Filtre de Kalman et Méthodes d'Ensemble pour l'Assimilation d'Observations en Météorogie et Océanographie

Olivier Talagrand
Laboratoire de Météorologie Dynamique, Paris

3èmes Rencontres Météo/MathAppli
Météo-France, Toulouse, 19 Septembre 2008

Purpose of assimilation : reconstruct as accurately as possible the state of the atmosphere (the ocean, or whatever the system of interest is), using all available appropriate information. The latter essentially consists of

- The observations.

- The physical laws governing the system, 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 of middle latitudes. 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 (Jaynes, E. T., 2007, *Probability Theory: The Logic of Science*, Cambridge University Press).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (unambiguously defined if a prior probability distribution is defined; see Tarantola, 2005).

Bayesian estimation impossible in practice 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 $n \approx 10^{6\text{-}7}$ of present NWP models.

- Probability distribution of errors affecting data is very poorly known (errors in assimilating model).

How to define in practice a probability distribution in a very large dimensional space ?

Only possible way seems to be through a finite ensemble, meant to sample the distribution.

$\Rightarrow$ *Ensemble methods* (used also for prediction)

Typical size of ensembles in present meteorological applications : O(10-100)

Exist at present in two forms

- *Ensemble Kalman Filter (EnKF)*.
- *Particle filters.*

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad k = 0, \ldots, K$$

$$E(\varepsilon_k) = 0 \quad ; \quad E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) \equiv R_k\, \delta_{kj}$$

- Evolution equation

$$x_{k+1} = M_k x_k + \eta_k \qquad k = 0, \ldots, K\text{-}1$$

$$E(\eta_k) = 0 \quad ; \quad E(\eta_k \eta_j^{\mathrm{T}}) \equiv Q_k\, \delta_{kj}$$

$$E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$$

- Background estimate at time 0

$$x^b{}_0 = x_0 + \zeta^b{}_0$$

$$E(\zeta^b{}_0) = 0 \quad ; \quad E(\zeta^b{}_0\, \zeta^b{}_0{}^{\mathrm{T}}) \equiv P^b{}_0$$

$$E(\zeta^b{}_0 \varepsilon_k{}^{\mathrm{T}}) = 0 \quad ; \quad E(\zeta^b{}_0 \eta_k{}^{\mathrm{T}}) = 0$$

Sequential assimilation assumes the form of *Kalman filter*

Background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^T [H_k P^b_k H_k^T + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^T [H_k P^b_k H_k^T + R_k]^{-1} H_k P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k x^a_k$$
$$P^b_{k+1} = M_k P^a_k M_k^T + Q_k$$

Kalman Filter produces at every step k the *Best Linear Unbiased Estimate* (*BLUE*) of real unknown state x_k from all data prior to k. In addition, it achieves bayesian estimation when the errors (ε_k, η_k, ζ^b_0) are globally gaussian.

How to update predicted ensemble with new observations ?

Predicted ensemble at time t : $\{x^b_n\}$, $n = 1, \ldots, N$
Observation vector at same time : $y = Hx + \varepsilon$

- Gaussian approach

  Produce sample of probability distribution for real observed quantity Hx
  $y_n = y - \varepsilon_n$
  where ε_n is distributed according to probability distribution for observation error ε.

  Then use Kalman formula to produce sample of 'analysed' states

  $$x^a_n = x^b_n + P^b H^T [HP^bH^T + R]^{-1} (y_n - Hx^b_n) , \qquad n = 1, \ldots, N \qquad (2)$$

  where P^b is 'exact' (not sample) covariance matrix of predicted ensemble $\{x^b_n\}$.

  In the linear case, and if errors are gaussian, (2) achieves Bayesian estimation, in the sense that $\{x^a_n\}$ is a sample of conditional probability distribution for x, given all data up to time t.

Ensemble Kalman Filter (EnKF, Evensen, 1994, and many others)

- In the forecast phase, ensemble is evolved according to model equations (with possible inclusion of random noise to simulate effect of model errors).

- In the analysis phase, ensemble is updated according to procedure that has just been described (equation 2 on previous slide), the matrix P^b being now the sample covariance matrix of the background ensemble $\{x^b{}_n\}$.

