Sequential methods for ocean data assimilation

From theory to practical implementations (I)

P. BRASSEUR

CNRS/LEGI, Grenoble, France

Pierre.Brasseur@hmg.inpg.fr

J. Ballabrera, L. Berline, F. Birol, J.M. Brankart, G. Broquet, V. Carmillet, F. Castruccio, F. Debost, D. Rozier, Y. Ourmières, T.Penduff, J. Verron

E. Blayo, S. Carme, Pham D.T., C. Robert

C.E. Testut, B. Tranchant, N. Ferry, E. Remy, M. Benkiran

F. Durand, L. Gourdeau, Ch. Maes, P. De Mey

A. Barth, C. Raick, M. Grégoire

L. Parent

Ocean data assimilation

❑ Data assimilation involves the optimal combination of measurements with the underlying dynamical principles governing the system under observation.

❑ Data assimilation can serve several oceanographic objectives:

➢ Ocean state estimation in space & time (4D) ;

➢ Detection of model errors ;

➢ Estimation of budgets & model parameters ;

➢ Initialisation, prediction, monitoring ;

➢ Optimal design of complex observation systems ;

➢ …

❑ Theories: optimal *control* (VAR) and optimal *estimation* (Kalman)

MERCATOR Assimilation Systems

- **Incremental implementation strategy**

	OI	Kalman filters	3D/4D-VAR
Research	1993 (SOFA)	1998 (SEEK)	1999 (OPAVAR)
R&D	1997	2002	2004
DEV	1999	2005	2008 ?
OP	2001	2007 ?	?
	SAM-1	**SAM-2**	**SAM-3**

LEGI
LABORATOIRE DES ECOULEMENTS GEOPHYSIQUES ET INDUSTRIELS

OUTLINE

❑ **State-of-the-art**

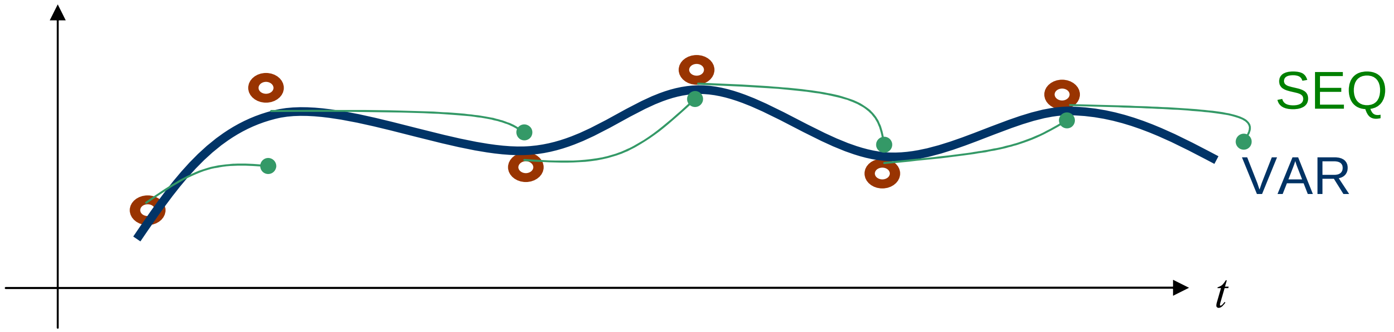
1. **Introduction**
2. **Kalman filter: fundamentals**
3. ***Applied* ocean data assimilation: specific issues**
4. **Simplifications of the KF – Optimal Interpolation**

❑ **Advanced issues**

5. **Space reduction: state and error sub-spaces**
6. **Low rank filters: SEEK and EnKF**
7. **Objective validation and adaptive schemes**
8. **Improved temporal strategies : FGAT and IAU**

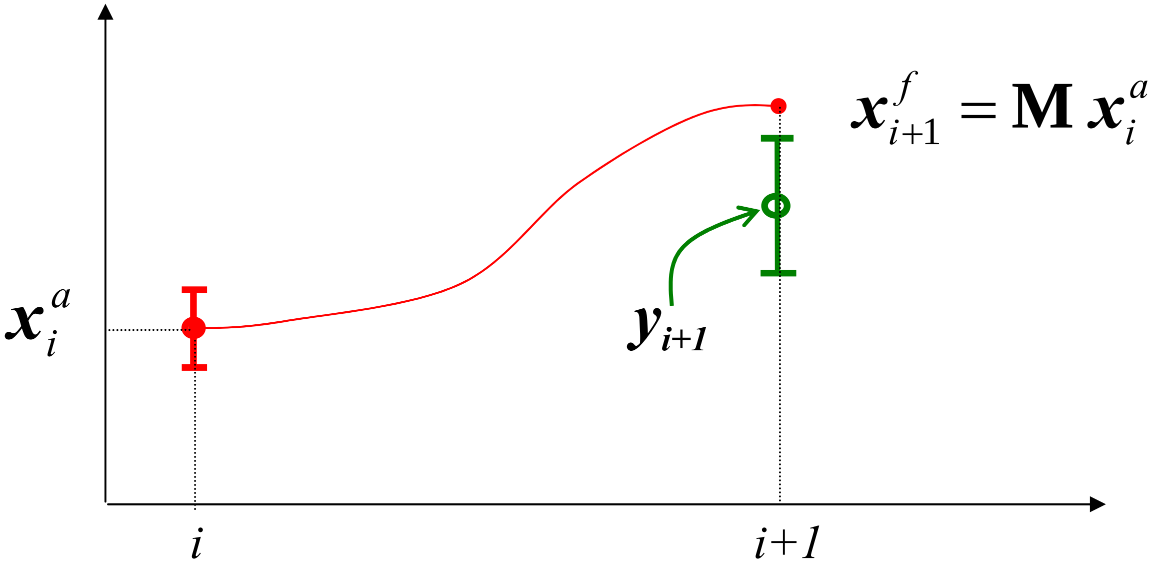
« Sequential » data assimilation ?

- ❑ **Methods/algorithms**
  - ➢ **« Variational » vs. « Sequential » (Talagrand 1997, Ide et al., 1997)**
  - ➢ **« Smoothers » vs. « Filters »**
  - ➢ **« Global problem » vs. « sub-problems » (Schröter 2004)**



2. Kalman Filter fundamentals
Problem definition

Notations: Ide *et al.* (1997)



$\boldsymbol{x}_{i+1}^{f} = \mathbf{M}\,\boldsymbol{x}_{i}^{a}$
$\boldsymbol{x}_{i}^{a}$
$\boldsymbol{y}_{i+1}$
i
$i+1$

$\boldsymbol{x}_{i}^{a}$: estimation of the « true » state vector $\boldsymbol{x}_{i}^{t}$ at time i , dimension n

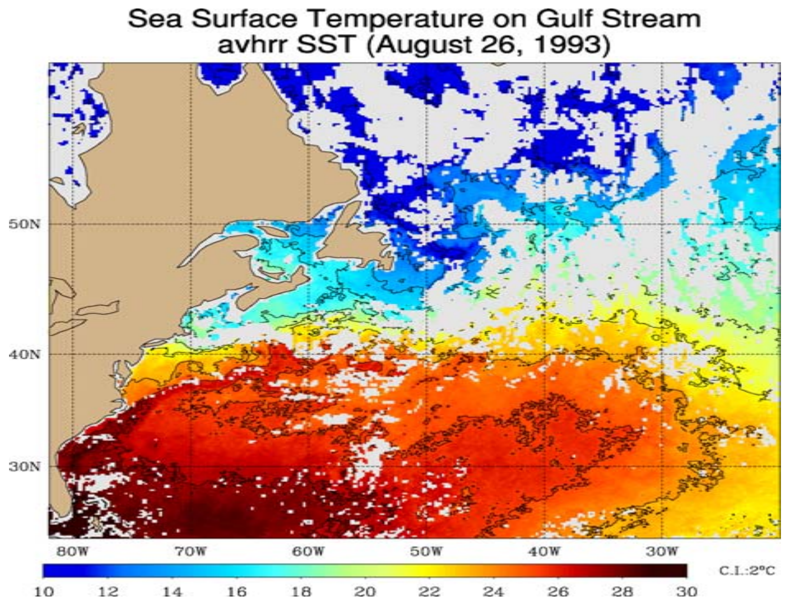
$\boldsymbol{x}_{i+1}^{f}$: forecast of the state vector at time $i+1$, using the linear model $\mathbf{M}$

$\boldsymbol{y}_{i+1}$: observations available at time $i+1$, dimension p

How can the true state be best estimated from this prior information ?

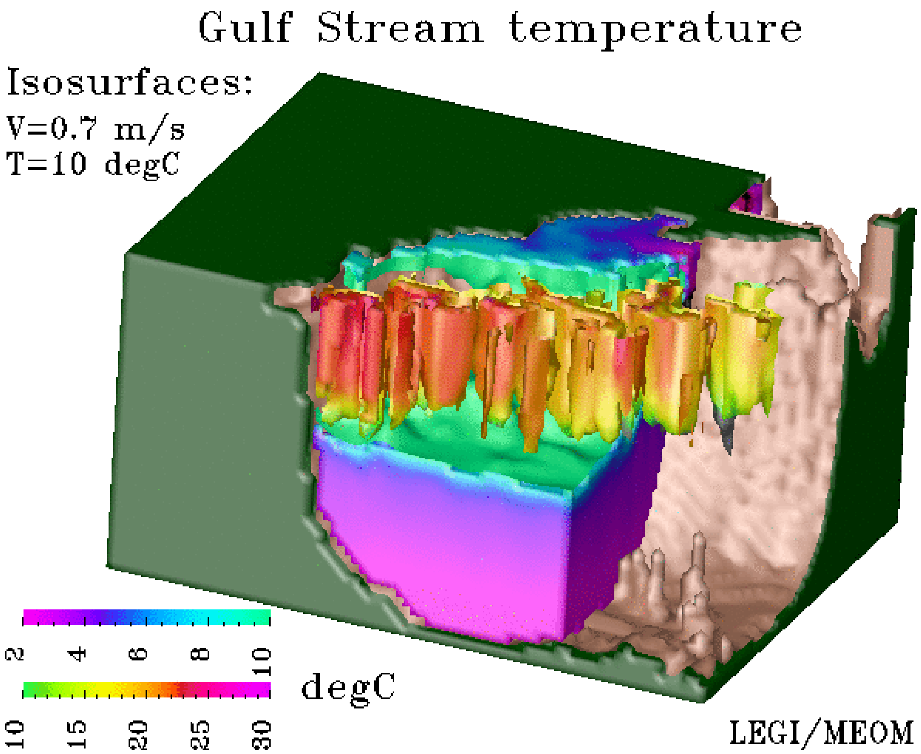
2. Kalman Filter fundamentals

Multivariate estimation


Sea Surface Temperature on Gulf Stream
avhrr SST (August 26, 1993)

$$y_{i+1}$$

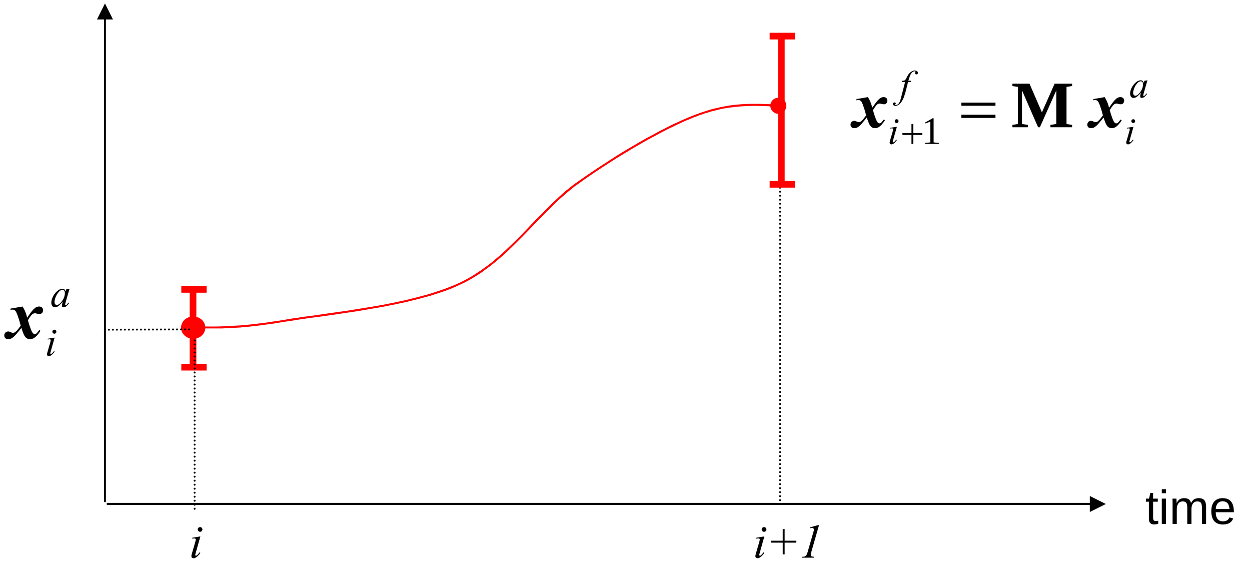
incomplete data


Gulf Stream temperature

$$x_{i+1}^{f} = \mathbf{M}\, x_{i}^{a}$$

imperfect model

2. Kalman Filter fundamentals
Uncertainties & PDFs



$$\boldsymbol{x}_{i+1}^{f} = \mathbf{M}\,\boldsymbol{x}_i^a$$

$\boldsymbol{\varepsilon}_i^a = \boldsymbol{x}_i^a - \boldsymbol{x}_i^t$: error on state estimate at time i ; unknown quantity, but assume

