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

Année Universitaire 2014-2015

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 7

4 Mai 2015

Variational assimilation (4DVar)

Available data

- 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
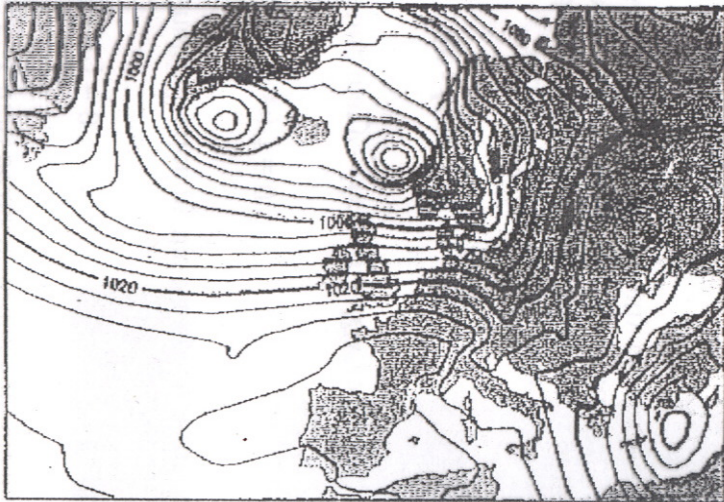
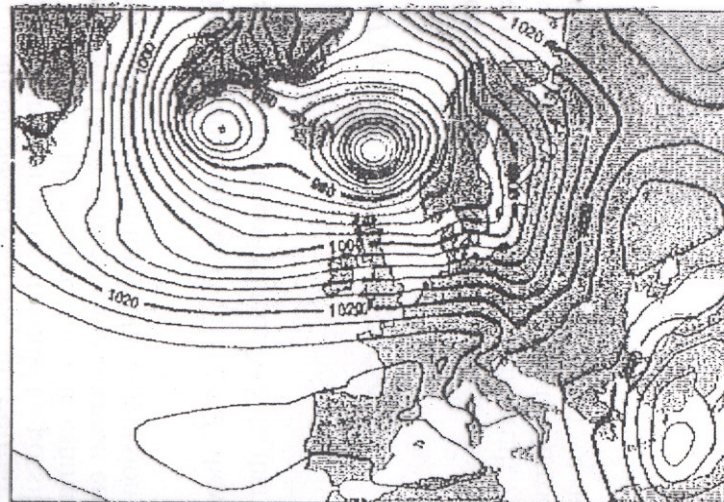Then objective function

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

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

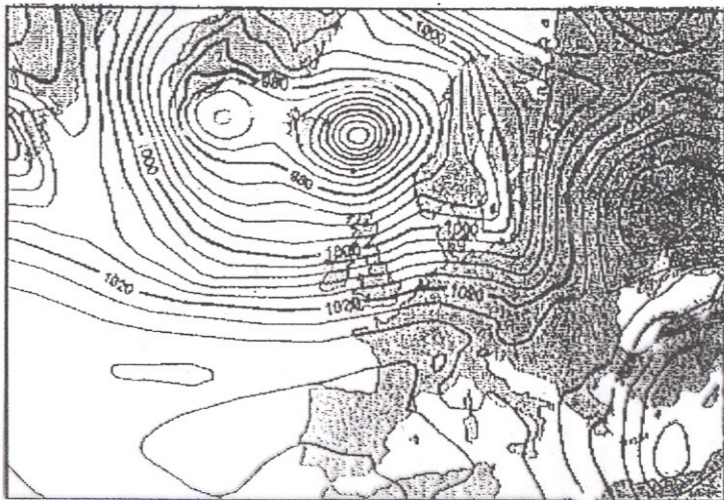subject to $\xi_{k+1} = M_k \xi_k, \qquad\qquad k = 0, \ldots, K\text{-}1$

