World-wide ocean optics database (WOOD)

Jeffrey H. Smart
Johns Hopkins University • *Laurel, Maryland USA*

Abstract

A very large volume of vertical profiles of optical properties, nutrients, and chlorophyll has been acquired in numerous data formats and from many institutions around the world. These data have been converted to a single data format, loaded into a relational Oracle© database, and made available to the public over the internet at http://wood.jhuapl.edu. The database also includes related oceanographic data such as temperature, salinity, and water depth. Users can search by parameter name, location, date, wavelength, and water depth. A graphical user interface is provided to access and display the data. On-screen data profile plots are linked to a chart of the data locations which includes coastlines.

Introduction

The Ocean Biogeographical Information System (OBIS) is intended to provide an internet accessible database with a graphical user interface (GUI) that is similar to that already provided by World-wide Oceans Optics Database (WOOD). This paper describes the design and functionality of WOOD, and it is hoped that this information will be useful to the design of OBIS. In addition, OBIS is being developed to support the Census of Marine Life, and the chlorophyll and nutrient data in WOOD are fundamental parameters to study because all marine life depends on them in one way or another.

Millions of dollars and thousands of hours of at-sea time have been spent collecting biological and optical oceanographic data, but until recently, very few organizations have made an effort to properly archive the data. When the data have been saved, they generally are kept in the originator's custom format, and often on media which all too rapidly become obsolete. Recently several government agencies and universities have begun to realize the wisdom and value of properly archiving their bio-optical data. For example, NASA has set up a password-controlled internet data bank called SEABASS which allows their Principal Investigators (PIs) shared access of optics-related data.

Recently several government agencies and universities have begun to realize the wisdom and value of properly archiving their bio-optical data.

Similarly, NOAA's Coastal Remote Sensing (CSC) office has been publishing their data via an internet homepage (see http://www.csc.noaa.gov/crs). In these examples, the data are all available in files having a standardized format, and they can be downloaded using FTP. However, this process can become very time consuming when one has to download many files and then manually extract the parameters, locations, and dates of interest. Moreover, if one wants data from a variety of web sites, custom software must be developed to read each site's data format.

Design of WOOD

A more comprehensive approach is to store all available data in a master relational database that utilizes general purpose processing software to extract, display, and analyze its contents. This is the approach taken in developing the World-wide Ocean Optics Database (WOOD). Major data archives were acquired from Bedford Institute of Oceanography (see Longhurst et al., 1995), from NOAA's chlorophyll profile datasets, from websites (NOAA/CSC, JGOFS, etc.), and from individual PIs. Oceanographers and database special-

---

TABLE I
Metadata parameters associated with each measurement station

Julian Date (0-366)	Parameter Recorded*
Time (0000-2359)	Instrument Type*
Time Type*: 0 = unknown,	Center Wavelength (in nm)
1 = GMT, 2 = Local	Bandwidth (FWHM in nm)
Latitude & Longitude	Angle (θ in degrees,
(decimal degrees)	0 if not applicable)
Originating Agency/University*	Vertical Resolution (in meters)
Dissemination Code*	Data Editing & Quality Flags
Ocean Bottom Depth (in meters)	

*Stored as an integer data flag. See the "Metadata Info" frame at the web site.

ists were surveyed to determine the optimum set of metadata to store with the actual measurements in order to optimize the utility of the data. Meetings were also held with NOAA and the U.S. Naval Oceanographic Office to discuss the database design and to finalize its key components. As a result of the survey and meetings, it was decided to store the parameters listed in Table 1 with each data measurement.

All data are stored as depth profiles with depths in meters. Currently time series at constant depth are stored as single point "profiles." The lookup tables for deciphering metadata stored as integer flags (i.e. parameter type, data units, data quality, data editing, cruise ID, etc.) are found under the "Metadata Info" frame at the WOOD website. The present list of variables intended to be included in WOOD are given in Table 2. The database design would permit the addition of other parameters as desired.

