Université de Liège

Université de Liège

Data-Interpolating Variational Analysis

Marine Observation and Data Expert Group

Jean-Marie Beckers and GHER group Aida Alvera, Alexander Barth, Luc Vandenbulcke, Charles Troupin, Damien Sirjacobs, Mohamed Ouberdous

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

Outline

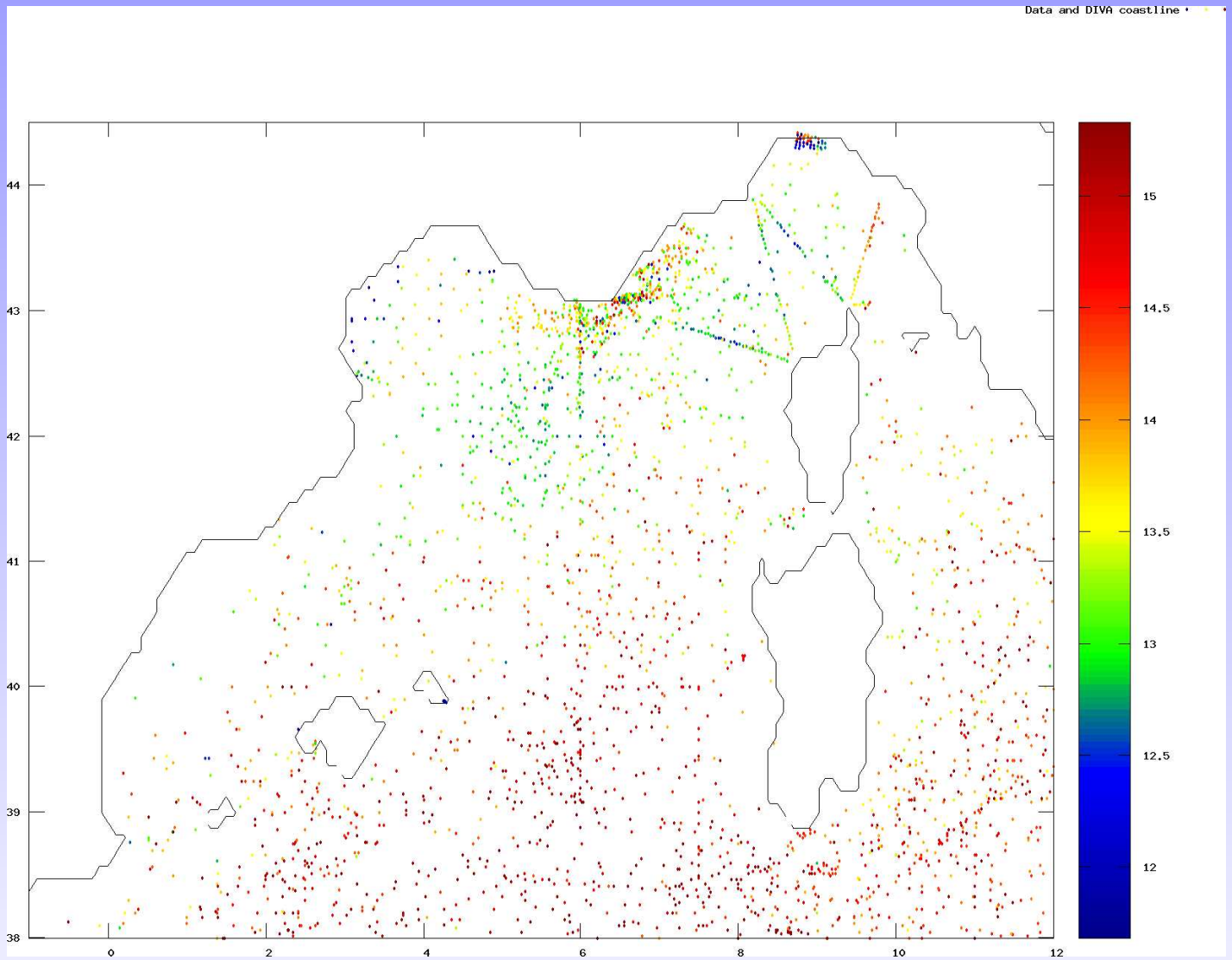
- Basics
- When to use
- Critical points
- Examples
- Plans

