INSTITUTO ESPAÑOL DE OCEANOGRAFIA

CEC/DG XII
MARINE SCIENCE AND TECHNOLOGY
MAST3

QCDAMAR

Quality Control of Oceanographic Data

María Jesús García
Andrei Nikouline

Users Guide

May 2005

MEDAR/MEDATLAS II

Mediterranean Data Archaeology and Rescue of Temperature, Salinity and Bio-chemical Parameters

Table of contents

1. Abstract

Software developped at IEO consists of three main and closely related three components dealing with data analysis, database storing and web access.

This manual describes processing methods, programs, and procedures used for **archiving and quality control** of oceanographic data in the form of Medatlas text format, i.e. it covers only first of these three parts focused on the data quality control.

2. Introduction

Since 1994 the Marine Department of the Spanish Oceanographic Institute develops system for archiving and quality control of oceanographic data. The work started in the frame of the European Marine Science & Technology Programme (MAST) when a consortium of several Mediterranean Data Centres began to work on the MEDATLAS project.

Here is a brief history of the project development:

In the 60's, the National Data Centres were created with the purpose to save and archive large data sets coming from different sources.
In 1994, in the frame of the European Marine Science & Technology Programme (MAST) a consortium of several Mediterranean Data Centres began to work on the MEDATLAS Project with the aim to create a hydrographic data bank for the Mediterranean Sea. A common protocols for exchanging data and for quality control was established according to the ICES/IOC recommendations. In the IEO the QCMED software was developed following the MEDATLAS protocols.
In December 1998, the MEDAR/MEDATLAS II concerned action started, with the aims of co-ordinate the tasks of Rescue and Safeguarding of national data sets of temperature, salinity and bio-chemical parameters, carry out by the National Oceanographic Data Centres (NODC) or the Designated National Agencies (DNA) of the IOC/IODE network around the Mediterranean Sea.

In order to enlarge the parameter and quality control requirements, a new version of the QCMED has been developed, named QCDAMAR, according to the MEDAR/MEDATLAS II protocols. In this version the use of new tools for the input data and visualization modules has been adopted for more friendly use. All old software modules for MS DOS were rewritten, improved and migrated to Windows environment. This version of QCDAMAR has been distributed between international oceanographic community as part of 4-CD's product "MEDATLAS 2002. Mediterranean and Black Sea database of temperature, salinity and bio-chemical parameters climatological atlas", also prepared in the frame of MAST.

Since 2003 the project was being financed directly by IEO. The software for automatic and visual oceanographic data quality control suffered many improvements and works nowadays not only with vertical profiles (mainly CTD and bottles observations) but also with time series of currents and sea level observations. New powerful routines for analysis and for graphic visualization were added. Data presented originally in ASCII format were organized recently in an open source MySQL database.

As the software is not necessarilly reduced now to its usage only in the Mediterranean region, it was renamed from QCMEDAR to QCDAMAR, following the abbreviation of “datos marinos” (marine data) in Spanish .

Both the Medatlas format and the methods used in the procedures of quality control are briefly described.

3. Format and Procedures

A complete description of main procedures used in the present software can be found in the following references: “MEDATLAS GROUP 1994”, “MEDAR-MEDATLAS PROTOCOLS 1999”, “UNESCO 1993” y “REINIGER R.F. and ROSS C.K., 1968”. This section contains only a brief summary of the Medatlas format and the quality controls as proceeded in the QCDAMAR software.

3.1 Medatlas Format

The data files are organized by cruises. The structure of the cruise files contains the cruise header and the profiles. Each profile is composed by a profile header, the data parameters at the observed levels and a final-default level. This is a typical structure of the format:

```
cruise header
    profile header
        data parameters
        data parameters
        .........................
         final-default level
     profile header
        data parameters
        data parameters
       .........................
        final-defaul level
       .........................
    profile header
       data parameters
       data parameters
      .........................
       final-default level
```

The cruise header information is based on the Cruise Summary Report. The information contained in cruise and profile headers occupy various lines of the text file. The file can be composed by many cruises with correspondent profles put together.

3.2 Quality Control Procedures

Here is the list of quality control procedures implemented in the programs.

⇒ <u>check for duplicates</u>

- **duplicate cruises**
- **duplicate profiles**
- **manual validation** : (keep only one)

⇒ <u>check stations date and the position</u>

- **check the date**
- **check the ship velocity:** (< 15 knots)
- **check the sounding:** (reference ETOPO5, interpolated from the 9 nearest points)
- **check the location:** (visualization)
- **check global header profile:** (time position and depth)
- **manual validation:** (Edit file and change flags)

⇒ <u>check data points</u>

- **check for acceptable data:** PRES must be present
- **check for impossible regional values**: value must be between the range given in the files "v_total.dat and v_region.dat for each parameter and each zone.
- **check for increasing depth:**
- **check for pressure:** reference profile sounding or ETOPO5 bathymetry.
  values<sounding 5% or between 0.5 or 2 when using reference statistics.
- **check for temperature and salinity** : reference profiles MEDATLAS climatology. (*)
  value must be between the data reference ± a range.
  data reference = values interpolated from the nearest reference profile at the measured pressure level.
  acceptable range of variation = 5, 4 or 3 times of the interpolated standard deviation of reference at the same level.
  (*) if you want to use another climatology (i.e. , Brasseur P) or bathymetry the format of the split files must be perserved.
- **check for constant profiles:** check constant values for each parameter.
  global parameter quality flag = 4.
- **check each pressure:** comparing with the bottom or the ETOPO5 reference
- **check for spikes:** the algorithms used to detect the spike are:
  for the top and bottom pressures: (v2-v1)/(p2-p1)>pic_borne
  for the other pressures depends on the type of data: bottles or CTD.
- **check for density inversion**: the algorithm used is (dens(n)-dens(n-1) ≤ EPS (*)
  where dens(n) is the potential density anomaly at level n computed from equations of the state
  of sea water. (ref: FOFONOFF and MILLARD,1983 and MILLERO and POISSON, 1981).
  (*) The noise value EPS is taken from the file "v_total.dat".
- **global quality check for a profile:** one quality flags per parameter in the header profile
  each flag value is computed from the corresponding parameter flags at each level.
- **check parameter distributions and flags**: visualization in colour for the flags.
- **manual validation**: changing the flags and writing the DM history by editing the file.

Time series header control:

Additionally to the quality control procedures for headers mentioned above, in case of time series some new checks were introduced:

- **Control of start date:** start date and time (year, month, hour, minute and seconds) in the header has to be the same or earlier than the date and time of the first (start) data record.
- **Control of end date:** end date and time in the header has to be the same or later than the date and time of the last (end) data record. In case of lacking of the end date in the header, last date of end record is written instead and the correspondent quality flag is changed to 5.
- **Control of end position:** if the distance between start and end positions > 1 nautical mile the flag is changed to 4; in case of no value of final position, it's changed to the start position value and its correspondent flag takes value 5.

For cases of movement, i.e. when latitude and logitude are present as parameters of a time serie, end position in the header is compared to the end time serie record position.

- **Control of "SENSOR DEPTH":** if its value is greater than DEPTH the flag = 4.
- **Control of "DISTANCE TO THE BOTTOM":** if the sum of "SENSOR DEPTH" + "DISTANCE TO THE BOTTOM" differs greatly from"DEPTH" the flag = 4.
- **Control of duration:** if the duration in the header is less than that calculated from data records, the flag = 4. If the duration value in the header is absent, real calculated duartion is written down instead and the flag is changed to 5. Deafault value = -9.

Time series data control:

These checks are mentioned and described in "5.2 Quality Control" of this document.

- **Date and time control:** all time serie records must be consecutive in time, i.e. each next record must have date and time equal to that of previous record plus time interval of the serie data. On the contrary the flag for all parameters of erroneous record is put to 4 and further analysis is stopped.
- **Pressure/Depth parameter control:** if absolute value of the difference between pressure/depth in each record and "SENSOR DEPTH" is greater than a value established by the user, the flag is changed to 3.
- **Broad range:** if parameter values don't enter in the established diapason, correspondent flag is = 4.
- **Delta check:** if absolute value of the difference between consecutive data records of any parameter is greater than limit established for that difference, the flag = 3.
- **Position parameters check:** if velocity calculated between consecutive records when latitude and longitude parameters are present (case of movement) is greater than established maximum ship velocity (20 knots), the flag is set to 3.

4.1 Software QCDAMAR. Installation & Structure.

4.1 System Requirements

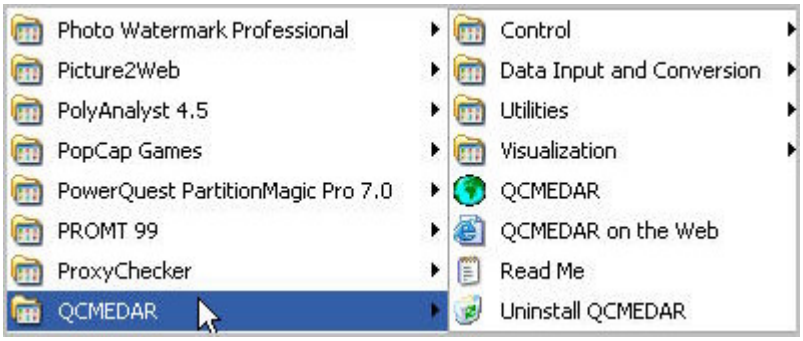
The software has been developed for PC compatible computers. Input data procedures, quality control and utilities modules are written in Visual Fortran, visualization modules – in Visual C++.

All Fortran modules work properly under Microsoft Windows 95/98/2000/XP and later; visualization modules in C++ are designed to work under Windows 98 and later.

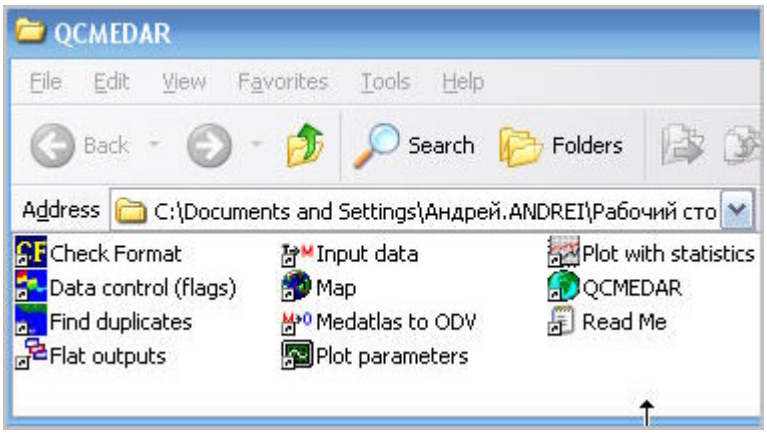
4.2 Installation

QCDAMAR can be downloaded freely from http://indamar.ieo.es/productos/qcdamar_setup.exe.

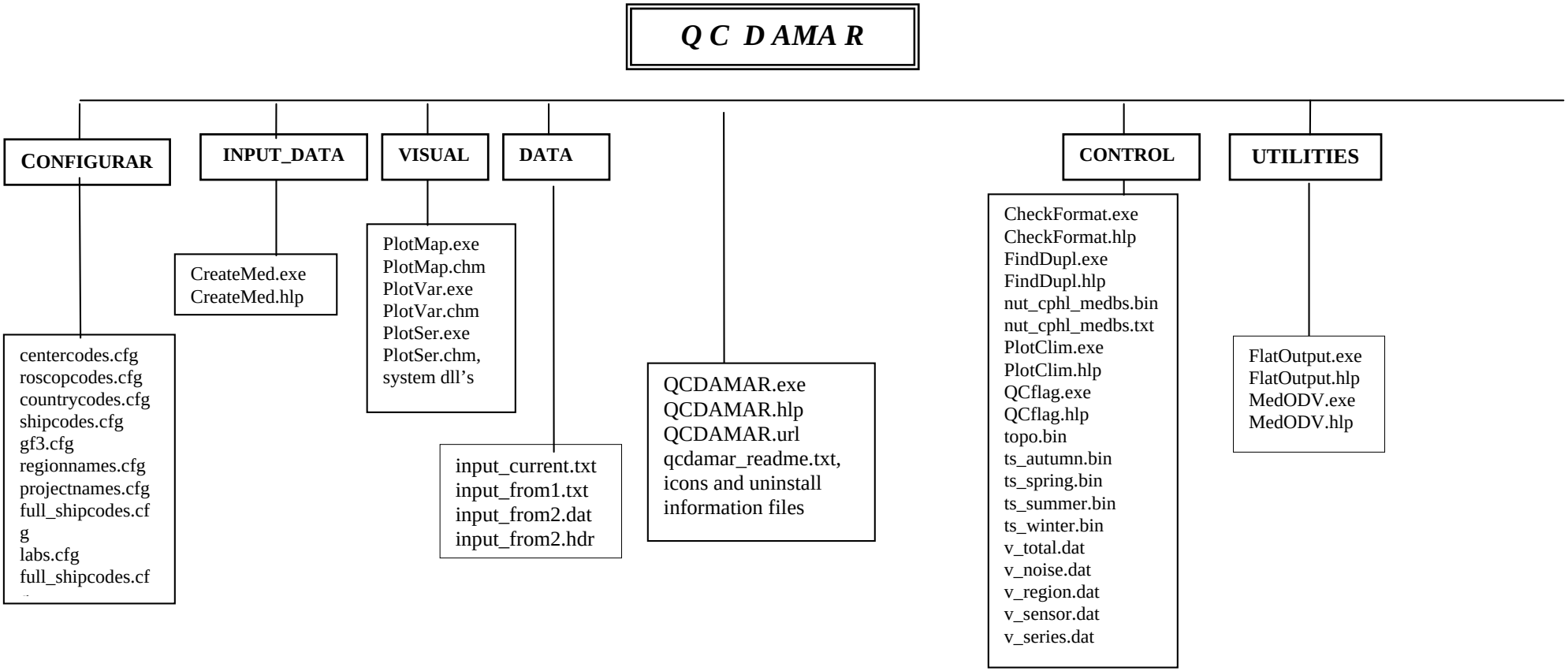
Installation program (file "qcdamar_setup.exe") puts all software components into C:\QCDAMAR folder. A group of icons is created in Start/Programs menu:



Programs can be also launched from desktop double-clicking correspondent icons in created QCMEDAR folder:



4.3 Folders Structure



4.4 Program Compilation.

QCDAMAR is distributed freely for scientific community, but if you want to modify some modules you might need to compile the routines and/or main programs using commercial compilers. These are the following that were used:

- *for Fortran modules:*

CVF (Compaq Visual Fortran) for Windows 95/98 and Windows NT/2000/XP systems is a complete development system that includes Microsoft's Visual Studio 6.0 development environment (same as MS Visual C++, VB,…) with integrated editor, build system, debugger and many other features.