The "Metadata Info" frame at the WOOD web site provides links to tables that contain the flags associated with:

1. Instrument types (e.g. for various radiometers, chlorophyll measuring techniques, beam transmissometers, backscatter meters, and absorption meters)
2. Data editing applied to the data (raw, spike edited, spike edited and low pass filtered, etc.)
3. Data quality (a somewhat subjective value based on known instrument calibration, drifts, noise, etc.).
4. Cruise ID numbers and Author Codes

The "Ocean Bottom Depth" metadata parameter is not a value obtained from the contributor of the profile data. (In most cases researchers do not collect bottom depth.) Rather, it comes from accessing the worldwide 5 nautical mile resolution Digital Bathymetric Data Base (DBDB5) and finding the closest point to the location of the profile. (In some cases the DBDB5 value is -10, which is a flag for missing data. Other values are all positive.) Occasionally the actual water depth exceeds that stored in DBDB5, so the profile may extend "beyond" the expected bottom. Similarly, especially close to a coast, the DBDB5 value may overestimate the true depth. In the future we hope to replace the 5 nautical mile resolution water depths with data

from a finer resolution database.

Oracle© Version 7.3.3 database software was selected to store all metadata and bio-optics-related data. The Oracle Webserver software Version 2.1 was selected to interface the database to the internet, and the entire system is installed on a Pentium PC running under the Windows NT Server operating system. This system allows the user to request data from one or many parameters simultaneously, using simple menus to select the desired variables and range of locations, dates, water depth, and wavelength. In order to greatly increase access time, the profile data are stored as string objects. That is, depth values are stored as one string, and the data parameter values as another. As a result, each string is treated as a single object, rather than a large collection of them. (This approach also means that only one set of metadata is needed for each profile. In the past, other Oracle© programmers have created databases wherein each depth/data pair is stored as a unique object, thus requiring duplication of the metadata for each point in the profile.) To facilitate parsing out the individual values after extracting data from

TABLE 2
Parameters to be included in the optics database
(Data are not necessarily available for all parameters at this time)

