## 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. .

## A daily example: numerical weather forecast




## 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
$\qquad$


## The best estimate

Several pieces of information:

- Model
- Prior (or background) value
$\longrightarrow$ Find the best possible estimate $\mathbf{x}^{a}$
- Observations
- Statistics
- ...


## The best estimate

Several pieces of information:

- Model
- Prior (or background) value
$\longrightarrow$ Find the best possible estimate $\mathbf{x}^{a}$
- Observations
- Statistics
- ...

What does the best possible estimate means?

- Estimate: deterministic value? pdf? some moments only of a pdf?
- Best: which criterion?


## 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
3. Some current challenges

## The simplest possible model problem

## Model problem: least squares approach

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

## Model problem: 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 $x^{b}=19^{\circ} \mathrm{C}$ and an observation $y=21^{\circ} \mathrm{C}$ of the (unknown) present temperature $x$.

- Let $J(x)=\frac{1}{2}\left[\left(x-x^{b}\right)^{2}+(x-y)^{2}\right]$
- $\operatorname{Min}_{x} J(x) \longrightarrow x^{a}=\frac{x^{b}+y}{2}=20^{\circ} \mathrm{C}$


## Model problem: 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 $x^{b}=19^{\circ} \mathrm{C}$ and an observation $y=21^{\circ} \mathrm{C}$ of the (unknown) present temperature $x$.

- Let $J(x)=\frac{1}{2}\left[\left(x-x^{b}\right)^{2}+(x-y)^{2}\right]$
- $\operatorname{Min}_{x} J(x) \longrightarrow x^{a}=\frac{x^{b}+y}{2}=20^{\circ} \mathrm{C}$

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

- Let $H(x)=\frac{9}{5} x+32 \quad$ observation operator

Let $J(x)=\frac{1}{2}\left[\left(H(x)-x^{b}\right)^{2}+(H(x)-y)^{2}\right]$
$\operatorname{Min}_{x} J(x) \longrightarrow x^{a}=20^{\circ} \mathrm{C}$

## Model problem: least squares approach

Drawback \# 1: if observation units are inhomogeneous

$$
\begin{aligned}
& x^{b}=19^{\circ} \mathrm{C} \text { and } y=69.8^{\circ} \mathrm{F} \\
& -J(x)=\frac{1}{2}\left[\left(x-x^{b}\right)^{2}+(H(x)-y)^{2}\right] \quad \longrightarrow x^{a}=20.53^{\circ} \mathrm{C}
\end{aligned}
$$

$$
\longrightarrow \text { adding apples and oranges !! }
$$

## Model problem: least squares approach

Drawback \# 1: if observation units are inhomogeneous

$$
\begin{aligned}
& x^{b}=19^{\circ} \mathrm{C} \text { and } y=69.8^{\circ} \mathrm{F} \\
& -J(x)=\frac{1}{2}\left[\left(x-x^{b}\right)^{2}+(H(x)-y)^{2}\right] \quad \longrightarrow x^{a}=20.53^{\circ} \mathrm{C}
\end{aligned}
$$

$\longrightarrow$ adding apples and oranges !!

Drawback \# 2: if observation accuracies are inhomogeneous
If $x^{b}$ is twice more accurate than $y$, one should obtain $x^{a}=\frac{2 x^{b}+y}{3}=19.67^{\circ} \mathrm{C}$
$\longrightarrow J$ should be $J(x)=\frac{1}{2}\left[\left(\frac{x-x^{b}}{1 / 2}\right)^{2}+\left(\frac{x-y}{1}\right)^{2}\right]$

## Model problem: linear statistical approach

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: $E\left(X^{a}\right)=x$
(it seems reasonable)
- of minimal variance: $\operatorname{Var}\left(X^{a}\right)$ minimum (optimal accuracy)
$\longrightarrow$ BLUE (Best Linear Unbiased Estimator)


## Model problem: linear statistical approach

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

## Hypotheses

- $E\left(\varepsilon^{b}\right)=E\left(\varepsilon^{0}\right)=0$ unbiased background and measurement device
- $\operatorname{Var}\left(\varepsilon^{b}\right)=\sigma_{b}^{2} \quad \operatorname{Var}\left(\varepsilon^{o}\right)=\sigma_{o}^{2}$
- $\operatorname{Cov}\left(\varepsilon^{b}, \varepsilon^{\circ}\right)=0$
known accuracies independent errors


## Model problem: linear statistical approach

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

## Hypotheses

- $E\left(\varepsilon^{b}\right)=E\left(\varepsilon^{o}\right)=0 \quad$ unbiased background and measurement device
- $\operatorname{Var}\left(\varepsilon^{b}\right)=\sigma_{b}^{2} \quad \operatorname{Var}\left(\varepsilon^{o}\right)=\sigma_{o}^{2}$
known accuracies
- $\operatorname{Cov}\left(\varepsilon^{b}, \varepsilon^{o}\right)=0$

