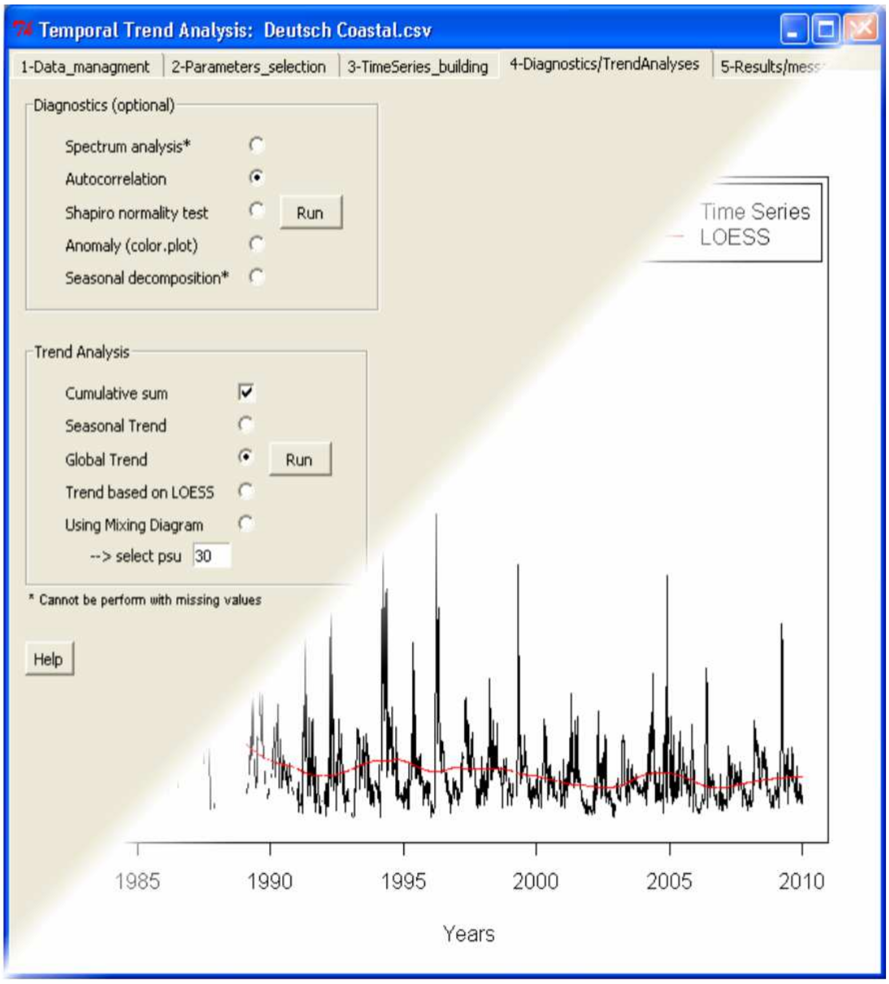
David Devreker
(David.Devreker@ifremer.fr)
(david_devreker@yahoo.fr)

Alain Lefebvre
(Alain.Lefebvre@ifremer.fr)

Centre IFREMER Manche - Mer du Nord
150 Quai Gambetta
62321 Boulogne sur Mer Cedex

TTAinterface



An Interface for Temporal Trend Analysis and Diagnostics

User guide

TTAinterface version 1.01 for R v2.14+

Contents

Introduction

The program TTAinterface is written with the R programming language in version 2.14+. It allows performing temporal trend analysis through a friendly user interface. The advantage of such kind of interface is the balanced choice it offer between the wide variety of analysis, the freedom that offer R through the console or its different packages but which avoid to perform routine analysis by a lambda user and the easiness of a clear interface with driven choice of well choose analysis and diagnostics tools that allow routine analysis in the frame of a common procedure. One advantage of the R language and its package is to be a free public licence (GPL) that allows to freely distributing the interface.

This document is a guide that shows you how to use the interface. The first part of this guide shows how to install and load the interface through the R console. The second part shows a rapid overview of the interface and its possibility. As an example, the dataset 'Deutsch Coastal.csv' will be used. This dataset regroup chlorophyll concentration, salinity and temperature at 7 different permanent coastal sampling stations from 1985 to 2009 with a fluctuated frequency sampling (~ between 7 days and 1 month). The third part shows a more detailed documentation of the interface functioning using the dataset 'North Sea.csv' as an example, this dataset contain more parameters (depth and nutrient) than the previous one but have also a lot more missing values.

A. Installing R and the TTAinterface package

Starting from the portable version, all the files you need are present in the 'TTA Portable Package' zip archive. First, you need to install the R software (v2.14.1), using R.2.14.1-win.exe, that comes with basic packages and a command console. The version present in the zip archive is the 2.14.1 for Microsoft Windows; any other version of R >= 2.14 should also work. You can download R for other platforms on the CRAN website (http://cran.r-project.org/), however the compatibility of the present package with other platforms that Microsoft Windows have not been tested. More complete instruction concerning R installation can be found on the CRAN website.

Installation of a portable version of the package

Once R install, run it and go in the 'Packages' menu of the console, click on 'Install package(s) from local zip files…' (step **1** in Fig. 1) and select the file 'TTAinterfaceTrendAnalysis_1.01.zip' that comes in the provided zip archive. The TTAinterface package will automatically download and install all other necessary packages if they are not already present in your computer (Fig. 2) (you obviously need an internet connection).

Installation from CRAN mirror

Alternatively, the package should be available on the CRAN mirror (need an internet connection). To install the interface though this way, click on "Install package(s)" in the 'Packages' menu of the console (step **2** in Fig. 1), select your mirror (your country), and follow the instructions to find the TTAinterfaceTrendAnalysis package.

Launch the interface

When everything has been install click on ‘Packages/Load package…’ in the ‘Packages’ menu of the console and select TTAinterfaceTrendAnalysis in the list (step **3** in Fig. 1). A small panel will appear inviting you to start the interface, click on the button. The step **1** or **2** need to be done only once to install the package, skip it and go directly to step **3** every time you need to load the interface.

In some case, if you close both the TTAinterface and the start panel, and try to reload them with step **3** without closing the R console it couldn’t work. Then you have to re-start the R console or if you want to stay in your R session, enter the line TTAinterface() in the console (Fig. 1).

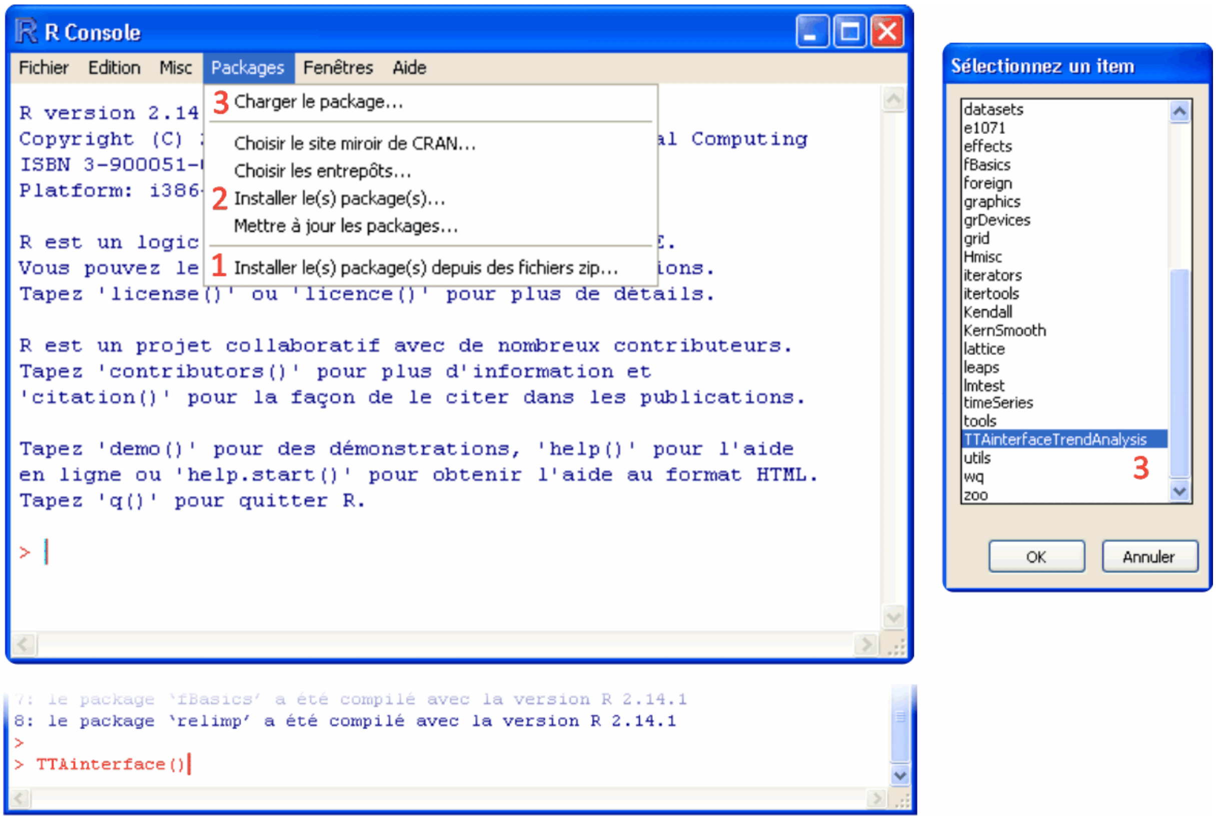


Figure 1. The R console and the different step to install and run the TTAinterface.

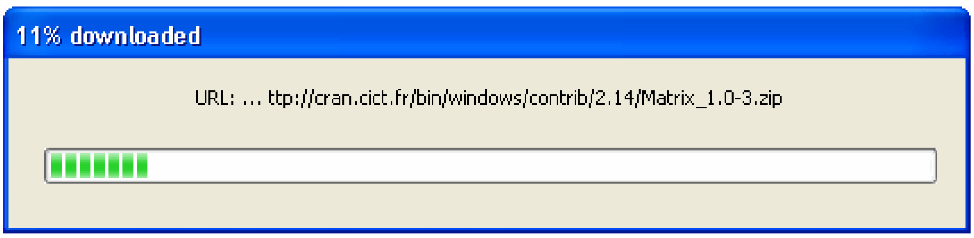
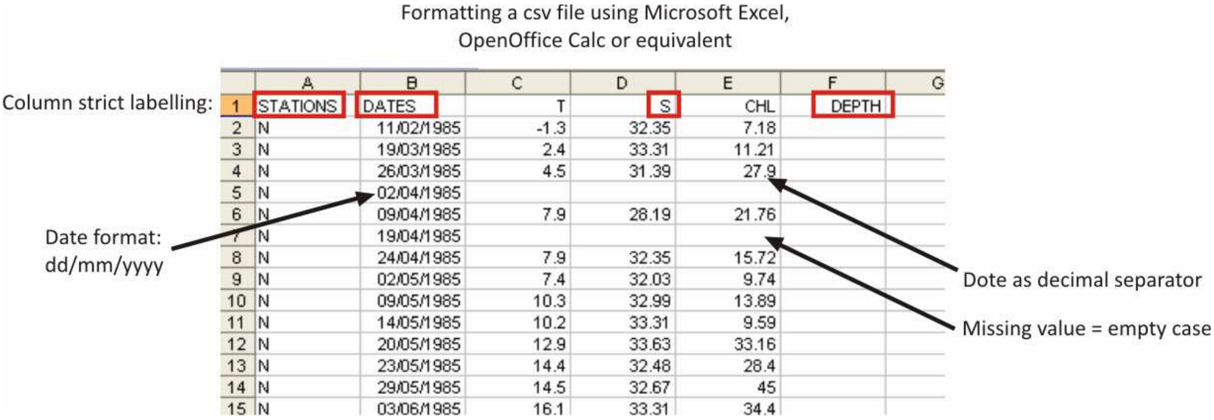


Figure 2. Packages auto-downloading window.

B. Create your dataset

The main 'difficulty' to use the interface is to prepare a csv file that will be correctly read by the program. Once in the interface, the guideline to perform analyses is very friendly.

The csv file has to be create with a spreadsheet software like Microsoft Excel or OpenOffice Calc. OpenOffice Calc is free (http://www.openoffice.org) and can manage files with 1 million lines whereas Microsoft Excel is limited to 65000 lines but is more intuitive to use. The csv file is a numeric table with column label in the first row. For interface needs, some of these labels have to be fixed. The column containing the sampling stations must be named STATIONS (with capital letter), DATES for the date column in format dd/mm/yyyy, DEPTH for the sampling depth column and S for the salinities column (Fig. 3). Columns with parameters values (chlorophyll, nutrient, phytoplankton…) can be freely labelled with the name of the parameters as a preference. All values in the same line must correspond to a unique sampling. Contrary to S and DEPTH columns, STATIONS and DATES columns are necessary to the interface work correctly (if you don't have stations, create the appropriate column and file it with a character).



	A	B	C	D	E	F	G
1	STATIONS	DATES	T	S	CHL	DEPTH	
2	N	11/02/1985	-1.3	32.35	7.18		
3	N	19/03/1985	2.4	33.31	11.21		
4	N	26/03/1985	4.5	31.39	27.9		
5	N	02/04/1985					
6	N	09/04/1985	7.9	28.19	21.76		
7	N	19/04/1985					
8	N	24/04/1985	7.9	32.35	15.72		
9	N	02/05/1985	7.4	32.03	9.74		
10	N	09/05/1985	10.3	32.99	13.89		
11	N	14/05/1985	10.2	33.31	9.59		
12	N	20/05/1985	12.9	33.63	33.16		
13	N	23/05/1985	14.4	32.48	28.4		
14	N	29/05/1985	14.5	32.67	45		
15	N	03/06/1985	16.1	33.31	34.4		

Figure 3. Summarized processes to formatting a csv file readable by the TTAinterface.

In your csv file you can keep columns that will be not used in the interface like coordinates or water masses labels (they will not be read by the program) but it is recommended to remove them to obtain the lighter csv file as possible.

Missing values must be empty case, 0 are read as values and characters (NA, NaN) can induce bugs (they are labelled by the program itself so you don't have to do it).

Decimal separator must be '.' (dote). Be careful with the value that appear like 6,00 (integer value with comma as decimal separator), they are not always replaced by 6.00 in Microsoft Excel.

Once your datasheet have been well prepared, save it using 'save as' (1 in Fig. 4A and Fig. 4B) and select CSV (Comma delimited) (*.csv) in the 'Save as type' option (2 in Fig. 4A and Fig. 4B). In OpenOffice Calc you have to check 'Edit filter settings' to select ';' as column separator in step 4 (Fig. 4B). In step 4, let also the 'Text delimiter' case empty.

Between step 2 and 4 OpenOffice Calc will ask you if you want to keep the actual format, check yes (3 in Fig. 4B).

Only one worksheet can be saved in a csv file (the active one by default).

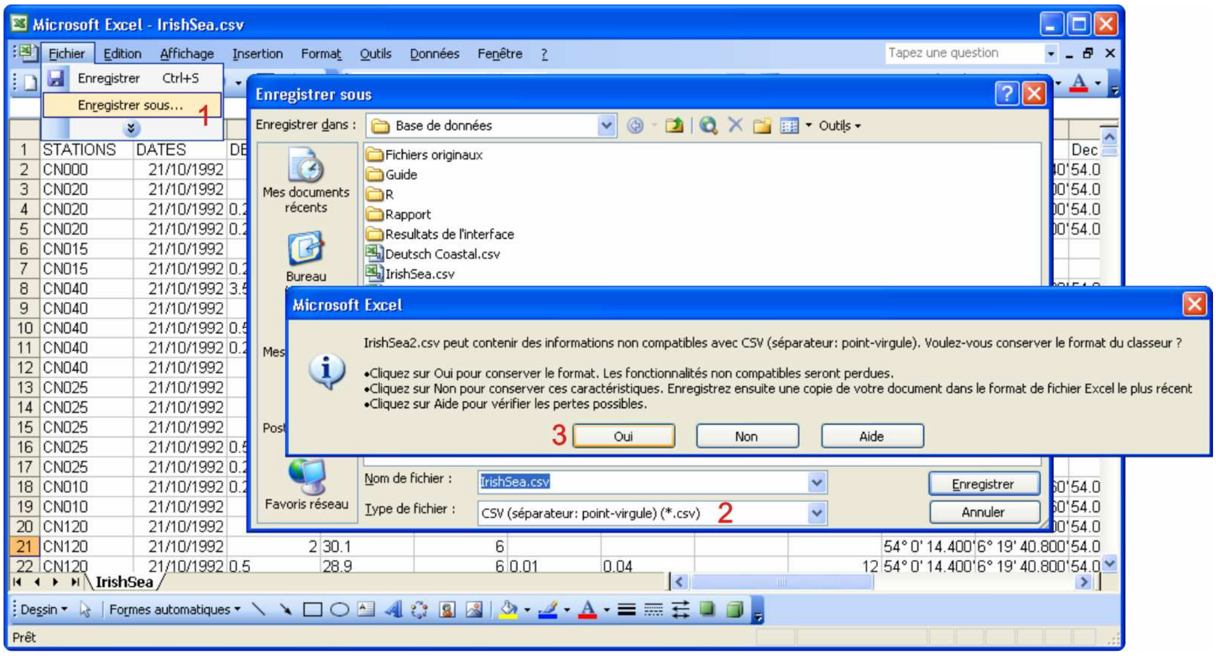


Figure 4A. Creating a csv file with Microsoft Excel.

