École Doctorale des Sciences de l'Environnement d'Île-de-France
Année 2008-2009

Modélisation Numérique de l'Écoulement Atmosphérique et Assimilation d'Observations

Olivier Talagrand
Cours 8
25 Juin 2009

Weak constraint variational assimilation allows for errors in the assimilating model

Data

- Background estimate at time 0

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

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

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

- Evolution equation

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

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

Then objective function

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

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

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

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

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

Can include nonlinear M_k and/or H_k.

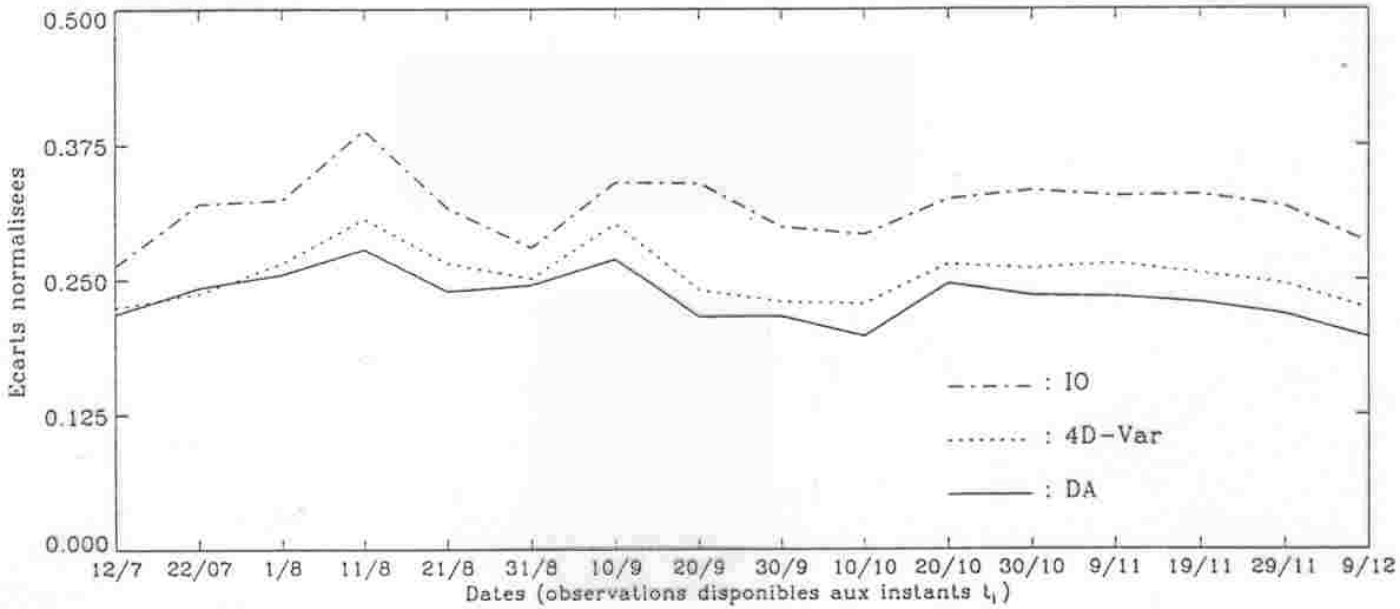


FIG. 9.11 – *Ecarts normalisés prévision/observations sur l'ensemble de la période étudiée*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999 nous permet d'observer les performances des différentes techniques d'assimilation.

