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

Année Universitaire 2020-2021

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 8

28 Mai 2021

- Particle filters. Principle. Variants (*Proposal Densities*). A few results.

- Ensemble Variational Assimilation. A few results.

- Conclusions and perspectives

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

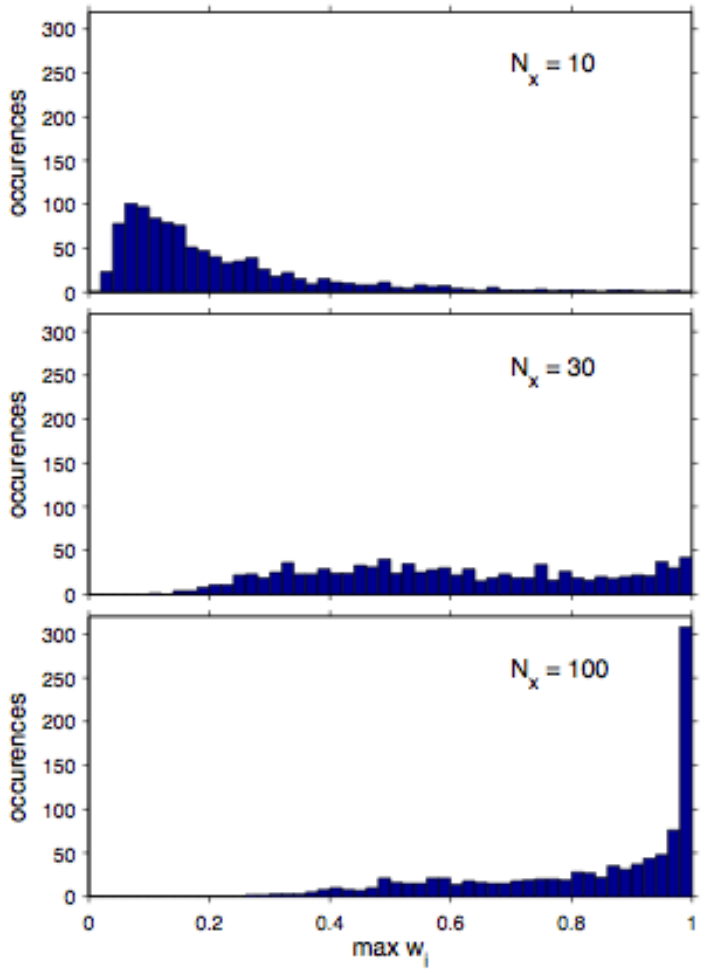
$$P(y|x^b_l) = P[\varepsilon = y - H(x^b_l)]$$

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 $= 5.5$

$\ldots = 25$

$\ldots = 127$

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.

Curse of dimensionality

Standard one-dimensional gaussian random variable X

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

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

.

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.

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

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

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

Chapman-Kolmogorov equation

Particular case

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

$y^n = H_n\, x^n + \varepsilon_n$ $\quad$ 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:

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

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

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

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:

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

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:

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

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

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

with weights w_i given by:

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

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

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

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*, …

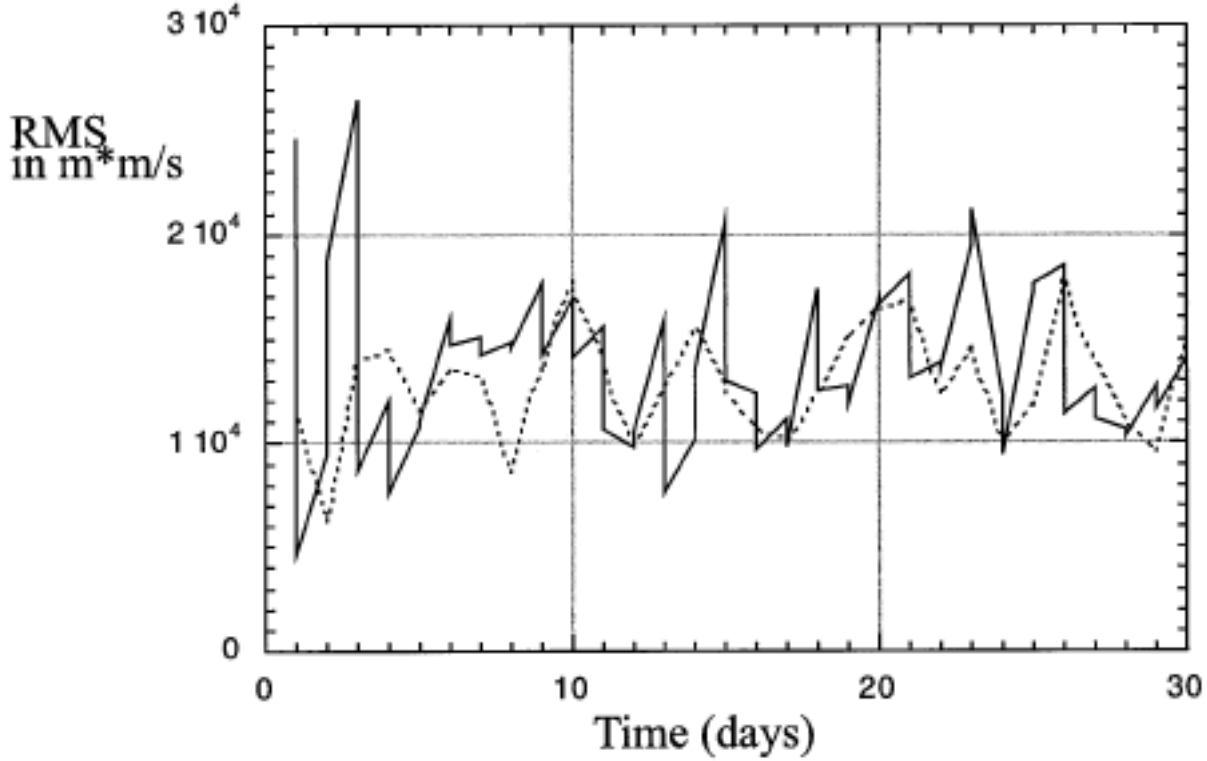


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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 $q(x^n \mid x^{n-1}{}_{1,\mathrm{L}}, y^n)$, 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.

$$\frac{\mathrm{D}(\zeta + f)}{\mathrm{D}\,t} = \eta$$

State-vector dimension $\approx$ 65,000

Decorrelation time: 25 timesteps

One complete noisy model field observed every 50 timesteps

24 particles

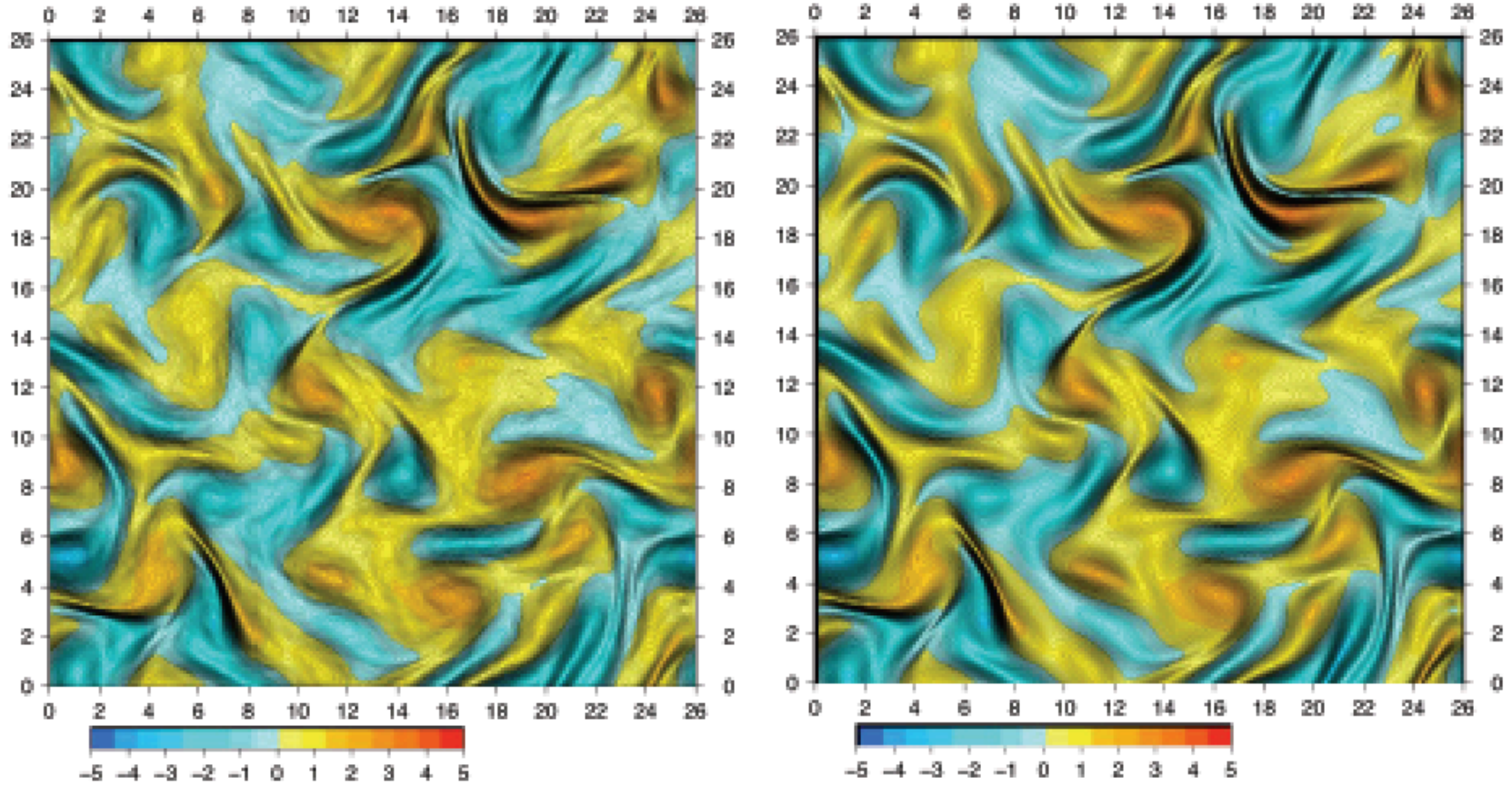


Figure 5.3. Snap shot of the vorticity field of the truth (right) and the particle filter mean (left) at time 25. Note the highly chaotic state of the fields, and the close to perfect tracking.

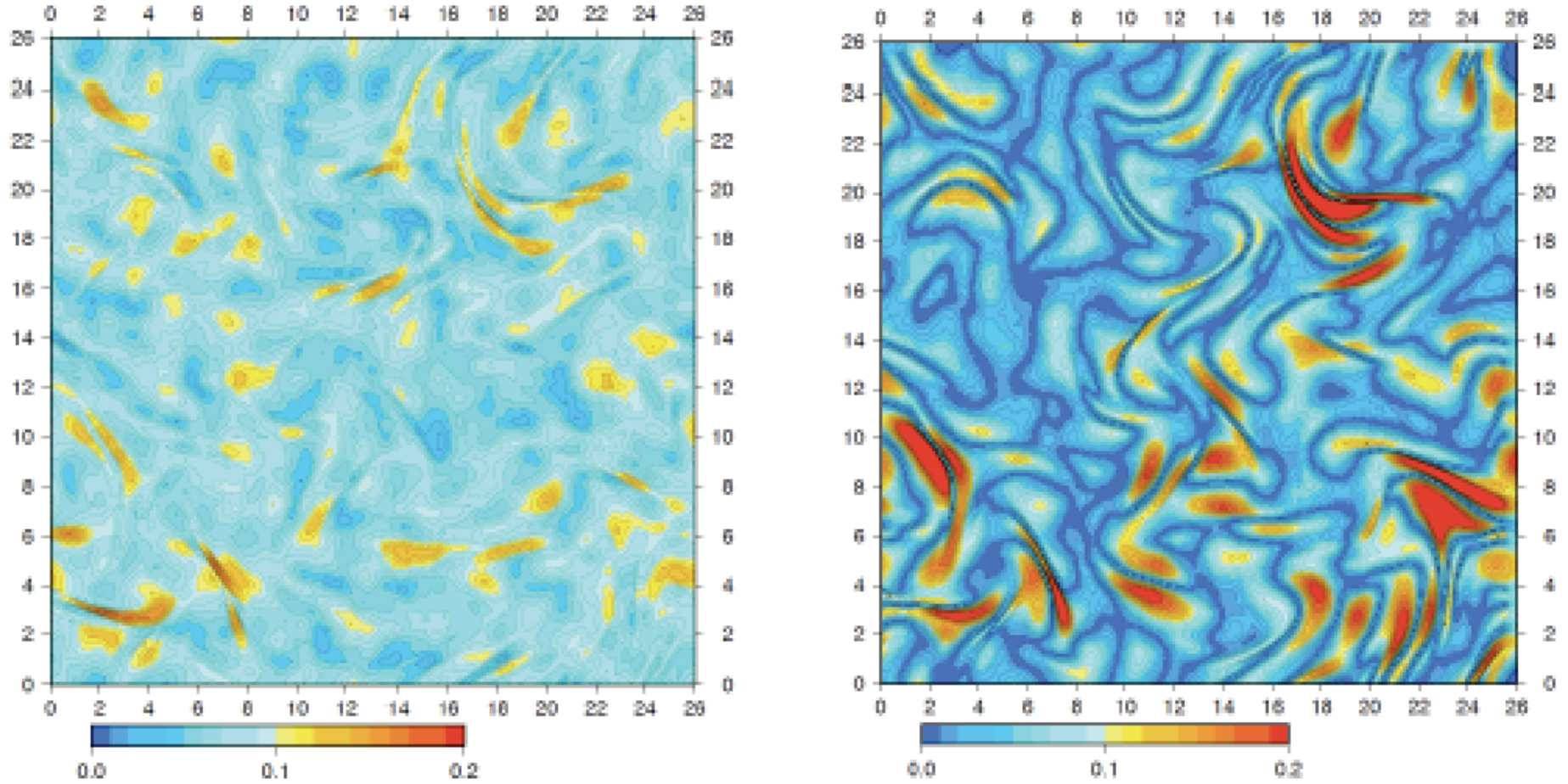


Figure 5.4. Snap shot of the absolute value of the mean-truth misfit and the standard deviation in the ensemble. The ensemble underestimates the spread at several locations, but averaged over the field it is slightly higher, 0.074 versus 0.056.

