Assimilation of Observations and Bayesianity

Olivier Talagrand and Mohamed Jardak
Laboratoire de Météorologie Dynamique, École Normale Supérieure, Paris, France

Workshop *Large-scale inverse problems in the earth sciences*
Special Semester on *Multiscale Simulation & Analysis in Energy and the Environment*

Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria
28 October 2011

Purpose of assimilation : reconstruct as accurately as possible the state of an observed (geophysical) system, using all available appropriate information. The latter essentially consists of

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

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

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

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

For some reason, uncertainty is conveniently described by probability distributions.

Assimilation can be stated as a problem in bayesian estimation, *viz*.,

Determine the conditional probability distribution for the state of the system, knowing everything we know.

Example

Available information in the form

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

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

Γ is known linear operator of $\mathcal{S}$ into $\mathcal{D}$
ζ is unknown random 'error', $\zeta \sim \mathcal{N}[\mu, S]$

Then conditional probability distribution is

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

where

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

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

Formula

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

is (still ?) at the basis of a large fraction of algorithms used operationaly for assimilation of observations in geophysical fluid applications. Those algorithms are heuristic extensions to mildly nonlinear (and non-gaussian) situations of a basically linear method.

Only exceptions so far

- *Ensemble Kalman Filter*, which does not need any linearity in dynamical evolution, but remains of form (1) in 'updating' phase.

- *Particle filters* (P. J. van Leeuwen, R. Miller), which are bayesian without any need for linearity or gaussianity, but (still ?) do not exist in real meteorology or oceanography.

Bayesian estimation is however impossible in practice in its general theoretical form in large dimensional problems because

- It is impossible to explicitly describe a probability distribution in a space with dimension even as low as $n \approx 10^3$, not to speak of the dimension $n \approx 10^{7\text{-}9}$ of present Numerical Weather Prediction models.

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

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

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

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

Data

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

Then conditional posterior probability distribution

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

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 producing sample of independent realizations of posterior probability distribution

Perturb data vector additively according to error probability distribution $\mathcal{N}[0, S]$, *and compute analysis* x^a *for each perturbed data vector*.

Work done here. Apply same recipe to nonlinear and non-gaussian cases, and look at what happens.

Everything synthetic, Two one-dimensional toy models : Kuramoto-Sivashinsky equation and Lorenz '96 model. Perfect model assumption

Variational assimilation.

There is no (and there cannot be) a general objective test of bayesianity. We use here as a substitute the much weaker property of ***reliability***.

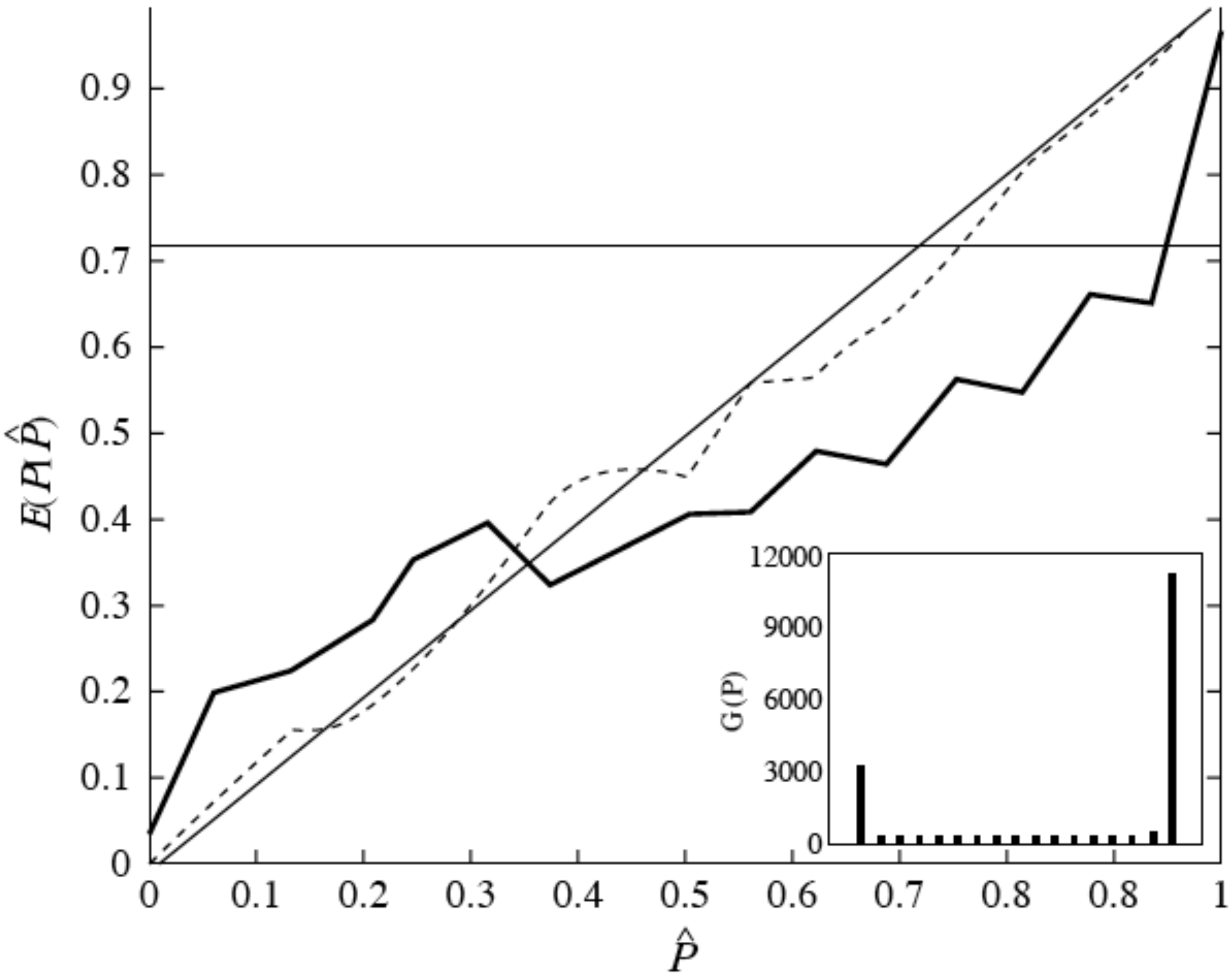
Reliability. Statistical consistency between predicted probability of occurrence and observed frequency of occurrence (it rains 40% of the time in circumstances when I predict 40%-probability for rain).

Observed frequency of occurrence $p'(p)$ of event, given that it has been predicted to occur with probability p, must be equal to p.

