Université de Liège

DIVA

Overview

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

http://modb.oce.ulg.ac.be/GHER

Université de Liège
MARE-GHER Sart-Tilman B5
4000 Liège, Belgique

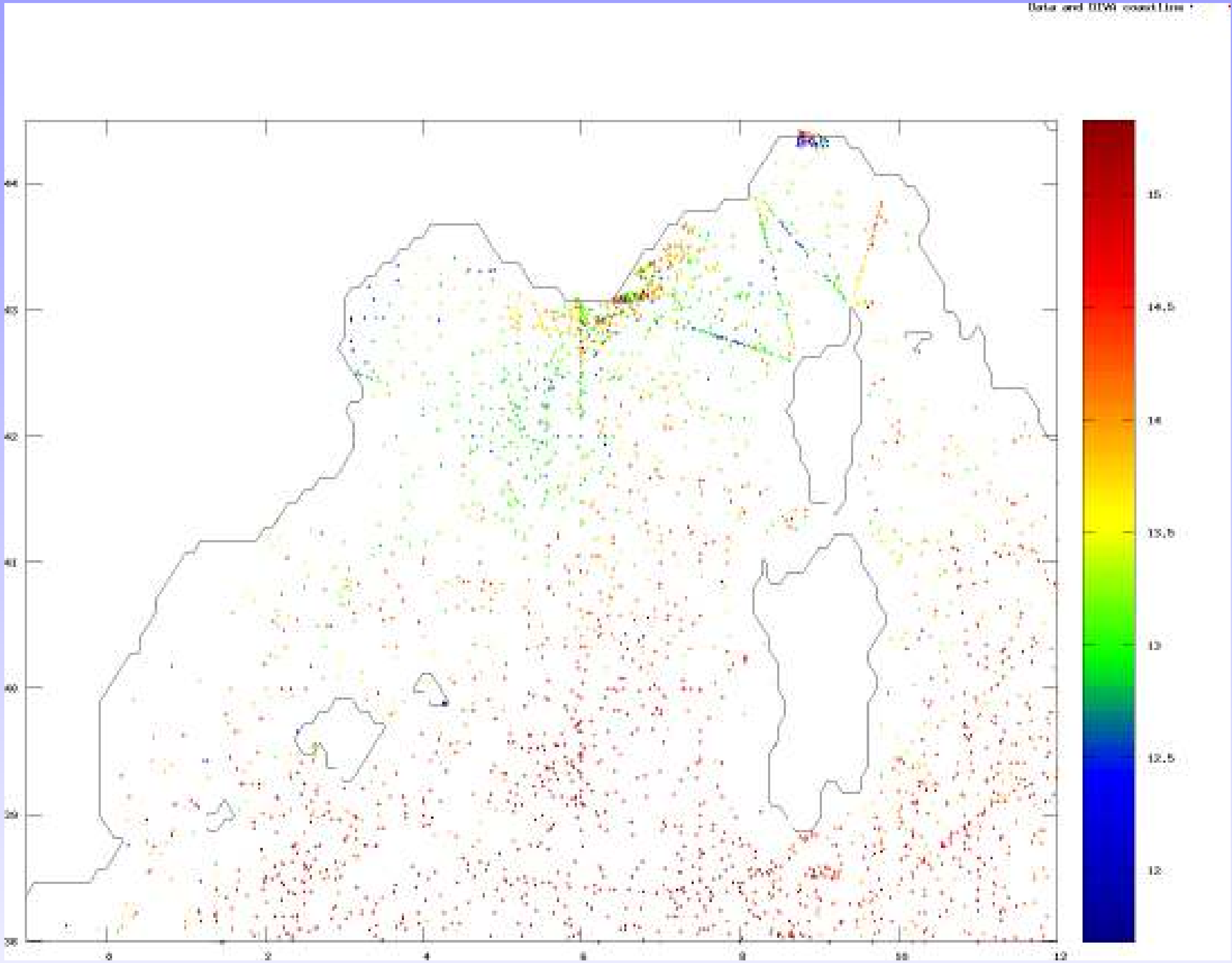
Outline

- Gridding
- Diva
- Implementations
- And for data managers ?
- Summary

Université de Liège

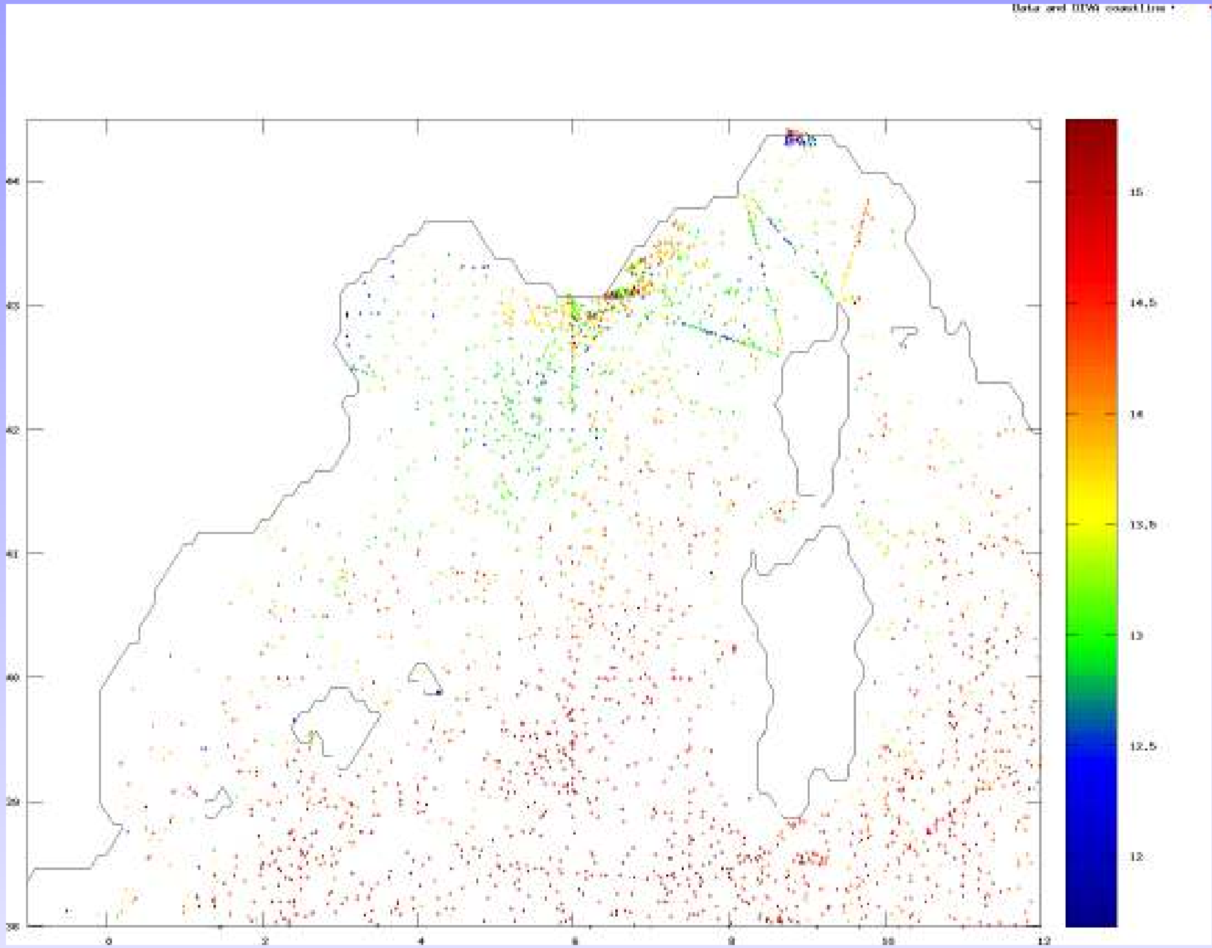
- **Gridding**
- Diva
- Implementations
- And for data managers ?
- Summary

Common problem

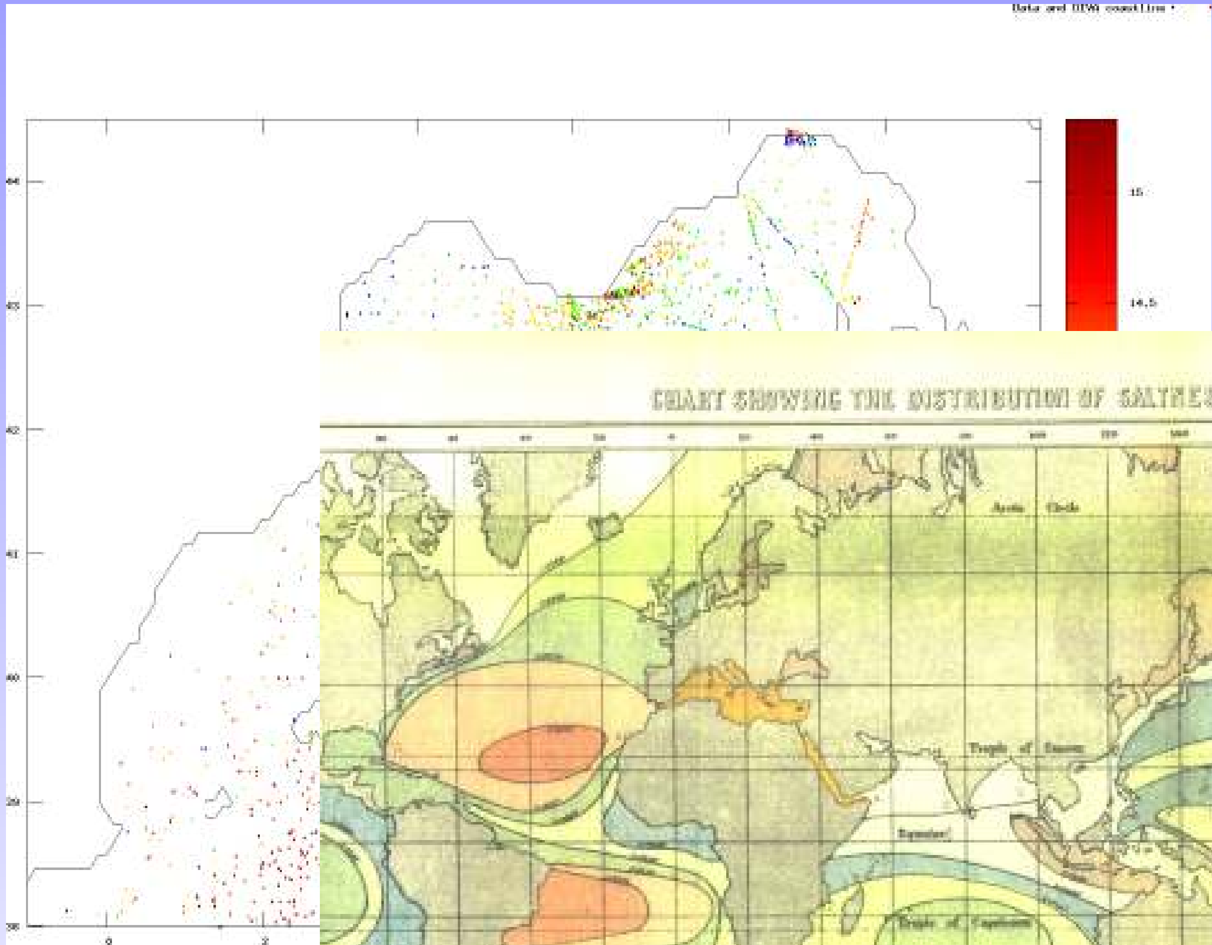


Appears when trying to produce maps, calculate volume averages, prepare initial conditions for models, quality control of data ...

