NASA/TM—2002–211617

The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS): Current Architecture and Implementation

Giulietta S. Fargion, Science Applications International Corporation, Beltsville, Maryland
Charles R. McClain, Goddard Space Flight Center, Greenbelt, Maryland

P. Jeremy Werdell, Science System and Applications Inc., Lanham, Maryland
Sean W. Bailey, Futuretech Corporation, Greenbelt, Maryland

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

September 2002

Available from:

NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
Price Code: A17

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Price Code: A10

613775
P52

NASA/TM—2002–211617

The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS): Current Architecture and Implementation

G.S. Fargion and C.R. McClain, Editors

P.J. Werdell and S.W. Bailey, Authors

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

September 2002

The NASA STI Program Office ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and mission, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at http://www.sti.nasa.gov/STI-homepage.html
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA Access Help Desk at (301) 621-0134
- Telephone the NASA Access Help Desk at (301) 621-0390
- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

PREFACE

The purpose of this technical report is to provide current documentation of the Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project activities, NASA Research Announcement (NRA) research status, satellite data processing, data product validation, and field calibration. This documentation is necessary to ensure that critical information is related to the scientific community and NASA management. This critical information includes the technical difficulties and challenges of validating and combining ocean color data from an array of independent satellite systems to form consistent and accurate global bio-optical time series products. This technical report is not meant as a substitute for scientific literature. Instead, it will provide a ready and responsive vehicle for the multitude of technical reports issued by an operational project.

Satellite ocean color missions require and abundance of high quality *in situ* measurements for bio-optical and atmospheric algorithm development and post-launch product validation and sensor calibration. To facilitate the assembly of a global data set, the Sea-viewing Wide Field-of-view (SeaWiFS) Project developed the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), a local repository for *in situ* data regularly used in their scientific analyses. The system has since been expanded to contain data sets collected by the SIMBIOS Project, as part of NRA-96-MTPE-04 and NRA-99-OES-99. SeaBASS is a well moderated and documented archive for bio-optical data with a simple, secure mechanism for locating and extracting data based on user inputs. Its holdings are available to the general public with the exception of the most recently collected data sets. Extensive quality assurance protocols, comprehensive data and system documentation, and the continuation of an archive and relational database management system (RDBMS) suitable for bio-optical data all contribute to the continued success of SeaBASS. This document provides and overview of the current operational SeaBASS system.

Giulietta S. Fargion
Charles R. McClain

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Motivation and Philosophy

Experiences with past and present satellite ocean color missions, such as the Coastal Zone Color Scanner (CZCS) and Sea-viewing Wide Field-of-view Sensor (SeaWiFS), demonstrate the need for high quality *in situ* measurements for bio-optical algorithm development and satellite data product validation (Gordon et al. 1983, Evans and Gordon 1994, McClain et al. 1998, Hooker and McClain 2000). The National Aeronautics and Space Administration (NASA) SeaWiFS Project, for example, is tasked with producing normalized water-leaving radiances with an absolute accuracy of 5% (Hooker and Esaias 1993), which requires comparative, globally distributed *in situ* radiometric measurements with accuracy finer than 5%. The advent of additional missions, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Medium Resolution Imaging Spectrometer (MERIS), and the approach of future missions, including the Global Imager (GLI) and the second Polarization and Directionality of the Earth's Reflectances (POLDER-2) instrument, further underline the need for accurate, temporally and geographically diverse samples of oceanographic and atmospheric data.

Historically, the amount of data suitable for algorithm development and satellite validation activities has been limited due to a paucity of simultaneous observations and the difficulty associated with obtaining globally distributed sampling (O'Reilly et al. 1998, Bailey et al. 2000). With regards to the latter, spatial biases are often undesirable for satellite missions with continuous global coverage. Due to their required accuracy, these data are additionally limited by biases introduced by varying measurement and data processing techniques (Hooker and Maritorena 2000, Hooker et al. 2001). As such, global, high quality, *in situ* data sets are invaluable and prerequisite to advance the field of ocean color.

To facilitate the assembly of a global bio-optical data set, the SeaWiFS Project developed the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), a local repository for *in situ* radiometric and phytoplankton pigment data used regularly in their scientific analyses (Hooker et al. 1994). The system has since been expanded to contain oceanographic and atmospheric data sets collected by the NASA Sensor Intercomparison and Merger for Interdisciplinary Biological and Oceanic Studies (SIMBIOS) Project (McClain et al. 2002), as part of NASA Research Announcements (NRA) NRA-96-MTPE-04 and NRA-99-OES-99, which has aided considerably in minimizing spatial bias and maximizing data acquisition rates (McClain and Fargion 1999a and 1999b, Fargion and McClain 2001 and 2002). The SeaWiFS and SIMBIOS Project Offices (SPO) currently share responsibility for the maintenance of SeaBASS, including all design modification and construction.

To develop consistency across multiple data contributors and institutions, the SPO has defined and documented a series of *in situ* data requirements and sampling strategies that ensure that any particular set of measurements will be acceptable for bio-optical and atmospheric correction algorithm development and ocean color sensor validation (Mueller and Austin 1995, Fargion et al. 2001, Mueller et al. 2002a and 2002b). In addition, the SPO has sponsored a series of round-robin activities to establish and advance the state of instrument calibration, protocols, and traceability to radiometric standards (Mueller 1993, Meister et al. 2002). Data prepared using these techniques are suitable for both verifying the radiometric precision and stability of satellite-borne ocean color sensors and validating the algorithms used to relate the radiances to other geophysical parameters.

SeaBASS was designed to be a well-moderated, maintained, and documented archive of bio-optical data, easily accessed by all authorized users, yet secure enough to restrict access when necessary, with a simple mechanism for locating and extracting data based on user inputs. Its success relies upon the application of sufficient quality assurance protocols, comprehensive data and system documentation, and the continuation of an archive and relational database management system (RDBMS) suitable for bio-optical data.

This document provides an overview of the current operational SeaBASS system. The design, protocols, and utilities described in this report supercede all other versions described in previous SeaBASS-related documents (Hooker and Firestone 1994, Firestone et al. 1994, Firestone and Hooker 2001, Werdell et al. 2000a, 2000b, 2002a, and 2002b).

1.2 Synopsis

As of this writing, SeaBASS includes data collected by researcher groups at 43 institutions (Figure 1.1), encompassing over 1,000 individual field campaigns and 30,000 bio-optical data files (Figure 1.2). These data include over 220,000 phytoplankton pigment concentrations, 10,000 continuous depth profiles, and 14,000 spectrophotometric scans (Figure 1.3). Atmospheric data sets are collected using instruments maintained in the SIMBIOS instrument pool*, several Fast-Rotating Shadow-band Radiometers (FRSR) deployed by the Brookhaven National Laboratory under SIMBIOS (Reynolds et al. 2001), and 14 CIMEL sun photometers contributed by the SIMBIOS Project to the NASA Aerosol Robotic Network (AERONET) (Holben et al. 2001) (Figure 1.1). These data include over 13,000 discrete measurements of aerosol optical thickness (AOT) and 111 FRSR campaigns (Figure 1.4). The volume of bio-optical data (hereafter used to describe both the oceanic and atmospheric data archived in SeaBASS) is rapidly increasing, however, as SIMBIOS US Science Team members are contractually obligated to provide data to SeaBASS (McClain and Fargion 1999a and 1999b, Fargion and McClain 2001 and 2002). The volume is expected to increase further as new and upcoming ocean color programs, for example, those for MODIS and MERIS, begin to require and collect validation data.

The full bio-optical data set includes measurements of apparent and inherent optical properties, phytoplankton pigment concentrations, and other related oceanographic and atmospheric data, such as water temperature, salinity, stimulated fluorescence, and AOT (refer to Chapter 3 for additional details on standard data parameters archived in SeaBASS). Data are collected using a number of instrument packages, such as profilers and handheld instruments, and manufacturers on a variety of platforms, including ships and moorings (Fargion and McClain 2002). Field data are collected and prepared, whenever possible, according to the protocols defined by the SPO, as referred to in section 1.1 and Chapter 4 of this document (see also Mueller and Austin 1995, Fargion et al. 2001, Mueller et al. 2002a and 2002b).

In brief, the SeaBASS system consists of: (1) the aforementioned data files, plus relevant documentation and instrument calibration files, (2) a directory tree structure used to house the files, built on a dedicated server at NASA Goddard Space Flight Center, and (3) a RDBMS developed to further catalog, locate, and distribute the files. Metadata from each data file is stored in the RDBMS, and, as such, the system may be queried to compile information about the full bio-optical data set or to locate specific data.

Through the use of online Common Gateway Interface (CGI) forms that interface with the RDBMS, the full bio-optical data set is queriable and available to authorized users via the World Wide Web. As of August 2002, all data collected prior to 31 December 1999 are available to the general public. A username and

* As of August 2002, the SIMBIOS instrument pool consists of 16 MicroTops sun photometers, 2 SIMBAD radiometers, 2 SIMBADA radiometers, 2 PREDE Mark II sun photometers, and 1 micropulse LIDAR.

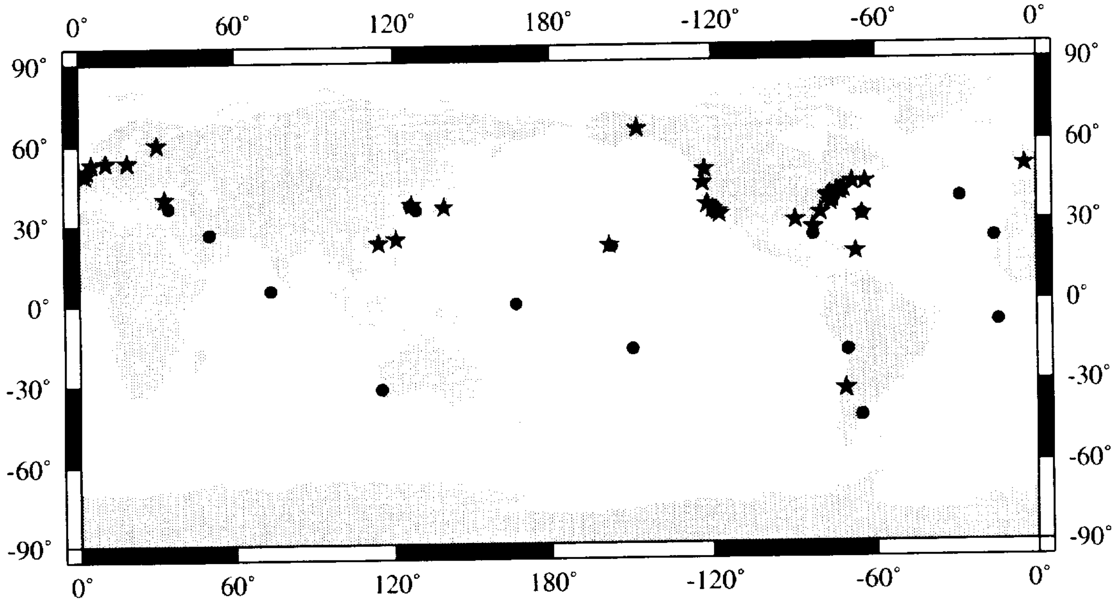


Figure 1.1 A map of the affiliations of SeaBASS data contributors (stars) and of relevant coastal and island CIMEL stations (circles).

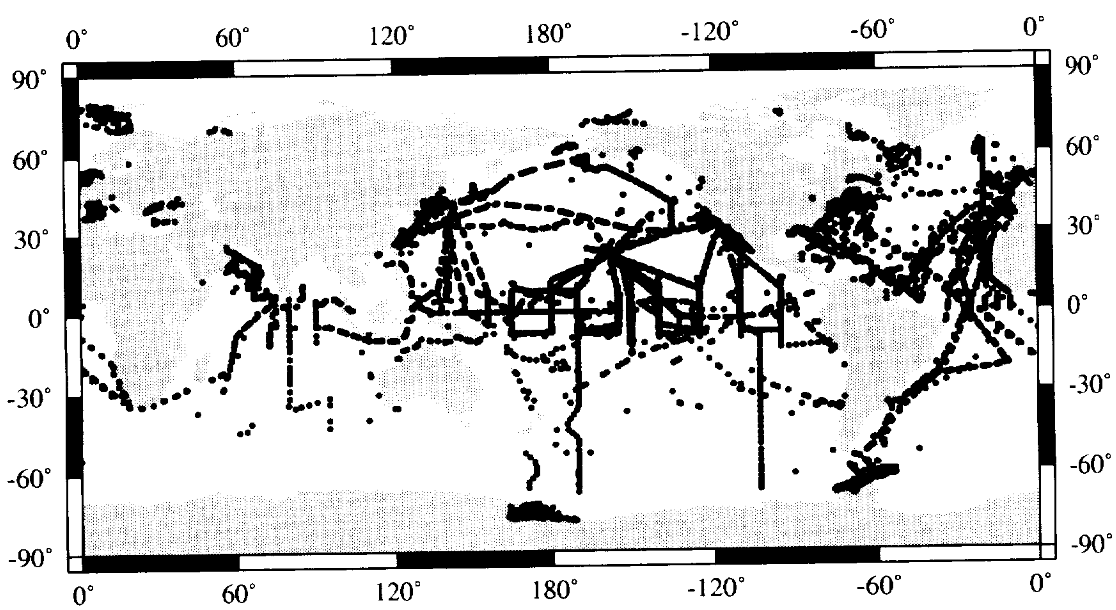


Figure 1.2 A map of all data points in the SeaBASS bio-optical data set.

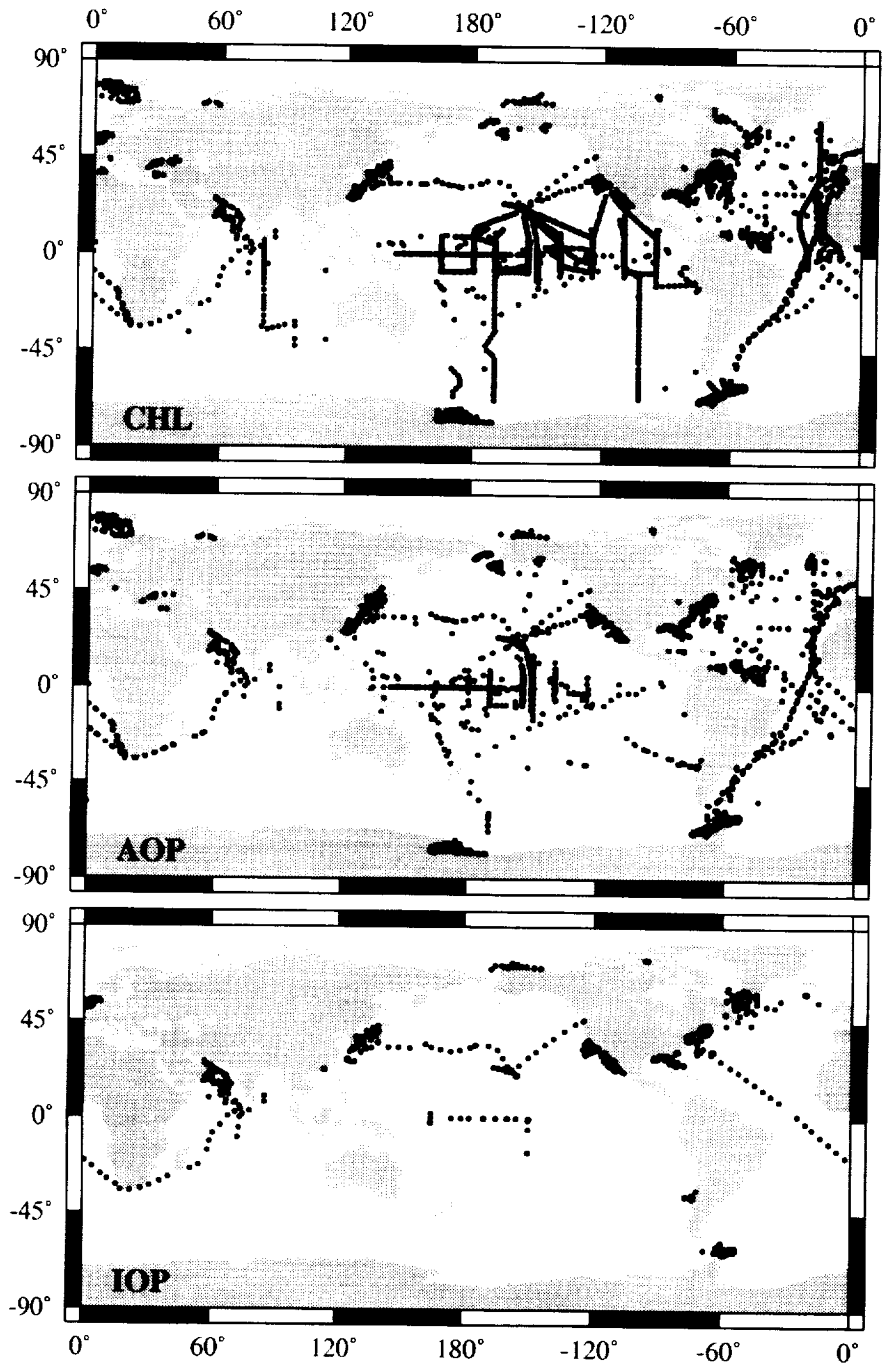


Figure 1.3 Maps of the distributions of common data parameters included in the bio-optical data set. From top to bottom, the maps include (1) chlorophyll *a* concentrations (CHL), (2) measurements of apparent optical properties (AOP), made using above water and profiling radiometers, and (3) measurements of inherent optical properties (IOP), made using absorption meters, backscattering sensors, and spectrophotometers.

password are required to access the remaining restricted data.