Starting from 2004 HP finished the development of CVF and recommends moving to another Intel® Visual Fortran (IVF) for Windows. IVF product specifications are fully compatible with CVF and it can be used to compile QCDAMAR Fortran modules even without any changes in the text sources. This compiler integrates with the Microsoft Visual Studio .NET environment for easy application development (however separate purchase and installation of Microsoft Visual C++.NET or Microsoft Visual Studio.NET is required for development on IA-32 platforms).

- *for C++ modules*:

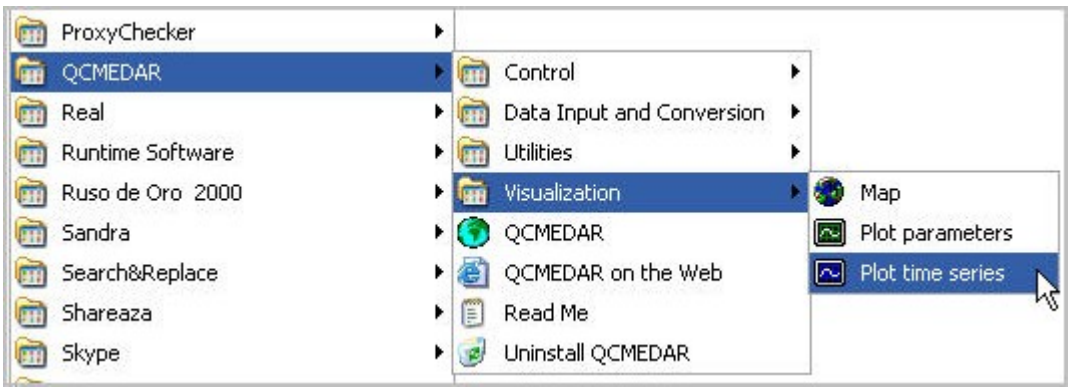
Microsoft Visual Studio was used for compilation.

Both Fortran and C++ modules have been compiled as Win32 applications creating standalone EXE files.

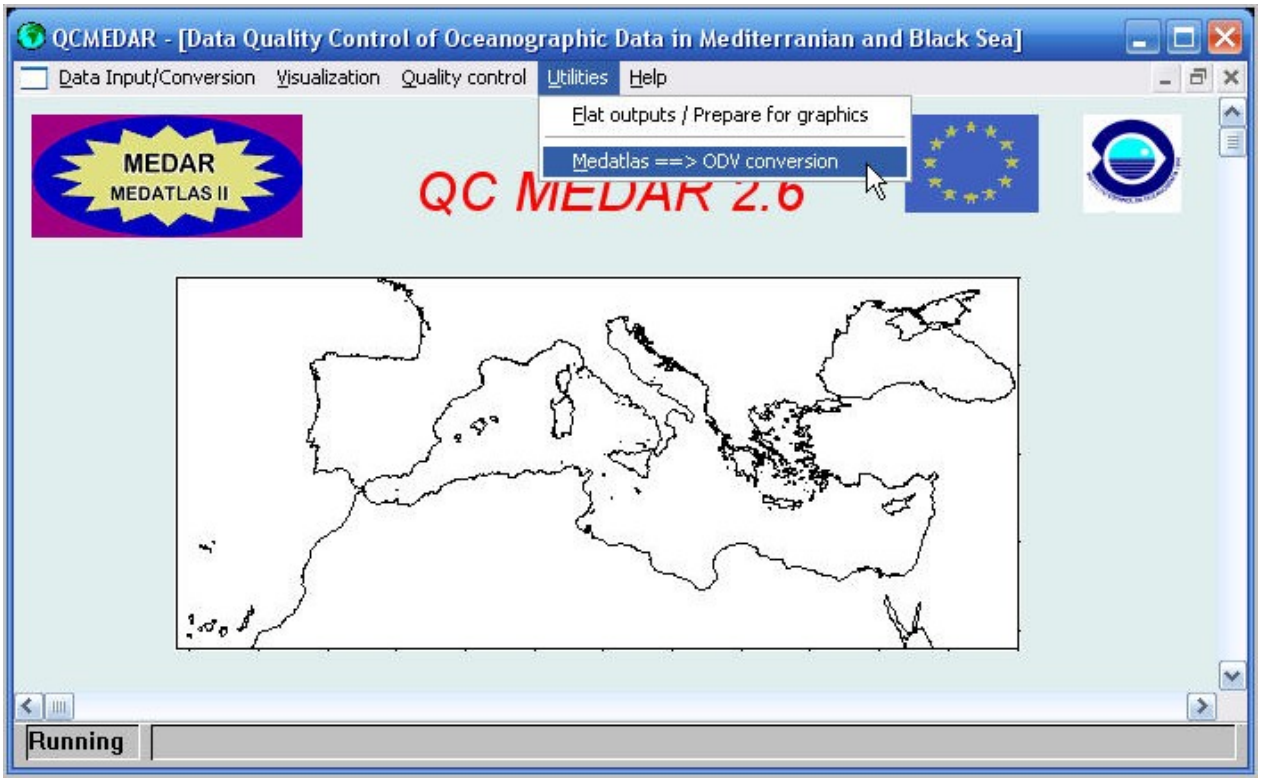
5. QCDAMAR Software. Operation & Sequence

Executable modules (programs) entering in QCDAMAR can be divided into four groups. Their purpose is different and actually they are located in different subfolders (see Figure "Folders Structure").

They can be run separately (recommended) from desktop icons or Start/Programs menu:



or called from the menu of the following envelope form that appears on the screen after running QCMEDAR.exe application:

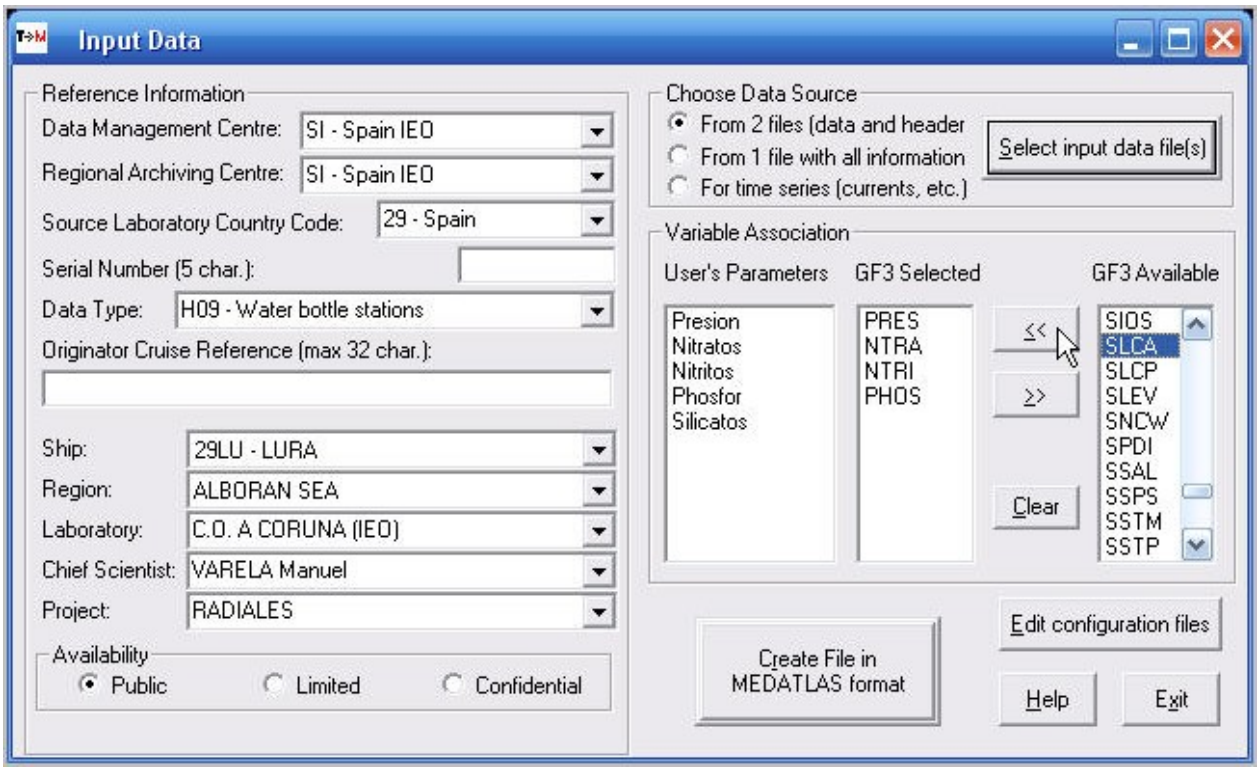


Here is the description of theses four groups of programs:

5.1 Input Data (folder "input_data").

In earlier versions of the software we could find here different filters to transcode data from different data sources to Medatlas format. Because of numerous formats of oceanografic data existent in the world, we left in this folder only one application "**CreateMed.exe**" that makes easier the process of conversion.

The program has the following interface:



Metadata (header information) are introduced via dialog box. It is possible now to create quickly cruise files by importing data automatically from one or two previously prepared ASCII spreadsheets that have very simple format. Only needed parameters can be chosen interactively and their free form names can be associated with existent GF3 codes.

<u>Create Medatlas file from one input text file:</u>

This is an example of the format when both header and data are found in one input text file.
First line contains description of the profile header information. It must start with the word 'Station' in order the program could detect the end of the previous and the beginning of next profile.
Second line has corresponding values for the profile header. The sign of degrees must be negative for south latitude and west longitude. Zeros are automatically added to the left part of string, so '1' for Time (hours) will be correctly interpreted as '01'.
Third line has the words describing the parameters found in the profile. Later, when these "user's parameters" will appear in the list box of the program dialog, you will be able to associate them with available GF3 codes. It's allowed to put one word for one parameter.
Next lines have numeric values of data divided by commas or tabulations.

Sample of such input file is shown below:

```
Station,Lat(gr),Lat(min),Lon(gr),Lon(min),Day,Month,Year(full),Time(hours),Time(min),Depth(m)
1111,60,11.1,24,13.5,1,4,2001,1,40,200
pressure1 nitratos1
0,0.14
5,-9
15,0.1734
25,0.08
50,0.08
75,1.48
100,2.61
Station,Lat(gr),Lat(min),Lon(gr),Lon(min),Day,Month,Year(full),Time(hours),Time(min),Depth(m)
2222,60,11.1,24,13.5,2,5,1999,2,40,200
pressure2 nitratos2
0,0.14
50,-9
150,0.1734
250,0.08
500,0.08
750,1.48
1000,2.61
```

Create Medatlas file from two separate text files:

Choose this option if your input data, values and header, are found in two separate text files. The files have same name, while the extensions of data file may be free (most commonly used are .DAT, .TXT, PRN), the header file must have .HDR extension.
In the standard Open File dialog you need to select only one file with data as the corresponding header file will be found automatically.
This is an example of the format when the data and header information are found in two separate files.
Data file:

```
Station Presion Nitratos Nitritos Phosfor Silicatos
0022,0,0.14,0.12,0.07,0.94
0022,5,-9,-9,-9,-9
0022,15,0.17,0.02,0.05,1.05
0022,25,0.08,0.01,0.04,1.03
0022,50,0.08,0.09,0.08,1.21
0022,75,1.48,0.19,0.06,1.66
0022,100,2.61,0.09,0.08,2.10
0333,0,0.08,0.303,0.07,0.29
0333,75,-9,-9,-9,-9
1001-2,0,0.3235,3.,0.9, 1234.126
1001-2,5,  0.13,7,0.6, 0.6
1001-2,15, 0.13,1,0.1, 0.8
```

First line begins with the word 'Station' followed by the words describing parameters found in the profile. Later these expressed in free form "user's parameters" will appear in the list box of the program dialog and you will be able to associate them with available GF3 codes. It's allowed to put only one word for each parameter.
Next lines have numeric values of data separated by commas, spaces or ";". Absent values for parameters must be equal to '-9'. Each line corresponds to one level and must start with station number of length 4. The cast is equal to 0 by default. In case of several casts for the same station, they are introduced preceded by hyphen after the station, p.e. '1001-2' in the example below means second cast of station 1001.
Data (and header) files are usually produced by export routines of widespread software, like database managers, spreadsheets, graphic packages, etc.