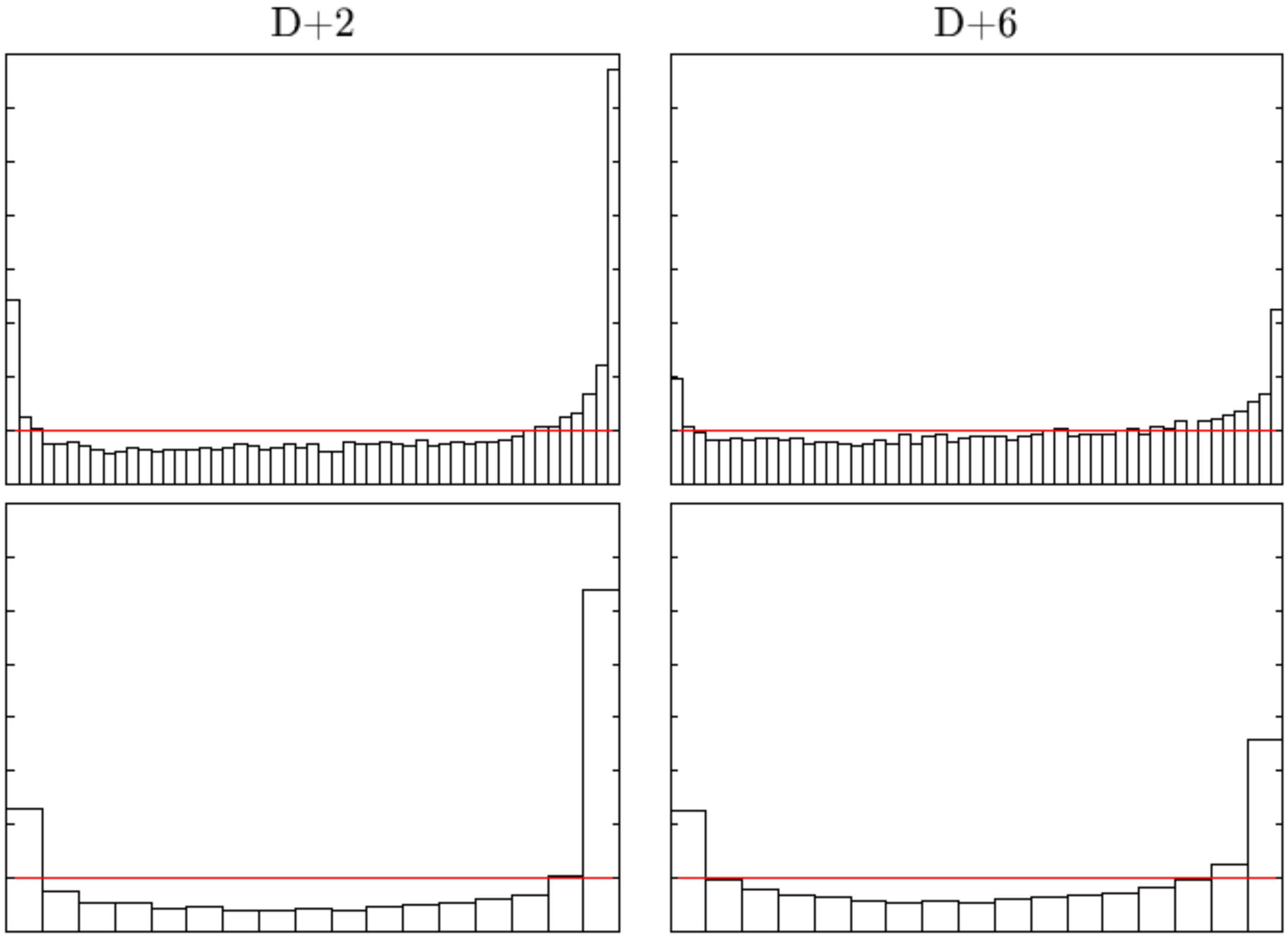
For any p, $p'(p) = p$

More generally, frequency distribution $F'(F)$ of reality, given that probability distribution F has been predicted for the state of the system, must be equal to F.

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



Reliability diagramme, NCEP, event $T_{850} > T_c$ - 4C, 2-day range, Northern Atlantic Ocean, December 1998 - February 1999



Rank histograms, T_{850}, Northern Atlantic, winter 1998-99

Top panels: ECMWF, bottom panels: NMC (from Candille, Doctoral Dissertation, 2003)

Ens/4D-Var methodology

- define the objective function.

$$\mathfrak{J}(X) = \frac{1}{2}\left\{(X - X_b)^t \mathfrak{B}^{-1}(X - X_b)\right\} \; +$$

$$\frac{1}{2}\sum_{iobs=0}^{Nobs}\left\{(Y_i - \mathfrak{H}_i(X_i))^t \mathfrak{R}_i^{-1}(Y_i - \mathfrak{H}_i(X_i))\right\}$$

1. $\mathfrak{B}$ background error covariance matrix.
2. $\mathfrak{R}$ observation error covariance matrix.
3. $\mathfrak{H} : \mathbb{R}^{state} \longrightarrow \mathbb{R}^{obs}$ observation operator.
4. X_b background state vector.
5. Y_i observation vector at time $t = t_i$.

- find the optimal initial solution X_0^{opt} and the optimal trajectory X^{opt}.

$$X_0^{opt} = \min_{X \in \mathfrak{X}} \mathfrak{J}(X) \;\; \text{and} \;\; X^{opt}(t) = \mathfrak{M}_{0\rightarrow t}(X_0^{opt}(x))$$

- Ens/4D-Var method
  - for iens = 1 : Nens
    1. sample
       - $X_b \in \mathcal{N}(\bar{X}_b, \mathfrak{B}^{1/2})$
       - $Y_i \in \mathcal{N}(\bar{Y}_i, \mathfrak{R}^{1/2})$
    2. form the ensemble member objective function $\mathfrak{J}(X(iens))$
    3. find the optimal initial ensemble member solution.

       $$X_0^{opt}(iens) = \min_{X \in \mathfrak{X}} \mathfrak{J}(X(iens))$$

    4. find the optimal ensemble member trajectory.

       $$X^{opt}(t, iens) = \mathfrak{M}_{0 \to t}(X_0^{opt}(iens))$$

  - end for
- Validation and verification of probabilistic prediction.
- Diagnosis and impacts of non-Gaussianity and non-linearity.

The models

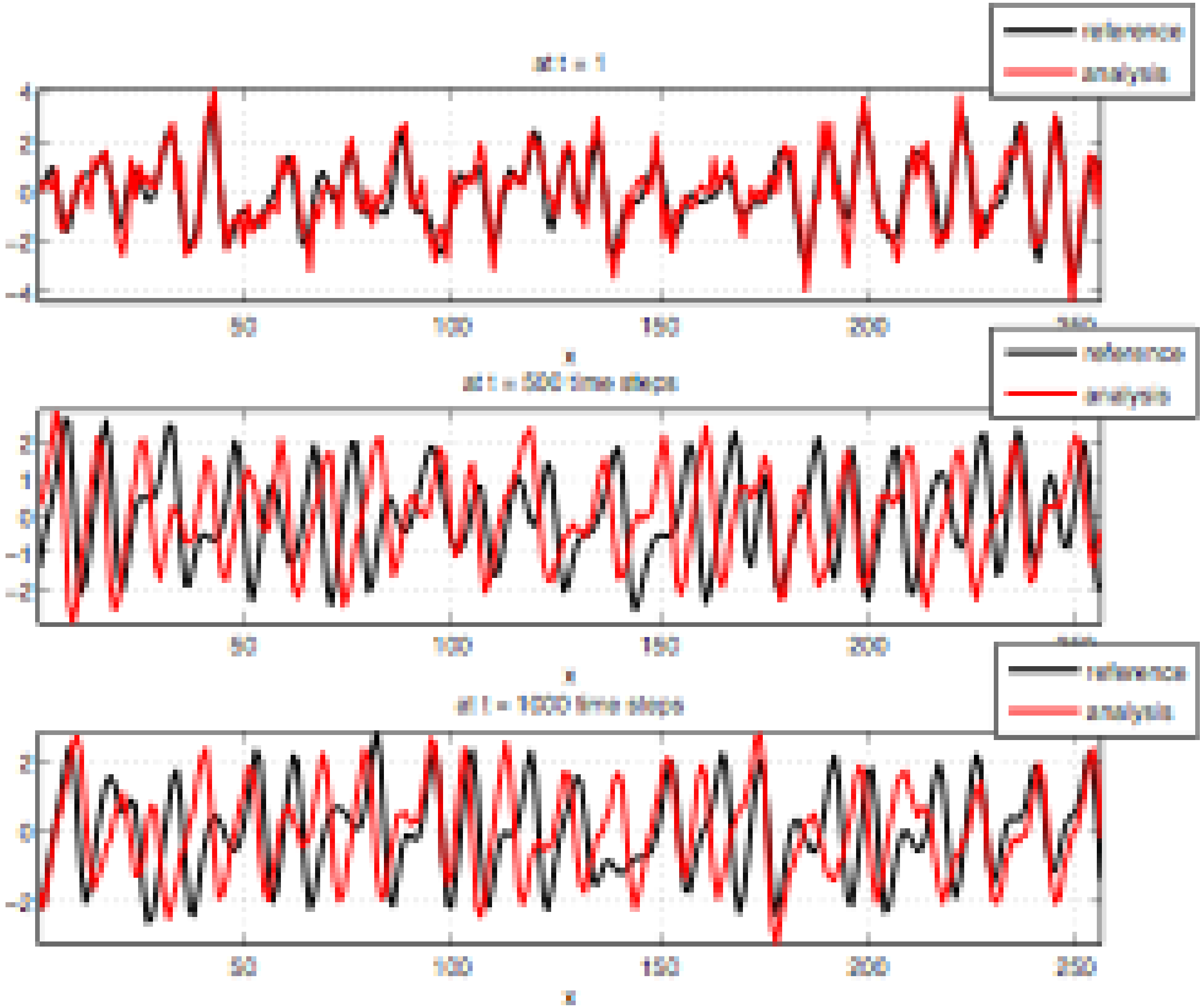
- The nonlinear model: the Kuramoto-Sivashinsky equation

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial^4 u}{\partial x^4} + \dfrac{\partial^2 u}{\partial x^2} + u\dfrac{\partial u}{\partial x} = 0, \quad x \in [0, L] \\[2ex] \dfrac{\partial^i u}{\partial x^i}(x+L, t) = \dfrac{\partial^i u}{\partial x^i}(x, t) \text{ for } i = 0, 1, \cdots, 4, \ \forall t > 0 \\[2ex] u(x, 0) = \cos(\dfrac{x}{32}) \left(1 + \sin(\dfrac{x}{32})\right) \end{cases}$$

