École Doctorale des Sciences de l'Environnement d'Île-de-France

Année Universitaire 2021-2022

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation de Données

Olivier Talagrand

Cours 4

14 Avril 2022

- Best Linear Unbiased Estimator. Complements

- How to introduce temporal dynamics in assimilation ? Kalman Filter. Theory. One didactic example.

- How to introduce nonlinearity ? Reduced Rank Kalman Filters. Ensemble Kalman Filter

- Kalman Smoother

Best Linear Unbiased Estimate

State vector $\boldsymbol{x}$, 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

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

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

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

  $\boldsymbol{H}$ 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 $\boldsymbol{B}$), $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) \equiv \boldsymbol{R}$

Best Linear Unbiased Estimate (continuation 3)

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

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

Equivalent set of formulæ

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

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

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

1200 GMT 19 January 1979

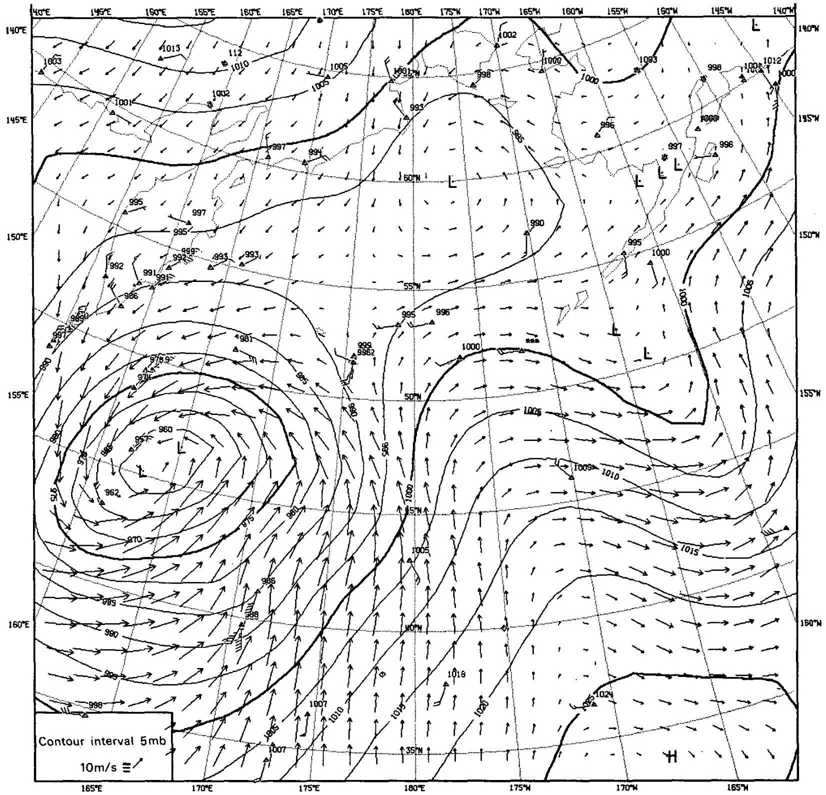


FIG. 14. Sea level pressure and wind forecast corresponding to the central area of Fig. 11, with plotted surface observations of sea level pressure and wind (each barb = 5 m s⁻¹).

1200 GMT 19 January 1979

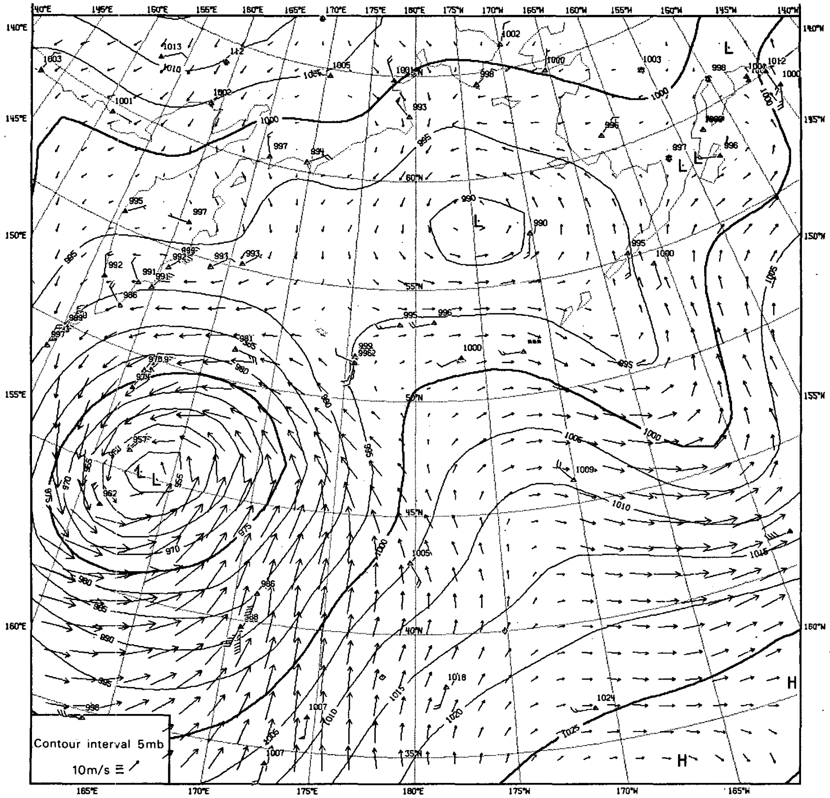


FIG. 15. As in Fig. 14 for the analysis in the data-assimilation cycle.

