# Diva workshop 2016 Diva in 2 dimensions

Alexander Barth, Aida Alvera-Azcárate, Mohamed Ouberdous, Charles Troupin, Sylvain Watelet & Jean-Marie Beckers

Acknowledgements: SeaDataNet, EMODnet Chemistry, EMODnet Biology, STARESO





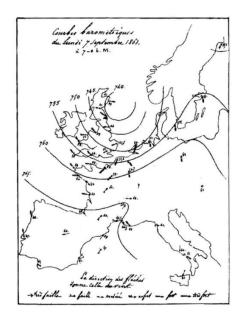








# Interpolation 150 years ago...





#### What is Diva?

Data Interpolating Variational Analysis



#### What is Diva?

- a method to produce gridded fields
- a set of bash scripts and Fortran programs

#### What is not Diva?

- a plotting tool
- a black-box
- a numerical model



#### Code development (1990-1996)

- Variational Inverse Method (VIM) (Brasseur, 1991, JMS, JGR)
- cross-validation (Brankart and Brasseur, 1996, JAOT)
- error computation (Brankart and Brasseur, 1998, JMS; Rixen et al., 2000, OM)



Code development (1990-1996) 2D-analysis (2006-2007)

- set of bash scripts (divamesh, divacalc,...)
- Fortran executables
- parameters optimization tools
- Matlab/Octave scripts for plotting



```
Code development (1990-1996)
2D-analysis (2006-2007)
3D-analysis (2007-2008)
```

- superposition of 2D layers
- automated treatment and optimization
- stability constraint (Ouberdous et al.)



```
Code development (1990-1996)

2D-analysis (2006-2007)

3D-analysis (2007-2008)

4D-analysis (2008-2009)

start from ODV spreadsheet

detrending (with J. Carstensen, DMU)
```

NetCDF 4-D climatology files



```
Code development (1990-1996)

2D-analysis (2006-2007)

3D-analysis (2007-2008)

4D-analysis (2008-2009)

Web tools

On-line analysis (Barth et al., 2010, Adv. Geosci.)

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

Climatology viewer: http:
//gher-diva.phys.ulg.ac.be/web-vis/clim.html
```



```
Code development (1990-1996)
2D-analysis (2006-2007)
3D-analysis (2007-2008)
4D-analysis (2008-2009)
Web tools
2011-2012
```

- multivariate approach
- data transformation tools
- 4-D graphical interface
- implementation of source/decay terms
- advanced error computation (Troupin et al., 2012, OM)



```
Code development (1990-1996)

2D-analysis (2006-2007)

3D-analysis (2007-2008)

4D-analysis (2008-2009)

Web tools

2011-2012

2013-2015
```

- Modernisation of the code structure
- n-dimensional generalisation
- optimized and approximate error calculations (clever poor man)



```
Code development (1990-1996)

2D-analysis (2006-2007)

3D-analysis (2007-2008)

4D-analysis (2008-2009)

Web tools

2011-2012

2013-2015

On-going:
```

- Analysis at a specific distance from the bottom
- Correlated observations errors (data weighting)
- ... **歩**



```
Code development (1990-1996)

2D-analysis (2006-2007)

3D-analysis (2007-2008)

4D-analysis (2008-2009)

Web tools

2011-2012

2013-2015

On-going:

General: user-driven developments
```



# Diva history

#### 4.6.7

Released in October 2014.

#### **New features**

■ Transformation of user relative length or advection fields files (ascii format) into the gher binary format, via a run of Diva (new script "asctobin")

#### New bug fixes

- Correction of time axis and climatology bounds in Netcdf output files (diva3Dwrt.F,diva4Dwrt.F,dv4DYRwrt.F,dv3DncYRw.F)
- Correction of some attributes in 4D netcdf (databins, snr, cl, varbak) (dv3DncYRw.F, diva3Dsub)
- Update of driver files (also in Example4D)

#### 4.6.6

Released in September 2014.

#### **New features**

- Check for severe errors in DIVA 3D/4D (script "godiva") + simple errors and warnings
- Possibility of binning the data before the parameters optimization (script "divabin" + program "binning\_lines.f90")
- Variable correlation length, depending on depth (script "divarlvardepth" + program "rlvardepth.f90")

#### New bug fixes

- Correction of the example in 4D (datasource)
- Correction of the script divaguessformODV4
- Exact match needed between variable name in "varlist" and its real name in the data file.

