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 6

14 Mai 2021

- Variational assimilation. Principle. The adjoint approach. Results

- The ‘Incremental approach’ to variational assimilation

- Weak constraint variational assimilation. Principle. The dual algorithm for variational assimilation. Examples.

or

- History of Numerical Weather Prediction

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

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

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

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

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

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

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

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.

Estimate x^a and associated error covariance matrix P^a are invariant in any invertible linear change of coordinates, either in data or state space.

Under determinacy condition rank$\Gamma = n$, it is possible to transform the data vector z into

$$x^b = x + \zeta^b$$
$$y = Hx + \varepsilon$$

$$E(\zeta^b \varepsilon^{\mathrm{T}}) = 0$$

Setting $E(\zeta^b \zeta^{bT}) \equiv P^b$, $E(\varepsilon\varepsilon^T) \equiv R$,

the expressions for x^a and P^a take either one of the two equivalent forms

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

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

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

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

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

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

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

'*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

  $x_0^b = x_0 + \zeta_0^b \qquad\qquad E(\zeta_0^b \zeta_0^{b\mathrm{T}}) = P_0^b$

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

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

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

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

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

Then objective function

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

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

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

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).

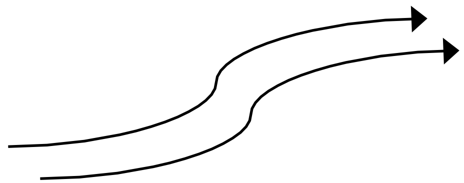
If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

A few basics

- Basic (nonlinear) model

$$x_{k+1} = M_k(x_k)$$

- Perturbation δx_0 at time 0. Resulting perturbation δx_k evolves in time according to



$$\delta x_{k+1} = M_k(x_k + \delta x_k) - M_k(x_k)$$

$$= M_k'(x_k)\ \delta x_k + o(\delta x_0)$$

where $M_k'(x_k)$ is jacobian of M_k at point x_k.

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

is *tangent linear model* along solution x_k.

A few basics (continuation)

Tangent linear model

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

Adjoint model

$$\lambda_k = [M_k'(x_k)]^{\mathrm{T}}\,\lambda_{k+1}$$

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

Adjoint Method (continued 3)

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

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

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

….

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

….

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

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

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

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

Adjoint equation

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

....

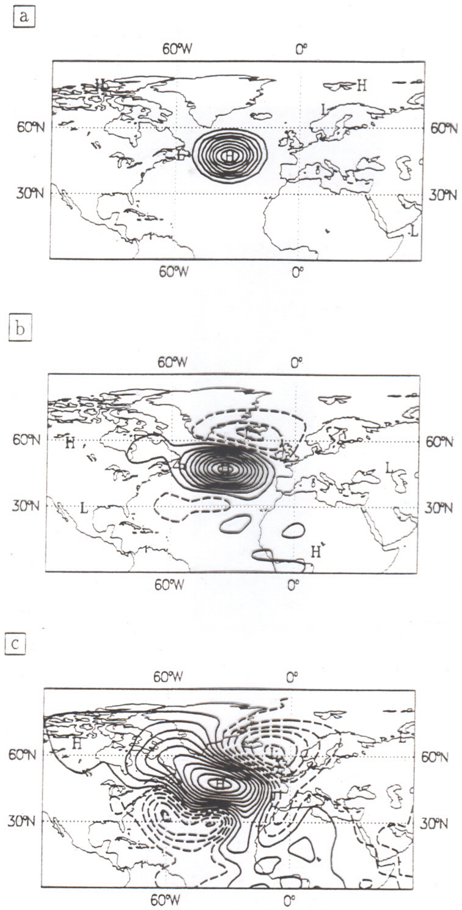
$\lambda_k = M_k{}'^{\mathrm{T}}\lambda_{k+1} + H_k{}'^{\mathrm{T}}\,R_k{}^{-1}\,[H_k(\xi_k) - y_k] \qquad k = K\text{-}1, \ldots, 1$

....

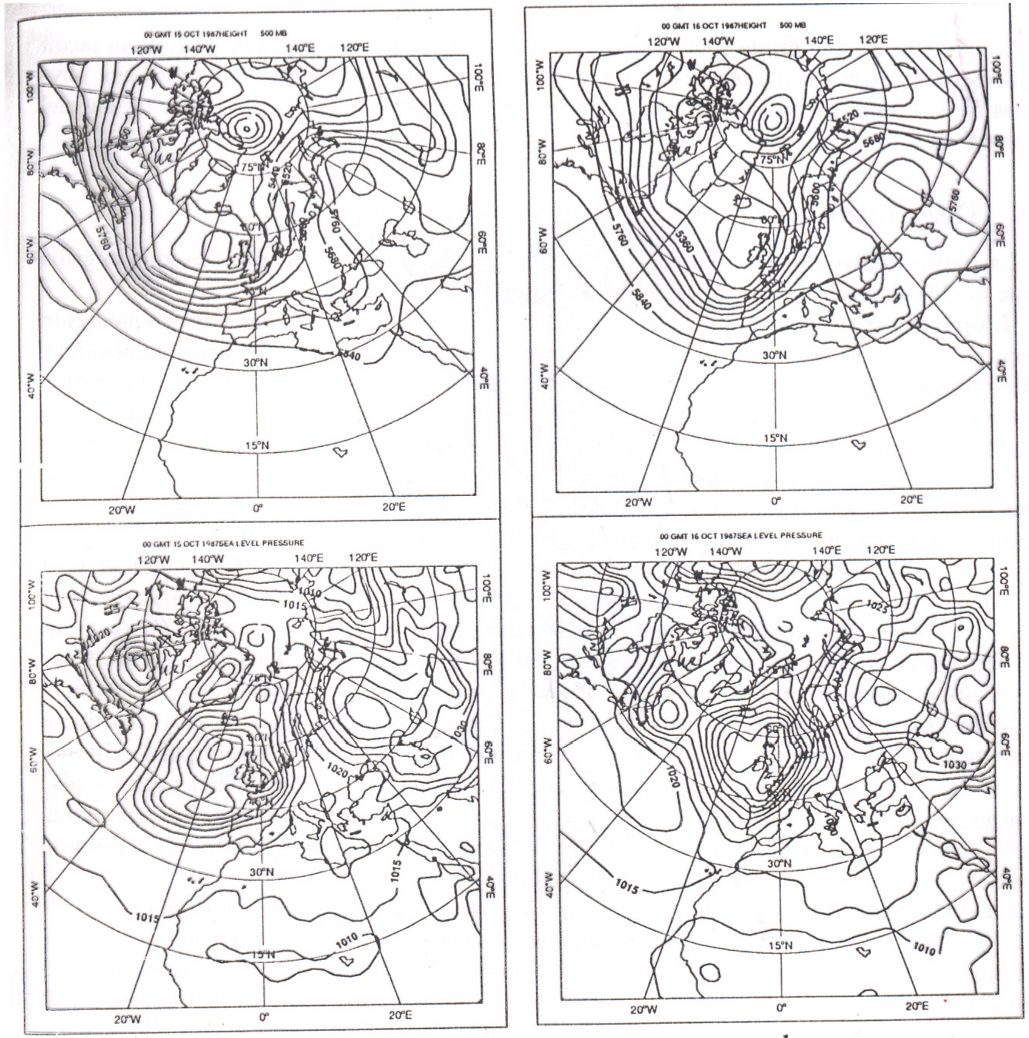
$\lambda_0 = M_0{}'^{\mathrm{T}}\lambda_1 + H_0{}'^{\mathrm{T}}\,R_0{}^{-1}\,[H_0(\xi_0) - y_0] + [P_0{}^b]^{-1}\,(\xi_0 - x_0{}^b)$

