École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2011-2012

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation d'Observations

Olivier Talagrand
Cours 8
25 Mai 2012

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

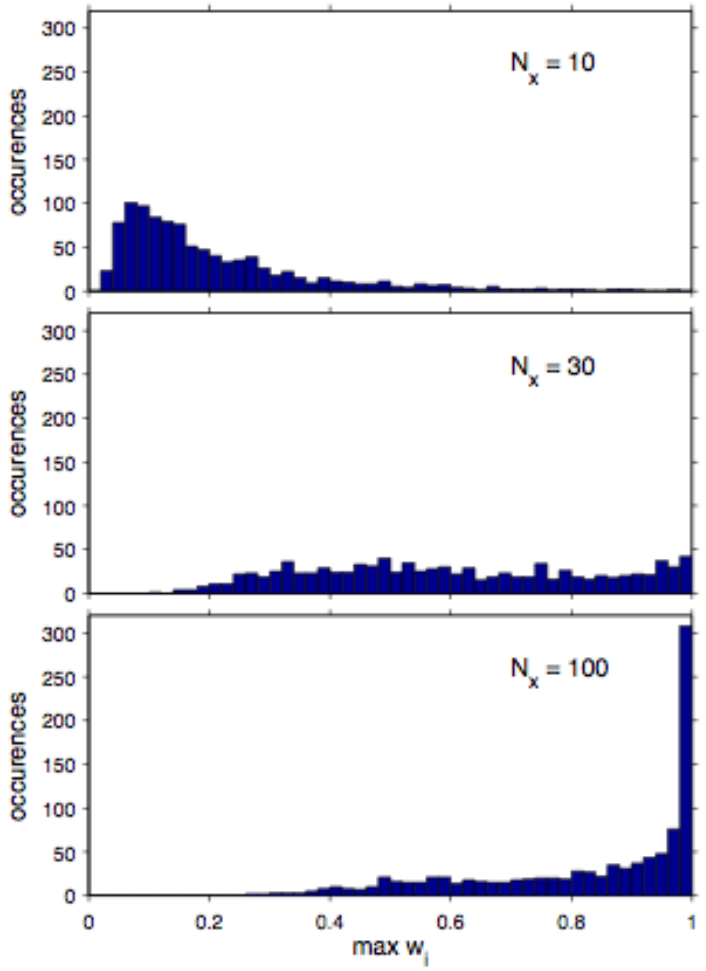
$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\, P(x^b{}_n)$$

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 $= 5.5$

$\ldots = 25$

$\ldots = 127$

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 $P(y|x)$ (or is it not ?)

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, or 'nudge' integration between times k-1 and k towards observation at time k.

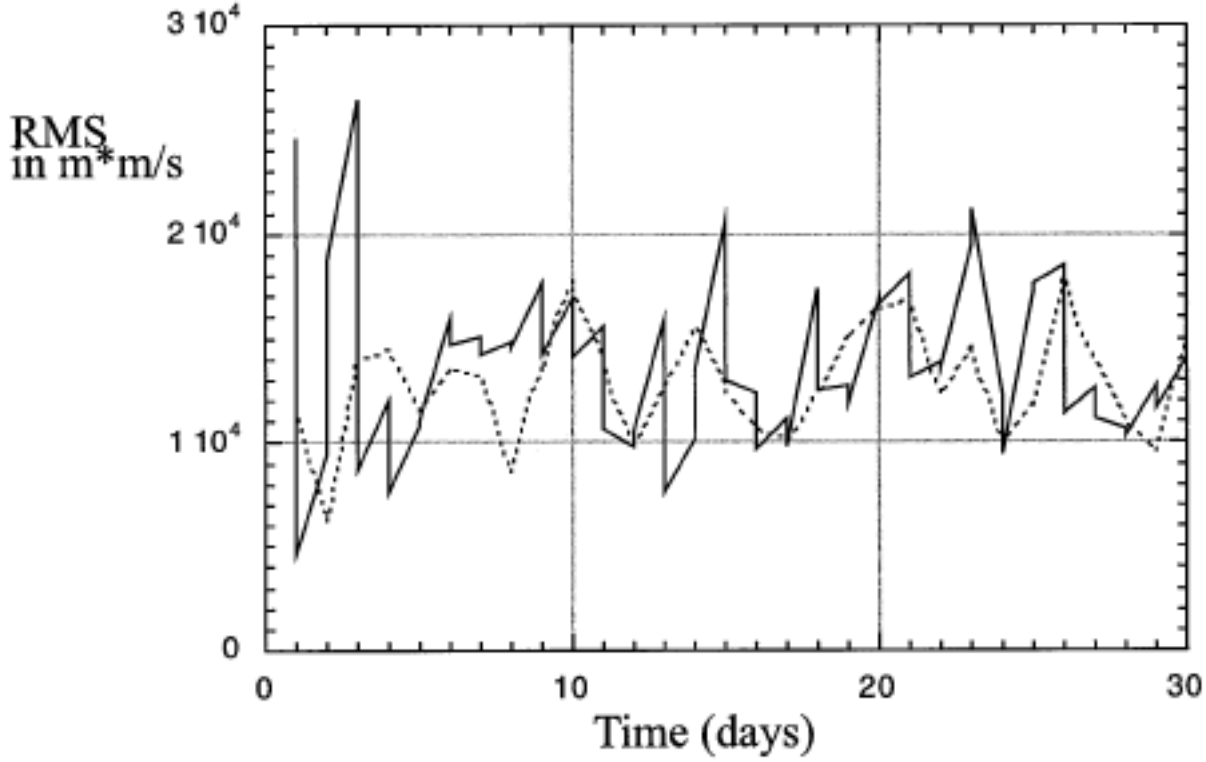
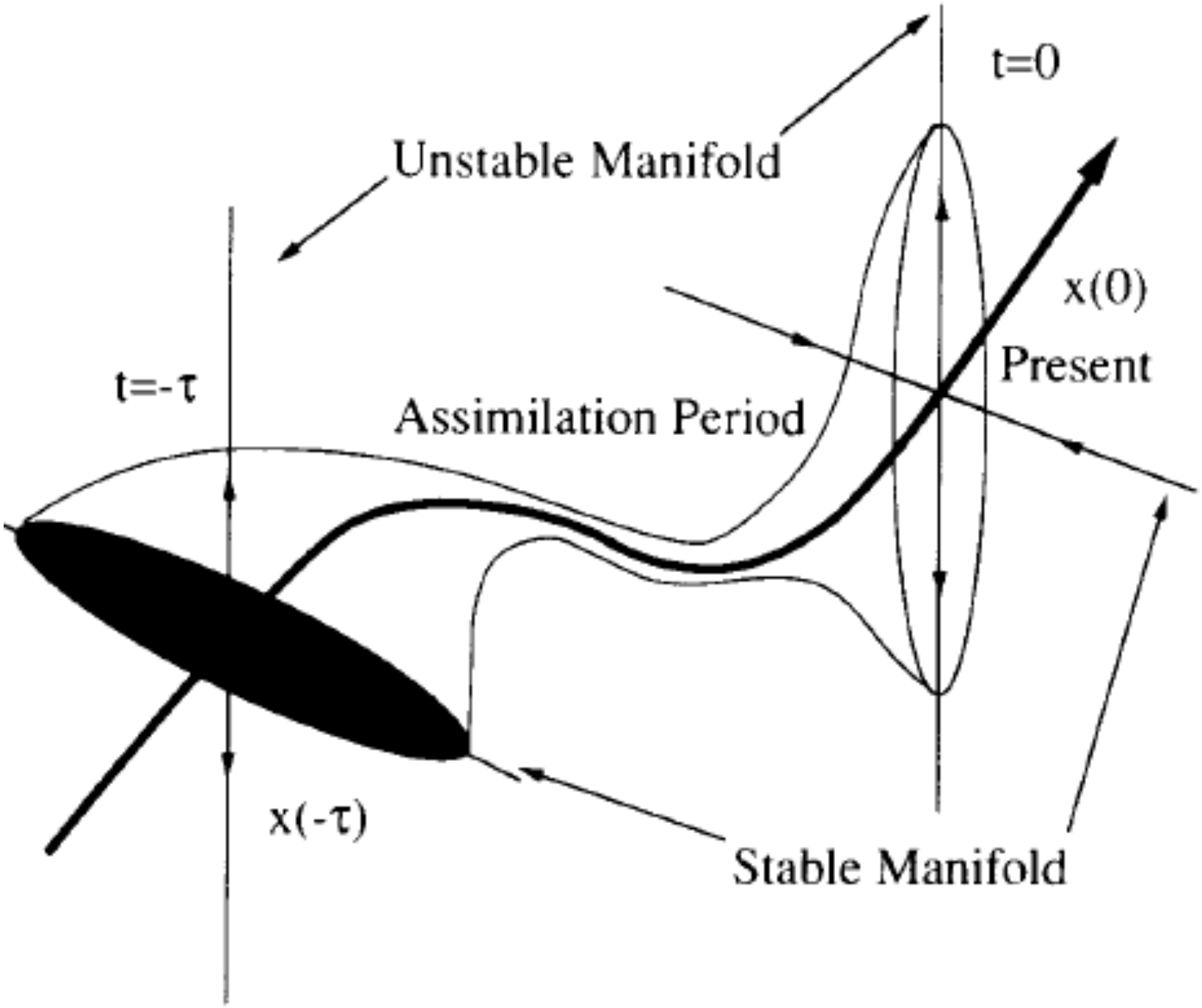