Historically, the data archived in SeaBASS have predominantly been used for satellite-data-product validation activities (Figure 1.5) and bio-optical algorithm development (O'Reilly et al. 1998, Bailey et al. 2000, Hooker and McClain 2000, Moore et al. 2001, Maritorena et al. 2002, Schwarz et al. 2002). As the number of viable satellite ocean-color data sets has increased, and the size and range of the community has grown, however, these data have also been used, by the SPO, for example, in support of international protocol workshops, data merger studies, and time series analyses.

In addition, all data collected prior to 31 December 1999 were submitted to the National Oceanic and Atmospheric Administration (NOAA) National Oceanic Data Center (NODC) for inclusion in their national archive. The development of a CD-ROM version of the full public bio-optical and atmospheric data set is planned for Fall 2002.

1.3 Report overview

The second chapter of this report describes the current SeaBASS data submission and access policies, including regulations for data distribution and acknowledgment. In this chapter, eligibility for accessing restricted data is also discussed and the schedule for making data public is outlined. In the third chapter, the format of SeaBASS data files is fully described. The fourth chapter outlines how a contributor verifies the format of their files and submits their data to SeaBASS. This chapter also describes how the SeaBASS Administrator verifies the format of a data file and evaluates the data within. The fifth chapter summarizes the architecture of the SeaBASS archive and its RDBMS. It also illustrates how data are stored and begins to describe how data are made available to the user community. Finally, in the sixth chapter, methods for accessing the bio-optical data set via the World Wide Web are presented in detail. In addition, supplementary online utilities are described in this chapter.

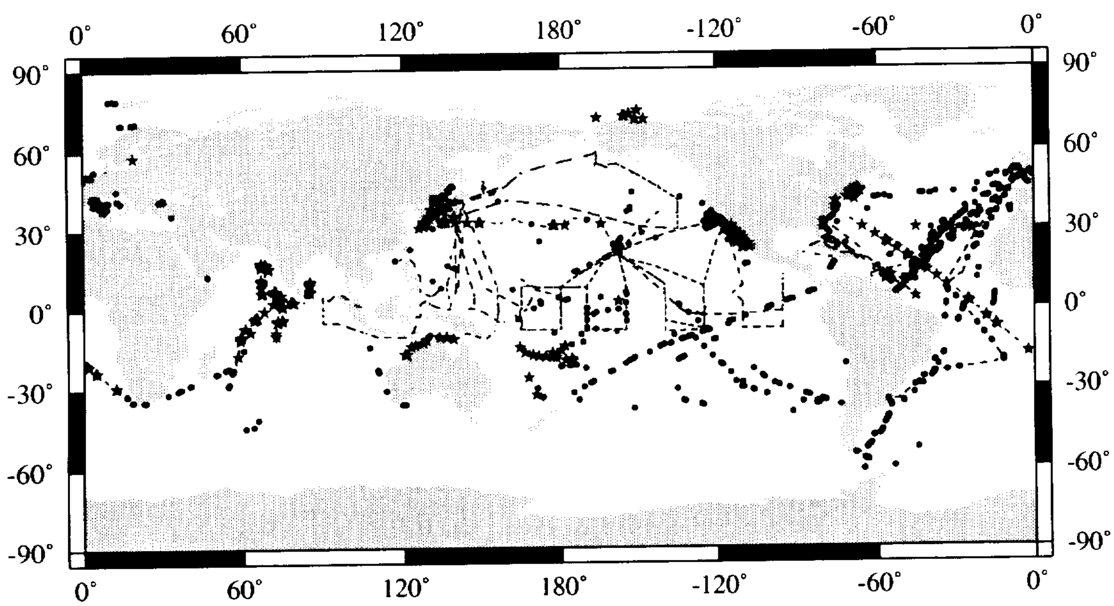


Figure 1.4 A map of all data points in the SeaBASS atmospheric data set. The locations of MicroTops sun photometer data are indicated with stars, SIMBAD radiometer data with circles, and FRSR data with single points.

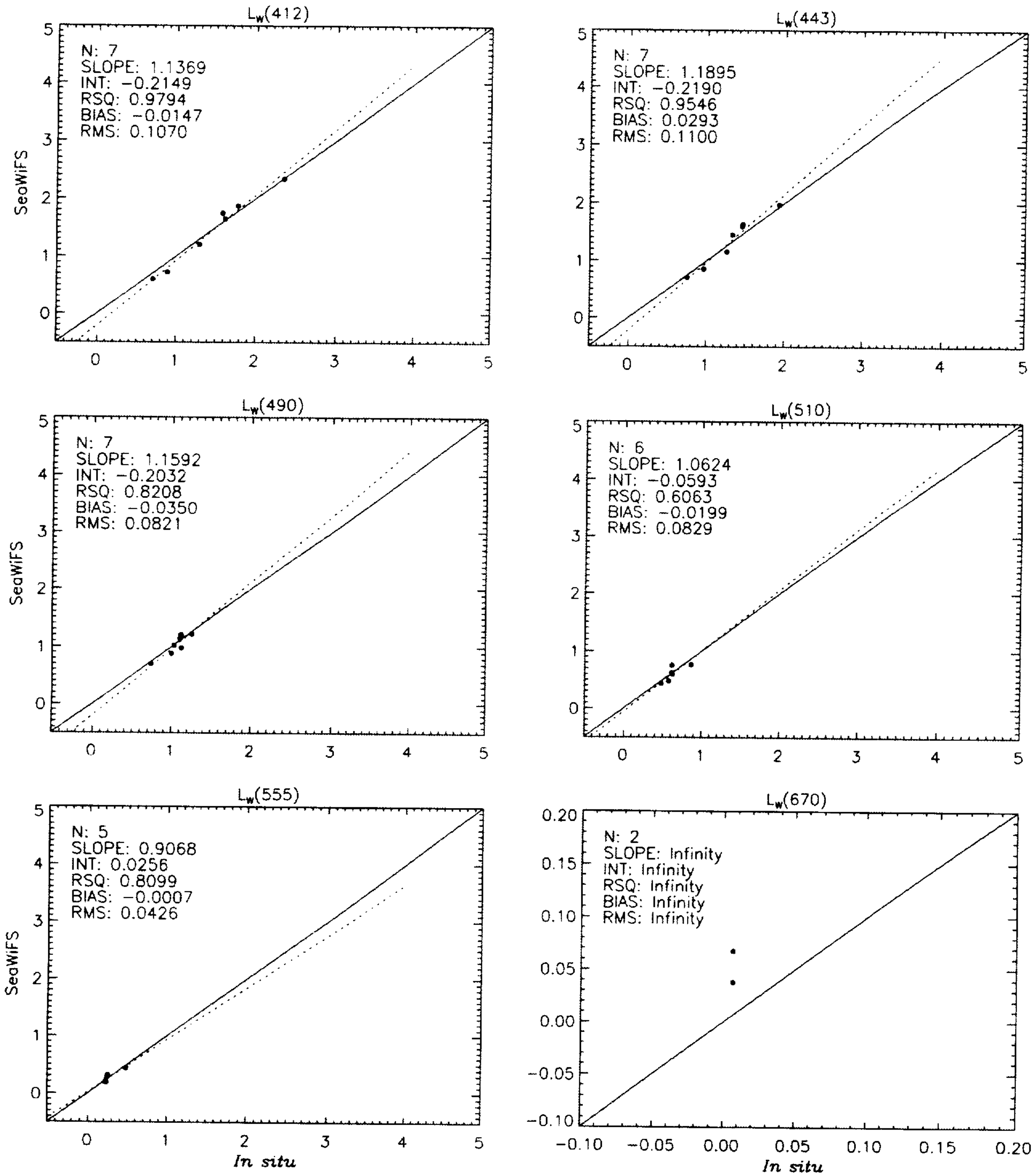


Figure 1.5 A representative example of standard output from the SPO's satellite data product validation activities. Specifically, *in situ* water-leaving radiances collected as part of the eighth Atlantic Meridional Transect (AMT-8) campaign (May - June 1999), are compared with coincident SeaWiFS data values.

Chapter 2

SeaBASS Data and Access Policies

2.1 Introduction

The SeaBASS data policies outlined in this report (and its references) apply to all data collected under the NASA Ocean Biogeochemistry Program at Goddard Space Flight Center for inclusion in the calibration and validation data set archived in SeaBASS. This includes all data submitted to the SPO and biological data collected under the NRA Ocean, Ice, and Climate, released in October 2001. Members of the SIMBIOS Science Team and those receiving funding under the Ocean Biogeochemistry Program must, at a minimum, comply with this data policy, although the SPO encourages a more open policy. Full detail on the scope and extent of all SeaBASS data and access policies is provided in the SIMBIOS NRA-99-OES-99, Appendix B and in Firestone and Hooker (2001). In addition, the details included in this report are available online via the SeaBASS World Wide Web site at:

<http://seabass.gsfc.nasa.gov/seabass_access.html>.

2.2 Data submission

Bio-optical algorithm development is observation limited, and, therefore, rapid turnaround and access to *in situ* data are fundamental to advance the field of ocean color. Principal Investigators (PI's) supported by SIMBIOS NRA-99-OES-09 contracts must submit data within six months of its collection. Data collected under funding from the NASA Ocean Biogeochemistry Program must be submitted within one year. International Science Team members and those involved with international ocean color missions, while not required to do so, are also encouraged to provide data to SeaBASS to foster additional collaboration.

2.3 Data access

Data archived in SeaBASS fall under one of two access categories, *restricted* or *public*. The former includes the full bio-optical data set. The latter excludes the most recently collected data. To protect the publication rights of contributors' data, restricted access is granted only to members of the SIMBIOS Science Team and other individuals approved on a case-by-case basis, as agreed upon by the SPO. The latter includes other NASA-funded researchers, international Science Team members, and regular voluntary contributors to the data archive. Other investigators are able to query for generic information about the restricted data, such as, PI's, data parameters, and temporal and spatial boundaries, but will be referred to the contributors for access to the data itself.

At the three-year anniversary of their collection, data are no longer restricted and are made available to the general public. Data contributors, however, may declare their data available for public access at any time prior to the three-year collection anniversary. As of August 2002, all data collected prior to 31 December 1999 are available to the public. These data have been released to the NODC and are available via their Web site at:

<http://www.nodc.noaa.gov/col/projects/access/seabass.html>.

A courtesy citation, naming the contributor and funding agency, accompanied these data.

The SPO will release additional data every three years at the conclusion of each SIMBIOS NRA. On occasion, special data sets for algorithm development will be made available to the research community without restrictions with the approval of the SIMBIOS Science Team.

2.3.1 Restricted access

The following individuals are eligible for access to the full bio-optical data set: (1) active PI's of the current SIMBIOS Science Team (i.e., those receiving funding under NRA-99-OES-99); (2) members of affiliated NASA-sponsored research programs or Science Teams, for example, MODIS Terra and Aqua; (3) members of international space agencies and associated Science Team members who are regularly contributing ocean color products to the SPO; and (4) regular, voluntary, contributors of *in situ* data, granted access on a case-by-case basis. With regards to the latter, periodic and substantial submission of *in situ* data is required for consideration, and renewal of this status is reviewed annually. An online application for full access to SeaBASS is available at:

<http://seabass.gsfc.nasa.gov/cgi-bin/register_seabass.cgi>.

Applicants are required to provide their affiliation, some general contact information, and the local Internet Protocol (IP) addresses from which they wish to access SeaBASS. Specific usernames and passwords are normally provided by the applicant, but will be generated by the SeaBASS Administrator upon request.

Each PI is eligible for a single account and a limited number of IP addresses from which SeaBASS may be accessed. The number of IP addresses allowed per PI is defined as two for funded SIMBIOS or NASA-sponsored Science Team members, and between three and five per international PI or space agency. The latter is determined at the discretion of the SPO and typically depends on the size of the organization and number of co-PI's named in the SIMBIOS proposal.

Each US and international PI is responsible for applying for access to SeaBASS, distributing their login information to the appropriate members of their staff and co-PI's, and informing the SeaBASS Administrator of all IP address and staffing changes. Likewise, the calibration / validation manager, or appropriate manager, of each space agency is responsible for the above, but also accepts responsibility for providing the SPO a list of all staff that will be accessing SeaBASS. The SeaBASS Administrator reserves the right to monitor all accounts, IP addresses, and passwords, and may implement additional security measures, as necessary.

2.3.2 Public access

Users without restricted access privileges who wish to access the public bio-optical data set may do so freely, but are prompted to log in each time they visit an online site where data may be accessed. The information obtained at this step is used only for statistical purposes and to determine visitors' interests, with the goal of providing better service. For additional information, read the NASA Website Privacy Statement, provided at:

<http://webmaster.gsfc.nasa.gov/policy/gsfc/privacy.html>.

Occasionally, this information will be provided to data contributors, upon request.

2.4 Distribution & acknowledgements

All users who are incorporating SeaBASS data into their research are expected to acknowledge both the data contributor and the funding agency, either the NASA SIMBIOS or SeaWiFS Project, or the NASA Ocean Biogeochemistry Program, as appropriate (Table 2.1). A citation should also be made of the SeaBASS data archive. For restricted data, the contributors have the right to be named as co-authors, and users are encouraged to discuss relevant findings with the contributor early in their research. All users are required to provide the contributors of data they are using a copy of any manuscript prior to initial submission for publication, thus presenting the contributor an opportunity to comment on the paper.

Restricted data accessed from SeaBASS are not to be distributed to unauthorized users. All users and contributors are required to report possible data errors and mislabels to the SeaBASS Administrator. All questions regarding data, however, should be addressed to the original data contributor and not to the Administrator. The SPO will not be held responsible for any data errors or misuse.

Table 2.1 Funded US SIMBIOS (under NRA-96-MTPE-04 and NRA-99-OES-99) and SeaWiFS PI's who have contributed to the SeaBASS bio-optical data set, as of August 2002.

Principal Investigator	Affiliation	Funding Agency
Robert Arnone	Naval Research Laboratory	NASA SIMBIOS
William Balch	Bigelow Laboratory for Ocean Sciences	NASA SIMBIOS
John Brock	United States Geological Survey	NASA SIMBIOS
Chris Brown	NOAA	NASA SIMBIOS
Douglas Capone	University of Maryland	NASA SIMBIOS
Kendall Carder	University of South Florida	NASA SIMBIOS
Francisco Chavez	Monterey Bay Aquarium and Research Institute	NASA SIMBIOS
Glenn Cota	Old Dominion University	NASA SIMBIOS
Tom Dickey	University of California, Santa Barbara	NASA SIMBIOS
David Eslinger	University of Alaska – Fairbanks	NASA SIMBIOS
Piotr Flatau	Scripps Institute of Oceanography, University of California, San Diego	NASA SIMBIOS
Robert Frouin	Scripps Institute of Oceanography, University of California, San Diego	NASA SIMBIOS
Lawrence Harding	University of Maryland	NASA SIMBIOS
Stanford Hooker	NASA Goddard Space Flight Center	NASA SeaWiFS, NASA SIMBIOS
Mark Miller	Brookhaven National Laboratory	NASA SIMBIOS
Greg Mitchell	Scripps Institute of Oceanography, University of California, San Diego	NASA SIMBIOS
Ru Morrison	Woods Hole Oceanographic Institute, Massachusetts Institute of Technology	NASA SIMBIOS
James Mueller	Center for Hydro-Optics and Remote Sensing, San Diego State University	NASA SIMBIOS
Frank Muller-Karger	University of South Florida	NASA SIMBIOS
Norman Nelson	University of California, Santa Barbara	NASA SIMBIOS
John Porter	University of Hawaii	NASA SIMBIOS
David Siegel	University of California, Santa Barbara	NASA SIMBIOS
James Spinhirne	NASA Goddard Space Flight Center	NASA SIMBIOS
Rick Stumpf	NOAA Center for Coastal Monitoring and Assessment	NASA SIMBIOS
Ajit Subramaniam	University of Maryland	NASA SIMBIOS
Ronald Zaneveld	Oregon State University	NASA SIMBIOS

Chapter 3

The SeaBASS Data File Format

3.1 Overview

The design of a SeaBASS data file was conceived based on a need for effortless data access, including online access, while accommodating a variety of computer operating systems. The objective was a simple and logical format that was easily expandable, portable across all computer platforms, web accessible, and manageable using a RDBMS. Accordingly, all SeaBASS data files are currently flat, two-dimensional text files that adhere to the basic American Standard Code for Information Interchange (ASCII) format. The SPO believes that the basic ASCII format most readily satisfies the prerequisite conditions, while also being the most approachable format by the widest user audience.