After A. Lorenc, MWR, 1981

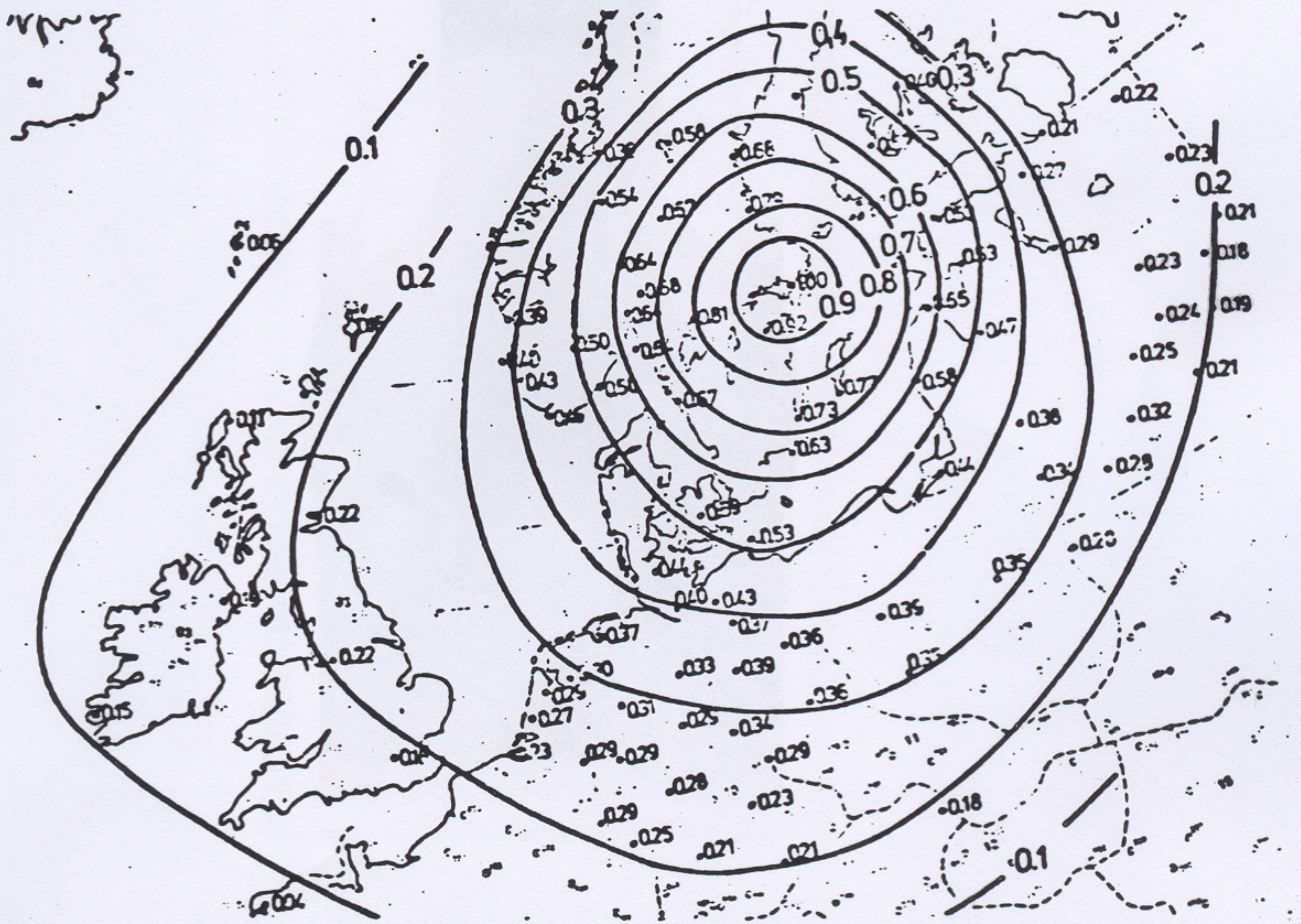


Figure 5.1.1.4.1 Auto-correlation of errors in 12h numerical forecasts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Each observation comes with its *representer*

Best Linear Unbiased Estimate (continuation 5)

BLUE is invariant in any invertible linear change of variables, in either state or observation space.

Equivalently, *BLUE* is independent of the possible choice of a scalar product in either one of the two spaces.

Best Linear Unbiased Estimate (continuation 6)

Variational form of the *BLUE*

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

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

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

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

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

'3D-Var'

Can easily, and heuristically, be extended to the case of a nonlinear observation operator $\boldsymbol{H}$.

Used operationally in USA, Australia, China, …

Best Linear Unbiased Estimate

The case of a nonlinear observation operator

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

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

Innovation $\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^b) = \boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{x}^b) + \boldsymbol{\varepsilon}$

$$\approx \boldsymbol{H}'(\boldsymbol{x} - \boldsymbol{x}^b) + \boldsymbol{\varepsilon}$$ if $\boldsymbol{x} - \boldsymbol{x}^b$ small

where $\boldsymbol{H}'$ is *Jacobian* matrix of $\boldsymbol{H}$ (matrix of partial derivatives) at point $\boldsymbol{x}^b$

Problem becomes linear in $\boldsymbol{x} - \boldsymbol{x}^b$

Tangent linear approximation

Best Linear Unbiased Estimate

$$0 = \boldsymbol{x} - \boldsymbol{x}^b + \xi^b$$

$$\boldsymbol{d} = \boldsymbol{H'}(\boldsymbol{x} - \boldsymbol{x}^b) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b \boldsymbol{H'}^{\mathrm{T}} [\boldsymbol{H'P}^b\boldsymbol{H'}^{\mathrm{T}} + \boldsymbol{R}]^{-1} [\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^b)]$$

$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b \boldsymbol{H'}^{\mathrm{T}} [\boldsymbol{H'P}^b\boldsymbol{H'}^{\mathrm{T}} + \boldsymbol{R}]^{-1} \boldsymbol{H'P}^b$$

Analogue for variational form. Minimize

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

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

- How to introduce temporal dynamics in assimilation ? Kalman Filter. Theory. One didactic example.

Question. How to introduce temporal dimension in estimation process ?

- Logic of Optimal Interpolation and of *BLUE* 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) $\boldsymbol{x}_k$. Belongs to state space $\mathcal{S}$ ($\dim \mathcal{S} = n$)

Evolution equation

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

$\boldsymbol{M}_k$ is (known) model, $\boldsymbol{\eta}_k$ 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. In the jargon of the trade, *Optimal Interpolation* designates an algorithm for sequential assimilation in which the matrix $\boldsymbol{P}^b$ is constant with time, and *3D-Var* an algorithm in which, in addition, the analysis $\boldsymbol{x}^a$ is obtained through a variational algorithm.

Variational Assimilation

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

Sequential Assimilation

Optimal Interpolation

- Observation vector at time k

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{\varepsilon}_k \qquad\qquad k = 0, \ldots, K$$

$$E(\boldsymbol{\varepsilon}_k) = 0 \quad ; \quad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = \boldsymbol{R}_k \, \delta_{kj}$$

$\boldsymbol{H}_k$ linear

- Evolution equation

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

Optimal Interpolation (2)

At time k, background $\boldsymbol{x}^b{}_k$ and associated error covariance matrix $\boldsymbol{P}^b$ known, assumed to be independent of k.

- Analysis step

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

In *3D-Var*, $\boldsymbol{x}^a{}_k$ is obtained by (iterative) minimization of associated objective function

- Forecast step

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

Sequential Assimilation. *Kalman Filter*

- Observation vector at time k

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{\varepsilon}_k \qquad k = 0, \ldots, K$$

$$E(\boldsymbol{\varepsilon}_k) = 0 \quad ; \quad E(\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = \boldsymbol{R}_k\, \delta_{kj}$$

$\boldsymbol{H}_k$ linear

- Evolution equation

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

$$E(\boldsymbol{\eta}_k) = 0 \quad ; \quad E(\boldsymbol{\eta}_k \boldsymbol{\eta}_j^{\mathrm{T}}) = \boldsymbol{Q}_k\, \delta_{kj}$$

$\boldsymbol{M}_k$ linear

- $E(\boldsymbol{\eta}_k \boldsymbol{\varepsilon}_j^{\mathrm{T}}) = 0$ (errors uncorrelated in time)

At time k, background $\boldsymbol{x}^b_k$ and associated error covariance matrix $\boldsymbol{P}^b_k$ known

- Analysis step

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

- Forecast step ($\boldsymbol{M}_k$ linear)

$$\boldsymbol{x}^b_{k+1} = \boldsymbol{M}_k \boldsymbol{x}^a_k$$
$$\boldsymbol{P}^b_{k+1} = E[(\boldsymbol{x}^b_{k+1} - \boldsymbol{x}_{k+1})(\boldsymbol{x}^b_{k+1} - \boldsymbol{x}_{k+1})^{\mathrm{T}}] = E[(\boldsymbol{M}_k \boldsymbol{x}^a_k - \boldsymbol{M}_k \boldsymbol{x}_k - \boldsymbol{\eta}_k)(\boldsymbol{M}_k \boldsymbol{x}^a_k - \boldsymbol{M}_k \boldsymbol{x}_k - \boldsymbol{\eta}_k)^{\mathrm{T}}]$$
$$= \boldsymbol{M}_k E[(\boldsymbol{x}^a_k - \boldsymbol{x}_k)(\boldsymbol{x}^a_k - \boldsymbol{x}_k)^{\mathrm{T}}] \boldsymbol{M}_k^{\mathrm{T}}$$
$$- E[\boldsymbol{\eta}_k (\boldsymbol{x}^a_k - \boldsymbol{x}_k)^{\mathrm{T}}] \boldsymbol{M}_k^{\mathrm{T}} - \boldsymbol{M}_k E[(\boldsymbol{x}^a_k - \boldsymbol{x}_k) \boldsymbol{\eta}_k^{\mathrm{T}}] + E[\boldsymbol{\eta}_k \boldsymbol{\eta}_k^{\mathrm{T}}]$$
$$= \boldsymbol{M}_k \boldsymbol{P}^a_k \boldsymbol{M}_k^{\mathrm{T}} + \boldsymbol{Q}_k$$

At time k, background $\boldsymbol{x}^b{}_k$ and associated error covariance matrix $\boldsymbol{P}^b{}_k$ known

- Analysis step

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

- Forecast step

$$\boldsymbol{x}^b{}_{k+1} = \boldsymbol{M}_k \boldsymbol{x}^a{}_k$$
$$\boldsymbol{P}^b{}_{k+1} = \boldsymbol{M}_k \boldsymbol{P}^a{}_k \boldsymbol{M}_k{}^{\mathrm{T}} + \boldsymbol{Q}_k$$

Kalman filter (KF, Kalman, 1960)

Must be started from some initial estimate $(\boldsymbol{x}^b{}_0, \boldsymbol{P}^b{}_0)$

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

If in addition errors are globally gaussian, the corresponding conditional probability distributions are the respective gaussian distributions $\mathcal{N}[\boldsymbol{x}^b{}_k, \boldsymbol{P}^b{}_k]$ and $\mathcal{N}[\boldsymbol{x}^a{}_k, \boldsymbol{P}^a{}_k]$.