#### 4.6.5

Released in April 2014.

#### **New features**

#### New bug fixes

- "end of line" problems under Windows (file "datasource")
- Portability of scripts using the "sort" command
- http://modb.oce.ulg.ac.be/mediawiki/index.php/New\_Diva\_Features



#### Diva related tools

Diva: base tool (command line), 2D analysis

Godiva: automatic repetition of 2D analysis

Diva-on-web: 2D analysis with your data on our server

OceanBrowser: visualisation tool of 4D NetCDF files

divand: multi-dimension analysis (lon, lat, time, depth)

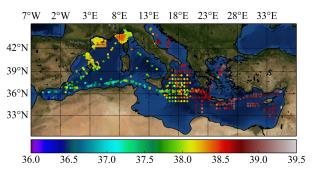
divaformatlab: wrapper to use in matlab

Clone-diva-x.x.x: virtual machine containing diva-x.x.x + other stuff

(gfortran, netcdf,...)



### Common problem



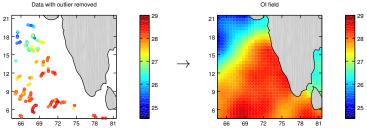
#### Appears when

- trying to produce maps
- calculate volume averages
- prepare initial conditions for models
- quality control of data
- **..**.



# The gridding problem

Gridding is the determination of a field  $\phi(r)$ , on regular grid of positions r based on arbitrarily located observations. Often the vector r is on a 2D, 3D or even 4D space.



- The fewer observations are available, the harder the gridding problem is
- In oceanography, in situ observations are sparse
- Observations are inhomogeneously distributed in space and time (more observations in the coastal zones and in summer)
- The variability of the ocean is the sum of various processes occurring at different spatial and temporal scales.



# The gridding problem

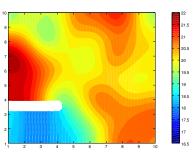


Figure 1: Example of oceanographic field.

- Figure 1 shows an idealized square domain with a barrier (*e.g.* a peninsula or a dike).
- This field is the true field that we want to reconstruct based on observations. Let's assume that the field represents temperature.
- The barrier suppresses the exchanges between each side of the barrier.
- The field varies smoothly over some length-scale

# Sampling locations

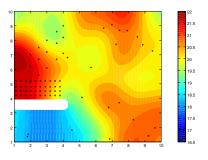


Figure 2: Sampling locations within the domain

- In regions where a measurement campaign has been carried out, a higher spatial coverage is achieved.
- Large gaps are also present.
- Based on the value of the field at the shown location, we will estimate the true field.

# True field at the sampling locations

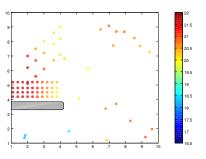


Figure 3: Value of the true field extract at the location of the observations.

- only the value of the observations is shown
- some information about the position of the structures and fronts is lost
- no method can provide exactly the true field.
- the more information about its structure and evolution we include in the analysis, to close we can get to the true field.



#### Observation errors

Observations are in general affected by different error sources and other "problems" that need to be taken into account:

- Instrumental errors (limited precision or possible bias of the sensor)
- 2 Representative errors: the observations do not necessarily corresponds to the field we want to obtain. For example, we want to have a monthly average, but the observations are instantaneous (or averages over a very short period of time).
- 3 Synopticity errors: all observations are not taken at the same time.
- 4 Other errors sources: human errors (e.g. permutation of longitude and latitude), transmission errors, malfunctioning of the instrument, wrong decimal separators...

Quality control is an important step to exclude suspicious data from the analysis. But since this is a subjective decision, the data should never be deleted but flagged as suspicious or bad data.

#### Observation errors

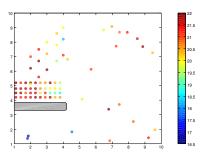
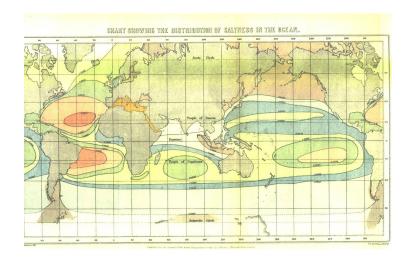


Figure 4: Observation with errors

In figure 4, a random perturbation was added to the observation shown in figure 3. This simulates the impact of the different error sources. To simplify matters, each observation was perturbed independently.



