A Brief History of Data Assimilation

It all Started from a Rather Menial Task

O. Talagrand

Laboratoire de Météorologie Dynamique, École Normale Supérieure, Paris, France

International Symposium on Data Assimilation – Online

11 February 2022

After the well-known (and unsuccessful, but nevertheless very instructive) attempt by L. F. Richardson in 1922, Numerical Weather Prediction started for good in the late 40s, at the Institute for Advanced Study in Princeton, under the leadership of John von Neumann (in a group comprising, among others, people such as J. Charney, N. A. Phillips, R. Fjörtoff and J. Smagorinsky)

Assimilation of observations, as it is known in meteorology and oceanography, originated from the need of defining initial conditions (ICs) for numerical weather prediction. That was considered at the start as a minor task, the ‘real’ thing being of course the forecast itself. Difficulties gradually arose

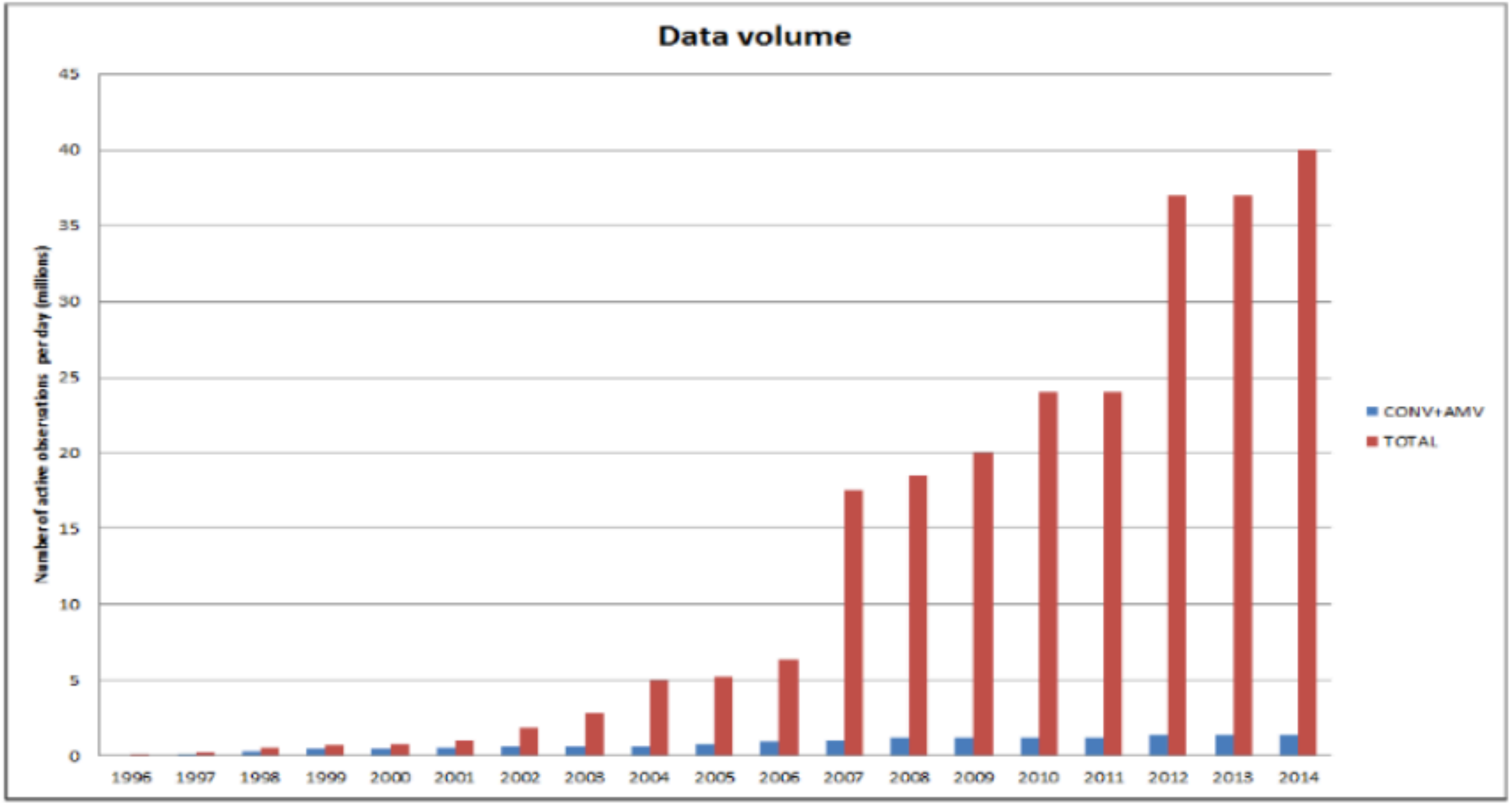
- Need for defining ICs with appropriate spatial scales ⇒ ‘*structure functions*‘ (now incorporated in background error covariance matrices)

- Need for defining ICs in approximate geostrophic balance ⇒ ‘*initialization’* (now also incorporated in background error covariance matrices)

- Realization that useful information was present in observations, particularly satellite observations, that were distributed over time ⇒ need to introduce the dynamical model in the process of definition of the ICs. Word *assimilation* was coined in 1967-68.

- Satellite observations, in addition to being distributed continuously in time, are ‘indirect’ ⇒ need for some form of ‘inversion’

- Very significant increase of the number of observations to be processed ⇒ need for powerful and robust algorithms



European Centre for Medium-Range Weather Forecasts (ECMWF, 2020)
Around 10 million observations are used per 6 h period

Proportion of resources devoted to assimilation in Numerical Weather Prediction has steadily increased over time. At ECMWF, about 40% of resources devoted to operational chain is now devoted to assimilation.

Originally, *analysis* of meteorological situation was perfomed by an experienced meteorologist on a chart.

When computers became available, they were used to perform the same task that was performed by hand. The main advantage was that the procedure was entirely automated, and could process rapidly a large amount of data. People spoke of *objective analysis,* by opposition to subjective analysis performed by individuals.

