NOAA Technical Report NESDIS 87

PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON OCEANOGRAPHIC BIOLOGICAL AND CHEMICAL DATA MANAGEMENT

Washington, D.C.
February 1997

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service

NOAA TECHNICAL REPORTS

National Environmental Satellite, Data, and Information Service

The National Environmental Satellite, Data, and Information Service (NESDIS) manages the Nation's civil Earth-observing satellite systems, as well as global national data bases for meteorology, oceanography, geophysics, and solar-terrestrial sciences. From these sources, it develops and disseminates environmental data and information products critical to the protection of life and property, national defence, the national economy, energy development and distribution, global food supplies, and the development of natural resources.

Publication in the NOAA Technical Report series does not preclude later publication in scientific journals in expanded or modified form. The NESDIS series of NOAA Technical Reports is a continuation of the former NESS and EDIS series of NOAA Technical Reports and the NESC and EDS series of Environmental Science Services Administration (ESSA) Technical Reports.

A limited number of copies are available by contacting Nancy Everson, NOAA/NESDIS, E/RA22, 5200 Auth Road, Washington D.C., 20233. Copies can also be ordered from the National Technical Information Service (NTIS), U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, VA. 22161, (703) 487-4650 (prices on request for paper copies or microfiche, please refer to PB number when ordering). A partial listing of more recent reports appear below:

NESDIS 31 Data Processing Alogorithms for Inferring Stratospheric Gas Concentrations from Balloon-Based Solar Occultation Data. I-Lok Chang (American University) and Michael P. Weinreb, April 1987. (PB87 196424)

NESDIS 32 Precipitation Detection with Satellite Microwave Data. Yang Chenggang and Andrew Timchalk, June 1988. (PB88 240239)

NESDIS 33 An Introduction to the GOES I-M Imager and Sounder Instruments and the GVAR Retransmission Format. Raymond J. Komajda (Mitre Corp) and Keith McKenzie, October 1987. (PB88 132709)

NESDIS 34 Balloon-Based Infrared Solar Occultation Measurements of Stratospheric 03, H20, HNo3, and CF2C12. Michael P. Weinreb and I-Lok Chang (American University), September 1987. (PB88 132725)

NESDIS 35 Passive Microwave Observing From Environmental Satellites, A Status Report Based on NOAA's June 1-4, 1987, Conference in Williamsburg, VA. James C. Fisher, November 1987. (PB88 208236)

NESDIS 36 Pre-Launch Calibration of Channels 1 and 2 of the Advanced Very High Resolution Radiometer. C. R. Nagaraja Rao, October 1987. (PB88 157169/AS)

NESDIS 39 General Determination of Earth Surface Type and Cloud Amount Using Multispectral AVHRR Data. Irwin Ruff and Arnold Gruber, February 1988. (PB88 199195/AS)

NESDIS 40 The GOES I-M System Functional Description. Carolyn Bradley (Mitre Corp), November 1988.

NESDIS 41 Report of the Earth Radiation Budget Requirements Review - 1987, Rosslyn, VA, 30 March-3 April 1987. Larry L. Stowe (Editor), June 1988.

NESDIS 42 Simulation Studies of Improved Sounding Systems. H. Yates, D. Wark, H. Aumann, N. Evans, N. Phillips, J. Sussking, L. McMillin, A. Goldman, M. Chahine and L. Crone, February 1989.

NESDIS 43 Adjustment of Microwave Spectral Radiances of the Earth to a Fixed Angle of Progagation. D. Q. Wark, December 1988. (PB89 162556/AS)

NESDIS 44 Educator's Guide for Building and Operating Environmental Satellite Receiving Stations. R. Joe Summers, Chambersburg Senior High, February 1989.

NESDIS 45 Final Report on the Modulation and EMC Consideration for the HRPT Transmission System in the Post NOAA-M Polar Orbiting Satellite ERA. James C. Fisher (Editor), June 1989. (PB89 223812/AS)

NESDIS 46 MECCA Program Documentation. Kurt W. Hess, September 1989.

NESDIS 47 A General Method of Using Prior Information in a Simultaneous Equation System. Lawrence J. Crone, David S Crosby and Larry M. McMillin, October 1989.

NESDIS 49 Implementation of Reflectance Models in Operational AVHRR Radiation Budget Processing. V. Ray Taylor, February 1990.

NESDIS 50 A Comparison of ERBE and AVHRR Longwave Flux Estimates. A. Gruber, R. Ellingson, P. Ardanuy, M. Weiss, S. K. Yang, (Contributor: S.N. Oh).

NESDIS 51 The Impact of NOAA Satellite Soundings on the Numerical Analysis and Forecast System of the People's Republic of China. A. Gruber and W. Zonghao, May 1990.

NESDIS 52 Baseline Upper Air Network (BUAN) Final Report. A. L. Reale, H. E. Fleming, D. Q. Wark, C. S. Novak, F. S. Zbar, J. R. Neilon, M. E. Gelman and H. J. Bloom, October 1990.

NESDIS 53 NOAA-9 Solar Backscatter Ultraviolet (SBUV/2) Instrument and Derived Ozone Data: A Status Report Based on a Review on January 29, 1990. Walter G. Planet, June 1990.

Proceedings of the International Workshop on Oceanographic Biological and Chemical Data Management

May 20 - 23, 1996

Sponsors:
Intergovernmental Oceanographic Commission
U.S. National Oceanographic Data Center
European Union MAST Programme

Hosts:
Bundesamt für Seeschiffahrt und Hydrographie
Hamburg, Germany

TABLE OF CONTENTS

SESSION III: (continued)

SESSION IV: BIOLOGICAL AND CHEMICAL DATA MANAGEMENT

SESSION IV: (continued)

APPENDIX I

APPENDIX II

PROLOGUE

The International Workshop on Oceanographic Biological and Chemical Data Management held in Hamburg, Germany from 20-23 May 1996 was a result of the decision of the Fifteenth Session of the IOC Committee on IODE to convene a meeting in recognition of the role that historical, digital archives of oceanographic biological, chemical and carbon dioxide data play in understanding the World Ocean's biogeochemical cycles. It was recommended that the Workshop bring together representatives from both government institutions (including data centres) and academic communities. The Workshop concentrated on a few parameters to ensure that progress was made in understanding how best to manage this data.

The problems of archiving oceanographic data magnify when the scope of the archive extends through the geochemical to the bio-geochemical. The challenge is to develop the database, data analysis and data visualization structures which will enable widely distributed, multidisciplinary investigators to work with each other's data and to collaborate with each other. In view of the need for oceanographic biological and chemical data, and of the problems in managing these data, it was decided to convene an International Workshop on Oceanographic Biological and Chemical Data Management to discuss the issues involved and identify ways to solve existing problems. In part, the Workshop was an outgrowth of the Ocean Climate Data Workshop held in Greenbelt, Maryland, USA during February 1992. The OCDW made a number of recommendations, among them the need for a follow-on workshop which "should be more narrowly focused with some specific recommendations".

The overall goal of the workshop was to improve the quantity and quality of chemical and biological data available to the scientific community. To outline the requirements for managing oceanographic chemical and biological data, the workshop started by:

- Identifying parameters that the IOC/IODE system can effectively handle;
- Describing minimum metadata requirements that make the data useful for future users of the data;
- Identifying problems that may limit the usefulness of historical data;
- Identifying users of these data and their requirements.

These objectives were accomplished in four sessions, as outlined in the Workshop Programme (Appendix I). Following each session, 'working groups' were lead by the session convenor to summarize and provide suggestions for the final panel discussion, which was held on the final day of the workshop.

In all, more than 40 experts from 15 countries and 3 international organizations (ICES, the Helsinki Commission, the Sir Alistair Hardy Foundation for Ocean Science) registered for the Workshop. The complete List of Participants is given in Annex II.

SESSION I: OVERVIEW

Convener - Trevor Platt (Canada)

BIOGEOCHEMICAL CYCLES

Trevor Platt and Shubha Sathyendranath

INTRODUCTION

The topic of biogeochemical cycles is a broad one. In the context of a meeting on data management, it is, perhaps, best condensed into the following two questions: How do we study ocean biogeochemical cycles at global and regional scales; and what are the data requirements for these studies? Alternatively, we could ask: Are the data managers doing the right thing from the point of view of scientists engaged in research on ocean biogeochemical cycles at large horizontal scales?

We shall take the view that results at large scale are arrived at by some kind of extrapolation procedure applied to observations taken at the restricted range of scales available to ships. In other words, we shall assume that the work to be done is computational in nature. The principal requirement from the data archive will then be the regional and seasonal distribution of oceanographic properties, both hydrographic and biogeochemical.

We can then envisage the following kinds of usage for the archive:

1. Use of the climatological fields to force numerical models.

2. Use of the climatological fields to compare with the output from numerical models.

These two applications are conventional enough. However, there is a third one that is, perhaps, less so.

3. Use of the archive as a basis to construct and test schemes for partition of the ocean into a suite of provinces to simplify analysis.

A DATA ARCHIVE FOR OCEAN BIOGEOCHEMISTRY

The problems of archiving ocean data magnify when the scope of the archive extends beyond the hydrographic through the geochemical to the biogeochemical. Generally speaking, biogeochemical data will be more complex, more variable with respect to measurement protocols, and less standardized with respect to how they should be interpreted. These differences arise partly from matters of definition, in particular from the translation of an absolute definition into an operational one, and partly from the related issue of lack of a primary standard against which some of the measurements can be compared.

Consider, for example, temperature, an important property in physical oceanography. Improvement in the methodology for measuring temperature at sea is not impeded by endless debate about what temperature "really is", or what the improved transducer is "really measuring", or whether temperature is what we should "really be measuring" in the first place. Neither will these issues arise when the results of ocean temperature measurements are applied to the vast body of physical-oceanographic theory in which temperature appears as a variable. It can be taken for granted that temperature has a universal significance that transcends the method used to measure it, and in any case, robust temperature standards are readily available. Once we are satisfied that the temperature measurements have been carefully made and carefully calibrated, we can concentrate on interpretation of the results, without agonizing about their limitations and meaning.

Things are much less clear-cut with respect to many biogeochemical variables. Especially in the ecological field, many of the variables in use have no universally-agreed definition. This is true even for the two most important ecological variables, biomass and production. And if the definitions could be agreed upon, it would not necessarily follow that these variables are, in fact, observables of the pelagic ecosystem, in the sense that acceptable operational equivalents of the basic definition could be found. It goes without saying that such a situation creates fertile ground for germination of the seeds of controversy.

Problems with settling on an operational definition spawn problems in standardization. In particular, it will often be difficult to trace procedures back to a primary standard. Usually, we will be fortunate if we have a secondary standard. Often, we will be even further removed from the absolute. An example is *in vivo* fluorescence, a proxy variable (with nonlinear response) for chlorophyll concentration, which is itself an imperfect proxy variable (conversion factor not constant) for particulate organic carbon, which many workers believe to be the correct way to index phytoplankton biomass. To put the implied uncertainty into perspective, bear in mind that phytoplankton biomass is perhaps the single most important variable in the marine ecosystem.

Another important quantity in ocean biogeochemistry is export production, the flux of organic carbon leaving the surface layer. It is measured by means of sediment traps, which are notoriously difficult to calibrate. Moreover, it is not easy to propose an independent method to measure the same thing and thus provide a check. We must fall back on proxies, such as the consumption of dissolved oxygen as the sedimenting organic matter is oxidized. This is an imperfect comparison at best. It becomes difficult, for example, to discuss issues of closure in local carbon budgets without worrying also about interpretation of the data. For a perspective, observe that export production is the most relevant variable in contemporary discussions of the role of the ocean as a sink for atmospheric carbon dioxide.

These difficulties, a reflection of the more complex nature of biogeochemical properties compared with hydrographic ones such as temperature, represent a considerable challenge to would-be archivers and managers of ocean biogeochemical data. At the least, they call for great attention in the provision of metadata to accompany the basic data files. They also present severe problems in the areas of quality assurance and quality control.

A related issue is that of coverage. If Walter Munk, speaking of data on physical oceanography, could refer to a "century of undersampling", we will find that for biogeochemical data, the situation is very much worse. The explanation lies partly in that the field has developed more slowly than hydrography, and partly in that the individual measurement methods for biogeochemistry are usually more time consuming than those for hydrography: there are few biogeochemical analogues of the conductivity-temperature-depth meter.

PARTITION OF THE OCEAN INTO PROVINCES

Notwithstanding the difficulties referred to above, extremely useful data compilations exist, that have made the names of Bunker, Levitus, Hastenrath and the like universally recognized wherever oceanography is practiced. These workers have done us a service beyond price. Their atlases can be used directly for the initiation (forcing) and testing of numerical models, or for any calculations in which we require to use climatological averages.

Particular problems arise in the case of sparse data sets for which the variable concerned cannot be derived as a universal function of an environmental co-variable. For example, to calculate primary production using models based on plant physiology, we need to know the initial slope of the photosynthesis-light curve. At present, we do not know how to calculate this property as a function of co-variables such as temperature and nutrients. One way to handle variables such as this is to partition the ocean into a suite of provinces, and to group the data according to their distribution by province and season. Then, values of the property could be assigned, when required, according to the same criteria of season and province, based on the relevant archive averages. Or, when province-specific rules are known, the values could be assigned as a function of local, environmental co-variables.

The partition is based on the premise that the ocean is not uniform: functional subsystems exist. One partition already in use seeks to capture the essential characteristics of these subsystems in so far as they affect primary production. The ocean is divided into regions having similar forcing with respect to the properties important for algal growth, under the premise that, in the pelagic ecosystem, the important biogeochemical rates are under physical control. Physical processes will regulate the environmental conditions that determine species assemblage, the vertical flux of nutrients and the manner of its delivery (constant or intermittent), the rate of vertical mixing and therefore the extent of photoadaptation, and the extent of stratification, important for the shape of the vertical profile of biogeochemical properties. If the ocean is mapped to show areas that share a common forcing, the resulting distribution of provinces will reflect the underlying oceanographic structure.

Data archives are extremely useful in construction and testing of such partitions. For example, the Arabian Sea has been subdivided into provinces based on the compilation of mixed-layer depth by Hastenrath.

ALLOWING FOR THE VARIANCE

One of the disadvantages of the atlas approach to data management is that it emphasizes the climatological averages at the expense of the variance. However, we would like to know how the biogeochemical "weather" extends, or averages to, the "climate" that we perceive at seasonal, annual, or longer time scales.

Again, we often require the average value of a function of environmental variables. In general, the average of the function will not be the same as the function of the averages. To take a simple, but important example, consider the calculation of new production ($\mathbf{P}_{new}$) as the product of total primary production ($\mathbf{P}_T$) and an **f**-ratio. The product of the averages is the unweighted product:

$$\mathbf{P}_{new} = \bar{\mathbf{f}}\,\bar{\mathbf{P}}_T,$$

whereas the average of the product is the weighted average

$$\mathbf{P}_{new} = \overline{\mathbf{f} \bullet \mathbf{P}_T}$$

These two equations will give different answers unless the covariance between **f** and $\mathbf{P}_T$ is zero. Usually, however, they are correlated, and the weighted calculation is to be preferred. For example, it was shown for the Sargasso Sea that the weighted calculation gave a value for new production that could be reconciled much more easily with previous data than that from the unweighted calculation.

These considerations indicate that we should not overlook the variance of properties when calculating their means. Nor should we overlook the chronological structure of data sets. We should analyze time series data to resolve the seasonal signal and the interannual variability. We should examine the response of the ecosystem to storms.

Such studies require that we integrate field data with remotely-sensed data. It will be a special challenge for data managers to reconcile remotely-sensed data with conventional data archives. The sheer magnitude of the data files can be daunting. Besides, the respective time and space scales of remotely-sensed and *in situ* data are not immediately compatible. Moreover, as is the case with *in situ* observations, the calibration and interpretation of remotely-sensed data on biological properties (such as chlorophyll), are more difficult than those on physical properties (such as temperature). For example, the relationship between ocean color and pigment concentration may vary, depending on the phytoplankton community structure. Also, the persistent presence of clouds will introduce bias into the averaged ocean-color fields. In short, remotely-sensed data is a solution that introduces its own kinds of problems. However, this is a challenge that must be addressed if we are to make the best use of the tools that we have.

SUMMARY

Data archives and data atlases for hydrographic properties have played an important role in the development of numerical modeling of the ocean system. Further problems arise in the archiving of biogeochemical data because of the sparseness of the data and because of obscurities surrounding interpretation. Hydrographic atlases have an additional function in aiding the construction and testing of schemes to partition the ocean into a suite of provinces to simplify the analysis of ocean biogeochemistry. It is important to emphasize the variance of ocean properties as well as their means if we are to exploit data bases to the full. An important entree into the variance is the use of remotely-sensed data. Methods must be found to reconcile the data archives from remote sensing with conventional ones from ships.

MANAGING THE DIVERSITY OF MARINE BIOLOGICAL DATA: PERSPECTIVES FROM A EUROPEAN UNION MARINE SCIENCES AND TECHNOLOGY (MAST) PROJECT

Gordon L J Paterson, Órla Ní Cheileachair and Yvonne McFadden

CHALLENGES TO MANAGING BIOLOGICAL OCEANOGRAPHIC DATA

Managing biological oceanographic data presents a number of challenges, which are not usually encountered when dealing with the physical sciences. These challenges can be summarized as:

- Diversity of biological data
- Project/ problem orientated data collection
- Scientific progress, i.e. new approaches, methods, gear
- Cultural practices and attitudes
- Funding and resources

Diversity of biological data. Biology is a subject with an enormous range of data types from biogeochemical data at one extreme to specimens of plants and animals on the other. The audiences for such data are also likely to be diverse. So making sure that the data are available to the appropriate group becomes an important issue.

Project / Problem centered data collection. Large-scale, multi-national, multidisciplinary projects are now an established mechanism for addressing the scientific questions of the day (e.g. JGOFS, GLOBEC, HELCOM, etc.). New approaches and novel types of data frequently result from the from such 'cutting-edge' programmes. As the scientific problems being tackled become more complex, the data will increasingly fall outside what is currently being catered for by data centres. Novel approaches and the data produced raises the issue about how they should be managed and made available to the scientific community.

Scientific Progress. Novel types of data and new methodologies often arise in response to a scientific question identified by research projects. Frequently little thought is given to providing some continuity between the new and old methods and thus providing a link and context between the new and archived data. In the worse cases new data becomes incompatible with previous work and the projects will have less general relevance. Development of new sampling and processing methodologies are particularly prone to this problem. Yet a balance has to be reached between progress and continuity.

Cultural practices and attitudes. Many biologists have little appreciation of data management, outside the narrow confines of their own project. As a result co-operative links between data centres and the data producers are weak or broken. Many data centres are not in a position to manage the diversity of biological data. So the issues raised by this revolve around: how to re-establish productive links between these two communities; how to market the skills and services that data management professionals can bring to biological projects, and the need to widen the scope of data management to include other areas of expertise such as Museums to manage specimens and DNA databases to manage sequence data from molecular biology.

Resources and funding. If the remit of data centres is broadened to encompass more biological oceanographic data types, then attitudes to data management must change. Researchers will have to take management seriously and treat it as an integral aspect of the project. To achieve this will require that funding is properly allocated to this task to ensure adequate resources can be provided by the data centres. In the current economic climate this is difficult to achieve as often the science is never fully funded and data management receives a low priority.

While there are such challenges, the situation is improving. Increasingly, funding agencies are insisting that all major research programmes make plans for data management. For example, current guidelines for projects in the latest EU MAST programme make it clear that data management is a priority and that detailed plans, including funding allocations, for data management of the project must be submitted with the science proposal. Such incentives will increasingly bridge the gap which currently exists between many biologists and data management professionals. However, to address fully the challenges presented by biological oceanographic data will require developing new approaches to manage and make available biological data.

ELECTRONIC DATA PUBLISHING - A POTENTIAL WAY FORWARD

The most obvious approach is to combine project orientated data management with easily accessed electronic data products. Project orientated data management is now common in most large projects (e.g. JGOFS, GLOBEC, and HELCOM). Several of these projects have produced CD ROMS of primary data, e.g. North Sea Community Project (Lowry *et al.* 1992) and U.S. Atlantic Continental Slope and Rise Program (Diaz *et al.* 1994). This need to reach and make data available to appropriate audiences prompted the MAST programme to fund a supporting initiative on electronic data publishing, the *EDAP* project. The objective of the EDAP project is to develop know-how on electronic data publishing for MAST research projects and to make available, by the end of the project, a guideline on electronic data publishing. Using the data from three selected MAST projects, EDAP takes a hands-on approach to investigating several ways of publishing marine data electronically.

One of these selected MAST projects included the *Community Structure and Processes* in the Deep-sea Benthic Community (Rice *et al.* 1996). The project's general objective was to determine the response of the abyssal benthic community to differing nutrient input regimes. Researchers from six nations and 16 institutes took part in the project. The diversity of data produced ranged from microbial molecular biology, community ecology data through to specimens and video footage. Managing this data, particularly in terms of quality control and assurance, was difficult, but perhaps the most pressing challenge was making the data available to the scientific community, as required by the MAST programme.

THE *PIRATE* PROJECT

As one of the EDAP approaches to developing electronic publishing expertise, the Irish Marine Data Centre developed a project-orientated system with the objective of collating such diverse data and presenting it in a coherent and comprehensive way. The system known as the PIRATE (Project Information Recorded About The Environment) system is a stand alone product available on CD-ROM and is designed to hold all of the data generated by a multidisciplinary project. The data and supporting information are accessed through a series of interactive user modules so that all information concerning a data set is readily available. PIRATE uses a relational database system to store the data.

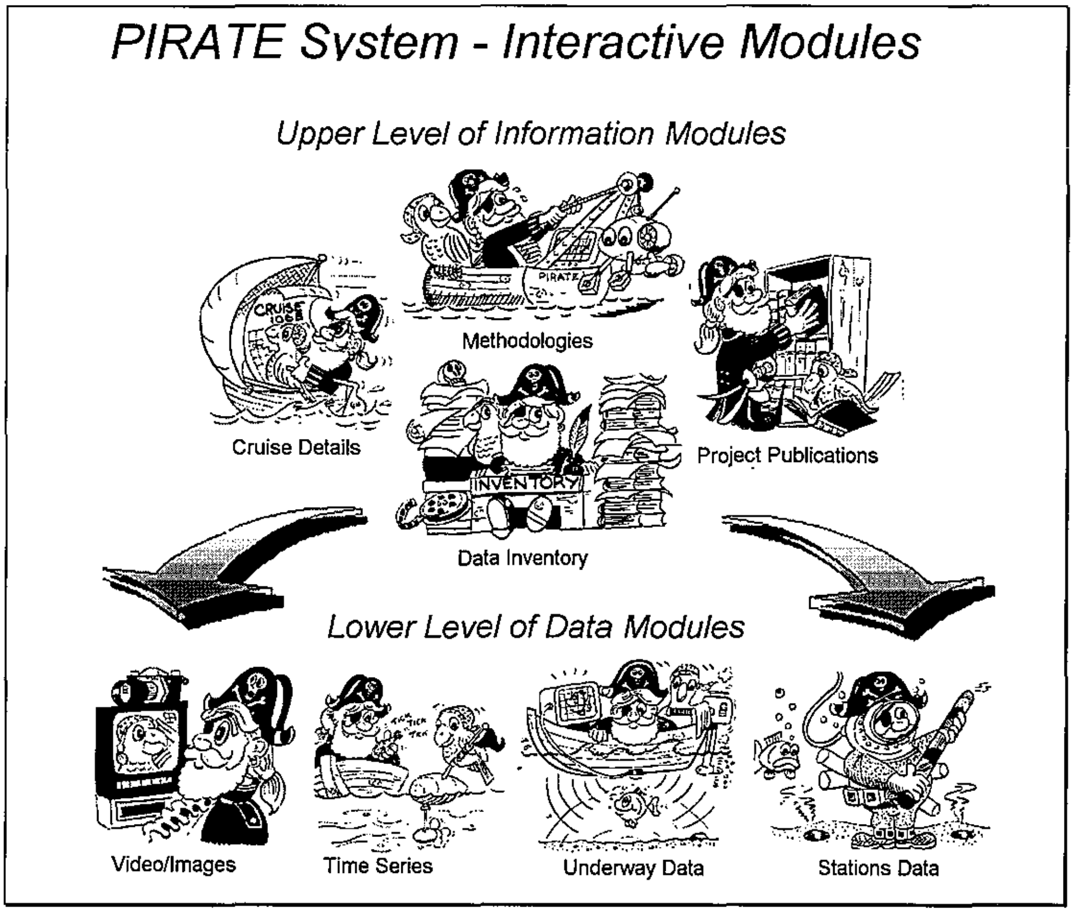


Figure 1.0 Overview of PIRATE system of interactive modules

This system was used for data generated by the *Community Structure and Processes in the Deep-sea Benthic Community* project. Given the diversity of data collection and generation techniques in the project, it was essential that as much of the supporting information or metadata was collected together with the primary data. A set of procedures for submitting the data was established. Each time data was submitted it was accompanied by a data inventory form describing the data set, sampling area, cruise details, and parameters sampled. In addition, all partners were asked at the start of the EDAP project, to produce standard protocols describing how their data was collected, processed and analyzed, together with what quality control and assurance procedures were used. Researchers were then able to refer to protocols when submitting data rather than have to re-describe the methodologies each time. This information was then included in the PIRATE system.

PIRATE is made up of a series of information and data modules (Fig.1). It has an upper information level comprising of a Data Inventory module, Cruise module, Methodologies module and Publications module. The Data Inventory module describes all of the data sets collected during the project, while the other modules describe associated cruise information, protocols and methodologies used and publications associated with each data set. Hence, all information starting from when the data was collected through to the final paper publication is readily accessible for each data set. The lower level holds the actual project data and is accessed through the Data Inventory Module of the upper level. The lower level consists of modules for the different types of data collected, e.g. station data, underway data, time series data, video, images and animation. All the data are geographically referenced and can be accessed from a map of sampling sites. Graphing and data visualization tools are included where appropriate.

The key to this system is flexibility through a modular approach to data access, coupled with interconnectivity between the various information modules. The upper level is designed so that all supporting documentation is linked to each data set in the Data Inventory module and integrated in a comprehensive manner. For example, by accessing an appropriate icon, the user can call up the cruise details or methodologies associated with each data set. Similarly, by accessing the Methodologies module it is possible to view all the data collected by a particular piece of gear or generated by a specific method.

ADVANTAGES OF STAND ALONE PRODUCTS FOR DIVERSE DATA SET

This stand alone approach has a number of advantages in managing diverse biological data. Firstly, it brings greater relevance by showing all the data together in a coherent way. Secondly, it can address a number of audiences without having to cater specifically for any one group. Thirdly, the user can examine how the data sets were produced and the measures taken to ensure their quality. Finally, such products are easily disseminated and if properly published should help overcome resistance to data management. However, care must be taken to ensure that no matter what system is used, the data remains available in a format, such as ASCII, which is widely accepted and, therefore, not at the mercy of changes in technology.

In summary, biological oceanographic data is often disjointed, diverse and a challenge to data management. Nevertheless, it is possible and certainly desirable to manage such data. An end-product that presents multidisciplinary project data in a visual, comprehensive and accessible manner, adds immediate and long-term value to the project and the research carried out. However, it may take non-conventional approaches to achieve this. In particular, data managers will need to play a more coordinating role in widening the scope of data management to include new technologies, such as the INTERNET and multimedia, as well as including the expertise of other data professionals such as Museum researchers.

ACKNOWLEDGMENTS

The authors wish to acknowledge the MAST programme for support in EDAP (MAS2-CT93-0072) and to thank many colleagues for fruitful discussions which contributed to the ideas expressed in this talk.

REFERENCES

Diaz, R.J., J.A. Blake and G. Randall Cutter Jr (Editors) 1994. Input, accumulation and cycling of materials on the continental slope off Cape Hatteras. *Deep-Sea Research Pt II* **41** (4-6). with CD-Rom Appendix.

Lowry, R. K. , R. N.Cramer, and L J Rickards 1992. *North Sea Project CD-Rom - user's guide.* NERC.

Rice, A. L. *et al.* 1996. *Community Structure and Processes in the Deep-Sea Benthos.* Final Report Contract no. MAS2-CT920033.

A REPORT ON "THE IOC/IODE GODAR PROJECT"

Sydney Levitus

INTRODUCTION

To understand the earth's climate system one must be able to describe both the mean state and the variability of the system, whether natural or anthropogenic. To determine the role of the world ocean as part of the earth's climate system, the international scientific community needs to have access to the most complete data bases of historical oceanographic data possible. Many historical oceanographic data are not available to the international community in digital form. The purpose of this manuscript is to describes a project to help alleviate this situation. "Data Archaeology" is the term used to describe the process of seeking out, restoring, evaluating, correcting, and interpreting historical data sets. "Rescue" refers to the effort to save data at risk of being lost to the science community.

Until recently there has been no observational system for the world ocean that even remotely compares to the meteorological observing system. Even now, with ship-of-opportunity programs, and drifting buoy programs, the data received in real-time and in delayed-mode are not sufficient to document the state of large parts of the world ocean and thus to provide initial conditions for ocean only or ocean-atmosphere prediction models. Thus, there is a need for placing all oceanographic data in digital databases.

In addition to scientific curiosity about the role of the world ocean as part of the earth's climate system, there are additional reasons for building the most complete oceanographic data bases possible.

1) It is the international scientific community that must advise national and international bodies on such issues as climate change and biodiversity. Hence, the international oceanographic and climate communities must have access to the most complete digital oceanographic data bases possible.

2) Substantial resources have been or may be allocated for national and international ocean and climate programs such as Climate and Global Change, Tropical Ocean and Global Atmosphere (TOGA), World Ocean Circulation Experiment, (WOCE), CLIVAR, and the Joint Global Ocean Flux Study (JGOFS), and for the establishment of a Global Ocean Observing System (GOOS). Planners of such programs should have access to all historical oceanographic data in order to optimize measurement strategies. Scientists analyzing data from such programs need historical data in order to place new data in perspective.

Based on personal observations and discussions with scientists and data managers from the international community, it is clear that substantial amounts of historical oceanographic data exist only in manuscript or analog form. In addition, there are data in digital form that are not available due to neglect or other reasons, for example, data that were used for one purpose by a scientist or institution, but never archived at an international data center. Effectively these data are not available to researchers and other users. Furthermore these data are at risk of being lost due to:

1) Media degradation such as fading ink or magnetic fields;

2) Environmental catastrophes such as fires and floods;

3) Simple neglect;

4) The retirement of individuals who know how to access these data or know the metadata associated with these data that make them useable to other scientists.

Therefore UNESCO's Intergovernmental Oceanographic Commission's (IOC) Committee on International Oceanographic Data Exchange (IODE) initiated a project known as the "Global Oceanographic Data Archaeology and Rescue" (GODAR) project.

Physical, chemical, and biological oceanographic data, as well as surface marine meteorological observations, are the specific types of data that the GODAR project focuses on. Initially, most data digitized or otherwise rescued have been physical parameters, particularly temperature data from bathythermograph profiles. In addition, data from the open ocean were given highest priority. Recognizing the importance of coastal regions to many countries, GODAR is also emphasizing data from these regions also. In particular, coastal time series records will be accorded priority.

As part of its leadership role of the GODAR project, WDC-A made a commitment to the institutions and countries who have made these data available. This commitment was to make all these data, plus all the data in the NODC/WDC-A digital profile archives available as ASCII files on CD-ROM as well as other electronic media. CD-ROM technology represents the least expensive and simplest way to exchange oceanographic data internationally, particularly to nations without an electronic network system. In accordance with GODAR and ICSU/IOC World Data Center principles, the data have been distributed internationally without restriction.

A publication entitled *"National Oceanographic Data Center Inventory of Physical Oceanographic Profiles: Global Distributions by Year for All Countries"* (Levitus and Gelfeld, 1992), served as the focal point for identifying the digital data held in the NODC/WDC-A archives. This publication has played an important role in discovering and acquiring historical data. Previous data distribution maps (Wust, 1935, Wyrtki, 1971; Gordon, 1982; Levitus, 1982, 1989) have already proven valuable for these efforts.

HISTORY OF GODAR

The IOC/IODE GODAR project had its origin at an *ad hoc* meeting held at NODC/WDC-A, Washington, D.C., in September 1990. The meeting was supported by the U.S. Climate and Global Change Program. Scientists and data managers from several countries and international centers including the Soviet Union, Republic of Korea, Japan, Chile, Australia, United States, and the International Council for the Exploration of the Seas (ICES) met informally to discuss the state of historical oceanographic data - in particular to discuss what should and could be done to stop the loss of irreplaceable data due to media degradation. The results of the meeting led to the establishment of various national and now an international project that are known generically as "Oceanographic Data Archaeology and Rescue" projects. For example as a result of this workshop, NODC received funding to begin ocean data archaeology and rescue projects. Similar projects were started at ICES and World Data Centers B and D for oceanography (WDC-B, WDC-D).

An international meeting known as the "Workshop on Ocean Climate Data" was arranged by the IOC and NOAA at Greenbelt, Maryland, U.S.A. (Churgin, 1992). The meeting was sponsored by the Commission of the European Communities (CEC), International Council of Scientific Unions (ICSU), World Meteorological Organization (WMO), ICES, and IOC. As a result of the demonstrated progress of various national data archaeology and rescue projects, the workshop recommended the expansion of these projects to band together under the umbrella of an existing international organization. Such support facilitates the exchange of data internationally.

A proposal for a GODAR project was submitted to the Fourteenth Session of the IODE meeting held in Paris, France from 1-9 December 1992. The IODE recommended to the IOC that this project be adopted as an IOC project. During the March 1993 IOC Assembly meeting, the IOC adopted the proposal for a GODAR project.

Specifically the GODAR project emphasizes:

a) Digitization of data now known to exist only in manuscript and/or analog form. This effort will have highest priority of all activities;

b) Rescue of digital data that are at risk of being lost due to media decay or neglect;

c) Ensuring that all oceanographic data available for international exchange are archived at two or more international data centers in digital form;

d) Preparing catalogues (inventories) of:

   1) Data now available only in manuscript form;
   2) Data now available only in analog form;
   3) Digital data not presently available to the international scientific community;

e) Performing quality control on all data and making all data accessible on various media including CD-ROM's as well as standard magnetic tape.

The first GODAR workshop was held in Obninsk, Russia, during May 1993. This meeting focused on data sets and activities in eastern and northern Europe. This region was chosen in particular because of the real possibility of the loss of substantial data sets due to economic conditions in eastern Europe. The report of the first GODAR workshop (IOC, 1993) gives some indication of the amounts of data that exist in manuscript form. The Russian delegation reported the existence of approximately 450,000 Mechanical Bathythermograph (MBT) profiles and 800,000 Oceanographic Station (OSD) profiles in manuscript form. Ukraine has data from at least 100 cruises. It should be noted that the data received at WDC-A from the Russian NODC over the last two years has been of excellent quality.

Additional regional GODAR meetings have been held in China (IOC, 1994a), India (IOC, 1994b), and Malta (IOC, 1995). Additional meetings are planned to be held in South America in October, 1996 and in West Africa in early 1997.

METHODS OF INVESTIGATION

From inception of the various national archaeology and rescue projects at various centers, efforts were coordinated to avoid duplication of effort and to maximize the use of scarce resources. Joint activities include the exchange of data, data distribution plots, catalogue information about data holdings, and the exchange of scientists and data managers between centers. An emphasis on "rescue" and exchange of data occurring simultaneously was for two reasons:

1) Some data are at risk of being permanently lost if not saved immediately;

2) The international research and administrative communities needed a demonstration of how quickly the project could act to actually make previously unavailable data accessible in digital format.

Perhaps the most valuable technique to quickly describe data holdings is to produce data distribution plots and tables of the number of profiles on a year-by-year basis for each major measurement type. Levitus and Gelfeld (1992) have done this for each of the major NODC digital archives. This work showed the distributions of NODC holdings for all countries. NODC/WDC-A has prepared similar plots on a country-by-country basis and distributed such summaries to data centers, scientists, and institutions in approximately 20 countries. These summaries have generated much interest and resulted in the exchange of more information and data, e.g., through the generation of such summaries the GODAR project was able to conclude that substantial amounts of Canadian MBT profiles (10,000) that already existed in digital form were not in the NODC/WDC-A files. These profiles have been acquired by NODC/WDC-A and will be distributed as part of the GODAR project and incorporated into the NODC/WDC-A archive.

Examination of cruise reports and monographs have also supplied both data and metadata not available in digital form. By examining the report series *"Special Scientific Reports...Fisheries"* of the Fish and Wildlife Service of the Department of Interior, we found chemical and biological data from U.S. research cruises that had never been digitized even though in some instances the physical parameters of these profiles are part of the NODC digital archive. Metadata about these biological and chemical parameters, such as measurement techniques, are available in these reports in detail. We also found observations of surface salinity and phosphate that had been measured at the same time that MBT profiles were made. We have begun digitizing all these data.

One example of a monograph providing substantial information about historical oceanographic observations is the work by Muromtsev (1958). He provides a list of "deep-water" observation made in the Pacific Ocean for the 19th and 20th centuries. In addition Muromtsev provides a table containing the actual temperature measurements from profiles made in the Pacific Ocean during the period 1804-1873. Although the accuracy of these measurements needs to be examined to determine whether the data are useful, at least the data have been tabulated in a single location with metadata that will allow the determination of the suitability of these data for scientific studies. Another example is the U.S. National Academy of Sciences report by Vaughn (1937). This document contains contributions by Defant, Sverdrup, Helland-Hansen, and Wust who, in total, describe a great deal of profiles made globally up to the date of the report. Some of this material has been published in other reports or atlases (e.g. Wust, 1935). Some of the material describes data not archived at NODC/WDC-A in digital form.

RESULTS

Table 1 documents how successful GODAR efforts have been. When one considers that the global historical data base of temperature data has increased by more than 1.2 million profiles in the past two years and that the results of the first GODAR workshop (IOC, 1993) have identified on the order of another one million profiles that are in manuscript form, it becomes clear that the international scientific community will have access to a much more complete data base than ever thought possible. This includes many MBT profiles that will help determine interannual variability of the upper ocean thermal structure. In addition, numerous Oceanographic Station Profiles that include temperature, salinity, and other parameters are now available.

Numerous institutions from the international community are now participating in the GODAR project. The navies of several countries have been declassifying oceanographic data and making these data available internationally without restriction. For example the Russian navy is participating in the GODAR project by making manuscript data available for digitization and distribution (IOC, 1993). Approximately 135,000 BT profiles have been made available to date from Russian naval vessels. During the last several years the U.S. Navy has declassified approximately 150,000 XBT profiles and made them available for international distribution. This in addition to more than six hundred thousand BT profiles previously made available. These data are now part of the NODC/WDC-A data bases and are available for international distribution. The British Navy has declassified approximately 171,000 bathythermograph profiles. Some of these observations were taken as early as 1945.

FUTURE WORK

Many of the data sets received to date have been processed at NODC/WDC-A and have been distributed internationally without restriction. CD-ROM media have proven to be the most popular way of disseminating these data. The CD-ROM technology has several advantages for the distribution of *in situ* oceanographic data:

1) There exists an international standard for reading and writing information to CD-ROMs;

2) CD-ROMs have relatively large storage capacity for the *in situ* oceanographic data sets we are distributing;

3) They are relatively inexpensive to master and duplicate;

4) Their small size, which makes physical distribution easier as compared with magnetic tapes;

5) A relatively small investment is required to obtain a personal computer and CD-ROM drive for accessing these data. Neither mainframes or even minicomputers are required for digital access to large data bases.

While we fully expect the data, and analyses of these data to be made available over electronic networks, we believe that CD-ROM technology best serves the international distribution of these data. Some countries simply do not yet have the required network capability. Even in many technologically advanced countries the transfer of large data sets over the Internet is not practical because of slowness associated with the phenomenal growth of the Internet.

The next generation of CD-ROM technology which should be available in 1997 has such high data recording densities that it appears that the entire known digital archive of ocean profiles will fit one CD-ROM in ASCII format. Thus, any scientists or citizen will have access to all these data.

More comprehensive data bases will lead to a better description of both the mean state and the variability of the world ocean. This will lead to better understanding of the role of the world ocean in the earth's climate system. Such work is international in nature. The atlas by Olbers *et al.* (1992) is an excellent example of an analysis of oceanographic data that has benefited by international cooperation resulting in the exchange of data.

It is important to emphasize that as part of its commitment to the institutions and countries who have made these data available, that the data gathered as a result of the GODAR project, and many products based on these data sets have been distributed internationally without restriction. These analyses are described by Conkright *et al.*, 1994; Levitus and Boyer, 1994a, 1994b, 1994c; Levitus, *et al.*, 1994. It is our hope that other investigators, research groups, and data centers from all countries with marine research and operational programs will participate in this project.

ACKNOWLEDGMENTS

We acknowledge the contribution of many individuals and organizations to this publication. Scientists and technicians studying the world ocean have undertaken the task of collecting and processing the data. Data centers in China, Japan, South Korea, India, and Russia have been particularly helpful through their participation in GODAR and GODAR type projects. Ms. Lin Shaohua and Professor Hou Wenfeng of WDC-D were instrumental in making data available from China.

Although many data managers and scientists from the international community have been instrumental in building global oceanographic data bases, we would like to thank in particular Harry Dooley, Shin Tani, Nick Mikhailov, and Yuri Sychev for their commitment to national and international oceanographic data archaeology and efforts. They were among the founders of international GODAR efforts and have continued to play a leading role in these efforts.

The organization of WDC-A data for digitization was due to the efforts of Ron Moffatt, Godfrey Trammell, and Ron Fauquet.

In addition to declassification of naval data there are data that have been gathered by navies that, while not classified in a formal sense, were not previously available. We appreciate the effort of these navies to make such data available.

Youri Oliounine of the IOC has provided valuable technical support in the formation and management of the IOC/IODE GODAR project. Ron Wilson, Chairman of the International Oceanographic Data Exchange Committee of the IOC and Director of the Canadian Marine Environmental Data Service has provided consistent and strong support for GODAR.

The NOAA Climate and Global Change Program and the NOAA Earth Science Data and Information System Management Program have supported the work of the NODC Data Archaeology and Rescue projects. The direct result of these projects is that valuable historical oceanographic data are now available to the international user community.

REFERENCES

Boyer, T. and S. Levitus, 1994: *Quality control of oxygen, temperature and salinity data.* **NOAA Technical Report No. 81,** National Oceanographic Data Center, Washington, D.C., 65 pp.

Churgin, J., 1992: *Proceeding of the Ocean Climate Data Workshop.* Available from Users Services Branch, NODC, E/OC2, Universal Bldg., 1825 Connecticut Ave., NW, Washington, DC.

Conkright, M.E., T. Boyer, and S. Levitus 1994b: ***Quality control and processing of historical oceanographic nutrient data.*** **NOAA Technical Report NESDIS 79,** National Oceanographic Data Center, Washington, D.C., 75 pp.

Conkright, M., S. Levitus, T. Boyer, 1994: World Ocean Atlas 1994, Vol. 1: Nutrients. NOAA Atlas Series.

Gordon, A.L., E. Molinelli, and T. Baker, 1982: Southern Ocean Atlas. *Columbia U. Press,* pp. 11, plates 233.

IOC, 1993: IOC-CEC-ICSU-ICES Regional Workshop for Member States of Eastern and Northern Europe (Global Oceanographic Data Archaeology and Rescue [GODAR] Project. **IOC Workshop Report 88. WDC-B, Oceanography Obninsk, Russia, 17-20 May 1993.**

IOC, 1994a: IOC-SOA-NOA Regional Workshop for Member States of the western Pacific-GODAR II (Global Oceanographic Data Archaeology and Rescue project). **IOC Workshop Report 100. World Data Centre D, Oceanography, Tianjin, China, 8-11 March, 1994.**

IOC, 1994b: IOC-ICSU-NIO-NOAA Regional Workshop for Member States of the Indian Ocean-GODAR III (Global Oceanographic Data Archaeology and Rescue Project. **IOC Workshop Report 107. Indian National Oceanographic Data Centre, Dona Paula, Goa, India, 6-9 December, 1994.**

IOC, 1995: IOC-ICSU-NOAA Regional Workshop for Member States of the Mediterranean Sea-GODAR IV (Global Oceanographic Data Archaeology and Rescue Project). **IOC Workshop Report 110. University of Malta, Foundation for International Studies, Valletta, Malta, 25-28 April, 1995**

IOC, 1992: *Summary Report of the Fourteenth Session of the IOC/IODE, 1-9 December 1992.* IOC/IODE-XIV/3, 32pp.

Levitus, S., 1982: *Climatological Atlas of the World Ocean,* NOAA Professional Paper No. 13, U.S. Gov. Printing Office, 173pp., w/microfiche attachments.

Levitus, S., 1988: Decadal and Pentadal Distributions of Hydrographic Stations at 1000 m Depth for the World Ocean. *Progress in Oceanography*, 20, 83-101.

Levitus, S., 1989: Interpentadal Variability of Temperature and Salinity at Intermediate Depths of the North Atlantic Ocean, 1970-74 Versus 1955-59. *Journal of Geophysical Research-Oceans*, 94, 6091-6131.

Levitus, S., and R. Gelfeld, 1992: National Oceanographic Data Center Inventory of Physical Oceanographic Profiles, Global Distributions by Year for All Countries. *Key to Oceanographic Records Documentation No. 18*, 242pp.

Levitus, S., M. Conkright, J.L. Reid, R. Najjar, and A. Mantyla, 1993: Global Distributions of Silicate, Nitrate, and Phosphate in the World Ocean. *Progress in Oceanography, 31, 245-273.*

Levitus, S., and J. Antonov, X. Zhou, H. Dooley, K. Selemenov, V. Tereschenkov, 1996: Decadal-Scale Variability of the North Atlantic Ocean. Accepted by National Academy of Sciences for publication in "Natural Climate Variability on Decade-to-Century Time Scales".

Conkright, S. Levitus, T. Boyer, 1994: ***World Ocean Atlas 1994, Vol. 1: Nutrients.*** *NOAA Atlas NESDIS 1*, U.S. Gov. Printing Office, Washington, D.C., 150 pp.

Levitus, S., T. Boyer, 1994: ***World Ocean Atlas 1994, Vol. 2: Oxygen.*** *NOAA Atlas NESDIS 2*, U.S. Gov. Printing Office, Washington, D.C., 186 pp.

Levitus, S., R. Burgett, T. Boyer, 1994: ***World Ocean Atlas 1994, Vol. 3: Salinity.*** *NOAA Atlas NESDIS 3*, U.S. Gov. Printing Office, Washington, D.C., 99 pp.

Levitus, S., T. Boyer, 1994: ***World Ocean Atlas 1994, Vol. 4: Temperature.*** *NOAA Atlas NESDIS 4*, U.S. Gov. Printing Office, Washington, D.C., 117 pp.

Levitus, S., T. Boyer, 1994: ***World Ocean Atlas 1994, Vol. 5: Interannual variability of upper ocean thermal structure.*** *NOAA Atlas NESDIS 5*. U.S. Gov. Printing Office, Washington, D.C., 176 pp.

Levitus, S., R. Gelfeld, T. Boyer, and D. Johnson, 1994a: ***Results of the NODC and IOC Data Archaeology and Rescue projects. Key to Oceanographic Records Documentation No. 19***, National Oceanographic Data Center, Washington, D.C., 67 pp.

Muromtsev, A.M., 1958: *The Principal Hydrological Features of the Pacific Ocean.* Main Administration of the Hydrometeorological Service of the USSR Council of Ministers, State Oceanographic Institute, (Published for National Science Foundation, Washington, D.C. by the Israel Program for Scientific Translations, Jerusalem, 1963, 417pp)

Olbers, D., V. Gouretzki, G. Seib, J. Schroeter, 1992: *The hydrographic atlas of the Southern Ocean*, Alfred-Wegener-Institute for Polar and Marine Research, Bremerhaven.

Robinson, M. K., and R.A. Bauer 1976: *Atlas of the North Pacific monthly mean temperatures and mean salinities of the surface layer*, U.S. Naval Oceanographic Office, Washington, D.C., 173 pp.

Robinson, M. K., R. A. Bauer, E. H. Schroeder, 1979: *Atlas of North Atlantic-Indian Ocean monthly mean temperatures.* U.S. Naval Oceanographic Office, Washington, D.C., 213pp.

Schroeder, E. H., 1965: Average monthly temperatures in the North Atlantic Ocean. *Deep Sea Res.*, 12,323343. Woods Hole, MA., pp 24.

Vaughn, T. W., 1937: International Aspects of Oceanography: Oceanographic Data and Provisions for Oceanographic Research. National Academy of Sciences, Washington, D.C., pp 225.

Wust, G., 1935: *The Stratosphere of the Atlantic Ocean*, English translation edited by W. J. Emery, Aprevind.

Wyrtki, K., 1971: *Oceanographic Atlas of the Indian Ocean Expedition*, National Science Foundation, 531pp.

GODAR DATA AND PROJECT ANNOUNCEMENTS:

1) Levitus, S., and R. Gelfeld, 1993: Oceanographic data archaeology project receives international support. ***NOAA/Earth System Monitor***, (2).

2) Levitus, S. and R. Gelfeld, 1993: Oceanographic data archaeology project receives international support. ***UNESCO, IMS Newsletter*** **No. 67**, pp. 4-5.

3) Levitus, S., R. Gelfeld, L. Stathoplos, and M.E. Conkright, 1993: Oceanographic Data Archaeology and Rescue Project is Underway. **U.S. JGOFS News**, 5(2).

4) Levitus, S., M.E. Conkright, R. Gelfeld, and T.P. Boyer, 1994: World Ocean Atlas 1994. **Global Change Newsletter**, No. 20, pp 4-6.

5) Gelfeld, R., and H. Dooley, 1994: Old Sea Data Help Scientists Today. ICES CIEM Information Newsletter, No. 24,p. 9.

6) Levitus, S., M.E. Conkright, R. Gelfeld, T.P. Boyer, 1994: World Ocean Atlas presents new Ocean climatologies. **NOAA/Earth System Monitor**, 5(2).

7) Levitus, S., M.E.Conkright, R. Gelfeld, and T.P. Boyer: World Ocean Atlas 1994 and CD-ROM Data sets, 1995. **Bulletin of the American Meteorological Society**, p. 405.

8) Anonymous, 1995a: New GODAR product available, ***UNESCO, IMS Newsletter*** **No. 73-74**, p. 11.

9) Anonymous, 1995b: New GODAR product available, **EOS Transactions, Section News**, 76(34), p. 339.

10) Anonymous, 1995c: Atlas of the World's Oceans on CD-ROM, **Scientific Computing**, Issue 13 November 1995.

Table 1: Data sets received through NODC and IOC Data Archaeology and Rescue projects (partial list as of May, 1996).

COUNTRY INST	DATA TYPE	NO. OF PROFILES	PERIOD OF OBS.	REMARKS
AUSTRALIA				
CSIRO	OSD	22,190	1929-1990	
CANADA				
MEDS	XBT	46,658	1968-1988	
	DBT	11,563	1982-1981	
	MBT	145,286	1943-1988	
CHINA				
WDC-D	OSD	6,817	1958-1990	
FRANCE	MBT	2,791	1964-1972	MBT+Surface Salinity
ICELAND	OSD	7,311	1938-1988	
ICES				
	OSD	8,850	1970-1992	DENMARK
	OSD	2,681	1971-1974	CINCECA
	OSD	432	1948-1990	OWS L
	OSD	1,171	1983-1990	OWS C
	OSD	7,989	1948-1988	OWS M
	OSD	245	1925-1930	ATLANTIC SLOPE
INDIA				
INODC	OSD	650	1976-1988	T, S, nutrients

JAPAN				
JODC	OSD	254,846	1965-1990	JAPAN FISH. AG.
	DBT	23,452	1979-1986	JAPAN FISH. AG.
	MBT	60,764	1965-1985	JAPAN FISH. AG.
	XBT	1,774	1979-1985	JAPAN FISH. AG.
	CURM	28,487	1964-1985	JAPAN FISH. AG.
RUSSIA				
POI	OSD	5,543	1947-1988	S. CHINA SEA
	CTD	4,249	1981-1988	S. CHINA SEA
WDC-B	MBT	333,000	1941-1987	
	OSD	64,560	1969-1991	N. Atlantic
SOUTH KOREA				
KODC	OSD	28,193	1961-1992	
UNITED KINGDOM				
BHO	MBT	63,187	1947-1981	
BHO	XBT	108,407	1969-1990	
IOS	OSD	105	1973	R.R.S. SHACKLETON, CRUISE 6 ICES Overflow73 Expedition
United States				
SI0	MBT	177,585	1941-1972	North Pacific
WHOI	MBT	242,264	1941-1961	U.S. Navy, Coast Guard, and research ships
WDC-A	OSD	1,169	1958-1978	Digitization of manuscrip data held at WDC-A
SI0	OSD	162	1972-1973	TASADAY Legs I-IV, Southtow, Climax
US COAST GUARD				
	OSD	634	1961	Surface T&S

SESSION II: APPLICATION OF HISTORICAL BIOLOGICAL AND CHEMICAL MEASUREMENTS

Convener - Tommy D. Dickey (USA)

USE OF TIME SERIES OF CHLOROPHYLL DATA

Y. Dandonneau

INTRODUCTION

Chlorophyll data are useful for two main reasons: as a component of algal cells, chlorophyll represents an estimate of the biomass of phytoplankton, and as the molecule which captures the photons and initiates the photosynthetic reactions, it is used as an impact in models of photosynthesis. A general formulation for these models is P = [chl] a* PAR , where P is the carbon fixation (moles of carbon per unit volume and unit time), [chl] is the chlorophyll concentration (unit mass of chlorophyll per unit volume a* is the chlorophyll specific light absorption coefficient (unit surface per unit mass of chlorophyll), PAR is the photosynthetic available radiation (moles of photons per unit surface and unit time) and is the quantum yield of photosynthesis (moles of carbon per moles of photons). This equation shows that, in addition to chlorophyll a, and light, there are two other inputs in light-photosynthesis models: the absorption of light by chlorophyll, and the quantum yield of photosynthesis. Both terms are highly variable at sea. Especially the absorption of light by chlorophyll should not be taken as a constant and varies according to the light or nutrients regimes, being smaller when the chlorophyll is concentrated into big, dense chloroplasts (package effect). Thus the capture of photons by chlorophyll cannot be known from the chlorophyll concentration alone . There is however the hope that knowledge of the ecological conditions will permit to predict a*, thus improving the accuracy of the models. A further step, if one wants to run models without forcing by chlorophyll data , is to convert the atoms of carbon fixed by photosynthesis into increases in chlorophyll concentration. This requires the knowledge of the carbon to chlorophyll ratio in phytoplankton, which is also highly variable (i.e. from 30 to 200). Lack of knowledge of a* and of c:chl is thus an important limit to the precision of models of photosynthesis and requires many further field experiments.

Another problem arises from the way in which a* is measured, and from the existence of pigments other than chlorophyll which are taken into account by this measurement. Some of these pigments are photo-protectants, are do not interact in the photosynthetic reactions (carotenoids). Some other ones broaden the light absorption spectrum of the algae: they capture photons at wavelengths where chl a cannot and transmit the energy to the photosystem (phycobiliproteins). The omnipresent and often abundant genus *Prochlorococcus* has no chlorophyll a but instead, contains divinyl-chlorophyll a which makes the photosynthesis, so that when measured separately, the sum chlorophyll a + divinyl-chlorophyll a must be used in models of photosynthesis, rather than chlorophyll a alone.

In spite of these shortcomings, chlorophyll has a unique advantage: it is responsible of the color of the sea, so that its concentration can be derived from the reflectance of the sea at certain wavelengths, measured by satellites. This property opens the way to global estimates of the primary production of the sea, on a nearly monthly basis. After the CZCS which collected the first sea color data, but stopped emitting eight years ago, SeaWiFS, OCTS, and POLDER will be launched soon. Satellites are powerful tools, and future use of chlorophyll measurements data will be oriented to compensate for the limits of satellite derived chlorophyll data :

-already existing chlorophyll data have been used to infer, on a regional basis, the vertical profile of [Chl] from the value at the surface.

-future data will help to improve the accuracy of the algorithms which are used to convert the reflectance into [chl].

-drifters with fluorescence sensors, or absorbance-meters at the wavelengths where chlorophyll absorbs light, can be deployed in those regions which are almost permanently covered by clouds and that the satellite cannot see.

Only using satellites detected chlorophyll will it be possible to address the questions raised by the role of biology in the evolution of climate.

Some existing data sets :

-Most marine laboratories located near the coast make regular sampling in a bay, or along a transect. The JGOFS time series stations: Hawaii (HOT), Bermuda (BATS), Villefranche sur mer (DYFAMED) are the most recent and sophisticated ones.

-Many modelers have made simulations of the ocean at weather station "P" in the North Pacific, where some chlorophyll data have been collected during a few years. The interest of station "P" lies in its location, where lateral advection is very small so that 1D models work in a realistic way, rather than in the quality of the chlorophyll time series.

- An impressive set of chlorophyll data has been collected during the CALCOFI cruises in the upwelling of California.

