École Doctorale des Sciences de l'Environnement d'Île-de-France
Année Universitaire 2011-2012

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 6

15 Mai 2012

Best Linear Unbiased Estimate

State vector x, belonging to *state space* $\mathcal{S}$ (dim$\mathcal{S} = n$), to be estimated.
Available data in the form of

- A '*background*' estimate (*e. g.* forecast from the past), belonging to *state space*, with dimension n

  $$x^b = x + \zeta^b$$

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

  $$y = Hx + \varepsilon$$

  H is known linear *observation operator*.

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

Best Linear Unbiased Estimate (continuation 1)

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

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

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

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

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

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

Best Linear Unbiased Estimate (continuation 2)

Apply formulæ for Optimal Interpolation

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

x^a is the *Best Linear Unbiased Estimate* (*BLUE*) of x from x^b and y.

Equivalent set of formulæ

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

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

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

Best Linear Unbiased Estimate (continuation 3)

H can be any linear operator

Example : (scalar) satellite observation

$$\boldsymbol{x} = (x_1, \dots, x_n)^{\mathrm{T}} \text{ temperature profile}$$

Observation $y = \Sigma_i h_i x_i + \varepsilon = \boldsymbol{H}\boldsymbol{x} + \varepsilon$, $\boldsymbol{H} = (h_1, \dots, h_n)$, $E(\varepsilon^2) = r$

Background $\boldsymbol{x}^b = (x_1{}^b, \dots, x_n{}^b)^{\mathrm{T}}$, error covariance matrix $P^b = (p_{ij}{}^b)$

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

$$[HP^b H^{\mathrm{T}} + R]^{-1} (\boldsymbol{y} - H\boldsymbol{x}^b) = (y - \Sigma_\iota h_\iota x_\iota{}^b) / (\Sigma_{ij} h_i h_j p_{ij}{}^b + r)^{-1} \equiv \mu$$ scalar !

- $P^b = p^b \boldsymbol{I}_n$ $\quad x_i^a = x_i^b + p^b h_i \mu$
- $P^b = \mathrm{diag}(p_{ii}{}^b)$ $\quad x_i^a = x_i^b + p_{ii}{}^b h_i \mu$
- General case $\quad x_i^a = x_i^b + \Sigma_j p_{ij}{}^b h_j \mu$

Each level i is corrected, not only because of its own contribution to the observation, but because of the contribution of the other levels to which its background error is correlated.