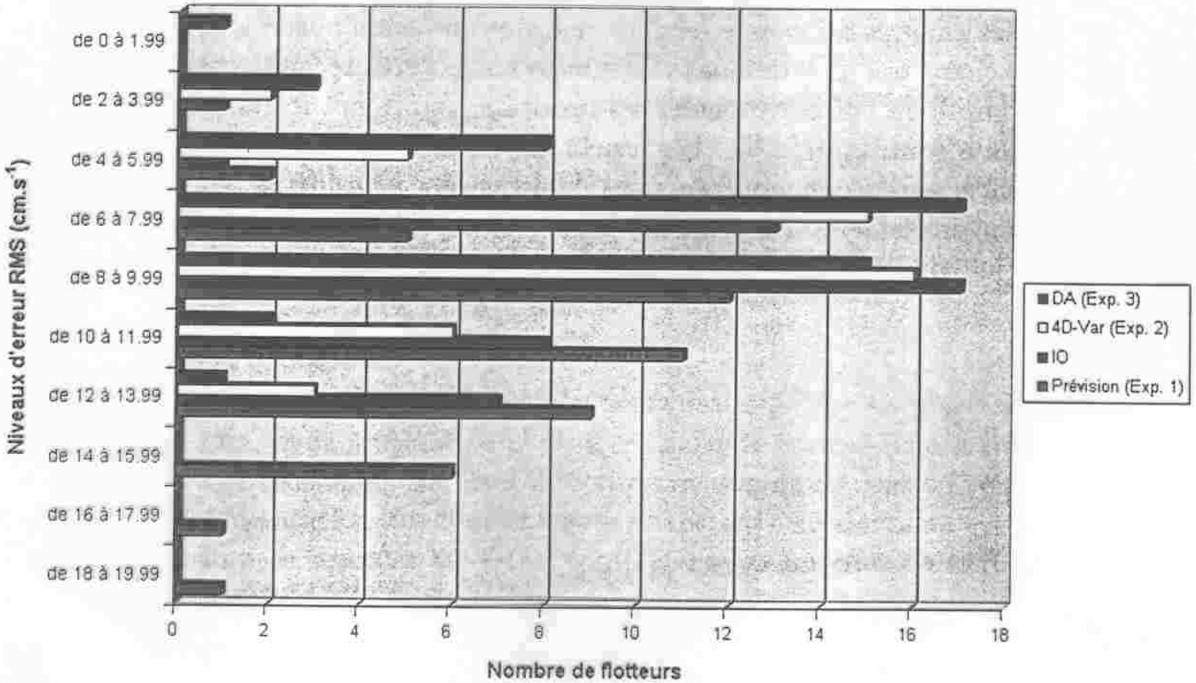


FIG. 9.15 – *Description des écarts flotteurs/modèle en terme de vitesse (à 150 m de profondeur) pour les différents algorithmes d'assimilation*

Louvel, Doctoral Dissertation, Université Paul-Sabatier, Toulouse, 1999

Sequential assimilation. Difficulty: carry in time uncertainty on the state of the flow.

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

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

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

  In the linear case, if errors are gaussian, and if P^b real (not sample) background error covariance matrix, (2) achieves Bayesian estimation, in the sense that $\{x^a{}_n\}$ is a sample of conditional probability distribution for x, given all data up to time t.

Assimilation ensemble tend to collapse (for reasons which are not fully understood).

Empirical remedies :

- Covariance inflation

- ‘Localization’ in physical space. Multiply covariances by function which becomes 0 at finite range (multiplying function must itself be of positive type).