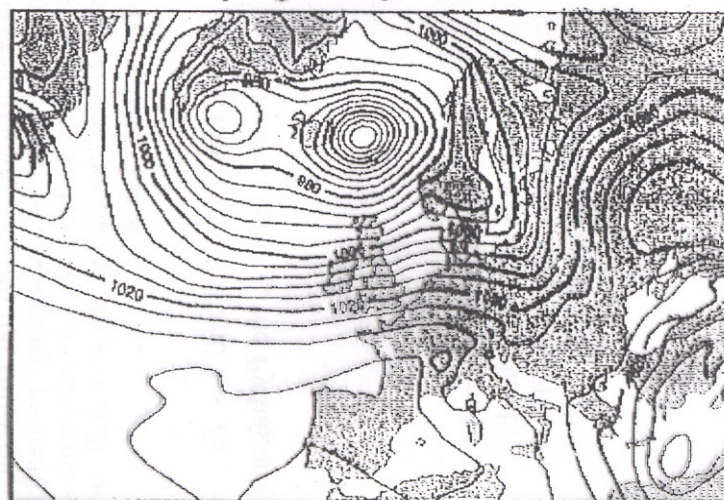3-day forecast from 3D-Var analysis 3-day forecast from 4D-Var analysis
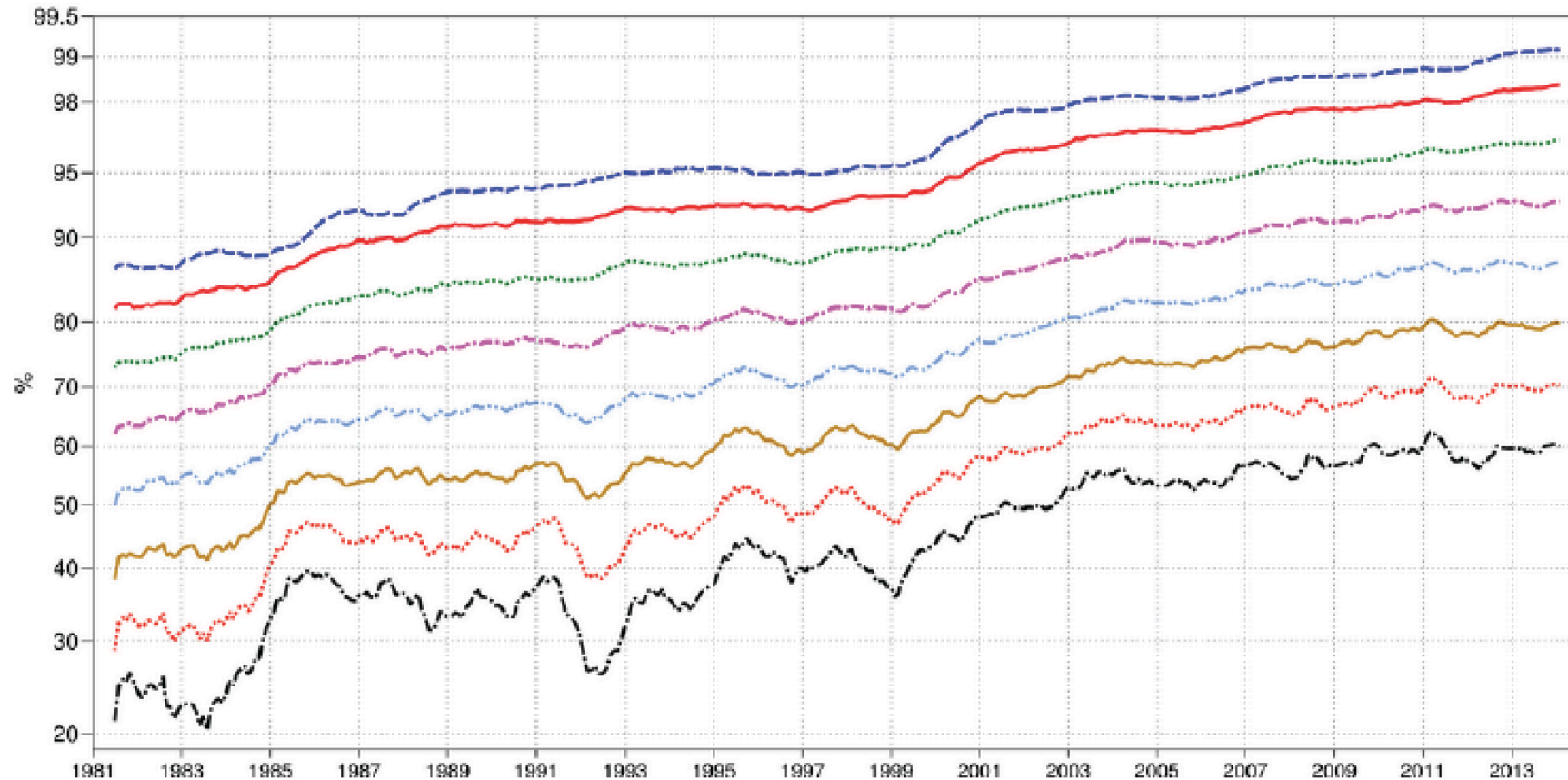
3D-Var verifying analysis 4D-Var verifying analysis

