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

Année Universitaire 2016-2017

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 7

15 Juin 2017

In the linear case, and if errors are uncorrelated in time, Kalman Smoother and Variational Assimilation are algorithmically equivalent. They produce the *BLUE* of the state of the system from all available data, over the whole assimilation window (Kalman Filter produces the *BLUE* only at the end of the final time of the window). If in addition errors are Gaussian, both algorithms achieve Bayesian estimation.

Incremental Method

Variational assimilation, as it has been described, requires the use of the adjoint of the full model.

Simplifying the adjoint as such can be very dangerous. The computed gradient would not be exact, and experience shows that optimization algorithms (and especially efficient ones) are very sensitive to even slight misspecification of the gradient.

Principle of *Incremental Method* (Courtier *et al*., 1994, *Q. J. R. Meteorol. Soc*.) : simplify simultaneously the (local tangent linear) dynamics and the corresponding adjoint.

Incremental Method (continuation 1)

- Basic (nonlinear) model

$\xi_{k+1} = M_k(\xi_k)$

- Tangent linear model

$\delta\xi_{k+1} = M_k' \, \delta\xi_k$

where M_k' is jacobian of M_k at point ξ_k.

- Adjoint model

$\lambda_k = M_k'^{T} \lambda_{k+1} + \dots$

Incremental Method. Simplify M_k' and $M_k'^{T}$.

Incremental Method (continuation 2)

More precisely, for given solution $\xi_k^{(0)}$ of nonlinear model, replace tangent linear and adjoint models respectively by

$$\delta\xi_{k+1} = L_k\, \delta\xi_k \qquad (2)$$

and

$$\lambda_k = L_k^{\mathrm{T}}\, \lambda_{k+1} + \ldots$$

where L_k is an appropriate simplification of jacobian M_k'.

It is then necessary, in order to ensure that the result of the adjoint integration is the exact gradient of the objective function, to modify the basic model in such a way that the solution emanating from $\xi_0^{(0)} + \delta\xi_0$ is equal to $\xi_k^{(0)} + \delta\xi_k$, where $\delta\xi_k$ evolves according to (2). This makes the basic dynamics exactly linear.

Incremental Method (continuation 3)

As concerns the observation operators in the objective function, a similar procedure can be implemented if those operators are nonlinear. This leads to replacing $H_k(\xi_k)$ by $H_k(\xi_k^{(0)}) + N_k\delta\xi_k$, where N_k is an appropriate 'simple' linear operator (possibly, but not necessarily, the jacobian of H_k at point $\xi_k^{(0)}$). The objective function depends only on the initial $\delta\xi_0$ deviation from $\xi_0^{(0)}$, and reads

$$\mathcal{J}_I(\delta\xi_0) = (1/2)\,(x_0^b - \xi_0^{(0)} - \delta\xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0^{(0)} - \delta\xi_0) + (1/2)\,\Sigma_k[d_k - N_k\delta\xi_k]^T\,R_k^{-1}\,[d_k - N_k\delta\xi_k]$$

where $d_k \equiv y_k - H_k(\xi_k^{(0)})$ is the innovation at time k, and the $\delta\xi_k$ evolve according to

$$\delta\xi_{k+1} = L_k\,\delta\xi_k \qquad (2)$$

With the choices made here, $\mathcal{J}_I(\delta\xi_0)$ is an exactly quadratic function of $\delta\xi_0$. The minimizing perturbation $\delta\xi_{0,m}$ defines a new initial state $\xi_0^{(1)} \equiv \xi_0^{(0)} + \delta\xi_{0,m}$, from which a new solution $\xi_k^{(1)}$ of the basic nonlinear equation is determined. The process is restarted in the vicinity of that new solution.

Incremental Method (continuation 4)

This defines a system of two-level nested loops for minimization. Advantage is that many degrees of freedom are available for defining the simplified operators L_k and N_k, and for defining an appropriate trade-off between practical implementability and physical usefulness and accuracy. It is the incremental method which, together with the adjoint method, makes variational assimilation possible.

First-Guess-At-the-right-Time 3D-Var (*FGAT 3D-Var*). Corresponds to $L_k = I_n$. Assimilation is four-dimensional in that observations are compared to a first-guess which evolves in time, but is three-dimensional in that no dynamics other than the trivial dynamics expressed by the unit operator is present in the minimization.

Buehner *et al*. (*Mon. Wea. Rev*., 2010)

For the same numerical cost, and in meteorologically realistic situations, Ensemble Kalman Filter and Variational Assimilation produce results of similar quality.

How to take model error into account in variational assimilation ?