Kalman filter. A simple example (Ghil *et al*.)

Shallow-water equations (aka *équations de Saint-Venant*)

$$\frac{\partial \varphi}{\partial t} + div(\varphi \boldsymbol{U}) = 0$$

$$\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{t}} + grad(\varphi + \frac{1}{2}\boldsymbol{U}^2) + k \wedge (f + \zeta)\boldsymbol{U} = 0$$

Periodic domain *D*. Equations conserve energy

$$E \equiv \frac{1}{2}\int_D (\varphi^2 + \varphi \boldsymbol{U}^2) dS$$

Equations linearized in the vicinity of state of rest

($\varphi = \Phi_0$, $\boldsymbol{U} = 0$)

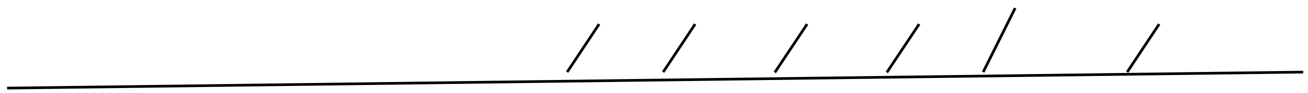
$$\frac{\partial \varphi}{\partial t} + \Phi_0 div \boldsymbol{U} = 0$$

$$\frac{\partial \boldsymbol{U}}{\partial t} + grad\varphi + k \wedge f\boldsymbol{U} = 0$$

Conserve quadratic energy

$$E \equiv \frac{1}{2} \int_D (\varphi^2 + \Phi_0 U^2) dS$$

Unidimensional domain



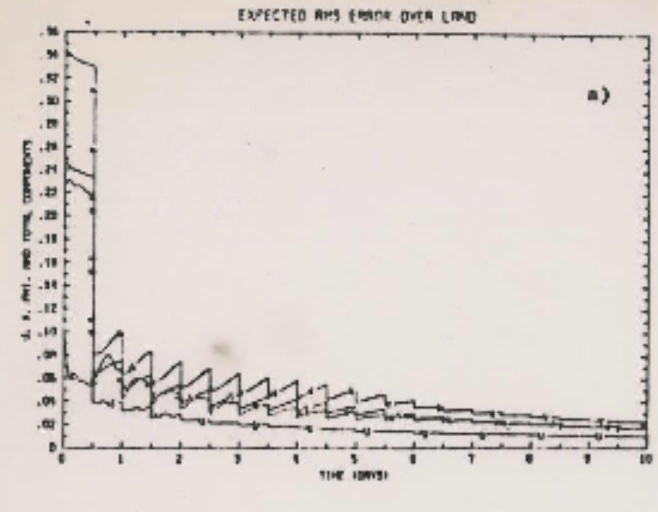
'Ocean'

(no observation)

'Continent'

(observations)

292



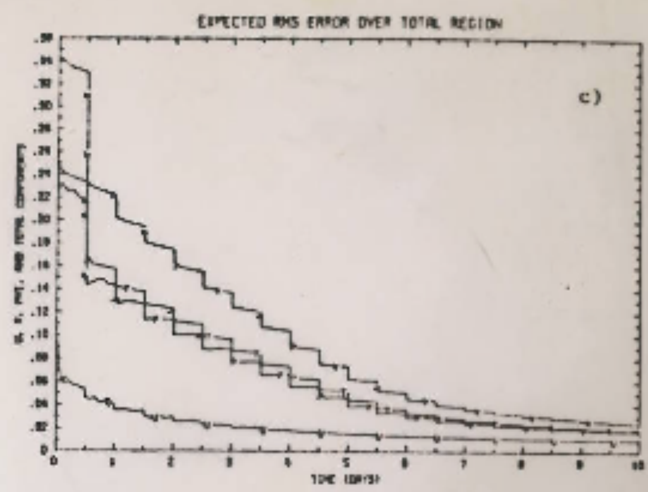
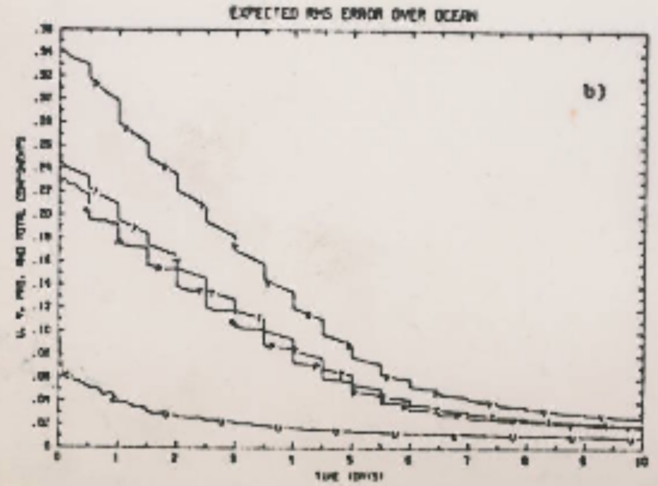


Fig. 2

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.



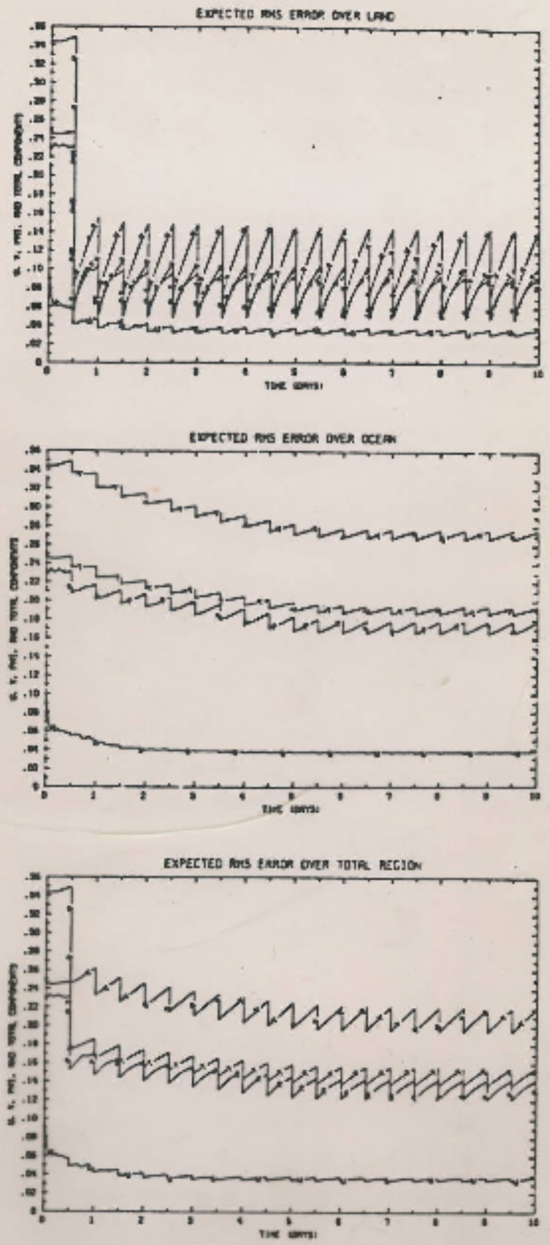


Fig. 6 This figure and the following ones show the properties of the estimated algorithm (2.11) in the presence of system noise, Q ≠ 0. This figure gives the Erms estimation error, and is homologous to Fig. 2. Notice the sharper increase of error over land between synoptic times, and the convergence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

Uncertainty evolves in time under the effect of

- Introduction of observations (decreases uncertainty)

- Model error (increases uncertainty)

- Dynamics of the system (increases or decreases uncertainty depending on stability of the state of the system) (dynamics is neutral in previous example)

Nonlinearities ?