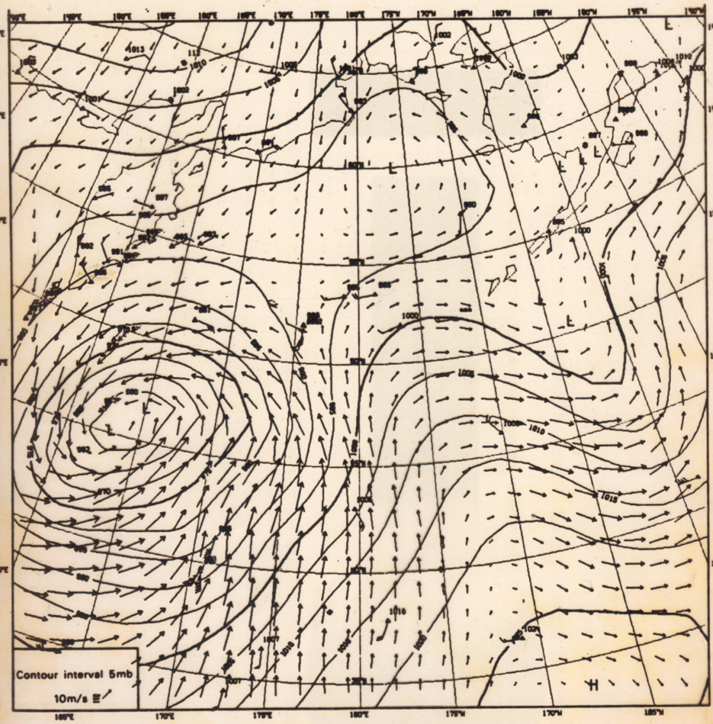


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

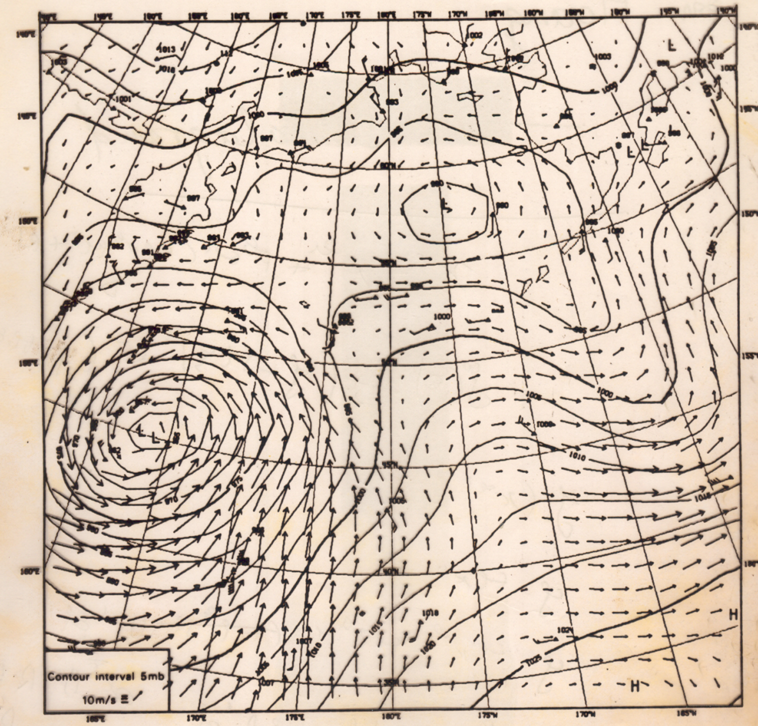


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

After A. Lorenc

Best Linear Unbiased Estimate (continuation 4)

Variational form of the *BLUE*

BLUE x^a minimizes following scalar *objective function*, defined on state space

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

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

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

'*3D-Var*'

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

Used operationally in USA, Australia, China, …

Question. How to introduce temporal dimension in estimation process ?

- Logic of Optimal Interpolation can be extended to time dimension.

- But we know much more than just temporal correlations. We know explicit dynamics.

Real (unknown) state vector at time k (in format of assimilating model) x_k. Belongs to state space $\mathcal{S}$ ($\dim \mathcal{S} = n$)

Evolution equation

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

M_k is (known) model, η_k is (unknown) model error

Sequential Assimilation

- Assimilating model is integrated over period of time over which observations are available. Whenever model time reaches an instant at which observations are available, state predicted by the model is updated with new observations.

Variational Assimilation

- Assimilating model is globally adjusted to observations distributed over observation period. Achieved by minimization of an appropriate scalar *objective function* measuring misfit between data and sequence of model states to be estimated.

- Observation vector at time k

$$y_k = H_k x_k + \varepsilon_k \qquad\qquad k = 0, \ldots, K$$

$$E(\varepsilon_k) = 0 \quad ; \quad E(\varepsilon_k \varepsilon_j^{\mathrm{T}}) = R_k\, \delta_{kj}$$

H_k linear

- Evolution equation

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

$$E(\eta_k) = 0 \quad ; \quad E(\eta_k \eta_j^{\mathrm{T}}) = Q_k\, \delta_{kj}$$

M_k linear

- $E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} H_k P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k x^a_k$$
$$\begin{aligned} P^b_{k+1} &= E[(x^b_{k+1} - x_{k+1})(x^b_{k+1} - x_{k+1})^{\mathrm{T}}] = E[(M_k x^a_k - M_k x_k - \eta_k)(M_k x^a_k - M_k x_k - \eta_k)^{\mathrm{T}}] \\ &= M_k E[(x^a_k - x_k)(x^a_k - x_k)^{\mathrm{T}}] M_k^{\mathrm{T}} - E[\eta_k (x^a_k - x_k)^{\mathrm{T}}] - E[(x^a_k - x_k)\eta_k^{\mathrm{T}}] + E[\eta_k \eta_k^{\mathrm{T}}] \\ &= M_k P^a_k M_k^{\mathrm{T}} + Q_k \end{aligned}$$

