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 5

7 Mai 2021

- Kalman Filter. Continuation

- Ensemble Kalman Filter.

- Kalman Smoother. Brief theory. An example.

- Variational assimilation. Principle. The adjoint approach. Results

Sequential Assimilation. *Kalman Filter*

- System evolves in time according to equation

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k + \boldsymbol{\eta}_k \qquad k = 0, \ldots, K\text{-}1$$

$$E(\boldsymbol{\eta}_k) = 0 \quad ; \quad E(\boldsymbol{\eta}_k \boldsymbol{\eta}_j^{\mathrm{T}}) = \mathbf{Q}_k\, \delta_{kj}$$

$\mathbf{M}_k$ linear

- Observation vector at time k

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\varepsilon}_k \qquad k = 0, \ldots, K$$

$$E(\boldsymbol{\varepsilon}_k) = 0 \quad ; \quad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = \mathbf{R}_k\, \delta_{kj}$$

$\mathbf{H}_k$ linear

- $E(\boldsymbol{\eta}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background $\boldsymbol{x}^b_k$ and associated error covariance matrix $\boldsymbol{P}^b_k$ known

- Analysis step

$$\boldsymbol{x}^a_k = \boldsymbol{x}^b_k + \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} (\boldsymbol{y}_k - \boldsymbol{H}_k \boldsymbol{x}^b_k)$$
$$\boldsymbol{P}^a_k = \boldsymbol{P}^b_k - \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{P}^b_k \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k \boldsymbol{P}^b_k$$

- Forecast step

$$\boldsymbol{x}^b_{k+1} = \boldsymbol{M}_k \boldsymbol{x}^a_k$$
$$\boldsymbol{P}^b_{k+1} = \boldsymbol{M}_k \boldsymbol{P}^a_k \boldsymbol{M}_k^{\mathrm{T}} + \boldsymbol{Q}_k$$

Kalman filter (KF, Kalman, 1960)

Must be started from some initial estimate $(\boldsymbol{x}^b_0, \boldsymbol{P}^b_0)$

If all operators are linear, and if errors are uncorrelated in time, Kalman filter produces at time k the *BLUE* $\boldsymbol{x}^b{}_k$ (resp. $\boldsymbol{x}^a{}_k$) of the real state $\boldsymbol{x}_k$ from all data prior to (resp. up to) time k, plus the associated estimation error covariance matrix $\boldsymbol{P}^b{}_k$ (resp. $\boldsymbol{P}^a{}_k$).

If in addition errors are globally gaussian, the corresponding conditional probability distributions are the respective gaussian distributions $\mathcal{N}[\boldsymbol{x}^b{}_k, \boldsymbol{P}^b{}_k]$ and $\mathcal{N}[\boldsymbol{x}^a{}_k, \boldsymbol{P}^a{}_k]$.

Kalman filter. A simple example (Ghil *et al.*)

Shallow-water equations (aka *équations de Saint-Venant*)

$$\frac{\partial \varphi}{\partial t} + div(\varphi \boldsymbol{U}) = 0$$

$$\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{t}} + grad(\varphi + \frac{1}{2}\boldsymbol{U}^2) + k \wedge (f + \zeta)\boldsymbol{U} = 0$$

Periodic domain *D*. Equations conserve energy

$$E \equiv \frac{1}{2}\int_D (\varphi^2 + \varphi \boldsymbol{U}^2) dS$$

Equations linearized in the vicinity of state of rest
($\varphi = \Phi_0$, $\boldsymbol{U} = 0$)

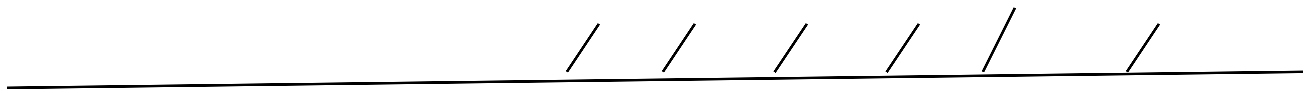
$$\frac{\partial \varphi}{\partial t} + \Phi_0 div \boldsymbol{U} = 0$$

$$\frac{\partial \boldsymbol{U}}{\partial t} + grad\varphi + k \wedge f\boldsymbol{U} = 0$$

Conserve quadratic energy

$$E \equiv \frac{1}{2} \int_D (\varphi^2 + \Phi_0 U^2) dS$$

Unidimensional domain

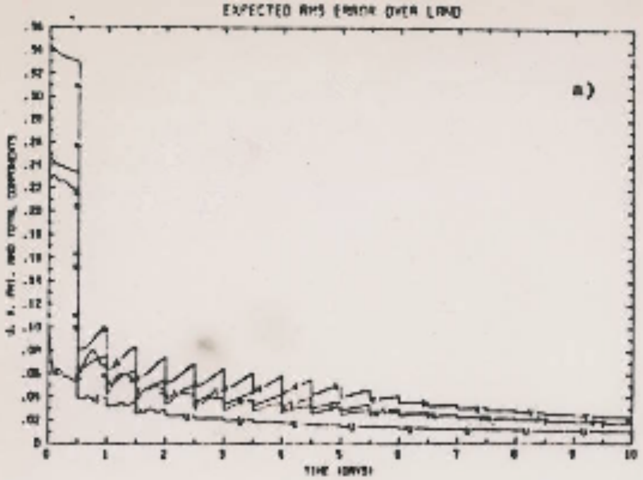


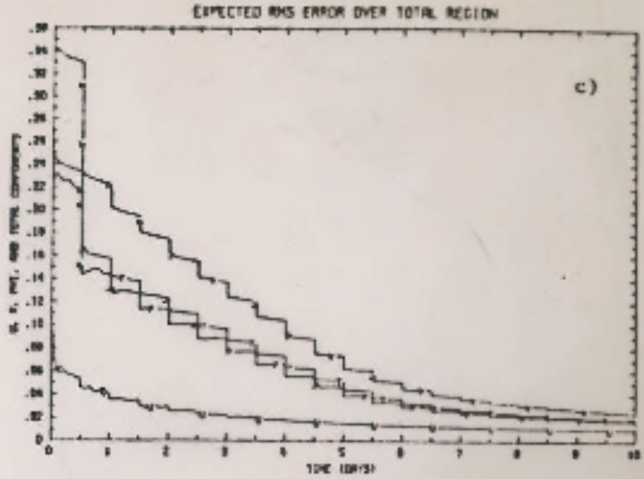
‘Ocean’

(no observation)

‘Continent’

(observations)





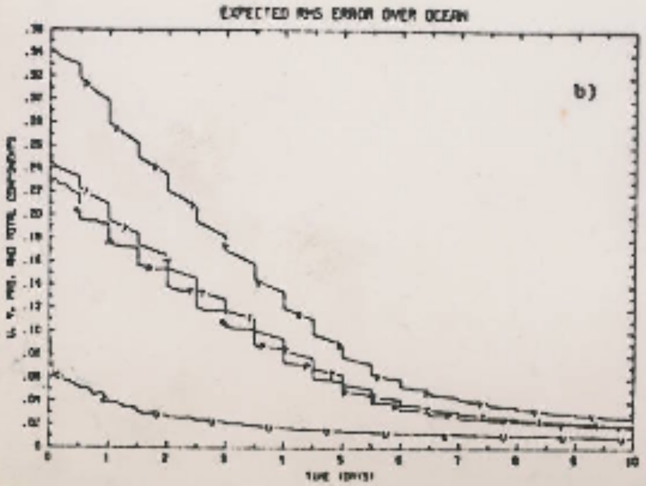


Fig. 2

The components of the total expected rms error (Erms), (trace: $P_k)^{1/2}$, in the estimation of solutions to the stochastic-dynamic system (Ψ, H), with Ψ given by (3.6) and H = (I 0). System noise is absent, Q = 0. The filter used is the standard K-B filter (2.11) for the model.

a) Erms over land; b) Erms over the ocean; c) Erms over the entire L-domain

In each one of the figures, each curve represents one component of the total Erms error. The curves labelled U, V, and P represent the u component, v component and ϕ component, respectively. They are found by summing the diagonal elements of P_k which correspond to u, v, and ϕ, respectively, dividing by the number of terms in the sum, and then taking the square root. In a) the summation extends over land points only, in b) over ocean points only, and in c) over the entire L-domain. The vertical axis is scaled in such a way that 1.0 corresponds to an Erms error of v_{max} for the U and V curves, and of ϕ_0 for the P curve. The observational error level is 0.089 for the U and V curves, and 0.080 for the P curve. The curves labelled T represent the total Erms error over each region. Each T curve is a weighted average of the corresponding U, V, and P curves, with the weights chosen in such a way that the T curve measures the error in the total energy $u^2 + v^2 + \phi^2/\bar{\phi}$, conserved by the system (3.1). The observational noise level for the T curve is then 0.088. Notice the immediate error decrease over land and the gradual decrease over the ocean. The total estimation error tends to zero.

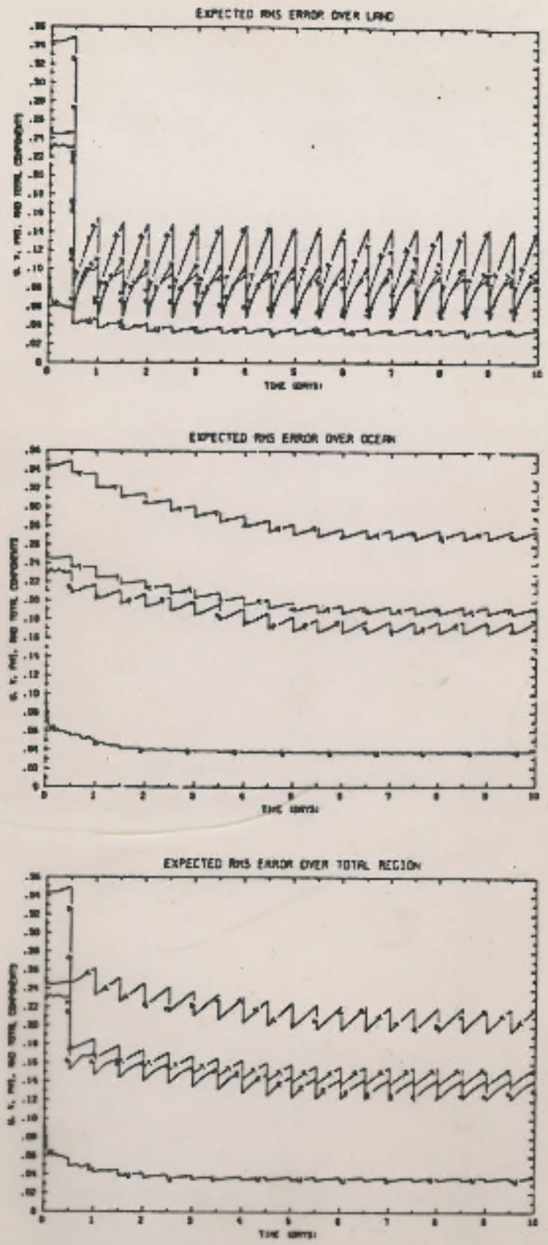


Fig. 6 This figure and the following ones show the properties of the estimated algorithm (2.11) in the presence of system noise, $Q \neq 0$. This figure gives the Erms estimation error, and is homologous to Fig. 2. Notice the sharper increase of error over land between synoptic times, and the convergence of each curve to a periodic, nonzero function.