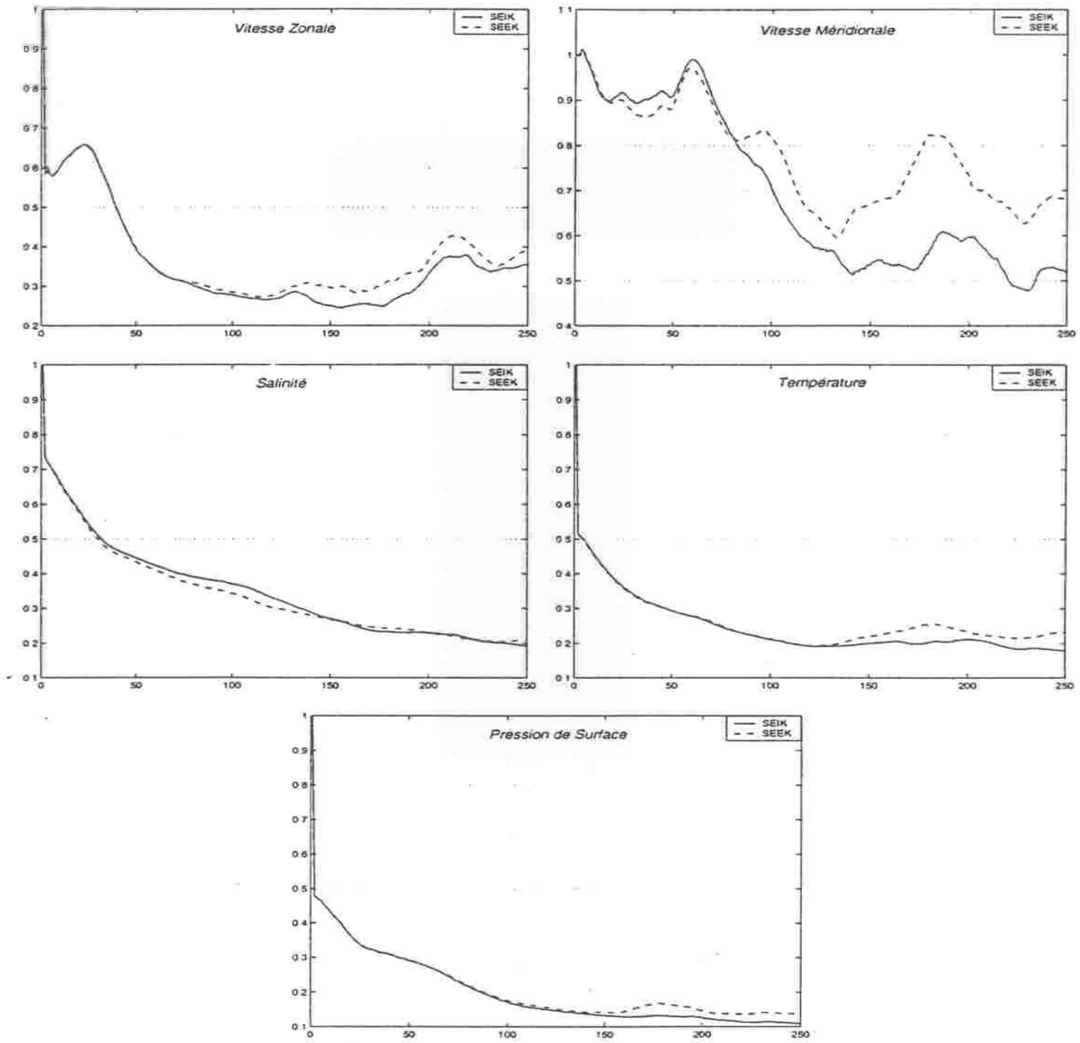


FIG. 5.2 – *Evolution dans le temps de la RRMS des filtres SEIK et SEEK*

I. Hoteit, Doctoral Dissertation, Université Joseph Fourier, Grenoble, 2001

Ensemble Kalman Filter (continuation)

Even if dynamical model is nonlinear, forecast phase is bayesian (provided errors are independent in time). Analysis phase will not in general because of

- Nonlinearity of observation operator

- Non-gaussianity of background and/or observation errors

- Sampling effects in P^b

Ensemble Kalman Filter is very commonly used in meteorological and oceanographical applications. Many variants exist, some of which do not require perturbations of the observations, but require previous analysis about which ensemble is evolved (Ensemble Transform Kalman Filter, ETKF, Bishop *et al*., 2001)

A general problem is *collapse of ensemble* in analysis phase. If dimension of ensemble is small (O(10-50)), spread of ensembles decreases in analysis. Since large ensembles are costly, *ad hoc* procedures are used to alleviate that effect :

- *Covariance inflation*. The spread of the ensemble about its mean is increased by an empirically determined numerical factor.

- '*Localization*'. Sampling effects in the background error covariance matrix create unrealistic correlations over large distances in physical space. These unrealistic correlations seem to contribute to the collapse of ensembles. They are eliminated by element-wise multiplication of the sample covariance matrix by another positive-definite matrix with compact support in physical space.

- *Double ensembles*. Two ensembles are evolved in parallel, the background error covariance matrix for updating either ensemble being determined from the other ensemble.

Origin of ensemble collapse ?