At time k, background x^b_k and associated error covariance matrix P^b_k known

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} (y_k - H_k x^b_k)$$
$$P^a_k = P^b_k - P^b_k H_k^{\mathrm{T}} [H_k P^b_k H_k^{\mathrm{T}} + R_k]^{-1} H_k P^b_k$$

- Forecast step

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

Kalman filter (KF, Kalman, 1960)

Must be started from some initial estimate (x^b_0, P^b_0)

If all operators are linear, and if errors are uncorrelated in time, Kalman filter produces at time k the *BLUE* $x^b{}_k$ (resp. $x^a{}_k$) of the real state x_k from all data prior to (resp. up to) time k, plus the associated estimation error covariance matrix $P^b{}_k$ (resp. $P^a{}_k$).

If in addition errors are gaussian, the corresponding conditional probability distributions are the respective gaussian distributions

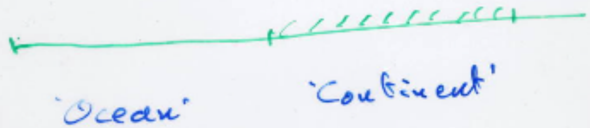
$\mathcal{N}[x^b{}_k, P^b{}_k]$ and $\mathcal{N}[x^a{}_k, P^a{}_k]$.

A didactic example (Ghil et al.)

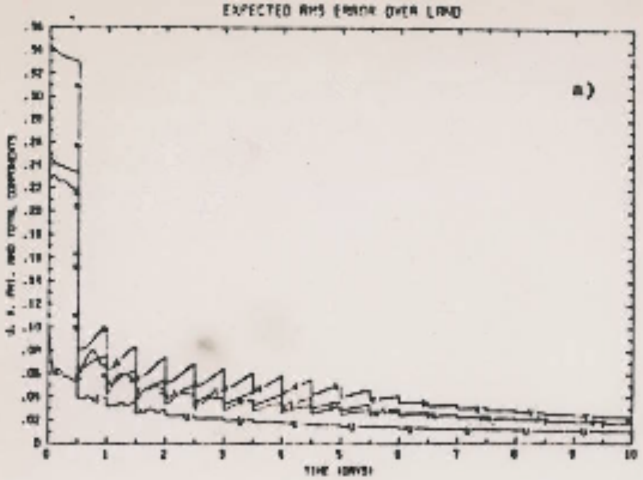
Barotropic model

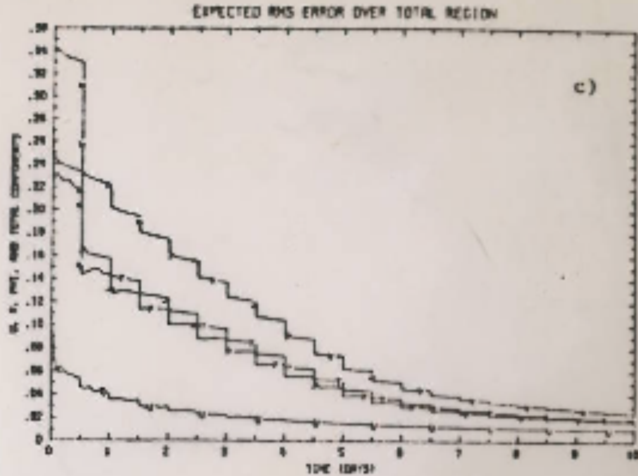
$$\begin{cases} \dfrac{\partial \varphi}{\partial t} + \operatorname{div}(\varphi\, \underline{U}) = 0 \\ \dfrac{\partial \underline{U}}{\partial t} + \underline{\operatorname{grad}}\left(\varphi + \frac{1}{2} U^2\right) + \underline{k} \times (f + \xi)\, \underline{U} = 0 \end{cases}$$

One dimension, periodic



Linearized (conserves energy)





