Added values to oceanographic multi-parameters data sets,
Example of European concerted action : MEDAR/MEDATLAS 2

Michèle FICHAUT [(1)], Maria Jésus GARCIA [(2),] , Gilbert MAUDIRE [(1)], Michel RIXEN [(3)],

Evgeny VYAZILOV[(4,)] and MEDAR Group

[(1)]IFREMER, Centre de Brest, BP 70, 29280 Plouzané, France

☎33 (0)2.9822.4643, Fax 33 (0)2.9822.4644)

michele.fichaut@ifremer.fr

gilbert.maudire@ifremer.fr

[(2)] IEO, c. Corazon de Maria 8, MADRID 28002, Spain

MJesus.garcia@md.ieo.es

[(3)] GHER, Univ. de Liège, BE4000, LIEGE, Belgium

M.Rixen@ulg.ac.be

[(4)] RIHMI-WDC, 6, Koroleva St., Kaluga Region, OBNINSK 024020, Russia

nodc@meteo.ru

Key words: Mediterranean, Black Sea, oceanography, software, data product

Abstract

The oceanographic data are currently collected by hundreds of laboratories and to access them may be neither easy nor fast. It was the overall objective of the EU concerted action MEDAR/MEDATLAS (MAS3-CT98-0174 & ERBIC20-CT98-0103) to make available a comprehensive data product of multi-disciplinary oceanographic *in-situ* data and information, collected in the Mediterranean and Black Sea, through a wide co-operation of the bordering countries. The result of this major effort of data rescue is to make available 286 000 qualified hydrographic stations of bathythermograph, CTD and bottle casts, from 150 different laboratories, and increases twofold the volume of available data compared to the previous 1997 release. To get a coherent data set from so many sources, and to prepare value added products, it was necessary to develop common standards, methodology and several software tools to process the data. The final data product includes observed data, meta-data, gridded data and maps. It includes also a user friendly data extraction and visualisation software to allow scientists, modellers, engineers, teachers, students and managers to process easily such a precious data set.

Résumé

Les données océanographiques sont couramment collectées par des centaines de laboratoires et il n'est ni facile ni rapide d'y avoir accès. Ce fut l'objectif de l'action concertée MEDAR/MEDATLAS de l'UE (MAS3-CT98-0174 & ERBIC20-CT98-0103) de donner accès à un jeu le plus exhaustif possible de données océanographiques *in situ* multidisciplinaires collectées en Méditerranée et en Mer Noire, par une large coopération des pays riverains. Il résulte de cet effort majeur de sauvegarde des données, 286 000 stations hydrologiques qualifiées de bathythermographes, bathysonde et palanquées de prélèvements bouteille provenant d'environ 150 laboratoires, ce qui multiplie par 2 les jeux de données disponibles par rapport à la précédente mise à jour de 1997. Afin d'obtenir une base de données intégrée cohérente à partir de tant de sources différentes, il était nécessaire de développer des normes communes, une méthode et des logiciels pour traiter les données. Le produit final contient les données observées, des méta-données, des données maillées et des logiciels pour permettre aux scientifiques, modélisateurs, ingénieurs, enseignants, étudiants et gestionnaires des territoires d'utiliser facilement ce précieux jeu de données.

1.Introduction

Basic oceanographic parameters like temperature, salinity and nutrients are needed for various scientific and technical studies. However most of the time, they remain dispersed among all the different organizations which carry out oceanographic cruises. The overall objective of the EU concerted action MEDAR/MEDATLAS (MAS3-CT98-0174 & ERBIC20-CT98-0103) was to make available a comprehensive data product of such multi-disciplinary in-situ data and information in the Mediterranean and Black Sea, through a wide co-operation of the Mediterranean and Black Sea countries.

