Data Assimilation: A Brief Overview

Olivier Talagrand

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

Special Session *Inverse Problems and Data Assimilation*

Multitemp 2015
8th International Workshop on the Analysis of Multitemporal Remote Sensing Images

Annecy, France
23 July 2015

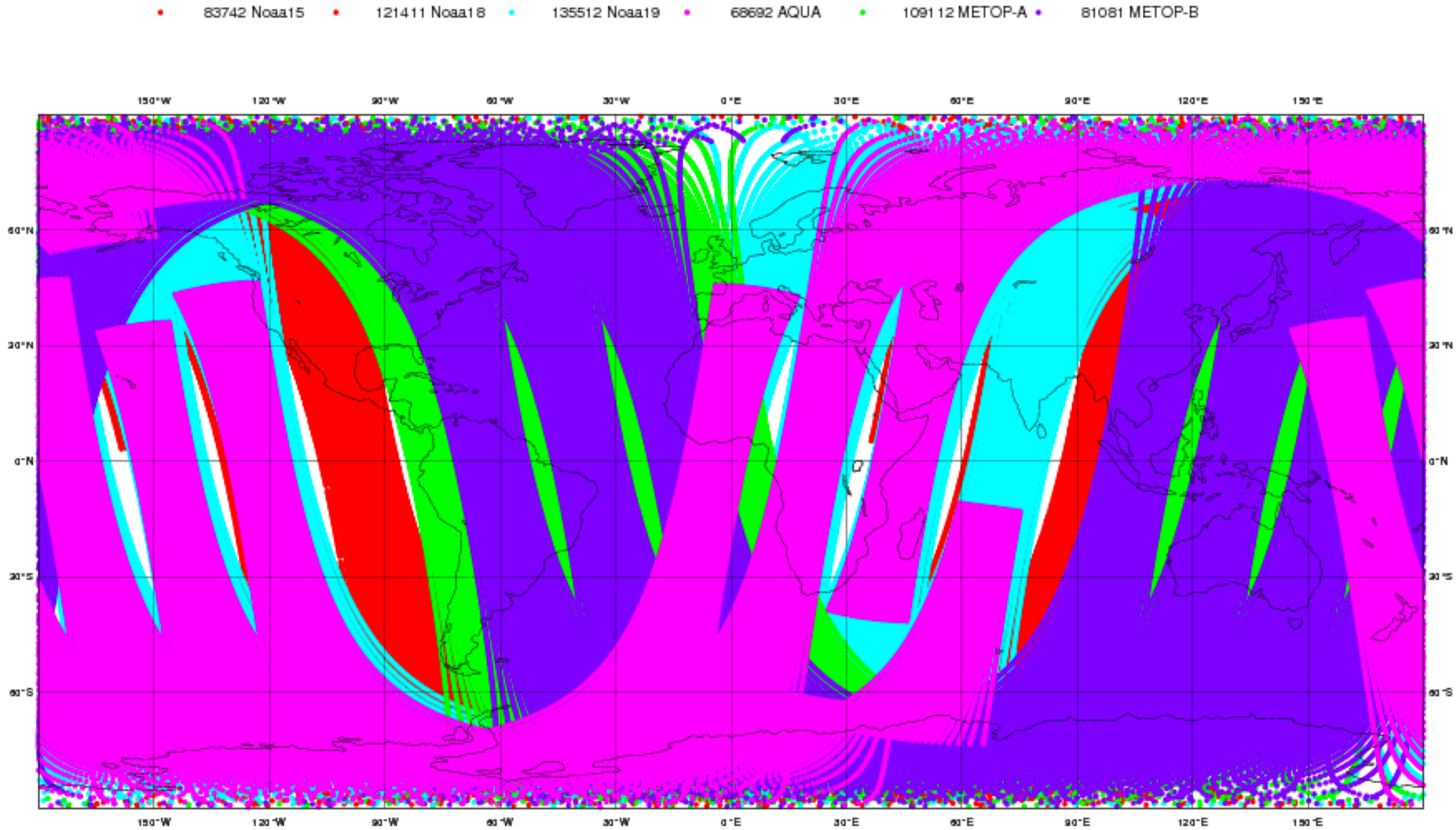
Figure 6 Hurricane Katrina mean-sea-level-pressure (MSLP) analysis for 12 UTC of 29 August 2005 and t+84h high-resolution and EPS forecasts started at 00 UTC of 26 August:

1st row: 1^{st} panel: MSLP analysis for 12 UTC of 29 Aug
2^{nd} panel: MSLP t+84h T_L511L60 forecast started at 00 UTC of 26 Aug
3^{rd} panel: MSLP t+84h EPS-control T_L255L40 forecast started at 00 UTC of 26 Aug
Other rows: 50 EPS-perturbed T_L255L40 forecast started at 00 UTC of 26 Aug.

The contour interval is 5 hPa, with shading patters for MSLP values lower than 990 hPa.

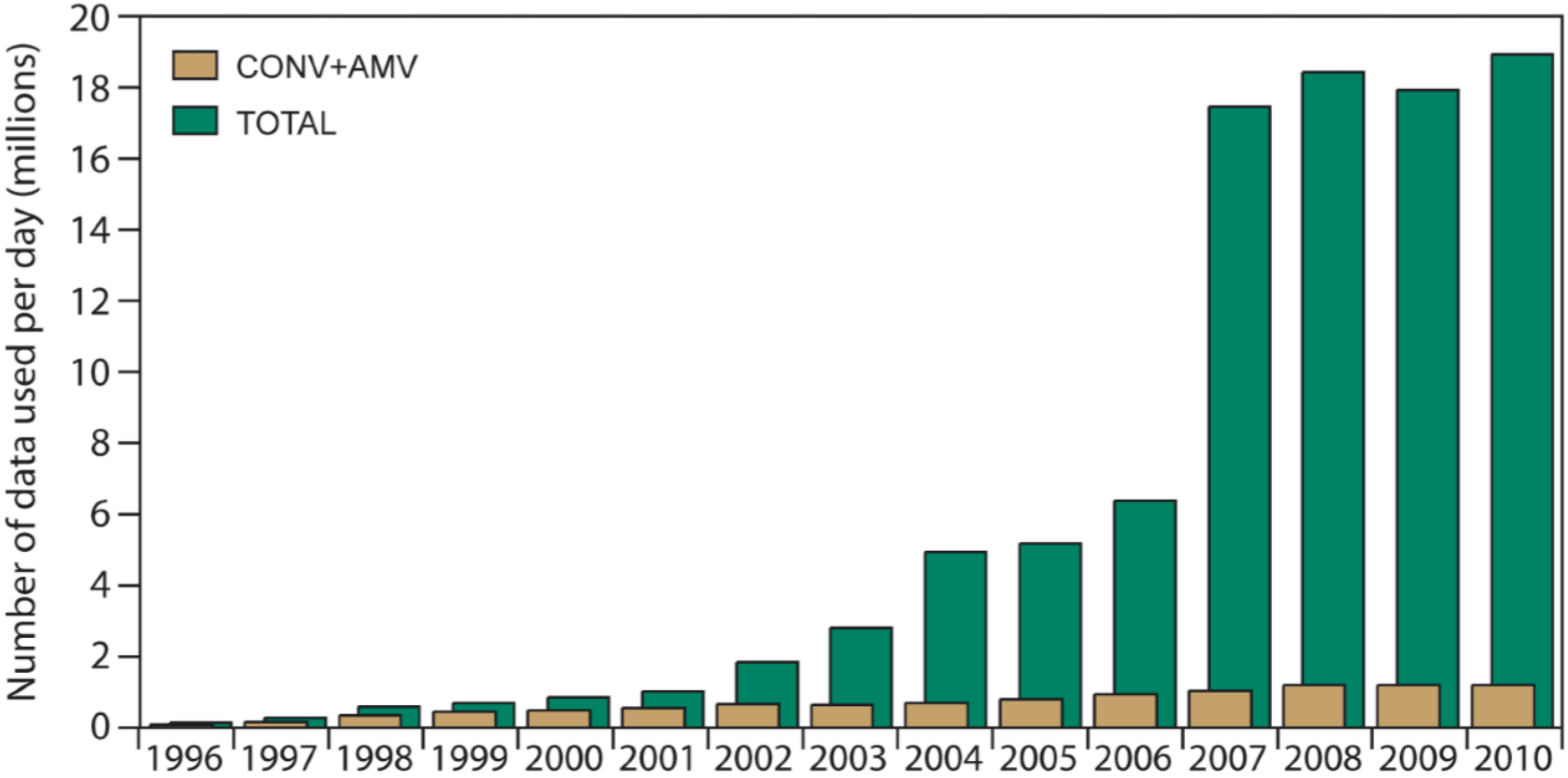
Assimilation of Observations originated from the need of defining initial conditions for Numerical Weather Predictions

ECMWF Data Coverage (All obs DA) - AMSU-A
19/Apr/2015; 00 UTC
Total number of obs = 599550



Magics 2.14.4 (64 bit)