Solutions



Solutions



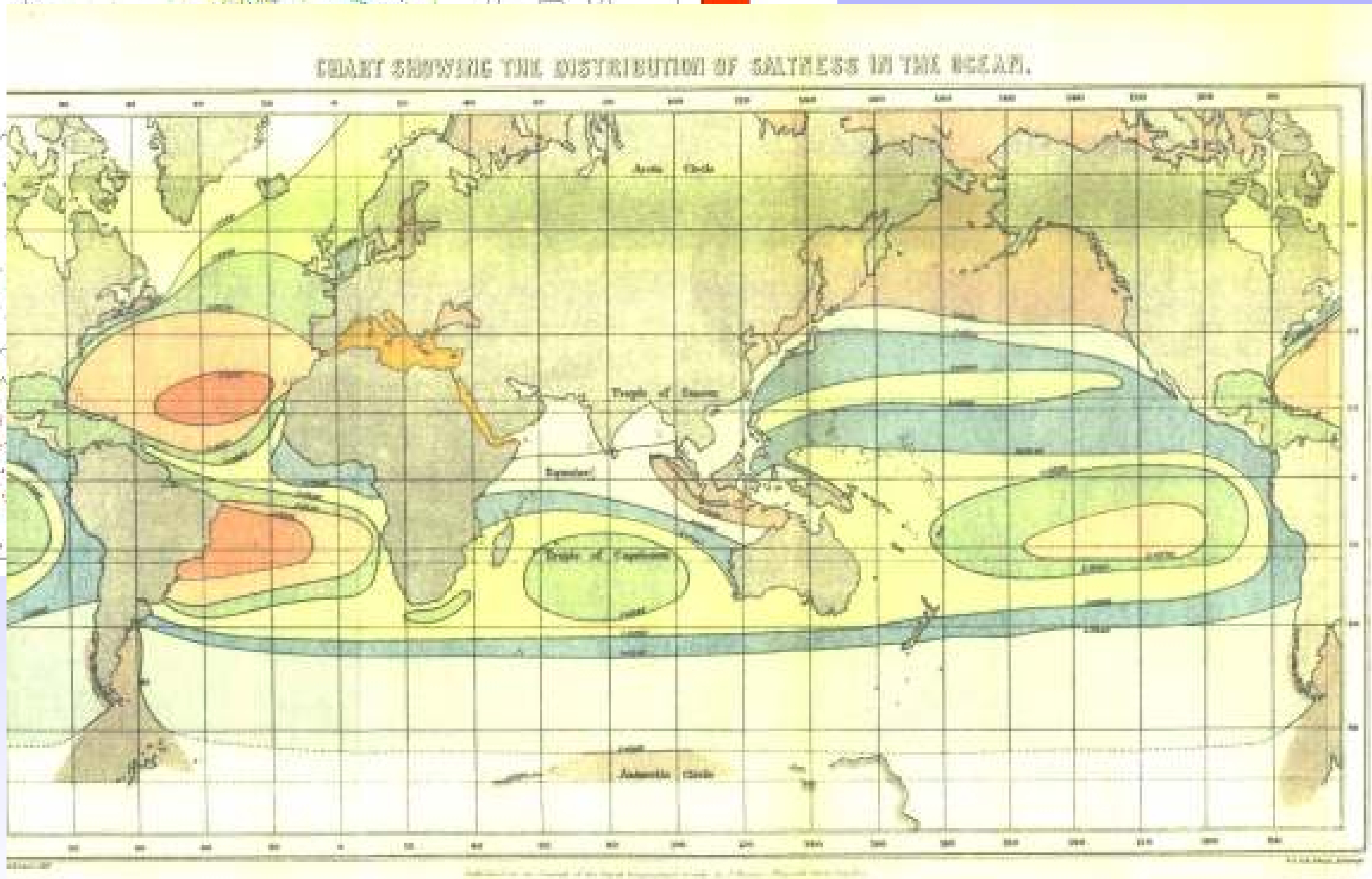
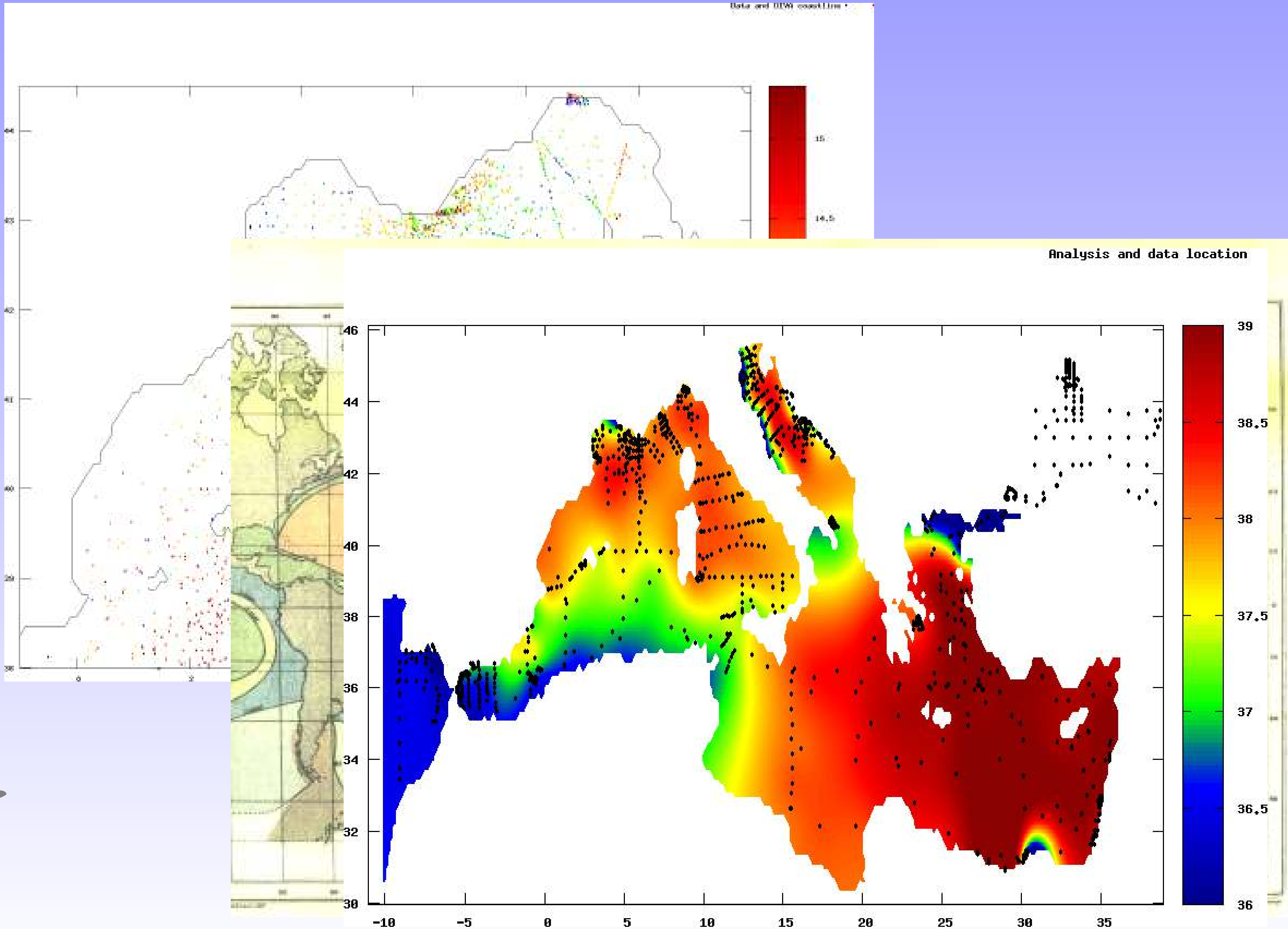
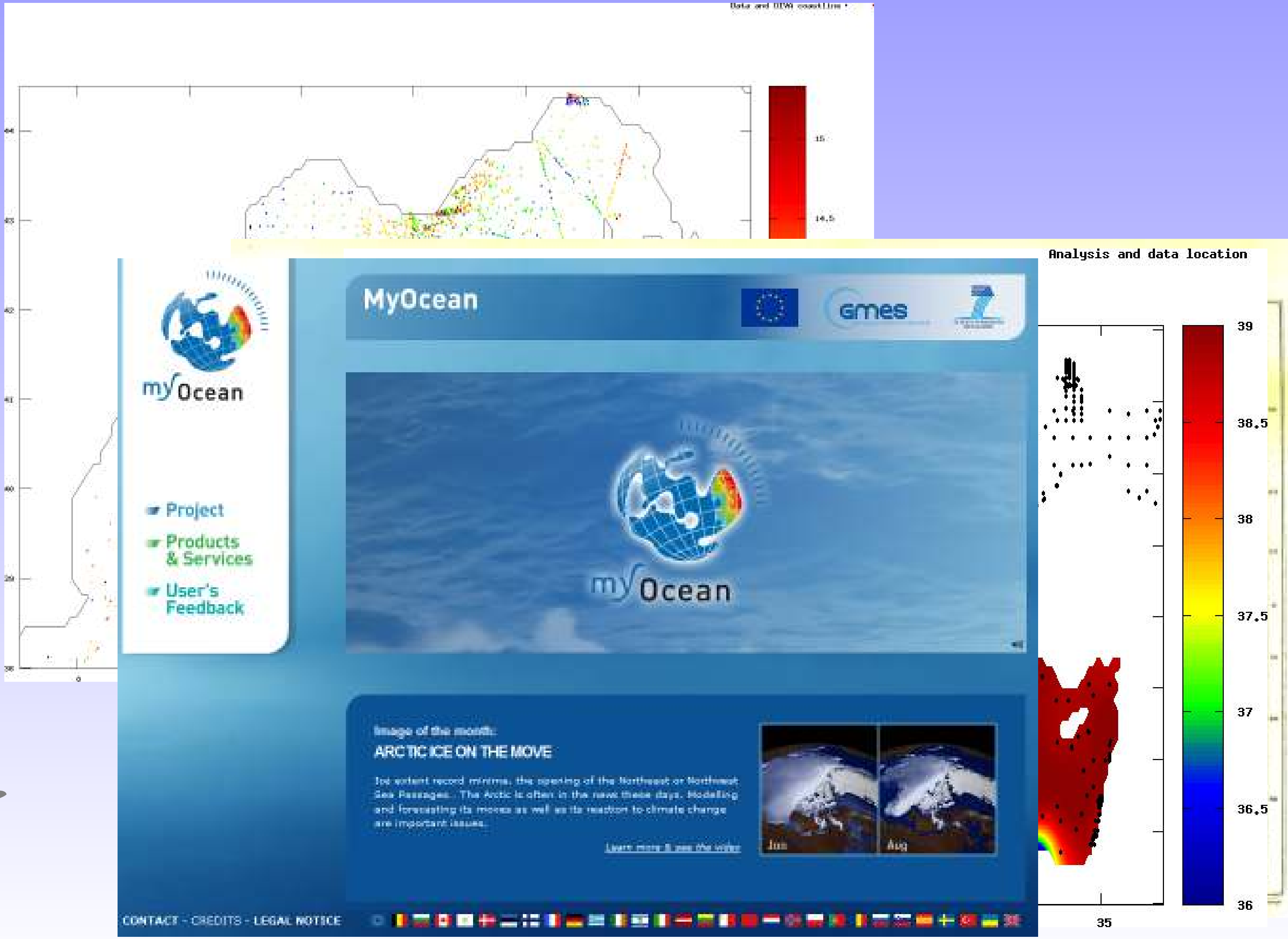


CHART SHOWING THE DISTRIBUTION OF SALTNESS IN THE OCEAN.
Arctic Circle
Tropic of Cancer
Equator
Tropic of Capricorn
Antarctic Circle

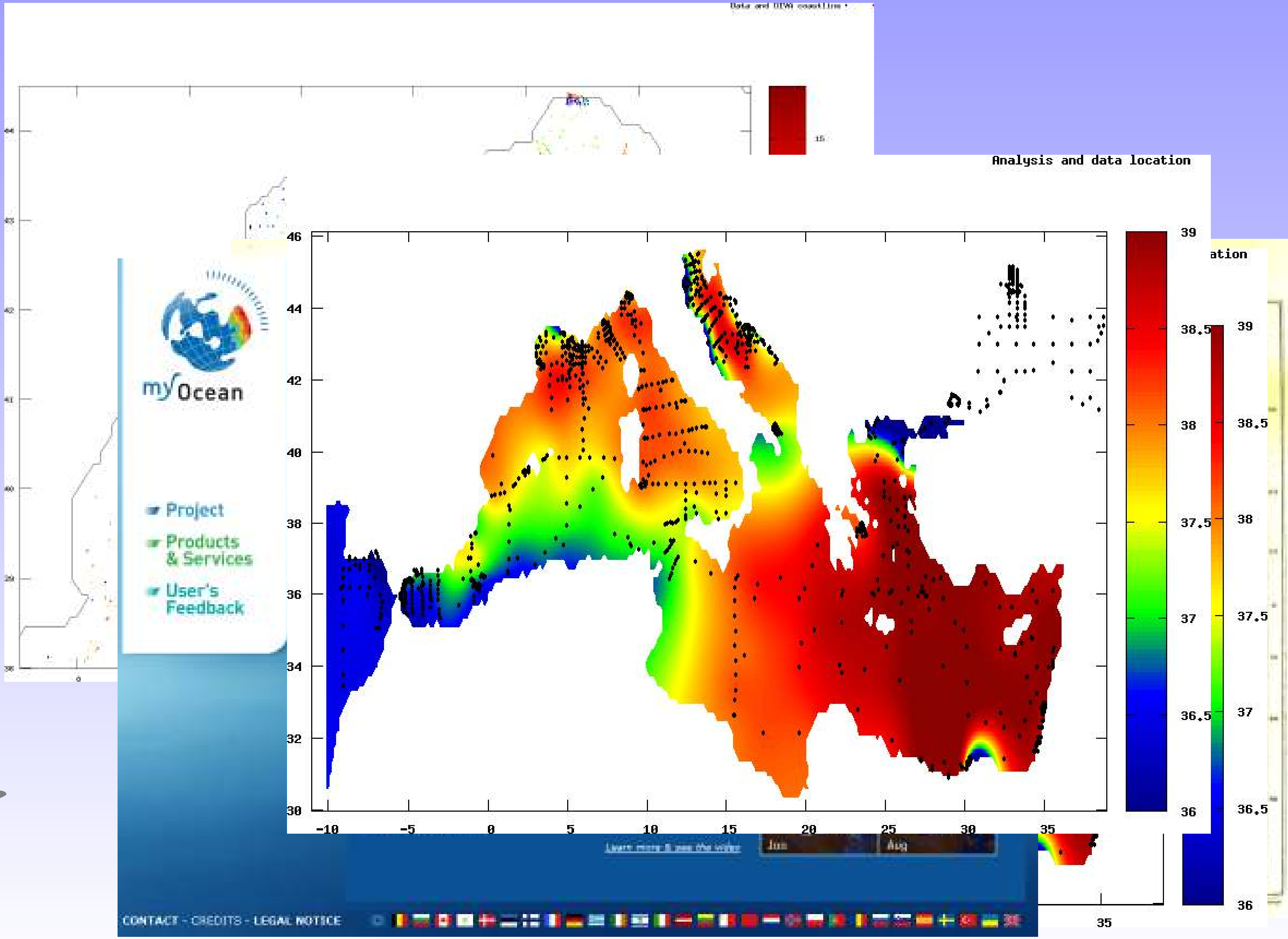
Solutions


Analysis and data location
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

Solutions


Data and DIVA coastline
MyOcean
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
Analysis and data location
39
38.5
38
37.5
37
36.5
36
35

Solutions



Analysis and data location
my Ocean
Project
Products & Services
User's Feedback
CONTACT - CREDITS - LEGAL NOTICE

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?

15°

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?
15°

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

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?
15°

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

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?
15°

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

Université de Liège

Estimation of today's temperature in Ostende

- Observer 1: 14°
- Observer 2: 16°

Your best guess ?
15°

But what if observer 1 in Sidney uses digital thermometer and observer 2 in Ostende his finger ?
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 Interpolation

- Need for covariance of the background field between data points: each element i, j of **B** provides the covariance between points in location i and j. Covariance between a given point on the analysis grid and all data points is also needed.
- Need for observational error information (not only instrumental). Equivalently signal-to-noise ratio

Analysis leads to spatial analysis and error field at any desired location if covariance between any two points is known.
Practical problems in addition to covariance specification: huge matrix inversions for large data sets.

Background covariance

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

- Normally obtained via statistics on data. Seldom possible (noticeable exception: satellite images).
- 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)

Note that real covariance are reduced by obstacles or increased along current.

Université de Liège

- *Gridding*
- ***Diva***
- *Implementations*
- *And for data managers ?*
- *Summary*

Université de Liège

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

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

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

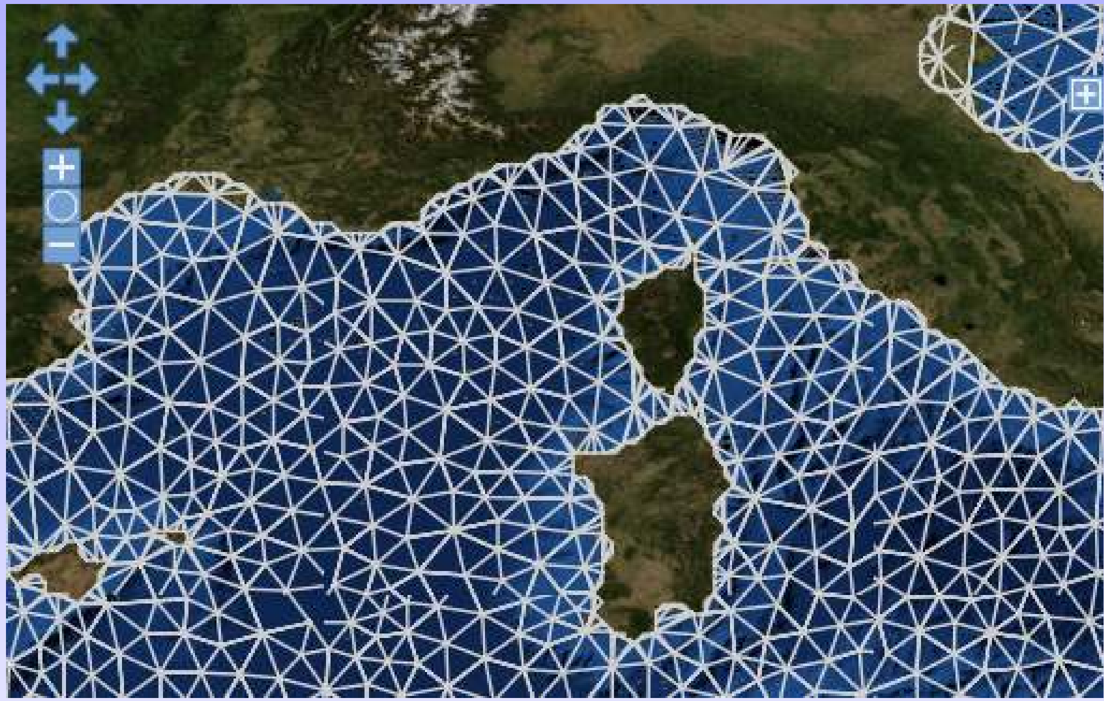
The background field φ_b is typically the data average value.

