École Doctorale des Sciences de l'Environnement d’Île-de-France
Année Universitaire 2024-2025

Modélisation Numérique de l’Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand
Cours 5

12 Mai 2025

From course 4

Kalman Filter

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.

Rudolf Kálmán (1930-2016)

- Lissage de Kalman.

- Assimilation variationnelle. Principe

- Méthode adjointe. Principe.

- Assimilation variationnelle. Résultats

- La Méthode incrémentale

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

Kalman smoother

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

Various possibilities

- Define new state vector $X^{\mathrm{T}} \equiv (x_0^{\mathrm{T}}, \ldots, x_K^{\mathrm{T}})$

  and use general *BLUE* formula from a background X^b and associated covariance matrix Π^b.

  Model equations, which bring information on the x_i 's, must be included in the observation vector and the associated observation operator.

  Can take into account temporal correlations

- Update sequentially vector $(x_0^{\mathrm{T}}, \ldots, x_k^{\mathrm{T}})^{\mathrm{T}}$ 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

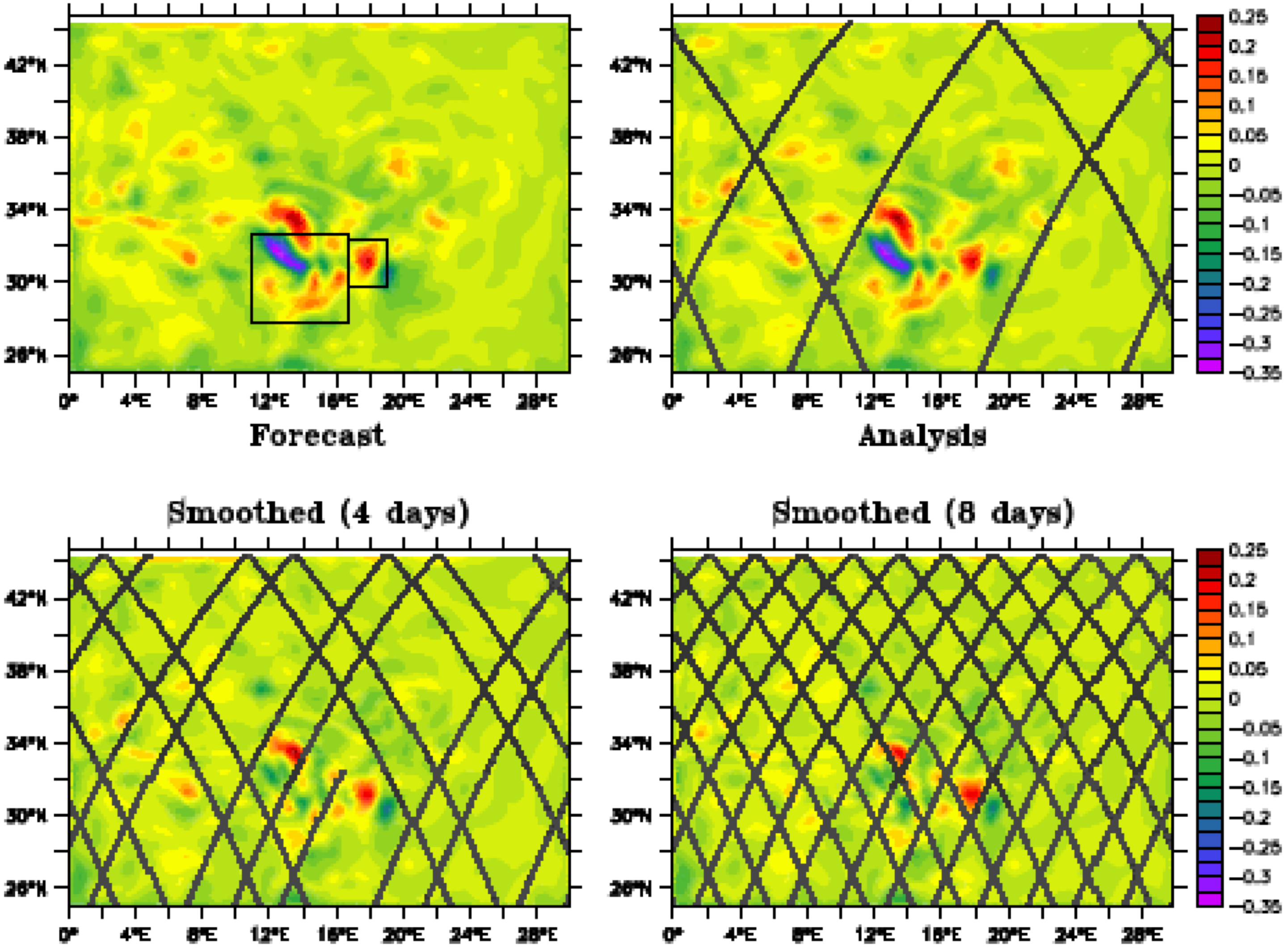


FIGURE 3.6 – Evolution du champ d’erreur en SSH du jour 38, au cours des étapes d’analyse successives. En haut à gauche : prévision du filtre ; en haut à droite : analyse du filtre. Les observations utilisées pour cette analyse sont distribuées le long des traces grises. En bas à gauche : analyse du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

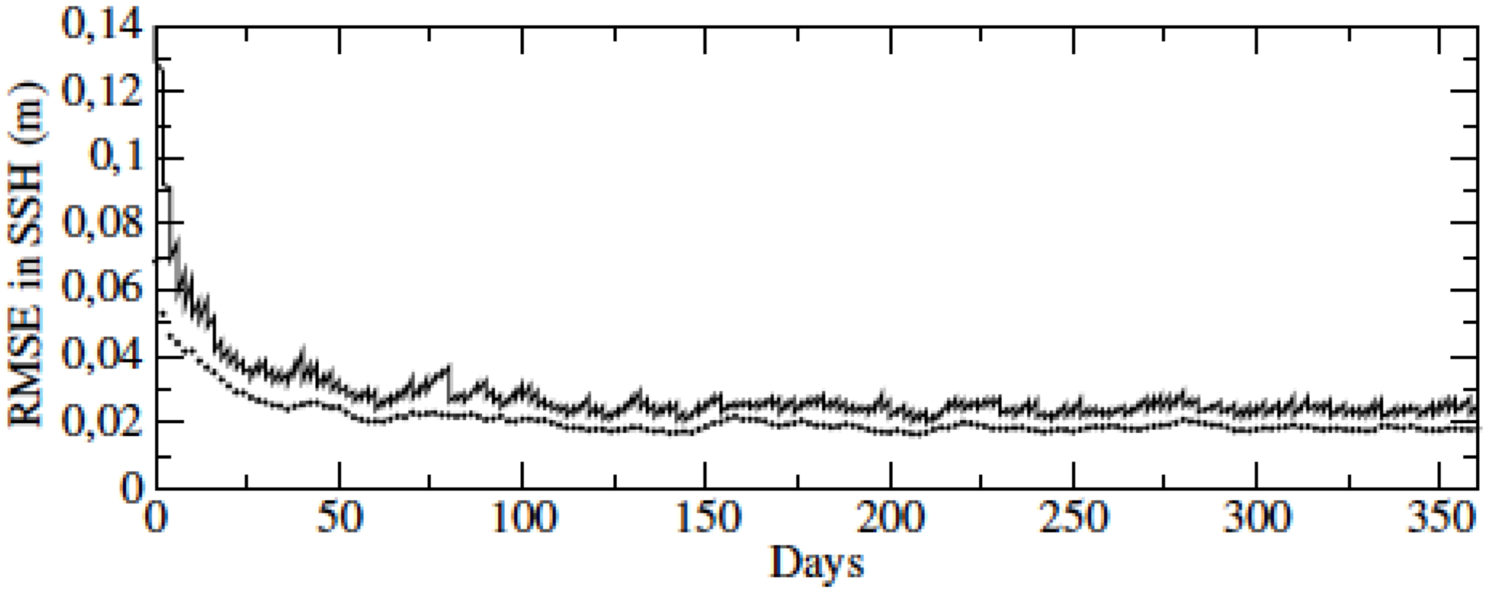


FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse) ; Points : lisseur à retard fixe de 8 jours.

