

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

Année Universitaire 2019-2020

Modélisation Numérique  
de l'Écoulement Atmosphérique  
et Assimilation de Données

Olivier Talagrand

Cours 5

21 Avril 2020

### ***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  $(\zeta^b, \varepsilon)$ .

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  $B$ ),  $E(\varepsilon \varepsilon^T) \equiv R$

## *Best Linear Unbiased Estimate* (continuation 3)

$$\begin{aligned}\mathbf{x}^a &= \mathbf{x}^b + \mathbf{P}^b \mathbf{H}^\top [\mathbf{H} \mathbf{P}^b \mathbf{H}^\top + \mathbf{R}]^{-1} (\mathbf{y} - \mathbf{H} \mathbf{x}^b) \\ \mathbf{P}^a &= \mathbf{P}^b - \mathbf{P}^b \mathbf{H}^\top [\mathbf{H} \mathbf{P}^b \mathbf{H}^\top + \mathbf{R}]^{-1} \mathbf{H} \mathbf{P}^b\end{aligned}$$

$\mathbf{x}^a$  is the *Best Linear Unbiased Estimate (BLUE)* of  $\mathbf{x}$  from  $\mathbf{x}^b$  and  $\mathbf{y}$ .

Equivalent set of formulæ

$$\begin{aligned}\mathbf{x}^a &= \mathbf{x}^b + \mathbf{P}^a \mathbf{H}^\top \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H} \mathbf{x}^b) \\ [\mathbf{P}^a]^{-1} &= [\mathbf{P}^b]^{-1} + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H}\end{aligned}$$

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

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

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

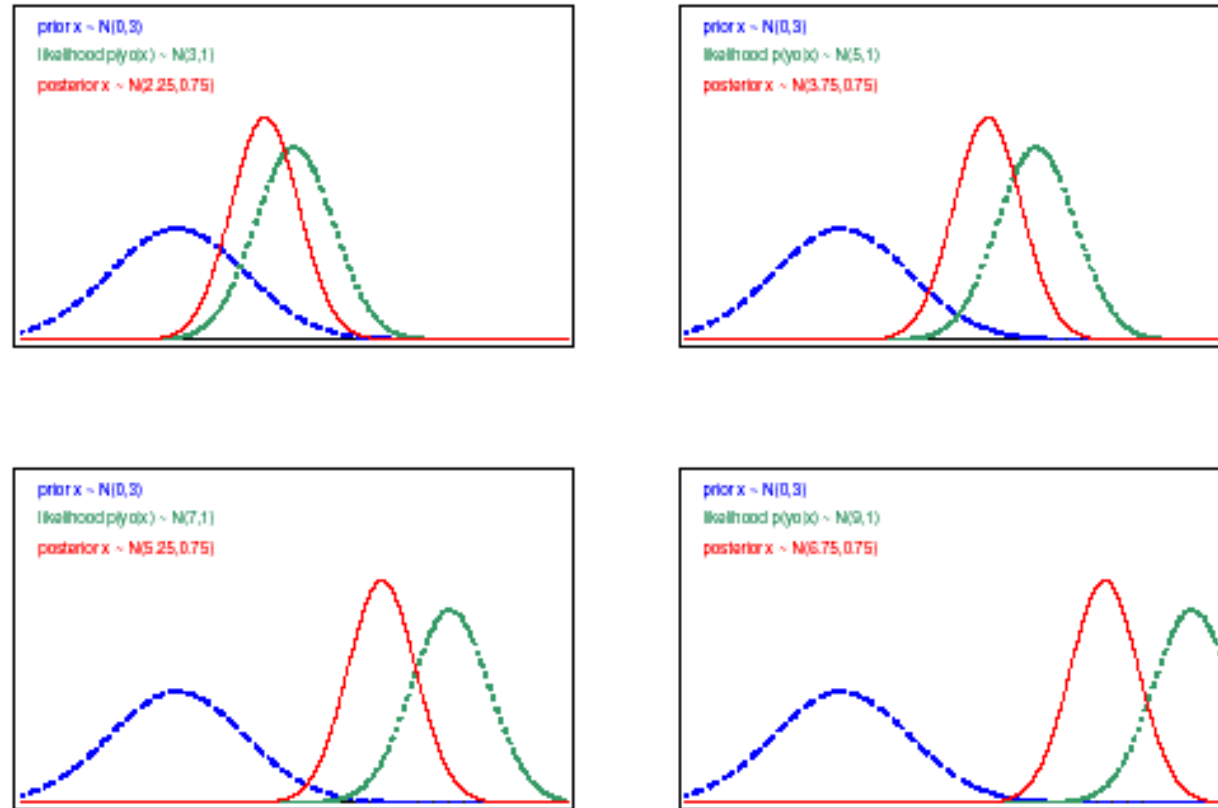


Fig. 1.1: Prior pdf  $p(x)$  (dashed line), posterior pdf  $p(x|y^o)$  (solid line), and Gaussian likelihood of observation  $p(y^o|x)$  (dotted line), plotted against  $x$  for various values of  $y^o$ . (Adapted from Lorenc and Hammon 1988.)

$$z_1 = x + \xi_1$$

$$z_2 = x + \xi_2$$

Same as before, but  $\xi_1$  and  $\xi_2$  are now distributed according to exponential law with parameter  $a$ , *i. e.*

$$p(\xi) \propto \exp[-|\xi|/a] \quad ; \quad \text{Var}(\xi) = 2a^2$$

Conditional probability density function is now uniform over interval  $[z_1, z_2]$ , exponential with parameter  $a/2$  outside that interval

$$E(x | z_1, z_2) = (z_1 + z_2)/2$$

$$\text{Var}(x | z_1, z_2) = a^2 (2\delta^3/3 + \delta^2 + \delta + 1/2) / (1 + 2\delta), \text{ with } \delta = |z_1 - z_2| / (2a)$$

Increases from  $a^2/2$  to  $\infty$  as  $\delta$  increases from 0 to  $\infty$ . Can be larger than variance  $2a^2$  of original errors (probability 0.08)

- Ensemble Kalman Filter. Continuation
- Kalman Smoother. Brief theory. An example.
- Variational assimilation. Principle. The adjoint approach. Results
- Weak constraint variational assimilation.  
Principle

## Ensemble Kalman Filter (EnKF) From course 4

How to update predicted ensemble with new observations ?

