MEDAS system for archiving, visualisation and validation of oceanographic data

V. Dadic and D. Ivankovic

Institute of oceanography and fisheries, Split Croatia, Mestrovicevo setliste 63, 21000 Split, Croatia
E-mail: dadic@izor.hr

Abstract

The Marine Environmental Database of the Adriatic Sea (MEDAS) serves for data management of different types of oceanographic parameters. The system was constructed to capture historical as well as real-time data and to make them available through the World Wide Web. A special subsystem was developed for validation of various oceanographic parameters using IOC protocols (IOC, 1993), as well as transforming data from/to different international formats.

Keywords: oceanographic database; historical and real-time data; data validation; data assimilation; numerical models.

Introduction

In the framework of the Croatian National monitoring programme of the Adriatic Sea MEDAS was developed using Oracle RDBMS, Java and ArcGIS tools. Its core is a relational database that includes general information on oceanographic data, responsible institutions, persons, etc. Further thematic subdatabases with measurement data related to physical and chemical oceanography, biology and fisheries also form part of the system (fig. 1). MEDAS has improved our management of oceanographic data, including archiving, dissemination, validation, visualisation and presentation of data through web pages in both text and graphical formats (Dadic *et al.*, 1995; Ivankovic, *et al.*, 2000).

A large volume of diverse oceanographic data has been stored in the database (more than a million records). As they have been measured during more than a century, their distribution in space (geographic position and layer in the water column) and time (year, season and month) is haphazard (Zore-Armanda *et al.*, 1991; Levitus and Boyer, 1998; MEDAR-MEDATLAS, 2002). Therefore, it was necessary to develop a procedure for data quality control and appropriate validation flagging prior to data processing.

Retrieving data from MEDAS is simple, so the user can retrieve any data in the following two steps:

- basic information about measured parameters (what, where, when, who, etc) can be obtained by searching the relational database

- through a connection to the thematic databases to get information through various menus. Depending on permissions, the user can access, insert or update records.

Creation of the database

The development of databases related to marine research is a specific task, in the first place because of the special requirements of oceanographic science. Users need to have easy and straight-forward access to the data, their derived parameters and their practical application. The basic considerations for the creation of marine environmental databases and secondary programmes are as follows:

- The majority of data have spatial and temporal components. The temporal component is simple to query (query period from-to), but for the spatial component it is necessary to develop a map interface to select data for analysis of a particular area
- The possibilities of research in a particular area depend on the quality of the data. To reach high-quality data, both human and material resources are needed. Data must be well protected from corruption and misuse, but without hampering ease of access for users
- While collecting data, it is possible to make different mistakes at various points. Hence, it is very important to develop constantly improved qualitative validation of data that will point to possible mistakes, hereby leaving the possibility for a final decision by the user
- To report on collected data, different processes and procedures have been used. It is very important to store these constantly changing procedures into a database. That way faster reporting can be achieved, and more than one model can be applied at the same time
- To present the results from the analyses it is necessary to develop a graphical presentation in order to better show the interdependence of different factors considered. The results have to be presented in their spatial context
- The database interface must be simple, accessible and, if is possible, platform-independent. Connection trough the Internet is a big advantage, and at present becomes imperative.

For all the above-mentioned points, it is obvious that it is a complex task that requires a multidisciplinary approach. It is often a never-ending process, and it demands constant step-by-step improvement to increase the usability of the database. It requires frequent feed-back from the end-user: there is the possibility of not having satisfactory results due to a too wide and theoretical approach.

Tools and technologies

The MEDAS database and web interface are based on Oracle 9i RDBMS and Application server. Java applets are used as mapping tools and for data visualisation. Special applications for automatic inserting and processing of real time data were

developed in the C++ program language. Oracle Graph and Matlab are used for data visualisation. The rationale behind these choices was as follows:

- Oracle 9i: Stability, security; easy installation and migration
- Oracle Application Server: no need for client side software (only browser), accessibility (different institutions, ship – GPRS), easy web publishing (default)
- Java applet – mapping tool and data visualisation: portability (cross platform), use of client-side resources (no need for powerful server)

These tools also have some disadvantages, such as software license cost, instability and bugs of some Java versions and sometimes need for manual JVM installation. In addition, the interface for transcoding data to ArcGIS 8.1 input for a spatial presentation of data was developed.