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

- $E\left(X^{a}\right)=\left(\alpha_{b}+\alpha_{o}\right) x+\alpha_{b} \underbrace{E\left(\varepsilon^{b}\right)}_{=0}+\alpha_{o} \underbrace{E\left(\varepsilon^{o}\right)}_{=0} \Longrightarrow \alpha_{b}+\alpha_{o}=1$


## Model problem: linear statistical approach

$$
\text { Let } X^{b}=x+\varepsilon^{b} \text { and } Y=x+\varepsilon^{o} \quad \text { with }
$$

## Hypotheses

- $E\left(\varepsilon^{b}\right)=E\left(\varepsilon^{0}\right)=0$ unbiased background and measurement device
- $\operatorname{Var}\left(\varepsilon^{b}\right)=\sigma_{b}^{2} \quad \operatorname{Var}\left(\varepsilon^{o}\right)=\sigma_{o}^{2}$
known accuracies
- $\operatorname{Cov}\left(\varepsilon^{b}, \varepsilon^{\circ}\right)=0$

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

- $E\left(X^{a}\right)=\left(\alpha_{b}+\alpha_{o}\right) x+\alpha_{b} \underbrace{E\left(\varepsilon^{b}\right)}_{=0}+\alpha_{o} \underbrace{E\left(\varepsilon^{o}\right)}_{=0} \Longrightarrow \alpha_{b}+\alpha_{o}=1$
$-\operatorname{Var}\left(X^{a}\right)=E\left[\left(X^{a}-x\right)^{2}\right]=E\left[\left(\alpha_{b} \varepsilon^{b}+\alpha_{o} \varepsilon^{o}\right)^{2}\right]=\alpha_{b}^{2} \sigma_{b}^{2}+\left(1-\alpha_{b}\right)^{2} \sigma_{o}^{2}$

$$
\frac{\partial}{\partial \alpha_{b}}=0 \quad \Longrightarrow \quad \alpha_{b}=\frac{\sigma_{o}^{2}}{\sigma_{b}^{2}+\sigma_{o}^{2}}
$$

## Model problem: linear statistical approach

## BLUE

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

## Model problem: linear statistical approach

## BLUE

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

## Model problem: linear statistical approach

## BLUE

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

Its accuracy: $\quad\left[\operatorname{Var}\left(X^{a}\right)\right]^{-1}=\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}} \quad$ accuracies are added

## Model problem: linear statistical approach

## BLUE

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

Its accuracy: $\quad\left[\operatorname{Var}\left(X^{2}\right)\right]^{-1}=\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}} \quad$ accuracies are added

- Hypotheses on the two first moments of $\varepsilon^{b}, \varepsilon^{o}$ lead to results on the two first moments of $X^{a}$.


## Model problem: linear statistical approach

## Variational equivalence

This is equivalent to the problem:

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

## Model problem: linear statistical approach

## Variational equivalence

This is equivalent to the problem:

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

## 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 $J$
- $\underbrace{J^{\prime \prime}\left(x^{a}\right)}_{\text {convexity }}=\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}}=\underbrace{\left[\operatorname{Var}\left(x^{a}\right)\right]^{-1}}_{\text {accuracy }}$


## Model problem: linear statistical approach

Geometric interpretation $E\left(\varepsilon^{o} \varepsilon^{b}\right)=0 \Longrightarrow E\left(\varepsilon^{a}\left(Y-X_{b}\right)\right)=0$

$\rightarrow$ orthogonal projection for the scalar product $\left\langle Z_{1}, Z_{2}\right\rangle=E\left(Z_{1} Z_{2}\right)$ for unbiased random variables.

## Model problem: Bayesian approach

- $x$ : a realization of a random variable $X$. What is the pdf $p(X \mid Y)$ ?
- Based on the Bayes rule:

$$
P(X=x \mid Y=y)=\frac{\overbrace{P(Y=y \mid X=x)}^{\text {likekihood }} \overbrace{P(X=x)}^{\text {prior }}}{\underbrace{P(Y=y)}_{\text {normalisation factor }}}
$$

## Model problem: Bayesian approach

- $x$ : a realization of a random variable $X$. What is the pdf $p(X \mid Y)$ ?
- Based on the Bayes rule:

$$
P(X=x \mid Y=y)=\frac{\overbrace{P(Y=y \mid X=x)}^{\text {likekihood }} \overbrace{P(X=x)}^{\text {prior }}}{\underbrace{P(Y=y)}_{\text {normalisation factor }}}
$$

- Back to our example:
- Background $X^{b} \sim \mathcal{N}\left(19, \sigma_{b}^{2}\right)$
- Observation $y=21^{\circ} \mathrm{C}$, and $Y=X+\varepsilon^{\circ}$ with $\varepsilon^{\circ} \leadsto \mathcal{N}\left(0, \sigma_{o}^{2}\right)$