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

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

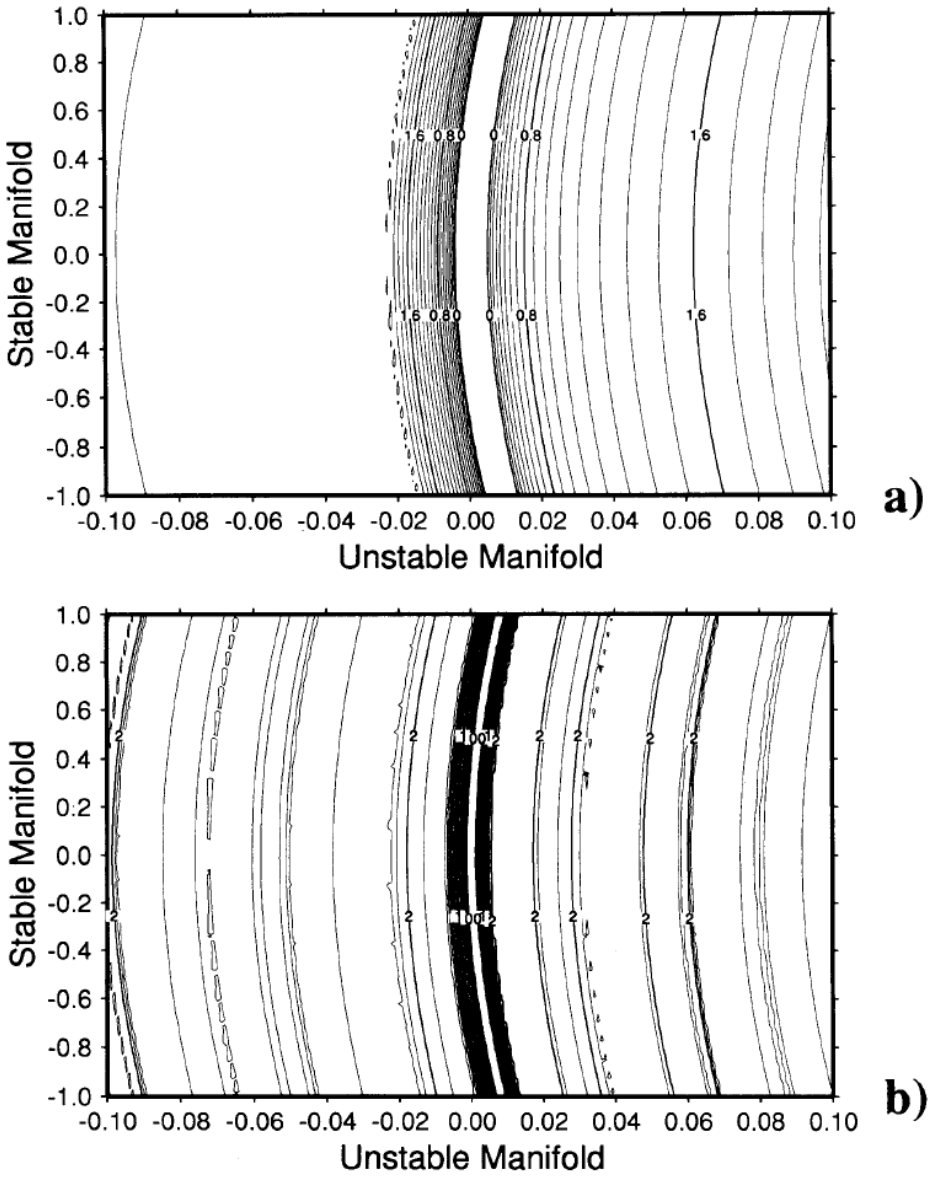


Fig. 3. Variations of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (Lorenz system) in the plane spanned by the stable and unstable directions, as determined from the tangent linear system (see text), and for $\tau = 6$ (panel (a)) and $\tau = 8$ (panel (b)) respectively. The metric has 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.

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.

