

Database and Database Management System of the TU-Black Sea Project

V.L. VLADIMIROV¹, S.T. BESIKTEPE², D.G. AUBREY³

¹ Marine Hydrophysical Institute, NASU, 2 Kapitanskaya St., Sevastopol,

335000 Ukraine

² Middle East Technical University, Institute of Marine Sciences, P.O.

Box 28, Erdemli, Icel, Turkey

³ Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

Abstract. Unique regional historical interdisciplinary multipurpose database was created in the framework of the NATO TU-Black Sea Project. Database includes all main physical, chemical and biological (including plankton) variables for the entire Black Sea basin. The database spans the most crucial period in the history of the Black Sea ecosystem starting from the "background" situation in 1960-th and over the drastic changes that occurred in recent years. The database is supplied with a powerful specially designed database management system.

1. Introduction

The purpose of the TU-BLACK SEA project is to develop and implement ecosystem models as a management tool for improving the health of the Black Sea, and to help to build the capacity within the Black Sea coastal countries for high quality oceanographic measurement and analysis as well as capabilities for carrying out long-term monitoring of the Black Sea following the completion of the project.

One of the main project objectives is to establish a database management system (DBMS) in all Black Sea countries, for environmental and oceanographic data pertinent to the goals of this Project.

To coordinate data base development a Working Group had been established with participation from all collaborators. This group followed through the implementation of this subproject. The group defined data base requirements and selected data base management system and computer platform. Once the DBMS is established the other tasks of this group were to collect contaminant input data, focusing on nutrients and eutrophication, and to coordinate the orderly and timely input of requisite data into the DBMS. This working group coordinated with national as well as regional efforts sharing at least a portion of this work.

The Working Group had its first meeting in Moscow during 22-23 November, 1993. The objectives of the meeting were: a) to familiarize the group with the TU-BLACK SEA project, b) to make sure that the responsibilities are understood with

regard to the identification of data base requirements, c) to discuss issues related to the selection of the data base management system and computer platform, and d) to make preliminary plans for procurement and implementation of DMBS, collection of contaminant data with special focus on nutrients leading to eutrophication.

It was recommended that the data included in the data base be restricted to the period from 1963 to present (approximately 30 years). This period of time encompasses the major ecosystem shifts in the Black Sea. It was also recommended that data entry and preliminary quality control procedures would occur at the country of origin of data. In order to establish a data inventory, forms, specially designed in accordance with project requirements, had been sent in April, 1994 to major cooperation partners.

Special Data Base Task Team was nominated by the Working Group in 1995 for the compilation of the Project database and database Inventory.

2. Data Base Inventory

The inventories from all the countries were completed by February, 1995. These inventories were accumulated and edited and the "Data Base Inventory of Time Series and Station Data" was prepared by the special subgroup of the Working Group 1 in February, 1995.

At the December 1995 Working Group 1 Task Team meeting it was decided that a Computerized Data Base Inventory (CDBI) should be developed and distributed among the participants instead of the printed version, to provide easy and comfortable access of users to the information collected in the inventory. The first version of the CDBI was designed and distributed to all participants in September 1996. It contains information about the datasets delivered to the database and about the datasets which are still not digitized or not delivered to the database due to different reasons.

The CDBI was developed on a basis of the special database management system for the oceanographic inventories previously designed in the Marine Hydrophysical Institute (MHI). It has the Windows-like interface and provides the following general features: running under both DOS and Windows; adjusting to hardware configuration; customizing screen view by user preference; has a context-sensitive hyper-text help system; has a high speed performance.

The CDBI provides next capabilities:

- map drawing:
 - selection of any geographic region in the Black Sea;
 - selections of any data set or station on the map;
 - display statistical information on data coverage on the map;
 - printing and saving a map to bitmap picture file;
 - customizing map view;
- data selection from internal data base for next searchable fields:
 - country;
 - institution;
 - name of research vessel;
 - year;

- month;
- data storage media;
- type of measured data;

- access to all information from "Data Base Inventory of Time Series and Station Data" and creation of brief reports on data selection.