## Model problem: Bayesian approach

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

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

## Model problem: Bayesian approach

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

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

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

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

## Model problem: Bayesian approach

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

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

- Hence

$$
\begin{aligned}
p(X=x) p(Y=21 \mid X=x)= & \frac{1}{\sqrt{2 \pi} \sigma_{b}} \exp \left(-\frac{(x-19)^{2}}{2 \sigma_{b}^{2}}\right) \frac{1}{\sqrt{2 \pi} \sigma_{o}} \exp \left(-\frac{(21-x)^{2}}{2 \sigma_{o}^{2}}\right) \\
= & K \exp \left(-\frac{\left(x-m_{a}\right)^{2}}{2 \sigma_{a}^{2}}\right) \\
& \text { with } m_{a}=\frac{\frac{1}{\sigma_{b}^{2}} 19+\frac{1}{\sigma_{o}^{2}} 21}{\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}}} \text { and } \sigma_{a}^{2}=\left(\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}}\right)^{-1}
\end{aligned}
$$

## Model problem: Bayesian approach

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

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

- Hence

$$
\begin{aligned}
& p(X=x) p(Y=21 \mid X=x)= \frac{1}{\sqrt{2 \pi} \sigma_{b}} \exp \left(-\frac{(x-19)^{2}}{2 \sigma_{b}^{2}}\right) \frac{1}{\sqrt{2 \pi} \sigma_{o}} \exp \left(-\frac{(21-x)^{2}}{2 \sigma_{o}^{2}}\right) \\
&= K \exp \left(-\frac{\left(x-m_{a}\right)^{2}}{2 \sigma_{a}^{2}}\right) \\
& \text { with } m_{a}=\frac{\frac{1}{\sigma_{b}^{2}} 19+\frac{1}{\sigma_{2}^{2}} 21}{\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}}} \text { and } \sigma_{a}^{2}=\left(\frac{1}{\sigma_{b}^{2}}+\frac{1}{\sigma_{o}^{2}}\right)^{-1} \\
& \longrightarrow X \mid Y=21 \sim \mathcal{N}\left(m_{a}, \sigma_{a}^{2}\right)
\end{aligned}
$$

## Model problem: Bayesian approach



## Model problem: Bayesian approach



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

 AlpesGeneralization: arbitrary number of unknowns and observations

To be estimated: $\mathbf{x}=\left(\begin{array}{c}x_{1} \\ \vdots \\ x_{n}\end{array}\right) \in \mathbf{R}^{n}$
Observations: $\mathbf{y}=\left(\begin{array}{c}y_{1} \\ \vdots \\ y_{p}\end{array}\right) \in \mathbf{R}^{p}$
Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H: \mathbf{R}^{n} \longrightarrow \mathbf{R}^{p}$

Generalization: arbitrary number of unknowns and observations

## A simple example of observation operator

$$
\begin{aligned}
& \text { If } \mathbf{x}=\left(\begin{array}{c}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4}
\end{array}\right) \text { and } \mathbf{y}=\binom{\text { an observation of } \frac{x_{1}+x_{2}}{}}{\text { an observation of } x_{4}} \\
& \text { then } \quad H(\mathbf{x})=\mathbf{H x} \quad \text { with } \mathbf{H}=\left(\begin{array}{llll}
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
0 & 0 & 0 & 1
\end{array}\right)
\end{aligned}
$$

Generalization: arbitrary number of unknowns and observations
To be estimated: $\mathbf{x}=\left(\begin{array}{c}x_{1} \\ \vdots \\ x_{n}\end{array}\right) \in \mathbf{R}^{n}$
Observations: $\mathbf{y}=\left(\begin{array}{c}y_{1} \\ \vdots \\ y_{p}\end{array}\right) \in \mathbf{R}^{p}$
Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H: \mathbf{R}^{n} \longrightarrow \mathbf{R}^{p}$
Cost function: $J(\mathbf{x})=\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|^{2} \quad$ with $\|$.$\| to be chosen.$

Generalization: arbitrary number of unknowns and observations
To be estimated: $\mathbf{x}=\left(\begin{array}{c}x_{1} \\ \vdots \\ x_{n}\end{array}\right) \in \mathbf{R}^{n}$
Observations: $\mathbf{y}=\left(\begin{array}{c}y_{1} \\ \vdots \\ y_{p}\end{array}\right) \in \mathbf{R}^{p}$
Observation operator: $\mathbf{y} \equiv H(\mathbf{x})$, with $H: \mathbf{R}^{n} \longrightarrow \mathbf{R}^{p}$
Cost function: $J(\mathbf{x})=\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|^{2} \quad$ with $\|$.$\| to be chosen.$

