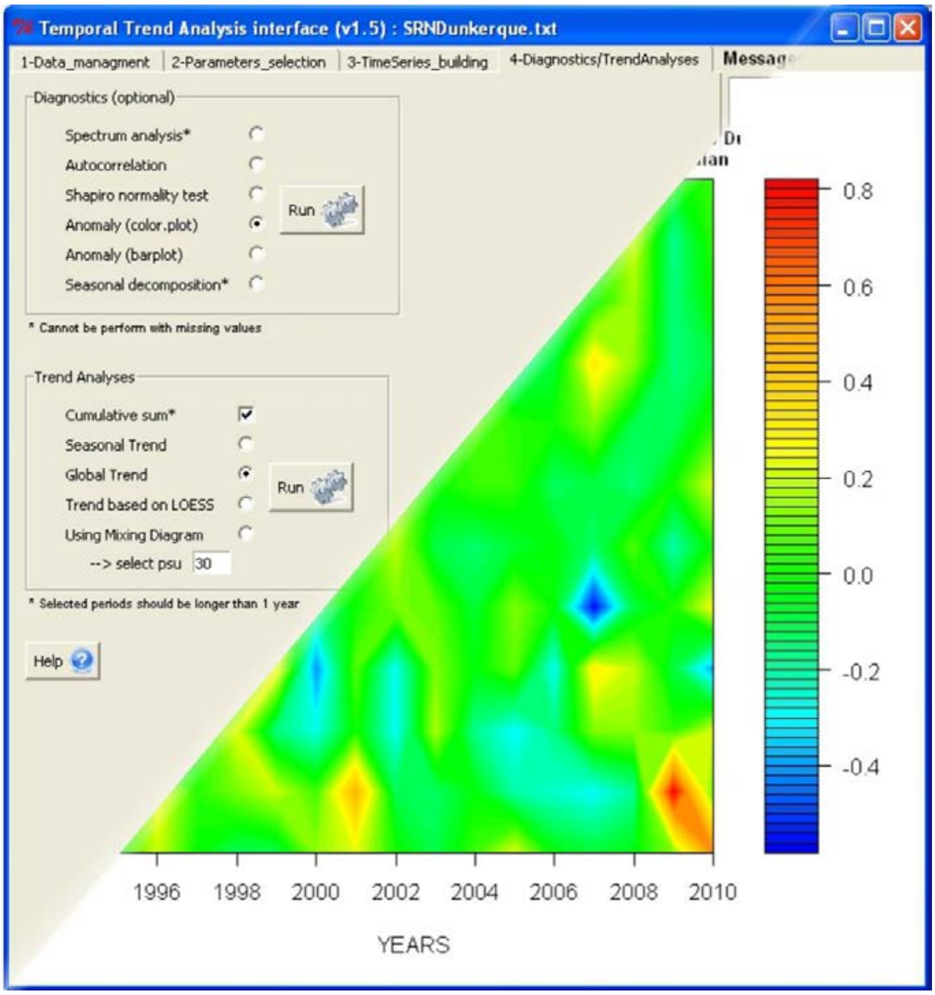
David Devreker
(david_devreker@yahoo.fr)
(David.Devreker@ifremer.fr)

Alain Lefebvre
(Alain.Lefebvre@ifremer.fr)

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

TTAinterfaceTrendAnalysis



Temporal Trend Analysis interface (v1.5) : SRNDunkerque.txt

Diagnostics (optional)

- Spectrum analysis* ○
- Autocorrelation ○
- Shapiro normality test ○
- Anomaly (color.plot) ●
- Anomaly (barplot) ○
- Seasonal decomposition* ○

* Cannot be perform with missing values

Trend Analyses

- Cumulative sum* ☑
- Seasonal Trend ○
- Global Trend ●
- Trend based on LOESS ○
- Using Mixing Diagram ○
- --> select psu 30

* Selected periods should be longer than 1 year

An Interface for *T*emporal *T*rend *A*nalysis and diagnostics

User guide

TTAinterface version 1.5.1 for R v3.0.0+

Contents

Introduction

The **TTAinterfaceTrendAnalysis** package is written with the R programming language in version 3.0.0+. It allows performing temporal trend analysis through a graphical user interface (GUI). The advantage of such kind of GUI 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. As an R coded interface this package is freely distributing (GPL Licence).

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 'SRNDunkerque.txt' will be used (data available in the package). This dataset regroup salinity, temperature and chlorophyll-a concentration sampled at 3 different stations (onshore to offshore) near the Gravelines Power Plant from 1995 to 2010 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 same dataset.

A. Installing R and the TTAinterfaceTrendAnalysis package

The **TTAinterfaceTrendAnalysis** package needs the basic R console to be installed and launched. It is written with R version 3.0.0+ and is compatible with the most recent version. R software (at least v3.0.0) comes with basic packages and a command console which can be downloaded from the CRAN website http://cran.r-project.org/. The **TTAinterfaceTrendAnalysis** package was created with the Tcl/Tk toolkit included in the tcltk package which is a part of the standard R installation for Windows, Linux and Unix platforms. For Mac OS X compatibility it is necessary to install an X Windows version of Tcl/Tk (http://cran.r-project.org/bin/macosx/tools/). More complete instruction concerning R installation can be found on the CRAN website.

Installation of a portable version of the package

If you have a portable version of the package, run R 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_1.5.zip' that comes in a zip archive. The TTAinterfaceTrendAnalysis 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).

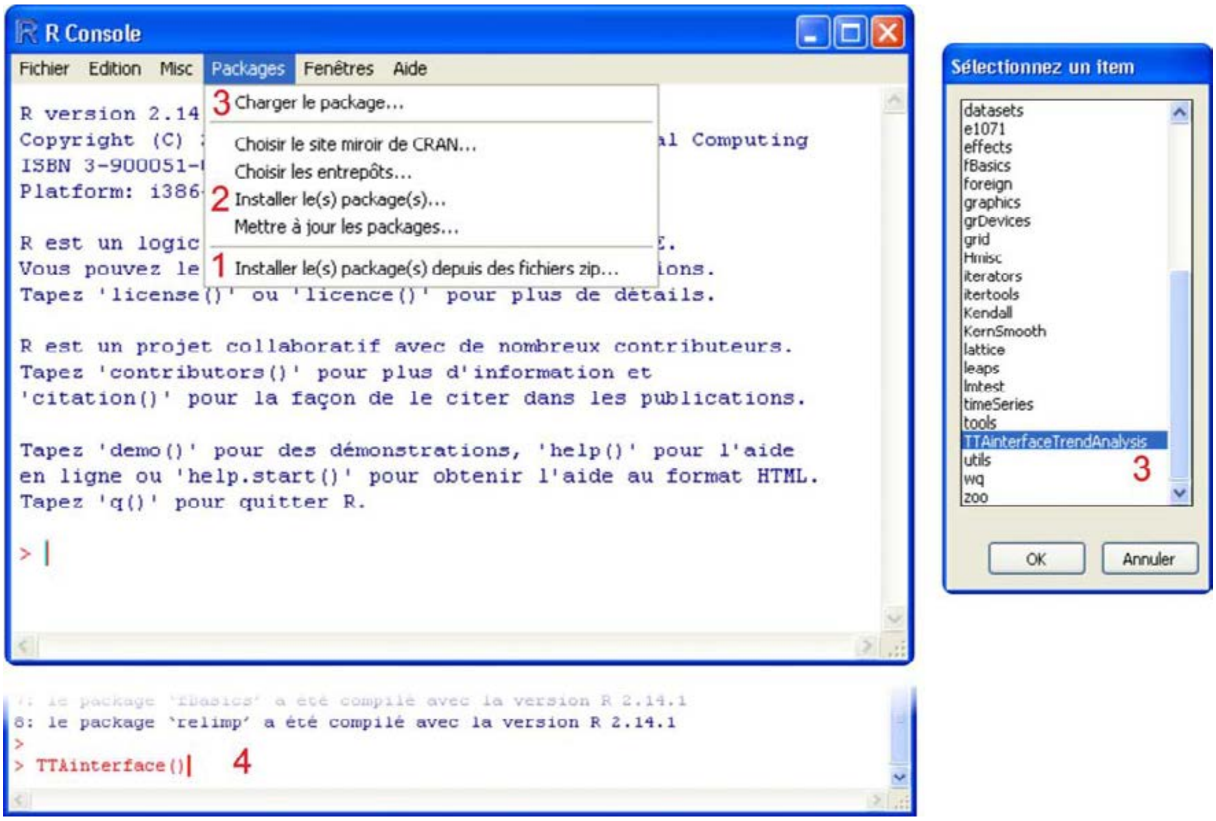
Installation from CRAN mirror

Alternatively, the package is available on the CRAN mirror (an internet connection is obviously needed). Open the R console and 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 is installed click on ‘Packages/Load package…’ in the ‘Packages’ menu of the console and select **TTAinterfaceTrendAnalysis** from the list (step **3** in Fig. 1). A small panel appears inviting you to start the interface (Fig. 2B), 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. If closed, the GUI can be directly re-launch using the start panel.

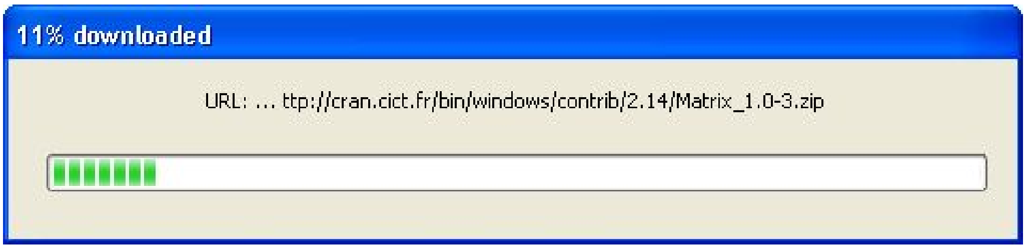
In some case, if you close both the GUI and the start panel, and try to reload them with step **3** without closing the R console, the GUI would not start. Then you have to re-start the R console or if you want to stay in your R session, enter the line TTAinterface() in the console (**4** in Fig. 1).



```
R version 2.14
Copyright (C)                    al Computing
ISBN 3-900051-
Platform: i386

R est un logic                   E.
Vous pouvez le                   ions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