# Solutions - Subjective methods





### Solutions - Subjective methods

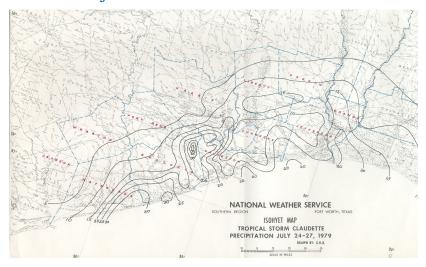


Figure 5: Isohyet (lines of constant precipitation) drawn by hand (from http://www.srh.noaa.gov/hgx/hurricanes/1970s.htm)



#### Interpolation or Analysis?

Because observations have errors, it is always better to produce a field approximation and never a strict interpolation.

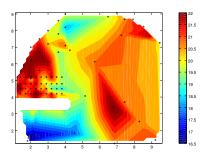


Figure 6: Gridded field using linear interpolation. This method is implemented in the function griddata of Matlab and GNU Octave.

- Figure 6 shows what would happen if the observations would have been interpolated linearly.
- The domain is decomposed into triangles where the vertices are the location of the data points based on the Delaunay triangulation.
- Within each triangle, the value is interpolated linearly.

# Solutions - Objective methods

- Subjective method is not sufficiently ... objective.
- Data Assimilation: region and model dependent.
- $\Rightarrow$  Objective analysis of data that are anomalies with respect to a background field  $\varphi_b(\mathbf{r})$ .

As opposed to the subjective method, objective analysis techniques aim to use mathematical formulations to infer the value of the field at unobserved locations based on the observation  $d_j$ . Most objective methods can be expressed as a linear combination of data anomalies  $d_j$  using weights  $w_j$ :

$$\varphi(\mathbf{r}) = \varphi_b(\mathbf{r}) + \sum_{j=1}^{N_d} w_j d_j$$
 (1)

The field  $\varphi(r)$  can be evaluated in any position r, hence gridding is possible. The background field (or first guess)  $\varphi_b$  is defined a priori and anomalies calculated with respect to this reference field (for example a climatological average). There are several ways to define the weighting function  $w_j$ , which result in different gridding techniques.

#### Cressman method

Cressman weights depend only on the distance r between the location r where the value of the field should be estimated and the location of the observation  $r_j$ :

$$r = |\boldsymbol{r} - \boldsymbol{r}_i| \tag{2}$$

The weights are then parameterized according to,

$$\tilde{w}(r) = \frac{R^2 - r^2}{R^2 + r^2} \quad \text{for} \quad r < R$$

$$= 0 \quad \text{for} \quad r \ge R$$
(3)

The weights as a function of distance are shown in figure 7. Weights must be scaled by their sum to ensure no bias.

$$w_j = \tilde{w}_j / \sum_j \tilde{w}_j \tag{4}$$



#### Cressman method

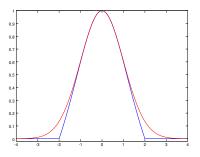


Figure 7: Cressman weights for R = 2 (blue) and Barnes weights for R = 1 (red).

The search radius R is the typical control parameter and defines the length-scale over which an observation is used. This length scale can be made to vary in space depending on data coverage and/or physical scales. This parameter is chosen by the users based on their knowledge of the domain and the problem.

#### Cressman method

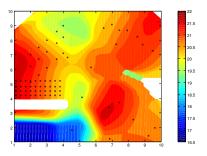


Figure 8: Gridded field by Cressman weighting

The Cressman weighting is a very simple and numerically quite efficient method. However, it suffers from some limitations which are apparent in figure 8.

- No estimate can be obtained at locations when no observation is located within the R.
- In regions with very few observations, the method can return a discontinuous field.
- The presence of barriers cannot be taken into account easily.
- All observations are assumed to have a similar error variance since the weighting is based only on distance.