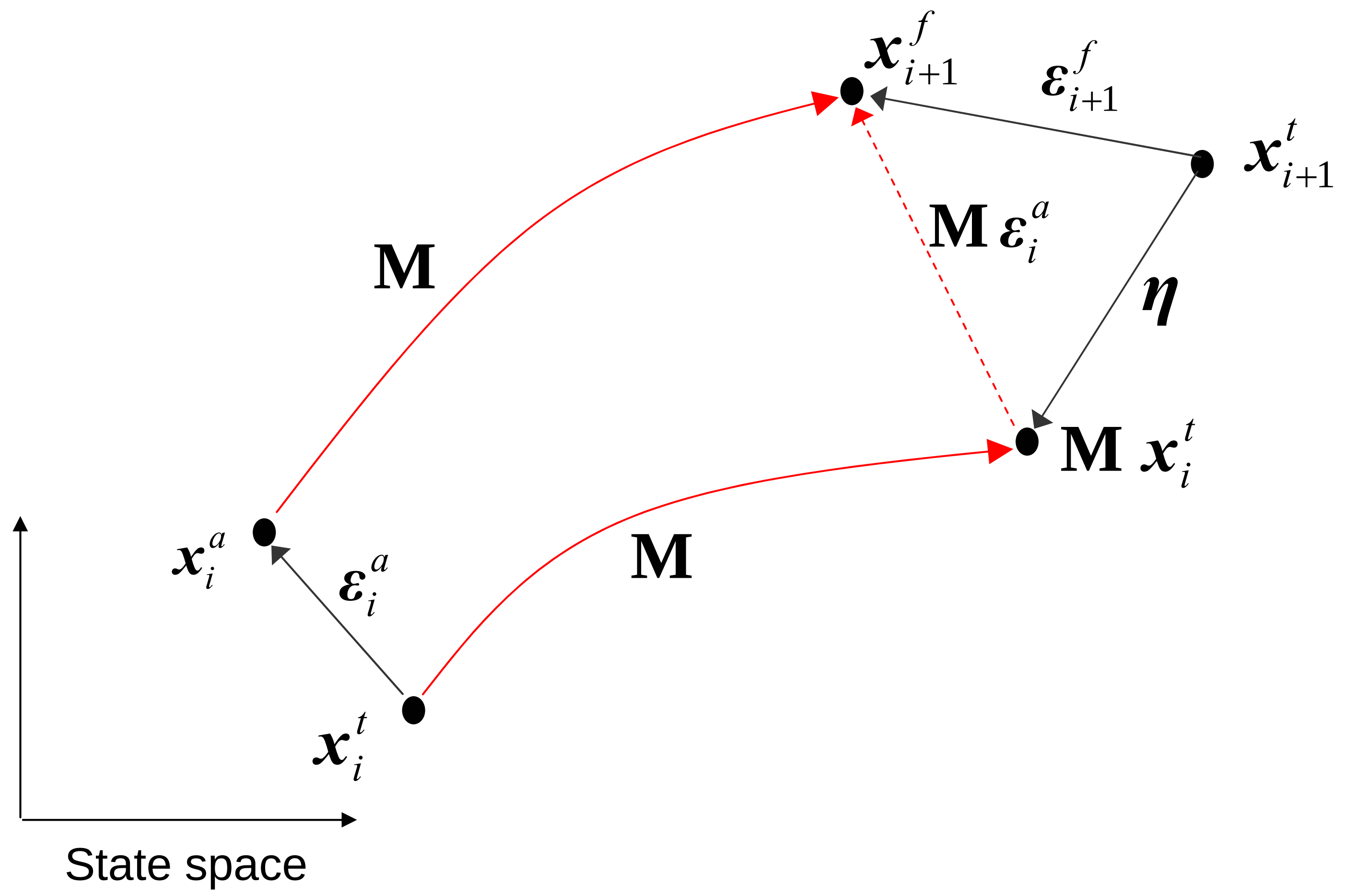
$$\boldsymbol{\varepsilon}_i^a \to N(0,\mathbf{P}_i^a) \sim \exp\left[-\frac{1}{2}\boldsymbol{\varepsilon}_i^{a\,T}\,\mathbf{P}_i^{a\,-1}\,\boldsymbol{\varepsilon}_i^a\right] \qquad (1)$$

$\boldsymbol{\eta} = \mathbf{M}\,\boldsymbol{x}_i^t - \boldsymbol{x}_{i+1}^t$: error on model forecast between time i and time $i+1$; assume:

$$\boldsymbol{\eta} \to N(0,\mathbf{Q}) \sim \exp\left[-\frac{1}{2}\boldsymbol{\eta}^{T}\,\mathbf{Q}^{-1}\,\boldsymbol{\eta}\right] \qquad (2)$$

2. Kalman Filter fundamentals

Error diagram



2. Kalman Filter fundamentals
Forecast error

$$\mathbf{M}\,\boldsymbol{\varepsilon}_i^a = \mathbf{M}\,\boldsymbol{x}_i^a - \mathbf{M}\,\boldsymbol{x}_i^t = \boldsymbol{x}_{i+1}^f - (\boldsymbol{x}_{i+1}^t + \boldsymbol{\eta}) = \boldsymbol{\varepsilon}_{i+1}^f - \boldsymbol{\eta}$$

Assuming pdf (1) and (2), model linearity and uncorrelated initial and modelling errors, the forecast error is distributed as :

$$\boldsymbol{\varepsilon}_{i+1}^f \rightarrow N(0, \mathbf{P}_{i+1}^f) \sim \exp\left[-\frac{1}{2}\boldsymbol{\varepsilon}_{i+1}^{f\;T}\,\mathbf{P}_{i+1}^{f\;-1}\,\boldsymbol{\varepsilon}_{i+1}^f\right] \qquad (3)$$

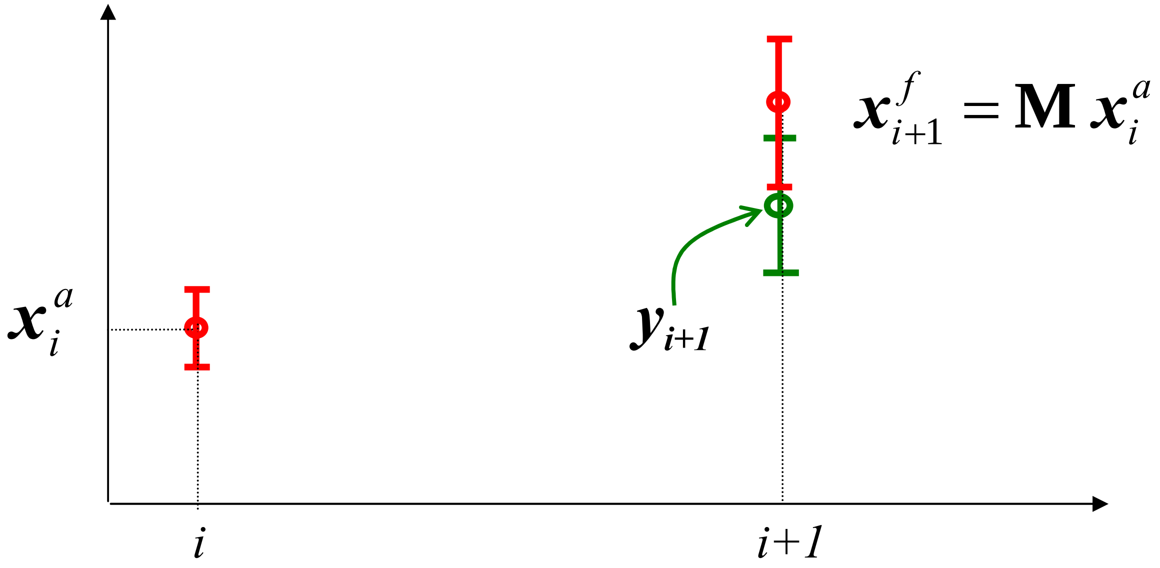
with

$$\begin{aligned}\boldsymbol{\varepsilon}_{i+1}^f = \mathbf{M}\,\boldsymbol{\varepsilon}_i^a + \boldsymbol{\eta} \quad \Rightarrow \quad & \overline{\boldsymbol{\varepsilon}_{i+1}^f \boldsymbol{\varepsilon}_{i+1}^{f\;T}} = \mathbf{M}\,\overline{\boldsymbol{\varepsilon}_i^a \boldsymbol{\varepsilon}_i^{a\;T}}\,\mathbf{M}^T + \overline{\boldsymbol{\eta}\boldsymbol{\eta}^T} \\ \Rightarrow \quad & \mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q}\end{aligned} \qquad (4)$$

The estimation error is amplified by:

- **unstable model dynamics (M) ;**
- **modelling errors Q .**

2. Kalman Filter fundamentals
Observations and errors



$\boldsymbol{x}_{i+1}^{f} = \mathbf{M}\,\boldsymbol{x}_{i}^{a}$
$\boldsymbol{x}_{i}^{a}$
$\boldsymbol{y}_{i+1}$
i
$i+1$

$$\boldsymbol{y}_{i+1} = \mathbf{H}\,\boldsymbol{x}_{i+1}^{t} + \boldsymbol{\varepsilon}_{i+1}^{o}$$: observation available at time $i+1$

A probability distribution for the observation error is assumed:

$$\boldsymbol{\varepsilon}_{i+1}^{o} \rightarrow N(0,\mathbf{R}) \sim \exp\left[-\frac{1}{2}\boldsymbol{\varepsilon}_{i+1}^{o\;T}\,\mathbf{R}^{-1}\,\boldsymbol{\varepsilon}_{i+1}^{o}\right] \qquad (5)$$

2. Kalman Filter fundamentals
Optimal estimation

Using Bayes rule at time $i+1$:

$$P\left(\boldsymbol{x}^t_{i+1}\middle|\boldsymbol{y}_{i+1}\right)=\frac{\overbrace{P\left(\boldsymbol{y}_{i+1}\middle|\boldsymbol{x}^t_{i+1}\right)}^{\text{given by (5)}}\cdot\overbrace{P\left(\boldsymbol{x}^t_{i+1}\right)}^{\text{given by (3)}}}{\underbrace{P(\boldsymbol{y}_{i+1})}_{\text{a scaling factor}}} \qquad (6)$$

$$P\left(\boldsymbol{x}^t_{i+1}\right)\cdot P\left(\boldsymbol{y}_{i+1}\middle|\boldsymbol{x}^t_{i+1}\right) \sim$$

$$\exp\left[-\frac{1}{2}(\boldsymbol{x}^f_{i+1}-\boldsymbol{x}^t_{i+1})^T\,\mathbf{P}^{f\;-1}_{i+1}(\boldsymbol{x}^f_{i+1}-\boldsymbol{x}^t_{i+1})\right]\cdot\exp\left[-\frac{1}{2}(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x}^t_{i+1})^T\,\mathbf{R}^{-1}(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x}^t_{i+1})\right]$$

$$=\exp\left[-\frac{1}{2}\left\{(\boldsymbol{x}^f_{i+1}-\boldsymbol{x}^t_{i+1})^T\,\mathbf{P}^{f\;-1}_{i+1}(\boldsymbol{x}^f_{i+1}-\boldsymbol{x}^t_{i+1})+(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x}^t_{i+1})^T\,\mathbf{R}^{-1}(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x}^t_{i+1})\right\}\right] \qquad (7)$$