Some of data values in the example above are not real. They were changed especially to demonstrate the program format capabilities shown in the output Medatlas file.

Header file:

```
Station,Lat(gr),Lat(min),Lon(gr),Lon(min),Day,Month,Year(full),Time(hours),Time(min),Depth(m)
0333,-40,30.2,-2,10.0,1,12,1998,14,50,300,
0022,60,11.1,24,13.5,1,3,2000,12,40,200,
1001-2,39,28.59,2,25.63,10,1,1997,1,3,100,
```

First line contains description of the profile header information. It starts with the word 'Station' and is skipped by the program.
Every next line has corresponding values for the profile header. The sign of degrees must be negative for south latitude and west longitude. Usually zeros are added to the left part of string automatically, so '1' for Time (hours) will be correctly interpreted as '01'. Station values notation is the same as in data file. The consecutive order is not important, because the program itself searches and associates headers and data by their station values.

Example of the output file in Medatlas format:

```
*SI29199700242 TEST                                29DB ODON DE BUEN
10/01/1997 01/03/2000 ALBORAN SEA
29 INSTITUTO ESPANOL DE OCEANOGRAFIA
CANO LUCAYA Natalio                        Project=ECOMALAGA-95
Regional Archiving= SI                     Availability=P
Data Type=H24 n=   3 QC=N
Data Type=H25 n=   3 QC=N
Data Type=H22 n=   3 QC=N
Data Type=H26 n=   3 QC=N
COMMENT
*SI2919970024210012 Data Type=H09
*DATE=10011997 TIME=0103 LAT=N39 28.59 LON=E002 25.63 DEPTH=   100 QC=0000
*NB PARAMETERS= 5 RECORD LINES=    3
*PRES SEA PRESSURE, sea surface=0   decibar=10000 pascals          def.= -999.9
*NTRA NITRATE (NO3-N) CONTENT       millimole/m3                   def.= 9.9999
*NTRI NITRITE (NO2-N) CONTENT       millimole/m3                   def.=     9.
*PHOS PHOSPHATE (PO4-P) CONTENT     millimole/m3                   def.=    9.9
*SLCA SILICATE (SiO4-Si) CONTENT    millimole/m3                   def.=9999.99
*GLOBAL PROFILE QUALITY FLAG=0 GLOBAL PARAMETERS QC FLAGS=00000
*DC HISTORY=
*
*DM HISTORY=
*
*COMMENT=
*
*SURFACE SAMPLES=
*
*PRES    NTRA NTRI PHOS    SLCA
   0.0 0.3235   3.  0.9 1234.13 00000
   5.0 0.1300   7.  0.6    0.60 00000
  15.0 0.1300   1.  0.1    0.80 00000
-999.9 9.9999   9.  9.9 9999.99 99999
```

The piece of the file shown above corresponds to the input information provided in the screenshots of two source files, data and header, used in this help file. The name of the output file, SI29199700242.sta is composed from Medatlas reference and the station and it has .STA extension.

Cruise start and end dates are calculated automatically, as well as data types and their occurencies.
The appropriate format for default values is chosen taking into account the number of digits before and after the comma at all levels for each parameter. Columns headers are aligned with data to make easier their 'readability'.

Create Medatlas file from one input text file that represents time series.

This is an example of the format of the input text file containing the information from current meter.
First line contains description of the profile header information. It must start with the word 'Station' in order the program could detect the end of the previous and the beginning of next profile.
Second line has corresponding values for the profile header, i.e. station number, coordinates and time of the beginning of observations. The sign of degrees must be negative for south latitude and west longitude. Zeros are automatically added to the left part of string, so '1' for Time (hours) will be correctly interpreted as '01', but the Year must be entered in the form of four digits, p.e. '2004'.
Third line describes the order of input of coordinates and time corresponding to the end of observations, which are introduced in the following fourth line in the same form as the information about the beginning of observations.
Fifth line is omitted by the program and describes only the order of input in the sixth line of sensor depth, its distance to bottom, duration of observations in days, sampling rate in seconds and magnetic declination in degrees.
Finally the seventh line has the words describing the parameters found in the profile. First four parameters (year, month number, day within month and time within day in the format HHMMSS) are always present for time series and must be entered in this order. Later on, after reading the input text file, when the "user's parameters" appear in the list box of the program dialog, you will be able to associate them with available GF3 codes. It's allowed to put one word for one parameter.
Next lines contain numeric values of data divided by commas or tabulations.

Sample of input file for time series (currents, sea level, etc.):

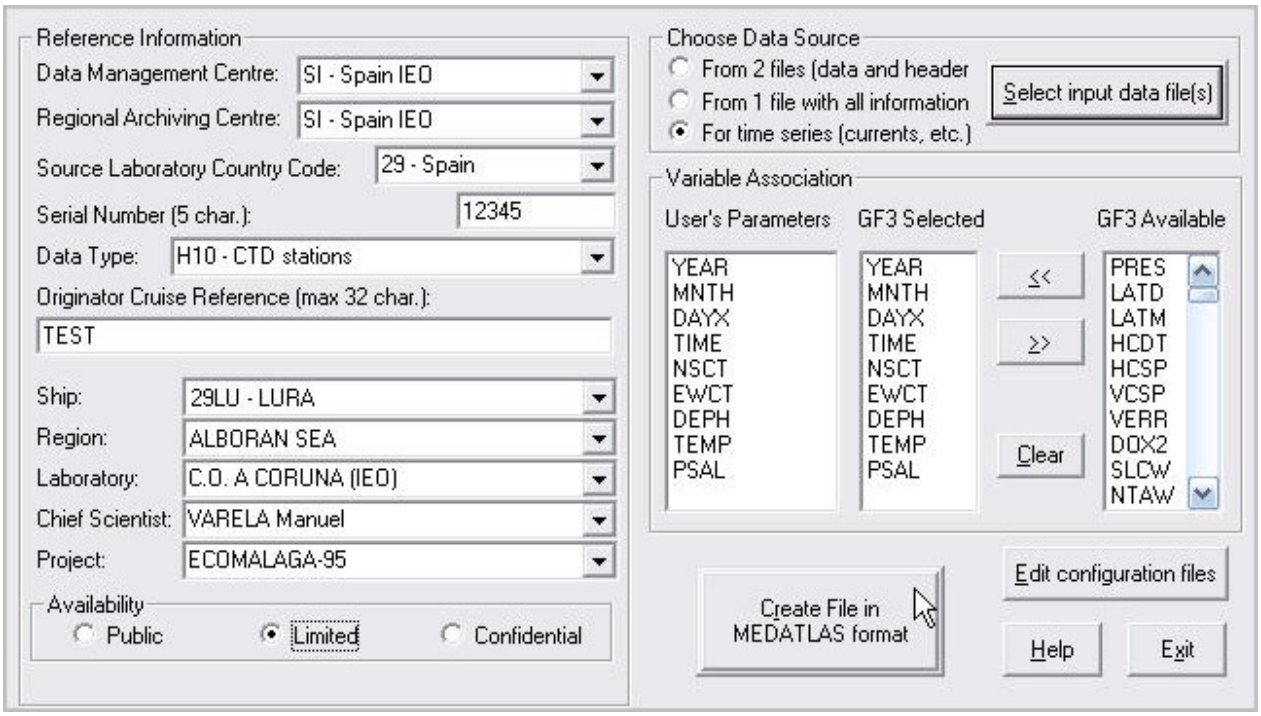
```
Station,Lat(degr),Lat(min),Lon(degr),Lon(min),Day,Month,Year(full),Time(hours),Time(min),Depth(m)
1111,60,11.1,24,13.3,1,4,2001,1,40,200
Elat(gr),Elat(min),Elon(gr),Elon(min),Eday,Emonth,Eyear(full),Etime(hours),Etime(min),Edepth(m)
60,21.1,34,19.6,3,5,2001,21,55,300
Sensor_depth(m),Distance_to_bottom(m),Duration(days),Sampling_Rate(sec),magnetic_declination(degr)
33,44,30,60,22.2
YEAR MNTH DAYX TIME   NSCT    EUCT    DEPH   TEMP  PSAL
 1997   05   11 181212  +0.174  -0.162  112.0 14.50 37.53
 1997   05   11 191313  +0.056  -0.006  106.0 14.47 37.55
 1997   05   11 201414  +0.051  +0.046  104.0 14.42 37.60
 1997   05   11 211515  +0.034  +0.126  102.0 14.45 37.56
 1997   05   11 221516  +0.024  +0.113  100.0 14.67 37.44
```

The file shown below corresponds to the input information provided in the screenshot above after clicking the input for time series.

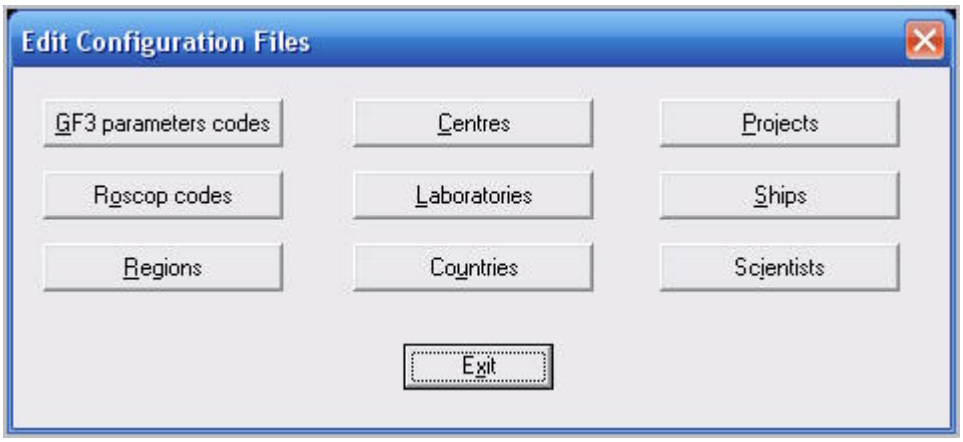
```
01/04/2001 01/04/2001 NE ATLANTIC OCEAN
29 C.O. A CORUNA (IEO)
VARELA Manuel                                Project=RADIALES
Regional Archiving= SI                       Availability=P
Data Type=D01 n=   1 QC=N
Data Type=H09 n=   1 QC=N
COMMENT
DM=
DM=
*SI2920010024211110 Data Type=H09
*DATE=01042001 TIME=0140 LAT=N60 11.10 LON=E024 13.50 DEPTH=   200 QC=0000
*NB PARAMETERS= 9 RECORD LINES=    5
*YEAR YEAR                          (yyyy)                          def.=   9999
*MNTH MONTH                         (mm)                            def.=     99
*DAYX DAY WITHIN MONTH              (dd)                            def.=     99
*TIME TIME WITHIN DAY               (hhmmss)                        def.= 999999
*NSCT CURRENT NORTH COMPONENT       (meter/second)                  def.= 99.999
*EWCT CURRENT EAST COMPONENT        (meter/second)                  def.= 99.999
*DEPH DEPTH BELOW SEA SURFACE       (meter)                         def.= -999.9
*TEMP SEA TEMPERATURE               (Celsius degree)                def.=  99.99
*PSAL PRACTICAL SALINITY            (P.S.U.)                        def.=  99.99
*GLOBAL PROFILE QUALITY FLAG=0 GLOBAL PARAMETERS QC FLAGS=000000000
*DC HISTORY=
*
*DM HISTORY=
*
*COMMENT=
*ADDITIONAL INFORMATION - TIME SERIES
*EDATE=02052001 ETIME=2155 ELAT=N60 21.10 ELON=E034 13.50 EDEPTH=   300 QC=0000
*SENSOR DEPTH=   33 (metre)            DISTANCE TO BOTTOM=   44 (metre) QC=00
*DURATION=   30 (day)                                                   QC=0
*SAMPLING RATE=   60 (second)          MAGNETIC DECLINATION=22.2 (degree)
*SURFACE SAMPLES=
*
*YEAR MNTH DAYX   TIME   NSCT   EWCT   DEPH  TEMP  PSAL
 1997   05   13 181212  0.174 -0.162  112.0 14.50 37.53 000000000
 1997   05   13 191313  0.056 -0.008  106.0 14.47 37.55 000000000
 1997   05   13 201414  0.051  0.046  104.0 14.42 37.60 000000000
 1997   05   13 211515  0.034  0.126  102.0 14.45 37.58 000000000
 1997   05   13 221616  0.024  0.113  100.0 14.67 37.44 000000000
 9999   99   99 999999 99.999 99.999 -999.9 99.99 99.99 999999999
```

Cruise start and end dates are calculated automatically, as well as data types and their occurencies.
The appropriate format for default values is chosen taking into account the number of digits before and after the comma at all levels for each parameter. Columns headers are aligned with data to make easier their 'readability'.