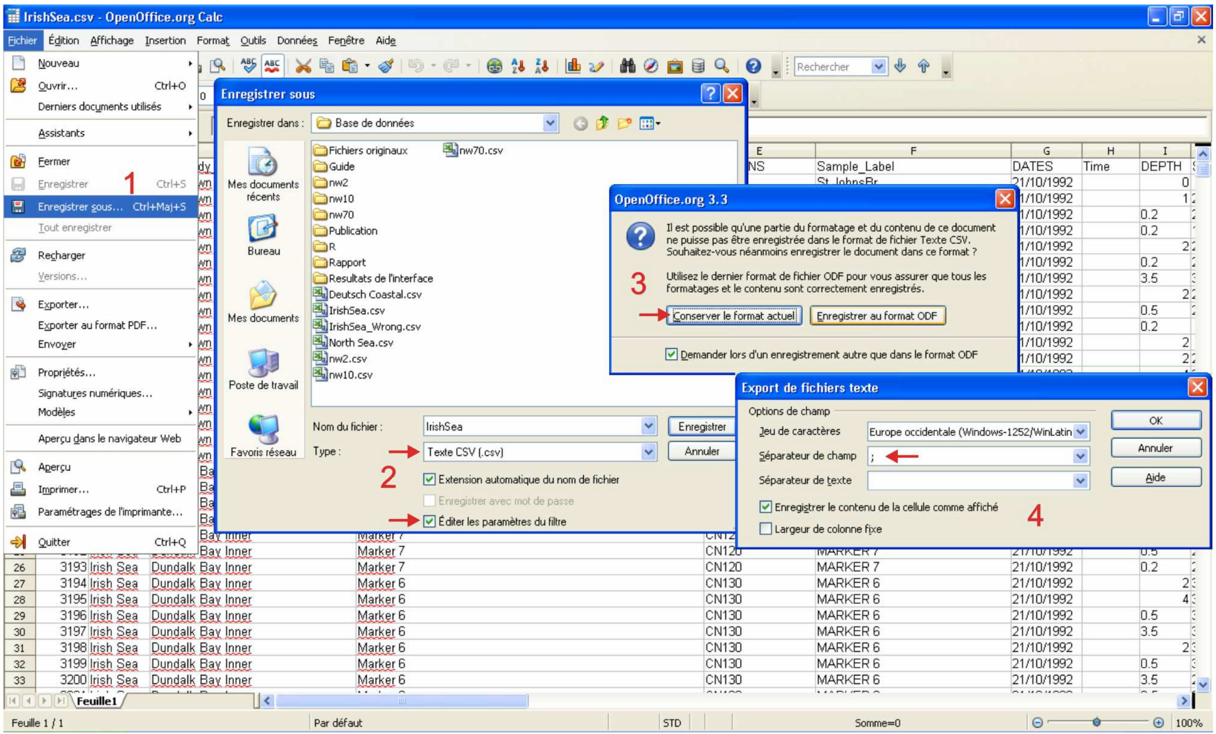


Figure 4B. Creating a csv file with Open Office Calc.

To summarize:

- Must be a .csv file
- Column labelling: STATIONS for sampling stations, DATES for dates, S for salinity and DEPTH for depths values.
- Dates format must be dd/mm/yyyy
- Dote as decimal separator
- Missing values must be empty cases

C. Quick steps

1. Managing your database

The first panel of the interface '1-Data_managment' allow to interact with the dataset

Import your database

This is the first page you can see when you start the interface (Fig. 5). You can only imports a .csv file and read the advice to correctly import a .csv file, all other options are disable.

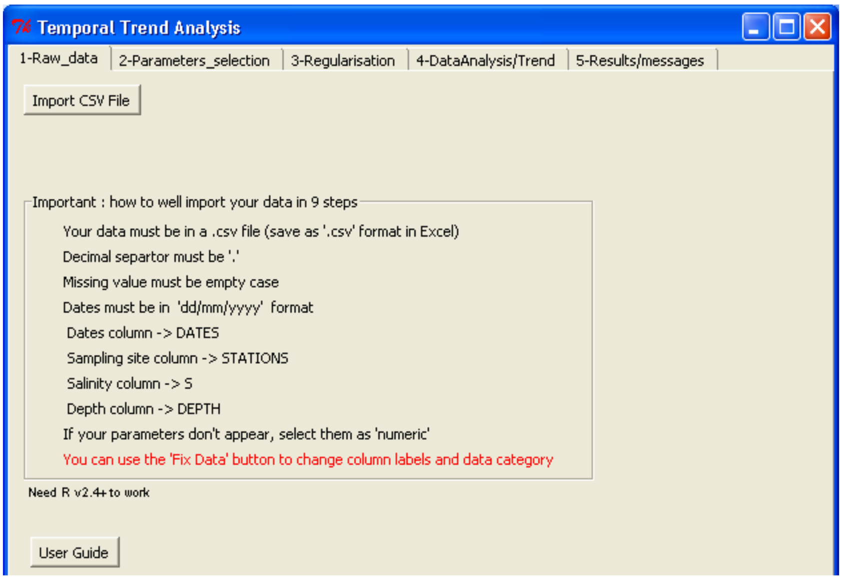


Figure 5. Panel 1 at the interface start.

Once your .csv file is imported more options appear:

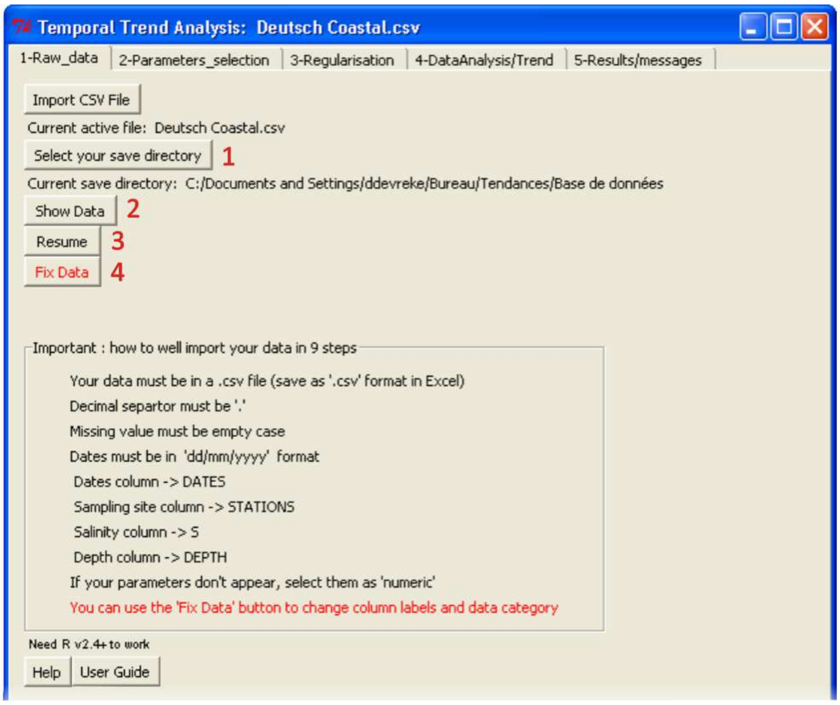


Figure 6. Panel 1 once csv file imported.

Change the save directory

The button ‘Select your save directory’ (1 in Fig. 6) will let you choose a folder on your computer to save the different figures and results of analysis you will obtain with the interface. By default this is the folder where is stock your imported .csv file. If you import a new .csv file the save directory will be reset to default.

Display your data

The button ‘Show Data’ (2 in Fig. 6) displays a table of your imported data.

Summarize your raw dataset

The button ‘Summary’ (3 in Fig. 6) displays a table with the main descriptive statistics of your raw data (Fig. 7) with Min. = minimum value of the distribution, 1st Qu. = first quantile (25%), Median = the median of the distribution (50%), Mean = the mean of the distribution, 3rd Qu. = third quantile (75%), Max. = maximum value of the distribution and NA’s = number of missing values in the distribution of the parameter.

Stat2

	STATIONS	DATES	T	S	CH
1	B : 684	01/09/2003: 7	Min. : -2.000	Min. : 0.12	Min. : -0.090
2	L15: 222	06/08/1997: 6	1st Qu.: 7.397	1st Qu.: 26.10	1st Qu.: 3.844
3	L16: 229	09/08/1995: 6	Median : 14.000	Median : 29.30	Median : 7.180
4	L2 : 164	10/06/2003: 6	Mean : 12.573	Mean : 28.34	Mean : 10.748
5	L9 : 267	18/08/2003: 6	3rd Qu.: 17.500	3rd Qu.: 31.00	3rd Qu.: 13.046
6	N :1187	(Other) :2922	Max. : 24.400	Max. : 54.50	Max. :265.320
7	W : 201	NA's : 1	NA's :310.000	NA's :180.00	NA's :171.000

Figure 7. Summary of the ‘Deutsch Coastal’ raw database.

Edit your data

The button ‘Fix Data’ (4 in Fig. 6) let you edit your dataset directly from the interface (Fig. 8). However you can only perform simple task such as modify column label, data category or edit case one by one:

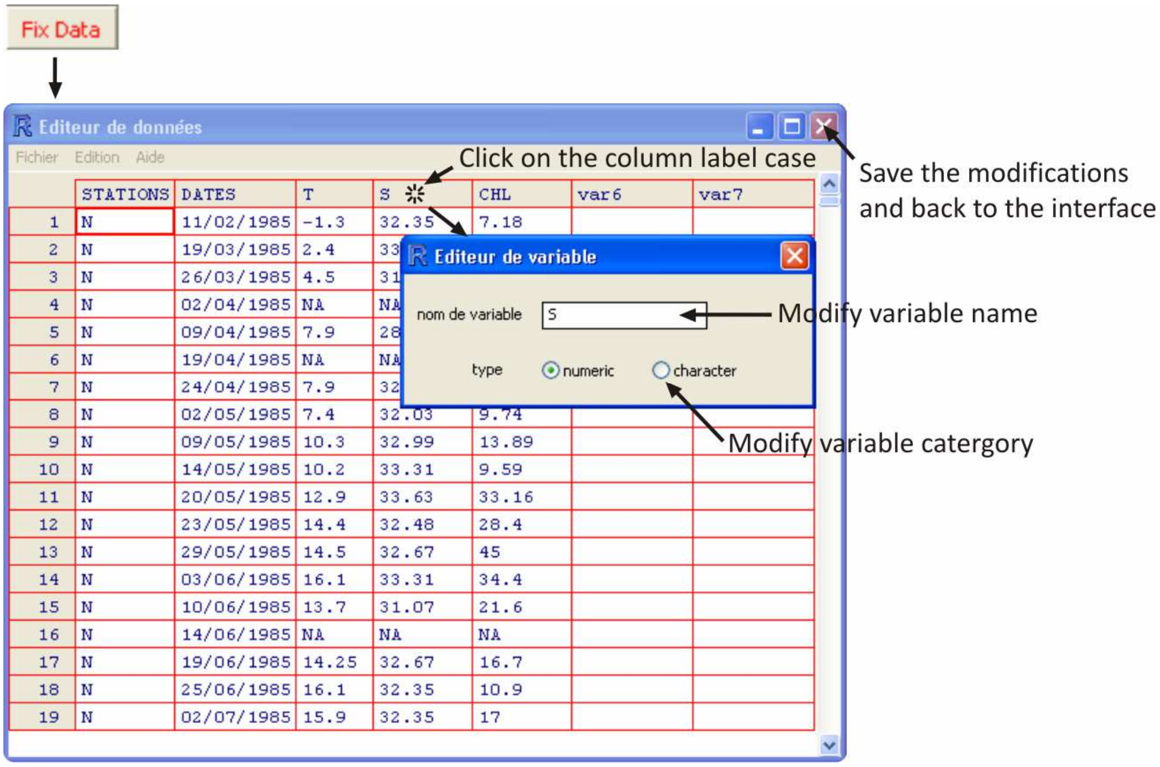
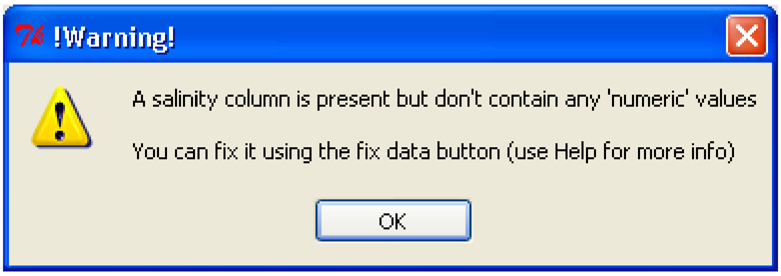


Figure 8. Data editor spreadsheet when using the ‘Fix Data’ button in panel 1.

If you import a csv file containing a column labelled ‘S’ for salinity but with values that have not been identified as ‘numeric’ a warning message appear (same thing for DEPTH column):



2. Select your parameters

The second panel of the interface '2-Parameters_selection' allows selecting the desire parameters for analysis (number corresponding to figure 9):

1. The different sampling stations
2. If data of each stations have to be melt during analysis
3. The parameter to analyse
4. Interval of salinity to analyse (sliders)
5. Interval of depth to analyse (if exist)
6. Interval of years to analyse
7. Months to analyse (must be spaced)

<u>Summary your selected parameters</u>

The button 'Summary' (8 in Fig. 9) displays descriptive statistics of the raw selected data. No mathematical treatment has been done on this data, only simple selection and sorting.

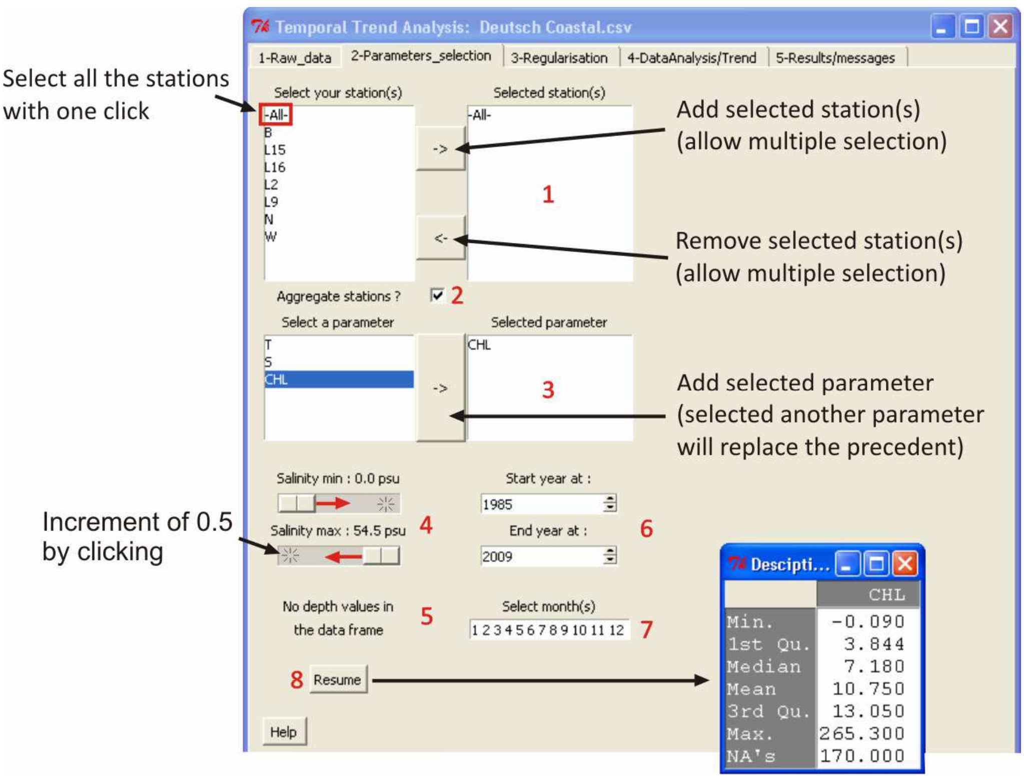
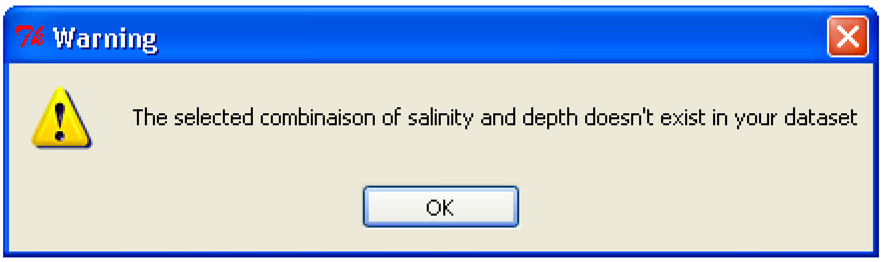


Figure 9. Options in panel 2.

If you didn't select any parameter or station before to perform mathematical treatment:

If the combination of salinity and depth selected in 4 and 5 is irrelevant:



3. Build your time series

The third panel '3-TimeSeries_building' allows to build a regularized time series necessary to perform temporal trend analysis (Fig. 10).

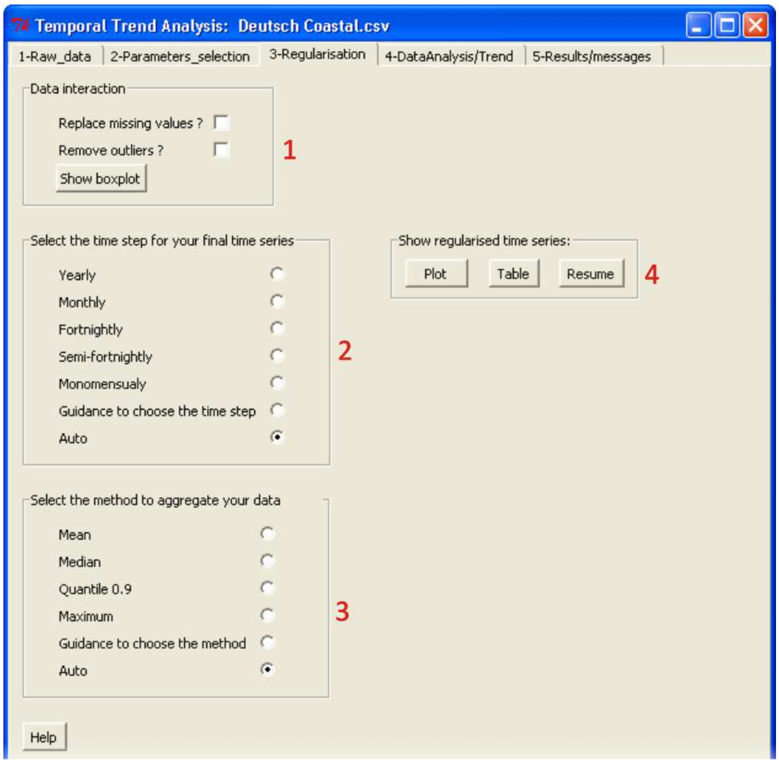
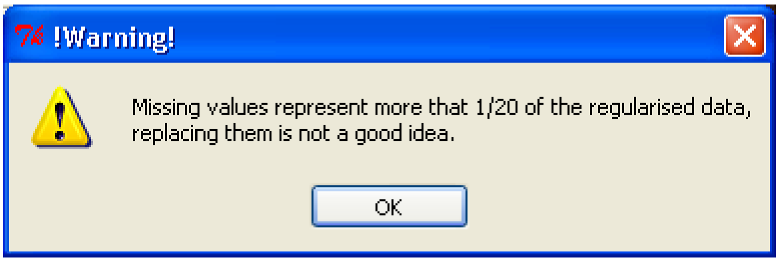


Figure 10. Options in panel 3.