ECMWF



Value as of early 2013 : around 25 millions per day

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

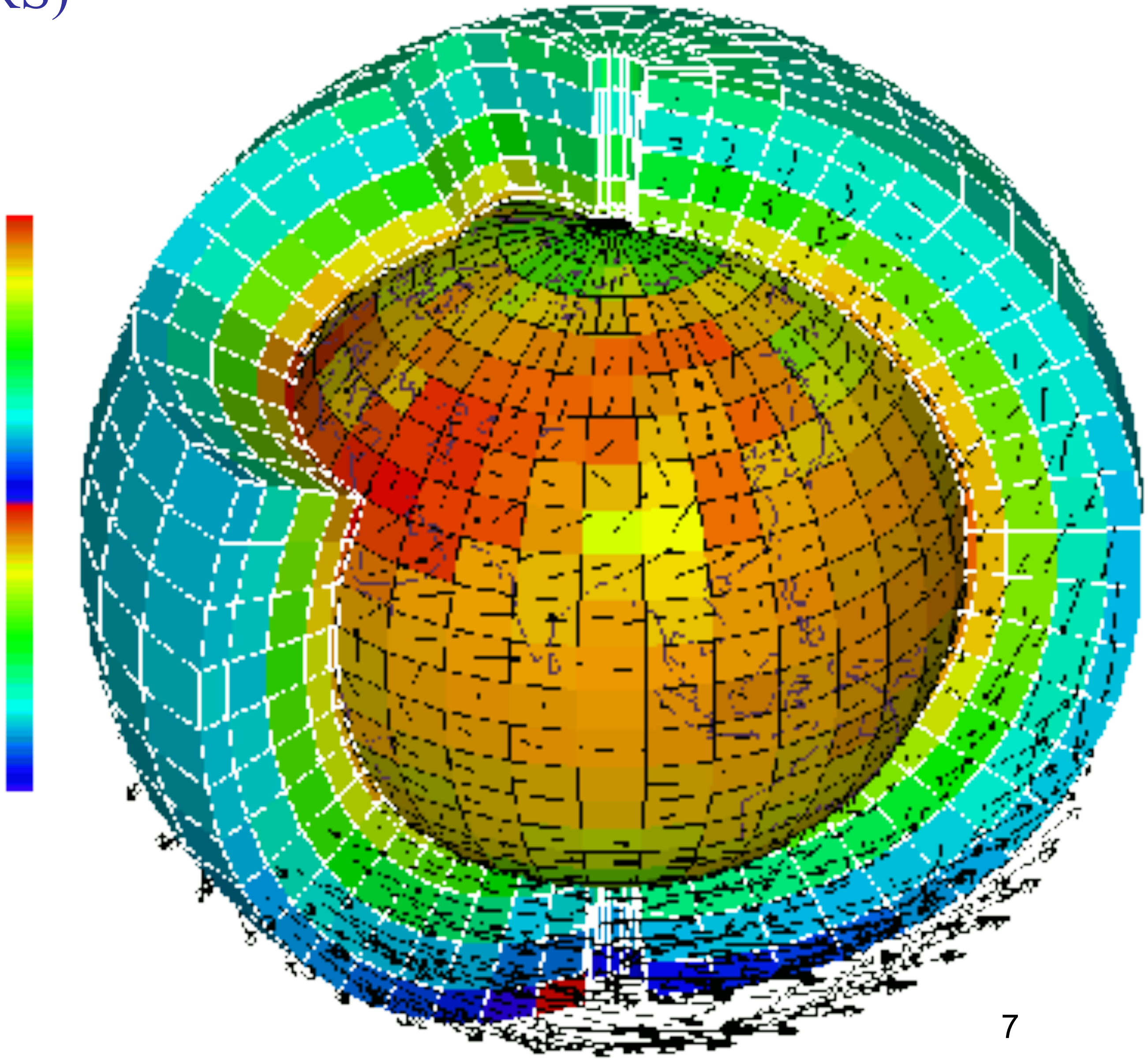
  $f(p, \rho, e) = 0$ $(p/\rho = rT, \, e = C_v T)$

- Conservation of mass of secondary components (water in the atmosphere, salt in the ocean, chemical species, ...)

  $Dq/Dt + q \, \mathrm{div}\underline{U} = S$

Physical laws available in practice in the form of a discretized (and necessarily imperfect) numerical model

A schematic of an Atmospheric General Circulation Model (L. Fairhead /LMD-CNRS)



European Centre for Medium-range Weather Forecasts (ECMWF, Reading, UK)

Horizontal spherical harmonics triangular truncation T1279 (horizontal resolution ≈ 16 kilometres)

137 levels on the vertical (0 - 80 km)

Dimension of state vector $n \approx 2.3\ 10^9$

Timestep (semi-implicit semi-Lagrangian scheme) ≈ 10 minutes

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, numerical 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.

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, ….)
- stellar and terrestrial magnetism
- …

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, $p \approx$ 1-3.10^7 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. The problem is largely to control the instabilities that can develop in the course of the assimilation.

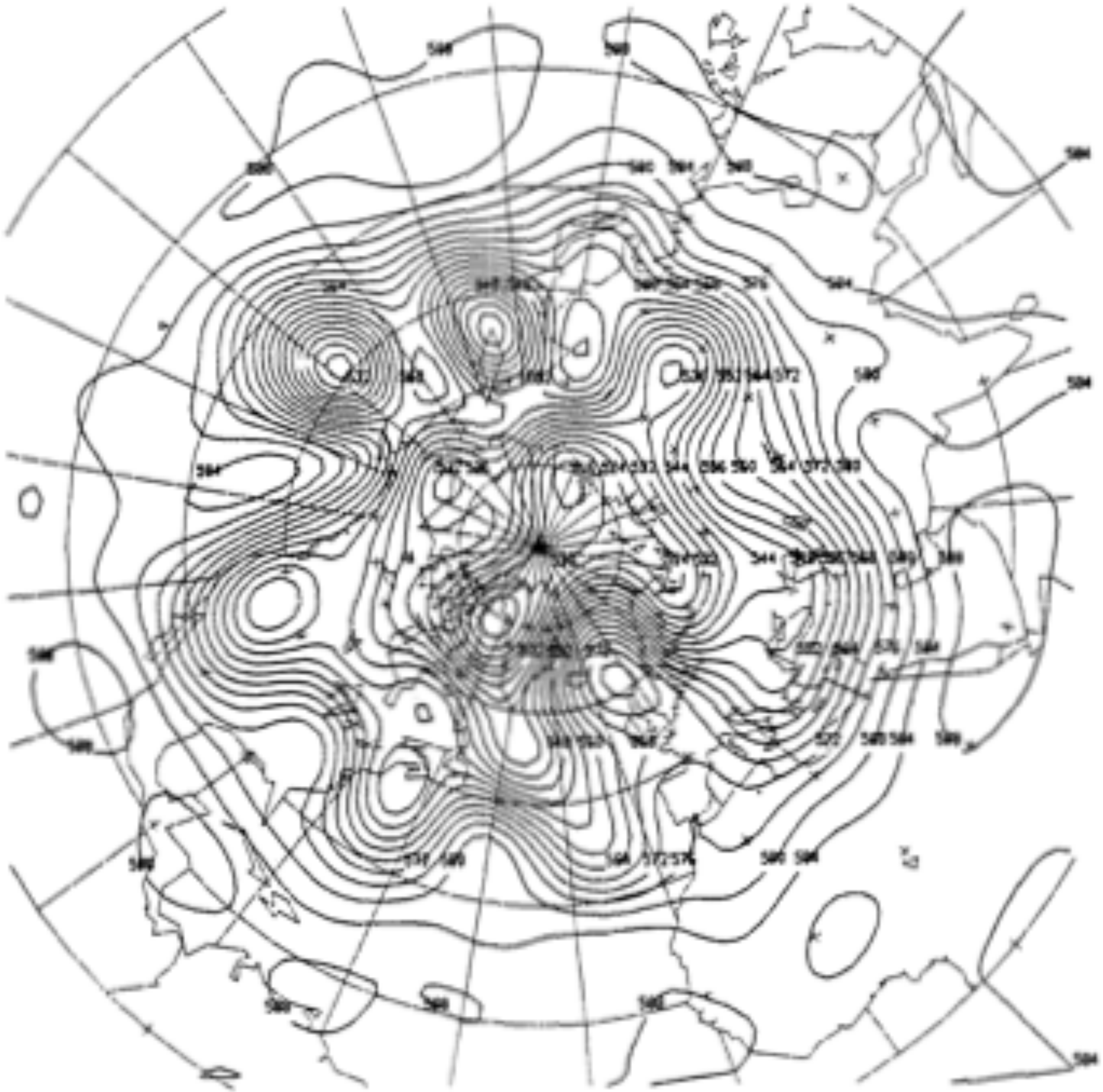


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

Courtier and Talagrand, *QJRMS*, 1987

(a)

Figure 1. Geographical distribution of the observations used for the assimilation experiments. (a): geopotential observations; (b): wind observations. At most of the points plotted, several observations were made at successive synoptic hours. On each of the two charts, the heavy line delineates the Aleutian depression (see Figure 2).

Courtier and Talagrand, *QJRMS*, 1987

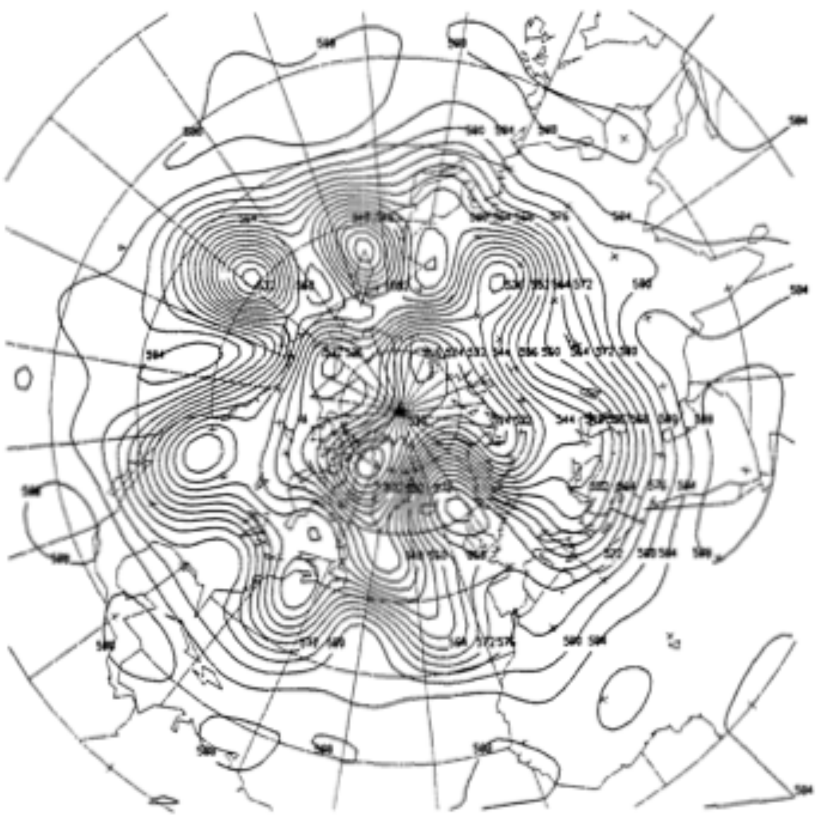


Figure 2. 500 mb height field produced by the operational analysis procedure of Direction de la Météorologie for 00 GMT, 26 April 1984. Units: dam, contour interval: 4 dam. The field has been truncated to the truncation of the model used for the experiments described in the article.