Predicted ensemble at time  $k$  :  $\{x_l^b\}$ ,  $l = 1, \dots, L$

Observation vector at same time :  $y = Hx + \varepsilon$

- Gaussian approach

Produce sample of probability distribution for real observed quantity  $Hx$

$$y_l = y - \varepsilon_l$$

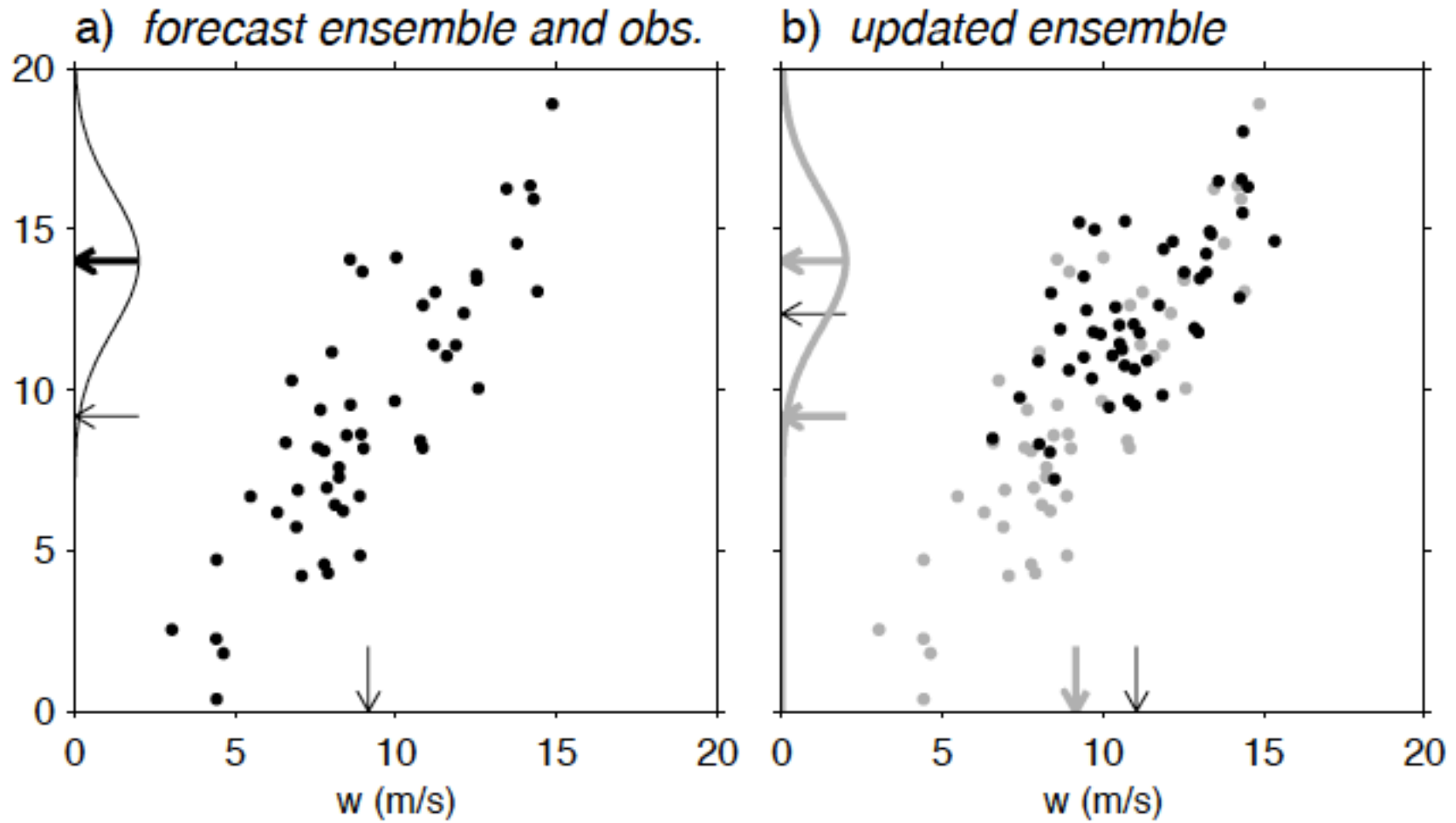
where  $\varepsilon_l$  is distributed according to probability distribution for observation error  $\varepsilon$ .

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

$$x_l^a = x_l^b + P^b H^T [HP^b H^T + R]^{-1} (y_l - Hx_l^b), \quad l = 1, \dots, L \quad (2)$$

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

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



C. Snyder



But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Collapse originates in the fact that gain matrix  $P^b H^T [HP^b H^T + R]^{-1}$  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

$$x_l^a = x_l^b + P^b H^T [HP^b H^T + R]^{-1} (y_l - Hx_l^b) , \quad l = 1, \dots, L$$

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

(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  $z_2$ .