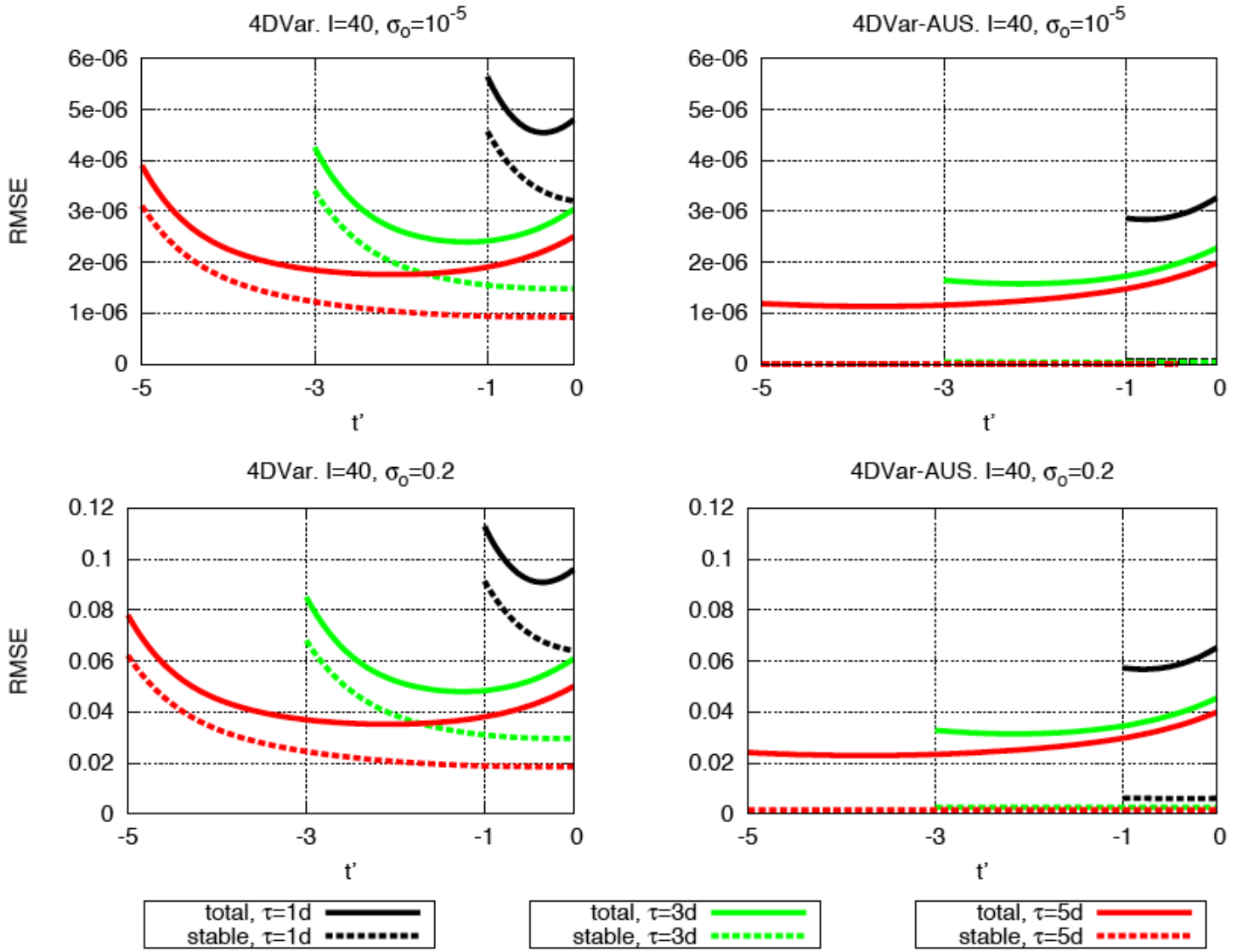


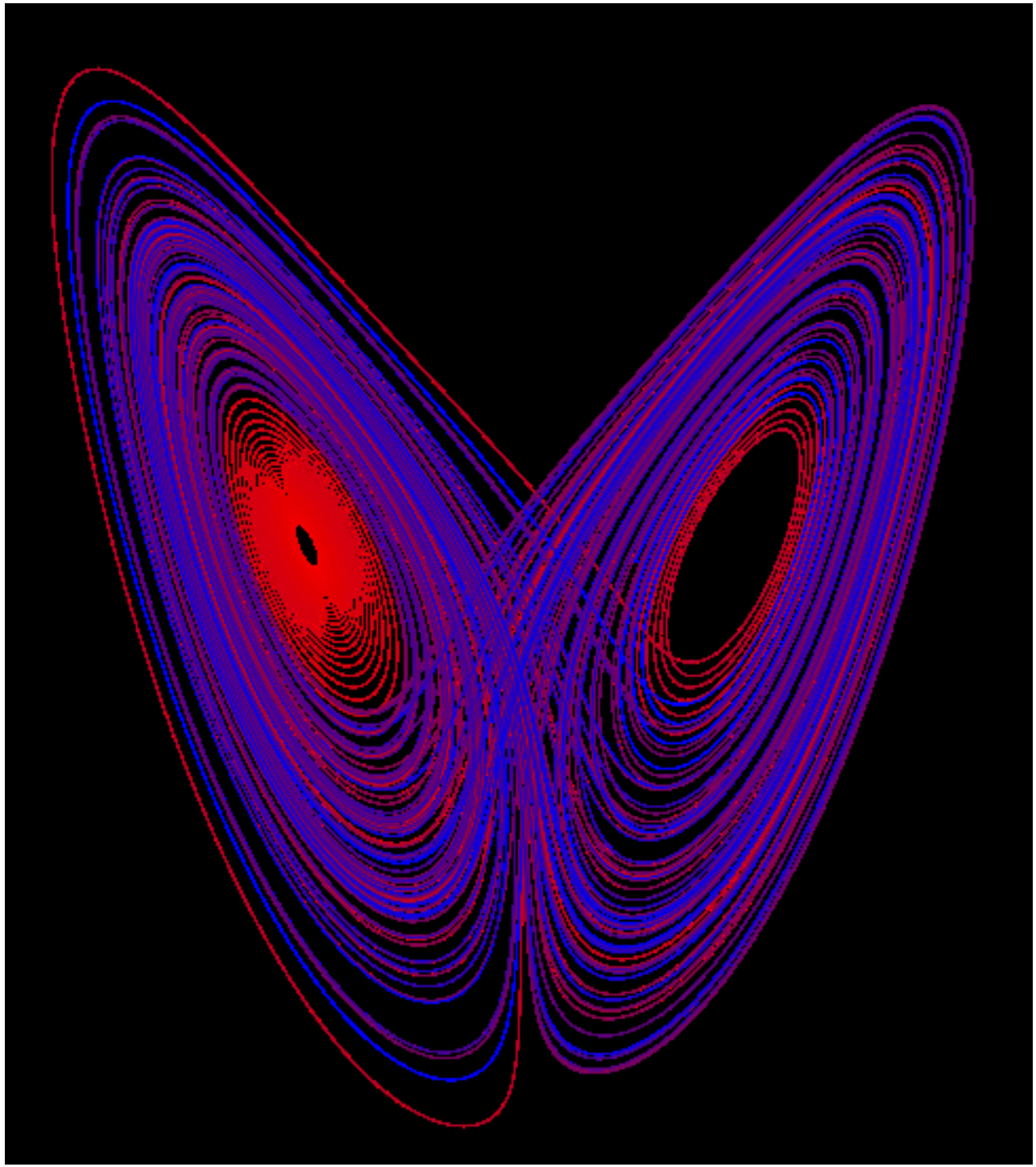
Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to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

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

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



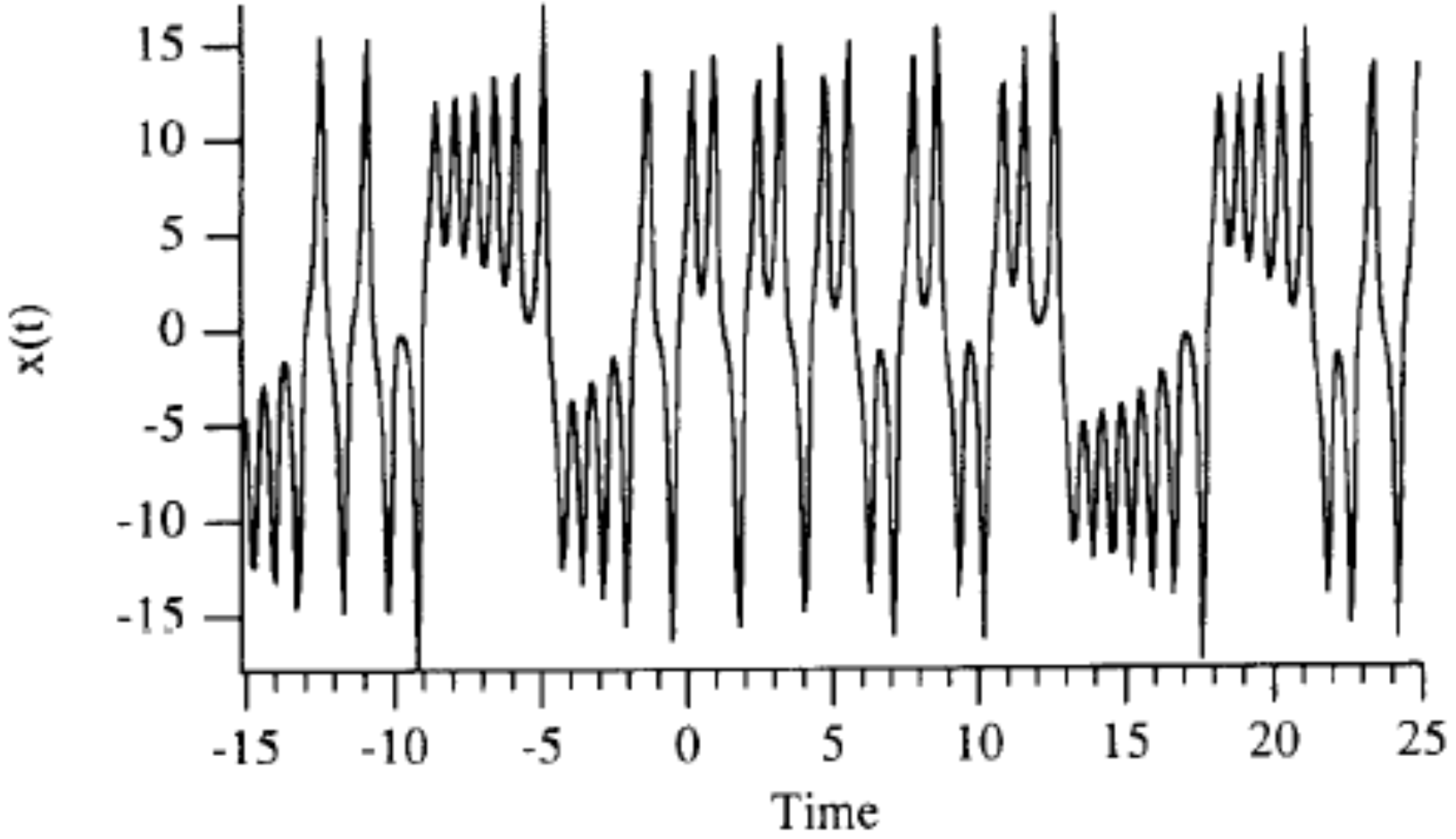


Fig. 2. Time variations, along the reference solution, of the variable $x(t)$ of the Lorenz system.

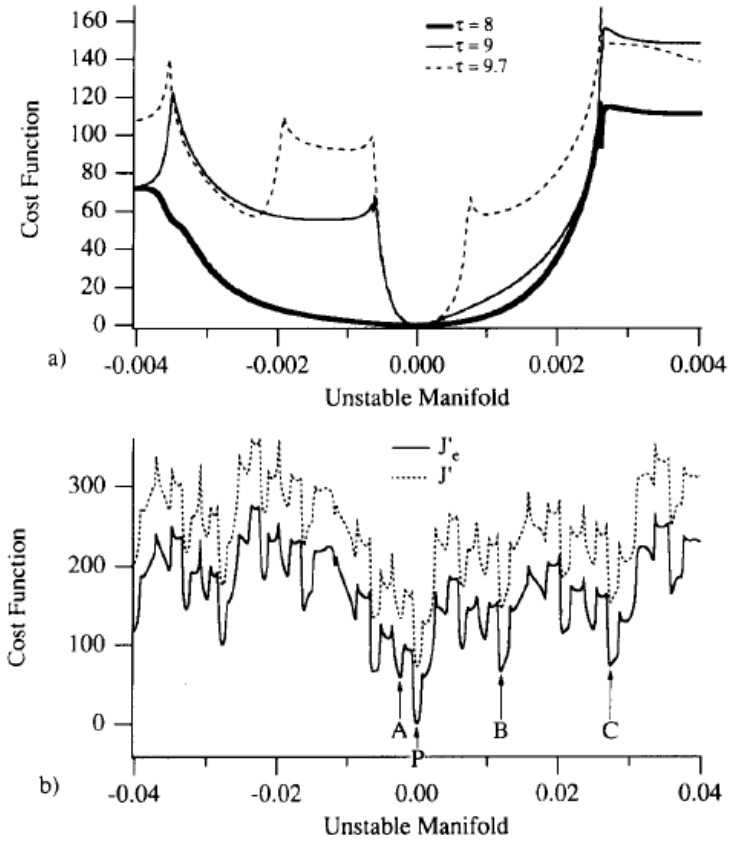


Fig. 4. Panel (a): Cross-section of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ along the unstable manifold, for various values of τ. Panel (b). As in panel (a), for $\tau = 9.7$, and with a display interval ten times as large, respectively for the error-free forward cost-function $J'_e(\tau,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

