Predictability and assimilation of observations in chaotic systems

O. Talagrand
Laboratoire de Météorologie Dynamique, École Normale Supérieure, Paris, France

Workshop
Exploring Complex Dynamics in High-Dimensional Chaotic Systems: From Weather Forecasting to Oceanic Flows

Max-Planck-Institut für Physik komplexer Systeme, Dresden, Germany
25 January 2010

With acknowledgments to F.-X. Le Dimet, C. Pires, R. Vautard, A. Trevisan and many others.

Assimilation of observations, as it is known in meteorology and oceanography, originated from the need of defining initial conditions (ICs) for numerical weather prediction. Difficulties progressively arose

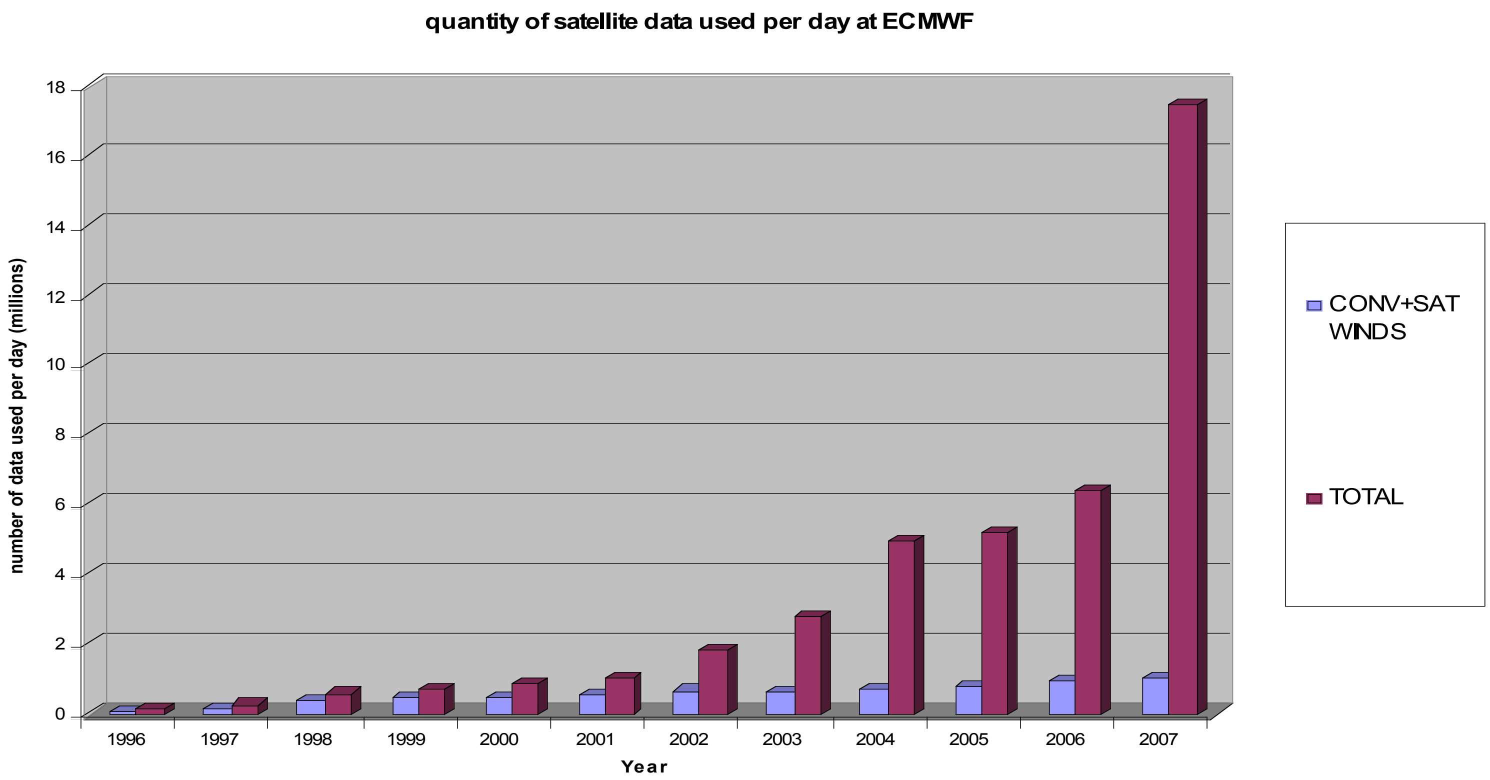
- Need for defining ICs with appropriate spatial scales ⇒ '*structure functions*' (now incorporated in background error covariance matrices)

- Need for defining ICs in approximate geostrophic balance ⇒ '*initialization*' (now also incorporated in background error covariance matrices)

- Realization that useful information was present in recent forecast ⇒ use of a background (word *assimilation* was coined in 1967-68)

- Use of satellite observations, which are
  - distributed continuously in time
  - indirect ⇒ need for some form of 'inversion'

December 2007: Satellite data volumes used: around 18 millions per day



European Centre for Medium-range Weather Forecasts
(ECMWF, Reading, UK)

Horizontal spherical harmonics triangular truncation T799 (horizontal resolution ≈ 28 kilometres)

91 levels on the vertical (average resolution 400 m)

Dimension of state vector $n \approx 2.3\ 10^8$

Timestep = 12 minutes

Purpose of assimilation : reconstruct as accurately as possible the state of the atmospheric or oceanic flow, using all available appropriate information. The latter essentially consists of

- The observations proper, which vary in nature, resolution and accuracy, and are distributed more or less regularly in space and time.

- The physical laws governing the evolution of the flow, available in practice in the form of a discretized, and necessarily approximate, numerical model.

- ‘Asymptotic’ properties of the flow, such as, *e*. *g*., geostrophic balance of middle latitudes. Although they basically are necessary consequences of the physical laws which govern the flow, these properties can usefully be explicitly introduced in the assimilation process.

Assimilation is one of many ‘*inverse problems*’ encountered in many fields of science and technology

- solid Earth geophysics
- plasma physics
- ‘nondestructive’ probing
- navigation (spacecraft, aircraft, ….)
- …

Solution most often (if not always) based on bayesian, or probabilistic, estimation. ‘Equations’ are fundamentally the same.

Difficulties specific to assimilation of meteorological observations :

- Very large numerical dimensions ($n \approx 10^6$-10^8 parameters to be estimated, $p \approx$ 1-2.10^7 observations per 24-hour period). Difficulty aggravated in Numerical Weather Prediction by the need for the forecast to be ready in time.

- Non-trivial, actually chaotic, underlying dynamics

Both observations and 'model' are affected with some uncertainty $\Rightarrow$ uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don't know too well why, but it works).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (unambiguously defined if a prior probability distribution is defined; see Tarantola, 2005).