It is a powerful tool that provides easy and comfortable user access to the information collected in the Inventory. Now CDBI includes information on the 377 data sets.

3. Data set formats

It was decided to develop simple formats for the program data sets to provide possibility to easy and comfortable data digitization by originators. Those formats were developed by the Working Group.

According to this decision, all information and data of the Program data sets have to be prepared using three types of files described below (Information File, Brief Description File, Data File).

Information File. This file contains the detailed description of the data set in the form of technical note with information on the recommended items.

Brief Description File. All the data files prepared for the NATO TU Black Sea Project are accompanied by brief description files containing concise information on their origin, instrumentation, method, units, data processing and other topics. Each item in a file starts from the name of the item prefixed by an asterisk and followed by a colon. Brief description file has the same name as the data file(s) or data directory if it relates to all data files in the directory.

Data File. Each data file contains data themselves according to one of the three formats (station data file, containing vertical series of measurements; header data file, containing observations on sea surface; and time series data file, containing time series at distinct points).

4. List of principal ecosystem variables included in the data base

List of principal ecosystem variables is a very important characteristic of the data base. It has to cover all main requirements of the Program. The first version of the list of variables to be included in the data base was prepared by Working Group in November 1993. Then it was adopted and updated in February 1994 by Coordination Meeting and in March 1994 by Workshop on Ecosystem Modeling of the Black Sea. The new revised list of the biological variables was prepared by the Working Group meeting, in January 1995.

The variables and groups of variables that were loaded into the final database:

- **Physics:** Temperature, Salinity (CTD-profiles, Nansen bottles, and CTD-thermograph), Fluorescence, Light attenuation (407nm, 422nm, 427nm, 457nm, 487nm, 540nm, 660nm), Photosynthetic active radiance (PAR), Secchi disk depth.
- **Chemistry:** Dissolved oxygen, Hydrogen sulfide, Thiosulfate, Sulfite, Sulfur-oxygen, Phosphate, Nitrite, Nitrate, Total amount of Nitrite+Nitrate, Urea, Ammonia, Total Alkalinity, Carbonate Alkalinity, pH, Particulate organic carbon, Total organic carbon, Particulate organic nitrogen, Total nitrogen, Dissolved organic phosphorus, Particulate phosphorus, Total phosphorus, Humic matter, Total suspended matter, Total manganese, Manganese, Total iron.
- **Biology:** Chlorophyll-a, Phaeopigments-a, Primary production, Bacterioplankton, Benthos, Microphytoplankton, Nanophytoplankton, Macrophytoplankton, Total phytoplankton, Macrozooplankton, Mesozooplankton, Ctenophora, Total zooplankton.

Other data that were included only in the ASCII data set:

- Climatic and averaged values of the variables.
- Meteorological data.
- Sea level.
- Spectra of transparency.
- Currents.

Data Quality Control

The importance of strict data quality control was emphasized by the Working Group. Data quality control for the data set, as agreed, took place on two levels: the originating institution provided quality control, and the Task Team prepared data quality control, as follows.

Originating Institution Data Quality Control. The following two methods for data quality control were strongly recommended:

- 1) Double data entry can be used, where each data point is entered by different data entry technicians independently. The data are then compared by computer automatically, to examine for mistakes. Mistakes have to be corrected.
- 2) The investigator who is the source of data can prepare the data file, and check each entry manually against the source to assure that no errors exist. Principal investigators also check the overall quality of provided data sets.

Task Team Quality Control: The following two steps had been taken by the Task Team as part of the overall quality control program:

- 1) All data sets were evaluated by the Task Team, by comparing all data and identifying outliers and trends. This automated evaluation provided some level of quality control on data entry, but not on overall data quality.

