Data assimilation in the geosciences

An overview

Alberto Carrassi[1], **Olivier Talagrand**[2] and **Marc Bocquet**[3]

(1): Nansen Environmental and Remote Sensing Center - Norway
(2): LMD, École Normale Supérieure, IPSL, France
(3): CEREA, joint lab École des Ponts ParisTech and EdF R$D, IPSL, France

European Geosciences Union
General Assembly 2018

- *Variational Methods*

- *Ensemble Kalman Filter Methods*

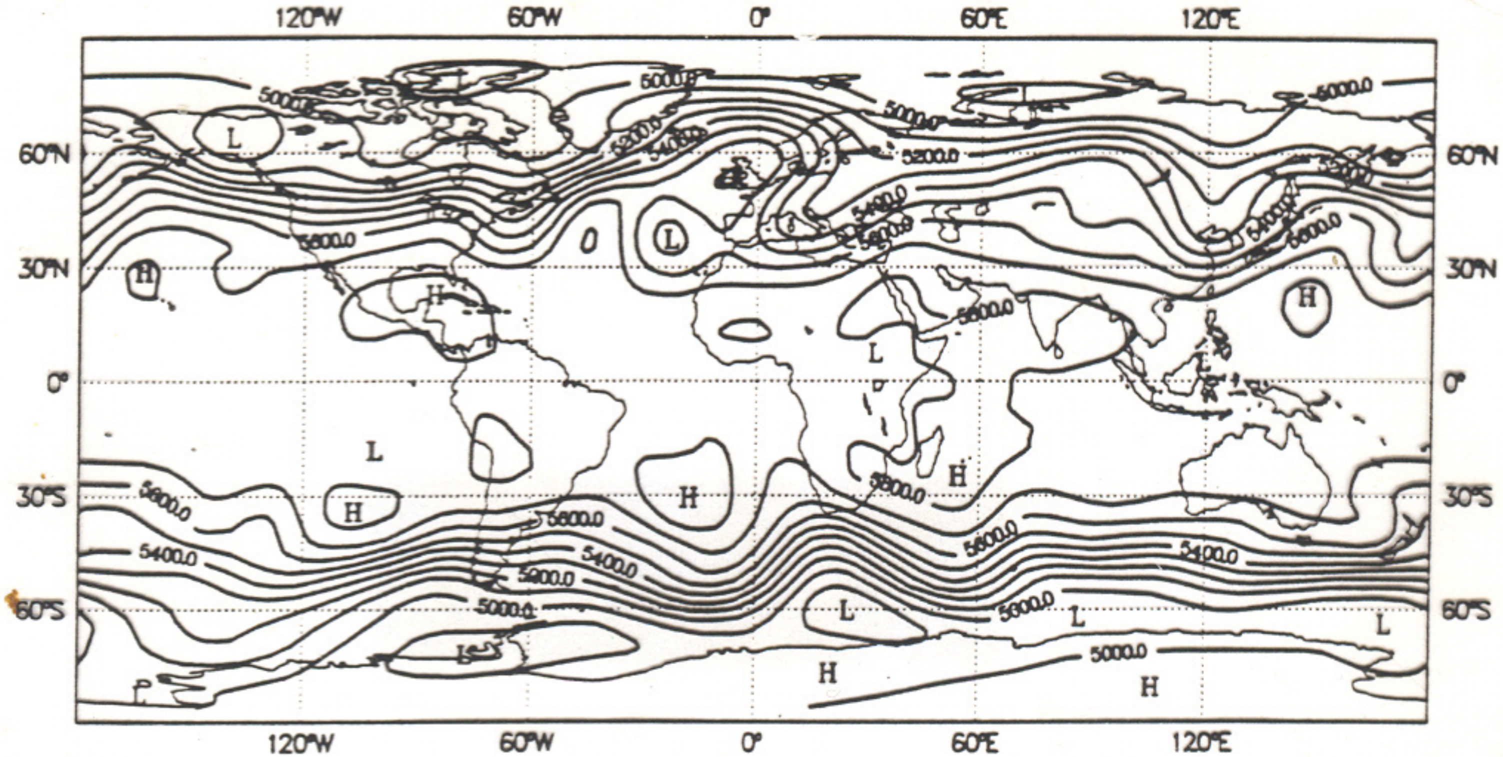
Kalman Filter, forecast step

$$P^b_{k+1} = M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

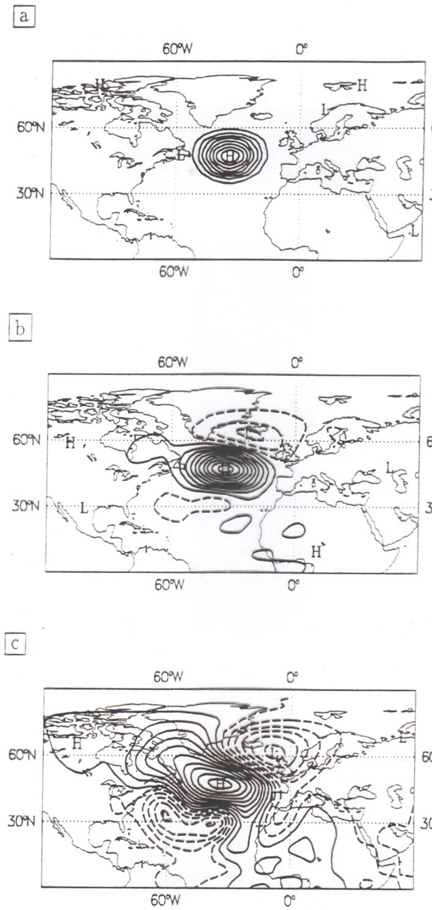
Multiplication by M_k = one integration of the model between times k and $k+1$.

Computation of $M_k P^a_k M_k^{\mathrm{T}} \approx 2n$ integrations of the model

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



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



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

Available data consist of

- Background estimate at time 0

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

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

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

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

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

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

Then objective function

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

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

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

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

Minimizing $\mathcal{J}(\xi_0)$ is exactly equivalent to smoothing (with exact model).

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

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 Approach

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

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

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

Adjoint equation, integrated backwards in time from from time K

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

….

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

….

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

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

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

Adjoint Approach (continued 2)

Nonlinearities ?

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

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

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

Adjoint equation

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

….

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

….