Uncertainty evolves in time under the effect of

- Introduction of observations (decreases uncertainty)

- Model error (increases uncertainty)

- Dynamics of the system (increases or decreases uncertainty depending on stability of the state of the system) (dynamics is neutral in previous example)

Nonlinearities ?

Linearity of observation and model operators have been explicitly used in

$\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{x}^b = \boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}^b) + \boldsymbol{\varepsilon} = - \boldsymbol{H}\boldsymbol{\zeta}^b + \boldsymbol{\varepsilon}$

$\boldsymbol{M}_k \boldsymbol{x}^a{}_k - \boldsymbol{M}_k \boldsymbol{x}_k = \boldsymbol{M}_k(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)$

If $\boldsymbol{H}$ nonlinear, and $\boldsymbol{x} - \boldsymbol{x}^b$ small
$\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{x}^b) \approx \boldsymbol{H}'(\boldsymbol{x} - \boldsymbol{x}^b)$
where $\boldsymbol{H}'$ is *Jacobian* matrix of $\boldsymbol{H}$ (matrix of partial derivatives) at point $\boldsymbol{x}^b$

Similarly, if $\boldsymbol{M}_k$ nonlinear, and $\boldsymbol{x}^a{}_k - \boldsymbol{x}_k$ small
$\boldsymbol{M}_k(\boldsymbol{x}^a{}_k) - \boldsymbol{M}_k(\boldsymbol{x}) \approx \boldsymbol{M}_k'(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)$
where $\boldsymbol{M}_k'$ is Jacobian matrix of $\boldsymbol{M}_k$ at point $\boldsymbol{x}^a{}_k$

Tangent Linear Approximation

Nonlinearities ?

Model is usually nonlinear, and observation operators (satellite observations) tend more and more to be nonlinear.

- Analysis step

$$\boldsymbol{x}^a{}_k = \boldsymbol{x}^b{}_k + \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}'^{\mathrm{T}} [\boldsymbol{H}_k{}' \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} [\boldsymbol{y}_k - \boldsymbol{H}_k(\boldsymbol{x}^b{}_k)]$$

$$\boldsymbol{P}^a{}_k = \boldsymbol{P}^b{}_k - \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}'^{\mathrm{T}} [\boldsymbol{H}_k{}' \boldsymbol{P}^b{}_k \boldsymbol{H}_k{}'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k{}' \boldsymbol{P}^b{}_k$$

- Forecast step

$$\boldsymbol{x}^b{}_{k+1} = \boldsymbol{M}_k(\boldsymbol{x}^a{}_k)$$

$$\boldsymbol{P}^b{}_{k+1} = \boldsymbol{M}_k{}' \boldsymbol{P}^a{}_k \boldsymbol{M}_k{}'^{\mathrm{T}} + \boldsymbol{Q}_k$$

Extended Kalman Filter (EKF, heuristic !)

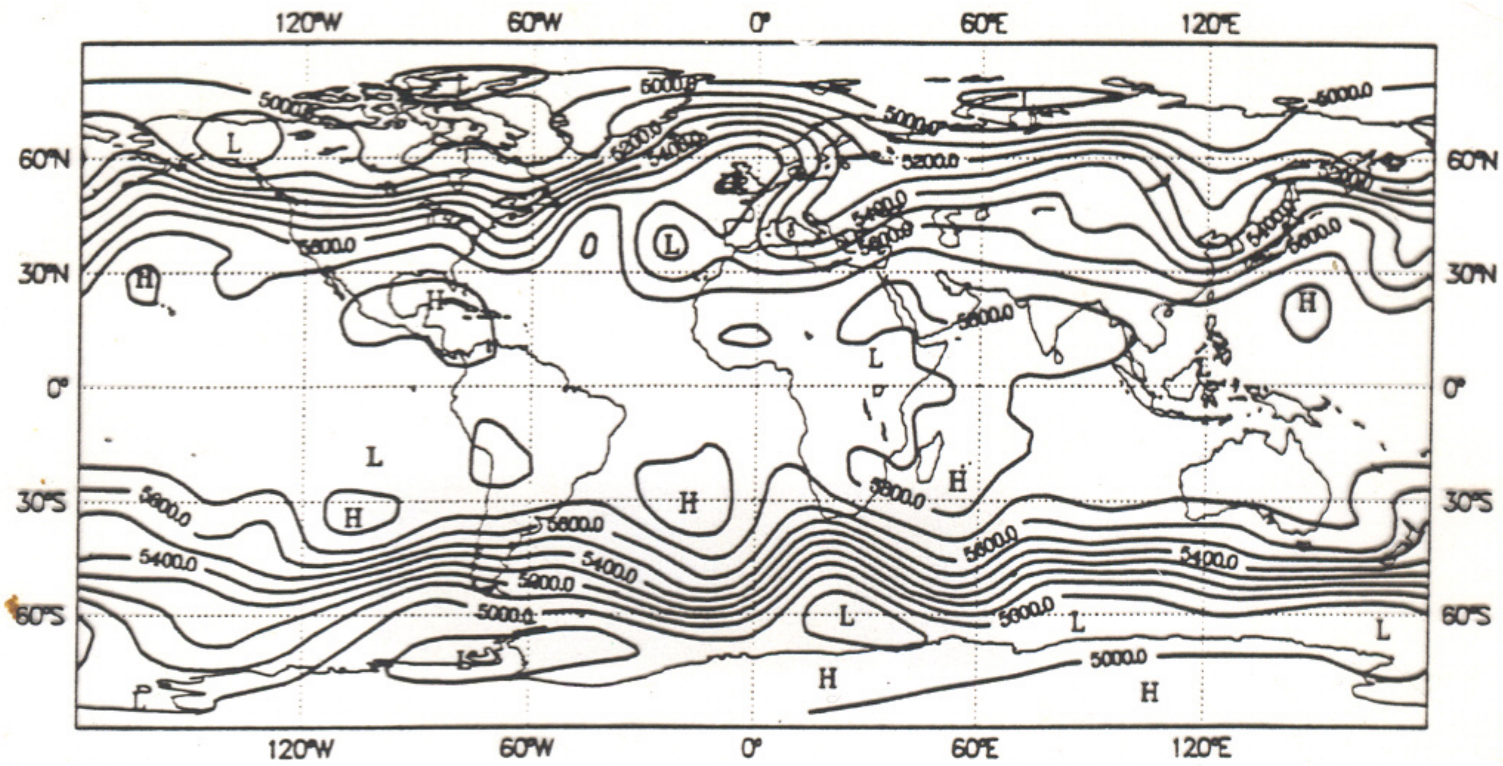
Costliest part of computation

$$\boldsymbol{P}^{b}{}_{k+1} = \boldsymbol{M}_k \boldsymbol{P}^{a}{}_k \boldsymbol{M}_k^{\mathrm{T}} + \boldsymbol{Q}_k$$

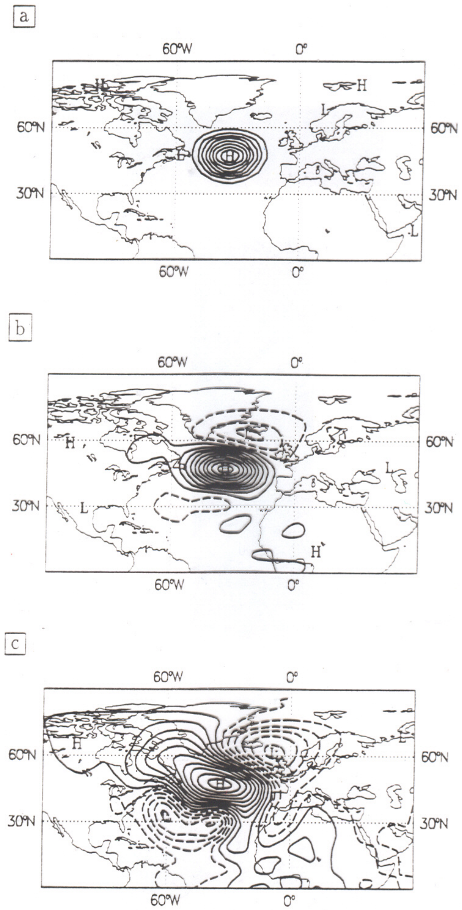
Multiplication of one vector by $\boldsymbol{M}_k$ = one integration of the model between times k and $k+1$

Computation of $\boldsymbol{M}_k \boldsymbol{P}^{a}{}_k \boldsymbol{M}_k^{\mathrm{T}} \approx 2n$ integrations of the model

Need for determining the temporal evolution of the uncertainty on the state of the system is the major difficulty in assimilation of meteorological and oceanographical observations



Analysis of 500-hPa geopotential for 1 December 1989, 00:00 UTC (ECMWF, spectral truncation T21, unit m. After F. Bouttier)



Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.

Two solutions :

- *Low-rank filters*

  Use low-rank covariance matrix, restricted to modes in state space on which it is known, or at least assumed, that a large part of the uncertainty is concentrated (this requires the definition of a norm on state space).

Reduced Rank Square Root Filters (*RRSQRT*, Heemink)

Singular Evolutive Extended Kalman Filter (*SEEK*, Pham)

….