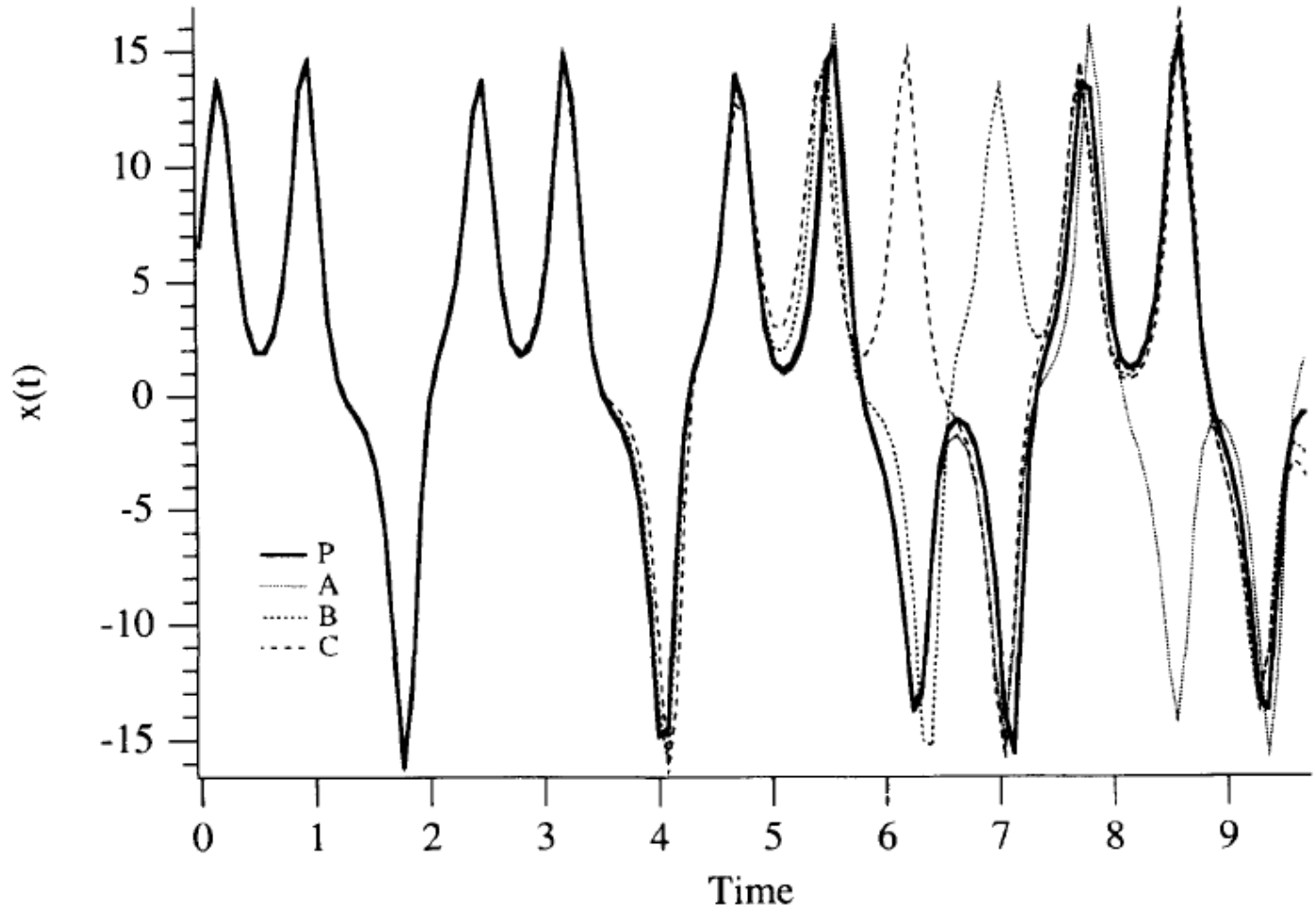


Fig. 5. Variations of the coordinate x along the orbits originating from the minima *P*, *A*, *B*, *C* (indicated in Fig. 4b) of the error-free cost-function.

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*.)

.

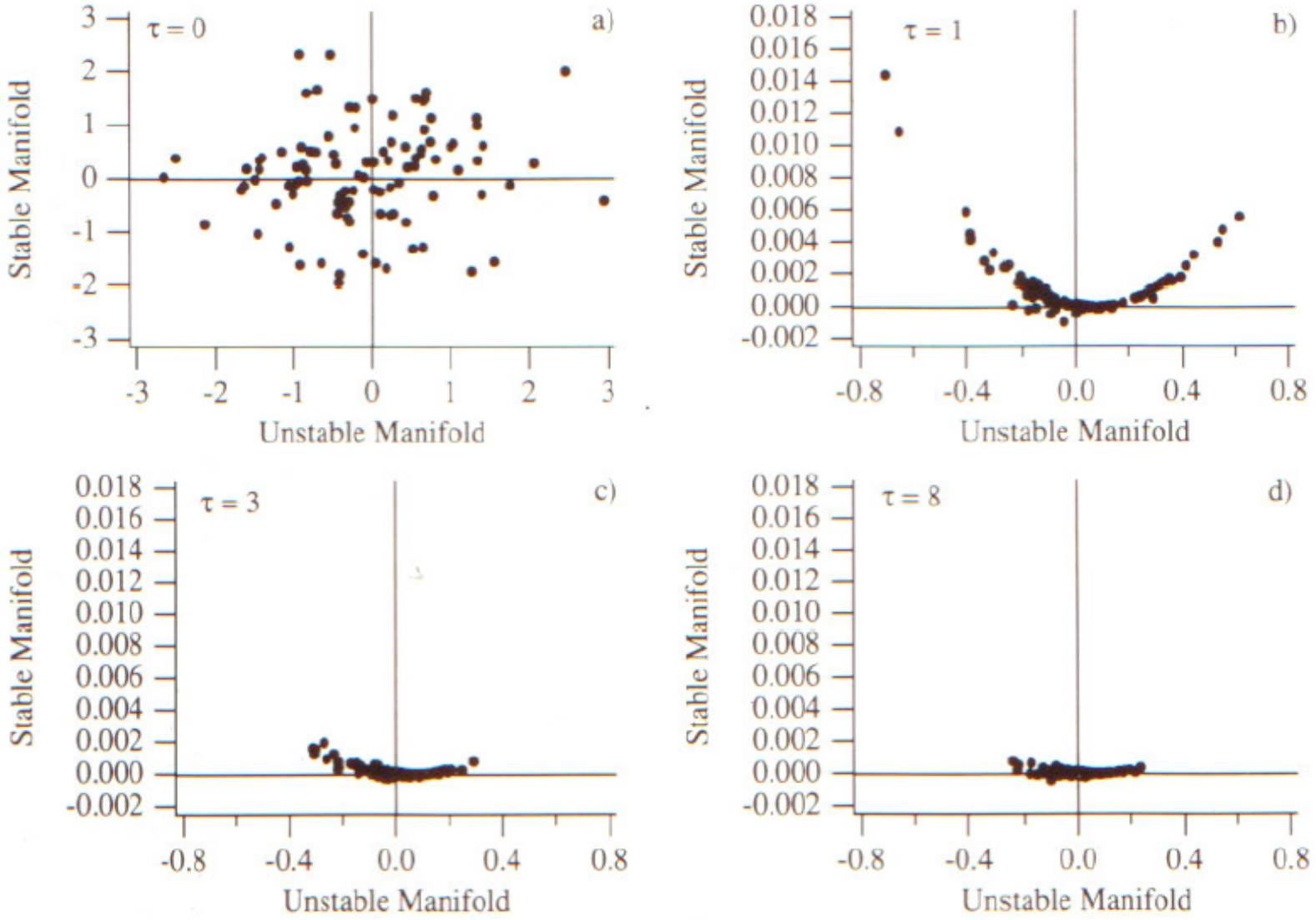


Fig. 7. Projection of the 100 minimizing solutions, at the end of the assimilation period, onto the plane spanned by the stable and unstable directions, defined as in Fig. 3. Values of τ are indicated on the panels. The projection is not an orthogonal projection, but a projection parallel to the local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

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

$\mu(C(\tau, x))$	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).

