OCEA0097-1

Data assimilation and inverse methods

Alexander Barth

a.barth@ulg.ac.be

Revision 1.2

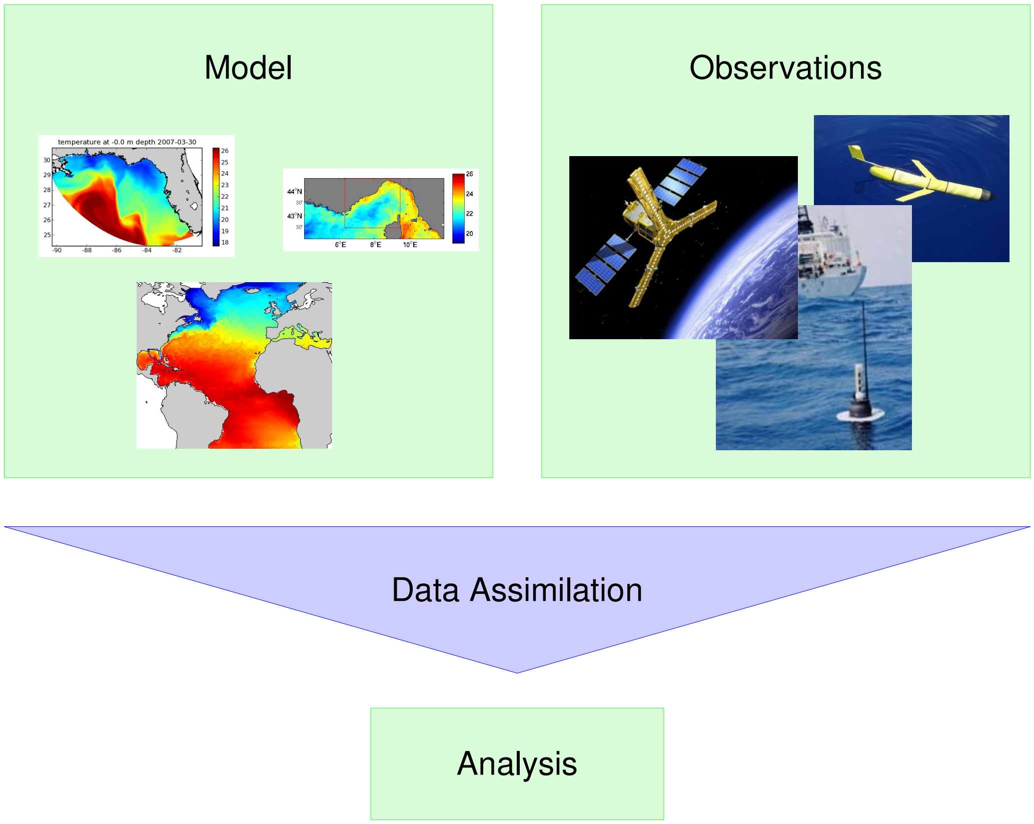
October 11, 2023

Data assimilation methods

.

> No one trusts a model except the man who wrote it; everyone trusts an observation except the man who made it
>
> (Harlow Shapley)

What is data assimilation ?



Outline

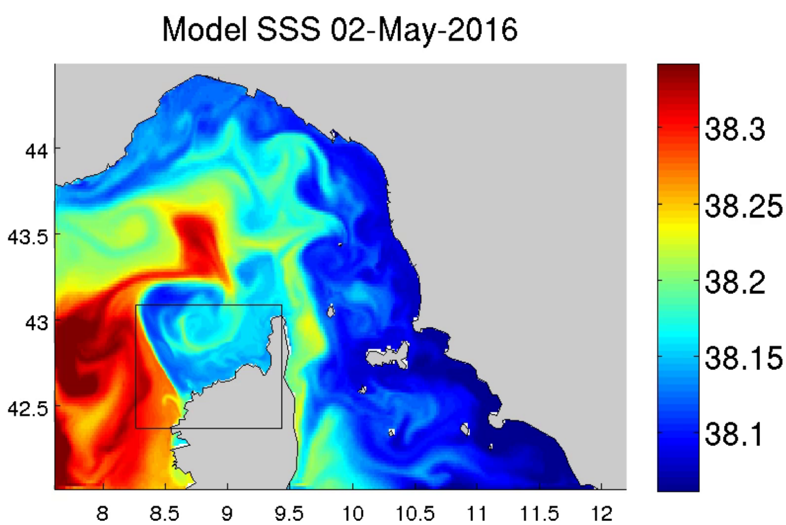
- Types of models
- Types of observations
- Data assimilation in a nutshell
- Basic concepts
- Sequential assimilation
    - Nudging
    - Successive corrections
    - Optimal Interpolation
    - 3D-Var
    - Kalman filter
    - Kalman smoother

- Non-Sequential assimilation
    - 4D-Var
    - Representer method

Types of ocean models

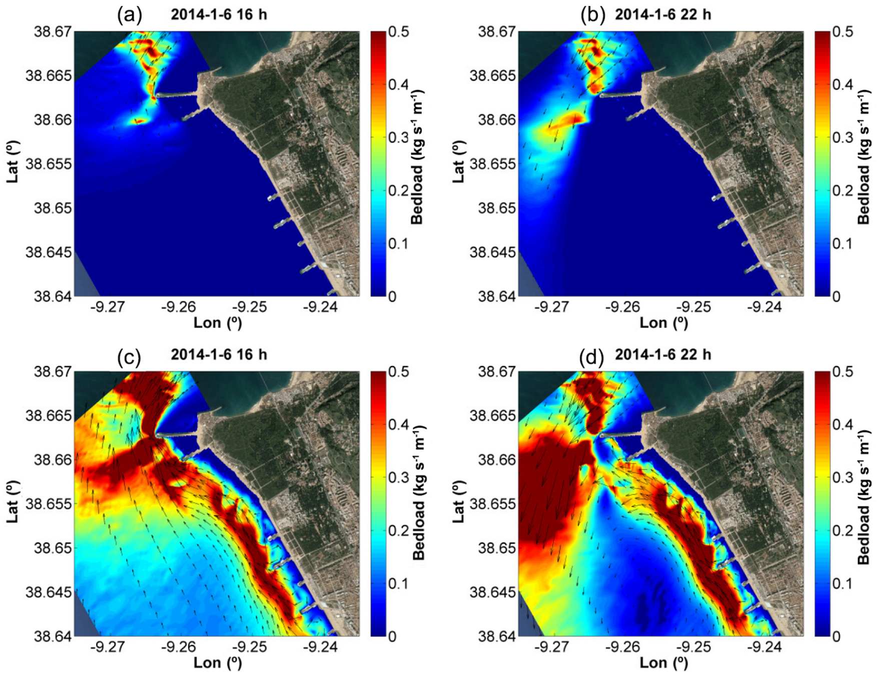
- hydrodynamical ocean
- sea-ice model
- wave model
- sediment model
- acoustic model
- biogeochemical model and ecosystem model

Hydrodynamical ocean



Surface salinity in a two-way nested model

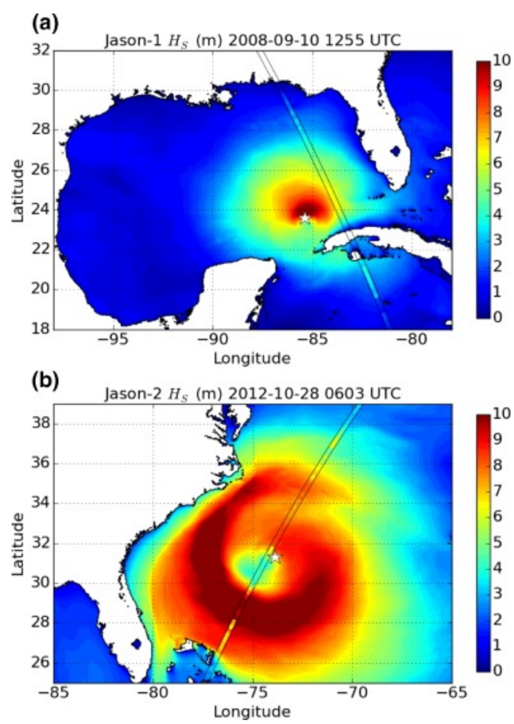
Sediment model



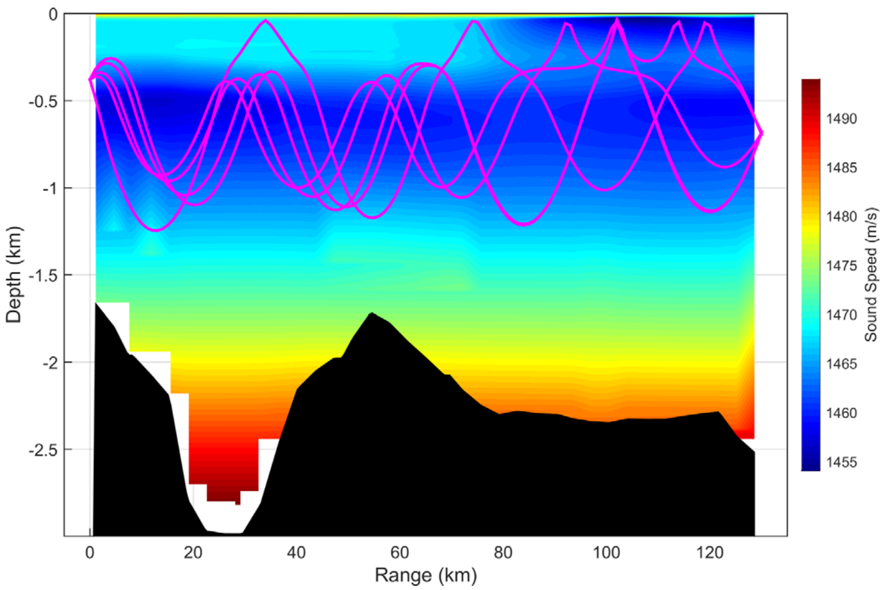
Bed load sediment transport for different wave scenarios (Franz *et al.*, 2017)

Wave model

Significant wave height for Hurricane Sandy (2012) Hurricane Ike (2008) (Chen and Curcic, 2016)

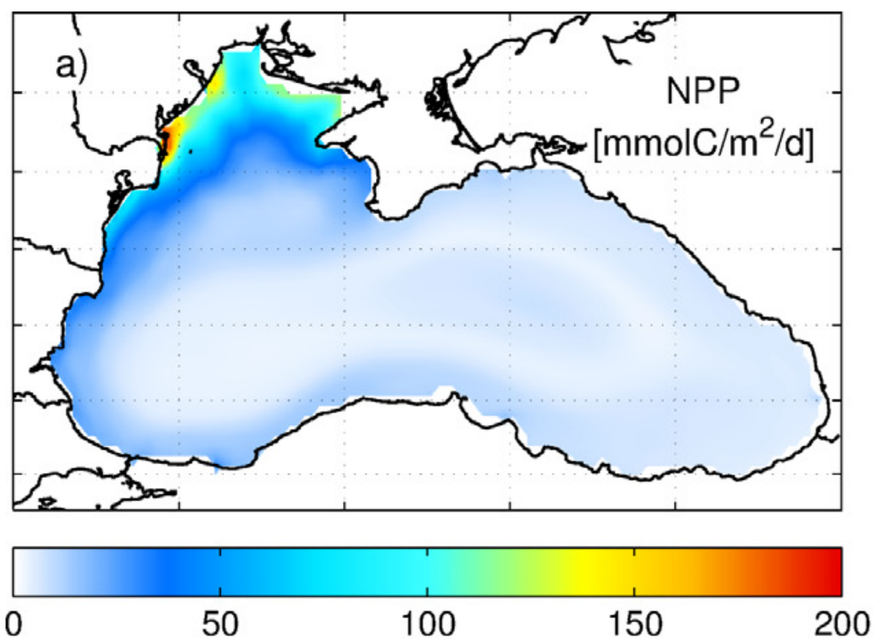


Acoustic model



Eigenrays obtained from the Fram Strait Model (Sagen *et al.*, 2016)

Biogeochemical model



Net primary production in the Black Sea (Capet *et al.*, 2016)

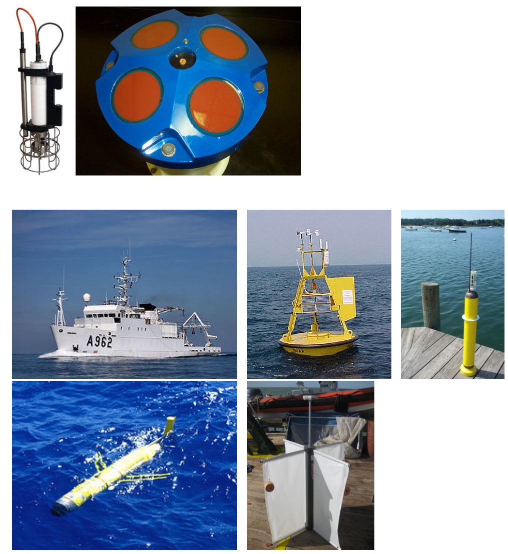
In situ observations

Instruments

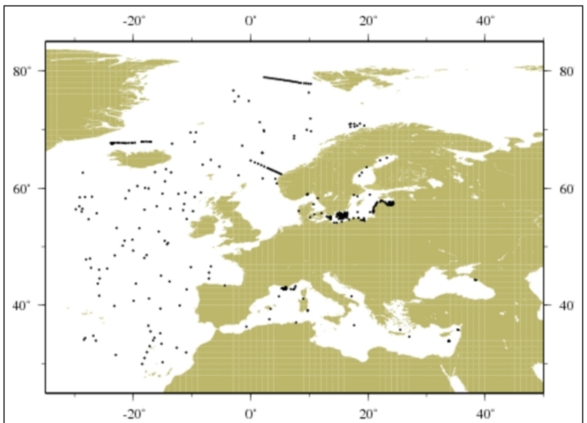
- Conductivity, Temperature, Depth (CTD) Sensors
- Acoustic Doppler Current Profiler (ADCP)
- Tide gauge
- Bottom pressure recorder,...

Platforms

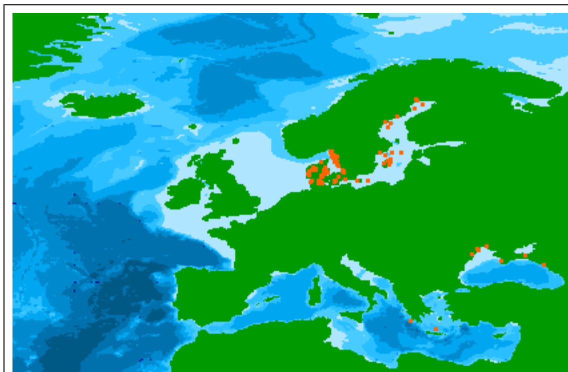
- Research ship
- Moored Buoys
- Profiling float, glider
- Drifter,...



Data distribution



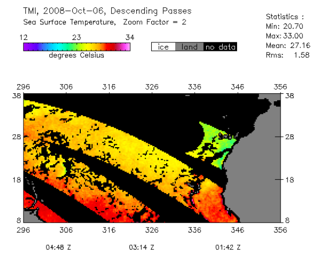
World Ocean Data Base, NODC, NOAA

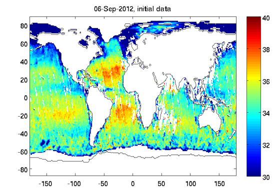


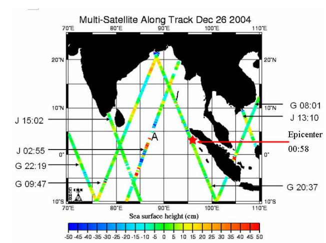
SeaDataNet, EU

- Distribution of salinity measurements from 1 to 7 August 2010.
- Distribution is very inhomogeneous, many gaps.
- Scale of variability (mesoscale) is the Rossby radius of deformation.
- In the Mediterranean Sea, this Rossy radius is about 10 km.
- Measurements in the open-ocean are costly, and problems in sharing ocean data

Remote sensing





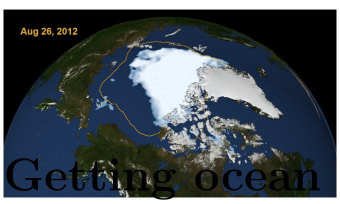


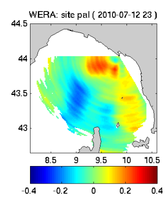
Parameters that can be measured from space/aircraft:

- sea surface temperature
- sea surface elevation (altimetry and gravimetry)
- sea surface salinity (since 2010)
- sea ice concentration, ice thickness
- ocean color and total suspended matter
- ...

Land-based remote sensing:

- sea surface currents





Getting ocean dat

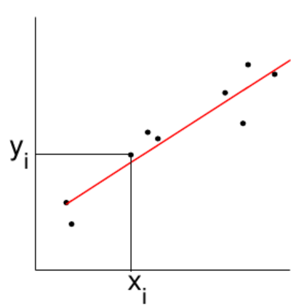
- Copernicus - Marine environment monitoring service (CMEMS) - model, in situ, satellite data
    - How to download CMEMS products?
    - MOTU, Using the MOTU client
    - Example notebook
- World Ocean Database - in situ
- SeaDataNet - in situ
- EMODNET - in situ
- Physical Oceanography Distributed Active Archive Center (PO.DAAC) - satellite data
- HYCOM - model
- ...