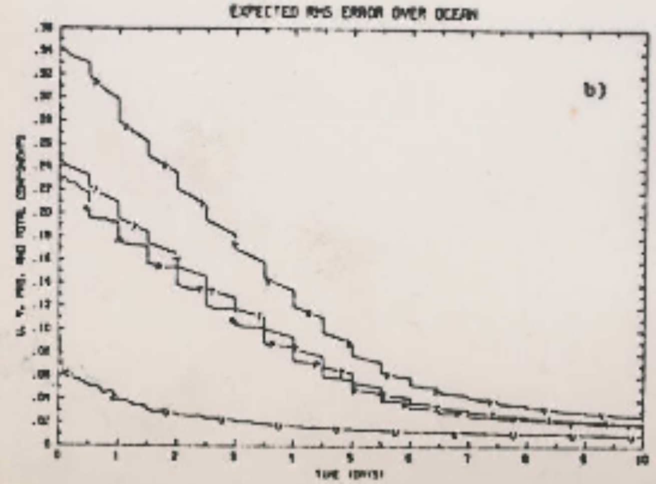
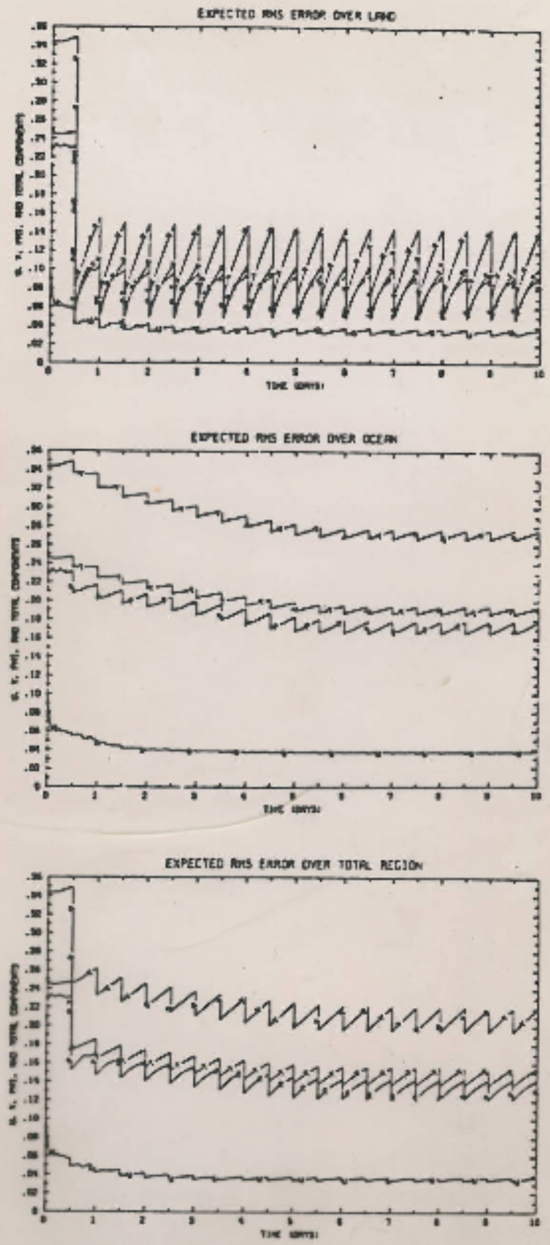


Fig. 2

The components of the total expected rms error (Erms), (trace: $P_k)^{1/2}$, in the estimation of solutions to the stochastic-dynamic system (Ψ, H), with Ψ given by (3.6) and $H = (I\ 0)$. System noise is absent, $Q = 0$. The filter used is the standard K-B filter (2.11) for the model.

a) Erms over land; b) Erms over the ocean; c) Erms over the entire L-domain

In each one of the figures, each curve represents one component of the total Erms error. The curves labelled U, V, and P represent the u component, v component and ϕ component, respectively. They are found by summing the diagonal elements of P_k which correspond to u, v, and ϕ, respectively, dividing by the number of terms in the sum, and then taking the square root. In a) the summation extends over land points only, in b) over ocean points only, and in c) over the entire L-domain. The vertical axis is scaled in such a way that 1.0 corresponds to an Erms error of v_{max} for the U and V curves, and of ϕ_0 for the P curve. The observational error level is 0.089 for the U and V curves, and 0.080 for the P curve. The curves labelled T represent the total Erms error over each region. Each T curve is a weighted average of the corresponding U, V, and P curves, with the weights chosen in such a way that the T curve measures the error in the total energy $u^2 + v^2 + \phi^2/\bar{\phi}$, conserved by the system (3.1). The observational noise level for the T curve is then 0.088. Notice the immediate error decrease over land and the gradual decrease over the ocean. The total estimation error tends to zero.