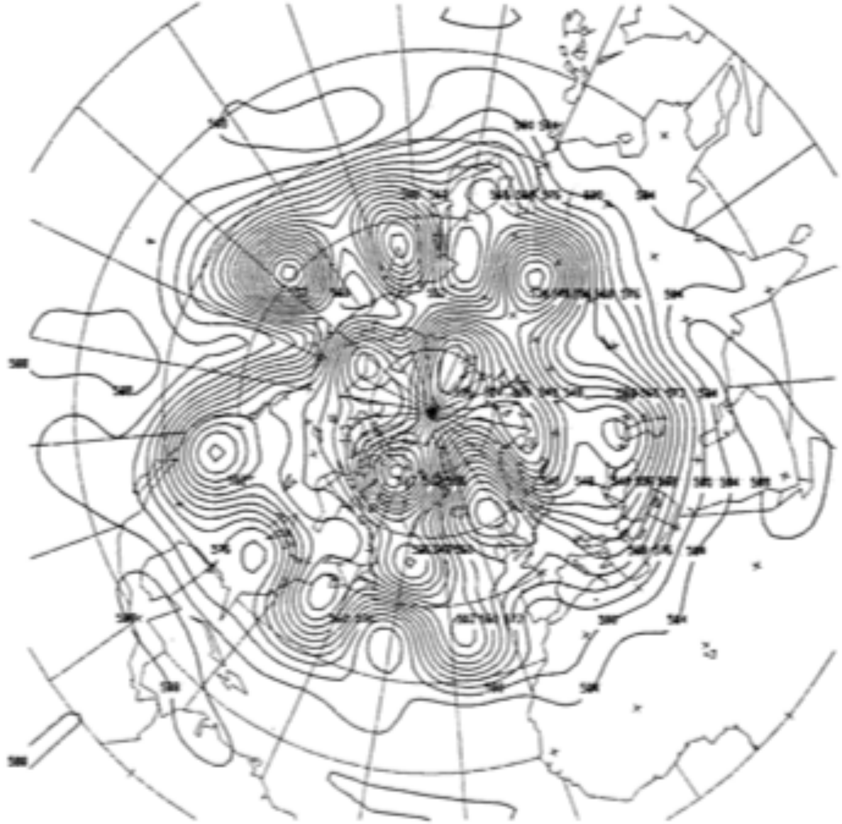


Figure 3. 500 mb height field produced for 00 GMT, 26 April 1984, by the variational analysis minimizing the distance function defined by Eqs. (1)–(2) over a 24-hour period. Units: dam; contour interval: 4 dam.

500-hPa geopotential field as determined by : (left) operational assimilation system of French Weather Service (3D, primitive equation) and (right) experimental variational system (2D, vorticity equation)

Courtier and Talagrand, *QJRMS*, 1987

Typical situation

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

  $x^b = x + \zeta^b \qquad\qquad E(\zeta^b \zeta^{bT}) \equiv \boldsymbol{P}^b$

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

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

  H is known *observation operator* (assumed to linear at this stage)

Least-variance linear estimate of $\boldsymbol{x}$ from $\boldsymbol{x}^b$ and $\boldsymbol{y}$

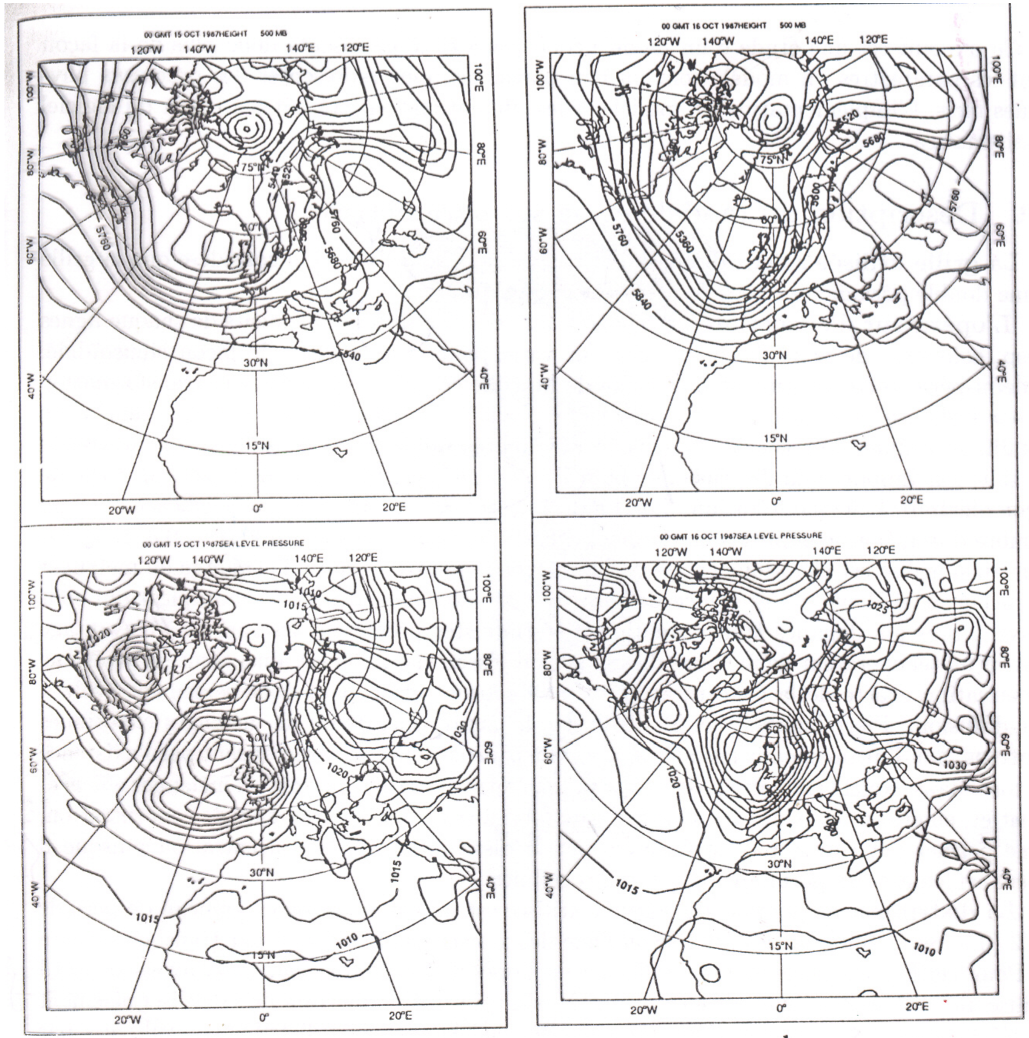
$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{P}^b\boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b)$$
$$\boldsymbol{P}^a = \boldsymbol{P}^b - \boldsymbol{P}^b\boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T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

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

The vector $\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b$ is the difference between the observation and what the background predicts for the observation. It is called the *innovation vector*.

The matrix $\boldsymbol{K} = \boldsymbol{P}^b\boldsymbol{H}^T [\boldsymbol{H}\boldsymbol{P}^b\boldsymbol{H}^T + \boldsymbol{R}]^{-1} = \boldsymbol{P}^a \boldsymbol{H}^T \boldsymbol{R}^{-1}$ is *gain matrix*.

If joint probability distribution of errors (ζ^b, ε) is gaussian, *BLUE* achieves bayesian estimation, in the sense that $P(\boldsymbol{x} \mid \boldsymbol{x}^b, \boldsymbol{y}) = \mathcal{N}[\boldsymbol{x}^a, \boldsymbol{P}^a]$.

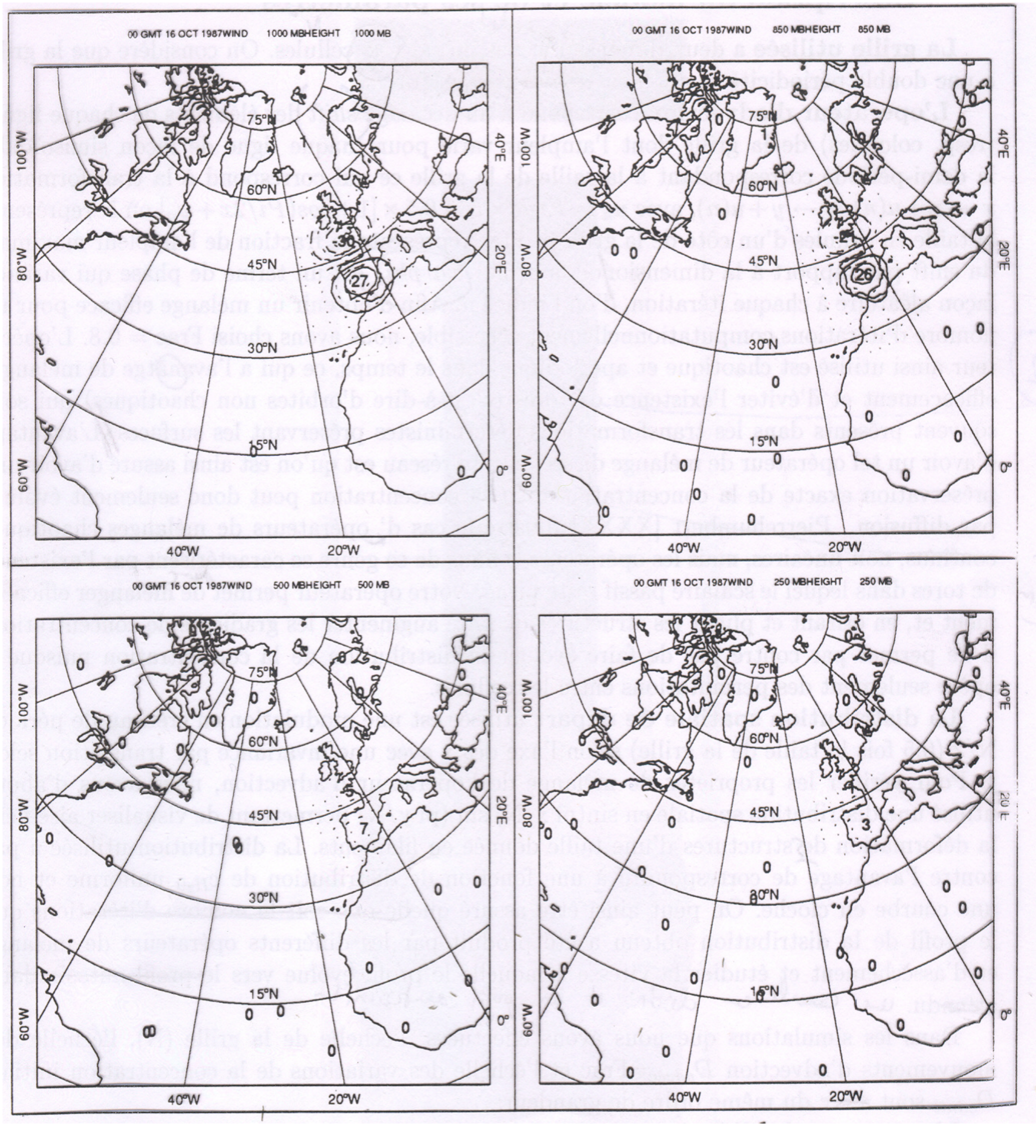


b

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



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

Best Linear Unbiased Estimate

Variational form of the *BLUE*

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

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

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

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

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

A large part of what is done in assimilation of observations, in geophysical applications as well as in other applications, is based on more or less empirical extensions of the *BLUE* to (moderately) nonlinear situations.