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

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

Observations

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

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

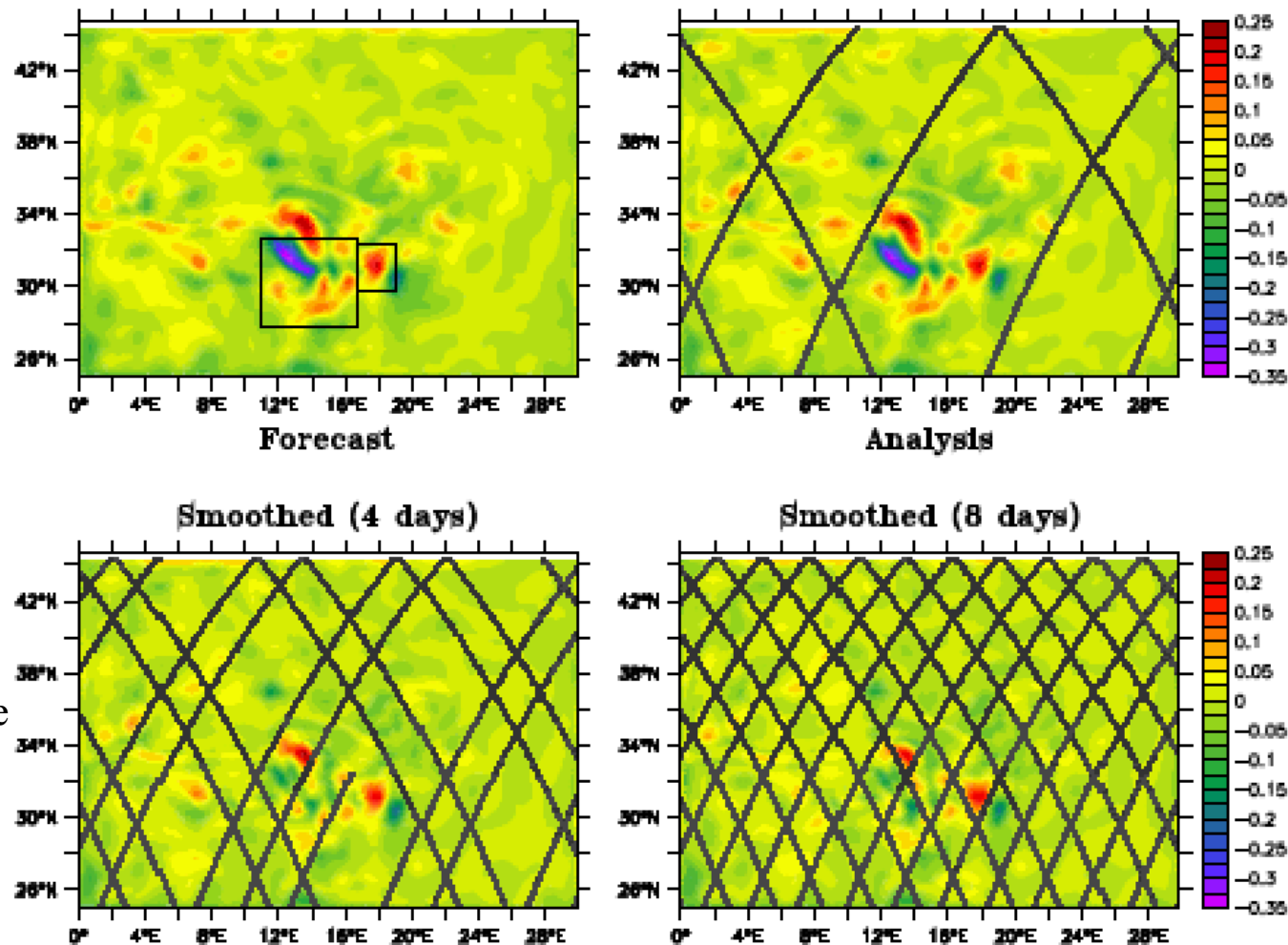
- Define new state vector  $x^T \equiv (x_0^T, \dots, x_K^T)$   
and use Kalman formula from a background  $x_b$  and associated covariance matrix  $\Pi_b$ .  
'Observation vector'  $y$  must include the model equations  
Can take into account temporal correlations
- Update sequentially vector  $(x_0^T, \dots, x_k^T)^T$  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.

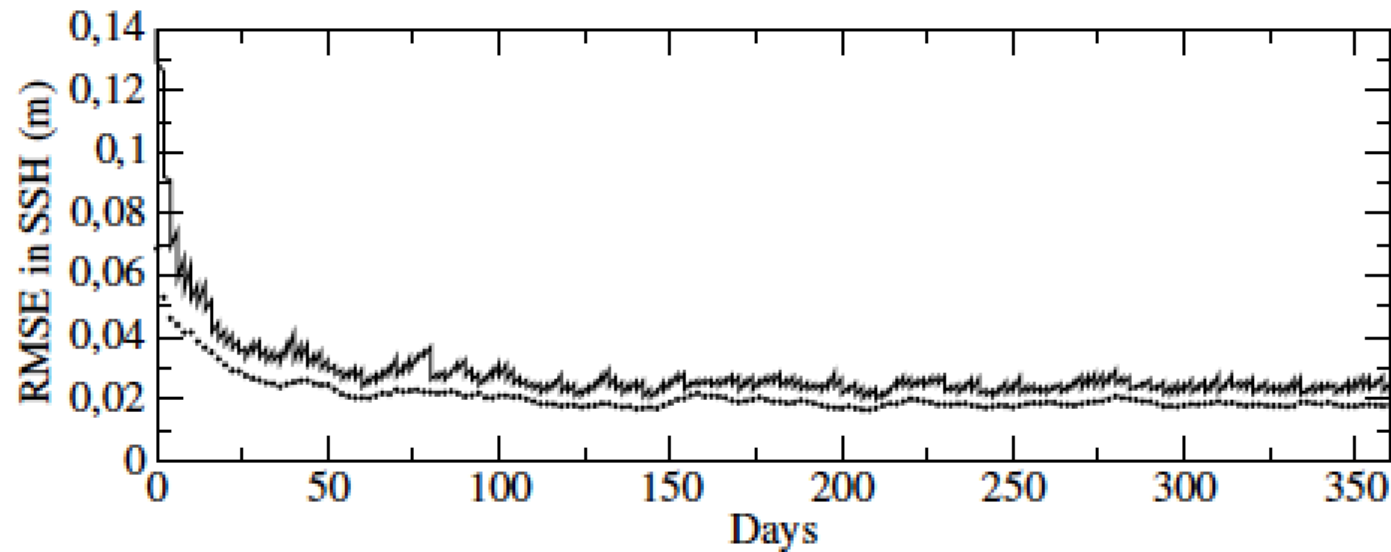


FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse); Points : lisseur à retard fixe de 8 jours.