Université de Liège

- ***Basics***
- *When to use*
- *Critical points*
- *Examples*
- *Plans*

Université de Liège

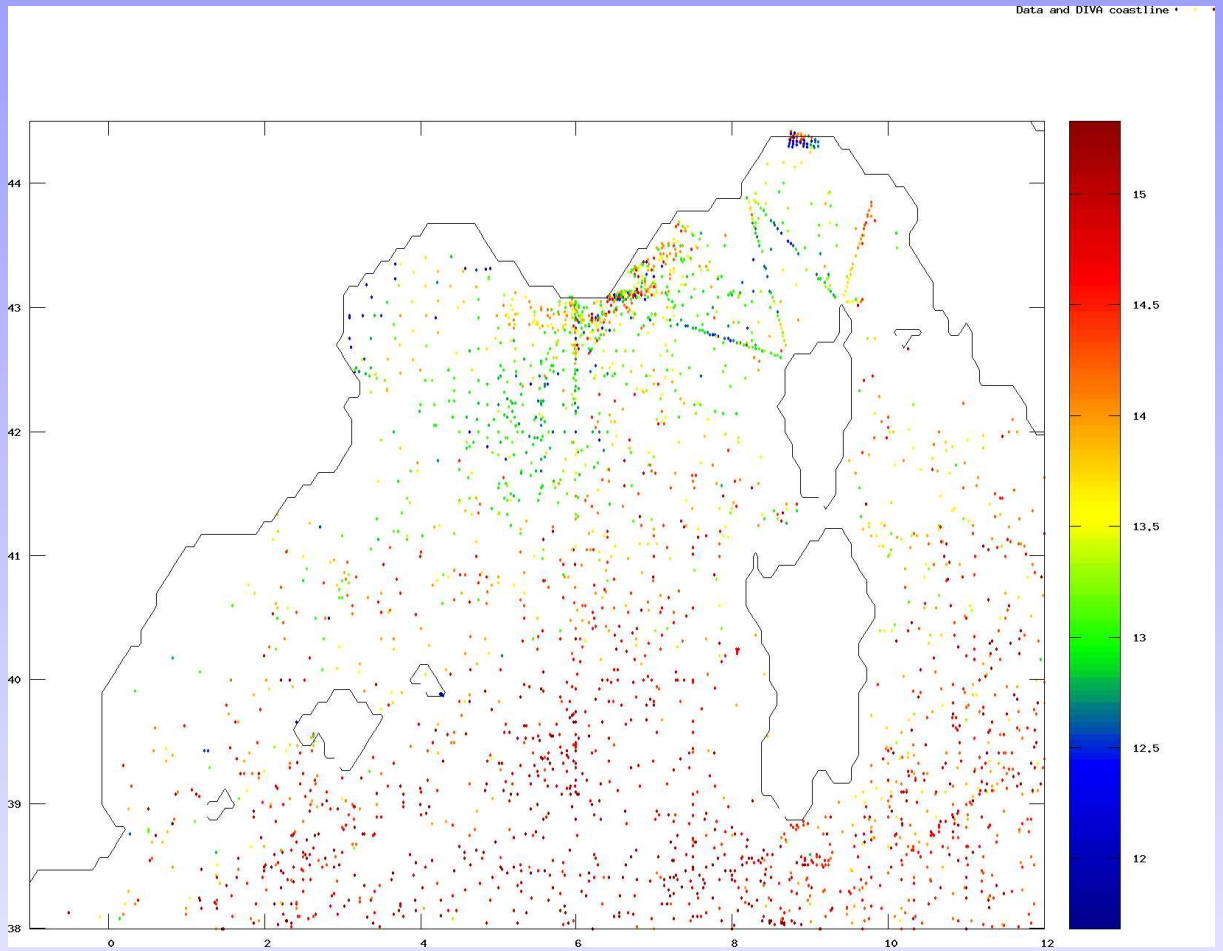
Common problem



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

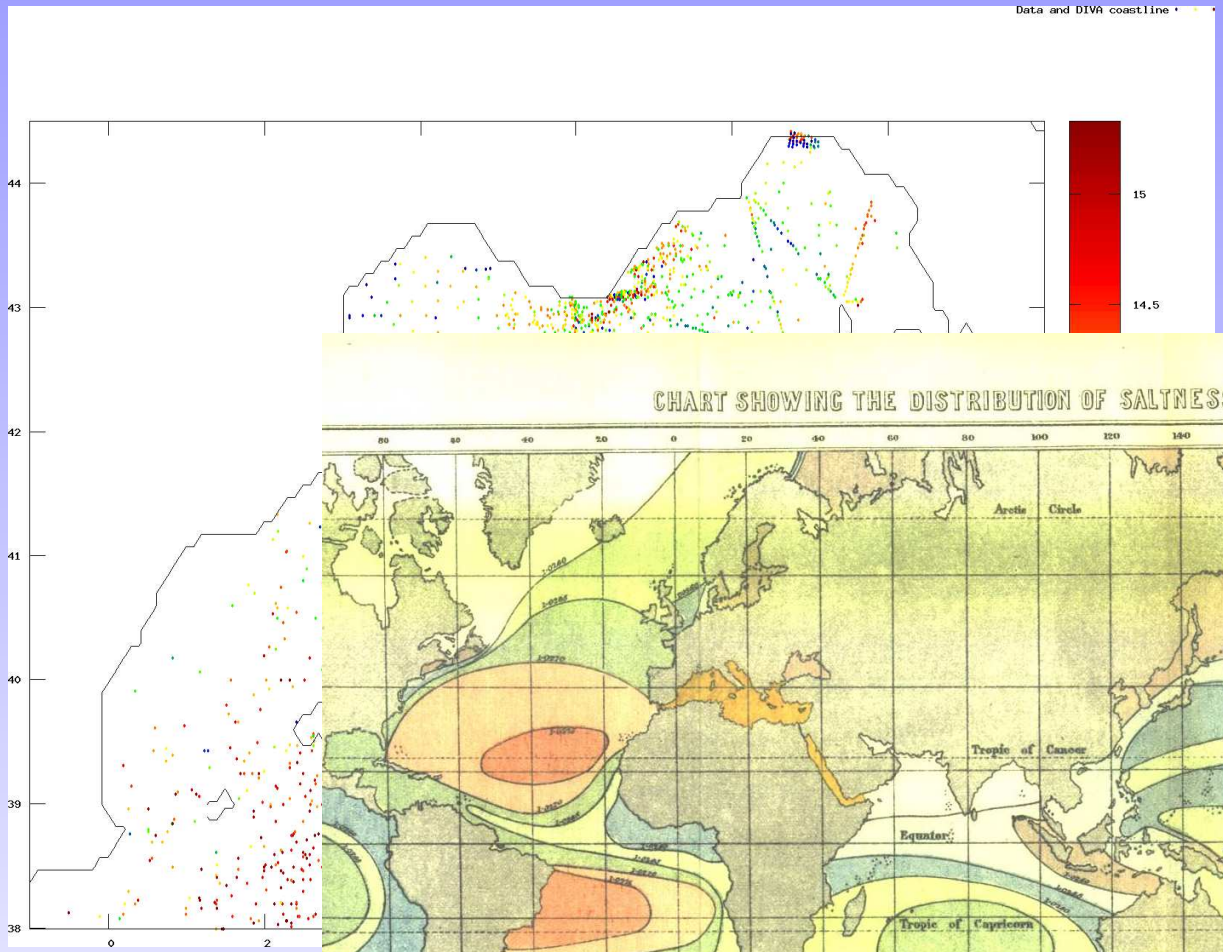
Université de Liège

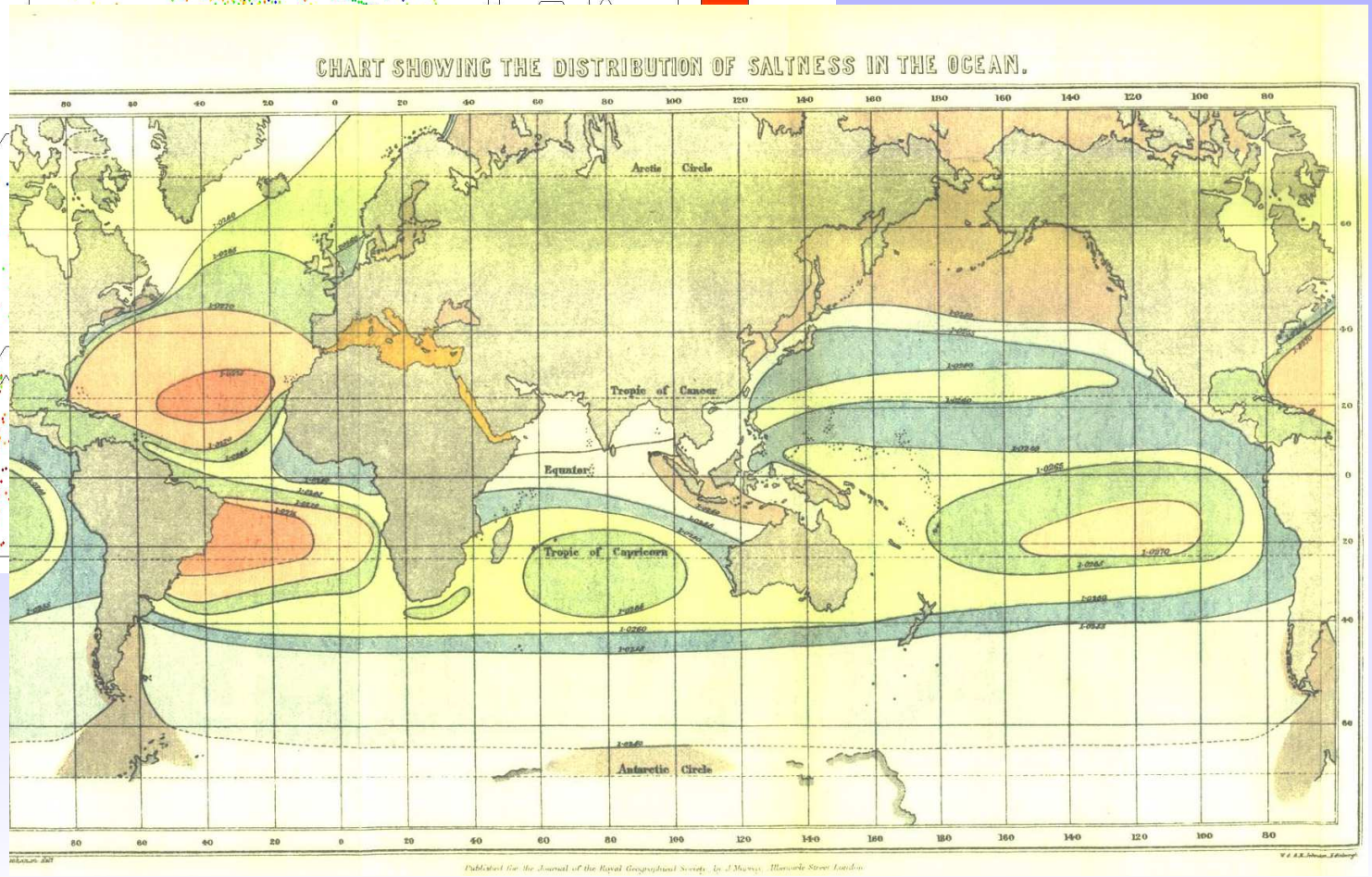
Solutions



Université de Liège

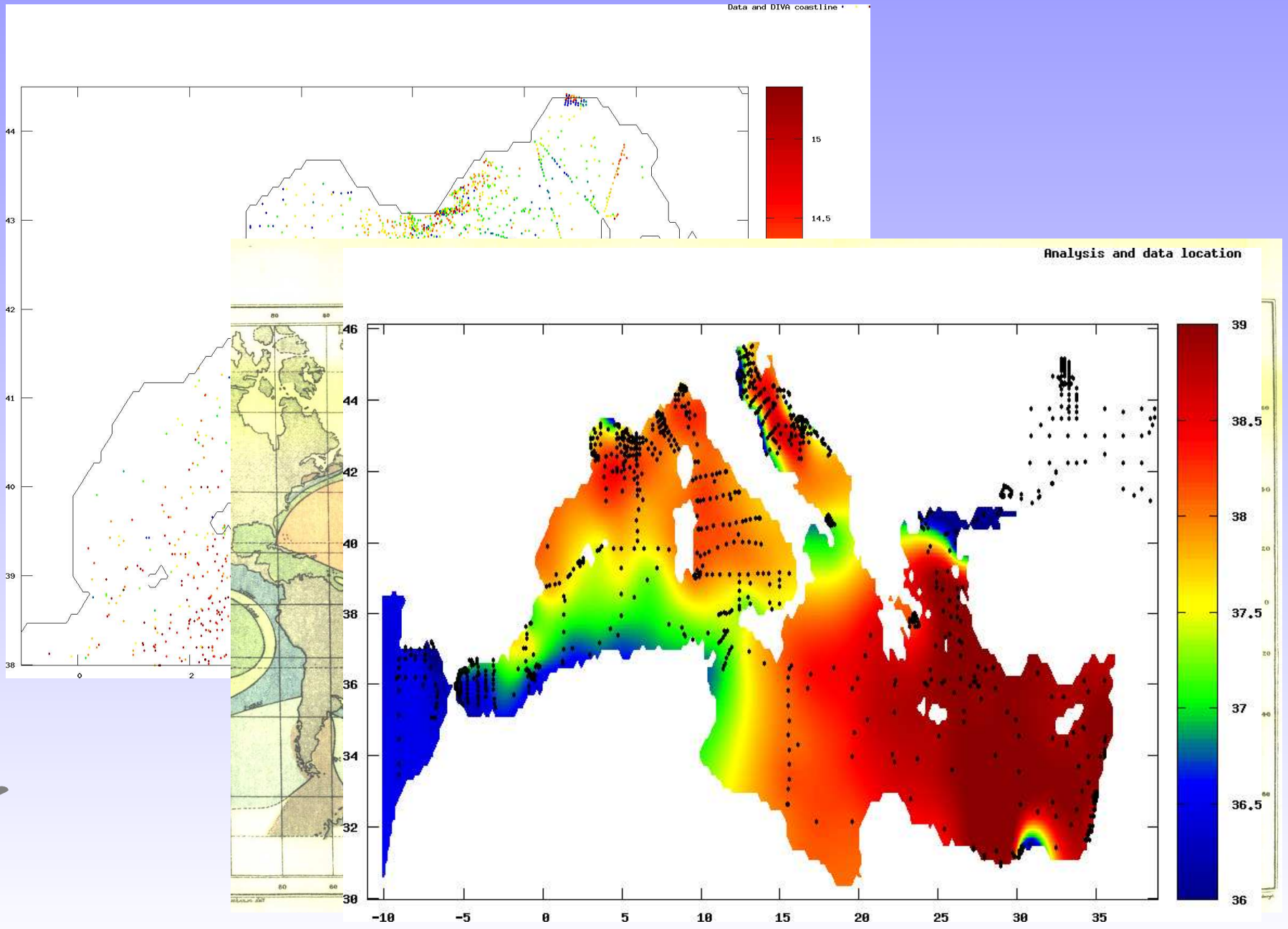
Solutions





Université de Liège

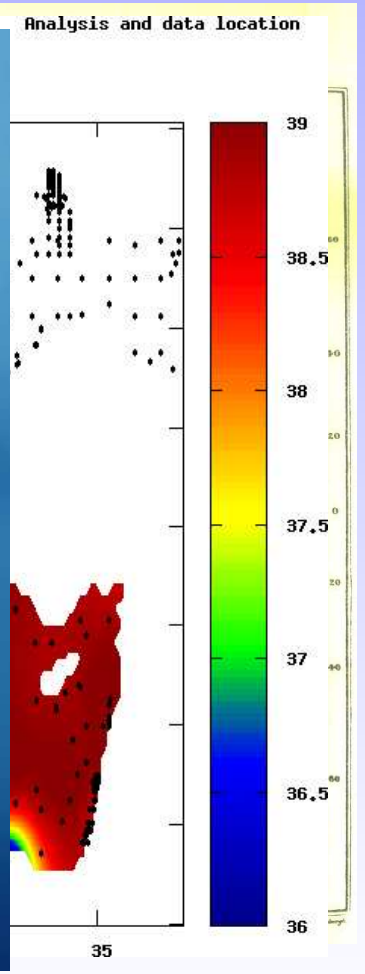
Solutions



Solutions

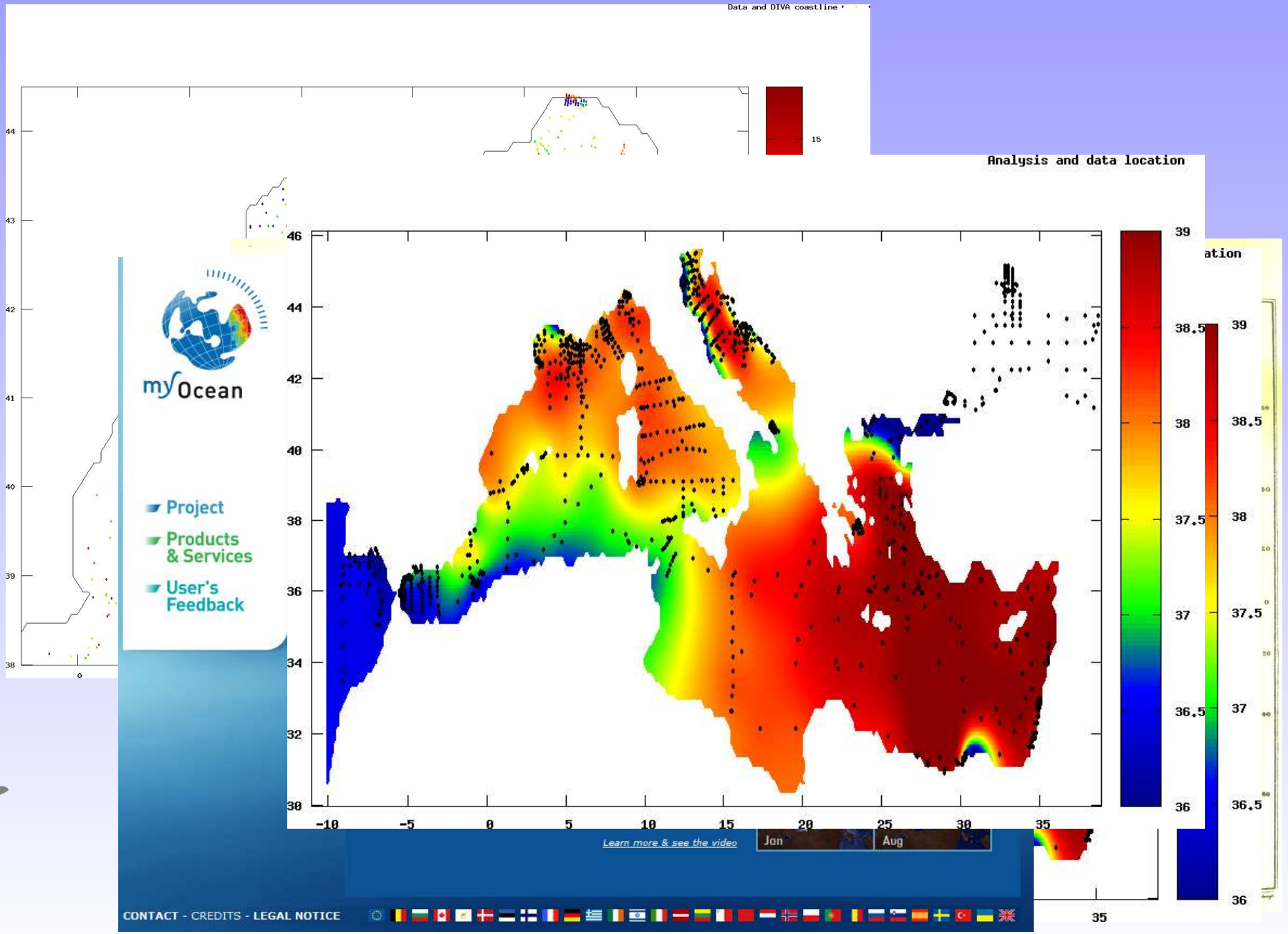

Data and DIVA coastline
15
14.5
44
43
42
41
40
39
38
0

MyOcean
GMES
myOcean
Project
Products & Services
User's Feedback
myOcean
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
Jan
Aug
CONTACT - CREDITS - LEGAL NOTICE


Analysis and data location
39
38.5
38
37.5
37
36.5
36
35

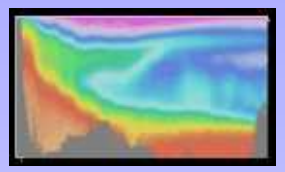
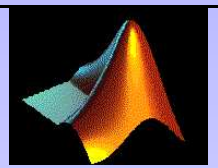
Université de Liège

Solutions



Université de Liège

Statistical spatial analysis

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

Mostly graphics oriented, without "oceanographic" knowledge.