These three examples can be qualified as "conventional" ones, as the chlorophyll data are (or have been) obtained using conventional techniques. This is not the case for some other important data sets:

-The SURTROPAC chlorophyll data (Dandonneau, 1992) which cover the tropical Pacific from 1978 to 1989 have been obtained from merchant ships using a specially adapted measurement technique (Figure 1).

-The Coastal Zone Color Scanner had not enough memory to globally store the measurements. Time series studies have however been made off California, in the Mediterranean Sea, and in the North Atlantic.

-A long time series of "chlorophyll" data has been built in the North Atlantic archiving the color of the silk of the continuous Plankton Recorder

-An even longer time series potentially exists in measurements of the transparency of seawater using the Secchi disk.

-More and more time series are being gathered from moored instruments (fluorescence, beam transmission, light absorption at 670 nm).

Under sampling and high quality data characterize the conventional time series, while oversampling can only be achieved with non conventional times series, in which proxies, rather than chlorophyll, are measured.

METHODS

The evolution of the techniques to measure chlorophyll concentration at sea is driven by two divergent objectives: to be able to measure more quickly and at a cheaper price, or to discriminate deeper into the variety of photosynthetic pigments.

Starting from the spectrophotometric measurements which were used in the 50's, the first pathway goes through the fluorometric determination of chlorophyll and pheophytin on extracts, to methods which do not require filtration: underway in vivo fluorescence, in situ fluorescence, beam transmission, and even: the satellite measurements.

The second pathway has moved to chromatography: HPLC can measure more than 20 different pigments, Analogous to HPLC, but faster and less expensive, spectrofluorometry can measure 10 different pigments.

The later techniques are not adapted to the objectives of mapping and monitoring, and are generally used during process studies.

CALIBRATIONS

Obviously, fluorescence or absorption data, or sea color detected from satellites need to be calibrated. They are bulk properties of a complex combination of pigments, and are generally calibrated versus chlorophyll a. Chlorophyll a can be purchased, per mg, which is a rather large amount and must be diluted before it gives fluorescence, or an optical density, comparable to what is generally obtained in routine work at sea. For accessory pigments which are measured by HPLC or spectrofluorimetry, pure pigments must be isolated and quantified to serve as standard, which represents a very complex task. Further, most pigments are degraded by light, so that operations must be conducted in dim light as quickly as possible. The difficulty in driving these calibrations may have resulted in errors for some data sets or subsets. There exist some substances which have optical properties similar to those of chlorophyll and may offer, easier ways to control the drift of fluorometers or spectrophotometers. Ethidium bromide for instance emits red light when excited with blue light. An ingenious standard has been made which can serve for chlorophyll a and pheophytin a, and consists of a cylinder of polymetacrylate of carbon into which pheophytin a has been incorporated prior to polymerization.

The above comments concern diluted pigments in acetone or alcohol. To obtain such solutions from natural phytoplankton populations at sea requires filtration and extraction, which is a critical step. For instance, it is still a controversy whether glass fiber GF/F filters, retain the picoplankton or not

(Chavez et al., 1995). The issue of this question has heavy implications: if it was ascertained that a significant and variable percentage of phytoplankton could pass through GF/F filters, we would have to abandon them and use expensive membrane filters which considerably lengthen the filtrations and are less convenient than the glass fiber filters for rapid, descriptive work.

Physiology must also be taken into account when calibrating bio-optical instruments. Figure 2 shows a record of in vivo fluorescence obtained in the equatorial Pacific. Obviously, there is a very pronounced diurnal cycle which was not found in the chlorophyll concentration and should be known and filtered out if one wants to use this record as a proxy for chlorophyll a.

LONG-TERM TRENDS IN CHLOROPHYLL CONCENTRATION

The evolution of the techniques to measure chlorophyll concentration have evolved, and should continue to evolve because we still do not handle a technique both precise, accurate, rapid and inexpensive. The consequence of the evolution, and of the problems discussed above, is that it is extremely difficult to find out significant long-term trends in the chlorophyll concentration in a given region. Venrick et al., (1987) suggest that an increase in chlorophyll concentration could have occurred from 1973 on in the central north Pacific. Their demonstration is based on a careful examination of all possible biases which could have resulted from changes in the measurements techniques.

On the opposite, in a global analysis, the SURTROPAC chlorophyll data have been "de-biased" assuming that the chlorophyll concentration should not change with time in the oligotrophic area between 12° and 15° latitude, north and south in the central and western Pacific (Dandonneau, 1992). A time dependent correction factor for the SURTROPAC data was thus proposed which made the average concentration in the two areas (taken together) constant.

Clearly, the problem of trends, which is very relevant to the problem of global warming, cannot easily be addressed with the existing chlorophyll data sets.

REFERENCES

Chavez, F.P., Buck, K.R., Bidigare, R.R., Karl, D.M., Hebel, D., Latasa, M., Campbell, L. (1995), on the chlorophyll a retention properties of glass-fiber GF/F filters. Limnology and Oceanography, **40** (2), 428-433.

Dandonneau, Y. (1992), Surface chlorophyll concentration in the Tropical Pacific Ocean: an analysis of data collected by merchant ships from 1978 to 1989. Journal of Geophysical Research, **97** (C3), 3581-3591.

Venrick, E.L., McGowan, J.A., Cayan, D.R., Hayward, T.L. (1987), climate and chlorophyll a: long-term trends in the central north Pacific Ocean. Science, **238** 70-72.

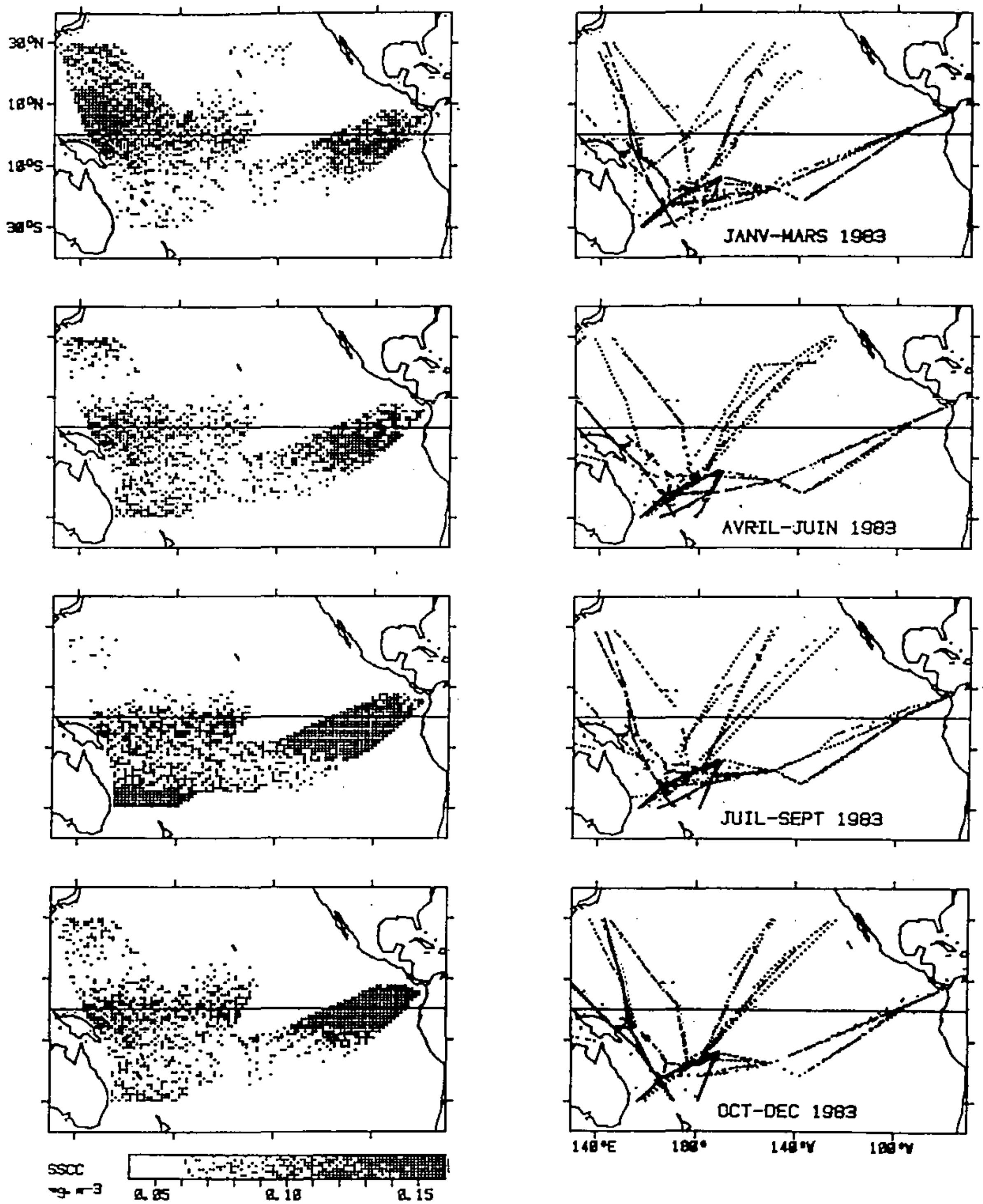


Figure 1: The SURTROPAC chlorophyll data: quarterly maps of Sea Surface Chlorophyll Concentration (SSCC) in the tropical Pacific (left column) and available data (right column).

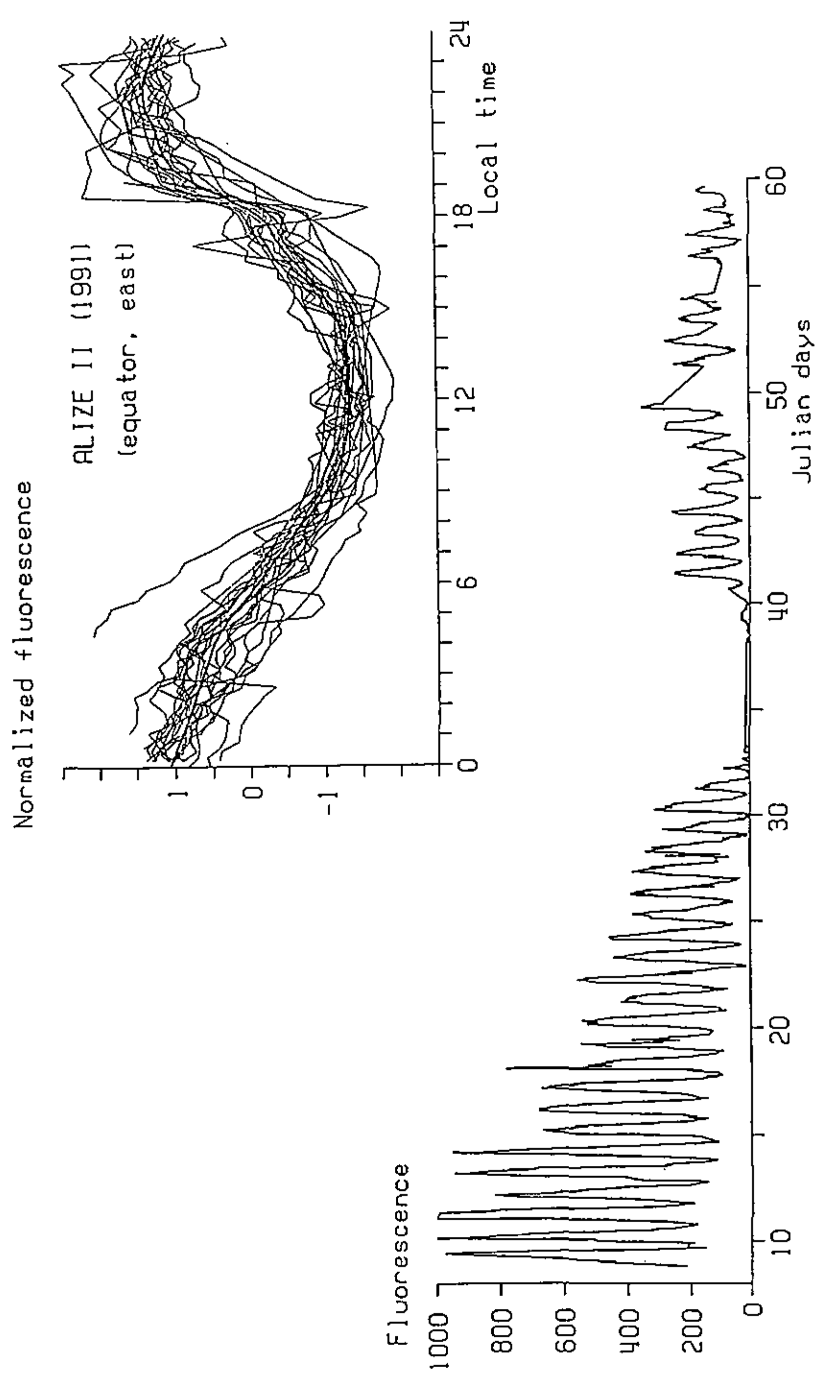


Figure 2: Underway record of *in vivo* fluorescence along the equator in the Pacific Ocean (bottom) and diurnal cycle of fluorescence (top). This diurnal cycle was caused by diurnal cycle of the fluorescence to chlorophyll ratio (the chlorophyll concentration did not exhibit any clear diurnal cycle).

OBJECTIVE ANALYSIS OF SURFACE CHLOROPHYLL DATA IN THE NORTHERN HEMISPHERE

M.E. Conkright and S. Levitus

INTRODUCTION

The Ocean Climate Laboratory (OCL) at the National Oceanographic Data Center (NODC) is supported by the NOAA Climate and Global Change program to produce scientifically quality controlled oceanographic databases. 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 using objective analysis techniques. Specifically, this project has produced five ocean atlases describing the global distributions of these parameters and two technical reports describing the quality control and processing procedures (Conkright et al., 1994a, 1994b; Boyer et al., 1994; Levitus and Boyer, 1994a, 1994b; Levitus et al., 1994a, 1994b). Observed and standard level profile data (along with quality control flags), objectively analyzed one-degree latitude-longitude mean fields for each of parameter, and five-degree square statistics of standard level values have been made available to the international oceanographic community on the World Ocean Atlas 1994 CD-ROM series.

We have expanded this work to include the quality control and analysis of historical chlorophyll- a data. This work presents an initial attempt at comparing objectively analyzed chlorophyll data, on a one-by-one degree, with global Coastal Zone Color Scanner data. Since most of the Arctic Ocean and Southern Hemisphere contain large areas where not enough data are available to produce representative fields, we confine the analysis to surface waters in the Northern Hemisphere and equatorial regions.

METHODS

Global composites of Coastal Zone Color Scanner (CZCS) ocean color images created by Gene Feldman at Goddard Space Flight Center, NASA (1978-1986) were used in this study (Feldman et al., 1989). The CZCS Level 3 data was subsampled to map on a 1x1 degree grid. The CZCS images are not shown in these figures since they require color in order to accurately map the distributions. For comparison of analyzed chlorophyll fields with CZCS images, refer to the SeaWiFS home page: http://seawifs.gsfc.nasa.gov/SEAWIFS.html.

To date, our digital chlorophyll database contains 92,933 profiles. The distribution of surface chlorophyll data at the Ocean Climate Laboratory is shown in Fig 1. Large gaps are observed in the central gyres of the Northern Hemisphere and south of the equator in the Southern Hemisphere for all basins. Fig. 2 shows the distribution of surface data as a function of year and season. For Fig. 2, all historical data were composited into four seasons regardless of year of observation. The seasons are defined as: Winter (January-March), Spring (April-June), Summer (July-September), and

Fall (October-December). The data spans from 1955 to 1990 with a big pulse in the 1980's. Most chlorophyll data, in our database, are for the spring season. These maps are important since they show possible bias in using an all season database to look at properties which have strong seasonal signals. Table 1 lists some of the contributors to the digital chlorophyll database.

Table 1. List of contributors to the digital chlorophyll database

SUBMITTING INVESTIGATOR	INSTITUTION	DESCRIPTION
W. Balch	Bigelow, U.S.	Historical/Global
C.R. McClain	GSFC/NASA, U.S.	Historical/Global
A. Longhurst	BIO, Canada	Historical/Global
H. Dooley	ICES, Denmark	Historical/N.Atlantic
J.S. Sarupria	NIO, India	Indian Ocean
J. O'Reilly	NOAA, U.S.	Eastern U.S.
S. Anderson	SIO, U.S.	CalCOFI
L. Codispoti	ODU, U.S.	Peru Upwelling
J. Sylvester	NODC, U.S.	North Atlantic
M. Luther	USF, U.S.	Historical/Indian
D. Karl	UH, U.S.	HOTS/N. Pacific
El-Sayed	TAMU, U.S.	Southern Ocean
BIOMASS	INTERNET	Southern Ocean
Cearex	NSIDC, U.S.	Eastern Arctic
BATS	NODC, U.S.	Sargasso Sea
JODC	Japan	Pacific Ocean
BOFS	BODC, U.K.	North Sea

The data were initially screened for extreme values to eliminate erroneous or unrepresentative data. Observed level data were then interpolated to standard depths and statistical checks were performed. All data at a standard level were zonally averaged in each one-degree latitude belt by individual ocean basin to produce analyzed mean chlorophyll values. Data were composited regardless of year/month of observation or method of collection.

This work is an attempt to represent large-scale permanent or semi-permanent features. North and equatorial Indian, Pacific, and Atlantic chlorophyll mean fields, on a one-by-one degree grid, are shown in Figs. 3-5. A grid point for which less than four one-degree square observations contributed to the analyzed value at the grid point is indicated by an "x" in the mean annual figures.

RESULTS AND DISCUSSION

Initial results show remarkable similarities between the satellite images and the annual mean analyzed fields, particularly in the North Indian Ocean (Fig. 3). High chlorophyll values are observed in the coastal regions decreasing with increasing latitudes and distance from the coast (Fig 3). Some of the differences observed are lower analyzed values in the equatorial regions (< 0.05 µg/l) and off the Somali coast. These differences could be due to a sampling bias since high chlorophyll values off the Somali coast are associated with upwelling during the monsoon period when sampling is infrequent. It is difficult to ascertain why these differences occur since not enough data are currently available.

The comparison between the CZCS images and the North Pacific mean fields are also very similar in the location of high and low pigment concentrations (see Fig. 4). Coastal and high latitude regions have a higher chlorophyll content than the central gyre which is depleted in chlorophyll a. Some of the discrepancies between the figures are lower annual mean chlorophyll values off the coast of Oregon, and higher analyzed values in the equatorial region. These differences may be due to the lack of *in-situ* data in many regions of the central North Pacific (see Fig. 1). Annual mean objectively analyzed fields for the North Atlantic are shown in Fig 5. Again, differences between this figure and CZCS images for the same region may be due to either a lack of *in-situ* data or bias in the seasonal coverage.

These results are based on data which have not been extensively quality controlled. Our aim is to fill in the data sparse areas by seeking out additional chlorophyll data and to develop adequate quality control procedures so we can more accurately compare the analyzed mean chlorophyll fields and satellite ocean color data. The ultimate goal is to produce seasonal maps showing the distributions of chlorophyll-a.

ACKNOWLEDGMENTS

We would like to thank all the contributors of data to the NODC, particularly W. Balch, C.W. McClain, A. Longhurst, M. Luther, H. Dooley, and J.S. Sarupria who have been compiling digital and manuscript pigment data from various principal investigators worldwide. We would also like to thank T. Boyer, T. O'Brien, L. Stathoplos, and D. Johnson for help in reformatting data containing chlorophyll. C. Forgy digitized some of the data in this database.

REFERENCES

Boyer, T.P. and S. Levitus. 1994. Quality control and processing of historical temperature, salinity and oxygen data. NOAA Technical Report NESDIS 81. 65 pp.

Conkright, M.E., T.P. Boyer and S. Levitus. 1994a. Quality control and processing of historical nutrient data. NOAA Technical Report NESDIS 79. 75 pp.

Conkright, M.E., S. Levitus and T.P. Boyer. 1994b. World Ocean Atlas 1994 Volume 1: Nutrients. NOAA Atlas NESDIS 1. 150 pp.

Feldman, G., N. Kuring, W. Esaias, C.R. McClain, J. Elrod, N. Maynard, D. Endres, R. Evans, J. Brown, S. Walsh, M. Carle, and G. Podesta. 1989. Ocean colour: Availability of the global data set. EOS 70:634-641.

Levitus S. and T.P. Boyer. 1994a. World Ocean Atlas 1994 Volume 2: Oxygen. NOAA Atlas NESDIS 2. 186 pp.

Levitus S. and T.P. Boyer. 1994b. World Ocean Atlas 1994 Volume 4: Temperature. NOAA Atlas NESDIS 4. 117 pp.

Levitus S., R. Burgett and T.P. Boyer. 1994a. World Ocean Atlas 1994 Volume 3: Salinity. NOAA Atlas NESDIS 3. 99 pp.

Levitus S., T.P. Boyer and J. Antonov. 1994b. World Ocean Atlas 1994 Volume 5: Interannual variability of upper ocean thermal structure. NOAA Atlas NESDIS 5. 176 pp.

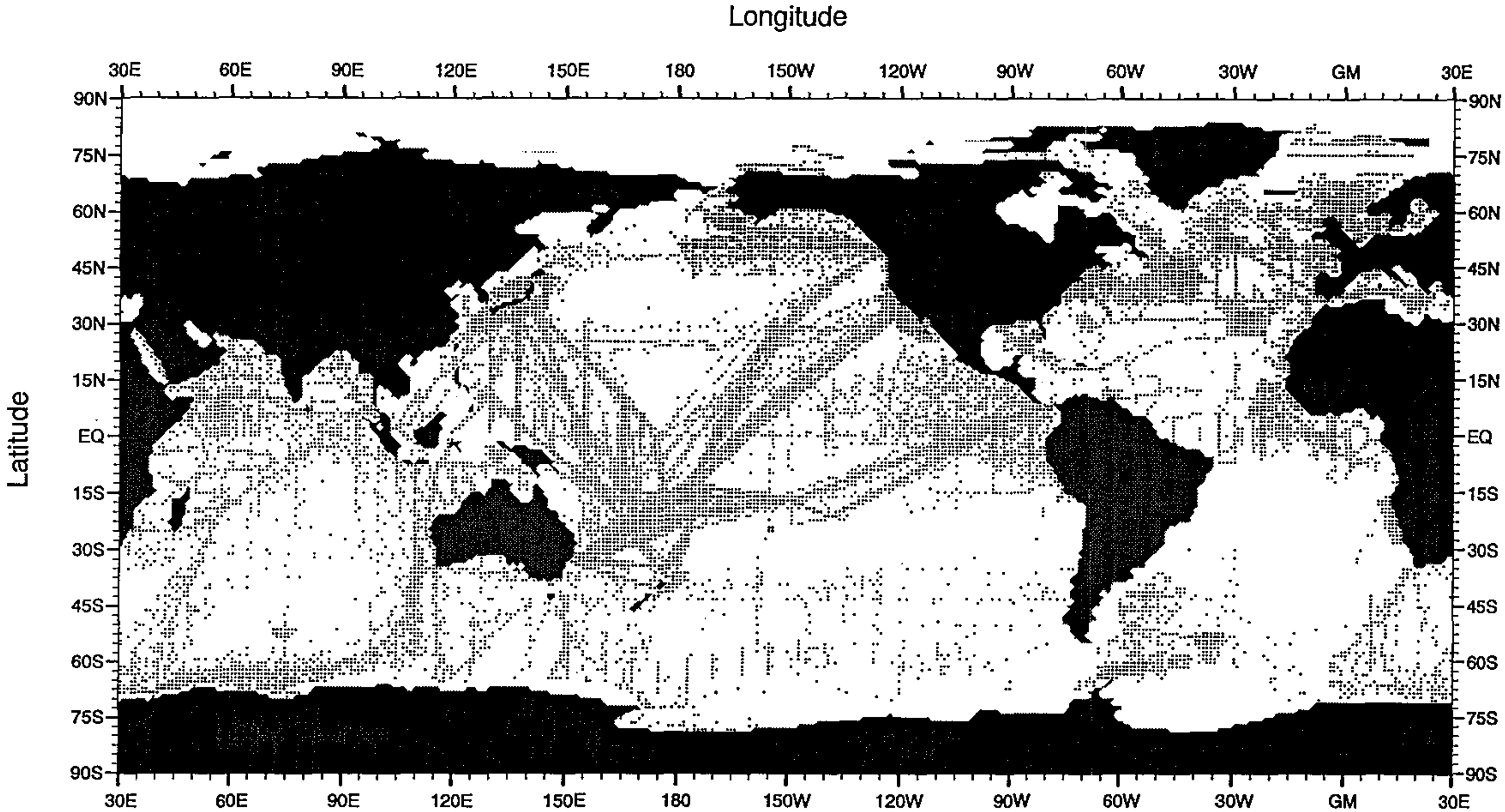


Fig 1. Distribution of Chlorophyll Data at the OCL

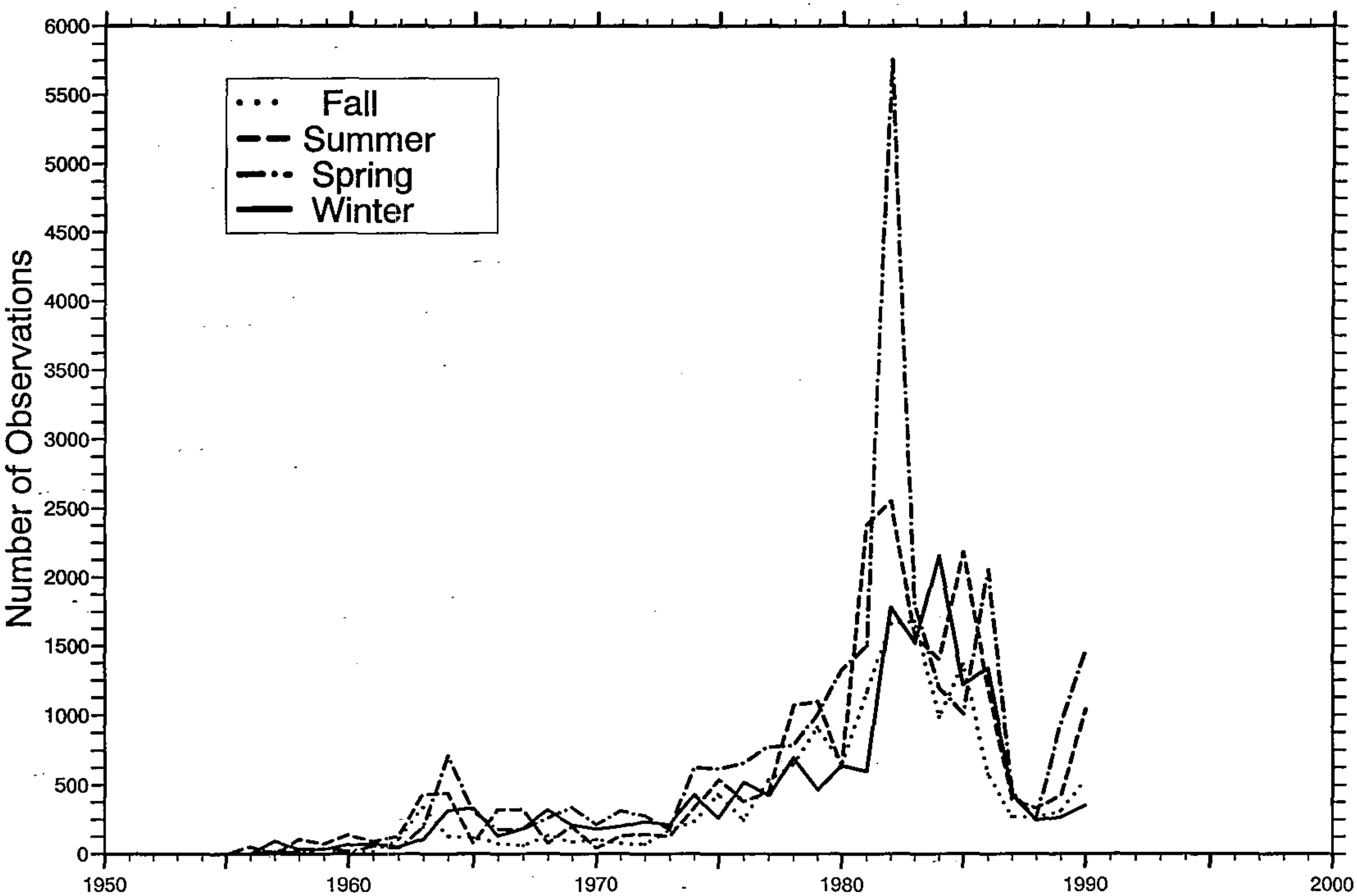


Figure 2: Time series of the total number of chlorophyll observations as a function of year and season for the sea surface

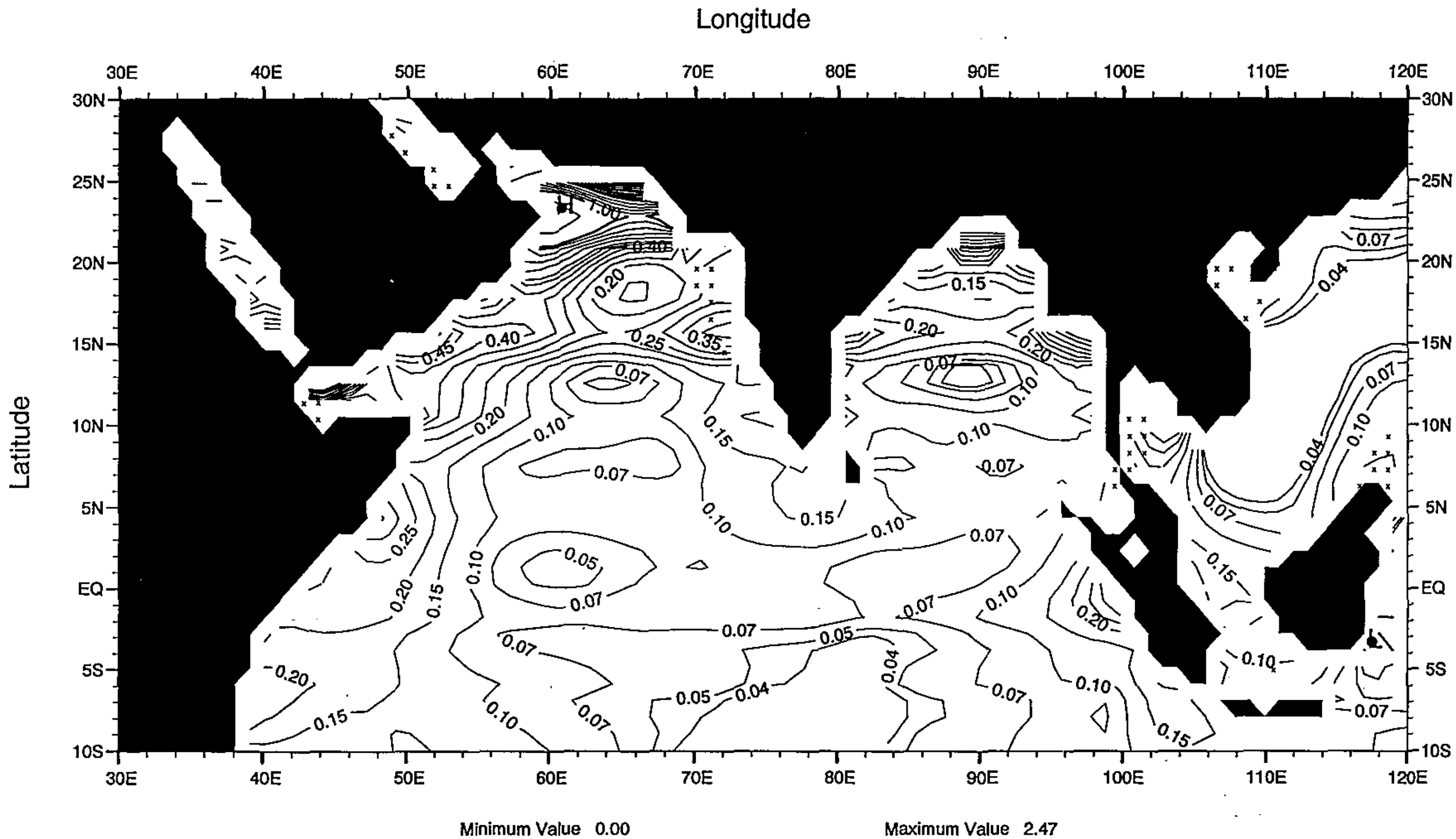


Fig 3. Annual mean surface cholorophyll in the North Indian Ocean

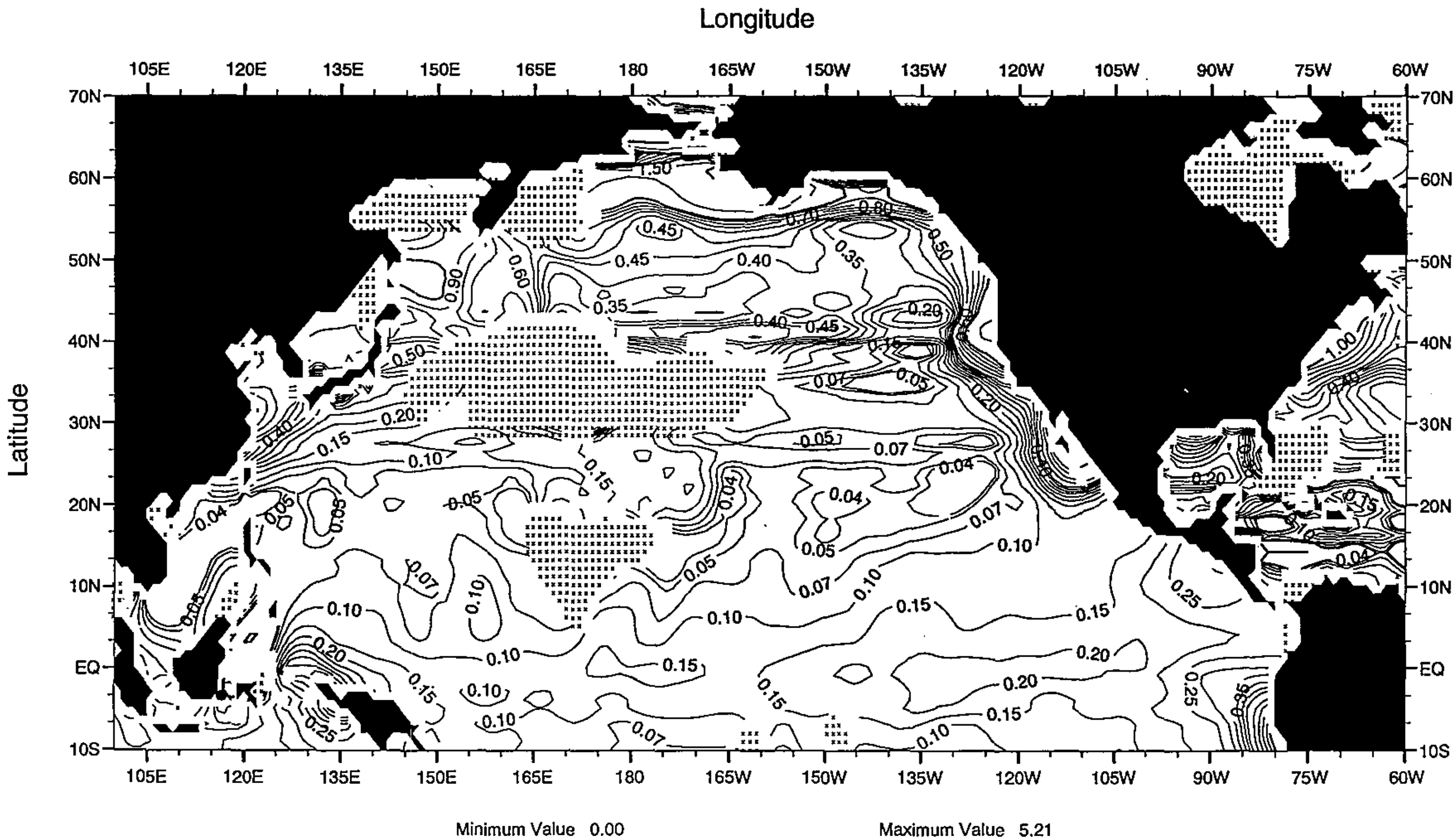


Fig 4. Annual mean surface cholorophyll in the North Pacific Ocean

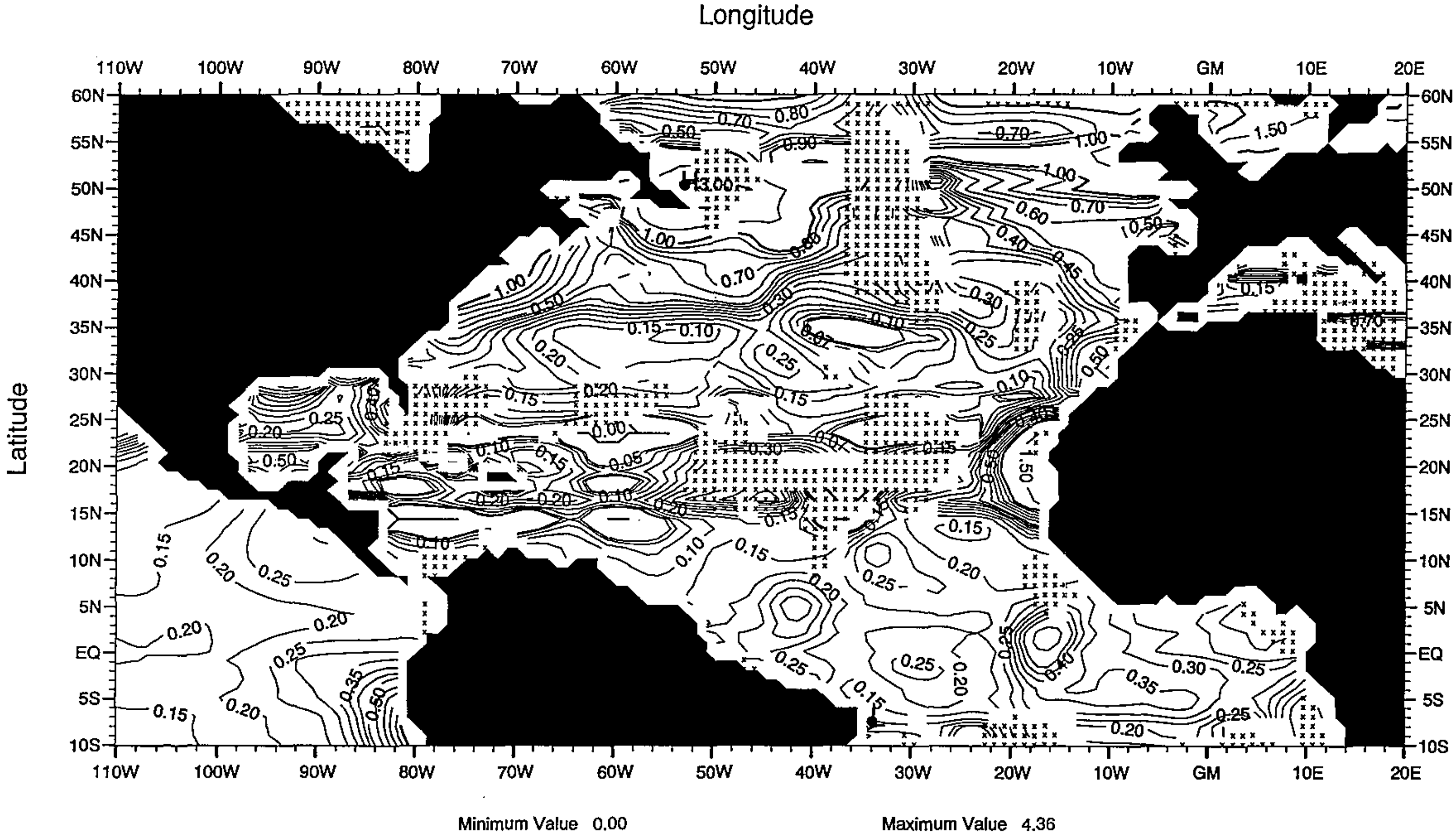


Fig 5. Annual mean surface cholorophyll in the North Atlantic Ocean

THE SPATIAL AND TEMPORAL VARIABILITY OF THE pCO_2 SYSTEM IN THE UPPER WATERS OF THE OCEAN

P.N. Makkaveev

INTRODUCTION

The carbon system in the waters of the World Ocean react to changes in both the physical and biochemical environment. Temperature, salinity and light intensity changes, as well as biota activity and biochemical processes, have an immediate influence on the equilibrium condition of the water, on the composition of the components in the carbon system, and on the correlation between the two.

The physical and biochemical characteristics of the water are subjected to changes at various scales which are influenced by the natural processes (i.e. seasonal changes and weather conditions). The changes can be periodic and they can be more or less accidental, and the composition of the carbon system components will reflect these changes.

The maintenance of dissolved inorganic carbon and its forms at every point in the Ocean must superimpose the exact daily, seasonal and other variations as well as the situational mixing. Due to the strong variability within the carbon system, any measurement of its components in the water characterizes only the state of the system at the exact moment of measurement (Bordovskiy, Makkaveev; 1995).

Now days, it is not sufficient to study only the spatial distribution of the carbon system. One must pay attention to the maintenance of the temporal variability and of the dissolved inorganic carbon, especially in the upper layer of the Ocean.

MAIN GOALS

There are three main reasons for this investigation:

1. Without knowing the main features of the natural temporal variability of the maintenance of the carbon system in the Ocean, it is difficult to judge about the spatial distribution of dissolved inorganic carbon and its forms because of the irregularity of the distribution of the observation over time.

2. As a result of man's activity, a considerable quantity of carbon dioxide is set free and is entranced into the atmosphere of Earth. About the half of the man-made carbon dioxide is absorbed by the Ocean, but we can not say definitely whether or not the changes in the condition of the carbon equilibrium in some regions of the Ocean are connected with this man-made influence, or if we are

just observing the natural variations of the carbon system. We need to be able to distinguish natural variation and the man-made components of these changes.

3. The knowledge of the peculiarity of the changes of carbon system in various regions of the Ocean can be extremely important for studying and precisely defining the quantity of the biological activity of the Ocean. Thermodynamical equations of the carbon equilibrium allow us (with sufficient quantity of data) to calculate possible changes caused by the hydrophysical factors and estimate the degree of influence of the biochemical factors with the bigger degree of reliability (Bordovskiy, Makkaveev; 1995).

DATA

For the studying of the peculiarities of variability of carbon equilibrium over time, it is desirable and important to examine the long time series of the observations. In spite of the availability of new direct methods for the determination of the carbon system components in recent years, the majority of historical observations have presented the estimations of pH and alkalinity data.

In spite of their imperfection, pH and alkalinity data covered the period from the almost beginning of this century and in some cases it can be of decisive importance for our investigation. However, when using historical data, a series of obstacles arise. It is necessary to compare and combine into the united massif data carried out from different time periods, by various methods, and by the application of different standards and different scales. For example, in the period of 1930-1960, scientific expeditions of the former USSR used four different practical pH scales. Great difficulties are also created by the existence of historical data of poor quality and those containing mistakes.

For more than 10 years, we have been processing archived, historical, hydrochemical data. This work has not been completed yet, but some preliminary results can be placed at your disposal. The data massif of pH and alkalinity in surface waters, including data from the 1920's until present, is practically completed (Table 1).

Table 1. The quantity of data of dissolved oxygen, pH & total alkalinity in the surface waters.

Ocean	Oxygen	pH	Alk
Atlantic	75,811	20,039	1,785
Indian	19,321	2,793	1,075
Pacific	90,363	24,950	5,673

At this time, I will not discuss in detail the methods used for quality-checking and correction of data. I'll say only that the work was carried out in several stages. During the first stage the data was let through a number of filters where the obvious data of poor quality or errors were cut off. On the other stages the rejection of spoilage and correction of data was done mainly by the manual methods.

Table 2. The quantity of data on the pH & total alkalinity in the surface waters.

	North Pacific		South Pacific		North Atlantic		South Atlantic		North Indian		South Indian	
Month	pH	Alk	pH	Alk	pH	Alk	pH	Alk	pH	Alk	pH	Alk
1	1378	307	730	121	807	43	255	15	62	49	272	147
2	2451	333	574	92	1233	54	174	47	122	30	258	124
3	2197	378	381	165	1562	81	187	106	140	26	182	50
4	1753	316	80	7	1977	131	171	22	176	2	81	58
5	2362	494	35	0	1713	128	109	0	86	47	169	60
6	1740	540	135	17	1898	182	180	0	223	128	147	43
7	2282	544	183	22	1797	163	151	0	44	42	77	29
8	3476	728	85	59	1646	213	38	0	38	34	144	30
9	1760	593	145	90	1552	128	107	23	68	0	56	0
10	2354	593	241	70	1614	206	258	6	1	0	157	52
11	2871	641	377	125	1244	157	270	0	0	0	107	46
12	1861	464	519	162	807	72	277	8	16	11	167	67

Analysis of the data distribution shows that the majority of data covers the summer season and the Northern Hemisphere (Table 2). Based on this massive, the values of pCO_2 and total carbon in surface waters were calculated (Millero, 1995).

SEASONAL MOTION

For studying the seasonal motion of carbon system, we calculated the average monthly data for latitude belts and for some areas where the quantity of date was enough. In Figures 1-3, you can see the variability of pCO_2 for the Atlantic, Indian and Pacific oceans. The dark areas in the figures mark areas where the pCO_2 in the water was more than in atmosphere. If we compare the changing of pCO_2 in Ocean surface waters and in atmosphere (Conway et al., 1994), the seasonal range of the pCO_2 in the surface waters is more than 10 times as much as in atmosphere.

The range of the seasonal variability of pCO_2 depends mostly on latitude. For all latitudes of the oceans, the existence of several minimums of the pCO_2 values is typical. The one or two minimums in the spring-summer period is connected with the plankton bloom, while the one minimum in winter is apparently connected with the increased dissolving of CO2 in the cooling water. The most complicated motion of the pCO_2 is present in the middle latitudes. In the low and high latitudes, the range of the interannual variability is decreased (Bordovskiy, Makkaveev, 1991).

In Figures 1-3, the characteristic features of all the oceans are also well seen. Most characteristic of the Pacific Ocean (Fig. 1) is the existence of a zone of constant high pCO_2 in the equatorial area (Bordovskiy, Makkaveev, 1991). For the Atlantic Ocean (Fig. 2), the sufficiently chaotic distribution of pCO_2 is typical. This is connected with the relatively short length of ocean and also with the great influence of the coasts and coastal currents. For the Indian ocean (Fig. 3), there is a sufficiently complicated structure of interior motion of pCO_2 in the northern part, where coastal influence and the influence of the monsoon climate is great.

Figure 1: Seasonal variability of the surface water pCO_2 at different latitudes in the Pacific ocean.

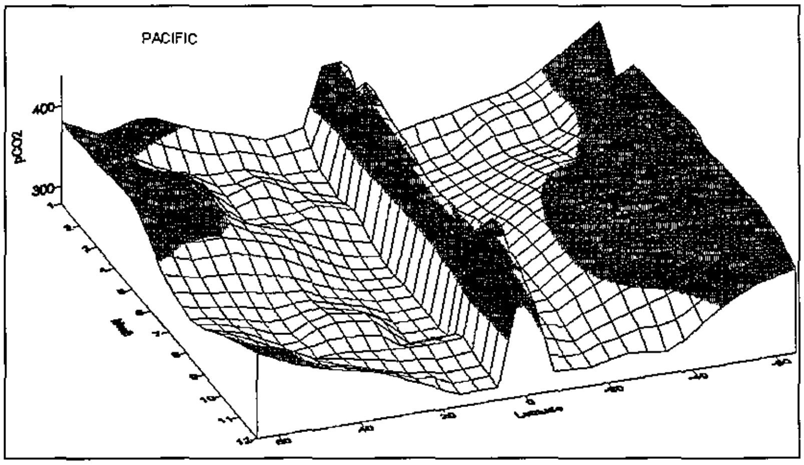


Figure 2: Seasonal variability of the surface water pCO_2 at different latitudes in the Atlantic ocean.

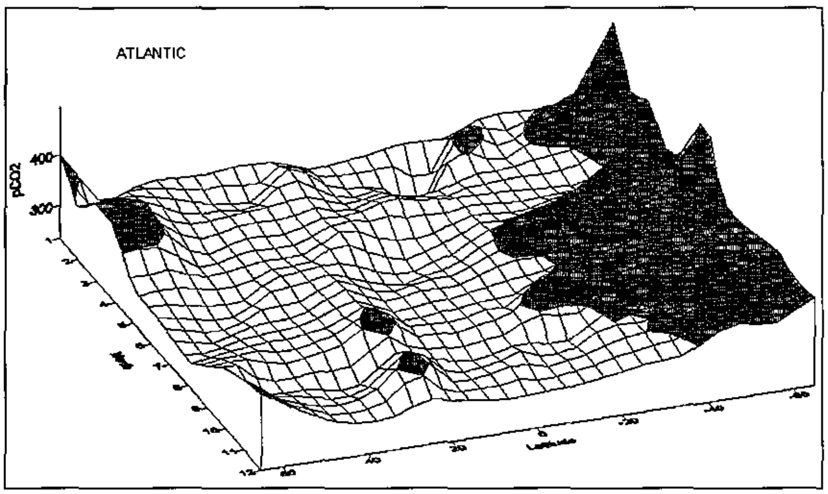
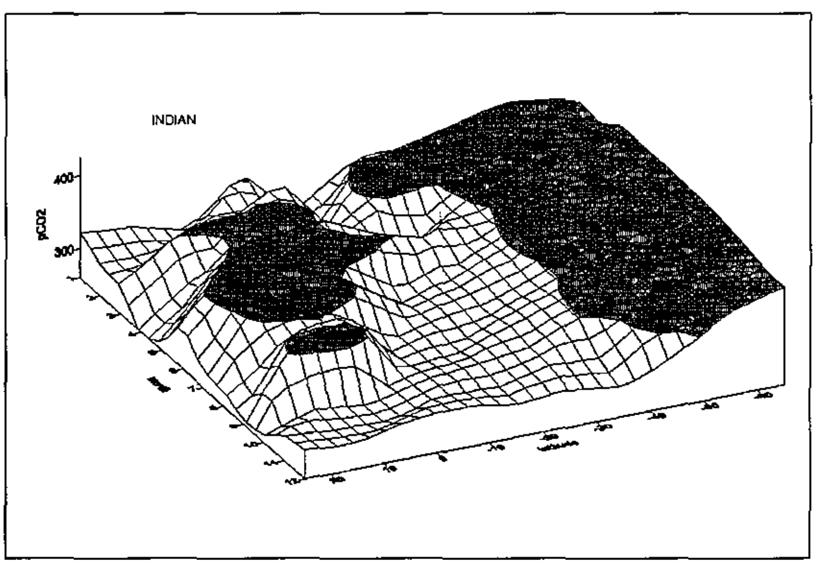


Figure 3: Seasonal variability of the surface water pCO_2 at different latitudes in the Indian ocean.



(For all figures, positive latitude corresponds to Northern hemisphere, negative to Southern).

DAILY MOTIONS

The apportionment of daily motion in the carbon system components is considerably complicated because it is, in a majority of the cases, disguised by the variations of the other time scales and non-tidal changes.

The range and the peculiarities of the daily variations of carbon system components change both from place and with the season. It is obvious that in the period of the active photosynthesis, the daily variations will be higher than in other periods, and its character in the higher latitudes will considerably differ from in the tropics. Thus, for its calculation, it is necessary to examine the peculiarities of the daily motion for different seasons and different latitudinal zones. It some cases, the quantity of the daily variations will probably be insignificant, but in other cases its quantity will be highly essential and excel not only the accuracy of the determination but possibly its spatial changes as well.

CONCLUSION

The continue this study requires two steps: First, the collection and analysis of all available data on the carbon system of the Ocean waters must be compiled and, in regions where these data are abundant, we must determine the general peculiarities of temporal variability at different scales for these regions. Second, these generalizations are to be used for the development of the mathematical models which will allow us to get information about the quantity of variability for these and other regions provided by future data of observation.

This work was supported by grant from Russian Science Foundation N 96-05-65819.

REFERENCES

Bordovskiy O.K., Makkaveev P.N. Variation of the CO2 in the Pacific ocean. Abstract of the 12th Japan - USSR Energy symposium. December 12-13, Tokyo, Japan. Tokai University, 1991, v.2, p.25-38.

Bordovskiy O.K., Makkaveev P.N. The Variability of the Carbon Dioxide in the Surface Ocean's Waters. Global Changes and Geography. The IGU conference. Moscow, Russia, August 14-18, 1995. Abstracts, p.49.

Millero F.J. Thermodynamics of the carbon dioxide system in oceans. 1995, Geochimica et Cosmochimica Acta, v.59, N.4, p.661-677.

Conway T.J., Tans P.P., Waterman L.S., Thoning K.W., Kitzis D.R., Masarie K.A. and Ni Zhang. Evidence for inter annual variability of carbon cycle from the National Oceanic and Atmospheric Administration/ Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network. Journal of Geophysical Research,1994, v. 99, N D11, p. 22831 - 22855.

INTERANNUAL-TO-DECADAL VARIABILITY OF THE TEMPERATURE-SALINITY STRUCTURE OF THE WORLD OCEAN

Sydney Levitus

ABSTRACT

1) During the past 20 years, observation-based scientific studies indicate that interannual-to-decadal-scale variability of T-S structure on gyre to basin scales has been observed in most major ocean basins:

    a) North Atlantic;
    b) Tropical Atlantic;
    c) North Pacific;
    d) South Pacific;
    e) Mediterranean.

    The success of some of these studies is due to data archaeology and rescue projects in various countries as described in IOC/GODAR reports. More than 1.2 million T profiles have been made available as a result of the IOC/GODAR project.

2) An additional 700,000 T profiles will be made available by the end of 1997 as a result of the GODAR project.

3) The size increase of the historical ocean data bases allows computation of quantities such as "annual heat storage anomalies" for the first time.

4) Historical data are being located at a faster rate then date are being digitized.

    Naval data from various countries are being declassified and are substantially increasing the size of the ocean data archives.

5) Thus the data base for identifying and describing interannual-to decadal-scale ocean variability can be expected to increase substantially.

REFERENCES

Antonov, J.: 1993, Linear trends of temperature at intermediate and deep layers of the North Atlantic Ocean and the North Pacific Ocean, *J. Climate*, 6, 1928-1942.

Bethoux, J.P., Gentili B., Raunet J., and Tailliez, D.: 1990, Warming trend in the western Mediterranean deep water, *Nature*, 347, 660-662.

Bindoff, N.L., and Church, J. A.: 1992, Warming of the water column in the southwest Pacific Ocean. *Nature*, 357, 59-62.

Bingham, F.M., Suga, T. and K. Hanawa, K.: 1992, Comparison of upper ocean thermal conditions in the western North pacific between two pentads: 1938-42 and 1978-92. *J. Oceanogr.*, 48:404-425.

Blindheim, J. and Skjoldal, H., R.: 1993, Effects of Climatic Changes on the Biomass yield of the Barents Sea, Norwegian Sea, and West Greenland Large Marine Ecosystem, *In* Large Marine Ecosystems ed. by K. Sherman, L.M. Alexander, and B.D. Gold, Amer. Assoc. Adv. Sci. Press, 185-198.

Brodeur, R.D. and Ware, D.M.,: 1992, Long-term variability in zooplankton biomass in the subarctic Pacific Ocean. *Fisheries Oceanogr.*, 1, 32-38.

Delcroix, T., and Henin, C.: 1989, Mechanisms of subsurface thermal structure and sea surface thermohaline variabilities in the southwestern tropical Pacific during 1975-85, *J. Mar. Res.*, 47, 777-812.

Delcroix, T., and Henin, C.: 1990, Sea surface thermohaline and subsurface thermal structure variabilities in the southwestern tropical Pacific Ocean during 1979-1985. *In* Air-Sea Interaction in Tropical western Pacific: Proceedings of U.S.-PRC International TOGA Symposium, 1988, Beijing.

Delcroix, T., and Henin, C.: 1991, Seasonal and interannual variations of sea surface salinity in the tropical Pacific Ocean, *J. Geophys. Res.-Oceans*, 96, 22135-22150.

Dickson, R., J. Meinke, S.A. Malmberg, and A.J. Lee: 1988, The great salinity anomaly in the northern North Atlantic, 1968-82. *Progr. Oceanogr.*, 20, 103-151.

Dooley, H.D., Martin, J.H.A., and Ellet, D.J.: 1984, Abnormal hydrographic conditions in the Northeast Atlantic during the 1970's. Rapp. P.v. Reun. *Cons. int. Explor. Mer.*, 185, 179-187.

Ezer, T., Mellor, G.L., and R.J. Greatbatch: 1995, On the interpentadal variability of the North Atlantic Ocean: Modelling changes in transport, meridional heat flux and coastal sea level between 1955-1959 and 1970-1974. Submitted to *J. Geophys. Res.-Oceans.*

Greatbatch, R., Fanning, A.F., Goulding, A.D., and Levitus, S.: 1991, A diagnosis of interpentadal circulation changes in the North Atlantic. *J. of Geophys. Res.-Oceans*, 96, 22009-220024.

Hanawa, K.: 1987, Interannual variations of the winter-time outcrop areas of Subtropical Mode Water in the western North Pacific, *Atmosphere-Ocean*, 25, 358-374.