Before creating output files you enter metadata information in the form:



The content of configuration files is displayed in combo boxes on the left side of the main program dialog.
You can choose needed value, introduce text or edit correspondent .CFG files located in 'c:\qcmedar\configurar' directory. WordPad supplied with Windows is used by the program, but you can always use any text editor of your choice.



When editing, you can move most frequently used items (lines) to the beginning of a .CFG file, because they will appear first (earlier) in the correspondent combo boxes of the main dialogue window. Adding of new lines in these files avoids unnecessary typing when you introduce cruise header information.
Some configuration files are called by other modules of QCMEDAR, so please, be careful when editing and preserve the original format of all .CFG files.
List of GF3 codes of the parameters found in the file 'gf3.cfg' of 'c:\qcmedar\configurar' directory is displayed in the box “Available” after reading input data files.
The box “Selected” displays GF3 codes selected and moved from 'GF3 Available' list box.

Directional buttons << and >> move parameter(s) from right/central list box (GF3 available/selected) to central/right (GF3 selected/available). You can choose parameters by clicking them. Multiple selection is supported.
Chosen parameters in ‘GF3 Selected’ box are associated with the correspondent parameters names in 'User's Parameters' box. The user gives these names in free form (words) in the input data file.

The format of “gf3.cfg” file:

```
#PRES - SEA PRESSURE, sea surface=0          (decibar=10000 pascals)        ***
YEAR - YEAR                                 (yyyy)                         D01
MNTH - MONTH                                (mm)                           D01
DAYX - DAY WITHIN MONTH                     (dd)                           D01
TIME - TIME WITHIN DAY                      (hhmmss)                       D01
NSCT - CURRENT NORTH COMPONENT              (meter/second)                 D01
EWCT - CURRENT EAST COMPONENT               (meter/second)                 D01
DEPH - DEPTH BELOW SEA SURFACE              (meter)                        D01
TEMP - SEA TEMPERATURE                      (Celsius degree)               ***
PSAL - PRACTICAL SALINITY                   (P.S.U.)                       ***
LATD - LATITUDE DEGREES                     (degree)                       D01
LATM - LATITUDE MINUTES                     (minute)                       D01
HCDT - DIRECTION REL. TRUE NORTH            (degree)                       D01
HCSP - HORIZONTAL CURRENT SPEED             (meter/second)                 D01
VCSP - VERTICAL CURRENT SPEED               (millimeter/second)            D01
VERR - VELOCITY ERROR                       (meter/second)                 D01
DOX2 - DISSOLVED OXYGEN                     (micromole/kg)                 H21
SLCW - SILICATE (SiO4-Si) CONTENT           (micromole/kg)                 H26
NTAW - NITRATE (NO3-N) CONTENT              (micromole/kg)                 H24
```

GF3 codes in the first column are displayed in 'GF3 available' list box of the program main dialogue. Information from the second and third columns is written in profiles headers. ROSCOP codes in the fourth column are used in calculating of occurence of different parameters. Put *** when ROSCOP code for the parameter is not yet defined or unknown.
Don't edit and don't move the first line (#PRES - ...). Don't change also *** in case of TEMP, PSAL and SSAL, because ROSCOP code for these parameters (H09 or H10) is chosen from 'Data Type' combo box.
Button “Clear” clears all three list boxes. The left and right list boxes will be filled again after selecting and reading by the program of input data file(s).
Push the button “Create File in MEDATLAS format” only after all parameters are associated and the necessary input information on the left pane of the dialog is provided.
When the exporting routine finishes the message 'Task complete' is displayed. In case of two input files the number of profiles headers is equal to the number of profiles processed when all 'stations' values in the header file were successfully found in the correspondent data file. Output file has .STA extension.
Information introduced in other boxes is also necessary to fill cruise and profile headers.
In the “Data Type” box H09 for bottles and H10 for CTD are used in most of cases.
“Originator cruise name/reference” - p.e. ECOMALA0500.
“Select input data file(s)” button displays standard file dialog to open an existing text file for transcoding in MEDATLAS format.
The “Serial Number” is a part of Medatlas reference of the cruise.
“Ship” and “Region” are not editable combo boxes. However you can add new values via editing correspondent configuration files.

5.2 Data Quality Control (folder “control”).

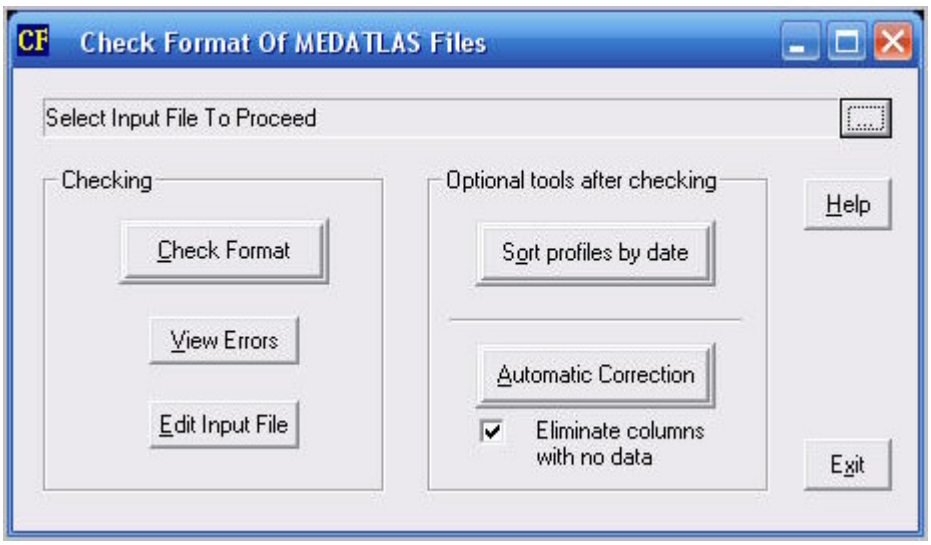
This is the most important group of QCDAMAR modules necessary to perform quality control of oceanographic data presented in Medatlas format.

The objective of the data quality control is to ensure the data set consistency within a single data set and within a collection of data sets. The complete quality control consists of the **automatic checking** where the data are not modified and only the quality flags are changed, and of the **visualizations** that help to the data manager to validate data by modifying some flags (if necessary) and writing manually any incidents in the “DM HISTORY” part of header of Medatlas format.

In this folder we find CheckFormat, FindDupl, QCFlag and PlotClim programs.

- **Check Format** (file “CheckFormat.exe”)

The main purpose of CheckFormat is to help to detect errors in the MEDATLAS files before performing full QC checking. Profiles are checked for valid date, time, coordinates, number and QC status of datatypes. Searching for duplicated profiles identifiers is performed, ship velocity is controlled, distance and time between profiles (within one mile, one hour) is checked.
Input files keep intact after this checking but some new files with different extensions (.err, .srt, .aqc) can be created.



Additional helpful functions are included:

- sorting profiles by date and time written in their headers;

During this action profiles of the input file are written to the hard drive as temporary binary files, then sorted and finally put together and written down in a new text file that resides in the same directory and has the same name as the input one but with the .SRT extension.

After this operation output file becomes input file for analysis that will be reflected in the label with input file name.
It's recommended to perform Check Format before sorting.

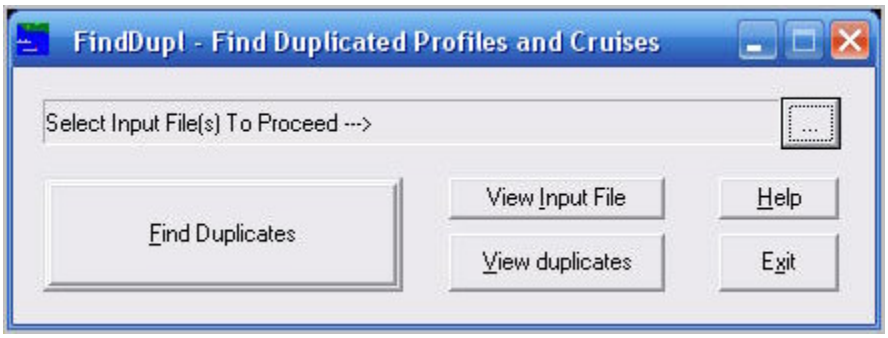
- automatic correction with the option of eliminating parameters in the files that have no data, i.e. when the columns contain only default values.

If cruise header is present, different datatypes, as well as their occurencies and QC status are found. Output formats for data, column titles and default values are composed taking into account the number of digits before and after decimal points.
As an option, you may check the box “Eliminate columns with no data” if you want during the process of automatic correction to eliminate from the profiles columns that have no data, i.e. when the parameter has only default values (p.e. 99.999) at all levels.
New created file has .AQC extension and same name as the input file.

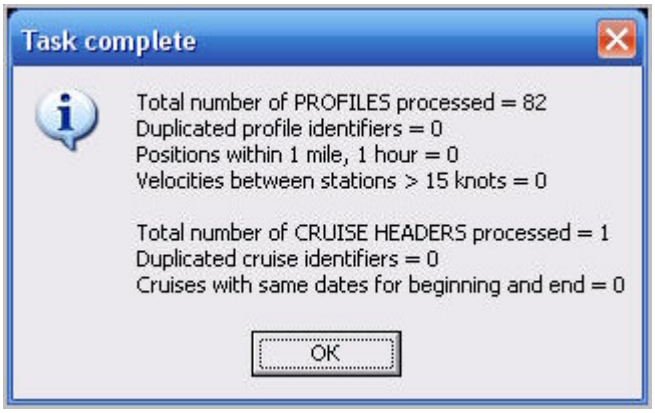
After this operation output file becomes input file for analysis that will be reflected in the label with input file name.
It's recommended to perform Check Format before Automatic Correction.

- **Find Duplicated Cruises and Profiles** (file “FindDupl.exe”)

This program searches for duplicated cruises and profiles identifiers in selected one or many files in MEDATLAS format. Existing file(s) are opened using the standard open file dialog. Multiple selection is possible using CTRL or SHIFT keys and LEFT mouse button.
In case of multiple selection input text files are merged into one temporary file '1tmp.tmp' created in the same directory.



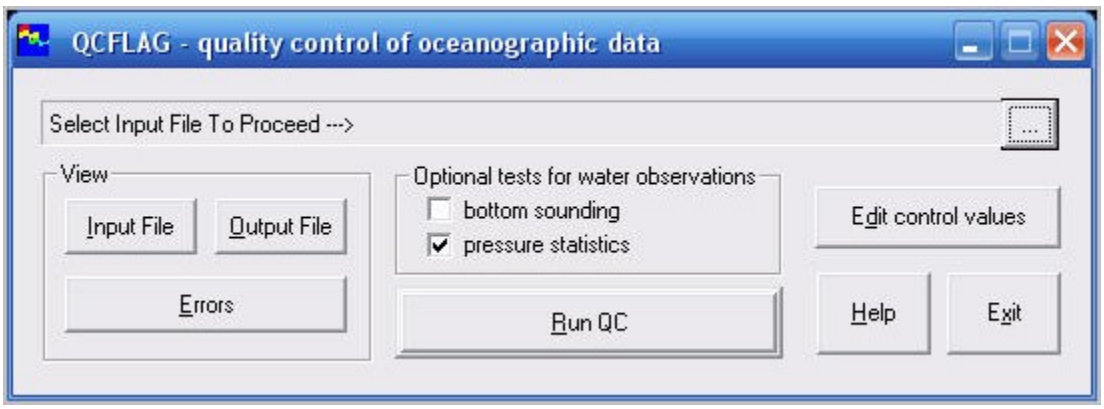
Start and end dates, ship velocities, positions (within one mile, hour) are also controlled.
When finished the message like following will inform you about the results:



Clicking on “View duplicates” WordPad is launched to view detected errors (if any) in the newly created text file that has same name as the input one but with .ERR extension.

- **Quality Control Of Oceanographic Data** (file “QCFlag.exe”)

This program performs full quality control of oceanographic data presented in MEDATLAS format. It's strongly recommended to analyze input file by CheckFormat and FindDupl programs before running QCflag.