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

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

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

The background field φ_b is typically the data average value.

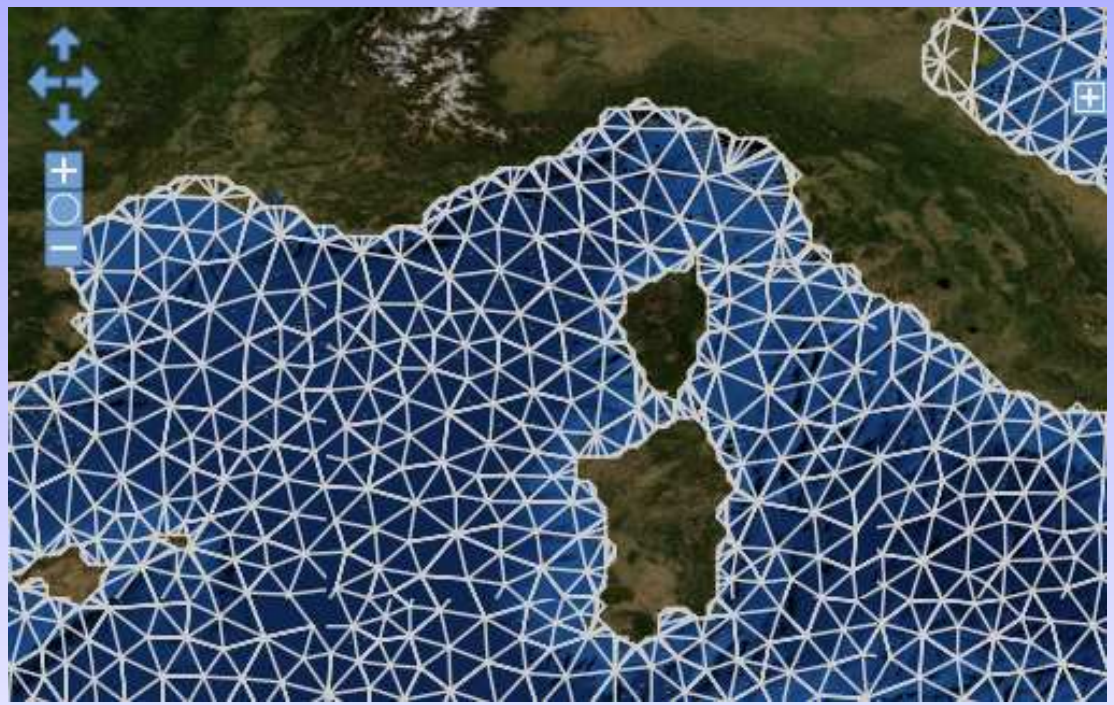
- α_0 penalizes the field itself (anomalies),
- α_1 penalizes gradients $\nabla\varphi$ (no trends) ,
- α_2 penalizes variability (regularization of second derivatives $\nabla\nabla\varphi$).
- α_* can be related to a length scale L of the analysis.
- μ_j penalizes data-analysis misfits (objective).

Université de Liège

DIVA basics

$$\mu = \frac{\sigma^2}{\epsilon^2}\frac{4\pi}{L^2} \tag{3}$$

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



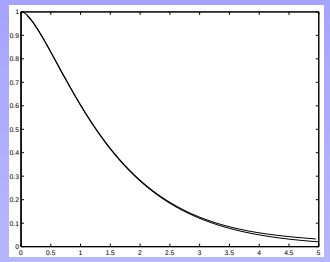
Solution by finite element method including DIVA mesh generator using topographic data. Note decoupling of subbasins.

Université de Liège

Université de Liège

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.

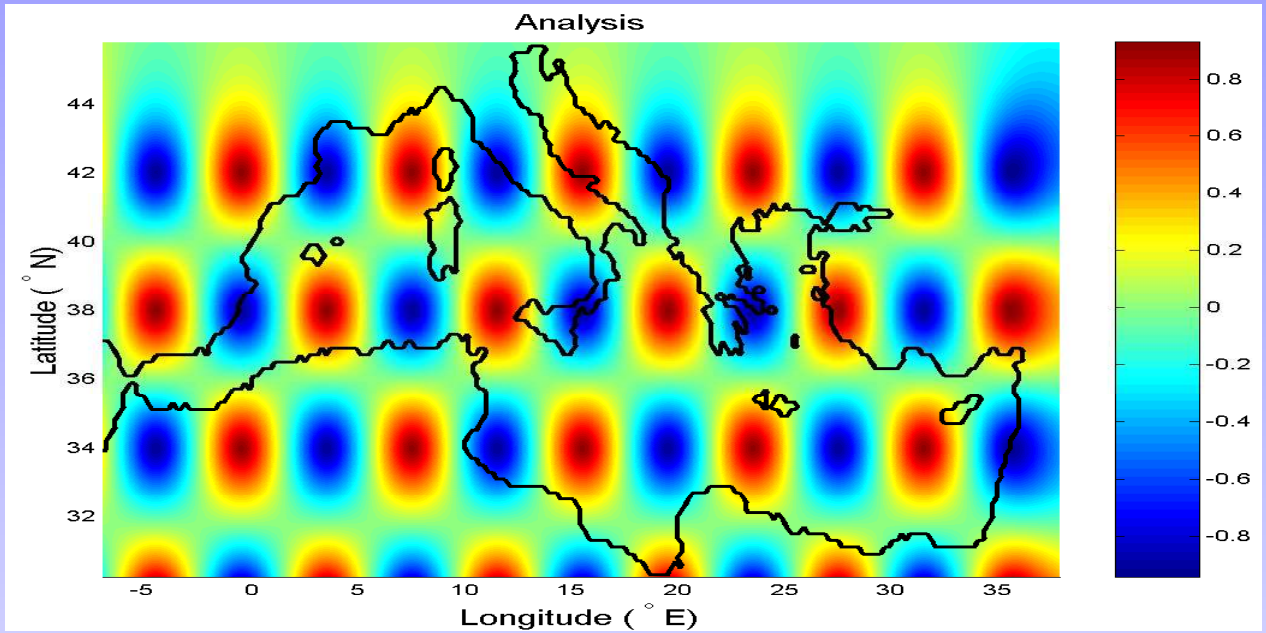
In this case, the OI solution=DIVA solution.

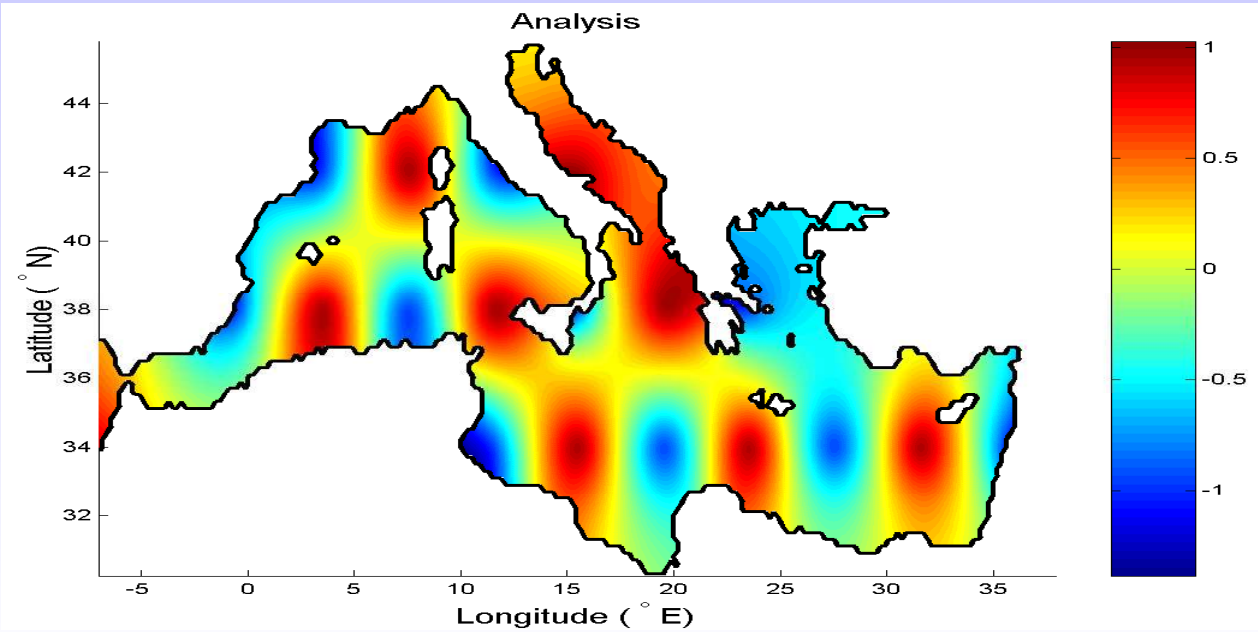
- Advantages of DIVA: regularization, fast finite element solution, boundary effects taken into account.
- Difficulties: generalizations to 3D, error estimates and multivariate versions are "hybrid".

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.

Université de Liège

Illustration of covariance functions





Université de Liège

Additions to basic tool

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

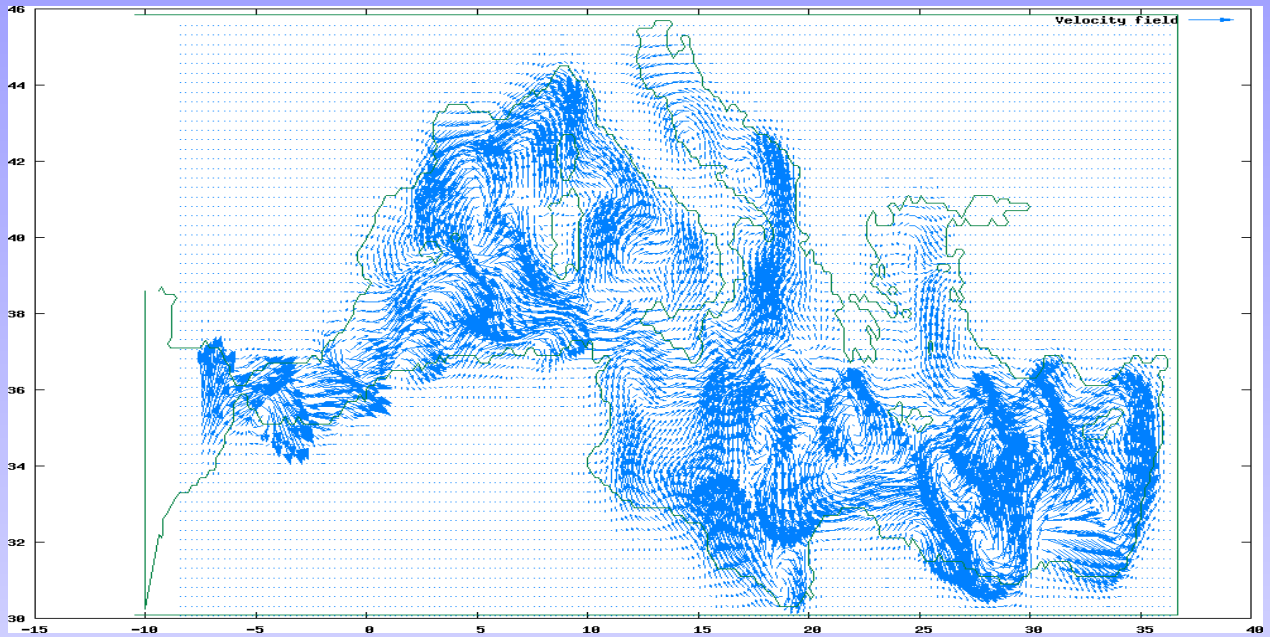
$$J_a = J(\varphi) + \theta \int_D \left[\boldsymbol{u} \cdot \nabla\varphi - \mathcal{A}\nabla\cdot\nabla\varphi\right]^2 \mathrm{d}D \qquad (4)$$