Harris, G.P.: 1988, Interannual variability in climate and fisheries in Tasmania. *Nature*, 333(6175), 754-757.

Hecht, A.: 1992, Abrupt changes in the characteristics of Atlantic and Levantine intermediate waters in the Southeastern Levantine basin. *Oceanol. Acta*, 15, 25-42.

Lacombe, H., Tchernia, P., and Gamberoni, L.: 1985: Variable bottom water in the western Mediterranean basin. *Prog. in Oceanogr.*, 14, 319-338.

Lazier, J.R.N.: 1994, The salinity decrease in the Labrador Sea over the past thirty years. National Academy of Sciences Press, Natural Climate Variability on decade-to-century time-scales. In press.

Levitus, S.: 1989a, Interpentadal variability of temperature and salinity at intermediate depths of the North Atlantic Ocean, 1970-74 versus 1955-59. *J. Geophys. Res.-Oceans*, 94, 6091-6131.

Levitus, S:. 1989b, Interpentadal variability of salinity in the upper 150m of the North Atlantic Ocean, 1970-74 versus 1955-59. *J. Geophys. Res.-Oceans*, 94, 9679-9685.

Levitus, S:. 1989c, Interpentadal variability of temperature and salinity in the deep North Atlantic Ocean, 1970-74 versus 1955-59. *J. Geophys. Res.-Oceans*, 94, 16125-16131.

Levitus, S.: 1990, Interpentadal variability of steric sea level and geopotential thickness of the North Atlantic Ocean, 1970-74 versus 1955-59. *J. Geophys. Res.-Oceans*, 95, 5233-5238.

Levitus, S., Antonov, J., and Boyer, T.P.: 1994a, Interannual variability of temperature at 125 m depth in the North Atlantic Ocean, *Science*, 266, 96-99.

Levitus, S., Boyer, T.P., and Antonov, J.: 1994b, *World Ocean Atlas 1994, Volume 5: Interannual Variability of Upper Ocean Thermal Structure*. NOAA NESDIS Atlas series, In Press.

Levitus, S., Gelfeld, R., Boyer, T.P., and Johnson,D.: 1994c, Results of the NODC and IOC Data Archaeology and Rescue projects. Key to Oceanographic Records Documentation No. 19, National Oceanographic Data Center, Washington, D.C., pp. 67.

Levitus, S., Antonov, J., Zhou, X., H. Dooley, Selemenov, K. Tereschenkov, V.: 1996, Decadal-Scale Variability of the North Atlantic Ocean. National Academy of Sciences Press, Natural Climate Variability on Decade-to-Century Time scales. In press.

Levitus, S., and J. Antonov, 1995: Observational evidence of interannual to decadal scale variability of the temperature-salinity structure of the world ocean. *Climatic Change: Special Issue on Long-term Climate Monitoring*, 31, 495-514.

Lie, H.J., and M. Endoh, M.: 1991, Seasonal and interannual variability in temperature of the upper layer of the northwest Pacific, 1964-1983, *J.Phys. Oceanogr.*, 21, 85-397.

Parilla, G., Lavin, A., Bryden, H., Garcia, M., and Millard, R.: 1994, Rising temperatures in the subtropical North Atlantic, *Nature*.

Petrie, B. and Drinkwater, K.F.: 1993, Temperature and salinity variability on the Scotian Shelf and in the Gulf of Maine 1945-1990. *J of Geophys. Res.-Oceans*, 20079-20089.

Read, J.F., and Gould, W.J.: 1992, Cooling and freshening of the subpolar North Atlantic Ocean since the 1960s. *Nature*, 55-57.

Rohling, E.J., and Bryden, H.L.: 1992, Man-induced salinity and temperature increases inn the western Mediterranean deep water, *J. Geophys. Res.-Oceans*, 97, 11191-11198.

Ridgway, K. and Godfrey, S.,: 1995, Long term temperature and circulation changes in the East Australian Current, *in preparation*.

Roemmich, D., and Wunsch, C.: 1984, Apparent changes in the climatic state of the deep North Atlantic, *Nature*, 307, 447-450.

Royer, T.C.: 1989, Upper ocean temperature variability in the northeast pacific Ocean: Is it an indicator of global warming?, *J. of Geophys. Res.-Oceans*, 94, 18175-18183.

Suga, T. and Hanawa, K.: 1995, Interannual variation of North Pacific Subtropical Mode water in the 137 E Section, *J. Phys. Oceanogr.*, 25, 1012-1017.

Venrick, E.L., McGowan, J.A., Cayan, D.R., Hatward, T.L.: 1987, Climate and chlorophyll, *Science*, 238, 70-73.

Watanabe, T., and Mizuno, K.: 1994, Decadal changes in the thermal structure in the North pacific. *International WOCE Newsletter, 15*, unpublished manuscript, 10-12.

BIOGEOCHEMICAL MODELING: REQUIREMENTS FOR BIOLOGICAL AND CHEMICAL DATA

Watson W. Gregg

Biogeochemical modeling involves mathematically simulating various biological and geochemical constituents in an attempt to understand the cycling of these constituents and the processes affecting their distributions. The primary advantage of the mathematical simulation modeling approach over conventional observational or experimental methods is that processes that interact among various components can be more easily understood. Although a precise knowledge of these processes may still be elusive, at least a general understanding can be sought that cannot be achieved using other methods, and the limits can be observed. Furthermore, simulation modeling holds the potential for prediction: both in terms of the impacts of environmentally- and anthropogenically-induced changes on the cycles and distributions of biogeochemical constituents themselves, but also in terms of climate and human populations.

However, biogeochemical models developed without an intrinsic dependence on data, and without data for validation, can quickly become philosophical exercises with limited applicability to the real world and thus limited benefit. Coupled closely with actual data, simulation modeling is a powerful tool to help understand the complex manner in which biological, geochemical, and even physical processes interact.

Thus a rich, comprehensive data archive is essential to the development and application of simulation models. The requirement on data archives deriving from the simulation modeling activity is large: diverse data constituents are required, and spatially and temporally complete fields are necessary.

BACKGROUND

The types of biogeochemical models currently in use are diverse, ranging form very simple efforts to complex multiple component-multidisciplinary investigations. There is quite a wide disparity in techniques, reflecting the relative youth of the activity. But generally, they all contain a low trophic level biological component (usually phytoplankton, often represented by chlorophyll), at least one nutrient which is required for growth, light inputs also required for growth and nutrient depletion, and a second trophic level (zooplankton or bacteria) to regenerate nutrients and deplete the biomass of phytoplankton (Figure 1). Losses/gains are represented typically in terms of nitrogen, carbon, or chlorophyll, and required conversions are internal to the model; the choice of the basis constituent is up to the modeler.

The dimensionally simplest models are those that represent a single point, and simulate nutrient uptake, phytoplankton growth, and nutrient regeneration/phytoplankton death in a single dimension, time. An example is the seasonal cycle evaluation by Evans and Parslow (1985), which included the three primary biological/ chemical constituents (nitrogen, phytoplankton, and herbivores), and arbitrary changes in the mixed layer depth, which in term affected the available light and the concentrations of substances by forced entrainment/detrainment. Sometimes these models are pieced together to simulate the processes of a larger regions (e.g. Glover et al., 1994; Wroblewski et al., 1988). The individual stations are horizontally isolated from one another, and thus are discontinuous, but still many meaningful insights can be gained from the assemblage of observations.

A second type of model involves 2 dimensions: time, and depth. In these representations the interactions of the chemical and biological constituents are coupled to a dynamical mixed layer depth, which is forced by the atmosphere (specifically, wind stress causing vertical circulation and turbulence). These are commonly referred to as mixed layer models. Some examples are McGillicuddy (1995) and Marra and Ho (1993).

Finally, there are fully coupled, interactive 4-dimensional simulations, including a full physical development. These are usually applied over a significant region and the emphasis is on the coupling between physical, biological, and geochemical processes, and the impacts of changes in one or more component on the others. Some examples include Sarmiento et al (1993), Gregg and Walsh (1992), and Walsh et al (1988).

Recent models have expanded on the biological and geochemical components, including multiple nutrient pathways (Sarmiento et al., 1993), and a detrital pathway. Some have begun to adopt multiple phytoplankton groups (Bissett et al., 1994; Gregg and Walsh, 1992).

USES OF DATA IN BIOGEOCHEMICAL MODELS

There are 5 primary uses of data in biogeochemical models:

1) Initialization
2) Boundary conditions
3) Forcing
4) Validation
5) Assimilation

Since time is a common dimension to virtually all biogeochemical models, they are initial value problems, therefore requiring initial conditions to begin. Data on nutrients and biological constituents can provide these initial conditions, but many modelers prefer to begin with arbitrarily specified fields, so that the robustness of the model can be evaluated. However, model stability analyses, and efforts focusing on specific processes or regions may require accurate initial conditions.

Boundary conditions are of major importance to numerical simulation models. These dictate the input conditions deriving from outside the model space, and can be integral in the solution.

Using chemical and biological data as forcing functions is an important an important modeling activity, especially when one is seeking refinement of an individual process that is otherwise difficult to isolate.

However, most modelers are interested in the interplay among the chemical and biological variables, and do not wish to force one against the other.

Validation is probably the most important use of data by modeling activities. It is the validation step that determines if the model is on the right track, whether is it adequately representing the conditions simulated, and whether new processes need to be added or old ones modified. Any spatial or temporal sampling is useful for validation, but obviously complete coverage is ideal.

Assimilation of data into models is an established technique for meteorology, but is relatively new to oceanographic models. It is even rarer in the field of biogeochemical modeling. Good assimilation requires comprehensive temporal and spatial sampling, and this condition is usually not met with in-situ data, given the sparseness of ship and buoy observations. Satellite data can potentially support the activity, but an operational biological satellite sensor is probably required.

DATA REQUIREMENTS FOR BIOGEOCHEMICAL MODELS

Given the data requirements for simulation models, the time and space scales can be established. For time, data are required as frequently as possible, but monthly and annual fields are necessary. Seasonal fields are often desired, but the difficulty of defining seasons, plus the practicality of using monthly fields to reasonably well define seasons suggests that they are not necessary for a routine data product. For space, global gridded data provide the regular sampling model models need, and the global nature enables the modeler to choose his area of emphasis. Modelers using 1-dimensional and mixed layer models might prefer localized data values, so it might be advantageous to also keep the raw values. As regards gridded data, it is essential to use a common and accurate land/ocean mask, and the National Geophysical Data Center (NGDC) provides such, with the ETOPO5 5-minute global elevation data file.

The relevant question is, however, what biological and chemical data are required for biogeochemical models? Given the known paucity of biological and chemical data (there is simply no analog in biological or chemical observation systems to temperature and salinity sampling), and given that most models are recent efforts, and the full data requirements are not completely established, the simplest answer is everything that is available. More specifically:

- all major nutrients;
- organic and inorganic carbon sources up to the second trophic level;
- biological rate measurements;
- oceanic optically active constituents.

However, this is an unsatisfactory answer to the developers of an incipient data archive and furthermore not a full one given knowledge of the state of science in biogeochemical modeling. Thus at a minimum we can say that the following data sets are of utmost importance.

Nutrients	Carbon	Rate Measurements
nitrate	chlorophyll	primary production
ammonium		

This is the minimum set needed most urgently for biogeochemical modeling. Although it excludes many important variables, and some whose importance has yet been realized, given the current state of simulation modeling, comprehensive data archives of these variables would be extremely helpful.

A list of secondarily important biological and chemical constituents is listed in Table 1. For most of these the importance has not been firmly established, or models are simply struggling with the limited state variables that they presently include, and cannot incorporate new variables. Or simply the models are emphasizing specific components in order to provide a better understanding of the interactions. In any event, production of data archives of the variables in Table 1 may be of great use in the future, and certainly of importance for limited studies presently, and thus acquisition and archival should be strongly considered.

A final area of data that is presently not used extensively in numerical models, but has importance with respect to light availability on the water column are optically-active oceanic constituents (Table 2). In addition to their use as helping to provide a refined underwater light field to enable more accurate assessments of biological activity, they are key variables for remote sensing. Thus an archive containing these variables can help provide better pigment and chlorophyll algorithms from spaceborne sensors, and thus assist modeling efforts that utilize these data.

Table 1. Biological and chemical data constituents of secondary importance for biogeochemical models.

Nutrients	Carbon	Rates
silicate	DOC (dissolved organic carbon)	export production
phosphate	accessory pigments	sinking fluxes
micro-nutrients	phycobilipigments	(sediment traps)
	carotenoids	
	phaeophytin	
	phytoplankton species/ group compositions	
	phytoplankton size fractions	
	detrital material	
	bacteria	
	zooplankton	
	calcium carbonate	

Table 2. Optical constituents of importance to biogeochemical models.

PAR (photosynthetically available radiation)
Spectral/integrated irradiance
CDOM (colored dissolved organic matter)
Total suspended material
Attenuation coefficients

REFERENCES

Bissett, W.P., M.B. Meyers, J.J. Walsh, and F.E. Muller-Karger, 1994. The effects of temporal variability of mixed layer depth on primary productivity around Bermuda. J. Geophys. Res. 99: 7539-7553.

Evans, G.T. and J.S. Parslow, 1985. A model of annual plankton cycles. Biol. Oceanogr. 3: 327-347.

Glover, D.M., J.S. Wroblewski, and C.R. McClain, 1994. Dynamics of the transition zone in coastal zone color scanner-sensed ocean color in the North pacific during oceanographic spring. J. Geophys. Res. 99: 7501-7512.

Gregg, W.W. and J.J. Walsh, 1992. Simulation of the 1979 spring bloom in the Mid-Atlantic Bight: A coupled physical/biological/optical model. J. Geophys. Res. 97: 5723-5743.

Marra, J. and C. Ho, 1993. Initiation of the spring bloom in the northeast Atlantic: A numerical simulation. Deep-Sea Res. 40: 55-73.

McGillicuddy, D.J., J.J. McCarthy, and A.R. Robinson, 1995. Coupled physical and biological modeling of the spring bloom in the North Atlantic (I): Model formulation and one-dimensional bloom processes. Deep-Sea Res. 42: 1313-1357.

Sarmiento, J.L., R.D. Slater, M.J.R. Fasham, H.W. Ducklow,, J.R. Toggweiler, and G.T. Evans, 1993. A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone. Glob. Biogeochem. Cycles 7: 417-450.

Walsh, J.J., D.A. Dieterle, and M.B. Meyers, 1988. A simulation analysis of the fate of phytoplankton within the Mid-Atlantic Bight. Cont. Shelf Res. 8: 757-788.

Wroblewski, J.S., J.L. Sarmiento, and G.R. Flierl, 1988. An ocean basin scale model of plankton dynamics in the North Atlantic 1. Solutions for the climatological oceanographic conditions in May. Glob. Biogeochem. Cycles 2: 199-218.

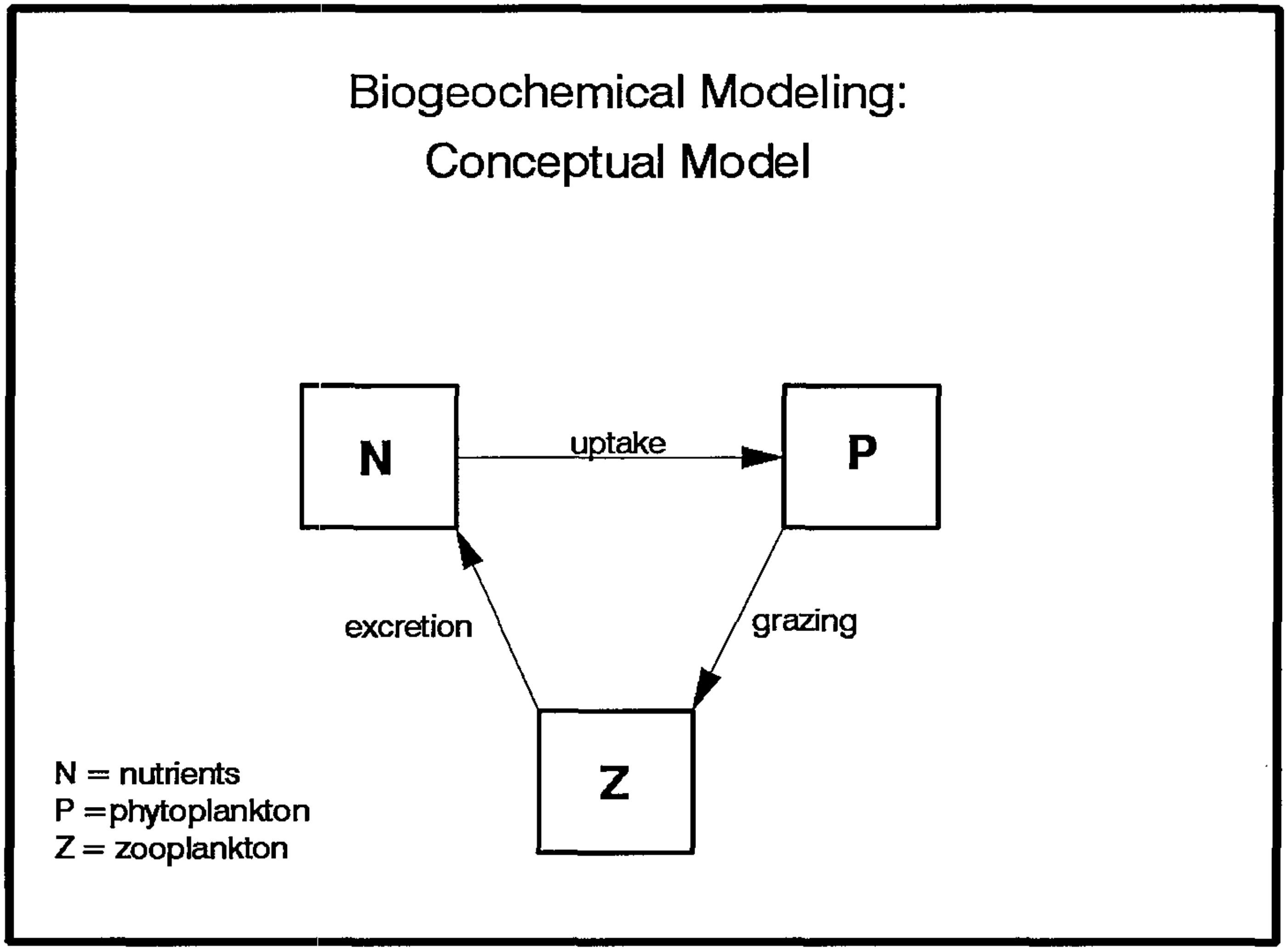


Figure 1. Generalized, simplified, conceptual diagram of most biogeochemical numerical simulation models. They typically contain three general components, that interact with one another: a nutrient group, a phytoplankton or producer group, and a zooplankton or grazer group.

SESSION III: INTERNATIONAL COOPERATION IN DATA MANAGEMENT OF BIOLOGICAL AND CHEMICAL OCEANOGRAPHIC DATA

Convener - William M. Balch (USA)

CURRENT AND PAST ICES ACTIVITIES IN CHEMICAL AND BIOLOGICAL OCEANOGRAPHIC DATA

H. D. Dooley

ABSTRACT

The Secretariat of the International Council for the Exploration of the Sea, based in Copenhagen, Denmark has housed an oceanographic data centre since the formation of ICES in 1902. The early activities of this centre focused very much on the physical, chemical and biological oceanographic data collected on surveys conducted by its member countries. These surveys extended across the North Atlantic, and also into the North Sea and Baltic Sea. The results of these surveys were published by ICES, latterly in the ICES Bulletin Hydrographique and Bulletin Planktonique, the former for more than 50 years at 3 monthly intervals. The Bulletin Planktonique survived until 1912, and this ended any formal and routine ICES activities in the management of biological oceanographic data. These Bulletins remain to this day a valuable record of all aspects of oceanographic data collections, and includes detailed records of techniques and methods deployed in the collection of these data.

During the period of the Bulletin Hydrographique, various chemical parameters were routinely measured, in particular oxygen, nitrogen, silicon and phosphorous and carbonic acid. In the late 1930s this parameter range was extended to include many other parameters, and this was extended further in the late 1960s. Presently most nutrient parameters, oxygen, alkalinity, pH, hydrogen sulphide, and chlorophyll a (which is strictly speaking a biological parameter) are managed by ICES. Some of this parameter list reflected the interests of the many Baltic countries that are ICES member countries. By this time (late 1960s)m ICES no longer published these data, but stored them in computer compatible form (punch cards) until the late 1970s, when they were maintained as a computer database. However to this day, the basic structure of the data set has remained unaltered.

In the late 1960s, following the formation of the World Data Centres, ICES activities in handling oceanographic data sets declined as this was now the responsibility of national data centres. Instead ICES was to concentrate on the reporting of information about the collection of data from research vessels, which evolved from the introduction of IOC's ROSCOP (Cruise Summary Report) system. This system provides the basis of meta-information about all types of marine data, including chemical and biological oceanographic data which comprise almost half of the seventy-five parameters included on ROSCOP. Of the ca 24,000 ROSCOP forms held at ICES, more than a third include information about chemical parameters, and a quarter have information about biological parameters. As well as providing a valuable data tracking tool for ICES, it is also a valuable source of information about the potential volume of data that is likely to be available should there be desire to create a global archive of particular parameters. The most common chemical oceanographic measurement is, unsurprisingly, oxygen which is reported on some 27% of cruises, whereas carbon dioxide is collected on less that 2% of cruises. The main nutrient salts are also well-represented, ranging from 26% of cruises for phosphate, to 20% for silicate. However other parameters for which there is current demand, for

example Total Nitrogen, Total Phosphorous, and Ammonia, are collected on less than 10% of cruises, and many of these originate from the Baltic. In the case of biological parameters, chlorophyll 'a' is reported on 16% of occasions, a similar amount for phyto and zoo-plankton. Biochemical observations are also well -represented (10%) whereas Dissolved and Particulate organic matter are observed on less than 5% of occasions. This information provides some insight into where efforts for historical data gathering maybe the most useful. However in most cases it is likely that many of these data, if not already submitted to data centres are either irretrievably lost, or not well enough documented to be useful. This is exacerbated by the fact that many data centres have not actively sought chemical oceanographic data for many years, preferring to concentrate on purely physical data types. As mentioned above, ICES has supported many chemical data types for many years and, although it has actively sought submission of all these types, it has proven difficult to obtain many of them. For example, although chlorophyll 'a' has been a reporting parameter for more than 30 years, it has only been possible to routinely acquire such data from two institutes, although many more have collected these data. ICES has however never sought other biological oceanographic data types, mainly because of potential problems arising from intercomparability of data. This has been particularly the case for Primary Productivity which, although of potentially great value to ICES needs, has never been considered to be reliable enough for collating into an international database.

Almost half of the oceanographic data reported to ICES at the present time include nutrient and oxygen data. Submission of any physical oceanographic data collected at the same time is also encouraged, as are ROSCOP forms completed with all relevant meta-information. These provided essential elements for quality control of chemical data. This is because a common problem is the submission of the wrong parameter name due to transcription errors by the originator. For example in a recent submission the silicate data turned out be in fact nitrite data, the denouement being aided by incompatible ROSCOP entries. The essential part of ICES quality control of chemical data is by way of property-property plots in which the physical oceanographic properties play an essential role. For open ocean data in particular plots of N:P and N:S is an excellent test for the quality of nutrient data. Such plots should exhibit low scatter and reflect the value of the corresponding Redfield ratio. Project data sets also require stringent quality control, and ICES experience suggest that currently there are many reasons for concern about the quality of chemical oceanographic data. For example in a recent Baltic research programme in which 22 research vessels were involved, calibration and field comparisons indicated a systematic offset of 52% in nitrate measurements. In the case of JGOFS, although many of the measurements are clearly of the highest quality, the overall quality of the data set is low because of the poor measurement capabilities of one of the participating countries. This occurred in spite of alleged strict adherence to JGOFS protocols, and careful intercalibration exercises.

For many years ICES has placed much emphasis on the need for intercalibration exercises for all parameters it maintains in its data banks. This has occurred most intensively for nutrient intercalibrations. Since 1978 five intercalibration exercises have been held, and all of these have demonstrated that there are good grounds for concern. Improvement is slow, however, as revealed by the JGOFS experience mentioned above, but the authors of the report of the latest ICES exercise voiced some note of optimism in their introductory statement: "More people than ever before are now inclined to enquire as to the quality of the results of chemical analyses before using them for their intended purpose, and 'Quality Assurance' is a phrase on everyone's lips.

What use is made of nutrient data in ICES Data banks? ICES data have in fact of restricted distribution for up to 10 years after collection. Thus all older data is publicly available and provides a substantial contribution to the data already published by WDC-A. Much of the newer data can be readily supplied to requesters, particularly if it is aggregated form. This is indeed a convenient form as it meets the need for numerical modelers engaged in studies, of, for example, nutrient trends. However the main user is ICES itself in order to help it meets its obligations in support of the requirements of the Pollution Regulatory Commissions. With this remit, analyses produced by ICES have provided a substantial contribution to the work of the North Sea Task Force which culminated in the publication of a Quality Status Report in 1993.

ICES is indebted to the co-operation and dedication of many data managers and scientists who actively and faithfully contribute the ICES data bank, and who patiently respond to sometimes painful questions about procedures and quality. Without the support of such individuals the maintenance of a large and diverse data bank would be unthinkable.

BIOLOGICAL AND CHEMICAL DATA MANAGEMENT IN JGOFS

Polly Machin and Roy Lowry

ABSTRACT

Since 1988, the British Oceanographic Data Centre (BODC) has practiced project oriented data management in support of large scale oceanographic field programmes in the U.K. Most of the work of adapting BODC's data management approach to encompass a very diverse biological and chemical field data set was pioneered during the Biogeochemical Ocean Flux Study (BOFS), the U.K. contribution to the Joint Global Ocean Flux Study (JGOFS).

The nature of the science in the BOFS programme required BODC to develop a strategy for the integration and organization of a very diverse and complex data set, arising from a number of different sampling regimes. This was achieved using relational database technology to create a logical model of the sampling environment encountered on a research cruise.

The data management practices established during BOFS have resulted in the creation of an on-line database which has subsequently been electronically published on CD-ROM. This is widely regarded throughout the international JGOFS community as a valuable resource.

1) INTRODUCTION

1.1) The JGOFS Programme

The Joint Global Ocean Flux Study was set up in the mid 1980's with the following scientific objectives:

- to determine and understand on a global scale the processes controlling the time-varying fluxes of carbon and associated biogenic elements in the ocean and to evaluate the related exchanges with the atmosphere, sea floor and continental boundaries.

- to develop a capacity to predict on a global scale the response of oceanic biogeochemical processes to anthropogenic perturbations, in particular those related to climate change.

In order to achieve these objectives field measurements of a wide range of biological and chemical parameters were required. A basic subset of 20 parameters, known as the 'Level 1 Parameters', were specified. However, in practice these were but a fraction of the parameters measured in JGOFS.

The JGOFS field programme was targeted on four areas with very different biogeochemical signatures:

- North Atlantic: a major area of deep water formation with a strong biological seasonal signal.

- Southern Ocean: a complicated region of large fluxes, which are thought to be approximately in balance..
- Arabian Sea: a very productive region with extreme seasonal variation driven by monsoon reversals.
- Equatorial Pacific: a large pool of unused nutrients and interannual variability associated with El Niño.

The major U.K. contribution to JGOFS was the Biogeochemical Ocean Flux Study (BOFS) which comprised 11 North Atlantic cruises in 1989-1991 and two cruises to the Southern Ocean in 1992. A further contribution, Arabesque, was made in 1994 with two cruises in the Arabian Sea.

1.2) Data Management in JGOFS

Data management in JGOFS is the responsibility of individual national programmes. As a result the problem of managing biological and chemical data in JGOFS has been addressed by a number of established national organizations founded on very different infrastructures and working practices. The inevitable consequence has been a variable level of success in managing national JGOFS data sets.

To overcome this through heightened international co-operation, the JGOFS Scientific Steering Committee set up the JGOFS Data Management Task Team (DMTT). Initially attempts were made to achieve harmonization through imposition of standards on the national operations. These were not a great success. For example, exchange formats were specified but not widely adopted because they required high level language programming support that was not universally available.

More recently some degree of harmonization has been achieved through technology transfer and the general availability of coding systems for biological and chemical parameters. This will be reinforced in the near future by the provision of centralized and nationally volunteered resources to create and electronically publish integrated basin scale JGOFS data sets.

It is generally accepted throughout the JGOFS community that the two national data management initiatives that have most successfully come to terms with the complexities of JGOFS data have been the operations in the U.S. and the U.K.. These have taken radically different approaches to the problem, both of which deserve equal attention from those wishing to address the problem of biological and chemical data management. The scope of this paper is, however, limited to covering the approach adopted in the U.K. by the British Oceanographic Data Centre (BODC).

1.3) BODC: A Brief Introduction

BODC was formed from the data banking service of the Marine Information and Advisory Service in April 1989. An important component of BODC's mandate was to provide data management support to the major marine science projects supported by the Natural Environment Research Council, such as BOFS and Arabesque.

U.K. JGOFS data management was handled by a project team comprising two members of staff (one fully dedicated) supported by software development resource from the BODC infrastructure and 3 industrial training students, each working on the project for a year.

The hardware environment was initially an IBM 4381 mainframe under VM/CMS, supported by two UNIX workstations to provide the high speed graphics which were essential for data screening. In 1993, the mainframe was replaced by a system of networked UNIX workstations and PC clones running Windows.

The key element of the software environment is the Oracle relational database management system (RDBMS). In addition, a large quantity of application software, written in Fortran77 and Pro-Fortran, has been developed in-house. The standard application package in use on PCS is Microsoft Office Professional.

2) BIOLOGICAL AND CHEMICAL DATA: THE DATA MANAGEMENT PROBLEM

Biological and chemical data may be considered from a data management point of view to come in two types: automatically logged data and discrete data. Automatically logged data are collected using instruments such as fluorometers, dissolved oxygen electrodes and AutoAnalysers running in continuous mode. These may be connected to a ship's pumped sea water supply or mounted on a range of fixed, profiling or towed platforms. Discrete data result from measurements or experiments on samples collected using a wide range of techniques such as water bottles, nets, sediment traps and cores. This latter category forms by far the most significant component of any biogeochemical data set by any measure other than a byte count.

The automatically logged data provide relatively familiar territory to data managers brought up on physical data as they may generally be classified as either time or depth series. However, the successful handling of these data requires a generalized system with no limitations on the parameters that may be included. Automatically logged data, whether they be physical, chemical or biological, are by nature high volume with the current trend being for volumes to increase. If data management is to include quality control then the problem of visualizing high volume data and identifying any problems needs to be addressed.

Discrete biological and chemical data is viewed as a daunting task for many data managers. Why is this? We believe that the answer to this question has two components. The first perceived problem is the complexity of the relationships within these data. This has two aspects: the first is that it is essential to know where the data came from to make sense of them. For example, consider handling a chlorophyll measurement on material taken from the gut of a copepod taken from the stomach of a fish taken by a net trawl. The second is that a single data set may need to be interpreted within the context of a number of diverse data sets that frequently cross discipline boundaries. Relational database technology was designed to manage the relationships between data. If this technology is adopted then problems of relationships may be easily solved.

The second problem associated with discrete data is the extent to which it is comparable with other data sets, given that there are very few standard techniques and new methods of analysis are continually being developed. This can be split into two aspects: the problem of variable data quality and the level of additional information or metadata that must accompany the data if they are to be of value to future generations of scientists. There is a technical problem to address here, namely how to link data and metadata together. However, a far more serious problem is how the data management operation can obtain the expertise required to specify just what should be included in the metadata and what the appropriate level of quality control should be.

In the sections that follow we will consider how BODC addressed the problems described above in its management of the U.K. JGOFS data.

3) BODC MANAGEMENT PROCEDURES FOR AUTOMATICALLY LOGGED DATA

BODC has developed a sophisticated system for handling and archiving many types of automatically logged oceanographic data. The software systems in use have been designed to provide high user productivity and data throughput. In addition they are flexible, easy to maintain, robust, reliable and wherever possible generalized so as to be independent of the type of data handled. They are therefore ideally suited to handling automatically logged biological and chemical data.

The process of quality assurance begins with the process known as Transfer (Loch 1993; Lowry and Loch 1995). This is basically a reformatting operation to convert the data into BODC's internal format (PXF). The reformatting process incorporates numerous checks to detect and correct any errors or unexpected variations in the data. Additionally, the data are standardized, which ensures that, once the data have been reformatted into PXF the units and processing history are known.

The problem of quality control is addressed through visual inspection of the data displayed as a time series, vertical profile or distribution plot. This is because in practice it has not been possible to develop an objective, automated procedure which is sufficiently generalized to cope with a highly diverse data and also rigorous enough to conform to the levels of precision specified by the JGOFS protocols. Experience has shown that whilst manual inspection of data displays can never be totally objective, it is both highly accurate and cost effective.

Once data have been examined a mechanism is required to mark any problems that have been identified. BODC maintain a policy of non-destructive editing. Data values are never changed: any suspect, null or interpolated data are discriminated from good data by having an associated quality control flag set to a particular value. In this way, the quality control is totally reversible as the quality control flag can be changed at any stage, for example in response to feedback from project scientists using the data.

The manual inspection of large quantities of high volume data conjures up images of major resource requirements. This has been addressed in BODC through the development of a sophisticated software package known as SERPLO (SERies PLOtting) which is described in more detail in Rickards (1994). This will plot any data which has either time or depth/pressure as the independent variable. It allows rapid inspection of the data through smooth and flexible panning and zooming facilities on the independent variable axis. Any of the parameters measured can be viewed either individually or simultaneously in any combination specified by the operator. Any problem data values may be flagged either individually or as a group enclosed by a box. The scaling of each parameter may be set to a predetermined level which is particularly useful for ensuring that parameters are viewed on a scale appropriate to their target level of precision.

SERPLO also incorporates features that allow quality control to be carried out in context. For example, positional data such as a cruise track may be displayed simultaneously in a second window with data and position linked through cursors. Knowing the position of the ship when the data were collected significantly enhances the level of understanding underpinning quality control decisions.

4) BODC PROCEDURES FOR HANDLING DISCRETE DATA

Discrete sample data sets accounted for most of the vast diversity of data types encountered on the BOFS project and included anything from potentiometric measurements of total alkalinity to measurements of the ^{18}O isotope ratio of foraminiferans in Kasten cores. BODC's problem was to devise a strategy to integrate them into a coherent database.

4.1) The Oracle Database

The key to addressing the problem of representing the relationships within the data was the adoption of relational database technology. The application of this technology to multidisciplinary data management had been pioneered during BODC's work on the NERC North Sea Project. The structure developed proved adequate for these data (essentially CTD casts with an associated limited parameter water bottle data set) but needed significant enhancement in order to encompass the vast increase in diversity present in the BOFS data set. The underlying philosophy, described in more detail in Lowry and Cramer (1995) was to dynamically develop data structures to match the data supplied. The ability to create new tables and modify existing ones under Oracle without the overhead of a database dump and reload was absolutely crucial for the success of this strategy.

The database was modeled on a three-level hierarchical structure. The top level is the event description. Events are sampling activities incorporating both over-the-side deployments and samples of water being drawn from the ship's continuously pumped "non-toxic supply". The second level are data linkage tables. These tables implement the one-to-many relationship between the events and the data. In addition they contain information specific to each major type of event. For example, the table NETINDX gives information on zooplankton net hauls, detailing the mesh size, mouth area, volume of water filtered and open air ratio, where known. The third level are the data storage tables. These are generally simple structures containing a column for each parameter measured. However, for water bottle data a fully normalized structure incorporating parameter coding has been developed and this enhancement will be extended to core profile and net catch data in due course.

Figure 1 illustrates the structure of part of the database, showing how the linkages are established between events and data. The examples shown are for samples taken from a zooplankton net haul, a Kasten core and a water bottle. The parameter code included in table BOTDATA is explained in more detail below. Each water bottle fired there may own a dozen or so records in BOTDATA: one for each parameter measured. The table MESOMASS contains size-fractionated mesozooplankton biomass, carbon and nitrogen data. It is slightly unusual, in that there is a one-to-one relationship between it and the linkage table NETINDX because each measurement refers to the whole of the catch for each net haul. The table KASCORE contains profile data along Kasten cores, such as dry bulk density, porosity and the ^{18}O isotope ratios of foraminiferans such as *G. bulloides*. This has a one-to-many relationship with the linkage table COREINDX because the profile from a single core includes more than one point.

The database represents a logical model of the sampling environment encountered during a research cruise with hooks provided for each sample collected. Each measurement submitted to BODC may therefore be linked to the water bottle, sediment trap cup, SAP filter, zooplankton net haul or core from which it was taken.

4.2) Sample Data Incorporation Procedures

Experience gained during the North Sea Project data management database showed that it was critically important to establish a definitive set of space/time co-ordinate information as soon as possible after the research ship had docked. It is so easy for mistakes in station numbering, depth registration of bottles or the time/position of samples taken to be made on board. This became apparent during the North Sea Project when samples were received from various sources for over 20 different depths from a CTD cast fitted with a 12-bottle rosette.

The first task was therefore to devote considerable effort to obtaining copies of all the log sheets and ancillary information available from a cruise. This has proved surprisingly difficult, particularly after scientists have left the ship. Consequently, BODC personnel now participate as shipboard data managers on supported research cruises wherever possible. By collating all available sources of information and resolving any discrepancies an authoritative version of where and when each sampling event occurred was built up. This basic information has proved to be of fundamental importance to the project scientists and is heavily requested, particularly in the early stages of projects.

Once analysis was completed, sample data sets were submitted to BODC by the principal investigators. This really did happen because the database was generally regarded as an essential part of the project and therefore contributed to willingly, thanks to the success of the novel data management strategy adopted by BODC and described in Lowry (1992). When a data set arrived, the first problem to address was quality control. This was generally achieved through careful scrutiny using spreadsheets and extensive consultation with the data originator who is primarily responsible for data quality. Further quality control was achieved by encouraging feedback from users of the data who subjected it to more rigorous scrutiny through interpretation than is possible by considering any data set in isolation.

When the data had been checked, they were loaded (using an Oracle utility) into temporary holding tables. Once they were within the relational environment, linkages were established by matching sample space and time co-ordinates to those already present in the database. This was achieved in practice by assigning values to primary key fields included in the holding tables. When all the primary keys had been successfully set, the data were copied into their permanent storage tables.

4.3) Addressing the Problem of Metadata

The technical problem of linking metadata to the data values to which they pertain has been addressed in two ways. The first of these has been to include as much metadata as possible into parameter codes. Each parameter measured is assigned an 8-character parameter code, which has a two-level hierarchical structure. The first level, held in the first 4 characters identifies the parameter being measured. Each parameter then owns a set of 4-character subcodes which are stored in characters 5-8 of the full parameter code.

The first two characters of the subcodes specify the analytical technique that was used. This is followed by an indication of the phase measured (dissolved or particulate) and the details of any filtration that was carried out on the sample. For example, the available parameter codes for different measurements of "chlorophyll-a" are given in Table 1.

CODE	DESCRIPTION	METHOD	UNITS
CPHLFLP1	Fluorometric chlorophyll-a	Fluo. assay acetone extract. (GFF filtered)	milligrams/cubic metre
CPHLFLP3	Fluorometric chlorophyll-a	Fluo. assay acetone extract. (GFC filtered)	milligrams/cubic metre
CPHLFLP4	Fluorometric chlorophyll-a	Fluo. assay acetone extract. (sum size fracs. >0.2u)	milligrams/cubic metre
CPHLFLP5	Fluorometric chlorophyll-a	Fluo. assay acetone extract. (0.2 um pore filtered)	milligrams/cubic metre
CPHLFLPC	Fluorometric chlorophyll-a	Fluor. assay of acetone extraction (centrifuged)	milligrams/cubic metre
CPHLFLPZ	Fluorometric chlorophyll-a	Fluo. assay of acetone extraction (filter unspec.)	milligrams/cubic metre
CPHLHPP1	HPLC chlorophyll-a	HPLC assay of acetone extraction (GO/F filtered)	milligrams/cubic metre
CPHLHPPZ	HPLC chlorophyll-a	HPLC assay of acetone extraction (filter unspec.)	milligrams/cubic metre
CPHLPR	CTD chlorophyll	Calibrated in-situ fluorometer	milligrams/cubic metre
CPHLSPP1	Spectrophotometric chlorophyll-a (Lorenzen)	Spect. assay of acetone extraction (GO/F filtered)	milligrams/cubic metre

CPHLSPP3	Spectrophotometric chlorophyll-a (Lorenzen)	Spec. assay acetone extraction (GO/C filtered)	milligrams/cubic metre
CPHLSPPC	Spectrophotometric chlorophyll-a (Lorenzen)	Spect. assay of acetone extraction (centrifuged)	milligrams/cubic metre
CPHLSPPZ	Spectrophotometric chlorophyll-a (Lorenzen)	Spect. assay of acetone extraction (filter unspec.)	milligrams/cubic metre
CPHLSSP1	Spectrophotometric chlorophyll-a (SCOR)	Spect. assay of acetone extraction (GO/F filtered)	milligrams/cubic metre
CPHLSSPC	Spectrophotometric chlorophyll-a (SCOR)	Spect. assay of acetone extraction (centrifuged)	milligrams/cubic metre

Table 1: Parameter codes for different measurements of chlorophyll a

The code system is maintained as a 'dictionary' within the relational database with user-accessible interrogation software. The advantage of the hierarchical nature of the code is that users of the data are offered considerable flexibility through wildcards. For example, someone undertaking a large scale survey of the depth of the chlorophyll maximum may not be concerned about the comparability of absolute values. This user may retrieve from the database by specifying the parameter code CPHL* which will retrieve all chlorophyll-a data irrespective of how the analyses were done. Conversely, a scientist may choose to discriminate between different methods of analysis (e.g. CPHLHP* for HPLC derived chlorophyll-a) or filtering techniques if they want to focus on the detailed data intercomparison.

The second mechanism of metadata delivery is through plain language documentation. To date, this has only become fully available when the data sets have been electronically published on CD-ROM, although it was passed on to scientists requesting data. The database was dumped to the CD-ROM in 'kit form' with each table in the database present as a simple ASCII file. Each dump file then had an associated plain language document describing the sampling and analytical protocols in as much detail as possible, particularly any deviation from the JGOFS core measurement protocols (UNESCO, 1994). Information was also given on the quality of the data and warnings were given where a particular analytical problem was identified.

This approach worked well for the BOFS data set where each parameter was stored as a column of one of a series of data tables. However, with the advent of the normalized structure for water bottle data storage this approach would result in a single massive document covering every protocol for every parameter measured. It is certain that such a document would never be read. Solutions to this problem are currently under consideration in preparation for the publication of the European Union OMEX (Ocean Margin Exchange) data set in June 1997. Possible candidates are linking documents to individual data values through a field in each record of BOTDATA or preparation of hypertext documents with "hot" parameter codes.

Having considered the technical aspects of metadata, the problem of obtaining the necessary scientific expertise to ensure that the data have been properly quality controlled and to prepare metadata documents must now be addressed. BODC recruitment has always focussed on scientists with an aptitude for computing rather than computing specialists. With the move into supporting multidisciplinary projects such as BOFS, a special effort was made to bring the widest possible range of appropriate scientific skills into the group. The resulting team comprises a geologist, a biological oceanographer, a biogeochemist, an atmospheric chemist and two physicists: all are graduates and half have PhDs.

Beyond the data management team there is a massive resource that may be tapped to help prepare high quality documentation, namely the scientific community. This is facilitated in the U.K. through the close working relationships that have been forged between scientists and data managers over the past 7 years. However, we firmly believe that the resources within the scientific community cannot be tapped by a data centre unless there is scientific expertise within the data centre. The problems inherent in a computer scientist liaising with a biologist to produce a document on chlorophyll analysis are too painful to contemplate.

5) HAS THE APPROACH SUCCEEDED?

To judge success in an enterprise such as BODC's management of biological and chemical data, two questions must be asked.

I. Has the database been able to incorporate all types of biological and chemical data submitted to the data centre?

II. Have scientists been able to make use of those data?

The answer to the first question is a simple "yes". By basing the approach on a flexible database schema, it has proved possible to adapt to whatever the scientists have produced in the way of field measurements. The second question is, however, far more important and must therefore be answered in a little more detail.

BOFS data have been delivered to the scientific community through three mechanisms:

- BOFS scientists were given access to the database via a suitably configured account on the Bidston computers supported by a Users' Guide which detailed the structure of the database and described the available retrieval software. This was complemented by a series of 3 day courses held at Bidston for scientists wishing to interrogate the database.

- Electronic publication of the BOFS North Atlantic data on CD-ROM in April including a 'kit-form database' (fully portable onto any system with a RDBMS), satellite images and X-ray images of Kasten cores, a software interface designed to smooth the passage from the data to commercial software and a comprehensive manual.

- Supplying data in response to specific requests for data from scientists to BODC.

Table 2 gives an indication of the extent to which the BOFS data has been disseminated through these mechanisms:

Year (Apr-March)	No. of requests	No. of user sessions	No. of CD-ROMs
1989-1990	20		
1990-1991	29	894	
1991-1992	24	1204	
1992-1993	29	500	
1993-1994	39	2000	
1994-1995	10	431	125
1995-1996	13	914	30

Table 2: The number of requests for BOFS data, remote logins to BODC project databases and BOFS North Atlantic CD-ROMs which have been disseminated. Note that the figure for database access covers access of Arabesque data in addition to BOFS.

These figures give us confidence that the biological and chemical data held by BODC are considered a valuable resource by the scientific community and are indeed being used.

6) CONCLUSIONS

The revolutionary approach pioneered for the data management of the North Sea Community Research Project has been successfully transferred to the U.K. JGOFS project. Maintaining a close working relationship with U.K. JGOFS scientists has been absolutely fundamental to this.

BODC has fully achieved its aims of supporting the U.K. JGOFS science and providing project scientists with timely access to the data. In addition, the U.K. JGOFS database is a major asset currently in heavy use which will also provide an important legacy to future generations of scientists.

ACKNOWLEDGMENTS

The work described here was only made possible by the unstinting dedication of a number of individuals. Special thanks are due to Ray Cramer who produced the software interface for the BOFS North Atlantic CD-ROM. The students Bill Cave, Peter Brocklehurst and Gareth Trevor have also made significant contributions to this project.

REFERENCES

Loch S.G. 1993 An Efficient, Generalized Approach to Banking Oceanographic Data. Proudman Oceanographic Lab. Report No. 24.

Lowry R.K. 1992 Data Management for Community Research Projects: A JGOFS Case Study. pp 251-273 in: Proceedings of the Ocean Climate Data Workshop, February 18-21, 1992.

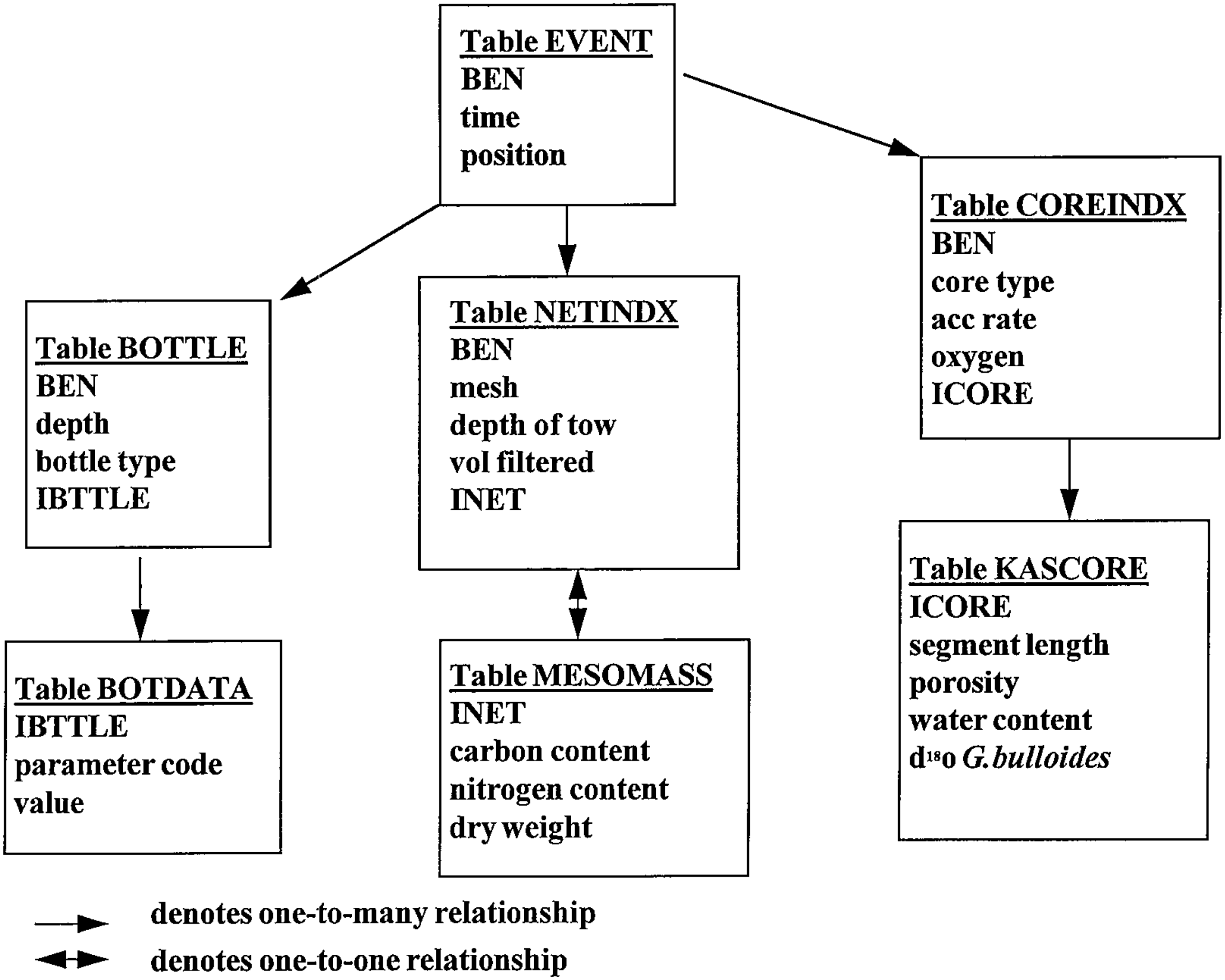
Lowry R.K. & Cramer R.N. 1995. Database applications supporting Community Research Projects in NERC marine sciences. Geological Data Management 97 pp 103-107

Lowry R.K. & Loch S.G. 1995. Transfer and SERPLO: powerful data quality control tools developed by the British Oceanographic Data Centre. Geological Data Management 97 pp 109-115

Rickards L.J. 1994. BODC Quality Assurance Procedures for Physical Oceanographic Data. ICES C.M. 1994/(C+E+L):1

UNESCO 1994. Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements. IOC/SCOR Manuals and Guides 29.

Figure 1 : An Overview of the Database Structures



MANAGEMENT OF BIOLOGICAL, PHYSICAL, AND CHEMICAL DATA WITHIN THE U.S. GLOBEC PROGRAM

Robert C. Groman and Peter H. Wiebe

ABSTRACT

The primary goal of the U.S. GLOBEC (**GLOB**al ocean **EC**osystems dynamics) Georges Bank Project is to understand the population dynamics of the target species on Georges Bank in terms of their coupling to the physical environment and their predators and prey. A key component of this project is the sharing of the field, retrospective, modeled, and derived data collected by the project's scientific investigators. We use the JGOFS (U.S. Joint Global Ocean Flux Study) data management software to serve these data and information to all investigators. The JGOFS systems takes advantage of standard World Wide Web browsers (such as Netscape, Mosaic, Internet Explorer, etc.) so that everyone can access our data base to look at, manipulate, and retrieve data stored within this distributed system.

INTRODUCTION

U.S. GLOBEC (**GLOB**al ocean **EC**osystems dynamics) is a research program organized by oceanographers and fisheries scientists to address the question of how global climate change may affect the abundance and production of animals in the sea. The U.S. GLOBEC Georges Bank Project, the first of several research modules with U.S. GLOBEC, is a large multi-national, multi-year, and multi-disciplinary oceanographic program involving over seventy scientific investigators from twenty-four organizations in the USA and Canada (Figure 1). The project's primary goal is to understand the population dynamics of the target species on the Bank, cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*), and the copepods, *Calanus finmarchicus* and *Pseudocalanus* spp., in terms of their coupling to the physical environment and their predators and prey. The ultimate goal is to predict changes in these populations as the physical and biotic environment change due to yearly variability and possible longer-term climatic changes.

In order to accomplish these goals, intensive field, laboratory, and retrospective studies were initiated in 1994, with major field efforts in 1995, 1997, and 1999. The effort is substantial, requiring broad-scale surveys of the entire Bank, and process studies which focus both on the links between the target species and their physical environment, and the determination of fundamental aspects of these species' life history (birth rates, growth rates, death rates, etc).

Equally important are the modeling efforts that are ongoing which seek to provide realistic predictions of the flow field and which utilize the life history information to produce an integrated view of the dynamics of the populations.

ORIGIN OF PROJECT DATA SETS

Broad-scale cruises carry out CTD (Salinity, Temperature, and Depth), zooplankton, larval fish, and acoustic surveys of Georges Bank and adjacent waters. Process cruises are designed to measure vital physiological rates of zooplankton and fish larvae, and to determine fine-scale vertical and horizontal distribution of zooplankton on Georges Bank, and to link these to particular physical processes under study. Zooplankton and fish larvae are collected at stations in a variety of ways, including pump, Bongo net, and MOCNESS (Multiple Opening/Closing Net and Environmental Sensing System). They are also remotely sensed in situ with the VPR (Video Plankton Recorder), TAPS (Tracor Acoustic Profiling System, a multi-frequency acoustic system that is lowered with the CTD), and BIOMAPER (a towed acoustic system). Microzooplankton is sampled with Goflo Bottles. Hydrography, and zooplankton abundance, distribution, stage frequency distribution, and condition are measured at stations between drifter stations and along selected transects. Moorings collect time-series data (hourly averages) from a variety of sensors including vector velocity (VACM) at two heights - 15m and 45m, and temperature sensors (VACM, SeaCats, and Branker temperature-probes) located at 1, 5, 10, 15, 20, 25, 30, 35, 40, and 45m. Salinity sensors (SeaBird SeaCats sampling at 2 Hz) are at 1, 10, 20, 30, and 40m. During cruises to set or recover moorings, hydrographic sections are made with a CTD. In addition, at strategic locations, satellite tracked Lagrangian drifters are deployed monthly on broad-scale, process, or mooring cruises to provide an additional measure of current flow. The data collected at sea are analyzed along with other available data such as satellite sea surface temperature images derived from full (1 km) resolution AVHRR satellite data for all of Georges Bank and the Gulf of Maine. Also analyzed are spatially averaged CZCS pigment images derived from full (1 km) resolution CZCS and SeaWiFS satellite data for all of Georges Bank and the Gulf of Maine.

The basic measurements spawn a number of derived computational products which also must be accommodated (see Table 1).

FOSTERING MULTI-DISCIPLINARY DATA EXCHANGE AND SYNTHESIS

Our challenge is to develop the database, data analysis, and data visualization structures which will enable widely distributed, multi-disciplinary investigators to work with each other's data and to collaborate with each other without the necessity of leaving their home institution (sometimes referred to as the collaboratory concept). An essential element of the project is the U.S. GLOBEC data policy which fosters quick sharing of data and information with other members of the program. Key elements of this policy (U.S. GLOBEC Data Policy, Report Number 10, February 1994) are:

- Data must be made available to the scientific community on a timely basis;
- Field data, retrospective data sets, and numerical experiments must all be included in the database;
- As soon as data might be useful to other researcher, the data should be released (note, the data need not be "final" for this to occur);

- Model results that would be useful for the interpretation of field data or comparison with later model studies should also be included;

- Documentation of the measurement and analysis techniques used to produce the data set must be submitted with the data to the Data Management Office;

- Data sets consist of both the actual measurements and also descriptive data, sometimes referred to as metadata. U.S. GLOBEC databases must include all relevant metadata;

- The primary responsibility of the Data Management Office is to accept data from U.S. GLOBEC investigators, to verify the data has been properly transmitted, to report on the status of data submissions to the Program Manager and the Steering Committee, and most importantly to facilitate the interdisciplinary exchange of data.

DATA MANAGEMENT STRUCTURE

We selected the U.S. Joint Global Ocean Flux Study (JGOFS) data management software [Flierl, et. al., in press] to implement our data storage and retrieval strategy. The JGOFS software provides a distributed, flexible, extensible, and data driven methodology to store and serve data and information about the data (metadata).

Distributed Functionality

The JGOFS system takes advantage of the **h**yper**t**ext **t**ransmission **p**rotocol (http) to exchange data between servers and clients. This enables the JGOFS system to use any UNIX or PC/Windows based computer as a server. The Georges Bank Program currently uses four data servers and one applications server. New data servers can be added easily; indeed we could have a ratio of one server per data contributor. Similarly, any networked computer system running a Web browser (such as Netscape, Internet Explorer, or Mosaic) is a supported client and has access to our on-line data and information. One does not need to know where the data are stored to access it; rather, the system takes care of automatically generating the necessary hypertext links on the Web page each time data are is requested. An additional benefit of this approach is that the data directory is automatically maintained. (See Data Driven, below.) To access our home page, use http://globec.whoi.edu.

