# Introduction to data assimilation applied to ocean models


Alexander Barth

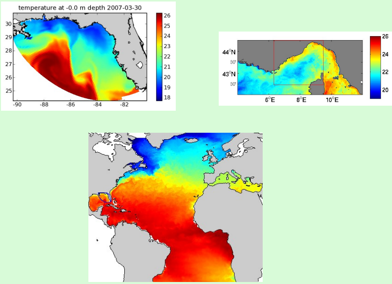
a.barth@ulg.ac.be


April 17, 2014

# Data assimilation methods

# What is data assimilation ?

Model




Observations

Data Assimilation

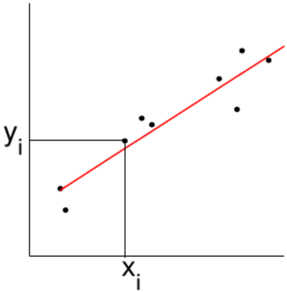
Analysis

# Outline

- Why data assimilation ?
- Basic concepts
- Sequential assimilation
  - Nudging
  - Successive corrections
  - Optimal Interpolation
  - 3D-Var
  - Kalman filter
  - Kalman smoother
- Non-Sequential assimilation
  - 4D-Var
  - Representer method

# Goal of data assimilation

- Calibration: choose model parameters coherent with observations. Example: linear regression.




$$J(a,b) = \sum_i \frac{1}{\sigma_i^2} \left[ y_i - (a x_i + b) \right]^2$$

$$\frac{\partial J}{\partial a} = \frac{\partial J}{\partial b} = 0$$

$$\frac{1}{\sigma_a^2} = \frac{1}{2} \frac{\partial^2 J}{\partial a^2} \qquad \frac{1}{\sigma_b^2} = \frac{1}{2} \frac{\partial^2 J}{\partial b^2}$$

- Improve the model accuracy with help of observations
- Data assimilation provides also a framework to identify model errors
- State estimation: determine the “best” (e.g. the most probable) state of a system

# Errors and uncertainty

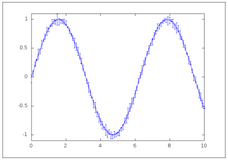
- Neither the model nor the observations are perfect.
- Both have **errors** (uncertainty).
- error $=$ systematic error (bias) $+$ random error
- How can we represent uncertainty?

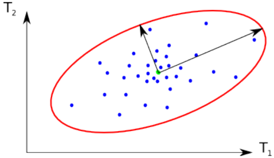
# Ways to represent uncertainty

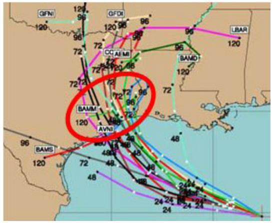
- For Gaussian-distributed errors.
  - Error bars (for scalar variables) (mean, standard deviation) or confidence interval
  - error covariance, example: $\mathbf{x} = (T_1, T_2)$

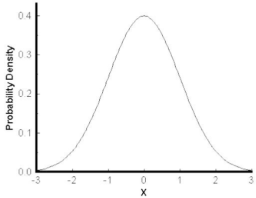
$$\mathbf{P} = \begin{pmatrix} \color{red}{P_{11}} & P_{12} \\ P_{12} & \color{red}{P_{22}} \end{pmatrix}$$

  - Error modes (EOF: empirical orthogonal functions)
  - Graphical representation: ellipsoid for more than one variable (vectors) (mean, error covariance): $\mathbf{x}\mathbf{P}^{-1}\mathbf{x} = 1$
- ensemble of possible values
- probability density function





# Errors in an ocean model

Errors in an ocean model might be due to

- errors in initial conditions
- errors in open ocean boundary conditions
- errors in atmospheric fields (wind, air temperature, ...)
- errors in bathymetry
- inappropriate parameterizations
- discretization error
- ...

What about model errors in other disciplines?

# Errors in your observations

Errors in your observations might be due to

- instrumental error (bias, drift, limited accuracy and precision)

Observations might not represent exactly the same as the model variables

- mismatch in resolved scale
- mismatch in resolved processes
- ...

# Notation

| | | |
|---|---|---|
| $n$ | scalar | number of state variables |
| $m$ | scalar | number of observations |
| $N$ | scalar | number of ensemble members |
| $k$ | scalar | ensemble index $k = 1, \ldots, N$ |
| $J$ | scalar | cost function |
| $f$ | function | model giving the model state vector at the next time step |
| $\mathbf{M}$ | matrix $n \times n$ | linear (or linearized) model |
| $\mathbf{x}^{f/a/t}$ | vector $n \times 1$ | the model forecast/analysis/truth |
| $\mathbf{P}^{f/a}$ | matrix $n \times n$ | error covariance of $\mathbf{x}^{f/a}$ |
| $\mathbf{S}^{f/a}$ | matrix $n \times N$ | square root decomposition of $\mathbf{P}^{f/a}$ |
| $\boldsymbol{\eta}_n$ | vector $n \times 1$ | the model error |
| $\mathbf{Q}$ | matrix $n \times n$ | error covariance of $\boldsymbol{\eta}_n$ |
| $\mathbf{y}^o$ | vector $m \times 1$ | observations |
| $\boldsymbol{\varepsilon}$ | vector $m \times 1$ | observation error |
| $\mathbf{R}$ | matrix $m \times m$ | error covariance of $\mathbf{y}^o$ |
| $\mathbf{H}$ | matrix $n \times m$ | observation operator |
| $E[\cdot]$ | | expectation |

The superscript $^f$ and $^a$ refer to forecast and analysis respectively.

# Basic concepts

- The **state vector** $\mathbf{x}_n$ at time $t_n$. For a primitive equation model, its dimension is about $n = 5 \times 5000 \times 20 = 5\ 10^6$.

- The **dynamical model** $f_n$:

$$\begin{aligned} \mathbf{x}_{n+1} &= f_n(\mathbf{x}_n) \quad [= \mathbf{M}_n\mathbf{x}_n + \mathbf{F}_n \text{ if the model is linear}] \\ \mathbf{x}_0 &= \mathbf{x}^i \end{aligned}$$

  model $\approx$ reality ($t$: true):

$$\begin{aligned} \mathbf{x}^t_{n+1} &= f_n(\mathbf{x}^t_n) + \boldsymbol{\eta}_n \\ \mathbf{x}^t_0 &= \mathbf{x}^i + \boldsymbol{\eta}^i \end{aligned}$$

- $\mathbf{x}^t$ is of course unknown for in a real application. The assimilation method do not require the knowledge of $\mathbf{x}^t$.

- The **observations**:

$$\mathbf{y}^o_n = h_n(\mathbf{x}^t_n) + \boldsymbol{\varepsilon}_n \quad \left[= \mathbf{H}_n\mathbf{x}^t_n + \boldsymbol{\varepsilon}_n \text{ if the model is linear}\right]$$

  in general: $\mathbf{H}_n$ = interpolation to observation grid $\circ$ variable transformation

$\boldsymbol{\varepsilon}_n$ = instrumental error + representativity error

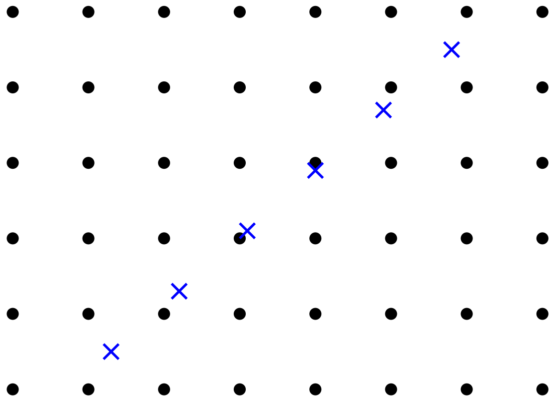


Figure 1: For example, an altimetry track and model grid points

## Assumptions

- All errors are zero in average (*i.e.* no bias):

$$E[\boldsymbol{\eta}_n] = E[\boldsymbol{\eta}^i] = E[\boldsymbol{\varepsilon}_n] = 0$$

- The covariances are known:

$$\begin{aligned} E[\boldsymbol{\eta}_n \boldsymbol{\eta}_{n'}^T] &= \mathbf{Q}_n \delta_{nn'} & E[\boldsymbol{\eta}_n {\boldsymbol{\eta}^i}^T] &= 0 \\ E[\boldsymbol{\eta}^i {\boldsymbol{\eta}^i}^T] &= \mathbf{P}^i & E[\boldsymbol{\eta}_n \boldsymbol{\varepsilon}_{n'}^T] &= 0 \\ E[\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_{n'}^T] &= \mathbf{R}_n \delta_{nn'} \end{aligned}$$

- Some assimilation methods are optimal if those assumptions are verified.
- If the assumptions are not verified (in particular biased model), the assimilation schemes can still given useful results.
- For some assimilation method the error covariance matrix of the model state $\mathbf{x}$ is assumed to be known:

$$E[\left(\mathbf{x} - \mathbf{x}^t\right) \left(\mathbf{x} - \mathbf{x}^t\right)^T] = \mathbf{P}$$

## Consistency check

- Innovation vector $\mathbf{d}_n$ (time index $n$ is dropped in the following):

$$\begin{aligned}\mathbf{d} &= \mathbf{y}^o - \mathbf{H}\mathbf{x}^f = \mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{H}\left(\mathbf{x}^f - \mathbf{x}^t\right) \\ E[\mathbf{d}] &= 0 \\ E[\mathbf{d}\mathbf{d}^T] &= \mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T\end{aligned}$$

- $\mathbf{H}\mathbf{P}\mathbf{H}^T$ is the error covariance of $\mathbf{H}\mathbf{x}$.
- One can use these relationships to test if the model is unbiased and if the error covariances are consistent.
- Normalized innovation $\mathbf{z} = \left(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T\right)^{-1/2} \mathbf{d}$ should follow a Gaussian distribution with zero mean unit covariance.
- Verification statistics:

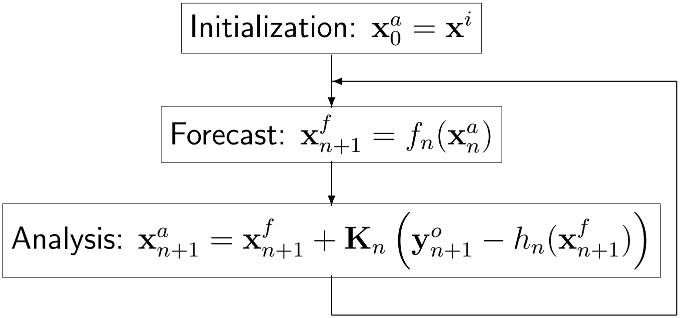
$$\mathrm{tr}(\mathbf{z}\mathbf{z}^T) = \chi^2_m$$

The left-hand side of the previous equation follow is a sum of $m$ Gaussian distributed variables squared. It follows thus a $\chi^2$ distribution with $m$ degrees of freedoms. This distribution has a mean of $m$ and a variance of $2m$ (Dee, 1995).

- One can also show that (Desroziers *et al.*, 2005):

$$\begin{aligned} E\left[\left(\mathbf{Hx}^a - \mathbf{Hx}^b\right)\left(\mathbf{y}^o - \mathbf{Hx}^b\right)\right] &= \mathbf{HP}^f\mathbf{H}^T \\ E\left[\left(\mathbf{y}^o - \mathbf{Hx}^a\right)\left(\mathbf{y}^o - \mathbf{Hx}^b\right)\right] &= \mathbf{R} \\ E\left[\left(\mathbf{Hx}^a - \mathbf{Hx}^b\right)\left(\mathbf{y}^o - \mathbf{Hx}^a\right)\right] &= \mathbf{HP}^a\mathbf{H}^T \end{aligned}$$

# Sequential assimilation




- $\mathbf{K}_n$: Kalman gain
- Analysis $=$ only unbiased estimation if $h$ is linear

# Unbiased linear combination

- Model forecast $\mathbf{x}^f$ and observations $\mathbf{y}^o$ are assumed unbiased
- Linear combination $\mathbf{x}^a$ should be unbiased too $E[\mathbf{x}^a] = \mathbf{x}^t$
- General form of linear combination

$$\begin{aligned}
\mathbf{x}^a &= \mathbf{J}\,\mathbf{x}^f + \mathbf{K}\,\mathbf{y}^o \\
E[\mathbf{x}^a] &= \mathbf{J}\,E[\mathbf{x}^f] + \mathbf{K}\,E[\mathbf{y}^o] \\
E[\mathbf{x}^a] &= \mathbf{J}\,\mathbf{x}^t + \mathbf{K}\mathbf{H}\,\mathbf{x}^t \\
E[\mathbf{x}^a] &= (\mathbf{J} + \mathbf{K}\mathbf{H})\,\mathbf{x}^t
\end{aligned}$$

therefore $\mathbf{J} + \mathbf{K}\mathbf{H} = \mathbf{I}$. If we choose $\mathbf{J} = \mathbf{I} - \mathbf{K}\mathbf{H}$,

- Analysis:

$$\begin{aligned}
\mathbf{x}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{x}^f + \mathbf{K}\mathbf{y}^o \\
\mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right)
\end{aligned}$$

# Direct insertion

- Part of the state vector is directly observed (e.g. SST)
- The observed part of the state vector is replaced by the observations.

$$\mathbf{x}^a_{nj'(i)} = \mathbf{y}^o_{ni}$$
$$\mathbf{x}^a_{nj} = \mathbf{x}^f_{nj} \qquad \text{on non-observed grid points}$$

- The $i$th observation corresponds to the $j'(i)$ element of the state vector
- The observation operator will be one for the observed elements of the state vector and zero otherwise ($\mathbf{H}_{j'(i),i} = 1$).

$$\mathbf{x}^a_n = \mathbf{x}^f_n + \mathbf{H}^T \left( \mathbf{y}^o - \mathbf{x}^f_n \right)$$

- Error in the model are assumed to be much larger than errors of the observations
- Problems
  - Updated part of the state vector is inconsistent relative to the part of the state vector which is not observed.
  - Adjustment processes (e.g. geostrophic adjustment creating barotropic waves, mixing) can degrade the model results

# Nudging

- As in direct insertion, a part of the state vector must be directly observed.
- Analysis:

$$\mathbf{x}^a_{nj'(i)} = \mathbf{x}^f_{nj'(i)} + r_i \left( \mathbf{y}^o_{ni} - \mathbf{x}^f_{nj'(i)} \right)$$

$$\mathbf{x}^a_{nj} = \mathbf{x}^f_{nj} \qquad \text{on non-observed grid points}$$

- In matrix form:

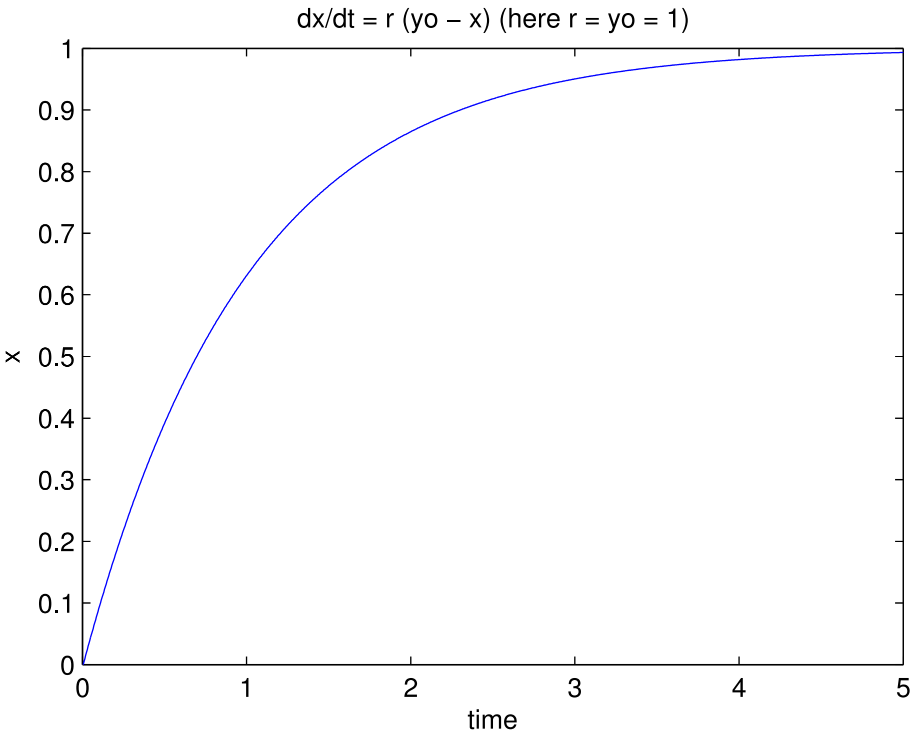
$$\mathbf{x}^a_n = \mathbf{x}^f_n + r_i \mathbf{H}^T \left( \mathbf{y}^o - \mathbf{x}^f_n \right)$$

- For a scalar variable: ($1/r =$ relaxation time scale)

$$\frac{dx}{dt} = f(x(t)) + r\left(y^o(t) - x(t)\right)$$

- Relaxation term is applied at the model time step.
- SST Nudging $\Rightarrow$ correction of surface heat flux.
- Nudging towards climatology to prevent drift of the model.
- Relaxation reduces the model variability.