Observations and data assimilation

- need to be related as closesly as possible to model variable
  - SST: derive bulk temperature from skin temperature
  - HF radar: (ideally) remove Stokes drift if model currents do not include the Stokes drift
- Estimate observation error covariance
- Identify and remove outliers
  - observation error is assumed to be Gaussian distributed (for most assimilation methods)
  - However, in practice extreme values (outliers) are often more common that expected from the pdf
  - Using flags from the HF radar processing (e.g. quality, count, minimum/maximum velocity during averaging,...)
  - Coherence with
    * other observations (risk: smoothing out gradients)
    * model forecast (risk: trusting too much a "wrong" model forecast)
    * model analysis

Goal of data assimilation

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



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

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

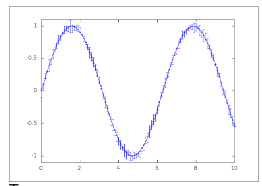
$$\frac{1}{\sigma_a^2} = \frac{1}{2} \frac{\partial^2 J}{\partial a^2} \qquad \frac{1}{\sigma_b^2} = \frac{1}{2} \frac{\partial^2 J}{\partial b^2}$$

- 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

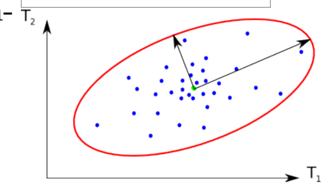
Errors and uncertainty

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

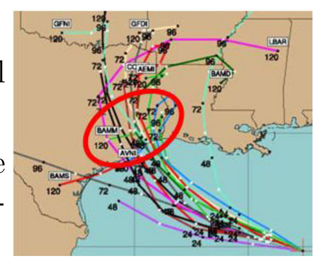
Ways to represent uncertainty



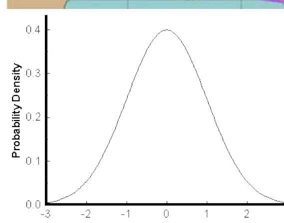
- For Gaussian-distributed errors.
  - Error bars (for scalar variables) (mean, standard deviation) or confidence interval
  - error covariance, example: $\mathbf{x} = (T_1, T_2)$



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



  - 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
- ...

Errors in your observations

Errors in your observations might be due to

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

Observations might not represent exactly the same as the model variables

- mismatch in resolved scales
- mismatch in resolved processes
- ...

In some cases the observation operator can be relative complex and might also involve same approximation (and thus potential errors).

Notation

n	scalar	number of state variables
m	scalar	number of observations
N	scalar	number of ensemble members
r	scalar	ensemble index $r = 1, \ldots, N$
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** $\mathbf{x}_k$ containing all **prognostic** variables at time t_k (time of the k-th time step). For a primitive equation model, its dimension is about $n = 5 \times 50000 \times 20 = 5\ 10^7$.

- The **dynamical model** f_k:

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

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

$$\begin{aligned} \mathbf{x}^t_{k+1} &= f_k(\mathbf{x}^t_k) + \boldsymbol{\eta}_k \\ \mathbf{x}^t_0 &= \mathbf{x}^i + \boldsymbol{\eta}^i \end{aligned}$$

- The difference between two successive times $t_{k+1} - t_k$ is not (necessarily) the time step of the model.

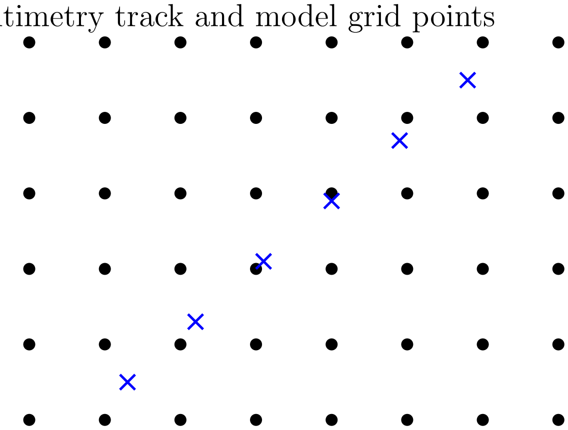
- $\mathbf{x}^t$ is of course unknown in a real application. The assimilation method does not require the knowledge of $\mathbf{x}^t$.

- The **observations**:

$$\mathbf{y}^o_k = h_k(\mathbf{x}^t_k) + \boldsymbol{\varepsilon}_k \quad \left[= \mathbf{H}_k \mathbf{x}^t_k + \boldsymbol{\varepsilon}_k \text{ if the obs. oper. is linear}\right]$$

Observation operator

- For example, an altimetry track and model grid points



- The observation operator includes often an interpolation and possibly a variable transformation

- The observation errors are composed by:

$$\boldsymbol{\varepsilon}_k = \text{instrumental error} + \text{representativity error}$$

Observation operator for HF radar data

- The HF radar systems operating at the frequency ν couple to a wave length of $\lambda_b = \frac{c}{2\nu}$ (where c is the speed of light)
- The HF radar current: weighted average with exponentially decreasing weights (Gurgel, 1994; Gurgel *et al.*, 1999):

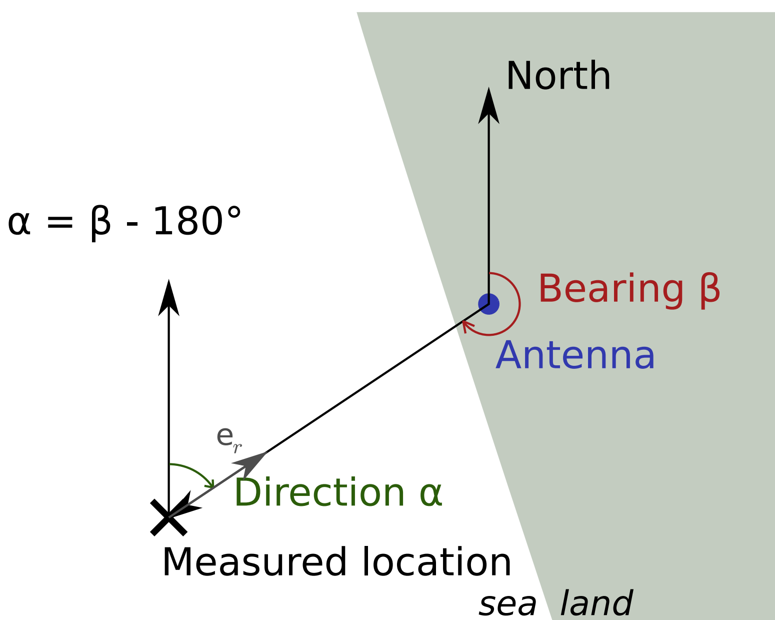
$$\mathbf{u}_{\text{surf}} = \frac{k_b}{1 - \exp(-k_b h)} \int_{-h}^{0} \mathbf{u}(z) \exp(k_b z) dz \tag{1}$$

  where $k_b = \frac{2\pi}{\lambda_b}$,
- Essentially represent an average over the upper meters.
- Radial velocity relative to the position of the HF radar site

$$u_{\text{HF}} = \mathbf{u}_{\text{surf}} \cdot \mathbf{e}_r \tag{2}$$

Observation operator for HF radar data

- Convention here:
  - $\mathbf{e}_r$ is the unit vector pointing in the towards to the location of the HF radar site.
  - u_{HF} is positive if velocity is pointing **towards** the site.
- Close to the HF radar site and far from the poles, direction α and bearing β are related by $\alpha = \beta - 180°$ (otherwise apply the azimuth formula)



Assumptions

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

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

- The covariances are known:

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

- 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 give useful results.
- For some assimilation methods, 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 $\mathbf{d}_k$ (time index k is dropped in the following):

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

- $\mathbf{H}\mathbf{P}\mathbf{H}^T$ 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 and covariance equal to the identity matrix.

Consistency check

- Verification statistics:

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

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

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

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

Assimilation in the simplest possible case

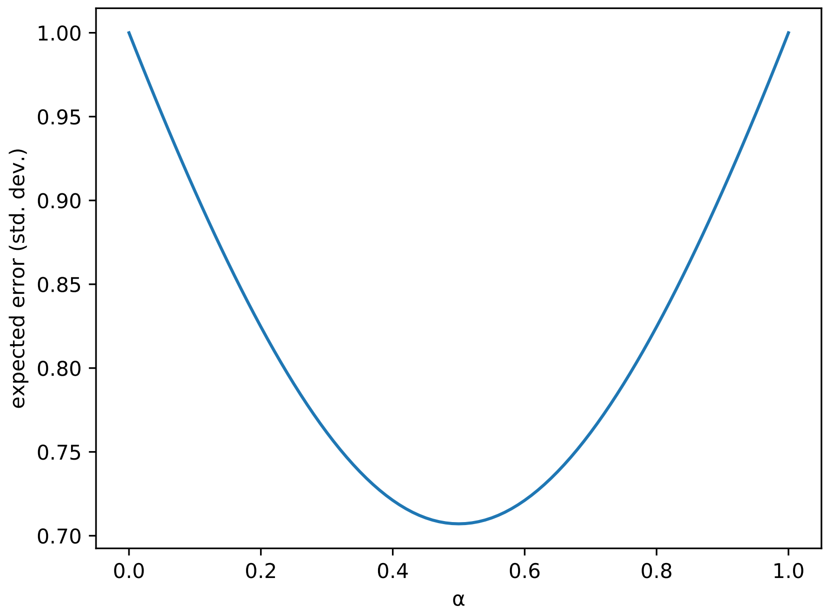
- model T_m (a scalar number) and the observations T_o (also a scalar number)
- both are approximation of the true value T_t
- mean squared error of T_m: $E\left[(T_m - T_t)^2\right] = \sigma_m^2$
- mean squared error of T_o: $E\left[(T_o - T_t)^2\right] = \sigma_o^2$
- the model and the observations are assumed to be unbiased in mutually independent
- weighted average: $T_a = (1-\alpha)T_m + \alpha T_o$
- mean square error of the weighted average:

$$\begin{aligned} \sigma_a^2 &= E\left[(T_a - T_t)^2\right] = E\left[\left((1-\alpha)(T_m - T_t) + \alpha(T_o - T_t)\right)^2\right] \\ &= (1-\alpha)^2\sigma_m^2 + \alpha^2\sigma_o^2 \end{aligned}$$

- at the minimum, we have $\frac{\partial \sigma_a}{\partial \alpha} = 0$
- we can show that: $\alpha = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_o^2}$

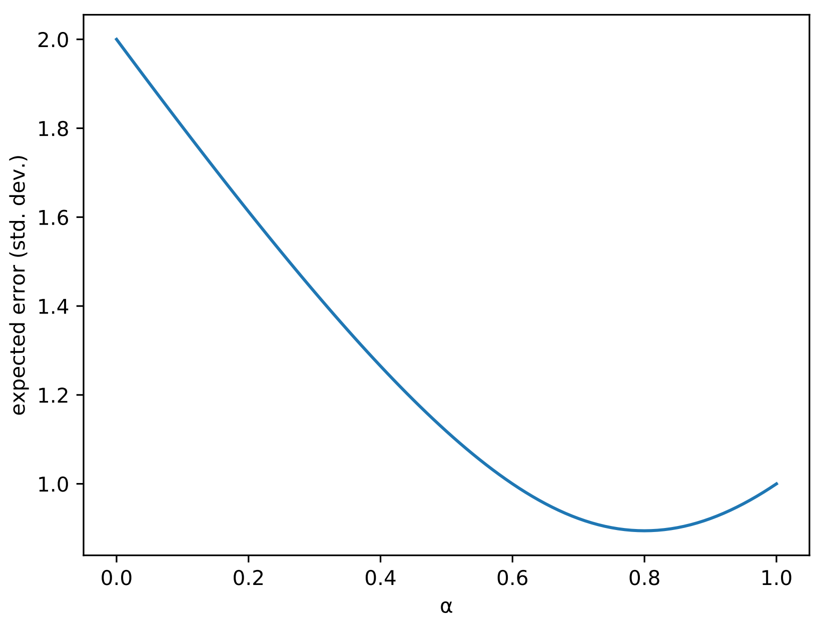
Assimilation in the simplest possible case

- Example for $\sigma_m = \sigma_o = 1$



Assimilation in the simplest possible case

- Example for $\sigma_m = 2$ and $\sigma_o = 1$



Sequential assimilation

$$\text{Initialization: } \mathbf{x}_0^a = \mathbf{x}^i$$

$$\text{Forecast: } \mathbf{x}_{k+1}^f = f_k(\mathbf{x}_k^a)$$

$$\text{Analysis: } \mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{K}_k \left(\mathbf{y}_{k+1}^o - h_k(\mathbf{x}_{k+1}^f)\right)$$

- $\mathbf{K}_k$: 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

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

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

- Analysis:

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

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.

$$\begin{aligned} \mathbf{x}^a_{kj'(i)} &= \mathbf{y}^o_{ki} \\ \mathbf{x}^a_{kj} &= \mathbf{x}^f_{kj} \qquad \text{on non-observed grid points} \end{aligned}$$

- The ith 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 ($\mathbf{H}_{j'(i),i} = 1$).

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

- Error in the model are assumed to be much larger than errors of the observations

Direct insertion

- 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:

$$\mathbf{x}^a_{kj'(i)} = \mathbf{x}^f_{kj'(i)} + r_i \left(\mathbf{y}^o_{ki} - \mathbf{x}^f_{kj'(i)} \right)$$
$$\mathbf{x}^a_{kj} = \mathbf{x}^f_{kj} \qquad \text{on non-observed grid points}$$

- In matrix form:

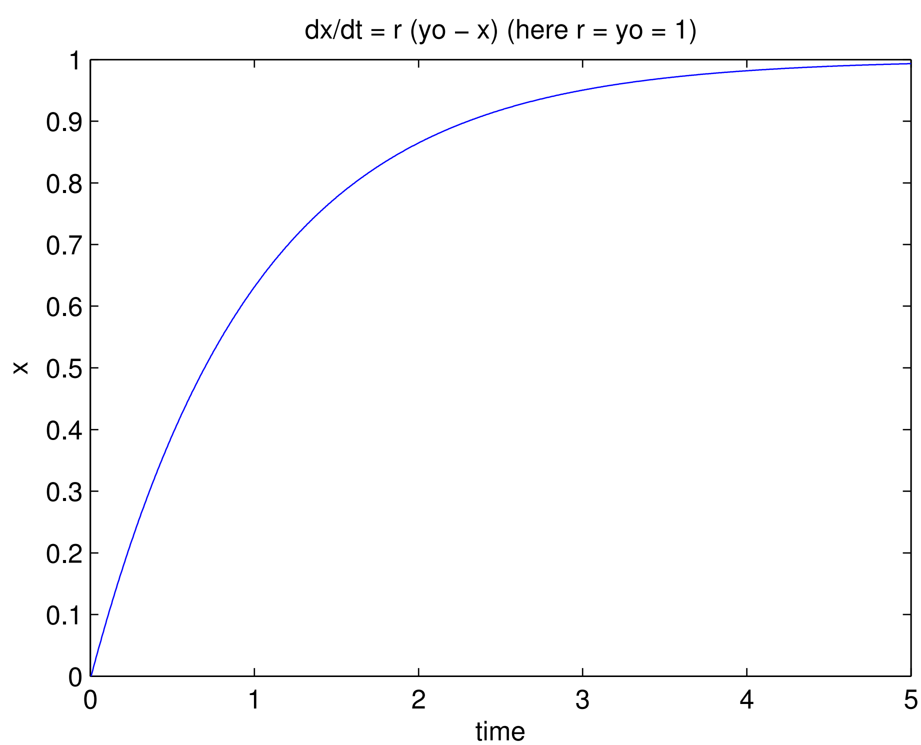
$$\mathbf{x}^a_k = \mathbf{x}^f_k + r_i \mathbf{H}^T \left(\mathbf{y}^o - \mathbf{x}^f_k \right)$$

- 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 different assimilation methods is available at `http://www.data-assimilation.net/Tools/AssimDemo/`.
- Review of what is a twin-experiment:
  - controlled model experiment
  - one model solution is declared as the "true" solution
  - pseudo-observation are extracted from this solution and noise is added
  - uncertain aspect of the model are perturbed
  - the pseudo-observations are assimilated into perturbed model
  - To which extend is the perturbed model similar to the "true" solution using data assimilation?
- 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:

$$\mathbf{x}_{n+1} = \mathbf{x}_n \tag{3}$$

- The model matrix $\mathbf{M}$ is thus the identity matrix.
- The two model variables are not dynamically coupled

1D advection in periodic domain

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

$$\mathbf{x}^{(n+1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \mathbf{x}^{(n)} \tag{4}$$

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

$$\frac{dx_1}{dt} = f x_2 \tag{5}$$
$$\frac{dx_2}{dt} = -f x_1 \tag{6}$$

- 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:

$$\mathbf{x}^{(n+1)} = \begin{pmatrix} \cos(f\Delta t) & \sin(f\Delta t) \\ -\sin(f\Delta t) & \cos(f\Delta t) \end{pmatrix} \mathbf{x}^{(n)} \tag{7}$$

- 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 allow us to find an analytic 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 $\Delta t = 0.1$.