<u>The following test are performed automatically when running QCflag:</u>

- Water observations

- check for valid ***date and time***
- check for valid ***coordinates***
- check the bottom ***sounding*** (compare with pre-existing statistics) - optional test.
  During this test the depths from ETOPO5 are used (5'x5' values grouped in squares of 1 x 1 degree with 169 points each). The depths of Mediterranean and Black Sea are stored in "c:\qcmedar\control\topo.bin" binary file. Statistic depth in our point is calculated too.
- search for title ***PRES*** in column headers of a profile
- check for ***impossible regional*** values. The ranges of the parameters and data for spikes control are taken from the text files "v_region.dat" and v_total.dat". You can edit these files very carefully keeping the format unchanged.
- search ***descending order of pressure*** values in a profile
- check ***pressure*** values (compare with sounding and pre-existing statistics)
- check if the parameter is ***constant on the vertical***
- ***spikes*** control
- ***density inversion*** test

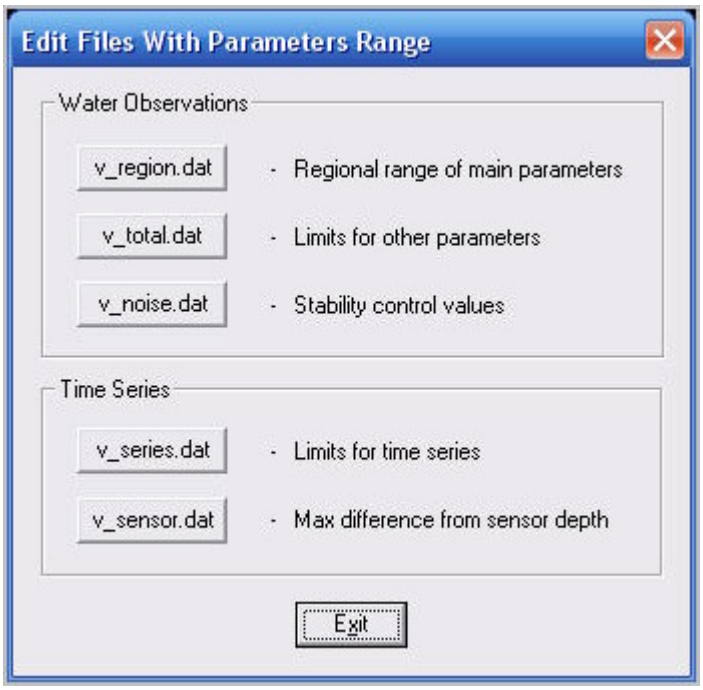
- Time Series

Apart of general headers control for valid date, time, etc., identical to water observations, some additional tests are performed in case of time series:

- ***beginning date*** in the header must be prior or equal to the date of the first measured data value
- ***end date*** must be greater than the beginning date, as well as it can't prior to the last measured data value. If end date is absent, the date of last measured data is assigned to the end date and the flag is changed to "5".
- ***duration in days*** given in the header is compared to its real value calculated from observations
- ***sensor depth*** in the header is compared with measured data pressures (PRES) or depths below sea surface (DEPH). Maximum allowed difference is read from editable text file "v_sensor.dat"
- value of the ***sum*** of sensor depth and distance to bottom minus 10% of this sum can't be greater than depth
- the values of maximum, minimum and delta, necessary for ***impossible regional*** values test of time series, are contained now in the file "v_series.dat"
- checking for ***descending time order*** is performed
- time continuity check is applied to each data measurement. ***Time shift*** between data values is compared with ***sampling rate***

If optional test for “bottom sounding” is checked, test is performed by comparing the bottom sounding value with -+20% of max and min values of ETOPO depths chosen from 9 nearest ETOPO points.
If optional test for “pressure statistics” is checked, test for pressure is performed during full QC. Pressure values at all levels are compared with sounding (if recorded) or with statistical depth value from ETOPO.
Detailed description of testing procedures implemented in QCflag can be found in Medar-Medatlas Project Protocol.

You can edit configuration text files located in 'c:\qcdamar\control' directory.



WordPad supplied with Windows is used by the program, but you can always use any text editor of your choice. Editing sometimes is necessary in order to adjust the ranges depending of the parameters variability in your working region.
Please, be carefiting and preserve the original format of all .DAT files, keeping the line order and positions unchanged.
Input observed data in MEDATLAS format are usually located in the C:\QCMEDAR\DATA directory. There is no restriction for the input file extensions in any option.

The output files with the same names and QC extension are found in the same directory. The files with the list of errors (if there any) have the ERR extension and the same names as the input files. The ETOPO5 data are interpolated and should be used for primary rough control of sounding.

The files "v_region.dat", v_total.dat" and "v_noise.dat" contain limit control values for water observations, while in the files "v_series.dat" and "v_sensor.dat" there are limit check values for the case of time series.
Below you can see some screenshots of these files.

v_total.dat:

GF3	MAX	MIN	BORNE	SPIKE
DOXY	650	0	50	50
DOX1	10	0	0.2	0.3
DOX2	650	0	50	50
NTRA	56	0	0.7	0.7
NTAW	99	0	2	2
NTRI	10	0	0.5	0.5
NTIW	100	0	2	2
NTRZ	100	0	10	10
AMON	10	0	1	1

v_region.dat:

Zone	Lat-Max	Lat-Min	Lon-Max	Lon-Min	Fondo	Cape	T1-max	T1-min	T2-max	T2-min	T-Spike	T-Spike	PSAL-Max	PSAL-Min
ANT	-50.00	-85.00	-20.00	-70.00	5000	200	15.0	-2.0	15.0	-2.0	0.5	1	38.0	33.0
DAT	50.00	40.00	0.00	-6.00	3000	200	25.0	5.0	15.0	0.0	0.5	1	39.0	34.0
DA1	50.00	40.00	8.00	30.00	7000	200	28.0	8.0	25.0	0.0	0.5	1	39.0	34.0
DF1	45.00	40.00	-5.00	-30.00	5000	200	30.0	22.0	25.0	0.0	0.5	1	39.0	34.5
DF2	42.00	40.00	9.00	4.50	3000	200	30.0	22.0	26.0	12.0	0.5	1	39.0	35.0
DF3	43.60	42.00	6.00	2.50	2750	200	26.0	20.0	25.0	12.0	0.5	1	39.0	34.0
DF4	44.50	42.00	9.40	6.10	2964	200	29.0	20.0	25.0	12.0	0.5	1	39.0	33.0
DF5	44.20	42.80	10.80	9.80	1650	200	27.0	22.0	26.0	12.0	0.5	1	39.0	33.0
DF6	41.70	35.30	37.70	33.70	4500	200	30.0	23.0	25.0	15.0	0.5	1	40.0	34.0

The hydrological regime of a particular area of study can be characterized by a great variability. To enhance the performance of quality control of imposible regional values for this area the user can introduce in the file ‘v_region.dat’ individual limits for parameters for that area. In the figure above each line corresponds to one specific known oceanographic region or any region defined by the user. P.e. the first line in our screenshot above (ANT) corresponds to Antarctics region, defined by the user, while the sixth line (DF2) correponds to the known region “Gulf Of Lions” in the Mediterranian Sea.

v_noise.dat:

```
200          - two layers limit in dbars/meters
.05          - noise level for upper layer
.03          - noise layer for lower layer
```

v_series.dat:

```
GFS      MAX      MIN   DELTA
-----------------------------------------------------------------------------------------------
YEAR    2020     1800       1      - YEAR                                  (yyyy|
MNTH      12        1       1      - MONTH                                 (mm)
DATX      31        1       1      - DAY WITHIN MONTH                      (dd)
TIME  246030   000000  016060      - TIME WITHIN DAY                       (hhmmss)
NSCT       5        0       3      - CURRENT NORTH COMPONENT               (meter/second|
EWCT       5        0       3      - CURRENT EAST COMPONENT                (meter/second|
HCDT     360        0     360      - DIRECTION REL. TRUE NORTH             (degree)
HCSP       5        0       3      - HORIZONTAL CURRENT SPEED              (meter/second|
VCSP     100        0      10      - VERTICAL CURRENT SPEED                (millimeter/second|
VERR       1        0     0.1      - VELOCITY ERROR                        (meter/second|
TEMP      30      10       5      - SEA TEMPERATURE                       (Celsius degree)
PSAL      40        0       2      - PRACTICAL SALINITY                    (P.S.U.)
PRES   11000        0     100      - SEA PRESSURE, sea surface=0           (decibar=10000 pascals|
```

v_sensor.dat:

```
10          -   Max difference of PRES or DEPH parameter from sensor depth in the header

Please be careful when editing this value, putting it first in the first line.
```

In case of a big amount of data to analyze the process can become lengthy and the progress dialogue can help to watch the program execution:

The operation is finished when the following dialogue appears on the screen:

When you run data tests the input file is kept unchanged, the resulting file with new flag values has .CLI extension, detected errors are always written into .ERR file.

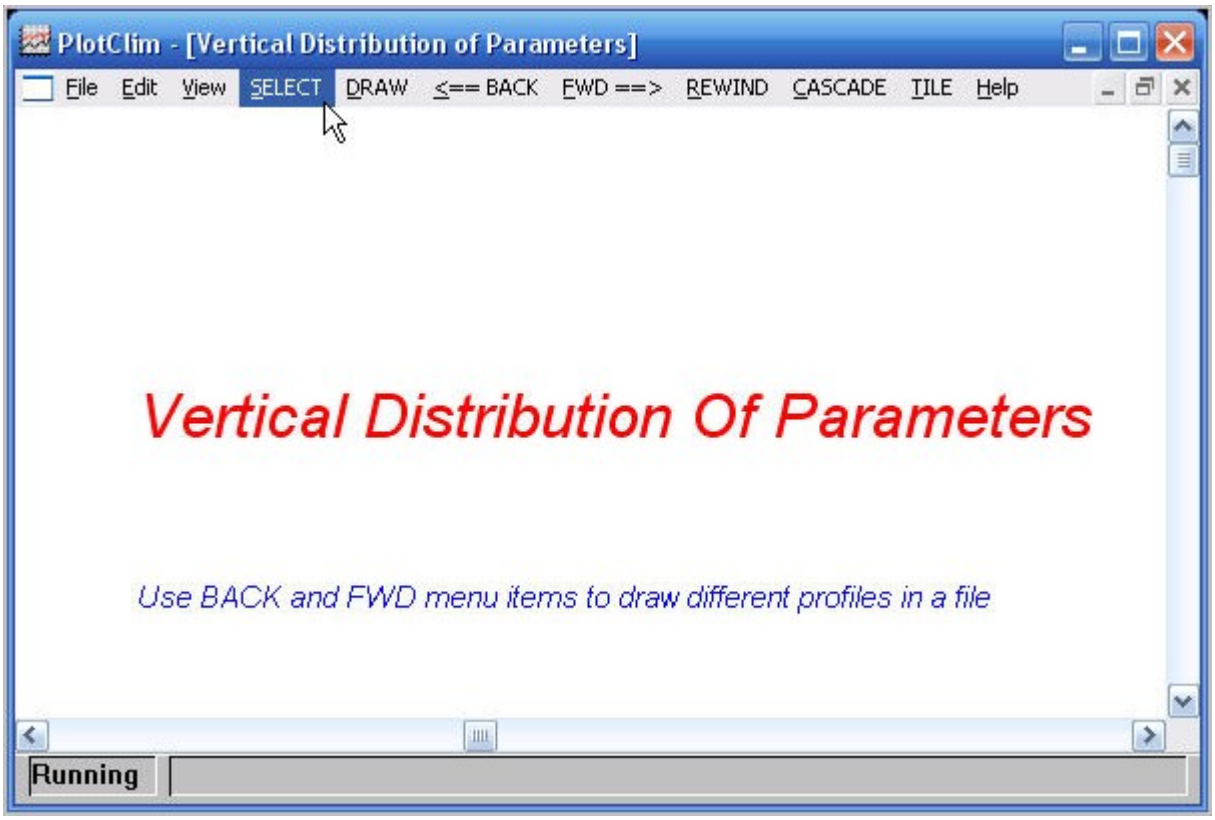
The flags indicating data quality are those currently used in the GTSPP real time quality control and recommended in the MAST/IOC Manual (p310):

0 = no QC performed on the element (NULL is equivalent to 0 here)
1 = QC performed; element correct
2 = QC performed; element inconsistent with statistics but no obvious anomaly)
3 = QC performed; element doubtful (= questionable)
4 = QC performed; element bad (= erroneous = wrong)
5 = the value was changed (=interpolated) after QC
6 -8 = reserved
9 = missing value

Each numerical value must have a quality flag.

