Data Processing Procedures and Conventions

Hydrographic Data

Hydrographic data are derived from instruments that profile the water column and are launched from platforms such as Research Vessels, Autonomous Profiling Floats (APEX and US Argo), and Aircraft. The most commonly used instruments are Conductivity, Temperature, Depth (CTD) sensors, eXpendable BathyThermographs (XBT), eXpendable Current Profilers (XCP), Ship-mounted Acoustic Doppler Current Profilers (ADP), and traditional water samplers such as Niskin Bottles that are most often attached to a CTD using a Rosette Sampler. There are similar aircraft deployable expendable probes that are named AXCTD, AXBT, and AXCP. All these instruments except the Niskin Bottles generate continuous profiles of the measured variables. Dataloggers usually digitize the depth profiles at 1-m intervals, though it can be less for high precision measurements. Hydrographic cast data are placed in one of 5 categories, shown below.

Data Type	Index Type Number	Description
CTD	1	Salinity and Temperature data from hydro casts or AXCTD's.
XBT	2	Temperature data from ship or aircraft deployed probes.
XCP	3	Relative Current Velocity and Temperature data from ship or aircraft deployed probes.
ADP	4	Current Velocity data from a downward looking ADCP usually mounted in the ship's hull.
NUT	5	Bottle data that include standard Nutrients. These are often combined with CTD data, and in some cases some data types (e.g. Oxygen) are from electronic sensors on the CTD.

Current velocity data from XCP's have an unknown, depth independent offset, and for this reason XCP's are often known as shear probes because the velocities are relative. The ADP currents on the other hand are stored as absolute velocities, and have undergone considerable processing to remove the effects of ship motion by using the bottom tracked velocity from the ADCP (preferred), or estimating the ship's velocity from precision navigation data. However, these methods have limitations and ADP current profile data will often be noisy, particularly if the ship is not on a steady heading. The Index Numbers in the table are used in the *netcdf* archive files discussed below. Cast data from a single cruise of a research vessel or aircraft flight are grouped under a unique ID that identifies the platform and cruise. Thus, all the data from a hydrographic survey of the De Soto Canyon taken by the R/V Pelican in December 1998 is grouped under the cruise ID: **PE9923.** This ID is taken from the cruise identification number assigned by the ship's operator, where **PE** is the ship ID (for Pelican), **99** is the fiscal year, and **23** is the consecutive cruise number of the fiscal year. Hydrographic data are divided into two parts:

1. **Station Data:** Position and time of each cast are identified by a *Station Number* that is usually, though not necessarily, numbered consecutively from the beginning of the cruise. Station numbers are always integers (i.e. stations 6a, 6b, etc. are not allowed). Station numbers are unique within each cruise.
2. **Profile Data:** Each profile, consisting of the depth and associated measured variables (e.g. Temperature and Salinity), is stored with depth increasing monotonically downwards, and identified by its station number. The depth variable may in some cases be replaced by pressure, measured in decibars (dbar). This is often the case for CTD systems.

Notes

Depth levels for expendable probe data (XBT and XCP) are calculated from elapsed time after launch using the manufacturer's formulae. These formula depths have been shown to have errors when compared with CTD's with precision depth sensors (Singer, 1990). No corrections have been applied to XBT and XCP depths in this archive. It is left to the user to apply his or her favorite correction formula.

QA/QC Procedures

Original data obtained from the instrument's data logger are edited for faulty values and interpolated to 1 m intervals. Where necessary near surface and near bottom values are deleted because of insufficient soak time (e.g. a CTD) or the instrument hitting the bottom (e.g. XBT probes), respectively. Salinity spikes, if they occur, are removed by a procedure that uses the equation of state and assumes the temperature record is correct and vertical gradients of sigma-t are positive (sigma-t increases downwards). Missing data values for a given depth level are flagged.

Archive Files

Hydrographic cast data from a single cruise are archived in a single netcdf file. The netcdf file name is the cruise ID (in lower case), sometimes appended by one of the 5 data types (e.g. "_ctd") if a subset of the original data is being provided, which is appended by ".nc". These netcdf files follow the COARDS conventions. Thus, *pe9923_ctd.nc* contains the CTD data from cruise **PE9923**. Data pertaining to the station, such as station number, position (latitude and longitude), time, water depth, instrument codes, etc., are stored in station arrays. These arrays also reference the start and stop indexes of the data arrays, so that data from a single station can be retrieved by the array slab operations of the netcdf API. This is done as follows:

For a particular station number, retrieve the *indextype*(1-5), the *start_index* and *stop_index* from the station arrays. The variables for the *indextype* are prefixed by the *data type* (see table above) followed by "_".Thus, if the *indextype* = 2, the variables are *XBT_Depth* and *XBT_Temperature*. The data is retrieved for the station by accessing the arrays *XBT_Depth(start_index : stop_index*) and *XBT_Temperature(start_index : stop_index)*.

- **Note:** *start_index* and *stop_index* are referenced to 1 (FORTRAN index conventions), whereas many software packages index with respect to 0 (C conventions). More than one type of data can be present in an archive as discussed above. Netcdf files are self-documenting, so the structure should be reasonably self-explanatory.

Reference

Singer, J.J., 1990. On the error observed in electronically digitized T-7 XBT Data. *J. Atmos. Ocean Technol., 7*: 603-611.