École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2022-2023

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 5

2 Mai 2023

From course 3

Best Linear Unbiased Estimate

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

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

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

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

$E(\boldsymbol{y}) = \boldsymbol{Hx}^b$, $\boldsymbol{y}' = \boldsymbol{y} - E(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{Hx}^b = \boldsymbol{\varepsilon} - \boldsymbol{H\zeta}^b$ (**$\boldsymbol{H}$** is linear)

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

From course 3

Best Linear Unbiased Estimate

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

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

Equivalent set of formulæ

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

Vector $\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b$ is *innovation vector*
Matrix $\boldsymbol{K} \equiv \boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} = \boldsymbol{P}^a \boldsymbol{H}^T \boldsymbol{R}^{-1}$ is *gain matrix*.

If couple (ζ^b, ε) is 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 3

Best Linear Unbiased Estimate

Variational form of the *BLUE*

BLUE $\boldsymbol{x}^a$ minimizes following scalar *objective function*, defined on state space

$\xi \in S \rightarrow$

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

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

$$\boldsymbol{P}^a = [\partial^2\mathcal{J}/\partial\xi^2]^{-1}$$ (inverse *Hessian*)

'3D-Var'

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

Used operationally in USA, Australia, China, …

- Assimilation variationnelle. Principe

- Méthode adjointe. Principe.

- Assimilation variationnelle. Résultats

- La Méthode incrémentale

- Compléments sur l'Estimation Statistique (*BLUE*)

Case of data that are distributed over time

Suppose for instance available data consist of

- Background estimate at time 0

  $\boldsymbol{x}_0{}^b = \boldsymbol{x}_0 + \zeta_0{}^b \qquad E(\zeta_0{}^b \zeta_0{}^{bT}) = \boldsymbol{P}_0{}^b$

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

  $\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{\varepsilon}_k \qquad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j{}^{\mathrm{T}}) = \boldsymbol{R}_k \, \delta_{kj}$

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

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

Errors assumed to be unbiased and uncorrelated in time, $\boldsymbol{H}_k$ and $\boldsymbol{M}_k$ linear

Then objective function

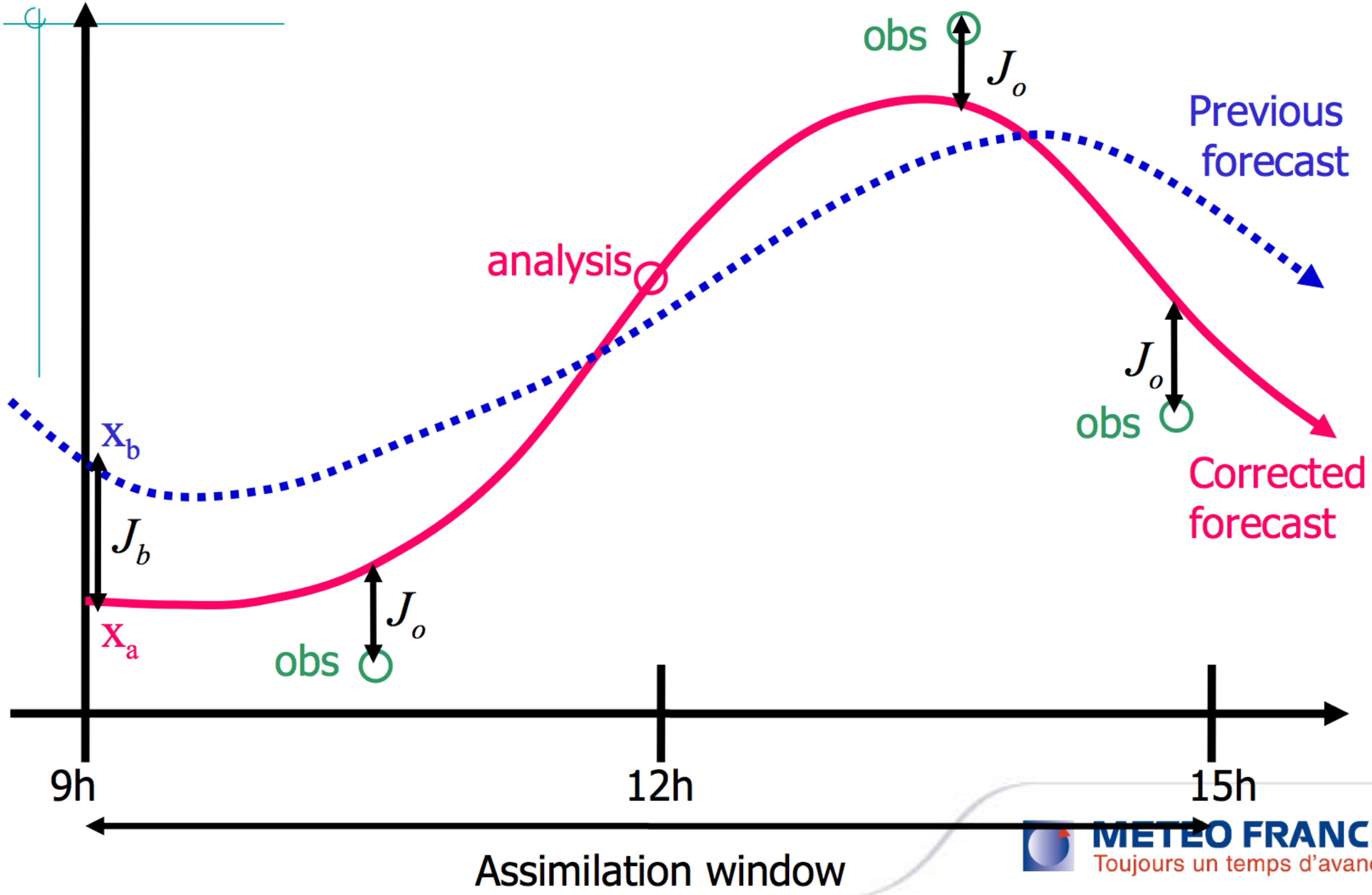
$\xi_0 \in S \rightarrow$

$$\mathcal{J}(\xi_0) \equiv (1/2)\,(\boldsymbol{x}_0{}^b - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0{}^b]^{-1}\,(\boldsymbol{x}_0{}^b - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k{}^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]$$

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

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

Principle of 4D-VAR assimilation



$$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0{}^b - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0{}^b]^{-1}\,(\boldsymbol{x}_0{}^b - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k{}^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]$$

subject to $\xi_{k+1} = \boldsymbol{M}_k\xi_k$, $\quad k = 0, \ldots, K\text{-}1$

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 $\boldsymbol{u} = \boldsymbol{\xi}_0$ ($\boldsymbol{u}$ 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 $\boldsymbol{u}$.

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 $\boldsymbol{u}$. Practically impossible.

Gradient computed by *adjoint method*.

- Méthode adjointe. Principe.

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).

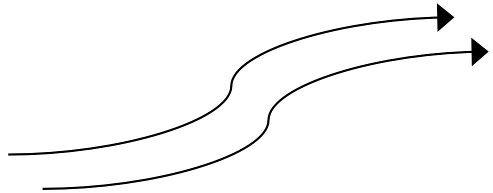
If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

A few basics

- Basic (nonlinear) model

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

- Perturbation $\delta\boldsymbol{x}_0$ at time 0. Resulting perturbation $\delta\boldsymbol{x}_k$ evolves in time according to



$$\delta x_{k+1} = \boldsymbol{M}_k(\boldsymbol{x}_k + \delta\boldsymbol{x}_k) - \boldsymbol{M}_k(\boldsymbol{x}_k)$$

$$= \boldsymbol{M}_k\text{'}(x_k)\ \delta\boldsymbol{x}_k + o(\delta\boldsymbol{x}_0)$$

where $\boldsymbol{M}_k\text{'}(\boldsymbol{x}_k)$ is jacobian of $\boldsymbol{M}_k$ at point x_k.

$$\delta\boldsymbol{\xi}_{k+1} = \boldsymbol{M}_k\text{'}(\boldsymbol{x}_k)\ \delta\boldsymbol{\xi}_k$$

is *tangent linear model* along solution x_k.