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 $k = 0, \ldots, K$

    $$y_k = H_k x_k + \varepsilon_k \qquad E(\varepsilon_k \varepsilon_k^T) = R_k$$

  - Model

    $$x_{k+1} = M_k x_k + \eta_k \qquad E(\eta_k \eta_k^T) = Q_k \qquad k = 0, \ldots, K\text{-}1$$

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

Then objective function

$$(\xi_0, \xi_1, \ldots, \xi_K) \to$$

$$\mathcal{J}(\xi_0, \xi_1, \ldots, \xi_K)$$

$$= (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0)$$

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

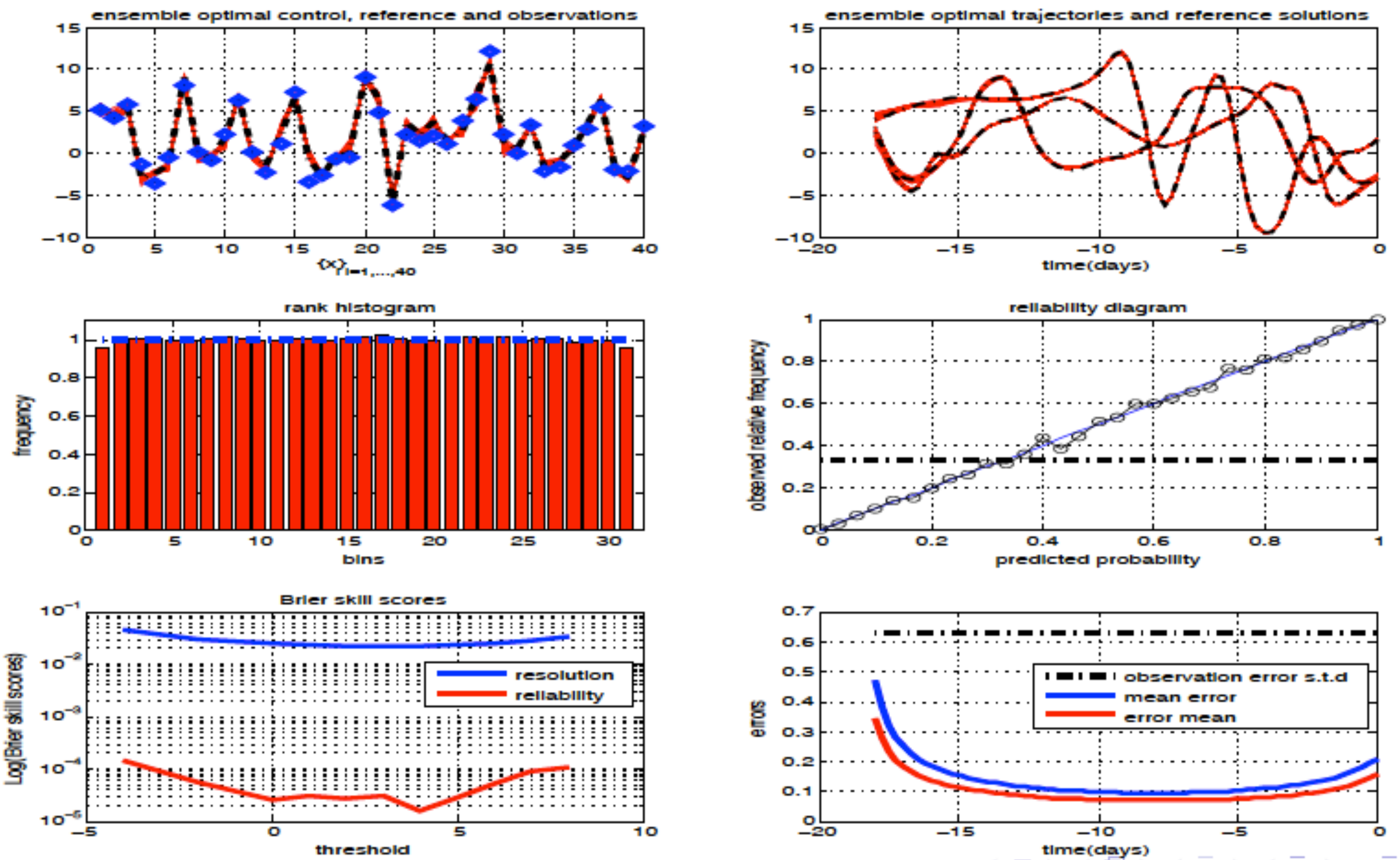
$$+ (1/2)\,\Sigma_{k=0,\ldots,K-1}[\xi_{k+1} - M_k\xi_k]^T\,Q_k^{-1}\,[\xi_{k+1} - M_k\xi_k]$$