Flexible Methods for Handling Data

The JGOFS system can handle any data format and data type necessary. This is because it uses the data object concept where data and the necessary method (i.e. software) to access it are linked into a "data object" (Figure 2). The networking and inter-operational software are common to every data object software module; only the code specific to accessing and reading the data needs to be written. In this way, we are able to serve both ASCII and binary data with equal ease, and serve image and video data using the same software. There are three "standard" methods distributed with the JGOFS system and these can be used to handle many situations without further programming. For example the default or **def** method handles ASCII flat files of the form:

leg	year	month	station	lat	lon	press	temp	sal	o2	sigth
1	81	6	3	38.28	-73.53	5.000	18.334	33.570	5.970	24.096
1	81	6	3	38.28	-73.53	25.000	12.848	34.159	6.990	25.773
1	81	6	3	38.28	-73.53	49.000	11.070	34.523	6.060	26.394
1	81	6	3	38.28	-73.53	99.000	11.093	35.090	5.340	26.831
1	81	6	3	38.28	-73.53	149.000	11.906	35.487	5.020	26.990
1	81	6	3	38.28	-73.53	199.000	10.819	35.435	4.210	27.152
1	81	6	3	38.28	-73.53	300.000	8.293	35.126	3.730	27.334
1	81	6	3	38.28	-73.53	400.000	6.363	35.046	4.640	27.546
1	81	6	3	38.28	-73.53	500.000	5.724	35.019	4.980	27.608

The data can be separated by either blanks, a comma, or tab. The columns do not need to line up. The **def** method can also be used to read hierarchically structured data where the slowest varying parameters are listed first. In fact, this is the preferred approach for handling such data.

Extensible Architecture

The JGOFS system is highly extensible since one is free to use any programming language or even scripting language to help serve data and information. The JGOFS architecture readily allows for new capabilities and new features without compromising its design or implementation. This feature is exploited, for example, when serving our static and video drifter track images as well as the AVHRR satellite images. In the later case, an existing satellite image is made available and converted, as needed, to a format suitable for display by the browser (i.e. gif image format) all within the context of data system.

Data Driven

The system lists available data objects based on what is currently available. These lists are generated each time a person requests a list and therefore they are always up to date (Figure 3). Furthermore, we attempt to serve data in the format and form used by the scientific investigator who collected or generated the data in the first place so that users of the system are actually using **the same data** from **the same computer** that the contributing investigator is using. Whenever the data are further processed or errors removed, the most up to date data are automatically made available to the others in the project. This is consistent with our policy of making data available whenever it is useful to others, even if the data are not yet final.

FUTURE VENTURES

Until recently our focus has been to provide access to the data; we have not addressed the data analysis needs of our researchers. However, we are expanding our ability to offer analysis tools to users of our system by adding additional display software and output format options. Currently, we offer basic x-y plotting (Figure 4) and basic mapping plots and a basic flat file listing of the data, in ASCII. We recently added the ability to download a Matlab formatted data file which can be loaded directly into Matlab using the standard load command. For Unix based machines, it is also possible to access data objects directly from the data system from within Matlab using the M-file command loadjg. This command, when given the data object name, accesses the data directly from the appropriate JGOFS data server over the network and creates a data vector for each field name in the object. We are also actively investigating other analysis options such as providing a scientific visualization capability and links to other display and analysis systems such as LinkWinds.

Other tasks include training of investigators with varying computer skills to make it possible for the collaboratory interactions to actually take place.

REFERENCES

Flierl, Glenn R., Bishop, James K.B., Glover, David M. and Paranjpe, Satish "A Data and Information System for JGOFS", in press. A different version of this paper is available on-line at http://lake.mit.edu/datasys/jgsys.html called "JGOFS Data System Overview."

LinkWinds User's Guide: The linked Windows Interactive Data System (Version 2.1), Allan S. Jacobson, et. al.

U.S. GLOBEC Data Policy, Report Number 10, February 1994. Available on-line from http://www.usglobec.berkeley.edu/usglobec/Reports/reports.home.html.

ACKNOWLEDGMENTS

We would like to thank Glenn Flierl for his primary contributions in designing and implementing the JGOFS data management system, and Chris Hammond and Warren Sass for their help in expanding on the system's capabilities. This project is supported by the National Science Foundation Grants OCE-9313674 and OCE-9417423. U.S. GLOBEC Contribution No. 55 and WHOI Contribution No. 9228.

Table 1: Summary of data sources, instruments, sensors, and derived measurements associated with the U.S. GLOBEC Georges Bank Program

Data Sources
- Broad-scale cruise
- Process Cruise
- Moorings
- Drifters
- Satellite
- Aircraft
- Modeling

Instruments
- CTD (MKV/MKIII/SeaBird)
- Rosette (1.4 & 5 liter bottles
- 1x1 meter MOCNESS (150 & 335 mm mesh)
- 10x10 meter MOCNESS (3 mm mesh)
- 60 cm Bongo (335 mm mesh)
- Acoustics 120/420 kHz
- Video Plankton Recorder
- ADCP (Ship and Mooring)
- Drifters (GPS/ARGOS)
- Turbulence
- MET Package
- Mooring
  - Current Meters
  - C/T
  - ADCP
- Infra-red radiometer (satellite)
- Air-borne laser (fluorescence)

Sensors
- **Sea**
  - Conductivity
  - Temperature
  - Pressure
  - Fluorescence
  - Transmittance
  - Acoustics (sv & TS)
  - Light (PAR)
  - Video
- **Air**
  - Temperature
  - Humidity
  - Light (PAR)
  - Wind Speed
  - Wind Direction
  - Precipitation
- **Space**
  - AVHRR

Derived Measurements
- Plankton Biomass
- Taxonomic Composition/Size Distribution
- Species
  - Counts
  - Size
  - Stage
  - Status
    - RNA/DNA
    - Cell growth potential (turn-over rate)
    - Gut fullness
    - Gut content
    - Gut fluorescence
    - Lipd content/structure
    - Genetics
  - Rates
    - Feeding
    - Egg production
    - Growth
    - Molting rate
  - Behavior
    - Orientation
    - Spacial relationships
- Density
- Currents
- Stratification
- Heat Flux (Air/Sea Exchange)
- Turbulence (Energy Dissipation)
- Nutrients (N, P Si)
- Oxygen 18/Oxygen 16
- Chlorophyll
- Phytoplankton/Microzooplankton

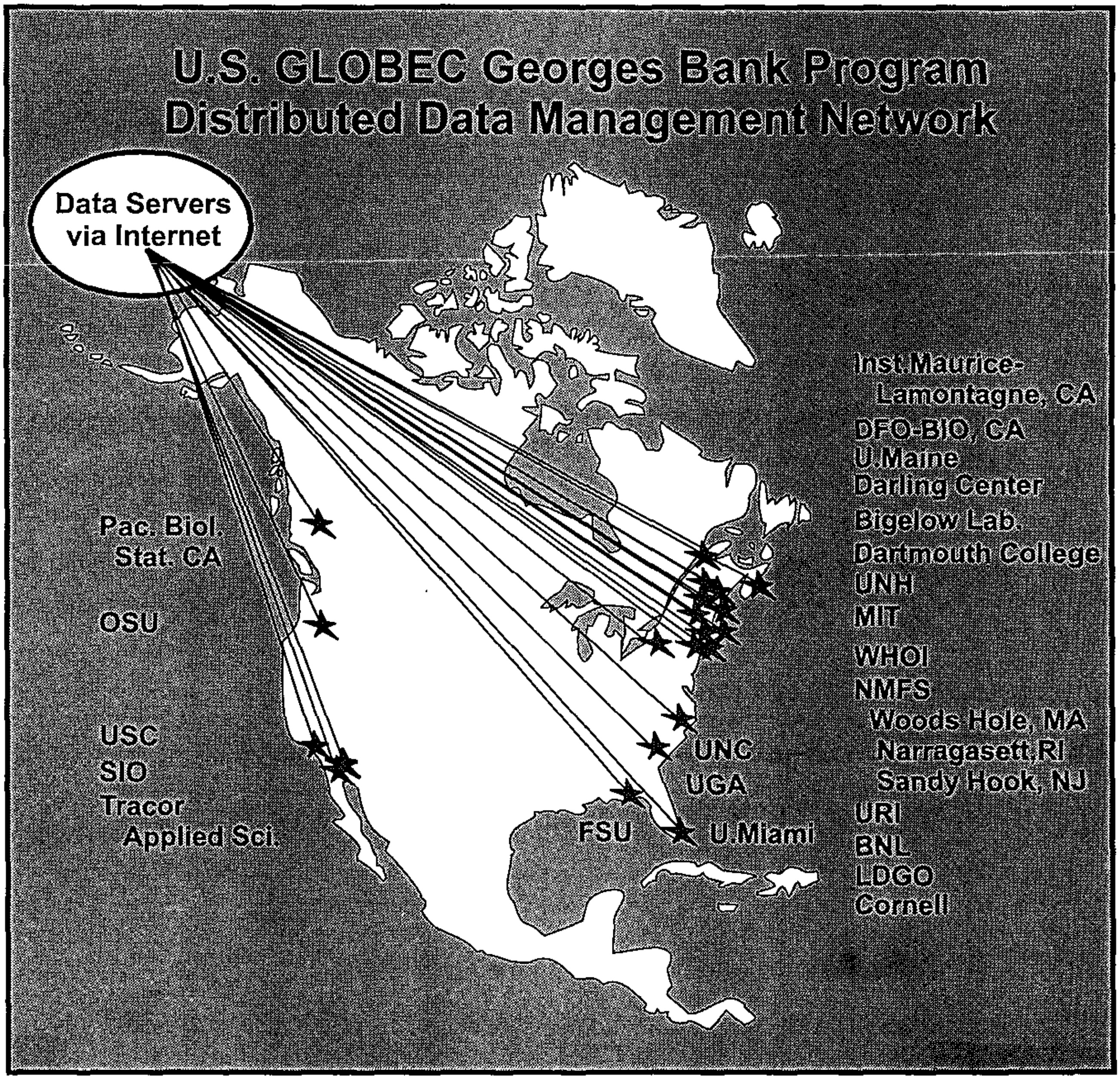


Figure 1: Seventy scientific investigators from twenty four organizations in Canada and the USA.

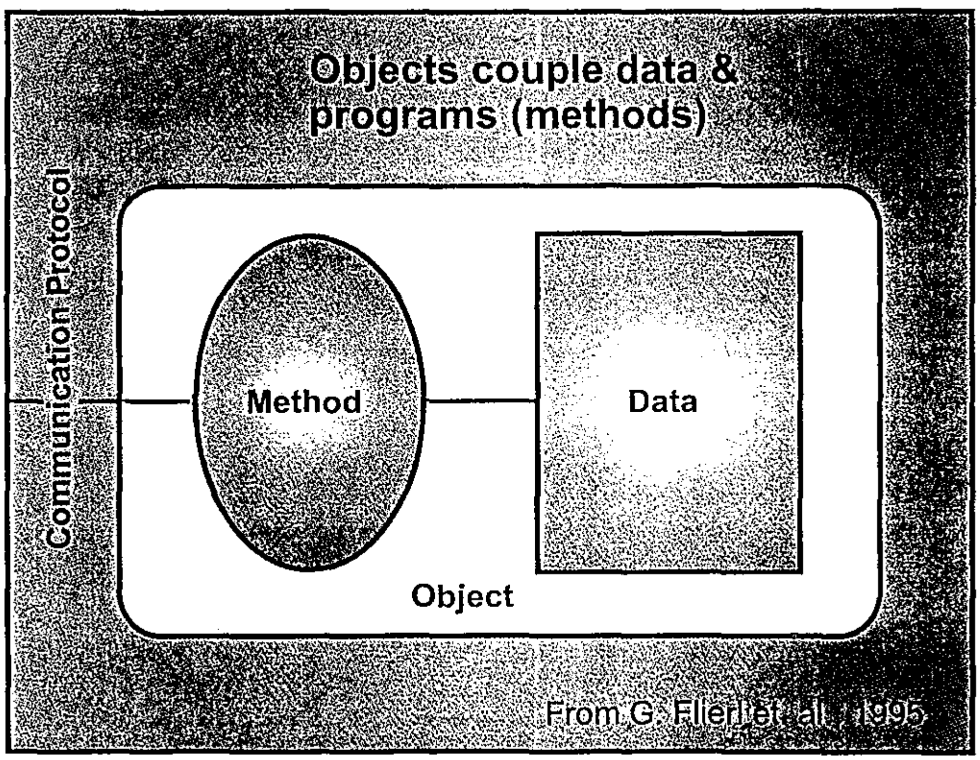


Figure 2: Data and the necessary software to access it are linked into a "data object".

/globec/gb/ -- data objects

SUBDIRECTORIES
- broadscale
  - 1992
  - 1993
  - 1994
  - 1995
  - 1996
- modeling
- mooring
  - 1993
  - 1994
  - 1995
  - 1996
- process
  - 1992
  - 1993
  - 1994
  - 1995
- satellite
  - 1993
  - 1994
  - 1995
  - 1996

OBJECTS		Click on this column for Documentation
bathymetry	Dave Greenberg	Georges Bank bathymetry data triplets
bathy ADCP	Charles Flagg	Georges Bank bathymetry from ADCP data
drifters	R. Limeburner	Drifter data (WHOI and other sources)
met 1 hour	Richard Payne	MET 1 hour data from Endeavor cruises
met 1 min	Richard Payne	MET 1 minute data from Endeavor cruises
MOCNESS	Ted Durbin	Preliminary biological counts
test		Test hydrographic data

Figure 3: Object lists are generated each time a person requests a list and therefore is always up to date. (Note: the format of this display is being revised.)

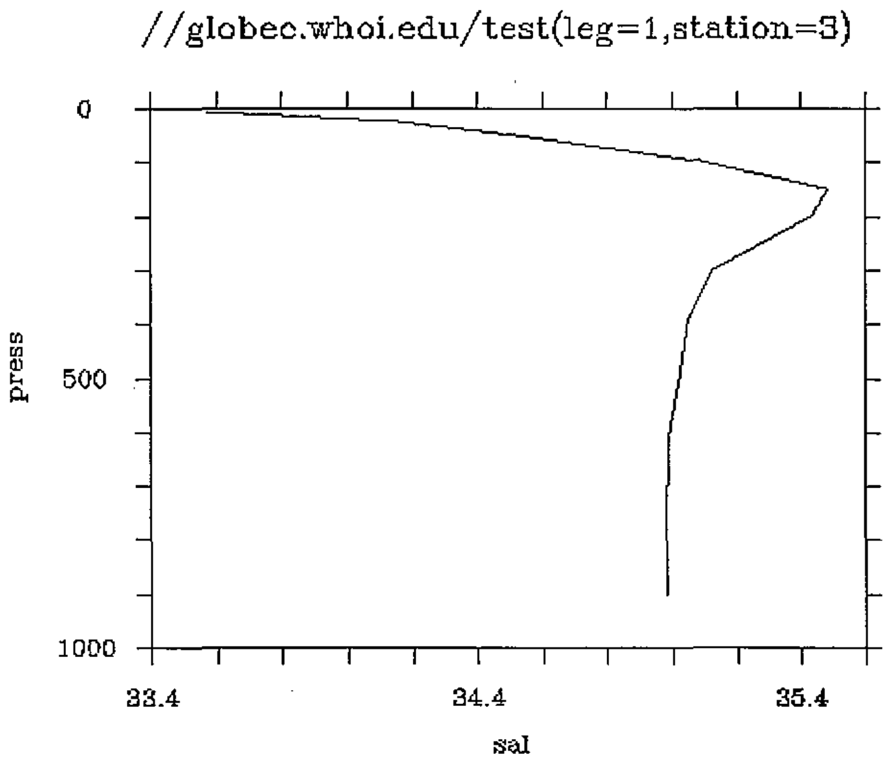


Figure 4: Basic x-y plot generated "on the fly" from data retrieved over the network.

THE SeaWiFS MISSION: REQUIREMENTS FOR BIOLOGICAL AND CHEMICAL DATA

Watson W. Gregg

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) is a follow-on, next-generation ocean color sensor to the Coastal Zone Color Scanner (CZCS). The CZCS provided sporadic ocean color data from 1978 until its demise in 1986. SeaWiFS, in contrast to the CZCS, is a dedicated, continuous operation sensor, whose main goal is to provide routine global coverage. The mission life is specified to be 5 years.

SeaWiFS also improves on the CZCS in data quality. SeaWiFS has 8 spectral bands (Table 1), instead of the functioning bands on the CZCS. The two additional visible bands (412 nm and 490 nm) allow better discrimination of pigments, and the two new near-infrared bands enable a better determination of aerosols in the atmospheric correction process. The spectral resolution of SeaWiFS can potentially provide more information about the oceans, due to the placement of the bands with respect to optical phenomena in the oceans (Figure 1). The SeaWiFS bands have a bi-linear gain response that enables them to avoid saturation over most land and cloud features yet retain a high signal-to-noise ratio for low-radiance ocean observations.

The primary scientific goal of the SeaWiFS mission is to provide global monitoring of ocean color. However, because the spacecraft contains limited storage capacity (125 x 10^6 Bytes), onboard subsampling of the data is required to obtain global coverage. Thus SeaWiFS will produce two different resolutions of data: Local Area Coverage (LAC) and Global Area Coverage (GAC). LAC data is full sensor resolution data (1.12 km at nadir) and is used for regional observations and calibration activities. GAC data is simply LAC data subsampled every fourth pixel along scan and along-track, and is used for the global data set. GAC data also contains only data within a 45° swath width, in contrast to the 58.3° swath for LAC data. Since both of these data are stored onboard the spacecraft data recorder, they are referred to as stored data. In addition to stored LAC and GAC data, SeaWiFS will broadcast real-time LAC data whenever the sensor is turned on. High Resolution Picture Transmission (HRPT) stations may capture these data whenever the spacecraft is visible. All data are 10-bit digitized, unlike the CZCS, which had 8-bit digitization. Also, SeaWiFS is capable of tilting ±20° along the velocity vector to avoid excessive sun glint contamination, like the CZCS. SeaWiFS is carried on orbit by the SeaStar spacecraft. SeaStar is in a circular, descending node, near-noon Sun-synchronous orbit at a nominal altitude of 705 km.

SEAWIFS DATA PRODUCTS

The SeaWiFS Project provides three general categories of data products: Level-1A, Level-2, and Level-3 data (Table 2). These definitions follow the standardized Earth Observing System (EOS) requirements. The Level-1A product is in count values, which indicate the data values actually stored in the spacecraft in 10-bits, which consequently range from 0 to 1023. The Level-2 product is where derived geophysical products first appear. They are reported in geophysical units, although they remain in spacecraft coordinates (scan direction by travel direction). The two primary products are chlorophyll-a and CZCS-pigment, which is a similar calculation used in CZCS processing to provide temporal continuity. The radiance products are used primarily to derive the chlorophyll and pigment values, but have importance in radiative transfer by themselves and also serve as the basis for the development of new algorithms. Epsilon is merely the ratio of the aerosol radiance at Band 7 to the aerosol radiance at Band 8, and provides an indication of the characteristics of the prevailing aerosols. Aerosol optical thickness at Band 8 is an indicator of the amount of aerosol at the given pixel.

The two Level-3 products are now in Earth-mapped, gridded coordinates. The two sub-categories, Binned Data and Standard Mapped Images (SMI) provide different levels of information for different user groups. The SMI is likely to be requested by most users. The Maximum Likelihood Estimator (MLE) value reported in the SMI nearly approximates the arithmetic/geometric mean (Campbell et al., 1995). The Binned Data product provides more detailed information for user interested in statistical dispersions of the data or error analyses.

BIOLOGICAL AND CHEMICAL DATA REQUIREMENTS FOR SEAWIFS

The SeaWiFS requirement for in-situ oceanographic data is strictly and solely for algorithm development and validation. Consequently, actual data points are required, with accurate Earth location and time information provided, rather than gridded or time-averaged data.

The types of data required for SeaWiFS algorithm development and validation fall into two general categories, based on the general categories of the processing algorithms. These are the atmospheric correction and bio-optical algorithms.

Atmospheric correction algorithms involve remove atmospheric radiative effects form the total radiance signal observed by the satellite. These include scattering of light in the direction of the satellite by molecules (Rayleigh scattering) and particles (aerosol scattering), as well as absorption of light by water vapor, ozone, and oxygen. The result of the atmospheric correction process is a suite of water-leaving radiances in the visible part of the spectrum viewed by SeaWiFS. This is the radiance that emits the ocean after interacting with optically active components in the water, and carries information about the abundance of these optically active substances, most importantly, chlorophyll and pigments. The radiances are actually normalized to nadir-viewing and converted to that which emits the surface just above the surface, producing what are called normalized water-leaving radiances. Validation of these radiances requires comparison with measured spectral radiances just above the sea surface. Radiances measured just below the sea surface are nearly as useful, since sea surface properties can be evaluated and the differences reconciled. Surface downwelling radiances and irradiances are useful, too, since they determine the amount of light available before interacting with the oceanic substances. These are most useful for algorithm development.

Bio-optical algorithms utilize the derived normalized water-leaving radiances to produce estimates of the primary SeaWiFS products, chlorophyll and CZCS-pigment, as well as other useful variables, such as the attenuation at 490 nm (K490). Validation of these algorithms requires information about the biological and optical state of the ocean at selected locations and times. Since the optical properties of water are well known and constant, the most important validation data are chlorophyll-a concentrations, phaeopigments, and colored dissolved organic matter (CDOM) concentrations. Chlorophyll-a is the substance of interest, phaeophytin is a degradation product of chlorophyll that interferes optically with observations of chlorophyll. CDOM (formerly known as gelbstoff) is another degradation product that may have either an oceanic or terrigenous source, that absorbs strongly in the blue wavelengths that consequently also interferes with good retrievals of chlorophyll from the water-leaving radiances. The placement of Band 1 (412 nm) was specifically selected to provide a method to remove this substance from the chlorophyll estimates. Co-located information about CDOM is very important for SeaWiFS algorithm validation.

In addition to the proposed standard product, new products are desired, and may be possible, for several other important optical constituents in the oceans. Some of these are:

- coccolithophore abundance;
- detrital absorption;
- CDOM concentration;
- phytoplankton speciation.

Data that can support the development of algorithms to retrieve these variables would be extremely useful. Such examples are:

- coccolith concentration;
- detrital material concentration and optical properties;
- CDOM concentration;
- accessory phytoplankton pigment concentrations;
- phytoplankton species;
- phytoplankton size distributions;
- inorganic and organic suspended materials and optical properties;
- humic and fulvic acids.

The latter two substances are considered the primary components of CDOM, and differ in the optical properties and stability in the marine environment. Note that CDOM concentrations are required here in support of the goal of retrieving CDOM itself, not just refining chlorophyll estimates, as discussed before.

Finally, it is important to note the efforts already begun toward developing a data archive to support SeaWiFS algorithms at the NASA Goddard Space Flight Center. The SeaWiFS Project, under the direction of Kevin Arrigo, has established the SeaWiFS Bio-Optical Archive and Storage System (SeaBASS; Hooker et al., 1994). This archive presently contains pigment, optical and atmospheric data from ship and buoy sites, and can be accessed from the Project. More information about this archive, and the SeaWiFS mission as well, can be obtained on the Internet, from the site:
http://seawifs.gsfc.nasa.gov/SEAWIFS.html.

Table 1. SeaWiFS spectral bands and their primary purpose.

Band 1 -- 412 nm -- Colored Dissolved Organic Matter (CDOM)
Band 2 -- 443 nm -- Chlorophyll
Band 3 -- 490 nm -- Chlorophyll at high concentrations
Band 4 -- 510 nm -- Accessory pigments/coccolith reflectance*
Band 5 -- 555 nm -- Chlorophyll (hinge point)
Band 6 -- 670 nm -- Currently unused
Band 7 -- 765 nm -- Aerosols
Band 8 -- 865 nm -- Aerosols

* Anticipated future product

Table 2. SeaWiFS data products.

Level-1A: Raw at-satellite radiance in all 8 bands, in count values with calibration and navigation information, and telemetry.

Level-2: Derived geophysical variables
Normalized water-leaving radiances
Bands 1-5 (mW cm^{-2} m^{-1} sr^{-1})
Aerosol radiances Bands 7 and 8 (mW cm^{-2} m^{-1} sr^{-1})
CZCS Pigment (mg m^{-3})
Chlorophyll-a (mg m^{-3})
Attenuation coefficient at 490 nm (m^{-1})
Epsilon Band 7/Band 8 (dimensionless)
Aerosol optical thickness band 8 (dimensionless)
(Total = 12 products)

Level-3: Mapped geophysical variables

Binned Data
9km x 9km equal area bin, Daily, 8-Day, Monthly, Yearly Time scales
Same variables as Level-2
Values represented as

$$\frac{\ln(x)}{n}, \frac{\ln(x)^2}{n}, n, n$$

and various other ancillary information for each bin

Standard Mapped Images (SMI)
- Single-byte values, 4096 x 2048, equiangular
- Same time domain as Binned Data
- Same variables except aerosol radiances and epsilon
- Values represented by Maximum Likelihood Estimator (MLE) of ln(x) for each bin

REFERENCES

Campbell, J.W., J.M. Blaisdell, and M. Darzi, 1995. Level-3 SeaWiFS data products: Spatial and temporal binning algorithms. NASA Technical Memorandum 104566, Vol. 32. 73 pp.

Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E. Yeh, and Y. Ge, 1994. The SeaWiFS bio-optical archive and storage system (SeaBASS). NASA Tech. Mem. 104566, Vol. 20. 40 pp.

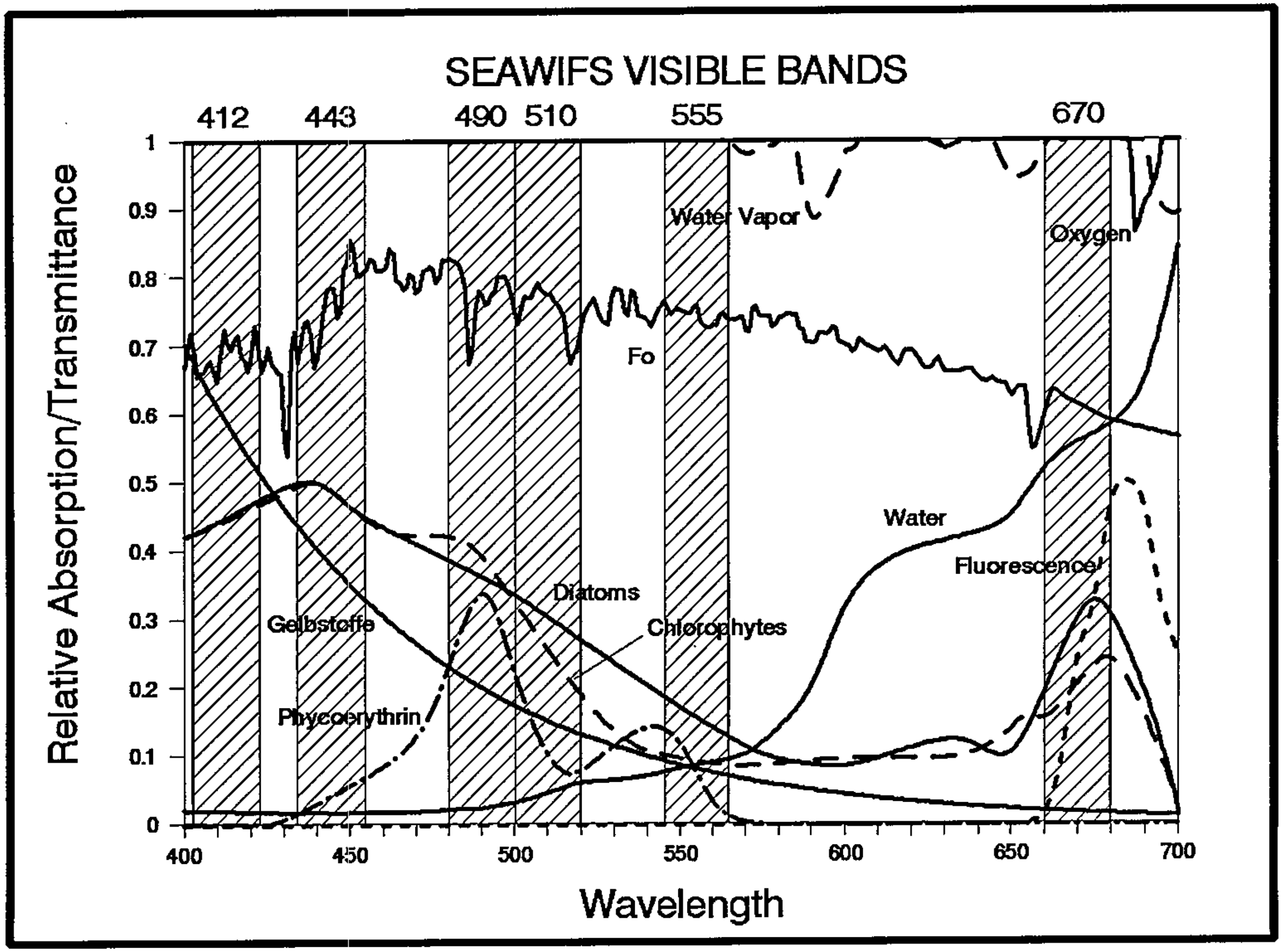


Figure 1. Relative absorption spectra of various oceanic and atmospheric optically active constituents, and the SeaWiFS visible bands. The bands are placed in regions of importance to optical constituents in the oceans. Fo indicates extraterrestrial irradiance.

MANAGEMENT OF THE HELCOM BMP BIOLOGICAL AND CHEMICAL DATA

Jouko Rissanen

ABSTRACT

The Finnish Environment Institute (FEI) is a national research and development centre and a part of the environment administration in Finland (http://www.vyh.fi/fei/fei.html). FEI promotes the quality of the environment according to the principles of sustained development and produces environmental information, models and services for administration, citizens and economic life. International cooperation in the environmental issues is one of the tasks. FEI has been responsible for the data bank of the Helsinki Commission s Baltic Monitoring Programme since 1991.

The Helsinki Convention was signed in 1974 by countries surrounding the Baltic Sea to protect the Baltic Sea from all kind of pollution. In 1992, a new Convention was signed by all nine countries bordering the Baltic Sea (Fig. 1) and by the European Economic Community.

The Baltic Marine Environment Protection Commission - the Helsinki Commission (HELCOM) (http://www.helcom.fi) was established by the Convention in 1974 to keep the implementation of the Convention under continuous observation and to make relevant decisions to fulfil the objectives and goals of the Convention such as the reduction and elimination of pollution of the marine environment of the Baltic Sea from all possible sources.

The Commission has established several expert committees. The Environment Committee (EC) is one of them and covering different sectors of the marine environment, the open sea and the coastal waters including issues related to nature conservation and biodiversity. A joint monitoring programme - Baltic Monitoring Programme (BMP) - has been coordinated by the Commission since 1979 to follow the long-term changes of selected determinants in the open sea ecosystem. The data are compiled into joint databases and evaluated at regular intervals by experts from the Baltic Sea States, in order to assess the environmental conditions and to predict possible man-introduced changes in the ecosystems of the open Baltic Sea (HELCOM 1990a, 1990b, 1996a, 1996b). The Commission has published special guidelines for the programme, for example about station networks, parameters, methods to be applied in sampling, analytical methods, frequency of sampling and the reporting of results (HELCOM 1988a, 1988b, 1988c, 1988d).

Each participating country has a national data coordinator (Fig. 2) who is responsible for data collecting, analyzing, processing and the submission of data according to timetables using agreed formats. The data originators are also responsible for the good quality of the delivered data. To ensure the best possible temporal sampling coverage in all parts of the Baltic Sea the national data coordinators prepare annually coordinated timetables for the BMP cruises.

The relational database of the HELCOM BMP consists of about 750,000 records of hydrographical and -chemical data and over 200,000 records of biological data from more than 50 stations. The list of the parameters included in the BMP and in the database are shown in Figure 3.

The HELCOM BMP data is not public. According to the contract the national coordinators still have all rights to the data and all data requests coming from outside of the HELCOM organization must be submitted to the national data coordinators for their approval. This is why there are no open links to public networks from the database. Data requests should be submitted to the Secretariat of HELCOM.

FEI prepares annual data reports to the HELCOM EC and EC MON meetings and the Contracting Parties. A map of nitrate concentrations on water layer near the bottom (Fig. 4) is an example of the content of the annual report.

During the preparation of the third periodic assessment of the state of the Baltic Sea FEI has produced over 80 data sets and about 800 draft figures. A map of the distribution of potentially toxic phytoplankton species (Fig. 5) is an example of this work.

The biological part of the database, the species data of phyto- and mesozooplankton and macrozoobenthos, is based on the Rubin code system which has been produced by the Nordic Code Centre (NCC). NCC was set up by the Nordic Environmental Data Group of the Nordic Council of Ministers in 1985 but the finance has now ended. NCC has developed and maintained several term lists on biological species, physico-chemical determinants and other terms relevant for describing environmental parameters (ftp://ftp.nrm.se/pub/nrm/it/termlist).The maintenance of the biological species thesaurus and distribution of material has been done by the Documentation Centre at the Swedish Museum of Natural History, Stockholm. The maintenance of the physico-chemical parameters has been a responsibility of FEI.

Changes in the scientific names and taxonomical position of species as well as new species are a problem because the Rubin code system is naturally always coming little bit behind, but there is a demand to use the new names immediately. This has in some cases lead to a situation were different countries use slightly different codes systems and names for the same species. To prevent this a small workgroup has now been established to standardize the codes and to improve the updating of Rubin code lists.

The data transfer, especially the data updates, between data originators and HELCOM data centre does not always work smoothly because of the old and very laborious data transfer format. The format used in HELCOM data transfer is a copy of an old punch card system and allows many different types of errors.

FEI is preparing for the national use a general format for the data transfer. This format could also be used for HELCOM data. FEI will publish the new format proposal in the Wold Wide Web. FEI will open a ftp-server for better data transfer in the near future, but it will not be a public server because the use of data is restricted to the HELCOM Contracting Participants only.

The HELCOM has now accepted a project proposal for improved use of information technology. The Finnish Environment Institute is cooperating in this project, which is managed by the German Federal Maritime and Hydrographic Agency (BSH). The aim of this project is to improve the information flow between contracting participants by means of enhancing reporting formats, improving the data quality by better feedback mechanisms between data originators and receivers, harmonizing data handling with other conventions, making use of distributed data banking and data archives, accelerating information delivery, distribution and retrieval by network techniques and establishing networked data management for integrated assessments.

In the last HELCOM EC Monitoring (1/96) meeting standardization and use of the data stored in HELCOM data base was discussed. It is evident that a need for exchange of environmental data, as well as a need to use and compare data collected at different times by different institutions for different purposes is increasing. To enable the re-use as well as to guarantee good quality of reported HELCOM data, it is necessary to provide it with supporting information.

To ensure the use of the HELCOM data for different purposes in the future "the standardization" of the data structure is needed. The use of common code lists improves the comparability and the use of the data for different purposes. In order to process data from different sources, uniform attribute definitions are necessary. List of the attributes consists of descriptions, codes and formats used in the data systems. This kind of list is useful when creating, upgrading or using data bases. If the HELCOM data base in the future will be replaced with a net connecting national data bases, uniform attribute definitions would be an advantage.

FEI expects that these activities will improve the management and use of the HELCOM data and cooperation with other international organizations like ICES, OSPARCOM and IOC.

REFERENCES

HELCOM 1994. Helsinki Commission, 20 Years of International Cooperation for the Baltic Marine Environment 1974-1994. Helsinki Commission, Helsinki, 40 p.

HELCOM 1988a. Guidelines for the Baltic Monitoring Programme for the Third Stage. Part A. Introductory Chapters. Baltic Sea Environment Proceedings No. 27A, 49 p.

HELCOM 1988b. Guidelines for the Baltic Monitoring Programme for the Third Stage. Part B. Baltic Sea Environment Proceedings No. 27B, 60 p.

HELCOM 1988c. Guidelines for the Baltic Monitoring Programme for the Third Stage. Part C. Harmful Substances in Biota and Sediments. Baltic Sea Environment Proceedings No. 27C, 154 p.

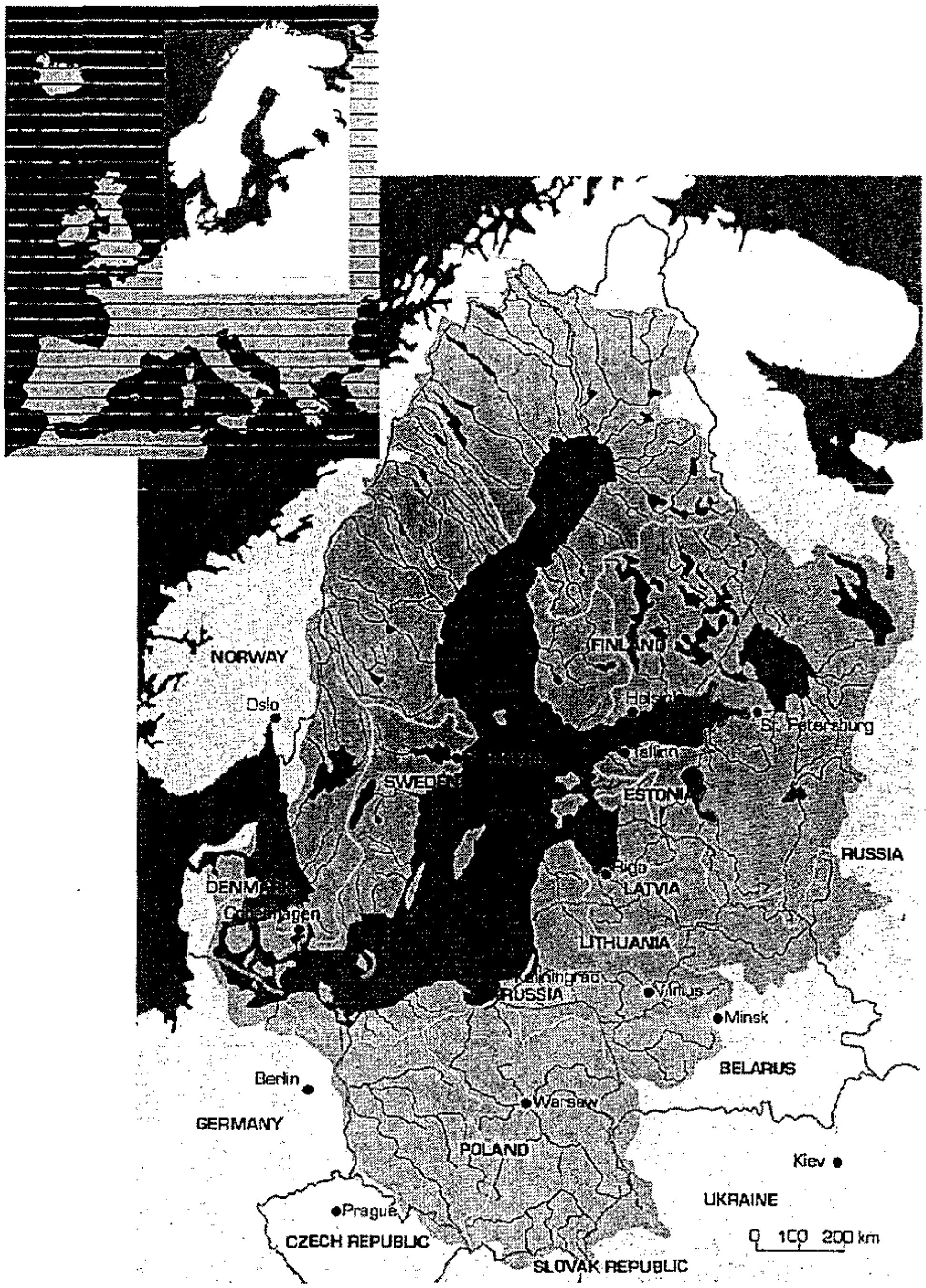
HELCOM 1988d. Guidelines for the Baltic Monitoring Programme for the Third Stage. Part D. Biological Determinants. Baltic Sea Environment Proceedings No. 27D, 161 p.

HELCOM 1990a. Second Periodic Assessment of the State of the Marine Environment of the Baltic Sea, 1994-1988; General Conclusions. Baltic Sea Environment Proceedings No. 35A.

HELCOM 1990b. Second Periodic Assessment of the State of the Marine Environment of the Baltic Sea, 1994-1988; Background Document. Baltic Sea Environment Proceedings No. 35B, 432 p.

HELCOM 1996a. The State of the Baltic Sea Marine Environment. Baltic Sea States Summit, Visby, Sweden 3-4 May 1996, Helsinki Commission, 20 p.

HELCOM 1996a. Protection of the Baltic Sea - results and experiences. Baltic Sea States Summit, Visby, Sweden 3-4 May 1996, Helsinki Commission, 32 p.



The area of the Baltic Sea is about 370,000 km², and its volume about 21,000 km³. The Baltic Sea drainage area comprises more than 1.7 million km². Fresh water is discharged into the sea from hundreds of rivers and water courses, and the annual fresh water input totals approximately 450 km³. Almost 80 million preople live within the drainage area, which includes not only the nine riparian states but also parts of five other states. The population density is by far the largest in the southernmost and south-eastern parts of the Baltic Region.

Figure 1: The drainage area of the Baltic Sea and the countries participating in HELCOM work (HELCOM 1994).

National coordinators of the HELCOM BMP

Ministry of the Environment
National Environment Research Institute
P.O.Box 358, DK-4000 ROSKILDE, **Denmark**

Estonian Marine Institute
Marine Research Center
Paldiski mnt. 1, EE-0001 Tallinn, **Estonia**

Finnish Institute of Marine Research
P.O.Box 33, FIN-00931 Helsinki, **Finland**

Federal Maritime and Hydrographic Agency of Germany
(Bundesamt für Seeschiffarhrt und Hydrographie)
P.O.Box 301220, D-20305 Hamburg, **Germany**

Marine Monitoring Centre
Latvian Hydrometeorological Agency
6 Daugavgrivas Str., LV-1007 Riga, **Latvia**

Center of Marine Research
Taikos 26, 5802 Klaipeda, **Lithuania**

Institute of Meteorology and Water Management
Waszyngtona 42, PL-81342 Gdynia, **Poland**

Hydrometeorological Institute
Malookhtinsky pr. 98, St. Petersburg 195196, **Russia**

Swedish Meteorological and Hydrological Institute
Byggnad 31 Nya Varvet, S-42671 Västra Frölunda, **Sweden**

Figure 2: The national coordinators of the HELCOM Baltic Monitoring Programme.

The selected hydrographical, -chemical and biological determinats of the HELCOM BMP

- Temperature
- Salinity
- Oxygen
- Hydrogen sulphide
- pH
- Alkalinity
- Phosphate
- Total phosphorus
- Ammonia
- Nitrate
- Nitrite
- Total nitrogen
- Silicate
- Phytoplankton primary production
- Chlorophyll-a
- Phytoplankton (species composition, abundance and biomass)
- Mesozooplankton (species composition, abundance and biomass)
- Macrozoobenthos (species composition, abundance and biomass on soft bottoms)

Figure 3: The selected hydrographic and chemical and biological parameters of the HELCOM Baltic Monitoring Programme.

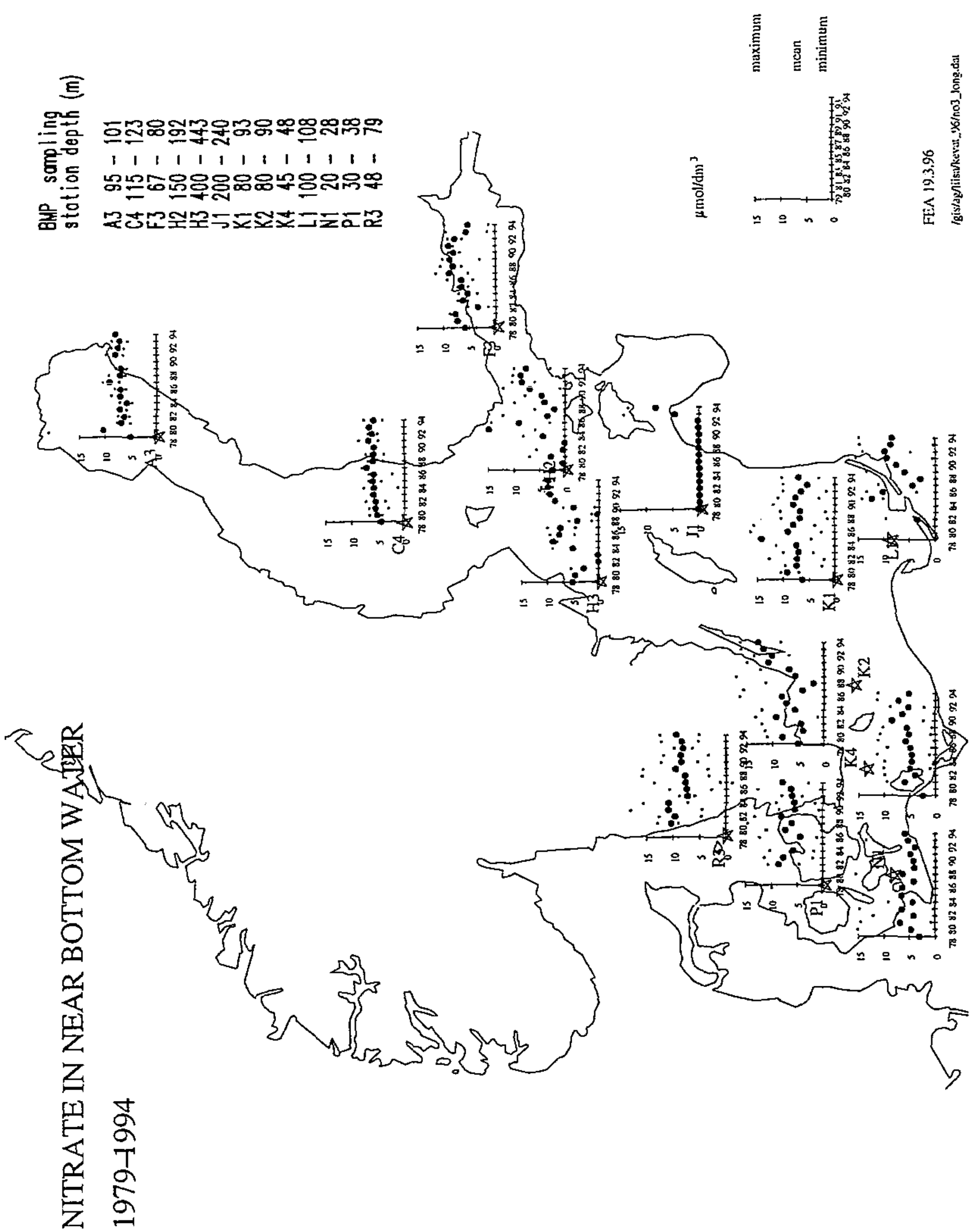


Figure 4: Nitrate concentrations in the water layer near the bottom in the period from 1979 to 1994. in the Baltic Sea. The end of the long stagnation period in 1993 and the oxygenizing of deep areas of the Baltic Sea can be seen in nitrate concentrations at station J1.

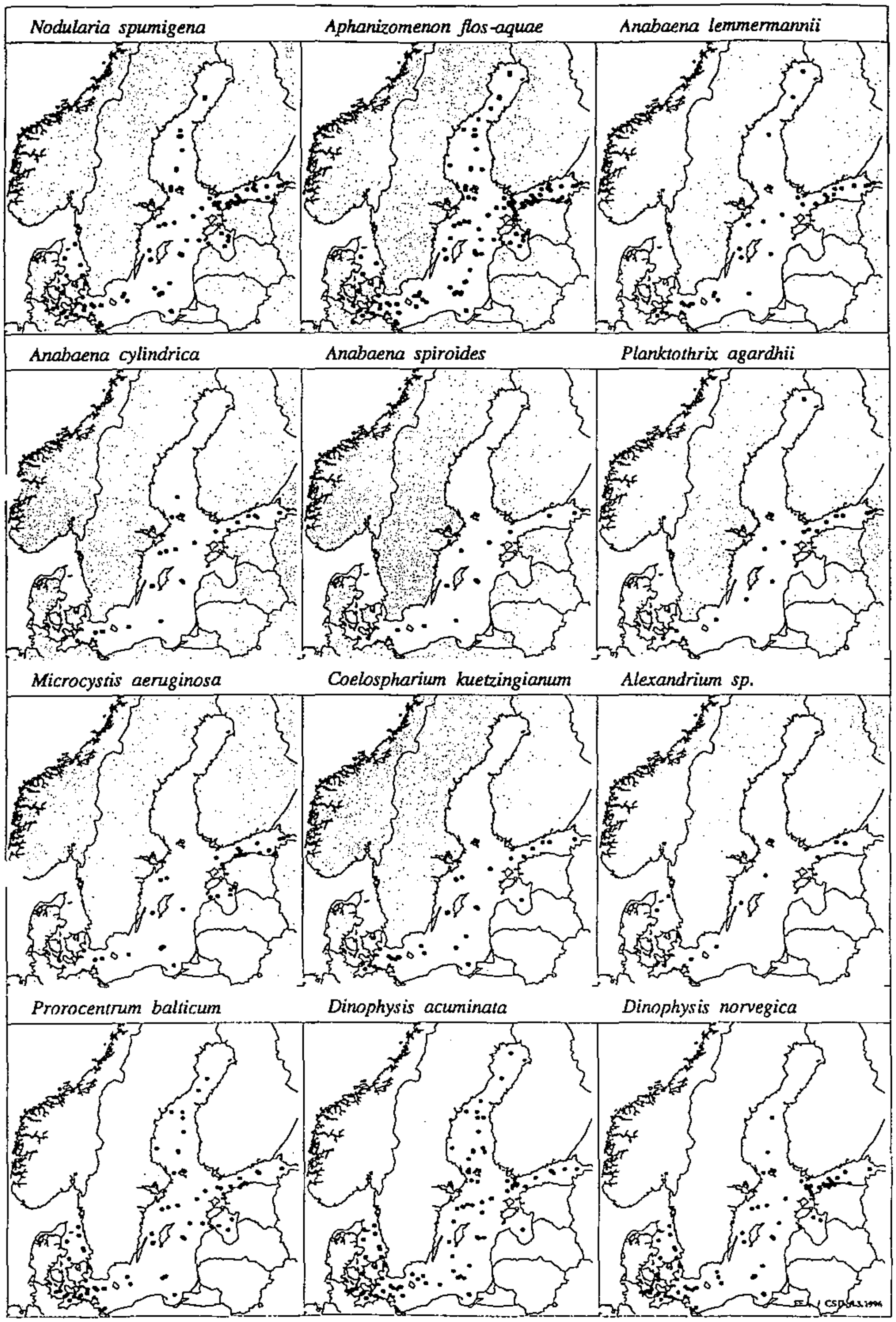


Figure 5: An example of the distribution maps of potentially toxic phytoplankton species.

DATA MANAGEMENT SUPPORT PROVIDED BY THE CARBON DIOXIDE INFORMATION ANALYSIS CENTER FOR THE U.S. DEPARTMENT OF ENERGY CO_2 OCEAN SURVEY

Alexander Kozyr and Thomas A. Boden

The Carbon Dioxide Information Analysis Center (CDIAC) at Oak Ridge National Laboratory provides data management support for the Joint Global Ocean Flux Studies (JGOFS) carbon dioxide (CO_2) measurements taken aboard research vessels during World Ocean Circulation Experiment (WOCE) Hydrographic Programme cruises. The U.S. Department of Energy's Global Change Research Program (DOE GCRP) sponsors both the CO_2 measurement operation and CDIAC's data support activities, which include development of a data management plan, data archival, data checking and evaluation, preparation of data documentation, and data dissemination.

An important part of the NDP process at the CDIAC involves the quality assurance (QA) of data before distribution. Data received at CDIAC are rarely in a condition that would permit immediate distribution, regardless of the source. To guarantee data of the highest possible quality, CDIAC conducts extensive QA reviews that involve examining the data for completeness, reasonableness, and accuracy. Although they have common objectives, these reviews are tailored to each data set, often requiring extensive programming efforts. In short, the QA process is a critical component in the value-added concept of supplying accurate, usable data for researchers.

CDIAC with cooperation from Taro Takahashi and his staff at LDEO developed a computer-assisted data evaluation system for assessing of the oceanic CO_2 measurements and associated data being obtained by the participating members of the DOE's Ocean CO_2 Measurement program. This system is used to rapidly identify inconsistencies among values which need to be corrected or removed, and to generate an internally coherent CO_2 and related chemistry data set (including the concentrations of oxygen and macro-nutrients) for global oceans.

During the last two years several data quality control computer programs, which were developed at LDEO and customized for execution on a UNIX operating environment, have been delivered and installed at CDIAC.

> **Figure 1:** PLOTNEST.C - This C program was written by Stewart C. Sutherland (LDEO) and identifies questionable data points from large numbers of observations on a regional scale. The program requires a specific data format and yields Postscript files containing nested depth profiles of measured ocean properties using the station number as an offset (i.e., the first station is defined at the beginning and subsequent stations are offset by a fixed interval).

Figure 2: PLOTSECT.C - This C program produces a Postscript file from an input file similar to PLOTNEST.C. It is used to display data values over an X-Y plane where X may be latitude, longitude or distance, and Y may be water depth or pressure. This program is used to examine the coherence of data with respect to their neighbors. Also, it is used to produce plot-section profiles for combinations of measured values.

Figure 3: To identify noisy data and possible systematic, methodological errors, property-property plots for all parameters are generated, using XVGR program (Version 2.09, P.J. Turner OGI-ESE), carefully examined, and compared with plots from previous expeditions. XVGR program is also used to generate pictures of sample depths at all hydrographic stations occupied during different expeditions.

As new measurements are made and submitted to CDIAC, it will be possible to compare measurements obtained at identical locations during different expeditions to assess their agreement. This is particularly important for measurements made in deep and abyssal waters. These checks will assess the geographical consistency of the CO_2-related parameters (TCO_2, pCO_2, alkalinity, and pH), salinity, and the concentrations of dissolved oxygen and macro-nutrients along constant seawater density surfaces.

One of the noteworthy developments is the willingness of scientists supported by the National Oceanic and Atmospheric Administration's Ocean-Atmosphere Carbon Dioxide Exchange program to archive their CO_2-related measurements at CDIAC for comparison, documentation and distribution purposes. Two documents that are contained data obtained during NOAA/PMEL cruises in Pacific and Indian oceans have been recently completed at CDIAC.

The end products of the data management plan are CDIAC's numeric data packages or NDPs. Each NDP contains written documentation and machine-readable data. CDIAC makes these NDPs available to anyone without charge in a variety of forms, including floppy diskettes, CD-ROM, nine track magnetic tape, eight-millimeter tape and quarter-inch tape cartridges. The data packages are also available from the CDIAC anonymous file transfer protocol (FTP) area via Internet:

FTP to cdiac.esd.ornl.gov (128.219.24.36)
Enter "ftp" or "anonymous" as the user ID
Enter your electronic mail address as the password
Change to the directory pub/ndp054 (or any NDP number you needed)
Acquire the files using the FTP "get" or "mget" command

or

World Wide Web URL: http://cdiac.esd.ornl.gov/cdiac/

Through the March of 1996, DOE GCRP-supported investigators had collected CO_2 measurements on 37 WOCE cruises. As of April 1996, CDIAC has received data from seventeen of these cruises. Ten of these databases have been checked and four fully documented.