A few basics (continuation)

Tangent linear model

$$\delta\boldsymbol{\xi}_{k+1} = \mathbf{M}_k'(x_k)\,\delta\boldsymbol{\xi}_k$$

Adjoint model

$$\boldsymbol{\lambda}_k = [\mathbf{M}_k'(\boldsymbol{x}_k)]^{\mathrm{T}}\,\boldsymbol{\lambda}_{k+1}$$

Describes evolution with respect to k of gradient of a scalar function $\mathcal{J}$ with respect to $\boldsymbol{x}_k$.

Adjoint Method (continued 3)

$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0^b - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0^b]^{-1}\,(\boldsymbol{x}_0^b - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k\xi_k]$

subject to $\xi_{k+1} = \boldsymbol{M}_k\xi_k$, $\quad k = 0, \ldots, K\text{-}1$

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

Adjoint equation

$\lambda_K = \quad \boldsymbol{H}_K^{\mathrm{T}}\,\boldsymbol{R}_K^{-1}\,[\boldsymbol{H}_K\,\xi_K - \boldsymbol{y}_K]$

….

$\lambda_k = \boldsymbol{M}_k^{\mathrm{T}}\lambda_{k+1} + \boldsymbol{H}_k^{\mathrm{T}}\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{H}_k\,\xi_k - \boldsymbol{y}_k] \quad k = K\text{-}1, \ldots, 1$

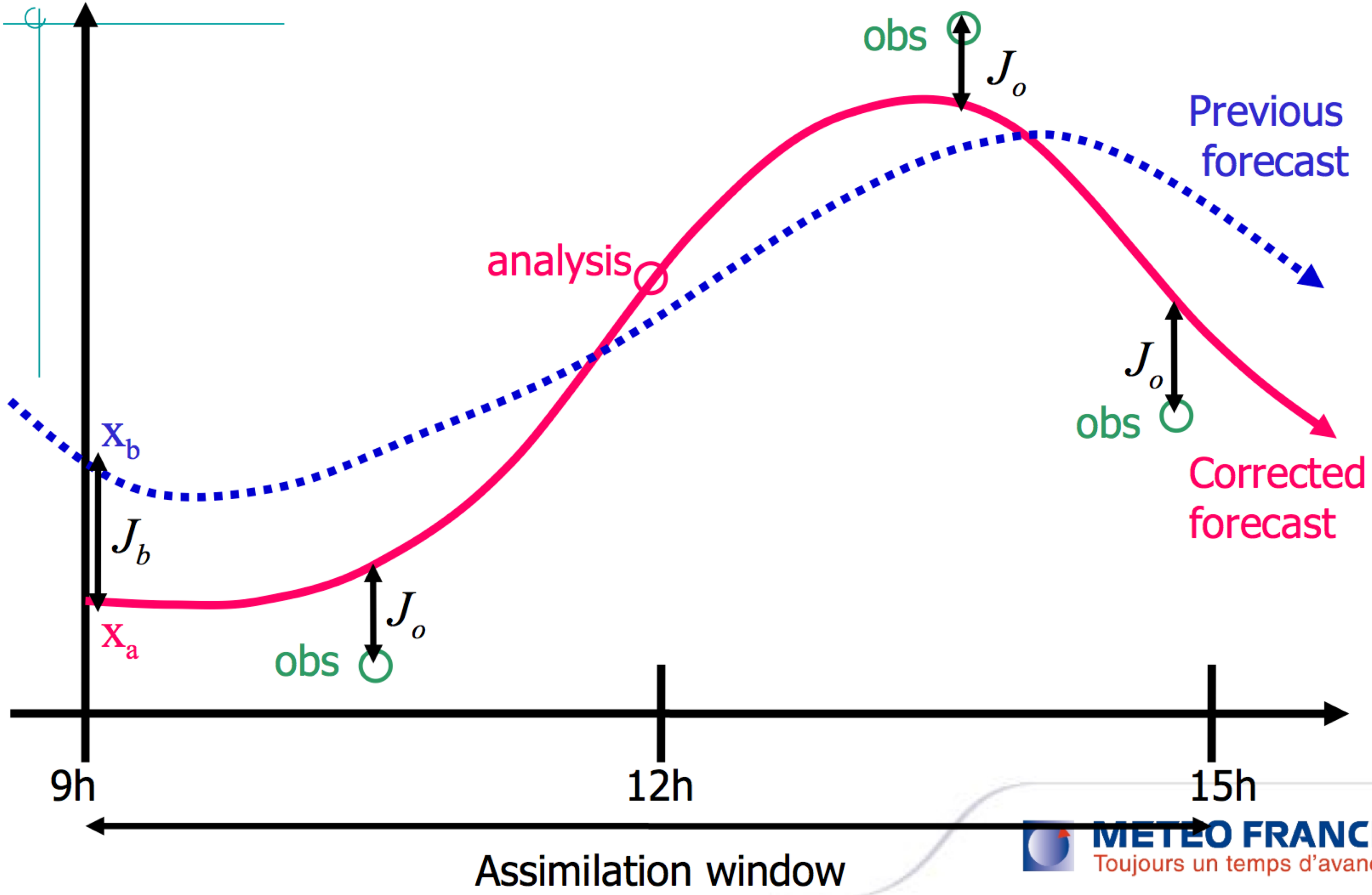
….

$\lambda_0 = \boldsymbol{M}_0^{\mathrm{T}}\lambda_1 + \boldsymbol{H}_0^{\mathrm{T}}\,\boldsymbol{R}_0^{-1}\,[\boldsymbol{H}_0\,\xi_0 - \boldsymbol{y}_0] + [\boldsymbol{P}_0^b]^{-1}\,(\xi_0 - \boldsymbol{x}_0^b)$

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

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

Principle of 4D-VAR assimilation



METEO FRANCE
Toujours un temps d'avance

Adjoint Method (continued 4)

Nonlinearities ?

$$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0^b - \xi_0)^T\,[\boldsymbol{P}_0^b]^{-1}\,(\boldsymbol{x}_0^b - \xi_0) + (1/2)\,\Sigma_k[\boldsymbol{y}_k - \boldsymbol{H}_k(\xi_k)]^T\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{y}_k - \boldsymbol{H}_k(\xi_k)]$$

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

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

Adjoint equation

$$\lambda_K = \boldsymbol{H}_K{}'^T\,\boldsymbol{R}_K^{-1}\,[\boldsymbol{H}_K(\xi_K) - \boldsymbol{y}_K]$$

….

$$\lambda_k = \boldsymbol{M}_k{}'^T\lambda_{k+1} + \boldsymbol{H}_k{}'^T\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{H}_k(\xi_k) - \boldsymbol{y}_k] \qquad k = K\text{-}1, \ldots, 1$$