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

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

$$\mathbf{x}_{n+1} = \mathbf{M}\mathbf{x}_n \tag{8}$$

Lorenz model

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

$$\frac{dx}{dt} = \sigma(y - x) \tag{9}$$
$$\frac{dy}{dt} = x(\rho - z) - y \tag{10}$$
$$\frac{dz}{dt} = xy - \beta z \tag{11}$$

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

Nudging demo

- Model: identity $\mathbf{x}_{k+1} = \mathbf{x}_k$
  - 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 (a system with two variables)
  - Based on the default values, try to find a good relaxation time-scale
    `http://data-assimilation.net/Tools/AssimDemo/?method=Nudging&model=oscillation`
  - How 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 of the state vector follows a Gaussian distribution
- The error covariance of the model state vector is assumed to be known and defined as:

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

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

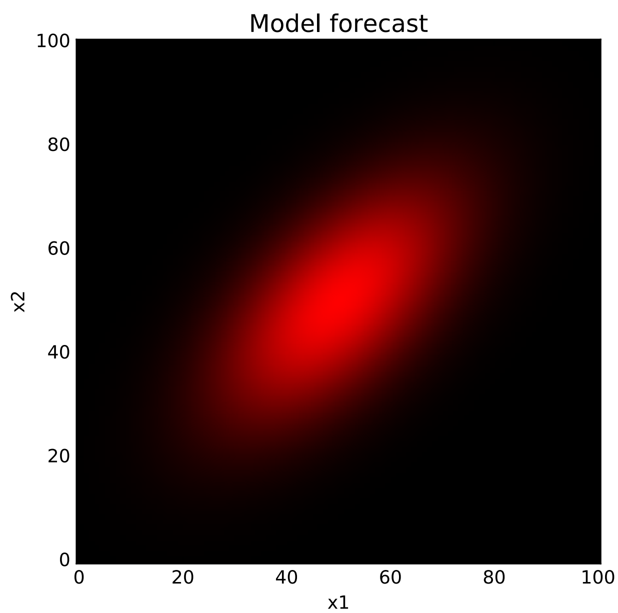
$$J(\mathbf{K}) = E[(\mathbf{x}_k^a - \mathbf{x}_k^t)^T \mathbf{W} (\mathbf{x}_k^a - \mathbf{x}_k^t)] = \mathrm{tr}(\mathbf{W}\mathbf{P}_k^a)$$

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

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

Analysis equation

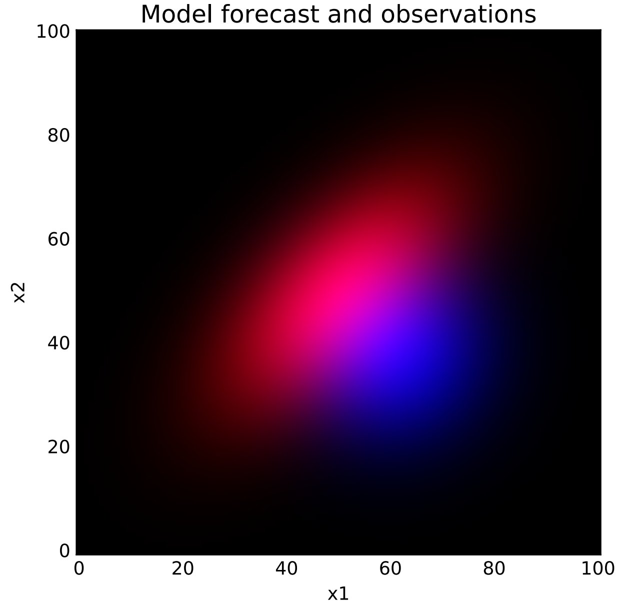
- Graphical representation
- Uncertainty is represented by the probability density function (pdf) of the model state vector
- Intensity of the color: $\rightarrow$ probability



$$p(\mathbf{x}) = A \exp(-(\mathbf{x} - \mathbf{x}^f)^T {\mathbf{P}^f}^{-1} (\mathbf{x} - \mathbf{x}^f))$$

Analysis equation

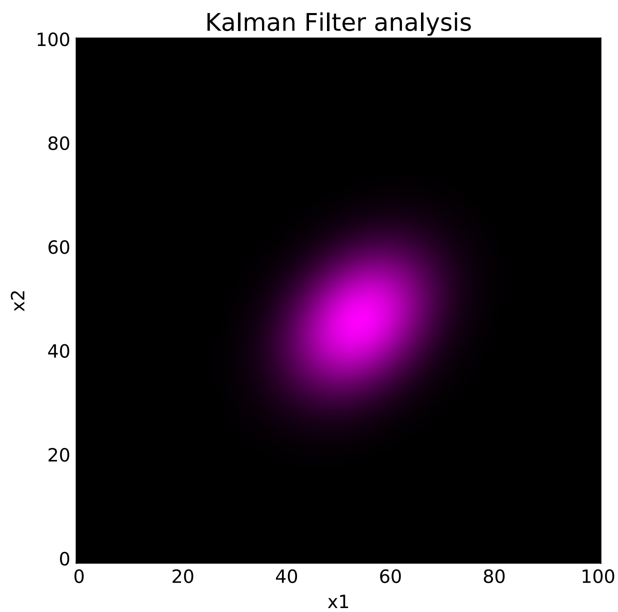
- pdf of the model state vector
- pdf of the observations given the model state vector
- Bayes rule:



$$p(\mathbf{x}|\mathbf{y}^o) = A' \exp(-(\mathbf{x}^f - \mathbf{x})^T {\mathbf{P}^f}^{-1}(\mathbf{x}^f - \mathbf{x}))$$
$$\exp(-(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}))$$

Analysis equation

- product of two Gaussian pdf is also a Gaussian pdf



$$p(\mathbf{x}|\mathbf{y}^o) = A' \exp(-(\mathbf{x}-\mathbf{x}^a)^T \mathbf{P}^{a-1}(\mathbf{x}-\mathbf{x}^a))$$

How to derive the Kalman gain?

- The analysis is given by:

$$\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) && (12)\\
&= (\mathbf{I}-\mathbf{K}\mathbf{H})\,\mathbf{x}^f + \mathbf{K}\mathbf{y}^o && (13)
\end{aligned}$$

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

$$\mathbf{P}^a(\mathbf{K}) = (\mathbf{I}-\mathbf{K}\mathbf{H})\,\mathbf{P}^f\,(\mathbf{I}-\mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \qquad (14)$$

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

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

Kalman gain

- 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 $\delta\mathbf{K}$.

$$\begin{aligned}
&\mathrm{tr}\left(\mathbf{W}\mathbf{P}^a(\mathbf{K}+\delta\mathbf{K})\right) - \mathrm{tr}\left(\mathbf{W}\mathbf{P}^a(\mathbf{K})\right) && (15)\\
&= 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{K}\ \mathbf{H}\mathbf{P}^f\mathbf{H}^T\,\delta\mathbf{K}^T\right) - 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{P}^f\mathbf{H}^T\delta\mathbf{K}^T\right) + 2\,\mathrm{tr}\left(\mathbf{W}\mathbf{K}\mathbf{R}\delta\mathbf{K}^T\right)\\
&= 2\,\mathrm{tr}\left(\mathbf{W}\left[\mathbf{K}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R}\right) - \mathbf{P}^f\mathbf{H}^T\right]\delta\mathbf{K}^T\right)
\end{aligned}$$

- Note that we used: $\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A})$ and $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^T)$
- Since the perturbation $\delta\mathbf{K}$ is arbitrary, the expression inside the brackets has to be zero.

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T+\mathbf{R}\right)^{-1} \qquad (16)$$

Error covariance of the analysis

Equation (14) can be expanded into:

$$
\begin{aligned}
\mathbf{P}^a &= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}^T && (17)\\
&= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T && (18)\\
&= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f && (19)
\end{aligned}
$$

where we used the optimal gain from equation (16).

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 (σ_m) with an observation T_o (σ_o), the analyzed temperature is:

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

Equivalent formulations

Equivalent formulations for the Kalman gain:

$$
\begin{aligned}
\mathbf{K} &= \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} && (20)\\
&= \left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\ \mathbf{H}^T\mathbf{R}^{-1} && (21)\\
&= \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} && (22)
\end{aligned}
$$

(proved using the Sherman-Morrison-Woodbury formula). For the analysis error covariance matrix:

$$
\mathbf{P}^{a^{-1}} = \mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \qquad (23)
$$

and the analysis update:

$$\mathbf{P}^{a^{-1}}\mathbf{x}^a \quad = \quad \mathbf{P}^{f^{-1}}\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{y}^o \tag{24}$$

Do you see a pattern in these two equations? What would be the analysis equation if you combine say 2 independent models and 3 independent observations vectors?

Example

- Compare the behavior of 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".

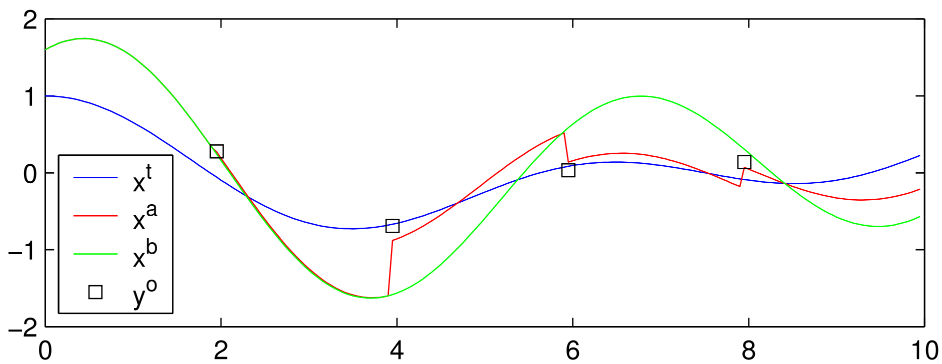


Figure 1: The observed part of a linear system with 4 state variables: the true state vector, $\mathbf{x}^t$, the analysis $\mathbf{x}^a$, the state of the system without assimilation $\mathbf{x}^b$ (b, background). The observations $\mathbf{y}^o$ are extracted from $\mathbf{x}^t$. The trajectories $\mathbf{x}^a$ and $\mathbf{x}^b$ start from a wrong initial condition.

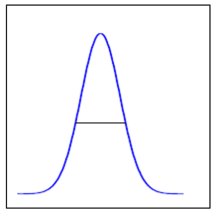
Covariances

- $\mathbf{P} : n \times n (n \approx 10^6)$. 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:

$$\mathbf{P} = E[\eta\eta^T] \tag{25}$$

$$\mathbf{P} = \mathbf{L}\mathbf{D}\mathbf{L}^T \qquad \mathbf{L} : n \times r,\ \mathbf{D} : r \times r \tag{26}$$

In general $r \approx 10 - 100$.

- Motivation: Empirical orthogonal functions have been shown to reduce the time variability of an ocean model and satellite data to a very small subspace defined by the EOFs.

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

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

- Related to SEEK analysis

Ensemble Optimal Interpolation

- Definition of error covariance

$$\mathbf{P} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \tag{27}$$

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

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

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

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

- But a reduced-rank covariance introduces an nonphysical 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):

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

- 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:

$$\mathbf{P}^{a^{-1}} = \frac{1}{2} \nabla_{\mathbf{x}} \nabla_{\mathbf{x}} J \quad (28)$$

$$= \mathbf{P}^{f^{-1}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \quad (29)$$

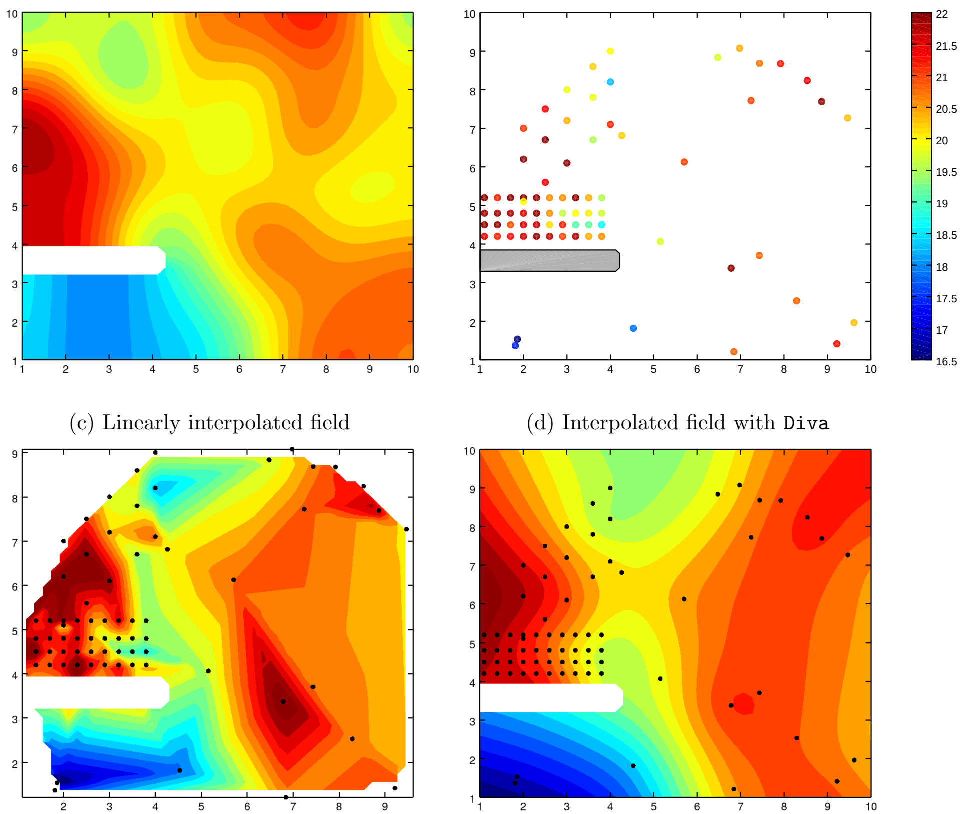
- Generalization of optimal interpolation to non-linear h

- 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:

$$\int_D \alpha_2 \nabla\nabla\varphi : \nabla\nabla\varphi + \alpha_1 \nabla\varphi \cdot \nabla\varphi + \alpha_0 \varphi^2 dD \quad (30)$$

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



(a) Example of oceanographic field (b) Extracted observations with errors

(c) Linearly interpolated field (d) Interpolated field with `Diva`

Relationship between 3D-var and optimal interpolation

For a linear $\mathbf{H}$,

$$\frac{1}{2}\nabla J(\mathbf{x}_a) = 0 \tag{31}$$

$$= \mathbf{P}^{f^{-1}}\left(\mathbf{x}^a - \mathbf{x}^f\right) - \mathbf{H}(\mathbf{x})^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^a\right) \tag{32}$$

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

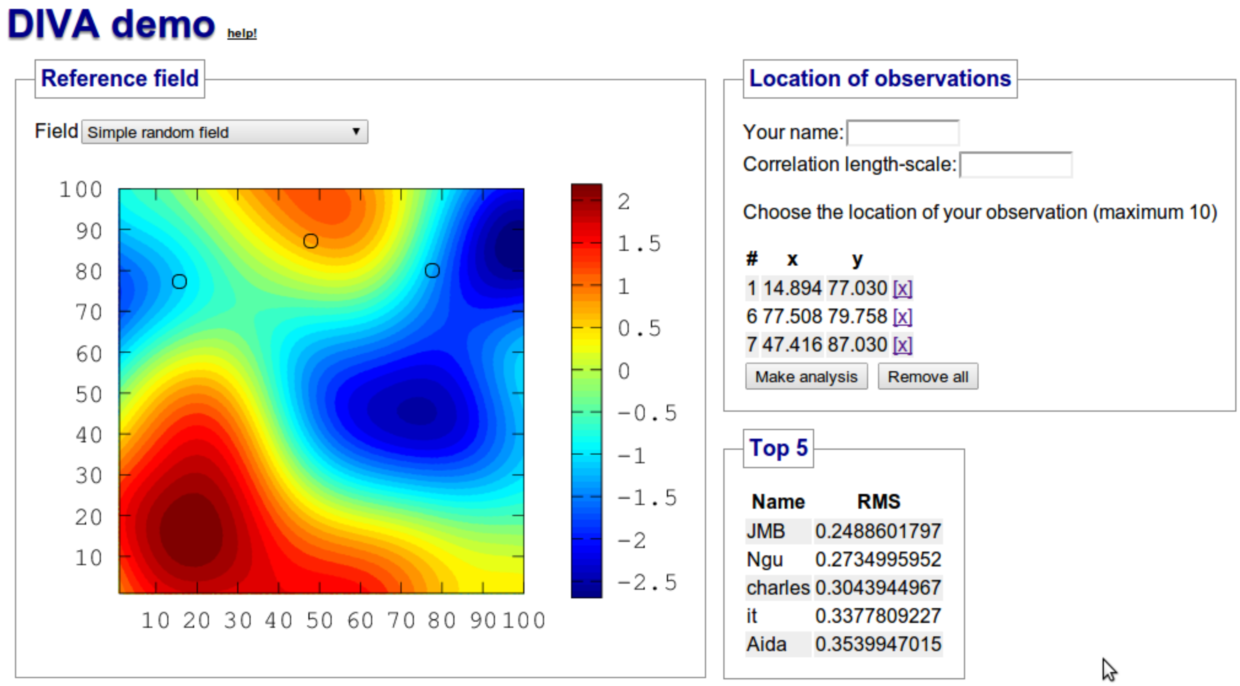
$$\left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{x}^a = \mathbf{P}^{f^{-1}}\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f + \mathbf{H}\mathbf{x}^f\right) \tag{33}$$

$$\mathbf{x}^a = \mathbf{x}^f + \left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\ \mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{34}$$