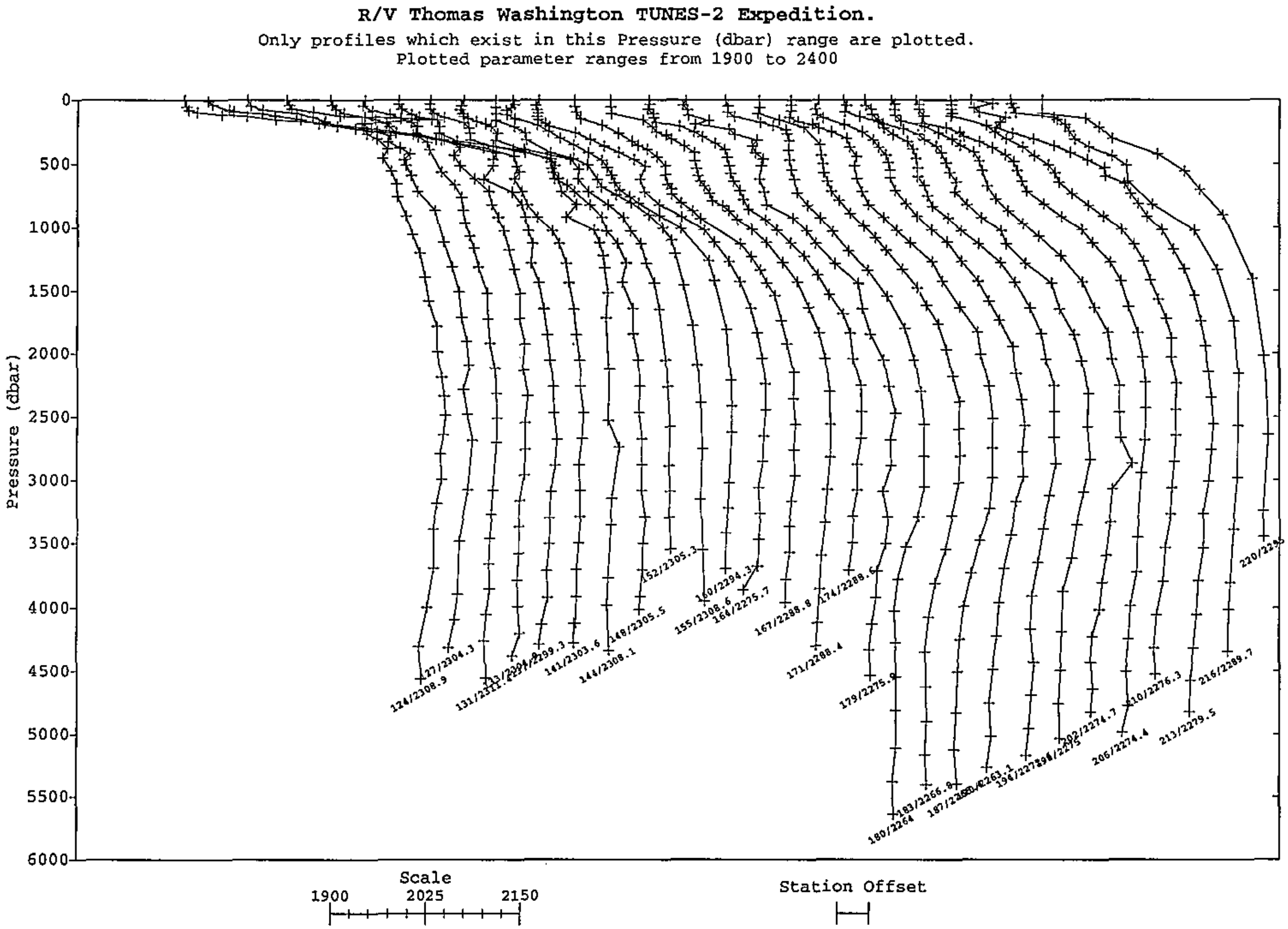


Figure 1. Nested profiles: total CO_2 (µmol/kg) vs pressure (dbar).

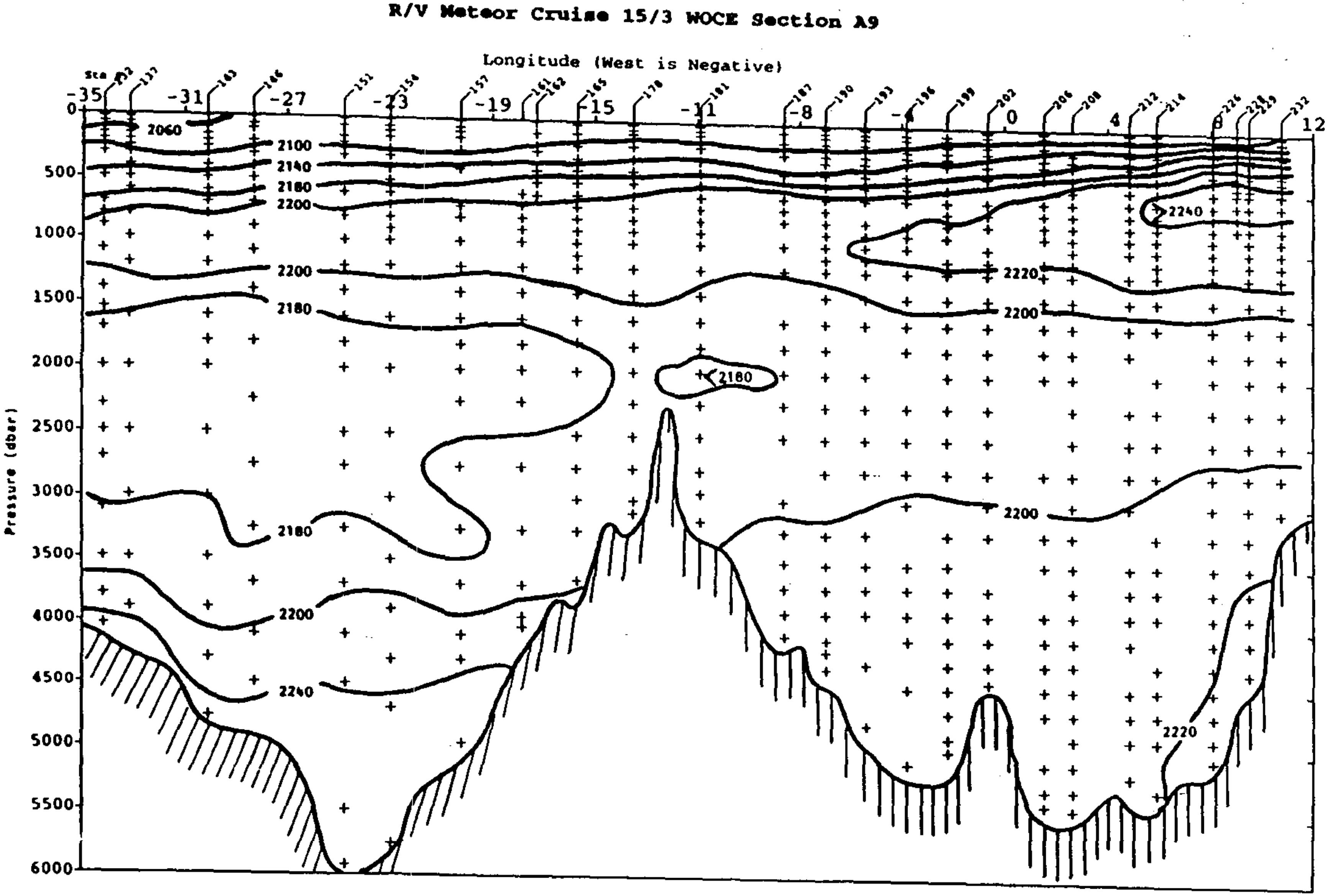


Figure 2. Distribution of the total CO_2 in seawater along WOCE Section A9.

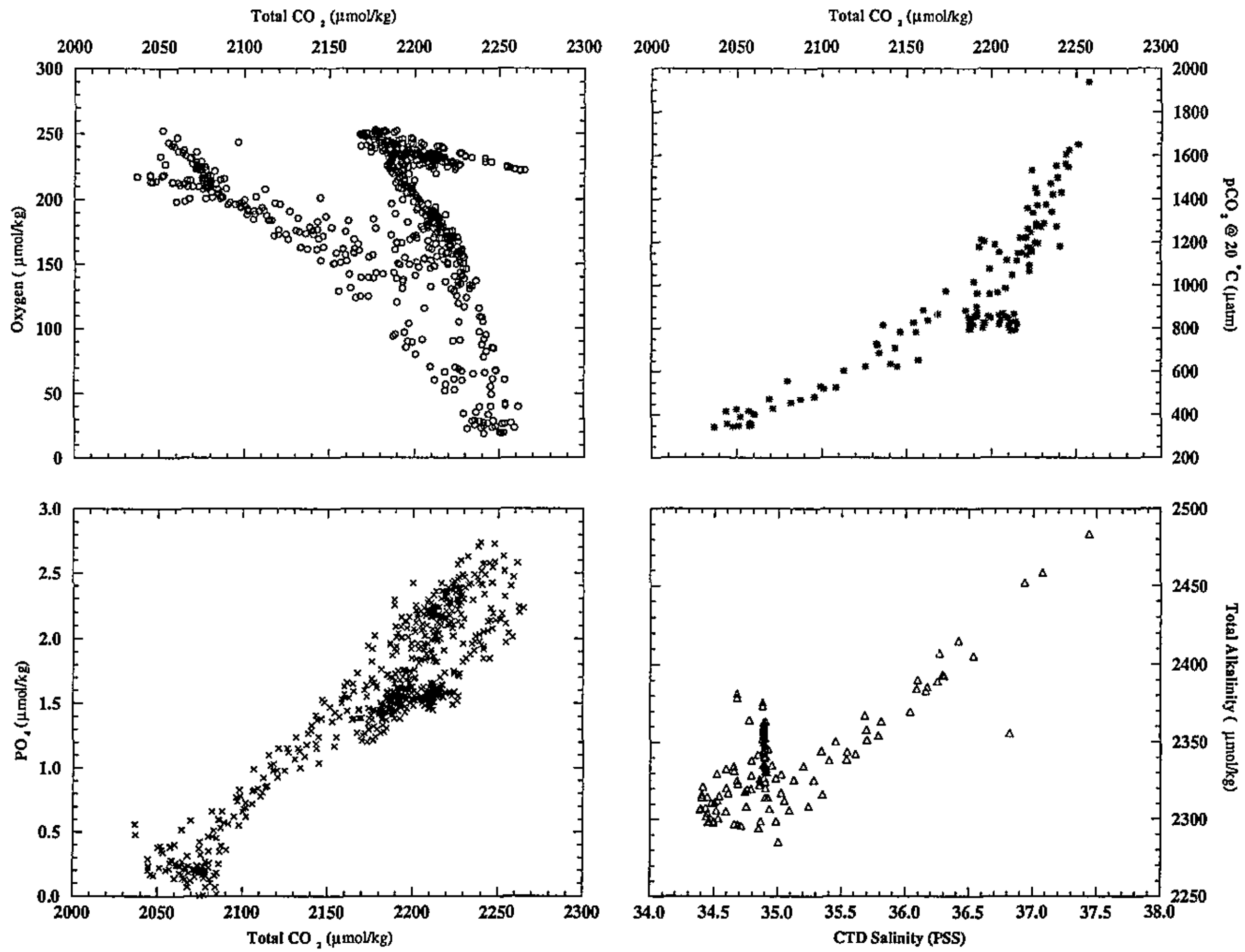


Figure 3. Property-property plots for all stations occupied during R/V *Meteor* Cruise 15/3.

CO_2 PARAMETERS: TOWARDS A DATA BASE?

Alain Poisson

The importance of the greenhouse effect of CO_2 led to an expansion in the number of scientists involved in the study of the fate of anthropogenic CO_2, and the number of measurements of the CO_2 parameters increased drastically in the nineties. Despite the present emphasis on studies of large scale and global phenomena, the original measurements of the various groups are usually archived separately rather than as a unified data set. In fact, relatively few such data are presently available in international data centers- The Joint IOC-JGOFS Advisory Panel on Carbon Dioxide recommended that an inventory of the $p(CO_2)$ measurements that had been made already be established and that such measurements be archived at an internationally recognized data center. During its meeting in La Jolla in June 1994, a Sub-Panel (Alain Poisson, chairman, Andrew Dickson and Hisayuki Inoue) was formed to set up an inventory of underway measurements of surface $p(CO_2)$ and to plan to establish a data base system that would make such measurements more widely available.

A simple strategy was followed:

- An international list of people that might be expected to have knowledge of such data was compiled;
- A letter was sent to the people that had been identified, explaining the goals of this project and asking for the location and dates of the $p(CO_2)$ measurements they had already made;
- The various national ocean data bases were asked to detail their current holdings of $p(CO_2)$ measurements;
- The inventory of the $p(CO_2)$ data was completed as a result of the individual inventories which were sent by 46 scientists from 19 countries;
- The CDIAC (Carbon Dioxide Information Analysis Center) in Oak Ridge, Tennessee, USA was approached to collaborate on this project and agreed to host the data base.

In order to store the data in a common unified form the sub-Panel proposed that the following information be archived:

- Date and time (UT) of measurement;
- Position of the measurement;

- $x(CO_2)$ - mole fraction of CO_2 - in dry air of air that would be in equilibrium with the surface sea water at the sea surface temperature and at a total pressure of 1 atm;
- $x(CO_2)$ in dry air Of atmospheric air (whenever measured);
- Fugacity of CO_2 in equilibrium with the surface sea water at 1 atm total pressure;
- Sea surface temperature;
- Sea surface salinity;
- Atmospheric pressure;
- Actual equilibration temperature;
- Measured wind speed and direction.

Each measurement in the final data base would be linked to the following "metadata" provided to ensure a permanent record of the original investigator's identity, as well as of the methodology used:

- Name and affiliation(s) of the investigator(s) responsible for the measurements;
- Identity of vessel used;
- Depth of sea water intake;
- Design of equilibrator;
- Method of CO_2 measurement (GC, IR, etc.);
- Information about the calibration gases used;
- Manufacturer;
- Date of certification;
- Traceability of certification (if known);
- Expected accuracy of certification;
- Concentrations of calibration gases used- Frequency of calibration;

- Information about the calibration of the pressure and temperature measuring instruments that were used. In particular, an estimate of the quality of sea surface temperature measurements;

- How were any data corrections (e.g. to sea surface temperature) made.;

- Estimate of the overall measurement precision and accuracy;

- References to other reports or papers that have been published describing this data set.

A more detailed request has been sent to the participants asking for specific information about their measurements (metadata) and for the measurement data themselves. CDIAC will compile the initial data base from the replies received, performing necessary conversions to put the $p(CO_2)$ data received into a common format, using the algorithms prepared by the Sub-Panel. A strategy will be designed for obtaining and incorporating recent and future measurements into the data base.

Is the archiving of the three other measurable CO_2 parameters an urgent need? Dissolved Inorganic Carbon (DIC) and Total Alkalinity (TA) are used by the modelers to calculate pCO_2; a great number of measurements of these two parameters were recently performed, mainly during the cruises of the WOCE and JGOFS programs. The techniques of measurement are now reliable and they could be archived easily in a near future. For the pH, the situation is less clear and more work is needed on the technique used to measure this parameter as well as on the standards to obtain accurate data, prior to archive this parameter.

EMERGING TECHNOLOGIES IN BIOLOGICAL, CHEMICAL, OPTICAL, AND PHYSICAL SAMPLING OF THE OCEANS

Tommy D. Dickey

INTRODUCTION

Global change involves a diverse and complex set of scientific problems related to the biology and chemistry of our planet as well as its physics. Thus, interdisciplinary research directed toward global scale problems has become an exciting new scientific focus. Future global ocean monitoring and modeling programs will require comprehensive sampling of the biological, chemical, optical, and physical environment. A diverse set of sampling platforms including ships (both on station and underway), moorings, drifters, autonomous underwater vehicles, and satellites along with sophisticated models will be necessary (Figs. 1 and 2). These platforms will require sensors capable of sampling a broad suite of interdisciplinary variables which, together with models, can be utilized to provide information covering up to nine orders of magnitude in time and space scales. International programs such as the Joint Global Ocean Flux Study (JGOFS) and the Global Ocean Ecosystem Dynamics study (GLOBEC are addressing key issues concerning the environment and its ecology in relation to climate change. JGOFS issues involve processes that control the fate of carbon and other biologically active elements , in the ocean whereas GLOBEC questions relate to the coupling of ocean physics with population dynamics of zooplankton and fish. New technologies are being developed for these efforts. These technologies involve emerging biological, chemical, optical, acoustical, and physical sensors integrated into systems which can measure several variables concurrently. Another important aspect is the telemetry of high frequency, interdisciplinary data in near real-time. Some highlights of technological developments and sampling strategies are presented here. Importantly, these collective activities may be used to assist in the planning and implementation of interdisciplinary measurement components of the planned International Global Ocean Observing System (GOOS).

SAMPLING CONSIDERATIONS

The selection and proper sampling (in space and time) of key variables for studies of the marine ecosystem including the carbon flux problem and the recruitment of larval fish and population dynamics are daunting tasks (e.g., Dickey, 1988, 1990, 1991, 1993; Jahnke, 1990; Dickey et al., 1993a). These studies encompass interdisciplinary processes whose scales of variability span at least nine orders of magnitude (e.g., Haury et al., 1978; Dickey, 1991). There are some dominant temporal and spatial scales of variability (e.g., the diurnal and seasonal cycles, inertial oscillations, fronts, and mesoscale eddies), however, episodic phenomena represent critical events for biogeochemical fluxes as well as for individual organisms and ultimately population dynamics. There are obvious nonlinear interactions to contend with as well. High temporal resolution measurements are needed for long periods of time. Analogously, sampling of patchiness in biogeochemical and physical properties and organismal groups on a broad range of scales is necessary (e.g., Cowles et al., 1990). Clearly, a variety of specialized sampling platforms, each with its own attributes and deficiencies, must be employed. A schematic illustrating a nested experimental array of such platforms is shown in Fig. 1. Utilizing data from such arrays, numerical models can be used to synthesize the data and to enable diagnostics and predictions (Fig. 2) (e.g., see GLOBEC Special Contribution No. 2, 1996). An important link between data collection and modeling is real-time or near real-time telemetry of data. Some of the presently utilized, emerging, and needed interdisciplinary sensors and systems are described below. More detailed reviews with extensive references to specific instruments and systems are given in Dickey (1988, 1990, 1991, 1993), U.S. GLOBEC Report 3 (1991), U.S. GLOBEC Report 4 (1991), ICES Study Group (1992), special volume of Arch. Hydrobiol. Ergebn. Limnol., 36 (1992), SCOR Working Group 90 (1993), Dickey et al. (1993a,b; 1996), GLOBEC Report 3 (1993), U.S. GLOBEC Report 8 (1993), and U.S. JGOFS Planning Report 18 (1993).

SENSORS

Advances in interdisciplinary instrumentation systems have been closely linked with the development of a variety of new sensors. Instruments created for use in upper ocean current studies include mechanical, acoustical, and electromagnetic current meters, acoustical Doppler current profiling systems (e.g., ADCP's) which can be used in either vertical- or horizontal-looking attitudes, surface HF radar systems, drifters, and drogues (see review by Dickey et al., 1996). Some of these instruments may be used in either moored or vertical profiling modes and can resolve time scales from minutes to several months or a few meters to a few hundred meters in the vertical depending on deployment strategy. Acoustic tomography involves the measurement of the field of sound speed fluctuations within a control volume by transmitting acoustic signals along several diverse paths. One of the attractive features of this technique is that a relatively large volume (on order of 100's of kilometers or more in the horizontal) of the ocean can be sampled synoptically. The measurement of vertical velocities has been most difficult, however measurements of vertical velocities have been done by using ADCP's and mechanical vector measuring current meters (VMCM's). Special velocity and temperature microstructure devices are capable of resolving vertical scales on order of 2 and 1 cm, respectively. These can also be used to estimate vertical diffusivity and turbulent fluxes. These special devices are expensive to operate and presently remain tools limited to a few experts. Nonetheless, their availability to the broader community is highly desirable.

New optical methods may be subdivided as follows. The first involves measurement of small scale (on order of a few manometers in wavelength and tens of microns in particle size) optical properties using various light sensors (e.g., Yentsch and Yentsch, 1984; Dickey, 1991, U.S. JGOFS Report No. 18, 1993). The second involves imagery of organisms or video techniques (scales of a few microns to a meter depending on imaging optics). Two primary functions of small scale in situ bio-optical measurements are 1) to enable the determination of the intensity and wavelength of light available for photosynthesis at depth and 2) to facilitate the identification and the quantification of phytoplankton populations (including growth rates) and their products. Some of the sensors described below may be used for one or both of these purposes and provide virtually continuous sampling with vertical resolution comparable to conductivity, temperature, depth (CTD) systems (few meters or less) and temporal resolution comparable to moored current meters (few minutes or less). Photosynthetically available radiation (PAR) sensors measure scalar irradiance in the visible waveband using a spherical light collector. More sophisticated optical instruments for quantifying the oceanic photo environment include multi-wavelength spectroradiometers. In particular, high spectral resolution (-2 manometers) radiometers are now being developed. In situ fluorometers are used to obtain nearly continuous records of fluorescence in order to estimate chlorophyll-a concentration and to infer phytoplankton pigment biomass. The recently developed fast repetition rate (two-pulse) fluorometer shows great promise for measurements of basic photosynthetic parameters as well as primary production. The beam transmissometer measures an inherent optical property of seawater, the beam attenuation coefficient, which relates to the volume of suspended matter or particle concentration in the water column.

A variety of sensors are being developed to measure a more comprehensive set of optical variables so that inherent optical properties (those independent of a natural light source) and apparent optical properties (those dependent on a natural light source) may be related. Emerging instrumentation which will facilitate characterization of inherent optical properties and will be useful for detailed study of various phytoplankton groups and primary productivity are spectral absorption and scattering meters and spectral fluorometers. The use of fiber optics to bring light signals from depth to the surface for shipboard (or surface buoy or drifter) signal processing and data analysis has been shown to be a viable option for several physical and bio-optical applications. A newly developed laser/fiber optic fluorometer which is used in parallel with a physical microstructure profiling instrument can provide vertical resolution of fluorescence on the centimeter scale. Expendable probes capable of measuring biologically and chemically relevant variables are likely to be developed. Such probes could enable broad geographic coverage and are especially useful for intensive regional studies as well as global surveys. Other in situ optical instruments for determining particle size distributions are being developed as well. Camera systems have been developed to examine settling rates of large aggregates of material in conjunction with sediment traps.

Various systems have been developed for the study of zooplankton and higher trophic level organisms. Promising optical methods, which have potential application for in situ small scale predator-prey studies, include Schlieren video systems and holographic systems (e.g., Dickey, 1988, 1990). Electronic and optical particle counters have been used to determine zooplankton abundance and physical dimensions indirectly (Sprules et al., 1992). The latter systems have been deployed in tow-yo and mooring modes. Video systems using CCD arrays are being utilized quite successfully from ships and specialized optics have been applied to view organisms down to scales of a few microns (e.g., Davis et al., 1992). Automated image analysis systems are being developed to attack the difficult problem of discriminating between taxonomic groups and larval stages.

Acoustical methods are attractive because they are non-intrusive and can be used for broad spatial coverage applications (e.g., Smith et al., 1992). They are generally more effective for larger scale organisms (roughly 100 microns to a meter or more, depending on the transducer used) than optical methods. Several different acoustical approaches have been suggested. For example, individual targets differing in size have different target strengths for differing frequencies. This principle has been exploited in the development of multi-frequency (up to 21 frequencies) and dual-beam acoustic systems which have been deployed in both shipboard (profile or towed) and moored modes. Another utilizes backscatter strength fluctuations to count the number of targets in a given volume. This technique was motivated by needs for fish stock assessment. Finally one of the more promising approaches to obtaining large volumes of data over broad areal extent is simultaneous sampling of currents and scattering intensity related to organisms with some ADCP's (e.g., Smith et al., 1992) which are increasingly being deployed from research and commercial vessels.

SYSTEMS

Many of the sensors described above may be considered to be modules, which can be interfaced with submersible packages including data acquisition systems and microprocessors. Two of the primary goals of interdisciplinary in situ measurement systems are: 1) to sample with complementary, interdisciplinary sensors as closely in space and time as possible and 2) to sample temporal and spatial scales of the ecosystem so as to avoid aliasing and undersampling. These include specialized profilers and tow-yo or undulating packages. They are often CTD-based packages which utilize extra data channels for optical and acoustical measurements. An important tool continues to be sophisticated multiple opening and closing net systems. These are needed to directly determine speciation and abundances of zooplankton and micronekton. These systems may now include: thermistors, conductivity sensors, fluorometers, dissolved oxygen sensors, and light sensors. A few investigators have also added acoustical sensors and video systems to their net systems to enable more detailed and quantitative analyses. Continuous plankton recording (CPR) systems have been utilized on ships of opportunity by the British for North Atlantic surveys for the past few decades (e.g., Quartley and Reid, 1996). This approach has been highly productive and the planned enhancement of these systems with emerging optical and acoustical sensors will likely lead to many new insights.

Multi-variable moored systems have been used to do time series measurements of several bio-optical and physical parameters as well as dissolved oxygen in several oceanic regions within the past few years. Recently, moored measurements relevant to zooplankton have been done using optical and acoustical systems. In addition, unattended multi-variable profilers have been used for measuring the vertical distributions of bio-optical and physical parameters.

In addition, new biogeochemical sensors for measurement of chemical variables of interest (e.g., nutrients such as nitrate, the partial pressure of carbon dioxide, and several other parameters) are being developed and tested on moorings as well as bottom tripods (benthic landers). Such tripods have been used to examine chemical reactions and fluxes in the uppermost sediments as related to the fate of biogenic materials produced in the upper layers of the ocean. Finally, trace element samplers are also being used on moorings in the open ocean.

For several years, drifters, drogues, and subsurface floats have been utilized by physical oceanographers for current measurements, however integration of biological and optical sensors with these systems is relatively recent. One of the principal attractions of drifters and drogues, which are equipped with bio-optical instrumentation, is that broad geographical extents can be sampled. In addition, acoustical and optical sensors placed on Lagrangian platforms can in principle provide data on the environment as experienced by a drifting organism. Some new drifters can also profile to depth and return data via satellite while at the surface. Another interesting approach being pursued at present involves a "smart drifter" which is intended to mimic larval behavior such as vertical migration.

Data have been transmitted in near real-time from interdisciplinary profilers and moorings, primarily in coastal studies or in proximity of ships (e.g., Dickey et al., 1993a,b) as well as in the open ocean (Frye and Owens, 1991). In order for this approach to be viable for interdisciplinary applications in the open ocean, broader bandwidth capabilities for data transmission will be required. Low earth orbit (LEO) satellites with such capability are being developed.

COMPLEMENTARY SAMPLING

Most of the in situ instrumentation and systems described above are particularly well-suited for shipboard (on-station and underway sampling), mooring, and drifter data acquisition. However, they cannot provide synoptic data with fine horizontal resolution over extensive geographical regions. For this reason, satellite-borne (and to some degree airplane-borne) remote sensing systems are especially important for oceanographers. The potential geographical coverage of satellite-borne sensors is virtually global with spatial resolution dependent upon the area observed or the footprint of the sensor. The altimeter being used for the Ocean Topography Experiment (TOPEX) is capable of resolving subtle currents of the world ocean with resolvable spatial and temporal scales of about 5 kilometers and 10 days, respectively. The anticipated data to be obtained from the satellite-based ocean color sensors of SeaWiFS (Sea-viewing, Wide-Field-of view Sensor) and several other ocean color imagers will be important for many coastal and open ocean programs. Synthetic aperture radar (SAR) sensors have the advantage of imaging even in the presence of clouds with even greater spatial resolution and can provide images of features such as fronts, slicks, and surface and internal gravity waves. Scatterometers provide important wind stress data. Airplane platforms may be used for deployment of many of these sensors as well. With these platforms, cloud problems are mitigated or eliminated and improved spatial resolution is derived. Many individual nations and groups of nations are applying remote sensing technologies to the ocean and significant advances are sure to follow.

Importantly, in situ data collected from moorings, drifters, and ships, provide a critical link for satellite sensor "ground truthing" as well as providing subsurface information and continuity of data during cloudy periods. Other potential ways to deploy in situ bio-optical, acoustical, and physical instrumentation packages in the future may include submarines and autonomous underwater vehicles (AUV's).

CONCLUSIONS

The global problems addressed by programs such as JGOFS and GLOBEC at present and by GOOS in the future require the consideration of biological, chemical, optical, and physical processes which can span scales up to nine orders of magnitude. Sampling remains a central issue in regard to the selection of key measurements as well as resolution and range. There has been a remarkable growth in the number of bio-optical, acoustical, and chemical sensors and systems which can be applied. Methods and instruments to obtain size distributions and taxonomic identification of phytoplankton and zooplankton (as well as their larval stages) are still needed. Techniques to observe organismal behavior and predator-prey interactions are important as well. Instrument standardization and calibration and interpretation of signals in a biological context remain as important issues. It is evident that intensive field testing and "ground truthing" of in situ sensors must be done periodically with standard shipboard sampling at representative oceanic sites. Simultaneous video and acoustical observations are likely to be quite effective. These efforts, which are somewhat analogous to the "ground truthing" of satellite-borne sensors, will be important for building our confidence in systems which will expand our interdisciplinary databases by several orders of magnitude. Further, selection of sensors and systems which are economical, as well as effective, is critical as large numbers must be deployed for global coverage of our presently undersampled environment.

Strides have been taken in interdisciplinary observational technology within the past decade. A most challenging aspect of emerging technologies is calibration, management, and optimal utilization of large volumes of optical and acoustical data (Dickey et al., 1993a). Careful consideration of time and space scales associated with the more energetic processes can be used to optimize both sampling and modeling. Yet the task of understanding processes whose scales lie between those scales which can more easily measured and modeled remains formidable. An important step for modelers and observationalists is to be able to accurately parameterize the smaller scale processes such as turbulence, particle motion, and animal feeding activities. It is imperative that modelers and technologists work together on these problems. Data assimilation modeling will be a vital aspect of GOOS (e.g., LMR-GOOS Planning Workshop, 1996).

REFERENCES

Cowles, T.J., R.A. Desiderio, J.N. Mourn, M.L. Myrick, and S.M. Angel, 1990, Fluorescence microstructure using a laser/fiber optic profiler, Proc. SPIE-Int. Soc. Opt. Eng., 1302, 336-345.

Davis, C.S., S.M. Gallager, M.S. Berman, L.R. Haury, and J.R. Stickler, 1992, The video plankton recorder (VPR): design and initial results, Arch. Hydrobiol. Beih. Ergebn. Limnol., 36, 67-81.

Dickey, T.D., 1988, Recent advances and future directions in multidisciplinary in situ oceanographic measurement systems, In Toward a Theory on Biological-Physical Interactions in the World Ocean, B.J. Rothschild, ed., Kluwer Academic, Dordrecht, The Netherlands, 555-5 9 8.

Dickey, T.D., 1990,Physical-optical-biological scales relevant to recruitment in largemarine ecosystems, in Large Marine Ecosystems: Patterns, Processes, and Yields, eds. K. Sherman, L.M. Alexander, and B.D. Gold, AAAS Press, Washington, DC, 82-98.

Dickey, T., 1991, The emergence of concurrent high resolution physical and bio-optical measurements in the upper ocean and their applications, Reviews of Geophysics, 29, 383-413.

Dickey, T., 1993, Technology and related developments for interdisciplinary global studies, Sea Technology, August, 1993, 47-53.

Dickey, T.D., T.C. Granata, and 1. Taupier-Letage, 1993a, Automated in situ observations of upper ocean biogeochemistry, bio-optics, and physics and their potential use for global studies, Proceedings of the Ocean Climate Data Workshop, Goddard Space Flight Center, Greenbelt, Maryland, Feb. 18-21, 1992, 317-352.

Dickey, T.D., R.H. Douglass, D. Manov, D. Bogucki, P.C. Walker, and P. Petrelis, 1993b, An experiment in duplex communication with a multi-variable moored system in coastal waters, Journal of Atmospheric and Oceanic Technology, 10, 637-644.

Dickey, T., A. Plueddemann, and R. Weller, 1996, Current and water property measurements in the coastal ocean, The Sea, in press.

Frye, D.E. and W.B. Owens, 1991, Recent developments in ocean data telemetry at Woods Hole Oceanographic Institution, IEEE Journal of Oceanic Engineering, 16, 350-359.

GLOBEC Report No. 3, 1993, Sampling and Observing Systems, T. Dickey (ed.), GLOBEC International, Chesapeake Biological Laboratory. Solomons, MD, 20688.

GLOBEC Special Contribution No. 2, 1996, An Advanced Modeling/Observation System (AMOS) for Physical-BiologicalChemical Ecosystem Research and Monitoring, A Working Paper/Technical Report prepared by the GLOBEC.INT Working Groups on Numerical Modeling and Sampling and Observational Systems, eds. A. Robinson and T. Dickey, GLOBEC International, in press.

Haury, L.R., J.A. McGowan, and P.H. Wiebe, 1978, Patterns and processes in the time-space scales of plankton distributions, in Spatial Pattern in Plankton Communities, ed. J.H. Steele, Plenum Press, New York, 277-314.

ICES Study Group, 1992, Report of the ICES Study Group on Zooplankton Production, Bergen Norway, March 23-26, ed. H.R. Skjoldal, 22pp.

Jahnke, R.A., 1990, Ocean flux studies: a status report, Reviews of Geophysics, 28, 381-398.

LMR-GOOS Planning Workshop Report, 1996, Report of the Planning Workshop for the Living Marine Resources Module of the Global Ocean Observing System, Center for Marine Sciences and Technology, University of Massachusetts, Dartmouth, March 1-5, 1996, SCOR, 68pp.

Quartley, C.P. and P.C. Reid, 1996, Long-term oceanographic data sets, Sea Technology, March 1996, 68-70.

SCOR Working Group 90, 1993, Emerging Technologies in Biological Sampling, a Report of SCOR Working Group 90, UNESCO Technical Paper in Marine Sciences, no. 66, ed. A. Herman, 48pp.

Smith, S., R. Pieper, M. Moore, L. Rudstam, C. Greene, C. Flagg, C Williamson, and J. Zamon, 1992, Acoustic techniques for the in situ observation of zooplankton, Arch. Hydrobiol. Beih. Ergebn. Limnol., 369 23-43.

Sprules, W.G., B. Bergstrom, H. Cyr, B.R. Hargraves, S.S. Kilham, H.J. MacIsaac, K. Matsushita, R.S. Stemberger, and R. Williams, 1992, Non-video optical instruments for studying zooplankton distribution and abundance, Arch. Hydrobiol. Beih. Ergebn. Limnol., 36, 45-58.

U.S. GLOBEC Report No. 3, 199 1, GLOBEC Workshop on Biotechnology Applications to Field Studies of Zooplankton, Division of Environmental Studies, University of California, Davis, CA, eds. L.S. Incze and P.J. Walsh, 29pp.

U.S. GLOBEC Report No. 4, 199 1, Workshop on Acoustical Technology and the Integration of Acoustical and Optical Sampling Methods, JOI, Washington, DC, eds. C. Greene, C. Greenlaw, V. Holliday, P. Ortner, R. Pieper, T. Stanton, and J. Traynor, 58pp.

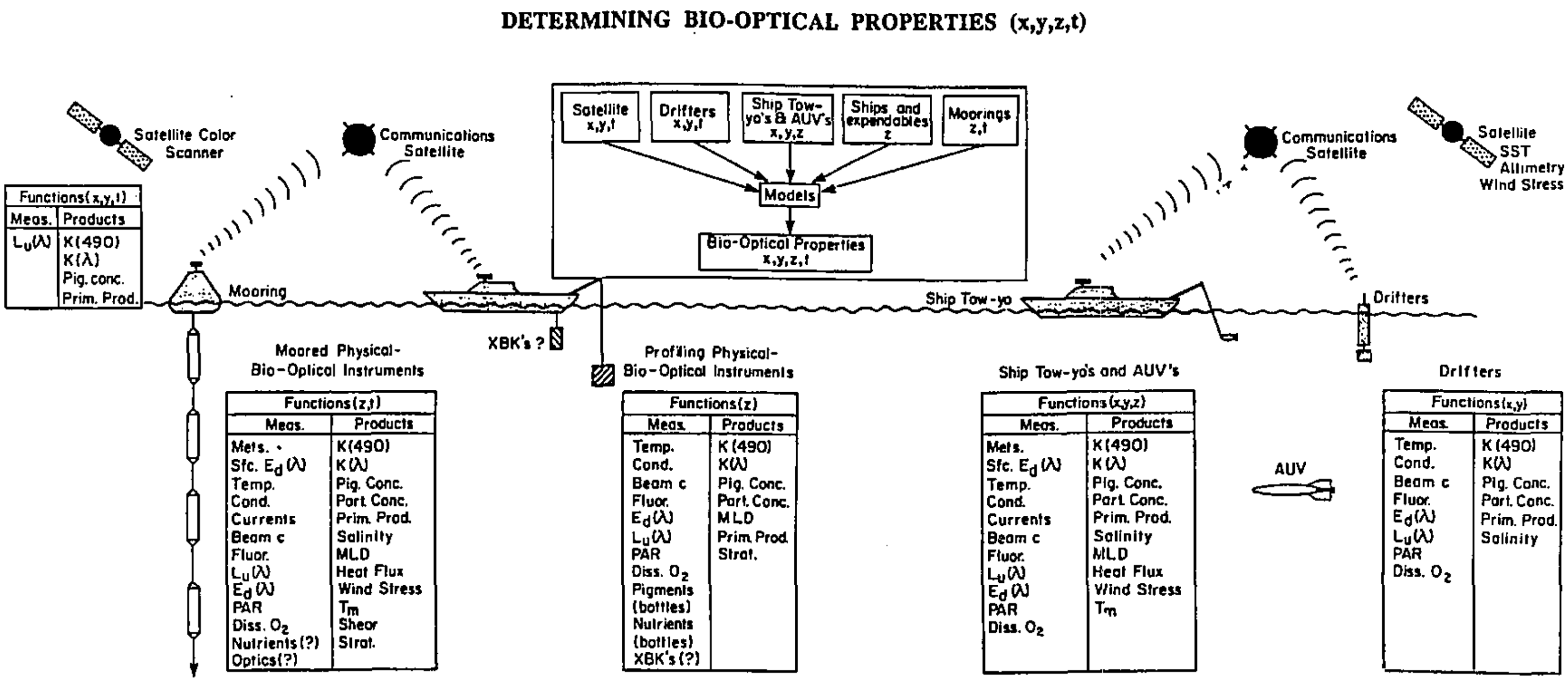


Figure 2: A schematic illustrating a methodology for determining the variability of the interdisciplinary variables (e.g., bio-optical properties as indicated here) in space and time on a global basis using satellite as well as *in situ* data sets along with appropriate models (after Dickey, 1991).

THE STATE OF THE HYDROCHEMICAL DATABASE IN RUSSIA

V. Sapozhnikov

The total hydrochemical database is comprised of 280,000 stations. These are all the stations, where at least one hydrochemical analysis was made, for example, oxygen determination in the surface layer.

As to the number of determinations, there are as follows:

Oxygen	230,000
pH	68,000
Alk	12,000
PO_4	92,900
P total	9,000
NO_2	26,000
NO_3	20,000
NH_4	29,500
Si	50,900

These figures are presented in *OCEANOLOGY*, 6, 1995, and do not include total organic nitrogen (TON), urea, total organic carbon (TOC), proteins, lipids, carbohydrates, enzymes (the activity of alkaline phosphatase and the ETS, lactate dehydrogenase, etc.). The direct oxygen determinations of primary production and biological oxygen demand (BOD) and others are also absent.

This database includes the USSR data and the data received on exchange conditions via RNODC in Moscow. There are still about 300 cruises which are still not digitized.

The database has a lot of faults, especially in the data obtained 20 years ago and even earlier, when the chemical analytical procedure was not perfect. But there are some approaches to correct data, introducing a respective coefficient, e.g. the coefficient of the column reduction for nitrates, either correcting erroneous titre, if there was a detailed description of the methods applied during the cruise.

The monograph *The Pacific Ocean* presented the data of about 7,000 stations checked by the following method: the diagrams of the vertical distribution of nutrient salts were plotted and compared on the Redfield-Richards stoichiometric ratio. Thus individual anomalies observed at some horizons could be marked out.

A lot of methodological problems have not been solved yet. Thus there is not an ultrapure deionized water, completely free from ammonium, dissolved organic nitrogen (DON), carbon (TOC), phosphorus (DOP), urea, consequently, there is no "pure" blank. In case of ammonium and DOP, one has to apply admixtures, or select water samples of the natural seawater (from various horizons) with rather low respective values. However now the problem is settled using new sorbents (a specially prepared activated coal).

The averaging of the results also causes problems. It seems reasonable to differentiate various climatic periods, e.g. El Niño and Anti-El Niño, which differ significantly even in such a stable area as the Subantarctic waters (30° - 50° S).

During the Anti-El Niño period, silica acid, nitrates, and phosphates are almost completely utilized in the surface waters of the Subantarctic waters up to 48° S, while in the El Niño years the isolines of the 10 uM nitrates and 15 uM silicates spread up to 32° - 35° S.

Similar features are observed in the northern Pacific: in the Sea of Okhotsk, the thickness of the surface mixed layer totals 8 - 15 m during cold periods and 20 - 35 m - in warm periods. If we average the thickness of the surface uniform layer, for example, to 20 m, we shall not see any difference in the nutrient stocks in the euphotic layer. Therefore we shall increase the values for cold periods, and decrease them for warm periods.

Now we have the following situation: the State Oceanographic Institute of the Hydrometeorological Committee are factually trying to do the quality control of the hydrochemical data, by stating reasonable limits of the maximal and minimal values. It is not yet clear what we shall get with this. A longer and a more complicated way, but I think, the right one, has been undertaken by Shirshov's Oceanology Institute of Russian Academy of Sciences and our Institute (VNIRO). We are plotting the vertical curves for all the hydrochemical characteristics, making an accurate quality control, and only then the data are stored in the database.

I would like to mention a very timely and urgent work undertaken by NODC within the framework of the GODAR program. If the old hydrochemical data were not found during the coming years (and sometimes the work means to find original log-books of the analytical determinations, introduce blank correction, count in the titres, etc.) they would be lost forever. We still have alive those researchers who performed the analyses and remember the details of the analytical procedures. Here is a typical example, the nitrate data of the 15th cruise of the RN Akademik Knipovich (the Bellingshausen Sea, the Pacific ocean, 1978) were multiplied by the coefficient of the column reduction, however the higher coefficients were due to the standards prepared on distilled water. The columns performed well. The real coefficient was about 1. After "interrogations" of the analyst we managed to rescue the whole set of the data of this cruise .

THE DATA BANK MANAGEMENT SYSTEM OF THE NODC OF GERMANY

Dr. K. Motamedi, R. Schwabe and Dr. H.R. Vatterrott

ABSTRACT

The Marine Environmental Data Base (MUDAB) is a joint project of the Federal Maritime and Hydrographic Agency (BSH) in Hamburg and the Federal Environmental Agency (UBA) in Berlin. It is located in the German National Oceanographic Data Centre and serves as the central German data base for marine data within the framework of international and national conventions for the protection of the North Sea and the Baltic Sea.

MUDAB has been developed at BSH and is based on the relational database management system INGRES. It is designed in client server architecture. The database server is running on a CRAY Super Server 6400.

There are two types of user interfaces for MUDAB - an alphanumeric and a graphical interface. The alphanumeric interface has been developed at BSH and is based on 4GL as programming language and MS Windows 3.11 as operation system. The graphical interface has been developed at Frauenhofer Institute for Computer Graphics, branch Rostock, and is based on PV-Wave as a software package for developing graphical interfaces.

The decision to build two types of user interfaces has been made in view of very different hard- and software conditions of potential users. The minimal requirement for using MUDAB in connection with the alphanumeric interface is a personal computer with a 486 processor and 8 MB RAM. Currently the database contains data and information of about 4,500 cruises covering about 150,000 stations, altogether more than two million records. The data comprise mainly physical parameters, such as temperature and salinity, chemical parameter such as nutrients, and organic, inorganic and radiochemical components of sea water as well as physical and chemical parameters in sediment.

The database schema considers a large amount of meta-information of different levels, e.g. Cruise Summery Report (CSR) for each cruise, information about measured parameters and measurement methods, intercomparisons and so on. Every value can be linked to any meta-information, even information of other structural levels.

The database design facilitates easy inclusion of new types of data. It is possible to define sub-databases for new data types, whereas type classification is dependent on the medium of measured data, space and time as well as the specific structure of the new data type. At present, sub-databases for sediment, water and biological determinants have been defined. The sub-database for sea water covers casts, time series as well as CTD/XBT data.

The alphanumeric interface is menu based and has been developed to guide the user in retrieving information. Starting point is either the cruise, the station or the data level. However, it is possible to retrieve information by formulation of SQL requests as well. Criteria for database retrieval are constraints of space and time as well as measured parameters and measurement methods. Currently, retrieved information is exported for further data processing in a data format understood by many PC based tools. In the near future data export in standard data formats and an interface for biological data will be implemented.

The graphical user interface consists of two main components: an easy to use database front end on one hand and an interactive system for fast visualization of marine environmental data (VISMAR) on the other hand. Data retrieval in MUDAB is carried out by defining basic requirements such as project, cruise, geographical category (e.g. geographical position and depth, time and measured parameters). VISMAR supports the database front end by providing graphical means for definition of geographical positions and areas. Retrieved data can be visualized by VISMAR immediately. This way the user is able to get a general idea about available information. For data visualization VISMAR provides a wide range of 2D and 3D visualization methods such as icons, box & whisker, isolines, different charts and wire frames.

USE OF OBJECTS IN THE MANAGEMENT OF MARINE DATA: EXAMPLE OF A DATABASE AMENABLE TO GEOSTATISTICAL ANALYSIS

Marek Ostrowski

Advancements in computer technology in the recent years have resulted in the emergence of a new approach towards the way in which the software is being designed and used. This is known as object orientation (OO). OO is based on assumption that the state and behavior of a unit of computerized information are non-separable. Within the framework of this approach, data and software managing the data are packaged together and are perceived as an object, namely an inseparable unit of information which exist at both the semantic and implementation levels. Objects already implemented may be combined together to provide a new functionality as well as new semantic meanings. By building hierarchies of new objects on top of already implemented objects, and hiding away (encapsulating) internal relationships redundant at the next, higher semantic level it is possible to create the complex systems which function at each level with a limited, hence manageable set of independent components. While the OO approach has been proven to be very successful in the development of complex software systems, for many other computer intensive jobs, including marine data handling, its role has yet to be defined.

This presentation showed a practical example of applying the OO approach in the domain of marine data handling. The example uses the data management system being developed to meet data management requirements in the EU sponsored project: "Geostatistic for Fish Stock Assessment". An abstract object that defines data retrieval methods applicable to a wide range of data analysis is identified and implemented for data sources with highly variable content and internal organization. Such an identification and the abstraction of a common object class, allow for subsequent treatment of all data types by the processing software as if they belong to the same file type. The exposed object class represents a query for time invariant trackline data (latitude, longitude, value per data cycle). Its implementations are built on top of three vastly diverse data storage systems: a spreadsheet, a DBMS, and the on-line GLOBEC data from the Internet. The data analysis tools, in this case a map visualization software and a variogram calculation program, retrieve the data on-line using the methods defined in abstract object protocol without any regard to how and where the actual data have been stored. The example does demonstrate a potential of the OO approach for achieving with a minimum effort an integrated solution featuring advanced processing capabilities. This can be done with almost no programming using the data that remain in their native formats and software pieces developed separately. In the view of the growing number of oceanographic data products published on CD-ROM's and available via the Internet, methodology similar to the one shown here may provide a least-effort approach to the integration of various marine data products in a way which will provide data access, analysis and presentation capabilities beyond the level anticipated by their originators.

SESSION IV: BIOLOGICAL AND CHEMICAL DATA MANAGEMENT

Convener - Gordon Paterson (U.K.)

QA/QC OF HISTORICAL NUTRIENT DATA

Chris and Jean Garside

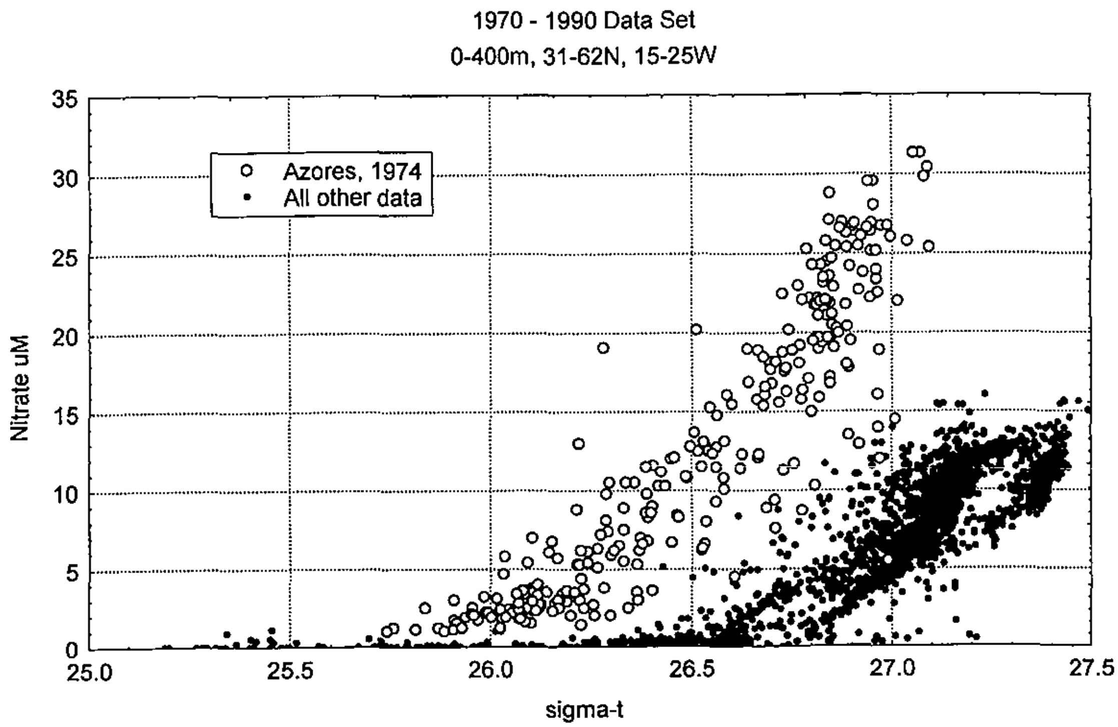
The provenance of historical nutrient data, especially those data in databases, is often difficult or impossible to establish. Databases and database programs often were not designed to accommodate metadata. The metadata that we expect to find in current data archives is generally either incomplete or absent. Sampling techniques, sample preparation and preservation, the analytical methodology, data reduction, accuracy, precision, correction and verification are increasingly rare with increasing age of the data set.

Frequently all that exists in a data record is a location/depth, date/time, T/S and some nutrient concentrations. With little or no metadata available the only tests we can apply to individual data points is how consistent they are with respect to other data presumed to be sufficiently similar to provide a meaningful comparison. Often there is not a high enough density of data and sufficiently similar data may not be available. As a result the comparison is often the rest of the available population of data the entire data set. A data set comprising 2600 hydrographic and nutrient data records from 0 - 400 m taken during all seasons for the last 25 years in an area from 31° - 62° N and 15° - 25° W will be used for illustration.

Knowing what to presume is close enough to sufficiently similar data is an art at best. A frequently used tool to visualize similarity or otherwise is plotting nutrient data against another property. Commonly used property / property plots include nutrient versus depth, temperature, salinity or density. The data user is obliged to make subjective decisions about what data are acceptable and what are to be rejected based on how well they plot relative to the majority of the rest of the data. Unfortunately, with large data sets spanning many years and a large geographical area, there is sufficient natural variance that questionable data can be difficult to identify.

Deep ocean samples are often compared to density because the relationship is well established for many oceanic regions. The time scales of nutrient and water cycling in the deep ocean basins are long and similar so that the relationship of nutrient concentration to density is sufficiently singular for it to be useful in some cases. In the test data, which is essentially an upper water column set, it is evident in a nitrate versus density plot that we have some data that appear to have too high concentrations compared to all other data. Once we have isolated them, we find that they are all from the fall of 1974, and the vicinity of the Azores. Since there are other data from the region that are lower and consistent with the rest of the data set we can set these aside as suspect. Unfortunately it is equally evident from this plot that low nutrient concentrations typical of the euphotic zone can be found at almost any density, so that this becomes an insensitive tool in the upper water column.

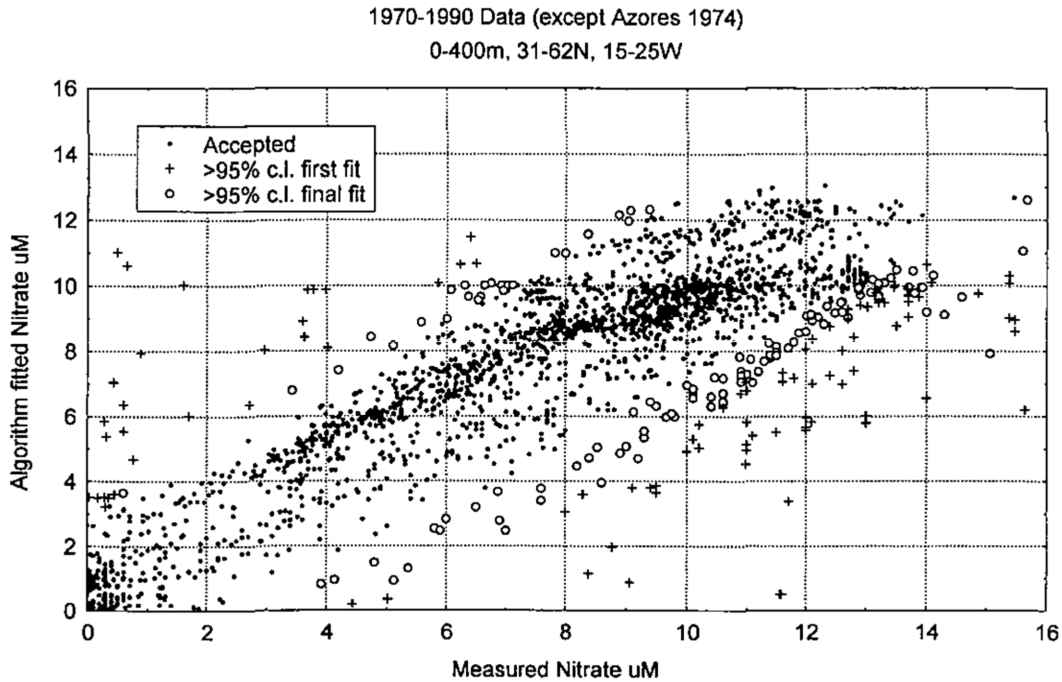
We have developed a novel quantitative approach to examining historical data which involves describing nutrient concentrations by a polynomial function in temperature and salinity. We hypothesized that any parcel of water, either newly upwelled or trapped in the euphotic zone by the developing seasonal pycnocline, would evolve new characteristics through time. It would warm (or cool) according to



the local heat budget, it would increase or decrease in salinity according to the net of precipitation and evaporation, and its nutrient concentration would decrease as a result of plant growth. Thus, nutrient concentrations as they change through time should be a unique function of T and S for this parcel. Over some spatial and temporal domains the trajectories of all of these nutrient concentrations in T/S space would be different, but may still plot up as an aggregate surface, equally describable by a single polynomial. This appears to be the case for the test data set, even though the spatial extent is 30% of the north Atlantic and all seasons for over 20 years!

The procedure for constructing these polynomials is given in a paper we published recently (Garside and Garside, 1995). Briefly it consists of fitting polynomials of increasing order, addinging one term at a time, and retaining new terms in the next fit only if they are statistically significant. The procedure quickly reveals which data fit the polynomial by plotting the difference between measured and predicted concentrations, against the ranked order of the size of the difference. These data can be tested statistically to see if they are probable members of the total data population. The standard error of the estimate from the fit can be applied as a measure to evaluate the differences between measured and calculated concentrations. Anything differing by more than some objective criterion, such as twice the standard error can be rejected and segregated for further examination.

After making a first fit to the test data using terms up to third order including cross products, only the TS^2 was not a significant term and the standard error of the prediction was 1.89 µM. 106 samples had differences between the actual and computed that were greater than the 95% confidence level and several that were beyond the 99.9% confidence level based on this S.E., and were rejected. A second polynomial was then calculated using fourth order terms , of which only T^4 was significant, and it had a standard error of the estimate of 1.47 μM. A further 125 samples were rejected, leaving 2073 samples, or about 80% of the original data.



Of the rejected data, much came from a few cruises, and within cruises the data were often systematically different from the algorithm predictions. For example, 47 of the 231 samples rejected based on the algorithm were deep samples from a cruise covering much of the area (marked "d" on graph), and they were consistently higher than expected from the algorithm prediction, suggesting that the analytical standard may have had a concentration somewhat less than supposed. For the data rejected based on their plot relative to density (277 samples), measured concentrations were on average fourfold the predicted, with a lot of scatter, and they represented almost half of the rejected data. In this case the standard must have been radically wrong, and additional factors (such as variable column efficiencies if manual methods were employed) contributed to the poor correlation of the measured and predicted values.

While traditional data QA/QC techniques provide a means of identifying historical data of doubtful provenance, predictive algorithms based on polynomials in T and S are both more sensitive and objective in their application to the task. They can also provide some insight into the problems that may exist in parts of the historical data, and ensure that a consistent data set with a well described variance remains for use.

REFERENCES

Garside C. and J.C. Garside. Euphotic-zone nutrient algorithms for the NABE and EqPac study sites. *Deep-Sea Research II* **42**(2-3): 335-347.

ACCURACY OF HISTORICAL PRIMARY PRODUCTION MEASUREMENTS

William M. Balch

ABSTRACT

I have spent several years compiling global primary productivity data. This paper focuses on a historical perspective of the ^{14}C bicarbonate technique (by far the most popular technique for measuring primary production), highlighting advantages and inherent problems. Precision of the ^{14}C technique is well quantified. Unfortunately, due to the many artifacts of sampling and algal manipulation, the accuracy of the ^{14}C technique remains poorly defined. With this in mind, I discuss what information will allow us to improve data quality, as well as the information required to "resurrect" historical data sets so that they are useable.