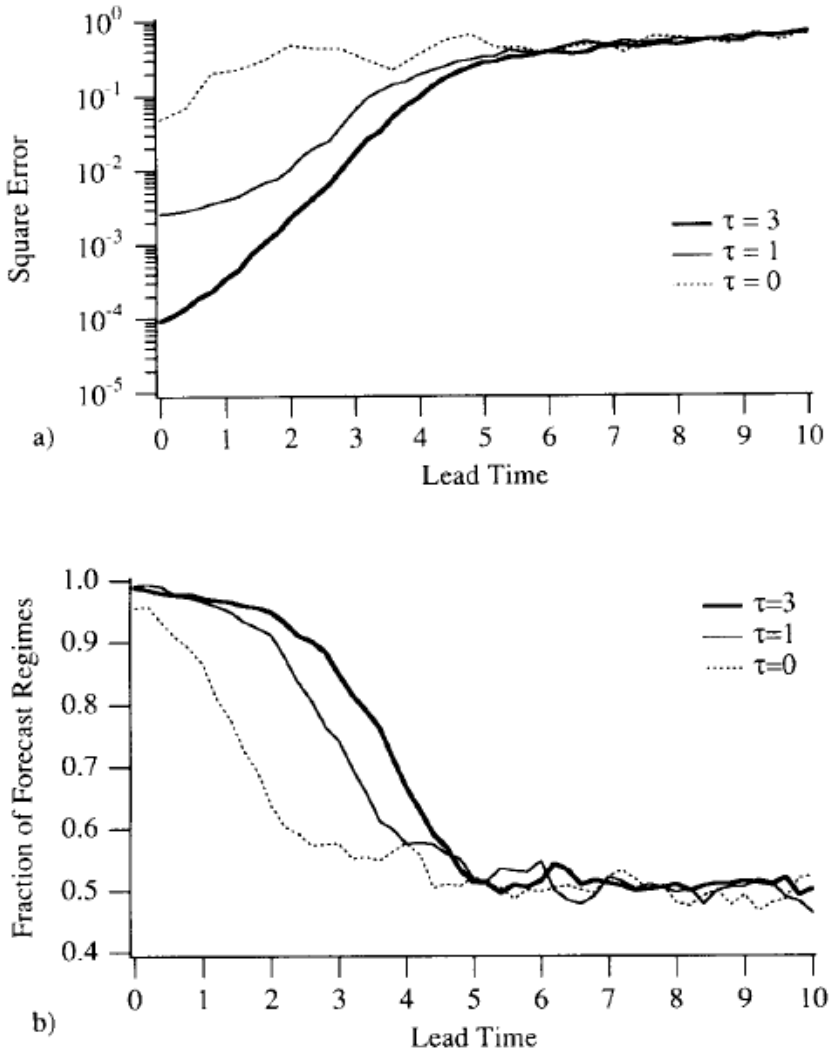


Fig. 11. (a) Squared normalized forecast error, averaged over the attractor, as a function of forecast time, and for the values $\tau = 0$, 1 and 3 of the length of the assimilation period (b). Fraction, over the attractor, of successful regime forecasts, as a function of forecast time, and for the values $\tau = 0$, 1 and 3 of the length of the assimilation period.

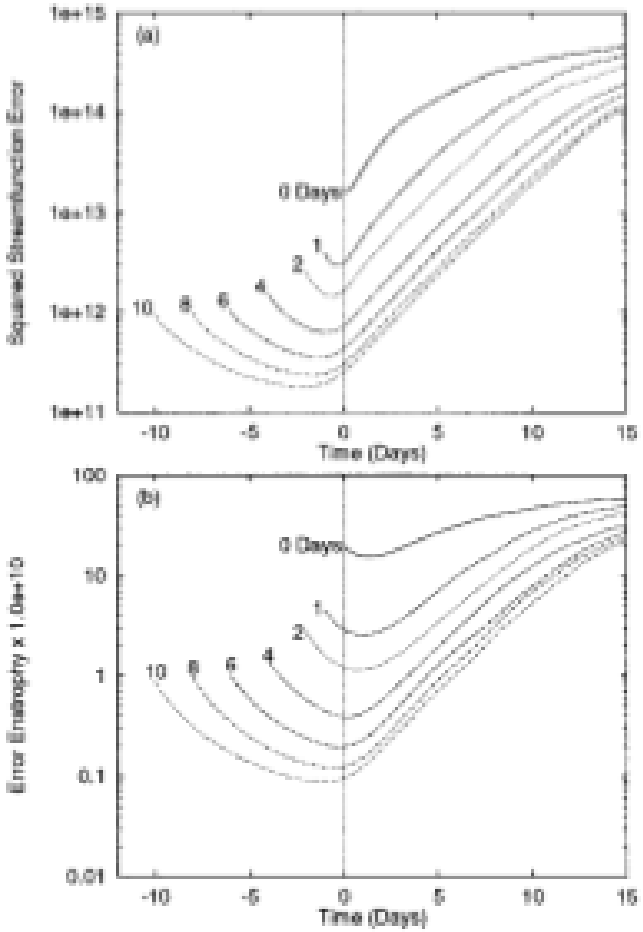


Fig. 5. Median values of the (a) streamfunction squared error, and (b) enstrophy error for the 200 forecast set as a function of forecast time and of the assimilation time T_a.

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

Since, after an assimilation has been performed over a period of time, uncertainty is likely to be concentrated in modes that have been unstable, it might be useful for the next assimilation, and at least in terms of cost efficiency, to concentrate corrections on the background in those modes.

Actually, 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*., **136**, 487-496.

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

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

$$\text{with } j = 1, ..., I.$$

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.

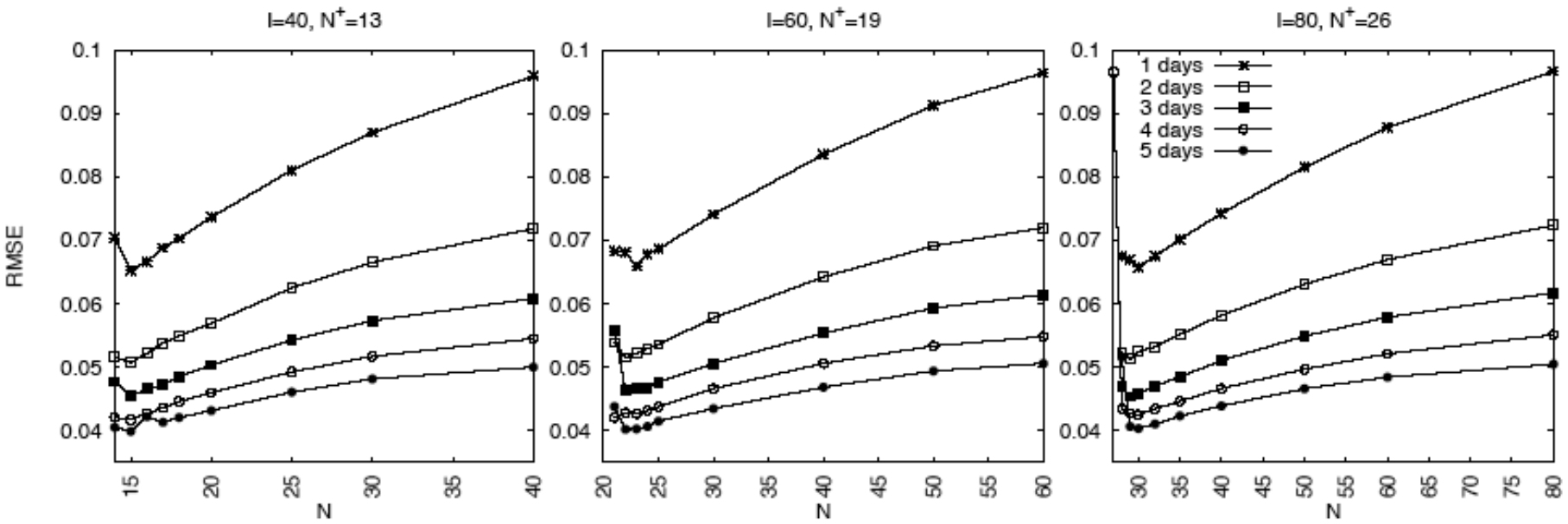


Figure 1. Time average RMS analysis error at $t = \tau$ as a function of the subspace dimension N for three model configurations: I=40, 60, 80. Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is $\sigma_o = 0.2$.

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.