Database design

Database design is a critical stage in the whole process of database development since the performance of the database depends upon its design. All required database objects should be carefully planned and co-ordinated at the beginning of the development. All the previous experience from that field is very welcome.

As it can be seen in figure 1, the main entity is 'measurements': the chart contains the facts about time and depth, while the other facts are gained from other entities. The measurement results are aligned according to smaller entities, into sub-tables that are referred to the measurement table. Figure 1 shows the entities so far. Further expansion of the number of entities (tables) with concrete data for other types of measurements (heavy metals, biochemistry, etc.), is planned.
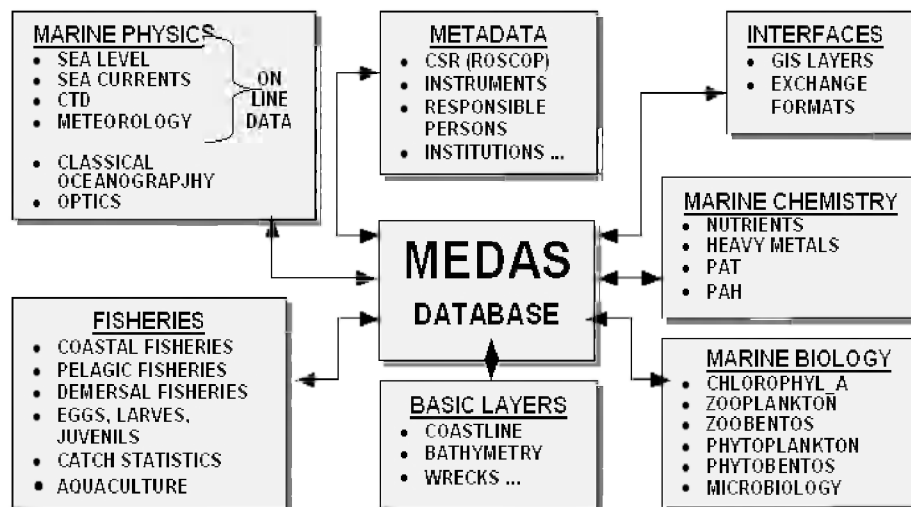

Fig. 1. The most important objects in database.

Database capabilities

The MEDAS system is used in the framework of the Croatian national monitoring programme (Systematic Research of the Adriatic Sea as a Base for Sustainable Development of the Republic of Croatia), in which various Croatian research institutions are involved. The database is used for storage, validation, and presentation of real-time measurements, and for running circulation models. In addition, the database is used for near real-time processing of data from research vessels. Visualisation and mapping tools are important parts of the database for both quality control and web presentation of the data. Various types of import and export formats are very useful for data sharing and processing. Figure 2 shows the main groups of MEDAS functionalities.
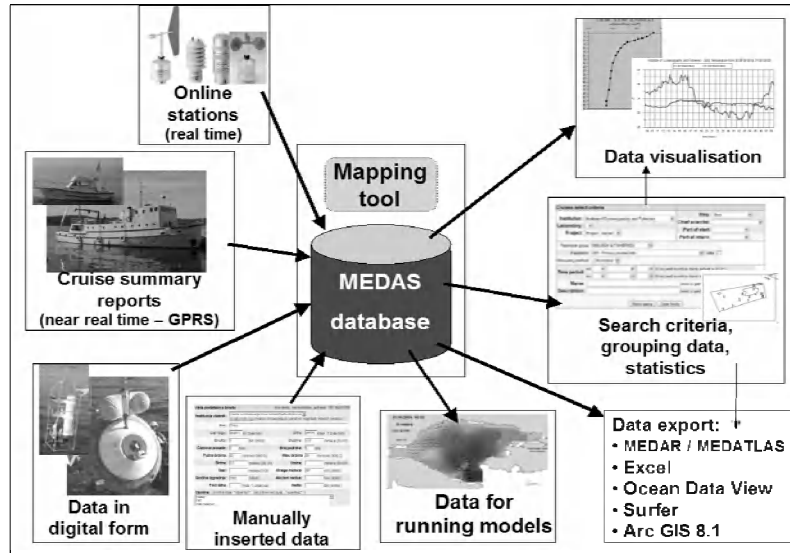

Fig. 2. Database functionalities.