THE NATURE OF THE PROBLEM

One of the goals of the original Global Ocean Flux program, was to determine the rates, controlling factors, means and fluctuating components of oceanic primary productivity (U.S. Joint Global Ocean Flux Study Long Range Plan, 1990). These goals were motivated by data sets such as those of Keeling et al. (1982) that showed that atmospheric carbon dioxide concentrations were cycling annually, superimposed on a long-term increase. There was (and still is) great interest in partitioning the variance in the atmospheric CO_2 cycle between anthropogenic causes and natural causes, as well as marine versus terrestrial components. Clearly, a thorough understanding of marine primary production is needed in order to model the global variability in atmospheric CO_2 with any degree of accuracy.

This paper describes the accuracy, precision and sensitivity of primary production estimates. Much has already been written on this topic, and the reader is referred to more comprehensive reviews than are possible here (such as Peterson, 1980). In this paper, I highlight some of the factors which are important for interpreting historical primary production data, and putting confidence limits on them.

METHODS FOR MEASURING PRIMARY PRODUCTION

There are many different techniques to measure primary production [e.g. fixation of radioactive bicarbonate, the splitting of $^{18}O\text{-}H_2O$ (Bender et al., 1987), solar stimulated fluorescence (Kiefer et al., 1989), and pump-probe fluorometry (Kolber and Falkowski, 1993)] and each technique is influenced by various parts of the autotrophic and heterotrophic food webs. In order to address the programmatic goals of the International Global Ocean Flux Program, the oceanographic community needed to decide which primary production technique was most appropriate. At the outset, this would appear to be a simple question, but as one delves deeper into the intricacies of each technique, it is apparent that each technique has its own specific applications and sometimes dramatic limitations.

By far, the most popular technique for measuring marine primary production is the dawn-to-dawn ^{14}C bicarbonate technique, first proposed by Steemann Nielsen (1952). This will be the focus of the remainder of this paper. The method essentially involves addition of trace amounts of ^{14}C-bicarbonate to seawater, which is then incubated for time scales of minutes to hours (usually a maximum of a day). The sample is filtered, rinsed with filtered, unlabeled seawater, and the radioactivity of the residual particulate material is measured using a liquid scintillation counter. After several hours, a fraction of the assimilated carbon is re-respired, and some is given off as dissolved organic carbon. Thus, the technique measures net-, not gross-, carbon fixation.

DEVELOPMENT AND APPLICATION OF THE ^{14}C-HCO_3 TECHNIQUE

The technique was first published in a 25 year period when other major laboratory discoveries about the photosynthetic apparatus were being made: Hill Oxidants (Hill, 1939), mechanisms for storage of light energy as chemical energy (Gaffron, 1960), "Z-scheme" model for photosynthesis (Hill and Bendall, 1960), cyclic vs. non-cyclic models of photosynthesis (Van Niel, 1941), maximum quantum yield and the debate between Warburg, Emerson, Franck, Gaffron, and Rabinowitch on its magnitude (Franck, 1957), energy flow from phycobilins to chlorophyll (Blinks, 1954; Duysen's et al., 1961), and Emerson enhancement (Emerson, 1943), to name just a few. The ^{14}C technique is deceptively simple, and could be easily applied to the controlled, laboratory studies of the time. Problems with the technique were not fully appreciated until it was applied to field populations in which monospecific populations were rarely observed, and a mixed autotrophic and heterotrophic community was the norm. Then, details of sampling, sample manipulation, and calculation became much more important, and associated errors were sizable.

The ^{14}C technique is typically used with bottles kept in on-deck incubators ("simulated in situ") or suspended from buoys at their depth of origin ("in situ"). Both techniques have been discussed at length by others (e.g. Lohrenz et al., 1992) and the reader is referred to the earlier work for details. The major difference between simulated in situ and in situ methods is that the former does not necessarily mimic the light quality of the originally water sample, nor its temperature. Another variation of the ^{14}C technique is the "photosynthesis-irradiance" curve which examines the photosynthetic response of phytoplankton as a function of light quantity. The technique also has been described in detail elsewhere (e.g. Platt et al., 1977), and many subtle variations exist. Indeed, the technique remains a "simulated in situ" incubation, and can suffer from the same issues of light quality and temperature noted above. The useful results from photosynthesis versus irradiance (PE) curves are that they provide a kinetic response of photosynthesis as a function of photon flux, which is useful for describing autotrophic carbon fixation as a function of depth.

For seagoing measurements of primary production, there are errors associated with collection of the samples, before any isotope is added. For example, in the 1980's, it was discovered that oceanic phytoplankton were highly sensitive to trace metal contamination of the water (Fitzwater et al., 1982). Trace metal contamination came from the hydrowire, metallic messengers used to close Niskin bottles, and spring closures inside Niskin bottles. This contamination could be significantly reduced by using

Kevlar-coated hydrowire, Teflon-coated messengers, and springs (Fitzwater et al., 1982). Price (1986) also discovered that latex tubing contaminated the collected water with toxic substances and killed 95% of the phytoplankton in four days. A piece of latex tubing as small as 100 mg can kill natural phytoplankton stored in standard incubation bottles in a few days!

Sample manipulation introduced even more potential errors into the ^{14}C technique (Table 1), ranging from the type of gloves one uses for sample handling, bottle composition, bottle cleaning, light shock, filter type, pre-screening, and incubation termination. Some investigators fume their filters before counting to eliminate any ^{14}C-labeled particulate inorganic carbon (calcite). We have shown elsewhere (Balch and Kilpatrick, in press) that this can account for 10-15% of the autotrophically-fixed carbon, on the same order as new production in such areas as the equatorial Pacific. Compounding the sampling errors are errors due to final calculation of the primary production. Differences in computation involve: dark bottle correction, calibration using external or internal standards, usage of a constant versus variable total CO2 concentration, and depth of integration. Table 2 provides a summary of the calculation plus the approximate errors associated with differences in how the computation is performed. While computation errors can be as high as 50%, they typically range between 0 and 15%.

The conclusion from the above potential errors is that the absolute accuracy of the ^{14}C technique used with phytoplankton is unknown; that is, the field technique has not yet been calibrated to a global standard due to the fact that it is a bio-assay, subject to poorly quantifiable errors of sampling, manipulation, and calculation. Moreover, due to differences in sensitivities to trace metals, these errors are as much a function of the species contained in the bottle, as how the sample is treated. The challenge is to define the absolute magnitude of carbon fixation estimates. As long as there are "bottle effects", the accuracy of the sea-truth data will be questioned.

PRECISION AND SENSITIVITY OF THE ^{14}C TECHNIQUE

The precision and sensitivity of the ^{14}C technique are better known than the accuracy. For example, Steemann Nielsen (1952; his table 1) showed that the counting precision for any one sample was just 2-4% and the 95% confidence limit of a single determination was about ±11-15%. Carpenter and Lively (1980) first showed the reaction of field populations of phytoplankton to the bottle type, and how the bottles were cleaned. Their data were included in the analysis of Fitzwater et al. (1982; their table 2), and the changes in precision could be directly compared. The ^{14}C counts dramatically showed that the mean productivity of an oceanic sample increased by 1.57X by changing from glass to polycarbonate bottles when using trace-metal clean ^{14}C bicarbonate solutions. Moreover, productivity measured with clean ^{14}C bicarbonate stock solution contained in polycarbonate bottles was 329% higher than when the standard ^{14}C solutions were used in polycarbonate bottles. Note, the coefficient of variation of the samples run with "dirty" ^{14}C was much higher (40%) than when clean ^{14}C solutions were used (7%). When the method was performed on coastal phytoplankton populations, such errors were still sizable but not as large.

A more recent examination of the precision of the ^{14}C technique was published by Lochte et al. (1993; his figure 1). In this case, three ships were essentially at the same place (within a few hundred yards), at the same time, during the north Atlantic Bloom experiment of the JGOFS expedition. The average integral production was 70 mmol C m^{-2} d^{-1} with a precision of ±35%. The average surface production was 4.25 mmol C m^{-3} d^{-1} with a precision of ±46%. Lochte et al. (1993) suggested that Zn contamination may have been responsible for the large variance. Nevertheless, these results highlight the inherent confidence limits that can be reasonably applied to contemporary primary productivity measurements.

Richardson (1991) described a comparison of the ^{14}C primary production technique made by 24 different laboratories from several countries. She concluded that even when the same filters were distributed to different laboratories, estimates of primary production (units of mg C m^{-2} d^{-1}) had a coefficient of variation of 10%. Laboratories that used different techniques to estimate the primary production (using common samples) showed coefficients of variation of 15%. Comparisons were also performed in which common field samples were analyzed using the same techniques. Interestingly, for three different experiments (with 5, 7, and 11 labs participating, respectively), the resulting mean integral production estimates had coefficients of variation of 25, 38, and 40%. For replicate determinations by any one laboratory, the coefficient of variation of production estimates was <10% with one exception, not much different than the precision of the technique as first described by Steemann Nielsen (1952). Errors in photo-adaptive parameters (derived from PE curves) were also substantial. Richardson (1991) found that when investigators used their own methods on a pooled water sample, their derived Pmax values (mg C m^{-3} h^{-1}) had a coefficient of variation of about ±20%. If each investigator collected their own water, and applied their own methods, the coefficient of variation was ±66-68%. Determinations of the slope of the light limited portion of the PE curve (a), had coefficients of variation of ±48-65%. These errors obviously propagate when calculating integral water column primary production from PE curve results.

One of the stated goals of the Richardson (1991) comparison was to determine the feasibility of a central primary production data bank; she concluded that the large differences in primary production (when made by different laboratories on the same water samples) made such a proposition unreasonable. Richardson's results, taken together with the results of Lochte et al. (1993), suggest that the coefficient of variation for field primary productivity measurements (±35% for integral production estimates) was typical for the general ^{14}C-primary productivity technique, when applied by different investigators on the same water sample. These results present global-change scientists with a dilemma. If total global marine primary production levels are estimated to be ~20-50 Gt y^{-1},(Steemann Nielsen, 1954; Fleming, 1957; Koblents-Mishke et al., 1970; Ryther, 1969; Leith and Whittaker, 1975; Platt and Subba Rao, 1975; Berger et al., 1987; and Longhurst et al., 1995), and there are about 2.5-3 Gt of "missing" anthropogenic carbon (or ~5-15% annually), then it will be difficult to find this "missing carbon" if the precision of our "sea-truth" integral ^{14}C-production estimates is ±35% (or ±48-68% as with photosynthesis-irradiance-based results).

INFORMATION NEEDED TO USE HISTORICAL DATA OR "WHAT IS NEEDED TO RESCUE OLD ^{14}C DATA?"

The rescue of historical primary production data is not trivial and will remain so, as long as samples have to be contained in bottles for significant amounts of time. It is currently impossible to exactly calculate the errors associated with sample handling, especially if there is no absolute global standard. A better approach is to increase the meta-data collected, so that the user can best judge the quality of the data. Table 3 outlines the types of metadata that should be included in any primary production data set. It will be the rare data set that contains all of the types of metadata, but consider the consequences of not having such data. In a method where incorrect ^{14}C stock preparation can cause a >300% change in primary production, knowledge of the stock preparation protocol can at least provide a first guess of data quality.

REFERENCES

Balch, W., M. and K. A. Kilpatrick. 1996. Calcification rates in the equatorial Pacific along 140 W. In press. Deep Sea Research

Bender, M., K. Grande, K. Johnson, J. Marra, P. J. LeB. Williams, J. Sieburth, M. Pilson, C. Langdon, G. Hitchcock, J. Orchardo, C. Hunt, P. Donaghay, and K. Heinemann. Limnol. Oceanogr. 32: 1085-1098.

Berger, W. H. 1989. Global Maps of Ocean Productivity. In W. H. Berger, V. S. Smetacek and G. Wefer (eds.). Productivity of the Ocean: Present and Past. John Wiley and Sons Limited. pp. 429-455.

Blinks, L. R. 1954. The photosynthetic function ofpigments other than chlorophyll. Annu. Rev. Plant Physiol. 5: 93-114

Carpenter, E. J. and J. S. Lively. 1980. Review of estimates of algal growth using ^{14}C tracer techniques. in Primary Productivity in the Sea. P. Falkowski (ed.) Plenum, NY.

Clayton, R. K. 1980. Photosynthesis: physical mechanisms and chemical patterns. IUPAB Biophysics Series. The International Union of Pure and Applied Biophysics. eds. F. Hutchinson, W. Fuller, and L. Mullins. Cambridge Univ. Press. 281 pp.

Duysens, L. N. M., Amesz, J., and B. M. Kamp. 1961. Two photocyemical systems in phytosynthesis. Nature (London) 190: 510-511.

Emerson, R. and C. M. Lewis. 1943. The dependence of the quantum yield of Chlorella photosynthesis on wavelength of light. Am. J. Bot. 30: 165-78.

Fitzwater, S. E., G. A. Knauer, and J. H. Martin. 1982. Metal contamination and its effect on primary production measurements. Limnol. Oceanogr. 27: 544-551.

Fleming, R. H. 1957. General features of the ocean. In: J. W. Hedgpeth (ed.) Treatise on Marine Ecology and Paleoecology. Geological Society of America Memoir 67: 87-107.

Franck, J. 1957. A theory of the photochemical part of photosynthesis. In Research in Photosynthesis, H. Gaffron et al., eds., pp. 142-6. Interscience, New York.

Gaffron, H. 1960. Energy storage: Photosynthesis. In Plant Physiology, F.C. Steward, ed. Vol. 1B, pp. 3-277. Academic Press, New York.

Hill, R. 1939. Oxygen production by isolated chlorophlsts. Proc. Roy. Soc. (London) B127: 192-210.

Hill, R. and Bendall, F. 1960. Function of two cytochrome componenst in chloroplasts: A working hypothesis. Nature (London) 186: 136-137.

Keeling, C., R. B. Bacastow, and T. P. Whorf. 1982. Measurements on the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii. In: Carbon Dioxide Review 1982, W. C. Clark (ed.), pp. 377-384.

Kiefer, D., W. S. Chamberlin, and C. R. Booth. 1989. Natural fluorescence of chlorophyll a: Relationship to photosynthesis and chlorophyll concentration in the western South Pacific gyre. Limnol. Oceanogr. 34: 868-881.

Koblentz-Mishke, O. I., V. V. Volkovinsky, and J. G. Kabanova. 1970. Plankton primary production of the world ocean. In: W. Wooster (ed.) Scientific Exploration of the South Pacific National Academy of Sciences, Washington, D. C. p. 183-193.

Kolber, Z. and P. G. Falkowski. 1993. Using active fluorescence to derive phytoplankton photosynthesis in situ. Limnol. Oceanogr. 38: 72-79.

Lieth, H., and R. H. Whittaker. 1975. Primary Productivity of the Biosphere. Springer-Verlag, New York, 339pp.

Lochte, K., H. W. Ducklow, M. J. R. Fasham and C. Stienen. 1993. Plankton succession and carbon cycling at 47 N 20 W during the JGOFS North Atlantic Bloom Experiment. Deep-Sea Research II. 40 (0.5): 91-114.

Lohrenz, S. E., D. A. Wiesenburg, C. R. Rein, R. A. Arnone, C. T. Taylor, G. A. Knauer, and A. H. Knap. 1992. A comparison of in situ and simulated in situ methods for estimating oceanic primary production. J. Plankton. Res. 14: 201-221.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995 An estimate of global primary production in the ocean from satellite radiometer data. J. Plank. Res. 17: 1245-1271.

Peterson, B. J. 1980. Aquatic primary productivity and the $^{14}CO_2$ method: a history of the productivity problem. Ann. Rev. Ecol. Syst. 11: 369-385.

Platt, T. and D. V. Subba Rao. 1975. Primary production of marine microphytes. Photosynthesis and productivity in different environments. In: International Biological Programme. Cambridge Univ. Press, Cambridge, England. 3: 249-279.

Platt, T., K. L. Denman, and A. D. Jassby. 1977. Modelling the productivity of phytoplankton, p. 807-856. In E. D. Goldberg [ed.], The sea: ideas and observations on progress in the study of the seas. John Wiley, New York, NY.

Price, N. M., P. J. Harrison, M. R. Landry, F. Azam, and K. J. F. Hall. 1986. Toxic effects of latex and Tygon tubing on marine phytoplankton, zooplankton, and bacteria. Mar. Ecol. Progr. Ser. 34: 41-49.

Richardson, K. 1991. Comparison of ^{14}C primary production determinations made by different laboratories. Mar. Ecol. Prog. Ser. 72: 189-201.

Ryther, J. H. 1969. Photosynthesis and fish production in the sea. Science. 166: 72-76.

Steemann Nielsen, E. 1952. The use of radioactive carbon (^{14}C) for measuring organic production in the sea, J. Conseil du Res. vol 18:117-140.

Steemann Nielsen, E. 1954. On organic production in the oceans. Conseil Perm. Intern. Explor. Mer., J. du Conseil, 19: 309-328.

U. S. Joint Global Ocean Flux Study- Long Range Plan. 1990. U. S. JGOFS Planning Office, Woods Hole Oceanographic Institution, Woods Hole, MA. 191 pp.

Van Niel, C. B. 1941. The bacterial photosyntheses and their importance for the general problem of photosynthesis. Adv. Enzymol. 1: 263-328.

TABLE 1: Sources of Error In ^{14}C Primary Productivity Measurements

Sources of error in ^{14}C primary productivity measurements due to water sampling:

-Use of metallic hydrowire
-Niskin bottles vs. GO-FLO's (black rubber O-rings, uncoated springs,
-Coated messengers

Sources of error in ^{14}C primary productivity measurements due to sample manipulation:

-Gloves (powdered, latex vs vinyl, etc)
-Bottle size
-Bottle composition (Borosilicate, Polycarb, Composition of caps, etc.)
-Optimal sample time
-Light shock
-Flow rate out of sampler
-Inadvertent silicate contamination with isotope addition
-Inadvertent trace metal contamination w/ isotope addition
-Inadvertent N contamination by ultra cleaning with nitric acid
-Filter type
-Pressure differential used in filtering
-Fuming filters to drive off bicarbonate
-Acidifying SW to drive off bicarbonate vs filtration
-Pre-screening to remove grazers
-Stopping incubations with preservatives

Sources of error in ^{14}C primary productivity measurements due to computational factors:

-Dark bottle correction
-Calibration (external vs internal standards)
-Constant vs variable CO2 with salinity
-Integration depth (0.1 vs 1%)

TABLE 2: Changes In Integrated Production Associated With Computation

$P = (DPM_s - DPM_{dk}) * CO_2 * Frac/(DPM_{tot} * Time)$

Computation	Int. Prod
Subtract dark bottle	-3.7%(1.5-8.0%; Banse '93)
	-50%(Morris et al., 1971)
$^{12}C/^{14}C$ frac.	+5 to 6%
Variable CO2	+5%[a]
Fuming(calcite loss)	0 to -15%[b]
0.1% vs 1% int.	+1 to 3%[c]

[a] This is based on variance in salinities over the water column in a coastal regime.
[b] Based on typical calcification rates in the euphotic zone.
[c] Approximate based on biomass and light levels found between these depths.

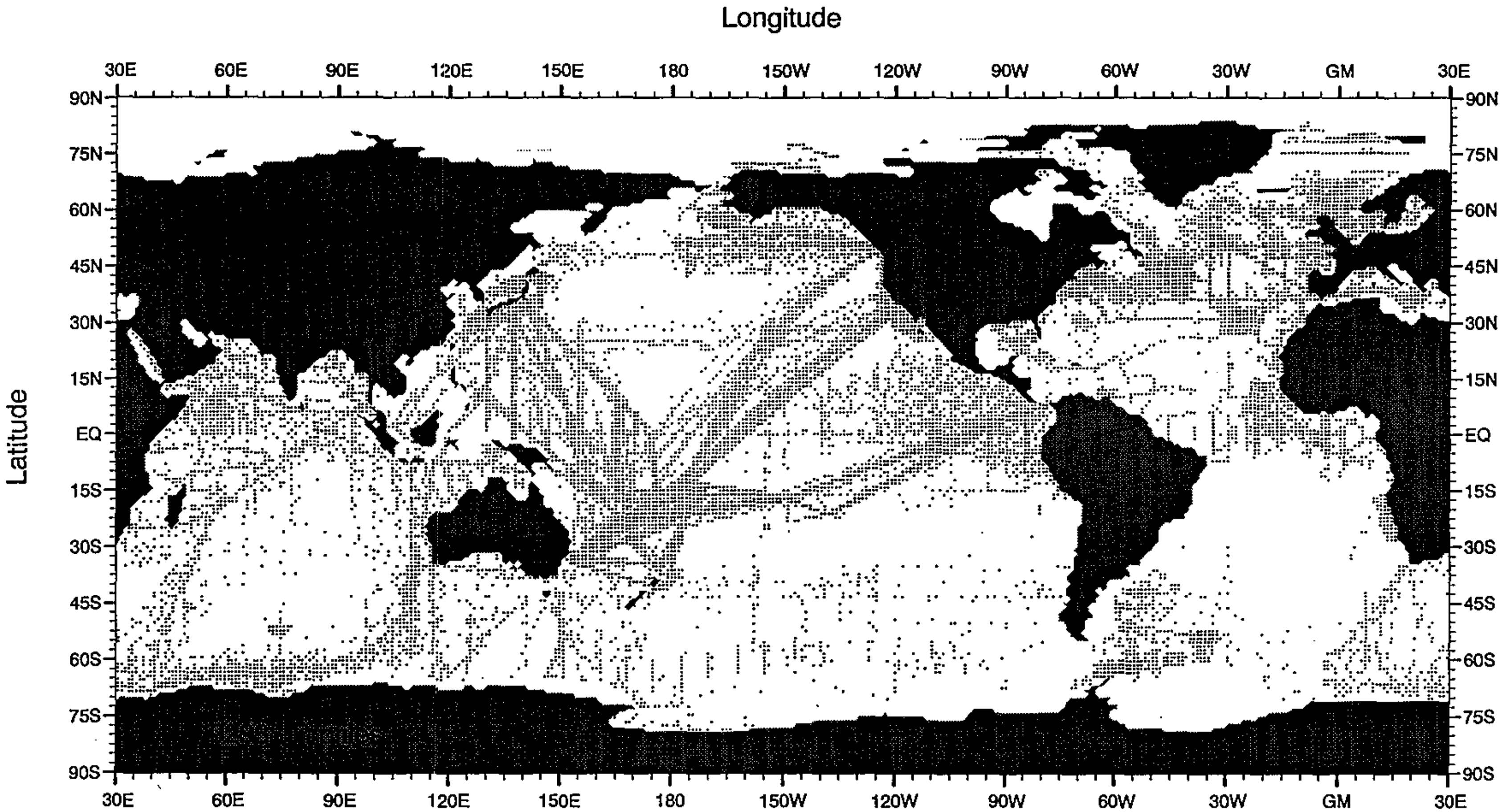


Fig 1. Distribution of Chlorophyll Data at the OCL

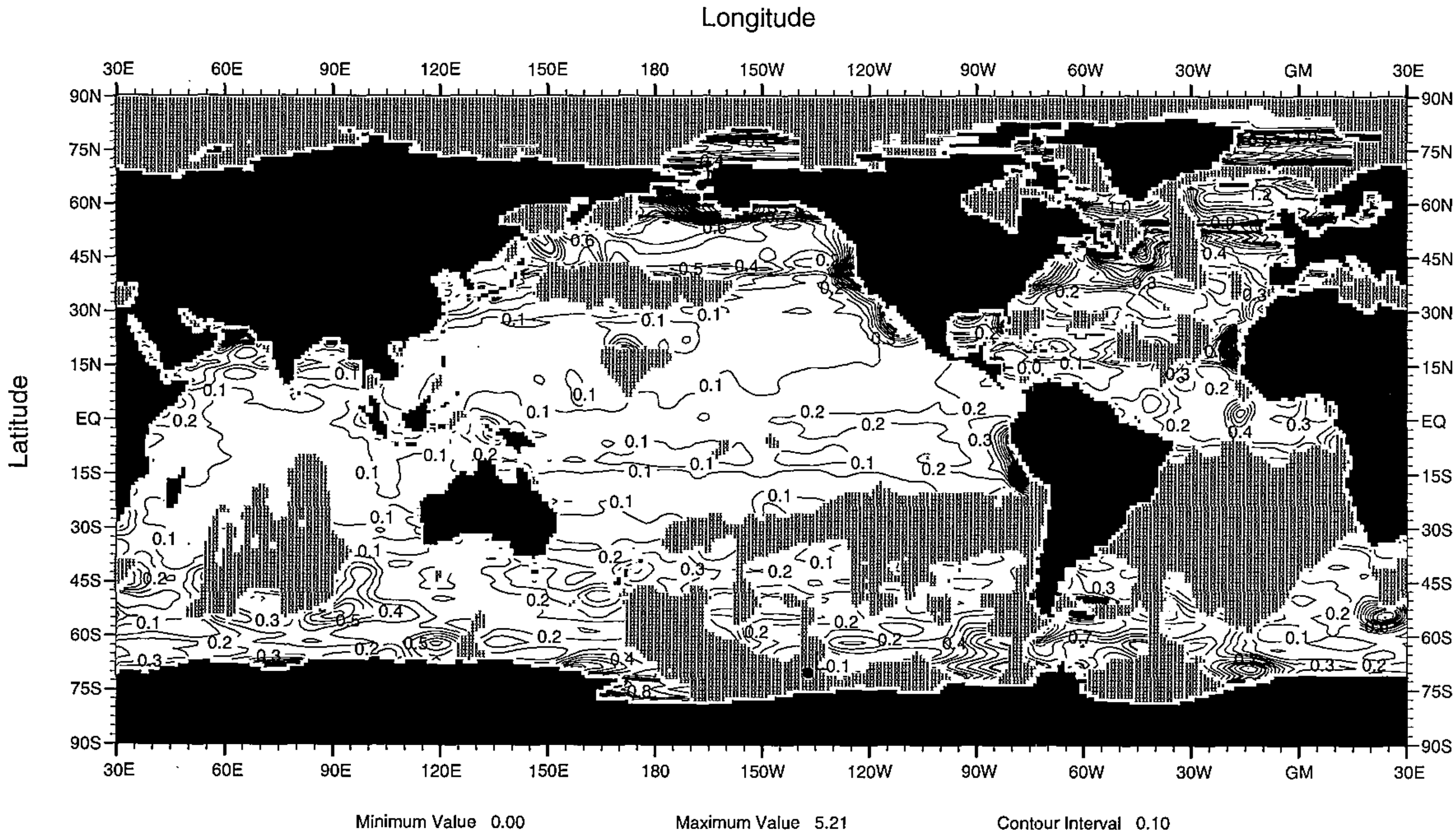


Fig 2. Annual mean surface chlorophyll

BIOLOGICAL DATA FROM THE SOUTH PACIFIC: SYNTHESIS OF THE WOCE AND HISTORICAL DATA

V. Gouretski
Max-Planck-Institute for Meteorology, Hamburg
K. Jancke
German Hydrographic Office, Hamburg

ABSTRACT

The WOCE Hydrographic Programme Special Analysis Centre is to prepare consistent and coherent hydrographic data sets from the WOCE hydrographic data. To provide estimates of how much the long-term changes in the hydrographic structure of the World Ocean are, and how much subsequently its circulation might change, consistent validated data sets produced from historical data can contribute as a sound and solid reference to compare those data WOCE is gathering in the 1990's.

A basin scale data set was compiled for the South Pacific Ocean, comprising totally about 4900 hydrographic profiles. A quality control procedure was developed to arrive at a validated data set. Six commonly measured oceanographic parameters are treated: temperature, salinity, oxygen, silicate, nitrate and phosphate. A description of the validation method and the analysis of its performance is given. A cruise by cruise comparison with the modem high-quality (WOCE) hydrographic data shows a rather poor quality of the historical data, especially that of the nutrient data (NO_3 and PO_4). A percentage of the data rejected by the validation procedure ranges from 15% for the sea-water density to 27% for Nitrate. The inter-cruise offsets of parameters are found by using the deep water temperature-parameter relationships as natural reference standards.

Gridded fields of the oceanographic parameters are produced by means of the optimum interpolation method applied to the data on the isopycnal surfaces. In general, isopycnal averaging results in distributions with sharper property gradients compared with the isobaric averaging. The differences are at their maximum in the vicinity of the Antarctic Circumpolar Current and the Antarctic continental slope.

THE ACCURACY OF HISTORICAL MEASUREMENTS OF PLANKTON DATA

Philip C. Reid

In this overview the term 'historic' has been given a flexible connotation; it has been applied differently to the oceanic regions of the world to reflect different rates of progress in plankton studies and is discussed for time series and plankton surveys. An attempt to extract information on the location and date type of time series observations has been made using the ASFA abstract service, using as a base 36,000 of publications containing the word 'plankton'. Approximately 50 time series have been identified, of which half are located in the North Atlantic region; they are virtually non-existent in the southern hemisphere. A time series is taken as any survey lasting at least 4 years with minimum observations at approximately monthly intervals. As standard methods are applied over a number of years in these studies the results are internally consistent and accurate, but in many cases the methods used are not standard, non-quantitative and difficult to compare with other surveys. Specific reference will be made to the CPR Survey, CALCOFI, the Helgoland dataset and the Dandonneau chlorophyll survey in the Pacific.

A historical account of the development of plankton surveys with time including observations on accuracy is outlined. Problems of intercalibration/splicing of historical surveys with modern surveys are raised. The importance of archived samples and the effects of long term storage are discussed in relation to accuracy. Few agencies/laboratories have produced catalogues of plankton samples held in their collections, and curation is generally given a very low priority. Conversion of species identifications to biomass may introduce errors into datasets. Much existing data is in the form of wet weight and thus inaccurate. Improved dry weight measurements are urgently needed. Examples of the large variability that may found in time and space for the dry weight of just one copepod species will be shown.

Manuals on techniques for zoo/phytoplankton sampling, fixation, preservation and analysis were publishing by UNESCO between 1968 and 1978. Updates of the publications, especially for zooplankton, are long overdue. Problems of accuracy in relation to methodology are well covered in these manuals so only brief comments are raised on this subject.

Problems of species identification and enumeration in plankton surveys have received little attention in the past. The vast majority of taxonomists are self taught and few surveys include quality control procedures. There are a number of coded species lists, but no international list has yet been agreed. ICES and a few other organizations produce identification sheets, but most texts used for identification are more than 20 years old. For many regions of the world the Plankton fauna has been poorly covered. A recent publication (1993) in Marine Biology on tropical plankton used as a basic reference Wickstead (1965). This document only contains 'coarse' line drawings of taxa and its use for identification places some doubt on the accuracy of the species identifications in the published study. It is a well known fact that many plankton analysts have a 'tunnel vision' when identifying certain taxa. They do not see these taxa in the samples. Groups such as soft/jelly taxa are frequently ignored and some taxonomists only use the specimens in samples for which they are experts, discarding other material. Attitudes/fashions may also influence results. For example,, there is a general belief that an 'epidemic' of algal blooms has occurred over the last two decades. It is still not clear if this is a real phenomenon or a consequence of an increasing global interest in blooms and their effects.

The presentation ends with a series of comments on changing fashions in marine science, how these are influenced by the publication policy of journals and the extent to which they may affect the accuracy and availability of plankton date. Much historical plankton sampling was undertaken in the absence of information/knowledge on hydrography. The advent of satellites and accurate hydrodynamic models and their identification of fronts and other features has revolutionized the 'accuracy' of sampling strategies. Much current plankton production theory is not backed up by appropriate field studies and modeling studies on the plankton are initiated on poor data with no data for validation or testing. The compilation of databases of zooplankton dry weight biomass for the 50 major marine 'biomes', as per Longhurst, by use of (interpolated) historic data that has undergone a check for accuracy is a priority for future research.

CHEMICAL AND BIOLOGICAL OCEANOGRAPHIC DATA AND INFORMATION MANAGEMENT AT THE INDIAN NODC

J. S. Sarupria and Aravind Ghosh K.

ABSTRACT

The Indian National Oceanographic Data Centre (INODC) is a national and international repository of oceanographic data and information for Indian ocean region. INODC has developed a system for the purpose of management of Chemical and Biological data and information. The system is in modular form and consists of three modules namely inventory, documentation and processed numerical data sets. The system is designed on the concept of open architecture so that new data sets that flow in future to INODC can be appended without disturbing the existing structure. The Chemical and Biological System (CBS) is linked with the environmental data sets for further analysis and to understand the biological cycle in the seas/oceans. The system provides geographical, seasonal and parameter-wise inventory information and retrieves data using multi-parameter select criteria. Hence CBS is useful for computer based storage, retrieval, selection and presentation of the data and information as per user's requirement.

METADATA REQUIREMENTS FOR PLANKTON DATA

Linda Stathoplos

ABSTRACT

At the U.S. National Oceanographic Data Center, plankton data are being added to the Ocean Climate Laboratory's Global Ocean Database to characterize and quantify plankton biomass, composition, and variability on a global scale. Measures of plankton biomass and taxonomic composition are stored with associated sampling and processing metadata (information about the data). Where metadata allows duplicate stations to be identified, plankton data are combined with physical and chemical data in the Global Ocean Database. Inclusion of metadata also allows for selective examination of data and appropriate quality control. Sufficient plankton data have been digitized and converted to a uniform format to begin development and application of quality control procedures to label the plankton data in the Global Ocean Database with specific quality flags.

INTRODUCTION

Plankton constitute a large proportion of the upper ocean biomass, storing and transforming large amounts of organic carbon and nutrients (Raymont, 1983; Barnes and Hughes, 1988; Omori and Ikeda, 1984). The transformation processes, their rates, and the size of the carbon pools they generate all depend on both the plankton taxonomic composition and biomass. Compilations of plankton taxa counts and biomass estimates are useful tools for describing large scale seasonal and interannual patterns of variability (Longhurst, 1985; Sherman, et al., 1983; 1990; McGowan, 1990). Difficulties with combining plankton data include: (1) many data are in non-digital form; (2) often, sampling documentation is missing or incomplete; (3) there are no standard formats, units, or sampling protocols; and (4) there are a variety of problems with taxonomic data (Stathoplos, this volume).

PLANKTON PARAMETERS IN THE GLOBAL OCEAN DATABASE

A uniform format has been developed to store observations of plankton biomass and taxonomic groups, along with associated sampling and processing information, to the Ocean Climate Laboratory's Global Ocean Database. Although plankton data are spatially and temporally limited compared with physical and chemical data, tens of thousands of stations with plankton observations have been merged (Fig. 1). To date, over 4,000 stations with plankton data have been digitized from holdings at the U.S. NODC and collocated World Data Center-A, Oceanography (WDC-A). At least 65,700 plankton observations held by WDC-A from 30 countries remain to be digitized. Additional data continue to be acquired, converted to digital form if necessary, and reformatted to expand the Global Ocean Database. The short-term goal is to include available plankton data with preliminary quality-control flags as part of the next release of the Global Ocean Database. The long term goal is to complete the quality-control labeling of the plankton data, and produce objectively analyzed plankton biomass maps, with accompanying taxonomic composition descriptions, on a global scale. Eventually, atlases will be produced and the combined database be made available on CD-ROM, comparable to those previously made available for physical and nutrient parameters in the World Ocean Atlas 1994 (Conkright, et al., 1994; Levitus and Boyer, 1994a, 1994b; Levitus, et al., 1994).

The plankton parameters of interest are biomass measurements as volumes, weights, or carbon, and for specific groups of plankton, concentrations or presence/absence. These observations can be per volume, per area, or per sample. Maintenance of metadata, the associated information about sample collection, identification, and measurement techniques, is required for appropriate application of quality control checks. Taxon descriptors, gear type, tow type, and methods to preserve, count or weigh samples are stored using code lists. Upper and lower sampling depths, sampling duration, sampler mesh size and mouth area are stored using numerical values.

QUALITY CONTROL

Quality control procedures applied to the combined data are of two types: the first are *a priori* checks, to detect grossly erroneous or 'impossible' values. The second are data-derived ranges that are set to identify atypical or 'outlier' values. The *a priori* checks include checks for impossible times, dates, locations, ship speeds, and so on. Next, duplicate records, which can result from duplicate data submissions, are identified and removed. Whenever there are sufficient metadata available to identify matching stations, plankton data are combined with associated physical and chemical data.

Checks specific to plankton observations are now being developed and applied to the Global Ocean Database (Table 1). While some historical data may be of low quality, the best way to develop criteria for recognizing and eliminating these data is to combine and evaluate them by documented quality control procedures. Since past ocean conditions can no longer be sampled, there is no alternative to use of historical data to increase temporal coverage in composite databases on decadal time scales.

The first four flags listed in Table 1 are of the *a priori* type, and do not depend on the properties of the combined data themselves. Depth errors are identified by flag 1. Published biogeographic data (e.g., Bé, 1977) will be used to apply flag 3. Reasonable organism sizes and volumes are used both to identify sampler incompatibilities to apply flag 2, and to set theoretical upper limits for group counts or volumes to apply flag 4.

To apply the data-based flag 5, simple approaches taking full advantage of exploratory data analysis are used to determine the data-based ranges (Liepins and Uppuluri, 1990; Tukey, 1977; Curran and Small, 1983). The ordered values are plotted, then transformed if necessary to improve their distribution. Biomass and count data distributions are typically log-normal, although power transformations have also been suggested (Downing, et al., 1987). Appropriate location and dispersion indices are chosen for the transformed data, which can then be used to identify outliers.

Exploratory data analysis is also used to identify data patterns and biases. Factors which can bias plankton samples include, for example, time of day (especially for vertically migrating species), net type, net mesh size, and type of tow. These factors have been discussed in detail in monographs on plankton methods (e.g. UNESCO, 1968; Omori and Ikeda, 1984). Algorithms have been constructed to model gear sampling bias (e.g. Barnes and Tranter, 1965). Some investigations have found no significant difference in biomass estimates between certain types of samplers (Roman, et al., 1986). Conversion factors between the different types of biomass measures have also been derived (Wiebe and others, 1975; 1988). The combined data can be examined for confirmation of these published conversion factors, and for the presence of other identifiable biases. The identification of atypical values and systematic bias is expected to be an iterative process. If sufficient data exist, appropriate ranges can be set regionally as well as globally.

CONCLUSION

For quality control of plankton data to give robust results, the inclusion of metadata is crucial. At a minimum, you must know where, when, and something about how the sample was obtained. If a particular group was observed, that group must also be adequately described. Application of *a priori* and data-derived quality control flags to the combined plankton data will enhance the value of these data to the research community.

ACKNOWLEDGMENT

Support for this project was provided to L. Stathoplos by NOAA's Office of Climate and Global Change (GC95-220). Carla Forgy digitized many plankton data, and Todd O'Brien converted many plankton data to the Global Ocean Database format; their assistance is gratefully acknowledged.

REFERENCES

Barnes, R. S. K. and R. N. Hughes. An Introduction to Marine Ecology. 2nd ed., Oxford: Blackwell Scientific Publications, 1988.

Barnes, H. and D. J. Tranter. A statistical examination of the catches, numbers and biomass taken by three commonly used plankton nets. Aust. J. Mar. Freshw. Res. 16 (1965): 293-306.

Bé, A.W.H. "An ecological, zoogeographic and taxonomic review of Recent planktonic foraminifera." In A.T.S. Ramsay, ed., Oceanic Micropaleontology, Vol. 1, pp. 1 - 100. London: Academic Press, 1977.

Conkright, Margarita E., S. Levitus and T. P. Boyer. "World Ocean Atlas 1994. Volume 1: Nutrients." NOAA Atlas NESDIS 1 (1994): 1 - 150.

Curran, III, Thomas A. and Robert D. Small. "On Using Exploratory Data Analysis as an Aid in Modeling and Statistical Forecasting." In T. Wright, ed., Statistical Methods and the Improvement of Data Quality, pp. 333 - 353. New York: Academic Press, 1983.

Downing, J.A., M. Pérusse, and Y. Frenette. "Effect of interreplicate variance on zooplankton sampling design and data analysis." Limnol. Oceanogr. 32 (3) (1987): 673-680.

Levitus, Sydney, R. Burgett and T. P. Boyer. "World Ocean Atlas 1994. Volume 3: Salinity." NOAA Atlas NESDIS 3 (1994): 1 - 99.

Levitus, Sydney and T. P. Boyer. "World Ocean Atlas 1994. Volume 2: Oxygen." NOAA Atlas NESDIS 2 (1994a): 1 - 186.

Levitus, Sydney and T. P. Boyer. "World Ocean Atlas 1994. Volume 4: Temperature." NOAA Atlas NESDIS 4 (1994b): 1-117.

Liepins, Gunar and V. R. R. Uppuluri. Data Quality Control: Theory and Pragmatics. New York: Marcel Dekker, Inc., 1990.

Longhurst, Alan R. "The Structure and Evolution of Plankton Communities." Progress in Oceanogr. 15(1985): 1-35.

McGowan, John. "Climate and Change in Oceanic Ecosystems: The Value of Time Series Data." Trends in Ecol. Evol. 5 (1990): 293-299.

Omori, Makoto and Tsutomu Ikeda. Methods in Marine Zooplankton Ecology. 332 p. New York: John Wiley & Sons, 1984.

Raymont, John E. G. Plankton and Productivity in the Oceans, Vol. 2 - Zooplankton. 2nd ed. 824 p. Oxford: Pergamon Press, 1983.

Roman, Michael R., C. S. Yentsch, A. L. Gauzens and D. A. Phinney. "Grazer Control of the Fine-Scale Distribution of Phytoplankton in Warm-Core Gulf Stream Rings." J. Marine Res. 44(1986): 795-813.

Sherman, Kenneth, J. R. Green, J. R. Goulet, and L. Ejsymont. "Coherence in Zooplankton of a Large Northwest Atlantic Ecosystem." Fisheries Bull. 81(1983):855-862.

Sherman, Kenneth, Lewis M. Alexander, and Barry D. Gold, eds. Large Marine Ecosystems: Patterns, Processes, and Yields. 242 p. Washington, DC: AAAS, 1990.

Tukey, J. W. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977.

UNESCO. Zooplankton sampling. Paris: The Unesco Press, 1968.

Wiebe, Peter H., Steven Boyd, and James L. Cox. "Relationships Between Zooplankton Displacement Volume, Wet Weight, Dry Weight, and Carbon." Fishery Bulletin 73 (1975): 777-786.

Wiebe, Peter H. "Functional Regression Equations for Zooplankton Displacement Volume, Wet Weight, Dry Weight, and Carbon: A Correction." Fishery Bulletin 86 (1988): 833-835.

Table 1: Proposed Quality Control Flags For Plankton Data

Proposed flag	Example
1) Depth problem	Missing upper or lower depth; lower <upper depth; depth >5000m
2) Taxon plus life stage incompatible with sampler	Portuguese man-of-war caught in Continuous Plankton Recorder sample
3) Taxon was observed outside normal biogeographical range	*Orbulina universa* (planktonic foraminifera) in Arctic Ocean
4) Parameter outside theoretical range, such as concentration exceeds number that can physically fit in volume	1000/ml for Ostracoda
5) Metadata-adjusted parameter outside global/regional range	(see text)

INFORMATION NECESSARY FOR ZOOPLANKTON BIOMASS STUDIES AND ITS POSSIBLE IMPLICATIONS

Jury A. Rudjakov

ABSTRACT

The data on zooplankton biomass are important for understanding and output assessment of many biological and geochemical processes in the ocean, such as biological productivity development, vertical and horizontal organic matter transfer, CO_2 budget maintenance, and potential bioluminosity of ocean water subsistence. To be of a value, biomass data have to be accompanied by the set of additional information which can be used in data quality control and make the data applicable in many kinds of broad ecological research. Besides of metadata which have been usually (but not always) recognized as important (geographical coordinates, date and time of sampling; type of the net used and its mesh size; layer sampled and volume of water filtered) the details of laboratory treatment of samples are very important as well. Of the latter, following are worthy of special attention: the porosity of the filters which were used in initial sample condensation and prior to biomass measurement; upper size limit of animals left in sample prior to biomass determination; whether gelatinous animals were present and, if they were, whether they and their parts were picked up before biomass determination; fixative type; whether biomass measurements were made after sample fixation, and if they were, how long was the period between fixation and measurement. Indications of phytoplankton bloom are very important and an example is given how these indications were implemented in the quality control of zooplankton biomass data.

INTRODUCTION

The understanding of biological and geochemical processes in the ocean demands for the knowledge of the numeric individual and integral characteristics of various groups of inhabitants of the upper layer of the ocean. The biomass of mesozooplankton, which comprises the animals of the size range from 0.2 to 20 mm, is one of the most critical. Only few of the possible applications of this parameter in marine ecology have been a matter of common knowledge. Namely, after the work by Vinogradov (1968) it became widely known that the plankton biomass of the upper layer of the ocean determines the feeding conditions of deeper living animals. The latter was demonstrated by formulated reasoning and calculations by Tseitlin (1986), who had shown also that mesoplankton biomass of the upper layer may be used to estimate biomass and production of organisms from other size groups, up to mesopelagic fishes and down to phytoplankton. The biomass content of the ocean water may be related to potential luminosity in different regions and depths (Rudjakov, 1968; Rudjakov, Voronina, 1967; Rudjakov, Tseitlin, 1989). Plankton biomass determines the feeding conditions of planktophagous animals, including not only the commercially important pelagic species but also benthic and benthopelagic ones, such as fishes and invertebrates inhabiting underwater mountains (Parin et al., 1985; Heinrich et al., 1993). Moreover, distribution of mesozooplankton biomass effects the vertical and horizontal fluxes of organic matter in the ocean (Vinogradov and Tseitlin, 1983; Rudjakov, 1986; Tseitlin, 1986; Rudjakov, 1987) and in this way influences the level of CO_2 concentration in the atmosphere. Finally, horizontal and vertical biomass distribution patterns form the basis for the estimates of the stability of pelagic ecosystems in relation to external impacts, since the stability can be expressed as characteristic or relaxation time of a distributed system (Zaikin, Rudjakov, 1989). These lines of enquiry roughly identify but not limit the possible users of plankton biomass data. The information above definitely says about the necessity to spend time and resources for the row biomass data, which are mostly in manuscript format, to be computerized and firmly saved: these efforts are clearly worthy of trying. It is evident, though, that to be of a value, the data must be reliable. For this reason two problems arise: how to perform the quality control of historical data sets and how to organize the collection of new data avoiding the mistakes which have been made by predecessors.

In this article I am going to sketch the set of facts which are referred here as metadata and which make the results of biomass determinations consistent and really applicable in the context of the large-scale ecological and oceanographical problems listed above. Some of them if only rarely were documented in historical data sets and, unfortunately, have not received much attention in modern bio-oceanographical research. Only the conventional methods of zooplankton biomass evaluation will be addressed here since others, such as *in situ* optical and acoustical biomass estimation, are not widely accepted and, so far, are mostly in a state of testing only; zooplankton sampling with pumping devices needs practically the same set of accompanying data as the sampling with nets and water bottles. The aim of the paper is two-fold: to outline the problems of the historical data management and thereby to attract the attention of planktonologists who can improve the quality of their data and to broaden the area of the data implementation.

METHODS OF ZOOPLANKTON BIOMASS DETERMINATION AND POSSIBLE DEFECTS OF DATA DOCUMENTATION

Results of biomass measurements have to be assigned to a unit of water volume or area bases. To perform these calculations the standard set of additional measurement has do be done. It incorporates either the direct measurements of filtered water by means of calibrated flow meter or the determination of the length of net path in water; the latter accompanied with the mouth area of the net allows to estimate the volume filtered. This information is nearly always available and sufficient to calculate biomass values. Documenting date and time of sampling, as well as the depths of layers sampled (or wire length and angle of its inclination) has been commonly recognized as divine duty of a planktonologist. These data are practically always accompany sample or station records (though, one cannot say the same about many published data, unfortunately). The data set encompassing station site, volume filtered, layer sampled, net type, mesh size, date, and time of sampling is required always. The information items are listed here in the order of their importance and if the lack of information about time of sampling only limits the sphere of data usage, the absence of knowledge about depth of sampling or volume filtered would nullify the data importance completely. In general, this data set is sufficient to bring the data gathered with various methods into better compatibility by means of conversion factors, exploiting the up-to-date concepts of plankton biomass vertical distribution, seasonal oscillations, and plankton organisms size distribution (Rudjakov et al., 1995; Rudjakov, Tseitlin, 1996; Rudjakov et al., 1996).

All conventional methods of biomass estimation have their origin in sampling with plankton nets or large capacity water bottles. Both include filtration as imperative: plankton net is a filter by itself, and, moreover, net and water bottle samples have to be condensed prior to fixation and storage. The filtration falls into the category of the most formidable procedures of quantitative chemistry and biology also, since the parameters of a filter are not stable and change in the course of filtration. For zooplanktonologists this problem is usually known as net clogging. This decline in net efficiency can be effectively overcome with the implementation of calibrated flow meters. But planktonologists pay less attention to the sample condensation always applied prior to sample fixation and storage and the porosity of filters used in this procedure if only rarely has been documented. This leads to uncertainty in determination of the lower limit of size spectrum of the sampled organisms, since in situations like this a reservation arises whether the pore size of the filter was smaller than the mesh size of the net. The situation becomes even worse if there is no information about upper size limit of the animals. As a result, the derived biomass value becomes meaningless since to be of a value, a biomass quantity has to be attributed to a certain size span of organisms. This reasoning is based on the character of size distribution of pelagic organisms which was firstly formulated by Sheldon with co-workers (Sheldon et al., 1972) and applicability of which to the wide size variety of pelagic creatures was demonstrated by Tseitlin (1986). This author showed (Tseitlin, 1981, 1985) that the biomass of pelagic organisms from the body size span from $\mathbf{L_{min}}$ to $\mathbf{L_{max}}$ can be expressed by the equation:

$$\mathbf{B = B_0 \ln(L_{max}/L_{min})} \qquad \mathbf{(1)}$$

... where $\mathbf{B_0}$ - a constant proportional to the local biomass value. The maximum size $\mathbf{L_{max}}$ is determined by the investigator who decides from which size on animals have to be removed before the biomass measurement will be done. In practice, of course, no one measures the animals before biomass determination which leads to some uncertainties. The minimum size $\mathbf{L_{min}}$ depends on the mesh size of the net applied and/or on the pore size of the filter which was used in the sample condensation. The formula **(1)** gives an idea about biomass dependance on sampling and laboratory technique and, moreover, can be used to improve the compatibility of data derived with different methods. Notably, in the situations when a flow meter is used (and, consequently, mesh size does not effect net efficiency), $\mathbf{L_{max}}$ is constant, and the average ratio of animals length and width equals to 3.3 (Saville, 1958; Timonin, 1983), the formula can be written (Rudjakov et al., 1996):

$$\mathbf{B(d_1) = k_d\, B(d_2)} \qquad \mathbf{(2)}$$

... where $\mathbf{k_d = \ln(4/d_1) / \ln(4/d_2)}$, $\mathbf{B}$ - biomass, $\mathbf{d_1}$ and $\mathbf{d_2}$ - mesh sizes of nets used (or pore sizes of filters applied if they are greater than the mesh size of the net).

Using formula **(2)** it is quite easy to calculate the biomass rate of different nets on the condition that the size distribution is close to the theoretical one. The latter condition can be valid for a large scale studies only, where, for instance, a net with mesh size 0.178 mm gives an estimate 1.25 times as large as the net with mesh size 0.33 mm.

There are four basic methods of biomass determination in plankton samples. Weight (carbon, dry, and wet) and displacement volume determinations are considered as effective enough (they are listed here approximately in order of their reliability), and the relationships between them are widely known and seem quite trustworthy (Wiebe et al., 1975; Wiebe, 1988; Vinogradov, Shushkina, 1987). The two other methods will be discussed here in more detail as their implementation evokes more doubts and questions.

a) Settling volume determination. The method basically consists of measuring the volume of the plankton sediment on the bottom of a jar with fixed sample. It is known that the density and resulting volume of the sediment depends on taxonomic composition and size spectrum of the plankton and, moreover, on the height and weight of the sediment itself. For this reason, the method is no longer recommended and its description is even left out of the papers devoted to the topic (Wiebe et al., 1975). Nevertheless, so many historical data sets contain the results of the settling volume determinations, and some of them were gathered in such purely investigated areas, that the question of their usefulness has to be reinvestigated. Judging from common knowledge, the conversion factors for settling volume determinations are region and even season dependent (M. Omori, personal communication) and their derivation demands for specially aimed investigations. Though, the possible benefits from using the data of such important expeditions as Dana (Jespersen, 1935) and Marchile (Fagetti, Fischer, 1964), indicate that these efforts worth trying.

b) Biomass calculations based on species counts and either individual weights or regressions connecting length and weight of animals are widely distributed. In Russian publications the results are often referred to as "zooplankton biomass" and opposed to "seston biomass", measured for the total catch, escaping the attention to the seston composition. The precision of the data depends on the accuracy of individual weight determination and/or on the applicability of the corresponding regression equations. As in the case of settling volume, these coefficients are region and season dependent and very often an investigator uses the coefficient set of his or her own, often unpublished. Nevertheless, this method of biomass evaluation must be considered as the best in the cases when the nets of mesh sizes smaller than 300 m were used. In such investigations seston biomass is, in an average, two times greater than zooplankton biomass and it is evident that incorporating in the weight or the volume of zooplankton the same of particles of unknown origin is unreasonable (Rudjakov, Tseitlin, 1992; Kitain et al., 1995). From this it follows that the sets of individual weights and regression coefficients must be accumulated at the data centers along with direct biomass measurements and species counts. Their regional changes and seasonal oscillations may compose either a topic of a targeted investigation, like that by Mizdalski (1988), and be useful in broad ecosystem research.

In respect of calculated biomass data the other questions persist. Speaking of the most important of them, one question is related to the development of taxonomy and concerns the taxonomical adherence of the person in charge of sample processing. The other one, closely connected with the first just mentioned, applies to the sphere of ethics, and, consequently, can be hardly solved satisfactorily. It is universally known that the results of species counts greatly depend on the planktonologist's qualification and diligence. In every scientific community an informal ordination list exists in relation to confidence to count results gained by different persons. The problem is how to preserve these lists and use them not coming into conflict with demands of open access to the data, not offending the feelings of workers, and not violating their career and thereby the possibility of future qualification improvement.

THE IMPORTANCE OF ADDITIONAL INFORMATION

Some observations which seemed to be of a little value or just considered as absolutely insignificant at the time of sampling or sample processing, might prove to be, sometimes, very important and here is an example of their implementation.

The excessively high, namely higher than 1 g m^{-3}, seston biomass values can be found in archive trip logs and articles. They were mostly observed in summer of temperate and high latitudes and in the upwelling regions. From the general experience, the most part of such cases is due to high phytoplankton development. The reason for that is given, for example, by the "Discovery" data on zooplankton biomass in the Southern Ocean which were derived taking special precautions aimed to eliminate phytoplankton before seston biomass determination (Foxton, 1956). As a result of these efforts, the values derived only rarely exceeded 0.5 g m^{-3} and never 1 g m^{-3}, while the Antarctic biomass data set at hand includes a lot of values in excess of 1 and even 10 or 20 grams per meter cube. The samples with such high biomass values which did not contain salps or medusae (they had been picked out earlier) were definitely taken in a patch of phytoplankton bloom. Unfortunately, cases of high phytoplankton content have not been routinely documented and if so, the doubts always persist whether the high values derived have to be attributed mostly to zoo- or phytoplankton.

The results of seston displacement volume determinations performed in the course of some Russian expeditions proved to be a happy exclusion to this practice of deficient data documentation. The diligent participants of those expeditions, as a rule, put down in the trip logs the eye-catching occasions of phytoplankton bloom, usually determined by the green color of samples, by the reduced rate of sample filtration, or by the direct microscopic investigation. There were about 200 seston biomass values which were taken at the sites where phytoplankton bloom was registered. Being mapped, the sites made the pattern coincident with the remotely sensed distribution of maximum phytoplankton abundance (Maslanik, Barry, 1990). These data showed close similarity to the log-normal distribution with parameters: mean = 3.175 and standard deviation = 0.526 (biomass was measured in wet weight mg m^{-3} and logarithms to base ten were taken). These parameters were used in the data quality control, in a computer program discriminating the samples and stations where seston presumably contained large amount of phytoplankton. The algorithm exploited the random number generation procedure and the idea that probability of phytoplankton contamination rises with seston biomass value in accordance with parameters above. The results of this work were used in the mapping of zooplankton biomass values in the Southern Ocean and, besides, they can be used to scale the remotely sensed data by ship observations. This example is very informative since it clearly demonstrates how important may be supplementary information which was never used earlier.

CONCLUSION

Summing up all above, it should be underlined here that the additional information or metadata is always needed and has to accompany the results of biomass measurements and calculations. This information may be subdivided into two categories:

Rudjakov J.A., Voronina N.M. 1967. Plankton and bioluminescence in the Red Sea and the Gulf of Aden - Okeanologya, 7, 6:1076-1088

Rudjakov J.A., Kitain V.J., Tseitlin B.B. 1996. The mean annual zooplankton biomass distribution in the upper layer of the Pacific - Okeanologya (in press).

Rudjakov J.A., Tseitlin V.B., Kitain V.J. 1995. Seasonal Variations of Mesoplankton Biomass in the upper Layer of the Bering Sea; understanding biomass oscillations in the ocean - ICES J. mar. Sci., 52, 747-753.

Saville A. 1958. Mesh selection in plankton nets - J.Cons.perm.intern.Explor.Mer., 23:192-201.