```
7: le package 'fBasics' a été compilé avec la version R 2.14.1
8: le package 'relimp' a été compilé avec la version R 2.14.1
>
> TTAinterface()|
```

Figure 1. The R console and the different step to install and run the **TTAinterfaceTrendAnalysis** package.



URL: … ttp://cran.cict.fr/bin/windows/contrib/2.14/Matrix_1.0-3.zip

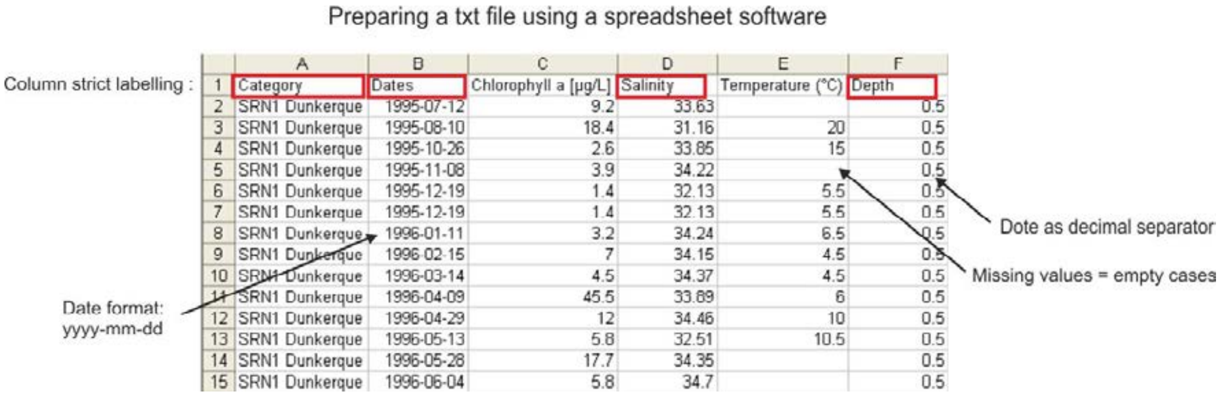
Figure 2A. Packages auto-downloading window.

Figure. 2B. Start panel of the **TTAinterfaceTrendAnalysis** package.

B. Create your dataset

The datasheet format that is read by the **TTAinterfaceTrendAnalysis** package is txt file (tab column separated). The main 'difficulty' to use the GUI is to prepare a txt file that is correctly read by the programme. Once in the interface, the guideline to perform analyses is very friendly.

The txt file has to be created with a spreadsheet software like Microsoft Excel or OpenOffice Calc. OpenOffice Calc is freely available (http://www.openoffice.org) and can manage files with 1 million lines whereas old versions of Microsoft Excel 32bits (before Excel 2007) are limited to 65500 lines but is more intuitive to use. The txt file is a numeric table with column label in the first row. For interface needs, some of these labels have to be defined and fixed. The column containing the categorical factors (sampling stations, taxa, chemical species…) must be named **Category** (with first letter in capital, remember that R is case sensitive), **Dates** for the date column in format yyyy-mm-dd (ISO 8601), Depth for the sampling depth column and **Salinity** for the salinities column (Fig. 3). Columns with parameters values (chlorophyll concentration, phytoplankton abundance, biomass…) can be freely labelled with the name of the parameters and its measurement unit as a preference. All values in the same line must correspond to a unique sampling (same dates and categorical factors). Contrary to **Salinity** and **Depth** columns, **Category** and **Dates** columns are necessary for the interface to work correctly (if you do not have categorical factors, create the appropriate column and file it with a character of your choice).



Preparing a txt file using a spreadsheet software

	A	B	C	D	E	F
1	Category	Dates	Chlorophyll a [µg/L]	Salinity	Temperature (°C)	Depth
2	SRN1 Dunkerque	1995-07-12	9.2	33.63		0.5
3	SRN1 Dunkerque	1995-08-10	18.4	31.16	20	0.5
4	SRN1 Dunkerque	1995-10-26	2.6	33.85	15	0.5
5	SRN1 Dunkerque	1995-11-08	3.9	34.22		0.5
6	SRN1 Dunkerque	1995-12-19	1.4	32.13	5.5	0.5
7	SRN1 Dunkerque	1995-12-19	1.4	32.13	5.5	0.5
8	SRN1 Dunkerque	1996-01-11	3.2	34.24	6.5	0.5
9	SRN1 Dunkerque	1996-02-15	7	34.15	4.5	0.5
10	SRN1 Dunkerque	1996-03-14	4.5	34.37	4.5	0.5
11	SRN1 Dunkerque	1996-04-09	45.5	33.89	6	0.5
12	SRN1 Dunkerque	1996-04-29	12	34.46	10	0.5
13	SRN1 Dunkerque	1996-05-13	5.8	32.51	10.5	0.5
14	SRN1 Dunkerque	1996-05-28	17.7	34.35		0.5
15	SRN1 Dunkerque	1996-06-04	5.8	34.7		0.5

Column strict labelling :

Date format: yyyy-mm-dd

Dote as decimal separator

Missing values = empty cases

Figure 3. Summarized processes to formatting a txt file readable in the **TTAinterfaceTrendAnalayis** package.

In your txt 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 txt file.

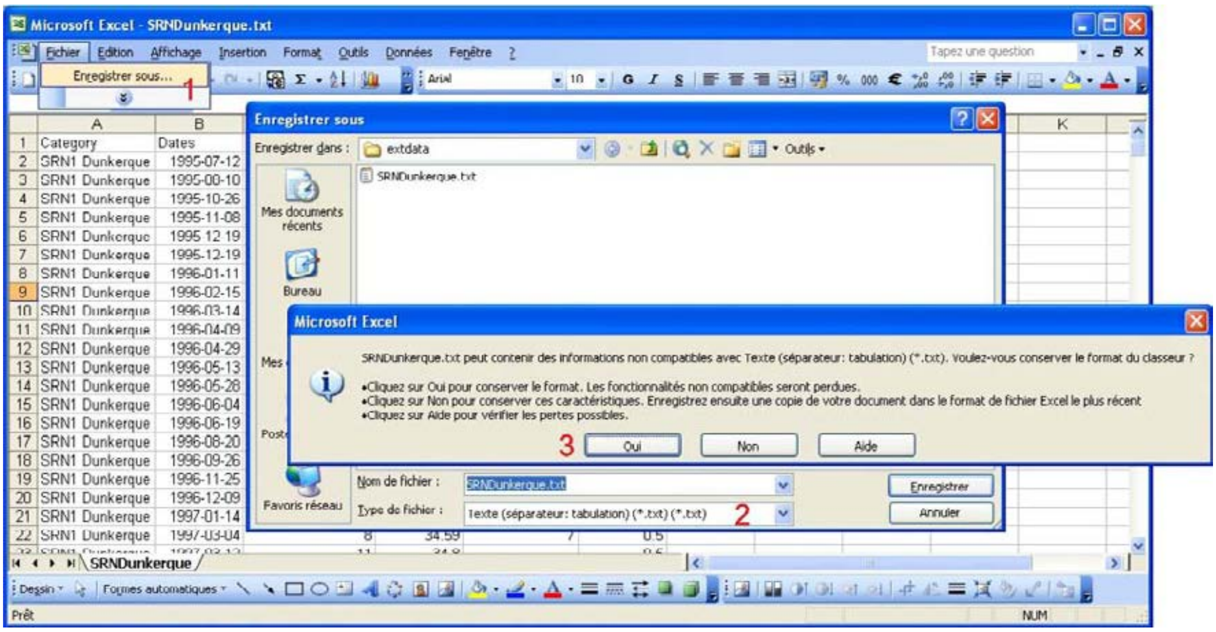
Missing values must be empty case, 0 are read as values and characters (NA, NaN) can induce bugs (they are labelled by the program itself so you don't have to do it).

Decimal separator must be '.' (dot). 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 TXT (Tab delimited) (*.txt) in the 'Save as type' option (**2** in Fig. 4A).

In LibreOffice Calc you have to select 'Text CSV (.csv)' (**2** in Fig. 4B), writing the .txt extension yourself, uncheck 'Automatic file extension' and check 'Edit filter settings' to select {tab} as column separator in step **3** (Fig. 4B).

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



	A	B
1	Category	Dates
2	SRN1 Dunkerque	1995-07-12
3	SRN1 Dunkerque	1995-08-10
4	SRN1 Dunkerque	1995-10-26
5	SRN1 Dunkerque	1995-11-08
6	SRN1 Dunkerque	1995 12 19
7	SRN1 Dunkerque	1995-12-19
8	SRN1 Dunkerque	1996-01-11
9	SRN1 Dunkerque	1996-02-15
10	SRN1 Dunkerque	1996-03-14
11	SRN1 Dunkerque	1996-04-09
12	SRN1 Dunkerque	1996-04-29
13	SRN1 Dunkerque	1996-05-13
14	SRN1 Dunkerque	1996-05-28
15	SRN1 Dunkerque	1996-06-04
16	SRN1 Dunkerque	1996-06-19
17	SRN1 Dunkerque	1996-08-20
18	SRN1 Dunkerque	1996-09-26
19	SRN1 Dunkerque	1996-11-25
20	SRN1 Dunkerque	1996-12-09
21	SRN1 Dunkerque	1997-01-14
22	SRN1 Dunkerque	1997-03-04

Microsoft Excel

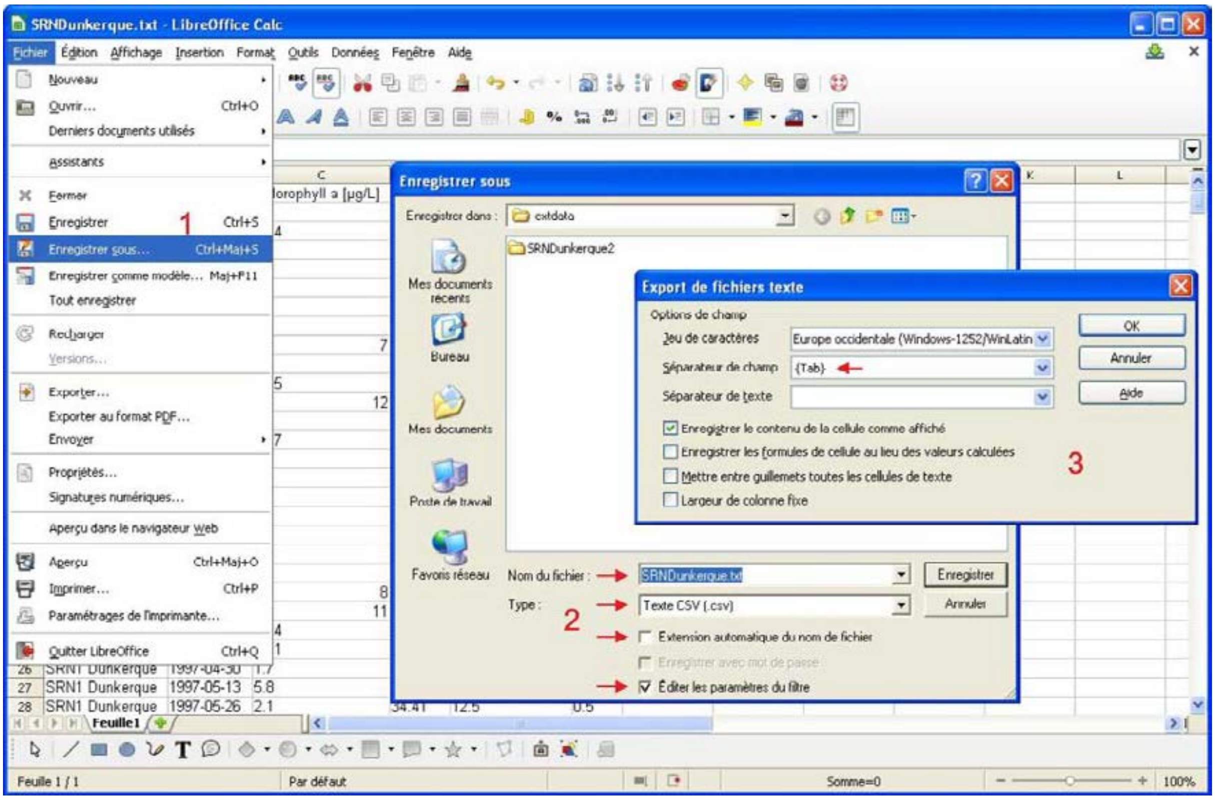
SRNDunkerque.txt peut contenir des informations non compatibles avec Texte (séparateur: tabulation) (*.txt). Voulez-vous conserver le format du classeur ?

- Cliquez sur Oui pour conserver le format. Les fonctionnalités non compatibles seront perdues.
- Cliquez sur Non pour conserver ces caractéristiques. Enregistrez ensuite une copie de votre document dans le format de fichier Excel le plus récent
- Cliquez sur Aide pour vérifier les pertes possibles.

Nom de fichier : SRNDunkerque.txt

Type de fichier : Texte (séparateur: tabulation) (*.txt) (*.txt)

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



Enregistrer dans : extdata

SRNDunkerque2

Nom du fichier : SRNDunkerque.txt

Type : Texte CSV (.csv)

☐ Extension automatique du nom de fichier

☐ Enregistrer avec mot de passe

☑ Éditer les paramètres du filtre

Options de champ

Jeu de caractères: Europe occidentale (Windows-1252/WinLatin

Séparateur de champ: {Tab}

Séparateur de texte:

☑ Enregistrer le contenu de la cellule comme affiché

☐ Enregistrer les formules de cellule au lieu des valeurs calculées

☐ Mettre entre guillemets toutes les cellules de texte

☐ Largeur de colonne fixe

Figure 4B. Creating a txt file with LibreOffice Calc.

<u>To summarize:</u>

- Must be a .txt file
- Column labelling: **Category** for categorical factors (sampling station, taxa, chemicals…), **Dates** for dates, **Salinity** for salinities and **Depth** for depths values.
- Dates format must be yyyy-mm-dd (ISO 8601)
- Dot as decimal separator
- Missing values must be empty cases

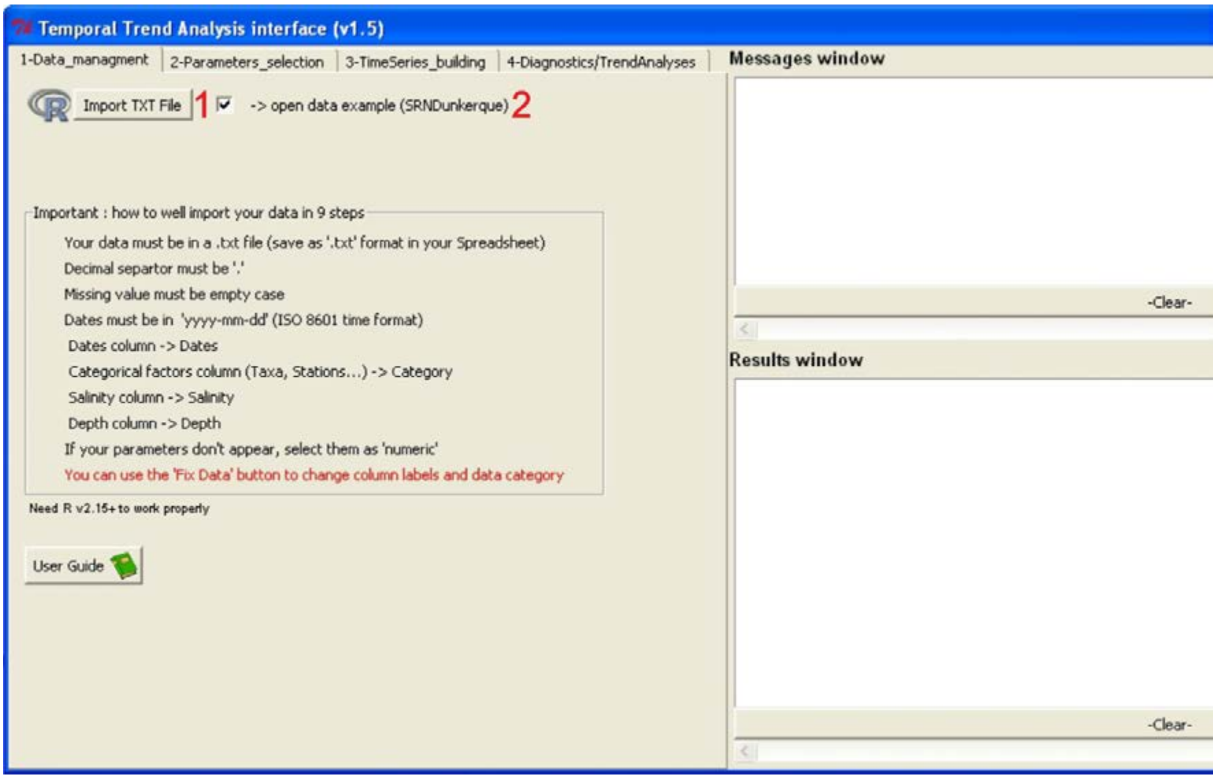
C. Quick steps

1. Managing your database

The first panel of the interface ‘1-Data_managment’ allows interactions with the dataset

Import your database

This is the first page you can see when you start the interface (Fig. 5). You can only import a txt file (or the data example file, **2** in Fig. 5) using the <Import TXT File> button (**1** in Fig. 5) and read the advice to correctly import a txt file. All other options are disabled.



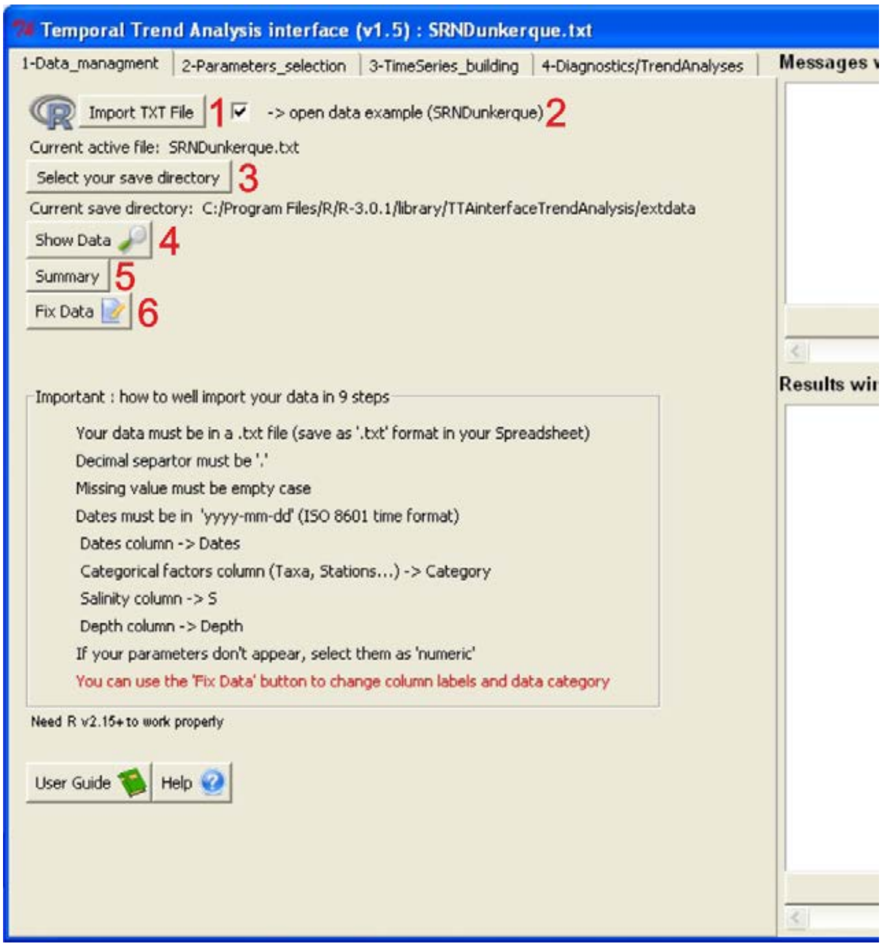
Important : how to well import your data in 9 steps

Your data must be in a .txt file (save as '.txt' format in your Spreadsheet)
Decimal separtor must be '.'
Missing value must be empty case
Dates must be in 'yyyy-mm-dd' (ISO 8601 time format)
Dates column -> Dates
Categorical factors column (Taxa, Stations...) -> Category
Salinity column -> Salinity
Depth column -> Depth
If your parameters don't appear, select them as 'numeric'
You can use the 'Fix Data' button to change column labels and data category

Need R v2.15+ to work properly

Figure 5. Panel 1 when the interface is launch.

Once your txt file is imported more options appear:



Temporal Trend Analysis interface (v1.5) : SRNDunkerque.txt
1-Data_managment
2-Parameters_selection
3-TimeSeries_building
4-Diagnostics/TrendAnalyses
Messages w
Import TXT File
1
☑ -> open data example (SRNDunkerque)
2
Select your save directory
3
Show Data
4
Summary
5
Fix Data
6
Results win
User Guide
Help

Current active file: SRNDunkerque.txt

Current save directory: C:/Program Files/R/R-3.0.1/library/TTAinterfaceTrendAnalysis/extdata

Important : how to well import your data in 9 steps

Your data must be in a .txt file (save as '.txt' format in your Spreadsheet)
Decimal separtor must be '.'
Missing value must be empty case
Dates must be in 'yyyy-mm-dd' (ISO 8601 time format)
Dates column -> Dates
Categorical factors column (Taxa, Stations...) -> Category
Salinity column -> S
Depth column -> Depth
If your parameters don't appear, select them as 'numeric'
You can use the 'Fix Data' button to change column labels and data category

Need R v2.15+ to work properly

Figure 6. Left part of the Panel 1 once a txt file is imported (here the data example).

Change the save directory

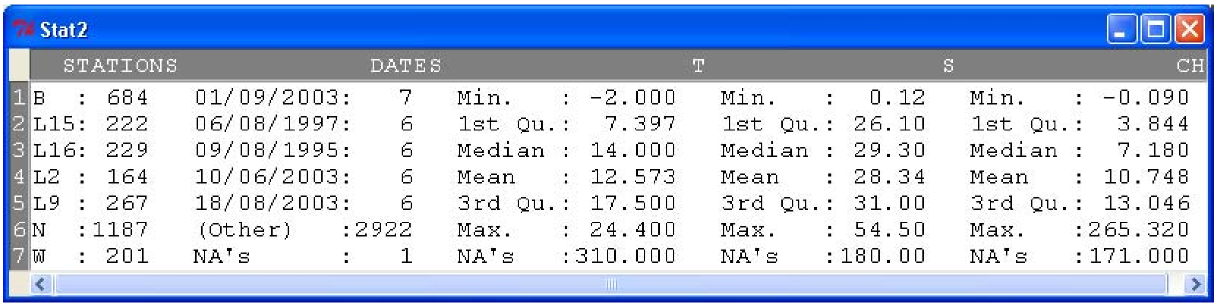
The button <Select your save directory> (**3** in Fig. 6) let you choose a folder on your computer to save the different figures and results of analysis you will obtain with the interface. By default this is the folder where your imported txt file is stock. If you import a new txt file the save directory is reset to default.

Display your data

The button <Show Data> (**4** in Fig. 6) displays a table of your imported data.

Summarize your raw dataset

The button <Summary> (**5** 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.

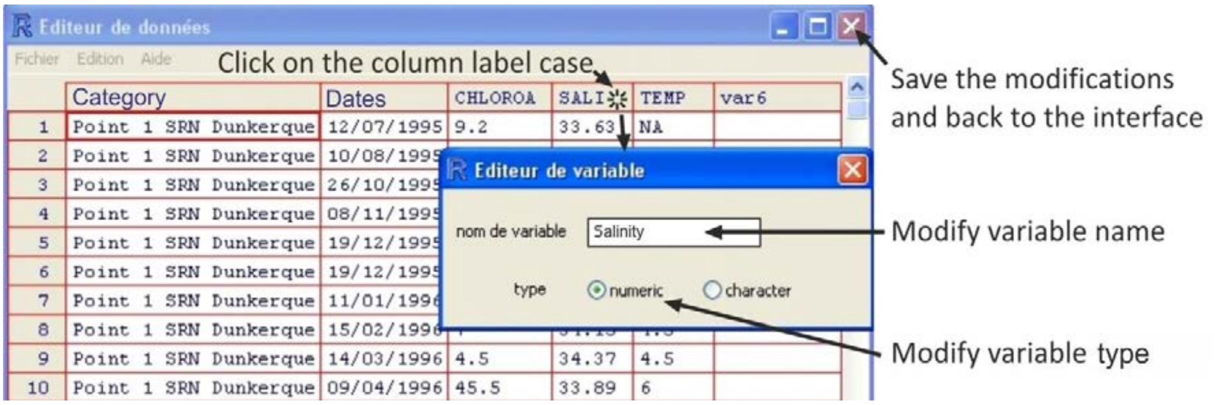


```
  STATIONS          DATES              T                  S                  CH
1 B  : 684   01/09/2003:   7   Min.   : -2.000   Min.   :  0.12   Min.   : -0.090
2 L15: 222   06/08/1997:   6   1st Qu.:  7.397   1st Qu.: 26.10   1st Qu.:  3.844
3 L16: 229   09/08/1995:   6   Median : 14.000   Median : 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. Resume of the ‘SRNDunkerque’ raw database.

Edit your data

The button <Fix Data> (**6** in Fig. 6) let you edit your dataset directly from the interface (Fig. 8). However you can only perform simple tasks such as modifying column label, modifying data type or editing cases one by one. For more complex modification, use your spreadsheet software.



Click on the column label case

	Category	Dates	CHLOROA	SALI	TEMP	var6
1	Point 1 SRN Dunkerque	12/07/1995	9.2	33.63	NA	
2	Point 1 SRN Dunkerque	10/08/1995				
3	Point 1 SRN Dunkerque	26/10/1995				
4	Point 1 SRN Dunkerque	08/11/1995				
5	Point 1 SRN Dunkerque	19/12/1995				
6	Point 1 SRN Dunkerque	19/12/1995				
7	Point 1 SRN Dunkerque	11/01/1996				
8	Point 1 SRN Dunkerque	15/02/1996				
9	Point 1 SRN Dunkerque	14/03/1996	4.5	34.37	4.5	
10	Point 1 SRN Dunkerque	09/04/1996	45.5	33.89	6	

nom de variable: Salinity

type ● numeric ○ character

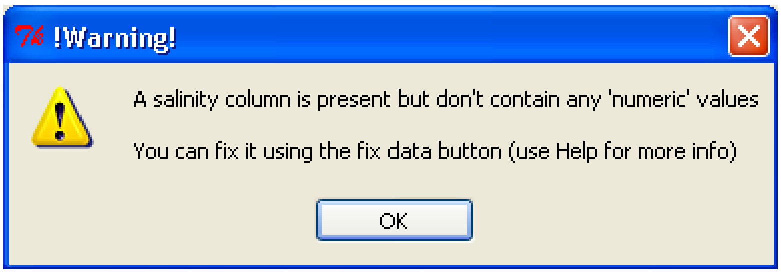
Save the modifications and back to the interface

Modify variable name

Modify variable type

Figure 8. Data editor spreadsheet when using the <Fix Data> button in panel 1.

If you import a txt file containing a column labelled **Salinity** for salinities but with values that are not identified as ‘numeric’ a warning message appear (same thing for Depth column):



A salinity column is present but don't contain any 'numeric' values

You can fix it using the fix data button (use Help for more info)

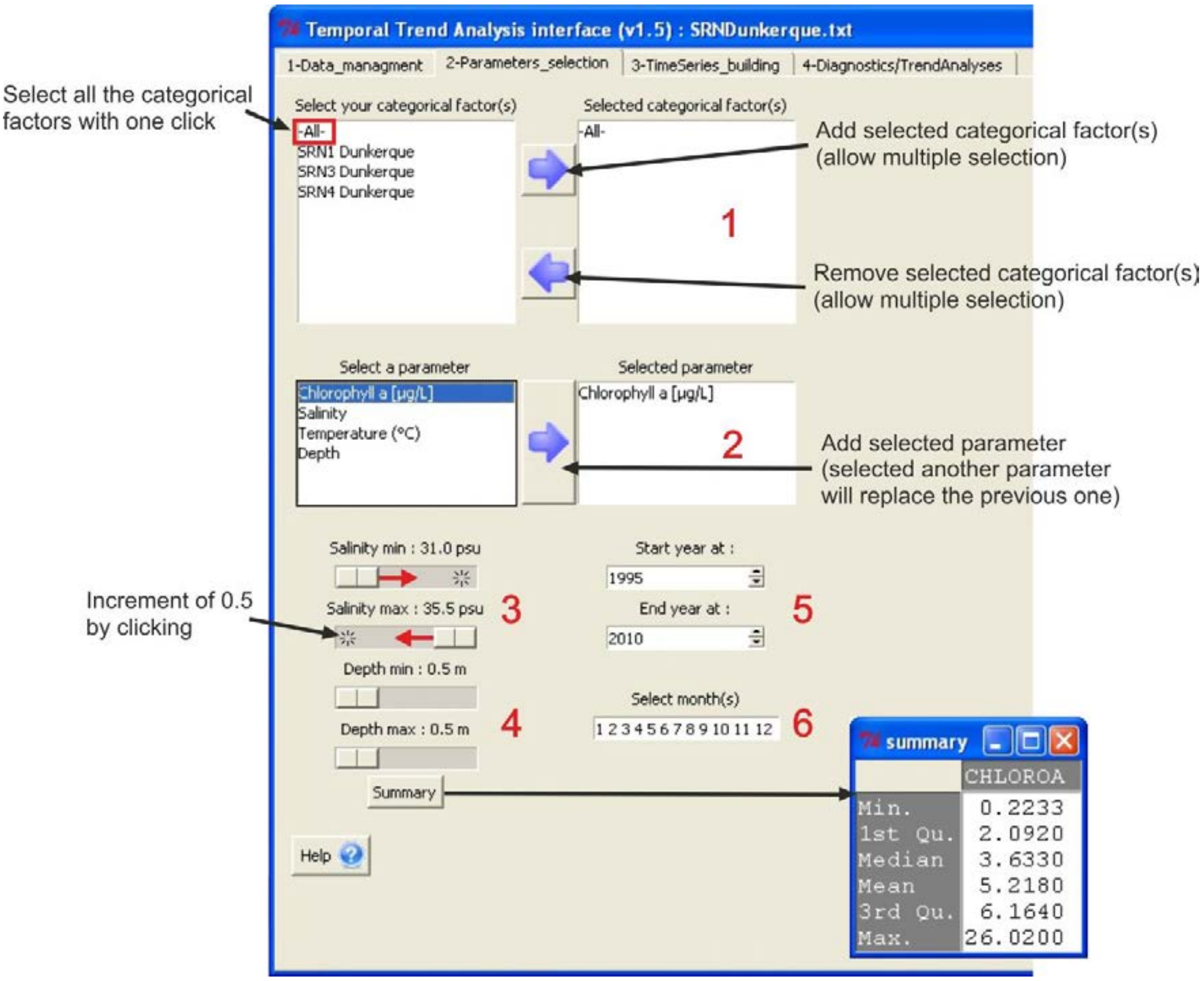
2. Select your parameters

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

1. The different categorical factors (here sampling stations)
2. The parameter to analyse
3. Interval of salinity to analyse (sliders)
4. Interval of depth to analyse (if exist)
5. Interval of years to analyse
6. Months to analyse (must be spaced)

<u>Summarize your selected parameters</u>

The button <Summary> displays descriptive statistics of the raw selected data. No mathematical treatment has been done on this data, only simple selection and sorting.



Salinity min : 31.0 psu

Salinity max : 35.5 psu

Depth min : 0.5 m

Depth max : 0.5 m

Start year at : 1995

End year at : 2010

Select month(s) 1 2 3 4 5 6 7 8 9 10 11 12

```
        CHLOROA
Min.     0.2233
1st Qu.  2.0920
Median   3.6330
Mean     5.2180
3rd Qu.  6.1640
Max.    26.0200
```

Select all the categorical factors with one click

Add selected categorical factor(s) (allow multiple selection)

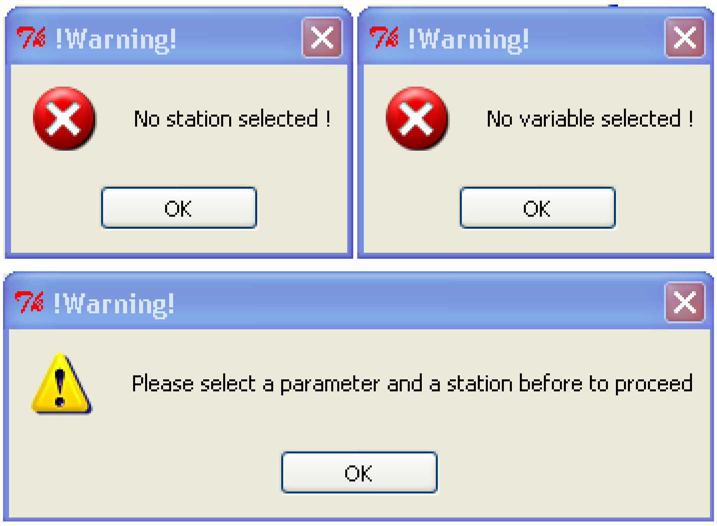
Remove selected categorical factor(s) (allow multiple selection)

Add selected parameter (selected another parameter will replace the previous one)

Increment of 0.5 by clicking

Figure 9. Options in the panel 2.

The following warnings appear if you didn't select any parameter or categorical factors before performing a mathematical treatment:

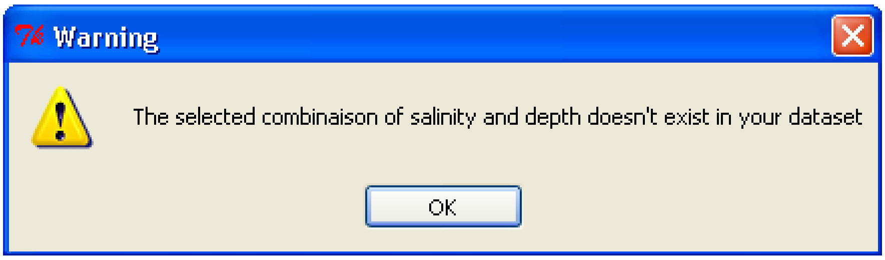


No station selected !

No variable selected !

