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

Année Universitaire 2021-2022

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 6

28 Avril 2022

Last course (April 21)

- Assimilation variationnelle. Principe

- Méthode adjointe. Principe.

- Assimilation variationnelle. Résultats

- ~~La Méthode incrémentale~~

- Assimilation à contrainte faible

This course

- Variational Assimilation. Complements
  - Incremental Method
  - Mahalanobis Norm
  - How to write (and validate) an adjoint code
  - Value of objective function at minimum. χ^2 test.

- Assimilation and (In)stability. Quasi-Static Variational Assimilation

Incremental Method for Variational Assimilation

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 both M_k' and $M_k'^{\mathrm{T}}$ consistently.

Incremental Method (continuation 2)

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

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

and

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

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 $\delta\xi_k$ evolves according to

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

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.

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 a strictly sequential assimilation (*i.e.*, any individual piece of information is discarded once it has been used), optimality is possible only if errors are uncorrelated 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 little used.

- Can be implemented in ensemble form (see course 7).

Bayesian Estimation (*see course 5*)

Data of the form

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

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

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

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 $\mathcal{X}$

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

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

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

$S = \mathrm{E}(\zeta\zeta^{T})$ is covariance matrix of data error ζ

Consider quantity $D = z_1^{T} S^{-1} z_2 = z_1^{T} [\mathrm{E}(\zeta\zeta^{T})]^{-1} z_2$

where z_1 and z_2 are any two vectors in data space

Change of coordinates $z \equiv Tw$

$$\zeta = T\chi \quad \Rightarrow S = \mathrm{E}(\zeta\zeta^{T}) = \mathrm{E}[T\chi(T\chi)^{T}] = T\,\mathrm{E}(\chi\chi^{T})T^{T}$$

$$D = w_1^{T}\, T^{T}\, [T\,\mathrm{E}(\chi\chi^{T})T^{T}]^{-1} T w_2$$

$$D = w_1^{T}\, [\mathrm{E}(\chi\chi^{T})]^{-1}\, w_2$$

Expression $D = z_1^{T} S^{-1} z_2$

defines proper scalar product, and associated norm, on data space

Mahalanobis norm

Prasanta Chandra Mahalanobis (1893 -1972)

Objective function

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

$$\mathcal{J}_{min} \equiv \mathcal{J}(\boldsymbol{x}^a) = (1/2)\,[\Gamma\boldsymbol{x}^a - \boldsymbol{z}]^{\mathrm{T}}\, S^{-1}\,[\Gamma\boldsymbol{x}^a - \boldsymbol{z}]$$

$$= (1/2)\,\boldsymbol{d}^{\mathrm{T}}\,[E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}})]^{-1}\,\boldsymbol{d}$$

where $\boldsymbol{d}$ is innovation

$$\Longrightarrow \qquad E(\mathcal{J}_{min}) = p/2 \qquad\qquad (p = \mathrm{dim}\boldsymbol{y} = \mathrm{dim}\boldsymbol{d})$$

If p is large, a few realizations are sufficient for determining $E(\mathcal{J}_{min})$

Remark. If in addition errors are gaussian, the quantity $2E(\mathcal{J}_{min})$ follows a χ^2-probability distribution of order p. For that reason the criterion $E(\mathcal{J}_{min}) = p/2$ is often called the χ^2 criterion. Also $Var(\mathcal{J}_{min}) = p/2$ in the gaussian case.

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

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

How to write the adjoint of a code ?

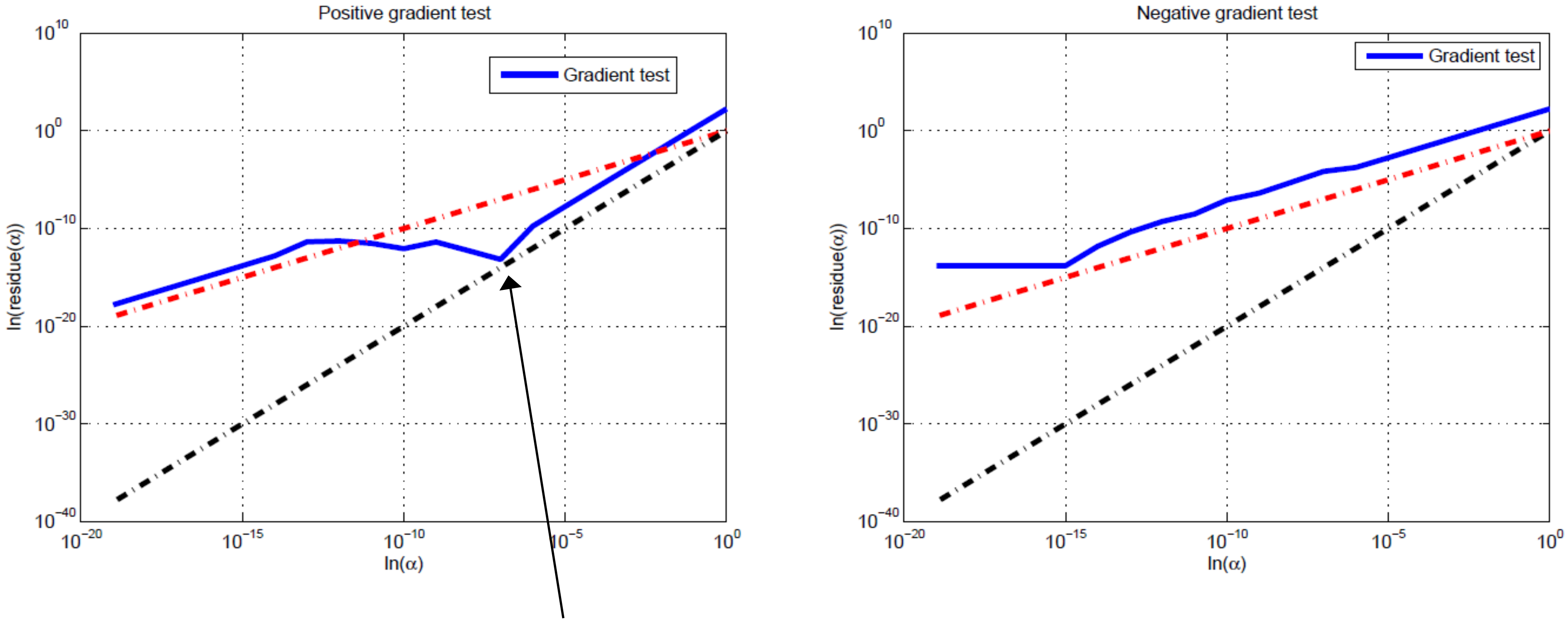
Adjoint compilers

TAPENADE (Laurent Hascoet, Institut national de recherche en informatique et en automatique)

FastOpt AD-Tool (Ralf Giering and Thomas Kaminski)

- ….