The specific project objectives were: 1. to compile and safeguard historical data; 2. to make available comparable and compatible data sets of : temperature, salinity, oxygen, nitrate, nitrite, ammonia, total nitrogen, phosphate, total phosphorus, silicate, H2S, pH, alkalinity, chlorophyll-a profiles by using a common protocol for formatting and quality checking; 3. to prepare and disseminate qualified value added products by using efficient gridding and mapping methodology; 4. to enhance communication between data managers and scientists to improve the data circulation.

A strong international co-operation was initiated, with data management responsibility at national and regional levels,, and specific tasks to prepare an integrated database and products. The location of the participating institutes is reported on Fig. 1 with colours referring to the regional or thematic activities.

The data management structure was distributed between 19 national data centres, 4 Regional Data Centres (RDC) respectively for the Black Sea, the Mediterranean Eastern Basin, the Central Mediterranean and the Mediterranean Western Basin, and one co-ordinating and Global Assembling Centre (GAC). At the national level, each participant, who represented the National Co-ordinator for International Oceanographic Data and Information Exchange (IODE) at the Intergovernmental Oceanographic Commission (IOC) of UNESCO, compiled and safeguarded copies of the data sets dispersed in the scientific laboratories of his country, and reformatted them at the common MEDATLAS format. Each National Oceanographic Data Centre (NODC) or Designated National Agencies (DNA) of the participating countries sent his data set to the corresponding RDC for regional expertise. After regional quality control of the data, each RDC sent the data sets to the GAC for final checking and merging in the final integrated database.
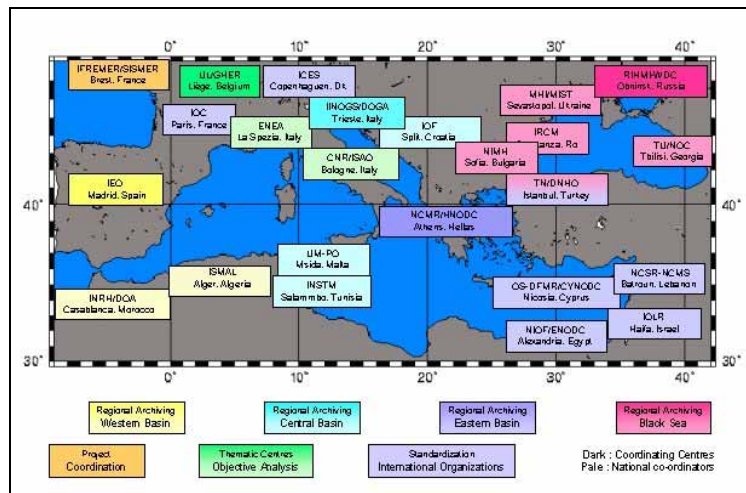

Fig. 1 : MEDAR Network

Thanks to this cooperation, the volume of available data represents now more than 286 000 stations (vertical profiles) from about 150 sources laboratories of 33 countries. The content of the database with regards to the different data types and parameters observed is given in Table 1. This data release doubles the volume of the previous one (MEDATLAS Group, 1997).

Table 1 : Content of the MEDAR 2002 Database

DATA TYPE	NB of STATIONS
BOTTLE	88453
CTD	36054
MBT	81465
XBT	80425
Thermistor string	29
Total	**286426**

PARAMETER NAME	NB OF PROFILES
SEA TEMPERATURE	284946
PRACTICAL SALINITY	118509
DISSOLVED OXYGEN	44989
NITRATE (NO3-N)	10588
NITRITE (NO2-N)	10561
AMMONIUM	5301
SILICATE	15936
PHOSPHATE	20808
ALKALINITY	2548
PH	14548
CHLOROPHYLL-A TOTAL	4716
HYDROGEN SULPHIDE (H2S)	1843
TOTAL NITROGEN	153
TOTAL PHOSPORUS	2381

To get a coherent integrated data set from so many sources, an important work for formatting, checking and assembling the data was necessary. This has been a key task of the project, including the development of a common standard protocol and the implementation of the quality checks procedure. Further tasks were dedicated to prepare gridded climatological fields and maps of the data. In parallel, a distributed web site (www.ifremer.fr/medar) was developed to exchange information and meta-data such as the reference protocol. The last task was to prepare and publish a data product on Cdrom with meta-data, observed data, gridded data and software. This publication is important to facilitate and broaden the use of the data, within the scientific community and beyond, especially for engineers, teachers students and managers.