The best estimate of $\boldsymbol{x}^t_{i+1}$ is the value of $\boldsymbol{x}$ which maximize (7), i.e. the minimum of :

$$J(\boldsymbol{x})=(\boldsymbol{x}^f_{i+1}-\boldsymbol{x})^T\,\mathbf{P}^{f\;-1}_{i+1}(\boldsymbol{x}^f_{i+1}-\boldsymbol{x})+(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x})^T\,\mathbf{R}^{-1}(\boldsymbol{y}_{i+1}-\mathbf{H}\,\boldsymbol{x}) \qquad (8)$$

2. Kalman Filter fundamentals
Kalman gain

$$\delta_{\boldsymbol{x}} J(\boldsymbol{x}) = 0 \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{x}_{i+1}^{f} + \mathbf{P}_{i+1}^{f} \mathbf{H}^{T} \mathbf{R}^{-1} (\boldsymbol{y}_{i+1} - \mathbf{H}\,\boldsymbol{x}) \quad (9)$$

Equation (9) can be solved for $\boldsymbol{x}$ using simple algebra (*), leading to:

$$\boldsymbol{x} = \boldsymbol{x}_{i+1}^{f} + \underbrace{\mathbf{P}_{i+1}^{f} \mathbf{H}^{T} (\mathbf{H}\mathbf{P}_{i+1}^{f} \mathbf{H}^{T} + \mathbf{R})^{-1}}_{\text{Kalman gain} = \mathbf{K}_{i+1}} (\boldsymbol{y}_{i+1} - \mathbf{H}\,\boldsymbol{x}_{i+1}^{f})$$

Note: the forecast and analysis equations can be extended to *weakly* non-linear models M and observation operator H.

(*) Hint : use matrix equality $\left[\mathbf{X}_1 + \mathbf{X}_{12}\, \mathbf{X}_2^{-1}\, \mathbf{X}_{21}\right]^{-1} = \mathbf{X}_1^{-1} - \mathbf{X}_1^{-1} \mathbf{X}_{12} \left[\mathbf{X}_2 + \mathbf{X}_{21} \mathbf{X}_1^{-1} \mathbf{X}_{12}\right]^{-1} \mathbf{X}_{21} \mathbf{X}_1^{-1}$

2. Kalman Filter fundamentals
Assimilation cycle (« sub-problem »)

$$\begin{cases} \boldsymbol{x}_i^a \\ \mathbf{P}_i^a \end{cases}$$

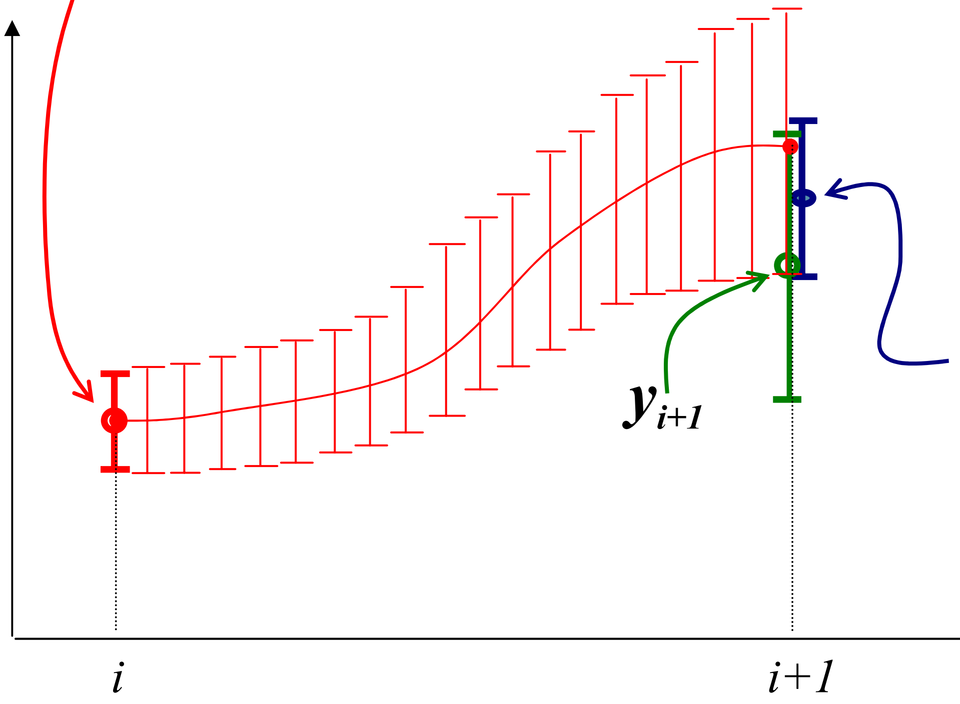
Forecast

$$\begin{cases} \boldsymbol{x}_{i+1}^f = M\left(\boldsymbol{x}_i^a\right) \\ \mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q} \end{cases}$$

Analysis

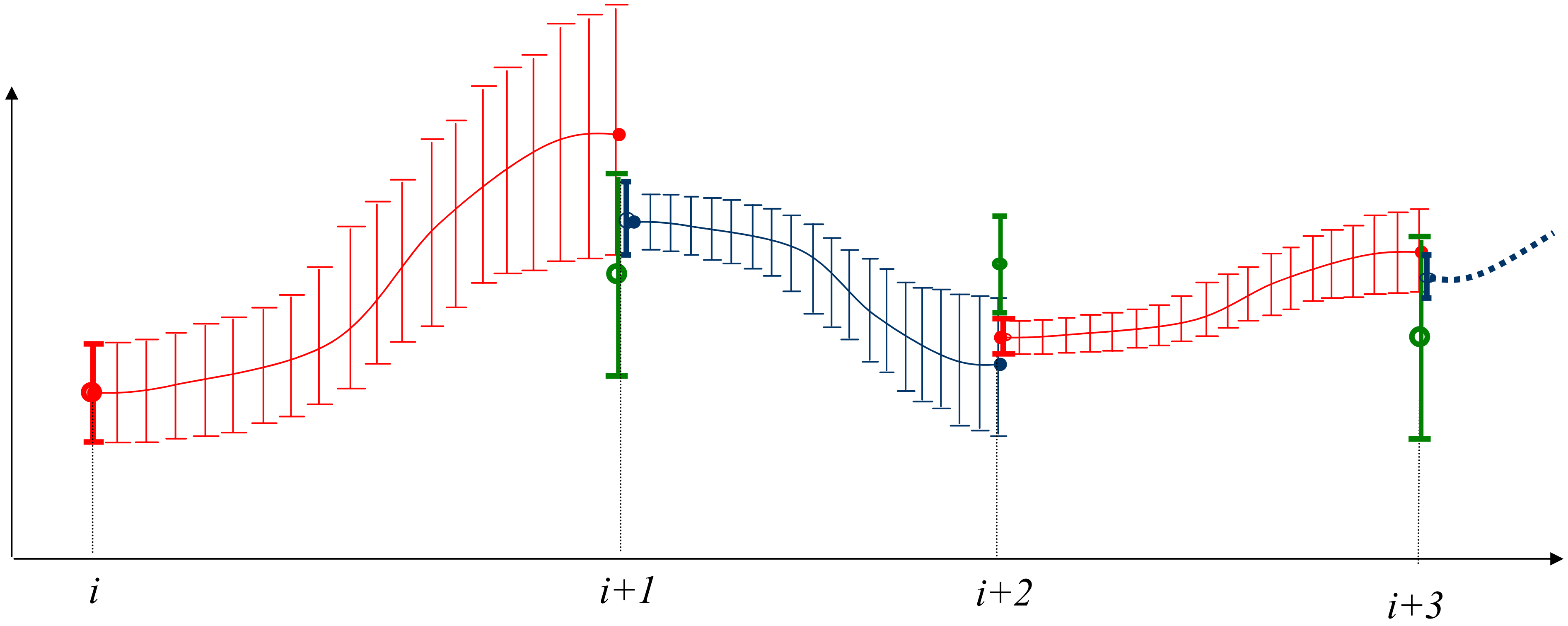
$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}_{i+1}^f\mathbf{H}^T + \mathbf{R})^{-1}$$

$$\begin{cases} \boldsymbol{x}_{i+1}^a = \boldsymbol{x}_{i+1}^f + \mathbf{K}_{i+1}\left(\boldsymbol{y}_{i+1} - H\left(\boldsymbol{x}_{i+1}^f\right)\right) \\ \mathbf{P}_{i+1}^a = \left(\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}\right)\mathbf{P}_{i+1}^f \end{cases}$$



2. Kalman Filter fundamentals
Assimilation sequence

Sequential assimilation = repeated forecast/analysis cycles



The best estimate at a given time is influenced by all previous observations (Kalman « filter »), and the analysis error covariance reflects the competition between this accumulation of past information and the error growth due to model imperfections .

3. Applied ocean data assimilation : *the art of making successful simplifications*

- ❑ **« The scientific difficulty of data assimilation is to find algorithms which simplify the BLUE (Best Linear Unbiased Estimation) to an affordable amount of computer resources, while preserving some of the essential characteristics. »**