Sherman-Morrison-Woodbury formula:

$$\left(\mathbf{P}^{f^{-1}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\ \mathbf{H}^T\mathbf{R}^{-1} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{35}$$

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



- `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:

$$\mathbf{P}_{k+1} = \mathbf{M}_k \mathbf{P}_k \mathbf{M}_k^T + \mathbf{Q}_k \quad \text{where } M_{kjj'} = \frac{\partial f_{kj}}{\partial x_{j'}}$$

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

Derivation of the error propagation equation

- The definition of the error covariance matrix of $\mathbf{x}_{k+1}$

$$\mathbf{P}_{k+1} = E\left[\left(\mathbf{x}_{k+1} - \mathbf{x}^t_{k+1}\right)\left(\mathbf{x}_{k+1} - \mathbf{x}^t_{k+1}\right)^T\right] \tag{36}$$

- The difference between the evolution equation of $\mathbf{x}_{k+1}$ and $\mathbf{x}^t_{k+1}$ yields

$$\mathbf{x}_{k+1} - \mathbf{x}^t_{k+1} = \mathbf{M}_k(\mathbf{x}_k - \mathbf{x}^t_k) - \eta_k \tag{37}$$

- Finally

$$\mathbf{P}_{k+1} = \mathbf{M}_k\mathbf{P}_k\mathbf{M}^T_k + \mathbf{Q}_k \tag{38}$$

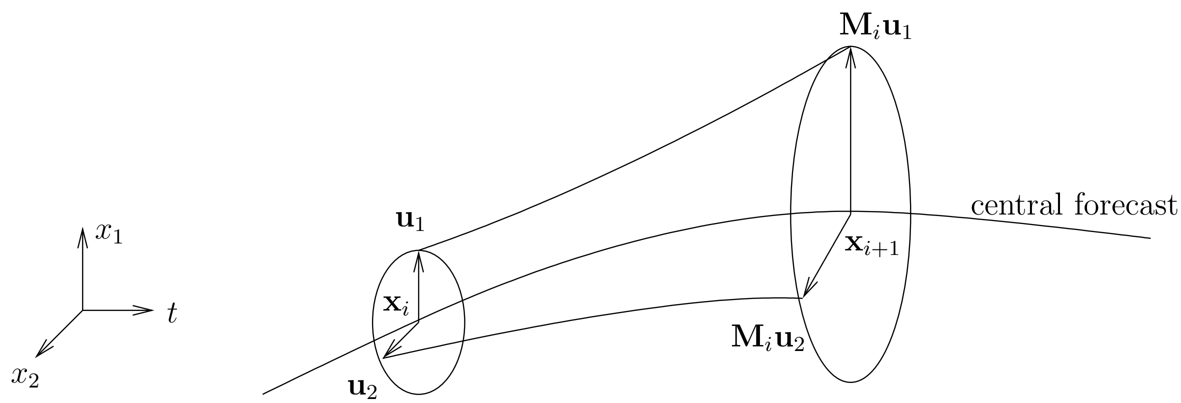


Figure 2: Forecast of the error covariance with the tangent linear model

- 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:
$$\mathbf{x}_{k+1}^f = f_k(\mathbf{x}_k^a)$$
$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

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

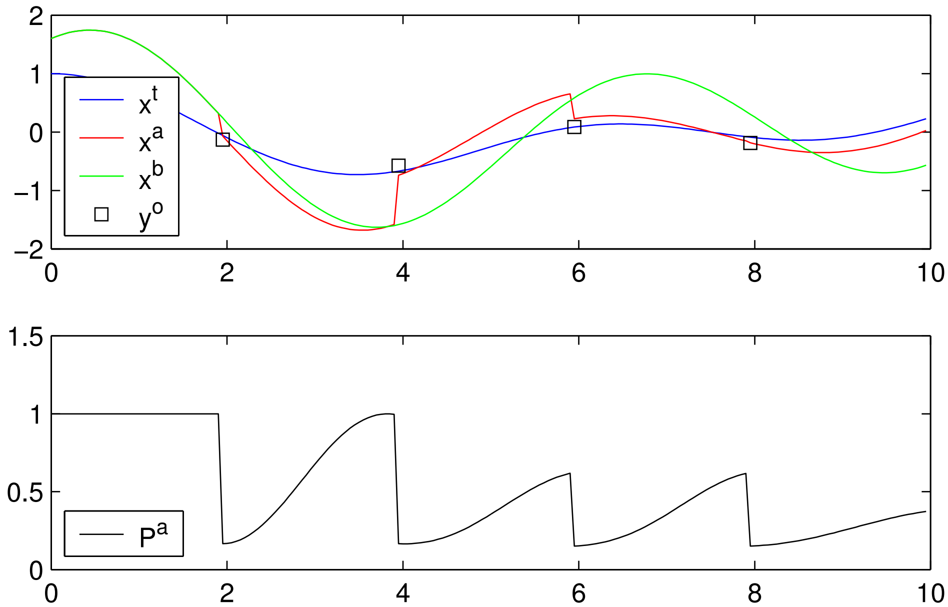


Figure 3: Example of a Kalman filter applied to a linear system. The curves from the first graph correspond to the observed part of the system. The lower panel shows the evolution of the error covariance. The error variance of the state vector is reduced at every assimilation cycle.

Numerical example: a water column

- Application of the Extended Kalman Filter
- Model represents a water column governed by:

$$\frac{\partial \mathbf{u}}{\partial t} + f\mathbf{e}_z \times \mathbf{u} = \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial \mathbf{u}}{\partial z}\right) \tag{39}$$

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial T}{\partial z}\right) \tag{40}$$

$$\frac{\partial S}{\partial t} = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial S}{\partial z}\right) \tag{41}$$

$$\frac{\partial k}{\partial t} = \tilde{\nu}\left(\frac{\partial \mathbf{u}}{\partial z}\right)^2 - \frac{\tilde{\nu}}{16}k^2 - \tilde{\nu}\frac{\partial b}{\partial z} + \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial k}{\partial z}\right) \tag{42}$$

- The prognostic variables $\mathbf{u}, T, S$ and k
- The diagnostic variables: buoyancy b, the Richardson number Ri the turbulent diffusion coefficient $\tilde{\nu}$ and $\tilde{\lambda}$:

$$b(T,S) = \frac{\rho(T,S) - \rho_0}{\rho_0} \tag{43}$$

$$Ri = \frac{\partial b}{\partial z}\left(\frac{\partial \mathbf{u}}{\partial z}\right)^{-2} \tag{44}$$

$$\tilde{\nu} = \tilde{\nu}(Ri, k) \tag{45}$$

$$\tilde{\lambda} = \tilde{\lambda}(Ri, k) \tag{46}$$

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

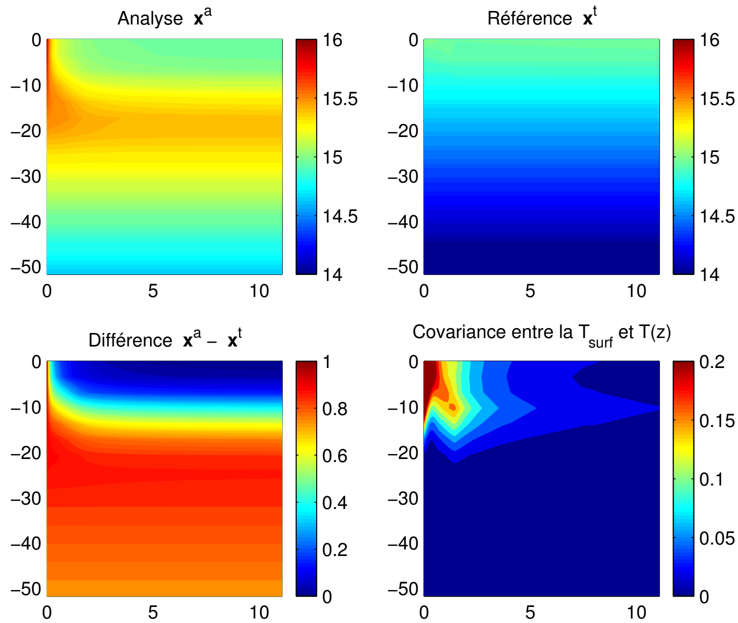


Figure 4: Temperature as a function of time (hours) and depth. Only the upper 50 meters are shown.

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

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 6 to 5. 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?

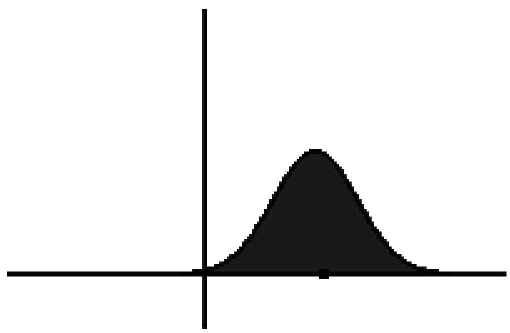
3. Discuss the correction by data assimilation of the variables x_1 and x_2 (not directly observed).

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

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

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

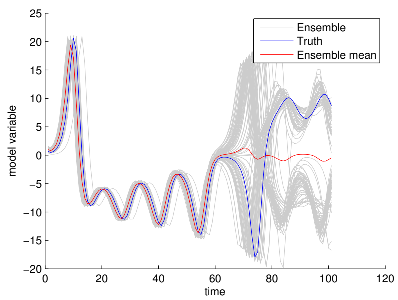


Propagation of uncertainty - Non-Gaussian errors

- 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 $\mathbf{x}^{f^{(r)}}$ $(r = 1, \ldots, N)$ one can compute the ensemble mean

$$\overline{\mathbf{x}^f} = \frac{1}{N}\sum_{r=1}^{N} \mathbf{x}^{f^{(r)}} \tag{47}$$

- The covariance around this mean is the ensemble covariance:

$$\mathbf{P}^f = \frac{1}{N-1}\sum_{r=1}^{N} \left(\mathbf{x}^{f^{(r)}} - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^{f^{(r)}} - \overline{\mathbf{x}^f}\right)^T \tag{48}$$

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

$$\left(\mathbf{S}^f\right)_r = \frac{\mathbf{x}^{f^{(r)}} - \overline{\mathbf{x}^f}}{\sqrt{N-1}} \tag{49}$$

  where $\mathbf{S}^f$ is a $n \times N$ matrix, which each column being the difference between each member and its ensemble mean.

- $\mathbf{P}^f$ is thus naturally decomposed in terms of this square root matrix $\mathbf{S}^f$:

$$\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{f^T} \tag{50}$$

- The original Ensemble Kalman Filter (Evensen, 1994; Burgers *et al.*, 1998) use perturbed observations

$$\mathbf{y}^{o(r)} = \mathbf{y}^o + \boldsymbol{\varepsilon}^{(r)}$$

  where $\boldsymbol{\varepsilon}^{(r)}$ is a random vector following a Gaussian distribution with zero mean and a covariance of $\mathbf{R}$.

- The ensemble mean $\mathbf{y}^{o(r)}$ is often forced to zero

- The added perturbation can be interpreted as perturbation of $\mathbf{H}\mathbf{x}^{f(r)}$.

- Every ensemble member is updated according to:

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

  The Kalman gain is based on the ensemble covariance of the state vector

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

  where $\mathbf{R}_e$ is the ensemble approximation of the observational error covariance matrix.

$$\mathbf{R}_e = \frac{1}{N-1}\sum_{r=1}^{N}\left(\boldsymbol{\varepsilon}^{(r)} - \overline{\boldsymbol{\varepsilon}}\right)\left(\boldsymbol{\varepsilon}^{(r)} - \overline{\boldsymbol{\varepsilon}}\right)^T \tag{51}$$

- Issues

  - Need to use perturbed observation introduces → additional source of error
  - Approach difficult to implement if m is large (→ sub-sampling or binning of the observations)

Ensemble Transform Kalman filter

- Formulation of the Kalman gain with the full observational error covariance matrix

$$\begin{aligned}
\mathbf{K} &= \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} && (52)\\
&= (\mathbf{S}^f\mathbf{S}^{f^T})\mathbf{H}^T\left[\mathbf{H}(\mathbf{S}^f\mathbf{S}^{f^T})\mathbf{H}^T + \mathbf{R}\right]^{-1} && (53)\\
&= \mathbf{S}^f(\mathbf{H}\mathbf{S}^f)^T\left[\mathbf{H}\mathbf{S}^f(\mathbf{H}\mathbf{S}^f)^T + \mathbf{R}\right]^{-1} && (54)\\
&= \mathbf{S}^f\left[\mathbf{I} + (\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f\right]^{-1}(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1} && (55)
\end{aligned}$$

- The Sherman-Morison-Woodbury identity has been applied from (54) to (55). This identity can be expressed as:

$$\mathbf{A}\mathbf{B}^T\left(\mathbf{C}+\mathbf{B}\mathbf{A}\mathbf{B}^T\right)^{-1} = \left(\mathbf{A}^{-1}+\mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{C}^{-1} \tag{56}$$

with $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = \mathbf{H}\mathbf{S}^f$, $\mathbf{C} = \mathbf{R}$.

- That is, instead of performing the inverse of an m by m matrix we need to perform only an inverse of a N by N matrix.

- The analysis covariance error $\mathbf{P}^a$:

$$\begin{aligned}
\mathbf{P}^a &= \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f && (57)\\
&= \mathbf{S}^f\mathbf{S}^{fT} - \mathbf{K}\mathbf{H}\mathbf{S}^f\mathbf{S}^{fT} && (58)\\
&= \mathbf{S}^f\mathbf{S}^{fT} - \mathbf{S}^f\left[\mathbf{I}+(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f\right]^{-1}(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}\mathbf{S}^{fT} && (59)\\
&= \mathbf{S}^f\left[\mathbf{I}-\left(\mathbf{I}+(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f\right)^{-1}(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}\right]\mathbf{S}^{fT} && (60)
\end{aligned}$$

- The goal is to find an expression like this:

$$\mathbf{P}^a = \mathbf{S}^a\mathbf{S}^{aT} \tag{61}$$

- This is possible when the following eigenvalue decomposition is made :

$$\left(\mathbf{H}\mathbf{S}^f\right)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{S}^f = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{62}$$

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 (62) in equation (60) one obtains:

$$\begin{aligned}
\mathbf{P}^a &= \mathbf{S}^f\left[\mathbf{I}-(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\right]\mathbf{S}^{fT} && (63)\\
&= \mathbf{S}^f\left[\mathbf{I}-(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T+\mathbf{I}-\mathbf{I}\right)\right]\mathbf{S}^{fT} && (64)\\
&= \mathbf{S}^f\left[\mathbf{I}-(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T+\mathbf{I}\right)+(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\right]\mathbf{S}^{fT} && (65)\\
&= \mathbf{S}^f\left[\mathbf{I}-\mathbf{I}+(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\right]\mathbf{S}^{fT} && (66)\\
&= \mathbf{S}^f(\mathbf{I}+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\mathbf{S}^{fT} && (67)\\
&= \mathbf{S}^f(\mathbf{U}\mathbf{U}^T+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1}\mathbf{S}^{fT} && (68)\\
&= \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1}\mathbf{U}^T\mathbf{S}^{fT} && (69)\\
&= \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1/2}(\mathbf{I}+\mathbf{\Lambda})^{-1/2}\mathbf{U}^T\mathbf{S}^{fT} && (70)
\end{aligned}$$

- 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.
- Solution: we multiply it by $\mathbf{U}^T$ (which does not change the product $\mathbf{S}^a\mathbf{S}^{aT}$):

$$\mathbf{S}^a = \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1/2}\mathbf{U}^T \tag{71}$$

- For a linear observation operator, the sum of all columns of $\mathbf{HS}^f$ is zero.

$$\mathbf{HS}^f\mathbf{1}_{N\times 1} = 0$$

- Thus $\mathbf{1}_{N\times 1}$ is a (unnormalized) eigenvector of $\left(\mathbf{HS}^f\right)^T\mathbf{R}^{-1}\mathbf{HS}^f$ with eigenvalue 0:

$$\left(\mathbf{HS}^f\right)^T\mathbf{R}^{-1}\mathbf{HS}^f\mathbf{1}_{N\times 1} = 0\ \mathbf{1}_{N\times 1} \tag{72}$$

- If eigenvalues are sorted in $\mathbf{\Lambda}$, then $\mathbf{1}_{N\times 1}$ is the smallest and Nth (last) eigenvalue as all eigenvalues positive:

$$\mathbf{U}\mathbf{e}_N = \frac{1}{\sqrt{N}}\mathbf{1}_{N\times 1} \tag{73}$$

$$\mathbf{\Lambda}_{N,N} = 0 \tag{74}$$

  where $\mathbf{e}_N$ is the a vector with all elements equal to zero except that last which is one. Therefore, it follows that

$$\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1/2}\mathbf{U}^T\mathbf{1}_{N\times 1} = \mathbf{1}_{N\times 1} \tag{75}$$

- Thus the mean of all columns of $\mathbf{S}^a$ is zero.