- 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

$$\mu = \frac{\sigma^2}{\epsilon^2}\frac{4\pi}{L^2} \qquad (3)$$

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



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

Bad news ☹

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

Bad news

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

- VIM brand already taken: 

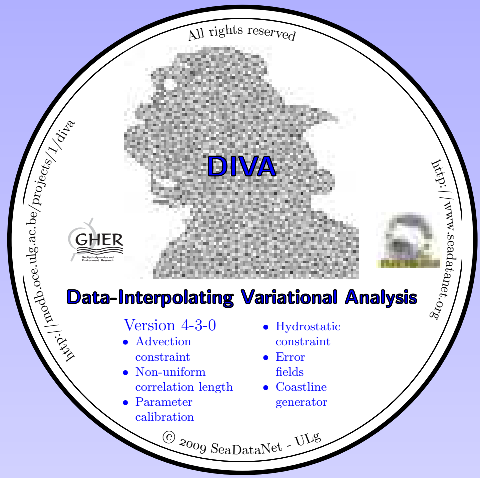
Bad news ☹

- 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 ☺

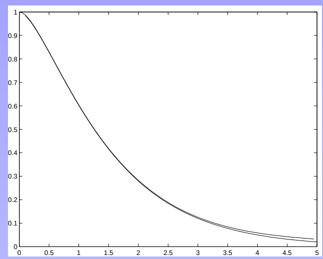
- Name easily changed: DIVA (Data-Interpolating Variational Analysis)



- 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



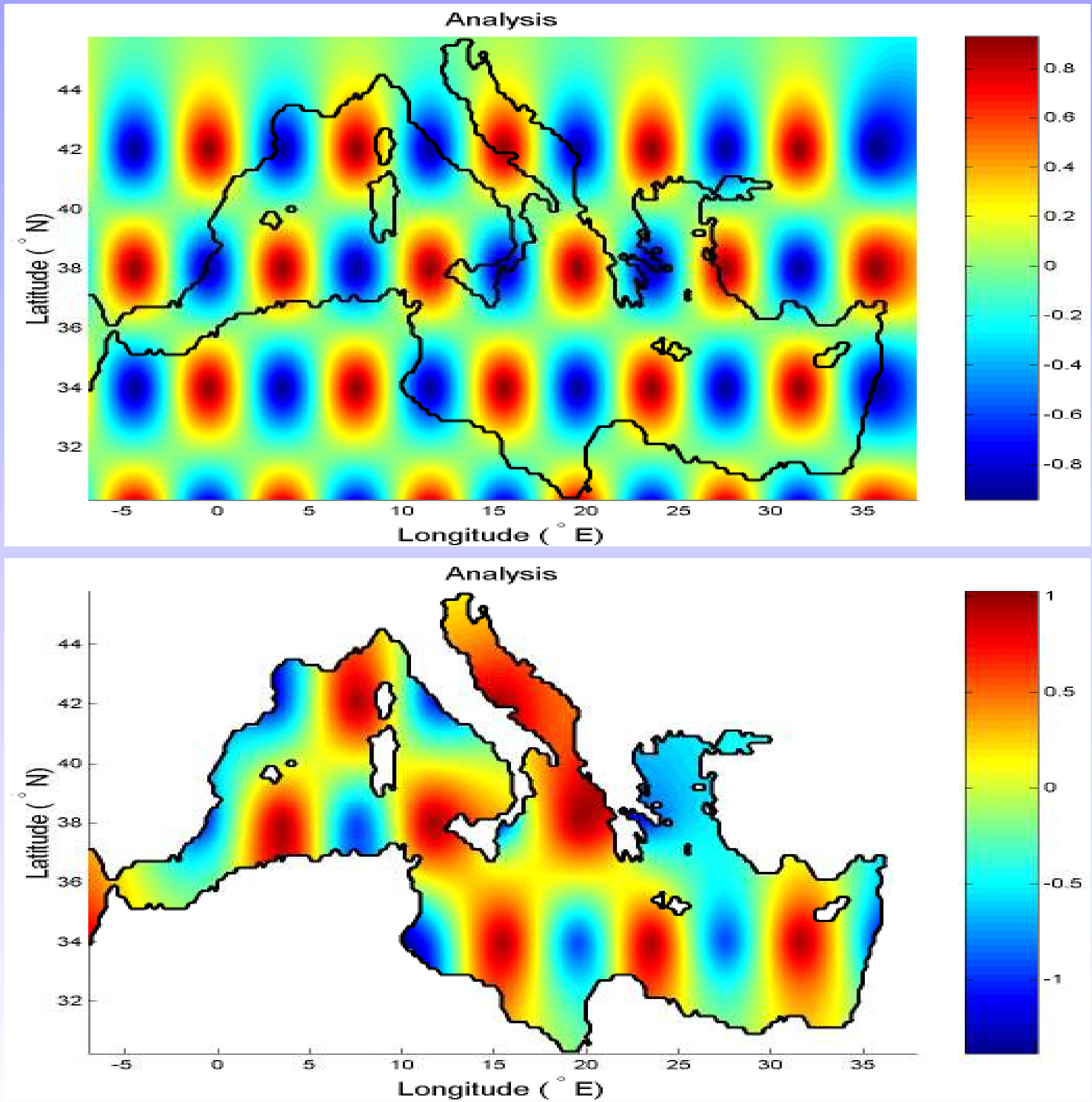
- if so-called reproducing kernel of the norm $=$ covariance function of OI,
- if the noise is random, spatially uncorrelated and the signal/noise ratio parameter is identical with OI.

In this case, the OI solution $=$ DIVA solution.

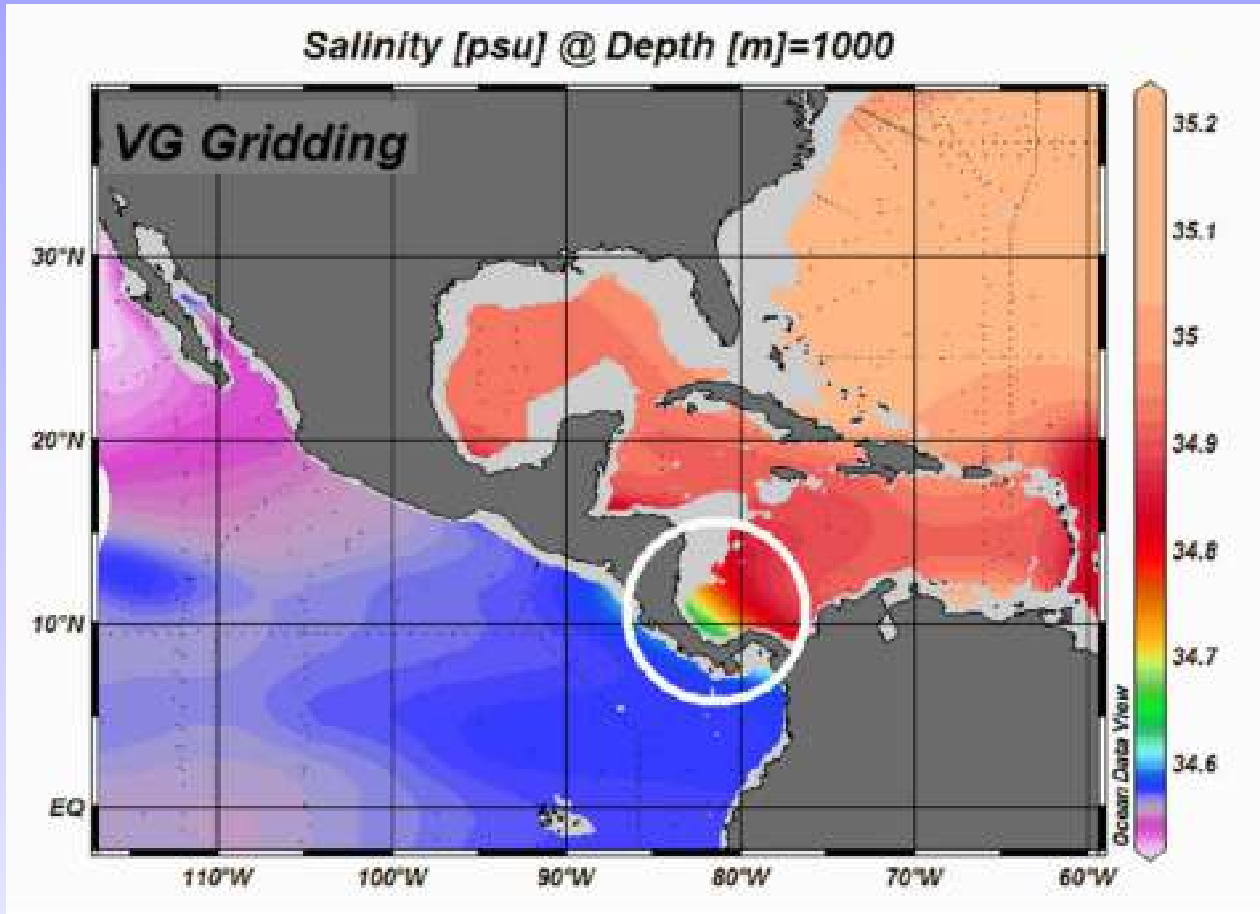
- Advantages of DIVA: regularization, fast finite-element solution, boundary effects taken into account.
- Difficulties: generalizations to 3D and multivariate versions are "hybrid".

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 minimisation.

Illustration of covariance functions

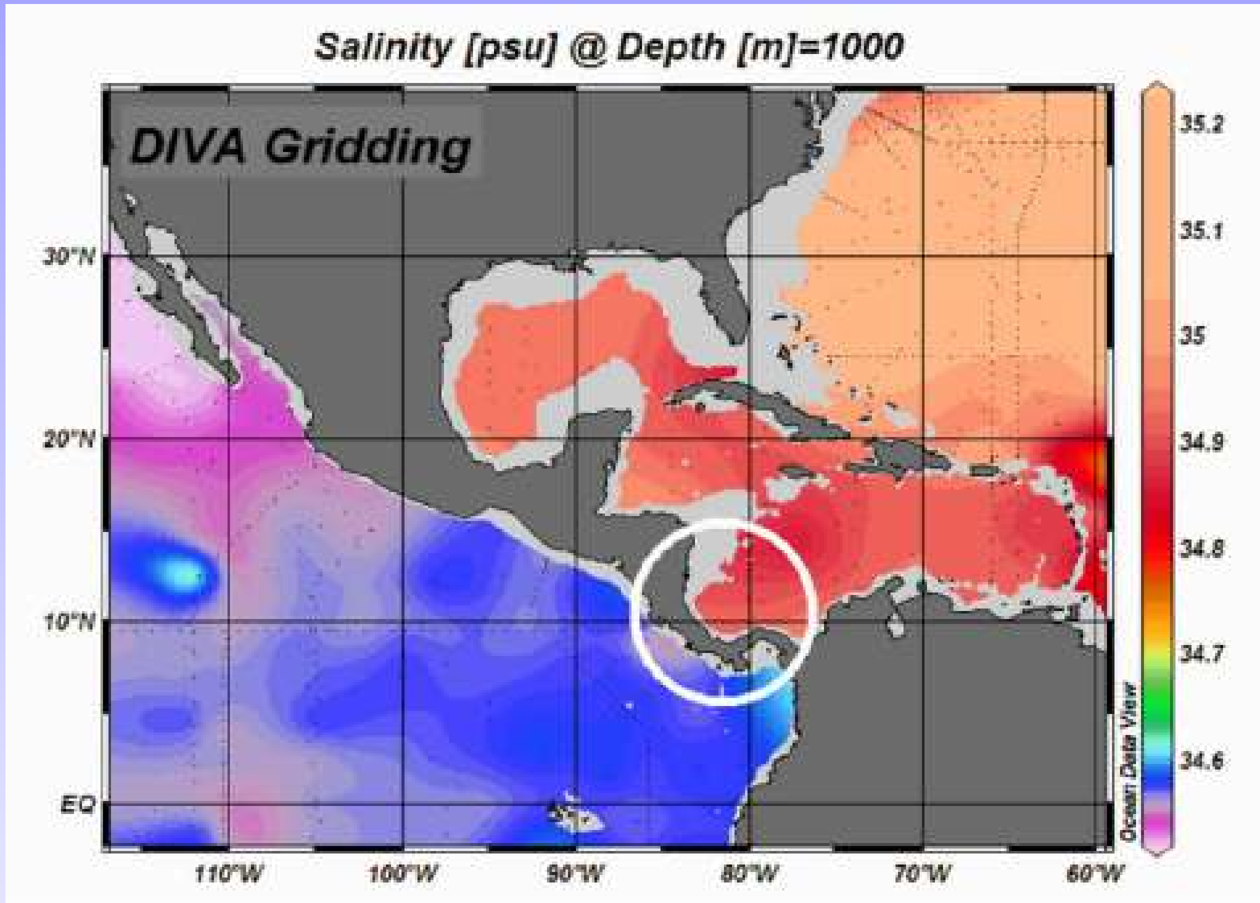


Standard ODV analysis



(R. Schlitzer ODV example)

ODV-DIVA analysis



(R. Schlitzer ODV example)

Additions to basic tool

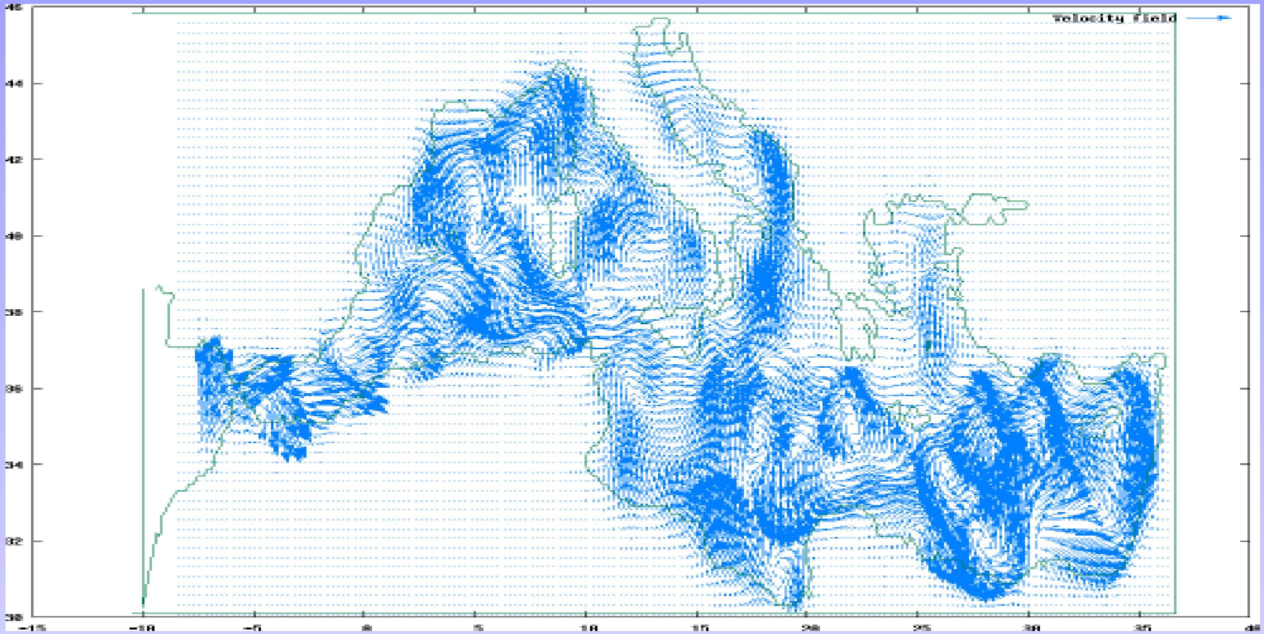
Advection constraint: Augmented cost function to deal with preferred correlation directions, eg, via advection with velocity $\boldsymbol{u}$ and diffusion $\mathcal{A}$

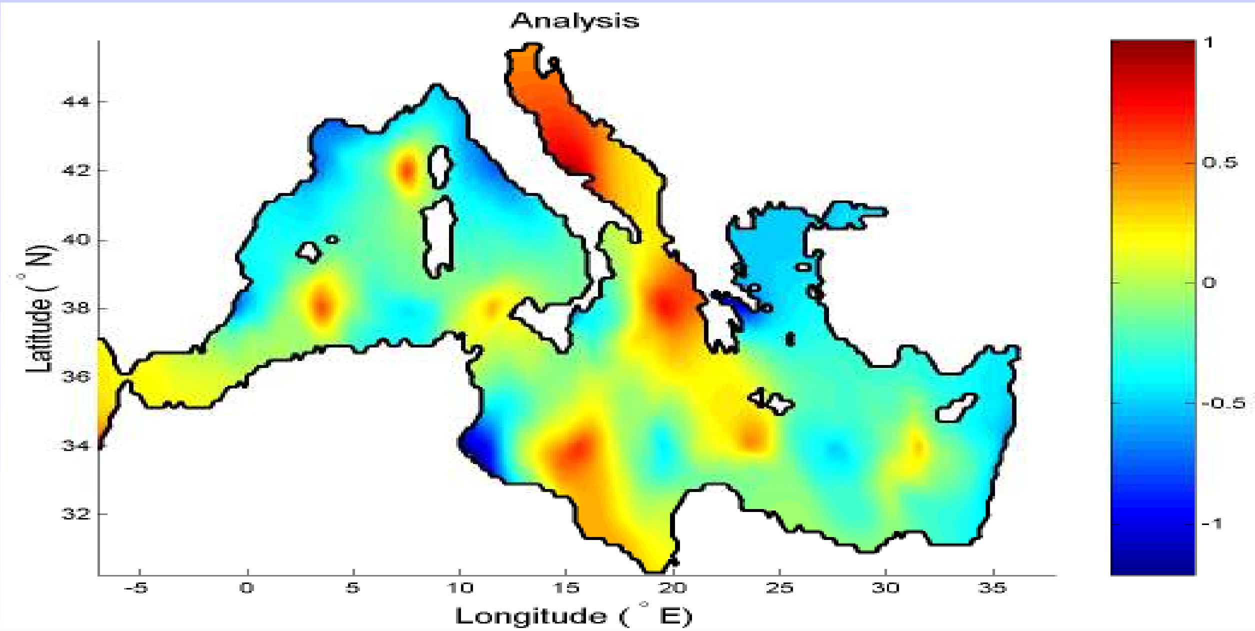
$$J_a = J(\varphi) + \theta \int_D \left[\boldsymbol{u} \cdot \nabla\varphi - \mathcal{A}\,\nabla\cdot\nabla\varphi\right]^2 \mathrm{d}D \tag{4}$$

