Interactive analysis and visualization of geoscience data with Ocean Data View [☆]

Reiner Schlitzer

Alfred Wegener Institute for Polar and Marine Research, Postfach 120161, 27515 Bremerhaven, Germany

Abstract

Ocean Data View (ODV) is a freeware package for the interactive exploration and graphical display of multi-parameter profile or sequence data. Although originally developed for oceanographic observations only, the underlying concept is more general, and data or model output from other areas of geosciences, like for instance geology, geophysics, geography and atmospheric research can be maintained and explored with ODV as well. The data format of ODV is designed for dense storage and direct data access, and allows the construction of very large datasets, even on affordable and portable hardware. ODV supports display of original data by colored dots or actual data values at the measurement locations. In addition, two fast and reliable variable-resolution gridding algorithms allow color shading and contouring of gridded fields along sections and on general 3D surfaces. A large number of derived quantities can be selected and calculated online. These variables are displayed and analyzed in the same way as the basic variables stored in disk files. ODV runs on PCs under Windows and on UNIX workstations under SUN Solaris. The software and extensive sets of coastline, topography, river-, lake- and border outlines as well as various gazetteers of topographic features are available at no cost over the Internet. In addition, the electronic atlas *eWOCE* that consists of oceanographic data from the World Ocean Circulation Experiment (WOCE) is also available free of charge over the Internet. A gallery of prepared plots of property distributions along WOCE sections provides a quick overview over hydrographic, nutrient, oxygen and transient tracer fields in the ocean and, apart from the scientific use for oceanographic research, can serve as tutorial material for introductory or advanced courses on oceanography.

1. Introduction

In the context of climate research and earth science, in general, there is a pressing need to better understand and quantify on a global scale the ongoing processes and property transports within and between the different compartments of the earth system. In the fields of oceanographic and atmospheric research, this recognition has led to the planning and implementation of observational programs which—in contrast to efforts in the past—are of global extent and for which the pooling and integration of individual datasets into large and comprehensive data resources is of greatest importance. An example of such a large observational program is the internationally coordinated World Ocean Circulation Experiment (WOCE), in which more than 30 countries participated and provided support for the measurement of many oceanographic parameters by means of ships, satellites or autonomous instruments. In the meantime, WOCE has been completed and a wealth of high-quality, multi-parameter datasets has been obtained. Large parts of this dataset are now publicly available and are being used for general oceanographic research and climate studies.

However, whereas the willingness of scientists to share data has increased considerably during the last decade, what is still lagging behind is the development and release of software tools for the optimal management and the flexible exploration and graphical analysis of the growing amount of available data. As a consequence, widespread use of the combined WOCE dataset, for example, is still hampered because of rather complex data formats and because of tedious and difficult search procedures when locating and subsetting data.

To facilitate the use of large, multi-parameter datasets like the WOCE data resource, and to provide easy-to-use yet powerful and fast methods for the quality control and graphical exploration of these data was the main motivation and objective for the development of the Ocean Data View (ODV) software, which is described in this paper. For WOCE, ODV was used to combine all currently available data of the WOCE hydrographic and upper ocean thermal programs (WHP and UOT) into an integrated dataset. When used with ODV, these datasets constitute an "electronic atlas of WOCE data" (*eWOCE*) that allows graphical display and interactive analysis of the data in many different ways (Schlitzer, 2000).

As the name implies, ODV was originally developed for the handling of oceanographic data, and the focus in this paper is mainly on marine data. It should be noted, however, that the underlying data model and the interactive analysis methods of ODV are more general, and scientists from other fields of earth sciences can benefit from ODV and analyze and explore other data types as well. Examples of such applications are presented in Section 4. ODV interfaces with the Pangaea database system (Diepenbroek et al., 2002), which holds and serves over the Internet a wide variety of geosciences datasets.