Sheldon R.W., Prakash A., Sutcliff W.H. 1972. The size distribution of particles in the ocean - Limnol.Oceanogr., 17,3:327-340.

Timonin A.G. 1983. Closing plankton nets for vertical catches of mesoplankton. In: Vinogradov M.E. (ed). Contemporary methods in quantitative appraisal of marine plankton distribution. Moscow: Nauka, p.158-173.

Tseitlin V.B. 1981. The size distribution of pelagic organisms in the tropical ocean - Okeanologya, 21, 4:125-131.

Tseitlin V.B. 1985. The size distribution of organisms in different ecosystems - Reports of the USSR Acad.Sci., 285, 5:1272-1276.

Tseitlin V.B. 1986. Energetics of deep-sea pelagic communities. Moscow: "Nauka", 102 p. (In Russian).

Vinogradov M.E. 1968. Vertical distribution of oceanic zooplankton. Moscow: Nauka, 320 p. (In Russian)

Vinogradov M.E., Shushkina E.A. 1987. Functioning of the epipelagic ocean plankton communities. Moscow: Nauka 240 p. (In Russian)

Vinogradov M.E., Tseitlin V.B. 1983. Deep-Sea Pelagic Domain (Aspects of Bioenergetics). In: Rowe G.T. (ed.). Deep-Sea Biology. The Sea, vol. 8. Jihn Wiley and Sons, N.Y., p.123-165.

Wiebe P.H. 1988. Functional regression equations for zooplankton displacement volume, wet weight, dry weight, and carbon: a correction - Fish. Bull., 86, 4:833-835.

Wiebe P.H., Boyd S., Cox J.L. 1975. Relationships between zooplankton displacement volume, wet weight, dry weight, and carbon - Fish.Bull., 73, 4:777-786.

Zaikin A.N., Rudjakov J.A. 1989. Distributivity criterion of pelagic communities and kinetic regimes of its heuristic models - Reports of the USSR Acad.Sci., 306, 4:977-980.

TAXONOMIC CODE SYSTEMS AND TAXONOMIC DATA MANAGEMENT

Linda Stathoplos

ABSTRACT

Combining plankton data requires managing taxa names and data associated with non-taxonomic groups. Problems with managing taxa names include that they are not unique, they can have synonyms, and that they are hierarchically related. Plankton data often include misspelled taxa names and non-taxonomic groups. Solutions to these problems, employed by the U.S. NODC Ocean Climate Laboratory to handle plankton data in their Global Ocean Database, are described. The NODC Taxonomic Code is used to: (1) associate each taxon name with a unique taxonomic serial number (TSN); (2) maintain taxon name synonymy (the variety of names that mean the same taxon); (3) distinguish among taxa with the same name; and (4) hierarchically group names. Non-taxonomic groups are characterized in the Global Ocean Database by a group of descriptors including trophic mode, realm, shape and size information. If provided, life stage and sex are also recorded. This suite of descriptors provides a flexible, robust and logically consistent way of combining a wide range of plankton observations.

INTRODUCTION

It is desirable to describe the taxonomic composition of plankton to understand pelagic ecosystems. To do this, several difficulties must be overcome. First, counts, relative abundance, and presence/absence data are commonly reported by non-taxonomic group. Second, even when these data are associated with a taxon name, that name may be modified with additional information, combined with another taxon name, or have a non-standard spelling. Finally, for data quality control, it is necessary to be able to group taxa hierarchically, for example to identify all the species that are copepods.

TAXONOMIC CODES

Taxa names are defined here as those issued according to the rules of the International Codes of Zoological and Botanical Nomenclature (Ride, et al., 1985; Lanjouw, et al., 1966) and their more recent amendments. Taxa names are distinct from other terms commonly used to describe groups of organisms, such as 'jellyfish', because there are defined rules for their application, and because of their hierarchical relationship. A number of schemes have been devised to assign codes (usually numbers) to taxa names. These codes are either 'intelligent' (meaningful) or arbitrary (meaningless). Here, the NODC Taxonomic Code provides one illustration of schemes used to codify taxa names.

The U.S. NODC has produced and maintained a taxonomic code for many years. Up until recently, 'intelligent' keys were assigned to taxa names, that is, numbers that reflect the position of that name in the Linnaean hierarchy. Thus, all euphausiid taxa names would begin with the number 6174, for example. While this approach had the advantage that taxonomic relationships were obvious from the numerical code, three problems emerged: certain portions of the hierarchy would run out of numbers, taxa would change number when they were reclassified, and multiple names had the same code (synonyms), so it was not possible to translate a number back to a unique name. Thus, the most recent version of the NODC Taxonomic Code, version 8.0 (available on CD-ROM from NODC User Services in June, 1996), contains the last set of 'intelligent' codes to be issued. Now, when taxa names are added to the NODC Taxonomic Code, they are issued arbitrary taxonomic serial numbers (TSNs). The hierarchical position of a taxon name is now maintained by pointers between these TSNs in a relational database.

It is the NODC Taxonomic Code TSNs that are used in the OCL Global Ocean Database to represent taxa names. The NODC Taxonomic Code database then keeps track of: (1) the TSN associated with a specific taxon name; (2) the set of names associated with a single taxon (synonymy); (3) the different meanings of a single name (e.g. Clionidae, both a sponge and a gastropod family), and (4) the hierarchical relationships of the taxa names (e.g. all the names that are copepods). In this way, records in the Global Ocean Database can be dynamically grouped by taxon for purposes of data analysis or quality control.

PLANKTON 'TAXONOMIC' DATA

Plankton abundance and presence/absence observations are usually reported by taxonomic group, for example the number of Copepoda per cubic meter. However, data from 'taxonomic' groups often includes non-taxonomic groups as well; examples from actual data are listed in Table 1. To enter these data into the OCL Global Ocean Database, the descriptors in Table 2 are used. The steps used to apply these descriptors are described below.

For those data with taxa names, if duplicates are found in the NODC Taxonomic Code, the correct TSN can usually be discerned from the sample type, e.g. Radiolaria in plankton samples are the protist order, not the insect genus. Spelling differences in a taxon name between originator's data and the NODC Taxonomic Code are resolved as follows: (1) the correct spelling is verified in the literature; (2) if the originator's spelling is preferred, the NODC Taxonomic Code spelling is update; (3) if the NODC Taxonomic Code spelling is preferred, then either (a) the PI is informed and consulted to verify that the two names are indeed variant spellings of the same taxa, or (b) if the PI is unknown or unavailable, NODC personnel verify the equivalence and keep a record of the variant spelling in a 'readme' file.

Taxon names that are verified as missing from the NODC Taxonomic Code (rather than spelled differently) are entered into the Global Ocean Database with temporary (negative) serial numbers. Once NODC Taxonomic Code taxonomic serial numbers are generated, these temporary negative numbers are replaced with TSNs in the Global Ocean Database. This is to avoid divergence between the set of taxonomic serial numbers in the NODC Taxonomic Code and the Global Ocean Database. Previously, NODC experienced difficulty with their plankton data formats containing data submitted with NODC Taxonomic Code-like numbers. These NODC Taxonomic Code-like numbers were generated by the data originator, but never were added to the NODC Taxonomic Code. The ability to translate these numbers back to a taxon name was lost, because no look-up table for these NODC Taxonomic Code-like numbers existed.

Non-taxonomic and modified taxonomic groups (Table 1) provide another challenge for data management. This information is stored by code lists using the descriptors listed in Tables 2 and 3.

As described above, non-taxonomic groups are assigned negative serial numbers. If taxa names are modified to indicate pieces of an organisms, such as frustules, spicules, etc., then counts are not recorded for that taxon. Rather, the taxon is recorded as 'present'. Information about realm, trophic mode, or sex is recorded under the appropriate descriptor by use of code lists.

Life stage terms are common in plankton lists. We have separated them from taxa names. This process is complicated by the fact that some life stage names are also taxa names, like Arachnactis, which is both a genus name, and the name of the larval stage of the family Cerianthidae. It is important for data managers to correctly distinguish the two. When a larval stage has a specific name, like Phyllosoma, the life stage is coded specifically as Phyllosoma rather than as the more generic term larva (Table 3).

The size descriptors (minimum and maximum size range, or a descriptive term) are used to capture both the size class of the group, e.g. 5 - 10 microns, and adjectives like 'small'. The shape descriptors include both descriptive terms, like 'rod-shaped' or 'spherical', and the dimensions of the shape, where for example the length and width are provided for an undescribed species of a particular phytoplankton genus.

CONCLUSION

The OCL has devised a successful method of combining historical plankton observations of both taxonomic and non-taxonomic groups of organisms. By using a small number of descriptors, and separating different types of information into distinct descriptors, a wide variety of data can be combined into a uniform format for quality control and data analysis.

ACKNOWLEDGMENT

Support for this project was provided to L. Stathoplos by NOAA's Office of Climate and Global Change (GC95-220).

REFERENCES

Ride, W.D.L., et al., 1985. International Code of Zoological Nomenclature, 3rd ed., adopted by the XX General Assembly of the International Union of Biological Sciences editorial committee. 338 p. Berkeley: University of California Press.

Lanjouw, J., et al., 1966. International Code of Botanical Nomenclature, adopted by the X International Botanical Congress, Edinburgh, August 1964. 402 p. Utrecht: International Bureau for Plant Taxonomy and Nomenclature, International Association for Plant Taxonomy.

TABLE 1: Examples of 'taxonomic' plankton data. These data include non-taxonomic groups, combined taxa names, life stages, modified and misspelled taxa names.

Group	Comment
jellyfish	non-taxonomic
algae (benthic)	non-taxonomic
5 micron coccoid paired cells	non-taxonomic
Phototrophic nanoplankton	non-taxonomic
Small truncated-conical flagellate	non-taxonomic
unspecified flagellate 10-15 microns	non-taxonomic
Salps + Doliolids	combined taxa
Phyllosoma	life stage
Pilidium	life stage
Other Mollusca	modified taxon
Biddulphia sp. (frustule)	modified taxon
Cocconeis sp. (epiphyte, 29 micron, 15 micron)	modified taxon
Dissodinium pseudolunula (cysts)	modified taxon
Rhizosolenia shrubsoleii (pieces)	modified taxon
Calanus finmarchius (males)	modified and misspelled taxon

TABLE 2: Descriptors for plankton groups used in the Ocean Climate Laboratory's Global Ocean Database. Only descriptor one (1) is required; others are entered when provided.

1) TSN (NODC taxonomic serial number) for taxon name
   - -OR-

   SN (negative serial number) for non-taxonomic group
2) Taxon name or group modifier
   - sp.
   - spp.
   - other
3) Realm
   - pelagic
   - benthic
   - epibiont
   - endobiont
4) Trophic mode
   - autotroph
     - chemotroph
     - phototroph
   - heterotroph
     - parasitic
     - saprophytic
   - plasitidic
   - non-plastidic
5) Life stage (see Table 3)
6) Sex
   - male
   - female
   - hermaphrodite
     - functional male
     - functional female
7) Size class, mm
   - minimum and maximum of size range
   - -OR-
   - descriptive term (negative numbered code list)
8) Shape
   - descriptor
   - -AND-
   - dimensions (microns)
     - radius
     - length
     - width

TABLE 3: Life stage code list. Where the larval stage name implies a particular taxon, that taxon name and taxonomic serial number (TSN) are listed as well. (Negative serial numbers, SNs, indicate non-taxonomic groups.)

Code	Life stage	(T)SN	Taxon name or non-taxonomic group
1	EGG/OVA		
2	NAUPLIUS		
3	ZOEA	83677	CRUSTACEA
4	MEGALOPA	98276	BRACHYURA
5	VELIGER	69459	GASTROPODA
6	LARVA		
7	JUVENILE		
8	ADULT		
9	LARVA+JUV+ADULTS (6-8)		
10	COPEPODITE V		
11	POSTLARVAE (7-8)		
12	CYPHONAUTES larva	155469	BRYOZOA
13	PHYLLOSOMA larva	97646	PALINURIDAE
14	PILIDIUM larva	57411	NEMERTEA (or NEMERTINEA)
15	TORNARIA larva	158617	ENTEROPNEUSTA
16	TROCHOPHORE larva	-5002	ZOOPLANKTON
17	ARACHNACTIS larva	51985	CERIANTHIDAE
18	ACTINOTROCHA larva	155457	PHORONIDAE
19	EMBRYO		
20	CYPRIS larva	89433	CIRRIPEDIA
21	BIPINNARIA larva	156862	ASTEROIDEA
22	OPHIOPLUTEUS larva	157325	OPHIUROIDEA
23	ECHINOPLUTEUS larva	157821	ECHINOIDEA
24	HYPNOSPORES		
25	COPEPODITE I		
26	COPEPODITE II		
27	COPEPODITE III		
28	COPEPODITE IV		
29	COPEPODITE		
30	CALYPTOPIS	95496	EUPHAUSIACEA
31	FURCILIA	95496	EUPHAUSIACEA
32	NAUPLIUS I		
33	NAUPLIUS II		
34	NAUPLIUS III		
35	NAUPLIUS IV		
36	NAUPLIUS V		
37	METANAUPLIUS		
38	POLYP		
39	MEDUSAE		
40	INDETERMINABLE		

COMPUTER TOOLS FOR MAPPING PELAGIC BIODIVERSITY

P.H. Schalk & R.P. Heijman

ABSTRACT

Protecting the world's biodiversity depends on reliable species identification and readily accessible documentation. To study biodiversity and to monitor changes in our biological environment high quality taxonomic and distributional data are imperative. At present access to species information and identification keys is limited by the fact that literature is scattered over a vast amount of sources. Exchange of biodiversity data between various researchers is hampered by the lack of compatible documentation tools.

ETI developed a standard computer tool (Linnaeus II) that combines the functionality of an interactive multimedia database, computer guided identification and a geographic information system. This software package is distributed to taxonomists and ecologists world-wide. It allows the specialist to create computer based biodiversity information and species identification systems on their own PC and exchange the standard data sets.

Most Linnaeus II users are organised in international networks. There data are stored in the World Biodiversity Database (WBD). Finished sections are released on CD-ROM, and concern typically taxonomic monographs or regional information systems and form the backbone of a modern digital library. An overview of the WBD contents will be made accessible on a World Wide Web site.

INTRODUCTION

All biodiversity studies have to be based on good taxonomy and reliable species inventories. Only reliable identifications will render high quality data sets on species diversity. Compared to the terrestrial biotope, marine biological sciences are still far from having a complete overview of ocean biodiversity and ecosystem structures. In general that what is known on species is spread over a huge amount of scattered literature and accessing these data is time consuming. There is an increasing need for accessible and reliable information on all known species and a strong desire to strengthen in-county user-friendly biodiversity databases (Braatz et al., 1992). A universal tool for biodiversity documentation is needed to promote marine biodiversity science and make existing knowledge (fast) accessible on a world-wide basis.

The only answer to this growing demand for directly accessible biological data is a universal standardised computer based biodiversity information system. This system must be flexible enough to cope with, and incorporate, the many small limited scope, databases in existence and must have the potential to produce (regional) master species databases (Bisby, 1993). For determining species composition, fast, standardised and universal identification tools are needed, based on expert's knowledge to ensure high quality of data. Only when a good inventory is made, monitoring studies can start, to follow and document changes. For monitoring studies it is evident that good and reliable identifications are imperative to be able to compare data sets of different working groups and geographical areas, especially when one aims at comparing biodiversity values (Raven & Wilson, 1992). A new computer approach for biodiversity surveys is the answer.

By using computer tools and modern interactive multimedia technology it is possible to create easy to use and cheaply updatable identification and information systems (Estep et al., 1989; Estep et al., 1993, Schalk & Los, 1992) that can be distributed to biologists world-wide. By including simple GIS systems that run on PC's, species distribution data can be stored and compared as well. In this article we discuss the possibilities and application of such a system, as a basic tool for biodiversity documentation.

THE EXPERT CENTER FOR TAXONOMIC IDENTIFICATION

Taxonomy and systematics form the basis for most biological sciences. Especially in the context of biodiversity studies and biomonitoring of ecosystem changes good taxonomy is an absolute necessity for obtaining reliable results. Despite this, the number of taxonomists is decreasing (May, 1990) and the documented species information and identification keys are difficult to access as they are scattered over a huge amount of literature (Schalk, 1992). Computer tools are an answer to preserve and exchange knowledge (Pankhurst, 1991, Schalk & Los, 1992).

The Expert center for Taxonomic Identification (ETI) is a non-profit UNESCO associated organisation at the University of Amsterdam. The center was founded in 1990 with financial support from the Dutch government, UNESCO and the University of Amsterdam and works together with major taxonomic institutions world-wide. ETI has three main aims: 1) to develop novel interactive multimedia software for biodiversity documentation and species identification; 2) to stimulate international networking of taxonomic specialists; 3) to build up a World Biodiversity Database containing detailed records of presently described and accepted species and disseminate this knowledge using modern electronic publishing tools.

ETI has developed a user friendly, easy to handle, universal biodiversity software shell (Linnaeus II) that allows specialists to store information on species in text, pictures, videos and sounds in a standard and exchangeable format using their own PC (Schalk, 1992; Estep et al., 1993). More important, this shell enables the user to construct "smart" (expert) identification systems giving both specialists and laymen fast access to the multimedia species information, as well as traditional interactive keys. A geographic information system is incorporated for storing and interactive use of biogeographic information. The Linnaeus II system has various support functions on-line, such as glossaries and methodology sections and a literature database. The easy use of the system will open up taxonomic keys also to laymen or students and these programs may also be used as a modern teaching tool.

The Linnaeus II system can be used on Macintosh and Windows PC's. It is an excellent and user friendly tool for biodiversity documentation. It allows the user to concentrate on the science instead of how to operate the program.

A BIODIVERSITY INFORMATION SYSTEM: LINNAEUS II

The Linnaeus II program contains six major sections: databases for species information and higher taxa, three programs for identification, a geographic information system, a references database with all relevant literature, a glossary section (with all terms) and an introduction section (with general information on the group treated and methodology). All parts are fully multimedia and may contain texts, photo's, sounds and videos. The simple graphic user interface and hyperlink system guarantees easy point-and-click navigation through the various parts of the program.

The species database can be accessed by a mouse click on its button on the navigator card of the program. It looks like a filecard with an overview picture of the species on the right and an abstract description on the left. From this point it is easy to page through the species cards for quick comparisons. A click on the detail "button" leads to the "detail card" of the species database, from which the various information fields can be accessed. There are fields for the description (with chapters for taxonomic, description, diagnosis, reproduction, ecology, practical importance), taxonomy (listing all higher taxa names which are hyperlinked to give access to the higher taxa database), synonyms (allowing searches on old names or common names), references (specific to this species), and multimedia clips (listing all illustrations, sound files, videos, which can be accessed by a simple mouse click). The higher taxa database is similar to the species cards with the same functionality but storing information on higher taxa such as family, order, class etc. Both these databases can be approached directly from the navigator section, but also through free text search functions or from the Identification program.

The IdentifyIt section is a multiple entry approach to identification using the computer's power to search and compare. Instead of a rigid binary choice system that one has to follow step by step, in IdentifyIt a data matrix is used to identify a species. The user has a free choice to select any character. Every character has a number of states, one or more states may be chosen, or even "unknown" may be selected. Characters and states are linked to descriptive texts, pictures, videos or sounds where appropriate, which can be called up by a button click. Once a choice has been made, the program sorts the species list in order of probability, shown as hit percentages. A click on the species name leads to the species database for a quick check, after which one can return to the identification program. It is also possible to compare species in the "hit list" for both common and exclusive characters. This way of identifying has great advantages compared to the conventional, printed, and rigid digotomous keys, as it allows the user to start with any known or available character. Thus, even damaged specimens may be identified as the system also works with incomplete character/state sets and answers with a list of probable species.

The geographic information section (MapIt) stores biogeographic data in a grid system. Data can be stored and viewed on a worldscale level or in several (linked) more detailed large scale maps. MapIt is fully interactive, meaning that the user can select one or more squares and do a query for species occurring there. It is also possible to compare distributions between species or search for patterns. MapIt also can present species diversity by giving an account of the number of species under a square (on any scale) and depicting this in a colour pattern.

May, R.K. 1992, Past efforts and future prospects towards understanding how many species there are. In: Solbrig, O.T., van Emden, H.M. & van Oordt, P.G.W.J. (eds). Biodiversity and global change: 71-81. (IUBS Monogr. 8)

May, R.M., 1988. How many species live on earth? Science 241: 1441-1449

May, R.M., 1990. Taxonomy as destiny. Nature 347: 129-130

Pankhurst, R.J., 1991. Practical Taxonomic Computing. Cambridge Univ.Press. 202 pp.

Pielou, E.C., 1966. The measurement of diversity in different types of biological collections. J. theor. Biol., 13: 131-144.

Raven, P.H. & E.O. Wilson, 1992. A fifty year plan for biodiversity surveys. Science 258: 1099-1100

Schalk, P.H., 1992. Computer-aided Taxonomy. Binary 4: 124-126

Wilson, E.O. 1992. The diversity of life: 1-424. (Belknap Press Harv. Univ. Press, Cambridge Mass.)

LONG-TIME SERIES OF HYDROLOGICAL AND PLANKTON DATA FROM LEBANESE WATERS (the Eastern Mediterranean)

S. Lakkis

ABSTRACT

Modifications in the pelagic ecosystem are presented under different aspects. Plankton samples and hydrobiological measurements collected from neritic and oceanic Lebanese waters (Eastern Mediterranean) since 1969 provided large amount of data useful for time series analysis study. These data showed seasonal variations and interannual fluctuations of planktonic species in relation to hydrological features. Spatial and temporal distributions of phytoplankton and zooplankton populations are related to environmental factors, namely the hydrology, the nutrient concentration and food supply. Most of the regular and common species are more affected by seasonal variations and modification of the community structure, while the "irregular" species exhibited annual and interannual fluctuations in abundance. Regular seasonality of hydrobiological parameters is an important feature of the Levantine Basin. Annual fluctuations of both phytoplankton and zooplankton are greatly related to seasonal thermic changes. The peak of abundance of phytoplankton, followed directly by that of zooplankton, occurred usually in spring time: April-June. During summer period, water layers stratification and nutrient's depletion induce the breakdown of the plankton populations manifested by paucity of the zooplankton biomass and a low rate of phytoplankton primary production. Normal interannual fluctuations and regular seasonal variations are a natural phenomenon within the planktonic community; such phenomenon is inherent to the physiological aspect of living organisms and to environment conditions. There is no chaos enhancing this variability in the ecosystem; certain deterministic reasons are probably responsible for the spatio-temporal variability within the pelagic ecosystem.

INTRODUCTION

The reasons of changes in the pelagic ecosystems and in the structure of marine communities are not well explained. In order to quantify the spatio-temporal variability of plankton community and ecological factors, we should consider the natural environmental factors like hydrology, climatology and hydroclimatology prevailing in the pelagic ecosystem. Although other factors like pollution and anthropogenic activities have great impact on the living communities, they should not be taken into consideration in this aspect and be eliminated. Variability within the plankton community may have biological reasons such as biological cycle, nutrition, food supply, breeding, growth, etc. or affected by environmental variables such as temperature, salinity, dissolved chemicals and oxygen. The study of the evolution of plankton community in the Lebanese sea water is carried out by using a long-time series analysis of hydrographic and plankton data. In this paper we try to present some aspects of these data gathered during the last 26 years.

Previous to 1968 few information on the hydrobiology and plankton of the Lebanese waters (eastern Mediterranean) were available. Gruvel (1931) in his monograph on the marine resources and fisheries from the Syrian and Lebanese coastal waters provided some results on commercial marine organisms. Rouch (1945) published few data on the surface sea water temperature off Beirut coast. George et al. (1964) gave an inventory of 350 species of marine fishes from Lebanon with emphasis to the most important commercial fishes. Mouneimné studied the biology of most common marine fishes of related to fisheries in Lebanon. As for hydrology and plankton, more regular investigations were carried out since 1968 several results and data are published in this respect (Lakkis, 1971; 1990a).

HYDROBIOLOGY OF THE AREA

The Levantine Basin (eastern Mediterranean) falls in a Mediterranean temperate area, having a moderate climate with certain subtropical affinity. The weather is characterized by a short winter period (December-March) showing an annual rainfall average of 1000 mm, a hot and dry long summer period (June-November) with a high temperature and strong evaporation. The spring short inter-season during April-May shows moderate sea water temperature (23°C ±2). The Levantine Sea, including the Lebanese area is considered as an oligotrophic and hypersaline water; the evaporation is higher than fresh water input. After the functioning of the Aswan high dam in 1964, the salinity trend of the Lebanese oceanic water increased 0.8‰ the annual average being actually 39.45‰ ±0.18 the highest value in the entire Mediterranean. The Levantine Sea is weakly affected by Atlantic temperate water through the Gibraltar strait current branch. This current directs towards the eastern Mediterranean parallel to north African coast up to the Nile delta area. The direction of the current changes to the north-east and became parallel to the Syro-Lebanese coast. Previous to 1965, the hydrology of this area was greatly influenced by the Nile flood, when the salinity dropped to 37‰ off Beirut coast (Rouch, 1945). After the regulation of the Nile river by Aswan high dam, the flood of the Nile that was occurring in August-September does not happen any more. The role of the Suez canal as a link between the Red Sea and the Mediterranean is evident, several species of Indo-Pacific species migrate through the pathway and established permanent populations in the eastern Mediterranean (Lakkis, 1980).

Plankton samples accompanied by hydrological measurements were gathered during 26 years: 1969-1995 (1976 is missing) from several coastal, neritic and oceanic sampling stations along the central and northern Lebanon (33° 52'N, 35° 29E and 34° 30'N, 35° 55'E) (Fig. 1). Plankton data are based on surface and vertical hauls using plankton nets of 50μ mesh size for microplankton collection and 100μ, 200μ, 300μ and 500μ bolting silk nets for meso and macrozooplankton organisms. Hydrobiological data comprised CTD probe measures for sea water temperature, salinity, and oxygen saturation; transparency was measured by Secchi disk and Nansen reversing bottles were used to collect water samples for laboratory chlorophyll a, phosphates and nitrates analysis. Five series of samples were collected from thirty stations during the whole period of survey, gathering 2850 samples distributed as follows:

-**Beirut Area**: 1969-1972 involving six stations situated on three different transects: three coastal and six offshores; vertical and surface monthly sampling series provided 500 samples.

-**Northern Lebanon**: 1972-1975 including four stations taken from different areas: coastal, industrial polluted and offshore waters: a series of 150 samples were collected seasonally and monthly by using horizontal and vertical net plankton tows.

-**Byblos Area**: 1972-1977. Ten stations were fixed: two inshore water, two estuarines (Nahr Ibrahim river) and six oceanic stations in deep water beyond the narrow continental shelf. Based on monthly sampling intervals, this area provided 750 phytoplankton and zooplankton samples.

-**Jounieh Area**: 1978-1988. Six stations were fixed covering coastal water, harbour, polluted and offshore stations ; series of 800 samples were gathered using 50μ and 200μ mesh size plankton net from horizontal and vertical tows.

-**St. Georges Bay**: 1988-1995 including seasonal series, monthly, weekly and diurnal series (night and day) from four stations: harbour, polluted, estuarine, coastal and offshore stations, they provided 650 samples.

The number and date of plankton samples are given in Figure 2. Additional details on the characteristics of the stations and the sampling procedure are described in previous studies (Lakkis, 1971;1990a).

SEASONAL VARIATIONS

The sea surface temperature ranges between a minimum of 16°C ±1.20 in February and a maximum of 30°C ±1.08 in August ; the annual mean being 23.50°C ±2.36. The seasonal variations of temperature in coastal waters are greater than in oceanic area. The annual thermal cycle is separated in two phases: a short cold winter phase (December-March) and a hot dry thermal phase in summer (June-November); an inter-season during April-May shows moderate values of temperature and salinity. During the cold period, there are homothermal conditions throughout the water column, during which turnover and mixing water layers occurred. Minimum water temperature of 16°C ±1.20 is registered throughout the water column to 150 deep; the salinity shows its minimum value of 39.10‰ ±0.25 at offshore stations. During the summer, a sharp thermocline is establishing between 35 and 75 m inducing a subsurface water layers stratification and a depletion in dissolved nutrients; a paucity in phytoplankton primary production and zooplankton biomass characterize the summer season. During this period the temperature increases up to a maximum of 30°C ±1.08 in August and September. At offshore stations oxygen saturation decreases down to 3-4 ml/l ±0.75 in August and September. Orthophosphate concentration shows its maximum value in September (0.11 μmol./l) and a minimum of 0.003 μmol/l in May, the annual mean being 0.07 ±0.08 mol./l. The N-NO_3 concentration has an annual average of 0.12 μmol/l ±0.14 with a maximum in September (0.37 μmol/l) and a minimum in January (0.01 μmol/l). Chlorophyll content has its peak during the spring bloom, coinciding with the highest phytoplankton density of 500,000 cells/l ±150,000. This biomass decreases in summer (July-August-September) to 0.08 mg/m^3 corresponding to minimum cell density of 40,000 ±20,000 The abundance of zooplankton shows the same seasonal variations pattern with a peak in April-May (1.98 cm^3/m^3) and a minimum in August (0.08 cm^3/m3), the annual average being 0.61cm^3/m^3 ±0.69. Variability of all hydrobiological factors are given in Table 1 and illustrated in Figures 3 and 4. At coastal water, the hydrobiological parameters are much higher than at oceanic water and are subject to great daily, weekly and monthly variations. Zooplankton biomass and the phytoplankton primary production are more important in neritic waters and coastal waters the input of nutrients with inflowing rivers being very important.

Table 1- Long-term seasonal and interannual variability of hydrobiological data from oceanic Lebanese water during 1969-1995 (combined data)

Parameters	Seasonal Variability Average±DS	%V.C.	Interannual Variability Average±DS	%V.C.
Temperature °C	23.68±6.3	27	23.68±2.36	10
Salinity S‰	39.00±1.5	4	39.28±0.18	0.04
Diss. Oxygen (ml/l-1)	6.46±0.78	12	6.46±0.5	8
Secchi disk (m.)	17.71±6.15	35	17.71±3.72	21
P-PO4 (µmol./l-1)	0.07±0.08	114	0.07±0.04	57
N-NO3(µmol./l)	0.12±0.14	83	0.12±0.08	66
Diatoms cell/l-1 x1000	157.0±163	104	157.0 ±96.0	71
Dinoflag. cells/l-1 x1000	3.83±2.3	60	3.83±1.67	49
Chlorophyll a mg/m3	0.68±0.75	109	0.68±0.57	84
Biomass Zoopl. cc/m3	0.61±0.69	103	0.61±0.45	73

Based on monthly data, matrix of correlation coefficient shows similarity level between the different hydrobiological factors. It is clear to note that the seasonal variations of temperature coincide with those of the salinity (r =0.85), transparency (r = 0.95) and are inversely proportional to dissolved oxygen saturation (r = -0.76).

INTERANNUAL FLUCTUATIONS

Hydrobiological data compiled throughout the whole period of survey show annual fluctuations in abundance and interannual variability. These fluctuations are more or less regular from year to year. The results of the variance analysis (Table 1) show the highest interannual variation coefficient within chlorophyll a concentration (84%), followed by zooplankton biomass (73%), diatoms (71%), nitrates (66%) and phosphates(57%). Salinity has the lowest variability (0.04%), while the temperature variations was 10%. In Figures 5 and 6, we illustrate the long-time series of some hydrobiological factors at oceanic water. Generally the seasonal variations and interannual fluctuations of all the parameters are higher within coastal waters than at oceanic area, except for temperature; the reasons are the influence and the irregularity of rivers freshwater discharge in coastal and estuarine waters.

Table 2- Matrix of correlation coefficient analysis of the ten selected hydrobiological parameters based on combined monthly data from Lebanese oceanic water during 1969-1995.
(* = Significant at 0.05 level)

	Temp	Sal	Oxy	Trans	PO_4	NO_3	Diat	Dino	Chlor a	Biom
Temp	1.00	0.85*	-0.76*	0.95*	-0.007	0.56	0.05	0.07	0.08	0.34
	Sal	1.00	-0.84*	0.52	0.24	0.74*	-0.07	0.15	-0.29	0.19
		Oxyg	1.00	-0.77*	-0.55	0.72*	0.68	0.53	0.51	0.16
			Trans	1.00	-0.83*	0.68	-0.55	0.62	-0.25	0.30
				PO_4	1.00	0.17	-0.24	0.40	-0.30	-0.72*
					NO_3	1.00	-0.52	0.60	-0.45	0.19
						Diat	1.00	-0.25	0.85*	0.52
							Dino	1.00	0.32	0.57
								Chlor a	1.00	0.35
									Biom	1.00

Principal component analysis of hydrobiological data throughout the period of survey was computed. The bi-plot of the ten parameters plots in relation to the 26 years is given on the two axes plan I and II (Fig. 7). Majority of the years showed small variability, their corresponding points are plotted close to the zero centre of XY plan; except the years 1972,73,74,75,77,78 and 1994, they had big variance and great interannual variability their plots are more distant from the center of the two first and second axes plan. Most of the variables are scattered close to the centre zero, the salinity is the nearest point, followed by chlorophyll a concentration and water transparency; those showed lower variance than the other farest points.

PLANKTON DATA

The identification up to the species (when possible) is done for all groups and the counting number of organisms was carried out for phytoplankton(cells/l) and zooplankton (individuals/m^3) for each determined species. The investigated area was divided in grids and all data were integrated in two main areas: coastal and oceanic Among the three main phytoplankton groups present in the area, the diatoms constituted 70% of the total biomass and 40% of the species, while the dinoflagellates counted less than 10% of the total phytoplankton biomass and 58% of the total identified species; the silicoflagellates being the smallest group: 4%. Taxonomical study provides 220 species of *Dinophyceae,* 151 *Bacillariophyceae*, 6 *Silicoflagellata* and two *Ebriidae*. Within zooplankton community, the *Copepoda* with 172 species come the first group by number of species and represent 50 to 75% of total zooplankton biomass (Fig. 8).

The species diversity, measured by the index of Shannon, is the highest for most of groups during fall and winter, corresponding to the lowest population density. On the contrary, the diversity index is the lowest in spring during the phytoplankton bloom and the zooplankton growth. This index generally increases from the coastal water toward the open sea and from the sea surface to the deep layers.

The biodiversity of the plankton community shows a high richness of species in comparison to other Mediterranean areas (Lakkis, 1996). Planktonic flora and fauna of Lebanese water are of temperate Mediterranean type, with certain subtropical affinity. Lessepsian migration from the Red Sea to the Eastern Mediterranean, enriches the marine environment and pelagic ecosystem with many Indo-Pacific species. Several migrant species settled themselves in the Levantine Basin forming permanent and regular populations (Lakkis, 1980; Ben Tuvia, 1985). During the last 26 years we noted at least 30 species of Erythrean and Indo-pacific origin; they established themselves in the Levantine Basin and formed regular populations. Among those we mention: *Calanopia elliptica, Labidocera pavo Ceratium egyptiacum f.suezensis, Rhopilema nomadica* etc...). During the same period we observed some changes in the plankton community, few species disappeared suddenly or became very rare such as: *Isias clavipes, Centropages typicus, Acartia italica, A.discaudata (Copepoda)*. While many other new introducing Lessepsian species appeared suddenly or progressively such as: *Ceratium breve, C.egyptiacum, Dinophysis* spp., *Labidocera pavo, Centropages furcatus,Rhopilema nomadica*,several unidentified decapod larvae. The jellyfish *Rhizostoma pulmo* is replaced by *Rhopilema nomadica*; *Centropages typicus* by *Centropages ponticus*.

In Figure 9, we can note the great oscillations of diatoms comparing to dinoflagellates and other zooplankton species. Interannual variability is the highest for *Siphonophora* (200%), followed by the Indo-Pacific origin copepod *Labidocera* (107%) and by Prosobranch larvae(100%). In general the regular groups and permanent species show moderate interannual variability, while the "irregular" species present great variability (Table 3).

DISCUSSION

Management of long-time series data is not an easy task, because we not only need to use hydrological and biological data, but also hydroclimatology and meteorology need to be considered (Goy et al., 1996). Unfortunately, the latter are not always available. Another approach for studying physiological responds of planktonic animals is breeding and developmental studies under controlled experiments.

By using simulation models, we can recreate environmental conditions for the major parameters in the Mediterranean (i.e. surface temperature, nutrients and food supply) and thus we may be informed about the modifications of juveniles and larval stages of planktonic animals (Morand et al., 1992). Malej and Malej (1992) have established a demographic model for *Pelagia noctiluca* to study long-time series in northern Adriatic sea.

Table 3-Interannual variability of selected planktonic species or groups from oceanic Lebanese waters during 1969-95. Averages are reported as ind./m^3 for zooplankton.

Species, Groups	Average ± S.D.	Variance	Interannual Variation
Diatoms cells/l x1000	111±22	484	20%
Dinoflagellates c/l x1000	4±1.4	2	35%
Tintinnids ind./m3	1316±286	81796	22%
Total Zooplankton	1433±576	331776	40%
Hydromedusae	15±12	144	80%
Siphonophora	26±52	2704	200%
Polychaeta	10±7	49	70%
Cladocera	54±45	2025	83%
Copepoda	954±277	76729	29%
Decapod larvae	33±23.4	547	71%
Pteropoda	19±15	225	79%
Appendicularia	52±29	841	56%
Ophioplutei	4±3	9	75%
Cirrhipedia larvae	22±8	64	36%
Lamellibranch larvae	21±17	289	81%
Ichtyoplankton (total)	25±20	400	80%
Prosobranch larvae	20±20	400	100%
Paracalanus parvus	28±14	196	50%
Clausocalanus spp.	21±11	121	52%
Temora stylifera	22±10	100	45%
Centropages ponticus	11±7	49	63%
Labidocera spp.	3.76±4.02	16	107%
Acartia spp.	20±13	169	68%
Euterpina	18±5	25	28%
Corycaeus spp.	27±10	100	37%
Leucifer	28±24	576	86%
Sagitta enflata	13±7	49	54%
S.friderici	21±14	196	66%
Thalia democratica	9±3	9	33%

Fluctuations of these key species would be an index of pelagic ecosystem modifications. The time-series observed on *Rhizostoma pulmo* and *Copepoda* (1990b) in the Lebanese sector had provided information on the structure and modifications on plankton community.

Time series analysis is an important method to study long term modifications and spatio-temporal variability of pelagic ecosystem. The study of one or more key species may provide information on this natural phenomenon of the sea environment. This was the aim of the "UNEP Jellyfish" project within the frame of MED POL Program for the Mediterranean (Goy et al., 1989a; Lakkis, 1991). In the same idea, Programme Hercules was proposed and adopted by the International Commission of the Scientific

Exploration of the Mediterranean in Monaco to study long-term variability of plankton in relation to global changes in all Mediterranean and its tributaries seas.(Nival et al., 1993).

CONCLUSION

Seasonal variations and interannual fluctuations in abundance are natural phenomenon occurring regularly and permanently in the pelagic ecosystem. The summer crash of phytoplankton and zooplankton biomass and their population density occurs in most of the Mediterranean regions and most of the temperate seas (Morand et al., 1992), while the spring phytoplankton bloom and the zooplankton production are also normal features of the Mediterranean ecosystem.

Over the 26 years of survey, there was a small trend in salinity augmentation (+0.8%) and in temperature (+0.3°C). The species diversity of phytoplankton and zooplankton showed a small increment; the two reasons of that are first: the continual migration of Erythrean and Indo-Pacific species through the Suez Canal into the Levantine Basin; the second reason is the decreasing in population density and in plankton biomass from the area. Although the amplitude of seasonal variations and annual fluctuations increased or decreased according to environmental conditions and species population biology, the pattern of the annual cycle of both the plankton and hydrology remain mostly the same. The majority of permanent species showed normal modifications in their seasonal variations related to hydrographical factors; while the irregular or occasional species showed higher seasonal variations as well as greater interannual fluctuations. It happened that the periodicity and the frequency of these oscillations increase or decrease without any apparent reason. Probably the periodicity within plankton community is not chaotic, there should be certain deterministic reasons responsible for these variations The explanation of these reasons may be partly done by long-time series data analysis and their variability in time and space.

ACKNOWLEDGMENTS

Thanks are due to IOC in the person of Dr. Oliounine for inviting me to participate at the International Workshop on Biological and Chemical Data Management in Hamburg, where this paper was presented. I am indebted to Miss R. Zeidane for her help in identification of the species and in plankton data analyzing.

REFERENCES

Ben Tuvia, A.1985- The impact of the Lessepsian(Suez Canal) fish migration in the eastern Mediterranean ecosystem. *In*: Mediterranean Marine Ecosystem (Editors: Moraitou-Apostolopoulou and Kiortsis):167-175. Plenum Press, N.Y.

Georges,C.J.,V.A. Athanassiou and I. Boulos, 1964- The fishes of the coastal waters of Lebanon Misc.pap.Nat.Sc., AUB, Beirut,4:1-27.

Goy,J.,S.Dallot et P.Morand,1989-Les proliférations de la méduse *Pelagia noctiluca* et les modifications associées de la composition du macroplancton gélatineux, Océanis 15,1:17-23.

Goy,J.,F.Menard,P.Morand and S.Dallot,1996-Evolution de l'écosysètme pélagique depuis le début du siècle en mer Ligure. MAP Technical Reports Series No 97, UNEP:1-22.

Goy,J., S.Lakkis et R.Zeidane,1991- Les Mduses (Cnidaria) des eaux libanaises. Ann.Inst. Oceanogr., Paris, 67, 2:99-128.

Gruvel, A. 1931-Les Etats de Syrie. Richesses marines et fluviales.Soc.Edit.Géogr.Marit.et Colon., Paris, 456pp.

Lakkis,S. 1971 Contribution à l'étude du zooplancton des eaux libanaises Mar. Biology, 11,2:138-148.

Lakkis,S.980-A comparative study of the plankton in the Red Sea and Lebanese waters. Proc. Symp.on the Coastal and Marine Environment of the Red Sea, Gulf of Aden and tropical western Indian Ocean. Vol.II:541-559. UNESCO, RSC, Khartoum.

Lakkis,S.,1990a-Vingt ans d'observation sur le plancton des eaux côtières libanaises. Structure et fluctuations interannuelles. Bull.Inst.Océanogr., No.Special, 7:51-66.

Lakkis,S.1990b-Composition, diversité et successions des copépodes planctoniques des eaux libanaises (Méditerranée orientale). Oceanologica Acta, 13,4:489-5-2.

Lakkis,S.1991-Aggregations of Scyphomedusa Rhizostoma pulmo in the Lebanese coastal waters during the summer of 1986. MAP Technical Reports Series No47,UNEP, Athens:119-127.

Lakkis,S.1996- Monographie Nationale sur la Biodiversité au Liban: Flore et Faune marines et côtières, espèces, populations et diversité génétique.Rapport UNEP (unpublished).

Malej and Malej (1992)- Population dynamics of the jellyfish Pelagia noctiluca (Forsskal,1975). *In*: Proceedings of the 25th EMBS on Marine Eutrophication and Population dynamics, eds. Olsen and Olsen, Fredensborg, Denmark,pp.215-219.

Morand,P.,J.Goy et S.Dallot,1992- Recrutement et fluctuations à long-terme de *Pelagia noctiluca (Cnidaria, Scyphozoa).* Ann.Inst.Océanogr., Paris, 68(1-2):151-158.

Nival,P.,S.Lakkis, J.Goy et S.Dallot,1993- Le Projet Hercule en Méditerranée. Programme inter-Méditerranée adopté par la CIESM, Monaco

Rouch,J.1945-Température et salinité de la mer à Beyrouth. Bull.Inst.Océanogr.Monaco, No.884:6pp.

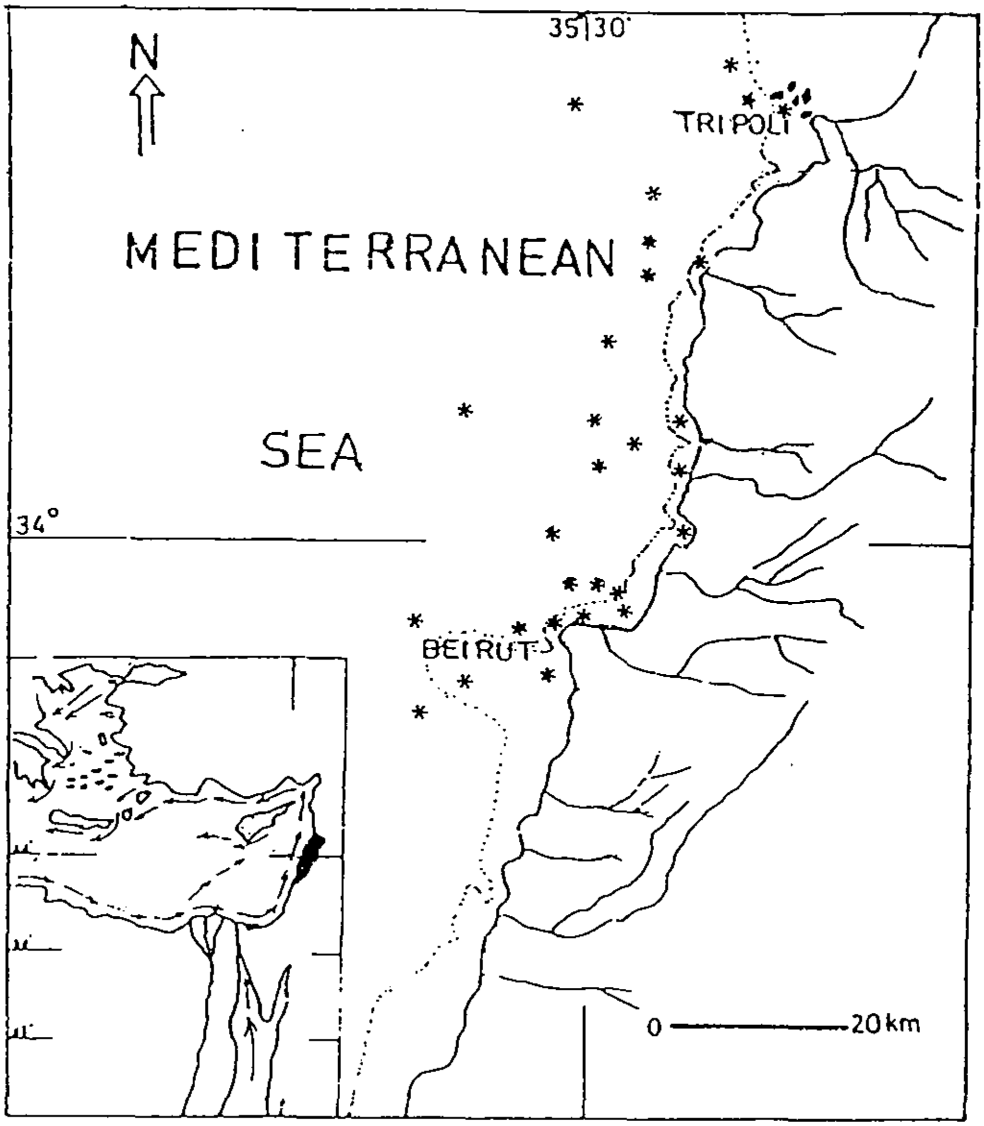


Figure 1: Location of sampling series collected from inshore and offshore the Lebanese coast during 1969-1995. The dotted line shows the limit of the narrow continental shelf. The insert indicates the surface circulation pattern in the eastern Mediterranean and Levantine Basin.

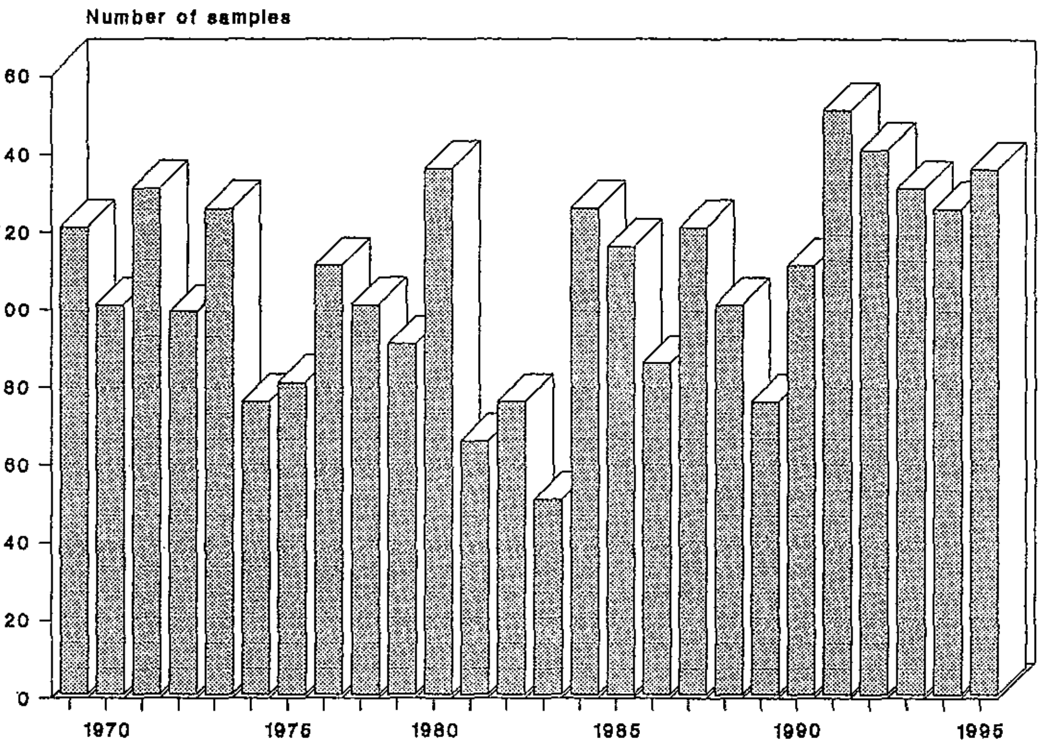


Figure 2: -Histograms showing the distribution and the number of samples collected from Lebanese waters during 1969-95.

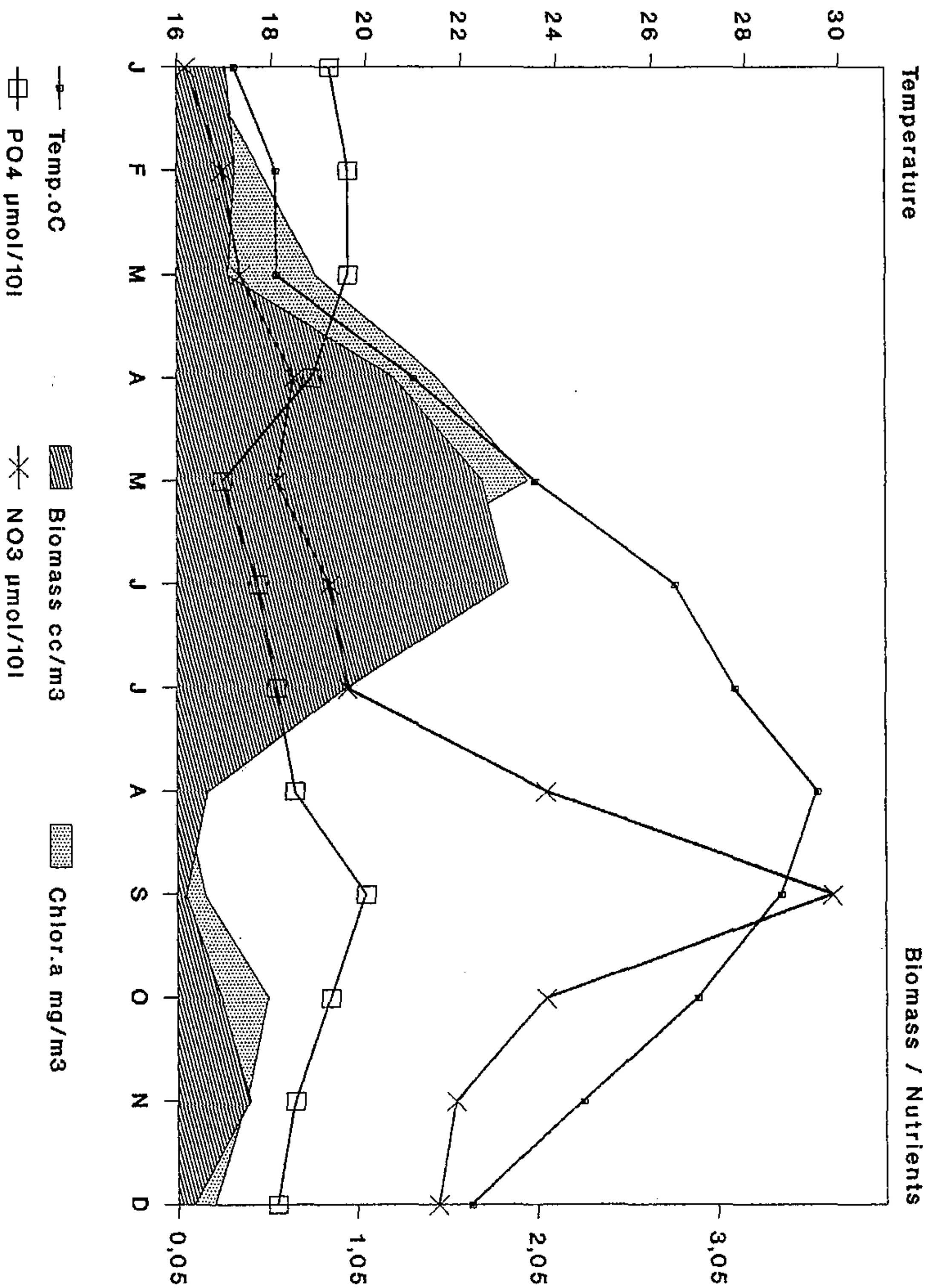


Figure 3: Seasonal variations of some hydrobiological factors from oceanic Lebanese water during 1986-95 (combined data).

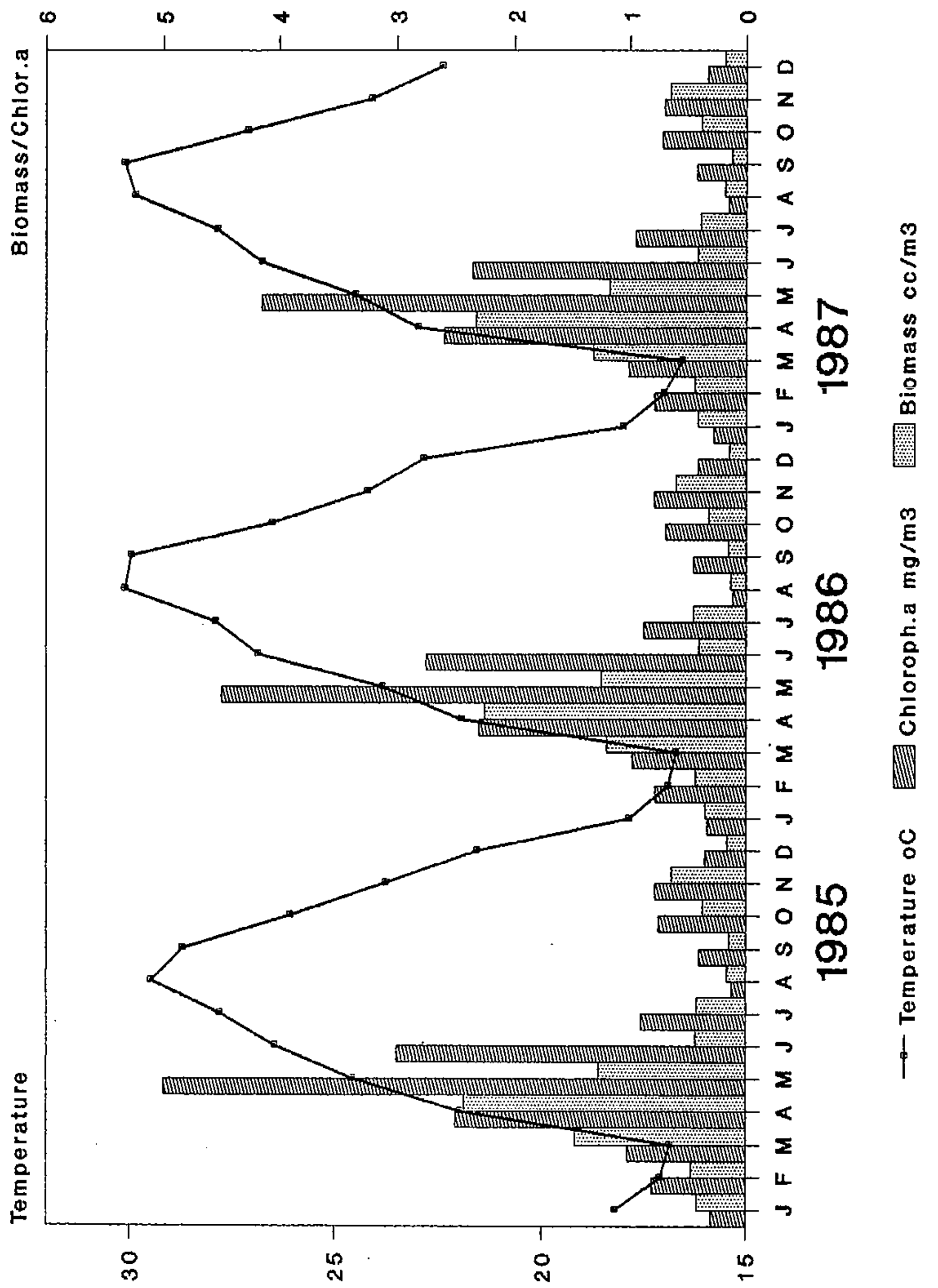


Figure 4: Seasonal variations of Chlorophyll a and plankton biomass in relation to temperature cycle during three consecutive years:1985-87.