$\mathbf{S}^a$ is the square root of $\mathbf{P}^a$:

$$\mathbf{P}^a = \mathbf{S}^a\mathbf{S}^{aT} \tag{76}$$

The decomposition (62) can also be used in the computation of the Kalman gain $\mathbf{K}$ by:

$$\begin{aligned}
\mathbf{K} &= \mathbf{S}^f\left[\mathbf{I}+(\mathbf{HS}^f)^T\mathbf{R}^{-1}\mathbf{HS}^f\right]^{-1}(\mathbf{HS}^f)^T\mathbf{R}^{-1} & (77)\\
&= \mathbf{S}^f\left[\mathbf{U}\mathbf{U}^T+\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\right]^{-1}(\mathbf{HS}^f)^T\mathbf{R}^{-1} & (78)\\
&= \mathbf{S}^f\mathbf{U}(\mathbf{I}+\mathbf{\Lambda})^{-1}\mathbf{U}^T(\mathbf{HS}^f)^T\mathbf{R}^{-1} & (79)
\end{aligned}$$

The ensemble after the analysis will have the following mean:

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}^f}\right) \tag{80}$$

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

$$\mathbf{x}^{a(r)} = \overline{\mathbf{x}^a} + \sqrt{N-1}\ \left(\mathbf{S}^a\right)_r \tag{81}$$

Overview of Kalman filters suitable for large systems

Deduced from the Extended Kalman Filter ($\rightarrow$ linearized model for the errors evolution):

- SEEK: Pham *et al.* (1998). Evolutive error space
- RRSQRT: reduced rank approximation of the square root filter (reformulation of the Kalman filter)

Ensemble Kalman filters:

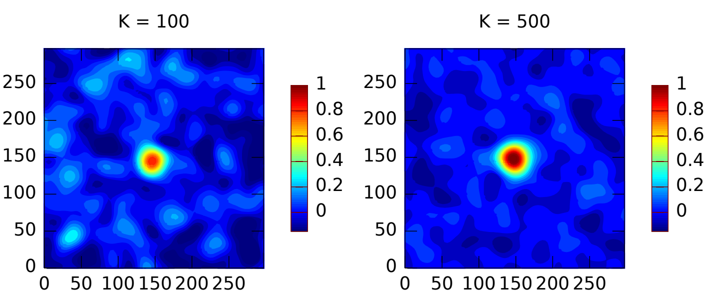
- SEIK Pham (2001). Evolutive error space (reformulation of SEEK using an ensemble)
- Ensemble Kalman filter Evensen (1994, 2007)
- Ensemble Transform Kalman Filter (Bishop *et al.*, 2001),
- Ensemble Adjustment Kalman Filter (Anderson, 2001)
- Error-subspace transform Kalman filter (Nerger *et al.*, 2012) (ESTKF)

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.

The need for localization and inflation

- For realistic ocean systems, only a relatively small number of ensemble members can be calculated.
- → systematically underestimated error variances (Whitaker and Hamill, 2002) (addressed by inflation)
- → spurious long-range correlations (addressed by localization)
- This can be illustrated also by using random perturbations whose spatial covariance decreases monotonically as a function of the distance.



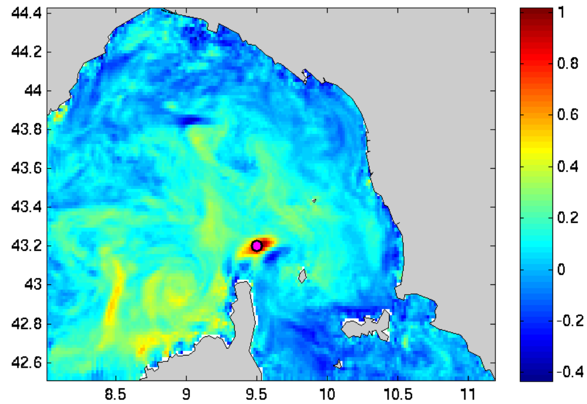
- Localization of the ensemble increment is therefore necessary to filter out spurious.

The need for localization

- Problematic spurious long-range correlations can be highlighted easily when assimilating a point measurement.

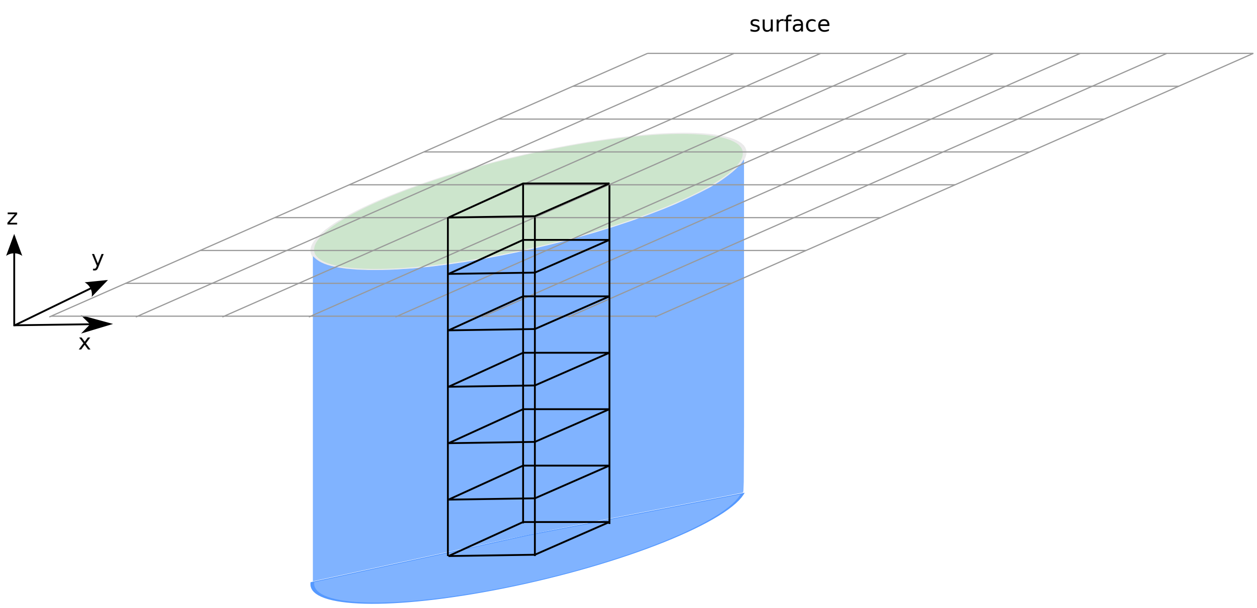
$$\mathbf{x}^a = \mathbf{x}^f + \underbrace{\mathbf{P}^f\mathbf{H}^T}_{\text{single column}} \underbrace{(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f)}_{\text{scalar}}$$

- Correction will be proportional to the covariance between the state vector at the observed location and all other model grid points
- Velocity covariance between location marked by the magenta circle and other model grid points



Localization

- **Domain localization**
  - the state vector is decomposed into sub-domains (e.g. single grid point or vertical column) where the assimilation is performed independently
  - Easily applied to parallel computers (Keppenne and Rienecker, 2003; Nerger and Hiller, 2013)
  - To avoid discontinuities in the analysis field, this approach is combined with the **observation localization** (Brankart *et al.*, 2003; Barth *et al.*, 2007; Hunt *et al.*, 2007)
  - The weight of distant observations (relative to the part of the state vector to be updated) is gradually decreased by increasing the error variance (observation localization or **R**-localization)



- **Covariance localization**:
  - Operates on the error covariance matrix $\mathbf{P}$ and it is sometimes called $\mathbf{P}$-localization
  - every single observation point is assimilated sequentially and the correction is filtered by a localization function.
  - Less suitable for parallel processing than the domain localization.

Inflation

$\mathbf{x}^{a\prime(r)}$ is the so-called r-th analysis ensemble perturbation: the different between the r-th analysis ensemble member and the analysis ensemble mean. $\mathbf{x}^{b\prime(r)}$ is defined in an analogous way for the background ensemble.

- multiplicative inflation (Anderson and Anderson, 1999):

$$\mathbf{x}^{a\prime(r)} \leftarrow \alpha \mathbf{x}^{a\prime(r)} \tag{82}$$

  where α is a positive inflation factor. Issue: where there are no observation, the error variance is still increased

- additive inflation (Houtekamer and Mitchell, 2005)

$$\mathbf{x}^{a\prime(r)} \leftarrow \mathbf{x}^{a\prime(r)} + \mathbf{d}^{\prime(r)} \tag{83}$$

  where the vector $\mathbf{d}'$ is a random vector of size n with a zero mean. In practice, it is difficult to come up with a suitable covariance matrix of the random vector $\mathbf{d}'$. The additive and multiplicative inflation scheme could also be applied to background ensemble.

- relaxation-to-prior perturbation (Zhang *et al.*, 2004):

$$\mathbf{x}^{a\prime(r)} \leftarrow (1-\alpha)\mathbf{x}^{a\prime(r)} + \alpha \mathbf{x}^{b\prime(r)} \tag{84}$$

  This approach avoids to increase the error variance where there are no observations.

- relaxation-to-prior spread (Whitaker and Hamill, 2012) is a form of multiplicative inflation with a space dependent inflation factor such that the ensemble standard deviation is relaxed towards the background ensemble standard deviation:

$$\boldsymbol{\sigma}^a \leftarrow (1-\alpha)\boldsymbol{\sigma}^a + \alpha \boldsymbol{\sigma}^b \tag{85}$$

  Every element of the vectors $\mathbf{x}^{a\prime(r)}$ is thus updated according to:

$$x_i^{a\prime(r)} \leftarrow x_i^{a\prime(r)} \left(\alpha \frac{\sigma_i^b - \sigma_i^a}{\sigma_i^a} + 1 \right) \tag{86}$$

Kalman smoother

- For a linear system, the Kalman filter provides the most accurate state given **all past observations**.
- The **Kalman smoother** provides the most accurate state for all past and future observations.
- Formally, one can derive the Kalman smoother by extending the state vector by the time dimension.
- The error covariances include also then space and time covariances.

$$\tilde{\mathbf{P}} = \begin{pmatrix} \mathbf{P}_{00} & \mathbf{P}_{01} & \dots & \mathbf{P}_{0N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{N_t0} & \mathbf{P}_{N_t1} & \dots & \mathbf{P}_{N_tN_t} \end{pmatrix}$$

- Every block $\mathbf{P}_{ij}$ is the covariance of the state vector at time i and j.

Kalman Smoother

- The Kalman Smoother analysis corresponds to the optimal interpolation analysis with the time-extended state vector.
- Extending the state vector by a time dimension can be implemented easily with ensemble methods: the vector $\mathbf{x}$ corresponds then to model trajectories
- For observations at high temporal frequency one can extend the state vector by several time instances within e.g. 24 hours and perform a smoother analysis within this time window and a simple filter analysis from one cycle to the next.

Particle filter

Bayes Theorem

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{p(\mathbf{y}^o|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}^o)} \tag{87}$$

- $p(\mathbf{x}|\mathbf{y}^o)$: 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:

$$p(\mathbf{y}^o|\mathbf{x}) = A \exp\left((\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x})) \right) \tag{88}$$

- $p(\mathbf{y}^o)$: 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) $\mathbf{x}^{(r)}$ ($r = 1, \dots, N$):

$$p(\mathbf{x}) = \frac{1}{N} \sum_{r=1}^{K} \delta(\mathbf{x} - \mathbf{x}^{(r)}) \tag{89}$$

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.

$$p(\mathbf{x}|\mathbf{y}^o) = \frac{1}{N} \sum_{r=1}^{N} w_r \delta(\mathbf{x} - \mathbf{x}^{(r)}) \tag{90}$$

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

- **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 tools

- 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)
- Empirical Gaussian Anamorphosis (the empirical transformation function such that a transformed variable follows a Gaussian distribution)
- Observation operator for HF radar surface currents
- Reference implementations of various ensemble analysis schemes

Non-sequential assimilation

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

4D-Var

- Minimization of the following cost function:

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

  with $\mathbf{x}_{k+1} = f_k(\mathbf{x}_k)$.

- 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 $\boldsymbol{\lambda}_k$:

$$\begin{aligned} \mathbf{x}_{k+1} &= f_k(\mathbf{x}_k) \\ \boldsymbol{\lambda}_{k-1} &= \mathbf{M}_k^T \boldsymbol{\lambda}_k + \mathbf{H}_k^T \mathbf{R}_k^{-1} \left(\mathbf{y}_k^o - h_k(\mathbf{x}_k)\right) \\ \boldsymbol{\lambda}_N &= 0 \end{aligned}$$

- The adjoint model is integrated backwards in time!

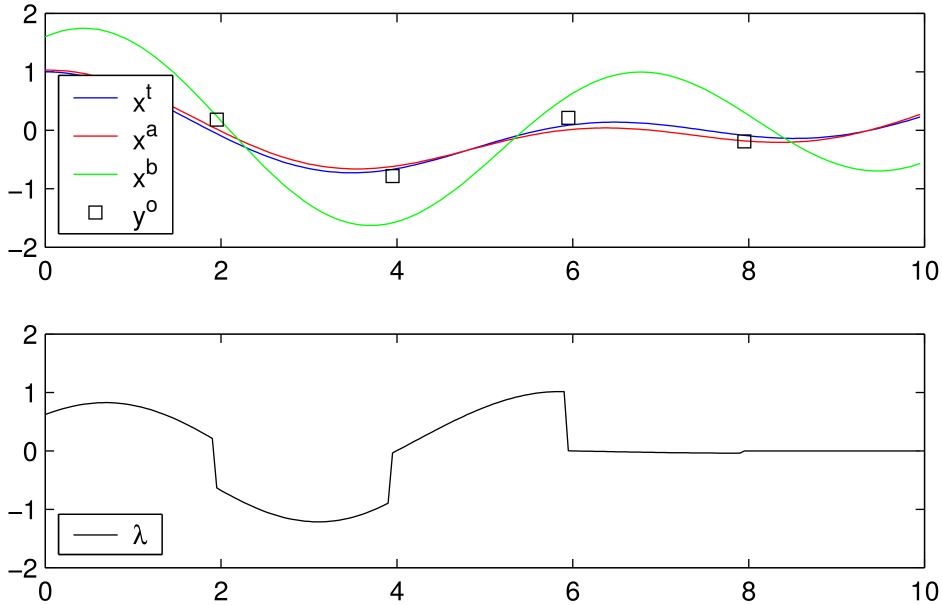


Figure 5: 4D-Var. The cost function is explicitly minimized; Upper panel: the observed component of the state vector $\mathbf{x}^t$ (truth), $\mathbf{x}^a$ (with assimilation) and $\mathbf{x}^b$ (without assimilation). After 20 iterations, the solution $\mathbf{x}^a$ is already quite close to the real 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 ($\mathbf{u}$) $\longrightarrow$ $\boxed{\text{Models } f_k \text{ and } h_k}$ $\longrightarrow$ 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:

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

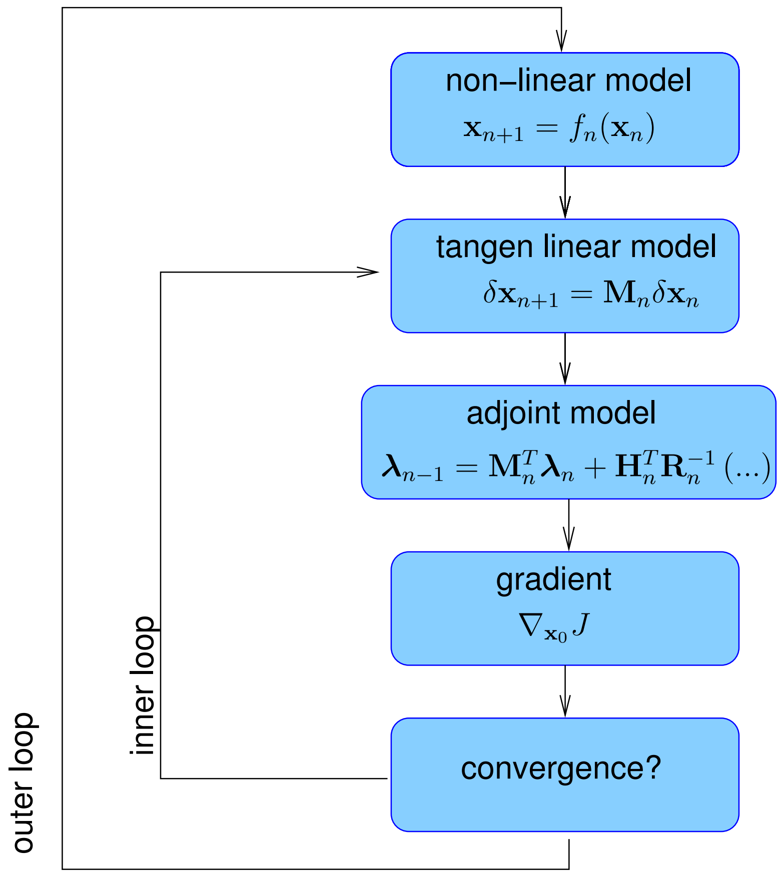
Incremental formulation

- Efficient algorithm to minimize a quadratic function
- Model and observation operators are linearized around first guess of the model trajectory → incremental formulation: (Courtier *et al.*, 1994; Courtier, 1997)

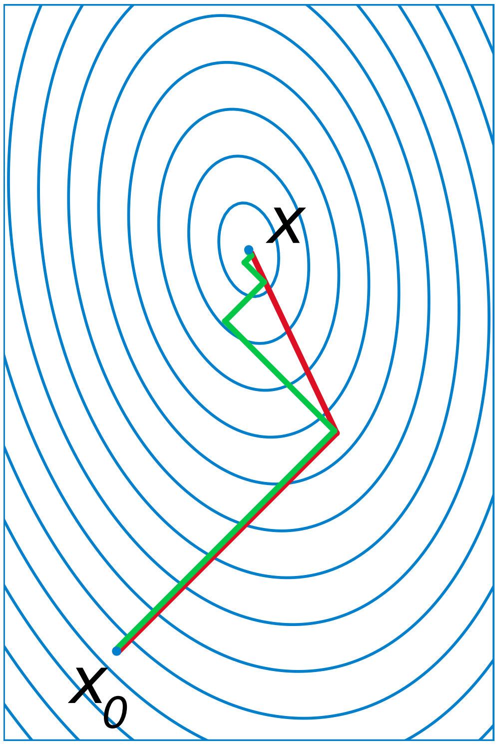
$$\begin{aligned} J(\delta \mathbf{x}_0) = & \left(\mathbf{x}_0 + \delta \mathbf{x}_0 - \mathbf{x}^i\right)^T \mathbf{P}^{i^{-1}} \left(\mathbf{x}_0 + \delta \mathbf{x}_0 - \mathbf{x}^i\right) \\ & + \sum_{k=1}^{N} \left(\mathbf{y}_k^o - h_k(\mathbf{x}_k) - \mathbf{H}_k \delta \mathbf{x}_k\right)^T \mathbf{R}_k^{-1}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

  with $\mathbf{x}_{k+1} = f_k(\mathbf{x}_k)$ and $\delta x_{k+1} = \mathbf{M}_k \delta x_k$.

