DIVA and DINEOF

Overview

Jean-Marie Beckers and GHER group Aida Alvera, Alexander Barth, Charles Troupin, Mohamed Ouberdous, Sylvain Watelet, Igor Tomazic

MARE INTERFACULTARY CENTER FOR MARINE RESEARCH - LIEGE UNIVERSITY http://modb.oce.ulg.ac.be/GHER MARE INTERFACULTARY CENTER FOR MARINE RESEARCH - LIEGE UNIVERSITY

Université de Liège

MARE-GHER Sart-Tilman B5

4000 Liège, Belgique

Université de Liège

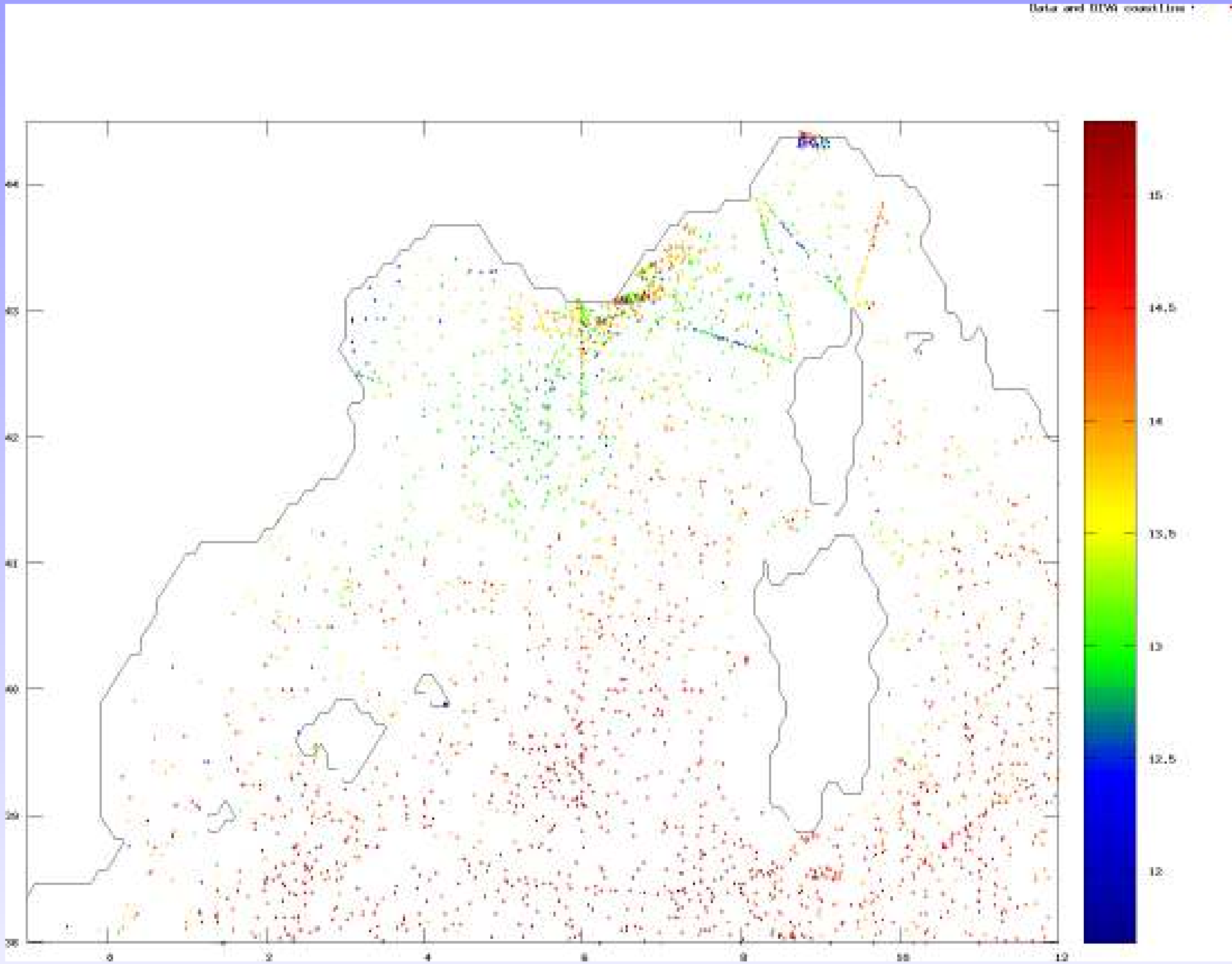
Outline

- ***Gridding***
- ***DIVA theory***
- ***Implementations and exercises***
- ***DINEOF theory***
- ***Implementations and exercises***

Université de Liège

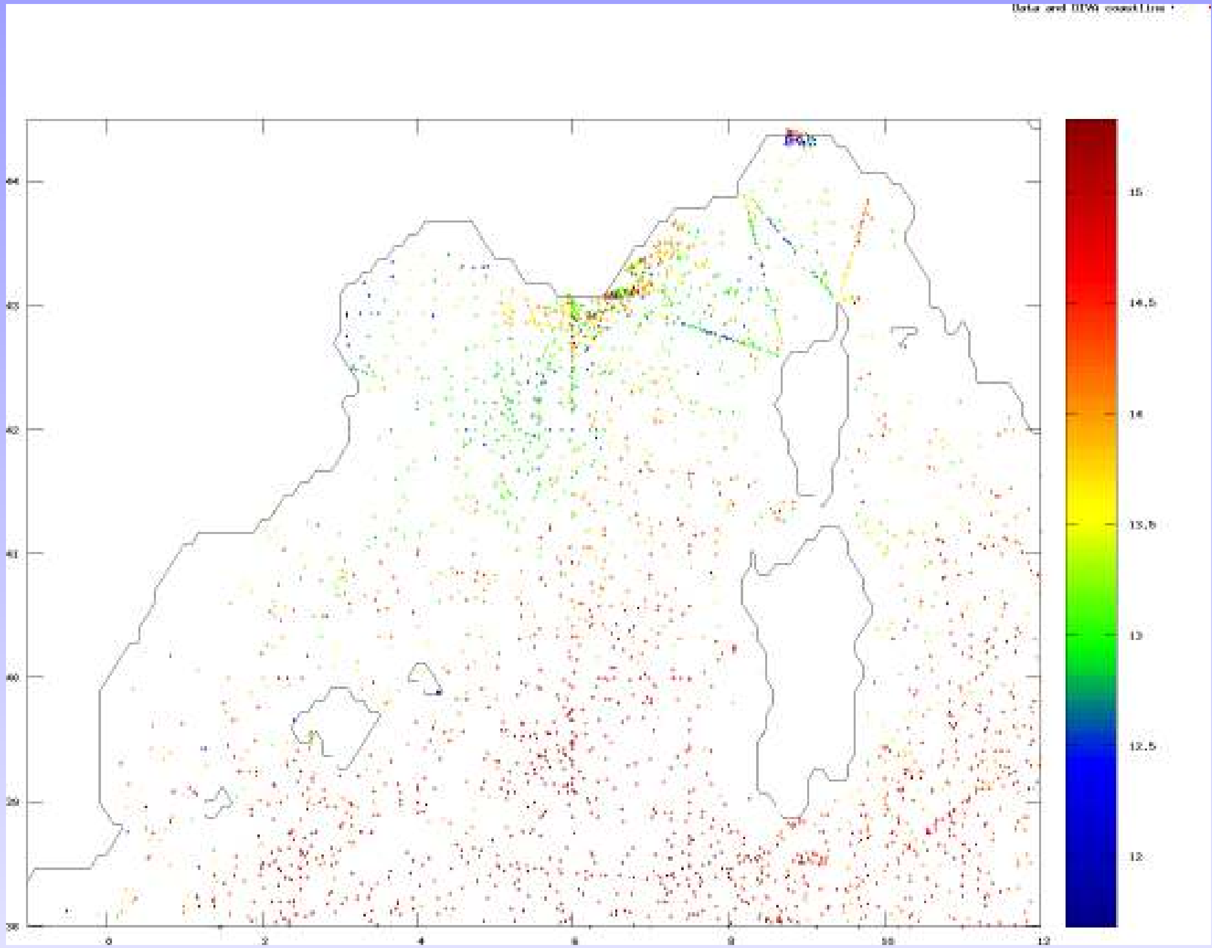
- ***Gridding***
- *DIVA theory*
- *Implementations and exercises*
- *DINEOF theory*
- *Implementations and exercises*

Common problem



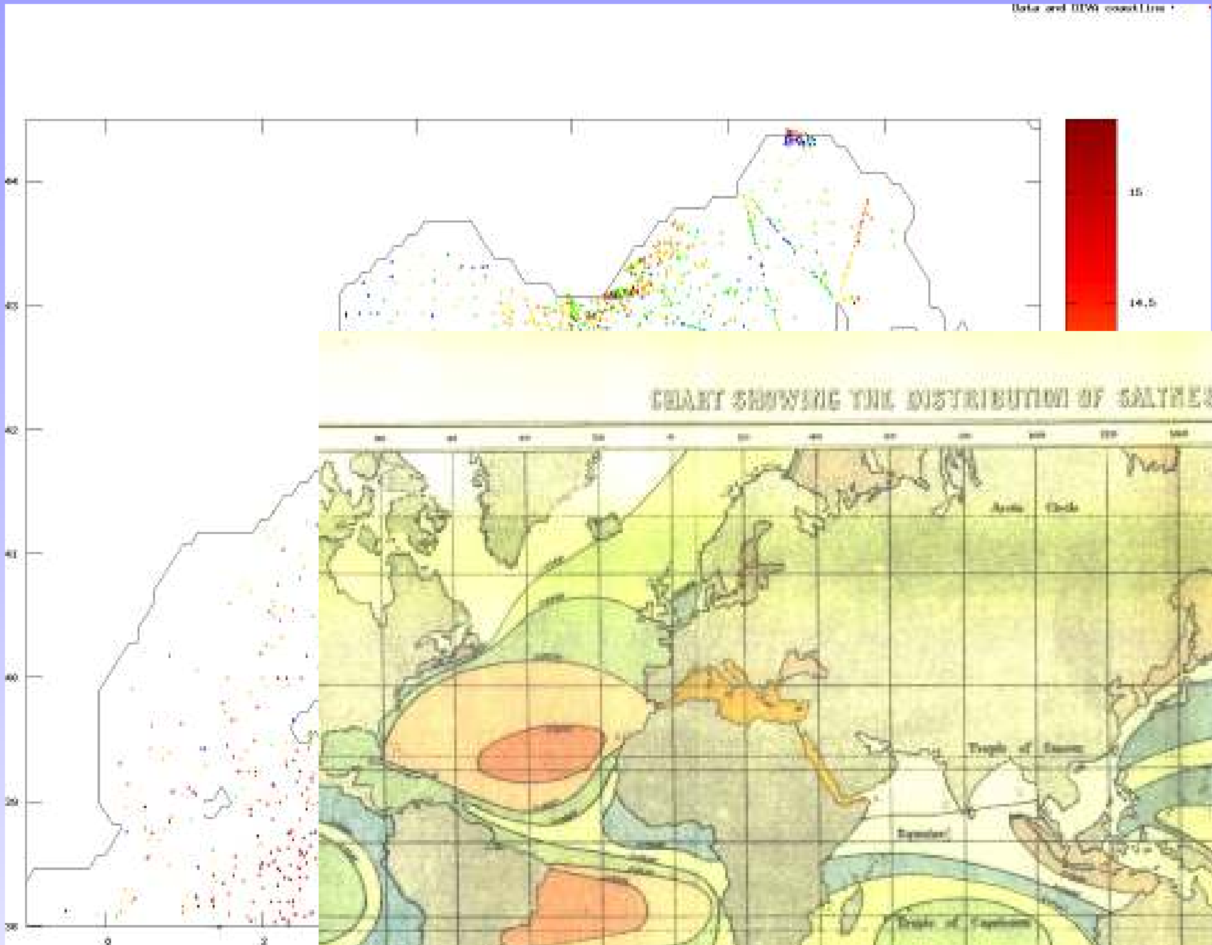
Appears when trying to produce maps, calculate volume averages, prepare initial conditions for models, quality control of data ...

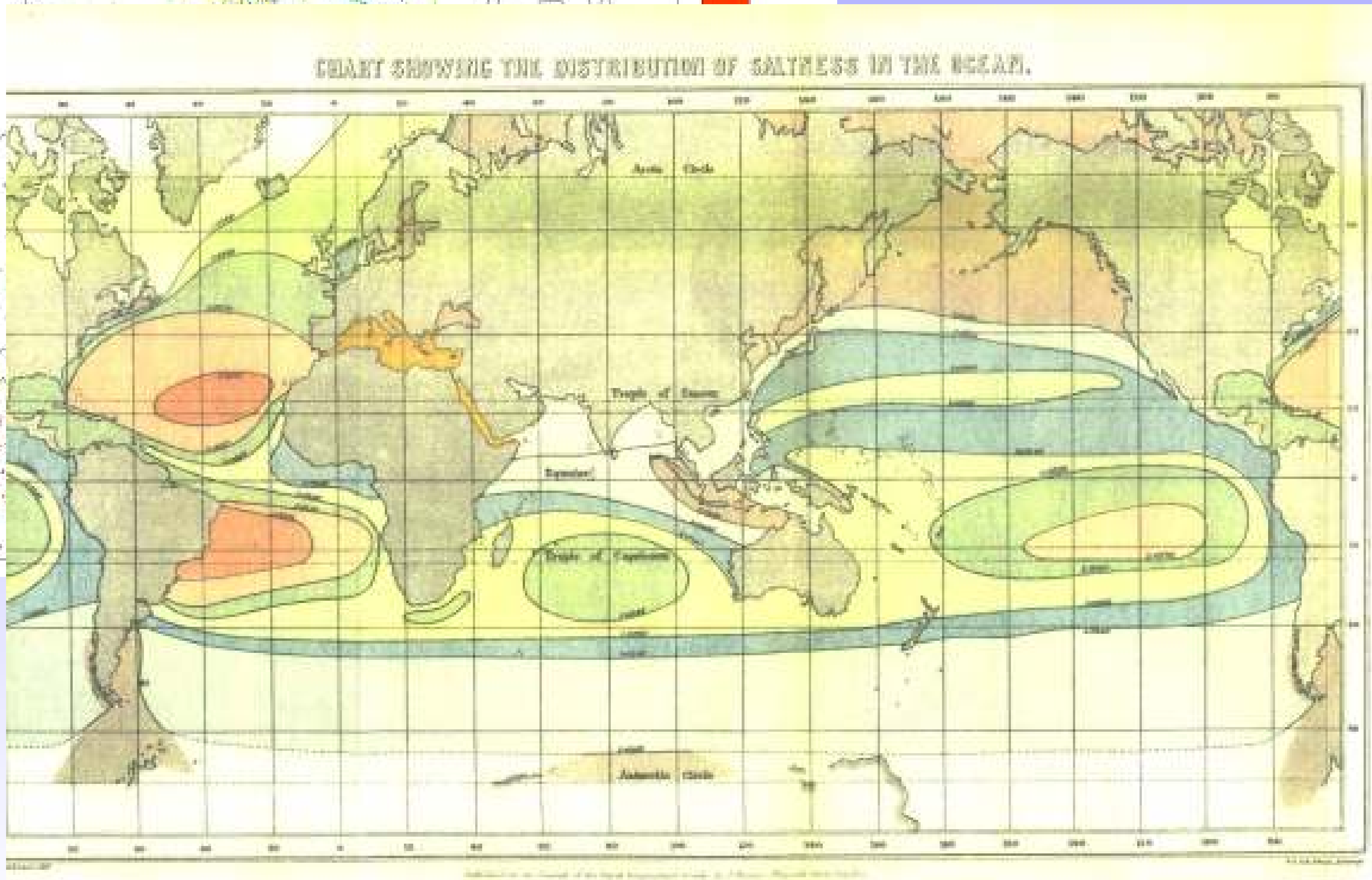
Solutions



Université de Liège

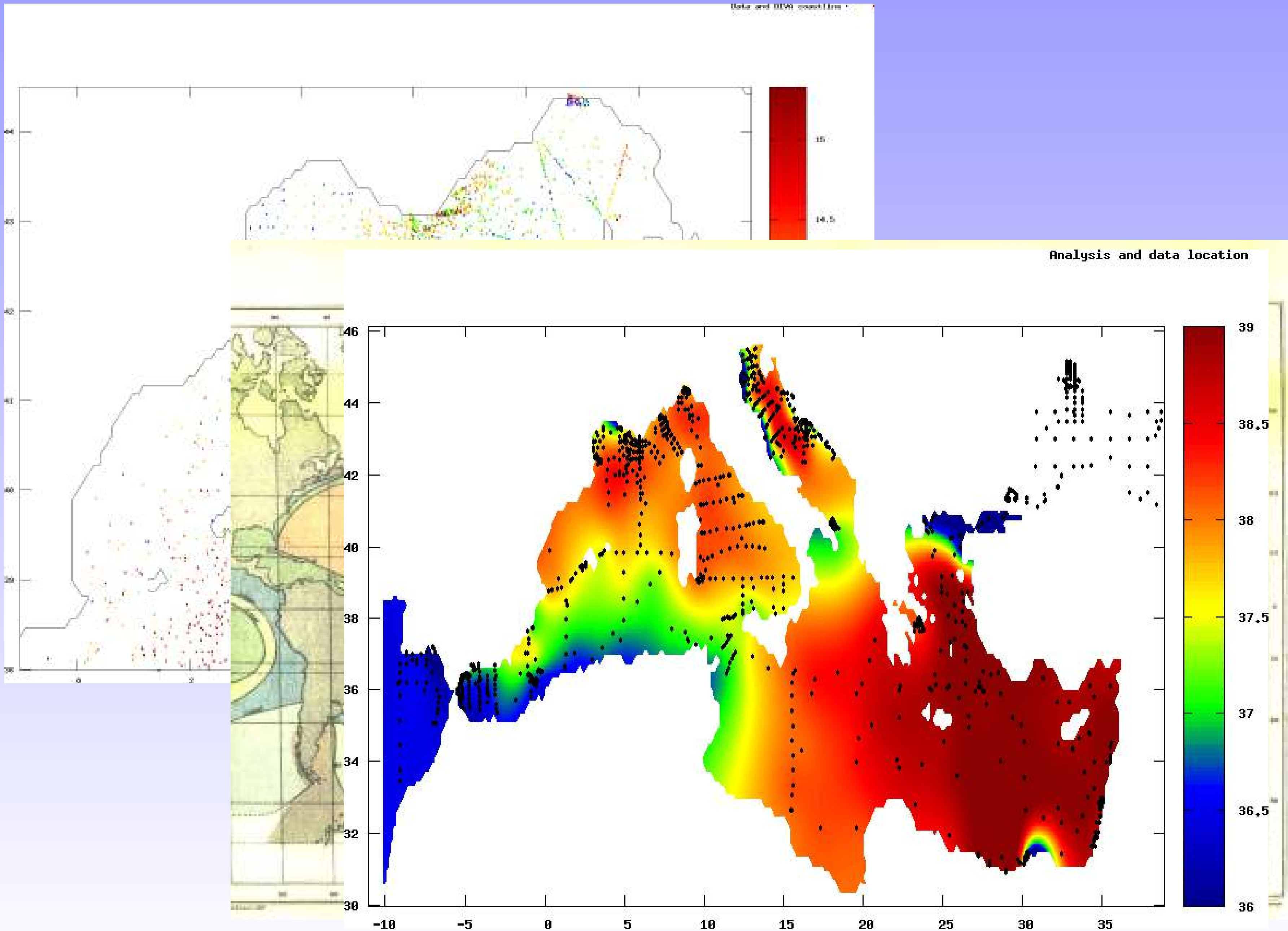
Solutions





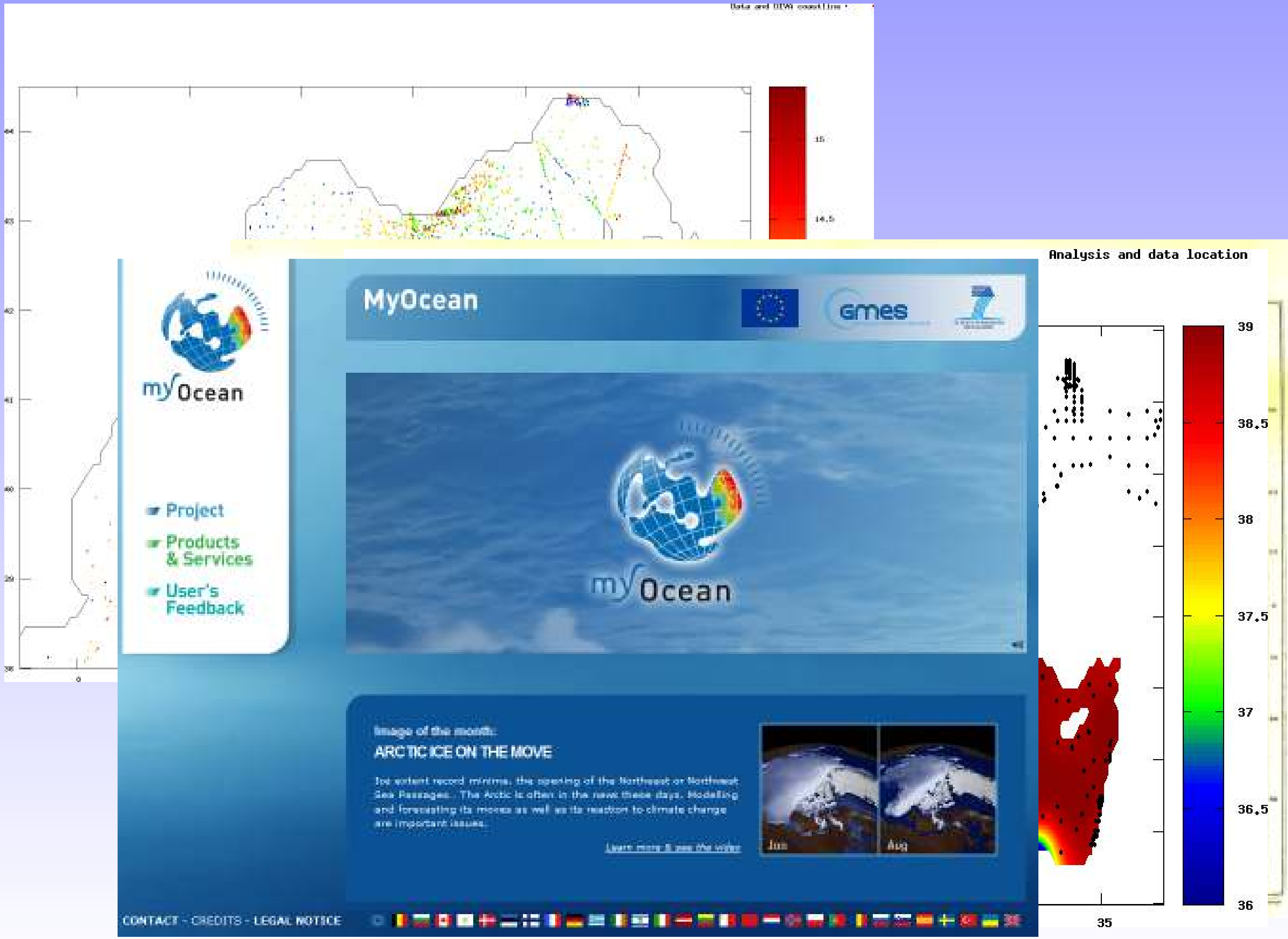
Université de Liège

Solutions

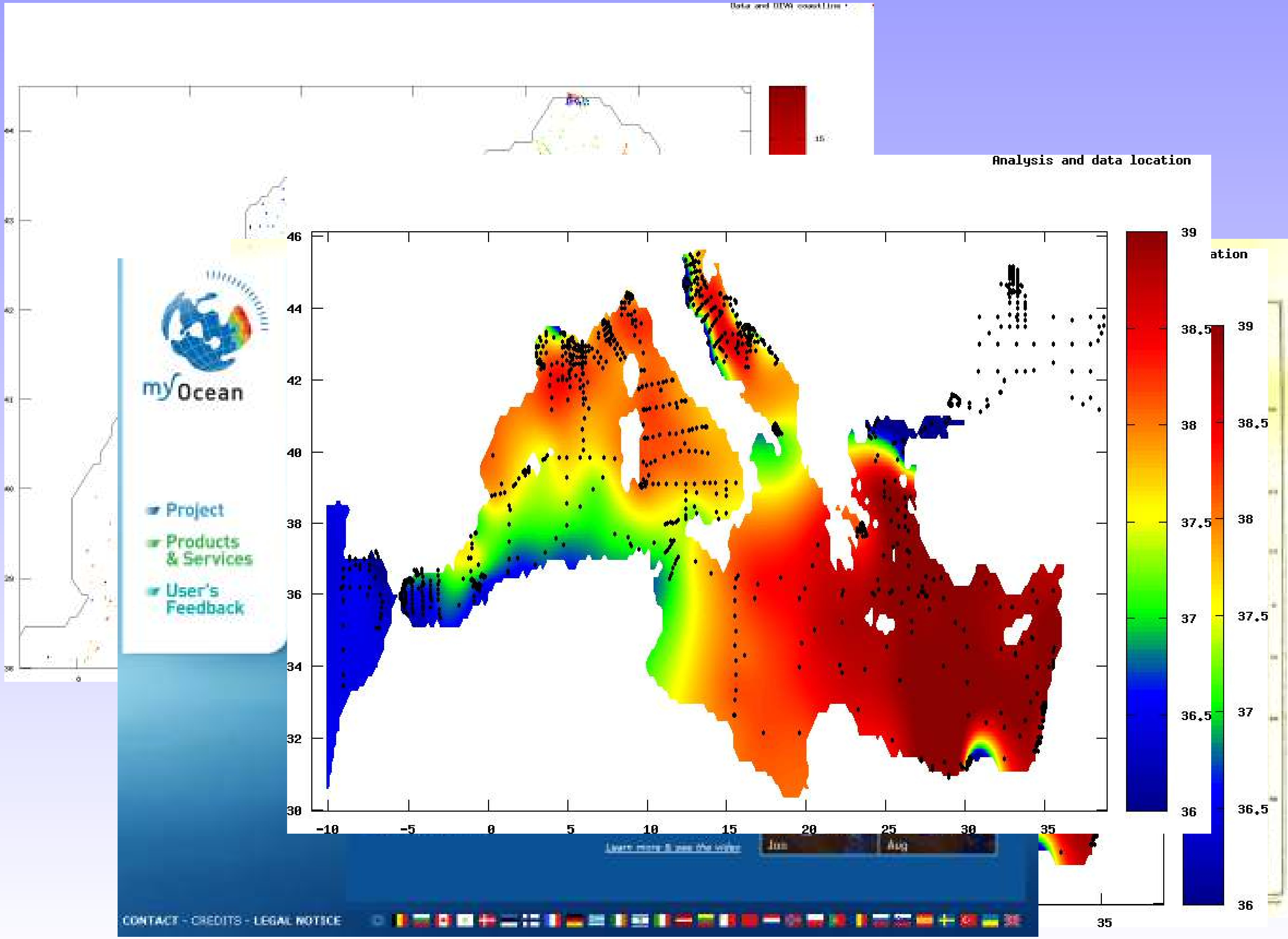


Université de Liège

Solutions


Analysis and data location
39
38.5
38
37.5
37
36.5
36
35
MyOcean
gmes
myOcean
Project
Products & Services
User's Feedback
Image of the month:
ARCTIC ICE ON THE MOVE
Ice extent record minima, the opening of the Northeast or Northwest Sea Passages. The Arctic is often in the news these days. Modelling and forecasting its moves as well as its reaction to climate change are important issues.
Learn more & see the video
Jun
Aug
CONTACT - CREDITS - LEGAL NOTICE

Solutions



Analysis and data location
my Ocean
Project
Products & Services
User's Feedback
CONTACT - CREDITS - LEGAL NOTICE
46
44
42
40
38
36
34
32
30
-10
-5
0
5
10
15
20
25
30
35
39
38.5
38
37.5
37
36.5
36

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

15°

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

But what if observer 1 uses digital thermometer and observer 2 his finger ?