# Example



# Demonstration

- A web-application showing the functioning of the Kalman filter is available at http://www.data-assimilation.net/Tools/AssimDemo/.
- Review of what is a twin-experiment
- Very simple models can be used:

## No time variation

The state vector $\mathbf{x}$ has two elements $(x_1, x_2)^T$ and there is no time variation:

$$\mathbf{x}_{n+1} = \mathbf{x}_n \tag{1}$$

The model matrix $\mathbf{M}$ is thus the identity matrix.

## 1D advection in periodic domain

The state vector $\mathbf{x}$ has four elements and it is subjected to the following dynamics

$$\mathbf{x}^{(n+1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \mathbf{x}^{(n)} \tag{2}$$

This simple system would be the result of a 1D advection scheme in a periodic domain with a constant velocity. The grid resolution over time step is equal to the velocity.

Without using the web-interface, what would be the model state after the 1st, 2nd,... time step?.

## Oscillations

- The state vector $\mathbf{x}$ has two elements and it is governed by:

$$\frac{dx_1}{dt} = f x_2 \quad (3)$$

$$\frac{dx_2}{dt} = -f x_1 \quad (4)$$

- The numerical example uses $f = 2\pi$ with a time step of $\Delta t = 0.1$. One can show that two successive states are related by:

$$\mathbf{x}^{(n+1)} = \begin{pmatrix} \cos(f\Delta t) & \sin(f\Delta t) \\ -\sin(f\Delta t) & \cos(f\Delta t) \end{pmatrix} \mathbf{x}^{(n)} \quad (5)$$

- What kind of oscillation would these equations describe in the ocean?

# Two oscillations

The state vector $\mathbf{x}$ has four elements and it is governed by:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} = \begin{pmatrix} 0 & 0 & -a & -b \\ 0 & 0 & -b & -a \\ a & b & 0 & 0 \\ b & a & 0 & 0 \end{pmatrix} \mathbf{x}$$

where $a = 2\pi$ and $b = \pi$. The eigenvectors and eigenvalues of the model matrix allows us to find an analytical solution:

$$\mathbf{x}(t) = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 \\ C_1 & C_2 & -C_3 & -C_4 \\ -C_2 & C_1 & -C_4 & C_4 \\ -C_2 & C_1 & C_4 & -C_3 \end{pmatrix} \begin{pmatrix} \cos(\omega t) \\ \sin(\omega t) \\ \cos(\omega' t) \\ \sin(\omega' t) \end{pmatrix}$$

where $\omega = a + b$ and $\omega' = a - b$.

In the numerical example, this equation is solved with a Crank-Nicholson schema and a time step $\Delta t = 0.1$.

$$\begin{aligned} \frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\Delta t} &= \mathbf{A} \, \frac{\mathbf{x}_{n+1} + \mathbf{x}_n}{2} \\ \left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{A}\right)\mathbf{x}_{n+1} &= \left(\mathbf{I} + \frac{\Delta t}{2}\mathbf{A}\right)\mathbf{x}_n \end{aligned}$$

The model matrix is thus:
$\mathbf{M} = \left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{A}\right)^{-1}\left(\mathbf{I} + \frac{\Delta t}{2}\mathbf{A}\right)$.

$$\mathbf{x}_{n+1} = \mathbf{M}\mathbf{x}_n \tag{6}$$

# Lorenz model

The classical Lorenz model (simplified mathematical model for atmospheric convection) with $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$.

$$\frac{dx}{dt} = \sigma(y - x) \tag{7}$$

$$\frac{dy}{dt} = x(\rho - z) - y \tag{8}$$

$$\frac{dz}{dt} = xy - \beta z \tag{9}$$

The system is discretized with a Runge Kutta time stepping scheme with $\Delta t = 0.05$.

# Nudging demo

- Model: identity $\mathbf{x}_{n+1} = \mathbf{x}_n$
  - Single observation (Model time steps between observations: 25)
    http://data-assimilation.net/Tools/AssimDemo/?method=Nudging&model=id&obs_tsteps=25
  - Relaxation term acts as low pass-filter (Model time steps between observations: 1)
    http://data-assimilation.net/Tools/AssimDemo/?method=Nudging&model=id&obs_tsteps=1
  - Over-fitting of observations if nudging relaxation time-scale is too short (Model time steps between observations: 1, relaxation time-scale: 2)
    http://data-assimilation.net/Tools/AssimDemo/?method=Nudging&model=id&obs_tsteps=1&nudging_
    ts=2
- Model: oscillation
  - Based on the default values, try to find a good relaxation time-scale
    http://data-assimilation.net/Tools/AssimDemo/?method=Nudging&model=oscillation
  - In which sense would you need to the change the other parameters to improve the solution with assimilation?

# Optimal Interpolation

- The observation operator must be linear
- The error covariance of the model state vector is defined as:

$$\mathbf{P}_n^{f,a} = E[(\mathbf{x}_n^{f,a} - \mathbf{x}_n^t)(\mathbf{x}_n^{f,a} - \mathbf{x}_n^t)^T]$$

- We assume that $\mathbf{P}_n^f$ is known.
- The Kalman gain is chosen such that the norm of $\mathbf{x}_n^a - \mathbf{x}_n^t$ is as small as possible:

$$J(\mathbf{K}) = E[(\mathbf{x}_n^a - \mathbf{x}_n^t)^T \mathbf{W} (\mathbf{x}_n^a - \mathbf{x}_n^t)] = \mathrm{tr}(\mathbf{W}\mathbf{P}_n^a)$$

- We introduce an error norm with the diagonal matrix $\mathbf{W}$
- The optimal value of $\mathbf{K}$ is independent of $\mathbf{W}$

$$\mathbf{K}_n = \mathbf{P}_n^f \mathbf{H}_n^T \left( \mathbf{H}_n \mathbf{P}_n^f \mathbf{H}_n^T + \mathbf{R}_n \right)^{-1}$$

# How to derive the Kalman gain?

The analysis is given by:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \quad (10)$$

$$= \left(\mathbf{I} + \mathbf{K}\mathbf{H}\right)\mathbf{x}^f + \mathbf{K}\mathbf{y}^o \quad (11)$$

The variance of the analysis $\mathbf{x}^a$ is a function of the gain matrix $\mathbf{K}$:

$$\mathbf{P}^a\left(\mathbf{K}\right) = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^f\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (12)$$

We want to have the overall smallest possible error on $\mathbf{x}^a$.

$$\mathrm{tr}\left(\mathbf{W}\mathbf{P}^a\left(\mathbf{K}\right)\right) = \mathrm{tr}\left(\mathbf{W}\mathbf{P}^f\right) - 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{K}\mathbf{H}\mathbf{P}^f\right) + \mathrm{tr}\left(\mathbf{W}\mathbf{K}\;\mathbf{H}\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T\right) + \mathrm{tr}\left(\mathbf{W}\mathbf{K}\mathbf{R}\mathbf{K}^T\right) \quad (13)$$

If $\mathbf{K}$ is the optimal gain, then a small increment of $\delta\mathbf{K}$ does not modify the total error variance in the first order of $\delta\mathbf{K}$.

$$\mathrm{tr}\left(\mathbf{W}\mathbf{P}^a\left(\mathbf{K} + \delta\mathbf{K}\right)\right) - \mathrm{tr}\left(\mathbf{W}\mathbf{P}^a\left(\mathbf{K}\right)\right) \quad (14)$$

$$= 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{K}\;\mathbf{H}\mathbf{P}^f\mathbf{H}^T\,\delta\mathbf{K}^T\right) - 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{P}^f\mathbf{H}^T\delta\mathbf{K}^T\right) + 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{K}\mathbf{R}\delta\mathbf{K}^T\right)$$

$$= 2\,\mathrm{tr}\left(\mathbf{W}\left[\mathbf{K}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right) - \mathbf{P}^f\mathbf{H}^T\right]\delta\mathbf{K}^T\right)$$

Since the perturbation $\delta\mathbf{K}$ is arbitrary, the expression inside the brackets has to be zero.

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{15}$$

## Error covariance of the analysis

Equation (12) can be expanded into:

$$\begin{aligned}
\mathbf{P}^a &= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}^T && (16)\\
&= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T && (17)\\
&= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f && (18)
\end{aligned}$$

where we used the optimal gain from equation (15).

# Optimal Interpolation analysis

- Analysis:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T(\underbrace{\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}}_{\text{covariance of the i.v.}})^{-1} \quad (\underbrace{\mathbf{y}^o - \mathbf{H}\mathbf{x}^f}_{\text{innovation vector}})$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f$$

- For scalars: if we want to combine the temperature predicted by a model $T_m$ ($\sigma_m$) with an observation $T_o$ ($\sigma_o$), the analyzed temperature is:

$$T_a = \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_o^2}\right)^{-1} \left(\frac{T_m}{\sigma_m^2} + \frac{T_o}{\sigma_o^2}\right)$$

$$\sigma_a^2 = \left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_o^2}\right)^{-1}$$

# Example

- Compare behavior of the variable $x_2$ of the model "identity matrix" and "oscillation".
- http://data-assimilation.net/Tools/AssimDemo/?method=OI&model=id
- http://data-assimilation.net/Tools/AssimDemo/?method=OI&model=oscillation
- Describe the behavior of the OI scheme if the error correlation of $x_1$ and $x_2$ is 0.9 for the model "identity matrix".

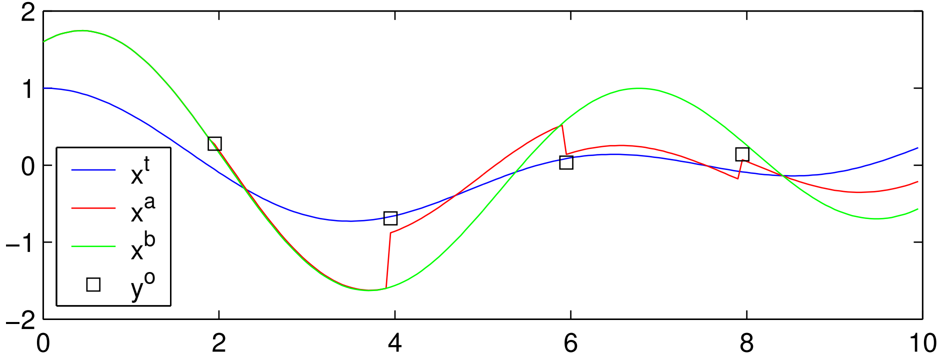


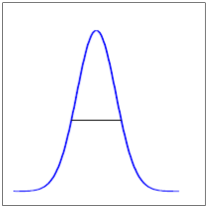

Figure 2: The observed part of a linear system with 4 state variables: the true state vector, $\mathbf{x}^t$, the analysis $\mathbf{x}^a$, the state of the system without assimilation $\mathbf{x}^b$ ($b$, background). The observations $\mathbf{y}^o$ are extracted from $\mathbf{x}^t$. The trajectories $\mathbf{x}^a$ and $\mathbf{x}^b$ start from a wrong initial condition.

# Covariances

- $\mathbf{P} : n \times n (n \approx 10^6)$. $10^{12}$ variables to determine and to store !?
- Constraints: fields are generally “smooth”, close to hydrostatic and geostrophic equilibrium (at sufficiently large scales) and obeying conservation laws,...
- Decomposition of $\mathbf{P}$ in variance $\mathbf{D}$ and correlation $\mathbf{C}$

$$\mathbf{P} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}$$

  - Correlation length $=$ typical spatial scale of the dominant process
  - $\Rightarrow$ “smooth” field



Correlation $\mathbf{C}$

## Reduced rank covariance matrices

- Representation of the covariances by the dominant eigenvectors and eigenvalues:

$$\mathbf{P} = E[\eta\eta^T] \tag{19}$$

$$\mathbf{P} = \mathbf{LDL}^T \qquad \mathbf{L} : n \times r, \;\; \mathbf{D} : r \times r \tag{20}$$

  In general $r \approx 10 - 100$.

- Ensemble representation: $\mathbf{x}^{(k)}, k = 1, \ldots, N$

$$\mathbf{P} =< (\mathbf{x}- < \mathbf{x} >)(\mathbf{x}- < \mathbf{x} >)^T >= \mathbf{XX}^T \qquad <>= \text{ensemble average}$$

  In general slower convergence $(N^{-1/2})$ if $N$ increases. $N \approx 100 - 500$.

- Consequence: The model error $\eta$ and the correction of the state vector $\mathbf{x}_n^a - \mathbf{x}_n^f$ belong to the vector subspace spanned by the columns of $\mathbf{L}$ (or $\mathbf{X}$).