## Remark

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

$$
p \geq n
$$

## Formalism "background value + new observations"

$$
\mathbf{z}=\binom{\mathbf{x}^{b}}{\mathbf{y}} \begin{aligned}
& \longleftarrow \text { background } \\
& \longleftarrow \text { new observations }
\end{aligned}
$$

The cost function becomes:

$$
J(\mathbf{x})=\underbrace{\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{b}\right\|_{b}^{2}}_{J_{b}}+\underbrace{\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|_{o}^{2}}_{J_{0}}
$$

## Formalism "background value + new observations"

$$
\mathbf{z}=\binom{\mathbf{x}^{b}}{\mathbf{y}} \begin{aligned}
& \longleftarrow \text { background } \\
& \longleftarrow \text { new observations }
\end{aligned}
$$

The cost function becomes:

$$
J(\mathbf{x})=\underbrace{\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{b}\right\|_{b}^{2}}_{J_{b}}+\underbrace{\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|_{o}^{2}}_{J_{0}}
$$

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

## If the problem is time dependent

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

$$
J_{o}(\mathbf{x})=\frac{1}{2} \sum_{i=0}^{N}\left\|H_{i}\left(\mathbf{x}\left(t_{i}\right)\right)-\mathbf{y}\left(t_{i}\right)\right\|_{o}^{2}
$$

## If the problem is time dependent

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

$$
J_{o}(\mathbf{x})=\frac{1}{2} \sum_{i=0}^{N}\left\|H_{i}\left(\mathbf{x}\left(t_{i}\right)\right)-\mathbf{y}\left(t_{i}\right)\right\|_{o}^{2}
$$

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

$$
J_{o}\left(\mathrm{x}_{0}\right)=\frac{1}{2} \sum_{i=0}^{N}\left\|H_{i}\left(\mathbf{x}\left(t_{i}\right)\right)-\mathbf{y}\left(t_{i}\right)\right\|_{o}^{2}=\frac{1}{2} \sum_{i=0}^{N}\left\|H_{i}\left(M_{0 \rightarrow t_{i}}\left(\mathbf{x}_{0}\right)\right)-\mathbf{y}\left(t_{i}\right)\right\|_{o}^{2}
$$

## If the problem is time dependent



## Uniqueness of the minimum ?

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

- If $H$ and $M$ are linear then $J_{0}$ is quadratic.


## Uniqueness of the minimum ?

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

- If $H$ and $M$ are linear then $J_{0}$ is quadratic.
- However $J_{0}$ generally does not have a unique minimum, since the number of observations is generally less than the size of $\mathbf{x}_{0}$ (the problem is underdetermined: $p<n$ ).

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

$$
J_{o}\left(x_{1}, x_{2}\right)=\frac{1}{2}\left(\frac{x_{1}+x_{2}}{2}-1.1\right)^{2}
$$



## Uniqueness of the minimum?

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

- If $H$ and $M$ are linear then $J_{0}$ is quadratic.
- However it generally does not have a unique minimum, since the number of observations is generally less than the size of $\mathbf{x}_{0}$ (the problem is underdetermined).
- Adding $J_{b}$ makes the problem of minimizing $J=J_{0}+J_{b}$ well posed.

$$
\begin{aligned}
& \text { Example: let }\left(x_{1}^{t}, x_{2}^{t}\right)=(1,1) \text { and } y=1.1 \text { an observa- } \\
& \text { tion of } \frac{1}{2}\left(x_{1}+x_{2}\right) \text {. Let }\left(x_{1}^{b}, x_{2}^{b}\right)=(0.9,1.05) \\
& J\left(x_{1}, x_{2}\right)= \\
& \quad \underbrace{\frac{1}{2}\left(\frac{x_{1}+x_{2}}{2}-1.1\right)^{2}}_{J_{0}}+\underbrace{\frac{1}{2}\left[\left(x_{1}-0.9\right)^{2}+\left(x_{2}-1.05\right)^{2}\right]}_{J_{b}} \\
& \longrightarrow\left(x_{1}^{a}, x_{2}^{a}\right)=(0.94166 \ldots, 1.09166 \ldots)
\end{aligned}
$$

## Uniqueness of the minimum ?

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

- If $H$ and/or $M$ are nonlinear then $J_{0}$ is no longer quadratic.

Example: the Lorenz system (1963)

$$
\left\{\begin{array}{l}
\frac{\mathrm{d} x}{\mathrm{~d} t}=\alpha(y-x) \\
\frac{\mathrm{d} y}{\mathrm{~d} t}=\beta x-y-x z \\
\frac{\mathrm{~d} z}{\mathrm{~d} t}=-\gamma z+x y
\end{array}\right.
$$



$$
J_{o}\left(y_{0}\right)=\frac{1}{2} \sum_{i=0}^{N}\left(x\left(t_{i}\right)-x_{\mathrm{obs}}\left(t_{i}\right)\right)^{2} d t
$$

## Uniqueness of the minimum ?

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

- If $H$ and/or $M$ are nonlinear then $J_{0}$ is no longer quadratic.