2. ODV software

2.1. Overview

ODV is a software package for the exploration, quality control and graphical analysis of irregularly distributed profile data that is available over the Internet at no cost for Windows PCs and SUN Solaris work-stations.[1] ODV is programmed in the C language and provides a maximum of execution power which is required for the fast gridding, shading and contouring of large datasets consisting of even hundred thousands of data points. ODV lets users interactively browse large sets of station data and produce high-quality station maps, general property–property plots of one or more

stations, scatter plots of selected stations, property sections along arbitrary cruise tracks, and property distributions on general 3D surfaces. ODV supports display of original data as colored dots or actual data values. In addition, two fast gridding algorithms allow color shading and/or contouring of gridded fields along sections and on surfaces (any plot with a Z variable can be gridded, color-shaded and/or contoured). A large number of derived quantities can be calculated quickly and can be displayed and analyzed in the same way as the basic variables stored on disk.

ODV is designed to be flexible and easy to use. Users are not required to know the details of the internal data storage format nor are they required to have programming experience. ODV always displays a map of available stations on the screen and facilitates navigation through the data by letting the user select stations, sections, and isosurfaces with the mouse. The screen layout and various other configuration features can be modified easily, and favorite settings can be stored in configuration files on disk for later use.

ODV can create and manage very large data collections. In the case of oceanography, it is possible to maintain the entire set of available global historical hydrographic data and to add newly arriving data. All this can be done on relatively inexpensive and widely available desktop or notebook computers. The ODV data collections are then ready for scientific analysis in the field or back home in the laboratory. In addition to the actual research applications, ODV is useful for data quality evaluation and for teaching and training. A review of the ODV software can be found in Brown (1998).

2.2. Modes

ODV can operate in five different modes thereby providing different analysis methods and display types commonly used in the scientific community.

MAP mode is intended for full-page station maps. This mode can be used to produce high-quality station or cruise maps of specific regions or the whole globe. ODV allows a choice between five map projections and provides bathymetry and land topography information as well as the outlines of rivers, lakes, sea-ice extent and national borders. This information is available at different levels of resolution and can be used by the user to compose various types of context maps according to specific needs. Individual stations can be high-lighted and arbitrary annotations can be added. As with ODV graphics in general, the maps can be printed on color or black and white printers or can be written to GIF, EMF, or postscript files for further processing. To view an example cruise map see.[2]

[1] Ocean Data View. 2001. http://www.awi-bremerhaven.de/GEO/ODV

[2] ODV Cruise Map. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_map.gif

STATION mode (like all following modes) provides a station map and one or more data plot windows. This mode is appropriate to produce X/Y property/property plots for one or more selected stations or profiles and to study differences between the stations.[3] The stations to be plotted can be selected using different methods, the simplest way being just to click on a station with the mouse.

In *SCATTER* mode the data of all stations shown in the map are displayed in the data plots. This provides an overview over all data from a given region, a specific cruise or a specifically selected station subset and is particularly useful for data quality checking. SCATTER mode (like all following modes), supports Z variables in addition to the X and Y variables. The value of a Z variable at a given X/Y point is displayed by ODV either by showing the actual numerical value or by filling with a value-dependent color. Plots with Z variables (this holds also for SECTION and SURFACE plots, see below) can be displayed in two ways: (1) by placing colored dots at the X/Y locations or (2) as continuous gridded fields estimated on the basis of the observed data. Gridded fields (see below for a description of the gridding algorithms) can be color-shaded and/or contoured.[4]

SECTION mode also supports Z variables on data plots and allows all plot types of the SCATTER mode, but the set of stations used for the plots is restricted to a section band usually following the given cruise tracks. Section bands can be defined arbitrarily and their width can be adjusted in order to properly select a set of stations. SECTION mode is appropriate to present property distributions and property/property plots for all stations along entire cruises[5] and to calculate and investigate geostrophic velocities perpendicular to the cross-section of a cruise.[6]

SURFACE mode allows the specification of surfaces in three-dimensional space (longitude/latitude/depth) defined as points of constant values of a given variable, like for instance, depth, density, or temperature, and then displays property distributions of other variables on this surface.[7] SURFACE mode also allows to draw arbitrary property/property plots for the given surface.