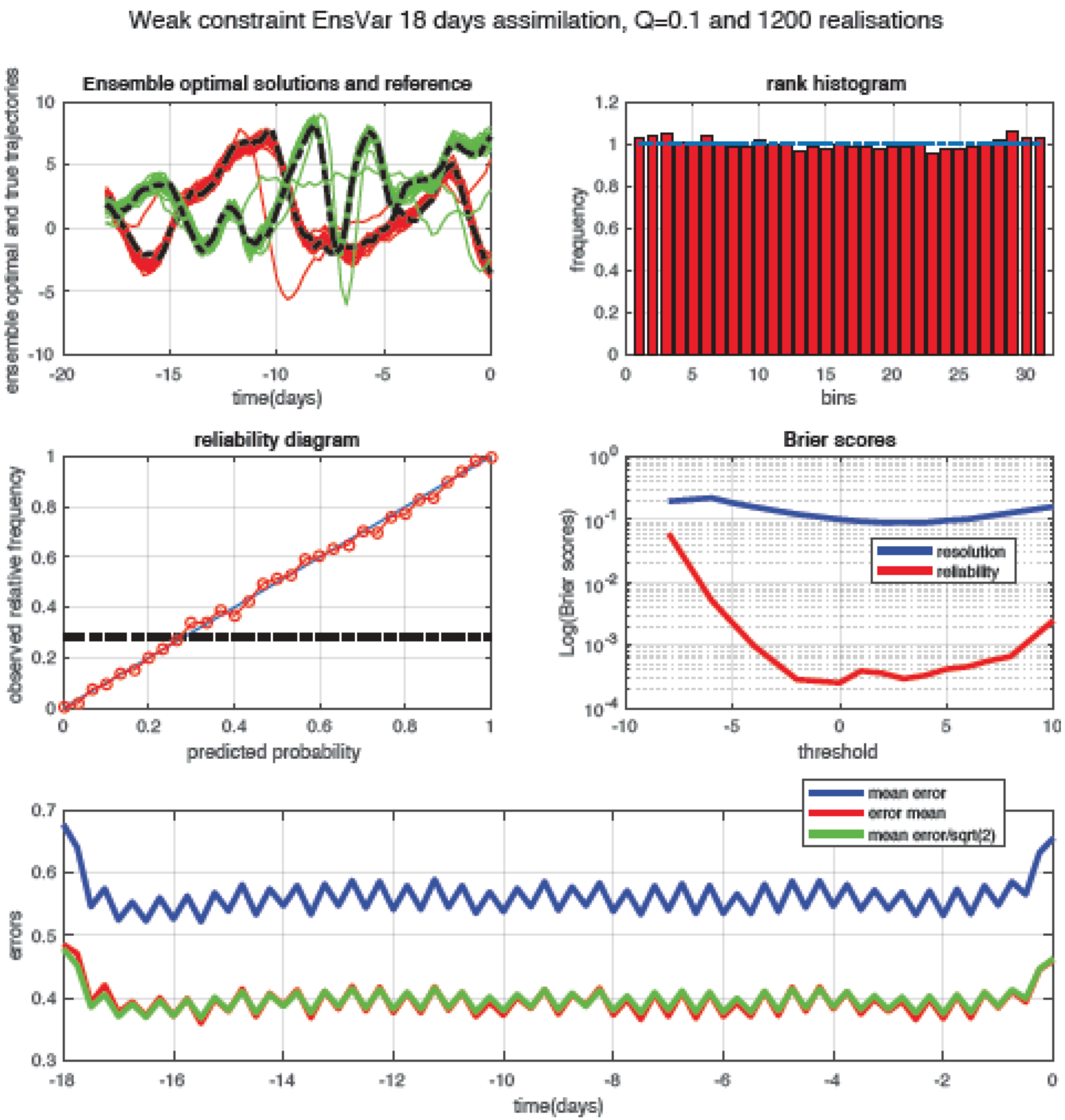
Can include nonlinear M_k and/or H_k.

Implemented operationally at ECMWF for the assimilation in the stratosphere.