Outliers and missing values

The 'Data interaction' frame (1 in Fig. 10) allows to remove outliers from the raw data distribution and to replace missing values from the time series (require for some diagnostic processes). If you choose to replace missing values but they represent a high part in your data distribution:



The button 'Show boxplot' displays a boxplot of your data distribution, by years, with outliers (Fig. 11).

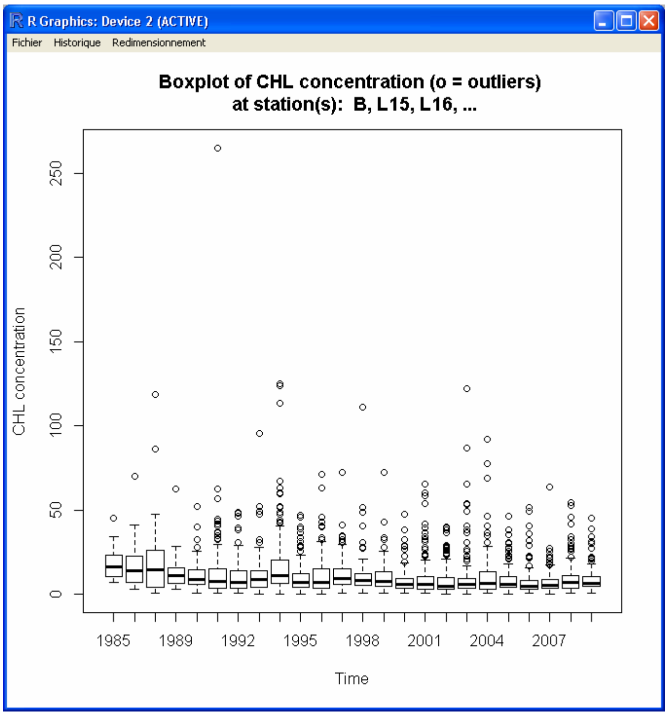


Figure 11. Boxplot of the ‘Deutsch Coastal’ raw database with outliers

Aggregation methods

Frame 2 and 3 in Fig. 10 show the options to build your regularized time series. You can select the time step and the method of data aggregation or let the interface select it automatically for you (default option). The auto option computes balanced choices, alternatively select guidance option to see the advices before choosing (Fig. 12).

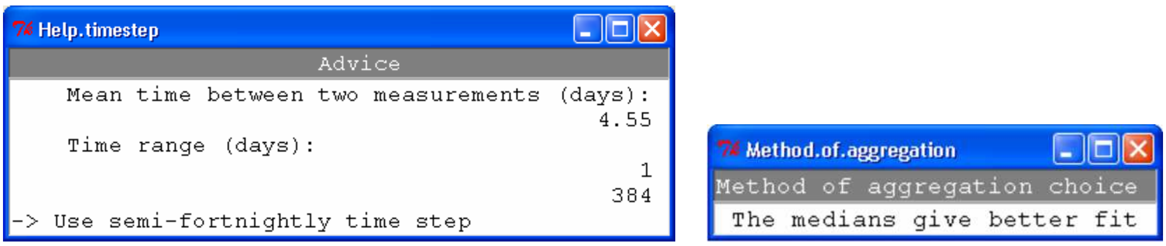


Figure 12. Guidance messages to choose balanced time step and method of aggregation to build time series.

Frame 4 allows to display a plot, a table or a summary of the regularized time series build with your selected options (Fig. 13). Plot and table are automatically saved when called.

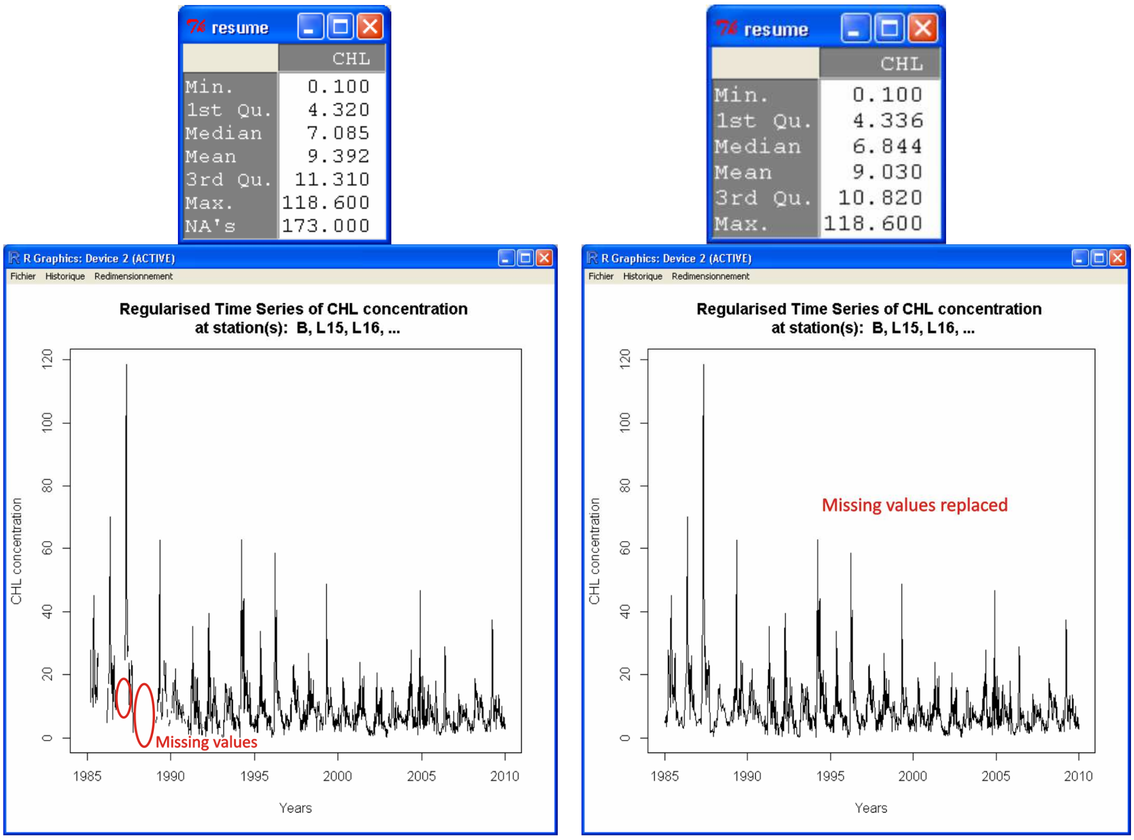


Figure 13. Summaries (top) and plots (bottom) of regularized time series with missing values kept (left) or replaced (right).

4. Perform diagnostics and trend analyses

The forth panel '4-Diagnostics/TrendAnalyses' allow to perform diagnostics and temporal trend analyses on your regularized time series (Fig. 14).

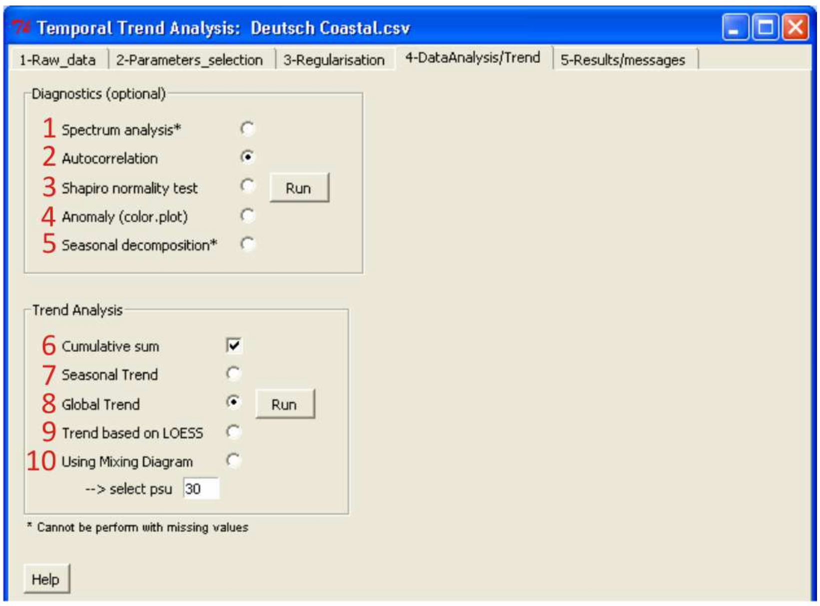
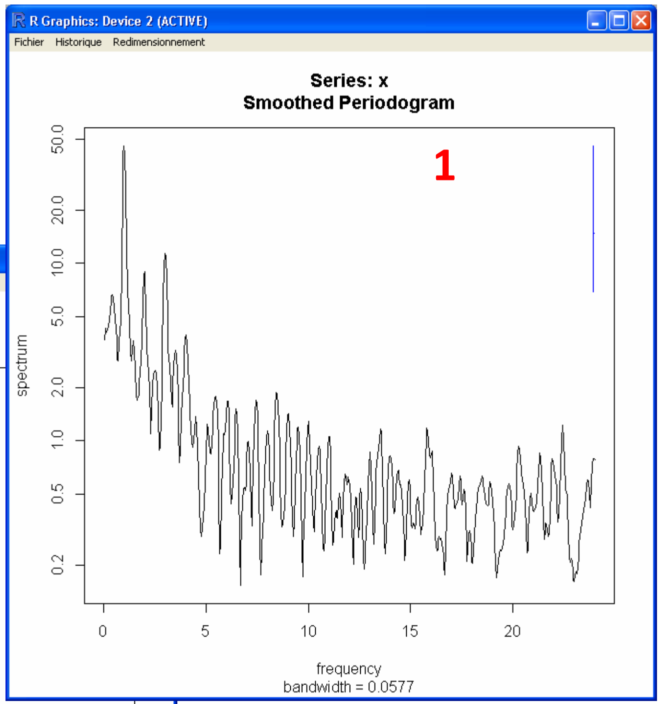
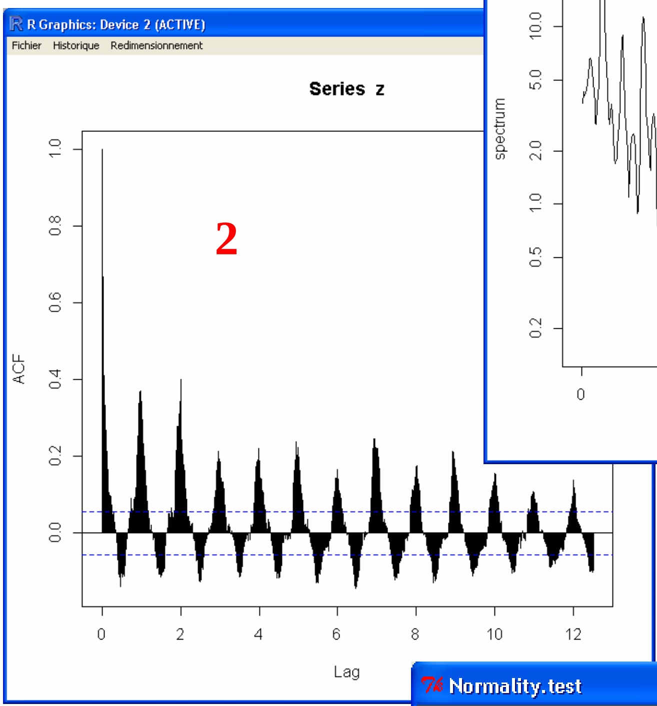


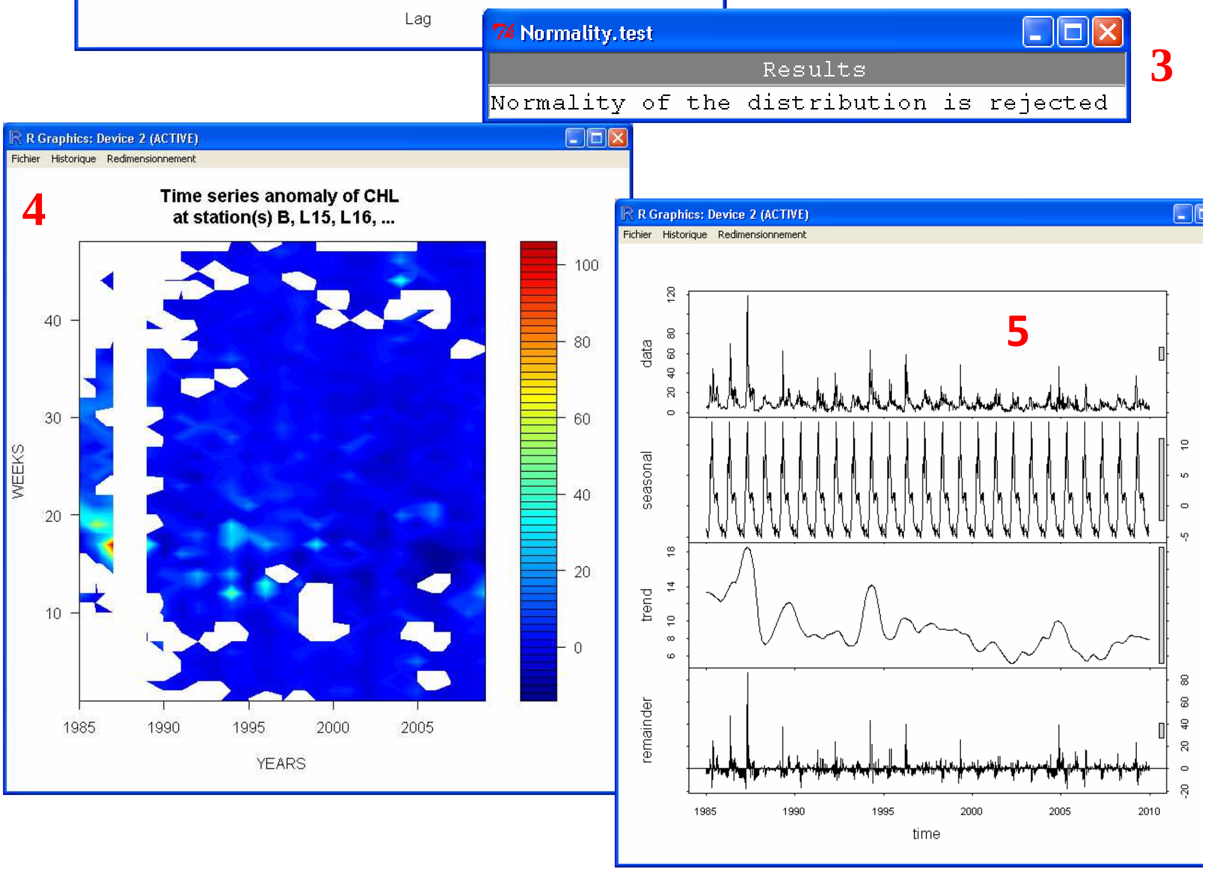
Figure 14. Options in panel 4.

Diagnostics tools (first frame), numbers correspond to Fig. 14.

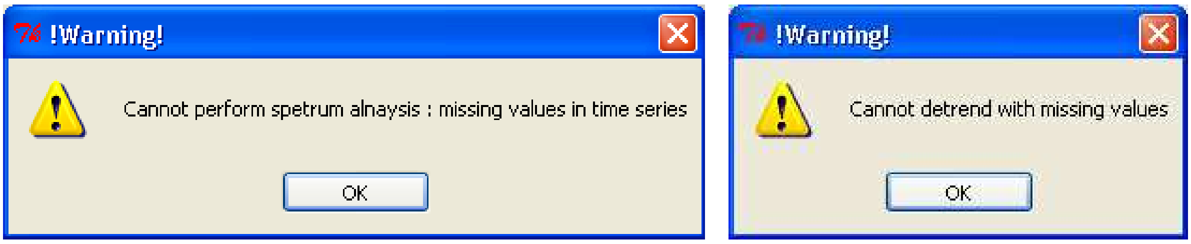
1. Fourier spectral density
2. Autocorrelation diagram
3. Shapiro Normality test
4. Anomaly by step time
5. Time series decomposition







Performing 1 and 5 while keeping missing values will display warning messages:



Temporal trend tests (Trend Analyses frame)

- Global trend (8 in Fig. 14 and 15) and Seasonal trend (7 in Fig. 14 and 15) (all based on Seasonal Kendall test) results are display on panel 5 'Results/Messages':

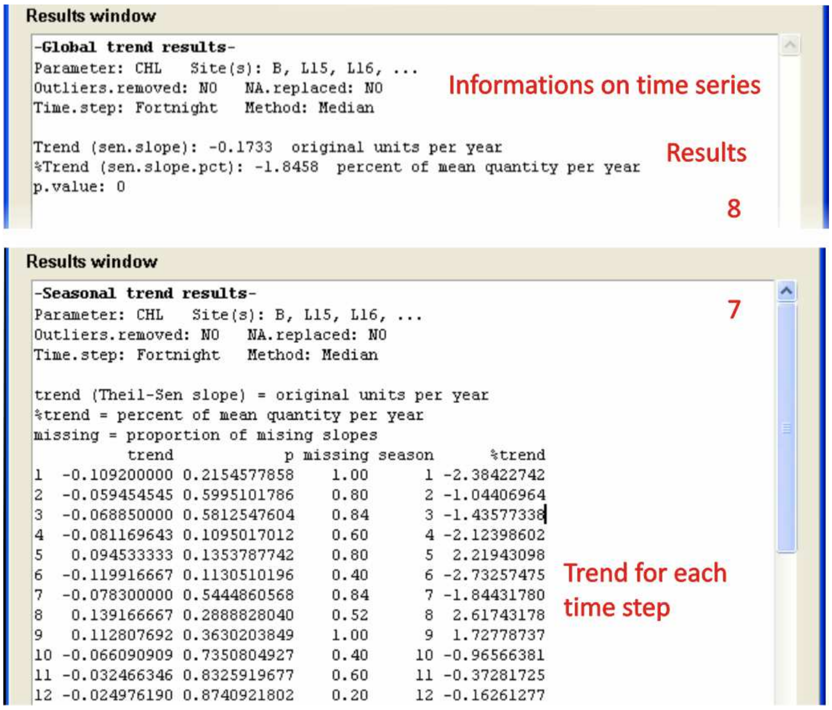


Figure 15. Results of temporal trend test (Global Trend in 8 and Seasonal Trend in 7) display in panel 5.