Bayesian estimation is however impossible in its general theoretical form in meteorological or oceanographical practice because

- It is impossible to explicitly describe a probability distribution in a space with dimension even as low as $n \approx 10^3$, not to speak of the dimension $n \approx 10^{6\text{-}8}$ of present Numerical Weather Prediction models.

- Probability distribution of errors on data very poorly known (model errors in particular).

One has to restrict oneself to a much more modest goal. Two approaches exist at present

- Obtain some 'central' estimate of the conditional probability distribution (expectation, mode, …), plus some estimate of the corresponding spread (standard deviations and a number of correlations).

- Produce an ensemble of estimates which are meant to sample the conditional probability distribution (dimension $N \approx O(10\text{-}100)$).

Proportion of resources devoted to assimilation in Numerical Weather Prediction has steadily increased over time.

At ECMWF, as much computing time devoted now to 24-hour assimilation as to 10-day high resolution forecast.

A large part of 'real life' assimilation algorithms still based on heuristic extension to nonlinear situations of statistical linear estimation

Data in the form

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

Known data vector z belongs to *data space* $\mathcal{D}$, $dim\mathcal{D} = m$,

Unknown state vector belongs to *state space* $\mathcal{S}$, $dim\mathcal{S} = n$

Γ known (mxn)-matrix, ζ unknown 'error'

Look for estimated state vector x^a of the form

$$x^a = \alpha + Az$$

subject to

- invariance in change of origin in state space $\Rightarrow$ $A\Gamma = I_m$
- quadratic estimation error $E[(x^a_i - x_i)^2]$ minimum for any component x_i.

Solution

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

$P^a \equiv E[(x^a - x)(x^a - x)^T] = (\Gamma^T S^{-1} \Gamma)^{-1}$

where $\mu \equiv E(\zeta)$

$S \equiv E\{[\zeta - \mu][\zeta - \mu]^T\}$

Best Linear Unbiased Estimator (*BLUE*) of x from z.

Requires *a priori* explicit knowledge of $E(\zeta)$ and $E\{[\zeta - E(\zeta)][\zeta - E(\zeta)]^T\}$

Unambiguously defined iff rank$\Gamma = n$. *Determinacy condition*. Requires $m \geq n$.

In case ζ is gaussian, $\zeta \sim \mathcal{N}[\mu, S]$, *BLUE* achieves bayesian estimation in the sense that $P(x \mid z) = \mathcal{N}[x^a, P^a]$

Perturbed data vector

$$z_p = z + \zeta_p \qquad , \qquad \zeta_p \sim \mathcal{N}[0, S]$$

and associated *BLUE*

$$x^a{}_p = (\Gamma^T S^{-1} \Gamma)^{-1} \Gamma^T S^{-1} [z_p - \mu]$$

$x^a{}_p$ is distributed according to conditional pdf $\mathcal{N}[x^a, P^a]$.

This defines a ready recipe for obtaining a sample of the conditional pdf. Perturb data vector N times, and independently, and compute corresponding *BLUE*s.

Variational form.

BLUE x^a minimizes following scalar *objective function*, defined on state space $\mathcal{S}$

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

BLUE is invariant in any invertible linear change of coordinates, either in state or data space.

If determinacy condition is verified, it is always possible to decompose data into

- A '*background*' estimate (*e. g.* forecast from the past), belonging to *state space*, with dimension n

$$x^b = x + \zeta^b ; \quad E(\zeta^b) = 0 \quad ; \; E(\zeta^b \zeta^{b\mathrm{T}}) \equiv P^b$$

- An additional set of data (*e. g.* observations), belonging to *observation space*, with dimension $m - n = p$

$$y = Hx + \varepsilon \quad ; \quad E(\varepsilon) = 0 \quad ; \; E(\varepsilon \varepsilon^{\mathrm{T}}) \equiv R$$

with $E(\varepsilon \zeta^{b\mathrm{T}}) = 0$

Then

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

$d \equiv y - Hx^b$ is *innovation vector*.

$$P^a = P^b - P^b H^T [HP^bH^T + R]^{-1} HP^b$$

Variational form

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

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

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

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad k = 0, \ldots, K$$

$$E(\varepsilon_k) = 0 \quad ; \quad E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) \equiv R_k \, \delta_{kj}$$

- Evolution equation

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

$$E(\eta_k) = 0 \quad ; \quad E(\eta_k \eta_j^{\mathrm{T}}) \equiv Q_k \, \delta_{kj}$$

$$E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$$

- Background estimate at time 0

$$x^b{}_0 = x_0 + \zeta^b{}_0$$

$$E(\zeta^b{}_0) = 0 \quad ; \quad E(\zeta^b{}_0 \, \zeta^b{}_0{}^{\mathrm{T}}) \equiv P^b{}_0$$

$$E(\zeta^b{}_0 \varepsilon_k{}^{\mathrm{T}}) = 0 \quad ; \quad E(\zeta^b{}_0 \eta_k{}^{\mathrm{T}}) = 0$$

Sequential assimilation assumes the form of *Kalman Filter*

Background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^T [H_k P^b_k H_k^T + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^T [H_k P^b_k H_k^T + R_k]^{-1} H_k P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k x^a_k$$
$$P^b_{k+1} = M_k P^a_k M_k^T + Q_k$$

In *Ensemble Kalman Filter* (*EnKF*), instead of evolving a covariance matrix, one evolves an ensemble of state vectors, from which the background error covariance matrix is determined.

Only exception so far, in real life assimilation, to linear (and gaussian) approach. Analysis step remains linear (and gaussian).

Bayesian character of *EnKF* ?

- If everything is linear and gaussian, distribution of ensemble produced by EnKF tends to underlying bayesian pdf when ensemble dimension tends to infinity.

- If not, distribution of ensemble tends to a limit which is not the bayesian distribution (F. Le Gland)

Variational assimilation leads to the following *weak constraint* objective function

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

$$\begin{aligned}\mathcal{J}(\xi_0, \xi_1, \ldots, \xi_K) \\ &= (1/2)\,(x^b{}_0 - \xi_0)^T\,[P^b{}_0]^{-1}\,(x^b{}_0 - \xi_0) \\ &+ (1/2)\,\Sigma_{k=0,\ldots,K}\,(y_k - H_k\xi_k)^T\,R_k^{-1}\,(y_k - H_k\xi_k) \\ &+ (1/2)\,\Sigma_{k=0,\ldots,K-1}\,(\xi_{k+1} - M_k\xi_k)^T\,Q_k^{-1}\,(\xi_{k+1} - M_k\xi_k)\end{aligned}$$

In present operational practice (ECMWF, Météo-France, UK Meteorogical Office, Canadian Meteorological Centre , …), model error is ignored. *Strong constraint* variational assimilation

$$\xi_0 \rightarrow \mathcal{J}(\xi_0) = (1/2)\,(x^b_0 - \xi_0)^T\,[P^b_0]^{-1}\,(x^b_0 - \xi_0)$$

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

subject to

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

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, recipe mentioned above

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

Provides bayesian sample of system orbits

- If not, very little can be said at present

How to take temporally correlated errors into account *(model or observation errors)* ***?*** Not possible in standard Kalman Filter, which always updates latest estimate at a given time.

