



# EMODNET Chemical lot

## - guidelines DIVA maps production

May 2010

Written by Anders Windelin and Marc Bassompierre (NERI)

Comments by Alexey Khaliulin (MHI), Sissy Iona (HCMR), Dick Schaap (MARIS) and Alexander Barth (UIG)

### Guidelines DIVA maps production

The aim for the guidelines for the DIVA maps production in relation to the EMODNET pilot project is to establish a common basis for the maps production with focus on the pre-processing of data and the validation of the produced maps.

The guidelines for the DIVA maps production is divided into four steps:

- 1) Data extraction
- 2) Pre-processing of data
- 3) Maps production
- 4) Validation of the produced maps

Figure 1 describes the overall concept for the DIVA maps production and is recommend as the working process for the maps production.

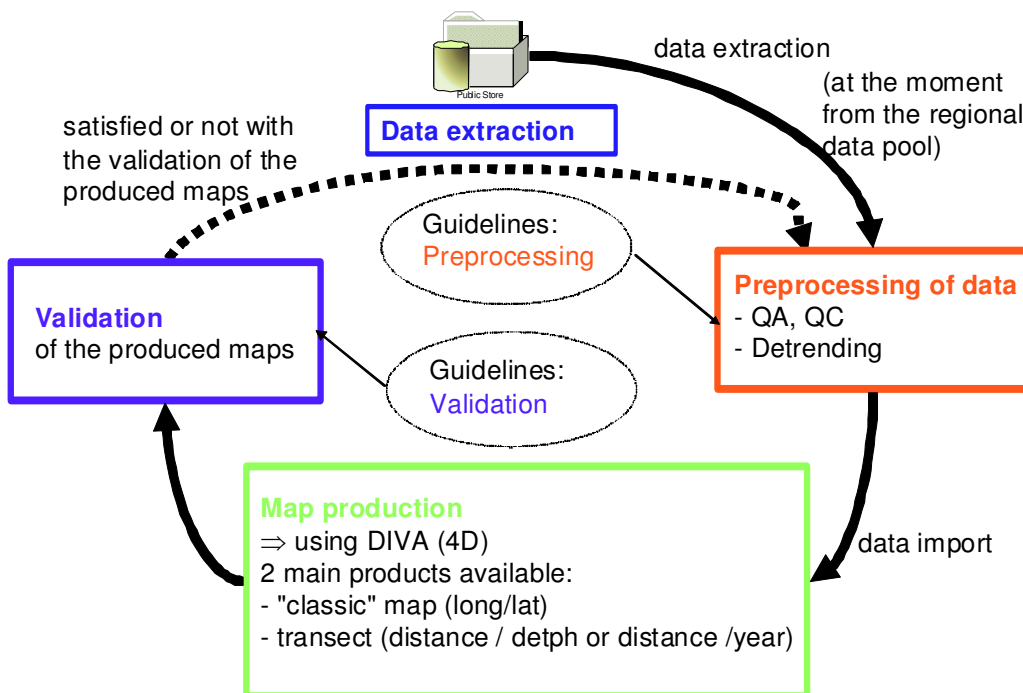


Figure 1: The overall concept for the DIVA maps production

Summarized in figure 2 in short headlines is the content of the 4 working steps for the DIVA maps production.