Linearity of observation and model operators have been explicitly used in

$$\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} - \boldsymbol{H}\boldsymbol{x}^b = \boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}^b) + \boldsymbol{\varepsilon} = - \boldsymbol{H}\boldsymbol{\zeta}^b + \boldsymbol{\varepsilon}$$

$$\boldsymbol{M}_k \boldsymbol{x}^a{}_k - \boldsymbol{M}_k\boldsymbol{x}_k = \boldsymbol{M}_k(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)$$

If $\boldsymbol{H}$ nonlinear, and $\boldsymbol{x} - \boldsymbol{x}^b$ small

$\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{x}^b) \approx \boldsymbol{H}'(\boldsymbol{x} - \boldsymbol{x}^b)$

where $\boldsymbol{H}'$ is *Jacobian* matrix of $\boldsymbol{H}$ (matrix of partial derivatives) at point $\boldsymbol{x}^b$

Similarly, if $\boldsymbol{M}_k$ nonlinear, and $\boldsymbol{x}^a{}_k - \boldsymbol{x}_k$ small

$\boldsymbol{M}_k(\boldsymbol{x}^a{}_k) - \boldsymbol{M}_k(\boldsymbol{x}) = \boldsymbol{M}_k'(\boldsymbol{x}^a{}_k - \boldsymbol{x}_k)$

where $\boldsymbol{M}_k'$ is Jacobian matrix of $\boldsymbol{M}_k$ at point $\boldsymbol{x}^a{}_k$

Tangent Linear Approximation

Nonlinearities ?

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

- Analysis step

$$\boldsymbol{x}^a_k = \boldsymbol{x}^b_k + \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} [\boldsymbol{H}_k' \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} [\boldsymbol{y}_k - \boldsymbol{H}_k(\boldsymbol{x}^b_k)]$$

$$\boldsymbol{P}^a_k = \boldsymbol{P}^b_k - \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} [\boldsymbol{H}_k' \boldsymbol{P}^b_k \boldsymbol{H}_k'^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \boldsymbol{H}_k' \boldsymbol{P}^b_k$$

- Forecast step

$$\boldsymbol{x}^b_{k+1} = \boldsymbol{M}_k(\boldsymbol{x}^a_k)$$

$$\boldsymbol{P}^b_{k+1} = \boldsymbol{M}_k' \boldsymbol{P}^a_k \boldsymbol{M}_k'^{\mathrm{T}} + \boldsymbol{Q}_k$$

Extended Kalman Filter (EKF, heuristic !)

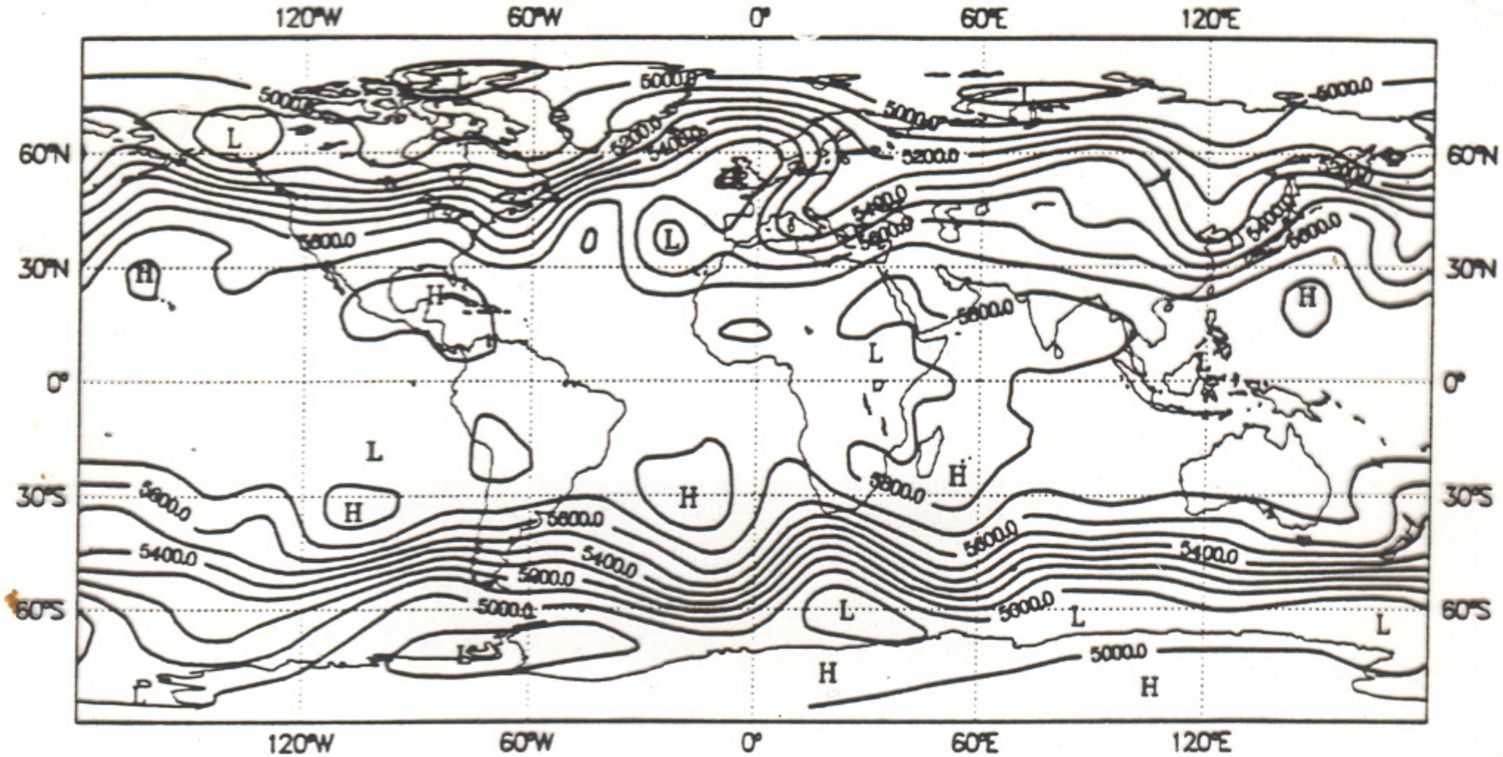
Costliest part of computation

$$\boldsymbol{P}^b{}_{k+1} = \boldsymbol{M}_k \boldsymbol{P}^a{}_k \boldsymbol{M}_k^{\mathrm{T}} + \boldsymbol{Q}_k$$

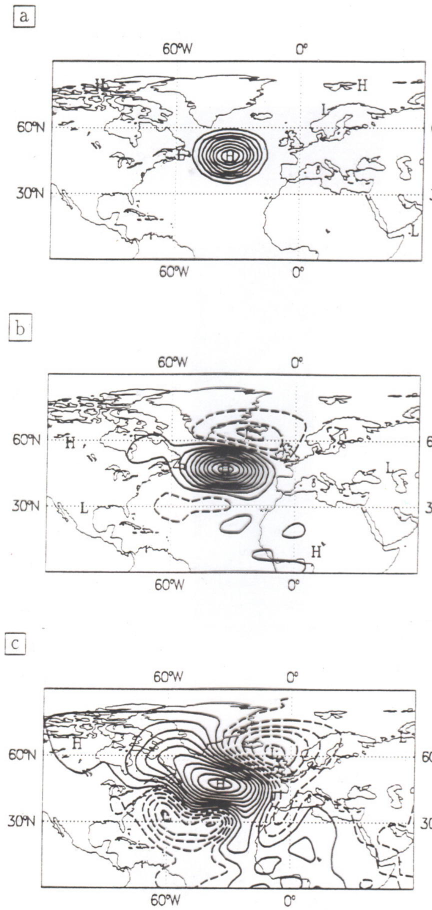
Multiplication of one vector by $\boldsymbol{M}_k$ = one integration of the model between times k and $k+1$

Computation of $\boldsymbol{M}_k \boldsymbol{P}^a{}_k \boldsymbol{M}_k^{\mathrm{T}} \approx 2n$ 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.