Reduced Rank Square Root Kalman Filter (***RRSQRT***, Verlaan and Heemink, 1997)

A covariance matrix P can be written as

$$P = S\,S^{\mathrm{T}}$$

where the column vectors of S are the (orthogonal) principal components (eigenvectors) of P (the modulus of each vector is the square root of the associated eigenvalue).

The principle of *RRSQRT* is to restrict the background error covariance matrix P^b to $r \ll n$ principal components, thereby approximating P^b by (the time index k is dropped)

$$P^b \approx S^b\,S^{b\mathrm{T}}$$

where S^b has dimensions n x r.

RRSQRT (continuation 1)

Setting $\Psi \equiv (HS^b)^{\mathrm{T}}$, the gain matrix of the Kalman filter and the analysis error covariance matrix respectively become

$$K = S^b\, \Psi\, (\Psi^{\mathrm{T}} \Psi + R)^{-1}$$

and

$$P^a = S^a\, S^{a\mathrm{T}}$$

with

$$S^a = S^b\, [I_r - \Psi\, (\Psi^{\mathrm{T}} \Psi + R)^{-1}\, \Psi^{\mathrm{T}}]^{1/2}$$

RRSQRT (continuation 2)

In the prediction phase, the column vectors of S^a are evolved by the tangent linear model (an evolution of a perturbed state by the full model is also possible). If a model error is to be introduced, that is done by reducing the order r of S^a to r-q, and introducing q new column vectors meant to represent the model error.

Orthogonality of the column vectors is lost in the prediction, and has to be reestablished. And, even if process is started from dominant column vectors, that dominance may of course be lost.

Advantages : in addition to reduced computational cost, numerical errors are smaller when dealing with square root covariance matrices, as done here, than with full matrices (better conditioning).

Singular Evolutive Extended Kalman Filter (***SEEK***, Pham, 1996)

Based on the fact that, because of the linearity of Kalman Filter, the rank of the covariance matrix P^a or P^b cannot increase in either the update or the model evolution. SEEK performs a linear filter starting from a low rank ${P^b}_0$, and so runs the exact Kalman filter in the case of a perfect model. The algorithmic implementation takes advantage of the rank-deficiency of the covariance matrix. The rank of the latter is conserved (or decreased), but the subspace spanned by the directions with non-zero error evolves, in both the update and the dynamic evolution.

In case model error is present, corresponding covariance matrix Q_k is projected onto the directions with non-zero error (this is of course an approximation).

Singular Evolutive Interpolated Kalman Filter (***SEIK***, Pham, 2001)

Non-trivial extension of SEEK to nonlinear model or observation operators. Rank deficiency is now forced.

Second solution :

- *Ensemble filters*

  Uncertainty is represented, not by a covariance matrix, but by an ensemble of point estimates in state space that are meant to sample the conditional probability distribution for the state of the system (dimension $L \approx O(10\text{-}100)$).

  Ensemble is evolved in time through the full model, which eliminates any need for linear hypothesis as to the temporal evolution.

Ensemble Kalman Filter (*EnKF*, Evensen, Anderson, …)

How to update predicted ensemble with new observations ?

Predicted ensemble at time k : $\{\boldsymbol{x}^b_l\}$, $l = 1, \ldots, L$

Observation vector at same time : $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$

- Gaussian approach

Produce sample of probability distribution for real observed quantity $\boldsymbol{H}\boldsymbol{x}$

$\boldsymbol{y}_l = \boldsymbol{y} - \boldsymbol{\varepsilon}_l$

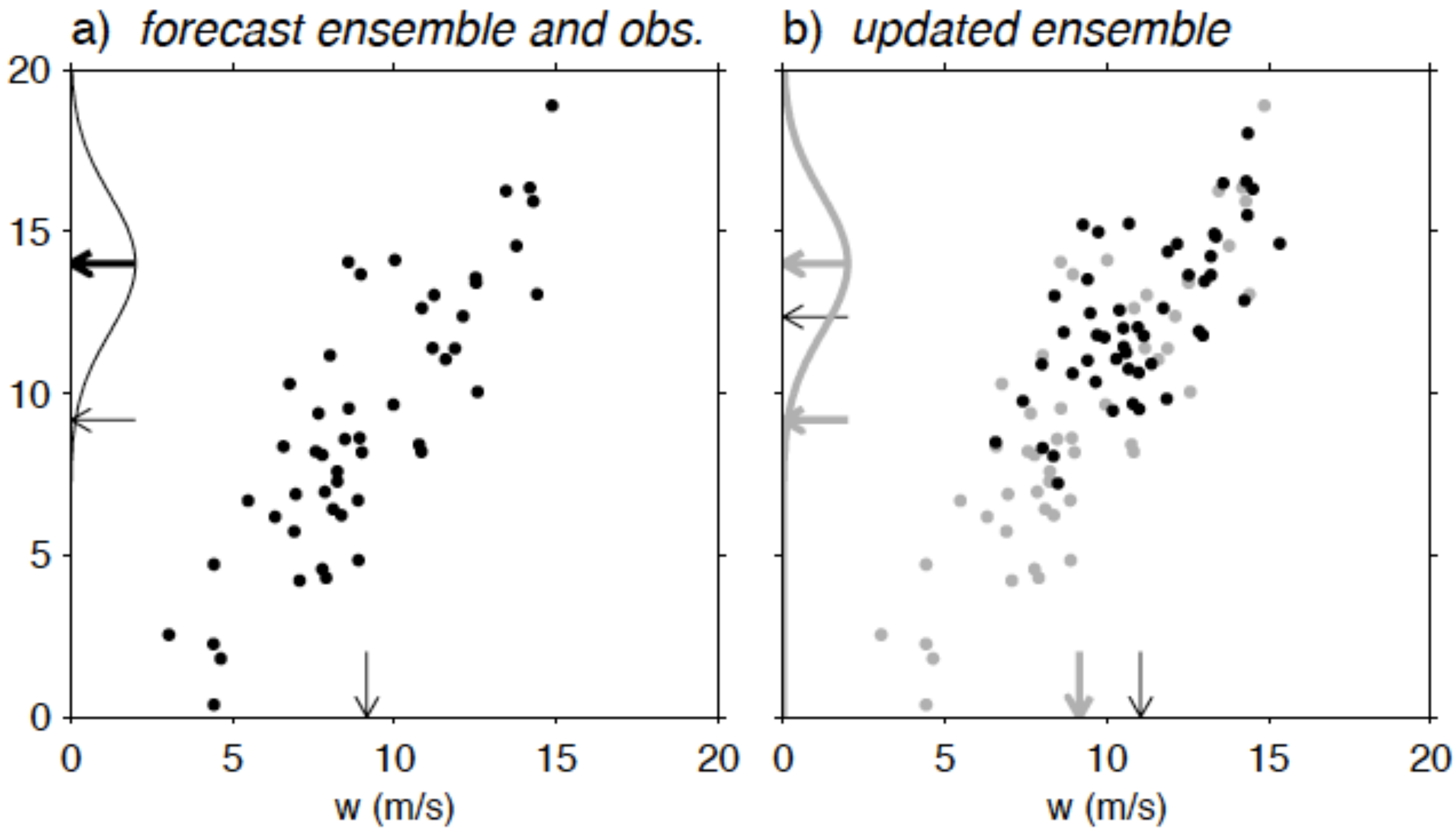
where $\boldsymbol{\varepsilon}_l$ is distributed according to probability distribution for observation error $\boldsymbol{\varepsilon}$.

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

$$\boldsymbol{x}^a_l = \boldsymbol{x}^b_l + \boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1} (\boldsymbol{y}_l - \boldsymbol{H}\boldsymbol{x}^b_l) , \qquad l = 1, \ldots, L \qquad (2)$$

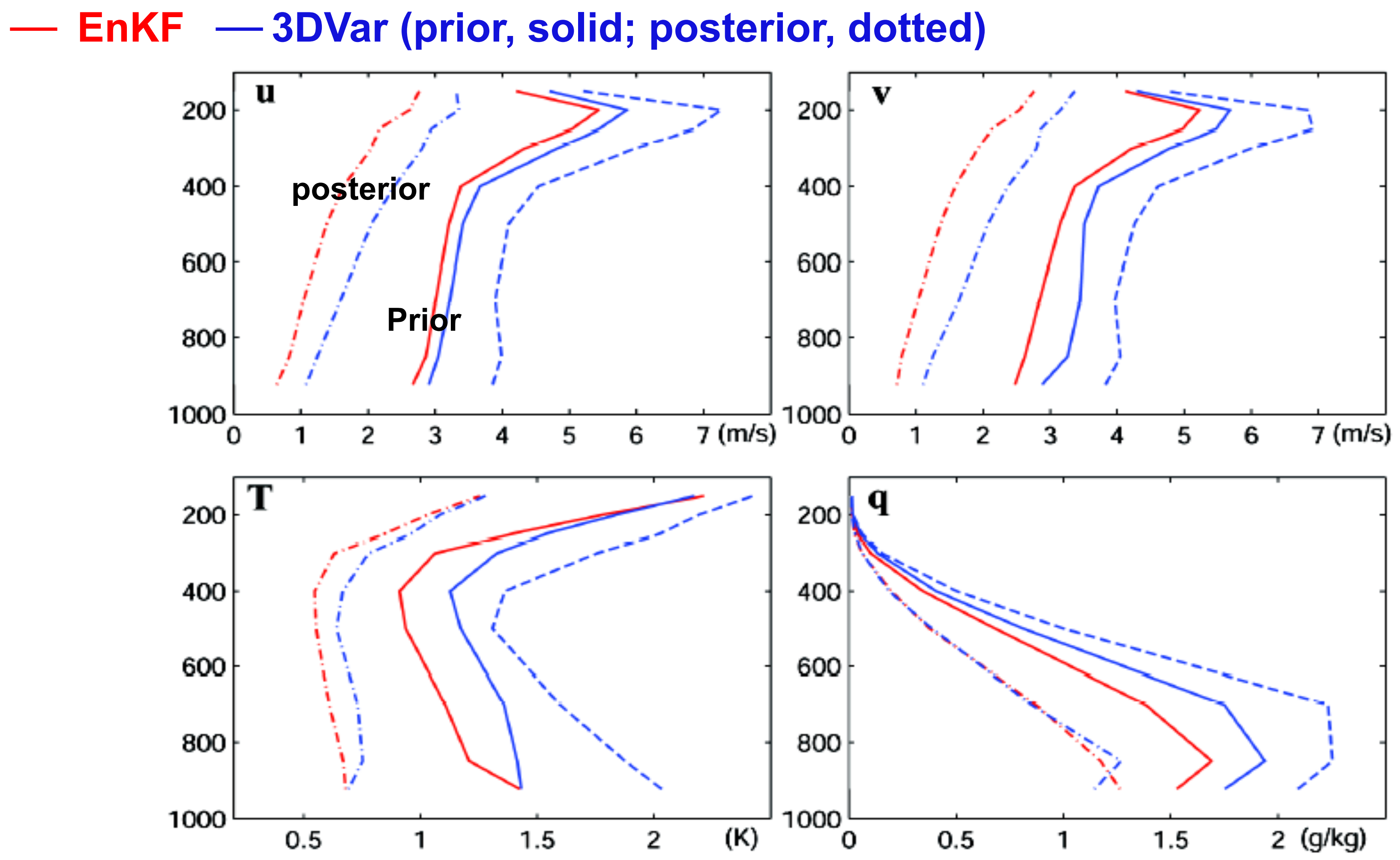
where $\boldsymbol{P}^b$ is the sample covariance matrix of predicted ensemble $\{\boldsymbol{x}^b_l\}$.