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

### Data of the form

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

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

Unknown state vector  $x$  belongs to *state space*  $\mathcal{X}$ ,  $\dim \mathcal{X} = n$

$\Gamma$  known ( $m \times n$ )-matrix,  $\zeta$  unknown 'error'

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

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

where

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

Then conditional probability distribution is

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

## Bayesian Estimation (continuation 1)

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

Then

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

with

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

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



## Variational form

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

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

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

but ‘error’  $\xi$ , 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  $(\zeta^b, \varepsilon)$ .

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)

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

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

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

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

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

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

From course 4

## Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

$$\begin{aligned}\mathbf{x}^a &= \mathbf{x}^b + P^b H^T [HP^b H^T + R]^{-1} (\mathbf{y} - H\mathbf{x}^b) \\ P^a &= P^b - P^b H^T [HP^b H^T + R]^{-1} HP^b\end{aligned}$$

$\mathbf{x}^a$  is the *Best Linear Unbiased Estimate (BLUE)* of  $\mathbf{x}$  from  $\mathbf{x}^b$  and  $\mathbf{y}$ .

Equivalent set of formulæ

$$\begin{aligned}\mathbf{x}^a &= \mathbf{x}^b + P^a H^T R^{-1} (\mathbf{y} - H\mathbf{x}^b) \\ [P^a]^{-1} &= [P^b]^{-1} + H^T R^{-1} H\end{aligned}$$

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

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

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

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

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

- $$\begin{aligned} \mathcal{J}(\xi) &\equiv (1/2) (x^b - \xi)^T [P^b]^{-1} (x^b - \xi) + (1/2) (y - H\xi)^T R^{-1} (y - H\xi) \\ &\equiv \mathcal{J}_b \quad + \quad \mathcal{J}_o \end{aligned}$$

‘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

$$x_0^b = x_0 + \xi_0^b \quad E(\xi_0^b \xi_0^{bT}) = P_0^b$$

- Observations at times  $k = 0, \dots, K$

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

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

$$x_{k+1} = M_k x_k \quad k = 0, \dots, K-1$$

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

Then objective function

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

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

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

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

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

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*  $\mathbf{u} = (u_i)$ ,  $\dim \mathbf{u} = n$

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

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

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

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

$$\delta v_j = \sum_i (\partial v_j / \partial u_i) \delta u_i$$

or, in matrix form

$$\delta \mathbf{v} = \mathbf{G}' \delta \mathbf{u}$$

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

## Adjoint Method (continued 1)

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

- Scalar function of output

$$J(v) = J[G(u)]$$

Gradient  $\nabla_u J$  of  $J$  with respect to input  $u$ ?

‘Chain rule’

$$\partial J / \partial u_i = \sum_j \partial J / \partial v_j (\partial v_j / \partial u_i)$$

or

$$\nabla_u J = G'^T \nabla_v J \quad (\text{A})$$

## Adjoint Method (continued 2)

$G$  is the composition of a number of successive steps

$$G = G_N \circ \dots \circ G_2 \circ G_1$$

'Chain rule'

$$G' = G_N' \dots G_2' G_1'$$

Transpose

$$G'^T = G_1'^T G_2'^T \dots G_N'^T$$

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

If  $G$  is nonlinear, local jacobian  $G'$  depends on local value of input  $u$ . 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)

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

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

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

Adjoint equation

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

....

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

....

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

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

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

## Adjoint Method (continued 3)

### Nonlinearities ?

$$\begin{aligned} \mathcal{J}(\xi_0) &= (1/2) (x_0^b - \xi_0)^T [P_0^b]^{-1} (x_0^b - \xi_0) + (1/2) \sum_k [y_k - H_k(\xi_k)]^T R_k^{-1} [y_k - H_k(\xi_k)] \\ &\text{subject to } \xi_{k+1} = M_k(\xi_k), \quad k = 0, \dots, K-1 \end{aligned}$$

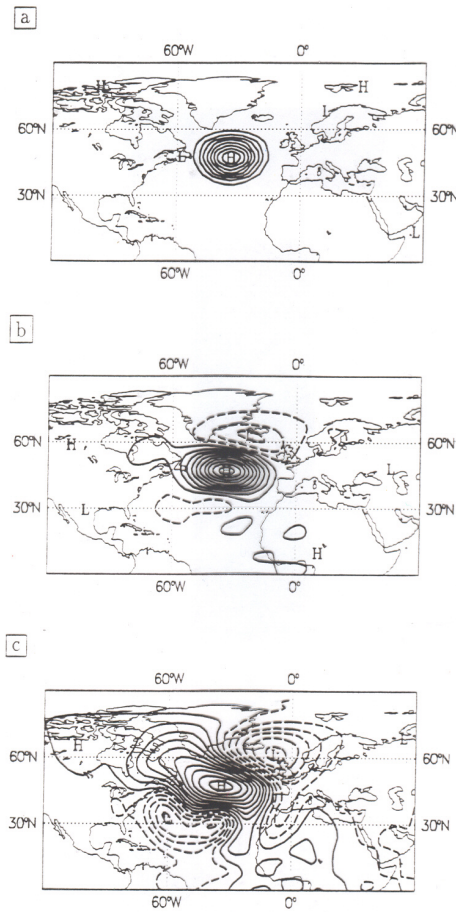
Control variable  $\xi_0 = u$

### Adjoint equation

$$\begin{aligned} \lambda_K &= H_K'^T R_K^{-1} [H_K(\xi_K) - y_K] \\ \dots & \\ \lambda_k &= M_k'^T \lambda_{k+1} + H_k'^T R_k^{-1} [H_k(\xi_k) - y_k] \quad k = K-1, \dots, 1 \\ \dots & \\ \lambda_0 &= M_0'^T \lambda_1 + H_0'^T R_0^{-1} [H_0(\xi_0) - y_0] + [P_0^b]^{-1} (\xi_0 - x_0^b) \end{aligned}$$

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

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.