#### Barnes method

As a variant of the Cressman weights, other weighting functions can be defined. In the Barnes scheme, the weights are defined using a Gaussian function:

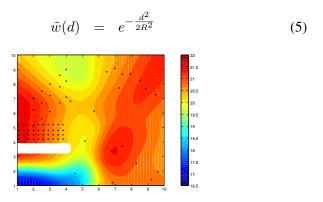
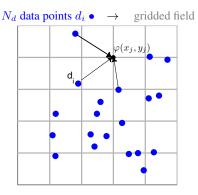


Figure 9: Gridded field using Barnes weights

Since the Barnes weights are never zero, in principle all observations are used for the gridding. An estimation can be obtained everywhere (which can be accurate or not). Artificial discontinuities are avoided using the Barnes weights (figure 9).

# DIVA: Data-Interpolating Variational Analysis



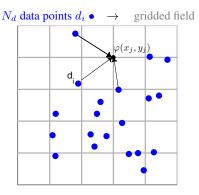
Formulation: minimize cost function  $J[\varphi]$ 

$$\min J[\varphi] = \sum_{i=1}^{N} \mu_i \left[ d_i - \varphi(x_i, y_i) \right]^2$$

$$+ \int_{D} \left( \nabla \nabla \varphi : \nabla \nabla \varphi + \alpha_1 \nabla \varphi \cdot \nabla \varphi + \alpha_0 \varphi^2 \right) dD$$



# DIVA: Data-Interpolating Variational Analysis



Formulation: minimize cost function  $J[\varphi]$ 

$$\begin{split} \min J[\varphi] &= \sum_{i=1}^N \mu_i \left[ d_i - \varphi(x_i, y_i) \right]^2 \qquad \text{data--analysis misfit} \\ &+ \quad \int_D \left( \boldsymbol{\nabla} \boldsymbol{\nabla} \varphi : \boldsymbol{\nabla} \boldsymbol{\nabla} \varphi + \alpha_1 \boldsymbol{\nabla} \varphi \cdot \boldsymbol{\nabla} \varphi + \alpha_0 \varphi^2 \right) \mathrm{d}D \qquad \text{field regularity} \end{split}$$

Non-dimensional version:

$$L = \text{length scale} \rightarrow \tilde{\nabla} = L\nabla$$

$$\rightarrow D = L^2 \tilde{D}$$
(6)

$$\rightarrow D = L^2 D \tag{7}$$

Non-dimensional version:

$$L = \text{length scale} \quad \to \quad \tilde{\nabla} = L\nabla \tag{6}$$

$$\rightarrow D = L^2 \tilde{D} \tag{7}$$

$$\tilde{J}[\varphi] = \sum_{i=1}^{N} \mu_{i} L^{2} [d_{i} - \varphi(x_{i}, y_{i})]^{2}$$

$$+ \int_{\tilde{D}} (\tilde{\nabla} \tilde{\nabla} \varphi : \tilde{\nabla} \tilde{\nabla} \varphi + \alpha_{1} L^{2} \tilde{\nabla} \varphi \cdot \tilde{\nabla} \varphi + \alpha_{0} L^{4} \varphi^{2}) d\tilde{D}$$



Non-dimensional version:

$$L = \text{length scale} \quad \to \quad \tilde{\nabla} = L\nabla \tag{6}$$

$$\rightarrow D = L^2 \tilde{D} \tag{7}$$

$$\tilde{J}[\varphi] = \sum_{i=1}^{N} \mu_{i} L^{2} [d_{i} - \varphi(x_{i}, y_{i})]^{2}$$

$$+ \int_{\tilde{D}} (\tilde{\nabla} \tilde{\nabla} \varphi : \tilde{\nabla} \tilde{\nabla} \varphi + \alpha_{1} L^{2} \tilde{\nabla} \varphi \cdot \tilde{\nabla} \varphi + \alpha_{0} L^{4} \varphi^{2}) d\tilde{D}$$

•  $\alpha_0 \to L$  for which data-analysis misfit  $\simeq$  regularity term:

$$\alpha_0 L^4 = 1$$



Non-dimensional version:

$$L = \text{length scale} \quad \to \quad \tilde{\nabla} = L\nabla \tag{6}$$

$$\rightarrow D = L^2 \tilde{D} \tag{7}$$

$$\tilde{J}[\varphi] = \sum_{i=1}^{N} \mu_{i} L^{2} [d_{i} - \varphi(x_{i}, y_{i})]^{2} 
+ \int_{\tilde{D}} (\tilde{\nabla} \tilde{\nabla} \varphi : \tilde{\nabla} \tilde{\nabla} \varphi + \alpha_{1} L^{2} \tilde{\nabla} \varphi \cdot \tilde{\nabla} \varphi + \alpha_{0} L^{4} \varphi^{2}) d\tilde{D}$$