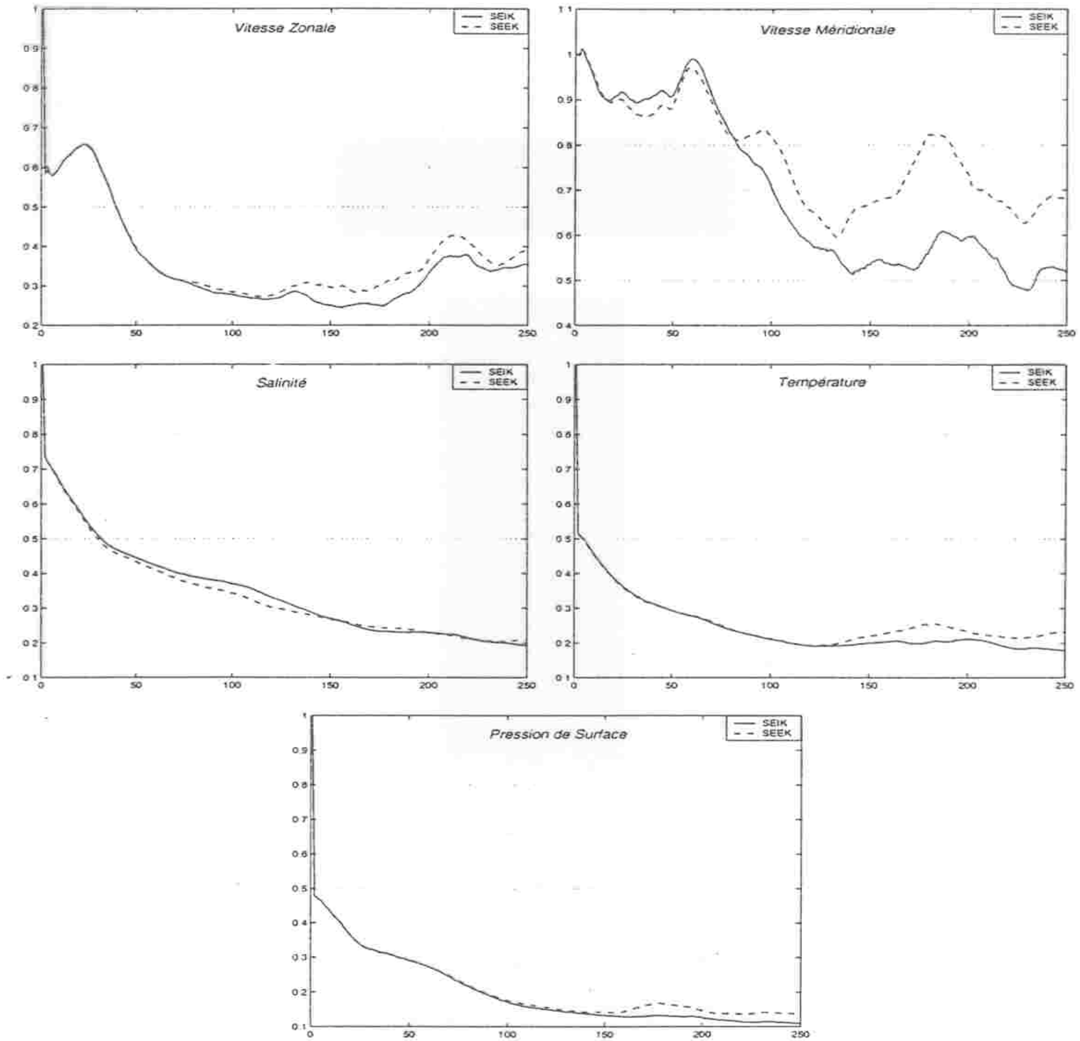


FIG. 5.2 – *Evolution dans le temps de la RRMS des filtres SEIK et SEEK*

I. Hoteit, Doctoral Dissertation, Université Joseph Fourier, Grenoble, 2001

Other approach

Ensemble Transform Kalman Filter (*ETKF*, Bishop *et al*.)
Does not require ‘perturbation’ of observations

Further evolution to *Local Ensemble Transform Kalman Filter* (*LETKF*, Kalnay *et al*.)
Analysis is performed locally in physical space

In any case, optimality always requires errors to be independent in time.

If errors are not gaussian, does not in general achieves bayesianity.

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

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

Defines updating of weights

Remarks

- Many variants exist, including possible 'regeneration' of ensemble elements
- If errors are correlated in time, explicit computation of $P(y \mid x^b{}_n)$ will require using past data that are correlated with y (same remark for evolution of ensemble between two observation times)

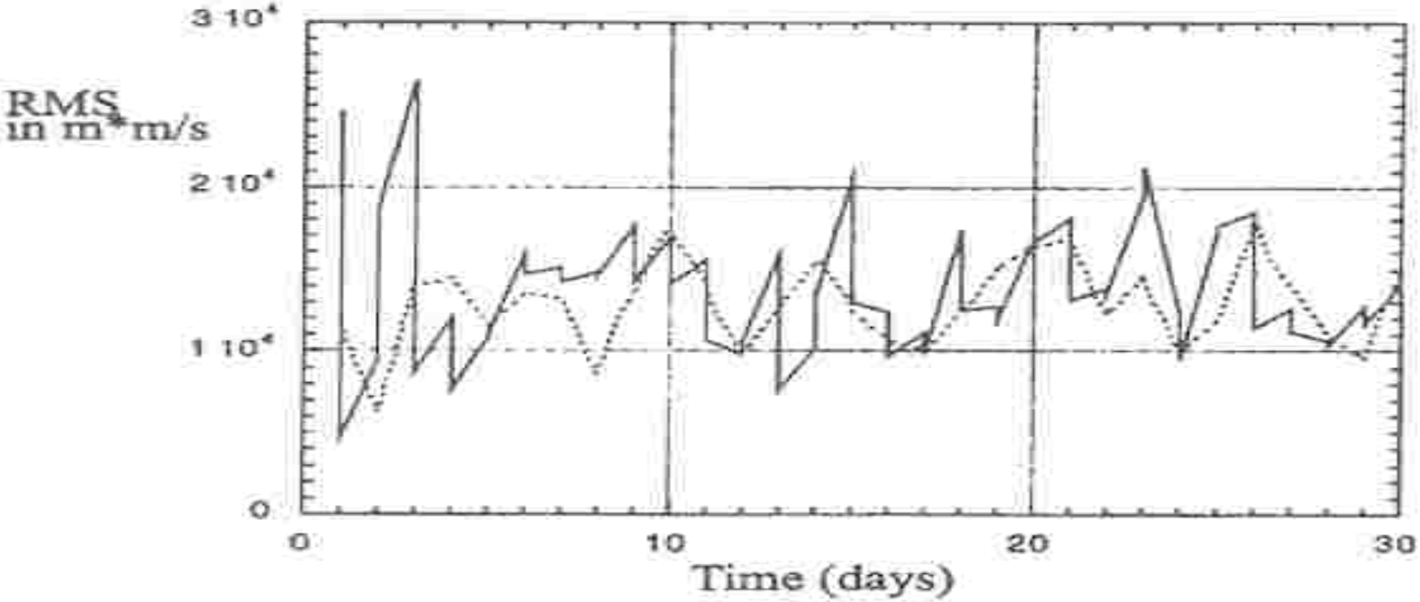


FIG. 12. Comparison of rms error ($m^2\ s^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

van Leeuwen, 2003, *Mon. Wea. Rev.*, **131**, 2071-2084

Exact bayesian estimation

Acceptation-rejection

Bayes' formula

$$f(x) \equiv \mathrm{P}(x \mid y) = P(y \mid x)\, P(x) \,/\, P(y)$$

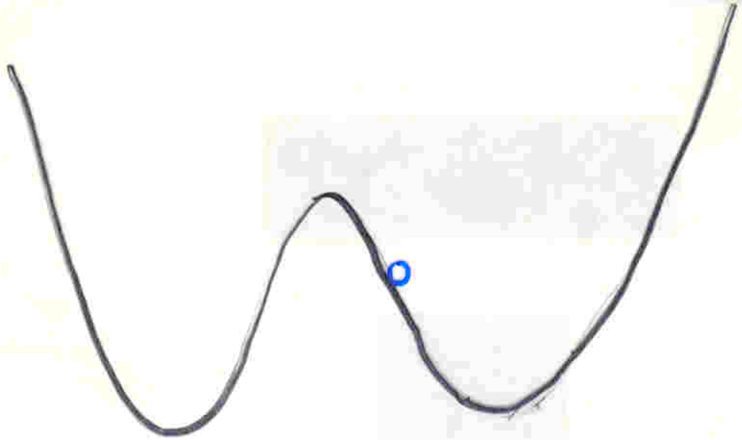
defines probability density function for x.