- **Plotting of Observed and Climatological Profiles** (file “PlotClim.exe”)

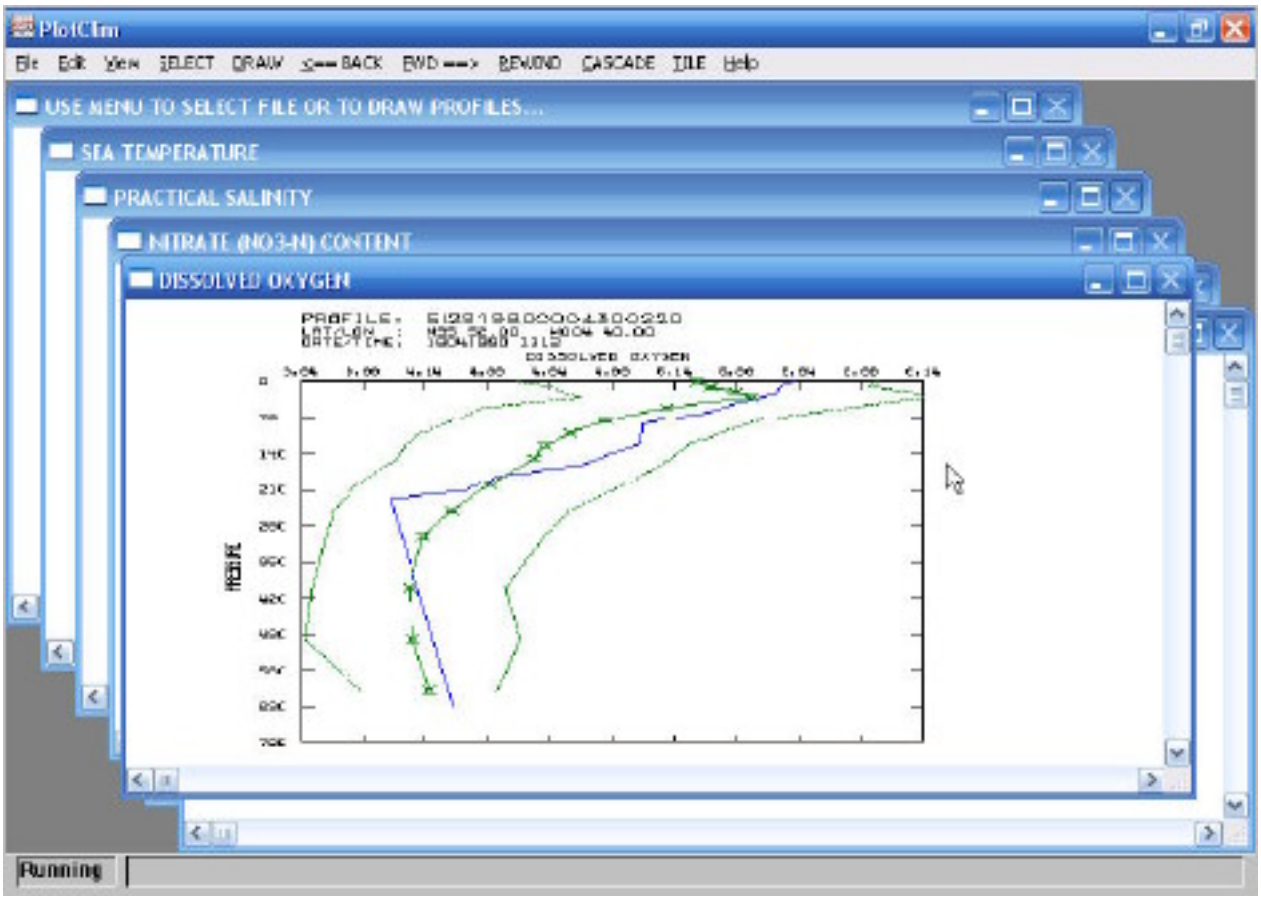
PlotClim produces graphics of vertical distribution of parameters together with statistical climatological profiles and their standard deviation.



“SELECT” button opens an existing file in Medatlas format using the standard open file dialog. At the same time the program reads this file and creates temporary unformatted file used to display graphics of the profiles.
“DRAW” - draws first graphic when the input Medatlas file is selected and prepared for drawing. Use it also after REWIND command.
<== BACK, FWD ==>, REWIND buttos are used to draw previous profile, next profile or to move to the beginning of the file.
The “CASCADE” command rearranges the application's open windows into an overlapping, cascaded pattern.
The “TILE” command rearranges the application's open windows into adjacent, non-overlapping windows.

Climatological profiles for temperature and salinity are taken by the application from previousely prepared ‘ts_autumn.bin’, ‘ts_spring.bin’, ‘ts_summer.bin’ and ‘ts_winter.bin’ binary files and correspond to mean seasonal values in squares of 1 x 1 degree. Statistical data for other available parameters are taken from the file 'nut_cphl_medbs.bin' and correspond to mean annual values in squares of 5 x 5 degrees. You can view the content of last binary file in its text analog 'nut_cphl_medbs.txt'. All these files reside in 'c:\qcmedar\control\' directory.

Here is an example of cascaded windows with the graphics of parameters.



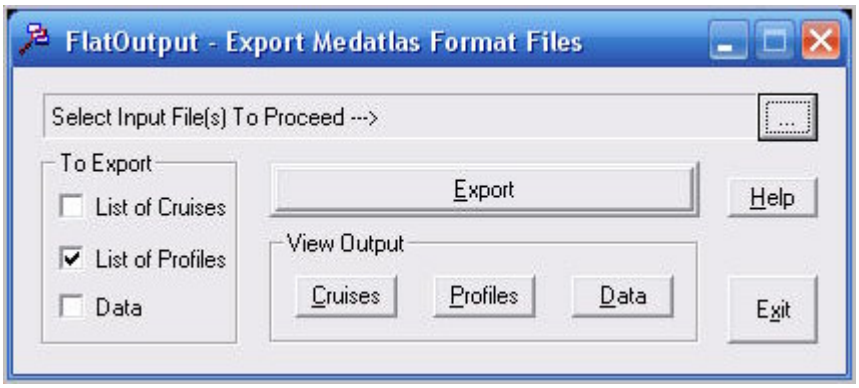
Blue curves correspond to observed parameters, while green lines show climatological or mean statistical profiles, the symbols on green lines are annual statistical values at standard levels. Dashed green lines show standard deviations for statistical profiles.
Climatological profiles are not drawn when the statistical profile for the profile location is not found (case of fluorescence in the screen shot below).
QC flags having values 2, 3 and 4 will be displayed on the blue line with different symbols and colours for each flag value.

5.3 Utilities (folder “utilities”).

This folder contains two programs to export data presented in Medatlas format into other text files necessary for further analysis or for import into MedODV software.

- **Export Medatlas Format Files** (file “FlatOutput.exe”)

This program exports oceanographic data presented in MEDATLAS format to different column-type ASCII files that can be used by other analysis or visualization software, such as Excel, Grapher, Surfer, etc.



Click the button "…" to open an existing file(s) in Medatlas format using the standard open file dialog. Multiple selection is possible using CTRL or SHIFT keys and LEFT mouse button.
In case of multiple selection, text files will be merged and the user is asked for the name of the resulting file that will be created in the same directory as input files.

On the left we can check squares to specify needed export files:

- *List of Cruises*

Output .LCR file with the following abbreviated information will be created:
- cruise reference
- originator cruise name
- ship code and name
- country code
- institution
- chief scientist
- project name
- start date, end date
- availability code (P/L/C)
- first data type
- number of profiles for this type
- QC for the type (Y/N)

- *List of Profiles*

Output .CRD file with following information will be created:
- profile identifiers
- latitude, longitude
- date, time
- bottom depth in meters
- data type (p.e. H09 or H10)
- number of measured parameters
- number of observations records
- global quality flag value
- global quality flags for the parameters

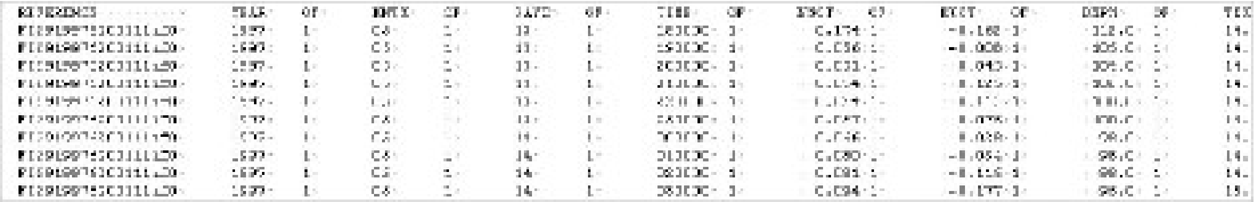
qc flags of latitude, longitude, date and time, depth go after their values.

- Data

Output .LVA file for water observations with the following information will be created:

- profile identifier
- data for all parameters at all levels
- QC flags for the parameters at all levels

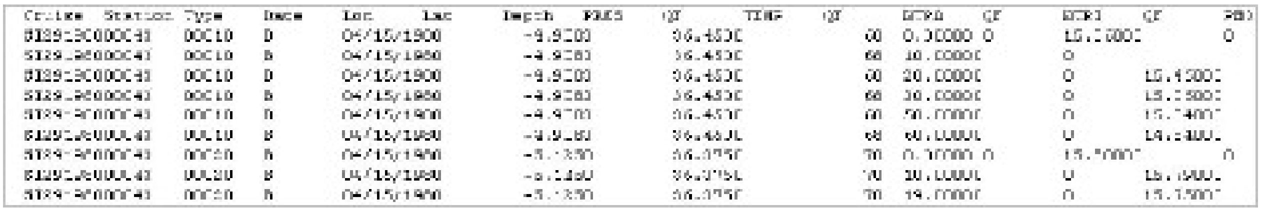
Output .LVA format for time series would be a little different. This is an example of .LVA file for current observations:



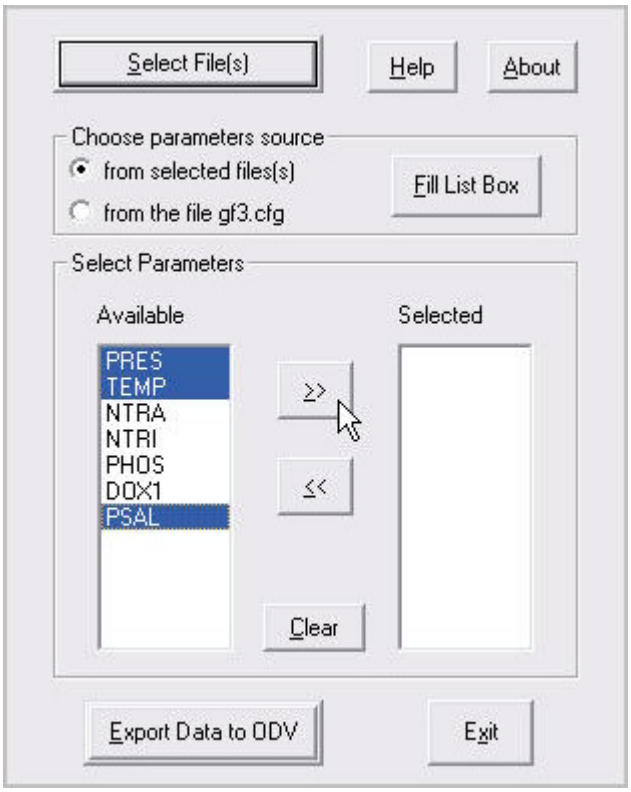
All columns in all output files are separated by tabs.

- **Export from Medatlas to MedODV format** (file "MedODV.exe")

This program converts oceanographic data from Medatlas format to specially formatted spreadsheet ASCII format, suitable for importing into Ocean Data View (ODV) software. The output files of this format can be used also with other tools, p.e. Excel.



ODV (http://www.awi-bremerhaven.de/GEO/ODV) is a computer program for the interactive exploration and graphical display of oceanographic profile data (bottle, CTD, XBT, etc.). The format specifications of TAB-separated spreadsheet file are given in ODV help documentation.



The program has simple interface similar to other components of QCDAMAR.
“Select File(s)” button opens an existing file(s) using the standard open file dialog. Multiple selection of files is realized with standard Windows methods, p.e. keeping pushed CTRL or SHIFT button and clicking files with LEFT mouse button. In case of multiple selection, Medatlas text files are merged into one text file with the name '1tmp.tmp' used as input for export routine and created in the same directory as input files.
“Fill List Box” starts the routine of filling left list box with the list of parameters available in the chosen Medatlas file(s) or with the list of all known parameters from the text file 'gf3.cfg'.
Buttons “>>” and “<<” move the parameter(s) of interest from the left list box (available) to the right (selected) and vice versa.
“Clear” button clears both left and right list boxes. Proved to be useful in combination with 'Fill List Box' button when you want to change quickly the list of selected parameters.
Parameters are chosen by clicking them. Multiple selection is always possible.
“Export Data to ODV” button performs the exporting. Output file has the same name as input, but the extension is changed to 'txt'. In case of several input files, you are always asked to give the output file name using standard Save File dialog.