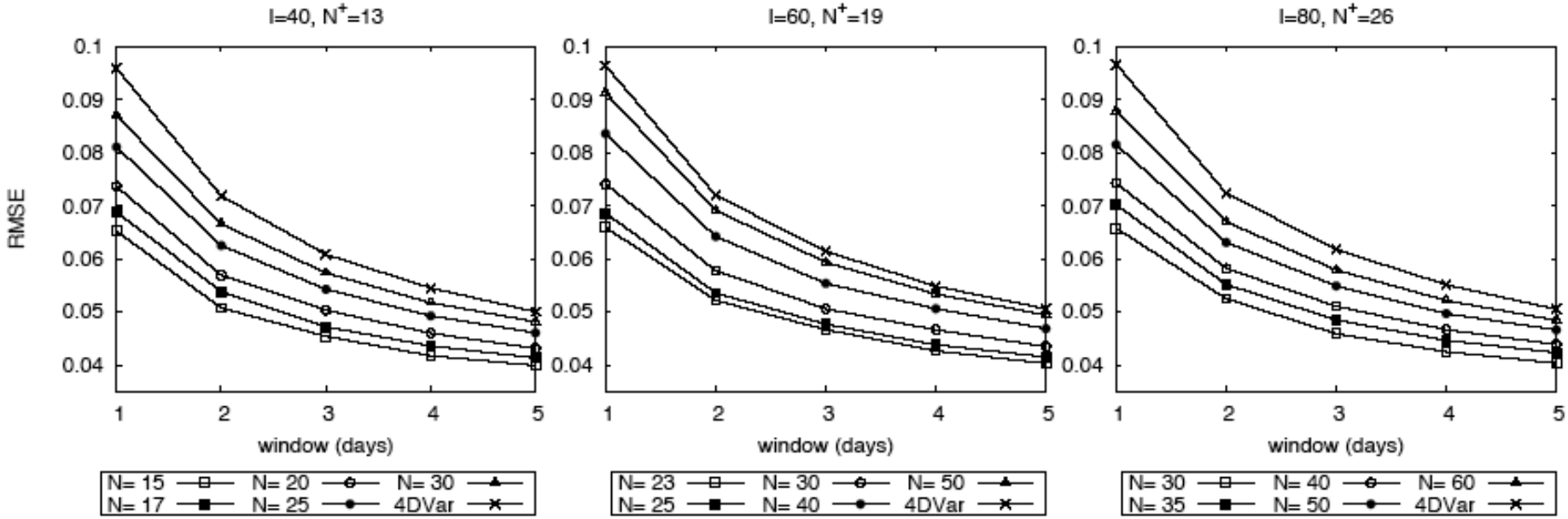


Figure 2. Time average RMS analysis error at $t = \tau$ as a function of the length of the assimilation window for three model configurations: I=40, 60, 80. Different curves in the same panel refer to a different subspace dimension N of 4DVar-AUS and to standard 4DVar. $\sigma_o = 0.2$.

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.

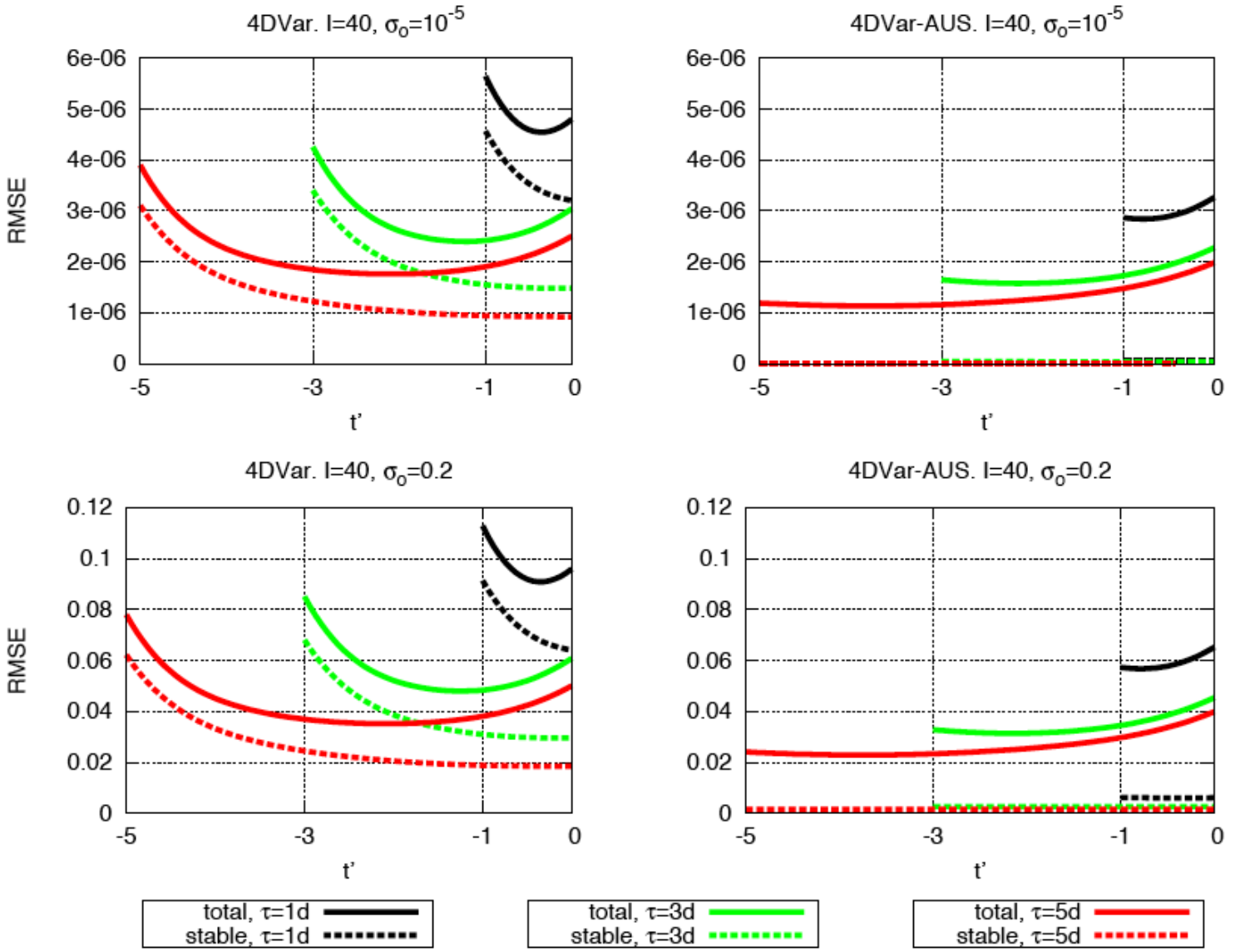


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to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

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 ?

Conclusions

- The main two classes of algorithms that presently exist for operational assimilation of observations in geophysical applications are Ensemble Kalman Filter (EnKF) and Variational Assimilation (4D-Var). They both are more or less empirical extensions to mildly nonlinear and nongaussian situations of algorithms which achieve Bayesian estimation in linear and gaussian situations.

- These two classes of algorithms produce useful results (and of comparable quality).

- They are far fom optimality. There is no obvious way for improving on them, but ensemble methods, meant to produce a sample of the sought-for conditional probability distribution, seem most promising. Fully bayesian particle filters are the subject of active research.

Conclusions (continued)

Assimilation, which originated from the need of defining initial conditions for numerical weather forecasts, has progressively extended to many diverse applications

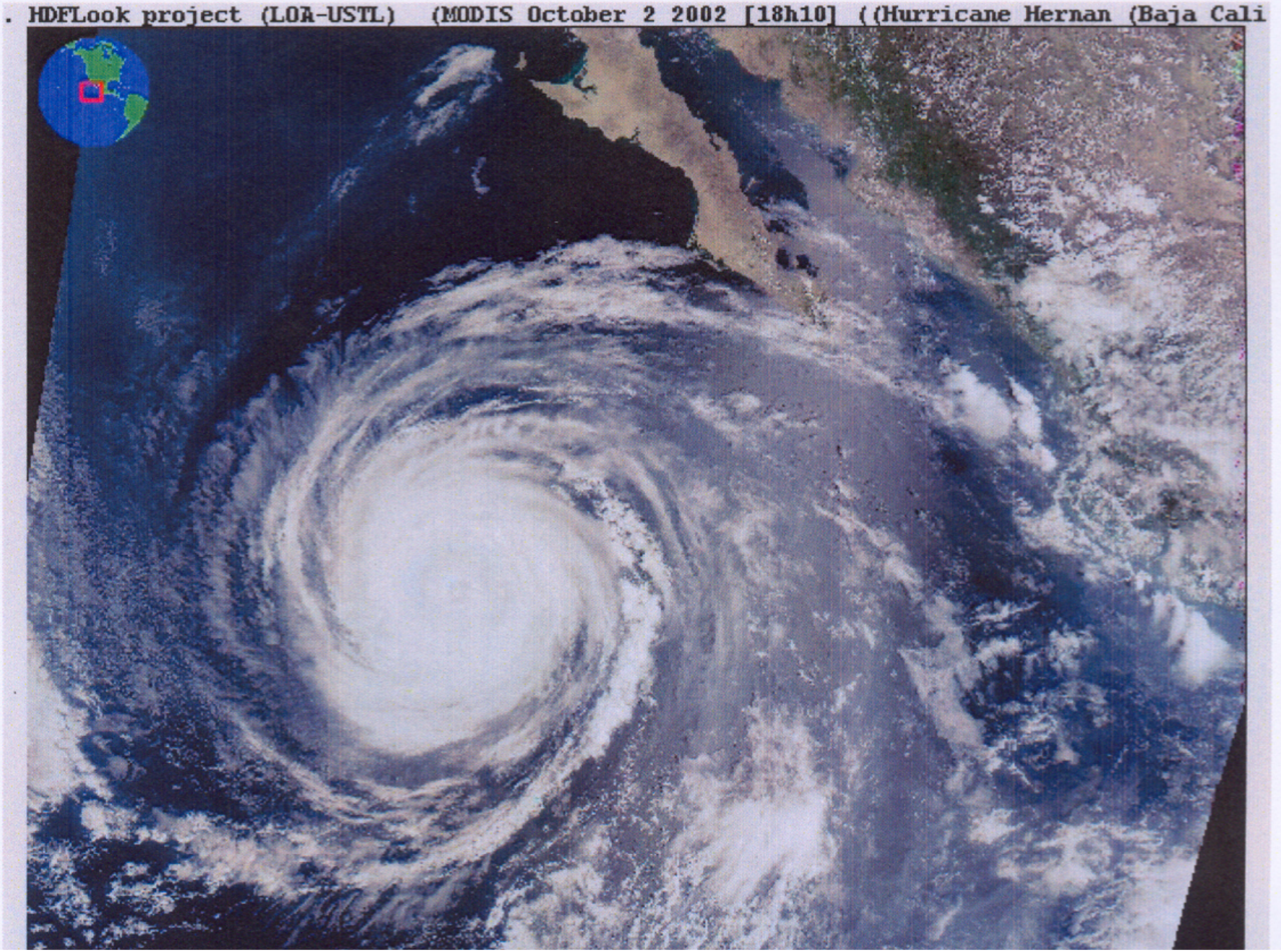
- Oceanography
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciology
- Magnetism (both planetary and stellar)
- Plate tectonics
- Planetary atmospheres (Mars, …)
- Reassimilation of past observations (mostly for climatological purposes, ECMWF, NCEP/NCAR)
- Identification of source of tracers
- Parameter identification
- *A priori* evaluation of anticipated new instruments
- Definition of observing systems *(Observing Systems Simulation Experiments)*
- Validation of models
- Sensitivity studies (adjoints)
- …