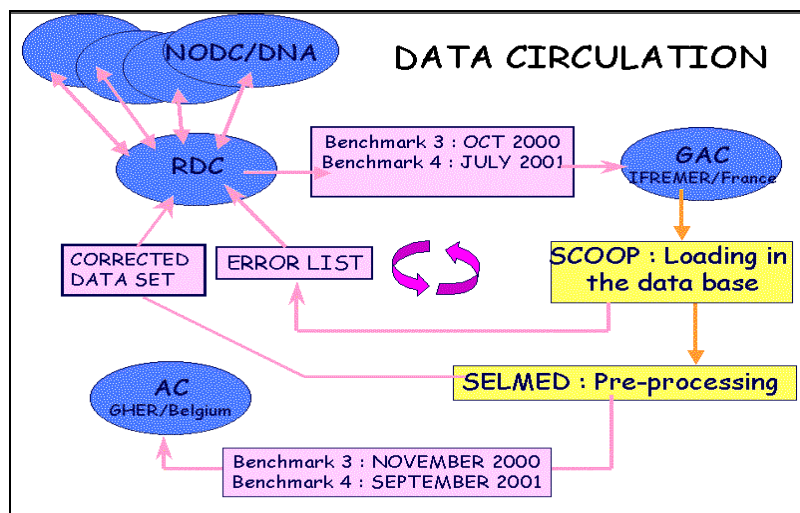The implementation of these interrelated tasks is schematised in Fig. 2.

Fig.2: Scheme of the Data processing in MEDAR
NODC/DNA= national data centres, **RDC**= regional data centres, **GAC**= global assembling centre, **AC**= analysis centre
SCOOP= data quality control software (system of control oriented physical oceanography)
SELMED = data selection and pre-processing software for the Mediterranean & Black Sea

At each stage of the project, an important development of complex software tools was necessary: expert software data checking and visualisation tools, software to perform the objective analysis (gridding) and mapping of the data, and specific software for the published data product for data retrieval and visualisation. This paper is focused on this final product on Cdrom and the software included (SELMED), which has also been used for data pre-processing, before the objective analysis.

2. Data Quality Control

The first data management task was to document and reformat the observation data. The exchange format was the auto descriptive MEDATLAS format (MEDAR Group, 2001). This format includes data points and

meta-data necessary to perform the quality checks such as a minimum information on the experimental conditions and source laboratories.

The data sets were afterwards checked for quality (QC) according to the common protocol based on the international CEC/DG XII, MAST and IOC/IODE (1993) recommendations, with automatic (objective) and visual (subjective) checks, which results in a quality flag added to each numerical value. Rather sophisticated software have been designed to perform these checks on location and date (Fig. 3) and data points (Fig. 4) respectively SCOOP on Unix (Bardet et al., 2000) and QCMEDAR (Garcia et al., 2001) on Windows platforms. With SCOOP software, all the checks are performed automatically first, and then an operator validates the flags, by checking the internal coherency of the observations of a same cruise and eventually the coherency with any other cruise carried out in the vicinity. The software allows also to search for duplicates, which are a major problem in the historical data sets. These quality checks have been performed on each profile in the RDC, and then sent to the GAC. The GAC finalised the last quality and duplicate checks, by using the same software, and integrated all the qualified data in a unique system. Several up and down transfers of subsets of data between the data centres of the structure have been necessary before getting a fully qualified integrated data set.