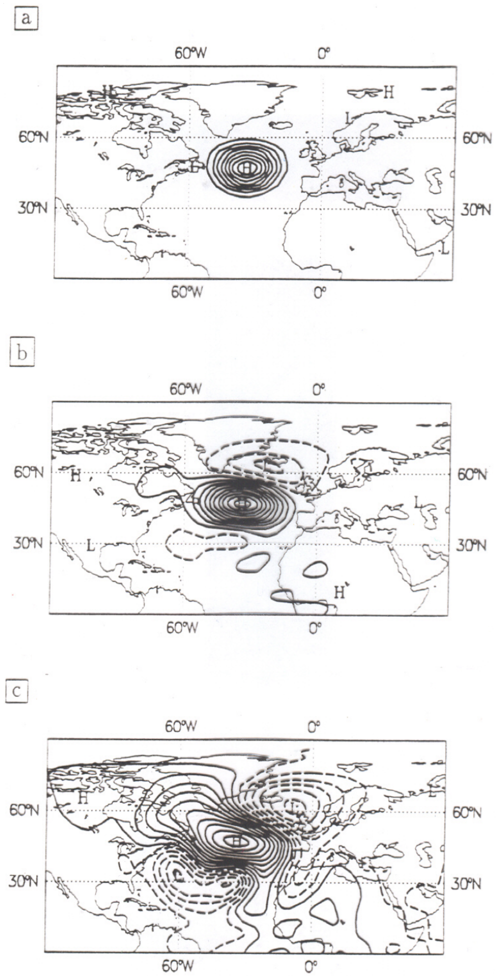
….

$$\lambda_0 = \boldsymbol{M}_0{}'^T\lambda_1 + \boldsymbol{H}_0{}'^T\,\boldsymbol{R}_0^{-1}\,[\boldsymbol{H}_0(\xi_0) - \boldsymbol{y}_0] + [\boldsymbol{P}_0^b]^{-1}\,(\xi_0 - \boldsymbol{x}_0^b)$$

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

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

\- Assimilation variationnelle. Résultats



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.

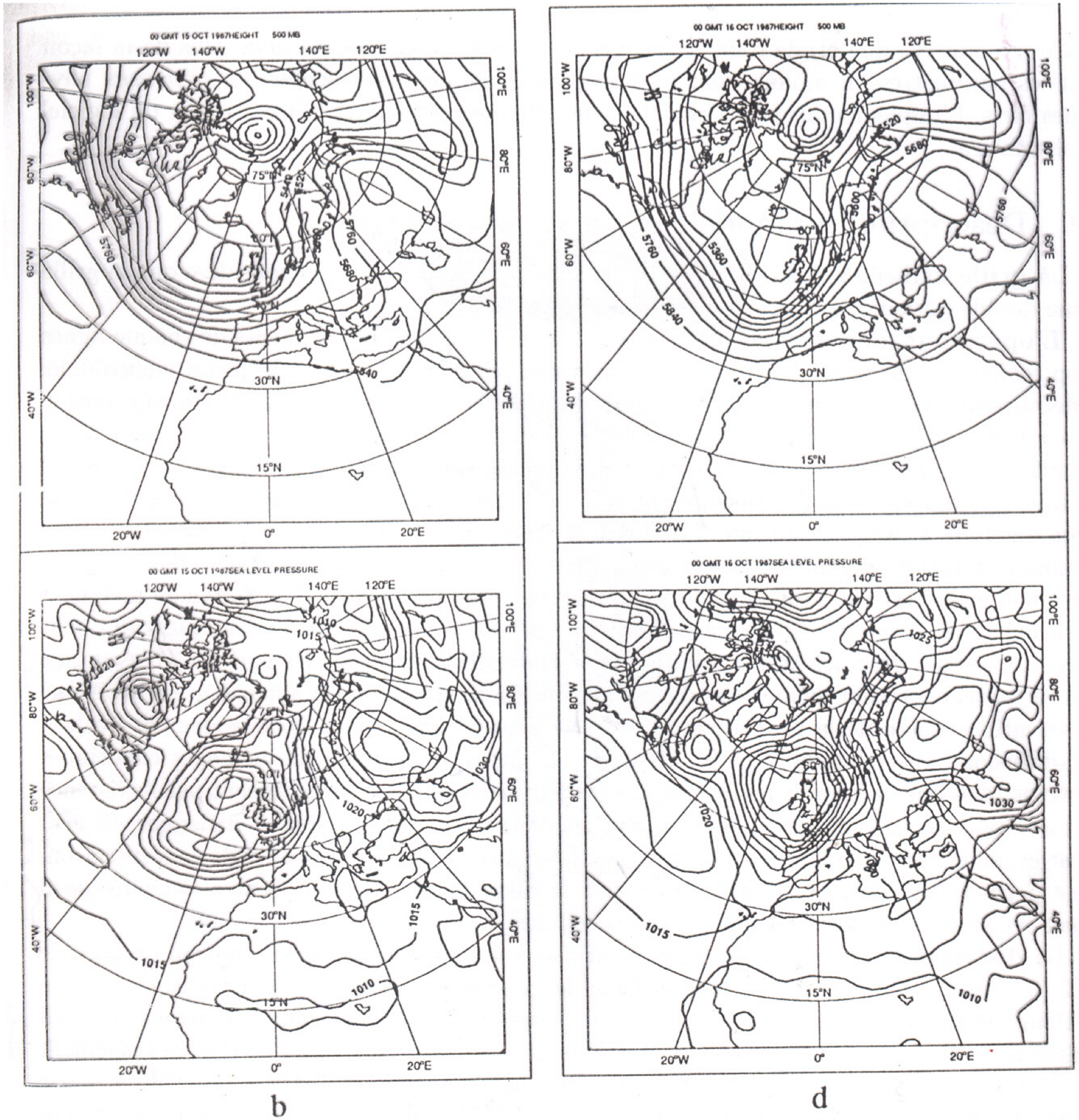
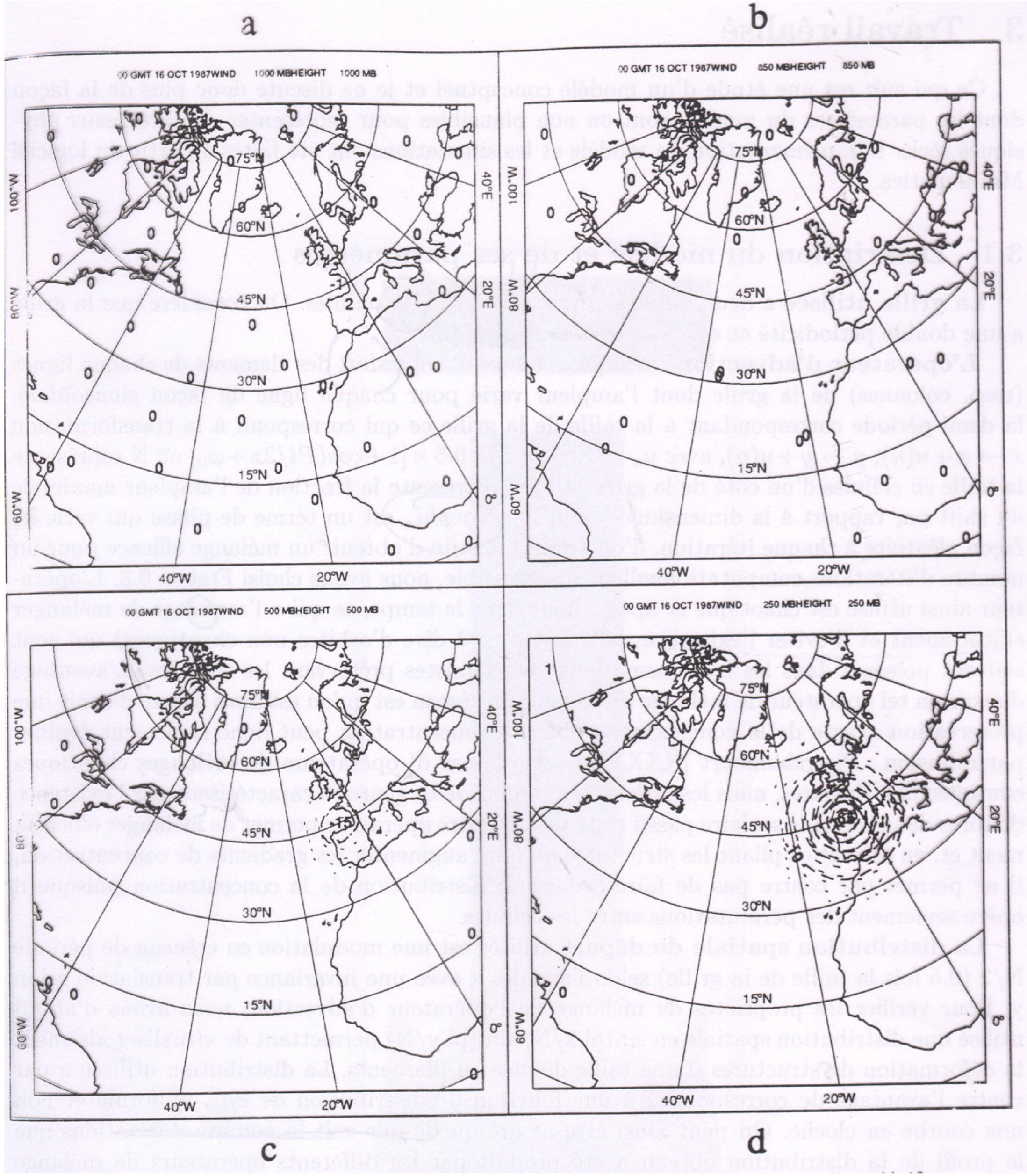