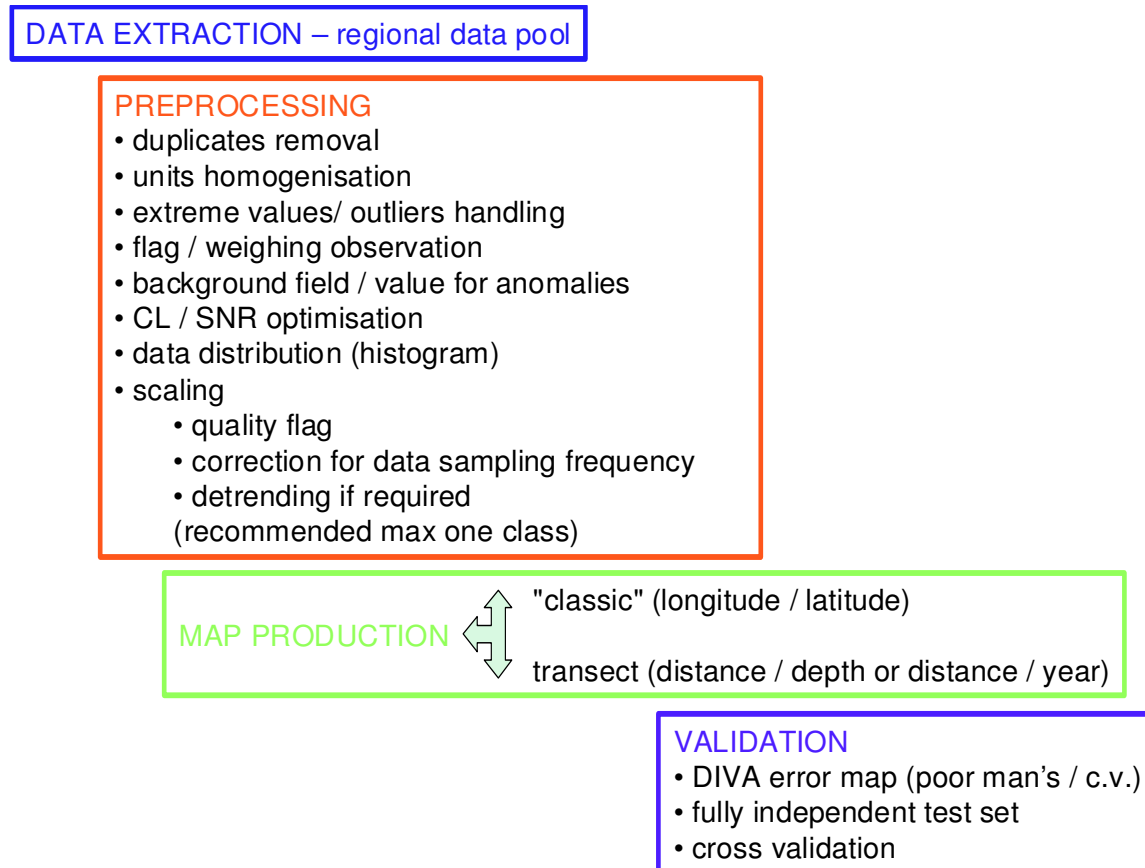


Figure 2: Working flow

In the following there is some text explanation for 4 working steps for the DIVA maps production.

## Data extraction

In the EMODNET chemical pilot project the data extraction will be from the three established regional data pool. In the future this could be done through the CDI (for a more the conceptual setup).

## Pre-processing of data

The pre-processing of the data is actually the most important step in the map production. It is recommended to go through the listed point when pre-processing data, see figure 2 the red box "pre-processing ". This will ensure a common and documented basis for the map production.

Some small comments or explanation to the listed points is shortly described here:

- Extreme values/outliers handling could e.g. be defined as being outside the range of [mean  $\pm$  standard deviation].
- Background field - if the mean background field is not suitable for converging, a reference field or value is to be submitted
- CL/SNR is to be optimised in an "automatic / objective" manner
- About scaling
  - Could be done with respect to the quality flag
  - A detrend function should be used for correcting years with respect to events, regime shift, trends, etc. This can be done in the DIVA batch mode, but might be best employed in the pre-processing phase. A similar processing could be applied to station classes: harbor, coastal & open water. Ditto if mixing

matrix classes or tissue classes. Generally it is recommended only to use detrending if required and only for max one class

- Variation through year classes in data sampling frequency is to be corrected (each year is to be weighed accordingly)

The sampling frequency by years in the water column for lead is shortly described as an example. The sampling is heterogeneous along the years (see figure 3), it is obvious that a climatology related product should correct for this data distribution:

- By weighing according to the years,
- By correcting eventual difference (detrending),

Here is 2006 higher than 2007 but with less observations it will be artificially weighed down.

