

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

Année Universitaire 2016-2017

Modélisation Numérique  
de l'Écoulement Atmosphérique  
et Assimilation de Données

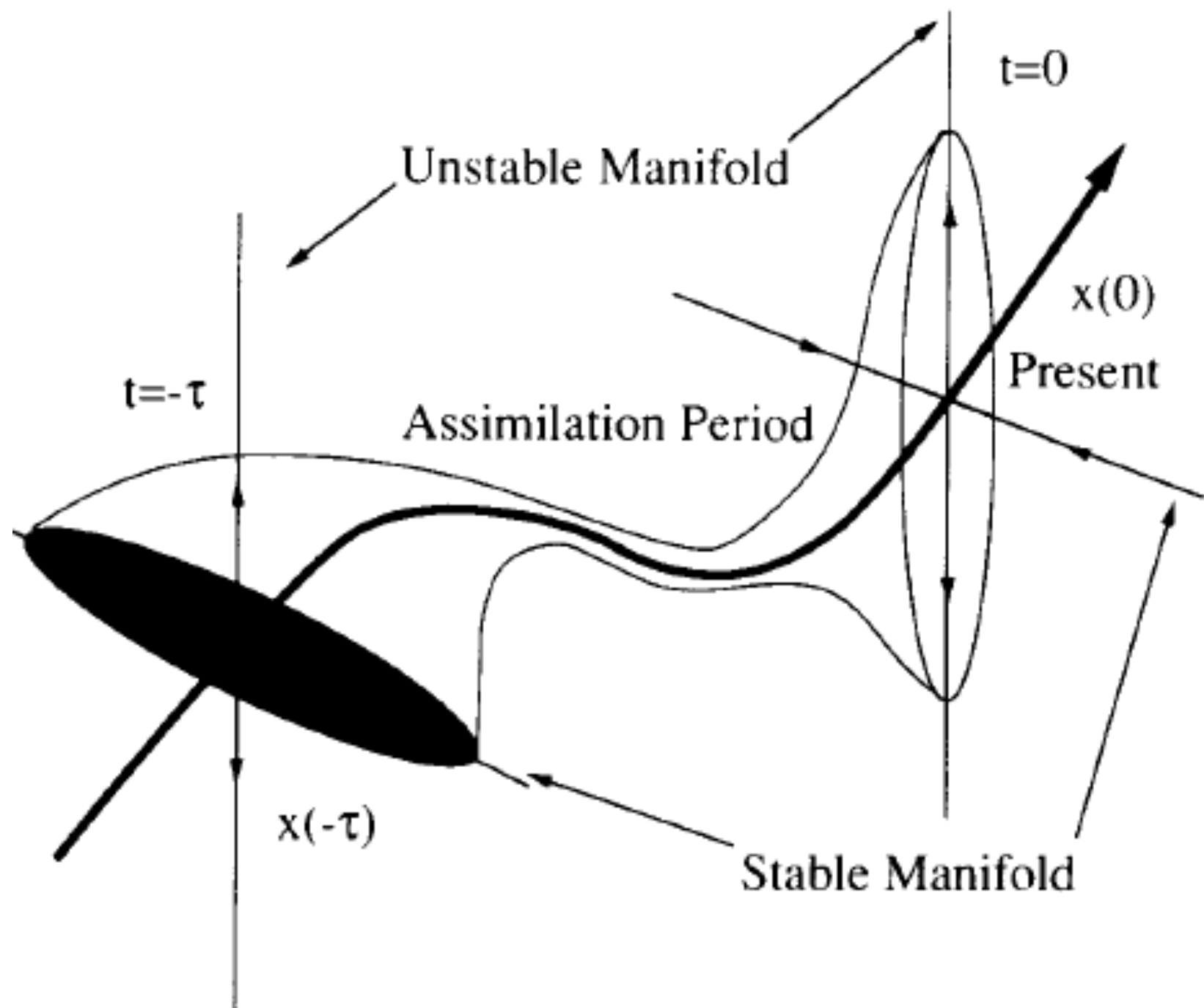
Olivier Talagrand

Cours 8

22 Juin 2017

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.



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.

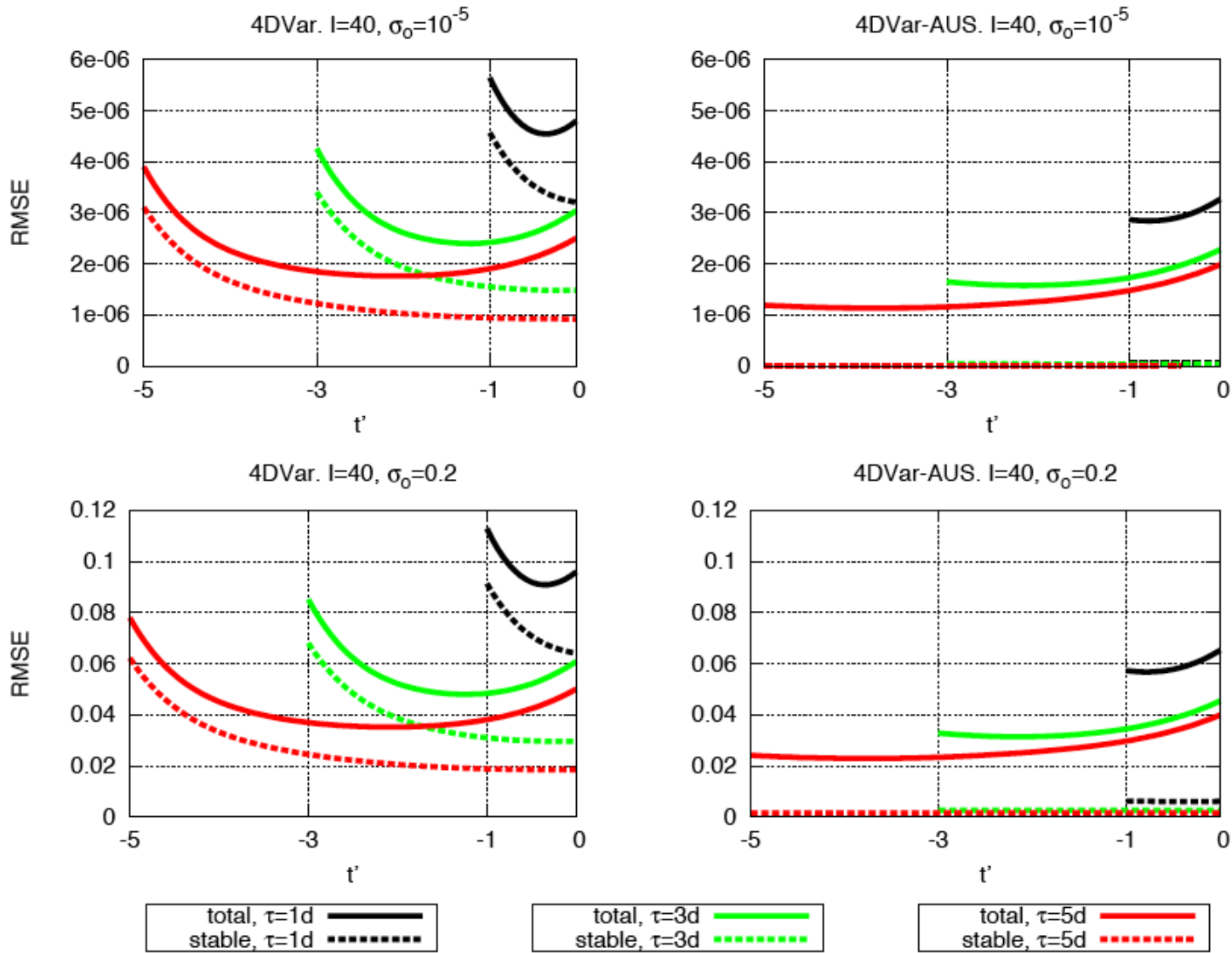
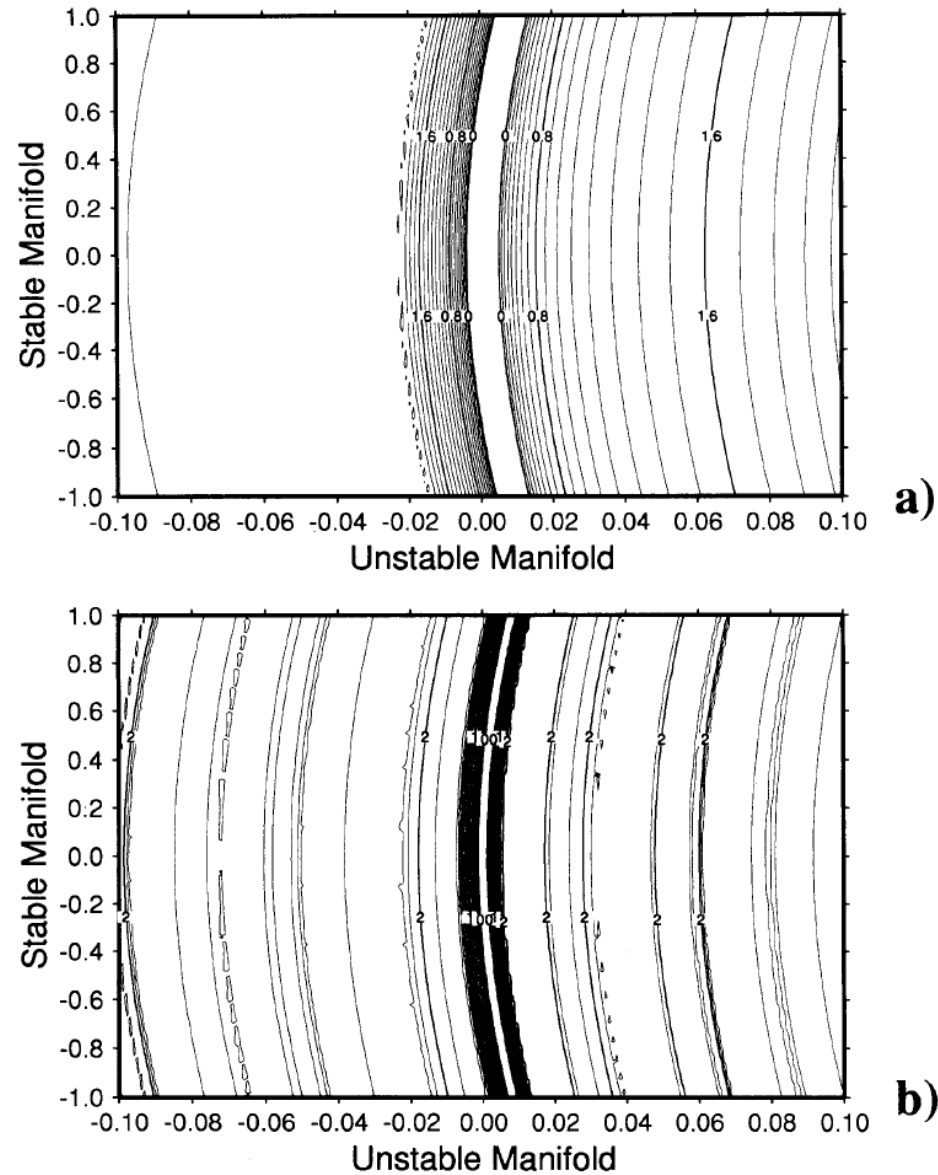


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of  $t' = t - \tau$ , with  $\sigma_0 = .2, 10^{-5}$  for the model configuration  $I = 40$ . Left panel: 4DVar. Right panel: 4DVar-AUS with  $N = 15$ . Solid lines refer to total assimilation error, dashed lines refer to the error component in the stable subspace  $e_{16}, \dots, e_{40}$ .



*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.

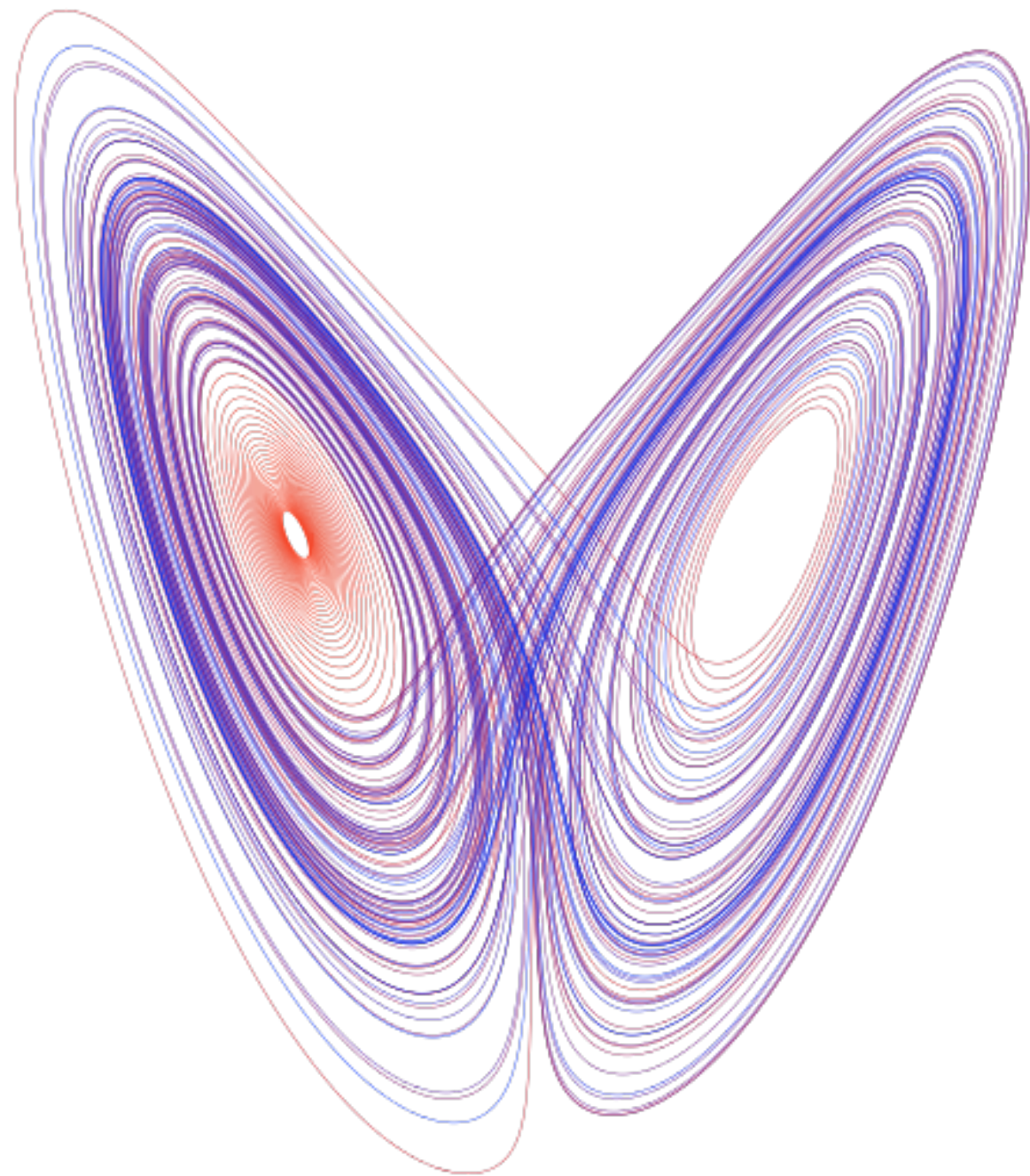
Lorenz (1963)

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

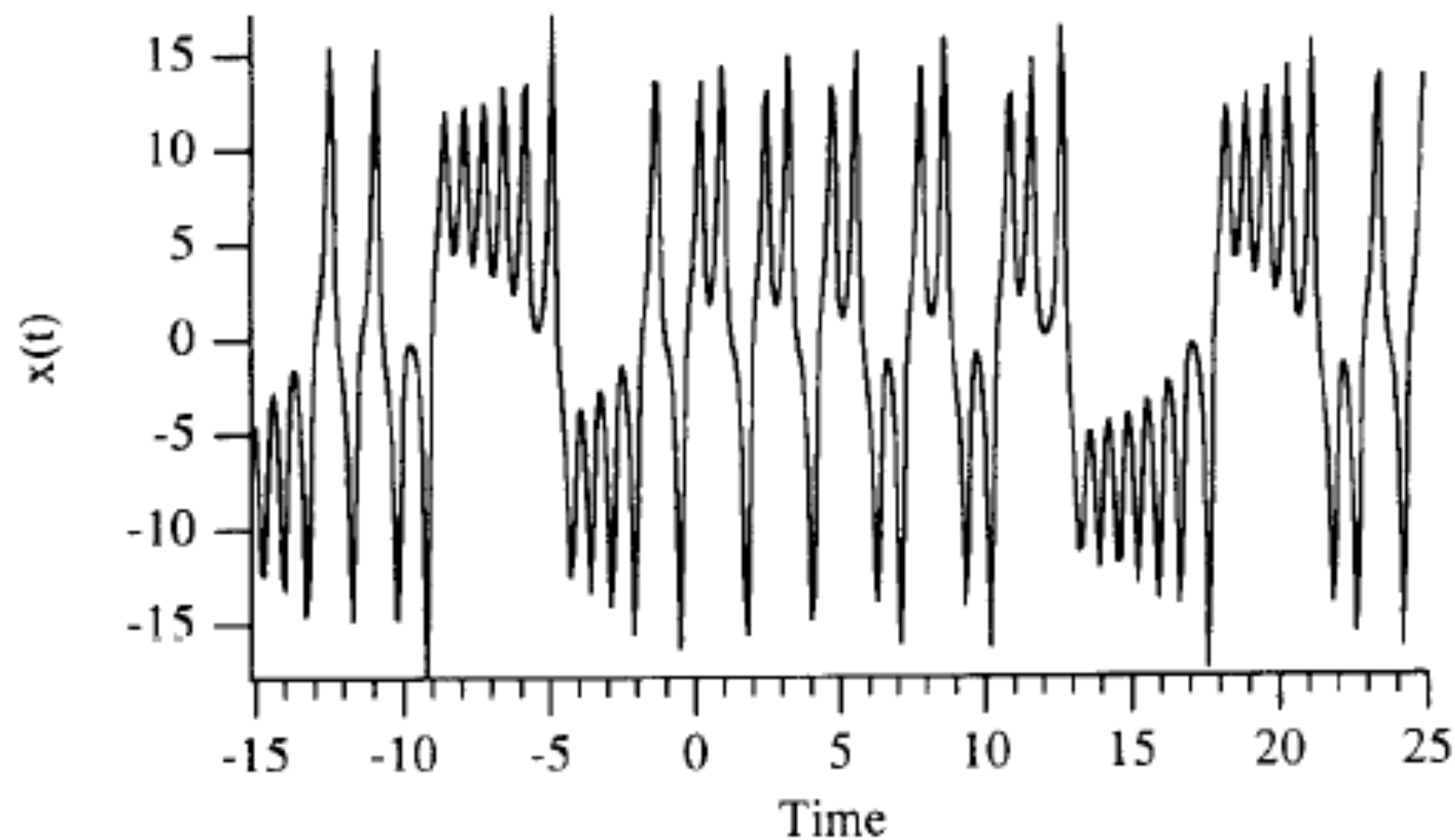
$$dy/dt = \rho x - y - xz$$

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

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





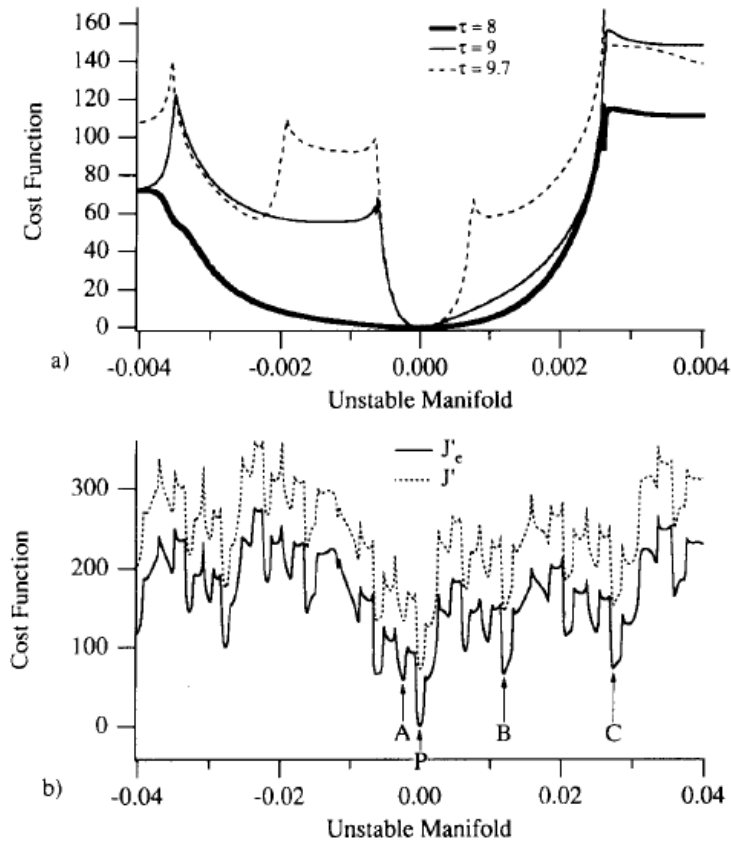


*Fig. 2.* Time variations, along the reference solution, of the variable  $x(t)$  of the Lorenz system.