Please select a parameter and a station before to proceed

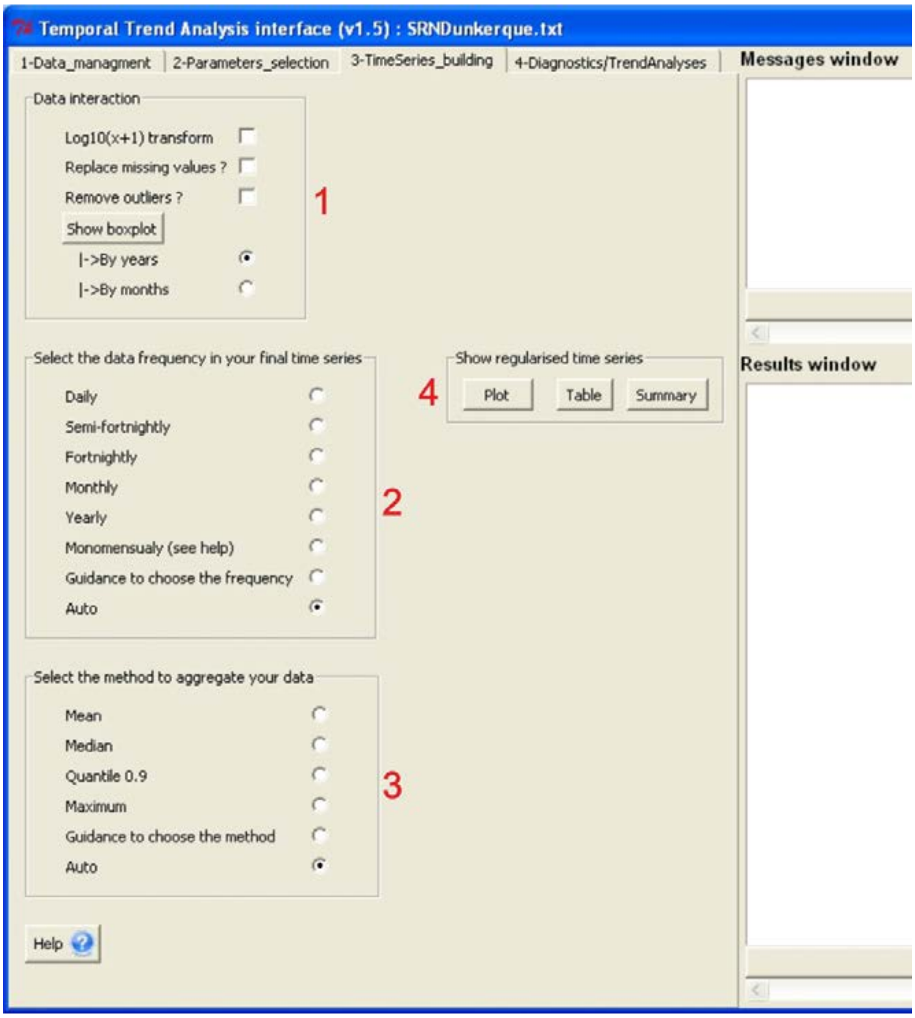
Also if the combination of salinity and depth selected in **4** and **5** (Fig. 9) is irrelevant:



The selected combinaison of salinity and depth doesn't exist in your dataset

3. Build your time series

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



Temporal Trend Analysis interface (v1.5) : SRNDunkerque.txt

1-Data_managment | 2-Parameters_selection | 3-TimeSeries_building | 4-Diagnostics/TrendAnalyses

Data interaction

- Log10(x+1) transform □
- Replace missing values ? □
- Remove outliers ? □
- Show boxplot
- |->By years ●
- |->By months ○

Select the data frequency in your final time series

- Daily ○
- Semi-fortnightly ○
- Fortnightly ○
- Monthly ○
- Yearly ○
- Monomensualy (see help) ○
- Guidance to choose the frequency ○
- Auto ●

Select the method to aggregate your data

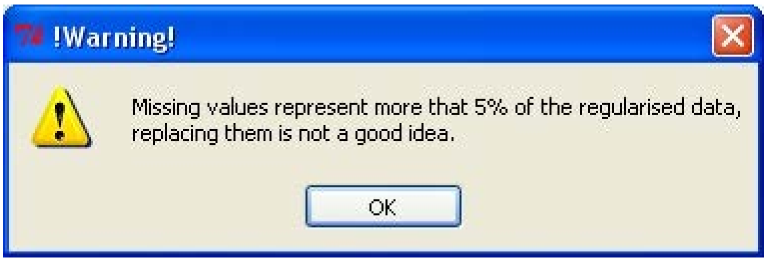
- Mean ○
- Median ○
- Quantile 0.9 ○
- Maximum ○
- Guidance to choose the method ○
- Auto ●

Show regularised time series: Plot | Table | Summary

Figure 10. Options in the panel 3.

Data interaction

The 'Data interaction' frame (**1** in Fig. 10) allows manipulating the raw data base. You can transform the selected parameter into log10(x+1) to build the regularized time series. You can also remove outliers from the raw data distribution and to replace missing values from the time series (require for some diagnostic processes). A warning message appears if you choose to replace missing values and if they represent more than 5% of your data:



!Warning!

Missing values represent more that 5% of the regularised data, replacing them is not a good idea.

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

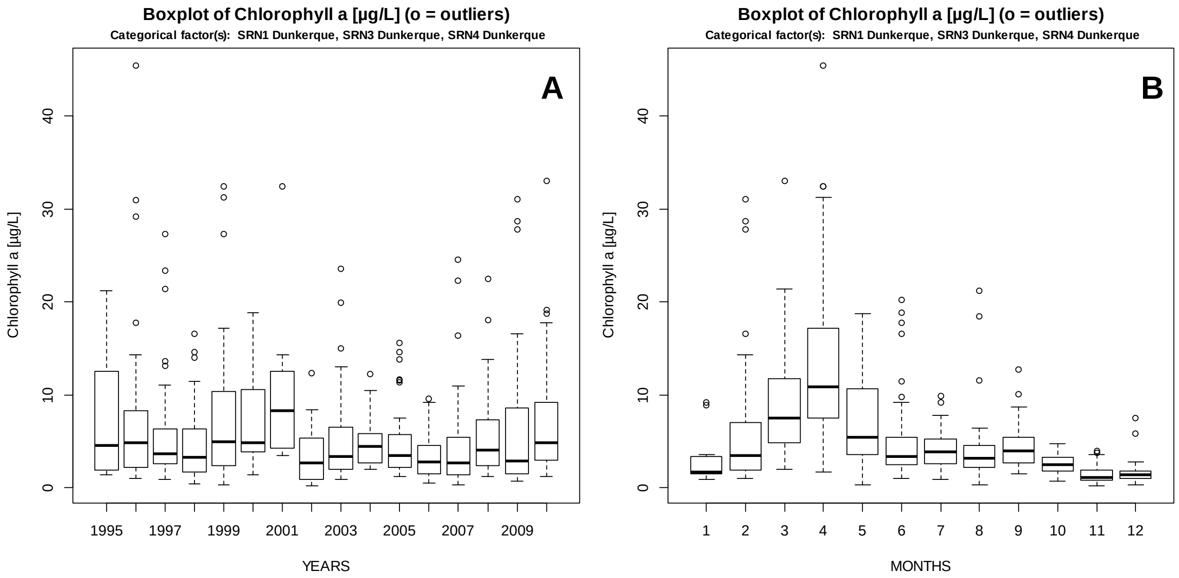
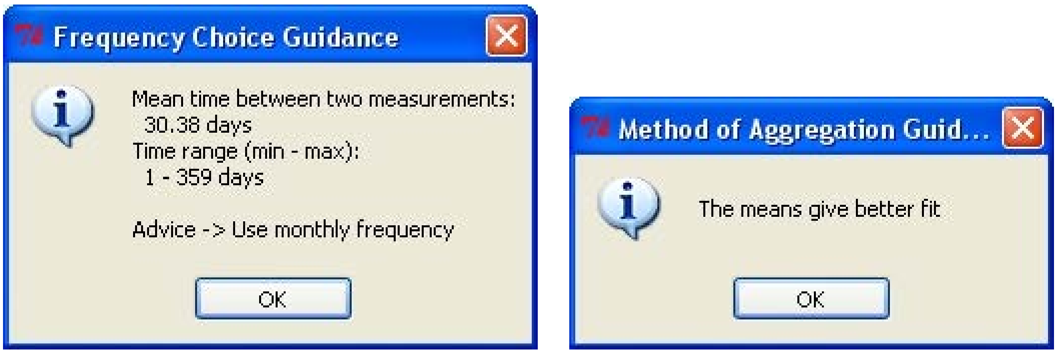


Figure 11. Boxplot by years (A) and by months (B) of the 'SRNDunkerque' raw database with outliers.

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).


Frequency Choice Guidance

Mean time between two measurements:
30.38 days
Time range (min - max):
1 - 359 days

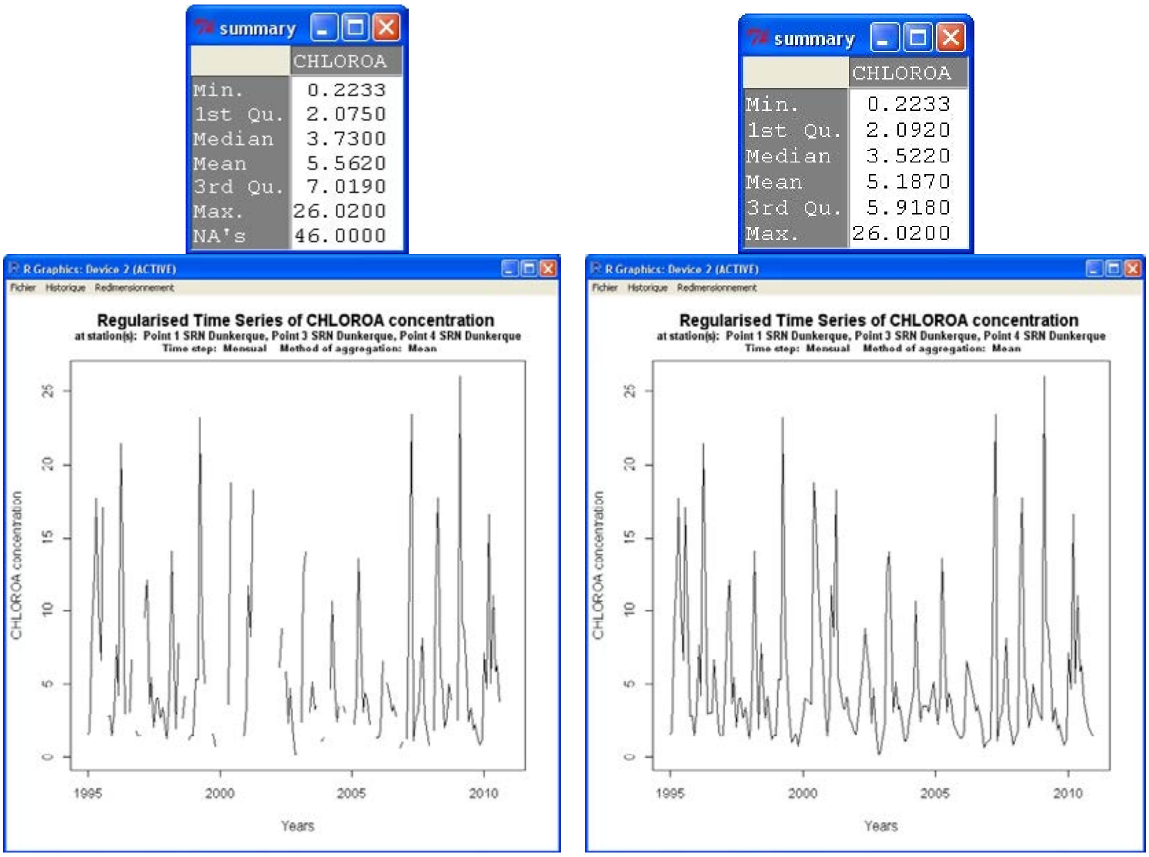
Advice -> Use monthly frequency

Method of Aggregation Guid...

The means give better fit

Figure 12. Messages to choose balanced frequency and method of aggregation to build time series.

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



summary

	CHLOROA
Min.	0.2233
1st Qu.	2.0750
Median	3.7300
Mean	5.5620
3rd Qu.	7.0190
Max.	26.0200
NA's	46.0000

summary

	CHLOROA
Min.	0.2233
1st Qu.	2.0920
Median	3.5220
Mean	5.1870
3rd Qu.	5.9180
Max.	26.0200

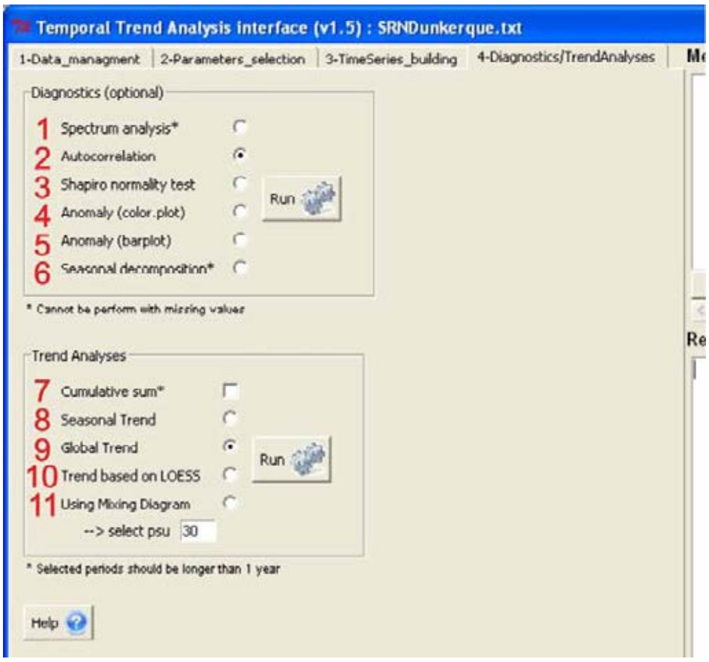
Regularised Time Series of CHLOROA concentration
at station(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4 SRN Dunkerque
Time step: Mensual Method of aggregation: Mean

R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
CHLOROA concentration
0
5
10
15
20
25
1995
2000
2005
2010
Years

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 analysis on your regularized time series (Fig. 14).



Temporal Trend Analysis interface (v1.5) : SRNDunkerque.txt

Diagnostics (optional)

1 Spectrum analysis* ○
2 Autocorrelation ●
3 Shapiro normality test ○
4 Anomaly (color.plot) ○
5 Anomaly (barplot) ○
6 Seasonal decomposition* ○

Run

* Cannot be perform with missing values

Trend Analyses

7 Cumulative sum* □
8 Seasonal Trend ○
9 Global Trend ●
10 Trend based on LOESS ○
11 Using Mixing Diagram ○
--> select psu 30

Run

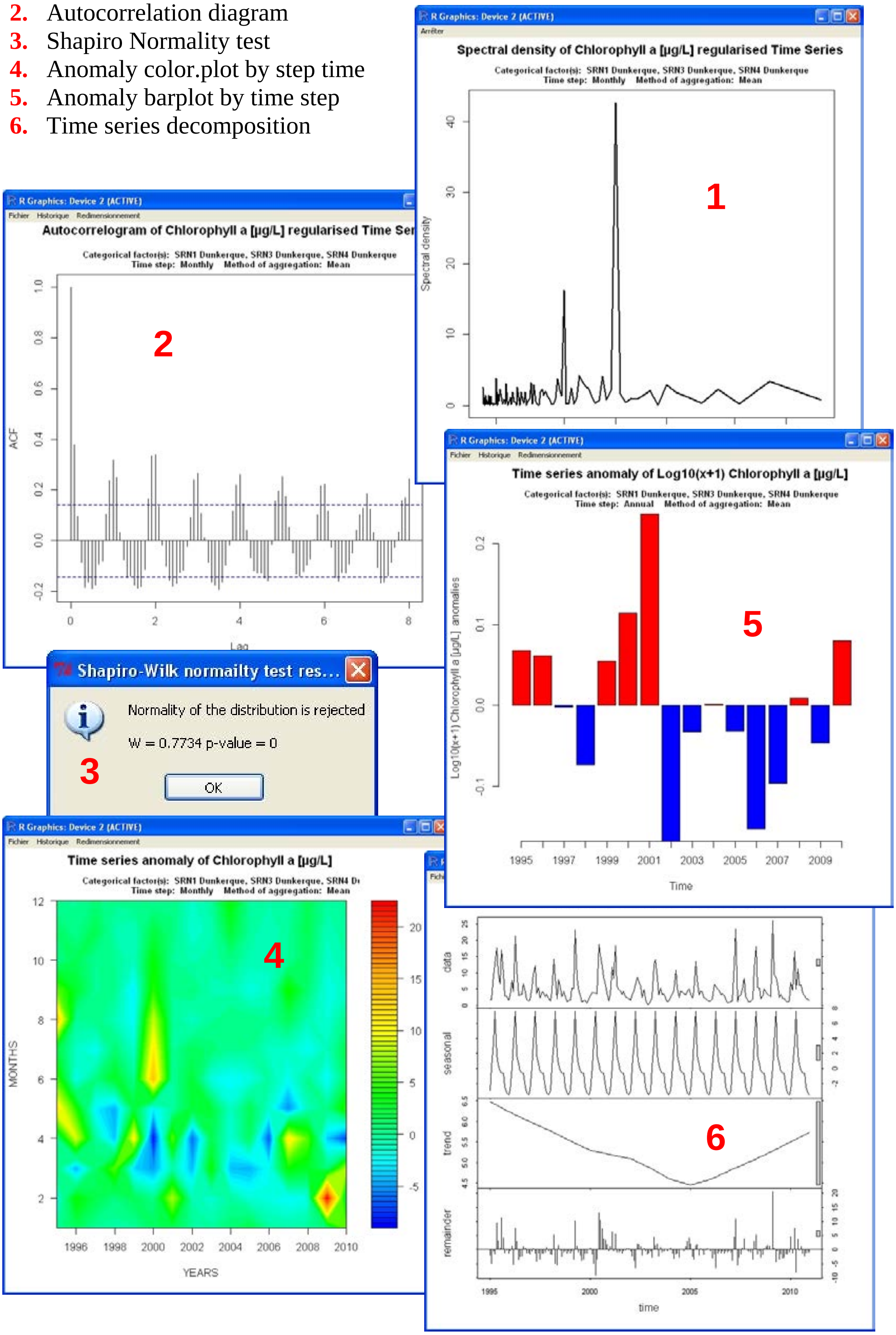
* Selected periods should be longer than 1 year

Help

Figure 14. Options in the panel 4.

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

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

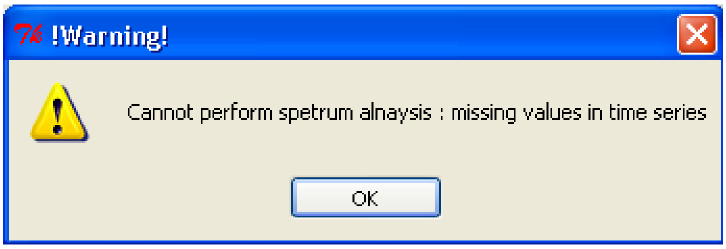

R Graphics: Device 2 (ACTIVE)
Arrêter
Spectral density of Chlorophyll a [µg/L] regularised Time Series
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Time step: Monthly Method of aggregation: Mean
Spectral density
0
10
20
30
40
1
R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Autocorrelogram of Chlorophyll a [µg/L] regularised Time Ser
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Time step: Monthly Method of aggregation: Mean
ACF
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
Lag
2
Shapiro-Wilk normailty test res...

Normality of the distribution is rejected

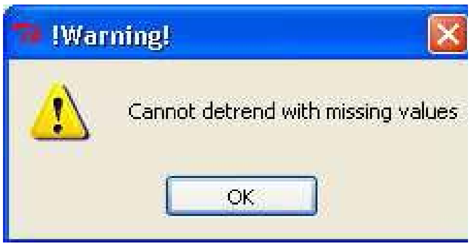
W = 0.7734 p-value = 0

OK
3
R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Time series anomaly of Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Di
Time step: Monthly Method of aggregation: Mean
MONTHS
2
4
6
8
10
12
1996
1998
2000
2002
2004
2006
2008
2010
YEARS
-5
0
5
10
15
20
4
R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Time series anomaly of Log10(x+1) Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Time step: Annual Method of aggregation: Mean
Log10(x+1) Chlorophyll a [µg/L] anomalies
-0.1
0.0
0.1
0.2
1995
1997
1999
2001
2003
2005
2007
2009
Time
5
data
seasonal
trend
remainder
6
1995
2000
2005
2010
time

Performing diagnostics **1** and **5** while keeping missing values displays warning messages:



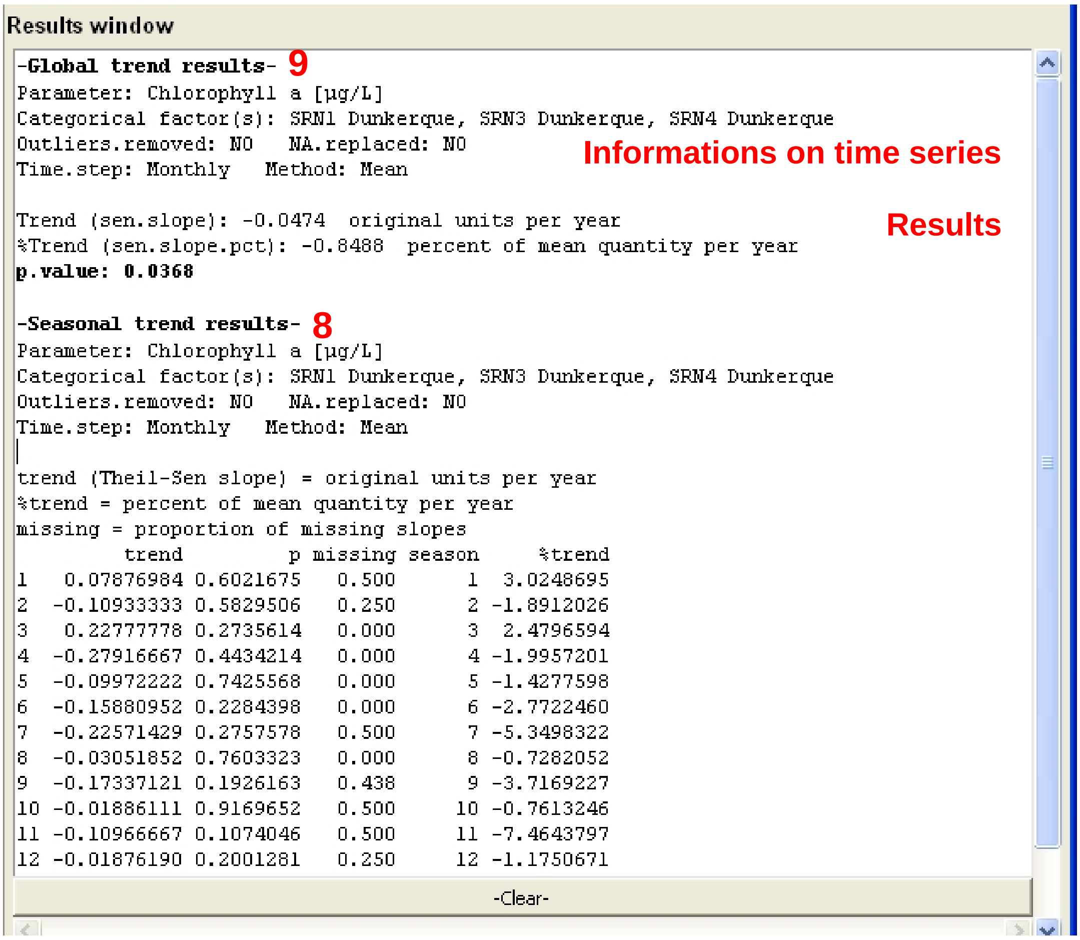
Cannot perform spetrum alnaysis : missing values in time series



Cannot detrend with missing values

Temporal trend tests (Trend Analyses frame)

- Global trend (**9** in Fig. 14 and 15) and Seasonal trend (**8** in Fig. 14 and 15) (both based on Seasonal Kendall test) results are display on the right par of the interface (Results/Messages):



```
-Global trend results-
Parameter: Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Outliers.removed: NO   NA.replaced: NO
Time.step: Monthly   Method: Mean

Trend (sen.slope): -0.0474  original units per year
%Trend (sen.slope.pct): -0.8488  percent of mean quantity per year
p.value: 0.0368

-Seasonal trend results-
Parameter: Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Outliers.removed: NO   NA.replaced: NO
Time.step: Monthly   Method: Mean

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   0.07876984 0.6021675   0.500      1  3.0248695
2  -0.10933333 0.5829506   0.250      2 -1.8912026
3   0.22777778 0.2735614   0.000      3  2.4796594
4  -0.27916667 0.4434214   0.000      4 -1.9957201
5  -0.09972222 0.7425568   0.000      5 -1.4277598
6  -0.15880952 0.2284398   0.000      6 -2.7722460
7  -0.22571429 0.2757578   0.500      7 -5.3498322
8  -0.03051852 0.7603323   0.000      8 -0.7282052
9  -0.17337121 0.1926163   0.438      9 -3.7169227
10 -0.01886111 0.9169652   0.500     10 -0.7613246
11 -0.10966667 0.1074046   0.500     11 -7.4643797
12 -0.01876190 0.2001281   0.250     12 -1.1750671
```

Figure 15. Results of temporal trend test (Global Trend in **9** and Seasonal Trend in **8**) display in the right part of the interface.