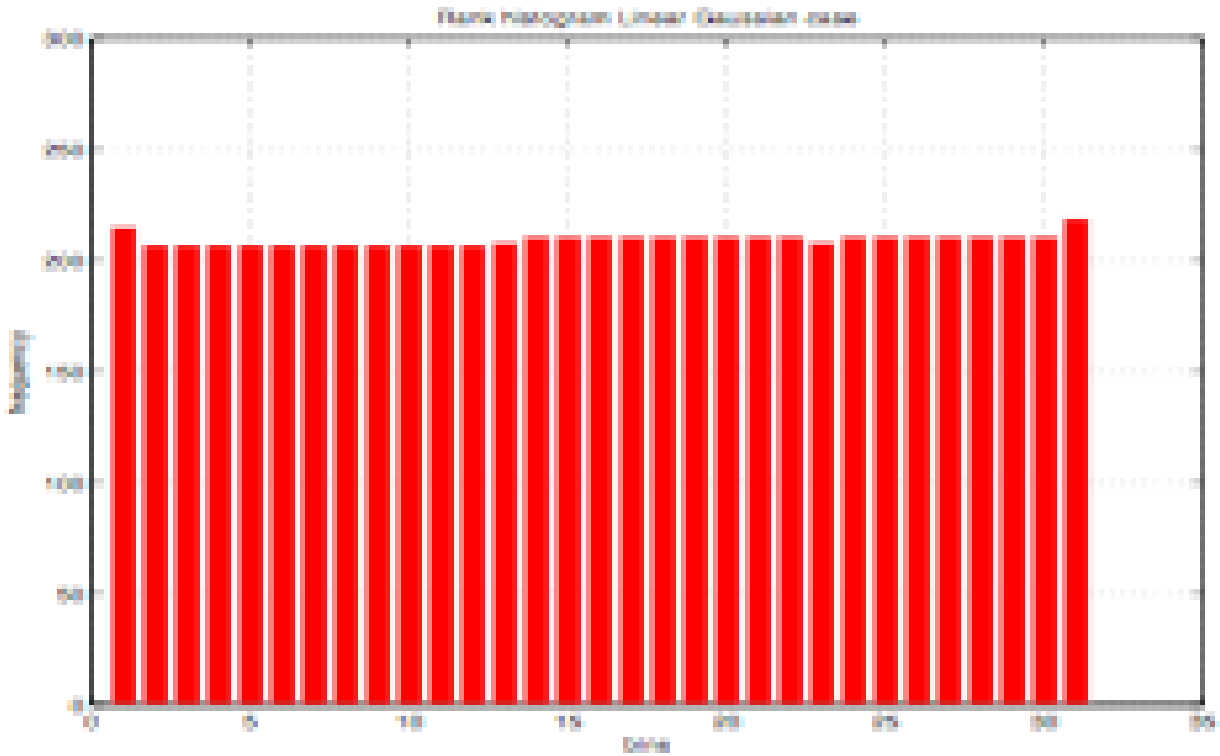
- The linear model: the tangent linear KS equation

$$\begin{cases} \dfrac{\partial v}{\partial t} + \dfrac{\partial^4 v}{\partial x^4} + \dfrac{\partial^2 v}{\partial x^2} + u\dfrac{\partial v}{\partial x} + v\dfrac{\partial u}{\partial x} = 0, \quad x \in [0, L] \\[2ex] \dfrac{\partial^i v}{\partial x^i}(x+L, t) = \dfrac{\partial^i v}{\partial x^i}(x, t) \text{ for } i = 0, 1, \cdots, 4, \ \forall t > 0 \\[2ex] v(x, 0) = \text{very small} \end{cases}$$

4D-Var : KS model



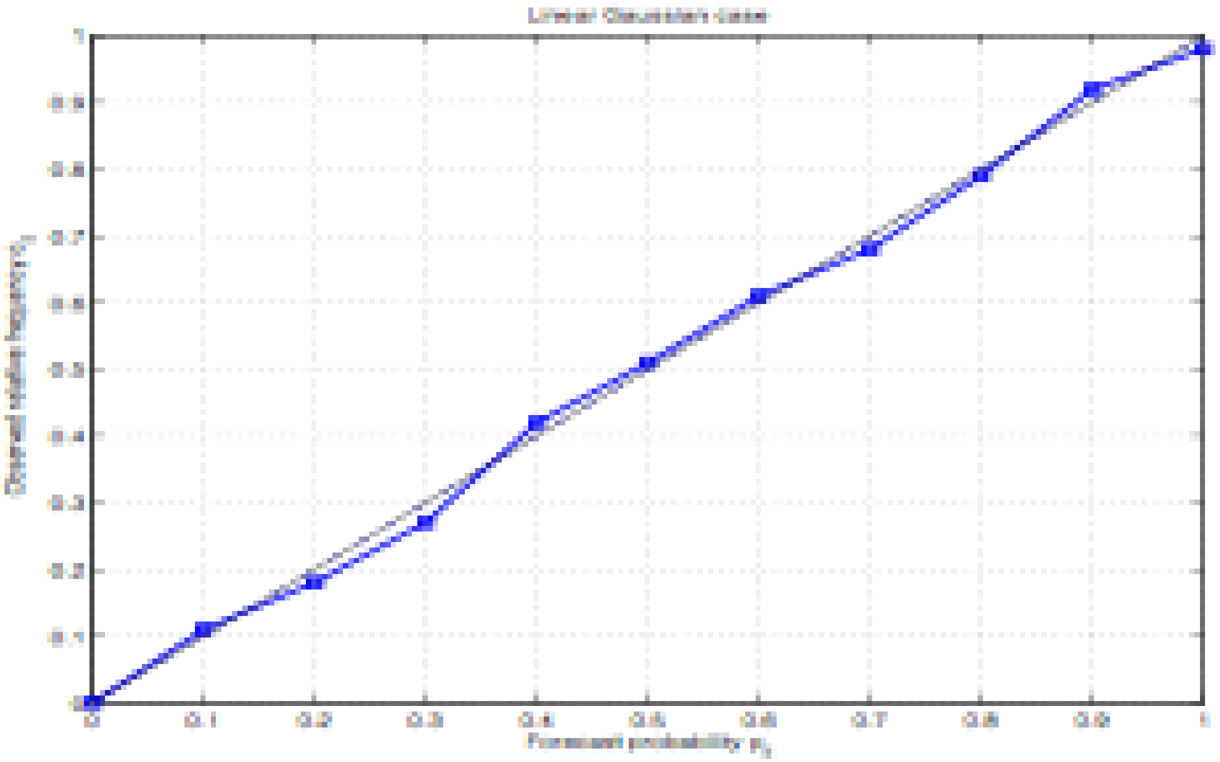
Rank Histogram: K-S model linear case



Size of ensembles : 30 ; number of ensemble assimilations :7000

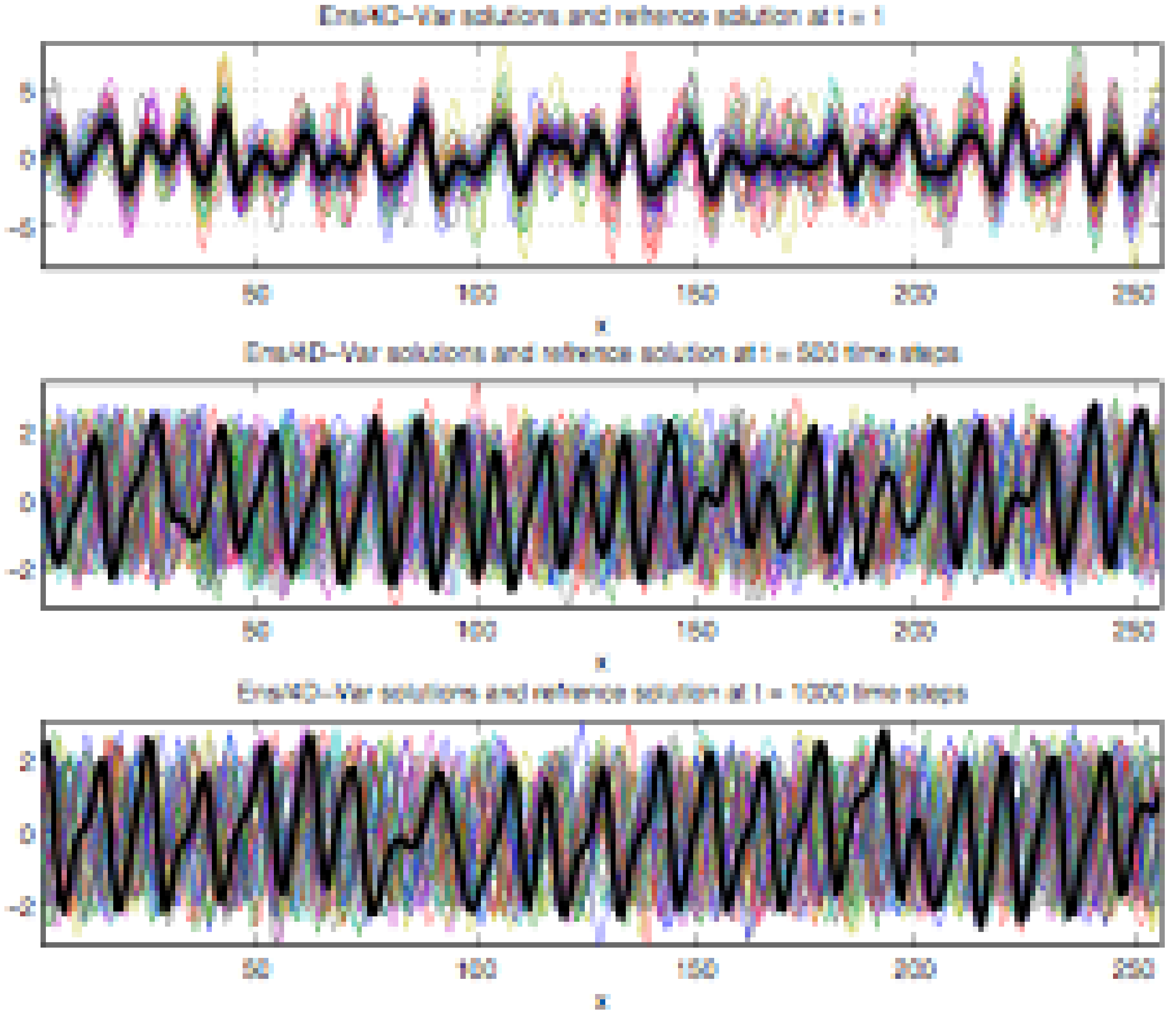
Reliability diagram: K-S model linear case