Principle : apply corrections to a first-guess field, by extrapolating differences between first guess and observations (Bergthórssen and Döös, 1955, Cressman, 1959)

One method gained wide acceptance

Optimal Interpolation (*OI*) (Gandin, 1963)

Observations y_j , $j = 1, \ldots, p$

Estimate of unknown quantity x looked for in the form

$$x^a = \alpha + \Sigma_j \beta_j y_j$$

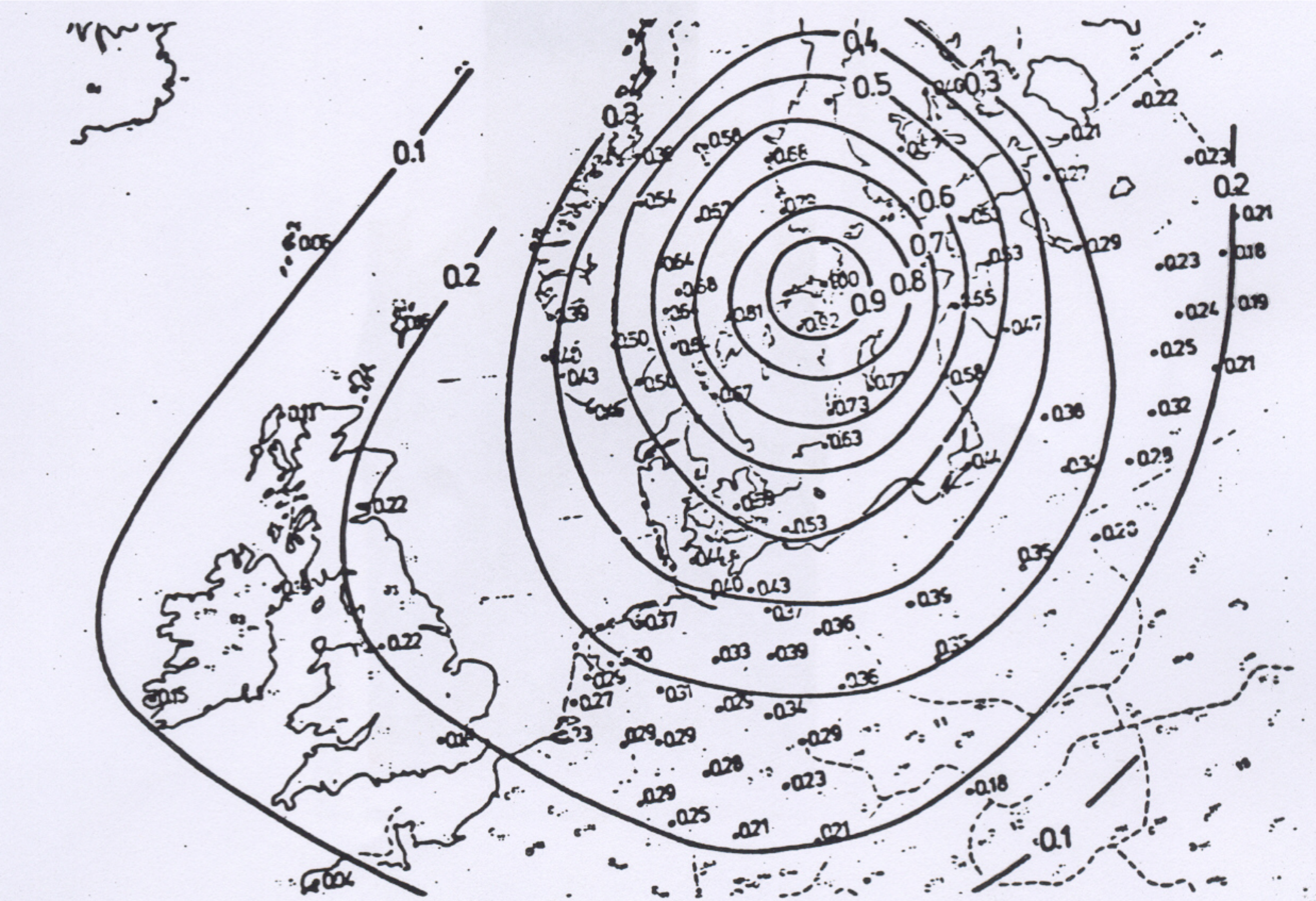
α and the β_j's being determined so as to minimize the expected quadratic estimation error $E[(x\text{-}x^a)^2]$

Requires the *a priori* explicit knowledge of first-order (expectations) and second-order (variances and covariances) statistical moments of all variables.

Expectations of quantities to be estimated make up *background* (or *first-guess*) to be updated by observations.

Each observation comes with its own influence function, or *representer* (its statistical covariance with variables to be estimated).

The correction performed on the background is a linear combination of all representers, with weights depending on values of observations, and on mutual covariances between observations.



<u>Figure 5.1.1.4.1</u> Auto-correlation of errors in 12h numerical fore-casts of surface pressure in a reference station (Stockholm) and other stations.

After N. Gustafsson

Schlatter's (1975) multivariate covariances

Specified as multivariate 2-point functions.

Not easy to ensure that specified functions are actually valid covariances.

Used in OI and related observation-space methods.