- For the analysis, $\mathbf{P} = \mathbf{LDL}^T$ doesn't have to be formed explicitly

$$\mathbf{K} = \mathbf{L}\left(\mathbf{D}^{-1} + \mathbf{L}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HL}\right)^{-1}\mathbf{L}^T\mathbf{H}^T\mathbf{R}^{-1}$$

- But a reduced-rank covariance introduces an unphysical long-range correlation

## Balanced covariances

- Conservation of *e.g.* salinity: $\int S d^3x =$ *const.*
  Geostrophic equilibrium: $\mathbf{v} = \frac{1}{\rho_0 f}\mathbf{e}_z \times \nabla p_h(T, S, \zeta)$
- General form (linear constraints):

$$\mathbf{Cx} = \textit{const.} \Rightarrow \mathbf{CP} = 0$$

- Example: $\sum_i \mathrm{cov}(S_i, S_j) = 0$
  In this case, the assimilation would not change the total salinity

# 3D-Var

- Minimization of the cost function:

$$J(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^f\right)^T \mathbf{P}^{f^{-1}} \left(\mathbf{x} - \mathbf{x}^f\right) + \left(\mathbf{y}^o - h(\mathbf{x})\right)^T \mathbf{R}^{-1} \left(\mathbf{y}^o - h(\mathbf{x})\right)$$

  using its gradient:

$$\nabla J(\mathbf{x}) = 2\mathbf{P}^{f^{-1}} \left(\mathbf{x} - \mathbf{x}^f\right) - 2\mathbf{H}(\mathbf{x})^T \mathbf{R}^{-1} \left(\mathbf{y}^o - h(\mathbf{x})\right) \quad \text{where } \mathbf{H}_{jm} = \frac{\partial h_m}{\partial x_j}$$

- Minimization: conjugate gradient, Newton-Raphson method,...
- The covariance of the analysis:

$$\mathbf{P}^{a^{-1}} = \frac{1}{2} \nabla_{\mathbf{x}} \nabla_{\mathbf{x}} J \qquad (21)$$

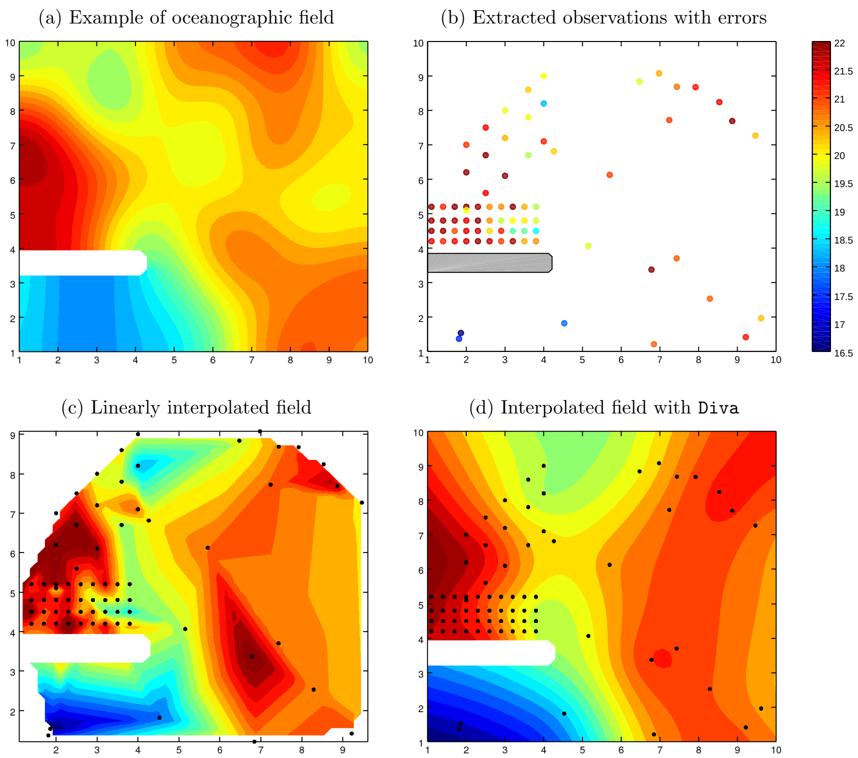
$$= \mathbf{P}^{f^{-1}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \qquad (22)$$

- Generalization of optimal interpolation to non-linear $h$
- No general inversion of $m \times m$ matrices.

- The term $\mathbf{x}^T {\mathbf{P}^f}^{-1} \mathbf{x}$ can be parameterized as “smoothness” constrain:

$$\int_D \alpha_2 \nabla\nabla\varphi : \nabla\nabla\varphi + \alpha_1 \nabla\varphi \cdot \nabla\varphi + \alpha_0 \varphi^2 dD \tag{23}$$

$\mathbf{x}$ is a discretization of the continuous field $\phi$.


(a) Example of oceanographic field
(b) Extracted observations with errors
(c) Linearly interpolated field
(d) Interpolated field with Diva
22
21.5
21
20.5
20
19.5
19
18.5
18
17.5
17
16.5

# Relationship between 3D-var and optimal interpolation

For a linear $\mathbf{H}$,

$$
\begin{aligned}
\frac{1}{2}\nabla J(\mathbf{x}_a) &= 0 && (24)\\
&= \mathbf{P}^{f^{-1}}\left(\mathbf{x}^a - \mathbf{x}^f\right) - \mathbf{H}(\mathbf{x})^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^a\right) && (25)
\end{aligned}
$$

Solving for $\mathbf{x}^a$:

$$
\left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{x}^a = \mathbf{P}^{f^{-1}}\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f + \mathbf{H}\mathbf{x}^f\right) \quad (26)
$$

$$
\mathbf{x}^a = \mathbf{x}^f + \left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \quad (27)
$$

Sherman-Morrison-Woodbury formula:

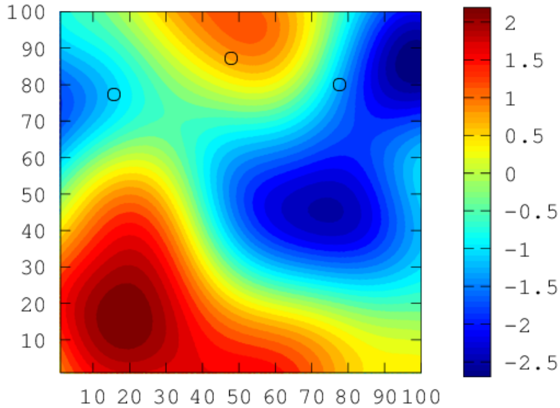
$$
\left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \quad (28)
$$

For a linear observation operator, $\mathbf{H}$, 3D-Var is thus equivalent to optimal interpolation!

# DIVA demo help!

## Reference field

Field: Simple random field




## Location of observations

Your name:

Correlation length-scale:

Choose the location of your observation (maximum 10)

| # | x | y | |
|---|---|---|---|
| 1 | 14.894 | 77.030 | [x] |
| 6 | 77.508 | 79.758 | [x] |
| 7 | 47.416 | 87.030 | [x] |

Make analysis | Remove all

## Top 5

| Name | RMS |
|---|---|
| JMB | 0.2488601797 |
| Ngu | 0.2734995952 |
| charles | 0.3043944967 |
| it | 0.3377809227 |
| Aida | 0.3539947015 |

- http://data-assimilation.net/Tools/divand_demo/html/
- First make some simple test with one and with two observations (one both sides of a gradient), change the correlation length.
- Try to make the "best" analysis with 10 observations at well chosen locations.

# The Kalman filter

- Error propagation through an algebraic expression such like $\rho = \rho(T, S)$:

$$\sigma_\rho^2 = \left(\frac{\partial \rho}{\partial T}\right)^2 \sigma_T^2 + \left(\frac{\partial \rho}{\partial S}\right)^2 \sigma_S^2$$
$$= \begin{pmatrix} \frac{\partial \rho}{\partial T} & \frac{\partial \rho}{\partial S} \end{pmatrix} \begin{pmatrix} \sigma_T^2 & 0 \\ 0 & \sigma_S^2 \end{pmatrix} \begin{pmatrix} \frac{\partial \rho}{\partial T} \\ \frac{\partial \rho}{\partial S} \end{pmatrix}$$

- For a model:

$$\mathbf{P}_{n+1} = \mathbf{M}_n \mathbf{P}_n \mathbf{M}_n^T + \mathbf{Q}_n \quad \text{where } M_{njj'} = \frac{\partial f_{nj}}{\partial x_{j'}}$$

- linear model: Kalman filter
- non-linear model: Extended Kalman filter (for error propagation the model is linearized)

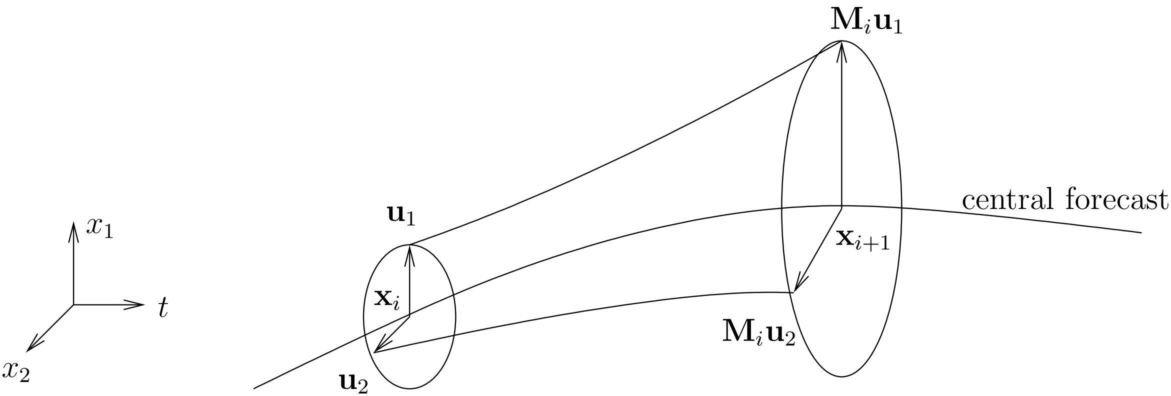



Figure 3: Forecast of the error covariance with the tangent linear model

- Discuss error propagation for $\mathbf{Q} = 0$ and $\mathbf{Q} \neq 0$ for models "identity matrix", and "oscillation" ($\mathbf{P}^i = \mathbf{I}$ and $\mathbf{P}^i = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$).
- Comment on error propagation with Lorenz model.

# (Extended) Kalman filter scheme

Initialization:

$$\mathbf{x}_0^a = \mathbf{x}^i$$
$$\mathbf{P}_0^a = \mathbf{P}^i$$

Forecast:

$$\mathbf{x}_{n+1}^f = f_n(\mathbf{x}_n^a)$$
$$\mathbf{P}_{n+1}^f = \mathbf{M}_n \mathbf{P}_n^a \mathbf{M}_n^T + \mathbf{Q}_n$$

Analysis:

$$\mathbf{x}_{n+1}^a = \mathbf{x}_{n+1}^f + \mathbf{K}_{n+1}\left(\mathbf{y}_{n+1}^o - h_{n+1}(\mathbf{x}_{n+1}^f)\right)$$
$$\mathbf{K}_{n+1} = \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T \left(\mathbf{H}_{n+1}\mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T + \mathbf{R}_{n+1}\right)^{-1}$$
$$\mathbf{P}_{n+1}^a = \mathbf{P}_{n+1}^f - \mathbf{K}_{n+1}\mathbf{H}_{n+1}\mathbf{P}_{n+1}^f$$

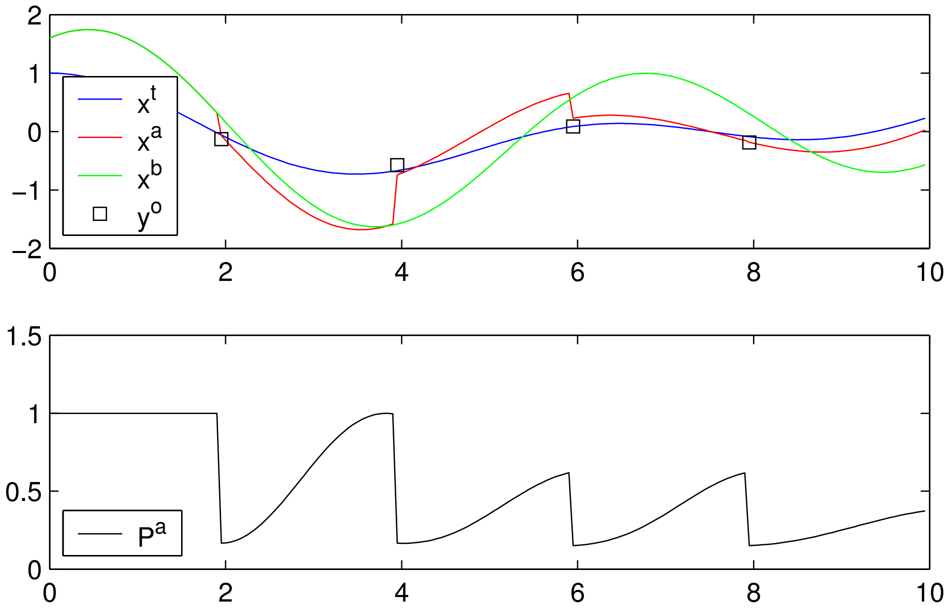



Figure 4: Example of a Kalman filter applied to a linear system. The curves from the first graph correspond to the observed part of the system. The lower panel shows the evolution of the error covariance. The error variance of the state vector is reduced at every assimilation cycle.

# Numerical example: a water column

- Application of the Extended Kalman Filter

- Model represents a water column governed by:

$$\frac{\partial \mathbf{u}}{\partial t} + f\mathbf{e}_z \times \mathbf{u} = \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial \mathbf{u}}{\partial z}\right) \quad (29)$$

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial T}{\partial z}\right) \quad (30)$$

$$\frac{\partial S}{\partial t} = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial S}{\partial z}\right) \quad (31)$$

$$\frac{\partial k}{\partial t} = \tilde{\nu}\left(\frac{\partial \mathbf{u}}{\partial z}\right)^2 - \frac{\tilde{\nu}}{16}k^2 - \tilde{\nu}\frac{\partial b}{\partial z} + \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial k}{\partial z}\right) \quad (32)$$

- The prognostic variables $\mathbf{u}, T, S$ et $k$

- The diagnostic variables: buoyancy $b$, the Richardson number $Ri$ the turbulent diffusibility $\tilde{\nu}$ and $\tilde{\lambda}$:

$$
\begin{aligned}
b(T,S) &= \frac{\rho(T,S) - \rho_0}{\rho_0} && (33)\\
Ri &= \frac{\partial b}{\partial z}\left(\frac{\partial \mathbf{u}}{\partial z}\right)^{-2} && (34)\\
\tilde{\nu} &= \tilde{\nu}(Ri, k) && (35)\\
\tilde{\lambda} &= \tilde{\lambda}(Ri, k) && (36)
\end{aligned}
$$

## Twin experiment

- Pseudo-observations = surface temperature generated by the model + noise
- For the assimilation, the model is started with a different initial condition than the model run that generated the observations
- Water column of 100 m depth and 30 vertical levels

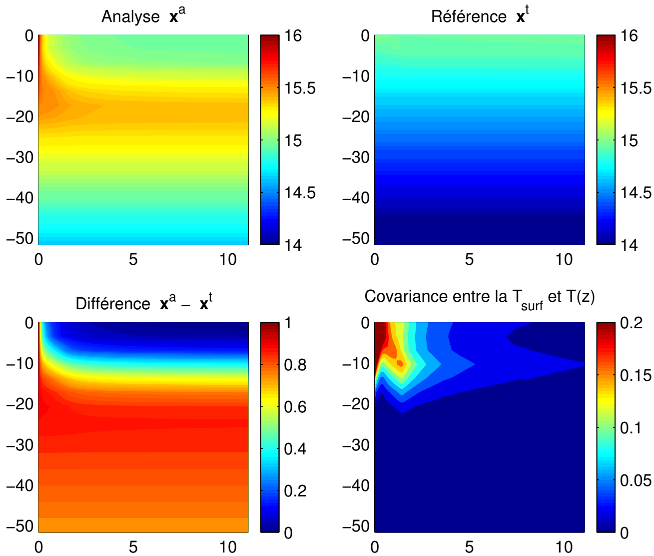



Figure 5: Temperature as a function of time (hours) and depth. Only the upper 50 meters are shown.