2.3. Data import and export

The ODV data format provides dense storage and allows instant access to any profile or sequence, even in large data collections. The data format is flexible and allows storage of data for up to 50 variables, where type and number of the variables is arbitrary and may vary from one collection to another. Typically, only about 1 MB of disk space is required to store about 1600 oceanographic bottle stations containing data for seven variables.

ODV allows easy import of new data into collections and also allows easy export of some or all the data from a collection. Hydrographic data in WOCE WHP format,[8] data from the World Ocean Atlas 1994,[9] the World Ocean Database 1998,[10] data in NODC SD2 format, and data in an ODV specific TAB-separated spreadsheet format can directly be incorporated into the ODV system. The ODV spreadsheet format is easy to use and is the preferred data import format for non-oceanographic datasets. ODV maintains quality flags associated with each individual data value. These quality flags can be used by ODV as a data quality filter to exclude bad or questionable values from the analysis.

2.4. Derived variables

In addition to the basic oceanographic variables stored in the data files, ODV can calculate and display a large number of derived variables. These derived variables are either coded in the ODV software (potential temperature, potential density, dynamic height, and many others) or are specified in user-defined macro files or "on-the-fly" expressions. The macro language is easy and general enough to allow a large number of applications. Use of macro files for new derived quantities broadens the scope of ODV considerably and allows easy experimentation with new quantities not yet established in the scientific community. A separate macro editor that can be invoked from ODV facilitates creation and modification of ODV macros. Any basic or derived variable can be displayed in ODV plots, and they all can be used to define 3D surfaces. Examples of 3D surfaces are depth horizons, isopycnals (surfaces of constant potential density), isothermals, isohalines, or property minimum or maximum layers (e.g., the intermediate water salinity minimum layer) defined by the zero-crossing of the vertical derivative (a derived quantity) of these variables.

[3] ODV STATION Plot. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_prop.gif

[4] ODV SCATTER Plot. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_scatter.gif

[5] ODV SECTION Plot. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_section.gif

[6] ODV Plot of Geostrophic Velocities. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_geostrophic.gif

[7] ODV SURFACE Plot. 2001. http://www.awi-bremerhaven.de/GEO/ODV/odv_surface.gif

[8] WOCE Hydrographic Data. 2001. http://whpo.ucsd.edu/whp_data.html

[9] World Ocean Atlas. 1994. http://www.nodc.noaa.gov/OC5/pr_woa4.html

[10] World Ocean Database. 1998. http://www.nodc.noaa.gov/OC5/indprod.html

2.5. Gridding algorithms

In addition to the simple display types of showing the actual numerical values or colored dots at the sample positions, ODV can also produce color-shaded and/or contoured gridded property distributions. For the gridding process ODV has two gridding algorithms built-in: *Quick Gridding* and *VG Gridding*. As the name implies, *Quick Gridding* is a fast method suitable for cases of good data coverage and yields results in a matter of seconds even for large datasets of several hundred thousand points. The underlying method is a weighted averaging scheme with separate user-supplied averaging length-scales in X and Y directions. To achieve high performance, the data are tile-sorted before the averaging process, and for the estimation at a given X/Y position only data values from a small neighborhood of the point are actually used for the averaging.

For poor or inhomogeneous data coverage *VG Gridding* is to be preferred over the *QuickGridding* method.

In contrast to *Quick Gridding* that uses an equidistant, rectangular grid for the estimation, *VG Gridding* analyzes the distribution of the data points and constructs a variable resolution, rectangular grid, where grid-spacing along X and Y directions varies according to data density. High resolution (small grid-spacing) is provided in regions with good data coverage, whereas in areas of sparse sampling the grid is coarse and resolution is limited. For typical hydrographic sections, for instance, this procedure leads to higher spatial resolution in the upper water column and in boundary current regions (data coverage is usually very good in these areas) as compared to the deep, open ocean regions.