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

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

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

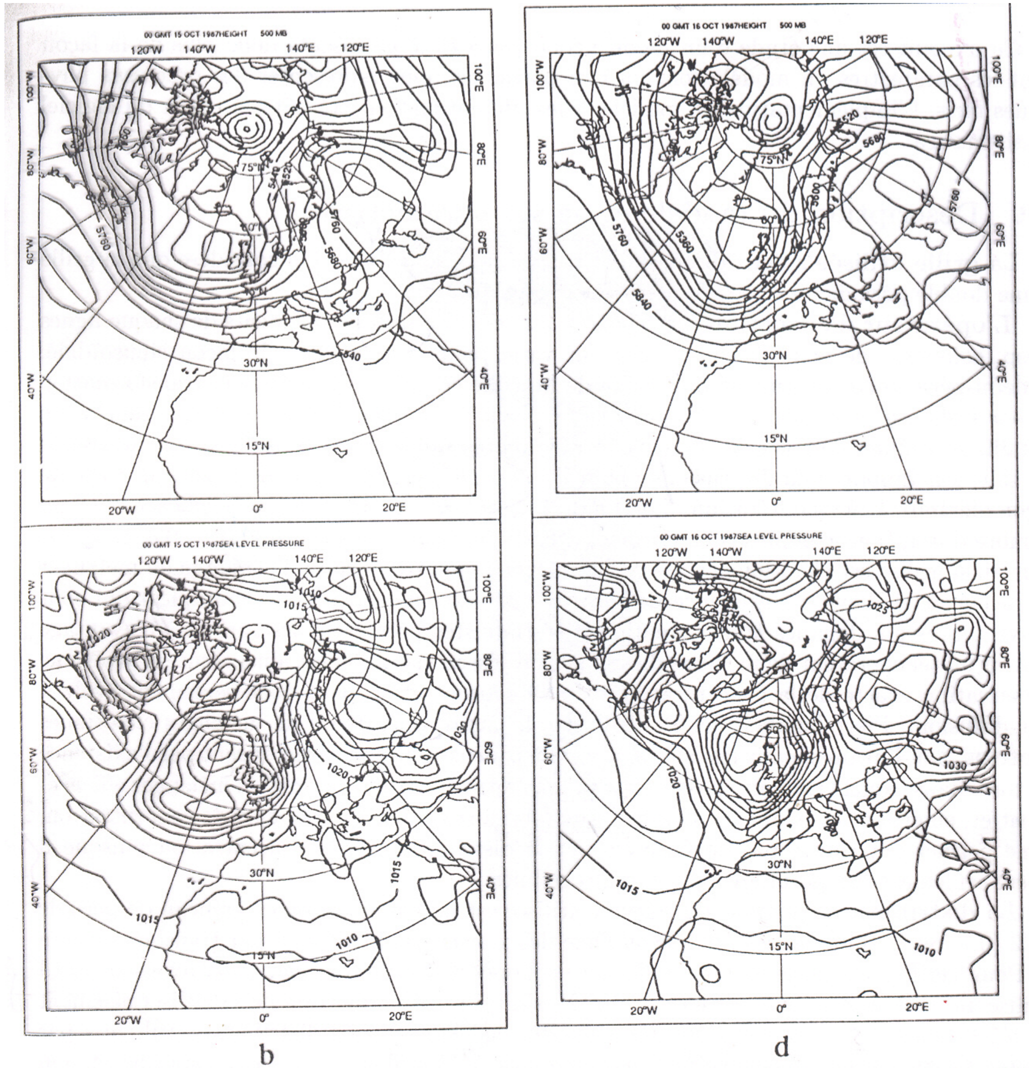
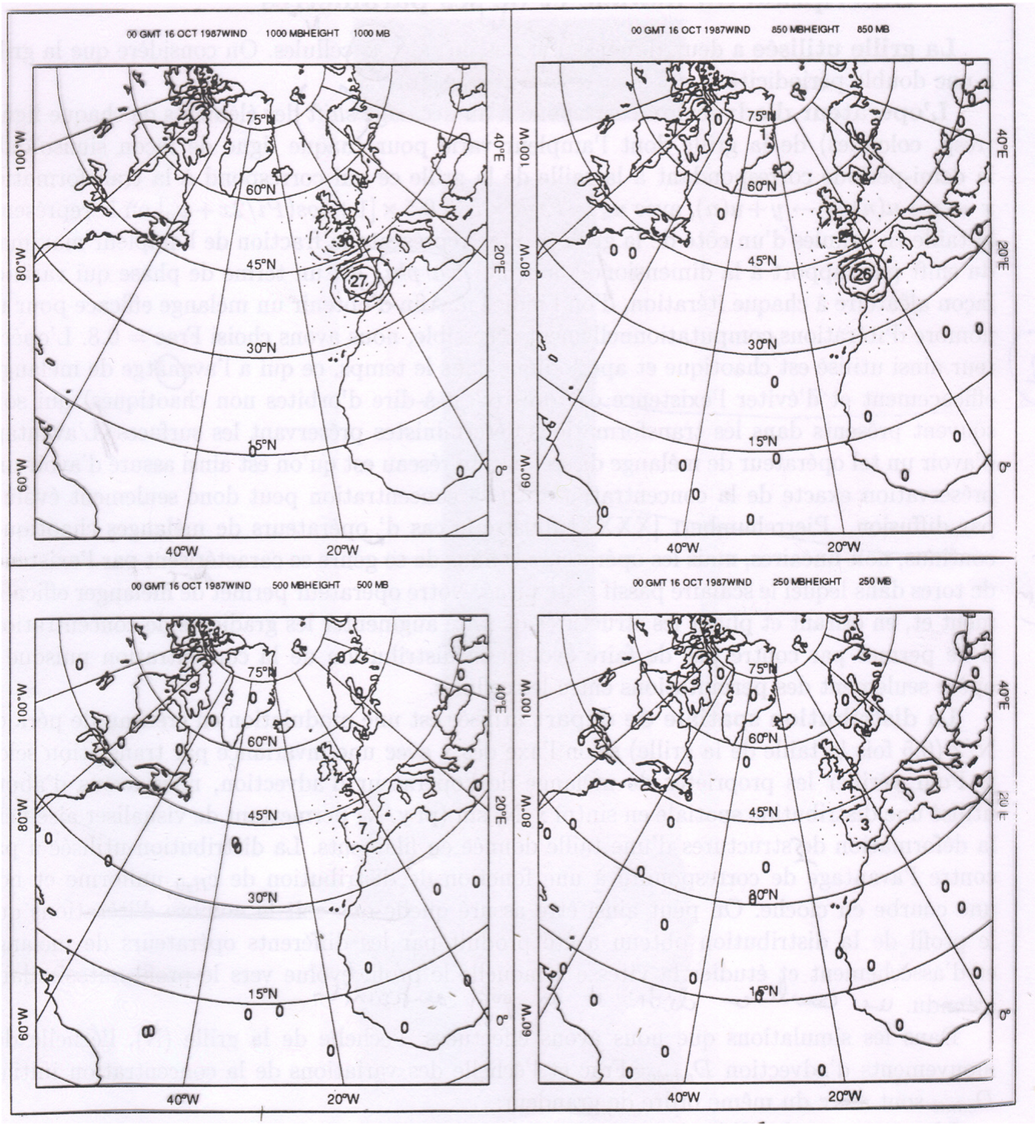