Assimilation is related to

- Estimation theory
- Probability theory
- Atmospheric and oceanic dynamics
- Atmospheric and oceanic predictability
- Instrumental physics
- Optimisation theory
- Control theory
- Algorithmics and computer science
- …

A few of the (many) remaining problems (in addition to problems raised by practical numerical implementation):

- Observability (data are noisy, system is chaotic !)
- More accurate identification and quantification of errors affecting data particularly the assimilating model (will always require independent hypotheses)
- Assimilation of images
- …

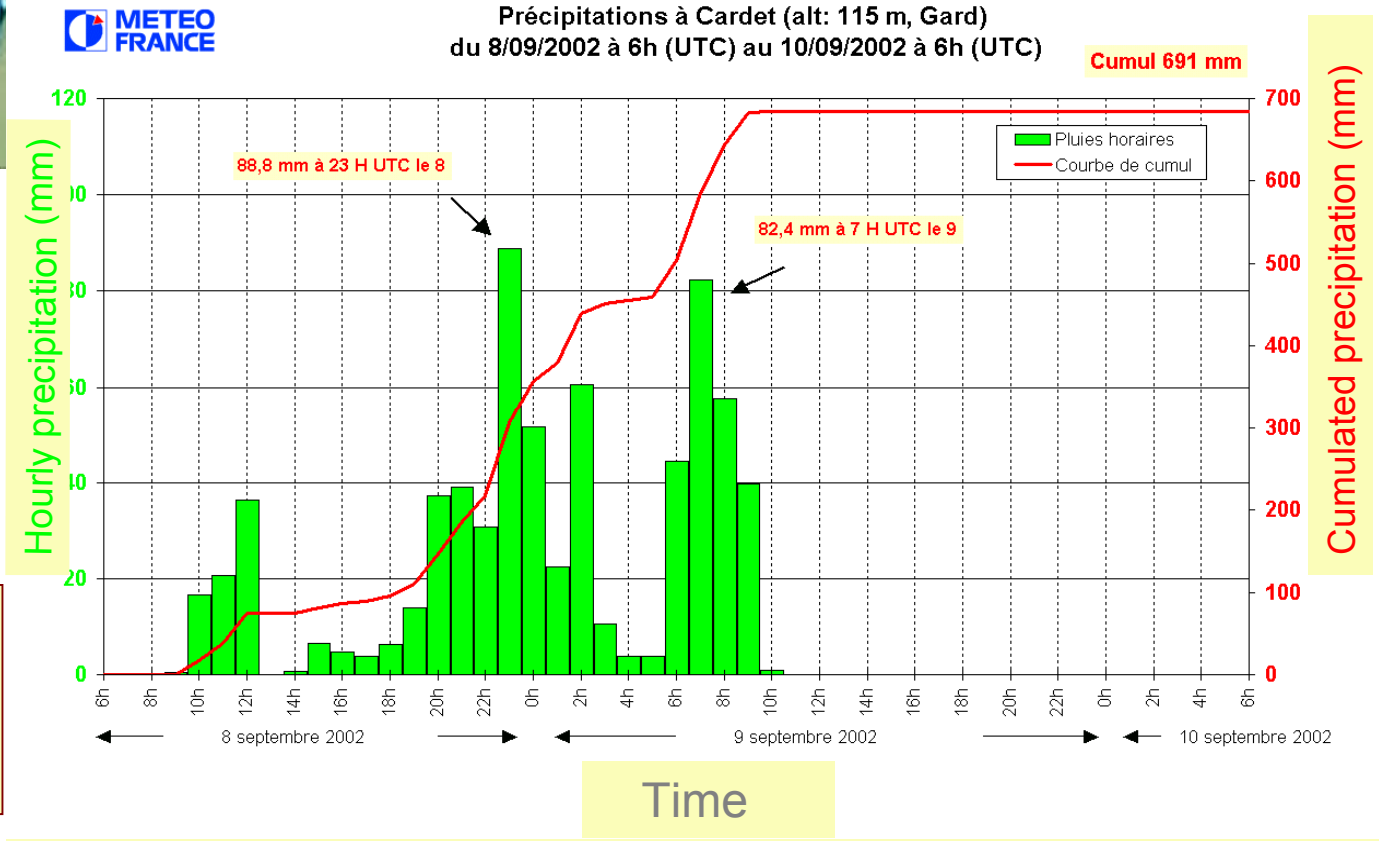

. HDFLook project (LOA-USTL) (MODIS October 2 2002 [18h10] ((Hurricane Hernan (Baja Cali

© Midi-Libre

8-9 September 2002 :