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)*

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

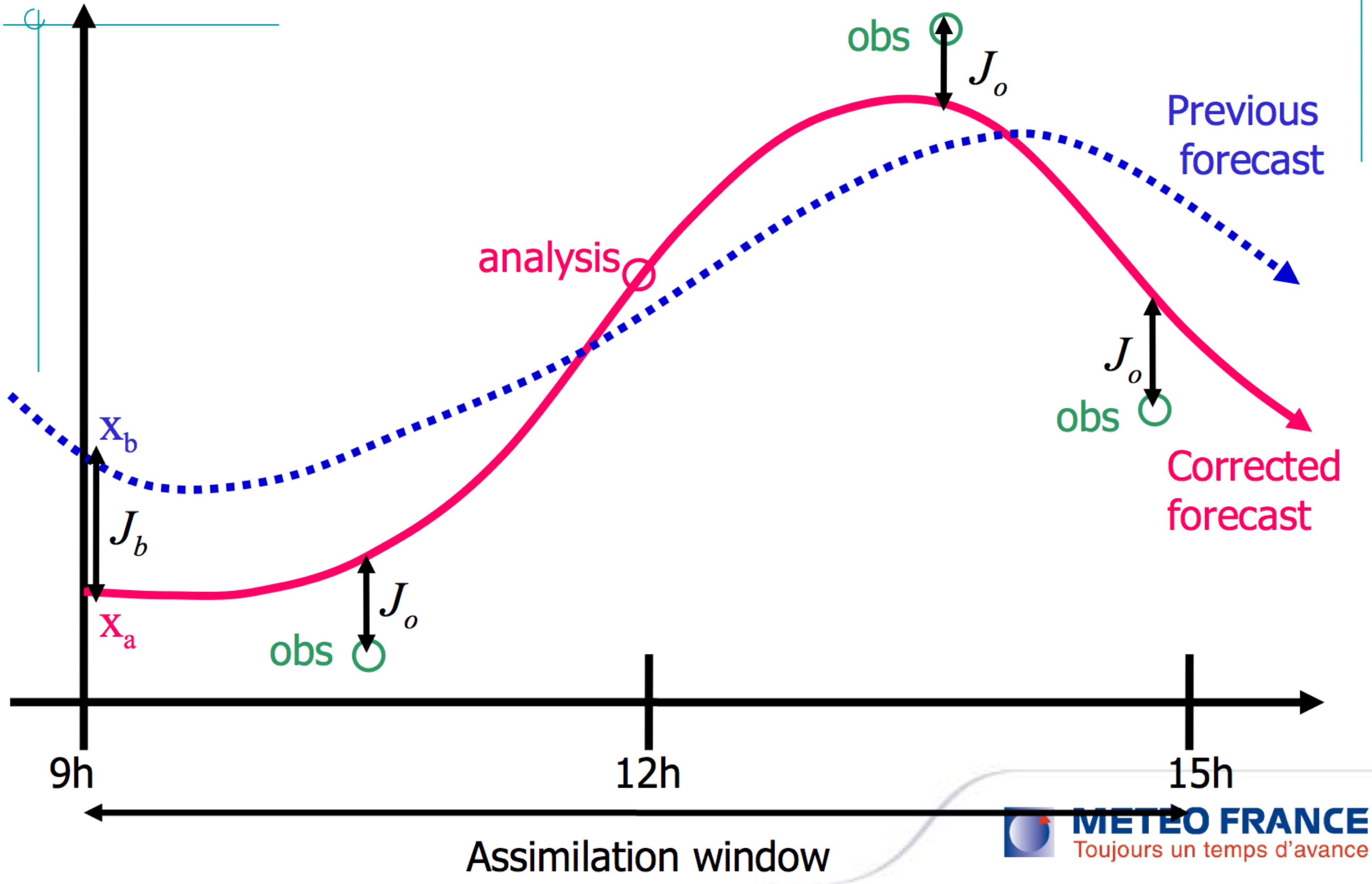
$\xi_0 \in S \rightarrow$

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

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

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

Principle of 4D-VAR assimilation



METEO FRANCE
Toujours un temps d'avance

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

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

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 $\boldsymbol{u} = \xi_0$ ($\boldsymbol{u}$ 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).

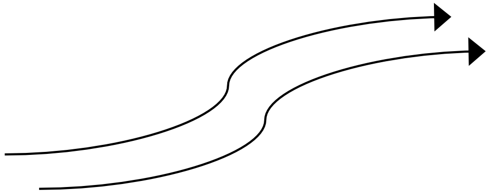
If everything is kept in memory, total operation count of adjoint computation is at most 4 times operation count of direct computation (in practice about 2).

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



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

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

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

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

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

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

Adjoint equation

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

….

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

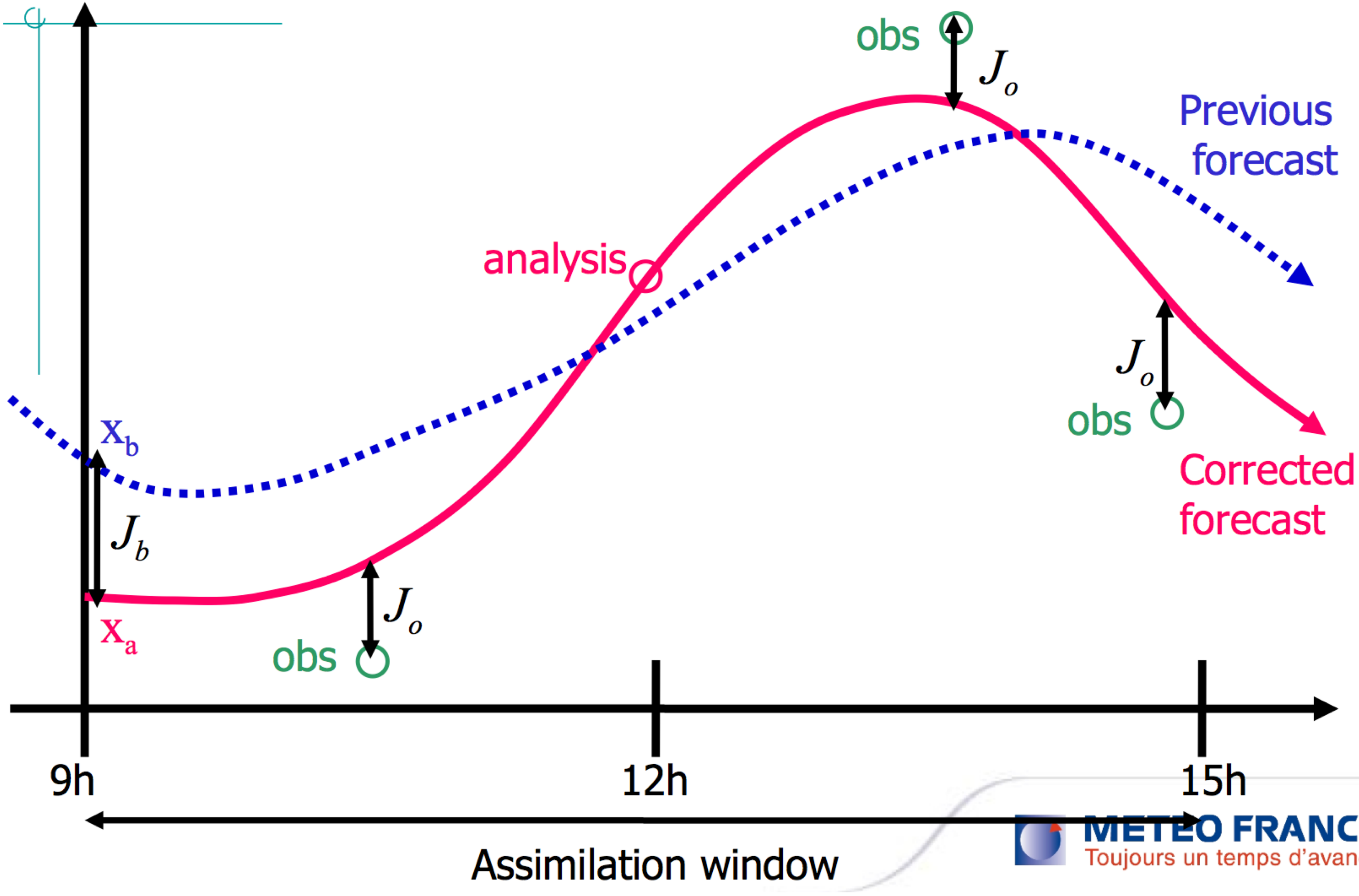
….