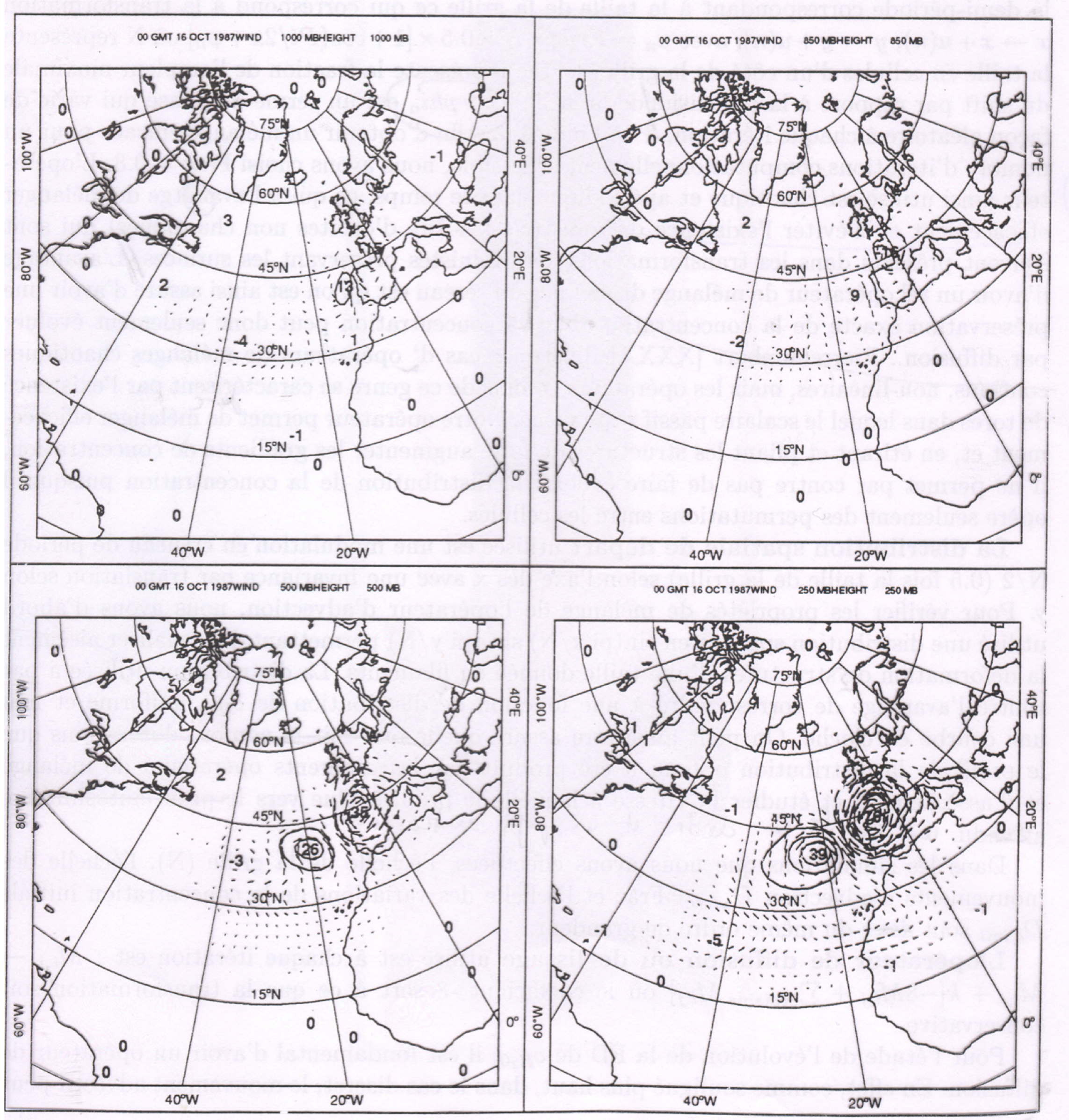
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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

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



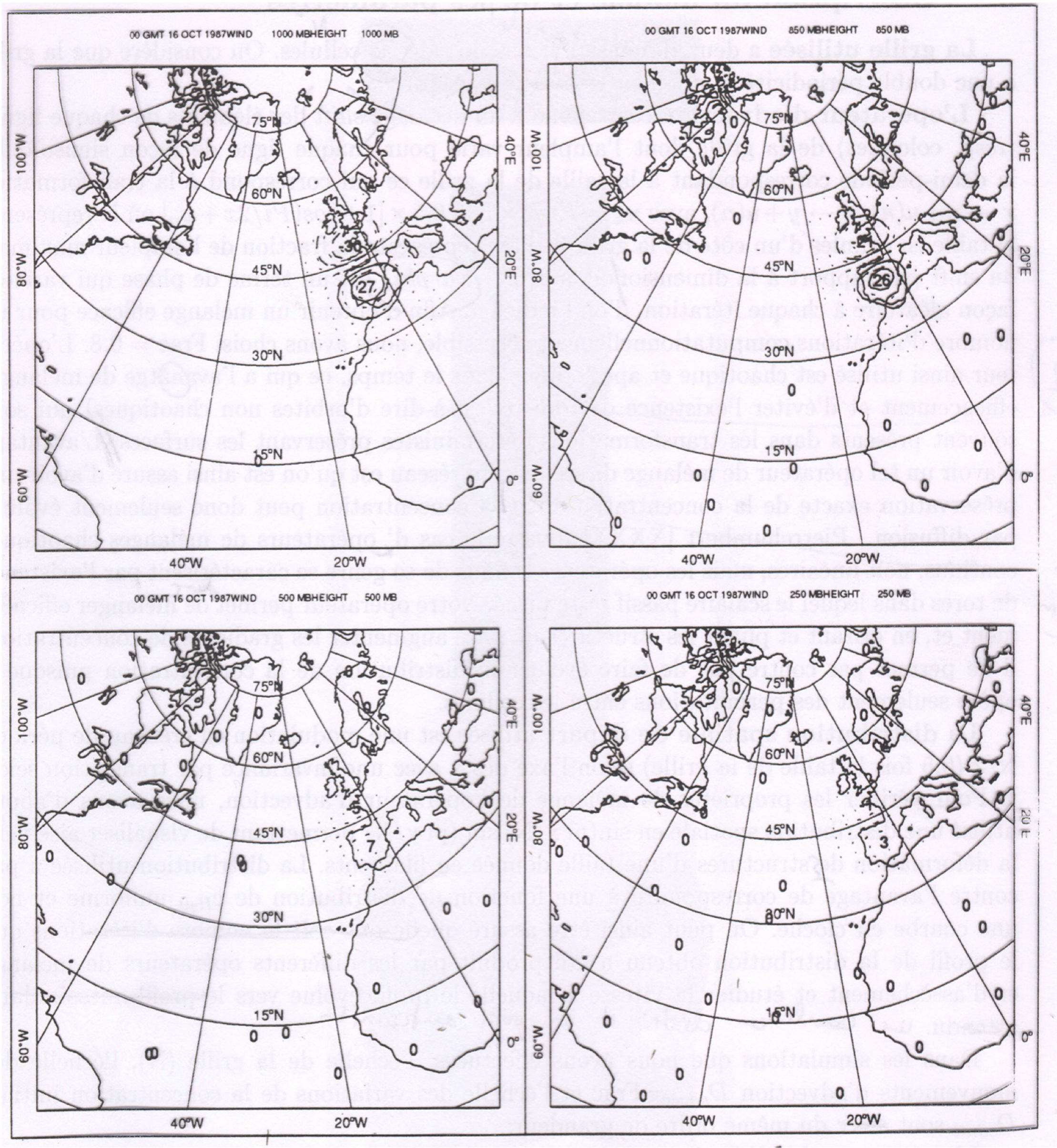
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