Visually identify and select different periods of trend with Cumulative sum (6 in Fig. 14)…

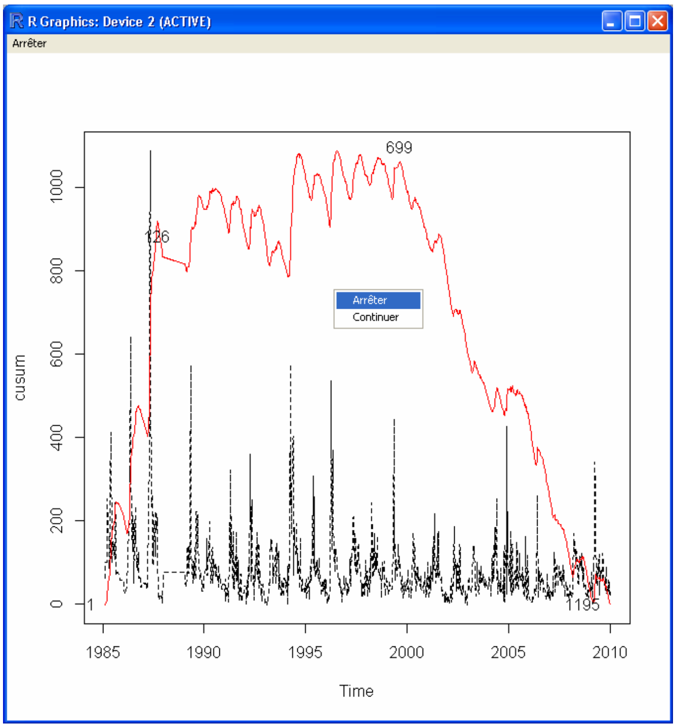
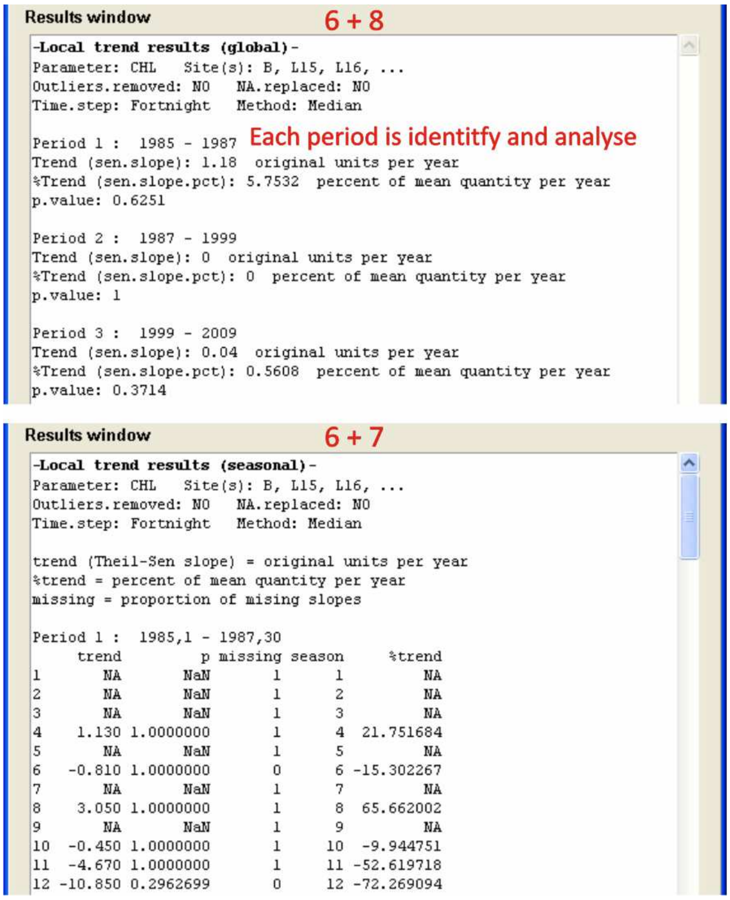


Figure 16. Cusum plot (cumulative sum in red), with periods manually identified (numbers).

…and perform Kendall test on each of these periods




```
-Local trend results (global)-
Parameter: CHL   Site(s): B, L15, L16, ...
Outliers.removed: NO   NA.replaced: NO
Time.step: Fortnight   Method: Median

Period 1 :  1985 - 1987
Trend (sen.slope): 1.18  original units per year
%Trend (sen.slope.pct): 5.7532  percent of mean quantity per year
p.value: 0.6251

Period 2 :  1987 - 1999
Trend (sen.slope): 0  original units per year
%Trend (sen.slope.pct): 0  percent of mean quantity per year
p.value: 1

Period 3 :  1999 - 2009
Trend (sen.slope): 0.04  original units per year
%Trend (sen.slope.pct): 0.5608  percent of mean quantity per year
p.value: 0.3714
```

```
-Local trend results (seasonal)-
Parameter: CHL   Site(s): B, L15, L16, ...
Outliers.removed: NO   NA.replaced: NO
Time.step: Fortnight   Method: Median

trend (Theil-Sen slope) = original units per year
%trend = percent of mean quantity per year
missing = proportion of mising slopes

Period 1 :  1985,1 - 1987,30
     trend         p missing season     %trend
1       NA       NaN       1      1         NA
2       NA       NaN       1      2         NA
3       NA       NaN       1      3         NA
4    1.130 1.0000000       1      4  21.751684
5       NA       NaN       1      5         NA
6   -0.810 1.0000000       0      6 -15.302267
7       NA       NaN       1      7         NA
8    3.050 1.0000000       1      8  65.662002
9       NA       NaN       1      9         NA
10  -0.450 1.0000000       1     10  -9.944751
11  -4.670 1.0000000       1     11 -52.619718
12 -10.850 0.2962699       0     12 -72.269094
```
Figure 17. Results of temporal trend tests perform on the different periods identified with cusum plot .

Selecting another test than Kendall after choosing the Cumulative sum option will display the warning message:

- Trend based on LOESS smoothing (9 in Fig. 14):

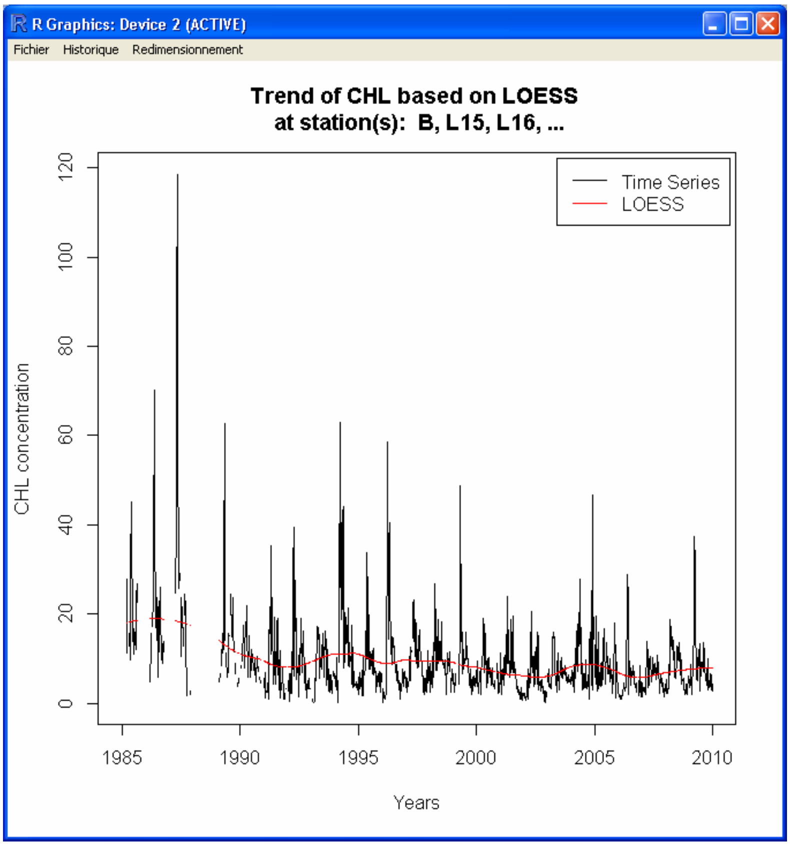


Figure 18. Plot of the regularized time series (black) with loess smoothing (red)

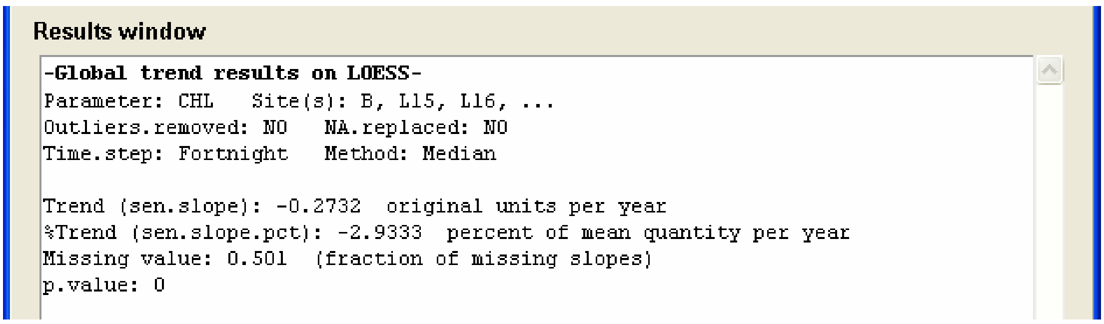


Figure 19. Results of temporal trend test (Global Trend) apply on loess smoothing, display in panel 5.

- Trend based on normalized concentration of nutrients at fixed salinity for each month (10 in Fig. 14):

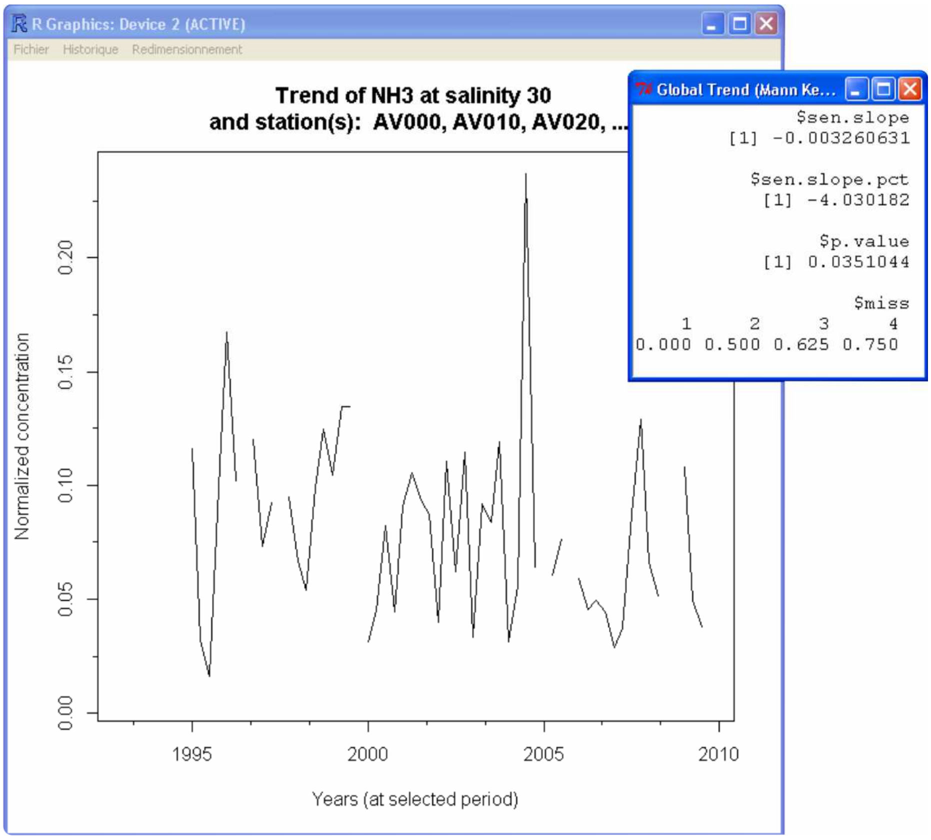
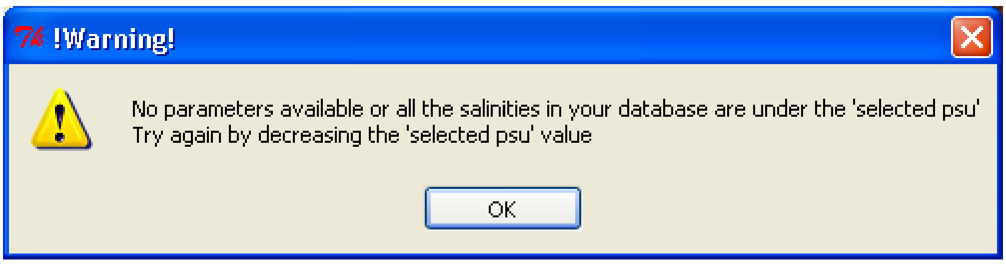


Figure 20. Plot of normalized NH3 concentration (from on another database) at salinity 30 and results (small text box) of Global Kendall perform on this time series.

If there is no salinity value in your data or if you choose a salinity to normalize nutrient over the maximum salinity of your dataset, a warning message will be display:



D. Detailed documentation

1. Interface organisation

The interface display all the options needed to perform temporal trend analysis through 5 successive panels (5 steps), from raw data managing to results display.
The panel 1 “1-Data_managment” focuses on the file and data management,
The panel 2 “2-Parameters_selection” focuses on the selection of the parameter and stations to analyse.
The panel 3 “3-TimeSeries_building” displays the option to build a regularised time series,
The panel 4 “4-Diagnostics/TrendAnalyses” focuses on diagnostic tools and statistics tests
The panel 5 “5-Results/Messages” displays results of analysis and divers messages and warnings.

Help buttons are available on each panel of the interface (except panel 5) to provide guideline on how to use options in their respective panel.

The top panel displays the name of the selected data file (once imported).

Figure 21. Top panel and panels’ titles of the interface.

The interest of panels against windowed menus is that all options are always visible and can be rapidly selected without going into multiple menus. This is only possible because the number of options has been optimized to the minimum needed to perform such analysis. Such interface can not be developed for more complex tools (which is not the objective of the interface).

2. Files and data managing panel

The first step of data analysis is the importation of your database in the interface. The interface will identify each column as a function of its label and category. In general columns with numeric values will be automatically identify as parameters. Other columns have to be manually labelled to facilitate the identification by the interface, such as sampling stations or depth (for further information see §B).

2.1 Import CSV file

To import a csv file containing your data just click on the button ‘Import CSV File’ in panel 1, it will display a window (a window generate by your OS in fact) where you can select and open csv files.

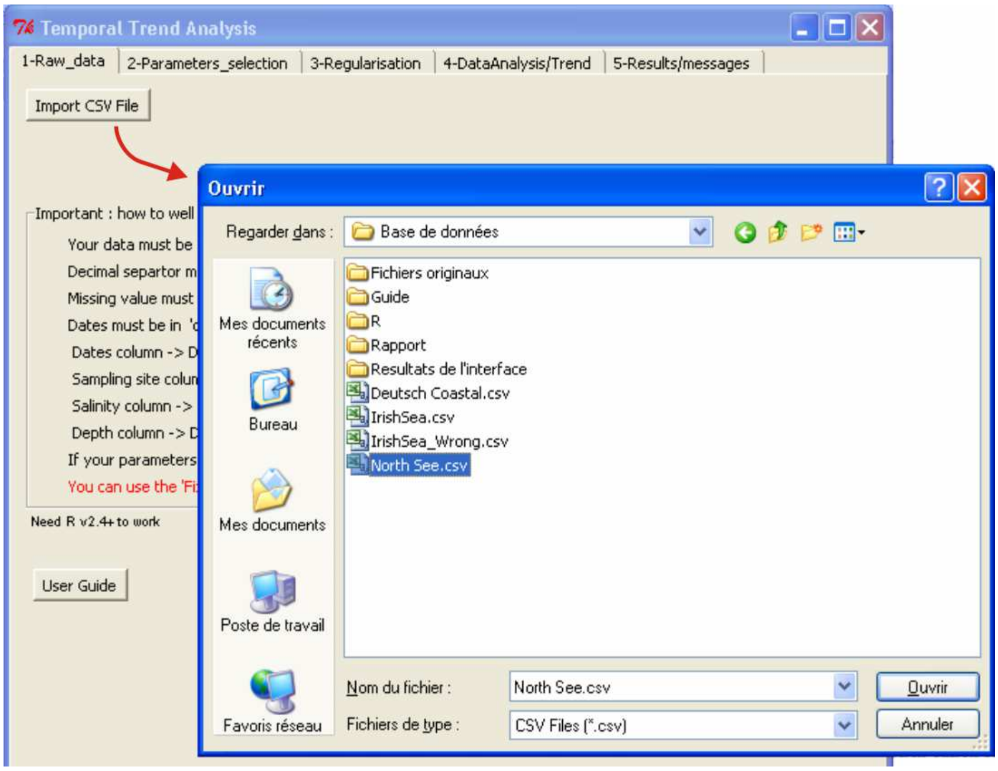


Figure 22. Panel 1 with csv file importation option and importation advices.

Until the file is imported panels 2 to 4 stay empty and panel 1 uncompleted. The other options will be available only when your data will be imported.

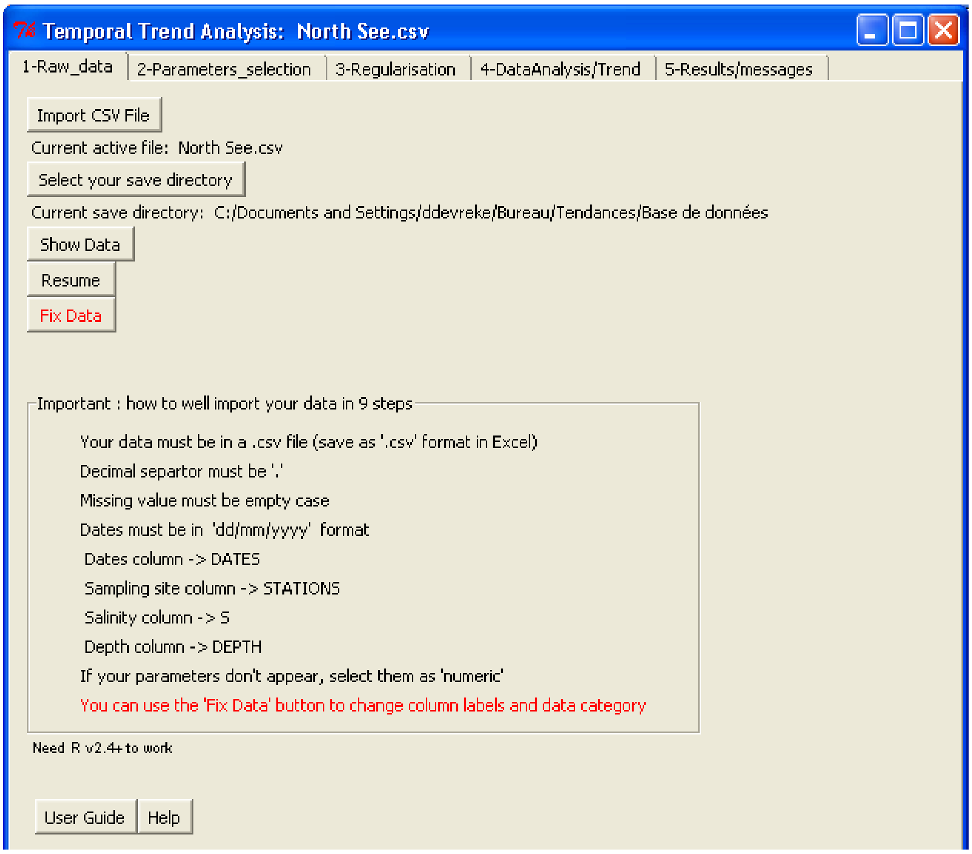


Figure 23. Panel one with all options available once csv file has been imported.