M. Ghil *et al*.

Fig. 6 This figure and the following ones show the properties of the estimated algorithm (2.11) in the presence of system noise, $Q \neq 0$. This figure gives the Erms estimation error, and is homologous to Fig. 2. Notice the sharper increase of error over land between synoptic times, and the convergence of each curve to a periodic, nonzero function.

Nonlinearities ?

Model is usually nonlinear, and observation operators (satellite observations) tend more and more to be nonlinear.

- Analysis step

$$x^a_k = x^b_k + P^b_k H_k'^T [H_k' P^b_k H_k'^T + R_k]^{-1} [y_k - H_k(x^b_k)]$$
$$P^a_k = P^b_k - P^b_k H_k'^T [H_k' P^b_k H_k'^T + R_k]^{-1} H_k' P^b_k$$

- Forecast step

$$x^b_{k+1} = M_k(x^a_k)$$
$$P^b_{k+1} = M_k' P^a_k M_k'^T + Q_k$$

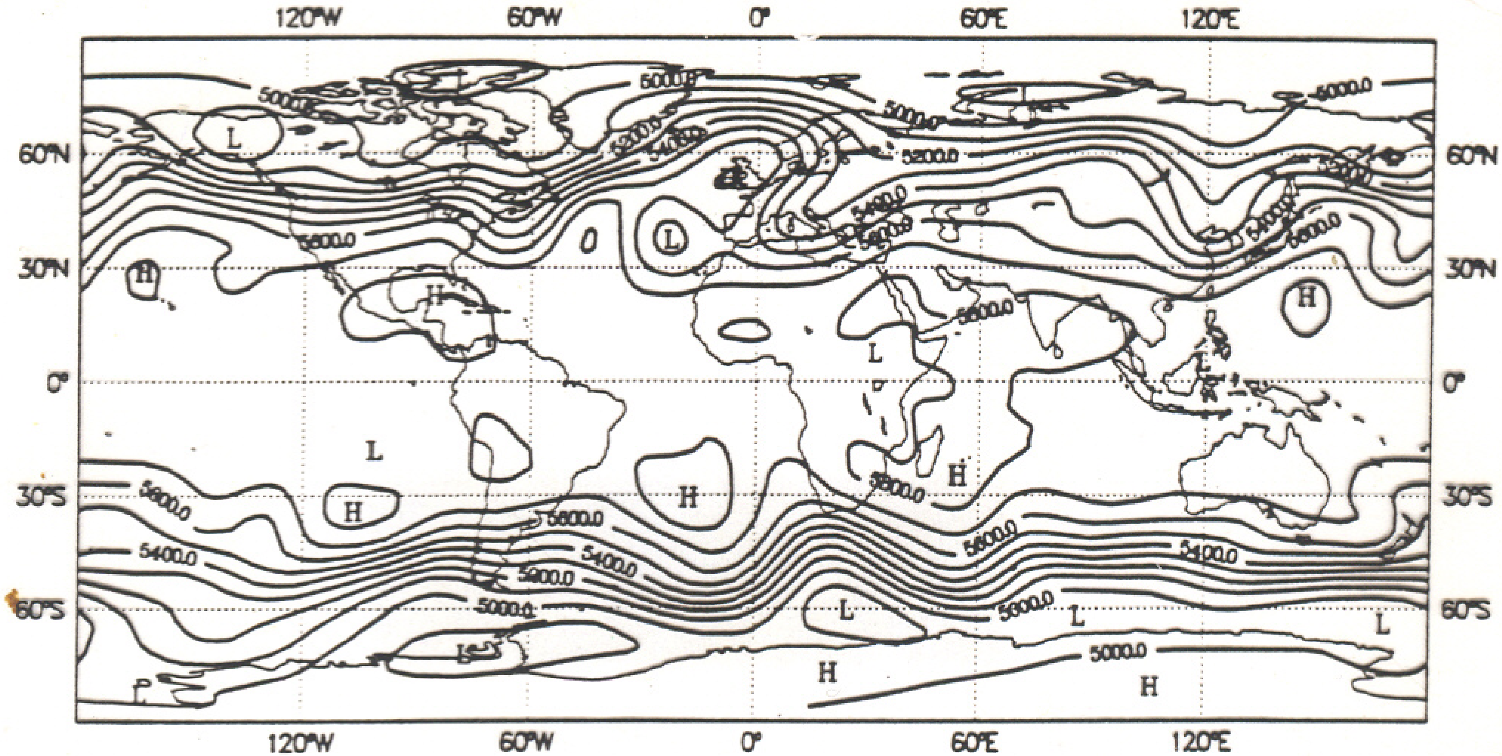
Extended Kalman Filter (EKF, heuristic !)