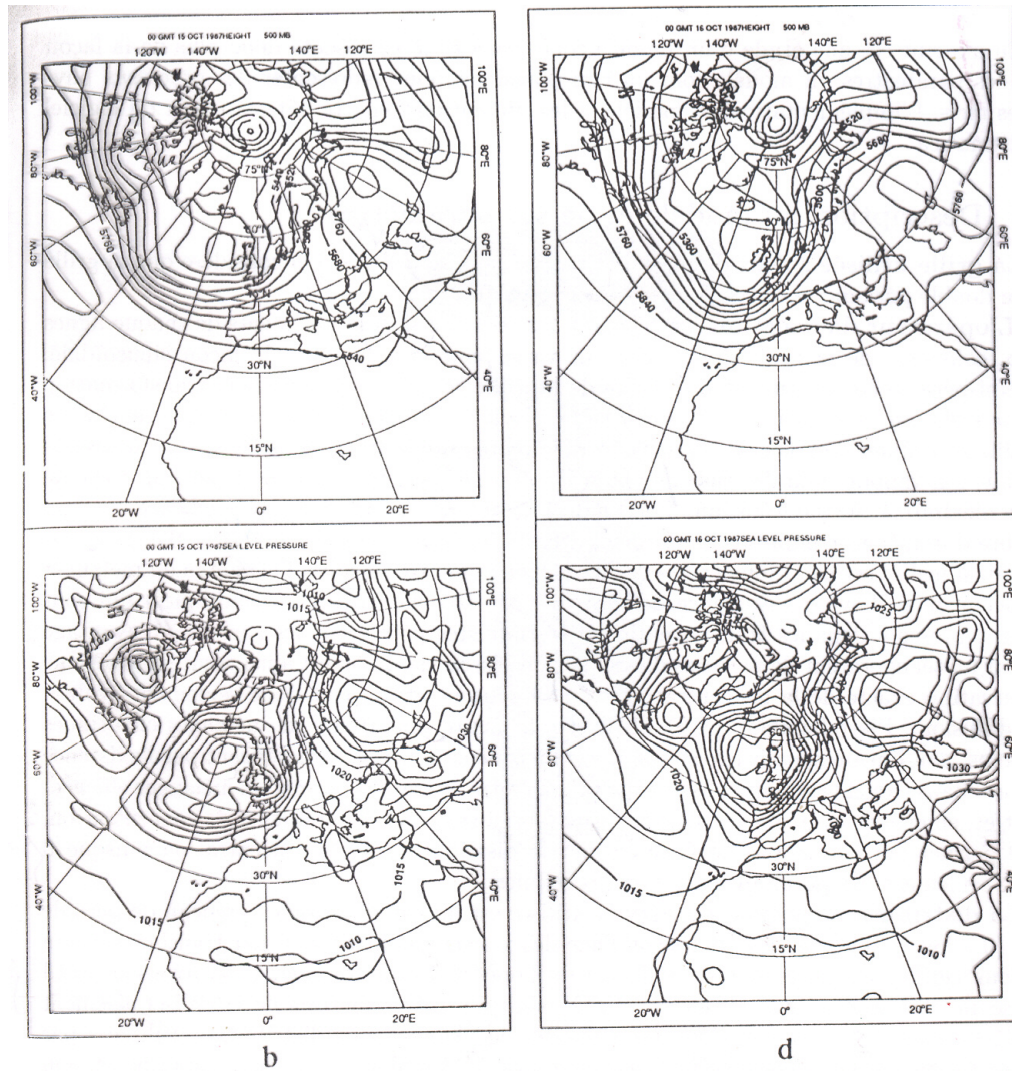
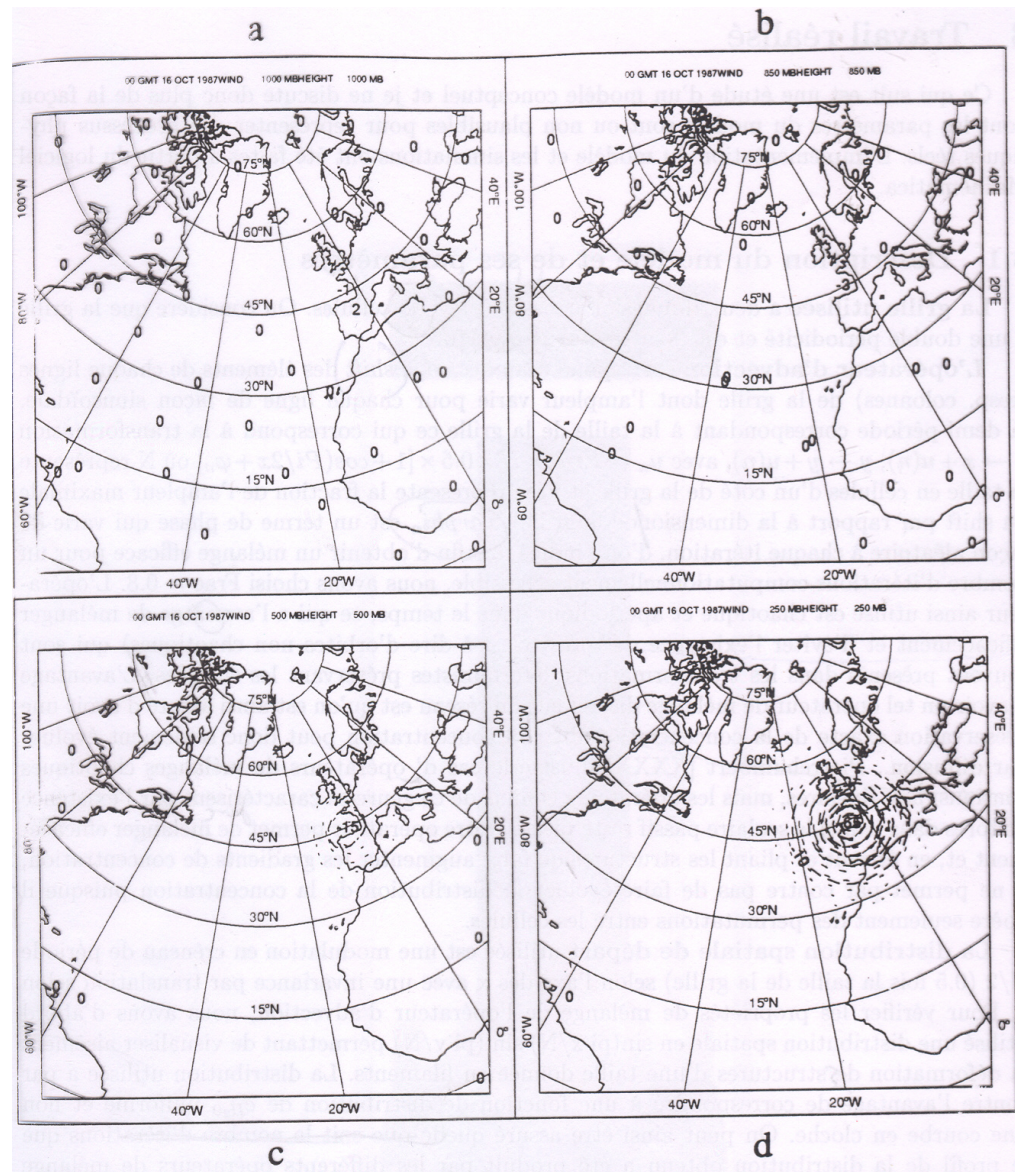
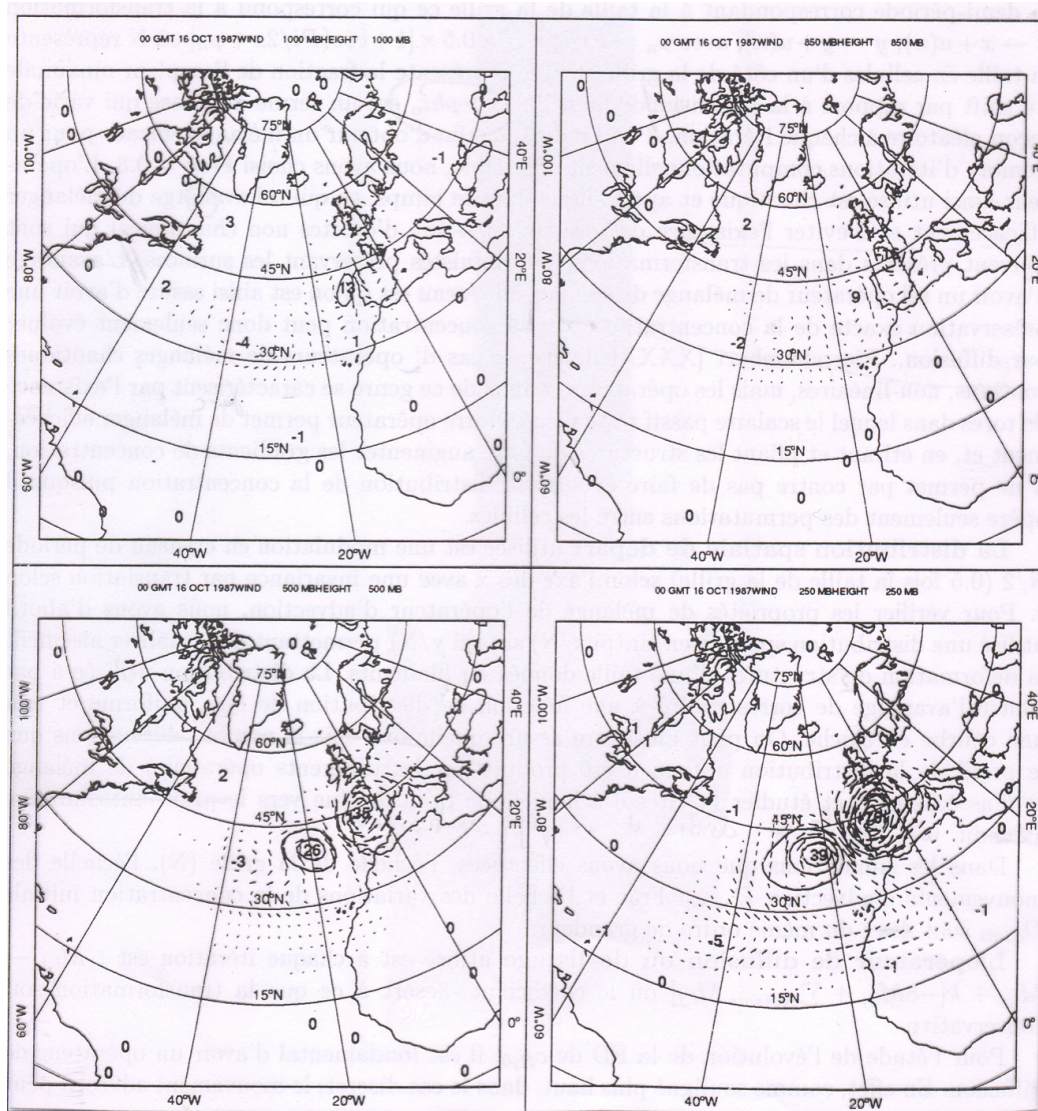