2.2 Select your save directory

By default, results of analysis and figures will be save in the same directory as your csv file, however if you want to save your results in a different folder just click on 'Select your save directory' and choose a folder as follow:

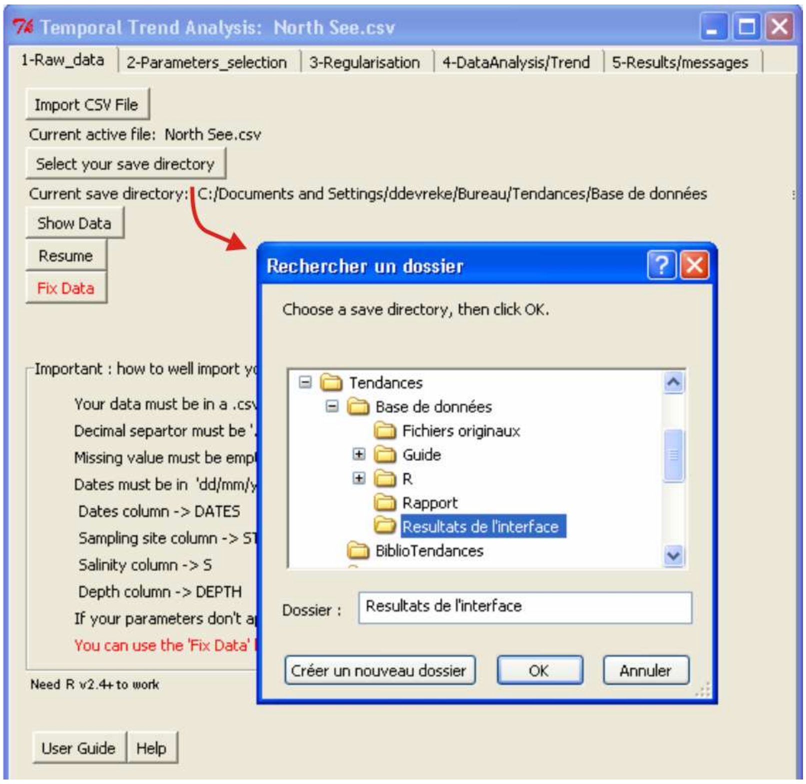


Figure 24. Panel 1 with select saved directory option.

You can also directly create a folder through this window.

From this basal directory, the program will automatically create an arborescence to save your files in function of the options you choose to perform analyses: SaveDirectory/FileName/Years/Parameter/DataInteraction/Aggregation.method/ (Fig. 25). Months, salinity and depths are not added in order to limit the arborescence declination and keeping a clear save path, therefore be careful to not overwrite your files if you change only these options between two analyses.

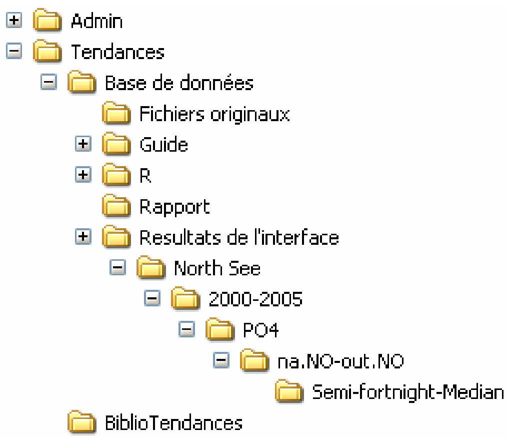


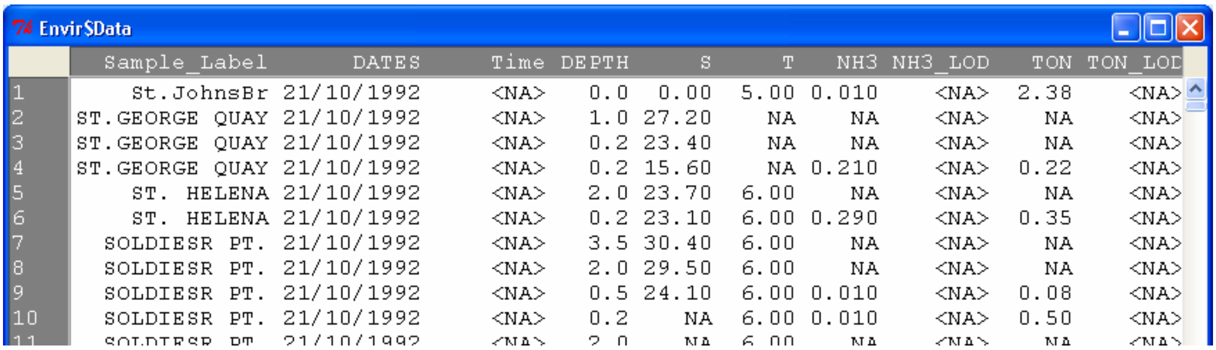
Figure 25. Save path arborescence as a function of selected options

Importing a new .csv file will reset your basal save directory to default.

Saved files are named with a specific nomenclature:
OriginalName_TestName_ Parameter.csv / .png.

2.3 Display your raw data

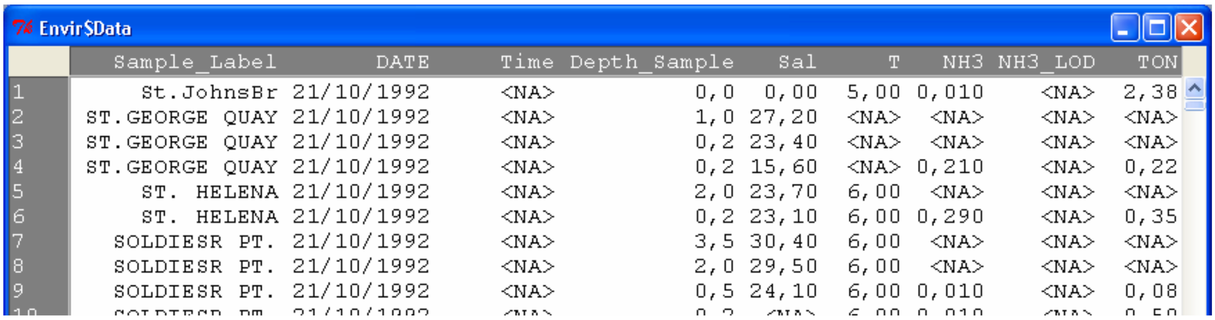
You can check your imported data by clicking on ‘ShowData’ button:



Envir$Data

	Sample_Label	DATES	Time	DEPTH	S	T	NH3	NH3_LOD	TON	TON_LOD
1	St.JohnsBr	21/10/1992	<NA>	0.0	0.00	5.00	0.010	<NA>	2.38	<NA>
2	ST.GEORGE QUAY	21/10/1992	<NA>	1.0	27.20	NA	NA	<NA>	NA	<NA>
3	ST.GEORGE QUAY	21/10/1992	<NA>	0.2	23.40	NA	NA	<NA>	NA	<NA>
4	ST.GEORGE QUAY	21/10/1992	<NA>	0.2	15.60	NA	0.210	<NA>	0.22	<NA>
5	ST. HELENA	21/10/1992	<NA>	2.0	23.70	6.00	NA	<NA>	NA	<NA>
6	ST. HELENA	21/10/1992	<NA>	0.2	23.10	6.00	0.290	<NA>	0.35	<NA>
7	SOLDIESR PT.	21/10/1992	<NA>	3.5	30.40	6.00	NA	<NA>	NA	<NA>
8	SOLDIESR PT.	21/10/1992	<NA>	2.0	29.50	6.00	NA	<NA>	NA	<NA>
9	SOLDIESR PT.	21/10/1992	<NA>	0.5	24.10	6.00	0.010	<NA>	0.08	<NA>
10	SOLDIESR PT.	21/10/1992	<NA>	0.2	NA	6.00	0.010	<NA>	0.50	<NA>

Figure 26. A correctly imported data set view with the ‘ShowData’ button in panel 1.



Envir$Data

	Sample_Label	DATE	Time	Depth_Sample	Sal	T	NH3	NH3_LOD	TON
1	St.JohnsBr	21/10/1992	<NA>	0,0	0,00	5,00	0,010	<NA>	2,38
2	ST.GEORGE QUAY	21/10/1992	<NA>	1,0	27,20	<NA>	<NA>	<NA>	<NA>
3	ST.GEORGE QUAY	21/10/1992	<NA>	0,2	23,40	<NA>	<NA>	<NA>	<NA>
4	ST.GEORGE QUAY	21/10/1992	<NA>	0,2	15,60	<NA>	0,210	<NA>	0,22
5	ST. HELENA	21/10/1992	<NA>	2,0	23,70	6,00	<NA>	<NA>	<NA>
6	ST. HELENA	21/10/1992	<NA>	0,2	23,10	6,00	0,290	<NA>	0,35
7	SOLDIESR PT.	21/10/1992	<NA>	3,5	30,40	6,00	<NA>	<NA>	<NA>
8	SOLDIESR PT.	21/10/1992	<NA>	2,0	29,50	6,00	<NA>	<NA>	<NA>
9	SOLDIESR PT.	21/10/1992	<NA>	0,5	24,10	6,00	0,010	<NA>	0,08

Figure 27. Dataset with labels and decimal separator issues view with the ‘ShowData’ button.

2.4 Edit your data – solve some importation issues

In case you have importations issues, you can edit your data with the ‘Fix Data’ button.

There are two situations where data importation can be corrupted:

- The csv file has been badly created; label and decimal separator don’t correspond with the interface standard (Fig. 27). Therefore only column labels can be modify with ‘Fix Data’. Back to your csv builder and check §B.

- The csv file has been well created following §B recommendations but parameters still don’t appear in panel 2. Therefore the problem is the category of the variable. This is an R importation issue that can be solve using ‘Fix Data’, then change the category to numeric (parameters) by clicking on column labels (if column with values with ‘,’ as decimal separator are selected as numeric you will obtain missing values). Dates and sampling site are character category.

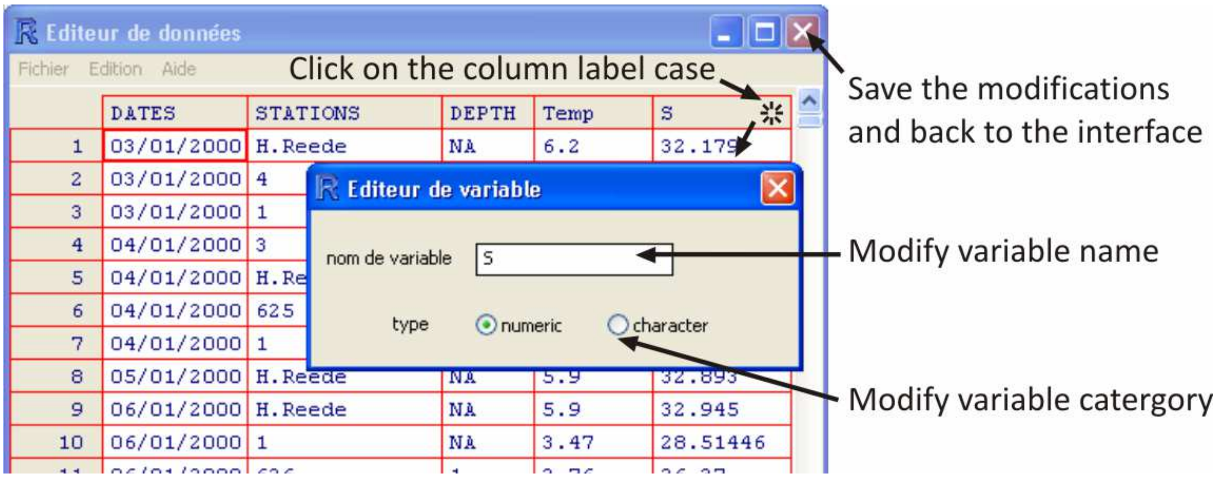


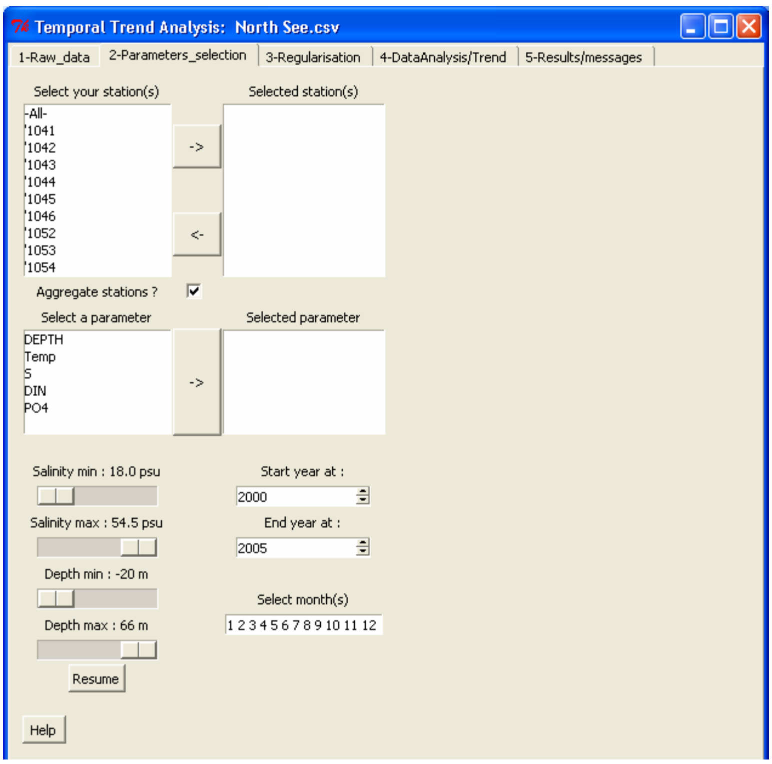
Figure 28. Data editor while 'FixData' button is used.

When you exit the Fix Data interface, your new dataset is automatically read by the interface and a new csv file is saved in your computer with the nomenclature FileName_fixed.csv. Unfortunately data category cannot be saved in csv file.

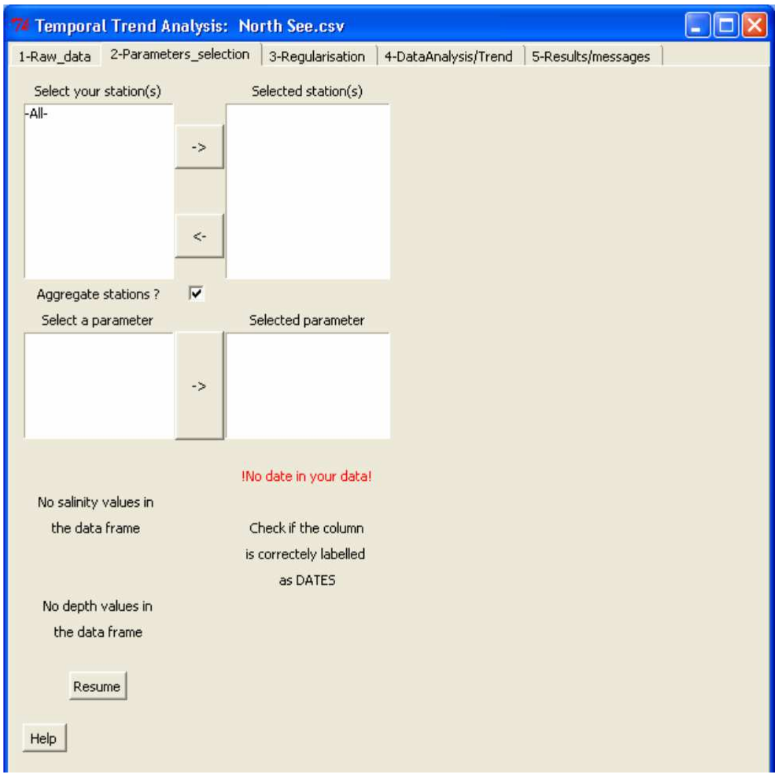
Editing your data will not change your save directory.

3. Parameters selection panel

If columns have been correctly labelled and categorised (see precedent chapters), lists, sliders and frames should be automatically filled with appropriate values.



Otherwise:



Let see details of each of these parts in the next chapters.

3.1 Sampling stations

Sampling station represent the area where your data have been measured, in your database they can be redundant (like for fixed station) or unique (like for transect). You may also want to perform analysis on all of your stations or only on a bunch of them. To select the stations you want to analyse you can select them using the arrows between the two selection boxes (Fig. 29): just click on the station in the left box and click on the top arrow, the selected station will appear in the right box. This support multiple selection using the 'Ctrl' key or by dragging the cursor. If you want analyse all stations just select -All- in the left box. To remove stations just select them in the right box and click on the bottom arrow, it also support multiple selection.

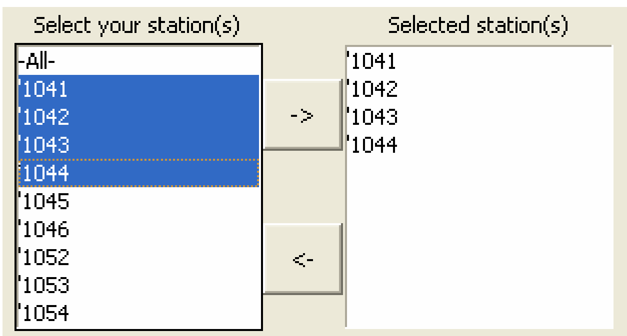


Figure 29. Selection boxes for sampling stations in panel 3, with selection arrows between.