Gradient test



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

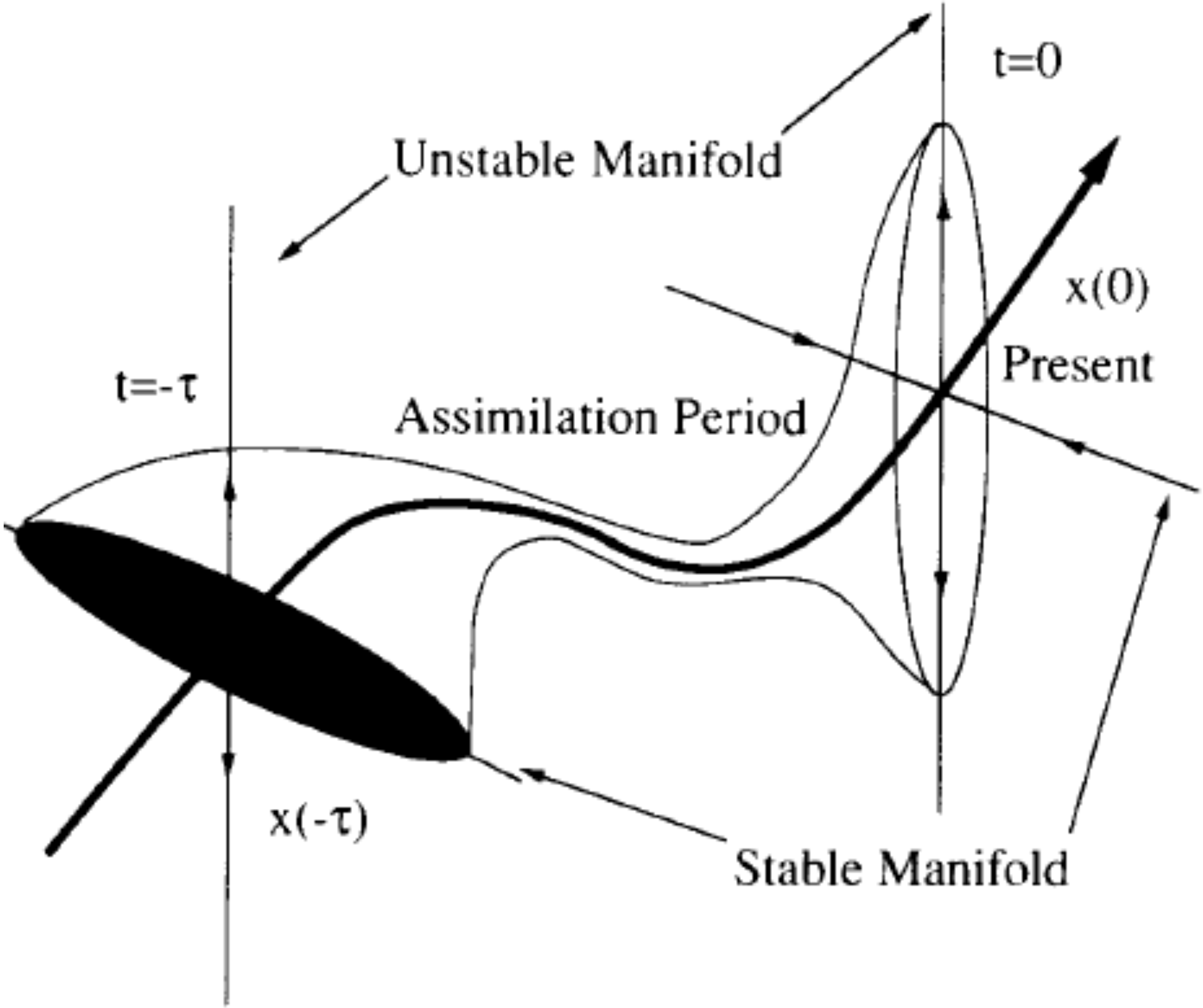
$\epsilon = 2^{-53}$ zero machine

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

- Assimilation and (In)stability

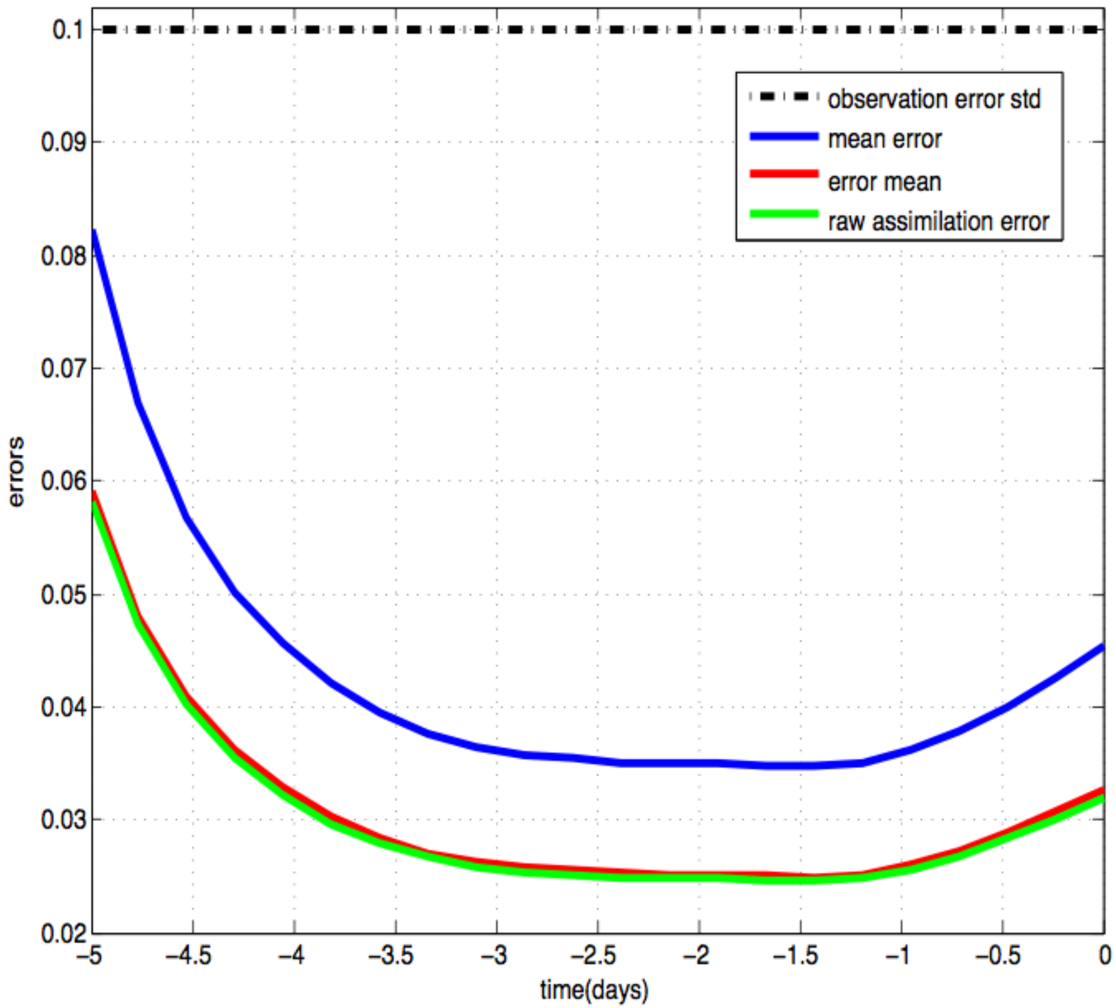
If there is uncertainty on the state of the system, and dynamics of the system is perfectly known, uncertainty on the state along stable modes decreases over time, while uncertainty along unstable modes increases.

Stable (unstable) modes : perturbations to the basic state that decrease (increase) over time.


t=0
Unstable Manifold
x(0)
Present
t=-τ
Assimilation Period
x(-τ)
Stable Manifold