Ensemble collapse generally attributed to the fact that ensemble size N is small in comparison with state dimension n (10-100 against 10^{6-7}). In particular, corrections made by analysis on background are limited to a space with dimension N.

Descamps (2007) has observed that collapse occurs in small dimension (n=1) with $N>n$. Sampling effects in the background error covariance matrix play a role.

Exact bayesian estimation

Particle filters

Predicted ensemble at time t : $\{x^b{}_n, n = 1, \ldots, N\}$, each element with its own weight (probability) $P(x^b{}_n)$
Observation vector at same time : $y = Hx + \varepsilon$

Bayes' formula

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

Defines updating of weights

Remarks

- Many variants exist, including possible 'regeneration' of ensemble elements
- If errors are correlated in time, explicit computation of $P(y \mid x^b{}_n)$ will require using past data that are correlated with y (same remark for evolution of ensemble between two observation times)

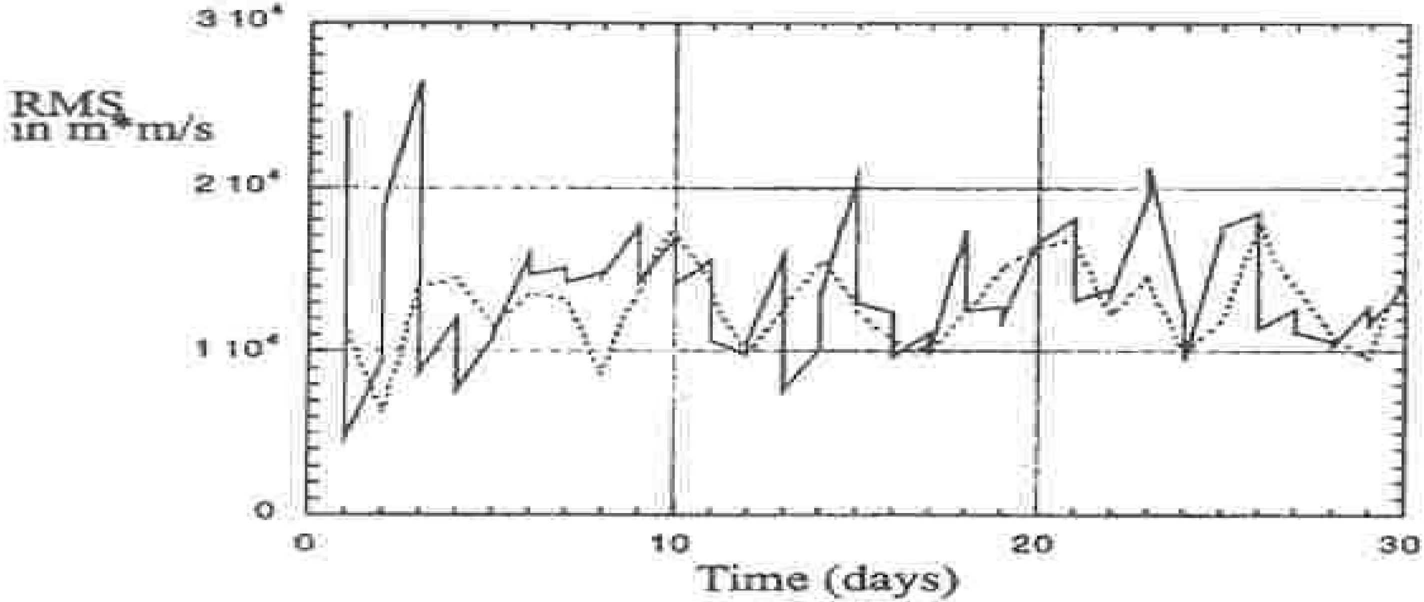


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

Exact bayesian estimation

Acceptation-rejection

Bayes' formula

$$f(x) \equiv \mathrm{P}(x \mid y) = P(y \mid x)\, P(x) \,/\, P(y)$$

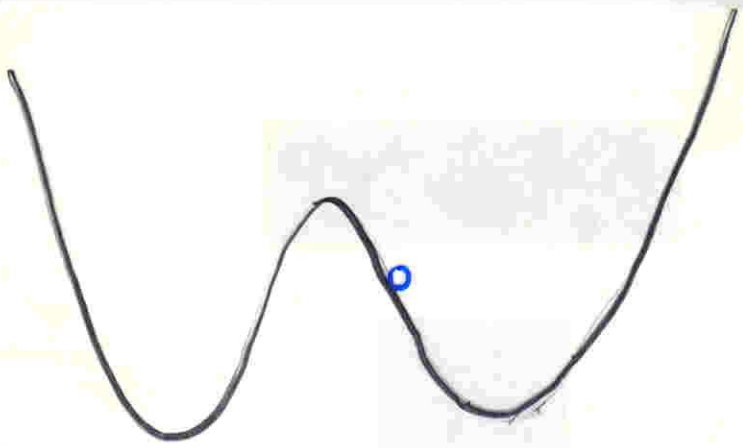
defines probability density function for x.