An important aspect, is the necessity to load the local statistics in the software. For temperature and salinity, estimates of these statistics like broad range scale of variation, climatological means and (to some extend) standard deviation were available from MEDATLAS 1997 release. For the nutrients, the first statistics were taken from the World Atlas (Levitus et al., 1998). These statistics were too broad for nutrients, and further preliminary studies were made at each regional data centre to improve them and feedback with new climatological computations has been necessary to finalise the checks.
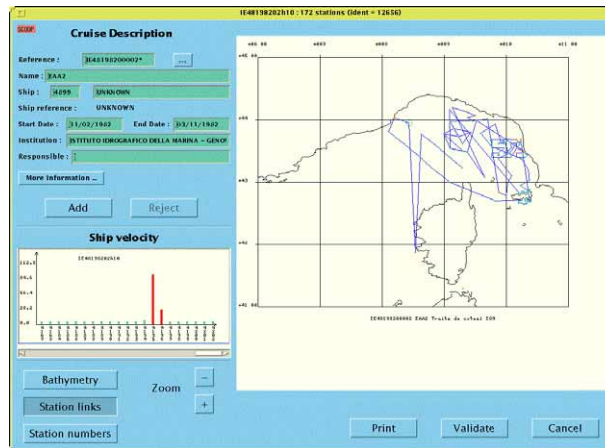

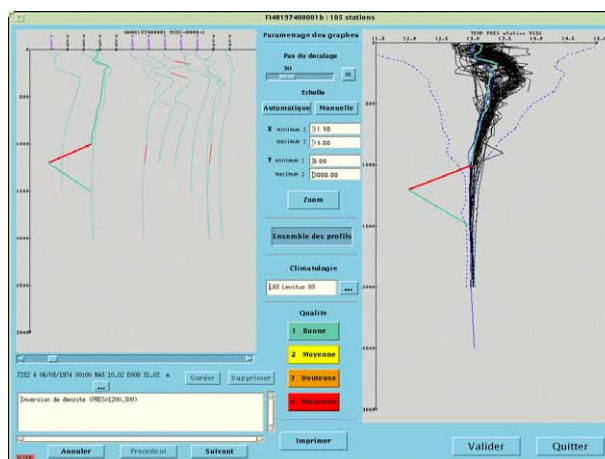**Fig. 3: Check of the location, date and duplicates**

Fig. 4: Check of the data points

3. Objective Analysis

Raw in-situ data sets are difficult to interpret and higher level products are required to offer a more complete and synthetic view of the Mediterranean and Black Sea bio-chemical systems. Climatological or gridded fields are also needed to provide initial or boundary conditions to numerical models. For these computations, the data without outliers have been interpolated at vertical standard levels and transferred to the Analysis

Centre in Liège, where an objective analysis has been performed. The software used to performed this pre-processing is disseminated with the data product described in the following chapter. For the objective analysis itself, instead of using the classical objective analysis scheme, gridded fields have been computed using the Variational Inverse Model (Brasseur 1991, Rixen et al. 2001), shown to be statistically equivalent to optimal interpolation, but with better efficiency. The basic idea of variational analysis is to determine a continuous field approximating the data and exhibiting small spatial variations. In other words, the target of the analysis is defined as the smoothest field that respects the consistency with the observed values over the domain of interest. It is also referred to as a spline interpolation method. Expressed in mathematical terms, the analysis is obtained as the minimum of a variational principle in a two-dimensional, horizontal space. The computations are made on a climatic, seasonal or monthly scale when relevant. The initial characteristic space length of the finite elements is 1 degree for the Mediterranean, 0.8 degree for the Black Sea and 0.5 degree for local analysis. This mesh is next divided by 3 and the final grid resolution is 0.2 degree for the Mediterranean, 0.1 degree for the Black sea and 0.04 to 0.08 degree for the local sub-region. However the results can be robust only when the space and time data coverage is sufficient.