Consequence : Consider 4D-Var assimilation, or any form of smoother, which carries information both forward and backward in time, performed over time interval $[t_0, t_1]$ over uniformly distributed noisy data. If assimilating model is perfect, estimation error is concentrated in stable modes at time t_0, and in unstable modes at time t_1. Error is smallest somewhere within interval $[t_0, t_1]$.

Similar result holds true for Kalman filter (or more generally any form of sequential assimilation), in which estimation error is concentrated in unstable modes at any time.



Linearized Lorenz'96. 5 days

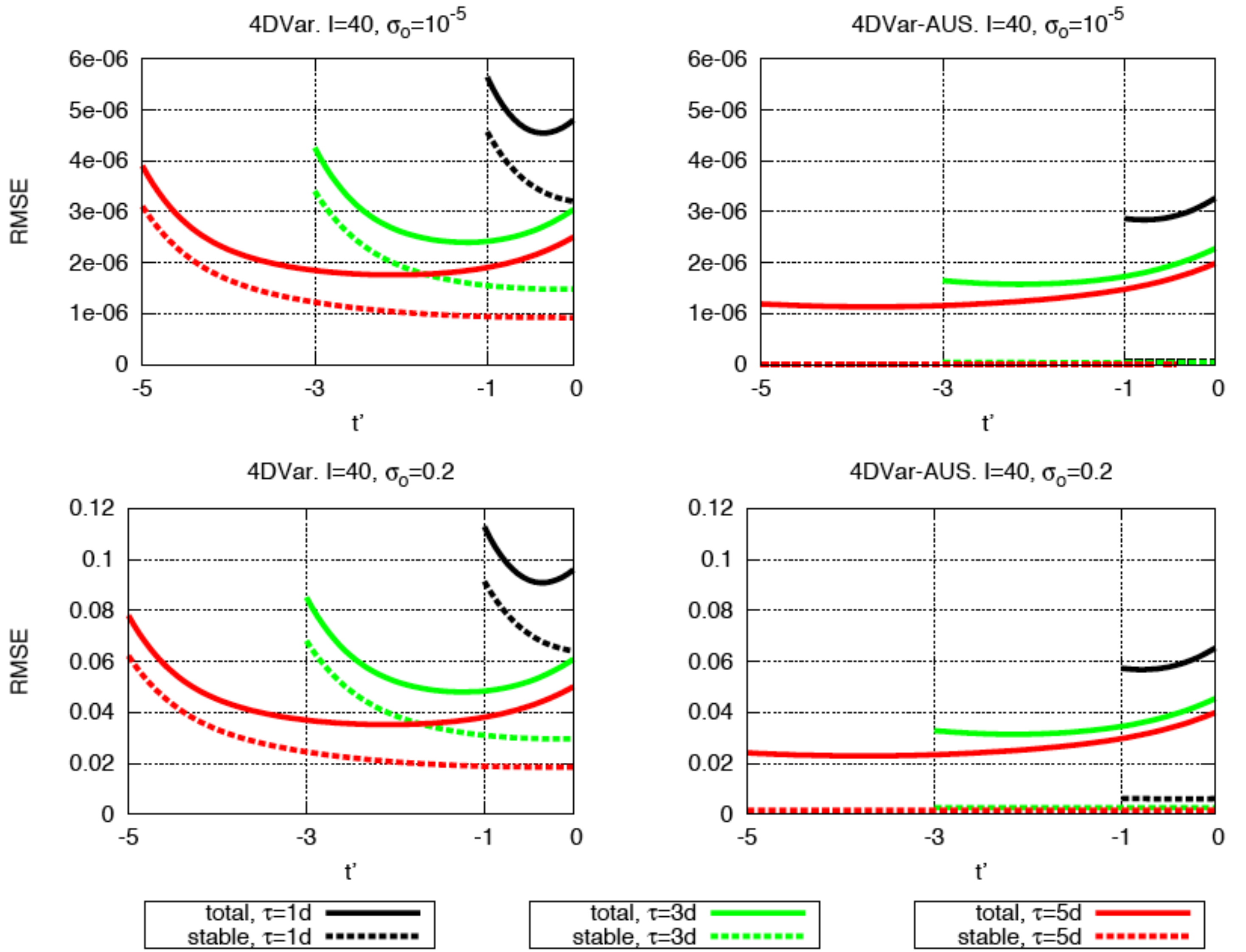


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to the error component in the stable subspace $\mathbf{e}_{16}, \ldots, \mathbf{e}_{40}$.