Courtier

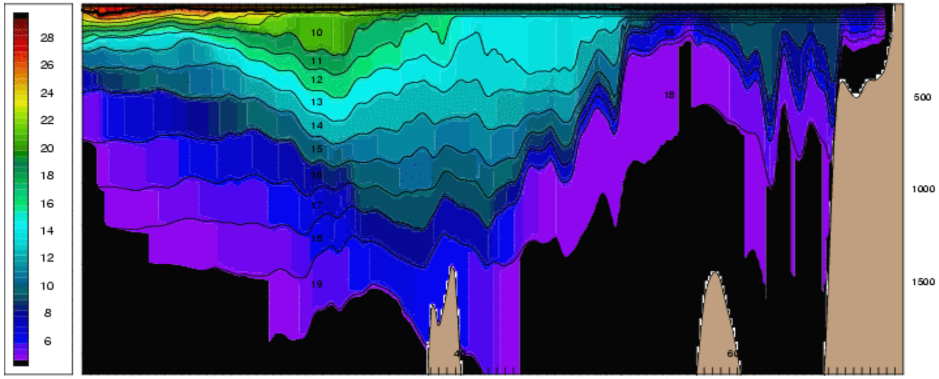
J. Meteor. Soc. Japan, 1997

3. Applied ocean data assimilation
Model complexity

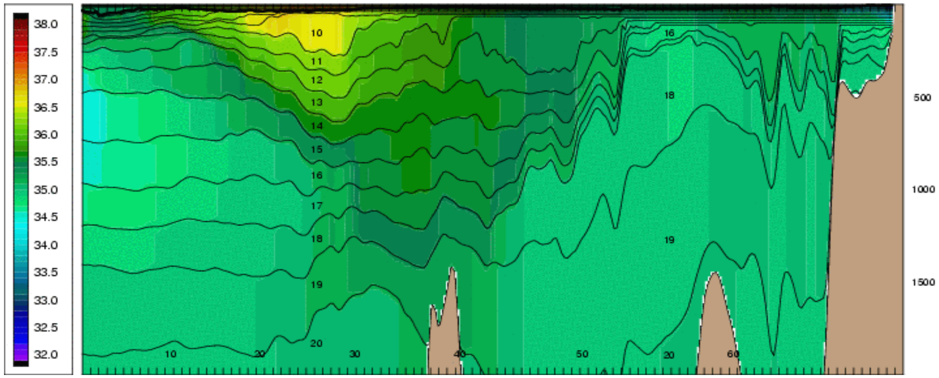
- **Model variables in HYCOM(*)**

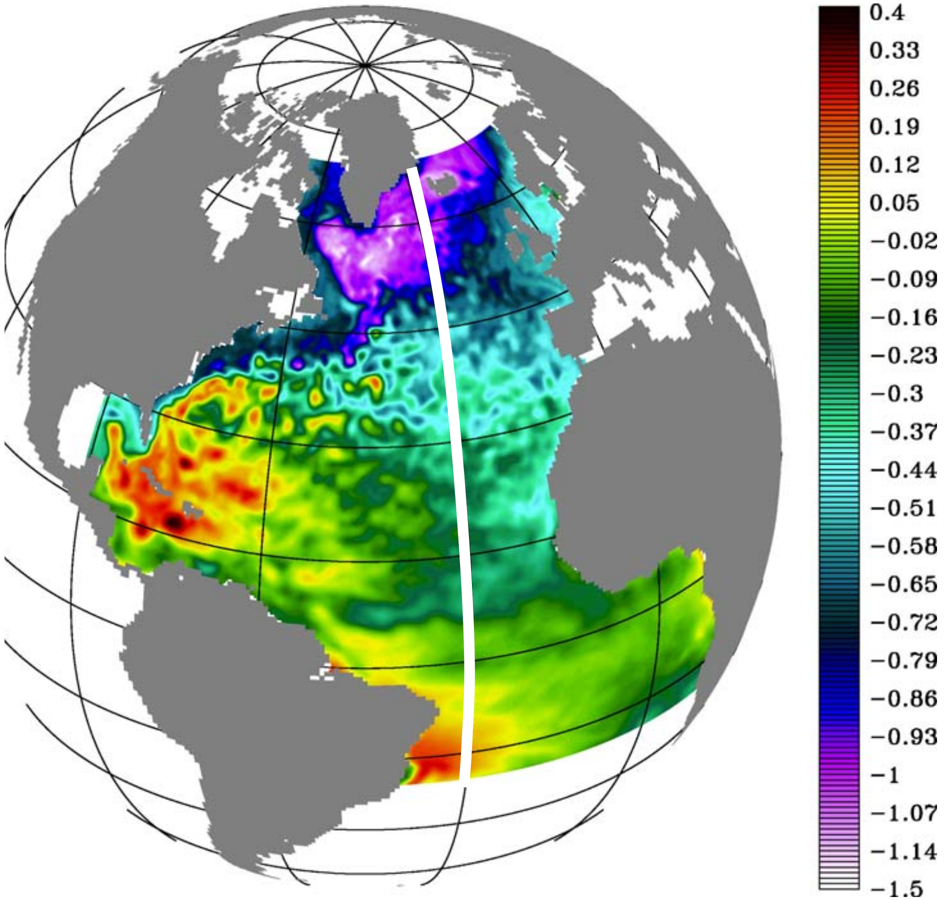
temperature, salinity, velocity, layer thicknesses, sea-surface height (SSH)

temperature merid.sec. 30.12w year 9.59 (aug 19) [08.1H]



salinity merid.sec. 30.12w year 9.59 (aug 19) [08.1H]





(*) ocean circulation model developed at Univ. Miami (RSMAS, E. Chassignet)

3. Applied ocean data assimilation
State vector dimension

- **HYCOM state vector x** : 3D grid of the 5 scalar model variables
  + 2D grid for SSH

- **R&D prototype**: 1/3° horizontal resolution , 19 hybrid layers

  $n \sim$ 5 x 350 x 350 x 19 ~ 1.1 x 10^7

  Operational prototype: 1/12° horizontal resolution , 26 hybrid layers

  $n \sim$ 5 x 1400 x 1400 x 26 ~ 2.5 x 10^8

- $\mathbf{M}$ **operator** : dim n x $n \sim$ 6 x 10^{16} real (i.e. ~ 6000 Earth Simulators !)

➢ **The state vector dimension can be huge**

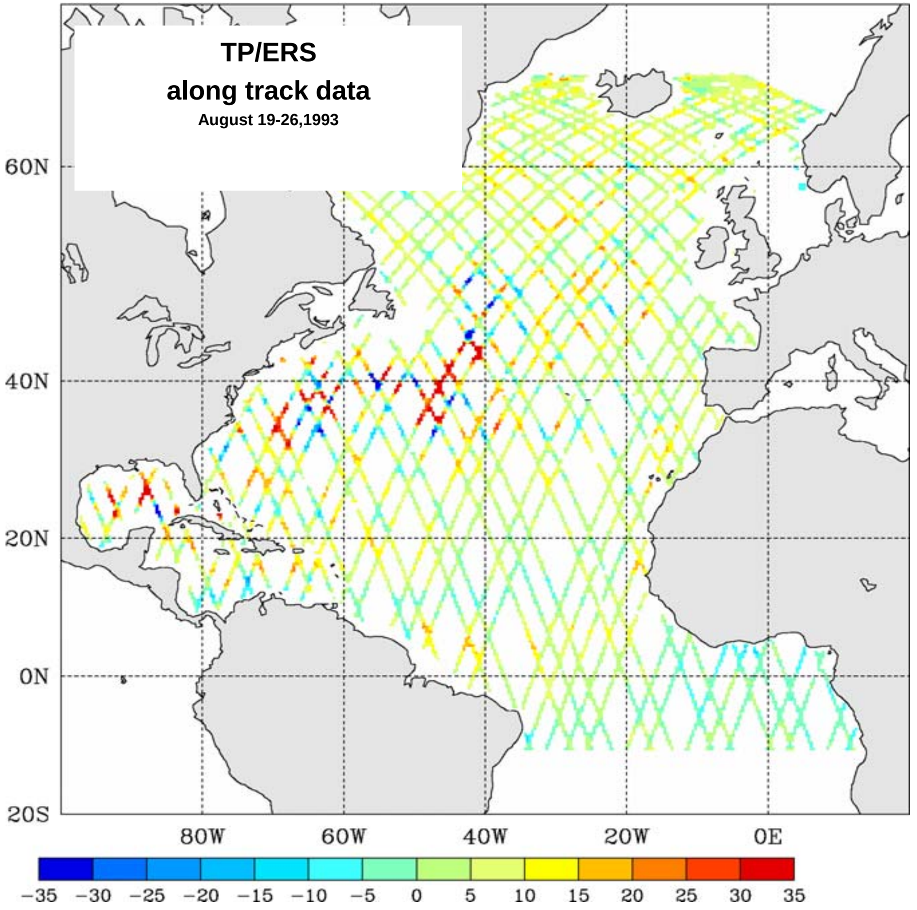
➢ **The model transition matrix « M » cannot be represented explicitely.**

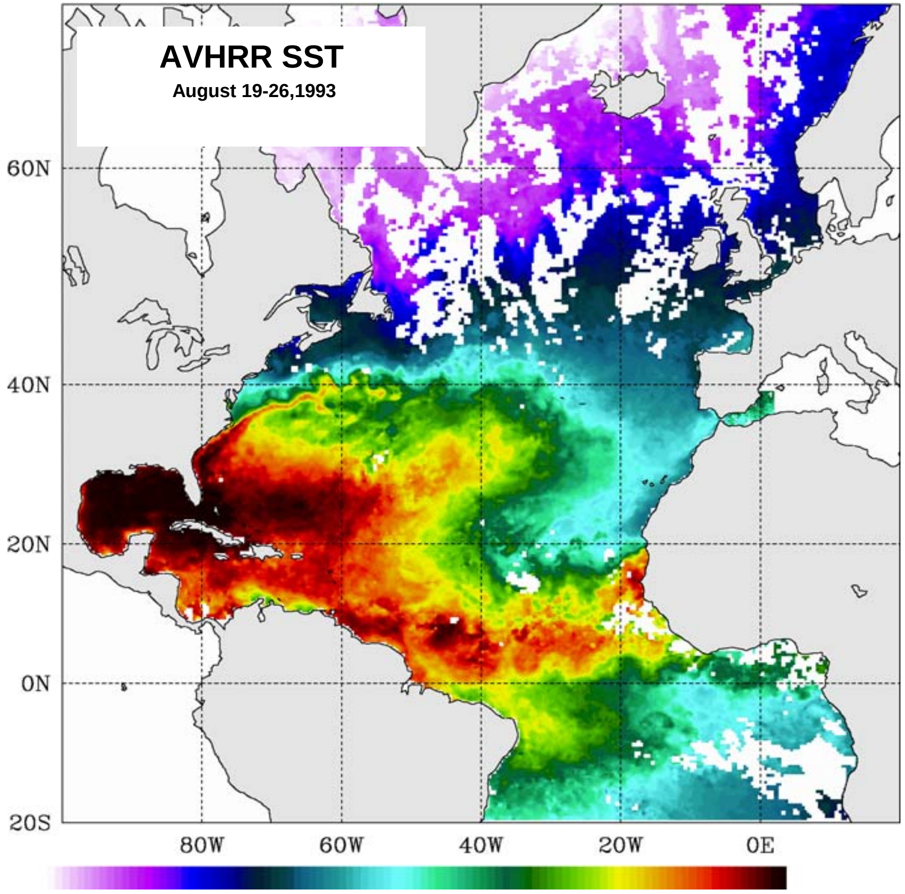
➢ **Instead, a computer code is used to transition « x » from time i to $i+1$**

3. Applied ocean data assimilation *Space observations*

- **Observed variables**:

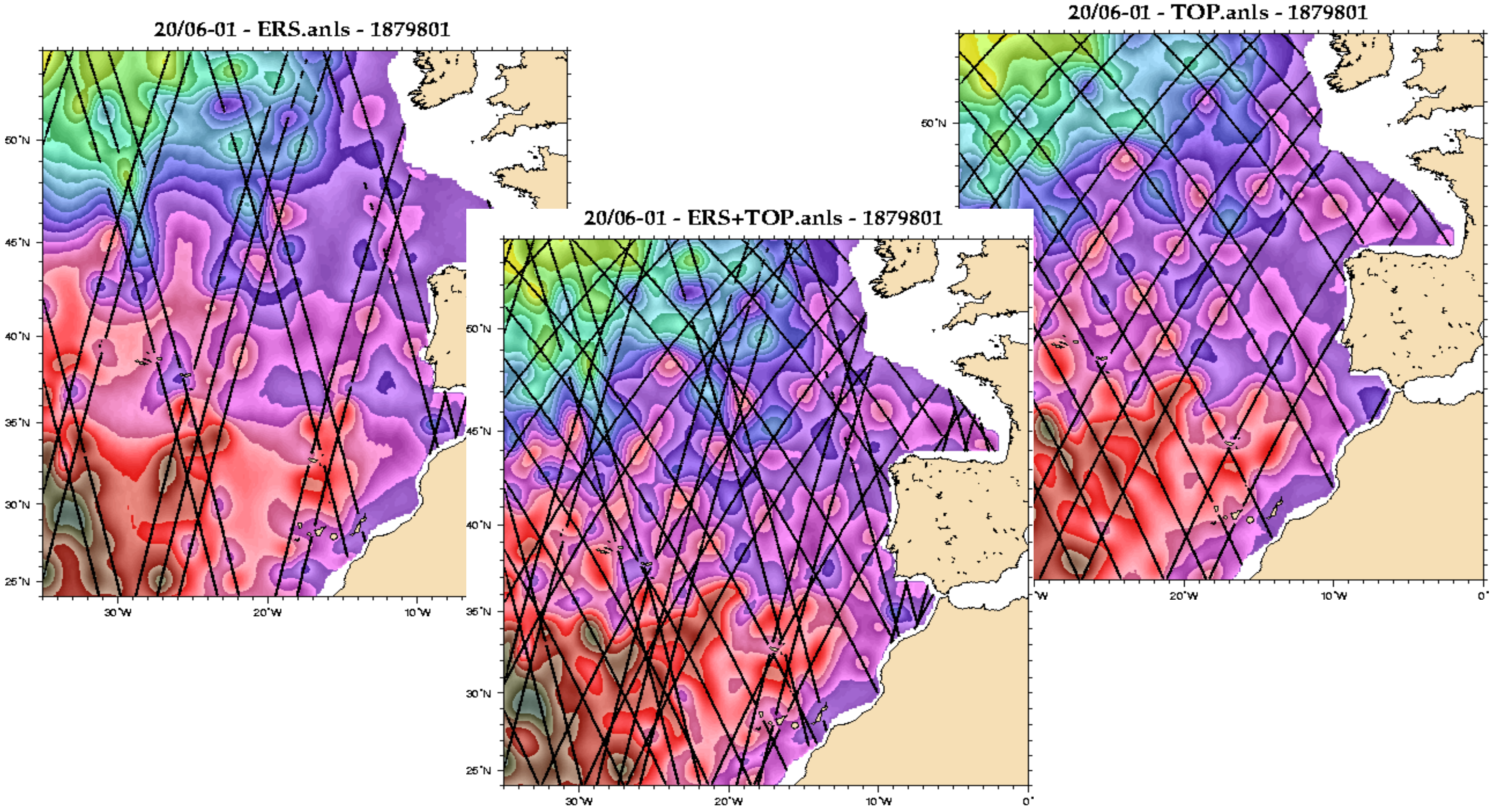
  - *from space*: sea-surface height (SSH), sea-surface temperature (SST)