The format was further designed to be self-describing. Each file is comprised of two parts, a header block followed by a data block (Figure 3.1). The former consists of a series of keywords and values intended to provide descriptive information about the data in the file, for example, the source of the data and its spatial and temporal limits. The latter contains a matrix of data values, similar to the organization of a spreadsheet. File names are not required to follow any specific naming convention, rather, they are left the discretion of the data contributors

3.2 The header block

The keyword-based approach was implemented to enhance automated processing, as the standard vocabulary permits data files to be easily parsed. Each keyword and its argument in the header block occupy one line in the block. The format is:

`/keyword=value,`

where *keyword* is an approved, case-insensitive keyword (Table 3.1) that must begin with a slash (/) and *value* is a string or number which assigns value to the keyword. The exceptions are `/begin_header` and `/end_header`, which do not have input arguments.

Every data file opens with `/begin_header`. The headers may then be listed in any order, so long as the list ends with `/end_header`. Commas separate multiple arguments for a given keyword, for example, in the case of multiple data contributors:

`/investigator=John_Smith,Mary_Jones.`

White space (blanks) and apostrophes ('), however, are invalid characters. Underscores are used to separate words, as indicated in the above example. For those keywords accepting numeric arguments, specific data units are required, as noted in Table 3.1.

In general, the International System of Units (SI) is used, except where traditional usage dictates otherwise. Units are not listed in the header block, with the exception of those keywords relating to time and location, for example, `/start_time` and `/north_latitude`, which require additional notation in the form of:

`/keyword=value[units],`

where *[units]* is set equal to "[GMT]" or "[DEG]", respectively, indicating units of Greenwich Mean Time (GMT) and decimal degrees.

Header keywords are divided into two groups, those required in every data file, and those used to provide additional, optional, information about the data in the file. A value of "NA" ("not available" or "not applicable") may be assigned to any keyword where information cannot be provided. Data files with missing required headers will not be accepted for submission to SeaBASS.

HEADER

```
/begin_header<cr>
.
.
.
/end_header<cr>
```

DATA

```
value(1,1)<delimiter>value(1,2)<delimiter>value(1,N)<cr>
value(2,1)<delimiter>value(2,2)<delimiter>value(2,N)<cr>
.
.
.
value(M,1)<delimiter>value(M,2)<delimiter>value(M,N)<cr>
```

Figure 3.1 The basic SeaBASS data file structure. Each file includes of two parts, a header block, which contains descriptive information about the file and its data, and a data block, which consists of a matrix of geophysical values.

Table 3.1. SeaBASS metadata headers, as of August 2002. (A previous version of this table was originally published in Werdell et al. (2002a)).

Header	Required	Description
/begin_header	Y	The first line of every data file, indicating the beginning of the header block. This header does not have an input argument.
/investigators	Y	The name of the principal investigator, followed by any associate investigators.
/affiliations	Y	A list of affiliations, e.g., university and laboratory, for each investigator.
/contact	Y	An electronic mail address for at least one of the investigators or point of contact for the data file.
/experiment	Y	The name of the long-term research project, e.g., CalCOFI and CARIACO. An entry of 'SIMBIOS' is not permitted.
/cruise	Y	The name of the specific cruise, or subset of the experiment, where the data in the file were collected e.g., cal9802 and car48. An entry of 'SIMBIOS' is not permitted.
/station	N	The name of the station or deployment area where data in the file were collected.
/data_file_name	Y	The current name of the data file.

/documents	Y	A list of cruise reports, station logs, digital images, and other associated documentation which provide additional information about the experiment and cruise. This documentation must accompany the data file at the time of submission.
/calibration_files	Y	A list of supplementary files containing coefficients and techniques used to calibrate the instruments used in data collection. This documentation must accompany the data files at the time of submission.
/data_type	Y	The general collection method, platform, or type of data found in the file. Acceptable values include: **cast** for vertical profiles, e.g., optical packages and CTD; **flow_thru** for continuous data, e.g., shipboard and underway flow through systems; **above_water** for above surface radiometry data, e.g., ASD, SIMBAD, and Satlantic SAS; **sunphoto** for sun photometry data, e.g., MicroTops and PREDE; **mooring** for moored and buoy data; **drifter** for drifter and drogue data; **scan** for discrete hyperspectral measurements; **lidar** for lidar and other active remote-sensing measurements, e.g., MPL; and **pigment** for laboratory measured pigment data, e.g., fluorometry and HPLC.
/data_status	N	The condition, or status, of the data file. The value **preliminary** indicates the data are new and the investigator intends to analyze the data further. The value **update** indicates the data are being resubmitted and informs the SPO that a resubmission will occur in the future. The value **final** indicates the investigator has no intention of revisiting the data set.
/start_date	Y	The earliest date data in the file were collected, in the form YYYYMMDD.
/end_date	Y	The latest date data in the file were collected, in the form YYYYMMDD.
/start_time	Y	The earliest time of day data in the file were collected, in the form HH:MM:SS. Values are required to be in GMT. This header requires a [GMT] trailer, e.g., /start_time=02:45:30[GMT].
/end_time	Y	The latest time of data in the file were collected, in the form HH:MM:SS. Values are required to be in GMT. This header requires a [GMT] trailer, e.g., /end_time=02:56:20[GMT].
/north_latitude	Y	The farthest north data in the file were collected, in decimal degrees. This header requires a [DEG] trailer, e.g., /north_latitude=45.223[DEG]. Coordinates south of the equator are set negative.
/south_latitude	Y	The farthest south data in the file were collected, in decimal degrees. This header requires a [DEG] trailer, e.g., /south_latitude=31.884[DEG]. Coordinates south of the equator are set negative.
/east_longitude	Y	The farthest east data in the file were collected, in decimal degrees. This header requires a [DEG] trailer, e.g., /east_longitude=170.225[DEG]. Coordinates set west of the Prime Meridian are set negative.
/west_longitude	Y	The farthest west data in the file were collected, in decimal degrees. This header requires a [DEG] trailer, e.g., /west_longitude=160.117[DEG]. Coordinates set west of the Prime Meridian are set negative.

`/cloud_percent`	N	Percent cloud cover for the entire sky, e.g., 0 for a cloud-free sky and 100 for a completely overcast sky.
`/measurement_depth`	N	The discrete depth at which data were collected, in meters. This header is required for bottle samples, shipboard flow-through systems, buoys, and moored radiometers.
`/secchi_depth`	N	The secchi depth at the station where the data were collected, in meters.
`/water_depth`	Y	The water depth at the station where the data were collected, in meters.
`/wave_height`	N	The wave height at the station where the data were collected, in meters.
`/wind_speed`	N	The wind speed at the station where the data were collected, in meters per second.
`! COMMENTS`	N	A space for additional comments. Common comments include additional ancillary information about the data file, sea and sky states, difficulties encountered during data collection, methods of data collection, instruments used, and a description of nonstandard SeaBASS field names included in the data file.
`/missing`	Y	The null value used as a numeric placeholder for any missing data in the data file. Each row of data must contain the same number of columns as defined in the `/fields` and `/units` headers. Only one missing value is allowed per file. It is required that this value be non-zero.
`/delimiter`	Y	The delimiter of the columns of data. Accepted delimiters include tab, space, and comma. Only a single delimiter is permitted per data file.
`/fields`	Y	A list of the fields, e.g., CHL, for each column of data included in the data file. Each entry describes the data in a single column, and every column must have an entry.
`/units`	Y	A list of the units, e.g., mg m^{-3}, for each column of data included in the data file. Every value in `/fields` must have an appropriate value listed in this header.
`/end_header`	Y	The final line of the header block, indicating the beginning of the data block. This header does not have an input argument.

Additional notation and information may be incorporated at any time in the header block as comment lines, which begin with an exclamation point (!). Unlike the keyword entries, comment lines are not restricted in format and may include both white spaces and apostrophes. An online description of each header keyword and its expected argument is available on the SeaBASS Web page at:

<http://seabass.gsfc.nasa.gov/seabass_header.html>.

3.3 Standard field names

The `/fields` and `/units` headers identify every column in the data block. Every value in the `/fields` header names a column in the data block, for example, "CHL", and every value in the `/units` header provides units for that column, for example, "mg m^{-3}". Fields and units are listed in the order the data are provided in the data block. Each column is required to have a corresponding `/fields` and `/units` entry; as such, the number of entries in the `/fields` and `/units` headers and the number of columns in the data block must be equal.

To ensure compatibility within the data archive, a standard set of case-insensitive field names and units has been adopted (Table 3.2). While the list is reasonably comprehensive, it does not account for all of the data types one might wish to provide to the archive. Data types

Table 3.2 The SeaBASS standardized parameters, with their appropriate abbreviations, units, and descriptions, as of August 2002. The notation ###.# indicates the parameter is wavelength specific, in nanometers, with the form of, for example, 490.6. The parameter abbreviations shown are mandated by the standard SeaBASS data file format. There are some limitations imposed on the format of the abbreviation because ASCII text is used, as described in the Section 3.3. (A previous version of this table was originally published in Werdell et al. (2002a)).

Abbreviation	Unit Abbreviation	Description
a###.#	m^{-1}	Total absorption coefficient
aaer###.#	m^{-1}	Absorption coefficient of atmospheric aerosols
ad###.#	m^{-1}	Absorption coefficient of detritus
adg###.#	m^{-1}	Absorption coefficient of detritus plus CDOM
ag###.#	m^{-1}	Absorption coefficient of CDOM
agp###.#	m^{-1}	Absorption coefficient of CDOM plus particles
altitude	m	Altitude above sea level
am	unitless	Air mass
angstrom	unitless	Ångström exponent
AOT###.#	unitless	Aerosol optical thickness
ap###.#	m^{-1}	Absorption coefficient of particles
aph###.#	m^{-1}	Absorption coefficient of phytoplankton
a*ph###.#	m^{-1}	Chlorophyll *a*-specific absorption coefficient of phytoplankton
At	degrees C	Air temperature
b###.#	m^{-1}	Total scattering coefficient
bb###.#	m^{-1}	Backscatter coefficient
bincount	none	Number of records averaged into a bin
bnw###.#	m^{-1}	Total scattering coefficient minus the scattering by water
bp###.#	m^{-1}	Particle scattering coefficient
c###.#	m^{-1}	Beam attenuation coefficient
cloud	%	Percent cloud cover
cnw###.#	m^{-1}	Beam attenuation coefficient minus the scattering by water
cond	mmho cm^{-1} •	Conductivity
date	yyyymmdd	Sample date
day	dd	Sample day
depth	m^{-1}	Depth of measurement
Ed###.#	uW cm^{-2} nm^{-1}	Downwelling irradiance
EdGND	Volts	Dark current values for E_d sensor
Elw	uW cm^{-2}	Downwelling irradiance over the infrared spectrum, 3 to 40 μm
Epar	uE cm^{-2} s^{-1} ♦	Profiled PAR
Es###.#	uW cm^{-2} nm^{-1}	Downwelling irradiance above the surface
EsGND	volts	Dark current values for E_s sensor
Esky###.#	uW cm^{-2} nm^{-1}	Downwelling sky irradiance
Esun###.#	uW cm^{-2} nm^{-1}	Downwelling direct normal sun irradiance
Esw	uW cm^{-2}	Downwelling irradiance over the solar spectrum, 0.3 to 3 μm
Eu###.#	uW cm^{-2} nm^{-1}	Upwelling irradiance
EuGND	volts	Dark current values for E_u sensor
F0###.#	uW cm^{-2} nm^{-1}	Extraterrestrial solar irradiance
hour	hh	Sample hour
It	degrees C	Instrument temperature

• The unit "mmho" (the so-called "milli-mho") is the traditional unit used in conductivity studies. In SI units, it is equivalent to the reciprocal of the ohm (or the siemens).

♦ The unit E, for Einstein, is the traditional unit used in PAR studies. In SI units, it is equivalent to 1 mole quanta, or 1 mole photons.

jd	jjj	Sequential day of year
Kd###.#	m^{-1}	Diffuse attenuation coefficient of downwelling irradiance
Kl###.#	m^{-1}	Diffuse attenuation coefficient of upwelling radiance
Knf###.#	m^{-1}	Diffuse attenuation coefficient of natural fluorescence of chlorophyll *a*
Kpar	m^{-1}	Diffuse attenuation coefficient of PAR
Ku###.#	m^{-1}	Diffuse attenuation coefficient of upwelling irradiance
lat	degrees	Sample latitude
lon	degrees	Sample longitude
Lsky###.#	uW cm^{-2} nm^{-1} sr^{-1}	Sky radiance
Lt###.#	uW cm^{-2} nm^{-1} sr^{-1}	Total water radiance
Lu###.#	uW cm^{-2} nm^{-1} sr^{-1}	Upwelling radiance
LuGND	volts	Dark current values for L_u sensor
Lw###.#	uW cm^{-2} nm^{-1} sr^{-1}	Water-leaving radiance
Lwn###.#	uW cm^{-2} nm^{-1} sr^{-1}	Normalized water-leaving radiance ($L_{WN} = L_W F_0 / E_s$)
minute	mn	Sample minute
month	mo	Sample month
natf	nE m^{-2} sr^{-1} s^{-1}	Natural fluorescence of chlorophyll *a*
nrb	photoelectrons μs^{-1} $shot^{-1}$	Normalized relative backscatter
Oz	Dobson units	Column ozone
PAR	uE cm^{-2} s^{-1}	PAR measured at the sea surface
pitch	degrees	Instrument pitch
PP	mg C/mg chla/h[*]	Primary productivity
pressure	dbar	Water pressure
pressure_atm	mbar	Atmospheric pressure
Q###.#	sr	E_u/L_u (equal to π in diffuse water)
quality	none	Analyst-defined data quality flag
R###.#	unitless	Irradiance reflectance ($R = E_u/E_d$)
RelAz	degrees	Sensor azimuth angle relative to the solar plane
Rl###.#	sr^{-1}	Radiance reflectance ($R_L = L_u / E_d$)
roll	degrees	Instrument roll
Rpi###.#	unitless	Radiance reflectance with π
Rrs###.#	sr^{-1}	Remote-sensing reflectance ($R_{rs} = L_W / E_s$)
sal	PSU	Salinity
sample	none	Sample number
SAZ	degrees	Solar azimuth angle
second	ss	Sample second
SenZ	degrees	Sensor zenith angle
sigmaT	kg m^{-3}	Density – 1000 kg m^{-3}
sigma_theta	kg m^{-3}	Potential density – 1000 kg m^{-3}
SN	none	Instrument serial number
SPM	g L^{-1}	Total suspended particulate material
SST	degrees C	Sea surface temperature
station	none	Sample station
stimf	volts	Stimulated fluorescence of chlorophyll *a*
SZ	m	Secchi disk depth
SZA	degrees	Solar zenith angle
tilt	degrees	Instrument tilt
time	hh:mm:ss	Sample time
trans	%	Percent transmission
volfilt	L	Volume filtered
waveheight	m	Wave height

[*] This parameter has the units of "milligrams of carbon per milligrams of chlorophyll *a* per hour". The individual units are separated with the solids (/), instead of the customary reciprocals, to avoid confusion as to how it is to be formatted.

