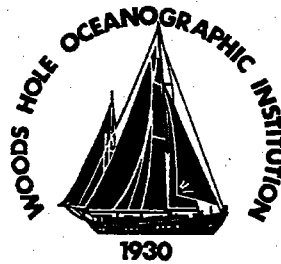


WHOI-96-01

COPY 2

**Woods Hole
Oceanographic
Institution**



DOCUMENT
LIBRARY

Woods Hole Oceanographic
Institution

HydroBase

**A Database of Hydrographic Stations and Tools
for Climatological Analysis**

by

Ruth Gorski Curry

March 1996

Technical Report

Funding was provided by the National Science Foundation through Grant OCE91-03364
and the National Oceanic and Atmospheric Administration through Contract Nos.
NA36GP0137 and NA46GP0303.

Approved for public release; distribution unlimited.

WHOI-96-01

DOCUMENT
LIBRARY
Woods Hole Oceanographic
Institution

HydroBase
A Database of Hydrographic Stations and Tools
for Climatological Analysis

by

Ruth Gorski Curry

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

March 1996

Technical Report

Funding was provided by the National Science Foundation through Grant OCE91-03364 and the National Oceanic and Atmospheric Administration through Contract Nos. NA36GP0137 and NA46GP0303.

Reproduction in whole or in part is permitted for any purpose of the United States Government. This report should be cited as Woods Hole Oceanog. Inst. Tech. Rept., WHOI-96-01.

Approved for public release; distribution unlimited.

Approved for Distribution:



Philip L. Richardson, Chair
Department of Physical Oceanography



Table of Contents

	Page
I. What is <i>HydroBase</i>?	1
A. Data Sources and Organization	2
B. Quality Control	3
1. Range Checking and Missing Data	3
2. Statistical Checking	3
3. Oxygen Quality	7
4. Further Quality Control	7
C. Comparison to <i>World Ocean Atlas 1994</i> Database	7
II. <i>HydroBase</i> Utilities	
A. An Overview	14
B. Module Descriptions & Examples	
1. Gridding Hydrographic Properties in 3-D: <i>grid3d</i>	17
2. Gridding on a 2-D surface: <i>gridsurf</i>	22
3. Smoothing a gridded surface: <i>smoothsurf</i>	25
4. Blanking gridnodes: <i>preblank, findblanks</i>	27
5. Extracting Stations from <i>HydroBase</i> : <i>extract</i>	28
6. Listing information from a cdf file: <i>cdfinfo cdf2asc</i>	30
7. Reading and Writing <i>HydroBase</i> files: <i>hydro_utils.c</i>	31
8. Reading and Writing cdf files: <i>hydro_cdf.c</i>	32
9. Computing hydrographic properties: <i>propcalc, prop_subs.c phyprops.f eos80d.f</i>	34
10. Sorting, listing and other tools: <i>ms*sort, getpos, surfdiff, timeprop, xy4gmt</i>	37
III. Appendices	
A. Structure of the <i>HydroBase</i> Data Files	41
B. Structure of the NetCDF files	42
C. Map of WMO squares: North Atlantic	44
IV. List of Figures	
1. Mean T-S relationship and standard deviation envelope used in quality control	4
2. Sample distribution with respect to depth	6
3. Potential Temperature at $\sigma_{\theta}=27.0$ <i>Levitus</i> vs. <i>HydroBase</i>	9
4. T-S of observations, <i>WOA</i> averages, and <i>HydroBase</i> averages : warm anomaly	10
5. T-S of observations, <i>WOA</i> averages, and <i>HydroBase</i> averages : cold anomaly	11
6. Deep pressure and temperature fields: <i>WOA</i> compared to <i>HydroBase</i>	12
7. T-S of observation, <i>WOA</i> averages, and <i>HydroBase</i> averages : deep anomaly	13
8. I/O diagram relating <i>HydroBase</i> modules	15
9. Isopycnal averaging at the sea surface	19

Abstract

This report documents the organization, functionality, and algorithms of a software package which operates as a database manager and toolset for climatological analysis of hydrographic station data. It details the methods of quality control used in construction of the small, but growing, database and discusses some of the improvements *HydroBase* methods offer over existing gridded databases, including a short comparison to Levitus's *World Ocean Atlas 1994* package. A large portion of this technical reference is devoted to describing the software modules and providing examples for their use.