How to introduce time dimension (that is the real problem in assimilation) ?

- Observation vector at time k

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

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

- Evolution equation

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

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

$$E(\eta_k \varepsilon_j^{\mathrm{T}}) = 0$$

- Background estimate at time 0

$$x^b{}_0 = x_0 + \zeta^b{}_0$$

$$E(\zeta^b{}_0) = 0 \quad ; \quad E(\zeta^b{}_0 \, \zeta^b{}_0{}^{\mathrm{T}}) \equiv P^b{}_0$$

$$E(\zeta^b{}_0 \varepsilon_k^{\mathrm{T}}) = 0 \quad ; \quad E(\zeta^b{}_0 \eta_k^{\mathrm{T}}) = 0$$

- Errors uncorrelated in time

Sequential assimilation assumes the form of ***Kalman filter (*** Kalman, 1960)

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

$$x^b_{k+1} = M_k x^a_k$$

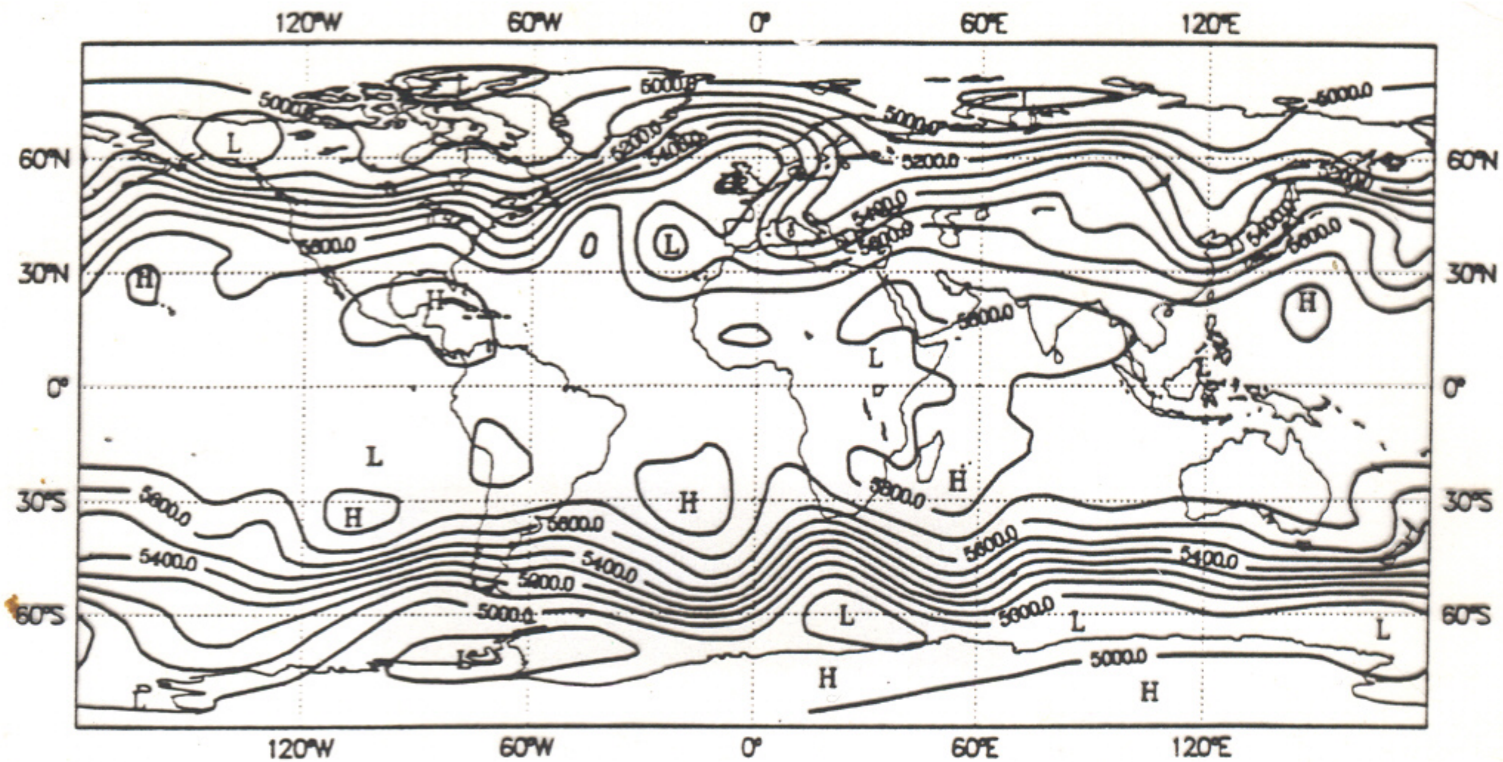
$$P^b_{k+1} = M_k P^a_k M_k^{\mathrm{T}} + Q_k$$

Costliest part of computation

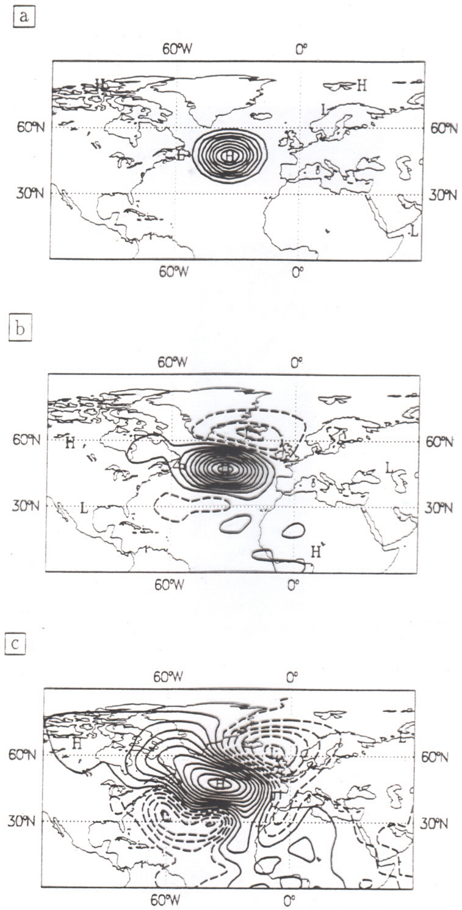
$$P^b{}_{k+1} = M_k P^a{}_k M_k{}^{\mathrm{T}} + Q_k$$

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. 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.

Most common approach at present is *Ensemble Kalman Filter* (*EnKF*)

- 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.

G. Evensen (Bergen), J. Anderson, C. Snyder, T. Hamill, (NCAR) ….

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.

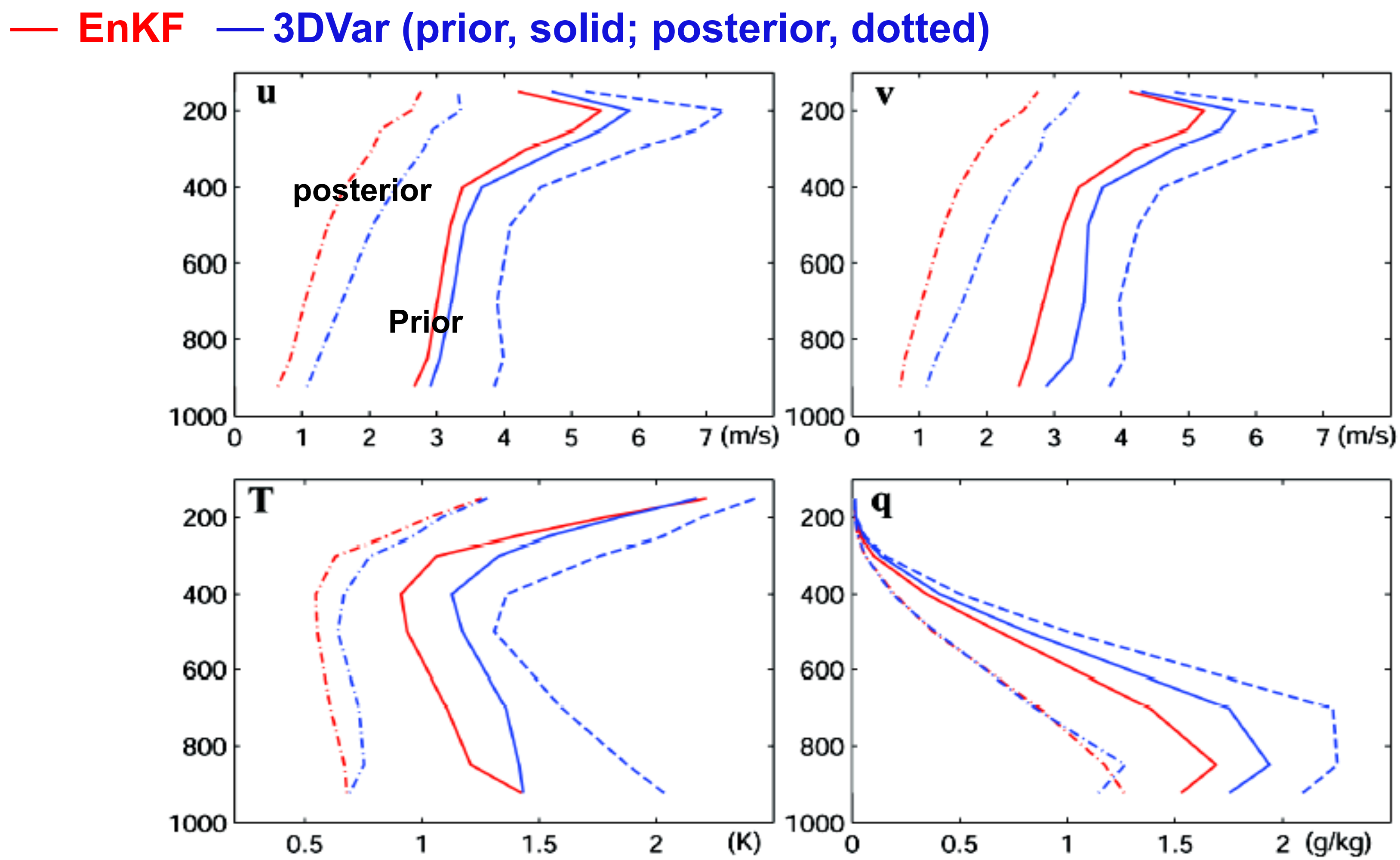
Then use Kalman formula to produce sample of 'analysed' states