Becomes singular in the strong constraint limit $Q_k \to 0$

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




Weak constraint EnsVar 18 days assimilation, Q=0.1 and 1200 realisations
Ensemble optimal solutions and reference
ensemble optimal and true trajectories
time(days)
rank histogram
frequency
bins
reliability diagram
observed relative frequency
predicted probability
Brier scores
Log(Brier scores)
threshold
resolution
reliability
mean error
error mean
mean error/sqrt(2)
errors
time(days)

Dual Algorithm for Variational Assimilation (aka *Physical Space Analysis System*, *PSAS*, pronounced *'pizzazz'*; see in particular book and papers by Bennett)

$$x^a = x^b + P^b H^T [HP^bH^T + R]^{-1} (y - Hx^b)$$

$$x^a = x^b + P^b H^T \Lambda^{-1} d = x^b + P^b H^T m$$

where $\Lambda \equiv HP^bH^T + R$, $d \equiv y - Hx^b$ and $m \equiv \Lambda^{-1} d$ maximises

$$\mu \rightarrow \mathcal{K}(\mu) = -(1/2)\, \mu^T \Lambda\, \mu + d^T \mu$$

Maximisation is performed in (dual of) observation space.

Dual Algorithm for Variational Assimilation (continuation 2)

Extends to time dimension, and to weak-constraint case, by defining state vector as

$$x \equiv (x_0^{\mathrm{T}}, x_1^{\mathrm{T}}, \ldots, x_K^{\mathrm{T}})^{\mathrm{T}}$$

or, equivalently, but more conveniently, as

$$x \equiv (x_0^{\mathrm{T}}, \eta_0^{\mathrm{T}}, \ldots, \eta_{K-1}^{\mathrm{T}})^{\mathrm{T}}$$

where, as before

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

The background for x_0 is x_0^b, the background for η_k is 0. Complete background is

$$x^b = (x_0^{b\mathrm{T}}, 0^{\mathrm{T}}, \ldots, 0^{\mathrm{T}})^{\mathrm{T}}$$

It is associated with error covariance matrix

$$P^b = \mathrm{diag}(P_0^b, Q_0, \ldots, Q_{K\text{-}1})$$

Dual Algorithm for Variational Assimilation (continuation 3)

Define global observation vector as

$$y \equiv (y_0^{\mathrm{T}}, y_1^{\mathrm{T}}, \ldots, y_K^{\mathrm{T}})^{\mathrm{T}}$$

and global innovation vector as

$$d \equiv (d_0^{\mathrm{T}}, d_1^{\mathrm{T}}, \ldots, d_K^{\mathrm{T}})^{\mathrm{T}}$$

where $d_k \equiv y_k - H_k\, x_k^b$, with $x_{k+1}^b \equiv M_k x_k^b$, $\quad k = 0, \ldots, K\text{-}1$

Dual Algorithm for Variational Assimilation (continuation 4)

For any state vector $\xi = (\xi_0^{\mathrm{T}}, \upsilon_0^{\mathrm{T}}, \ldots, \upsilon_{K-1}^{\mathrm{T}})^{\mathrm{T}}$, the observation operator H

$$\xi \rightarrow H\xi = (u_0^{\mathrm{T}}, \ldots, u_K^{\mathrm{T}})^{\mathrm{T}}$$

is defined by the sequence of operations

$$u_0 = H_0 \xi_0$$

then for $k = 0, \ldots, K\text{-}1$

$$\xi_{k+1} = M_k \xi_k + \upsilon_k$$
$$u_{k+1} = H_{k+1} \xi_{k+1}$$

The observation error covariance matrix is equal to

$$R = \mathrm{diag}(R_0, \ldots, R_K)$$

Dual Algorithm for Variational Assimilation (continuation 5)

Maximization of dual objective function

$$\mu \rightarrow \mathcal{K}(\mu) = -(1/2)\ \mu^T \Lambda\ \mu + d^T\mu$$

requires explicit repeated computations of its gradient

$$\nabla_\mu \mathcal{K} = -\ \Lambda\mu + d = -\ (HP^bH^T + R)\mu + d$$

Starting from $\mu = (\mu_0{}^T, \ldots, \mu_K{}^T)^T$ belonging to (dual) of observation space, this requires 5 successive steps

- Step 1. Multiplication by H^T. This is done by applying the transpose of the process defined above, *viz.*,

Set $\chi_K = 0$

Then, for $k = K\text{-}1, \ldots, 0$

$$\nu_k = \chi_{k+1} + H_{k+1}{}^T \mu_{k+1}$$
$$\chi_k = M_k{}^T \nu_k$$

Finally $\lambda_0 = \chi_0 + H_0{}^T \mu_0$

The output of this step, which includes a backward integration of the adjoint model, is the vector $(\lambda_0{}^T, \nu_0{}^T, \ldots, \nu_{K\text{-}1}{}^T)^T$

Dual Algorithm for Variational Assimilation (continuation 6)

- Step 2. Multiplication by P^b. This reduces to

$$\xi_0 = P_0^b \lambda_0$$
$$\upsilon_k = Q_k \nu_k \quad , \quad k = 0, \ldots, K\text{-}1$$

- Step 3. Multiplication by H. Apply the process defined above on the vector $(\xi_0^{\mathrm{T}}, \upsilon_0^{\mathrm{T}}, \ldots, \upsilon_{K-1}^{\mathrm{T}})^{\mathrm{T}}$, thereby producing vector $(u_0^{\mathrm{T}}, \ldots, u_K^{\mathrm{T}})^{\mathrm{T}}$.