Two solutions :

- *Low-rank filters*

  Use low-rank covariance matrix, restricted to modes in state space on which it is known, or at least assumed, that a large part of the uncertainty is concentrated (this requires the definition of a norm on state space).

Reduced Rank Square Root Filters (*RRSQRT*, Heemink)

Singular Evolutive Extended Kalman Filter (*SEEK*, Pham)

….

Reduced Rank Square Root Kalman Filter (***RRSQRT***, Verlaan and Heemink, 1997)

A covariance matrix P can be written as

$$P = S\,S^{\mathrm{T}}$$

where the column vectors of S are the (orthogonal) principal components (eigenvectors) of P (the modulus of each vector is the square root of the associated eigenvalue).

The principle of *RRSQRT* is to restrict the background error covariance matrix P^b to $r \ll n$ principal components, thereby approximating P^b by (the time index k is dropped)

$$P^b \approx S^b\,S^{b\mathrm{T}}$$

where S^b has dimensions n x r.

RRSQRT (continuation 1)

Setting $\Psi \equiv (HS^b)^T$, the gain matrix of the Kalman filter and the analysis error covariance matrix respectively become

$$K = S^b \, \Psi (\Psi^T \Psi + R)^{-1}$$

and

$$P^a = S^a \, S^{aT}$$

with

$$S^a = S^b \, [I_r - \Psi (\Psi^T \Psi + R)^{-1} \Psi^T]^{1/2}$$

RRSQRT (continuation 2)

In the prediction phase, the column vectors of S^a are evolved by the tangent linear model (an evolution of a perturbed state by the full model is also possible). If a model error is to be introduced, that is done by reducing the order r of S^a to r-q, and introducing q new column vectors meant to represent the model error.

Orthogonality of the column vectors is lost in the prediction, and has to be reestablished. And, even if process is started from dominant column vectors, that dominance may of course be lost.

Advantages : in addition to reduced computational cost, numerical errors are smaller when dealing with square root covariance matrices, as done here, than with full matrices (better conditioning).

Singular Evolutive Extended Kalman Filter (***SEEK***, Pham, 1996)

Based on the fact that, because of the linearity of Kalman Filter, the rank of the covariance matrix P^a or P^b cannot increase in either the update or the model evolution. SEEK performs a linear filter starting from a low rank ${P^b}_0$, and so runs the exact Kalman filter in the case of a perfect model. The algorithmic implementation takes advantage of the rank-deficiency of the covariance matrix. The rank of the latter is conserved (or decreased), but the subspace spanned by the directions with non-zero error evolves, in both the update and the dynamic evolution.

In case model error is present, corresponding covariance matrix Q_k is projected onto the directions with non-zero error (this is of course an approximation).

Singular Evolutive Interpolated Kalman Filter (***SEIK***, Pham, 2001)

Non-trivial extension of SEEK to nonlinear model or observation operators. Rank deficiency is now forced.

Second solution :

- *Ensemble filters*

  Uncertainty is represented, 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 : $\{\boldsymbol{x}^b{}_l\}$, $l = 1, \ldots, L$

Observation vector at same time : $\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{\varepsilon}$

- Gaussian approach

Produce sample of probability distribution for real observed quantity $\boldsymbol{Hx}$

$\boldsymbol{y}_l = \boldsymbol{y} - \boldsymbol{\varepsilon}_l$

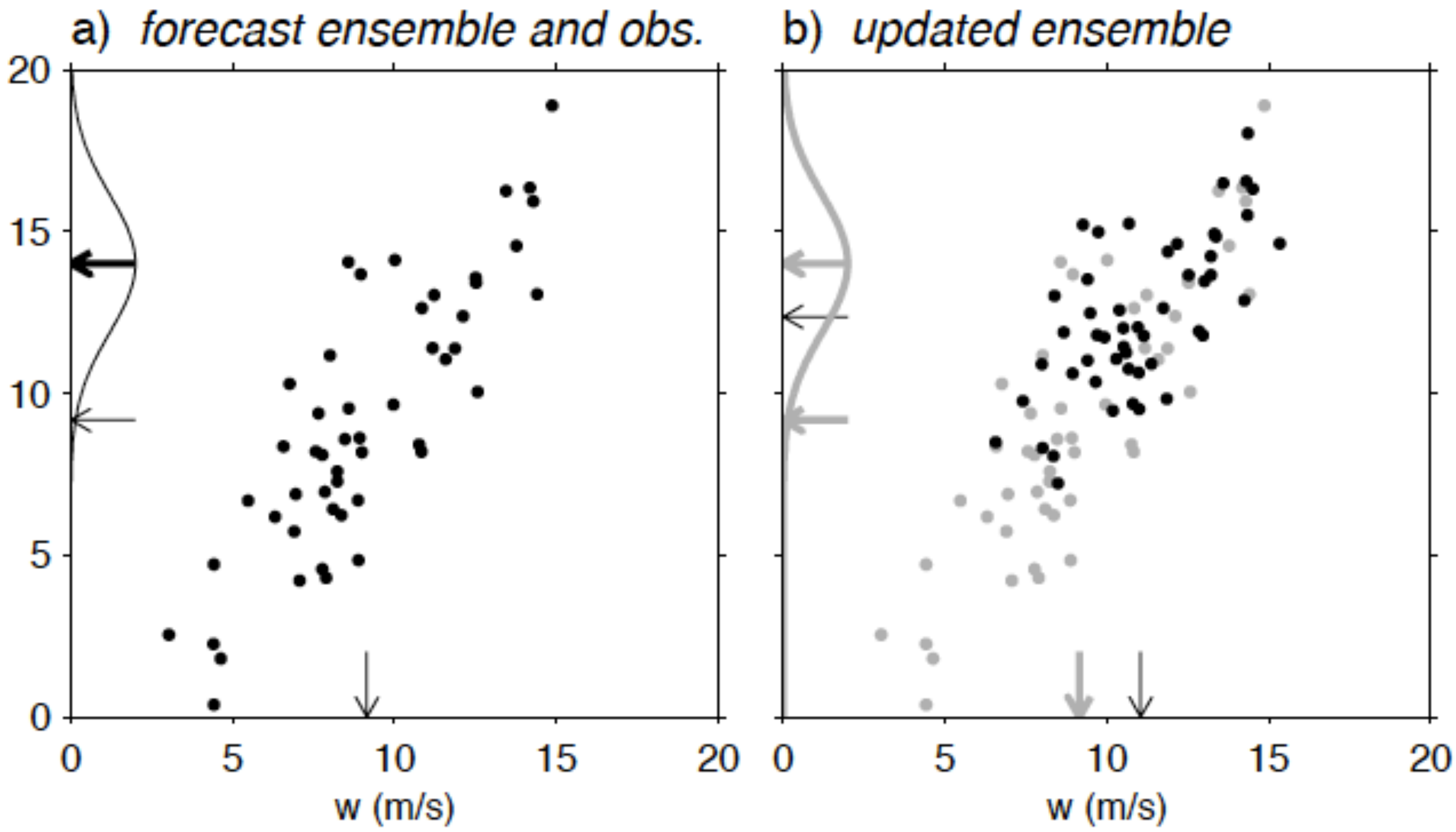
where $\boldsymbol{\varepsilon}_l$ is distributed according to probability distribution for observation error $\boldsymbol{\varepsilon}$.