- Minimize this function using the conjugate gradient method (inner loops)
- After the minimum is reached, a new model trajectory is computed with the full nonlinear 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
$\boldsymbol{\lambda}_{n-1} = \mathbf{M}_n^T\boldsymbol{\lambda}_n + \mathbf{H}_n^T\mathbf{R}_n^{-1}\,(...)$
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

$$f(\mathbf{x}) = f^{(p)}(\dots f^{(2)}(f^{(1)}(\mathbf{x}))) \tag{91}$$

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

$$\mathbf{F} = \mathbf{F}^{(p)} \dots \mathbf{F}^{(2)} \mathbf{F}^{(1)} \tag{92}$$

  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

$$d = ab + c \tag{93}$$

- 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:

$$\delta f = \frac{\partial f}{\partial a}\delta a + \frac{\partial f}{\partial b}\delta b + \frac{\partial f}{\partial c}\delta c \tag{94}$$

- For the example statement, one obtains:

$$\delta d = b\delta a + a\delta b + \delta c \tag{95}$$

Derivation of adjoint

- The adjoint is the transpose of the tangent linear model

$$\mathbf{F}^T = \mathbf{F}^{(1)^T}\mathbf{F}^{(2)^T}\ldots\mathbf{F}^{(p)^T} \tag{96}$$

- 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 δa^* 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

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

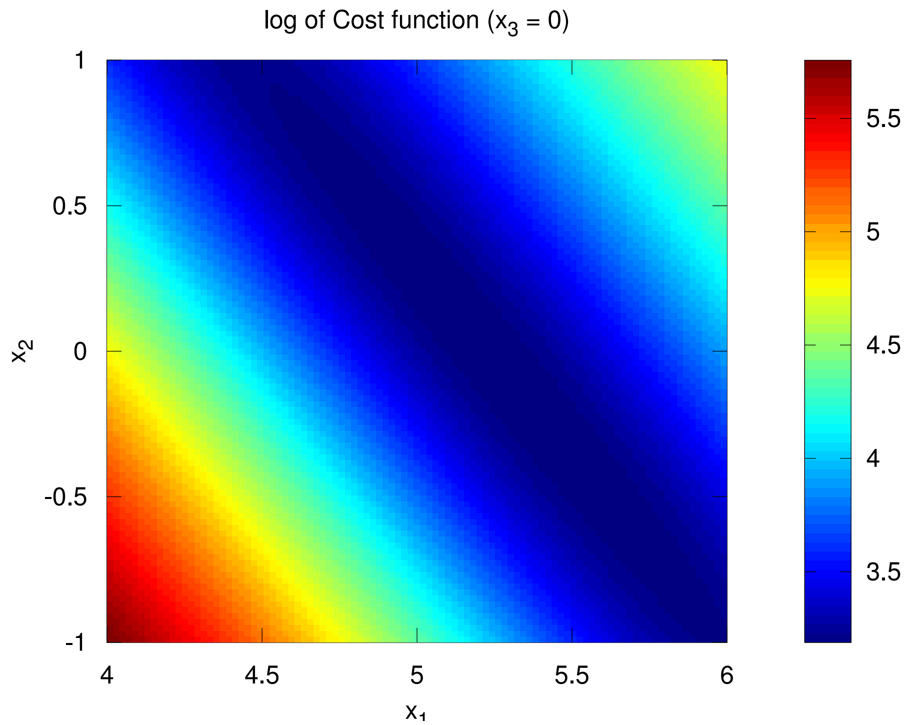
- For an adjoint integration one needs to have the state of the non-linear forward model
- All loops are reversed
- Automatic adjoint generators exist TAF, ADIFOR, (TAMC, OpenAD, ODYSSEE), Tapenade

Demo

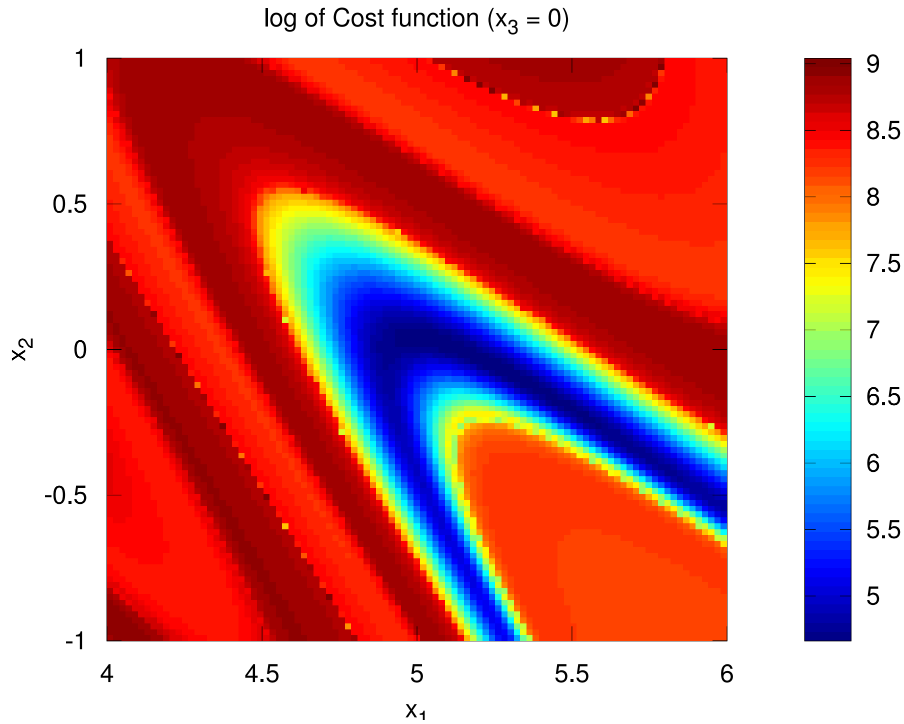
- Use the 4DVar method with the linear model and compare the results to the Kalman Filter
- Use the Lorenz model by varying number of time steps
- `http://data-assimilation.net/Tools/AssimDemo/`

Cost function of the Lorenz Model

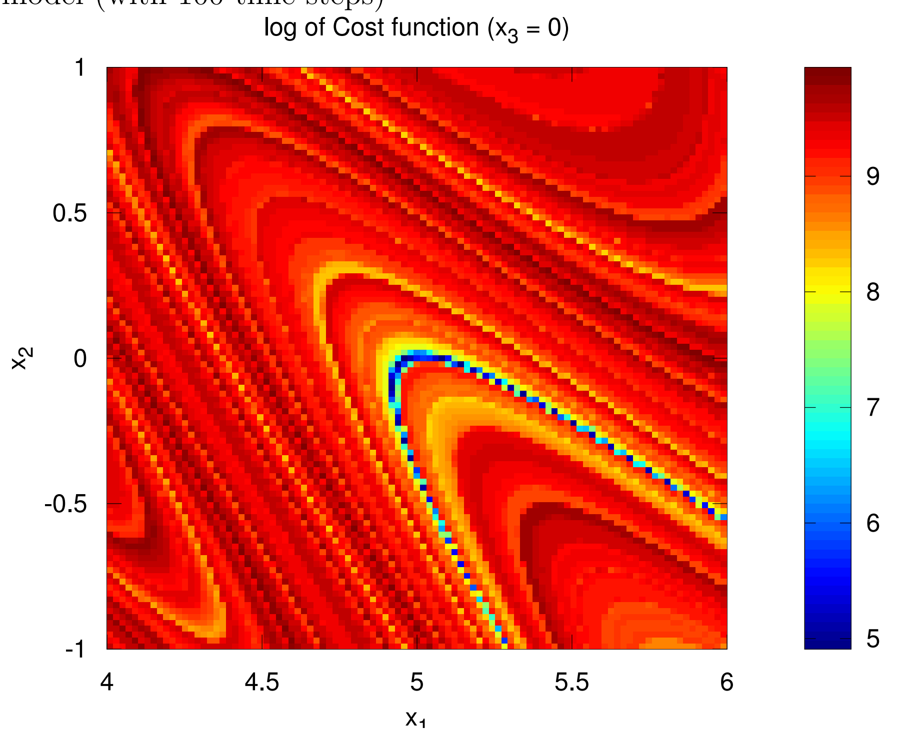
- Long assimilation window: more observations, but complex cost function
- Short assimilation window: less observations, but easier to minimize cost function
- Test: assimilate every model time step, true initial condition is $\mathbf{x} = (5, 0, 0)$
- Error variance of initial condition and observations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

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

Dual formulation

The optimal interpolation update:

$$
\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) && (97) \\
\mathbf{K} &= \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} && (98)
\end{aligned}
$$

Inverting a matrix of size $m \times m$ can be prohibitive. Inverting this matrix and multipling the inverse by the innovation vector $\mathbf{y}^o - \mathbf{H}\mathbf{x}^f$ is equivalent in solving the following linear system for the vector $\mathbf{w}$:

$$
(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})\mathbf{w} = \mathbf{y}^o - \mathbf{H}\mathbf{x}^f \qquad (99)
$$

The system can be solved by minimizing the following cost function for $\mathbf{w}$:

$$
J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{w} - \mathbf{w}^T\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \qquad (100)
$$

The final analysis in then obtained by:

$$
\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T\mathbf{w} \qquad (101)
$$

The Dual 4D Var problem can be formally derived from the Dual 3D Var assembling all observations into a single observation vector and the observation operator would then model operator so that the observation are compared to the model results from the right time. Applying this observation operator to a given vector $\mathbf{x}$ would thus require to run the foreward model and extracting the observed part of the model state vector. The product between the adjoint of the operation operator amounts of essentially running the adjoint of the model.

In addition the weak-constrained 4D Var problem can be derived by formally extending the vector $\mathbf{x}$ with the model error $\boldsymbol{\eta}_n$.

The representer method

First guess $\mathbf{x}_k^b$

$$\begin{aligned}\mathbf{x}_{k+1}^b &= \mathbf{M}_k\mathbf{x}_k^b + \mathbf{F}_k \\ \mathbf{x}_0 &= \mathbf{x}^i\end{aligned}$$

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

$$\begin{aligned}\boldsymbol{\Lambda}_{k-1n'} &= \mathbf{M}_k^T\boldsymbol{\Lambda}_{kk'} + \mathbf{H}_k^T\delta_{kk'} \\ \boldsymbol{\Lambda}_{Nk'} &= 0\end{aligned}$$

Representers $\tilde{\mathbf{R}}_{kk'}$

$$\begin{aligned}\tilde{\mathbf{R}}_{k+1n'} &= \mathbf{M}_k\tilde{\mathbf{R}}_{kk'} + \mathbf{Q}_k\boldsymbol{\Lambda}_{kk'} \\ \tilde{\mathbf{R}}_{0n'} &= \mathbf{P}^i\mathbf{M}_0^T\boldsymbol{\Lambda}_{0n'}\end{aligned}$$

Corrections $\mathbf{b}_k$

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

Analysis $\mathbf{x}_n$

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{M}_k\mathbf{x}_k + \mathbf{F}_k + \mathbf{Q}_k\boldsymbol{\lambda}_k \\ \mathbf{x}_0 &= \mathbf{x}^i + \mathbf{P}^i\mathbf{M}_0^T\boldsymbol{\lambda}_0 \\ \boldsymbol{\lambda}_{k-1} &= \mathbf{M}_k^T\boldsymbol{\lambda}_k + \mathbf{H}_k^T\mathbf{b}_k \\ \boldsymbol{\lambda}_N &= 0\end{aligned}$$

Interpretation

- The representers are covariances:

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

- 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 / 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:

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

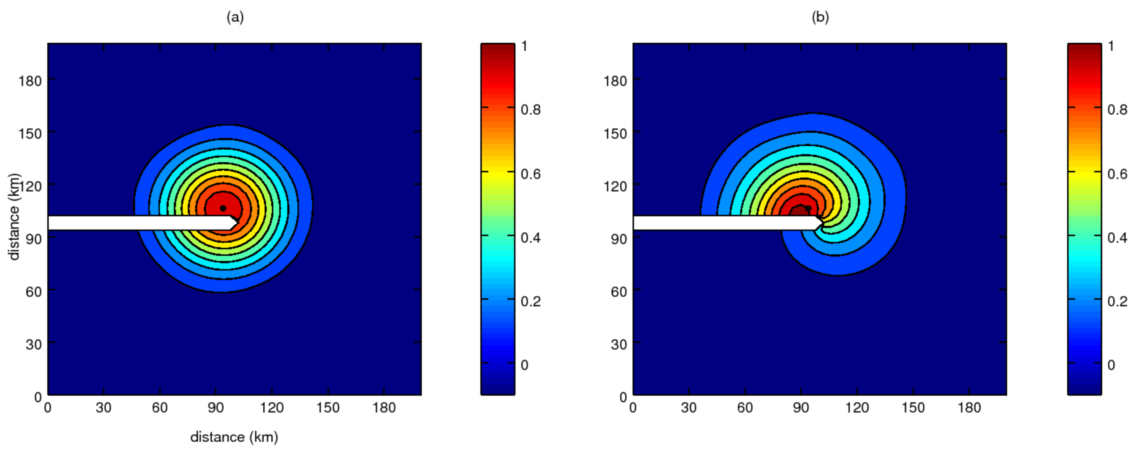
- The cost function can be used to define the probability of a perturbation $\mathbf{x}$ (*e.g.* Kalnay, 2002):

$$p(\mathbf{x}) = \alpha \exp\left(-J(\mathbf{x})\right) \tag{102}$$

- Perturbations are derived from the Hessian matrix of J (Barth *et al.*, 2009).

