DIVA Manual

Contents

1 Detrending data by defining groups and classes

When analysing climatological data one is very often faced with data sets that have heterogeneous coverage in time in space. This can lead to misinterpretations of the analysis if for example there have been much more measurements during a specially warm year than during other years. It is also not uncommon that there are much less cruise data-sets in stormy periods than in calm periods. Here we present a method to deal with such problems by defining classes.

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.

1.1 Theory

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

$$\sum_i \mu_i \left[d_i - \psi(x_i, y_i)\right]^2 \tag{1}$$

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 $\psi(x, y)$.

If we define one group, each data point is in one and only one class of this group. Hence when calculating the misfit in the minimisation part of DIVA, we include now an (unknown) trend value for each class ($d_{C_1}, d_{C_2}...$).

$$\sum_{i \in C_1} \mu_i \left[d_i - d_{C_1} - \psi(x_i, y_i)\right]^2 + \sum_{i \in C_2} \mu_i \left[d_i - d_{C_2} - \psi(x_i, y_i)\right]^2 + ... \tag{2}$$

If we assume we know the function $\psi(x, y)$, minimisation with respect to each of the unkowns d_{Cj} yields

$$d_{C_1} = \frac{\sum_{i \in C_1} \mu_i \left[d_i - \psi(x_i, y_i)\right]}{\sum_{i \in C_1} \mu_i} \tag{3}$$

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 the ψ is not known since it is also the result of the minimisation process. However, we can iterate and start with an analysis without detrending. Then, using the field of ψ, we can calculate a first guess of the trends in each group and subtract if from the original data. Then a new analysis can be performed, the trends recalculated and so on until convergence.

1.2 Implementation

Here we generalize by allowing several groups of classes.

The detrending is done hierarchically: First trends for the first group are calculated and removed from the data. Then the second group is treated and so on. Once the data has been detrended, a new diva analysis is performed. 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.

1.3 Use

Simply provide `./input/data.dat` with additional fifth, sixth ... columns. If you do not want to use variable data weight, column 4 must contain the value of 1. Column 5, 6 etc contain the information in which class the data point falls. Classes must be numbered starting with 1.

Example:

- column 5 contains value 1 for a data point of the year 1975, 2 for 1976, 3 for 1977 and so on
- column 6 contains 1 for a data corresponding to month 01-03, 2 for the month 04-06 and so on
- column 7 contains 1 for day values, 2 for night values
- column 8 contains 1 for points that have a density below 1025 kg/m^3, 2 for points that have a density above it.

Execute `divadetrend ngroups [niterations]`. The parameter `ngroups` specifies that the first `ngroups` will be used for the detrending. (You might create for exemple 5 groups and try with detrending on the first one only using `divadetrend 1`. The optional parameter `niterations` tells how many iterations are to be performed for the detrending. Default value is 10 iterations.

Outputs: `./output/rmsmisfit.dat` contains the evolution during the iterations of the misfit (after detrending). It should decrease if the detrending works well. `trends.all.1.dat` deals with group 1 and contains on column 1 the class number and on columns 2 the final trend value associated with it. Columns 3 and 4 correspond to the next to last iteration and the last columns to the first iteration.

`divagnu` procudes plots for the trend in each group.

Note: presently you can define at maximum 5 groups with each group having 50 classes (members). You can increase these limits by editing and compiling `detrend.f`

NOTE: it is assumed that the mesh already exists, otherwise execute `divamesh` before.

1.4 Interna

`detrend.a` creates a new `./input/data.dat`. Input: `fort.88` original data file, `fort.89` analysis at data points. Output: `fort.90` modified data file with data detrended by groups and classes. `trends.1.dat` trends for classes of group 1. `trends.2.dat` trends for classes of group 2 etc

1.5 Example

`./examples/Trends` contains data sampled from a spatial pattern over which was added a seasonal cycle, dayly cycle and interannual variations. Groups are years, month and hours. Look at matlab file `pseudodata.m` how the data were generated and look how `divadetrend` is able to reconstruct the time variations. Also look at the analysed field, with and without detrending.