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

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



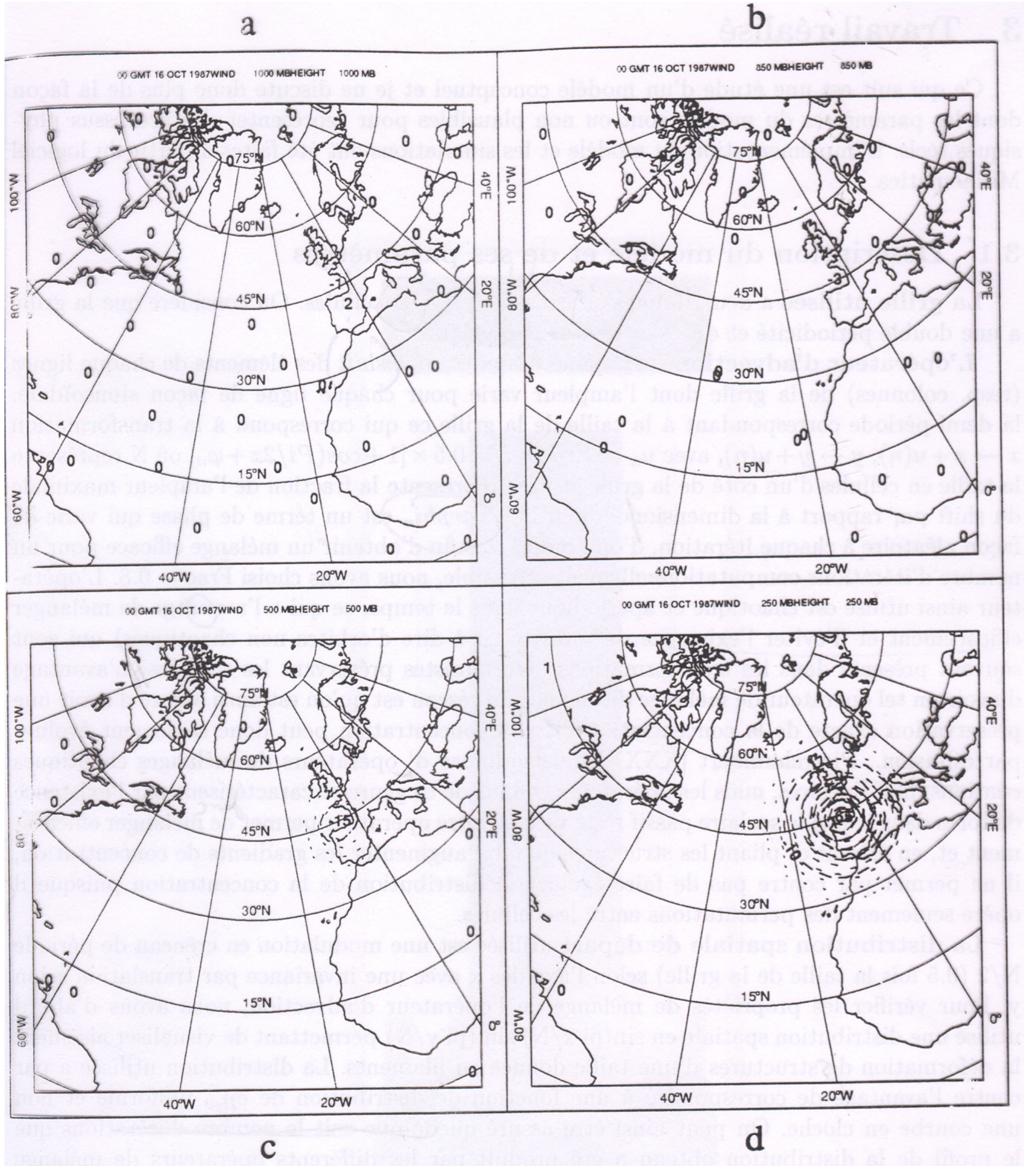
Temporal evolution of the 500-hPa geopotential autocorrelation with respect to point located at 45N, 35W. From top to bottom: initial time, 6- and 24-hour range. Contour interval 0.1. After F. Bouttier.