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

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

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 ?

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

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

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

Adjoint equation

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

....

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

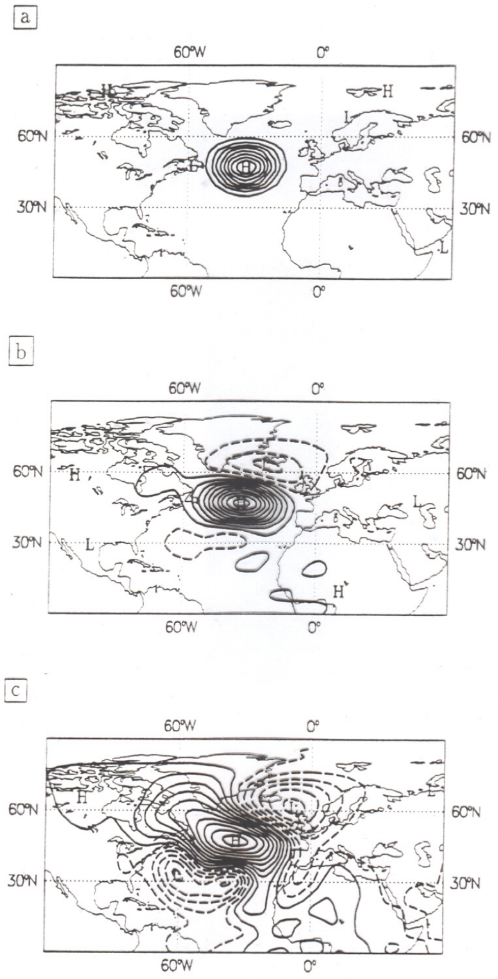
....

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

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

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.

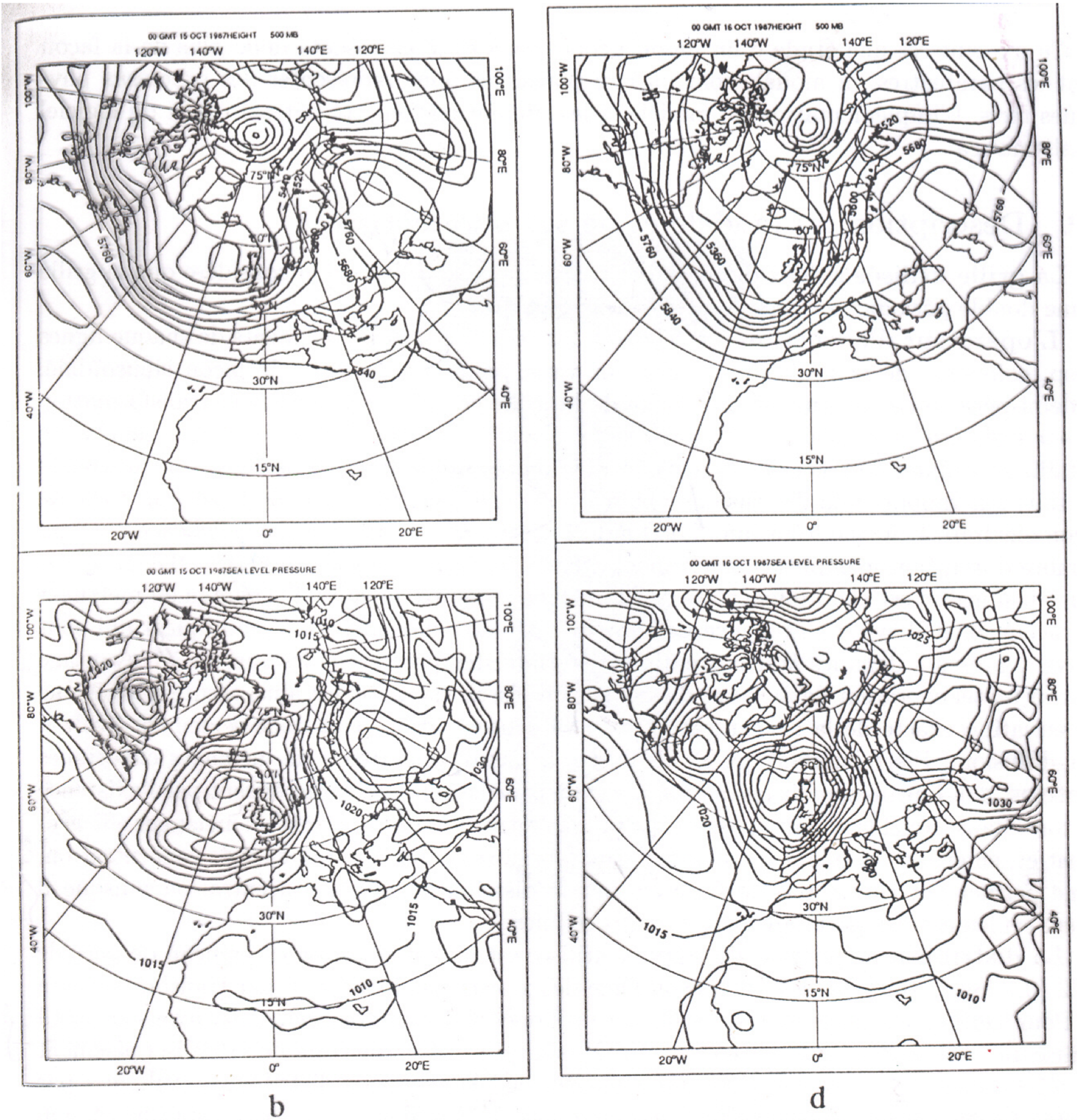
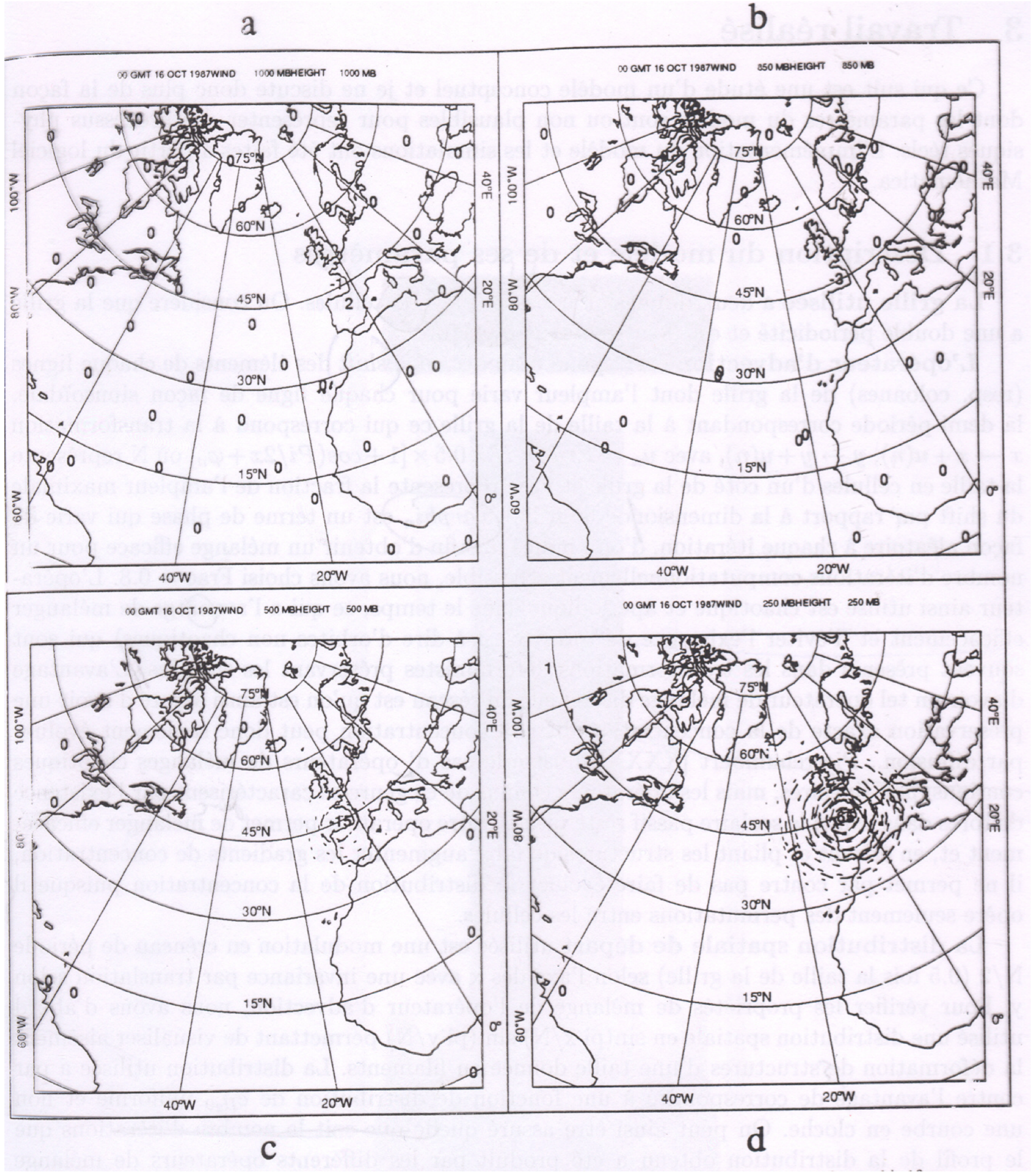