Solution. Augment state vector to temporal dimension over time interval over which correlations are significant.

This is what variational assimilation does. Solutions are being developed for Kalman Filter.

Exact bayesian estimation ?

Particle filters

Predicted ensemble at time t : $\{x^b{}_n, n = 1, \ldots, N\}$, each element with its own weight (probability) $P(x^b{}_n)$

Observation vector at same time : $y = Hx + \varepsilon$

Bayes' formula

$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\, P(x^b{}_n)$$

Defines updating of weights

Bayes' formula

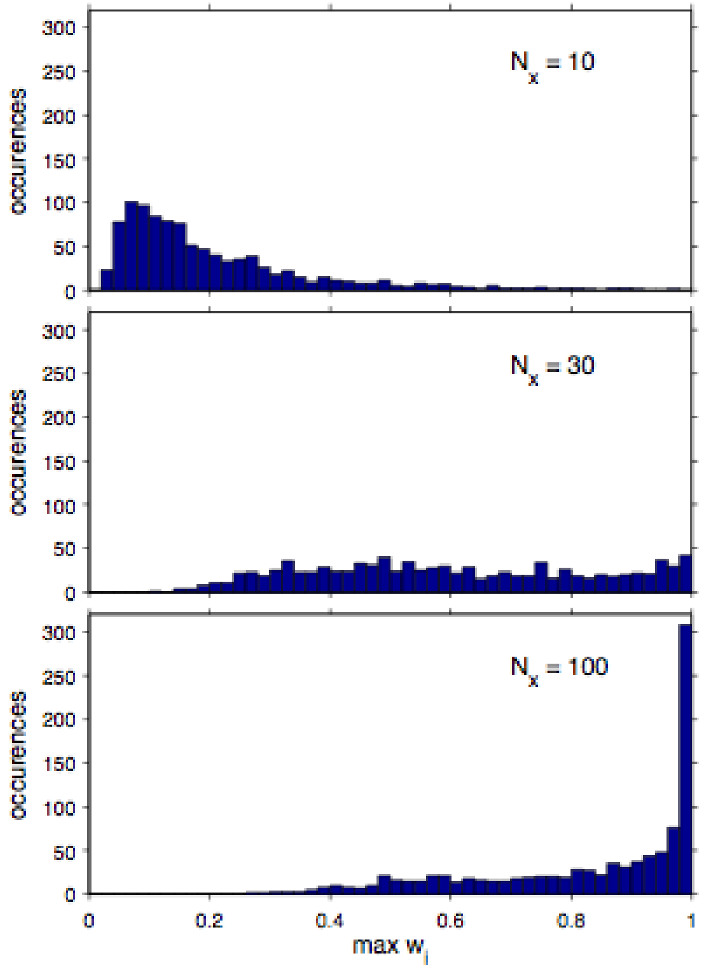
$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\, P(x^b{}_n)$$

Defines updating of weights; particles are not modified. Asymptotically converges to bayesian pdf. Very easy to implement.

Observed fact. For large state dimension, ensemble tends to collapse.

Behavior of $\max w^i$

▷ $N_e = 10^3$; $N_x = 10, 30, 100$; 10^3 realizations



average squared error of posterior mean = 5.5

… = 25

… = 127

C. Snyder,
http://www.cawcr.gov.au/staff/pxs/wmoda5/Oral/Snyder.pdf

Problem originates in the 'curse of dimensionality' Large dimension pdf's are very diffuse, so that very few particles (if any) are present in areas where conditional probability ('*likelihood*') $P(y|x)$ is large.

Bengtsson *et al*. (2008) and Snyder *et al*. (2008) evaluate that stability of filter requires the size of ensembles to increase exponentially with space dimension.

Alternative possibilities (review in van Leeuwen, 2009, *Mon. Wea. Rev.*, 4089-4114)

Resampling. Define new ensemble.

Simplest way. Draw new ensemble according to probability distribution defined by the updated weights. Give same weight to all particles. Particles are not modified, but particles with low weights are likely to be eliminated, while particles with large weights are likely to be drawn repeatedly. For multiple particles, add noise, either from the start, or in the form of 'model noise' in ensuing temporal integration.

Random character of the sampling introduces noise. Alternatives exist, such as *residual sampling* (Lui and Chen, 1998, van Leeuwen, 2003). Updated weights w_n are multiplied by ensemble dimension N. Then p copies of each particle n are taken, where p is the integer part of Nw_n. Remaining particles, if needed, are taken randomly from the resulting distribution.

Importance Sampling.

Use a *proposal density* that is closer to the new observations than the density defined by the predicted particles (for instance the density defined by EnKF, after the latter has used the new observations). Independence between observations is then lost in the computation of likelihood $P(y|x)$.

In particular, *Guided Sequential Importance Sampling* (van Leeuwen, 2002). Idea : use observations performed at time k to resample ensemble at some timestep anterior to k.