CSR (Cruise summary report)

The CSR web interface includes forms for data input and also a java applet as mapping tool. The objective is to develop a fully web-based database interface. The advantage of a web-based database interface is that is platform-independent, and only requires a browser and access to the Internet. With this facility we can insert CSR data directly from research vessels (GPRS, satellite internet connection) and from various different institutions. The system contains procedures for automatic checks of CSR data (time-spatial crosscheck). All CSR data are available online (http://www.izor.hr/roscop/eng/) with detailed information about time, station position, measured parameters, responsible persons, projects, institution and ship details. Cruise information includes a mapping tool that is implemented using frames with a form – applet interaction. CSR data are linked with measurements; together with an interface for authorisation, this system is used for data quality control and processing.

Real-time data

The real-time part of MEDAS includes programmes for automatic upload of data from real-time meteo-ocean stations, and an automatic visualization and Internet publication. On a daily basis, model results are automatically calculated and published on the Internet. The real-time part is available online at http://www.izor.hr/eng/online/. Figures 4 to 7 shows some web examples of real-time and aggregated data.

Real-time data are used as input for circulation models. For example, the model assimilates data from Punta Jurana online station. The model is initiated with hydrostatic conditions; after that, the results from any day serve as initial conditions for the following day. Visualization of model results is generated by a C++ application. The first application generates a dynamic Matlab script, which is run, to dynamically create the html pages. Some examples are shown on figures 8 to 10.
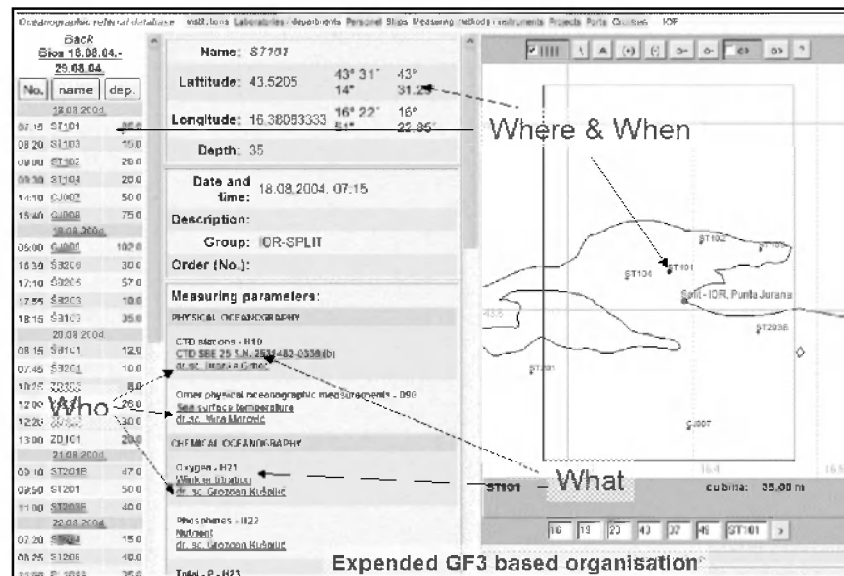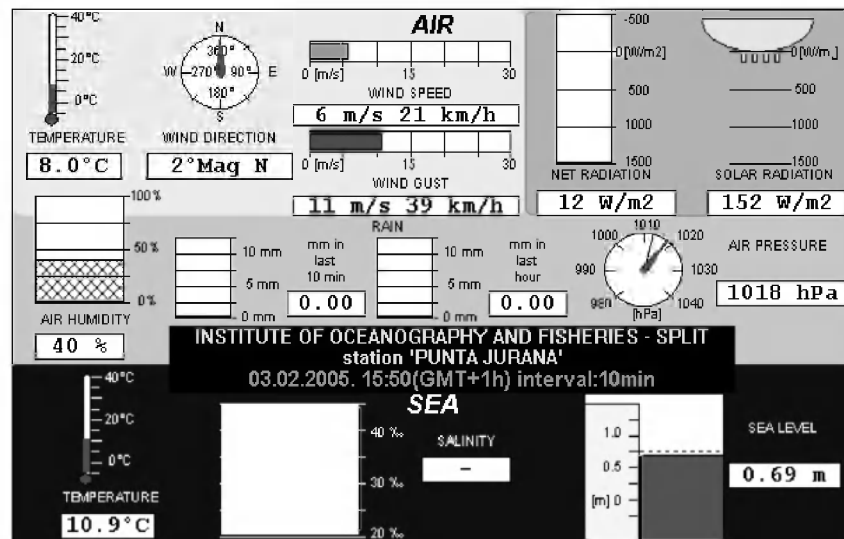

Fig. 3. Example of extended on-line CSR (browser).