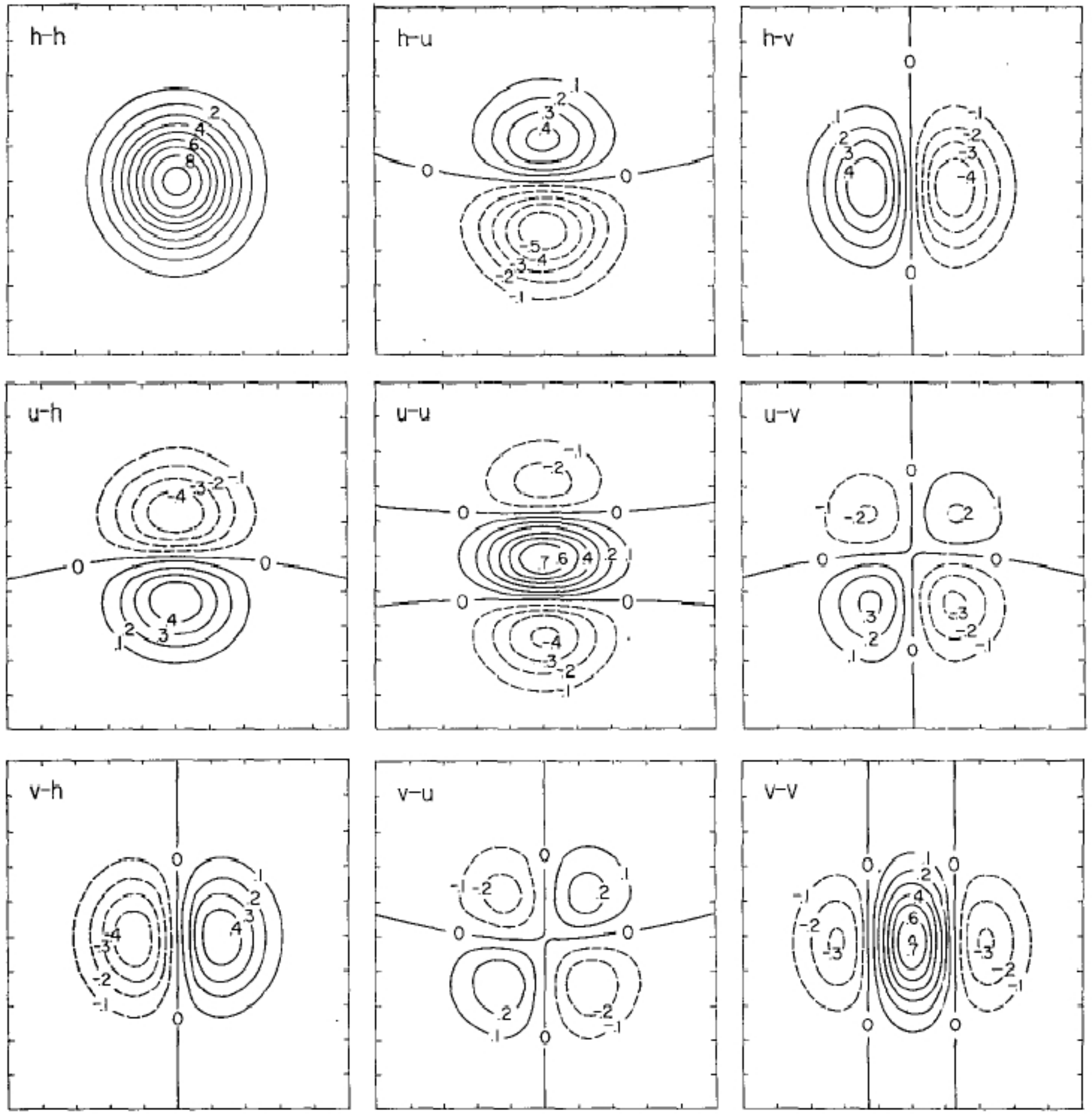


FIG. 3. Correlations among the variables h, u, and v based upon the expression $\mu = 0.95 \exp(-1.24s^2)$ for height-height correlation and the geostrophic relations. Diagrams centered at 110°W, 35°N. Tick marks 500 km apart.

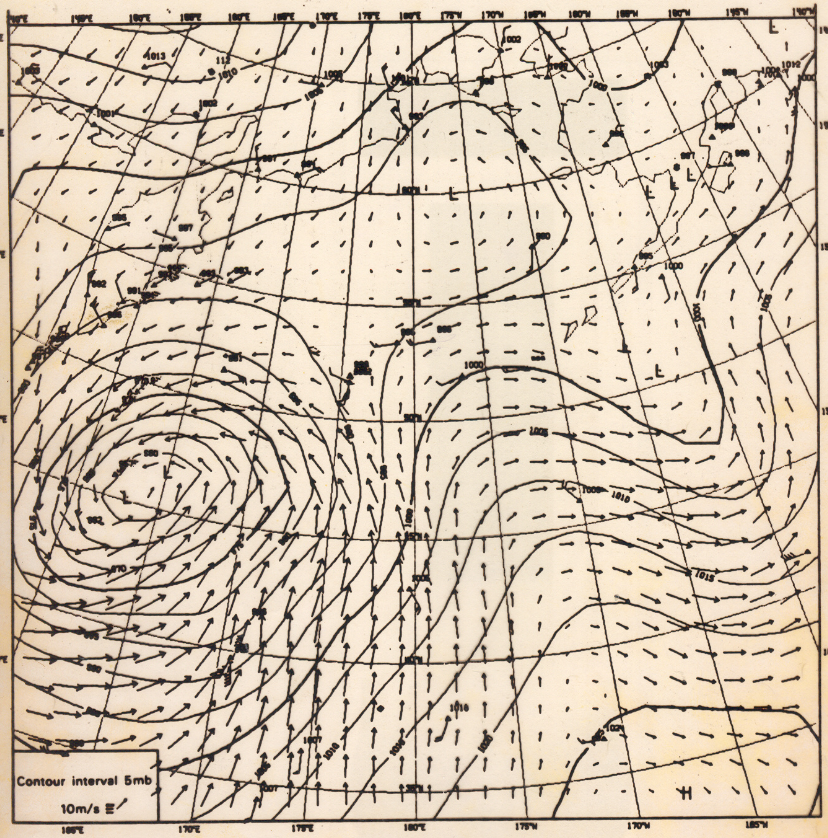


Fig. 4 Sea level pressure and wind forecast corresponding to the central area of Fig. 1, with plotted surface observations of pressure and wind (each fleche = 5 m/s).

