



## An introduction to data assimilation

Eric Blayo Univ. Grenoble Alpes and Inria Data assimilation, the science of compromises

Context characterizing a (complex) system and/or forecasting its evolution, given several heterogeneous and uncertain sources of information



Widely used for geophysical fluids (meteorology, oceanography, atmospheric chemistry...), but also in other numerous domains (e.g. nuclear energy, medicine, agriculture planning...)

Closely linked to inverse methods, control theory, estimation theory, filtering...



E. Blayo - An introduction to data assimilation

#### A daily example: numerical weather forecast



UNIVERSITÉ



## Data assimilation, the science of compromises

Numerous possible aims:

- Forecast: estimation of the present state (initial condition)
- Model tuning: parameter estimation
- Inverse modeling: estimation of parameter fields
- Data analysis: re-analysis (model = interpolation operator)
- OSSE: optimization of observing systems





#### The best estimate

Several pieces of information:

- Model
- Prior (or background) value

 $\longrightarrow$  Find the best possible estimate  $\mathbf{x}^a$ 

- Observations
- Statistics



Alpes



#### The best estimate

Several pieces of information:

- Model
- Prior (or background) value

 $\longrightarrow$  Find the best possible estimate  $x^a$ 

- Observations
- Statistics

. . .

What does the best possible estimate means?

Estimate: deterministic value? pdf? some moments only of a pdf?

Best: which criterion?



Workshop TDMF, Orsay, 2017 5/57

#### Objectives for this lecture

- introduce data assimilation and its several points of view
- give an overview of the main families of methods
- point out the main difficulties and current corresponding answers



#### Objectives for this lecture

- introduce data assimilation and its several points of view
- give an overview of the main families of methods
- point out the main difficulties and current corresponding answers

#### Outline

- 1. Data assimilation for dummies: a simple model problem
- 2. Generalization: linear estimation theory, variational and sequential approaches

Workshop TDMF, Orsav, 2017

3. Some current challenges



# The simplest possible model problem







Two pieces of information on a single quantity. Which estimation for its true value ?  $\longrightarrow$  least squares approach





Two pieces of information on a single quantity. Which estimation for its true value ?  $\longrightarrow$  least squares approach

**Example** a prior value  $x^b = 19^{\circ}$ C and an observation  $y = 21^{\circ}$ C of the (unknown) present temperature x.

• Let 
$$J(x) = \frac{1}{2} \left[ (x - x^b)^2 + (x - y)^2 \right]$$

• 
$$\operatorname{Min}_{x} J(x) \longrightarrow x^{a} = \frac{x^{b} + y}{2} = 20^{\circ} \mathrm{C}$$



E. Blayo - An introduction to data assimilation

Two pieces of information on a single quantity. Which estimation for its true value ?  $\longrightarrow$  least squares approach

**Example** a prior value  $x^b = 19^{\circ}$ C and an observation  $y = 21^{\circ}$ C of the (unknown) present temperature x.

• Let 
$$J(x) = \frac{1}{2} \left[ (x - x^b)^2 + (x - y)^2 \right]$$

• 
$$\operatorname{Min}_{x} J(x) \longrightarrow x^{a} = \frac{x^{b} + y}{2} = 20^{\circ} \mathrm{C}$$

If  $\neq$  units:  $x^b = 66.2^{\circ}$ F and  $y = 69.8^{\circ}$ F

► Let 
$$H(x) = \frac{9}{5}x + 32$$
 observation operator  
Let  $J(x) = \frac{1}{2} [(H(x) - x^b)^2 + (H(x) - y)^2]$   
Min<sub>x</sub>  $J(x) \longrightarrow x^a = 20^\circ C$ 

E. Blayo - An introduction to data assimilation



**Drawback # 1:** if observation units are inhomogeneous

$$x^{b} = 19^{\circ}\text{C and } y = 69.8^{\circ}\text{F}$$

$$J(x) = \frac{1}{2} \left[ (x - x^{b})^{2} + (H(x) - y)^{2} \right] \longrightarrow x^{a} = 20.53^{\circ}\text{C}$$

$$\longrightarrow adding apples and oranges !!$$



E. Blayo - An introduction to data assimilation

**Drawback # 1:** if observation units are inhomogeneous

$$x^{b} = 19^{\circ}\text{C and } y = 69.8^{\circ}\text{F}$$

$$J(x) = \frac{1}{2} \left[ (x - x^{b})^{2} + (H(x) - y)^{2} \right] \longrightarrow x^{a} = 20.53^{\circ}\text{C}$$

$$\longrightarrow adding apples and oranges !!$$

**Drawback # 2:** if observation accuracies are inhomogeneous

If  $x^b$  is twice more accurate than y, one should obtain  $x^a = \frac{2x^b + y}{3} = 19.67^{\circ}\text{C}$ 

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



Workshop TDMF, Orsay, 2017

Informatics mathematics

Reformulation in a probabilistic framework:

- the goal is to find an estimator  $X^a$  of the true unknown value x
- $x^b$  and y are realizations of random variables  $X^b$  and Y
- One is looking for an estimator (i.e. a r.v.)  $X^a$  that is
  - ► linear:  $X^a = \alpha_b X^b + \alpha_o Y$  (in order to be simple)
  - **unbiased**:  $E(X^a) = x$  (it seems reasonable)
  - ▶ of minimal variance: Var(X<sup>a</sup>) minimum (optimal accuracy)

→ BLUE (Best Linear Unbiased Estimator)



Let 
$$X^b = x + \varepsilon^b$$
 and  $Y = x + \varepsilon^o$  with

#### Hypotheses

•  $E(\varepsilon^{b}) = E(\varepsilon^{o}) = 0$  unbiased background and measurement device

• 
$$Var(\varepsilon^b) = \sigma_b^2$$
  $Var(\varepsilon^o) = \sigma_o^2$  known accuracies

• 
$$Cov(\varepsilon^b, \varepsilon^o) = 0$$
 independent errors





Let 
$$X^b = x + \varepsilon^b$$
 and  $Y = x + \varepsilon^o$  with

#### Hypotheses

•  $E(\varepsilon^{b}) = E(\varepsilon^{o}) = 0$  unbiased background and measurement device

Var(ε<sup>b</sup>) = σ<sub>b</sub><sup>2</sup> Var(ε<sup>o</sup>) = σ<sub>o</sub><sup>2</sup> known accuracies
 Cov(ε<sup>b</sup>, ε<sup>o</sup>) = 0 independent errors

Since 
$$X^a = \alpha_b X^b + \alpha_o Y = (\alpha_b + \alpha_o) x + \alpha_b \varepsilon^b + \alpha_o \varepsilon^o$$
:

$$E(X^a) = (\alpha_b + \alpha_o)x + \alpha_b \underbrace{E(\varepsilon^b)}_{=0} + \alpha_o \underbrace{E(\varepsilon^o)}_{=0} \implies \alpha_b + \alpha_o = 1$$



Let 
$$X^b = x + \varepsilon^b$$
 and  $Y = x + \varepsilon^o$  with

#### Hypotheses

•  $E(\varepsilon^{b}) = E(\varepsilon^{o}) = 0$  unbiased background and measurement device

Var(ε<sup>b</sup>) = σ<sub>b</sub><sup>2</sup> Var(ε<sup>o</sup>) = σ<sub>o</sub><sup>2</sup> known accuracies
 Cov(ε<sup>b</sup>, ε<sup>o</sup>) = 0 independent errors

Since 
$$X^{a} = \alpha_{b}X^{b} + \alpha_{o}Y = (\alpha_{b} + \alpha_{o})x + \alpha_{b}\varepsilon^{b} + \alpha_{o}\varepsilon^{o}$$
:

$$E(X^a) = (\alpha_b + \alpha_o)x + \alpha_b \underbrace{E(\varepsilon^b)}_{=0} + \alpha_o \underbrace{E(\varepsilon^o)}_{=0} \implies \alpha_b + \alpha_o = 1$$

Var
$$(X^a) = E\left[(X^a - x)^2\right] = E\left[(\alpha_b\varepsilon^b + \alpha_o\varepsilon^o)^2\right] = \alpha_b^2\sigma_b^2 + (1 - \alpha_b)^2\sigma_o^2$$

$$\frac{\partial}{\partial \alpha_b} = 0 \quad \Longrightarrow \quad \alpha_b = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2}$$

E. Blayo - An introduction to data assimilation

# BLUE $X^{a} = \frac{\frac{1}{\sigma_{b}^{2}}X^{b} + \frac{1}{\sigma_{o}^{2}}Y}{\frac{1}{\sigma_{b}^{2}} + \frac{1}{\sigma_{o}^{2}}}$





### BLUE

$$X^{a} = \frac{\frac{1}{\sigma_{b}^{2}}X^{b} + \frac{1}{\sigma_{o}^{2}}Y}{\frac{1}{\sigma_{b}^{2}} + \frac{1}{\sigma_{o}^{2}}} = X^{b} + \underbrace{\frac{\sigma_{b}^{2}}{\sigma_{b}^{2} + \sigma_{o}^{2}}}_{\text{gain}} \underbrace{(Y - X^{b})}_{\text{innovation}}$$









E. Blayo - An introduction to data assimilation



► Hypotheses on the two first moments of ε<sup>b</sup>, ε<sup>o</sup> lead to results on the two first moments of X<sup>a</sup>.