Flag	Parameter & Units	Acronym
1	Absorption Coefficient (/m)	a
2	Scattering Coefficient (/m)	b
3	Beam Attenuation Coefficient (/m)	c
4	Diffuse Attenuation Coefficient (/m)	K_d
5	Backscatter (/m)	b_b
6	Angular Backscattering Coefficient (/SR/m * 10^{-4})	$\beta(\theta)$
7	Downwelling Irradiance ($\mu W/cm^2$)	E_d
8	Upwelling Irradiance ($\mu W/cm^2$)	E_u
9	Scalar Irradiance ($\mu W/cm^2$)	E_o
10	Upwelling Radiance ($\mu W/SR/cm^2$)	L_u
11	Radiance Reflectance	$R_I = L_u/E_d$
12	Pigment Concentration ($\mu g/l$)	$C = Chl\ a + Phaeophytin$
13	Chlorophyll a ($\mu g/l$)	Chl a
14	Chlorophyll b ($\mu g/l$)	Chl b
15	Chlorophyll c ($\mu g/l$)	Chl c
16	Phaeophytin ($\mu g/l$)	Phaeo
	Related Variables of Potential Interest:	
17	Nutrients (NO_2, NO_3, PO_4, SiO_4, NH_4 etc.)	N
18	Oxygen (ml/l)	O_2
19	Temperature (°C)	T
20	Salinity (ppt)	S
21	Sigma T	σ_T
22	BV Frequency (cph)	BV
23	Bioluminescence (photons/s/cc x 10^{-8})	Biolum
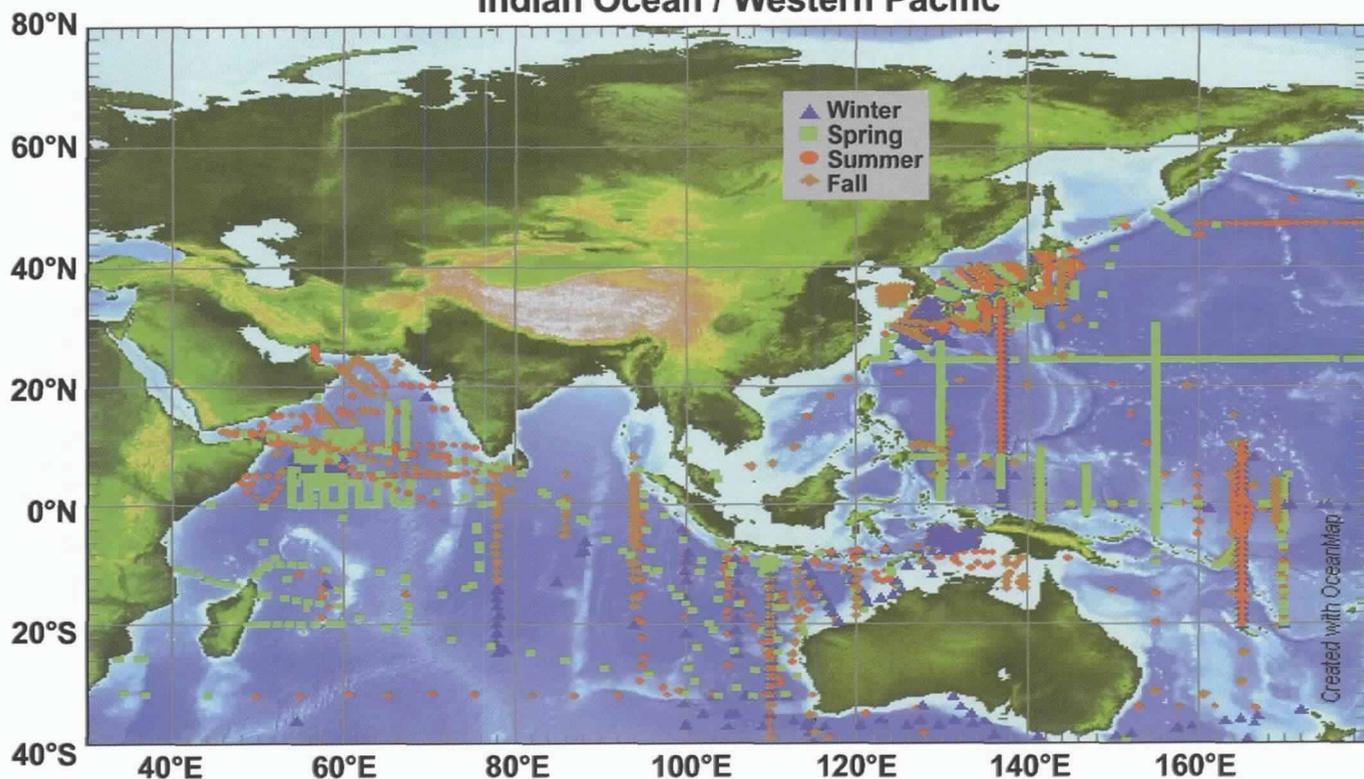
Figure 1. Example of WOOD K_d Locations in the Indian and Western Pacific Oceans.

WOOD, a space is stored in the strings between each value. The only problem with storing profiles as strings is that the present version of Oracle© limits strings to 2000 bytes. (As discussed later, Oracle© 8 removes this limitation.) This restriction means that a maximum of about 333 values could be stored per profile. Only a small percentage of the original data now stored in WOOD exceeded this limit (most were finely sampled fluorometer profiles). Nevertheless, it was important to preserve the information contained in these original data, so several special pre-processing programs were written that solved the problem. For high quality (noise free) densely sampled K_d or chlorophyll profiles, an inflection point sampling routine was used that ensured resultant accuracy of 0.005 (units of m^{-1} and mg/m^3, respectively.) For noisier fluorometer data, a 1-m interpolation routine was applied.

Results and Applications

The WOOD website has been operational since May 1997 and has been providing scientists around the world with a "one stop shopping" site to quickly and easily extract optics data from specific regions of interest. Over 57,000 profiles of K_d, primarily at 490 nm, have been loaded, and Figure 1 shows an example of their locations and seasonal distribution. (The only problem with overlaying symbols for multiple profiles at a given location is that the underlying symbols are obscured by the last one written to the image file.) The database is accessed virtually every week at JHU/APL for such diverse applications as future field test planning, optical vulnerability studies, mine warfare utility assessments, and optical climatology validations.

Each data parameter (e.g. chlorophyll, the diffuse attenuation coefficient, "K_d", backscatter, "$\beta(170°)$", and the beam attenuation coefficient, "c") is stored in a separate master file that is part of the relational database. The user can choose what data to retrieve from WOOD by selecting a desired wavelength and range for latitude/longitude, year, Julian date, and water depth. The user has the option to receive the complete set of metadata listed in Table 1 or a "Standard" set which includes only the most essential data. Multiple parameters can be retrieved simultaneously. In short, the complete database is readily accessible and easy to search.