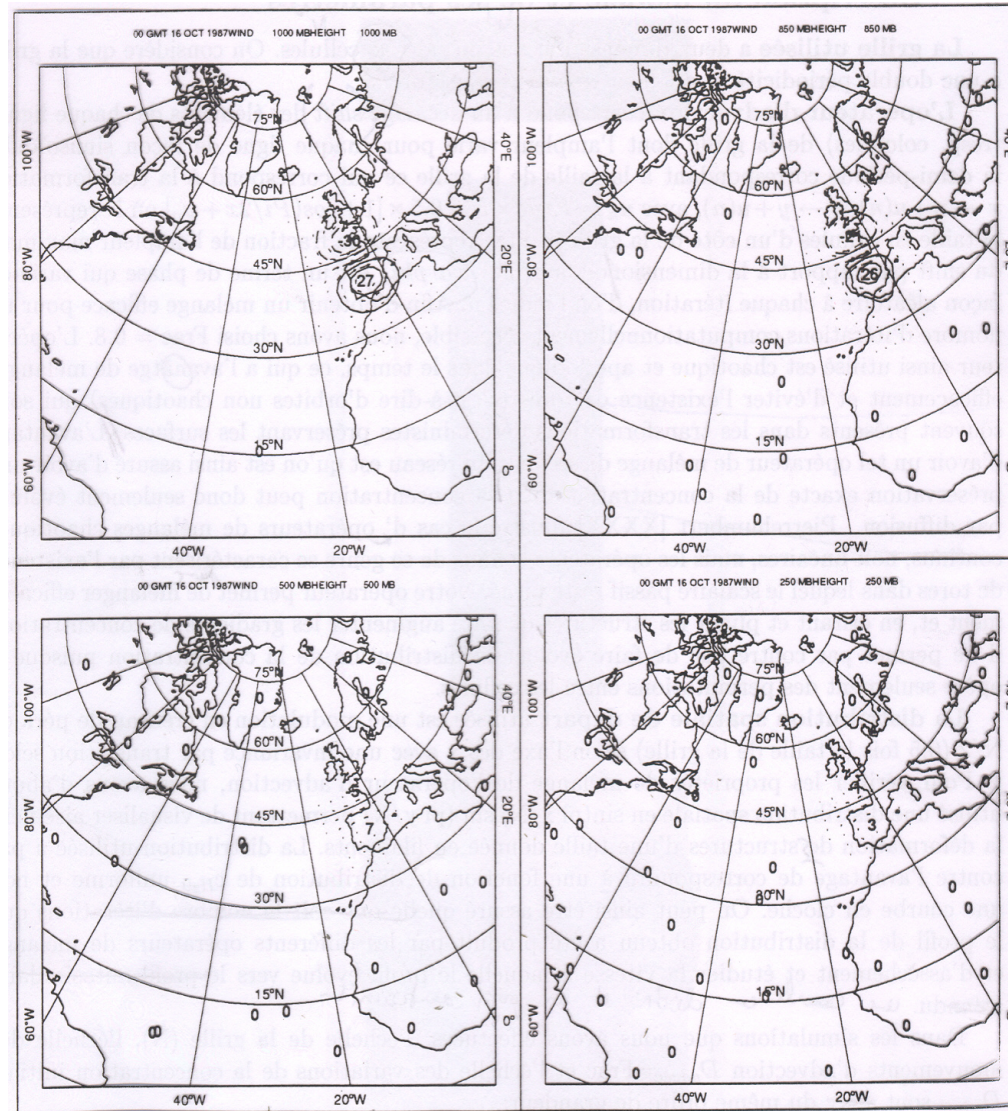
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October 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.