After construction of the grid, the property under consideration (temperature, salinity, etc.) is estimated at every grid-point by applying a weighted-average scheme using data values from a neighborhood of the grid-point. Weights decrease with increasing distance from the grid-point, and separate (e-folding) length-scales in X and Y directions specified by the user are applied. Averaging length-scales are proportional to grid-spacing, e.g., in areas of higher grid resolution (upper water column, boundary currents, etc.) smaller averaging length-scales are used automatically. This overall approach allows to resolve small-scale features in areas of dense data coverage and at the same time provides smooth and stable fields in other regions where observations might be sparse. Special methods are implemented that increase execution speed and allows field estimations within a few seconds even for thousands of data points. Once a property field has been estimated, the results are passed to shading and contouring routines for display of the fields on the screen or on a printer.

2.6. Gazetteers of geographical features

ODV can be used to identify undersea features like seamounts, ridges, fracture zones, troughs, basins or other geographical features in the ocean or on land. Currently, the ODV distribution contains (1) the GEBCO gazetteer of undersea features, (2) a gazetteer of oceanic features approved by the United States Board on Geographic Names (BGN), (3) a gazetteer of river names, (4) a gazetteer of the world's largest cities and (5) a gazetteer of oceanographic sections occupied during the WOCE experiment. In gazetteer mode, ODV loads the information from the respective gazetteer file and marks the feature locations in the map. Moving the mouse close to a feature point invokes a popup window displaying the name of the feature. Feature subsets can be produced by specifying feature type and/or feature name substrings. The format of the ODV gazetteer files is documented and users can extent existing gazetteers or produce new ones.

3. eWOCE atlas

To facilitate the use of the global WOCE dataset, all data released by the WOCE Hydrographic Programme (WHP) have been compiled into an integrated dataset. When used with the ODV visualization software for Windows and SUN Solaris, this dataset constitutes an *eWOCE* that permits graphical display and interactive analysis of the data in many different ways. With extensive interactive controls such as user-defined plot configuration, zooming, auto-scaling, color adjustment, and station/sample selection, this electronic atlas complements and surpasses printed atlases that are now in preparation.

More than 200 property distributions along WHP sections are provided with *eWOCE*. Starting from these template plots, users can easily produce arbitrary property/property plots, distributions on general iso-surfaces, property difference distributions between repeats, time-series plots, geostrophic velocity sections, and many other plot types.

eWOCE is part of the "WOCE Global Data, Version 2.0" CD set (WOCE Data Products Committee, 2000). The latest versions of the WHP bottle data and the ODV software are also available over the Internet (see footnote 1).[11]

3.1. Gallery

Plots of more than 200 property distributions along WHP lines are provided in the *eWOCE*

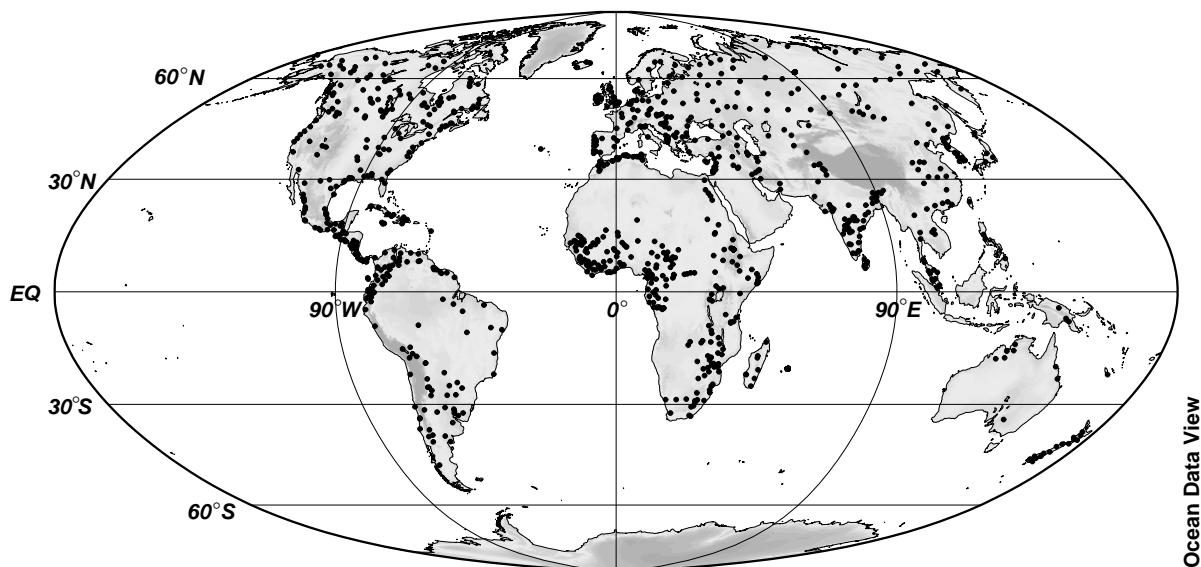[11] eWOCE Atlas. 2001. http://www.awi-bremerhaven.de/GEO/eWOCE