Remark. In case of Gaussian errors, if $\boldsymbol{P}^b$ was exact covariance matrix of background error, (2) would achieve Bayesian estimation, in the sense that $\{\boldsymbol{x}^a_l\}$ would be a sample of conditional probability distribution for $\boldsymbol{x}$, given all data up to time k.


a) forecast ensemble and obs.
b) updated ensemble
20
15
10
5
0
0
5
10
15
20
w (m/s)
w (m/s)

C. Snyder

Month-long Performance of EnKF vs. 3Dvar with WRF



Better performance of EnKF than 3DVar also seen in both 12-h forecast and posterior analysis in terms of root-mean square difference averaged over the entire month

(Meng and Zhang 2007c, MWR, in review)

The case of a nonlinear observation operator ?

Predicted ensemble at time k : $\{x^b_l\}$, $l = 1, \ldots, L$

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

Two possibilities

1. Take tangent linear approximation (as in Extended KF) and introduce jacobian $\boldsymbol{H}'$

2. Come back to original formula

$$\boldsymbol{x}^a = E(\boldsymbol{x}) + \boldsymbol{C}_{\boldsymbol{xy}} [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} [\boldsymbol{y} - E(\boldsymbol{y})]$$

That formula does not require any other link between $\boldsymbol{x}$ and $\boldsymbol{y}$ than the one defined by the covariances matrices $\boldsymbol{C}_{\boldsymbol{xy}}$ and $\boldsymbol{C}_{\boldsymbol{yy}}$.

Here, as shown on the occasion of the derivation of the *BLUE*, $E(\boldsymbol{x})$ is the backgound $\boldsymbol{x}^b$, and $\boldsymbol{y} - E(\boldsymbol{y})$ is the innovation $\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^b)$

Solution. Compute $\boldsymbol{C}_{\boldsymbol{xy}}$ and $\boldsymbol{C}_{\boldsymbol{yy}}$ as sample covariances matrices of the ensembles $\{x^b_l\}$ and $\{\boldsymbol{y}_l - \boldsymbol{H}(\boldsymbol{x}^b_l)\}$, where the $\boldsymbol{y}_l$'s are, as before, the perturbed observations $\boldsymbol{y}_l = \boldsymbol{y} - \boldsymbol{\varepsilon}_l$.

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Collapse originates in the fact that gain matrix $\boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1}$ is nonlinear wrt background error matrix P^b, resulting in a systematic sampling effect. Solution : empirical 'covariance inflation'.

- Spurious correlations appear at large geographical distances. Empirical 'localization' (see Gaspari and Cohn, 1999, *Q. J. R. Meteorol. Soc.*)

- In formula

$$\boldsymbol{x}^a{}_l = \boldsymbol{x}^b{}_l + \boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} (\boldsymbol{y}_l - \boldsymbol{H}\boldsymbol{x}^b{}_l) , \qquad l = 1, \dots, L$$

$\boldsymbol{P}^b$, which is covariance matrix of an L-size ensemble, has rank L-1 at most. This means that corrections made on ensemble elements are contained in a subspace with dimension L-1. Obviously very restrictive if $L \ll p$, $L \ll n$.

Houtekamer and Mitchell (1998) use two ensembles, the elements of each of which are updated with covariance matrix of other ensemble.

There exist many variants of Ensemble Kalman Filter

Ensemble Transform Kalman Filter (***ETKF***, Bishop *et al*., *Mon. Wea. Rev*., 2001)

Requires a prior 'control' analysis $x_c{}^a$, emanating from a background $x_c{}^b$. An ensemble is evolved about that control without explicit use of the observations (and without feedback to control)

More precisely, define L x L matrix T such that, given $P^b = ZZ^T$, then $P^a = ZTT^TZ^T$ (not trivial, but possible). Then the background deviations $x^b{}_l - x_c{}^b$ are transformed through $Z \rightarrow ZT$ into an ensemble of analysis deviations $x^a{}_l - x_c{}^a$.

(does not avoid collapse of ensembles)

Local Ensemble Transform Kalman Filter (***LETKF***, Hunt *et al*., *Physica D*, 2007)

Each gridpoint is corrected only through the use of neighbouring observations.

Other variants of Ensemble Kalman Filter

'*Unscented*' *Kalman Filter* (Wan and van der Merve, 2001, Wiley Publishing)

Weighted Kalman Filter (Papadakis *et al*., 2010, *Tellus A*)

Inflation-free Ensemble Kalman Filters (Bocquet and Sakov, 2012, *Nonlin. Processes Geophys*.)

An iterative ensemble Kalman filter in the presence of additive model error (Sakov *et al*., 2017, *Q. J. R. Meteorol. Soc.)*

Bayesian properties of Ensemble Kalman Filter ?

Very little is known.

Le Gland *et al*. (2011). In the linear and gaussian case, the discrete pdf defined by the filter, in the limit of infinite sample size L, tends to the bayesian gaussian pdf.

No result for finite size (note that ensemble elements are not mutually independent)

In the nonlinear case, the discrete pdf tends to a limit which is in general not the bayesian pdf.

Situation still not entirely clear

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. The weights given to z_1 and z_2. will remain equal in sequential assimilation.

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

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

Observations

$y_k = x_k + \varepsilon_k$, $k = 0, 1, 2$ $E(\varepsilon_k{}^2) = r$, 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)

Two questions

- *How to propagate information backwards in time ?* (useful for reassimilation of past data)

- *How to take into account possible dependence in time ?*

Kalman Filter, whether in its standard linear form or in its Ensemble form, does neither.

Kalman smoother

Propagates information both forward and backward in time, as does 4DVar, but uses Kalman-type formulæ

Various possibilities

- Define new state vector $x^T \equiv (x_0^T, \ldots, x_K^T)$
  and use Kalman formula from a background x^b and associated covariance matrix Π^b.
  'Observation operator' must include the model equations
  Can take into account temporal correlations

- Update sequentially vector $(x_0^T, \ldots, x_k^T)^T$ for increasing k
  Cannot take into account temporal correlations

Algorithms exist in ensemble form

E. Cosme (2015)

Ensemble smoother based on *Singular Evolutive Extended Kalman Filter* (*SEEK*)

Of second type above. Retropropagates corrections on fields backwards in time, but without modifying relative weights given to previous data, *i.e.* cannot be optimal in case of temporal dependence between errors.