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
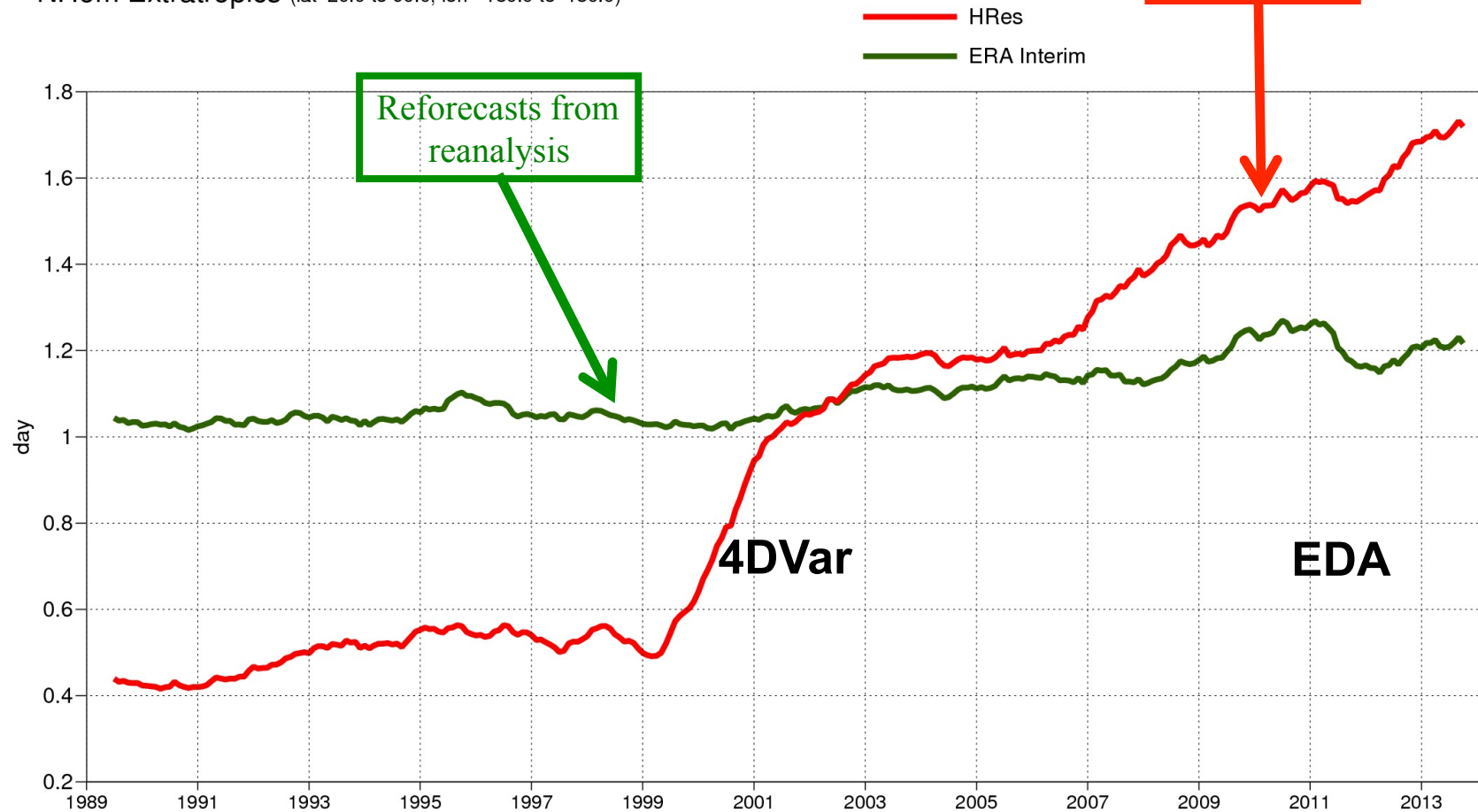
Initial state error reduction

HRes and ERA Interim 00,12UTC forecast skill
500hPa geopotential
Lead time of Anomaly correlation reaching 99.5%
NHem Extratropics (lat 20.0 to 90.0, lon -180.0 to 180.0)

Credit E. Källén, ECMWF

Incremental Method

Variational assimilation, as it has been described, requires the use of the adjoint of the full model.

Simplifying the adjoint as such can be very dangerous. The computed gradient would not be exact, and experience shows that optimization algorithms (and especially efficient ones) are very sensitive to even slight misspecification of the gradient.

Principle of *Incremental Method* (Courtier *et al.*, 1994, *Q. J. R. Meteorol. Soc*.) : simplify simultaneously the (local tangent linear) dynamics and the corresponding adjoint.

Incremental Method (continuation 1)

- Basic (nonlinear) model

$$\xi_{k+1} = M_k(\xi_k)$$

- Tangent linear model

$$\delta\xi_{k+1} = M_k{}' \, \delta\xi_k$$

where $M_k{}'$ is jacobian of M_k at point ξ_k.

- Adjoint model

$$\lambda_k = M_k{}'^{\mathrm{T}} \lambda_{k+1} + \ldots$$

Incremental Method. Simplify $M_k{}'$ and $M_k{}'^{\mathrm{T}}$.