Many investigations of bio-optical characteristics require supporting data such as profiles of nutrients, temperature, salinity, and water depth, and WOOD is designed to provide all of these kinds of data. At present, however, very little data have been received containing simultaneous measurements of each one of these parameters. Therefore, until institutions and PIs begin to provide such multi-parameter datasets, the usefulness of WOOD for answering questions about the controlling mechanisms for the observed optical

March 1984 K_d Profile Locations

Figure 2. Example of WOOD used to assess spatial variability in K_d off Southern California.
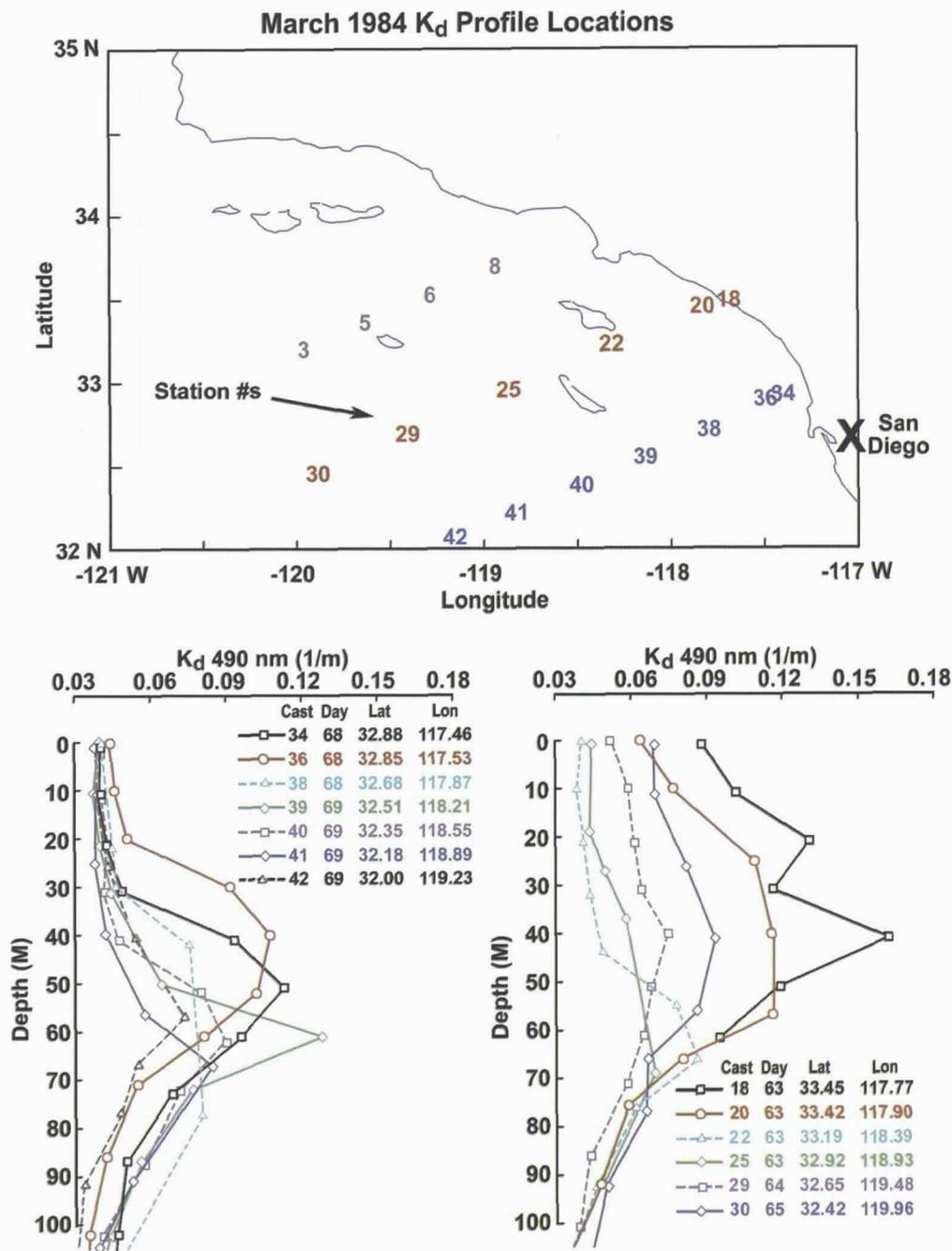
Notes: Data from March 1984. The bottom panels show K_d profiles from Stations # 34-42 (left) and from # 18-30 (right)