- Visually identify and select different periods of trend with Cumulative sum (7 in Fig. 14, Fig. 16)…

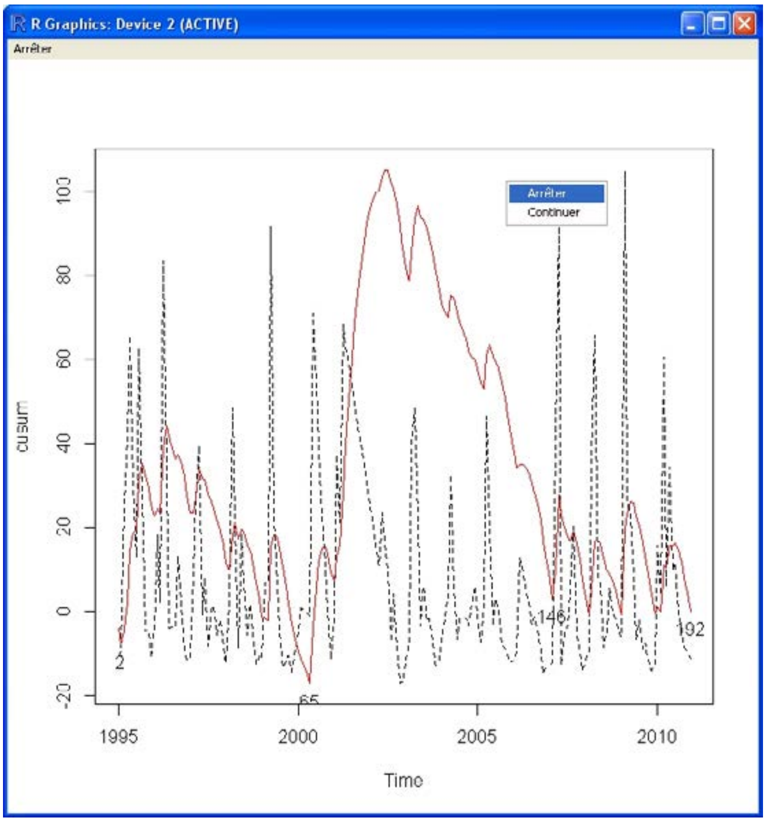
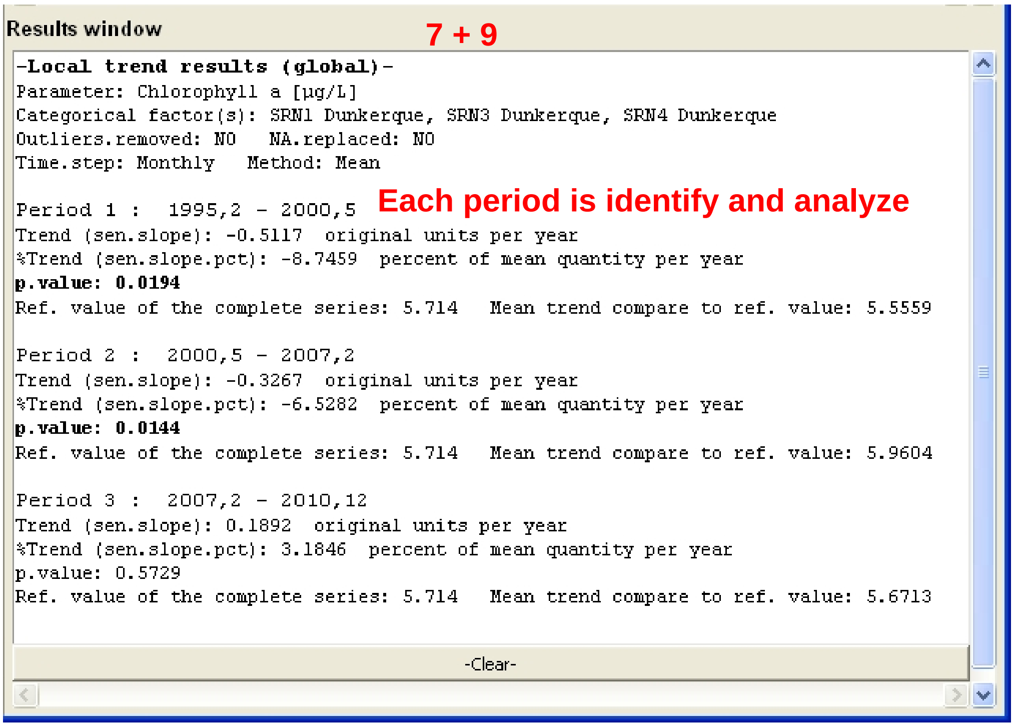


Figure 16. Cusum plot (cumulative sum in red), with periods manually identified (numbers).

…and perform Kendall test (Global or Seasonal) on each of these periods (Fig. 17):

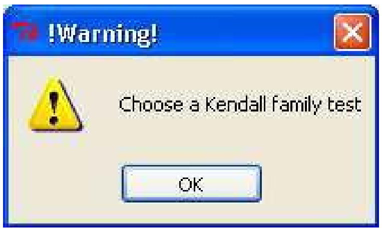


```
-Local trend results (global)-
Parameter: Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Outliers.removed: NO   NA.replaced: NO
Time.step: Monthly   Method: Mean

Period 1 :  1995,2 - 2000,5
Trend (sen.slope): -0.5117  original units per year
%Trend (sen.slope.pct): -8.7459  percent of mean quantity per year
p.value: 0.0194
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.5559

Period 2 :  2000,5 - 2007,2
Trend (sen.slope): -0.3267  original units per year
%Trend (sen.slope.pct): -6.5282  percent of mean quantity per year
p.value: 0.0144
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.9604

Period 3 :  2007,2 - 2010,12
Trend (sen.slope): 0.1892  original units per year
%Trend (sen.slope.pct): 3.1846  percent of mean quantity per year
p.value: 0.5729
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.6713
```

Each period is identify and analyze

Figure 17. Results of Global Kendall test perform on the different periods identified with cusum plot .

The following warning message is display if you select another test than Kendall after choosing the Cumulative sum option:

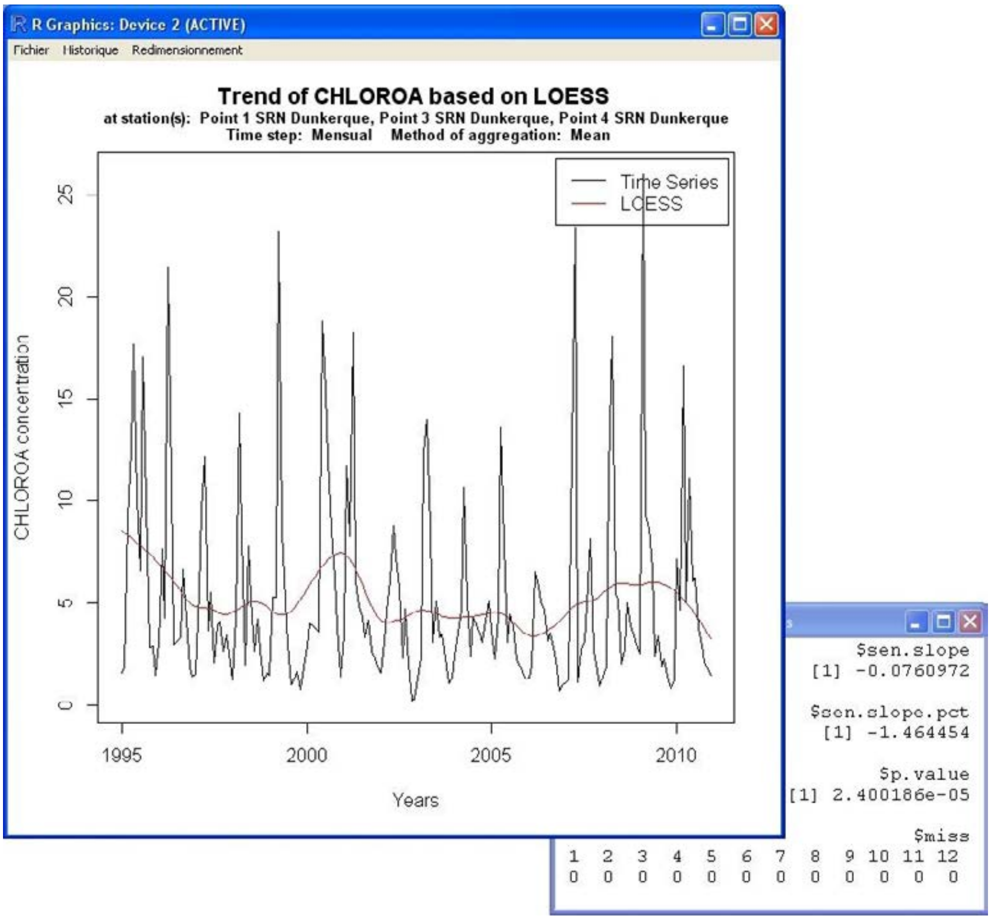


!Warning!

Choose a Kendall family test

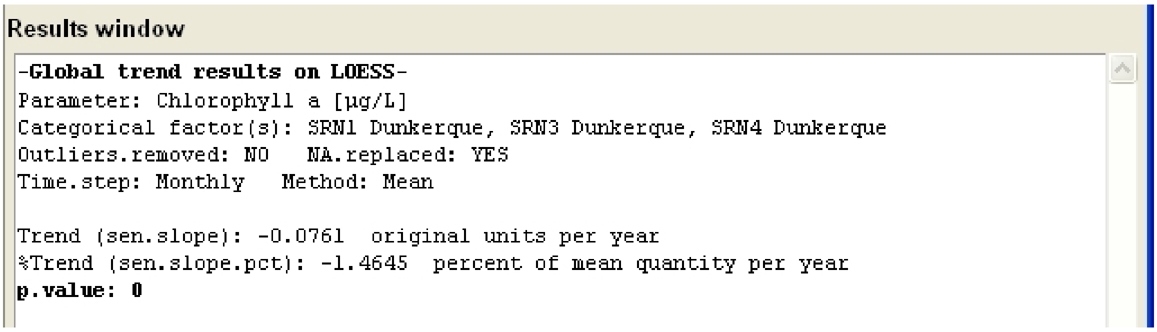
OK

- Trend based on LOESS smoothing (**10** in Fig. 14):



```
$sen.slope
[1] -0.0760972

$sen.slope.pct
[1] -1.464454

$p.value
[1] 2.400186e-05

$miss
 1  2  3  4  5  6  7  8  9 10 11 12
 0  0  0  0  0  0  0  0  0  0  0  0
```

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

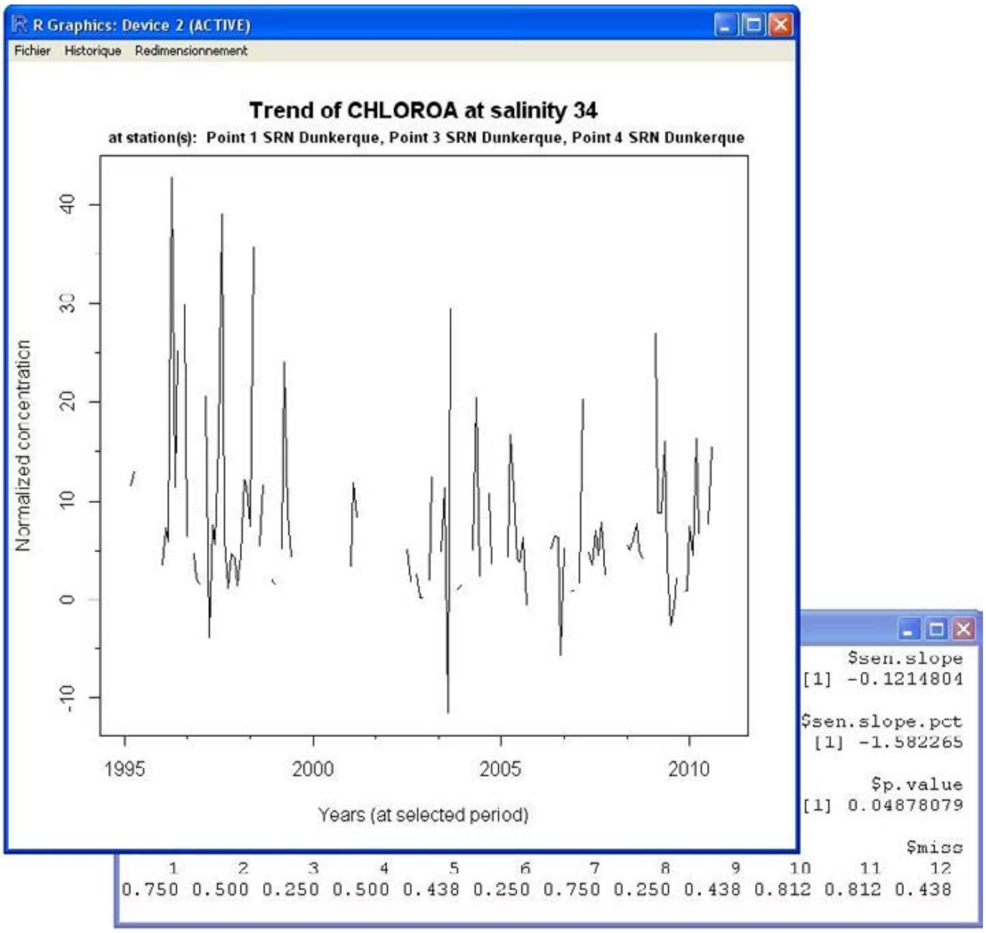


Results window

```
-Global trend results on LOESS-
Parameter: Chlorophyll a [µg/L]
Categorical factor(s): SRN1 Dunkerque, SRN3 Dunkerque, SRN4 Dunkerque
Outliers.removed: NO   NA.replaced: YES
Time.step: Monthly   Method: Mean

Trend (sen.slope): -0.0761  original units per year
%Trend (sen.slope.pct): -1.4645  percent of mean quantity per year
p.value: 0
```

Figure 19. Results of temporal trend test (Global Trend) apply on loess smoothing, display in right part of the interface.

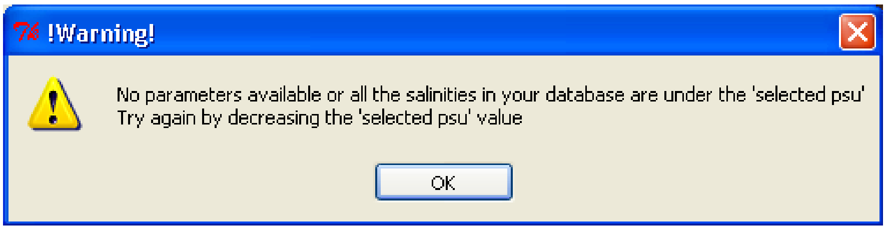
- Trend based on normalized concentration of nutrients at fixed salinity for each month (**11** in Fig. 14):


R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Trend of CHLOROA at salinity 34
at station(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4 SRN Dunkerque
Normalized concentration
-10
0
10
20
30
40
1995
2000
2005
2010
Years (at selected period)

```
$sen.slope
[1] -0.1214804

$sen.slope.pct
[1] -1.582265

$p.value
[1] 0.04878079

$miss
    1     2     3     4     5     6     7     8     9    10    11    12
0.750 0.500 0.250 0.500 0.438 0.250 0.750 0.250 0.438 0.812 0.812 0.438
```

Figure 20. Plot of normalized chlorophyll-a concentration at salinity 34 and results (text box in the background) of Global Kendall perform on this time series.

The following warning message is display 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:



No parameters available or all the salinities in your database are under the 'selected psu'
Try again by decreasing the 'selected psu' value

D. Detailed documentation

1. Interface organisation

The interface display all the options needed to perform temporal trend analysis through 4 successive panels (4 steps), from raw data managing to results display.
The panel 1 “1-Data_managment” focuses on the file and data management, it is the pre-processing part.
The panel 2 “2-Parameters_selection” focuses on the selection of the parameter from the categorical factors 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.

Results are displayed in the right part of the interface and are always visible.

Help buttons are available on each panel of the interface 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).

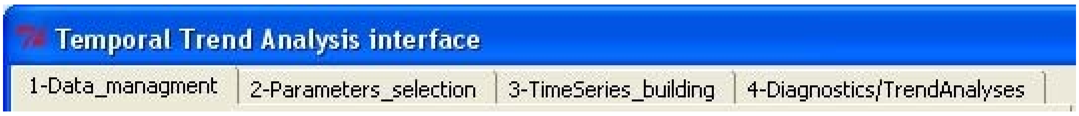

Temporal Trend Analysis interface
1-Data_managment
2-Parameters_selection
3-TimeSeries_building
4-Diagnostics/TrendAnalyses

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

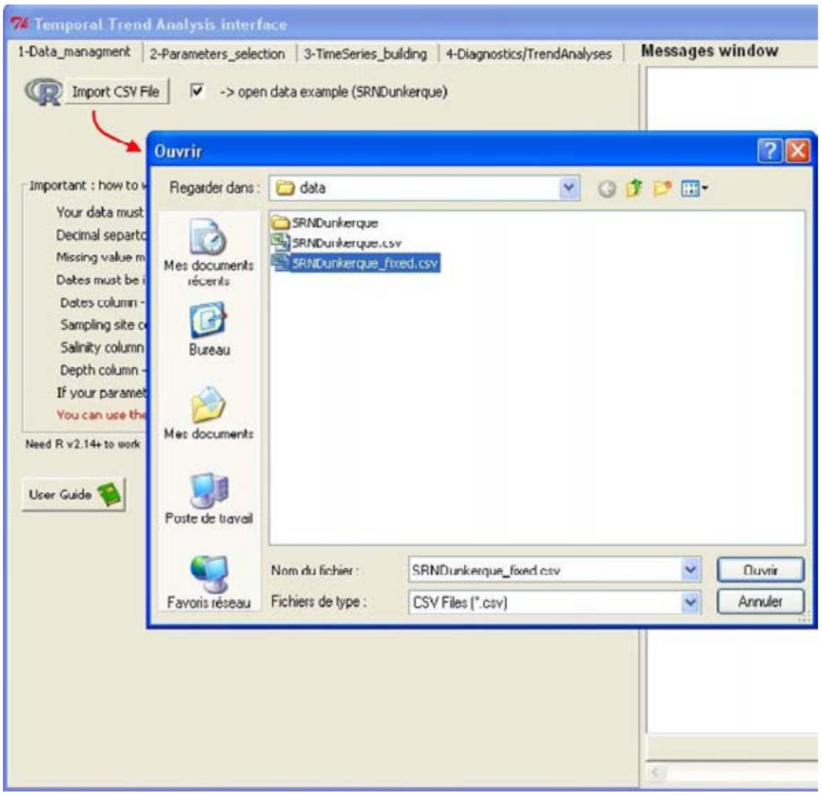
The advantage of having panels against windowed menus is that all options are always visible and can be rapidly selected without going into multiple menus. This is only possible because the number of options is optimised to the minimum needed to perform such analysis. Such interface cannot be developed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

2. Files and data managing panel

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

2.1 Import TXT file

To import a txt file containing your data just click on the button <Import TXT File> in panel 1, a window generates by your OS is display where you can select and open txt files (Fig. 22).