I. What Is *HydroBase*?

HydroBase is a set of utilities designed to enable the user to access and manipulate profiled station data plus a limited, but growing, database of hydrographic and CTD stations. It is *not* a graphics package. The products of *HydroBase* are ASCII or netCDF files which the user can input to his/her own favorite graphics software. The source code for these utilities is also provided so that individuals can build their own applications from the basic modules. The strengths of this package lie in its flexibility to produce gridded datasets on scales specified by the user and its ability to handle any number of measured or derived properties. Its most important feature, perhaps, and one which distinguishes this from other databases is its implementation of isopycnal averaging. Although most other gridded datasets are produced by averaging properties on standard depth surfaces, *HydroBase* performs its gridding on closely spaced density surfaces and then transforms the averaged matrix from density to depth coordinates to ultimately produce a three dimensional matrix gridded in latitude, longitude and depth. Recent analysis by Lozier, McCartney, and Owens (1994) demonstrates that averaging on depth or pressure surfaces produces water properties that are dissimilar to the observed water masses; and that these artifacts are avoided by averaging/gridding in density coordinates.

The *HydroBase* utilities permit the user to :

- extract station subsets from the database using a variety of parameters;
- compute hydrographic properties from observations;
- produce 3-D gridded datasets of any hydrographic properties;
- specify all gridding parameters (latitude, longitude, depth, time);
- project hydrographic properties onto surfaces;
- smooth the gridded data on a surface;
- analyze data by groups of years.

The source code includes libraries to facilitate the reading and writing of *HydroBase* files and the computation of hydrographic properties for C programmers who wish to utilize the data in other ways.

The style of these programs was patterned after the GMT-SYSTEM¹ (Wessel and Smith, 1991), a popular and freely distributed graphics mapping package, because it has proved to be relatively easy to understand and use. In addition to its graphics capabilities, GMT provides several modules for interpolating and low-pass filtering 2-dimensional gridded data which are particularly useful in the context of filling data holes in hydrographic property fields. I have achieved excellent results with their module, *surface*, which implements a minimum

1. At this writing, GMT-SYSTEM is available by sending email to listserv@soest.hawaii.edu containing the single message: information gmtgroup.

curvature spline under tension algorithm (Smith and Wessel, 1990). Rather than reinventing the wheel and writing a *HydroBase* module to perform the same task, I recommend making use of this tool (or other statistical tools) to accomplish "objective mapping" of the property fields when necessary. Several script files have been included in the *.extras* subdirectory of the software distribution which use GMT to produce plots in *PostScript* format from the *HydroBase* files.

A. Data Sources and Organization

The station data used in this synthesis were obtained from the National Oceanographic Data Center (NODC) and represent all hydrographic station data (pressure, temperature, salinity, and oxygen) available as of April 1990. At the time this document is being produced, only one ocean basin has been included into *HydroBase*:

North Atlantic (0° - 75° N, 85° W - 20° E) : 131635 stations

As time and funding permit, the South Atlantic, Pacific and Indian Ocean basins will be added.

The stations are arranged geographically by 10-degree WMO (Marsden) squares and stored in files named *msq10.extent*, where *msq10* is the 4-digit WMO Square designation:

$$\text{WMO Square} = \text{hem} * 1000 + (\text{lat}/10) * 100 + (\text{lon}/10);$$