Construct sample of that pdf as follows.

Draw randomly couple $(\xi, \psi) \in \mathcal{S}$ x $[0, \sup f]$.

Keep ξ if $\psi < f(\xi)$. ξ is then distributed according to $f(x)$.

Miller, Carter and Blue, Tellus, 1999



$$\frac{d^2x}{dt^2} = -\frac{d\phi}{dx} - \alpha\frac{dx}{dt} + \text{Noise}$$

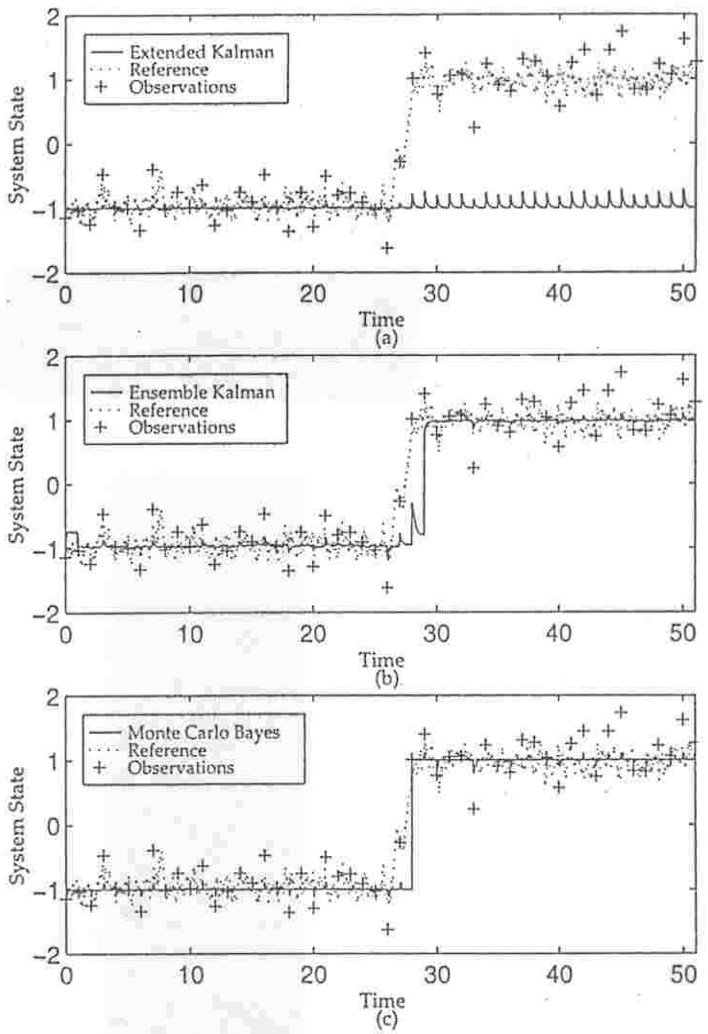


Fig. 4. Comparison of the EKF, the ensemble method and nonlinear filtering by Bayes' theorem for the double-well problem.

Miller, Carter and Blue, 1999, *Tellus*, **51A**, 167-194

Acceptation-rejection

Seems costly.

Requires capability of permanently interpolating probability distribution defined by finite sample to whole state space.

Conclusion on Sequential Assimilation

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)
Optimality is possible only if errors are independent in time

- Ensemble sequential assimilation is most promising.

- Is it worth the trouble (and the cost) of implementing non-Gaussian updating methods ? Not clear. Will presumably depend of particular problem at hand.

- Seems desirable to evaluate predicted ensembles as ensembles (as is done for ensemble prediction), and not only through the accuracy of their mean.

Variational Assimilation

Pros

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

Can 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.

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

- Oceanography
- Atmospheric chemistry (both troposphere and stratosphere)
- Oceanic biogeochemistry
- Ground hydrology
- Terrestrial biosphere and vegetation cover
- Glaciolology
- 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)
- …