Other features

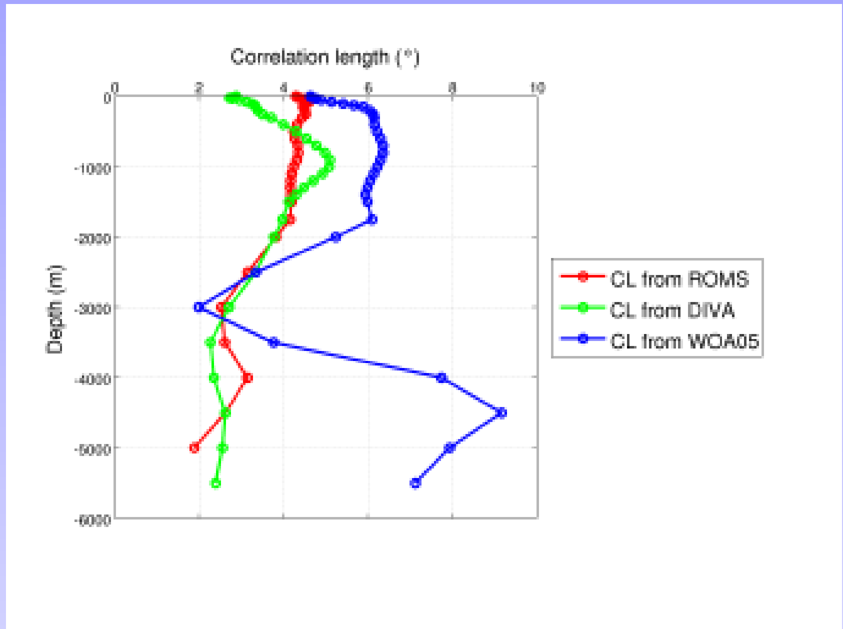
- Error fields taking data distribution into account.
- Toolbox approach allowing to design own versions.
- 3D and 4D modes by looping, hydrostatic constraint in 3D mode.
- Cross validation tools to infer statistical parameters and error estimates.
- Climatology production version with heterogeneous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

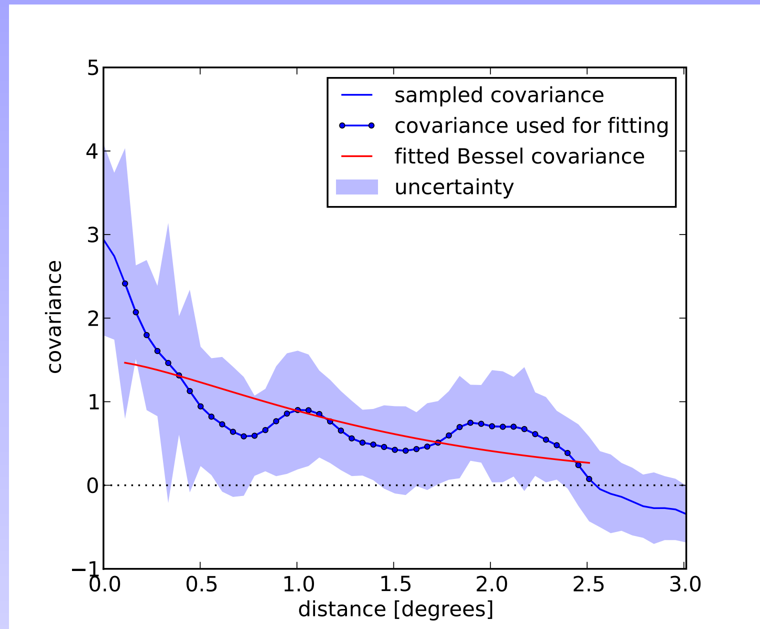
Covariances with advection





Parameter calibration

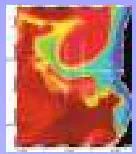




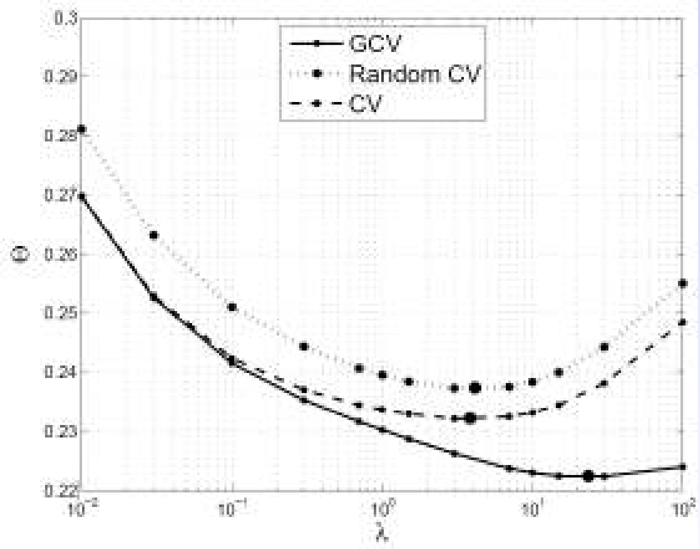
Spatial coherence of parameters: here correlation length obtained with covariance fitting (Troupin *et al.*, 2010).

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.

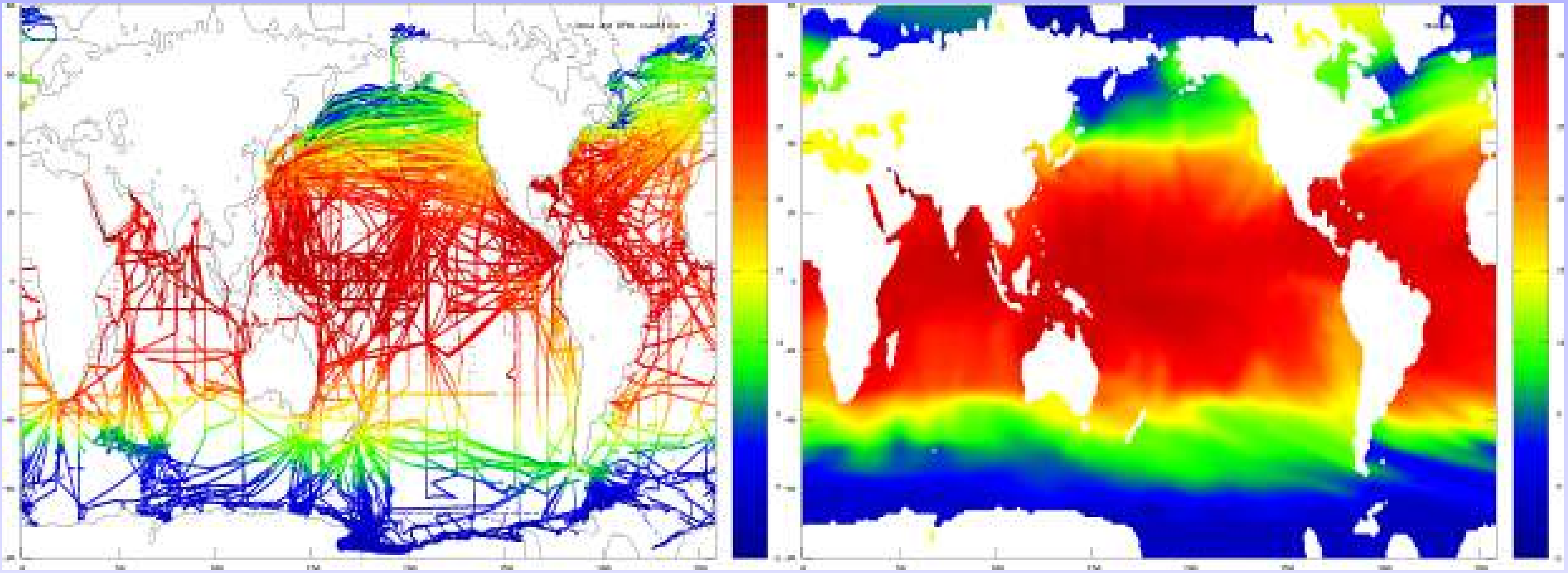


Université de Liège

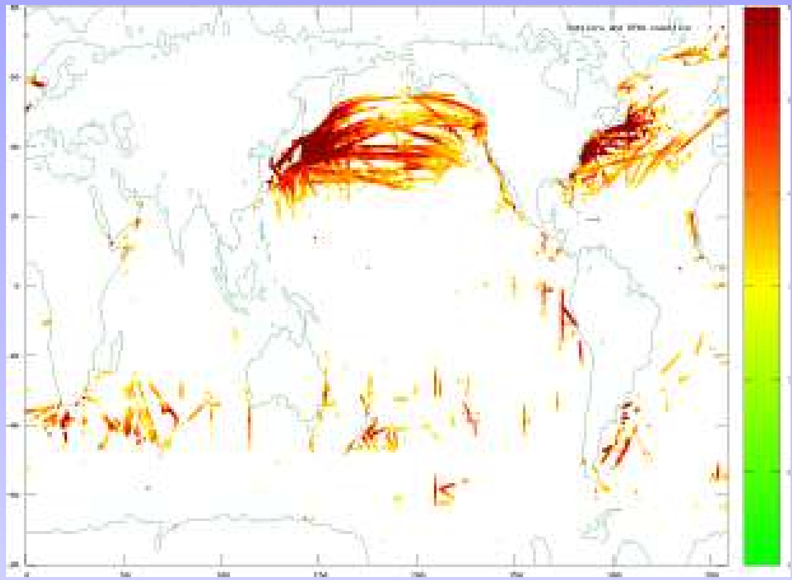
? !

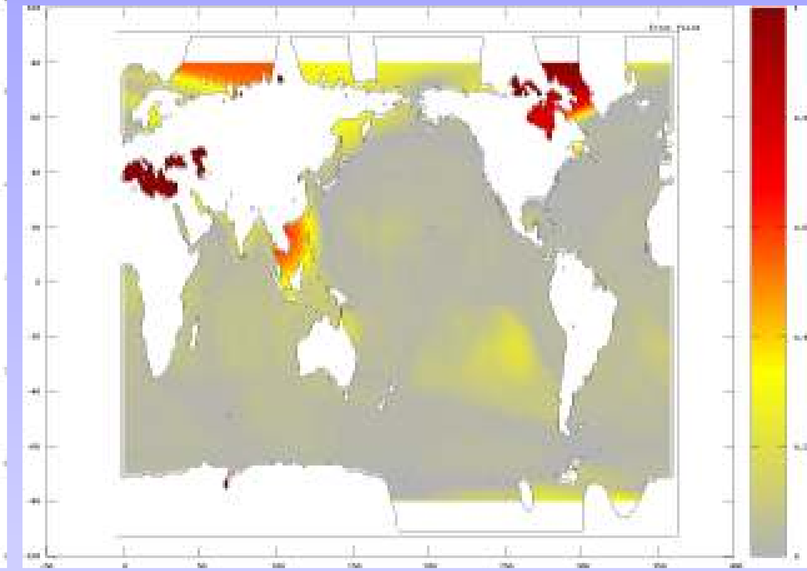
Huge problems

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

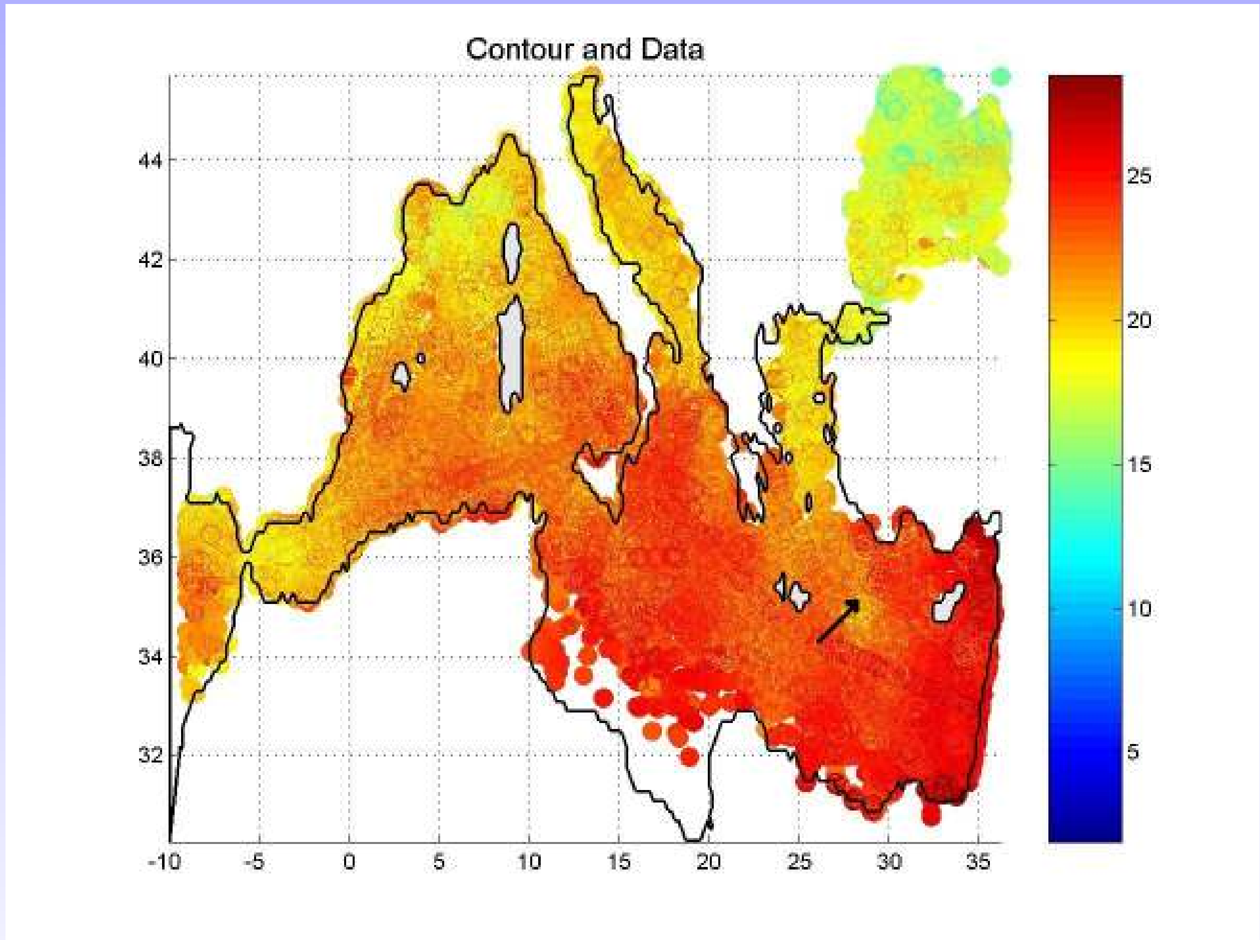




Outliers detected via comparison of statistically expected residuals (value provided by the DIVA analysis) and actual residuals (note that expected residuals decrease for large S/N).
Error field (on the right) can be used to mask regions with large uncertainties (low data coverage and/or large errors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

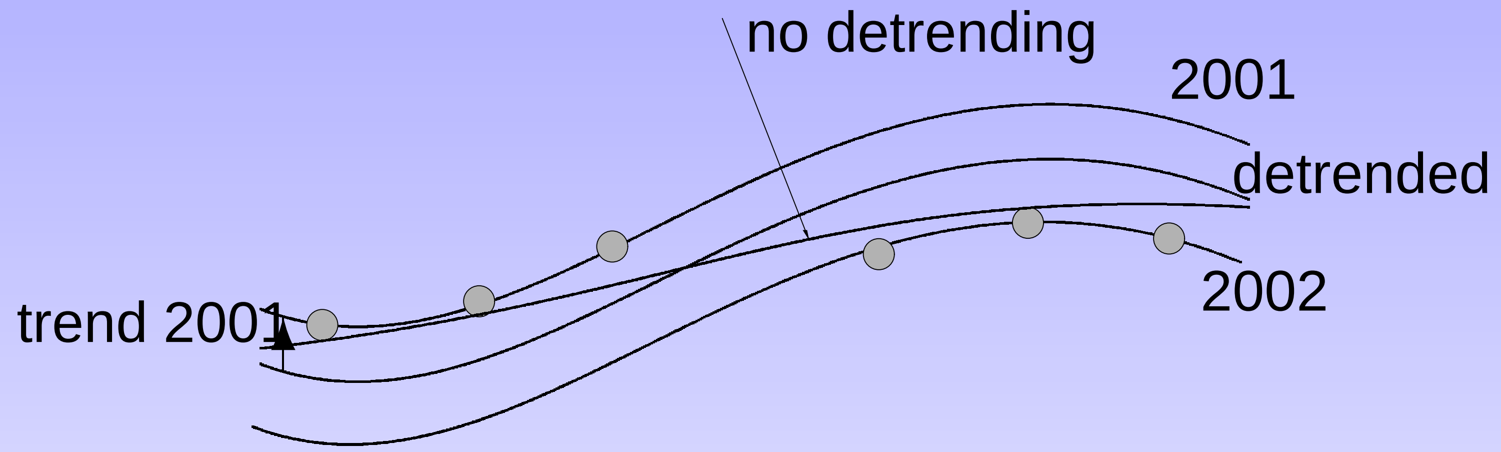


Outlier with `divaqc`

```
/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
```