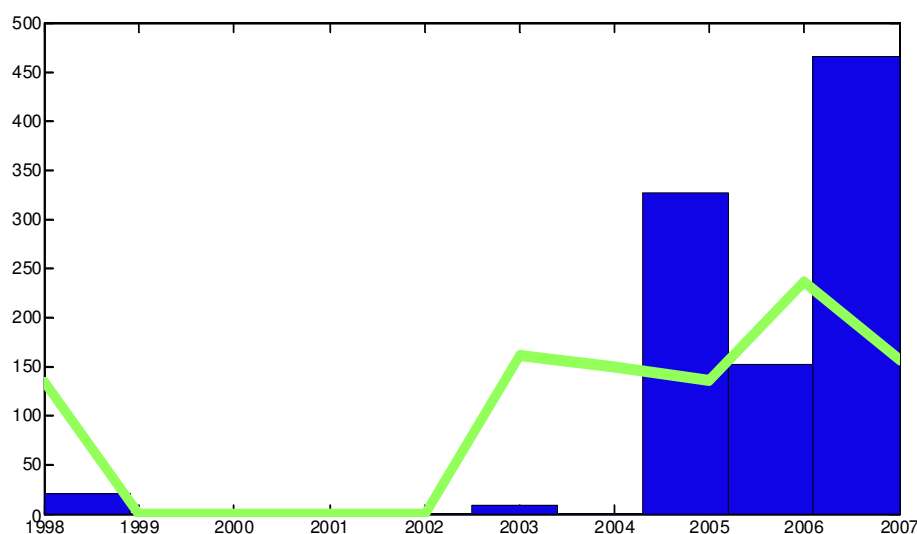


Figure 3: Explanation: green line: mean / year and blue histograms: data distribution / year

## Maps production

For the maps production the DIVA-software DIVA 4D (GODIVA) will be used. The maps will be the "classic" maps (longitude / latitude – "surface maps"), but for some data transect (section) along the coast could be relevant.

Using the DIVA software for the maps production actually means that you have a DIVA field production with the following sub products:

- bin field
- analysed field
- error field
- masked analysed field with 0.3 x the error field (EU management maps for the MSFD)
- masked analysed field with 0.5 x the error field (EU management maps for the MSFD)
- bottom field

for map generation.

The produce maps for the 3 regional data pools will be uploaded to the EDMODNET portal. At the moment the viewing and downloading service is available at: <http://gher-diva.phys.ulg.ac.be/emodnet/>

It should be notice that the current viewing system only supports the "traditional" 4D products (longitude, latitude, depth and time). Analyses made on a section (along a coastline for example) cannot be included with the current version. It requires a quite large change to the way the current viewing system is built since the structure of the underlying net-cdf file would be different.

## **Validation of the produced maps**

It is recommended that the validation process contains the following three elements:

- 1) DIVA error map (poor man's / c.v.)
- 2) fully independent test set
- 3) cross analyses

Ad 1)

DIVA propose different options with regards to error field generation, the "poor man's error" is used by default, a cross validated or "real error" might be more suitable: According to time cost (ispec 11,  $\pm 1$ ,  $\pm 4$  or  $\pm 17$ )

Ad 2)

Keeping a sub set (5%) of the data outside the DIVA analyses will provide a fully independent test set in order to calculate a Root Mean Square Error and iterate that ten times

## **Summery**

In short terms the present workflow is as follows:

- Data extraction from the regional data pools
- Pre-processing (MatLab or SAS or ODV or ...)
- DIVA analyses by the DIVA 4D (GODIVA)
- Validation (DIVA / MatLab / ...)

It is suggested to work through the listed processes for 2 or 3 parameters at least, for the three different matrix sediment, biota, and water column (surface layer). Furthermore it is recommended only to make maps covering areas where there is a reasonable density of data sets .