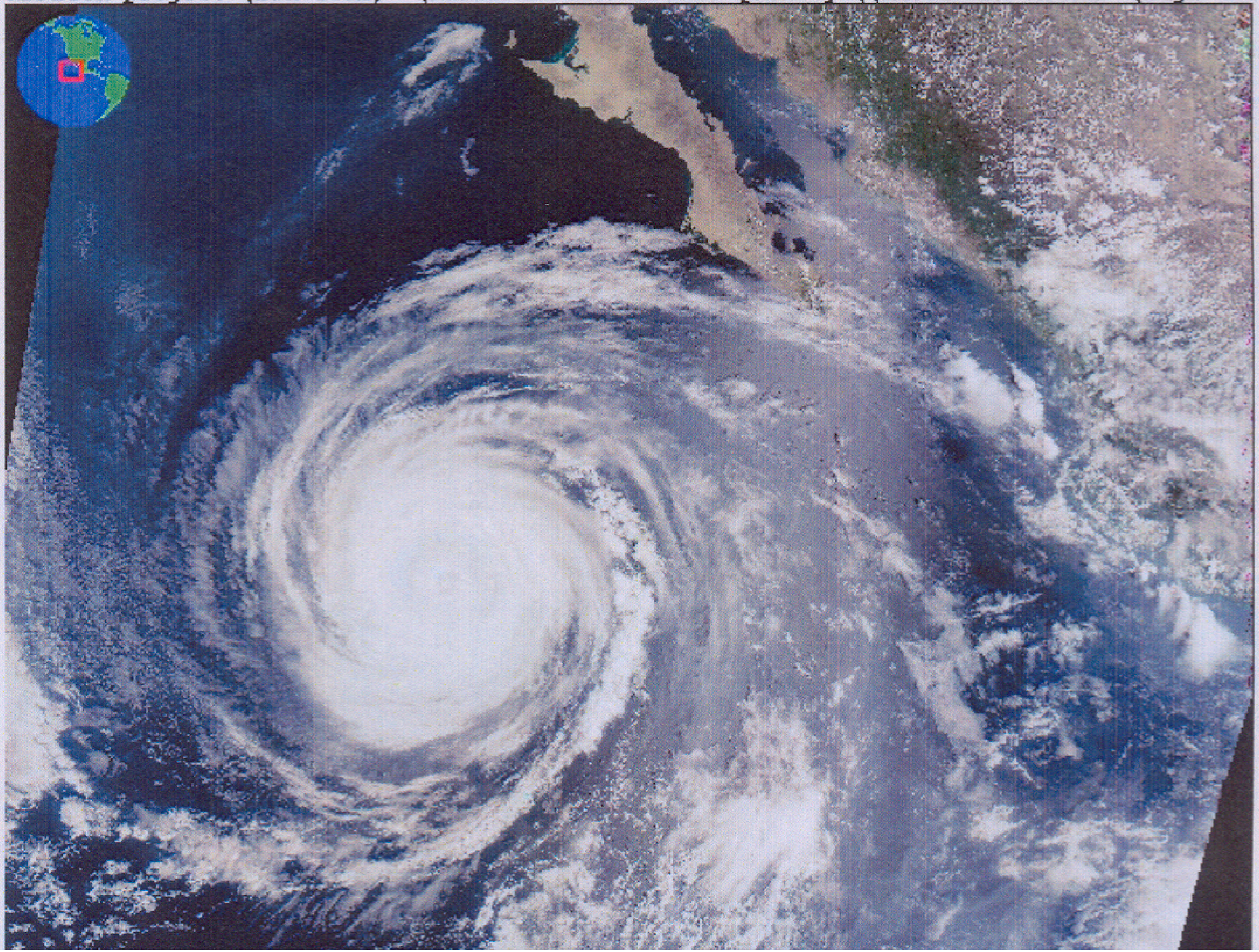
Assimilation is related to

- Estimation theory
- Probability theory
- Atmospheric and oceanic dynamics
- Atmospheric and oceanic predictability
- Instrumental physics
- Optimisation theory
- Control theory
- Algorithmics and computer science
- …

A few of the (many) remaining problems :