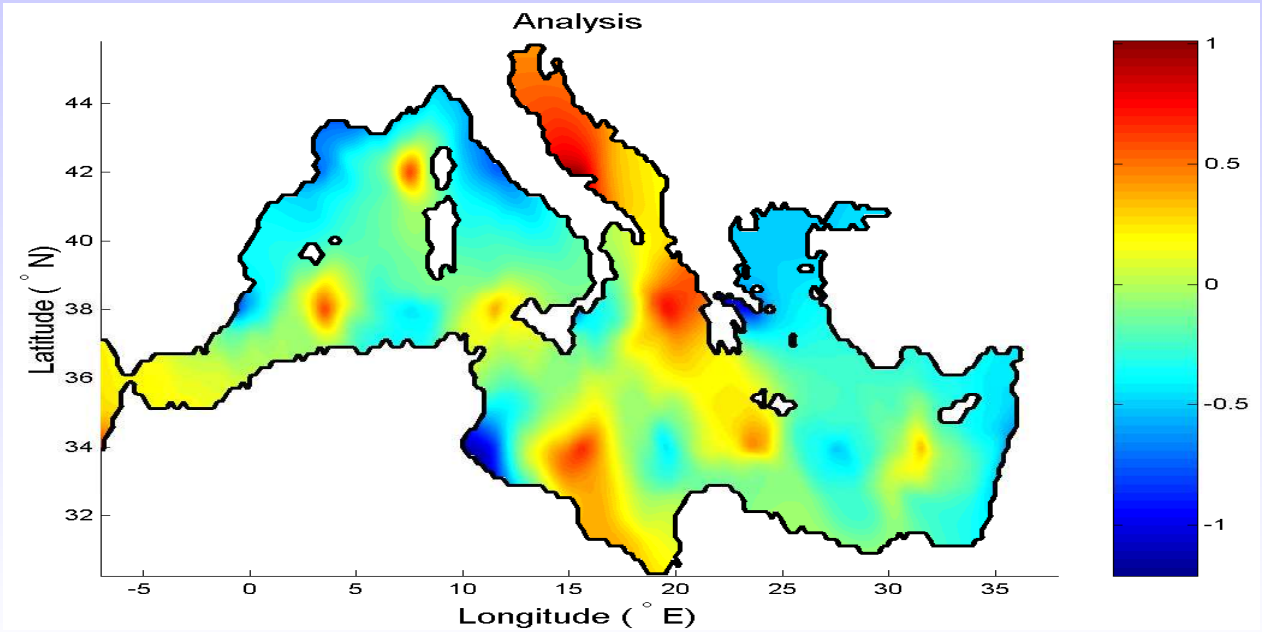
Other features

- 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.
- ...

Université de Liège

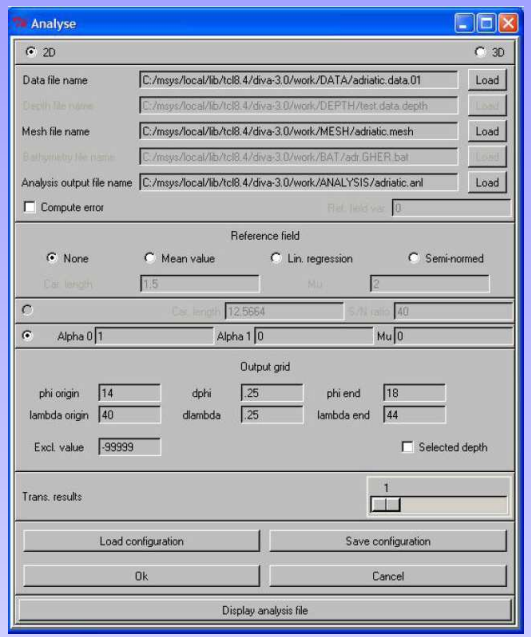
Covariances with advection

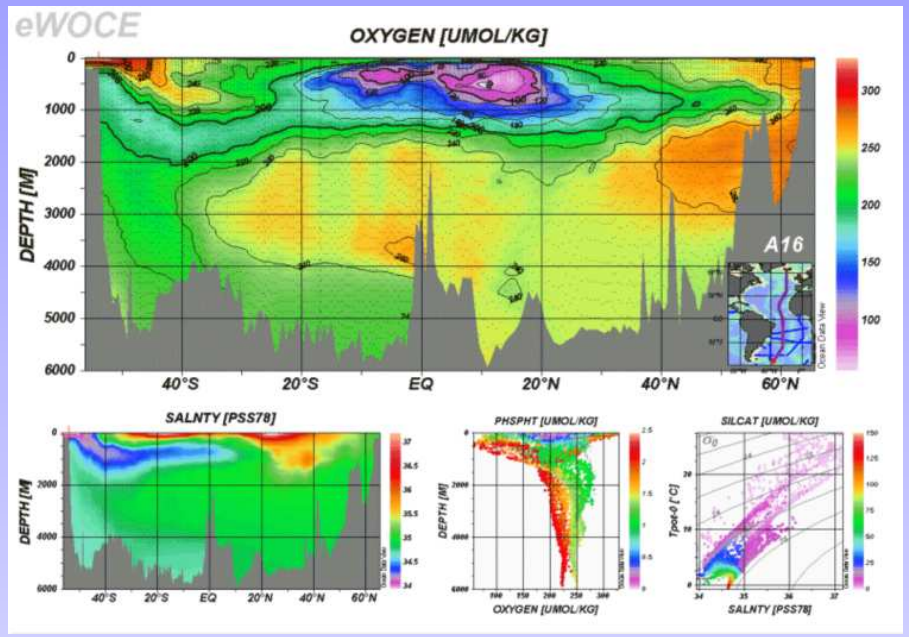


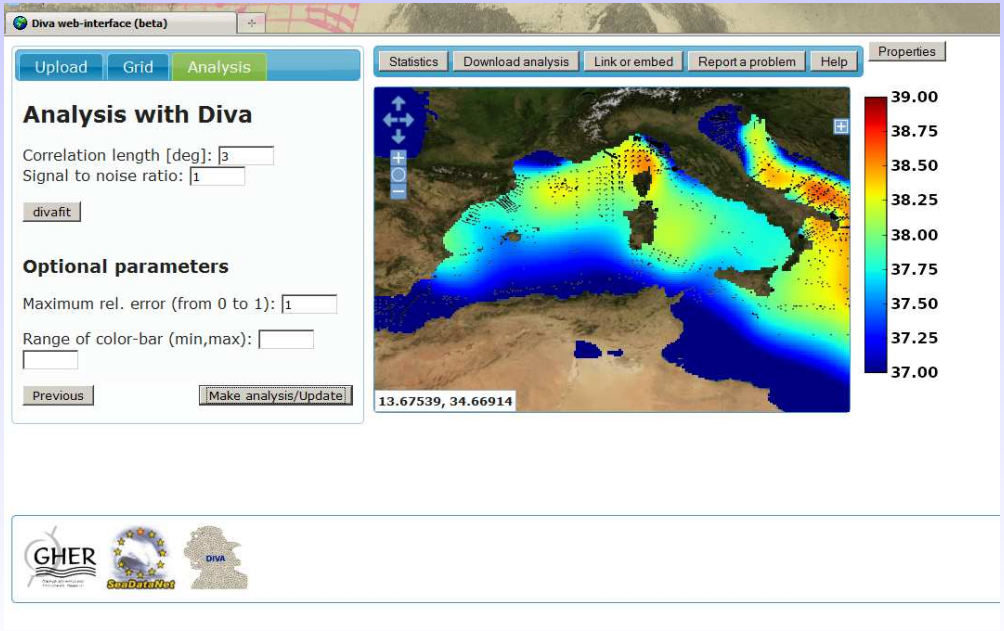


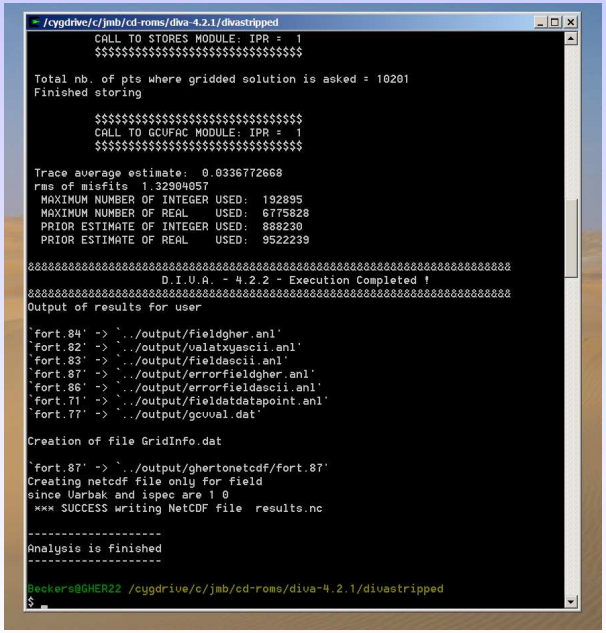
Université de Liège

How to use DIVA?

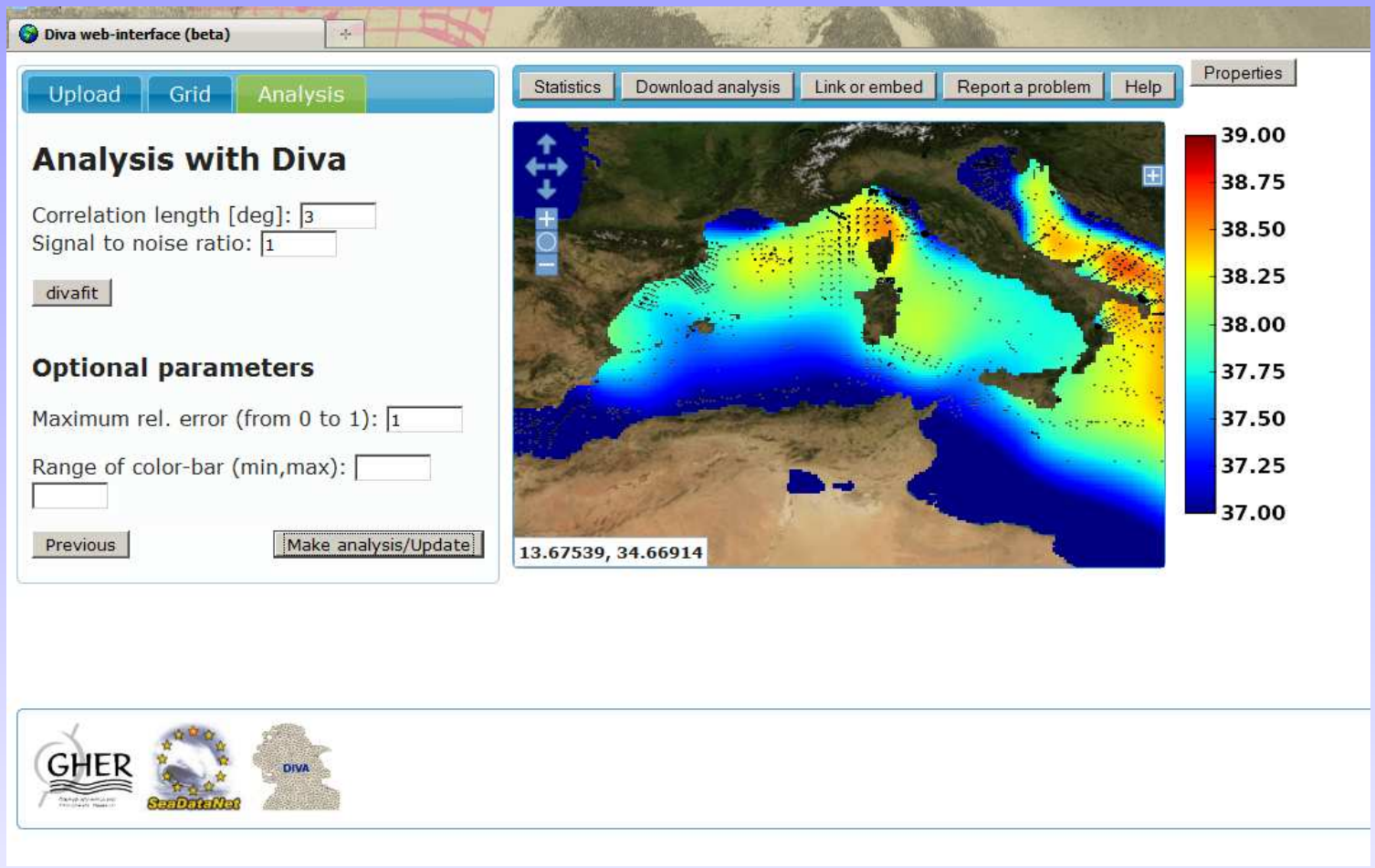








DIVA on WEB



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

For occasional use or quick data exploration

DIVA on WEB

- A first prototype of DIVA on WEB was developped (not aimed as a tool for climatology production but as an easy way to access the tools for occasional uses).
- User uploads its 2D data in simple ascii format (test version now with ONE ODV4 spreadsheet support).
- Based on OpenLayer, OGC-compliant using Phyton. Maps are generated internally by a request to a server located in Liege. The request itself is prepared by a Web interface. Results are shown in OpenLayers.
- Requests could come from other servers. (ICES and VLIZ expressed interest).
- Additional outputs as netCDF files, Matlab (or Octave) files and KML files (Google Earth).