The gridded fields have been mapped horizontally and vertically. On these maps such as the salinity distribution (Fig. 5 and 6), the main features of the region, such as the Atlantic inflow, the mesocale gyres and meanders appears very clearly at the resolution of the climatological computation. The results of this new climatology are expected to improve the quality check methodology in the future.
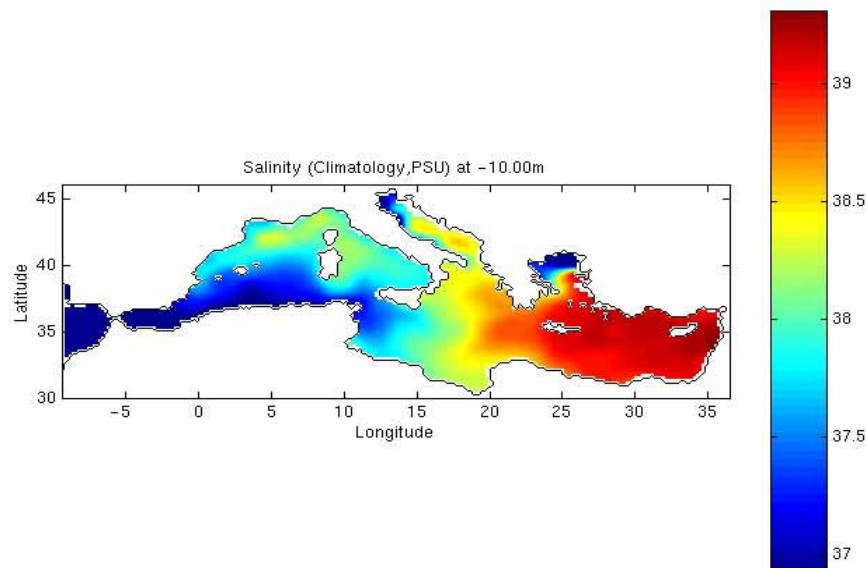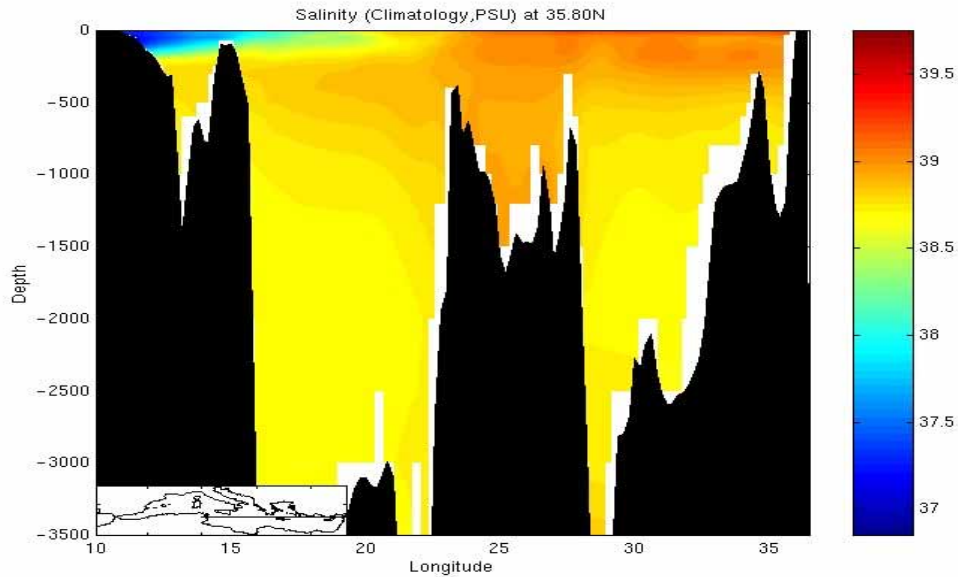

Fig. 5: Salinity distribution at 10 dbar

Fig. 6: Vertical Salinity distribution in the Mediterranean Eastern Basin

3. Data Products

An important component of the project, to facilitate and accelerate the access to the data, is the publication of the integrated database. The observed and gridded climatological data, the climatological maps and all the meta-data and documentation about the project are in publication on a set of four CDRoms (Fig. 7). This package represents a real scientific and educational tool, as it contains also user friendly and expert software to extract, visualise, plot, check for quality and make some scientific computations on the data.