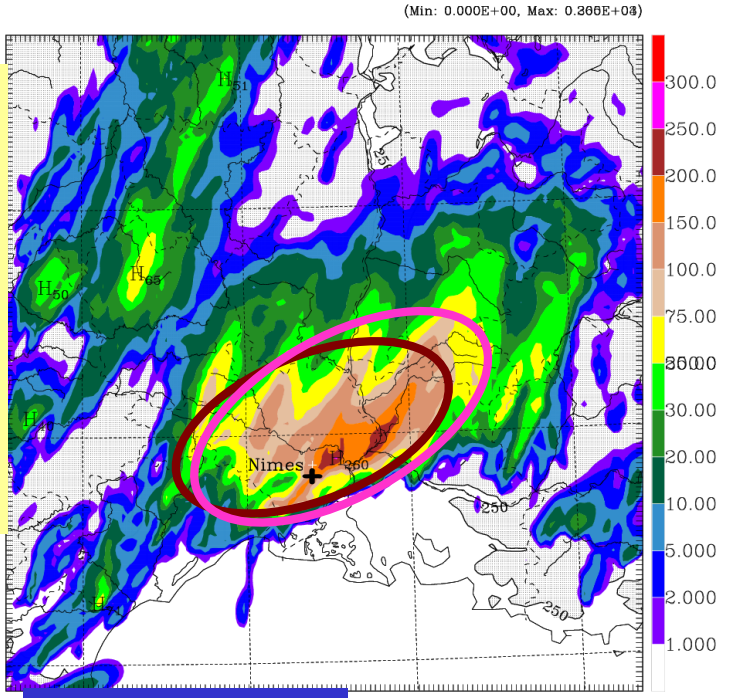
More than 20 deaths

Economic loss : 1,6 billion €

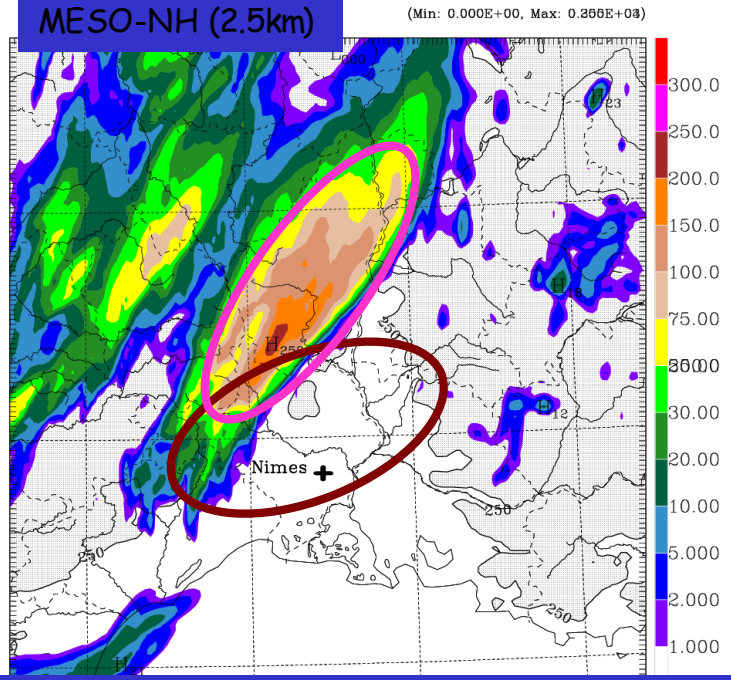


The accumulated rainfall totals have locally reached 600-700 mm

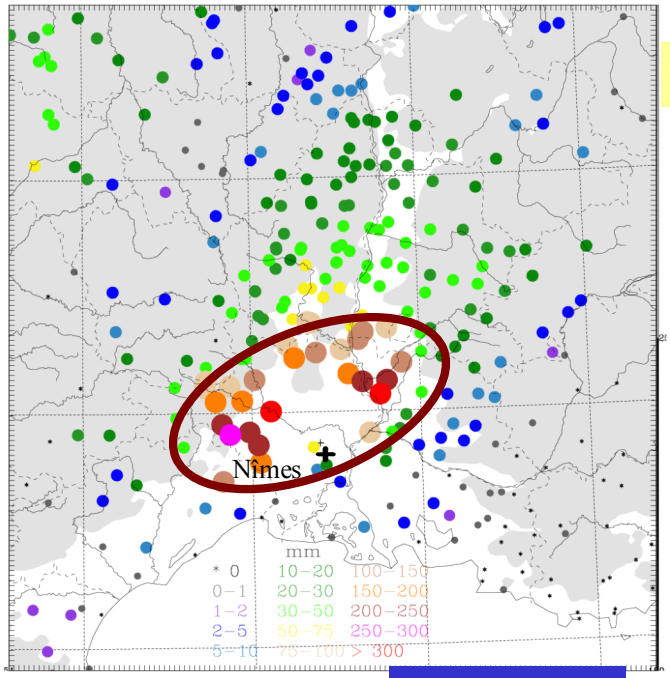
Initial conditions : mesoscale analysis (surface obs, radar, satellite) for 12UTC, 8th Sept. 2002



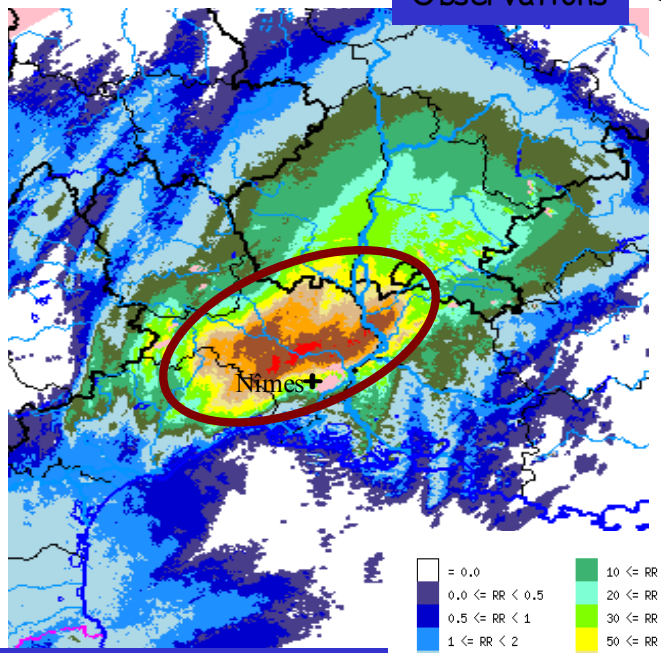
MESO-NH (2.5km)



Raingauges



Observations



Ducrocq et al, 2003

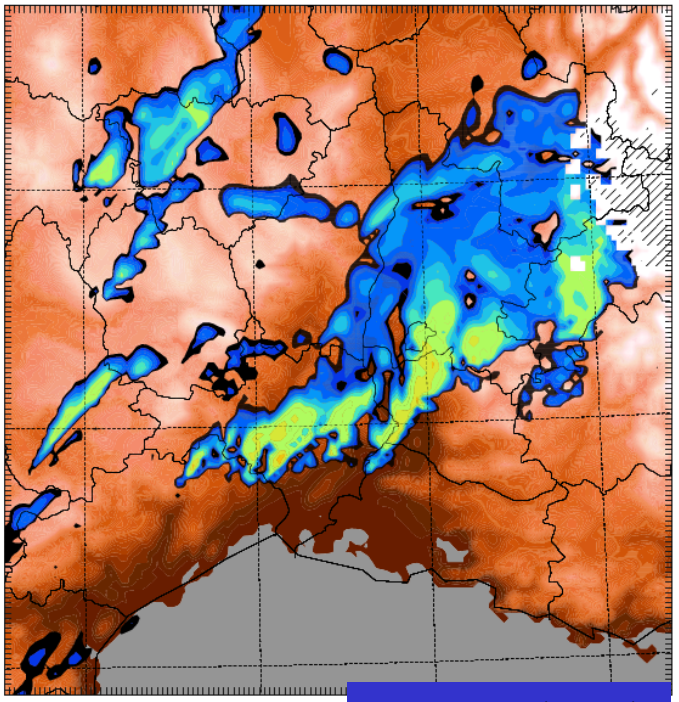
Initial conditions : large scale ARPEGE analysis for 12UTC, 8th Sept. 2002

Nîmes radar

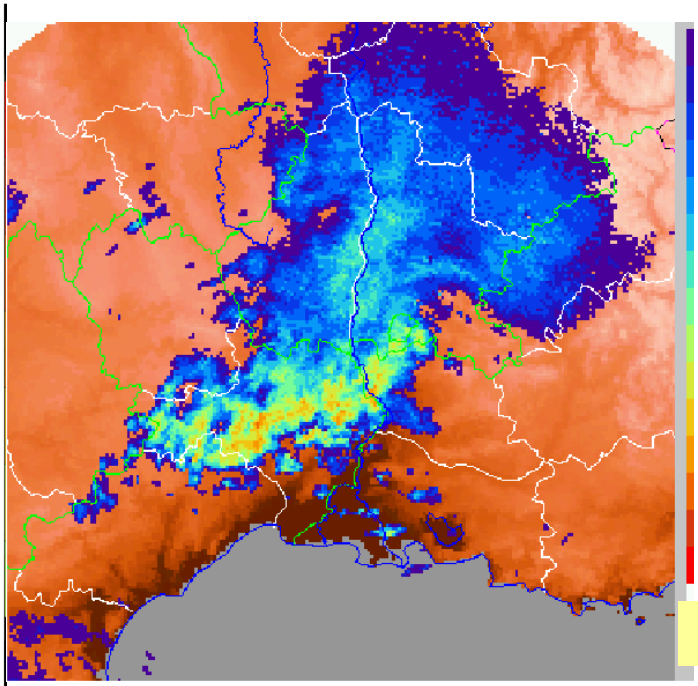
12-h accumulated rainfall from 12 UTC, 8 Sept to 0 UTC, 9 Sept 2002

Initial conditions : mesoscale analysis (surface obs, radar, satellite) for

12UTC, 8th Sept. 2002

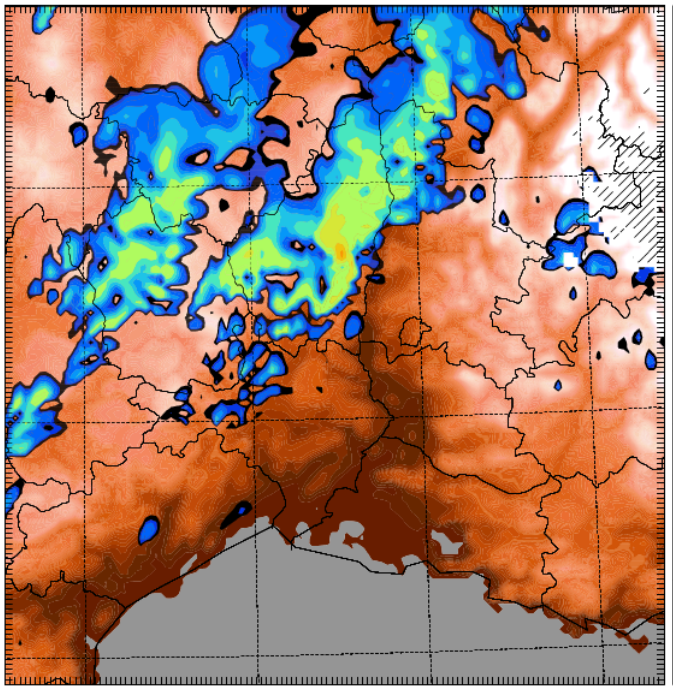


MESO-NH (2.5km)



Nîmes radar

Observations



Initial conditions : large scale ARPEGE analysis for

12UTC, 8th Sept. 2002

Reflectivities at 21 UTC, 8 Sept.