#### Variational equivalence

This is equivalent to the problem:

Minimize 
$$J(x) = \frac{1}{2} \left[ \frac{(x - x^b)^2}{\sigma_b^2} + \frac{(x - y)^2}{\sigma_o^2} \right]$$





#### Variational equivalence

This is equivalent to the problem:

Minimize 
$$J(x) = rac{1}{2} \left[ rac{(x-x^b)^2}{\sigma_b^2} + rac{(x-y)^2}{\sigma_o^2} 
ight]$$

#### **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

• 
$$\underbrace{J''(x^a)}_{\text{convexity}} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} = \underbrace{[\operatorname{Var}(x^a)]^{-1}}_{\operatorname{accuracy}}$$

**Geometric interpretation**  $E(\varepsilon^{o}\varepsilon^{b}) = 0 \implies E(\varepsilon^{a}(Y - X_{b})) = 0$ 



 $\rightarrow$  orthogonal projection for the scalar product  $\langle Z_1, Z_2 \rangle = E(Z_1Z_2)$  for unbiased random variables.



• x: a realization of a random variable X. What is the pdf p(X|Y)?

Based on the Bayes rule:





E. Blayo - An introduction to data assimilation

• x: a realization of a random variable X. What is the pdf p(X|Y)?

Based on the Bayes rule:



Back to our example:

- Background  $X^b \sim \mathcal{N}(19, \sigma_b^2)$
- Observation  $y = 21^{\circ}$ C, and  $Y = X + \varepsilon^{o}$  with  $\varepsilon^{o} \rightsquigarrow \mathcal{N}(0, \sigma_{o}^{2})$

• Background 
$$X^b \rightsquigarrow \mathcal{N}(19, \sigma_b^2)$$

• Observation  $y = 21^{\circ}$ C, and  $Y = X + \varepsilon^{o}$  with  $\varepsilon^{o} \rightsquigarrow \mathcal{N}(0, \sigma_{o}^{2})$ 

$$P(X = x | Y = 21) = \frac{P(Y = 21 | X = x) P(X = x)}{P(Y = y)}$$



E. Blayo - An introduction to data assimilation

• Prior: 
$$P(X = x) = P(X^b = x) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(\frac{(x-19)^2}{2\sigma_b^2}\right)$$

Likelihood:

Alpes

$$p(Y = 21 | X = x) = p(\varepsilon^{o} = 21 - x | X = x)$$
  
=  $p(\varepsilon^{o} = 21 - x) \quad \varepsilon^{o} \text{ is assumed independent from } X$   
=  $\frac{1}{\sqrt{2\pi} \sigma_{o}} \exp\left(-\frac{(21 - x)^{2}}{2 \sigma_{o}^{2}}\right)$ 

• Background 
$$X^b \sim \mathcal{N}(19, \sigma_b^2)$$

▶ Observation  $y = 21^{\circ}$ C, and  $Y = X + \varepsilon^{o}$  with  $\varepsilon^{o} \rightsquigarrow \mathcal{N}(0, \sigma_{o}^{2})$ 

$$P(X = x | Y = 21) = \frac{P(Y = 21 | X = x) P(X = x)}{P(Y = y)}$$

Hence

$$p(X = x) p(Y = 21 | X = x) = \frac{1}{\sqrt{2\pi\sigma_b}} \exp\left(-\frac{(x - 19)^2}{2\sigma_b^2}\right) \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{(21 - x)^2}{2\sigma_o^2}\right)$$
$$= K \exp\left(-\frac{(x - m_a)^2}{2\sigma_a^2}\right)$$
with  $m_a = \frac{\frac{1}{\sigma_b^2} \frac{19 + \frac{1}{\sigma_o^2}}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}}$  and  $\sigma_a^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)^{-1}$ 



- Background  $X^b \sim \mathcal{N}(19, \sigma_b^2)$
- ▶ Observation  $y = 21^{\circ}$ C, and  $Y = X + \varepsilon^{o}$  with  $\varepsilon^{o} \rightsquigarrow \mathcal{N}(0, \sigma_{o}^{2})$

$$P(X = x | Y = 21) = \frac{P(Y = 21 | X = x) P(X = x)}{P(Y = y)}$$

Hence

$$p(X = x) p(Y = 21 | X = x) = \frac{1}{\sqrt{2\pi\sigma_b}} \exp\left(-\frac{(x - 19)^2}{2\sigma_b^2}\right) \frac{1}{\sqrt{2\pi\sigma_o}} \exp\left(-\frac{(21 - x)^2}{2\sigma_o^2}\right)$$
$$= K \exp\left(-\frac{(x - m_a)^2}{2\sigma_a^2}\right)$$
with  $m_a = \frac{\frac{1}{\sigma_b^2} 19 + \frac{1}{\sigma_o^2} 21}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}}$  and  $\sigma_a^2 = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)^{-1}$ 

$$\longrightarrow X \mid Y = 21 \rightsquigarrow \mathcal{N}(m_a, \sigma_a^2)$$









Same as the BLUE because of Gaussian hypothesis





### Model problem: synthesis

Data assimilation methods are often split into 2-3 families:

- Variational methods: minimization of a cost function (least squares approach)
- Linear statistical approach: computation of the BLUE (with hypotheses on the first two moments)
- Bayesian approach: approximation of pdfs (with hypotheses on the pdfs)
- There are strong links between those approaches, depending on the case (linear, Gaussian...)