- Step 4. Add vector $R\mu$, *i. e.* compute

$$\varphi_0 = \xi_0 + R_0 \mu_0$$
$$\varphi_k = \upsilon_{k\text{-}1} + R_k \mu_k \quad , \quad k = 1, \ldots,$$

- Step 5. Change sign of vector $\varphi = (\varphi_0^{\mathrm{T}}, \ldots, \varphi_K^{\mathrm{T}})^{\mathrm{T}}$, and add vector $d = y - Hx^b$,

Dual Algorithm for Variational Assimilation (continuation 7)

Dual algorithm remains regular in the limit of vanishing model error. Can be used for both strong- and weak-constraint assimilation.

No significant increase of computing cost in comparison with standard strong constraint variational assimilation (Courtier, Louvel)

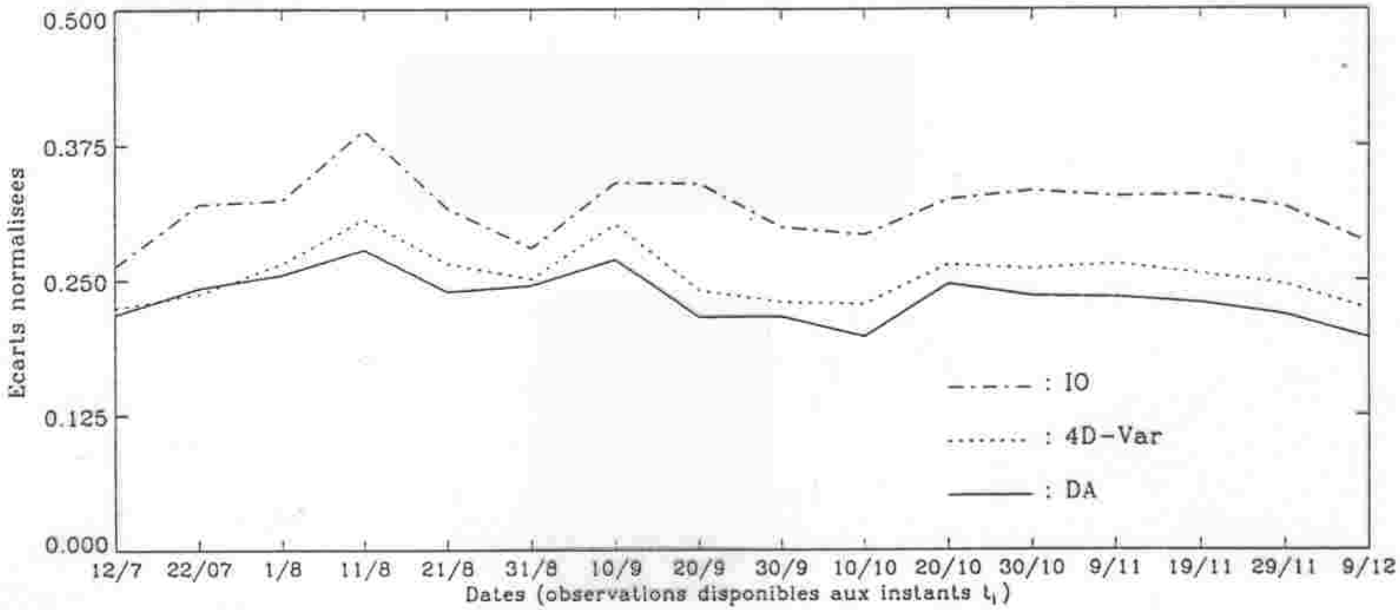


FIG. 9.11 – *Ecarts normalisés prévision/observations sur l'ensemble de la période étudiée*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999 nous permet d'observer les performances des différentes techniques d'assimilation.