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

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

Variational form

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

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

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

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

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

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

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.

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

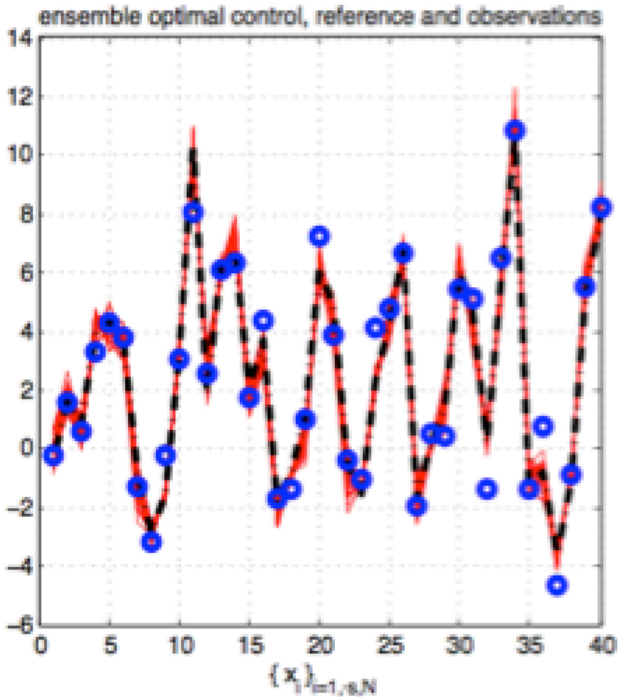
- Forward model

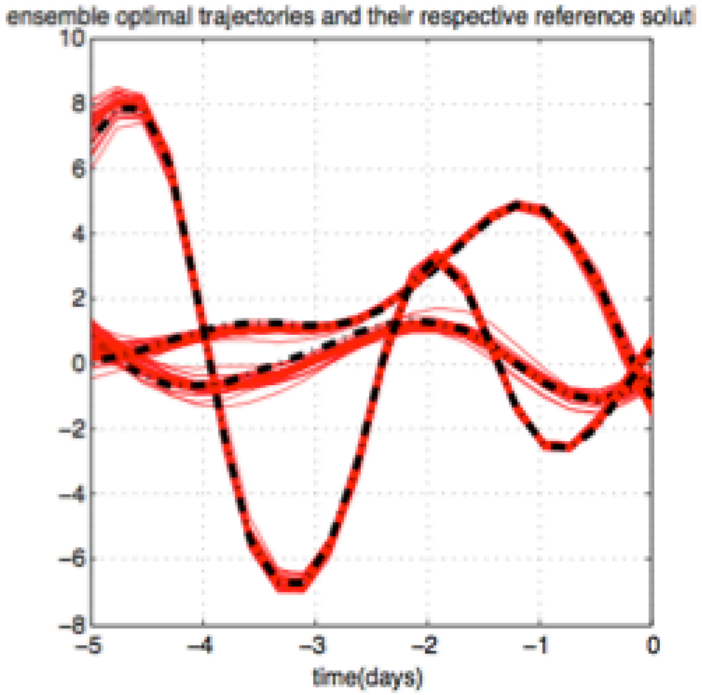
$$\frac{dx_k}{dt} = (x_{k+1} - x_{k-2})x_{k-1} - x_k + F \quad \text{for} \quad k = 1, \cdots, N$$

- 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. $\frac{1}{\lambda_{max}} \simeq 2.5 days$, λ_{max} the largest Lyapunov exponent.
  7. $\Delta t = 0.05 = 6 hours$, 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

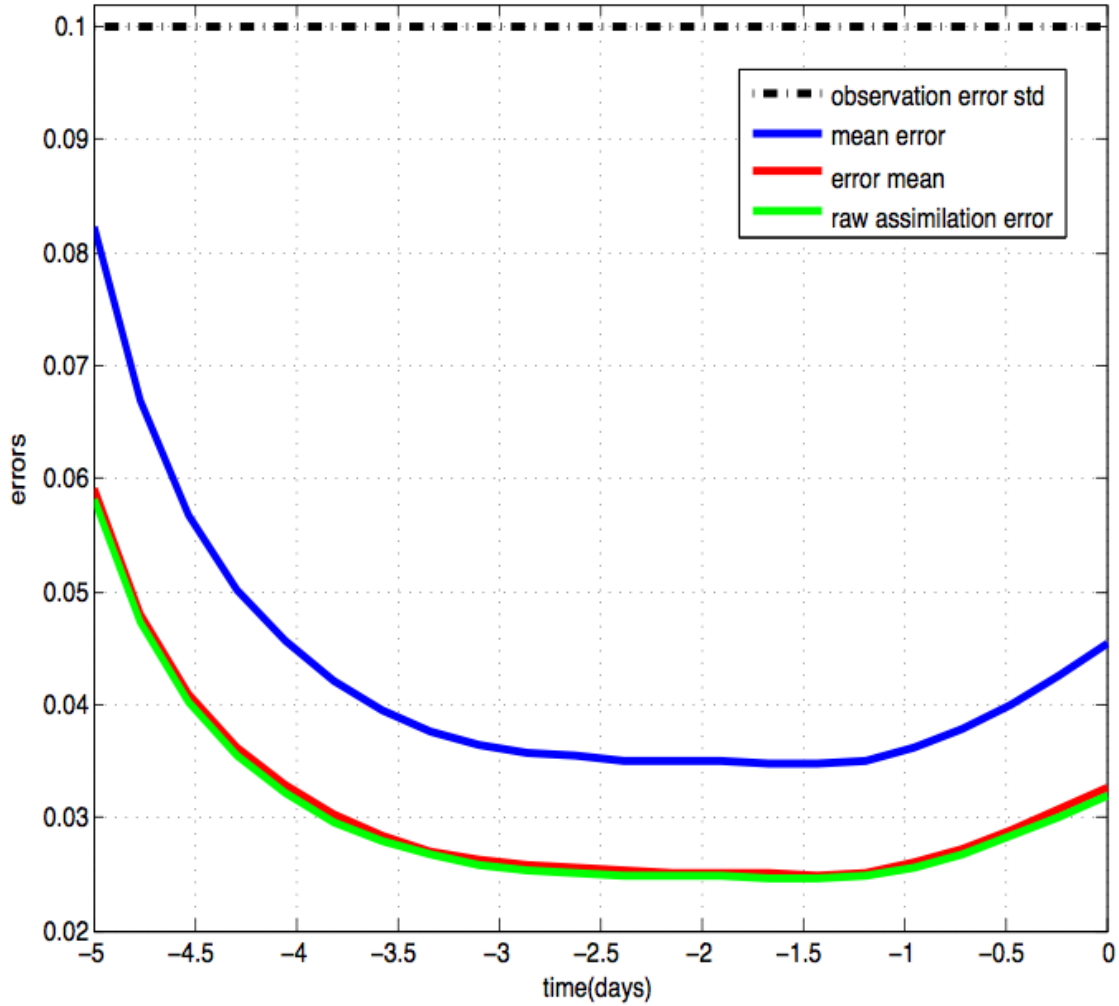
$$y_k \rightarrow \; z_k = y_k + \delta_k$$

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

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

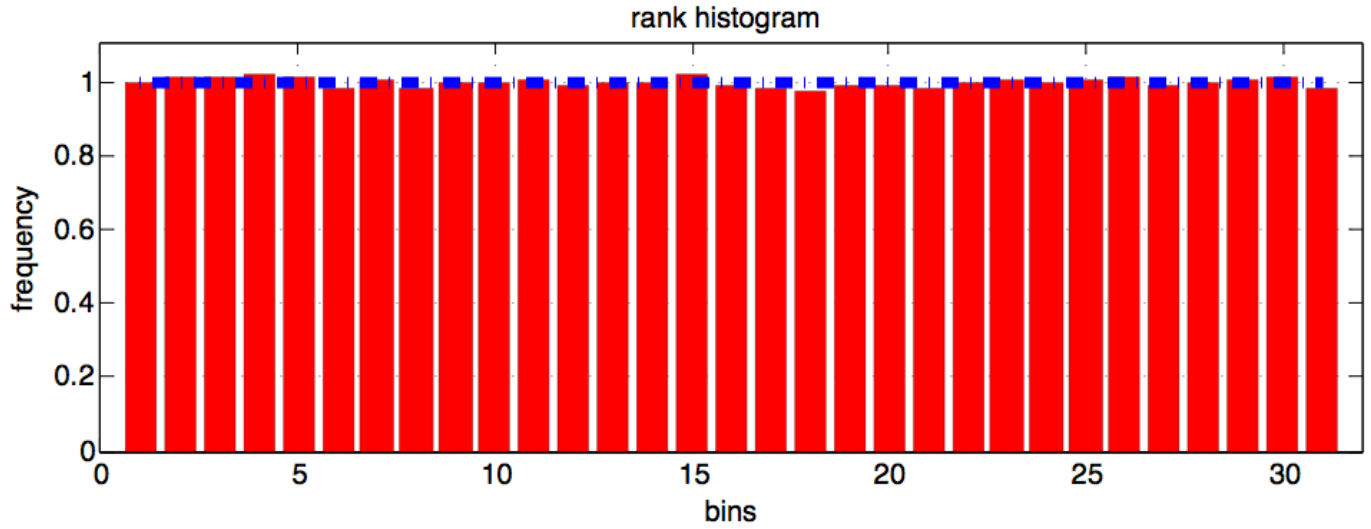
Reliability can be objectively validated, provided a large enough sample of realizations of the estimation system is available.

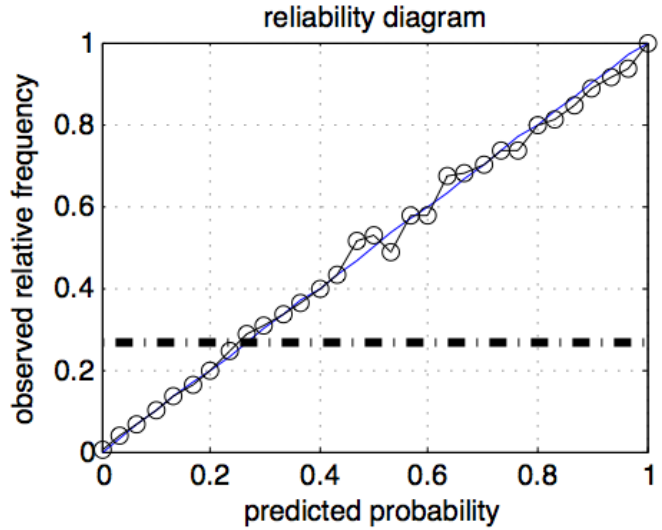
Bayesianity implies reliability, the converse not being true.

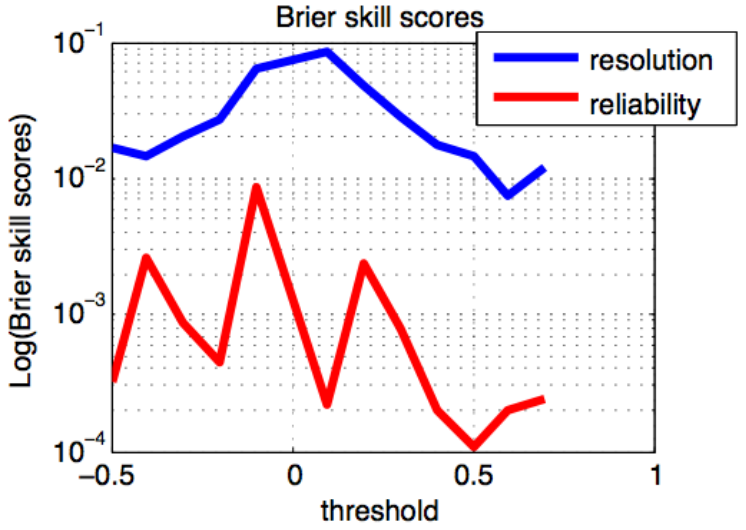
In addition, we evaluate *resolution* (also called *sharpness*), which bears no direct relation to bayesianity, and is the capability of the estimation system to *a priori* distinguish between different situations. It is best defined as the degree of statistical dependence between X and Y (J. Bröcker).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