Trevisan *et al*., 2010, *Q. J. R. Meteorol. Soc.*

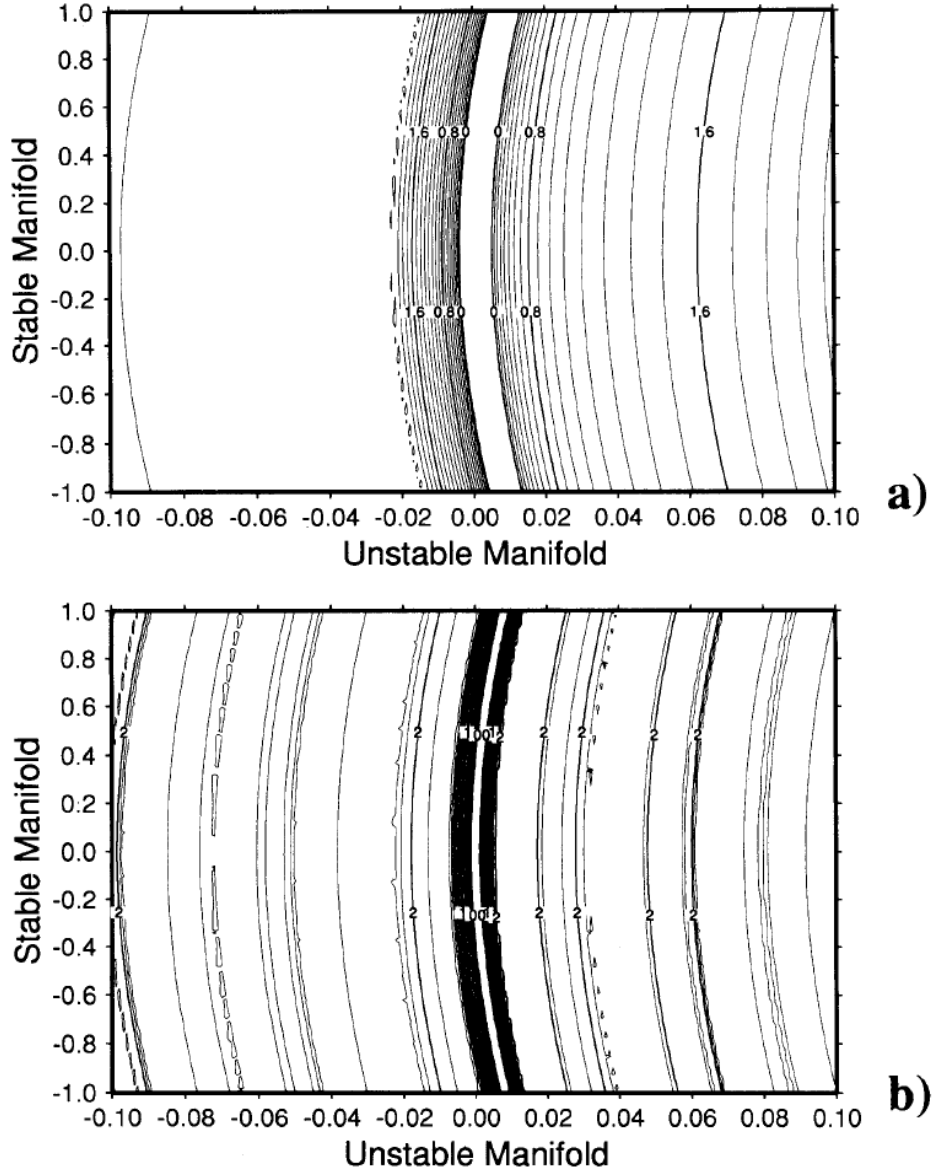


Fig. 3. Variations of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (Lorenz system) in the plane spanned by the stable and unstable directions, as determined from the tangent linear system (see text), and for $\tau = 6$ (panel (a)) and $\tau = 8$ (panel (b)) respectively. The metric has been distorted in order to make the stable and unstable manifolds orthogonal to each other in the figure. The scale on the contour lines is logarithmic (decimal logarithm). Contour interval: 0.1. For clarity, negative contours, which would be present only in the central "valley" directed along the stable manifold, have not been drawn.