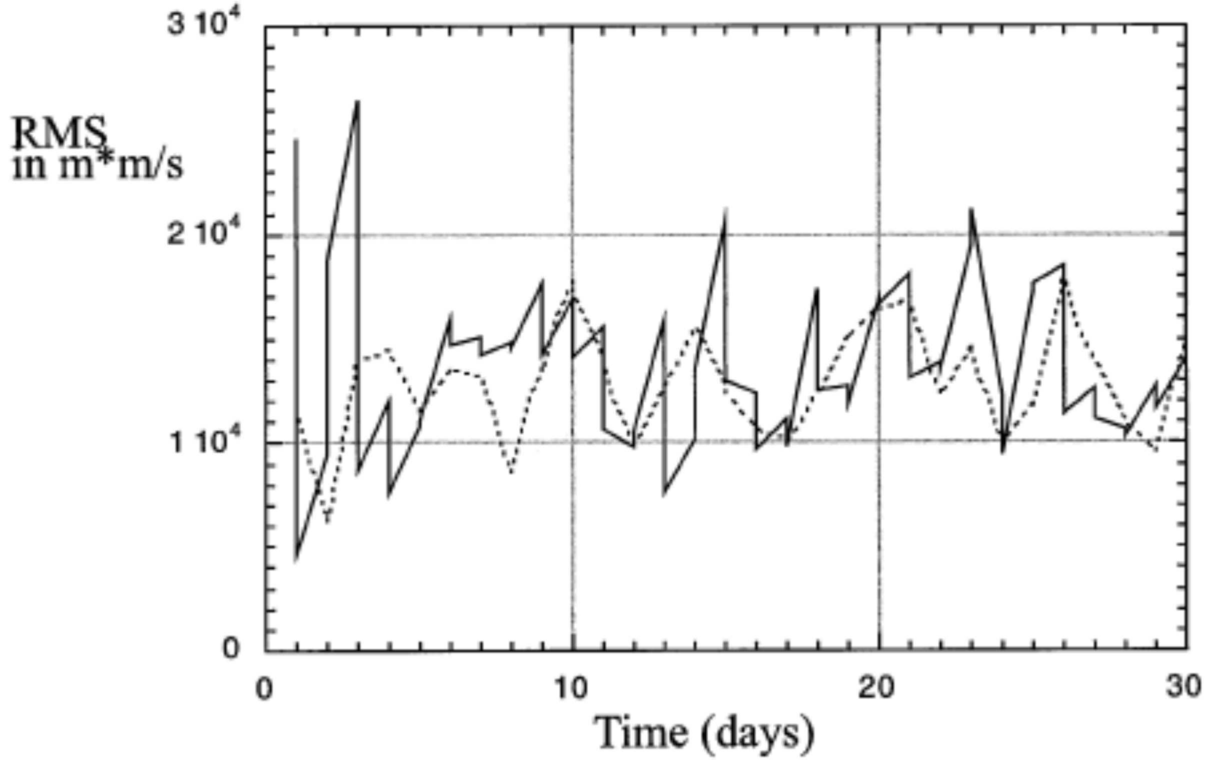


FIG. 12. Comparison of rms error ($m^2 s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

van Leeuwen, 2003, *Mon. Wea. Rev.*, **131**, 2071-2084

Exact bayesian estimation (continuation)

Acceptation-rejection

Bayes' formula

$$f(x) \equiv \mathrm{P}(x \mid y) = P(y \mid x)\, P(x) \,/\, P(y)$$

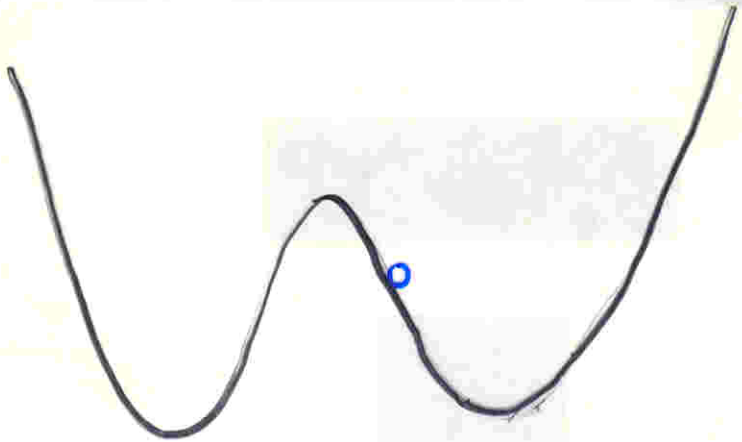
defines probability density function for x.

Construct sample of that pdf as follows.

Draw randomly couple $(\xi, \psi) \in \mathcal{S} \times [0, \mathrm{sup} f]$.

Keep ξ if $\psi < f(\xi)$. ξ is then distributed according to $f(x)$.

Miller, Carter and Blue, Tellus, 1999



$$\frac{d^2x}{dt^2} = -\frac{d\phi}{dx} - \alpha\frac{dx}{dt} + \text{Noise}$$

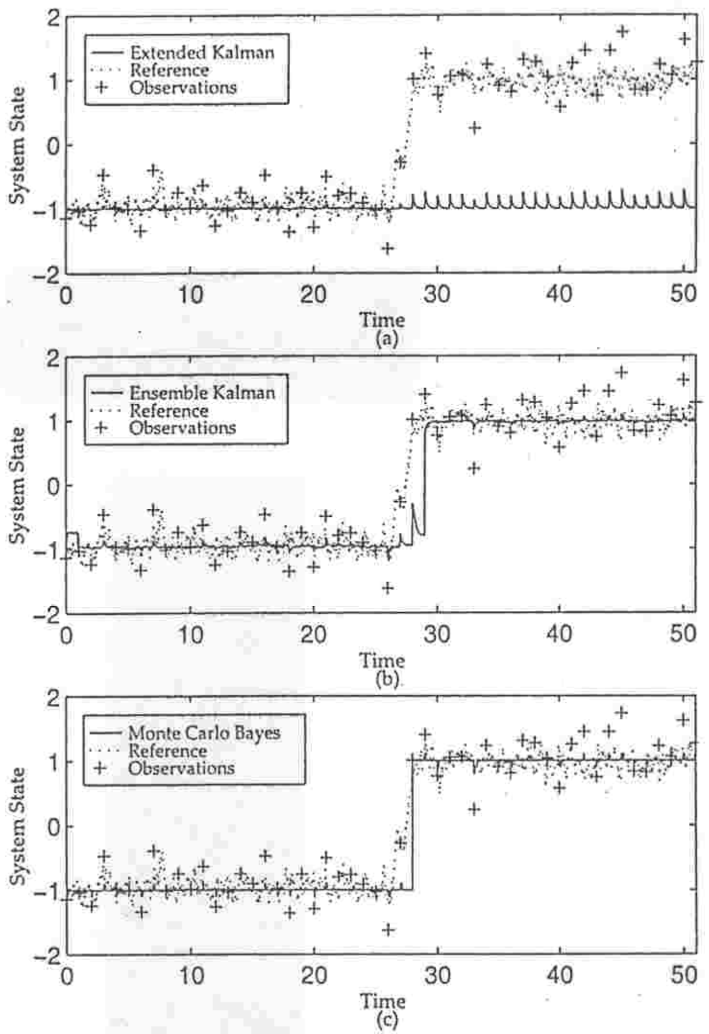


Fig. 4. Comparison of the EKF, the ensemble method and nonlinear filtering by Bayes' theorem for the double-well problem.

Miller, Carter and Blue, 1999, *Tellus*, **51A**, 167-194

Acceptation-rejection

Seems costly.

Requires capability of permanently interpolating probability distribution defined by finite sample to whole state space.