Important : how to w

Your data must
Decimal separto
Missing value m
Dates must be i
Dates column -
Sampling site c
Salinity column
Depth column -
If your paramet
You can use the

Regarder dans : data

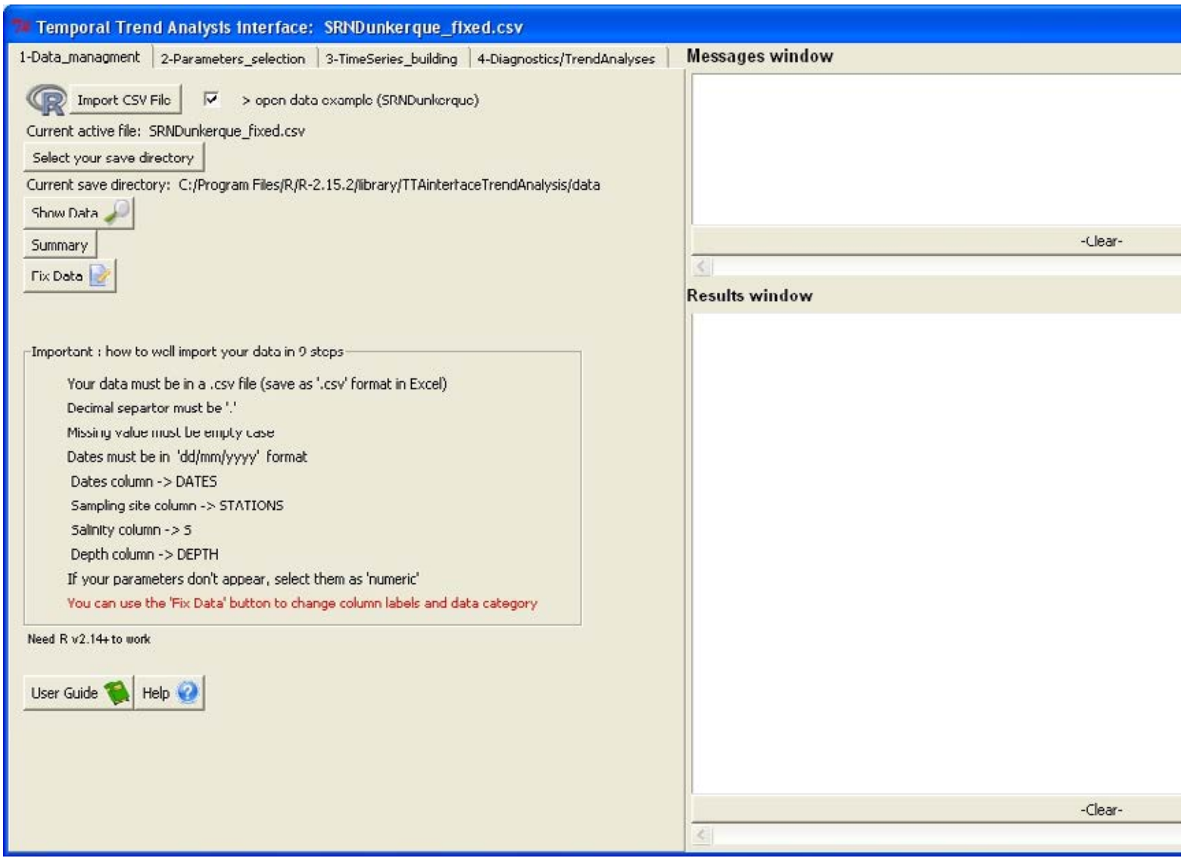
SRNDunkerque
SRNDunkerque.csv
SRNDunkerque_fixed.csv

Nom du fichier : SRNDunkerque_fixed.csv

Fichiers de type : CSV Files (*.csv)

Figure 22. Panel 1 with the <Import TXT File> option and importation advices.

Until the file is imported panels 2 to 4 stay empty and panel 1 uncompleted. The other options are available only when your data are imported (Fig. 23).



Current active file: SRNDunkerque_fixed.csv

Current save directory: C:/Program Files/R/R-2.15.2/library/TTAinterfaceTrendAnalysis/data

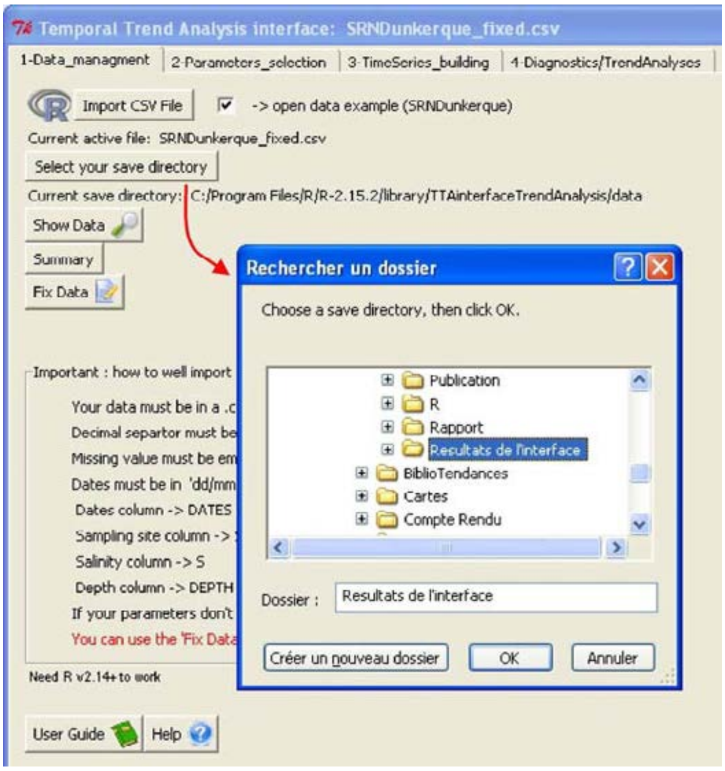
Important : how to well import your data in 9 steps

Your data must be in a .csv file (save as '.csv' format in Excel)
Decimal separtor must be '.'
Missing value must be empty case
Dates must be in 'dd/mm/yyyy' format
Dates column -> DATES
Sampling site column -> STATIONS
Salinity column -> S
Depth column -> DEPTH
If your parameters don't appear, select them as 'numeric'
You can use the 'Fix Data' button to change column labels and data category

Figure 23. Panel 1 with all options available once txt file has been imported.

2.2 Select your save directory

By default, results of analysis and figures are saved in the same directory as your txt 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 follows:



☑ -> open data example (SRNDunkerque)

Current active file: SRNDunkerque_fixed.csv

Current save directory: C:/Program Files/R/R-2.15.2/library/TTAinterfaceTrendAnalysis/data

Important : how to well import

Your data must be in a .c

Decimal separtor must be

Missing value must be em

Dates must be in 'dd/mm

Dates column -> DATES

Sampling site column ->

Salinity column -> S

Depth column -> DEPTH

If your parameters don't

You can use the 'Fix Data

Need R v2.14+ to work

Choose a save directory, then click OK.

Dossier : Resultats de l'interface

Figure 24. Panel 1 with the <Select your saved directory> option.

You can also directly create a folder through this window.

From this basal directory, the programme automatically creates an arborescence to save your files based on the options you choose to perform analyses (Fig. 25). Other options, such as Months, salinity and depths are not added in order to limit the arborescence declination and keeping a clear save path, therefore the user has to be careful to not overwrite its files (by changing the save directory destination) if these options are changed between two analyses.

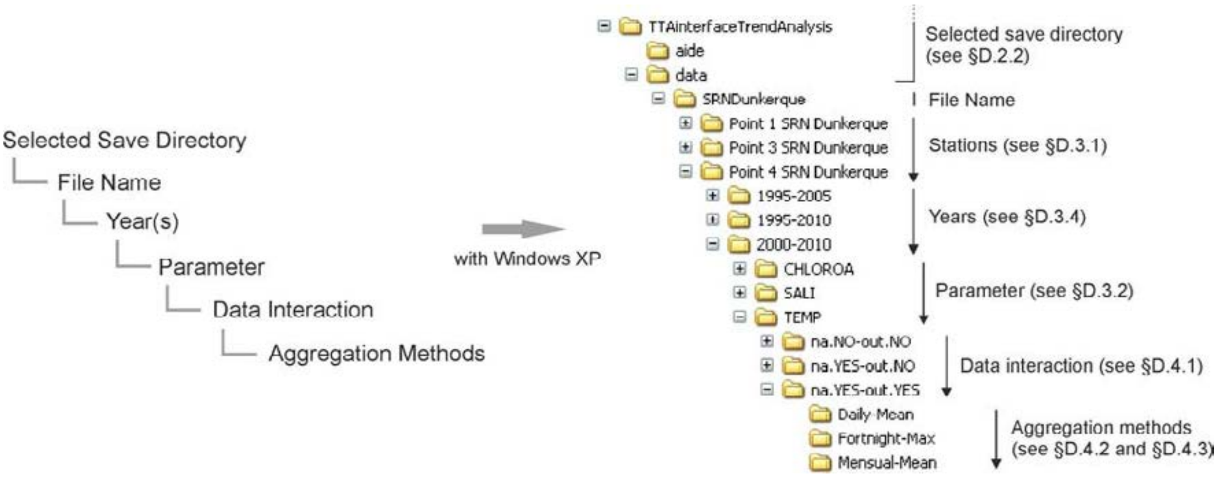


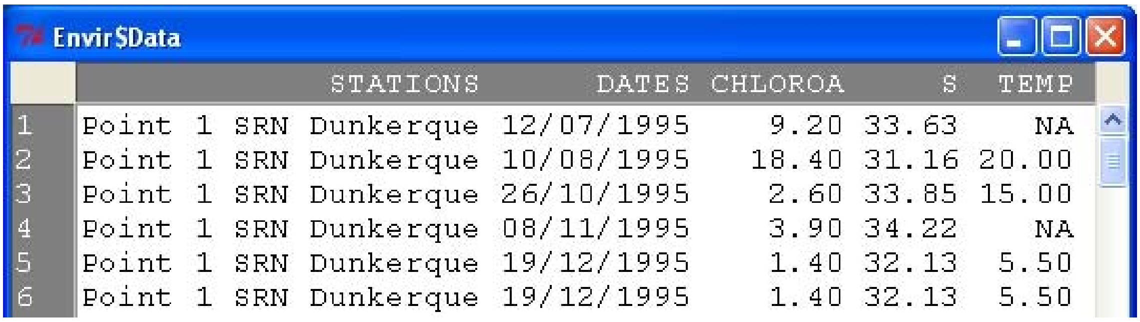
Figure 25. Example of a save path arborescence created as a function of selected options

Note that due to Windows OS limitation the save path cannot exceed ~255 characters, choose short station and parameters name if possible. Also, importing a new txt file will reset your basal save directory to default.

Saved files are named with a specific nomenclature:
OriginalFileName_TestName_ ParameterName.txt / .png.

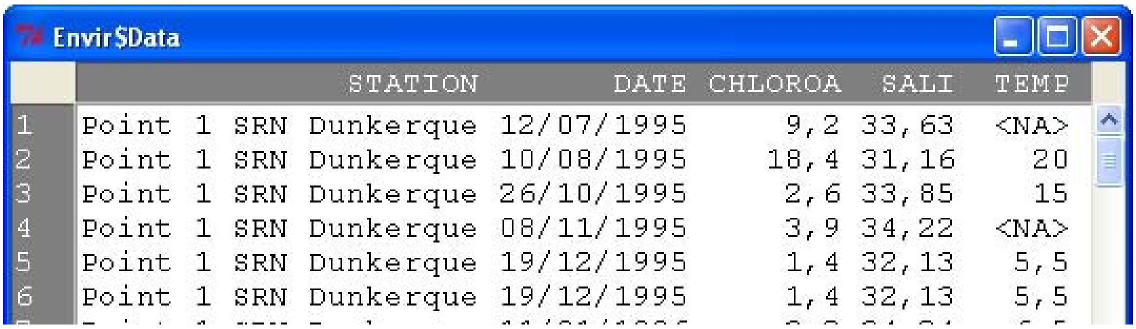
2.3 Display your raw data

You can check your imported data by clicking on <Show Data> button:



	STATIONS	DATES	CHLOROA	S	TEMP
1	Point 1 SRN Dunkerque	12/07/1995	9.20	33.63	NA
2	Point 1 SRN Dunkerque	10/08/1995	18.40	31.16	20.00
3	Point 1 SRN Dunkerque	26/10/1995	2.60	33.85	15.00
4	Point 1 SRN Dunkerque	08/11/1995	3.90	34.22	NA
5	Point 1 SRN Dunkerque	19/12/1995	1.40	32.13	5.50
6	Point 1 SRN Dunkerque	19/12/1995	1.40	32.13	5.50

Figure 26. A correctly imported data set view with the <Show Data> button in panel 1.



	STATION	DATE	CHLOROA	SALI	TEMP
1	Point 1 SRN Dunkerque	12/07/1995	9,2	33,63	<NA>
2	Point 1 SRN Dunkerque	10/08/1995	18,4	31,16	20
3	Point 1 SRN Dunkerque	26/10/1995	2,6	33,85	15
4	Point 1 SRN Dunkerque	08/11/1995	3,9	34,22	<NA>
5	Point 1 SRN Dunkerque	19/12/1995	1,4	32,13	5,5
6	Point 1 SRN Dunkerque	19/12/1995	1,4	32,13	5,5

Figure 27. Dataset with labels and decimal separator issues view with the <Show Data> button.

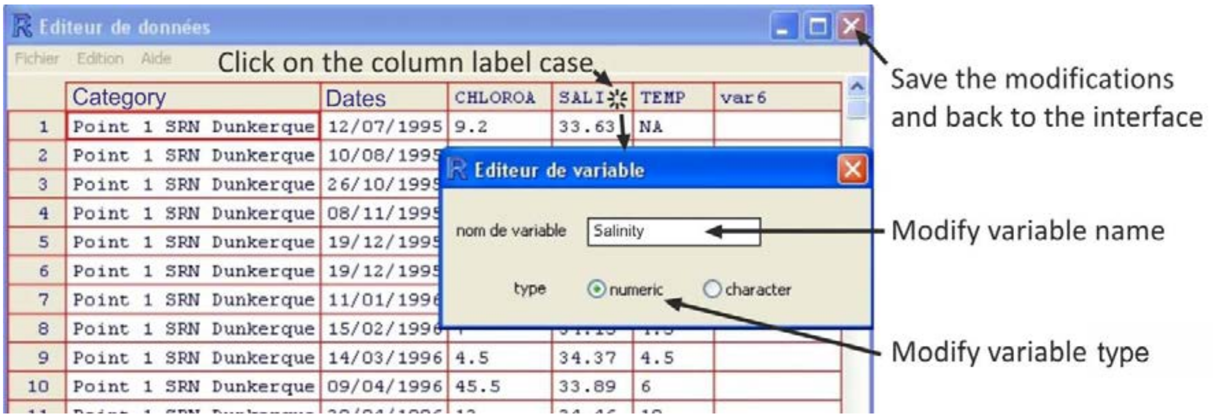
2.4 Edit your data – solve some importation issues

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 txt file is badly created; label and decimal separator do not correspond with the interface standard (Fig. 27). Therefore only column labels can be modified with <Fix Data> then you have to go back to the txt builder and check §B.

- Although the txt file is created following recommendations in §B parameters do not appear in panel 2. Therefore the problem is the category of the variable (numeric or character). 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 categorical factors site are character type.



Editeur de données
Fichier Edition Aide
Click on the column label case
Save the modifications and back to the interface
Modify variable name
Modify variable type

	Category	Dates	CHLOROA	SALI	TEMP	var6
1	Point 1 SRN Dunkerque	12/07/1995	9.2	33.63	NA	
2	Point 1 SRN Dunkerque	10/08/1995				
3	Point 1 SRN Dunkerque	26/10/1995				
4	Point 1 SRN Dunkerque	08/11/1995				
5	Point 1 SRN Dunkerque	19/12/1995				
6	Point 1 SRN Dunkerque	19/12/1995				
7	Point 1 SRN Dunkerque	11/01/1996				
8	Point 1 SRN Dunkerque	15/02/1996				
9	Point 1 SRN Dunkerque	14/03/1996	4.5	34.37	4.5	
10	Point 1 SRN Dunkerque	09/04/1996	45.5	33.89	6	

Editeur de variable

nom de variable: Salinity

type ● numeric ○ character

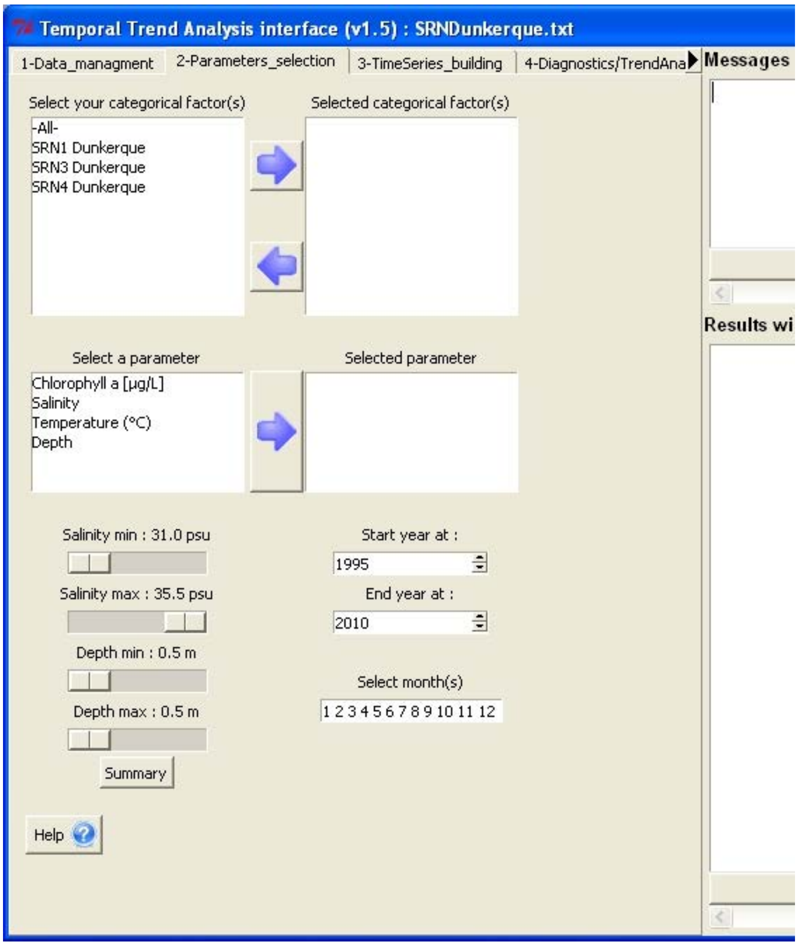
Figure 28. Data editor while <Fix Data> button is used.

By quitting the Fix Data panel, the new dataset is automatically read by the interface and a new txt file is saved with the nomenclature FileName_fixed.txt. Unfortunately data types cannot be saved in a txt file.

Editing your data do not change your save directory.

3. Parameters selection panel

If columns have been correctly labelled and categorised (see previous section), lists, sliders and frames should be automatically filled with appropriate values:



Select your categorical factor(s)

-All-
SRN1 Dunkerque
SRN3 Dunkerque
SRN4 Dunkerque

Selected categorical factor(s)

Select a parameter

Chlorophyll a [µg/L]
Salinity
Temperature (°C)
Depth

Selected parameter

Salinity min : 31.0 psu

Salinity max : 35.5 psu

Depth min : 0.5 m

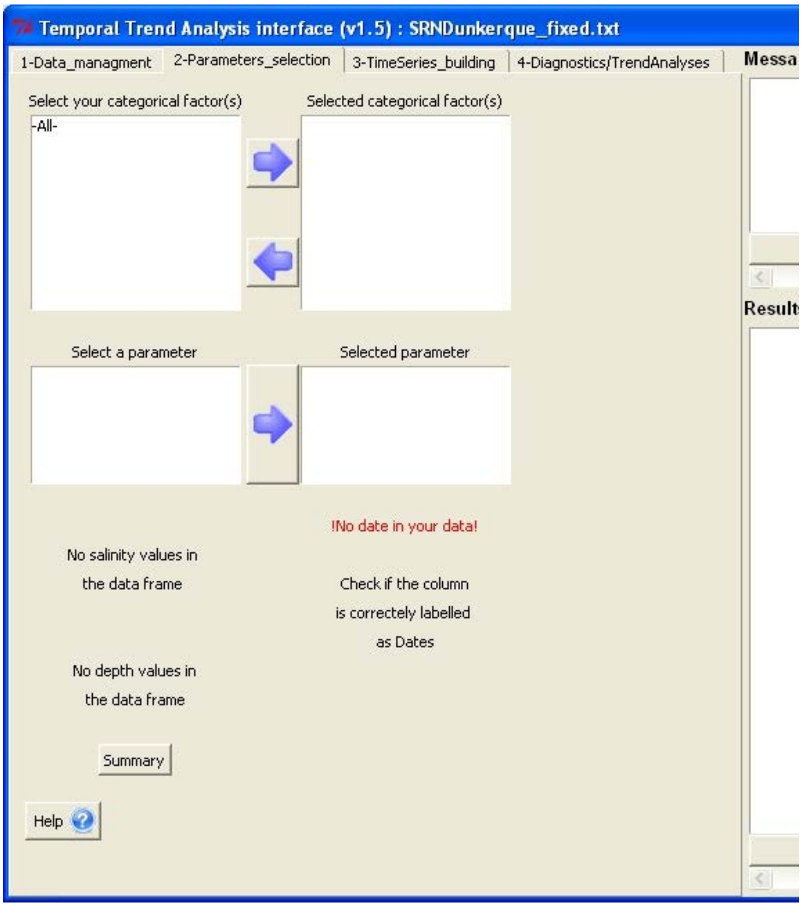
Depth max : 0.5 m

Start year at : 1995

End year at : 2010

Select month(s): 1 2 3 4 5 6 7 8 9 10 11 12

Otherwise:


Temporal Trend Analysis interface (v1.5) : SRNDunkerque_fixed.txt
1-Data_managment
2-Parameters_selection
3-TimeSeries_building
4-Diagnostics/TrendAnalyses
Messa
Select your categorical factor(s)
Selected categorical factor(s)
-All-
Select a parameter
Selected parameter
Result
Summary
Help

!No date in your data!

No salinity values in the data frame

Check if the column is correctely labelled as Dates

No depth values in the data frame

Let see details of each of these parts in the next section (§D.3.1 to §D.3.4).