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

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.

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



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)

EnKF is used operationally at several meteorological and oceanographical centres (Deutscher Wetterdienst, Canadian Meteorological Center, …).

Variational Assimilation.

Available 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 $k = 0, \dots, K$

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

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

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

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

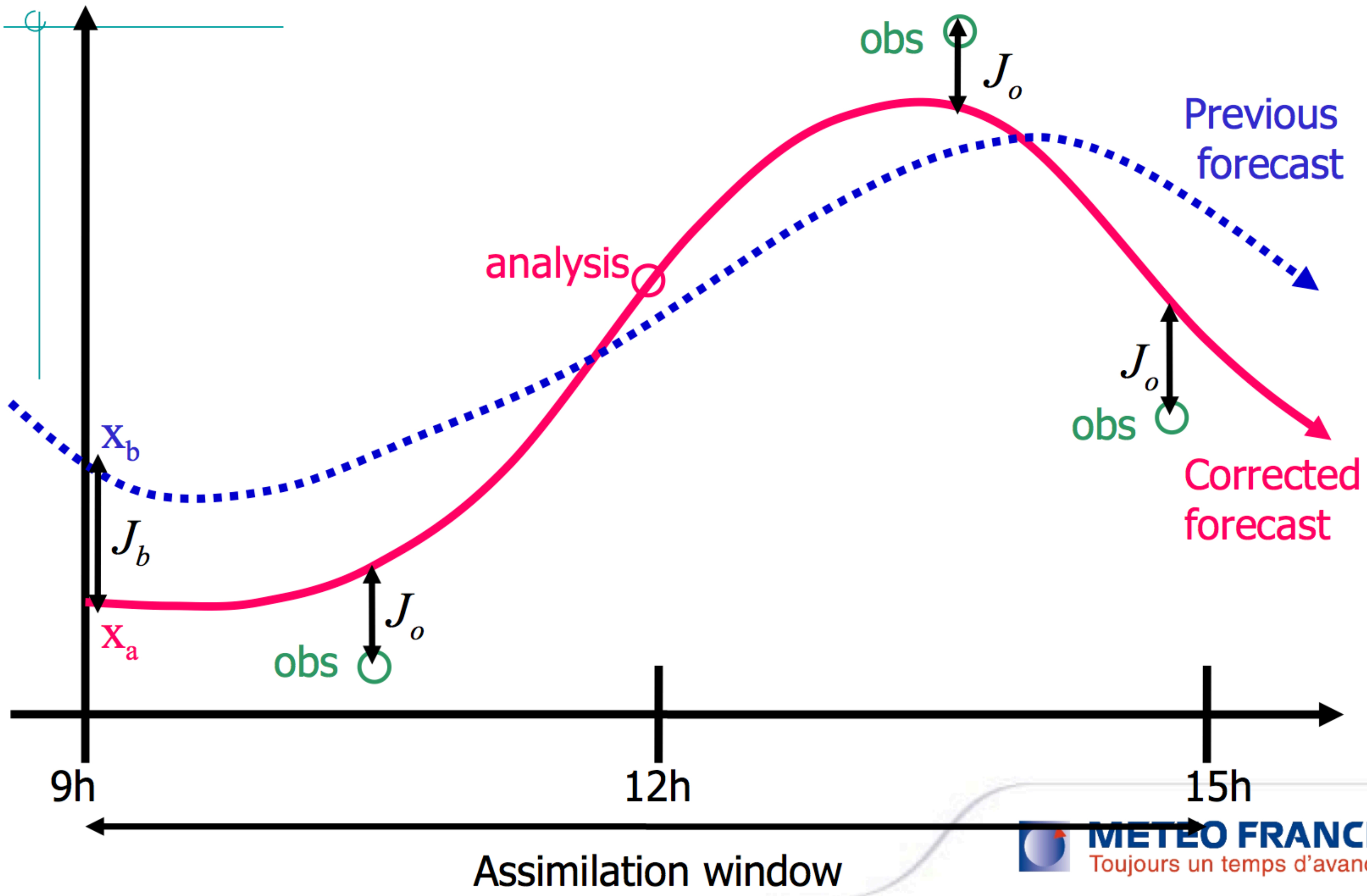
Then objective function

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

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

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

Principle of 4D-VAR assimilation



METEO FRANCE
Toujours un temps d'avance

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

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

Propagates information both forward and backward in time.

Background is not necessary, if observations are in sufficient number to overdetermine the problem. Nor is 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 gradient $\nabla_u\mathcal{J}$?

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 $\boldsymbol{\nabla}_{\boldsymbol{u}}\mathcal{J}$ 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

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\mathcal{J} \;=\; \boldsymbol{G}'^{\mathrm{T}} \, \boldsymbol{\nabla}_{\boldsymbol{v}}\mathcal{J} \qquad \text{(A)}$$

Evolution equation

$$d\boldsymbol{x}/dt = F(\boldsymbol{x}(t))$$

Tangent linear equation

$$d\delta\boldsymbol{x}/dt = \boldsymbol{F}'(\boldsymbol{x}(t), t)\ \delta\boldsymbol{x}$$

describes to first order evolution of small perturbation $\delta\boldsymbol{x}$ on $\boldsymbol{x}$

Adjoint equation

$$d\boldsymbol{x}'/dt = -\ \boldsymbol{F}'^{*}(\boldsymbol{x}(t), t)\ \boldsymbol{x}'$$

describes (exact) evolution of gradient $\boldsymbol{x}'$ of (any) scalar function with respect to $\boldsymbol{x}$. Adjoint equation is to be integrated baclward in time.