# Applications outside oceanography

First applied to the trajectory estimation for the Apollo program

- Attitude and Heading Reference Systems
- Autopilot
- Battery state of charge estimation
- Brain-computer interface
- Chaotic signals
- Tracking of objects in computer vision
- Dynamic positioning
- Economics, in particular macroeconomics, time series, and econometrics
- Inertial guidance system
- Orbit Determination
- Radar tracker
- Satellite navigation systems
- Seismology
- Sensorless control of AC motor variable-frequency drives
- Simultaneous localization and mapping
- Speech enhancement
- Weather forecasting
- Structural health monitoring

## Some approximations of the Kalman filter

- Pham *et al.* (1998); Pham (2001). Evolutive error space
- Evensen (1994, 2007): Ensemble Kalman filter
- RRSQRT: reduced rank approximation of the square root filter (reformulation of the Kalman filter)
- Ensemble Transform Kalman Filter (Bishop *et al.*, 2001), Ensemble Adjustment Kalman Filter (Anderson, 2001),...

## Kalman Filter Demonstration

### No time variation

Test to carry out:

1. Only the first variable $x_1$ is observed, $\mathbf{P}^i = \mathbf{I}$, $\mathbf{R} = 0.2$ and no model noise $\mathbf{Q} = 0$ is assumed. Explain the behavior of $x_1$, $x_2$ in time and their error covariance matrix.
2. How to change the previous setup, to increase the rate of convergence of $x_1$ to the true state?

3. Use default values, except assuming that initially $x_1$ and $x_2$ are perfectly correlated. Explain the behavior of $x_2$.

4. Use default values, except assuming that $\mathbf{Q} = 0.1\mathbf{I}$ ("random walk"). Discuss first the free run (state vector and its error covariance/error standard deviation) and then the results with assimilation.

## 1D advection in periodic domain

1. Using the default value, explain the behavior of the observed variables $x_1$ and $x_3$ (and their error covariance). Why do the non-observed variables get corrected too?

2. Using the default values, except reducing the model time step between observations from 5 to 4. We increase the frequency of assimilation, yet no variable converges anymore. Why? Can this happen in oceanography? Think of an example.

3. Use default values, except assuming that $\mathbf{Q} = 0.1\mathbf{I}$. How could you use the error covariance of the results with assimilation to justify the use of optimal interpolation?

## Oscillations

1. Using the default values, why does the error covariance $\mathbf{P}$ remains equal to the identity matrix of the free run?

2. What different changes to the default values are necessary to make the run with assimilation converge to the true solution?

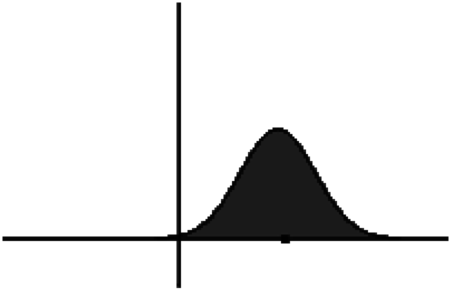
3. Discuss the correction by data assimilation of the variables $x_1$ and $x_2$ (not directly observed).

# Propagation of uncertainty - Non-Gaussian errors

- The probability density $p(\mathbf{x},t)$ for the random vector $\mathbf{X}_t$ satisfies the Fokker-Planck equation

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = \underbrace{-\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[f_i(\mathbf{x})p(\mathbf{x},t)\right]}_{\text{"advection"}} + \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2}{\partial x_i\,\partial x_j}\left[Q_{ij}(\mathbf{x})p(\mathbf{x},t)\right]}_{\text{"diffusion"}}$$

- $\boldsymbol{\eta}_n$ is assumed to be normally distributed $N(0,\mathbf{Q})$

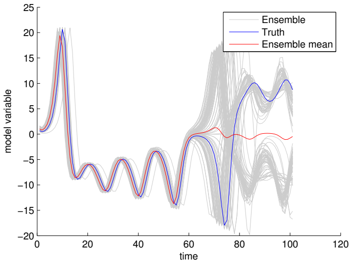


# Propagation of uncertainty - Non-Gaussian errors

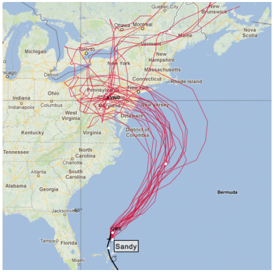
- Even if the model is non-linear, the Fokker-Planck equation is linear (not always non-chaotic)!
- Impossible to solve for large geophysical problems
- If every dimension of $\mathbf{x}$ would be discretized with 100 grid points, then pdf $p$ would represent $100^n$ numbers.
- Equation is similar to an advection-diffusion dimension in fluid dynamics (however in a very high dimensional space)
- If this equation represents the Eulerian view, what would be the equivalent Lagrangian view?

# Ensemble simulation

- Lagrangian approach to the Fokker Planck simulation $\rightarrow$ ensemble simulation
- In an ensemble simulation, a model is run a large number of times with different forcings, initial condition, parametrization,... within the uncertainty limit of the perturbed variable
- The spread of the ensemble reflects the resulting uncertainty in the model results




Ensemble simulation of a Lorenz model




Ensemble simulation for tracks of Hurricane Sandy by NCEP

## Ensemble Kalman filter

From the ensemble of forecast states $\mathbf{x}^{f^{(k)}}$ where $k = 1, \ldots, N$ one can compute the ensemble mean

$$\overline{\mathbf{x}^f} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}^{f^{(k)}} \tag{37}$$

and ensemble covariance:

$$\mathbf{P}^f = \frac{1}{N-1} \sum_{k=1}^{N} \left( \mathbf{x}^{f^{(k)}} - \overline{\mathbf{x}^f} \right) \left( \mathbf{x}^{f^{(k)}} - \overline{\mathbf{x}^f} \right)^T \tag{38}$$

We construct the columns of the matrix $\mathbf{S}^f$ by:

$$\left( \mathbf{S}^f \right)_k = \frac{\mathbf{x}^{f^{(k)}} - \overline{\mathbf{x}^f}}{\sqrt{N-1}} \tag{39}$$

where $\mathbf{S}^f$ is a $n \times N$ matrix, which each column being the difference between each member and its ensemble mean. Its mean over all columns it thus zero. As many other assimilation schemes (SEEK, RRSQRT, ESSE, EnKF), $\mathbf{P}^f$ is decomposed in terms of this square root matrix $\mathbf{S}^f$:

$$\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{f^T} \tag{40}$$

Typically, the number of ensemble members $N$ is much smaller than the state vector size $n$. We rewrite the Kalman Filter analysis, by avoiding any matrix of the size $n \times n$:

$$\begin{aligned}
\mathbf{K} &= (\mathbf{S}^f {\mathbf{S}^f}^T)\mathbf{H}^T \left[\mathbf{H}(\mathbf{S}^f {\mathbf{S}^f}^T)\mathbf{H}^T + \mathbf{R}\right]^{-1} && (41)\\
&= \mathbf{S}^f (\mathbf{H}\mathbf{S}^f)^T \left[\mathbf{H}\mathbf{S}^f (\mathbf{H}\mathbf{S}^f)^T + \mathbf{R}\right]^{-1} && (42)\\
&= \mathbf{S}^f \left[\mathbf{I} + (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}^f\right]^{-1} (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (43)
\end{aligned}$$

Where the Sherman-Morison-Woodbury identity has been applied from (42) to (43). This identity can be expressed as:

$$\mathbf{A}\mathbf{B}^T\left(\mathbf{C} + \mathbf{B}\mathbf{A}\mathbf{B}^T\right)^{-1} = \left(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{C}^{-1} \qquad (44)$$

with $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = \mathbf{H}\mathbf{S}^f$, $\mathbf{C} = \mathbf{R}$. That is, instead of performing the inverse in space of matrix $\mathbf{A}$ the inverse is done in the space of the matrix $\mathbf{C}$. We also substitute $\mathbf{P}^f$ in the expression of the analysis covariance error $\mathbf{P}^a$ by its square root matrices:

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f \tag{45}$$

$$= \mathbf{S}^f \mathbf{S}^{fT} - \mathbf{KHS}^f \mathbf{S}^{fT} \tag{46}$$

$$= \mathbf{S}^f \mathbf{S}^{fT} - \mathbf{S}^f \left[\mathbf{I} + (\mathbf{HS}^f)^T \mathbf{R}^{-1} \mathbf{HS}^f\right]^{-1} (\mathbf{HS}^f)^T \mathbf{R}^{-1} \mathbf{HSS}^{fT} \tag{47}$$

$$= \mathbf{S}^f \left[\mathbf{I} - \left(\mathbf{I} + (\mathbf{HS}^f)^T \mathbf{R}^{-1} \mathbf{HS}^f\right)^{-1} (\mathbf{HS}^f)^T \mathbf{R}^{-1} \mathbf{HS}\right] \mathbf{S}^{fT} \tag{48}$$

In order to avoid to form $\mathbf{P}^a$ explicitly, we need to express $\mathbf{P}^a$ also in terms of the square root matrices $\mathbf{S}^a$.

$$\mathbf{P}^a = \mathbf{S}^a \mathbf{S}^{aT} \tag{49}$$

This is possible when the following eigenvalue decomposition is made :

$$\left(\mathbf{HS}^f\right)^T \mathbf{R}^{-1} \mathbf{HS}^f = \mathbf{U \Lambda U}^T \tag{50}$$

where $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ and where $\mathbf{\Lambda}$ is diagonal. $\mathbf{U}$ and $\mathbf{\Lambda}$ are both of size $r \times r$.

Using the decomposition (50) in equation (48) one obtains:

$$
\begin{aligned}
\mathbf{P}^a &= \mathbf{S}^f \left[\mathbf{I} - (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1}\mathbf{U\Lambda U}^T\right] \mathbf{S}^{fT} & (51)\\
&= \mathbf{S}^f \left[\mathbf{I} - (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1} \left(\mathbf{U\Lambda U}^T + \mathbf{I} - \mathbf{I}\right)\right] \mathbf{S}^{fT} & (52)\\
&= \mathbf{S}^f \left[\mathbf{I} - (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1} \left(\mathbf{U\Lambda U}^T + \mathbf{I}\right) + (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1}\right] \mathbf{S}^{fT} & (53)\\
&= \mathbf{S}^f \left[\mathbf{I} - \mathbf{I} + (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1}\right] \mathbf{S}^{fT} & (54)\\
&= \mathbf{S}^f (\mathbf{I} + \mathbf{U\Lambda U}^T)^{-1} \mathbf{S}^{fT} & (55)\\
&= \mathbf{S}^f (\mathbf{UU}^T + \mathbf{U\Lambda U}^T)^{-1} \mathbf{S}^{fT} & (56)\\
&= \mathbf{S}^f \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{U}^T \mathbf{S}^{fT} & (57)\\
&= \mathbf{S}^f \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} (\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{U}^T \mathbf{S}^{fT} & (58)
\end{aligned}
$$

So we found a square root decomposition of $\mathbf{P}^a$ in terms of $\mathbf{S}^f \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1/2}$. But in order to construct an ensemble from the columns of $\mathbf{S}^a$, its mean has to be zero. Therefore we multiply it by $\mathbf{U}^T$ (which does not change the product $\mathbf{S}^a \mathbf{S}^{aT}$):

$$
\mathbf{S}^a = \mathbf{S}^f \mathbf{U}(1 + \mathbf{\Lambda})^{-1/2} \mathbf{U}^T \tag{59}
$$

The decomposition (50) can also be used in the computation of the Kalman gain $\mathbf{K}$ by:

$$
\begin{aligned}
\mathbf{K} &= \mathbf{S}^f \left[\mathbf{I} + (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}^f\right]^{-1} (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (60) \\
&= \mathbf{S}^f \left[\mathbf{U}\mathbf{U}^T + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\right]^{-1} (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (61) \\
&= \mathbf{S}^f \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{U}^T (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (62)
\end{aligned}
$$

The ensemble after the analysis will have the following mean:

$$
\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}^f}\right) \tag{63}
$$

Based on the mean $\overline{\mathbf{x}^a}$ and the columns of $\mathbf{S}^a$, an ensemble can be reconstructed:

$$
\mathbf{x}^{a(k)} = \mathbf{x}^a + \sqrt{N-1}\ (\mathbf{S}^a)_k \tag{64}
$$

# Exercise

- Compare the results of the linear models using the Extended Kalman Filter and the Ensemble Kalman Filter.
- Compare the results of the Lorenz 1963 model using the Extended Kalman Filter and the Ensemble Kalman Filter.

# Particle filter

## Bayes Theorem

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)} \tag{65}$$

- $p(\mathbf{x}|\mathbf{y}^o)$: a posteriori pdf, pdf of the model state $\mathbf{x}$ given the observations $\mathbf{y}^o$.
- $p(\mathbf{x})$: a priori pdf, pdf of the model state $\mathbf{x}$ before knowing the observations $\mathbf{y}^o$.
- $p(\mathbf{y}^o|\mathbf{x})$: probability of a measurement $\mathbf{y}^o$ if the system is in the state $\mathbf{x}$. For Gaussian observations errors:

$$p(\mathbf{y}^o|\mathbf{x}) = A \exp\left((\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x}))\right) \tag{66}$$

- $p(\mathbf{y}^o)$: The denominator is just a normalization to ensure that the pdf integrates to one.

## Particle filter

The model pdf is represented by an ensemble (or by particles) $\mathbf{x}^{(k)}$ $(k = 1, \ldots, N)$:

$$p(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^{K} \delta(\mathbf{x} - \mathbf{x}^{(k)}) \tag{67}$$

Initially all particles are equally probable, but by comparison to the observations, the particles who are closer to the observations are more likely than the particles who a further away from the observations.

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{1}{N} \sum_{k=1}^{N} w_k \delta(\mathbf{x} - \mathbf{x}^{(k)}) \tag{68}$$

where the weights are given by: $w_k = \frac{p(\mathbf{y}^o|\mathbf{x}^{(k)})}{\sum_{k=1}^{N} p(\mathbf{y}^o|\mathbf{x}^{(k)})}$

- **Re-sampling**: Particles with very low probability are ignored and particles with high probability are duplicated.
- No Gaussian assumption of the model error is necessary.
- **Curse of dimensionality**: Large number of particles are needed for high-dimensional problems.

# Sangoma diagnostics

- Sangoma project: http://data-assimilation.net/
- Provides several diagnostics and utilities mainly related to ensemble-based data assimilation:
  - Ensemble rank histograms, mutual information, relative entropy
  - Ensemble sensitivity of posterior mean to observations in a particle filter
  - Array modes and associated quantities
  - Brier skill score, CRPS, RCRV
  - Spatially correlated ensemble perturbations
  - Perturbation based on EOFs
  - Weakly constrained ensemble perturbations (ensemble perturbations that have to satisfy an a priori linear constraint)
  - Dominant POD modes from an ensemble of snapshots.
  - Empirical Gaussian Anamorphosis (the empirical transformation function such that a transformed variable follows a Gaussian distribution)
  - Observation operator for HF radar surface currents

# Non-sequential assimilation

- Strong constraints ($\mathbf{Q}_n = 0$). 4D-Var, adjoint methods
- Weak constraints ($\mathbf{Q}_n \neq 0$). Representer method

# 4D-Var

- Minimization of the following cost function:

$$\begin{aligned} J(\mathbf{x}_0) &= \left(\mathbf{x}_0 - \mathbf{x}^i\right)^T \mathbf{P}^{i^{-1}} \left(\mathbf{x}_0 - \mathbf{x}^i\right) \\ &+ \sum_{n=1}^{N} \left(\mathbf{y}_n^o - h_n(\mathbf{x}_n)\right)^T \mathbf{R}_n^{-1} \left(\mathbf{y}_n^o - h_n(\mathbf{x}_n)\right) \end{aligned}$$

with $\mathbf{x}_{n+1} = f_n(\mathbf{x}_n)$.