Incremental Method (continuation 2)

More precisely, for given solution $\xi_k^{(0)}$ of nonlinear model, replace tangent linear and adjoint models respectively by

$$\delta\xi_{k+1} = L_k\,\delta\xi_k \qquad\qquad (2)$$

and

$$\lambda_k = L_k^{\mathrm{T}}\,\lambda_{k+1} + \dots$$

where L_k is an appropriate simplification of jacobian M_k'.

It is then necessary, in order to ensure that the result of the adjoint integration is the exact gradient of the objective function, to modify the basic model in such a way that the solution emanating from $\xi_0^{(0)} + \delta\xi_0$ is equal to $\xi_k^{(0)} + \delta\xi_k$, where $\delta\xi_k$ evolves according to (2). This makes the basic dynamics exactly linear.

Incremental Method (continuation 3)

As concerns the observation operators in the objective function, a similar procedure can be implemented if those operators are nonlinear. This leads to replacing $H_k(\xi_k)$ by $H_k(\xi_k^{(0)}) + N_k \delta\xi_k$, where N_k is an appropriate 'simple' linear operator (possibly, but not necessarily, the jacobian of H_k at point $\xi_k^{(0)}$). The objective function depends only on the initial $\delta\xi_0$ deviation from $\xi_0^{(0)}$, and reads

$$\mathcal{J}_I(\delta\xi_0) = (1/2)\,(x_0^b - \xi_0^{(0)} - \delta\xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0^{(0)} - \delta\xi_0)$$
$$+\,(1/2)\,\Sigma_k[d_k - N_k\delta\xi_k]^T\,R_k^{-1}\,[d_k - N_k\delta\xi_k]$$

where $d_k \equiv y_k - H_k(\xi_k^{(0)})$ is the innovation at time k, and the $\delta\xi_k$ evolve according to

$$\delta\xi_{k+1} = L_k\,\delta\xi_k \qquad\qquad (2)$$

With the choices made here, $\mathcal{J}_I(\delta\xi_0)$ is an exactly quadratic function of $\delta\xi_0$. The minimizing perturbation $\delta\xi_{0,m}$ defines a new initial state $\xi_0^{(1)} \equiv \xi_0^{(0)} + \delta\xi_{0,m}$, from which a new solution $\xi_k^{(1)}$ of the basic nonlinear equation is determined. The process is restarted in the vicinity of that new solution.

Incremental Method (continuation 4)

This defines a system of two-level nested loops for minimization. Advantage is that many degrees of freedom are available for defining the simplified operators L_k and N_k, and for defining an appropriate trade-off between practical implementability and physical usefulness and accuracy. It is the incremental method which, together with the adjoint method, makes variational assimilation possible.

First-Guess-At-the-right-Time 3D-Var (FGAT 3D-Var). Corresponds to $L_k = I_n$. Assimilation is four-dimensional in that observations are compared to a first-guess which evolves in time, but is three-dimensional in that no dynamics other than the trivial dynamics expressed by the unit operator is present in the minimization.

Weak constraint variational assimilation

Allows for errors in the assimilating model

- Data

  - Background estimate at time 0

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

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

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

  - Model

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

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

Then objective function

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

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

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

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

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

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 $Q_k \rightarrow 0$

Dual Algorithm for Variational Assimilation (aka *Physical Space Analysis System*, *PSAS,* pronounced *'pizzazz';* see in particular book and papers by Bennett)

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

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

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

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

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

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

or, equivalently, but more conveniently, as

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

where, as before

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

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

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

It is associated with error covariance matrix

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

Dual Algorithm for Variational Assimilation (continuation 3)

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

$$\xi \to H\xi = (u_0^{\mathrm{T}}, \ldots, u_K^{\mathrm{T}})^{\mathrm{T}}$$

is defined by the sequence of operations

$$u_0 = H_0\xi_0$$

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

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

The observation error covariance matrix is equal to

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

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 $\qquad \chi_K = 0$

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

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

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

- Step 3. Multiplication by H. Apply the process defined above on the 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

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

- 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)