3. Applied ocean data assimilation
Space observations

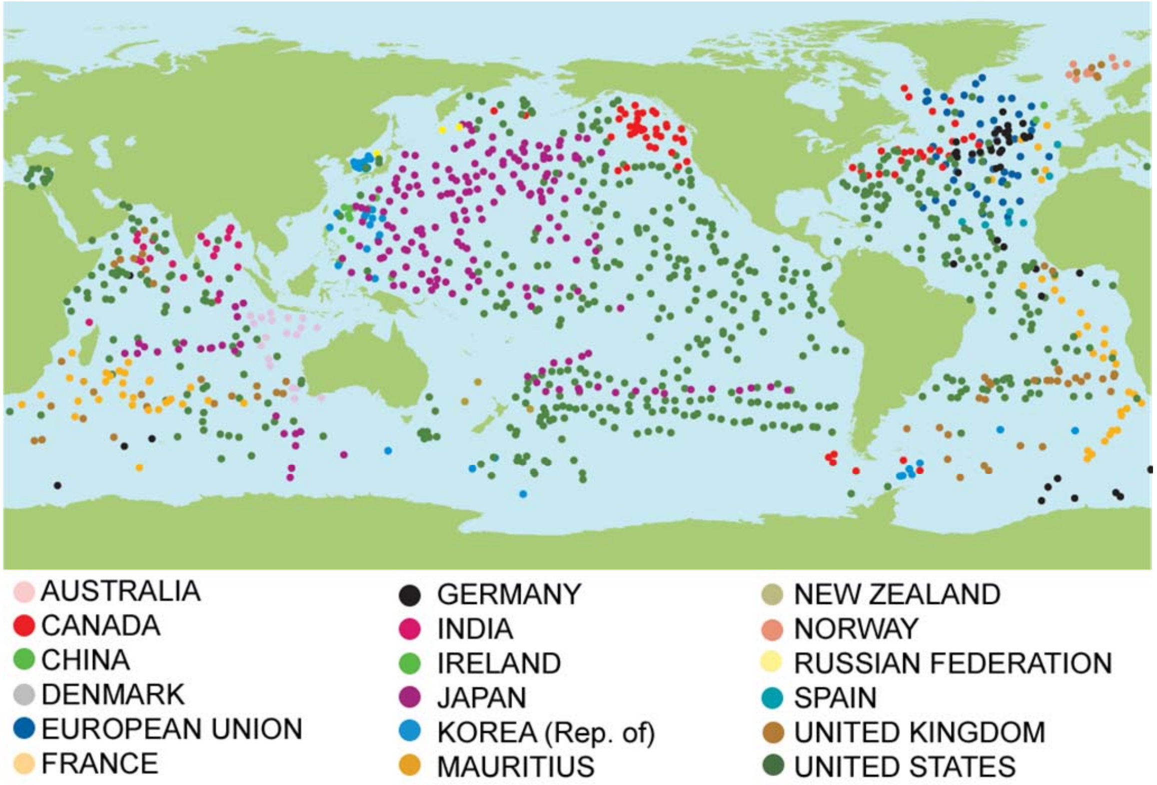
- ❑ One week of altimeter data

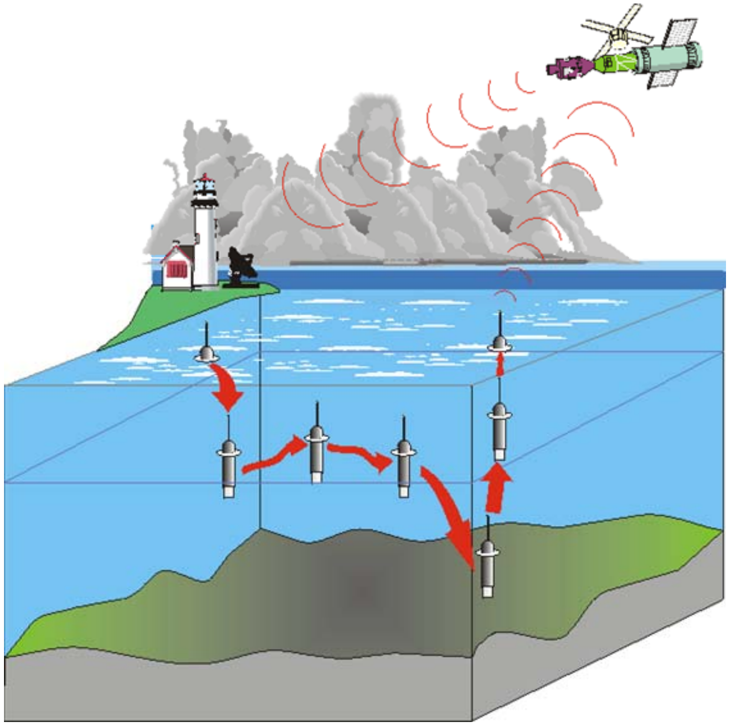

20/06-01 - ERS.anls - 1879801
50°N
45°N
40°N
35°N
30°N
25°N
30°W
20°W
10°W
20/06-01 - ERS+TOP.anls - 1879801
50°N
45°N
40°N
35°N
30°N
25°N
30°W
20°W
10°W
0°
20/06-01 - TOP.anls - 1879801
50°N
W
20°W
10°W
0°

3. Applied ocean data assimilation
In situ observations

- **Observed variables**:

  - *in situ*: T/S profiles (drifting floats, field campaigns, …)





ARGO, July 2004

LEGI
LABORATOIRE DES ECOULEMENTS GEOPHYSIQUES ET INDUSTRIELS

3. Applied ocean data assimilation
Observations

- **Observation vector $\mathbf{y}$** : data from various sources, at different time and space resolution
- **Radar Altimetry** : along-track measurements of SSH anomalies
  (JASON: 1 obs. / 7 km, ~ 300 km equatorial tracks separation, repeated every 10 days ;
  ERS/ENVISAT: ~ 80 km equatorial tracks separation, repeated every 35 days)
- **AVHRR SST** : weekly composite images at 4 km resolution (if no clouds)
- **ARGO flots** : 3° x 3° horizontal resolution (targetted), profiles (between 2000 m depth to surface) every 10 days with 1 obs / m along vertical

- $p = \dim \mathbf{y}$ is much smaller than $n = \dim \mathbf{x}$: too few observations !
- The ocean surface relatively well observed by satellites: vertical extrapolation of data assimilated at the surface into the ocean's interior has to be consistent with vertical data profiles
- The observed variables are closely related to model variables:
  $\mathbf{H}$ is mainly an interpolation operator (~ simple)

3. Applied ocean data assimilation
Error covariance matrix

- **<u>Specification of error covariance matrix</u>** $\mathbf{P}_0^a$ **?**

- Assume a background state $\boldsymbol{x}_0$ and associated error covariance $\mathbf{P}_0$
Consider the analysis step with only one data η at a model grid point
and the associated observation error $\boldsymbol{\varepsilon}$.
  - p = 1 , $\boldsymbol{y}$ is a scalar and $\mathbf{H}$ is a vector of the form $\mathbf{H} = \left[0,...,0,1,0,...,0\right]$
  - The Kalman gain is then a (n x 1) vector:

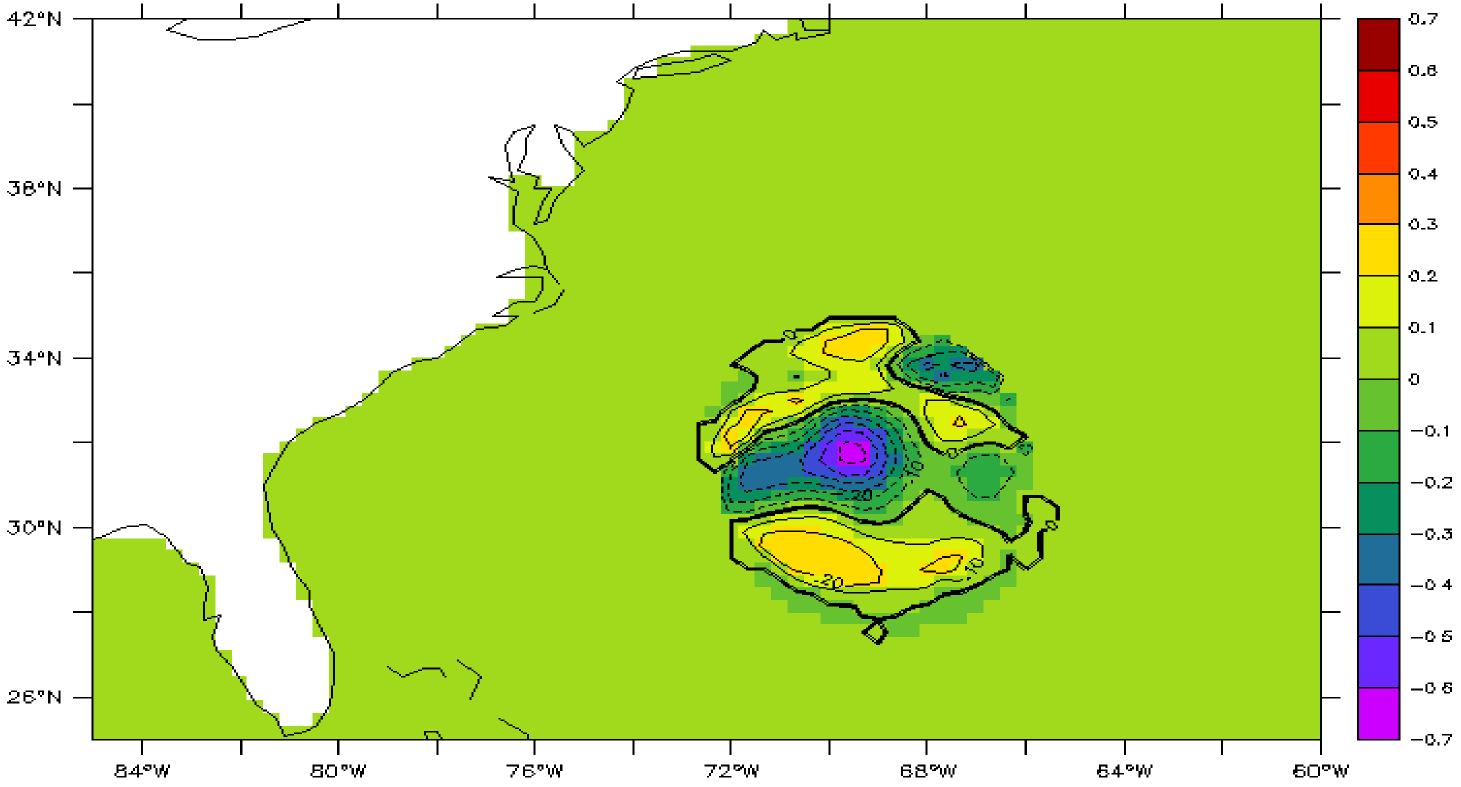
$$\mathbf{K} = \mathbf{P}_0\mathbf{H}^T(\mathbf{H}\mathbf{P}_0\mathbf{H}^T + \mathbf{R})^{-1} = \frac{1}{(p_{\eta\eta} + \varepsilon^2)}\left\{\mathbf{P}_0\right\}_\eta \quad \text{with} \quad p_{\eta\eta} = \left\{\mathbf{P}_0\right\}_{\eta\eta}$$