- The constrain is introduced in the cost function with the Lagrangian multiplier

# 4D-Var

- Gradient of the cost function:

$$\nabla_{\mathbf{x}_0} J = 2\,\mathbf{P}^{i^{-1}}\left(\mathbf{x}_0 - \mathbf{x}^i\right) - 2\,\mathbf{M}_0^T \boldsymbol{\lambda}_0$$

  is calculated using the adjoint variable $\boldsymbol{\lambda}_n$:

$$\begin{aligned} \mathbf{x}_{n+1} &= f_n(\mathbf{x}_n) \\ \boldsymbol{\lambda}_{n-1} &= \mathbf{M}_n^T \boldsymbol{\lambda}_n + \mathbf{H}_n^T \mathbf{R}_n^{-1}\left(\mathbf{y}_n^o - h_n(\mathbf{x}_n)\right) \\ \boldsymbol{\lambda}_N &= 0 \end{aligned}$$

- The adjoint model is integrated backwards in time!

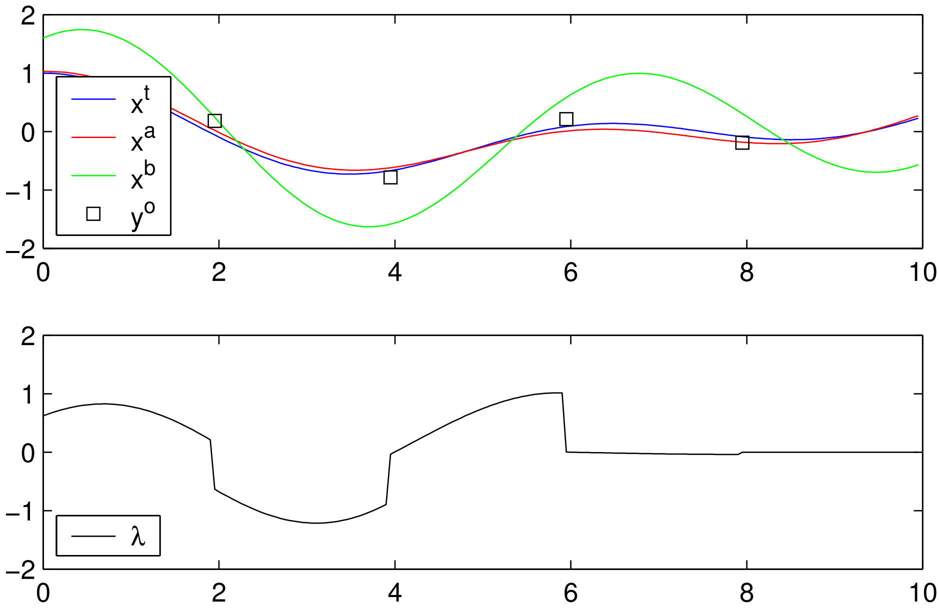



Figure 6: 4D-Var. The cost function is explicitly minimized; Upper panel: the observed component of the state vector $\mathbf{x}^t$ (truth), $\mathbf{x}^a$ (with assimilation) and $\mathbf{x}^b$ (without assimilation). After 20 iterations, the solution $\mathbf{x}^a$ is already quite close to the real trajectory $\mathbf{x}^t$. Lower panel: The adjoint variable $\boldsymbol{\lambda}$ corresponding to the observed part of the state vector.

# 4D-Var

Init. Cond., bound. Cond., parameters $(\mathbf{u})$ → Models $f_n$ and $h_n$ → Observable model predictions $(\mathbf{v})$

$$\mathbf{v} = g(\mathbf{u})$$

A perturbation on the inputs $\delta\mathbf{u}$ is linked to the perturbations on the outputs $\delta\mathbf{v}$ by:

$$\delta\mathbf{v} = \mathbf{G}\ \delta\mathbf{u} \quad \text{with} \quad \mathbf{G}_{ij} = \frac{\partial \mathbf{v}_i}{\partial \mathbf{u}_j}$$

# 4D-Var

Cost function:

$$J(\mathbf{v}) = J\left[g(\mathbf{u})\right]$$

The sensitivity of $J$ relative to $\mathbf{u}$ is obtained by the gradient of the cost function $J$:

$$\nabla_{\mathbf{u}} J = \mathbf{G}^T \, \nabla_{\mathbf{v}} J$$

For a time integration, one has:

$$\begin{aligned} g &= g_N \circ \ldots g_2 \circ g_1 \\ \mathbf{G} &= \mathbf{G}_N \ldots \mathbf{G}_2 \mathbf{G}_1 \\ \mathbf{G}^T &= \mathbf{G}_1^T \mathbf{G}_2^T \ldots \mathbf{G}_N^T \end{aligned}$$

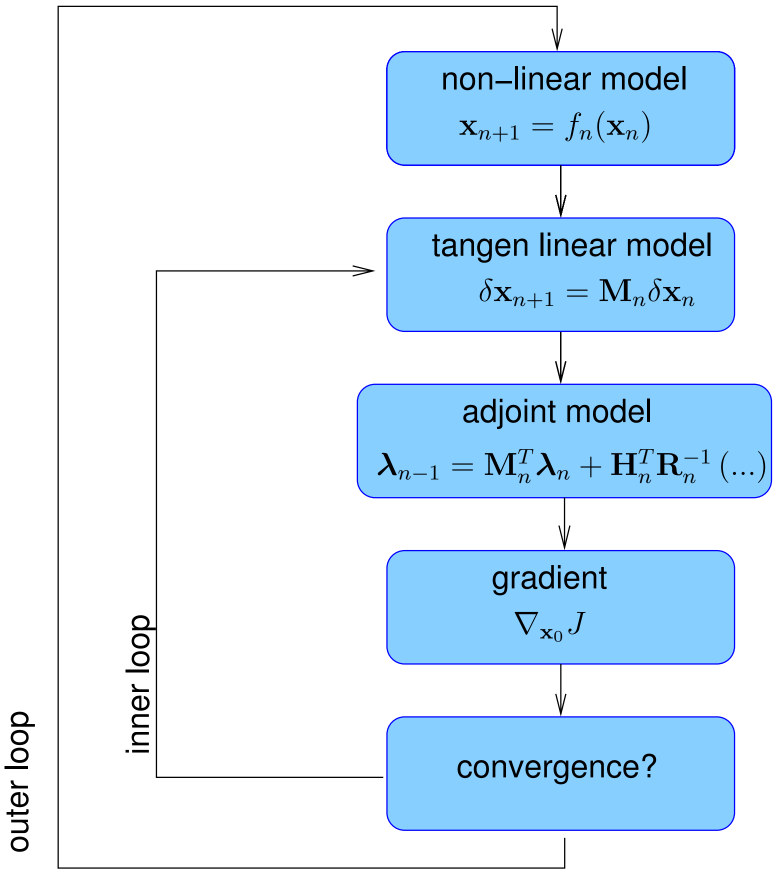
# Incremental formulation

- Efficient algorithm to minimize a quadratic function
- Model and observation operators are linearized around first guess of the model trajectory $\rightarrow$ incremental formulation:

$$
\begin{aligned}
J(\delta\mathbf{x}_0) &= \left(\mathbf{x}_0 + \delta\mathbf{x}_0 - \mathbf{x}^i\right)^T \mathbf{P}^{i-1} \left(\mathbf{x}_0 + \delta\mathbf{x}_0 - \mathbf{x}^i\right) \\
&\quad + \sum_{n=1}^{N} \left(\mathbf{y}_n^o - h_n(\mathbf{x}_n) - \mathbf{H}_n\delta\mathbf{x}_n\right)^T \mathbf{R}_n^{-1} \left(\mathbf{y}_n^o - h_n(\mathbf{x}_n) - \mathbf{H}_n\delta\mathbf{x}_n\right)
\end{aligned}
$$

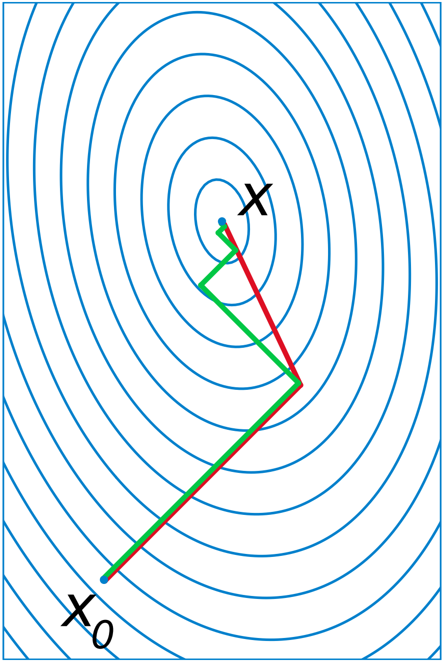
with $\mathbf{x}_{n+1} = f_n(\mathbf{x}_n)$ and $\delta x_{n+1} = \mathbf{M}_n \delta x_n$.

- Minimize this function using the conjugate gradient method (inner loops)
- After the minimum is reached, a new model trajectory is computed with the full non-linear model
- The model and observation operator are linearized around this new trajectory and the whole procedure is repeated (outer loops)



non–linear model
$\mathbf{x}_{n+1} = f_n(\mathbf{x}_n)$
tangen linear model
$\delta\mathbf{x}_{n+1} = \mathbf{M}_n \delta\mathbf{x}_n$
adjoint model
$\boldsymbol{\lambda}_{n-1} = \mathbf{M}_n^T \boldsymbol{\lambda}_n + \mathbf{H}_n^T \mathbf{R}_n^{-1} (\ldots)$
gradient
$\nabla_{\mathbf{x}_0} J$
convergence?
inner loop
outer loop

# Conjugate gradient method




- Minimizing $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{b}$ is equivalent to solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$.
- $\mathbf{A}$ is a symmetric and positive defined matrix.
- All search directions $\mathbf{p}_i$ are "conjugate" ($\mathbf{p}_i^T\mathbf{A}\mathbf{p}_j = 0$ if $i \neq j$).
- Conjugate gradient method converges faster than the steepest descent method.

# Derivation of tangent linear

- Model can be broken down to a series instructions $f^{(p)}$ where every instruction corresponds to a line of code

$$f(\mathbf{x}) = f^{(p)}(\dots f^{(2)}(f^{(1)}(\mathbf{x}))) \tag{69}$$

- By applying the chain-rule, the tangent linear of $f$ is:

$$\mathbf{F} = \mathbf{F}^{(p)} \dots \mathbf{F}^{(2)} \mathbf{F}^{(1)} \tag{70}$$

where $\mathbf{F}_{ij} = \frac{\partial f_i}{\partial x_j}$ and $\mathbf{F}^{(p)}_{ij} = \frac{\partial f^{(p)}_i}{\partial x_j}$

- Let's consider a simple statement

$$d = ab + c \tag{71}$$

- This statement can be seen as a function $f$ with input $a$, $b$ and $c$.
- The tangent linear code is obtained by differentiation of $f$:

$$\delta f = \frac{\partial f}{\partial a}\delta a + \frac{\partial f}{\partial b}\delta b + \frac{\partial f}{\partial c}\delta c \tag{72}$$

- For the example statement, one obtains:

$$\delta d = b\delta a + a\delta b + \delta c \tag{73}$$

# Derivation of adjoint

- The adjoint is the transpose of the tangent linear model

$$\mathbf{F}^T = \mathbf{F}^{(1)^T} \mathbf{F}^{(2)^T} \dots \mathbf{F}^{(p)^T} \tag{74}$$

- The example statement can also be written in matrix form:

$$\begin{pmatrix} \delta a \\ \delta b \\ \delta c \\ \delta d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ b & a & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta a \\ \delta b \\ \delta c \\ \delta d \end{pmatrix}$$

- The adjoint variables $\delta a^*$ are governed by the transpose of this matrix:

$$\begin{pmatrix} \delta a^* \\ \delta b^* \\ \delta c^* \\ \delta d^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & b \\ 0 & 1 & 0 & a \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \delta a^* \\ \delta b^* \\ \delta c^* \\ \delta d^* \end{pmatrix}$$

or

$$
\begin{aligned}
\delta a^* &= \delta a^* + b\,\delta d^* \\
\delta b^* &= \delta b^* + a\,\delta d^* \\
\delta c^* &= \delta c^* + \delta d^* \\
\delta d^* &= 0
\end{aligned}
$$

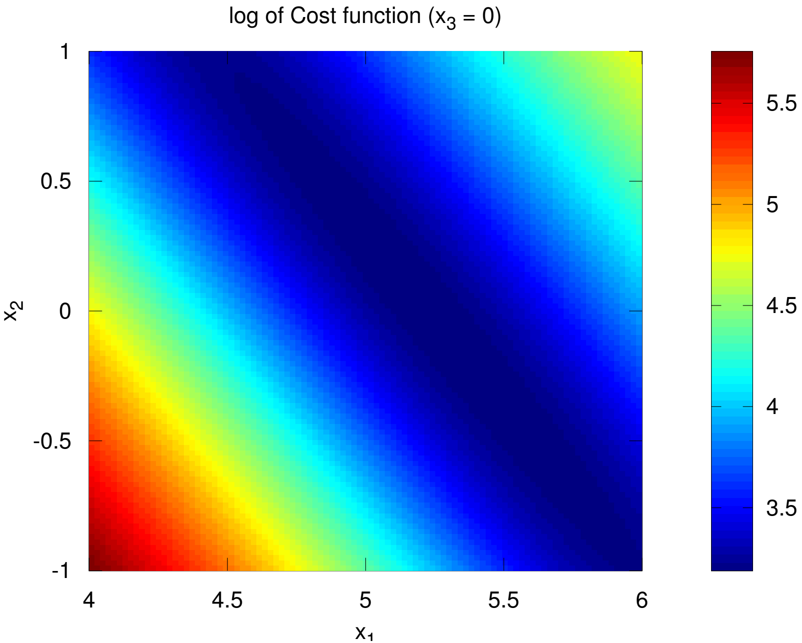
- For an adjoint integration one needs to have the state of the non-linear forward model
- All loops are reversed
- Automatic adjoint generators exist (TAMC, OpenAD, ODYSSEE)