1. Threshold $\tau = 0.5$.
2. 7000 realizations.

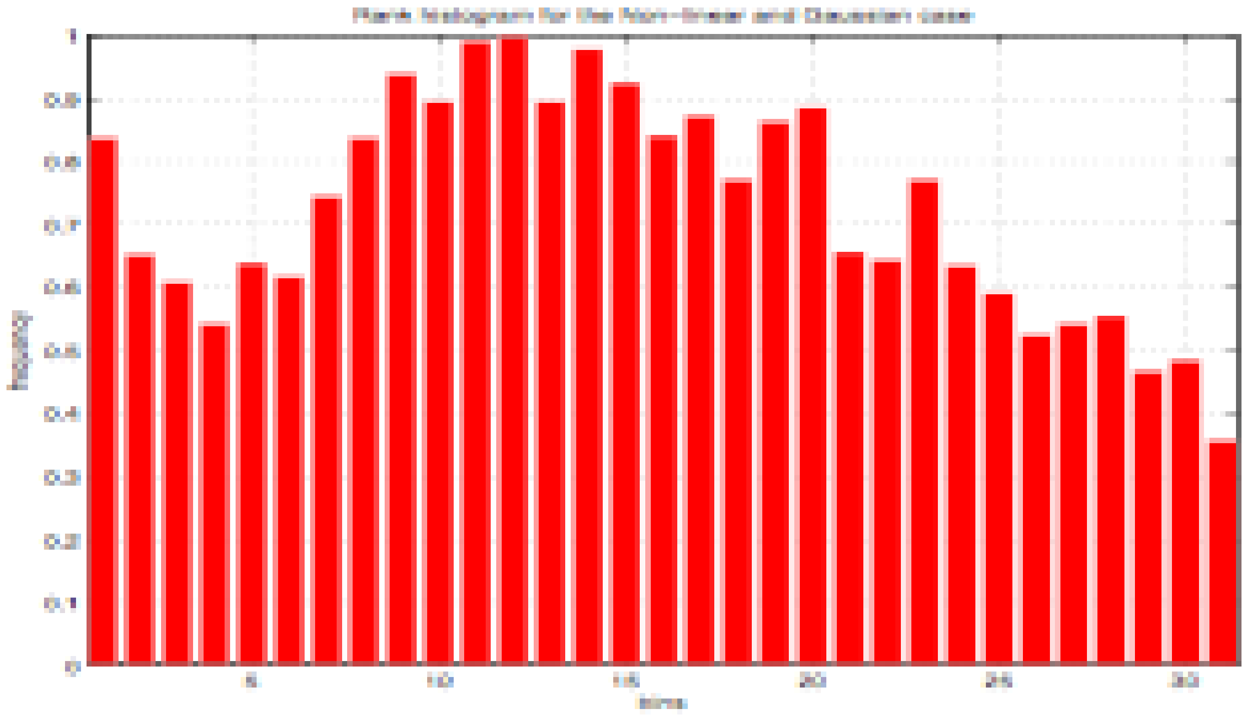


Non-gaussianity of the error, if model is kept linear, has no significant impact on scores.

Ens/4Dvar solutions: KS model

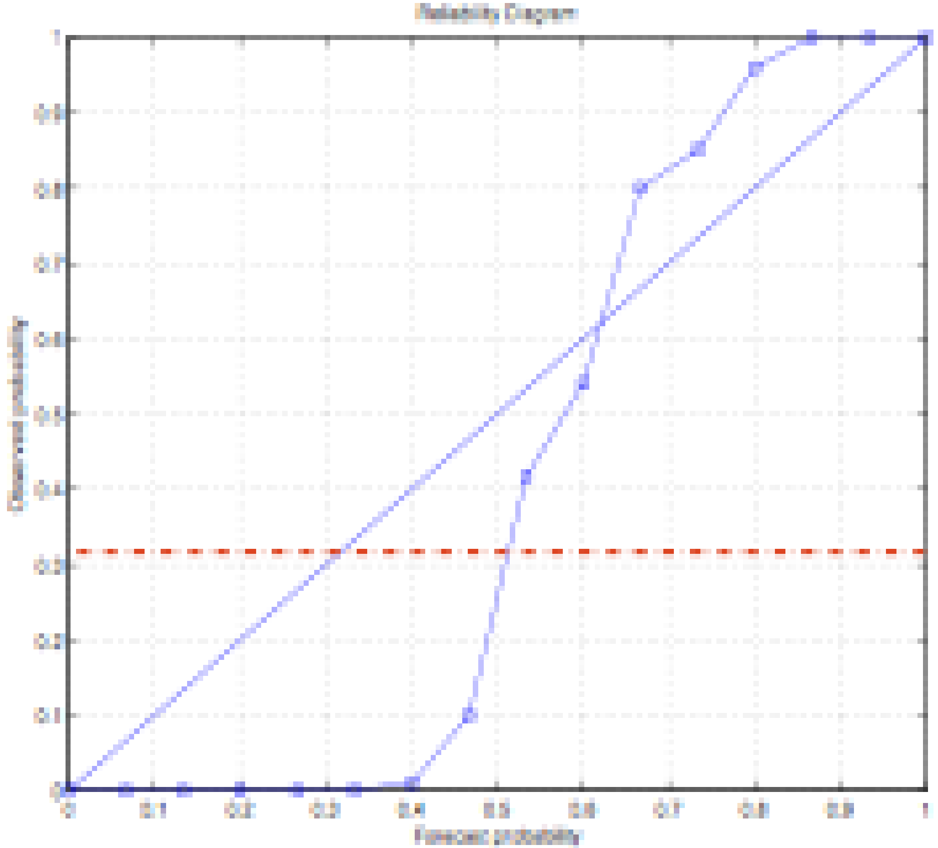


Rank Histogram: KS model



Reliability diagram: KS model

1. Threshold $\tau = 0.5$.
2. 7000 realizations.



Brier Score (Brier, 1950), relative to binary event $\mathcal{E}$

$$\mathcal{B} \equiv E[(p - p_o)^2]$$

where p is predicted probability of occurrence, $p_o = 1$ or 0 depending on whether $\mathcal{E}$ has been observed to occur or not, and E denotes average over all realizations of the prediction system.