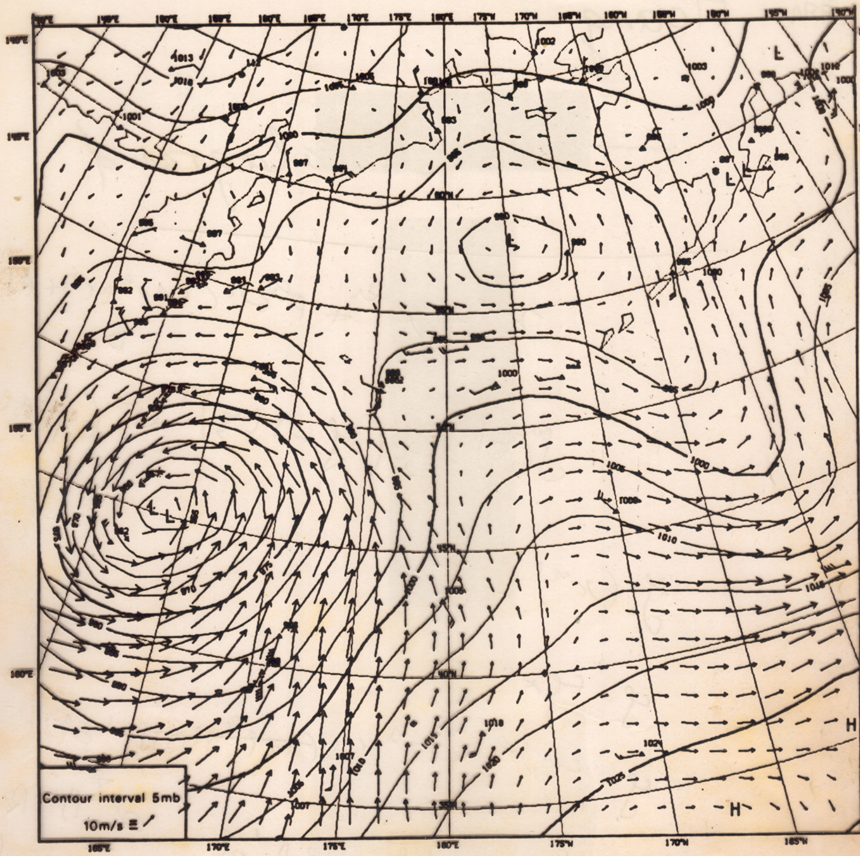


Fig. 5 As Fig. 4 for the analysis in the data assimilation cycle

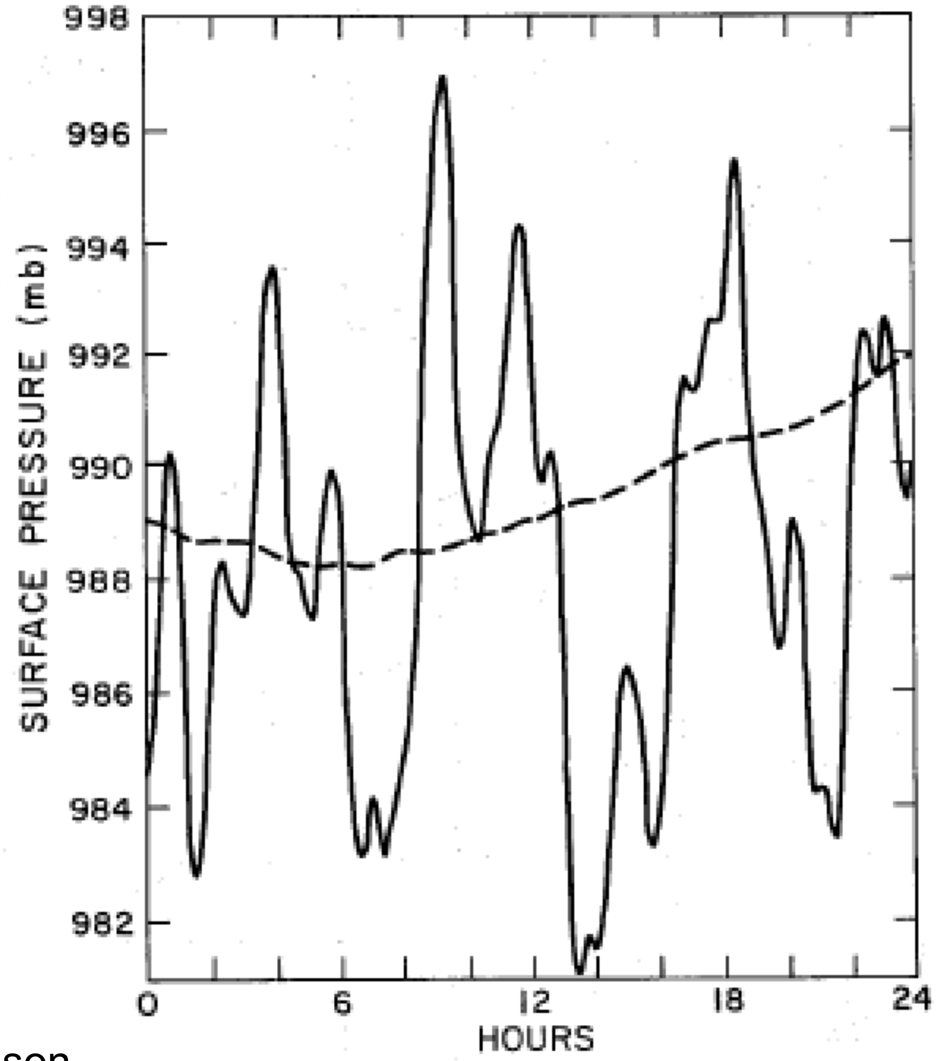
After A. Lorenc, MWR, 1981

In what is called in practice optimal interpolation, the background, usually denoted $\boldsymbol{x}^b$, is a recent forecast coming from the past. The associated covariance matrix that is required by the algorithm is usually taken as a climatological estimate (or an estimate varying with season, but not on the current estimate of the flow). It is denoted $\boldsymbol{P}^b$ (or simply $\boldsymbol{B}$).

Optimal interpolation has been abundantly used (and is still in many applications), and was the standard method of assimilation of meteorological observations for a number of years.

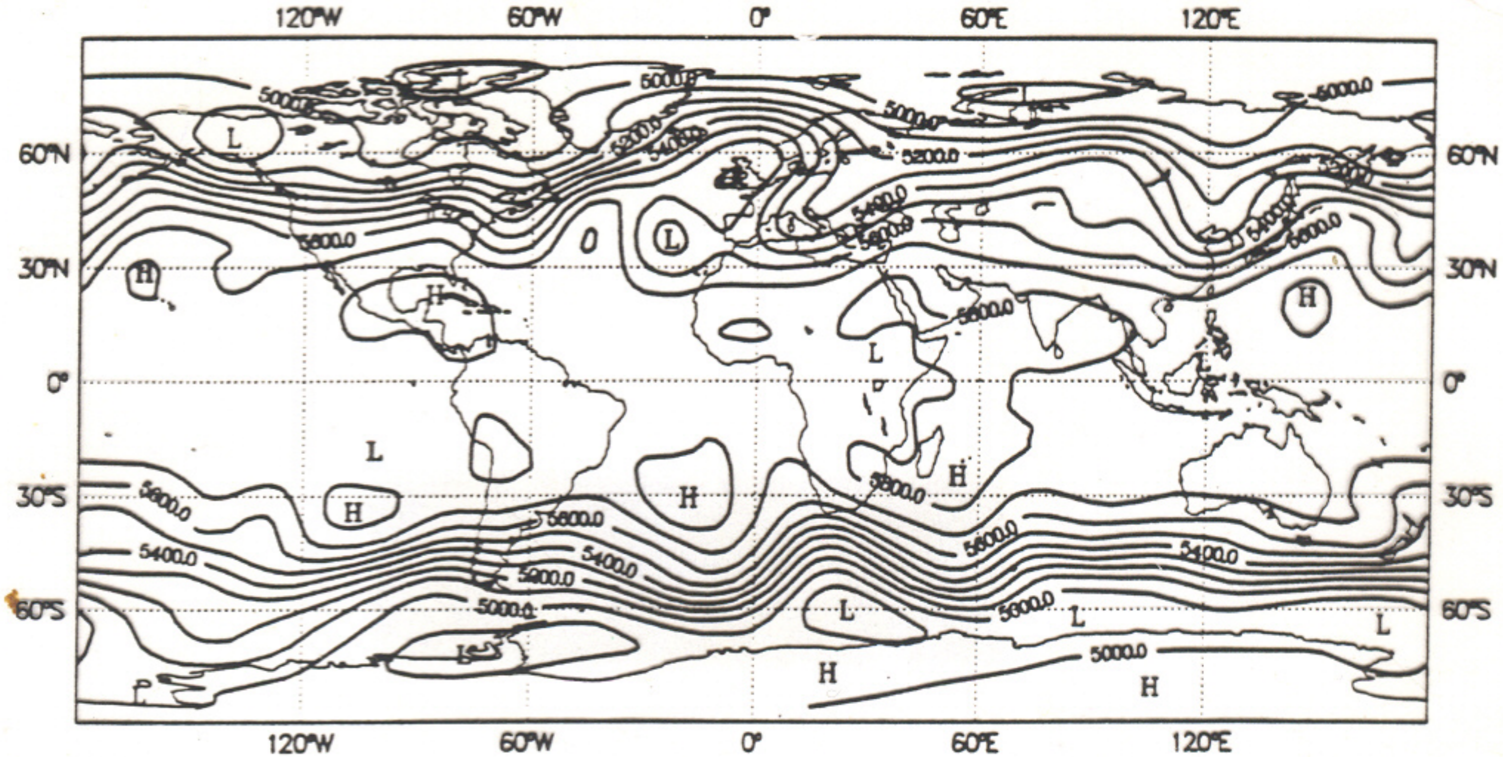
In later years, the analysis $\boldsymbol{x}^a$ has been obtained through minimization of a scalar objective function. It is then called *3DVar*.