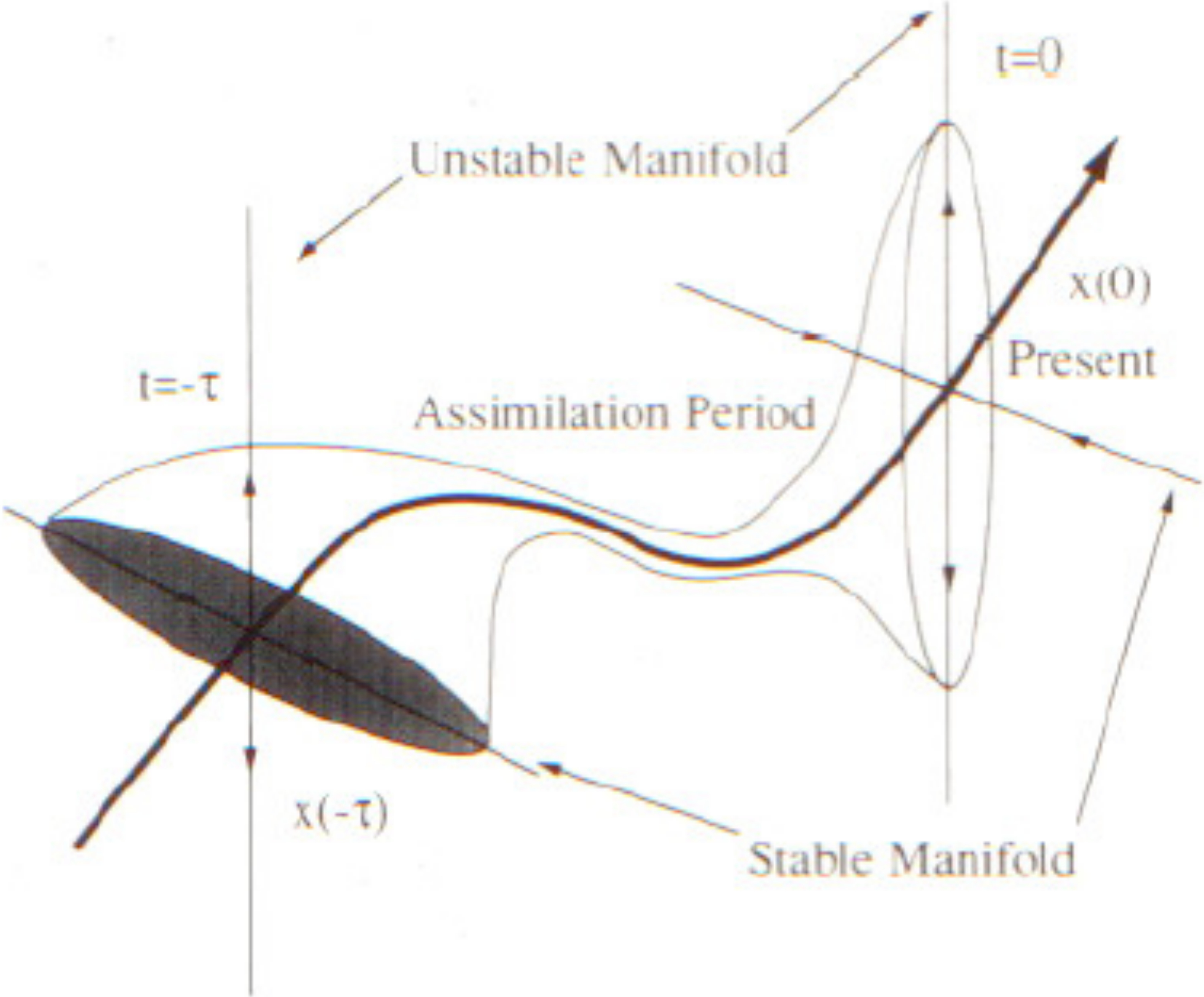
Those alternative possibilities are beneficial in that they reduce the occurrence of ensemble collapse. General conclusion from experiments performed in geophysical applications is that particle filters can produce better results than, say, Ensemble Kalman Filter. But that is obtained at the price of using ensemble dimensions that are prohibitive (100-200).

Particle filters are now a 'hot' research topic, studied in many places (C. Snyder, P. J. van Leeuwen, S. Nakano, C. Baehr, …)

A particular question, as always with sequential estimation, is the possibility of taking temporal error dependence into account. This can in principle be done by augmenting the state vector to the time dimension.

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 (Pires *et al*., *Tellus*, 1996).

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 : 4D-Var assimilation, 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 asimilation), in which estimation error is concentrated in unstable modes at any time.

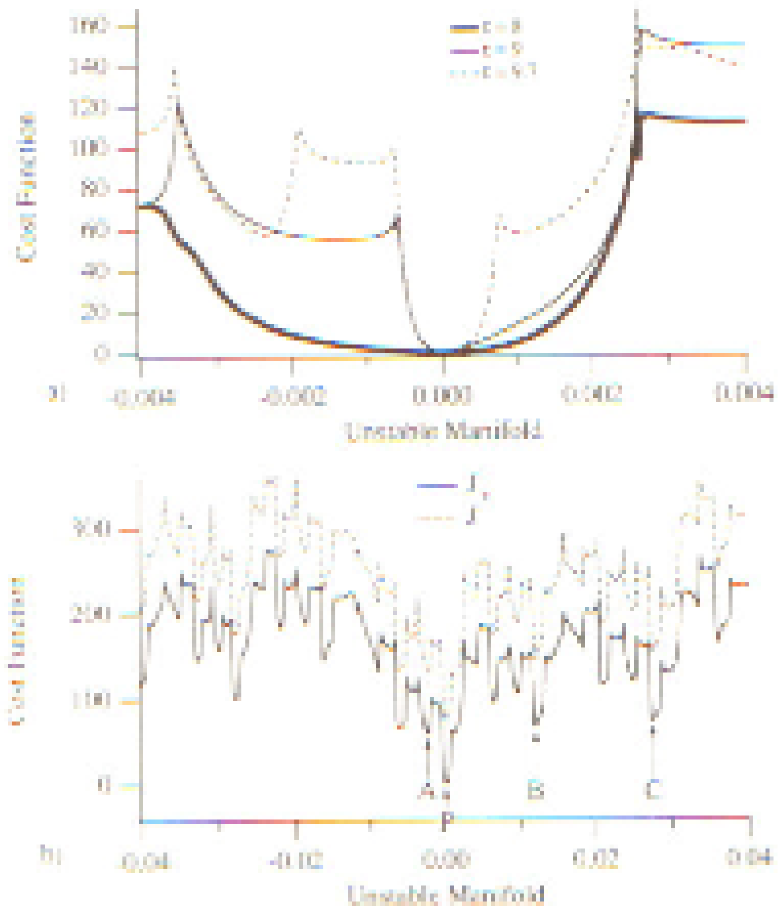


Fig. 4. Panel (a): Cross-section of the error-free forward cost-function $J'_c(\tau, \hat{x}, x)$ along the unstable manifold, for various values of τ. Panel (b). As in panel (a), for $\tau = 9.7$, and with a display interval ten times as large, respectively for the error-free forward cost-function $J'_c(\tau, \hat{x}, x)$ (solid curve) and for the error-contaminated cost-function $J_c(\tau, \hat{x}, x)$ (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)