Construct sample of that pdf as follows.

Draw randomly couple $(\xi, \psi) \in S \times [0, \sup f]$.

Keep ξ if $\psi < f(\xi)$. ξ is then distributed according to $f(x)$.

Miller, Carter and Blue, Tellus, 1999



$$\frac{d^2x}{dt^2} = -\frac{d\phi}{dx} + \alpha\frac{dx}{dt} + \text{Noise}$$

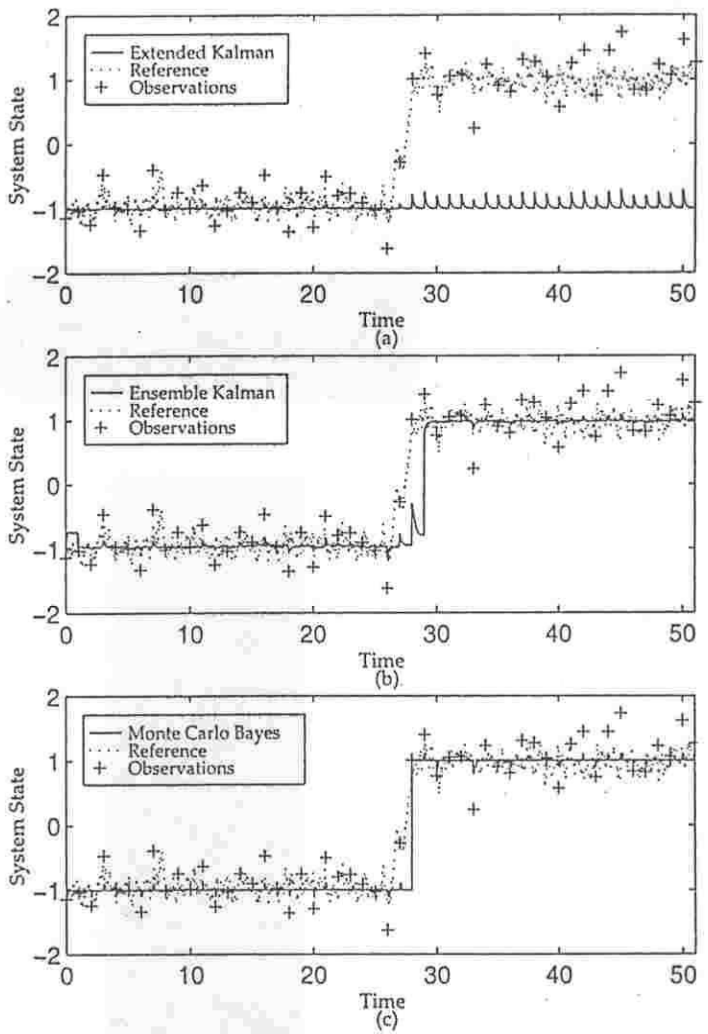


Fig. 4. Comparison of the EKF, the ensemble method and nonlinear filtering by Bayes' theorem for the double-well problem.

Miller, Carter and Blue, 1999, *Tellus*, **51A**, 167-194

Acceptation-rejection

Seems costly.

Requires capability of permanently interpolating probability distribution defined by finite sample to whole state space.

Time-correlated Errors

Sequential methods, whether of the Kalman or particle filter type cannot be Bayesian if errors are not independent in time. This extends to ‘smoothers’, in which updating by new observation is performed, not only on estimate at observation time, but also on estimates at previous times.

Time-correlated Errors

Example of time-correlated observation errors

$z_1 = x + \zeta_1$
$z_2 = x + \zeta_2$

$E(\zeta_1) = E(\zeta_2) = 0$; $E(\zeta_1^2) = E(\zeta_2^2) = s$; $E(\zeta_1\zeta_2) = 0$

BLUE of x from z_1 and z_2 gives equal weights to z_1 and z_2.

Additional observation then becomes available

$z_3 = x + \zeta_3$
$E(\zeta_3) = 0$; $E(\zeta_3^2) = s$; $E(\zeta_1\zeta_3) = cs$; $E(\zeta_2\zeta_3) = 0$

BLUE of x from (z_1, z_2, z_3) has weights in the proportion $(1, 1+c, 1)$

Time-correlated Errors (continuation 1)

Example of time-correlated model errors

Evolution equation

` $x_{k+1} = x_k + \eta_k \qquad E(\eta_k^2) = q$

Observations

$y_k = x_k + \varepsilon_k\,, \qquad k = 0, 1, 2 \qquad E(\varepsilon_k^2) = r$, errors uncorrelated in time

Sequential assimilation. Weights given to y_0 and y_1 in analysis at time 1 are in the ratio $r/(r+q)$. That ratio will be conserved in sequential assimilation. All right if model errors are uncorrelated in time.