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 (Courtier, Louvel)
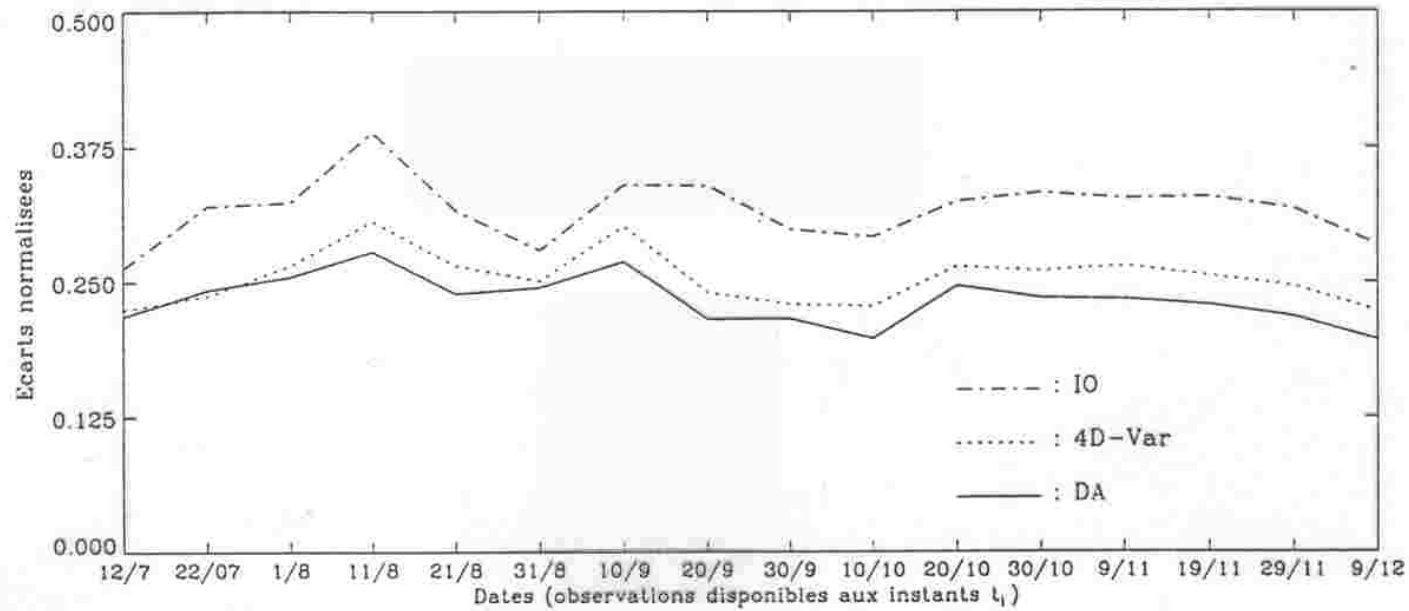
FIG. 9.11 – *Ecarts normalisés prévision/observations sur l'ensemble de la période étudiée*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999
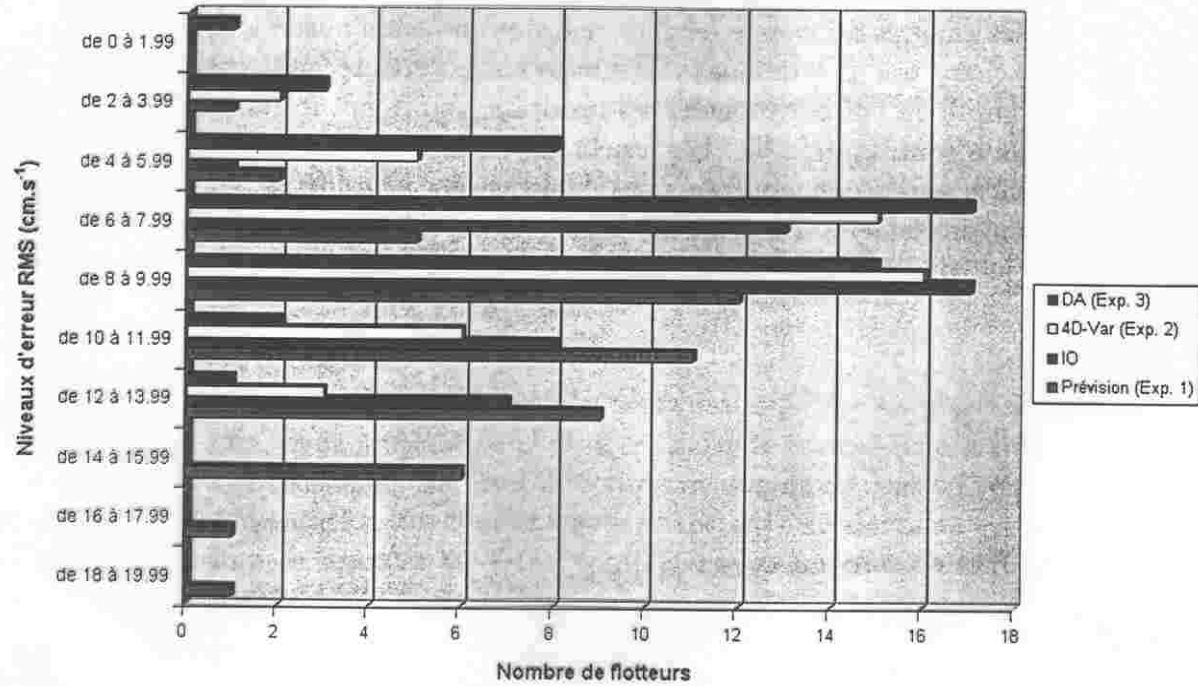
FIG. 9.15 – *Description des écarts flotteurs/modèle en terme de vitesse (à 150 m de profondeur) pour les différents algorithmes d'assimilation*