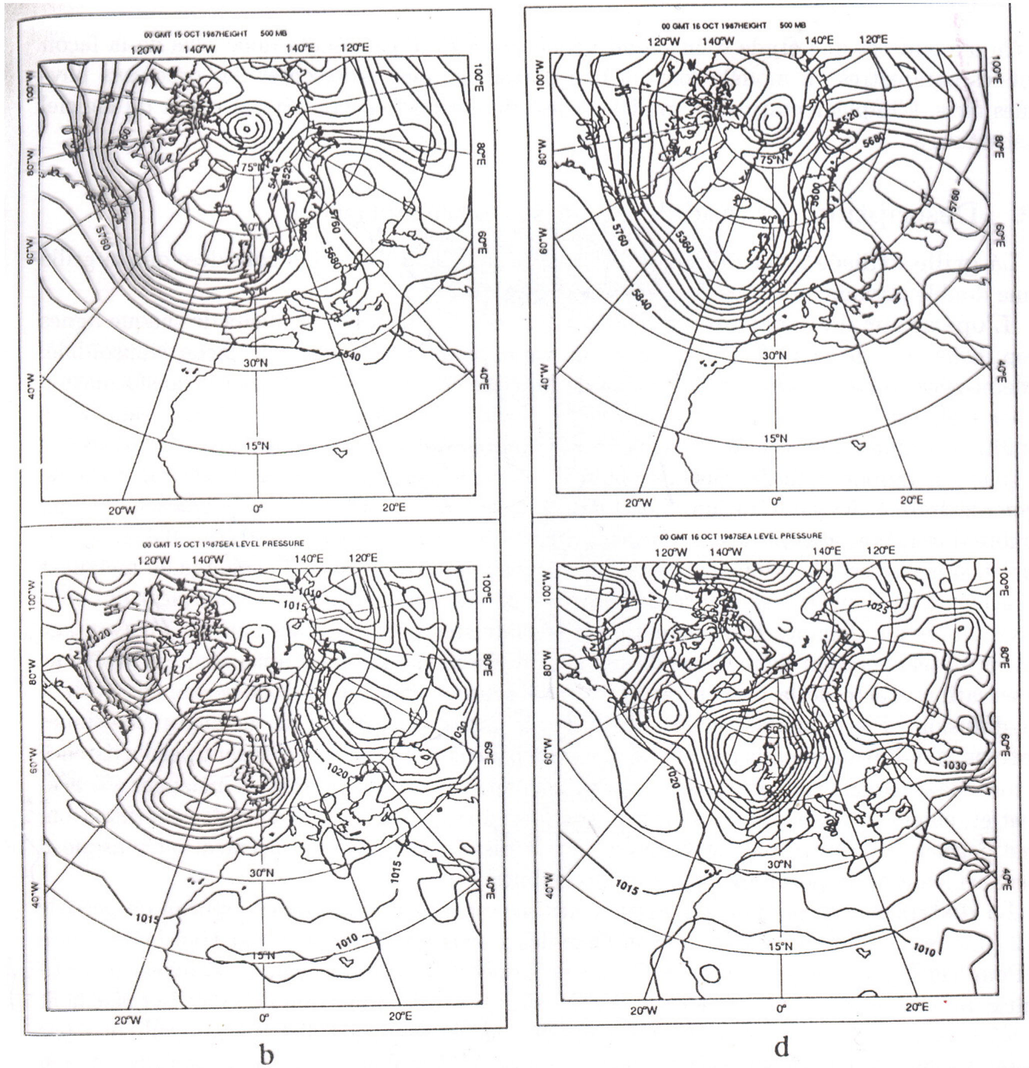
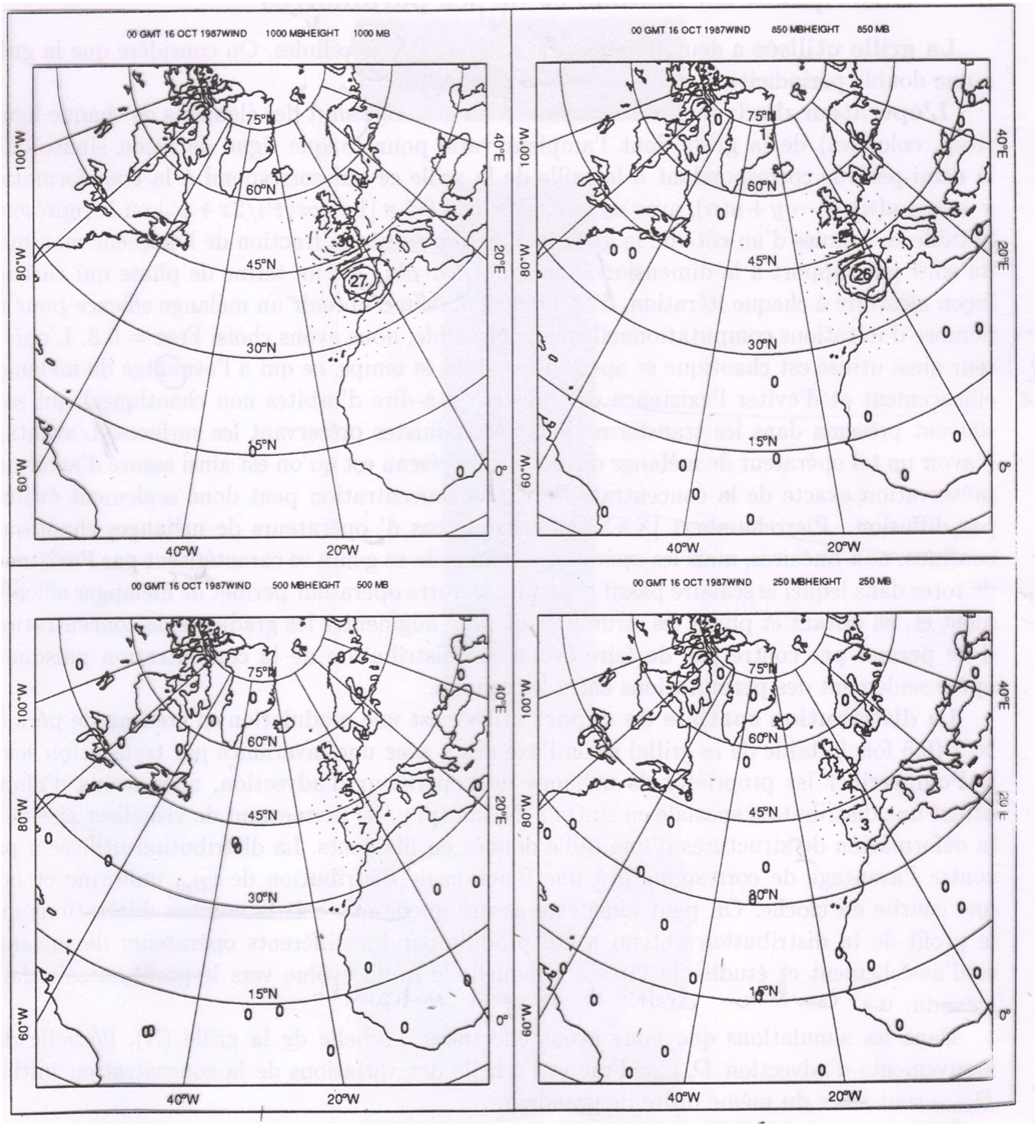


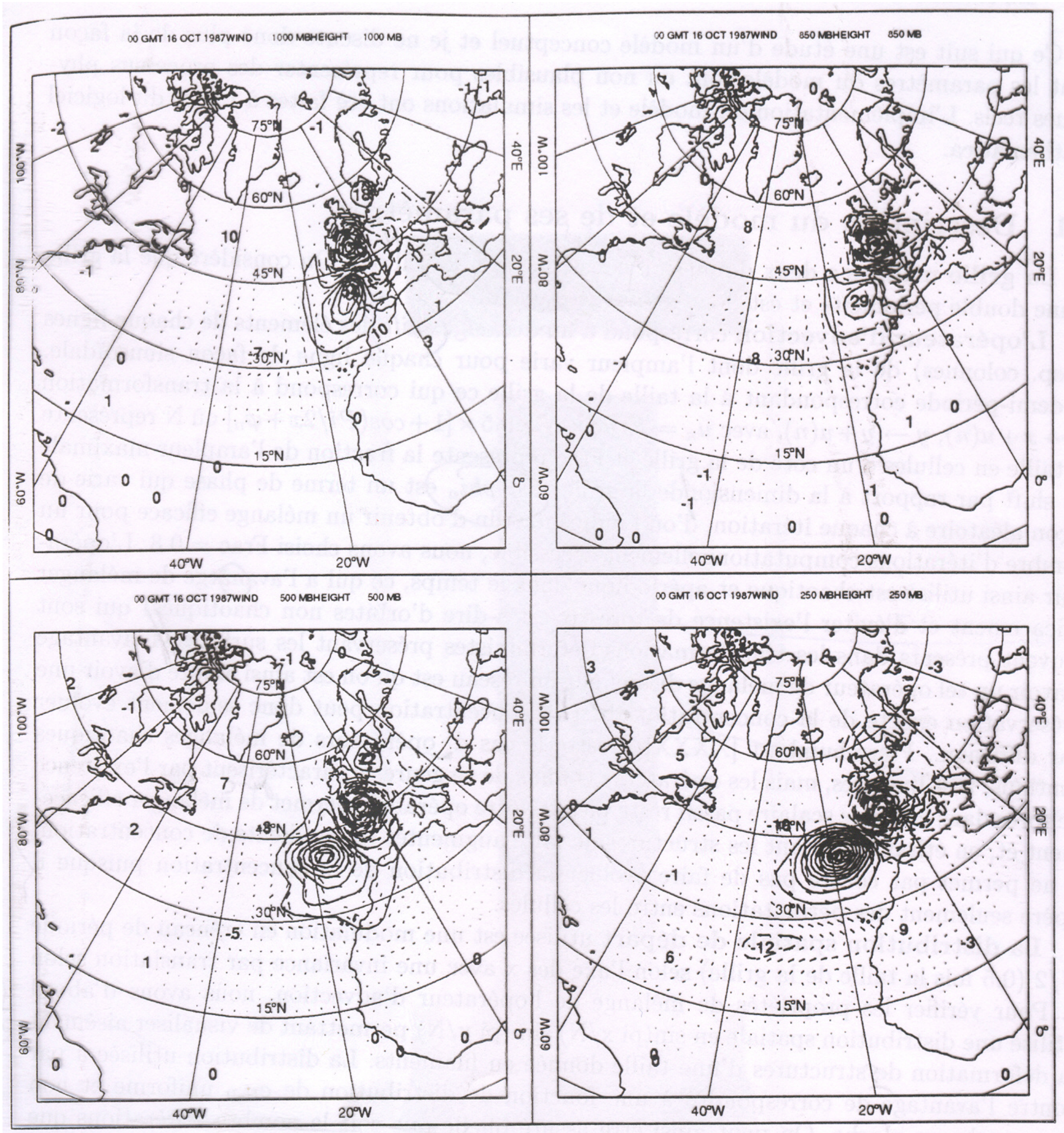
FIG. 1. Background fields for 0000 UTC 15 October–0000 UTC 16 October 1987. Shown here are the Northern Hemisphere (a) 500-hPa geopotential height and (b) mean sea level pressure for 15 October and the (c) 500-hPa geopotential height and (d) mean sea level pressure for 16 October. The fields for 15 October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