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

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

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

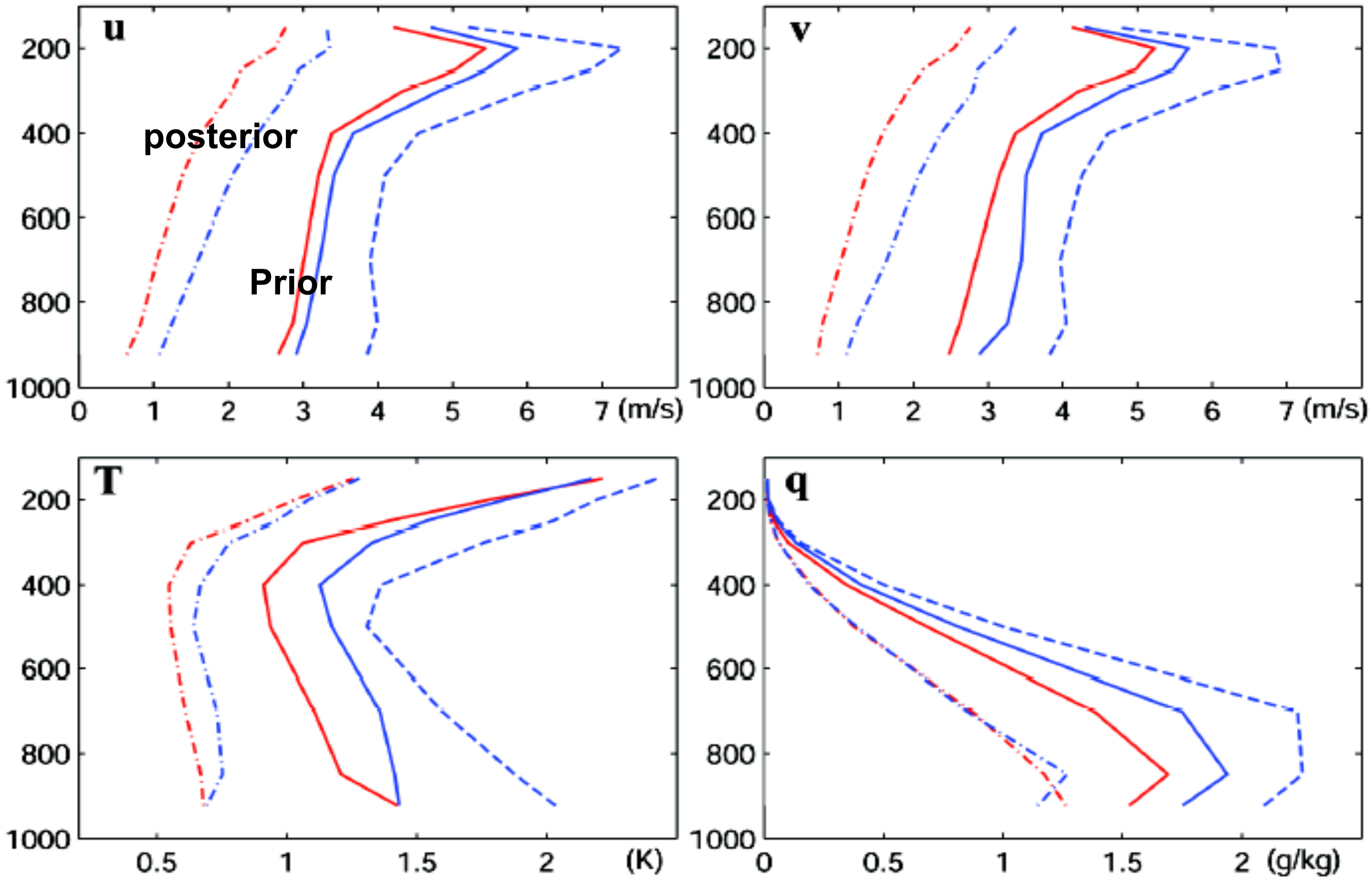
Remark. In case of Gaussian errors, if $\boldsymbol{P}^b$ was exact covariance matrix of background error, (2) would achieve Bayesian estimation, in the sense that $\{\boldsymbol{x}^a{}_l\}$ would be a sample of conditional probability distribution for $\boldsymbol{x}$, given all data up to time k.


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

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 : $\boldsymbol{y} = \boldsymbol{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon}$ $\boldsymbol{H}$ nonlinear

Two possibilities

1. Take tangent linear approximation (as in Extended KF) and introduce jacobian $\boldsymbol{H}'$

2. Come back to original formula

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

That formula does not require any other link between $\boldsymbol{x}$ and $\boldsymbol{y}$ than the one defined by the covariances matrices $\boldsymbol{C}_{xy}$ and $\boldsymbol{C}_{yy}$.

Here, as shown on the occasion of the derivation of the *BLUE*, $E(\boldsymbol{x})$ is the backgound $\boldsymbol{x}^b$, and $\boldsymbol{y} - E(\boldsymbol{y})$ is the innovation $\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^b)$

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

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Collapse originates in the fact that gain matrix $\boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1}$ is nonlinear wrt background error matrix $\boldsymbol{P}^b$, resulting in a systematic sampling effect. Solution : empirical 'covariance inflation'.

- Spurious correlations appear at large geographical distances. Empirical 'localization' (see Gaspari and Cohn, 1999, *Q. J. R. Meteorol. Soc.*)

- In formula

$$\boldsymbol{x}^a_l = \boldsymbol{x}^b_l + \boldsymbol{P}^b \boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} (\boldsymbol{y}_l - \boldsymbol{H}\boldsymbol{x}^b_l) , \qquad l = 1, \ldots, L$$

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

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

There exist many variants of Ensemble Kalman Filter

Ensemble Transform Kalman Filter (***ETKF***, Bishop *et al.*, *Mon. Wea. Rev.*, 2001)

Requires a prior 'control' analysis x_c^a, emanating from a background x_c^b. An ensemble is evolved about that control without explicit use of the observations (and without feedback to control)

More precisely, define L x L matrix T such that, given $P^b = ZZ^T$, then $P^a = ZTT^TZ^T$ (not trivial, but possible). Then the background deviations $x^b_l - x_c^b$ are transformed through $Z \rightarrow ZT$ into an ensemble of analysis deviations $x^a_l - x_c^a$.

(does not avoid collapse of ensembles)

Local Ensemble Transform Kalman Filter (***LETKF***, Hunt *et al.*, *Physica D*, 2007)

Each gridpoint is corrected only through the use of neighbouring observations.

Other variants of Ensemble Kalman Filter

'*Unscented*' *Kalman Filter* (Wan and van der Merve, 2001, Wiley Publishing)

Weighted Kalman Filter (Papadakis *et al*., 2010, *Tellus A*)

Inflation-free Ensemble Kalman Filters (Bocquet and Sakov, 2012, *Nonlin. Processes Geophys*.)

An iterative ensemble Kalman filter in the presence of additive model error (Sakov *et al*., 2017, *Q. J. R. Meteorol. Soc.)*

Bayesian properties of Ensemble Kalman Filter ?

Very little is known.

Le Gland *et al*. (2011). In the linear and gaussian case, the discrete pdf defined by the filter, in the limit of infinite sample size L, tends to the bayesian gaussian pdf.

No result for finite size (note that ensemble elements are not mutually independent)

In the nonlinear case, the discrete pdf tends to a limit which is in general not the bayesian pdf.