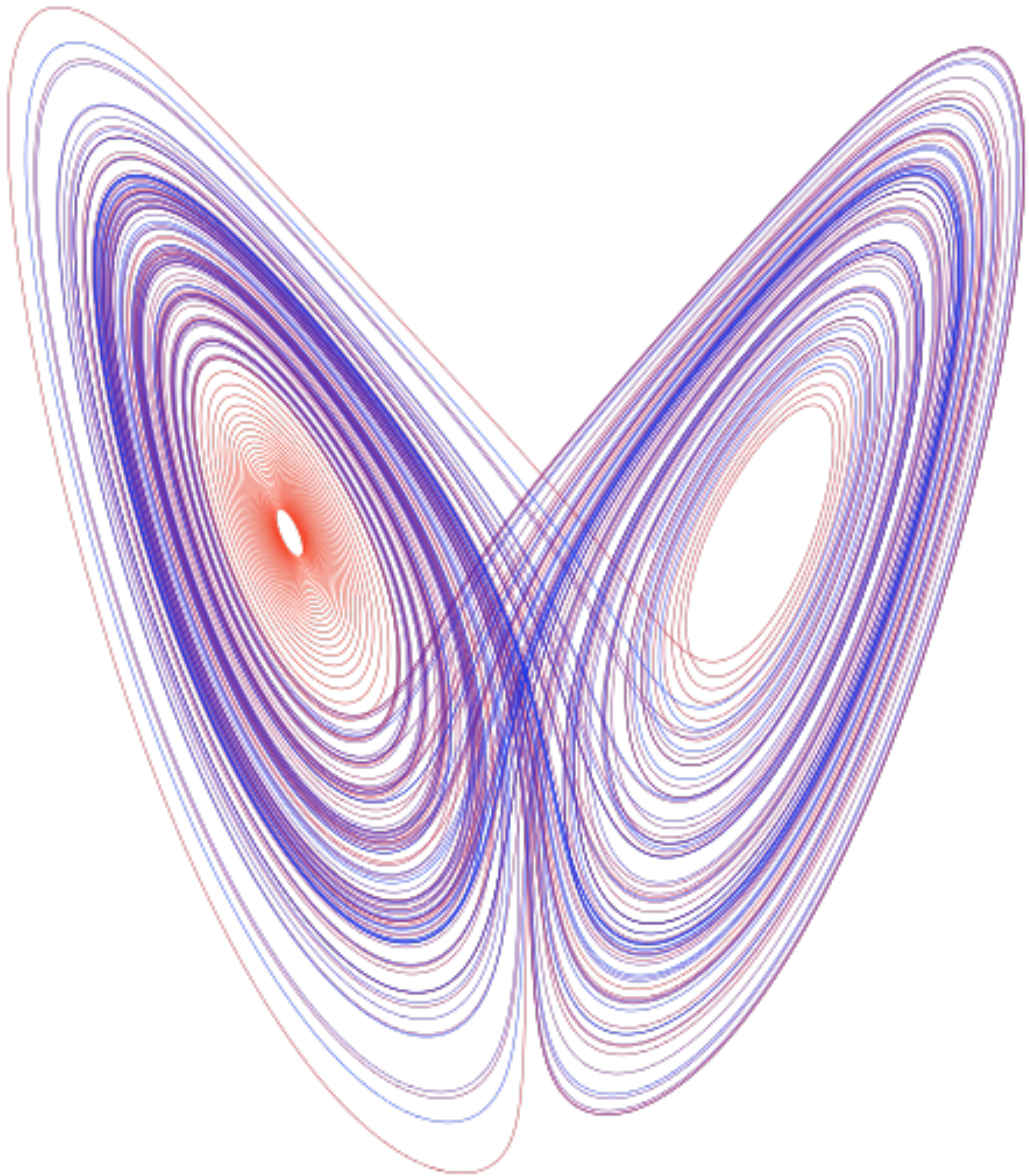
Lorenz (1963)

$$dx/dt = \sigma(y - x)$$

$$dy/dt = \rho x - y - xz$$

$$dz/dt = -\beta z + xy$$

with parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3 \Rightarrow$ chaos



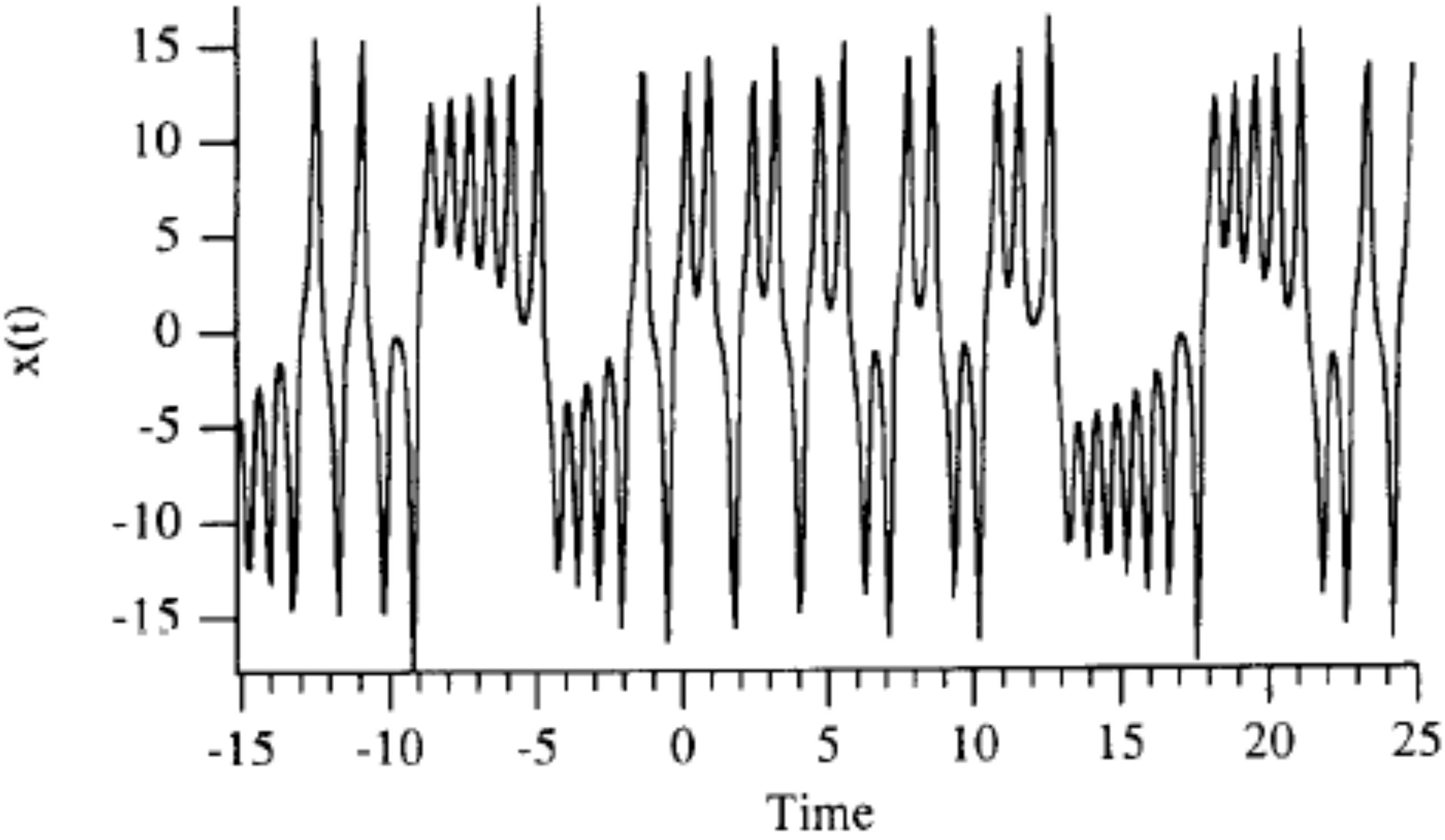


Fig. 2. Time variations, along the reference solution, of the variable $x(t)$ of the Lorenz system.