Impact of barriers

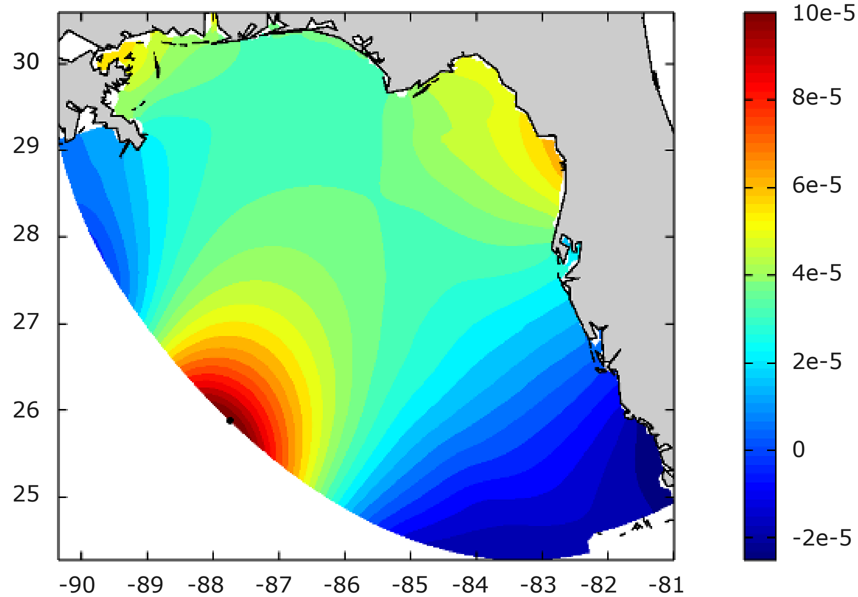
- The "smoothness" constraint is implemented through a diffusion operator (laplacian), it takes thus the land-sea mask into account



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

Harmonic shallow water equations

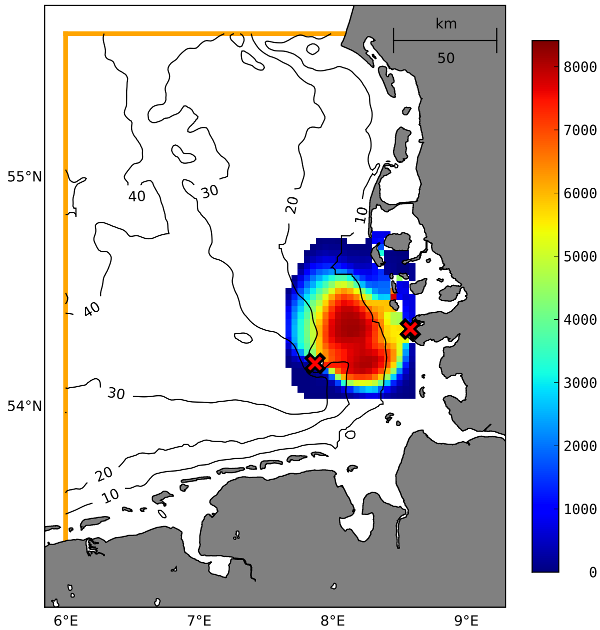
- For tidal models, perturbations should be approximately a harmonic solution to the shallow water equations



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

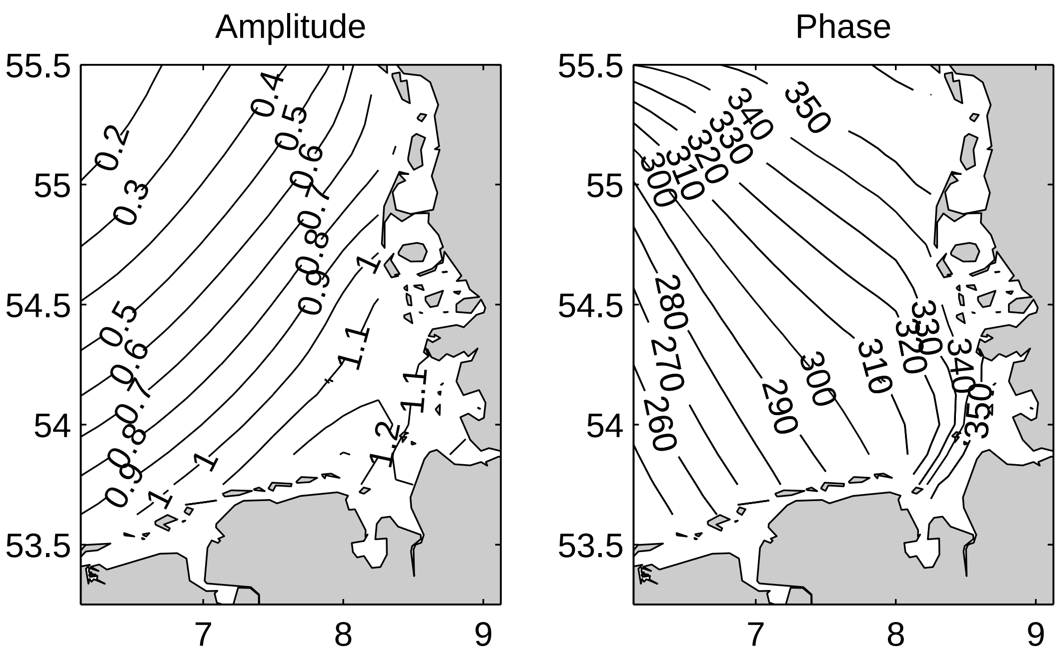
German Bight model

- General Estuarine Ocean Model (GETM Burchard and Bolding, 2002)
- 3-D primitive equations with a free-surface
- 21 σ 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).



- 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)

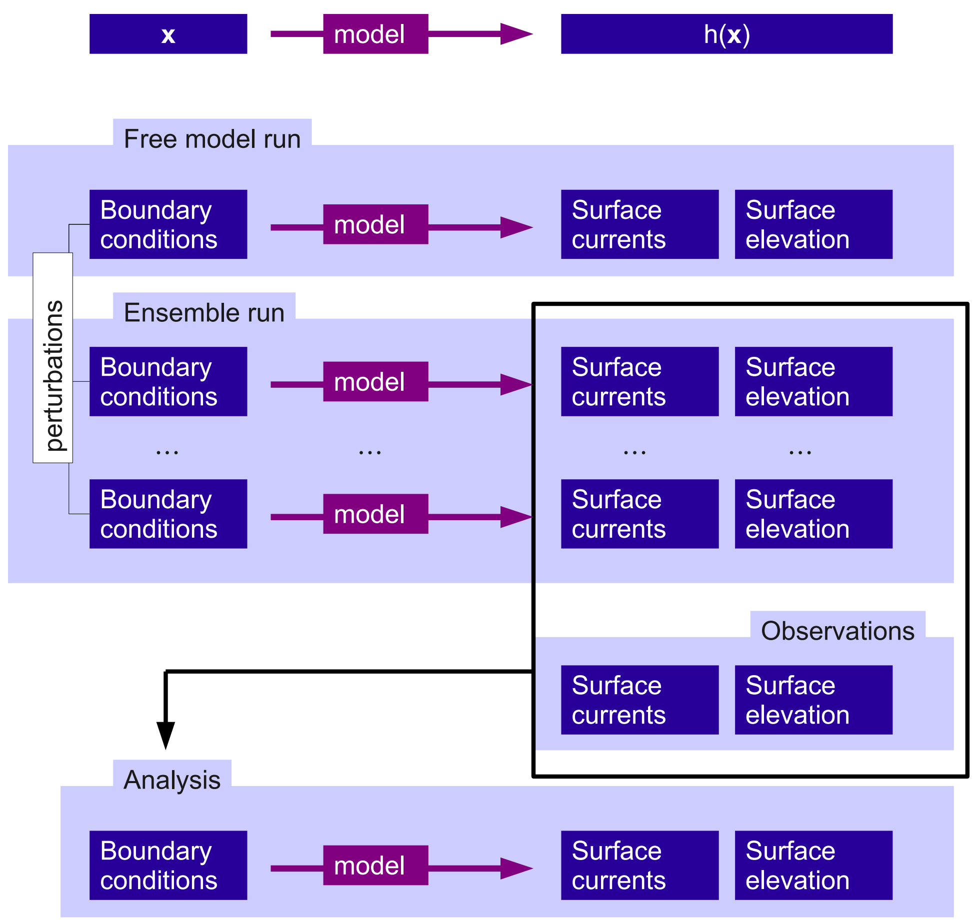


- M2 amplitude (in m) and phase (in degrees) of EOT08a for the German Bight based on altimetry.
- complex tidal parameters are assimilated

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:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed forcing} \tag{103}$$



Smoother scheme

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

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A}\left(\mathbf{B} + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - h(\mathbf{x}^b)\right) \tag{104}$$

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

$$\begin{aligned}
\mathbf{A} = \mathrm{cov}(\mathbf{x}^b, h(\mathbf{x}^b)) &= \left\langle \left(\mathbf{x} - \langle \mathbf{x} \rangle\right) \left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right)^T \right\rangle & (105)\\
\mathbf{B} = \mathrm{cov}(h(\mathbf{x}^b), h(\mathbf{x}^b)) &= \left\langle \left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right) \left(h(\mathbf{x}) - \langle h(\mathbf{x}) \rangle\right)^T \right\rangle & (106)
\end{aligned}$$

  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 (104) minimizes,

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T {\mathbf{P}^b}^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^o - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^o - h(\mathbf{x})) \tag{107}$$

  or

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T {\mathbf{P}^b}^{-1} (\mathbf{x} - \mathbf{x}^b) + \sum_n (\mathbf{y}_n^o - (h(\mathbf{x})_n))^T {\mathbf{R}_n}^{-1} (\mathbf{y}_n^o - (h(\mathbf{x})_n)) \tag{108}$$

  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.

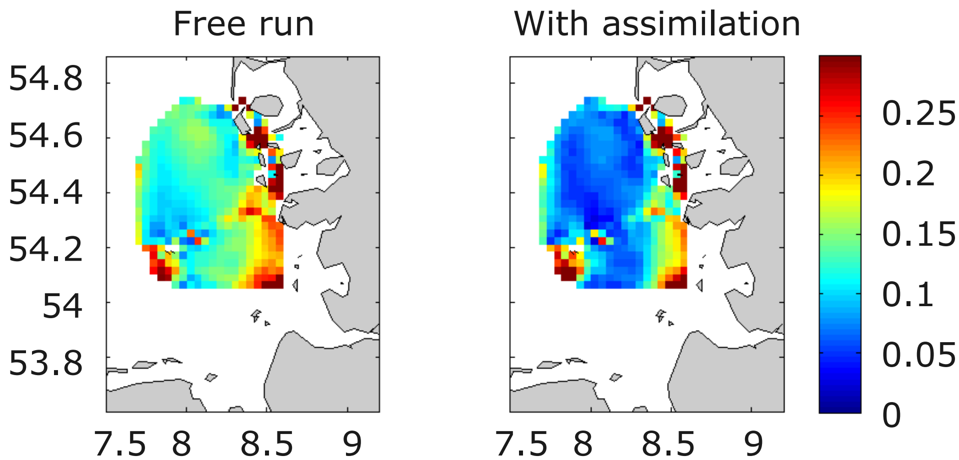
- Approach is closely related to Ensemble Smoother (van Leeuwen, 2001), 4D-EnKF (Hunt *et al.*, 2007) and AEnKF (Sakov *et al.*, 2010) where model trajectories instead of model states are optimized and to the Green's method with stochastic "search directions"

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

RMS difference

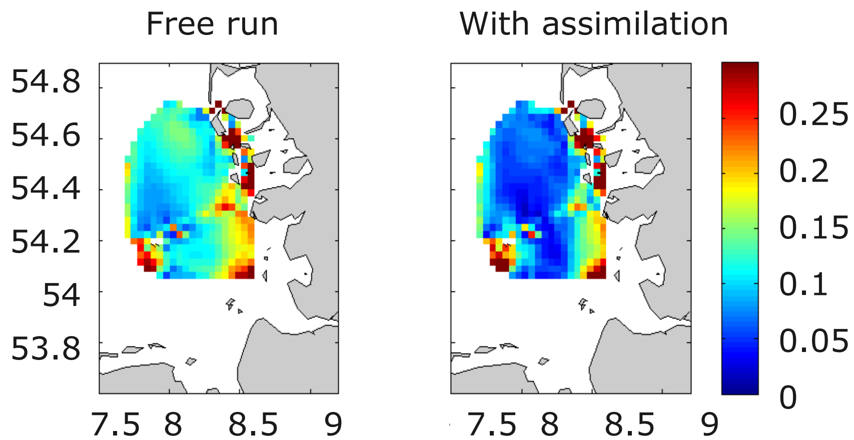
$$\mathrm{RMS}^2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T (A\cos(\omega t-\phi) - A'\cos(\omega t-\phi'))^2 dt \tag{109}$$

$$= \frac{A^2 + A'^2}{2} - AA'\cos(\phi-\phi') \tag{110}$$



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

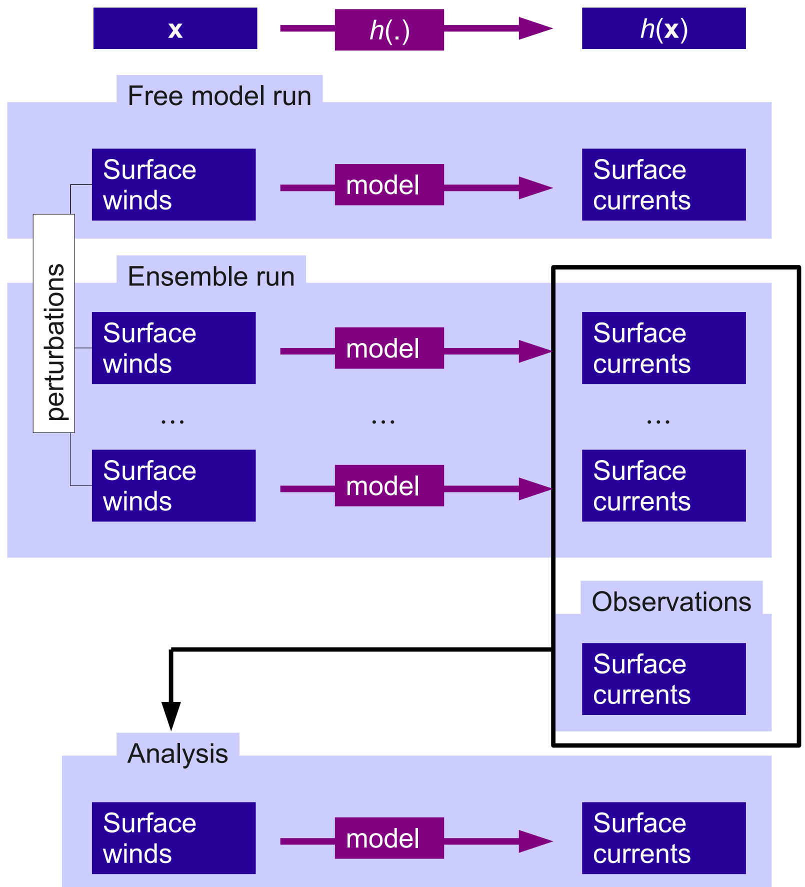
Table 1: Comparison with tide gage observations. Amplitude is in m and phase in degrees.

- 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 (Barth *et al.*, 2010).

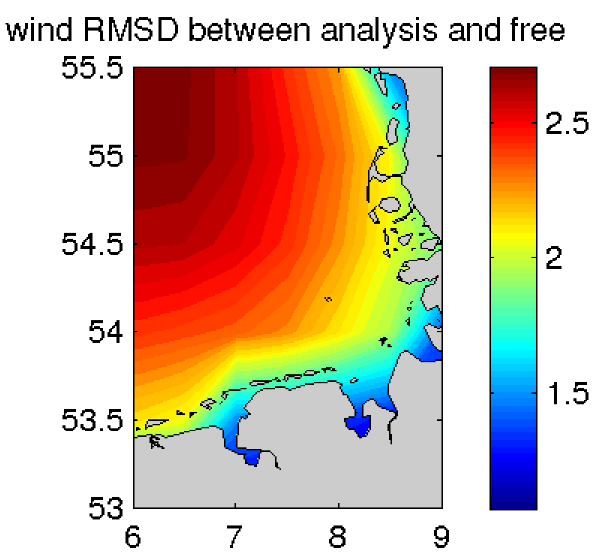
Wind estimation from HF radar observations

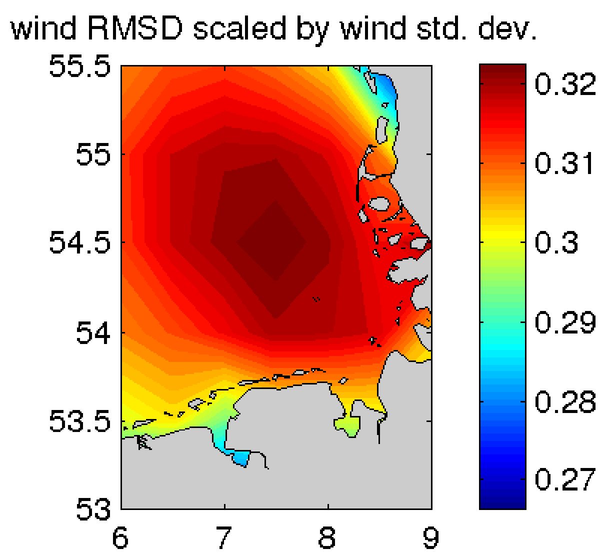
- Ensemble of 100 wind forcings are created (by using a Fourier decomposition)
- estimation vector $\mathbf{x}$: u- and v- component of wind forcing
- observations: $\mathbf{y}^o$: surface currents
- "observation operator" $h(\cdot)$:

$$h(\cdot) = \text{Interpolation to obs. location} \circ \text{Model integration with perturbed wind} \quad (111)$$



Time-averaged wind correction statistics





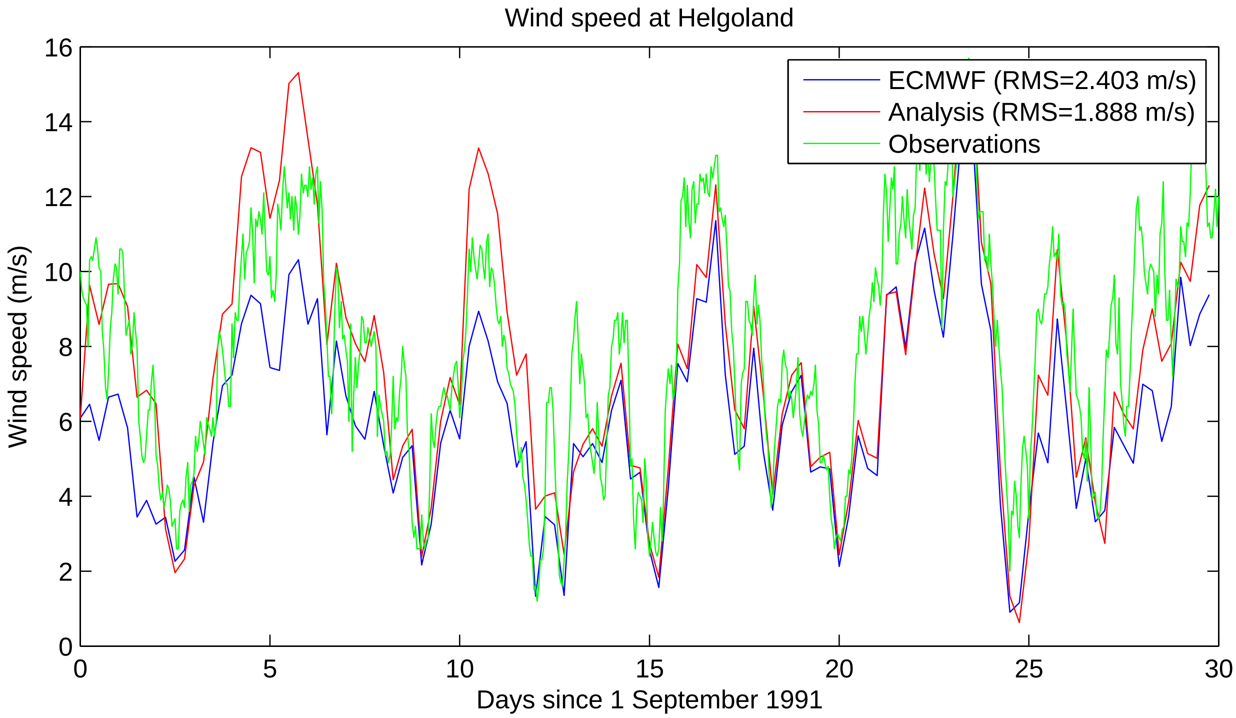


Figure 6: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Helgoland. Units are m/s.