- $\alpha_0 \to L$  for which data-analysis misfit  $\simeq$  regularity term:
- $\alpha_0 L^4 = 1$

 $\alpha_1 \rightarrow \text{influence of gradients:}$ 

$$\alpha_1 L^2 = 2\xi, \qquad \xi = 1$$



Non-dimensional version:

$$L = \text{length scale} \quad \to \quad \tilde{\nabla} = L\nabla \tag{6}$$

$$\rightarrow D = L^2 \tilde{D} \tag{7}$$

$$\tilde{J}[\varphi] = \sum_{i=1}^{N} \mu_{i} L^{2} [d_{i} - \varphi(x_{i}, y_{i})]^{2}$$

$$+ \int_{\tilde{D}} (\tilde{\nabla} \tilde{\nabla} \varphi : \tilde{\nabla} \tilde{\nabla} \varphi + \alpha_{1} L^{2} \tilde{\nabla} \varphi \cdot \tilde{\nabla} \varphi + \alpha_{0} L^{4} \varphi^{2}) d\tilde{D}$$

- $\alpha_0 \to L$  for which data-analysis misfit  $\simeq$  regularity term:
- $\alpha_0 L^4 = 1$

 $\bullet$   $\alpha_1 \rightarrow$  influence of gradients:

$$\alpha_1 L^2 = 2\xi, \qquad \xi = 1$$

 $\mu_i L^2 \rightarrow$  weight on data:

$$\mu_i L^2 = 4\pi \frac{\text{signal}}{\text{noise}_i}$$



# Analysis parameters are related to data

Non-dimensional version:

$$L = \text{length scale} \quad \to \quad \tilde{\nabla} = L\nabla \tag{6}$$

$$\rightarrow D = L^2 \tilde{D} \tag{7}$$

$$\tilde{J}[\varphi] = \sum_{i=1}^{N} \mu_{i} L^{2} [d_{i} - \varphi(x_{i}, y_{i})]^{2}$$

$$+ \int_{\tilde{D}} (\tilde{\nabla} \tilde{\nabla} \varphi : \tilde{\nabla} \tilde{\nabla} \varphi + \alpha_{1} L^{2} \tilde{\nabla} \varphi \cdot \tilde{\nabla} \varphi + \alpha_{0} L^{4} \varphi^{2}) d\tilde{D}$$

Coefficients  $\alpha_0$ ,  $\alpha_1$  and  $\mu_i$  related to

- $\blacksquare$  Correlation length L
- 2 Signal-to-noise  $\lambda$
- 3 Observational noise standard deviation  $\epsilon_i^2$



# Main analysis parameters

### Correlation length *L*:

- Measure of the *influence* of data points
- Estimated by a least-square fit of the covariance function

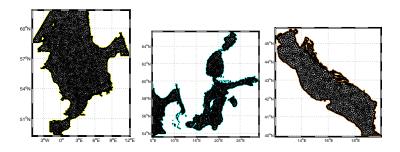
### Signal-to-noise ratio $\lambda$ :

- Measure of the *confidence* in data
- Estimated with Generalized Cross Validation techniques



#### Minimization with a finite-element method

Field regularity  $\rightarrow$  plate bending problem  $\rightarrow$  finite-element solver



#### Advantages:

- boundaries taken into account
- numerical cost (almost independent on data number)
- no *a posteriori* masking (except if based on error level)



### Minimization with a finite-element method

Triangular FE only covers sea: 
$$J[\varphi] = \sum_{e=1}^{N_e} J_e(\varphi_e)$$
 (8)

In each element: 
$$\varphi_e(\mathbf{r_e}) = \mathbf{q_e}^T \mathbf{s}(\mathbf{r_e})$$
 with 
$$\begin{cases} \mathbf{s} & \to \text{ shape functions} \\ \mathbf{q} & \to \text{ connectors} \\ \mathbf{r_e} & \to \text{ position} \end{cases}$$
 (9)