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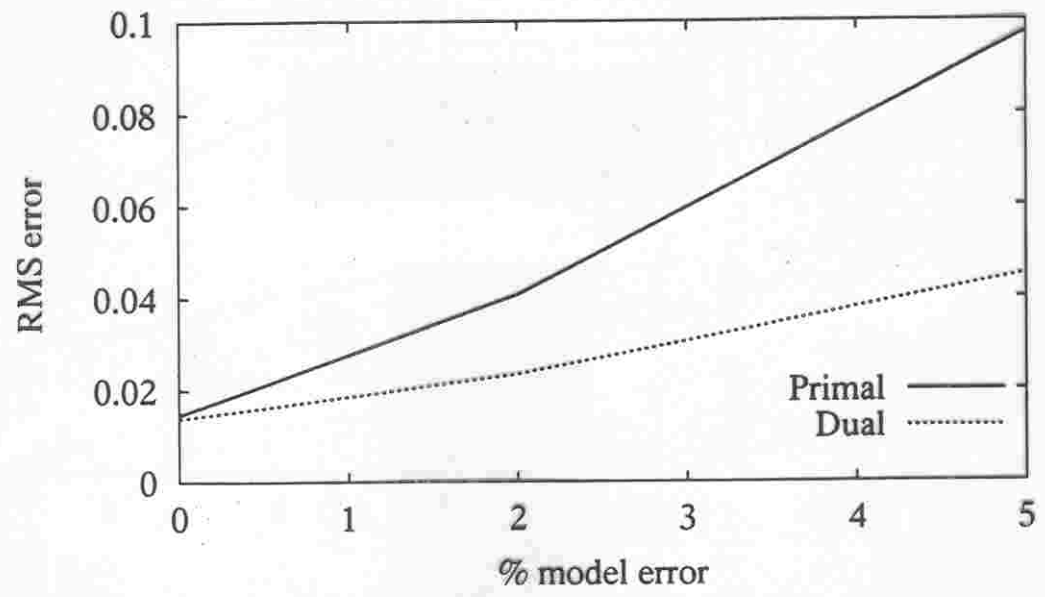
FIG. 6.13 – Normes RMS des erreurs d'assimilation obtenues pour les deux méthodes en fonction de l'erreur introduite dans le modèle au cours de la période d'assimilation.

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

Variational assimilation has been extended to non Gaussian probability distributions (lognormal distributions), the unknown being the mode of the conditional distribution (M. Zupanski, Fletcher).

Bayesian character of variational assimilation ?

- If everything is linear and gaussian, ready recipe for obtaining bayesian sample

Perturb data (background, observations and model) according to their error probability distributions, do variational assimilation, and repeat process

 Sample of system orbits thus obtained is bayesian

- If not, very little can be said at present

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.