‘*Initialization*’

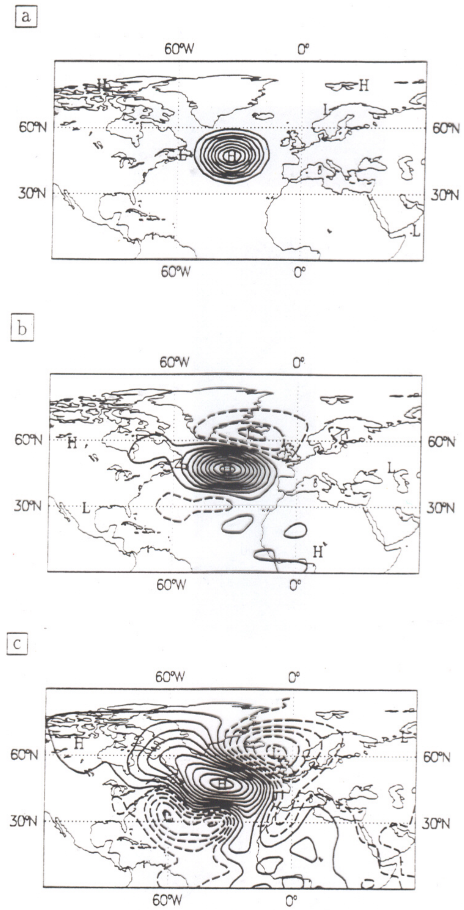


Temperton and Williamson, ECMWF, 1979

Surface pressure vs time at 40 °N, 90 °W during forecasts with new Jacobian, before (solid) and after (dashed) nonlinear initialization (2 iterations, 5 vertical modes, using horizontal modes based on old Jacobian).



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.

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

Two classes of algorithms have been, and are still, largely used for that purpose

- *Kalman Filter*, especially in its Ensemble form

- *Variational Assimilation*

Purpose of assimilation : reconstruct as accurately as possible the state of the atmospheric or oceanic flow, using all available appropriate information. The latter essentially consists of

- The observations proper, which vary in nature, resolution and accuracy, and are distributed more or less regularly in space and time.

- The physical laws governing the evolution of the flow, available in practice in the form of a discretized, and necessarily approximate, dynamical model.

- ‘Asymptotic’ properties of the flow, such as, *e*. *g*., geostrophic balance of middle latitudes. Although they basically are necessary consequences of the physical laws which govern the flow, these properties can usefully be explicitly introduced in the assimilation process.

Both observations and ‘model’ are affected with some uncertainty ⇒ uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don’t know too well why, but it works; see, *e.g*. Jaynes, 2007, *Probability Theory: The Logic of Science,* Cambridge University Press).

Assimilation is a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know (see Tarantola, A., 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM).

Assimilation is one of many '*inverse problems*' encountered in many fields of science and technology

- solid Earth geophysics
- plasma physics
- 'nondestructive' probing
- navigation (spacecraft, aircraft, ….)
- …

Solution most often (if not always) based on Bayesian, or probabilistic, estimation. 'Equations' are fundamentally the same.

Difficulties specific to assimilation of meteorological observations :

- Very large numerical dimensions ($n \approx 10^6$-10^9 parameters to be estimated, $p \approx 4$-5.10^7 observations per 24-hour period). Difficulty aggravated in Numerical Weather Prediction by the need for the forecast to be ready in time.

- Non-trivial, actually chaotic, underlying dynamics

If errors are additive and Gaussian, Optimal Interpolation achieves Bayesian estimation.

Bayesian estimation is actually impossible in its general theoretical form in meteorological or oceanographical practice because

- It is impossible to explicitly describe a probability distribution in a space with dimension even as low as $n \approx 10^3$, not to speak of the dimension $n \approx 10^{6\text{-}9}$ of present Numerical Weather Prediction models (the *curse of dimensionality)*.

- Probability distribution of errors on data very poorly known (model errors in particular).

One has to restrict oneself to a much more modest goal. Two approaches have been developed

- Obtain some 'central' estimate of the conditional probability distribution (expectation, mode, …), plus some estimate of the corresponding spread (standard deviations and a number of correlations).

- Produce an ensemble of estimates that are meant to sample the conditional probability distribution (dimension $N \approx O(10\text{-}100)$).

Kalman Filter

- Observation vector at time k

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

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

$\boldsymbol{H}_k$ linear

- Evolution equation

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

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

$\boldsymbol{M}_k$ 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

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

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

- 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, Jones, 1965, Petersen, 1968, Ghil *et al*., 1981, …)

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

In case observation and model errors $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are globally Gaussian, KF produces at any time k the (Gaussian) pdf for $\boldsymbol{x}_k$, conditioned by all data up to that time.

Equation

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

describes temporal evolution, between two observation instants, of uncertainty on the state of the flow. Much too costly for practical implementation.

Two solutions :

- ***Low-rank filters*** (Verlaan and Heemink, 1997, Pham *et al*., 1998, …)

  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.

- ***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, 1994, Houtekamer and Mitchell, 1998, …)

How to update predicted ensemble with new observations ?