E. Cosme, HDR, 2015, Lissage d'ensemble SEEK

Données synthétiques

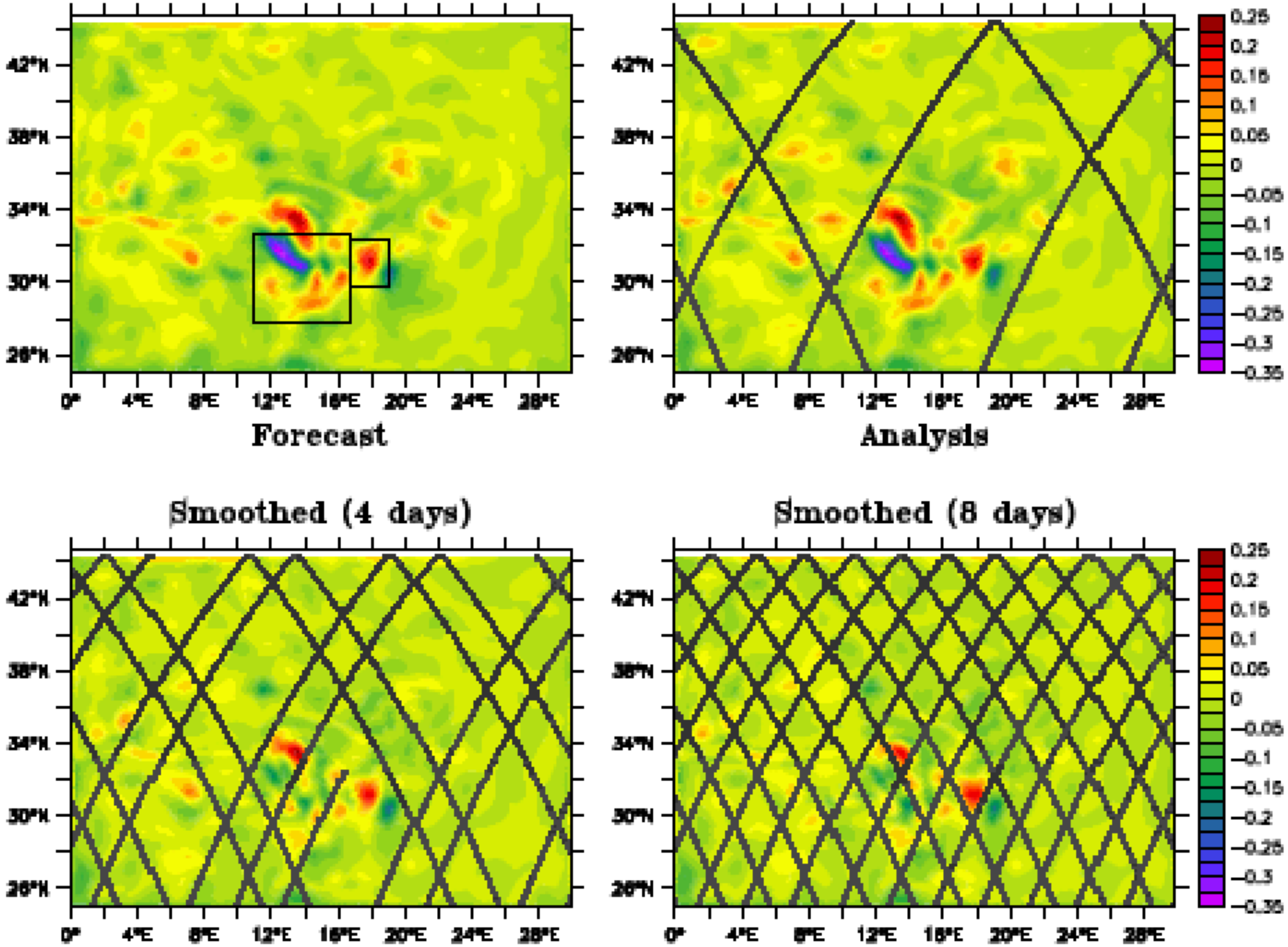


FIGURE 3.6 – Evolution du champ d'erreur en SSH du jour 38, au cours des étapes d'analyse successives. En haut à gauche : prévision du filtre ; en haut à droite : analyse du filtre. Les observations utilisées pour cette analyse sont distribuées le long des traces grises. En bas à gauche : analyse du lisseur après introduction des observations des jours 40 et 42 ; En bas à droite : analyse du lisseur après introduction des observations des jours 40 à 46.

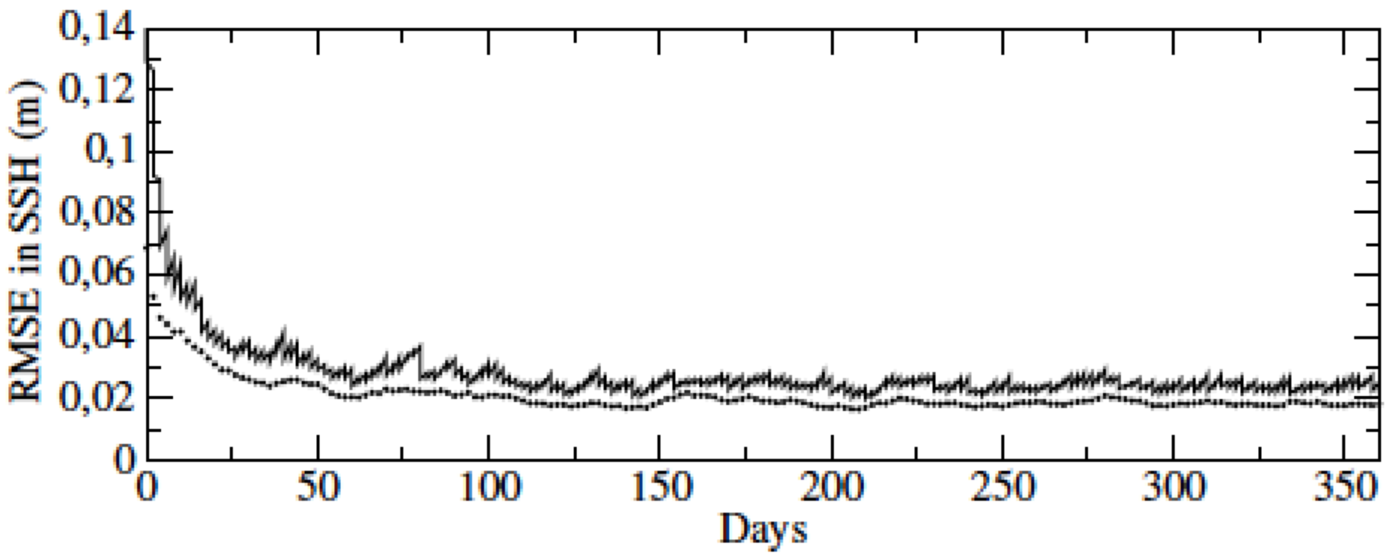


FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse) ; Points : lisseur à retard fixe de 8 jours.

E. Cosme, HDR, 2015, Lissage d'ensemble SEEK

Other variants of Ensemble Kalman Smoothers

*An iterative ensemble Kalman smoother (*Bocquet and Sakov, 2014. *Q. J. R. Meteorol. Soc.*)

An Iterative Ensemble Kalman Smoother in Presence of Additive Model Error (Fillion *et al.*, 2019, *SIAM/ASA J. Uncertainty Quantification)*

Bayesian Estimation (*see course 2*)

Data of the form

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

Known data vector z belongs to *data space* $\mathcal{D}$, $dim\mathcal{D} = m$,
Unknown state vector x belongs to *state space* $\mathcal{X}$, $dim\mathcal{X} = n$
Γ known (mxn)-matrix, ζ unknown 'error'

Probability that $x = \xi$ given ? $\quad x = \xi \Rightarrow \zeta = z - \Gamma\xi$

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

where

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

Then conditional probability distribution is

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

Bayesian Estimation (continuation 1)

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

Then

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

with

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

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

Variational form

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

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

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

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

If data still of the form

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

but 'error' ζ , which still has expectation 0 and covariance S, is not Gaussian, expressions

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

do not achieve Bayesian estimation, but define least-variance linear estimate of x from z (*Best Linear Unbiased Estimator, BLUE*), and associated estimation error covariance matrix.

From course 4

Best Linear Unbiased Estimate

State vector x, belonging to *state space* $\mathcal{S}$ ($\dim \mathcal{S} = n$), to be estimated.

Available data in the form of

- A '*background*' estimate (*e. g.* forecast from the past), belonging to *state space*, with dimension n

  $$x^b = x + \zeta^b$$

- An additional set of data (*e. g.* observations), belonging to *observation space*, with dimension p

  $$y = Hx + \varepsilon$$

  H is known linear *observation operator.*

Assume probability distribution is known for the couple (ζ^b, ε).

Assume $E(\zeta^b) = 0$, $E(\varepsilon) = 0$, $E(\zeta^b \varepsilon^T) = 0$ (not restrictive)

Set $E(\zeta^b \zeta^{bT}) \equiv P^b$ (also often denoted B), $E(\varepsilon\varepsilon^T) \equiv R$

From course 4

Best Linear Unbiased Estimate (continuation 1)

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b \qquad (1)$$

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{\varepsilon} \qquad (2)$$

A probability distribution being known for the couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$, eqs (1-2) define probability distribution for the couple $(\boldsymbol{x}, \boldsymbol{y})$, with

$$E(\boldsymbol{x}) = \boldsymbol{x}^b \,, \;\; \boldsymbol{x}' = \boldsymbol{x} - E(\boldsymbol{x}) = - \boldsymbol{\zeta}^b$$

$$E(\boldsymbol{y}) = H\boldsymbol{x}^b \,, \;\; \boldsymbol{y}' = \boldsymbol{y} - E(\boldsymbol{y}) = \boldsymbol{y} - H\boldsymbol{x}^b = \boldsymbol{\varepsilon} - H\boldsymbol{\zeta}^b$$

$\boldsymbol{d} \equiv \boldsymbol{y} - H\boldsymbol{x}^b$ is called the *innovation vector*.

From course 4

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

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

x^a is the *Best Linear Unbiased Estimate* (*BLUE*) of x from x^b and y.

Equivalent set of formulæ

$$\boldsymbol{x}^a = \boldsymbol{x}^b + P^a H^{\mathrm{T}} R^{-1} (\boldsymbol{y} - H\boldsymbol{x}^b)$$
$$[P^a]^{-1} = [P^b]^{-1} + H^{\mathrm{T}} R^{-1} H$$

Vector $d \equiv \boldsymbol{y} - H\boldsymbol{x}^b$ is *innovation vector*

Matrix $K \equiv P^b H^{\mathrm{T}} [HP^bH^{\mathrm{T}} + R]^{-1} = P^a H^{\mathrm{T}} R^{-1}$ is *gain matrix*.

If probability distributions are *globally* gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

From course 4

Best Linear Unbiased Estimate (continuation 6)

Variational form of the *BLUE*

BLUE x^a minimizes following scalar *objective function*, defined on state space

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

- $\mathcal{J}(\xi) \equiv (1/2)\,(x^b - \xi)^{\mathrm{T}}\,[P^b]^{-1}\,(x^b - \xi) + (1/2)\,(y - H\xi)^{\mathrm{T}}\,R^{-1}\,(y - H\xi)$

$\equiv \quad \mathcal{J}_b \quad + \quad \mathcal{J}_o$

'*3D-Var*'

Can easily, and heuristically, be extended to the case of a nonlinear observation operator H.

Used operationally in USA, Australia, China, …

Case of data that are distributed over time

Suppose for instance available data consist of

- 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 $k = 0, \ldots, K$

  $y_k = H_k x_k + \varepsilon_k$ $\qquad E(\varepsilon_k \varepsilon_j^T) = R_k \, \delta_{kj}$

- Model (supposed for the time being to be exact)

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

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

Then objective function

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

$$\mathcal{J}(\xi_0) \equiv (1/2)\,(x_0^b - \xi_0)^T [P_0^b]^{-1} (x_0^b - \xi_0) + (1/2)\,\Sigma_k [y_k - H_k \xi_k]^T R_k^{-1} [y_k - H_k \xi_k]$$

$$\equiv \qquad \mathcal{J}_b \qquad + \qquad \mathcal{J}_o$$

subject to $\xi_{k+1} = M_k \xi_k$, $\qquad k = 0, \ldots, K\text{-}1$

$$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^T\,R_k^{-1}\,[y_k - H_k\xi_k]$$

Background is not necessary, if observations are in sufficient number to overdetermine the problem. Nor is strict linearity.

Four-Dimensional Variational Assimilation

‘*4D-Var*’

How to minimize objective function with respect to initial state $u = \xi_0$ (u is called the *control variable* of the problem) ?

Use iterative minimization algorithm, each step of which requires the explicit knowledge of the local gradient $\nabla_u \mathcal{J} \equiv (\partial \mathcal{J}/\partial u_i)$ of $\mathcal{J}$ with respect to u.