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

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







Linearized Lorenz'96. 5 days

Objective function

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

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

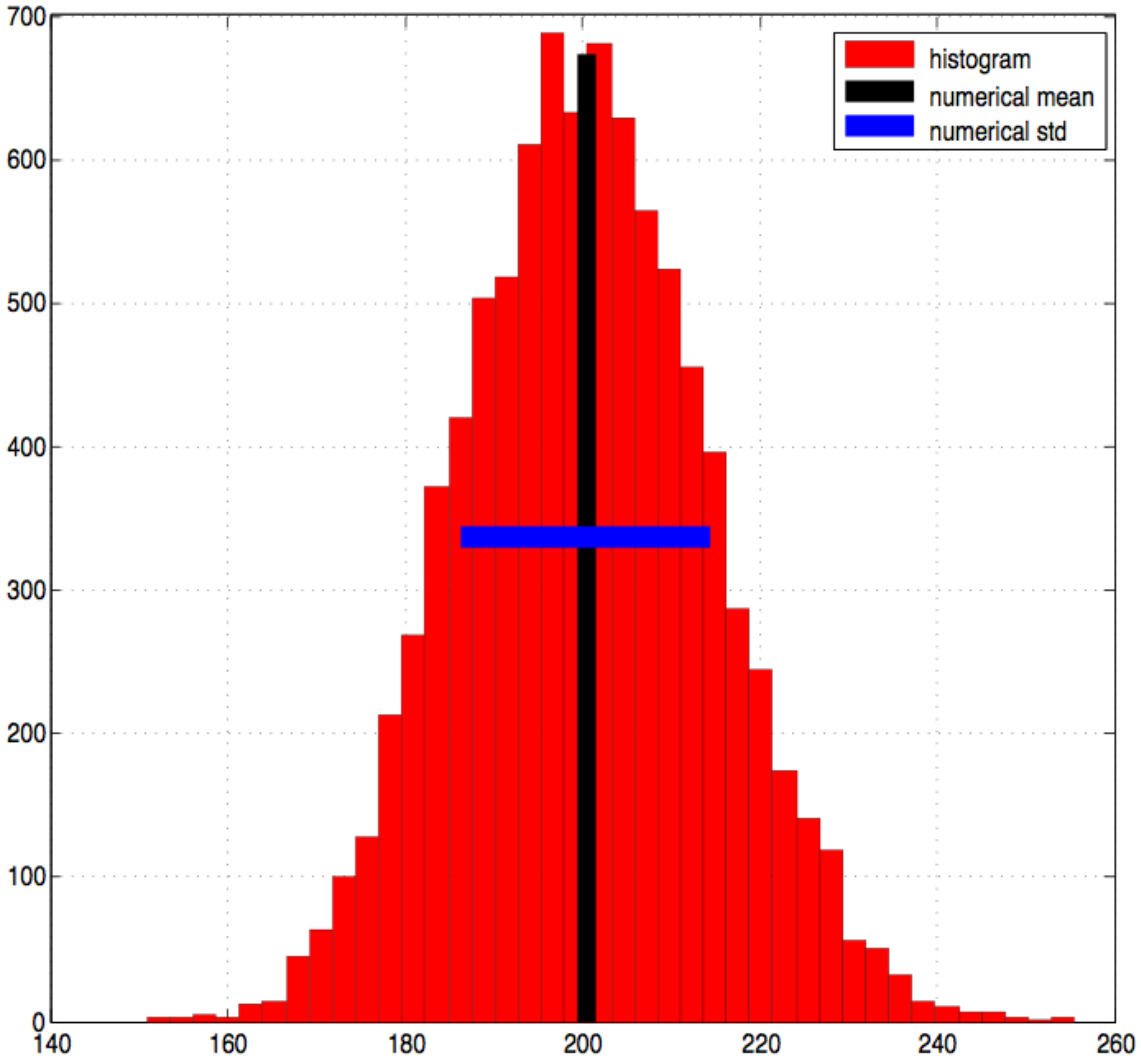
$$= (1/2)\,\boldsymbol{d}^{\mathrm{T}}\,[E(\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}})]^{-1}\,\boldsymbol{d}$$

where $\boldsymbol{d}$ is innovation

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

If p is large, a few realizations are sufficient for determining $E(\mathcal{J}_{min})$
Often called χ^2 criterion.

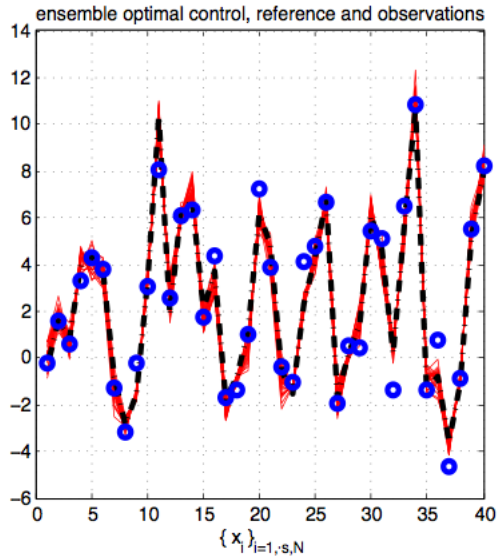
Remark. If in addition errors are gaussian $Var(\mathcal{J}_{min}) = p/2$

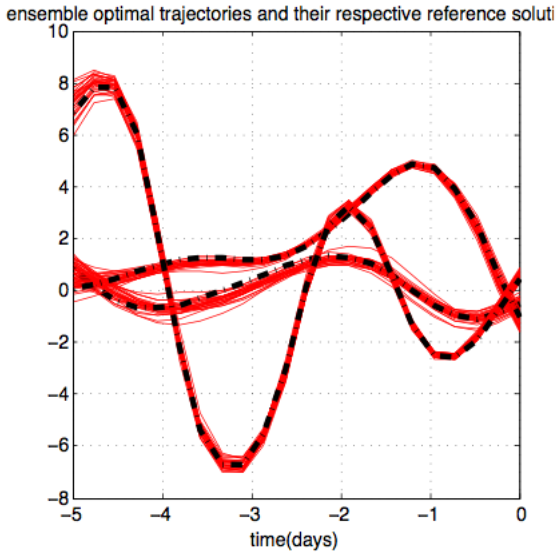


Linearized Lorenz'96. 5 days. Histogram of $\mathcal{J}_{min}$

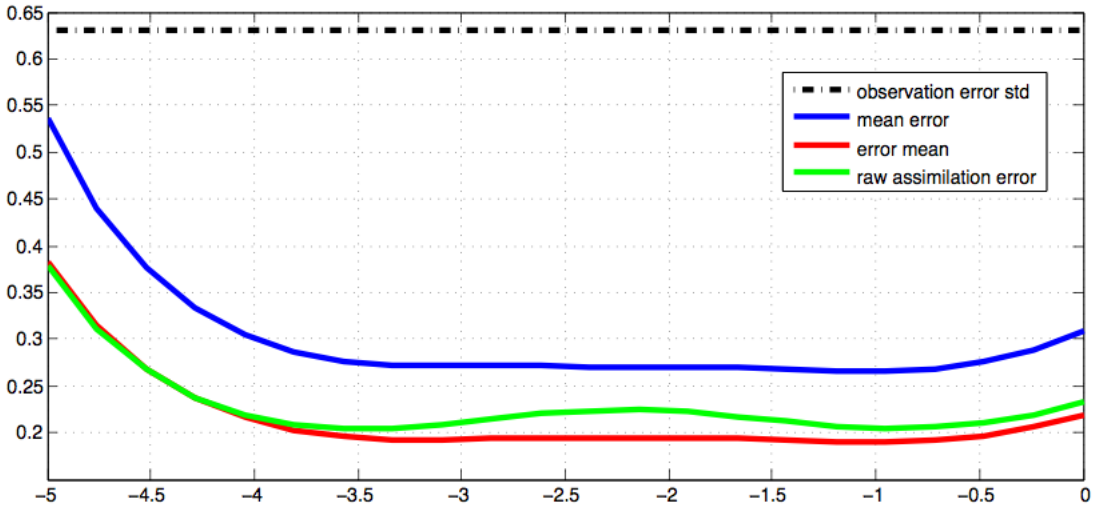
$E(\mathcal{J}_{min}) = p/2$ (=200) ; $\sigma(\mathcal{J}_{min}) = \sqrt{(p/2)}$ (≈14.14)

Observed values 199.39 and 14.27

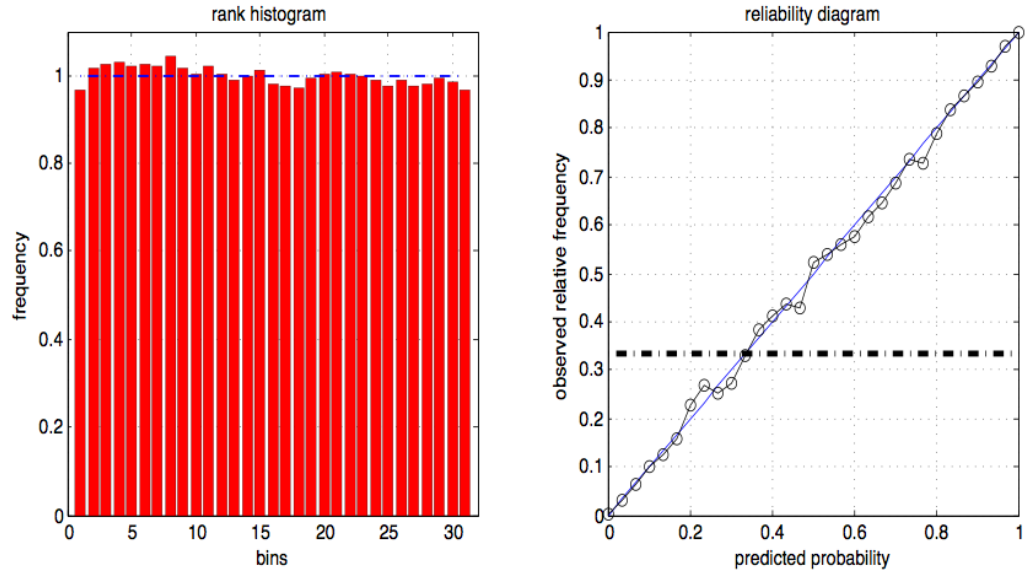


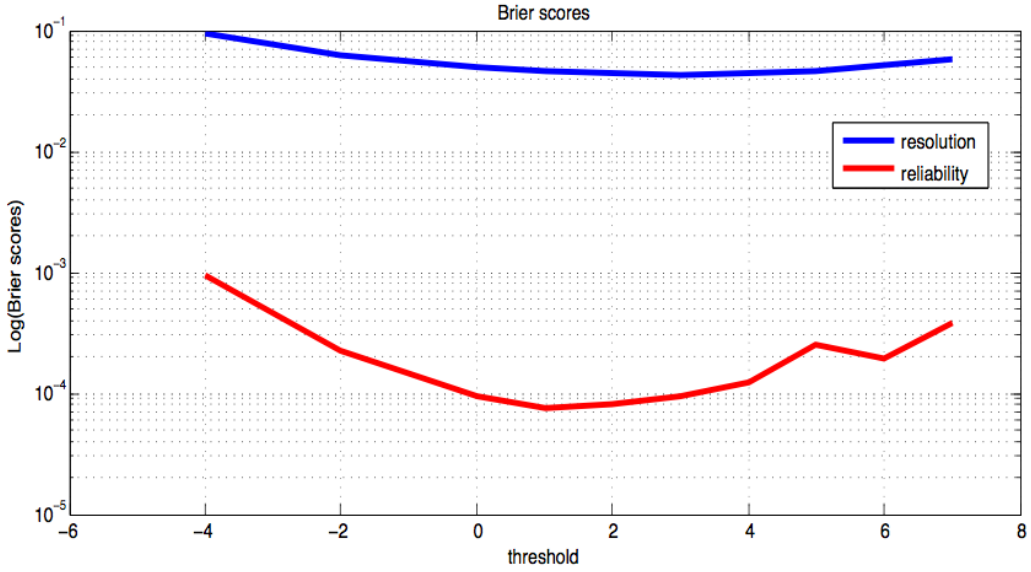


(initial time of assimilation window)