Predicted ensemble at time k : $\{\boldsymbol{x}^b_l\}$, $\qquad l = 1, \ldots, L$
Observation vector at same time : $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$

- Gaussian approach

Produce sample of probability distribution for real observed quantity $\boldsymbol{H}\boldsymbol{x}$
$\boldsymbol{y}_l = \boldsymbol{y} - \boldsymbol{\varepsilon}_l$
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

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

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.

But problems

- Collapse of ensemble for small ensemble size (less than a few hundred). Collapse originates in the fact that gain matrix $\boldsymbol{P}^b \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-1}$ is nonlinear wrt background error matrix 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.*)

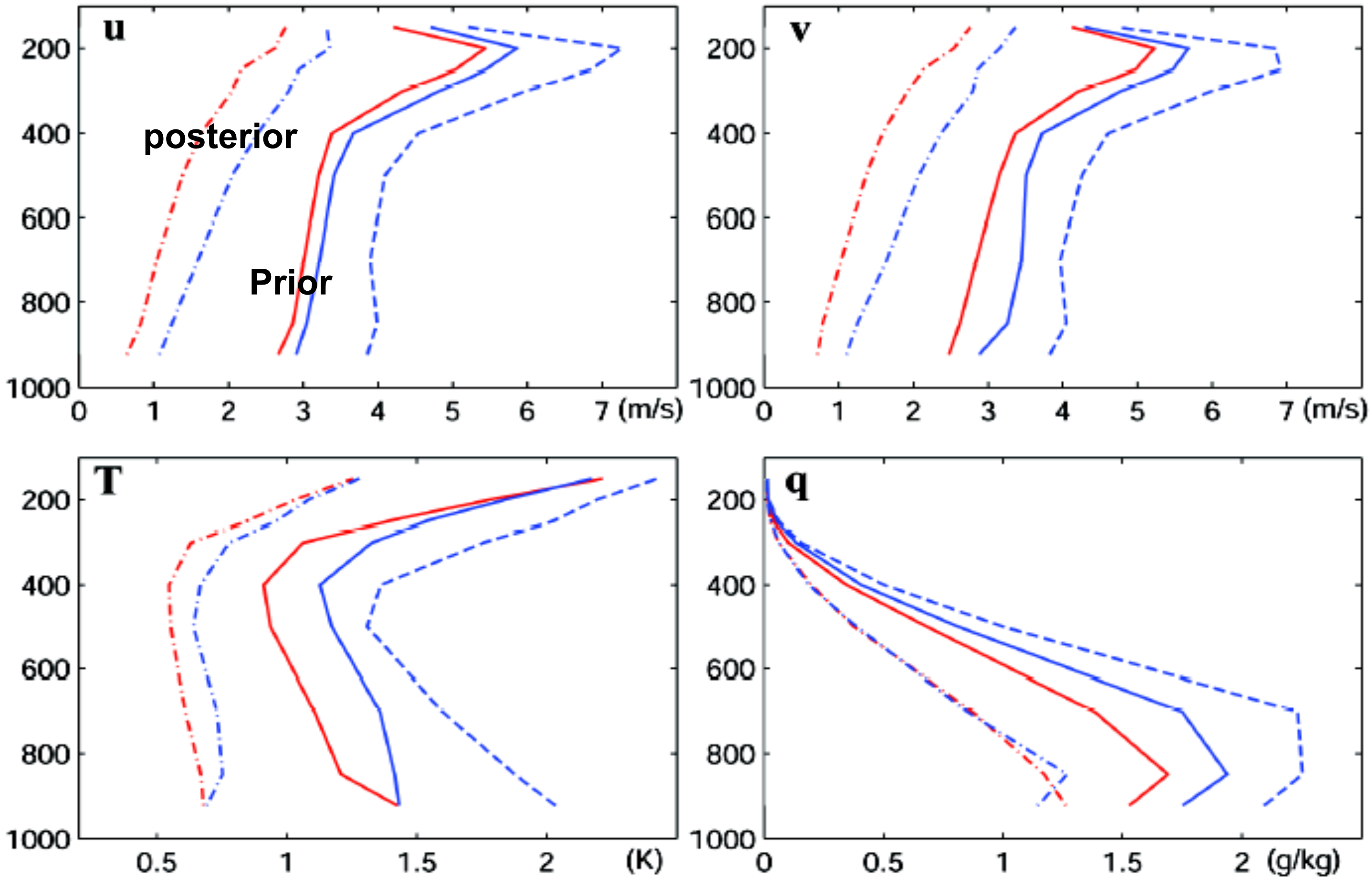
- In formula

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

$\boldsymbol{P}^b$, 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 $L \ll p$, $L \ll n$.

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, MWR, 2008)

Size of ensembles ?

Must be fundamentally sufficient to prevent growth of errors in neutral and unstable modes of the system. Size of ensembles must be at least as large as the number of those neutral and unstable modes (typically from a few tens to a few hundreds for large scale meteorological flow).

Ensemble Kalman Filter exists in many variants and is used for many applications, both in operations and in research.

Variational Assimilation

- 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 $k = 0, \ldots, K$

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

- Model

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

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

Sequence of model states $(\xi_0, \xi_1, ..., \xi_K)$ over observation period

Then scalar objective function

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

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

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

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

measures misfit between sequence $(\xi_0, \xi_1, ..., \xi_K)$ and data

Purpose of variational assimilation : find sequence $(\xi_0, \xi_1, ..., \xi_K)$ which minimizes $\mathcal{J}(\xi_0, \xi_1, ..., \xi_K)$ (*4DVar*)

In linear case, solves the same problem as Kalman Filter. States produced by both algorithms at the end of the assimilation window are the same. In addition, if errors are Gaussian, algorithm, like KF, is Bayesian.

Propagates information both forward and backward in time. Is a *smoother*.

Can include nonlinear M_k and/or H_k, as well as temporal correlations.

Contrary to KF, does not produce an explicit estimate of uncertainty on the final estimate.

And, contrary to KF, there is no obvious cycling of estimation error from one assimilation window to the next.