00 GMT 15 OCT 1987HEIGHT 500 MB
00 GMT 16 OCT 1987HEIGHT 500 MB
00 GMT 15 OCT 1987SEA LEVEL PRESSURE
00 GMT 16 OCT 1987SEA LEVEL PRESSURE
120°W
140°W
140°E
120°E
100°W
80°W
60°W
40°W
100°E
80°E
60°E
40°E
75°N
60°N
45°N
30°N
15°N
20°W
0°
20°E
5760
5680
5520
5840
5360
5600
5540
1010
1015
1020
1023
1030
b
d

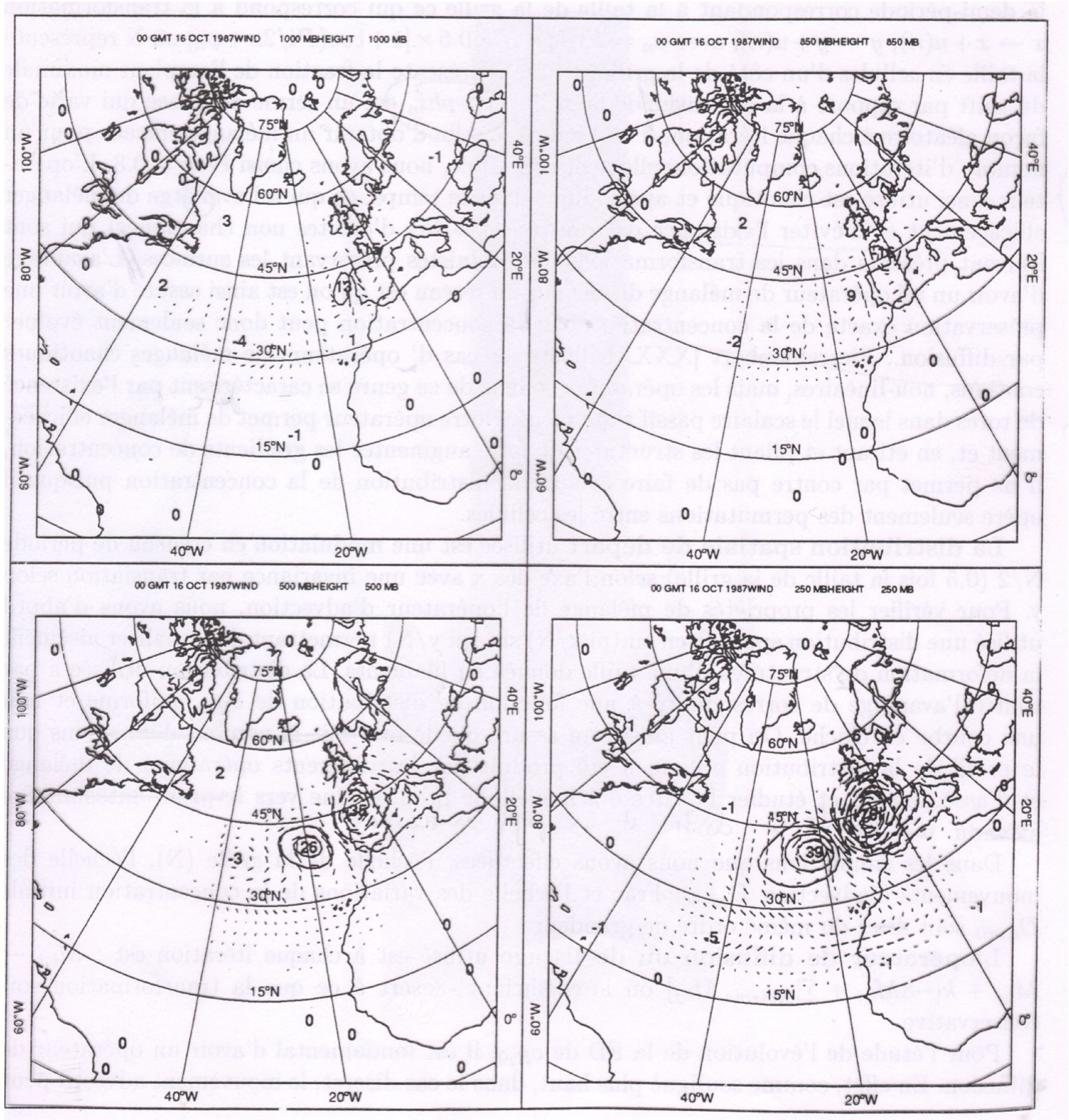
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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


a
b
c
d
00 GMT 16 OCT 1987WIND 1000 MBHEIGHT 1000 MB
00 GMT 16 OCT 1987WIND 850 MBHEIGHT 850 MB
00 GMT 16 OCT 1987WIND 500 MBHEIGHT 500 MB
00 GMT 16 OCT 1987WIND 250 MBHEIGHT 250 MB
75°N
60°N
45°N
30°N
15°N
100°W
80°W
60°W
40°W
20°W
0°
20°E
40°E