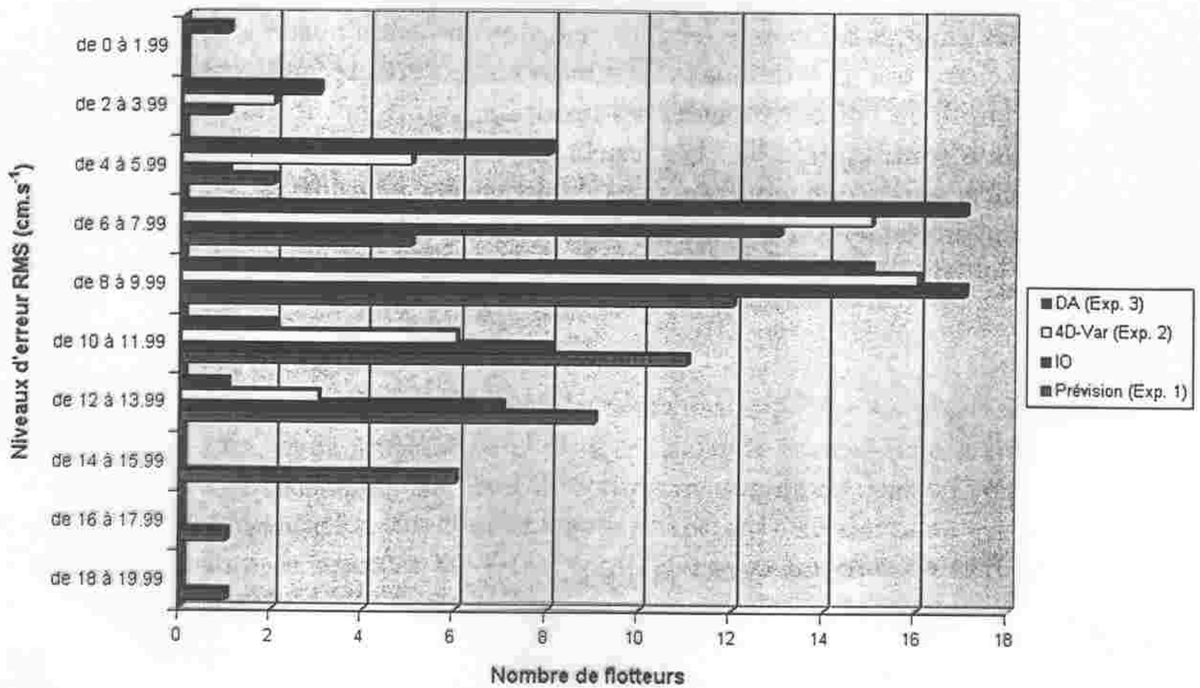


FIG. 9.15 – *Description des écarts flotteurs/modèle en terme de vitesse (à 150 m de profondeur) pour les différents algorithmes d'assimilation*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999

Dual Algorithm for Variational Assimilation (continuation)

Requires

- Explicit background (not much of a problem)
- Exact linearity (much more of a problem). Definition of iterative nonlinear procedures is being studied (Auroux, …)

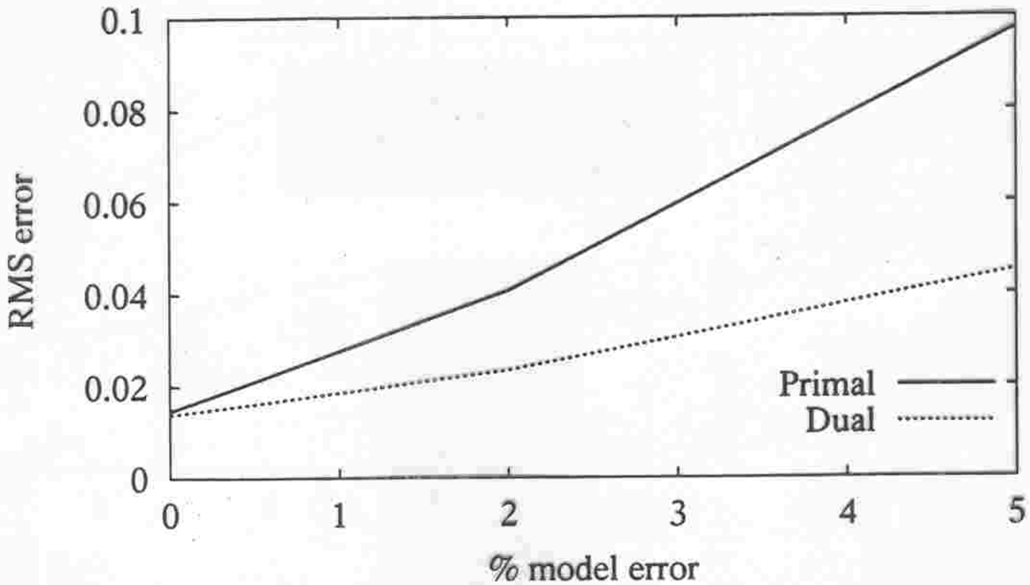


FIG. 6.13 – Normes RMS des erreurs d'assimilation obtenues pour les deux méthodes en fonction de l'erreur introduite dans le modèle au cours de la période d'assimilation.