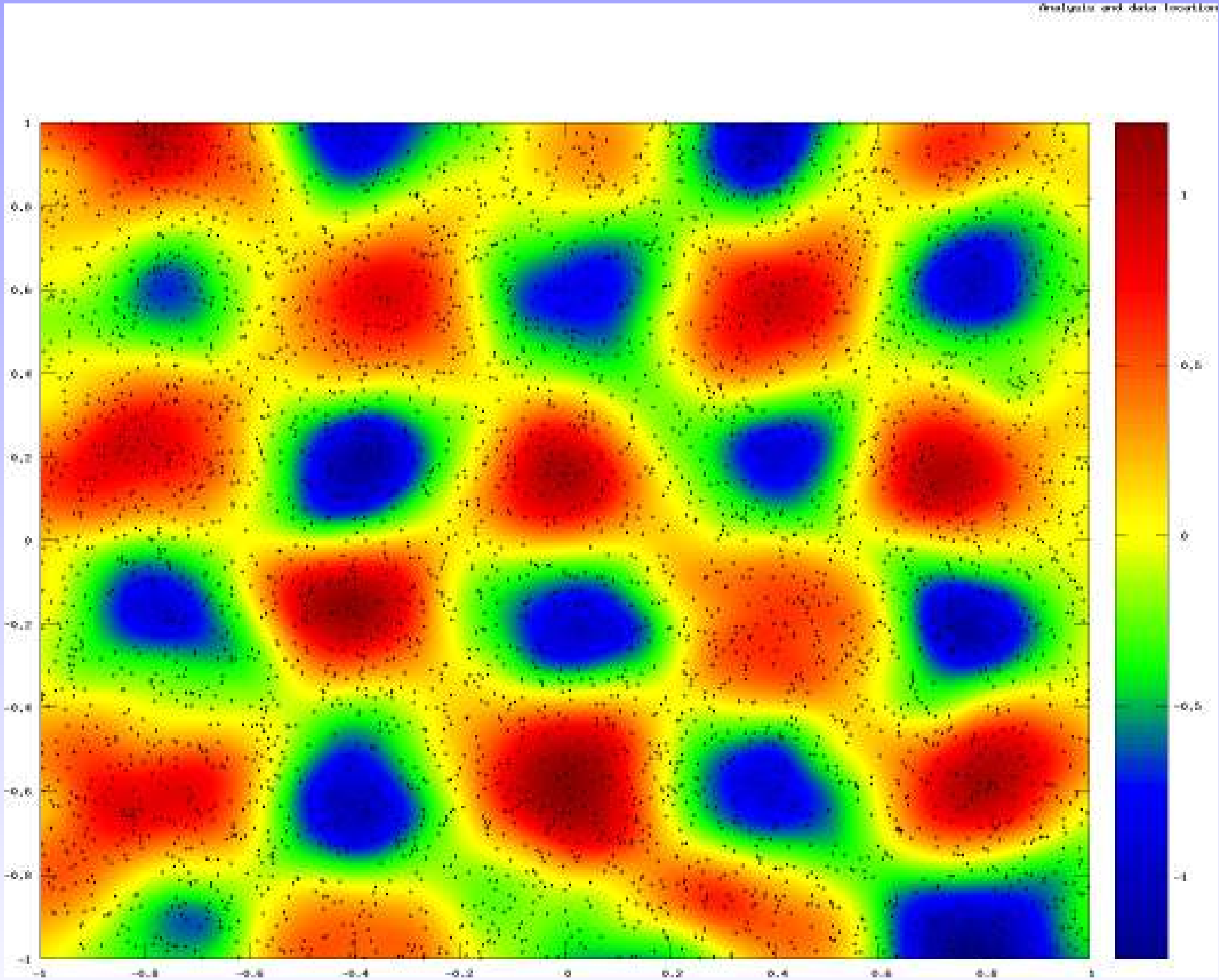
Detrending

Heterogeneous data distribution:



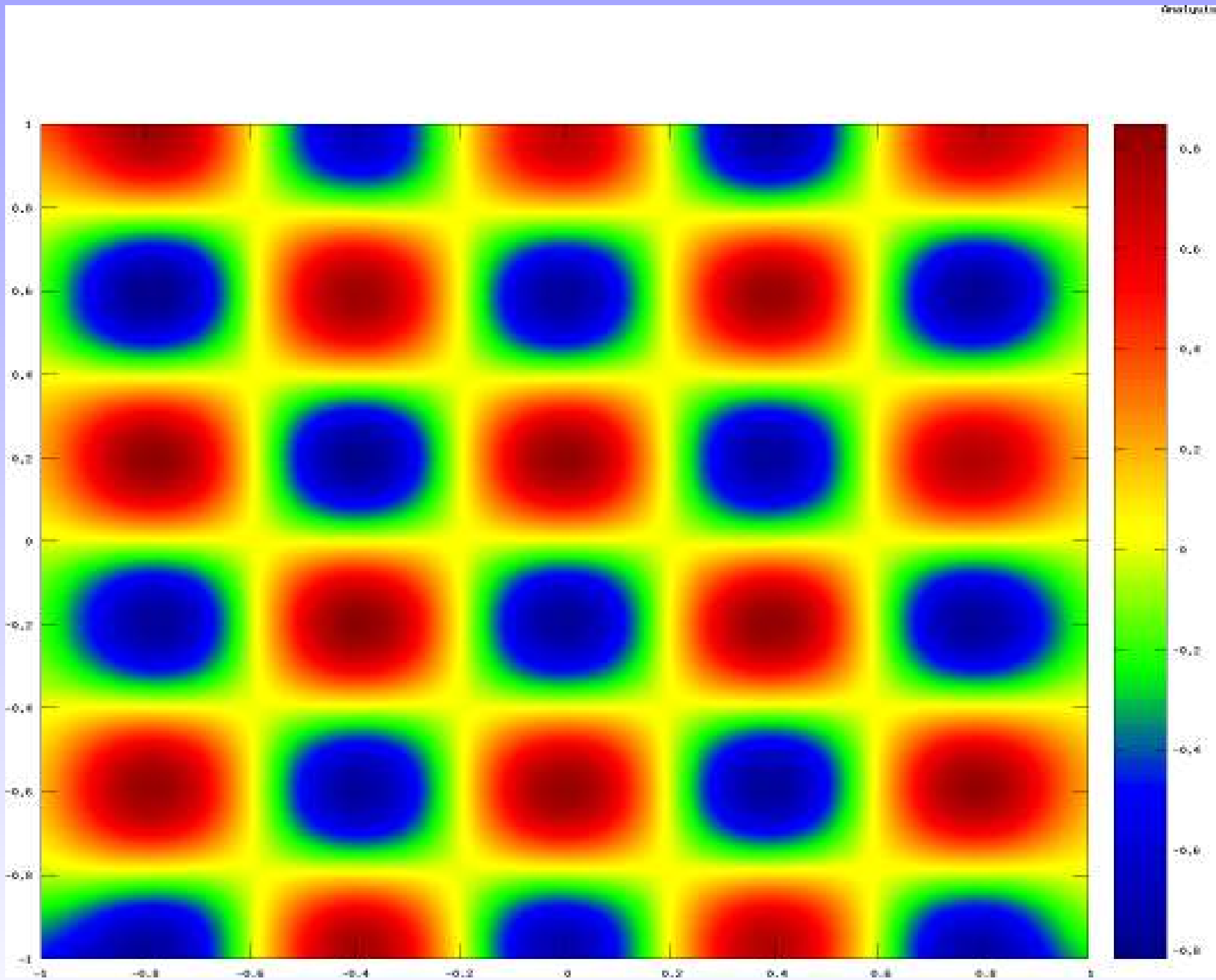
First analysis show a bias for each year's data with respect to first analysis. Subtract the bias estimate and redo the analysis, accumulating the bias. After convergence, detrended analysis+bias of the year.

Example without detrending



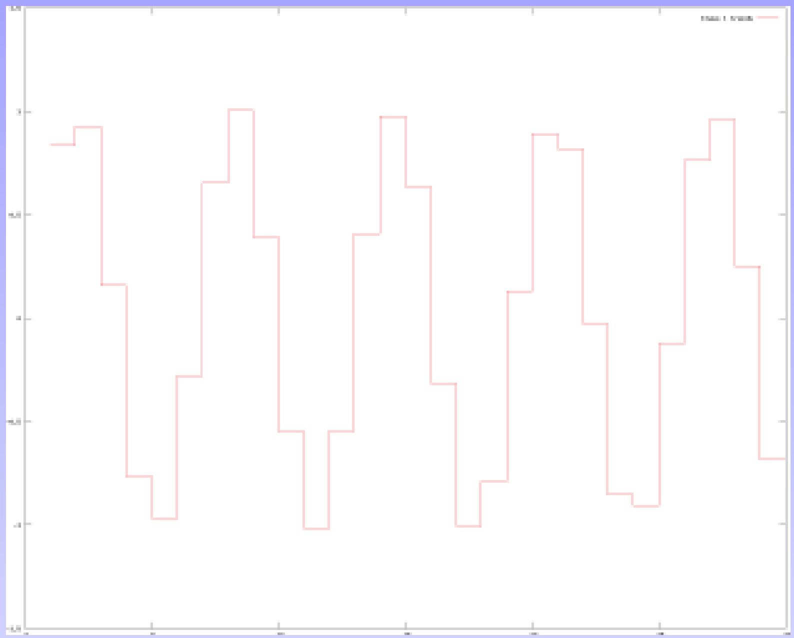
DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.

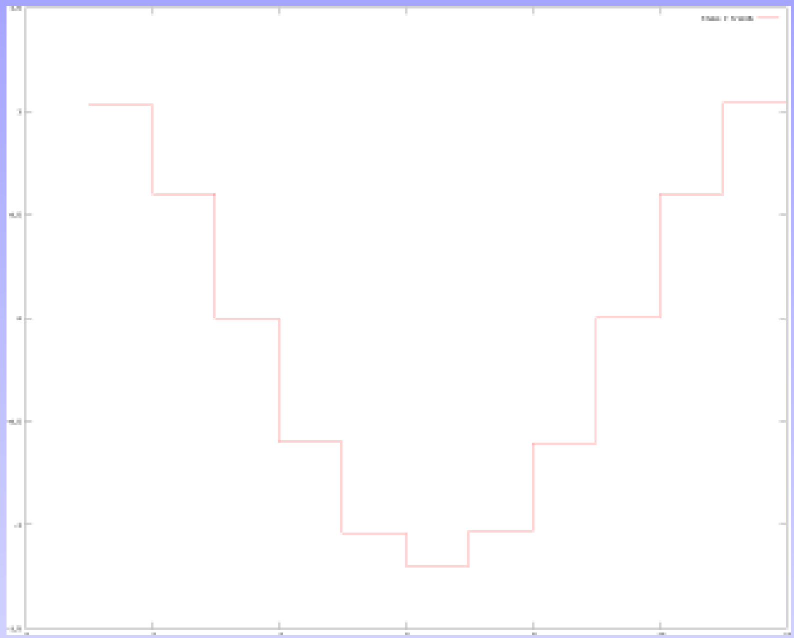
Example with detrending

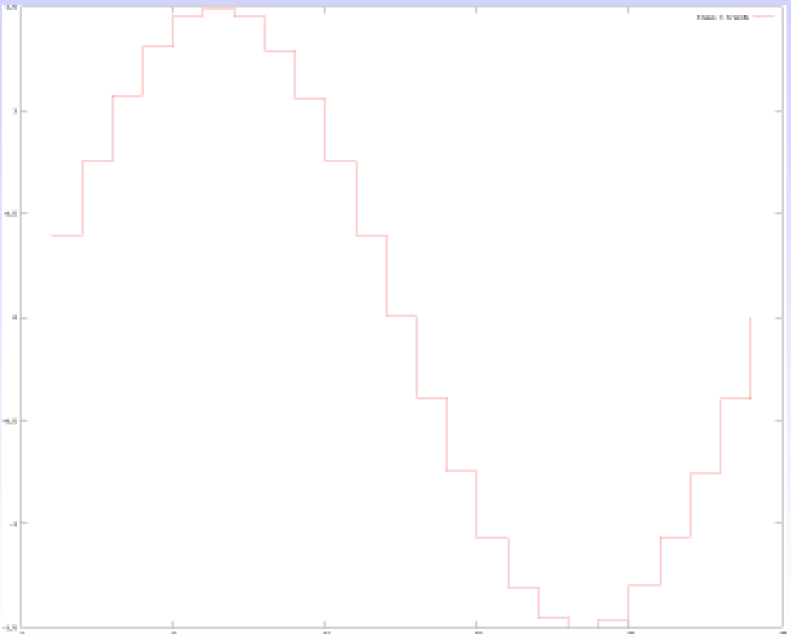


DIVA analysis of sin-cosine spatial structure with superimposed decadal, seasonal and daily cycles and noise.