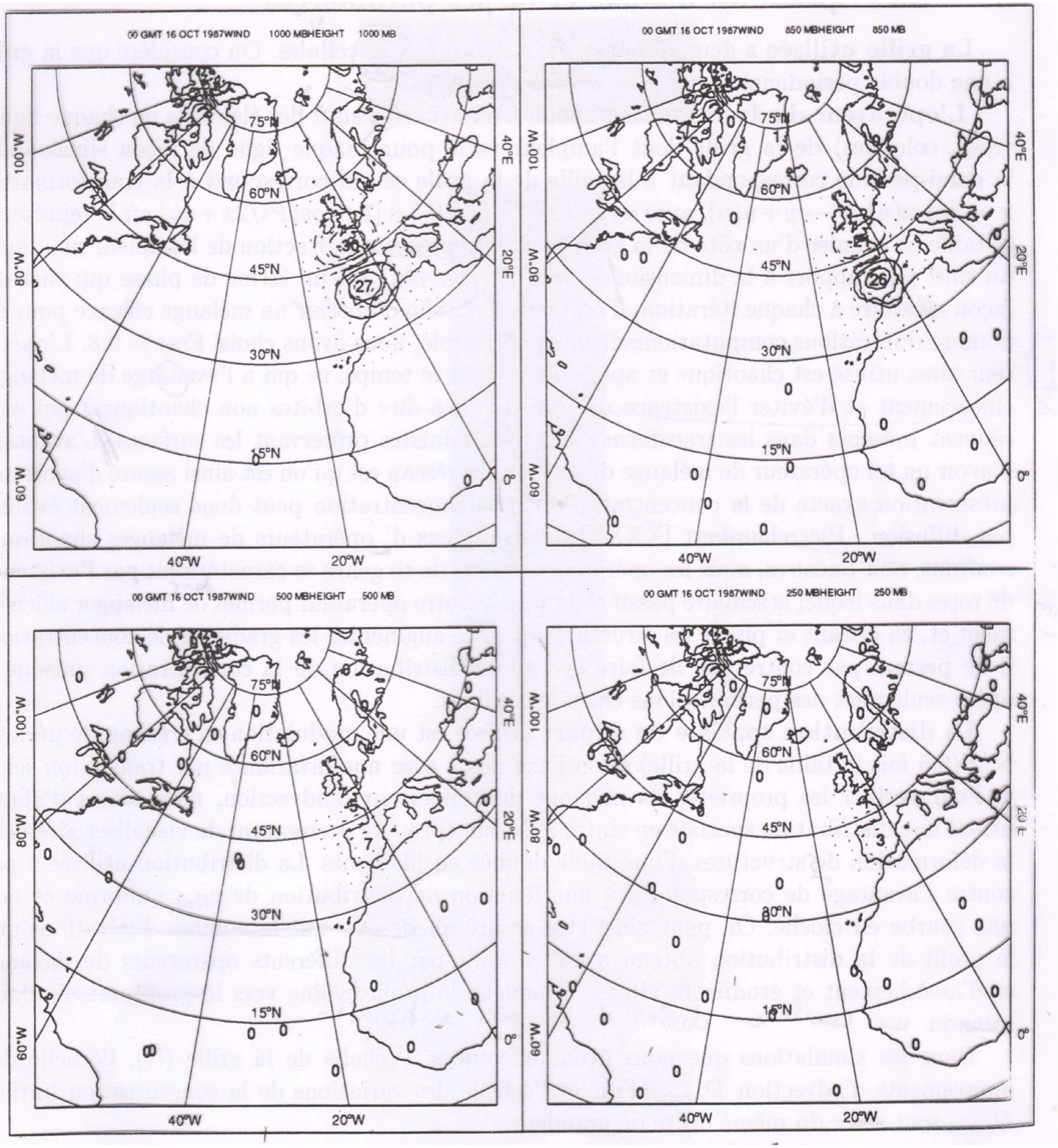
Analysis increments in a 3D-Var corresponding to a height observation at the 250-hPa pressure level (no temporal evolution of background error covariance matrix)

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


00 GMT 16 OCT 1987WIND 1000 MBHEIGHT 1000 MB
00 GMT 16 OCT 1987WIND 850 MBHEIGHT 850 MB
00 GMT 16 OCT 1987WIND 500 MBHEIGHT 500 MB
00 GMT 16 OCT 1987WIND 250 MBHEIGHT 250 MB
75°N
60°N
45°N
30°N
15°N
100°W
80°W
60°W
40°W
20°W
0°
20°E
40°E
26
39

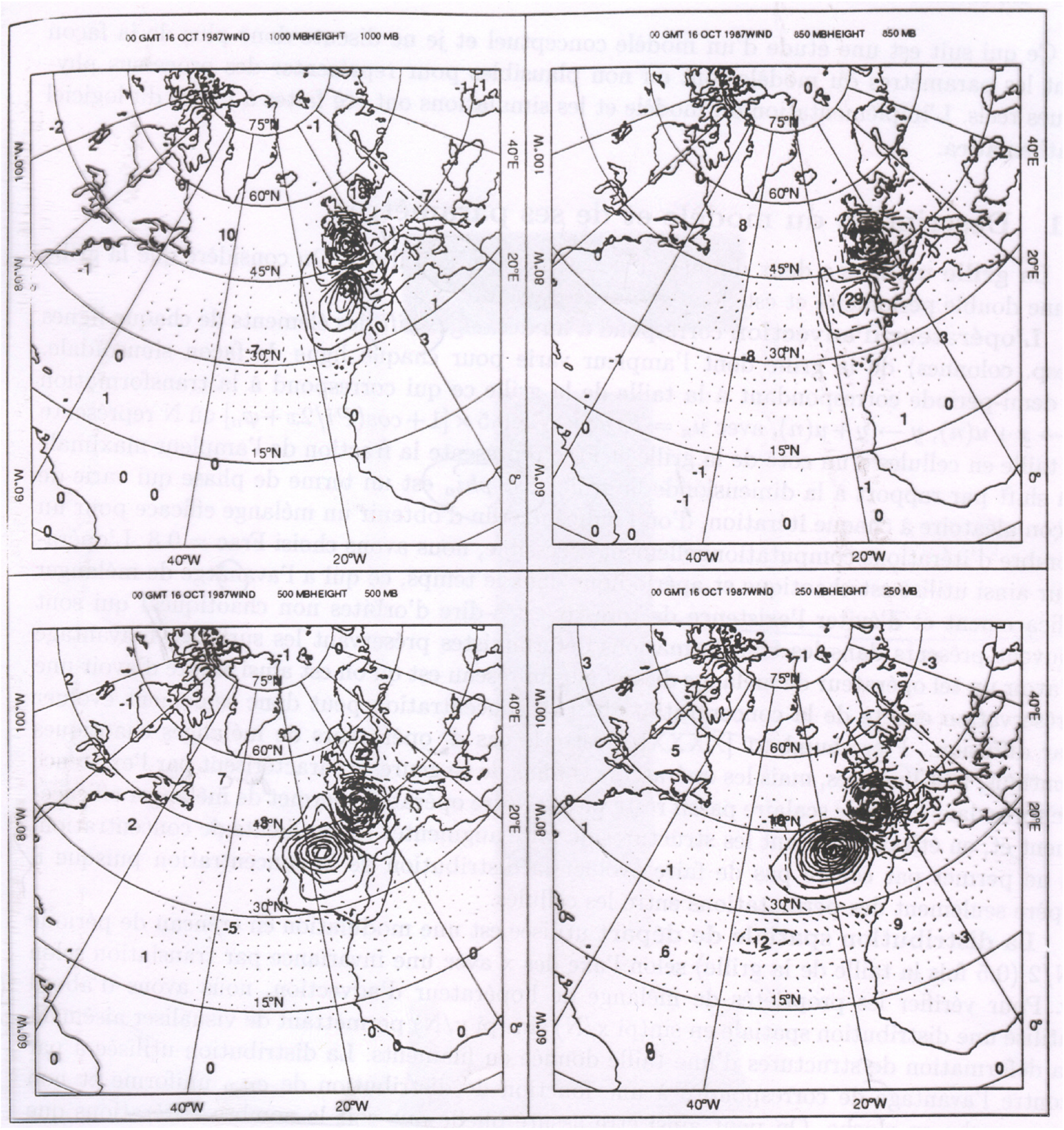
Same as before, but at the end of a 24-hr 4D-Var