- Observability (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
- …

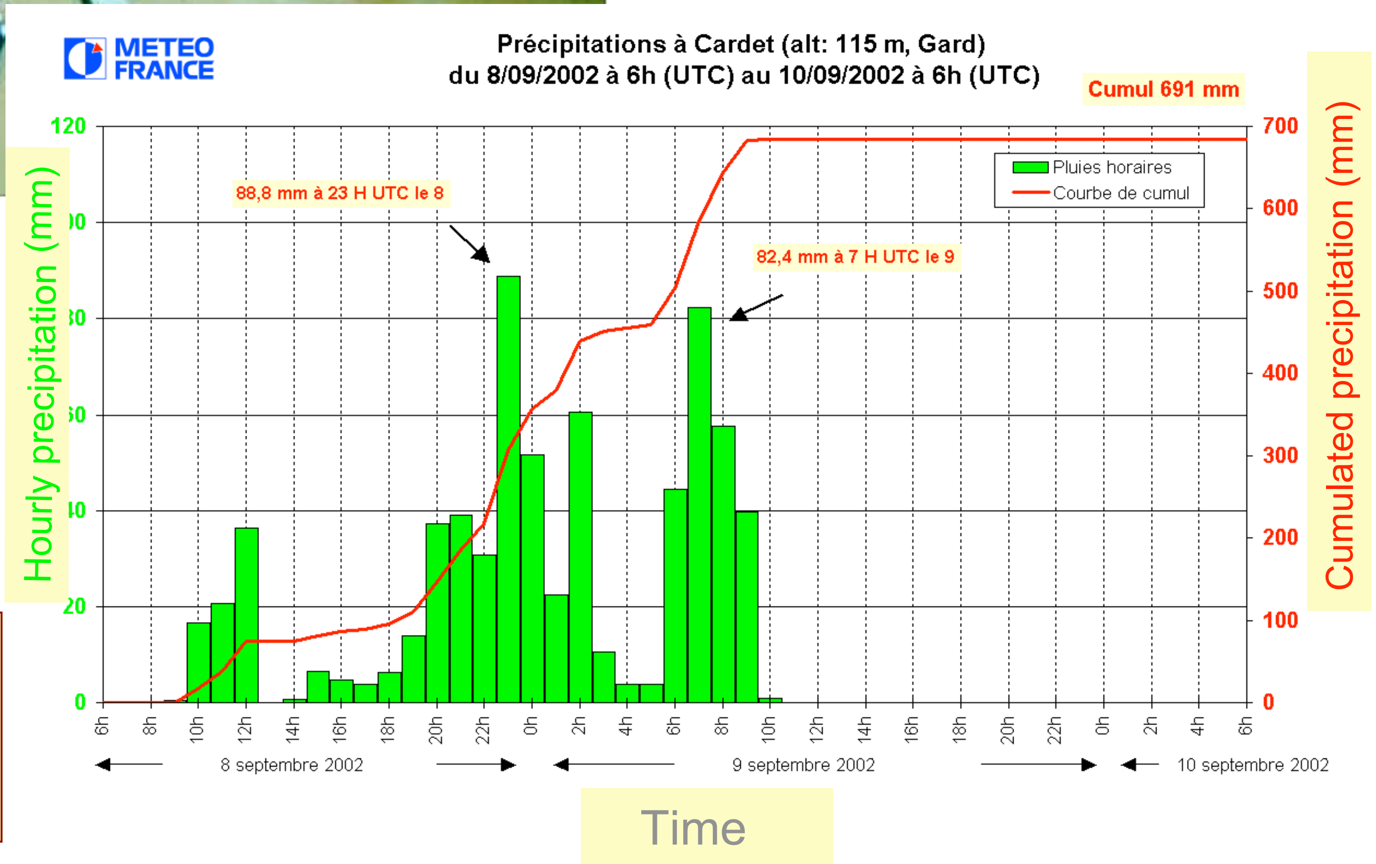
. HDFLook project (LOA-USTL) (MODIS October 2 2002 [18h10] ((Hurricane Hernan (Baja Cali



8-9 September 2002 :