- RMS difference between analyzed winds and ECMWF winds (averaged over time)
- RMS difference scaled by wind standard deviation (Barth *et al.*, 2011)

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

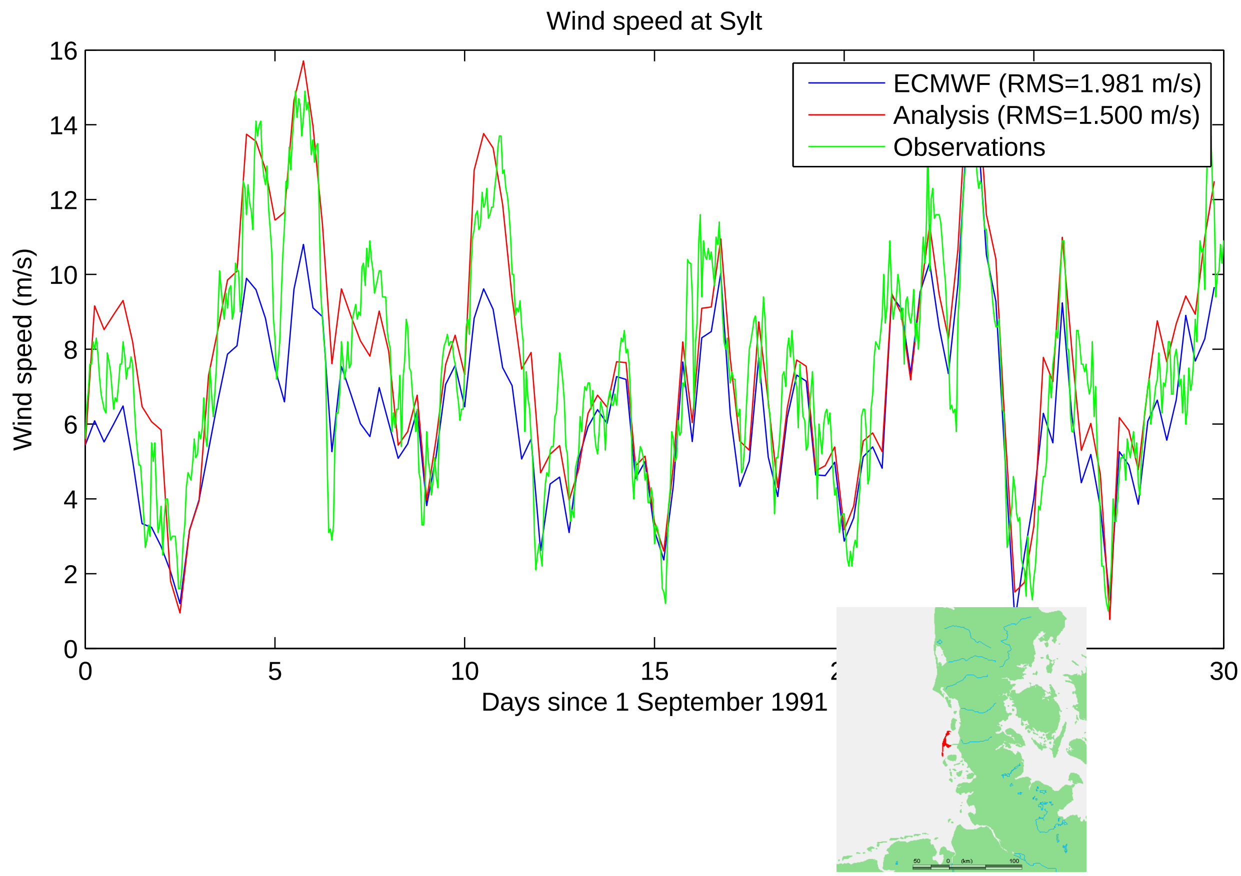
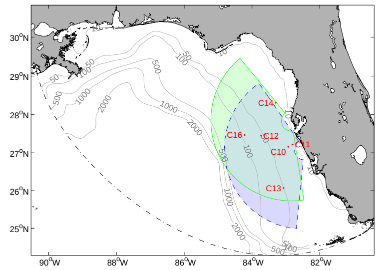
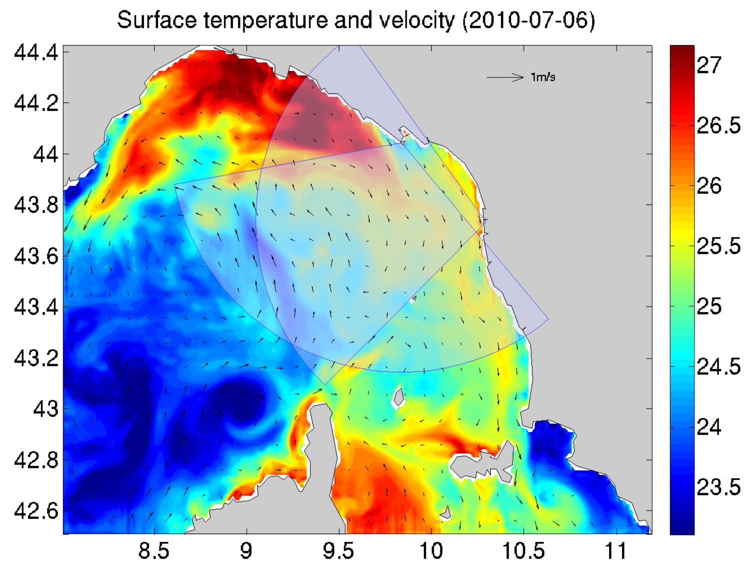


Figure 7: Measured wind speed, wind speed from ECMWF and analyzed wind speed at Sylt. Units are m/s.

Ligurian Sea Model

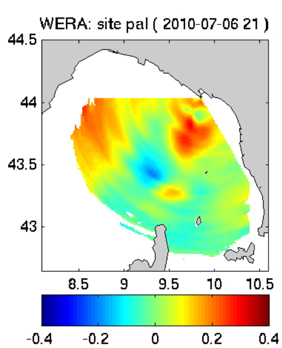
- ROMS nested (off-line) in Mediterranean Ocean Forecasting System
- 1/60 degree resolution and 32 vertical levels
- Currents: Western & Eastern Corsican Current, Northern Current, inertial oscillation, mesoscale currents
- Two WERA HF radar systems (Palmaria, San Rossore) by NATO-CMRE from 2009 to 2010.

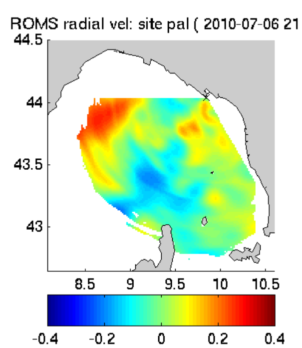




Observations

- Frequency of $\nu = 12.359$ MHz and coupled to a wave length of $\lambda_b = 12.13$ m,
- Radial currents are measured and used for the assimilation
- Angular resolution of 6 degrees, radial resolution of 2.4 km
- Currents are averaged over 1 h

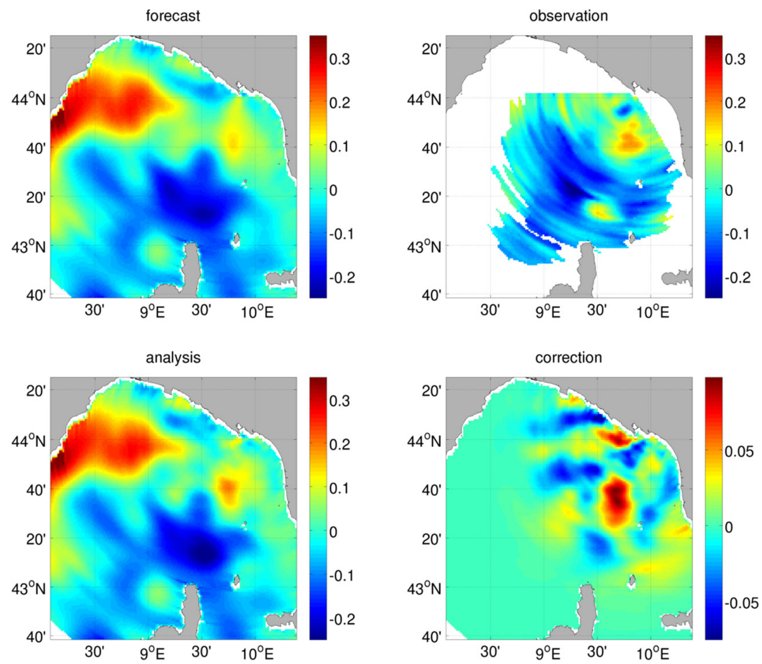




Radial currents on 2010-07-06 21:30 relative to the Palmaria site: left panel shows WERA measurements and right panel shows ROMS results without assimilation.

Corrections

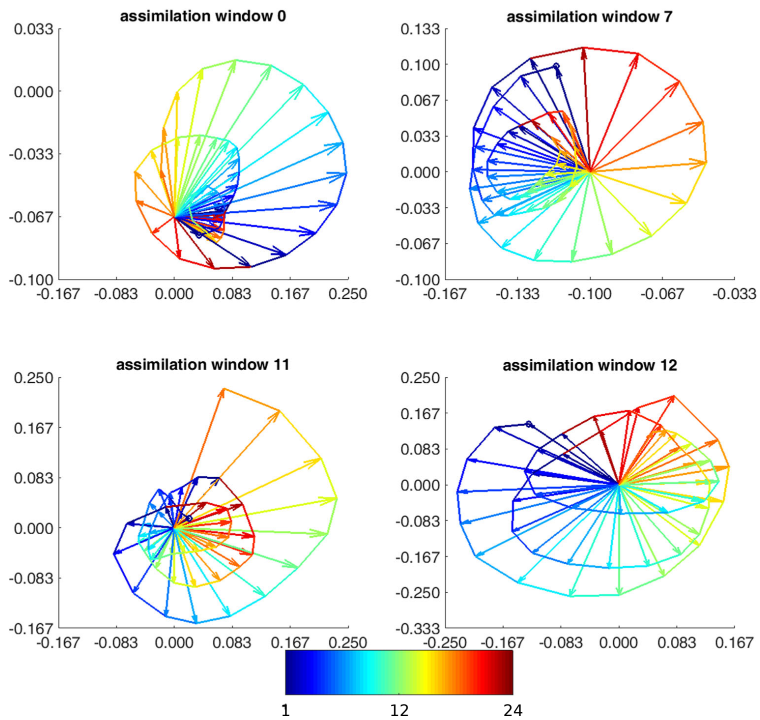
- To correct the eddy field relatively large ensembles are needed
- If no ensemble member predicts an eddy at a given location, the ensemble analysis will not be able to introduce an eddy at a given location



(Vandenbulcke *et al.*, 2017)

Temporal covariances

- By including several time instances into a state-vector, one can also analyse the temporal covariance
- Here: inertial oscillations



Summary of sequential methods

Kalman smoother, 4D-Var, Representer method

$\downarrow$ *only past data*

Kalman filter

$\downarrow$ *constant* $\mathbf{P}_k^f$

Optimal Interpolation, 3D-Var

$\downarrow$ $\mathbf{R}$ *and* $\mathbf{P}_k^f$ *diagonal*

Nudging

$\downarrow$ $\mathbf{R} \to 0$

Direct Insertion

Bibliography

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

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127**, 2741–2758, doi:10.1175/1520-0493(1999)127¡2741:AMCIOT¿2.0.CO;2.

Barth, A., A. Alvera-Azcárate, J.-M. Beckers, M. Rixen, and L. Vandenbulcke, 2007: Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea. *Journal of Marine Systems*, **65**, 41–59, doi:10.1016/j.jmarsys.2005.07.006.
URL `http://hdl.handle.net/2268/4260`

Barth, A., A. Alvera-Azcárate, J.-M. Beckers, J. Staneva, E. V. Stanev, and J. Schulz-Stellenfleth, 2011: Correcting surface winds by assimilating High-Frequency Radar surface currents in the German Bight. *Ocean Dynamics*, **61**, 599–610, doi:10.1007/s10236-010-0369-0.
URL `http://hdl.handle.net/2268/83330`

Barth, A., A. Alvera-Azcárate, K.-W. Gurgel, J. Staneva, A. Port, J.-M. Beckers, and E. V. Stanev, 2010: Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents - application to the German Bight. *Ocean Science*, **6**, 161–178, doi:10.5194/os-6-161-2010.

Barth, A., A. Alvera-Azcárate, J.-M. Beckers, R. H. Weisberg, L. Vandenbulcke, F. Lenartz, and M. Rixen, 2009: Dynamically constrained ensemble perturbations - application to tides on the West Florida Shelf. *Ocean Science*, **5**, 259–270, doi:10.5194/os-5-259-2009.

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, doi:10.1175/1520-0493(2001)129¡0420:ASWTET¿2.0.CO;2.

Brankart, J.-M., C.-E. Testut, P. Brasseur, and J. Verron, 2003: Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: application to a 1993-96 hindcast of the North Atlantic Ocean circulation. *Journal of Geophysical Research*, **108**, 3074, doi:10.1029/2001JC001198.

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

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126¡1719:ASITEK¿2.0.CO;2.

Capet, A., F. J. Meysman, I. Akoumianaki, K. Soetaert, and M. Grégoire, 2016: Integrating sediment biogeochemistry into 3D oceanic models: A study of benthic-pelagic coupling in the Black Sea. *Ocean Modelling*, **101**, 83 – 100, doi:https://doi.org/10.1016/j.ocemod.2016.03.006.
URL `http://www.sciencedirect.com/science/article/pii/S146350031630004X`

Chen, S. S. and M. Curcic, 2016: Ocean surface waves in hurricane ike (2008) and superstorm sandy (2012): Coupled model predictions and observations. *Ocean Modelling*, **103**, 161 – 176, doi:10.1016/j.ocemod.2015.08.005, waves and coastal, regional and global processes.

Chen, Y. and C. Snyder, 2007: Assimilating Vortex Position with an Ensemble Kalman Filter. *Monthly Weather Review*, **135**, 1828–1845, doi:10.1175/MWR3351.1.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **123**, 2449–2461, doi:10.1002/qj.49712354414.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367–1387, doi:10.1002/qj.49712051912.

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, doi:10.1029/94JC00572.

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

Franz, G., M. T. Delpey, D. Brito, L. Pinto, P. Leitão, and R. Neves, 2017: Modelling of sediment transport and morphological evolution under the combined action of waves and currents. *Ocean Science*, **13**, 673–690, doi:10.5194/os-13-673-2017.
URL `https://www.ocean-sci.net/13/673/2017/`

Gurgel, K.-W., 1994: Shipborne measurement of surface current fields by HF radar (extended version). *L'Onde Electrique*, **74**, 54–59.

Gurgel, K.-W., H.-H. Essen, and S. P. Kingsley, 1999: HF radars: Physical limitations and recent developments. *Coastal Engineering*, **37**, 201–218, doi:10.1016/S0378-3839(99)00026-5.

Houtekamer, P. L. and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3269–3289, doi:10.1256/qj.05.135.

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, doi:10.1016/j.physd.2006.11.008.

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

Keppenne, C. L. and M. M. Rienecker, 2003: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *Journal of Marine Systems*, **40-41**, 363 – 380, doi:10.1016/S0924-7963(03)00025-3.

Nerger, L. and W. Hiller, 2013: Software for ensemble-based data assimilation systems-Implementation strategies and scalability. *Computers & Geosciences*, **55**, 110 – 118, doi:10.1016/j.cageo.2012.03.026.

Nerger, L., T. Janjić, J. Schröter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based kalman filters. *Quarterly Journal of the Royal Meteorological Society*, **138**, 802–812, doi:10.1002/qj.945.

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.

Sagen, H., B. D. Dushaw, E. K. Skarsoulis, D. Dumont, M. A. Dzieciuch, and A. Beszczynska-Möller, 2016: Time series of temperature in fram strait determined from the 2008–2009 damocles acoustic tomography measurements and an ocean model. *Journal of Geophysical Research: Oceans*, **121**, 4601–4617, doi:10.1002/2015JC011591.

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.

Vandenbulcke, L., J.-M. Beckers, and A. Barth, 2017: Correction of inertial oscillations by assimilation of HF radar data in a model of the Ligurian Sea. *Ocean Dynamics*, 117–135, doi:10.1007/s10236-016-1012-5.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed obsevations. *Monthly Weather Review*, **130**, 1913–1924, doi:10.1175/1520-0493(2002)130¡1913:EDAWPO¿2.0.CO;2.

— 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132¡1238:IOIEAO¿2.0.CO;2.