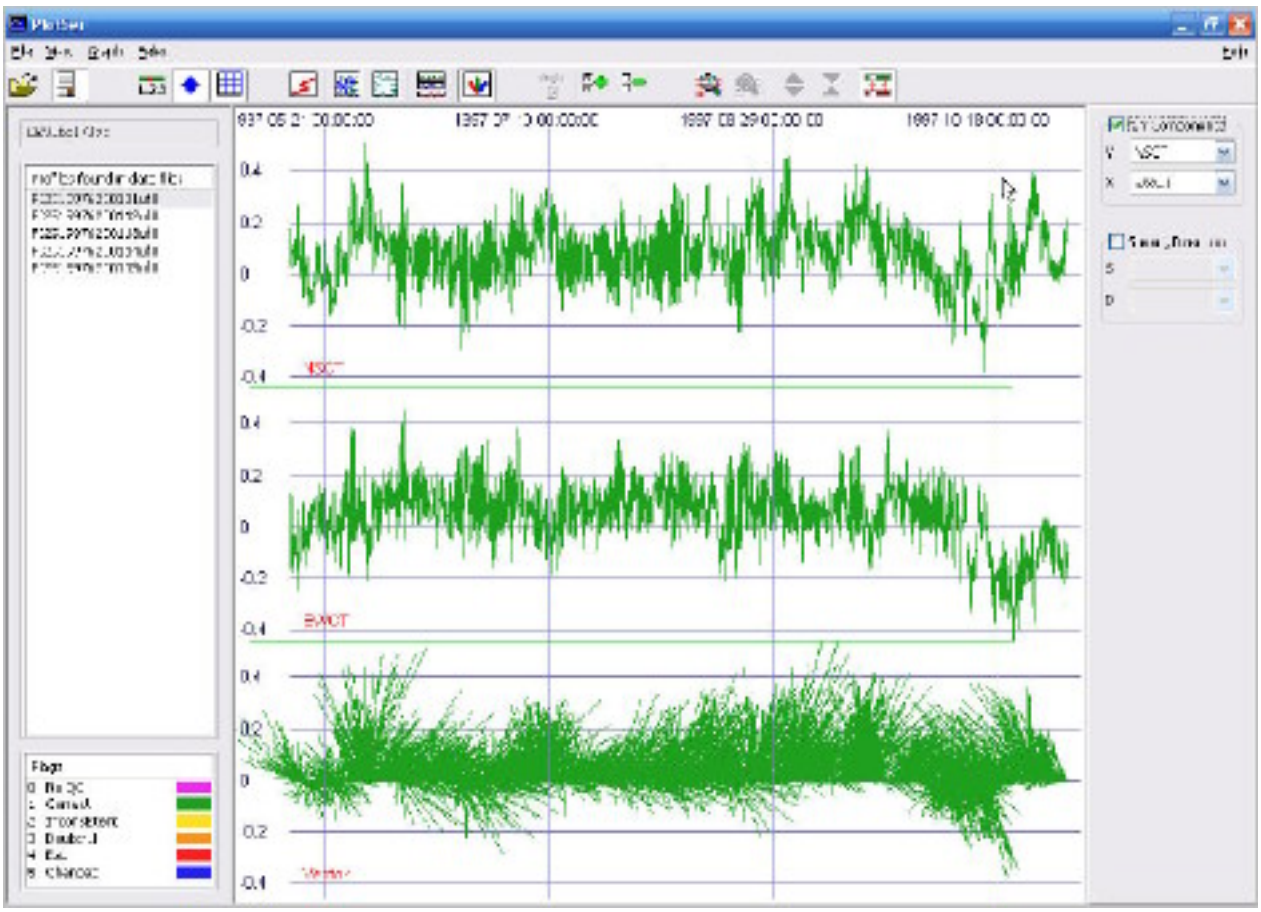
The process is finished when the message like this is displayed:

5.4 Visualization (folder “visual”).

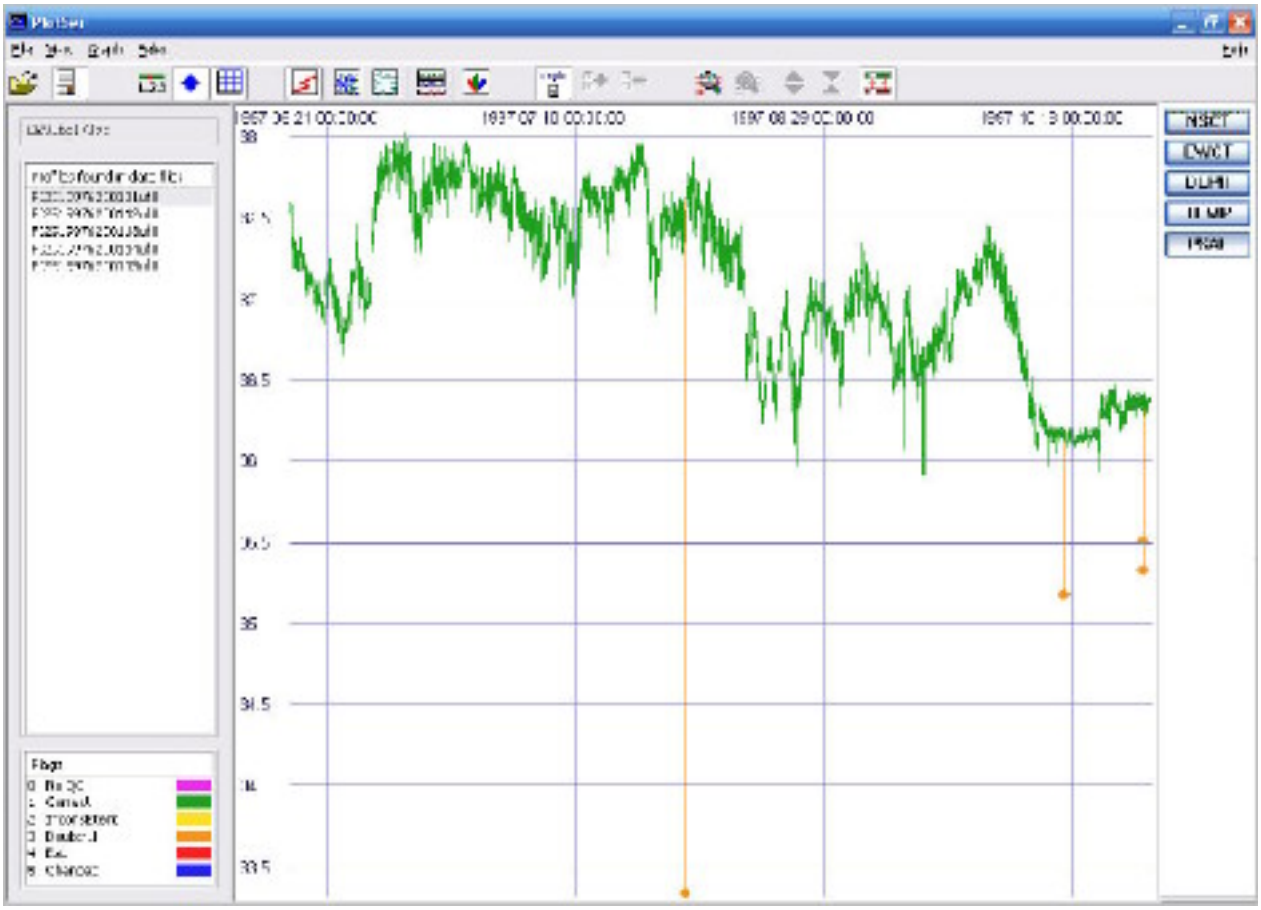
To help with data quality control and make it more flexible and reliable a visualization module were implemented. Three programs PlotMap, PlotVar and PlotSer reside in this folder. Data exported by FlatOutput utility are used as input for these programs. Recently the possibilities of the PlotMap an PlotVar modules were greatly expanded as these two programs were incorporated in SelAVI software working with oceanographic data in MySQL database format.

- **Plot data series** (file PlotSer.exe)

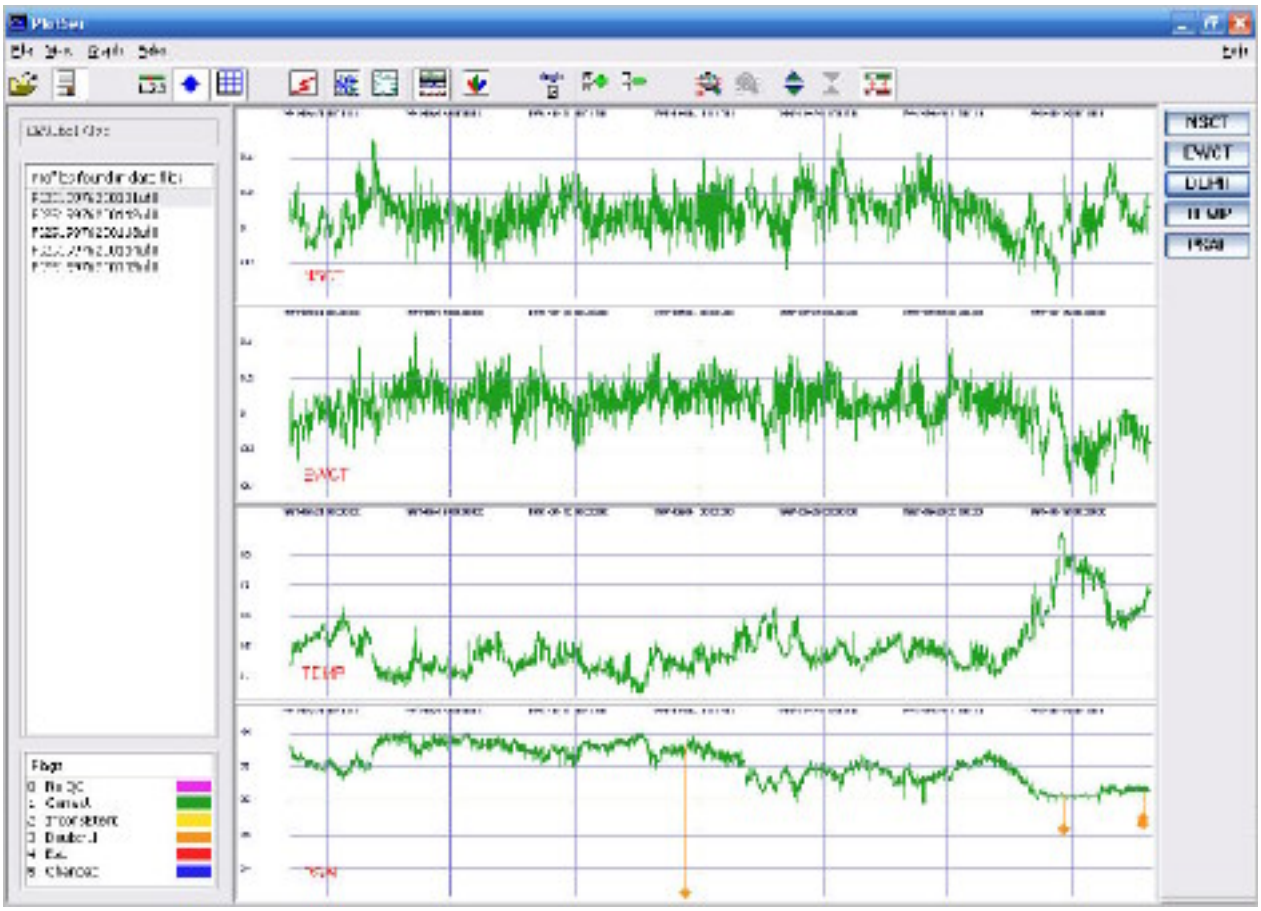
This program displays on the screen several types of graphics of time series. It´s especially helpful for “visual” analysis of the results of automatic quality control of currents. Thus both components and vectors of currents can be displayed together. When vertical and horizontal component are present, the speed and direction necessary to plot vectors of currents are calculated. And on the other side, when we have only vectors, the visualization of components is possible.