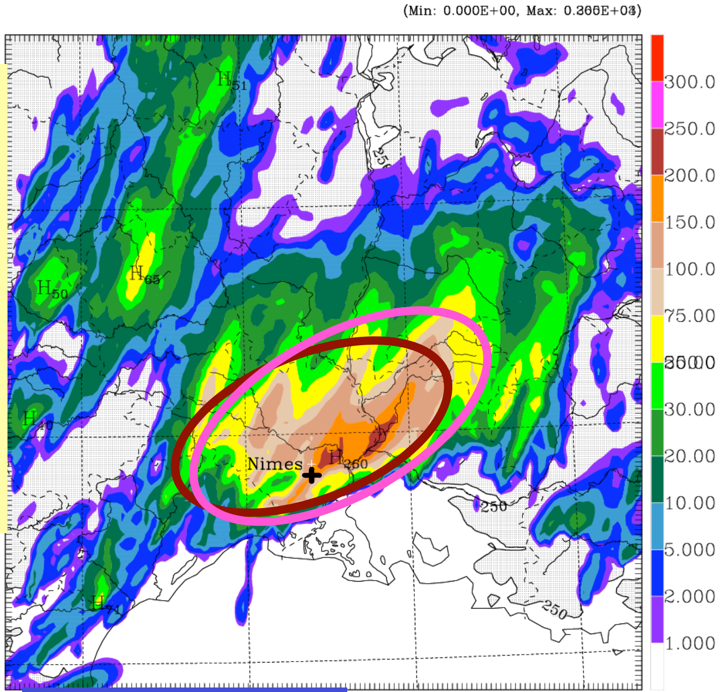
More than 20 deaths

Economic loss : 1,6 billion €

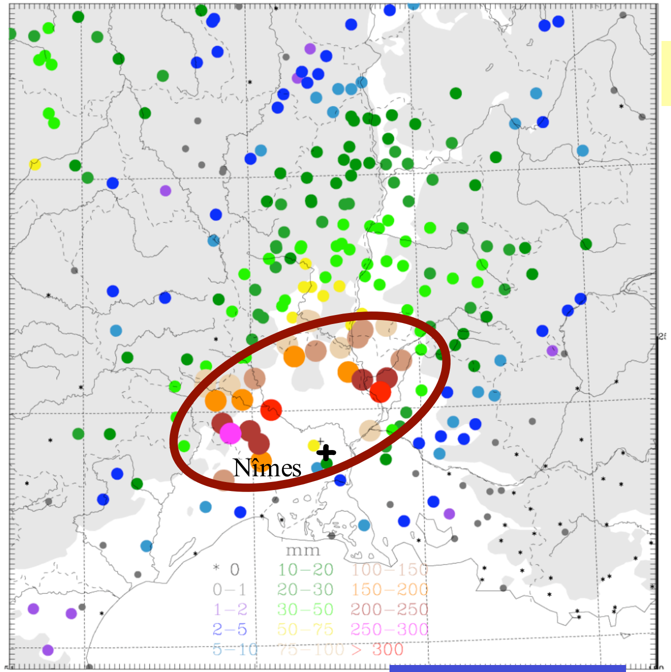


The accumulated rainfall totals have locally reached 600-700 mm

Initial conditions : mesoscale analysis (surface obs, radar, satellite) for 12UTC, 8th Sept. 2002



MESO-NH (2.5km)