For the moment only restrictions on CPU time and data quantity.

Université de Liège

- *Basics*
- ***When to use***
- *Critical points*
- *Examples*
- *Plans*

Comparison

Method	$\min(\epsilon^2)$	3D	Multivar	Ops/image	$\epsilon(\boldsymbol{r})$	a priori	C.V.	anisotropy
Cressman		$\star$	$\star$	$N_d N_a$		$w(r/L)$	(L)	$(\star)$
O.I.	$\star$	$\star$	$\star$	$N_d^3 + N_d N_a$	$\star$	$c(r/L)$	$L, \sigma^2/\mu^2$	$(\star)$
DIVA	$\star$	$(\star)$	$(\star)$	$N_e^{5/2}$	$\star$	$K(r/L)$	$L, \sigma^2/\mu^2$	$\star$
DINEOF	$(\star)$	$\star$	$\star$	$N_a^{5/4}$	$(\star)$	stat.	N	$\star$

N_d : number of data points

N_a : number of grid points for analysis

N_e : number of finite elements

N : number of EOFs

L : correlation length

σ^2/ϵ^2 : signal to noise ratio

$\star$: available feature

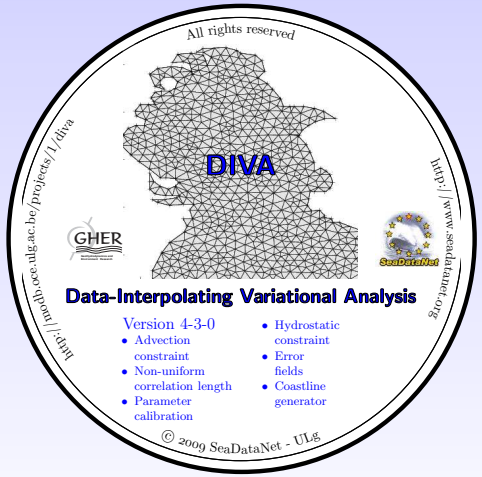
$(\star)$: available with some adaptations

Université de Liège

Université de Liège

Reference tool for SeaDataNet:

- Large data sets expected
- Error fields requested
- GUI, ODV-interface, WEB-interface (beginners) or scripts (expert mode)
- 3D (= stacking of 2D)
- G.C.V. for calibration of L and S/N.
- Land and underwater islands/obstacles taken into account (covariance changed).



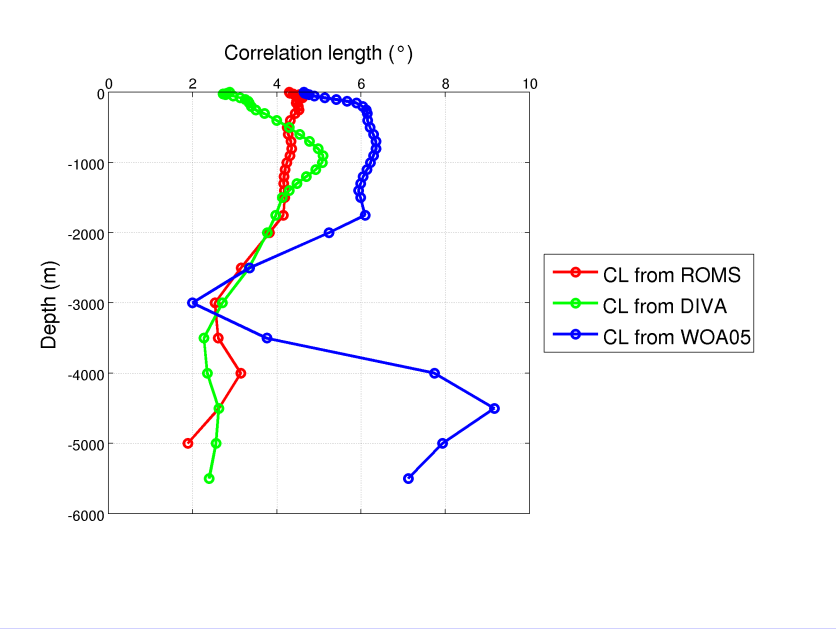
`http://modb.oce.ulg.ac.be/projects/1/diva`

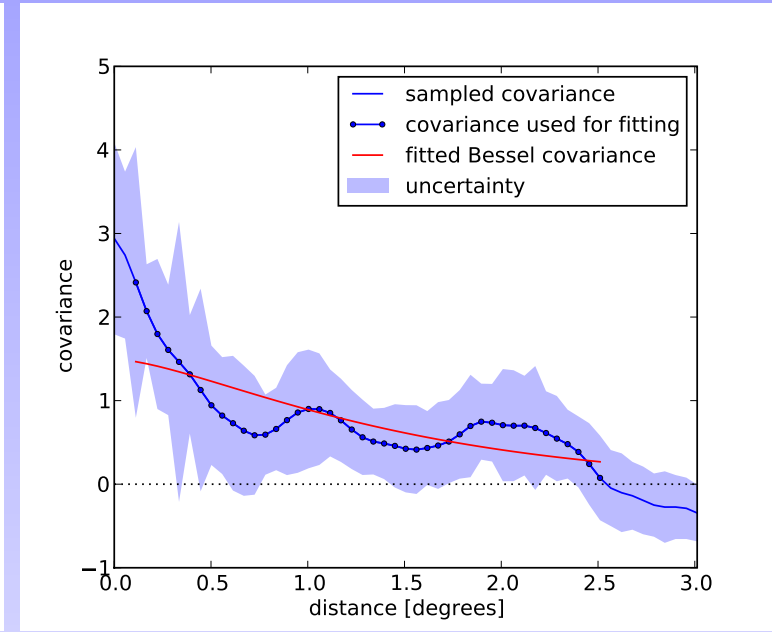
Université de Liège

- *Basics*
- *When to use*
- ***Critical points***
- *Examples*
- *Plans*

Université de Liège

Parameter calibration



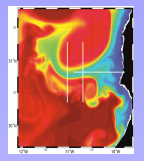


Spatial coherence of parameters: here correlation length obtained with covariance fitting (Troupin *et al.*, 2010).

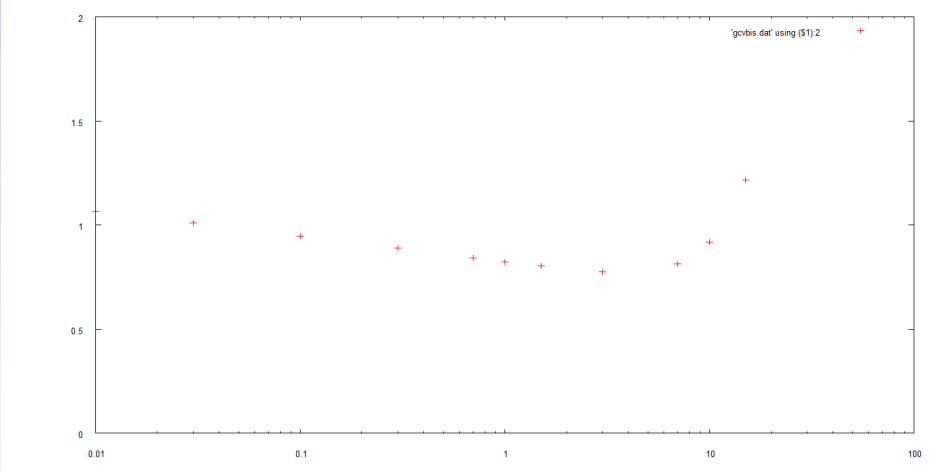
Université de Liège

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.

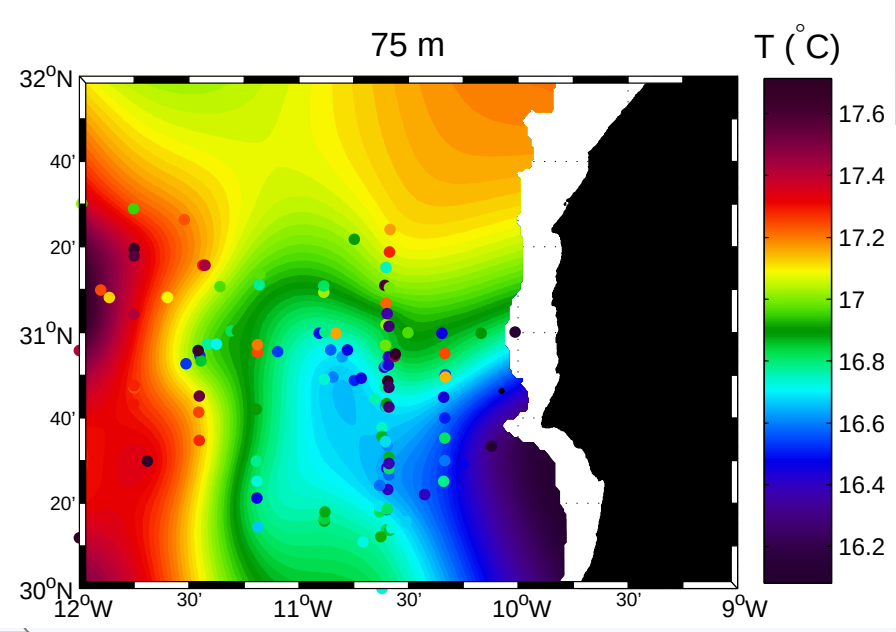


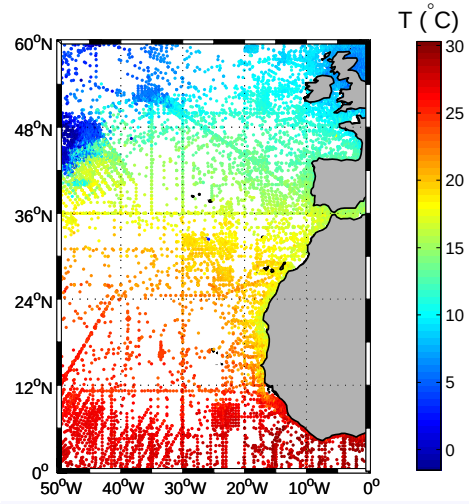
Université de Liège

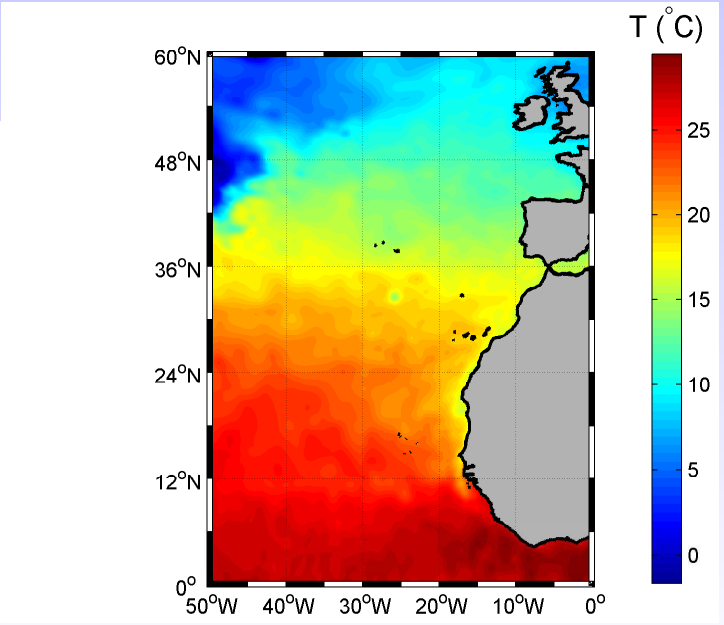
Data availability

THE problem.