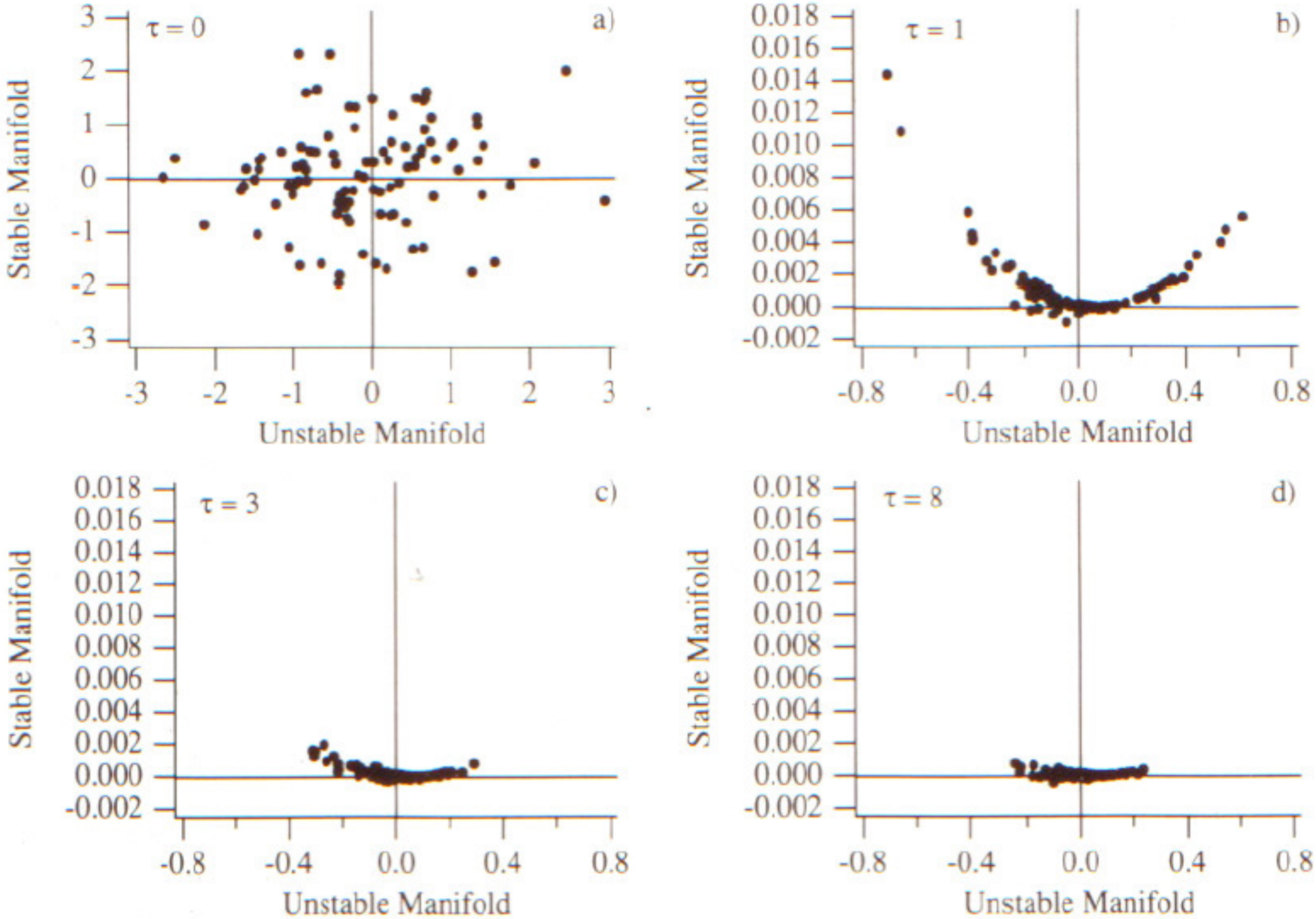


Fig. 7. Projection of the 100 minimizing solutions, at the end of the assimilation period, onto the plane spanned by the stable and unstable directions, defined as in Fig. 3. Values of τ are indicated on the panels. The projection is not an orthogonal projection, but a projection parallel to the local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

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

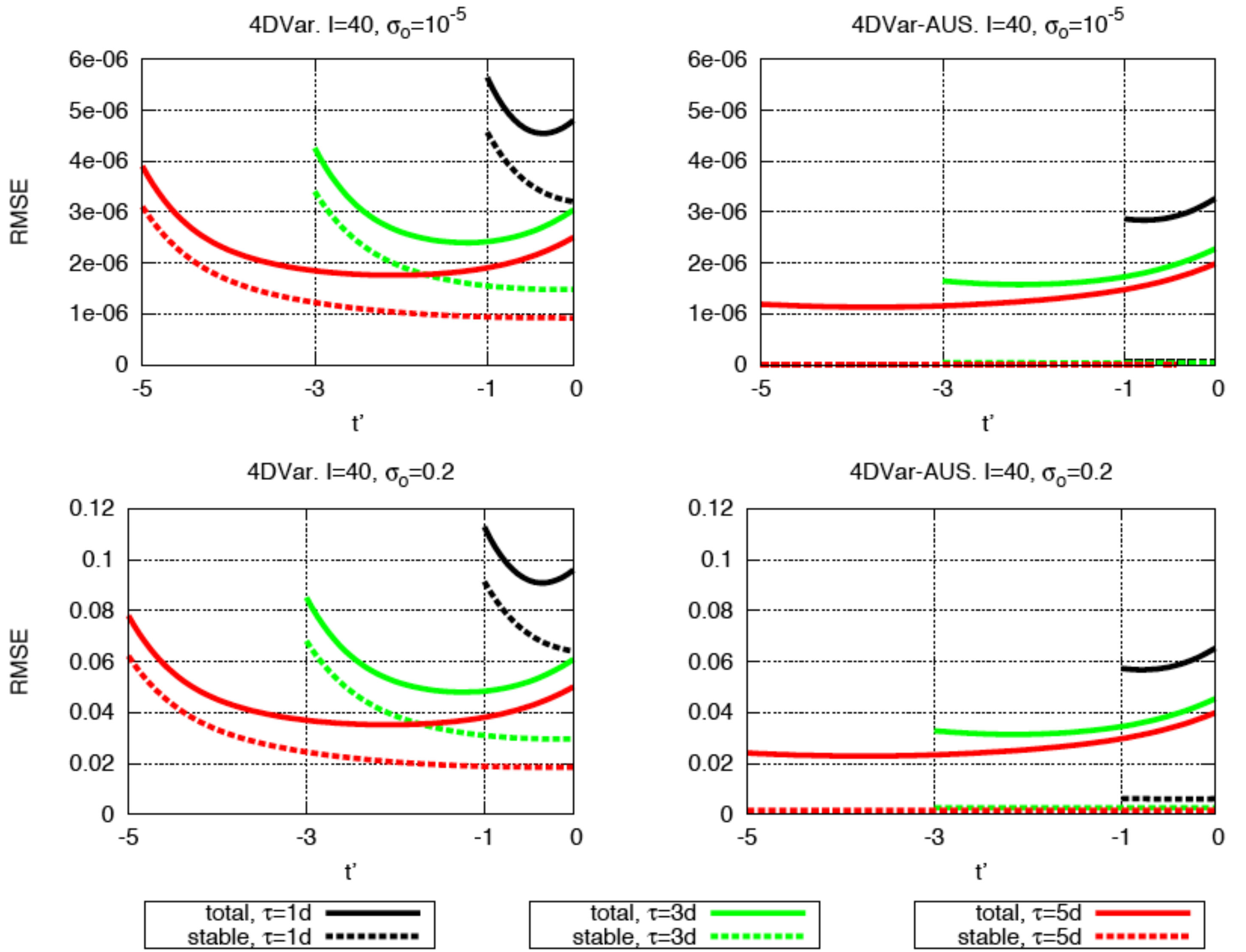


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to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

Trevisan *et al.*, 2010, in press

Consequence : it might be useful, at least in terms of cost efficiency, to concentrate assimilation in modes that have been unstable in the recent past, where uncertainty is likely to be largest.