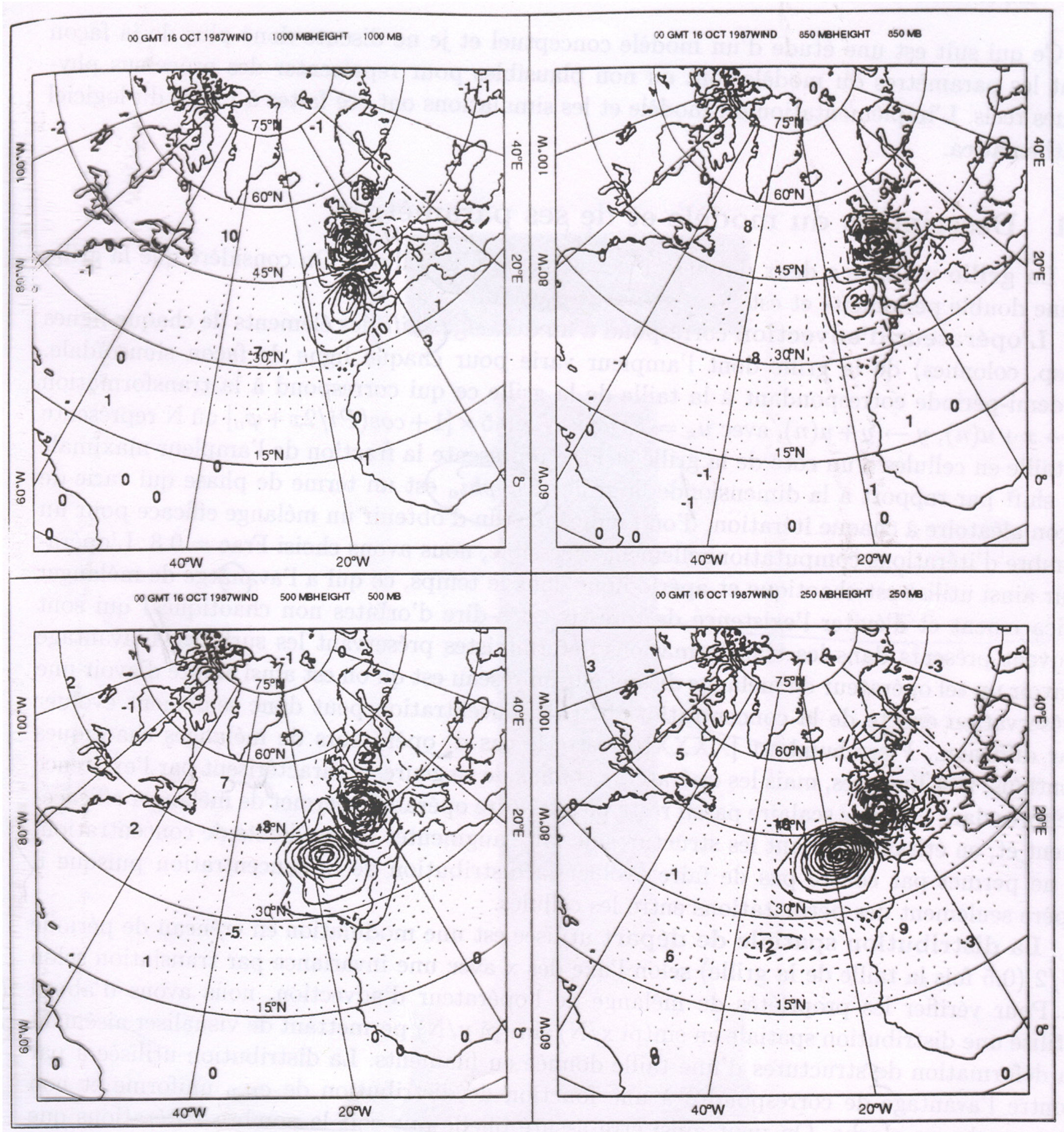
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October 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 *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

