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 8

12 Mai 2022

Last course (May 5)

- Assimilation dans l'espace instable

-Filtres particulaires

-Assimilation Variationnelle d'Ensemble

This course (May 12)

- A few Basics about Dynamical Systems. Lyapunov exponents

- A few variants of Filters and Smoothers, among many

- Artificial Intelligence and Data Assimilation

- Conclusions and Perspectives

Dynamical system

State vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. Evolves in time according to equation

$$d\boldsymbol{x}/dt = \boldsymbol{F}(\boldsymbol{x}) \qquad (1)$$

or, componentwise

$$dx_i/dt = F_i(\boldsymbol{x}), \quad i = 1, \ldots, n$$

Purely deterministic (no stochastic component)

$$d\boldsymbol{x}/dt = \boldsymbol{F}(\boldsymbol{x}) \qquad (1)$$

Initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ defines unique solution (or *orbit*)

$$\boldsymbol{x}(t) = \boldsymbol{R}(t, t_0)\ (\boldsymbol{x}_0)$$

$\boldsymbol{R}(t, t_0)$ is the *resolvent* of Eq. (1) between times t_0 and t.

System can be discretized in time

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

Typical questions about dynamical systems

- *Stationary points* ($\boldsymbol{F}(\boldsymbol{x}) = 0$) and *associated stability ?*

- *Stability of orbits ?*

- *Long term behaviour of orbits* (convergence to fixed points, periodicity, convergence to limit cycle, divergence to infinity , non-periodic oscillations, …) *?*

- *Uncertainty in initial conditions. How does it evolve ?*

$$d\boldsymbol{x}/dt = \boldsymbol{F}(\boldsymbol{x}) \qquad (1)$$

Solution $\boldsymbol{x}(t)$. Perturbation $\delta\boldsymbol{x}(t)$. Evolves according to

$$d\delta\boldsymbol{x}/dt = \boldsymbol{F}[\boldsymbol{x}(t) + \delta\boldsymbol{x}] - \boldsymbol{F}[\boldsymbol{x}(t)] \approx \mathbf{F'}(t)\ \delta\boldsymbol{x}$$

where $\mathbf{F'}(t)$ is *Jacobian* (matrix of partial derivatives) of operator $\boldsymbol{F}$ at point $\boldsymbol{x}(t)$

$$d\delta\boldsymbol{x}/dt = \mathbf{F'}(t)\ \delta\boldsymbol{x} \qquad \text{(TLM)}$$

is *tangent linear system* of system (1) along solution $\boldsymbol{x}(t)$. Describes evolution of perturbation $\delta\boldsymbol{x}$ on $\boldsymbol{x}(t)$ to first order wrt initial value of perturbation.

$\delta\boldsymbol{x}(t) = \boldsymbol{F}[\boldsymbol{x}(t)]$ is solution of (TLM)

$$d\delta\boldsymbol{x}/dt = \mathbf{F}'(t)\ \delta\boldsymbol{x} \qquad \text{(TLM)}$$

Adjoint equation

$$d\lambda/dt = -\ [\mathbf{F}'(t)]^{T}\ \lambda \qquad \text{(ADJ)}$$

For system discretized in time

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

$$\delta\boldsymbol{x}_{k+1} = \mathbf{M}_k'\,\delta\boldsymbol{x}_k \qquad \text{(TLM)}$$

Adjoint

$$\lambda_k = [\mathbf{M}_k']^{\mathrm{T}}\,\lambda_{k+1} \qquad \text{(ADJ)}$$

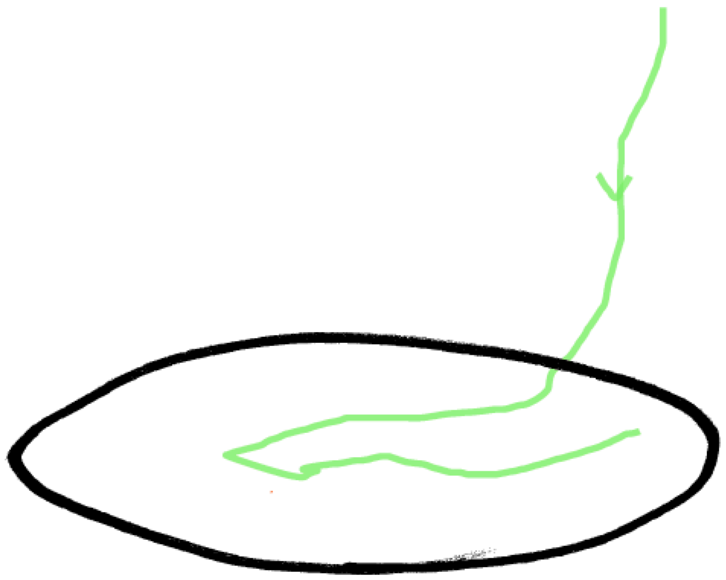
Lorenz (1963)

$$dx/dt = \sigma(y-x)$$

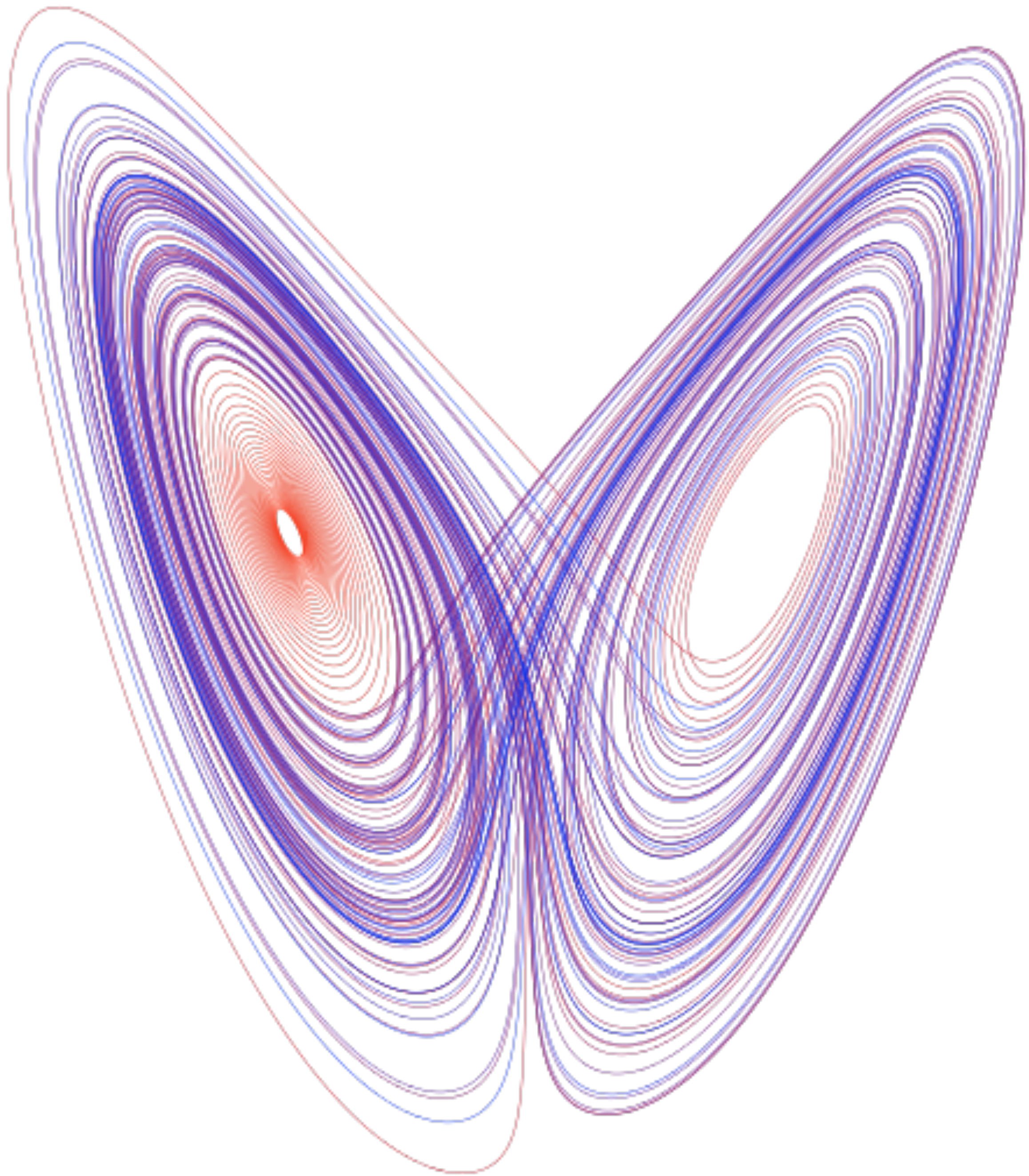
$$dy/dt = \rho x - y - xz$$

$$dz/dt = -\beta z + xy$$

with parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ $\Rightarrow$ chaos



All orbits end up trapped in the same neighbourhood, within which they have accumulation points, which consist themselves of full orbits



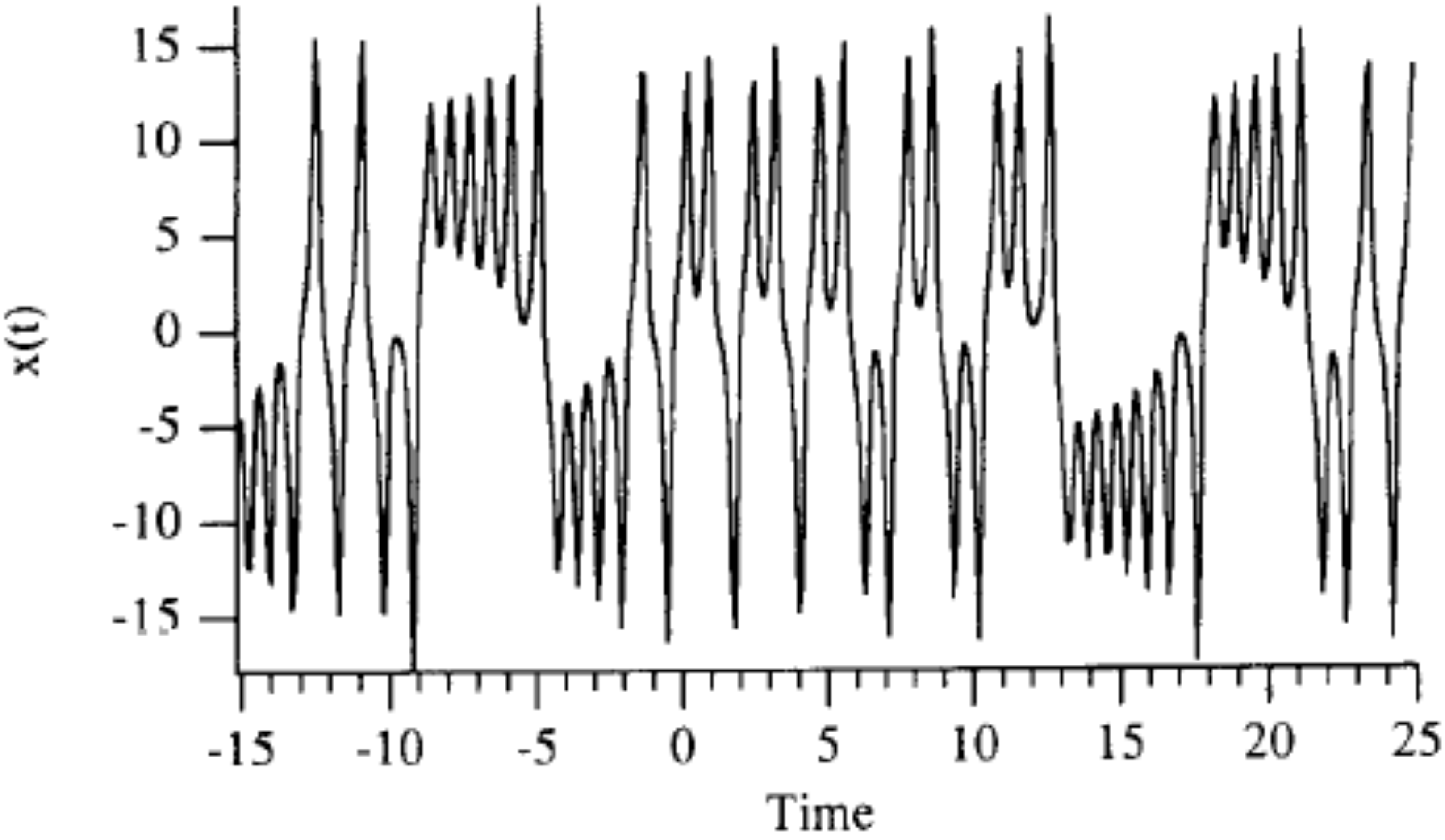


Fig. 2. Time variations, along the reference solution, of the variable $x(t)$ of the Lorenz system.

Probability Density Function (*PDF*) $p(\boldsymbol{x}, t)$ for state vector. Evolves in time according to equation

$$Dp/Dt + p\ \mathrm{div}\boldsymbol{F} = 0$$

which expresses conservation of probability in the flow F. It is fundamentally the same equation as the 'continuity' equation, which expresses conservation of mass in physical motion. It is called in the present context the *Liouville equation*.

$dx/dt = \sigma(y - x)$

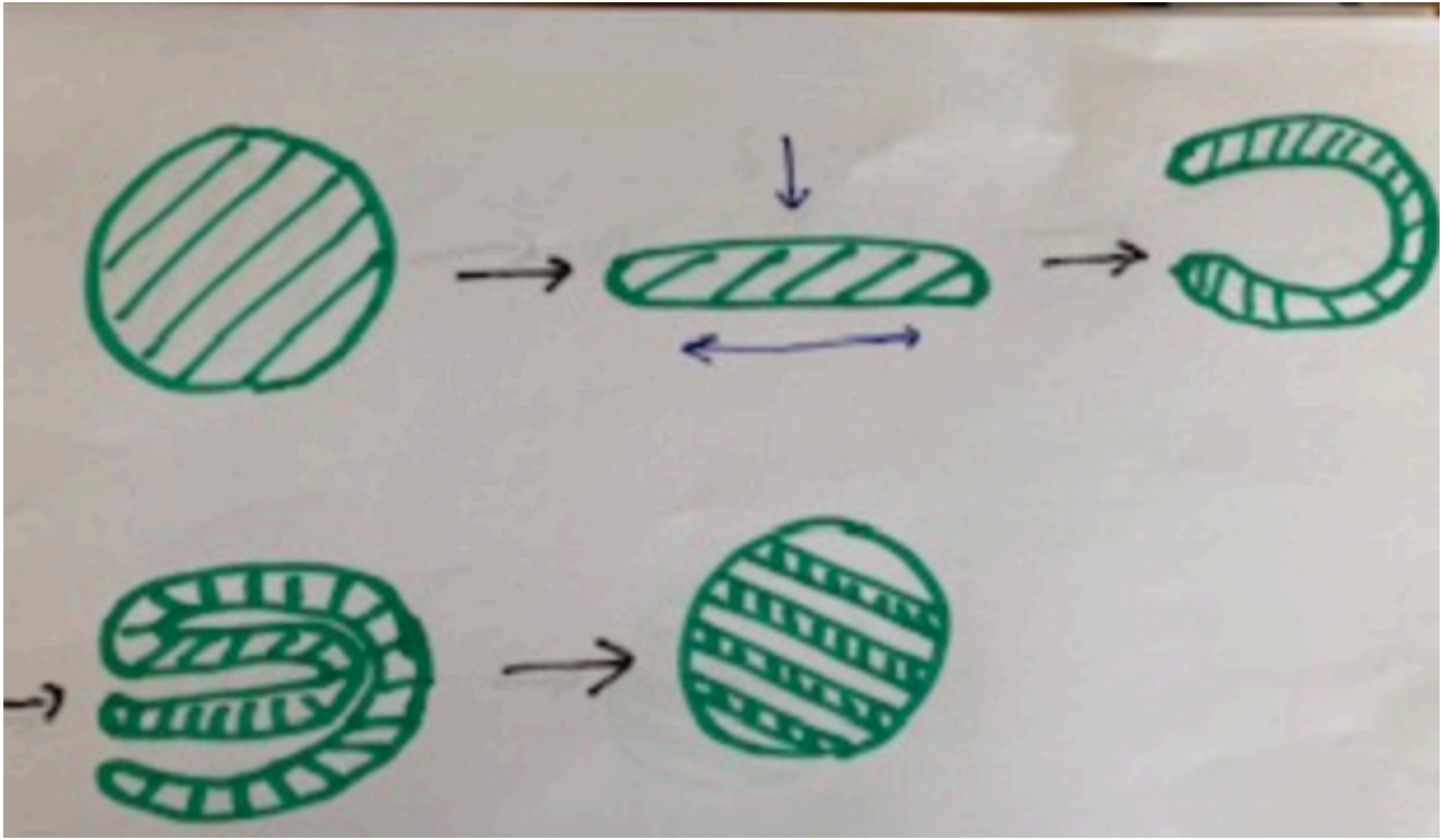
$dy/dt = \rho x - y - xz$

$dz/dt = -\beta z + xy$

with parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ ($\Rightarrow$ chaos)

$$\mathrm{div}F = \partial(dx/dt)/\partial x + \partial(dy/dt)/\partial y + \partial(dz/dt)/\partial z$$
$$= -(\sigma + 1 + \beta) = -13.666\ldots < 0$$

Volume element $V(t) = V(0) \exp[-13.67\, t]$ decreases exponentially with time



Loss of predictability in dissipative chaos

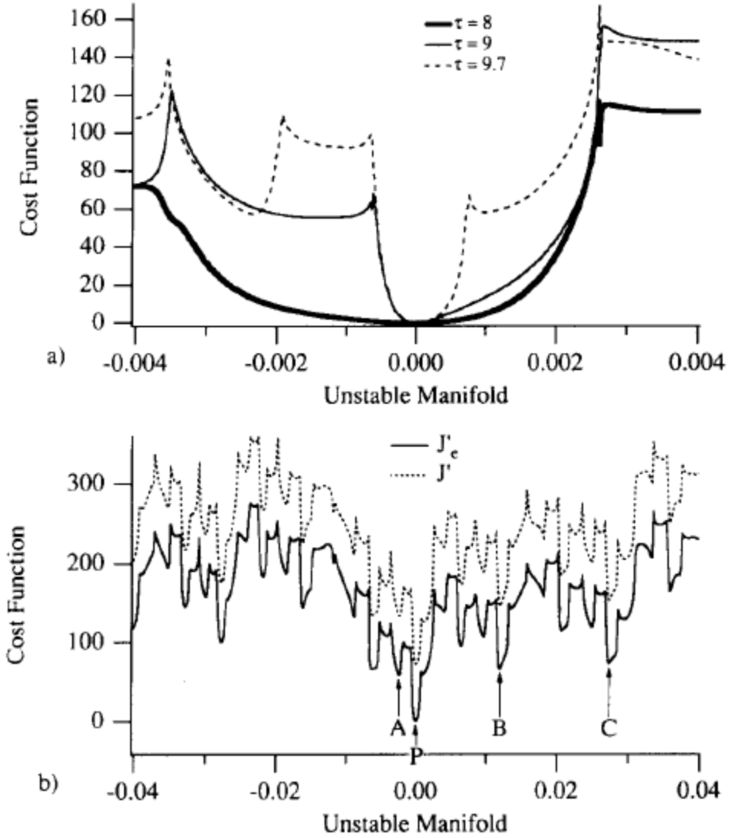


Fig. 4. Panel (a): Cross-section of the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ along the unstable manifold, for various values of τ. Panel (b). As in panel (a), for $\tau = 9.7$, and with a display interval ten times as large, respectively for the error-free forward cost-function $J'_e(\tau, \hat{x}, x)$ (solid curve) and for the error-contaminated cost-function $J_e(\tau, \hat{x}, x)$ (dashed curve). In the latter case, the total variance of the observational noise is $E^2 = 75$.

Pires *et al., Tellus*, 1996 ; Lorenz system (1963)

Linear constant coefficient system with dimension n

$$d\boldsymbol{x}/dt = \boldsymbol{A}\boldsymbol{x}$$

Eigenvectors $\boldsymbol{e}_j,\ j = 1, \ldots, n$

Eigenvalues $\mu_j = \lambda_j + i\nu_j,\ \lambda_1 > \ldots > \lambda_n$

$$\boldsymbol{x}(t_0) = \Sigma_j\, x_j(t_0)\, \boldsymbol{e}_j$$

$$\begin{aligned} \boldsymbol{x}(t_0+\tau) &= \Sigma_j \exp(\mu_j \tau)\, x_j(t_0)\, \boldsymbol{e}_j \\ &= \exp(\mu_1 \tau)\, x_1(t_0) \\ &\qquad \text{x } [\boldsymbol{e}_1 + \Sigma_{j>1} \exp((\mu_j - \mu_1)\tau)\, x_j(t_0)\, \boldsymbol{e}_j] \end{aligned}$$

$$\boldsymbol{x}(t_0+\boldsymbol{\tau}) \;=\; \exp(\mu_1 \boldsymbol{\tau})\, x_1(t_0)\, [\boldsymbol{e}_1 + o(1)]$$

$$\| \boldsymbol{x}(t_0+\boldsymbol{\tau}) \| = \exp(\lambda_1 \boldsymbol{\tau})\; |x_1(t_0)|\; \| \boldsymbol{e}_1 \|$$

$$\lim_{t \to \infty} [(1/t) \ln \| \boldsymbol{x}(t) \|] = \lambda_1$$

If $x_1(t_0) = 0$, $x_2(t_0) \neq 0$, the limit is λ_2, and so on …

There exist a sequence of real numbers (real parts of eigenvalues of matrix $\boldsymbol{A}$)

$$\lambda_1 > \ldots > \lambda_m \qquad m \leq n$$

and sequence of (sub)spaces of $\mathcal{R}^n$

$$\mathcal{E}_{m+1} = \varnothing \subset \mathcal{E}_m \subset \ldots \subset \mathcal{E}_j \subset \ldots \subset \mathcal{E}_1 = \mathcal{R}^n$$

such that $\lim_{t \to \infty} [(1/t) \ln \| \boldsymbol{x}(t) \|] = \lambda_j$ when $\boldsymbol{x}(t_0) \in \mathcal{E}_j / \mathcal{E}_{j+1}$

The same is fundamentally true for dynamical systems with attractors (solutions constantly return to the vicinity of same points → *ergodicity*)

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

Solution $\boldsymbol{x}(t)$. Associated TLM

$$d\delta\boldsymbol{x}/dt = \mathbf{F}'(t)\ \delta\boldsymbol{x} \qquad \text{(TLM)}$$

There exist a sequence of real numbers (*Lyapunov exponents*)

$$\lambda_1 > \ldots > \lambda_m \qquad m \leq n$$

and sequence of (sub)spaces of $\mathcal{R}^n$

$$\mathcal{E}_{m+1} = \varnothing \subset \mathcal{E}_m \subset \ldots \subset \mathcal{E}_j \subset \ldots \subset \mathcal{E}_1 = \mathcal{R}^n$$

such that $\lim_{t \to \infty} [(1/t) \ln \| \delta\boldsymbol{x}(t) \|] = \lambda_j$ when $\delta\boldsymbol{x}(t_0) \in \mathcal{E}_j / \mathcal{E}_{j+1}$

Lorenz 1963

$dx/dt = \sigma(y-x)$

$dy/dt = \rho x - y - xz$

$dz/dt = -\beta z + xy$

with parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ ($\Rightarrow$ chaos)

Lyapunov exponents

0.9056, 0, −14.5723 (sum = -13.6667)

[div$F$ = - (σ +1 +β) = - 41/3 = - 13.6666.. < 0]

Lyapunov exponents

$$0.9056,\ 0, -14.5723$$

Lyapunov exponents measure rate of growth of perturbations, averaged over the whole attractor

Presence of at least one positive Lyapunov exponent is signature of chaos.

In an ergodic system, one exponent is equal to 0. It corresponds to perturbations in the direction of the motion, which will be neither amplified nor damped over long periods.

Experiments performed on the Lorenz (1996) model

$$\frac{d}{dt}x_j = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F$$

with $j = 1, ..., I.$

with periodic conditions in j, and value $F = 8$, which gives rise to chaos.

Three values of I have been used, namely $I = 40, 60, 80$, which correspond to respectively $N^+ = 13$, 19 and 26 positive Lyapunov exponents.

In all three cases, the largest Lyapunov exponent corresponds to a doubling time of about 2 days (with 1 'day' = 1/5 model time unit).

Identical twin experiments (perfect model)

Lyapunov exponents

$$\lambda_1 > \ldots > \lambda_m \qquad m \leq n$$

associated (sub)spaces of $\mathcal{R}^n$

$$\mathcal{E}_{m+1} = \emptyset \subset \mathcal{E}_m \subset \ldots \subset \mathcal{E}_j \subset \ldots \subset \mathcal{E}_1 = \mathcal{R}^n$$

$$\lim_{t \to \infty} [(1/t) \ln \| \delta \boldsymbol{x}(t) \|] = \lambda_j \text{ when } \delta \boldsymbol{x}(t_0) \in \mathcal{E}_j / \mathcal{E}_{j+1}$$

Modulus $\| \delta \boldsymbol{x}(t) \|$ depends on choice of norm, but asymptotic exponential rate of growth or decay does not. Lyapunov exponents do not depend on position on orbit, and are the same for all orbits with the same attractor. Subspaces $\mathcal{E}_j$ depend on position on orbit, but evolve with the motion.

Lyapunov vectors

At a given point along an orbit, *forward Lyapunov vectors* are vectors which will concentrate most rapidly on the Lyapunov rate of growth or decay.

Similarly *bacward Lyapunov vectors* are vectors that have concentrated most rapidly in the past on the Lyapunov rate of growth or decay. In assimilation, they tend to dominate the background error.

These vectors depend on the choice of a norm, are orthogonal with respect to the chosen norm, and do not evolve with the motion.

One also defines *covariant Lyapunov vectors*, which are exactly amplified or damped according to the Lyapunov exponents, and evolve with the motion. They do not depend on the choice of a norm, and are not orthogonal wrt to a time-independent norm.

The notions of Lyapunov exponents and vectors have turned out to be very useful for the study of the dynamics of the atmosphere and the ocean, They are relatively easy to determine (identifying them does not require long numerical integrations, which means that the atmosphere and the ocean have in a sense 'good ergodicity'). They more or less explicitly underlie the approach of Assimilation in the Unstable Subspace

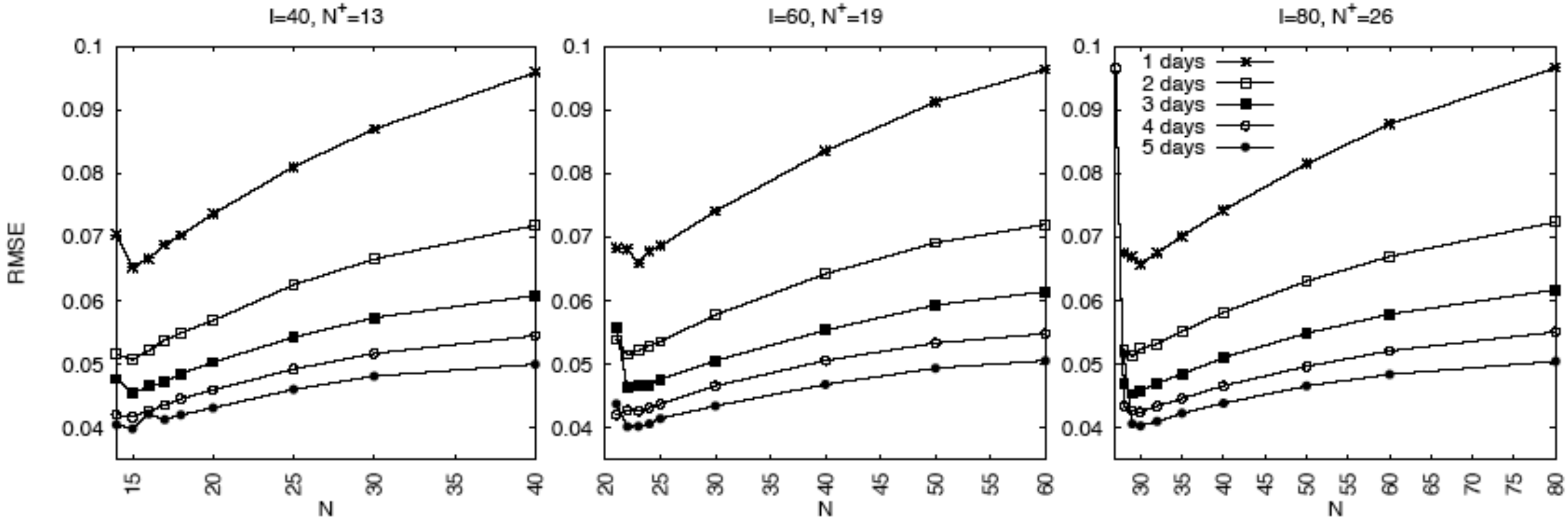


Figure 1. Time average RMS analysis error at $t = \tau$ as a function of the subspace dimension N for three model configurations: I=40, 60, 80. Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is $\sigma_o = 0.2$.

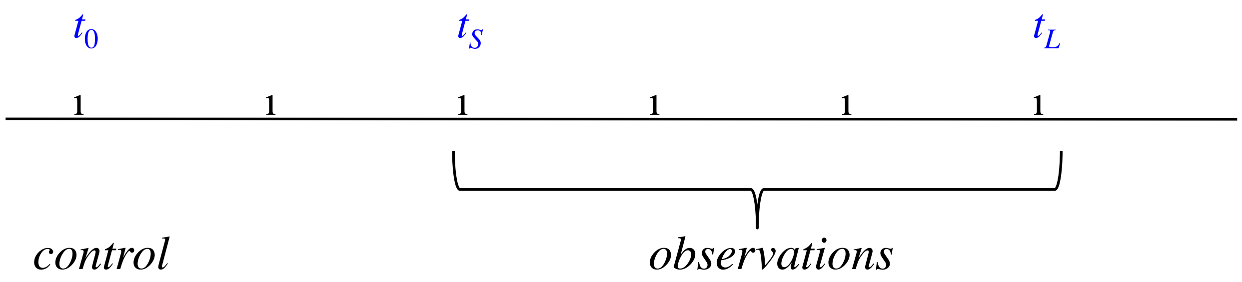
No explicit background term (*i. e.*, with error covariance matrix) in objective function : information from past lies in the background to be updated, and in the N perturbations which define the subspace in which updating is to be made.

Best performance for N slightly above number N^+ of positive Lyapunov exponents.

Trevisan *et al*., 2010

Iterative Ensemble Kalman Smoother (*IEnKS*, Bocquet and Sakov, 2014)

Minimization performed at time t_0, in an appropriately chosen reduced subspace, assimilating observations performed between times t_S and t_L, with $t_0 \leq t_S \leq t_L$



If the dimension of the reduced subspace is small enough, gradient of objective function can be computed by finite differences, and approximate Hessian can be determined. Once the minimization has been achieved, a new ensemble of perturbations can be obtained by transport of the approximate inverse Hessian.

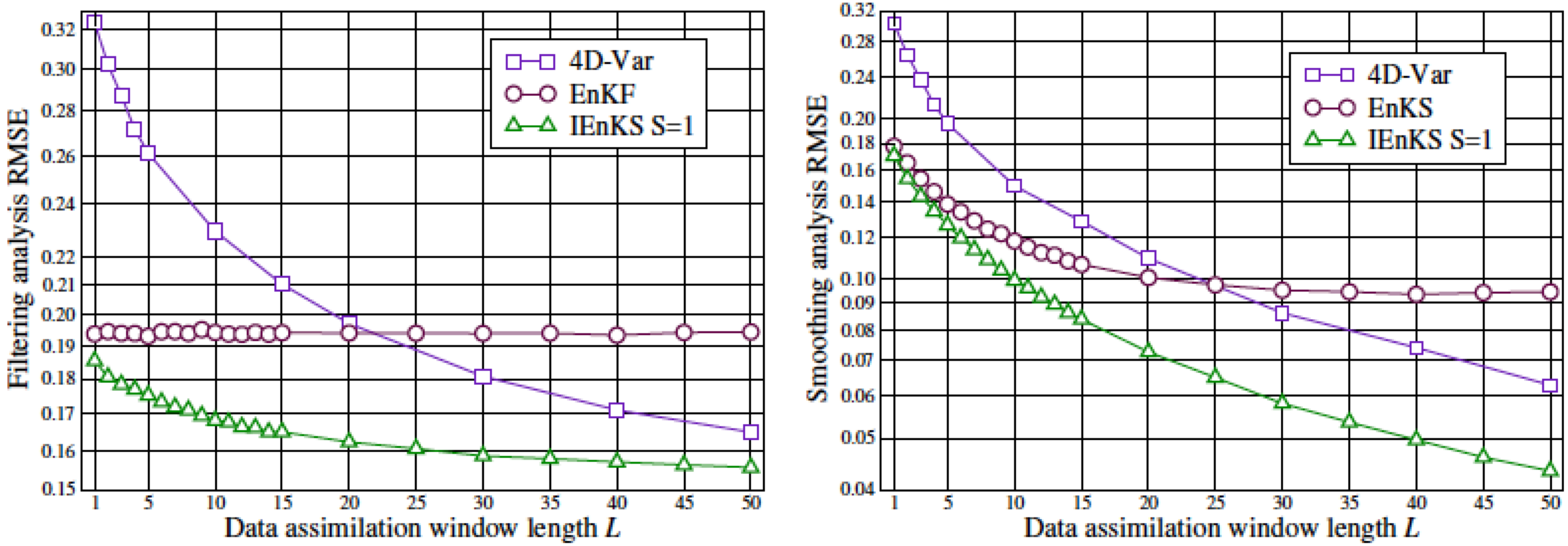


Figure 5: Average root mean square error of several DA methods computed from synthetic experiments with the Lorenz-96 model. The left panel shows the filtering analysis root mean square error of optimally tuned EnKF, 4DVar, IEnKS assimilation experiments, as a function of the length of the DAW. The right panel shows the smoothing analysis root mean square error of optimally tuned EnKS, 4DVar and IEnKS as a function of the length of their data assimilation window. The optimal RMSE is chosen within the window for 4DVar and it is taken at the beginning of the window for the IEnKS. The EnKF, EnKS and IEnKS use an ensemble of $N = 20$, which avoids the need for localization but requires inflation. The length of the DAW is $L \times \Delta t$, where $\Delta t = 0.05$.

Carrassi *et al.*, 2018

Unknown $\boldsymbol{x}$ to be determined. Belongs to *state space* $\mathcal{S}$, with dimension n
Data, belonging to *data space* $\mathcal{D}$, with dimension m, available in the form

$$z = \boldsymbol{\Gamma x} + \zeta$$

where $\boldsymbol{\Gamma}$ is a known (mxn)-matrix, rank$\boldsymbol{\Gamma} = n$ and ζ is 'error'

Best Linear Unbiased Estimate (*BLUE*)

$$\boldsymbol{x}^a \equiv (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} [z - \boldsymbol{\mu}]$$

with $\boldsymbol{\mu} = E(\zeta)$ and $\boldsymbol{S} = E[(\zeta - E(\zeta)) (\zeta - E(\zeta))^{\mathrm{T}}]$.

$E(\boldsymbol{x}^a - \boldsymbol{x}) = 0$ $\qquad$ $E[(\boldsymbol{x}^a - \boldsymbol{x}) (\boldsymbol{x}^a - \boldsymbol{x})^{\mathrm{T}}] \equiv \boldsymbol{P}^a = (\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{\Gamma})^{-1}$

Determinacy condition : rank$\boldsymbol{\Gamma} = n$. Data contain information, directly or indirectly, on every component of state vector x. Requires $m \geq n$.

BLUE is invariant in any change of origin, or in any invertible linear transformation, in either data or state space. In particular, it is independent of the choice of a scalar product or norm in either of those spaces. *BLUE* minimizes the quadratic estimation error on any generalized component of $\boldsymbol{x}$.

If error $\boldsymbol{\zeta}$ is gaussian, $\zeta \sim \mathcal{N}[\boldsymbol{\mu}, S]$, *BLUE* achieves bayesian estimation in the sense that

$$P(x \mid z) = \mathcal{N}[x^a, P^a]$$

Any assumed probability distribution $P(\zeta)$ defines a conditional probability distribution $P(x \mid z)$ for x. In case the distribution $P(\zeta)$ is known only through its expectation $\boldsymbol{\mu}$ and covariance matrix S, the gaussian distribution $\mathcal{N}[\boldsymbol{\mu}, S]$ leads for $\boldsymbol{x}$ to the conditional probability distribution $P(x \mid z)$ with the largest entropy. The gaussian choice is in that sense the 'least-committing' choice.

BLUE is the simplest of non-simplicist algorithms.

The *BLUE* can be obtained by minimization of the following scalar objective function, defined on state space $\mathcal{X}$

$$\xi \in \mathcal{X} \rightarrow \mathcal{J}(\xi) \equiv (1/2)\,[\Gamma\xi - (z-\boldsymbol{\mu})]^{T} S^{-1} [\Gamma\xi - (z-\boldsymbol{\mu})]$$

And in case of nonlinearity ?

$$\boldsymbol{z} = \boldsymbol{\Gamma}(\boldsymbol{x}) + \boldsymbol{\zeta}$$

Variational approach can be heuristically implemented

$$\xi \in \mathcal{X} \rightarrow \mathcal{J}(\xi) \equiv (1/2)\,[\Gamma(\xi) - (z-\boldsymbol{\mu})]^{T} S^{-1}\, [\Gamma(\xi) - (z-\boldsymbol{\mu})]$$

It works !

If data are of the form (after possibly an appropriate transformation)

$$x^b = x + \zeta^b$$
$$y = H(x) + \varepsilon$$

Transformation

$$x^b = x + \zeta^b$$
$$y - H(x^b) = H(x) - H(x^b) + \varepsilon \approx H'(x - x^b) + \varepsilon$$

where H' is jacobian of H, makes the estimation problem linear in the deviation $x - x^b$ (*tangent linear approximation*)

All algorithms that have been presented in the course, with the exception of particle filters, are empirical heuristic extensions of the BLUE approach to approximate nonlinear and non-gaussian situations.

Artificial Intelligence

(aka ***Machine Learning*** or ***Deep Learning***)

Numerical modelling of the atmospheric and oceanic flow, as presented in the course, fundamentally built on known physical laws (conservation of mass, momentum and energy).

Why not directly use observations (for instance, in the case of a weather forecast, why not look for analogues in the past, and make the forecast from those analogues) ?

E. N. Lorenz (1960s). Sample of past observations will never be large enough for competing with physically-based models.

But :

- there is no incompatibility between the two approaches

- there remain many processes in numerical models which we do not know how to describe on the basis of well-established physical laws (interactions between atmosphere and underlying medium, such as *e.g.* vegetation, all kinds of subgrid scale processes, …)

- amount of data of all kinds, as well as computing power, are increasing very rapidly.

Artificial Intelligence (aka ***Machine Learning***) (continuation)

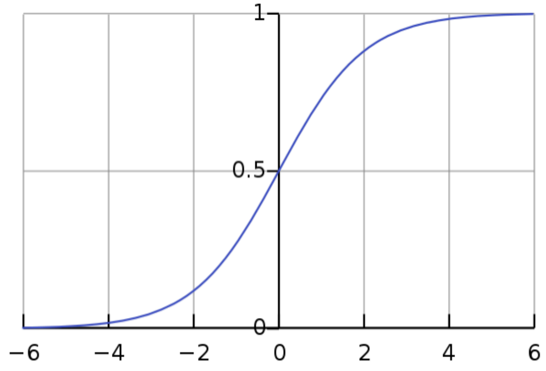
Powerful numerical tools have been developed for the exploitation of very large sets of data (*big data*)

Neural networks. Define an explicit numerical link between an *input set* and an *output set*. Define function F such that, to some useful degree of approximation

$$y = F(x)$$

where x and y belong to the input and output set respectively.

The function F is typically built as a composition of *sigmoid functions*



Artificial Intelligence (aka ***Machine Learning***) (continuation 2)

Neural networks have turned out to be extremely efficient in many applications. In the context of assimilation of observations, they have been used for defining for instance the observation operators (***H***) corresponding to satellite observations. But they have been used more recently, in evaluation studies and on idealized situations, but with some success, for determining 'dynamical laws'.

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

- Oceanography
- Palaoclimatology
- 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)
- Mathematical studies, independently of direct real life applications
- …

It has now become a major tool of numerical environmental science, and a subject of mathematical study in its own right.

A few of the (many) remaining problems :

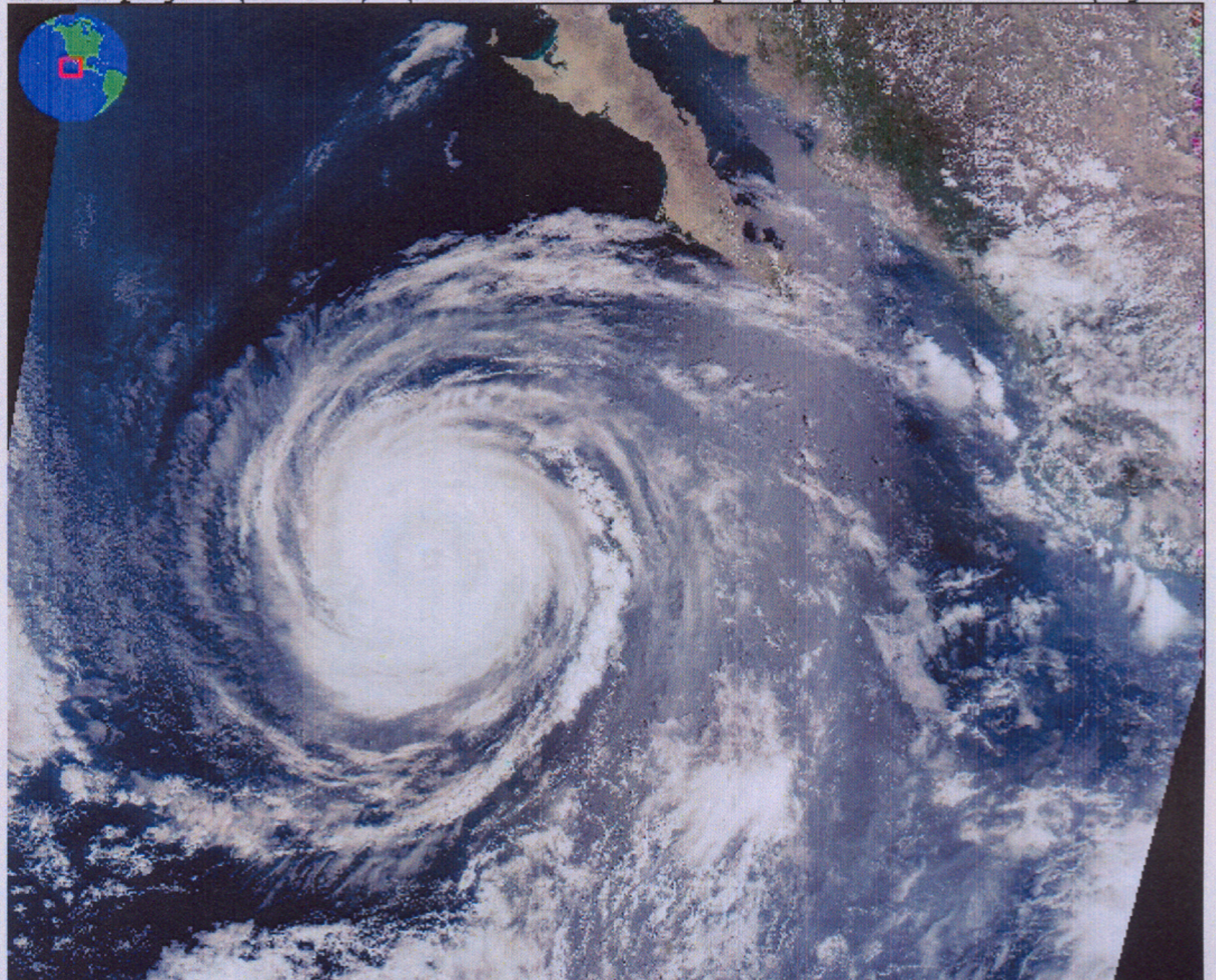
- Observability (what to observe in order to know what we want to know ? Data are noisy, system is chaotic !)

- 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



La Fin du Cours …