2) Data sets were evaluated by Task Team members who are experts in that particular discipline. Biological, Chemical, and Physical data had all been evaluated this way, and results of this quality evaluations will be published in technical reports.

6. Expert Groups

The decision was made by the Working Group during its January 1995 meeting that "When and if a valuable data set on any type of characteristics are delivered, the Task Team will meet in Erdemli together with a group of experts to merge and make quality control of the combined data set, and produce a technical report." A series of Expert Groups had therefore been identified. It was agreed that the Expert Groups include the scientists that participated in collecting and primary processing of data subsets delivered to the database. Five Expert Groups were formed according to the characteristics of the data sets:

1. Physics
2. Chemistry
3. Biology 1 (zooplankton)
4. Biology 2 (phytoplankton)
5. Biooptics (chlorophyll, PP, fluorescence, light attenuation, PAR)

Each Expert group met two times for two weeks. The first meeting was held in October-December 1996 and the second one was in January-March 1997. They evaluated all delivered data and metadata, put the quality flags for physical, chemical and biooptical data and prepared technical reports.

7. Software for data quality control

The CruBase DBMS [1] designed in the MHI for the oceanographic data and intended for interdisciplinary multiparameter data of an oceanographic cruise or a multiship survey was used by the Task Team and Expert Groups for the quality control and preparation of the final data set.

To provide the possibility of the effective quality control of the data loaded into the CruBase system, special subsystem of the CruBase DBMS was developed (CruBase Data Control System) and improved during the Task Team meetings. The format of the CruBase files was updated. Now this format permits to store quality flag with each data value. During loading the quality flag 0 is assigned to each data value (data was not checked).

There are many schemes used to flag data quality [2]. Many of them are acceptable. We have chosen the scheme employed for data reported in real-time, as given in the GETADE Formatting Guidelines for Oceanographic Data Exchange [3], devised by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE). This scheme is universal and can be applied to all types of data in the Project database. It uses a one character field with the following interpretation:

- 0 = data are not checked
- 1 = data are checked and appear correct
- 2 = data are checked and appear inconsistent but correct
- 3 = data are checked and appear doubtful
- 4 = data are checked and appear to be wrong
- 5 = data are checked and the value has been altered
- 6 = data are checked and appear to be assigned to the wrong depth

Flag 6 was added according to request of the Expert Groups.

8. Structure of the data base.

A work on the compilation of the first version of the NATO TU-Black Sea data base (presented in the ASCII files) was finished in April, 1997. For this purpose, the data sets delivered by participants were preliminary checked, transformed to the NATO TU-Black Sea formats and were included in the joint data set. The structure of this data set was chosen to be in common with the structure of the data base inventory and most part of data received from participants are arranged according to country and institution of their origin. Brief description and information files accompany each data file (or data directory) in cases where such information was provided by the participants. A map with the positions of stations included into the database is shown in Figure 1.

At present data base includes **12,790 files (149.4 MB)**. All physical and chemical data sets were loaded into the CruBase system for quality control, which was done by the Expert Groups.

The final project database contains the following parts:

- Main TU-Black Sea Data Base (loaded into the OceanBase-II system, quality checked),
 - Description of the Data Base and OceanBase system,
 - Technical reports of the Expert Groups and Task Team,
 - Raw data sets as they were delivered by participants,
 - Climatic and averaged data,
 - Additional and special data sets,
 - Intercalibrated data sets of the recent joint Black Sea cruises.

9. Comparison with other data bases

Brief information is given below on some well-known recent regional and global data sets to compare them with the TU-Black Sea Data Base.

The North Sea Tidal Data CD-ROM is a pilot project within the MAST Program of the Commission of European Communities. The CD-ROM contains tidal data (waterlevels, currents and sets of tidal constants) of the North Sea, with emphasis on data from the 10-years period 1984-1993. The data have been collected from relevant data sources in Netherlands and by a search of international gray literature. All data have been quality controlled and formatted.