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

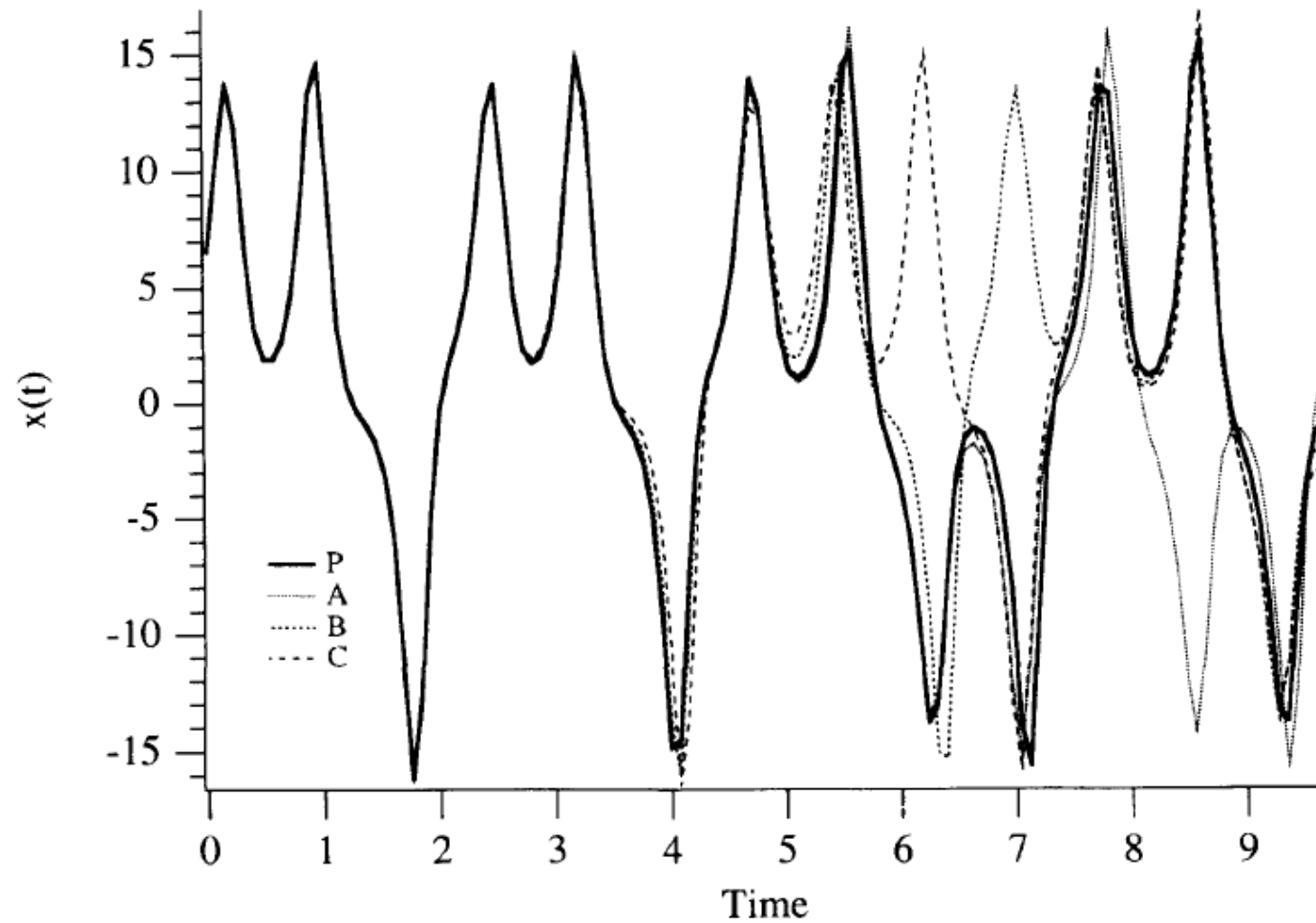
$$J(\xi_0) = (1/2) \sum_k [y_k - H_k \xi_k]^T R_k^{-1} [y_k - H_k \xi_k]$$

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  $\tau$ . 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.*)

.

| $\mu(C(\tau, x))$ | 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        |
| $\tau = 2$        | $5.9 \times 10^{-2}$ | 5.74                             | $4.5 \times 10^{-2}$  | 0.401       |
| $\tau = 3$        | $3.3 \times 10^{-2}$ | 29.4                             | $2.9 \times 10^{-2}$  | 0.397       |
| $\tau = 8$        | $1.4 \times 10^{-2}$ | 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).

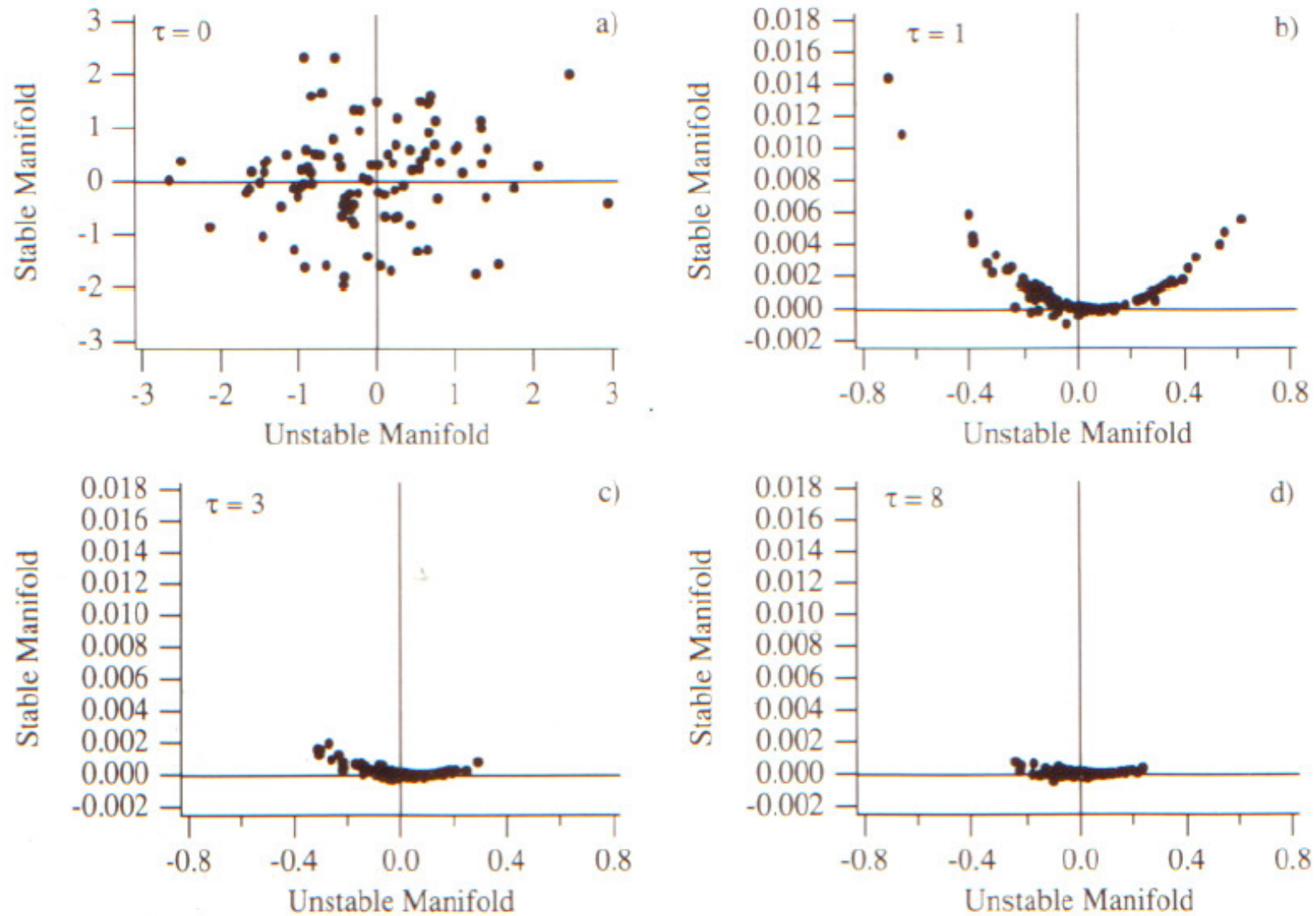


Fig. 7. Projection of the 100 minimizing solutions, at the end of the assimilation period, onto the plane spanned by the stable and unstable directions, defined as in Fig. 3. Values of  $\tau$  are indicated on the panels. The projection is not an orthogonal projection, but a projection parallel to the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

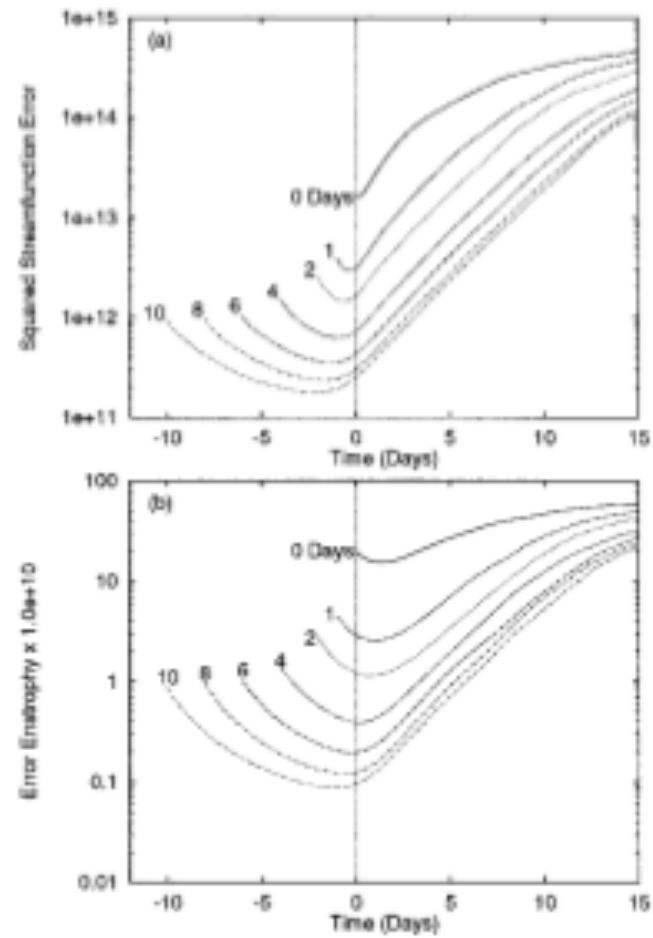


Fig. 5. Median values of the (a) streamfunction squared error, and (b) enstrophy error for the 200 forecast set as a function of forecast time and of the assimilation time  $T_a$ .

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

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

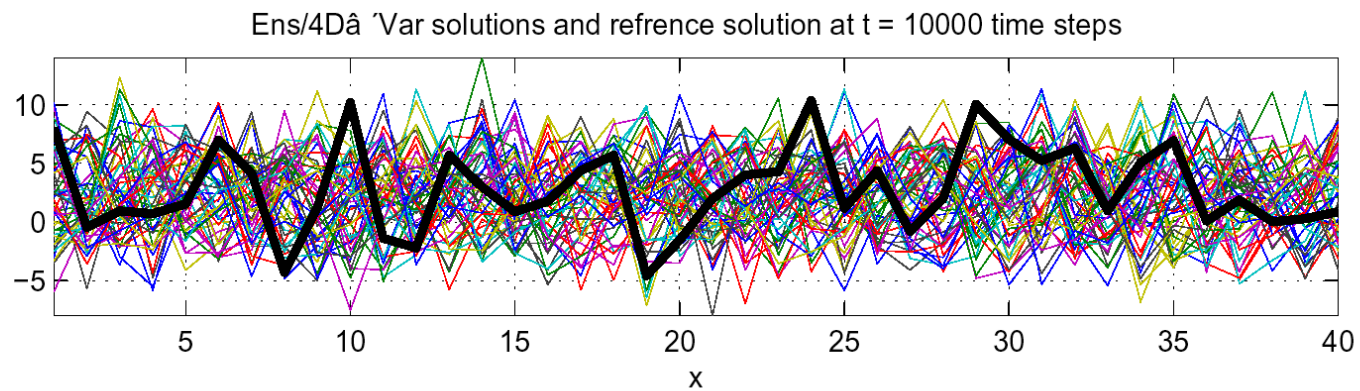
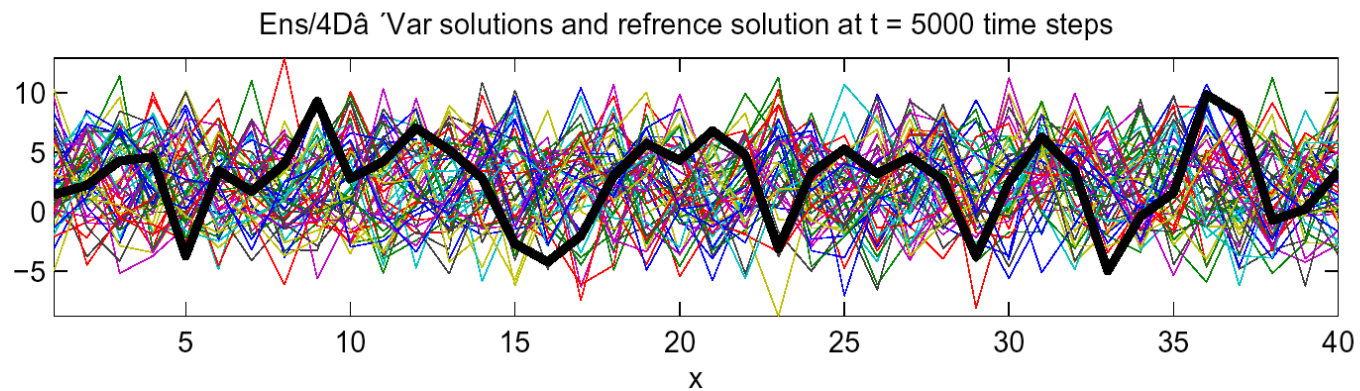
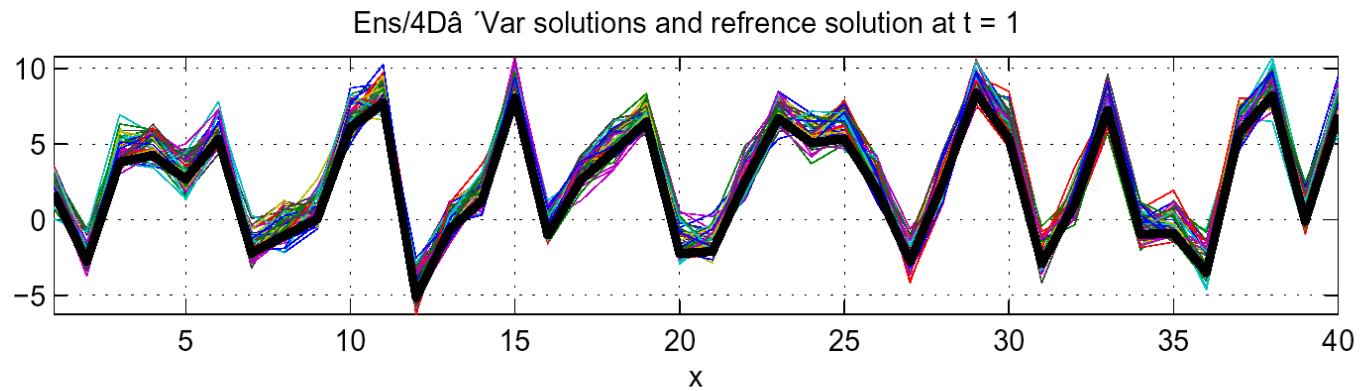
with  $j = 1, \dots, I$ .

with periodic conditions in  $j$ , and 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).

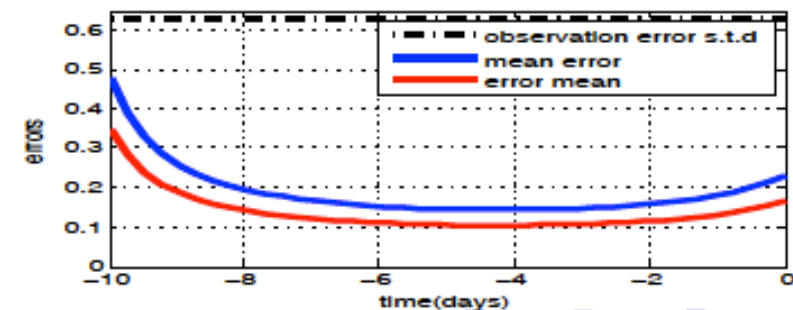
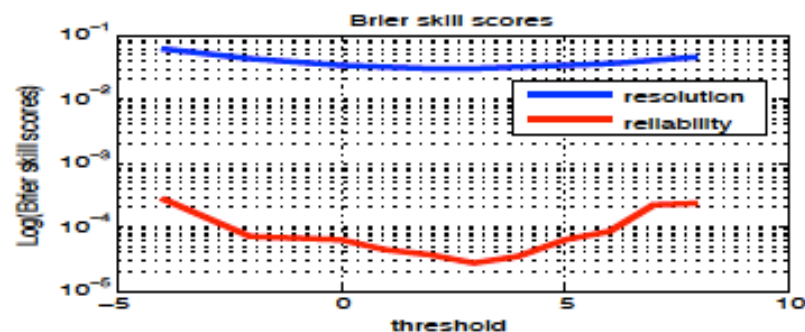
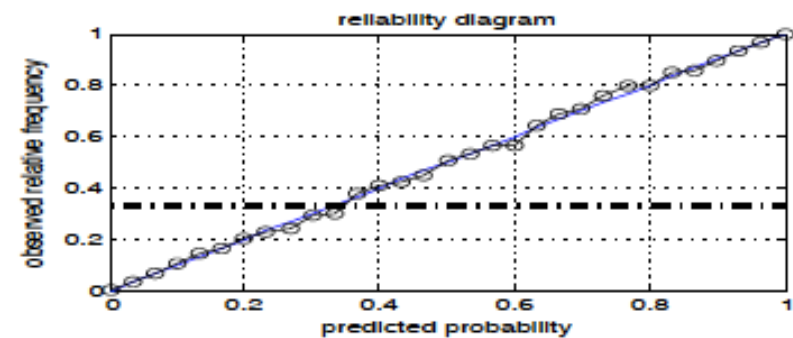
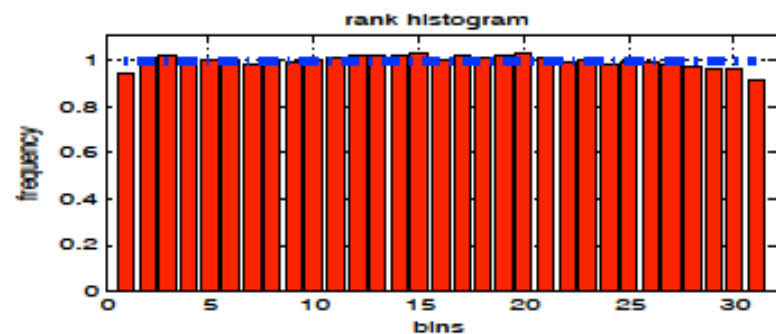
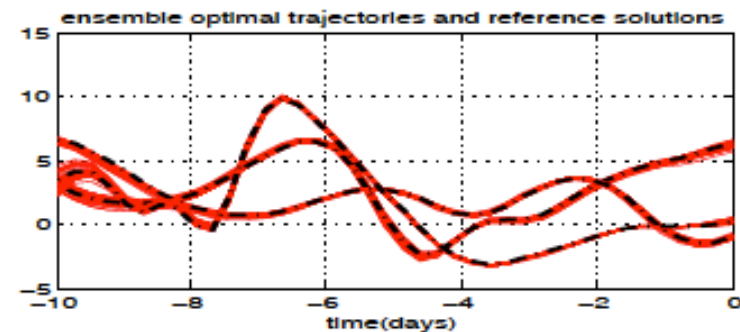
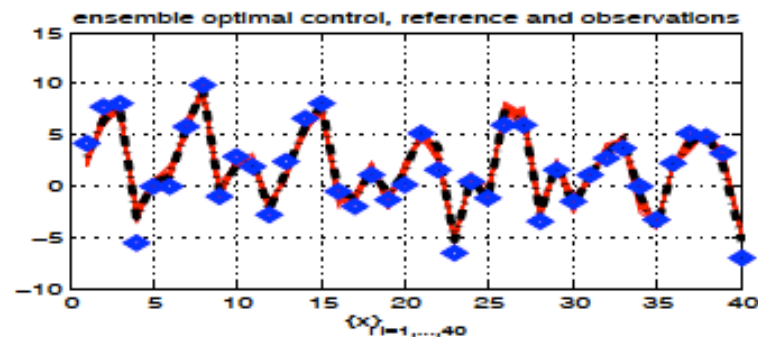
Identical twin experiments (perfect model)



Lorenz'96 model (M. Jardak)

---

# EnsVar : the non-linear Lorenz96 model 10 days with QSVA



‘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  $\tau = 1, \dots, 5$  days. Results are averaged over 5000 successive windows.