Fig. 7: The set of four auto mounted MEDAR CDroms

CDROM 1 contains the html documentation about the project (project summary, MEDAR network partnership, data sources by country, location maps of the collected data, format and codes description, quality assurance, manuals and documents published during the project, documentation about delivered software: SELMEDAR, ODV, QCMEDAR).

- The cruise inventory, a GIS developed by RIHMI – World Data Centre, Russian Federation (Vyazylov and Puzova, 2001) (Fig. 8)

- QCMEDAR software developed by IEO, Spain (Garcia et al., 2001)

- Ocean Data View Software, a very powerful software to process and plot data developed for the international WOCE experiment at the Bremerhaven University (Schlitzer, 2000)
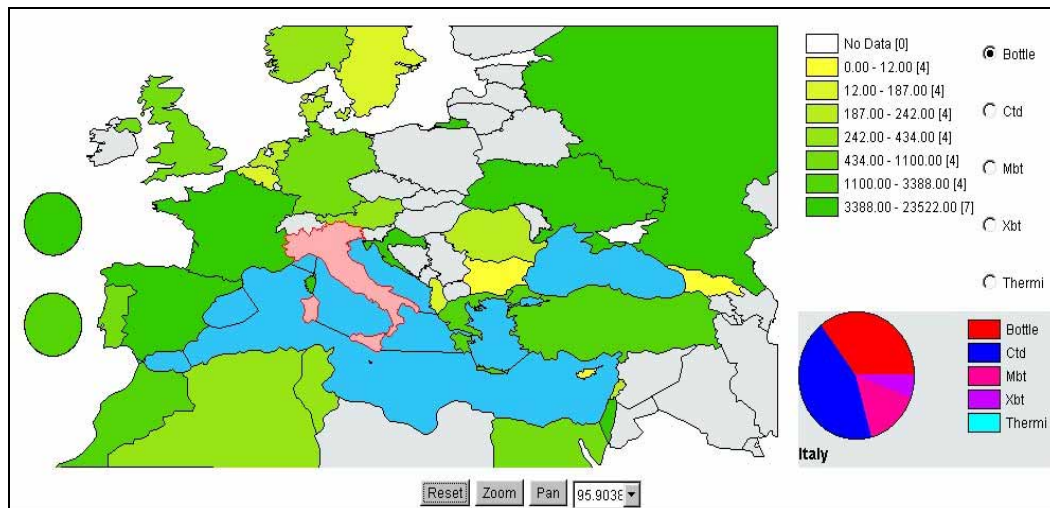

Fig. 8 GIS MEDAR Cruise Inventory -
Distribution of the total number of profiles collected by country
and numbers by data type for the current country (here Italy)

CDROM 2 is the core of the database as it contains all the observed data files (around 650 Mo). The data at MEDATLAS format are organised in hierarchical directories :

- Level 1 – Data Type: Bottle, Ctd, Xbt, Mbt, Thermi
- Level 2 – Time period
- Level 3 – Year or data files
- Level 4 – Data files

These data are accessible through a software designed as a user-friendly interface so that users can extract data from the whole data set, following several criteria. This software, available for PC/WINDOWS, allows:

- extraction and display of any subset of data selected according to any combination of the following criteria: geographical location (Fig. 9), data type (bottle, CTD, bathythermograph), cruise name and reference, time period, month, ship, country, parameter and quality flags

- extraction at three output formats: MEDATLAS, Comma Separated Values for spreadsheet, Ocean Data View (WOCE/Bremerhaven University visualisation software)

- interpolation at 25 pre-defined standard levels or at user defined standard levels