## Uniqueness of the minimum ?

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

- If $H$ and/or $M$ are nonlinear then $J_{0}$ 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.


## 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

$$
\begin{aligned}
J(\mathbf{x}) & =J_{b}(\mathbf{x})+J_{o}(\mathbf{x}) \\
& =\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{b}\right)+\frac{1}{2}(\mathbf{H} \mathbf{x}-\mathbf{y})^{T} \mathbf{R}^{-1}(\mathbf{H} \mathbf{x}-\mathbf{y})
\end{aligned}
$$

Optimal estimation in the linear case

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

## Minimizing J

$$
\begin{aligned}
J(\mathbf{x}) & =J_{b}(\mathbf{x})+J_{o}(\mathbf{x}) \\
& =\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{b}\right)+\frac{1}{2}(\mathbf{H} \mathbf{x}-\mathbf{y})^{T} \mathbf{R}^{-1}(\mathbf{H} \mathbf{x}-\mathbf{y})
\end{aligned}
$$

## Optimal estimation in the linear case

$$
\begin{aligned}
\mathbf{x}^{a}=\mathbf{x}^{b}+ & \underbrace{\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{T} \mathbf{R}^{-1}}_{\text {gain matrix }} \underbrace{\left(\mathbf{y}-\mathbf{H} \mathbf{x}^{b}\right)}_{\text {innovation vector }} \\
J\left(\mathbf{x}_{0}\right)= & \frac{1}{2}\left(\mathbf{x}_{0}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}_{0}-\mathbf{x}^{b}\right) \\
& +\frac{1}{2} \sum_{i=1}^{N}\left[\mathbf{y}_{i}-\mathbf{H}_{i} \mathbf{M}_{0, i} \mathbf{x}_{0}\right]^{T} \mathbf{R}_{i}^{-1}\left[\mathbf{y}_{i}-\mathbf{H}_{i} \mathbf{M}_{0, i} \mathbf{x}_{0}\right]
\end{aligned}
$$

With a linear evolution model

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

## Minimizing J

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

- even in the linear case (for which we have an explicit expression for $\mathbf{x}^{a}$ ), the computation of $\mathbf{x}^{a}$ 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.

$$
\mathbf{x}_{k+1}=\mathbf{x}_{k}+\alpha_{k} \mathbf{d}_{k}
$$


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

## Getting the gradient is not obvious

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

## Example:

$$
\left\{\begin{array}{l}
\frac{d \mathbf{x}(t))}{d t}=M(\mathbf{x}(t)) \quad t \in[0, T] \quad \text { with } \mathbf{u}=\left(\begin{array}{c}
u_{1} \\
\vdots \\
\mathbf{x}(t=0)=\mathbf{u} \\
u_{n}
\end{array}\right)
\end{array}\right.
$$

$$
\begin{aligned}
& J(\mathbf{u})=\frac{1}{2} \int_{0}^{T}\left\|\mathbf{x}(t)-\mathbf{x}^{\mathrm{obs}}(t)\right\|^{2} \longrightarrow \text { requires one model run } \\
& \nabla J(\mathbf{u})=\left(\begin{array}{c}
\frac{\partial J}{\partial u_{1}}(\mathbf{u}) \\
\vdots \\
\frac{\partial J}{\partial u_{n}}(\mathbf{u})
\end{array}\right) \simeq\left(\begin{array}{c}
{\left[J\left(\mathbf{u}+\alpha \mathbf{e}_{1}\right)-J(\mathbf{u})\right] / \alpha} \\
\vdots \\
{\left[J\left(\mathbf{u}+\alpha \mathbf{e}_{n}\right)-J(\mathbf{u})\right] / \alpha}
\end{array}\right) \\
& \longrightarrow n+1 \text { model runs }
\end{aligned}
$$

## Getting the gradient is not obvious

In actual large scale applications like meteorology / oceanography, $n=[\mathbf{u}]=\mathcal{O}\left(10^{6}-10^{9}\right) \quad \longrightarrow$ 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

$$
\left\{\begin{array}{l}
\left.\frac{\partial u}{\partial t}+u \frac{\partial u}{\partial x}-\nu \frac{\partial^{2} u}{\partial x^{2}}=f \quad x \in\right] 0, L[, t \in[0, T] \\
u(0, t)=\psi_{1}(t) u(L, t)=\psi_{2}(t) \quad t \in[0, T] \\
u(x, 0)=u_{0}(x) \quad x \in[0, L]
\end{array}\right.
$$