Decomposes into

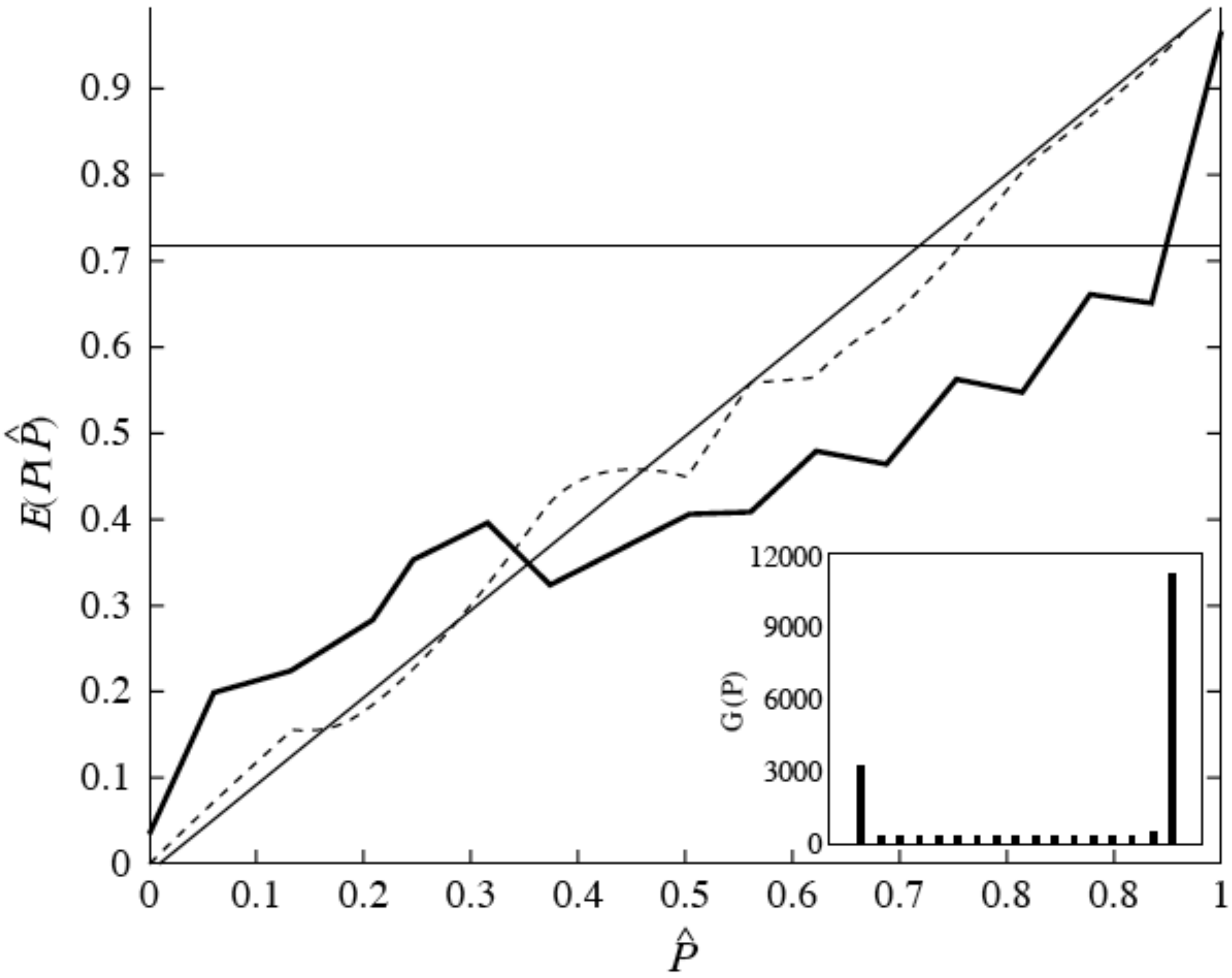
$$\mathcal{B} = E[(p-p')^2] \;-\; E[(p'-p_c)^2] \;+\; p_c(1-p_c)$$

where $p_c \equiv E(p_o) = E(p')$ is observed frequency of occurrence of $\mathcal{E}$.

First term $E[(p-p')^2]$ measures reliability.

Second term $E[(p'-p_c)^2]$ measures dispersion of *a posteriori* calibrated probabilities p'. The larger that dispersion, the more discriminating, or resolving, and the more useful, the prediction system. That term measures the *resolution* of the system.

Third term, called *uncertainty* term, depends only on event $\mathcal{E}$, not on performance of prediction system.



Reliability diagramme, NCEP, event $T_{850} > T_c$ - 4C, 2-day range, Northern Atlantic Ocean, December 1998 - February 1999

Brier Skill Score

$$\mathcal{B}_{SS} \equiv 1 - \mathcal{B}/p_c(1-p_c)$$

(positively oriented)

and components

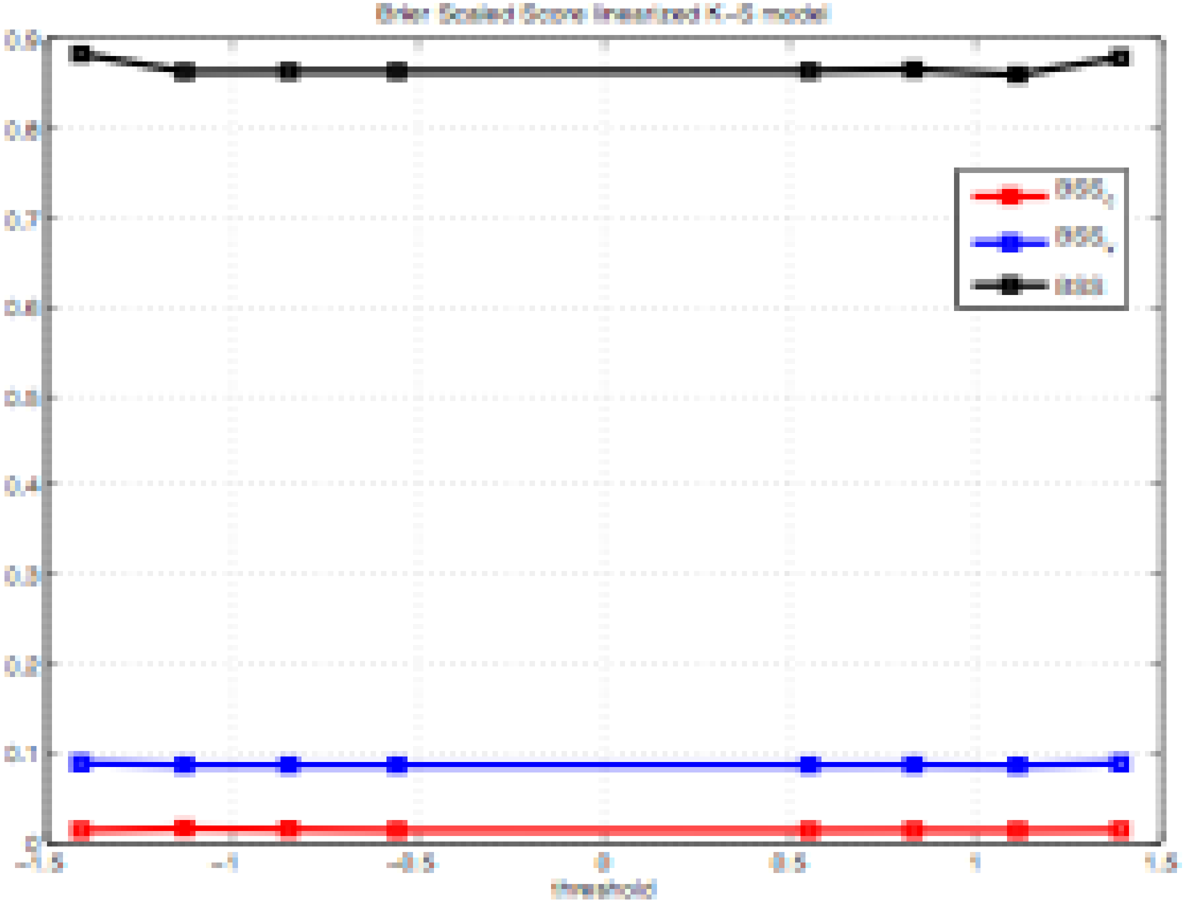
$$\mathcal{B}_{rel} \equiv E[(p-p')^2]/p_c(1-p_c)$$

$$\mathcal{B}_{res} \equiv 1 - E[(p'-p_c)^2]\,/p_c(1-p_c)$$

(negatively oriented)