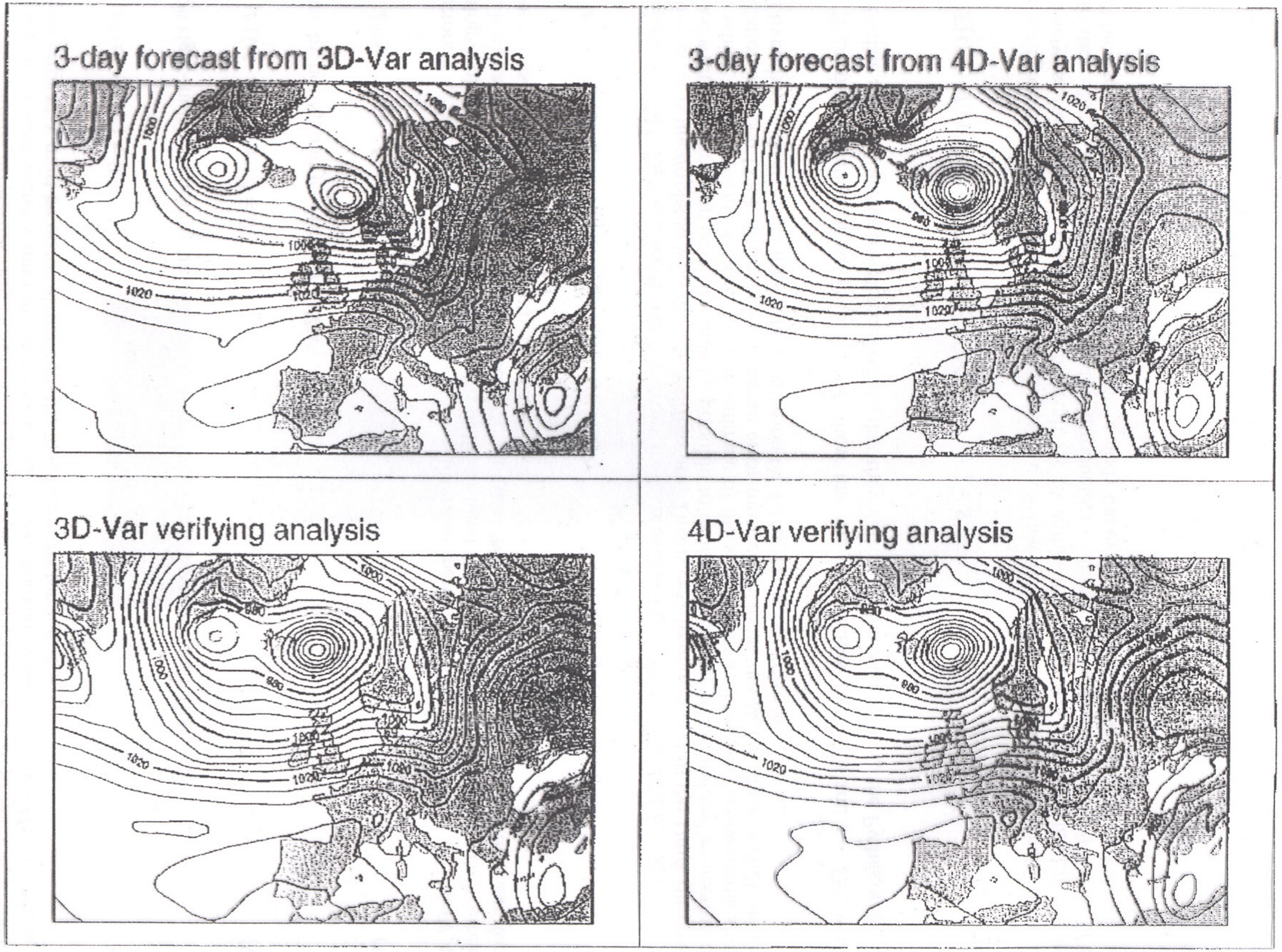
Analysis increments in a 3D-Var corresponding to a *u*-component wind observation at the 1000-hPa pressure level (no temporal evolution of background error covariance matrix)

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



ECMWF, Results on one FASTEX case (1997)

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

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.

Conclusion on Ensemble Kalman Filter

Pros

‘Natural’, and well adapted to many practical situations

Provides, at least relatively easily, explicit estimate of estimation error

Cons

Carries information only forward in time (of no importance if one is interested only in doing forecast)

In present form, optimality is possible only if errors are independent in time

Conclusion on Variational Assimilation

Pros

Carries information both forward and backward in time (important for reassimilation of past data).

Can easily take into account temporal statistical dependence (Järvinen *et al*.)

Does not require explicit computation of temporal evolution of estimation error

Very well adapted to some specific problems (*e. g.*, identification of tracer sources)

Cons

Does not readily provide estimate of estimation error

Requires development and maintenance of adjoint codes. But the latter can have other uses (sensitivity studies).

- Dual approach seems most promising. But still needs further development for application in non exactly linear cases.

- Is ensemble variational assimilation possible ? Probably yes. But also needs development.

Exact bayesian estimation ?

Particle filters

Predicted ensemble at time t : $\{x^b{}_n, n = 1, \ldots, N\}$, each element with its own weight (probability) $P(x^b{}_n)$

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

Bayes' formula

$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\ P(x^b{}_n)$$

Defines updating of weights

Bayes' formula

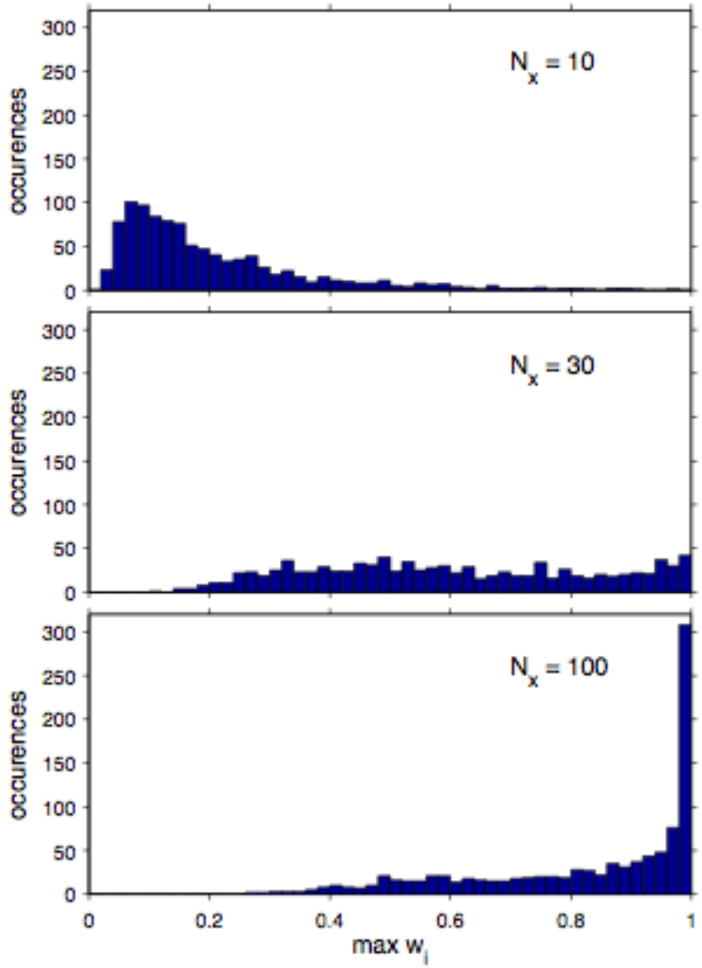
$$P(x^b{}_n|y) \sim P(y|x^b{}_n)\ P(x^b{}_n)$$

Defines updating of weights; particles are not modified. Asymptotically converges to bayesian pdf. Very easy to implement.

Observed fact. For large state dimension, ensemble tends to collapse.

Behavior of $\max w^i$

- $N_e = 10^3$; $N_x = 10, 30, 100$; 10^3 realizations



average squared error of posterior mean $= 5.5$

... $= 25$

... $= 127$

C. Snyder, http://www.cawcr.gov.au/staff/pxs/wmoda5/Oral/Snyder.pdf

Problem originates in the 'curse of dimensionality'. Large dimension pdf's are very diffuse, so that very few particles (if any) are present in areas where conditional probability ('*likelihood*') $P(y|x)$ is large.

Bengtsson *et al*. (2008) and Snyder *et al*. (2008) evaluate that stability of filter requires the size of ensembles to increase exponentially with space dimension.

Importance Sampling.

Use a *proposal density* that is closer to the new observations than the density defined by the predicted particles (for instance the density defined by EnKF, after the latter has used the new observations). Independence between particles is then lost in the computation of likelihood $P(y|x)$ (or is it ?)

In particular, *Guided Sequential Importance Sampling* (van Leeuwen, 2002). Idea : use observations performed at time k to resample ensemble at some timestep anterior to k, or 'nudge' integration between times k-1 and k towards observation at time k.

Particle filters are actively studied (van Leeuwen, Morzfeld, …)

Particle filters are actively studied (van Leeuwen, Morzfeld, …)