- "Enough data" different for cruise snapshots or climatologies
- Recurrent discussion: if and how to mask analyses. Scientific approach (provide field and error estimate) vs large public approach (danger of misinterpretation or misuse).
- Validation and QC of products (independent data, other climatologies).





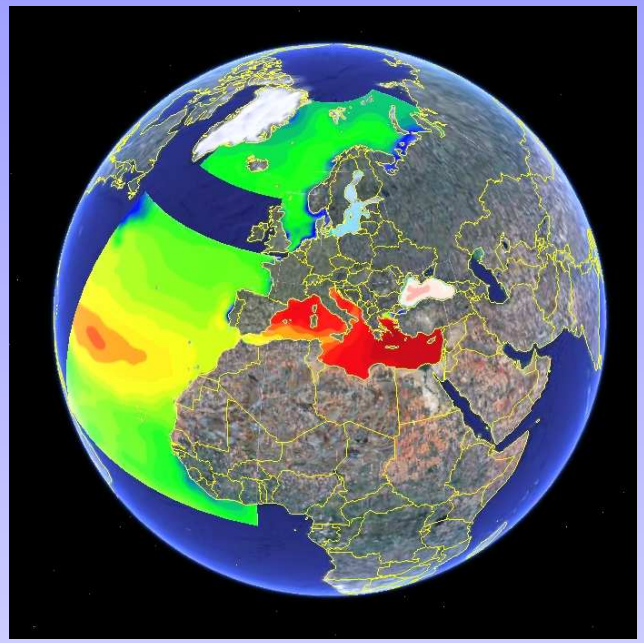


Université de Liège

- *Basics*
- *When to use*
- *Critical points*
- ***Examples***
- *Plans*

Université de Liège

Regional products within SeaDataNet

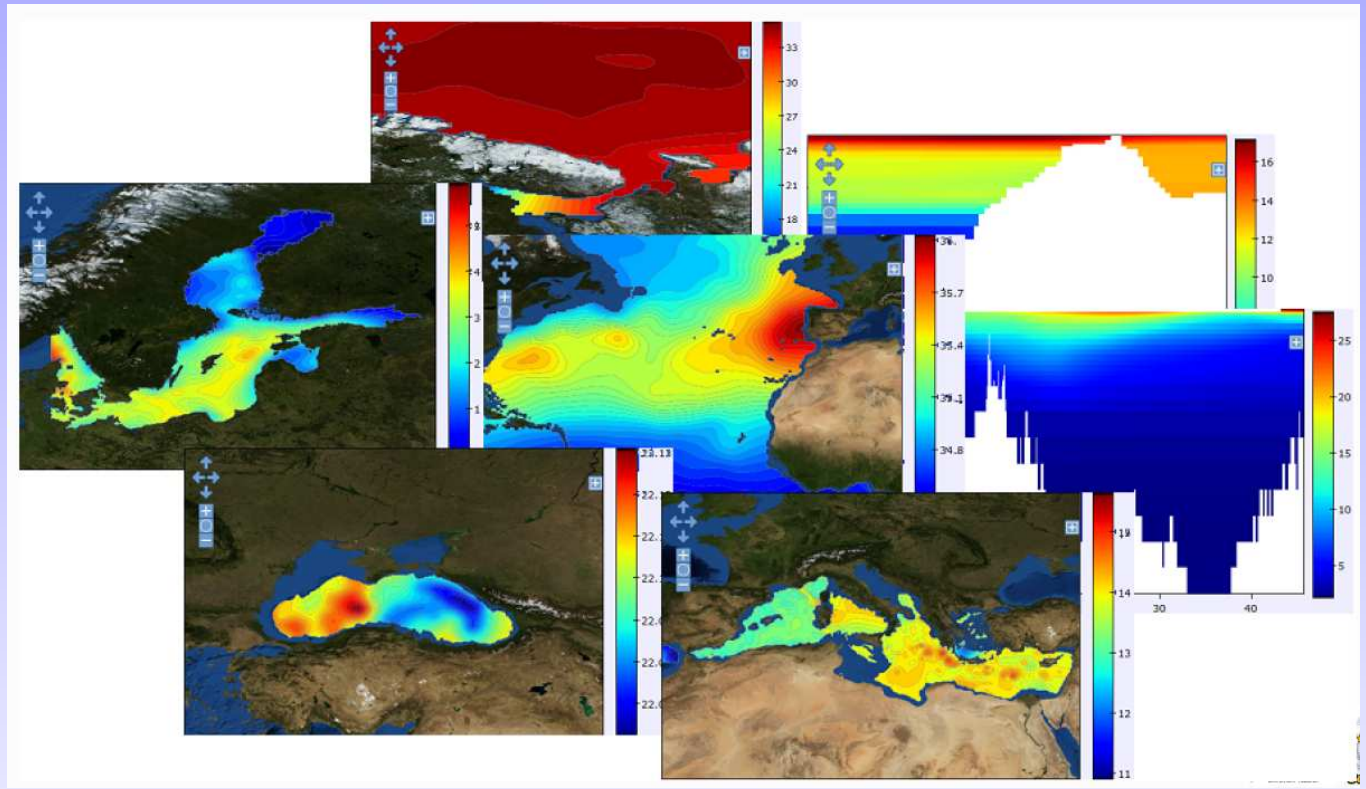


Installation of latest DIVA version and additional climatology-production tools.

Université de Liège

Climatology on WEB

Regional products uniformly and automatically prepared as 4D netCDF files. Hence possibility for common interface:

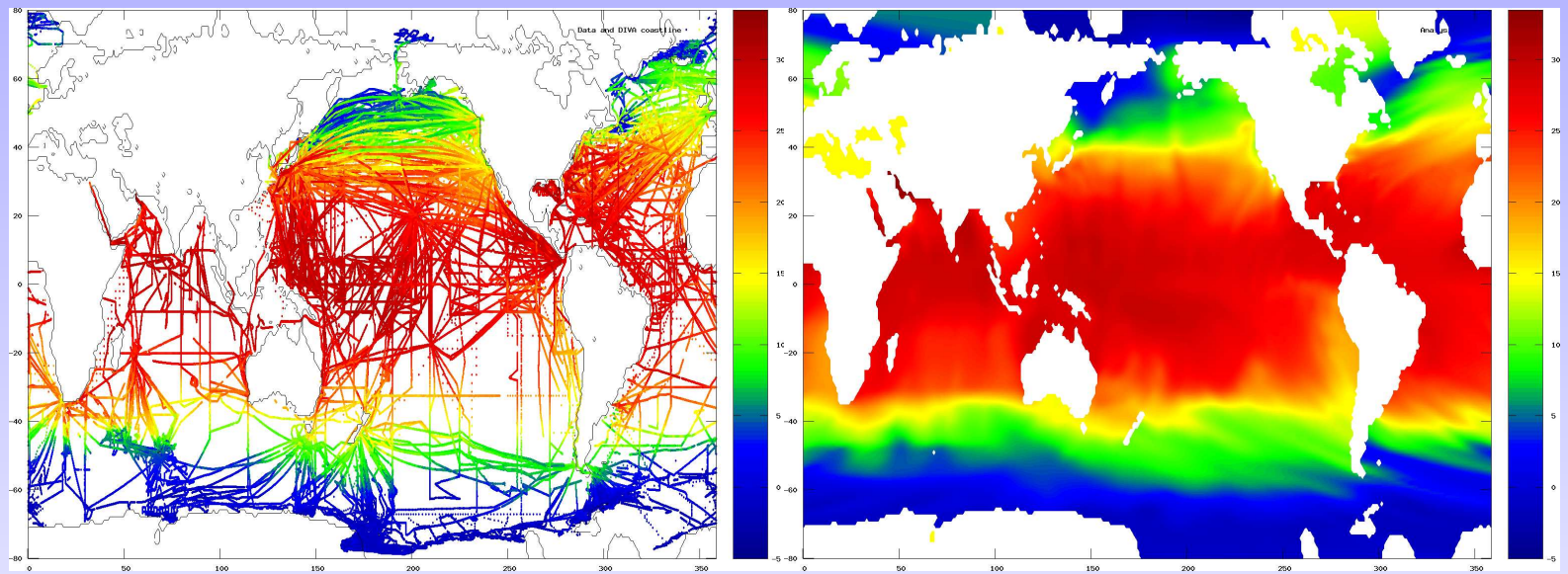


`http://gher-diva.phys.ulg.ac.be`

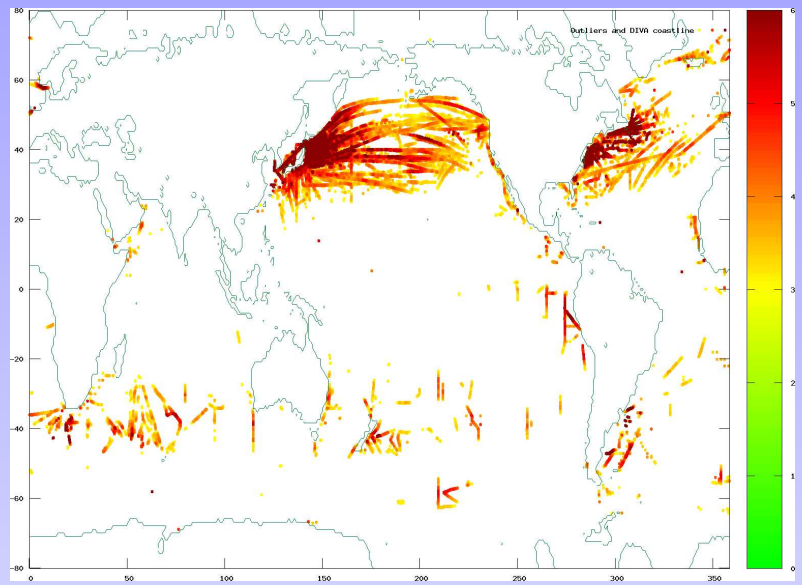
Includes vertical transect tool.

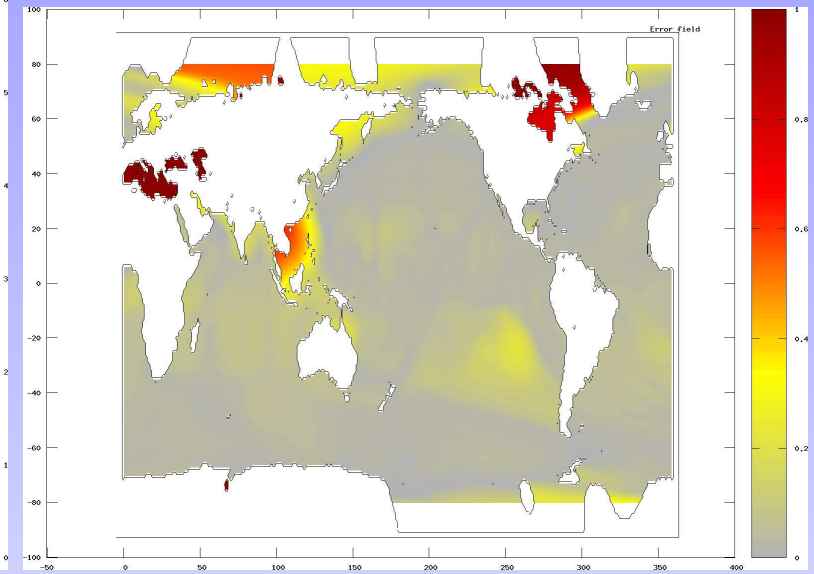
Huge problems

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.