wavelength	nm	Wavelength of measurement
windspeed	m s^{-1}	Wind speed
Wt	degreesC	Water temperature
Wvp	mm	Water vapor
year	yyyy	Sample year
Pigments:		
Allo	mg m^{-3}	HPLC alloxanthin
Anth	mg m^{-3}	HPLC antheraxanthin
Asta	mg m^{-3}	HPLC astaxanthin
Beta-beta-Car	mg m^{-3}	HPLC ββ-carotene (β-carotene)
Beta-epi-Car	mg m^{-3}	HPLC βε-carotene (α-carotene)
Beta-psi-Car	mg m^{-3}	HPLC βψ -carotene (γ-carotene)
But-fuco	mg m^{-3}	HPLC 19′-butaonoyloxyfucoxanthin
Cantha	mg m^{-3}	HPLC canthaxanthin
CHL	mg m^{-3}	Fluorometrically or spectrophotometrically-derived chlorophyll *a*
Chl_a	mg m^{-3}	HPLC chlorophyll *a*
Chl_a_allom	mg m^{-3}	HPLC chlorophyll *a* allomers
Chl_a_prime	mg m^{-3}	HPLC chlorophyll *a* epimer
Chl_b	mg m^{-3}	HPLC chlorophyll *b*
Chl_c	mg m^{-3}	HPLC chlorophyll *c*
Chl_c1c2	mg m^{-3}	HPLC chlorophyll c_1 and c_2
Chl_c3	mg m^{-3}	HPLC chlorophyll c_3
Chlide_a	mg m^{-3}	HPLC chlorophyllide *a*
Chlide_b	mg m^{-3}	HPLC chlorophyllide *b*
Croco	mg m^{-3}	HPLC crocoxanthin
Diadchr	mg m^{-3}	HPLC diadinochrome
Diadino	mg m^{-3}	HPLC diadinoxanthin
Diato	mg m^{-3}	HPLC diatoxanthin
Dino	mg m^{-3}	HPLC dinoxanthin
DV_Chl_a	mg m^{-3}	HPLC divinyl chlorophyll *a*
DV_Chl_b	mg m^{-3}	HPLC divinyl chlorophyll *b*
Echin	mg m^{-3}	HPLC echinenone
epi-epi-Car	mg m^{-3}	HPLC εε-carotene (ε-carotene)
Et-8-carot	mg m^{-3}	HPLC ethyl-apo-8′-carotene
Et-chlide_a	mg m^{-3}	HPLC ethyl chlorophyllide *a*
Et-chlide_b	mg m^{-3}	HPLC ethyl chlorophyllide *b*
Fuco	mg m^{-3}	HPLC fucoxanthin
Hex-fuco	mg m^{-3}	HPLC 19′-hexanoyloxyfucoxanthin
Lut	mg m^{-3}	HPLC lutein
Lyco	mg m^{-3}	HPLC lycopene
Me-chlide_a	mg m^{-3}	HPLC methyl chlorophyllide *a*
Me-chlide_b	mg m^{-3}	HPLC methyl chlorophyllide *b*
MV_Chl_a	mg m^{-3}	HPLC monovinyl chlorophyll *a*
MV_Chl_b	mg m^{-3}	HPLC monovinyl chlorophyll *b*
Mg_DVP	mg m^{-3}	HPLC Mg 2,4-divinyl phaeoporphyrin a_5 monomethyl ester
Monado	mg m^{-3}	HPLC monadoxanthin
Neo	mg m^{-3}	HPLC neoxanthin
P-457	mg m^{-3}	HPLC P-457
Perid	mg m^{-3}	HPLC peridinin
PHAEO	mg m^{-3}	Total phaeopigment concentration
Phide_a	mg m^{-3}	HPLC phaeophorbide *a*
Phide_b	mg m^{-3}	HPLC phaeophorbide *b*
Phide_c	mg m^{-3}	HPLC phaeophorbide *c*
Phythl-chl_c	mg m^{-3}	HPLC phytylated chlorophyll *c*

Phytin_a	mg m-3	HPLC phaeophytin *a*
Phytin_b	mg m-3	HPLC phaeophytin *b*
Phytin_c	mg m-3	HPLC phaeophytin *c*
Pras	mg m-3	HPLC prasinoxanthin
Pyrophytin_a	mg m-3	HPLC pyrophaeophytin *a*
Pyrophytin_b	mg m-3	HPLC pyrophaeophytin *b*
Pyrophytin_c	mg m-3	HPLC pyropheophytin *c*
Siphn	mg m-3	HPLC siphonein
Siphx	mg m-3	HPLC siphonaxanthin
Tot_Chl_a	mg m-3	HPLC divinyl chlorophyll *a* plus monovinyl chlorophyll *a* plus chlorophyllide *a* plus chlorophyll *a* allomers plus chlorophyll *a* epimer
Tpg	mg m-3	Total pigment concentration
Vauch	mg m-3	HPLC vaucheriaxanthin-ester
Viola	mg m-3	HPLC violaxanthin
Zea	mg m-3	HPLC zeaxanthin

that do not fall under one of the predefined standard field names may be included in a submitted data file. The contributor, however, is asked to define each non-standard data type as a comment in the header block. If there are frequent queries for non-standard data types, then the new field names and associated units will be included in the standard list. An online, regularly updated, version of the standard field names list is available via the SeaBASS Web page at:

<http://seabass.gsfc.nasa.gov/cgi-bin/stdfields.cgi>.

Note that data values are required to be in meaningful geophysical units (e.g., providing voltages with conversion coefficients is unacceptable). Note also that there are some limitations and restrictions imposed on the format of the unit abbreviations because ASCII text is used. For example, although "per meter" is classically represented as "m^{-1}", the format to input would be "m-1" or "1/m", the latter being the reciprocal of the unit. In addition, the letter "u" is used in the unit abbreviations (e.g., "uW cm-2 nm-1") instead of the Greek letter μ, again, because Greek letters cannot be used in an ASCII file.

Whenever possible, the standard field names and units assigned to each data parameter were specified based on traditional oceanographic and atmospheric abbreviations as listed in current literature. The standard names assigned to high performance liquid chromatography (HPLC) derived pigments are based on abbreviations defined by the Scientific Committee on Oceanic Research (SCOR) Working Group 78, as listed in Jeffrey et al. (1997).

3.4 The data block

In the data block, values are provided as a matrix (i.e., in columns), similar to a spreadsheet. Spaces, tabs, or commas may be used to delimit each column, provided a single, consistent delimiter is used throughout the data file and its appropriate value is listed in the `/delimiter` header. Each row of data is terminated with a carriage return. Data that are missing, bad, or unavailable are replaced with a numeric blank, for example, "-999", whose value is non-zero and a non-observable value for the particular data type. Only a single numeric blank may be used per file and this value must be listed in the `/missing` header. Exponential notation is acceptable, in the form: `<n>e<x>`, where <*n*> is a numeric entry and <*x*> is the exponential value, for example, "3.33e-4". Latitude and longitude are provided in decimal degrees, with coordinates north of the equator and east of the Prime Meridian set positive. Water depth and pressure values are listed as increasing positive with increasing depth.

Each file should be segmented into a logical grouping, such as, by station, date, or parameter, or based on the measurement or instrument type.

For example, all parameters collected in a single depth profile or spectrophotometric analysis of a discrete water sample should be incorporated into a single data file (Figures 3.2 and 3.3). On the contrary, multiple discrete measurements collected at several stations may be incorporated into a single file (Figure 3.4), provided the appropriate metadata (e.g., date, time, latitude, longitude, depth, and station) are also included as data columns.

```
/begin_header
/investigators=John_Smith,Mary_Johnson
/affiliations=MBARI,State_University
/contact=jsmith@mbari.org,mary@state.edu
/experiment=TAO_Moorings
/cruise=gp1-02-ka
/station=341
/data_file_name=cast_example.txt
/documents=TAO_README.txt
/calibration_files=ocp14a.cal
/data_type=cast
/data_status=preliminary
/start_date=19971215
/end_date=19971215
/start_time=21:15:39[GMT]
/end_time=21:19:30[GMT]
/north_latitude=-0.016[DEG]
/south_latitude=-0.016[DEG]
/east_longitude=-170.02[DEG]
/west_longitude=-170.02[DEG]
/cloud_percent=10.0
/measurement_depth=NA
/secchi_depth=NA
/water_depth=2100
/wave_height=0.5
/wind_speed=5
!
! Downcast better than upcast.
!
/missing=-999
/delimiter=space
/fields=depth,Lu412.2,Lu443.4,Lu489.7,Lu510.0
/units=m,uW/cm^2/nm/sr,uW/cm^2/nm/sr,uW/cm^2/nm/sr,uW/cm^2/nm/sr
/end_header@
0.0 5.856900 5.989949 5.787405 4.898884 4.280903
1.0 1.244184 1.066594 0.852400 0.461248 0.177923
2.0 1.299710 1.113997 0.884608 0.457049 0.159074
3.0 1.298214 1.113140 0.886502 0.455522 0.155225
...
```

Figure 3.2 A representative example of a SeaBASS cast data file.

```
/begin_header
/investigators=Greg_Smith,Mary_Johnson
/affiliations=UCSD_SIO
/contact=gsmith@ucsd.edu
/experiment=CALCOFI
/cruise=CAL9910
/station=93.28
/data_file_name=scan_example.txt
/documents=README.cal9910
/calibration_files=cal9910_scanlog.txt
/data_type=scan
/data_status=final
/start_date=19991003
/end_date=19991003
/start_time=19:20:00[GMT]
/end_time=19:20:00[GMT]
/north_latitude=34.392[DEG]
/south_latitude=34.392[DEG]
/east_longitude=-124.327[DEG]
/west_longitude=-124.327[DEG]
/cloud_percent=20
/measurement_depth=10
/secchi_depth=22.1
/water_depth=230
/wave_height=1
/wind_speed=3.7
!
! Method of estimating particulate absorption:
! Mitchell, B.G., Ocean Optics X, p.137-148, 1990
!
! The spectral range is 400 nm - 750 nm with 2 nm step
!
/missing=-999
/delimiter=space
/fields=wavelength,ad,ap,ag
/units=nm,1/m,1/m,1/m
/end_header@
400 0.00533 0.01217 0.01162
402 0.00528 0.01246 0.02508
404 0.00525 0.01282 0.02931
406 0.00523 0.01323 0.02713
408 0.0052  0.01369 0.02098
410 0.00516 0.01416 0.01852
412 0.0051  0.0146  0.00693
414 0.00502 0.01502 0.01381
416 0.00492 0.01539 0.00973
418 0.00481 0.01572 0.0126
420 0.00469 0.01603 0.00691
...
```

Figure 3.3 A representative example of a SeaBASS scan data file.

```
/begin_header
/investigators=John_Smith,Mary_Johnson
/affiliations=Goddard_Space_Flight_Center,State_University
/contact=jsmith@simbios.gsfc.nasa.gov,mary@state.edu
/experiment=AMT
/cruise=AMT07
/station=NA
/data_file_name=chl_example.txt
/documents=A7OPSLOG.TXT
/calibration_files=A7CHL.cal
/data_type=pigment
/data_status=preliminary
/start_date=19981016
/end_date=19981016
/start_time=12:11:08[GMT]
/end_time=15:25:45[GMT]
/north_latitude=36.1234[DEG]
/south_latitude=31.8823[DEG]
/east_longitude=-51.2363[DEG]
/west_longitude=-55.1125[DEG]
/cloud_percent=NA
/measurement_depth=NA
/secchi_depth=NA
/water_depth=NA
/wave_height=NA
/wind_speed=NA
!
! Turner fluorometer last calibrated 27 Mar 2000 JMW
!
/missing=-999
/delimiter=space
/fields=date,time,station,lat,lon,depth,CHL
/units=yyyymmdd,hh:mm:ss,none,degrees,degrees,m,mg/m^3
/end_header@
19981016 14:33:22 st001 32.3234 -53.1624 0.5 0.32
19981017 13:01:56 st002 33.1122 -53.1276 0.5 0.33
19981018 15:25:45 st003 36.1234 -51.2363 0.5 0.45
19981019 12:11:08 st004 31.8823 -55.1125 0.5 0.22
19981020 14:13:14 st005 34.2341 -52.3545 0.5 0.11
...
```

Figure 3.4 A representative example of a SeaBASS pigment data file.

Chapter 4

Data Submission and Quality Control

4.1 Introduction

Both bio-optical algorithm development and ocean color sensor validation analyses are observation limited, and, therefore, access to high quality *in situ* data is fundamental for their progress. Accordingly, the SPO has defined a series of standard oceanographic and atmospheric parameters, instrument specification and calibration techniques, field measurement methods, and data submission protocols to ensure that data archived in SeaBASS would be acceptable for such activities (Mueller et al. 2002a and 2002b). While the latter requires a specific data file format (as outlined in Chapter 3 of this report), the data submission protocols were designed to be as straightforward and effortless on the part of the contributor, while still offering a useful format for internal SPO efforts. This chapter describes how data files are submitted to SeaBASS and the path of these data through the SeaBASS system (Figure 4.1).

4.2 FCHECK

The SPO developed feedback software to evaluate the format of submitted data files, the principal component of which is known as FCHECK. FCHECK consists of Practical Extraction and Report Language (PERL) code, several lookup tables, and UNIX / LINUX mail handling utilities. It evaluates the header and data blocks of an input file by parsing each and comparing the metadata and data with a series of prerequisite criteria. The output (Figure 4.2) is a series of warnings and errors, where warnings are used to indicate that some desirable condition was not met, for example, an optional header was not provided, and each error references a point where the file format is invalid. With regards to the header block, FCHECK verifies that the following conditions are met:

- all required headers are provided;
- the `/begin_header` and `/end_header` keywords are included;
- the arguments for the time keywords are in the range and format 00:00:00 to 23:59:59;
- the arguments for the date keywords are in the format YYYYMMDD, where *YYYY* is the four-digit year, and *MM* and *DD* are the two-digit month and day, respectively;
- the arguments for the latitude keywords are in the range and format –90.0 to 90.0;
- the arguments for the longitude keywords are in the range and format –180.0 to 180.0;
- bracketed arguments, e.g., "[GMT]", are provided for each time and location keyword;
- the arguments for the `/data_type` and `/data_status` keywords are recognized options;
- the arguments for the `/delimiter` keyword is a recognized option and applicable for the data block;
- the `/missing` keyword has only one argument;
- the number of `/fields` and `/units` arguments are equal; and
- the associated `/units` argument is valid for each standardized `/fields` argument.

For the data block, FCHECK verifies that the following conditions are met:

- the number of columns in each row is equal to the number of `/fields` and `/units` arguments;
- all date and time data are within acceptable ranges, e.g., hour values between 00 and 23;
- all location data are within acceptable ranges, e.g., latitude values between –90.0 and 90.0; and

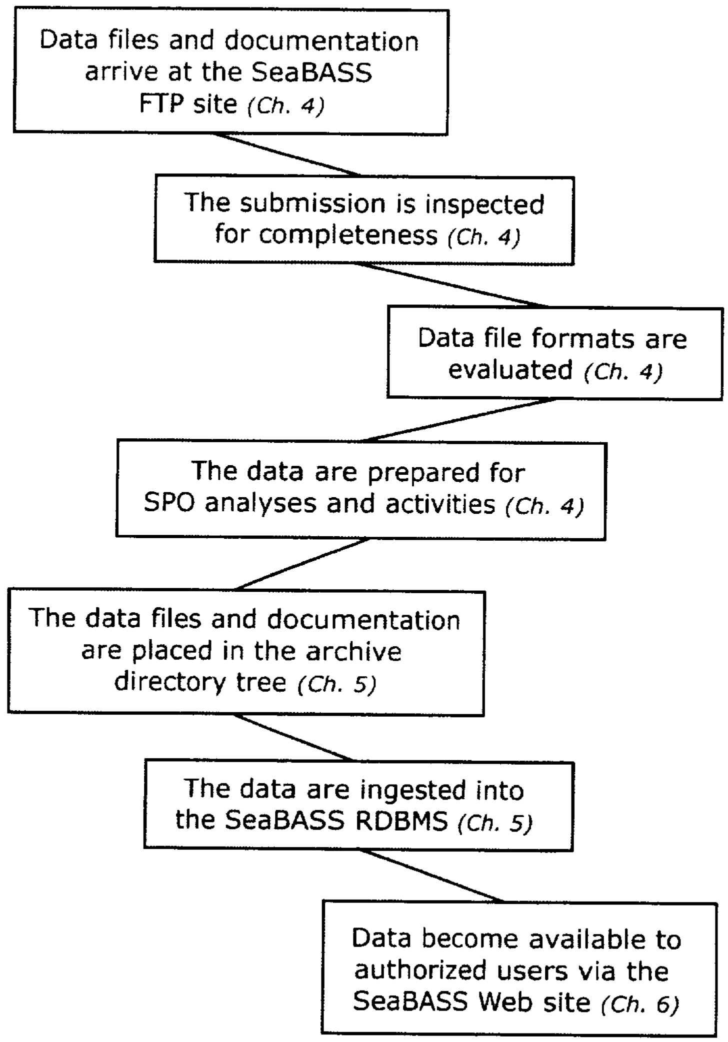


Figure 4.1 An illustrative example of how data files and documentation are handled after their submission to SeaBASS.

- all other data are within reasonable ranges, e.g., non-negative pigment concentrations.

FCHECK is made available to the community using electronic mail and file transfer protocol (FTP). The electronic mail address for FCHECK is:

<fcheck@seabass.gsfc.nasa.gov>.