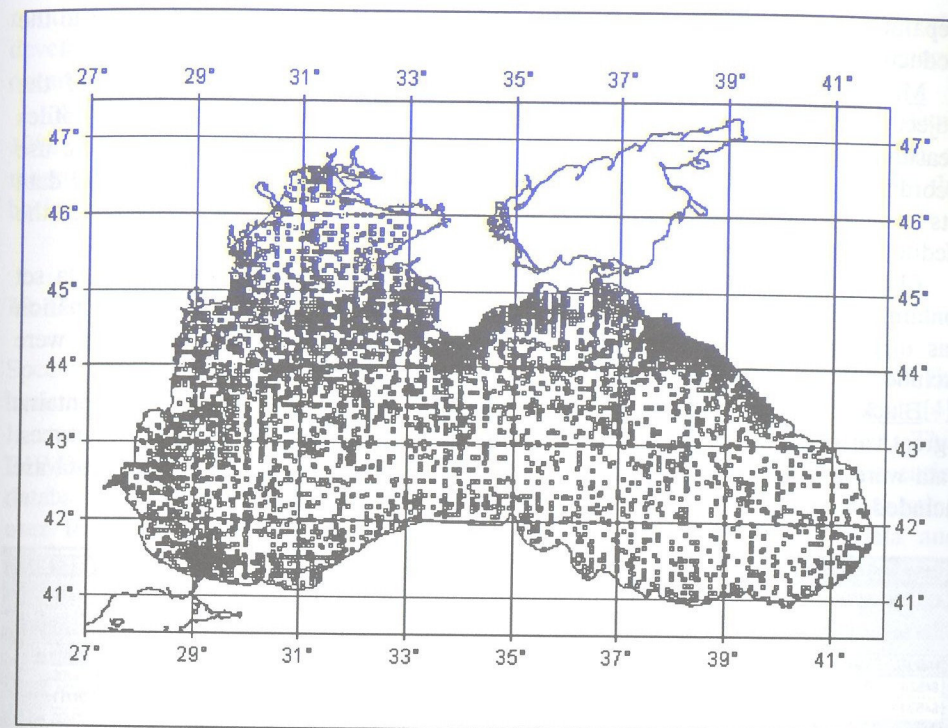


Figure 1. Positions of stations included into the database (the picture was obtained using one of the OceanBase possibilities).

NERC North Sea project Data Set CD-ROM was created by British Oceanographic Data Center in 1992. It contains the multidiscipline data set for the region for the 1987-1992 period.

The BOFS CD-ROM (the Biogeochemical Ocean Flux Study) contains 98% of all data series collected during the BOFS field programme (1989-1993) in the North Atlantic and covers a great diversity of biological, chemical and physical measurements both in the water column and in the air-sea and benthic boundary layers.

World Ocean Atlas CD-ROM Series 1994. The Ocean Climate Laboratory (OCL) at NODC is supported by NOAA to produce scientifically quality controlled databases of the world ocean. Work to date includes quality control of historical in situ temperature, salinity, oxygen, phosphate, nitrate, and silicate data and the preparation of one-degree latitude-longitude mean fields for each of these parameters. Enclosed and semienclosed seas are of the lowest priority for this project.

MODB. The Mediterranean Oceanic Data Base. The general objective of MODB is to deliver advanced data products to the Mediterranean research projects supported by the MAST programme of the European Union. A special effort is being devoted to the

preparation and archiving of quality checked *in situ* hydrographic data sets, and to the production of gridded climatological fields for temperature and salinity.

MEDATLAS. Mediterranean Hydrographic Atlas. The main objectives of the project are: (1) To update the available data sets of temperature and salinity profiles measured in the Mediterranean Sea. (2) To check the data quality for scientific use according to the IOC and MAST recommendations. (3) To merge the compiled data sets and update the climatological statistics with a resolution adapted to the Mediterranean space scales.