Assume $E(\eta_0\eta_1) = cq$

Weights given to y_0 and y_1 in estimation of x_2 are in the ratio

Time-correlated Errors (continuation 2)

Moral. If data errors are correlated in time, it is not possible to discard observations as they are used while preserving optimality of the estimation process. In particular, if model error is correlated in time, all observations are liable to be reweighted as assimilation proceeds.

Variational assimilation can take time-correlated errors into account.

Example of time-correlated observation errors. Global covariance matrix

$$\boldsymbol{R} = \left(R_{kk'} = E(\varepsilon_k \varepsilon_{k'}^{\mathrm{T}})\right)$$

Objective function

$$\xi_0 \in \mathcal{S} \ \rightarrow$$

$$J(\xi_0) = (1/2)\,(x_0^b - \xi_0)^{\mathrm{T}}\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_{kk'}[y_k - H_k\xi_k]^{\mathrm{T}}\,[\boldsymbol{R}^{-1}]_{kk'}\,[y_{k'} - H_{k'}\xi_{k'}]$$

where $[\boldsymbol{R}^{-1}]_{kk'}$ is the kk'-subblock of global inverse matrix $\boldsymbol{R}^{-1}$.

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 3)

Time correlation of observational error has been introduced by ECMWF (Järvinen *et al.*, 1999) in variational assimilation of high-frequency surface pressure observations (correlation originates in that case in representativeness error).

Identification and quantification of temporal correlation of errors, especially model errors ?

Q. Is it possible to have at the same time the advantages of both ensemble estimation and variational assimilation (propagation of information both forward and backward in time, and, more importantly, possibility to take temporal dependence into account) ?

Same approach that underlies EnKF. Perturb all data (model and observations) according to the corresponding error probability distribution and, for each set of perturbed data, perform a variational assimilation. In the linear and gaussian case, this will produce a sample of conditional probability distribution for the orbit of the system, subject to the data.

Still to be done.

Evaluation of assimilation ensembles

Ensembles must be evaluated as descriptors of probability distributions (and not for instance on the basis of properties of individual elements). This implies, among others

- Validation of the expectation of the ensembles

- Validation of the spread *(spread-skill relationship)*

Reduced Centred Random Variable (RCRV, Candille *et al*., 2006)