conditions will be hindered. Meanwhile, a primary use for WOOD will be the generation of regionally-specific empirical fits that capture the monthly or seasonal optical characteristics of particular locations. For example, Longhurst et al. (1995) have already used large portions of the data contained in WOOD to provide a world-wide chlorophyll profile climatology. A similar approach is frequently carried out on small-

er geographic regions at APL. For example, while planning a proposed Navy field test, the local variability of optical profiles was examined off the coast of Southern California in March, as shown in Figure 2. In this example, within about 100 nautical miles of the coast, K_d values in the upper 40 m were generally significantly higher as one progressed north from San Diego. (The WOOD Data were examined for numer-

TABLE 3
Example of level of detail associated with WOOD data parameter flags.
The Morel88 algorithm is described in Morel, 1988.

Flag	Description
41	K_d from Sunlight
42	K_d from Moon/Starlight (applies to photomultiplier tube systems)
43	K_d from Irradiance-calibrated absorption meter
44	K_d from bottle sampled total pigment using Morel88 algorithm
45	K_d from Lidar (K_{sys})
46	K_d from Sunlight-calibrated Fluorometer
47	K_d from CZCS
48	K_d from SeaWiFS
49	reserved for future assignment
430	K_d from AC-9 using a,c 715 nm to correct for scattering effects
431	K_d from AC-9 using no scattering corrections
440	K_d from calibrated fluorometer chlor_a using Morel88 algorithm
441	K_d from chl_a only using Morel88 algorithm

ous years of March data; Figure 2 is a typical set of data from 1994.)

Conclusions And Future Work

A very large volume of bio-optics-related data exist at many oceanographic institutions around the world. Even within a given institution, numerous data formats and media have been used to archive the data. The primary goal of this work has been to establish a unified database that collects, in one place and in one format, as many of these data sets as possible. The database is available as an international resource and has proven to be useful for a variety of applications, such as field test planning, model development and validation, and area characterizations. The database is also an invaluable resource to the U.S. Navy for applications such as detectability assessments, tactical guidance, and environmental predictions.

The volume of oceanographic optics data is rapidly increasing. To maximize the utility and availability of these data, it is vital that funding agencies put more emphasis on requiring that PIs contribute their data to unified, quality-controlled databases like the WOOD.

In the future there are several improvements that will be added to WOOD. First, the database must be frequently updated with data from recent at-sea mea-surements. Eventually it may even include worldwide coverage from satellites, such as Coastal Zone Color Scanner (CZCS) and SeaWIFS, and climatology data products (such as those developed by Longhurst et al., 1995) that combine remote sensing and in situ data. Second, a new class of data will be generated called "derived parameters." These derived data will consist of optical parameters that have been estimated from existing WOOD data using published empirical algorithms (such as those provided in Morel, 1988) that relate one variable to another. Third, another feature being developed is the collation of multi-parameter data so that simultaneous measurements are provided in blocks. (At present such data requests are serviced by displaying all of the data from one parameter, then all of the data from a second parameter, etc.) Fourth, the database is being upgraded to Oracle© Version 8 in order to eliminate several of the Version 7 limitations, including the limit of only 2000 bytes per string array. Finally, a more concerted effort is underway to obtain datasets from foreign data centers such as the British Ocean Data Center (BODC), the Japanese Ocean Data Center (JODC), etc.

Acknowledgements:

REFERENCES

Longhurst, A., et al., 1995: An estimate of global primary production in the ocean from satellite radiometer data. *J. Plankton Res.*, 17(6), 1245-1271.

Morel, A., 1988: Optical Modeling of the Upper Ocean in Relation to its Biogenous Matter Content (Case I Waters). *JGR*, 93(C9), 10, 749-10, 768.