Fig. 1. Map of river discharge monitoring stations from RivDis1.1 dataset of Vörösmarty et al. (1998).

Gallery.[12] You view these plots with your Internet browser via easy to use interactive map interfaces, and there is no need to download the *eWOCE* data files or software first. Visit the *eWOCE* Gallery (see footnote 12), choose one of the ocean basins and point the mouse to one of the WHP section identifiers. Then choose a property from the list that appears.

3.2. Datasets

eWOCE provides global data collections of WHP bottle and CTD data (*WoceBtl* and *WoceCTD*) and basin-wide datasets from the Upper Ocean Thermal programme (*WoceAtlUOT*, *WocePacUOT*, *WoceIndUOT*).

The *WoceBtl* collection contains hydrographic, nutrient and tracer data for more than 14,000 stations. Most of these stations contain oxygen, phosphate, nitrate, and silicate data in addition to temperature and salinity. CFC observations are provided for about 5700 stations (about 41%), and almost 3000 stations contain data for carbon parameters. About 75% of the stations were occupied during the WOCE period between 1987 and 1998, while the rest of the stations are pre-WOCE and included for reference and for the analysis of temporal changes on decadal time scales. The *WoceCTD* collection currently contains high-resolution temperature and salinity data for more than 4300 stations.

The *eWOCE* collections *WoceAtlUOT, WocePacUOT,* and *WoceIndUOT* contain hydrographic data (mostly temperature) for the upper 500 to 800 m of the water column. There are more than 185,000 stations for the Atlantic, almost 425,000 stations for the Pacific and 63,000 for the Indian Ocean. Except for the Southern Ocean, the spatial and temporal coverage is excellent, and these datasets allow detailed investigations of climate variability during the 1990s. The collections *WoceAtlUOT, WocePacUOT,* and *WoceIndUOT* are very large and are only available on the "WOCE Global Data, Version 2.0" CD set (WOCE Data Products Committee, 2000).

4. Other ODV Applications

In addition to oceanographic data, ODV can handle profile, sequence or gridded data or model output from other fields of research. Three such examples are described below.

