(1/2)\ \mu^T \Lambda\ \mu + d^T\mu$$

Maximisation is performed in (dual of) observation space.

Dual Algorithm for Variational Assimilation (continuation 2)

Extends to time dimension, and to weak-constraint case, by defining state vector as

$$x \equiv (x_0^{\mathrm{T}}, x_1^{\mathrm{T}}, \ldots, x_K^{\mathrm{T}})^{\mathrm{T}}$$

or, equivalently, but more conveniently, as

$$x \equiv (x_0^{\mathrm{T}}, \eta_0^{\mathrm{T}}, \ldots, \eta_{K\text{-}1}^{\mathrm{T}})^{\mathrm{T}}$$

where, as before

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

The background for x_0 is x_0^b, the background for η_k is 0. Complete background is

$$x^b = (x_0^{b\mathrm{T}}, 0^{\mathrm{T}}, \ldots, 0^{\mathrm{T}})^{\mathrm{T}}$$

It is associated with error covariance matrix

$$P^b = \mathrm{diag}(P_0^{\ b}, Q_0, \ldots, Q_{K\text{-}1})$$

Dual Algorithm for Variational Assimilation (continuation 3)

For any state vector $\xi = (\xi_0^{\mathrm{T}}, \upsilon_0^{\mathrm{T}}, \dots, \upsilon_{K-1}^{\mathrm{T}})^{\mathrm{T}}$, the observation operator H

$$\xi \rightarrow H\xi = (u_0^{\mathrm{T}}, \dots, u_K^{\mathrm{T}})^{\mathrm{T}}$$

is defined by the sequence of operations

$$u_0 = H_0 \xi_0$$

then for $k = 0, \dots, K\text{-}1$

$$\xi_{k+1} = M_k \xi_k + \upsilon_k$$
$$u_{k+1} = H_{k+1} \xi_{k+1}$$

The observation error covariance matrix is equal to

$$R = \mathrm{diag}(R_0, \dots, R_K)$$

Dual Algorithm for Variational Assimilation (continuation 4)

Maximization of dual objective function

$$\mu \rightarrow \mathcal{K}(\mu) = -(1/2)\ \mu^{\mathrm{T}}\ \Lambda\ \mu + d^{\mathrm{T}}\mu$$

requires explicit repeated computations of its gradient

$$\nabla_{\mu}\mathcal{K} = -\ \Lambda\mu + d = -\ (HP^{b}H^{\mathrm{T}} + R)\mu + d$$

Starting from $\mu = (\mu_0{}^{\mathrm{T}}, \ldots, \mu_K{}^{\mathrm{T}})^{\mathrm{T}}$ belonging to (dual) of observation space, this requires 5 successive steps

- Step 1. Multiplication by H^{T}. This is done by applying the transpose of the process defined above, *viz*.,

Set $\chi_K = 0$

Then, for $k = K\text{-}1, \ldots, 0$

$$\nu_k = \chi_{k+1} + H_{k+1}{}^{\mathrm{T}}\ \mu_{k+1}$$
$$\chi_k = M_k{}^{\mathrm{T}}\ \nu_k$$

Finally $\lambda_0 = \chi_0 + H_0{}^{\mathrm{T}}\ \mu_0$

The output of this step, which includes a backward integration of the adjoint model, is the vector $(\lambda_0{}^{\mathrm{T}}, \nu_0{}^{\mathrm{T}}, \ldots, \nu_{K\text{-}1}{}^{\mathrm{T}})^{\mathrm{T}}$

Dual Algorithm for Variational Assimilation (continuation 5)

- Step 2. Multiplication by P^b. This reduces to

$$\xi_0 = P_0^b \lambda_0$$
$$\upsilon_k = Q_k \nu_k \quad , \quad k = 0, \ldots, K\text{-}1$$

- Step 3. Multiplication by H. Apply process defined above to vector $(\xi_0^T, \upsilon_0^T, \ldots, \upsilon_{K-1}^T)^T$, thereby producing vector $(u_0^T, \ldots, u_K^T)^T$.

- Step 4. Add vector $R\mu$, *i. e.* compute

$$\varphi_0 = \xi_0 + R_0 \mu_0$$
$$\varphi_k = \upsilon_{k\text{-}1} + R_k \mu_k \quad , \quad k = 1, \ldots, K$$

- Step 5. Change sign of vector $\varphi = (\varphi_0^T, \ldots, \varphi_K^T)^T$, and add vector $d = y - Hx^b$,

Dual Algorithm for Variational Assimilation (continuation 6)

The model error covariance matrix Q_k is present in the algorithm only in its direct (not inverse form). Dual algorithm remains regular in the limit of vanishing model error. Can be used for both strong- and weak-constraint assimilation.