By default the data of all stations will be melt for analyse (that certainly why you select them together for analyse). However if you don't want to melt data but keep stations independently from each other you can uncheck the appropriate box (Fig. 30). Therefore you will be unable to perform analysis as distinct regularized time series will not be built from separate stations. You will be only able to summarise the aggregated data table, and display a table and a plot of the aggregated data.

Figure 30. Aggregate stations check box in panel 3.

3.2 The parameters

The process to select the parameter you want analyse is the same as for the stations, except that only one parameter can be selected (don't support multiple selection) and there is only one arrow for selection (no remove arrow) (Fig. 31). To replace an already selected parameter by another one just select the new parameter in the left box and click on the arrow, it will automatically replace the previous one in the right box.

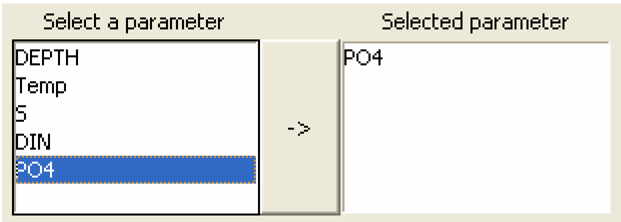


Figure 31. Selection boxes for parameter in panel 3, with selection arrow between.

3.3 Depth and salinity

In some cases you will need to perform analyses at specific depth or salinity (which characterize water masses). In the panel 2, there is 4 sliders to select these salinities and depths (Fig. 32). By default these sliders display the maximum and minimum values of salinity and depth in your dataset (if they exist). If you don't modify these values all data will be take into account for analyse, including data at missing salinity and depth. You can modify these values by sliding left or right the cursors or by clicking on the area next to the cursor to have a more accurate increment (+/- 0.5 unit), data at missing salinity or depth will be exclude. You can perform analyse at a unique depth or salinity by giving the same min and max values.

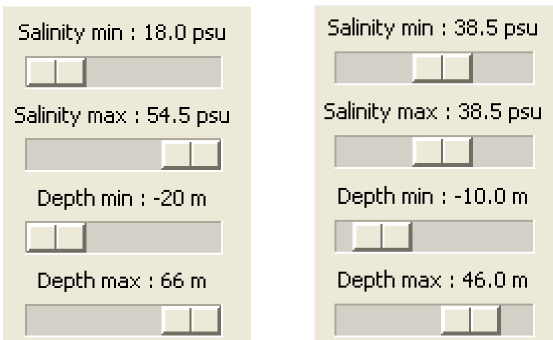


Figure 32. Sliders to select salinity and depth at which you want to perform analysis. Left figure: default options; right figure: selection of only one salinity and a range of depth (4 to 25.5m)

3.4 Years and months

As for salinity and depth you can modify the years and months at which you want analyse your data. By default the two lists in panel 2 display first and last years of you dataset and the 12 numeric months (Fig. 33). You can modify years just by clicking on the arrows or by typing it. You can delete or add months (no matter the order) and there must be a space between months.

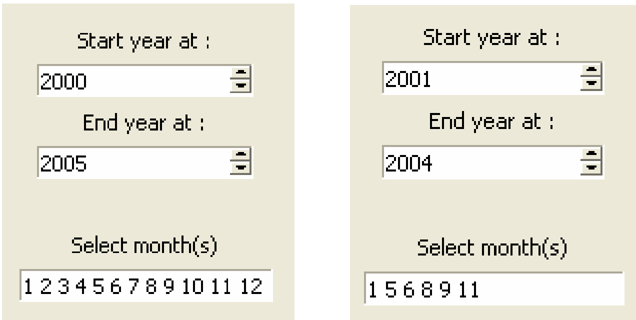


Figure 33. Boxes to select year range and months at which you want to perform analysis. Left figure: default options; right figure: analysis between 1995 and 2006 at 7 different months.

4. Time series building panel

The third panel focused on time series regularization before proceeding to temporal trend analysis (Fig. 34).

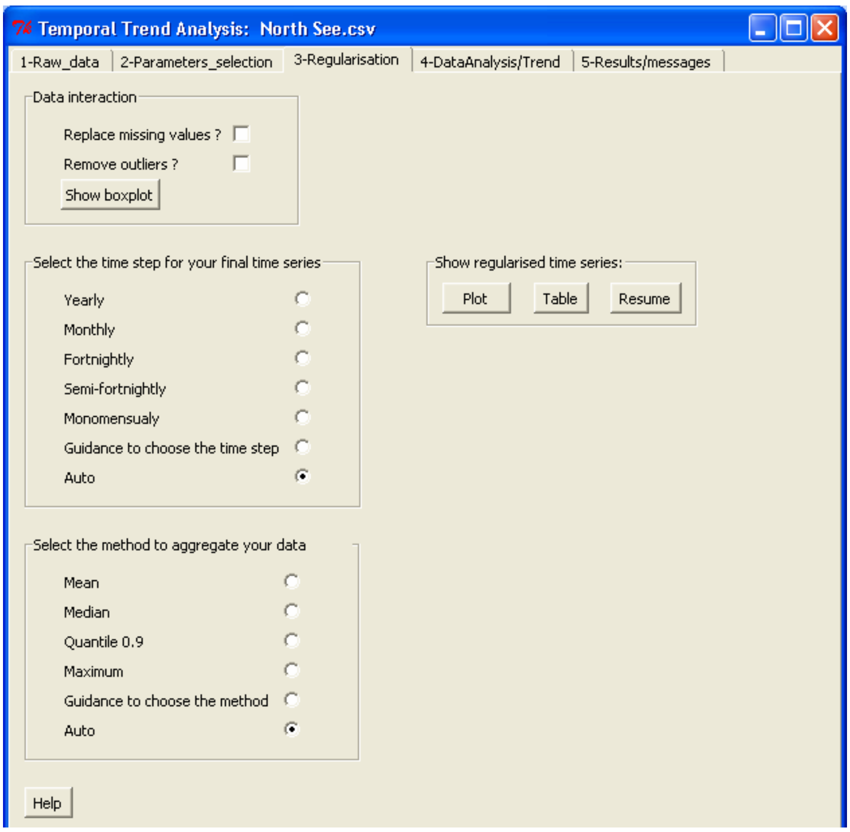


Figure 34. Panel 3 with default options.

4.1 Missing values and outliers

The first frame of the third panel shows 3 different options to deal with missing values and outliers of your data (Fig. 35).

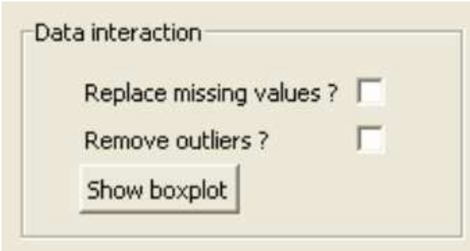


Figure 35. Data interaction frame options in panel 3.

Some statistics and diagnostic tools require that time series are regular and thus have regularly spaced measurement and contains no missing values. The option, 'Replace missing values', will replace this missing values from the time series (and not from the raw data) by values calculated from the data distribution. The final time series will depend on the method of aggregation (see §4.2 and §4.3); therefore missing values will be calculated from aggregated data in two successive steps.

- Time series generally present strong autocorrelation (see 4.1), in this case value at time *t* depend on values at time *t*+1. Therefore missing values can be estimates (predicated) from linear regression (or more complex regression) between values around the missing value, however this is relevant only when few data are missed, long period of missing values cannot be replaced using this method (cannot efficiently extrapolate seasonal fluctuations).
- When missing values are present over a long period, and that there is a really need to replace them, they can be replace by the median of data from the same cycle (e.g. month, week, year, depending on the time step choose), inversely this method is less relevant than regression for shorter period of missing values (it loose the dependency due to autocorrelation).

The present interface used a combination of both methods; the linear regression method to replace missing values over 3 successive unit of time and the median method for longer period of missing values. The median method act first to reduce the lag between missing values and to allow the regression method more frequently. Missing values at the beginning and at the end of the series are replaced using the median method if possible or are ignored.

Data distribution frequently contains outliers; these outliers are due, for example, to error of measurements than have not been corrected or extreme natural event. In some cases these outliers can greatly influenced statistical analysis comparatively to the rest of values and it should be interesting to remove them. The second option present in the frame (Fig. 35) allows you to remove these outliers and to save them in a separate csv file (in case you need to identify them). The method used to identify outliers is the boxplot method by years. For each year, values over Q3+1.5(Q3-Q1) and under Q1-1.5(Q3-Q1) are considered as outliers. The 'Show boxplot' button displays the box and whiskers plot with outliers.

Both options, missing values and outliers, can be check together or independently, outliers will be always removed first and missing values in second places.

4.2 Time step selection of your time series

Raw database generally display discontinuous time series, with missing values and variable measurement frequency between values. Temporal trend analyses generally need regularised time series to be perform. To build such regularised time series, the interface will aggregate raw data from the same selected period (week, month, year…) using a selected method (mean, median, max value…). These different options are available in frame 2 and 3 of the third panel (Fig. 36 and 37).

Frame 2 'Select the time step for your final time series' let you select the time step at which the interface will aggregate your data to build the regularized time series. Seven options are available (Fig. 36); the 5 first options are classic frequencies, yearly (all data by years), monthly (all data by month), fortnightly (all data by 15 days), semi-fortnightly (all data by 7 days) and monomensualy (all data by month, all years including).

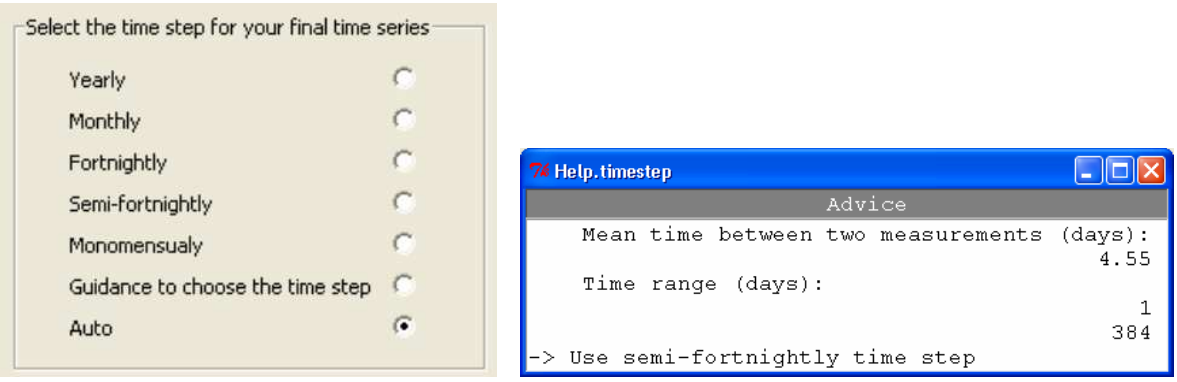


Figure 36. Time step selection frame and time step advice (guidance option).

Just remember: you have to choose a time step in relation with the sampling frequency of your data in order to keep the maximum of information without creating to many missing values. The option 'Guidance to choose the time step' will suggest a balanced choice by computing the mean time and the minimum and maximum period that separates two successive measurements in your database. Arbitrary, if mean time between two measurements is under 10 days, the interface advice the semi-fortnight time step; if mean time is between 10 and 23 days, fortnight time step is advice; between 23 and 60 days, monthly time step is advice and over 60 days annual time step is advice. Monomensual time step is only available in manual choice. You are free to follow these suggestions or selecting another time step. The auto option (default option) will automatically apply the advice without display the suggestion. In some case, this method of aggregation is enough to removed all missing values in the regularised time series without using the 'Replace missing values' option (§4.1).

4.3 Method of aggregation of your time series

The second frame 'Select the method to aggregate your data' (Fig. 37) let you choose the method with which data will be aggregated at the time step previously set. Four methods are available: by averaging the data, by selecting the median of the data, by selecting the quantile 90% of the data or the maximum of the data of the same time step. The guidance option will also suggest the method that fit the best the original data distribution. The interface will compare data distribution obtain with each method (at selected time step) with the raw data distribution using a Wilcoxon non parametric test, the comparison with the highest p-value (less significant difference) will determine the best method. You are also free to follow these suggestions or selecting another method. The auto option (default option) will automatically apply the advice without display the suggestion.

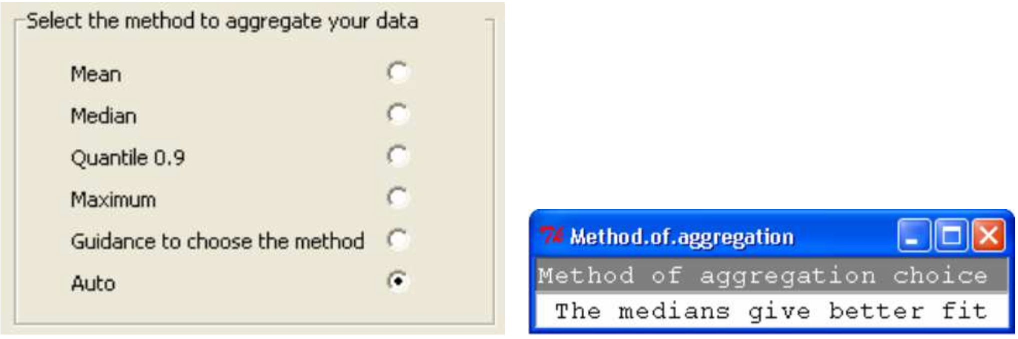


Figure 37. Aggregation method selection frame and method advice (guidance option).

4.4 Visualised your regularized time series

The fourth frame 'Show regularised time series' (Fig. 38) let you display your newly build regularised time series through a plot, a table or a summary. Plot and table will be saved in your computer. The table display column labels which vary as a function of the time step you selected. For all time steps, the fist column label is the parameter you selected for analysis, other column contain temporal indications. The years and months are indicated in the eponym columns. The week number within months (2 weeks/month with fortnight time step and 4 weeks/month with semi-fortnight time step) are indicated in the 'week.month' column. The week number within year (base on fortnight week) is indicated in the 'week.year' column (24 weeks/year for forthright time step and 48 for semi-fortnight time step). The 'time' column count the time between regularised measurements (the value of unit depend on the selected time step).

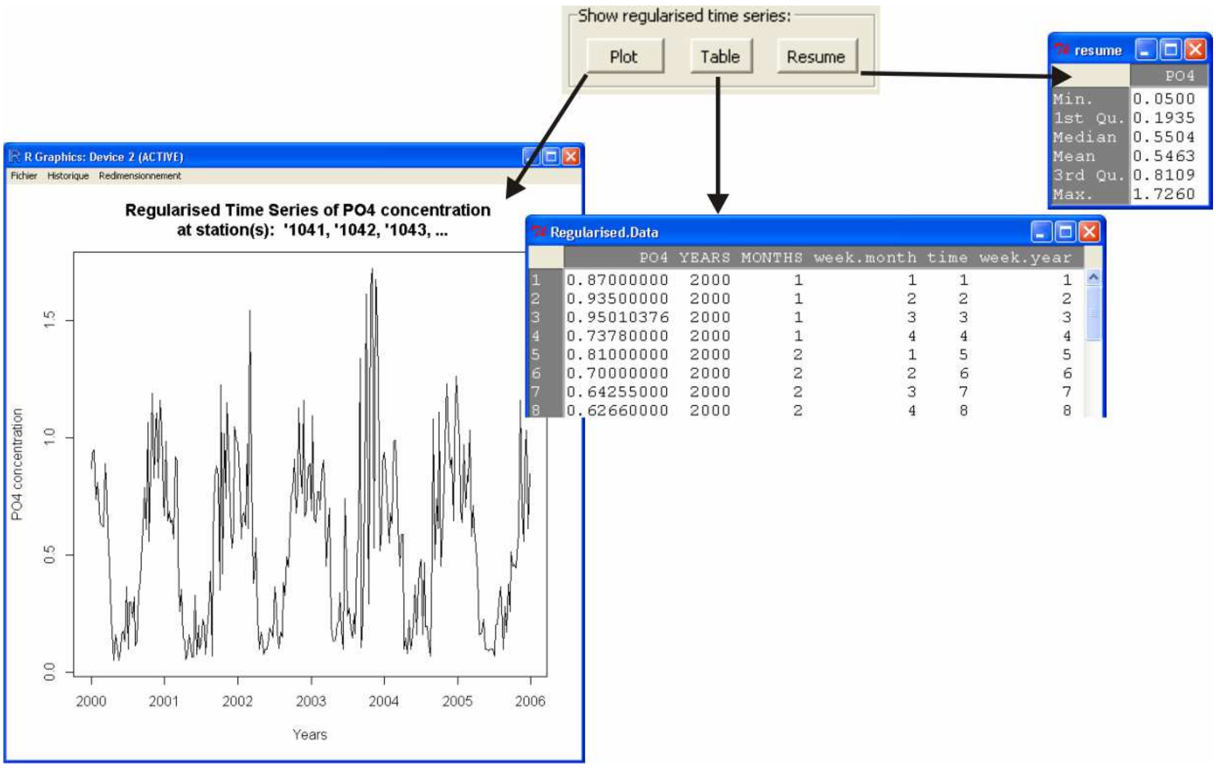


Figure 38. Regularised time series display options frame.

5. Diagnostics, statistics and results panel

5.1 Diagnostic tools

The options present in the first frame of the forth panel ‘Diagnostics (optional)’ (Fig. 39) are not require to perform temporal trend analysis but give additional information that can be useful to explain variations in the time series.

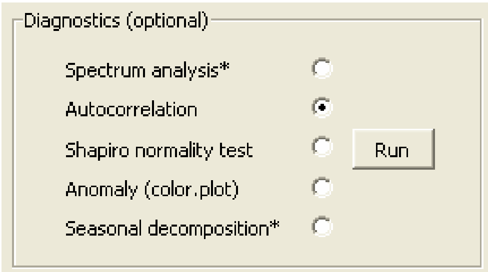


Figure 39. Diagnostic frame in panel 4

- **Spectrum analysis**: it allows to estimates the spectral density (discrete Fourier transform) of a time series and to display a periodogram with confidence interval (Fig. 40). Spectrum analysis is concerned with the exploration of cyclical patterns of data. The purpose of the analysis is to decompose a complex time series with cyclical components into a few underlying sinusoidal (sine and cosine) functions of particular wavelengths. Then you can determine the frequency of each cycle (spectrum) present in the time series from the most important to the less. In our program the basic unit is one year. Spectrum analysis cannot be perform with missing values and result is not influenced by the time step at which you build the time series. This is the spectrum() function of the ‘stats’ package (http://127.0.0.1:20999/library/stats/html/00Index.html). For more information about Fourier transform and periodogram in R see also Shummway and Stoffer (2006).