# Demo

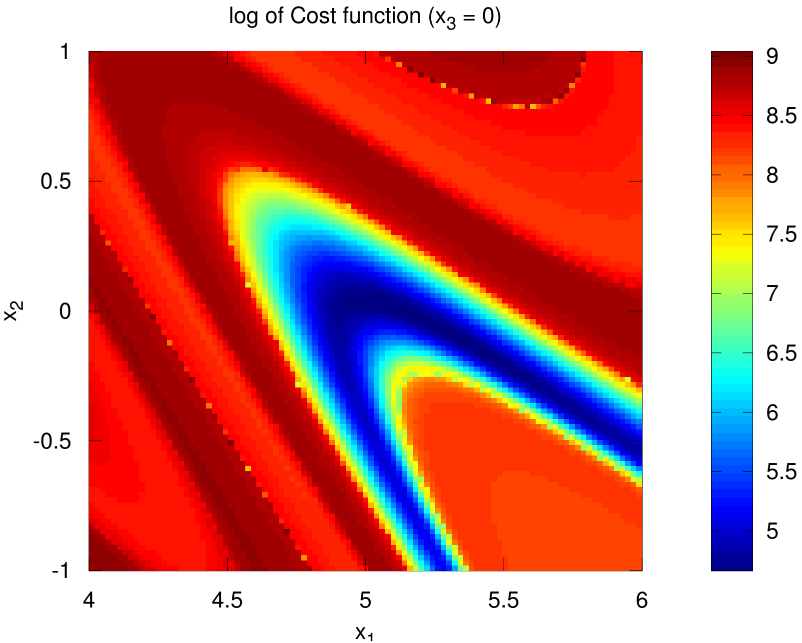
- Use the 4DVar method with the linear model and compare the results to the Kalman Filter
- Use the Lorenz model by varying number of time steps
- http://data-assimilation.net/Tools/AssimDemo/

# Cost function of the Lorenz Model

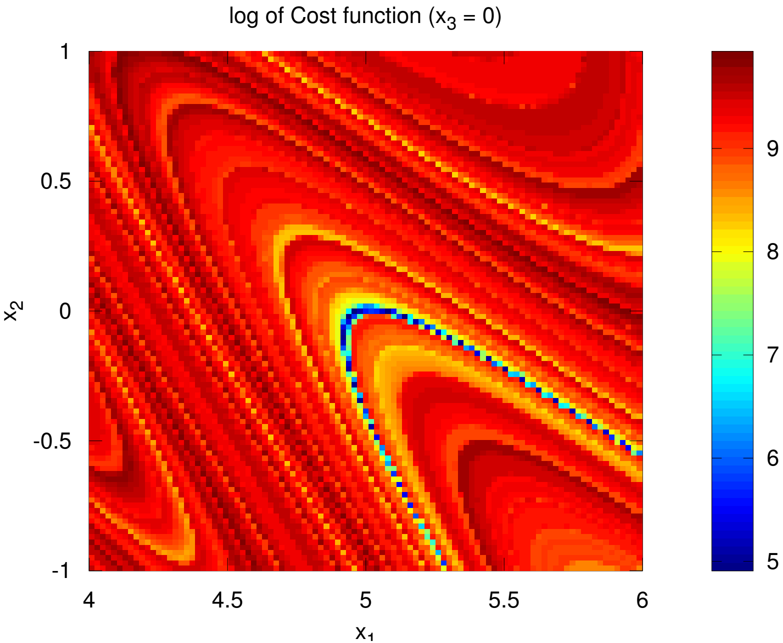
- Long assimilation window: more observations, but complex cost function
- Short assimilation window: less observations, but easier to minimize cost function
- Test: assimilate every model time step, true initial condition is $\mathbf{x} = (5, 0, 0)$
- Error variance of initial condition and observations is 1.




Lorenz model (with 20 time steps)




Lorenz model (with 100 time steps)




Lorenz model (with 150 time steps)

# The representer method

- Hypothesis: $f_n$ and $h_n$ are linear
- Cost function:

$$
\begin{aligned}
J(\mathbf{x}_0, \dots, \mathbf{x}_N) = & \left(\mathbf{x}_0 - \mathbf{x}^i\right)^T \mathbf{P}^{i^{-1}} \left(\mathbf{x}_0 - \mathbf{x}^i\right) \\
& + \sum_{n=1}^{N} \left(\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n\right)^T \mathbf{R}_n{}^{-1} \left(\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n\right) \\
& + \sum_{n=0}^{N-1} \left(\mathbf{x}_{n+1} - \mathbf{M}_n \mathbf{x}_n - \mathbf{F}_n\right) \mathbf{Q}_n{}^{-1} \left(\mathbf{x}_{n+1} - \mathbf{M}_n \mathbf{x}_n - \mathbf{F}_n\right)
\end{aligned}
$$

# The representer method

**First guess $\mathrm{x}_n^b$**

$$\mathbf{x}_{n+1}^b = \mathrm{M}_n \mathbf{x}_n^b + \mathbf{F}_n$$
$$\mathbf{x}_0 = \mathbf{x}^i$$

**Adjoint of the representers $\boldsymbol{\Lambda}_{nn'}$**

$$\boldsymbol{\Lambda}_{n-1n'} = \mathrm{M}_n^T \boldsymbol{\Lambda}_{nn'} + \mathbf{H}_n^T \delta_{nn'}$$
$$\boldsymbol{\Lambda}_{Nn'} = 0$$

**Representers $\tilde{\mathrm{R}}_{nn'}$**

$$\tilde{\mathrm{R}}_{n+1n'} = \mathrm{M}_n \tilde{\mathrm{R}}_{nn'} + \mathbf{Q}_n \boldsymbol{\Lambda}_{nn'}$$
$$\tilde{\mathrm{R}}_{0n'} = \mathbf{P}^i \mathrm{M}_0^T \boldsymbol{\Lambda}_{0n'}$$

**Corrections $\mathrm{b}_n$**

$$\underline{\mathrm{b}} = \left(\underline{\mathbf{R}} + \underline{\mathbf{H}\tilde{\mathbf{R}}}\right)^{-1} \left(\underline{\mathbf{y}}^o - \underline{\mathbf{H}\mathbf{x}^b}\right)$$

$$\underline{\mathbf{y}}^{o\,T} = (\mathbf{y}_1^{o\,T}, \ldots, \mathbf{y}_N^{o\,T})$$
$$\underline{\mathbf{H}\mathbf{x}}^{b\,T} = (\mathbf{H}_1 \mathbf{x}_1^{b\,T}, \ldots, \mathbf{H}_N \mathbf{x}_N^{b\,T})$$
$$\underline{\mathbf{b}}^T = (\mathbf{b}_1^T, \ldots, \mathbf{b}_N^T)$$
$$\underline{\mathbf{R}} = \mathrm{diag}\,(\mathbf{R}_1, \ldots, \mathbf{R}_N)$$
$$\underline{\mathbf{H}\tilde{\mathbf{R}}} = \begin{pmatrix} \mathbf{H}_1 \tilde{\mathbf{R}}_{11} & \cdots & \mathbf{H}_N \tilde{\mathbf{R}}_{N1} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_1 \tilde{\mathbf{R}}_{1N} & \cdots & \mathbf{H}_N \tilde{\mathbf{R}}_{NN} \end{pmatrix}$$

**Analysis $\mathrm{x}_n$**

$$\mathbf{x}_{n+1} = \mathrm{M}_n \mathbf{x}_n + \mathbf{F}_n + \mathbf{Q}_n \boldsymbol{\lambda}_n$$
$$\mathbf{x}_0 = \mathbf{x}^i + \mathbf{P}^i \mathrm{M}_0^T \boldsymbol{\lambda}_0$$
$$\boldsymbol{\lambda}_{n-1} = \mathrm{M}_n^T \boldsymbol{\lambda}_n + \mathbf{H}_n^T \mathrm{b}_n$$
$$\boldsymbol{\lambda}_N = 0$$

## Interpretation

- The representers are covariances:

$$\tilde{\mathbf{R}}_{nn'} = E[(\mathbf{x}_n^b - \mathbf{x}_n^t)(\mathbf{H}_{n'}\mathbf{x}_{n'}^b - \mathbf{H}_{n'}\mathbf{x}_{n'}^t)^T]$$

- Analysis with the representer method $=$ optimal interpolation with the time coordinate included in the state vector

- $nm+2$ integrations with numerical model $nm+1$ integration with the adjoint model.

- Method becomes prohibitive if $m$ is large (satellite data)

# Application to assimilation of HF Radar currents / West Florida Shelf

# Observations

- HF radar radial surface currents maps (CODAR) detided
- 2-day averaged
- error variance estimate provided by instrument is used

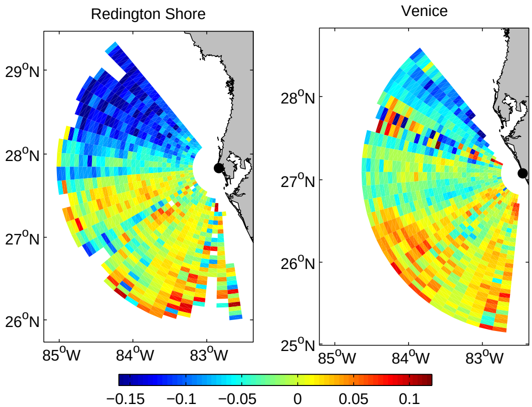



Figure 7: Radial velocities measured from the Redington and Venice sites (West Florida Shelf) on December 9, 2005. Positive values represent a current towards the antenna.

# State vector

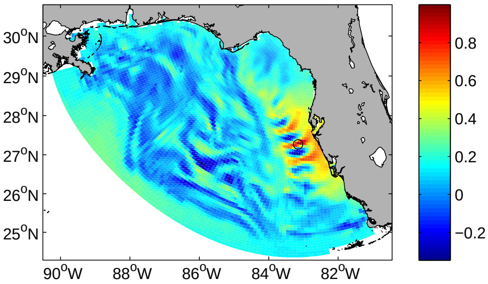
The state vector includes:

- elevation
- horizontal velocity
- temperature and salinity
- 2-day averaged wind stress

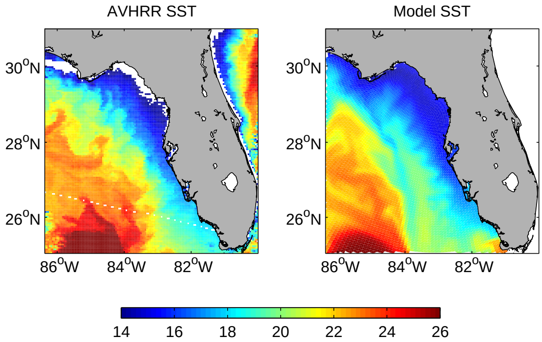
# Model error covariance

- The error covariance defines how two different variables at different locations are related
- 100-member ensemble of wind fields
  - EOF analysis of the u and v wind components
  - random perturbations proportional to spatial EOFs
- For each wind field, the WFS ROMS model was integrated for 30 days

- The resulting ensemble was used for the assimilation of HF Radar currents
- Error covariance assumed constant in time ⇒ “OI-approximation”.




(a) Correlation between the u-velocity at a specific location marked by the circle and the u-velocity at all other model grid points.

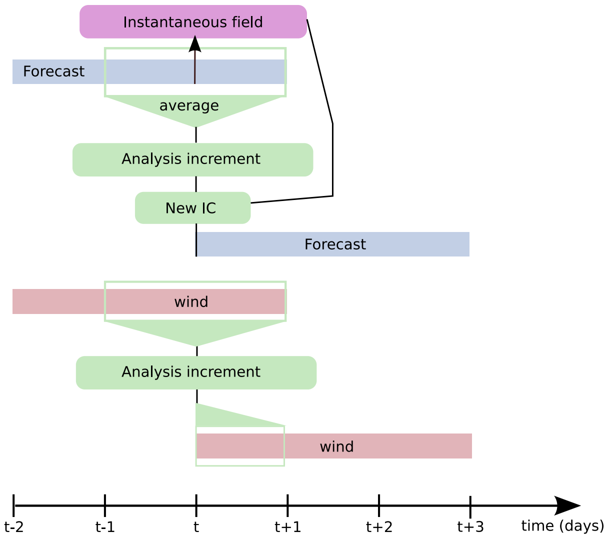



(b) AVHRR SST and model SST on January 29, 2004

- The velocity error covariance on the shelf is closely related to the presence of the meandering front on the shelf. The covariance structure is a superposition of various ensemble members with different phase.

# Sequential algorithm

- Data is assimilated every 2 days
- Model is started at t-2 and run for 3 days
- Currents are averaged over t-1 and t+1
- Wind stress is also averaged over t-1 and t+1
- Analysis increment is computed based on the model error covariance expressed as an ensemble
- This correction is added to the instantaneous model field at t to produce a new initial condition (IC)
- The wind stress correction is applied uniformly to the wind forcing between t and t+1



# Comparison with independent observations

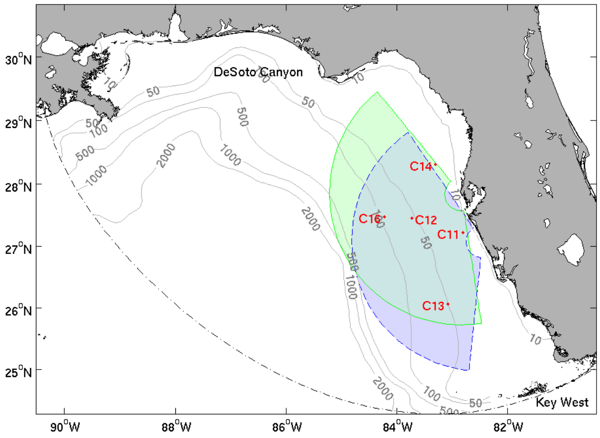
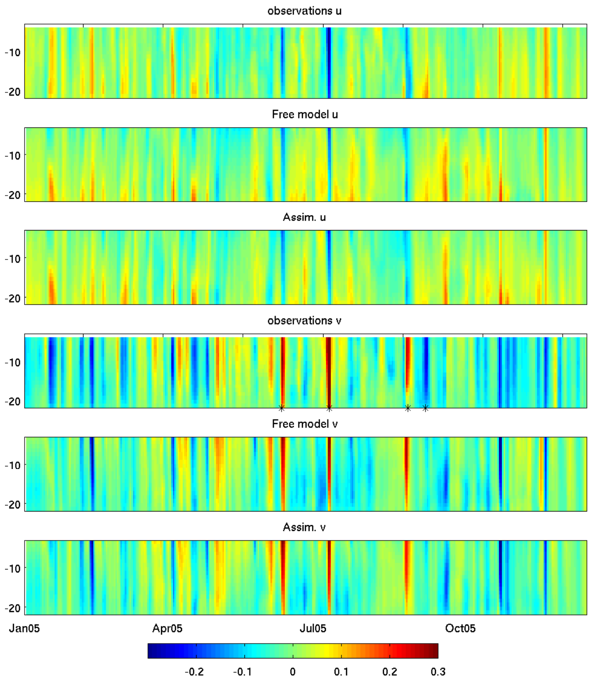
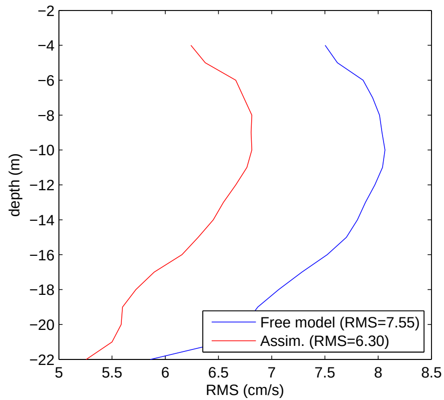



Figure 8: Location of ADCP sites

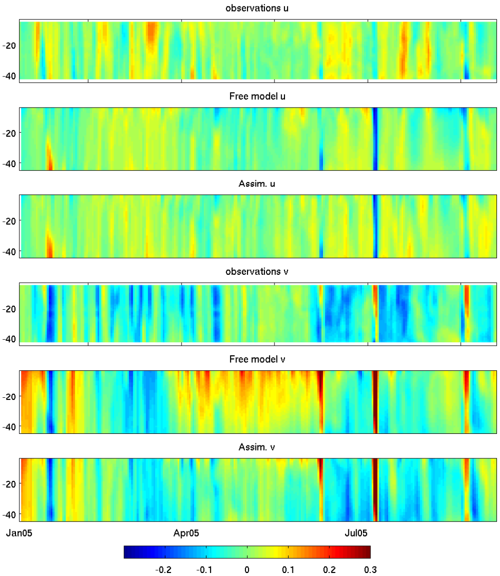



(a) ADCP observations and model runs. The asterisks on the time axis of the 4th panel represent Tropical Storm Arlene, Hurricane Dennis, Hurricane Katrina and Hurricane Ophelia.