Black Sea Hydrographic Data Set of the Moscow State University. This data set contains only temperature and salinity data for the 1900-1990. No header information was included except the date and coordinates. Only evidently erroneous data were excluded from this data set.

Black Sea Data Base of the Marine Hydrophysical Institute. This data base contains a great variety of physical and chemical data from the MHI archive and other sources. Data were passed through some stages of quality control. None of biological data are included.

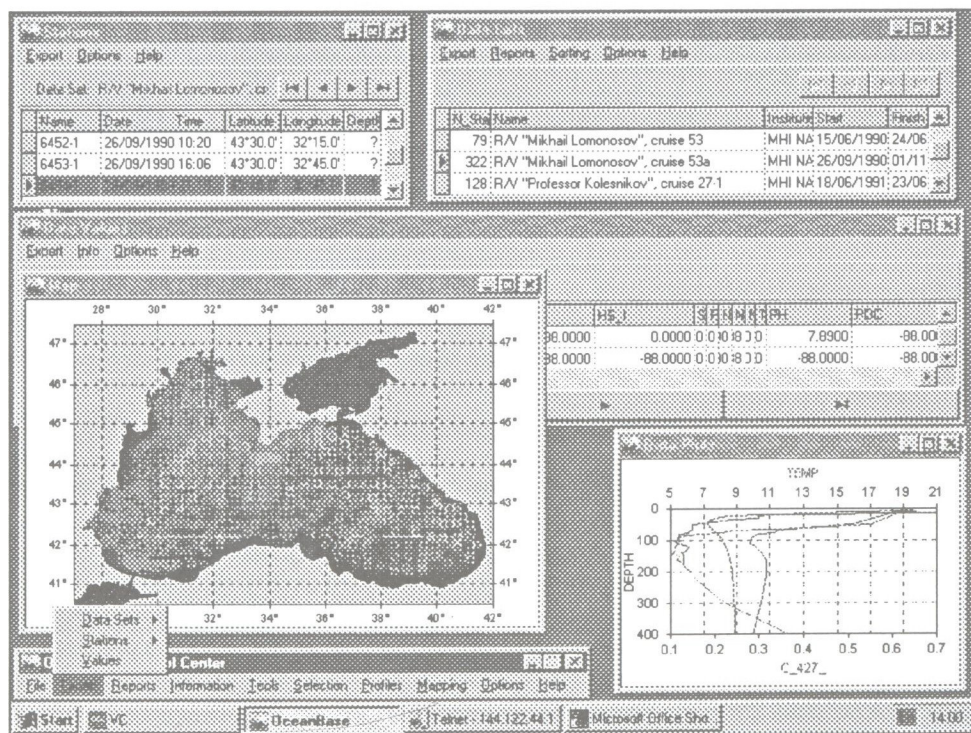


Figure 2. One of the OceanBase-II working desktops created by user

Comparing the TU-Black Sea Data Base and these data bases, one can conclude, that it is the first successful attempt to create a regional historical interdisciplinary

multipurpose database supplied with the special powerful DBMS. To our mind, the development of TU-Black Sea Data Base is a unique event in modern regional oceanography.

According to our estimation, the TU-Black Sea database includes now from about 10% (e.g., temperature and salinity) to 30% (e.g., transparency) of all existing data of measurements performed in the Black Sea since the beginning of Ocean Science in the basin.

10. OceanBase-II DBMS.

Special unique DBMS (OceanBase-II) was developed for the project database on the basis of the OceanBase DBMS system developed by the MHI Database Laboratory [4]. It was designed using Borland Delphi Developer to run under Windows-95. This DBMS provides a lot of possibilities for quick and comfortable work with the entire database. One can view, sort, select and export all necessary data and metadata using user friendly interface. User can open as many windows as he want. Data and information in these windows are synchronized. For example, clicking at the station position on the map, user obtains immediately the cursor placed at the corresponding dataset name in the "Data Sets" window, the cursor placed at this station line in the "Stations" window, values of data at this station in the "Values" window, etc. One of the OceanBase-II working desktops created by user is shown in Figure 2.