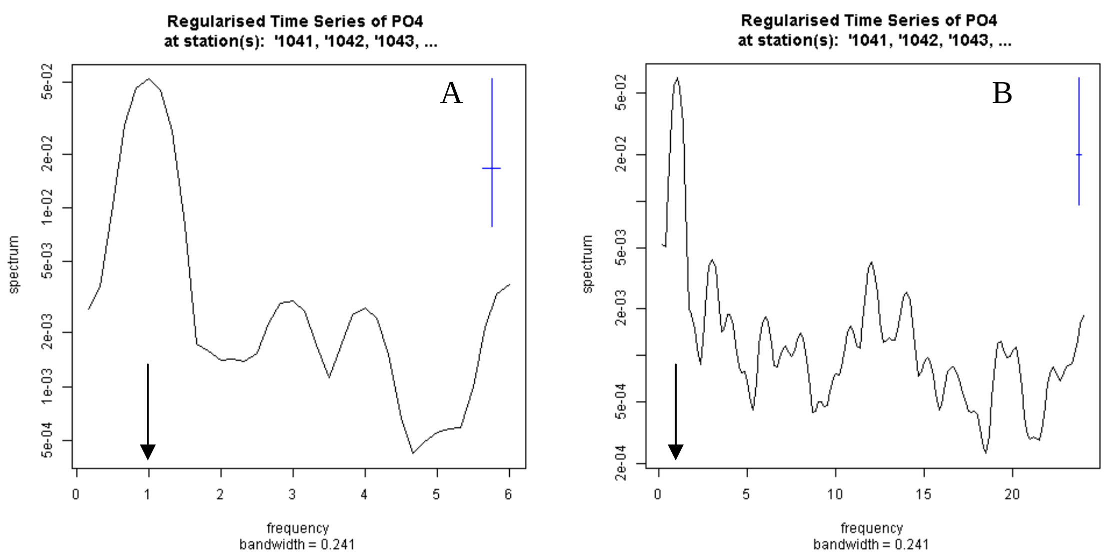


Figure 40. A. Spectrum density of monthly regularized time series of PO4 concentration in the North Sea between 2000 and 2005. B. Same at semi-fortnightly scale. Blue vertical bar is confidential interval. Save as “North See_Spectrum_PO4.png”.

Interpretation:

In this case the identify cycle (by arrow) is the annual cycle that show the highest spectrum at frequency = 1. It's likely the most common frequency you will certainly obtain as it is due to seasonality from year to year frequent in biological processes. Varying the time step of data aggregation will not degrade the signal as far as it stay under the mean sampling frequency of your data.

- **Autocorrelation**: computes (and plots with confidence interval at 0.95) estimates of the autocorrelation function (Fig. 41). As for spectrum analysis the most frequently highest autocorrelation will be observed at lag 1 (1 year) whatever the time step, only the number of subdivision between lags are determined by the time step selected to build the time series. This is the acf() function of the 'stats' package. For more information about autocorrelation function in R see also Shummway and Stoffer (2006).

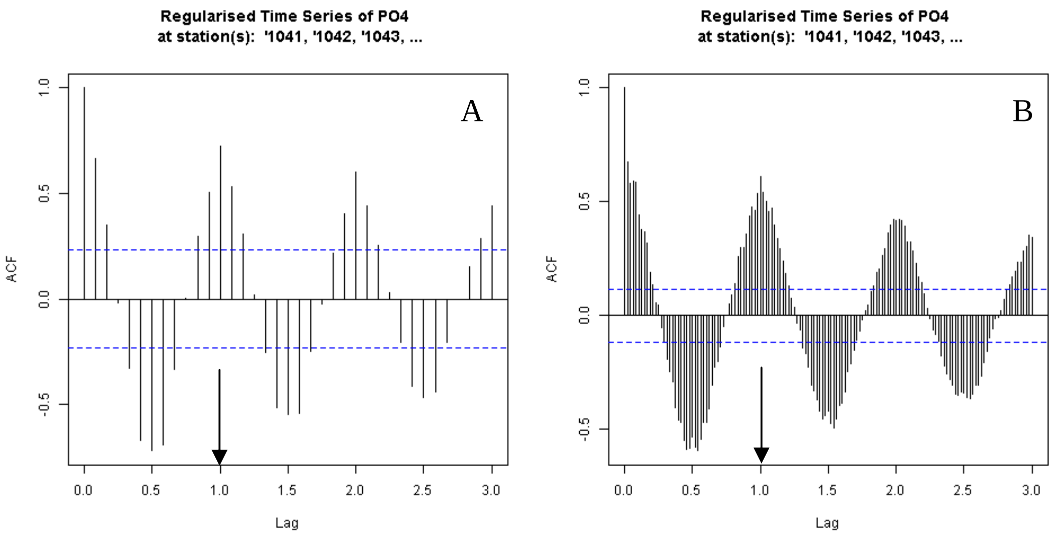


Figure 41. A. Autocorrelation of monthly regularized time series of PO4 concentration in the North Sea between 2000 and 2005. B. Same at semi-fortnightly scale. Save as "North See_AutoCorr_PO4.png".

Interpretation:

In this case maximum positive autocorrelation is obtain for a lag of 1 (=1 year in our program) whatever the time step choose to aggregate the data. Autocorrelation is significant every 0.5 lag (6 months) up to 3 lags (3 years).

- **Shapiro normality test**: Shapiro–Wilk test tests the null hypothesis that a sample came from a normally distributed population (Null hytpothesis: follow a normal distribution, thus if the p-value is lower than the chosen alpha level (0.05 in our program), the sample don't follow a normal distribution).

- **Anomaly (color.plot)**: Computes time series anomalies by X_{ij} - X_i, with X_{ij} value of the parameter X at the period i of the year j and X_i the median of the parameter X for the period i (all year mixed) (Fig. 42). The option will produces a contour plot with the areas between the contours filled in solid colour. A key showing how the colours map is shown to the right of the plot. Red colours show positive anomaly and blue colours negative anomalies. White areas occur when there are missing values. Works only with time series build at monthly, semi-fortnight and fortnight time step.

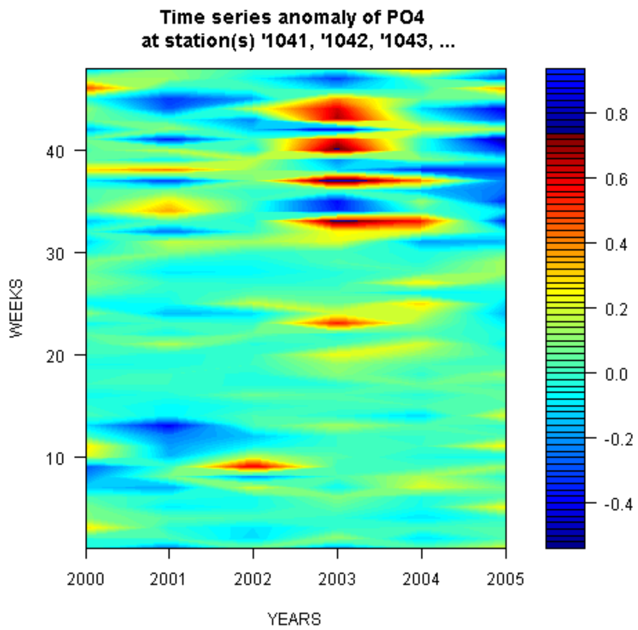


Figure 42. Color plot of PO4 concentration anomalies in the Irish Sea between 2000 and 2005 at monthly scale. Cold and hot areas represent respectively negative and positive anomaly. Very strong positive anomalies can be represented by blue colours (like here), it is due to the labelling range method actually use by the function. Save as "North See_ColorPlot_PO4.png".

Interpretation:

Anomalies (lower than 'normal') of PO4 concentration can be observed in 2001, 2003 and 2005 at the beginning and at the en of the year. Positive anomalies (higher than 'normal') are observed in 2002 and particularly in 2003.

- **Seasonal decomposition**: Decompose and plot a time series into seasonal, trend and irregular components using loess (Fig. 43). The seasonal component is found by loess smoothing (locally weighted scatterplot smoothing) the seasonal sub-series. The remainder component is the residuals from the seasonal plus trend fit. The seasonal values are removed, and the remainder smoothed to find the trend. This is the function stl() of the 'stats' package.

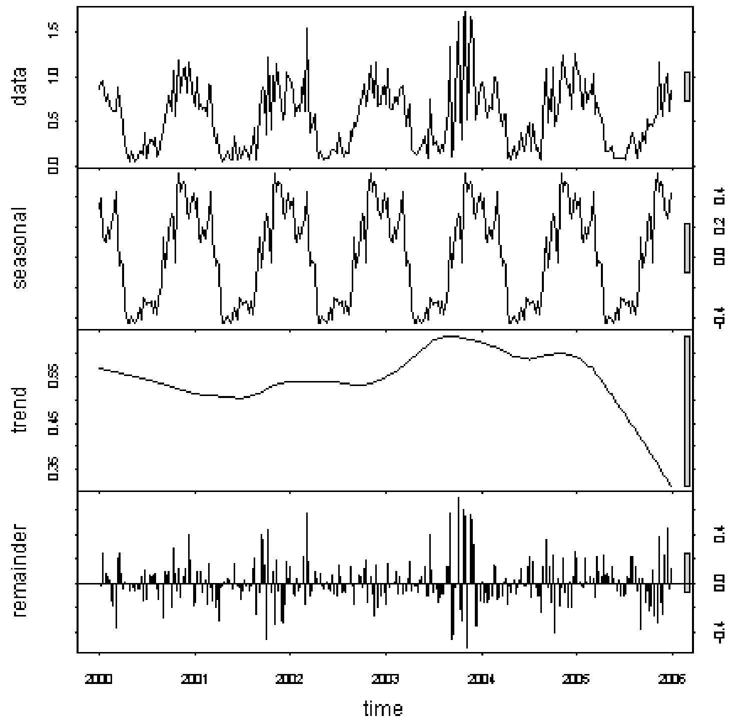


Figure 43. Plots obtain with the decomposition option with from top plot to bottom plot: regularized time series, seasonal component, global trend, and remainder.

Interpretation:

In this case PO4 concentration variations show a very high seasonal signal (2d plot). The overall trend shows a slight increase in 2004, the strong decrease during the last year should be an artefact due to the smoothing method (periodic). Remainder (4th plot) show few seasonality pattern, so the major part of PO4 concentration variation is due to the seasonality.

5.2 Temporal trend tests

The second frame of the fourth panel display the available tests to perform temporal trend analysis (Fig. 44).

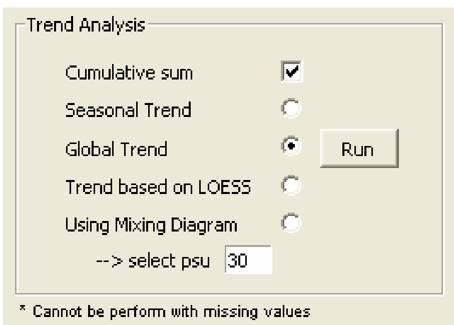
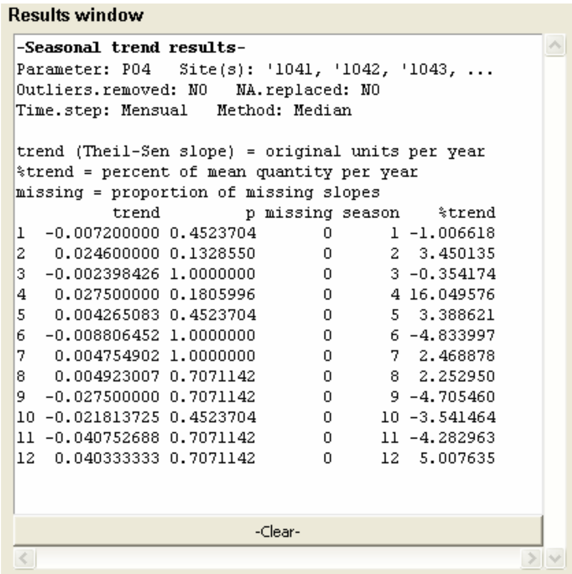


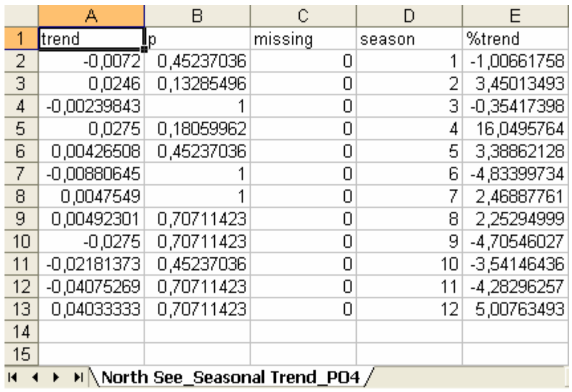
Figure 44. Trend analysis frame in panel 4.

- **Seasonal Trend**: perform a Seasonal Kendall test on the time series with details of trend between months (Fig. 45). The Seasonal Kendall test takes the seasonal variability into account during trend calculation. This seasonality is not limited to a cycle of 12 months but is extended to the time step you choose to build your time series. The trend value is obtained by calculating a Mann-Kendall tests between season (test the presence of a trend) and performing a Sen's Slope estimator to estimate a value of this trend (median between ranks). This method comes from the 'wq' package (http://cran.r-project.org/web/packages/wq/index.html). For more information about Kendall test see Hirsch et al. (1982) and Hirsch and Slack (1984). Results are display in panel 5 and saved in a csv file with: trend column = trend of the parameter at the selected time step (season column) in original unit per year; p column = significance of the slope; missing column = proportion of missing slope at the time step; %trend column = percent of mean quantity per year at the selected time step; season column = counting of time step succession (1 to 12 for monthly, 1 to 24 for fortnightly etc…).



```
Results window
-Seasonal trend results-
Parameter: PO4   Site(s): '1041, '1042, '1043, ...
Outliers.removed: NO   NA.replaced: NO
Time.step: Mensual   Method: Median

trend (Theil-Sen slope) = original units per year
%trend = percent of mean quantity per year
missing = proportion of missing slopes
          trend         p missing season    %trend
1  -0.007200000 0.4523704       0      1 -1.006618
2   0.024600000 0.1328550       0      2  3.450135
3  -0.002398426 1.0000000       0      3 -0.354174
4   0.027500000 0.1805996       0      4 16.049576
5   0.004265083 0.4523704       0      5  3.388621
6  -0.008806452 1.0000000       0      6 -4.833997
7   0.004754902 1.0000000       0      7  2.468878
8   0.004923007 0.7071142       0      8  2.252950
9  -0.027500000 0.7071142       0      9 -4.705460
10 -0.021813725 0.4523704       0     10 -3.541464
11 -0.040752688 0.7071142       0     11 -4.282963
12  0.040333333 0.7071142       0     12  5.007635
-Clear-
```



	A	B	C	D	E
1	trend	p	missing	season	%trend
2	-0,0072	0,45237036	0	1	-1,00661758
3	0,0246	0,13285496	0	2	3,45013493
4	-0,00239843	1	0	3	-0,35417398
5	0,0275	0,18059962	0	4	16,0495764
6	0,00426508	0,45237036	0	5	3,38862128
7	-0,00880645	1	0	6	-4,83399734
8	0,0047549	1	0	7	2,46887761
9	0,00492301	0,70711423	0	8	2,25294999
10	-0,0275	0,70711423	0	9	-4,70546027
11	-0,02181373	0,45237036	0	10	-3,54146436
12	-0,04075269	0,70711423	0	11	-4,28296257
13	0,04033333	0,70711423	0	12	5,00763493
14					
15					

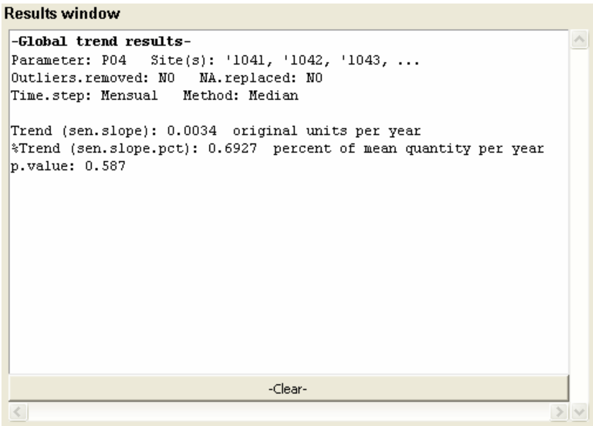
North See_Seasonal Trend_PO4

Figure 45. Results of the Seasonal Trend analysis (monthly scale) display in the panel 5 (left) and saved in a csv file (right) as "OriginalName_Seasonal Trend_Parameter.csv".

Interpretation:

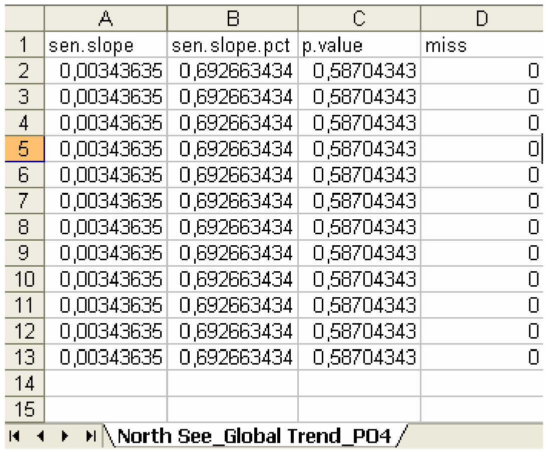
The trend of PO4 concentration vary as a function of the considered month, season 1 (first row = January in our example) show a low trend of -0.007 ng/L/year (trend column) that correspond to -1% ng/L/year (column %trend) with a p-value of 0.45 (quite good). No missing values are observed in our time series (missing column). The most important trend is observed at season 4 (April) that shows a increasing trend of 16% per year but have the lowest p-value (0.18).