4.1. Global river discharge database

Vörösmarty et al. (1998) have compiled historical river discharge data for more than 1350 monitoring stations world-wide. These data have been imported into the ODV system and can be obtained from the author. Fig. 1 shows a map of the monitoring stations produced with ODV. Note that full color support is provided by ODV; here a black and white version was produced because of printing requirements. As an example of the rich information content of the discharge database, Fig. 2 shows runoff data for the Siberian river Ob which
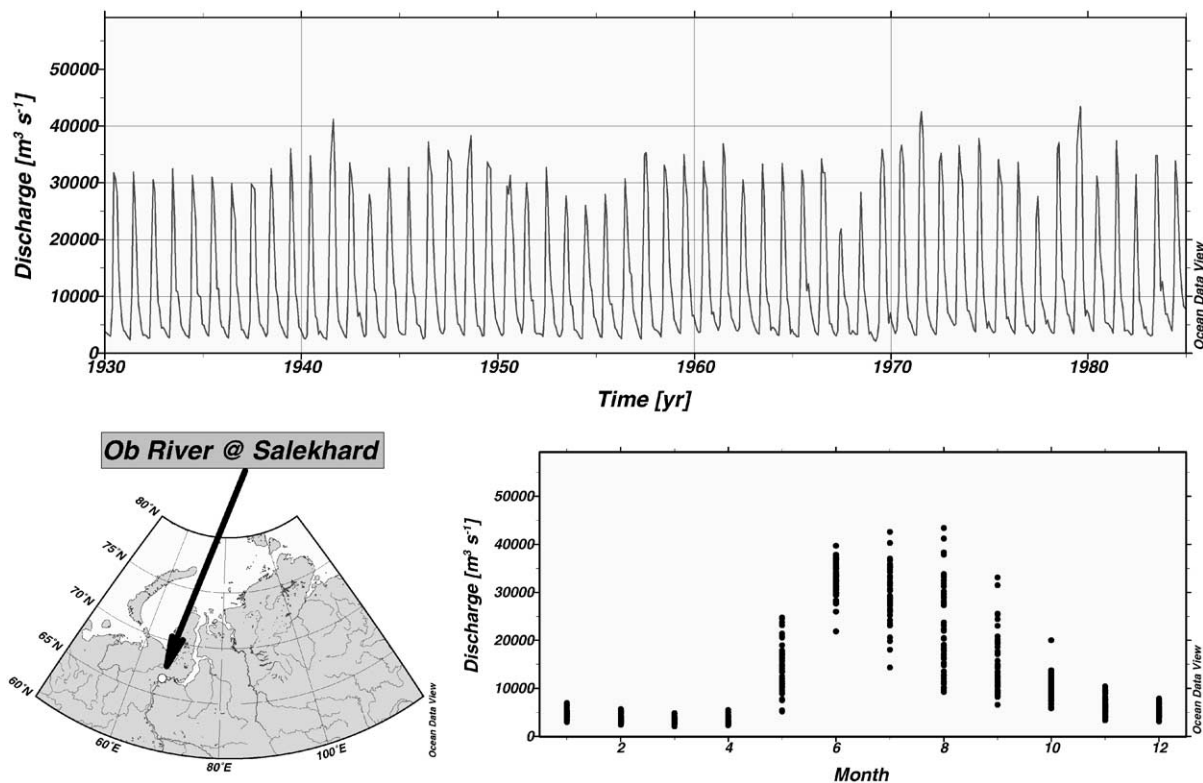
Fig. 2. Water discharge of river Ob at station Salekhard as function of time for period between 1930 and 1985 (data from Vörösmarty et al., 1998). Also shown is discharge versus month of year which reveals strong seasonality and interannual variability of runoff.
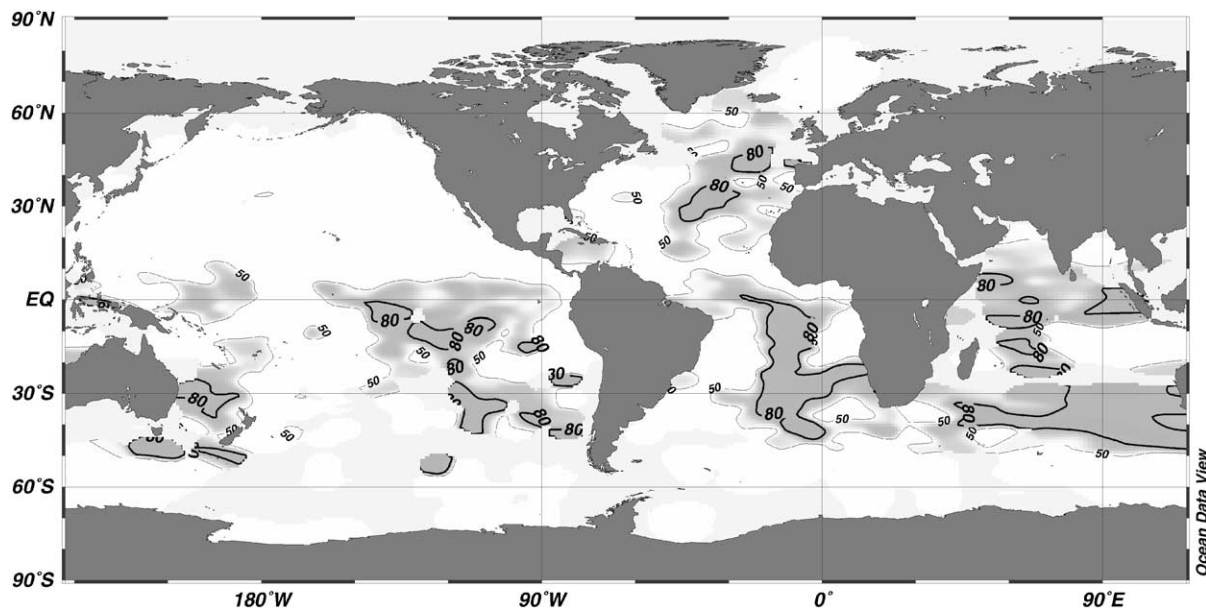