  - The posterior estimate is a correction of the background using the η–column of $\mathbf{P}_0$

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_0 + \frac{1}{(p_{\eta\eta} + \varepsilon^2)}\left\{\mathbf{P}_0\right\}_\eta(\eta - \eta_0) \quad \text{with} \quad \eta_0 = \left\{\boldsymbol{x}_0\right\}_\eta$$

3. Applied ocean data assimilation
Horizontal covariance structures

Example: Horizontal covariance relative to a SSH (η) point at (32ºN,70ºW)
MERCATOR Assimilation System - Testut *et al*.(2004)

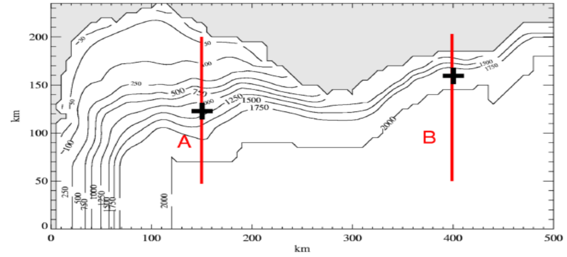


Influence function of a single altimeter measurement in the sub-tropical gyre

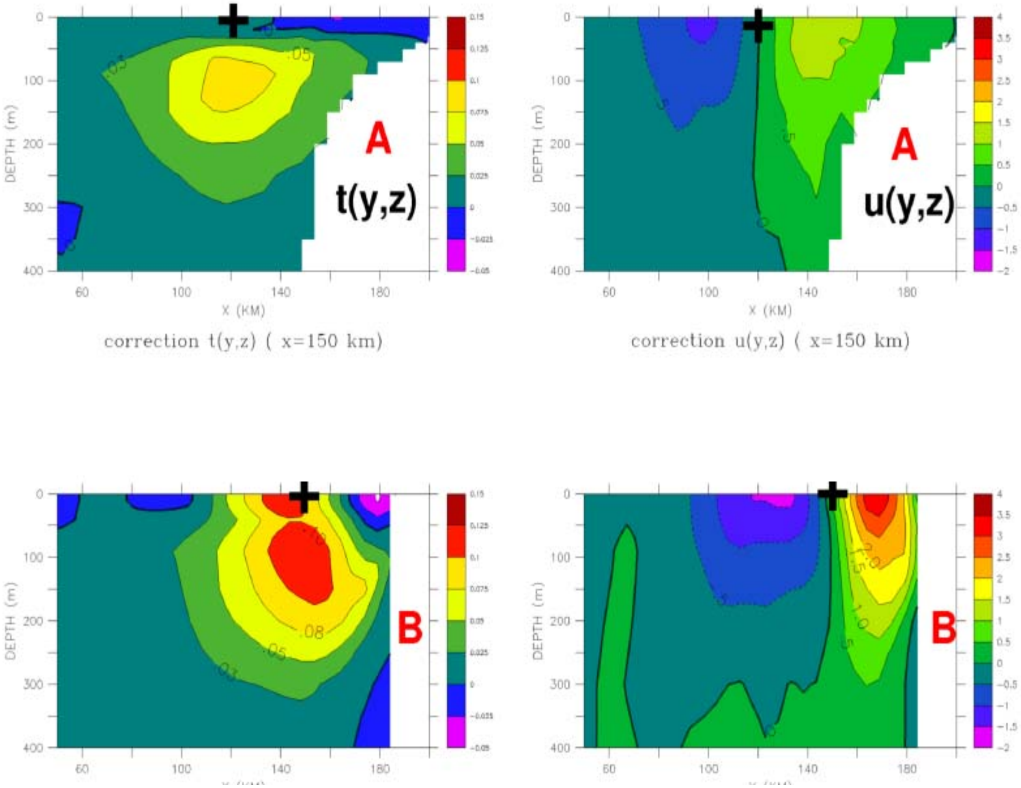
3. Applied ocean data assimilation
Vertical covariance structures

- **Multivariate representers in 3D space, showing covariance structures consistent with the model dynamics**

Temperature and velocity corrections to a free surface elevation misfit δη at two locations



(+): location of sea surface measurement, δη=5 cm



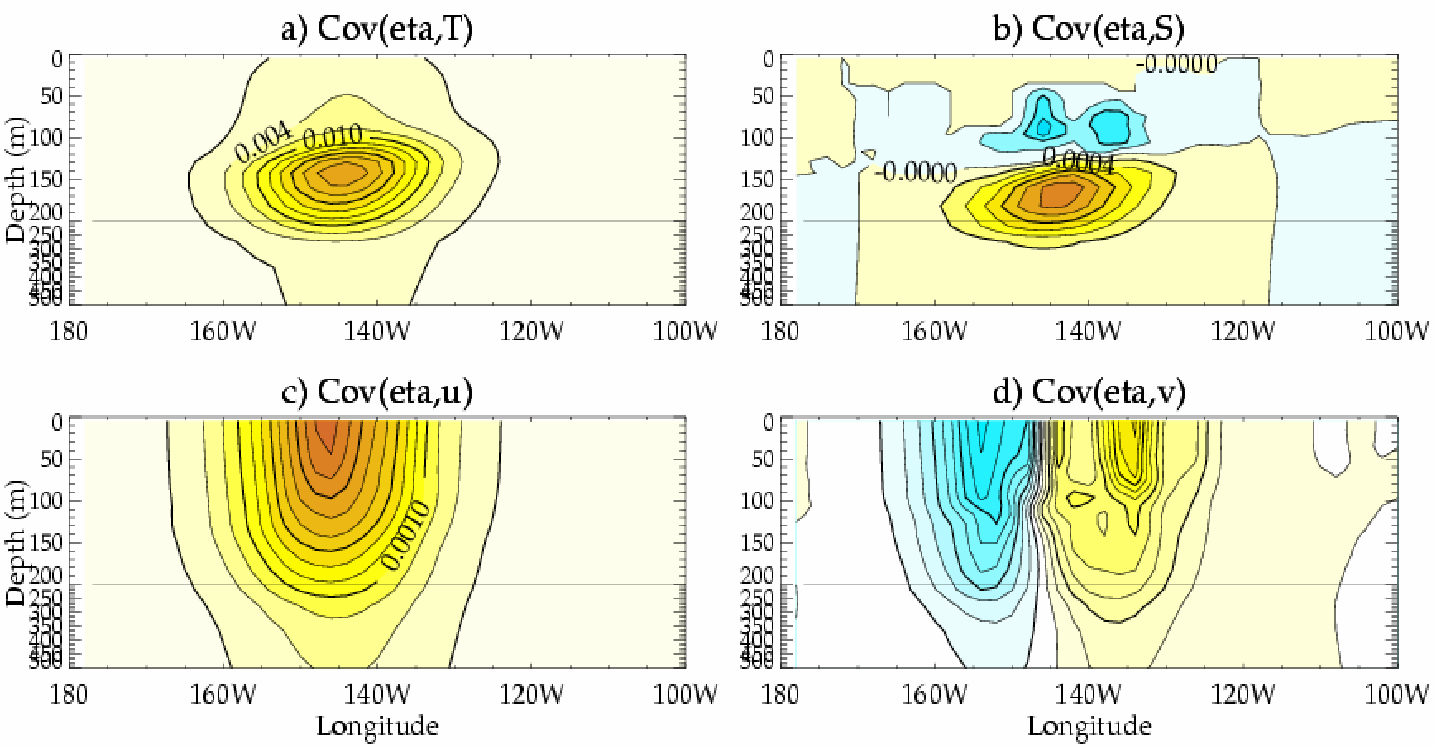
correction t(y,z) (x=150 km)

correction u(y,z) (x=150 km)

Representer functions of SSH in a free-surface coastal model
(*Echevin et al., JPO, 2000)*

3. Applied ocean data assimilation
Multivariate covariance structures

Example: covariance relative to a SSH (η) point at (0º,144ºW) - Weaver *et al*.(2003)



- The rows/colums of $\mathbf{P}$ should be « balanced » dynamically.
- This requires multivariate covariances
- A full-rank representation of $\mathbf{P}$ (dim n x n) is still impossible !

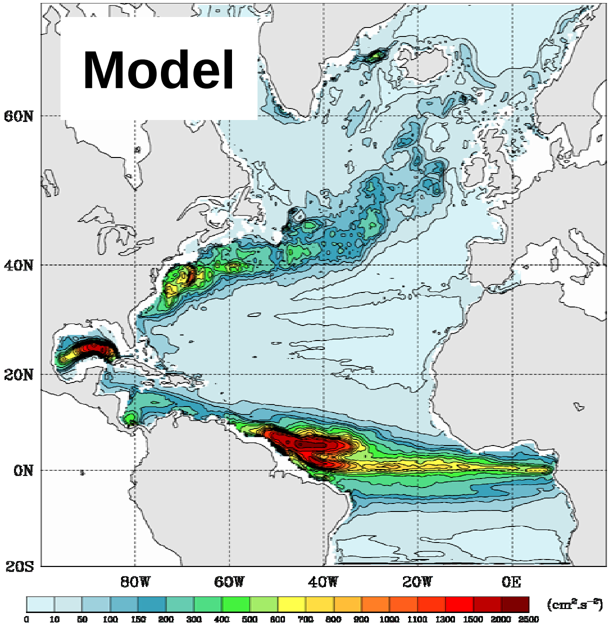
LEGI
LABORATOIRE DES ECOULEMENTS GEOPHYSIQUES ET INDUSTRIELS

3. Applied ocean data assimilation
Model errors Q

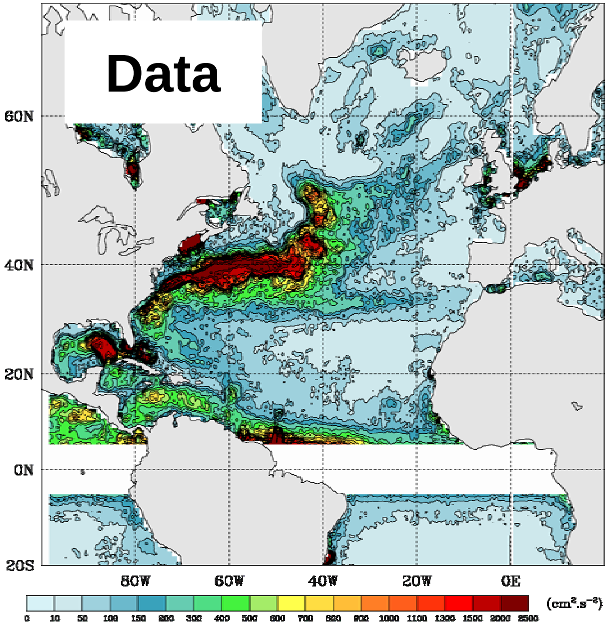
- **Model variability differs from observed variability**

EKE example :

Eddy Kinetic Energy at 50m depth on North Atlantic
NATL3 free simulation (1990-1999)



Eddy Kinetic Energy on North Atlantic
TP/ERS (Oct 92 to Oct 97)