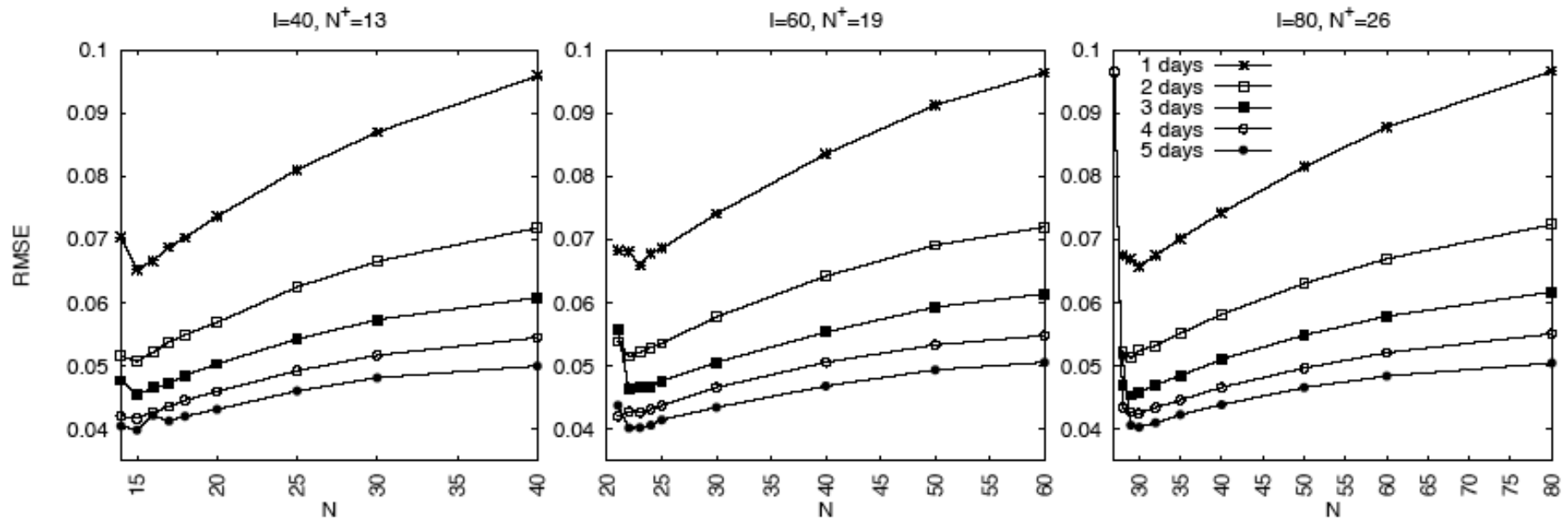


Figure 1. Time average RMS analysis error at  $t = \tau$  as a function of the subspace dimension  $N$  for three model configurations:  $I=40, 60, 80$ . Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is  $\sigma_o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

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.



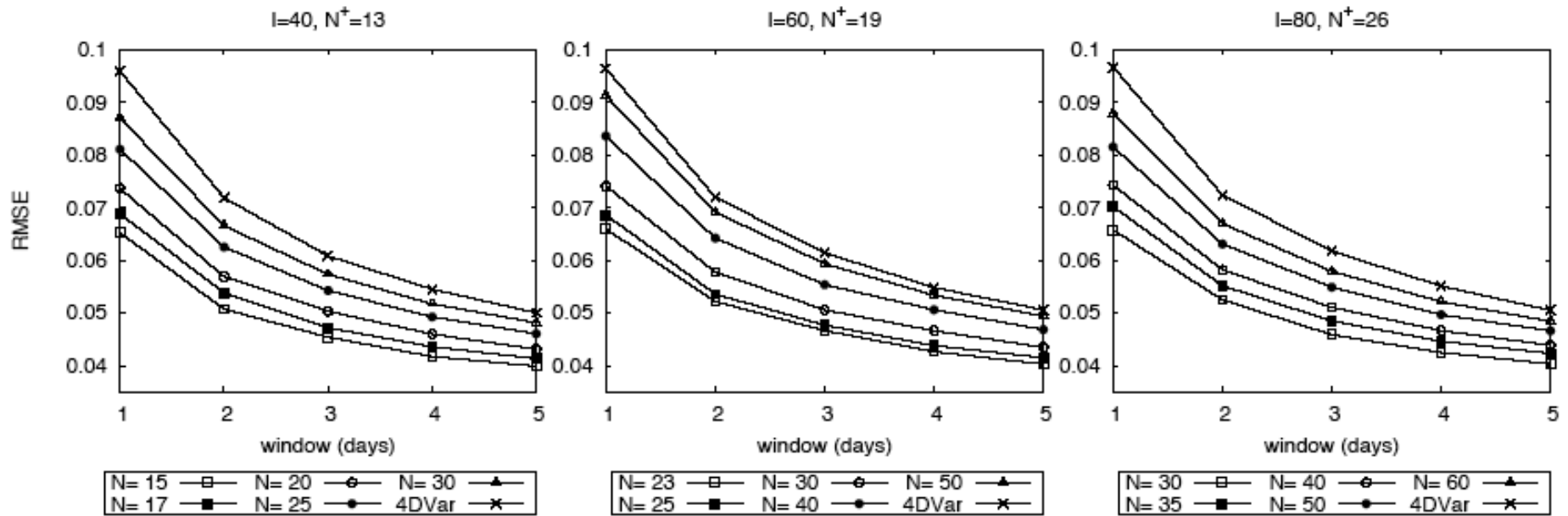


Figure 2. Time average RMS analysis error at  $t = \tau$  as a function of the length of the assimilation window for three model configurations:  $I=40, 60, 80$ . Different curves in the same panel refer to a different subspace dimension  $N$  of 4DVar-AUS and to standard 4DVar.  $\sigma_o = 0.2$ .

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  $\tau$ .

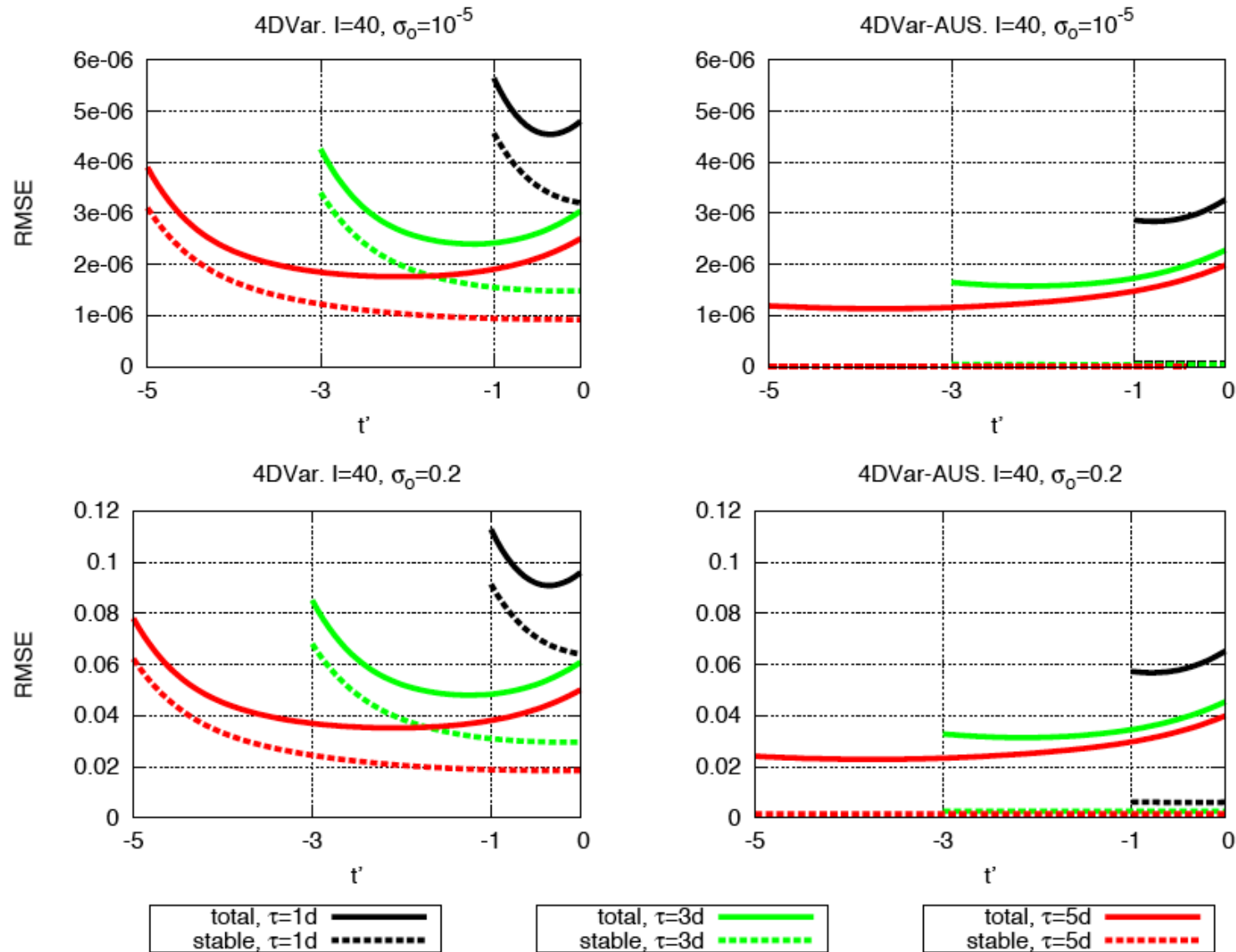


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of  $t' = t - \tau$ , with  $\sigma_0 = .2, 10^{-5}$  for the model configuration  $I = 40$ . Left panel: 4DVar. Right panel: 4DVar-AUS with  $N = 15$ . Solid lines refer to total assimilation error, dashed lines refer to the error component in the stable subspace  $e_{16}, \dots, e_{40}$ .

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).

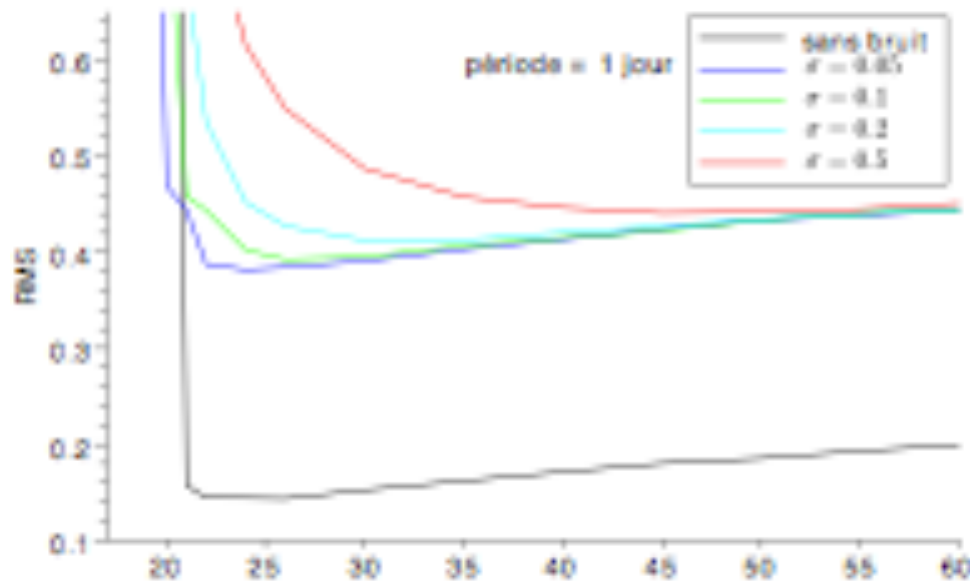
Results show that, if all degrees of freedom that are available to the model are used, the minimization process introduces components along the stable modes of the system, in which no error is present, in order to ensure a closer fit to the observations. This degrades the closeness of the fit to reality.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

These results apply because no explicit background is available at the initial time of the assimilation window (only the unstable subspace is known). A proper background (obtained for instance from a properly implemented Kalman Filter, or from an Ensemble Variational Assimilation) would not only say the the uncertainty is restricted to the unstable space, but how it is distributed in that subspace. The ‘restriction’ to the unstable subspace would be automatically made.

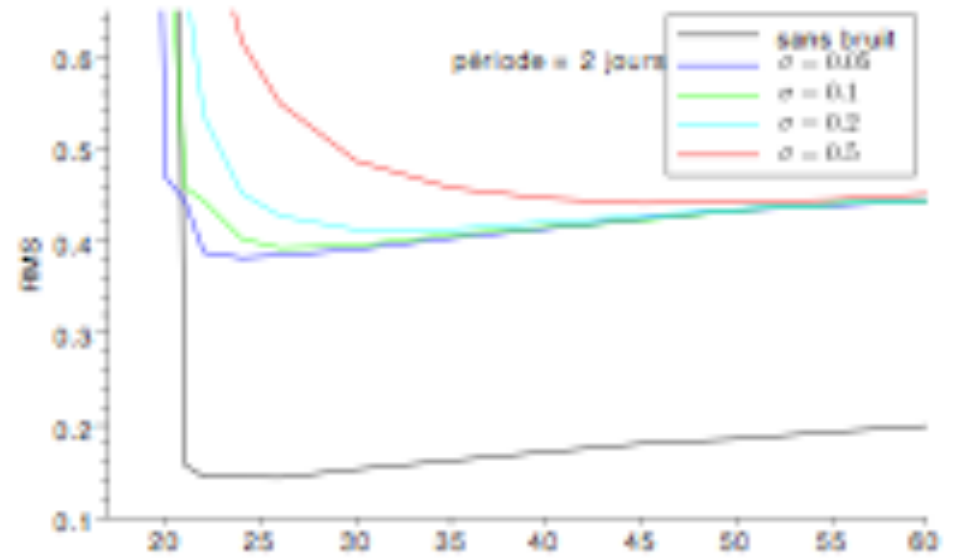
Can have major practical algorithmic implications.

## Questions.

- Degree of generality of results ?
- Impact of model errors ?



$\tau = 1$  day



$\tau = 2$  days

Time averaged rms analysis error at the end of the assimilation window (with length  $\tau$ ) as a function of increment subspace dimension ( $I = 60, N^+=19$ ), for different amplitudes of white model noise.

(W. Ohayon and O. Pannekoucke, 2011).

## Conclusions

Error concentrates in unstable modes at the end of assimilation window. It must therefore be sufficient, at the beginning of new assimilation cycle, to introduce increments only in the subspace spanned by those unstable modes.

In the perfect model case, assimilation is most efficient when increments are introduced in a space with dimension slightly above the number of non-negative Lyapunov exponents.

In the case of imperfect model (and of strong constraint assimilation), preliminary results lead to similar conclusions, with larger optimal subspace dimension, and less well marked optimality. Further work necessary.

In agreement with theoretical and experimental results obtained for Kalman Filter assimilation (Trevisan and Palatella, McLaughlin).

# *Ensemble Variational Assimilation and Bayesian Estimation*

Olivier Talagrand<sup>1</sup> and Mohamed Jardak<sup>1,2</sup>

1. Laboratoire de Météorologie Dynamique/IPSL

École Normale Supérieure, Paris, France

2. Meteorological Office, Exeter, UK

Workshop

*Theoretical aspects of ensemble data assimilation  
for the Earth system*

Les Houches, France

6 April 2015



## Assimilation considered as a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (the *data*), viz.,

- observations proper
- physical laws governing the system ('*model*')
- ...

Jaynes, E. T., 2003, *Probability theory: the logic of science*, Cambridge University Press

Tarantola, A., 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics (<http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf>)

## Data of the form

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

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

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

$\Gamma$  known ( $m \times n$ )-matrix,  $\zeta$  unknown 'error'

Then conditional probability distribution is

$$P(x | z) = \mathcal{N}[x^a, P^a]$$

where

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

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

*Determinacy condition* :  $\text{rank} \Gamma = n$ . Requires  $m \geq n$ . Set  $p \equiv m - n$

## Variational form.

Conditional expectation  $x^a$  minimizes following scalar *objective function*, defined on state space  $\mathcal{X}$

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

*Variational assimilation*, implemented heuristically in many places on (not too much) nonlinear data operators  $\Gamma$ .

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

Conditional probability distribution

$$P(x | z) = \mathcal{N}[x^a, P^a]$$

with

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

Ready recipe for determining Monte-Carlo sample of conditional pdf  $P(x | z)$  :

- Perturb data vector  $z$  according to its own error probability distribution

$$z \rightarrow z' = z + \delta, \quad \delta \sim \mathcal{N}[0, S]$$

and compute

$$x'^a = (\Gamma^T S^{-1} \Gamma)^{-1} \Gamma^T S^{-1} [z' - \mu]$$

$x'^a$  is distributed according to  $\mathcal{N}[x^a, P^a]$

*Ensemble Variational Assimilation (EnsVar)* implements that algorithm, the expectations  $x'^a$  being computed by standard variational assimilation (optimization)

## Purpose of the present work

- Objectively evaluate EnsVar as a probabilistic estimator in nonlinear and/or non-Gaussian cases.
- Objectively compare with other existing ensemble assimilation algorithms : *Ensemble Kalman Filter (EnKF)*, *Particle Filters (PF)*
- Simulations performed on two small-dimensional chaotic systems, the Lorenz'96 model and the Kuramoto-Sivashinsky equation

Purely heuristic !

*Conclusion.* Works very well, at least on small dimension chaotic systems, and using *Quasi-Static Variational Assimilation (QSVA)* over long assimilation periods

## Experimental procedure (1)

0. Define a *reference solution*  $x_k$  by integration of the numerical model

1. Produce ‘observations’ at successive times  $t_k$  of the form

$$y_k = H_k x_k + \varepsilon_k$$

where  $H_k$  is (usually, but not necessarily) the unit operator, and  $\varepsilon_k$  is a random (usually, but not necessarily, Gaussian) ‘observation error’.

## Experimental procedure (2)

2. For given observations  $y_k$ , repeat  $N_{ens}$  times the following process

- ‘Perturb’ the observations  $y_k$  as follows

$$y_k \rightarrow z_k = y_k + \delta_k$$

where  $\delta_k$  is an independent realization of the probability distribution which has produced  $\varepsilon_k$ .

- Assimilate the ‘perturbed’ observations  $z_k$  by variational assimilation

This produces  $N_{ens}$  (=30) model solutions over the assimilation window, considered as making up a tentative sample of the conditional probability distribution for the state of the observed system over the assimilation window.

The process 1-2 is then repeated over  $N_{real}$  successive assimilation windows. Validation is performed on the set of  $N_{real}$  (=9000) ensemble assimilations thus obtained.