Fig. 3. Map of CaCO₃ content (in %) of surface sediments in world ocean [data from Archer (1996) retrieved from PANGAEA system (Diepenbroek et al., 2002)].

clearly shows the strong seasonal and interannual variability for this river. Note that corresponding data for any other river or comparisons and correlations between selected sets of rivers could be produced instantly with ODV.

4.2. Global database of surface sediment characteristics

Archer (1996) has produced a global dataset of surface sediment composition consisting of more than 5200 irregularly spaced data points. This dataset was retrieved from the PANGAEA information system (Diepenbroek et al., 2002) and imported into an ODV collection. As an example, Fig. 3 shows the distribution of $CaCO_3$ content, which was obtained by gridding, shading and contouring using ODV on the basis of the original data values. Geochemists will recognize the low $CaCO_3$ content in deep ocean basins due to dissolution in high pressure environments.
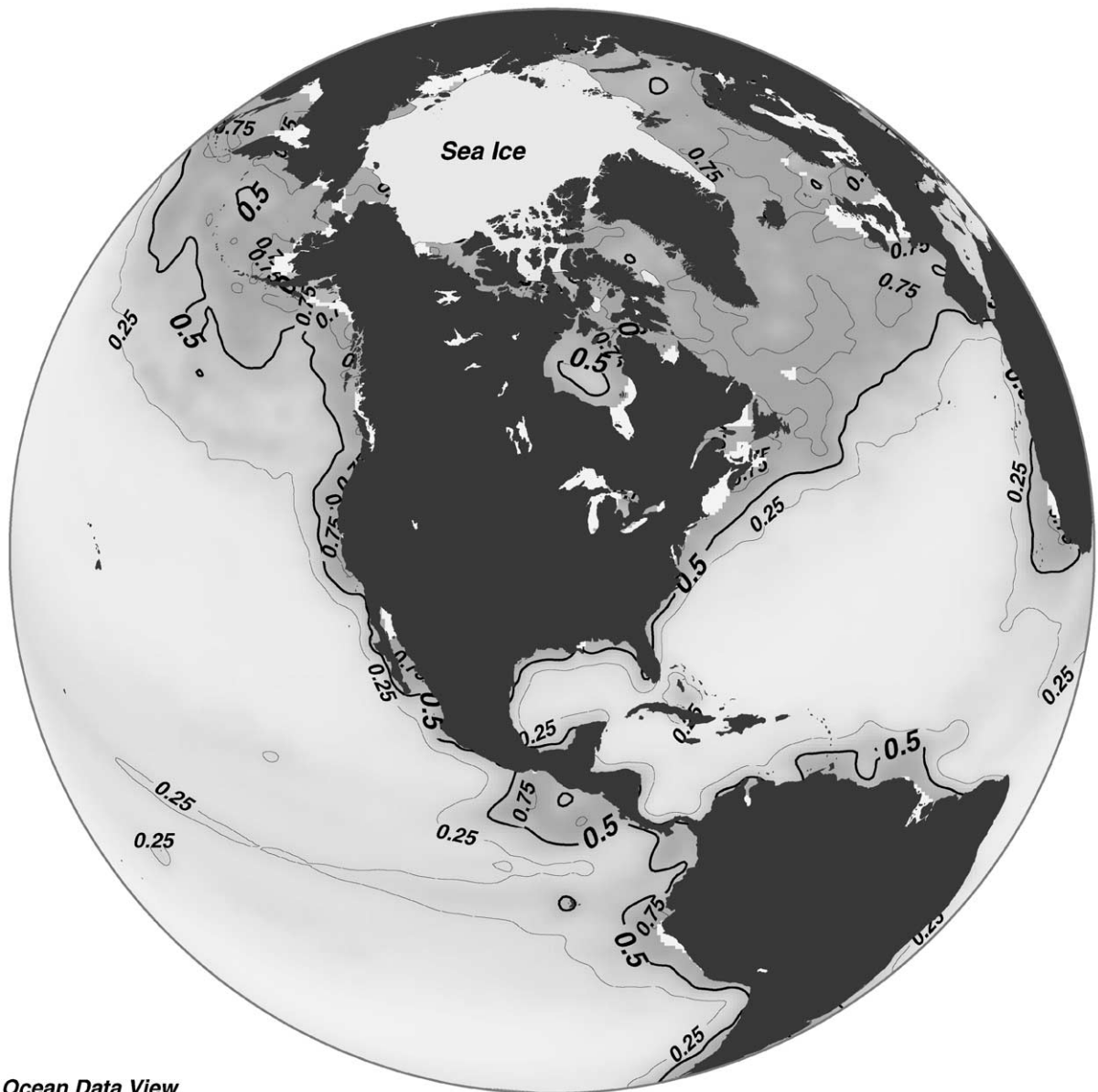

