École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2022-2023

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 7

23 Mai 2023

Last course (May 6)

- Weak-constraint Variational Assimilation.
  Dual Algorithm for Variational Assimilation
- Complements on Variational Assimilation.
    - Mahalanobis Norm
    - How to write (and validate) an adjoint code
    - Value of objective function at minimum. χ^2 test

- Compared qualities of Sequential and Variational Assimilation

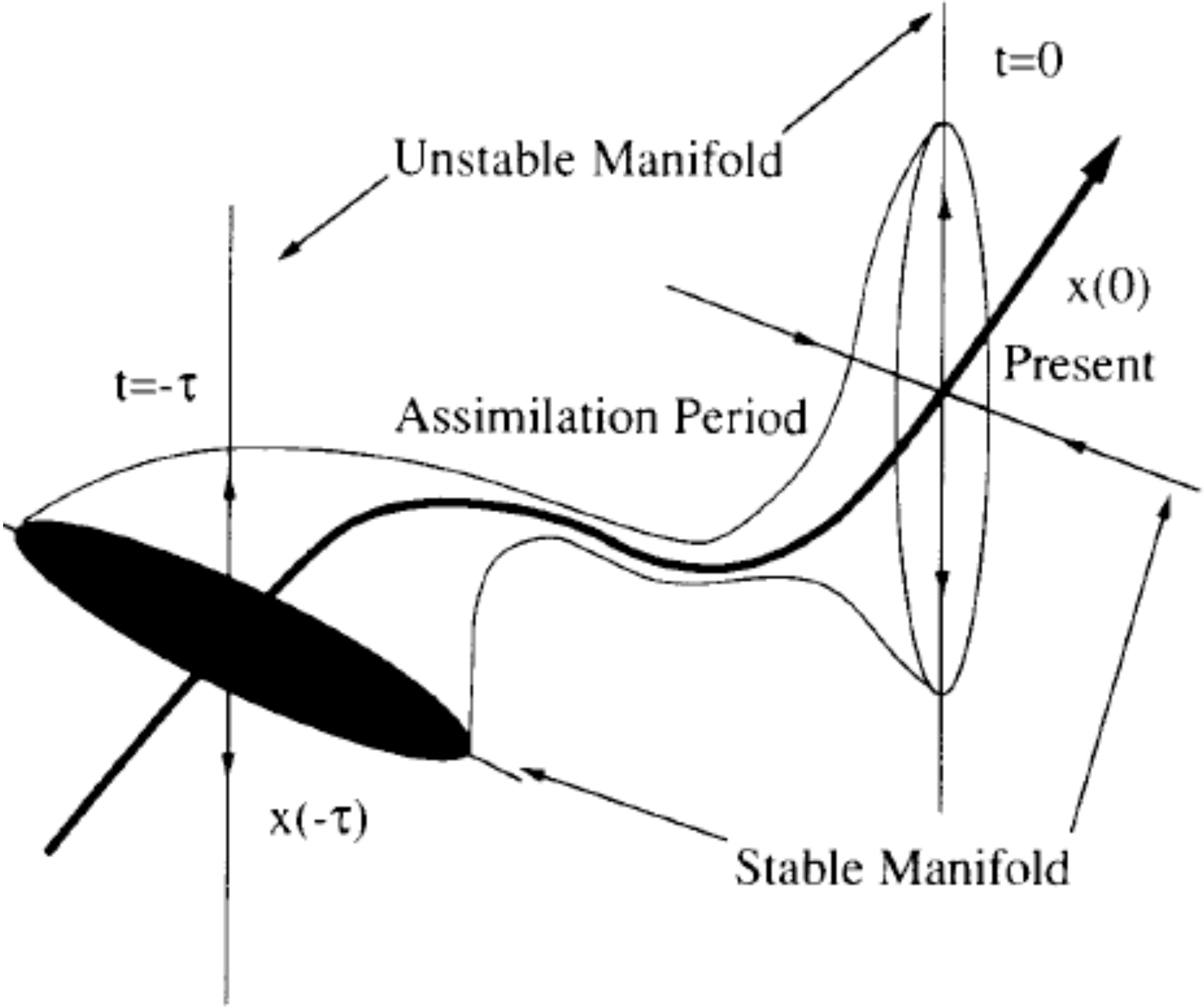
- Assimilation and (In)stability. Quasi-Static Variational Assimilation

This course

- Assimilation dans l'espace instable

- Filtres particulaires

- Assimilation Variationnelle d'Ensemble



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

Gurumoorthy *et al*. (2017*a*, 2017*b*) have shown that in the linear perfect model case, the error covariance matrix of the Kalman filter converges to the neutral-unstable subspace of the system (space spanned by the non-negative Lyapunov exponents of the system)

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.

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 periodic conditions in j, and value $F = 8$, which gives rise to chaos.

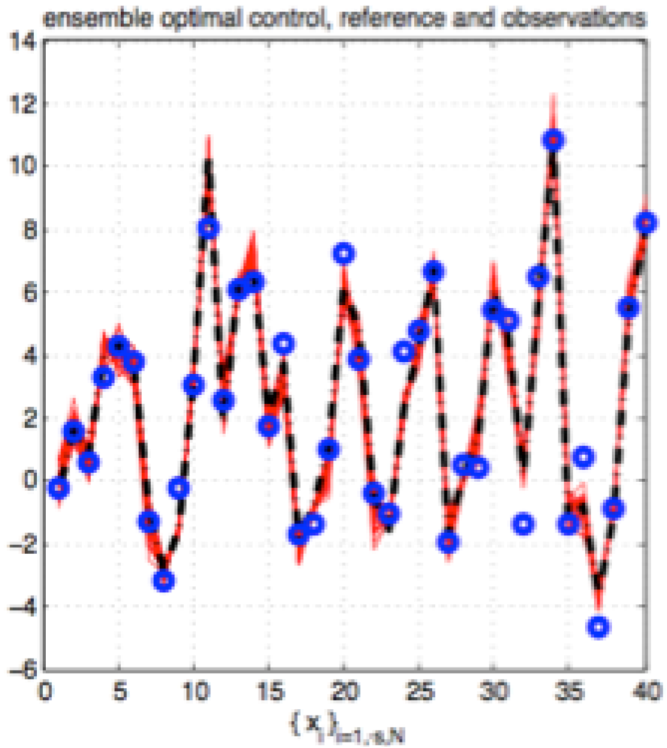
Three values of I have been used, namely $I = 40, 60, 80$, which correspond to respectively $N^+ = 13, 19$ and 26 positive Lyapunov exponents.

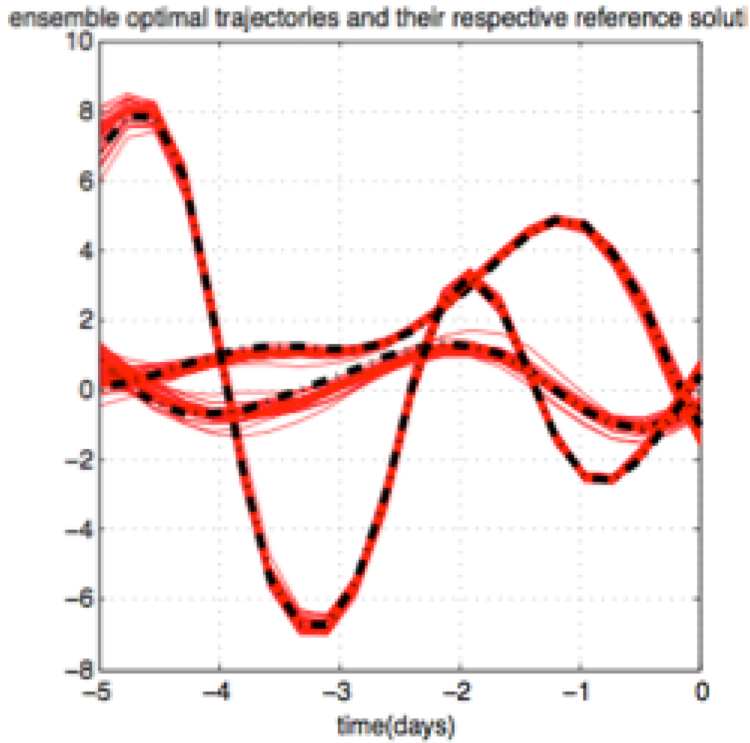
In all three cases, the largest Lyapunov exponent corresponds to a doubling time of about 2 days (with 1 'day' = 1/5 model time unit).

Identical twin experiments (perfect model)

System produces wavelike chaotic motions, with properties similar to those of midlatitude atmospheric waves

- generally westward phase velocity
- typical predictability time : 5 ‘days’
- in addition, quadratic terms conserve ‘energy’





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).

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.

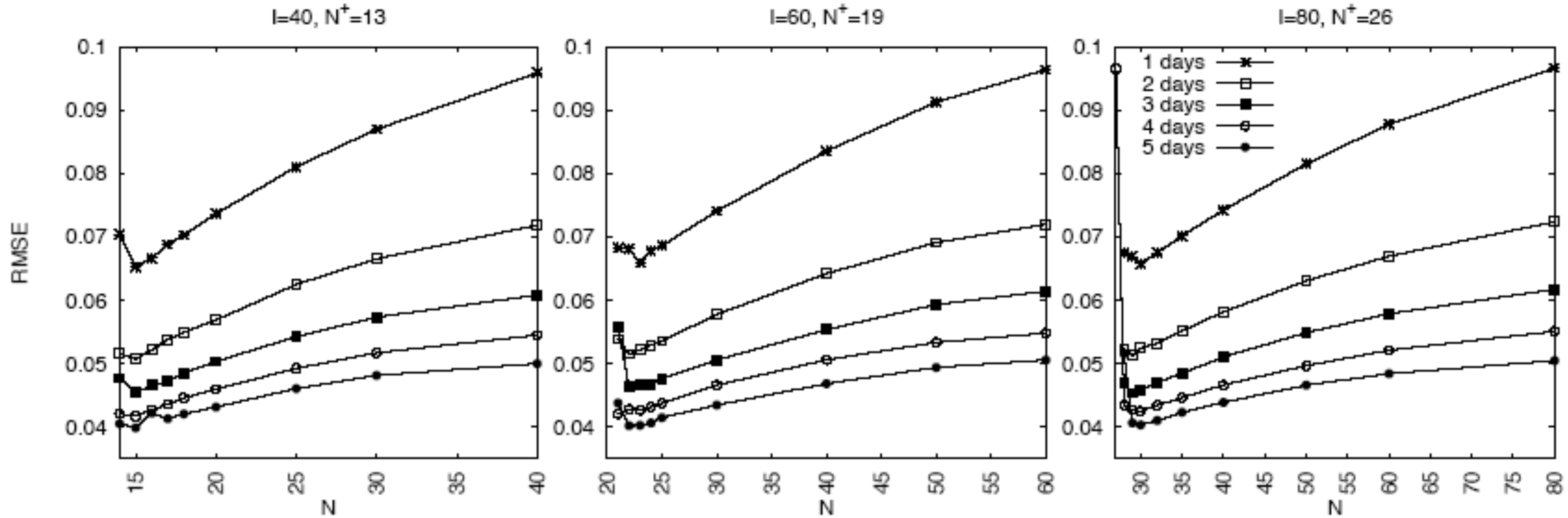


Figure 1. Time average RMS analysis error at $t = \tau$ as a function of the subspace dimension N for three model configurations: I=40, 60, 80. Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is $\sigma_o = 0.2$.

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.

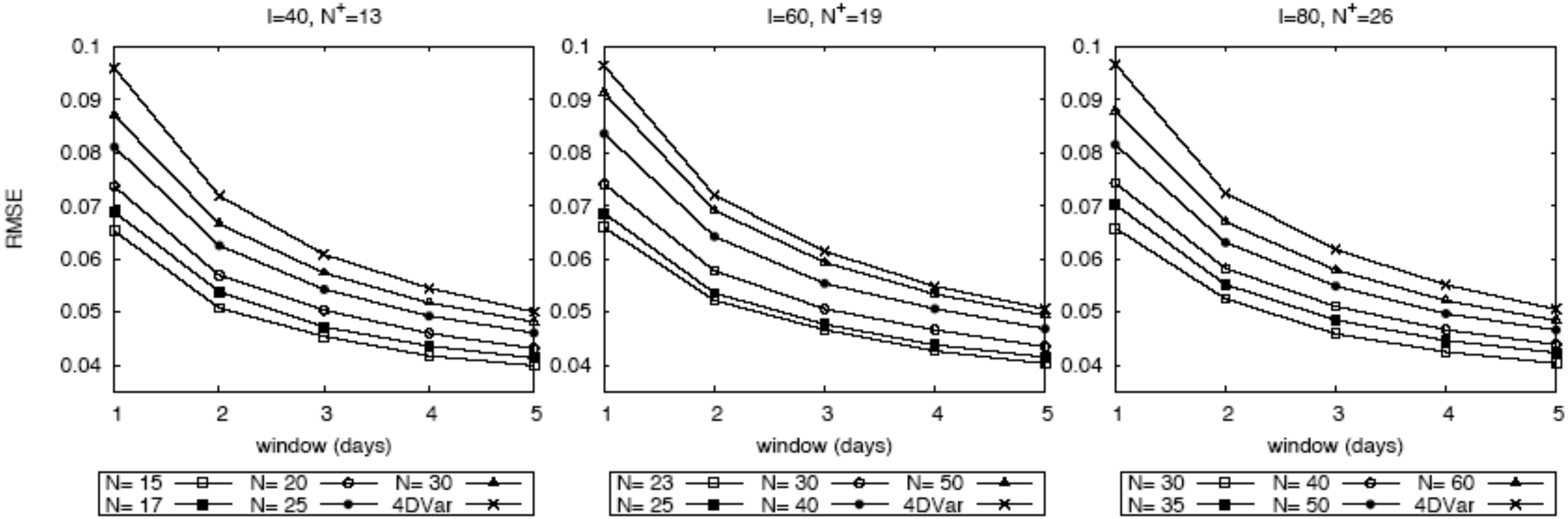


Figure 2. Time average RMS analysis error at $t = \tau$ as a function of the length of the assimilation window for three model configurations: I=40, 60, 80. Different curves in the same panel refer to a different subspace dimension N of 4DVar-AUS and to standard 4DVar. $\sigma_o = 0.2$.

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.