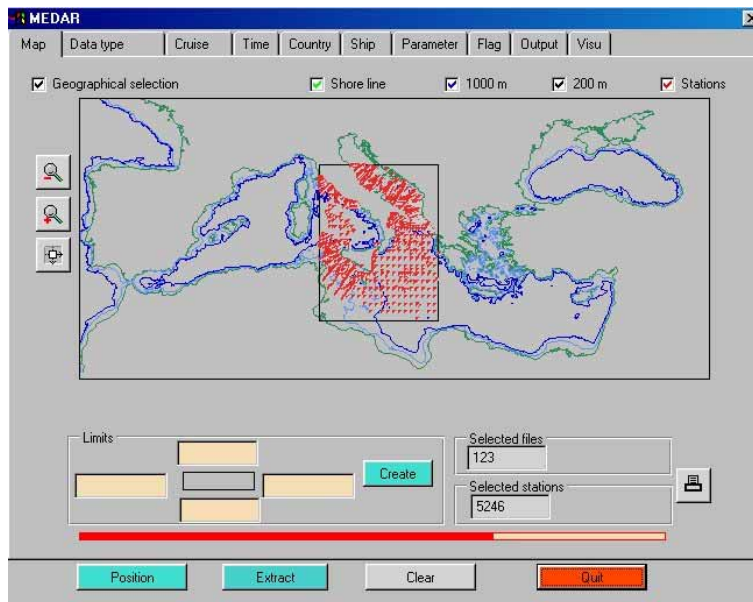- visualisation of the selected data on parameter/ parameter plots (Fig. 10).

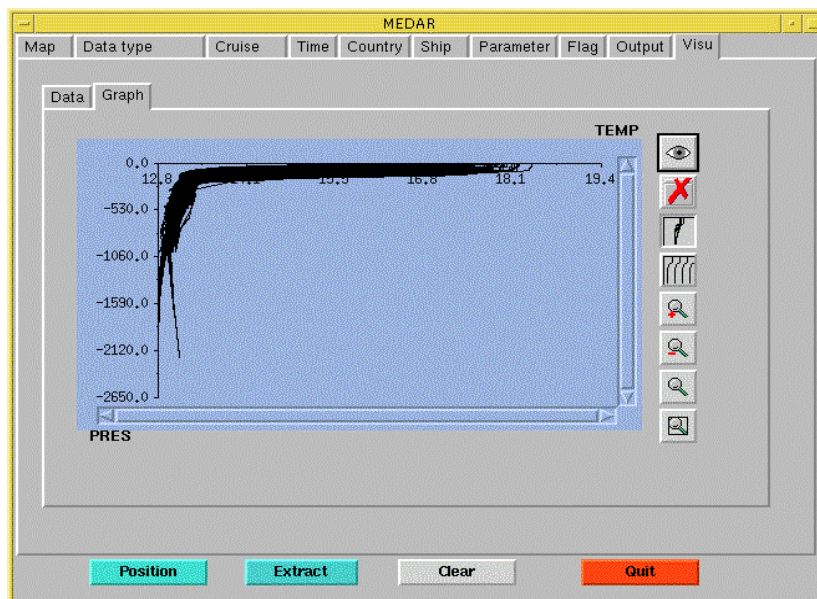**Fig. 9: SELMEDAR selection software – extraction of data by geographic area**

Fig. 10: SELMEDAR selection software – Plot of temperature profiles

CDROM 3 contains all the maps and figures drawn from the gridded data resulting from the objective analysis.

CDROM 4: contains all the numerical fields from the objective analysis, in binary NETCDF files.

The variational inverse model software developed to compute these fields and its documentation can be downloaded from the Liège University web site (http://modb.oce.ulg.ac.be/Medar/).