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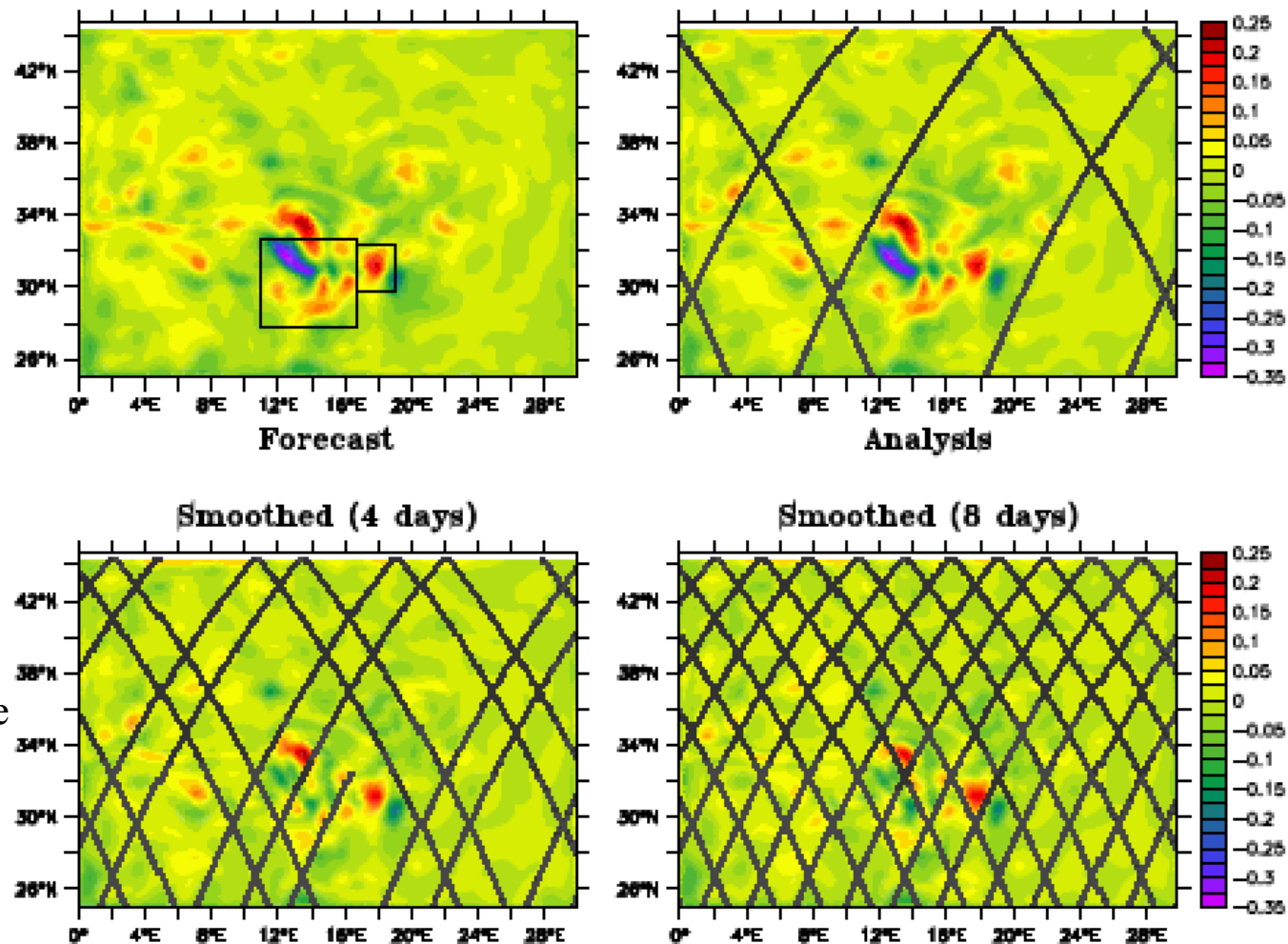
E. Cosme, HDR, 2015, Lissage d'ensemble SEEK

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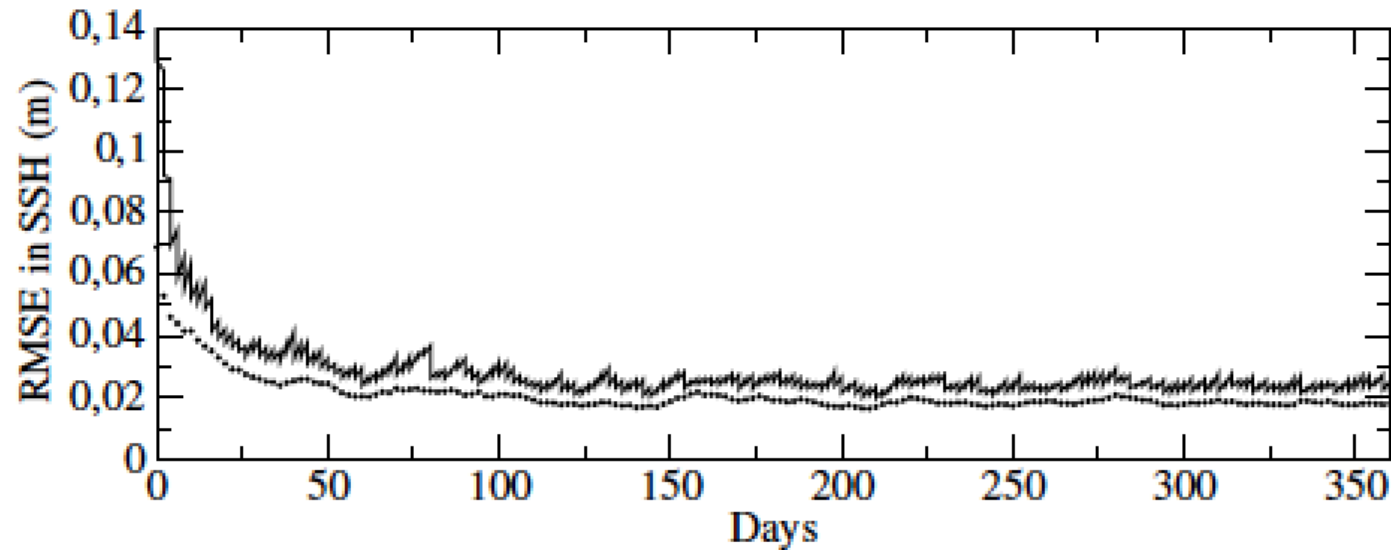
FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse) ; Points : lisseur à retard fixe de 8 jours.

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

Conclusion on Sequential Assimilation

Pros

'Natural', and well adapted to many practical situations

Provides, at least relatively easily, explicit estimate of estimation error

Cons

Carries information only forward in time (of no importance if one is interested only in doing forecast)

In present form, optimality is possible only if errors are independent in time

Conclusion on Variational Assimilation

Pros

Carries information both forward and backward in time (important for reassimilation of past data).