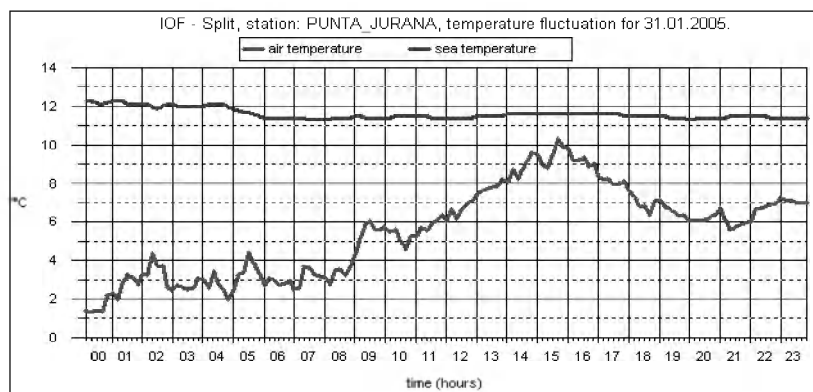*Fig. 4 Visualisation of real-time data.*

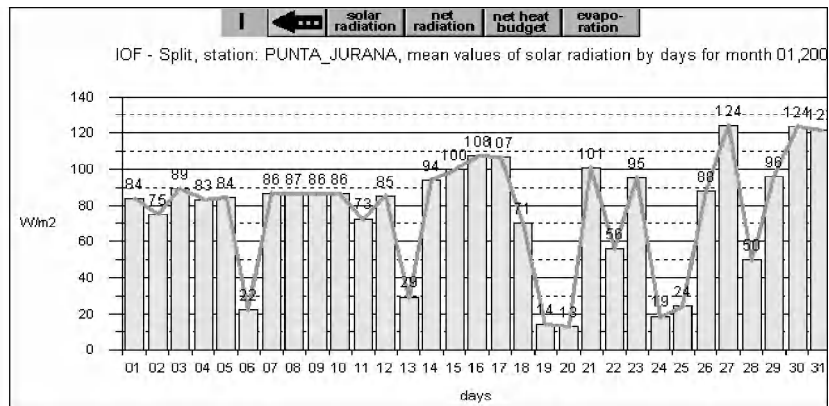*Fig. 5. Graphical presentation of 24h air pressure and humidity fluctuation.*

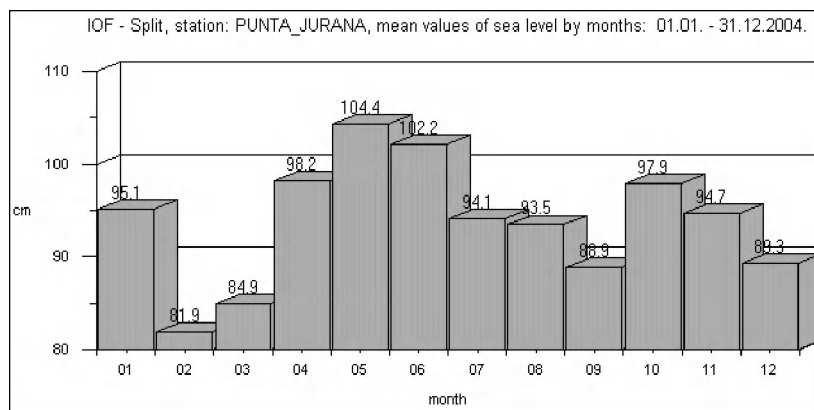*Fig. 6. Monthly mean solar radiation per day.*