Fig. 4. Annual average chlorophyll a concentration ($mg\,m^{-3}$) for year 1999 derived from measurements of SeaWiFS satellite (see footnote 13).

4.3. Satellite chlorophyll maps

In recent years, satellites are routinely used for monitoring the land/ocean surfaces and for the detection of temporal variability and changes. One such application is the measurement by the SeaWiFS satellite of surface water color as an indicator for chlorophyll content and thus ocean productivity. Data from this mission are released on a regular grid[13] and are used by many scientists. Here, these data have been imported into an ODV collection, and the chlorophyll a concentrations in Northern Hemisphere waters are displayed in Fig. 4. Note that for this plot ODV uses an orthographic projection with the viewer position at $93°W/40°N$ (can be set arbitrarily by the user) and with sea ice extent added as an overlay.

5. Summary

ODV is a flexible and easy-to-use tool for the interactive exploration, quality control and graphical display of a wide range of geoscience data types. The package is available for Windows PCs and SUN Solaris workstations and can be downloaded over the Internet (see footnote 1). In addition to the software, extensive datasets for ocean bathymetry, land topography, rivers, lakes, sea-ice extent and national borders are provided. Several gazetteer databases provide context information for various types of geographical features. ODV is already used in more than one hundred research institutes world-wide, and a number of major datasets are released to the public in ODV format. Initiatives are under way to employ ODV for teaching in schools and for oceanography courses at universities.

References

Archer, D.E., 1996. An atlas of the distribution of calcium carbonate in sediments of the deep sea. Global Biogeochemical Cycles 10 (1), 159–174.

Brown, M., 1998. Ocean Data View 4.0. Oceanography 11 (2), 19–21.

Diepenbroek, M., Grobe, H., Reinke, M., Schlitzer, R., Sieger, R., Wefer, G., 2002. PANGAEA—an information system for environmental sciences. Computers & Geosciences 28 (10), 1201–1210.

Schlitzer, R., 2000. Electronic atlas of WOCE hydrographic and tracer data now available. EOS Transactions of American Geophysical Union 81 (5), 45.

Vörösmarty, C.J., Fekete, B., Tucker, B.A., 1998. River Discharge Database, Version 1.1 (RivDIS v1.0 supplement). Available through the Institute for the Study of Earth, Oceans, and Space/University of New Hampshire, Durham NH (USA) at http://pyramid.sr.unh.edu/csrc/hydro/.

WOCE Data Products Committee, 2000. WOCE Global Data, Version 2.0. WOCE International Project Office. WOCE Report No. 171/00, Southampton, UK.

---

[13] SeaWiFS Ocean Color Observations. http://seawifs.gsfc.nasa.gov/SEAWIFS/IMAGES/IMAGES.html