Also, presence of residual noise in stable modes can be damageable for analysis and subsequent forecast.

Assimilation in the Unstable Subspace (AUS) (Carrassi *et al*., 2007, 2008, for the case of 3D-Var)

Four-dimensional variational assimilation in the unstable subspace (*4DVar-AUS*)

Trevisan *et al*., 2010, Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension, *Q. J. R. Meteorol. Soc.*, in press

4D-Var-AUS

Algorithmic implementation

Define *N* perturbations to the current state, and evolve them according to the tangent linear model, with periodic reorthonormalization in order to avoid collapse onto the dominant Lyapunov vector (same algorithm as for computation of Lyapunov exponents).

Cycle successive 4D-Var's, restricting at each cycle the modification to be made on the current state to the space spanned by the *N* perturbations emanating from the previous cycle (if *N* is the dimension of state space, that is identical with standard 4D-Var).

Experiments performed on the Lorenz (1996) model

$$\frac{d}{dt}x_j = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F$$

with $j = 1, ..., I.$

with value F = 8, which gives rise to chaos.

Three values of I have been used, namely I = 40, 60, 80, which correspond to respectively N^+ = 13, 19 and 26 positive Lyapunov exponents.

In all three cases, the largest Lyapunov exponent corresponds to a doubling time of about 2 days (with 1 ‘day’ = 1/5 model time unit).

Identical twin experiments (perfect model)

‘Observing system’ defined as in Fertig *et al*. (*Tellus*, 2007):

At each observation time, one observation every four grid points (observation points shifted by one grid point at each observation time).

Observation frequency : 1.5 hour

Random gaussian observation errors with expectation 0 and standard deviation $\sigma_0 = 0.2$ (‘climatological’ standard deviation 5.1).

Sequences of variational assimilations have been cycled over windows with length $\tau = 1, \ldots, 5$ days. Results are averaged over 5000 successive windows.

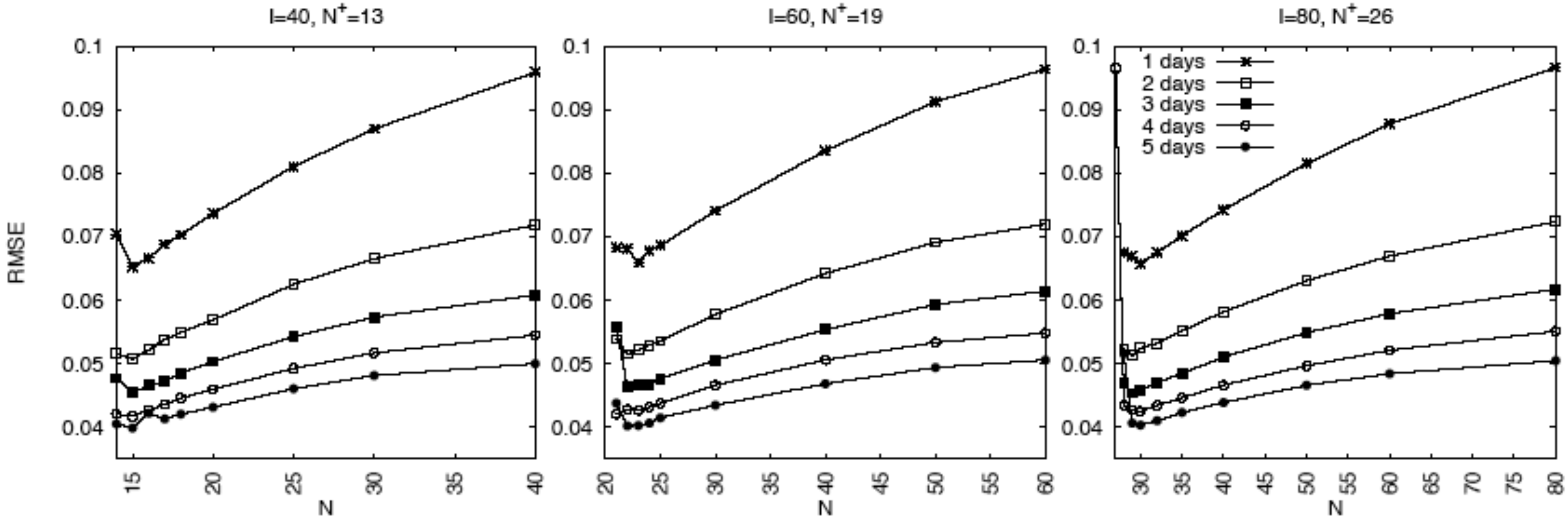


Figure 1. Time average RMS analysis error at $t = \tau$ as a function of the subspace dimension N for three model configurations: I=40, 60, 80. Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is $\sigma_o = 0.2$.

No explicit background term (*i. e.*, with error covariance matrix) in objective function : information from past lies in the background to be updated, and in the N perturbations which define the subspace in which updating is to be made.

Best performance for N slightly above number N^+ of positive Lyapunov exponents.