(9) in (8) + variational principle

$$J_e(\mathbf{q_e}) = \mathbf{q_e}^T \mathbf{K_e} \mathbf{q_e} - 2\mathbf{q_e}^T \mathbf{g_e} + \sum_{i=1}^{N_{d_e}} \mu_i d_i$$
 (10)

where 
$$\left\{ \begin{array}{ll} \mathbf{K_e} & \rightarrow \text{local stiffness matrix} \\ \mathbf{g} & \rightarrow \text{vector depending on local data} \end{array} \right.$$



### Minimization with a finite-element method

On the whole domain: 
$$J(\mathbf{q}) = \mathbf{q}^T \mathbf{K} \mathbf{q} - 2 \mathbf{q}^T \mathbf{g} + \sum_{i=1}^{N_d} \mu_i d_i$$
 (8)

Minimum:  $\mathbf{q} = \mathbf{K}^{-1}\mathbf{g}$  (9)

$$\mathbf{q} = \mathbf{K}^{-1} \mathbf{g} \tag{10}$$

- Connectors (new unknowns) -
- Stiffness matrix —
- Charge vector —

Mapping of data on FEM 
$$\rightarrow$$
 transfer operator  $\mathbf{T_2} \rightarrow$   $\mathbf{g} = \mathbf{T_2}(\mathbf{r})\mathbf{d}$   
Solution at any location  $\rightarrow$  transfer operator  $\mathbf{T_1} \rightarrow$   $\boldsymbol{\varphi}(\mathbf{r}) = \mathbf{T_1}(\mathbf{r})\mathbf{q}$ 

# Diva Cocktail Recipe

#### Ingredients:

- 1 1/2 oz vodka
- 1/2 oz passion-fruit juice
- 1/2 oz lime juice
- 1 tbsp cherry juice
- fill with soda





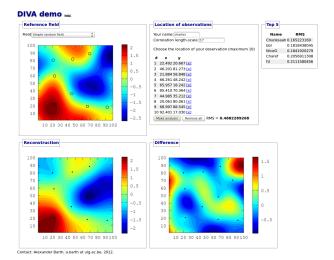
# Diva Cocktail Recipe





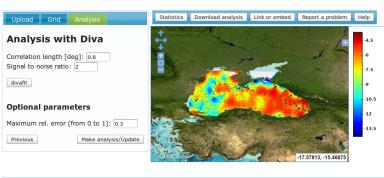
#### Want to use Diva?

### Playing...



#### Want to use Diva?

With your own data...





#### Want to use Diva?

#### For serious work:

2D version (for production), open source, GPL nD version (for research), open source, GPL

```
File Edit View Search Terminal Help
             D.I.V.A. - 4.6.1 - Execution track ...
       ***********
      CALL TO MATHPR MODULE: IPR = 1
      There are 1 data localized in the mesh (and resorted)
      CALL TO SOLVER MODULE: IPR = 1
```

# Running Diva in 2D: input files

1 data.dat: contains the observations

xlylvalue

```
36.5500 45.163 17.7138
33.7500 44.167 18.135
32.7500 44.167 18.51
36.2500 43.833 18.5892
33.2500 45.083 18.2326
32.7833 43.917 18.477
32.7500 43.500 18.59
37.2433 44.833 18.1555
36.5000 44.000 18.19
35.8333 43.750 18.62
34.2500 43.832 18.29
35.6500 44.000 18.75
38.0000 44.000 18.155
37.8200 44.368 17.1916
39.0000 42.500 18.23
33.1333 44.433 18.001
33.0500 44.433 18.09
33.2500 44.167 18.231
32.5333 44.833 18.014
38.0167 44.447 18.0568
```

# Running Diva in 2D: input files

1 data.dat: contains the observations

xlylvalue (coastline or isobaths)

2 coast.cont: delimits land and sea

```
П
   27.4375000
                     40.3499985
   27.4500008
                     40.3375015
   27.4666672
                     40.3375015
   27.4833336
                     40.3375015
   27.5000000
                     40.3375015
   27.5166664
                     40.3375015
   27.5333328
                     40.3375015
   27.5499992
                     40.3375015
   27.5666676
                     40.3375015
   27.5791664
                     40.3499985
   27.5833340
                     40.3541679
   27.6000004
                     40.3541679
   27.6124992
                     40.3666649
   27.6166668
                     40.3708344
                     40.3833351
   27.6291676
   27.6291676
                     40.4000015
   27.6291676
                     40.4166679
   27.6291676
                     40.4333344
   27.6166668
                     40.4458351
   27.6124992
                     40.4500008
```