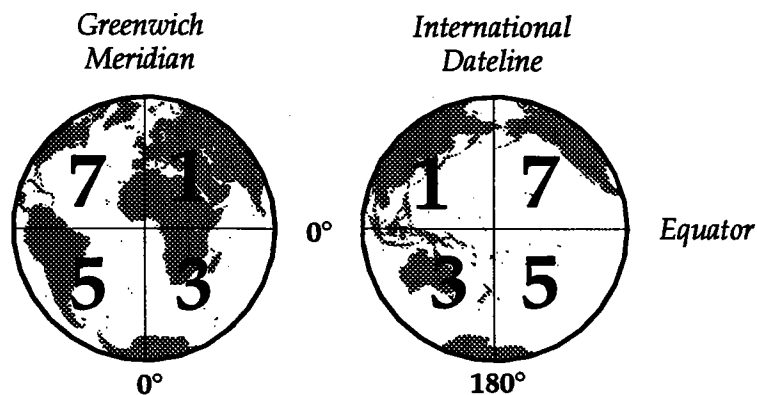
where *hem* = 1 or 7 for Northern hemisphere, (see diagram below)

3 or 5 for Southern hemisphere,

1 or 3 for Eastern hemisphere,

5 or 7 for Western hemisphere;

lat = (integer) latitude, *lon* = (integer) longitude



and *extent* denotes the type of data in the file, which can be any combination of:

- .qc** : quality controlled
- .shall** : stations < 200 meters depth
- .ctd** : CTD high resolution data
- .raw** : original data (not quality checked)

There is no particular ordering of stations within each file. See Appendix A for information on the structure of a *HydroBase* station.

The original database (Lozier, Owens, and Curry., 1996) contained only hydrographic stations >200 meters water depth. Subsequently, the coastal data, other shallow stations, and some CTD sections have been put into *HydroBase* format. These files are designated *msq10.shall* and *msq10.ctd*. Because *HydroBase* will continue to grow and evolve with time, a README file is stored with the data to log all updates and changes to the database.

B. Quality Control

Since producing a dataset suitable for climatological study was central to the goals of this project, the methods of quality control were designed to identify and eliminate unrealistic data without destroying the natural variation inherent to ocean circulation. The stations which successfully passed through our quality control steps may not meet everyone's standards because quality control is a very subjective process. Nor do we claim to have identified every "bad" station or observation and kept every "good" one. These quality control methods are best viewed as a preliminary process to deal with substantial numbers of unrealistic values scattered throughout a dataset of enormous size. Your usage of this dataset will dictate the need for additional examination of individual data points.

1. Range Checking and Missing Data

A preliminary sieve of the data eliminated values outside of broad property ranges which were defined for temperature, salinity and oxygen as a function of depth and latitude using atlases and synoptic section for guidance. Because subsequent quality control steps would require the ability to compute potential density, any observation level missing either temperature or salinity was entirely eliminated. Temperature and salinity were retained for any level missing only an oxygen value.

2. Statistical Checking

To further identify questionable data points, we used the fact that potential temperature (θ) - salinity (S) and θ - oxygen (O_2) relationships are locally well defined in the world oceans. Plots of θ -S and θ - O_2 for all stations within a discrete geographic area illustrate these relationships and graphically reveal data points that deviate significantly from them. These plots are often used to identify erroneous points visually in data as they are collected and analyzed. Because hand-picking outliers from visual representations of several hundred thousand stations was not a viable option, a method was devised to identify these outliers computationally.

The θ -S curve for a group of stations with similar property profiles can be approximated by vertically subdividing the observed points into smaller groups as a function of density and then connecting the mean θ ,S point for each density bin with straight lines (figure 1). As the number of density bins increases, these line segments converge on the shape of the θ -S curve. A line tangent to this curve at the mean θ ,S point of each density bin approximates the θ -S relation for that bin. The standard deviation of salinity as a function of potential temperature can be easily computed as the least squares distance from the tangent line to the corresponding observed θ ,S combination. Figure 1 illustrates that a distance of approximately ± 2 standard deviations from the tangent line adequately characterizes the natural spread of observations; most points outside of this envelope appear to be errant. Although density variations are often dominated by potential temperature, in some locations, notably the subpolar oceans and near the sea-surface, salinity has a greater influence. In these instances, it is more sensible to compute the standard deviation of potential temperature as a function of