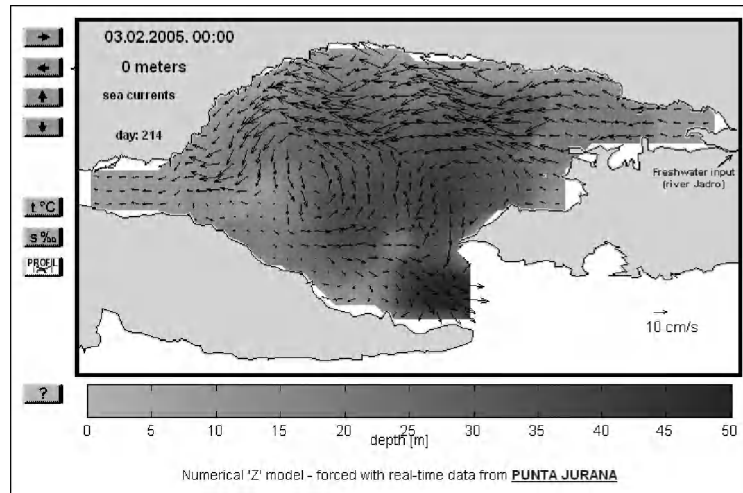*Fig. 7. Yearly mean air temperatures per month.*

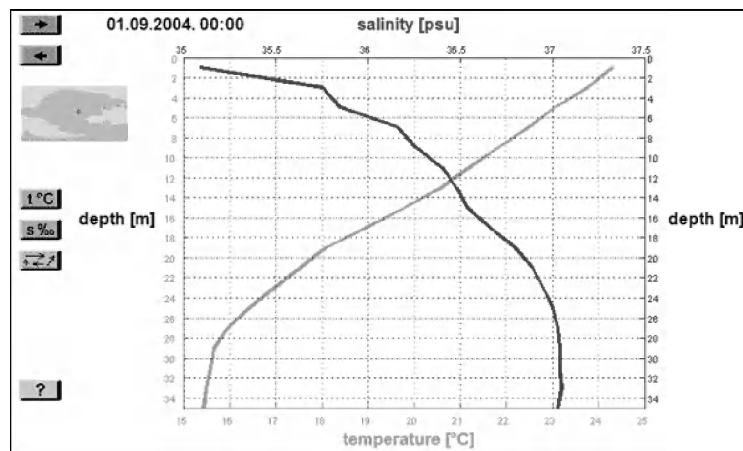*Fig. 8. Example of model results (sea currents).*

Fig. 9. Example of model results (temperature – salinity profile).

Tools for data validation

Oceanographic data stored in MEDAS are randomly distributed in space (geographic position and layer in the water column) and time (year, season and month). Therefore, it was necessary to calculate monthly and seasonally climatological values of each parameter (mean value, minimum, maximum and standard deviation) at standard oceanographic depths and specific marine areas, which most often are defined as squares. As there is a large amount of data in our database, manual validation is practically impossible. Therefore, based on protocols of IOC-UNESCO (IOC, 1993) and World Data Centre (Boyer and Levitus, 1994) a procedure was defined, which includes comparison of different parameters and an iterative method following several steps:

- Calculation of monthly and seasonal means and standard deviations from all data for each parameter at the same measuring level for the entire marine area
- Comparison of each reading with the calculated mean value and its exclusion from processing if the difference exceeds a predefined range (from 1 to 5 standard deviations, depending parameter and entire area)
- Interpolation of values from observed levels to standard oceanographic levels
- Comparison of each interpolated value with the calculated mean value and its exclusion from processing if the difference exceeds a predefined range (from 1 to 3 standard deviations depending parameter and entire area)
- Visual check of profiles based on interpolated values at standard levels
- Estimating interpolated data in square nodes of geographic grid
- Presentation of output results in graphic form and check for 'bull's eyes'
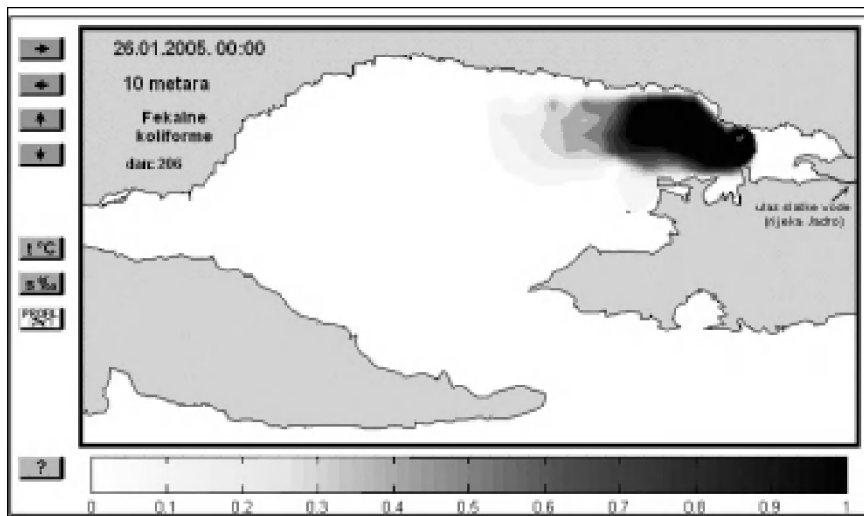- Repeating procedure if necessary