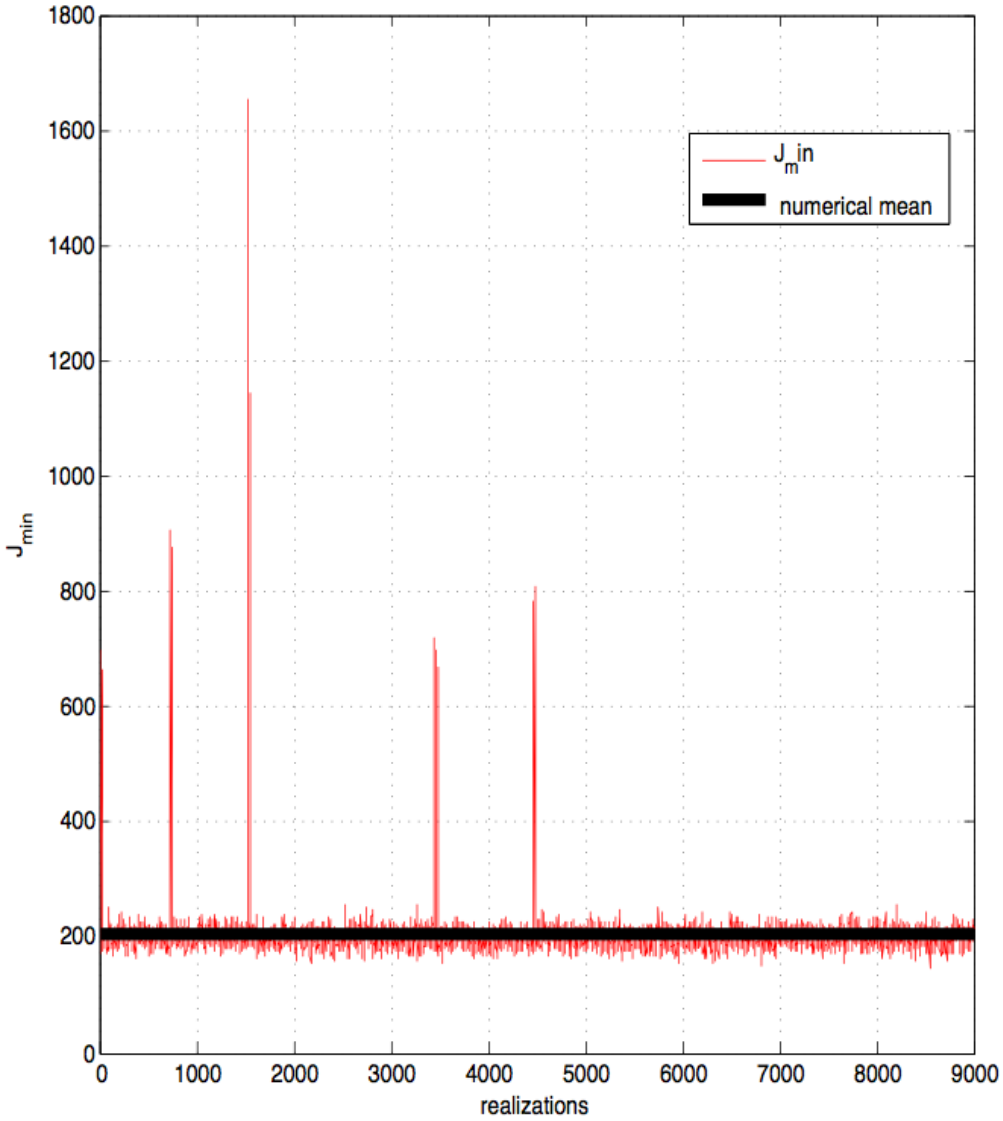


Nonlinear Lorenz'96. 5 days



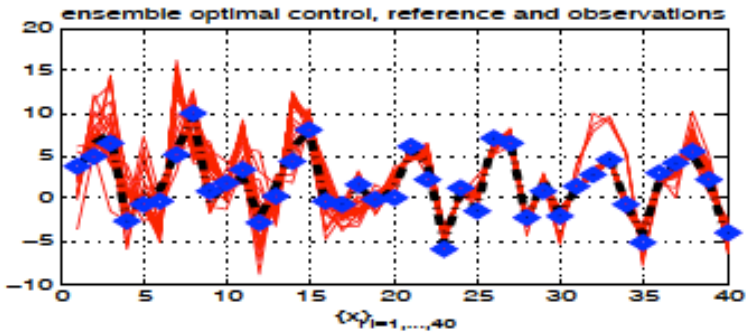


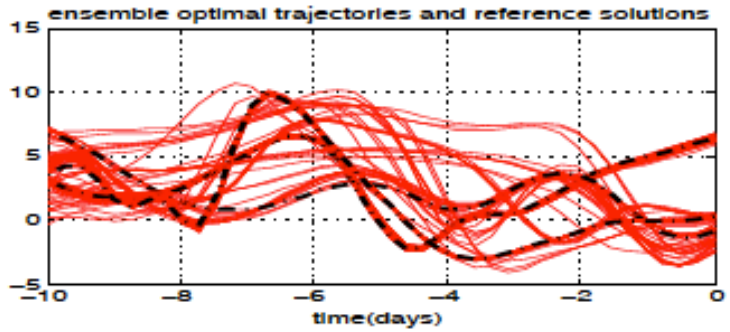
Nonlinear Lorenz'96. 5 days

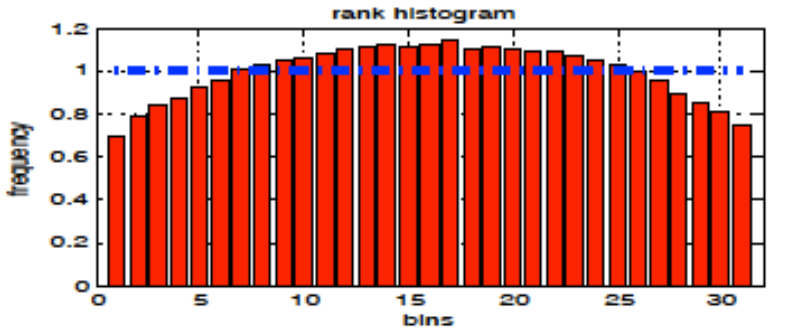


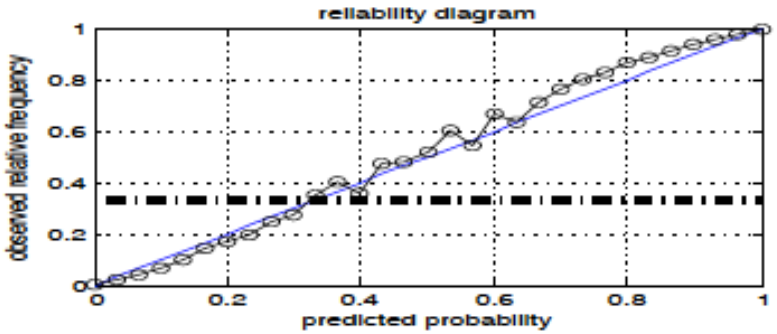
Nonlinear Lorenz'96. 5 days. Histogram of $\mathcal{J}_{min}$

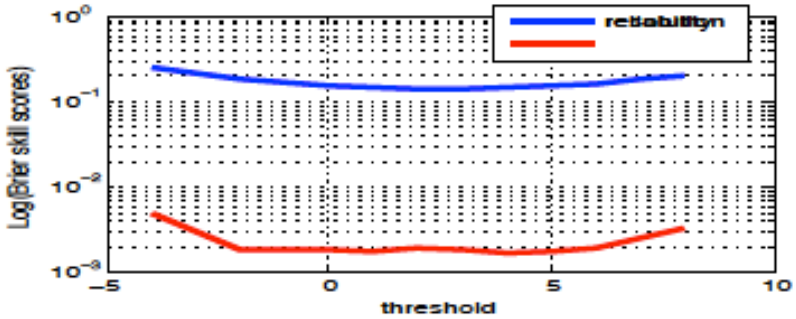
EnsVar : the non-linear Lorenz96 model (10 days $\simeq$ 2 TU)

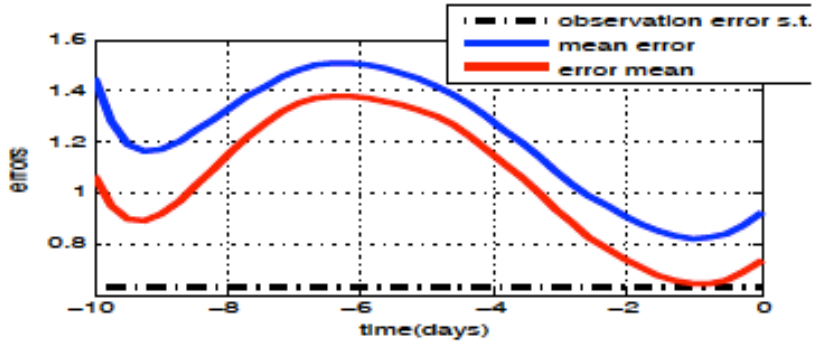




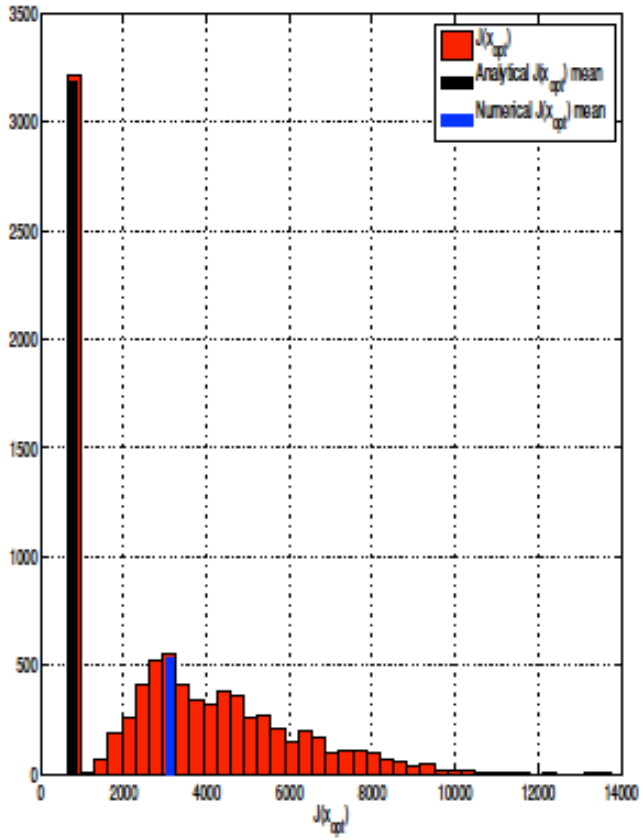






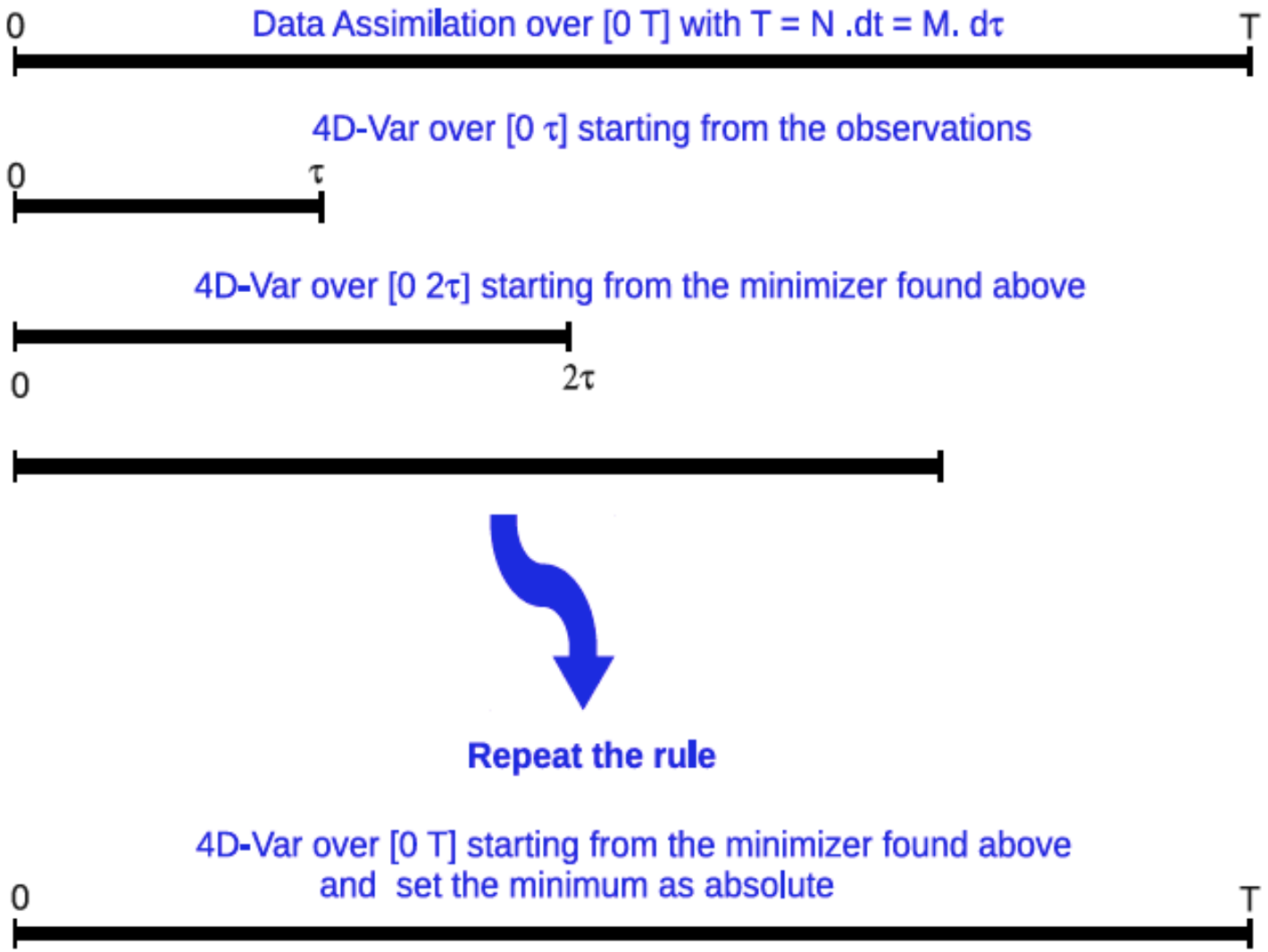


EnsVar : consistency

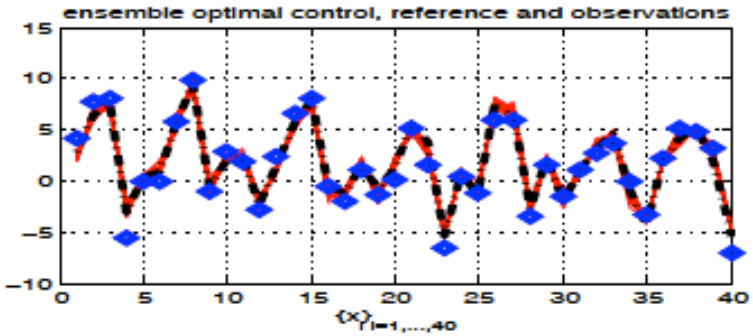


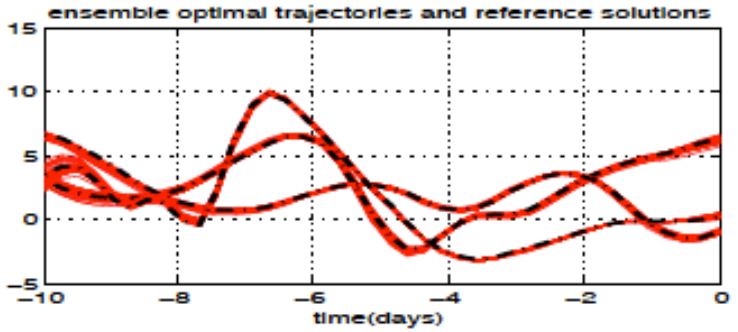
Nonlinear Lorenz'96. 10 days. Histogram of $\mathcal{J}_{min}$