How to numerically compute the gradient $\nabla_u \mathcal{J}$?

Direct perturbation, in order to obtain partial derivatives $\partial\mathcal{J}/\partial u_i$ by finite differences ? That would require as many explicit computations of the objective function $\mathcal{J}$ as there are components in u. Practically impossible.

Gradient computed by *adjoint method*.

Adjoint Method

Input vector $\boldsymbol{u} = (u_i)$, $\dim \boldsymbol{u} = n$

Numerical process, implemented on computer (*e. g.* integration of numerical model)

$$\boldsymbol{u} \rightarrow \boldsymbol{v} = \boldsymbol{G}(\boldsymbol{u})$$

$\boldsymbol{v} = (v_j)$ is *output vector*, $\dim \boldsymbol{v} = m$

Perturbation $\delta\boldsymbol{u} = (\delta u_i)$ of input. Resulting first-order perturbation on $\boldsymbol{v}$

$$\delta v_j = \Sigma_i \, (\partial v_j/\partial u_i) \; \delta u_i$$

or, in matrix form

$$\delta\boldsymbol{v} \;=\; \boldsymbol{G}' \delta\boldsymbol{u}$$

where $\boldsymbol{G}' \equiv (\partial v_j/\partial u_i)$ is local matrix of partial derivatives, or *jacobian matrix*, of $\boldsymbol{G}$.

Adjoint Method (continued 1)

$$\delta \boldsymbol{v} = \boldsymbol{G}' \delta \boldsymbol{u} \qquad \text{(D)}$$

- Scalar function of output

$$\mathcal{J}(\boldsymbol{v}) = \mathcal{J}[\boldsymbol{G}(\boldsymbol{u})]$$

Gradient $\nabla_{\boldsymbol{u}} \mathcal{J}$ of $\mathcal{J}$ with respect to input $\boldsymbol{u}$?

'Chain rule'

$$\partial \mathcal{J} / \partial u_i = \Sigma_j \, \partial \mathcal{J} / \partial v_j \, (\partial v_j / \partial u_i)$$

or

$$\nabla_{\boldsymbol{u}} \mathcal{J} = \boldsymbol{G}'^{\mathrm{T}} \, \nabla_{\boldsymbol{v}} \mathcal{J} \qquad \text{(A)}$$

Adjoint Method (continued 2)

$\boldsymbol{G}$ is the composition of a number of successive steps

$$\boldsymbol{G} = \boldsymbol{G}_N \circ \ldots \circ \boldsymbol{G}_2 \circ \boldsymbol{G}_1$$

'Chain rule'

$$\boldsymbol{G}' = \boldsymbol{G}_N' \ldots \boldsymbol{G}_2' \boldsymbol{G}_1'$$

Transpose

$$\boldsymbol{G}'^{\mathrm{T}} = \boldsymbol{G}_1'^{\mathrm{T}} \boldsymbol{G}_2'^{\mathrm{T}} \ldots \boldsymbol{G}_N'^{\mathrm{T}}$$

Transpose, or *adjoint*, computations are performed in reversed order of direct computations.

If $\boldsymbol{G}$ is nonlinear, local jacobian $\boldsymbol{G}'$ depends on local value of input $\boldsymbol{u}$. Any quantity which is an argument of a nonlinear operation in the direct computation will be used again in the adjoint computation. It must be kept in memory from the direct computation (or else be recomputed again in the course of the adjoint computation).

If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

Adjoint Method (continued 3)

$\mathcal{J}(\xi_0) = (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k\xi_k]^T\,R_k^{-1}\,[y_k - H_k\xi_k]$

subject to $\xi_{k+1} = M_k\xi_k$, $k = 0, \ldots, K\text{-}1$

Control variable $\xi_0 = \boldsymbol{u}$

Adjoint equation

$\lambda_K = H_K^T R_K^{-1}\,[H_K\,\xi_K - y_K]$

....

$\lambda_k = M_k^T\lambda_{k+1} + H_k^T R_k^{-1}\,[H_k\,\xi_k - y_k]$ $k = K\text{-}1, \ldots, 1$

....

$\lambda_0 = M_0^T\lambda_1 + H_0^T R_0^{-1}\,[H_0\,\xi_0 - y_0] + [P_0^b]^{-1}\,(\xi_0 - x_0^b)$

$$\nabla_u \mathcal{J} = \lambda_0$$

Result of direct integration (ξ_k), which appears in quadratic terms in expression of objective function, must be kept in memory from direct integration.

Adjoint Method (continued 3)

Nonlinearities ?

$$\mathcal{J}(\xi_0) \; = \; (1/2)\,(x_0^b - \xi_0)^{\mathrm{T}}\,[P_0^b]^{-1}\,(x_0^b - \xi_0) + (1/2)\,\Sigma_k[y_k - H_k(\xi_k)]^{\mathrm{T}}\,R_k^{-1}\,[y_k - H_k(\xi_k)]$$

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

Control variable $\qquad \xi_0 = \boldsymbol{u}$

Adjoint equation

$$\lambda_K = \qquad H_K{}'^{\mathrm{T}}\,R_K^{-1}\,[H_K(\xi_K) - y_K]$$

....

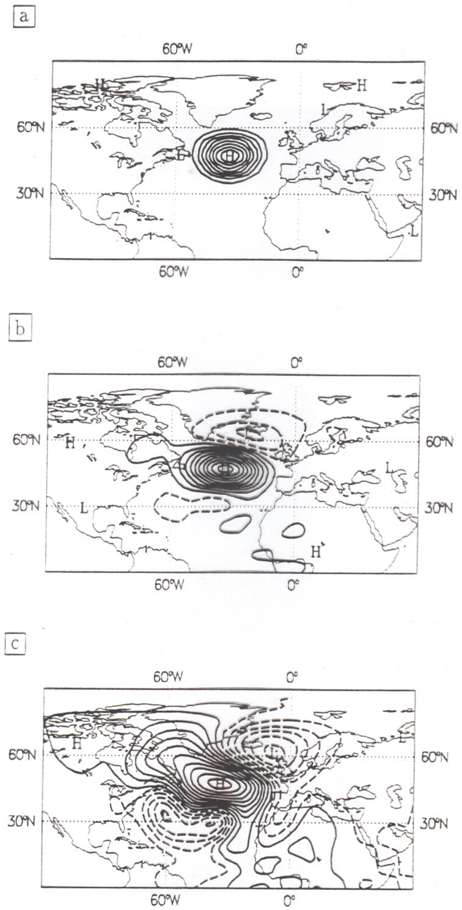
$$\lambda_k = \; M_k{}'^{\mathrm{T}}\lambda_{k+1} + H_k{}'^{\mathrm{T}}\,R_k^{-1}\,[H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$$

....

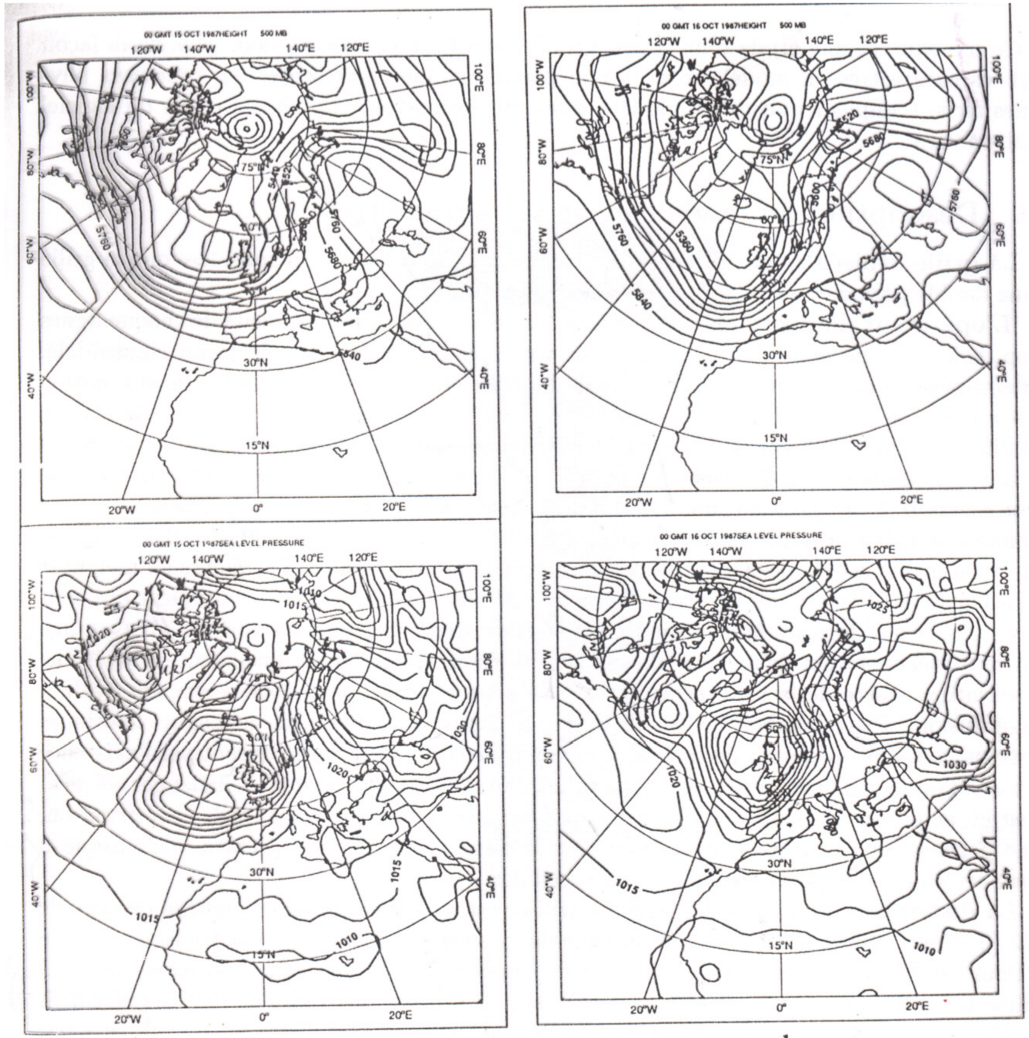
$$\lambda_0 = \; M_0{}'^{\mathrm{T}}\lambda_1 \; + H_0{}'^{\mathrm{T}}\,R_0^{-1}\,[H_0(\xi_0) - y_0] \; + \; [P_0^b]^{-1}\,(\xi_0 - x_0^b)$$

$$\nabla_u \mathcal{J} \; = \lambda_0$$

Not approximate (it gives the exact gradient $\nabla_u \mathcal{J}$), and really used as described here.