No significant increase of computing cost in comparison with standard strong-constraint variational assimilation (Louvel)

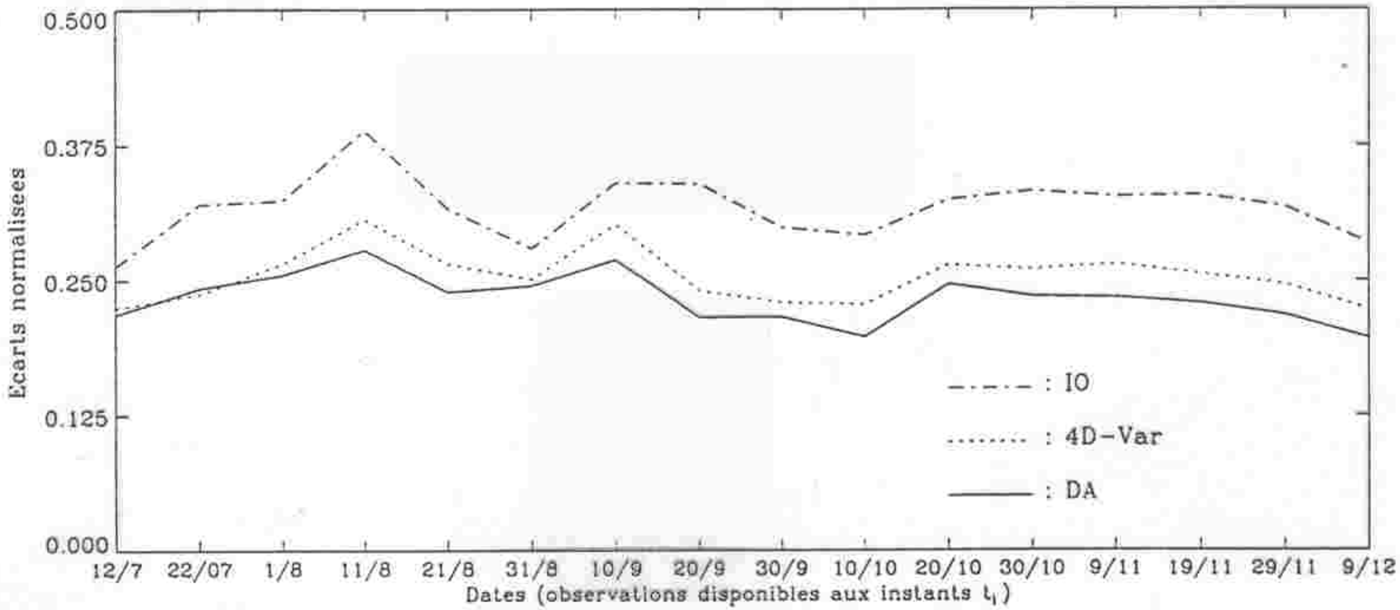


FIG. 9.11 – *Ecarts normalisés prévision/observations sur l'ensemble de la période étudiée*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999 nous permet d'observer les performances des différentes techniques d'assimilation.