Costliest part of computation

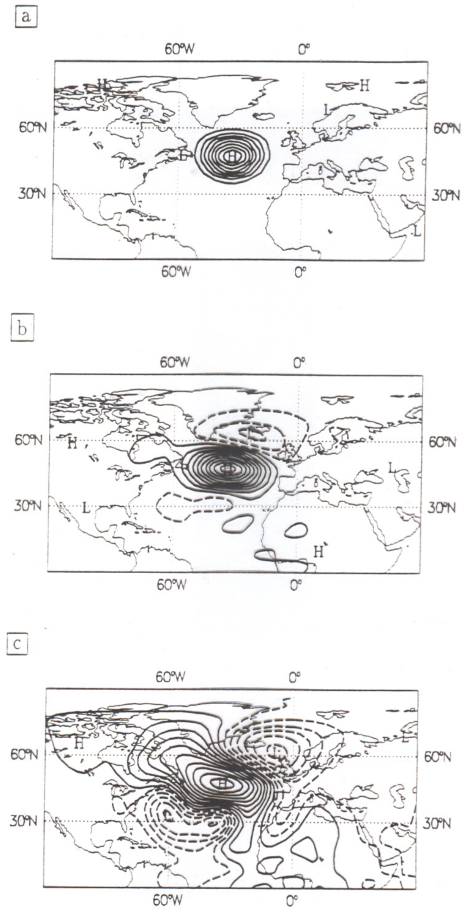
$$P^b_{k+1} = M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

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



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



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.

Two solutions :

- Low-rank filters (Heemink, Pham, …)

  Reduced Rank Square Root Filters, Singular Evolutive Extended Kalman Filter, ….

- Ensemble filters (Evensen, Anderson, …)

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

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

How to update predicted ensemble with new observations ?

Predicted ensemble at time t : $\{x^b{}_n\}$, $n = 1, \ldots, N$
Observation vector at same time : $y = Hx + \varepsilon$

- Gaussian approach

  Produce sample of probability distribution for real observed quantity Hx
  $y_n = y - \varepsilon_n$
  where ε_n is distributed according to probability distribution for observation error ε
  .

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

  $$x^a{}_n = x^b{}_n + P^b H^T [HP^bH^T + R]^{-1} (y_n - Hx^b{}_n) , \qquad n = 1, \ldots, N \qquad (2)$$

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

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

Called ***Ensemble Kalman Filter*** (*EnKF*)

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Empirical 'covariance inflation'

- Spurious correlations appear at large geographical distances. Empirical 'localization'.