- $u^{\mathrm{obs}}(x, t)$ an observation of $u(x, t)$
- Cost function: $J\left(u_{0}\right)=\frac{1}{2} \int_{0}^{T} \int_{0}^{L}\left(u(x, t)-u^{\mathrm{obs}}(x, t)\right)^{2} d x d t$


## Example: an adjoint for the Burgers' equation

$$
\left\{\begin{array}{l}
\left.\frac{\partial u}{\partial t}+u \frac{\partial u}{\partial x}-\nu \frac{\partial^{2} u}{\partial x^{2}}=f \quad x \in\right] 0, L[, t \in[0, T] \\
u(0, t)=\psi_{1}(t) u(L, t)=\psi_{2}(t) \quad t \in[0, T] \\
u(x, 0)=u_{0}(x) \quad x \in[0, L]
\end{array}\right.
$$

- $u^{\mathrm{obs}}(x, t)$ an observation of $u(x, t)$
- Cost function: $J\left(u_{0}\right)=\frac{1}{2} \int_{0}^{T} \int_{0}^{L}\left(u(x, t)-u^{\mathrm{obs}}(x, t)\right)^{2} d x d t$


## Adjoint model

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

## Example: an adjoint for the Burgers' equation

$$
\left\{\begin{array}{l}
\left.\frac{\partial u}{\partial t}+u \frac{\partial u}{\partial x}-\nu \frac{\partial^{2} u}{\partial x^{2}}=f \quad x \in\right] 0, L[, t \in[0, T] \\
u(0, t)=\psi_{1}(t) u(L, t)=\psi_{2}(t) \quad t \in[0, T] \\
u(x, 0)=u_{0}(x) \quad x \in[0, L]
\end{array}\right.
$$

- $u^{\text {obs }}(x, t)$ an observation of $u(x, t)$
- Cost function: $J\left(u_{0}\right)=\frac{1}{2} \int_{0}^{T} \int_{0}^{L}\left(u(x, t)-u^{\mathrm{obs}}(x, t)\right)^{2} d x d t$


## Adjoint model

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

## Gradient of $J$

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

## Getting the gradient is not obvious

In actual large scale applications like meteorology / oceanography, $n=[\mathbf{u}]=\mathcal{O}\left(10^{6}-10^{9}\right) \quad \longrightarrow$ 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
- there exists automatic differentiation softwares:
$\longrightarrow$ cf http://www.autodiff.org


## Generalization: linear statistical approach

## Generalization: linear statistical approach

To be estimated: $\mathbf{x}=\left(\begin{array}{c}x_{1} \\ \vdots \\ x_{n}\end{array}\right) \in \mathbf{R}^{n} \quad$ Observations: $\mathbf{y}=\left(\begin{array}{c}y_{1} \\ \vdots \\ y_{p}\end{array}\right) \in \mathbf{R}^{p}$
Linear observation operator: $\mathbf{y} \equiv H(\mathbf{x})=\mathbf{H} \mathbf{x}$
Statistical framework:

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


## Generalization: linear statistical approach

To be estimated: $\mathbf{x}=\left(\begin{array}{c}x_{1} \\ \vdots \\ x_{n}\end{array}\right) \in \mathbf{R}^{n} \quad$ Observations: $\mathbf{y}=\left(\begin{array}{c}y_{1} \\ \vdots \\ y_{p}\end{array}\right) \in \mathbf{R}^{p}$
Linear observation operator: $\mathbf{y} \equiv H(\mathbf{x})=\mathbf{H} \mathbf{x}$
Statistical framework:

- $\mathbf{y}$ is a realization of a random vector $\mathbf{Y}$
- One is looking for the BLUE, i.e. a r.v. $\mathbf{X}^{a}$ that is
- linear: $\mathbf{X}^{a}=\mathbf{A Y}$ with $\operatorname{size}(\mathbf{A})=(n, p)$
- unbiased: $E\left(X^{a}\right)=\mathbf{x}$
- of minimal variance: $\operatorname{Var}\left(\mathbf{X}^{a}\right)=\sum_{i=1}^{n} \operatorname{Var}\left(X_{i}^{a}\right)$ minimum Gauss-Markov theorem: $\mathbf{A}=\left(\mathbf{H}^{\top} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{\top} \mathbf{R}^{-1}$


## Generalization: linear statistical approach

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

## Hypotheses:

- $H(\mathbf{x})=\mathbf{H x}$
linear observation operator
- $E\left(\varepsilon^{b}\right)=0$ and $E\left(\varepsilon^{o}\right)=0$ unbiased background and observations
- $\operatorname{Cov}\left(\varepsilon^{b}, \varepsilon^{o}\right)=0 \quad$ independent background and observation errors
- $\operatorname{Cov}\left(\varepsilon^{b}\right)=\mathbf{B}$ and $\operatorname{Cov}\left(\varepsilon^{o}\right)=\mathbf{R}$ known accuracies and covariances