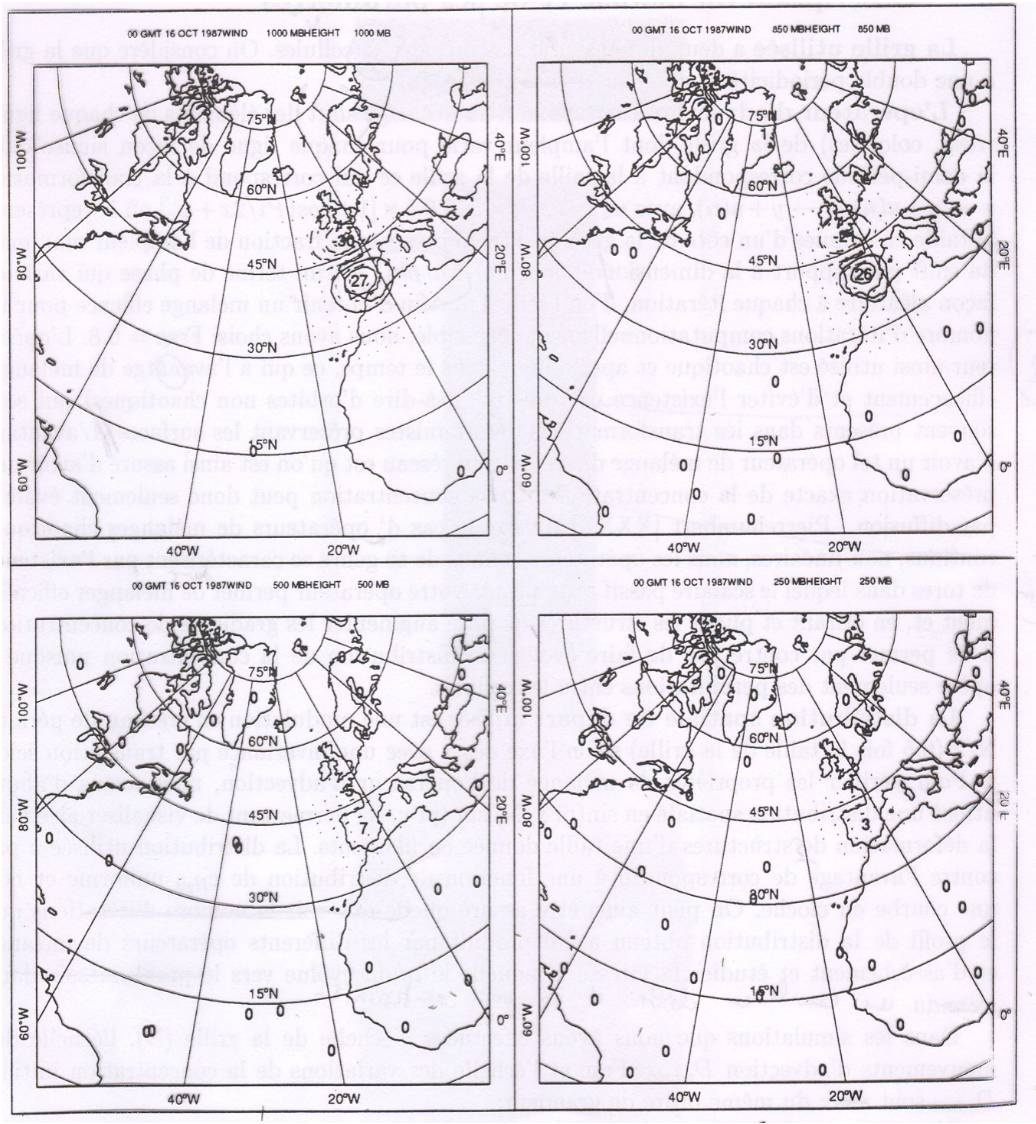
First (and still now largely) implemented assuming dynamical model to be exact (*strong constraint variational assimilation*, by opposition to *weak constraint variational assimilation*, which takes model error into account)

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

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

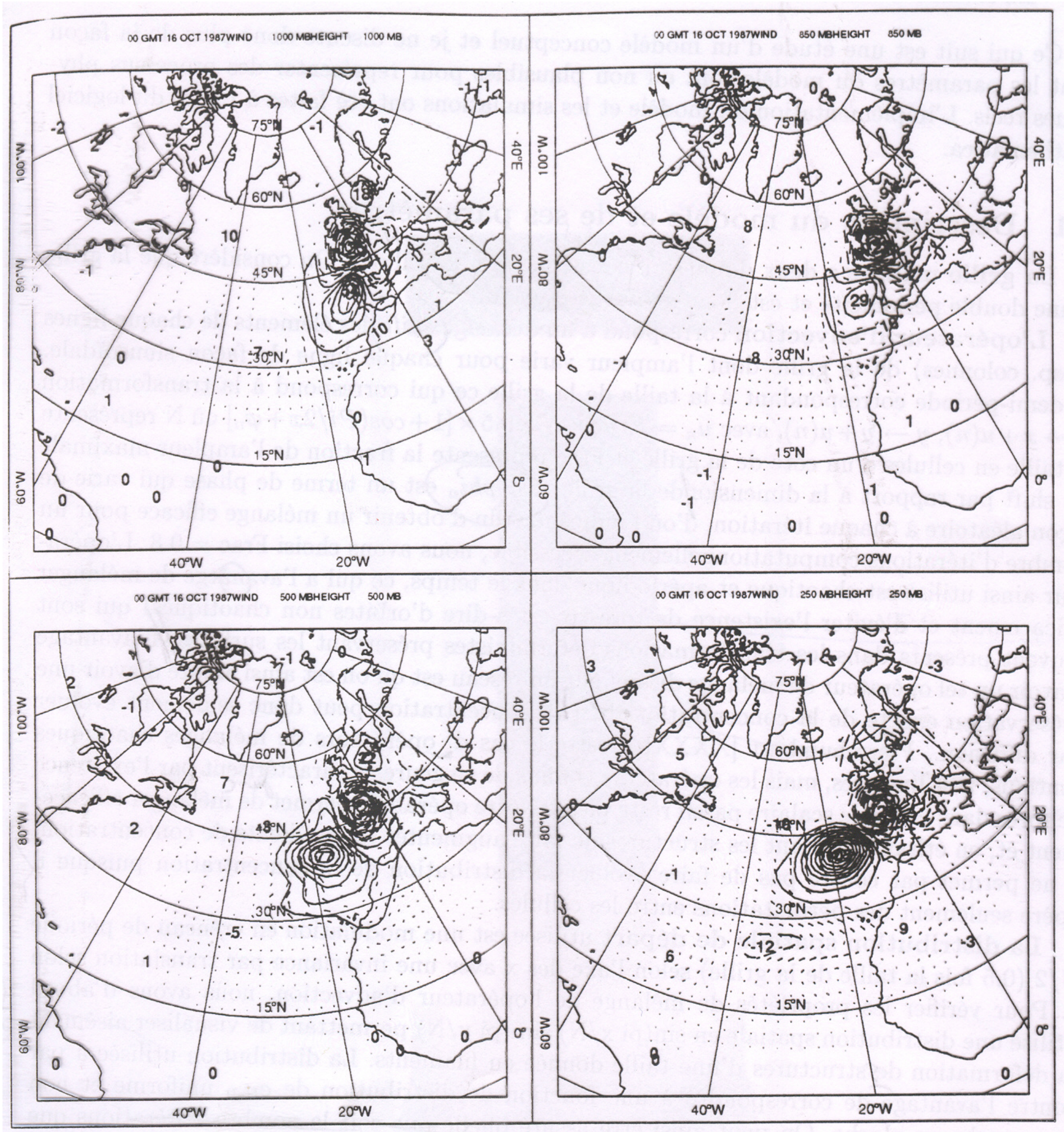
subject to $\xi_{k+1} = M_k\xi_k$, $k = 0, \dots, K\text{-}1$