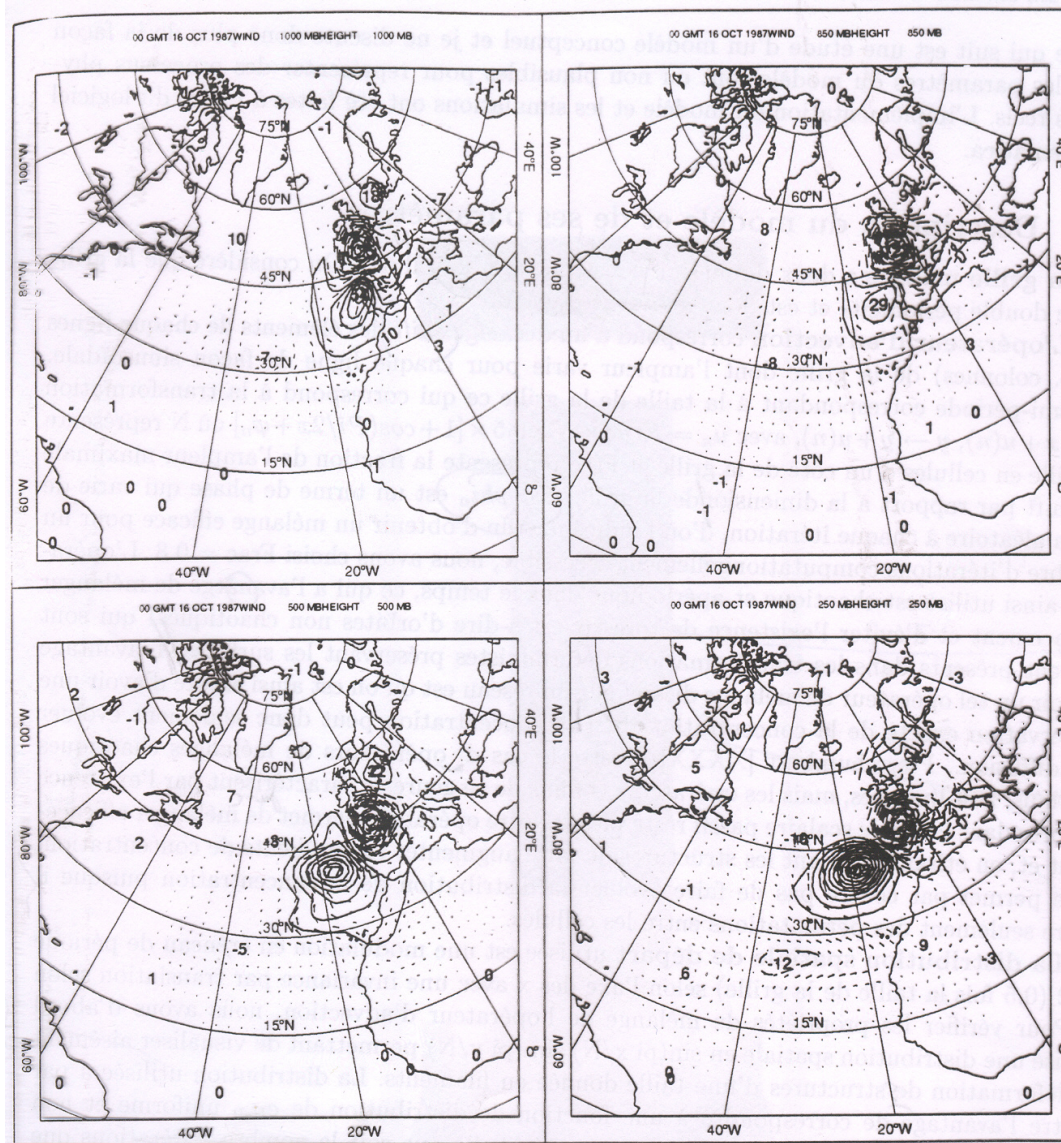
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)<sub>42</sub>