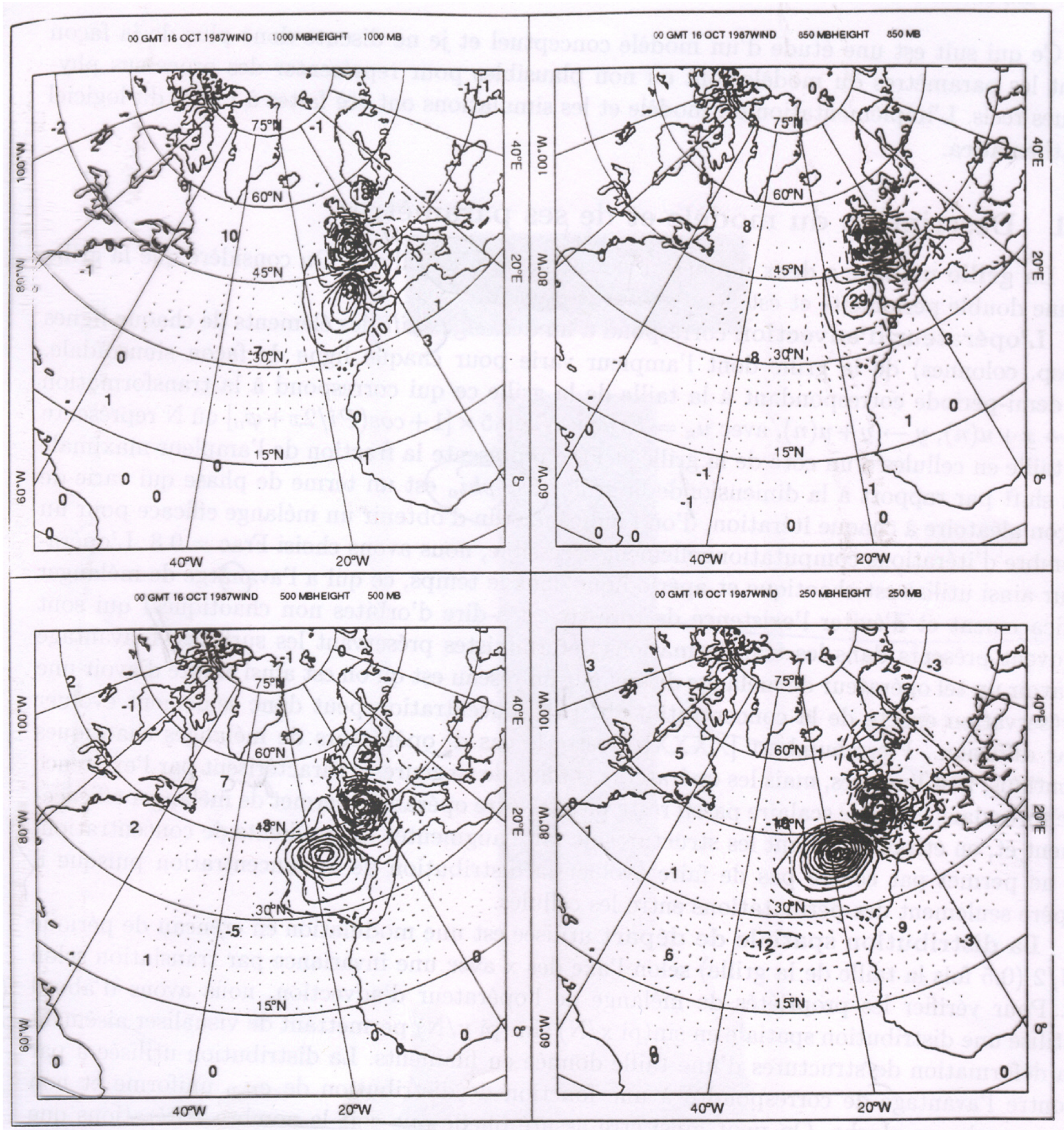
Same as before, but at the end of a 24-hr 4D-Var