For some scalar variable x, ensemble has mean μ and standard deviation σ. Ratio

$$s = \frac{\xi - \mu}{\sigma}$$

where ξ is verifying observation. Over a large number of realizations

$$E(s) = 0 \qquad , \qquad E(s^2) = 1$$

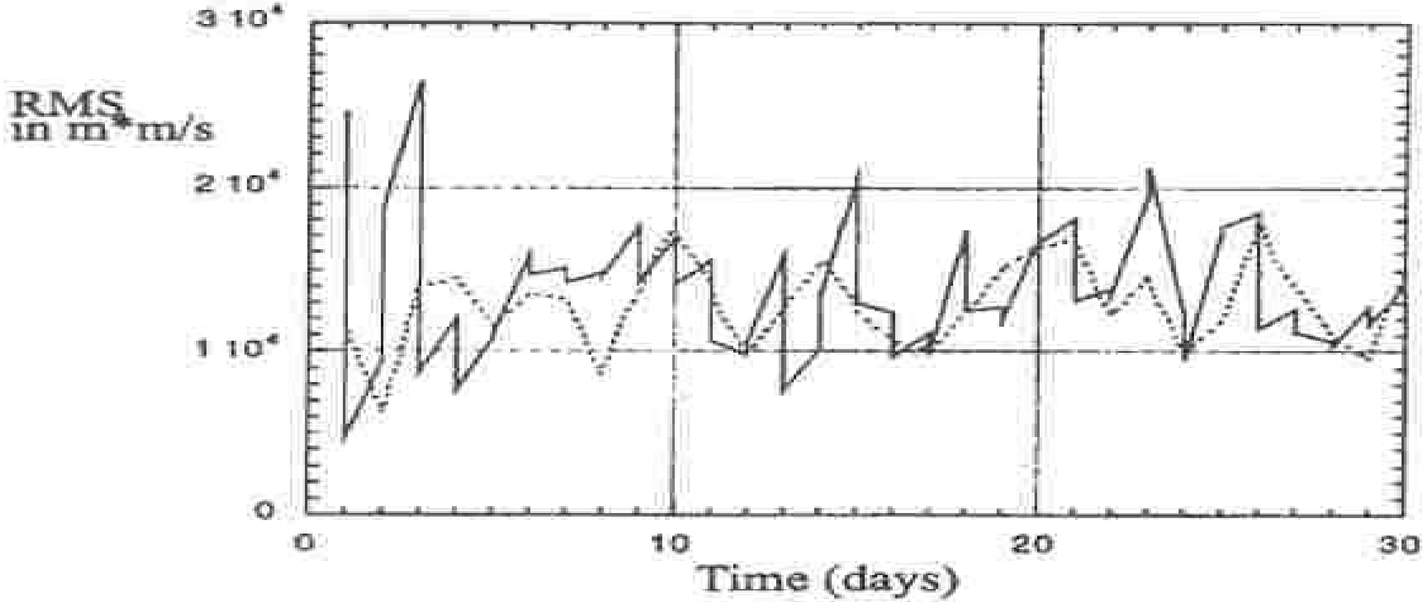


FIG. 12. Comparison of rms error ($m^2 s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

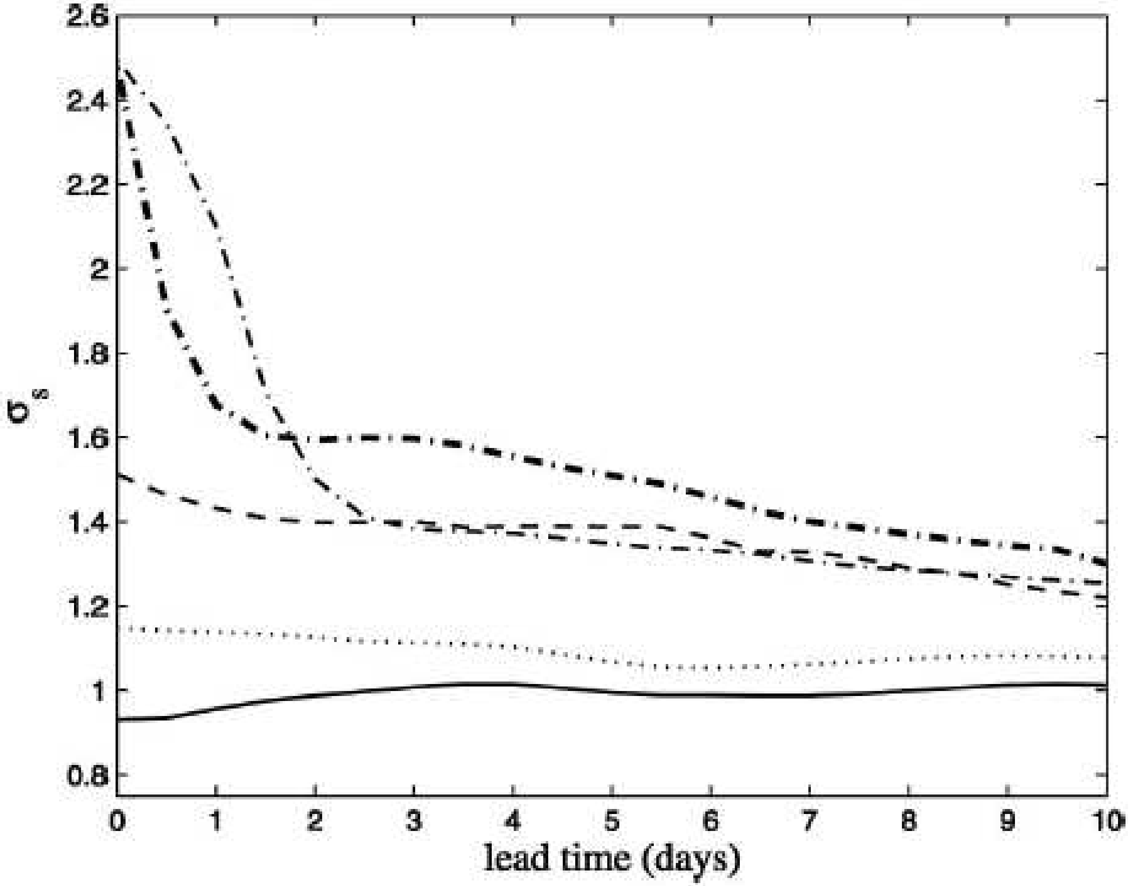


FIG. 4. Evolution of the std dev of the RCRV, as a function of lead time, for the four different methods: EnKF (solid line), ETKF (dotted line), BM (dashed line), and SV computed over a 24-h optimization period (heavy dashed–dotted line) and SV computed over a 48-h optimization period (thin dashed–dotted line).