Auroux, Doctoral Dissertation, Université de Nice-Sophia Antipolis, Nice, 2003

Variational assimilation has been extended to non Gaussian probability distributions (lognormal distributions), the unknown being the mode of the conditional distribution (M. Zupanski, Fletcher).

Bayesian character of variational assimilation ?

- If everything is linear and gaussian, ready recipe for obtaining bayesian sample

Perturb data (background, observations and model) according to their error probability distributions, do variational assimilation, and repeat process

Sample of system orbits thus obtained is bayesian

- If not, very little can be said at present

Time-correlated Errors

Example of time-correlated observation errors

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

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

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

Additional observation then becomes available

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

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

$$\rho = \frac{r - qc}{r + q + qc}$$

Conclusion

Sequential assimilation, in which data are processed by batches, the data of one batch being discarded once that batch has been used, cannot be optimal if data in different batches are affected with correlated errors. ***This is so even if one keeps trace of the correlations.***

Solution

Process all correlated in the same batch (4DVar, some smoothers)

Time-correlated Errors (continuation 3)

Moral. If data errors are correlated in time, it is not possible to discard observations as they are used. 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

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

Objective function

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

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

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

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 4)

Temporal 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 time correlation of errors, especially model errors ?

Conclusion on Sequential Assimilation

Pros

'Natural', and well adapted to many practical situations

Provides, at least relatively easily, explicit estimate of estimation error

Cons

Carries information only forward in time (of no importance if one is interested only in doing forecast)

In a strictly sequential assimilation (*i.e.*, any individual piece of information is discarded once it has been used), optimality is possible only if errors are independent in time.

Conclusion on Variational Assimilation

Pros

Carries information both forward and backward in time (important for reassimilation of past data).

Can easily take into account temporal statistical dependence (Järvinen *et al.*)

Does not require explicit computation of temporal evolution of estimation error

Very well adapted to some specific problems (*e. g.*, identification of tracer sources)

Cons

Does not readily provide estimate of estimation error

Requires development and maintenance of adjoint codes. But the latter can have other uses (sensitivity studies).

- Dual approach seems most promising. But still needs further development for application in non exactly linear cases.

- Is ensemble variational assimilation possible ? Probably yes. But also needs development.

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

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

Defines updating of weights

Bayes' formula

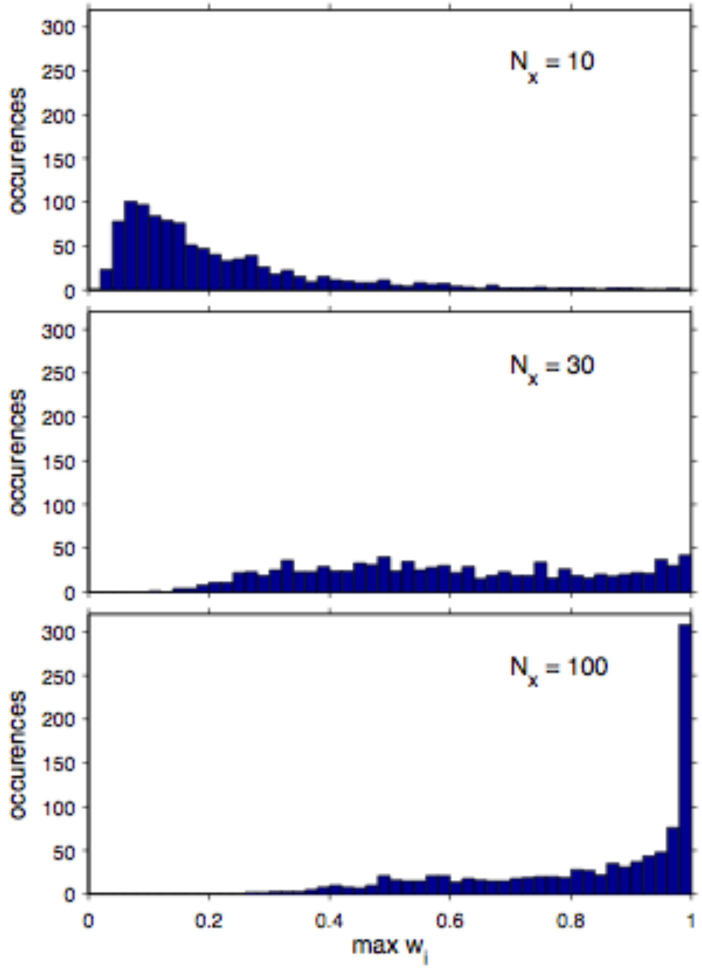
$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\ P(x^b{}_n)$$

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.