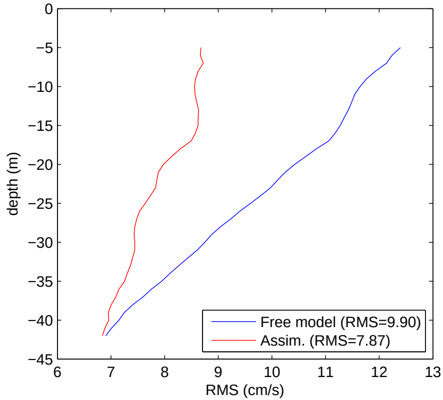



(b) Time-averaged RMS error between ADCP observations and the model run without assimilation (free model) and with assimilation.

Figure 9: Currents at station C10




(a) ADCP observations and model runs.




(b) Time-averaged RMS error.

- Free model shows an unrealistic Northwestward current at C12 during summer which is corrected through the assimilation
- The time averaged RMS error is reduced.

# Application to assimilation of HF Radar currents / German Bight

# Outline

- Weakly constrained ensemble perturbations
- Example 1: Estimation of tidal boundary conditions using HF radar observations
- Example 2: Estimation of wind forcing using HF radar observations

# Weakly constrained ensemble perturbations

- For ensemble schemes, unknown initial and boundary conditions, parameters, ... have to be perturbed within their range of uncertainty.
- By validation of the model with observations one can obtain an estimate of the magnitude of the perturbation.
- But which spatial structure?
- Method to create ensemble perturbation that satisfy *a priori* linear constraints
- Example of constraints:
  - geostrophic equilibrium
  - zero horizontal divergence of surface winds
  - stationary solution to the advection-diffusion equation
  - the linear shallow water equations
  - perturbations should be close to a subspace defined by *e.g.* empirical orthogonal functions (EOFs).
  - ...

# Probability of a perturbation

- To describe our *a priori* knowledge of what a realistic perturbation is, we introduce a cost function $J$, similar to the cost function used in variational analysis techniques:

$$J(\mathbf{x}) = \text{“linear balance”}^2 + \text{“smooth”}^2 + \text{“limited amplitude”}^2$$

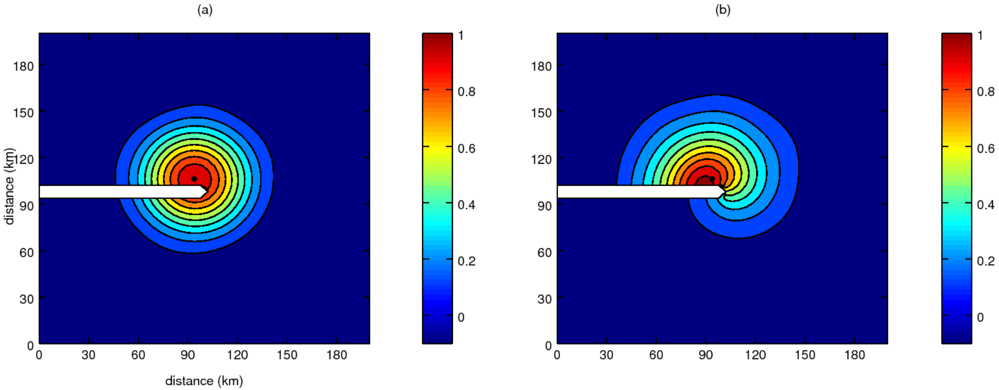
- The cost function can be used to define the probability of a perturbation $\mathbf{x}$ (*e.g.* Kalnay, 2002):

$$p(\mathbf{x}) = \alpha \exp\left(-J(\mathbf{x})\right) \tag{75}$$

- Perturbations are derived from the Hessian matrix of $J$.
- Article and source code (for MATLAB and GNU Octave) is available at http://modb.oce.ulg.ac.be/mediawiki/index.php/WCE
- “Dynamically constrained ensemble perturbations - application to tides on the West Florida Shelf”. Ocean Science, 5(3):259–270, 2009. http://www.ocean-sci.net/5/259/2009.

# Impact of barriers

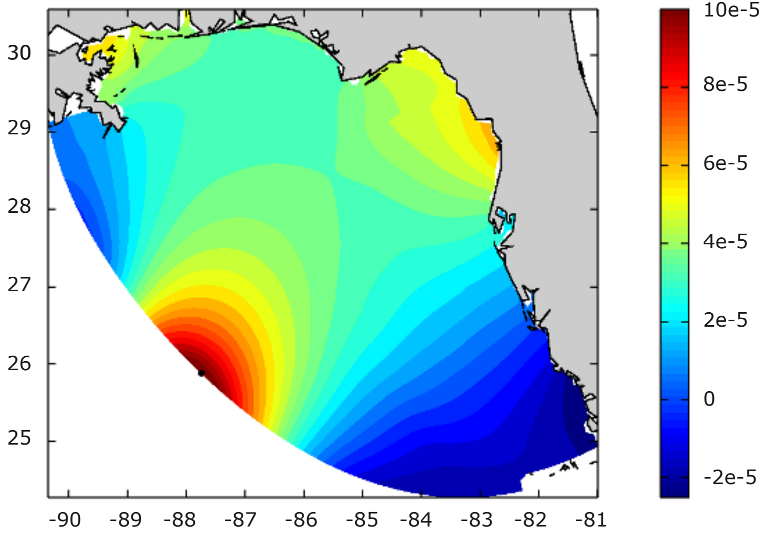
- The "smoothness" constraint is implemented through a diffusion operator (laplacian), it takes thus the land-sea mask into account




- Ensemble covariance using "classical" Fourier modes (a) and constrained perturbations based on the land-sea mask (b).

# Harmonic shallow water equations

- For tidal models, perturbations should be approximately a harmonic solution to the shallow water equations

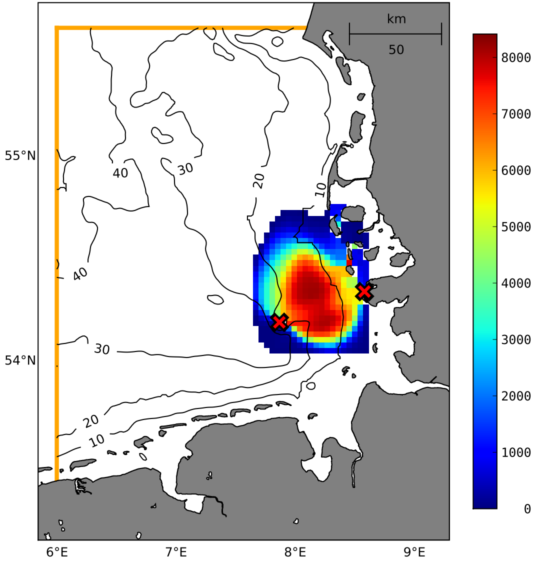



- Horizontal covariance of the constrained perturbations between the point near the open boundary marked by a black dot and all other grid points.

# German Bight model

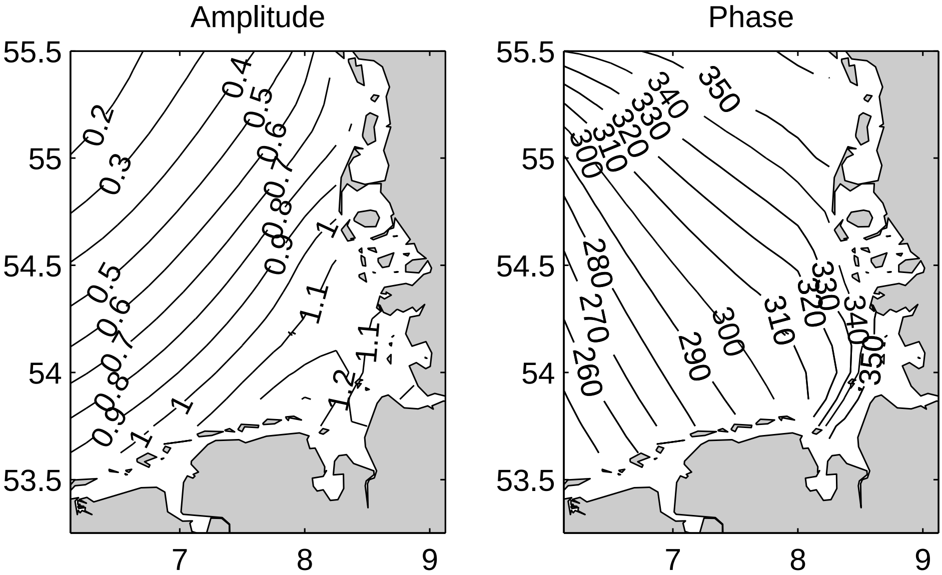
- General Estuarine Ocean Model (GETM Burchard and Bolding, 2002)
- 3-D primitive equations with a free-surface
- 21 $\sigma$ levels, resolution of about 0.9 km.
- nested in a 5-km resolution North Sea-Baltic Sea model
- ETOPO-1 topography with observations from BSH
- Atmospheric fluxes are estimated by the bulk formulation using 6-hourly ECMWF re-analysis
- Implementation by GKSS (Staneva *et al.*, 2009).

# HF radar observations




- Spatial coverage of the HF radar zonal and meridional surface velocity observations
- The number of samples available at each observation grid point is color–coded according to the color-bar.
- The crosses show the location of HF radar antennas.
- The operating frequency: 29.85 MHz (coupling to 5.02 m long ocean waves).
- HF Radar measurements from University of Hamburg (PRISMA project)

# Empirical Ocean Tides (EOT08a)

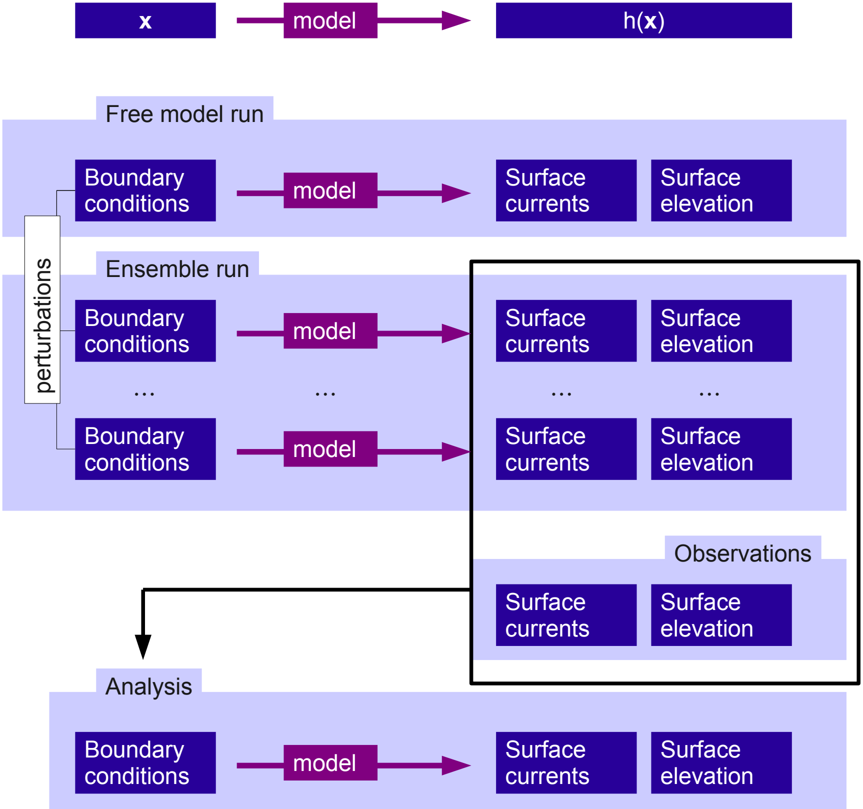



- M2 amplitude (in m) and phase (in degrees) of EOT08a for the German Bight based on altimetry.
- complex tidal parameters are assimilated

# Smoother scheme

- M2 tidal boundary conditions are perturbed within the range of their uncertainty to create a ensemble with 51 members. Perturbations are constrained by the linear shallow water equations.
- The GETM model is run for 40 days with each of those perturbed boundary values.
- All HF radar observations at any time instance within the integration period and the EOT parameters are grouped in the observation vector (vector $\mathbf{y}^o$) with their corresponding error covariance (matrix $\mathbf{R}$) estimated by cross-validation.
- Observations are extracted from perturbed model solution (vector $h(\mathbf{x}^{(k)})$).
- Schematically, the non-linear operator $h(\cdot)$ performs the following operations:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed forcing} \tag{76}$$


x
model
h(x)
Free model run
Boundary conditions
model
Surface currents
Surface elevation
perturbations
Ensemble run
Boundary conditions
model
Surface currents
Surface elevation
...
...
...
...
Boundary conditions
model
Surface currents
Surface elevation
Observations
Surface currents
Surface elevation
Analysis
Boundary conditions
model
Surface currents
Surface elevation

# Smoother scheme

- The optimal perturbation is given the Kalman analysis (using non-linear observation operators as in Chen and Snyder (2007)):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\left(\mathbf{B} + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \quad (77)$$

- where the matrices $\mathbf{A}$ and $\mathbf{B}$ are covariances estimated from the ensemble.

$$\mathbf{A} = \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) = \left\langle \left(\mathbf{x} - \langle \mathbf{x} \rangle\right)\left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right)^T \right\rangle \quad (78)$$

$$\mathbf{B} = \mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) = \left\langle \left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right)\left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right)^T \right\rangle \quad (79)$$

where $\langle \cdot \rangle$ is the ensemble average.

- But covariance matrices do not need to be formed explicitly. Analysis is performed in the subspace defined by the ensemble members.

# Smoother scheme

- For a linear model and an infinite large ensemble, equation (77) minimizes,

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b^{-1}} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x})) \tag{80}$$

  or

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{b^{-1}} (\mathbf{x} - \mathbf{x}^b) + \sum_n (\mathbf{y}_n^o - (h(\mathbf{x})_n))^T \mathbf{R}_n{}^{-1} (\mathbf{y}_n^o - (h(\mathbf{x})_n)) \tag{81}$$

  where $n$ references to the indexed quantifies at time $n$. This is the cost function from which 4D-Var and Kalman Smoother can be derived.