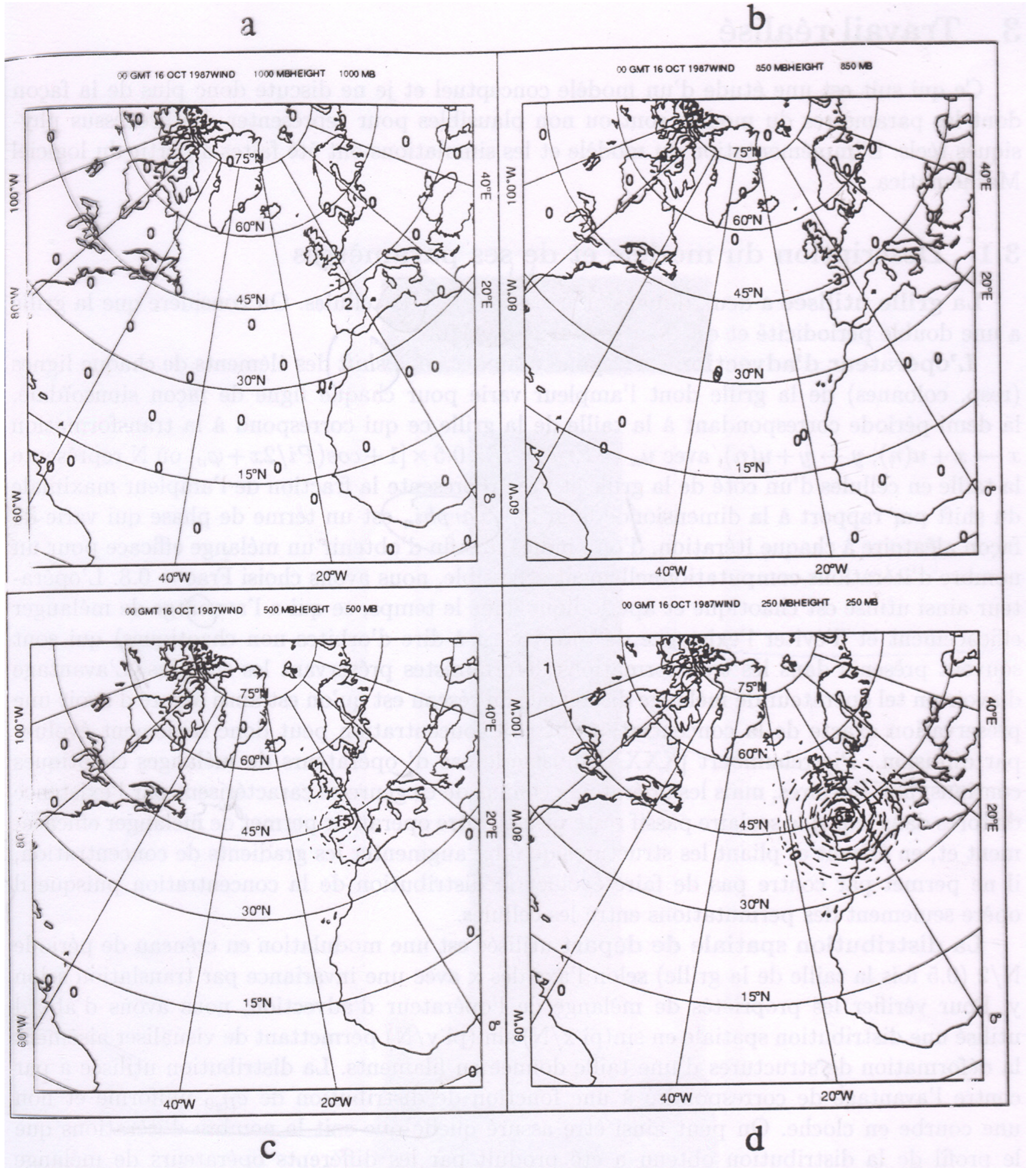
Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.



b d

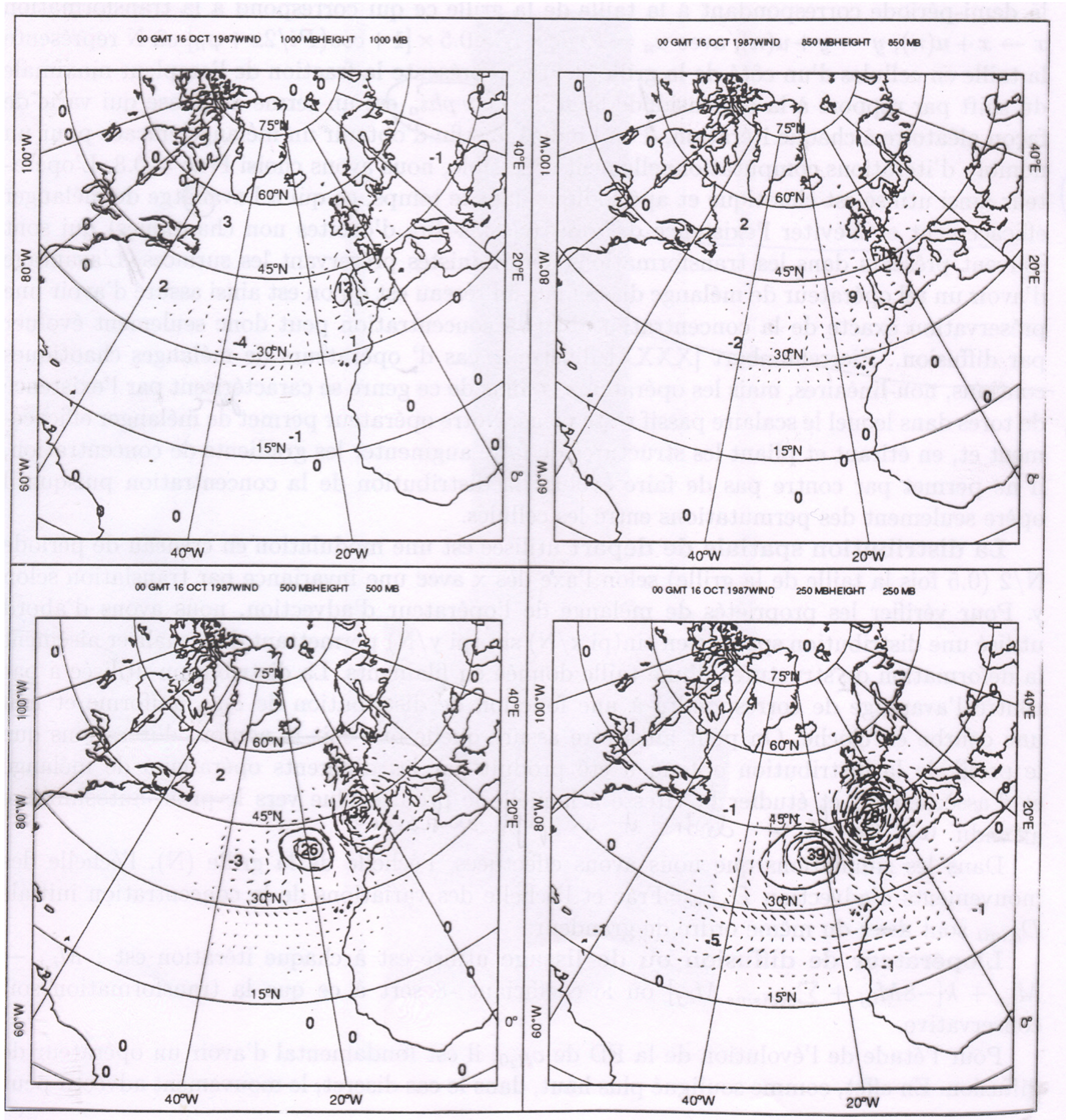
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October are from the initial estimate of the initial conditions for the 4DVAR minimization. The fields for 16 October are from the 24-h T63 adiabatic model forecast from the initial conditions. Contour intervals are 80 m and 5 hPa.

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



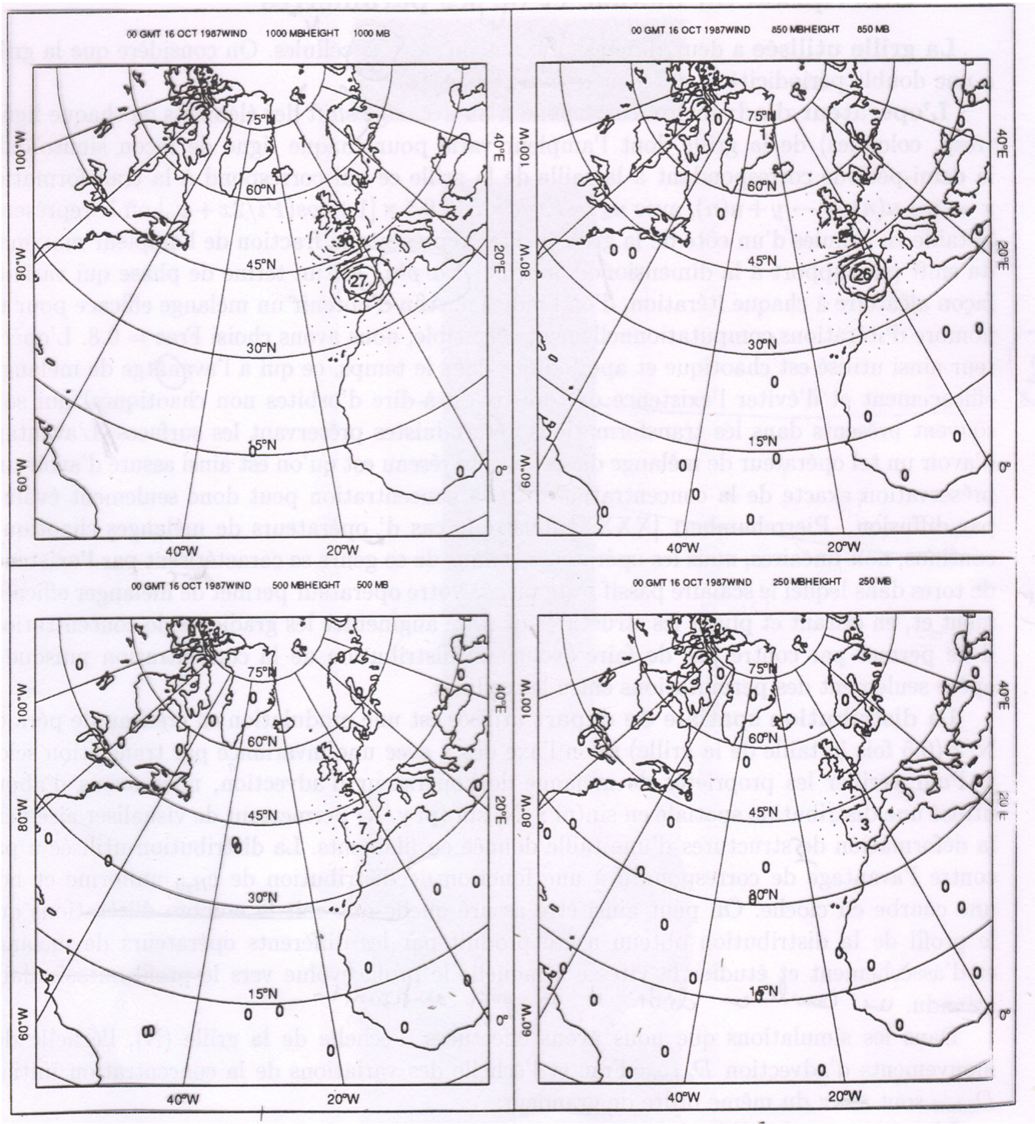
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