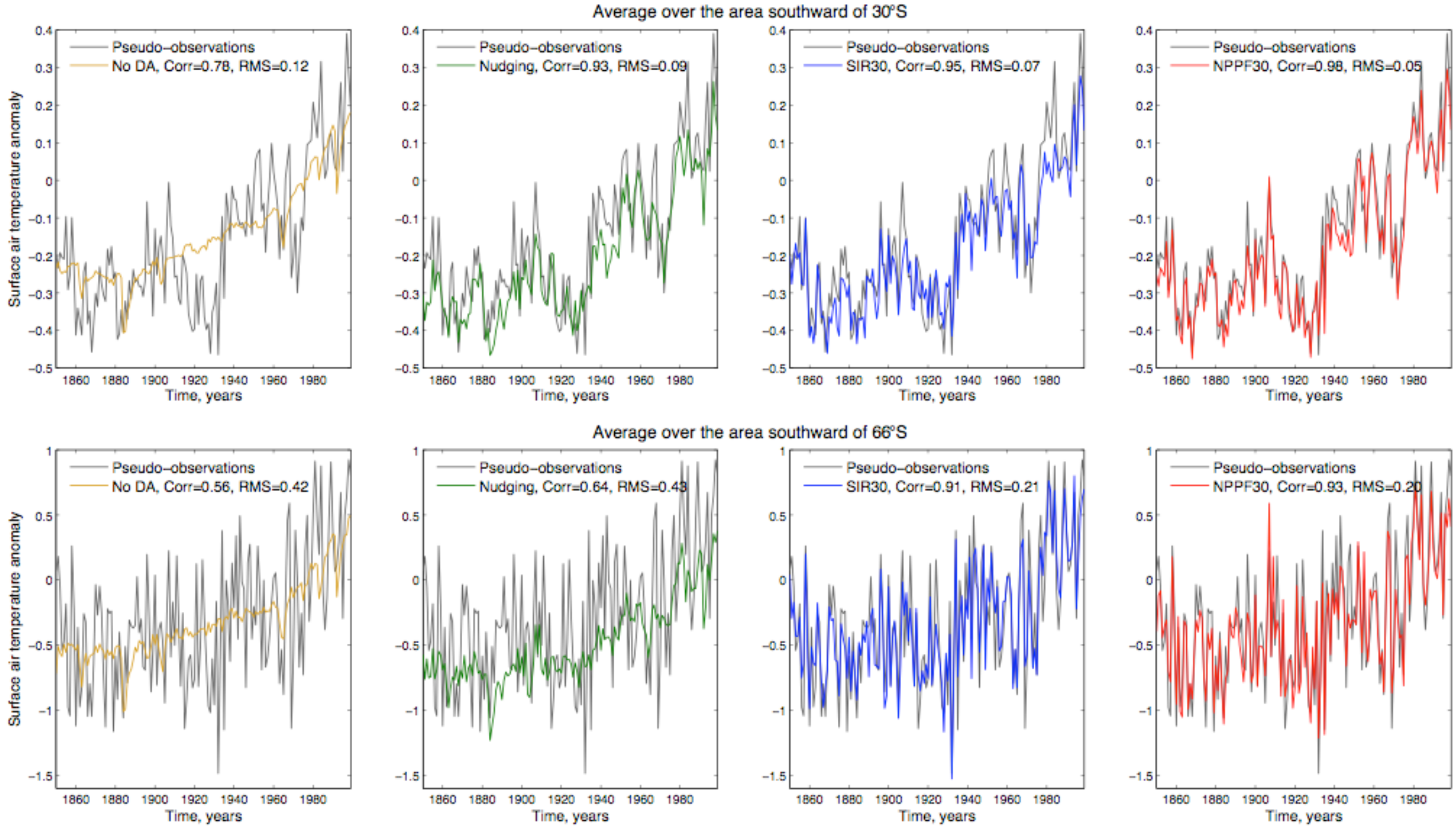
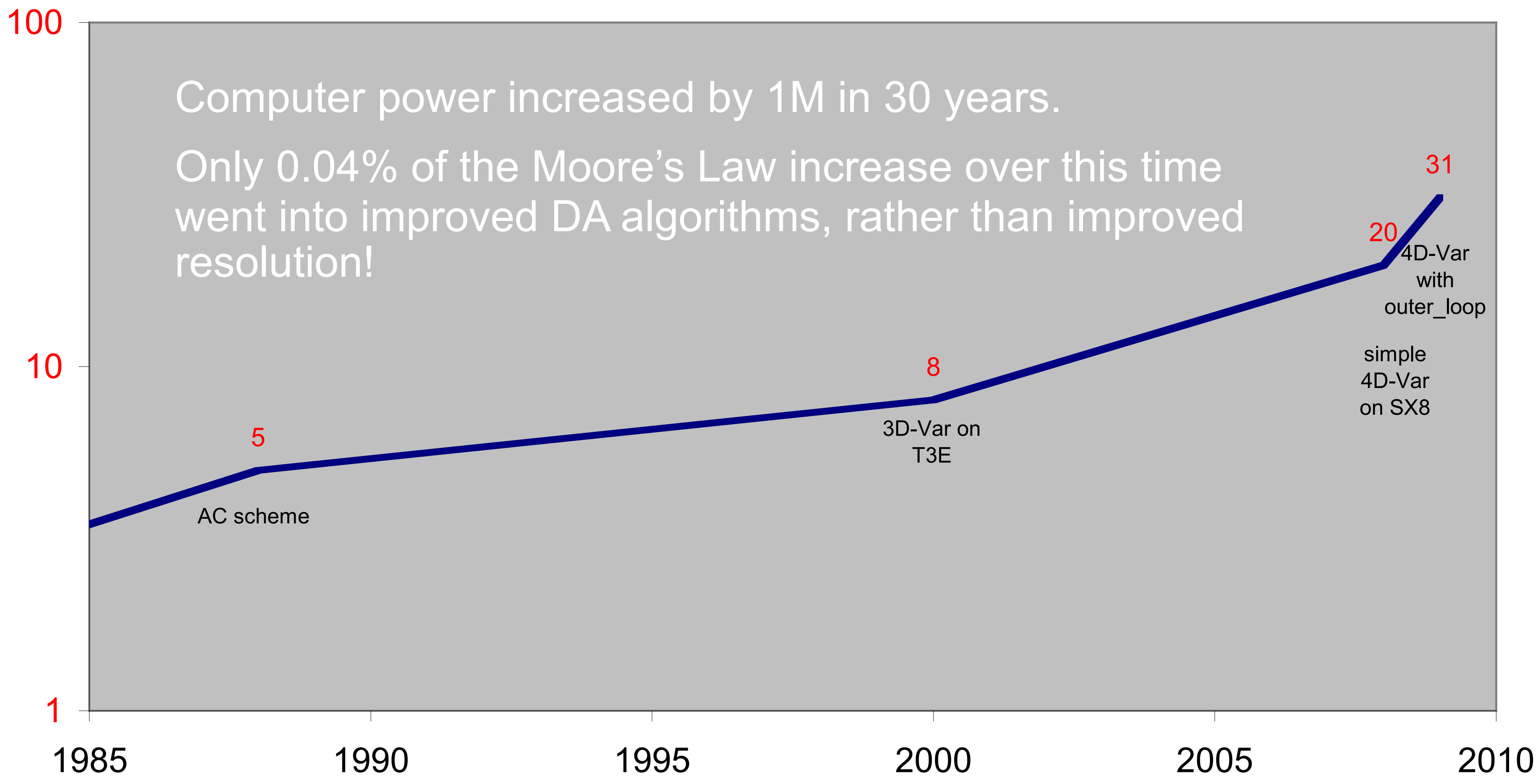


Fig. 1. Reconstructions of surface air temperature anomalies (°C) averaged over the area southward of 30° S (top panel) and over the area southward of 66° S (bottom panel), when the dense pseudo-observations are assimilated. Gray line: pseudo-observations; yellow line: model simulations without data assimilation; green line: the nudging; blue line: the sequential importance resampling applied over the polar cap southward of 30° S; red line: the nudging proposal particle filter applied over the polar cap southward of 30° S. Correlations and the RMS errors are displayed in upper left corners.

Dense observations (from Dunbinkina and Goosse, *Clim. Past*, **9**, 2013)

ratio of supercomputer costs:
1 day's assimilation / 1 day forecast



Courtesy A. Lorenc

Assimilation, which originated from the need of defining initial conditions for numerical weather forecasts, has gradually extended to many diverse applications

- Oceanography
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciology
- Magnetism (both planetary and stellar)
- Plate tectonics
- Planetary atmospheres (Mars, …)
- Reassimilation of past observations (mostly for climatological purposes, ECMWF, NCEP/NCAR)
- Identification of source of tracers
- Parameter identification
- *A priori* evaluation of anticipated new instruments
- Definition of observing systems *(Observing Systems Simulation Experiments)*
- Validation of models
- Sensitivity studies (adjoints)
- …

It has now become a major tool of numerical environmental science

A few of the (many) remaining problems :

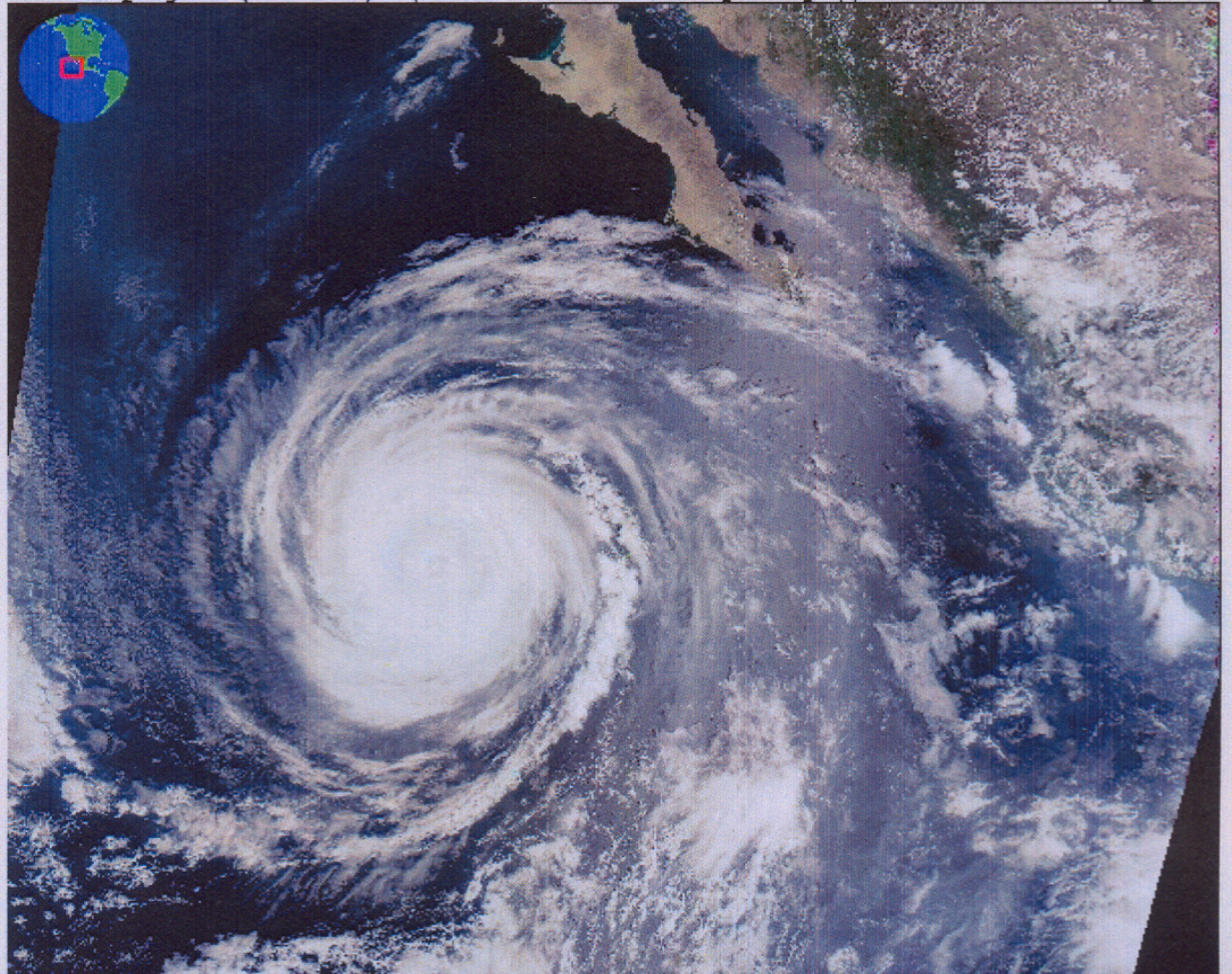
- Observability (what to observe in order to know what we want to know ? Data are noisy, system is chaotic !)

- More accurate identification and quantification of errors affecting data particularly the assimilating model (will always require independent hypotheses)

- Assimilation of images

- …

. HDFLook project (LOA-USTL) (MODIS October 2 2002 [18h10] ((Hurricane Hernan (Baja Cali



Thanks !