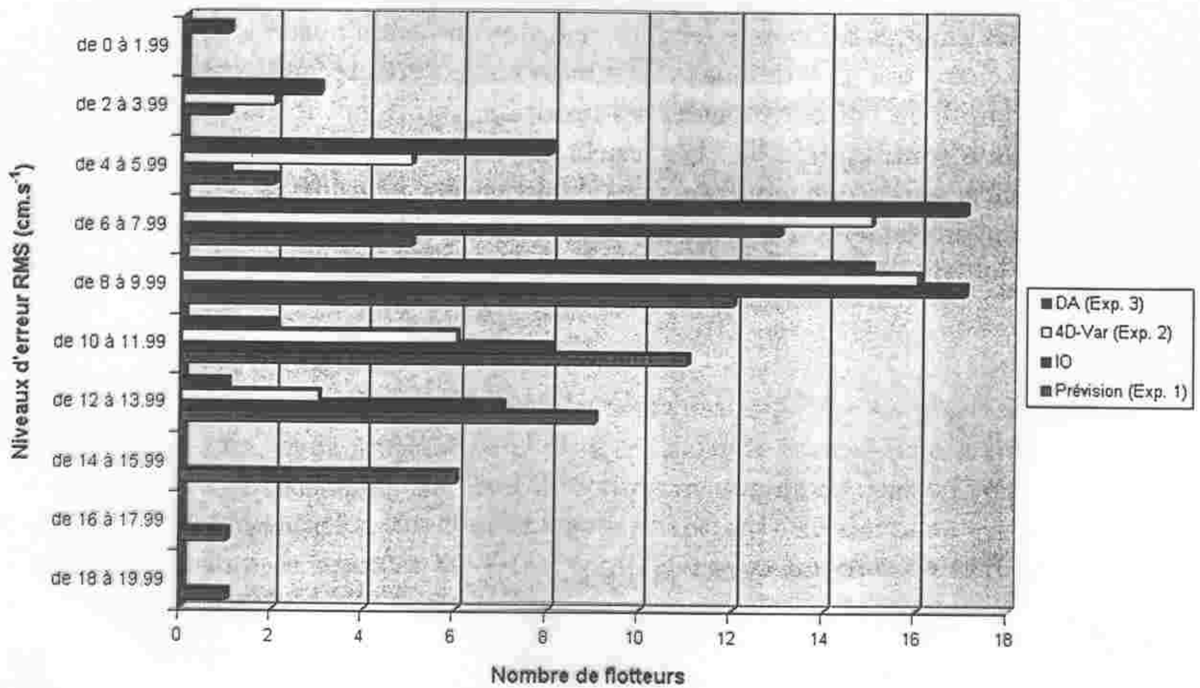


FIG. 9.15 – *Description des écarts flotteurs/modèle en terme de vitesse (à 150 m de profondeur) pour les différents algorithmes d'assimilation*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999

Dual Algorithm for Variational Assimilation (continuation)

Requires

- Explicit background (not much of a problem)

- Exact linearity (much more of a problem). Definition of iterative nonlinear procedures is being studied (Auroux, …)

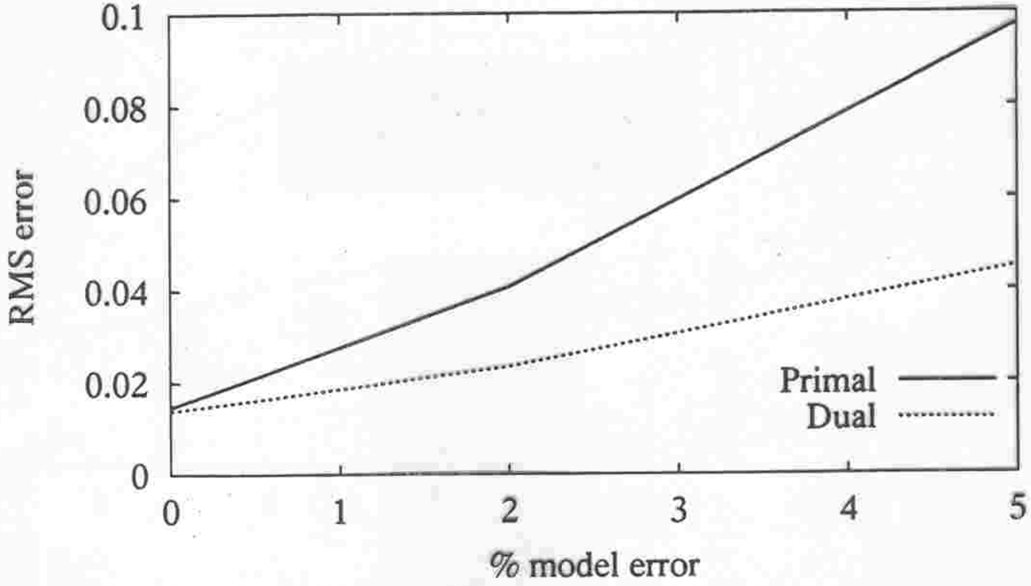


FIG. 6.13 – Normes RMS des erreurs d'assimilation obtenues pour les deux méthodes en fonction de l'erreur introduite dans le modèle au cours de la période d'assimilation.

Auroux, Doctoral Dissertation, Université de Nice-Sophia Antipolis, Nice, 2003

Conclusion on Variational Assimilation

Pros

Carries information both forward and backward in time (important for reassimilation of past data. Kalman smoother also does it).

Can 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.

- Dual approach seems most promising (see also, Fisher, ECMWF). But still needs further development for application in non exactly linear caes.

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