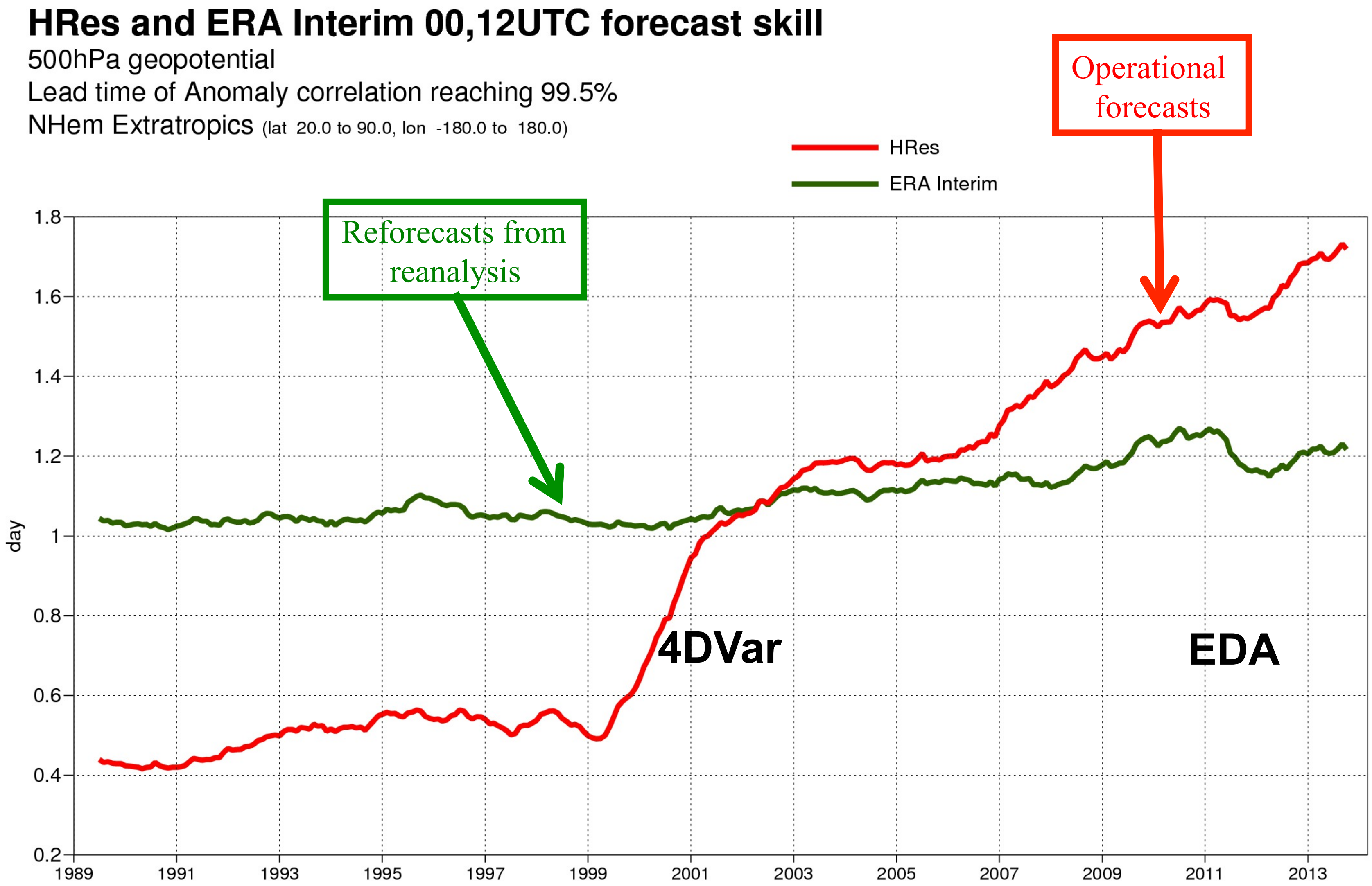
Thépaut *et al.*, 1993, *Mon. Wea. Rev.*, **121**, 3393-3414