Descamps and Talagrand, *Mon. Wea. Rev.*, 2007

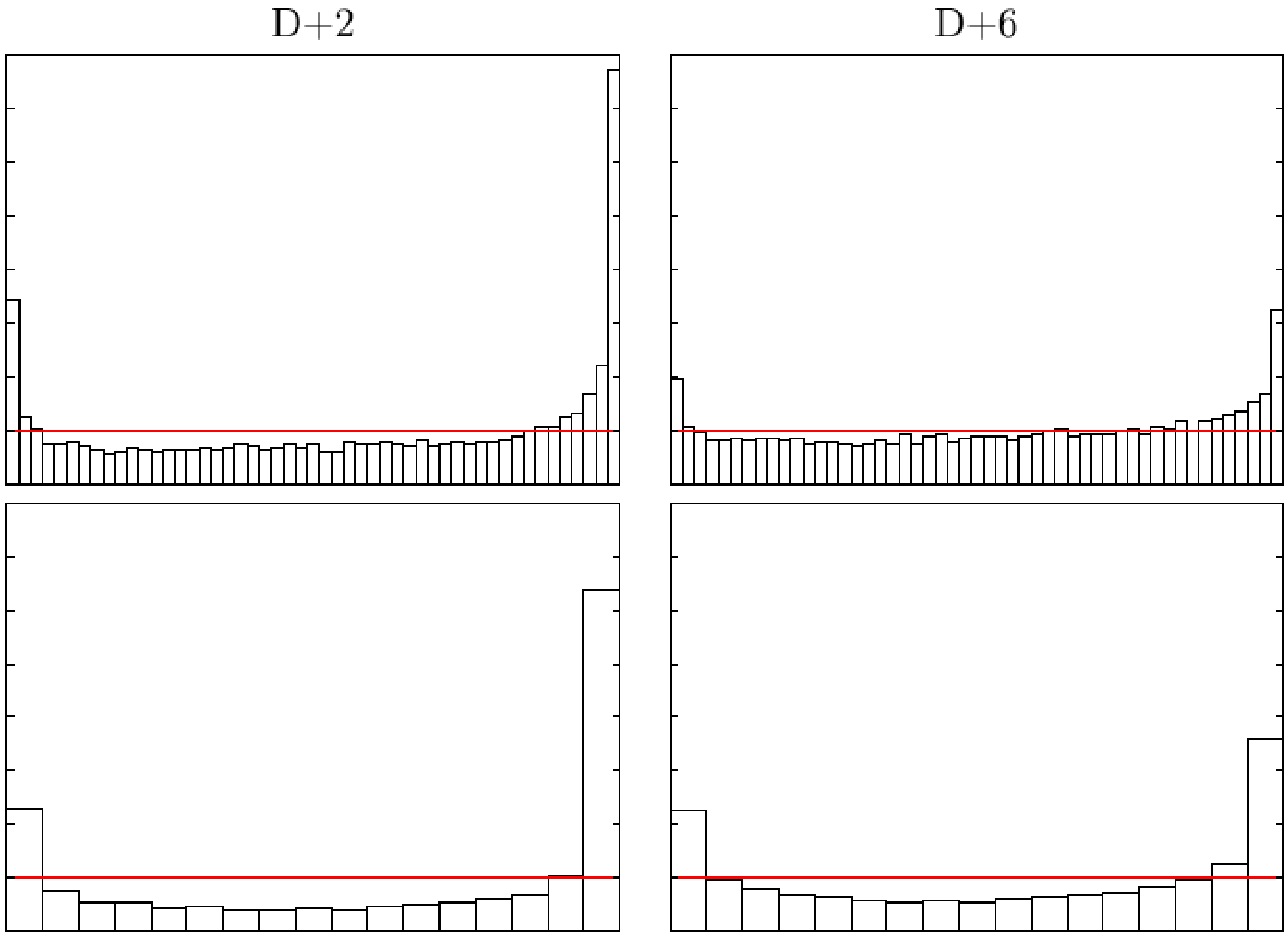
Rank Histograms

For some scalar variable x, N ensemble values, assumed to be N independent realizations of the same probability distribution, ranked in increasing order

$$x_1 < x_2 < \ldots < x_N$$

Define $N+1$ intervals.

If verifying observation ξ is an $N+1$st independent realization of the same probability distribution, it must be statistically undistinguishable from the x_i's. In particular, must be uniformly distributed among the $N+1$ intervals defined by the x_i's.



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

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

Two properties make the value of an ensemble estimation system (either for assimilation or for prediction)

Reliability is statistical consistency between estimated probability distributions and verifying observations. Is objectively and quantitatively measured by a number of standard diagnostics (among which Reduced Centred Random Variable and Rank Histograms, reliability component of Brier and Brier-like scores).

