

Diva Workshop 2016

Advanced features

Sylvain Watelet, Örjan Back, Alexander Barth & Jean-Marie Beckers

Acknowledgements: SeaDataNet, EMODnet Chemistry,
EMODnet Biology, STARESO



Data weighting I

1. The problem

Since DIVA works with a diagonal observation error covariance matrix, it is assumed that the observation errors are uncorrelated in space and time. In practice, this assumption is not always valid especially when dealing e.g. with cruise measurements (same instrument) or with time series at a fixed geographic point (representativity error).

2. The solution

The data weighting option proposes to decrease the weights in the analysis of such observations. These weights are based on an exponential function using a 3D (x,y,t) distance between several observations:

$$W_i = \frac{1}{\sum_{j=1}^{N_{box}} \exp\left(-\left(\left(\frac{\Delta x_{ij}}{L_X}\right)^2 + \left(\frac{\Delta y_{ij}}{L_Y}\right)^2 + \left(\frac{\Delta t_{ij}}{L_T}\right)^2\right)\right)} \quad (1)$$

Data weighting II

where N_{box} is the number of data in a box defined by the characteristic length scales $(3L_X, 3L_Y, 3L_T)$ surrounding each data. Δx_{ij} , Δy_{ij} and Δt_{ij} are the distances between 2 data in this box. A simple example is presented in Fig. 1.

Data weighting III

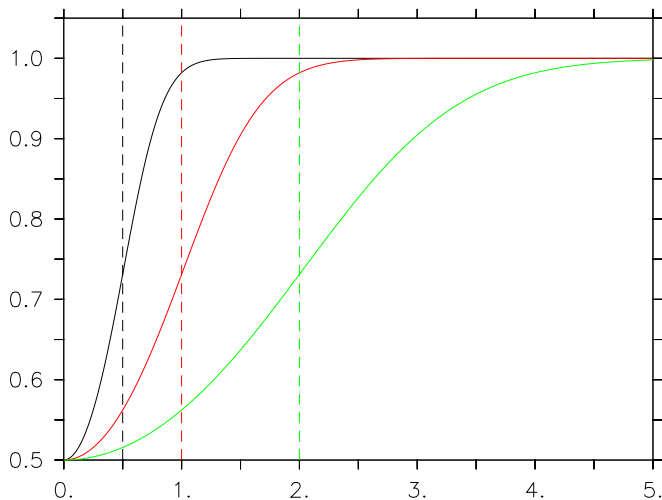


Figure 1: Weights of 2 points in fonction of the distance and the length scale chosen by the user

Data weighting IV

3. Application

A comparison between not-weighted and weighted analyses is shown in Figure 2.

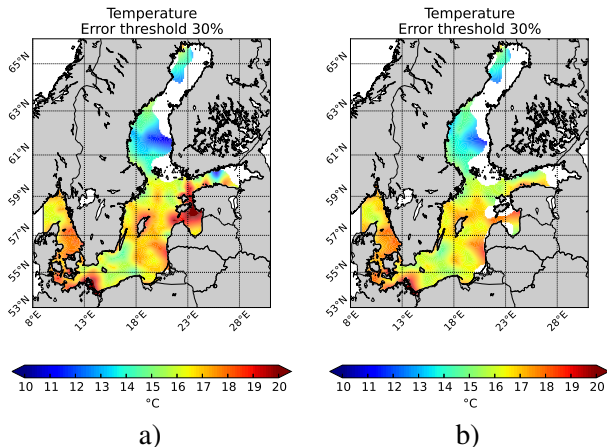


Figure 2: a) Not weighted data. b) Data weighting enabled.

Detrending I

Theory & Implementation

- Data sets have heterogeneous coverage in time and/or in space
- => Misinterpretations of the analysis
- Example 1 : more measurements during a specially warm year than during other years
- Example 2 : much less cruise data sets in stormy periods than in calm periods.

Detrending II

Theory & Implementation

Here we present a method to deal with such problems by defining *classes* and *groups*.

A *group* is simply one way of subdividing the data into different members. For example a group can be based on years and the classes are 1995, 1996, 1997 if we are looking at this period. Another group could be based on seasons and classes could be winter, spring, summer and autumn.

In the functional to be minimised by Diva, there is a data-analysis misfit term :

$$\sum_{i=1}^{N_d} = \mu_i [d_i - \varphi(x_i, y_i)]^2 \quad (2)$$

where μ_i is the data weight on data d_i found in location x_i, y_i . The solution of the minimisation is the analysed field $\varphi(x, y)$.

If we define one group, each data point is in one and only one class C_j of this group. Hence when calculating the misfit in the

Detrending III

Theory & Implementation

minimisation part of Diva, we include now an (unknown) trend value for each class ($d_{C_1}, d_{C_2} \dots$):

$$\sum_{i \in C_1} \mu_i [d_i - d_{C_1} - \varphi(x_i, y_i)]^2 + \sum_{i \in C_2} \mu_i [d_i - d_{C_2} - \varphi(x_i, y_i)]^2 + \dots \quad (3)$$

If we assume we know the function $\varphi(x, y)$, minimisation with respect to each of the unknowns d_{C_j} yields

$$d_{C_1} = \frac{\sum_{i \in C_1} \mu_i [d_i - \varphi(x_i, y_i)]}{\sum_{i \in C_1} \mu_i} \quad (4)$$

and similarly for the other classes. Hence we see that the trend for each class is the weighted misfit of the class with respect to the overall analysis. The problem is of course that φ is not known since it is also the result of the minimisation process. However, we can iterate

Detrending IV

Theory & Implementation

and start with an analysis without detrending. Then, using the field of φ , we can calculate a first guess of the trends in each group and subtract it from the original data. Then a new analysis can be performed, the trends recalculated and so on until convergence.

Detrending V

Theory & Implementation

Here we generalize by allowing several groups of classes.

The detrending is done hierarchically:

- 1 Trends for the first group are calculated and removed from the data.
- 2 The second group is treated and so on.
- 3 Once the data has been detrended, a new diva analysis is performed.
- 4 With the new analysis, the data-analysis misfit (or residual) can be reused to calculate better estimates of the trends.

This loop is repeated a predefined number of times.

Detrending

In practice

In `input` directory provide:

- `divadata` directory where data set files have more than five columns (5th, 6th, ... contain the information in which class the data point belongs)
- same other inputs as for normal run

In `Climatology` directory provide the usual input text files and:

- Edit the `driver` file and
- choose a flag number for detrending a value (less or equal to the number of groups) present in your data set

Run `divadoall` script file.

- Results will be stored in

`output/3Danalysis` directory

Data transformation

See other presentation