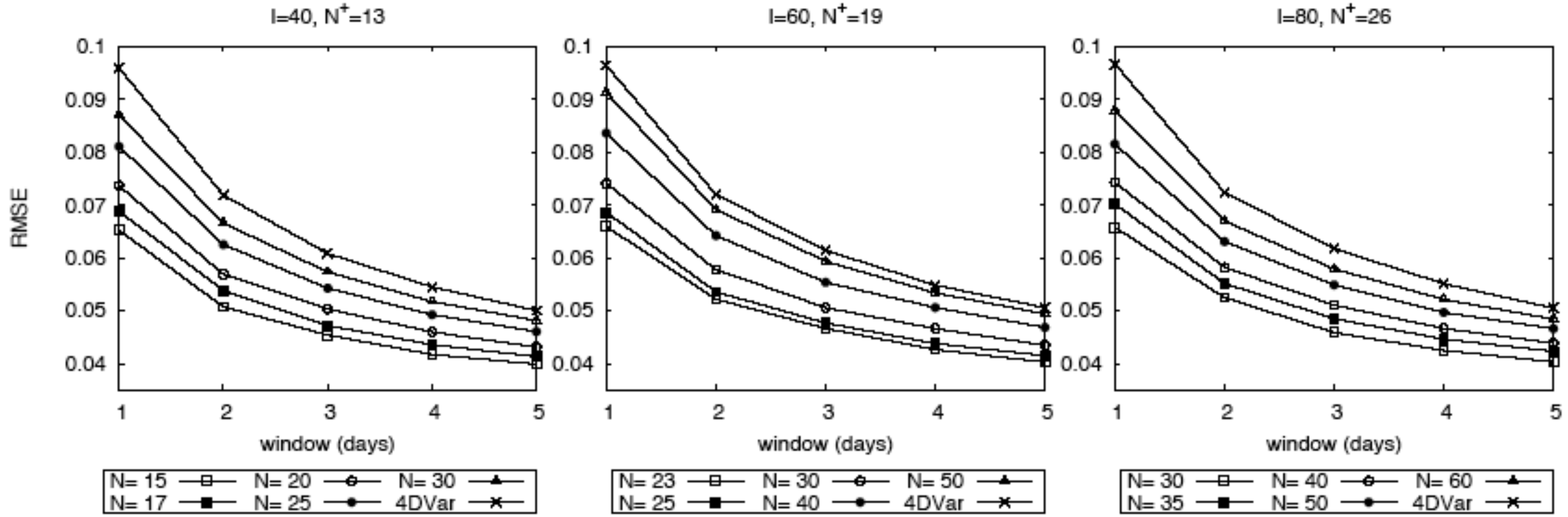


Figure 2. Time average RMS analysis error at $t = \tau$ as a function of the length of the assimilation window for three model configurations: I=40, 60, 80. Different curves in the same panel refer to a different subspace dimension N of 4DVar-AUS and to standard 4DVar. $\sigma_o = 0.2$.

Different curves are almost identical on all three panels. Relative improvement obtained by decreasing subspace dimension N to its optimal value is largest for smaller window length τ.

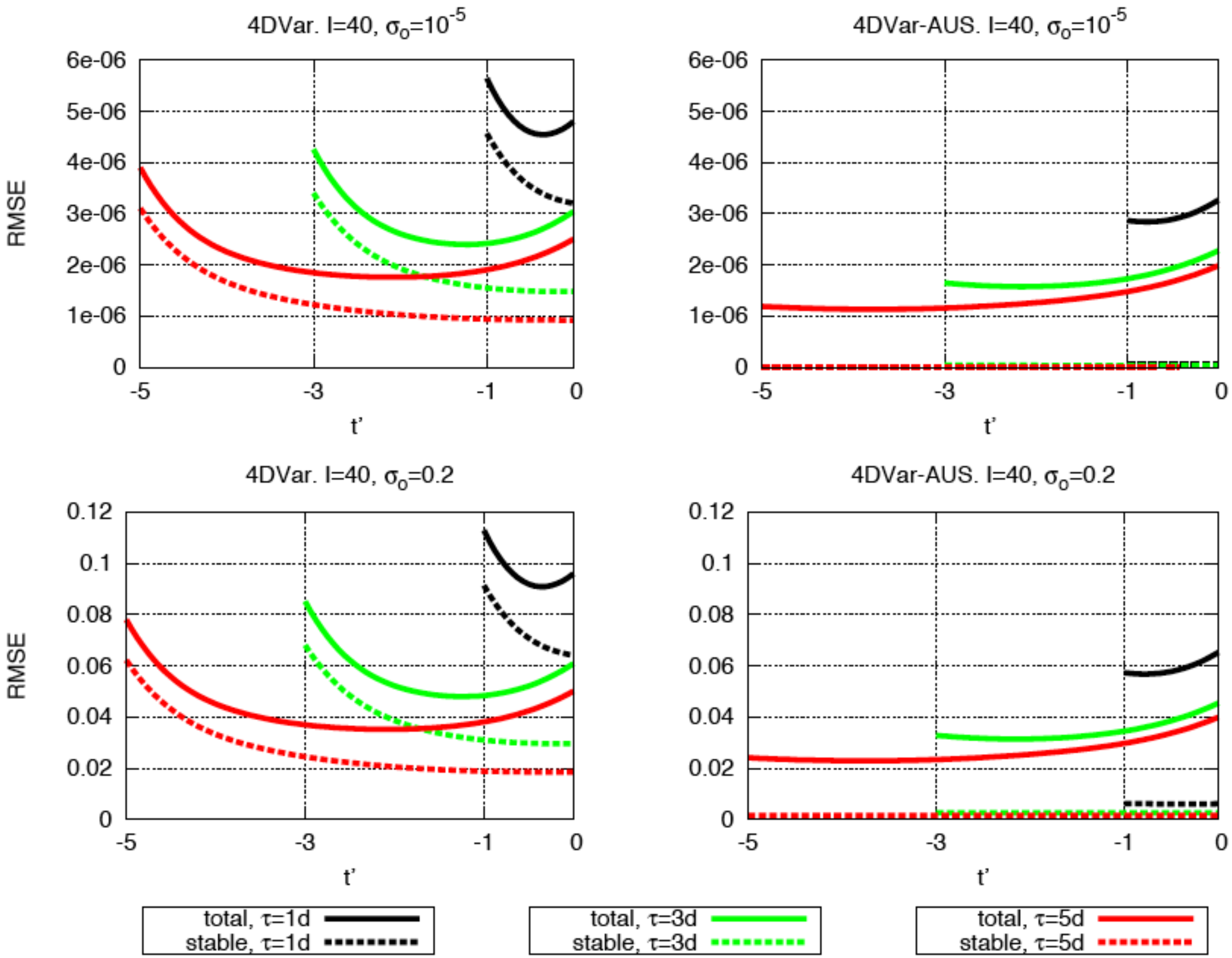


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to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

Experiments have been performed in which an explicit background term was present, the associated error covariance matrix having been obtained as the average of a sequence of full 4D-Var's.

The estimates are systematically improved, and more for full 4D-Var than for 4D-Var-AUS. But they remain qualitatively similar, with best performance for 4D-Var-AUS with N slightly above N^+.

Minimum of objective function cannot be made smaller by reducing control space. Numerical tests show that minimum of objective function is smaller (by a few percent) for full 4D-Var than for 4D-Var-AUS. Full 4D-Var is closer to the noisy observations, but farther away from the truth. And tests also show that full 4D-Var performs best when observations are perfect (no noise).

Results show that, if all degrees of freedom that are available to the model are used, the minimization process introduces components along the stable modes of the system, in which no error is present, in order to ensure a closer fit to the observations. This degrades the closeness of the fit to reality. The optimal choice is to restrict the assimilation to the unstable modes.

Can have major practical algorithmic implications.

Questions.

- Degree of generality of results ?

- Impact of model errors ?

Thanks !