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


00 GMT 16 OCT 1987WIND 1000 MBHEIGHT 1000 MB
00 GMT 16 OCT 1987WIND 850 MBHEIGHT 850 MB
00 GMT 16 OCT 1987WIND 500 MBHEIGHT 500 MB
00 GMT 16 OCT 1987WIND 250 MBHEIGHT 250 MB
75°N
60°N
45°N
30°N
15°N
100°W
80°W
60°W
40°W
20°W
0°
20°E
40°E
27
26
3
0

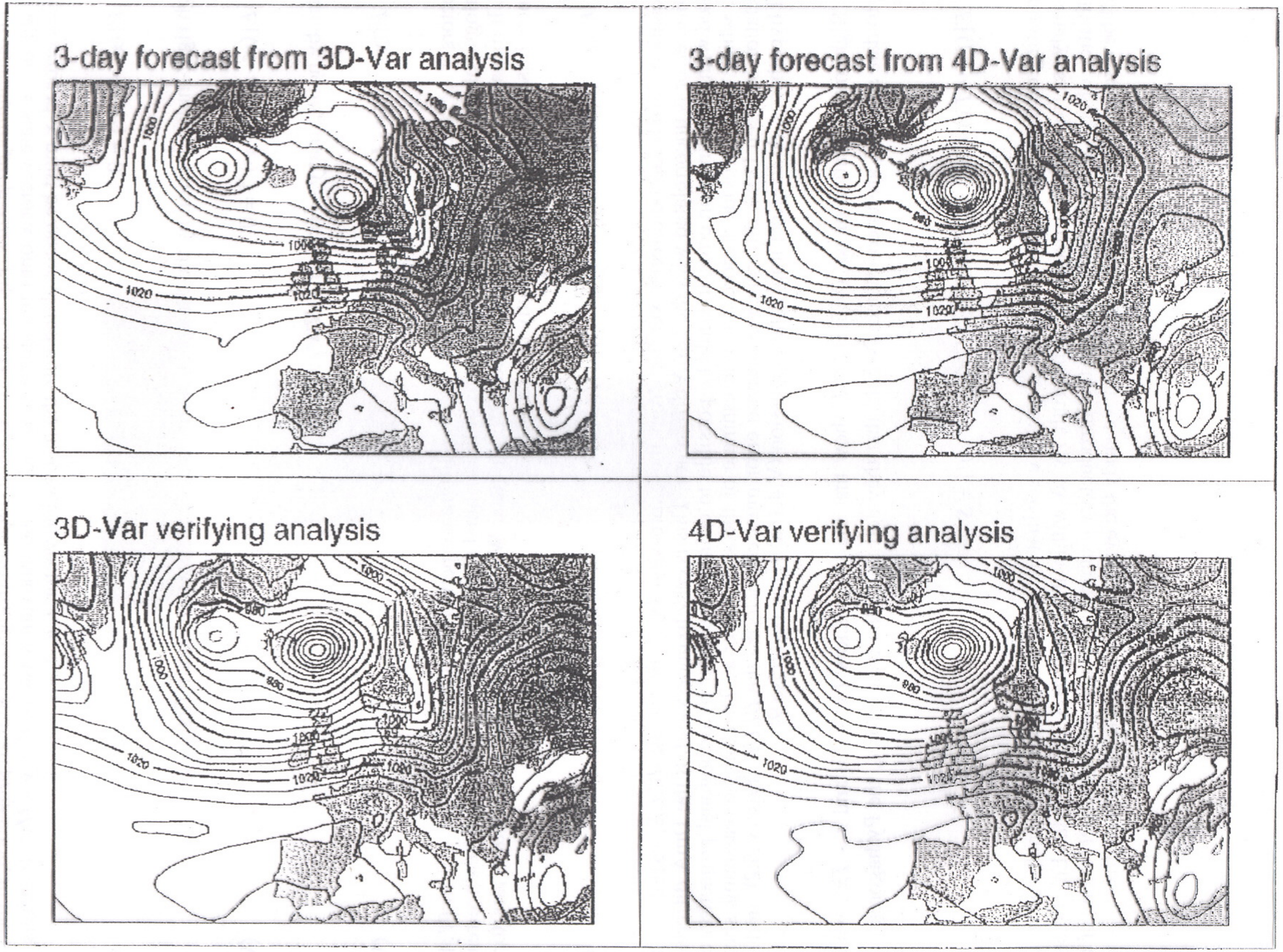
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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


00 GMT 16 OCT 1987WIND 1000 MBHEIGHT 1000 MB
00 GMT 16 OCT 1987WIND 850 MBHEIGHT 850 MB
00 GMT 16 OCT 1987WIND 500 MBHEIGHT 500 MB
00 GMT 16 OCT 1987WIND 250 MBHEIGHT 250 MB
75°N
60°N
45°N
30°N
15°N
100°W
80°W
60°W
40°W
20°W
40°E
20°E
0°

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

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



3-day forecast from 3D-Var analysis

3-day forecast from 4D-Var analysis

3D-Var verifying analysis

4D-Var verifying analysis

ECMWF, Results on one FASTEX case (1997)