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

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

Initial state error reduction



Credit E. Källén, ECMWF

Costliest part of computation

$$P^b_{k+1} = M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

Multiplication by M_k = one integration of the model between times k and $k+1$.

Computation of $M_k P^a_k M_k^{\mathrm{T}} \approx 2n$ integrations of the model

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

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

Weak constraint variational assimilation

Allows for errors in the assimilating model

- Data

  - Background estimate at time 0

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

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

  $$y_k = H_k x_k + \varepsilon_k \qquad E(\varepsilon_k \varepsilon_k^T) = R_k$$

  - Model

  $$x_{k+1} = M_k x_k + \eta_k \qquad E(\eta_k \eta_k^T) = Q_k \qquad k = 0, \ldots, K\text{-}1$$

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

Then objective function

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

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

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

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

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

Exactly equivalent to smoothing in the linear case

Can include nonlinear M_k and/or H_k.

Implemented operationally at ECMWF for the assimilation in the stratosphere.

In the linear case, and if errors are uncorrelated in time, Kalman Smoother and Variational Assimilation are algorithmically equivalent. 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 final time of the window). If in addition errors are Gaussian, both algorithms achieve Bayesian estimation.

Variational assimilation can easily take into account temporal correlations between errors (done oprationally by Järvinen *et al*., 1999).

Requires adjoint of the assimilating model. Must be developed in case one must implement variational assimilation on a model which has not been written with that purpose in mind. But adjoint, once it is available, can be used to other applications (sensitivity studies)

$$P^b{}_{k+1} = M_k P^a{}_k M_k^{\mathrm{T}} + Q_k$$

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

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

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

How to update predicted ensemble with new observations ?

Predicted ensemble at time k : $\{x^b{}_l\}$, $l = 1, \ldots, L$
Observation vector at same time : $y = Hx + \varepsilon$