3.1 The categorical factors

Categorical factors represent the category of your time series like sampling stations, taxa or chemical species. They can be redundant or unique in a database. The categorical factors to be analysed can be selected using the arrows between the two selection boxes (Fig. 29): just select the factors in the left box and click on the top arrow and the selected factors will appear in the right box. This supports multiple selection using the 'Ctrl' key or by dragging the cursor. All factors can be analysed at once by selecting -All- in the left box. To remove factors just select them in the right box and click on the bottom arrow, it also supports multiple selection.

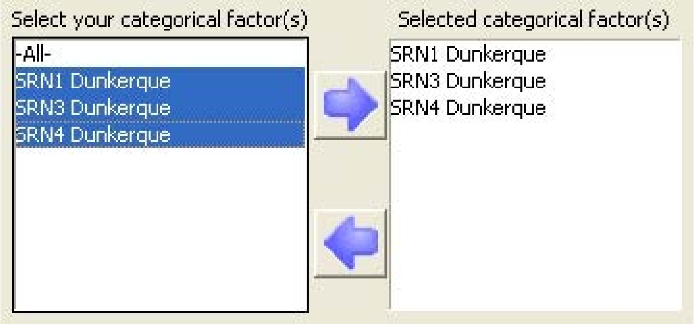
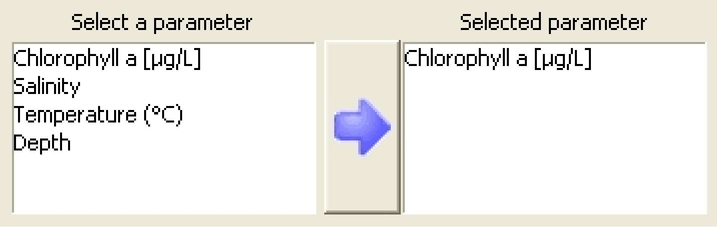


Figure 29. Selection boxes for categorical factors in panel 3, with selection arrows between.

3.2 The parameters

The process of selecting the parameter to be analysed is the same as for the categorical factors, except that only one parameter can be selected (it does not support multiple selection) and there is only one arrow for selection (no remove arrow) (Fig. 30). To replace a parameter already selected 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.

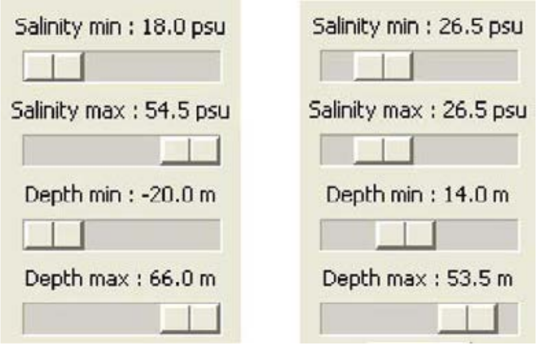


Chlorophyll a [µg/L]

Figure 30. Selection boxes for parameter in panel 3, with selection arrow between.

3.3 Depth and salinity

In some cases analyses have to be performed at specific depth or salinity (which characterise the water masses). In the panel 2, there are 4 sliders to select these salinities and depths (Fig. 31). By default these sliders display the maximum and minimum values of salinity and depth in your dataset (if they exist). By keeping these values unchanged all data are taken into account for analysis, including data at missing salinity and depth. These values can be modified by sliding the cursors on the left or right or by clicking on the area next to the cursor to have a more accurate increment (+/- 0.5 unit), therefore data at missing salinity or depth are excluded from the analysis. Analysis can be performed at a unique depth or salinity by giving the same min and max values.



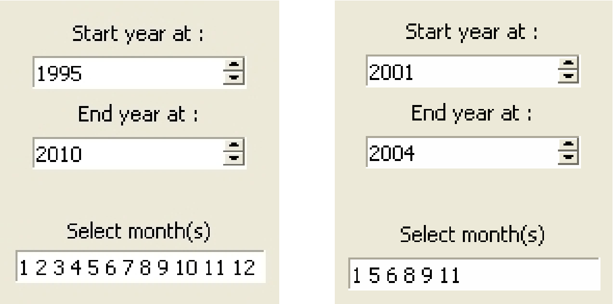
Salinity min : 18.0 psu
Salinity max : 54.5 psu
Depth min : -20.0 m
Depth max : 66.0 m

Salinity min : 26.5 psu
Salinity max : 26.5 psu
Depth min : 14.0 m
Depth max : 53.5 m

Figure 31. sliders to select specific salinity and depth before processing data. Left figure: default options; right figure: example of a selection of a given salinity (26.5) and a range of depth (14 to 53.5m)

3.4 Years and months

As for salinity and depth, years and months to be analysed can be modified. By default the two lists in panel 2 display the first and last years and the months present in your dataset (Fig. 32). Years can be modified just by clicking on the arrows or by typing it. Months can be deleted or added (the order is irrelevant) and there must be a space between the months. This can be useful to process data for a given period of a year, for example, to compare the productive period (in terms of phytoplankton development) versus the non-productive period of an area such as within the WFD.



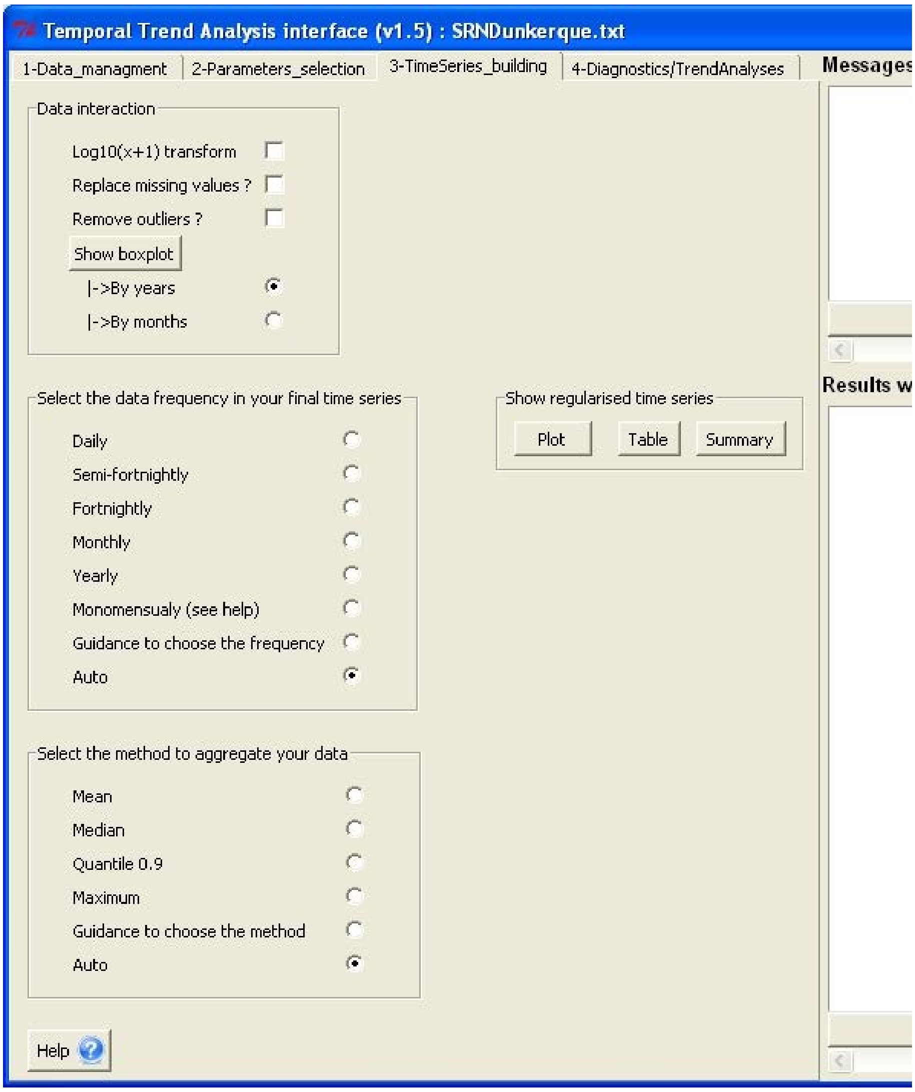
Start year at : 1995
End year at : 2010
Select month(s) 1 2 3 4 5 6 7 8 9 10 11 12

Start year at : 2001
End year at : 2004
Select month(s) 1 5 6 8 9 11

Figure 32. Boxes to select year range and months at which you want to perform analysis. Left figure: default options; right figure: analysis between 2001 and 2004 at 6 different months.

4. Time series building panel

The third panel focused on time series regularization before proceeding to temporal trend analysis (Fig. 33).



Temporal Trend Analysis interface (v1.5) : SRNDunkerque.txt

1-Data_managment | 2-Parameters_selection | 3-TimeSeries_building | 4-Diagnostics/TrendAnalyses

Messages

Results w

Data interaction
- Log10(x+1) transform □
- Replace missing values ? □
- Remove outliers ? □
- Show boxplot
  - |->By years ●
  - |->By months ○

Select the data frequency in your final time series
- Daily ○
- Semi-fortnightly ○
- Fortnightly ○
- Monthly ○
- Yearly ○
- Monomensualy (see help) ○
- Guidance to choose the frequency ○
- Auto ●

Show regularised time series
Plot | Table | Summary

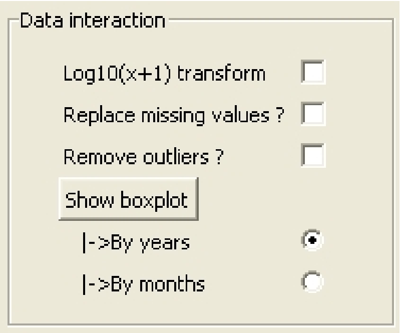
Select the method to aggregate your data
- Mean ○
- Median ○
- Quantile 0.9 ○
- Maximum ○
- Guidance to choose the method ○
- Auto ●

Help

Figure 33. 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. 34).



Data interaction

Log10(x+1) transform □

Replace missing values ? □

Remove outliers ? □

|->By years ●

|->By months ○

Figure 34. Data interaction frame options in panel 3.

Some diagnostic tools (spectrum analysis and seasonal decomposition, see §D.5.1) available in the GUI require regular time series and thus have regularly spaced measurement and contains no missing values. The option ‘Replace missing values’ replace the missing values from the time series (and not from the raw data) by values calculated from the data distribution. The final time series depend on the method of aggregation (see next paragraph); therefore missing values are calculated from aggregated data in two successive steps.

- Time series generally present strong autocorrelation, 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) of 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 uses a combination of both methods; the linear regression method to replace missing values in 3 successive units of time and the median method for longer period of missing values. The median method acts 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 or extreme natural event. In some cases these outliers can greatly influence statistical analysis comparatively to the rest of values and it should be interesting to remove them. The second option present in the frame (Fig. 34) allows you to remove these outliers and to save them in a separate txt 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, with Q1 being the first quartile (25% of data distribution) and Q3 the third quartile (75% of data distribution). The <Show boxplot>

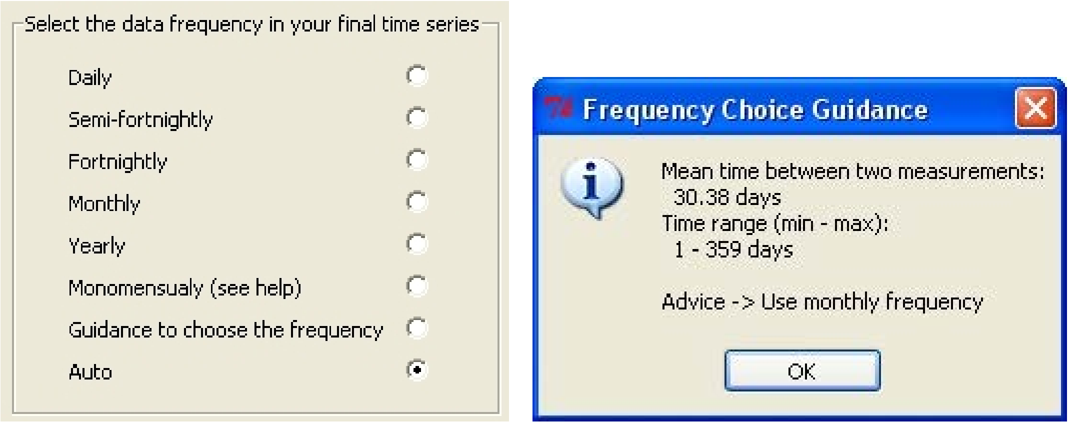
button displays the box and whiskers plot with outliers (Fig. 11), this is the boxplot() function of the {graphics} package.

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 Data frequency 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 performed. To build such regularised time series, the interface will aggregate raw data from the same selected period (day, 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. 35 and 36).

The second frame in panel 3 ‘Select the data frequency in your final time series’ allows the selection of the time step at which the interface aggregates the data to build the regularised time series. Eight options are available (Fig. 35); the 5 first options are classic frequencies, daily (all data by day), semi-fortnightly (all data by 7 days), fortnightly (all data by 15 days), monthly (all data by month), yearly (all data by years). Monomensualy time step aggregate all data by month, all years including.



Select the data frequency in your final time series

- Daily ○
- Semi-fortnightly ○
- Fortnightly ○
- Monthly ○
- Yearly ○
- Monomensualy (see help) ○
- Guidance to choose the frequency ○
- Auto ●

Frequency Choice Guidance

Mean time between two measurements:
30.38 days
Time range (min - max):
1 - 359 days

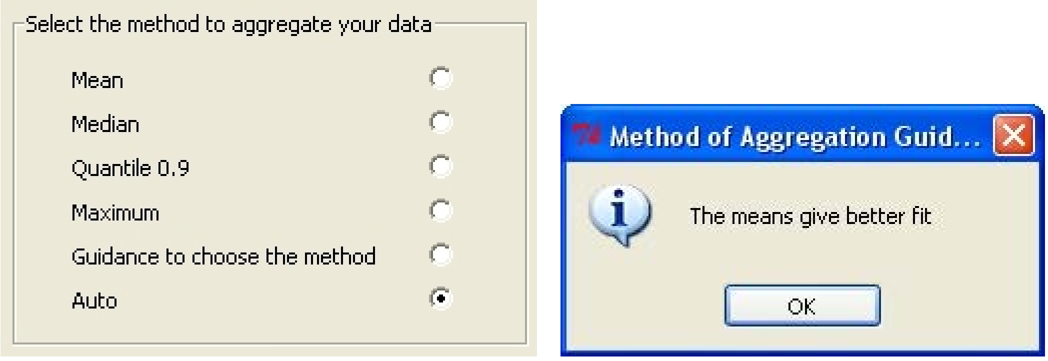
Advice -> Use monthly frequency

Figure 35. Data frequency selection frame and advice (guidance option).

It is better to choose a time step in relation with the theoretical sampling frequency of your data in order to keep the maximum of information without creating too many missing values. The option <Guidance to choose the time step> suggest a balanced choice by computing the mean time and the minimum and maximum period that separates two successive measurements in your database. This method is inspired from the {pastec} package (Grosjean and Ibanez, 2002). 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 to select another time step knowing all the consequences that this could have for the interpretation of the results. The auto option (default option) will automatically apply the advice without displaying the suggestion. In some case, this method of aggregation is sufficient to remove all missing values in the regularised time series without using the <Replace missing values> option in §D.4.1.

4.3 Method of aggregation of your time series

The third frame ‘Select the method to aggregate your data’ (Fig. 36) allow the method with which data will be aggregated at the time step previously set to be chosen. Four methods are available: by averaging the data (Mean), 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 best fits the original data distribution. The interface compares data distribution obtained with each method (at the selected time step) with the raw data distribution using an ANOVA with Dunnett’s post-hoc test. The comparison with the highest p-value (less significant difference) determines the best method. You are also free to follow these suggestions or to select another method. The auto option (default option) automatically apply the advice without displaying the suggestion.



Select the method to aggregate your data

Mean ○
Median ○
Quantile 0.9 ○
Maximum ○
Guidance to choose the method ○
Auto ●

Method of Aggregation Guid...

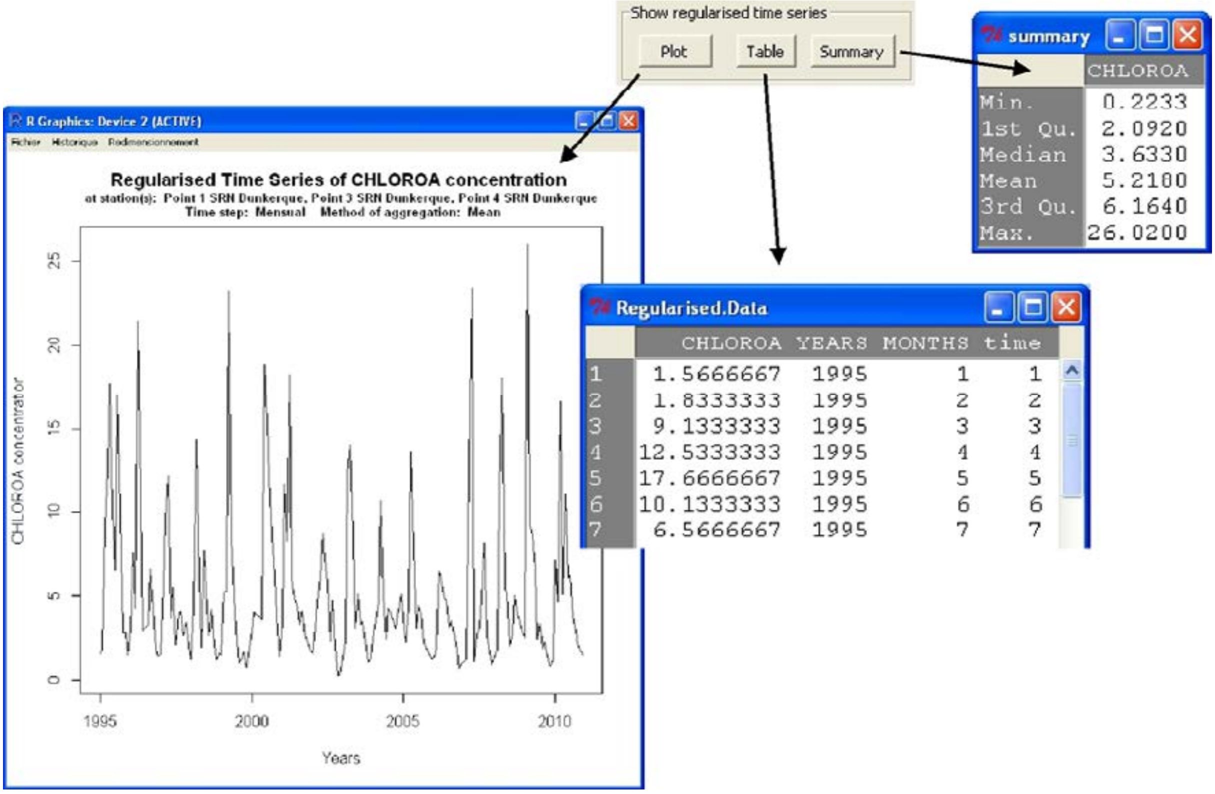
The means give better fit

OK

Figure 36. Aggregation method selection frame and method advice (guidance option).

4.4 Visualised your regularized time series

The fourth frame ‘Show regularised time series’ (Fig. 37) display the 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 first column label is the parameter 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). Week.year and week.month column are present only if fortnight or semi-fortnight frequency is selected. The DayYears column (present only if you choose daily frequency) show the number of the day within a year; in the interface a year has 366 days, non bissextile years have an extra day at the end of December with <NA> as parameter value. The ‘time’ column count the time between regularised measurements (the value of unit depend on the selected time step).


Show regularised time series
Plot
Table
Summary
R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Regularised Time Series of CHLOROA concentration
at station(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4 SRN Dunkerque
Time step: Mensual Method of aggregation: Mean
CHLOROA concentrator
0
5
10
15
20
25
1995
2000
2005
2010
Years
summary
Regularised.Data

	CHLOROA
Min.	0.2233
1st Qu.	2.0920
Median	3.6330
Mean	5.2100
3rd Qu.	6.1640
Max.	26.0200

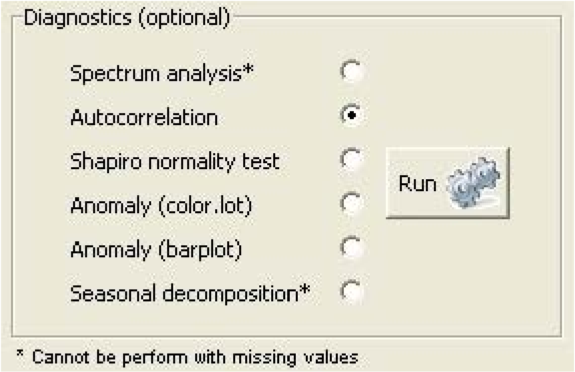
	CHLOROA	YEARS	MONTHS	time
1	1.5666667	1995	1	1
2	1.8333333	1995	2	2
3	9.1333333	1995	3	3
4	12.5333333	1995	4	4
5	17.6666667	1995	5	5
6	10.1333333	1995	6	6
7	6.5666667	1995	7	7

Figure 37. Regularised time series display options frame.

5. Diagnostics and statistics panel

5.1 Diagnostic tools

The options present in the first frame of the forth panel ‘Diagnostics (optional)’ (Fig. 35) are not required to perform temporal trend analysis but give additional information that can be useful to explain some patterns in the time series.



Diagnostics (optional)

Spectrum analysis* ○
Autocorrelation ●
Shapiro normality test ○
Anomaly (color.lot) ○
Anomaly (barplot) ○
Seasonal decomposition* ○

* Cannot be perform with missing values

Figure 38. Diagnostic frame in panel 4

- **Spectrum analysis** allows to estimates the spectral density (discrete Fourier transform) of time series and to display a periodogram (Fig. 39). 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 performed with missing values. Use the spectrum() function of the {stats} package. For more information about Fourier transform and periodogram in R see also Shummway and Stoffer (2006).