Twin (strong constraint) experiment. Observations $y_k = H_k x_k + \varepsilon_k$ at successive times k, and objective function of form

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

x_k denotes here the complete state vector, and H_k is the unit operator (all three components of x_k are observed)

No 'background' term from the past, but observation y_0 at time $k = 0$.

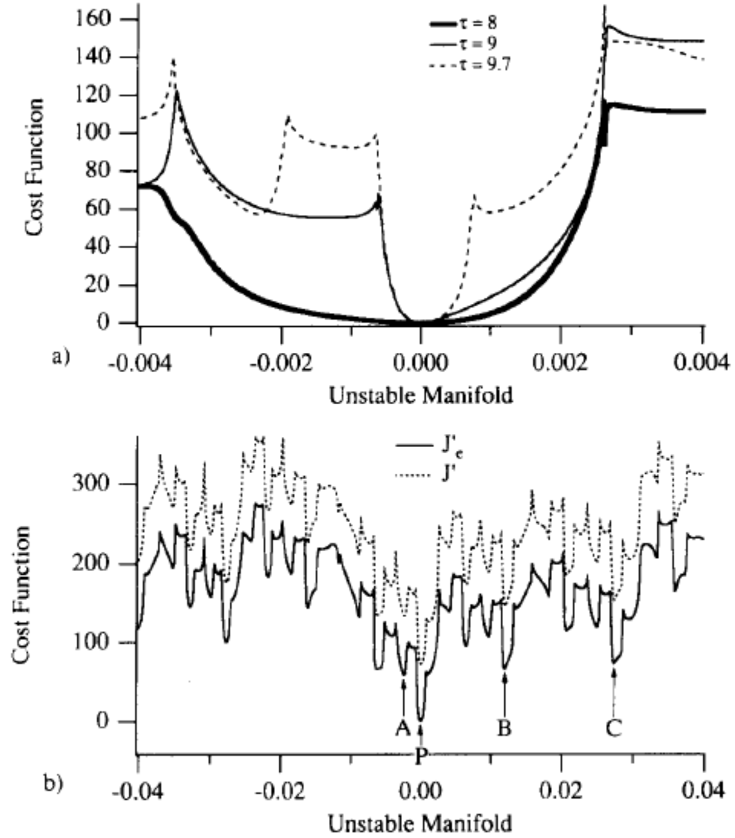


Fig. 4. Panel (a): Cross-section of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ along the unstable manifold, for various values of τ. Panel (b). As in panel (a), for $\tau = 9.7$, and with a display interval ten times as large, respectively for the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (solid curve) and for the error-contaminated cost-function $J_e(\tau, \hat{x}, x)$ (dashed curve). In the latter case, the total variance of the observational noise is $E^2 = 75$.

Pires *et al., Tellus*, 1996 ; Lorenz system (1963)