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

But what if observer 1 which uses digital thermometer is in Bruges and observer 2 which uses his his finger is in Delft ?

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

But what if observer 1 which uses digital thermometer is in Sidney and observer 2 which uses his his finger is in Delft ?

Estimation of today's temperature in Delft

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

But what if observer 1 which uses digital thermometer is in Sidney and observer 2 which uses his his finger is in Delft ?

Best guess probably 16° but with a higher error bar.

Exploit knowledge of errors and distance both for the estimate itself but also the error bars

Optimal interpolation

Exploit knowledge of errors and distance both for the estimate itself but also the error bars

When done mathematically searching for the estimate with lowest expected error: OPTIMAL INTERPOLATION

⚠ Optimal, but needs some statistical information

Optimal estimate of true state ^{t}

Université de Liège

With two observations T_1 and T_2 deviation from the (unknown) truth T^t by (unknown) observational errors ϵ_1 and ϵ_2:

$$T_1 = T^t + \epsilon_1, \quad \langle \epsilon_1 \rangle = 0, \quad T_2 = T^t + \epsilon_2, \quad \langle \epsilon_2 \rangle = 0 \tag{1}$$

statistical average, denoted by $\langle \quad \rangle$ with unbiased estimates $\langle \epsilon_* \rangle = 0$
Linear estimate of the field

$$T = w_1 T_1 + w_2 T_2 = (w_1 + w_2) T^t + (w_1 \epsilon_1 + w_2 \epsilon_2) \tag{2}$$

$$\langle T \rangle = (w_1 + w_2) T^t, \tag{3}$$

we obtain an unbiased estimate of the true state if we take $w_1 + w_2 = 1$. This leaves one parameter free to chose: w_2

Exploit knowledge on errors to find optimal value of w_2

Choice of weighting ?

Analyzed field T^a calculated as

$$T^a = (1 - w_2)T_1 + w_2 T_2 \ = \ T_1 + w_2(T_2 - T_1) \qquad (4)$$

while in reality there is an error

$$T^a - T^t = (1 - w_2)\epsilon_1 + w_2\epsilon_2, \qquad (5)$$

This error is zero on average but its variance is not zero:

$$\left\langle (T^a - T^t)^2 \right\rangle = (1 - w_2)^2 \left\langle \epsilon_1^2 \right\rangle + w_2^2 \left\langle \epsilon_2^2 \right\rangle + 2(1 - w_2)\, w_2 \left\langle \epsilon_1 \epsilon_2 \right\rangle \qquad (6)$$

The actual errors ϵ_1 and ϵ_2 are not known, but the error variance $\left\langle \epsilon_1^2 \right\rangle$ are. Often we can reasonably suppose that the errors ϵ_1 and ϵ_2 are uncorrelated $\langle \epsilon_1 \epsilon_2 \rangle = 0$. The error variance $\left\langle \epsilon^2 \right\rangle$ of the analysis is then

$$\left\langle \epsilon^2 \right\rangle = (1 - w_2)^2 \left\langle \epsilon_1^2 \right\rangle + w_2^2 \left\langle \epsilon_2^2 \right\rangle. \qquad (7)$$

So what ?

Minimisation

$$\langle \epsilon^2 \rangle = (1 - w_2)^2 \langle \epsilon_1^2 \rangle + w_2^2 \langle \epsilon_2^2 \rangle . \tag{8}$$

Naturally, the best estimate for T is the one with the lowest expected error variance and we will use w_2, which minimizes the right-hand side:

$$w_2 = \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \tag{9}$$

$$T^a = \frac{\langle \epsilon_1^2 \rangle \langle \epsilon_2^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \left(\frac{T_1}{\langle \epsilon_1^2 \rangle} + \frac{T_2}{\langle \epsilon_2^2 \rangle} \right). \tag{10}$$

Université de Liège

Best estimate

With (9) we obtain the minimal error variance

$$\langle \epsilon^2 \rangle = \frac{\langle \epsilon_1^2 \rangle \langle \epsilon_2^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} = \left(1 - \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \right) \langle \epsilon_1^2 \rangle, \tag{11}$$

while the estimate of the temperature itself reads also

$$T^a = T_1 \; + \; \left(\frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \right) (T_2 - T_1)\,. \tag{12}$$

Error variance on the combination of T_1 and T_2 is smaller than both $\langle \epsilon_1^2 \rangle$ and $\langle \epsilon_2^2 \rangle$.

VAR approach

Same solution by

$$\min_T J = \frac{(T - T_1)^2}{2\left\langle \epsilon_1^2 \right\rangle} + \frac{(T - T_2)^2}{2\left\langle \epsilon_2^2 \right\rangle}. \quad (13)$$

Optimal interpolation

Analysis x^a as a linear combination of the forecast x^f and the observations y:

$$\mathrm{x}^a = \mathrm{x}^f + \mathbf{K}\left(\mathrm{y} - \mathbf{H}\mathrm{x}^f\right) \tag{14}$$

H observation operator and innovation vector

$$\boldsymbol{d} = \mathrm{y} - \mathbf{H}\mathrm{x}^f \tag{15}$$

Objective: prescribe an optimal matrix **K**: Kalman gain matrix

Knowledge on error distribution

$$\boldsymbol{\epsilon} = \mathrm{x} - \mathrm{x}^t \tag{16}$$

$$\boldsymbol{\epsilon}^o = \mathrm{y} - \mathrm{y}^t. \tag{17}$$

Error-covariance matrix

$$\mathbf{R} = \left\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{o\,\mathsf{T}} \right\rangle \tag{18}$$

is semi-positive defined since $\boldsymbol{z}^\mathsf{T} \mathbf{R} \boldsymbol{z} = \left\langle \left(\boldsymbol{z}^\mathsf{T} \boldsymbol{\epsilon}^o \right)^2 \right\rangle$

The analysis step (14) reads

$$\mathrm{x}^t + \boldsymbol{\epsilon}^a = \mathrm{x}^t + \boldsymbol{\epsilon}^f + \mathbf{K} \left(\boldsymbol{\epsilon}^o - \mathbf{H} \boldsymbol{\epsilon}^f \right) \; + \; \mathbf{K} \underbrace{\left(\mathrm{y}^t - \mathbf{H} \mathrm{x}^t \right)}_{=0} \tag{19}$$

Kalman gain

$$\boldsymbol{\epsilon}^a = \boldsymbol{\epsilon}^f + \mathbf{K}\left(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f\right). \tag{20}$$

Construct the error covariance $\left\langle \boldsymbol{\epsilon}^a {\boldsymbol{\epsilon}^a}^\mathsf{T} \right\rangle$ of the analysis by multiplying (20) by its transposed and take the statistical average or expectation

$$\begin{aligned}
\left\langle \boldsymbol{\epsilon}^a {\boldsymbol{\epsilon}^a}^\mathsf{T} \right\rangle &= \left\langle \boldsymbol{\epsilon}^f {\boldsymbol{\epsilon}^f}^\mathsf{T} \right\rangle + \mathbf{K}\left\langle (\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f){\boldsymbol{\epsilon}^f}^\mathsf{T} \right\rangle + \left\langle \boldsymbol{\epsilon}^f ({\boldsymbol{\epsilon}^o}^\mathsf{T} - {\boldsymbol{\epsilon}^f}^\mathsf{T}\mathbf{H}^\mathsf{T}) \right\rangle \mathbf{K}^\mathsf{T} \\
&+ \mathbf{K}\left\langle (\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^f)({\boldsymbol{\epsilon}^o}^\mathsf{T} - {\boldsymbol{\epsilon}^f}^\mathsf{T}\mathbf{H}^\mathsf{T}) \right\rangle \mathbf{K}^\mathsf{T}.
\end{aligned} \tag{21}$$

Define covariance matrices

$$\mathbf{P} = \left\langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\mathsf{T} \right\rangle \tag{22}$$

and assume that observational errors and model errors are not correlated, $\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^\mathsf{T} \rangle = 0$.

Kalman gain

The error-covariance matrix after analysis can then be written as

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^\mathsf{T}\mathbf{K}^\mathsf{T} + \mathbf{K}\left(\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T}\right)\mathbf{K}^\mathsf{T}$$
$$= \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^\mathsf{T}\mathbf{A}^{-1}\mathbf{H}\mathbf{P}^f + \left(\mathbf{P}^f\mathbf{H}^\mathsf{T} - \mathbf{K}\mathbf{A}\right)\mathbf{A}^{-1}\left(\mathbf{H}\mathbf{P}^f - \mathbf{A}\mathbf{K}^\mathsf{T}\right) \quad (23)$$

where we define matrix

$$\mathbf{A} = \mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T} + \mathbf{R} \quad (24)$$

which is symmetric and we suppose that it can be inverted. Global error estimate:

$$\epsilon^a = \left\langle \boldsymbol{\epsilon}^{a\mathsf{T}} \boldsymbol{\epsilon}^a \right\rangle = \operatorname{trace} \mathbf{P}^a. \quad (25)$$

Search for an optimal **K** which minimizes this trace or for which

$$\epsilon^a(\mathbf{K} + \mathbf{L}) - \epsilon^a(\mathbf{K}) = 0 \quad (26)$$

for any small departure matrix **L**

Kalman gain

$$\mathrm{trace} - \mathbf{L}\left(\mathbf{H}\mathbf{P}^f - \mathbf{A}\mathbf{K}^\mathsf{T}\right) - \left(\mathbf{P}^f\mathbf{H}^\mathsf{T} - \mathbf{K}\mathbf{A}\right)\mathbf{L}^\mathsf{T} = 0, \tag{27}$$

where we neglected quadratic terms in **L**. The two terms are the transposed version of each other and since the trace of the matrix and its transposed are identical, we must request

$$\mathrm{trace}\left(\mathbf{P}^f\mathbf{H}^\mathsf{T} - \mathbf{K}\mathbf{A}\right)\mathbf{L}^\mathsf{T} = 0.$$

Since **L** is arbitrary, the optimal solution with minimum error is obtained when

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

Kalman gain

The error covariance of the analysis is obtained by injecting (28) into (23)

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^f = \left(\mathbf{I} - \mathbf{P}^f\mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T} + \mathbf{R}\right)^{-1}\mathbf{H}\right)\mathbf{P}^f \qquad (29)$$

which is the analogue of (11). Note that both the Kalman gain matrix and the error covariance after the analysis do not depend on the *value* of the observations or the forecasted state vector but only on their statistical error covariances. The only field that depends on the actual values is of course the state vector itself:

$$\mathrm{x}^a = \mathrm{x}^f + \mathbf{P}^f\mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T} + \mathbf{R}\right)^{-1}\left(\mathrm{y} - \mathbf{H}\mathrm{x}^f\right). \qquad (30)$$

The use of (28) in (14) to combine the forecast and observation with prescribed error covariance $\mathbf{P}^f$ and $\mathbf{R}$ is known as optimal interpolation (OI)

3D-Var

Find the state vector that minimizes the error measure J given by

$$J(\mathrm{x}) = \frac{1}{2}(\mathrm{x}-\mathrm{x}^f)^{\mathsf{T}}\mathbf{P}^{f^{-1}}(\mathrm{x}-\mathrm{x}^f) + \frac{1}{2}(\mathbf{H}\mathrm{x}-\mathrm{y})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{H}\mathrm{x}-\mathrm{y}) \qquad (31)$$

Yields the same optimal state.

3D-Var

Find the state vector that minimizes the error measure J given by

$$J(\mathrm{x}) = \frac{1}{2}(\mathrm{x}-\mathrm{x}^f)^{\mathsf{T}}\mathbf{P}^{f^{-1}}(\mathrm{x}-\mathrm{x}^f) + \frac{1}{2}(\mathbf{H}\mathrm{x}-\mathrm{y})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{H}\mathrm{x}-\mathrm{y}) \tag{32}$$

Yields the same optimal state.

VAR minimizes a sum of a quadratic term of the state-variables and a quadratic term of misfits (residuals)

3D-Var

Find the state vector that minimizes the error measure J given by

$$J(\mathrm{x}) = \frac{1}{2}(\mathrm{x}-\mathrm{x}^f)^T \mathbf{P}^{f^{-1}}(\mathrm{x}-\mathrm{x}^f) + \frac{1}{2}(\mathbf{H}\mathrm{x}-\mathrm{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathrm{x}-\mathrm{y}) \qquad (33)$$

Yields the same optimal state.

VAR minimizes a sum of a quadratic term of the state-variables and a quadratic term of misfits (residuals)

Kalman filter and 3DVAR yield same results when used with same covariances in linear case.

Université de Liège

Spatial interpolation

- "Model forecast": Background field (data mean, global spatial south-north gradient, climatological mean, ...)
- $\mathbf{B} = \mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T}$: covariance of the background field between data points: element i, j of $\mathbf{B}$ is the covariance between points in location i and j. Covariance between a given point and all data points is stored in column vector $\mathbf{c}$ with local variance σ^2.
- Analysis ϕ of anomaly y with respect to background leads to spatial analysis at any desired location of covariance between any two points is known.

$$\phi = \mathbf{c}^\mathsf{T} (\mathbf{B} + \mathbf{R})^{-1} \mathrm{y} \tag{34}$$

with a local error variance of the analysis

$$\epsilon_a^2 = \sigma^2 - \mathbf{c}^\mathsf{T} (\mathbf{B} + \mathbf{R})^{-1} \mathbf{c} \tag{35}$$

⚠ Note that inversion of matrix is needed (cost $\propto$ cube of number of data points).

Background covariance

Problem, how to specify background covariances (between all data points and between data points and the desired analysis location).

- c_i= covariance between location of the analysis and data location of point i = $C(x, x_i)$
- B_{ij}=covariance between location of data point i and location of point j= $C(x_i, x_j)$

Approaches

- Normally obtained via statistics on data. Seldom possible (noticable exception: satellite images, see DINEOF).
- Standard OI: via functions $B_{ij} = f(r/L)$ where r is the distance between points i and j, but still function f needs to be determined. L is the so-called correlation length. Here statistics on all data couples as a function of distance. Example: $f = \sigma^2 \exp(-r^2/L^2)$.
- Via functionals (see Kernel of DIVA later)

Signal to noise ratio

$$\mathbf{B} = \sigma^2 \tilde{\mathbf{B}} \tag{36}$$

$$\mathbf{R} = \epsilon^2 \tilde{\mathbf{R}} \tag{37}$$

$$\mathbf{c} = \sigma^2 \tilde{\mathbf{c}} \tag{38}$$

with non-dimensional correlation matrices $\tilde{\mathbf{B}}$, $\tilde{\mathbf{R}}$, $\tilde{\mathbf{c}}$

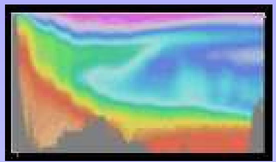
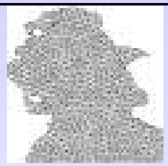
$$\phi = \tilde{\mathbf{c}}^{\mathsf{T}} \left(\tilde{\mathbf{B}} + \frac{1}{\lambda} \tilde{\mathbf{R}} \right)^{-1} \mathrm{y} \tag{39}$$

with signal-to noise ratio

$$\lambda = \frac{\sigma^2}{\epsilon^2} \tag{40}$$

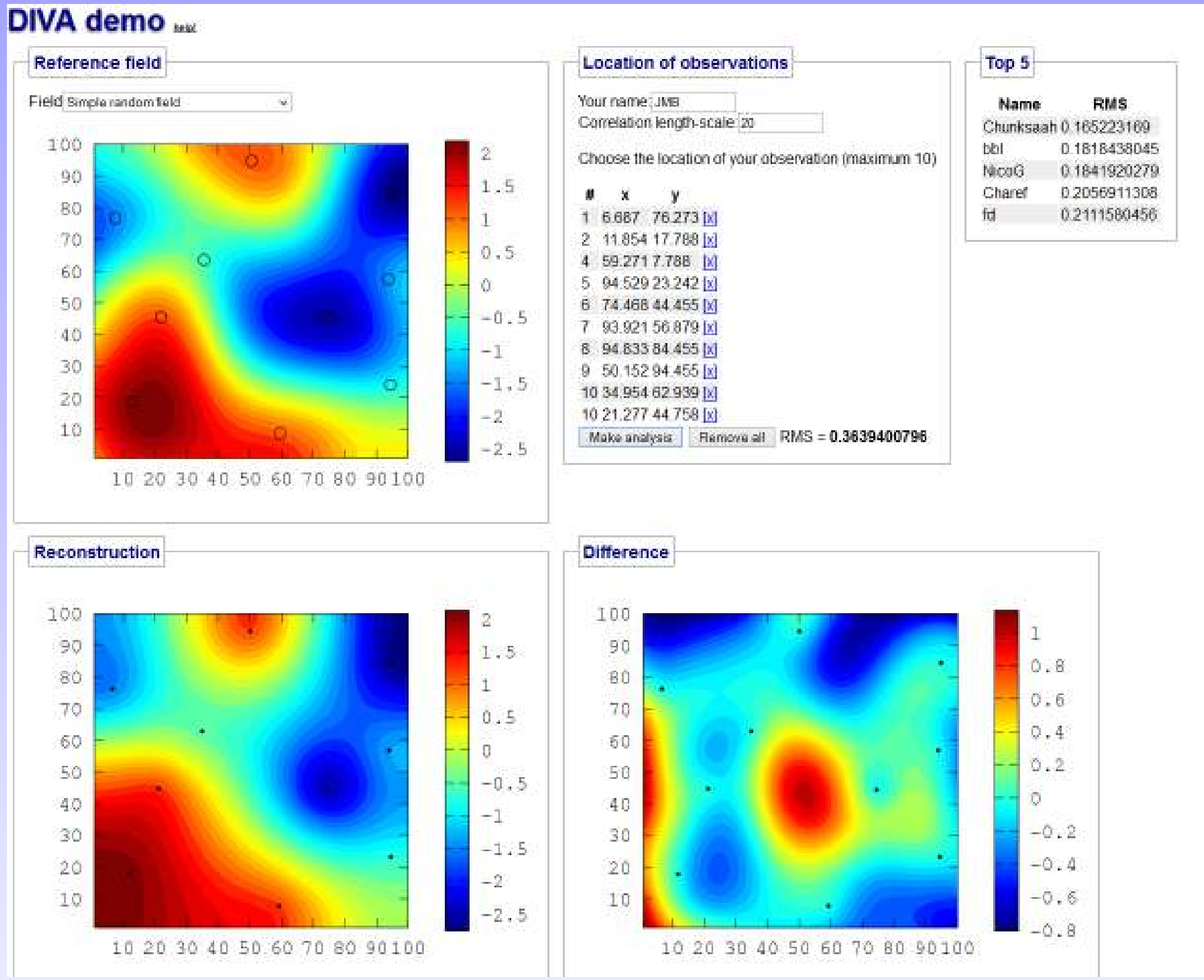
Also the error field is only depending on the ratio.

Statistical spatial analysis

Tools	Formats	Method
	ODV spreadsheet WOCE WOA ...	Dist. weighting, **VIM**
	netCDF (toolbox), CSV ascii, ...	Polynomial interpolation ...
	CSV ascii, ...	Kriging, OI, ...
	ODV spreadsheet	Variational

Mostly graphics oriented, without "oceanographic" knowledge.

Demo tool to feel effect of L



http://www.data-assimilation.net/Tools/divand_demo/html/

- *Gridding*
- ***DIVA theory***
- *Implementations and exercises*
- *DINEOF theory*
- *Implementations and exercises*

DIVA Basics

Variational Inverse Method, (Brasseur *et al.*, 1996). Knowing data d_j at location (x_j, y_j), search the field φ which minimizes

$$J[\varphi] = \sum_{j=1}^{Nd} \mu_j [d_j - \varphi(x_j, y_j)]^2 + \|\varphi - \varphi_b\|^2 \tag{41}$$

$$\|\varphi\| = \int_D (\alpha_2 \nabla\nabla\varphi : \nabla\nabla\varphi + \alpha_1 \nabla\varphi \cdot \nabla\varphi + \alpha_0 \varphi^2) \mathrm{d}D \tag{42}$$

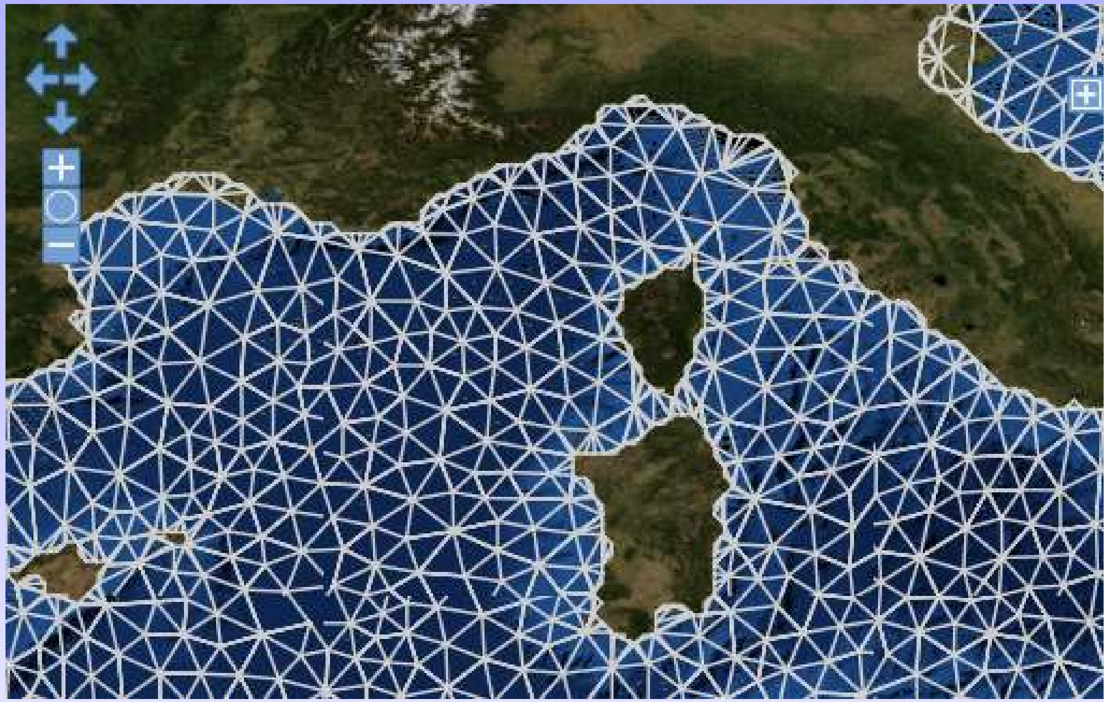
The background field φ_b is typically the data average value.

- α_0 penalizes the field itself (anomalies),
- α_1 penalizes gradients $\nabla\varphi$ (no spatial trends),
- α_2 penalizes variability (regularization of second derivatives $\nabla\nabla\varphi$),
- α_* can be related to a length scale L of the analysis,
- μ_j penalizes data-analysis misfits (objective).

Basics

$$\mu = \frac{\sigma^2}{\epsilon^2}\frac{4\pi}{L^2} \qquad (43)$$

where the σ^2/ϵ^2 is known as a signal to noise ratio S/N.



Solution by finite element method. Note decoupling of sub-basins. (Each element is in fact composed by three sub-elements, each one with cubic functions)

Bad news ..⌢

- No error estimate comes with the method, only an indicator of data-coverage.