- 'Gaussian' approach

Produce sample of probability distribution for real observed quantity Hx

$y_l = y - \varepsilon_l$

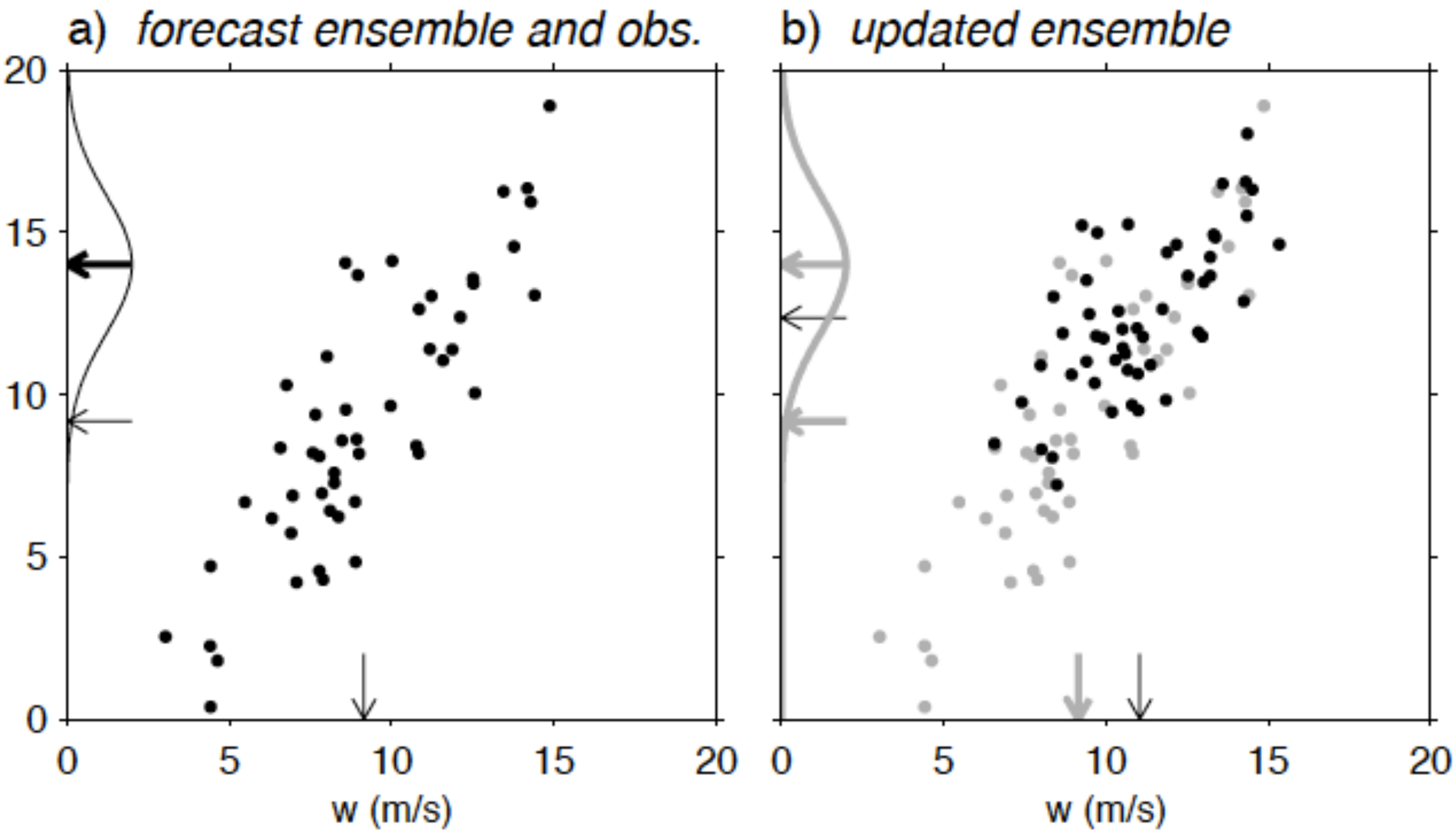
where ε_l is distributed according to probability distribution for observation error ε.

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

$$x^a{}_l = x^b{}_l + P^b H^T [HP^bH^T + R]^{-1} (y_l - Hx^b{}_l) , \qquad l = 1, \ldots, L \qquad (2)$$