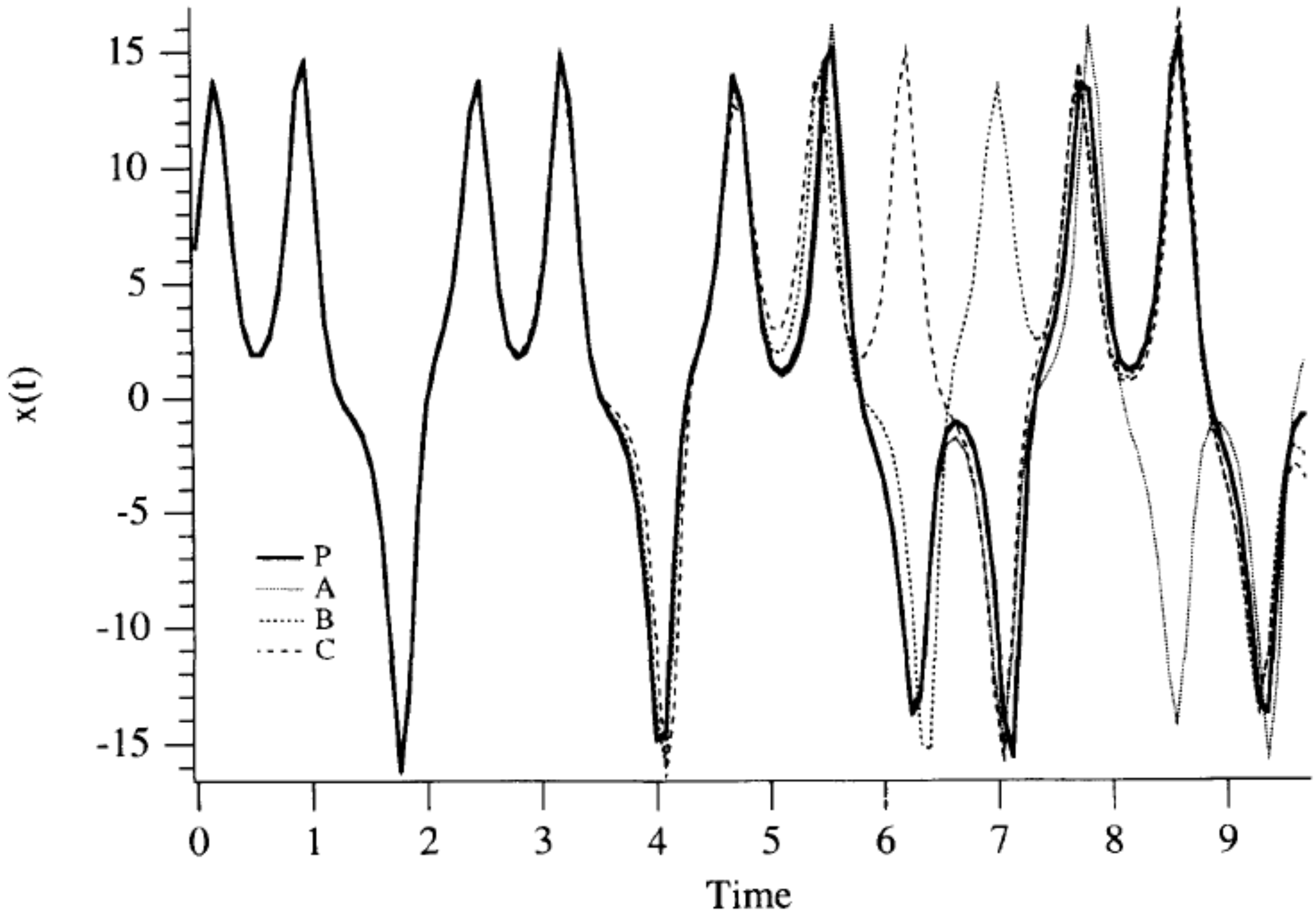


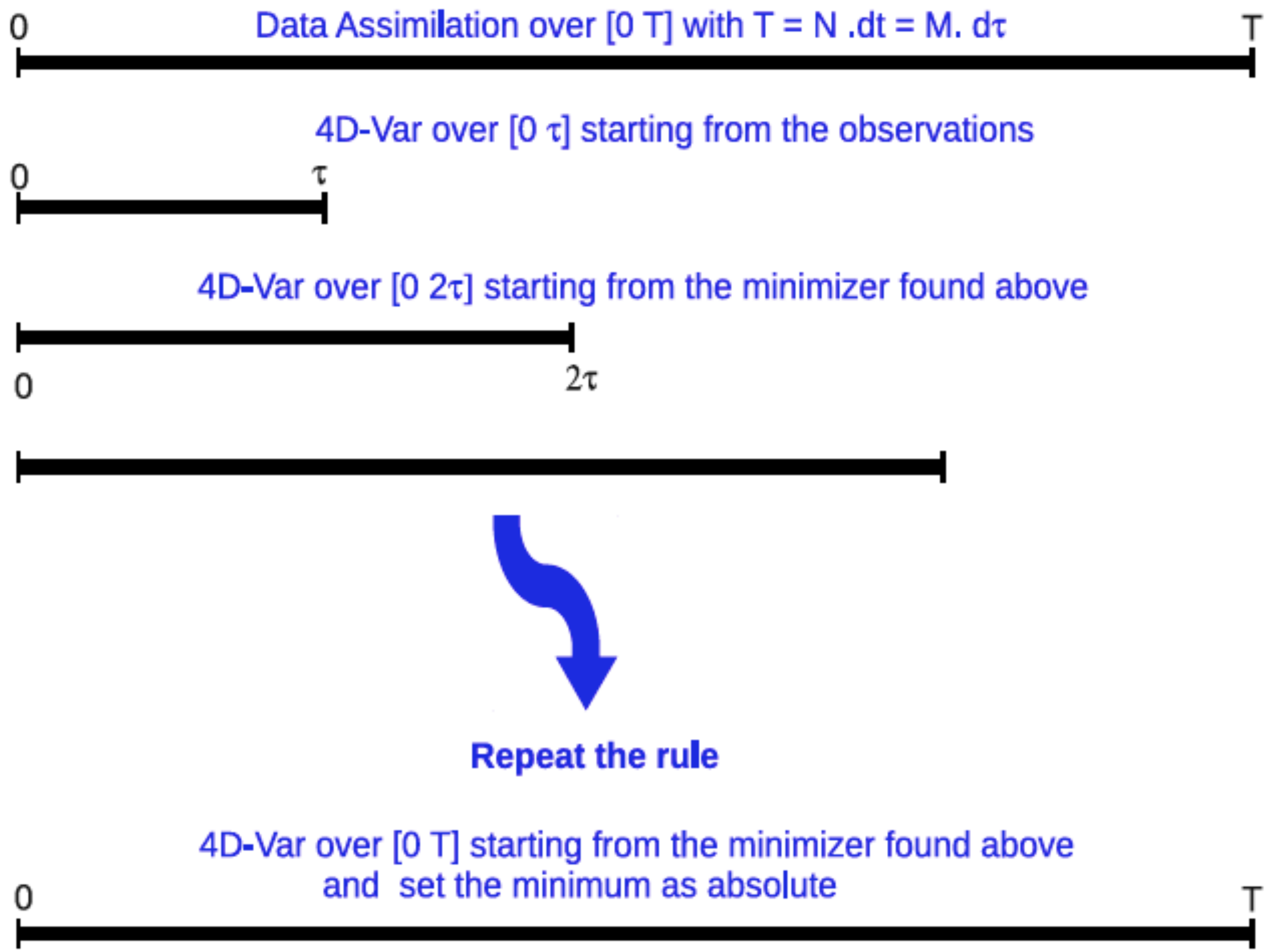
Fig. 5. Variations of the coordinate x along the orbits originating from the minima *P*, *A*, *B*, *C* (indicated in Fig. 4b) of the error-free cost-function.

Minima in the variations of objective function correspond to solutions that have bifurcated from the observed solution, and to different folds in state space.

Quasi-Static Variational Assimilation (*QSVA*). Increase progressively length of the assimilation window, starting each new assimilation from the result of the previous one. This should ensure, at least if observations are in a sense sufficiently dense in time, that current estimation of the system always lies in the attractive basin of the absolute minimum of objective function (Pires *et al*., Swanson *et al*., Luong, Järvinen *et al*.)

.

Quasi-Static Variational Assimilation (QSVA)



$\mu(C(\tau, x))$	Cloud of points QSVA	Cloud of points raw assimilation	Linear tangent system	Upper bound
$\tau = 0$	1	1	1	1
$\tau = 1$	0.36	0.37	0.39	0.46
$\tau = 2$	5.9×10^{-2}	5.74	4.5×10^{-2}	0.401
$\tau = 3$	3.3×10^{-2}	29.4	2.9×10^{-2}	0.397
$\tau = 8$	1.4×10^{-2}	59.9	*	0.396

In the left column, the estimates are calculated from the ensemble of 100 assimilations (see also Fig. 7). The 2nd column contains the values obtained from the raw assimilation. In the 3rd column, the estimates are obtained from the tangent linear system and eqs. (3.5–3.9) (the star indicates a computational overflow). The estimates in the right-hand column are the upper bounds defined by eq. (3.13).