Situation still not entirely clear

Time-correlated Errors

Example of time-correlated observation errors

$z_1 = x + \zeta_1$

$z_2 = x + \zeta_2$

$E(\zeta_1) = E(\zeta_2) = 0$; $E(\zeta_1{}^2) = E(\zeta_2{}^2) = s$; $E(\zeta_1\zeta_2) = 0$

BLUE of x from z_1 and z_2 gives equal weights to z_1 and z_2. The weights given to z_1 and z_2. will remain equal in sequential assimilation in the successive background and analyzed estimates x^b and x^a

Additional observation then becomes available

$z_3 = x + \zeta_3$

$E(\zeta_3) = 0$; $E(\zeta_3{}^2) = s$; $E(\zeta_1\zeta_3) = cs$; $E(\zeta_2\zeta_3) = 0$

BLUE of x from (z_1, z_2, z_3) has weights in the proportion $(1, 1{+}c, 1)$

Time-correlated Errors (continuation 1)

Example of time-correlated model errors

Evolution equation

` $x_{k+1} = x_k + \eta_k$ $E(\eta_k^2) = q$

Observations

$y_k = x_k + \varepsilon_k$, $k = 0, 1, 2$ $E(\varepsilon_k^2) = r$, errors uncorrelated in time

Sequential assimilation. Weights given to y_0 and y_1 in analysis at time 1 are in the ratio $r/(r+q)$. That ratio will be conserved in sequential assimilation. All right if model errors are uncorrelated in time.

Assume $E(\eta_0\eta_1) = cq$

Weights given to y_0 and y_1 in estimation of x_2 are in the ratio

$$\rho = \frac{r - qc}{r + q + qc}$$

Conclusion

Sequential assimilation, in which data are processed by batches, the data of one batch being discarded once that batch has been used, cannot be optimal if data in different batches are affected with correlated errors. ***This is so even if one keeps trace of the correlations.***

Solution

Process all correlated in the same batch (4DVar, some smoothers)

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.

Kalman smoother

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

Various possibilities

- Define new state vector $x^T \equiv (x_0^T, \ldots, x_K^T)$
  and use Kalman formula from a background x^b and associated covariance matrix Π^b.
  'Observation operator' must include the model equations
  Can take into account temporal correlations

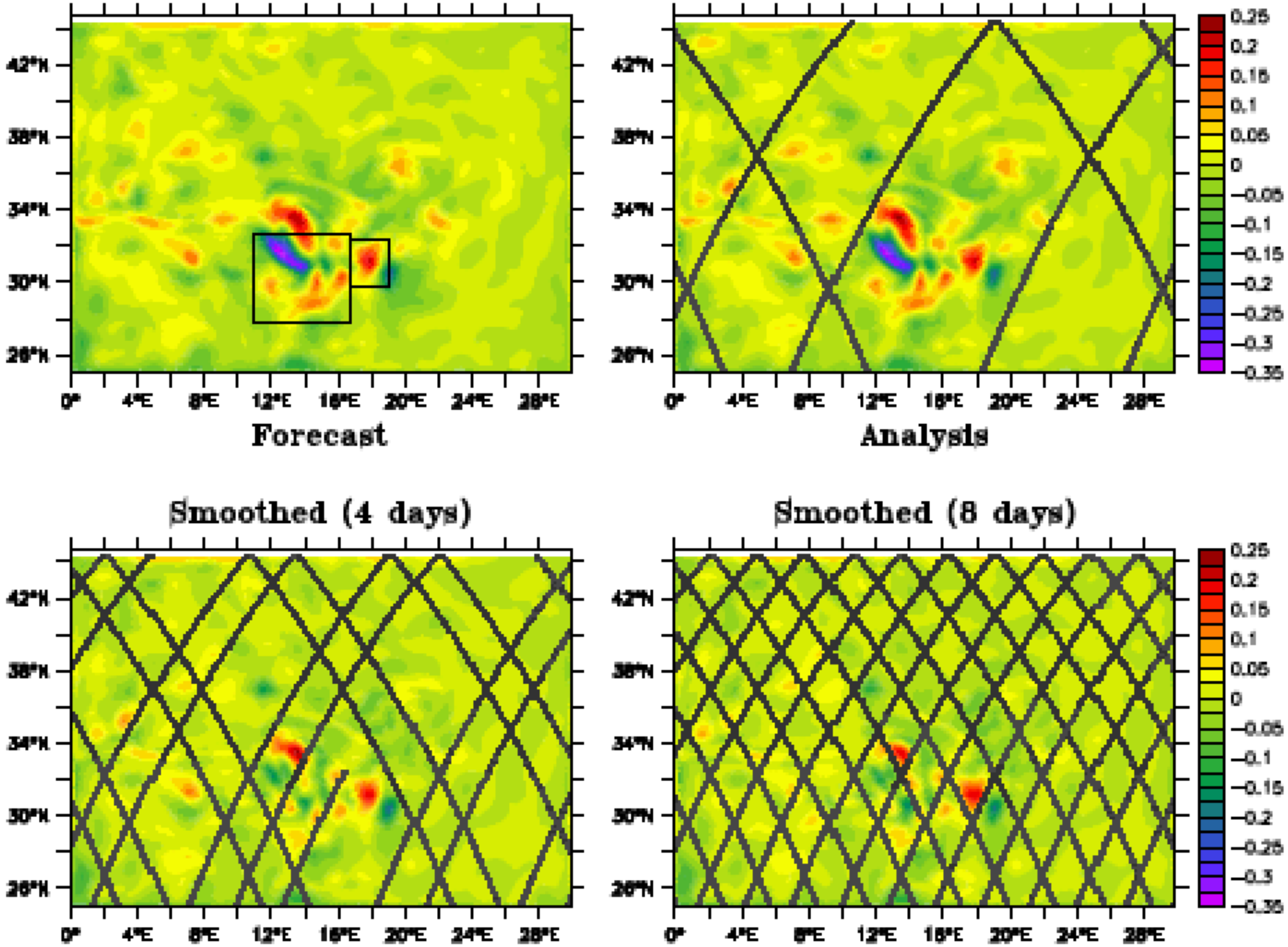
- Update sequentially vector $(x_0^T, \ldots, x_k^T)^T$ for increasing k
  Cannot take into account temporal correlations

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.

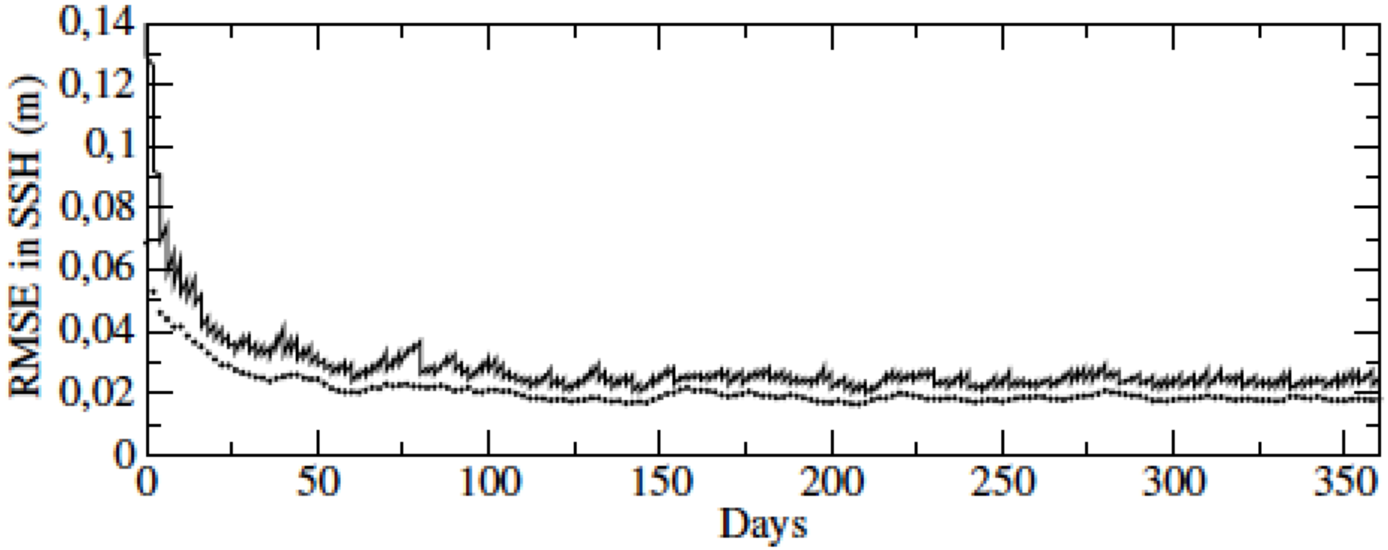


FIGURE 3.7 – Evolution de l'erreur RMS de SSH au cours du temps. Ligne continue : Résultat du filtre (les dents de scie reflètent l'alternance des étapes de prévision et d'analyse) ; Points : lisseur à retard fixe de 8 jours.