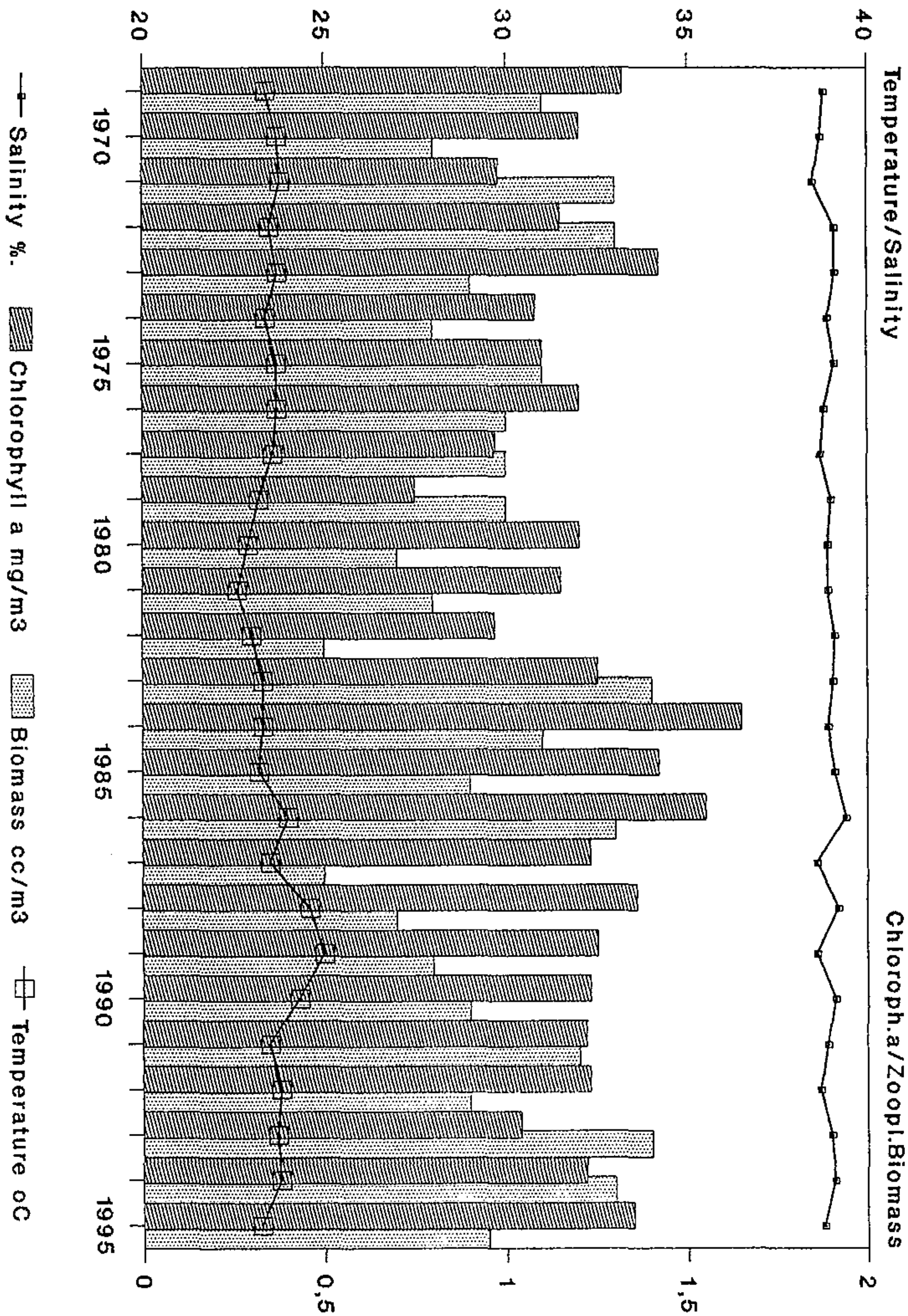


Figure 5: Interannual fluctuations of four hydrobiological parameters: temperature, salinity, chlorophyll a and plankton biomass.

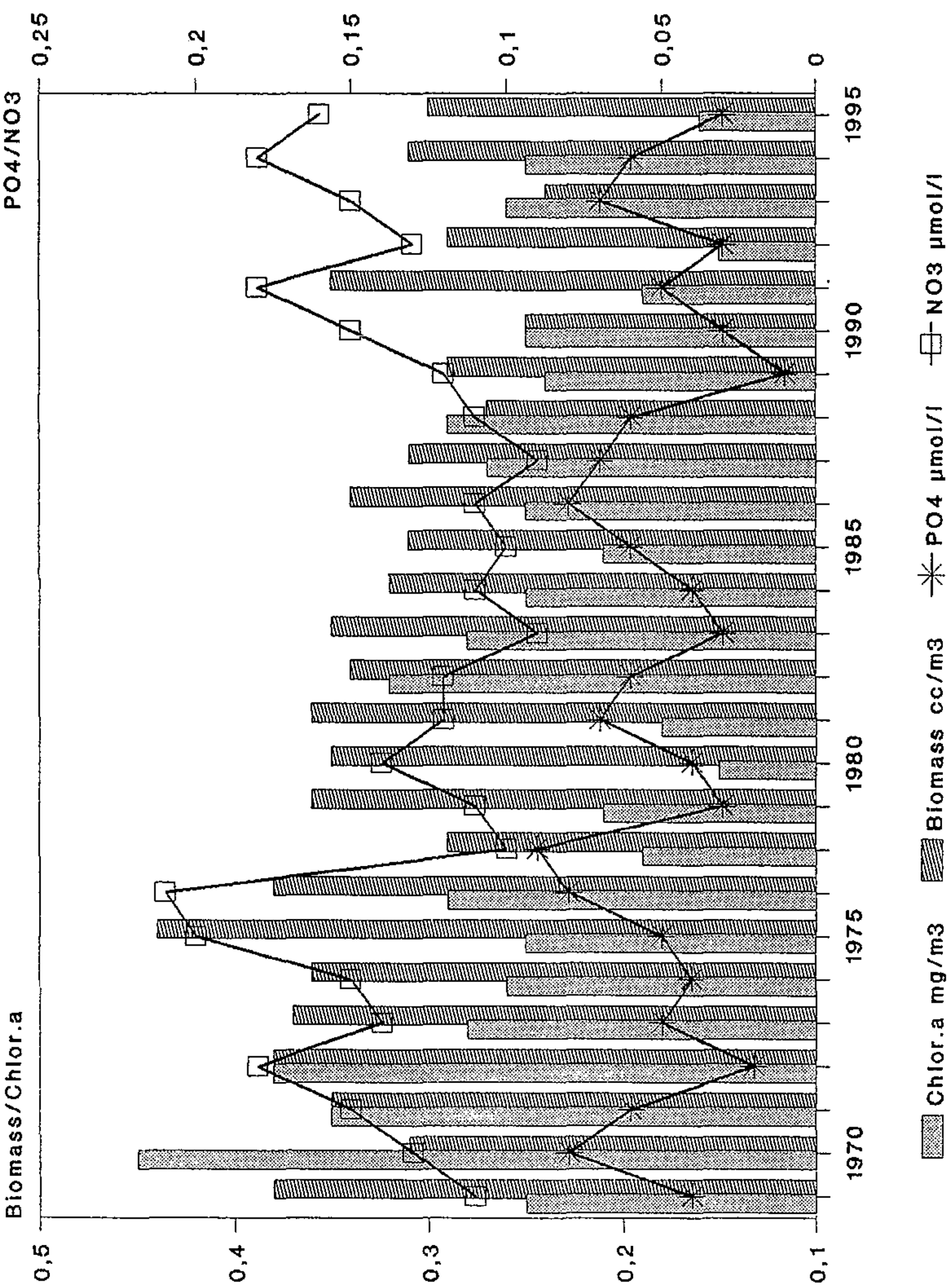


Figure 6: Interannual fluctuations in abundance of plankton biomass and nutrients.

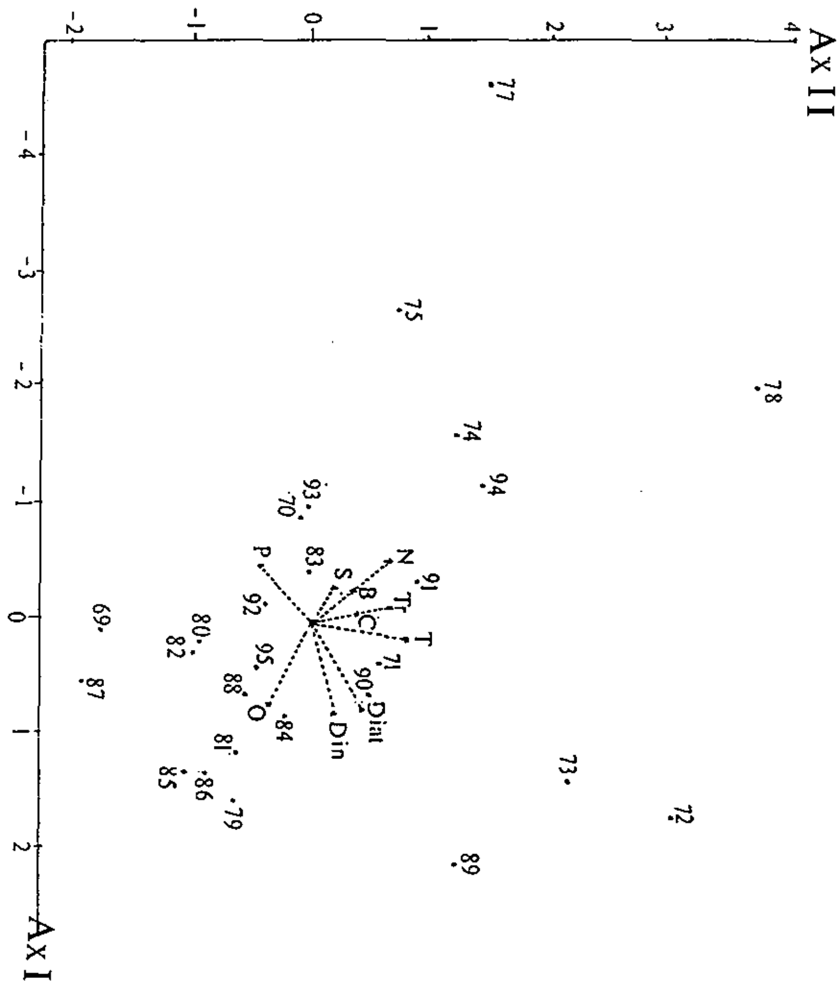


Figure 7: Principal component analysis of ten hydrobiological factors during 26 years of survey. Plots are shown on the two main axes I and II. The label of each years is given by last two figures, hydrobiological factors by the initials for each parameter .

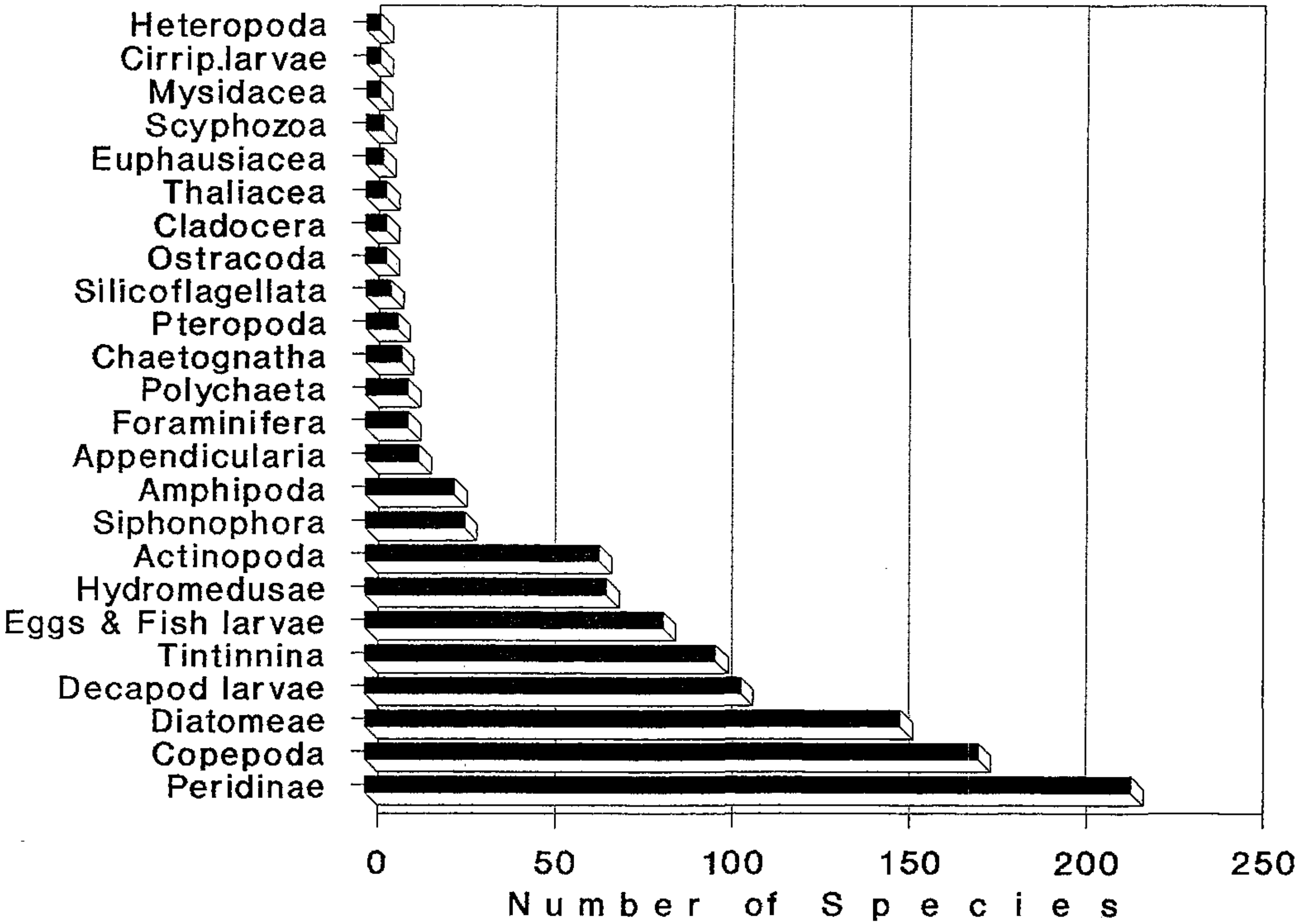


Figure 8: -Histogram showing the compiled number of taxa found for each planktonic group in Lebanese water during 1969-1995.

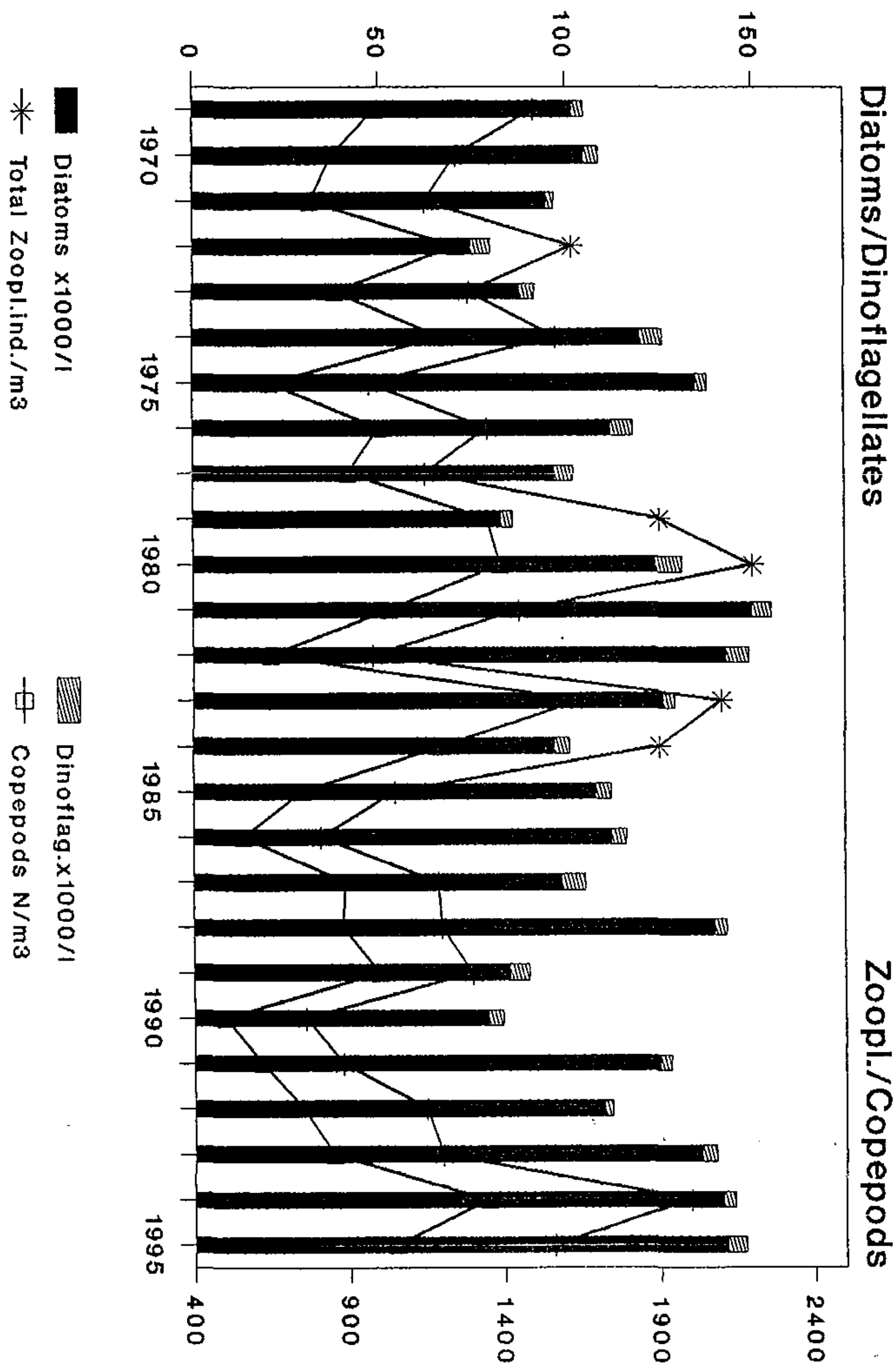


Figure 9: Compared Interannual fluctuations of phytoplankton, zooplankton and planktonic copepods during 1969-1995.

TAXONOMY MANAGER: A SYSTEM FOR THE SPECIFICATION, MANAGEMENT AND USE OF PALEONTOLOGICAL KNOWLEDGE

Jacqueline R. Reich

ABSTRACT

The software system of Taxonomy Manager represents the paleontological knowledge based on real texts in a consistent and non-redundant way. The elements of knowledge are divided into smaller or larger textual information units. These units are qualified, linked and put into specified contexts by dedicated rules and relationships representing together a very fine network of paleontological relevant information. Any set of information units may be extended, especially the taxon knowledge based on several parts of diagnoses. The system is divided into components to represent, to specify, to manage, and to retrieve the paleontological knowledge. The user interfaces of Taxonomy Manager are text-editors, browsing tools - as menus, lists or graphs -, and message or interactive dialogs. The activities of different users of Taxonomy Manager are controlled by an internal system observer to allow access for multiple users. The paleontological knowledge represented by Taxonomy Manager is offered to external software or human users. The mutually interdependent data are stored by Taxonomy Manager in separated databases, which are linked at the same time in a reliable way to maintain the data consistent.

INTRODUCTION

The mini-world of paleontological research is mainly based on three parts: the fossils, the paleontologists and the research literature. The paleontologists try to identify and to classify the fossils. They want to determine the fossils nature and to give sorted structure to the diversity and uniqueness of their phenomena. One part of the discussed information about a fossil sample concerns the morphological description[1] of the specimen; besides of biometry, biostratigraphy, and ecology, which are other information categories of paleontology. In a first step morphological characters or specific features are selected, and in a second step relevant differences comparing these characters are defined. New information or extended knowledge[2] will be represented in research literature, which is on one side an information resource, and on the other side scientific information that has to be criticized.

---

[1] The basic question for any morphological study is *what* to describe and *how* to describe it. For the definition and classification of a new taxon, the description should be as complete as possible. But what is 'completeness'? There are unlimited characters that could be described. On one side the species description should be the most appropriate paraphrase of the significant characters, on the other side it must be comparable with descriptions of other taxa. Furthermore, a practical and meaningful mechanism to express the differences and interdependences between any attributes and their minor components must be developed.

[2] 'knowledge' understood as the justified true belief excluding conjecture.

The literature can be grouped into monographies, catalogues and articles in different journals. Its content is not limited to common text, but contains scientific terms, definitions, information links and references, numerical values, tables, pictures and photos. Knowledge represented in literature is rarely consistent, and detailed relationships between smaller or larger information units are not represented. It is impossible to access information at the same time efficiently and extensively, and to get complete answers for questions.

TAXONOMY MANAGER

The computer based representation of paleontological knowledge is not a trivial task[3]. The system of Taxonomy Manager will play the role of the research literature previously. That means Taxonomy Manager has to represent the paleontological knowledge and make it available to external software tools or human users. The information units of text, scientific terms, definitions, information links and references, numerical values, tables, pictures and photos require a dedicated[4] environment to be correctly specified, managed and used. The tools of Taxonomy Manager enable an intelligent information retrieval and a quick data access. Specified behavior and relationships between different information units increase their semantic expressiveness representing their closer or larger context.

The factoring of Taxonomy Manager into five well distinguished *components* is the basis for the consistent and reliable data representation: the databases of paleontological knowledge, the knowledge specification part, the data-management mechanism, the data-retrieval tools and the internal system observer.

Accordingly, special *interfaces* exist for the knowledge specification part, for several kinds of intelligent information retrieval, and to open Taxonomy Manager to other computer based software-tools communicating by a defined data-access protocol.

The *databases* of paleontological knowledge contain the data of the scientific glossary, the data of the smaller or larger textual description units, the data of the complete taxon definitions, the data of the so-called maturation communities[5], the data of geological age, and the data of any rules[6] or relationships[7] defined for any information unit mentioned above.

---

[3] An environment for the specification and management of paleontological knowledge is obliged to the concepts of *database management systems*, the theory of *distributed system architectures*, and to the philosophy of *knowledge based systems*. A database management system organizes the structure and implementation of data-bases, protects the stored, durable data, and offers controlled data-access to different users. A distributed system architecture represents a system that runs on a collection of machines that do not have shared memory, yet look to its users like a single computer. A knowledge based system includes databases assumed to represent additional information about a domain of discourse or world. Following these fundamental concepts Taxonomy Manager has to meet the requirements concerning data structuring, data independence, control of data redundancy and consistency etc.

[4] To call an element to be 'dedicated' is a widespread expression in computer science. It means an especiallly designed and developed element which very often plays a special role in a wider context. To 'dedicate' an element presupposes a profound analysis of the task to be solved.

[5] Maturation communities are groups of taxa according to their ecological environment of coastal depth zonation and their space of geological age. All taxa show some typical morphological characters still under development during the geological space of time.

The *knowledge specification* part contains dedicated editors to enter new data, each supplied with an appropriate formalism[8] which builds up the default window subdivided by keywords into different data entry sections. Notes of comment are supported at any place close to the specification of data. For any formalism used in Taxonomy Manager an appropriate parser exists to control the syntax and completeness of data specification. Any new specified data will be tested for data consistency either by a testing step directly invoked by the user or by the internal system data-manager. An update of existing databases is only possible after proved data consistency.

The *data-management* refuses inconsistent data. The mechanism itself is not visible to the user. In order to support consistent data, the strongly and mutually interdependent data are stored without - or at least as few as possible - redundancy. Taxonomy Manager communicates by error-message dialogs with the user pointing to the place of an erroneous element causing data inconsistency.

The *data-retrieval* tools support query-answer dialogs and a mouse-driven browser to access stored knowledge. Formulating a query the user will get a list of found features with the path of their location, or the item searched for will be selected and highlighted in the appropriate window. Using a browser the user can freely move through the stored knowledge as far as he or she is allowed to go.

The internal *system observer* enables several users to contact Taxonomy Manager at the same time. It prevents, that read data, which may be used for a longer time interval, will be changed at the same moment from another side. The included history functionality of the system observer allows the user to revert to a previous place, piece of work or search.

As mentioned above, the *knowledge* concerning the definition of taxa, based on morphological description, is subdivided into the strongly interdependent glossar-terms, description-units and proper taxa descriptions.

There are three categories of *glossar-terms*. A glossar-term may be either a normal-term[9], a descriptive-term[10] or an auxiliary-term[11]. The description of any type of glossar-terms consists of a citation-form[12],

---

6 Taxonomy Manager supports following rules concerning one or at least two information units. Some of the rules may be combined. For one information unit: negation, option, weight (= degree of importance). For at least two information units: reason, contradiction, selection, replacement, combination.

7 Taxonomy Manager supports following relationships: precondition, dependency, membership, exclusivity, local arrangement, component, contrast, embrionic gradient, gradient values, old/new, restriction, synonymy. The amount of the reserved words for the relationships and the rules of the above footnote is not too high, and should not require too much time to be learned.

8 A formalism is a special language for the specification of the morphological knowledge and tasks. The underlying complex structures of the formalism and its processes supported by Taxonomy Manager are hidden from the users. The users implicitly refer to the hidden structures extending the specific system-domains according to the tasks they want to be solved. The formalism can adjust its expressiveness according to the authority of specific types of users.

9 exp.: alveole, beam, chamber, test etc.

10 exp.: basal, ovoid, sutural, marginal etc.

11 exp.: likely, mostly, more or less etc.

a definition-text, and optional remarks and any comments. Rules and relationships may be known additionaly[13]. The categories of glossar-terms are represented - individually or all together - in form of a tree to the user. The links of the tree belong to one of the four qualities 'is_a', 'has_a', specified_by', or 'described_by', which name the reference of both linked nodes.

It exists only one type of *description-unit*. The description of description-units consists of a citation-form, a short-name[14], a description-text, and again optional remarks and any comments. Rules and relationships may be known additionaly concerning the description-unit itself, or any glossar-term or description-unit mentioned inside the description-text. The latter one must not contradict to already established rules or relationships. The description-units are represented in an analogous tree form as mentioned above for the glossar-terms.

The description of *taxa* consists of a taxon name[15], an author, the date of the first taxon description, a type-species name, a type-species author, a type-species date, a geological age, a very detailed diagnose part[16], and also optional remarks and any comments. In addition rules and relationships may be defined concerning the taxon itself, or any glossar-term or description-unit mentioned inside the taxon diagnose part, again without contradicting to already established rules or relationships. As the tree representation symbolizes the hierarchy of paleontological taxonomy, the link qualities are reduced to the so-called 'is_a'.

EXTERNAL SOFTWARE TOOLS

A special request formulated by paleontologists is an interface to a software environment, the so-called *Catalogue Text Analyzer* (CTA), which is under development. CTA will be able to analyze, to compare and to classify paleontological text - freely formatted or structured by SGML[17] - based on the knowledge offered by Taxonomy Manager. The paleontologist will be able to control and to correct the result of the analysis, if necessary. Any correction of new specified data is again checked by Taxonomy Manager.

---

12 The citation-form is the basic sring-representation of a one or a multiple words long information-unit. The citation-form may be changed by word inflection according to the linguistic morphological rules of the appropriate language.

13 For normal-terms: rules or relationships may be defined or not. For descriptive-terms: a relationship with a normal-term is obligatory. For auxiliary-terms: relationships may be defined but no rules.

14 The short-name is an abbreviation which can be used instead of the citation-form or the whole description-unit text.

15 The problem of taxa nomenclature is given by the history of paleontology. The taxon name must include the author name. Taxonomy Manager supports for all taxa a synonymy list, and allows information retrieval seperated for genus- and/or species-name elements.

16 The taxon diagnose is subdivided in following parts: shell texture, wall surface texture, chambers, chamber arrangement & growth, chamberlet arrangement & growth, particular chamber feature, foraminal feature, endoskeleton, exoskeleton, supplemental skeleton, reproduction feature, proloculus, embrionic apparatus, nepionic stage, brute chamber, di-/trimorphism.

17 SGML = Standard General Markup Language. It can be used to code a wide variety of texts in a way that is independent of the input and output devices being used.

To point to the relevance of textual information, let me cite the following sentences as they are valid for the paleontological knowledge: "One of the simplest, natural types of information representation is by means of written texts. Data to be processed often does not decompose into independent records. This type of data is characterized by the fact that it can be written down as a long sequence of characters. Such linear sequence is called a *text.*"[18].

To gather related information parts freely distributed over a text, or to understand the real statement of negated sentences, a lot of problems have to be solved.

How far may information units be separated, that they are still valid to be related to by a single or multiple relationship supported by Taxonomy Manager? The information units may be separated into different parts of a sentence, into different sentences of a paragraph or into different paragraphs of a whole text.

The other problem of negated statements demand for an appropriate knowledge specification, that contradiction or similarity of differently formulated statements can be recognized.

CONCLUSION

Taxonomy Manager contributes to the general discussion of the representation of descriptive, more or less intuitive, and subjective knowledge. I am convinced, that this knowledge is worth to be managed by appropriate means. By a reliable management system the nowadays low reputation of this non-experimental knowledge could be corrected.

The aquisition of paleontological morphological knowledge and the process of species classification is very time consuming and expensive. In a similar degree, the implementation and maintenance of such complex knowledge is expensive, too. By the concepts of Taxonomy Manager a wide range of other research domains linked with taxonomical information can access identical knowledge represented in a unique way without the investigation of money and time in the same knowledge acquisition again and again.

The project of Taxonomy Manager is a rather interdisciplinary work. The tasks and problems are formulated by the *paleontologists*, interesting ideas and research approaches are found in *computer-linguistics,* tools and concepts to resolve the tasks are offered by various domains of *computer-science.*

Taxonomy Manager is implemented in *Macintosh Common Lisp* (MCL 2.0 / MCL 3.0), the *Common Lisp Object Oriented System* (CLOS) of the Macintosh. Till the end of 1996 the different parts of Taxonomy Manager and of Catalogue Text Analyzer will be implemented in a final version. All research approaches will be described in a PhD-Thesis book, and in addition the concept and system of Taxonomy Manager will be represented in a graphical *Visual Design Language* (VDL) based on the software development methodology of *Solution-Based Modeling* (SBM).

---

[18] Crochemore/ Rytter 1994, p.1

LITERATURE

Bryan, Martin, 1992. *SGML An Author's Guide to the Standard Generalized Markup Language* Addison-Wesley, ISBN 0-201-17535-5.

Crochemore, Maxime and Rytter Wojciech 1994. *Text Algorithms,* Oxford University Press, ISBN 0-19-50809-0.

Goldstein, Neal and Alger Jeff 1992. *Developing Object-Oriented Software for the Macintosh. Analysis, Design, and Programming.* Addison-Wesley, ISBN 0-201-57065-3.

Salton, Gerald 1989. *Automatic Text Processing* Addison-Wesley.

APPENDIX I

LIST OF PARTICIPANTS

Dr. William M. Balch
Bigelow Laboratory for Ocean Sciences
P.O. Box 475 - McKown Point
W. Boothbay Harbor, ME 04576 USA
Phone: 207-633-9600
Fax: 207-633-9641
E-mail: bbalch@bigelow.org

Dr. Ante Baric
Institute of Oceanography and Fisheries
Mestrovicevo Setaliste 63
PO Box 500
21000 Split Croatia
Phone: 385 21 358 688
Fax: 385 21 358 650
E-mail: baric@jadran.izor.hr

Dr. H. Baumert
Institut für Meeresforschung
Troplowitzstraße 7
Hamburg, Germany
Phone: 41 23 -9221

Dr. Margarita Conkright
NOAA/NESDIS/NODC
OCL - E/OC52
1315 East West Hwy Sta. 4350
Silver Spring, MD 20910 USA
Phone: 301-713-3292 ext 193
Fax: 301-713-3303
E-mail: mconkright@nodc.noaa.gov

Dr. Yves Dandonneau
LODYC CNRS-ORSTOM
Université Pierre et Marie Curie
Tour 14, 2ème étage, 4, place Jussieu
75252 Paris cedex 05 France
Phone: 1 44273 481
Fax: 1 44273 805
E-mail: yd@lodyc.jussieu.fr

Prof. Tommy D. Dickey
ICESS/Department of Geography
Head Ocean Physics Laboratory
Univ of Calif. at Santa Barbara
Santa Barbara, CA 93106-3060 USA
Phone: 805 893-7354
Fax: 805 967-5704
E-mail: tommy@icess.ucsb.edu

Dr. Harry Dooley
International Council for the
Exploration of the Sea (ICES)
Palegade 2-4
1261 Copenhagen Denmark
Phone: 45 33 15 4225
Fax: 45 33 93 4215
E-mail: harry@ices.dk

Dr. Michèle Fichaut
IFREMER-SISMER
BP 70
29280 Plouzane France
Phone: 33 98 22 4643
Fax: 33 98 22 4644
E-mail: Michèle.Fichaut@ifremer.fr

Dr. Chris Garside
Bigelow Laboratory for Ocean Sciences
McKown Point
West Bothbay Harbor, ME 04575 USA
Phone: 207-633-9600
Fax: 207-633-9641
E-mail: cgarside@bigelow.org

Aravind Kolli Ghosh
INDIAN NODC
National Institute of Oceanography
Dona Paula
Goa 403 004 India
Phone: 91 832 226253
Fax: 91-832-2233401360
E-mail: garvind@csnio.ren.nic.in

Dr. Viktor Gouretski
Max-Planck-Institut
Bundesstraße 55
20146 Hamburg Germany
Phone: 3190 3541
E-mail: gouretski@m5.hamburg.bsh.d400.de

Dr. Watson Gregg
NASA/GSFC
Code 902.3
Greenbelt, MD 20771 USA
Phone: 301-286-3464
Fax: 301-286-1775
E-mail: gregg@salmo.gsfc.nasa.gov

Robert C. Groman
Swift House, MS No. 38
Woods Hole Oceanographic Inst.
Woods Hole, MA 02543-1127 USA
Phone: 508 289-2409
Fax: 508 457-2169
E-mail: rgroman@whoi.edu

R. P. Heijman
Expert Center for Taxonomic Identification
University of Amsterdam
Mauritskade 61
1092 AD Amsterdam Netherlands
Phone: 31 (20) 525 7239
Fax: 31 (20) 525 7238
E-mail: RHeijman@ETI.BIO.uva.nl

Wilfried Horn
Bundesamt für Seeschiffahrt und Hydrographie
Deutsches Ozeanographisches Datenzentrum
Bernhard-Nocht-Straße 78
POB 30 12 20
20305 Hamburg
Phone: 49-40-3190-3535
Fax: 49-40-3190-5035 or 5000
E-mail:
wilfried.horn@m5.hamburg.bsh.d400.de

Alex Kozyr
Carbon Dioxide Information Analysis Center
Oak Ridge National Laboratory - Box 2008
Oak Ridge Tennesee 37831 USA
Phone: 615 574-0390
Fax: 615 574-2232
E-mail: cdiac@ornl.gov
or alex@alex.esd.ornl.gov

Prof. Sami Lakkis
Lebanese Univ. and LNCSR
Marine Research Centre
P.O.Box 123
Jounieh Lebanon
Phone: 961-9-918570
Fax: 961-9-943166

Sydney Levitus
NOAA/NESDIS/NODC/OCL - E/OC5
1315 East West Hwy. Sta. 4362
Silver Spring, MD 20910 USA
Phone: 301-713-3294 ext 194
Fax: 301-713-3303
E-mail: slevitus@nodc.noaa.gov

Ms Polly Machin
Britsh Oceanographic Data Centre
Bidston Observatory, Birkenhead, Wirral
Meerseyside L43 7RA United Kingdom
Phone: 44 - 151 - 653 8633
Fax: 44 - 151 - 652 3950
E-mail: pom@pol.ac.uk

Dr. P. N. Makkaveev
Shizshov's Institute of Oceanology - RAS
23, Krasikova Str.
Moscow 117218 Russia
Fax: 1 - 095 - 1245983
Phone (W): 095 - 124-77-42
Phone (H): 095 - 412-43-34
E-mail: 400t@hehem.iozan.zh

Adrian Mallia
Environmental Management Unit
Planning Authority
St. Francis Ravelin CMR 01,
Floriana Malta
Phone: 356 - 2409 76
Phone: 356 - 2290 1603
Fax: 356 - 224846

Prof. You. V. Morozov
Moscow Medical Academy
4th Parkstr. 28-152
105043 Moscow Russia
Phone: 1 095 46527 88

Dr. Khosro Motamedi
Bundesamt für Seeschiffahrt und Hydrographie
Deutsches Ozeanographisches Datenzentrum
Bernhard-Nocht-Straße 78
POB 30 12 20
20305 Hamburg
Phone: 49-40-3190-3531
Fax: 49-40-3190-5035 or 5000
E-mail: motamedi@m5.hamburg.bsh.d400.de

Friedrich Nast
Bundesamt für Seeschiffahrt und Hydrographie
Deutsches Ozeanographisches Datenzentrum
Bernhard-Nocht-Straße 78
POB 30 12 20
20305 Hamburg
Phone: 49-40-3190-3530
Fax: 49-40-3190-5035 or 5000
E-mail: nast@m5.hamburg.bsh.d400.de

Todd O'Brien
NOAA/NESDIS/NODC/OCL - E/OC52
1315 East West Hwy. Sta. 4253
Silver Spring, MD 20910 USA
Phone: 301 713-3291 ext 182
Fax: 301 713-3303
E-mail: tobrien@nodc.noaa.gov

Dr. Iouri Oliounine
UNESCO/IOC
1, Rue Miollis
Paris 75015 France
Phone: 456 839 63
Fax: (33) (1) 405 693 16
E-mail: i.oliounine@unesco.org

Dr. Erdogan Okus
University of Istanbul
Marine Science and Management Institute
Müsküle sk.
Vefa - 34470
Istanbul Turkey
Phone: 212 528 2539
Fax: 212 526 8433

Marek Ostrowski
Institute of Marine Research
Nordeparken 2
Bergen Norway
Phone: 47 - 55 23 68 48
E-mail: marek@imr.no

Dr. Gordon Paterson
Nematode and Polychaete Res. Group
Department of Zoology
The Natural History Museum
Cromwell Road
London SW7 58D United Kingdom
Phone: 44 171 938 9414
Fax: 44 171 938 9158
E-mail: gljp@nhm.ac.uk

Trevor Platt
Bedford Institute of Oceanography
Dartmouth, Nova Scotia B2Y 4A2 Canada
Fax: 902 426 93 88
E-mail: tplatt@ac.dal.ca

Dr. Alain Poisson
Laboratoire de Physique et Chimie Marines
Université P. et M. Curie, case 134
4 Place Jussieu
75252 Paris Cedex 05, France
Phone: 1 - 44 27 4869
E-mail: apoisson@ccr.jussieu.fr

Mrs. Jacqueline Reich
Geologisch-Paläontologisches
Institut der Universität Basel
Bernoullistr. 32
4056 Basel Switzerland
Phone: 0041 - 61/267 36 39
Fax: 0041 - 61/267 36 13
E-mail: reich@ubaclu.unibas.ch

Dr. Philip C. Reid
Sir Alister Hardy Foundation for Ocean Science
The Laboratory, Citadel Hill
Plymouth PL1 2PB UK
Phone: 44 1752 22 1112
Phone: 44 1752 63 3100
Fax: 44 1752 63 3102
E-mail: PCRE@WPO.NERC.AC.UK

Jouko Rissanen
Finnish Environment Institute
P.O.Box 140
FIN-00251 Helsinki Finland
Phone: 357 - 0 - 403 00 357
Fax: 357 - 0 - 403 00 391
E-mail: jouko.rissanen@vyh.fi

Dr. Jury A. Rudjakov
Laboratory of Plankton Ecology & Distribution
P.P. Shirshov Institute of Oceanology
Russian Academy of Sciences
23 Krasikov Str.,
Moscow 117218 Russia
E-mail: grud@ocean.comcp.msk.su

Dr. Viktor Sapozhnikov
Russian Federal Research Institution of
Fisheries & Oceanography
Chief of Marine Ecology Laboratory
17 V. Krasnoselskaya
Moscow, 107140 Russia
Phone: 095264-8392
Fax: 7.095-264-66-85 (Office)
Fax: 7.095-330-39-78 (Home)
E-mail: dscom@sovam.com

Reinhard Schwabe
Bundesamt für Seeschiffahrt und Hydrographie
Deutsches Ozeanographisches Datenzentrum
Bernhard-Nocht-Straße 78
POB 30 12 20
20305 Hamburg
Phone: 49-40-3190-3536
Fax: 49-40-3190-5035 or 5000
E-mail: schwabe@m5.hamburg.bsh.d400.de

Kjell Seglem
Institute of Marine Research
P.O.Box 1870
N-5024 Bergen Norway
Phone: 0047 - 55 2385 00
Fax: 0047 - 55 2385 84
E-mail: kjell@imr.no

Dr. Linda Stathoplos
NOAA/NESDIS/NODC/OCL - E/OC52
1315 East West Hwy Sta. 4239
Silver Spring MD 20910 USA
Phone: 301-713-3292 ext. 180
Fax: 301-713-3303
E-mail: lstathoplos@nodc.noaa.gov

Jan Szaron
Swedish Met. and Hydr. Institute
Oceanographical Laboratory
Byggnad 31
NYA VARVET
SE-42671 Västra Frölunda Sweden
Phone: 46-31-69 65 32
Fax: 46-31-69 04 18
E-mail: jszaron@smhi.se

John Wallace
Irish Marine Data Centre
Marine Institute
80 Arcourt St.
Dublin 2 Ireland
Phone: 353 14 757 100
Fax: 353 14 757 104
E-mail: john.wallace@marine.ir

Dr. Heide-Rose Vatterrott
Frauenhufer Institute for Computer Graphics
Rostock
Loachim-Jungius-Str. 9
D-18059 Rostock Germany
Phone: 49 381 4024 132
Fax: 49 381 4024 199
E-mail: vh@egd.igd.fng.de

Sunhild Wilhelms
Bundesamt für Seeschiffahrt und Hydrographie
Deutsches Ozeanographisches Datenzentrum
Bernhard-Nocht-Straße 78 POB 30 12 20
20305 Hamburg
Phone: 49-40-3190-3537
Fax: 49-40-3190-5035 or 5000
E-mail: wilhelms@m5.hamburg.bsh.d400.de

Dr. Victor Zubarevich
VNIRO
V. Krasnoselskaya, 17
Moscow
Lomonosovskiy pr. 23-89
Moscow 117311 Russia
Phone: 095 - 264 83 92
Fax: 095 - 264 91 87
E-mail: dscom@sovam.com

APPENDIX II

INTERNATIONAL WORKSHOP ON OCEANOGRAPHIC BIOLOGICAL AND CHEMICAL DATA MANAGEMENT

20-23 May 1996
Bundesamt für Seeschiffahrt und Hydrographie, Hamburg, Germany

VENUE

The Workshop will be held in Hamburg, Germany, in the Bundesamt für Seeschiffahrt und Hydrographie, BernhardNocht.Strasse 78, Room: Grosseer Sitzungssaal.

SPONSORS

- National Oceanographic Data Center of the United States of America / World Data Centre A for Oceanography;
- Intergovernmental Oceanographic Commission of UNESCO;
- European Union MAST Programme;
- Bundesamt für Seeschiffahrt und Hydrographie.

GENERAL OBJECTIVES

In order to conduct research studies on bio-geochemical cycles on the earth's ocean-atmosphere system, scientists need access to quality controlled digital oceanographic data sets of chemical and biological parameters. With the advent of new technology in the measurements of biological and chemical parameters, the precision of instruments has increased several fold. Despite the increase in precision, we should not automatically eliminate earlier measurements from historical archives.

The overall goal of the workshop is to improve the quantity and quality of chemical and biological data available to the scientific community. The specific purpose of the Workshop is to provide recommendations to guide management of chemical and biological oceanographic data by the IOC/IODE system.

The topic 'Biological and Chemical Data Management' encompasses many parameters, from bacteria to large mammals and from tracer gases to complex organic compounds. However, in order to focus on data management issues, case studies will be limited to parameters routinely sampled.

SPECIFIC OBJECTIVES

Determine the requirements for managing oceanographic chemical and biological data, for example:

- Identify parameters that the IOC/IODE system can effectively handle;
- Describe minimum metadata requirements that make data useful for future users;
- Identify problems that may limit the usefulness of historical data;
- Identify the users and their requirements.

The following problems have been identified as starting points from which to address the issues of data management:

- Most biological data, and to a lesser extent chemical data, exist only in manuscript form;
- Data exists in diverse formats;
- Availability of metadata is critical;
- Data containing taxonomic information requires taxonomic data management;
- The scientific evaluation of historical biological and chemical data and metadata needs to be supported by governmental and non-governmental agencies.

METHODS

A mixture of scientific and data management presentations will be used to stimulate discussion. Three discussion topics have been selected as a vehicle to meet the Workshop objectives. They are:

- Application of historical biological and chemical measurements;
- International co-operation in data management of biological and chemical oceanographic data;
- Biological and chemical data management - how can the problems be tackled?

The workshop will be introduced by speakers who will give attendees a broad overview of both on-going and planned activities. It is hoped that these speakers, as well as representatives of each of the topical areas, will lead a final panel discussion bringing together suggestions made during the meeting into requirements for future actions.

WORKSHOP PROGRAMME

Day 1: 20 May 1996

8.30 - 9.00	REGISTRATION
9.00 - 9.30	*Opening Remarks* Chairman: S. Levitus BSH President: P. Ehlers

SESSION I: OVERVIEW

Convener: T. Platt (Canada)
Rapporteurs: L. Stathoplos, T. O'Brien (USA)

9.30 - 10.15	*Biogeochemical Cycles:* T. Platt (Canada)
10.30 - 11.15	*Managing the Diversity of Marine Biological Data: Perspectives from a European Union Marine Sciences and Technology (MAST) Project:* G. Paterson (UK)
11.30 - 12.00	Coffee Break
12.00 - 12.45	*Data Archaeology and Rescue of Historical Oceanographic Data: A Report on the IOC/IODE GODAR Project:* S. Levitus (USA)
13.00 - 14.00	Lunch

SESSION II: APPLICATION OF HISTORICAL BIOLOGICAL AND CHEMICAL MEASUREMENTS

Convener: T. Dickey (USA)
Rapporteurs: L. Stathoplos, T. O'Brien (USA)

14.00 - 14.20	*Use of Time Series Chlorophyll Data:* Y. Dandonneau (France)
14.30 - 14.50	*Objective Analysis of Surface Chlorophyll Data in the Northern Hemisphere:* M. Conkright (USA)
15.00 - 15.30	Coffee Break
15.30 - 15.50	*The Spatial and Temporal Variability of the pCO_2 System in the Upper Waters of the Ocean:* P. Makkaveev (Russia)

16.00 - 16.20	*Interannual-to-Decadal Variability of the Temperature-Salinity Structure of the World Ocean:* S. Levitus (USA)
16.30 - 16.50	*Biogeochemical Modeling: Requirements for Biological and Chemical Data:* W. Gregg (USA)
17.00	Working Groups

SESSION III: INTERNATIONAL COOPERATION IN DATA MANAGEMENT OF BIOLOGICAL AND CHEMICAL OCEANOGRAPHIC DATA

Convener: W. Balch (USA)
Rapporteurs: L. Stathoplos, T. O'Brien (USA)

Day 2: 21 May 1996

9.00 - 9.20	*Current and Past ICES Activities in Chemical and Biological Oceanographic Data:* H. Dooley (ICES)
9.30 - 9.50	*Management of Biological and Chemical Data in the JGOFS:* P. Machin (UK)
10.00 - 10.30	Coffee Break
10.30 - 10.50	*Management of Biological, Physical and Chemical Data within the U.S. GLOBEC Project:* R. Groman (USA)
11.00 - 11.20	*The SeaWiFS Mission: Requirements for Biological and Chemical Data:* W. Gregg (USA)
11.30 - 11.50	*Management of the HELCOM BMP Biological and Chemical Data:* J. Rissanen (Helsinki Commission)
12.00 - 13.00	Lunch
13.00 - 13.20	*Management of CO_2 data at CDIAC:* A. Kozyr (USA)
13.30 - 13.50	*CO_2 Parameters: Towards a Database?:* A. Poisson (France)
14.00 - 14.20	*Examples of Availability of Biological and Chemical Data:* F. Nast (Germany)
14.30 - 15.00	Coffee Break

15.00 - 15.20	*Emerging Technologies in Biological, Chemical, Optical, and Physical Sampling in the Oceans:* T. Dickey (USA)
15.30 - 15.50	*The State of the Hydrochemical Database in Russia:* V. Sapoznikov (Russia)
16.00 - 16.20	*The Data Bank Management System of the NODC of Germany:* K. Motamedi, R. Schwabe (Germany)
16.30 - 16.50	*Use of Objects in the Management of Marine Data: Examples of a Database Amenable to Geostatistical Analysis:* M. Ostrowski (Norway)
18.00	Social Event (sponsored by BSH)

SESSION IV: BIOLOGICAL AND CHEMICAL DATA MANAGEMENT

Convener: G. Paterson (UK)
Rapporteurs: L. Stathoplos, T. O'Brien (USA)

Day 3: 22 May 1996

9.00 - 9.20	*QA/QC of Historical Nutrient Data:* C. Garside (USA)
9.30 - 9.50	*Accuracy of Historical Primary Production Measurements:* W. Balch (USA)
10.00 - 10.30	Coffee Break
10.30 - 10.50	*Quality Control of Historical Chlorophyll Data:* M. Conkright (USA)
11.00 - 11.20	*Nutrient data from the South Pacific: Synthesis of the WOCE and Historical Data:* V. Gouretskii (Russia)
11.30 - 11.50	*The Accuracy of Historical Measurements of Plankton Data:* C. Reid (UK)
12.00 - 12.20	*Chemical and Biological Data and Information Management at the Indian NODC:* A. K. Ghosh (India)
12.30 - 13.30	Lunch

13.30 - 13.50	*Metadata Requirements for Plankton Data:* L. Stathoplos (USA)
14.00 - 14.20	*Information Necessary for Zooplankton Biomass Studies and its Possible Implications:* J. Rudjakov (Russia)
14.30 - 15.00	Coffee Break
15.00 - 15.20	*Taxonomic Code Systems and Taxonomic Data Management:* L. Stathoplos (USA)
15.30 - 15.50	*Computer Tools for Mapping Pelagic Biodiversity:* R. P. Heijman (Netherlands)
16.00 - 16.20	*Long Time Series of Hydrobiological and Plankton Data from Lebanese Waters (Eastern Mediterranean):* S. Lakkis (Lebanon)
16.30	Working Groups

Day 4: 23 May 1996

9.30 - 9.50	*Taxonomy Manager: A System for the Specification, Management and Use of Paleontological Knowledge:* J. Reich (Switzerland)
10.00 - 12.20	Round table discussions led by the Workshop Chairman and Conveners of Sessions
12.30 - 13.30	Lunch
13.30 - 15.00	Formulation of conclusions and recommendations

NESDIS 54 Evaluation of Data Reduction and Compositing of the NOAA Global Vegetation Index Product: A Case Study. K. P. Gallo and J. F. Brown, July 1990.

NESDIS 55 Report of the Workshop on Radiometric Calibration of Satellite Sensors of Reflected Solar Radiation, March 27-28, 1990, Camp Springs, MD. Peter Abel (Editor), July 1990.

NESDIS 56 A Noise Level Analysis of Special 10-Spin-Per-Channel VAS Data. Donald W. Hillger, James F. W. Purdom and Debra A. Lubich, February 1991.

NESDIS 57 Water Vapor Imagery Interpretation and Applictions to Weather Analyis and Forecasting. Roger B. Weldon and Susan J. Holmes, April 1991.

NESDIS 58 Evaluating the Design of Satellite Scanning Radiometers for Earth Radiation Budget Measurements with System, Simulations. Part 1: Instantaneous Estimates. Larry Stowe, Philip Ardanuy, Richard Hucek, Peter Abel and Herbert Jacobowitz, October 1991.

NESDIS 59 Interactive Digital Image Display and Analysis System (IDIDAS) User's Guide. Peter J. Celone and William Y. Tseng, October 1991.

NESDIS 60 International Dobson Data Workshop Summary Report. Robert D. Hudson (University of Maryland) and Walter G. Planet, February 1992.

NESDIS 61 Tropical Cyclogenesis in the Western North Pacific. Raymond M. Zehr, July 1992.

NESDIS 62 NOAA Workshop on Climate Scale Operational Precipitation and Water Vapor Products. Ralph Ferraro (Editor), October 1992.

NESDIS 63 A Systematic Satellite Approach for Estimating Central Pressures of Mid-Latitude Oceanic Storms. Frank J. Smigielski and H. Michael Mogil, December 1992.

NESDIS 64 Adjustment of TIROS Operational Vertical Sounder Data to a Vertical View. David Q. Wark, March 1993.

NESDIS 65 A Noise Level Analysis of Special Multiple-Spin VAS Data During Storm-fest. Donald W. Hillger, James F.W. Purdom and Debra A. Molenar, April 1993.

NESDIS 66 Catalogue of Heavy Rainfall Cases of Six Inches or more over the Continental U.S. during 1992. Charles Kadin, April 1993.

NESDIS 67 The Relationship between Water Vapor Plumes and Extreme Rainfall Events during the Summer Season. Wassila Thiao, Roderick A. Scofield and Jacob Robinson, May 1993.

NESDIS 68 AMSU-A Engineering Model Calibration. Tsan Mo, Michael P. Weinreb, Norman C. Grody and David Q. Wark, June 1993.

NESDIS 69 Nonlinearity Corrections for the Thermal Infrared Channels of the Advanced Very High Resolution Radiometer: Assessment and Recommendations. C.R. Nagaraja Rao (Editor), June 1993.

NESDIS 70 Degradation of the Visible and Near-Infrared Channels of the Advanced Very High Resolution Radiometer on the NOAA-9 Spacecraft: Assessment and Recommendations for Corrections. C.R. Nagaraja Rao (Editor), June 1993.

NESDIS 71 Spectral Radiance-Temperature Conversions for Measurements by AVHRR Thermal Channels 3,4,5. Paul A. Davis, August 1993.

NESDIS 72 Summary of the NOAA/NESDIS Workshop on Development of a Global Satellite in Situ Environmental Database. K.P. Gallo and D.A. Hastings (Eds), August 1993.

NESDIS 73 Intercomparison of the Operational Calibration of GOES-7 and METEOSAT-3/4. W. Paul Menzel, Johannes Schmetz, Steve Nieman, Leo Van de Berg, Volker Gaertner, and Timothy J. Schmit, September 1993.

NESDIS 74 Dobson Data Re-Evaluation Handbook. Robert D. Hudson and Walter G. Planet (Eds), October 1993.

NESDIS 75 Detection and Analysis of Fog at Night Using GOES Multispectral. Gary P. Ellrod, February 1994.

NESDIS 76 Tovs Operationa Sounding Upgrades: 1990-1992. A. Reale, M. Chalfant, R. Wagoner, T. Gardner and L. Casey, March 1994.

NESDIS 77 NOAA Polar Satellite Calibration: A System Description. Cecil A. Paris, April 1994.

NESDIS 78 Post-Launch Calibration of the Visible and Near Infrared Channels of the Advanced Very High Resolution Radiometer on NOAA-7,-9, and -11 Spacecraft. C. R. Nagaraja Rao and Jianhua Chen, April 1994.

NESDIS 79 Quality Control and Processing of Historical Oceanographic Nutrient Data. Margarita E. Conkright, Timothy P. Boyer and Sydney Levitus, April 1994.

NESDIS 80 Catalogue of Heavy Rainfall Cases of Six Inches or More Over the Continental U.S. During 1993. Richard Borneman and Charles Kadin, August 1994.

NESDIS 81 Quality Control of Oxygen Temperature and Salinity Data. T.P. Boyer and S. Levitus, May 1994.

NESDIS 82 An Introduction to the GOES I-M Imager and Sounder Instruments and the GVAR Retransmission Format. Raymond J. Komajda (Mitre) and Keith McKenzie, November 1994.

NESDIS 83 Tropical Cyclone Motion Forecasting Using Satellite Water Vapor Imagery. Vernon F. Dvorak and H. Michael Mogil, December 1994.

NESDIS 84 Spurious Semi-Diurnal Variation in the E.R.B.E. Outgoing Longwave Radiation. Chandrasekhara R. Kondragunta and Arnold Gruber, June 1995.

NESDIS 85 Calibration of the Advanced Microwave Sounding Unit-A for NOAA-K. Tsan Mo, June 1995.