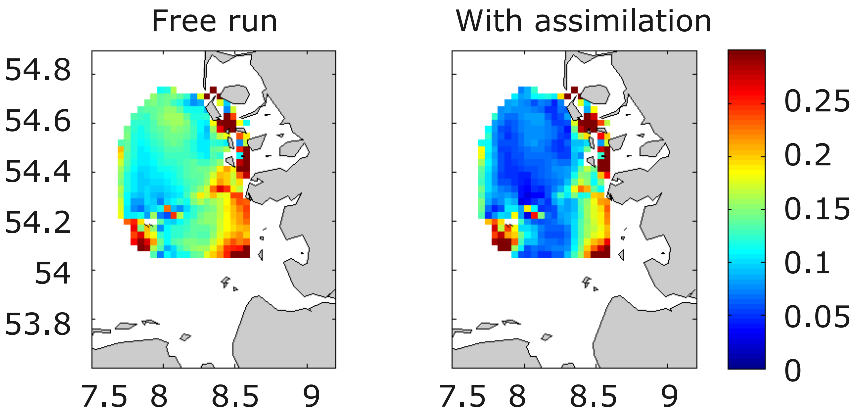
- Approach is closely related to Ensemble Smoother (van Leeuwen, 2001), 4D-EnKF (Hunt *et al.*, 2007) and AEnKF (Sakov *et al.*, 2010) where model trajectories instead of model states are optimized and to the Green's method with stochastic "search directions"

- The model is rerun with the optimized boundary values for 60 days.

# RMS difference

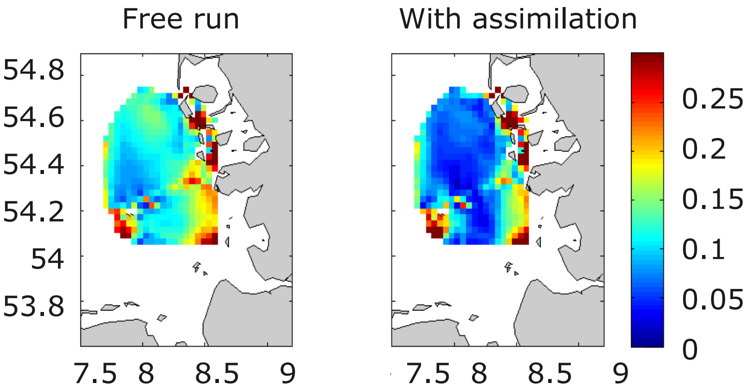
$$\text{RMS}^2 = \lim_{T\to\infty} \frac{1}{T}\int_0^T (A\cos(\omega t - \phi) - A'\cos(\omega t - \phi'))^2 dt \quad (82)$$

$$= \frac{A^2 + A'^2}{2} - AA'\cos(\phi - \phi') \quad (83)$$




RMS difference between surface current observations due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).

# Comparison with un-assimilated observations (M2)




- RMS difference between surface current observations (not used in the assimilation) due to the M2 tides and the corresponding model results without (left panel) and with assimilation (right panel).
- Analysis RMS compared to unassimilated data is only 0.002 m/s larger than compared to assimilated data

# Tide gage observations

| | Helgoland | | | Cuxhaven | | |
|---|---|---|---|---|---|---|
| | amplitude | phase | RMS | amplitude | phase | RMS |
| Observations | 1.13 | 304 | | 1.36 | 334 | |
| Free | 0.81 | 318 | 0.28 | 0.95 | 15 | 0.63 |
| Assimilation | 0.97 | 302 | 0.12 | 1.08 | 2 | 0.46 |

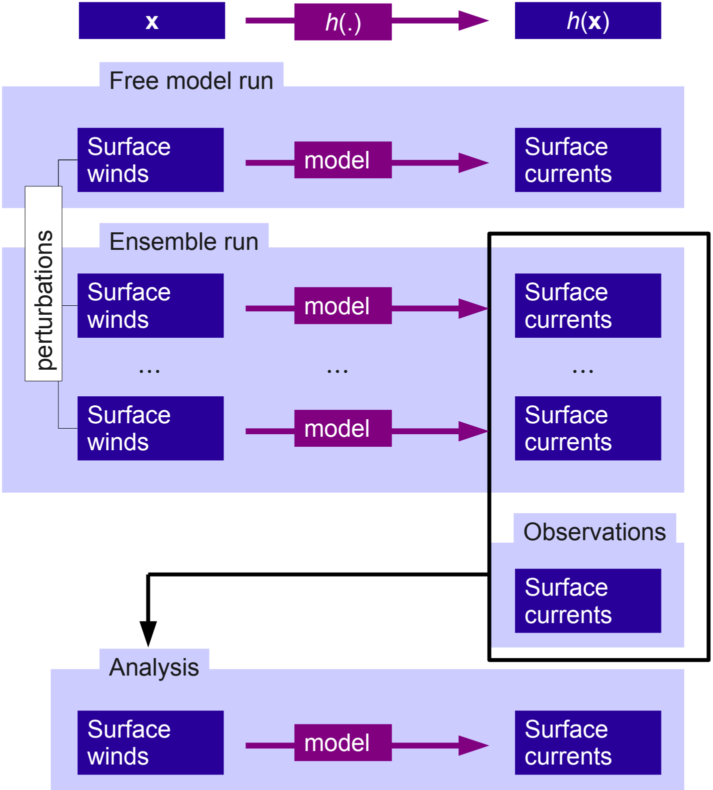
Table 1: Comparison with tide gage observations. Amplitude is in m and phase in degrees.

- Tide gage observations from different time period → only comparison of tidal parameters
- Helgoland within the area covered by radar, but not Cuxhaven
- The assimilation reduces the RMS error by a factor of 2 for Helgoland and by a factor of 1.4 for Cuxhaven.
- Ocean Science, 6, 161–178, 2010 http://www.ocean-sci.net/6/161/2010/os-6-161-2010.pdf.

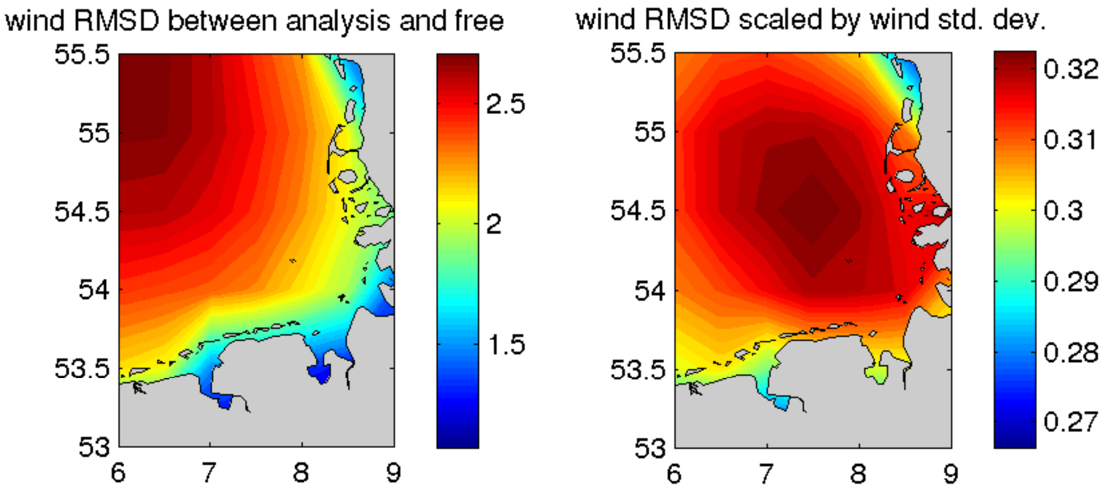
# Wind estimation from HF radar observations

- Ensemble of 100 wind forcings are created (by using a Fourier decomposition)
- estimation vector $\mathbf{x}$: u- and v- component of wind forcing
- observations: $\mathbf{y}^o$: surface currents
- "observation operator" $h(\cdot)$:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed wind} \quad (84)$$


x
h(.)
h(x)
Free model run
Surface winds
model
Surface currents
perturbations
Ensemble run
Surface winds
model
Surface currents
...
...
...
Surface winds
model
Surface currents
Observations
Surface currents
Analysis
Surface winds
model
Surface currents

# Time-averaged wind correction statistics




- RMS difference between analyzed winds and ECMWF winds (averaged over time)
- RMS difference scaled by wind standard deviation

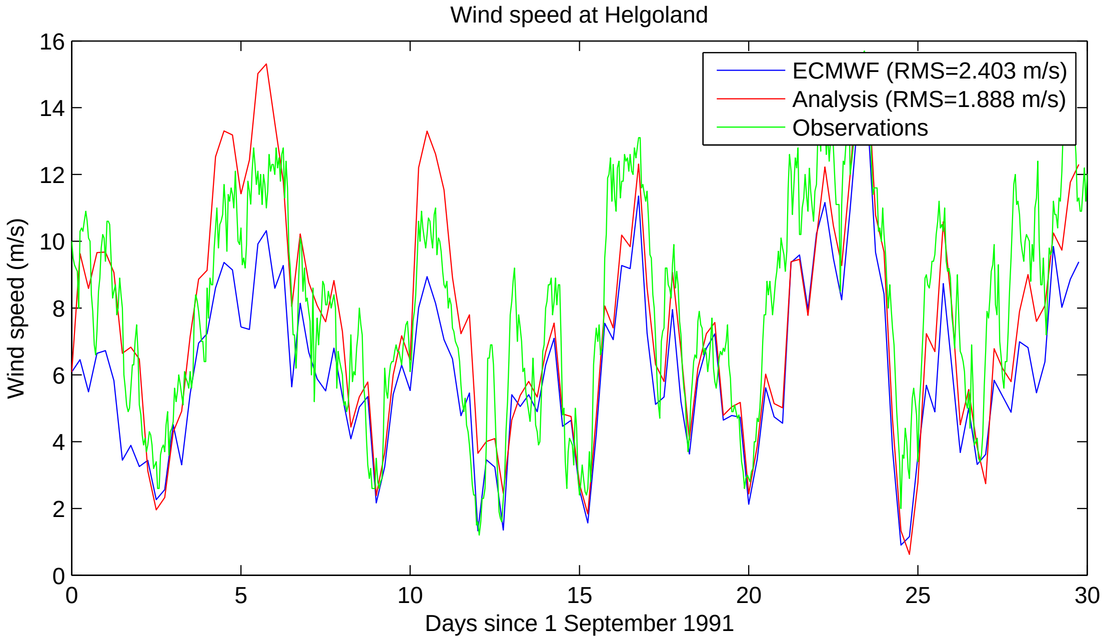



Figure 10: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Helgoland. Units are m/s.

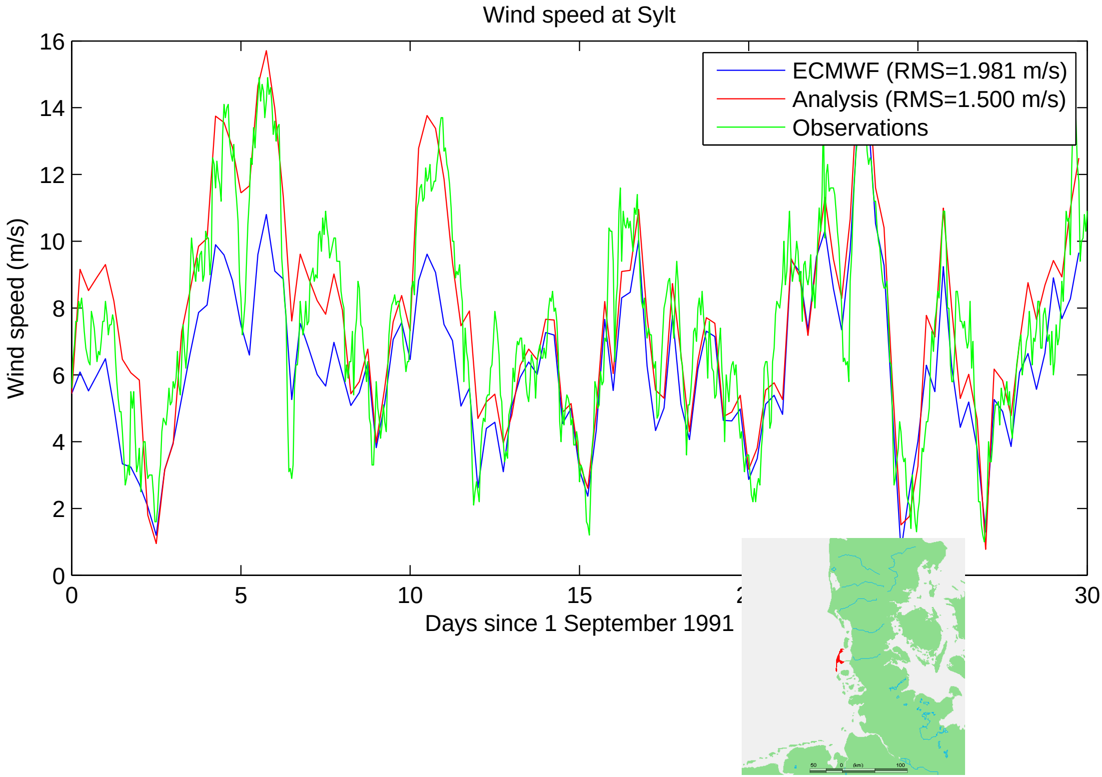
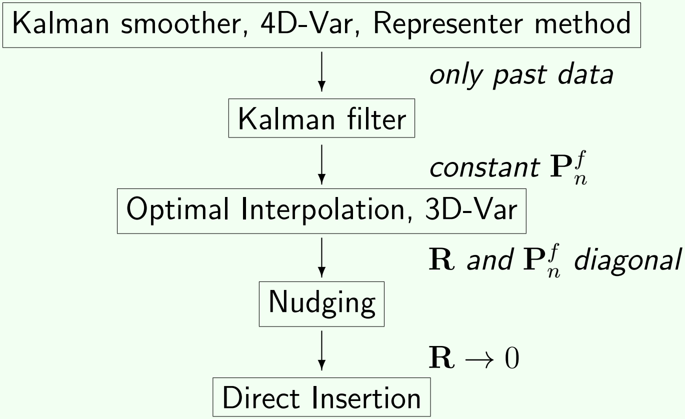



Figure 11: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Sylt. Units are m/s.

# Summary

- Ensemble assimilation methods require realistic perturbation schemes (error covariances)
- Use of dynamical relationships (similar to Variational analysis)
- Optimizing tidal boundary conditions and wind forcing with a smoother scheme
- HF radar observation is a very valuable data set for constraining regional and coastal models

# Summary of sequential methods




Slides on http://modb.oce.ulg.ac.be/wiki (Lecture → Introduction to data assimilation)

# Bibliography


Anderson, J. L., 2001: An Ensemble Adjustment Filter for Data Assimilation. *Monthly Weather Review*, **129**, 2884–2903.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129**, 420–436.

Burchard, H. and K. Bolding, 2002: GETM – a general estuarine transport model. Scientific documentation. Technical Report EUR 20253 EN, European Commission.

Chen, Y. and C. Snyder, 2007: Assimilating Vortex Position with an Ensemble Kalman Filter. *Monthly Weather Review*, **135**, 1828–1845.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, **123**, 1128–1145.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3385–3396, doi:10.1256/qj.05.108.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10143–10162.

— 2007: *Data assimilation: the Ensemble Kalman Filter*. Springer, 279pp.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 1 edition.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**, 1194–1207.

Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16**, 323–340.

Sakov, P., G. Evensen, and L. Bertino, 2010: Asynchronous data assimilation with the EnKF. *Tellus*, **62A**, 24–29.

Staneva, J., E. V. Stanev, J.-O. Wolff, T. H. Badewien, R. Reuter, B. Flemming, A. Bartholomä, and K. Bolding, 2009: Hydrodynamics and sediment dynamics in the German Bight. A focus on observations and numerical modelling in the East Frisian Wadden Sea. *Continental Shelf Research*, **29**, 302–319.

van Leeuwen, P. J., 2001: An Ensemble Smoother with Error Estimates. *Monthly Weather Review*, **129**, 709–728.