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



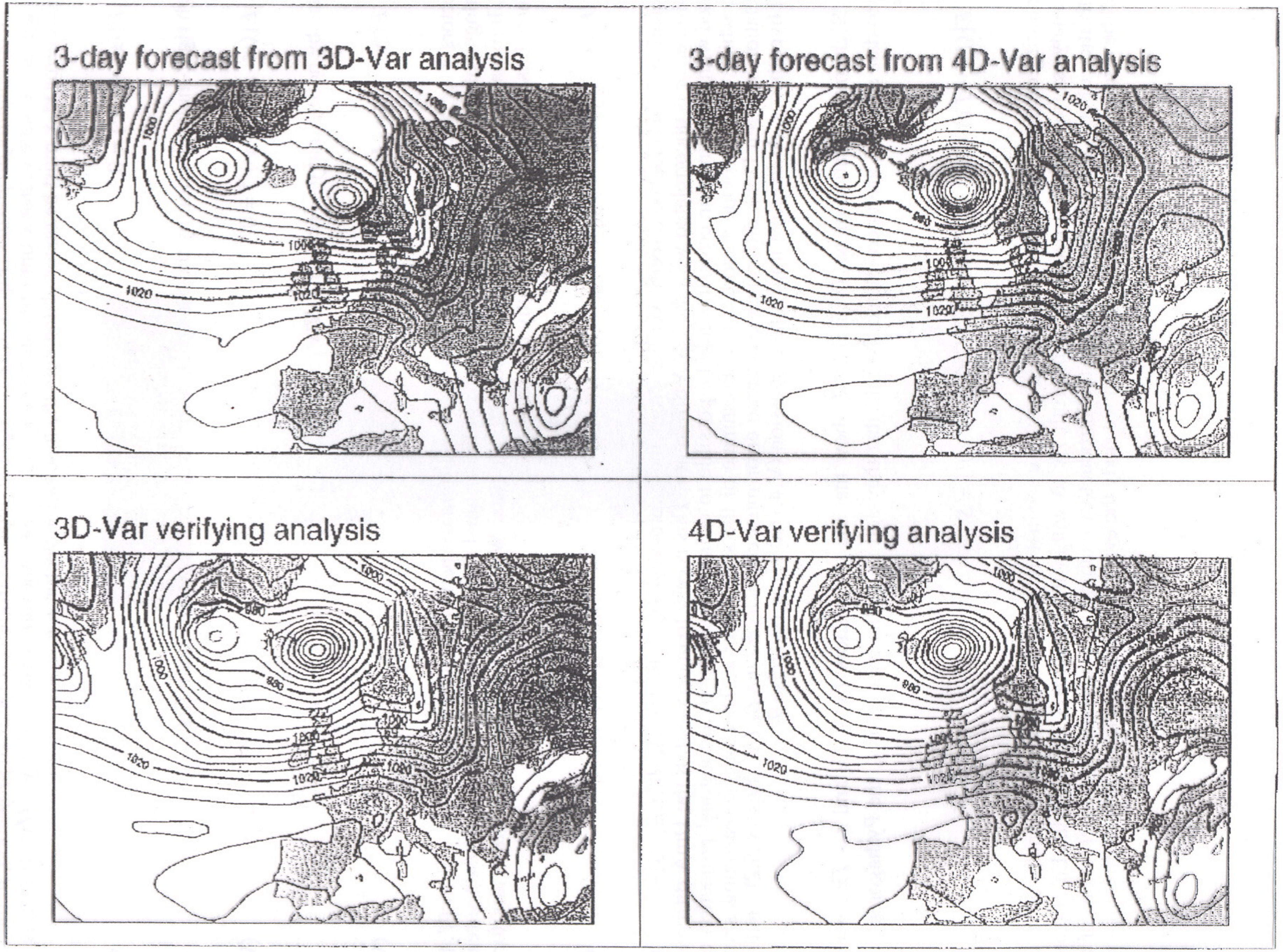
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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)

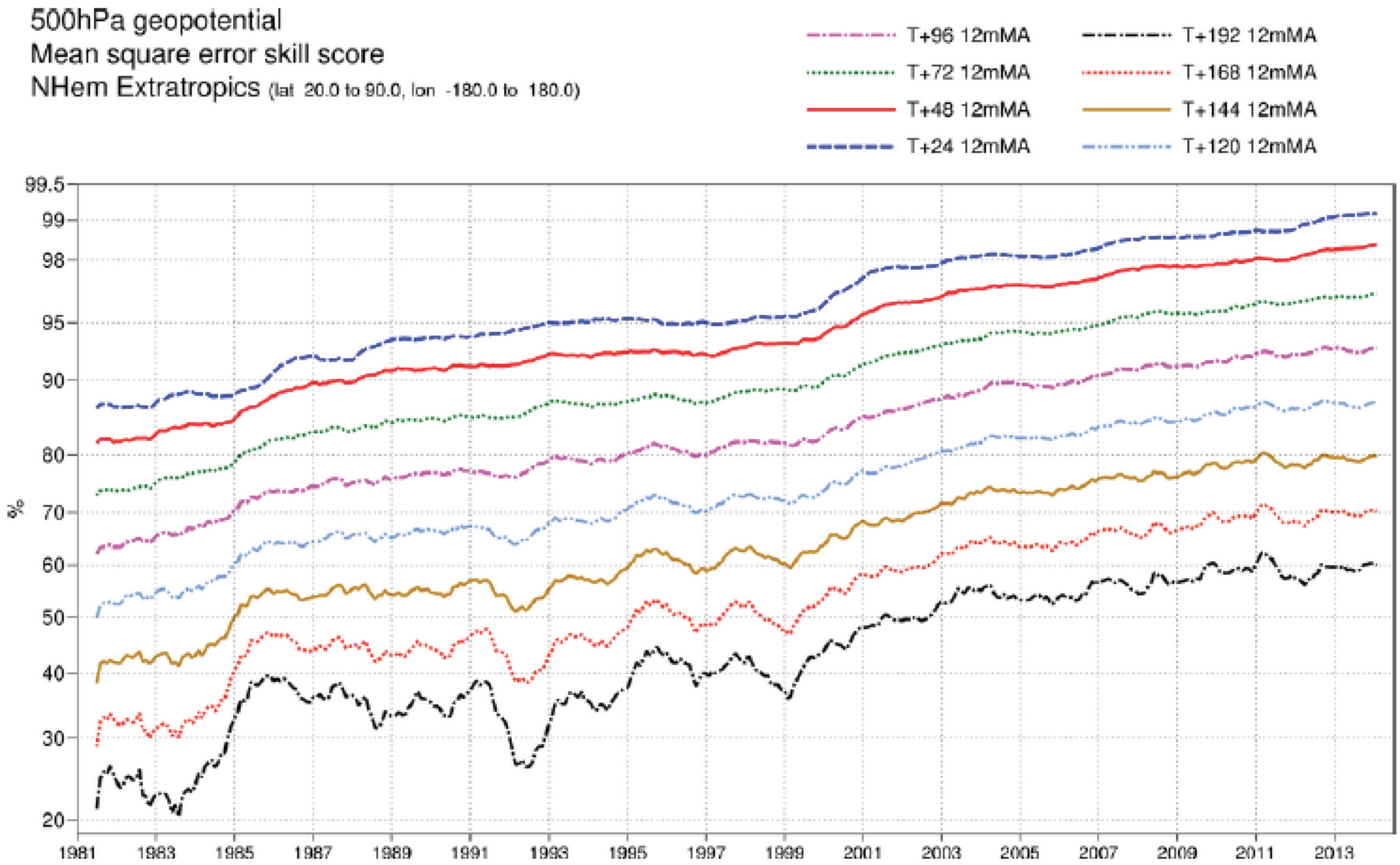
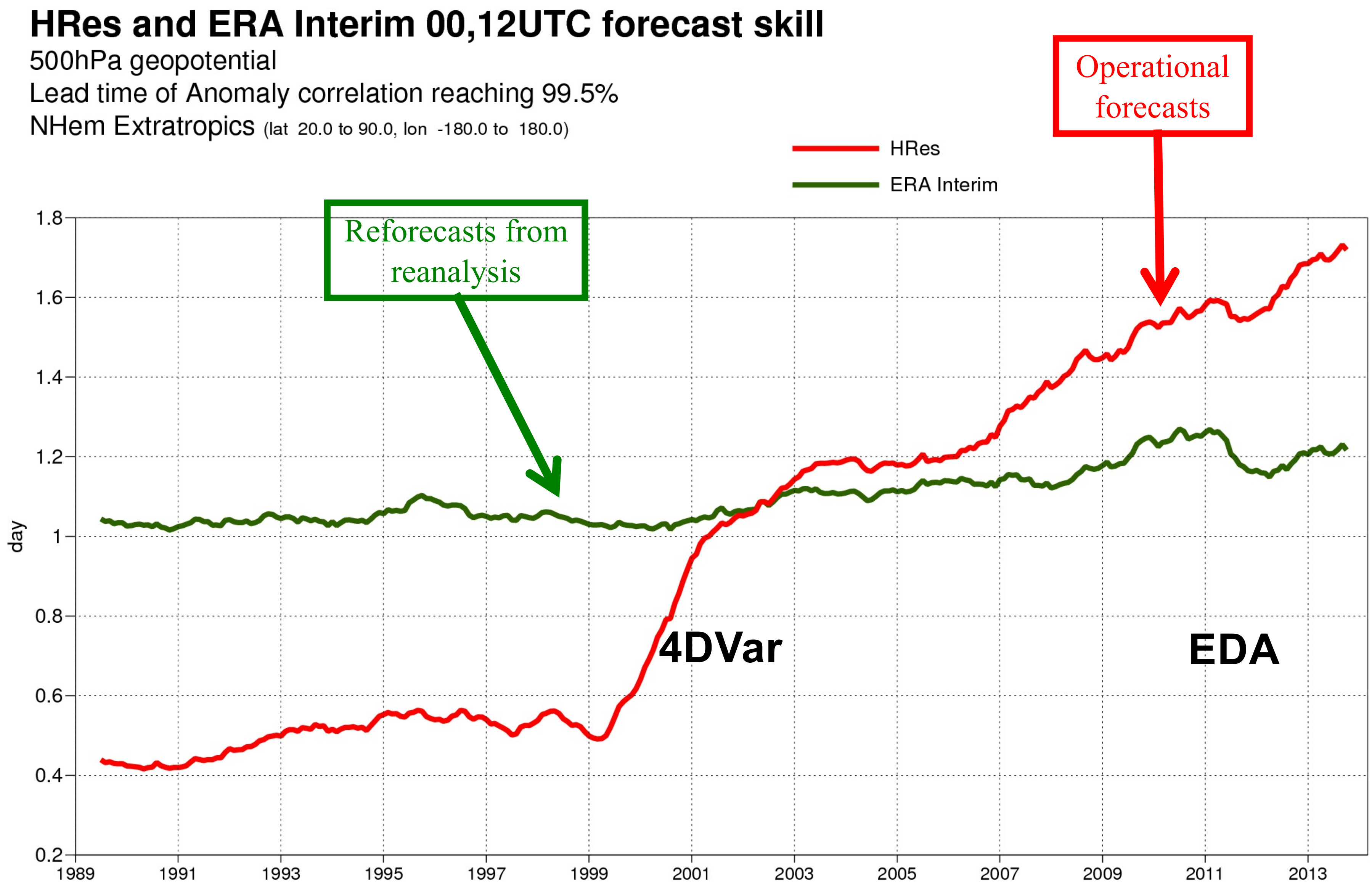