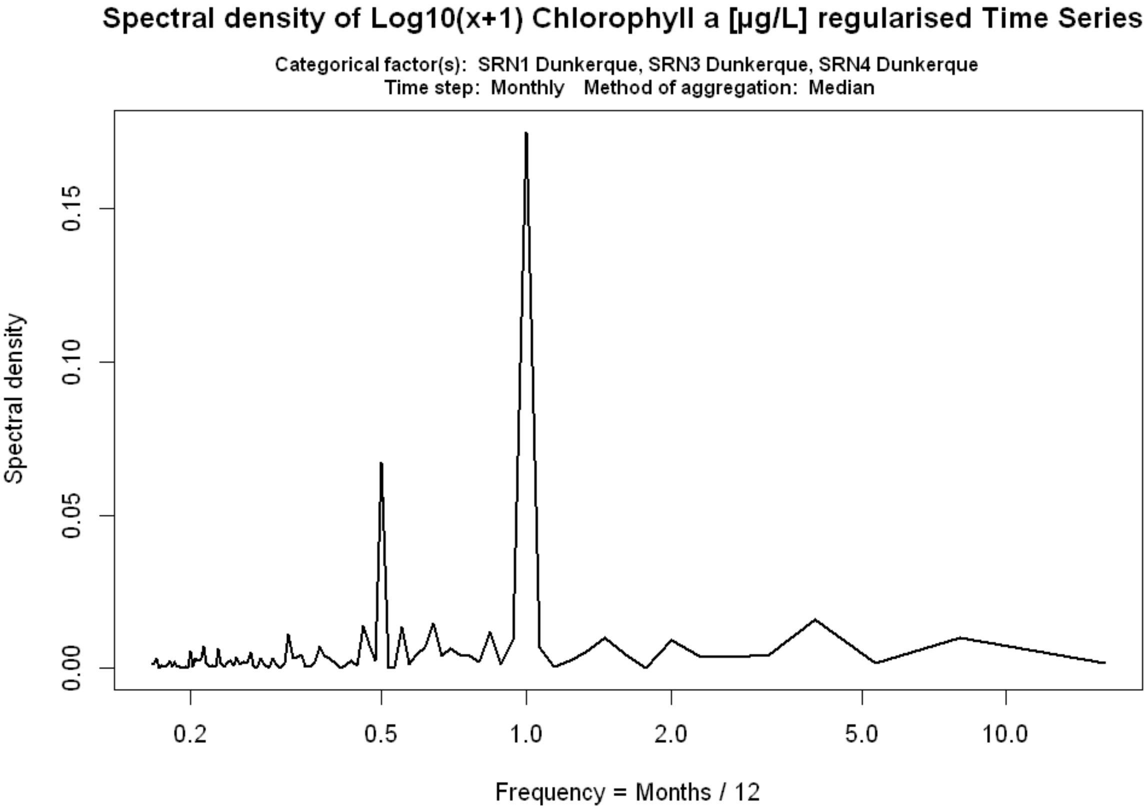


Figure 39. Spectrum density of monthly regularized time series of chlorophyll-a concentration (log10(x+1) transformed) at Dunkerque between 1995 and 2010. Save as “SRNDunkerque_Spectrum_Chlorophyll a [µg.png”.

Interpretation:

In this case the identify cycle (by arrow) is the annual cycle that show the highest spectrum at frequency = 1 (12 months). 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. The second frequency you can detect is 0.5 = 12*0.5 = 6 months.

- **Autocorrelation** computes (and plots with confidence interval at 0.95) estimates of the autocorrelation function (Fig. 40). As for spectrum analysis the most frequently highest autocorrelation are 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).

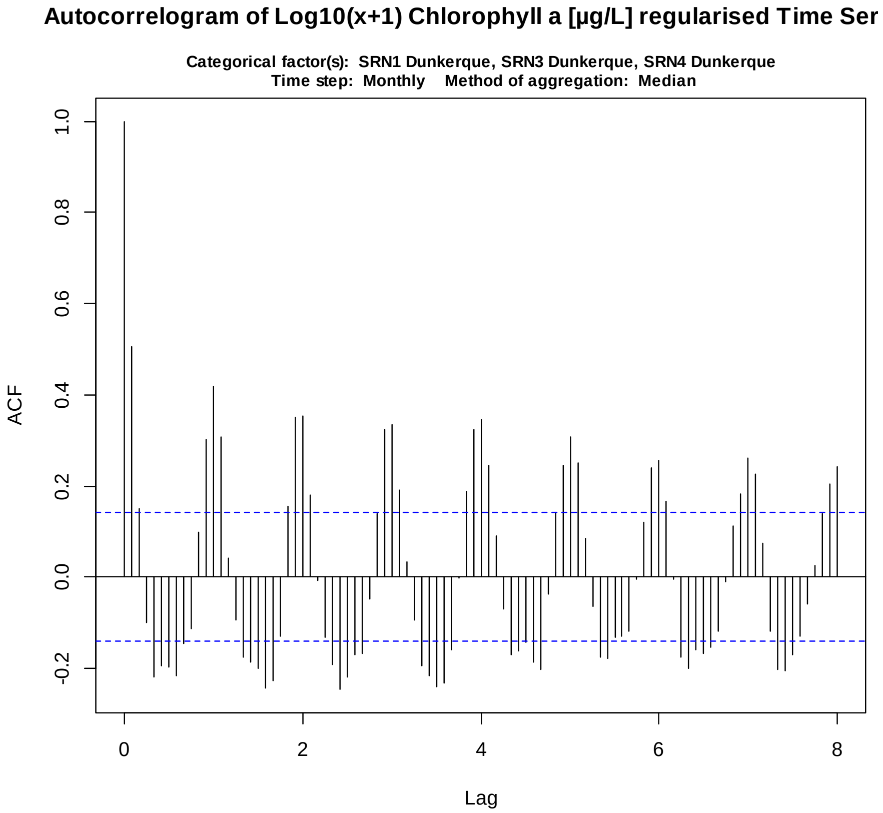


Figure 40. A. Autocorrelation of monthly regularized time series of chlorophyll-a concentration (log10(x+1) transformed) at Dunkerque between 1995 and 2010. Save as “SRNDunkerque_AutoCorr_Chlorophyll a [µg.png”.

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). This is the shapiro.test() function of the {stats} package.

- **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. 41). It produces a contour plot with the areas between the contours filled in solid colour. Red colours show positive anomaly and blue colours negative anomalies. White areas occur when there are missing values. This option works only with time series build at monthly, semi-fortnight and fortnight time step. Use the filled.contour() function of the {graphics} package.

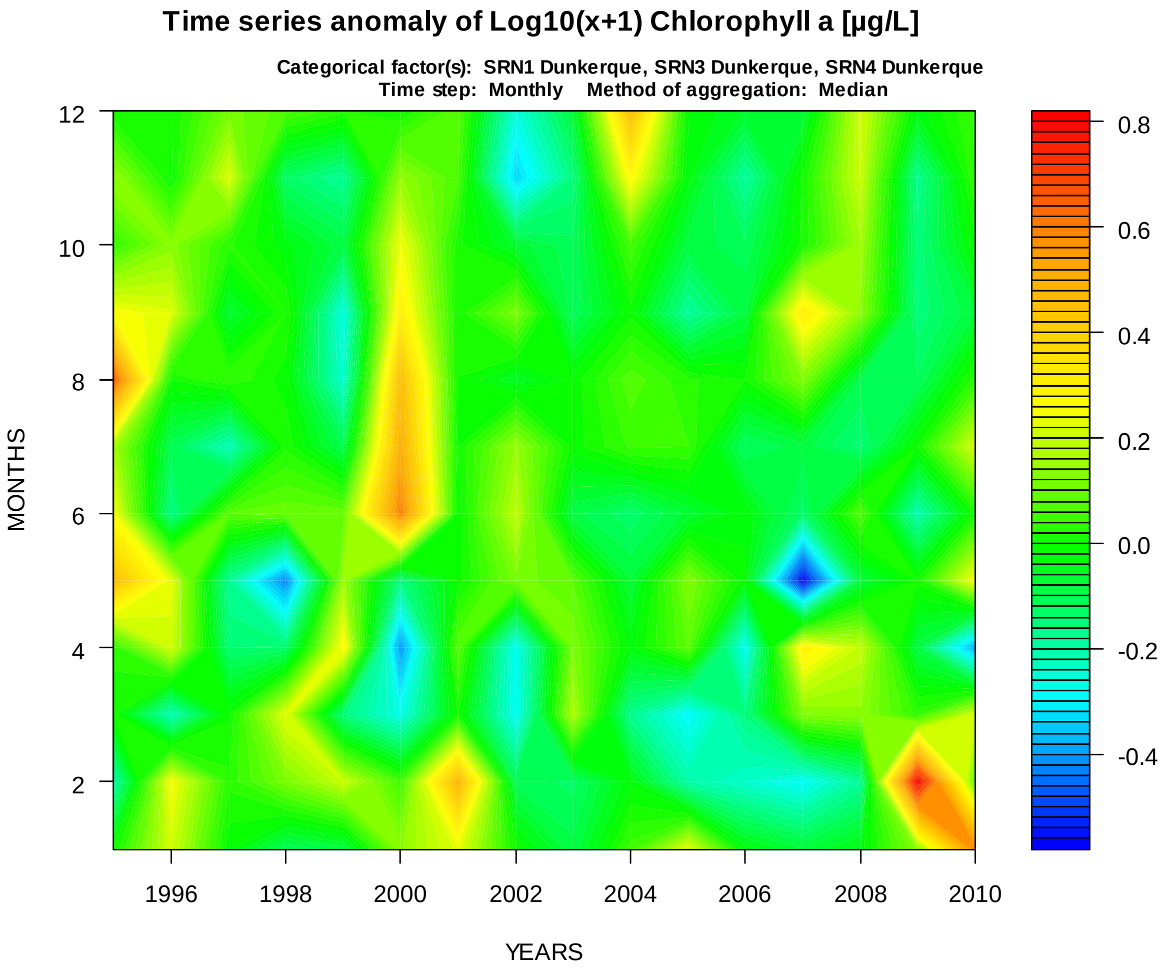


Figure 41. Color plot of chlorophyll-a concentration (log10(x+1) transformed) anomalies at Dunkerque between 1995 and 2011 at monthly scale (missing values are replaced here). Cold and hot areas represent respectively negative and positive anomaly. Save as “SRNDunkerque_ColorPlot_Chlorophyll a [µg.png”.

Interpretation:

Positive anomalies (higher than ‘normal’) of chlorophyll-a concentration can be observed in 2000 and 2009 in summer and February respectively. Negative anomalies (lower than ‘normal’) are observed in early spring 2000 and 2002.

- **Anomaly (barplot)** displays a bar plot that show the anomalies of the time series calculated for each time step. Each anomaly is the difference between the value at the time step and the median of the entire regularized time series. Values that are under this median are negative anomalies (blue bars in the figure) and values over this median are positive anomalies (red bars in the figure). Use the barplot() function of the {graphics} package.

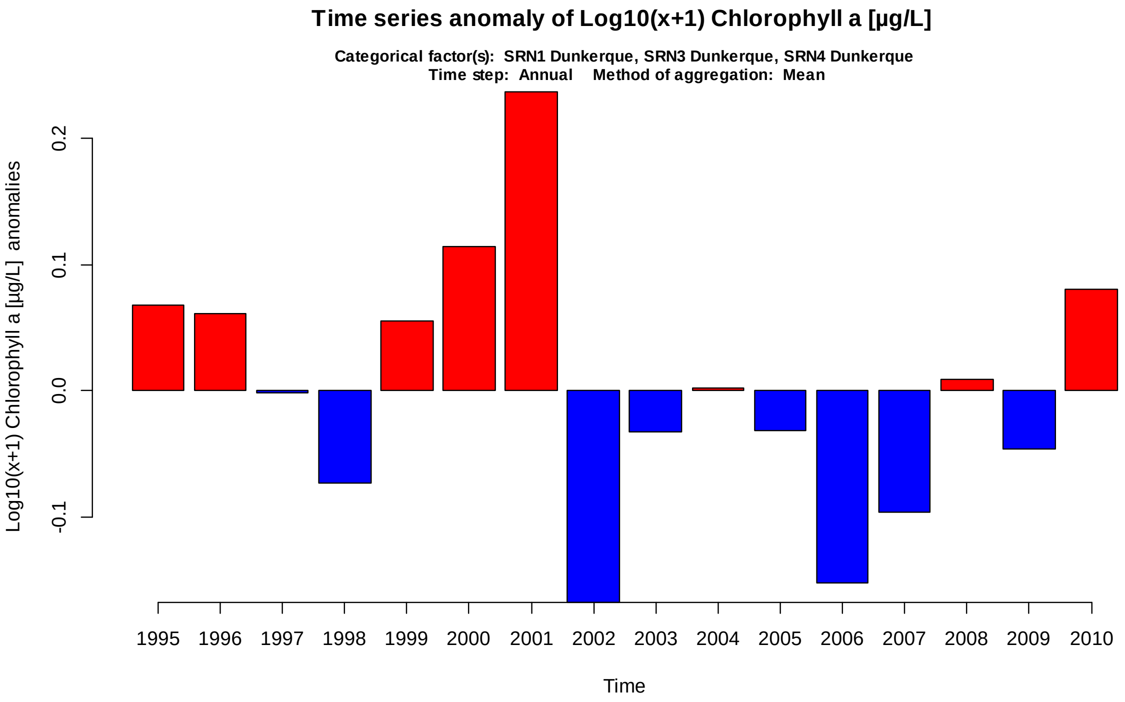


Figure 42. Bar plot of annual chlorophyll a concentration (log10(x+1) transformed) anomalies at Dunkerque between 1995 and 2010. Save as “SRNDunkerque_Anomaly BarPlot_Chlorophyll a [µg.png”.

Interpretation:

Chlorophyll a concentrations show succession of periods of negative and positive anomalies. The period of the time series before 2002 is characterised by more positive anomalies whereas after 2002 there is more negative anomalies showing an overall trend to decreasing chlorophyll a concentration (with a small increase at the end of the time series).

- **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.

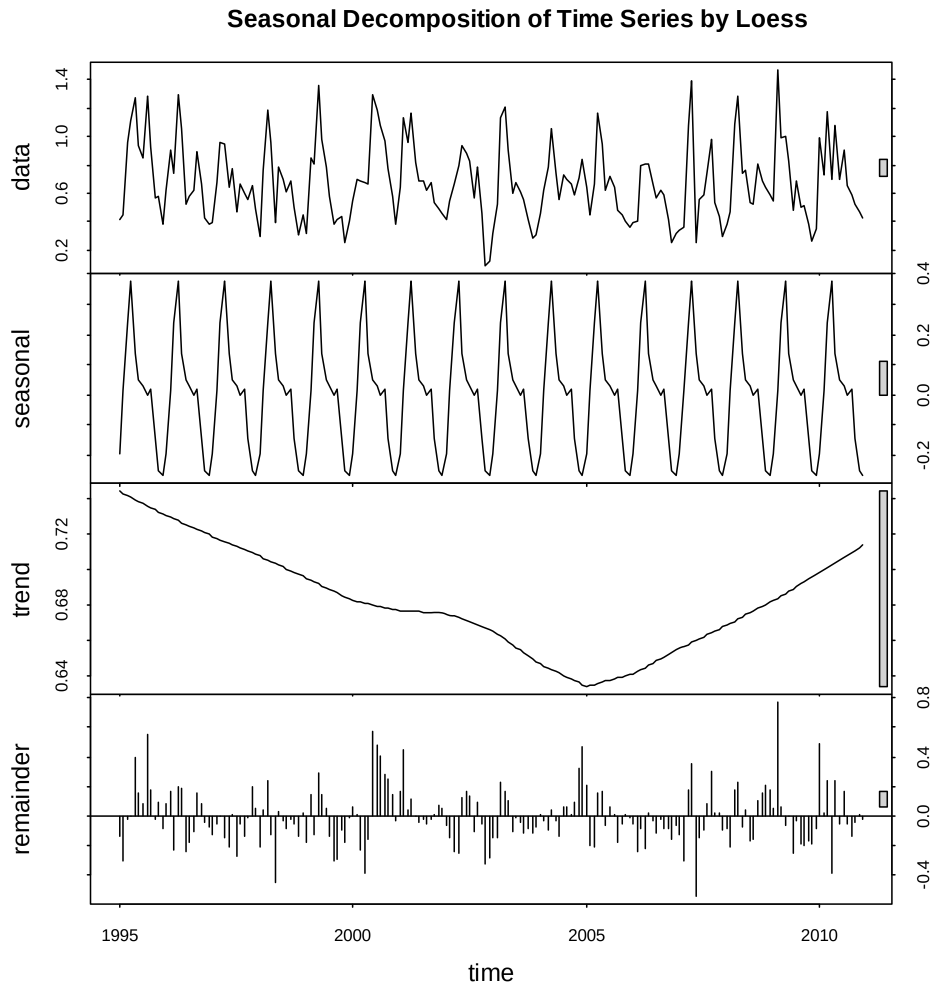


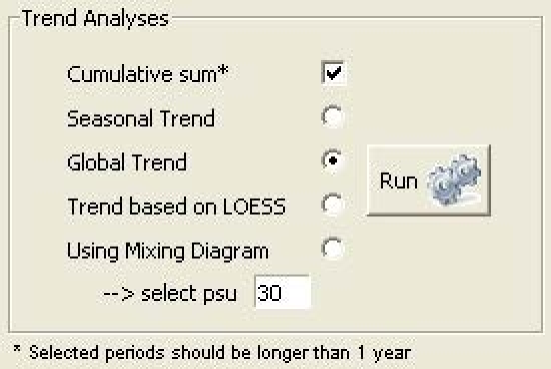
Figure 43. Plots obtain with the loess decomposition of the chlorophyll a regularized time series with from top plot to bottom plot: the regularized time series, the seasonal component, the global trend, and the remainder.

Interpretation:

In this case chlorophyll-a concentration variations show a seasonal cycle characterised by a strong peak value followed by a smaller one (2d plot). The overall trend shows an increase from 1995 to 2005 and an increase from 2005 to 2011. Remainders (4th plot) show few seasonality patterns, the major part of chlorophyll-a concentration variation is due to the seasonality and inter-annual pattern.

5.2 Temporal trend tests

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



Trend Analyses

Cumulative sum* ☑

Seasonal Trend ○

Global Trend ●

Trend based on LOESS ○

Using Mixing Diagram ○

--> select psu 30

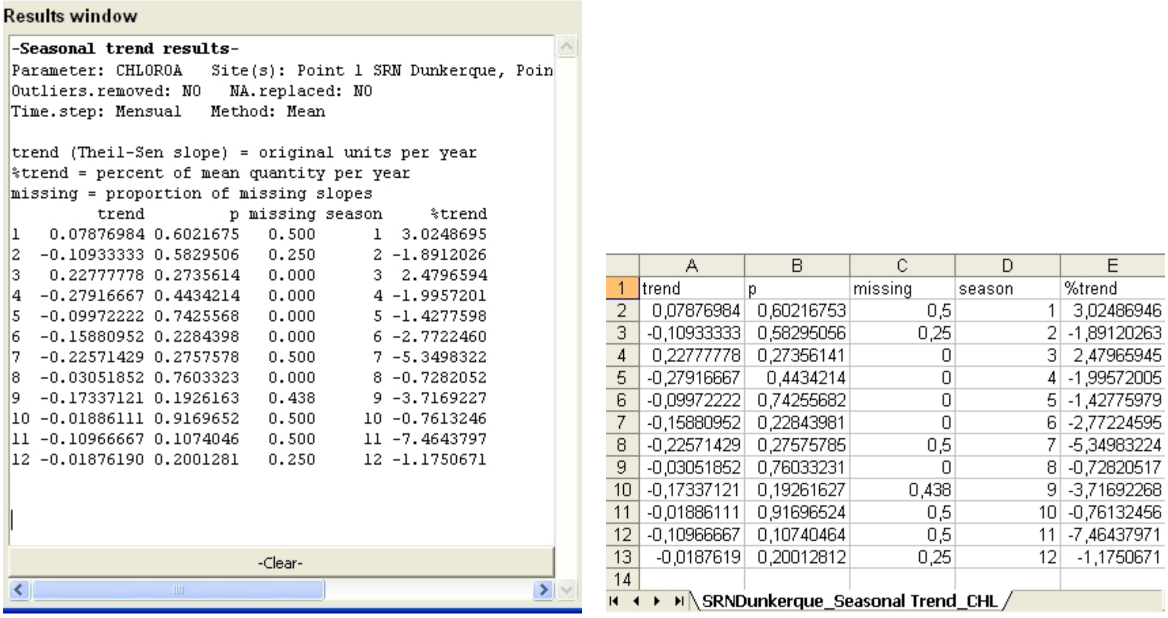
Run

* Selected periods should be longer than 1 year

Figure 44. Trend analysis frame in panel 4.

- **Seasonal Trend** performs 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 assessment. 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 test between seasons and performing a Sen's Slope estimator to estimate a value of this trend (median between ranks). This is the seasonTrend() function of the {wq} package. For more information about Kendall test see Hirsch et al. (1982) and Hirsch and Slack (1984).

  Results are displayed in the right part of the interface and saved in a txt 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; season column = counting of time step succession (1 to 12 for monthly, 1 to 24 for fortnightly etc…); %trend column = percent of mean quantity per year at the selected time step.



```
-Seasonal trend results-
Parameter: CHLOROA   Site(s): Point 1 SRN Dunkerque, Poin
Outliers.removed: NO   NA.replaced: NO
Time.step: Mensual   Method: Mean

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   0.07876984 0.6021675   0.500      1  3.0248695
2  -0.10933333 0.5829506   0.250      2 -1.8912026
3   0.22777778 0.2735614   0.000      3  2.4796594
4  -0.27916667 0.4434214   0.000      4 -1.9957201
5  -0.09972222 0.7425568   0.000      5 -1.4277598
6  -0.15880952 0.2284398   0.000      6 -2.7722460
7  -0.22571429 0.2757578   0.500      7 -5.3498322
8  -0.03051852 0.7603323   0.000      8 -0.7282052
9  -0.17337121 0.1926163   0.438      9 -3.7169227
10 -0.01886111 0.9169652   0.500     10 -0.7613246
11 -0.10966667 0.1074046   0.500     11 -7.4643797
12 -0.01876190 0.2001281   0.250     12 -1.1750671
```

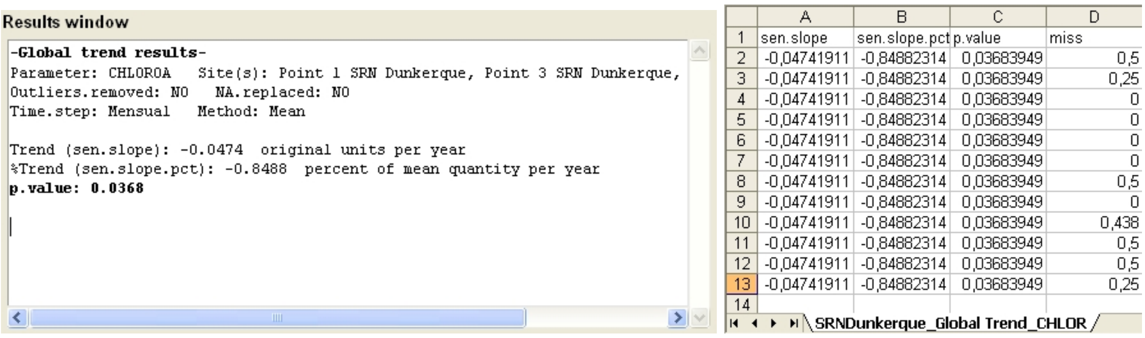
	A	B	C	D	E
1	trend	p	missing	season	%trend
2	0,07876984	0,60216753	0,5	1	3,02486946
3	-0,10933333	0,58295056	0,25	2	-1,89120263
4	0,22777778	0,27356141	0	3	2,47965945
5	-0,27916667	0,4434214	0	4	-1,99572005
6	-0,09972222	0,74255682	0	5	-1,42775979
7	-0,15880952	0,22843981	0	6	-2,77224595
8	-0,22571429	0,27575785	0,5	7	-5,34983224
9	-0,03051852	0,76033231	0	8	-0,72820517
10	-0,17337121	0,19261627	0,438	9	-3,71692268
11	-0,01886111	0,91696524	0,5	10	-0,76132456
12	-0,10966667	0,10740464	0,5	11	-7,46437971
13	-0,0187619	0,20012812	0,25	12	-1,1750671
14					

Figure 45. Results of the Seasonal Trend analysis (monthly scale) display in the right part of the interface (left) and saved in a txt file (right) as “OriginalName_Seasonal Trend_Parameter.txt”.