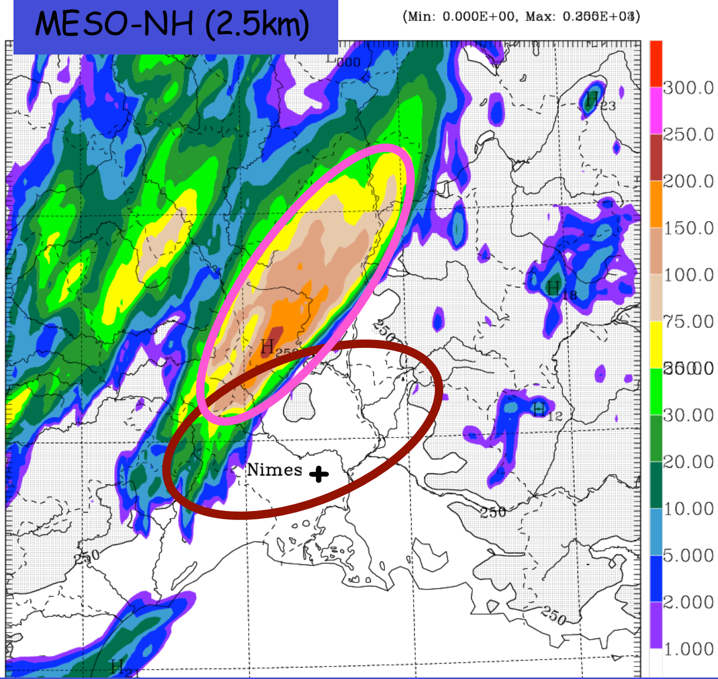


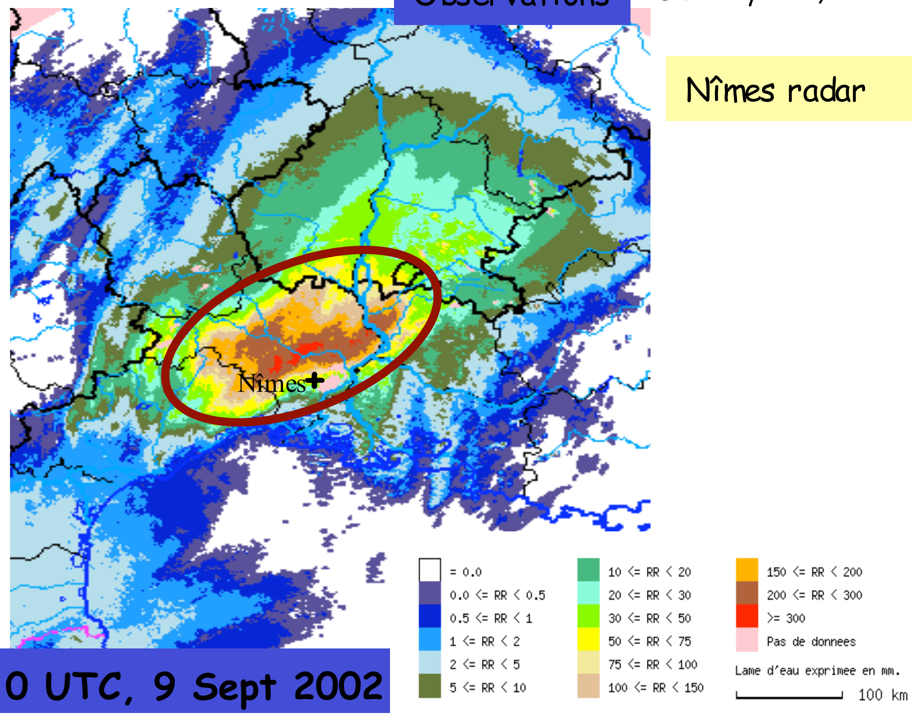
Raingauges

Observations

Ducrocq et al, 2003

Initial conditions : large scale ARPEGE analysis for 12UTC, 8th Sept. 2002



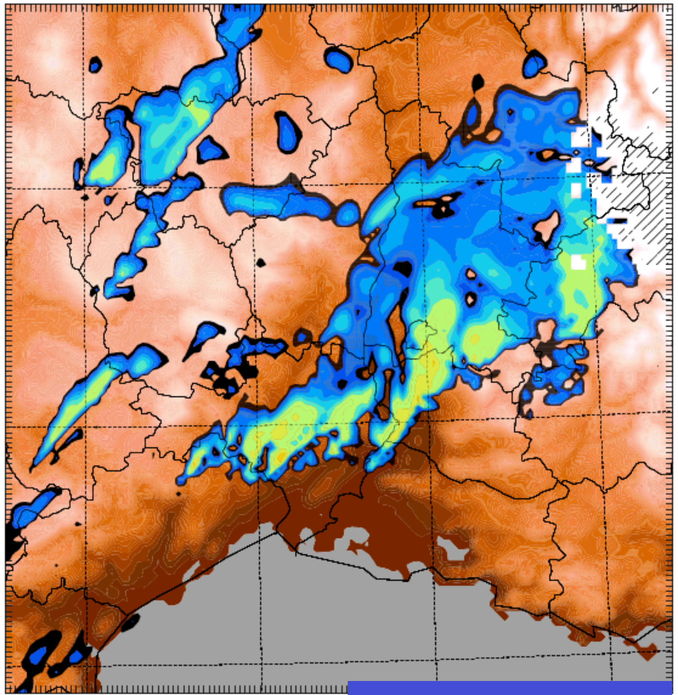


Nîmes radar

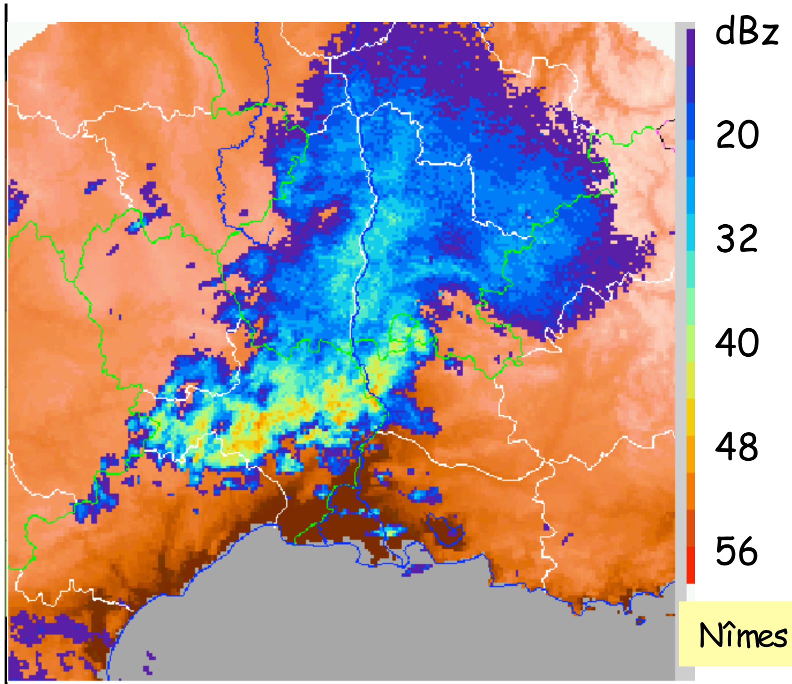
12-h accumulated rainfall from 12 UTC, 8 Sept to 0 UTC, 9 Sept 2002

Initial conditions : mesoscale analysis (surface obs, radar, satellite) for

12UTC, 8th Sept. 2002

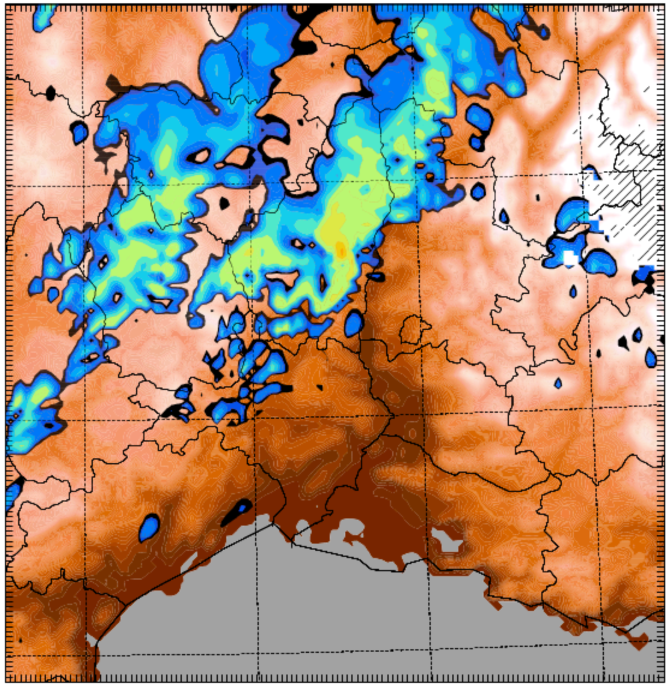


MESO-NH (2.5km)



Nîmes radar

Observations



Initial conditions : large scale ARPEGE analysis for

12UTC, 8th Sept. 2002

Reflectivities at 21 UTC, 8 Sept.