A contributor may evaluate a single file by electronically mailing the data file to this address. The simultaneous evaluation of multiple data files requires several steps. First, the contributor creates a new directory in the SeaBASS FTP site at:

<ftp://samoa.gsfc.nasa.gov/seabass/fcheck>,

and uploads the files to be evaluated to this directory. The contributor then sends electronic

```
FCHECK Ver. 3.0  last modified: Jul 26 2002 09:49:07
File: chl20020110_1400.sb
This file has passed the FCHECK program.


FCHECK Ver. 3.0  last modified: Jul 26 2002 09:49:07
File: iop20020110_1400.sb
This file has passed the FCHECK program but 1 warnings were issued.

*************************** WARNINGS: ******************************
1) Negative value detected in data block for field(s):
          aph ap ad


FCHECK Ver. 3.0  last modified: Jul 26 2002 09:49:07
File: spmr20020110_1400.sb
This file has failed the FCHECK program.

2 errors were found.

**************************** ERRORS: ********************************
1) Missing value not allowed for depth!

2) Header /fields:
      solar_time_gmt water_temp is(are) not found in the names
      list at http://seabass.gsfc.nasa.gov/cgi-bin/stdfields.cgi

      This may be due to one of the following:
      a) The fieldname is incorrectly formatted [Lw_490 rather than
      the required Lw490]
      b) The fieldname is not typical for standard SeaBASS
      submission, i.e. it's new to us!  If the fieldname does not
      have an equivalent standardized name, please contact the
      SeaBASS administrator to discuss how to submit the information
```

Figure 4.2 Several examples of the standard output of FCHECK, the SeaBASS data file format verification software. From top to bottom, the three evaluated files (1) fully passed FCHECK, (2) passed FCHECK, but warnings were issued, and (3) failed to pass FCHECK.

mail to FCHECK with the specific subject line "FTP: <new directory>" where *<new directory>* is the name of the directory created in the first step. The receipt of this electronic mail triggers FCHECK to evaluate each file in *<new directory>*. For both methods of invoking of FCHECK, the results of the analyses are electronically mailed back to the contributor and to the SeaBASS Administrator. Additional help is available by sending electronic mail to FCHECK with the word "HELP" in the subject line and via the SeaBASS Web site.

4.3 Documentation

An often unspoken, yet underlying, objective of the SPO is to maintain sufficient information and detail about the archived data as to make the full bio-optical set as complete as possible and, therefore, maximally approachable to an outside user. The SPO feels that complete documentation both reinforces accurate use of the data and encourages future data corrections and updates, when necessary. Further, complete documentation enhances the preservation of these data when submitted to the NODC. As such, the SeaBASS Administrator will accept any and all supporting documentation deemed relevant by the data contributor.

At a minimum, the SPO requires that the supporting documentation include two files, a cruise report and an instrument report. The format of the former is left to the discretion of the contributor, so long as it includes a station log that provides ancillary information or comments for each measurement, such as date, location, sea and sky states, and other observations. Contributors are encouraged to include additional documentation, such as digital photographs of sea and sky states. The latter should be either a report that lists the instruments used and the data processing methods, equations, and any relevant references, or the instrument calibration files themselves. Calibration files must include both calibration coefficients and the date each instrument was calibrated. In cases where this information is not available or relevant per se, a brief document describing any calibration techniques or reasons for the former is acceptable.

4.4 Data submission

Data files and supporting documentation are submitted to SeaBASS via FTP. Note that using the FTP version of FCHECK does not constitute a submission to SeaBASS. Once a data contributor has finished preparing their data files, and each of these files have passed FCHECK, both the data files and the supporting documentation and instrument calibration files are uploaded to the SeaBASS FTP site at:

<ftp://samoa.gsfc.nasa.gov/seabass/incoming>.

The data contributor will need to create a unique directory for their data, if one does not already exist. Once the data are uploaded, the data contributor is asked to inform the SeaBASS Administrator via electronic mail that data have been submitted. As described in Chapter 2, all US SIMBIOS Science Team members and researchers receiving funding under the Ocean, Ice, and Climate NRA are obligated to submit data within six months and one year, respectively, of its collection.

4.5 Evaluation of data file formats

The SeaBASS Administrator routinely collects data uploaded to the FTP site. Prior to ingestion into the data archive, however, the submission and the data files' format are evaluated using several standard quality control (QC) procedures. First, the submission is inspected for completeness. All available supporting documentation and instrument calibration files must accompany the submitted data files. These files must match exactly the arguments listed for the `/documents` and `/calibration_files` header keywords.

Submissions that do not include supporting documentation and calibration files are considered incomplete and the data files will not be archived. Next, the format of each file is evaluated using FCHECK. The Administrator will occasionally waive certain reported errors, if unavoidable for a given file or data type. Otherwise, the data contributor must address and resolve all errors reported by FCHECK.

Once a data file has passed FCHECK, its header block is further evaluated with the following guidelines. The arguments for the `/experiment` and `/cruise` keywords must be consistent amongst all files in the submission and all previously submitted data from the specific cruise. The arguments for the `/data_type` and `/data_status` keywords must be valid, for example, either "preliminary", "update", or "final" for `/data_status`, and accurate. Certain optional headers must be present for specific data types for which they provide relevant information, such as, the keyword `/measurement_depth` for spectrophotometric analyses of discrete water samples. And, the single argument for the `/missing` header must be numeric and a nonvalid data value for the given data type.

Additionally, all white space is removed from the header block, with the exception of any comment lines, and all non-standard field names must be defined as a comment line in each file in which they are used. With regards to both the header and data blocks, the date, time, latitude, and longitude data values reported in each file must be within an acceptable range, for example, ±90 for latitude.

The Administrator regularly creates a map of data points for each specific cruise to verify that the data are reasonably continuous and oceanic (Figure 4.3). All missing values in both blocks are located and verified. Finally, the Administrator adds one or two additional pieces of information to each data file. An additional header keyword and argument, `/received=YYYYMMDD`, is added to each file to document the date of submission of the data file. For the header argument, *YYYY* is the four-digit year, *MM* is the two-digit month, and *DD* is the two-digit day. And, comment lines describing changes are added to any data file whose content was modified by the Administrator.

4.6 Evaluation of data

The SPO has developed an additional series of QC protocols, and accompanying software, to assist in evaluating the quality of submitted radiometric data and to prepare the data for various satellite-validation activities. These practices were developed to further aid in (1) verifying the radiometric accuracy of the *in situ* data, (2) evaluating the radiometric stability of the field instruments, particularly for long term time series submitted to SeaBASS, and (3) developing consistency amongst the data used to validate the satellite-derived data products.

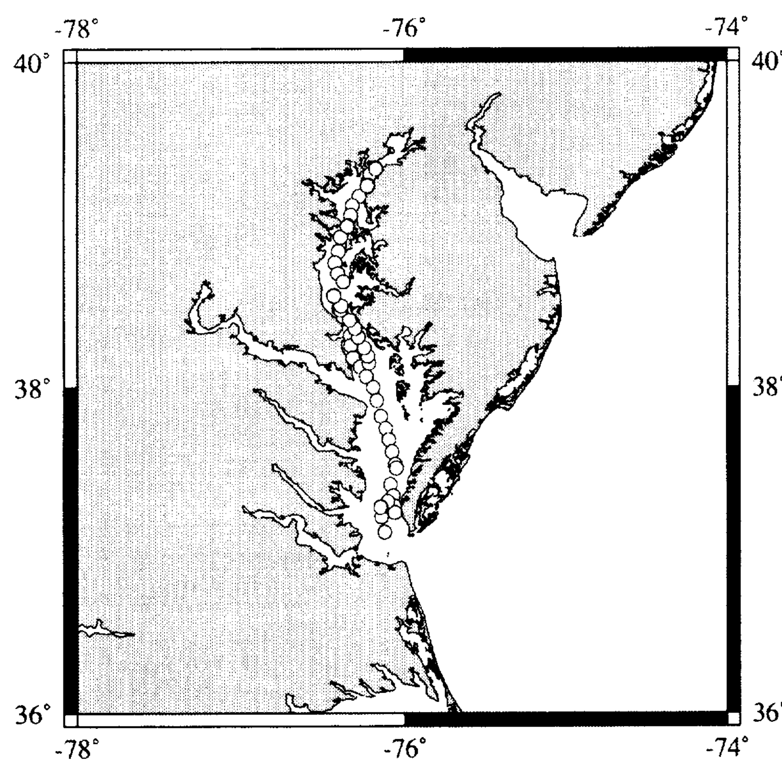


Figure 4.3 A regional map generated by the SeaBASS Administrator to verify the accuracy of the location coordinates provided in a series of submitted SeaBASS data files.

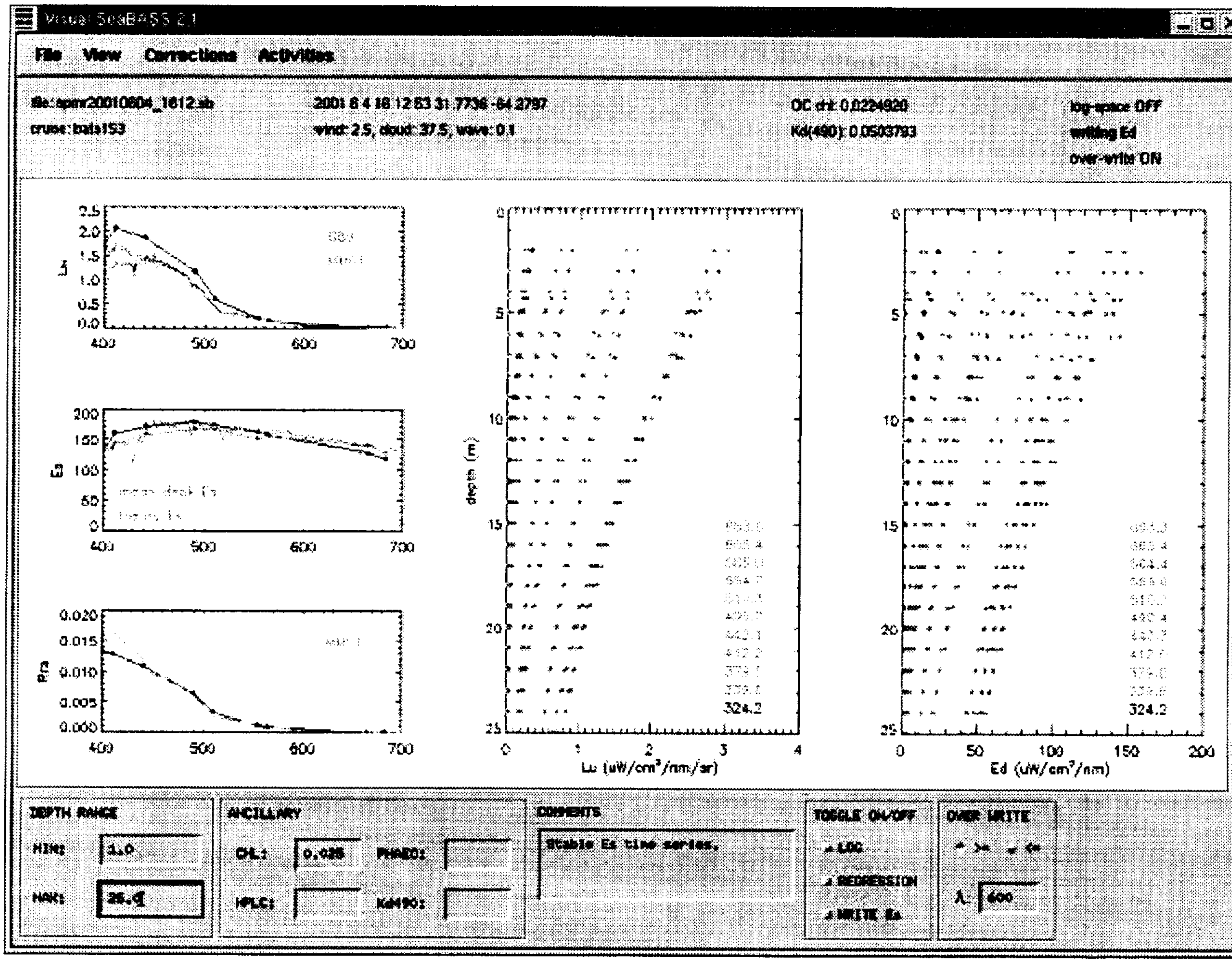


Figure 4.4 A screen-capture of the interactive software, *Visual SeaBASS*, used by the SPO to evaluate radiometric profiles and prepare data for their satellite validation and algorithm development activities.

With regards to the latter, recent work has suggested that the uncertainty associated with derived parameters, such as water-leaving radiance, may be significantly reduced when a single processor prepares the data (Hooker et al. 2001).

In-water optical properties, such as downwelling irradiance and upwelling radiance, are analyzed using interactive software (Figure 4.4), internally known by the SPO as *Visual SeaBASS*, written in the Interactive Data Language (IDL) programming environment developed by Research Systems, Inc. Visual SeaBASS was developed to display and evaluate profiles of optical properties and to estimate near-surface values from these profiles (Mueller 2002). Users are provided a variety of options when defining the extrapolation interval for the computation of near-surface values, including the ability to remove outliers through statistical filtering. Other features include a capability to adjust radiance values to minimize instrument self-shading effects (Zibordi and Ferrari 1995) and the ability to load additional oceanographic data (for example, water temperature profiles) for comparison with the radiometric casts.

The main Visual SeaBASS window displays both the radiometric depth profiles and the estimated near-surface radiometric spectra. For qualitative comparison, the spectral plots also include theoretical surface values, calculated using three well-validated bio-optical algorithms. Currently, water-leaving radiance, clear sky downwelling irradiance, and remote-sensing reflectance spectra are estimated using the models described in Gordon et al. (1988), Frouin et al. (1989), and Morel and Maritorena (2001), respectively. The theoretical in-water spectra are used for qualitative comparison only, as these algorithms were developed for Case 1 waters (Morel and Prieur 1977) and are not always appropriate for direct comparison with the near-shore or colored dissolved organic matter (CDOM) dominated water studied by many SeaBASS contributors. Such comparisons of estimated values with theoretical clear-water and clear sky maxima, however, have proven valuable in distinguishing erroneous or contaminated profiles.

Surface radiances estimated using Visual SeaBASS are further evaluated by comparing band ratios of the estimated values with those of the theoretical values. Normally, all wavebands shorter than 555 nm are normalized to 555 nm, a wavelength thought to be one of the most invariant for the widest range of water types. The shapes of the estimated and theoretical spectra are further compared by normalizing each spectra to its integrated area (i.e., area under the curve). Such integration eliminates differences due to magnitude so that the spectral shapes may be directly compared. Both analyses assist in distinguishing erroneous or contaminated wavebands, as they illustrate variations or inconsistencies in the full spectra. When long running time series of data are available (for a particular locality, station, or region of interest) the newly submitted data are compared with these historical values (Figure 4.5), a practice which assists in evaluating the radiometric stability of the field instruments. Note that outliers located in the latter analyses are not excluded from ingestion into the data archive, as such evaluations are purely qualitative and subjective. Rather, the data are kept intact and the contributor is contacted in an attempt to resolve any possible source(s) of contamination and explain any differences.

Last, submitted radiometric and pigment data are prepared for the SPO's satellite-to-*in situ* match-up analyses (Bailey et al. 2000). In the event of multiple radiometric profiles per station, the data are further evaluated to select single representative spectra for each station (Figure 4.6). A governing philosophy of the match-up analyses is the need to compare the satellite data product with a single ground truth measurement per date and location. Normally, this single spectrum is defined as that collected under the clearest or most stable sky conditions (determined using downwelling irradiance data collected by a shipboard reference sensor) or as the data with the clearest water-leaving radiance value at 490 nm, another indication of favorable atmospheric conditions. Replicate pigment measurements are normally averaged.

Atmospheric data submitted to SeaBASS, in particular, AOT, are subject to a similar series of QC standards. These protocols and analyses are documented elsewhere (Fargion et al. 2001, Mueller et al. 2002a and 2002b) and will not be discussed further in this report.

4.7 Data ingestion

Once the SeaBASS Administrator finishes evaluating the incoming data set, the data files are organized, cataloged and ingested into the SeaBASS archival system (as described in Chapter 5 of this report). Upon completion of the latter, the data and data files are freely available to authorized users via the World Wide Web (as outlined in Chapter 6 of this report).

The contributor is contacted if problems occur or the data set cannot be archived for any reason. Data prepared by the SPO for their validation and merger activities are generally not made available to the community, but may be distributed to interested, appropriate parties for additional analysis upon request.