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

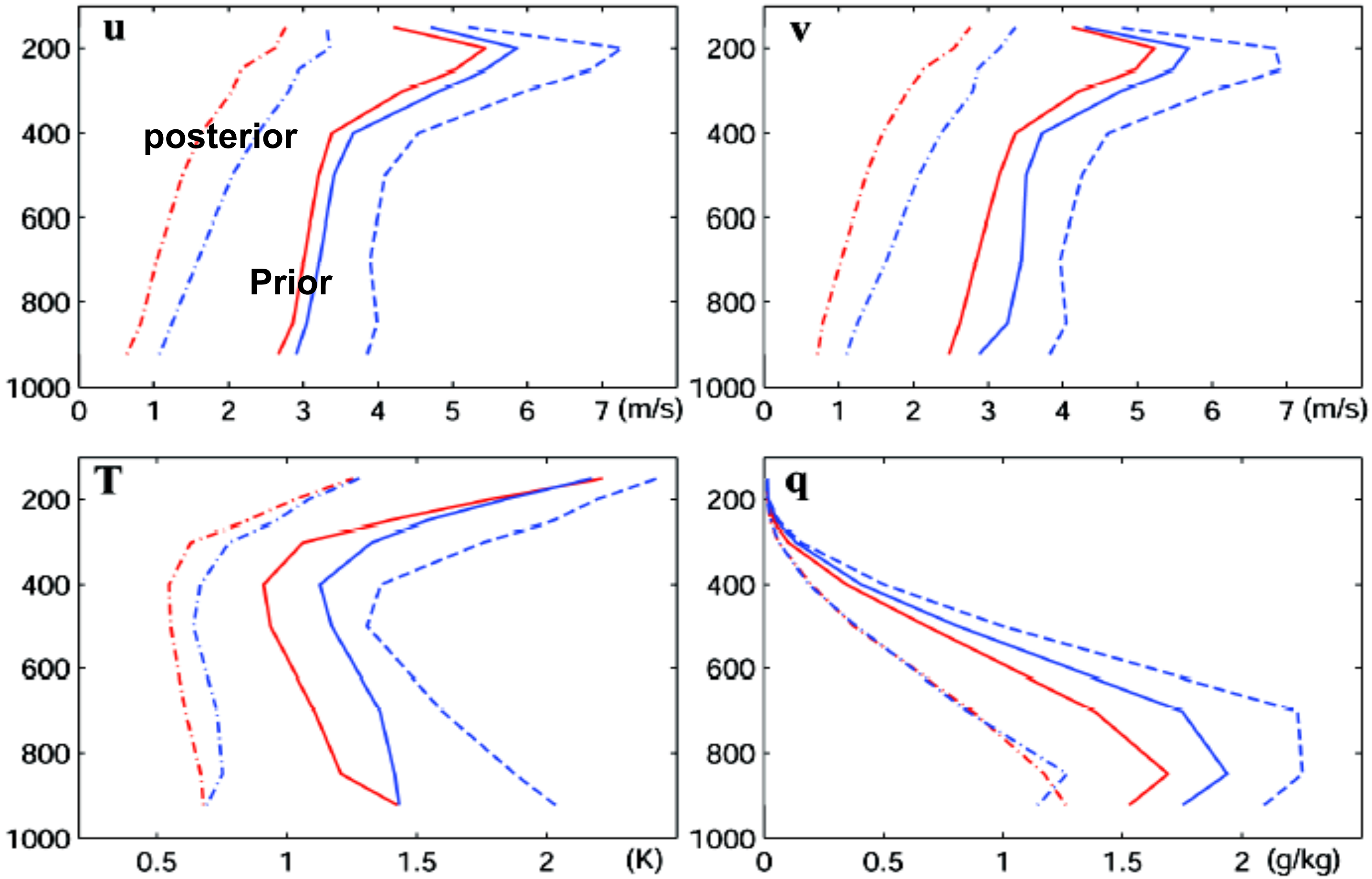
Remark. In case of Gaussian errors, if P^b was exact covariance matrix of background error, (2) would achieve Bayesian estimation, in the sense that $\{x^a{}_l\}$ would be a sample of conditional probability distribution for x, given all data up to time k.


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

C. Snyder

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

— EnKF — 3DVar (prior, solid; posterior, dotted)



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

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

The case of a nonlinear observation operator ?

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

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

Two possibilities

1. Take tangent linear approximation (as in Extended KF) and introduce jacobian H'

2. Formula

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

is basically

$$\boldsymbol{x}^a = x^b + \boldsymbol{C}_{\boldsymbol{xy}} [\boldsymbol{C}_{\boldsymbol{yy}}]^{-1} [\boldsymbol{y} - Hx^b]$$

where $\boldsymbol{C}_{\boldsymbol{xy}}$ is the covariance matrix between the background error and the innovation $y - Hx_b$, and $\boldsymbol{C}_{\boldsymbol{yy}}$ is the covariance matrix of the innovation.

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

There are specific problems with EnKF, resulting from the fact that the background sample covariance matrix has low rank.

There are many variants for it.

Does not easily take into account temporal correlations between errors.

Does not require an adjoint (makes things easier if one wants to use an existing model that does not already have an adajoint).

See next lecture by Marc Bocquet.