Can easily 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. But the latter can have other uses (sensitivity studies).

- Dual approach seems most promising. But still needs further development for application in non exactly linear cases.

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

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

$\partial J/\partial b = (\partial J/\partial a)(\partial a/\partial b) + (\partial J/\partial b')(\partial b'/\partial b) + (\partial J/\partial c')(\partial c'/\partial b)$
                c 1 0
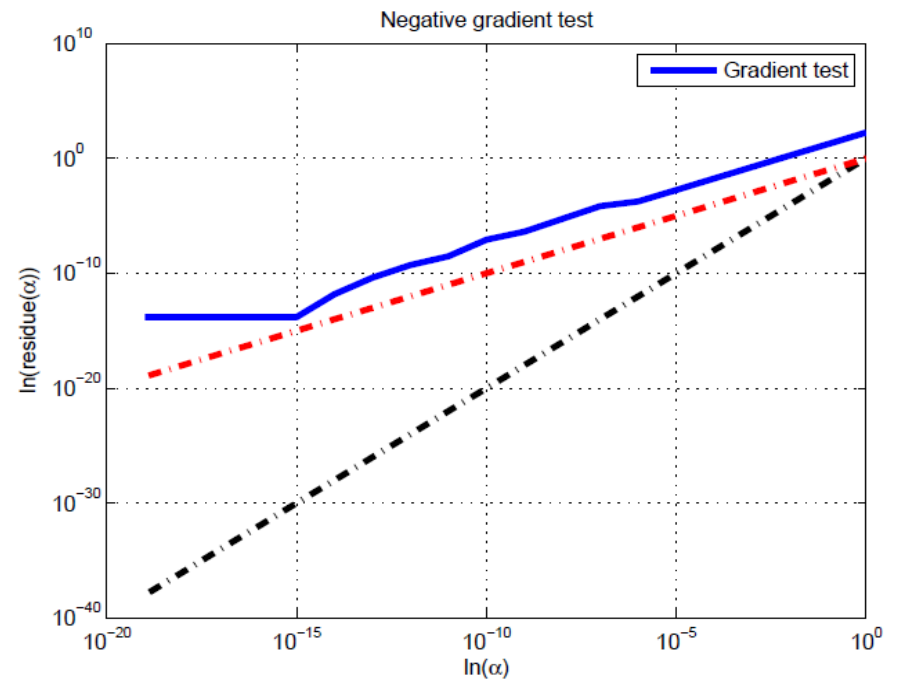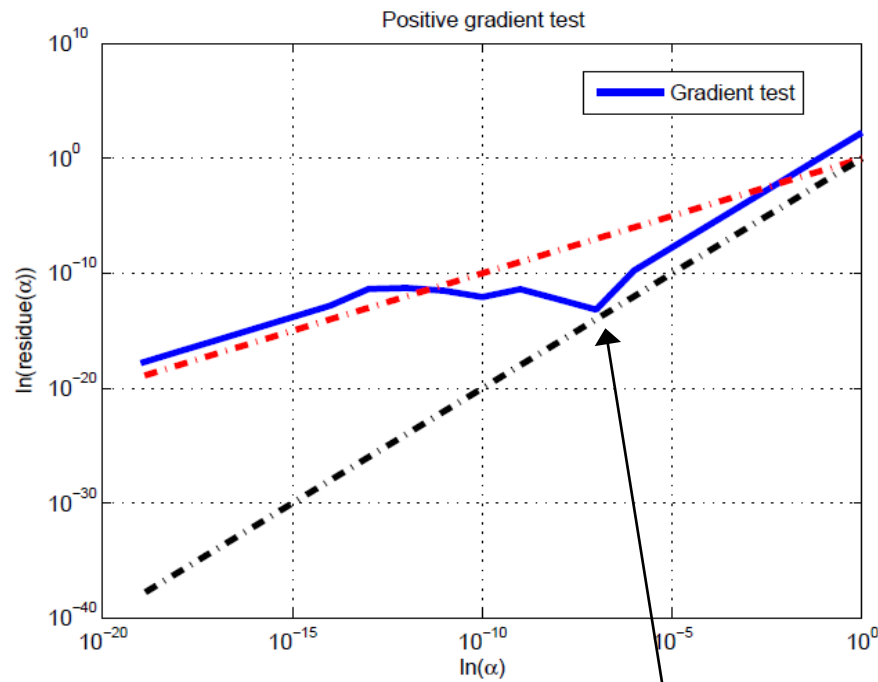
$\partial J/\partial b = (\partial J/\partial a)\, c + \partial J/\partial b'$

Similarly

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

Gradient test

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

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

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

M. Jardak