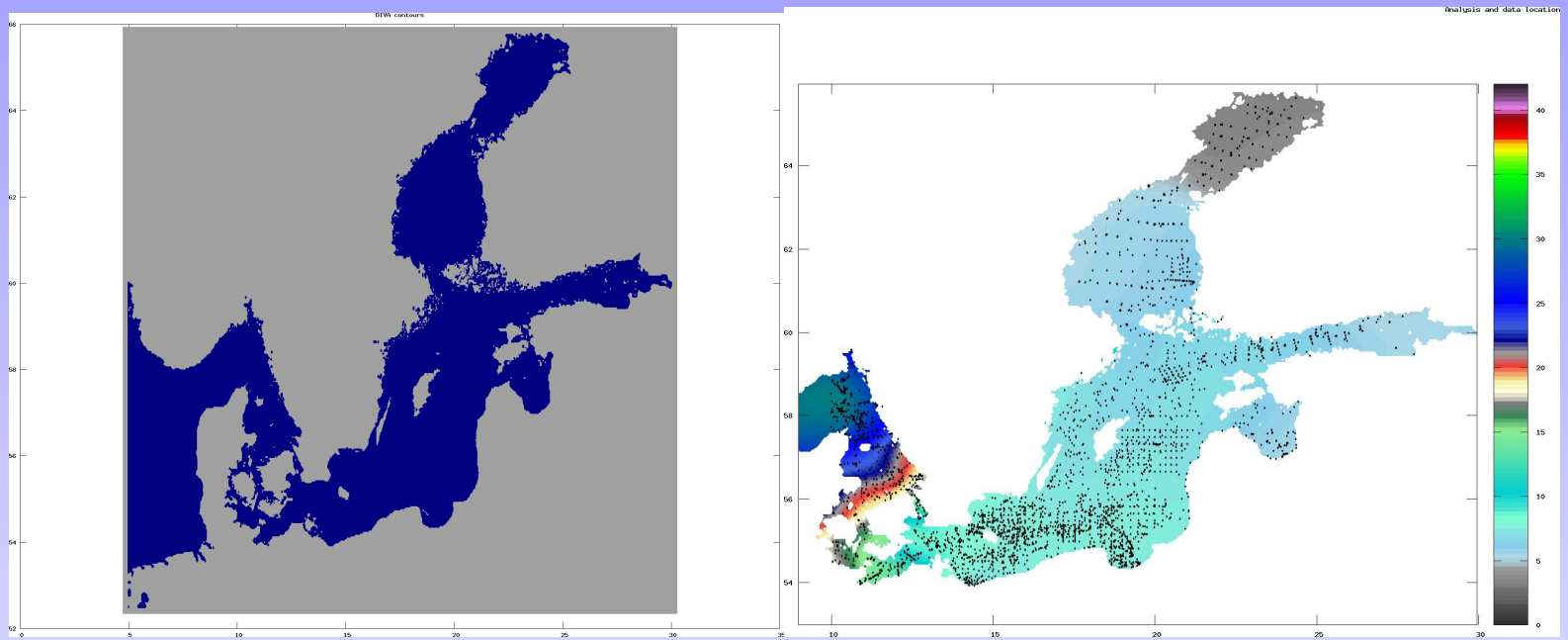
Huge problems, outliers and relative error field





Outliers detected via comparison of statistically expected residuals (value provided by the DIVA analysis) and actual residuals.

Heterogeneous case

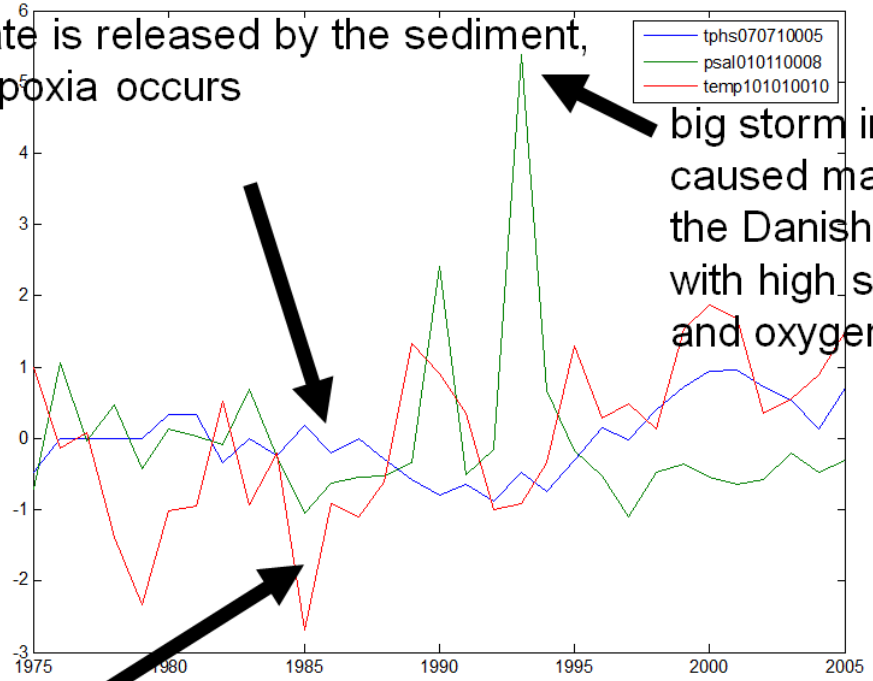


Baltic Salinity Climatology (Bassompierre *et al.*, 2010)

Université de Liège

Heterogeneous case, trends

X= nutrient & climatic trends from Kattegat to Bothnian Bay (SDN products)



phosphate is released by the sediment, when hypoxia occurs

big storm in January 1993, caused massive water through the Danish straits inflows with high salinity and oxygenated the Baltic

increased temperature increases risk of hypoxia

tphs0707005= total phosphate July -100 m
psal 0101108=salinity January level -40 m
temp101010010=temperature October -20 m

(Bassompierre *et al.*, 2010)

Université de Liège

- *Basics*
- *When to use*
- *Critical points*
- *Examples*
- ***Plans***

Computational aspects

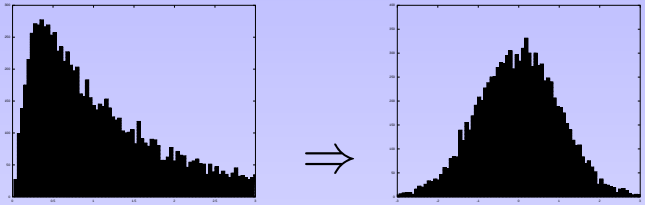
- Migration to F95 with dynamic memory allocation.
- Open source tools for large matrix calculations and parallel machines. (But best parallelization remains on parameters, seasons ...)
- Feeding of very large data arrays (10^7).
- Web visualization (both single DIVA analysis and 4D netCDF files).

In all cases, backward compatibility, portability and independence on proprietary software

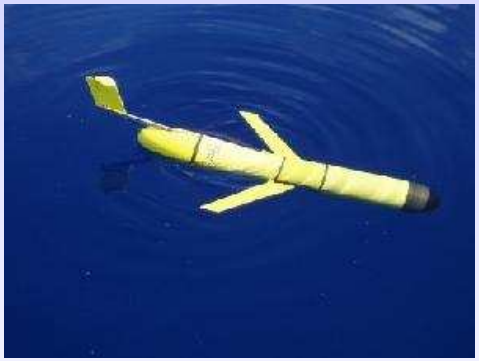
Université de Liège

Functional aspects

- Common problem in SeaDataNet and other projects: lack of biochemical data, strange statistical distributions.
  - ⋆ Multivariate approach with non-collocated data.
  - ⋆ Data transformation tools including anamorphosis.
  - ⋆ Absence-presence data for probability analysis.



- Adapting input possibilities.



- Investigate solutions to calibration with dependent data.

Very profound changes?

- N-Dimensional generalization. Easy in finite differences, awful in FEM, with need to rewrite completely the code.
- More complicated physical laws (source terms, vertical advection...).
- Diva-on-web to be used from within CDI interface for analyses on the fly ? (Including possibility to process restricted data without actually delivering them?).
- Rethinking of folder structure for better support of multiple users in a single installation.

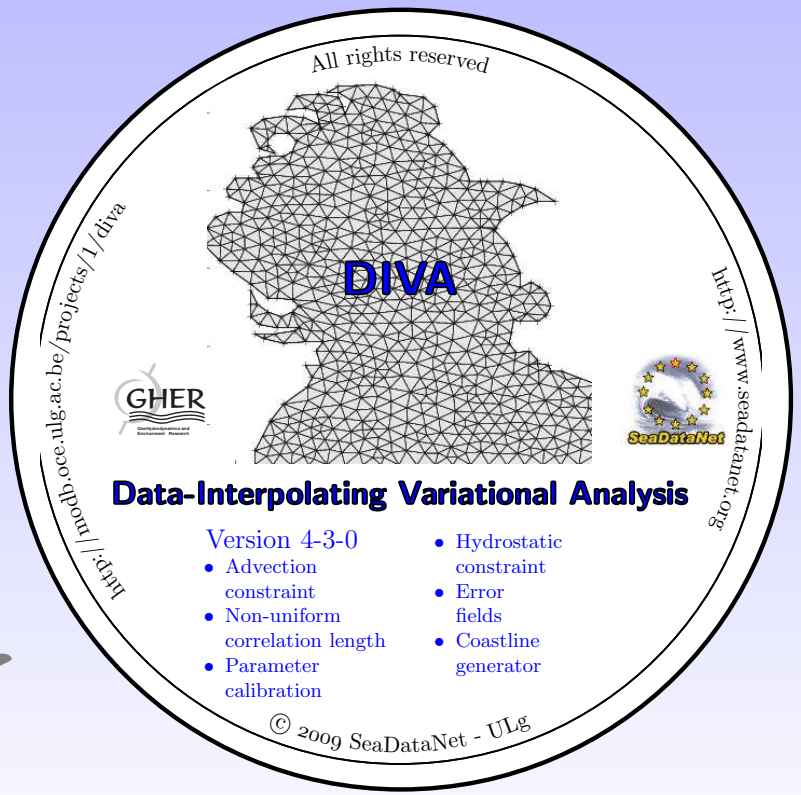
Some kind of conservatism here: robustness is priority (10^2 to 10^4 spatial analyses for a 4D product). Funding ?

Proud of

- SeaDataNet regional centers now well working with DIVA4D.
- Group that was most sceptical at Athens SDN annual meeting now the most happy user.
- DIVA development nicely user driven (annual workshop).
- ODV4 spreadsheet format gaining acceptance and easying up operations.
- DIVA incorporation into ODV very stable.
- Product preparation very effective at spotting outliers (visually or by automated a posteriori outlier detection) and testing data flow.