Figure 3: 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

Persistence = 0 ; climatology = 50 at long range

Initial state error reduction



Credit E. Källén, ECMWF

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.

Time-correlated Errors (*continuation from course 4*)

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

$$\mathcal{R} = (\boldsymbol{R}_{kk'} = E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_{k'}^{\mathrm{T}}))$$

Objective function

$$\xi_0 \in S \rightarrow$$

$$\mathcal{J}(\xi_0) = (1/2)\,(\boldsymbol{x}_0^b - \xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0^b]^{-1}\,(\boldsymbol{x}_0^b - \xi_0) + (1/2)\,\Sigma_{kk'}[\boldsymbol{y}_k - \boldsymbol{H}_k \xi_k]^{\mathrm{T}}\,[\mathcal{R}^{-1}]_{kk'}\,[\boldsymbol{y}_{k'} - \boldsymbol{H}_{k'} \xi_{k'}]$$

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 ?

In the linear case, Kalman Smoother and Variational Assimilation are algorithmically equivalent. If errors are uncorrelated in time, 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 globally Gaussian, both algorithms achieve Bayesian estimation.

If errors are correlated in time, only some Kalman Smoothers are equivalent with Variational Assimilation.

- La Méthode incrémentale

Incremental Method for Variational Assimilation

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

$\xi_{k+1} = \boldsymbol{M}_k(\xi_k)$

- Tangent linear model

$\delta\xi_{k+1} = \boldsymbol{M}_k{}' \, \delta\xi_k$

where $\boldsymbol{M}_k{}'$ is jacobian of $\boldsymbol{M}_k$ at point ξ_k.

- Adjoint model

$\lambda_k = \boldsymbol{M}_k{}'^{\mathrm{T}} \lambda_{k+1} + \ldots$

Incremental Method. Simplify both $\boldsymbol{M}_k{}'$ and $\boldsymbol{M}_k{}'^{\mathrm{T}}$ consistently.

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 $\boldsymbol{H}_k(\xi_k)$ by $\boldsymbol{H}_k(\xi_k^{(0)}) + \boldsymbol{N}_k\delta\xi_k$, where $\boldsymbol{N}_k$ is an appropriate 'simple' linear operator (possibly, but not necessarily, the jacobian of $\boldsymbol{H}_k$ at point $\xi_k^{(0)}$). The objective function depends only on the initial $\delta\xi_0$ deviation from $\xi_0^{(0)}$, and reads

$$\mathcal{J}_{\mathrm{I}}(\delta\xi_0) = (1/2)\,(\boldsymbol{x}_0^b - \xi_0^{(0)} - \delta\xi_0)^{\mathrm{T}}\,[\boldsymbol{P}_0^b]^{-1}\,(\boldsymbol{x}_0^b - \xi_0^{(0)} - \delta\xi_0)$$
$$+ (1/2)\,\Sigma_k[\boldsymbol{d}_k - \boldsymbol{N}_k\delta\xi_k]^{\mathrm{T}}\,\boldsymbol{R}_k^{-1}\,[\boldsymbol{d}_k - \boldsymbol{N}_k\delta\xi_k]$$

where $\boldsymbol{d}_k \equiv \boldsymbol{y}_k - \boldsymbol{H}_k(\xi_k^{(0)})$ is the innovation at time k, and the $\delta\xi_k$ evolve according to

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

With the choices made here, $\mathcal{J}_{\mathrm{I}}(\delta\xi_0)$ 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 $\boldsymbol{L}_k$ and $\boldsymbol{N}_k$, 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 $\boldsymbol{L}_k = \boldsymbol{I}_n$. 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.

- Compléments sur l'Estimation Statistique (*BLUE*)

From course 3

Best Linear Unbiased Estimate

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b$$

$$\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon}$$

$\boldsymbol{H}$ is known linear *observation operator.*

Assume probability distribution is known for the couple $(\boldsymbol{\zeta}^b, \boldsymbol{\varepsilon})$.
Assume $E(\boldsymbol{\zeta}^b) = 0$, $E(\boldsymbol{\varepsilon}) = 0$, $E(\boldsymbol{\zeta}^b \boldsymbol{\varepsilon}^T) = 0$ (not restrictive)
Set $E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{bT}) \equiv \boldsymbol{P}^b$ (also often denoted $\boldsymbol{B}$), $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) \equiv \boldsymbol{R}$

From course 3

Best Linear Unbiased Estimate

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

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

Equivalent set of formulæ

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

Vector $\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b$ is *innovation vector*
Matrix $\boldsymbol{K} \equiv \boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} = \boldsymbol{P}^a \boldsymbol{H}^T \boldsymbol{R}^{-1}$ is *gain matrix*.

If couple (ζ^b, ε) is Gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

Condition $E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}}) = 0$ is not mathematically restrictive. Setting $E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}}) \equiv \boldsymbol{D}$ (possibly $\neq 0$), and coming back to the general formula

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

with again $\boldsymbol{x}' = \boldsymbol{x} - E(\boldsymbol{x}) = -\,\boldsymbol{\zeta}^b$, $E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}}) = \boldsymbol{P}^b$

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

$$\boldsymbol{C}_{xy} = E(\boldsymbol{x}'\boldsymbol{y}'^{\mathrm{T}}) = E[-\boldsymbol{\zeta}^b(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)^{\mathrm{T}}] = -\underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\varepsilon}^{\mathrm{T}})}_{\boldsymbol{D}^{\mathrm{T}}} + \underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{P}^b}\boldsymbol{H}^{\mathrm{T}} = -\,\boldsymbol{D}^{\mathrm{T}} + \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{xy} = -\,\boldsymbol{D}^{\mathrm{T}} + \boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{yy} = E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}}) = E[(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)\,(\boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{\zeta}^b)^{\mathrm{T}}]$$
$$= \underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})}_{\boldsymbol{R}} - \underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{D}}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{H}\underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{D}^{\mathrm{T}}.} + \boldsymbol{H}\underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{P}^b}\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{yy} = \boldsymbol{R} - \boldsymbol{D}\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{H}\boldsymbol{D}^{\mathrm{T}} + \boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}$$

Leading to expressions

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

$$\boldsymbol{P}^a = \boldsymbol{P}^b - [\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{D}^{\mathrm{T}}]\ [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{D}\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{H}\boldsymbol{D}^{\mathrm{T}}+\boldsymbol{R}]^{-1}\ [\boldsymbol{H}\boldsymbol{P}^b-\boldsymbol{D}]$$

This is equivalent to replacing the observation vector $\boldsymbol{y}$ with the vector $\boldsymbol{v} \equiv \boldsymbol{y} - \boldsymbol{D}[\boldsymbol{P}^b]^{-1}\boldsymbol{x}^b$, the error of which is uncorrelated with ζ^b, and then using the formulæ for the case of no correlation between background and observation errors.

But the hypothesis of no correlation is almost always made in practice, although it is certainly not always verified (observations performed by a same satellite instrument, which have been through a same post-processing, are very likely to have correlated errors).

Now, taking into account correlations between backgound and observation errors does not render, as shown in course 4, the corresponding estimate optimal. That would require to modify the weights that have been given to previous data.

Bayesian Estimation

Data of the form

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

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

Probability that $\boldsymbol{x} = \xi$ given in $\mathcal{X}$? $\quad \boldsymbol{x} = \xi \Rightarrow \zeta = \boldsymbol{z} - \boldsymbol{\Gamma}\xi$

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

where

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

Then conditional probability distribution is

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

Bayesian Estimation (continuation 1)

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

Then

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

with

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

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

Variational form

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

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

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

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

If data still of the form

$$\boldsymbol{z} = \boldsymbol{\Gamma}\boldsymbol{x} + \boldsymbol{\zeta}$$

but 'error' $\boldsymbol{\zeta}$, which still has expectation 0 and covariance $\boldsymbol{S}$, is not Gaussian, expressions

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

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

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

Expressions

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

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

are invariant in linear invertible change of coordinates, in either data or state space. If determinacy condition is verified, data vector $\boldsymbol{z}$ can be transformed, through linear invertible change of coordinates in data space, into

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b$$

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

from which the formulæ derived previously can be obtained (in both cases $E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{b\mathrm{T}}) = 0$ and $\neq 0$) .

Expressions

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

are valid in both the Gaussian case and the general linear (*BLUE*) case. But, although, they are algebraically identical, they do not have the same significance. In the Gaussian case, as said, they solve entirely the problem of Bayesian estimation. For any data vector $\boldsymbol{z}$, $\boldsymbol{x}^a$ and $\boldsymbol{P}^a$ are respectively the expectation and covariance of the conditional (Gaussian) probability distribution $P(\boldsymbol{x} \mid \boldsymbol{z})$. In the general linear case, $\boldsymbol{x}^a$ and $\boldsymbol{P}^a$ are expectations over all possible realizations of $\boldsymbol{z}$ (*i.e.* of the error $\boldsymbol{\zeta}$). For a given $\boldsymbol{z}$, $\boldsymbol{x}^a$ and $\boldsymbol{P}^a$ can be very different from the corresponding Bayesian values.

Entropy of a probability distribution

Probability distribution over domain described by coordinate ξ, with probability density $p(\xi)$. *Entropy*

$$S \equiv -\int p \ln p \, d\xi$$

Entropy of a probability distribution is a measure of the associated uncertainty. The larger the entropy, the larger the uncertainty. A uniform probability distribution over an interval of length a has entropy $\ln a$, which tends to -∞ as a tends to zero. A one-dimensional Gaussian probability distribution with variance s has entropy $\ln\sqrt{(2\pi e s)}$.

For given variance s, entropy is largest for the Gaussian distribution.

Entropy of a probability distribution (continuation)

Data of the form (see slide 47)

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta}$$

The knowledge of a probability distribution for $\boldsymbol{\zeta}$ defines a conditional probability distribution $P(\boldsymbol{x}|\boldsymbol{z})$ for $\boldsymbol{x}$. Assuming that only the expectation and covariance matrix $\boldsymbol{S}$ of $\boldsymbol{\zeta}$ are known, for which distribution of $\boldsymbol{\zeta}$ is the entropy of $P(\boldsymbol{x}|\boldsymbol{z})$ largest ?

Response. The entropy of $P(\boldsymbol{x}|\boldsymbol{z})$ is largest when $\boldsymbol{\zeta}$ is Gaussian.

If the probability distribution for $\boldsymbol{\zeta}$ is unknown, assuming that it is Gaussian is in a sense the 'least committing' choice.

Convective Instability

In dry atmosphere in hydrostatic balance, adiabatic lapse rate (vertical gradient of temperature)

$$(dT/dz)_{ad} = -g/C_p$$

$g \approx 10\ m\ s^{-2}$, $C_p \approx 10^3$ SI, $-g/C_p \approx -10°\ C/km$

Water vapour is present in the atmosphere, and will usually condense, and emit heat, in an ascending motion. In practice, dT/dz is observed to have value about $-6°\ C/km$, which is close to *its adiabatic wet* value.

Reminder

Potential temperature

$\theta \equiv T(p_0 / p)^\kappa$ with $\kappa \equiv r/C_p$ (≈ 0.285 for dry air)

Potential temperature is conserved in adiabatic transformation

Stratified atmosphere at rest with temperature gradient dT/dz and associated gradient of potential temperature $d\theta/dz$.

Particle displaced adiabatically upward from its equilibrium position. Expands at pressure of background stratification.

- if background temperature larger than temperature of displaced particle, *i. e.* $dT/dz > (dT/dz)_{ad}$ (potential temperature increases with altitude), buoyancy force will pull particle back to its original position. Stratification is said to be *convectively stable*. Particle will oscillate with *Brunt-Väisälä frequency* N

$$N^2 \equiv (g/\theta)\,(d\theta/dz)$$

In the atmosphere, the corresponding period has typical value of a few minutes.

- if background temperature lower than temperature of displaced particle, *i. e.* $dT/dz < (dT/dz)_{ad}$ (potential temperature decreases with altitude), particle will move farther away from its original position $\Rightarrow$ *convective instability*

Convective instability is at the origin of intense *convective cells* (cumulus clouds, thunderstorms), with core of intense ascending motion surrounded by slower subsiding motion. Convective instability is the main process through which energy is carried from the lower surface (continents, oceans) into the atmosphere. It also carries water and momentum.

Convection occurs also in the ocean, when the upper surface is cooled by radiation.

Another similar phenomenon occurs in the ocean (but with no thermodynamical effects involved) when dense water (whose salinity has been increased by evaporation) is transported (for instance by wind) above less dense water. This phenomenon is a component of the *thermohaline* circulation.

Convective instability, which results from a deviation of hydrostatic balance, is not represented in hydrostatic models, and has to be appropriately parametrised in those models.

Cours à venir

~~Mardi 21 mars~~
~~Mardi 28 mars~~
~~Mardi 4 avril~~
~~Mardi 11 avril~~
~~Mardi 2 mai~~
Mardi 9 mai
Mardi 23 mai
Mardi 30 mai