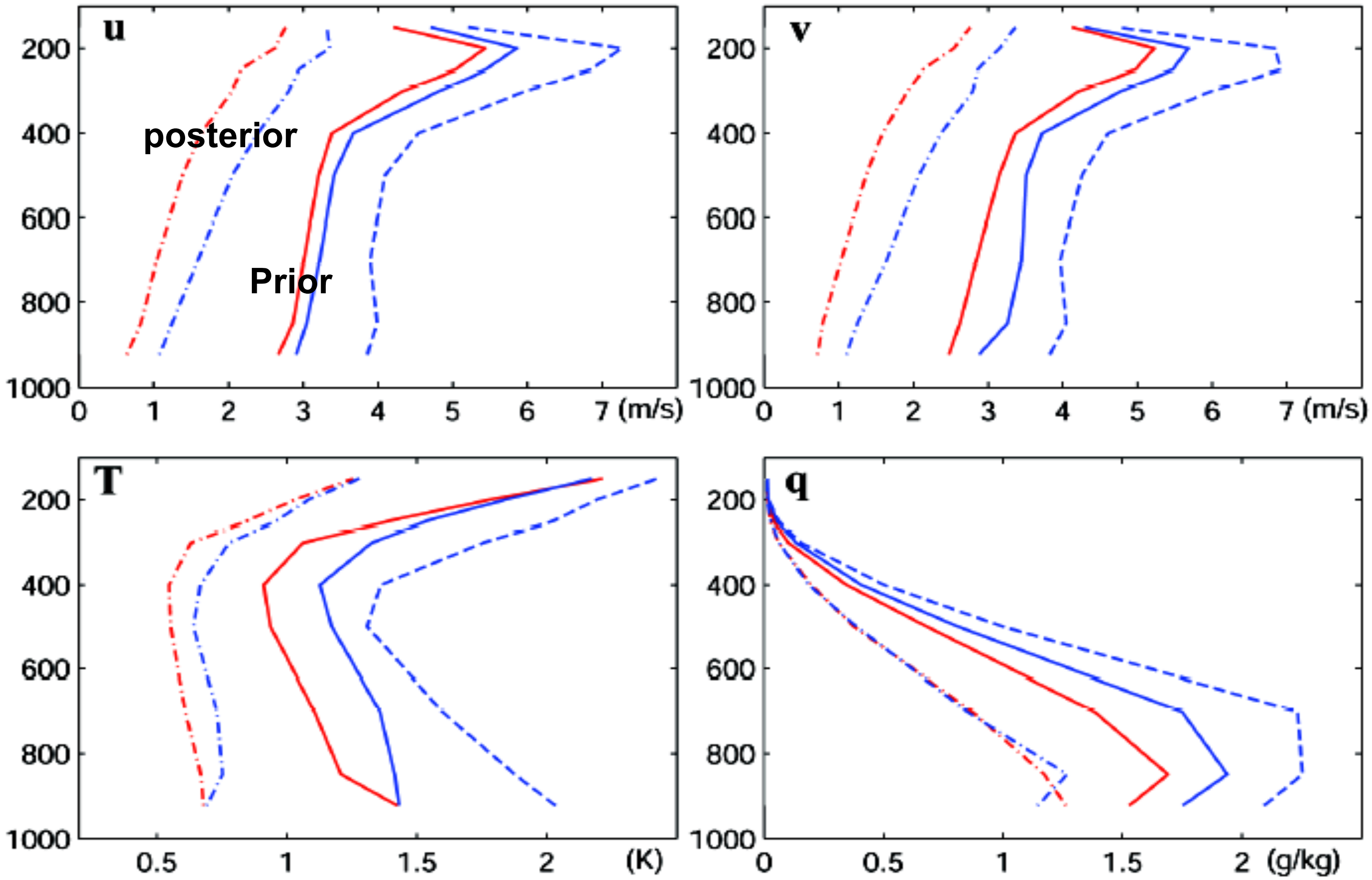
In formula

$$x^a_n = x^b_n + P^b H^T [HP^b H^T + R]^{-1} (y_n - Hx^b_n) , \qquad n = 1, \ldots, N$$

P^b, which is covariance matrix of an N-size ensemble, has rank N-1 at most. This means that corrections made on ensemble elements are contained in a subspace with dimension N-1. Obviously very restrictive if $N \ll p$, $N \ll n$.

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)

Situation still not entirely clear.

Houtekamer and Mitchell (1998) use two ensembles, the elements of each of which are updated with covariance matrix of other ensemble.

Local Ensemble Transform Kalman Filter (*LETKF*) defined by Kalnay and colleagues. Correction is performed locally in space on the basis of neighbouring observations.

In any case, optimality always requires errors to be independent in time. In order to relax that constraint, it is necessarily to augment the state vector in the temporal dimension.