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)*

Best Linear Unbiased Estimate

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

$$\boldsymbol{x}^b = \boldsymbol{x} + \boldsymbol{\zeta}^b \qquad E(\boldsymbol{\zeta}^b \boldsymbol{\zeta}^{b\mathrm{T}}) \equiv \boldsymbol{P}^b \quad \dim \boldsymbol{P}^b = n \times n$$

Observation vector $\boldsymbol{y}$, belonging to *observation space* $\mathcal{O}$ (dim $\mathcal{O} = p$)

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon} \qquad E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) \equiv \boldsymbol{R} \quad \dim \boldsymbol{R} = p \times p$$

$\boldsymbol{H}$ linear operator from $\mathcal{S}$ into $\mathcal{O}$ $\quad$ dim $\boldsymbol{H} = p \times n$

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

$$\mathcal{S} \leftarrow \mathcal{S}^* \leftarrow \mathcal{O}^* \leftarrow \mathcal{O}$$

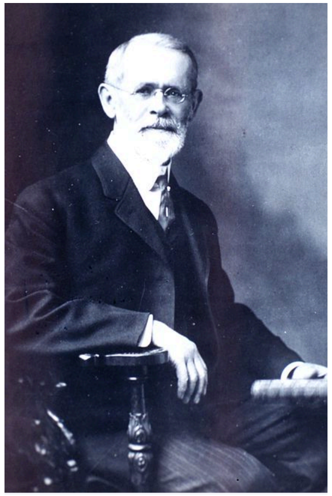
Alternative form of gain matrix

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

structure is the same

History of Numerical Weather Prediction

Cleveland Abbe



The Physical Basis of Long Range Weather Forecasts, 1901,

Monthly Weather Review

Wilhelm Bjerknes

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

V. Bjerknes at the origin of the 'Bergen School of Meteorology'

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

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

- 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

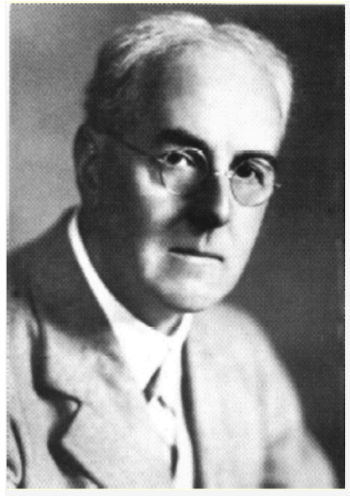
History of Numerical Weather Prediction (continuation)

Lewis Fry Richardson

Weather Prediction by Numerical Process, 1922

Cambridge University Press *

Forecast Factory



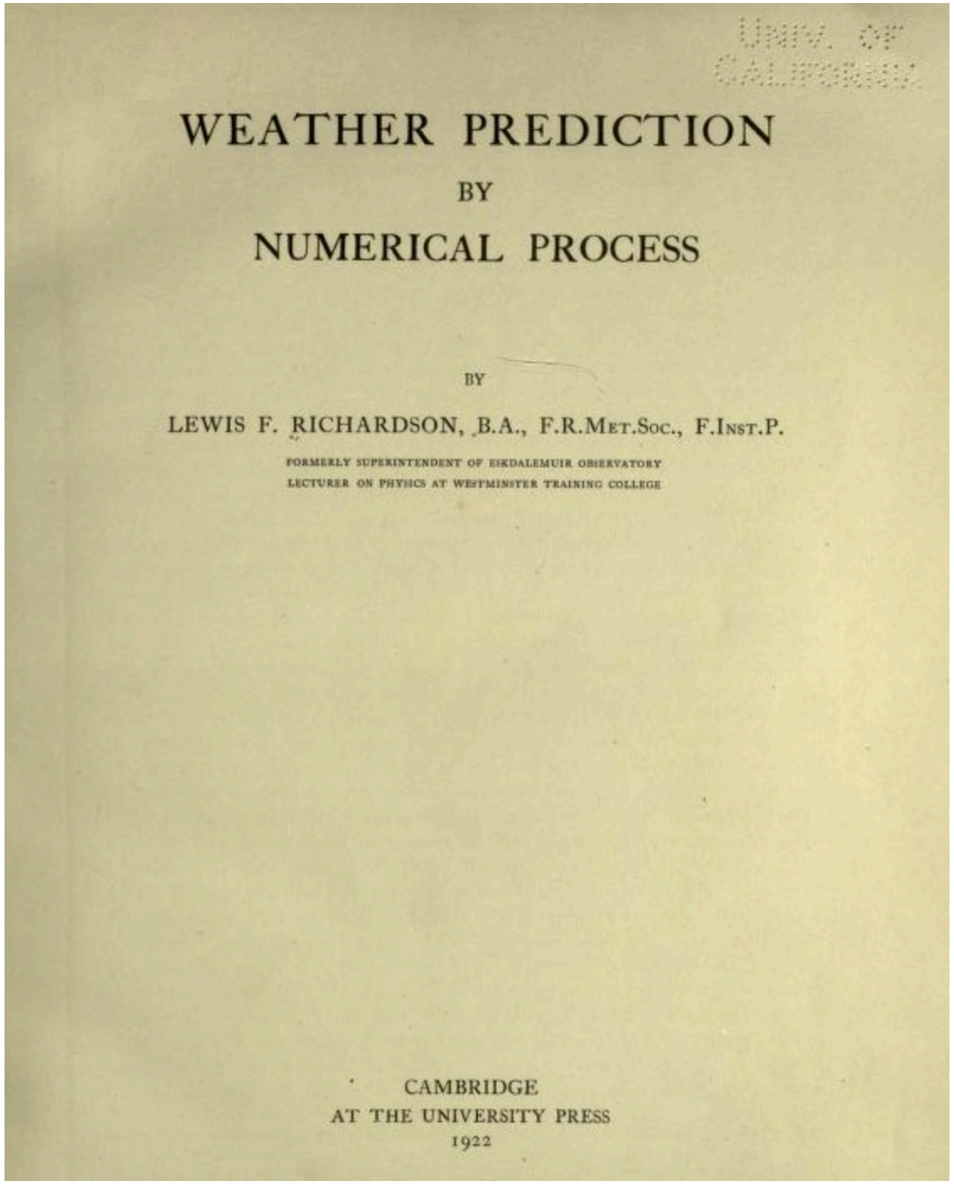
Richardson number, fractals, pacifism

* 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

History of Numerical Weather Prediction (continuation 2)

https://energy4climate.pages.in2p3.fr/public/education/ensemble_data_assimilation_tutorial/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 3)

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, …)

History of Numerical Weather Prediction (continuation 4)

Institute for Advanced Study, about 1948-50. J. von Neumann is second from left, J. Charney first on right (R. Fjørtoft third from right ?)

History of Numerical Weather Prediction (continuation 5)


Jule Gregory Charney en 1978.

Charney developed vorticity barotropic model

First simulation of real atmospheric situation in 1950

First operational numerical forecast performed in 1954 in Sweden

(C. G. Rossby)

History of Numerical Weather Prediction (continuation 6)

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

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

Ensemble prediction

History of Numerical Weather Prediction (continuation 7)

Extension to simulation of oceanic circulation and climate

(early 1970's, S. Manabe and K. Bryan, GFDL).

Climatic simulations (S. Manabe, R. Wetherald)

S. Manabe awarded Nobel Prize in Physics in 2021

History of Numerical Weather Prediction (continuation 8)

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 in 2021 and 2022

Cours à venir

~~Jeudi 17 mars~~
~~Jeudi 24 mars~~
~~Jeudi 31 mars~~
~~Jeudi 14 avril~~
Jeudi 21 avril
Jeudi 28 avril
Jeudi 5 mai
Jeudi 12 mai