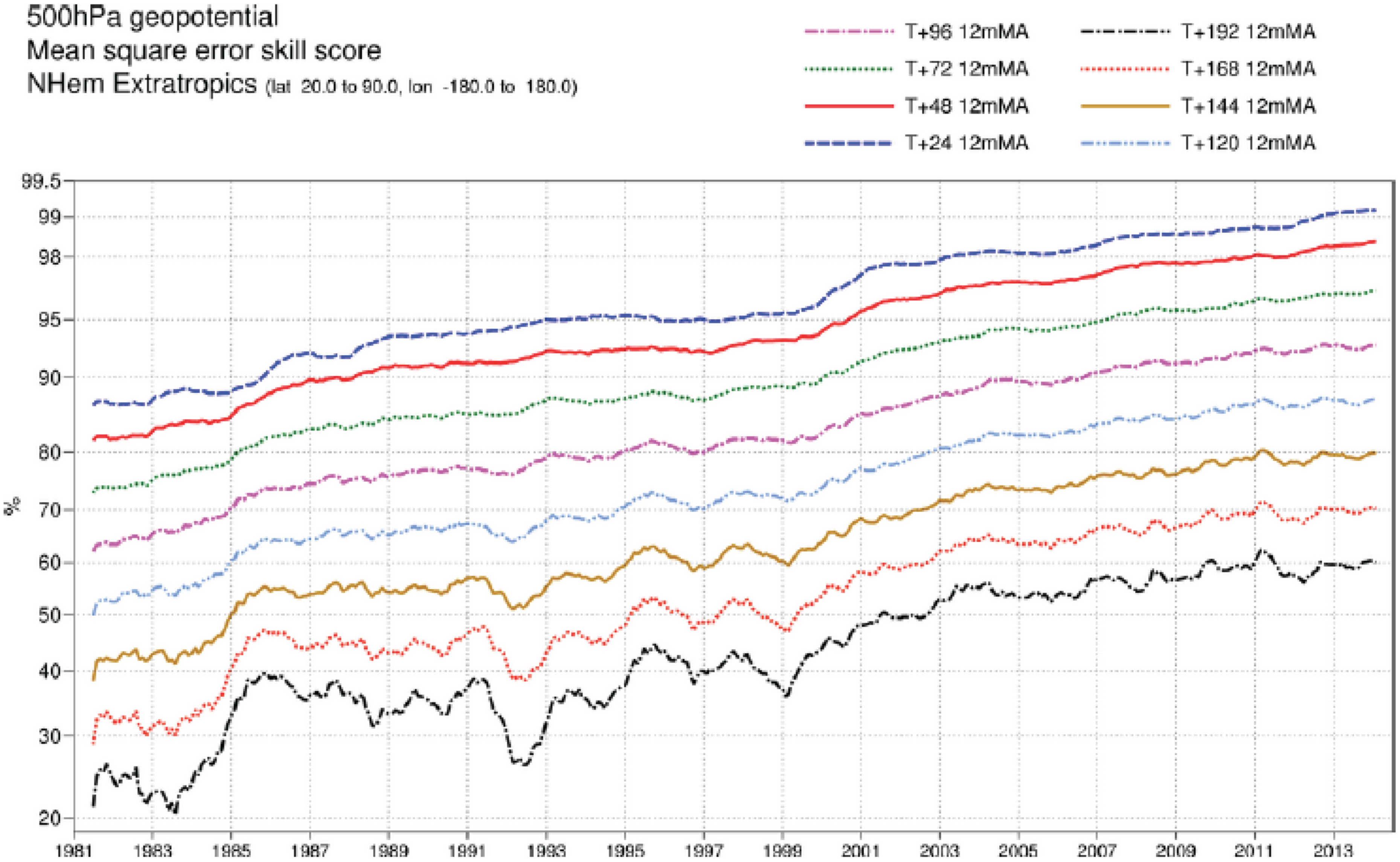
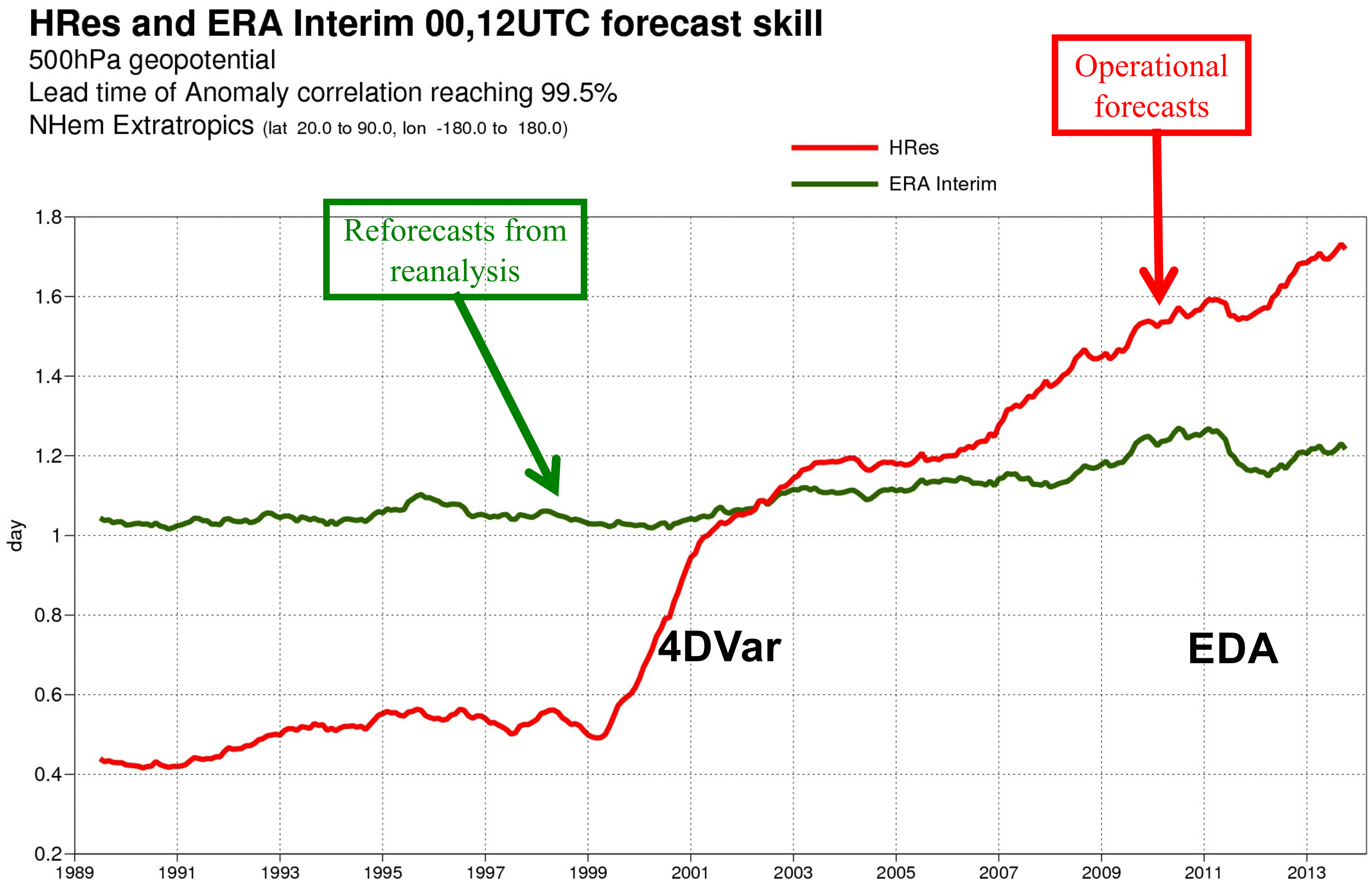


Figure 3: 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

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, and if errors are uncorrelated in time, Kalman Smoother and Variational Assimilation are algorithmically equivalent, and produce the *BLUE* of the state of the system from all data available over the 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 can take into account the corresponding correlations, and be 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

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

- 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 $\boldsymbol{\xi}_k^{(0)}$ of nonlinear model, replace tangent linear and adjoint models respectively by

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

and

$$\boldsymbol{\lambda}_k = \boldsymbol{L}_k^{\mathrm{T}}\, \boldsymbol{\lambda}_{k+1} + \ldots$$

where $\boldsymbol{L}_k$ is an appropriate simplification of jacobian $\boldsymbol{M}_k'$.

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 $\boldsymbol{\xi}_{\mathbf{0}}^{(0)} + \delta\boldsymbol{\xi}_0$ is equal to $\boldsymbol{\xi}_k^{(0)} + \delta\boldsymbol{\xi}_k$, 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

State vector $\boldsymbol{x}$, belonging to *state space* S ($\dim S = n$), 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

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

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

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

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

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

From course 3

Best Linear Unbiased Estimate

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

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

Equivalent set of formulæ

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

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

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

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

$$\boldsymbol{P}^a = \boldsymbol{C}_{\boldsymbol{xx}} - \boldsymbol{C}_{\boldsymbol{xy}}\,[\boldsymbol{C}_{\boldsymbol{yy}}]^{-1}\,\boldsymbol{C}_{\boldsymbol{yx}}$$

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

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

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

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

$$\boldsymbol{C}_{\boldsymbol{yy}} = E(\boldsymbol{y}'\boldsymbol{y}'^{\mathrm{T}}) \;=\; E[(\boldsymbol{\varepsilon} - \boldsymbol{H\zeta}^b)\,(\boldsymbol{\varepsilon} - \boldsymbol{H\zeta}^b)^{\mathrm{T}}]$$

$$= \underbrace{E(\boldsymbol{\varepsilon\varepsilon}^{\mathrm{T}})}_{\boldsymbol{R}} - \underbrace{E(\boldsymbol{\varepsilon\zeta}^{b\mathrm{T}})}_{\boldsymbol{D}}\boldsymbol{H}^{\mathrm{T}} - \;\boldsymbol{H}\underbrace{E(\boldsymbol{\varepsilon\zeta}^{b\mathrm{T}})}_{\boldsymbol{D}^{\mathrm{T.}}} + \boldsymbol{H}\underbrace{E(\boldsymbol{\zeta}^b\boldsymbol{\zeta}^{b\mathrm{T}})}_{\boldsymbol{P}^b}\boldsymbol{H}^{\mathrm{T}}$$

$$\boldsymbol{C}_{\boldsymbol{yy}} = \;\boldsymbol{R} - \boldsymbol{DH}^{\mathrm{T}} \;-\boldsymbol{HD}^{\mathrm{T}} + \;\boldsymbol{HP}^b\boldsymbol{H}^{\mathrm{T}}$$

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

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

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, $\boldsymbol{\zeta}$ unknown ‘error’

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

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

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

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

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 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 $\boldsymbol{P}^a$ have no necessary Bayesian meaning. In particular, for a given $\boldsymbol{z}$, $\boldsymbol{P}^a$, which is the covariance matrix of the estimation error over all possible realizations of $\boldsymbol{z}$ (*i.e.* of the error $\boldsymbol{\zeta}$), can be very different from the corresponding Bayesian 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 $\neq 0$) .