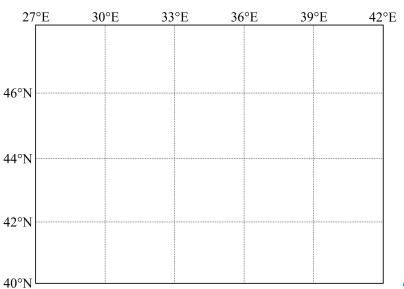
## Running Diva in 2D: input files

- 1 data.dat: contains the observations
- 2 coast.cont: delimits land and sea
- 3 param.par: analysis parameters

x|y|value (coastline or isobaths)

 $L, \lambda$ , resolution, ...

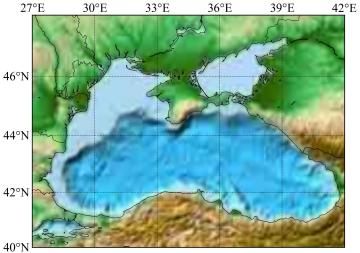
### Select region of study





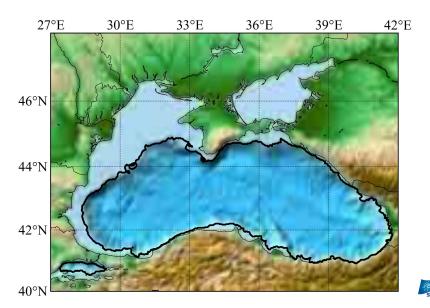
### Extract topography, for example via

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

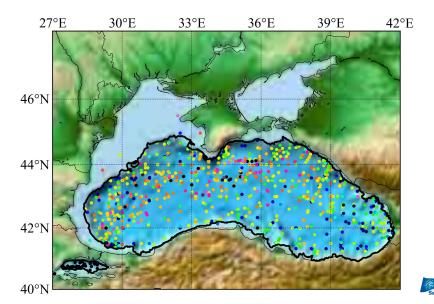




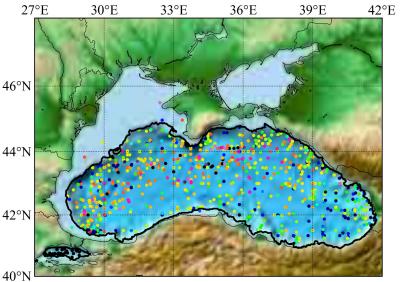
#### Generate contour



### Extract data

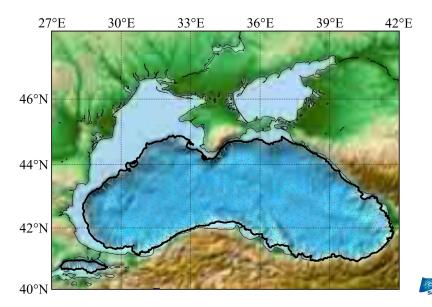


### Evaluate analysis parameters

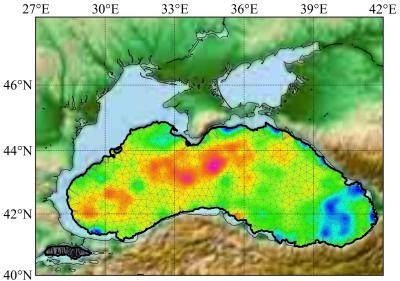




#### Create finite-elementmesh

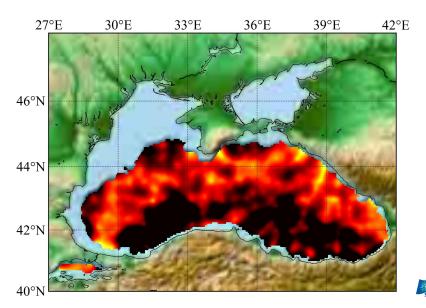


### Generate analysis





#### Generate error field



#### When to use 2D version

- occasional use
- 2D fields like benthic properties
- for implementation of special features by your own (eg multiplicative bias correction, special background field creation based on habitats
- ..

otherwise: use 3D or 4D version directly



Next...

Diva in 4 dimensions