Interpretation:

The trend of chlorophyll-a concentration vary as a function of the considered month, for example season 1 (first row = January in our example) show a trend of 0.078 µg/L/year (trend column) that correspond to - 3% µg/L/year (column %trend). However p.values shows that no trend is significant (all p-value > 0.05). Missing values are observed in our time series (missing column).

- **Global Trend** does the same test as above but gives 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 is the seaKen() function of the {wq} package.



```
-Global trend results-
Parameter: CHLOROA   Site(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque,
Outliers.removed: NO   NA.replaced: NO
Time.step: Mensual   Method: Mean

Trend (sen.slope): -0.0474  original units per year
%Trend (sen.slope.pct): -0.8488  percent of mean quantity per year
p.value: 0.0368
```

	A	B	C	D
1	sen.slope	sen.slope.pct	p.value	miss
2	-0,04741911	-0,84882314	0,03683949	0,5
3	-0,04741911	-0,84882314	0,03683949	0,25
4	-0,04741911	-0,84882314	0,03683949	0
5	-0,04741911	-0,84882314	0,03683949	0
6	-0,04741911	-0,84882314	0,03683949	0
7	-0,04741911	-0,84882314	0,03683949	0
8	-0,04741911	-0,84882314	0,03683949	0,5
9	-0,04741911	-0,84882314	0,03683949	0
10	-0,04741911	-0,84882314	0,03683949	0,438
11	-0,04741911	-0,84882314	0,03683949	0,5
12	-0,04741911	-0,84882314	0,03683949	0,5
13	-0,04741911	-0,84882314	0,03683949	0,25
14				

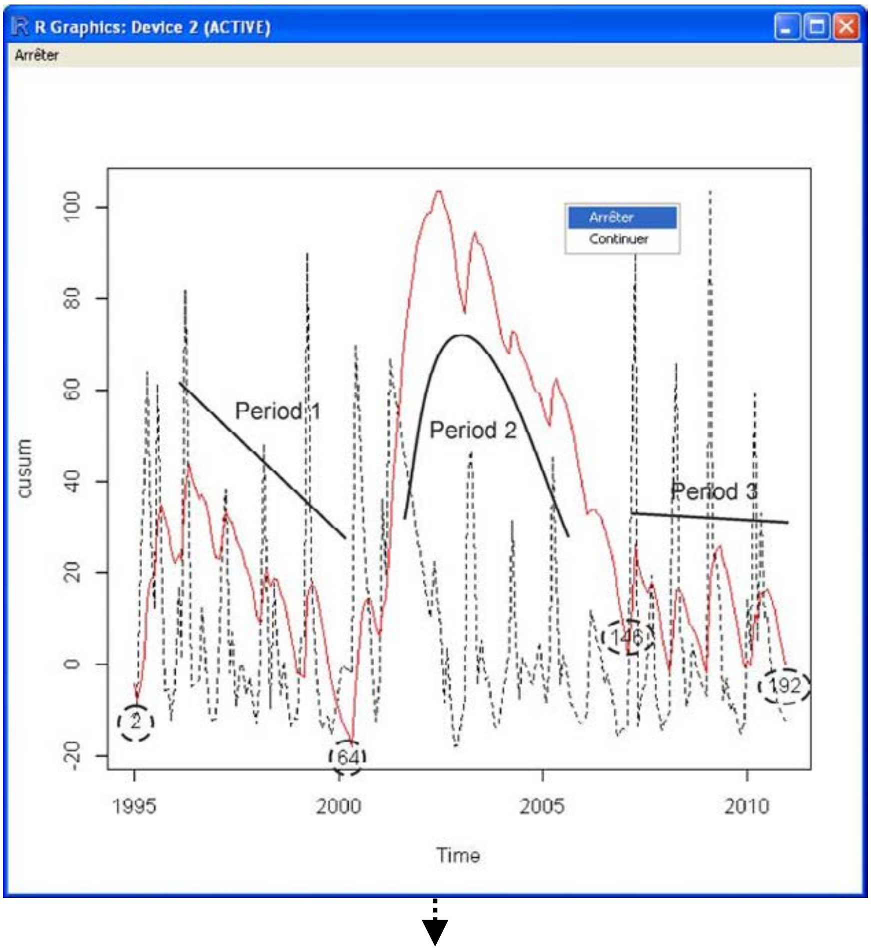
Figure 46. Results of the Global Trend analysis (monthly scale) display in the right part of the interface (left) and saved in a txt file (right) as “OriginalName_Global Trend_Parameter.txt”.

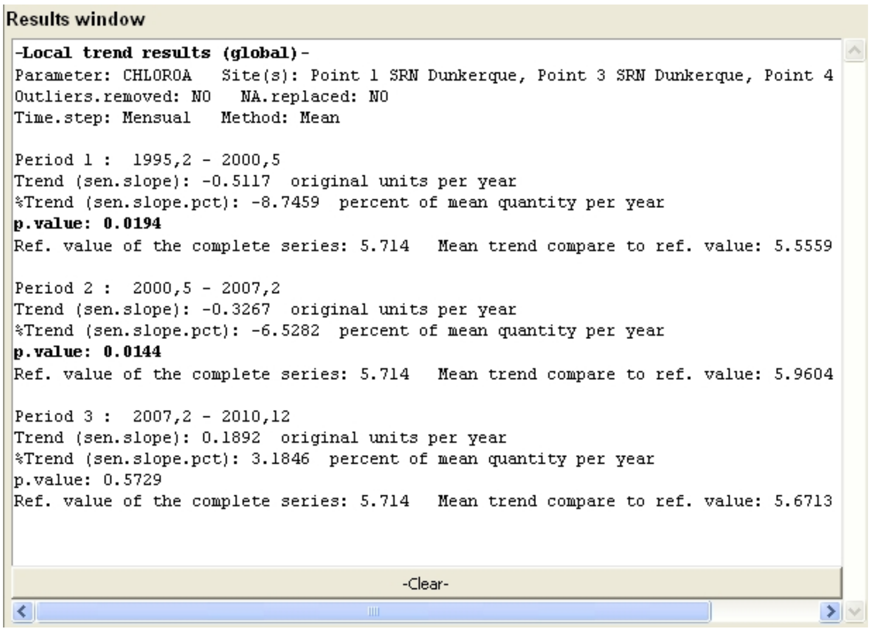
Interpretation:

The global trend of chlorophyll-a concentration is significantly (p.value < 0.05) low with -0.0474 µg/L/year, resulting from the strong variation among months (Fig. 43). It represents -0.8% µg/L/year. The time series shows missing value (miss column in the right figure).

- **Cumulative sum** plots a cumulative sum curve of the time series and allow to manually identifying changes of the pattern (shift, trend) in your time series (Fig. 47). This method comes from the cusum function of the pastecs package. For more information about cusum function see also Ibanez et al. (1993). Once the periods are identified, the program performs 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 returns 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 missing values), 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, a window with cusum curve in red will appear. The different points where there are changes in tendency have to be manually selected by left-clicking on the curve (Fig. 46). 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.





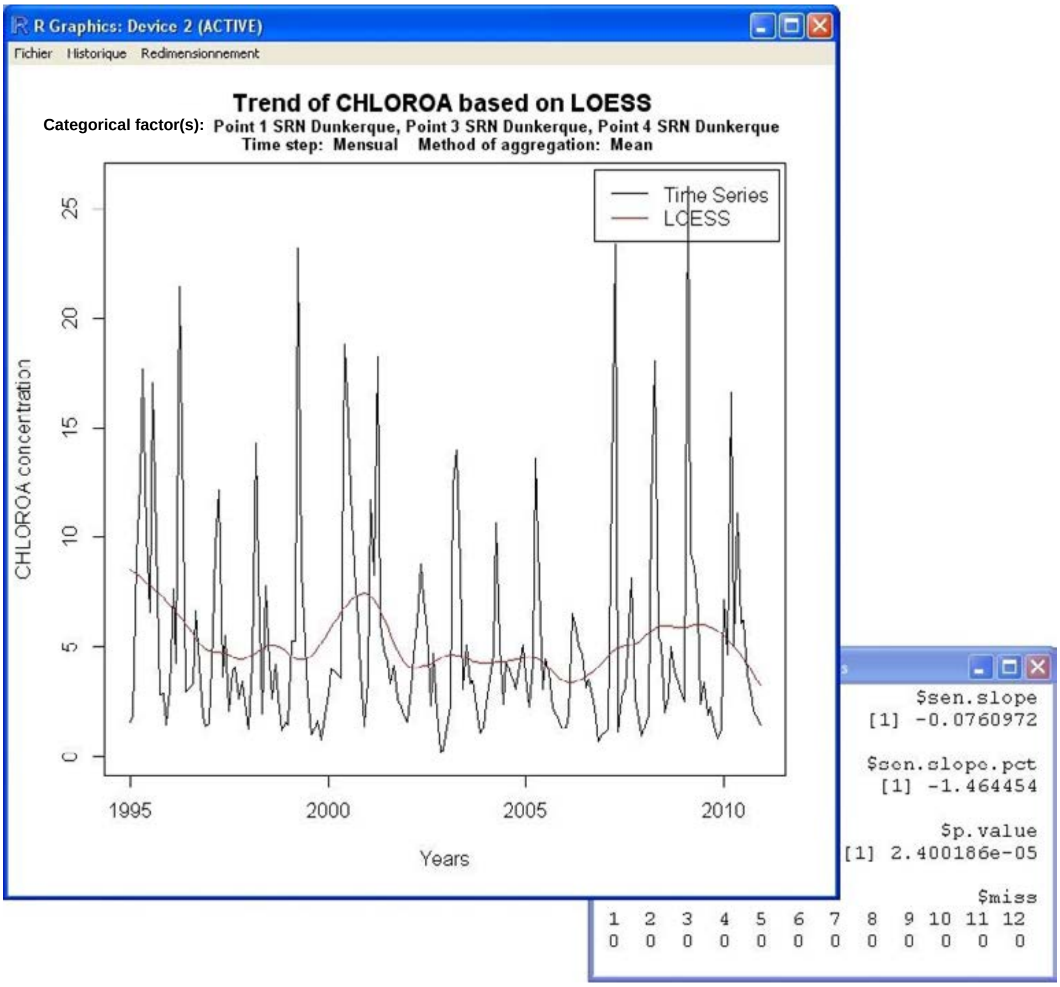
```
-Local trend results (global)-
Parameter: CHLOROA   Site(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4
Outliers.removed: NO   NA.replaced: NO
Time.step: Mensual   Method: Mean

Period 1 :  1995,2 - 2000,5
Trend (sen.slope): -0.5117  original units per year
%Trend (sen.slope.pct): -8.7459  percent of mean quantity per year
p.value: 0.0194
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.5559

Period 2 :  2000,5 - 2007,2
Trend (sen.slope): -0.3267  original units per year
%Trend (sen.slope.pct): -6.5282  percent of mean quantity per year
p.value: 0.0144
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.9604

Period 3 :  2007,2 - 2010,12
Trend (sen.slope): 0.1892  original units per year
%Trend (sen.slope.pct): 3.1846  percent of mean quantity per year
p.value: 0.5729
Ref. value of the complete series: 5.714   Mean trend compare to ref. value: 5.6713
```

Figure 47. Top figure: chlorophyll-a concentration (µg/l) variations near Gravelines (hatched black line) with cusum plot (red line) and different periods identified (solid black lines). Bottom figure: results of a global trend apply on each of these periods display in the right part of the interface.

Interpretation:

The global trend of chlorophyll-a concentration is significantly negative during the two first periods (1995-2000 and 2000-2007), with -8.7% and -6.5% of µg/L/year. The positive trend during the last period is not significant (p.value > 0.05). These decreases are observed only during these periods and not between periods.

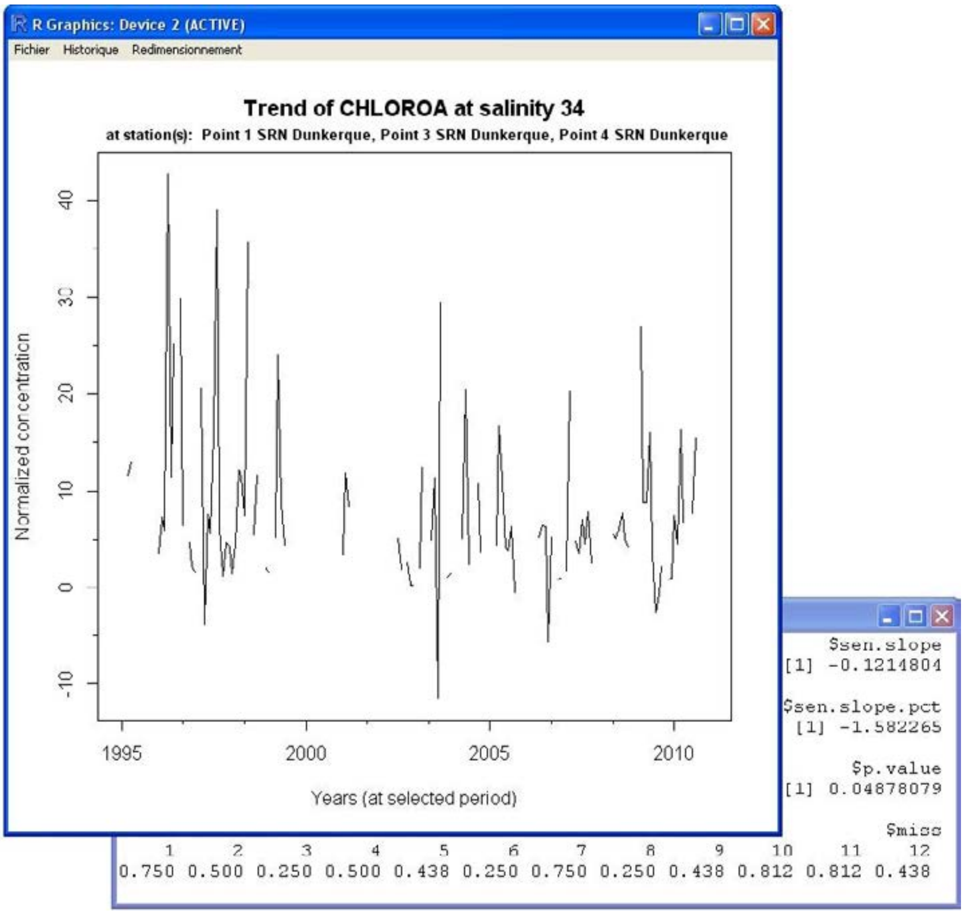
- **Trend based on LOESS**: A loess smoothed curve of the regularised time series is considered to perform a Global Trend test instead of the time series itself (Fig. 48). This is the loess() function of the {stats} package.


R Graphics: Device 2 (ACTIVE)
Fichier Historique Redimensionnement
Trend of CHLOROA based on LOESS
Categorical factor(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4 SRN Dunkerque
Time step: Mensual Method of aggregation: Mean
Time Series
LOESS
CHLOROA concentration
0
5
10
15
20
25
1995
2000
2005
2010
Years

```
       $sen.slope
     [1] -0.0760972

   $sen.slope.pct
      [1] -1.464454

         $p.value
[1] 2.400186e-05

            $miss
 1  2  3  4  5  6  7  8  9 10 11 12
 0  0  0  0  0  0  0  0  0  0  0  0
```

Figure 48. Plot of the regularised time series of chlorophyll-*a* concentration (µg/l) (black line) and loess smoothing (red line). Background table displays result of global trend test perform on loess smoothed data, results are also shown in the right part of the interface.

- **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 performed 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.


R Graphics: Device 2 (ACTIVE)
Fichier
Historique
Redimensionnement
Trend of CHLOROA at salinity 34
at station(s): Point 1 SRN Dunkerque, Point 3 SRN Dunkerque, Point 4 SRN Dunkerque
Normalized concentration
-10
0
10
20
30
40
1995
2000
2005
2010
Years (at selected period)

```
$sen.slope
[1] -0.1214804

$sen.slope.pct
[1] -1.582265

$p.value
[1] 0.04878079

$miss
    1     2     3     4     5     6     7     8     9    10    11    12
0.750 0.500 0.250 0.500 0.438 0.250 0.750 0.250 0.438 0.812 0.812 0.438
```

Figure 49. Plot of normalized concentration of chlorophyll-*a* at salinity 34 between 1995 and 2011 near Gravelines (not enough data in winter for winter normalization). Table displays results of global trend test on these data, results are also shown in the right size of the interface.

All mixing diagram (example from another database in 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 txt table containing all normalized concentration of nutrient per months/years is also generated and saved.

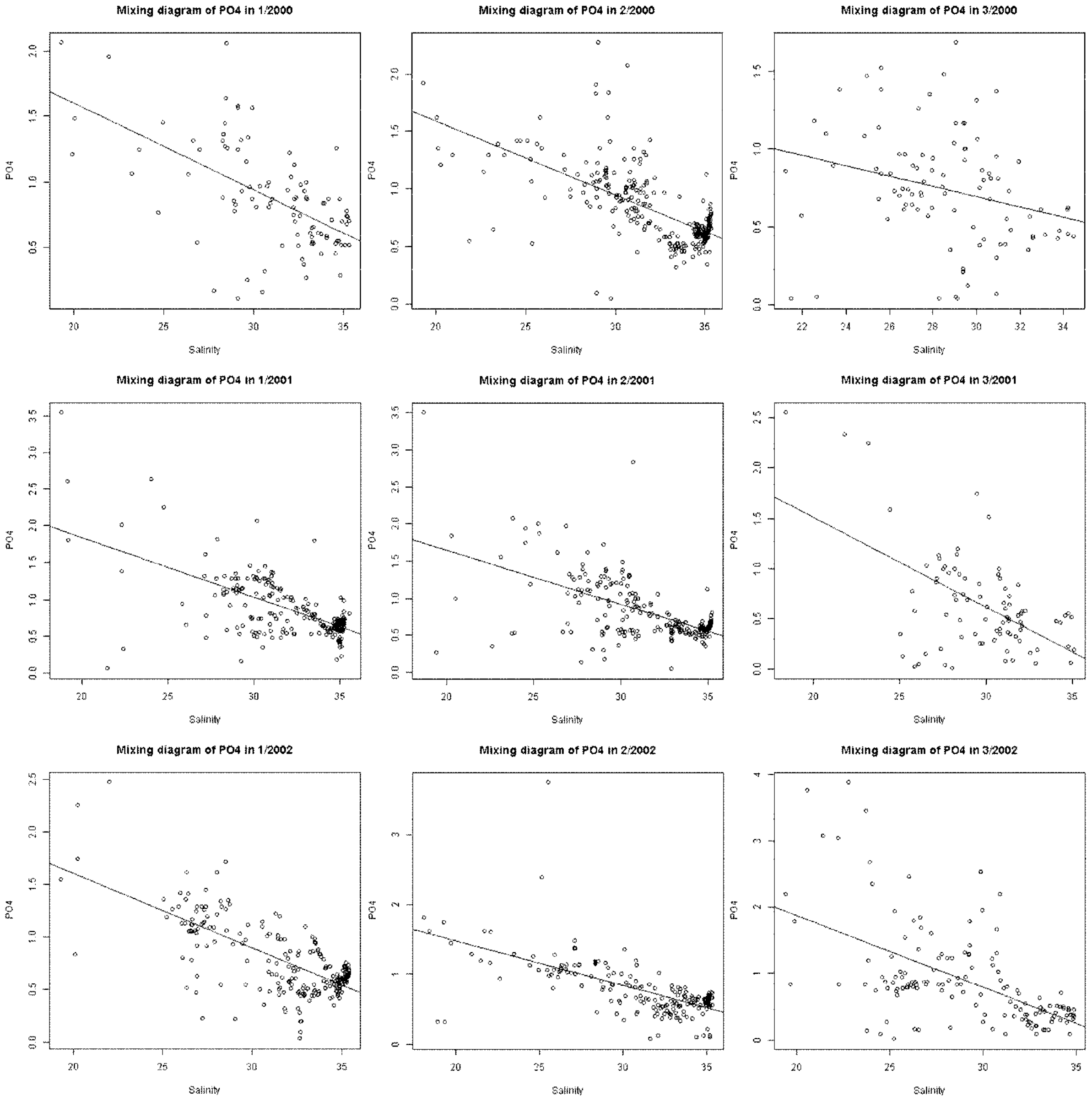


Figure 50. Plots of PO4 (Orthophosphate) concentration against salinity by month/year with linear regressions. Data from another database (just for the example).

6. Results and messages

The right part of the interface 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 txt file like the option you chose to build the time series. Here significant result of Kendall tests are display in bold.

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 a spreadsheet.

7. Extra advices

Importing a new txt file, changing the save directory or editing your data with <Fix data> will reset all your options to default. Changing categorical factors 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 categorical factors in panel 2 and passed directly at the panel 4, balanced options in panel 3 are reselected 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.