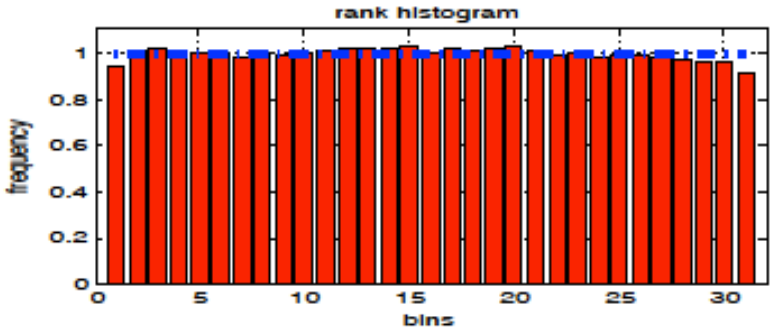
Quasi-Static Variational Assimilation (QSVA)

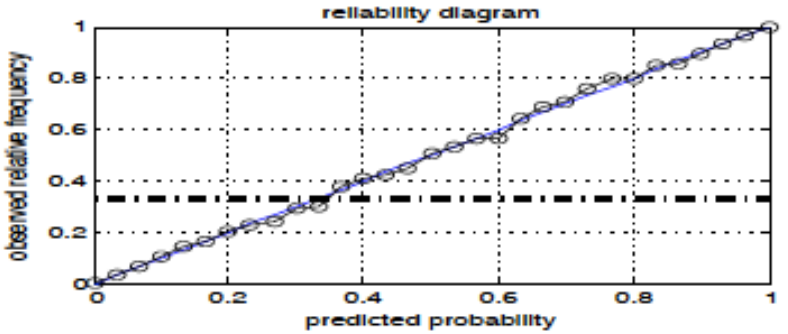


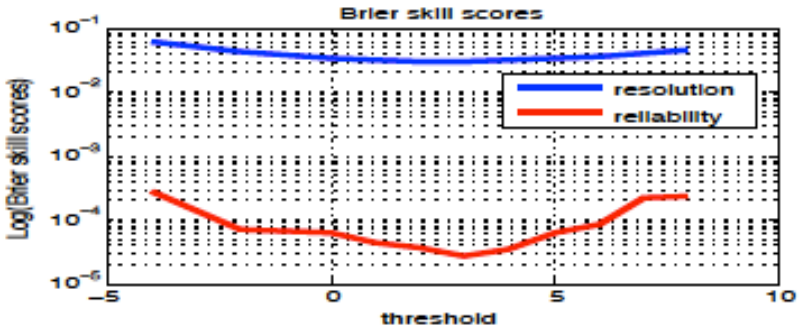
EnsVar : the non-linear Lorenz96 model 10 days with QSVA

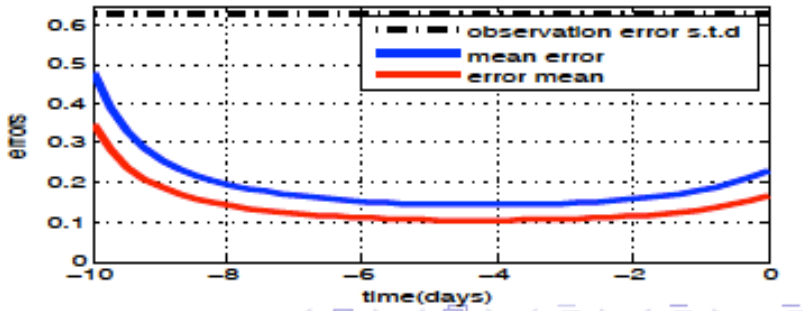




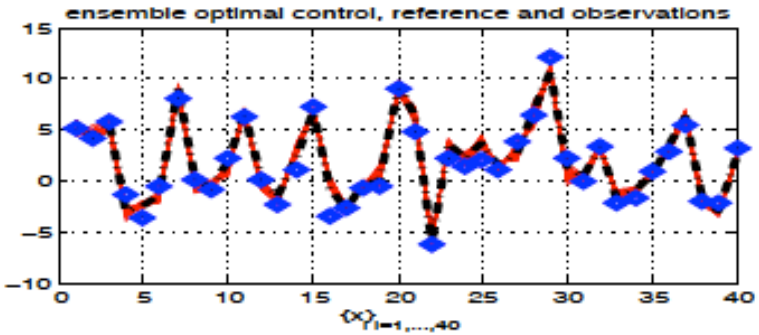


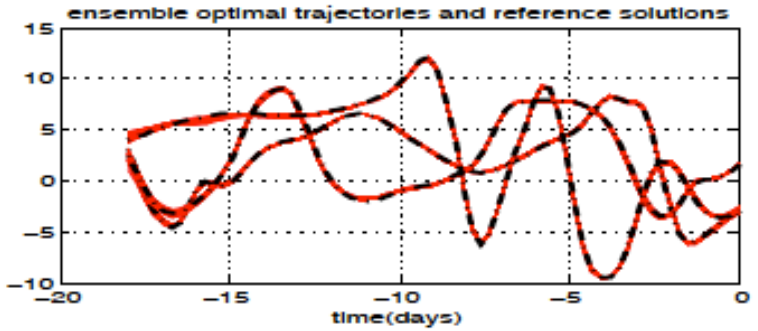


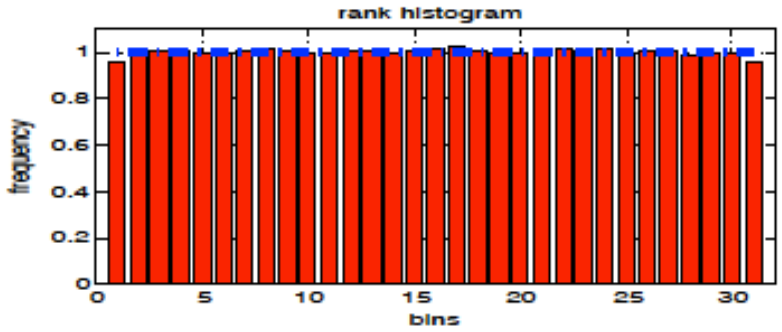


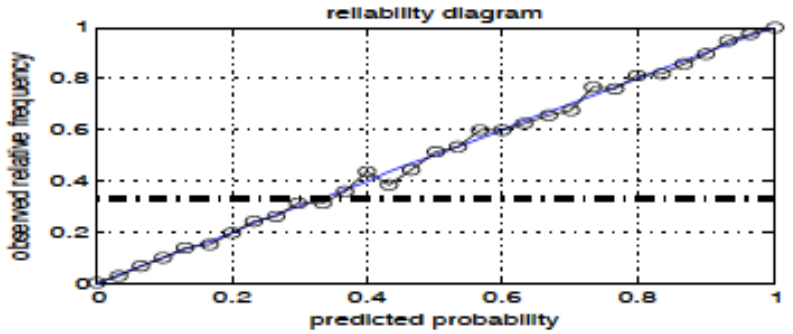


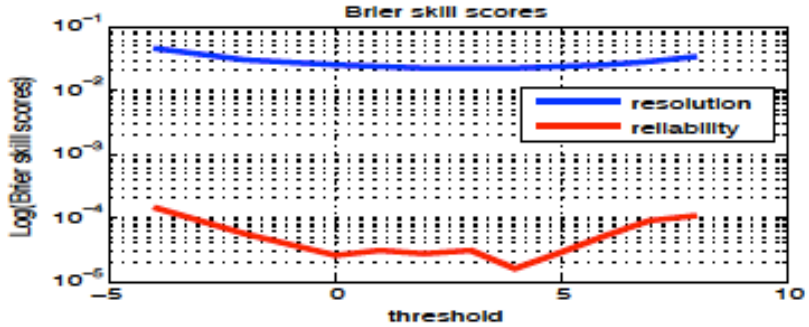
EnsVar : the non-linear Lorenz96 model 18 days with QSVA

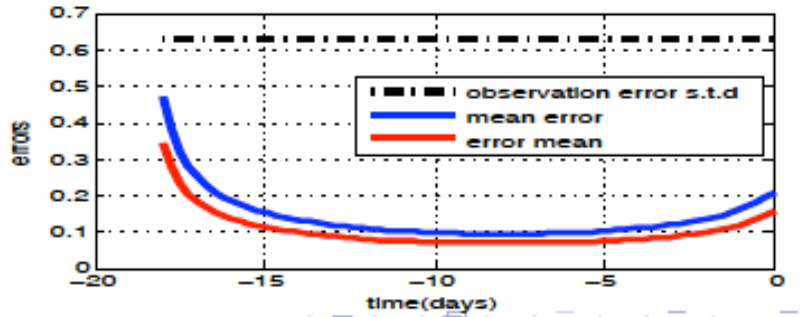










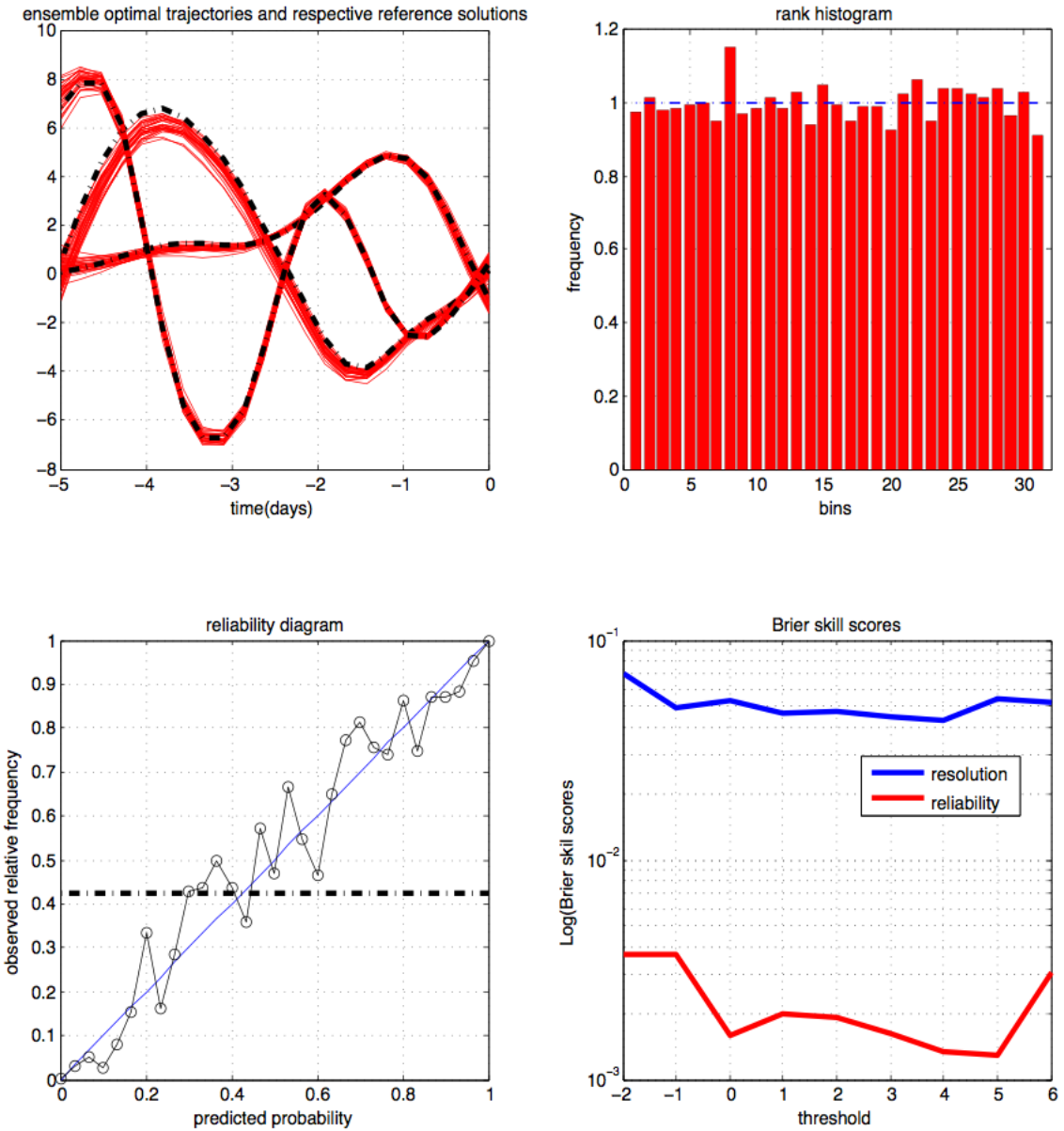


- Results are independent of the Gaussian character of the observation errors (trials have been made with various probability distributions)

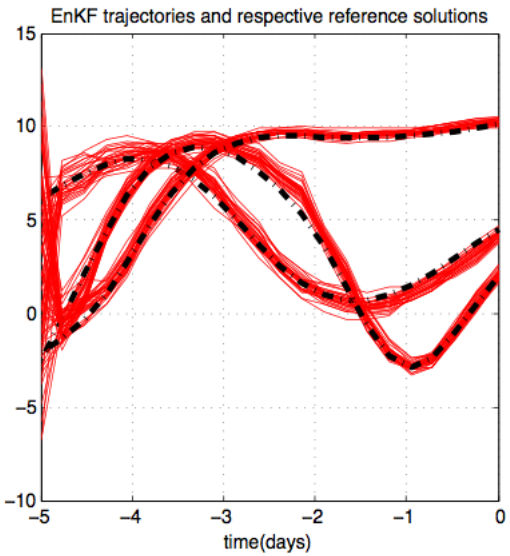
- Ensembles produced by EnsVar are very close to Gaussian, even in strongly nonlinear cases.