4. Conclusive remarks

Such an integrated database has been developed by the joint work of data managers for the data compilation and processing, and software engineers, to get tools to do it. Various kind of technologies and skills have been used, from commercial RDBMS and GIS systems, local interfaces and dedicated data processing systems, for different kinds of platforms. All this software appeared rather powerful to meet the project objectives and deadlines, and provide qualified data to users, however there is also a demand to get similar possibilities on line. For the meta data, this exists already, and to some extend for the gridded data. For the observation data, there are some development underway (Maudire et al., 2002) but it should be noticed that the difficulty is not uniquely technical but operational. The data remains under the responsibility of the source countries which collected them, and they are not always willing to delegate it to a remote server. Furthermore, the problem of synchronisation of continuous updating between several servers is not trivial. For these reasons, and also for many users who have not always access to fast communication networks and with limited computer facilities, the publication on Cdrom of MEDAR database remains the basic result of the project. It corresponds to a real value added product, compared to the data themselves, and the addition of user friendly dedicated free run time software facilitates their use by non specialists. Therefore it can be expected that, even with future direct on line access to data, the publication on Cdrom as a full package of data and processing tools, will remain a necessary task to improve the use and the promotion of the database.

5. References

[1]   Brasseur, P. (1991). *A Variational Inverse Method for the reconstruction of general circulation fields in the Northern Bering Sea*. Journal of Geophysical Research, 96(C3):4891-4907.

[2]   CEC/DG XII, MAST and IOC/IODE, 1993. Manual of Quality Control Procedures for validation of Oceanographic Data. UNESCO Manual and Guides N° 26, 436 P.

[3]   Bardet D. and Fichaut M., 2000. System of Control Oriented Oceanographic Parameters SCOOP, V4.2, User Manual. Rap Int. IFREMER/TMSI/IDM/SISMER/02-072.

[4]   Fichaut M., G. Maudire, M.Larour, 2001. *SELMEDAR, Data selection from the MEDAR database*. Rap. Int. R.INT.TMSI/SISMER-ISIB/SIS/01-042, 28 p.

[5]   Garcia M-J.., Nikouline A.., Alcantara P., Molinero J., 2001. QCMEDAR,. *MEDATLAS Format & Quality Control (V1.1)*. Users guide. Rap. Int. Instituto Espanol de Oceanografia IEODC.

[6]   Levitus, S., T.P. Boyer, M.E. Conkright, T. O' Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson, R. Gelfeld, 1998: NOAA *Atlas NESDIS 18,World Ocean Database 1998*, U.S. Gov. Printing Office, Wash., D.C., 346pp + set of Cdroms.

[7]   Maudire G., Le Duff M., Moussat E., 2002. *Marine Databases : a step forward integrated requests. Symposyum Ensuring Long-Term Preservation and Adding Value to scientific and technical data*. Toulouse 5-7 November 2002, France

[8]   MEDAR Group, 2001. Medar-Medatlas Protocol *(Version 3) Part I: Exchange Format and Quality Checks for Observed Profiles*; Rap. Int. IFREMER/TMSI/IDM/SIS002-006, 50 P.

[9]   MEDAR Group, 2002. MEDAR/MEDATLAS 2001 Database. *Mediterreanean and Black Sea cuise Inventory, Observed Data, Analysed Data and Climatological Atlas* (4 Cdrom*, in preparation).*

[10] MEDATLAS Group, 1997. *A Mediterranean Hydrographic Atlas from a composite Quality checked temperature and salinity data* (3 CD-ROM), IFREMER Ed..

[11] Rixen, M., Beckers, J.-M., Brankart, J.-M., and Brasseur, P. (2001). *A numerically efficient data analysis method with error map generation*. Ocean Modelling, 2(1-2):45-60.

[12] Schlitzer R., 2000 - *Electronic Atlas of WOCE Hydrographic and Tracer Data Now Available, Eos Trans. AGU*, 81(5), 45.

[13] Vyazilov E., Puzova N., 2001. *Oceanographic research vessel cruises for the Mediterranean and Black Sea*. RIHMI-WDC Rep.