b d

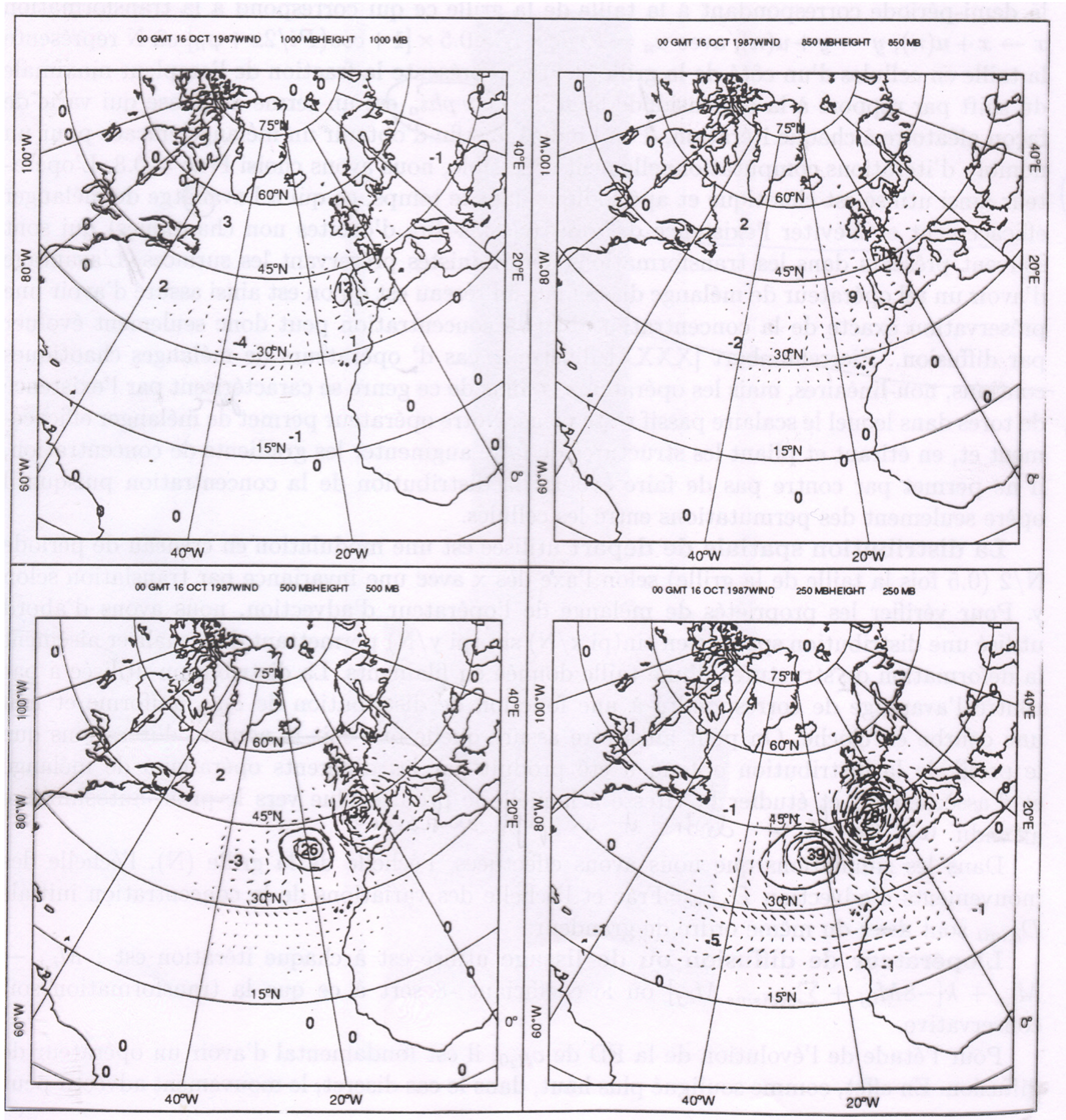
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



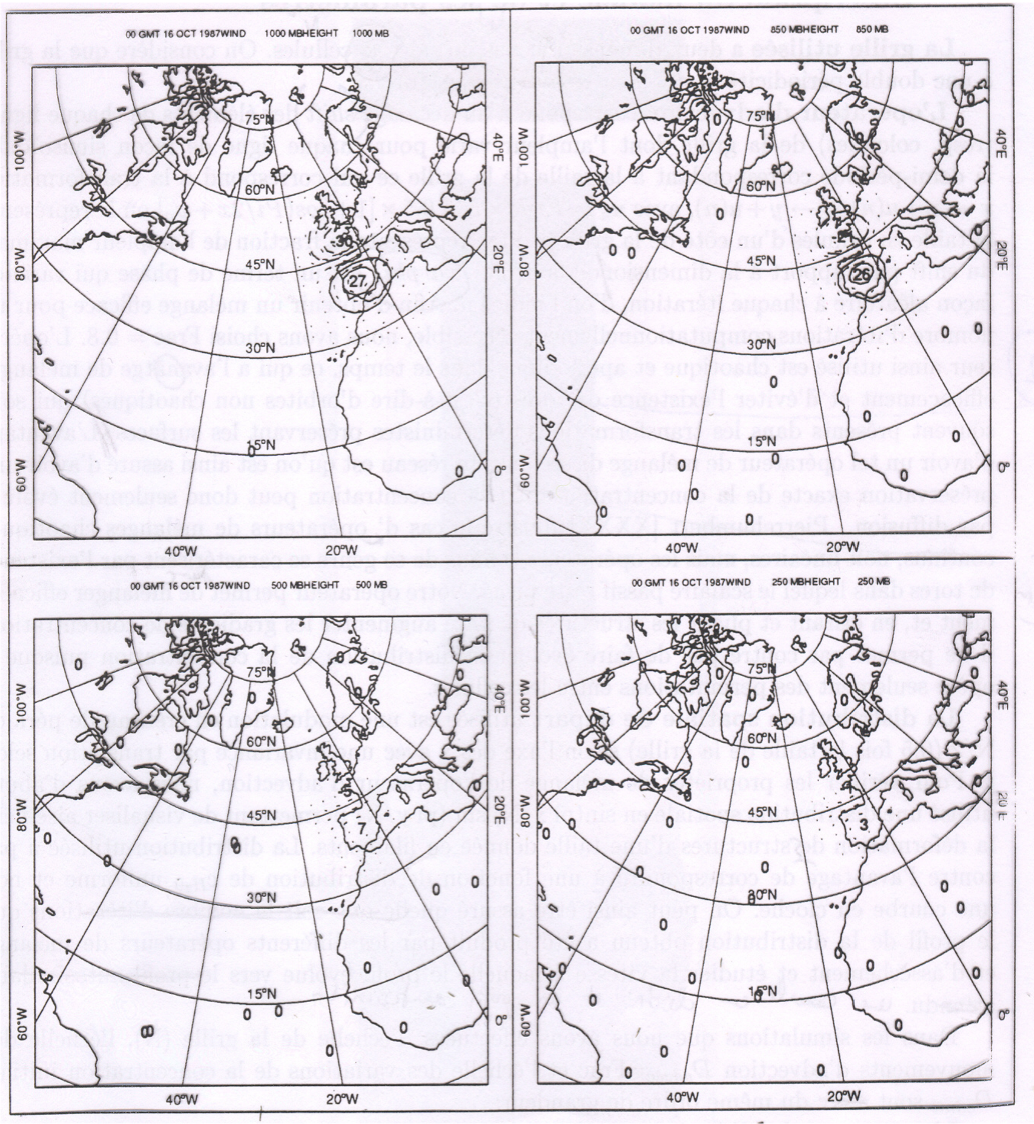
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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