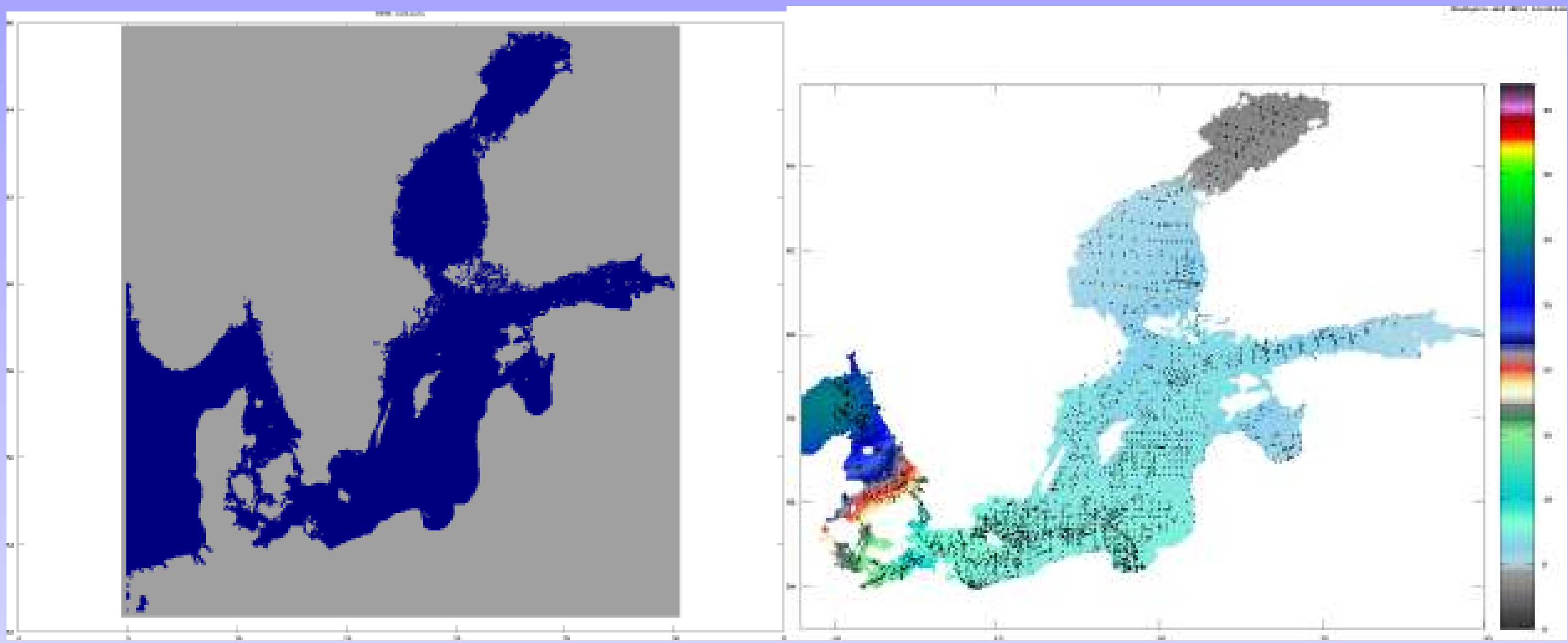
Trends can also be retrieved





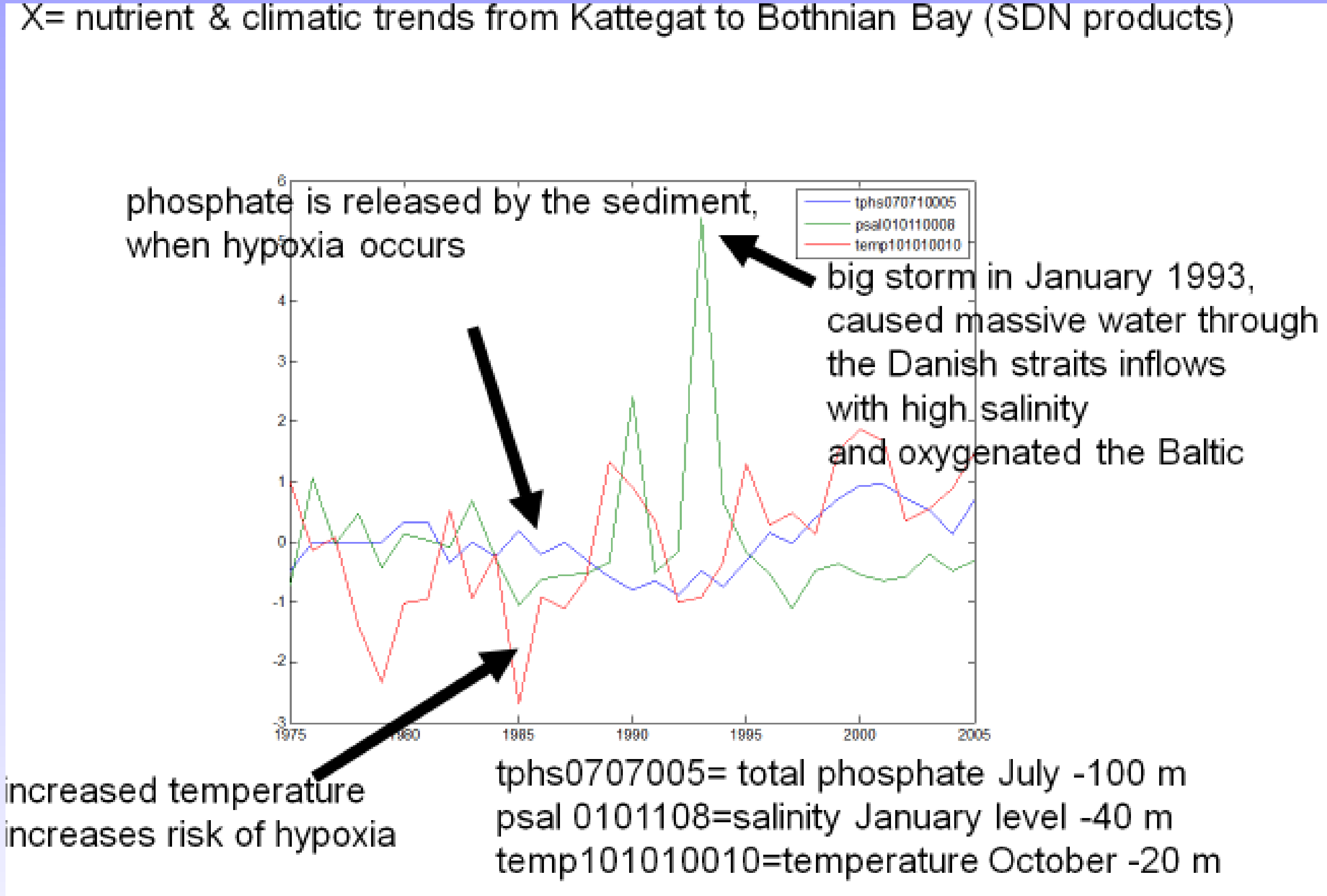


Heterogeneous case



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

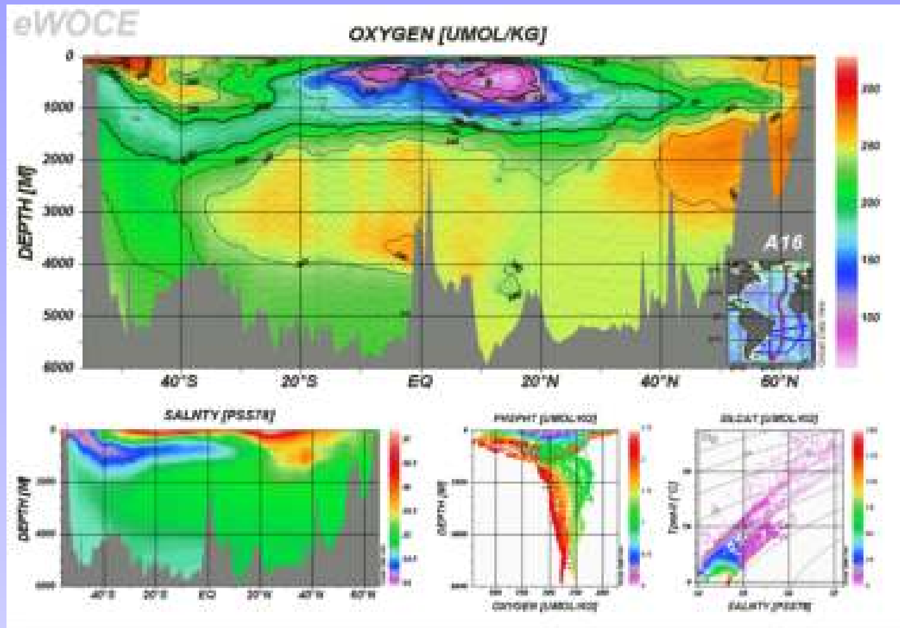
Heterogeneous case, trends

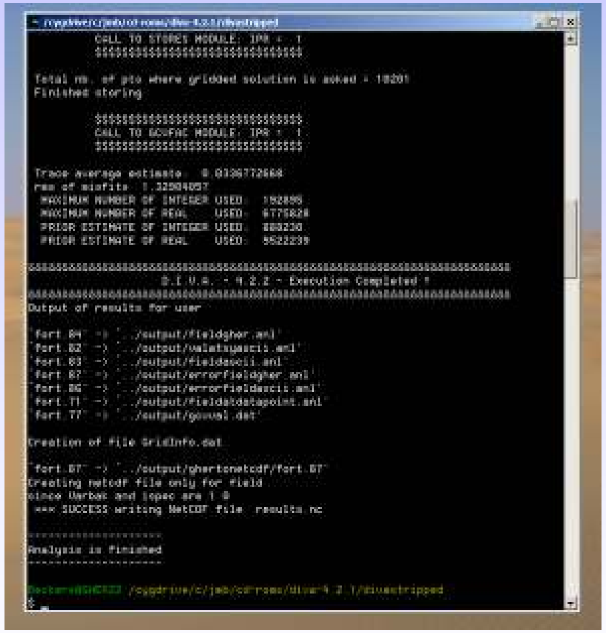


(Bassompierre *et al.*, 2010)

- *Gridding*
- *Diva*
- ***Implementations***
- *And for data managers ?*
- *Summary*

How to use DIVA?

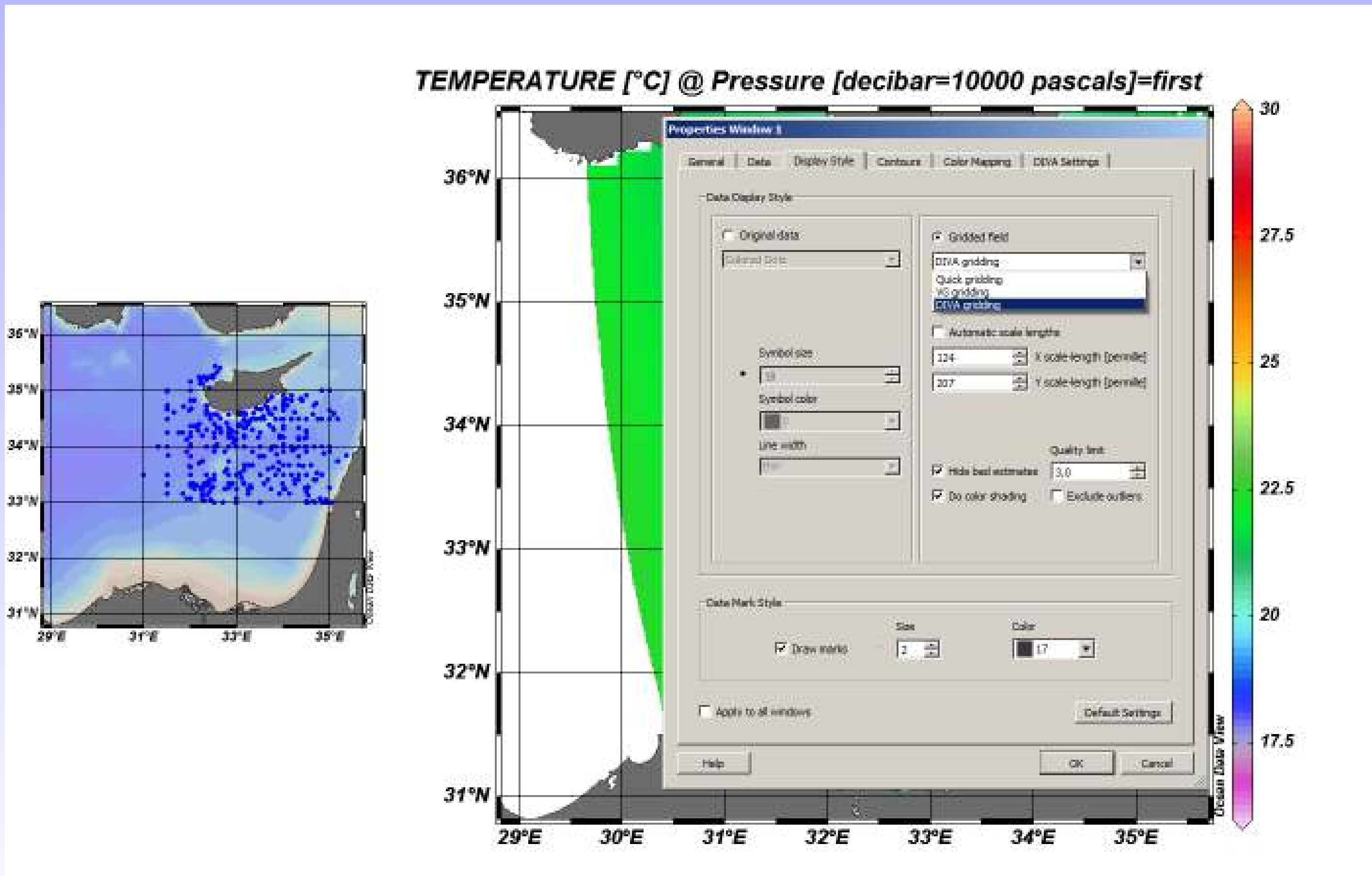




Try parameters of gridding in ODV

On `Display Style`

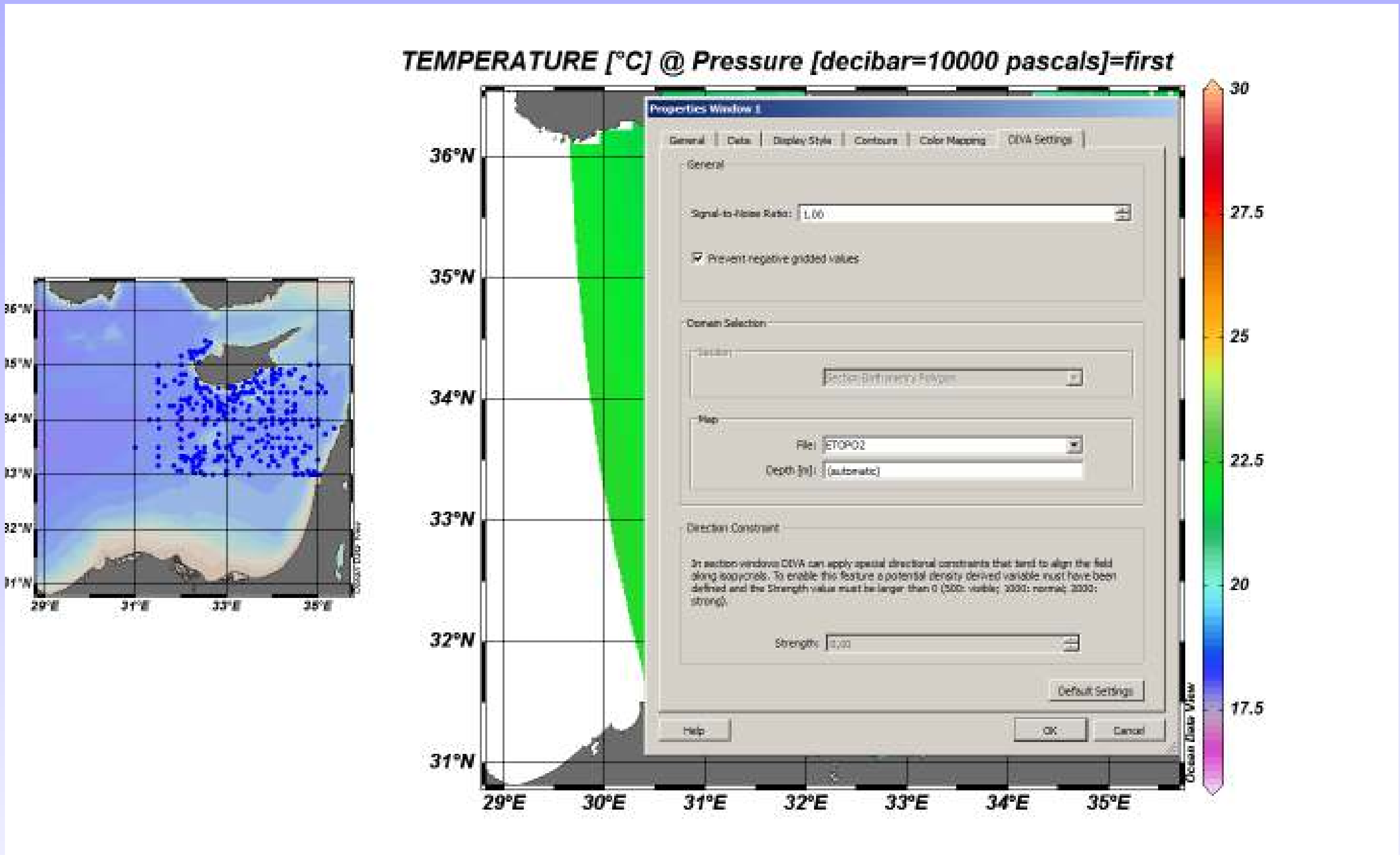
- change analysis method (quick, VG, DIVA)
- change X scale-length and Y scale-length