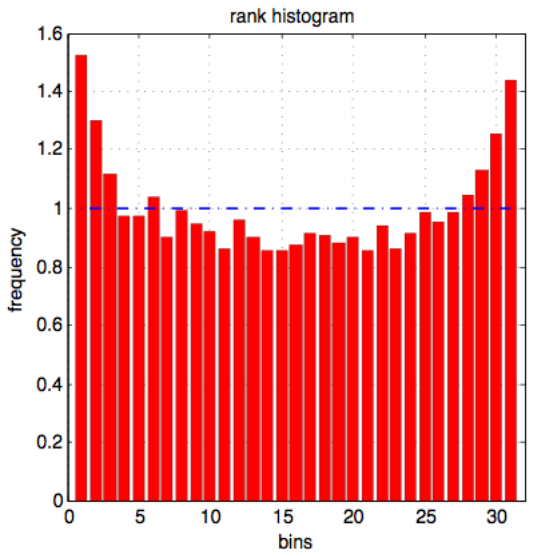
- Comparison *Ensemble Kalman Filter* (*EnKF*) and *Particle Filters* (*PF*)

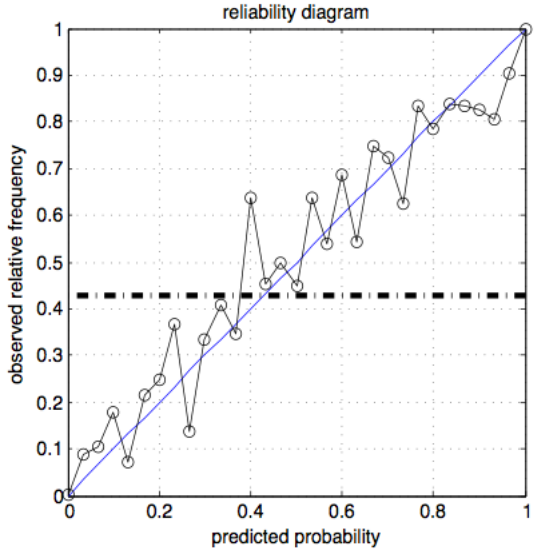
  Both of these algorithms being sequential, comparison is fair only at end of assimilation window

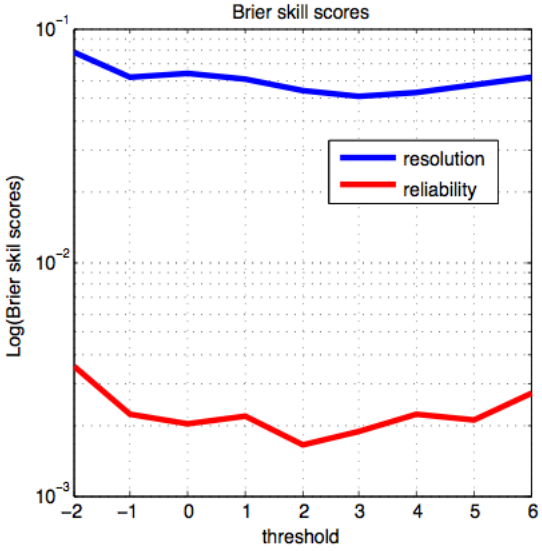


Nonlinear Lorenz'96. 5 days. Diagnostics at end of assimilation window

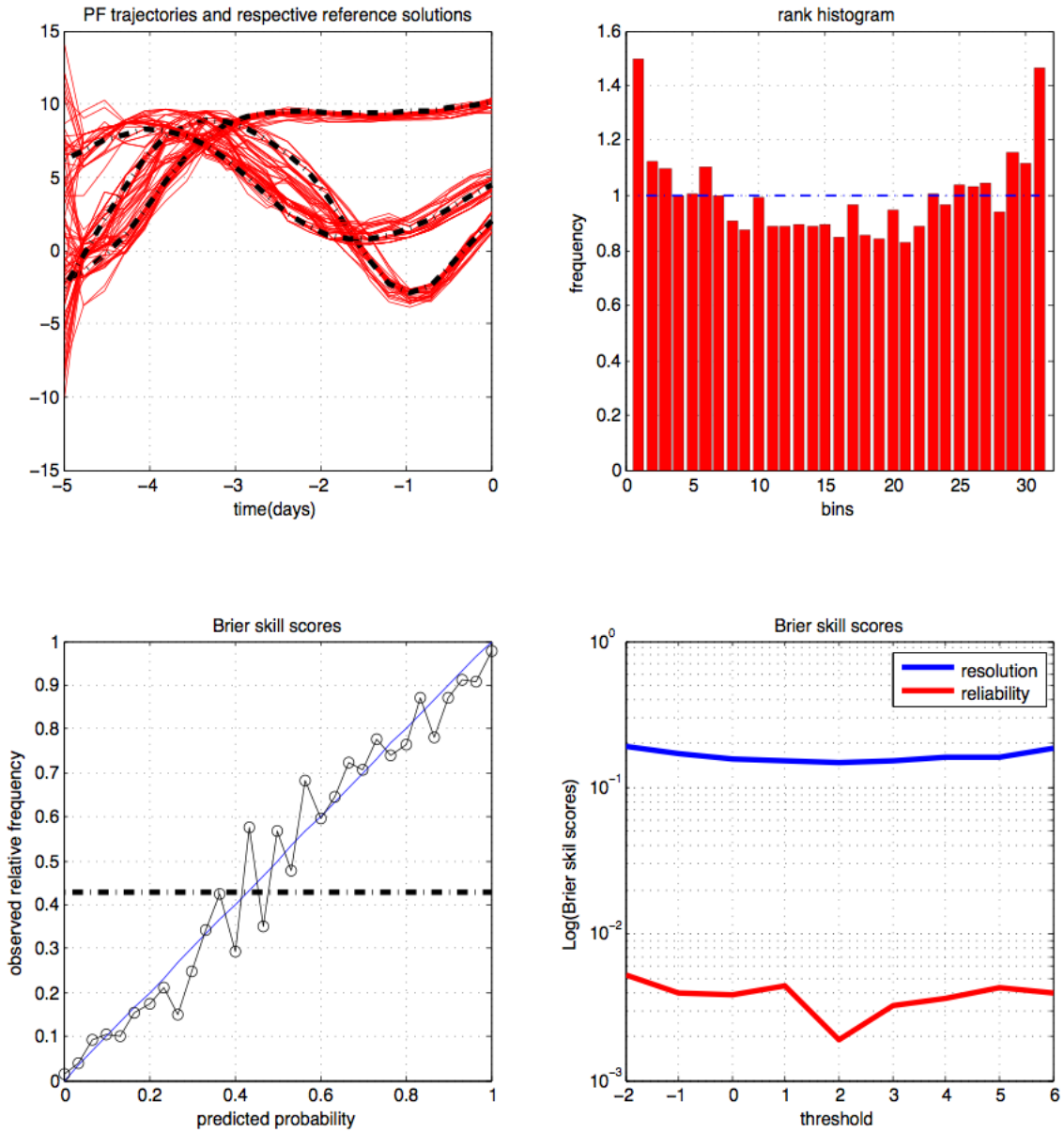




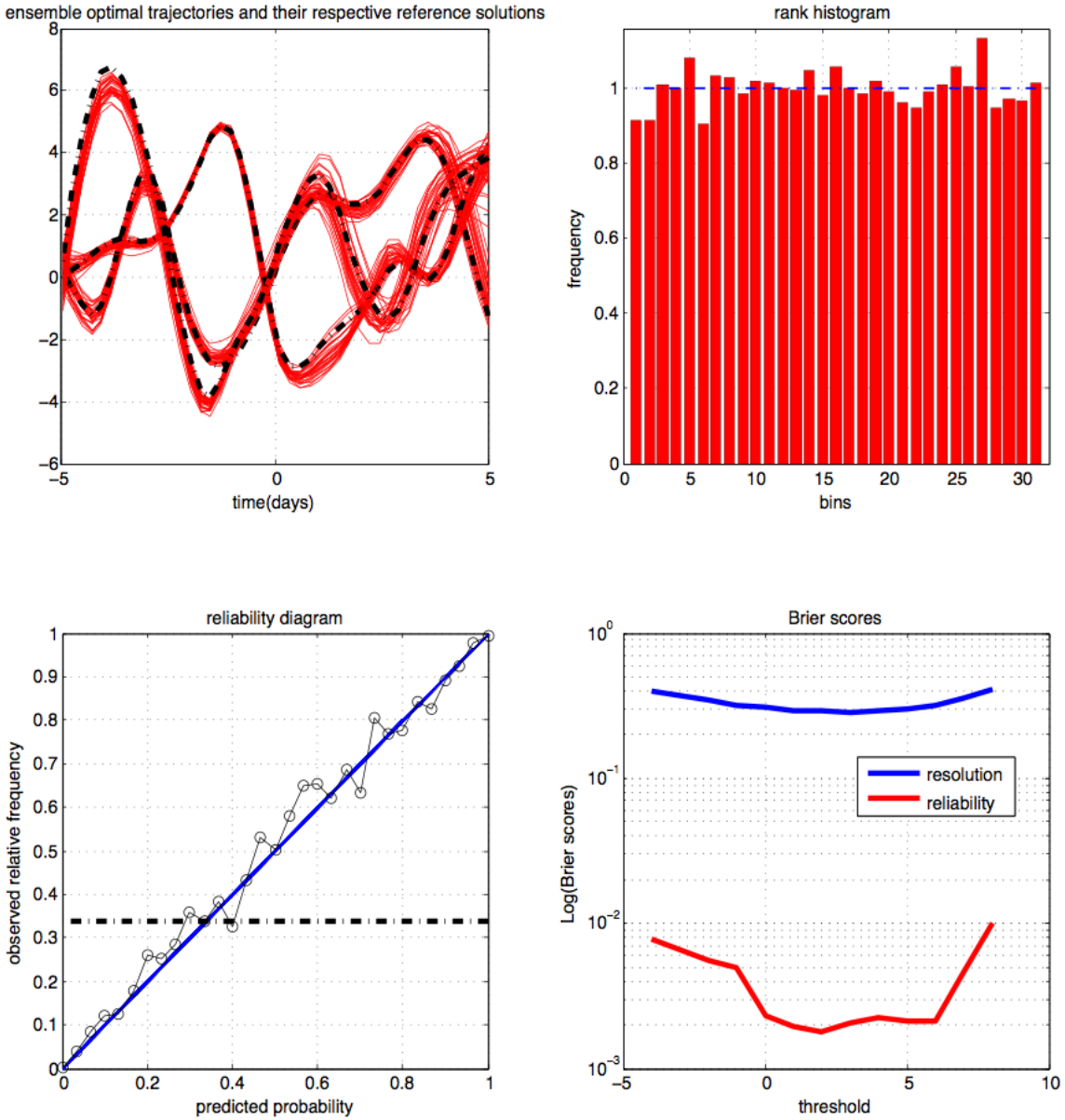




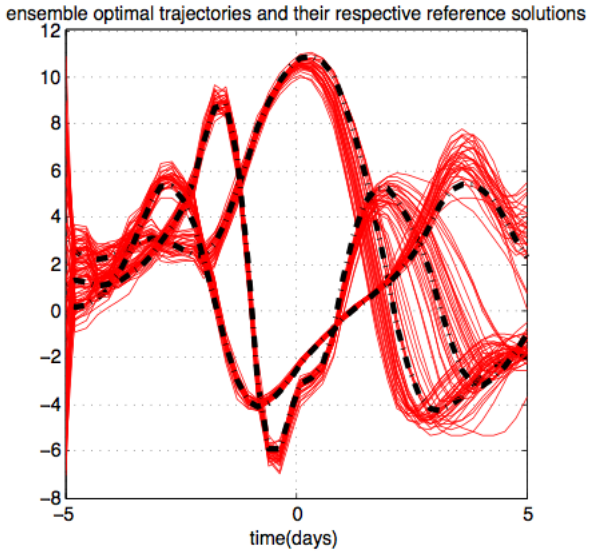
Nonlinear Lorenz'96. EnKF. Diagnostics after 5 days of assimilation

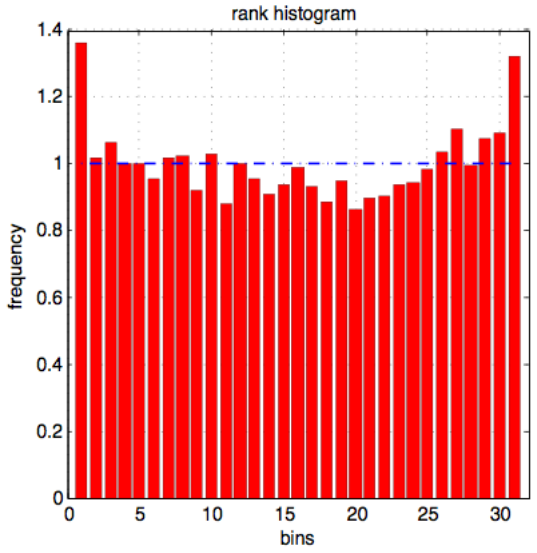


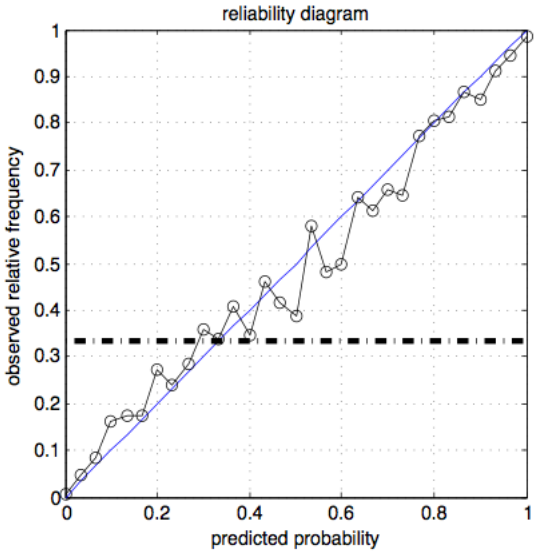
Nonlinear Lorenz'96. PF. Diagnostics after 5 days of assimilation