Université de Liège

Bad news :(

- No error estimate comes with the method, only an indicator of data-coverage.
- VIM brand already taken: 

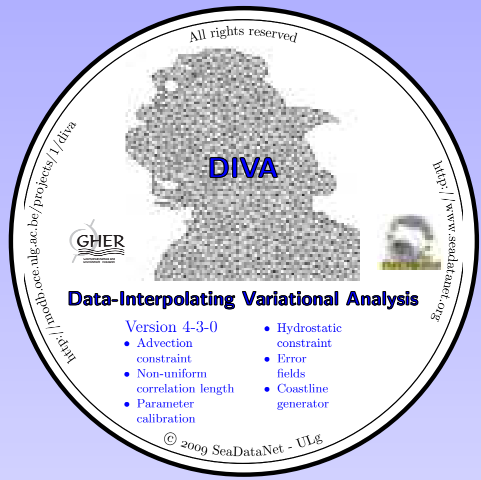
Bad news

- No error estimate comes with the method, only an indicator of data-coverage.
- VIM brand already taken: 

- Method is equivalent to well established existing optimal interpolation (OI)

Good news ⌣

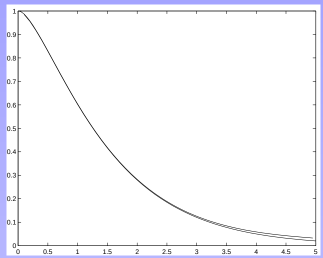
- Name easily changed: DIVA (Data-Interpolating Variational Analysis)



- Optimal interpolation (OI) provides error estimates, so DIVA can also provide it via equivalence
- DIVA has some practical advantages over OI

DIVA as OI

DIVA is identical to the well known Optimal Interpolation



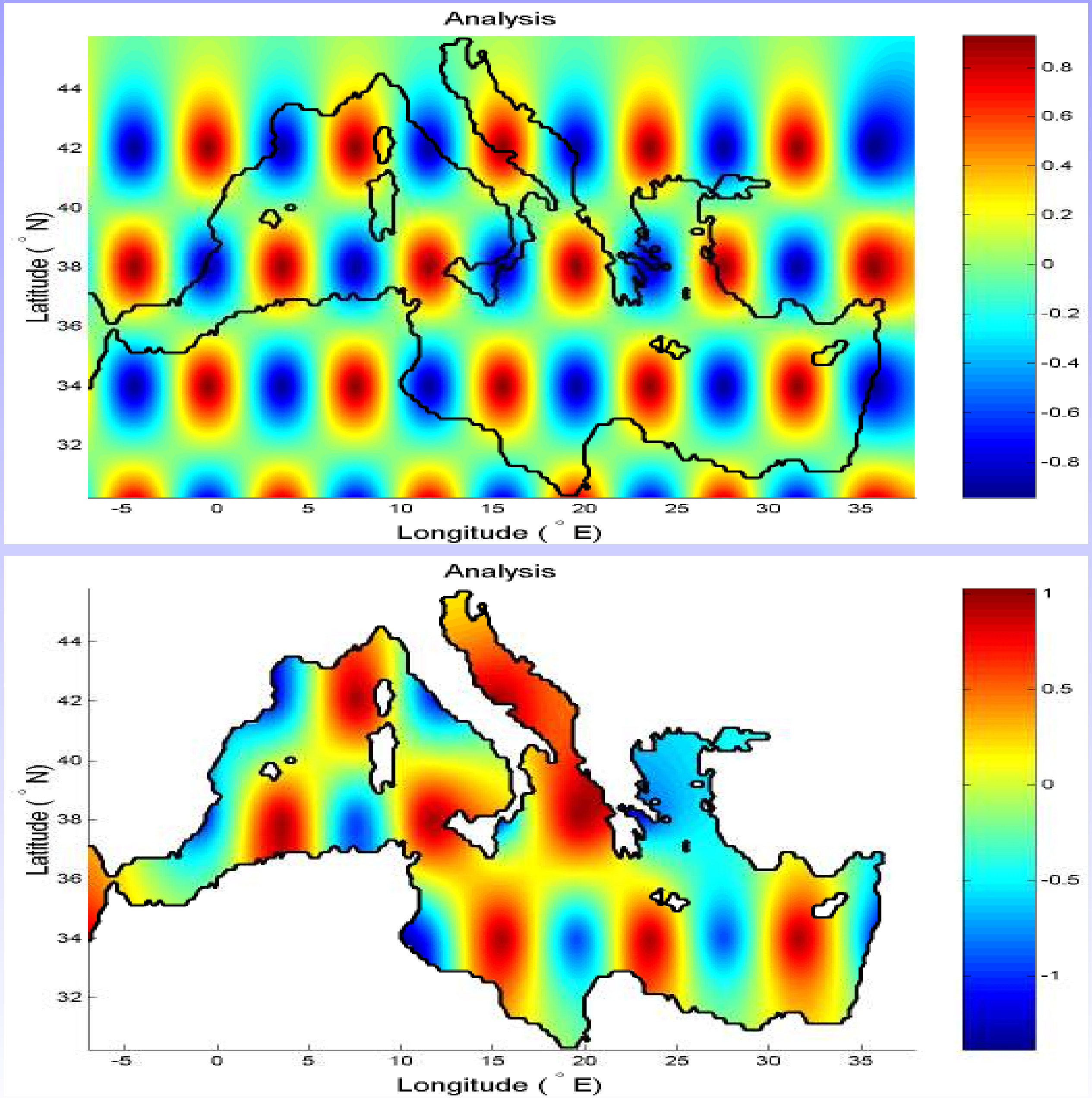
- if so-called reproducing kernel of the norm = covariance function of OI,
- if the noise is random, spatially uncorrelated and the signal/noise ratio parameter is identical with OI.

In this case, the OI solution = DIVA solution.

- Advantages of DIVA: regularization, fast finite-element solution, boundary effects taken into account.
- Difficulties: generalizations to 3D and multivariate versions are "hybrid".

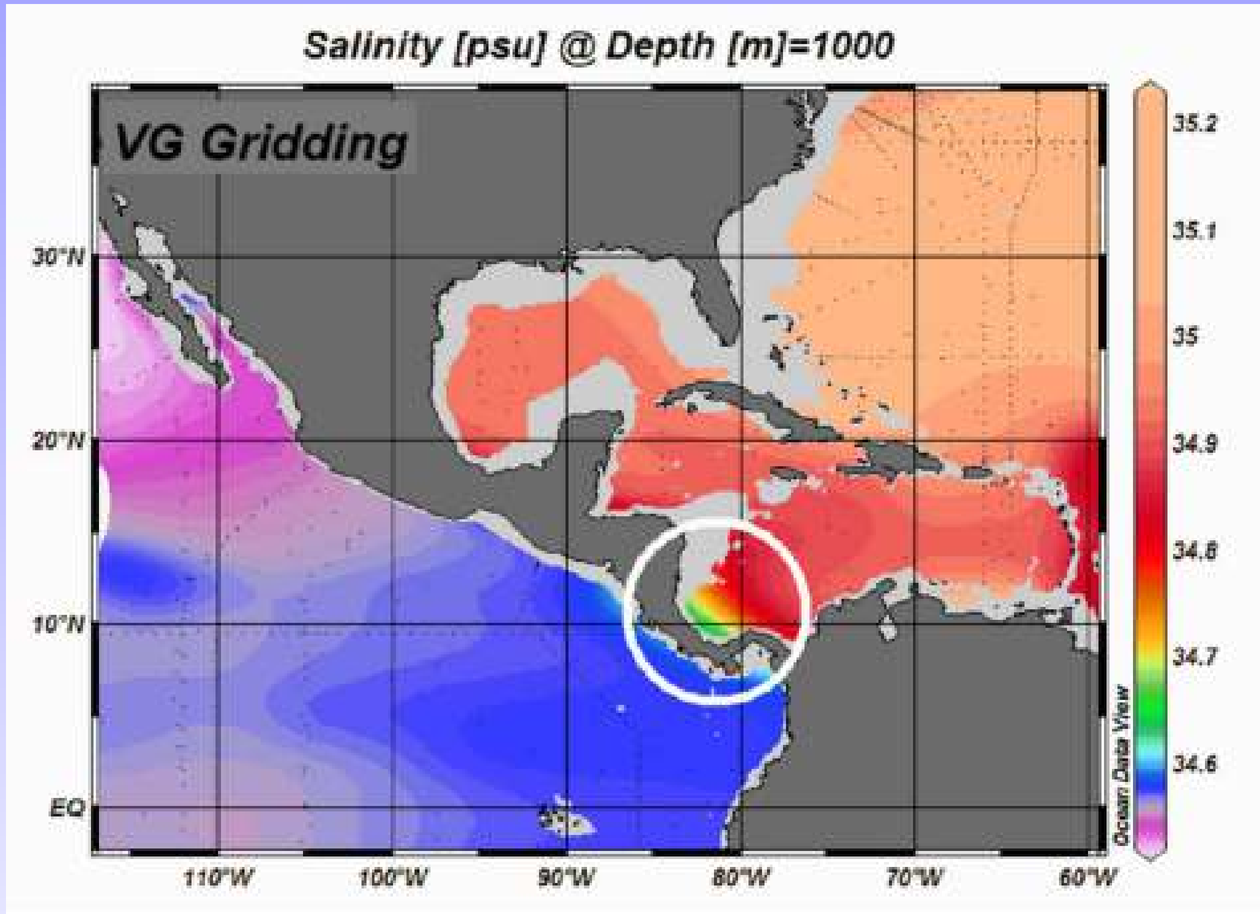
Major direct advantage of DIVA: matrix to invert is related to the finite-element mesh, NOT the number of data. Useful for large data sets (Rixen *et al.*, 2000). Equivalence allows to calculate error fields with DIVA even if formulation does not rely on error minimization.

Illustration of covariance functions



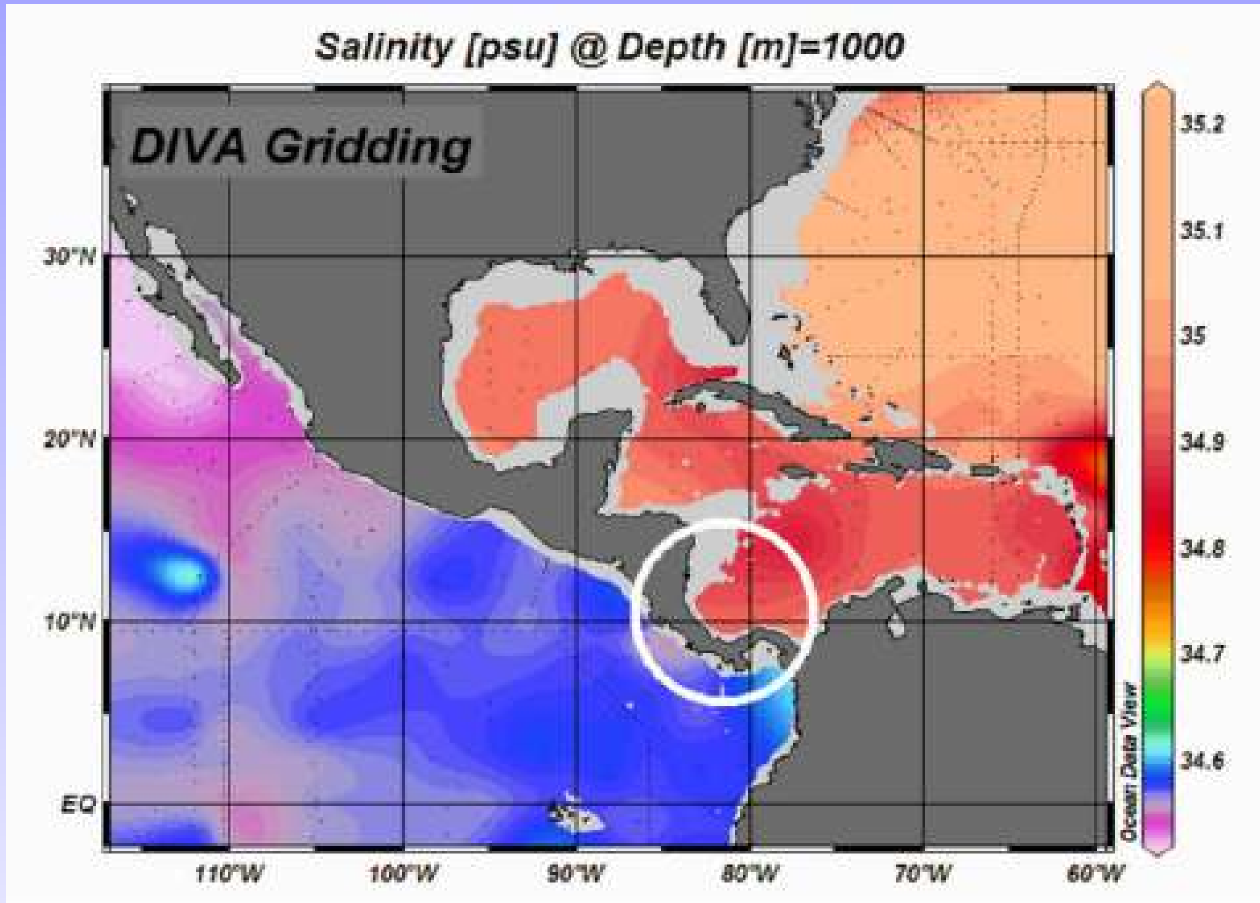
Université de Liège

Standard ODV analysis



(R. Schlitzer ODV example)

ODV-DIVA analysis



(R. Schlitzer ODV example)

Additions to basic tool

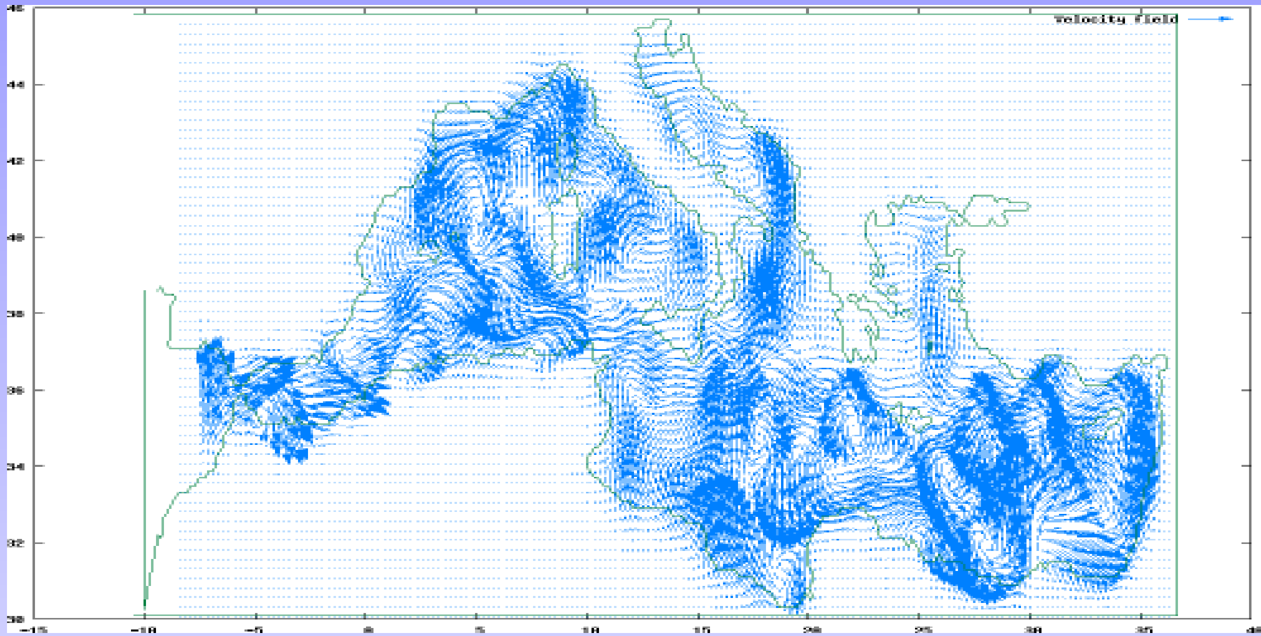
Advection constraint: Augmented cost function to deal with preferred correlation directions, eg, via advection with velocity $\boldsymbol{u}$ and diffusion $\mathscr{A}$

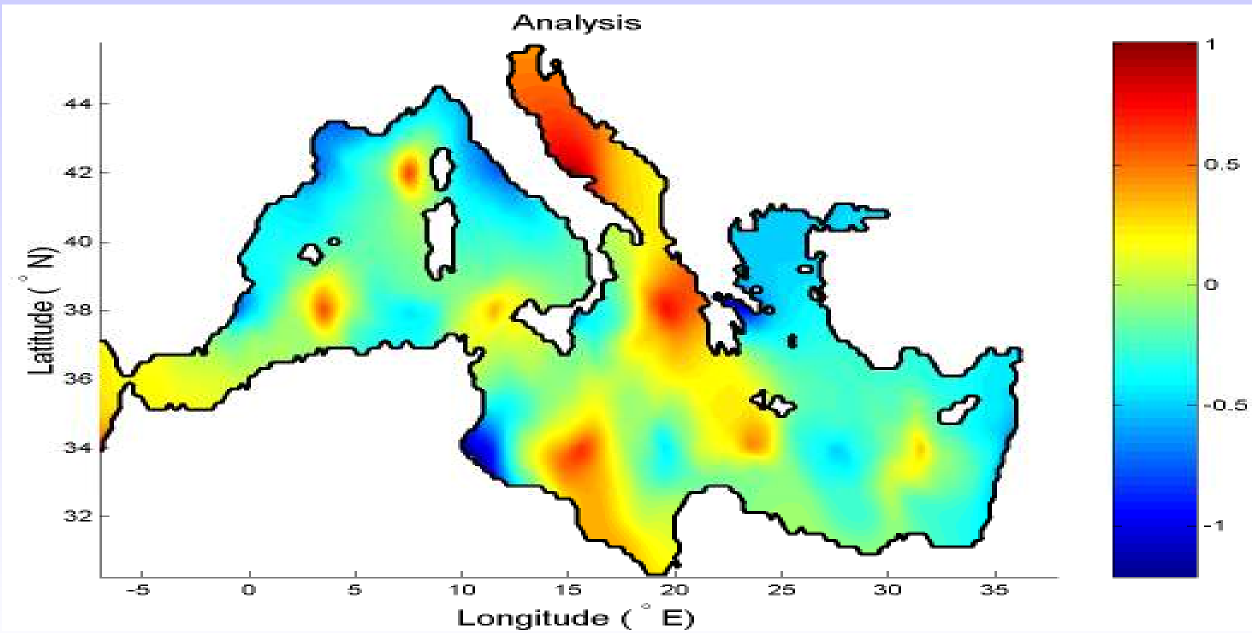
$$J_a = J(\varphi) + \theta \int_D \left[\boldsymbol{u} \cdot \nabla \varphi - \mathscr{A} \nabla \cdot \nabla \varphi\right]^2 \mathrm{d}D \qquad (44)$$

Other features

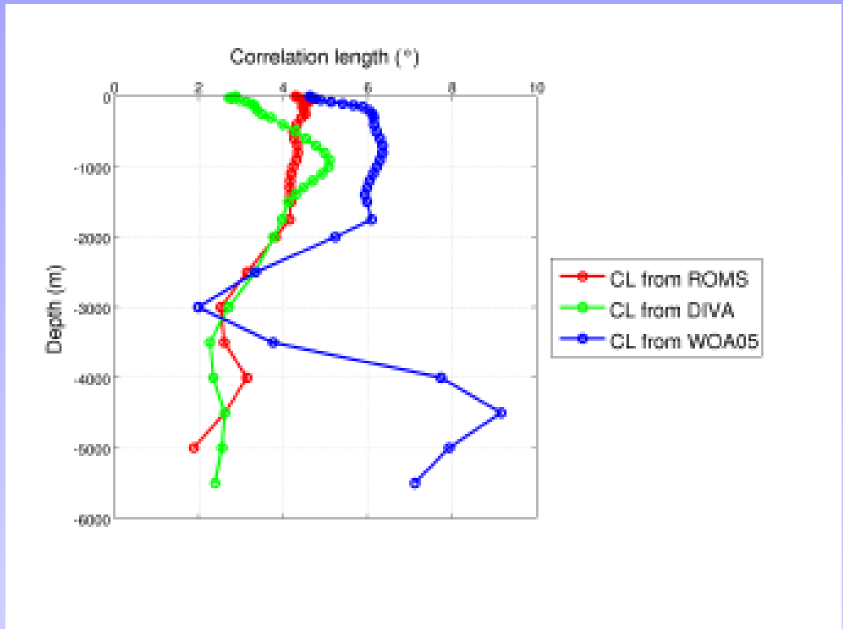
- Error fields taking data distribution into account.
- Toolbox approach allowing to design own versions.
- 3D and 4D modes by looping, hydrostatic constraint in 3D mode.
- Cross validation tools to infer statistical parameters and error estimates.
- Climatology production version with heterogeneous data distributions (detrending) .
- Outlier detection.
- Variable correlation length.
- ...

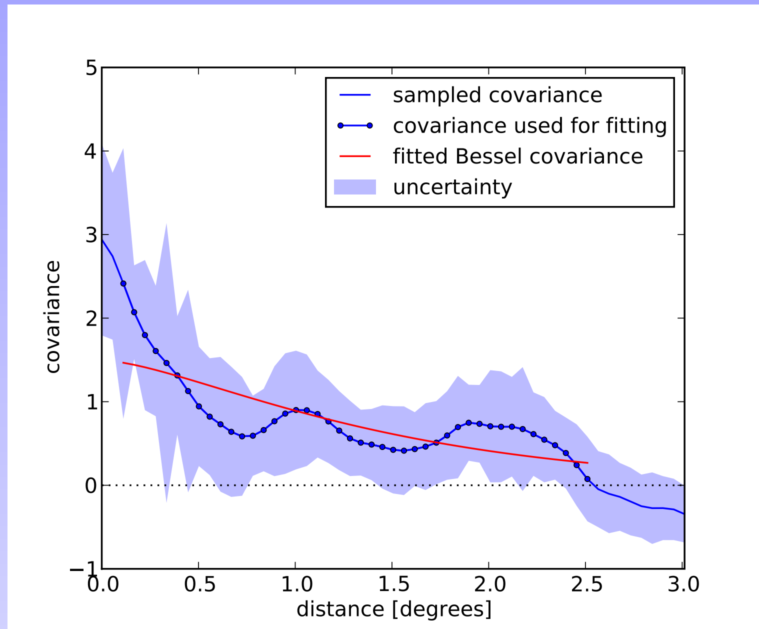
Covariances with advection





Parameter calibration

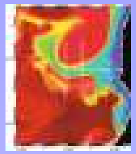




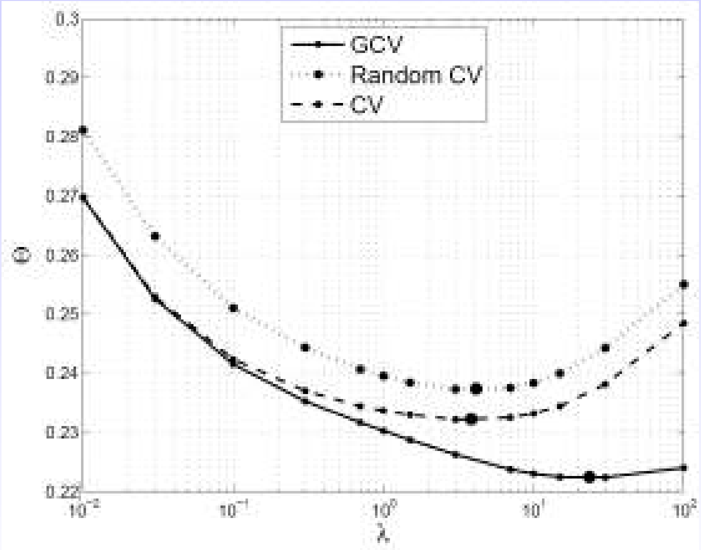
Spatial coherence of parameters: here correlation length obtained with covariance fitting (Troupin *et al.*, 2010).

Signal to noise ratio

The most elusive parameter.

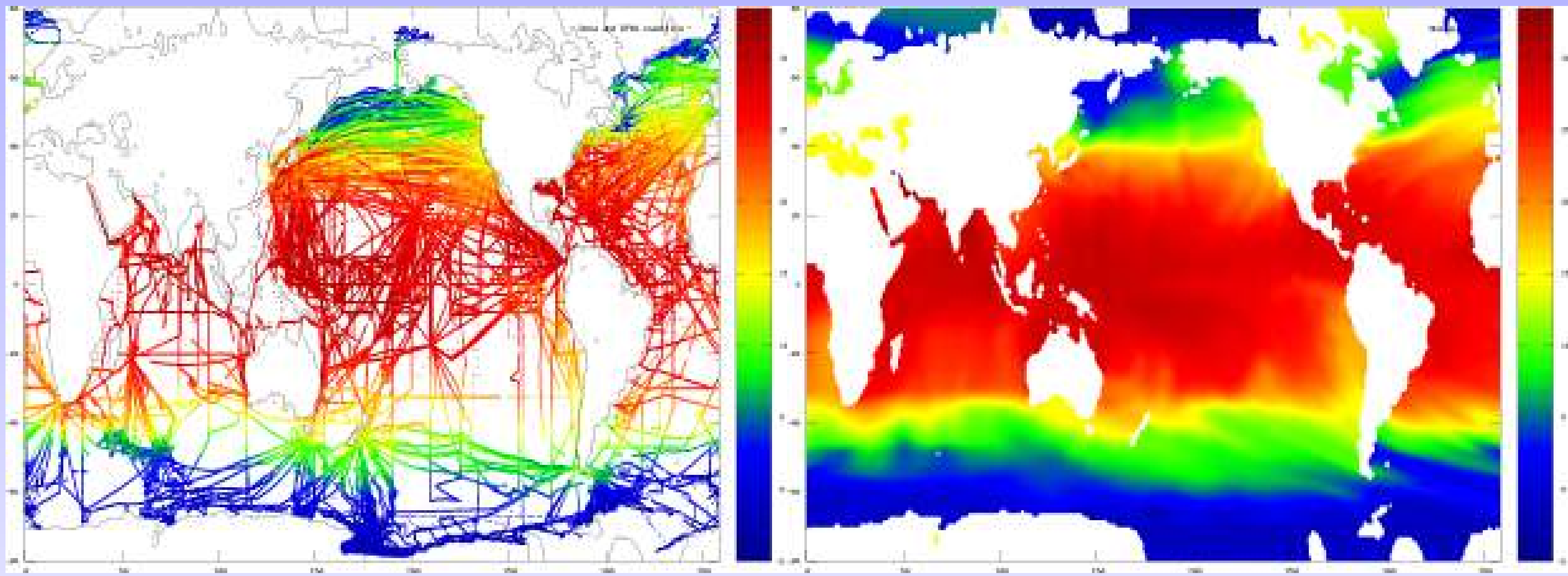
- Noise is not only instrumental error: 
- Very hard problem to decide on value with dependent data (cross-validation approaches fail).

A series of estimation tools are provided with DIVA, but here the experience of oceanographers is critical. A posteriori analysis of residuals allows to verify coherence. With reasonable amount of data, parameter not critical for analysis but for error estimates.