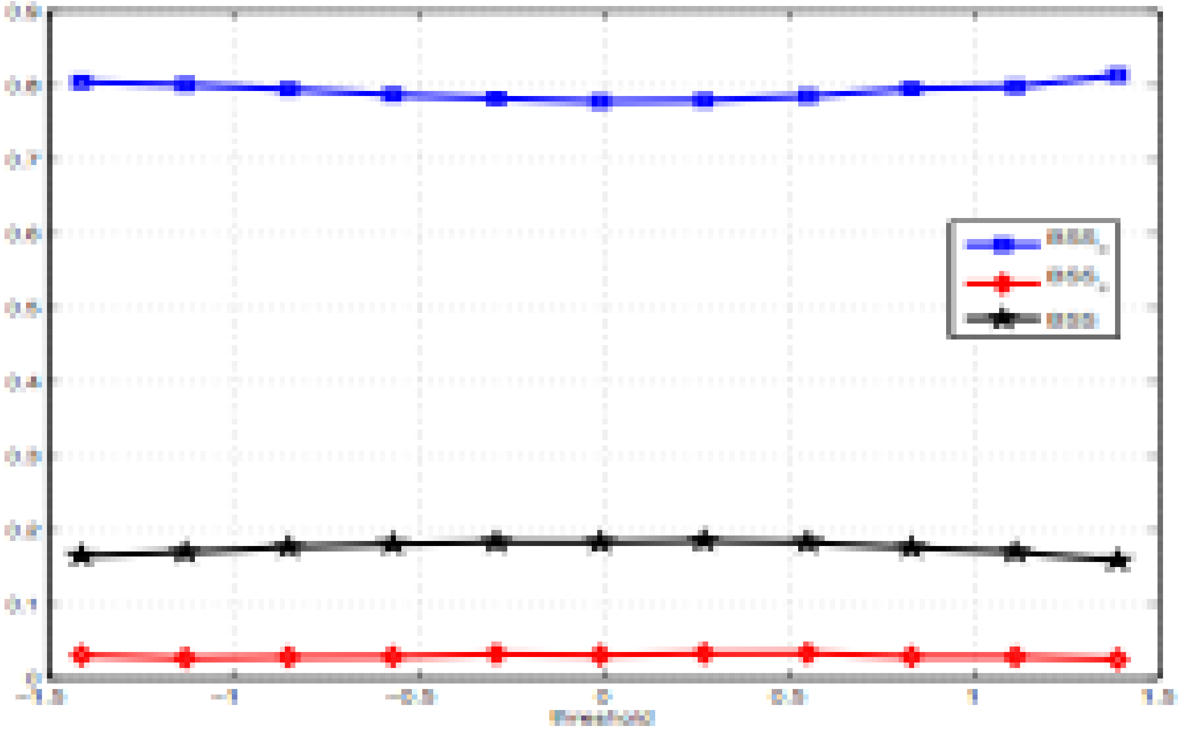
Brier Skill Score: TL KS model

1. Threshold $\tau \in [-1.5 \;\; 1.5]$.
2. 7000 realizations.



Brier Skill Score: KS model nonlinear case

1. Threshold $\tau \in [-1.5 \;\; 1.5]$.
2. 7000 realizations.
3. $\mathfrak{C} = \mathfrak{C}(t = 500 \;\; \text{time steps})$.



- Forward model

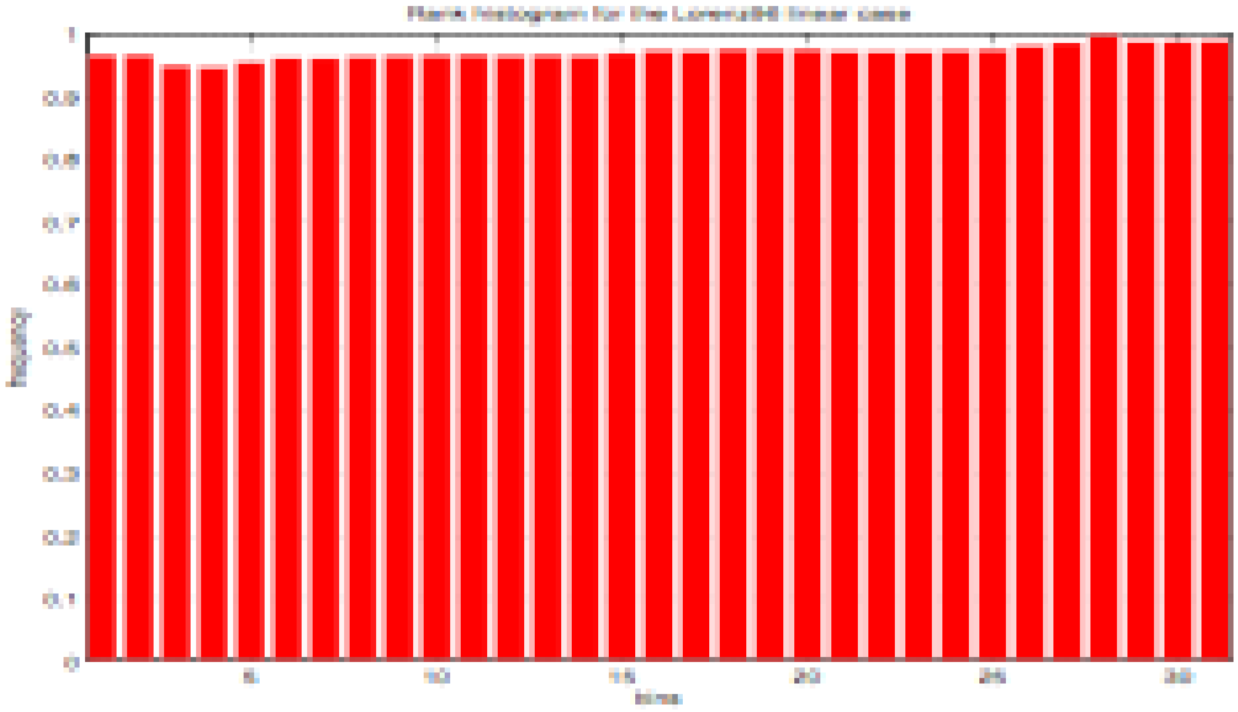
$$\frac{dX_k}{dt} = (X_{k+1} - X_{k-2})X_{k-1} - X_k + F \quad \text{for} \quad k = 1, \cdots, N$$

- Tangent linear model

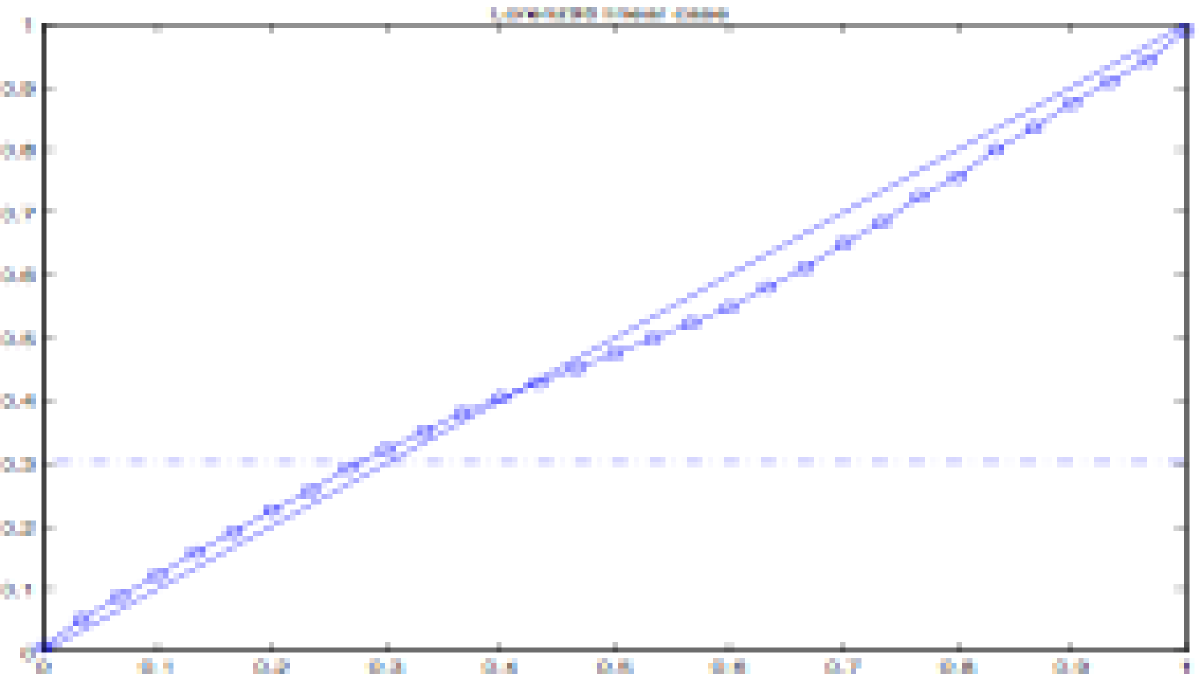
$$\begin{cases} \frac{d\delta X_k}{dt} = X_{k-1}\delta X_{k+1} + (X_{k+1} - X_{k-2})\delta X_{k-1} - X_k\delta X_{k-2} - \delta X_k \\ \qquad \text{for} \quad k = 1, \cdots, N \end{cases}$$

- the index k is cyclic so that $X_{k-N} = X_{k+N} = X_k$
- $F = 8$ external driving force
- X_k a damping term
- $N = 40$ the system size