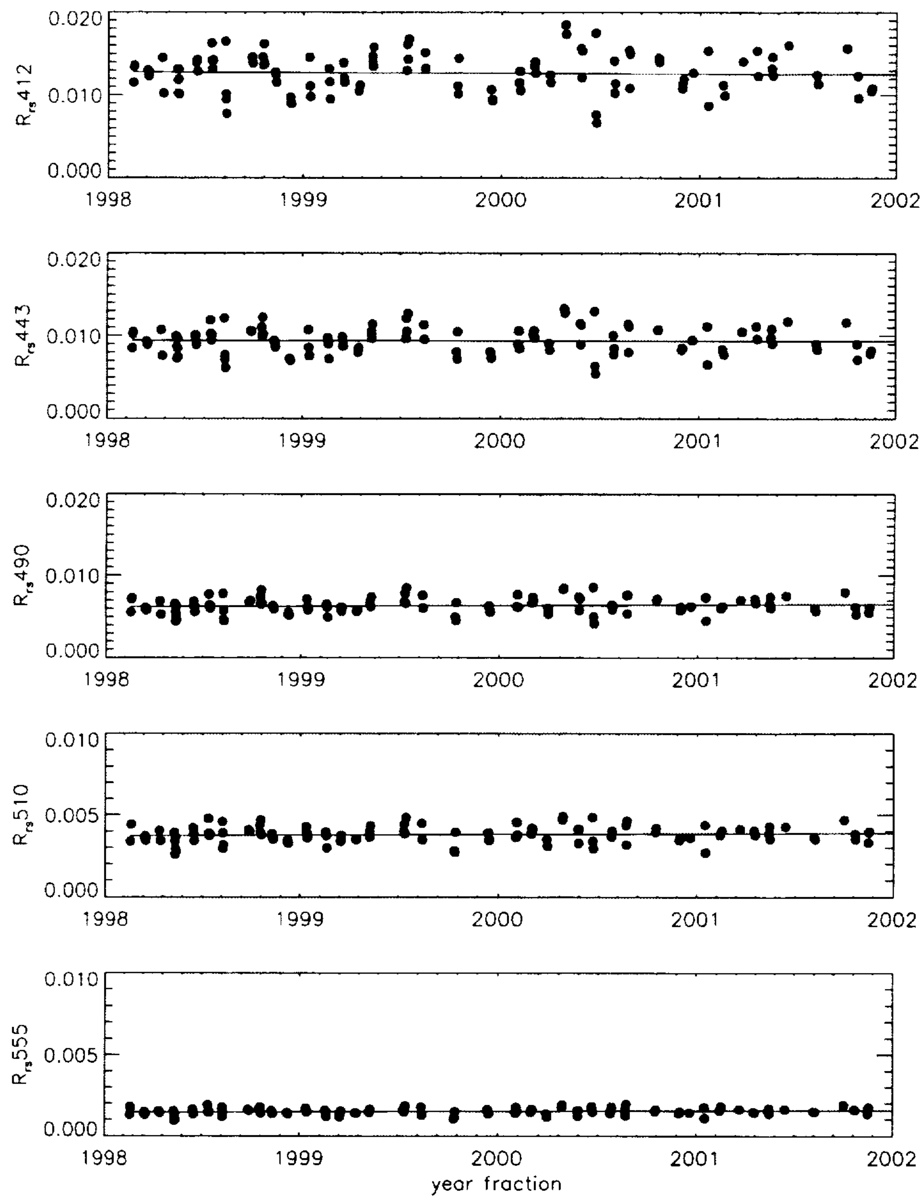


Figure 4.5 A representative example of a time series analysis of radiometric data, specifically, remote-sensing reflectance, collected on a long-term field campaign. Such analyses are used to help verify the radiometric stability of data archived in SeaBASS.

4.8 Updates

The SPO encourages all contributors to update and correct their submitted data sets. Records are maintained of updates and corrections and the updated data are stored offline. A summary of new and updated data is posted online at:

<http://seabass.gsfc.nasa.gov/cgi-bin/seabass_news.cgi>.

The contributor maintains the responsibility of ensuring that the current data in the archive is identical to that used in the contributor's most recent publications or current research.

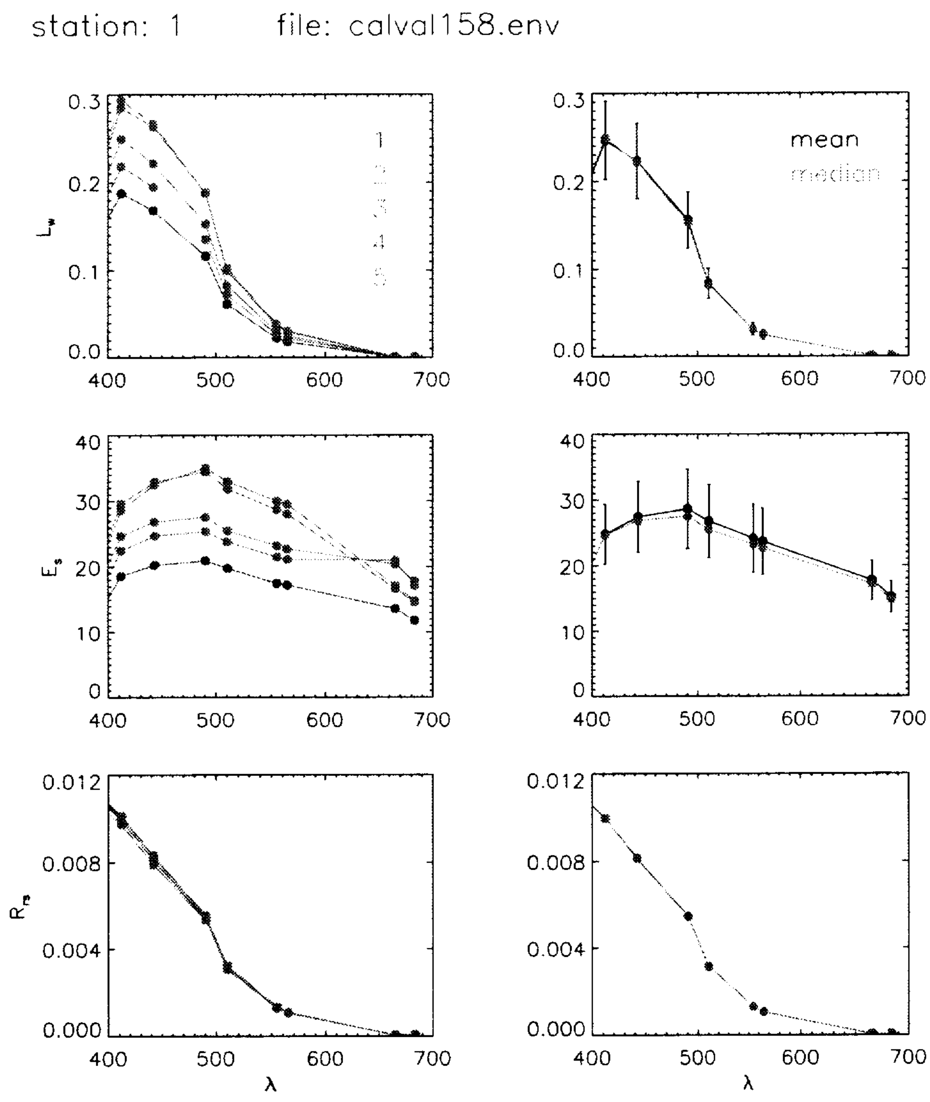


Figure 4.6 A representative example of how replicate surface spectra measured at a single station are compared with the purpose of assigning a single representative spectrum to the station. Stations are defined using narrow temporal and spatial boundaries.

Chapter 5

SeaBASS Architecture and RDBMS Design

5.1 Introduction

In building SeaBASS, the SPO was tasked with creating a system to archive, catalog, and distribute bio-optical data and relevant documentation. It required a repository with holdings that were queriable and available via the World Wide Web, secure enough to limit access to authorized users, and accessible by all computer platforms. Further, the design needed to be easily expandable and flexible enough to accommodate large data sets and multiple data types. As discussed in Chapter 3 of this report, to satisfy several of these prerequisite conditions, SeaBASS data files adhere to the basic ASCII format. The current chapter outlines how such data files are ingested, organized, and stored at NASA Goddard Space Flight Center.

The current architecture of the SeaBASS system consists of two principal components: a directory tree structure, where the data files and documentation are organized and stored, and a RDBMS used to further catalog, archive, and distribute the data. All submitted data files and documentation reside permanently in the directory tree. For each file, metadata from the header arguments, its location in the directory tree, and some geophysical data are stored in the RDBMS. As such, the RDBMS may be queried to compile information about the bio-optical data set or to locate certain data. The tandem use of these components provides a proficient means for the SPO to archive and catalog their data holdings. In conjunction with a series of online CGI forms that interface with the RDBMS (described in detail in Chapter 6 of this report), these components also provide an effective means for authorized users to search the bio-optical holdings and obtain specific data.

5.2 Data ingestion

Once the evaluation of a submitted data set is complete, the SeaBASS Administrator moves the data files and documentation from the SeaBASS FTP site to an appropriate location in the directory tree. Data and metadata included in the data files are then ingested into the RDBMS using additional software, written in PERL and Transact-SQL, an extension of the Structured Query Language database programming environment. The software system gathers information from each data file by parsing the header and data blocks. This information is loaded into a series of linked RDBMS tables using SQL stored procedures and bulk copy commands. Upon completion, the data are fully included in the bio-optical data set and immediately available to authorized users (as described in Chapter 6 of this report).

5.3 The data archive

All data files, related documents, and instrument calibration files are stored in a directory tree structure (Figure 5.1), residing on a dedicated server at NASA Goddard Space Flight Center. The directory tree is organized by affiliation of the contributing PI, experiment, and specific cruise. Each cruise directory has additional subdirectories where the data files, documentation, and other related files are stored. Typically, the data files reside in a subdirectory named "archive", and supporting documents, images, global and regional maps, and calibration files in subdirectory named "documents". A third subdirectory, named "raw", is occasionally created by the SeaBASS Administrator to store administrative comments, SPO processing code, and previous versions of updated data files.

Individuals with full access to the bio-optical archive may peruse the SeaBASS directory tree by pointing their Web browsers to:

<http://seabass.gsfc.nasa.gov/SEABASS_ARCHIVE>.

As valid SeaBASS username and password is required, this online index is not available to the general public. Note, also, that the "archive" and "documents" subdirectories are available for perusal by authorized users, however, the "raw" subdirectory is not.

5.4 The relational database management system

The SeaBASS RDBMS was built using the SQL Server product from Sybase, Inc. Its design consists of a series of tables (Figure 5.2, Table 5.1), in the third normal form, used to store both metadata from each header keyword of a file and certain geophysical data values from the file's data block. To enhance query performance, the server is configured for parallel data access using multiple computer processors. In addition, most database objects reside on their own physical device and data are frequently distributed across multiple partitions. The SeaBASS Administrator regularly addresses RDBMS performance and tuning issues and implements new logic, as the standard user base continues to expand. The table architecture is such that users may query the RDBMS for metadata or data from a single file, or all data archived for a given institution, contributor, experiment, data type, data parameter, or spatial and temporal range. The architecture may be easily expanded to accommodate new data types or metadata

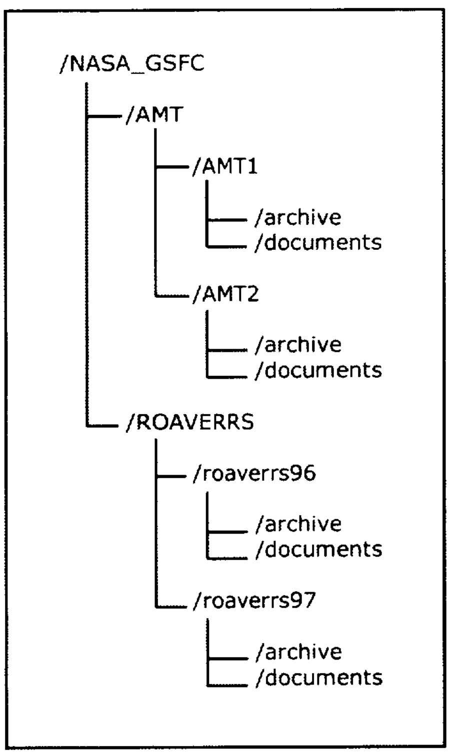


Figure 5.1 A representative example of the organization of the SeaBASS archive directory tree. Data are organized by the affiliation of the contributing PI, experiment, and specific cruise. The subdirectories "archive" and "document" are used to store data files and supporting documentation, respectively.

elements and can comfortably manage large data sets.

Additional metadata tables are occasionally added to the system to accommodate new online resources (refer to Chapter 6 of this report for additional details) or SPO administrative software. The administrative RDBMS tables and software exceed the scope of this report and will not be discussed further.

The RDBMS ingestion software first parses the header block, evaluating each header keyword and argument, and loads this information into the appropriate table. Each header keyword has an analogous column in one or more RDBMS tables (Table 5.1). The system is normalized such that separate tables exist for contributor affiliations, contact information, cruise particulars, and details relating to each data file. The latter includes both ancillary data, such as wind speed and wave height, and other statistics about the data file, including the date the data were archived and the data status, if, for example, the data had been resubmitted. Once the headers have been examined, the data block is parsed. The date, time, latitude, longitude, station name, and measurement depth of each data record is stored in an additional RDBMS table. Certain geophysical data values are loaded into additional RDBMS tables.

As of August 2002, all fluorometric and HPLC pigment concentrations, sun photometer data, including AOT, column water vapor, and total column ozone, and discrete hyperspectral spectrophotometer data are included in additional, separate, RDBMS data tables. All other data are available only as complete data files (those residing in the SeaBASS archive directory tree). The system, however, may be easily expanded in the future to include RDBMS tables for other geophysical data, such as that from depth profiles and flow-through and moored systems.

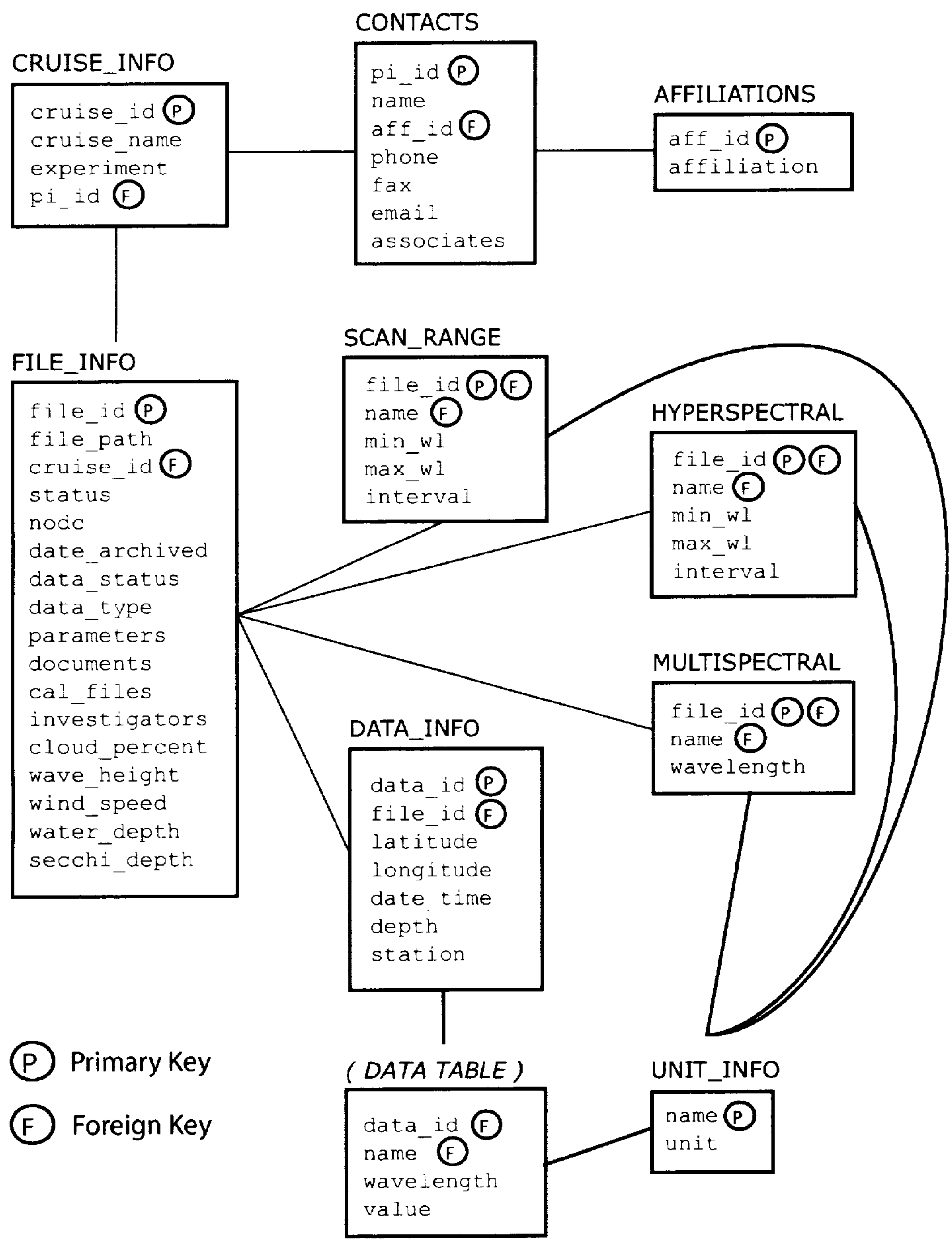


Figure 5.2 The data model for the SeaBASS RDBMS. Each box represents a table, with the table name listed on top (in capital letters) and column names provided inside the box. A circled "P" and "F" designate primary and foreign keys, respectively. The box "(DATA TABLE)" is a generic listing for several tables that hold geophysical data values.