## BLUE

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

## Link with the variational approach

## Statistical approach: BLUE

$$
\begin{aligned}
& \quad \mathbf{X}^{a}=\mathbf{X}^{b}+\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{T} \mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{H} \mathbf{X}^{b}\right) \\
& \text { with } \operatorname{Cov}\left(\mathbf{X}^{a}\right)=\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1}
\end{aligned}
$$

Variational approach in the linear case

$$
\begin{aligned}
J(\mathbf{x}) & =\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{b}\right\|_{b}^{2}+\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|_{o}^{2} \\
& =\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{b}\right)+\frac{1}{2}(\mathbf{H} \mathbf{x}-\mathbf{y})^{T} \mathbf{R}^{-1}(\mathbf{H} \mathbf{x}-\mathbf{y}) \\
\min _{\mathbf{x} \in \mathbf{R}^{n}} J(\mathbf{x}) & \longrightarrow \mathbf{x}^{a}=\mathbf{x}^{b}+\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{T} \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H} \mathbf{x}^{b}\right)
\end{aligned}
$$

## Link with the variational approach

## Statistical approach: BLUE

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

with $\operatorname{Cov}\left(\mathbf{X}^{a}\right)=\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1}$
Variational approach in the linear case

$$
\begin{aligned}
J(\mathbf{x}) & =\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^{b}\right\|_{b}^{2}+\frac{1}{2}\|H(\mathbf{x})-\mathbf{y}\|_{o}^{2} \\
& =\frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{b}\right)+\frac{1}{2}(\mathbf{H} \mathbf{x}-\mathbf{y})^{T} \mathbf{R}^{-1}(\mathbf{H} \mathbf{x}-\mathbf{y}) \\
\min _{\mathbf{x} \in \mathbf{R}^{n}} J(\mathbf{x}) & \longrightarrow \mathbf{x}^{a}=\mathbf{x}^{b}+\left(\mathbf{B}^{-1}+\mathbf{H}^{T} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{T} \mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H} \mathbf{x}^{b}\right)
\end{aligned}
$$

## Same remarks as previously

- The linear statistical approach rationalizes the choice of the norms for $J_{0}$ and $J_{b}$ in the variational approach.
- $\underbrace{\left[\operatorname{Cov}\left(\mathbf{X}^{a}\right)\right]^{-1}}_{\text {accuracy }}=\mathbf{B}^{-1}+\mathbf{H}^{\top} \mathbf{R}^{-1} \mathbf{H}=\underbrace{\operatorname{Hess}(J)}_{\text {convexity }}$


## If the problem is time dependent

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

## If the problem is time dependent

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

- Direct application of the BLUE on $\left[t_{0}, t_{N}\right]$ (hyp: $\varepsilon^{m}=0$ ):

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

$\longrightarrow$ 4D-Var algorithm

## If the problem is time dependent

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

- Direct application of the BLUE on $\left[t_{0}, t_{N}\right] \quad$ (hyp: $\varepsilon^{m}=0$ ):

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

- Sequential application of the BLUE every observation time:



## Kalman filter

## Hypotheses

- $\varepsilon^{m}\left(t_{i}\right)$ is unbiased, with covariance matrix $\mathbf{Q}_{i}$
- $\varepsilon^{m}\left(t_{i}\right)$ and $\varepsilon^{m}\left(t_{j}\right)$ are independent $(i \neq j)$
- Unbiased observation $\mathbf{y}_{i}$, with error covariance matrix $\mathbf{R}_{i}$
- $\varepsilon^{m}\left(t_{i}\right)$ and analysis error $\mathbf{x}^{a}\left(t_{i}\right)-\mathbf{x}^{t}\left(t_{i}\right)$ are independent

Evolution of the first two moments - Kalman filter
Initialization: $\quad \mathbf{x}^{\mathbf{a}}\left(t_{0}\right)=\mathbf{x}^{b}$

$$
\mathbf{P}^{\mathbf{a}}\left(t_{0}\right)=\mathbf{B}
$$

Step i: (prediction - correction, or forecast - analysis)