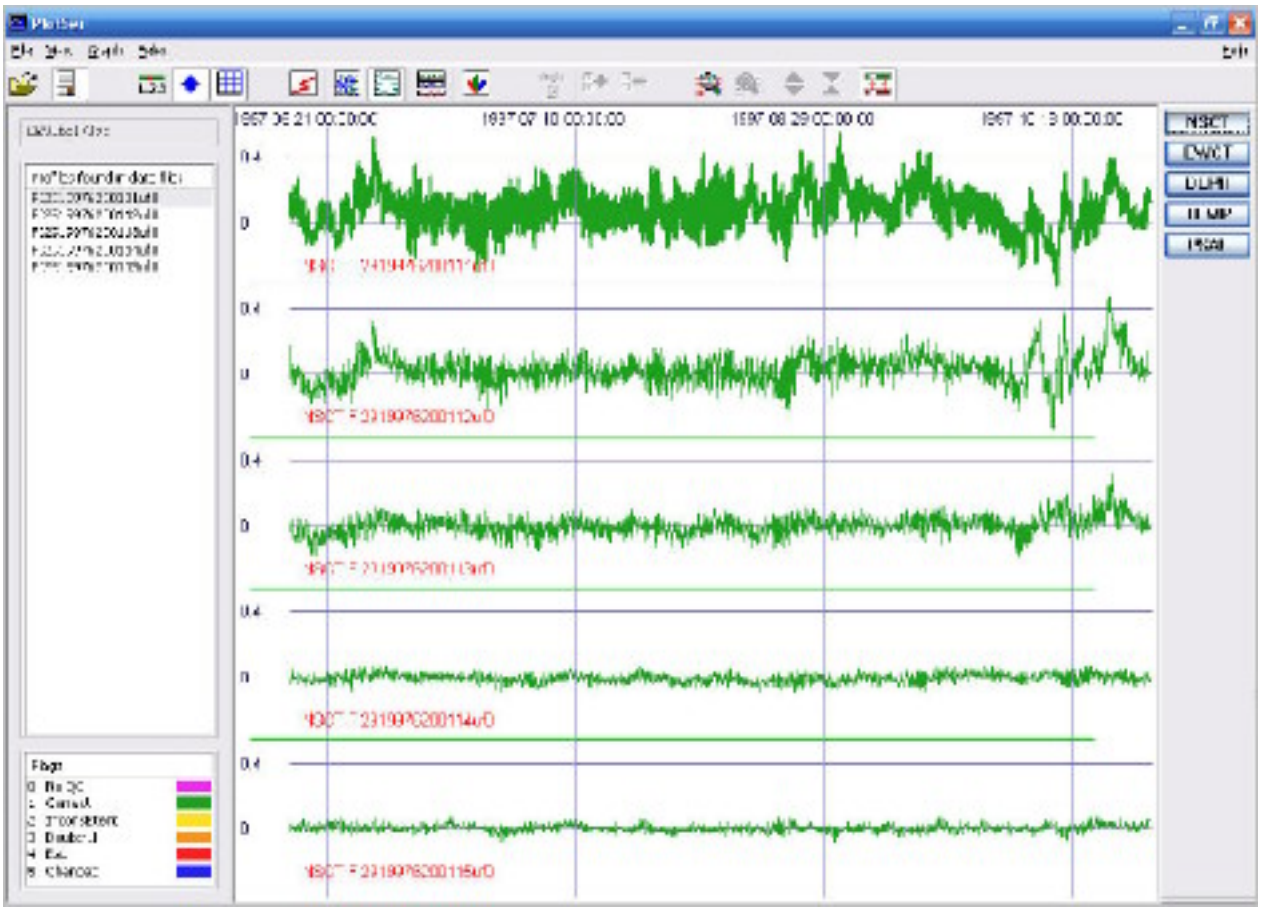
In a “single” type of graphic we can see any parameter alone or in the common scale together with those chosen by clicking correspondent buttons having parameters names on the rightmost column.



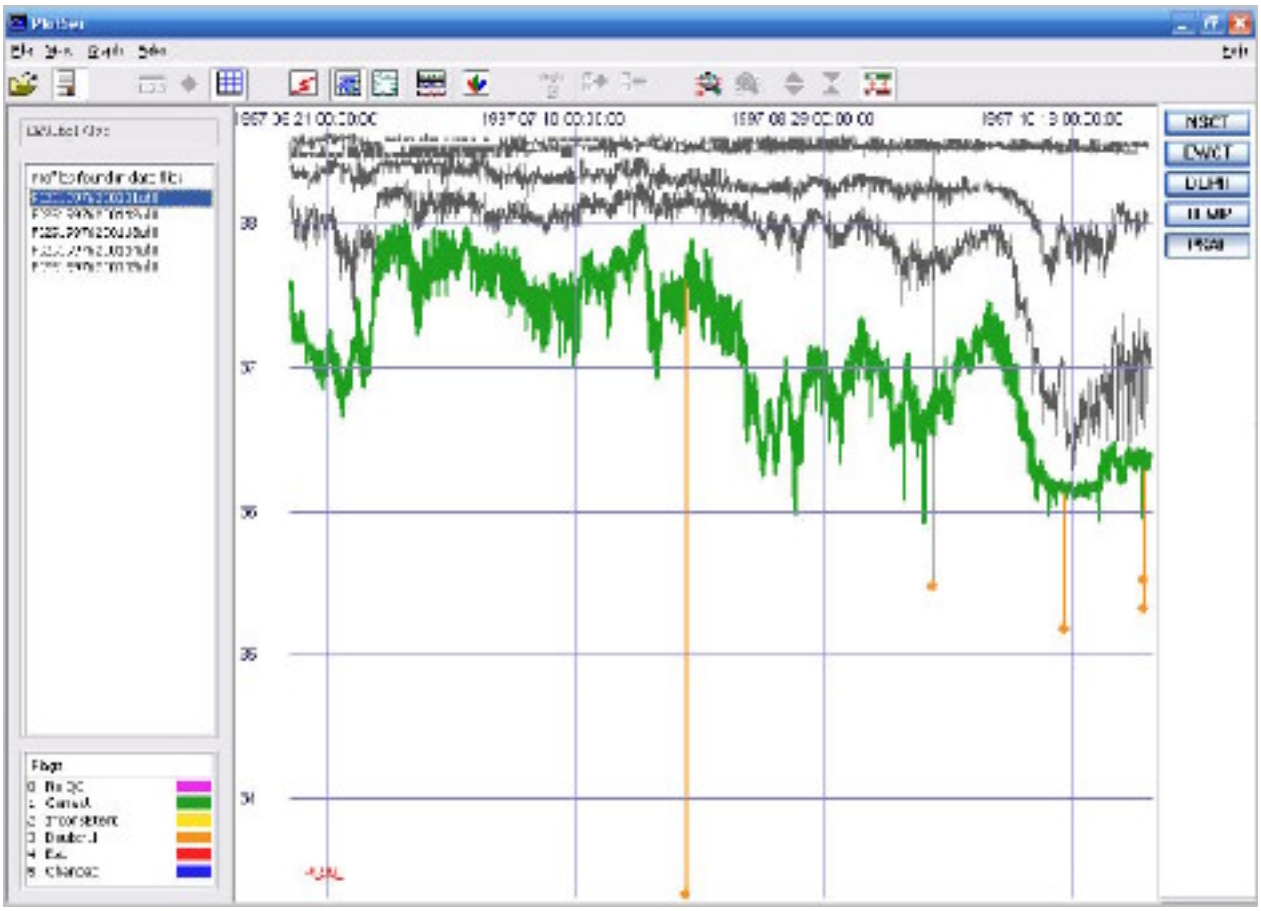
On the other hand “multiwindow” type of plot presents chosen parameters together with individually scaled graphics.



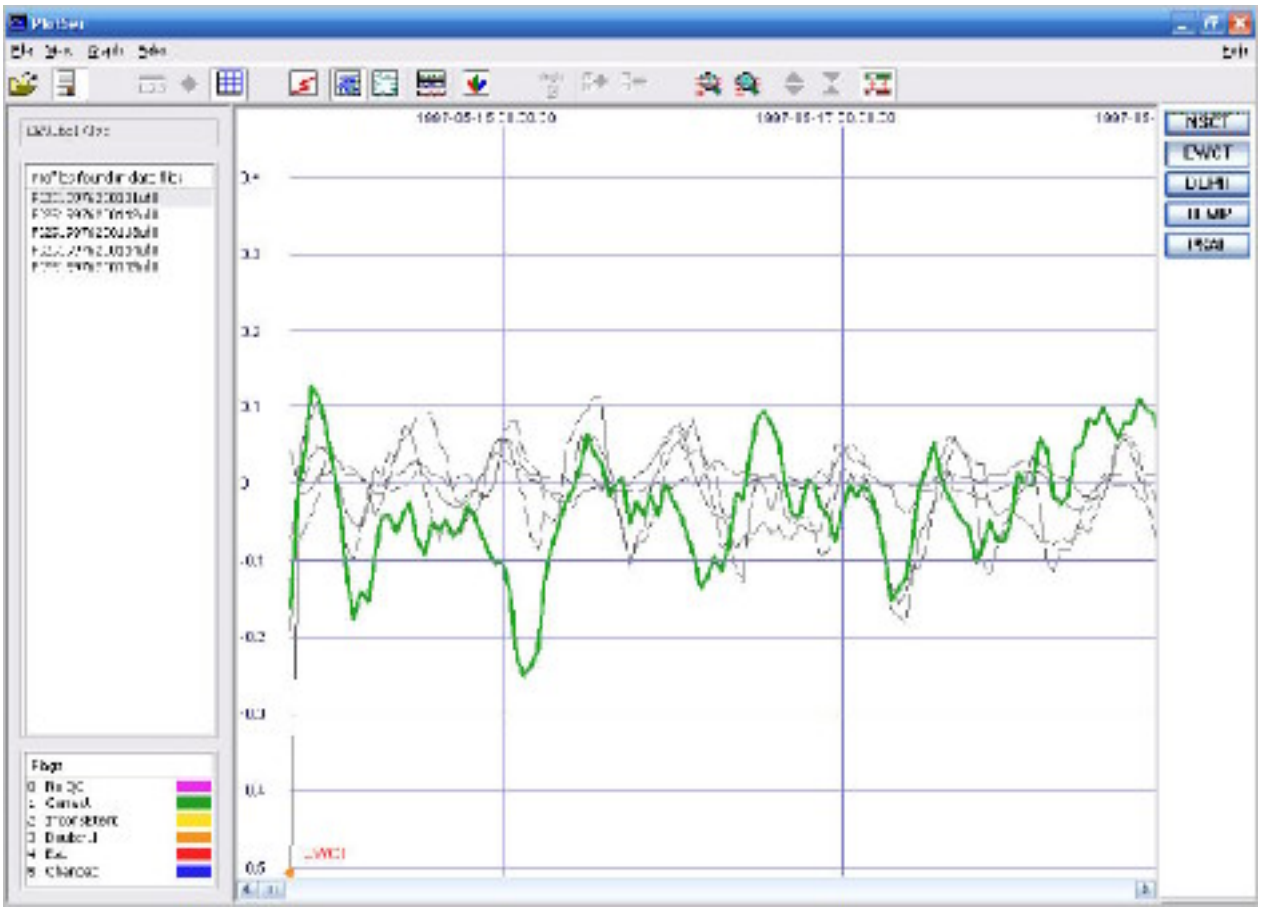
In a “multiserie” type of plot we can observe changes in time of any parameter in different time series. Next plot shows changes of the current north-south component with depth.



“Superposed” window is very useful to see changes of any chosen parameter in the same scale:



The program has great options of zoom and scroll synchronizing. Next plot shows “zoomed in” graphic of the east-west component of current to see better details of the its first values.



As you may see the horizontal scroll bar has appeared automatically in this case.

6. SUMMARY

The QCDAMAR makes possible to **archive and exchange** data in the same format and with the same **quality control** level. Afterwards the data will be suitable for sending to any National Data Center.

Later the data from different sources can be integrated into a large data set for using in future investigations and for developing long time series and products as climatologies, temporal and space variability.

Because QCDAMAR does not change the data values and preserves the DC and DM history in the headers, the Quality Control can be repeated at any time from the beginning.

If you find some errors please report them to IEO. Comments about improvement of the sofware, bugs, usability or something else will be welcome.

7. References

BRASSEUR P., SCHOENAUEN R.. Seasonal Data Fields for the Mediternean sea. University of Liege, GHER internal Report.

CURE D., FICHAUT M., MAUDIRE G., 1995. Dossier de spécification du logiciel SCOOP. SISMER/SCOOP/DSL/010,VI.2.Rap. int. IFREMER/SISMER.

FOFONOFF N.P. and MILLARD R.C., 1983. Algorithms for computation of fundamental properties of seawater Unescop technical papers in marine science 44,53 p.

ICES. Oceanographic Data Center. Data Exchange Format

ICES/IOC 1995. GETADE Formating Guidelines for Oceanographic Data Exchange (Preliminary report).

JONES M.T., BODC, 1994. General Bathymetric chart of the Ocean, the GEBCO digital Atlas, CDROM.

LEVITUS S. Climatological Atlas of the World Ocean, 1994 Edition (in preparation).

MEDATLAS GROUP, 1994 Specifications for the Mediterranean Sea Data Banking and Quality Controls. SISMER/IS/94-0014, V 5.

MILLARD C., FICHAUT M., MEDAR/MEDATLAS GROUP, 1999 MEDAR/MEDATLAS II protocols., DITI/DT SISMER.

MILLERO F. J. and POISSON A., 1981. Internacional one-atmosphere equation of state of sewater. Deep-sea ressearch, Vol28A, No.6, pp 625-629.

REINIGER R.F. and ROSS C.K., 1968. a method of interpolation with application to oceanographic data. Deep-Sea Research, Vol 15, pp 185-193

NOAA/NODC, 1984. Key to Oceanographic Records Documenatation No.14 User Guide

NOAA/NGDC, 1993. ETOPO5. 5-minute gridded elevation/bathymetry for the word. Global Relief Data CDROM.

UNESCO 1989. A General Formatting System for Geo-reference Data. IOC. Manuals and Guides NO. 17.

UNESCO 1991. Manual on International Ocanographic Data Exchange. Manuals and Guides NO. 9

UNESCO 1993. Manual of Quality Control Procedures for validation of Oceanographic Data, prepared by CEC:DG: XII, MAST and IODE. Manuals and Guides NO. 26

UNESCO 1994. Protocols for the Joint Global Ocean Flux Study (JGOFS) core measurements. IOC, Manual and Guides 29. 170 p.

UNESCO/IOC, 1995. OceanPC Manual. Soltware for Management and International Exchange of Oceanographic Data. IOC/INF-943. 130 P.