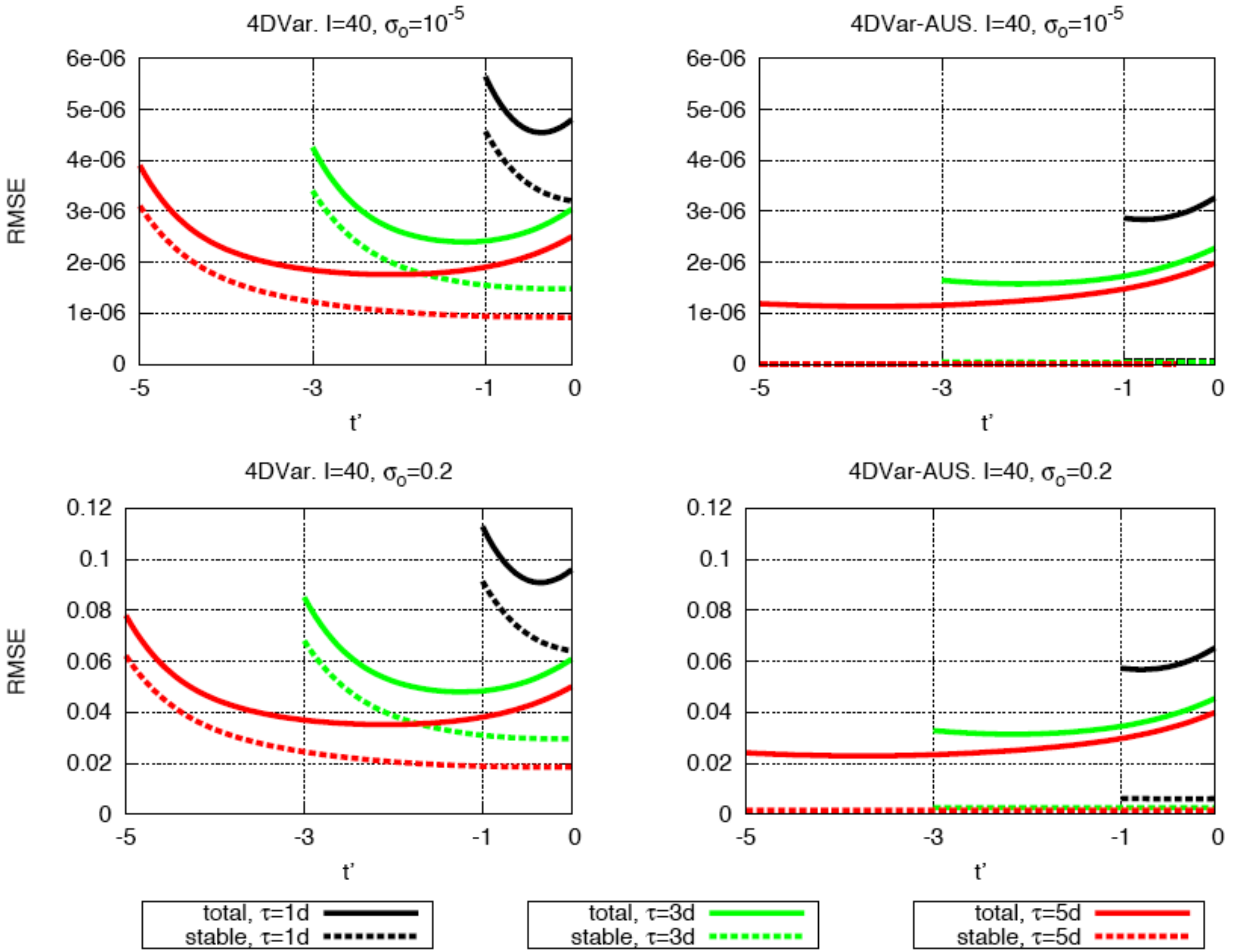


Figure 3. Time average RMS error within 1, 3, 5 days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = .2, 10^{-5}$ for the model configuration $I = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to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

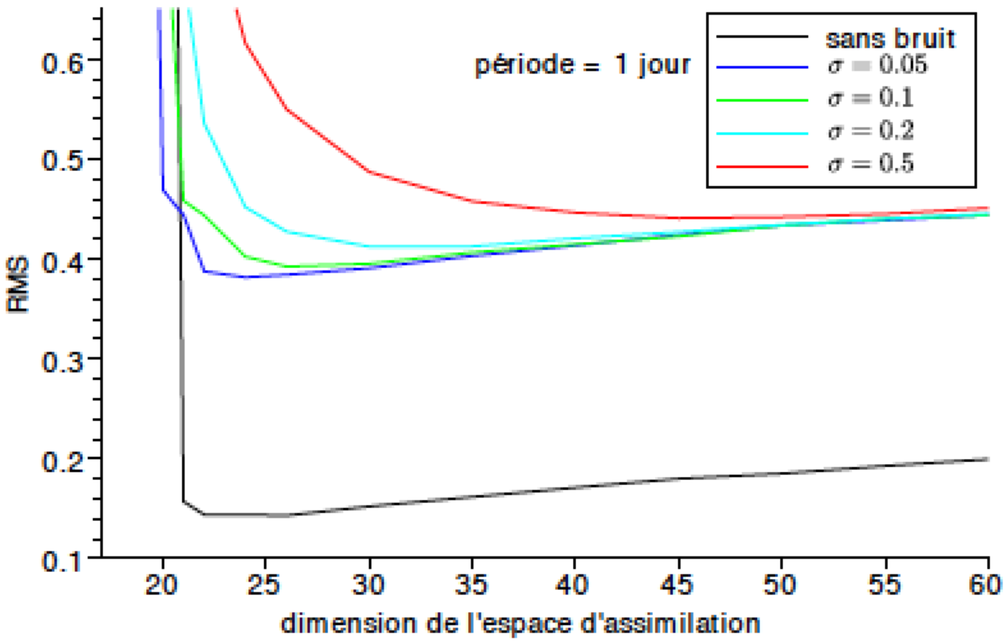
These results apply because no explicit background is available at the initial time of the assimilation window (only the unstable subspace is known). A proper background (obtained for instance from a properly implemented Kalman Filter, or from an Ensemble Variational Assimilation) would not only say that 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.

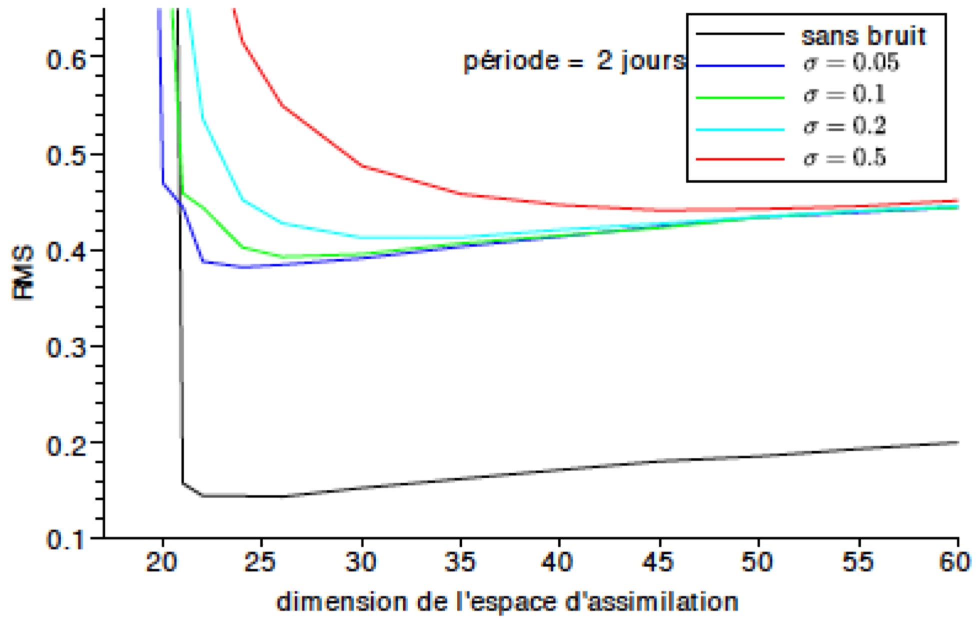
Questions.

- 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 τ) 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).

Exact bayesian estimation ?

Particle filters

Predicted ensemble at time t : $\{x^b_l, l = 1, \ldots, L\}$, each element with its own weight (probability) $P(x^b_l)$

Observation vector at same time : $y = H(x) + \varepsilon$

Bayes' formula

$$P(x^b_l|y) = P(y|x^b_l)\, P(x^b_l) \,/\, P(y)$$

Defines updating of weights

Bayes' formula

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

If error ε is independent of all previous data

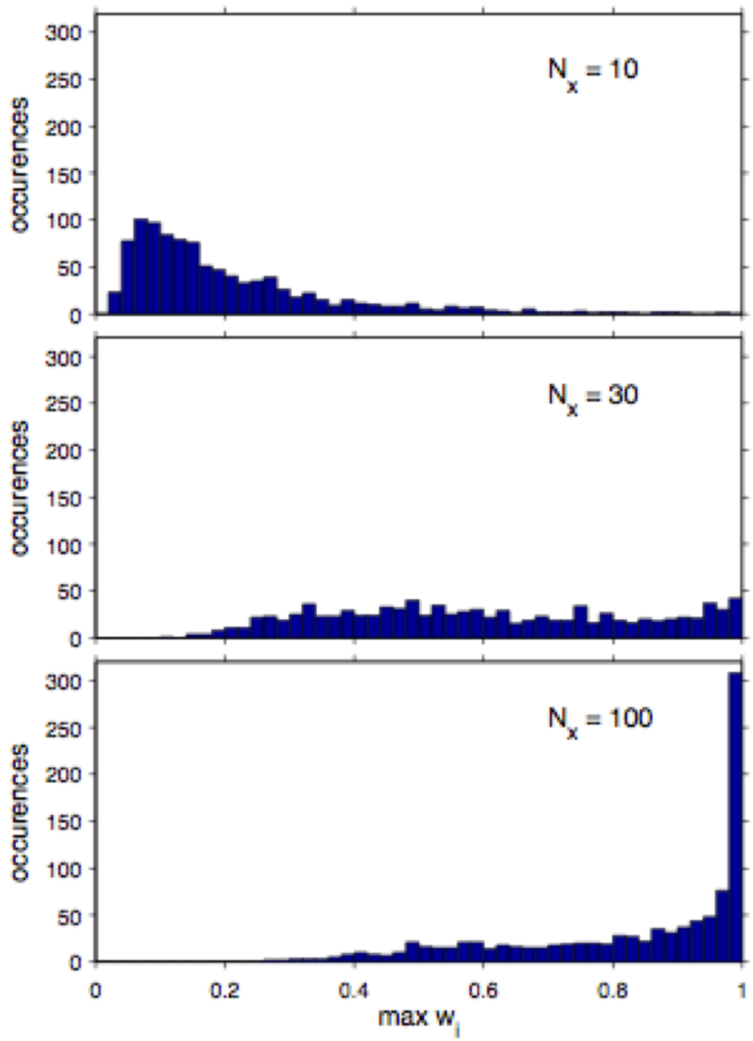
$$P(y|x^b{}_l) = P[\varepsilon = y - H(x^b{}_l)]$$

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.

Curse of dimensionality

Standard one-dimensional gaussian random variable X

$$P[\,|X| < \sigma\,] \approx 0.84$$

In dimension $n = 100$, $0.84^{100} = 3.10^{-8}$

.

χ^2-probability distribution of order p

$\chi^2(p) \sim \Sigma_p [\mathcal{N}(0, 1)]^2$

Expectation $m = p$, variance $\sigma^2 = 2p$

$\sigma/m = \sqrt{(2/p)}$

for large p, distribution is extremely peaked

Recall that, in gaussian variational assimilation, $2E(\mathcal{J}_{min})$, where $\mathcal{J}_{min}$ is minimum of objective function, follows a χ^2-probability distribution of order p

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, 2017, *Annales de la faculté des sciences de Toulouse Mathématiques*, **26** (4), 1051-1085)

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_l are multiplied by ensemble dimension L. Then p copies of each particle l are taken, where p is the integer part of Lw_l. Remaining particles, if needed, are taken randomly from the resulting distribution.

However, resampling is not sufficient to avoid degeneracy of filters.

Markov chain Monte Carlo (***MCMC***) ***Methods***

Sequence of random vectors $\{x^n, n = 0, \ldots\}$

Assume $P(x^n \mid x^{n-1}, \ldots, x^0) = P(x^n \mid x^{n-1})$

Markovianity. Verified in particular if $x^n = F(x^{n-1}, \eta)$, where F is deterministic, and η is random with *a priori* known probability distribution.

Sequence of observations $\{y^n, n = 0, \ldots\}$

Assume $P(y^n \mid x^n, x^{n-1}, \ldots, x^0) = P(y^n \mid x^n)$

Verified in particular if $y^n = G(x^n, \varepsilon)$, where G is deterministic, and ε is random with *a priori* known probability distribution.

We want to estimate $P(x^n \mid y^n, \ldots, y^0) \equiv P(x^n \mid y^{0:n})$

$$P(x^n \mid y^{0:n}) = P(x^n \mid y^n, y^{0:n\text{-}1}) = P(y^n \mid x^n, y^{0:n\text{-}1})\, P(x^n \mid y^{0:n\text{-}1}) \,/\, P(y^n \mid y^{0:n\text{-}1})$$
$$= P(y^n \mid x^n)\, P(x^n \mid y^{0:n\text{-}1}) \,/\, P(y^n)$$

$$P(x^n \mid y^{0:n\text{-}1}) = \int P(x^n \mid x^{n\text{-}1})\, P(x^{n\text{-}1} \mid y^{0:n\text{-}1})\, dx^{n\text{-}1}$$

Chapman-Kolmogorov equation

Particular case

$x^n = M_n\, x^{n\text{-}1} + \eta_n$ $\quad M_n$ linear, η_n Gaussian with *a priori* known pdf

$y^n = H_n\, x^n + \varepsilon_n$ $\quad H_n$ linear, ε_n Gaussian with *a priori* known pdf

$\Rightarrow$ Kalman filter

Idea :

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).

We are now to discuss a very interesting property of particle filters that has received little attention in the geophysical community. We start from Bayes:

$$p(x^{0:n}|y^{0:n}) = \frac{p(y^n|x^n)p(x^n|x^{n-1})}{p(y^n)} p(x^{0:n-1}|y^{0:n-1}). \tag{5.1}$$

To simplify the analysis, and since we concentrate on a filter here, let us first integrate out the past, to get:

$$p(x^n|y^{0:n}) = \frac{p(y^n|x^n)}{p(y^n)} \int p(x^n|x^{n-1})p(x^{n-1}|y^{0:n-1})\, \mathrm{d}x^{n-1}. \tag{5.2}$$

This expression does not change when we multiply and divide by a so-called proposal transition density $q(x^n|x^{n-1}, y^n)$, so:

$$\begin{aligned} &p(x^n|y^{0:n}) \\ &= \frac{p(y^n|x^n)}{p(y^n)} \int \frac{p(x^n|x^{n-1})}{q(x^n|x^{n-1}, y^n)} q(x^n|x^{n-1}, y^n)p(x^{n-1}|y^{0:n-1})\, \mathrm{d}x^{n-1}. \end{aligned} \tag{5.3}$$

van Leeuwen, 2017, *Annales de la faculté des sciences de Toulouse Mathématiques*, **26** (4), 1051-1085

As long as the support of $q(x^n|x^{n-1}, y^n)$ is equal to or larger than that of $p(x^n|x^{n-1})$ we can always do this. This last condition makes sure we don't divide by zero. Let us now assume that we have an equal-weight ensemble of particles from the previous analysis at time $n-1$, so

$$p(x^{n-1}|y^{0:n-1}) = \sum_{i=1}^{N} \frac{1}{N}\delta_{x_i^{n-1}}. \tag{5.4}$$

Using this in the equation above gives:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} \frac{1}{N}\frac{p(y^n|x^n)}{p(y^n)}\frac{p(x^n|x_i^{n-1})}{q(x^n|x_i^{n-1}, y^n)}q(x^n|x_i^{n-1}, y^n). \tag{5.5}$$

As a last step, we run the particles from time $n-1$ to n, i.e. we sample from the transition density. However, instead of drawing from $p(x^n|x_i^{n-1})$, so running the original model, we sample from $q(x^n|x_i^{n-1}, y^n)$, so from a modified model. Let us write this modified model as

$$x^n = g(x^{n-1}, y^n) + \hat{\beta}^n \tag{5.6}$$

so that we can write for the transition density, assuming $\hat{\beta}^n$ is Gaussian distributed with covariance $\hat{Q}$:

$$q(x^n|x^{n-1}, y^n) = N(g(x^{n-1}, y^n), \hat{Q}). \tag{5.7}$$

van Leeuwen, 2017, *ibid*.

Drawing from this density leads to:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} \frac{1}{N} \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)} \delta(x^n - x_i^n) \qquad (5.8)$$

so the posterior pdf at time n can be written as:

$$p(x^n|y^{0:n}) = \sum_{i=1}^{N} w_i \delta_{x_i^n} \qquad (5.9)$$

with weights w_i given by:

$$w_i = \frac{1}{N} \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}. \qquad (5.10)$$

We recognise the first factor in this expression as the likelihood, and the second as a factor related to using the proposal transition density instead of the original transition density to propagate from time $n-1$ to n, so it is related to the use of the proposed model instead of the original model. Note that because the factor $1/N$ and $p(y^n)$ are the same for each particle and we are only interested in relative weights, we will drop them from now on, so

$$w_i = p(y^n|x_i^n) \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}. \qquad (5.11)$$

van Leeuwen, 2017, *ibid*.