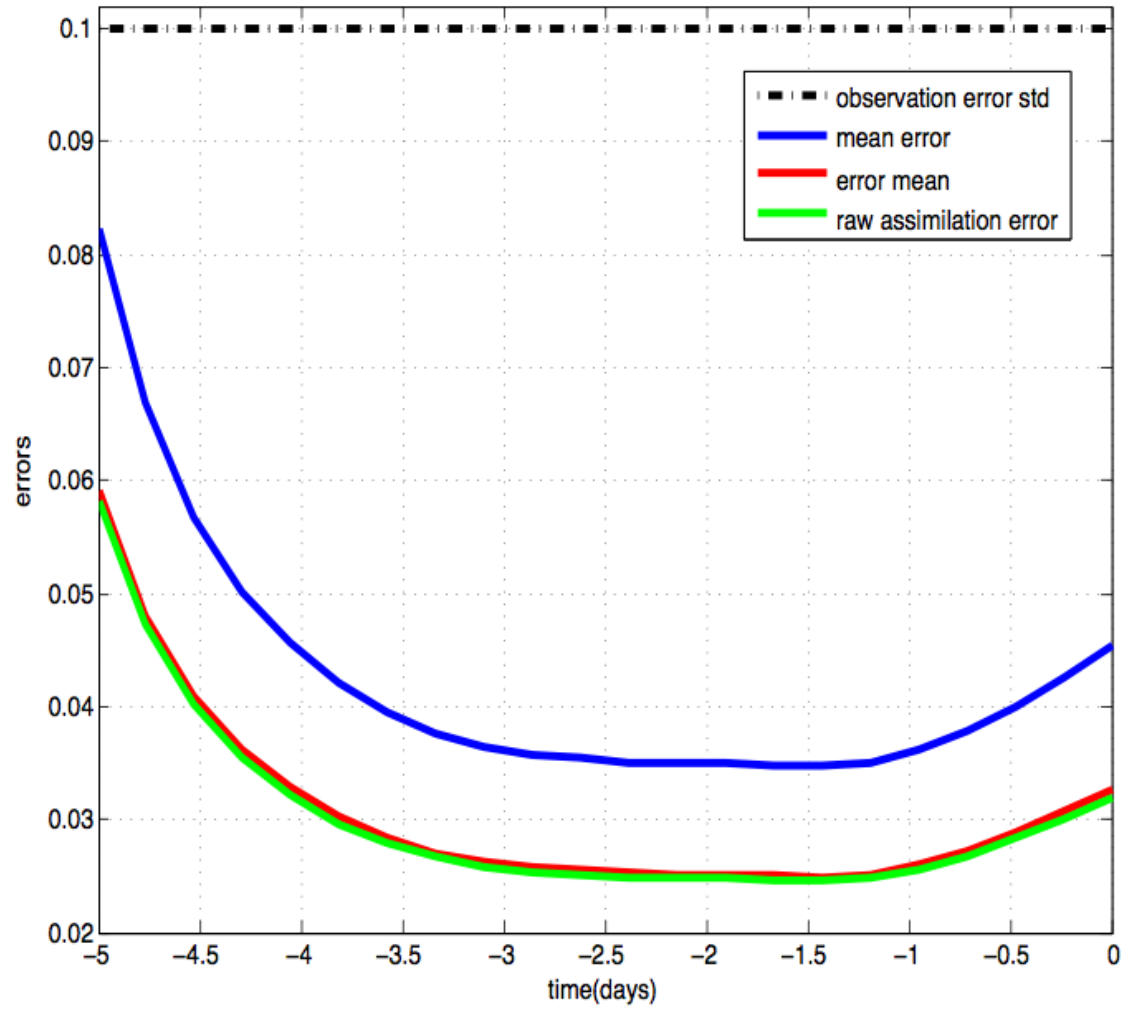
# The Lorenz96 model

- Forward model

$$\frac{dx_k}{dt} = (x_{k+1} - x_{k-2})x_{k-1} - x_k + F \quad \text{for } k = 1, \dots, N$$

- Set-up parameters :

- ① the index  $k$  is cyclic so that  $x_{k-N} = x_{k+N} = x_k$ .
- ②  $F = 8$ , external driving force.
- ③  $-x_k$ , a damping term.
- ④  $N = 40$ , the system size.
- ⑤  $N_{ens} = 30$ , number of ensemble members.
- ⑥  $\frac{1}{\lambda_{max}} \simeq 2.5days$ ,  $\lambda_{max}$  the largest Lyapunov exponent.
- ⑦  $\Delta t = 0.05 = 6hours$ , the time step.
- ⑧ frequency of observations : every 12 hours.
- ⑨ number of realizations : 9000 realizations.



Linearized Lorenz'96. 5 days

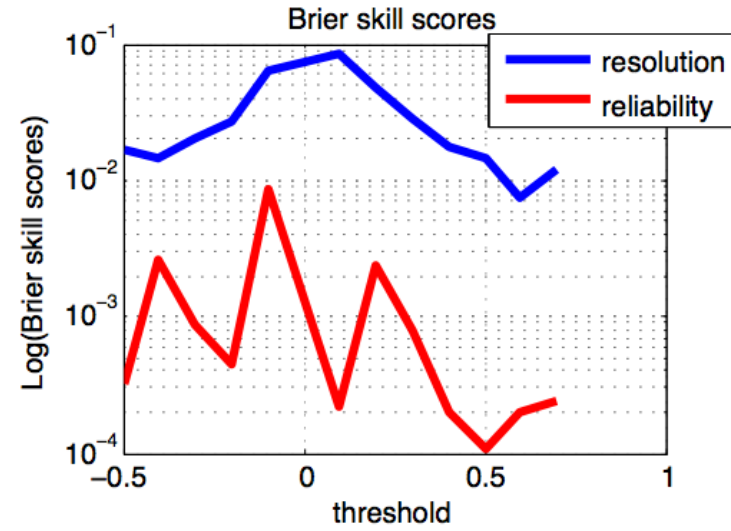
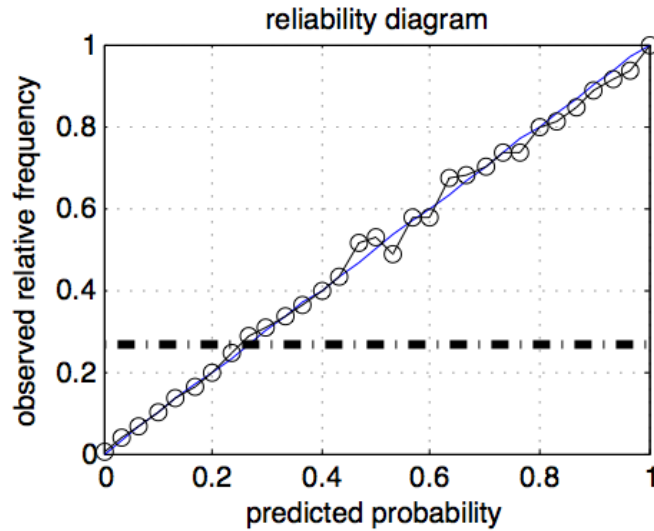
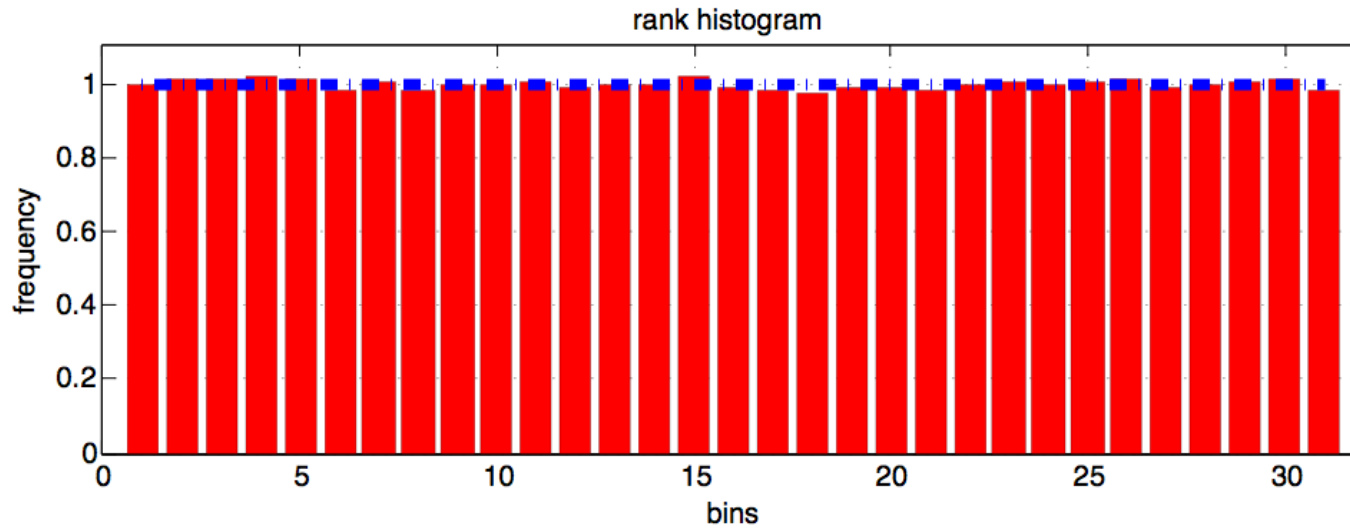
## How to objectively evaluate the performance of an ensemble (or more generally probabilistic) estimation system ?

- There is no general objective criterion for Bayesianity
- We use instead the weaker property of *reliability*, *i. e.* statistical consistency between predicted probabilities and observed frequencies of occurrence (it rains with frequency 40% in the circumstances where I have predicted 40% probability for rain).

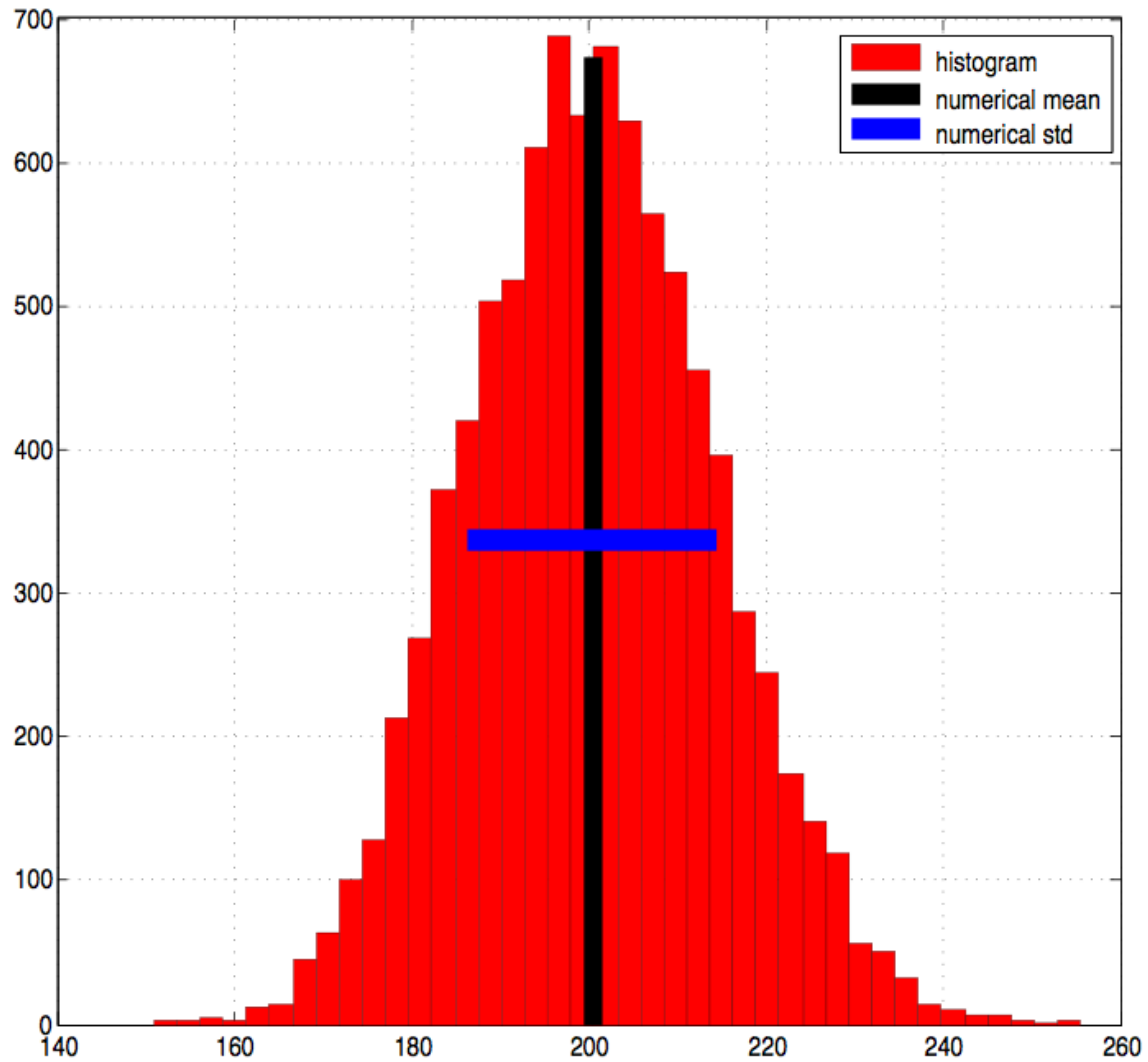
Reliability can be objectively validated, provided a large enough sample of realizations of the estimation system is available.

Bayesianity implies reliability, the converse not being true.

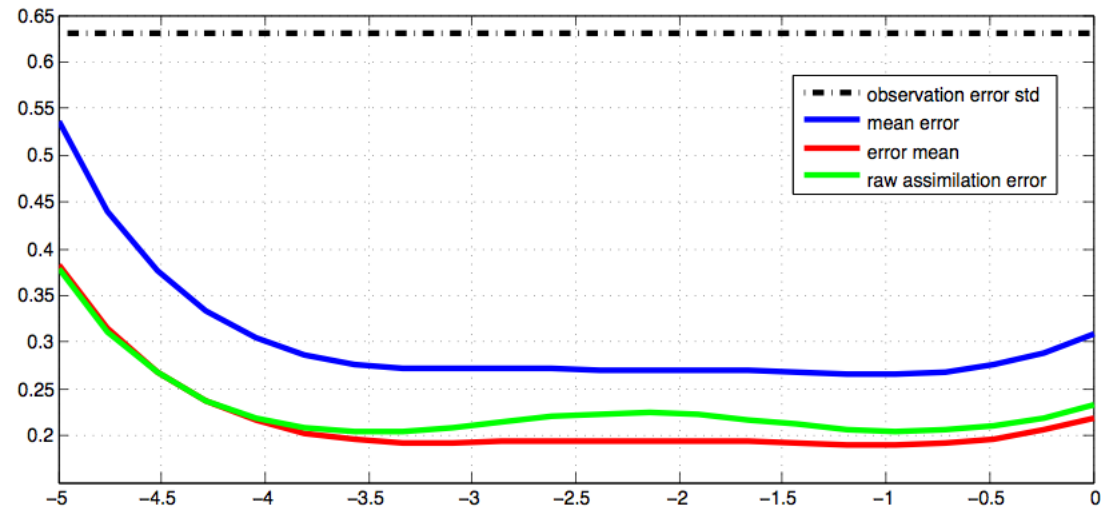
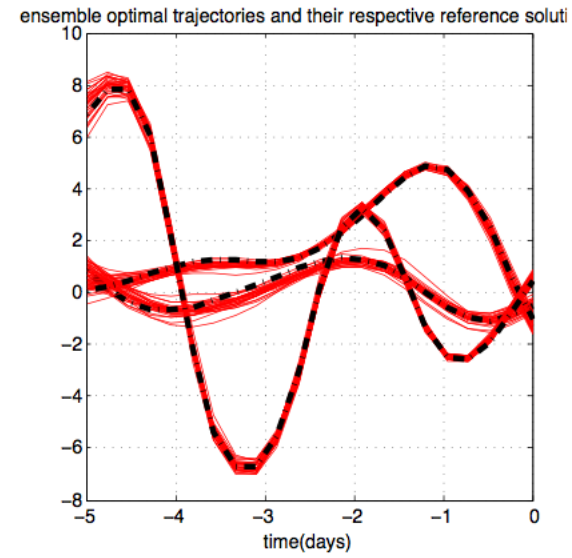
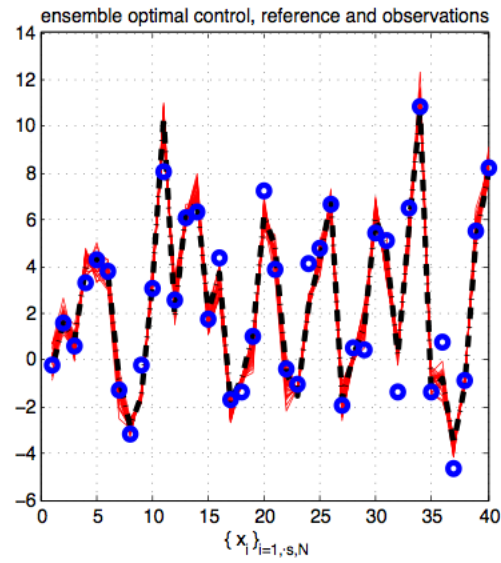
- We also evaluate *resolution*, which bears no direct relation to bayesianity, and is best defined as the degree of statistical dependence between the predicted probability distribution and the verifying observation (J. Bröcker). Resolution, beyond reliability, measures the degree of practical accuracy of the ensembles.



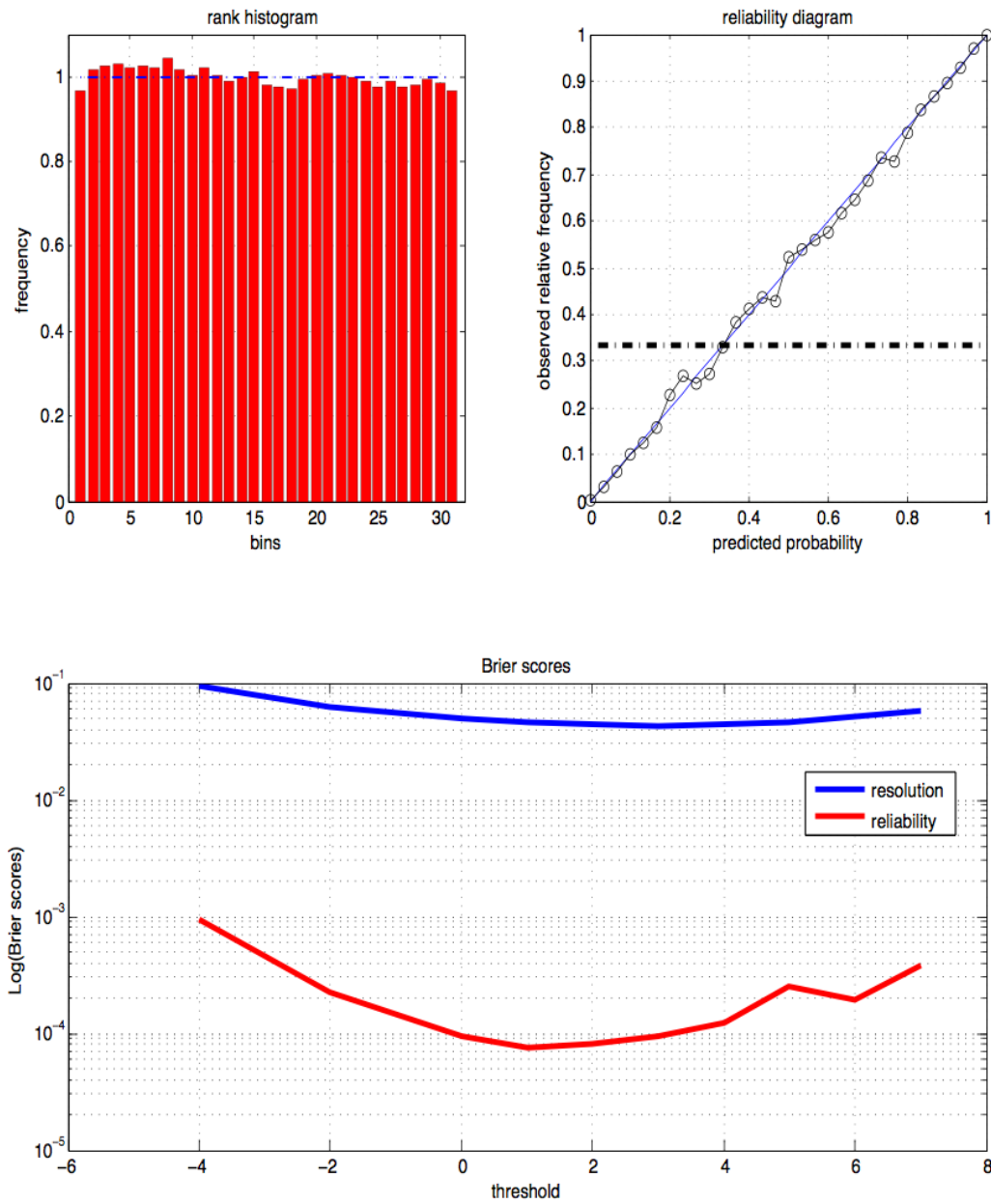
Linearized Lorenz'96. 5 days



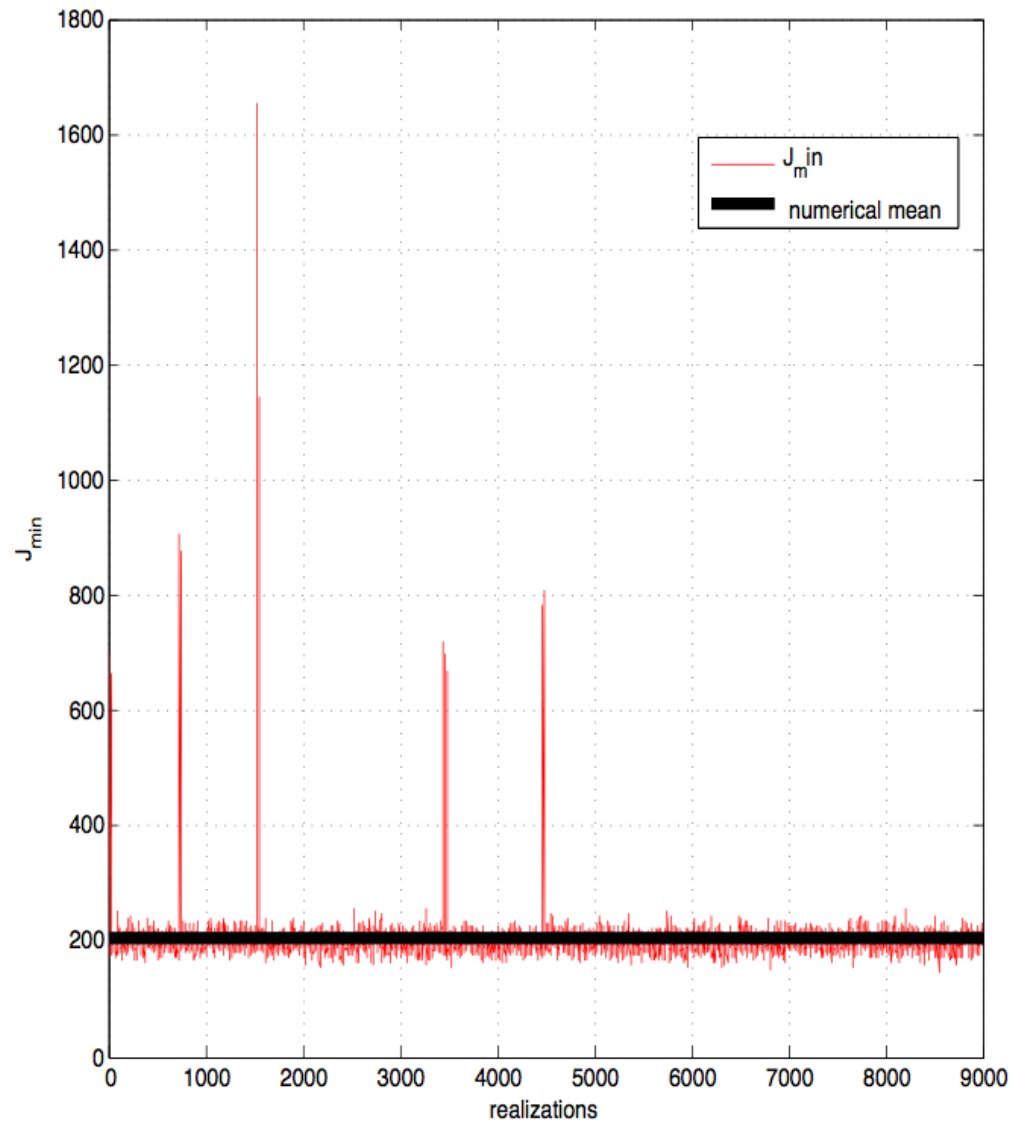
Linearized Lorenz'96. 5 days. Histogram of  $\mathcal{J}_{min}$   
 $E(\mathcal{J}_{min}) = p/2 (=200)$  ;  $\sigma(\mathcal{J}_{min}) = \sqrt{(p/2)} (\approx 14.14)$