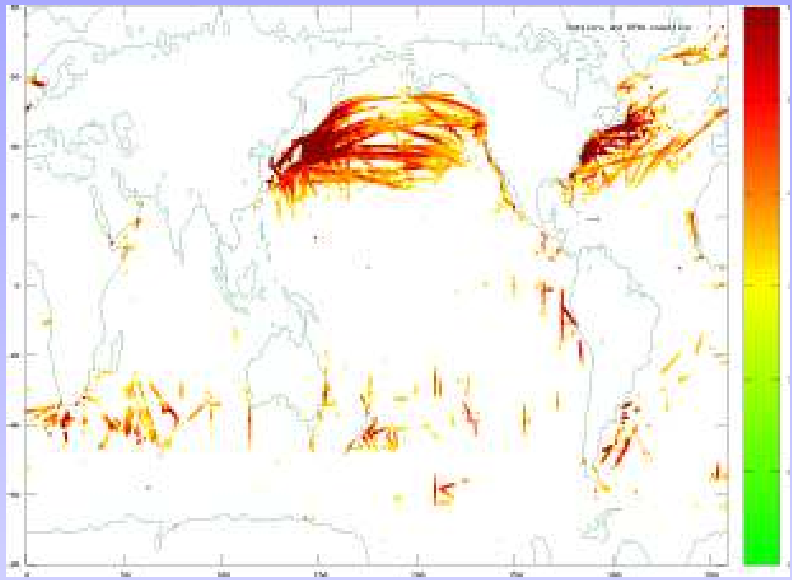


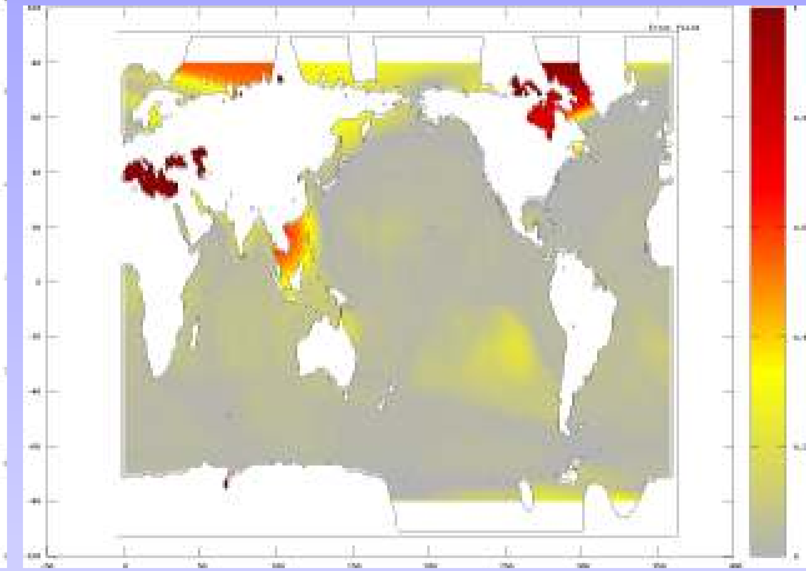
Huge problems

LDEO data base with 4.5 10^6 measurements (Takahashi *et al.*, 2009). Running on a laptop within a few minutes. Shown here, temperature fields.



Huge problems, outliers and relative error field

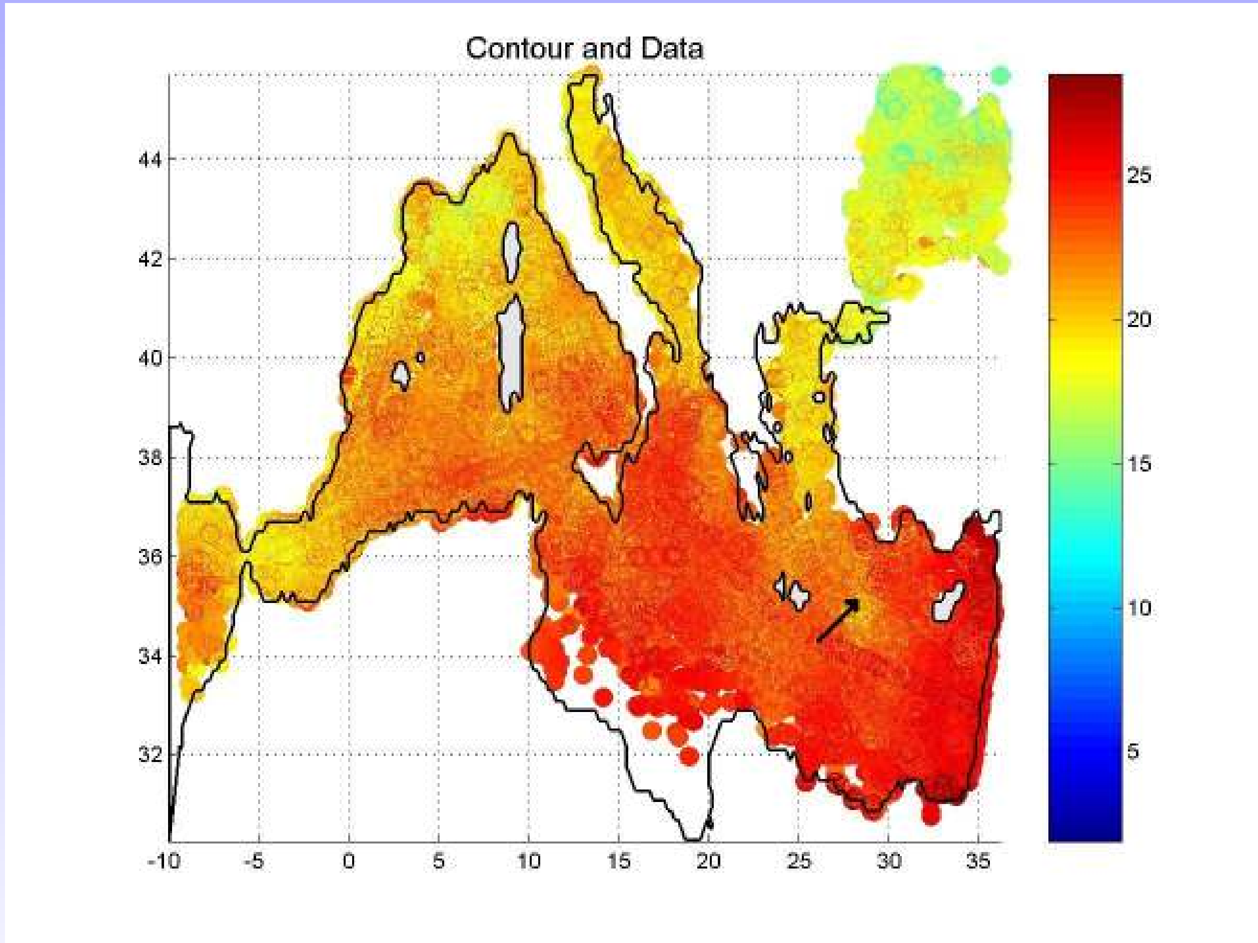




Outliers detected via comparison of statistically expected residuals (value provided by the DIVA analysis) and actual residuals (note that expected residuals decrease for large S/N).
Error field (on the right) can be used to mask regions with large uncertainties (low data coverage and/or large errors on data)

Outlier detection to detect encoding problems

Add value of 27 in 29°E , 35°N (21788th data point): corresponds to a displaced profile within a subregion



Université de Liège

Outlier with divaqc

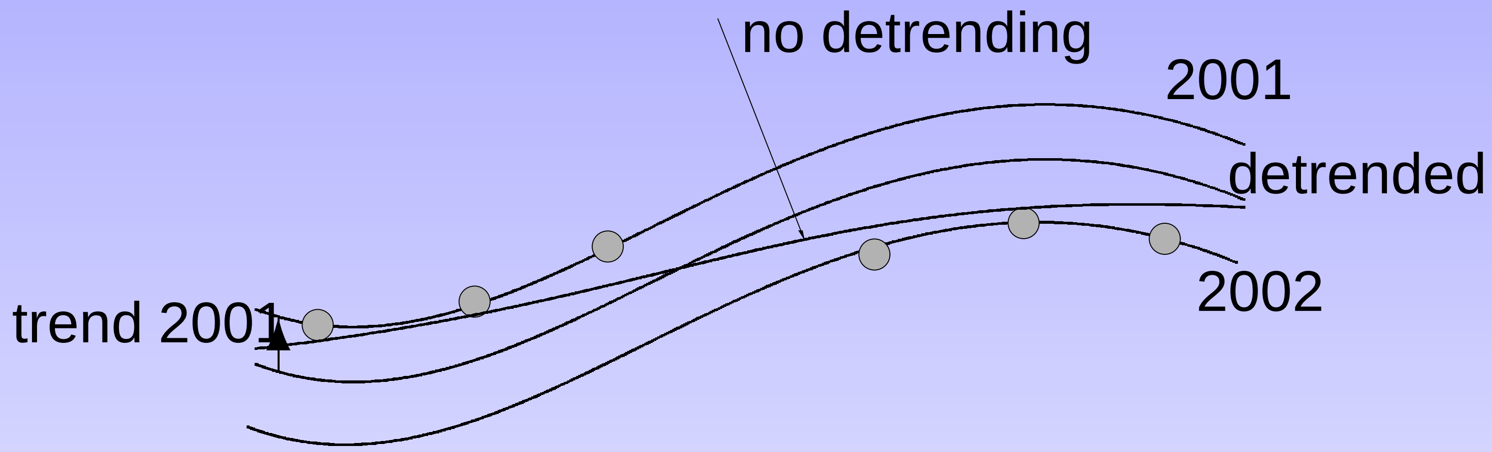
/cygdrive/c/jmb/cd-roms/modb/Utilities/Diva/Diva4.0/divastripped

```
Beckers@GHER22 /cygdrive/c/jmb/cd-roms/modb/Utilities/Diva/Diva4.0/divastripped
$ head -40 ./output/outliers.normalized.dat
  7.83717633 6416  29.  33.6669998  13.7980003  23.3453999  0.696789324
  7.75217195 2837  28.2999992  40.7750015  7.67000008  16.995903  0.688102722
  7.42754382 6418  25.7666664  34.7669983  14.007  23.0623131  0.697402358
  5.68016422 831 -0.773333311  37.7874985  27.6000004  20.6183472  0.699612796
  4.35829445 20260  7.19999981  41.7999992  15.1000004  20.4000015  0.69670248
  4.32356833 9578  7.  43.0666656  14.  19.3166256  0.704522073
  4.26528099 822 -0.786666691  37.75  26.  20.7408123  0.701173782
  4.25726546 460 -2.91333342  35.2050018  27.6000004  22.4128113  0.692871869
  4.2435566 1379  6.0333333  42.8166656  13.8999996  19.1117496  0.703695774
  4.23628475 21788  29.  35.  27.  21.8075123  0.697001696
  4.23542502 18193  10.9666662  38.  17.7299995  22.9030018  0.699810088
  4.15811381 18117  7.  43.  14.2299995  19.3404331  0.70424962
  4.15148862 1524  7.3166666  43.6861649  14.7609997  19.8653717  0.704541206
  4.13344054 8310 -0.781666696  36.4329987  15.1999998  20.2505627  0.700169802
  4.06387505 828 -0.758333325  37.6833344  26.  20.9775677  0.702664018
  4.05206397 825 -0.819999993  37.7050018  25.8999996  20.8979683  0.701841593
  3.94206592 9927  4.2833333  42.8833351  15.  19.7462311  0.690053046
  3.87482896 20258  7.43166685  40.0999985  16.6000004  21.26474  0.690022349
  3.84992195 3213  6.5333333  41.3166656  15.5  20.1702461  0.695324957
  3.80394796 1409  8.33333302  43.7999992  15.3800001  20.0317707  0.700980067
  3.79860263 20263  7.19999981  42.9669991  14.8000002  19.4583645  0.702965677
```

? !

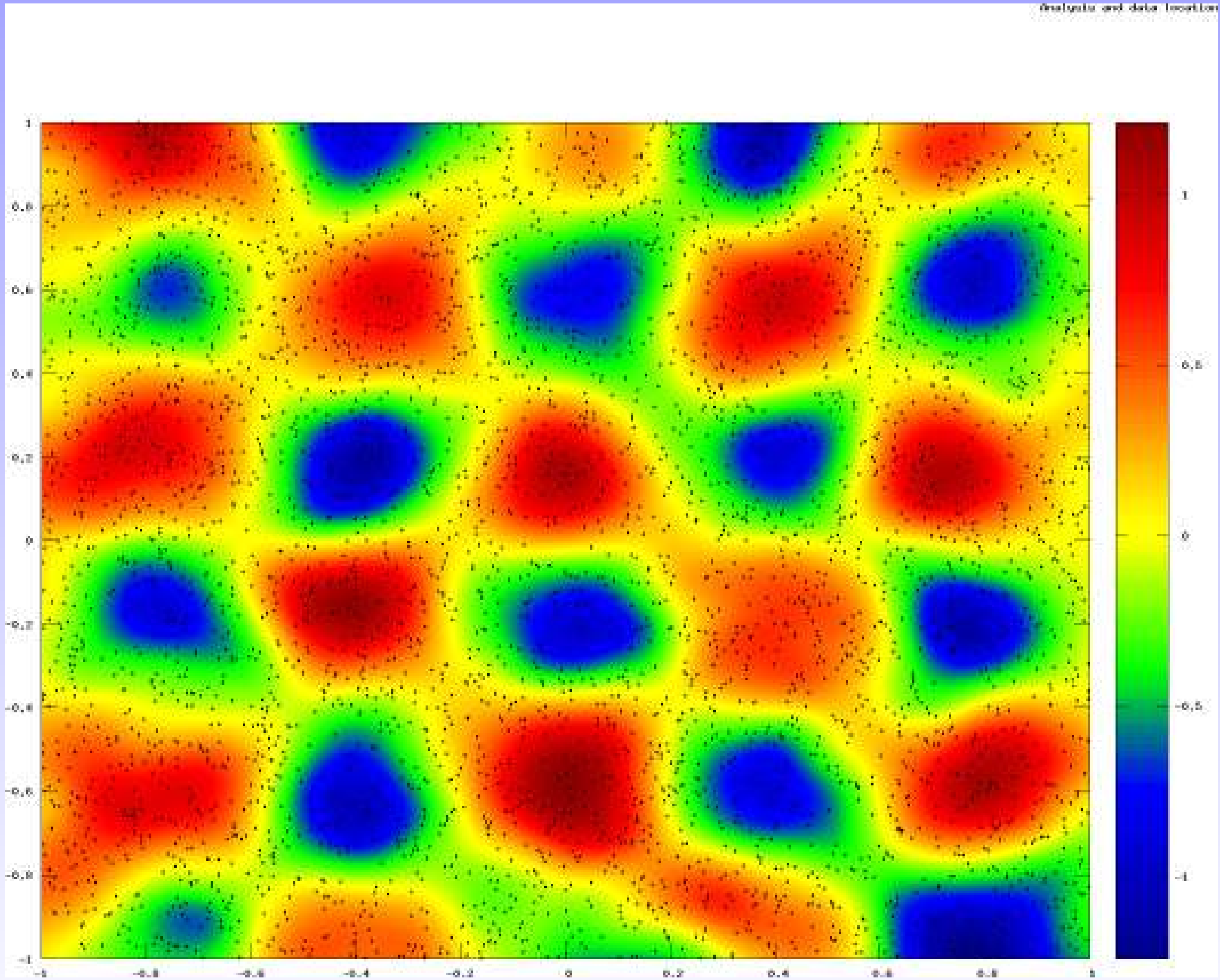
Detrending

Heterogeneous data distribution:



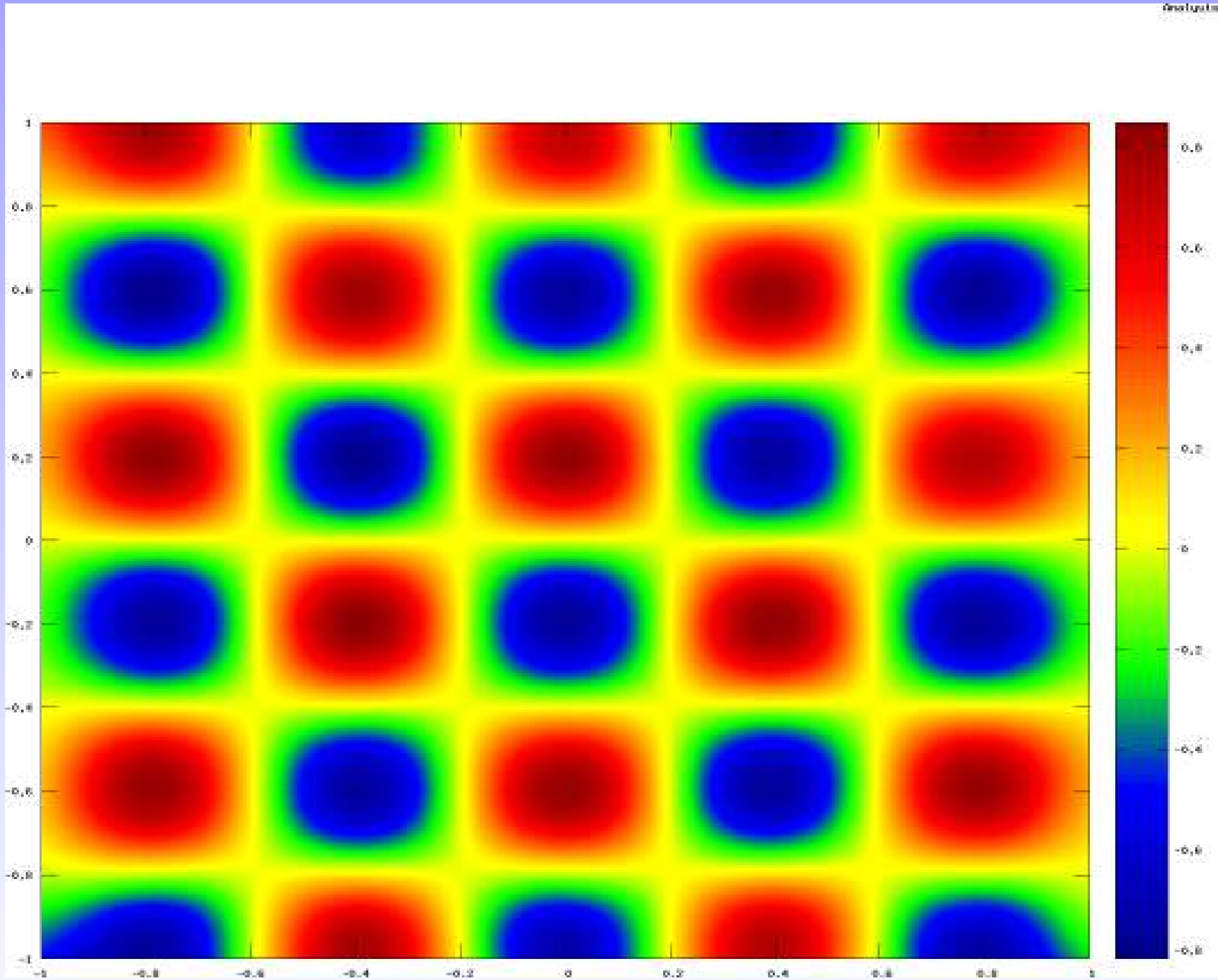
First analysis show a bias for each year's data with respect to first analysis. Subtract the bias estimate and redo the analysis, accumulating the bias. After convergence, detrended analysis+bias of the year.

Example without detrending



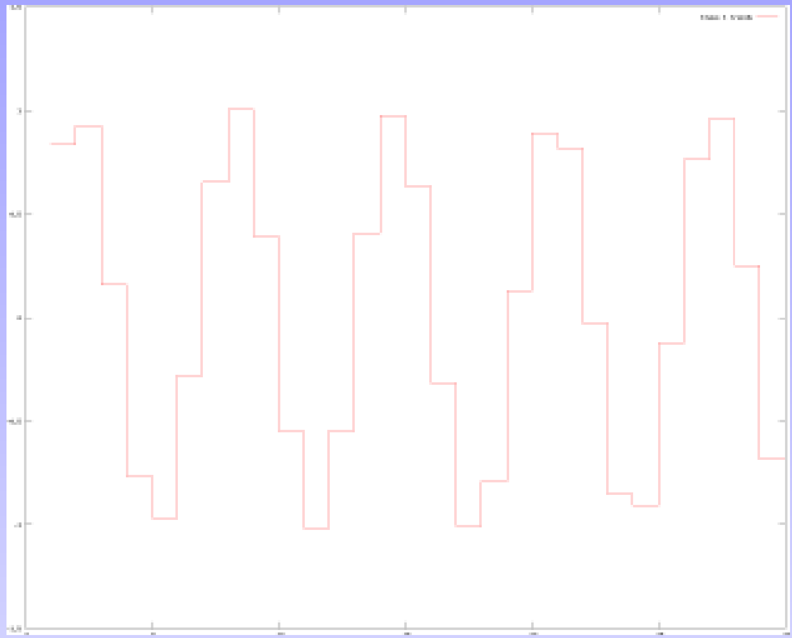
DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.

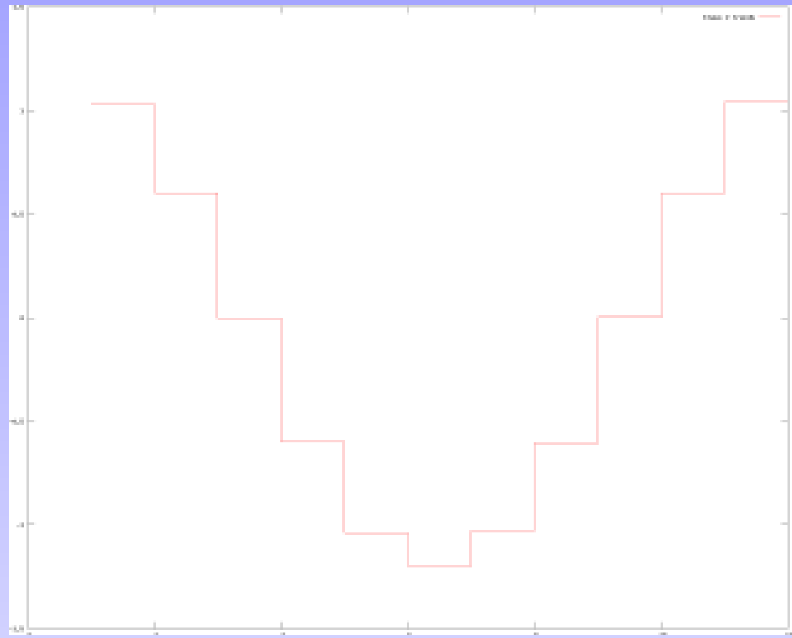
Example with detrending

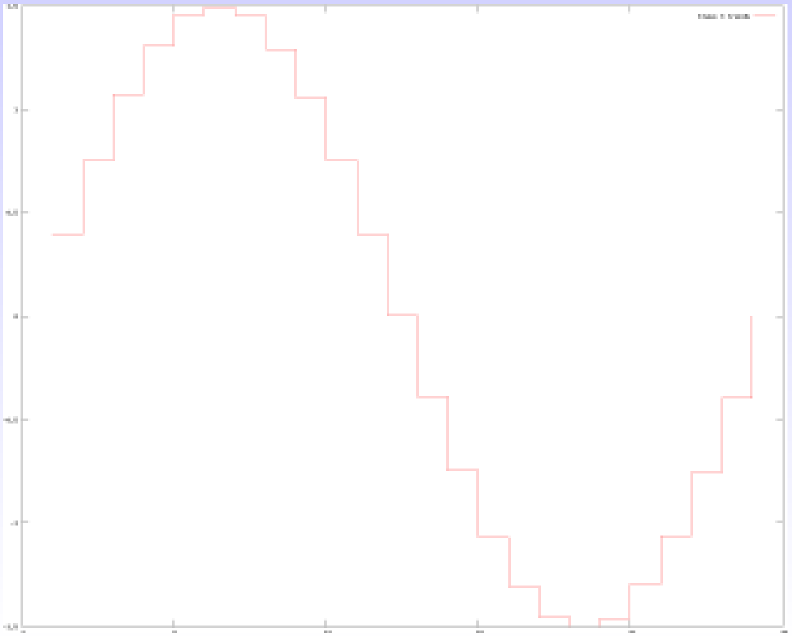


DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.

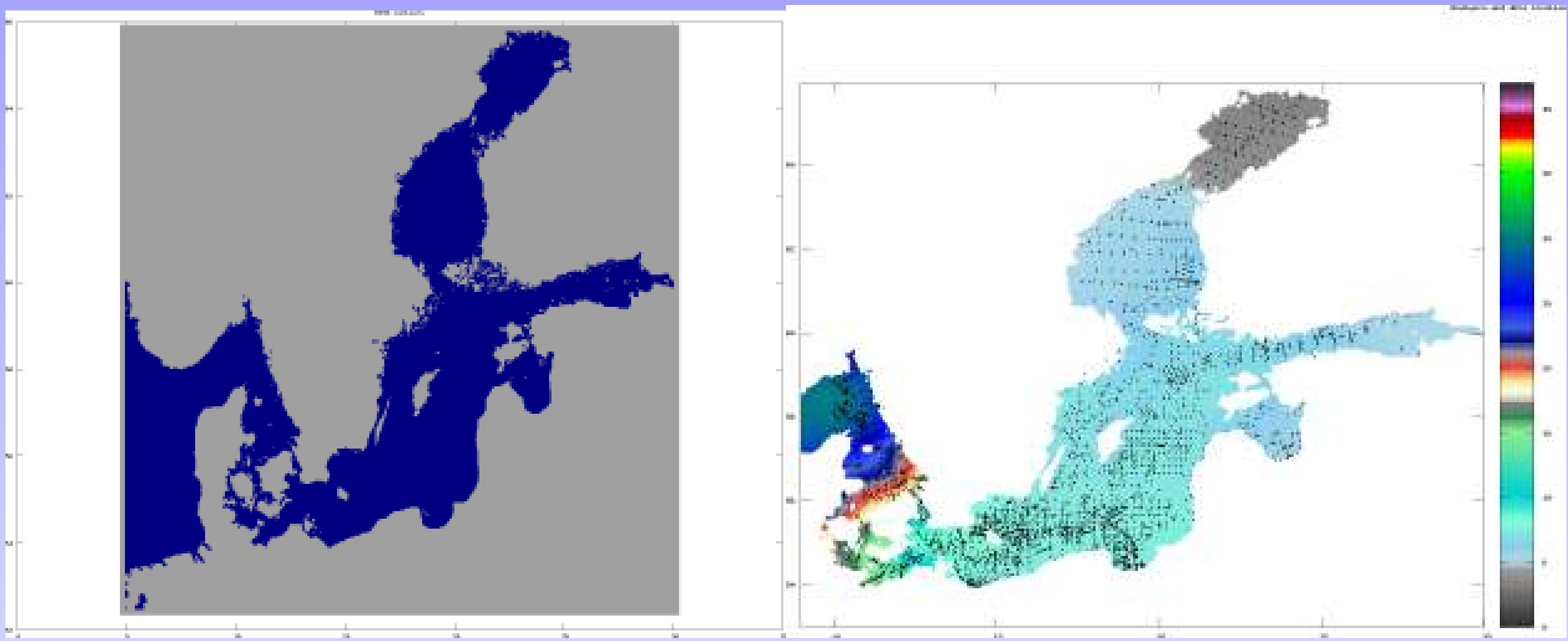
Trends can also be retrieved





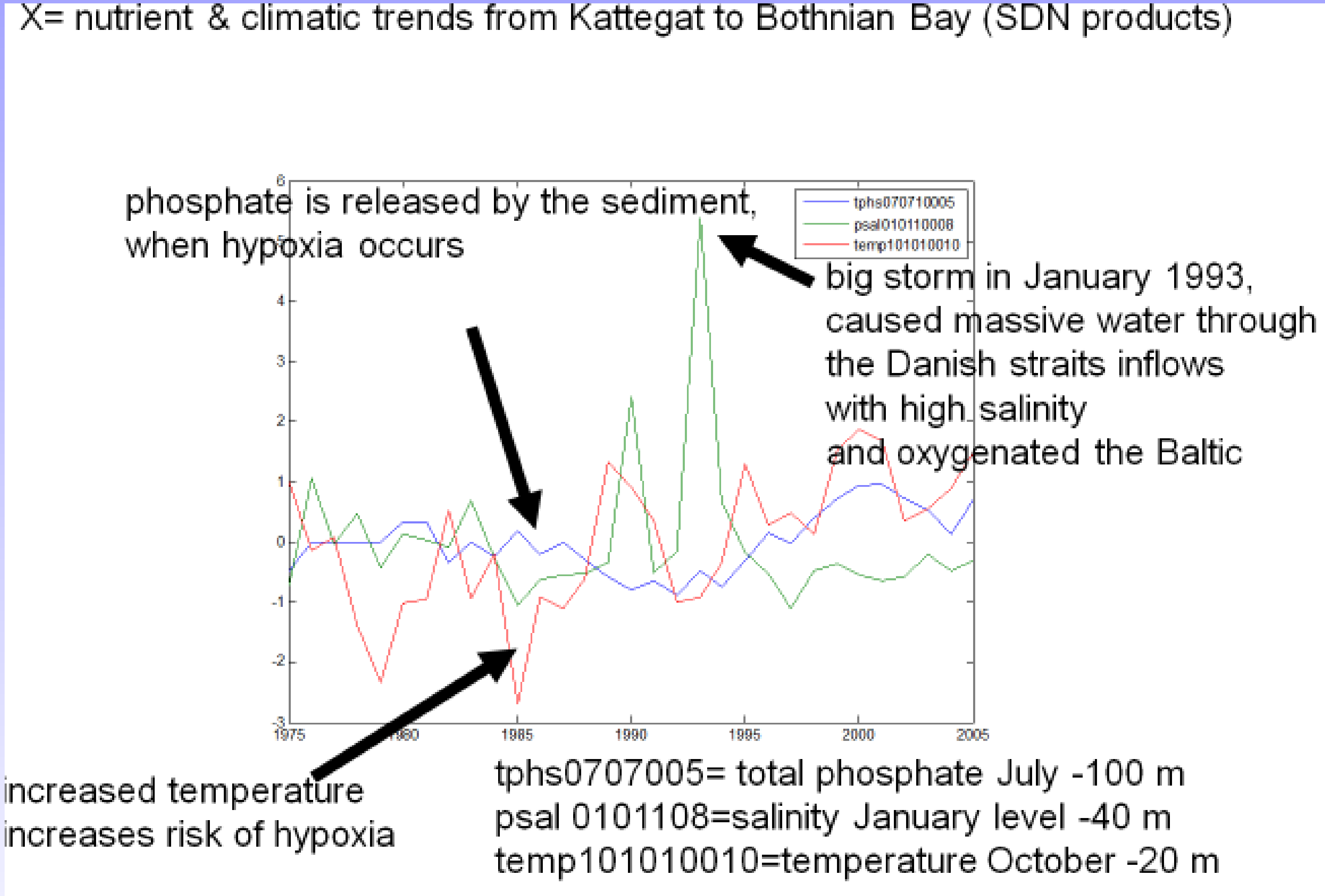


Heterogeneous case



Baltic Salinity Climatology (Bassompierre *et al.*, 2010)

Heterogeneous case, trends



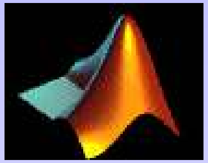
(Bassompierre *et al.*, 2010)

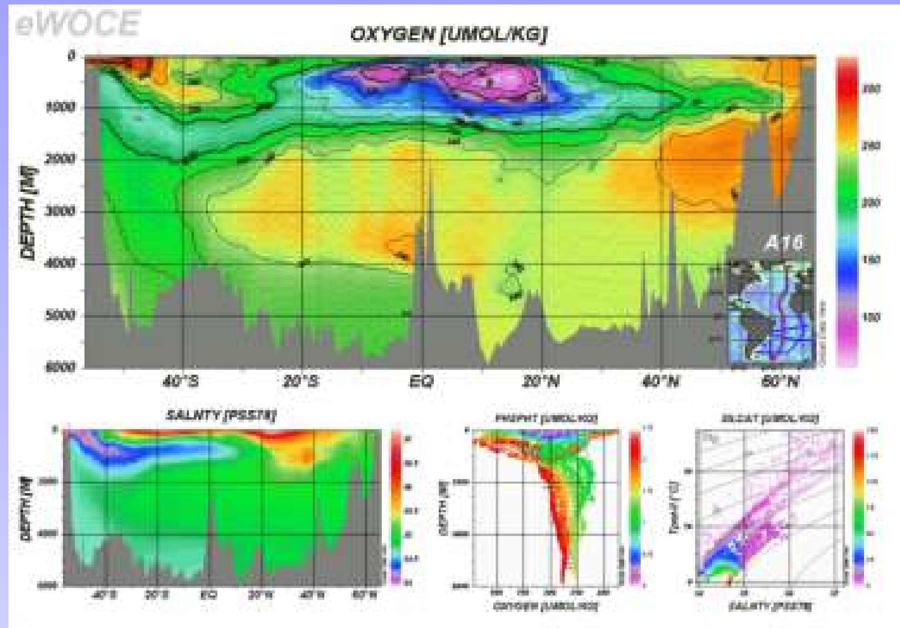
Université de Liège

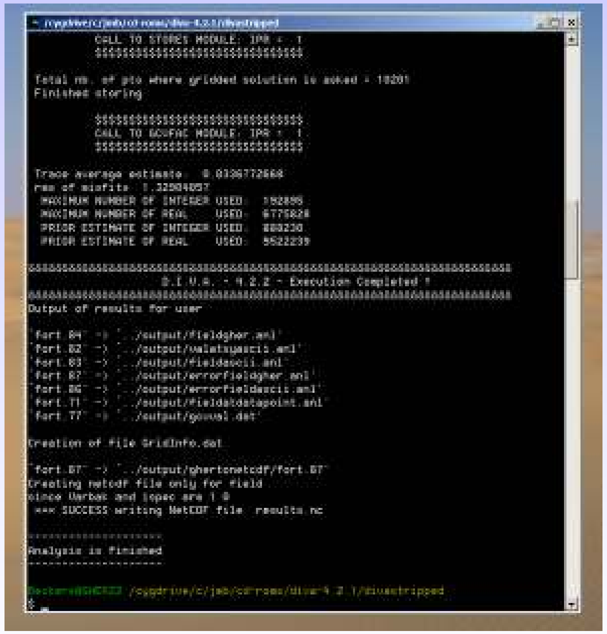
- *Gridding*
- *DIVA theory*
- ***Implementations and exercises***
- *DINEOF theory*
- *Implementations and exercises*

Université de Liège

How to use DIVA?



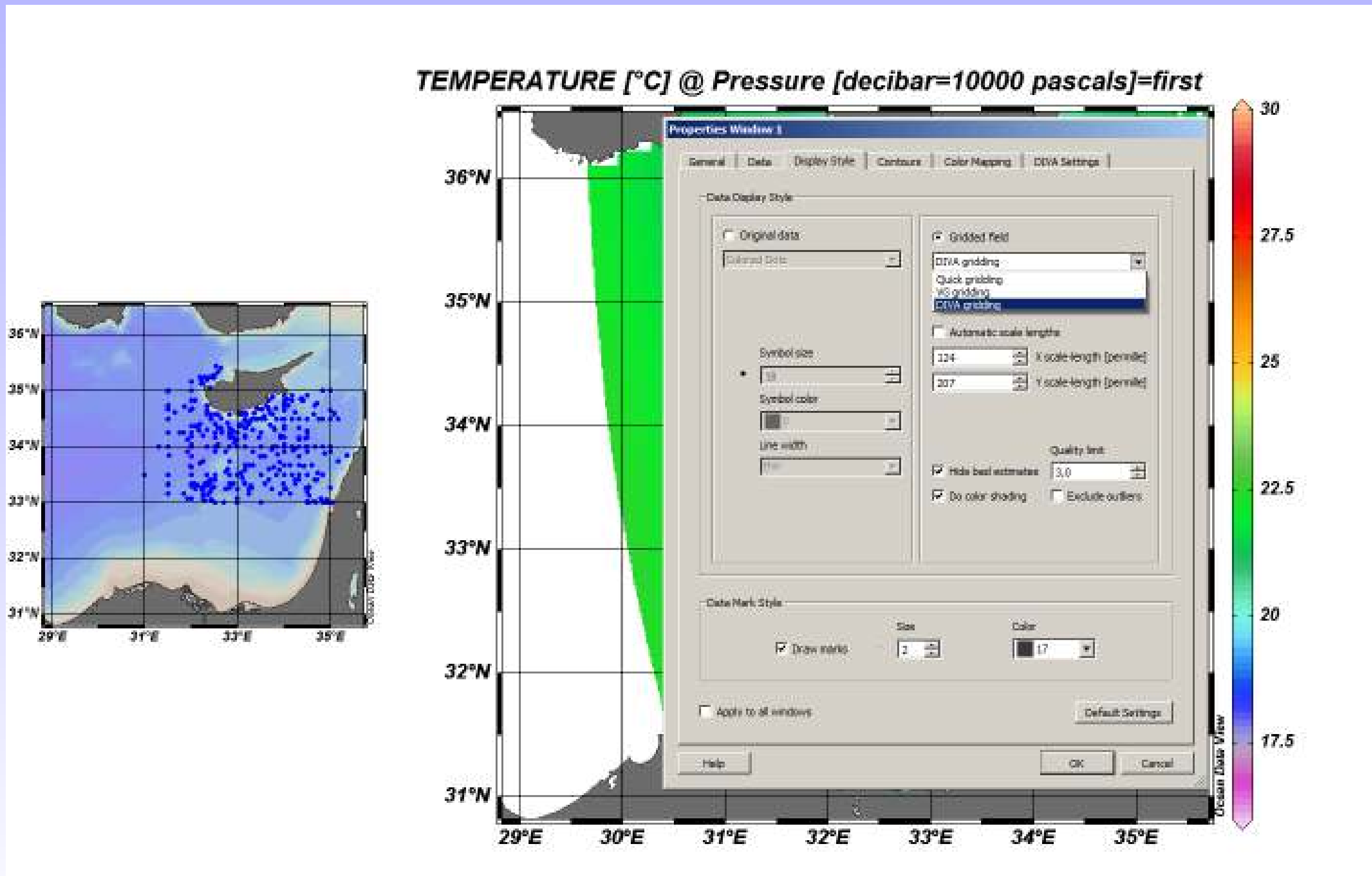




Try parameters of gridding in ODV

On `Display Style`

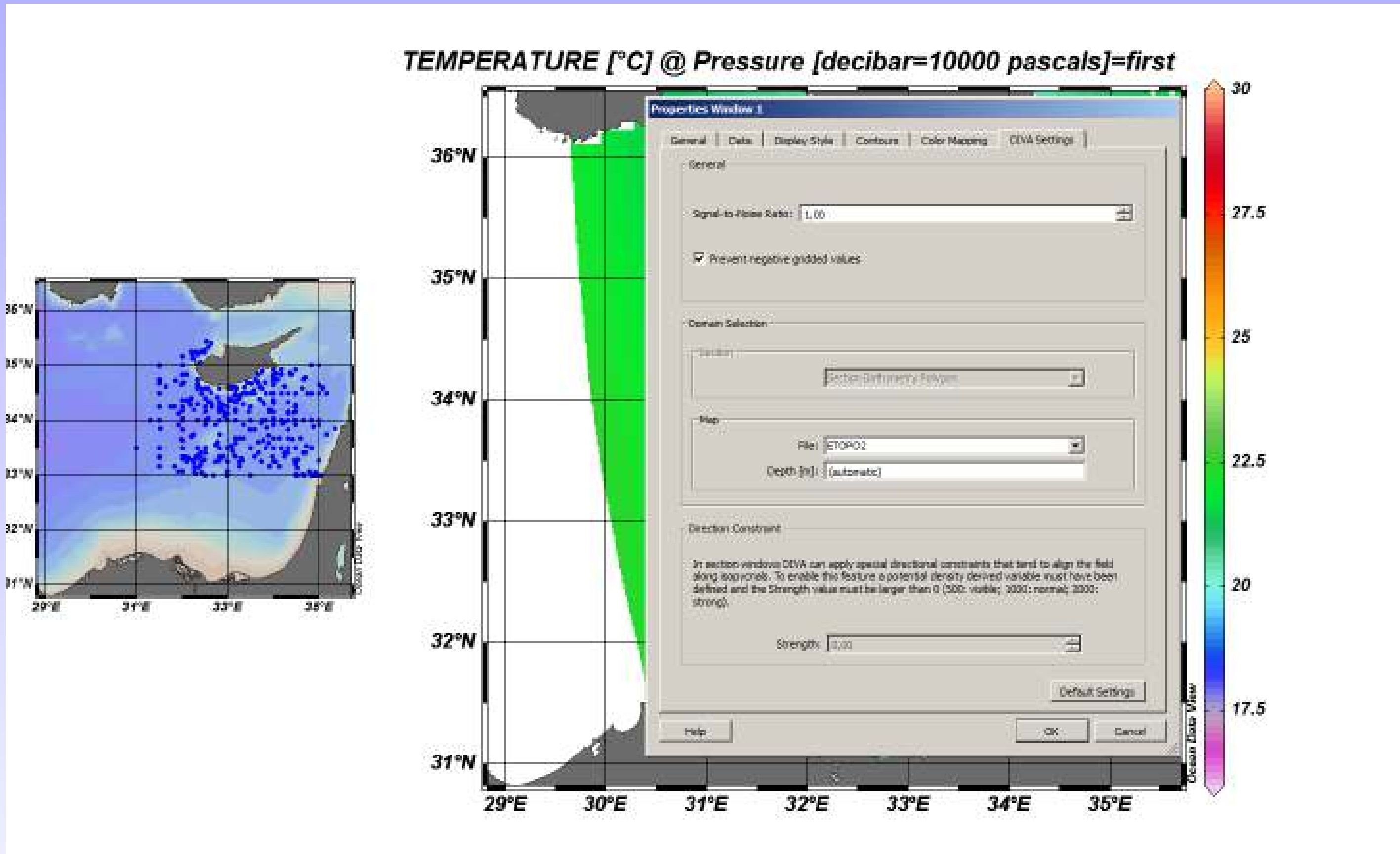
- change analysis method (quick, VG, DIVA)
- change X scale-length and Y scale-length



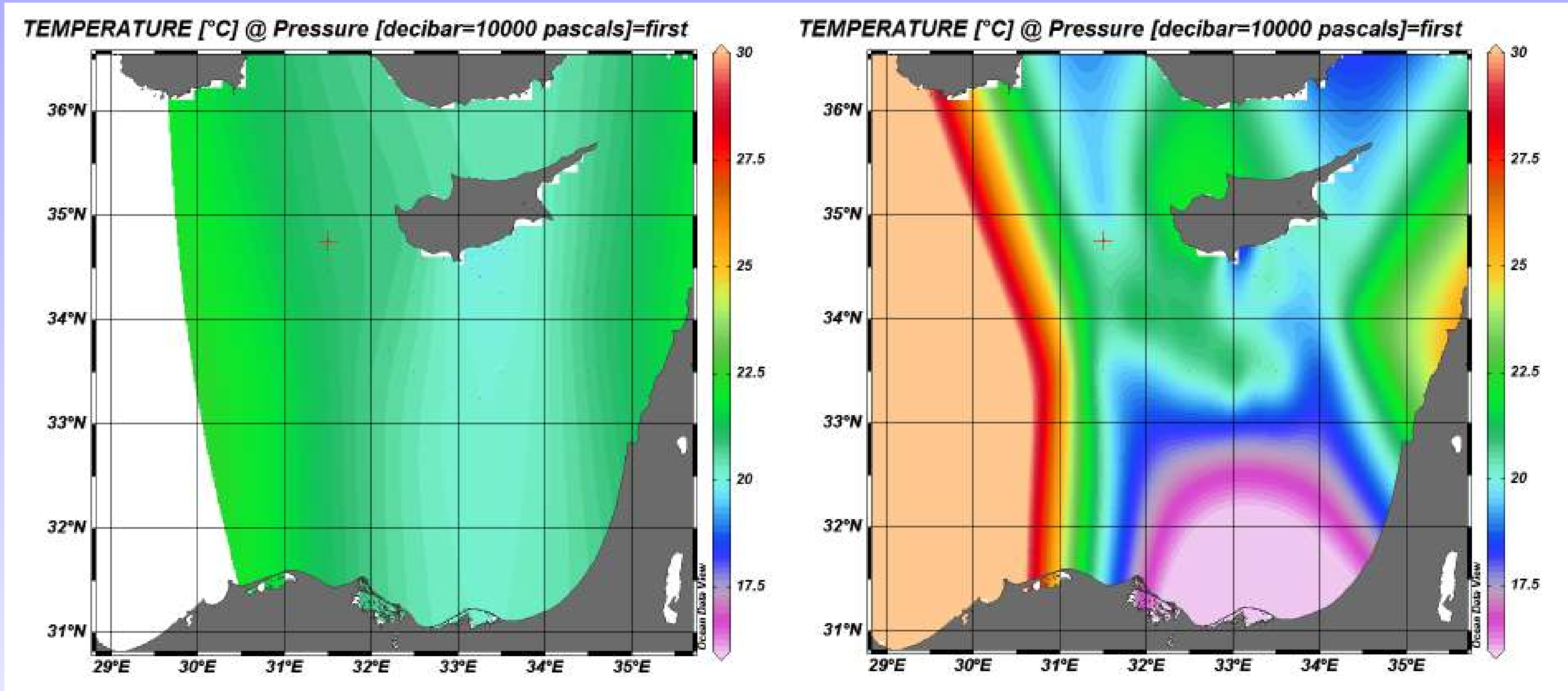
Try parameters of gridding in ODV

On `DIVA Settings`

- Change Signal-to-Noise Ratio

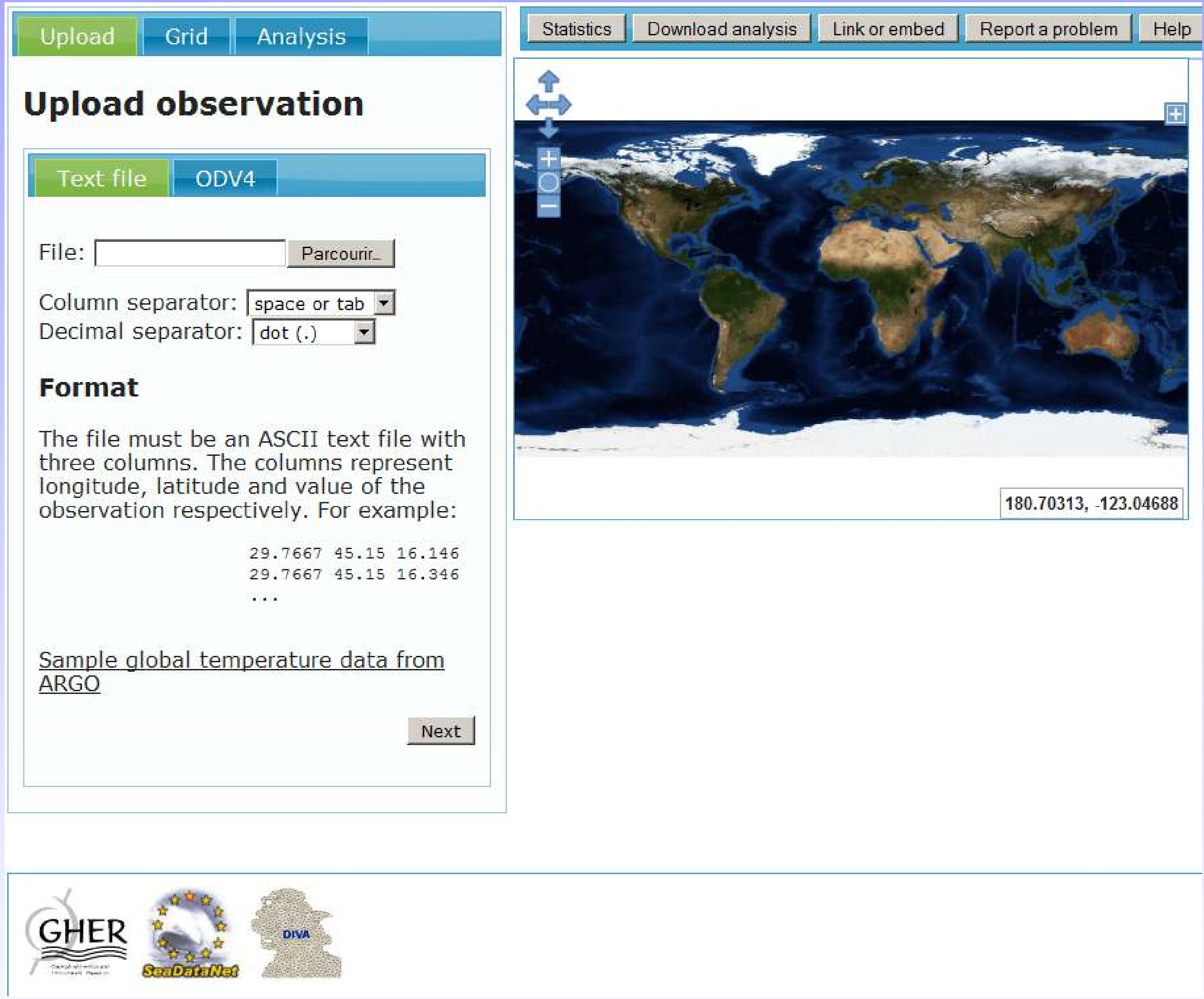


Observed changes in analyses due to changes in parameters (here signal-to-noise ratio from 1 to 100)



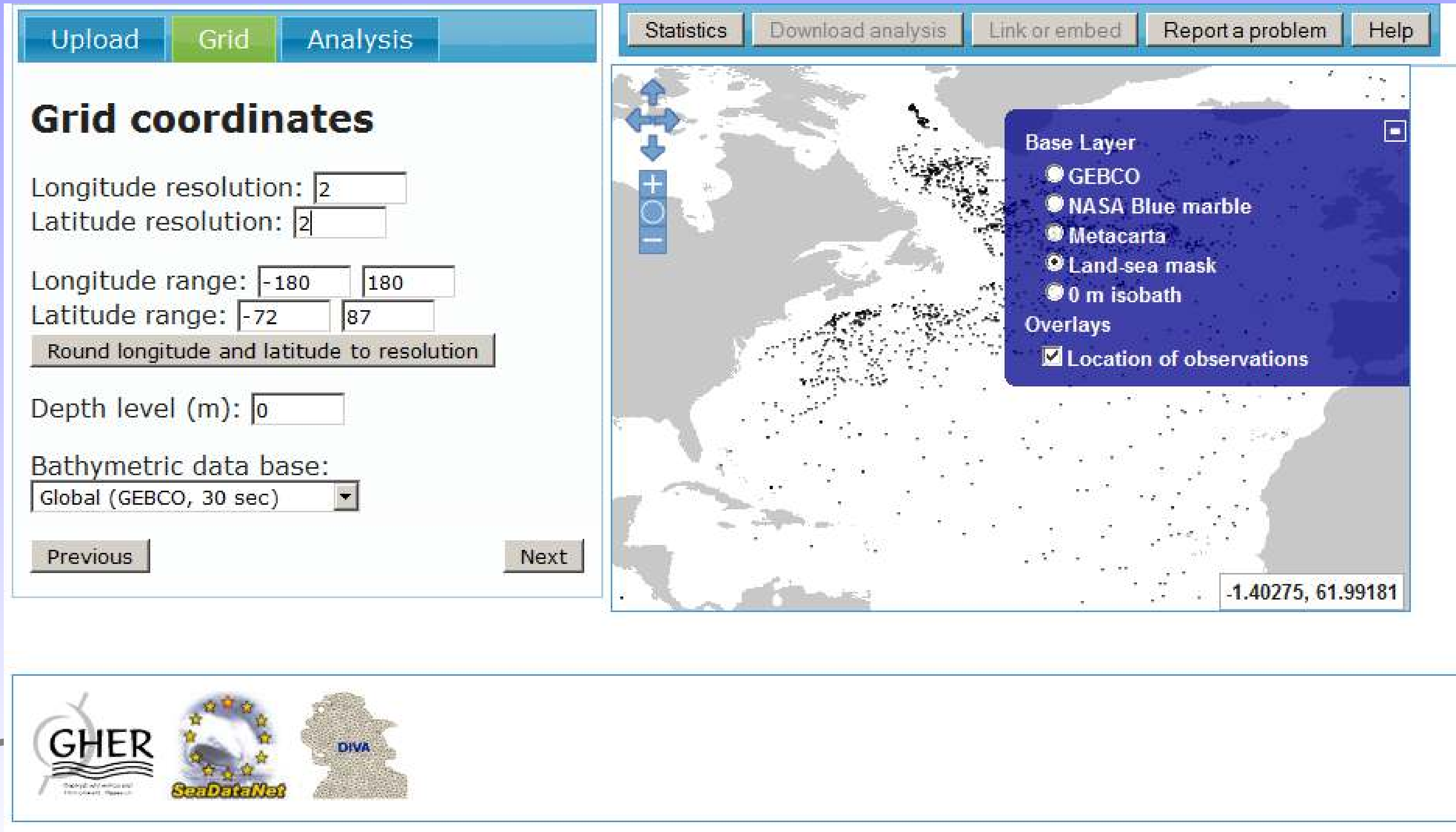
Diva-on-web

http://gher-diva.phys.ulg.ac.be/web-vis/diva.html



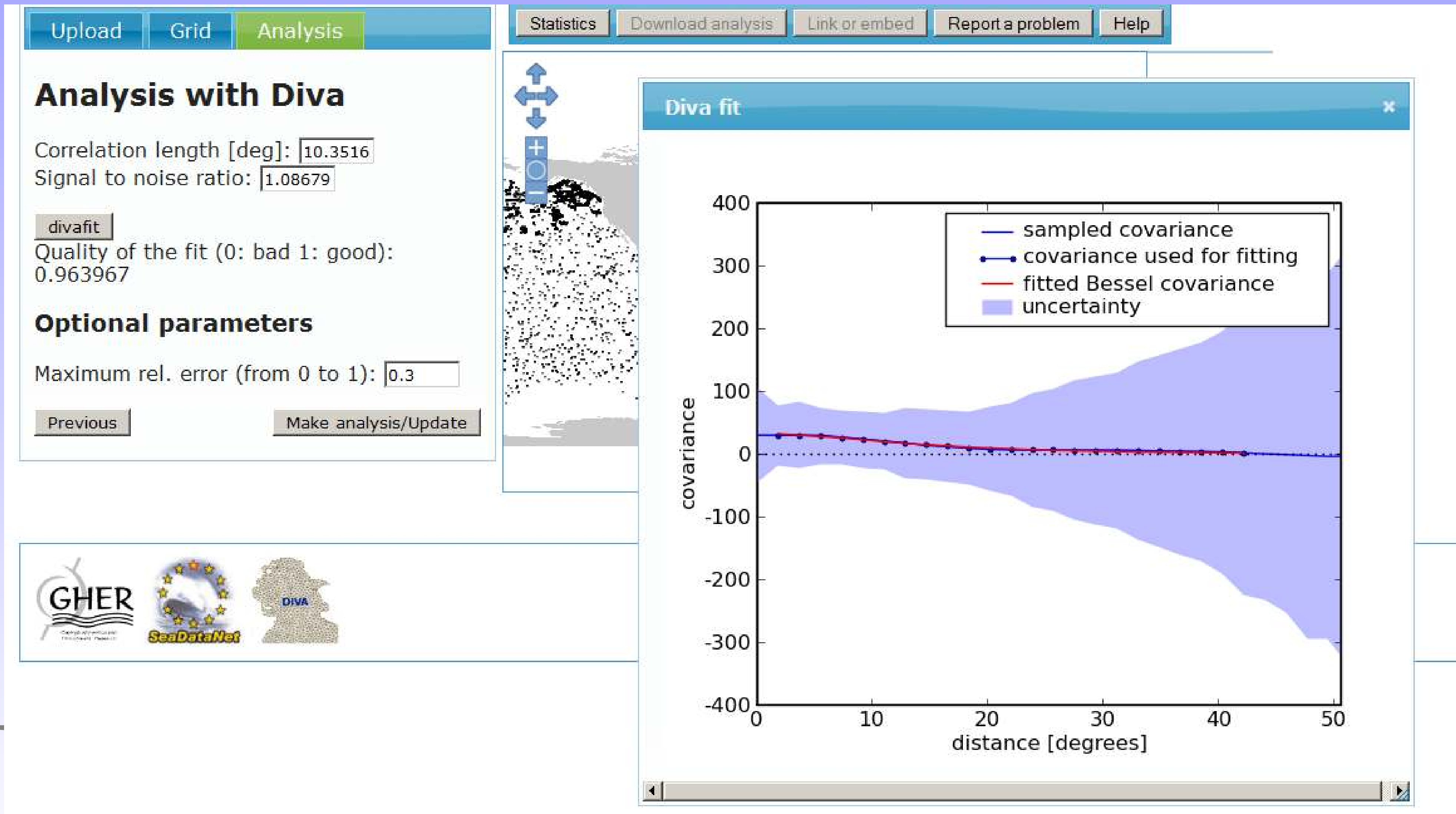
Diva-on-web

Data upload (3 column ascii file or ODV4) and output grid definition



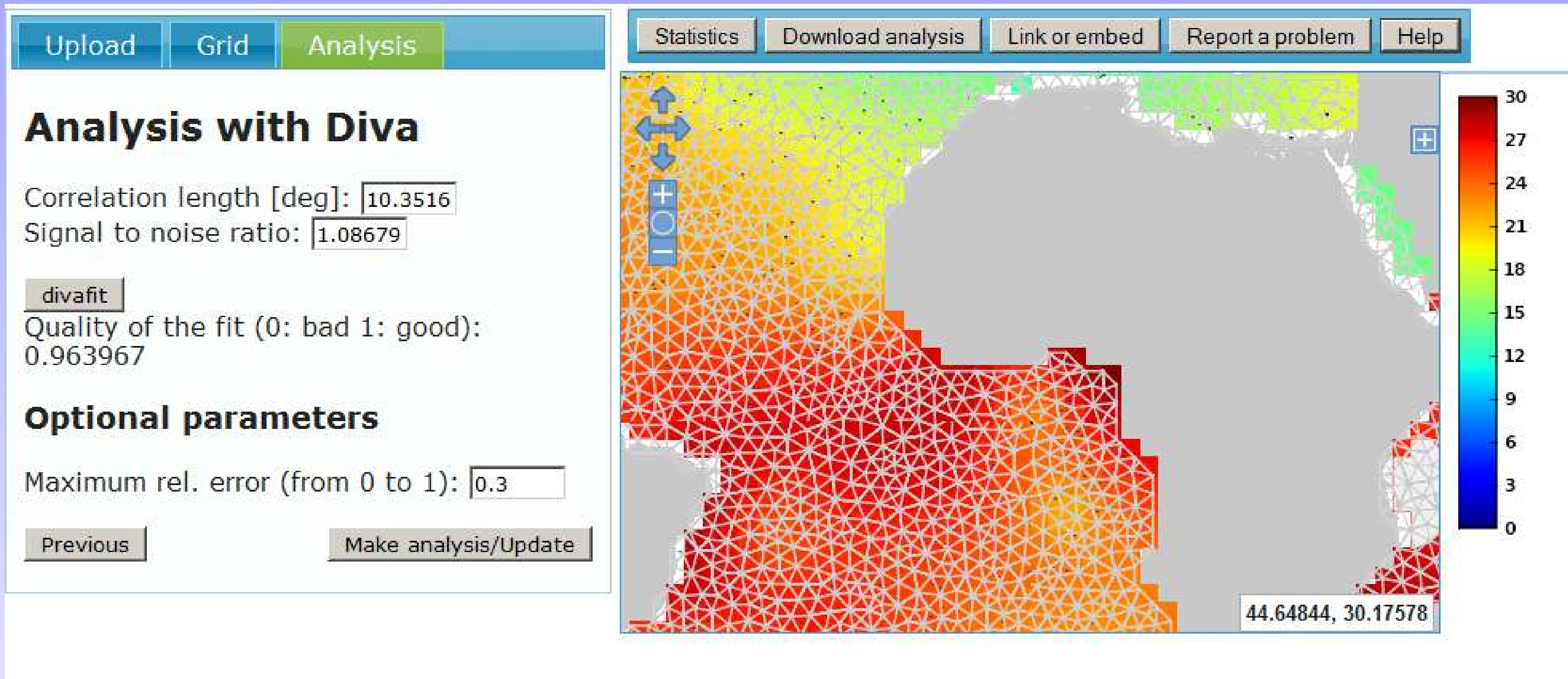
Diva-on-web

Analysis parameter definition (or fit)



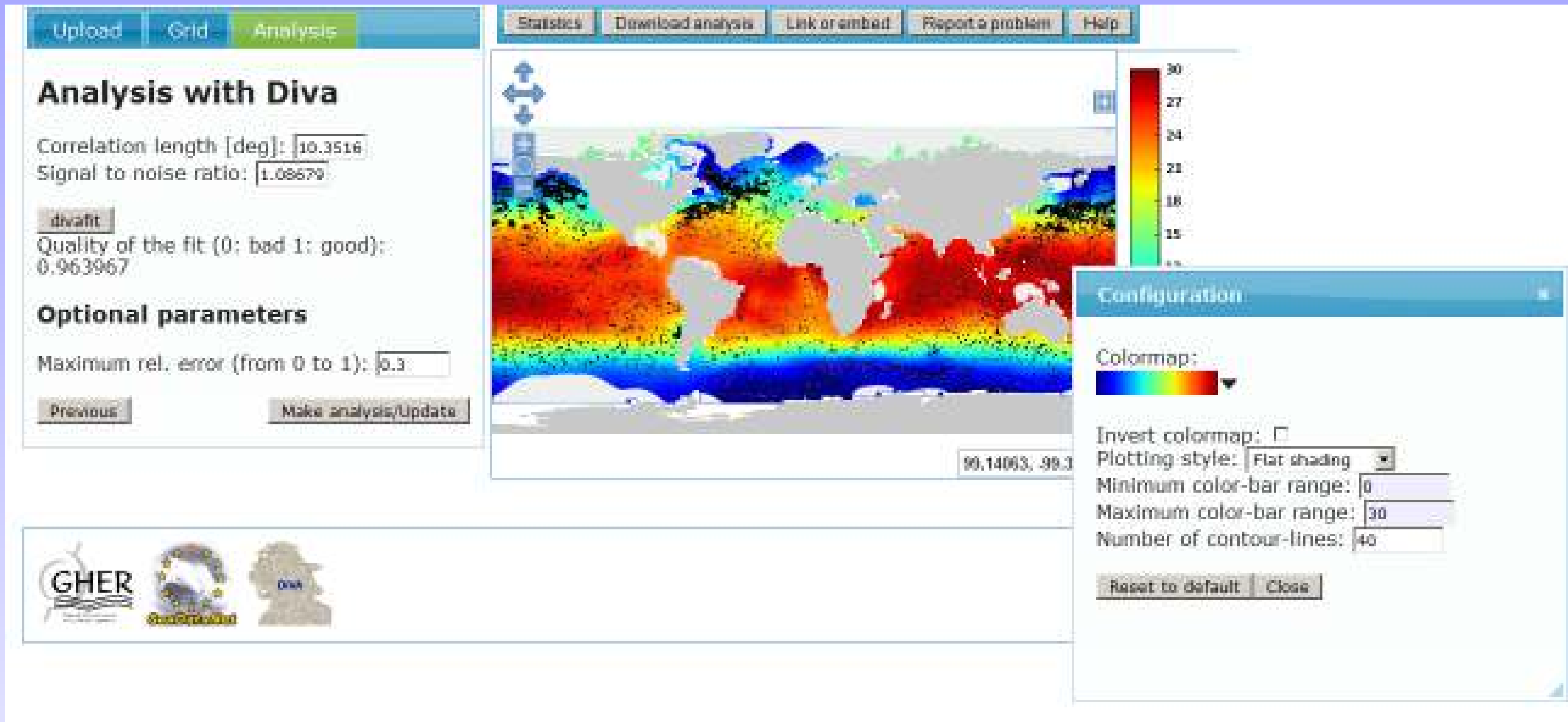
Diva-on-web

Analysis (and numerical grid)



Diva-on-web

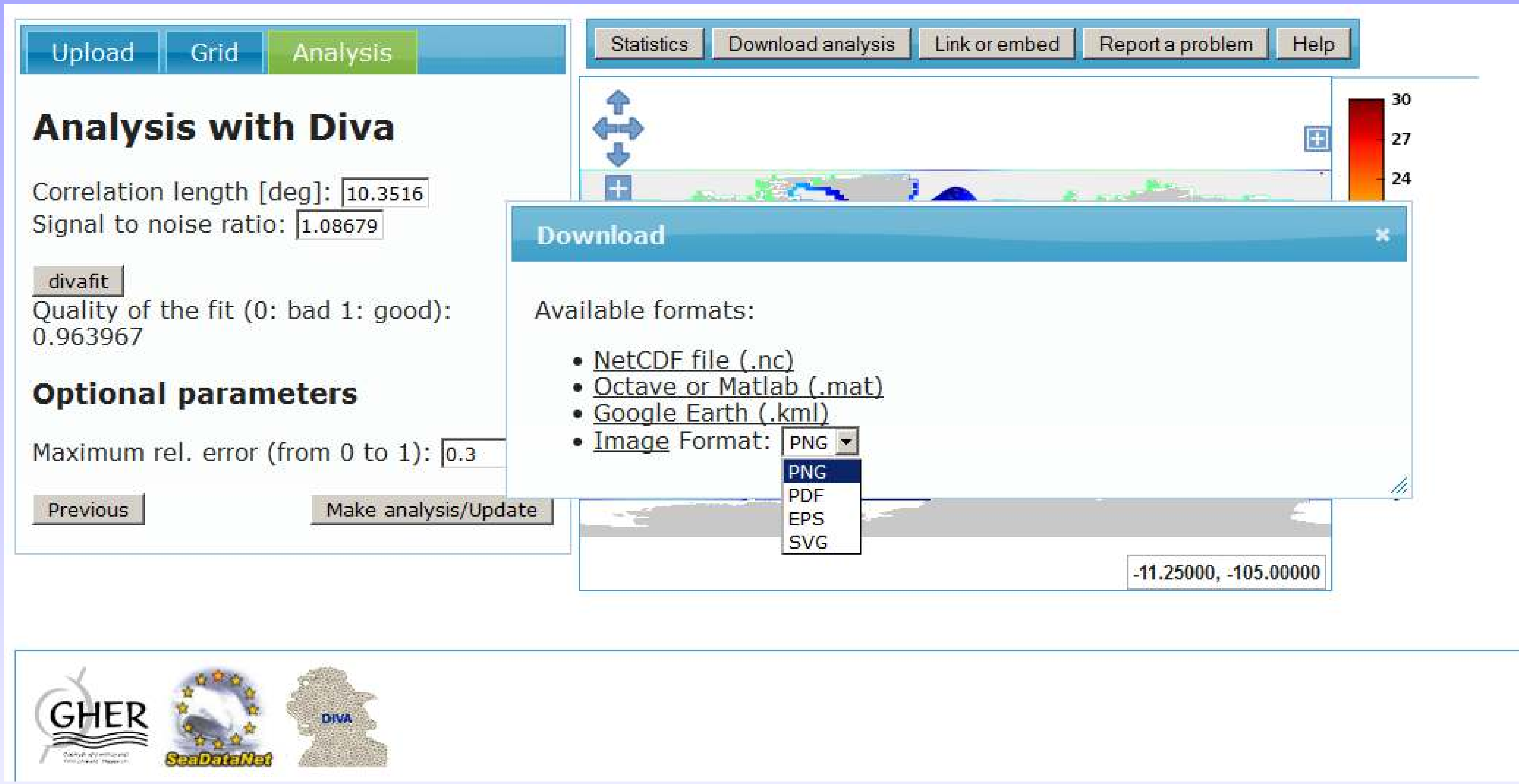
Plotting options



Diva-on-web

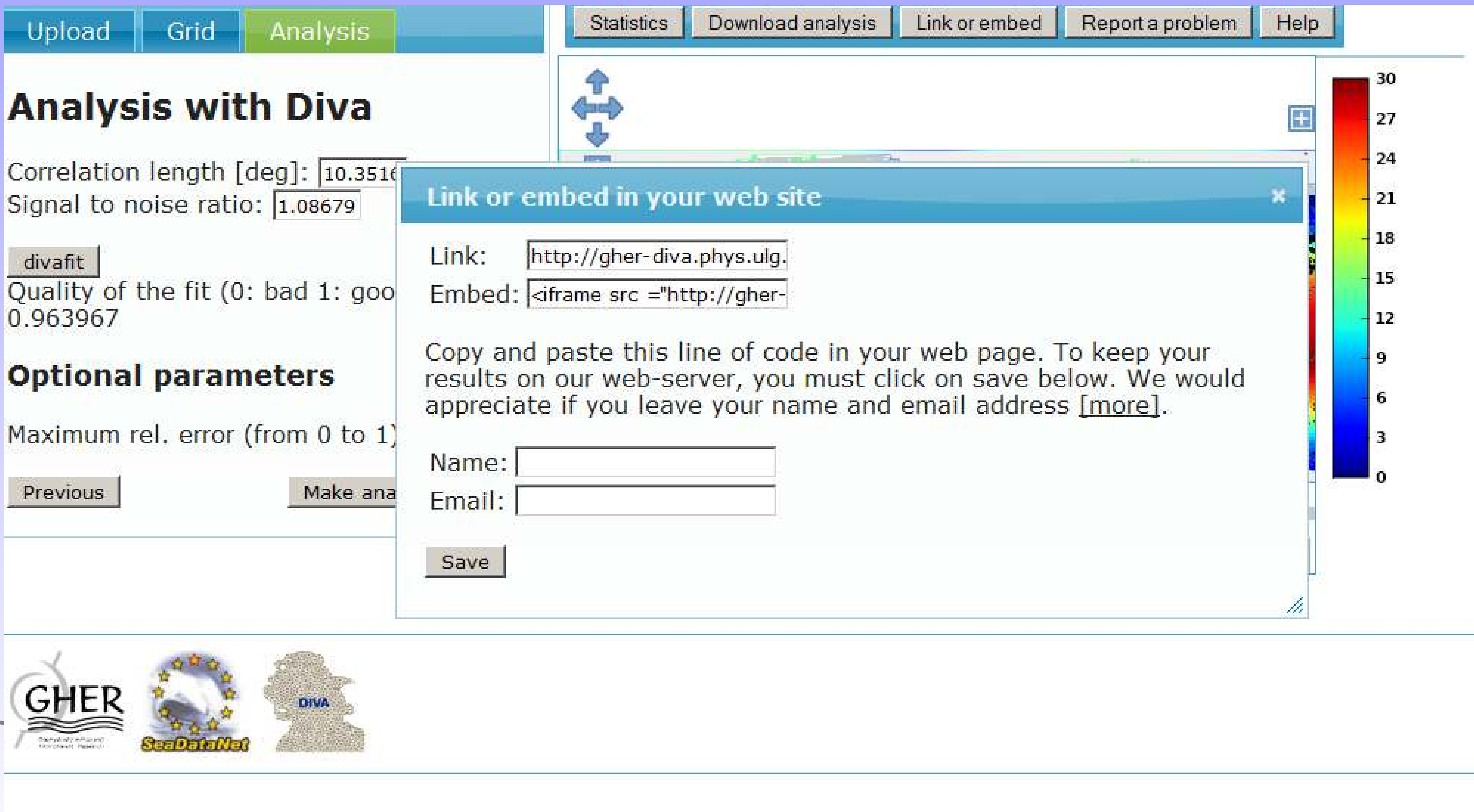
Université de Liège

Download options



Diva-on-web

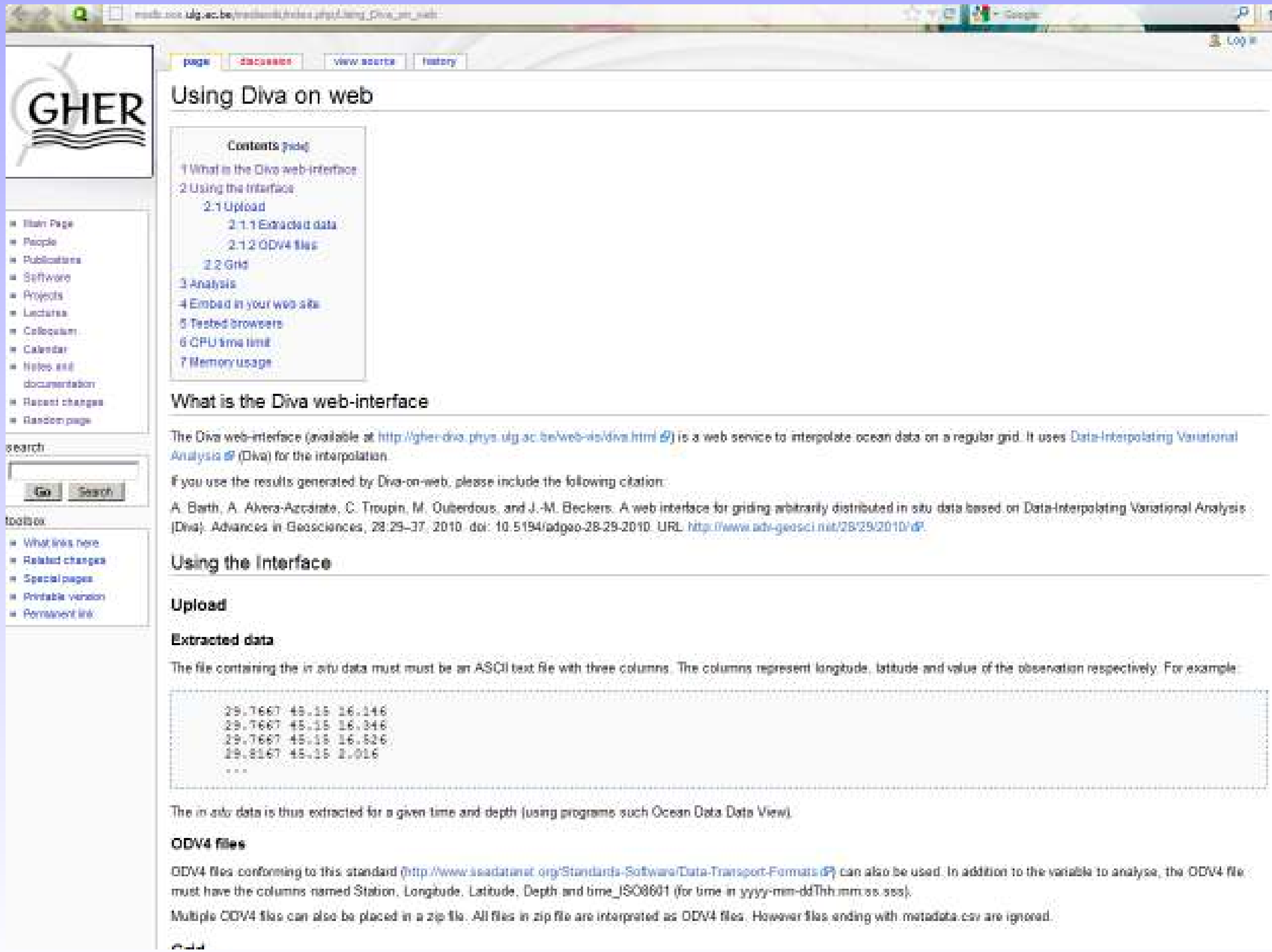
Exploitation of WMS-OGC layering techniques



Help

More information

http://modb.oce.ulg.ac.be/mediawiki/index.php/Using_Diva_on_web



page | discussion | view source | history

Using Diva on web

Contents [hide]

- Main Page
- People
- Publications
- Software
- Projects
- Lectures
- Colloquium
- Calendar
- Notes and documentation
- Recent changes
- Random page

search

Go Search

toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

What is the Diva web-interface

The Diva web-interface (available at http://gher-diva.phys.ulg.ac.be/web-vis/diva.html) is a web service to interpolate ocean data on a regular grid. It uses Data-Interpolating Variational Analysis (Diva) for the interpolation.

If you use the results generated by Diva-on-web, please include the following citation:

A. Barth, A. Alvera-Azcárate, C. Troupin, M. Ouberdous, and J.-M. Beckers. A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (Diva). Advances in Geosciences, 28:29–37, 2010. doi: 10.5194/adgeo-28-29-2010. URL http://www.adv-geosci.net/28/29/2010/.

Using the Interface

Upload

Extracted data

The file containing the *in situ* data must must be an ASCII text file with three columns. The columns represent longitude, latitude and value of the observation respectively. For example:

```
29.7667 45.15 16.146
29.7667 45.15 16.346
29.7667 45.15 16.526
29.8167 45.15 2.016
...
```

The *in situ* data is thus extracted for a given time and depth (using programs such Ocean Data Data View).

ODV4 files

ODV4 files conforming to this standard (http://www.seadatanet.org/Standards-Software/Data-Transport-Formats) can also be used. In addition to the variable to analyse, the ODV4 file must have the columns named Station, Longitude, Latitude, Depth and time_ISO8601 (for time in yyyy-mm-ddThh:mm:ss.sss).

Multiple ODV4 files can also be placed in a zip file. All files in zip file are interpreted as ODV4 files. However files ending with metadata.csv are ignored.

Grid

Software download for advanced users: GODIVA 4D

http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA

GODIVA

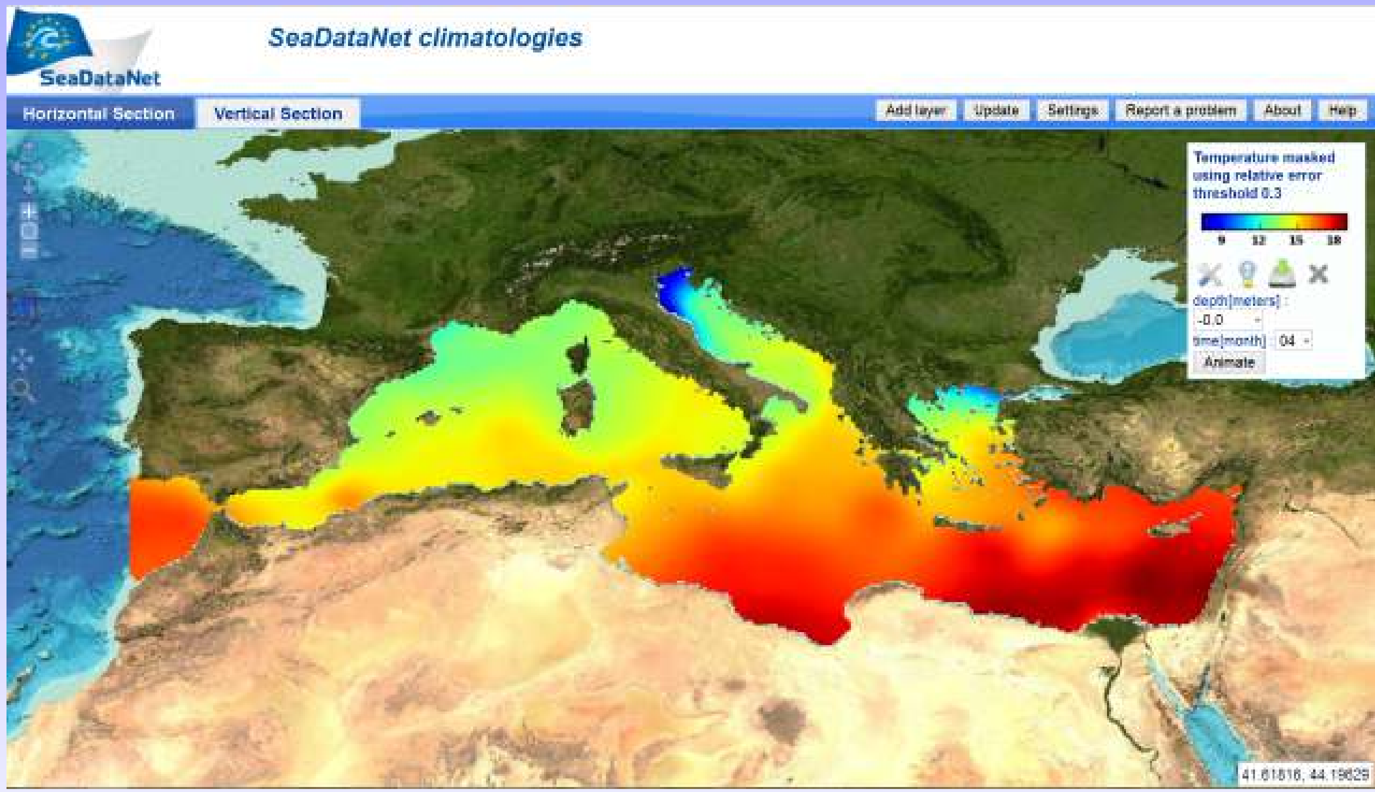
Command line version:
`http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA`
currently version 4.6.5

- All features
- Customizable
- Automatized
- Linux or Cygwin needed
- A few driver files allow automatic creation of climatologies using ODV files.
- Recent developments: parallel solver, more efficient grid generator, approximate and quick error calculations...

Google user group
`http://groups.google.com/group/diva_users` for feedback, questions, suggestions etc; developer's contribution via SVN

OceanBrowser

Results of GODIVA (4D netCDF files with CF conventions), ready for upload on server with OceanBrowser visualization (including vertical sections). OGC-WMS approach.



`http://gher-diva.phys.ulg.ac.be/web-vis/`

Université de Liège

- P. Brasseur, J.-M. Beckers, J.-M. Brankart, and R. Schoenauen. Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of a historical data set. *Deep Sea Research*, 43:159–192, 1996. http://orbi.ulg.ac.be/handle/2268/4788
- Ch. Troupin, F. Machin, M. Ouberdous, D. Sirjacobs, A. Barth, J.-M. Beckers. High-resolution Climatology of the North-East Atlantic using Data-Interpolating Variational Analysis (Diva). *Journal of Geophysical Research*, **114**, 2010. http://orbi.ulg.ac.be/handle/2268/68400
- M. Rixen, J.-M. Beckers, J.-M. Brankart, and P. Brasseur. A numerically efficient data analysis method with error map generation. *Ocean Modelling*, 2:45–60, 2000. http://orbi.ulg.ac.be/handle/2268/40371
- A. Karafistan, J.-M. Martin, M. Rixen, and J.-M. Beckers. Space and time distributions of phosphates in the Mediterranean Sea. *Deep Sea Research*, 49:67–82, 2002. http://orbi.ulg.ac.be/handle/2268/4289
- Tyberghein, L., Verbruggen, H., Klaas, P., Troupin, C., Mineur, F., De Clerck, O., 2011. ORACLE: a global environmental dataset for marine species distribution modeling. *Global Ecology and Biogeography.* http://orbi.ulg.ac.be/handle/2268/112937
- Barth, A., Alvera-Azcárate, A., Troupin, C., Ouberdous, M., Beckers, J.M., 2010. A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA). *Adv. Geosci.* 28, 29-37. http://orbi.ulg.ac.be/handle/2268/81401

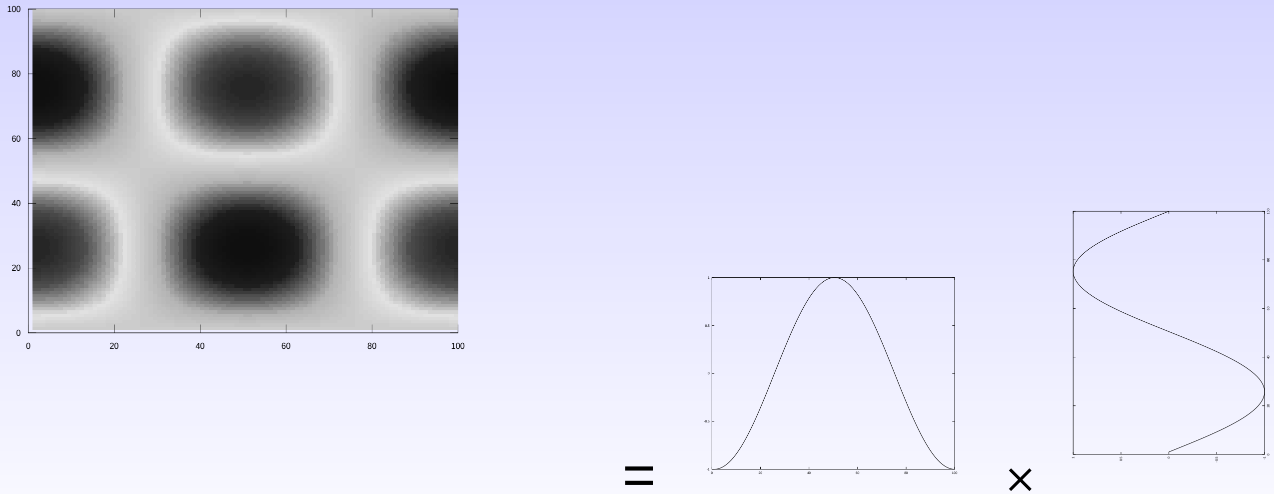
- M. Rixen, J.-M. Beckers, S. Levitus, J. Antonov, T. Boyer, C. Maillard, M. Fichaut, E. Balopoulos, S. Iona, H. Dooley, M.-J. Garcia, B. Manca, A. Giorgetti, G. Manzella, N. Mikhailov, N. Pinardi, M. Zavatarelli, and the Medar Consortium. The Western Mediterranean Deep Water: a proxy for global climate change. *Geophysical Research Letters*, 32, 2005. `http://orbi.ulg.ac.be/handle/2268/4299`
- C. Troupin, A. Barth, D. Sirjacobs, M. Ouberdous, J.-M. Brankart, P. Brasseur, M. Rixen, A. Alvera-Azcárate, M. Belounis, A. Capet, F. Lenartz, M.-E. Toussaint, J.-M. Beckers. Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA). *Ocean Modelling*, **52**, 90-101, 2012. `http://orbi.ulg.ac.be/handle/2268/125731`

Université de Liège

- *Gridding*
- *DIVA theory*
- *Implementations and exercises*
- ***DINEOF theory***
- *Implementations and exercises*

EOFs: Generalisations and other names

- Empirical orthogonal functions (EOFs)
- Proper orthogonal decomposition (POD)
- Karhunen-Loève decompositions (best to start from for non-uniform data distribution)
- Proper orthogonal modes (POM)
- Principal component analysis (PCA)



Université de Liège

Classical approach for EOFs

- We assume that we have a matrix $\mathbf{X}$ containing the observations, which is arranged such that the element i, j of the matrix is called $(\mathbf{X})_{ij}$ and is given by the value of the field $f(\boldsymbol{r}, t)$ at location $\boldsymbol{r}_i$ and moment t_j :

$$(\mathbf{X})_{ij} = f(\boldsymbol{r}_i, t_j). \tag{45}$$

- The field f is an observational field and contains thus all errors (instrumental, unresolved structures, etc).
- We can then write the matrix as a succession of n column vectors

$$\mathbf{X} = \left(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\right), \tag{46}$$

each of the column vectors $\mathbf{x}_j$ being the discrete state vector of size m at moment t_j.

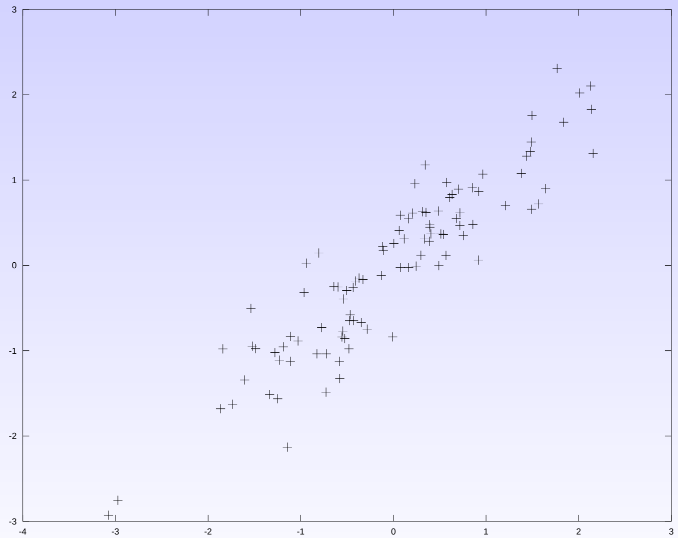
Defining a mode

Try to find a spatial structure $\mathbf{u}$ which represents at best the data:

> Maximize norm of $\mathbf{X}^\star\mathbf{u}$ with normalization constraint on $\mathbf{u}$.
>
> Try to find the direction in which the data have their largest component.

Find extrema of functional J with Lagrange multiplier

$$J = \mathbf{u}^\star \mathbf{X}\mathbf{X}^\star \mathbf{u} - \lambda\left(\mathbf{u}^\star\mathbf{u} - 1\right) \tag{47}$$



Euler-Lagrange equation

- Variations in (47) on $\mathbf{u}$:

$$\mathbf{X}\mathbf{X}^{\star}\mathbf{u} = \lambda\mathbf{u} \tag{48}$$

- Variations on λ

$$\mathbf{u}^{\star}\mathbf{u} = 1 \tag{49}$$

EOFs are normalized eigenvectors of "covariance"[a] matrix $\mathbf{X}\mathbf{X}^{\star}$which is a symmetric positive defined matrix: real positive eigenvalues and orthogonal eigenvectors:

$$\lambda_i = \rho_i^2 \qquad \text{conventionally } \rho_i > 0 \text{ and } \rho_{i+1} \leq \rho_i \tag{50}$$

$$\mathbf{u_i}^{\star}\mathbf{u_j} = \delta_{ij} \tag{51}$$

Storing $\mathbf{u_i}$ as columns in $\mathbf{U}$ yields $\mathbf{U}^{\star}\mathbf{U} = \mathbf{I}$

[a] " " because missing 1/n. Observe time summation to get spatial covariances

Temporal amplitudes

What is the temporal evolution $\mathbf{g_j}$ of the amplitudes of a mode $\mathbf{u_j}$?

Temporal amplitudes

What is the temporal evolution $\mathbf{g_j}$ of the amplitudes of a mode $\mathbf{u_j}$?
See it as minimizing the norm of $\mathbf{X} - \mathbf{u_j}\mathbf{g_j}^\star$

Temporal amplitudes

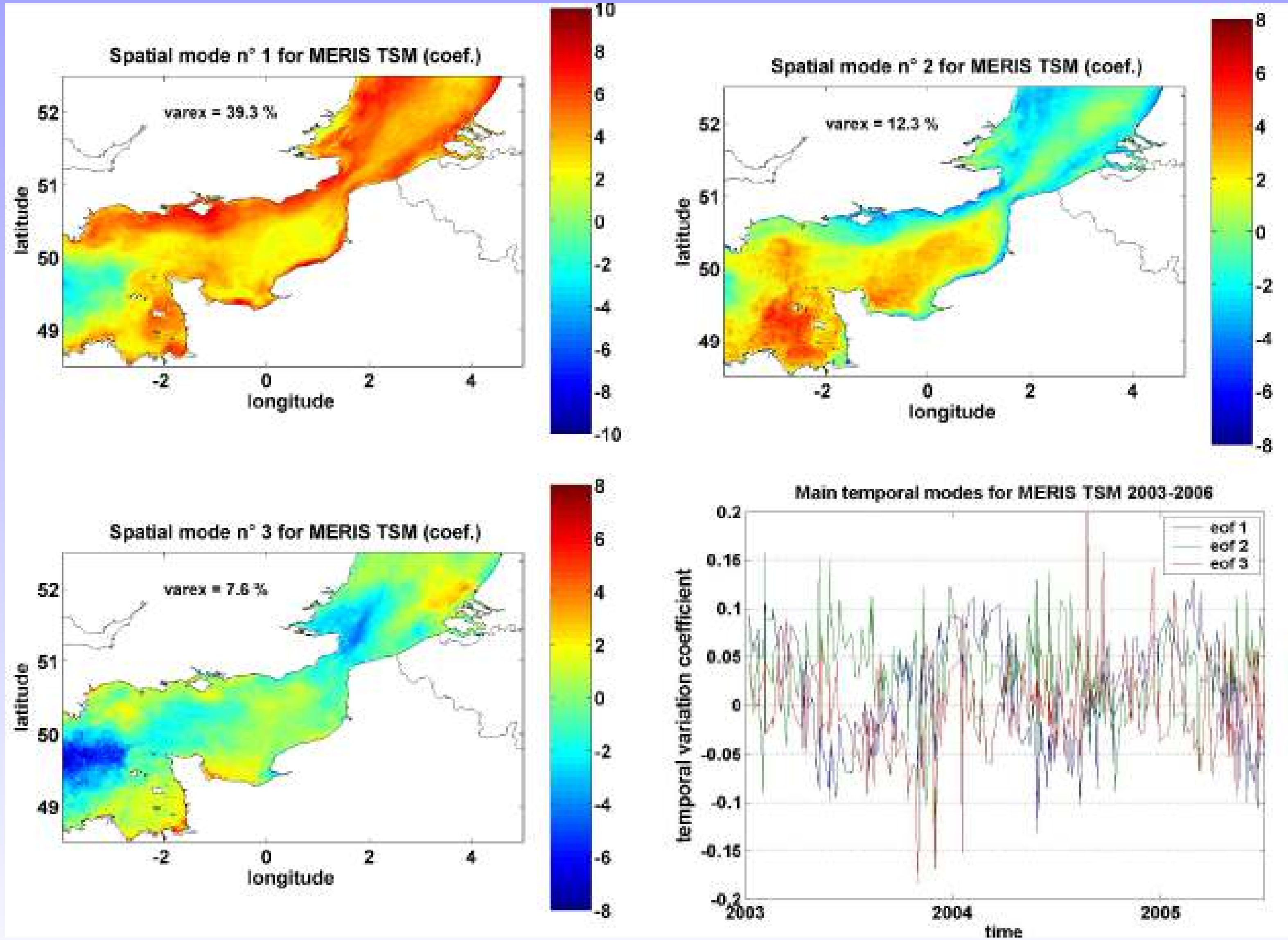
What is the temporal evolution $\mathbf{g_j}$ of the amplitudes of a mode $\mathbf{u_j}$?
See it as minimizing the norm of $\mathbf{X} - \mathbf{u_j g_j}^\star$

$$\mathbf{g_j} = \mathbf{X}^\star \mathbf{u_j} \tag{52}$$

Projection of data of a given moment onto the spatial mode $\mathbf{u_j}$

Université de Liège

Example of EOF



Temporal modes

Using temporal covariances $\mathbf{X}^\star\mathbf{X}$ instead of spatial covariances $\mathbf{X}\mathbf{X}^\star$ yields temporal EOFs :

$$\mathbf{X}^\star\mathbf{X}\,\mathbf{v} = \mu\mathbf{v} \tag{53}$$

$$\mathbf{v}^\star\mathbf{v} = 1 \tag{54}$$

$$\mathbf{V}^\star\mathbf{V} = \mathbf{I} \tag{55}$$

Link between projection of data and temporal modes? Yes,

$$\mathbf{g_j} = \mathbf{X}^\star\mathbf{u_j} = \rho_j\mathbf{v_j} \tag{56}$$

easily proven via SVD decomposition of a matrix

Université de Liège

SVD

Singular value decomposition of a data matrix **X**:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\star = \sum_{k=1}^{q} \rho_k \mathbf{u}_k \mathbf{v}_k^\star \tag{57}$$

leads to

- spatial (**U**) and temporal (**V**) EOFs (orthonormal)
- singular values ρ_i (stored on the diagonal of **D**)

They are also solution of

$$\begin{aligned} \mathbf{X}\mathbf{v} &= \rho\mathbf{u} \\ \mathbf{X}^\star\mathbf{u} &= \rho\mathbf{v}, \end{aligned} \quad \Rightarrow \quad \begin{aligned} \mathbf{X}\mathbf{X}^\star\mathbf{u} &= \rho^2\mathbf{u} \\ \mathbf{X}^\star\mathbf{X}\mathbf{v} &= \rho^2\mathbf{v} \end{aligned} \tag{58}$$

Spatial and temporal modes can be obtained via SVD decomposition!

Properties

- Conventionally (positive) singular values are ordered by decreasing value.
- A given mode j contributes to the explained variance as the squared norm of $\mathbf{u_j}\rho_j\mathbf{v_j}^\star$, ie: ρ_j^2.
- The total variance in the data (squared norm of $\mathbf{X}$) equals the sum of all squared singular values:

$$\mathrm{trace}(\mathbf{X}\mathbf{X}^\star) = \mathrm{trace}(\mathbf{X}^\star\mathbf{X}) = ||\mathbf{X}||_2^2 = \sum_{k=1}^{q} \rho_k^2, \tag{59}$$

is a measure of the total variance (also sometimes abusively called energy) in the system. The ratio $f_k = \rho_k^2 / \sum_{k=1}^{q} \rho_k^2$ is thus a measure of the variance contained in mode k compared to the overall energy and one often says that mode k explains $100\, f_k\,\%$ of the variance and that the first K modes explain $100 \sum_{k=1}^{K} \rho_k^2 / \sum_{k=1}^{q} \rho_k^2\,\%$ of the total variance.

Properties

- The first p modes define thus the best base using p base vectors in which the data can be expressed with minimum loss of information.
- Truncation can be used to filter data by rejecting modes.
- First modes often have physical meaning, the following less (due to orthogonality constraint).
- EOF very efficient for standing patterns, less for propagating features (try to generate synthetic data and look at SVD decomposition).
- Space and time can be interchanged.
- Calculate eigenvalues on smallest covariance matrix, eigenvalues are the same !
- No explicit information on "distance" or "time", only covariances are relevant.
- Data can be reordered without changing the EOFs.
- 2D spatial data can be packed into 1D arrays.

Truncation

$$\mathbf{X_N} = \mathbf{U_N}\mathbf{D_N}\mathbf{V_N}^\star = \sum_{k=1}^{N} \rho_k \mathbf{u}_k \mathbf{v}_k^\star \quad (60)$$

is the best approximation to $\mathbf{X}$ using only N spatial modes
The covariance matrix $\mathbf{X}\mathbf{X}^\star$ can also be approximated by using the truncated representation of $\mathbf{X}$

$$\mathbf{X_N}\mathbf{X_N}^\star = \mathbf{U_N}\mathbf{D_N}\mathbf{V_N}^\star\mathbf{V_N}\mathbf{D_N}^\star\mathbf{U_N}^\star = \mathbf{U_N}\mathbf{D_N}\mathbf{D_N}^\star\mathbf{U_N}^\star = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^\star \quad (61)$$

where $\tilde{\mathbf{U}}$ is matrix $\mathbf{U_N}$ where each column is multiplied by the associated singular value.
This reduced rank covariance matrix is specially useful in Data Assimilation or Optimal Interpolation

> Truncation rejects some data as noise, hence SVD can be used as filtering tool

Generalisations

- Values of the state vector can accommodate complex values, allowing complex EOF on 2D vectors (horizontal velocities, gradients ...), Fourier transforms, Hilbert transforms, etc
- Multivariate EOFs $\mathbf{x}^\star = \left(\mathbf{T}^\star, \mathbf{S}^\star, \mathbf{p}^\star, \ldots\right)$ (needs proper non dimensional form)
- SSA, lagged covariances to detect autocorrelations in time

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_n \\ x_2 & x_3 & x_4 & \ldots & x_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_k & \ldots & & \ldots & x_{k-1} \end{pmatrix}, \tag{62}$$

- MSSA, temporally lagged spatial fields

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \ldots & \mathbf{x}_n \\ \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \ldots & \mathbf{x}_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathbf{x}_k & \ldots & & \ldots. & \mathbf{x}_{k-1} \end{pmatrix}, \tag{63}$$

Applications

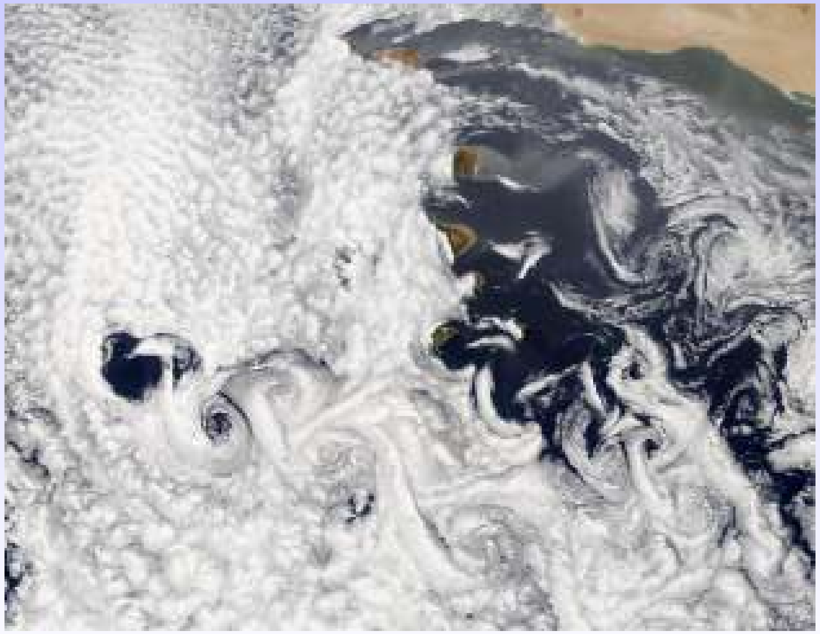
- Analysis of variability
- Data compression
- Filtering
- Data synthesizing
- Model intercomparison (time-space error distinctions, phase-amplitude error distinctions)
- Objective analysis (or optimal interpolation) of in-situ data: use of vertical EOFs and horizontal objective analysis of their amplitudes to reduces the number of data during the costly optimal interpolation
- Reducing the size of a problem by projecting (complex) reference model equations onto a small number of Principal Components
- Simplify forecasts by using any extrapolation method (neural networks, genetic algorithms etc) on temporal modes
- Ensemble preparation for assimilations
- Calculation of covariances (remember OI ?):

Université de Liège

The Problem

EOFs and friends are powerful tools to analyze data and reduce the dimensionality of a problem. However, the decomposition assumes the data matrix **X** (or continuous function f) to be known.

Missing data must be imputed before performing EOF analysis



Data INterpolating EOF: DINEOF interpolation

- Large scale EOFs should not be influenced by local changes in the values of a few points
- Large scale EOFs can thus be estimated by using a first guess of missing data
- Then, once the larger scale EOFs and their amplitudes are estimated, they can serve to calculate the value of the field at the missing points by

$$(\mathbf{X}_a)_{ij} = \left(\mathbf{U_N D_N V_N}^\star\right)_{ij} = \sum_{k=1}^{N} \rho_k (\mathbf{u}_k)_i \left(\mathbf{v}_k^\star\right)_j, \tag{64}$$

- EOFs themselves can be re-evaluated and the process can be repeated until convergence.
- What is the optimal number N of relevant EOFs to be retained to recompose the signal at the missing data points ?

Optimal number N of EOFs: Cross validation technique

Université de Liège

- Set aside a random set of valid data (random points or random clouds).
- Use the EOF interpolation and calculate an error estimate based on the rms distance between the interpolated field at these points and the data set aside there
- Start with 1 EOF, fill in, calculate rms error, continue by filling in with a second EOF until convergence, calculate rms error ...
- Provides reconstruction error as a function of number of EOF retained.
- The optimal truncation is the one that minimises the diagnosed error.

Once the optimal number know, perform a last iteration with this number of EOFs reinjecting the data set aside for cross-validation.

DINEOF

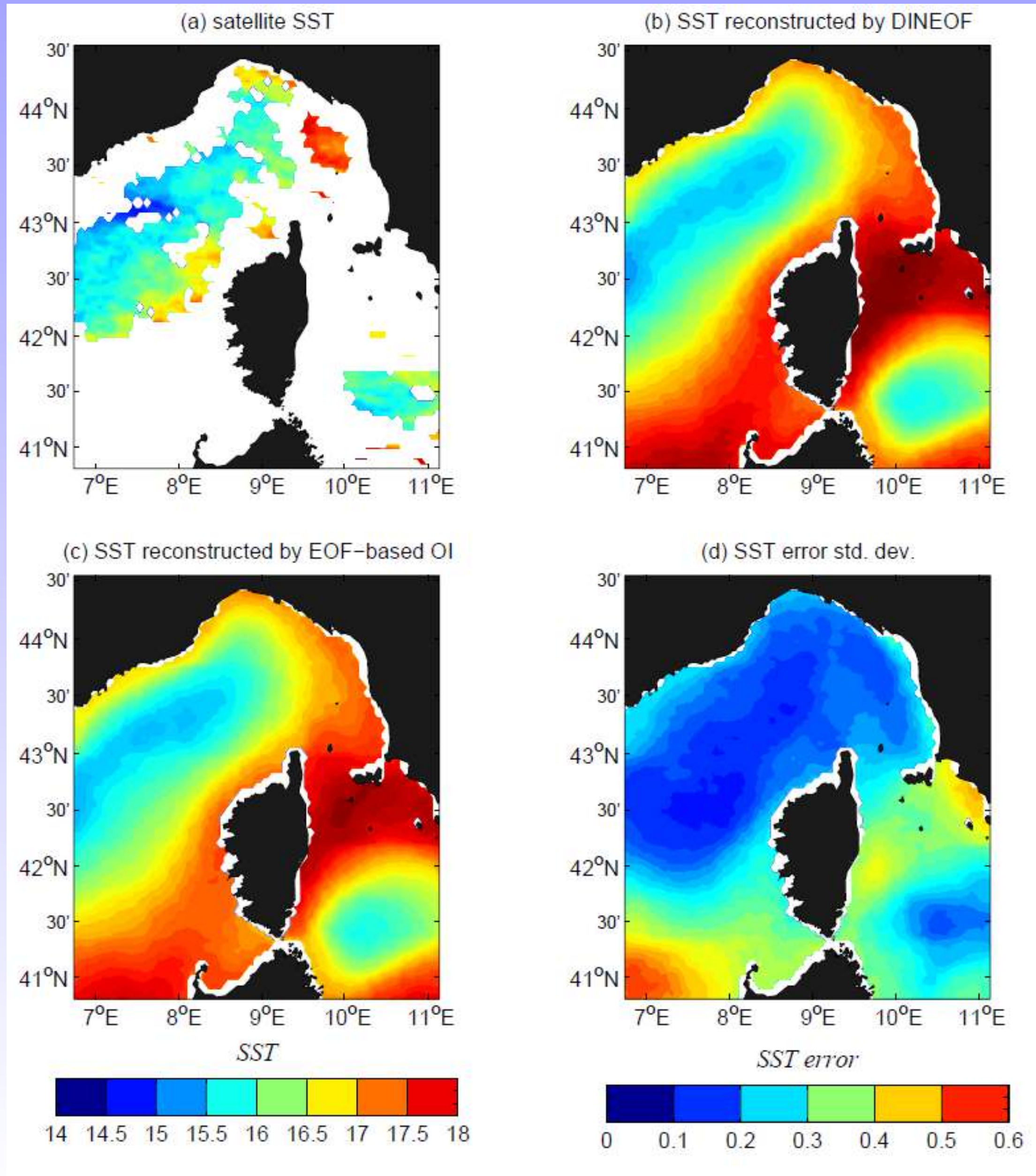
- Implementation much more rapid than OI version.
- Problem: no natural error maps in DINEOFs

Error maps:

- Use error estimate from OI version (35)
- Numerical verification that difference OI-DINEOF is smaller than this error
- Can also be exploited to detect strange pixels in original data (outlier detection using spatial covariances)

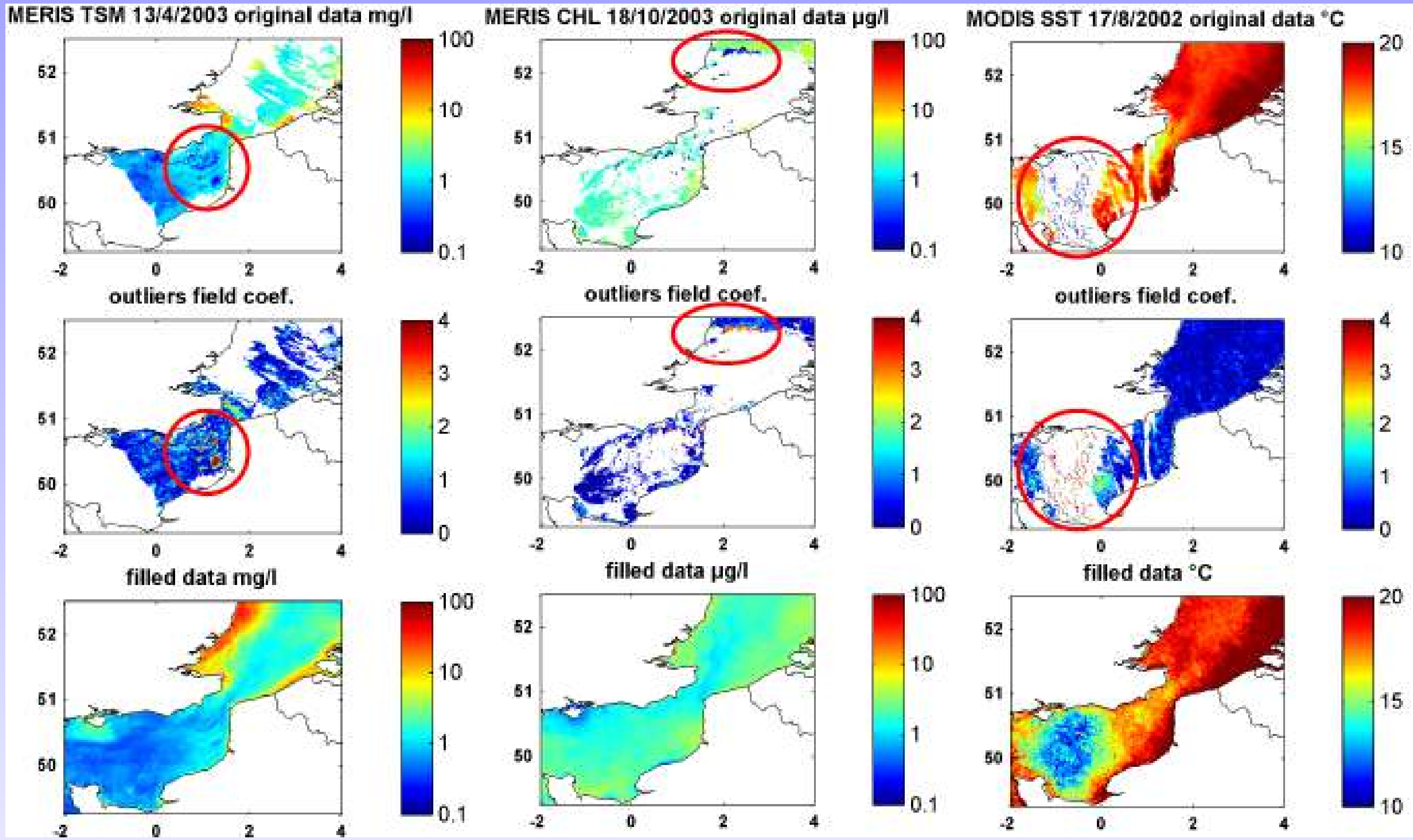
CANNOT see things under clouds which are due to patterns never seen before (EOF are exploiting past pattern)

Error maps



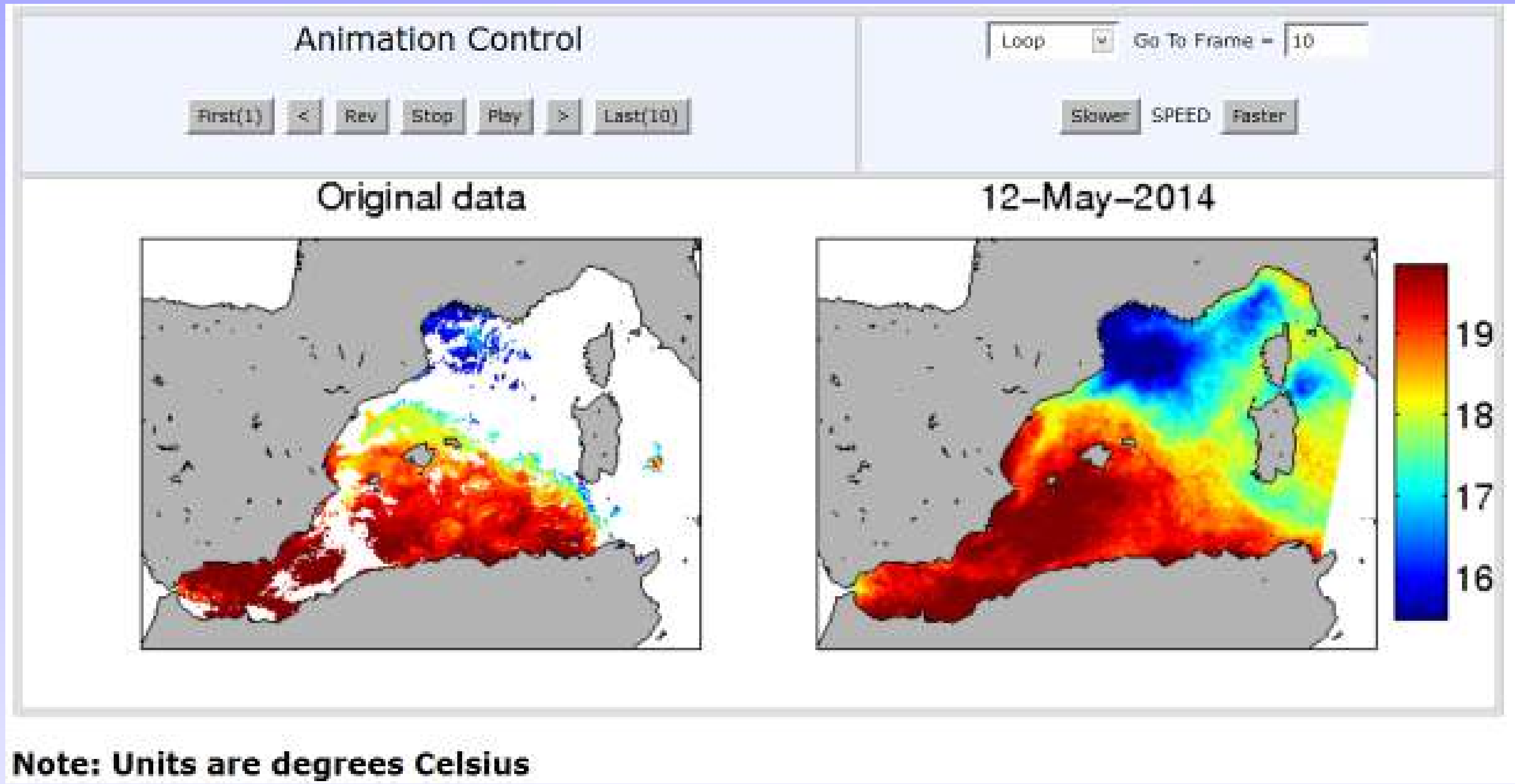
Université de Liège

Outlier detection



Operational use example

Daily 10 day period based on last six month EOFs



`http://gher-diva.phys.ulg.ac.be/DINEOF/dineof.html`

Université de Liège

- *Gridding*
- *DIVA theory*
- *Implementations and exercises*
- *DINEOF theory*
- ***Implementations and exercises***

On line version: beta version

http://gher-dineof01.phys.ulg.ac.be:8081/

Make it work

"Parameter free" method, yet some things to decide

- Time and space window for data: neither too small nor too large: Cover the scale you are interested in with several repetitions in time (EOFs!)
- Subtraction of mean field (remember background in OI?): The better your first guess, the better the result
- Number of EOFs to retain a priori for Krylov space before cross validation: small value reduces CPU time but should be larger than final number of EOFs retained
- Reinforce temporal coherence (EOFs do not know explicitly about "distance"): avoid spikes
- How to validate: incorporated cross-validation, additional cross-validation (add clouds on your side), validation with independent (in-situ) data
- Keep original data in unclouded regions or analyzed data: Most coherent approach is to use analyzed field everywhere

Université de Liège

Even more decisions

- Number of iterations: neither too few nor too many
- Criteria for convergence: neither too severe nor too relaxed
- Elimination of images or pixels with too few data: threshold typically ~ 3%
- Normalisation in case of multivariate approaches
- Exploitation of delayed information (*eg.* wind a few hours before TSM field)

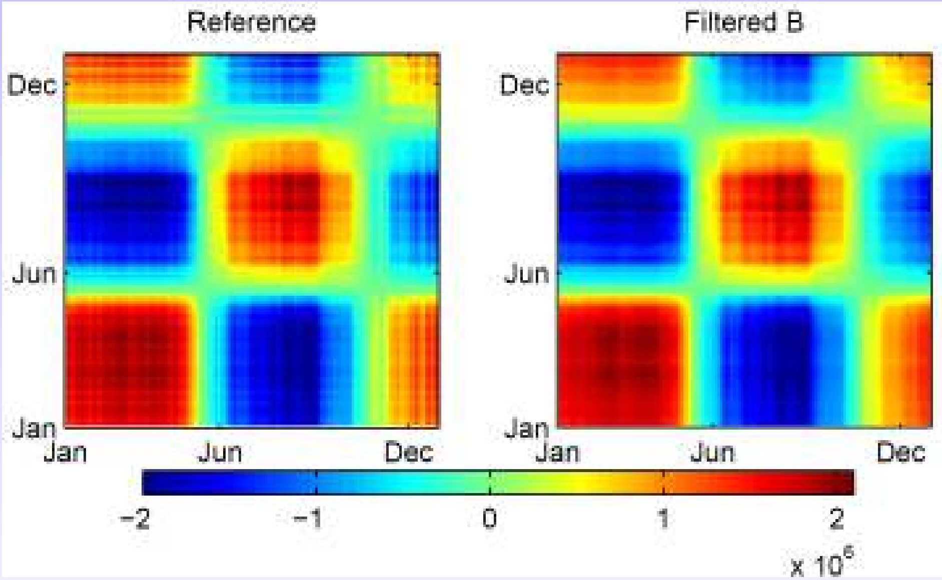
? !

Temporal coherence

Application of time-filtering matrix **F** to data **X** would lead to filtered data $\tilde{\mathbf{X}}$

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{F} \tag{65}$$

Incorporation into the DINEOF iterations?
Instead of working with covariance $\mathbf{X}^{\star}\mathbf{X}$ work with filtered covariance matrix (EOFs are the eigenvectors) $\mathbf{F}^{\star}\mathbf{X}^{\star}\mathbf{X}\mathbf{F}$.



Construction of filtering operator F

How to construct filter matrix (operator) **F**: by mimicking effect of diffusion in space

$$\frac{\partial c}{\partial t} = \kappa \frac{\partial^2 c}{\partial x^2} \tag{66}$$

translated into time filtering (diffusion in time direction) as iterations

$$\boldsymbol{x_i}^{(s+1)} = \boldsymbol{x_i}^{(s)} + \frac{\boldsymbol{G_{i+1}} - \boldsymbol{G_i}}{t'_{i+1} - t'_i} \tag{67}$$

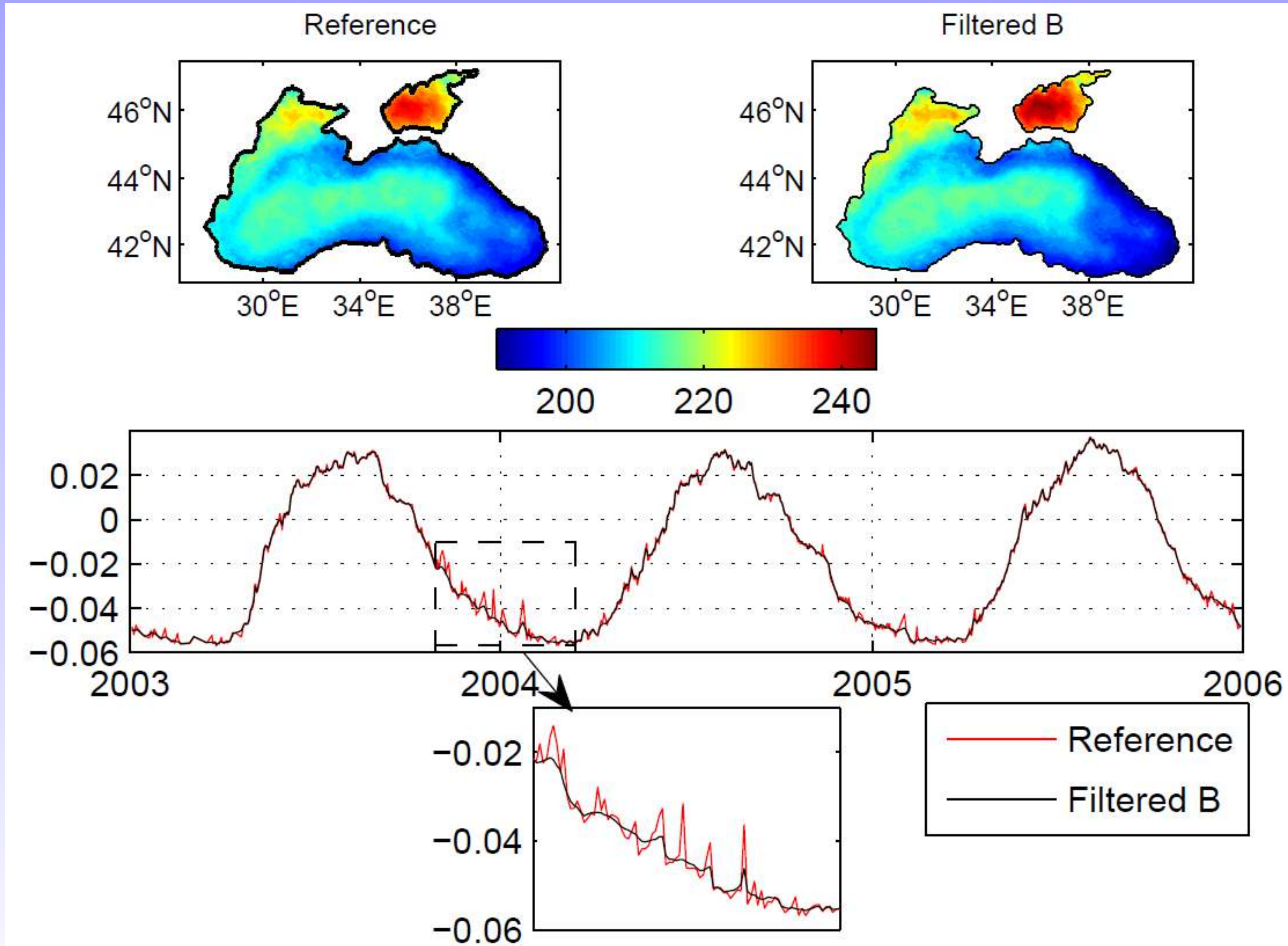
$$\boldsymbol{G_{i+1}} = \alpha \frac{\boldsymbol{x_{i+1}}^{(s)} - \boldsymbol{x_i}^{(s)}}{t_{i+1} - t_i} \tag{68}$$

$$t'_i = \frac{t_i + t_{i-1}}{2} \tag{69}$$

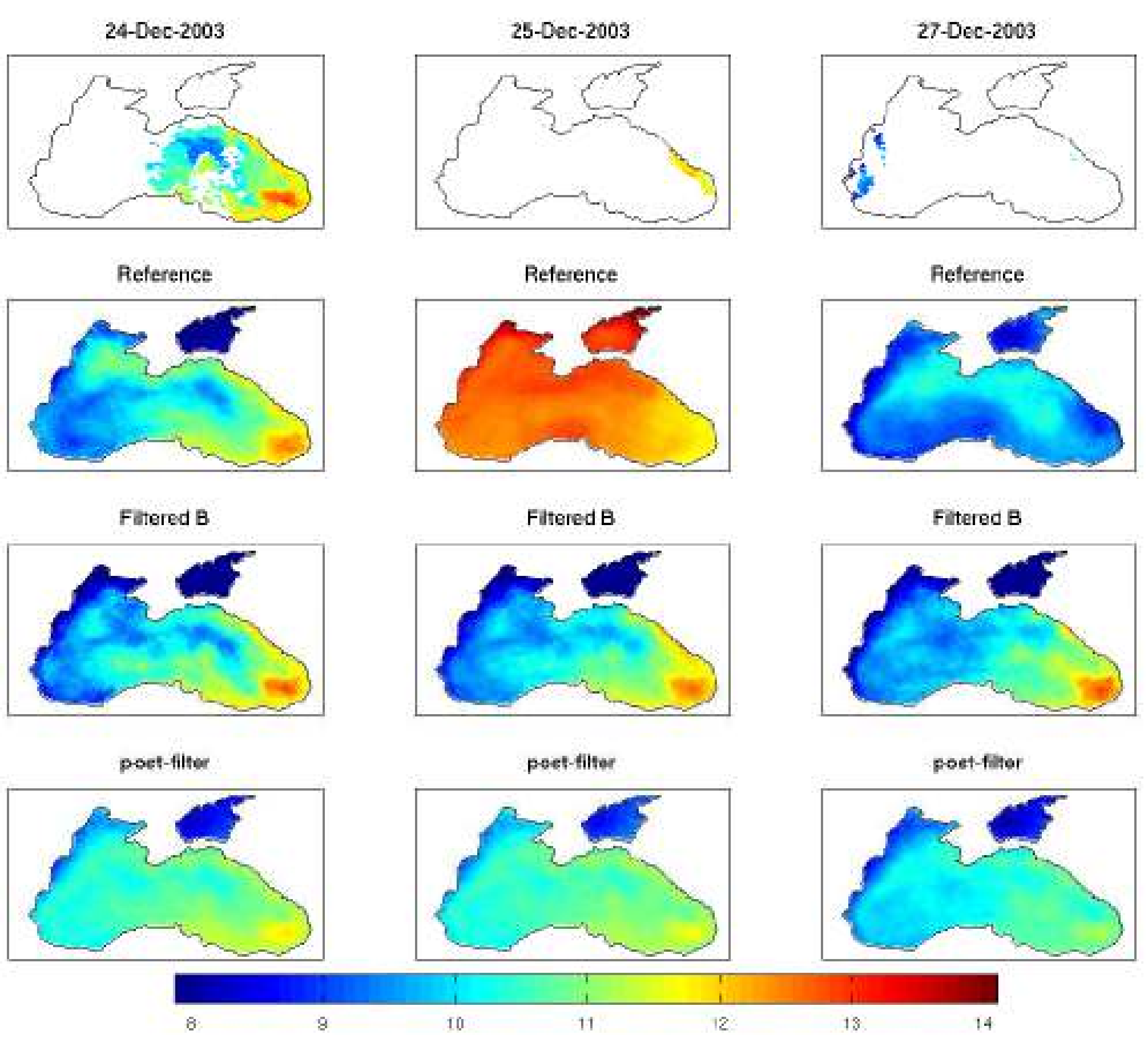
with iterations $s = 1, 2, 3..., p$ and strength α of filter defining filter caracteristics (*Alvera et al 2009*)

Université de Liège

Effect of filtering

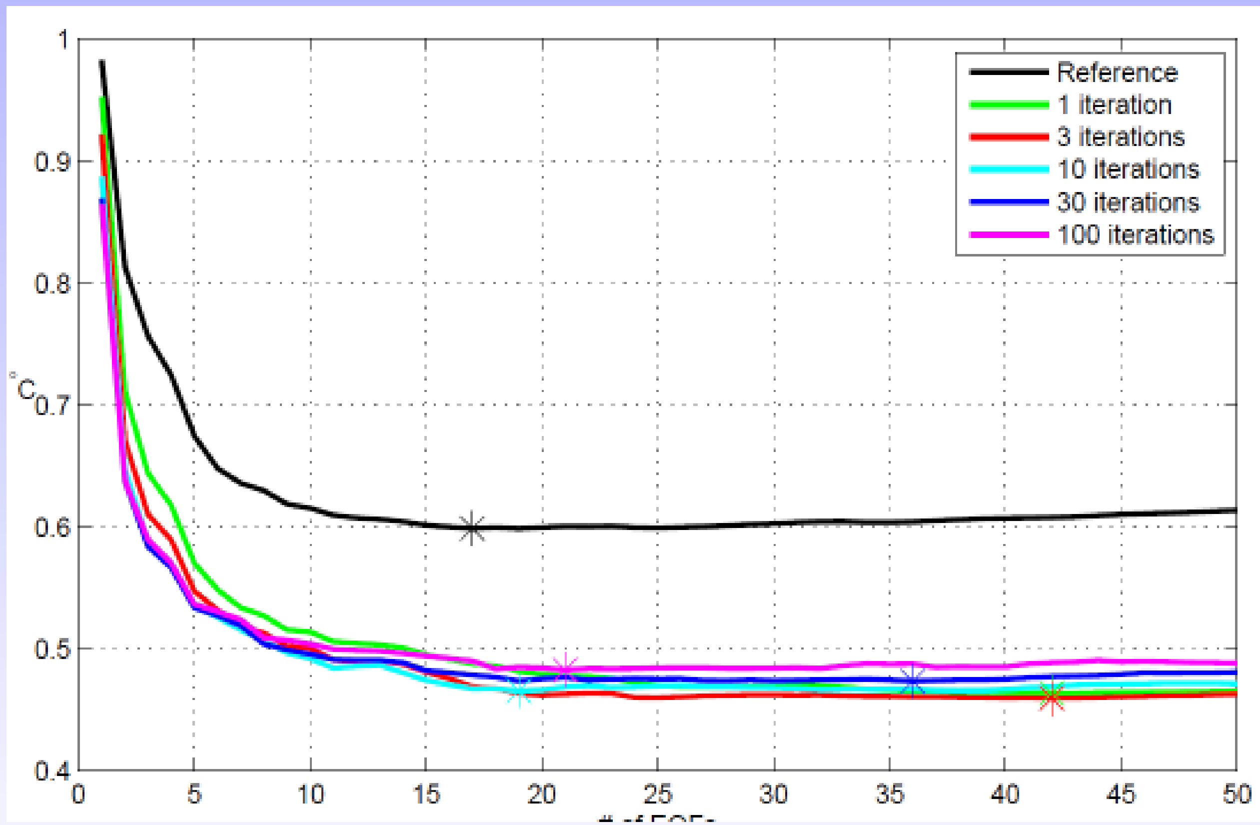


Effect of filtering



How to decide on values for filtering

- Your decision to filter out processes shorter than T: $2\pi\sqrt{\alpha p} \sim T$ and α such that filtering iterations remain stable $\alpha < \min(\Delta t)^2/2$
- By cross validation again !

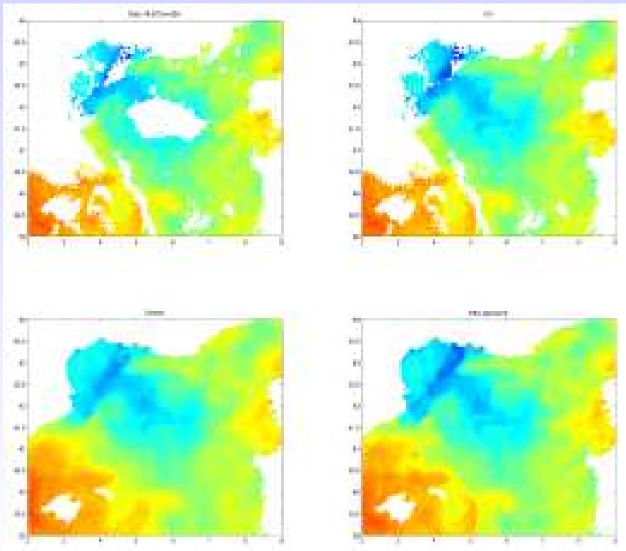


Command line version

`http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF`

- DINEOF Fortran code with a lot of matlab/octave tools around.
- Next version with better netCDF support
- Google usergroup
  `http://groups.google.com/group/dineof`
- SVN for contributors

In next version: possibility to add small scale OI on top of DINEOF interpolation



A. Alvera-Azcárate, D. Sirjacobs, A. Barth, and J.-M. Beckers. Outlier detection in satellite data using spatial coherence. Remote Sensing of Environment, 119:84-91, 2012.

D. Sirjacobs, A. Alvera-Azcárate, A. Barth, G. Lacroix, Y. Park, B. Nechad, K. Ruddick, J.-M. Beckers. Cloud filling of ocean color and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology. Journal of Sea Research, 65(1):114-130. 2011.

A. Alvera-Azcárate, A. Barth, D. Sirjacobs, J.-M. Beckers. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. Ocean Science, 5, 475-485, 2009

A. Alvera-Azcárate, A. Barth, J.-M. Beckers, and R. H. Weisberg. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll and wind satellite fields. Journal of Geophysical Research, 112:C03008, 2007. doi:10.1029/2006JC003660.

J.-M. Beckers, A. Barth, and A. Alvera-Azcarate. DINEOF reconstruction of clouded images including error maps. Application to the Sea Surface Temperature around Corsican Island. Ocean Science, 2(2):183-199, 2006.

A. Alvera-Azcárate, A. Barth, M. Rixen, and J.-M. Beckers. Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea. Ocean Modelling, 9:325-346, 2005.

J.-M. Beckers and M. Rixen. EOF calculations and data filling from incomplete oceanographic data sets. Journal of Atmospheric and Oceanic Technology, 20(12):1839-1856, 2003