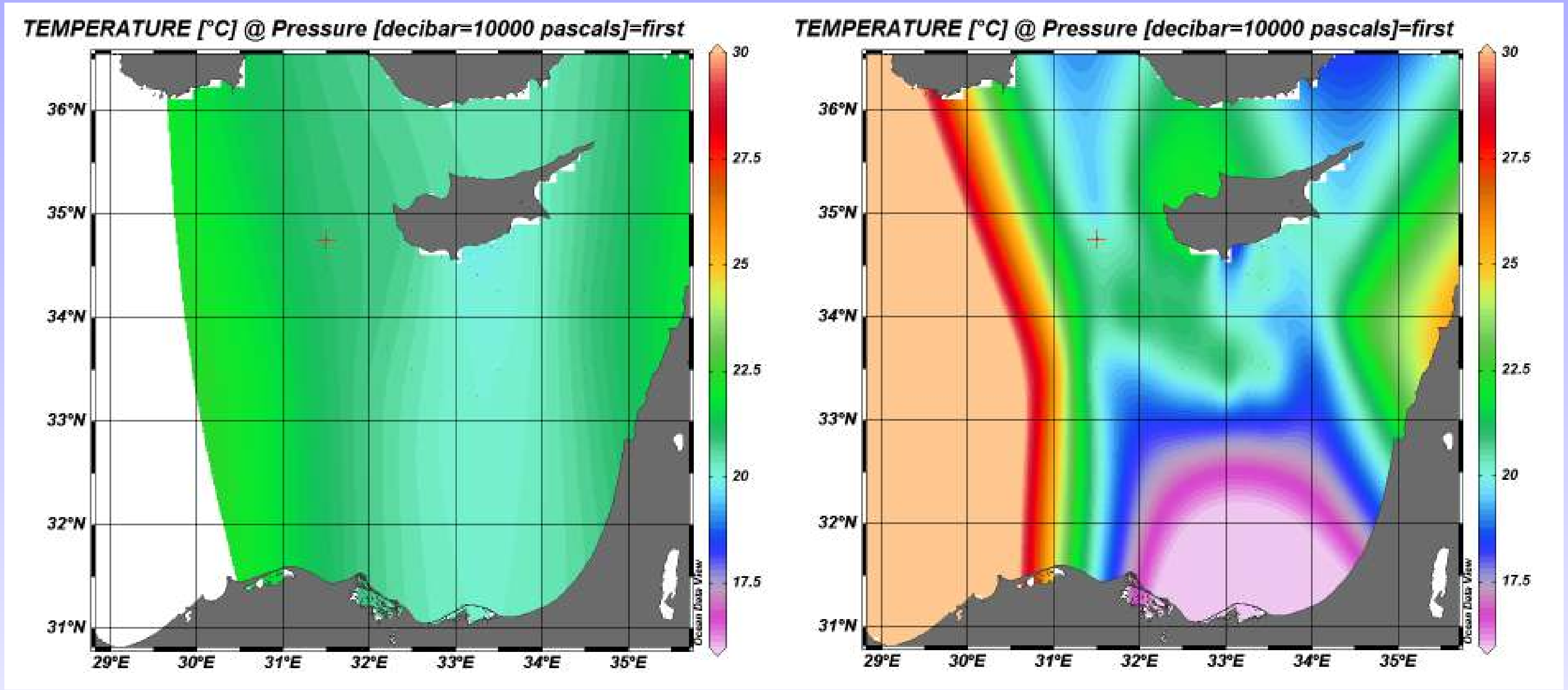
Try parameters of gridding in ODV

On DIVA Settings

- Change Signal-to-Noise Ratio

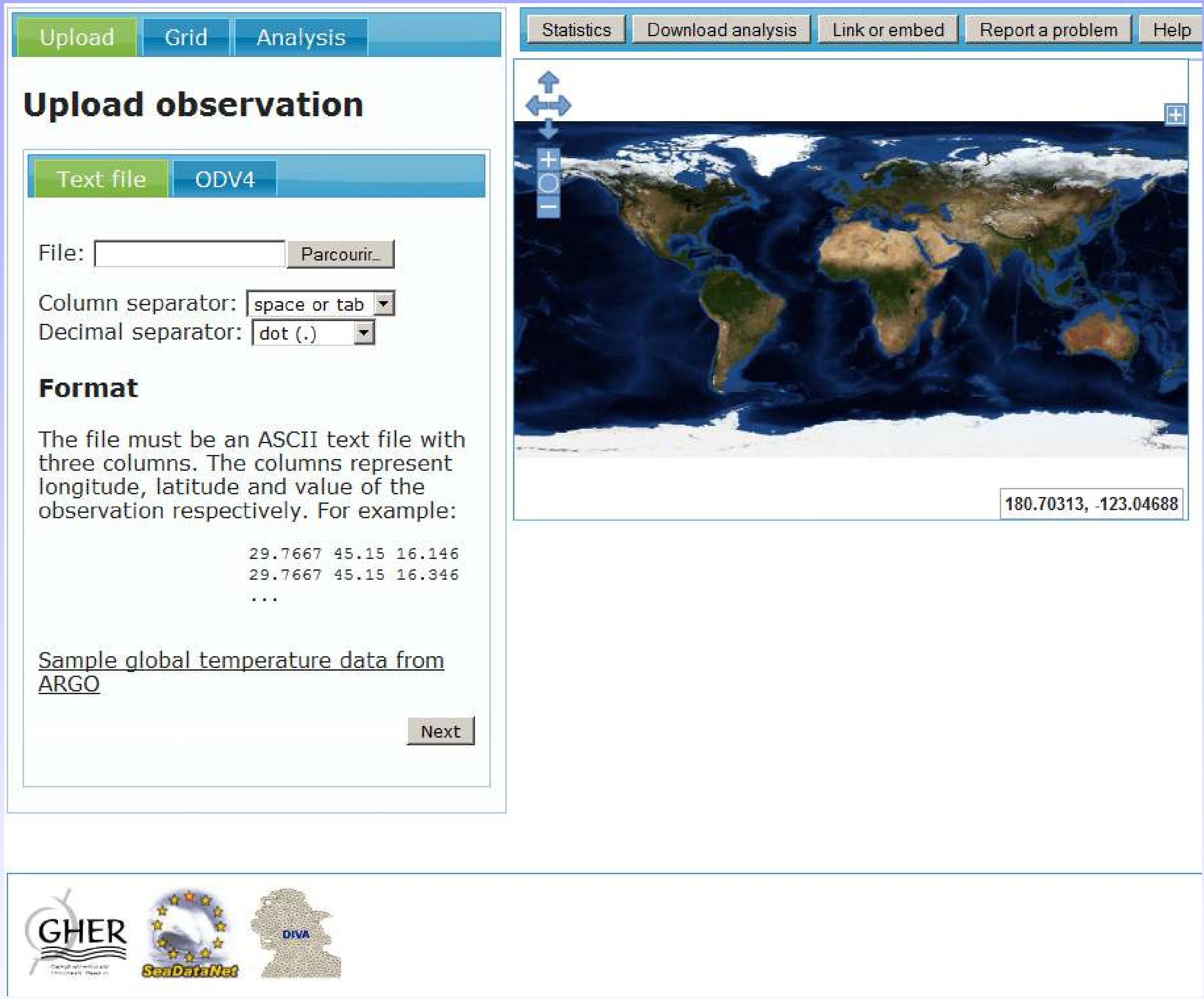


Observed changes in analyses due to changes in parameters (here signal-to-noise ratio from 1 to 100)



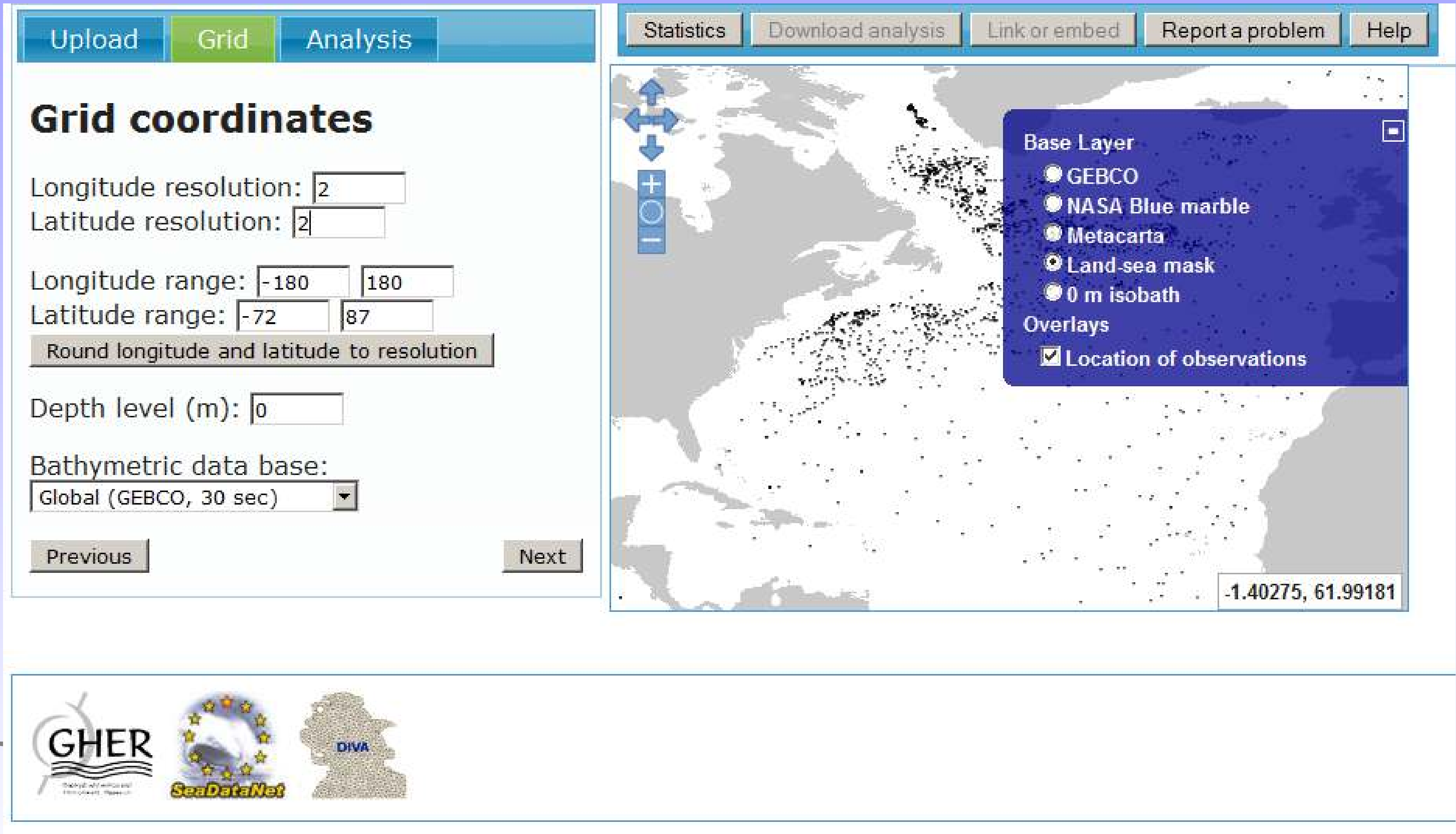
Diva-on-web

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



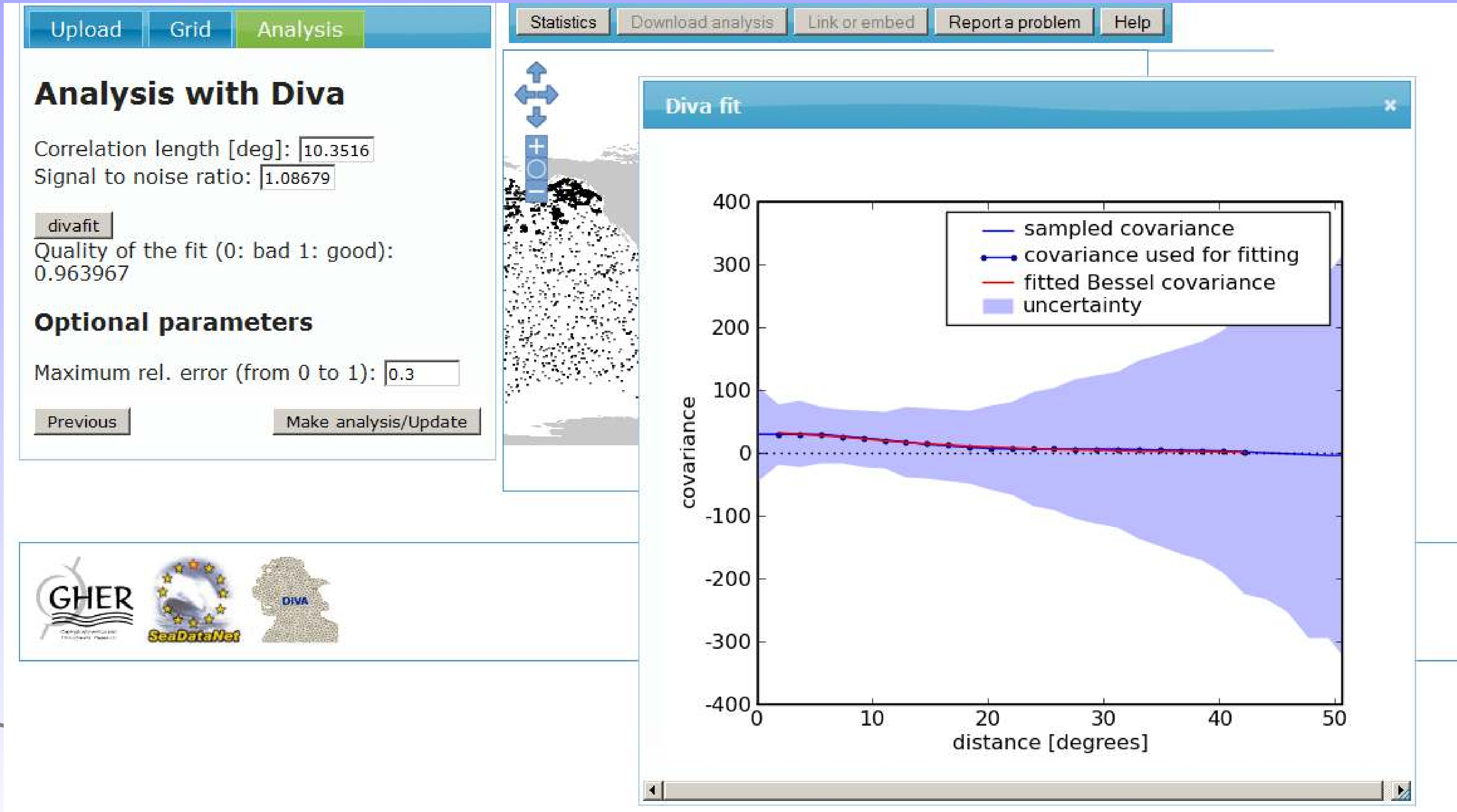
Diva-on-web

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



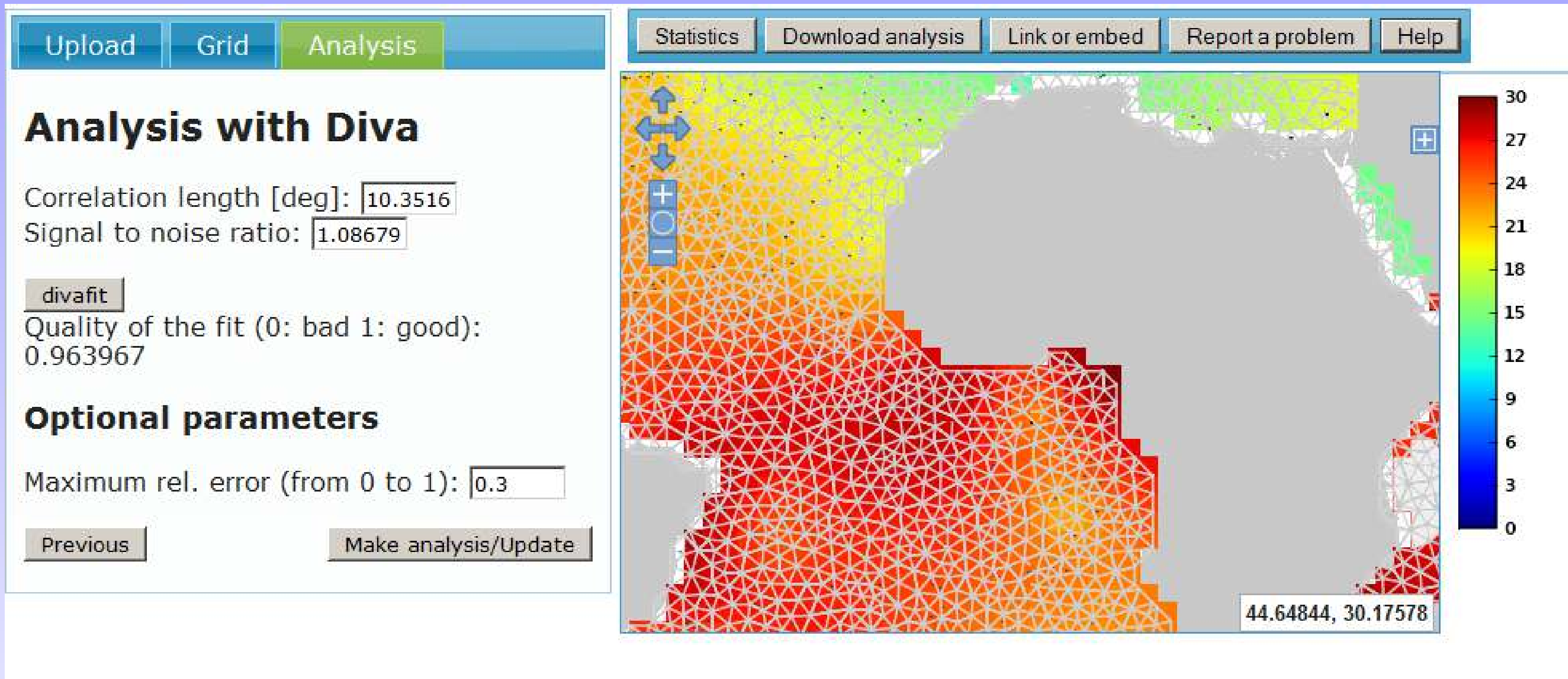
Diva-on-web

Analysis parameter definition (or fit)