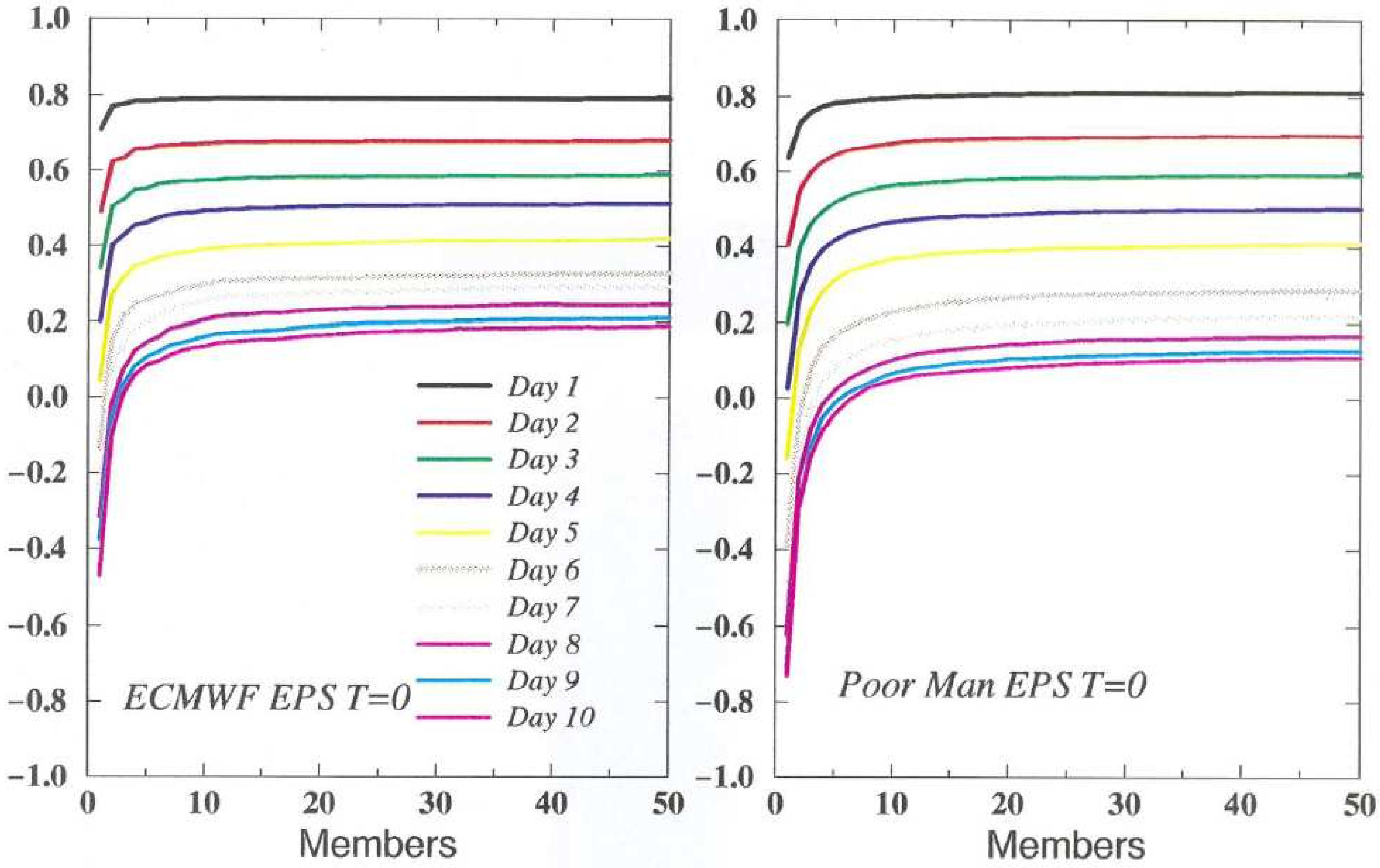
Resolution (aka *skewness*) is the property that reliably predicted probability distributions are useful (essentially have small spread). Also measured by a number of standard diagnostics (resolution component of Brier and Brier-like scores).

.

To-day's message. Evaluate assimilation ensembles in terms of reliability and resolution.

Size of Ensembles ?

o Observed fact : in ensemble prediction, present scores saturate for value of ensemble size N in the range 30-50, independently of quality of score.



Impact of ensemble size on Brier Skill Score
ECMWF, event $T_{850} > T_c$ Northern Hemisphere

(Talagrand *et al.*, ECMWF, 1999)

Theoretical estimate (raw Brier score)

Size of Ensembles (continued) ?

- o Relatively large ensembles seem to be necessary for numerical implementation of EnKF, but not for simply evolving probability distributions in a prediction process.

- o Q. What about particle filters ?