Table 5.1 The SeaBASS RDBMS tables with their associated header keywords and descriptions, as of August 2002. The record "(DATA TABLE)" is a generic listing for several tables that hold geophysical data values. Currently, data tables exist for phytoplankton pigment concentrations, sun photometer data, and discrete hyperspectral spectrophotometer data.

Table name	Related header keywords	Table contents and description
affililiations	`/affiliation`	The universities, laboratories, and organizations of all contributors to SeaBASS.
contacts	`/contacts` `/investigators`	Contact information for each principal contributor of data to the archive.
cruise_info	`/cruise` `/experiment`	A list of experiments and specific cruises with data in the archive. References the principal contributor of data for each record.

data_info	`/east_longitude`
	`/end_date`
	`/end_time`
	`/measurement_depth`
	`/north_latitude`
	`/south_latitude`
	`/start_date`
	`/start_time`
	`/station`
	`/west_longitude`
		The date, time, latitude, longitude, depth, and station of each archived data measurement.

file_info	`/calibration_files`
	`/cloud_percent`
	`/data_status`
	`/data_type`
	`/documents`
	`/fields`
	`/investigators`
	`/secchi_depth`
	`/water_depth`
	`/wave_height`
	`/wind_speed`
		One record for every file in the bio-optical data set. References the cruise on which data in the file were collected. Notes, using a status index, which data were loaded into data tables, and indicates if the data were submitted to the NODC. Includes general information about each file, such as the date data were archived, the status of the data (if, for example, the data are updates of previously submitted data), and the type of data in the file. Also includes a full list of contributors and data parameters, such as CHL and AOT, and references the documentation and instrument calibration files that accompany the data file. Finally, provides additional oceanographic and atmospheric data associated with the data in a given file.

hyperspectral	`/fields`	An inventory of data parameters and the range and interval of wavelengths for each hyperspectral radiometric measurement.
multispectral	`/fields`	An inventory of data parameters and nominal wavelengths for each multispectral radiometric measurement.
scan_range	`/fields`	An inventory of data parameters and the range and interval of wavelengths for each discrete hyperspectral scan, including laboratory spectrophotometer and above water radiometer scans.
unit_info	`/fields` `/units`	A list of standard data parameter names and their respective units.
(DATA TABLE)	`/fields`	Geophysical data values.

Chapter 6

Database Access and Online Resources

6.1 Introduction

The purpose of this chapter is twofold: to present methods of accessing the bio-optical data set and to describe available online resources. The SeaBASS World Wide Web home page, located at

<http://seabass.gsfc.nasa.gov>,

provides a complete description of the system's architecture, comprehensive documentation of access policies and submission protocols, and direct access to the bio-optical data set. Most of the details documented in this report are also posted on the Web site, where they may be updated and amended as the need arises. Note that, when appropriate, the addresses of associated and relevant Web pages have been provided in other chapters. Many of these descriptive Web pages are self-describing and will not be discussed further in this report.

The Web site also includes supplementary resources, all of which are updated regularly, such as tables of new and updated archived data, answers to frequently asked questions, detailed contact information, and interactive mapping routines. In addition, results from the SPO-sponsored satellite-to-*in situ* match-up analyses (Bailey et al. 2000) and the SeaWiFS Bio-optical Algorithm Mini-workshop (SeaBAM) (O'Reilly et al. 1998) are available via the SeaBASS home page. The methods and results from the latter activities are described in the references provided and will also not be further addressed.

Every resource listed above is freely available for public use, with the exception of access to the restricted data set, currently defined as all data collected after 31 December 1999. Web pages providing access to restricted data require a username and password. The SeaBASS Data Access Policy and the process of applying for a username and password are both described in detail in Chapter 2 of this report.

6.2 Bio-optical data set access

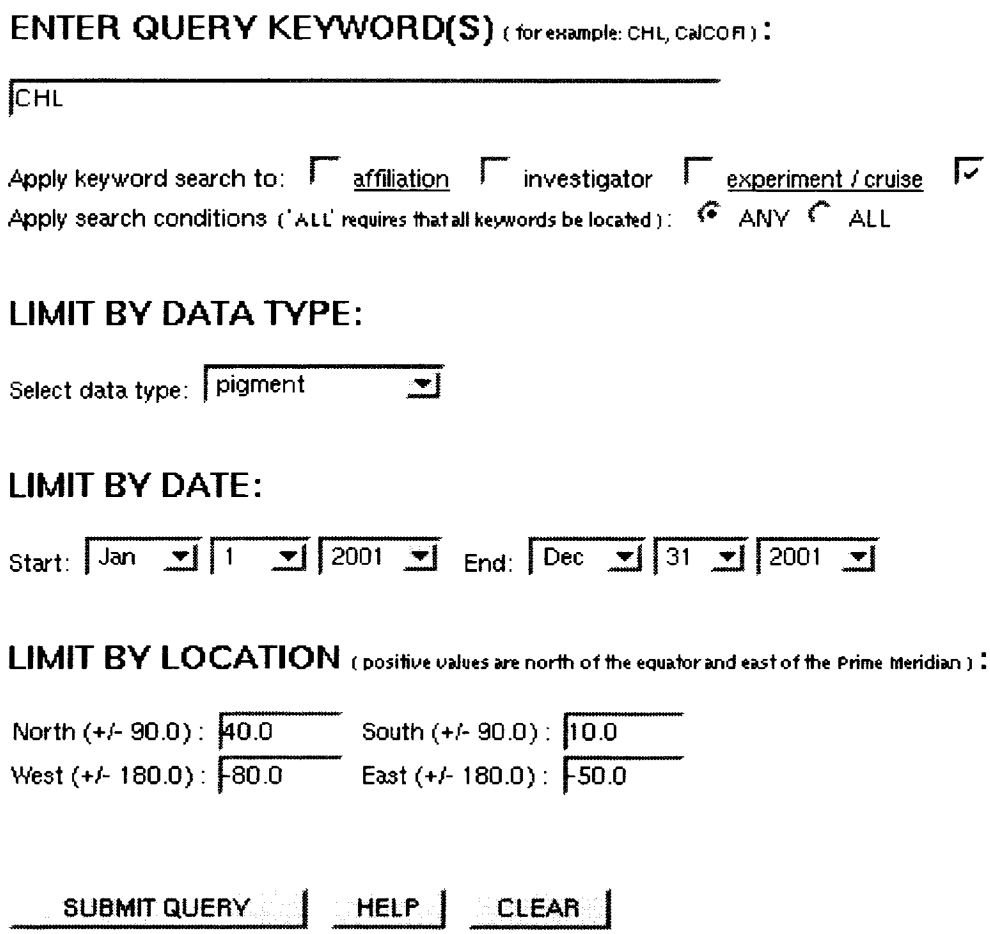
Data and files from the bio-optical data set may be accessed and saved using a series of online search engines (Table 6.1) provided at:

<http://seabass.gsfc.nasa.gov/dataordering.html>.

Each search engine is a CGI form (Figure 6.1) that interfaces with the SeaBASS RDBMS. Using the form, visitors may limit queries to particular experiments, contributors, date and location ranges, and data types (for a description of SeaBASS data types, refer to Chapter 3 of this report). Linked to each search engine are a series of supplementary Web pages with tables listing additional relevant information, for example, the names of archived experiments or data types, to assist users narrow or tailor their queries. On occasion, Java pop up windows are used to provide definitions or explanations of an online feature.

When a user submits a query, the software polls the RDBMS and generates a list of corresponding data files or values which meet the search criteria. Items in the returned list may subsequently be viewed using the client's Web browser, saved locally, and in some cases, mapped or plotted using additional online software. In several instances, two versions of a given CGI form exist, one of which is password-protected and another which accessible by the general public. Both versions are identical, with the exception that only data collected prior to 31 December 1999 are available to the public.

Upon the completion of a query, an electronic mail message is sent to the SeaBASS Administrator. The message notifies the



ENTER QUERY KEYWORD(S) (for example: CHL, CalCOFI):

CHL

Apply keyword search to: affiliation investigator experiment / cruise data field

Apply search conditions ('ALL' requires that all keywords be located): ANY ALL

LIMIT BY DATA TYPE:

Select data type: pigment

LIMIT BY DATE:

Start: Jan 1 2001 End: Dec 31 2001

LIMIT BY LOCATION (positive values are north of the equator and east of the Prime Meridian):

North (+/- 90.0): 40.0 South (+/- 90.0): 10.0

West (+/- 180.0): -80.0 East (+/- 180.0): -50.0

SUBMIT QUERY HELP CLEAR

Figure 6.1 A representative example of an online CGI form used to query the SeaBASS bio-optical data archive.

Administrator that a online query of the RDBMS has been executed and describes the particular search parameters. The Web server and the search engine software also log all user activity. This information is made available to all data contributors upon request.

When a visitor selects data files to be saved, these files are compiled into a single tar (UNIX tape archive) bundle which is created on the SeaBASS FTP site at:

<ftp://samoa.gsfc.nasa.gov/seabass/outgoing>.

Likewise, all geophysical data to be saved are written to a single file placed on this FTP site. Both may be downloaded locally using standard FTP or via the client's Web browser. Users operating on Microsoft Windows and Macintosh platforms may need to install additional third-party software to extract the individual data files from the tar bundle.

6.2.1 Data search engines

Several search engines are available to locate and extract data files and geophysical data values from the bio-optical data set. The most comprehensive search engine, the SeaBASS Bio-optical Search Engine, available at:

Table 6.1. Online resources for extracting and manipulating data from the SeaBASS bio-optical data set. "Restricted" indicates the data are part of the restricted data set and a username and password is required for access. "Public" indicates the data are available to the general public. As of August 2002, all data collected prior to 31 December 1999 are publicly available. Each resource is available via the SeaBASS Web site at <http://seabass.gsfc.nasa.gov/dataordering.html>.

Utility	Access	Description
Bio-optical Search Engine	Restricted, Public	The principal utility for searching the full bio-optical data set. Queries return a list of files, which are available to view and save. Additional online resources may be used to plot and map the data in a returned file(s).
Pigment Locator	Restricted, Public	The search engine used to access directly the fluorometric and HPLC phytoplankton pigment data included in the bio-optical data set. Queries return a list of geophysical data values, which are available for download.
Aerosol Locator	Restricted, Public	The search engine used to access directly the AOT and other sun photometer data included in the bio-optical data set. Queries return a list of geophysical data values, which are available for download.
General Search Engine	Public	The principal utility used to compile, from the bio-optical data set, generic information about a cruise, contributor, data type, or date or location range. Queries return a table of archived cruises and a list of data contributors and parameters associated with each. Additional online resources may be used to map data from a specific cruise.
Cruise Search Engine	Public	A search engine used to compile generic information about cruises included in the bio-optical data set. Queries return date and location ranges, data contributors, and the data types and parameters collected.
Validation Cruise Search Engine	Public	The search engine used to locate potential validation data sets for a specific, user-defined satellite mission. Queries return a list of cruises and their date range, center latitude and longitude, data contributors, and data parameters.
Archive Directory Tree	Restricted	The directory structure used to organize and store the data files included in the bio-optical data set. The data files, supporting documentation, and instrument calibration files are all available for perusal.
Mapping Utility	Public	A utility for generating maps of SeaBASS data points based on customized date, location, and parameter inputs. Additional resources include a global map of all data included in the bio-optical data set and mission-specific maps for OCTS/POLDER, SeaWiFS, MODIS Terra and Aqua, and MERIS. All are updated daily.

<http://seabass.gsfc.nasa.gov/cgi-bin/seabass_search.cgi>,

provides full access to the bio-optical data set. Queries return a list of matching data files, which are available to view or download. In addition, users may generate online maps and plots of data from one or more files or download relevant documentation. For this search engine, users may not only limit queries to particular date ranges, but also to defined monthly climatologies (i.e., specific months for any given number of years). This utility provides access to all data in the bio-optical data set.

Currently, two search engines are available for extracting geophysical values from the bio-optical data set, the SeaBASS Pigment Locator, found at

<http://seabass.gsfc.nasa.gov/cgi-bin/pigment_search.cgi>,

and the SeaBASS Aerosols Locator, located at:

<http://seabass.gsfc.nasa.gov/cgi-bin/aerosol_search.cgi>.

These utilities provide direct access to phytoplankton pigment concentrations and sun photometer data, respectively, included in the bio-optical data set. Queries return a table of geophysical data values, including ancillary information, such as date, time, location, cruise name, and contributor, which are available to download. For both of these utilities, users define the data parameters for which they'd like to search by navigating through a series of checkboxes.

Finally, visitors with full SeaBASS accounts may access the data archive directly via their Web browser. Such users may peruse the SeaBASS directory tree by visiting

<http://seabass.gsfc.nasa.gov/SEABASS_ARCHIVE>.

Here, all data files, supporting documentation, and instrument calibration files comprising the bio-optical data set are available to view or download.

6.2.2 Metadata search engines

Other search engines are available for compiling metadata relating to the bio-optical data set. These provide generic information about the data, such as cruise and experiment names, date and location ranges, data parameters collected, and contributor names. They do not, however, provide access to geophysical data values, and are therefore available to the general public.

The utility of the metadata search engines is twofold: visitors may compile general information about data archived in SeaBASS without directly accessing the geophysical data, and queries return a summary or overview of archived cruises, for example, date and location ranges, whereas the data search engines do not. The former is thought to be particularly useful for (1) visitors without full access to the bio-optical data set, who are interested in what has been archived, (2) SeaBASS data contributors desiring a simple, yet comprehensive, list of their archived data, and (3) other researchers searching for potential validation cruises for their satellite validation activities.

Currently, the most versatile metadata search utility is the SeaBASS General Search Engine, available at:

<http://seabass.gsfc.nasa.gov/cgi-bin./general_info.cgi>.

Like the data search engines, queries may be limited by particular experiments, contributors, date and location ranges, and data types. Queries generate a table of archived cruises and the data types and parameters collected on each. Subsequent Web pages, hyperlinked to each cruise in the table, display more detailed information about the given cruise, for example, date and location ranges, and provide the user the option of generating a regional map of data points.

The SeaBASS Cruise Search Engine, located at:

<http://seabass.gsfc.nasa.gov/cgi-bin/cruise_search.cgi>,

is a significantly simplified version of the General Search Engine. Its design was arranged to be more efficient for visitors interested in particular archived cruises. Here, visitors enter only a cruise name, or a search string. Queries return the date and location range, data contributor(s), and data types and parameters for each matching cruise. In addition, users are provided the option of generating a regional map of data points for each cruise.

The last metadata search engine, the SeaBASS Validation Cruise Search Engine, available at:

<http://seabass.gsfc.nasa.gov/cgibin/validation_cruises.cgi>,

operates somewhat differently than the others. This utility was designed specifically to assist researchers requiring potential validation cruises for their satellite calibration and validation activities. As such, queries may be limited only by satellite mission. Current options include the Ocean Color and Temperature Scanner (OCTS) / POLDER, SeaWiFS, MODIS Terra and Aqua, and MERIS. Queries return a table of potential validation cruises. Each record in the table includes the data contributor(s), the data type collected, the start and end dates, and the center latitude and longitude coordinates for the given cruise. As with the others, users are provided the option of generating a regional map of data points for each cruise.

6.3 Other online utilities

6.3.1 Maps

Visitors wishing to generate maps of SeaBASS data may do so interactively using the SeaBASS Mapping Utility, available at:

<http://seabass.gsfc.nasa.gov/cgi-bin/seabass_map.cgi>.

The default map is global, however, users are provided the option of customizing latitude and longitude boundaries. Mapped data points may be further limited by user-defined date ranges and specific data parameters (e.g., "chl" and "AOT"). In the event more than one data parameter is specified, users may restrict mapped points to those where all parameters were collected; otherwise, all data where any matching parameter was collected will be included in the map.

Several other global maps are linked to this Web site: (1) a map of all data included in the bio-optical data set, and (2) mission-specific maps of pigment, radiometer, and sun photometer data points. Currently, the latter includes maps for OCTS / POLDER, SeaWiFS, MODIS Terra and Aqua, and MERIS. All of the above are updated daily and are available for download.

6.3.2 News and updates

All changes made to SeaBASS, including both system updates and the archival of new and updated data, are documented online at the SeaBASS New and Updates page, located at:

<http://seabass.gsfc.nasa.gov/cgi-bin/seabass_news.cgi>.

The first half of this site is reserved for news of recent additions and modifications to the SeaBASS RDBMS and Web site. A brief summary of each update, the date of each update, and hyperlinks to relevant Web sites, when available, are all posted.