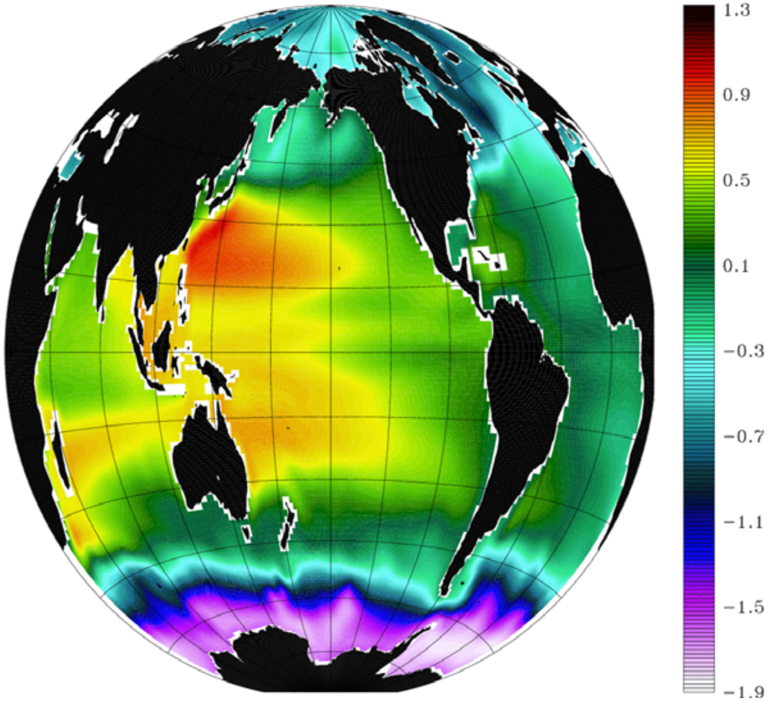


- Many different model error sources (finite discretizations, representation of bottom topography, atmospheric forcings, etc …) which cannot be easily quantified in terms of a **Q** matrix

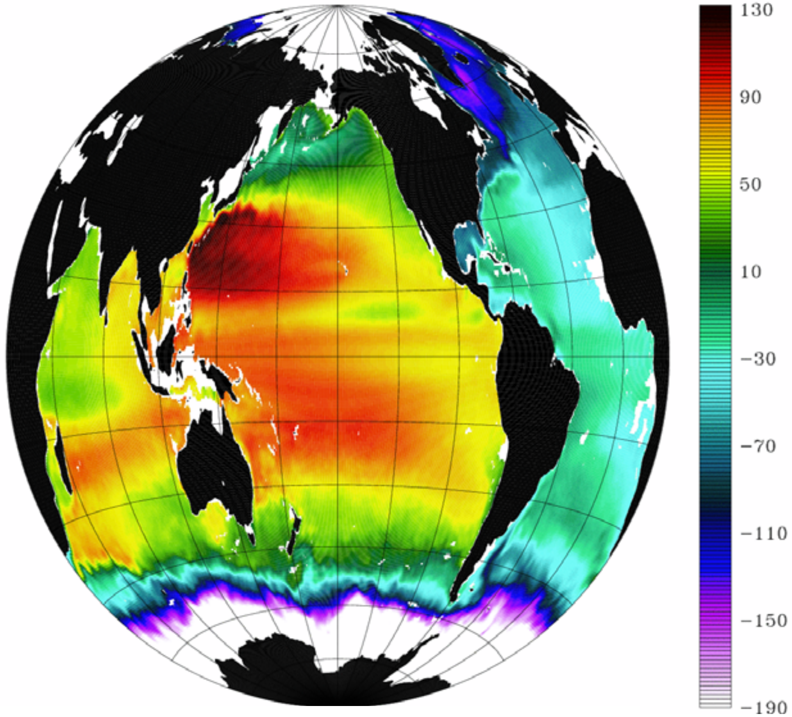
3. Applied ocean data assimilation *Systematic model errors*

- **Model mean SSH differs from observed mean SSH**

Mean sea-level difference between Pacific and Atlantic systematically too small in the model



OPA model (Madec et al. 2003)



Data (Niiler *et al*., 2003)

Optimality properties of the KF only in the absence of biases !

4. Simplifications of the Kalman Filter
Sub-optimal filters

- **« Optimal analysis is obtained when …the statistics …of error fields associated with forecasts and observations are known and accurately specified. Since these statistics are not generally available, actual implementations of assimilation algorithms are always sub-optimal»**

Dee and Da Silva

Q. J. R. Meteorol. Soc., 1998

Error dynamics :

- The forecast error requires ~ n model integrations !!!

$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\mathbf{P}_{i}^{a}\mathbf{M}^{T} + \mathbf{Q} = \mathbf{M}\left(\mathbf{M}\mathbf{P}_{i}^{a}\right)^{T} + \mathbf{Q}$$

- The ~ n model integrations are useless if $\mathbf{Q}$ is poorly known

4. Simplifications of the Kalman Filter
Optimal Interpolation

Simplification of the Kalman filter: « Optimal Interpolation »

To save cost and memory requirements, the KF can be simplified drastically by using time-independent « background » covariance matrix

$$\mathbf{P}_{i+1}^{f} = \mathbf{B} = \mathbf{D}^{1/2}\,\mathbf{C}\,\mathbf{D}^{1/2}$$

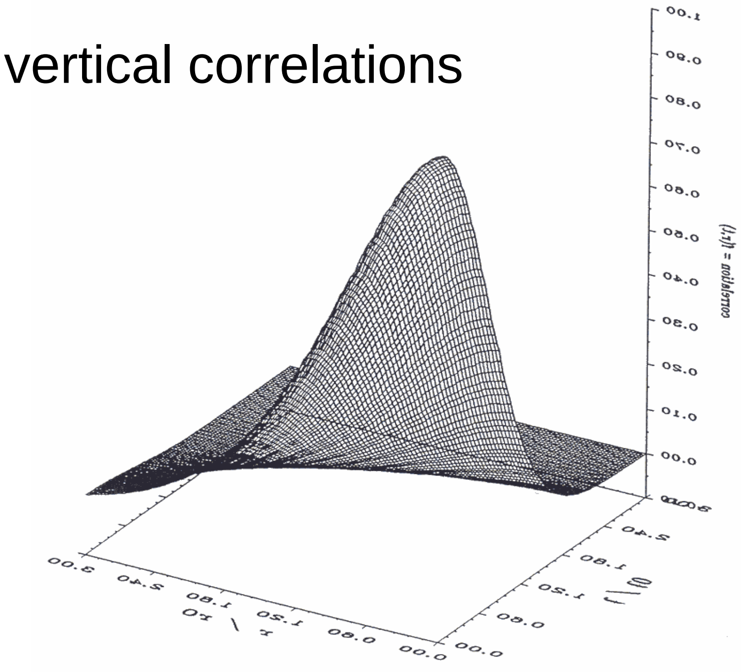
correlations

variances

- ❑ $\mathbf{C}$: expressed as a product of horizontal and vertical correlations

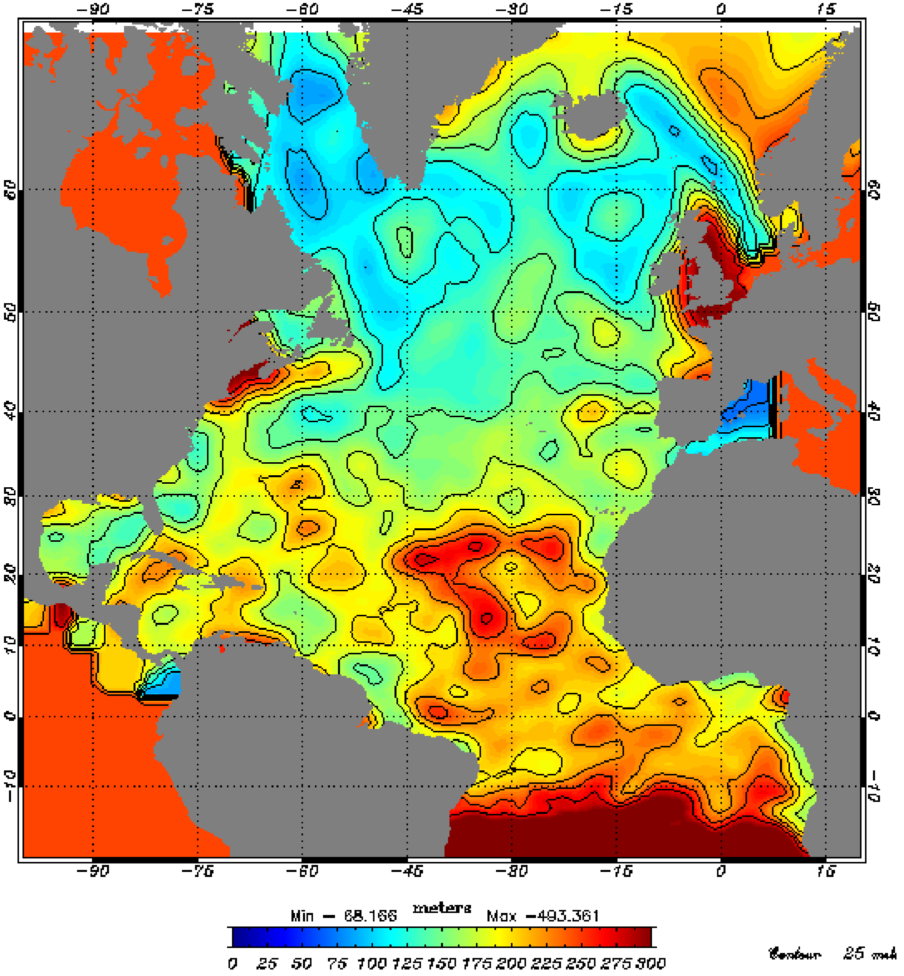
$$\{\mathbf{C}\}_{i,j} = c^{h}(l)\,.\,c^{v}(d)$$

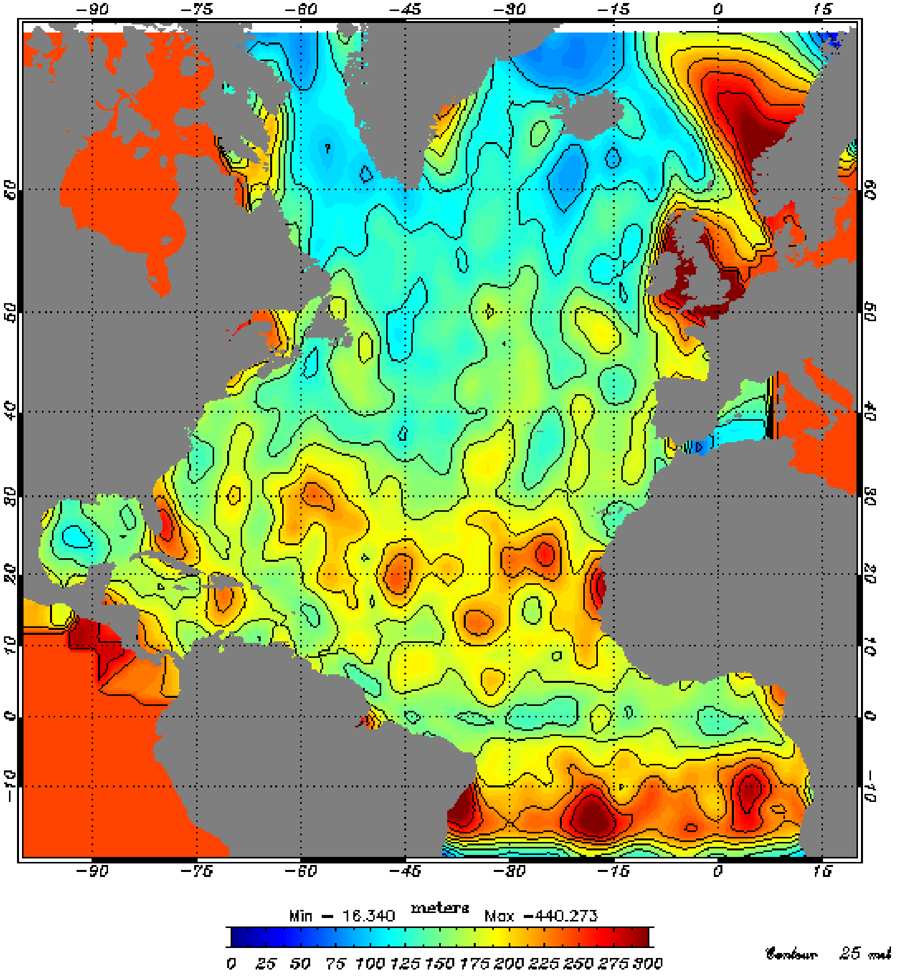
- ❑ Example: $c^{h}(l) = \left(1 + al + \frac{1}{3}a^{2}l^{2}\right)e^{-al}$



4. Simplifications of the Kalman Filter
Optimal Interpolation – SAM1

- ❑ **Meridional and zonal correlation scales prescribed in SAM1**





4. Simplifications of the Kalman Filter
Asymptotic filters

Steady-state filters *(Fukumori et al., 1993)*

The KF can be simplified using a time-independent covariance matrix computed as the asymptotic limit of the **Riccati equation**:

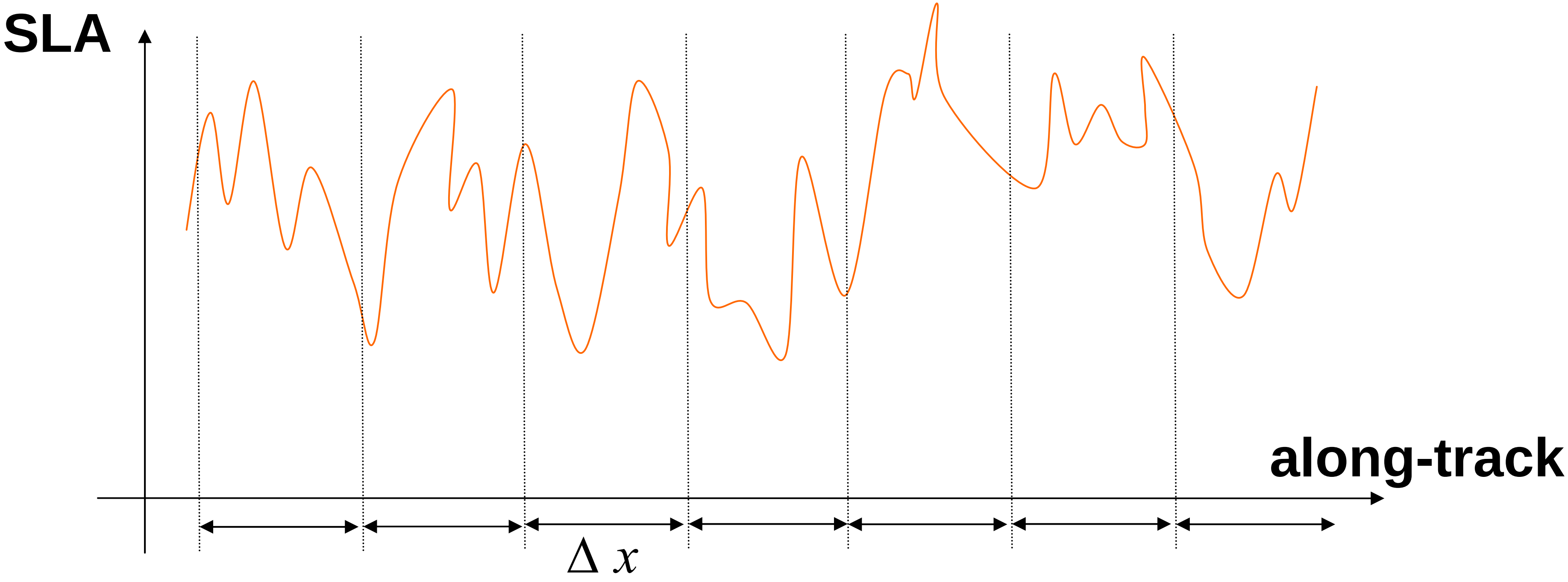
$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\left\{\underbrace{\mathbf{P}_i^{f} - \mathbf{P}_i^{f}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_i^{f}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{P}_i^{f}}_{\mathbf{P}_i^{a}}\right\}\mathbf{M}^{T} + \mathbf{Q}$$

Unicity of optimal estimates ?

The same model with the same set of observations can provide different sequences of « optimal estimates », depending on the « target » field !

- Let M : « perfect eddy-resolving » ocean model

  y : « perfect filament-resolving » ocean observations

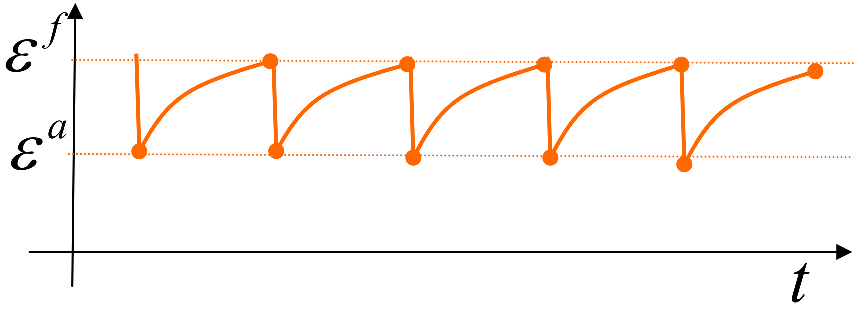


Assimilation of sub-grid scale data

Target = eddies (~ 50 km)

$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\mathbf{P}_{i}^{a}\mathbf{M}^{T} + 0$$

$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^{f}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}_{i+1}^{f}\mathbf{H}^{T} + \mathbf{R}^{\varphi})^{-1}$$



$\mathbf{R}^{\varphi}$: representativeness error due to the observation of filaments

Example : scalar, linear model $x_{i+1}^f = m x_i^a$

- Target = model resolution
- KF equations

$$p_{i+1}^f = m p_i^a m + 0 = m^2 p_i^a$$

$$k_{i+1} = p_{i+1}^f (p_{i+1}^f + r^\varphi)^{-1} = m^2 p_i^a (m^2 p_i^a + r^\varphi)^{-1} \qquad (0 < k_{i+1} < 1)$$

$$p_{i+1}^a = (1 - k_{i+1}) p_{i+1}^f = r^\varphi m^2 p_i^a (m^2 p_i^a + r^\varphi)^{-1} \qquad (0 < p_{i+1}^a < r^\varphi)$$

Example : scalar, linear model $x_{i+1}^{f} = m x_i^a$

- Target = observation resolution
- KF equations

$$p_{i+1}^{f} = m p_i^a m + q^{\varphi} = m^2 p_i^a + q^{\varphi}$$

$$k_{i+1} = p_{i+1}^{f}\left(p_{i+1}^{f} + 0\right)^{-1} = 1 \quad \Rightarrow \quad x_{i+1}^{a} = x_{i+1}^{f} + k_{i+1}\left(y_{i+1} - x_{i+1}^{f}\right) = y_{i+1} \; !$$

$$p_{i+1}^{a} = \left(1 - k_{i+1}\right) p_{i+1}^{f} = 0$$

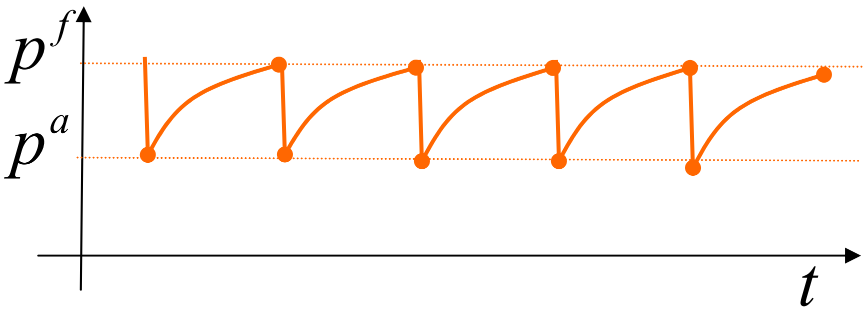
Example : scalar, linear model $x^{f}_{i+1} = m x^{a}_{i}$

- Numerical example: $m = \sqrt{2} \quad q^{\varphi} = r^{\varphi} = \varepsilon$

Target = model resolution

$$p^{a} = 0.5\,\varepsilon$$

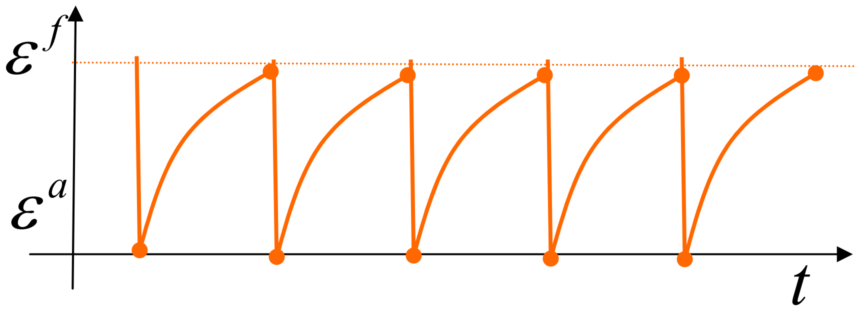
$$p^{f} = \varepsilon$$



Target = data resolution

$$p^{a} = 0$$

$$p^{f} = \varepsilon$$



Example : scalar, linear model $x_{i+1}^{f} = mx_{i}^{a}$

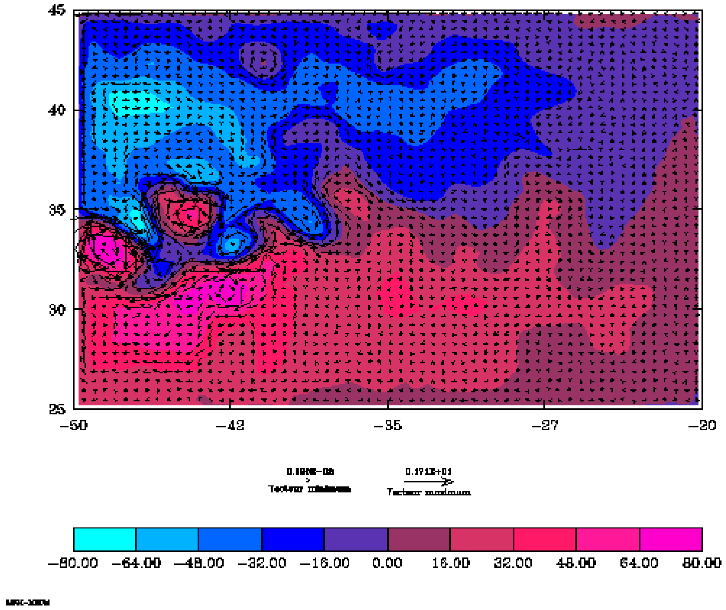
- Misleading implementation : $q^{\varphi} = r^{\varphi} = 0$
- Initialization : x_0, p_0

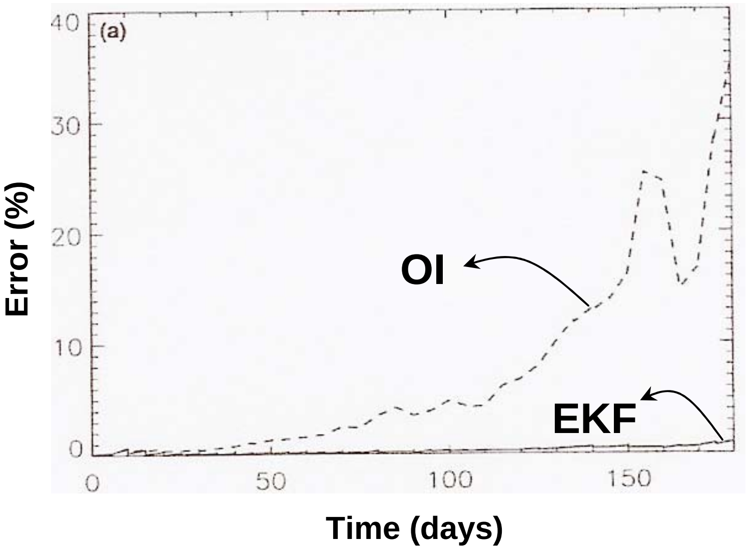
OI vs. EKF

- ❑ A full Kalman filter cannot be implemented into realistic ocean models (error forecast and analysis equations too expensive in CPU and memory requirements)

- ❑ « Optimal Interpolation » over-simplifies the propagation of errors by neglecting dynamical principles and statistical information

Idealized double-gyre model (Ballabrera *et al.*, 2001)


45
40
35
30
25
−50
−42
−35
−27
−20
−80.00
−64.00
−48.00
−32.00
−16.00
0.00
16.00
32.00
48.00
64.00
80.00


(a)
Error (%)
Time (days)
OI
EKF