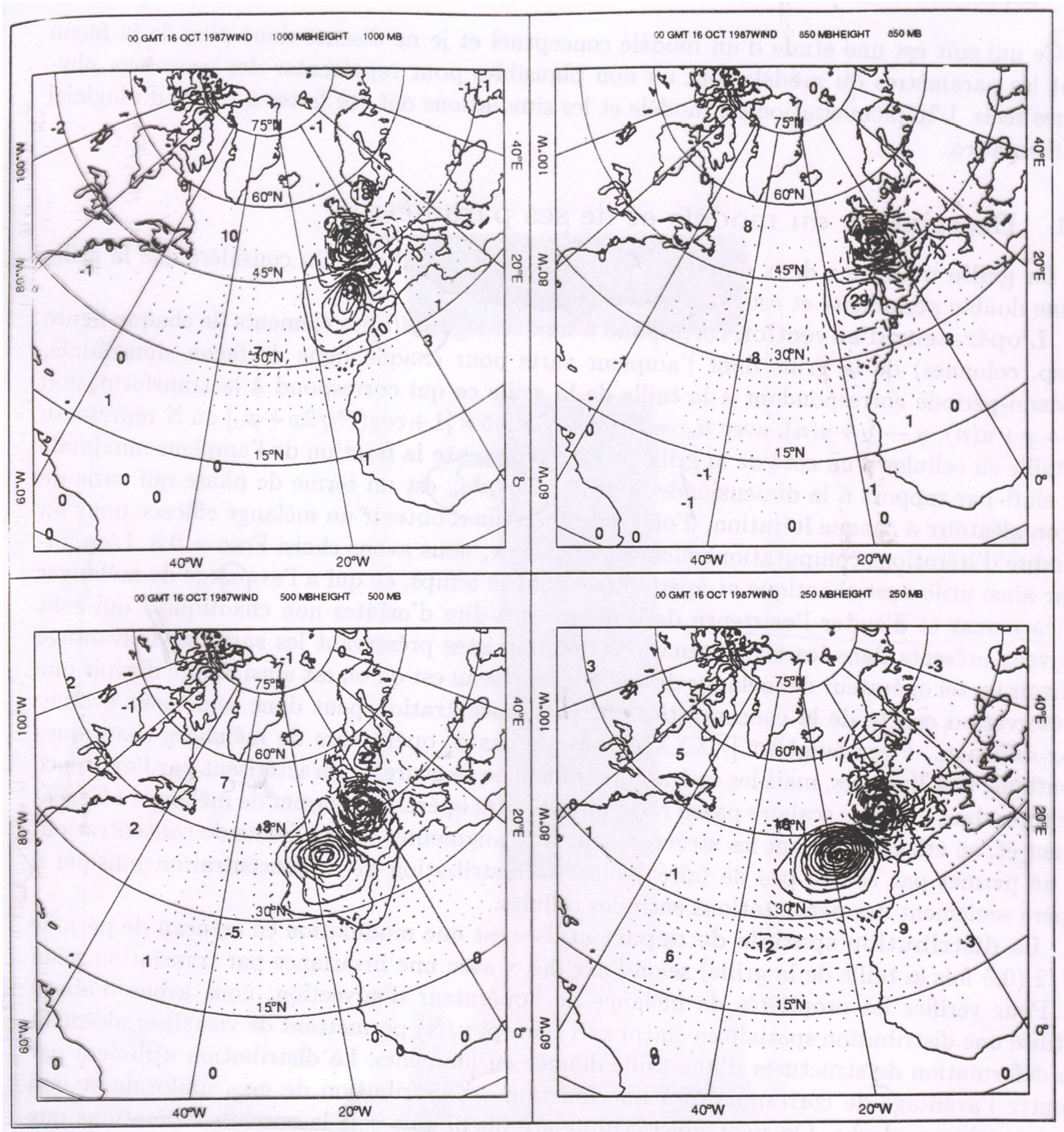
Same as before, but at the end of a 24-hr 4D-Var

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



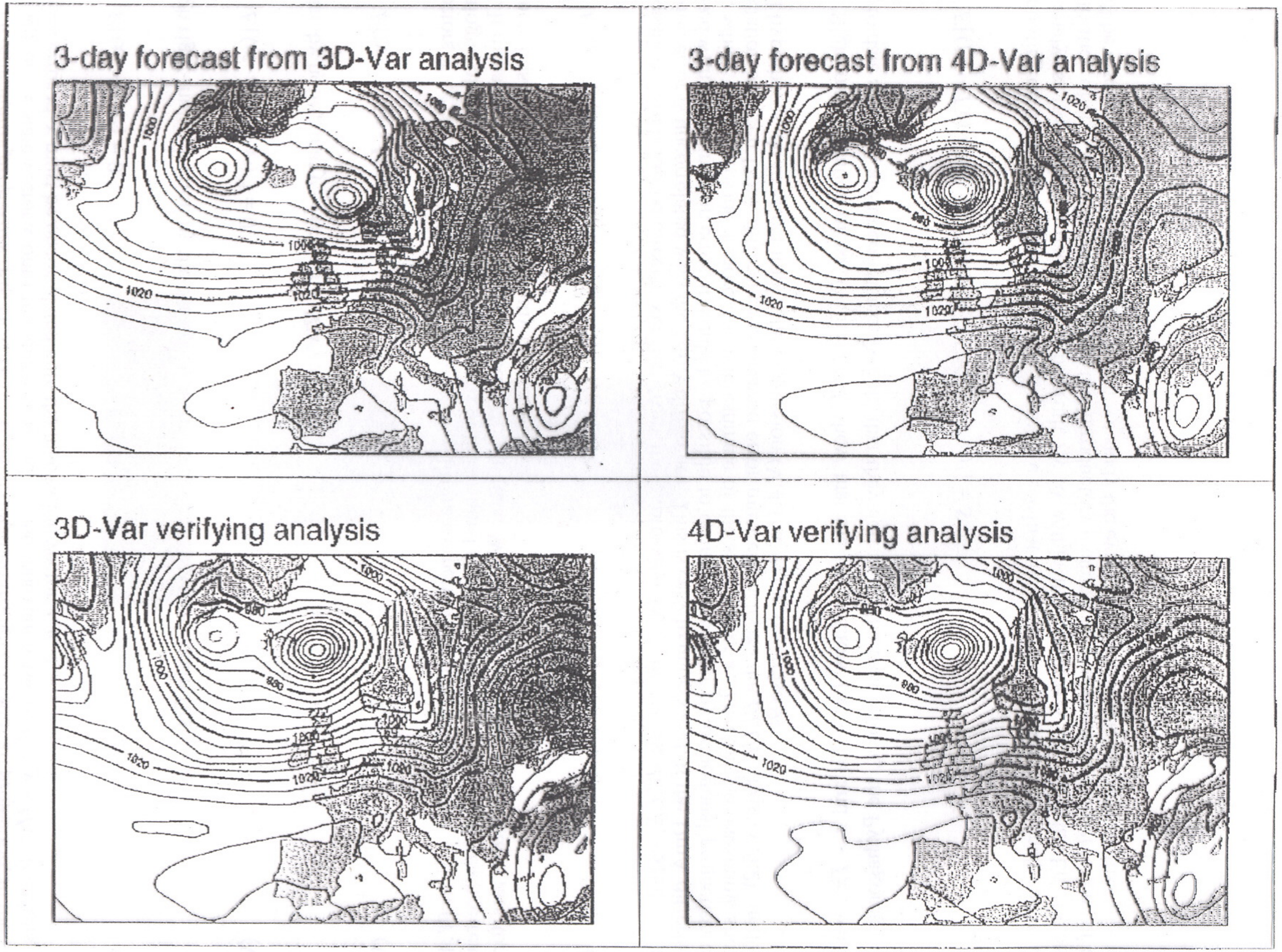
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

Thépaut *et al*., 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



Same as before, but at the end of a 24-hr 4D-Var

Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



ECMWF, Results on one FASTEX case (1997)

Strong Constraint 4D-Var is now used operationally at several meteorological centres (Météo-France, UK Meteorological Office, Canadian Meteorological Centre, Japan Meteorological Agency, ...) and, for a number of years, at ECMWF. The latter now has a 'weak constraint' component in its operational system.

Cours à venir

~~Vendredi 26 mars~~
~~Vendredi 2 avril~~
~~Vendredi 9 avril~~
~~Vendredi 16 avril~~
~~Vendredi 7 mai~~
Vendredi 14 mai
Vendredi 21 mai
Vendredi 28 mai