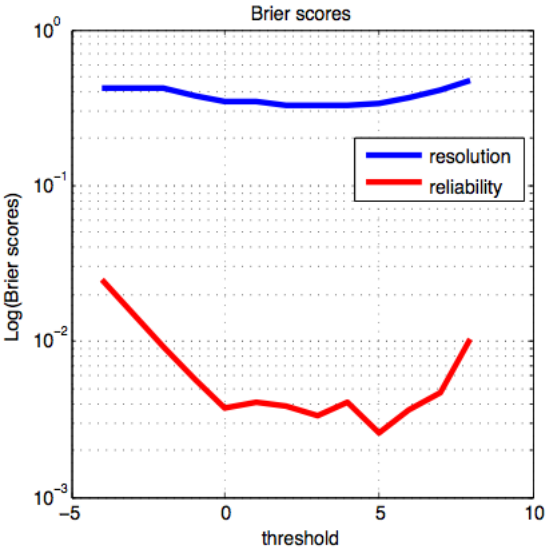


EnsVAR. Diagnostics for 5-day forecasts

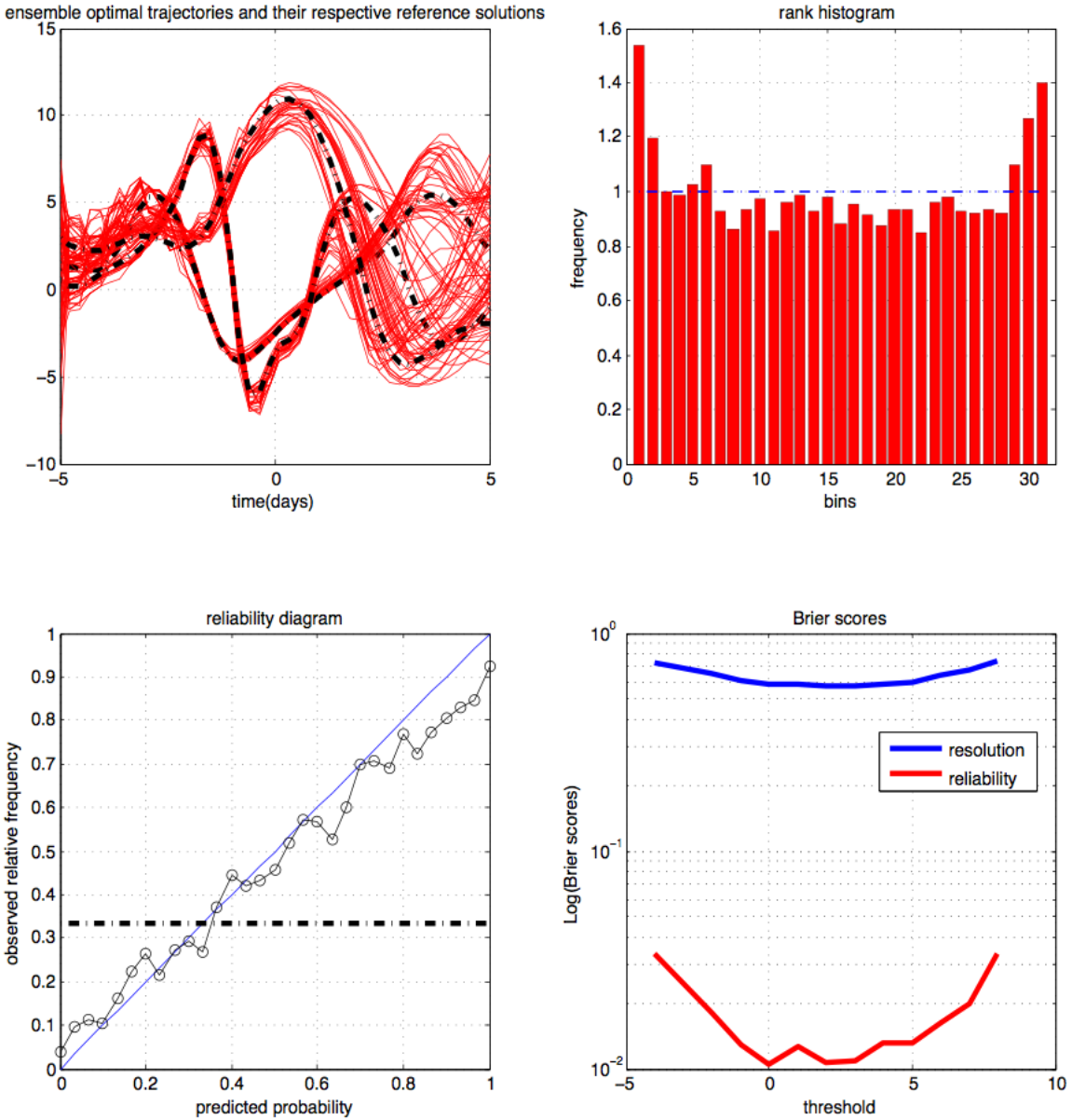








EnKF. Diagnostics for 5-day forecasts



PF. Diagnostics for 5-day forecasts

method \ *DA procedure*	*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 7

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

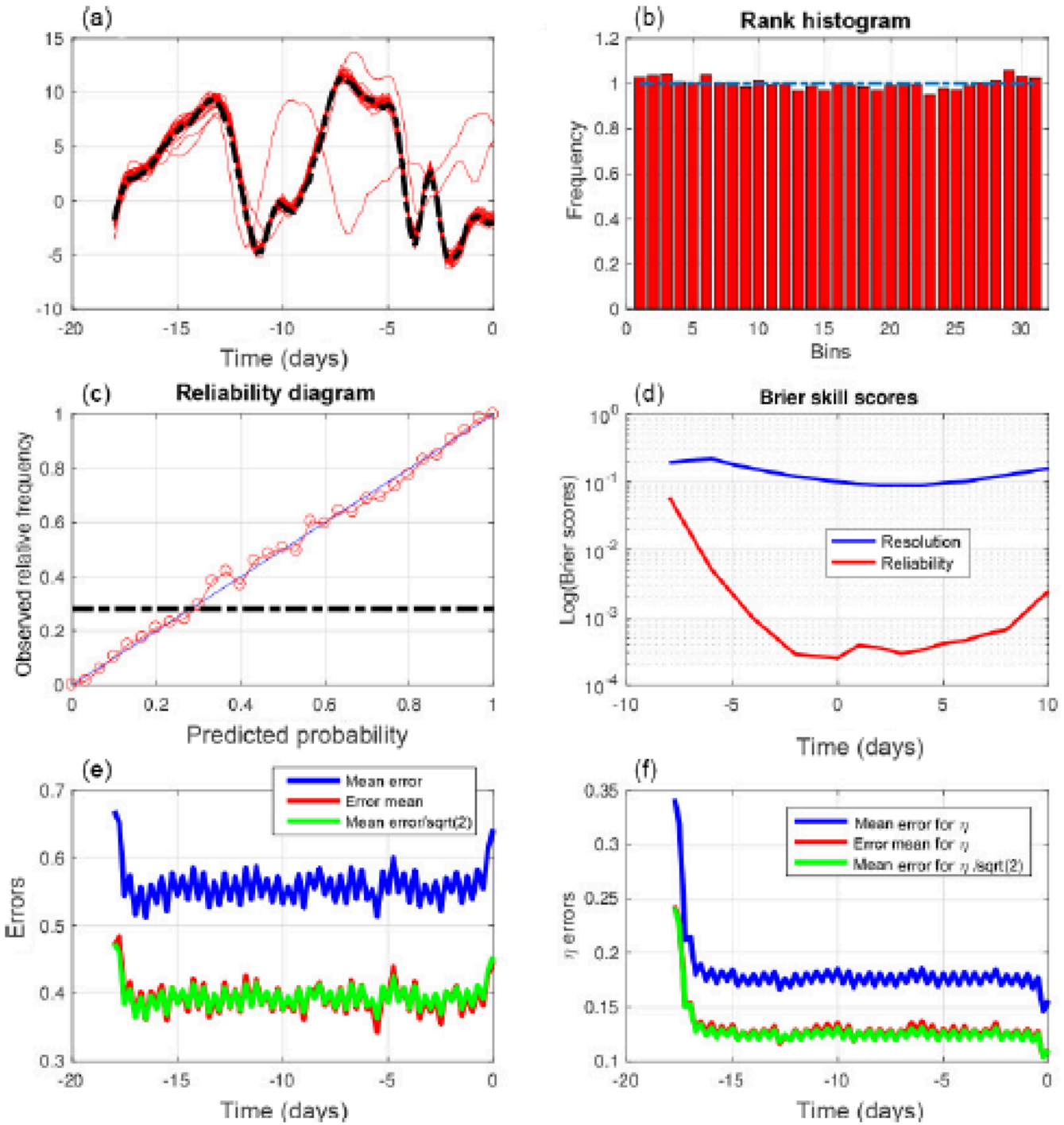
$$\xi_{k+1} = M_k(\xi_k) + \eta_k \quad , \quad k = 0, \ldots, K\text{-}1$$

'Observations' consist of

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

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



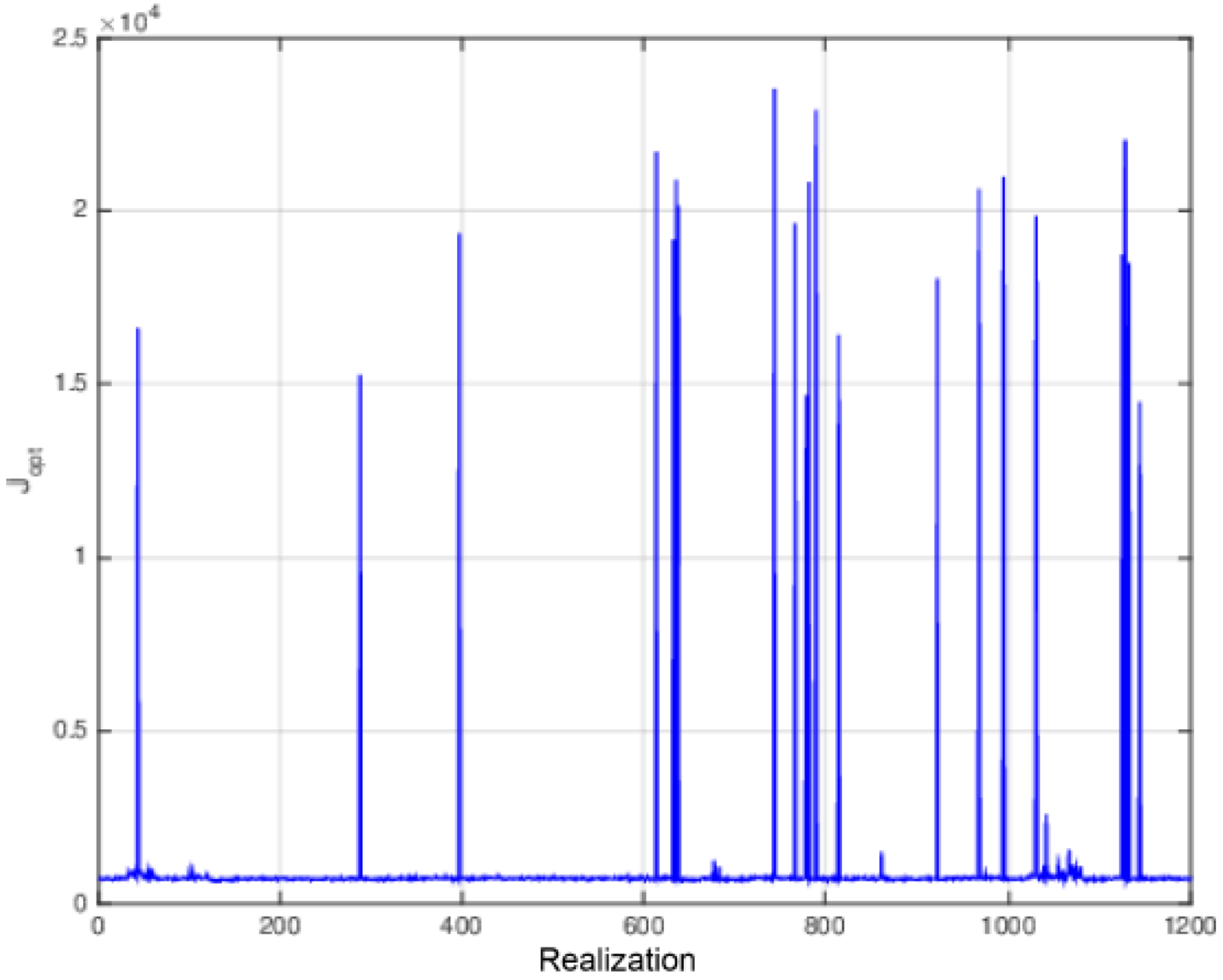
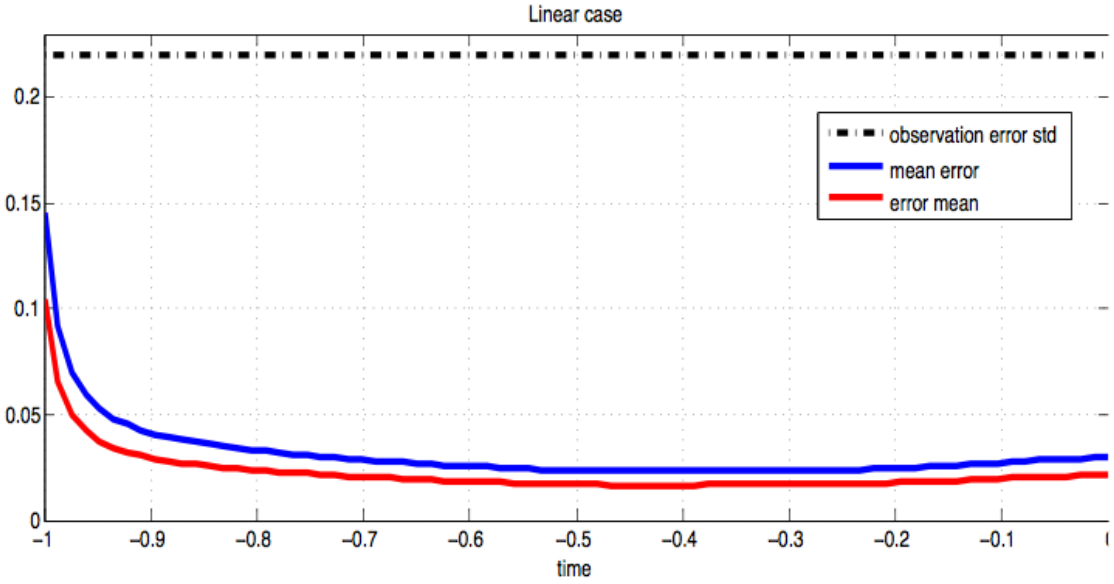