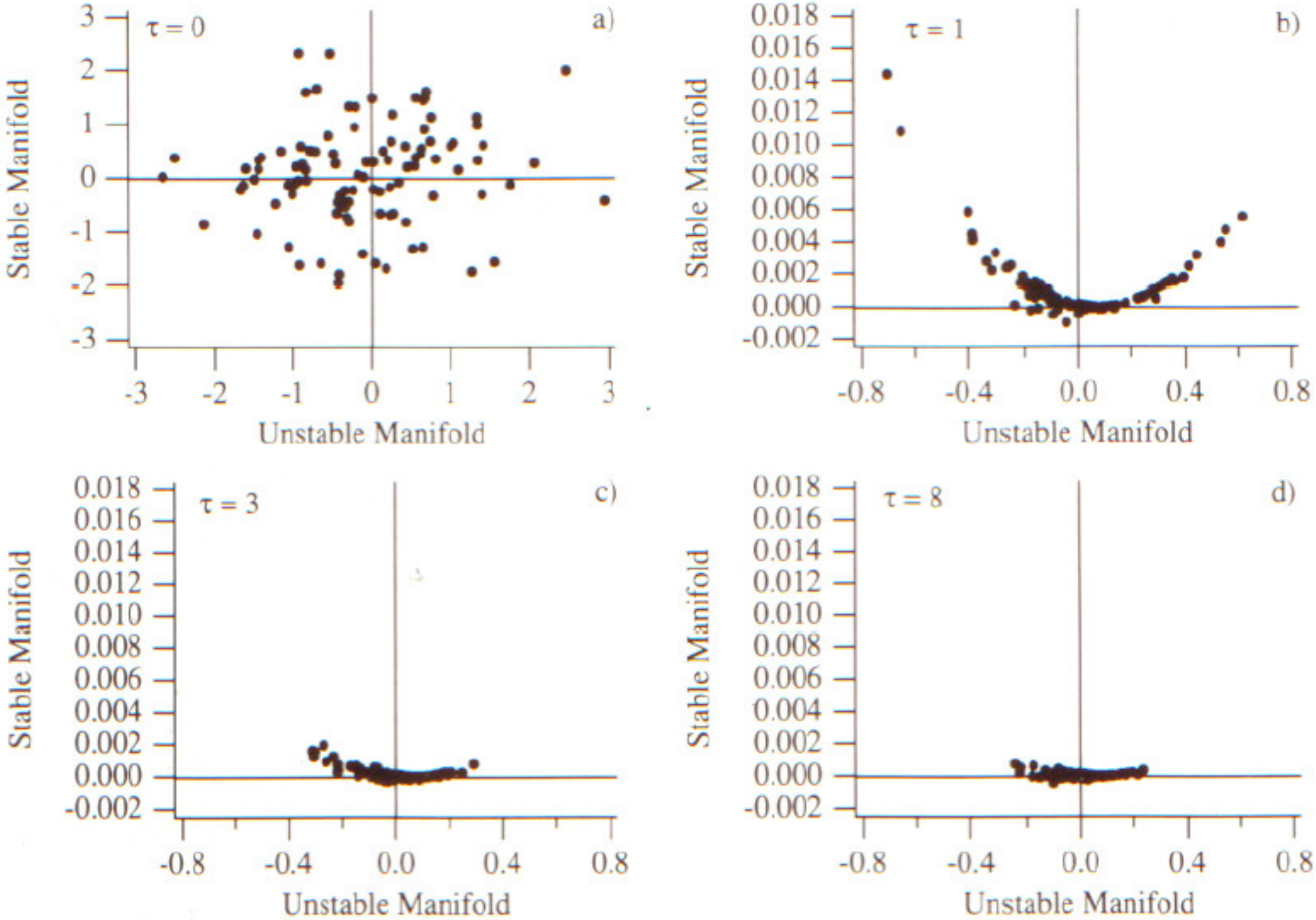


Fig. 7. Projection of the 100 minimizing solutions, at the end of the assimilation period, onto the plane spanned by the stable and unstable directions, defined as in Fig. 3. Values of τ are indicated on the panels. The projection is not an orthogonal projection, but a projection parallel to the local velocity vector $(dx/dt, dy/dt, dz/dt)$ (central manifold).

Pires *et al., Tellus*, 1996 ; Lorenz system (1963)

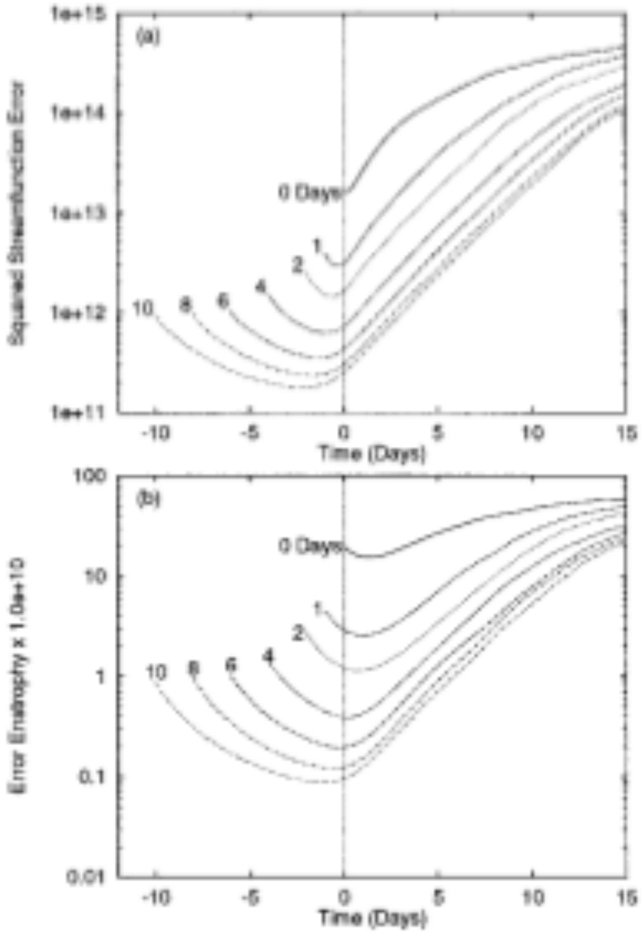


Fig. 5. Median values of the (a) streamfunction squared error, and (b) enstrophy error for the 200 forecast set as a function of forecast time and of the assimilation time T_a.

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

Cours à venir

~~Jeudi 17 mars~~
~~Jeudi 24 mars~~
~~Jeudi 31 mars~~
~~Jeudi 14 avril~~
~~Jeudi 21 avril~~
~~Jeudi 28 avril~~
Jeudi 5 mai
Jeudi 12 mai