11. Conclusions

The main features of the TU-Black Sea Data Base are as follows:

- Database includes all main physical, chemical and biological (including plankton) variables for the entire Black Sea basin;
- It primarily covers the time period from 1963 with extensive data sets for 1973-1994 (and in some cases earlier data sets) and it includes data obtained at 26,000 oceanographic stations (Fig. 1.);
- Database covers the most crucial period in the history of the Black Sea ecosystem starting from the "background" situation in 1960-th till the drastic changes occurred in recent years;
- It includes data from all main regional and international sources;
- All data included into the database were quality checked by the qualified groups of regional experts, well acquainted with the Black Sea data;
- Each value of physical, chemical and biooptical data are accompanied with the quality flag (the same as in the MEDATLAS project, that provides the regional compatibility);
- Special and powerful DBMS has been designed for the database to provide easy and comfortable access to data;

- Although the main objective of the database is to serve the development and implementation of the ecosystem models, due to its structure and contents it can be used for many other oceanographic and environmental purposes;

- It includes also the initial set of the ASCII files delivered by participants, that provides the possibility to an interested person to select the subset, to perform the quality control according to other principles and to create an alternative database;

- Database will be delivered to participants on the CD-ROM.

According to these features, the database of the TU Black Sea project can be used as a basis for the databases of any new oceanographic and environmental projects in the Black Sea region.

Acknowledgments. We would like to thank Vladimir Miroschnichenko and Vyacheslav Lyubartsev for the titanic work on creation of the new software. We wish to thank also all members of the Task Team: Dr. A. Mishonov, N. Kapustina, V. Tsiganok, L. Galkovskaya; heads of the Expert groups: L. Ivanov, S. Konovalov, V. Melnikov, A. Mikaelyan, and O. Yunev; all members of Expert groups and all scientists, who performed the huge work on the data base compilation and data quality control. Our special thanks to Professors Umit Unluata and Valery Eremeev for their permanent support of this work.

References

1. Vladimirov, V.L. (1992) Integrated data bank for the cruise data of the research vessel, in *Automated systems for the monitoring of the marine environment*, Marine Hydrophysical Institute, Sevastopol, pp. 126-131 (in Russian).
2. UNESCO (1993) *Manual of Quality Control Procedures for Validation of Oceanographic Data*. Prepared by: CEC: DG: XII, MAST and IOC/IODE. Manual and Guides No.24.
3. ICES/IOC (1995) GETADE Formatting Guidelines for Oceanographic Data Exchange. Preliminary report.
4. Vladimirov V.L., Miroschnichenko V.V. (1997) Multipurpose database management systems for the marine environmental research, in Nilgun B. Harmancioglu, M. Necdet Alpaslan, Sevinc D. Ozgul and Vijay P. Singh (Eds.), *Integrated Approach to Environmental Data Management Systems. Proceedings of the NATO ARW*. Kluwer Academic Publishers, Dordrecht, pp. 355-364.

PHYSICAL, CHEMICAL AND BIOLOGICAL BLACK SEA DATA BANK

IVANOV

A. MIKAELYAN

O. BAŞTUĞ

N. BODE

KAMBUR

N. NEZLİ

E. SHUSKIN

V. VEDER

¹ Marine Institute

335000,

² Institute

335011,

³ P.P. Shchegolev

117851,

⁴ Institute

⁵ Marine

Sevastopol

⁶ Romania

⁷ Institute

⁸ Southern

Gelendzhik

Abstract. The data set of the Data Bank for the Black Sea, which data are freely accessible, is presented in the paper constitutes a basis for the development of the (physical, biogeochemical and biological) coverage as well as of satellite data.

1. Hydrographic data of the Black Sea

1.1 GENERAL INFORMATION