Fig. 10. Reconstruction of faecal coliform field in Kastela Bay obtained by Z-model with assimilation of various data from the station Punta Jurana.

Interpolation of values on standard oceanographic levels was done by third order Newton method of finite differences modified by referral curves (Reinger and Ross, 1968, Dadic *et al.*, 2002). This method is specially recommended for oceanographic parameters sampled at the discrete oceanographic levels (*e.g.* by water samplers and nets). Ordinary kriging method and semi-variogram were used for spatial analysis of data (Journel and Huijbregts, 1995).

Climatologic values derived by above procedure serves for validation of new data received in the database. This process cannot be fully automated and expert opinion is needed to decide on the quality of each datum after checking by visualisation. Depending on quality, each datum is assigned a quality flag (MEDAR-MEDATLAS Group, 2002).

As the main objective of any spatial investigation is to simplify the analysis of spatially-distributed data to end users, a special interface between Oracle and ArcGIS 8.1 tool was developed. This system enables overlaying layers of different parameters and their presentation through maps. Based on these maps data validation and various analyses of oceanographic properties of the Adriatic Sea have been possible.

Some results of data analysis

As a result of data analysis using the MEDAS system, many duplications, uncertainties and erroneous historical data were identified. For example, about 49.7% of BOT data were duplicated, 3.2% outside climatologically range, 0.7% of oceanographic stations were attributed to wrong position, and about 17.4% of BOT data attributed as MBT data.

Based on the data analysis, four different sub-regions of the Adriatic Sea with similar oceanographic properties were recognised and 41 standard oceanographic levels defined as suitable for climatological analysis.
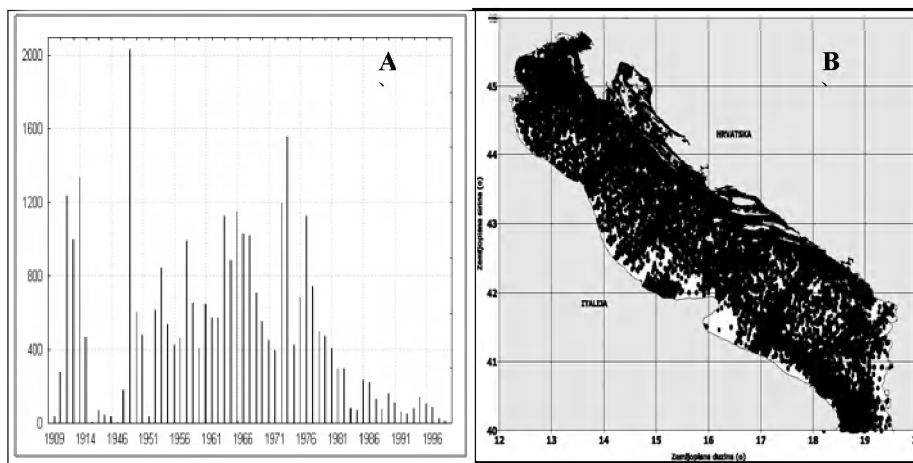

Fig. 11. Number of measured temperature profiles by year (A) and spatial distribution of station in the Adriatic Sea (B).

After validation and harmonisation of data, numerous unique and corrected data profiles were included in statistical calculations, *e.g.* more than 100,000 temperature readings. The importance of validation procedures and harmonisation of classical oceanographic observations is shown in figure 12. The temperature spatial field gathered from original data looks very artificial, but looks much more realistic when calculated from partly validated data. Many of the unrealistic features, like very high gradients and bull's eyes, disappeared from the objective analysis after the data validation.