Minimization performed through iterative descent algorithm, each step of which requires explicit knowledge of local gradient of objective function. Computations made possible through use of the *adjoint model* of the dynamical model. Local gradient is obtained by one forward integration of the direct model over the assimilation window, followed by one backward integration of the adjoint model (Penenko and Obraztsov, 1976, Le Dimet and Talagrand, 1986)



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

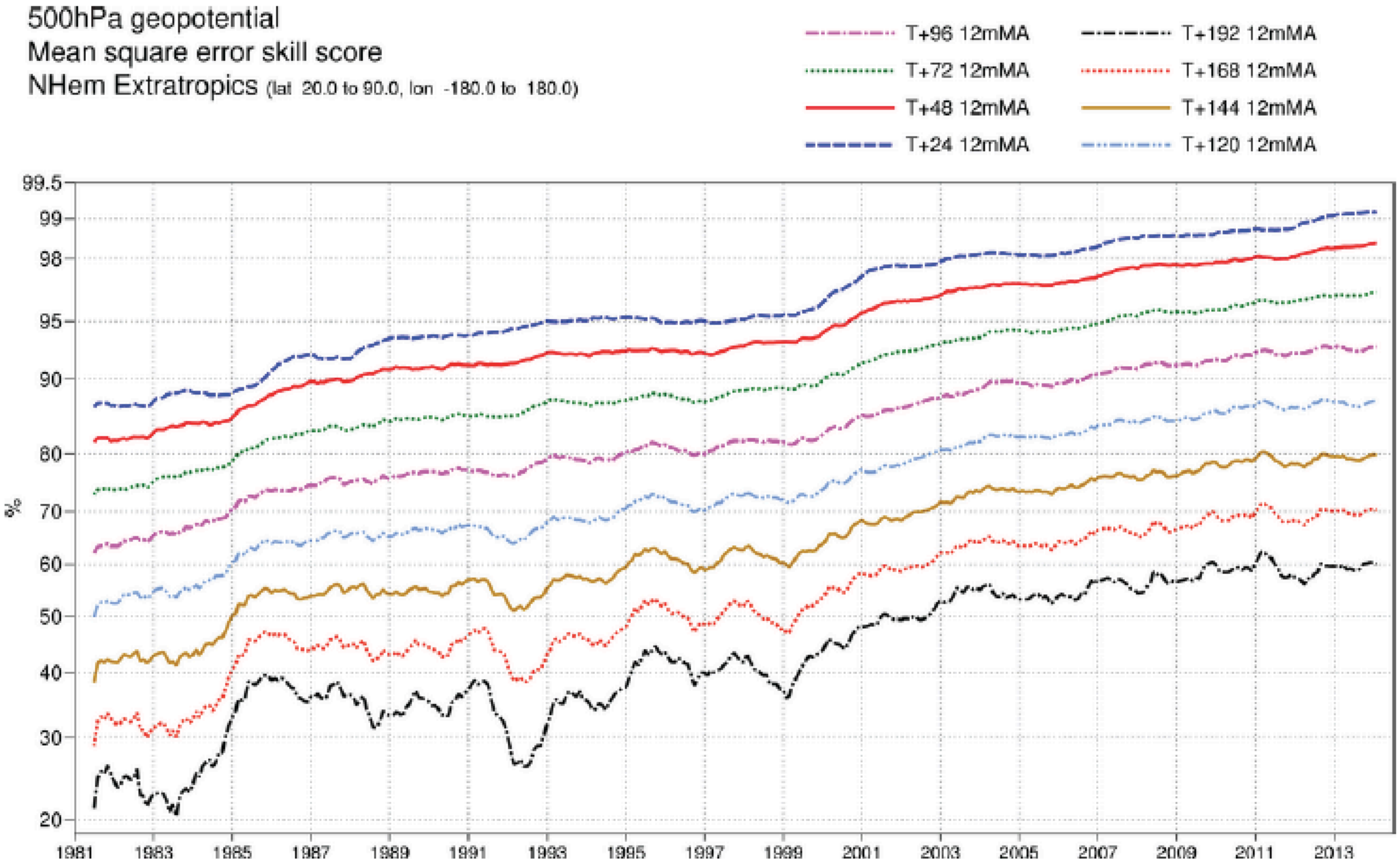


Figure 3: 500 hPa geopotential height mean square error skill score for Europe (top) and the northern hemisphere extratropics (bottom), showing 12-month moving averages for forecast ranges from 24 to 192 hours. The last point on each curve is for the 12-month period August 2013–July 2014.

Persistence = 0 ; climatology = 50 at long range

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.

Is some form of Ensemble Variational Assimilation possible ?

Objective function

$$
\begin{aligned}
\mathcal{J}(\xi_0, \xi_1, \ldots, \xi_K) \\
&= (1/2)\,(x_0^b - \xi_0)^T\,[P_0^b]^{-1}\,(x_0^b - \xi_0) \\
&\quad + (1/2)\,\Sigma_{k=0,\ldots,K}[y_k - H_k\xi_k]^T\,R_k^{-1}\,[y_k - H_k\xi_k] \\
&\quad + (1/2)\,\Sigma_{k=0,\ldots,K-1}[\xi_{k+1} - M_k\xi_k]^T\,Q_k^{-1}\,[\xi_{k+1} - M_k\xi_k]
\end{aligned}
$$

perturb the data x_0^b, y_k, 0 $(= x_{k+1} - M_k x_k)$ according to their error pdf, and perform a variational assimilation on each set of perturbed data. In linear and Gaussian case, that process is exactly Bayesian, in the sense that it produces independent realizations of the pdf of the state of the system, conditioned to the data.

Ensemble of Data Assimilations (EDA)

Used at ECMWF and Météo-France for defining (in part) background error covariance matrix P_0^b.

EnsVar : the non-linear Lorenz96 model 10 days with QSVA