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.

Curse of dimensionality

Standard one-dimensional gaussian random variable X

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

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

.

Alternative possibilities (review in van Leeuwen, 2009, *Mon. Wea. Rev.*, 4089-4114)

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_n are multiplied by ensemble dimension N. Then p copies of each particle n are taken, where p is the integer part of Nw_n. Remaining particles, if needed, are taken randomly from the resulting distribution.

Importance Sampling.

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). This however leads to using twice the same observations.

In particular, *Guided Sequential Importance Sampling* (van Leeuwen, 2002). Idea : use observations performed at time k to resample ensemble at some timestep anterior to k, or 'nudge' integration between times k-1 and k towards observation at time k.

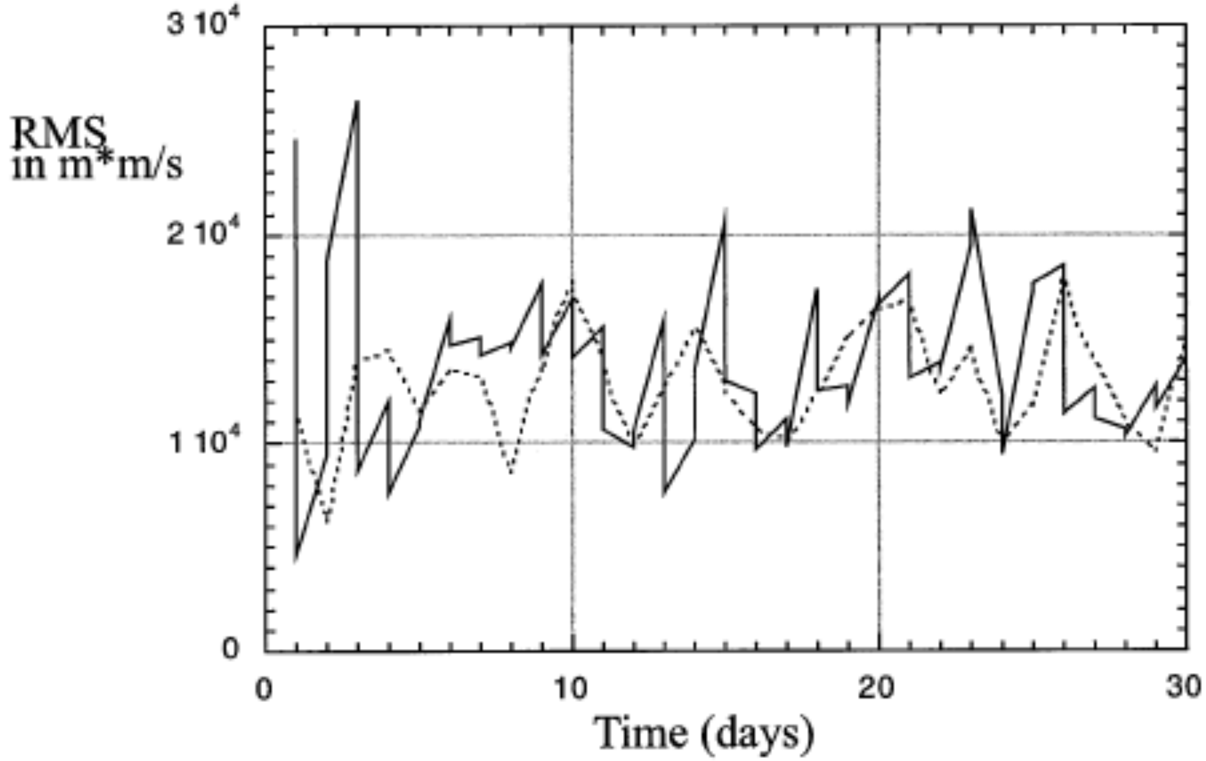


FIG. 12. Comparison of rms error ($m^2 s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

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

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

Cours à venir

~~Jeudi 6 avril~~

~~Jeudi 13 avril~~

~~Jeudi 20 avril~~

~~Jeudi 11 mai~~

~~**Lundi** 29 mai~~

~~Jeudi 1 juin~~

~~Jeudi 15 juin~~

Jeudi 22 juin

De 10h00 à 12h30, Salle de la Serre, 5ième étage, Département de Géosciences, École Normale Supérieure, 24, rue Lhomond, Paris 5