Nonlinear Lorenz'96. 5 days



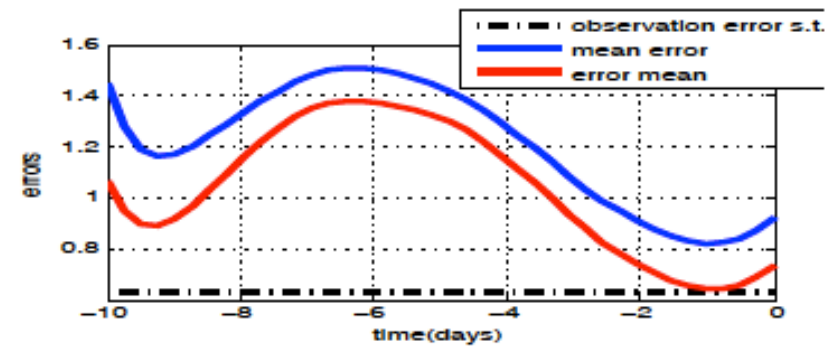
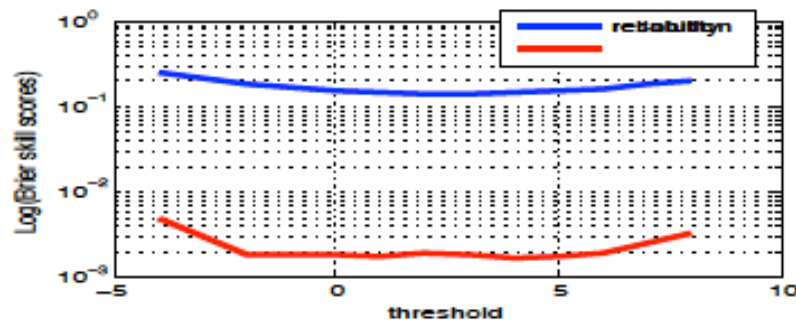
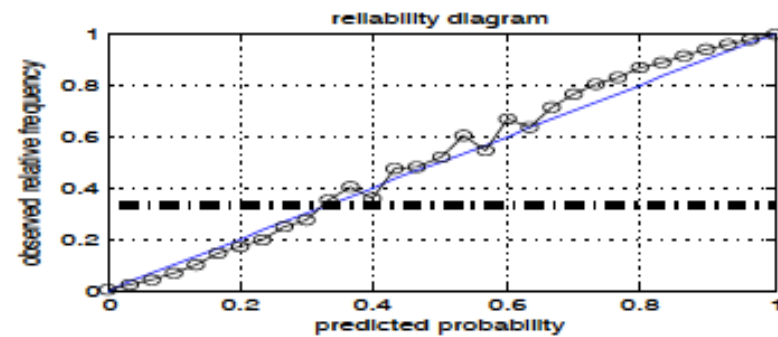
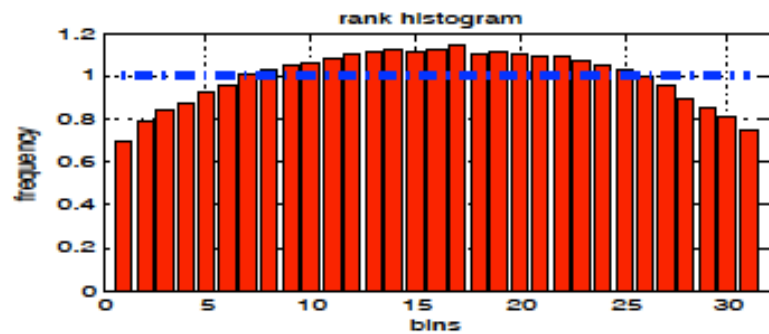
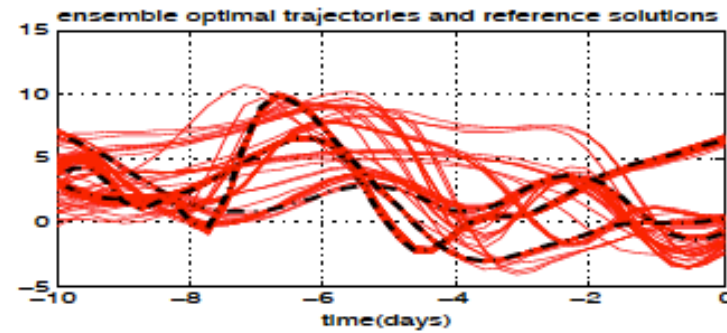
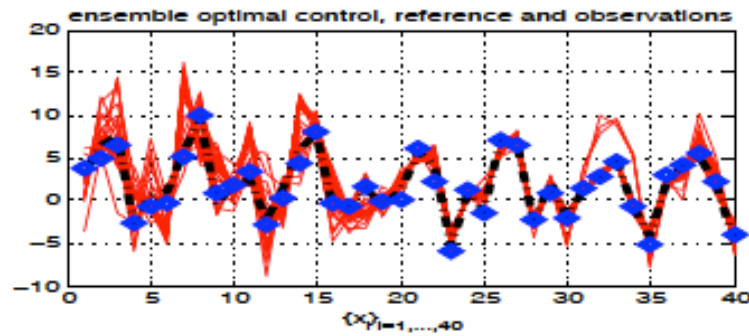
Nonlinear Lorenz'96. 5 days



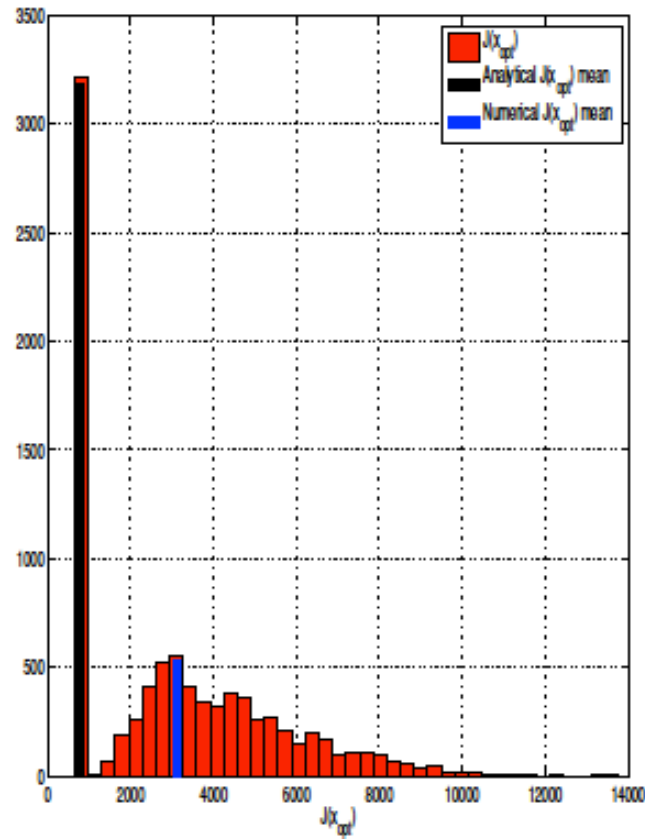
Nonlinear Lorenz'96. 5 days. Histogram of  $J_{\min}$



# EnsVar : the non-linear Lorenz96 model (10 days $\simeq$ 2 TU)

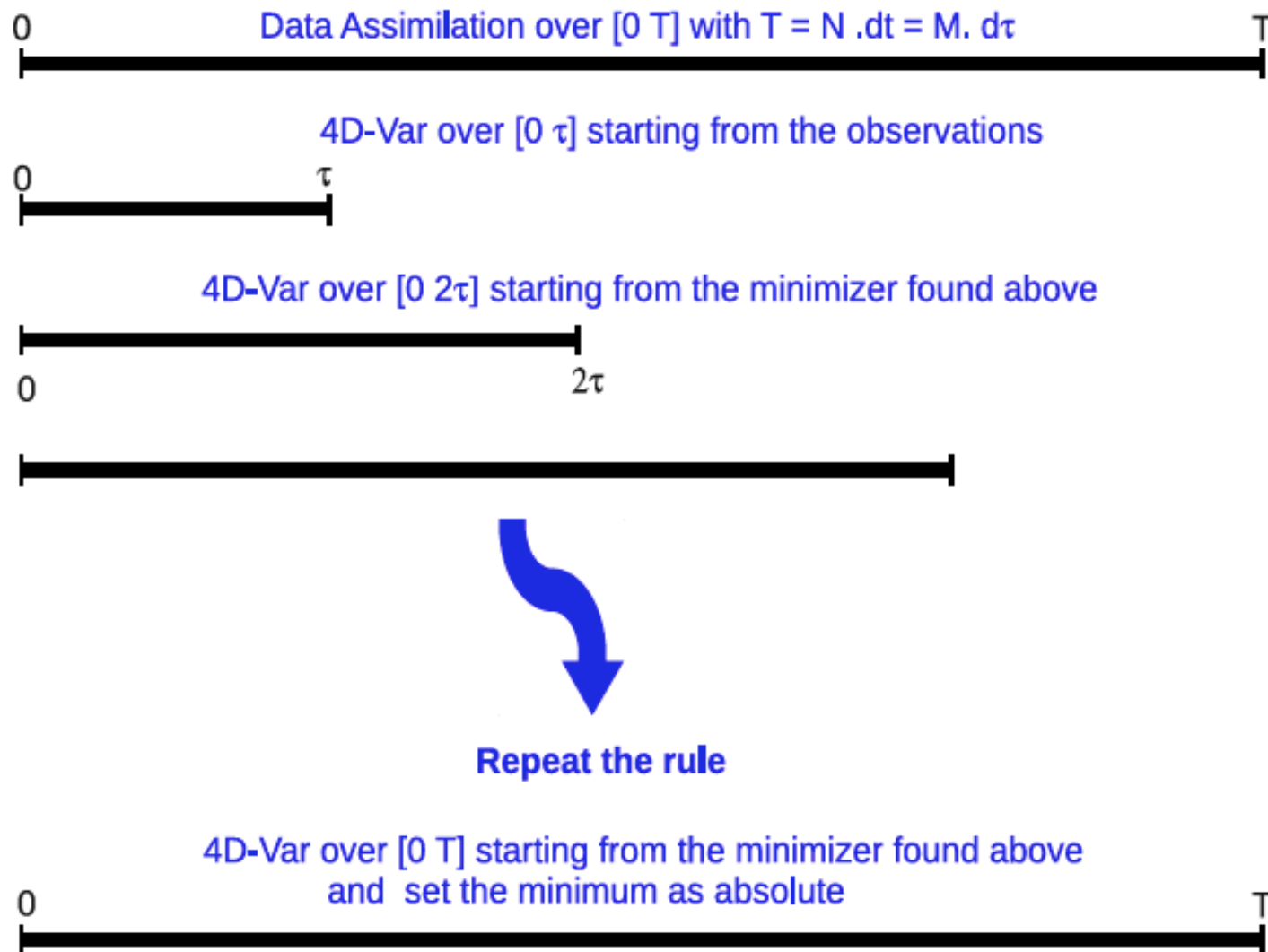


# EnsVar : consistency

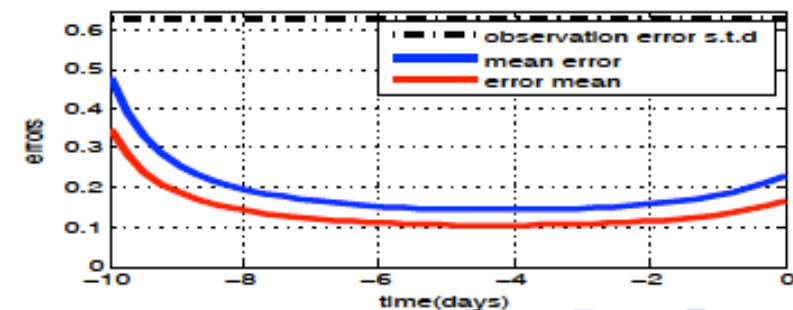
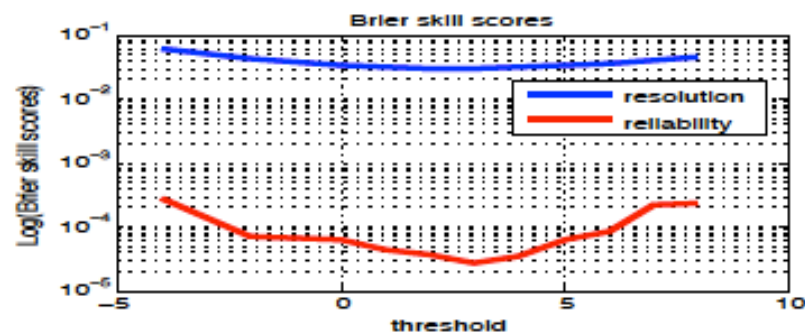
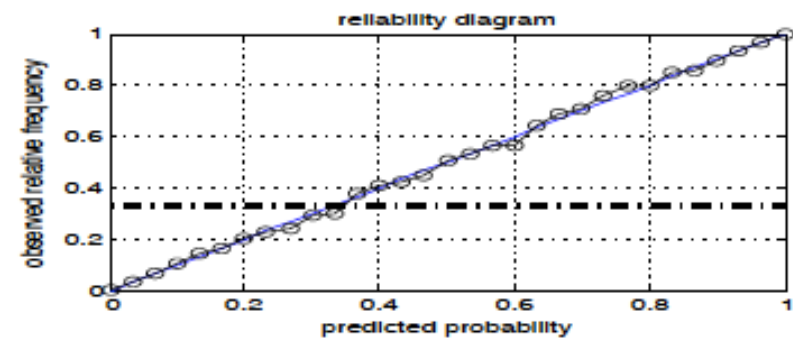
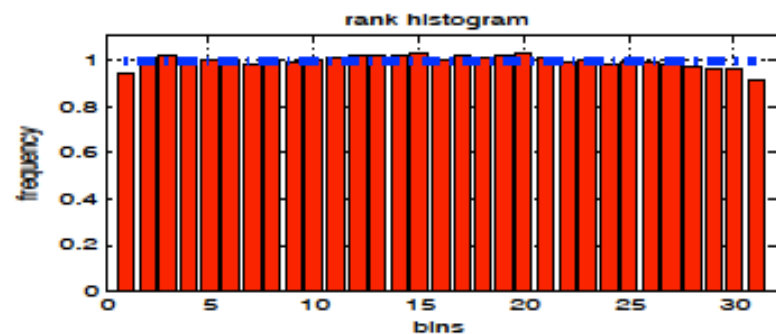
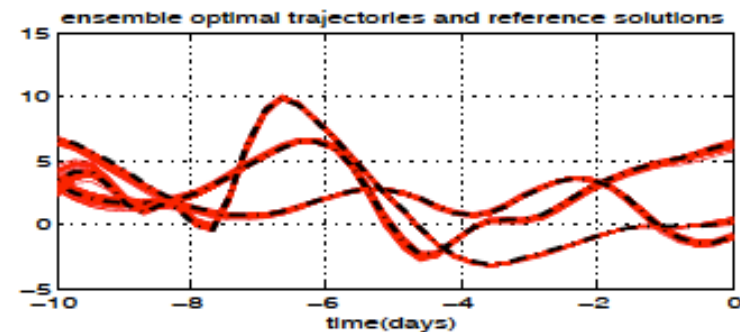
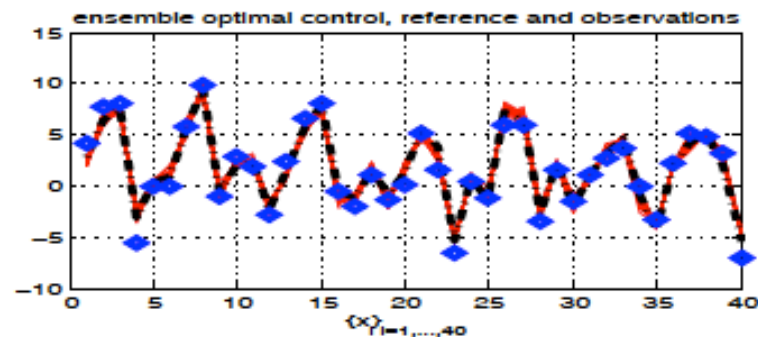


Nonlinear Lorenz'96. 10 days. Histogram of  $J_{min}$

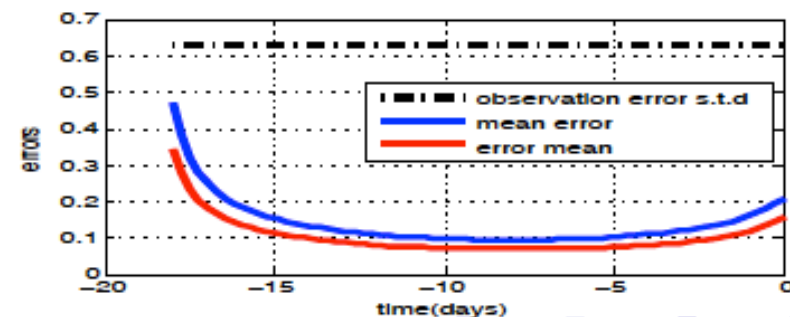
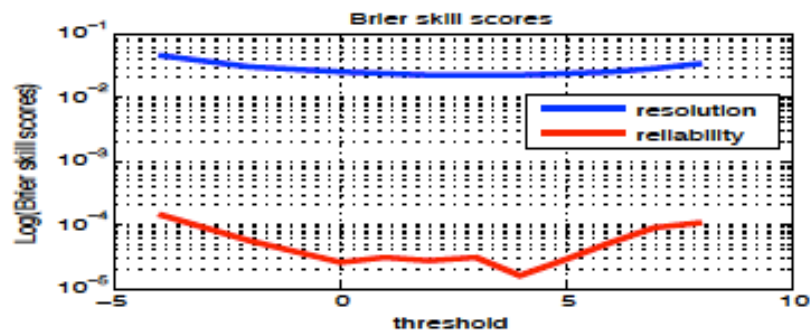
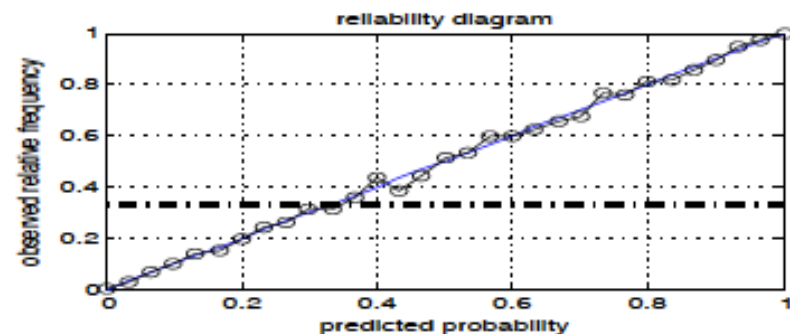
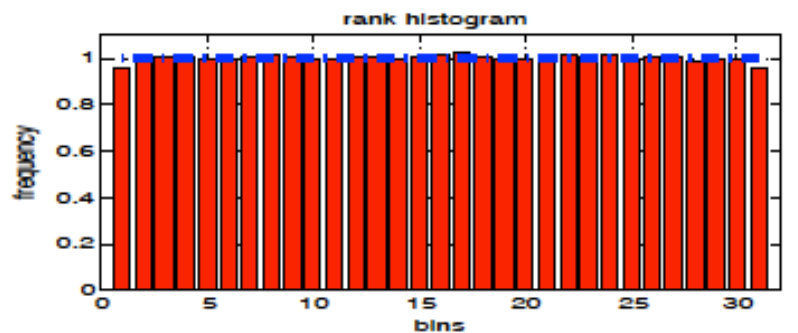
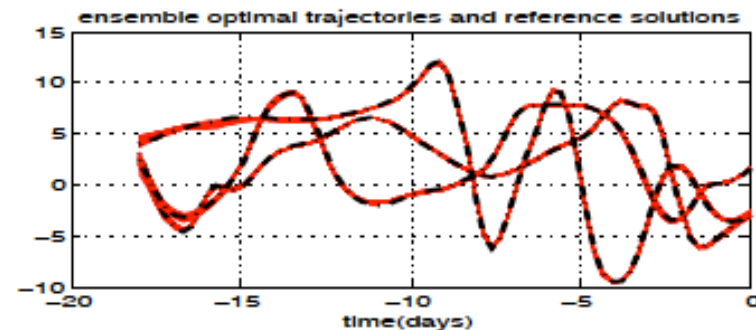
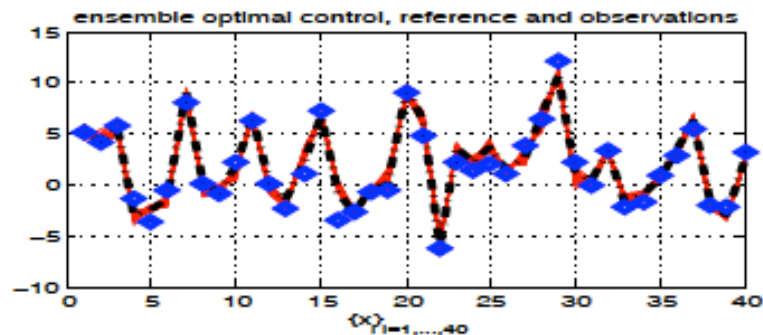
# Quasi-Static Variational Assimilation (QSVA)



# EnsVar : the non-linear Lorenz96 model 10 days with QSVA

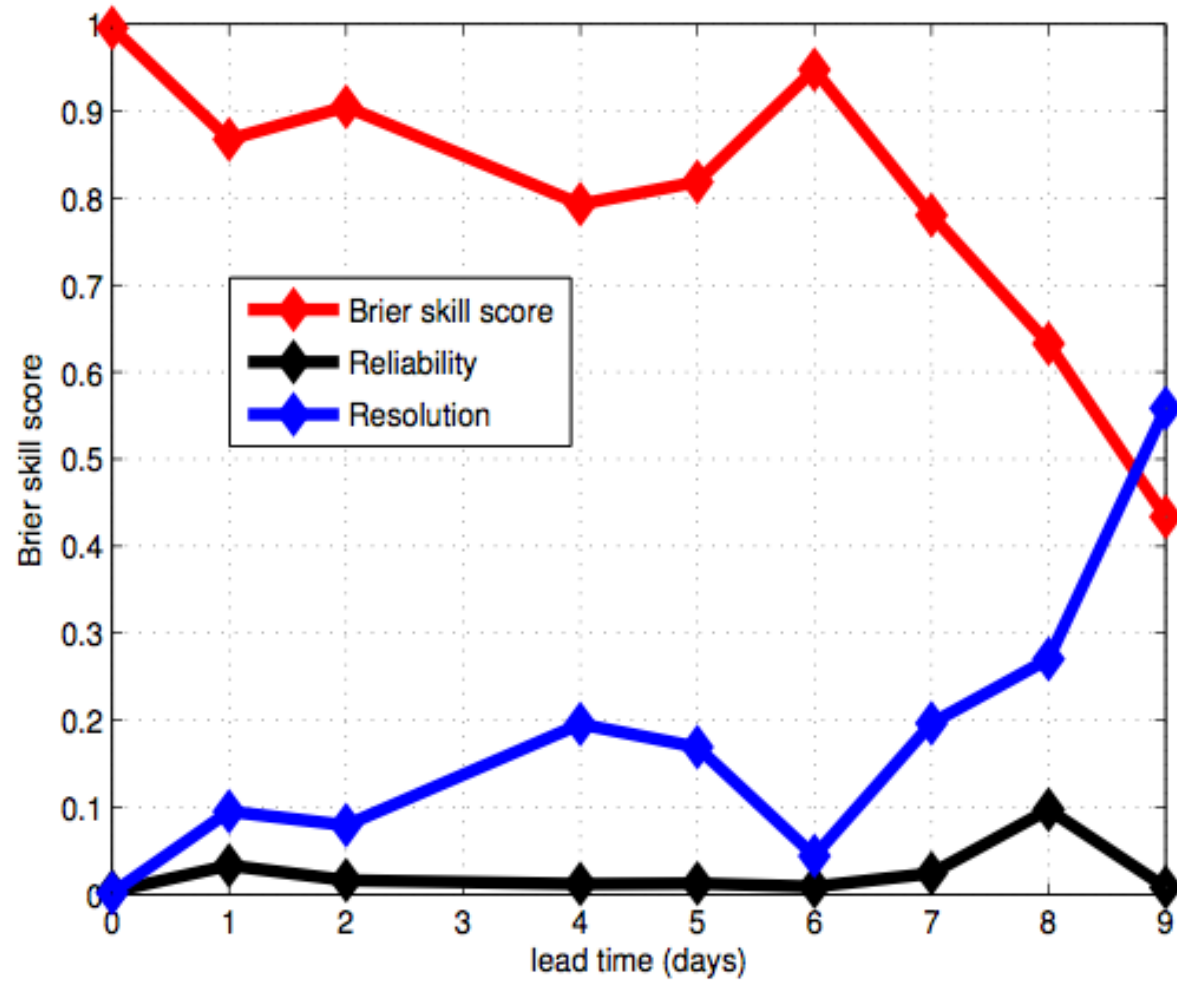


# EnsVar : the non-linear Lorenz96 model 18 days with QSVA



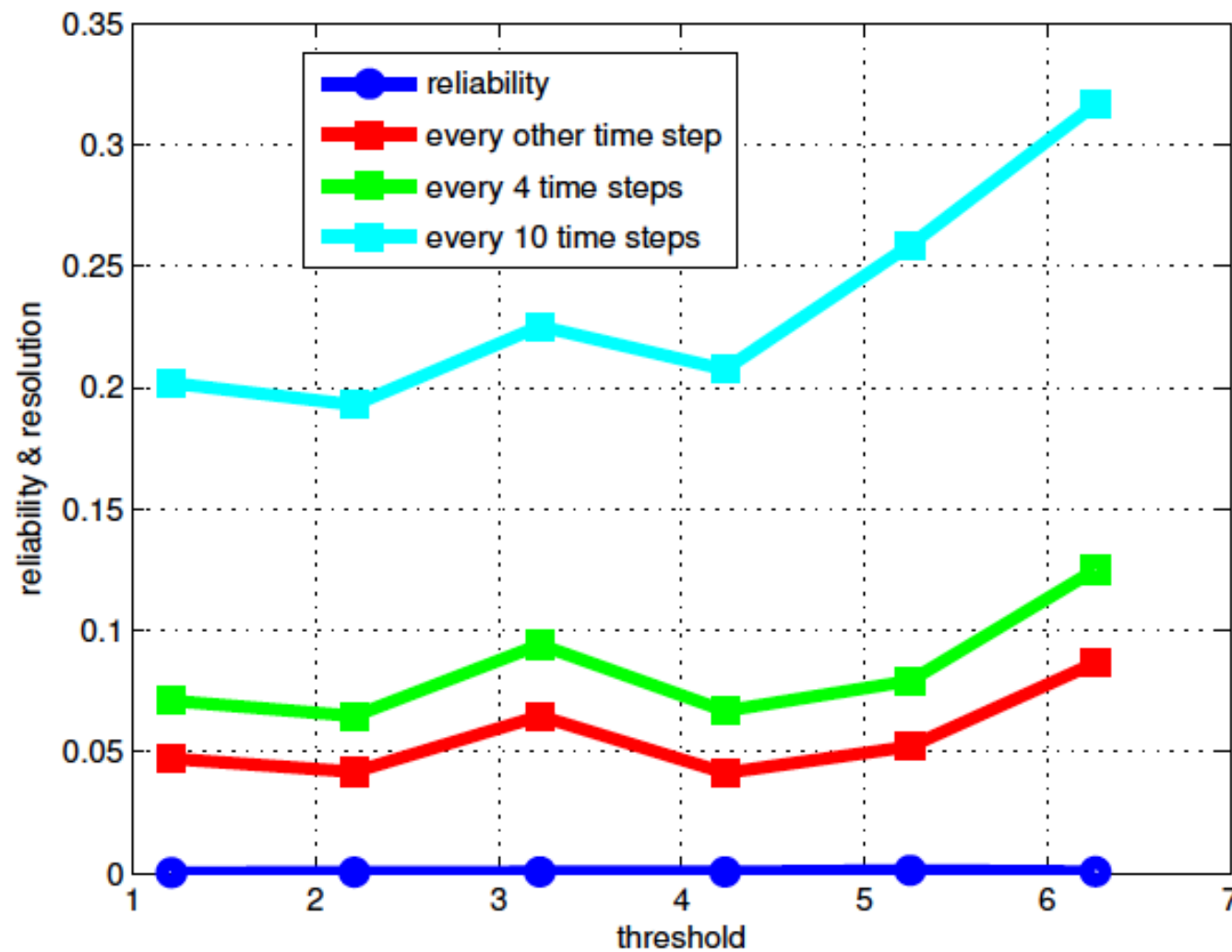
# EnsVar : forecasting

## Brier Skill Scores as a function of the lead time



# EnsVar : observation frequency impact

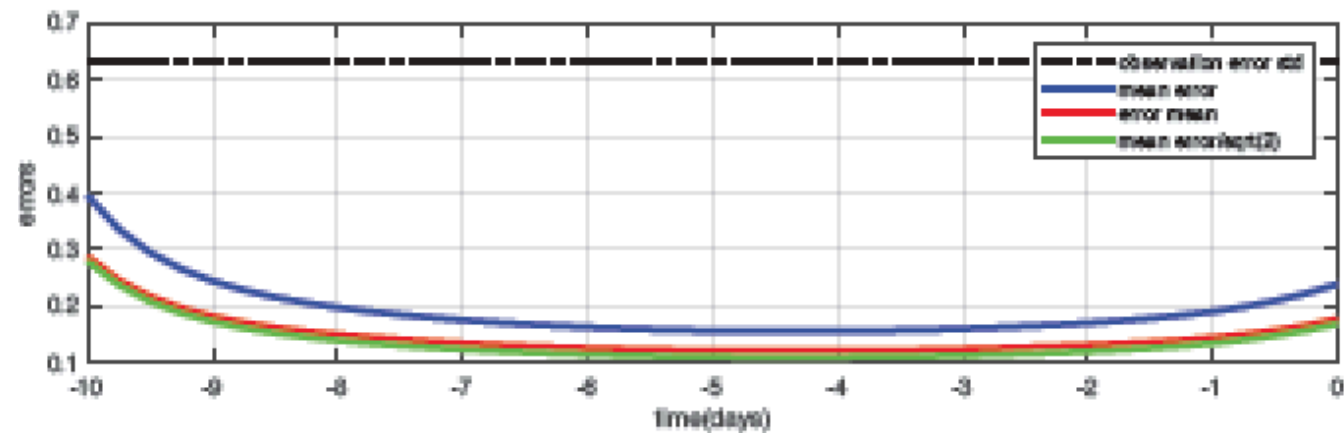
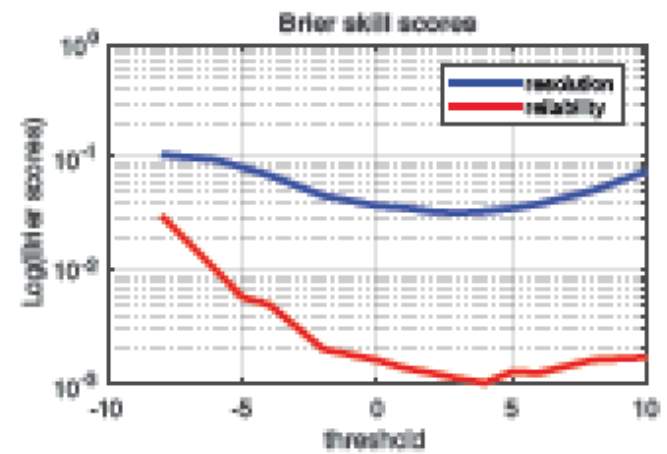
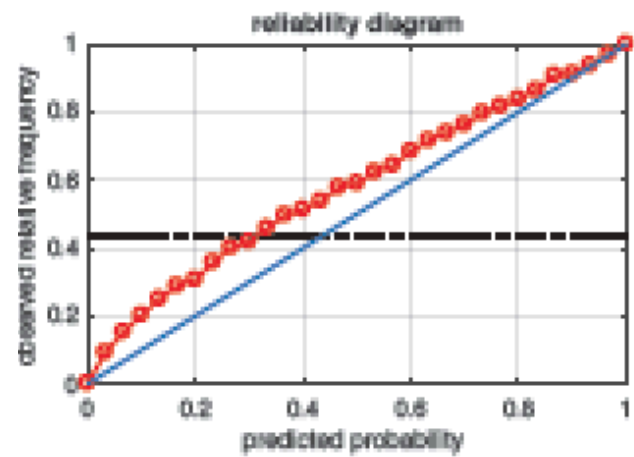
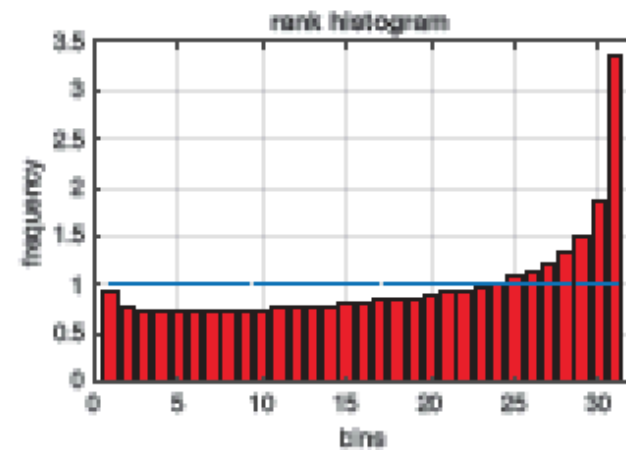
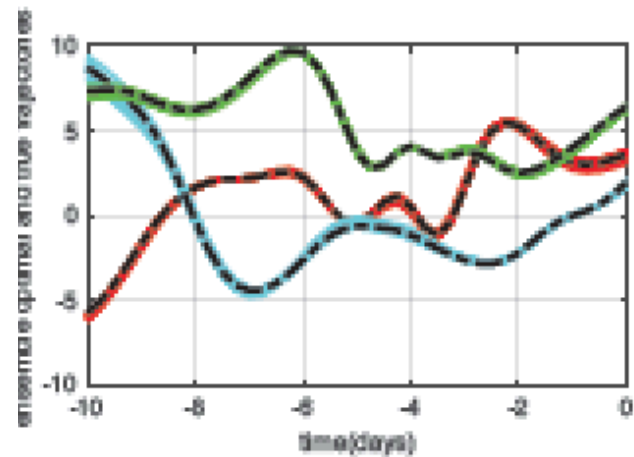
## Impact on the reliability and resolution



- Results are independent of the Gaussian character of the observation errors (trials have been made with various probability distributions)
- Ensembles produced by EnsVar are very close to Gaussian, even in strongly nonlinear cases.

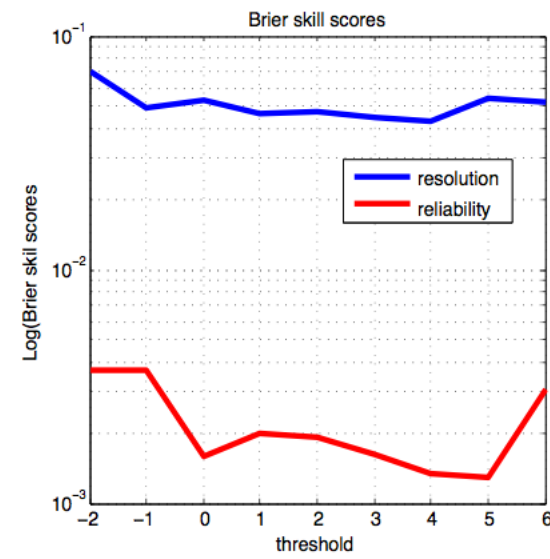
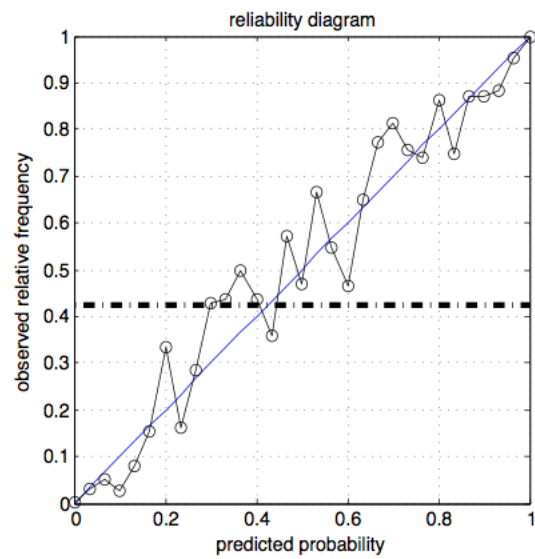
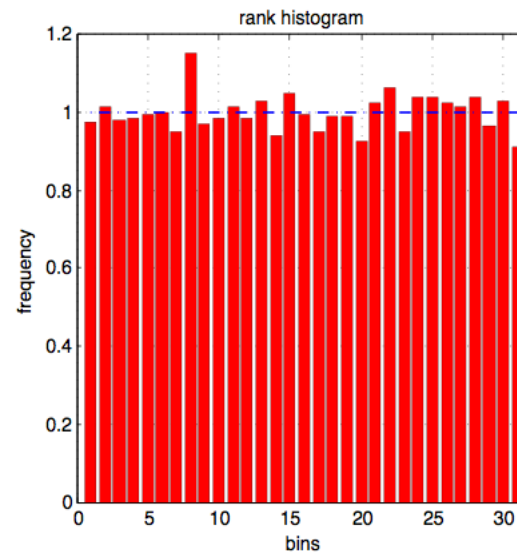
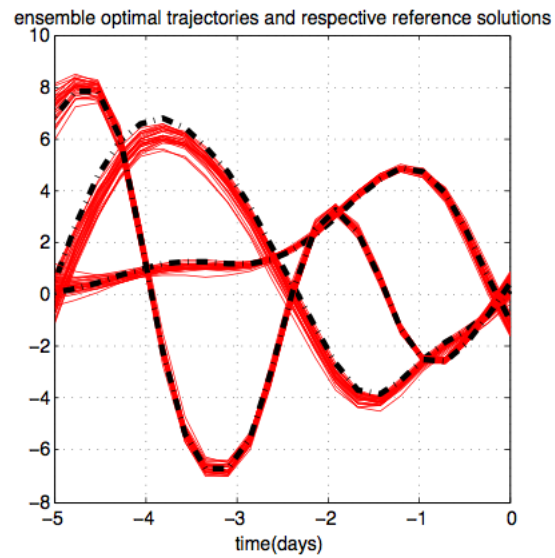


Selection  
made on left  
mode of  
histogram

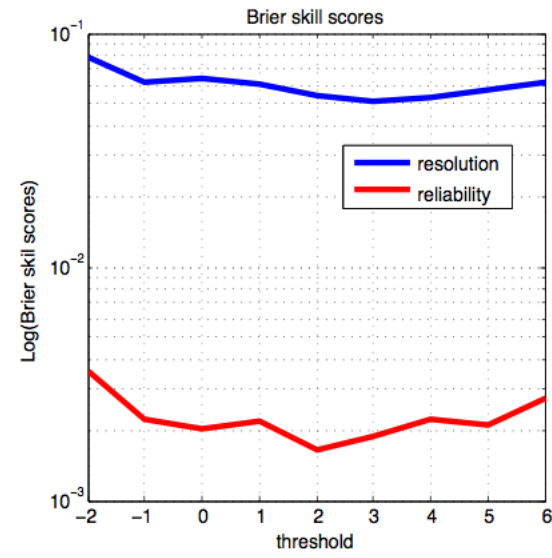
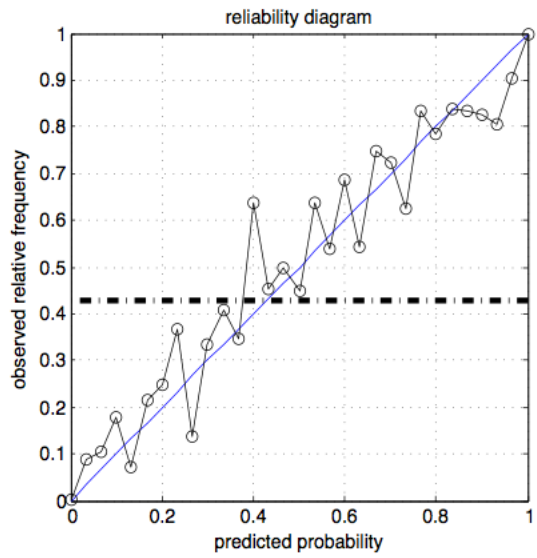
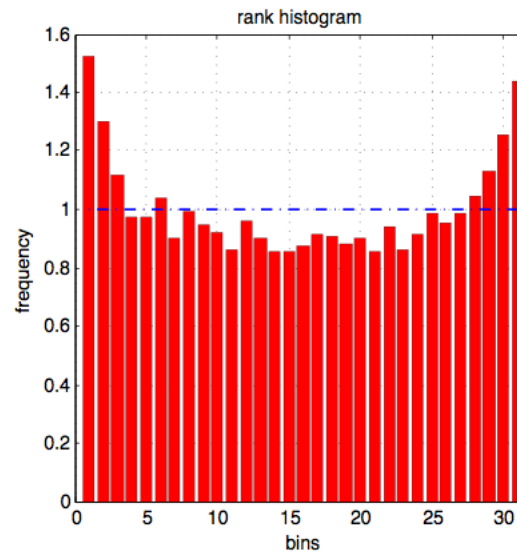
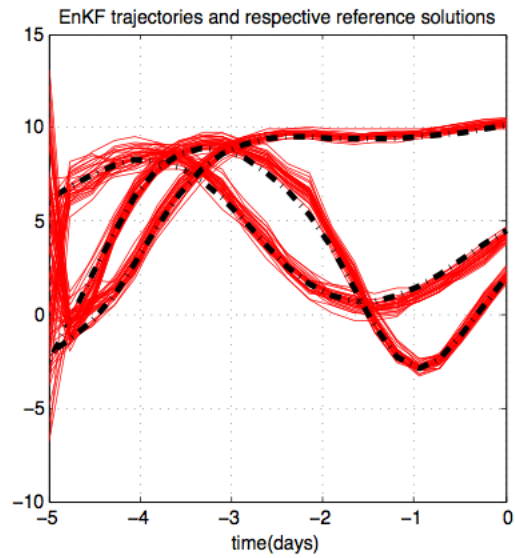


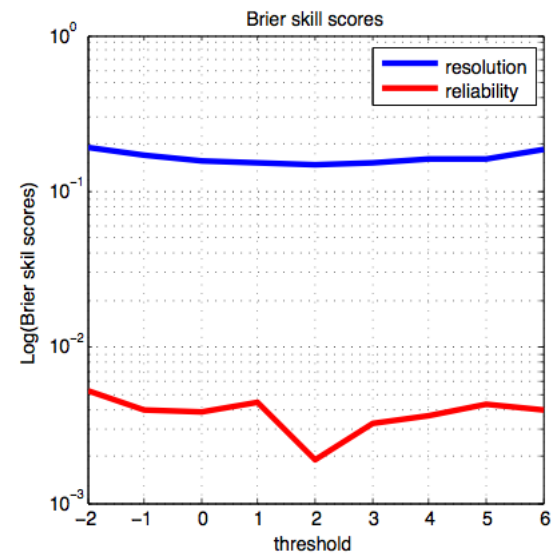
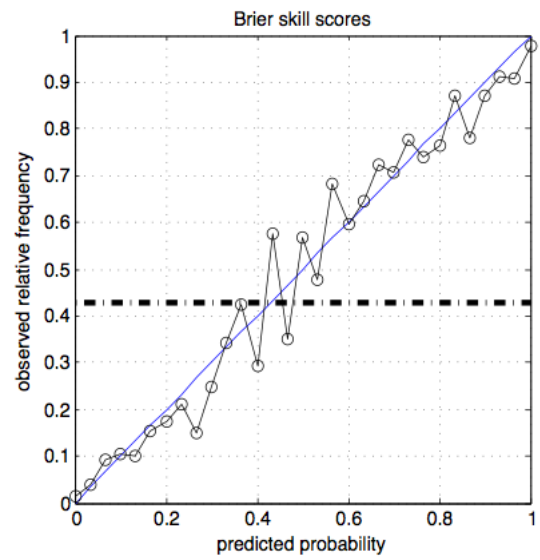
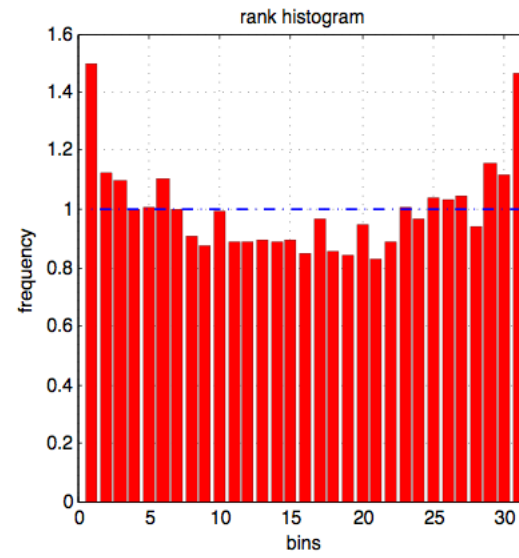
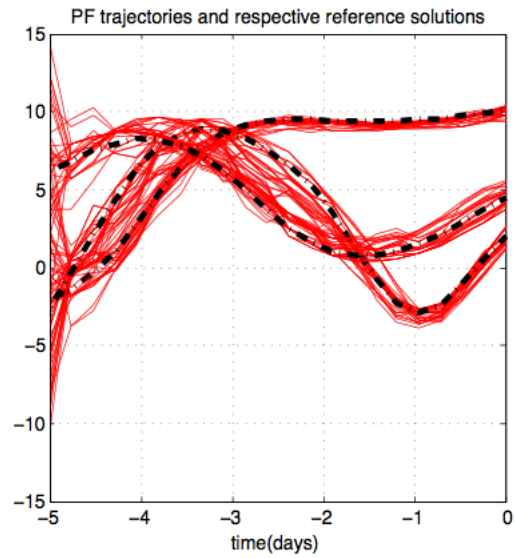
- Comparison *Ensemble Kalman Filter (EnKF)* and *Particle Filters (PF)*

Both of these algorithms being sequential, comparison is fair only at end of assimilation window

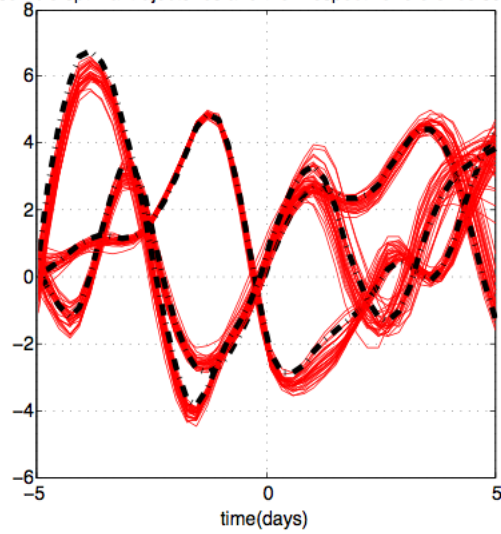


Nonlinear Lorenz'96. 5 days. Diagnostics at end of assimilation window

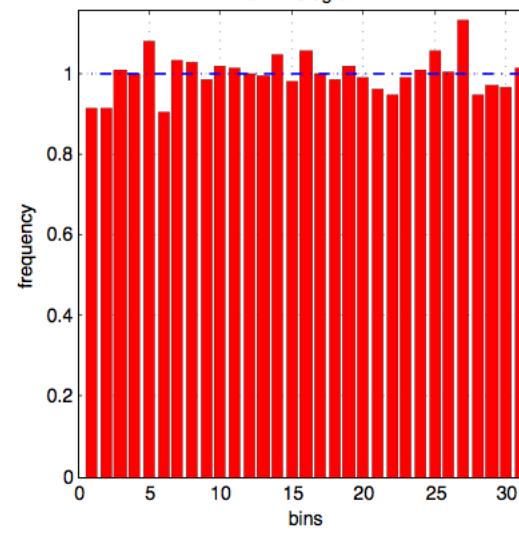




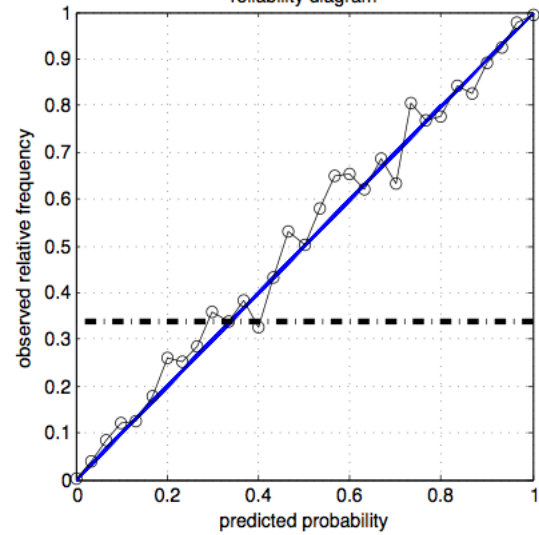
ensemble optimal trajectories and their respective reference solutions



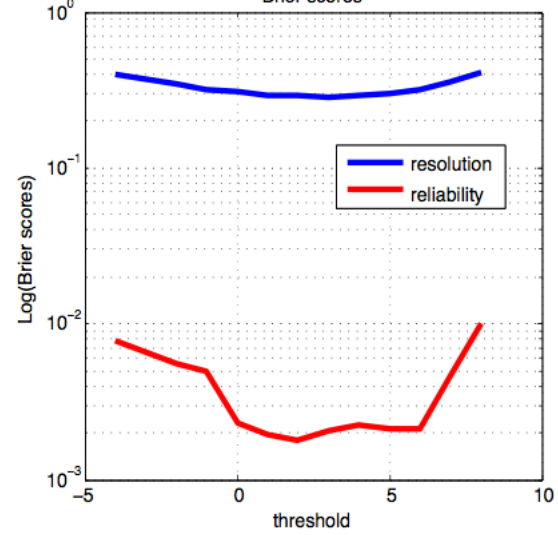
rank histogram



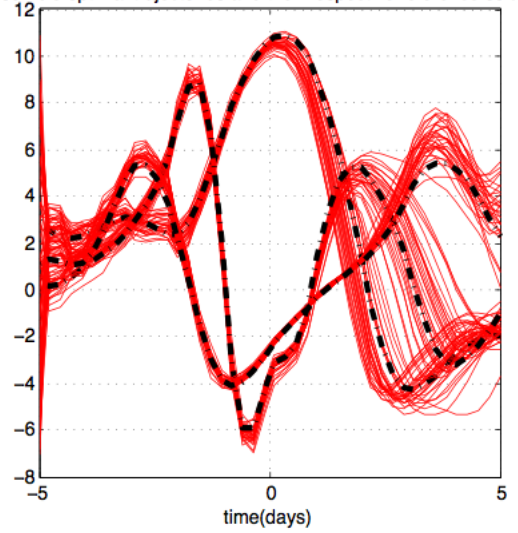
reliability diagram



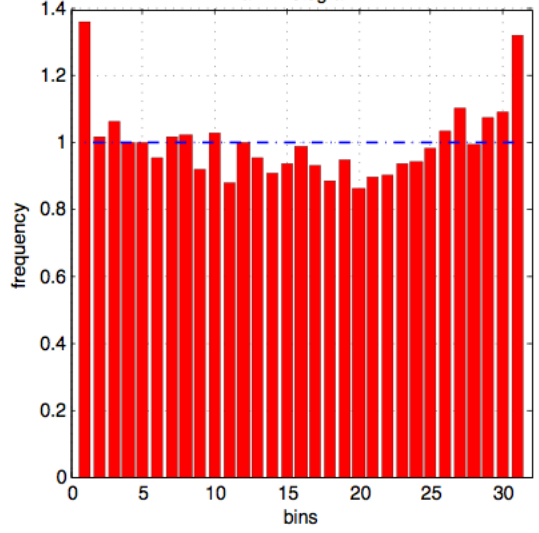
Brier scores



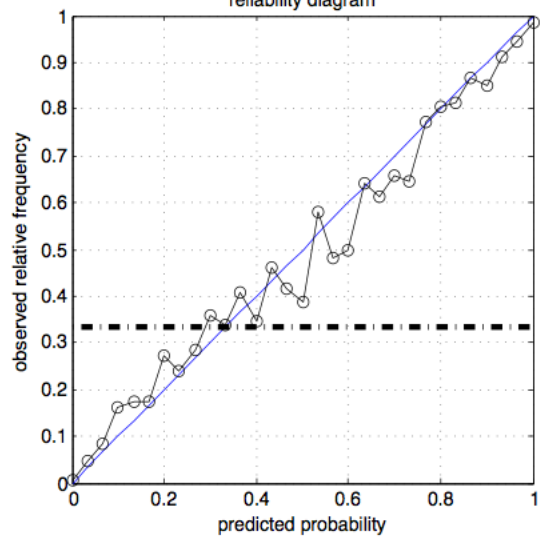
ensemble optimal trajectories and their respective reference solutions



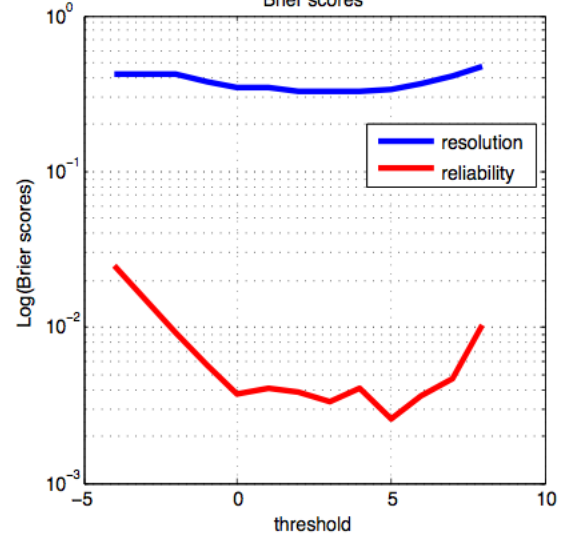
rank histogram

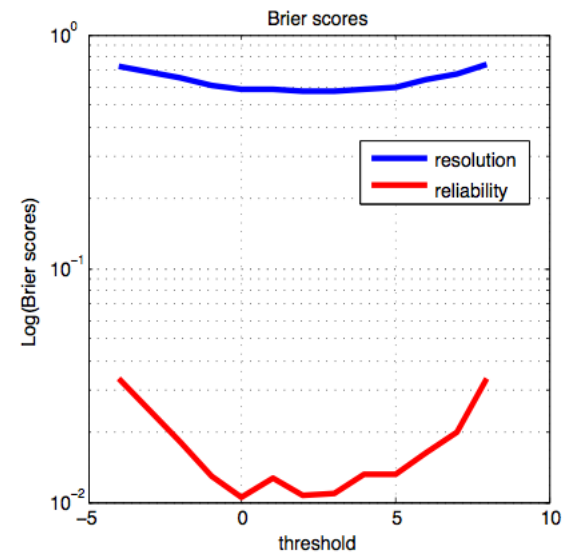
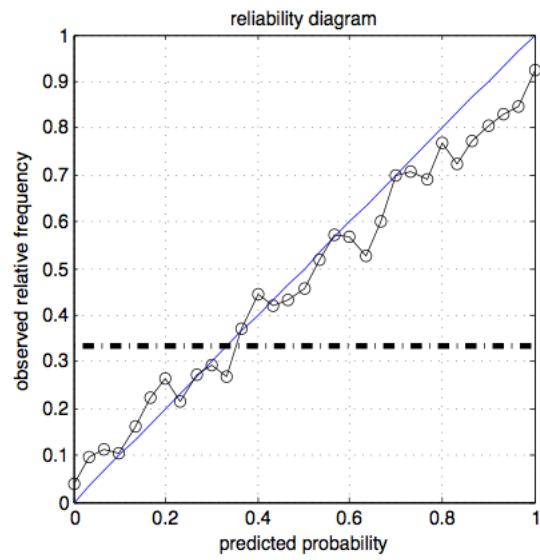
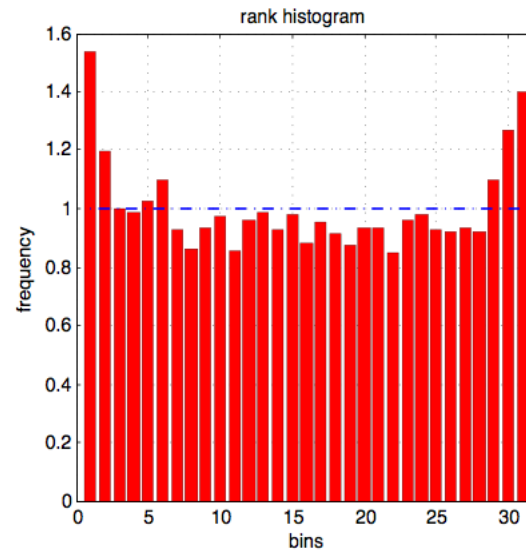
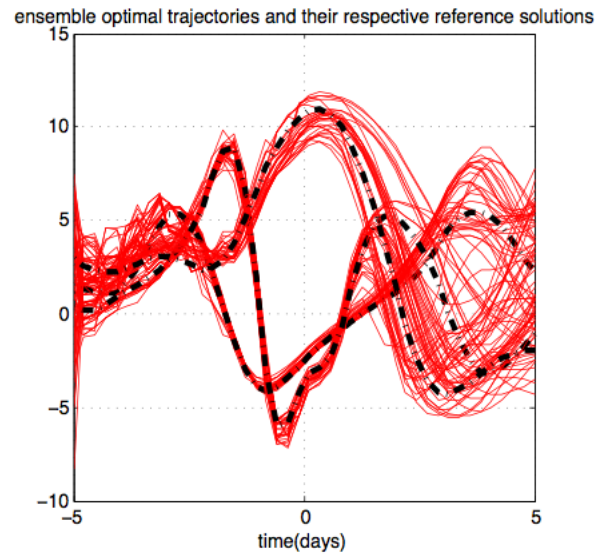


reliability diagram



Brier scores







| <i>method</i> | <i>DA procedure</i> | <i>Assimilation</i> | <i>Forecasting</i> |
|---------------|---------------------|---------------------|--------------------|
| EnsVAR        |                     | 0.2193510           | 1.49403506         |
| EnKF          |                     | 0.2449690           | 1.67176110         |
| PF            |                     | 0.7579790           | 2.62461295         |

RMS errors at the end of 5-day assimilations and 5-day forecasts

# Weak constraint EnsVar

- define the objective function.

$$\mathfrak{J}(x, \eta_1, \eta_2, \dots, \eta_{N-1}, \eta_N) = \frac{1}{2} \{ (x - x_b)^T B^{-1} (x - x_b) \} +$$

$$\frac{1}{2} \sum_{i=0}^N \{ (y_i - H_i(x_i))^T R_i^{-1} (y_i - H_i(x_i)) \} + \frac{1}{2} \sum_{i=1}^N \eta_i^T Q_i^{-1} \eta_i$$

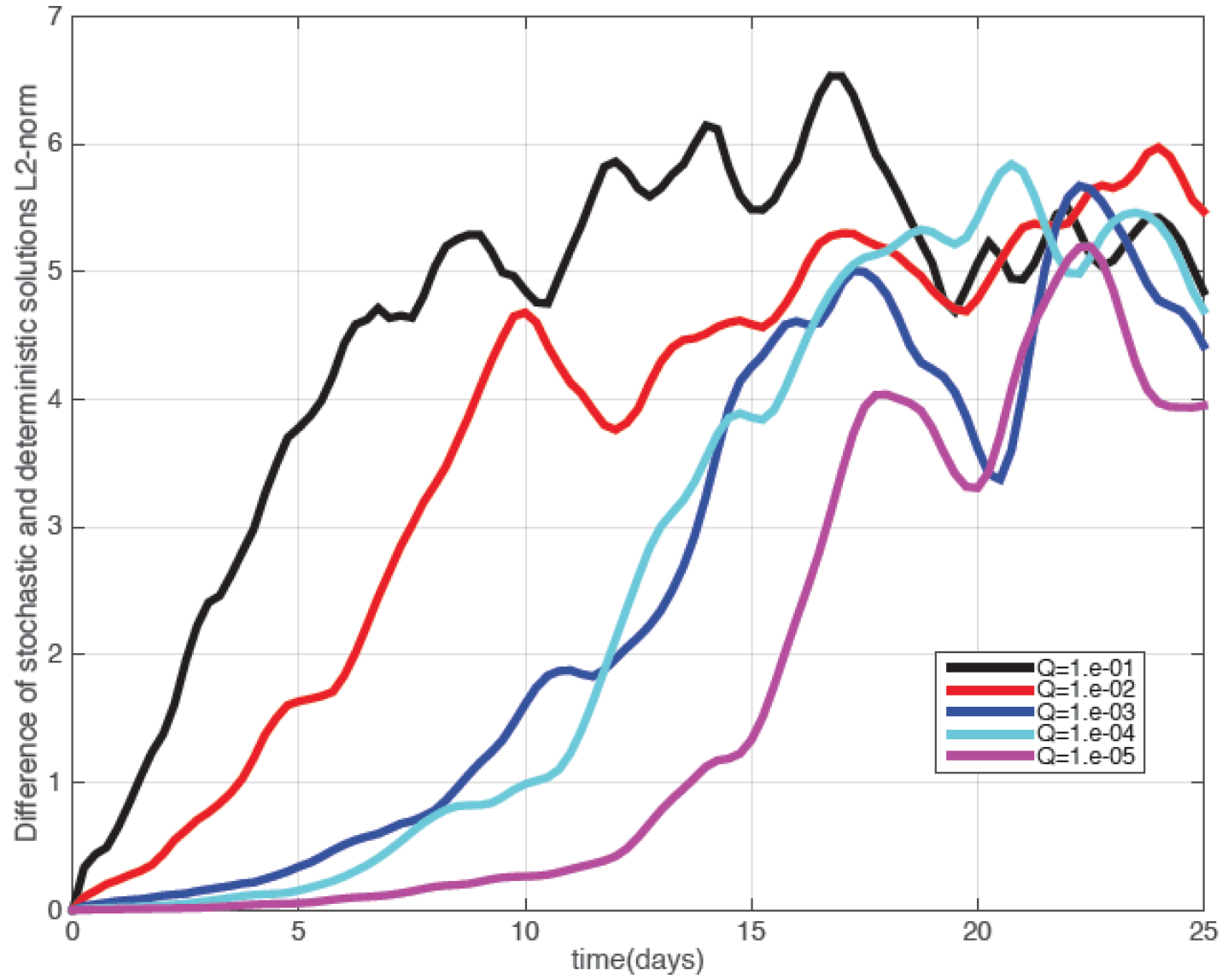
- 1 B background error covariance matrix and R observation error covariance matrix.
  - 2 Q model error covariance matrix.
  - 3  $H : \mathbb{R}^{state} \rightarrow \mathbb{R}^{obs}$  observation operator.
  - 4  $x_b$  background state vector and  $y_i$  observation vector at time  $t = t_i$ .
  - 5  $\eta_i$  model error vector at  $t = t_i$  with  $x(t_i) = \mathfrak{M}_{t_i \leftarrow t_{i-1}}(x(t_{i-1})) + \eta_i$
- find the optimal control variable  $(x_0^{opt}, \eta_1^{opt}, \eta_2^{opt}, \dots, \eta_N^{opt})$  and the optimal trajectory  $x^{opt}$ .

$$(x_0^{opt}, \eta_1^{opt}, \eta_2^{opt}, \dots, \eta_N^{opt}) = \min_{x, \eta_1, \eta_2, \dots, \eta_N \in \mathfrak{A}} \mathfrak{J}(x, \eta_1, \eta_2, \dots, \eta_N)$$

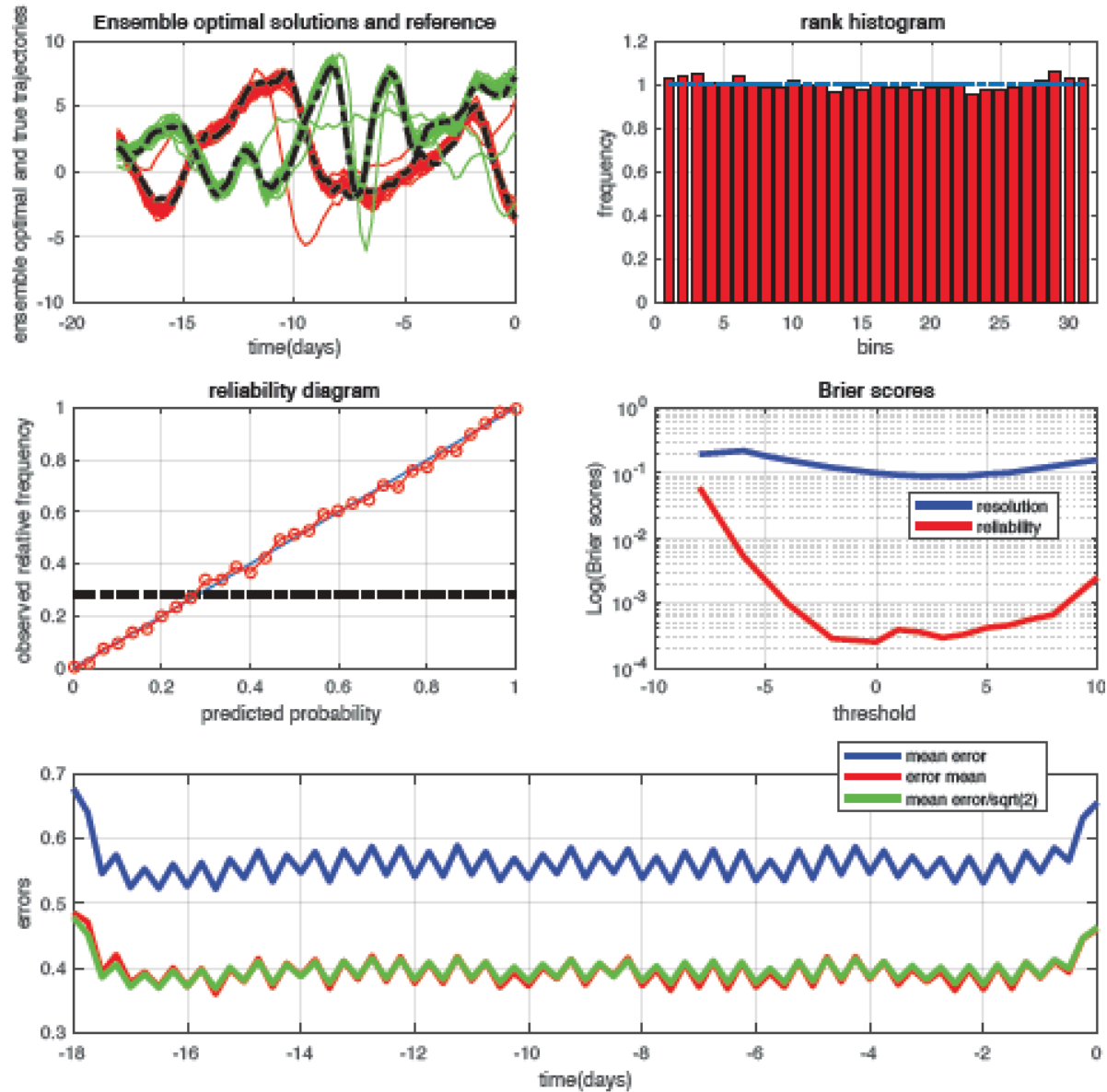
$$x_i^{opt} = \mathfrak{M}_{t_i \leftarrow t_{i-1}}(\mathfrak{M}_{t_{i-1} \leftarrow t_{i-2}}(\dots(\mathfrak{M}_{t_2 \leftarrow t_1}(\mathfrak{M}_{t_1 \leftarrow t_0}(x_0^{opt}) + \eta_1^{opt}) + \eta_2^{opt}) \dots + \eta_{i-1}^{opt})) + \eta_i^{opt}$$



### Divergence between stochastic and deterministic solutions

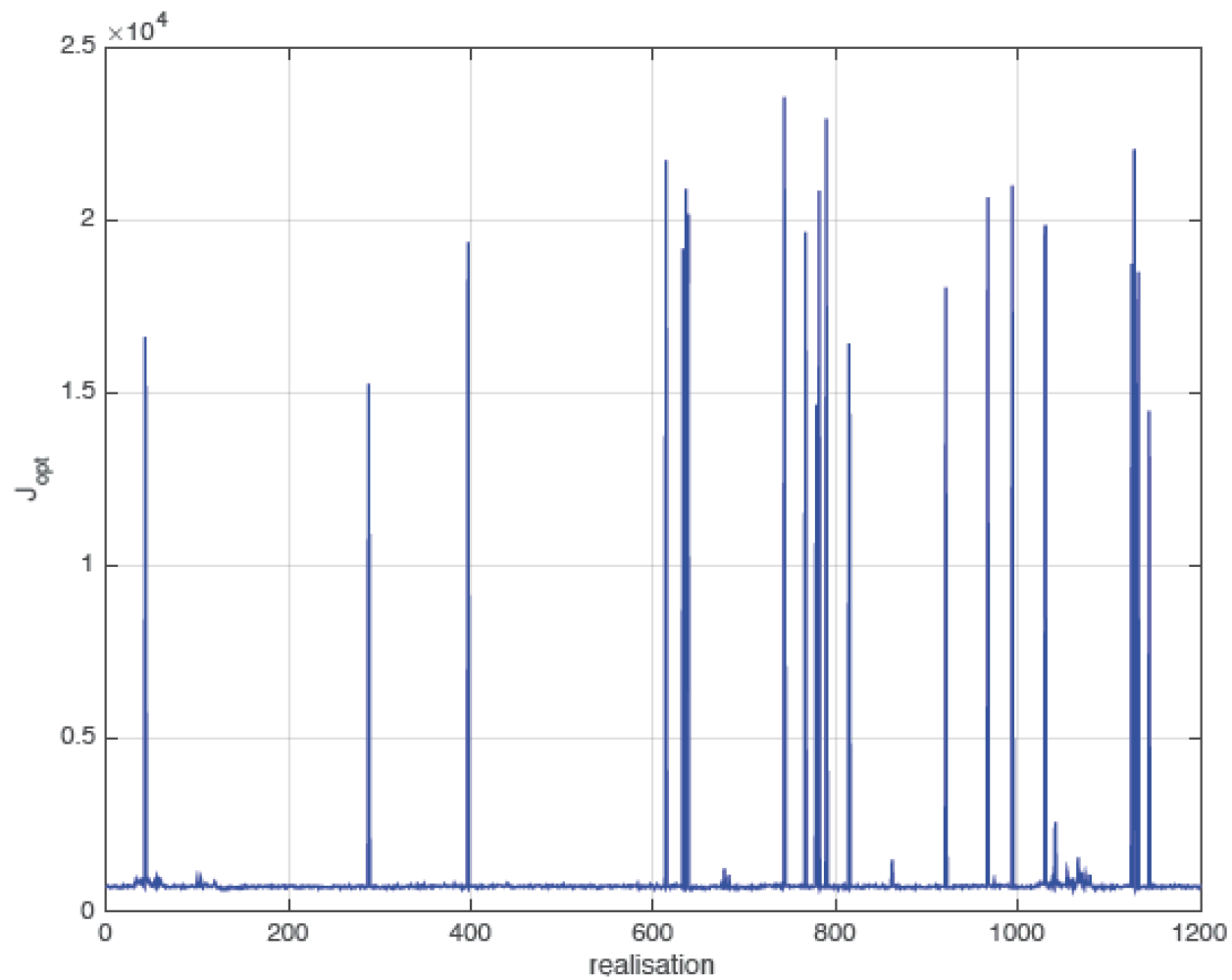


Weak constraint EnsVar 18 days assimilation, Q=0.1 and 1200 realisations



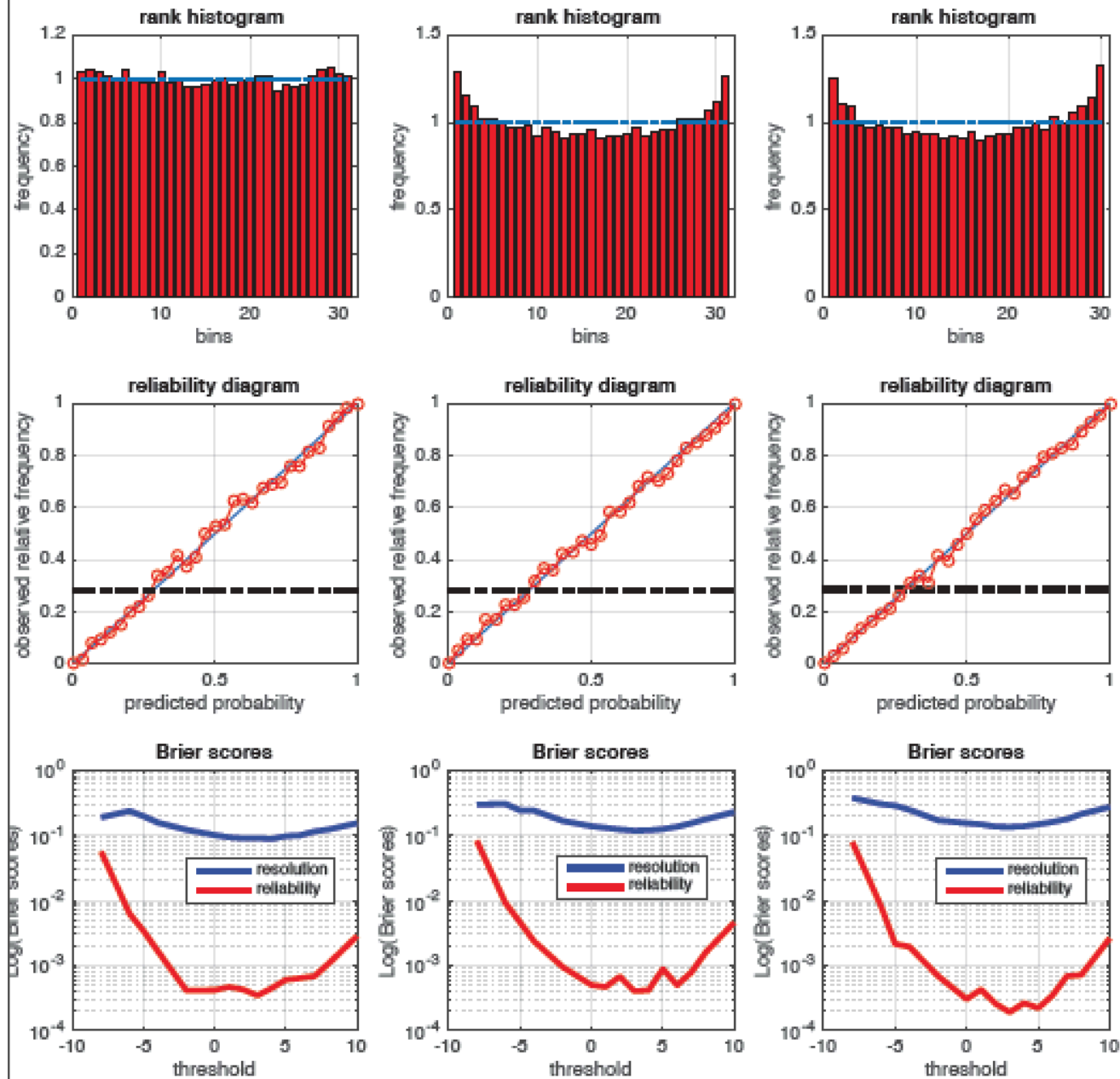
Q=0.1 R=0.4

Weak constraint nonlinear case, 18 days (no QSVA) and 1200 realisations



EnsVar, EnKF and PF diagnostics, 18 days assimilation, Q=0.1 and 1200 realisations

RCRV = (-0.0053, 1.0168)    RCRV = (8.83e-04, 1.1359)    RCRV = (-0.0123, 1.1065)



## Summary

- Under non-linearity and non-Gaussianity the EnsVar is a reliable and consistent ensemble estimator (provided the QSVA is used for long DA windows) .
- EnsVar is at least as good an estimator as EnKF and PF.
- Similar results have been obtained for the Kuramoto-Sivashinsky model.

Ensembles obtained are Gaussian, even if errors in data are not

Produces Monte-Carlo sample of (probably not) bayesian pdf

# EnsVar : Pros and cons

## Pros

- Easy to implement when having a 4D-Var code
- Highly parallelizable
- No problems with algorithm stability (i.e. no ensemble collapse, no need for localization and inflation, no need for weight resampling)
- Propagates information in both ways and takes into account temporally correlated errors

## Cons

- Costly (Nens 4D-Var assimilations).
- Empirical.
- Cycling of the process (**work in progress**).



## **And now ?**

- Implementation on physically more realistic models (QG, Shallow water, ...)
- Comparison with other ensemble algorithms (IEnKS)
- Cycling and/or overlap
- Minimisation in unstable space (AUS, Trevisan *et al.*)
- ...

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

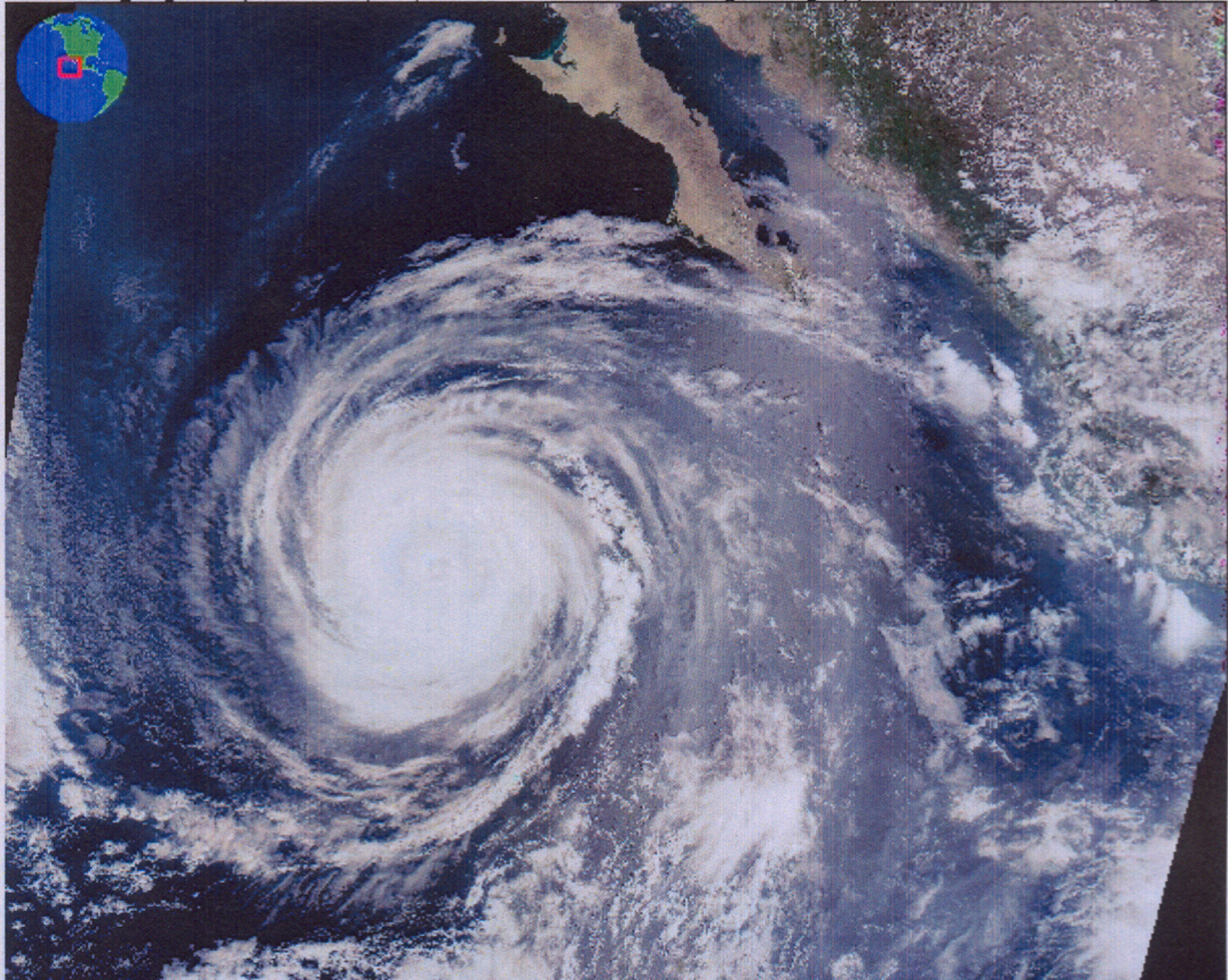
- Oceanography
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciology
- Magnetism (both planetary and stellar)
- Plate tectonics
- Planetary atmospheres (Mars, ...)
- Reassimilation of past observations (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)
- ...

It has now become a major tool of numerical environmental science

A few of the (many) remaining problems :

- Observability (what to observe in order to know what we want to know ? 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



## Cours à venir

~~Jeudi 6 avril~~

~~Jeudi 13 avril~~

~~Jeudi 20 avril~~

~~Jeudi 11 mai~~

**Lundi 29 mai**

~~Jeudi 1 juin~~

~~Jeudi 15 juin~~

~~Jeudi 22 juin~~

Contrôle le mardi 12 juillet, de 14h00 à 16h00, Salle de la Serre, 5ième étage, Département de Géosciences, École Normale Supérieure, 24, rue Lhomond, Paris 5

The End