Several variants of proposal densities have been defined and studied : perform an EnKF up to observation time, and then use the obtained ensemble as proposal density, *nudge* the model integration between times n-1 and n towards the observations at time n, perform a 4D-Var on each particle, *optimal proposal density*, …

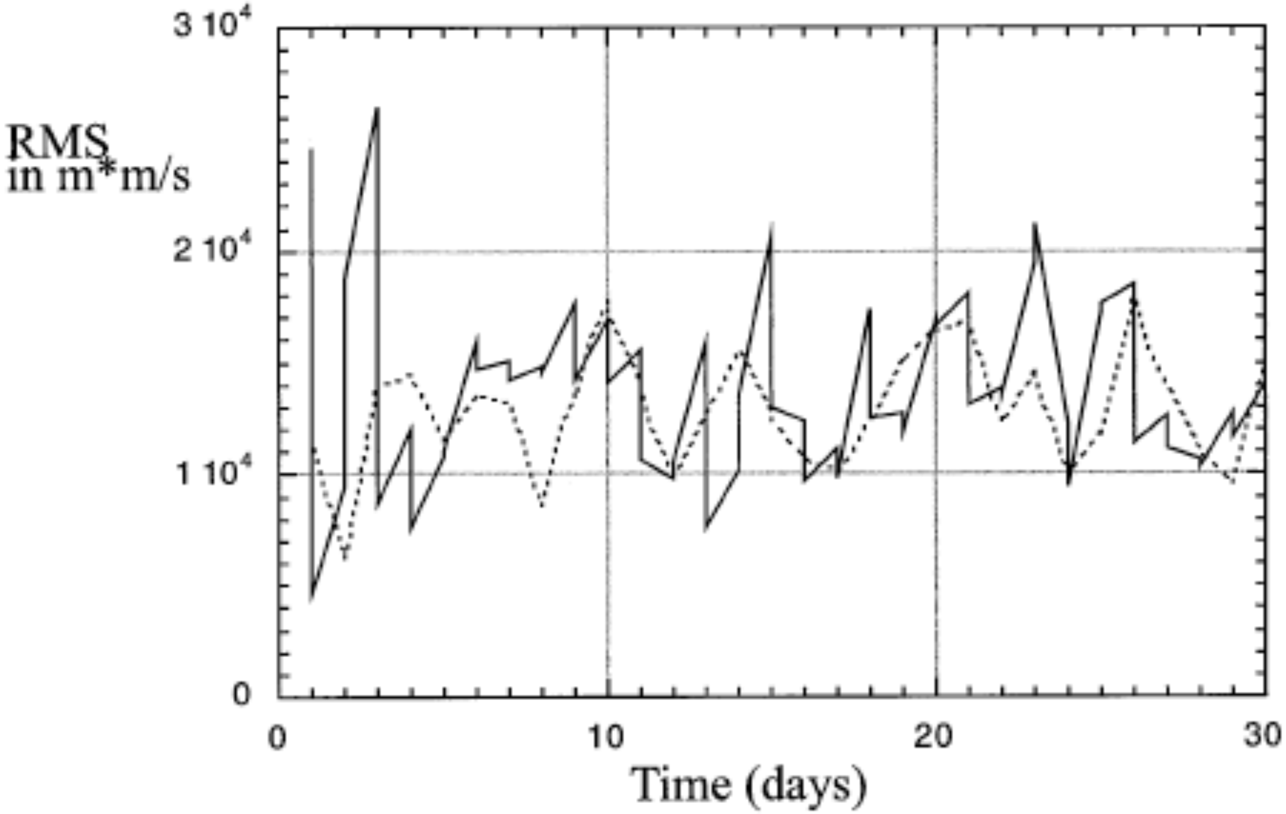


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

The Equivalent-Weights Particle Filter (Ades and van Leeuwen, *QJRMS*, 2013).

Make the proposal density depend on the whole ensemble at time n-1, and not only on x_l^{n-1}, use density of the form $q(x^n \mid x^{n-1}{}_{1,L}, y^n)$, where 1,L denotes all ensemble indices, rather than of the more restrictive form $q(x^n \mid x_l^{n-1}, y^n)$. This gives many degrees of freedom which can be exploited for obtaining at time n an ensemble with almost equal weights.

Example Vorticity equation model with random error η.

$$\frac{D(\zeta + f)}{Dt} = \eta$$

State-vector dimension ≈ 65,000

Decorrelation time: 25 timesteps

One complete noisy model field observed every 50 timesteps

24 particles

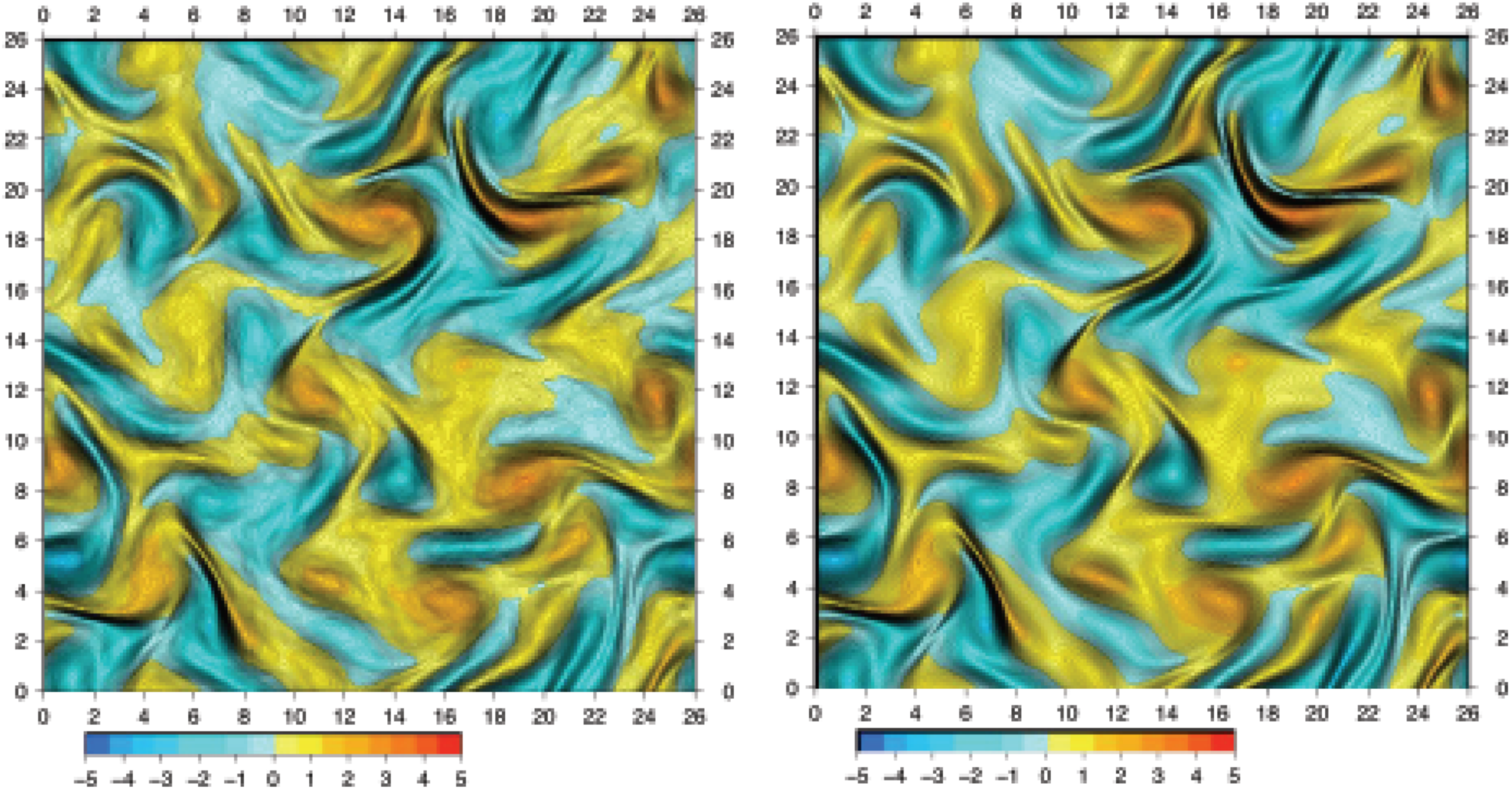


Figure 5.3. Snap shot of the vorticity field of the truth (right) and the particle filter mean (left) at time 25. Note the highly chaotic state of the fields, and the close to perfect tracking. (12 observations)

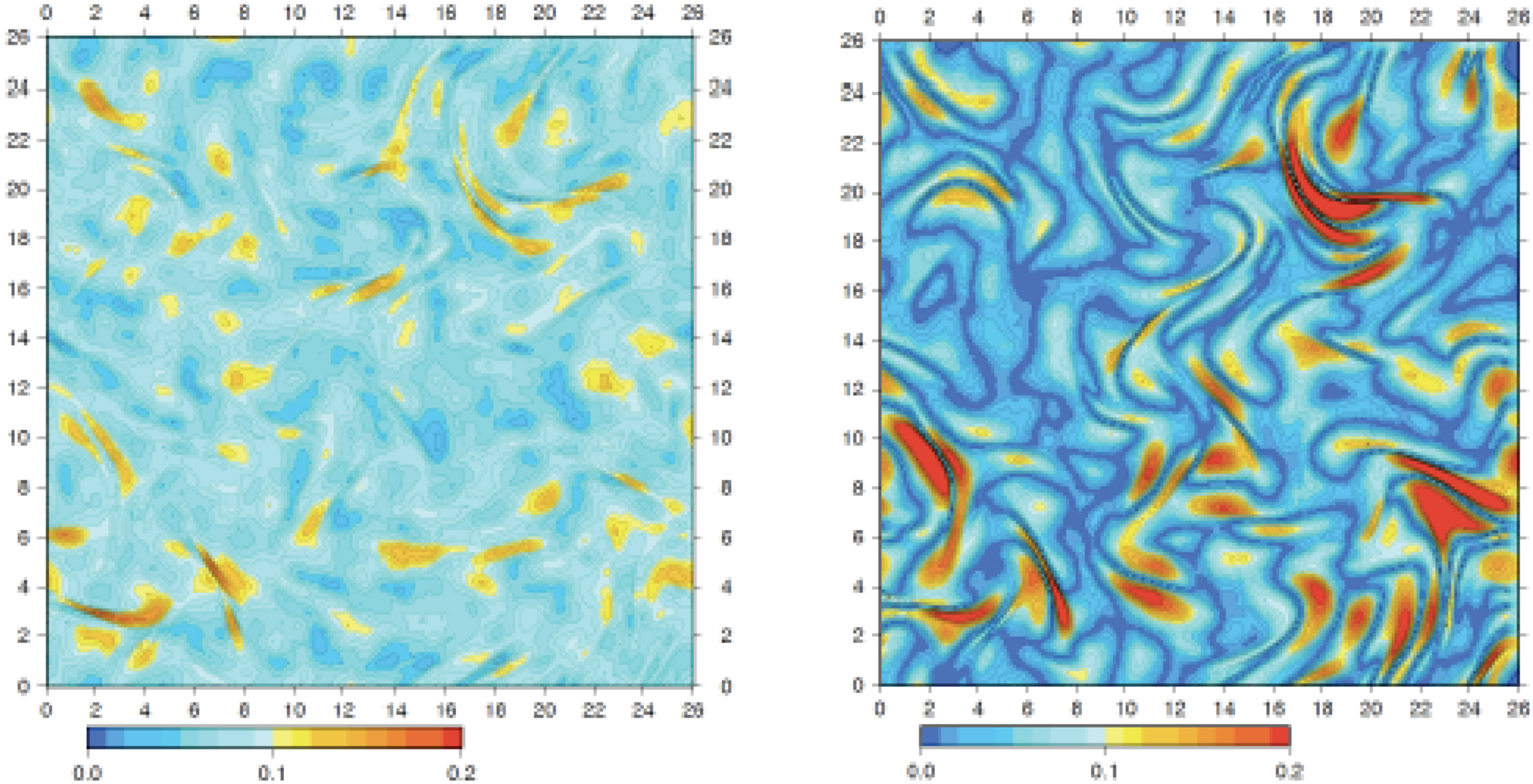


Figure 5.4. Snap shot of the absolute value of the mean-truth misfit and the standard deviation in the ensemble. The ensemble underestimates the spread at several locations, but averaged over the field it is slightly higher, 0.074 versus 0.056.

Bayesianity : experts say all these filters are bayesian (in the limit of infinite ensemble size)

Possible difficulties : numerical implementation, numerical cost

Particle filters are actively studied (van Leeuwen, Morzfeld, …)

- Ensemble Variational Assimilation (*EnsVAR*).

  (work with M. Jardak, 2018)

Ensemble Variational Assimilation

Data of the form

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

Conditional probability distribution is

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

with

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

Variational form

$$P(x \mid z) \propto \exp[\, -(z - \Gamma\xi)^T S^{-1} (z - \Gamma\xi)/2 \,] \propto \exp[\, -(\xi - x^a)^T (P^a)^{-1} (\xi - x^a)/2 \,]$$

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

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

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

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

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

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

and compute

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

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.

Used at ECMWF and Météo-France (under the name *Ensemble of Data Assimilations*, *EDA*) for defining initial conditions of ensemble prediction, and also for defining background error covariance matrix in 4D-Var, but not for assimilation *per se*.

Present purpose

Evaluate EnsVar as a probabilistic estimator when implemented in nonlinear and/or non-Gaussian cases, *i. e.*, through minimization of

$$\xi \in \mathcal{X} \rightarrow \mathcal{J}(\xi) \equiv (1/2)\,[\Gamma(\xi) - z']^{\mathrm{T}} S^{-1} [\Gamma(\xi) - z']$$

where Γ may be nonlinear, and errors affecting data z may be non-Gaussian.

- 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

The Lorenz96 model

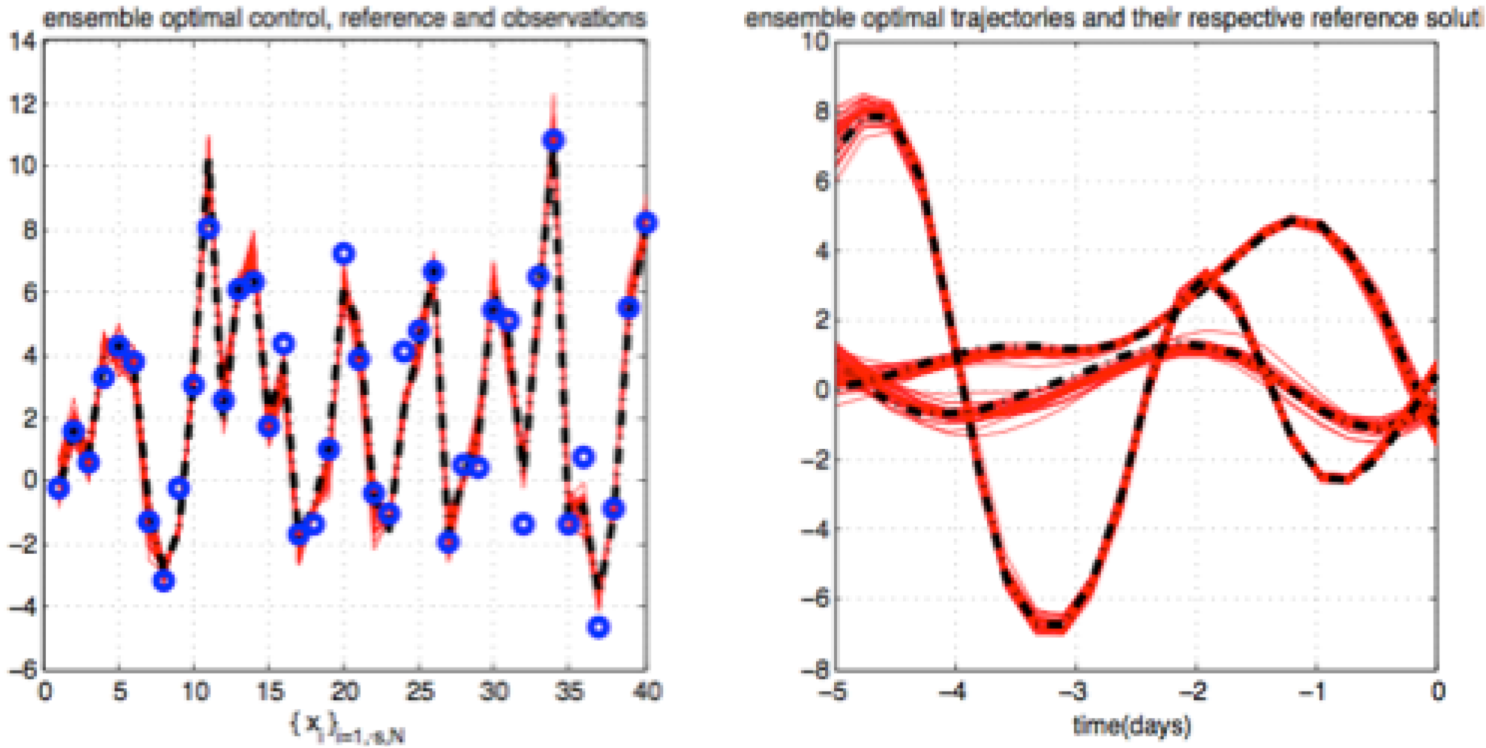
- Forward model

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

- 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. $Nens = 30$, 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.

System produces wavelike chaotic motions, with properties similar to those of midlatitude atmospheric waves

- generally westward phase velocity
- typical predictability time : 5 ‘days’
- in addition, quadratic terms conserve ‘energy’



Experimental procedure (1)

0. Define a *reference solution* x_t^r by integration of the numerical model

1. Produce 'observations' at successive times t_k of the form

$$y_k = H_k x_k^r + \varepsilon_k$$

where H_k is (usually, but not necessarily) the unit operator, and ε_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

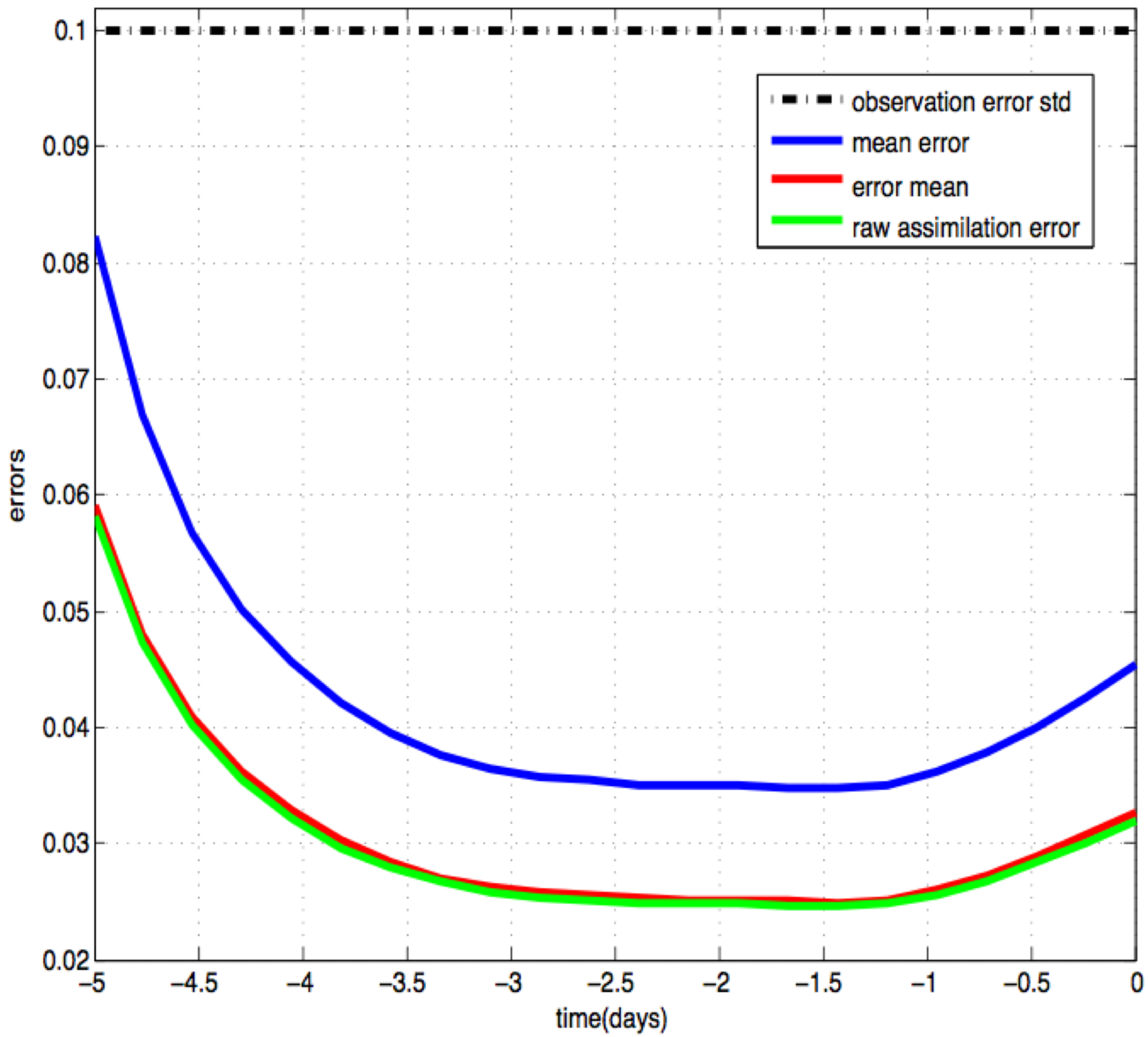
$$y_k \rightarrow z_k = y_k + \delta_k$$

where δ_k is an independent realization of the probability distribution which has produced ε_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.



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 circonstances where I have predicted 40% probability for rain).

Denote Y the predicted probability distribution, and X the verifying reality. Consider the probability distribution for the couples (X, Y) (that probability distribution can be obtained empirically). Reliability is the property that

$$P(X \mid Y) = Y \quad \text{for any } Y$$

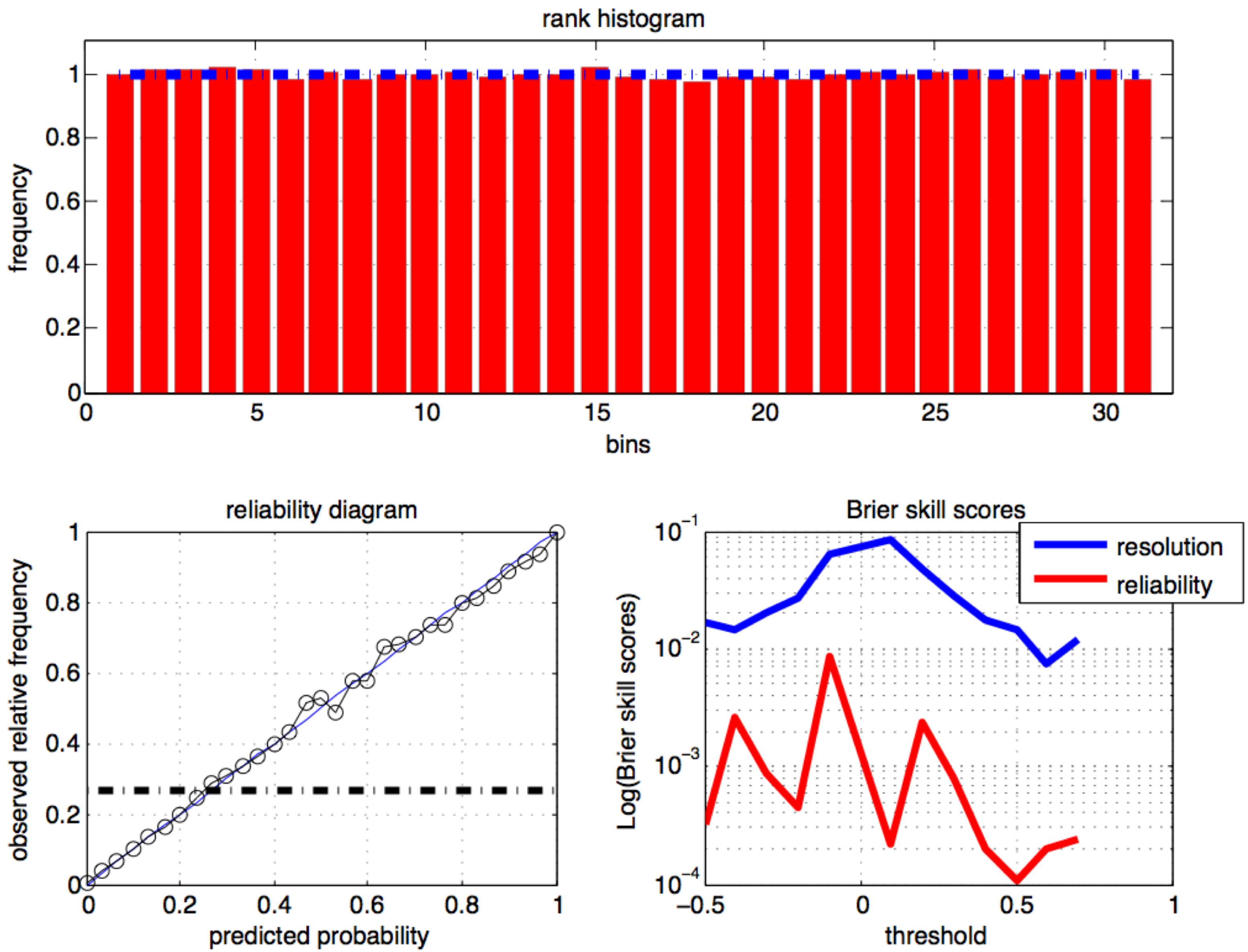
Reliability can be objectively validated, provided a large enough sample of realizations of the estimation system is available.

Bayesianity implies reliability, the converse not being true.

In addition, we evaluate *resolution* (also called *sharpness*), which bears no direct relation to bayesianity, and is the capability of the estimation system to *a priori* distinguish between different situations. It is best defined as the degree of statistical dependence between X and Y (J. Bröcker). Total absence of resolution is independence between X and Y, *viz*.

$$P(X \mid Y) = P(X) \quad \text{for any } Y$$

Resolution, beyond reliability, measures the degree of usefulness of the ensembles.



Linearized Lorenz'96. 5 days

Objective function

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

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

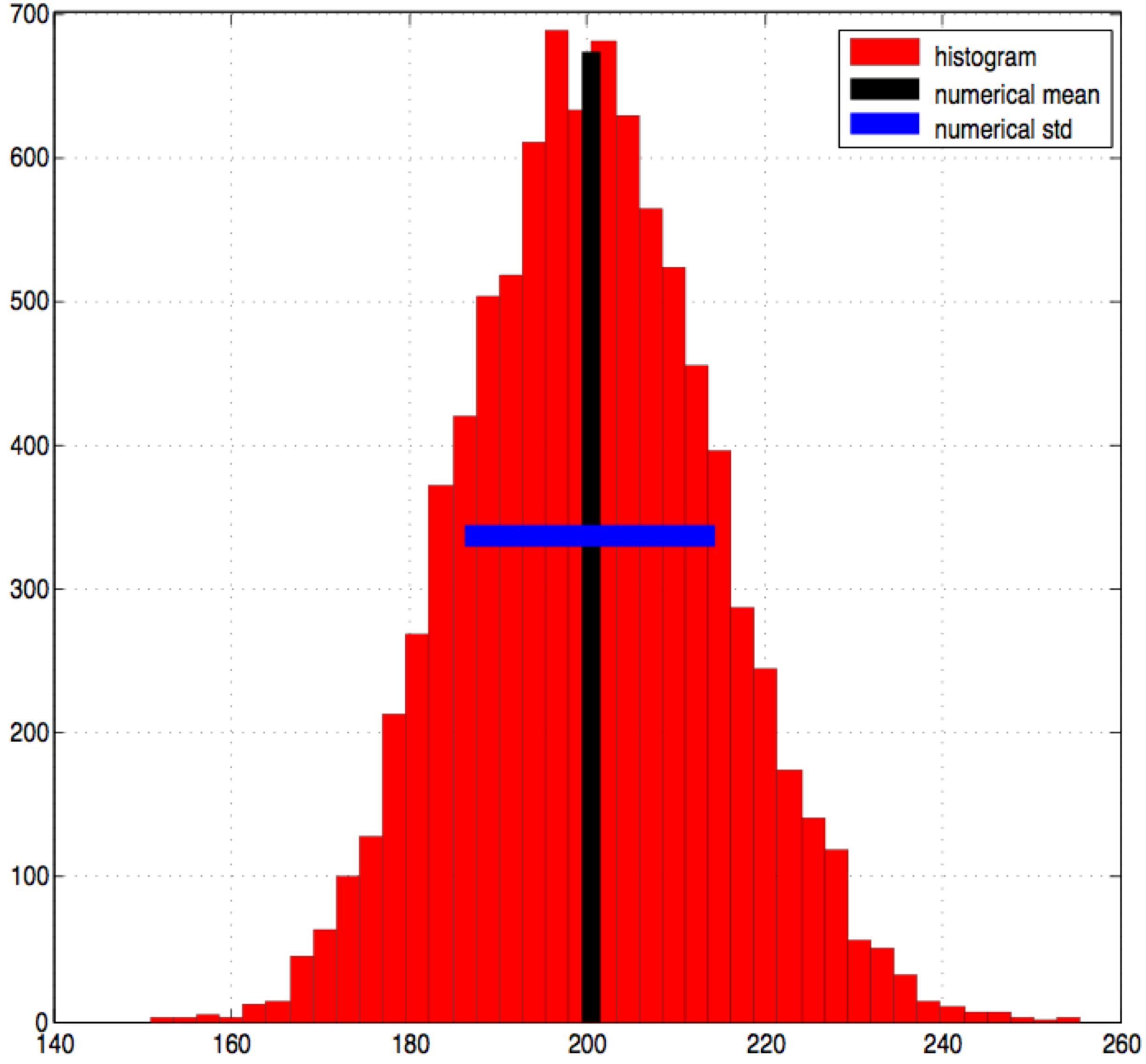
$$= (1/2) \, d^T [E(dd^T)]^{-1} d$$

where d is innovation

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

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

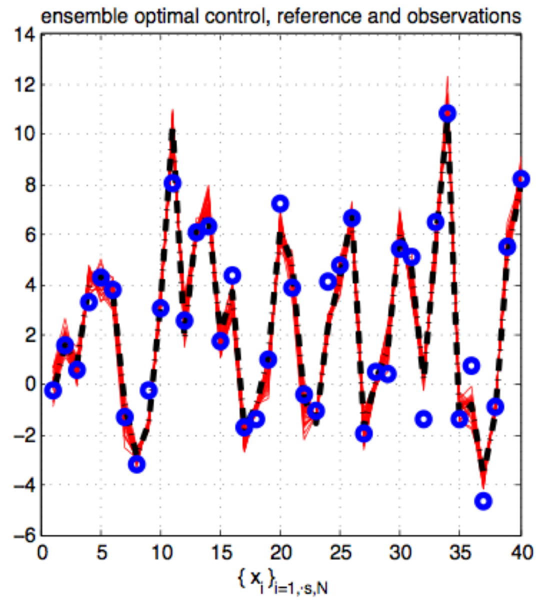
Remark. If in addition errors are gaussian, the quantity $2E(\mathcal{J}_{min})$ follows a χ^2-probability distribution of order p. For that reason the criterion $E(\mathcal{J}_{min}) = p/2$ is often called the χ^2 criterion. Also $Var(\mathcal{J}_{min}) = p/2$ in the gaussian case.



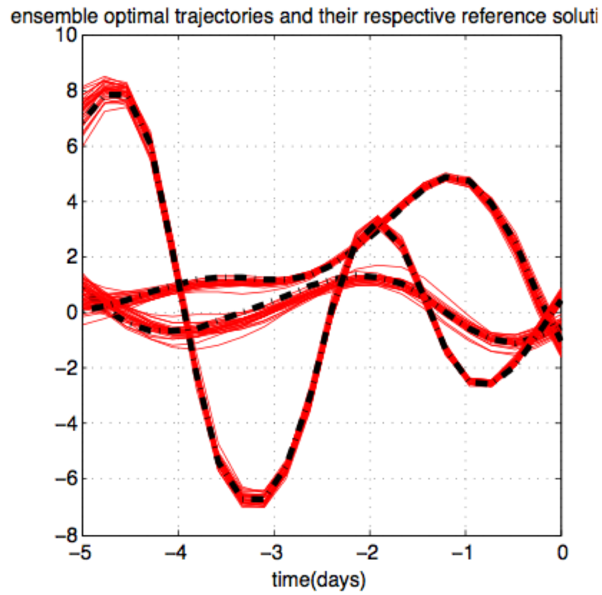
Linearized Lorenz'96. 5 days. Histogram of $\mathcal{J}_{min}$

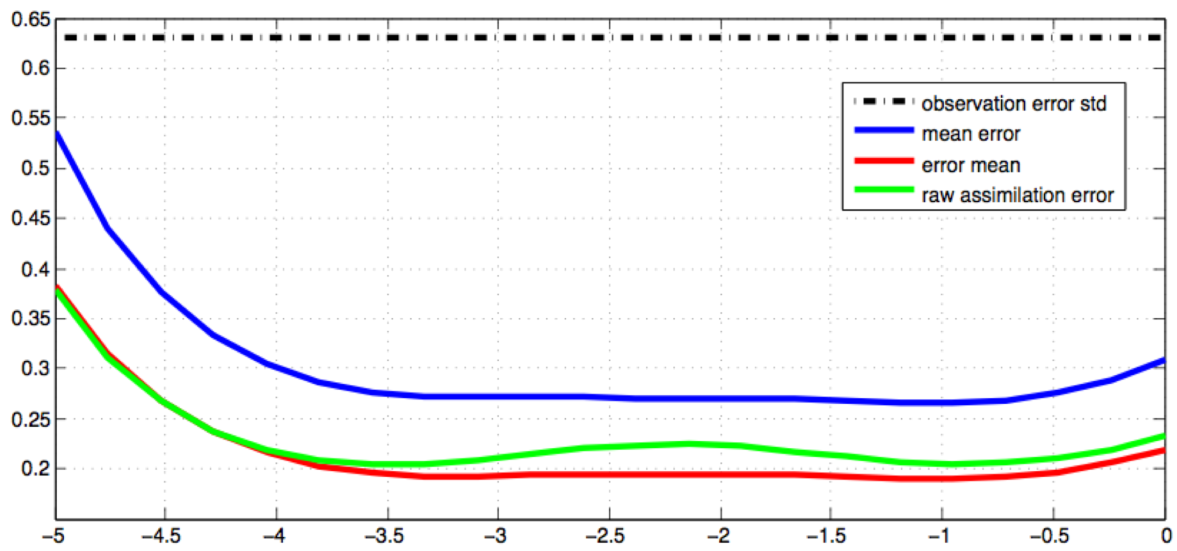
$E(\mathcal{J}_{min}) = p/2$ (=200) ; $\sigma(\mathcal{J}_{min}) = \sqrt{(p/2)}$ (≈14.14)

Observed values 199.39 and 14.27

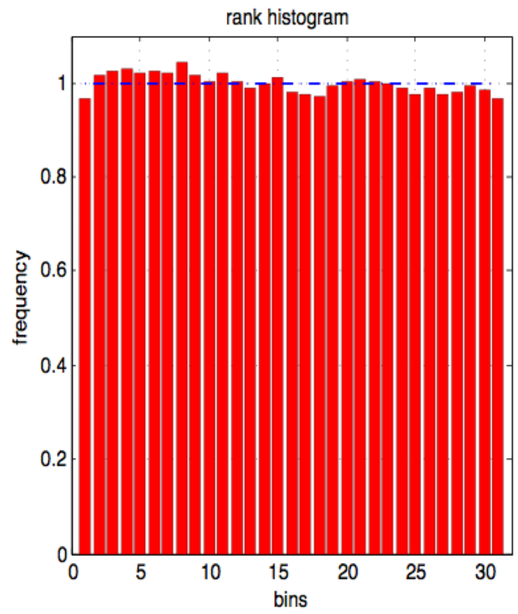


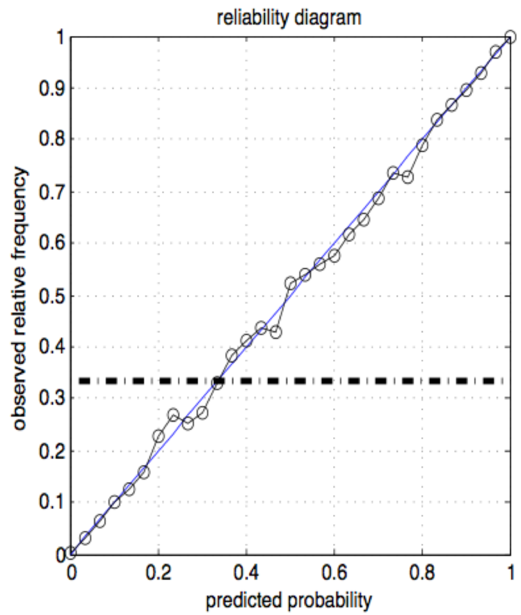
(initial time of assimilation window)

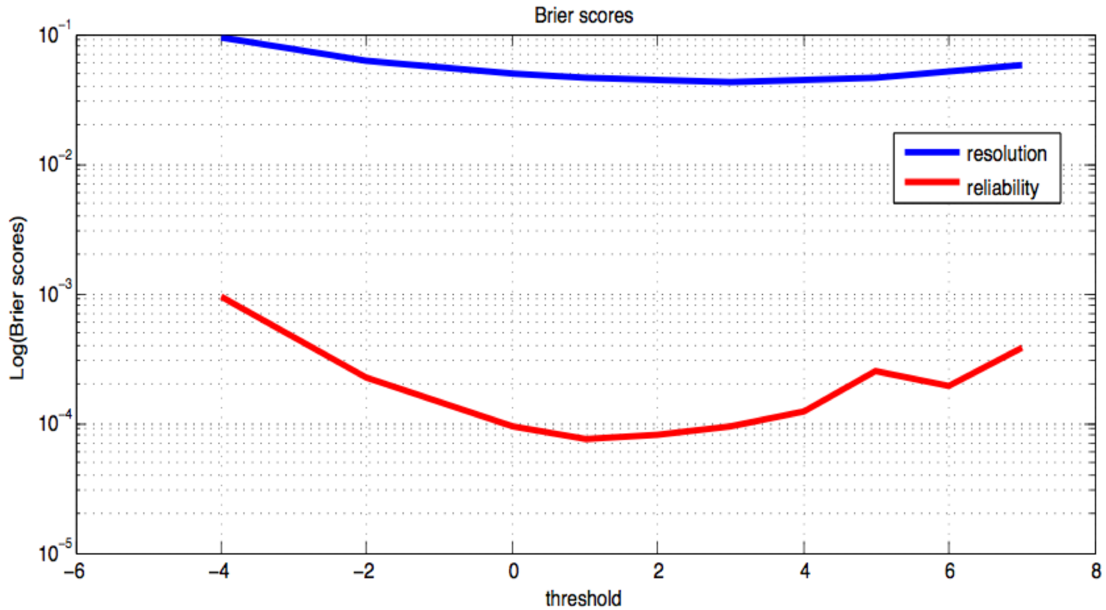




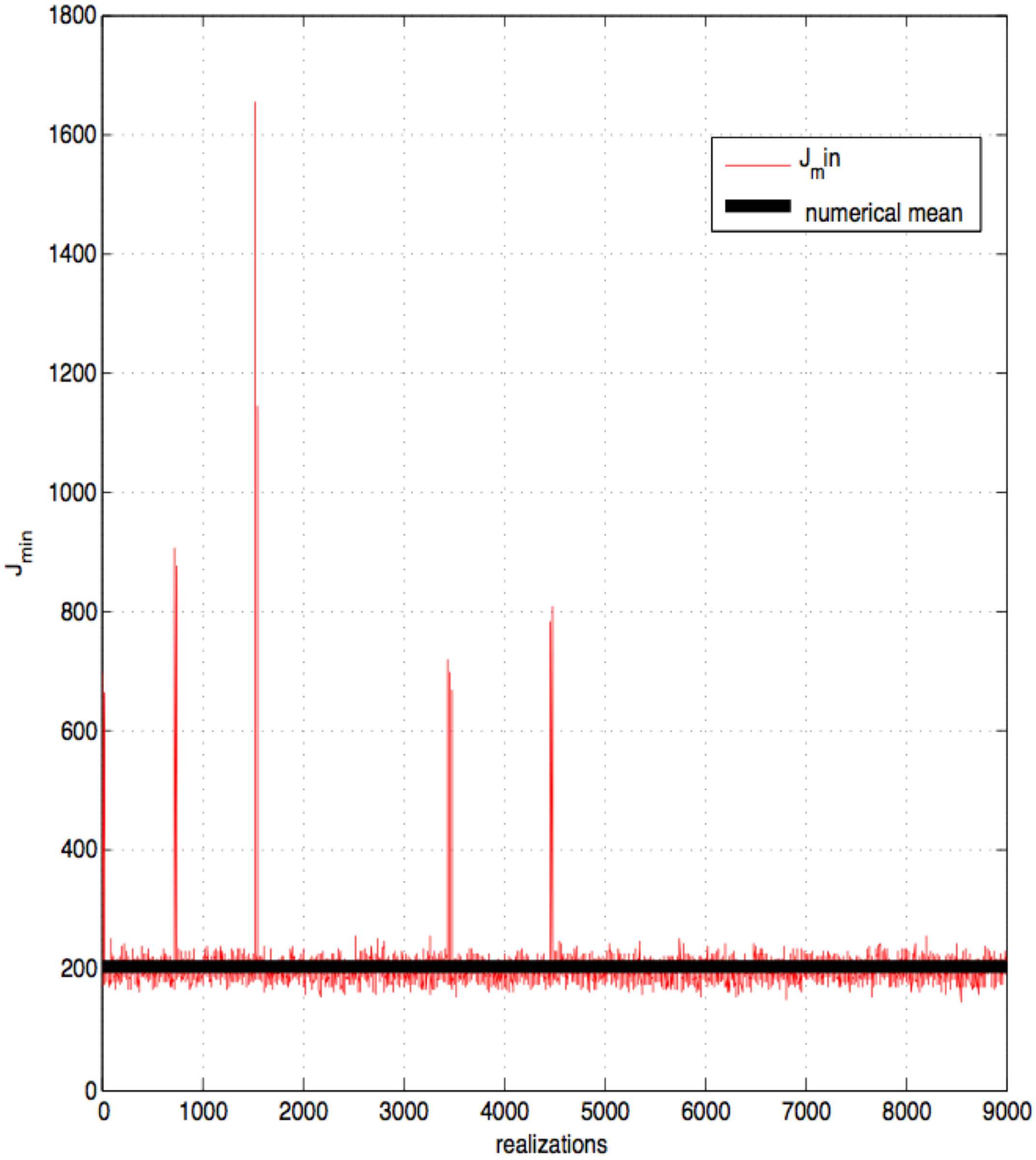
Nonlinear Lorenz'96. 5 days





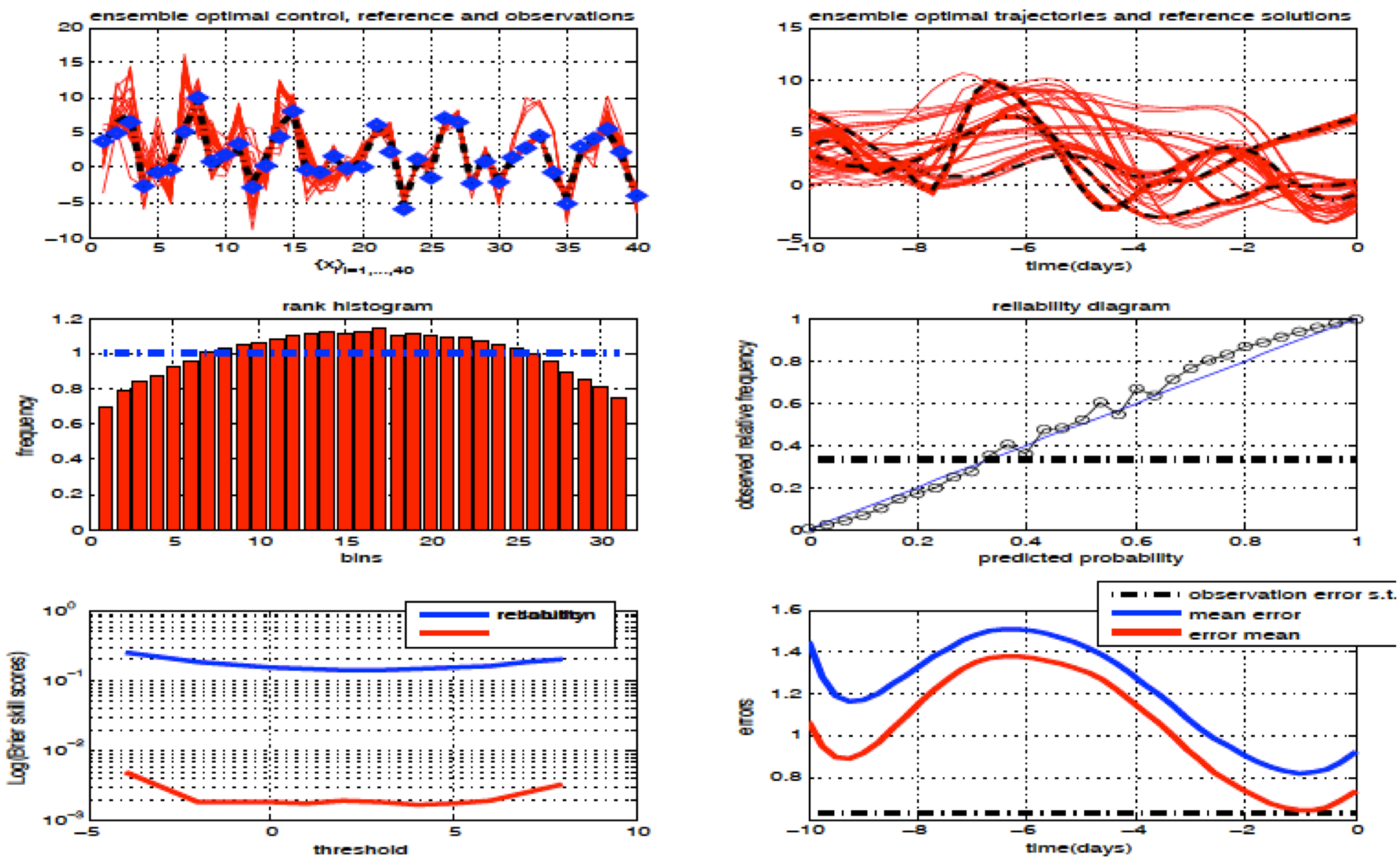


Nonlinear Lorenz'96. 5 days

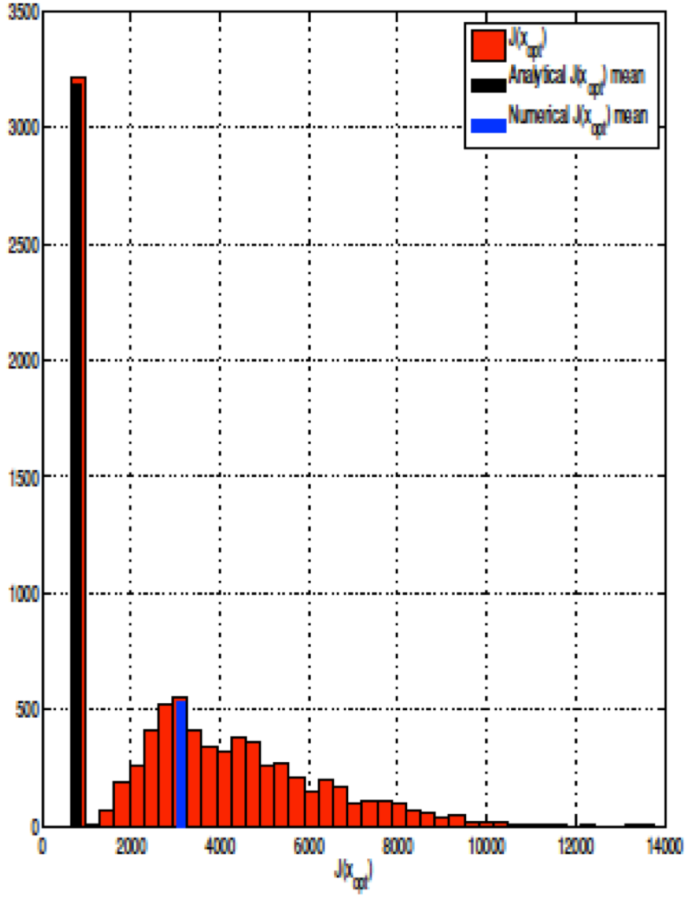


Nonlinear Lorenz'96. 5 days. Histogram of $\mathcal{J}_{min}$

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

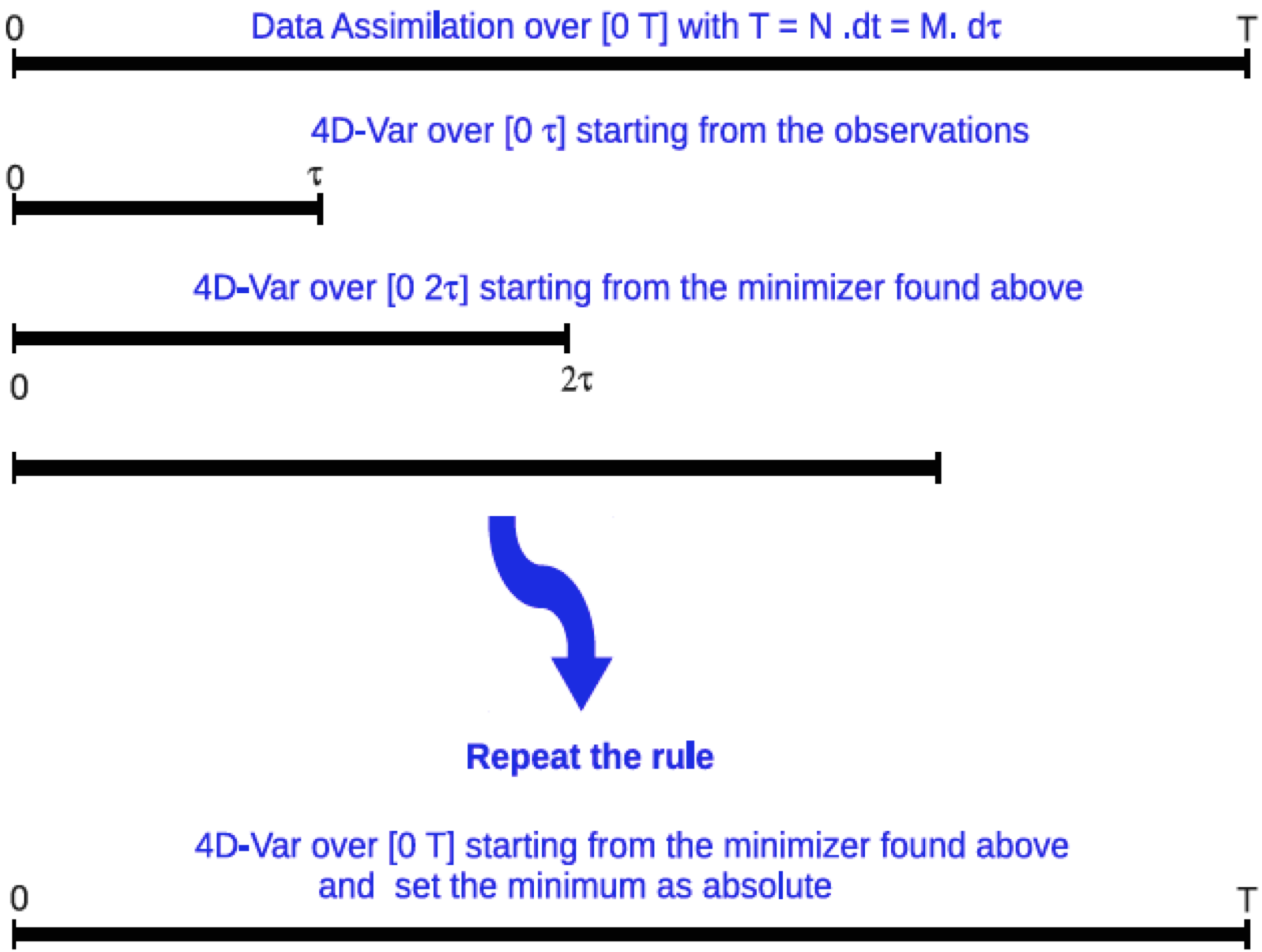


EnsVar : consistency

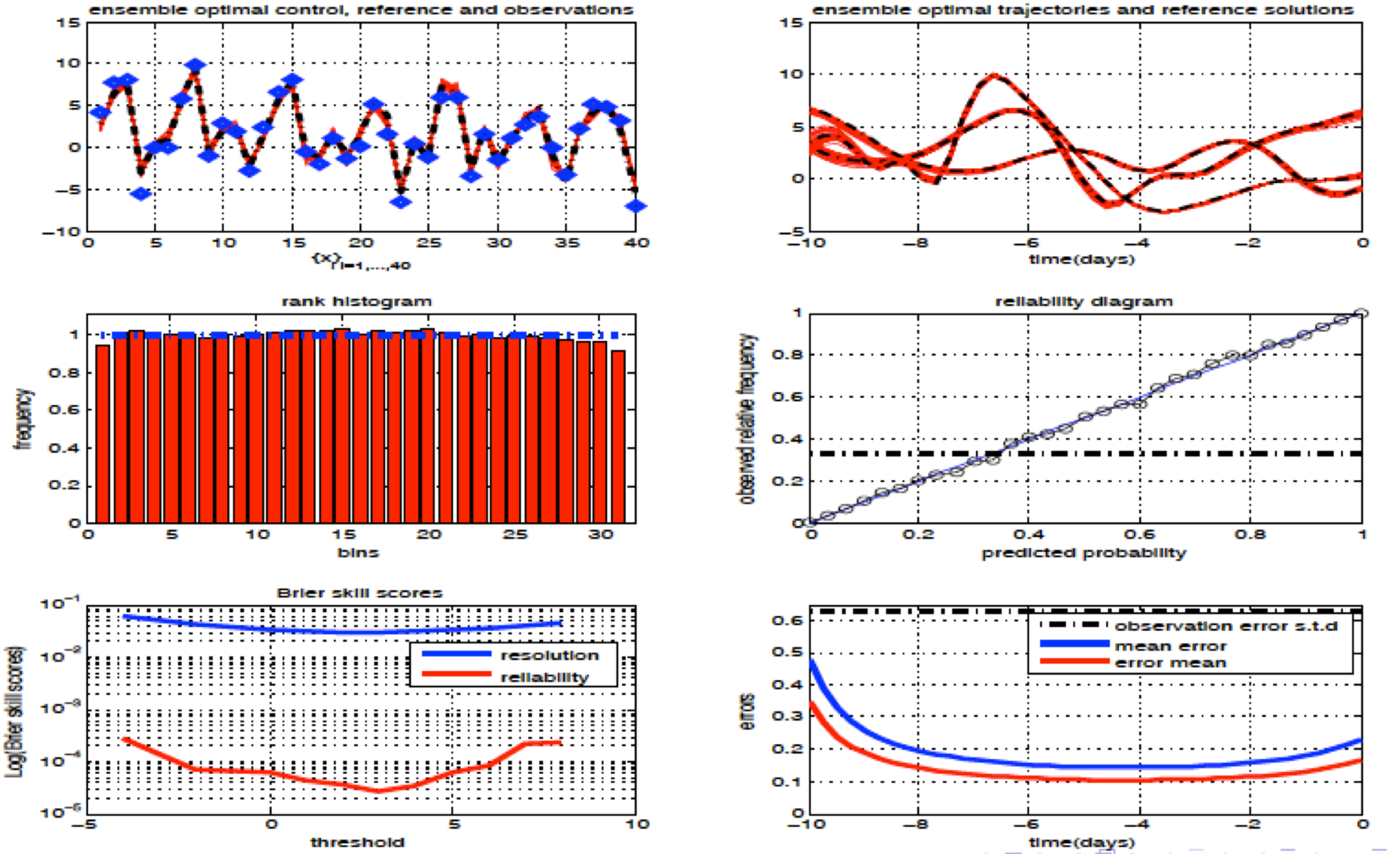


Nonlinear Lorenz'96. 10 days. Histogram of $\mathcal{J}_{min}$

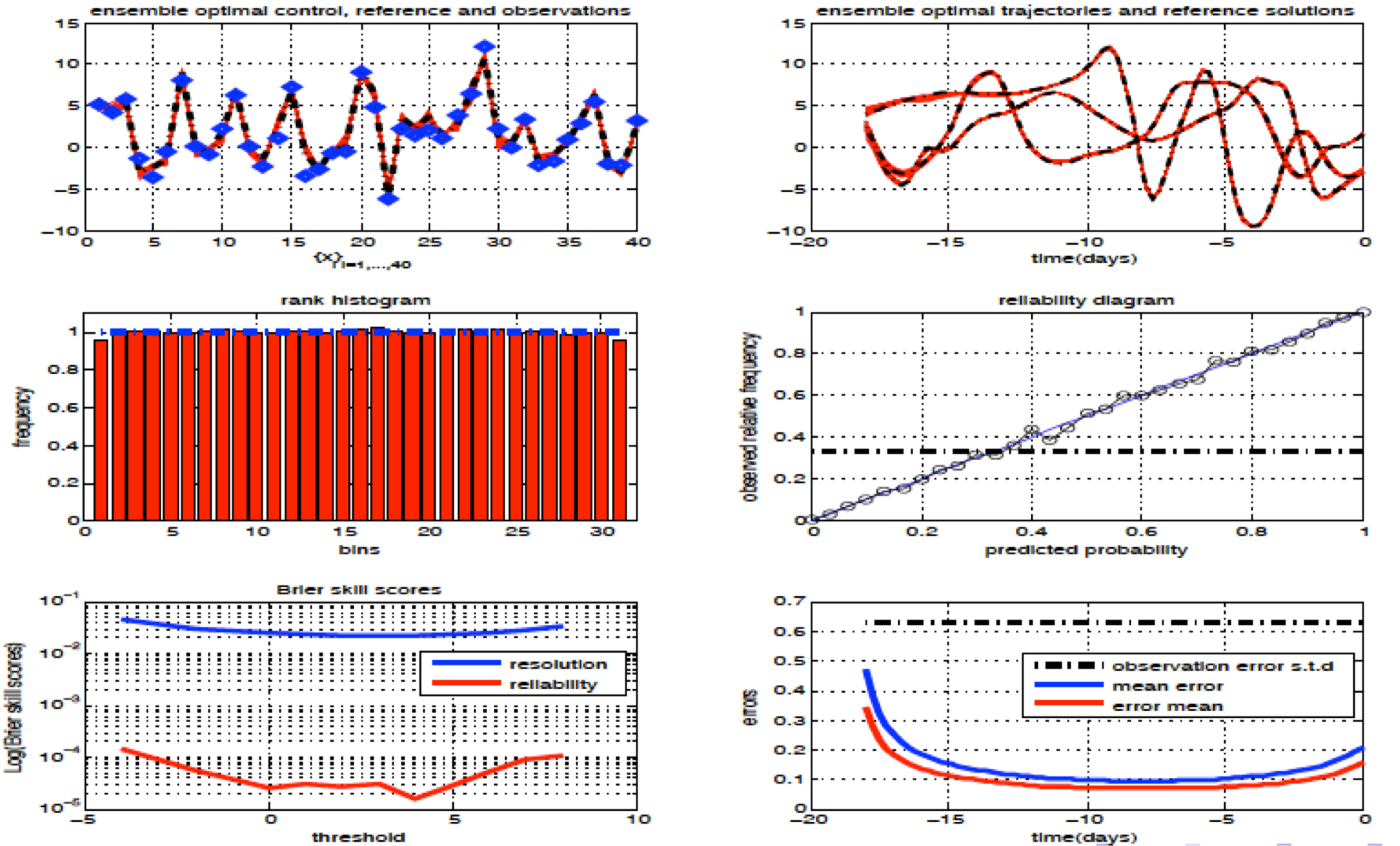
Quasi-Static Variational Assimilation (QSVA)



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



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

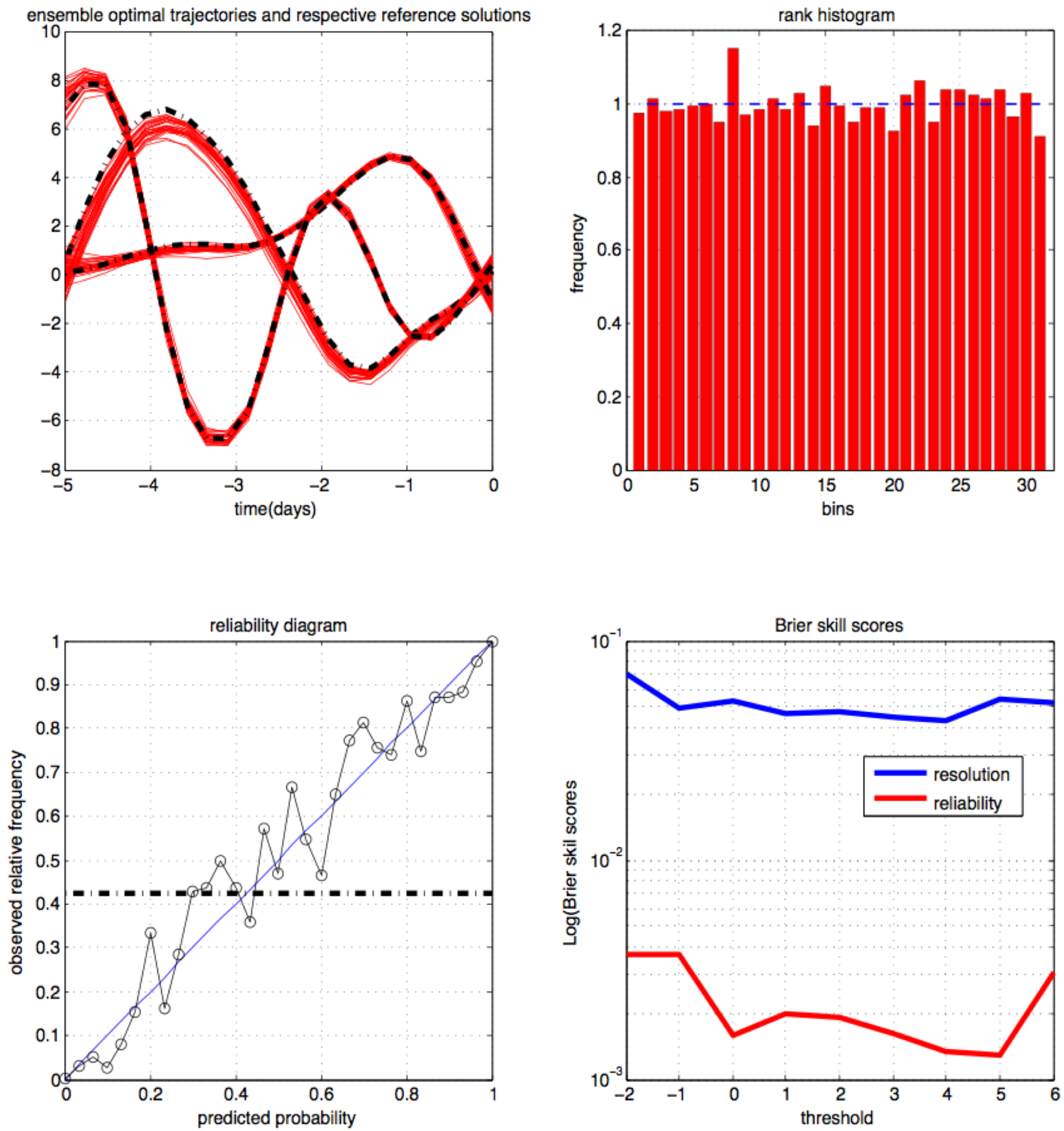


- Results are independent of the Gaussian character of the observation errors (trials have been made with various probability distributions)

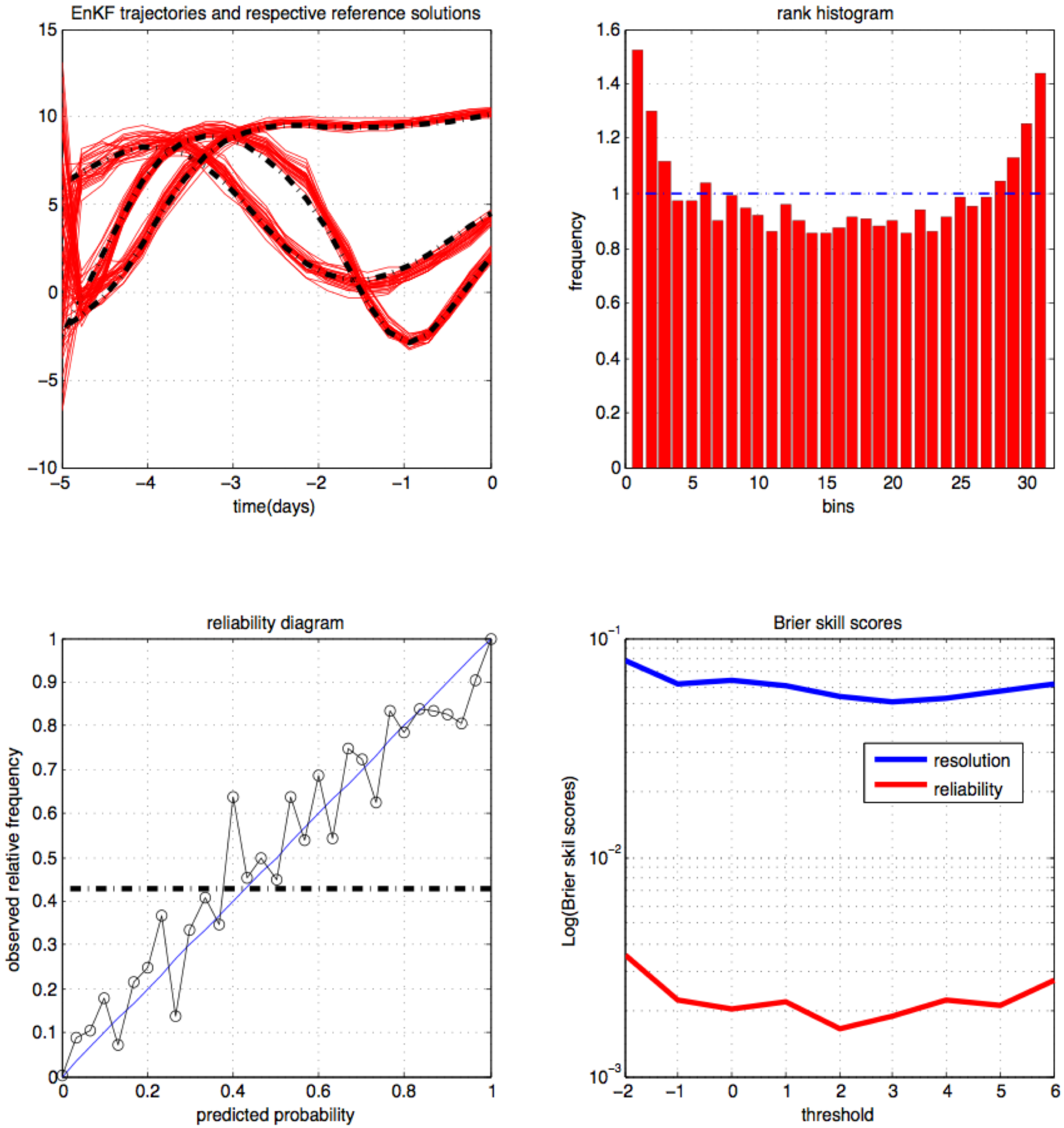
- Ensembles produced by EnsVar are very close to Gaussian, even in strongly nonlinear cases.

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

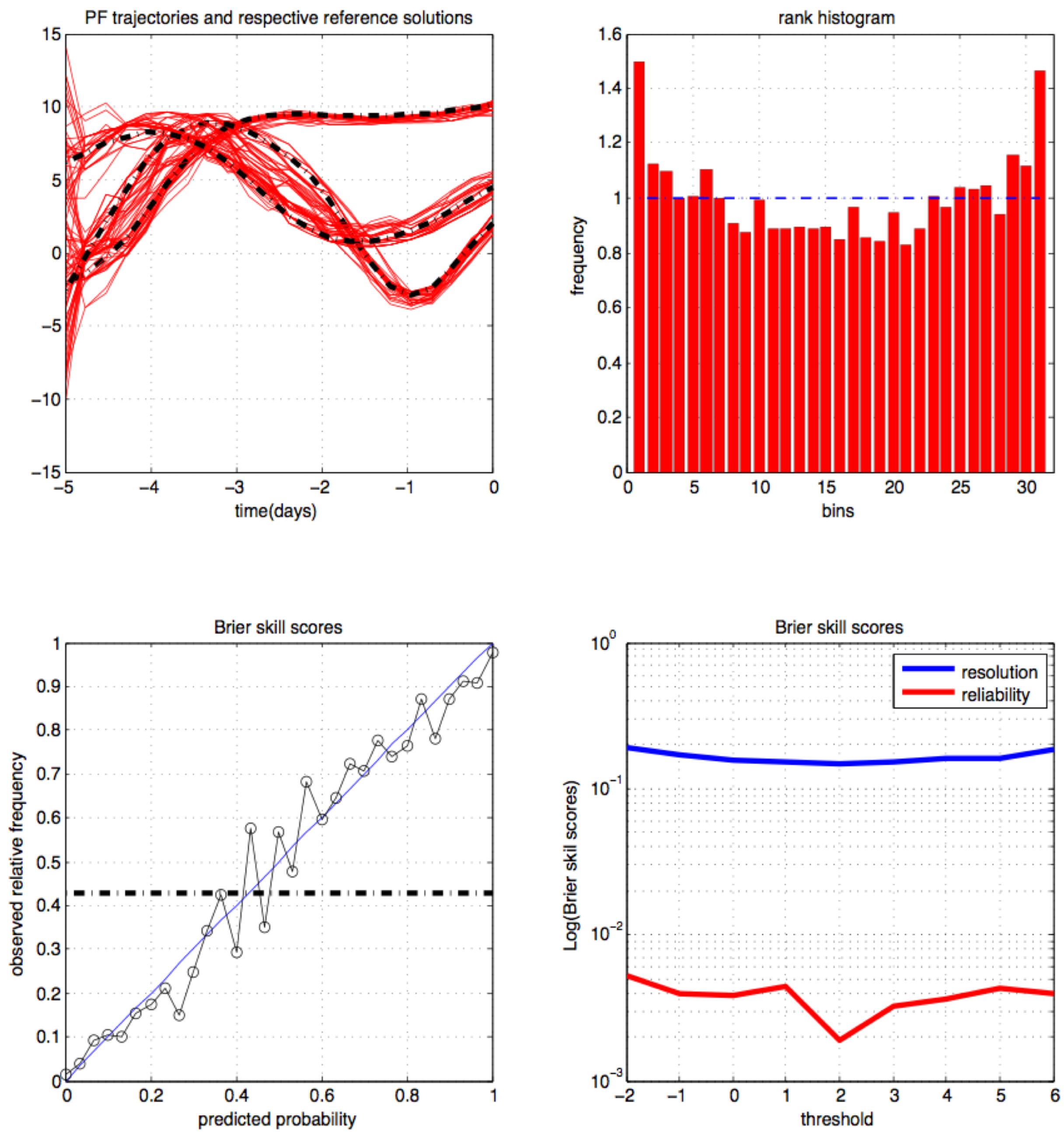
  Both of these algorithms being sequential, comparison is fair only at end of assimilation window



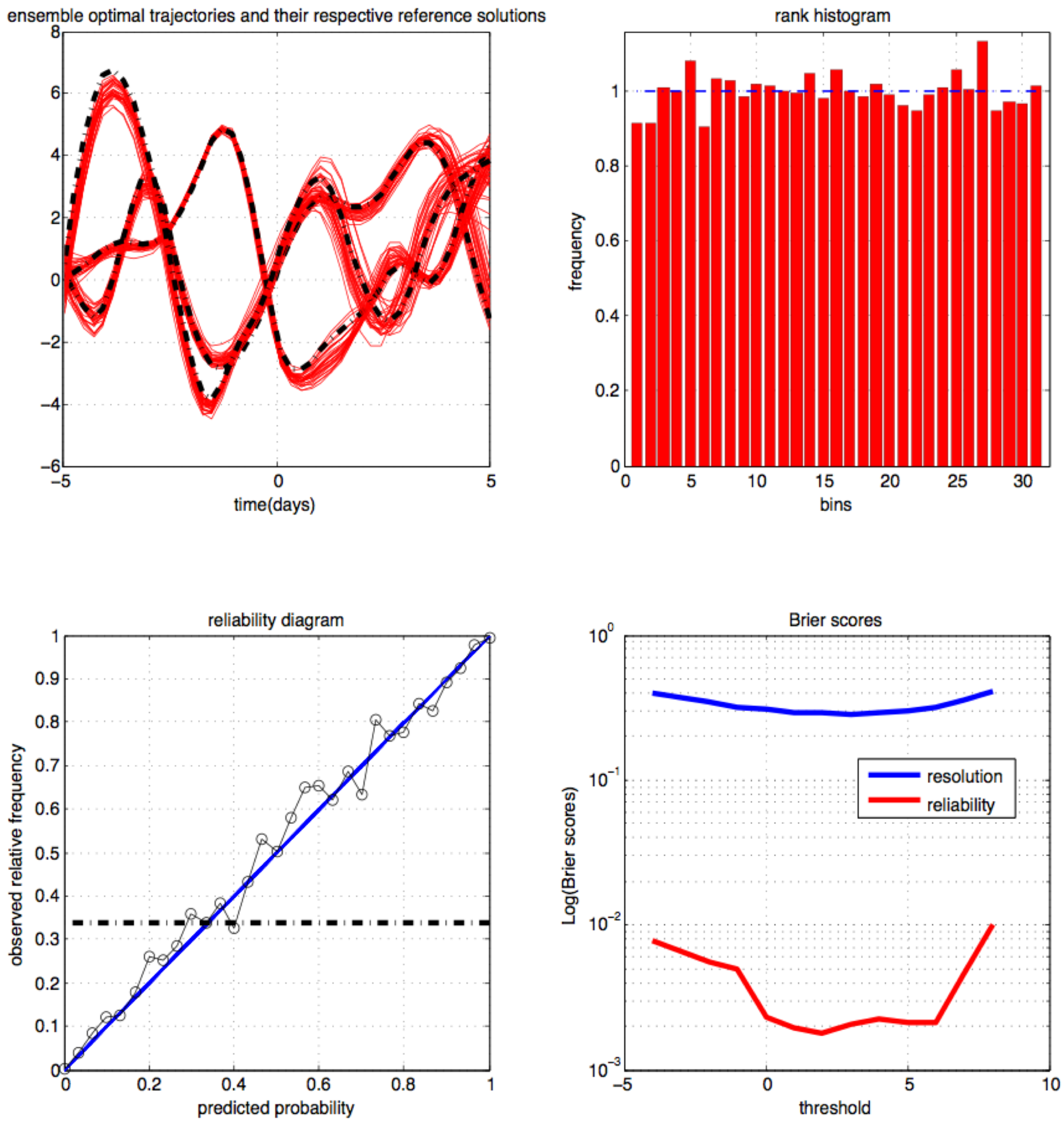
Nonlinear Lorenz'96. EnsVAR. Diagnostics after 5 days of assimilation



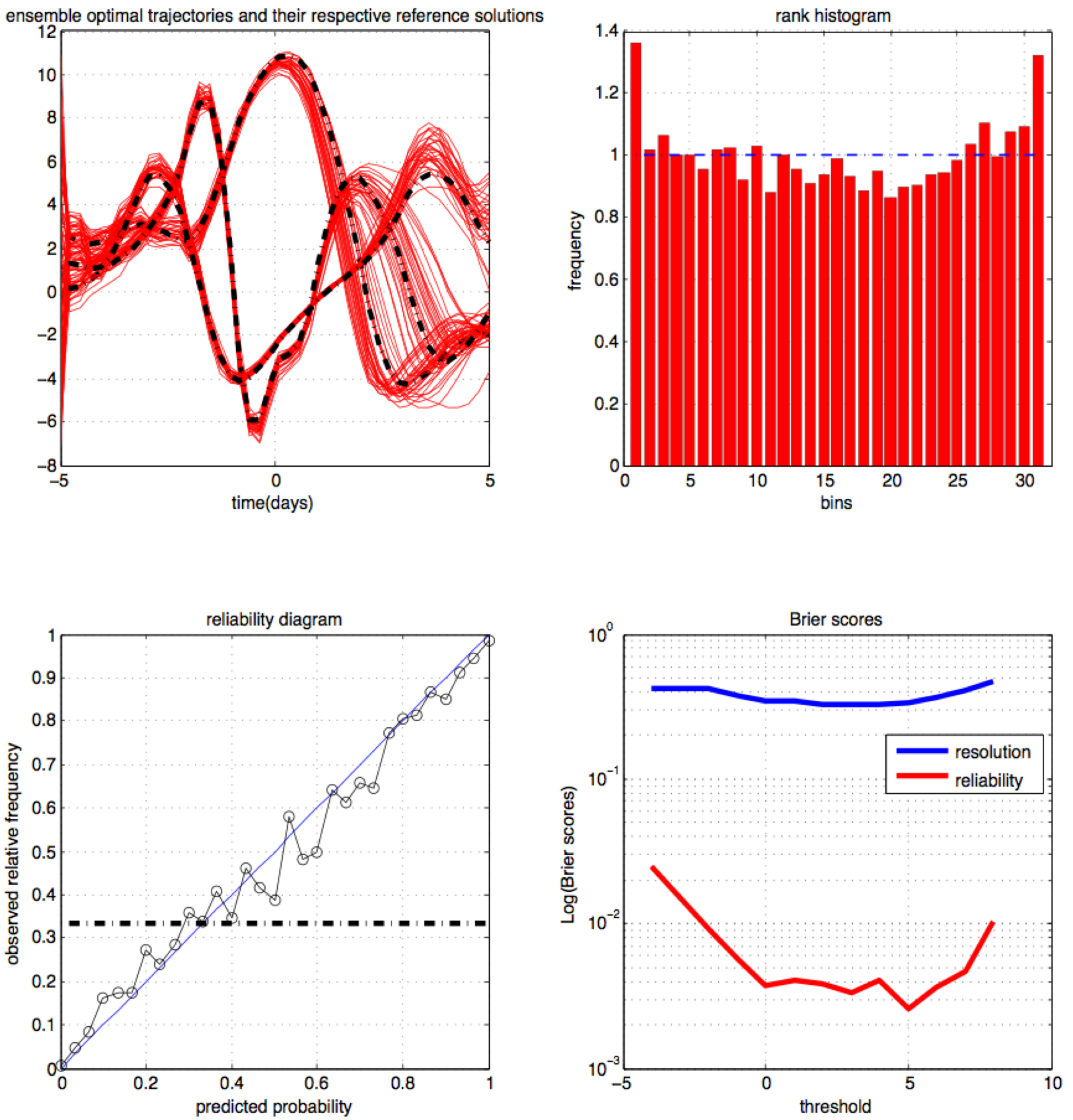
Nonlinear Lorenz’96. EnKF. Diagnostics after 5 days of assimilation



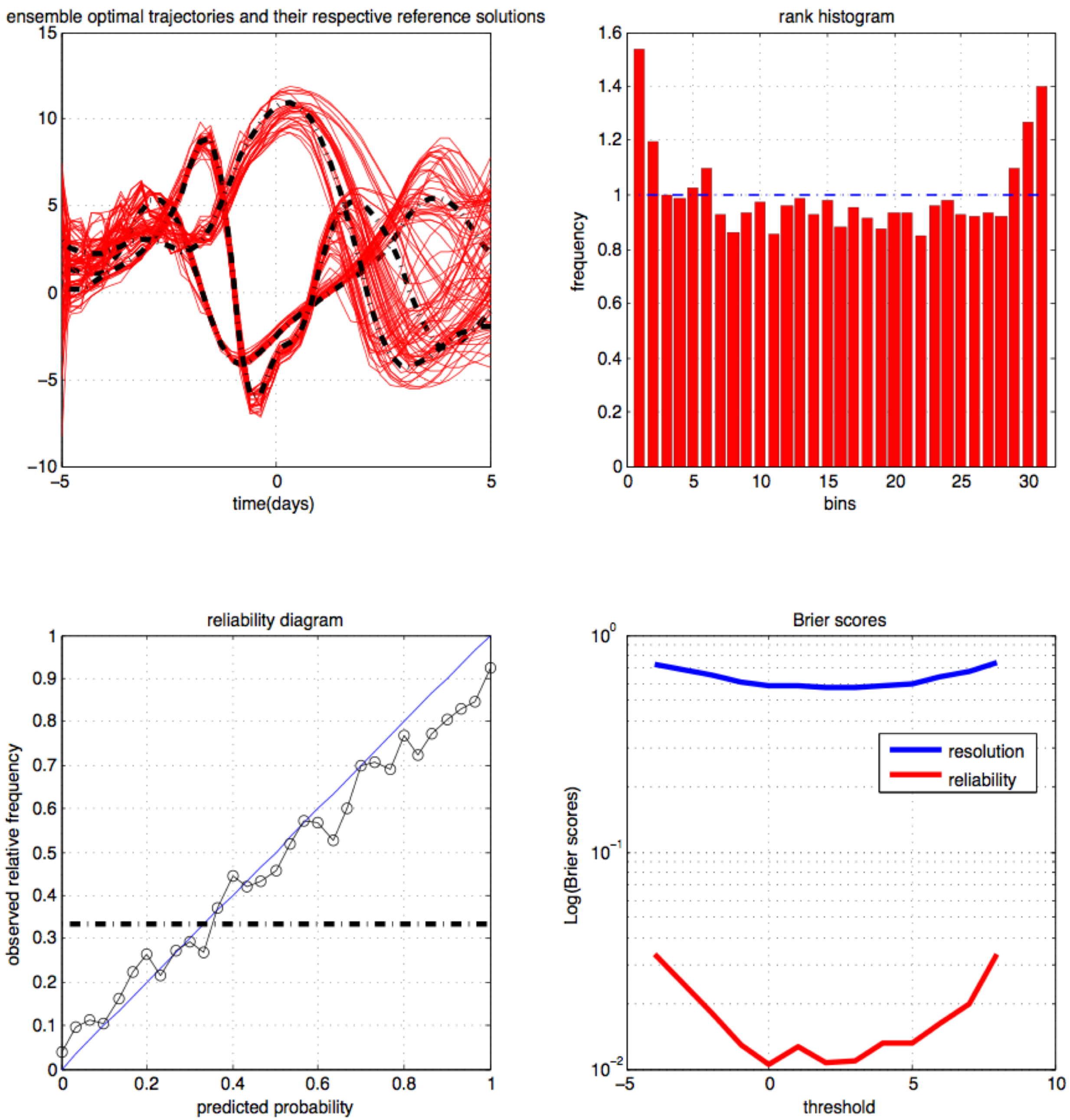
Nonlinear Lorenz’96. PF. Diagnostics after 5 days of assimilation



EnsVAR. Diagnostics for 5-day forecasts



EnKF. Diagnostics for 5-day forecasts



PF. Diagnostics for 5-day forecasts

DA procedure / *method*	*Assimilation*	Forecasting
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

From course 6

Weak constraint variational assimilation

Allows for errors in the assimilating model

Data

- Background estimate at time 0

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

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

$y_k = H_k x_k + \varepsilon_k \qquad E(\varepsilon_k \varepsilon_{k'}^T) = R_k \delta_{kk'}$

- Model

$x_{k+1} = M_k x_k + \eta_k \qquad E(\eta_k \eta_{k'}^T) = Q_k \delta_{kk'} \qquad k = 0, \ldots, K\text{-}1$

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

In the present case, objective function of the form

$$(\xi_0, \eta_1, \ldots, \eta_{K-1}) \rightarrow$$

$$\mathcal{J}(\xi_0, \eta_1, \ldots, \eta_{K-1})$$

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

$$+ (1/2)\ \Sigma_{k=0,\ldots,K-1}\, \eta_k^T Q_k^{-1} \eta_k$$

subject to

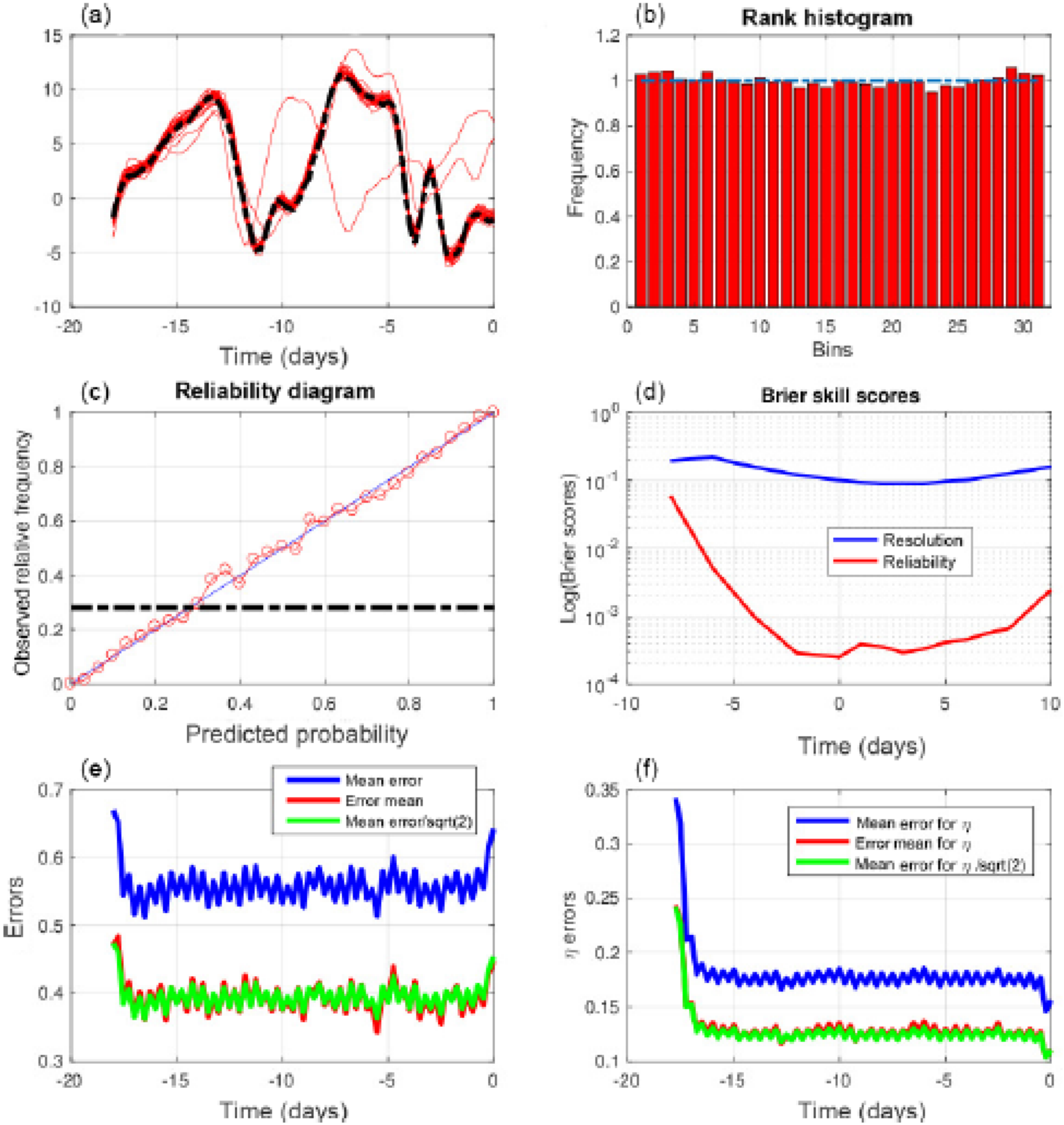
$$\xi_{k+1} = M_k(\xi_k) + \eta_k\ , \qquad k = 0, \ldots, K-1$$

'Observations' consist of

- sequence $\{y_k\}$, $k = 0, \ldots, K$ (with unit observation operator H_k)
- observations 0 for η_k , $k = 0, \ldots, K-1$

It turns out that QSVA is no more necessary. The model error term in the objective function has a regularizing effect which makes the function much smoother.

Weak-constraint ensemble variational assimilation
18 days, Q = 0.1
1200 realizations



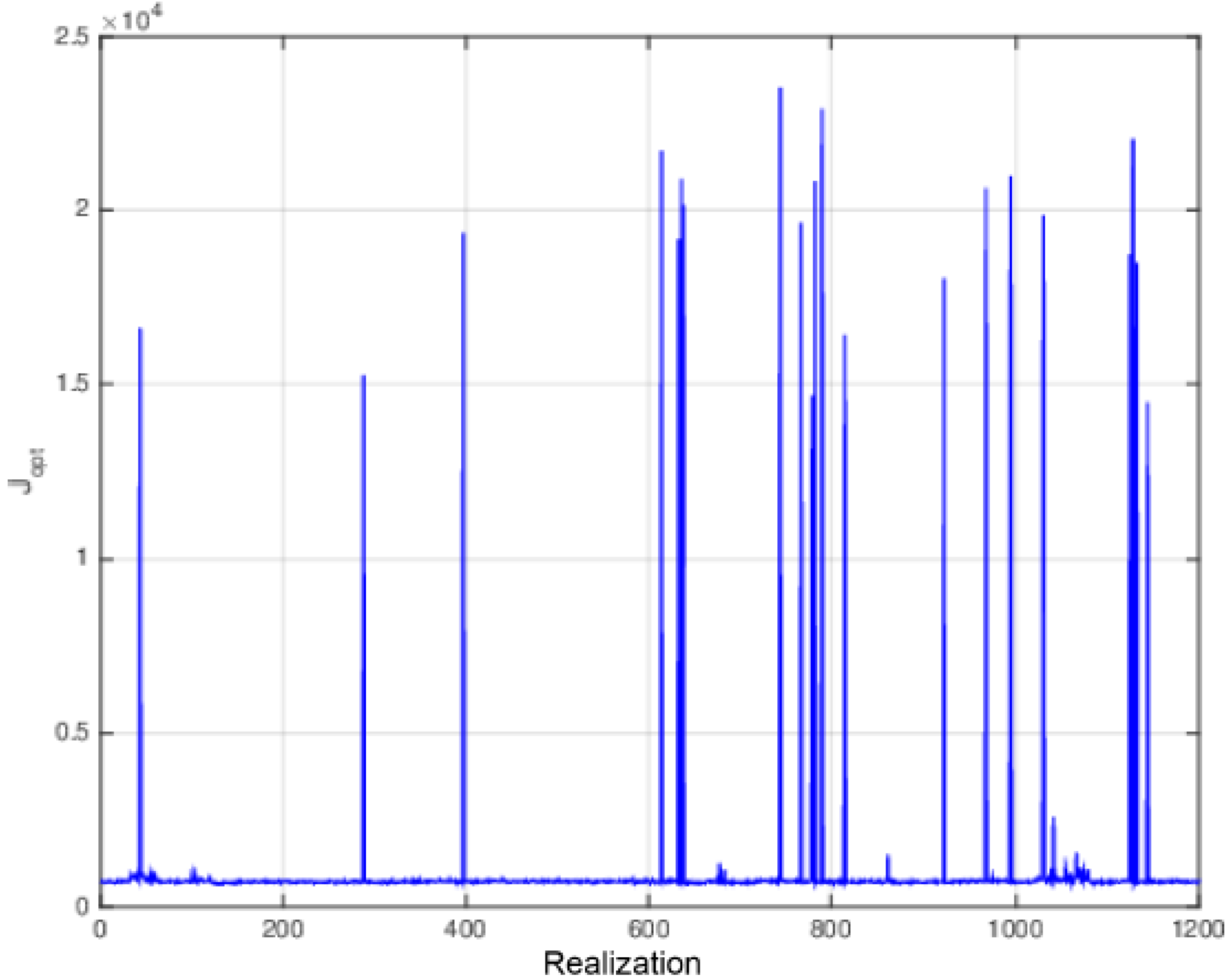
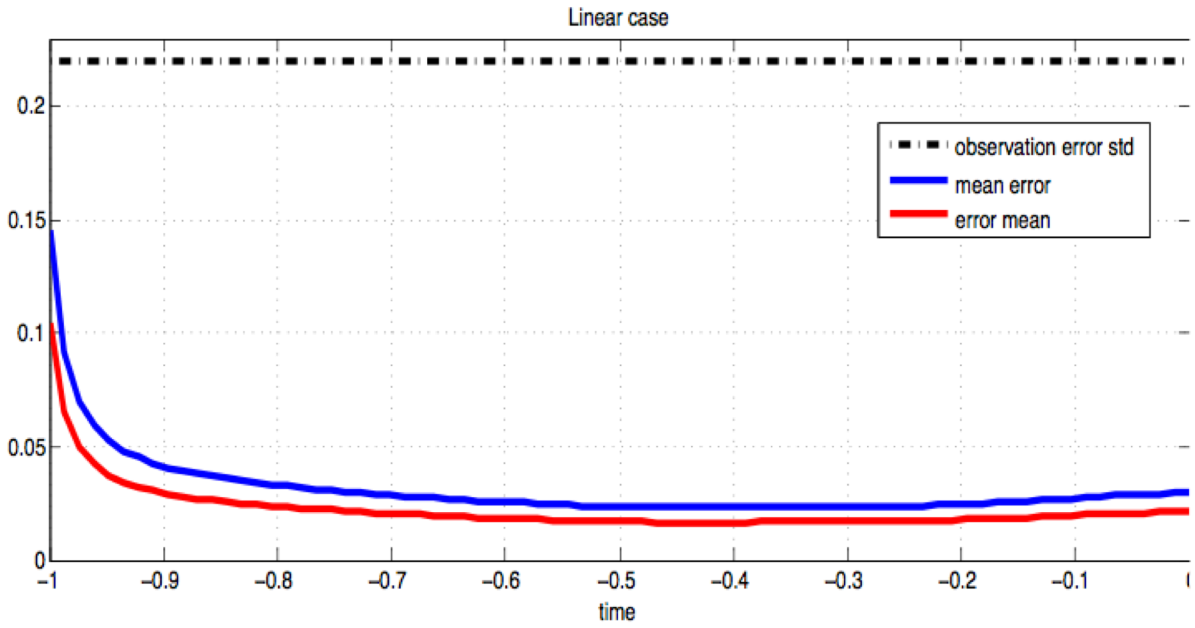


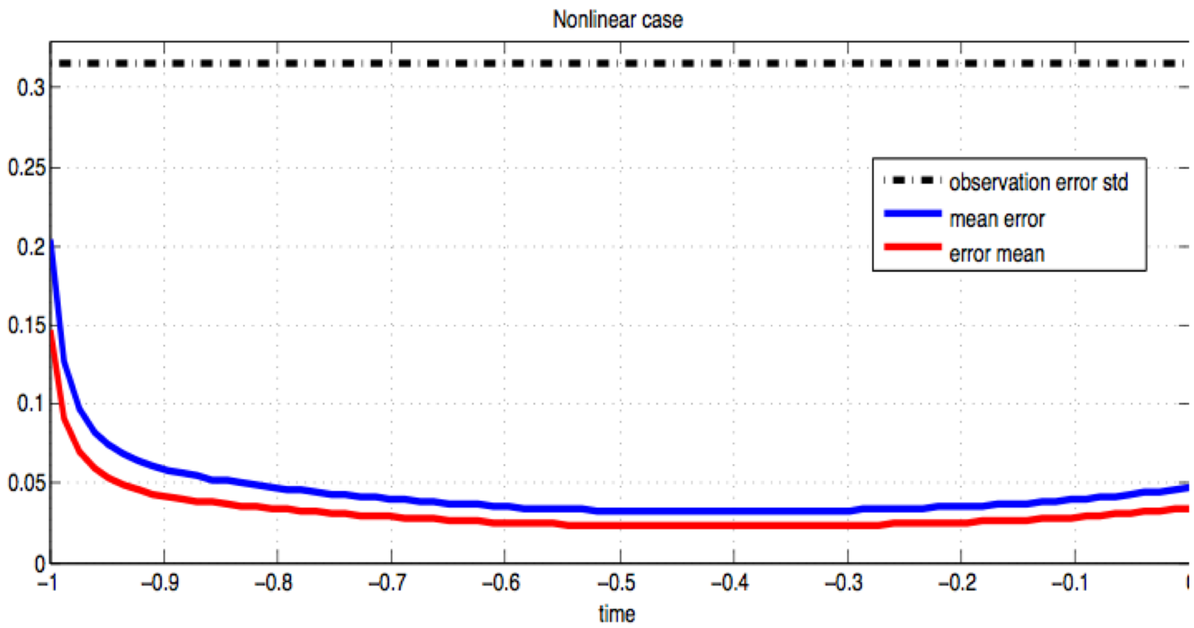
Figure 11. Values of (half) the minima of the objective function for all realizations of the weak-constraint assimilations over 18-day windows.

Kuramoto-Sivashinsky equation

$$\frac{\partial u}{\partial t} + \frac{\partial^4 u}{\partial x^4} + \frac{\partial^2 u}{\partial x^2} + u\frac{\partial u}{\partial x} = 0, \quad x \in [0, L]$$

with periodicity in x, $L = 32\pi$


Linear case
observation error std
mean error
error mean
0.2
0.15
0.1
0.05
0
−1
−0.9
−0.8
−0.7
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
time


Nonlinear case
observation error std
mean error
error mean
0.3
0.25
0.2
0.15
0.1
0.05
0
−1
−0.9
−0.8
−0.7
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
time

Summary

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 Kuramuto-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).

Cours à venir

~~Mardi 21 mars~~
~~Mardi 28 mars~~
~~Mardi 4 avril~~
~~Mardi 11 avril~~
~~Mardi 2 mai~~
~~Mardi 9 mai~~
~~Mardi 23 mai~~
Mardi 30 mai