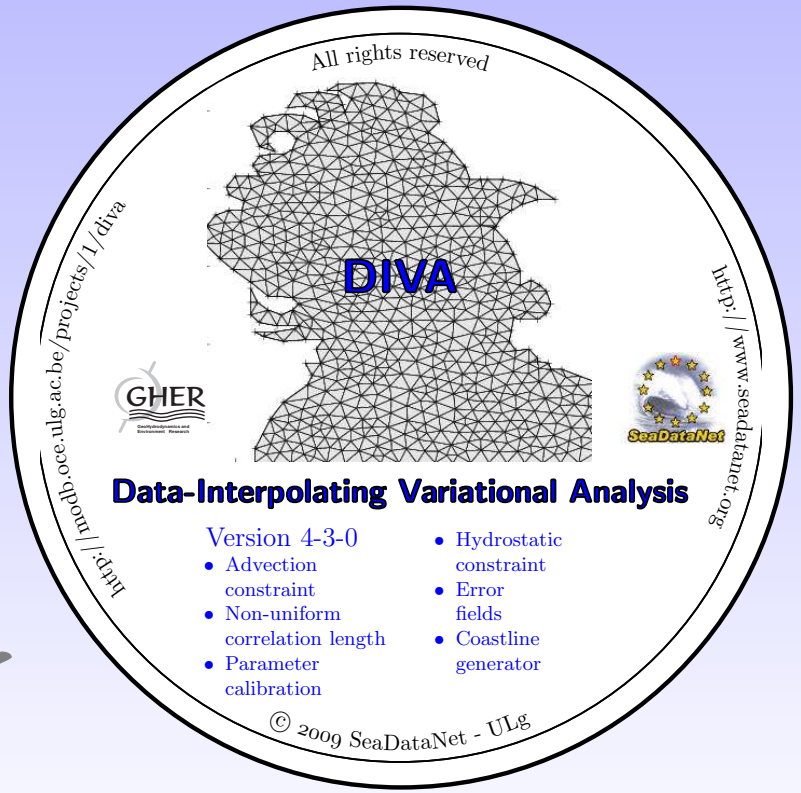
Acknowledgements

Special thanks to Helge Sagen, Marina Tonani, Sissy Iona, Cristina Tronconi, Jacob Carstensen, Marc Bassompierre, Serge Scory, Øivind Østensen, Pierre Brasseur, Jean-Michel Brankart and Michel Rixen for comments on DIVA evolution. Reiner Schlitzer designed and implemented the ODV coupling. SeaDataNet financed the continued development. The support from the Fonds pour la Formation la Recherche dans l'Industrie et dans l'Agriculture (FRIA) and National Fund for Scientific Research is also acknowledged.



Acknowledgements

Special thanks to Helge Sagen, Marina Tonani, Sissy Iona, Cristina Tronconi, Jacob Carstensen, Marc Bassompierre, Serge Scory, Øivind Østensen, Pierre Brasseur, Jean-Michel Brankart and Michel Rixen for comments on DIVA evolution. Reiner Schlitzer designed and implemented the ODV coupling. SeaDataNet financed the continued development. The support from the Fonds pour la Formation la Recherche dans l'Industrie et dans l'Agriculture (FRIA) and National Fund for Scientific Research is also acknowledged.



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.

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*, accepted, 2010.

J.-M. Beckers and M. Rixen. EOF calculations from incomplete oceanographic data sets. *Journal of Atmospheric and Ocean Technologies*, 20:1839–1856, 2003.

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.

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.

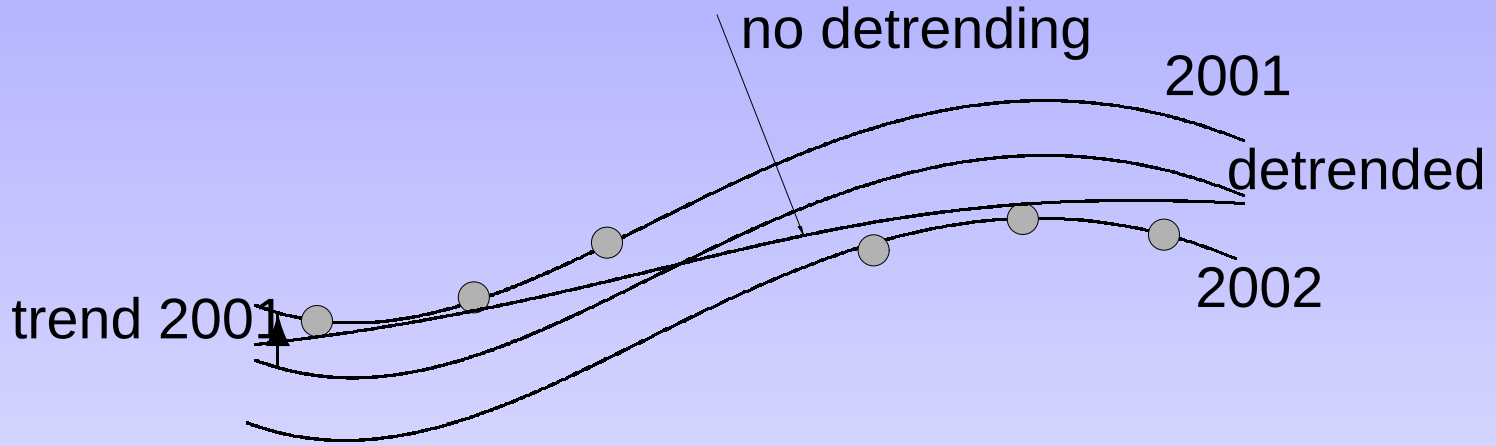
A. Alvera-Azcárate, A. Barth, M. Rixen, and J.-M. Beckers. Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic sea. *Ocean Modelling*, 9:325–346, 2005.

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.

Université de Liège

Detrending

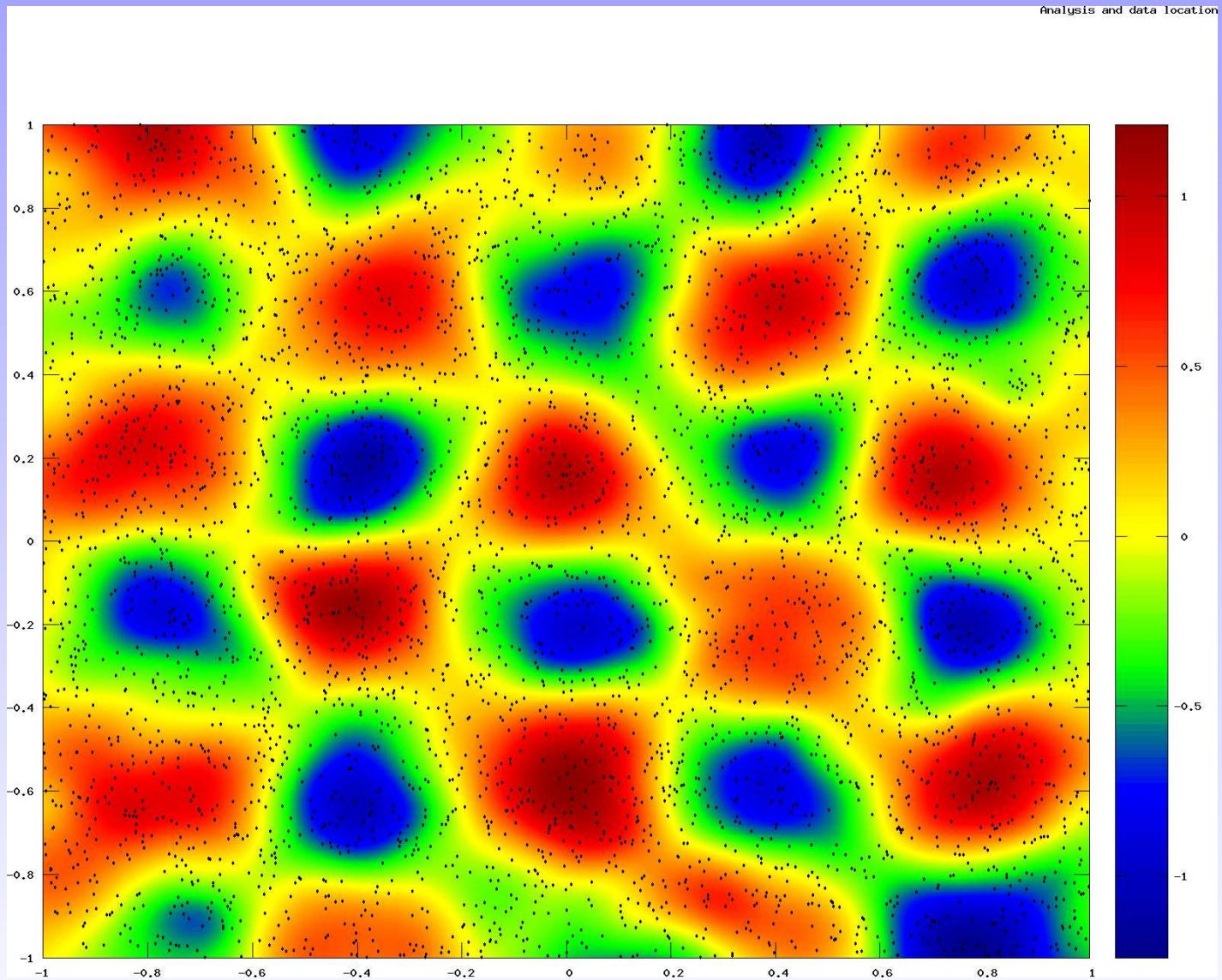
Heterogeneous data distribution:



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

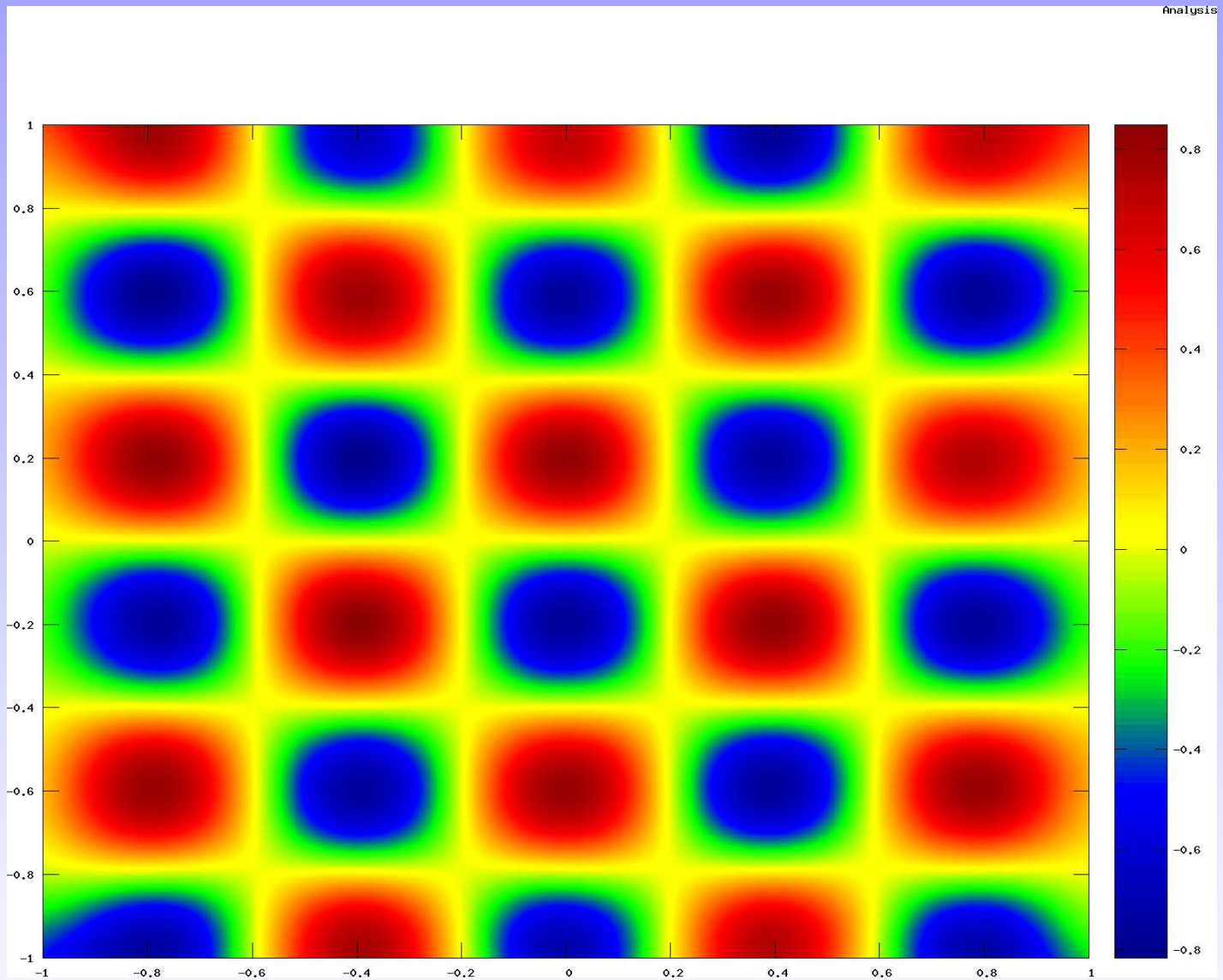
Université de Liège

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