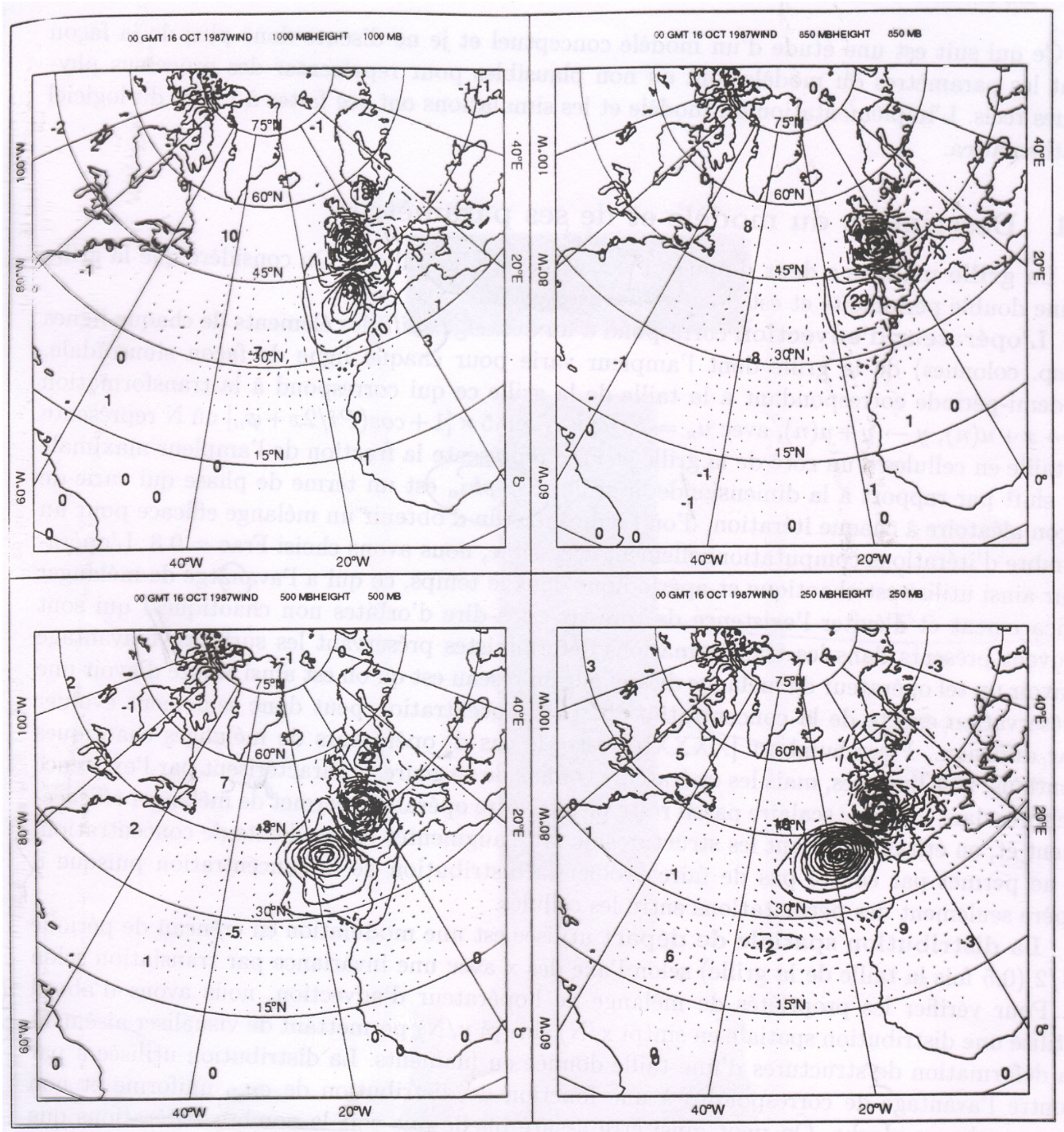
Same as before, but at the end of a 24-hr 4D-Var

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



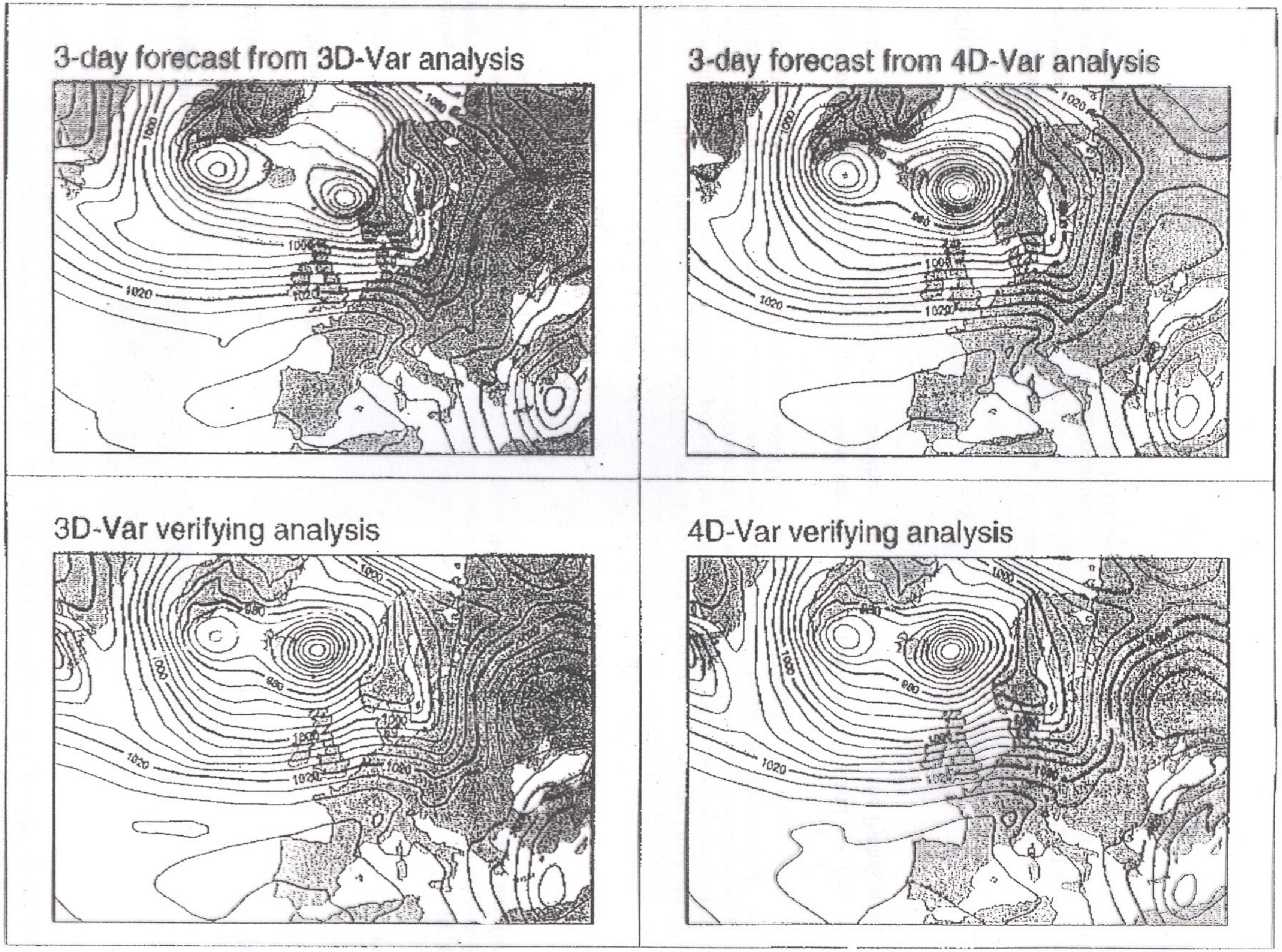
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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)