The remainder of the site is dedicated to posting information about recently submitted bio-optical data sets. The principle component is a table that lists details about data submitted within the past three months. The table includes the data contributor, the date data were submitted, the data status (either new or update), the experiment and cruise names(s), and the data parameters provided. Each record also includes two hyperlinks: one to the data's position in the archive directory tree, which is accessible only to visitors with full access to the bio-optical data set; and another to an additional Web page that lists more detailed information about the cruise(s), for example, date and location ranges, and provides a regional map of data points. With the exception of the link to the archive directory tree, the information on this Web page is available to the general public.

Also provided on this site is a link to the SeaBASS Email Notification Service, located at:

<http://seabass.gsfc.nasa.gov/cgi-bin/seabass_mail.cgi>.

Interested parties may join the SeaBASS Updates mailing list by completing and submitting the electronic form provided at this site. An electronic message listing recently submitted data is sent to members of this list once a week. Only cruises archived in the past week are included in the electronic message.

6.3.3 Historical data sets

The SPO maintains two additional Web sites relating to other global bio-optical data sets, both of which are available via the SeaBASS Web site. The first, the SeaWiFS calibration and validation historical pigment database (Firestone and McClain 1994), was originally assembled to assist in evaluating satellite pigments retrivals. It is located at:

<http://seabass.gsfc.nasa.gov/cgi-bin/pigment_query.cgi>,

and includes only measurements of chlorophyll *a* and phaeophytin pigment concentrations.

The second, the SeaWiFS Bio-optical Mini-workshop (SeaBAM) data set (O'Reilly et al. 1998), was assembled to develop and evaluate the operational SeaWiFS chlorophyll *a* and CZCS-like pigment algorithms. It is located at:

<http://seabass.gsfc.nasa.gov/seabam/seabam.html>.

and includes coincident radiometric observations and chlorophyll *a* concentrations. The SeaBAM Web site also includes details about the candidate algorithms and results from the workshop. The two data sets are described in detail in the provided references and will not be discussed further in this report. Note, however, that most data from both sets has subsequently been archived in the SeaBASS bio-optical data set.

6.3.4 NODC

In December 2001, the public data from the bio-optical data set, those collected prior to 31 December 1999, were released to the NODC for inclusion in their national archive. This submission is available via their Web site at:

<http://www.nodc.noaa.gov/col/projects/access/ seabass.html>.

Additional details regarding the access and acknowledgement policies for these data are provided in Chapter 2 of this report.

6.4 User statistics

As of August 2002, 35 research groups outside of the SPO have been granted full access to SeaBASS. From 1 January 2002 through 31 July 2002, these groups queried SeaBASS over 800 times and downloaded more than 42,000 data files from the bio-optical data set. During the same time period, 97 research groups searched the public data set 375 times and downloaded over 22,000 files.

Glossary

AERONET	Aerosol Robotic Network
AOT	Aerosol Optical Thickness
ASCII	American Standard Code for Information Interchange
CDOM	Colored Dissolved Organic Matter
CGI	Common Gateway Interface
CZCS	Coastal Zone Color Scanner
FTP	File Transfer Protocol
GLI	Global Imager
HPLC	High Performance Liquid Chromotography
IDL	Interactive Data Language
IP	Internet Protocol
LINUX	A UNIX-type operating system developed under the GNU General Public License
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NRA	NASA Research Announcement
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center
OCTS	Ocean Color and Temperature Scanner
PERL	Practical Extraction and Report Language
POLDER	Polarization and Directionality of the Earth's Reflectances
QC	Quality Control
RDBMS	Relational Database Management System
SCOR	Scientific Committee on Oceanic Research
SeaBAM	SeaWiFS Bio-optical Mini-workshop
SeaBASS	SeaWiFS Bio-optical Archive and Storage System
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SI	International System of Units
SIMBIOS	Sensor Intercomparison & Merger for Biological & Oceanic Interdisciplinary Studies
SPO	SeaWiFS / SIMBIOS Project Office
SQL	Structured Query Language
UNIX	Uniplexed Information and Computing System

References

Bailey, S.W., C.R. McClain, P.J. Werdell, and B.D. Schieber, 2000: Normalized water-leaving radiance and chlorophyll *a* match-up analyses. In: McClain, C.R., R.A. Barnes, R.E. Eplee, Jr., B.A. Franz, C. N.C. Hsu, F.S. Patt, C.M. Pietras, W.D. Robinson, B.D. Schieber, G.M. Schmidt, M. Wang, S.W. Bailey, and P.J. Werdell, 2000: SeaWiFS postlaunch calibration and validation analyses, part 2. *NASA Tech. Memo. 2000-206892, Vol. 10*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 45-52.

Evans, R.H., and H.R. Gordon, 1994: Coastal Zone Color Scanner "system calibration": A retrospective examination. *J. Geophys. Res.*, 99, 7293-7307.

Fargion, G. S. and C.R. McClain 2001: SIMBIOS Project 2000 Annual Report. *NASA Tech. Memo. 2001-209976*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 164 pp.

Fargion, G. S. and C.R. McClain 2002: SIMBIOS Project 2001 Annual Report. *NASA Tech. Memo. 2002-210005*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 184 pp.

Fargion, G. S., R. Barnes and C.R. McClain 2001: In Situ Aerosol Optical Thickness Collected by the SIMBIOS Program (1997-2000): Protocols, and Data QC and Analysis. *NASA Tech. Memo. 2001-209982*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 103 pp.

Firestone, E.R., and S.B. Hooker, 2001: SeaWiFS postlaunch technical report series cumulative index: Volumes 1-11. *NASA Tech. Memo. 2001-206892 Vol. 12*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp.

Firestone, J.K., and C.R. McClain, 1994: The historical pigment database. In: Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-N. Yeh, and Y. Ge, 1994: The SeaWiFS bio-optical archive and storage system (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 16-21.

Firestone, J.K., E-N. Yeh, T.L. Westphal, C.R. McClain, S.B. Hooker, and Y. Ge, 1994: The bio-optical data system. In: Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-N. Yeh, and Y. Ge, 1994: The SeaWiFS bio-optical archive and storage system (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 6-15.

Frouin, R., D.W. Lingner, C. Gautier, K.S. Baker, and R.C. Smith, 1989: A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance at the ocean surface. *J. Geophys. Res.*, 94, 9731-9742.

Gordon, H.R., D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, and W.W. Broenkow, 1983: Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparisons of ship determinations and CZCS estimates. *Appl. Opt.*, 22, 20-36.

Gordon, H.R., O.B. Brown, R.H. Evans, J.W. Brown, R.C. Smith, K.S. Baker, and D.K. Clark 1988: A semianalytic radiance model of ocean color. *J. Geophys. Res.*, 93, 10909-10924.

Holben, B.N., D. Tanre, A. Smirnov, T.F. Eck, I. Slutsker, N. Abuhassan, W.W. Newcomb, J. Schafer, B. Chatenet, F. Lavenue, Y.J. Kaufman, J. Vande Castle, A. Setzer, B. Markham, D. Clark, R. Frouin, R. Halthore, A. Karnieli, N.T. O'Neill, C. Pietras, R.T. Pinker, K. Voss, and G. Zibordi, 2001: An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET. *J. Geophys. Res.*, 106, 12 067-12 097.

Hooker, S.B., and W.E. Esaias, 1993: An overview of the SeaWiFS project. *EOS Trans. AGU*, 74, 241-246.

Hooker, S.B., and J.K. Firestone, 1994: Overview of the SeaBASS file architecture. In: Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-N. Yeh, and Y. Ge, 1994: The SeaWiFS bio-optical archive and storage system (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 3-5.

Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-N. Yeh, and Y. Ge, 1994: The SeaWiFS bio-optical archive and storage system (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 37 pp.

Hooker, S.B., and C.R. McClain, 2000: The calibration and validation of SeaWiFS data. *Prog. Oceanogr.*, 45, 427-465.

Hooker, S.B., and S. Maritorena, 2000: An evaluation of oceanic radiometers and deployment methodologies. *J. Atmos. Ocean. Tech.*, 17, 811-830.

Hooker, S.B., G. Zibordi, J-F. Berthon, D. D'Alimonte, S. Maritorena, S. McLean, and J. Sildam, 2001: Results of the second SeaWiFS data analysis round robin, March 2000 (DARR-00). *NASA Tech. Memo. 2001-206892, Vol. 15*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 71 pp.

Jeffrey, S.W., R.F.C. Mantoura, and T. Bjornland, 1997: Data for the identification of 47 phytoplankton pigments. In: *Phytoplankton Pigments in Oceanography*, S.W. Jeffrey, R.F.C. Mantoura, and S.W. Wright, Eds., UNESCO Publishing, France, 449-559.

Maritorena, S., D.A. Siegel, and A.R. Peterson, 2002: Optimization of a semianalytical ocean color model for global-scale applications. *Appl. Opt.*, 41, 2705-2714.

McClain, C.R., M.L. Cleave, G.C. Feldman, W.W. Gregg, and S.B. Hooker, 1998: Science quality SeaWiFS data for global biosphere research. *Sea Technol.*, 39, 10-15.

McClain, C.R., and G.S. Fargion, 1999a: SIMBIOS Project 1998 Annual Report. *NASA Tech. Memo. 1999-208645*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 105 pp.

McClain, C.R., and G.S. Fargion, 1999b: SIMBIOS Project 1999 Annual Report. *NASA Tech. Memo. 1999-209486*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 128 pp.

McClain, C.R., W. Esaias, G. Feldman, R. Frouin, W. Gregg and S. Hooker, 2002: The Proposal for the NASA Sensor Intercalibration and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Program, 1995. *NASA Tech. Memo. 2002-210008*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 54 pp.

Meister, G., P. Abel, R. Barnes, J. Cooper, C. Davis, M. Godin, D. Goebel, G. Fargion, R. Frouin, D. Korwan, R. Maffione, C. McClain, S. McLean, D. Menzies, A. Poteau, J. Robertson, and J. Sherman, 2002: The first SIMBIOS radiometric intercomparison (SIMRIC-1), April – September 2001. *NASA Tech. Memo. 2002-210006*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

Moore, T.S., J.W. Campbell, and H. Feng, 2001: A fuzzy logic classifications scheme for selecting and blending satellite ocean color algorithms. *IEEE T. Geosci. Remote*, 39, 1764-1776.

Morel, A., and L. Prieur, 1977: Analysis of variations in ocean color. *Limnol. Oceanogr.*, 22, 709-722.

Morel, A., and S. Maritorena, 2001: Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.*, 106, 7163-7180.

Mueller, J.L., 1993: The first SeaWiFS intercalibration round-robin experiment, SIRREX-1, July 1992. *NASA Tech. Memo. 104566, Vol. 14.*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

Mueller, J.L., and R.W. Austin, 1995: Ocean Optics Protocols for SeaWiFS Validation, Revision 1. *NASA Tech. Memo. 104566, Vol. 25*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 65 pp.

Mueller, J.L., 2002: In-water radiometric profile measurements and data analysis protocols. In: Mueller, J.L. and 39 co-authors, 2002: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Part 1. *NASA Tech. Memo. 2002-210004/Rev3-Vol1*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 123-137.

Mueller, J.L., C. Pietras, S.B. Hooker, D.K. Clark, A. Morel, R. Frouin, B.G. Mitchell, R.R. Bidigare, C. Trees, J. Werdell, G.S. Fargion, R. Arnone, R.W. Austin, S. Bailey, W. Broenkow, S.W. Brown, K. Carder, C. Davis, J. Dore, M. Feinholz, S. Flora, Z.P. Lee, B. Holben, B.C. Johnson, M. Kahru, D.M. Karl, Y.S. Kim, K.D. Knobelspiesse, C.R. McClain, S. McLean, M. Miller, C.D. Mobley, J. Porter, R.G. Steward, M. Stramska, L. Van Heukelem, K. Voss, J. Wieland, M.A. Yarbrough, and M. Yuen, 2002a: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Part 1. *NASA Tech. Memo. 2002-210004/Rev3-Vol1*, Mueller, J.L., and G.S. Fargion, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 137 pp.

Mueller, J.L., C. Pietras, S.B. Hooker, D.K. Clark, A. Morel, R. Frouin, B.G. Mitchell, R.R. Bidigare, C. Trees, J. Werdell, G.S. Fargion, R. Arnone, R.W. Austin, S. Bailey, W. Broenkow, S.W. Brown, K. Carder, C. Davis, J. Dore, M. Feinholz, S. Flora, Z.P. Lee, B. Holben, B.C. Johnson, M. Kahru, D.M. Karl, Y.S. Kim, K.D. Knobelspiesse, C.R. McClain, S. McLean, M. Miller, C.D. Mobley, J. Porter, R.G. Steward, M. Stramska, L. Van Heukelem, K. Voss, J. Wieland, M.A. Yarbrough, and M. Yuen, 2002b: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Part 2. *NASA Tech. Memo. 2002-210004/Rev3-Vol2*, Mueller, J.L, and G.S. Fargion, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 171 pp.

O'Reilly, J.E., S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M. Kahru, and C. McClain, 1998: Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.*, 103, 24937-24953.

Reynolds, M.R., M.A. Miller, and M.J. Bartholomew, 2001: A fast-rotating, spectral shadow-band radiometer for marine applications. *J. Atmos. Ocean. Tech.*, 18, 200-214.

Schwarz, J.N., P. Kowalczuk, S. Kaczmarek, G.F. Cota, B.G. Mitchell, M. Kahru, F.P. Chavez, A. Cunningham, D. McKee, P. Gege, T. Kishino, D.A. Phinney, and R. Raine, 2002: Two models for absorptions by coloured dissolved organic matter (CDOM). *Oceanologia*, 44, 209-241.

Werdell, P.J., S.W. Bailey, and G.S. Fargion, 2000a: Appendix B: SeaBASS data file format. In: Fargion, G.S. and J.L. Mueller, 2000a: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 2. *NASA Tech. Memo. 2000-209966*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 176-181.

Werdell, P.J., S.W. Bailey, and G.S. Fargion, 2000b: SeaBASS data protocols and policy. In: Fargion, G.S. and J.L. Mueller, 2000a: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 2. *NASA Tech. Memo. 2000-209966*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 170-172.

Werdell, P.J., S.W. Bailey, and G.S. Fargion, 2002a: Appendix B: SeaBASS data file format. In: Mueller, J.L. and 39 co-authors, 2002a: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Part 2. *NASA Tech. Memo. 2002-210004/Rev3-Vol2*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 292-298.

Werdell, P.J., S.W. Bailey, and G.S. Fargion, 2002b: SeaBASS data protocols and policy. In: Mueller, J.L. and 39 co-authors, 2002a: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Part 2. *NASA Tech. Memo. 2002-210004/Rev3-Vol2*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 284-287.

Zibordi, G., and G.M. Ferrari, 1995: Instrument self-shading in underwater optical measurements: experimental data. *Appl. Opt.*, 34, 2750-2754.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 2002

3. REPORT TYPE AND DATES COVERED
Technical memorandum

4. TITLE AND SUBTITLE
The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS): Current Architecture and Implementation

5. FUNDING NUMBERS
970.2

6. AUTHOR(S)
P. Jeremy Werdell and Sean W. Bailey
Giulietta S. Fargion and Charles R. McClain, Editors

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PEFORMING ORGANIZATION REPORT NUMBER
2002-03543-0

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS (ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
TM—2002–211617

11. SUPPLEMENTARY NOTES
G.S. Fargion, Science Applications International Corp., Beltsville, Maryland; P. J. Werdell, Science System and Applications Inc., Lanham, Maryland; S.W. Bailey, Futuretech Corporation, Greenbelt, Maryland

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified–Unlimited
Subject Category: 48
Report available from the NASA Center for AeroSpace Information, 7121 Standard Drive, Hanover, MD 21076-1320; (301) 621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Satellite ocean color missions require an abundance of high-quality in situ measurements for bio-optical and atmospheric algorithm development and post-launch product validation and sensor calibration. To facilitate the assembly of a global data set, the NASA Sea-viewing Wide Field-of-view (SeaWiFS) Project developed the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), a local repository for in situ data regularly used in their scientific analyses. The system has since been expanded to contain data sets collected by the NASA Sensor Intercalibration and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project, as part of NASA Research Announcements NRA-96-MTPE-04 and NRA-99-OES-99. SeaBASS is a well moderated and documented archive for bio-optical data with a simple, secure mechanism for locating and extracting data based on user inputs. Its holdings are available to the general public with the exception of the most recently collected data sets. Extensive quality assurance protocols, comprehensive data and system documentation, and the continuation of an archive and relational database management system (RDBMS) suitable for bio-optical data all contribute to the continued success of SeaBASS. This document provides an overview of the current operational SeaBASS system.

14. SUBJECT TERMS
SeaWiFS, SIMBIOS, SeaBASS, bio-optical data, biological oceanography, bio-optical algorithms, database, remote sensing, satellite validation.

15. NUMBER OF PAGES
45

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL