An introduction to data assimilation

Episode 2

Eric Blayo
University of Grenoble and INRIA

Previously...

Model problem: least squares approach

Two different available measurements of a single quantity. Which estimation for its true value ? $\longrightarrow$ least squares approach

Example 2 obs $y_1 = 19°C$ and $y_2 = 21°C$ of the (unknown) present temperature x.

- Let $J(x) = \frac{1}{2}\left[(x-y_1)^2 + (x-y_2)^2\right]$
- $\text{Min}_x\ J(x) \quad \longrightarrow \hat{x} = \dfrac{y_1+y_2}{2} = 20°C$

Model problem: least squares approach

Observation operator If $\neq$ units: $y_1 = 66.2°$F and $y_2 = 69.8°$F

- Let $H(x) = \frac{9}{5}x + 32$
- Let $J(x) = \frac{1}{2}\left[(H(x) - y_1)^2 + (H(x) - y_2)^2\right]$
- $\text{Min}_x\ J(x) \quad \longrightarrow \hat{x} = 20^\circ\text{C}$

Drawback # 1: *if observation units are inhomogeneous*
$y_1 = 66.2°$F and $y_2 = 21°$C

- $J(x) = \frac{1}{2}\left[(H(x) - y_1)^2 + (x - y_2)^2\right] \quad \longrightarrow \hat{x} = 19.47^\circ\text{C}$!!

Drawback # 2: *if observation accuracies are inhomogeneous*
If y_1 is twice more accurate than y_2, one should obtain $\hat{x} = \frac{2y_1 + y_2}{3} = 19.67^\circ\text{C}$

$\longrightarrow J$ should be $J(x) = \frac{1}{2}\left[\left(\frac{x - y_1}{1/2}\right)^2 + \left(\frac{x - y_2}{1}\right)^2\right]$

Model problem: statistical approach

Reformulation in a **probabilistic framework**:

- the goal is to estimate a scalar value x
- y_i is a realization of a random variable Y_i
- One is looking for an estimator (i.e. a r.v.) $\hat{X}$ that is
  - **linear**: $\hat{X} = \alpha_1 Y_1 + \alpha_2 Y_2$ (in order to be simple)
  - **unbiased**: $E(\hat{X}) = x$ (it seems reasonable)
  - **of minimal variance**: $\text{Var}(\hat{X})$ minimum (optimal accuracy)

$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)

Model problem: statistical approach

Let $Y_i = x + \varepsilon_i$ with

Hypotheses

- $E(\varepsilon_i) = 0 \qquad (i = 1, 2)$ — unbiased measurement devices
- $\mathrm{Var}(\varepsilon_i) = \sigma_i^2 \qquad (i = 1, 2)$ — known accuracies
- $\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = 0$ — independent measurement errors

BLUE

$$\hat{X} = \frac{\dfrac{1}{\sigma_1^2} Y_1 + \dfrac{1}{\sigma_2^2} Y_2}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}}$$

Its accuracy: $\left[\mathrm{Var}(\hat{X})\right]^{-1} = \dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}$ — accuracies are added

Model problem: statistical approach

Variational equivalence

This is equivalent to the problem:

$$\text{Minimize } J(x) = \frac{1}{2}\left[\frac{(x-y_1)^2}{\sigma_1^2} + \frac{(x-y_2)^2}{\sigma_2^2}\right]$$

Remarks:

- This answers the previous problems of sensitivity to inhomogeneous units and insensitivity to inhomogeneous accuracies
- This gives a rationale for choosing the norm for defining J
- $$\underbrace{J''(\hat{x})}_{\text{convexity}} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \underbrace{[\mathrm{Var}(\hat{x})]^{-1}}_{\text{accuracy}}$$

Model problem: formulation background + observation

If one considers that y_1 is a prior (or *background*) estimate x_b for x, and $y_2 = y$ is an independent observation, then:

$$J(x) = \underbrace{\frac{1}{2}\frac{(x-x_b)^2}{\sigma_b^2}}_{J_b} + \underbrace{\frac{1}{2}\frac{(x-y)^2}{\sigma_o^2}}_{J_o}$$

and

$$\hat{x} = \frac{\dfrac{1}{\sigma_b^2}x_b + \dfrac{1}{\sigma_o^2}y}{\dfrac{1}{\sigma_b^2} + \dfrac{1}{\sigma_o^2}} = x_b + \underbrace{\frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}}_{\text{gain}}\;\underbrace{(y-x_b)}_{\text{innovation}}$$

Model problem: Bayesian approach

One can also consider x as a realization of a r.v. X, and be interested in the pdf $p(X|Y)$.

Several optimality criteria

- **minimum variance**: $\hat{X}_{MV}$ such that the spread around it is minimal
  $$\longrightarrow \hat{X}_{MV} = E(X|Y)$$
- **maximum a posteriori**: most probable value of X given Y
  $$\longrightarrow \hat{X}_{MAP} \text{ such that } \frac{\partial p(X|Y)}{\partial X} = 0$$
- **maximum likelihood**: $\hat{X}_{ML}$ that maximizes $p(Y|X)$

- Based on the Bayes rule:
  $$P(X = x \,|\, Y = y) = \frac{P(Y = y \,|\, X = x)\, P(X = x)}{P(Y = y)}$$
- requires additional hypotheses on prior pdf for X and for $Y|X$
- In the Gaussian case, these estimations coincide with the BLUE

Generalization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$

Observations: $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$

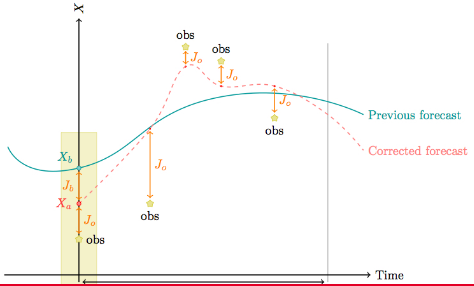
Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$

Generalization: variational approach

Stationary case: $$J(\mathbf{x}) = \underbrace{\frac{1}{2}\|\mathbf{x}-\mathbf{x}_b\|_b^2}_{\text{background term } J_b} + \underbrace{\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|_o^2}_{\text{observation term } J_o}$$

Time dependent case:

$$\begin{aligned} J(\mathbf{x}_0) &= \frac{1}{2}\|\mathbf{x}_0-\mathbf{x}_0^b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(\mathbf{x}(t_i))-\mathbf{y}(t_i)\|_o^2 \\ &= \frac{1}{2}\|\mathbf{x}_0-\mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0))-\mathbf{y}(t_i)\|_o^2 \end{aligned}$$



Generalization: statistical approach

Let $\mathbf{X}_b = \mathbf{x} + \varepsilon_b$ and $\mathbf{Y} = \mathbf{H}\mathbf{x} + \varepsilon_o$

Hypotheses:

- $E(\varepsilon_b) = 0$ — unbiased background
- $E(\varepsilon_o) = 0$ — unbiased measurement devices
- $\text{Cov}(\varepsilon_b, \varepsilon_o) = 0$ — independent background and observation errors
- $\text{Cov}(\varepsilon_b) = \mathbf{B}$ et $\text{Cov}(\varepsilon_o) = \mathbf{R}$ — known accuracies and covariances

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + \underbrace{(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}}_{\text{gain matrix}} \quad \underbrace{(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)}_{\text{innovation vector}}$$

with $\left[\text{Cov}(\hat{\mathbf{X}})\right]^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ — accuracies are added

Links between both approaches

Statistical approach: BLUE

$$\hat{\mathbf{X}} = \mathbf{X}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}_b)$$

with $\mathrm{Cov}(\hat{\mathbf{X}}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$

Variational approach in the linear stationary case

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\|\mathbf{x} - \mathbf{x}_b\|_b^2 + \frac{1}{2}\|H(\mathbf{x}) - \mathbf{y}\|_o^2 \\ &= \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}) \end{aligned}$$

$$\min_{\mathbf{x}\in\mathbf{R}^n} J(\mathbf{x}) \longrightarrow \hat{\mathbf{x}} = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

Same remarks as previously

- The statistical approach rationalizes the choice of the norms for J_o and J_b in the variational approach.
- $$\underbrace{\left[\mathrm{Cov}(\hat{\mathbf{X}})\right]^{-1}}_{\text{accuracy}} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \underbrace{\mathrm{Hess}(J)}_{\text{convexity}}$$

If the problem is time dependent

Dynamical system: $\mathbf{x}^t(t_{k+1}) = \mathbf{M}(t_k, t_{k+1})\,\mathbf{x}^t(t_k) + \mathbf{e}(t_k)$

- $\mathbf{x}^t(t_k)$ true state at time t_k
- $\mathbf{M}(t_k, t_{k+1})$ model assumed linear between t_k and t_{k+1}
- $\mathbf{e}(t_k)$ model error at time t_k

Observations $\mathbf{y}_k$ distributed in time.

Hypotheses

- $\mathbf{e}(t_k)$ is unbiased, with covariance matrix $\mathbf{Q}_k$
- $\mathbf{e}(t_k)$ and $\mathbf{e}(t_l)$ are independent $(k \neq l)$
- Unbiased observation $\mathbf{y}_k$, with error covariance matrix $\mathbf{R}_k$
- $\mathbf{e}(t_k)$ and analysis error $\mathbf{x}^a(t_k) - \mathbf{x}^t(t_k)$ are independent

If the problem is time dependent

Kalman filter (Kalman and Bucy, 1961)

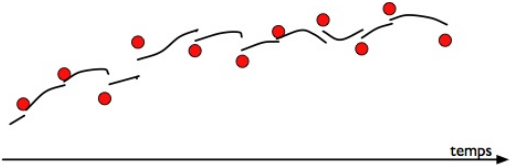
Initialization:

$$\mathbf{x}^a(t_0) = \mathbf{x}_0 \quad \textit{approximate initial state}$$
$$\mathbf{P}^a(t_0) = \mathbf{P}_0 \quad \textit{error covariance matrix}$$

Step k: *(prediction - correction, or forecast - analysis)*

$$\begin{aligned} \mathbf{x}^f_{k+1} &= \mathbf{M}_{k,k+1}\,\mathbf{x}^a_k \qquad \text{Forecast} \\ \mathbf{P}^f_{k+1} &= \mathbf{M}_{k,k+1}\mathbf{P}^a_k\mathbf{M}^T_{k,k+1} + \mathbf{Q}_k \end{aligned}$$

$$\begin{aligned} \mathbf{x}^a_{k+1} &= \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}\left[\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}^f_{k+1}\right] \qquad \text{BLUE analysis} \\ \mathbf{K}_{k+1} &= \mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1}\left[\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}\right]^{-1} \\ \mathbf{P}^a_{k+1} &= \mathbf{P}^f_{k+1} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f_{k+1} \end{aligned}$$



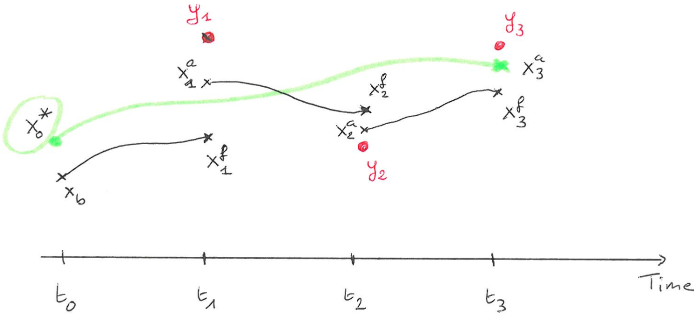
If the problem is time dependent

Equivalence with the variational approach

If $\mathbf{H}_k$ and $\mathbf{M}(t_k, t_{k+1})$ are linear, and if the model is perfect ($\mathbf{e}_k = 0$), then the Kalman filter and the variational method minimizing

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{x}_0) + \frac{1}{2}\sum_{k=0}^{N}(\mathbf{H_k M}(t_0,t_k)\mathbf{x}-\mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathbf{H_k M}(t_0,t_k)\mathbf{x}-\mathbf{y}_k)$$

lead to the same solution at $t = t_N$.



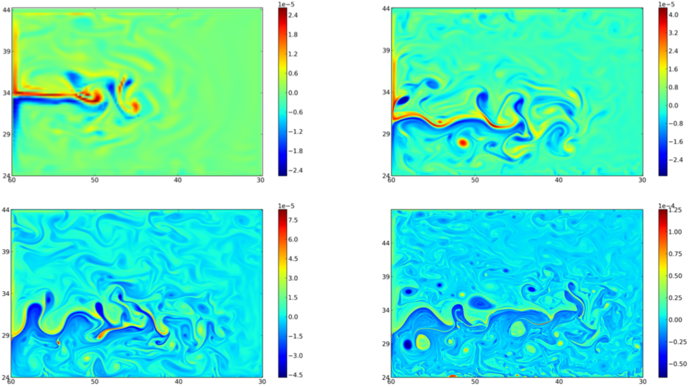
Common main methodological difficulties

- Non linearities: J non quadratic / what about Kalman filter ?
- Huge dimensions $[\mathbf{x}] = \mathcal{O}(10^6 - 10^9)$: minimization of J / management of huge matrices
- Poorly known error statistics: choice of the norms / $\mathbf{B}, \mathbf{R}, \mathbf{Q}$
- Scientific computing issues (data management, code efficiency, parallelization...)

$\longrightarrow$ TODAY's LECTURE

Towards larger dimensions and stronger nonlinearities

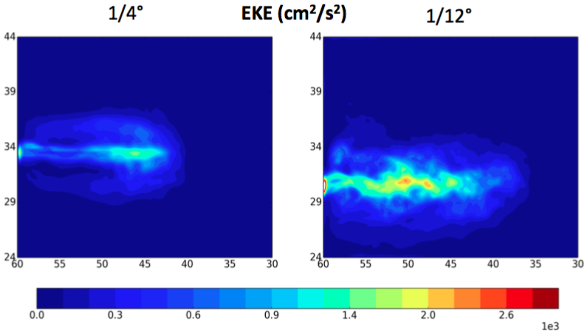
Increasing the model resolution increases the size of the state variable and, for a number of applications, allows for stronger scale interactions.



Snapshots of the surface relative vorticity in the SEABASS configuration of NEMO, for different model resolutions: $1/4^\circ$, $1/12^\circ$, $1/24^\circ$ and $1/100^\circ$.

Towards larger dimensions and stronger nonlinearities

This results in increased turbulent energy levels and nonlinear effects.

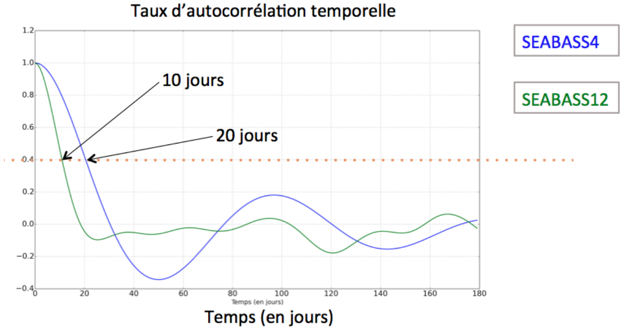


Max EKE

- 1/4° : 1650 cm^2/s^2
- 1/12° : 3000 cm^2/s^2
- Jason-1 Gulf Stream: 3000 cm^2/s^2

Towards larger dimensions and stronger nonlinearities

This results in increased turbulent energy levels and nonlinear effects.



Statistical approach

Non linearities: extended Kalman filter

The Kalman filter assumes that M and H are linear. If not: linearization

Reminder: derivatives and gradients

$f : E \longrightarrow \mathbf{R}$ (E being of finite or infinite dimension)

- Gradient (or Fréchet derivative): E being an Hilbert space, f is Fréchet differentiable at point $x \in E$ iff

$$\exists p \in E \text{ such that } f(x+h) = f(x) + (p,h) + o(\|h\|) \quad \forall h \in E$$

  p is the derivative or gradient of f at point x, denoted $f'(x)$ or $\nabla f(x)$.

- $h \rightarrow (p(x), h)$ is a linear function, called differential function or tangent linear function or Jacobian of f at point x

Non linearities: extended Kalman filter

The Kalman filter assumes that M and H are linear. If not: linearization

$$\mathbf{x}^f_{k+1} = M_{k,k+1}(\mathbf{x}^a_k) \simeq M_{k,k+1}(\mathbf{x}^t_k) + \mathbf{M}_{k,k+1} \underbrace{(\mathbf{x}^a_k - \mathbf{x}^t_k)}_{\mathbf{e}^a_k}$$

$$\Longrightarrow \quad \mathbf{x}^f_{k+1} - \mathbf{x}^t_{k+1} = \mathbf{e}^f_{k+1} = \underbrace{M_{k,k+1}(\mathbf{x}^t_k) - \mathbf{x}^t_{k+1}}_{\mathbf{e}_k} + \mathbf{M}_{k,k+1}\mathbf{e}^a_k$$

$$\Longrightarrow \quad \mathbf{P}^f_{k+1} = \mathrm{Cov}(\mathbf{e}^f_{k+1}) = \mathbf{M}_{k,k+1}\mathbf{P}^a_k\mathbf{M}^T_{k,k+1} + \mathbf{Q}_k$$

and similarly for the other equations of the filter

Non linearities: extended Kalman filter

Extended Kalman filter

Initialization:
$$\begin{array}{rcll} \mathbf{x}^a(t_0) & = & \mathbf{x}_0 & \textit{approximate initial state} \\ \mathbf{P}^a(t_0) & = & \mathbf{P}_0 & \textit{error covariance matrix} \end{array}$$

Step k: *(prediction - correction, or forecast - analysis)*

$$\begin{array}{rcll} \mathbf{x}^f_{k+1} & = & M_{k,k+1}(\mathbf{x}^a_k) & \text{Forecast} \\ \mathbf{P}^f_{k+1} & = & \mathbf{M}_{k,k+1}\mathbf{P}^a_k\mathbf{M}^T_{k,k+1} + \mathbf{Q}_k & \end{array}$$

$$\begin{array}{rcll} \mathbf{x}^a_{k+1} & = & \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}\left[\mathbf{y}_{k+1} - H_{k+1}(\mathbf{x}^f_{k+1})\right] & \text{BLUE analysis} \\ \mathbf{K}_{k+1} & = & \mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1}\left[\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}\right]^{-1} & \\ \mathbf{P}^a_{k+1} & = & \mathbf{P}^f_{k+1} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f_{k+1} & \end{array}$$

Non linearities: extended Kalman filter

Extended Kalman filter

Initialization:

$$\begin{array}{rcll} \mathbf{x}^a(t_0) & = & \mathbf{x}_0 & \textit{approximate initial state} \\ \mathbf{P}^a(t_0) & = & \mathbf{P}_0 & \textit{error covariance matrix} \end{array}$$

Step k: *(prediction - correction, or forecast - analysis)*

$$\begin{array}{rcll} \mathbf{x}^f_{k+1} & = & M_{k,k+1}(\mathbf{x}^a_k) & \text{Forecast} \\ \mathbf{P}^f_{k+1} & = & \mathbf{M}_{k,k+1}\mathbf{P}^a_k\mathbf{M}^T_{k,k+1} + \mathbf{Q}_k & \end{array}$$

$$\begin{array}{rcll} \mathbf{x}^a_{k+1} & = & \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}\left[\mathbf{y}_{k+1} - H_{k+1}(\mathbf{x}^f_{k+1})\right] & \text{BLUE analysis} \\ \mathbf{K}_{k+1} & = & \mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1}\left[\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}\right]^{-1} & \\ \mathbf{P}^a_{k+1} & = & \mathbf{P}^f_{k+1} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f_{k+1} & \end{array}$$

- OK if nonlinearities are not too strong
- Requires the availability of $\mathbf{M}_{k,k+1}$ and $\mathbf{H}_k$
- More sophisticated approaches have been developed $\longrightarrow$ unscented Kalman filter (exact up to second order, requires no tangent linear model nor Hessian matrix)

Huge dimension: reduced order filters

As soon as $[\mathbf{x}]$ becomes huge, it's no longer possible to handle the covariance matrices.

Idea: a large part of the system variability can be represented (or is assumed to) in a reduced dimension space.

$\longrightarrow$ RRSQRT filter, SEEK filter, SEIK filter...

Huge dimension: reduced order filters

Example: Reduced Rank SQuare Root filter

- $\mathbf{P}_0^f \simeq \mathbf{S}_0^f \left(\mathbf{S}_0^f\right)^T$ with $\text{size}(\mathbf{S}_0^f) = (n, r)$ (r leading modes, $r \ll n$)
- This is injected in the filter equations. This leads for instance to $\mathbf{P}_k^a = \mathbf{S}_k^a \left(\mathbf{S}_k^a\right)^T$, with

$$\mathbf{S}_k^a = \underbrace{\mathbf{S}_k^f}_{(n,r)} \left(\underbrace{\mathbf{I}_r - \mathbf{\Psi}_k^T [\mathbf{\Psi}_k \mathbf{\Psi}_k^T + \mathbf{R}_k]^{-1} \mathbf{\Psi}_k}_{(r,r)} \right)^{1/2} \qquad \text{where } \mathbf{\Psi}_k = \underbrace{\mathbf{H}_k \mathbf{S}_k^f}_{(p,r)}$$

Pros: most computations in low dimension

Cons: choice and time evolution of the modes

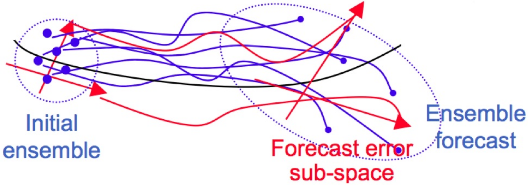
A widely used filter: the Ensemble Kalman filter

- addresses both problems of non linearities and huge dimension
- rather simple and intuitive

Idea: generation of an ensemble of N trajectories, by N perturbations of the set of observations (consistently with $\mathbf{R}$). Standard extended Kalman filter, with covariance matrices computed using the ensemble:

$$\mathbf{P}_k^f = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_{j,k}^f - \bar{x}_k^f)(\mathbf{x}_{j,k}^f - \bar{x}_k^f)^T \text{ with } \bar{x}_k^f = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_{j,k}^f$$

$$\mathbf{P}_k^a = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_{j,k}^a - \bar{x}_k^a)(\mathbf{x}_{j,k}^f - \bar{x}_k^f)^T \text{ with } \bar{x}_k^a = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_{j,k}^a$$



Variational approach

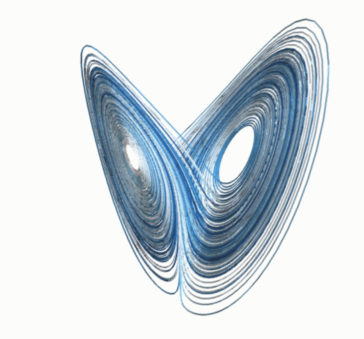
Cost function and non linearities

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.

  Example: the Lorenz system (1963)

$$\begin{cases} \dfrac{dx}{dt} = \alpha(y-x) \\ \dfrac{dy}{dt} = \beta x - y - xz \\ \dfrac{dz}{dt} = -\gamma z + xy \end{cases}$$

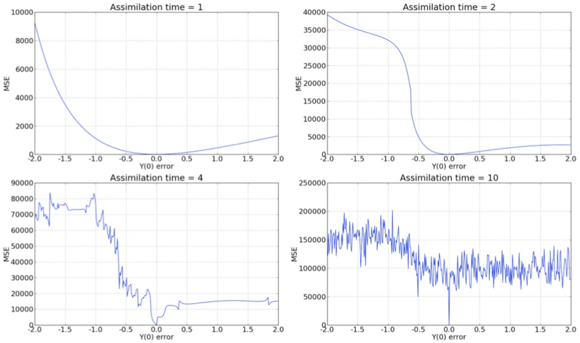


$$J_o(y_0) = \frac{1}{2}\sum_{i=0}^{N}\left(x(t_i) - x_{\mathrm{obs}}(t_i)\right)^2\,dt$$

Cost function and non linearities

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N}\|H_i(M_{0\rightarrow t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

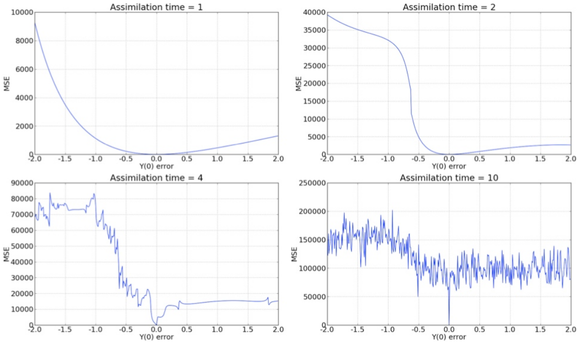
- If H and/or M are nonlinear then J_o is no longer quadratic.



Cost function and non linearities

$$J(\mathbf{x}_0) = J_b(\mathbf{x}_0) + J_o(\mathbf{x}_0) = \frac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}_b\|_b^2 + \frac{1}{2}\sum_{i=0}^{N} \|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$$

- If H and/or M are nonlinear then J_o is no longer quadratic.



- Adding J_b makes it "more quadratic" (J_b is a regularization term), but $J = J_o + J_b$ may however have several (local) minima.

4D-Var

4D-Var algorithm corresponds to the minimization of

$$J(\mathbf{x}_0) = \frac{1}{2}\,(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=0}^{N}(H_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i)$$

Preconditioned cost function

Defining $\mathbf{v} = \mathbf{B}^{-1/2}\,(\mathbf{x} - \mathbf{x}^b)$, J becomes

$$J(\mathbf{v}_0) = \frac{1}{2}\,\mathbf{v}_0^T\mathbf{v}_0 + \frac{1}{2}\sum_{i=0}^{N}(H_i(\mathbf{B}^{1/2}\mathbf{v}_i + \mathbf{x}_i^b) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{B}^{1/2}\mathbf{v}_i + \mathbf{x}_i^b) - \mathbf{y}_i)$$

4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var

The problem is written in terms of $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^b$, and

$$J(\mathbf{x}_0) = \frac{1}{2}\,(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\,\sum_{i=0}^{N} (H_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (H_i(\mathbf{x}_i) - \mathbf{y}_i)$$

is approximated by a series of quadratic cost functions:

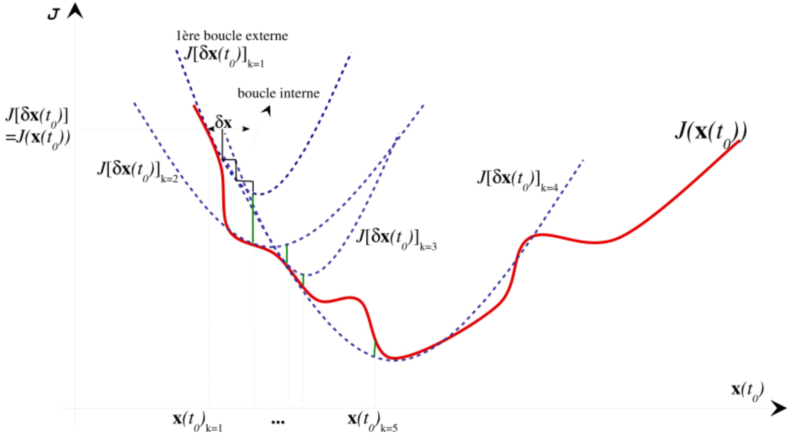
$$J^{(k+1)}(\delta\mathbf{x}_0) = \frac{1}{2}\,\delta\mathbf{x}_0^T \mathbf{B}^{-1} \delta\mathbf{x}_0 + \frac{1}{2}\,\sum_{i=0}^{N} (\mathbf{H}_i^{(k)} \delta\mathbf{x}_i - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i^{(k)} \delta\mathbf{x}_i - \mathbf{d}_i)$$

with $\delta\mathbf{x}_{i+1} = \mathbf{M}_{i,i+1}^{(k)} \delta\mathbf{x}_i$ and $\mathbf{d}_i = \mathbf{y}_i - H_i(\mathbf{x}_i^{(k)})$

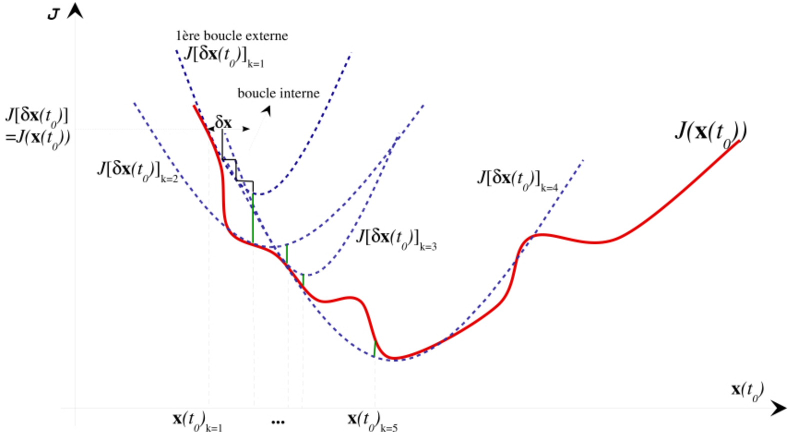
- Kind of Gauss-Newton algorithm
- Tangent linear hypotheses must be satisfied:

$$M(\mathbf{x}_0^{(k)} + \delta\mathbf{x}_0) \simeq M(\mathbf{x}_0^{(k)}) + \mathbf{M}^{(k)} \delta\mathbf{x}_0$$
$$H_i(\mathbf{x}_i^{(k)} + \delta\mathbf{x}_i) \simeq H_i(\mathbf{x}_i^{(k)}) + \mathbf{H}_i^{(k)} \delta\mathbf{x}_i$$

4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var



4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var



Multi-incremental 4D-Var: inner loops can be made using some simplified physics and/or coarser resolution (Courtier et al. 1994, Courtier 1995, Veersé ıd Thépaut 1998, Trémolet 2005).

4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var

The 3D-FGAT (First Guess at Appropriate Time) is an approximation of incremental 4D-Var where the tangent linear model is replaced by identity:

$$J^{(k+1)}(\delta \mathbf{x}_0) = \frac{1}{2}\, \delta \mathbf{x}_0^T \mathbf{B}^{-1} \delta \mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^{N} (\mathbf{H}_i^{(k)} \delta \mathbf{x}_0 - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i^{(k)} \delta \mathbf{x}_0 - \mathbf{d}_i)$$

$\longrightarrow$ something between 3D and 4D

Pros:

- much cheaper, does not require the adjoint model (see later)
- algorithm is close to incremental 4D-Var
- innovation is computed at the correct observation time

Cons: approximation !

4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var

3D-Var: all observations are gathered as if they were all at time t_0.

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2} \sum_{i=0}^{N} (H_i(\mathbf{x}_0) - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (H_i(\mathbf{x}_0) - \mathbf{y}_i)$$

Pros: still cheaper
Cons: approximation !!

4D-Var / Incremental 4D-Var / 3D-FGAT / 3D-Var

3D-Var: all observations are gathered as if they were all at time t_0.

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=0}^{N}(H_i(\mathbf{x}_0) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_0) - \mathbf{y}_i)$$

Pros: still cheaper
Cons: approximation !!

Remark: 3D-Var = Optimal Interpolation = Krigging

Summary: simplifying $J \to$ a series of methods

4D-Var:

$$J(\mathbf{x}_0) = \frac{1}{2}\,(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\,\sum_{i=0}^{N}(H_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i)$$

Incremental 4D-Var: $M(\mathbf{x}_0 + \delta\mathbf{x}_0) \simeq M(\mathbf{x}_0) + \mathbf{M}\delta\mathbf{x}_0$

$$J^{(k+1)}(\delta\mathbf{x}_0) = \frac{1}{2}\,\delta\mathbf{x}_0^T \mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\,\sum_{i=0}^{N}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_i - \mathbf{d}_i)^T \mathbf{R}_i^{-1}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_i - \mathbf{d}_i)$$

Multi-incremental 4D-Var: $M(\mathbf{x}_0 + \delta\mathbf{x}_0) \simeq M(\mathbf{x}_0) + \mathbf{S}^{-l}\mathbf{M}^L\delta\mathbf{x}_0^L$

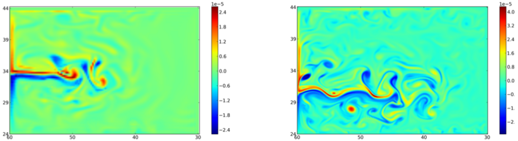
$$J^{(k+1)}(\delta\mathbf{x}_0^L) = \frac{1}{2}\,(\delta\mathbf{x}_0^L)^T \mathbf{B}^{-1}\delta\mathbf{x}_0^L + \frac{1}{2}\,\sum_{i=0}^{N}(\mathbf{H}_i^{(k),L}\delta\mathbf{x}_i^L - \mathbf{d}_i)^T \mathbf{R}_i^{-1}(\mathbf{H}_i^{(k),L}\delta\mathbf{x}_i^L - \mathbf{d}_i)$$

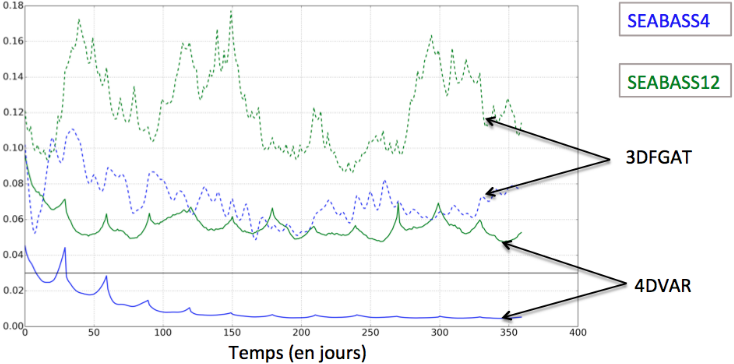
3D-FGAT: $M(\mathbf{x}_0 + \delta\mathbf{x}_0) \simeq M(\mathbf{x}_0) + \delta\mathbf{x}_0$

$$J^{(k+1)}(\delta\mathbf{x}_0) = \frac{1}{2}\,\delta\mathbf{x}_0^T \mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\,\sum_{i=0}^{N}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_0 - \mathbf{d}_i)^T \mathbf{R}_i^{-1}(\mathbf{H}_i^{(k)}\delta\mathbf{x}_0 - \mathbf{d}_i)$$

3D-Var: $M(\mathbf{x}_0 + \delta\mathbf{x}_0) \simeq \mathbf{x}_0 + \delta\mathbf{x}_0$

$$J(\mathbf{x}_0) = \frac{1}{2}\,(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\,\sum_{i=0}^{N}(H_i(\mathbf{x}_0) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_0) - \mathbf{y}_i)$$





3DFGAT inefficace au 1/12°

4DVAR >> 3DFGAT

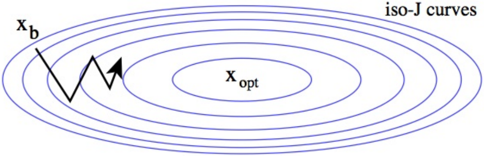
Given the size of n and p, it is generally impossible to handle explicitly H, $\mathbf{B}$ and $\mathbf{R}$. So, even in the simplest case (3D-Var + H linear, for which we have an explicit expression for $\hat{\mathbf{x}}$) the direct computation of the gain matrix is impossible.

► the computation of $\hat{\mathbf{x}}$ is performed using an optimization algorithm.

Descent methods

Descent methods for minimizing the cost function require the knowledge of (an estimate of) its gradient.

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\, \mathbf{d}_k$$



$$\text{with } \mathbf{d}_k = \begin{cases} -\nabla J(\mathbf{x}_k) & \text{gradient method} \\ -\left[\text{Hess}(J)(\mathbf{x}_k)\right]^{-1} \nabla J(\mathbf{x}_k) & \text{Newton method} \\ -\mathbf{B}_k\, \nabla J(\mathbf{x}_k) & \text{quasi-Newton methods (BFGS, \ldots)} \\ -\nabla J(\mathbf{x}_k) + \frac{\|\nabla J(\mathbf{x}_k)\|^2}{\|\nabla J(\mathbf{x}_{k-1})\|^2} d_{k-1} & \text{conjugate gradient} \\ \ldots & \ldots \end{cases}$$

Reminder: derivatives and gradients

$f : E \longrightarrow \mathbf{R}$ (E being of finite or infinite dimension)

- Directional (or Gâteaux) derivative of f at point $x \in E$ in direction $d \in E$:

$$\frac{\partial f}{\partial d}(x) = \hat{f}[x](d) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

Example: partial derivatives $\dfrac{\partial f}{\partial x_i}$ are directional derivatives in the direction of the members of the canonical basis ($d = e_i$)

Reminder: derivatives and gradients

$f : E \longrightarrow \mathbf{R}$ (E being of finite or infinite dimension)

- Gradient (or Fréchet derivative): E being an Hilbert space, f is Fréchet differentiable at point $x \in E$ iff

$$\exists p \in E \text{ such that } f(x+h) = f(x) + (p,h) + o(\|h\|) \quad \forall h \in E$$

  p is the derivative or gradient of f at point x, denoted $f'(x)$ or $\nabla f(x)$.

- $h \rightarrow (p(x), h)$ is a linear function, called differential function or tangent linear function or Jacobian of f at point x

Reminder: derivatives and gradients

$$f : E \longrightarrow \mathbf{R} \qquad (E \text{ being of finite or infinite dimension})$$

- Gradient (or Fréchet derivative): E being an Hilbert space, f is Fréchet differentiable at point $x \in E$ iff

$$\exists p \in E \text{ such that } f(x+h) = f(x) + (p,h) + o(\|h\|) \quad \forall h \in E$$

p is the derivative or gradient of f at point x, denoted $f'(x)$ or $\nabla f(x)$.

- $h \rightarrow (p(x), h)$ is a linear function, called differential function or tangent linear function or Jacobian of f at point x

- Important (obvious) relationship: $\dfrac{\partial f}{\partial d}(x) = (\nabla f(x), d)$

Getting the gradient is not obvious

The computation of $\nabla J(\mathbf{x}_k)$ may be difficult if the dependency of J with regard to the control variable $\mathbf{x}$ is not direct.

Example:

- $u(x)$ solution of an ODE
- K a coefficient of this ODE
- $u^{\text{obs}}(x)$ an observation of $u(x)$
- $J(K) = \dfrac{1}{2} \left\| u(x) - u^{\text{obs}}(x) \right\|^2$

Getting the gradient is not obvious

The computation of $\nabla J(\mathbf{x}_k)$ may be difficult if the dependency of J with regard to the control variable $\mathbf{x}$ is not direct.

Example:

- $u(x)$ solution of an ODE
- K a coefficient of this ODE
- $u^{\text{obs}}(x)$ an observation of $u(x)$
- $J(K) = \frac{1}{2} \left\| u(x) - u^{\text{obs}}(x) \right\|^2$

$$\hat{J}[K](k) = (\nabla J(K), k) = < \hat{u}, u - u^{\text{obs}} >$$

$$\text{with } \hat{u} = \frac{\partial u}{\partial k}(K) = \lim_{\alpha \to 0} \frac{u_{K+\alpha k} - u_K}{\alpha}$$

Getting the gradient is not obvious

It is often difficult (or even impossible) to obtain the gradient through the computation of growth rates.

Example:

$$\left\{ \begin{array}{l} \dfrac{d\mathbf{x}(t))}{dt} = M(\mathbf{x}(t)) \quad t \in [0, T] \\ \mathbf{x}(t=0) = \mathbf{u} \end{array} \right. \qquad \text{with } \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}$$

$$J(\mathbf{u}) = \frac{1}{2} \int_0^T \|\mathbf{x}(t) - \mathbf{x}^{\text{obs}}(t)\|^2 \qquad \longrightarrow \quad \text{requires one model run}$$

$$\nabla J(\mathbf{u}) = \begin{pmatrix} \dfrac{\partial J}{\partial u_1}(\mathbf{u}) \\ \vdots \\ \dfrac{\partial J}{\partial u_N}(\mathbf{u}) \end{pmatrix} \simeq \begin{pmatrix} [J(\mathbf{u} + \alpha\, \mathbf{e}_1) - J(\mathbf{u})] / \alpha \\ \vdots \\ [J(\mathbf{u} + \alpha\, \mathbf{e}_N) - J(\mathbf{u})] / \alpha \end{pmatrix}$$

$\longrightarrow$ $N + 1$ model runs

Getting the gradient is not obvious

In actual applications like meteorology / oceanography,
$N = [\mathbf{u}] = \mathcal{O}(10^6 - 10^9) \quad \longrightarrow$ this method cannot be used.

Alternatively, the adjoint method provides a very efficient way to compute ∇J.

Getting the gradient is not obvious

In actual applications like meteorology / oceanography,
$N = [\mathbf{u}] = \mathcal{O}(10^6 - 10^9) \quad \longrightarrow$ this method cannot be used.

Alternatively, the adjoint method provides a very efficient way to compute ∇J.

On the contrary, do not forget that, if the size of the control variable is very small (< 10), ∇J can be easily estimated by the computation of growth rates.

Reminder: adjoint operator

- General definition:
  Let $\mathcal{X}$ and $\mathcal{Y}$ two prehilbertian spaces (i.e. vector spaces with scalar products).
  Let $A : \mathcal{X} \longrightarrow \mathcal{Y}$ an operator.
  The adjoint operator $A^* : \mathcal{Y} \longrightarrow \mathcal{X}$ is defined by:

$$\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \qquad < Ax, y >_{\mathcal{Y}} = < x, A^* y >_{\mathcal{X}}$$

  In the case where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces and A is linear, then A^* always exists (and is unique).

Reminder: adjoint operator

- General definition:
  Let $\mathcal{X}$ and $\mathcal{Y}$ two prehilbertian spaces (i.e. vector spaces with scalar products).
  Let $A : \mathcal{X} \longrightarrow \mathcal{Y}$ an operator.
  The adjoint operator $A^* : \mathcal{Y} \longrightarrow \mathcal{X}$ is defined by:

$$\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \qquad < Ax, y >_{\mathcal{Y}} = < x, A^* y >_{\mathcal{X}}$$

  In the case where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces and A is linear, then A^* always exists (and is unique).

- Adjoint operator in finite dimension:
  $A : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ a linear operator (i.e. a matrix). Then its adjoint operator A^* (w.r. to Euclidian norms) is A^T.

The adjoint method: a simple example (continuous case)

The assimilation problem

- $\begin{cases} -u''(x) + c(x)\,u'(x) = f(x) & x \in]0,1[\\ u(0) = u(1) = 0 \end{cases}$ $\qquad f \in L^2(]0,1[)$
- $c(x)$ is unknown
- $u^{\text{obs}}(x)$ an observation of $u(x)$
- Cost function: $J(c) = \dfrac{1}{2}\displaystyle\int_0^1 \left(u(x) - u^{\text{obs}}(x)\right)^2\, dx$

The adjoint method: a simple example (continuous case)

The assimilation problem

- $\begin{cases} -u''(x) + c(x)\,u'(x) = f(x) & x \in]0,1[\\ u(0) = u(1) = 0 \end{cases}$ $\qquad f \in L^2(]0,1[)$
- $c(x)$ is unknown
- $u^{\text{obs}}(x)$ an observation of $u(x)$
- Cost function: $J(c) = \dfrac{1}{2} \displaystyle\int_0^1 \left(u(x) - u^{\text{obs}}(x)\right)^2 \, dx$

$\nabla J \rightarrow$ Gâteaux-derivative: $\hat{J}[c](\delta c) = < \nabla J(c), \delta c >$

$$\hat{J}[c](\delta c) = \int_0^1 \hat{u}(x) \left(u(x) - u^{\text{obs}}(x)\right) dx \quad \text{with } \hat{u} = \lim_{\alpha \to 0} \frac{u_{c+\alpha\delta c} - u_c}{\alpha}$$

What is the equation satisfied by $\hat{u}$?

The adjoint method: a simple example (continuous case)

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 \end{cases}$$

tangent linear model

The adjoint method: a simple example (continuous case)

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 \end{cases}$$

tangent linear model

Going back to $\hat{\mathrm{J}}$: scalar product of the TLM with a variable p

$$-\int_0^1 \hat{u}''p + \int_0^1 c\,\hat{u}'p = -\int_0^1 \delta c\,u'p$$

The adjoint method: a simple example (continuous case)

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 \end{cases}$$

tangent linear model

Going back to $\hat{\mathrm{J}}$: scalar product of the TLM with a variable p

$$-\int_0^1 \hat{u}''p + \int_0^1 c\,\hat{u}'p = -\int_0^1 \delta c\,u'p$$

Integration by parts:

$$\int_0^1 \hat{u}\,(-p'' - (c\,p)') = \hat{u}'(1)p(1) - \hat{u}'(0)p(0) \; - \int_0^1 \delta c\,u'p$$

The adjoint method: a simple example (continuous case)

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 \end{cases}$$

tangent linear model

Going back to $\hat{\mathrm{J}}$: scalar product of the TLM with a variable p

$$-\int_0^1 \hat{u}''p + \int_0^1 c\,\hat{u}'p = -\int_0^1 \delta c\,u'p$$

Integration by parts:

$$\int_0^1 \hat{u}\,(-p'' - (c\,p)') = \hat{u}'(1)p(1) - \hat{u}'(0)p(0) \; - \int_0^1 \delta c\,u'p$$

$$\begin{cases} -p''(x) - (c(x)\,p(x))' = u(x) - u^{\mathrm{obs}}(x) & x \in]0,1[\\ p(0) = p(1) = 0 \end{cases}$$

adjoint model

The adjoint method: a simple example (continuous case)

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 & \end{cases}$$ tangent linear model

Going back to $\hat{J}$: scalar product of the TLM with a variable p

$$-\int_0^1 \hat{u}''p + \int_0^1 c\,\hat{u}'p = -\int_0^1 \delta c\,u'p$$

Integration by parts:

$$\int_0^1 \hat{u}\,(-p'' - (c\,p)') = \hat{u}'(1)p(1) - \hat{u}'(0)p(0) \; - \int_0^1 \delta c\,u'p$$

$$\begin{cases} -p''(x) - (c(x)\,p(x))' = u(x) - u^{\text{obs}}(x) & x \in]0,1[\\ p(0) = p(1) = 0 & \end{cases}$$ adjoint model

Then $\nabla J(c(x)) = -u'(x)\,p(x)$

The adjoint method: a simple example (continuous case)

Remark

Formally, we just made

$$(\,TLM(\hat{u}), p) = (\hat{u}, TLM^*(p))$$

We indeed computed the adjoint of the tangent linear model.

The adjoint method: a simple example (continuous case)

Remark

Formally, we just made

$$(TLM(\hat{u}), p) = (\hat{u}, TLM^*(p))$$

We indeed computed the adjoint of the tangent linear model.

Actual calculations

- Solve for the direct model

$$\begin{cases} -u''(x) + c(x)\, u'(x) = f(x) & x \in]0,1[\\ u(0) = u(1) = 0 \end{cases}$$

- Then solve for the adjoint model

$$\begin{cases} -p''(x) - (c(x)\, p(x))' = u(x) - u^{\text{obs}}(x) & x \in]0,1[\\ p(0) = p(1) = 0 \end{cases}$$

- Hence the gradient: $\nabla J(c(x)) = -u'(x)\, p(x)$

The adjoint method: a simple example (discrete case)

Model

$$\begin{cases} -u''(x) + c(x)\, u'(x) = f(x) & x \in]0,1[\\ u(0) = u(1) = 0 \end{cases}$$

$$\longrightarrow \begin{cases} -\dfrac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + c_i \, \dfrac{u_{i+1} - u_{i-1}}{2h} = f_i & i = 1 \dots N \\ u_0 = u_{N+1} = 0 \end{cases}$$

Cost function

$$J(c) = \frac{1}{2} \int_0^1 \left(u(x) - u^{\text{obs}}(x) \right)^2 dx \qquad \longrightarrow \frac{1}{2} \sum_{i=1}^{N} \left(u_i - u_i^{\text{obs}} \right)^2$$

Gâteaux derivative:

$$\hat{J}[c](\delta c) = \int_0^1 \hat{u}(x) \left(u(x) - u^{\text{obs}}(x) \right) dx \qquad \longrightarrow \sum_{i=1}^{N} \hat{u}_i \left(u_i - u_i^{\text{obs}} \right)$$

The adjoint method: a simple example (discrete case)

Tangent linear model

$$\begin{cases} -\hat{u}''(x) + c(x)\,\hat{u}'(x) = \; -\delta c(x)\,u'(x) & x \in]0,1[\\ \hat{u}(0) = \hat{u}(1) = 0 \end{cases}$$

$$\begin{cases} -\dfrac{\hat{u}_{i+1} - 2\hat{u}_i + \hat{u}_{i-1}}{h^2} + c_i\,\dfrac{\hat{u}_{i+1} - \hat{u}_{i-1}}{2h} = -\delta c_i\,\dfrac{u_{i+1} - u_{i-1}}{2h} & i = 1 \ldots N \\ \hat{u}_0 = \hat{u}_{N+1} = 0 \end{cases}$$

Adjoint model

$$\begin{cases} -p''(x) - (c(x)\,p(x))' = u(x) - u^{\text{obs}}(x) & x \in]0,1[\\ p(0) = p(1) = 0 \end{cases}$$

$$\begin{cases} -\dfrac{p_{i+1} - 2p_i + p_{i-1}}{h^2} - \dfrac{c_{i+1}\,p_{i+1} - c_{i-1}p_{i-1}}{2h} = u_i - u_i^{\text{obs}} & i = 1 \ldots N \\ p_0 = p_{N+1} = 0 \end{cases}$$

Gradient

$$\nabla J(c(x)) = -u'(x)\,p(x) \longrightarrow \begin{pmatrix} \vdots \\ -p_i\,\dfrac{u_{i+1} - u_{i-1}}{2h} \\ \vdots \end{pmatrix}$$

The adjoint method: a simple example (discrete case)

Remark: with matrix notations

What we do when determining the adjoint model is simply transposing the matrix which defines the tangent linear model

$$(\mathbf{M}\hat{\mathbf{U}}, \mathbf{P}) = (\hat{\mathbf{U}}, \mathbf{M}^T \mathbf{P})$$

In the preceding example:

$$\mathbf{M}\hat{\mathbf{U}} = \mathbf{F} \quad \text{with } \mathbf{M} = \begin{bmatrix} 2\alpha & -\alpha+\beta & 0 & \cdots & 0 \\ -\alpha-\beta & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -\alpha+\beta \\ 0 & \cdots & 0 & -\alpha-\beta & 2\alpha \end{bmatrix}$$

$\alpha = 1/h^2, \beta = c_i/2h$

The adjoint method: a simple example (discrete case)

Remark: with matrix notations

What we do when determining the adjoint model is simply transposing the matrix which defines the tangent linear model

$$(\mathbf{M}\hat{\mathbf{U}}, \mathbf{P}) = (\hat{\mathbf{U}}, \mathbf{M}^T \mathbf{P})$$

In the preceding example:

$$\mathbf{M}\hat{\mathbf{U}} = \mathbf{F} \quad \text{with } \mathbf{M} = \begin{bmatrix} 2\alpha & -\alpha+\beta & 0 & \cdots & 0 \\ -\alpha-\beta & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -\alpha+\beta \\ 0 & \cdots & 0 & -\alpha-\beta & 2\alpha \end{bmatrix}$$

$\alpha = 1/h^2$, $\beta = c_i/2h$

But $\mathbf{M}$ is generally not explicitly built in actual complex models...

A more complex (but still linear) example: control of the coefficient of a 1-D diffusion equation

$$\left\{\begin{array}{l} \dfrac{\partial u}{\partial t} - \dfrac{\partial}{\partial x}\left(K(x)\dfrac{\partial u}{\partial x}\right) = f(x,t) \qquad x \in]0,L[,\, t \in]0,T[\\ u(0,t) = u(L,t) = 0 \quad t \in [0,T] \\ u(x,0) = u_0(x) \quad x \in [0,L] \end{array}\right.$$

- $K(x)$ is unknown
- $u^{\text{obs}}(x,t)$ an available observation of $u(x,t)$

$$\text{Minimize } J(K(x)) = \frac{1}{2}\int_0^T \int_0^L \left(u(x,t) - u^{\text{obs}}(x,t)\right)^2 \, dx\, dt$$

Gâteaux derivative

$$\hat{J}[K](k) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\, dx\, dt$$

Gâteaux derivative

$$\hat{J}[K](k) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{obs}(x,t)\right)\, dx\, dt$$

Tangent linear model

$$\begin{cases} \dfrac{\partial \hat{u}}{\partial t} - \dfrac{\partial}{\partial x}\left(K(x)\dfrac{\partial \hat{u}}{\partial x}\right) = \dfrac{\partial}{\partial x}\left(k(x)\dfrac{\partial u}{\partial x}\right) & x \in]0,L[,\ t \in]0,T[\\ \hat{u}(0,t) = \hat{u}(L,t) = 0 \qquad t \in [0,T] \\ \hat{u}(x,0) = 0 \qquad x \in [0,L] \end{cases}$$

Gâteaux derivative

$$\hat{J}[K](k) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\,dx\,dt$$

Tangent linear model

$$\begin{cases} \dfrac{\partial \hat{u}}{\partial t} - \dfrac{\partial}{\partial x}\left(K(x)\dfrac{\partial \hat{u}}{\partial x}\right) = \dfrac{\partial}{\partial x}\left(k(x)\dfrac{\partial u}{\partial x}\right) \qquad x \in]0,L[,\ t \in]0,T[\\ \hat{u}(0,t) = \hat{u}(L,t) = 0 \qquad t \in [0,T] \\ \hat{u}(x,0) = 0 \qquad x \in [0,L] \end{cases}$$

Adjoint model

$$\begin{cases} \dfrac{\partial p}{\partial t} + \dfrac{\partial}{\partial x}\left(K(x)\dfrac{\partial p}{\partial x}\right) = u - u^{\text{obs}} \qquad x \in]0,L[,\ t \in]0,T[\\ p(0,t) = p(L,t) = 0 \qquad t \in [0,T] \\ p(x,T) = 0 \quad x \in [0,L] \quad \text{final condition !!} \rightarrow \text{backward integration} \end{cases}$$

Gâteaux derivative of J

$$\hat{J}[K](k) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\, dx\, dt$$
$$= \int_0^T \int_0^L k(x)\frac{\partial u}{\partial x}\frac{\partial p}{\partial x}\, dx\, dt$$

Gradient of J

$$\nabla J = \int_0^T \frac{\partial u}{\partial x}(.,t)\frac{\partial p}{\partial x}(.,t)\, dt \qquad \text{function of } x$$

Discrete version:

same as for the preceding ODE, but with $\displaystyle\sum_{n=0}^{N}\sum_{i=1}^{I} u_i^n$

Matrix interpretation: $\mathbf{M}$ is much more complex than previously !!

A nonlinear example: the Burgers' equation

The assimilation problem

$$\left\{ \begin{array}{ll} \dfrac{\partial u}{\partial t} + u \, \dfrac{\partial u}{\partial x} - \nu \, \dfrac{\partial^2 u}{\partial x^2} = f & x \in]0, L[, t \in [0, T] \\ u(0,t) = \psi_1(t) & t \in [0, T] \\ u(L,t) = \psi_2(t) & t \in [0, T] \\ u(x,0) = u_0(x) & x \in [0, L] \end{array} \right.$$

- $u_0(x)$ is unknown
- $u^{\text{obs}}(x,t)$ an observation of $u(x,t)$
- Cost function: $J(u_0) = \dfrac{1}{2} \displaystyle\int_0^T \int_0^L \left(u(x,t) - u^{\text{obs}}(x,t) \right)^2 \, dx \, dt$

Gâteaux derivative

$$\hat{J}[u_0](h_0) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\, dx\, dt$$

Gâteaux derivative

$$\hat{J}[u_0](h_0) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\, dx\, dt$$

Tangent linear model

$$\left\{\begin{array}{ll} \dfrac{\partial \hat{u}}{\partial t} + \dfrac{\partial (u\hat{u})}{\partial x} - \nu \dfrac{\partial^2 \hat{u}}{\partial x^2} = 0 & x \in]0,L[,\ t \in [0,T] \\ \hat{u}(0,t) = 0 & t \in [0,T] \\ \hat{u}(L,t) = 0 & t \in [0,T] \\ \hat{u}(x,0) = h_0(x) & x \in [0,L] \end{array}\right.$$

Gâteaux derivative

$$\hat{J}[u_0](h_0) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{obs}(x,t)\right)\, dx\, dt$$

Tangent linear model

$$\begin{cases} \dfrac{\partial \hat{u}}{\partial t} + \dfrac{\partial (u\hat{u})}{\partial x} - \nu \dfrac{\partial^2 \hat{u}}{\partial x^2} = 0 & x \in]0, L[,\, t \in [0, T] \\ \hat{u}(0,t) = 0 \qquad t \in [0, T] \\ \hat{u}(L,t) = 0 \qquad t \in [0, T] \\ \hat{u}(x,0) = h_0(x) \qquad x \in [0, L] \end{cases}$$

Adjoint model

$$\begin{cases} \dfrac{\partial p}{\partial t} + u \dfrac{\partial p}{\partial x} + \nu \dfrac{\partial^2 p}{\partial x^2} = \left(u - u^{obs}\right) & x \in]0, L[,\, t \in [0, T] \\ p(0,t) = 0 \qquad t \in [0, T] \\ p(L,t) = 0 \qquad t \in [0, T] \\ p(x,T) = 0 \qquad x \in [0, L] & \text{final condition !!} \rightarrow \text{backward integration} \end{cases}$$

Gâteaux derivative of J

$$\hat{J}[u_0](h_0) = \int_0^T \int_0^L \hat{u}(x,t)\left(u(x,t) - u^{\text{obs}}(x,t)\right)\, dx\, dt$$
$$= -\int_0^L h_0(x)p(x,0)\, dx$$

Gradient of J

$$\nabla J = -p(.,0) \qquad \text{function of } x$$

Derivation and validation of an adjoint code

Writing an adjoint code

- obeys systematic rules
- is not the most interesting task you can imagine
- there exists automatic differentiation softwares:
  $\longrightarrow$ cf http://www.autodiff.org

Validation tests

- of the tangent linear model: compare $M(x+\delta x) - M(x)$ with $\mathbf{M}[x](\delta x)$ for small values of $\|\delta x\|$
- of the adjoint model: compare $(\mathbf{M}x, z)$ with $(x, \mathbf{M}^* z)$
- of the gradient: compare the directional derivative $(\nabla J(x), d)$ with the growth rate $\dfrac{J(x+\alpha d) - J(x)}{\alpha}$ (where $\nabla J(x)$ is the gradient given by the adjoint code)

Possible other uses for an adjoint model

The (local) sensitivity problem

How much is a particular model output Z_{out} sensitive to any change in a particular model input c_{in} ? $\longrightarrow \nabla_{c_{in}} Z_{out}$

$$J_o(\mathbf{x}_0) = \frac{1}{2} \sum_{i=0}^{N} \| H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i) \|_o^2 \quad \text{is replaced by } Z_{out}(c_{in}).$$

Possible other uses for an adjoint model

The (local) sensitivity problem

How much is a particular model output Z_{out} sensitive to any change in a particular model input c_{in} ? $\longrightarrow \nabla_{c_{in}} Z_{out}$

$J_o(\mathbf{x}_0) = \frac{1}{2} \sum_{i=0}^{N} \|H_i(M_{0\to t_i}(\mathbf{x}_0)) - \mathbf{y}(t_i)\|_o^2$ is replaced by $Z_{out}(c_{in})$.

The stability problem

Let consider a dynamical system: $\mathbf{x}(t)$ the state vector, $M_{t_1\to t_2}$ the model between t_1 and t_2.

Find the optimal perturbation $\mathbf{z}_1^*(t_1)$ that maximizes

$$\rho(\mathbf{z}(t_1)) = \frac{\|M_{t_1\to t_2}(\mathbf{x}(t_1) + \mathbf{z}(t_1)) - M_{t_1\to t_2}(\mathbf{x}(t_1))\|}{\|\mathbf{z}(t_1)\|}$$

$\longrightarrow$ leading eigenvectors of $\mathbf{M}^*_{t_1\to t_2}\mathbf{M}_{t_1\to t_2}$ (singular vector theory)

Summary

In summary

- Several methods, either variational or statistical, that faces the same difficulties: non linearities, huge dimension, poorly known error statistics...
- Variational methods:
  - a series of approximations of the cost function, corresponding to a series of methods
  - the more sophisticated ones (4D-Var, incremental 4D-Var) require the tangent linear and adjoint models (the development of which is a real investment)
- Statistical methods:
  - extended Kalman filter handle (weakly) non linear problems (requires the TL model)
  - reduced order Kalman filters address huge dimension problems
  - a quite efficient method, addressing both problems: ensemble Kalman filters (EnKF)
  - these are so called "Gaussian filters"

  - particle filters: currently being developed - fully Bayesian approach - still limited to low dimension problems

Some present research directions

- new methods: less expensive, more robust w.r.t. nonlinearities and/or non gaussianity (particle filters, En4DVar, BFN...)
- better management of errors (prior statistics, identification, a posteriori validation...)
- "complex" observations (images, Lagrangian data...)
- new application domains (often leading to new methodological questions)
- definition of observing systems, sensitivity analysis...

Some references

1. Blayo E. and M. Nodet, 2012: Introduction à l'assimilation de données variationnelle. *Lecture notes for UJF Master course on data assimilation*.
   https://team.inria.fr/moise/files/2012/03/Methodes-Inverses-Var-M2-math-2009.pdf
2. Bocquet M., 2014 : Introduction aux principes et méthodes de l'assimilation de données en géophysique. *Lecture notes for ENSTA - ENPC data assimilation course*.
   http://cerea.enpc.fr/HomePages/bocquet/Doc/assim-mb.pdf
3. Bocquet M., E. Cosme and E. Blayo (Eds.), 2014: Advanced Data Assimilation for Geosciences. *Oxford University Press*.
4. Bouttier F. and P. Courtier, 1999: Data assimilation, concepts and methods. *Meteorological training course lecture series ECMWF*, European Center for Medium range Weather Forecast, Reading, UK.
   http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts21.html
5. Cohn S., 1997: An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, **75**, 257-288.
6. Daley R., 1993: Atmospheric data analysis. *Cambridge University Press*.
7. Evensen G., 2009: Data assimilation, the ensemble Kalman filter. *Springer*.
8. Kalnay E., 2003: Atmospheric modeling, data assimilation and predictability. *Cambridge University Press*.
9. Lahoz W., B. Khattatov and R. Menard (Eds.), 2010: Data assimilation. *Springer*.
10. Rodgers C., 2000: Inverse methods for atmospheric sounding. *World Scientific, Series on Atmospheric Oceanic and Planetary Physics*.
11. Tarantola A., 2005: Inverse problem theory and methods for model parameter estimation. *SIAM*.
    http://www.ipgp.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf

Two announcements

- CNA 2014: 5ème Colloque National d'Assimilation de données
  Toulouse, 1-3 décembre 2014

- Doctoral course "Introduction to data assimilation"
  Grenoble, January 5-9, 2015