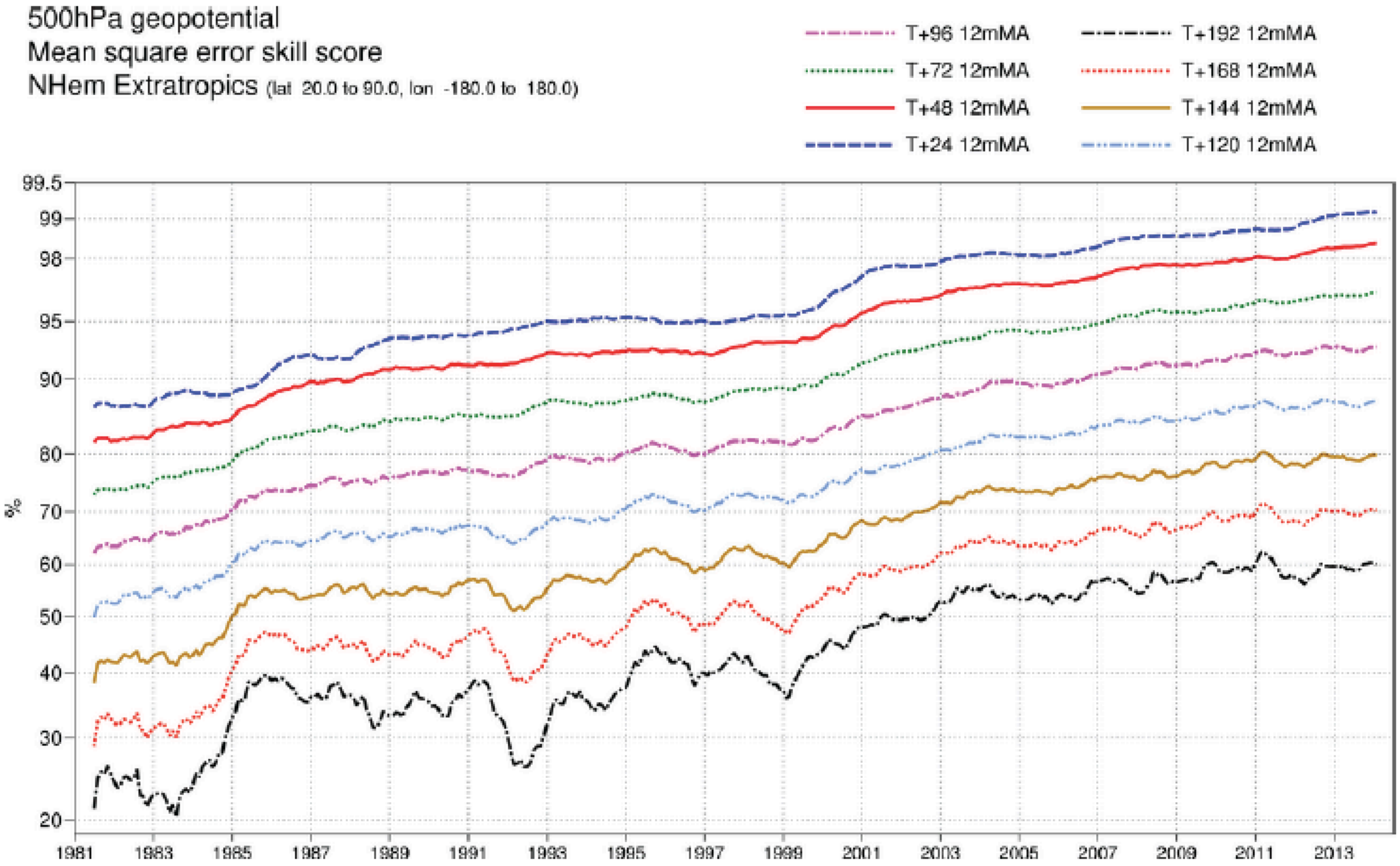
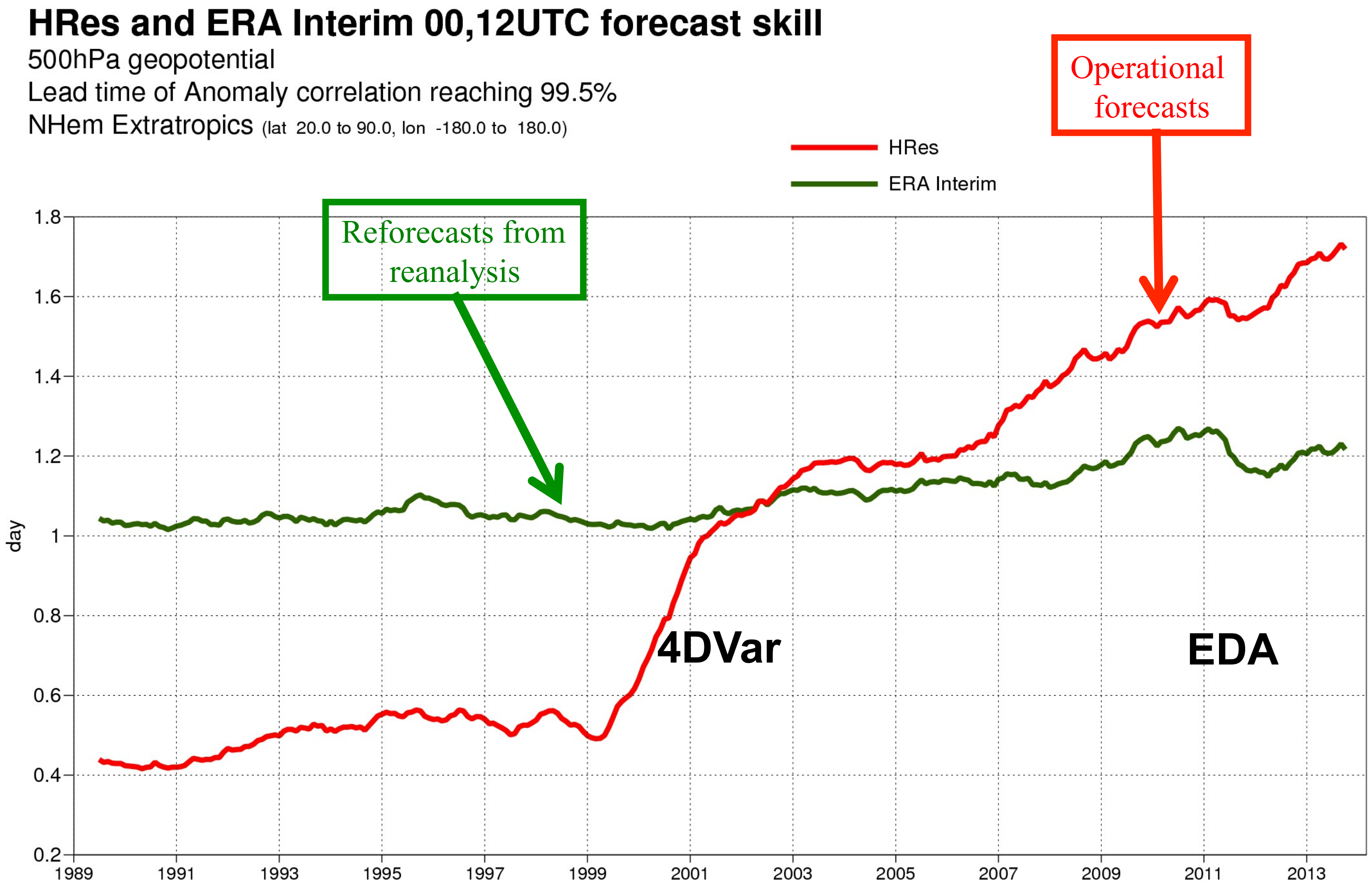


Figure 3: 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

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 3)

If data errors are correlated in time, it is not possible to discard observations as they are used. In particular, if model error is correlated in time, all observations are liable to be reweighted as assimilation proceeds.

Variational assimilation can take time-correlated errors into account.

Example of time-correlated observation errors. Global covariance matrix

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

Objective function

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

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

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

Similar approach for time-correlated model error.

Time-correlated Errors (continuation 4)

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

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

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.

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

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

- Tangent linear model

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

where M_k' is jacobian of M_k at point ξ_k.

- Adjoint model

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

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

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

where L_k is an appropriate simplification of jacobian M_k'.

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

Incremental Method (continuation 3)

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

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

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

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

With the choices made here, $\mathcal{J}_{\mathrm{I}}(\delta\xi_0)$ 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 L_k and N_k, and for defining an appropriate trade-off between practical implementability and physical usefulness and accuracy. It is the incremental method which, together with the adjoint method, makes variational assimilation possible.

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

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

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

How to take model error into account in variational assimilation ?

Weak constraint variational assimilation

Allows for errors in the assimilating model

- Data

  - Background estimate at time 0

  $$x_0^b = x_0 + \zeta_0^b \qquad\qquad E(\zeta_0^b \zeta_0^{bT}) = P_0^b$$

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

  $$y_k = H_k x_k + \varepsilon_k \qquad\qquad E(\varepsilon_k \varepsilon_{k'}^{T}) = R_k \delta_{kk'}$$

  - Model

  $$x_{k+1} = M_k x_k + \eta_k \qquad\qquad E(\eta_k \eta_{k'}^{T}) = Q_k \delta_{kk'} \qquad\qquad k = 0, \ldots, K\text{-}1$$

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

Then objective function

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

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

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

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

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

Can include nonlinear M_k and/or H_k.

Implemented operationally at ECMWF for the assimilation in the stratosphere.

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

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

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

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

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

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

Maximisation is performed in (dual of) observation space.

Dual Algorithm for Variational Assimilation (continuation 2)

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

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

or, equivalently, but more conveniently, as

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

where, as before

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

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

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

It is associated with error covariance matrix

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

Dual Algorithm for Variational Assimilation (continuation 3)

Define global observation vector as

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

and global innovation vector as

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

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

Dual Algorithm for Variational Assimilation (continuation 4)

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

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

is defined by the sequence of operations

$$u_0 = H_0 \xi_0$$

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

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

The observation error covariance matrix is equal to

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

Dual Algorithm for Variational Assimilation (continuation 5)

Maximization of dual objective function

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

requires explicit repeated computations of its gradient

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

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

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

Set $\chi_K = 0$

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

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

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

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

Dual Algorithm for Variational Assimilation (continuation 6)

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

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

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

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

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

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

Dual Algorithm for Variational Assimilation (continuation 7)

Temporal correlations can be introduced.

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

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

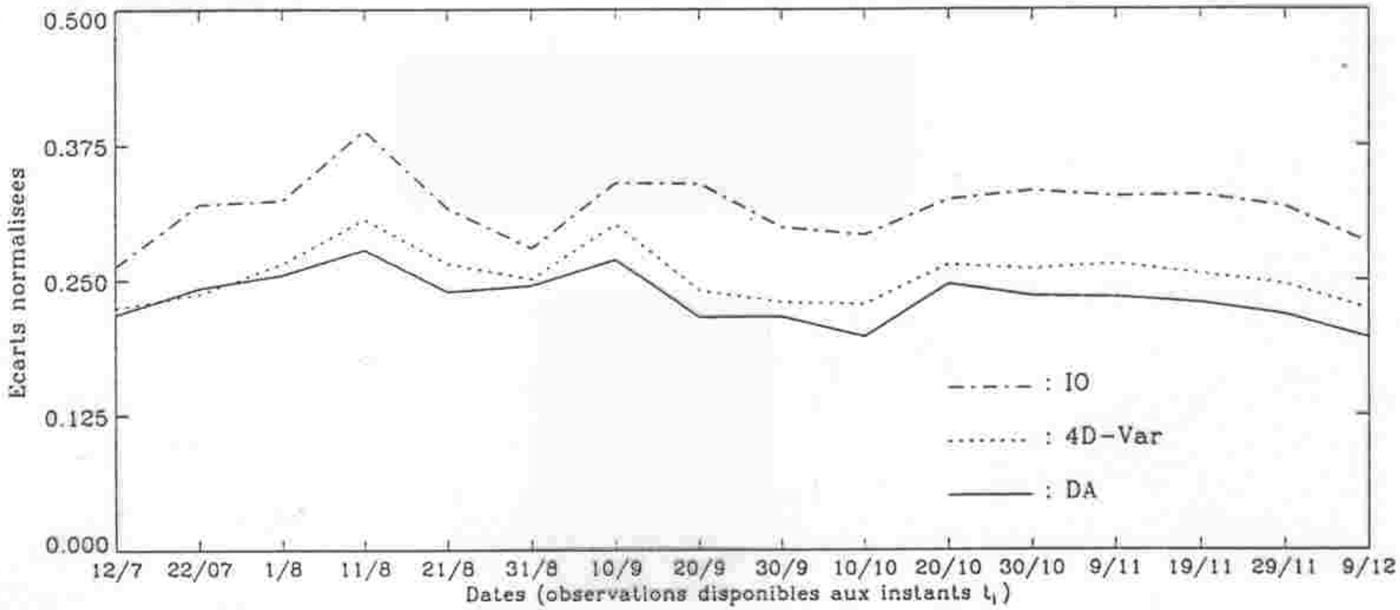


FIG. 9.11 – *Ecarts normalisés prévision/observations sur l'ensemble de la période étudiée*