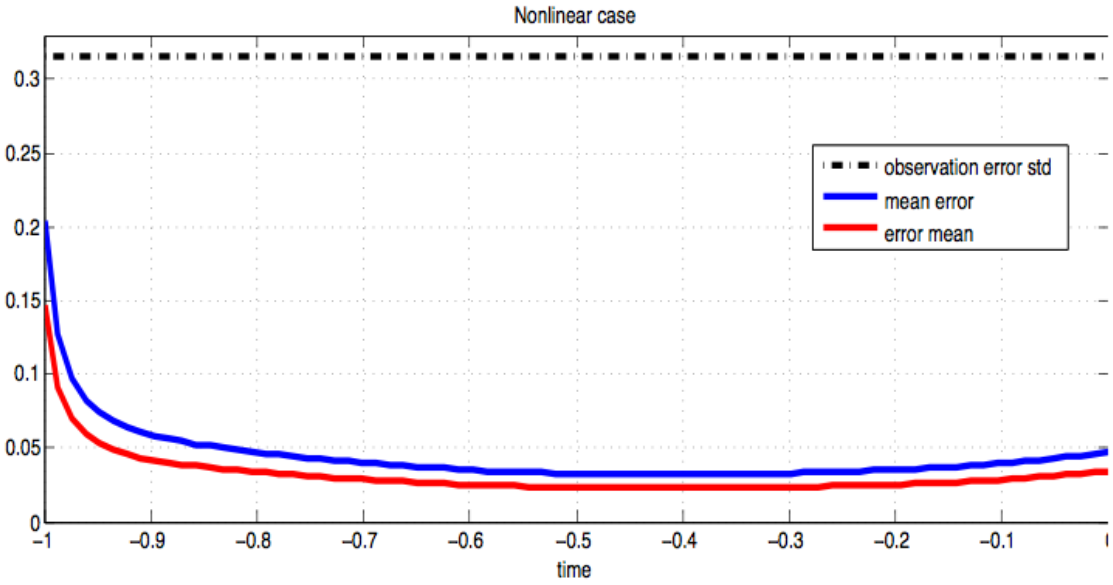
Figure 11. Values of (half) the minima of the objective function for all realizations of the weak-constraint assimilations over 18-day windows.

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).

Unknown $\boldsymbol{x}$ to be determined. Belongs to *state space* $\mathcal{S}$, with dimension n
Data, belonging to *data space* $\mathcal{D}$, with dimension m, available in the form

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta}$$

where $\boldsymbol{\Gamma}$ is a known $(m \times n)$-matrix, rank$\boldsymbol{\Gamma} = n$ and $\boldsymbol{\zeta}$ is 'error'

Best Linear Unbiased Estimate (BLUE)

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

with $\boldsymbol{\mu} = E(\boldsymbol{\zeta})$ and $\boldsymbol{S} = E[(\boldsymbol{\zeta} - E(\boldsymbol{\zeta}) (\boldsymbol{\zeta} - E(\boldsymbol{\zeta})^{\mathrm{T}}]$.

$E(\boldsymbol{x}^a - \boldsymbol{x}) = 0$ $\qquad$ $E[(\boldsymbol{x}^a - \boldsymbol{x}) (\boldsymbol{x}^a - \boldsymbol{x})^{\mathrm{T}}] \equiv \boldsymbol{P}^a = (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1}$

Determinacy condition : rank$\boldsymbol{\Gamma} = n$. Data contain information, directly or indirectly, on every component of state vector x. Requires $m \geq n$.

BLUE is invariant in any change of origin, or in any invertible linear transformation, in either data or state space. In particular, it is independent of the choice of a scalar product or norm in either of those spaces. *BLUE* minimizes the quadratic estimation error on any component of $\boldsymbol{x}$.

If error ζ is gaussian, $\zeta \sim \mathcal{N}[\boldsymbol{\mu}, S]$, *BLUE* achieves bayesian estimation in the sense that

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

Any assumed probability distribution $P(\zeta)$ defines a conditional probability distribution $P(x \mid z)$ for x. In case the distribution $P(\zeta)$ is known only through its expectation $\boldsymbol{\mu}$ and covariance matrix S, the gaussian distribution $\mathcal{N}[\boldsymbol{\mu}, S]$ leads for $\boldsymbol{x}$ to the conditional probability distribution $P(x \mid z)$ with the largest entropy. The gaussian choice is in that sense the 'least-committing' choice.

BLUE is the simplest of non-simplicist algorithms.

The *BLUE* can be obtained by minimization of the following scalar objective function, defined on state space $\mathcal{X}$

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

And in case of nonlinearity ?

$$z = \boldsymbol{\Gamma}(\boldsymbol{x}) + \boldsymbol{\zeta}$$

Variational approach can be heuristically implemented

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

It works !

If data are of the form (after possibly an appropriate transformation)

$$x^b = x + \zeta^b$$
$$y = H(x) + \varepsilon$$

Transformation

$$x^b = x + \zeta^b$$
$$y - H(x^b) = H(x) - H(x^b) + \varepsilon \approx H'(x - x^b) + \varepsilon$$

where H' is jacobian of H, makes the estimation problem linear in the deviation $x - x^b$
(*tangent linear approximation*)

All algorithms that have been presented in the course, with the exception of particle filters, are empirical heuristic extensions of the BLUE approach to approximate nonlinear and non-gaussian situations.

Artificial Intelligence

(aka ***Machine Learning*** or ***Deep Learning***)

Numerical modelling of the atmospheric and oceanic flow, as presented in the course, fundamentally built on physical laws known from physics (conservation of mass, momentum and energy).

Why not directly use observations (for instance, in the case of a weather forecast, why not look for analogues in the past, and make the forecast from those analogues) ?

E. N. Lorenz (1960s). Sample of past observations will never be large enough for competing with physically-based models.

But :

- there is no incompatibility between the two approaches

- there remain many processes in numerical models which we do not know how to describe on the basis of well-established physical laws (interactions between atmosphere and underlying medium, such as *e.g.* vegetation, all kinds of subgrid scale processes, …)

Artificial Intelligence (aka ***Machine Learning***) (continuation)

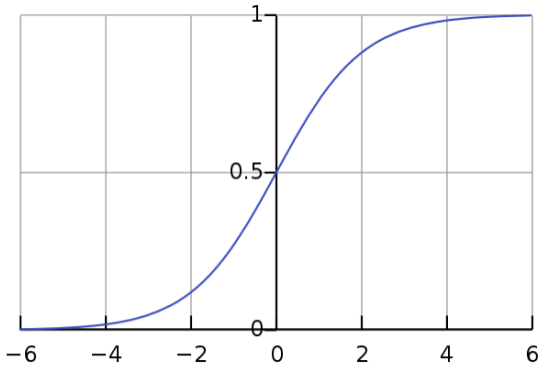
Powerful numerical tools have been developed for the exploitation of very large sets of data (*big data*)

Neural networks. Define an explicit numerical link between an *input set* and an *output set*. Define function F such that, to some useful degree of approximation

$$y = F(x)$$

where x and y belong to the input and output set respectively.

The function F is typically built as a composition of *sigmoid functions*



Artificial Intelligence (aka ***Machine Learning***) (continuation 2)

Neural networks have turned out to be extrermely efficient in many applications. In the context of assimilation of observations, they have been used for defining for instance the observation operators (***H***) corresponding to satellite observations. But they have been used more recently, in evaluation studies, and with some success, for determining 'dynamical laws'.

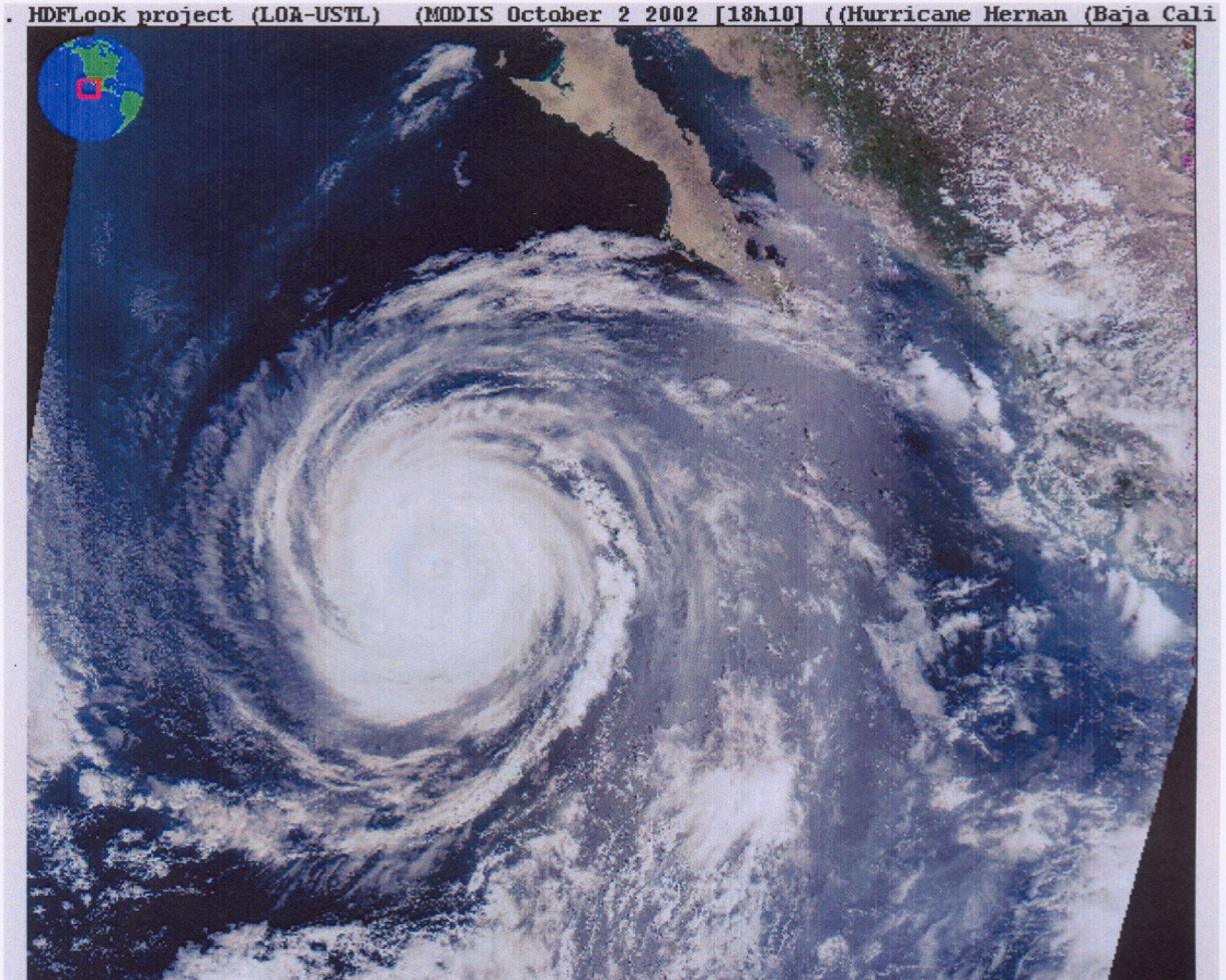
Assimilation, which originated from the need of defining initial conditions for numerical weather forecasts, has gradually extended to many diverse applications

- Oceanography
- Palaoclimatology
- 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)
- Mathematical studies, independently of direct real life applications
- …

It has now become a major tool of numerical environmental science, and a subject of mathematical study in its own right.

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

La Fin du Cours …