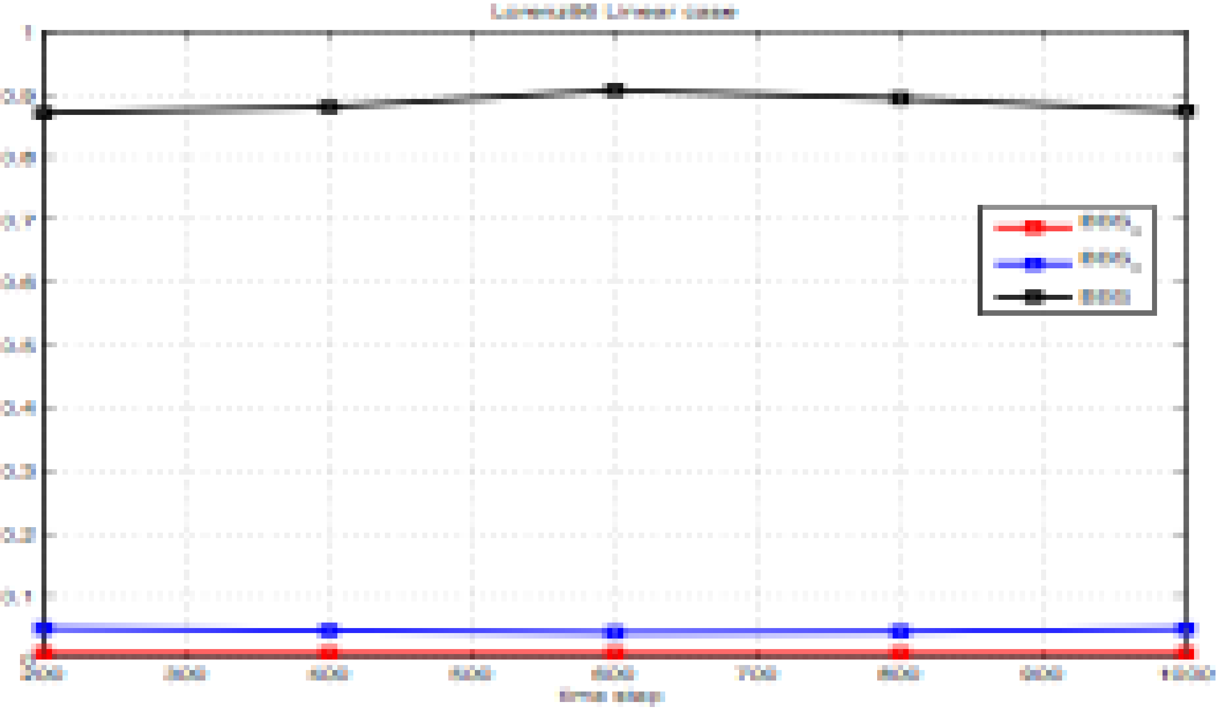
rank histogram: Lorenz96 model linear case



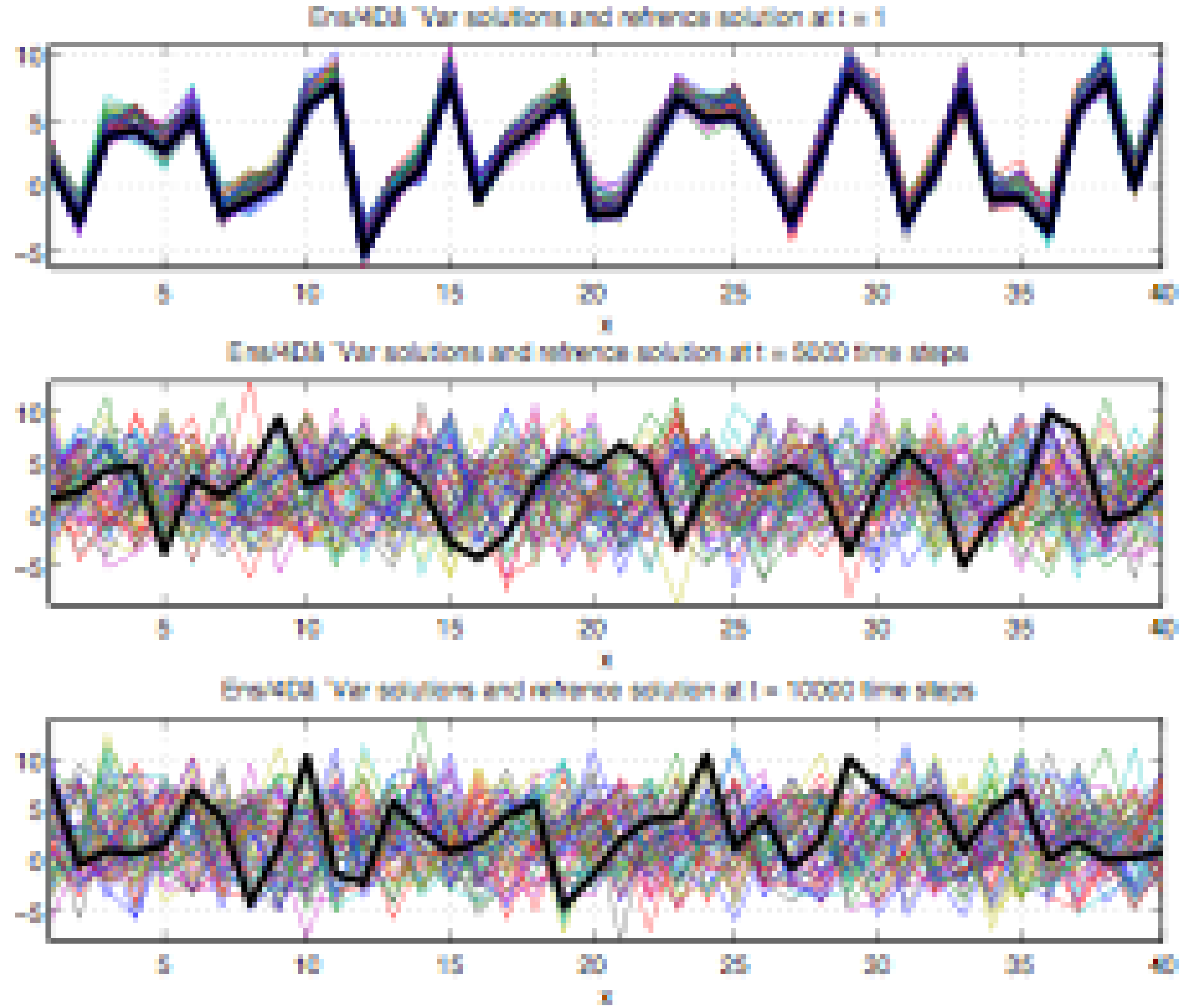
reliability diagram: Lorenz96 model linear case



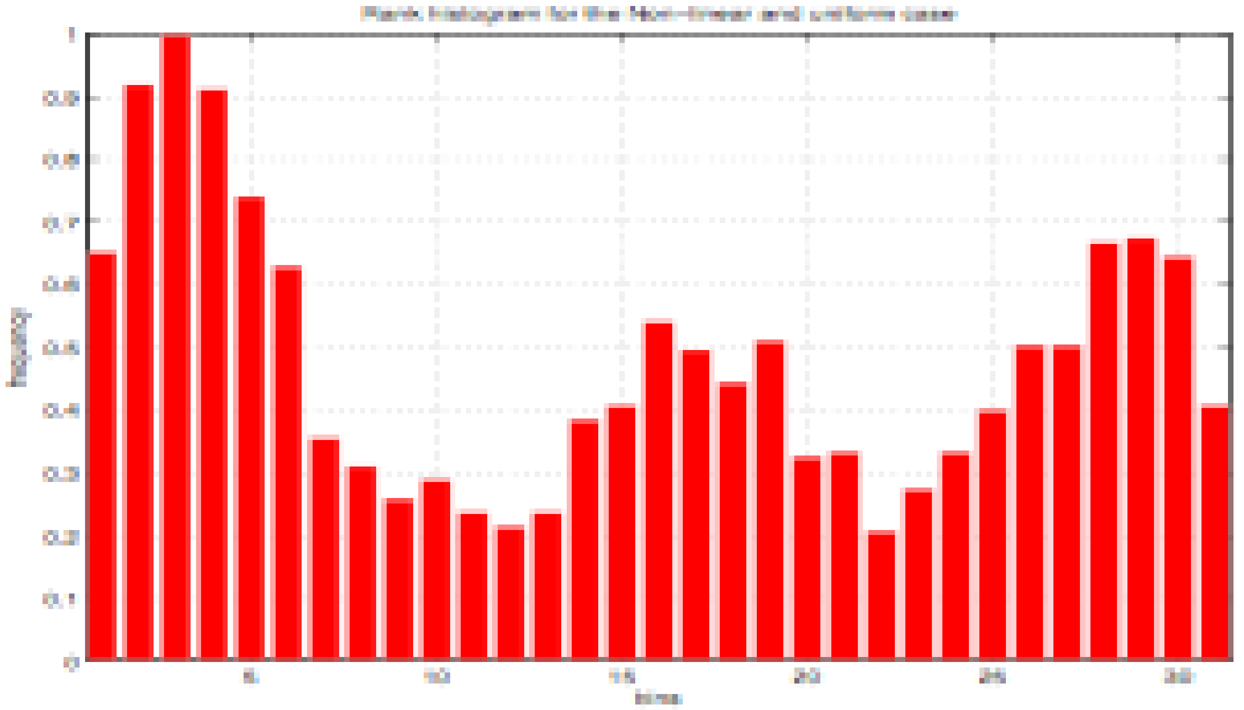
Brier Skill score: Lorenz96 model linear case