#### Model problem: synthesis

Data assimilation methods are often split into 2-3 families:

- Variational methods: minimization of a cost function (least squares approach)
- Linear statistical approach: computation of the BLUE (with hypotheses on the first two moments)
- Bayesian approach: approximation of pdfs (with hypotheses on the pdfs)
- There are strong links between those approaches, depending on the case (linear, Gaussian...)

#### Theorem

If you have understood this previous stuff, you have understood (almost) everything on data assimilation.



## Generalization: variational approach



E. Blayo - An introduction to data assimilation


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  $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$ 



E. Blayo - An introduction to data assimilation

#### A simple example of observation operator

If 
$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$
 and  $\mathbf{y} = \begin{pmatrix} \text{an observation of } \frac{x_1 + x_2}{2} \\ \text{an observation of } x_4 \end{pmatrix}$   
then  $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$  with  $\mathbf{H} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ 



E. Blayo - An introduction to data assimilation

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  $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$ 

Cost function:  $J(\mathbf{x}) = \frac{1}{2} ||H(\mathbf{x}) - \mathbf{y}||^2$  with ||.|| to be chosen.



E. Blayo - An introduction to data assimilation

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  $H : \mathbf{R}^n \longrightarrow \mathbf{R}^p$ 

Cost function: 
$$J(\mathbf{x}) = \frac{1}{2} ||H(\mathbf{x}) - \mathbf{y}||^2$$
 with  $||.||$  to be chosen.

#### Remark

(Intuitive) necessary (but not sufficient) condition for the existence of a unique minimum:

$$p \ge n$$

## Formalism "background value + new observations"

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}^b \\ \mathbf{y} \end{pmatrix} \quad \xleftarrow{} \begin{array}{c} \mathsf{background} \\ \longleftarrow \ \mathsf{new \ observations} \\ \end{array}$$

The cost function becomes:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{x} - \mathbf{x}^b\|_b^2}_{J_b} + \underbrace{\frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|_o^2}_{J_o}$$



E. Blayo - An introduction to data assimilation

## Formalism "background value + new observations"

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}^b \\ \mathbf{y} \end{pmatrix} \stackrel{\longleftarrow}{\longleftarrow} \mathbf{background}$$

The cost function becomes:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{x} - \mathbf{x}^b\|_b^2}_{J_b} + \underbrace{\frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|_o^2}_{J_o}$$

The necessary condition for the existence of a unique minimum  $(p \ge n)$  is automatically fulfilled.



• Observations are distributed in time:  $\mathbf{y} = \mathbf{y}(t)$ 

The observation cost function becomes:

$$J_o(\mathbf{x}) = rac{1}{2} \sum_{i=0}^N \|H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i)\|_o^2$$



E. Blayo - An introduction to data assimilation

- Observations are distributed in time:  $\mathbf{y} = \mathbf{y}(t)$
- The observation cost function becomes:

$$J_o(\mathbf{x}) = rac{1}{2} \sum_{i=0}^N \|H_i(\mathbf{x}(t_i)) - \mathbf{y}(t_i)\|_o^2$$

► There is a model describing the evolution of **x**:  $\frac{d\mathbf{x}}{dt} = M(\mathbf{x})$  with  $\mathbf{x}(t=0) = \mathbf{x}_0$ . Then *J* is often no longer minimized w.r.t. **x**, but w.r.t.  $\mathbf{x}_0$  only, or to some other parameters.

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





E. Blayo - An introduction to data assimilation

Grenoble

Alpes

$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

• If H and M are linear then  $J_o$  is quadratic.





$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

- If H and M are linear then  $J_o$  is quadratic.
- However J<sub>o</sub> generally does not have a unique minimum, since the number of observations is generally less than the size of x<sub>0</sub> (the problem is underdetermined: p < n).</p>

Example: let  $(x_1^t, x_2^t) = (1, 1)$  and y = 1.1 an observation of  $\frac{1}{2}(x_1 + x_2)$ .

$$J_o(x_1, x_2) = \frac{1}{2} \left( \frac{x_1 + x_2}{2} - 1.1 \right)^2$$





E. Blayo - An introduction to data assimilation

$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

- If H and M are linear then J<sub>o</sub> is quadratic.
- However it generally does not have a unique minimum, since the number of observations is generally less than the size of x<sub>0</sub> (the problem is underdetermined).
- Adding  $J_b$  makes the problem of minimizing  $J = J_o + J_b$  well posed.



$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

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

Example: the Lorenz system (1963)

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



$$J_o(y_0) = rac{1}{2} \sum_{i=0}^{N} (x(t_i) - x_{obs}(t_i))^2 dt$$

E. Blayo - An introduction to data assimilation



$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

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







$$J(\mathbf{x}_{0}) = J_{b}(\mathbf{x}_{0}) + J_{o}(\mathbf{x}_{0}) = \frac{1}{2} \|\mathbf{x}_{0} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \sum_{i=0}^{N} \|H_{i}(M_{0 \to t_{i}}(\mathbf{x}_{0})) - \mathbf{y}(t_{i})\|_{o}^{2}$$

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



► Adding J<sub>b</sub> makes it "more quadratic" (J<sub>b</sub> is a regularization term), but J = J<sub>o</sub> + J<sub>b</sub> may however have several local minima.

Grenoble

Alpes

30/57

# A fundamental remark before going into minimization aspects

Once J is defined (i.e. once all the ingredients are chosen: control variables, norms, observations...), the problem is entirely defined. Hence its solution.



The "physical" (i.e. the most important) part of variational data assimilation lies in the definition of J.

The rest of the job, i.e. minimizing J, is "only" technical work.

Implementation: 3D-VAR, 3D-FGAT, 4D-VAR, incremental 4D-VAR...



## Minimizing J

Alpes

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x})$$
  
=  $\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$ 





E. Blayo - An introduction to data assimilation

## Minimizing J

Alpes

$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x})$$
  
=  $\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$ 

#### Optimal estimation in the linear case

$$\mathbf{x}^{a} = \mathbf{x}^{b} + \underbrace{(\mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{T}\mathbf{R}^{-1}}_{\text{gain matrix}} \underbrace{(\mathbf{y} - \mathbf{H}\mathbf{x}^{b})}_{\text{innovation vector}}$$

$$J(\mathbf{x}_{0}) = \frac{1}{2} (\mathbf{x}_{0} - \mathbf{x}^{b})^{T} \mathbf{B}^{-1} (\mathbf{x}_{0} - \mathbf{x}^{b}) \\ + \frac{1}{2} \sum_{i=1}^{N} [\mathbf{y}_{i} - \mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}_{0}]^{T} \mathbf{R}_{i}^{-1} [\mathbf{y}_{i} - \mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}_{0}]$$

#### With a linear evolution model

$$\mathbf{x}^{a} = \mathbf{x}^{b} + \left[\mathbf{B}^{-1} + \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} \mathbf{H}_{i} \mathbf{M}_{0,i}\right]^{-1} \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} (\mathbf{y}_{i} - \mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}^{b})$$



E. Blayo - An introduction to data assimilation



Given the size of n and p, it is generally impossible to handle explicitly **H**, **B** and **R**. So the direct computation of the gain matrix is impossible.

▶ even in the linear case (for which we have an explicit expression for  $x^a$ ), the computation of  $x^a$  is performed using an optimization algorithm.





## Minimizing J: descent methods

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

. . .

with 
$$\mathbf{d}_{k} = \begin{cases} -\nabla J(\mathbf{x}_{k}) \\ -[\operatorname{Hess}(J)(\mathbf{x}_{k})]^{-1} \nabla J(\mathbf{x}_{k}) \\ -\mathbf{B}_{k} \nabla J(\mathbf{x}_{k}) \\ -\nabla J(\mathbf{x}_{k}) + \frac{\|\nabla J(\mathbf{x}_{k})\|^{2}}{\|\nabla J(\mathbf{x}_{k-1})\|^{2}} d_{k-1} \\ \dots \end{cases}$$

gradient method Newton method quasi-Newton methods (BFGS, ...) conjugate gradient



iso-I curves

E. Blayo - An introduction to data assimilation

## Getting the gradient is not obvious

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

#### Example:

$$\begin{pmatrix} \frac{d\mathbf{x}(t)}{dt} = M(\mathbf{x}(t)) & t \in [0, T] \\ \mathbf{x}(t=0) = \mathbf{u} & \text{with } \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

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

$$\nabla J(\mathbf{u}) = \begin{pmatrix} \frac{\partial J}{\partial u_1}(\mathbf{u}) \\ \vdots \\ \frac{\partial J}{\partial u_n}(\mathbf{u}) \end{pmatrix} \simeq \begin{pmatrix} [J(\mathbf{u} + \alpha \, \mathbf{e}_1) - J(\mathbf{u})] / \alpha \\ \vdots \\ [J(\mathbf{u} + \alpha \, \mathbf{e}_n) - J(\mathbf{u})] / \alpha \end{pmatrix} \longrightarrow n + 1 \text{ model runs}$$

E. Blayo - An introduction to data assimilation

## Getting the gradient is not obvious

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

In such cases, the adjoint method provides an efficient way to compute  $\nabla J$ .







Example: an adjoint for the Burgers' equation

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f \quad x \in ]0, L[, t \in [0, T] \\ u(0, t) = \psi_1(t) \quad u(L, t) = \psi_2(t) \quad t \in [0, T] \\ u(x, 0) = u_0(x) \quad x \in [0, L] \end{cases}$$

•  $u^{obs}(x, t)$  an observation of u(x, t)

4

• Cost function: 
$$J(u_0) = \frac{1}{2} \int_0^T \int_0^L (u(x,t) - u^{obs}(x,t))^2 dx dt$$



E. Blayo - An introduction to data assimilation

Example: an adjoint for the Burgers' equation

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f \quad x \in ]0, L[, t \in [0, T] \\ u(0, t) = \psi_1(t) \quad u(L, t) = \psi_2(t) \quad t \in [0, T] \\ u(x, 0) = u_0(x) \quad x \in [0, L] \end{cases}$$

•  $u^{obs}(x,t)$  an observation of u(x,t)

• Cost function: 
$$J(u_0) = \frac{1}{2} \int_0^T \int_0^L (u(x,t) - u^{obs}(x,t))^2 dx dt$$

#### Adjoint model

4

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



Example: an adjoint for the Burgers' equation

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f \quad x \in ]0, L[, t \in [0, T] \\ u(0, t) = \psi_1(t) \quad u(L, t) = \psi_2(t) \quad t \in [0, T] \\ u(x, 0) = u_0(x) \quad x \in [0, L] \end{cases}$$

•  $u^{obs}(x,t)$  an observation of u(x,t)

• Cost function: 
$$J(u_0) = \frac{1}{2} \int_0^T \int_0^L (u(x,t) - u^{obs}(x,t))^2 dx dt$$

#### Adjoint model

4

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

#### Gradient of J

Alpes

$$abla J = -p(.,0)$$
 function

E. Blayo - An introduction to data assimilation

of x

37/57

## Getting the gradient is not obvious

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

In such cases, the adjoint method provides an efficient way to compute  $\nabla J$ .

It requires writing a tangent linear code and an adjoint code (beyond the scope of this lecture):

- obeys systematic rules
- is not the most interesting task you can imagine



## Generalization: linear statistical approach







Generalization: linear statistical approach To be estimated:  $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$  Observations:  $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$ 

Linear observation operator:  $\mathbf{y} \equiv H(\mathbf{x}) = \mathbf{H}\mathbf{x}$ 

#### Statistical framework:

- > y is a realization of a random vector Y
- ▶ One is looking for the BLUE, i.e. a r.v. X<sup>a</sup> that is
  - linear:  $X^a = AY$  with size(A) = (n, p)
  - unbiased:  $E(\mathbf{X}^a) = \mathbf{x}$
  - of minimal variance:  $Var(\mathbf{X}^a) = \sum_{i=1}^{n} Var(X_i^a)$  minimum



Generalization: linear statistical approach To be estimated:  $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$  Observations:  $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \in \mathbf{R}^p$ 

Linear observation operator:  $\mathbf{y} \equiv H(\mathbf{x}) = \mathbf{H}\mathbf{x}$ 

#### Statistical framework:

- > y is a realization of a random vector Y
- ▶ One is looking for the BLUE, i.e. a r.v. X<sup>a</sup> that is
  - linear:  $X^a = AY$  with size(A) = (n, p)
  - unbiased:  $E(\mathbf{X}^a) = \mathbf{x}$
  - of minimal variance:  $Var(\mathbf{X}^a) = \sum_{i=1}^{n} Var(X_i^a)$  minimum

Gauss-Markov theorem:  $\mathbf{A} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}$ 

## Generalization: linear statistical approach

Background:  $\mathbf{X}^b = \mathbf{x} + \varepsilon^b$  and new observations:  $\mathbf{Y} = H(\mathbf{x}) + \varepsilon^o$ 

#### Hypotheses:

- $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$  linear observation operator
- $E(\varepsilon^b) = 0$  and  $E(\varepsilon^o) = 0$  unbiased background and observations
- $Cov(\varepsilon^b, \varepsilon^o) = 0$  independent background and observation errors

• 
$$Cov(\varepsilon^b) = \mathbf{B}$$
 and  $Cov(\varepsilon^o) = \mathbf{R}$  known accuracies and covariances

#### BLUE

$$\mathbf{X}^{a} = \mathbf{X}^{b} + \underbrace{(\mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{T}\mathbf{R}^{-1}}_{\text{gain matrix}} \underbrace{(\mathbf{Y} - \mathbf{H}\mathbf{X}^{b})}_{\text{innovation vector}}$$
  
with  $[\operatorname{Cov}(\mathbf{X}^{a})]^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H}$  accuracies are added



## Link with the variational approach

#### Statistical approach: BLUE

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

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

Variational approach in the linear case

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{x}^b\|_b^2 + \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|_o^2$$
  
=  $\frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$   
min  $J(\mathbf{x}) \longrightarrow \mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^b)$ 



E. Blayo - An introduction to data assimilation

Alpes

## Link with the variational approach

#### Statistical approach: BLUE

$$\mathbf{X}^{a} = \mathbf{X}^{b} + (\mathbf{B}^{-1} + \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{H} \mathbf{X}^{b})$$

1

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

Variational approach in the linear case

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{x}^{b}\|_{b}^{2} + \frac{1}{2} \|H(\mathbf{x}) - \mathbf{y}\|_{o}^{2}$$
  
$$= \frac{1}{2} (\mathbf{x} - \mathbf{x}^{b})^{T} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^{b}) + \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y})^{T} \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$$
  
$$\min_{e \in \mathbf{R}^{n}} J(\mathbf{x}) \longrightarrow \mathbf{x}^{a} = \mathbf{x}^{b} + (\mathbf{B}^{-1} + \mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^{T} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H} \mathbf{x}^{b})$$

#### Same remarks as previously

The linear statistical approach rationalizes the choice of the norms for J<sub>o</sub> and J<sub>b</sub> in the variational approach.

Workshop TDMF, Orsay, 2017

42/5

$$\underbrace{[\operatorname{Cov}(\mathbf{X}^{a})]^{-1}}_{\operatorname{accuracy}} = \mathbf{B}^{-1} + \mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H} = \underbrace{\operatorname{Hess}(J)}_{\operatorname{convexity}}$$

E. Blayo - An introduction to data assimilation

Dynamical system:  $\mathbf{x}^{t}(t_{i+1}) = \mathbf{M}_{i,i+1}\mathbf{x}^{t}(t_{i}) + \varepsilon^{m}(t_{i})$ 





Dynamical system:  $\mathbf{x}^{t}(t_{i+1}) = \mathbf{M}_{i,i+1}\mathbf{x}^{t}(t_{i}) + \varepsilon^{m}(t_{i})$ 

• Direct application of the BLUE on  $[t_0, t_N]$  (hyp:  $\epsilon^m = 0$ ):

$$\mathbf{x}^{a} = \mathbf{x}^{b} + \left[\mathbf{B}^{-1} + \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} \mathbf{H}_{i} \mathbf{M}_{0,i}\right]^{-1} \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} (\mathbf{y}_{i} - \mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}^{b})$$

 $\longrightarrow$  4D-Var algorithm



E. Blayo - An introduction to data assimilation

Dynamical system:  $\mathbf{x}^{t}(t_{i+1}) = \mathbf{M}_{i,i+1}\mathbf{x}^{t}(t_{i}) + \varepsilon^{m}(t_{i})$ 

• Direct application of the BLUE on  $[t_0, t_N]$  (hyp:  $\epsilon^m = 0$ ):

$$\mathbf{x}^{a} = \mathbf{x}^{b} + \left[\mathbf{B}^{-1} + \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} \mathbf{H}_{i} \mathbf{M}_{0,i}\right]^{-1} \sum_{i=1}^{N} \mathbf{M}_{0,i}^{T} \mathbf{H}_{i}^{T} \mathbf{R}_{i}^{-1} (\mathbf{y}_{i} - \mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}^{b})$$
$$\longrightarrow 4D \text{-Var algorithm}$$

Sequential application of the BLUE every observation time:



## Kalman filter

#### Hypotheses

- $\varepsilon^m(t_i)$  is unbiased, with covariance matrix  $\mathbf{Q}_i$
- $\varepsilon^m(t_i)$  and  $\varepsilon^m(t_j)$  are independent  $(i \neq j)$
- Unbiased observation y<sub>i</sub>, with error covariance matrix R<sub>i</sub>
- $\varepsilon^m(t_i)$  and analysis error  $\mathbf{x}^a(t_i) \mathbf{x}^t(t_i)$  are independent

#### Evolution of the first two moments - Kalman filter

 $\frac{\text{Initialization}}{\mathbf{P}^{a}(t_{0})} = \mathbf{x}^{b}$  $\mathbf{P}^{a}(t_{0}) = \mathbf{B}$ 

 $\underbrace{ \underbrace{Step i:}_{i \in I} (prediction - correction, or forecast - analysis) \\ x^{f}(t_{i+1}) &= M_{i,i+1} x^{a}(t_{i}) & Forecast \\ P^{f}(t_{i+1}) &= M_{i,i+1} P^{a}(t_{i}) M_{i,i+1}^{T} + Q_{i} \\ \\ x^{a}(t_{i+1}) &= x^{f}(t_{i+1}) + K_{i+1} [y_{i+1} - H_{i+1} x^{f}(t_{i+1})] \\ K_{i+1} &= P^{f}(t_{i+1}) H_{i+1}^{T} [H_{i+1} P^{f}(t_{i+1}) H_{i+1}^{T} + R_{i+1}]^{-1} \\ P^{a}(t_{i+1}) &= P^{f}(t_{i+1}) - K_{i+1} H_{i+1} P^{f}(t_{i+1}) \\ \end{aligned}$ 


### Kalman filter and 4D-Var



E. Blayo - An introduction to data assimilation

Alpes

#### Kalman filter and 4D-Var

Alpes

If  $\mathbf{H}_i$  and  $\mathbf{M}_{i,i+1}$  are linear, and if the model is perfect ( $\varepsilon^m(t_i) = 0$ ), then the Kalman filter and the variational method minimizing

$$J(\mathbf{x}_{0}) = \frac{1}{2} (\mathbf{x}_{0} - \mathbf{x}^{b})^{T} \mathbf{B}^{-1} (\mathbf{x}_{0} - \mathbf{x}^{b}) + \frac{1}{2} \sum_{i=0}^{N} (\mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}_{0} - \mathbf{y}_{i})^{T} \mathbf{R}_{i}^{-1} (\mathbf{H}_{i} \mathbf{M}_{0,i} \mathbf{x}_{0} - \mathbf{y}_{i})$$
  
lead to the same solution at  $t = t_{N}$ .





E. Blayo - An introduction to data assimilation



Several types of problem:

Filtering: 
$$p(\mathbf{X}_N | \mathbf{Y}_{1:N} = \mathbf{y}_{1:N})$$

Forecast: 
$$p(\mathbf{X}_{l}|\mathbf{Y}_{1:N} = \mathbf{y}_{1:N})$$
  $(l > N)$ 

- Smoothing: all other cases.
  - ▶  $p(\mathbf{X}_{l}|\mathbf{Y}_{1:N} = \mathbf{y}_{1:N})$  (l < N) : fixed-point smoothing
  - $p(\mathbf{X}_{0:N}|\mathbf{Y}_{1:N} = \mathbf{y}_{1:N})$  : fixed-interval smoothing



...

Workshop TDMF, Orsay, 2017 48/57

Two tools:

Bayes theorem: 
$$P(X = x | Y = y) = \underbrace{\frac{P(Y = y | X = x)}{P(Y = y)} P(X = x)}_{primalisation factor}$$

• Marginalization rule: 
$$p(\mathbf{X}) = \int p(\mathbf{X} | \mathbf{Z}) p(\mathbf{Z}) d\mathbf{Z}$$

And some usual hypotheses:

- $\varepsilon^{o}$  is independent from past and present states
- $\blacktriangleright \ \varepsilon^m$  is dependent at most of the present state, but not from past states

49/57

Filtering and forecast problems can be solved by a sequential algorithm alternating two phases:

Analysis at time t<sub>i</sub> :

$$p(\mathbf{X}_i = \mathbf{x}_i | \mathbf{Y}_{1:i} = \mathbf{y}_{1:i}) \propto p(\mathbf{X}_i = \mathbf{x}_i | \mathbf{Y}_{1:i-1} = \mathbf{y}_{1:i-1}) p(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{X}_i = \mathbf{x}_i)$$

► Forecast from t<sub>i</sub> to t<sub>i+1</sub> :

$$p(\mathbf{X}_{i+1} = \mathbf{x}_{i+1} | \mathbf{Y}_{1:i} = \mathbf{y}_{1:i}) = \int p(\mathbf{X}_{i+1} = \mathbf{x}_{i+1} | \mathbf{X}_i = \mathbf{x}_i) p(\mathbf{X}_i = \mathbf{x}_i | \mathbf{Y}_{1:i} = \mathbf{y}_{1:i}) d\mathbf{x}_i$$

E. Blayo - An introduction to data assimilation

#### Links with previous methods

- ▶ The Kalman filter corresponds to the general Bayesian sequential algorithm in the case where errors are Gaussian,  $\varepsilon^m$  is independent from the present state, H and M are linear, and  $X_0 \rightsquigarrow \mathcal{N}(\mathbf{x}^b, \mathbf{B})$ .
- Minimizing J in the variational approach is equivalent to looking for the mode of  $p(\mathbf{X}_0|\mathbf{Y}_{1:N} = \mathbf{y}_{1:N})$  if  $\varepsilon^m = 0$ ,  $\mathbf{X}_0 \rightsquigarrow \mathcal{N}(\mathbf{x}^b, \mathbf{B})$ , and  $\varepsilon^o_i \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$ .



E. Blayo - An introduction to data assimilation

#### Tens of implementations...

UNIVERSITÉ Grenoble

Alpes

- ▶ Two main groups: particle filters and Ensemble Kalman Filters
- Differ mainly by their analysis step Mainly: resampling methods and transformation methods



# To go further...







### Common main methodological difficulties

- ▶ Non linearities: J non quadratic / what about Kalman filter ?
- ► Huge dimensions [x] = O(10<sup>6</sup> 10<sup>9</sup>): minimization of J / management of huge matrices / approximation of covariance matrices / computation cost
- Poorly known error statistics: choice of the norms / B, R, Q
- ▶ HPC issues: data management, code efficiency, parallelization...



E. Blayo - An introduction to data assimilation

### In short

#### Variational methods:

- a series of approximations of the cost function, corresponding to a series of methods: 4DVar, incremental 4DVar, 3DFGAT, 3DVar
- the more sophisticated ones (4DVar, incremental 4DVar) require the tangent linear and adjoint models (the development of which is a real investment). En4DVar methods try to avoid it.
- Statistical methods:
  - extended Kalman filter handles (weakly) non linear problems (requires the tangent linear 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: fully Bayesian approach still limited to low dimension problems



#### Some present research directions

- improved methods: more robust w.r.t. nonlinearities and/or non gaussianity, or without adjoint, or less expensive...
- 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...



#### Two announcements

- Doctoral course "Introduction to data assimilation" Grenoble, January 8-12, 2018
- CNA 2018: 7ème Colloque National d'Assimilation de données Rennes, 26-28 septembre 2018