Three sets of equivalent equations

$$\boldsymbol{z} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta},\ E(\boldsymbol{\zeta}) = 0,\ E(\boldsymbol{\zeta\zeta}^{\mathrm{T}}) \equiv \boldsymbol{S}^{b}$$

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

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

$\boldsymbol{x}^{b} = \boldsymbol{x} + \boldsymbol{\zeta}^{b}$, $E(\boldsymbol{\zeta}^{b}) = 0,\ E(\boldsymbol{\zeta}^{b}\boldsymbol{\zeta}^{b\mathrm{T}}) \equiv \boldsymbol{P}^{b}$,

$\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon}$, $E(\boldsymbol{\varepsilon}) = 0$, $E(\boldsymbol{\varepsilon\varepsilon}^{\mathrm{T}}) \equiv \boldsymbol{R}$

$E(\boldsymbol{\zeta}^{b}\boldsymbol{\varepsilon}^{\mathrm{T}}) = 0$ (not restrictive)

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

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

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

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

The three equations have the same algebraic structure

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

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

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b, \ \boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon}$$

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

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

Data vector

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

Analysis $\boldsymbol{x}^a$ (whatever its exact meaning) minimizes following scalar *objective function*, defined on state space $\mathcal{X}$

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

where $\boldsymbol{S} = \mathrm{E}(\zeta\zeta^{\mathrm{T}})$ is covariance matrix of data error ζ

(for example 4D-Var

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

Consider quantity $D = z_1{}^{\mathrm{T}}\, S^{-1}\, z_2 = z_1{}^{\mathrm{T}}\, [\mathrm{E}(\zeta\zeta^{\mathrm{T}})]^{-1}\, z_2$

where z_1 and z_2 are any two vectors in data space

Change of coordinates $z \equiv Tw$

$$\zeta = T\chi \quad \Rightarrow S = \mathrm{E}(\zeta\zeta^{\mathrm{T}}) = \mathrm{E}[T\chi(T\chi)^{\mathrm{T}}] = T\,\mathrm{E}(\chi\chi^{\mathrm{T}})T^{\mathrm{T}}$$

$$D = w_1{}^{\mathrm{T}}\, T^{\mathrm{T}}\, [T\,\mathrm{E}(\chi\chi^{\mathrm{T}})T^{\mathrm{T}}]^{-1} Tw_2 = w_1{}^{\mathrm{T}}\, T^{\mathrm{T}}\, T^{-\mathrm{T}}\, [\mathrm{E}(\chi\chi^{\mathrm{T}})]^{-1}\, T^{-1}\, Tw_2$$

$$D = w_1{}^{\mathrm{T}}\, [\mathrm{E}(\chi\chi^{\mathrm{T}})]^{-1}\, w_2$$

Expression $D = z_1^{\mathrm{T}} S^{-1} z_2$

defines proper scalar product, and associated norm, on data space

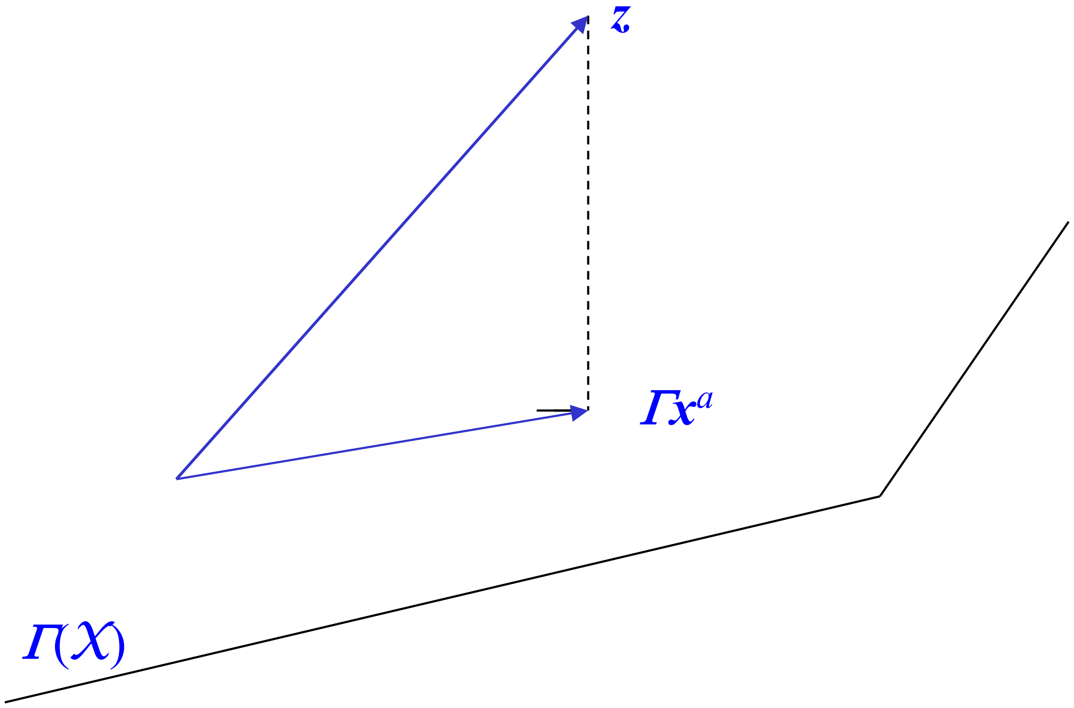
Called ***Mahalanobis norm***

Prasanta Chandra Mahalanobis (1893 -1972)

Minimizing objective function

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

amounts to orthogonal projection onto space $\boldsymbol{\Gamma}(\mathcal{X})$, followed by inversion through $\boldsymbol{\Gamma}$ (generalized inverse)



Gaussian variables

Unidimensional

$$\mathcal{N}[m, a] \sim (2\pi\, a)^{-1/2} \exp\,[-\,(1/2a)\,(\xi - m)^2]$$

Dimension n

$$\mathcal{N}[\boldsymbol{m}, \boldsymbol{A}] \sim [(2\pi)^n \det \boldsymbol{A}]^{-1/2} \exp\,[-\,(1/2)\,\underbrace{(\boldsymbol{\xi} - \boldsymbol{m})^{\mathrm{T}} \boldsymbol{A}^{-1} (\boldsymbol{\xi} - \boldsymbol{m})}_{\textit{Mahalanobis norm}}]$$

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 $\ln\surd(2\pi es)$.

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

Entropy of a probability distribution (continuation)

Data of the form (see slide 51)

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

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

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

If the probability distribution for ζ is unknown, assuming that it is Gaussian is in a sense the 'least committing' choice.

Cours à venir

~~Mercredi 2 avril~~

~~Vendredi 11 avril~~

~~Vendredi 18 avril~~

~~Mercredi 23 avril~~

~~Lundi 12 mai~~

Mercredi 28 mai

Mercredi 11 juin

Mercredi 18 juin