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

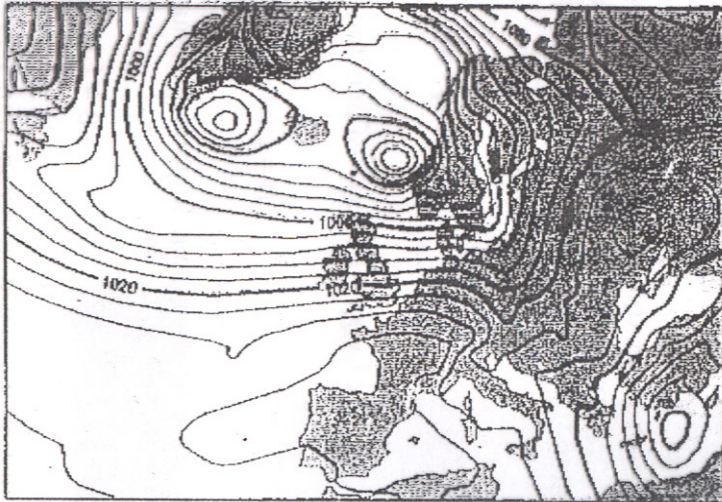


Analysis increments in a 3D-Var corresponding to a  $u$ -component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

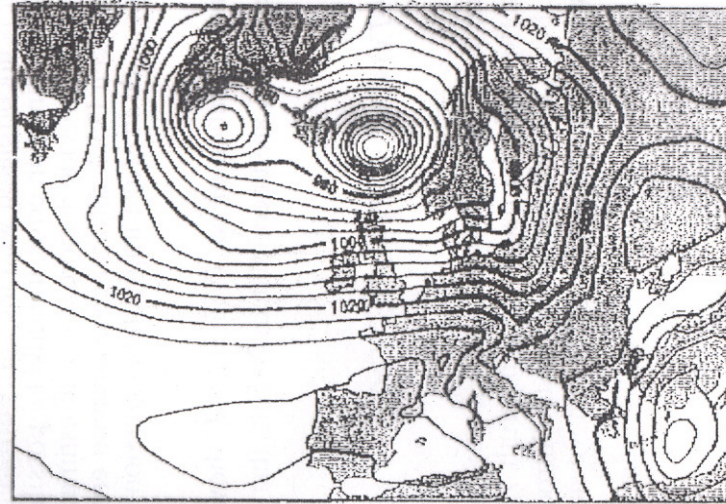


Same as before, but at the end of a 24-hr 4D-Var

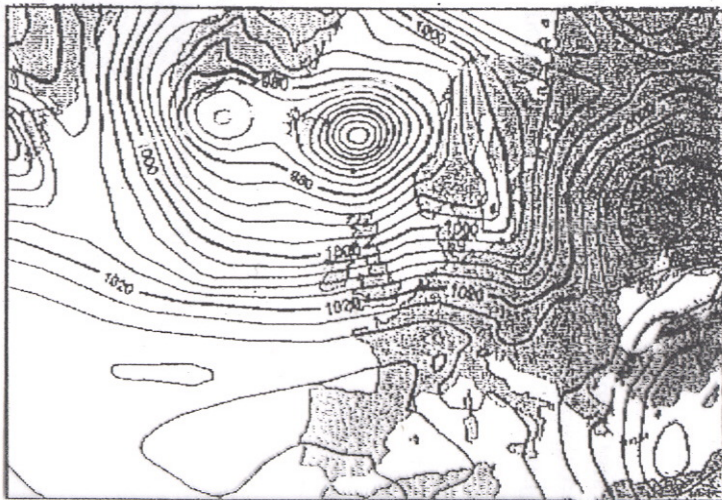
3-day forecast from 3D-Var analysis



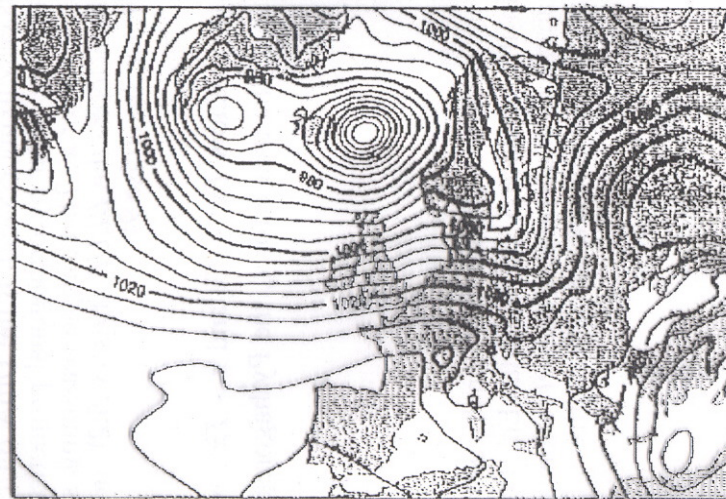
3-day forecast from 4D-Var analysis



3D-Var verifying analysis



4D-Var verifying analysis



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

## Cours à venir

~~Jeudi 19 Mars~~

~~Jeudi 26 mars~~

~~Jeudi 02 avril~~

~~Jeudi 09 avril~~

~~Mardi 21 avril, 14h00~~

Mardi 28 avril, 14h00

Mardi 12 mai, 14h00

Dernier cours XXX