Based on climatology analyses of more oceanographic parameters (temperature, salinity, oxygen, nutrients, chlorophyll_a), 41 standard levels and two maximum depth difference criteria (inner and outer) were defined for each standard level (Dadic *et al.*, 2002).

As there was anisotropy in spatially distributed data in transversal and longitudinal axes of the Adriatic Sea, different radius influence and grid mesh were used. So, influence radius of 7.5km and interpolated data on a 5 x 5km grid were used in transversal direction of the Adriatic Sea and in the coastal area and influence radius of 12km and 7.5 x 7.5km grid in longitudinal direction at the open sea during kriging interpolations.
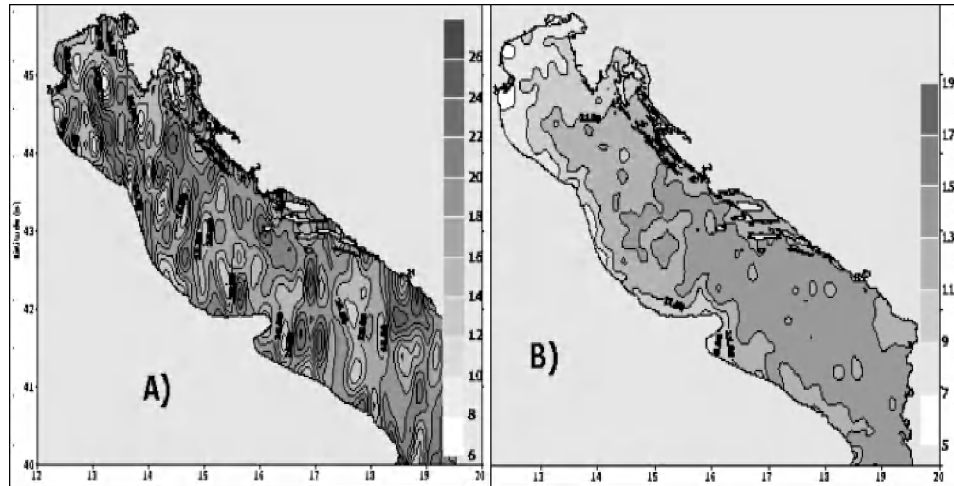

Fig. 12. Interpolated field of mean value of temperature at the sea surface gathered from the original data (A) and data passed quality control (B).

References

Boyer T. and S. Levitus. 1994. QC and processing of historical oceanographic temperature, salinity and oxygen data. SilverSpring. USA-NOAA technical report NESDIS, No. 81, 112p.

Dadic V., M. Srdelic and A. Grubisic. 1995. Oceanographic information system and databases of the Adriatic Sea. *Proc. of 17th International Conference ITI'95*, Pula, 289-294.

Dadic V., M. Srdelic and Z. Grzetic. 2002. Analysis of oceanographic properties of the Adriatic Sea by GIS technique. Cartography and information 1:46-59.

IOC. 1993. Manual of Quality Control Procedures for validation of Oceanographic Data, prepared by CEC-DG: XII, MAST and IODE. Paris. IOC (UNESCO) Manuals and Guides NO. 26:436p.

Ivankovic D., V. Dadic and M. Srdelic. 2000. Marine environmental database of the Adriatic Sea with application for managing and visualisation of data. ENVIROSOFT International Conference: Computer techniques to Environmental Studies VIII, 28-30 June 2000, Bilbao, Spain. Proceedings:398-406.

Journel A.G. and Ch.J. Huijbregts. 1993. Mining geostatistics, Academic Press, New York-London. 600pp

Levitus S., and T. P. Boyer, 1998. World ocean atlas 1988. SilverSpring. NOAA Atlas NESDIS 4, 117 p + 7 CD.

MEDAR-MEDATLAS Group. 2002. Mediterranean and Black Sea database of temperature, salinity and bio-chemical parameters climatological atlas. EU-MAST Programme. Brest. IFREMER Ed. 4 CD.

Reinger R.F. and C.K. Ross. 1968. A method of interpolation with application to oceanographic data. Deep-Sea Research 15:185-193.

Zore-Armanda M., M. Bone, V. Dadic, M. Morovic, D. Ratkovic, L. Stojanovski and I. Vukadin. 1991. Hydrographic properties of the Adriatic Sea in the period from 1971 through 1983. Acta Adriat. 32 (1):1-547.