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

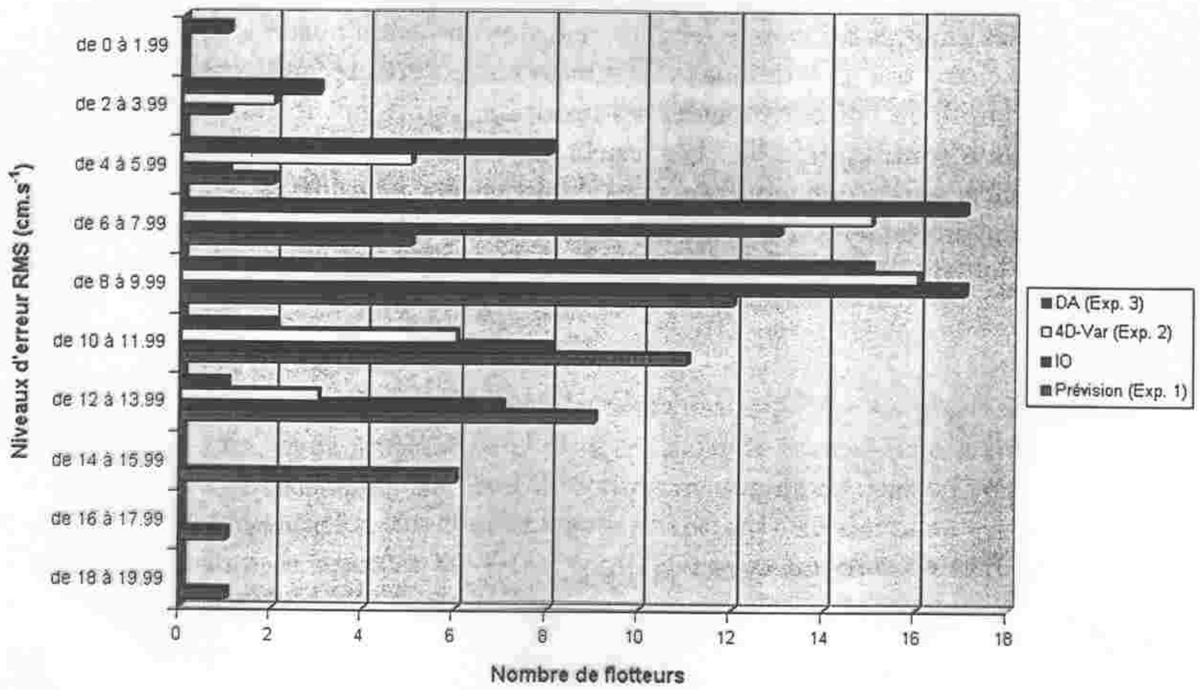


FIG. 9.15 – *Description des écarts flotteurs/modèle en terme de vitesse (à 150 m de profondeur) pour les différents algorithmes d'assimilation*

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

Dual Algorithm for Variational Assimilation (continuation)

Requires

- Explicit background (not much of a problem)

- Exact linearity (much more of a problem). Definition of iterative nonlinear procedures is being studied (Auroux, …)

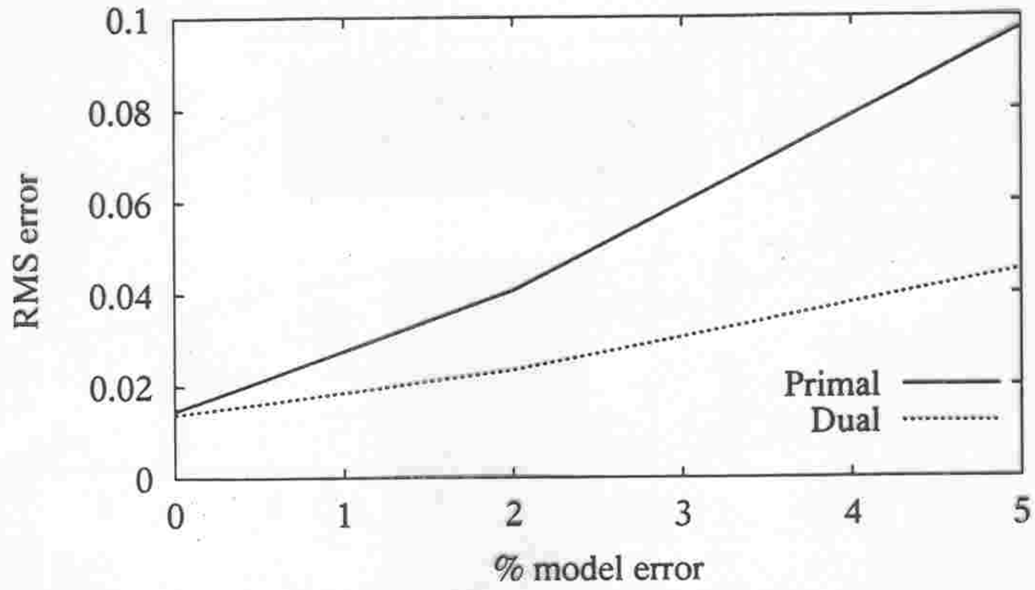


FIG. 6.13 – Normes RMS des erreurs d'assimilation obtenues pour les deux méthodes en fonction de l'erreur introduite dans le modèle au cours de la période d'assimilation.

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

Dual Algorithm for Variational Assimilation is now used, in the weak-constraint form, at *Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique* (*CERFACS*) in Toulouse (A. Weaver, S. Gürol) for assimilation of oceanographical observations.

Conclusion on Sequential Assimilation

Pros

‘Natural’, and well adapted to many practical situations

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

Cons

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

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

Conclusion on Variational Assimilation

Pros

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

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

Does not require explicit computation of temporal evolution of estimation error

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

Cons

Does not readily provide estimate of estimation error

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

- Dual approach seems most promising. But little used.

- Can be implemented in ensemble form (see course 7).

History of Numerical Weather Prediction

Wilhelm Bjerknes

Das Problem der Wettervorhersage, betrachtet von Standpunkt der Mechanik und Physik, 1904, *Meteorologische Zeitschrift*

École de Météorologie de Bergen

From course 2

Physical laws governing the flow

- Conservation of mass

  $D\rho/Dt + \rho\,\mathrm{div}\underline{U} = 0$

- Conservation of energy

  $De/Dt - (p/\rho^2)\, D\rho/Dt = Q$

- Conservation of momentum

  $D\underline{U}/Dt + (1/\rho)\,\underline{\mathrm{grad}p} - \underline{g} + 2\,\underline{\Omega} \wedge \underline{U} = \underline{F}$