- **Global Trend**: same as above but give the general trend without detail (Fig. 46). Also take the seasonal variability into account. Sen's Slope estimator is for the totality of the time series. This method comes from the 'wq' package (http://cran.r-project.org/web/packages/wq/index.html).



```
Results window
-Global trend results-
Parameter: PO4   Site(s): '1041, '1042, '1043, ...
Outliers.removed: NO   NA.replaced: NO
Time.step: Mensual   Method: Median

Trend (sen.slope): 0.0034  original units per year
%Trend (sen.slope.pct): 0.6927  percent of mean quantity per year
p.value: 0.587
-Clear-
```



	A	B	C	D
1	sen.slope	sen.slope.pct	p.value	miss
2	0,00343635	0,692663434	0,58704343	0
3	0,00343635	0,692663434	0,58704343	0
4	0,00343635	0,692663434	0,58704343	0
5	0,00343635	0,692663434	0,58704343	0
6	0,00343635	0,692663434	0,58704343	0
7	0,00343635	0,692663434	0,58704343	0
8	0,00343635	0,692663434	0,58704343	0
9	0,00343635	0,692663434	0,58704343	0
10	0,00343635	0,692663434	0,58704343	0
11	0,00343635	0,692663434	0,58704343	0
12	0,00343635	0,692663434	0,58704343	0
13	0,00343635	0,692663434	0,58704343	0
14				
15				

North See_Global Trend_PO4

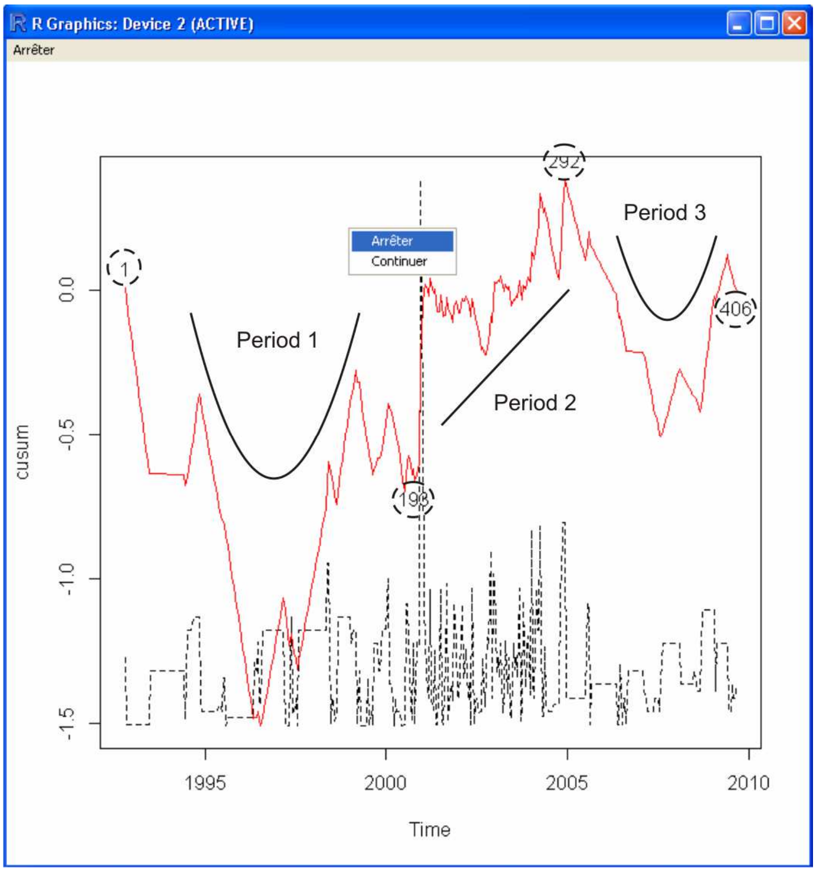
Figure 46. Results of the Global Trend analysis (monthly scale) display in the panel 5 (left) and saved in a csv file (right) as "OriginalName_Global Trend_Parameter.csv".

Interpretation:

The global trend of PO4 concentration is low with 0.0034 ng/L/year, resulting from the strong variation among months (Fig. 44). It's represent 0.7% ng/L/year with a p-value of 0.58. The time series shows no missing value (miss column in the right figure).

- **Cumulative sum**: plot a cumulative sum curve of the time series and allow to manually identifying changes in the tendency in your time series (create successive periods) (Fig. 47). This method come from the cusum function of the 'pastecs' package (http://cran.r-project.org/web/packages/pastecs/). For more information about cusum function see also Ibanez et al. (1993). Once the periods have been identify, the program perform the Global or Seasonal Trend test (as selected by the user) on each period. Only Global or Seasonal Trend test can be perform, if you select another test the interface will return a warning message asking you to choose a Kendall family test. The cumulative sum curve is automatically calculated from your time series with missing values removed (cannot work with outliers), however the trend calculations are perform on the time series build with your own options (so even with no replacement of missing values).

  Once the option selected and the button 'Run' clicked, the plot window with cusum curve in red will appear. Then you have to left-click on this curve to identify the different points where the tendency changes. Once all points have been identify, right-click on the plot and check 'Arrêter'/'Stop', analysis will start automatically and result will be display on panel 5. You can close the plot window. Be careful to not close the plot window before checking right-click-'Arrêter'/'Stop', it will cause the interface to stop functioning.



Results window

```
-Local trend results (global)-
Parameter: NH3   Site(s): AV000, AV010, AV020, ...
Outliers.removed: NO   NA.replaced: NO
Time.step: Fortnight   Method: Median

Period 1 :  1992 - 2000
Trend (sen.slope): -1e-04  original units per year
%Trend (sen.slope.pct): -0.1986  percent of mean quantity per year
p.value: 0.8333

Period 2 :  2000 - 2004
Trend (sen.slope): 6e-04  original units per year
%Trend (sen.slope.pct): 1.075  percent of mean quantity per year
p.value: 0.725

Period 3 :  2004 - 2008
Trend (sen.slope): 0.005  original units per year
%Trend (sen.slope.pct): 10.645  percent of mean quantity per year
p.value: 0.2207
```

Figure 47. Top figure: NH_3 concentration variations in the Irish Sea (hatched black line) with cumsum plot (red) and different periods identified (solid black lines). Bottom figure: results of a global trend apply on each of these periods display in panel 5.

Interpretation:

The global trend of NH_3 concentration is very low during the two first periods, it show a slightly higher increase during the last period, 2004-2008, with 10% mg/L/year. This decrease is observed only during this period and not between periods. In fact the trend could be a decrease between period 2 and 3 with an increase during period 3.

Trend based on LOESS: A loess smoothed curve of the regularised time series is considered to perform a Global Trend test instead of the time series itself (Fig. 48). This is the loess() function of the 'stats' package.

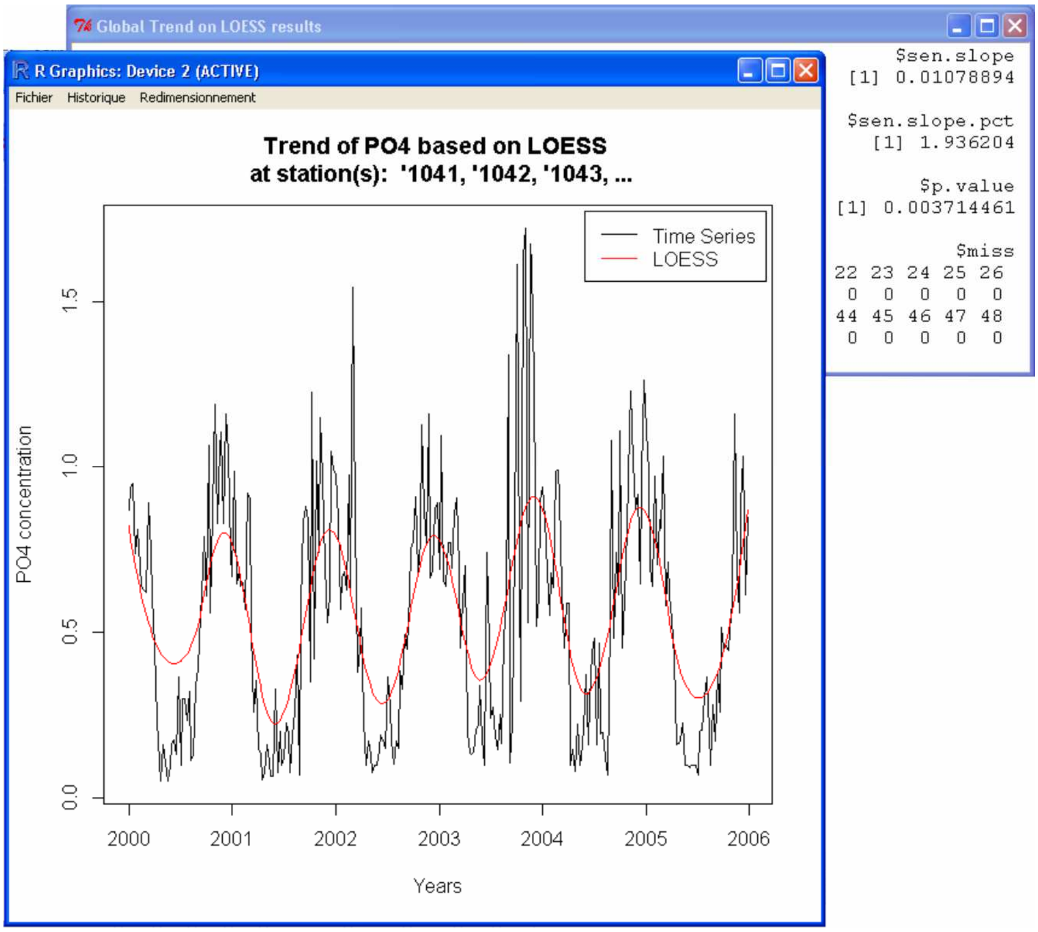


Figure 48. Plot of regularised time series of PO4 concentration (black line) and loess smoothing (red). Background table displays result of global trend test perform on loess smoothed data, results are also shown in panel 5.

- **Mixing Diagram**: To consider temporal trend of nutrient concentration in a salinity gradient, a widely used method consist to used monthly normalized concentration of nutrient at fixed salinity (generally 30) instead of raw data to perform temporal analyses (OSPAR, 2002). To normalize, a monthly linear regression is done between raw salinity and nutrient concentration (one regression per month). From these linear regression equations, normalized concentrations of nutrient are estimated at the salinity you enter in the text box 'select psu' (Fig. 44). Thus, a monthly time series is build using the new normalized concentrations instead of the aggregated raw data (this test is independent from the time step and aggregation method selected on panel 3). A Global Trend analysis is perform on this time series. Such method is generally used to analyse variation in winter concentration of nutrient. This can be easily obtain in the interface by selecting winter months (1 2 3) in panel 2 and perform the Mixing Diagram analysis.

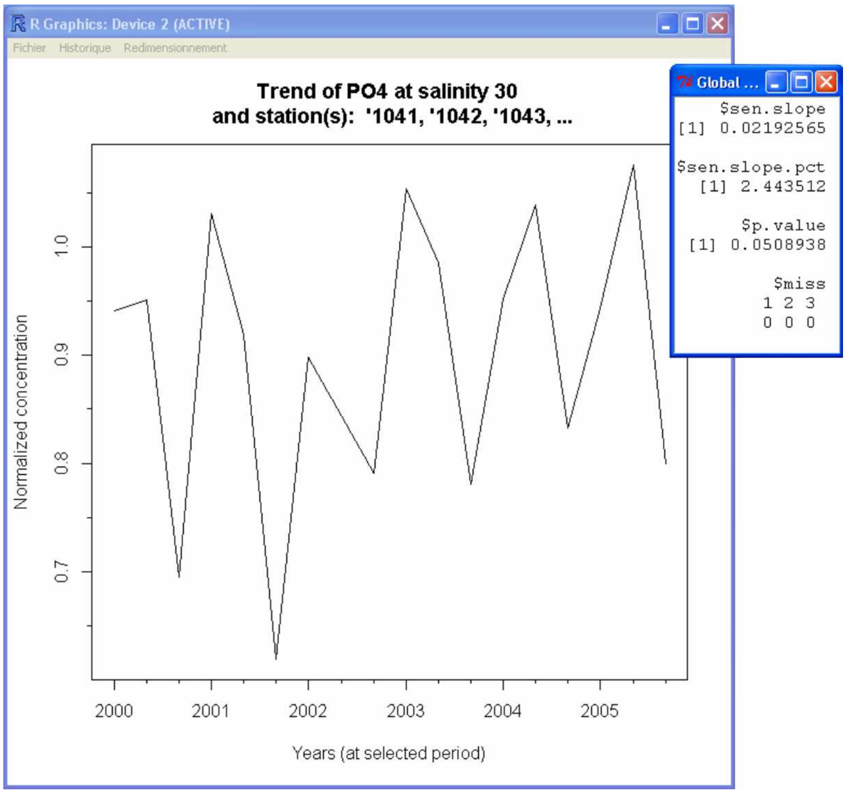


Figure 49. Plot of normalized winter concentration of PO4 (3 data per year) at salinity 30 between 2000 and 2005 from the North Sea. Table displays results of global trend test on these data, results are also shown in panel 5.

All mixing diagram (linear regression between nutrient and salinity, Fig. 50) are saved for each months and year (if possible) but not display, only the final results (plot of the time series and Global Trend results) is display by the interface (Fig. 49). A csv table containing all normalized concentration of nutrient per months/years is also generated and saved.

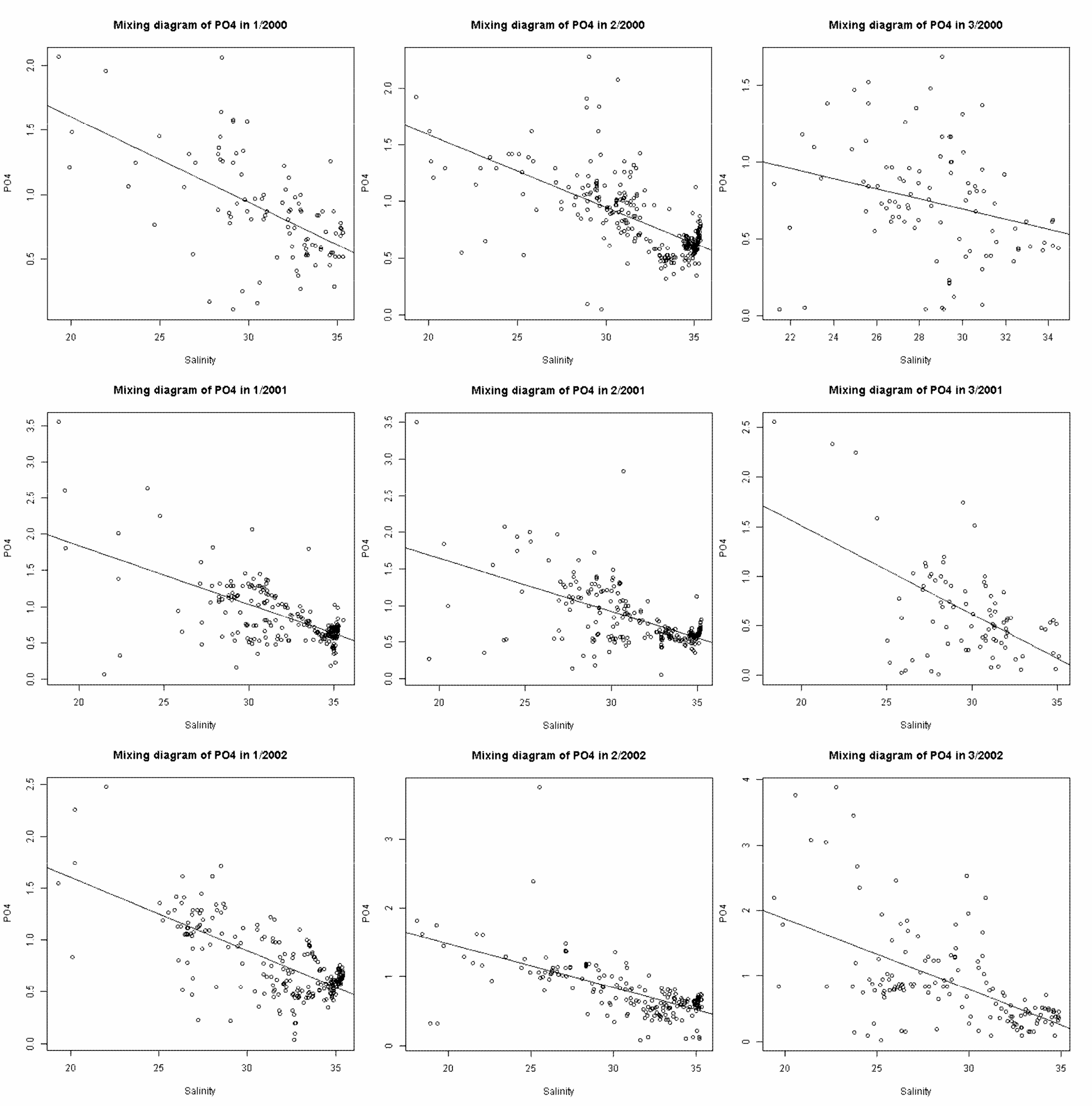


Figure 50. Plot of PO4 concentration against salinity by month/year with linear regressions.

6. Panel results and messages

The last panel just displays results of your analysis (as mention in the other sections) in the bottom text box and other messages in the top text box. Results text box displays more information than the saved csv file like the option you chose to build the time series.

There is an option to clear the text boxes, but be careful you cannot cancel this clean up action. So if you want to save some specific results, do it before cleaning the box. So you can save the text in the box by simply copy-paste (ctrl-c / ctrl-v) in the bloc-note and then import the data in Microsoft Excel.

7. Extra advices

Importing a new .csv file, changing the save directory or editing your data with 'Fix data' will reset all your options to default. Changing stations or any other parameters in panel 2 will not change options in panel 3 and 4, which allow performing rapid analysis among the different parameters of your dataset. Also for rapid analysis you can choose your parameter and stations in panel 2 and passed directly at the panel 5, balanced options in panel 3 are pre-selected by default.

References

Hirsch, R.M., J.R. Slack, and R.A. Smith, 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18(1):107-121

Hirsch, R.M. and J.R. Slack, 1984. A non parametric trend test for seasonal data with serial dependence. Water Resources Research 20(6)727-732.

Ibanez, F., J.M. Fromentin and J. Castel, 1993. Application of the cumulated function to the processing of chronological data in oceanography. Comptes Rendus de l'Académie des Sciences Serie III - Sciences de La Vie - Life Sciences. 316:745–748.

OSPAR, 2002, Common assessment criteria, their (region specific) assessment levels and guidance on their use in the area classification within the comprehensive procedure of the common procedure. OSPAR 02/8/2-E.

Shumway R.H. and D.S. Stoffer, 2006. Time series analysis and its applications with R examples, 2nd edn. Springer.