$$
\begin{array}{rlr}
\mathbf{x}^{f}\left(t_{i+1}\right) & =\mathbf{M}_{i, i+1} \mathbf{x}^{a}\left(t_{i}\right) \quad \text { Forecast } \\
\mathbf{P}^{f}\left(t_{i+1}\right) & =\mathbf{M}_{i, i+1} \mathbf{P}^{a}\left(t_{i}\right) \mathbf{M}_{i, i+1}^{T}+\mathbf{Q}_{i} \\
& & \\
\mathbf{x}^{a}\left(t_{i+1}\right) & =\mathbf{x}^{f}\left(t_{i+1}\right)+\mathbf{K}_{i+1}\left[\mathbf{y}_{i+1}-\mathbf{H}_{i+1} \mathbf{x}^{f}\left(t_{i+1}\right)\right] \\
\mathbf{K}_{i+1} & =\mathbf{P}^{f}\left(t_{i+1}\right) \mathbf{H}_{i+1}^{T}\left[\mathbf{H}_{i+1} \mathbf{P}^{f}\left(t_{i+1}\right) \mathbf{H}_{i+1}^{T}+\mathbf{R}_{i+1}\right]^{-1} \quad \text { BLUE } \\
\mathbf{P}^{a}\left(t_{i+1}\right) & =\mathbf{P}^{f}\left(t_{i+1}\right)-\mathbf{K}_{i+1} \mathbf{H}_{i+1} \mathbf{P}^{f}\left(t_{i+1}\right)
\end{array}
$$

## Kalman filter and 4D-Var



## Kalman filter and 4D-Var

If $\mathbf{H}_{i}$ and $\mathbf{M}_{i, i+1}$ are linear, and if the model is perfect $\left(\varepsilon^{m}\left(t_{i}\right)=0\right)$, then the Kalman filter and the variational method minimizing
$J\left(\mathbf{x}_{0}\right)=\frac{1}{2}\left(\mathbf{x}_{0}-\mathbf{x}^{b}\right)^{T} \mathbf{B}^{-1}\left(\mathbf{x}_{0}-\mathrm{x}^{b}\right)+\frac{1}{2} \sum_{i=0}^{N}\left(\mathbf{H}_{i} \mathbf{M}_{0, i} \mathbf{x}_{0}-\mathbf{y}_{i}\right)^{T} \mathbf{R}_{i}^{-1}\left(\mathbf{H}_{i} \mathbf{M}_{0, i} \mathbf{x}_{0}-\mathbf{y}_{i}\right)$
lead to the same solution at $t=t_{N}$.


## Generalization: Bayesian approach

## Generalization: Bayesian approach

Several types of problem:

- Filtering: $p\left(\mathbf{X}_{N} \mid \mathbf{Y}_{1: N}=\mathbf{y}_{1: N}\right)$
- Forecast: $p\left(\mathbf{X}_{I} \mid \mathbf{Y}_{1: N}=\mathbf{y}_{1: N}\right) \quad(I>N)$
- Smoothing: all other cases.
- $p\left(\mathbf{X}_{,} \mid \mathbf{Y}_{1: N}=\mathbf{y}_{1: N}\right) \quad(I<N):$ fixed-point smoothing
- $p\left(\mathbf{X}_{0: N} \mid \mathbf{Y}_{1: N}=\mathbf{y}_{1: N}\right)$ : fixed-interval smoothing
- ...


## Generalization: Bayesian approach

Two tools:

- Bayes theorem: $P(X=x \mid Y=y)=\frac{\overbrace{P(Y=y \mid X=x)}^{\text {likekihood }} \overbrace{P(X=x)}^{\text {prior }}}{\underbrace{P(Y=y)}_{\text {normalisation factor }}}$
- Marginalization rule: $p(\mathbf{X})=\int p(\mathbf{X} \mid \mathbf{Z}) p(\mathbf{Z}) d \mathbf{Z}$

And some usual hypotheses:

- $\varepsilon^{0}$ is independent from past and present states
- $\varepsilon^{m}$ is dependent at most of the present state, but not from past states


## Generalization: Bayesian approach

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

- Analysis at time $t_{i}$ :

$$
p\left(\mathbf{X}_{i}=\mathbf{x}_{i} \mid \mathbf{Y}_{1: i}=\mathbf{y}_{1: i}\right) \propto p\left(\mathbf{X}_{i}=\mathbf{x}_{i} \mid \mathbf{Y}_{1: i-1}=\mathbf{y}_{1: i-1}\right) p\left(\mathbf{Y}_{i}=\mathbf{y}_{i} \mid \mathbf{X}_{i}=\mathbf{x}_{i}\right)
$$

- Forecast from $t_{i}$ to $t_{i+1}$ :

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

## Generalization: Bayesian approach

## 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 $\mathbf{X}_{0} \leadsto \mathcal{N}\left(\mathbf{x}^{b}, \mathbf{B}\right)$.
- Minimizing $J$ in the variational approach is equivalent to looking for the mode of $p\left(\mathbf{X}_{0} \mid \mathbf{Y}_{1: N}=\mathbf{y}_{1: N}\right)$ if $\varepsilon^{m}=0, \mathbf{X}_{0} \leadsto \mathcal{N}\left(\mathbf{x}^{b}, \mathbf{B}\right)$, and $\varepsilon_{i}^{o} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{i}\right)$.


## Tens of implementations...

- 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 $[\mathrm{x}]=\mathcal{O}\left(10^{6}-10^{9}\right)$ : minimization of $\mathrm{J} /$ 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...


## 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