- Equation of state

  $f(p, \rho, e) = 0$ $\quad\quad\quad\quad$ $(p/\rho = rT, e = C_v T)$

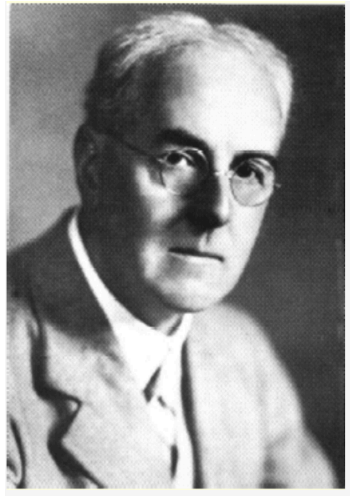
- Conservation of mass of secondary components (water in the atmosphere, salt in the ocean, chemical species, …)

  $Dq/Dt + q\,\mathrm{div}\underline{U} = S$

These physical laws must be expressed in practice in discretized (and necessarily imperfect) form, both in space and time

History of Numerical Weather Prediction (continuation)

Lewis Fry Richardson



Weather Prediction by Numerical Process, 1922

Cambridge University Press *

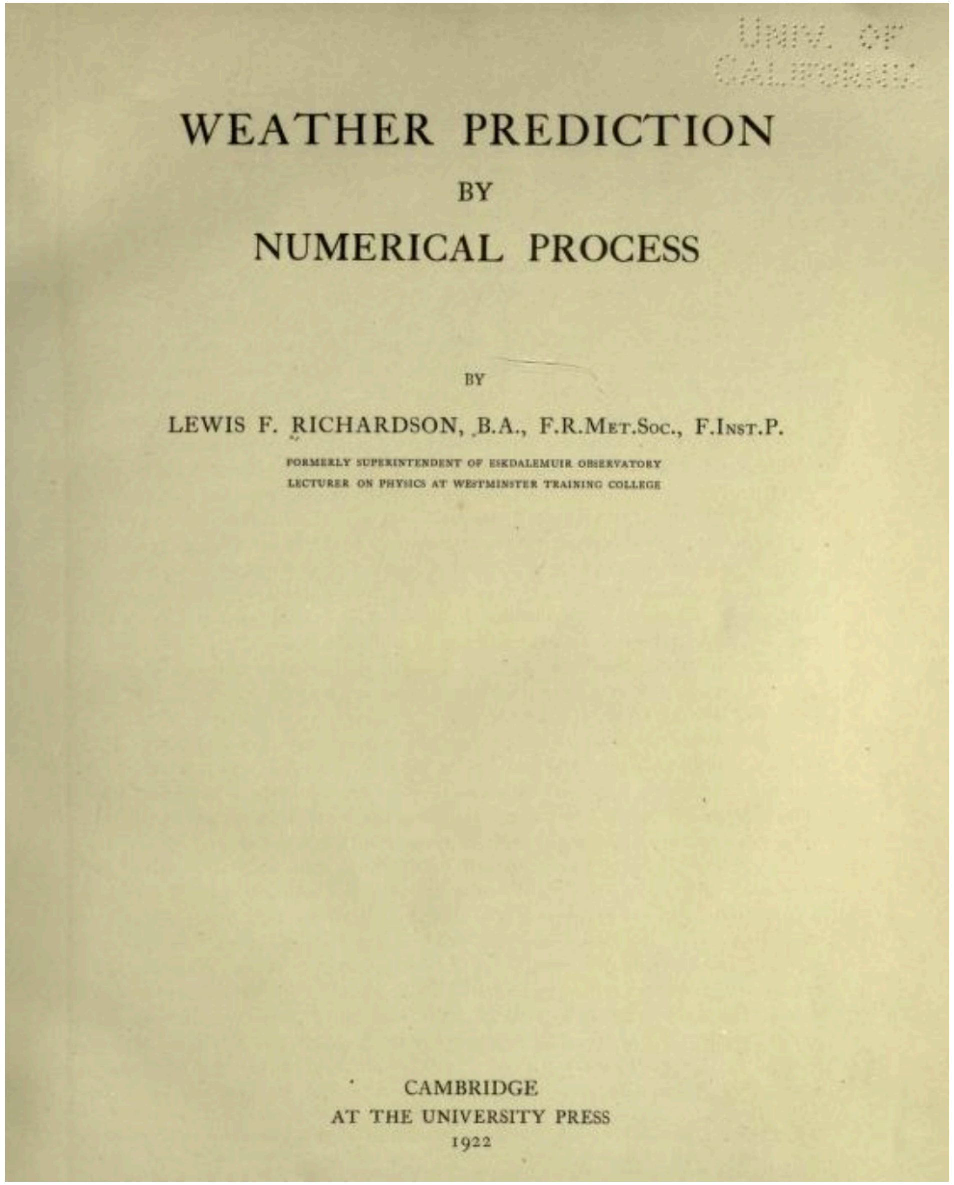
Forecast Factory

Richardson number, fractals, pacifism

* Accessible at URL

https://energy4climate.pages.in2p3.fr/public/education/ensemble_data_assimilation_tutorial/notebooks/T1%20-%20Introduction%20to%20Ensemble%20Data%20Assimilation%20for%20Numerical%20Weather%20Prediction.html

https://
energy4climate.pages.in2p3.fr/
public/education/
ensemble_data_assimilation_tut
orial/notebooks/T1%20-
%20Introduction%20to
%20Ensemble%20Data
%20Assimilation%20for
%20Numerical%20Weather
%20Prediction.html



WEATHER PREDICTION
BY
NUMERICAL PROCESS

BY

LEWIS F. RICHARDSON, B.A., F.R.MET.SOC., F.INST.P.
FORMERLY SUPERINTENDENT OF ESKDALEMUIR OBSERVATORY
LECTURER ON PHYSICS AT WESTMINSTER TRAINING COLLEGE

CAMBRIDGE
AT THE UNIVERSITY PRESS
1922

History of Numerical Weather Prediction (continuation 2)

John von Neumann

Institute for Advanced Studies, Princeton, 1946-1950

First electronic computers (ENIAC)

(J. Charney, N. A. Phillips, R. Fjørtoft, C. G. Rossby,

J. Smagorinsky, …)

Jule Gregory Charney en 1978.

Charney developed barotropic model

First operational numerical forecast around 1955 in Sweden

(C. G. Rossby)

History of Numerical Weather Prediction (continuation 3)

Numerical prediction has gradually been implemented in more and more meteorological services around the world.

Extension to simulation of oceanic circulation and climate (early 1970's, S. Manabe and K. Bryan, GFDL).

Climatic simulations (S. Manabe, R. Wrtherald)

European Centre for Medium-Range Weather Forecasts (ECMWF, 1975)

Ensemble prediction

History of Numerical Weather Prediction (continuation 4)

A large variety of models covering different spatial and temporal scales and phenomena (small-scale convection, monthly and seasonal prediction, atmospheric chemistry, …) have been developed over the years and are used for research and operational applications.

Intergovernmental Panel on Climate Change (IPCC, 1988)

Publishes reports that describe the state of climate science and presents ‘projections’ largely based on numerical simulations

First report in 1990

Fifth report in 2014

Sixth report to be published in 2021-2022

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