Ens/4D-Var: Lorenz96 model nonlinear case

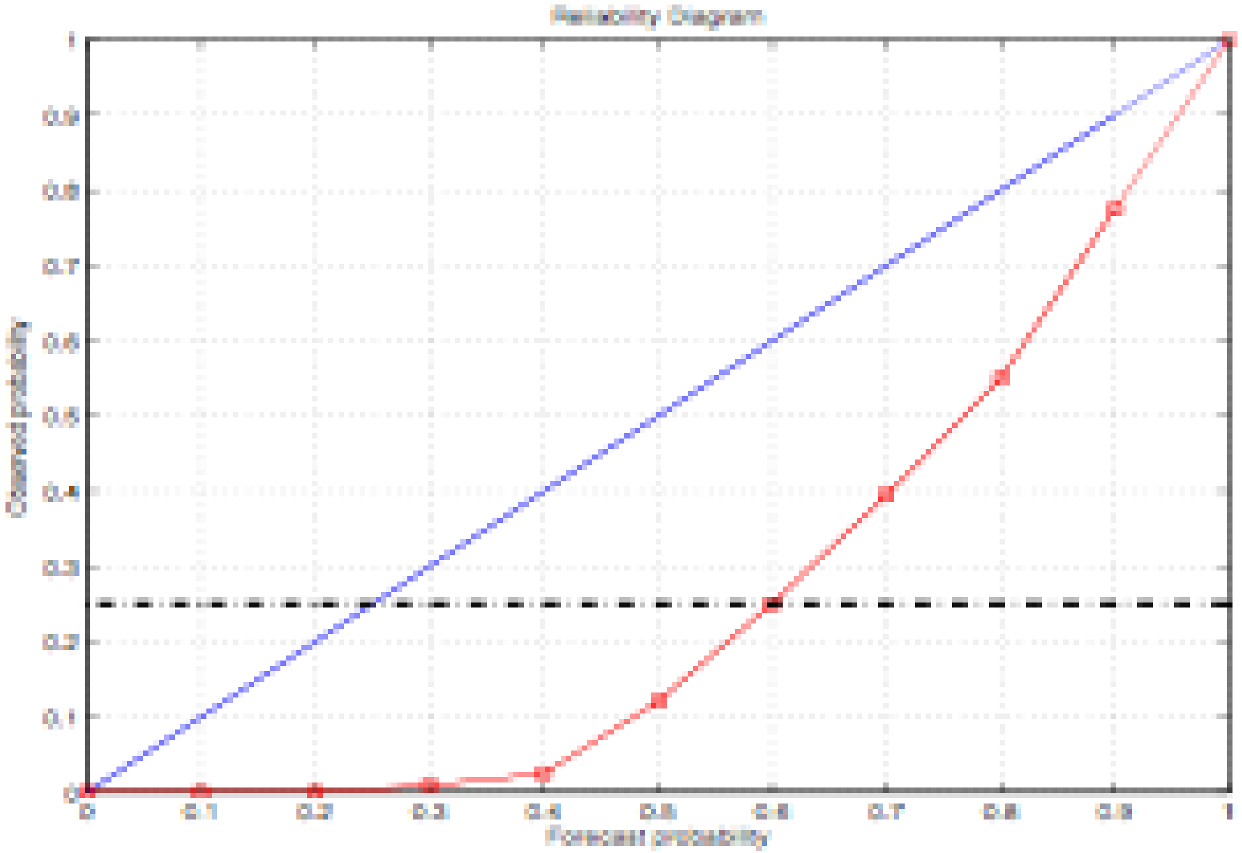


rank histogram: Lorenz96 model nonlinear case

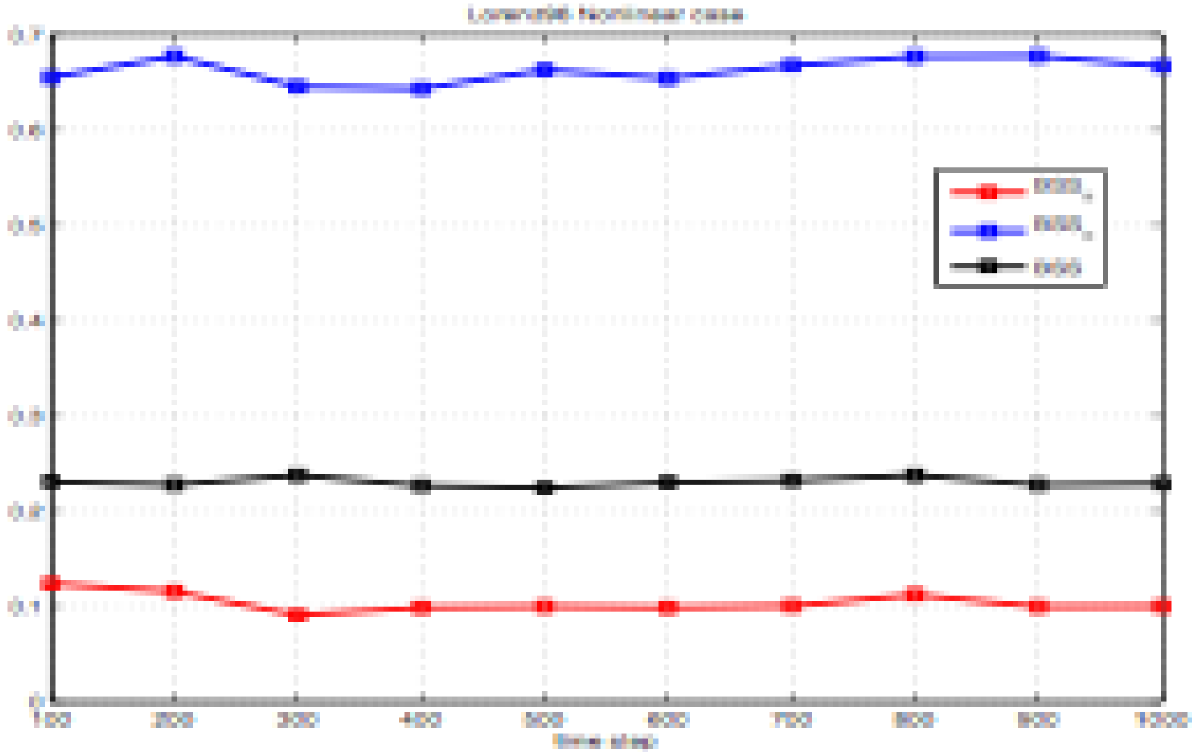


Reliability diagram: Lorenz96 model nonlinear case

1. Threshold $\tau = 0.5$.
2. 3000 realizations.



Brier Score: Lorenz96 model nonlinear case



Preliminary conclusions

- In the linear case, ensemble variational assimilation produces, in agreement with theory, reliable estimates of the state of the system.

- Nonlinearity significantly degrades reliability (and therefore bayesianity) of variational assimilation ensembles. Resolution (*i. e.*, capability of ensembles to reliably estimate a broad range of different probabilities of occurrence, is also degraded. Similar results (not shown) have been obtained with Particle Filter, which produces ensembles with low reliability.

Perspectives

- Perform further studies on physically more realistic dynamical models (shallow-water equations)

- Further comparison with performance Particle Filters, and with performance of Ensemble Kalman Filter.

Announcement

International Conference on Ensemble Methods in Geophysical Sciences

Supported by World Meteorological Organization, Météo-France, Agence Nationale de la Recherche (France)

Dates and location : 12-16 November 2012, Toulouse, France

Forum for study of all aspects of ensemble methods for estimation of state of geophysical systems (atmosphere, ocean, solid earth, …), in particular in assimilation of observations and prediction : theory, algorithmic implementation, practical applications, evaluation, …

Scientific Organizing Committee is being set up. Announcement will be widely circulated in coming weeks.