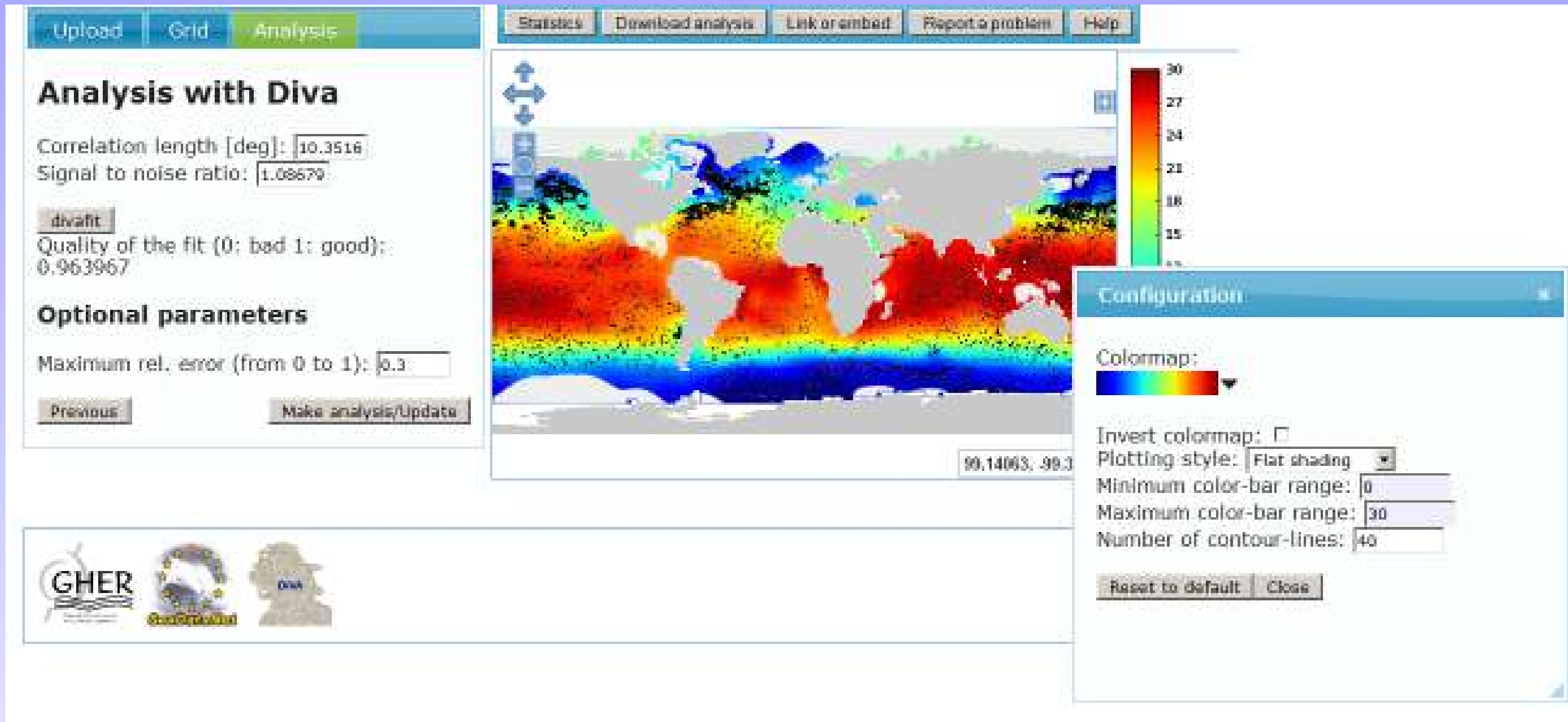
Diva-on-web

Analysis (and numerical grid)



Diva-on-web

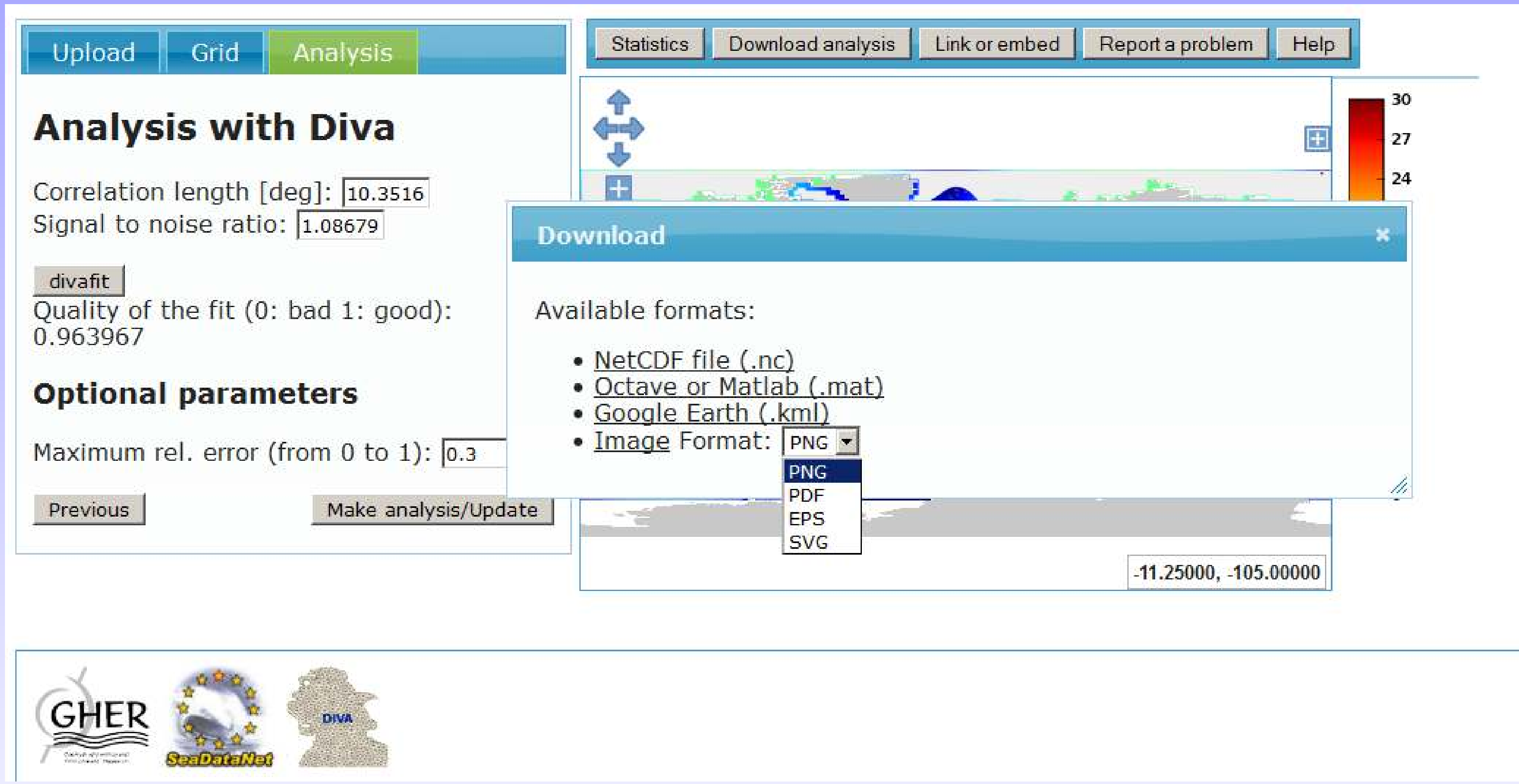
Plotting options



Diva-on-web

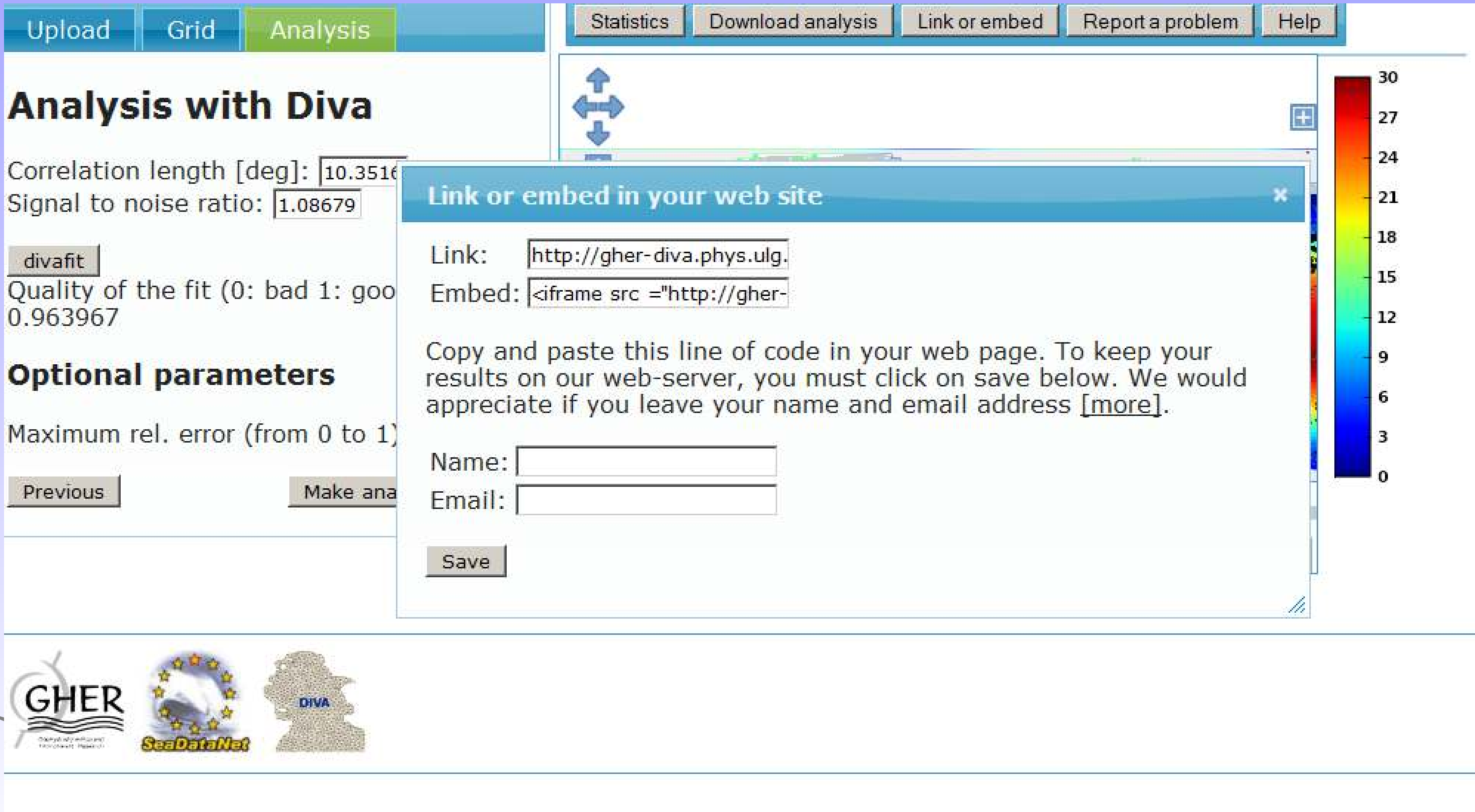
Université de Liège

Download options



Diva-on-web

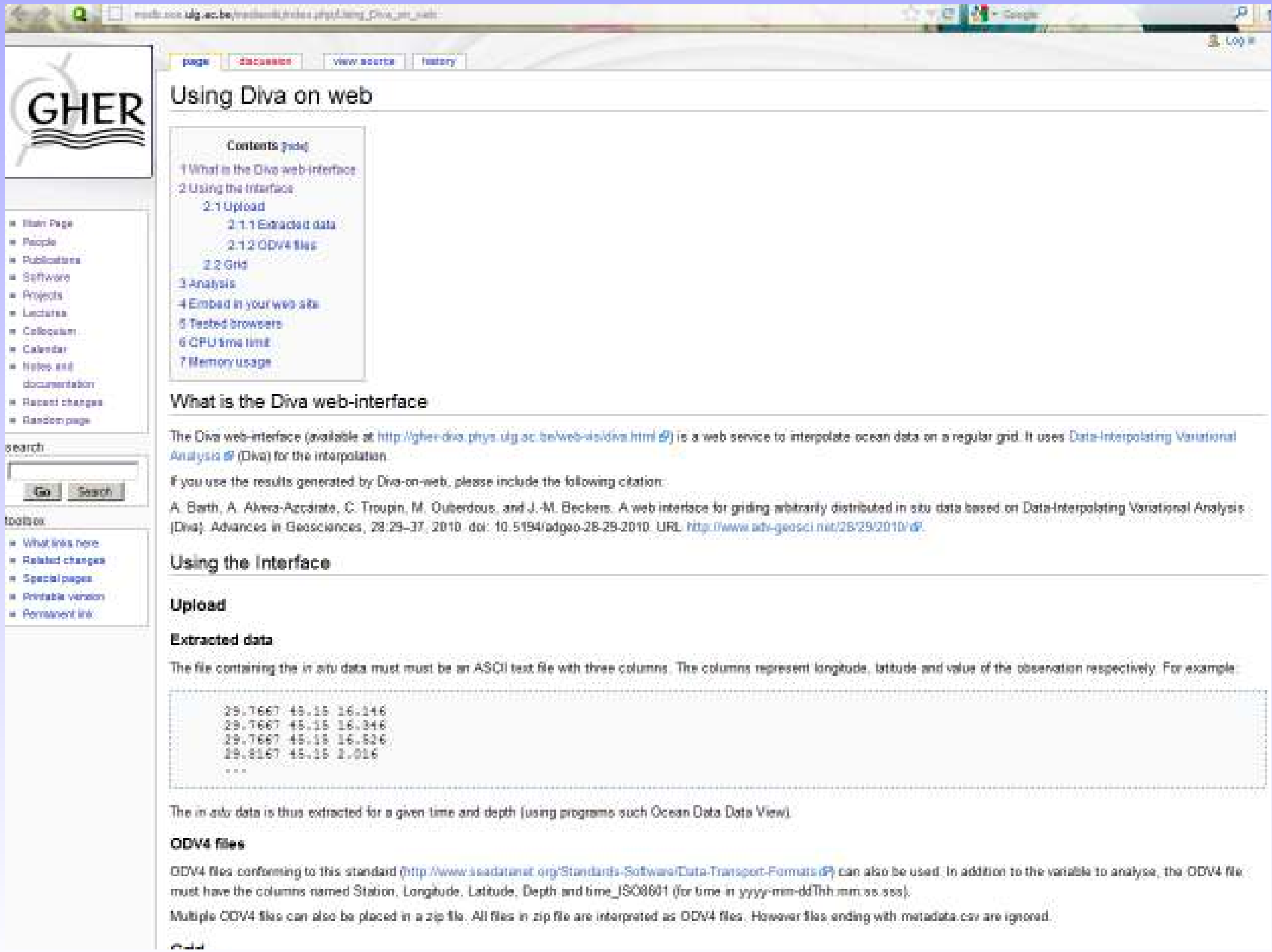
Exploitation of WMS-OGC layering techniques



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]

1 What is the Diva web-interface
2 Using the interface
2.1 Upload
2.1.1 Extracted data
2.1.2 ODV4 files
2.2 Grid
3 Analysis
4 Embed in your web site
5 Tested browsers
6 CPU time limit
7 Memory usage

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.

Software download for advanced users: GODIVA 4D

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

Université de Liège

Université de Liège

- *Gridding*
- *Diva*
- *Implementations*
- ***And for data managers ?***
- *Summary*

? !

Université de Liège

DIVA and data-managers as providers

Scientist rarely use single profiles

- Need for standardisations to simplify automated processing. If not strict adherence
  - ⋆ risk of missing data without user noting the profile was not used
  - ⋆ risk of crashing a complex calculation because a single profile causes problems
- Need for correct metainformation (units, names etc): otherwise again incorrect selections, mixing of different units etc
- QUALITY information: standardised flags for data selection AND weighting of information

Users do NOT want to go through checking of all profiles again and if they want to merge with other data they want to automize transformations.

DIVA as a tool for data-managers

DIVA can also be used by data-managers themselves

- Checking data flow and conformity to standards by trying to use DIVA (diva-on-web or GODIVA)
- Visual representation and verification of data set (not only profiles but spatial structures)
- Exploit outlier detection using spatial information on top of standard QC
- For interested centers: diva analysis as exploratory tool for data bases (mapping on the fly with diva-on-web engine or batch mode of ODV)

- *Gridding*
- *Diva*
- *Implementations*
- *And for data managers ?*
- ***Summary***

Université de Liège

Possible uses for you

DIVA can be used in several ways. As data managers you can work with DIVA as format checking tool and gridding tool (gridding is not plotting) . When using DIVA

- Report non-standard ODV4 files retrieved from SDN interface to SDN-TTT
- Report diva runtime errors (be it in ODV or diva-on-web or GODIVA) to GHER
- For installation of GODIVA (4D batch mode), contact GHER
- For bugs on diva-on-web, use web interface for reporting
- Do not hesitate to add things to a wishlist
- Strong interactions during DIVA workshops
  `http://modb.oce.ulg.ac.be/mediawiki/index.php/Diva_workshop_2014`

but you can also point scientists to DIVA if they use your data base.

- 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`