Hydratools Manual Version 1.0—Documentation for a MATLAB®-Based Post-Processing Package for the Sontek Hydra

By Marinna A. Martini[1], Christopher R. Sherwood[1], Rachel Horwitz[1], Andree Ramsey[1], Frances S. Lightsom[1], Jessica R. Lacy[2] and Jingping Xu[2]

[1]USGS Woods Hole Science Center

[2]USGS Pacific Science Center

Open-File Report 2005-1026

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Meyers, Director

U.S. Geological Survey, Reston, Virginia: 2006

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Martini, M.A., Sherwood, C.R., Horwitz, Rachel, Ramsey, Andree, Lightsom, F.S., Lacy, J. R., and Xu, Jingping, 2006, Hydratools manual version 1.0—Documentation for a MATLAB®-based post-processing package for the Sontek Hydra: U.S. Geological Survey Open-File Report 2005-1026, 49 p. plus CD.

Contents

Figures

Hydratools Manual Version 1.0, Documentation for a MATLAB®-Based Post-Processing Package for the Sontek Hydra

By Marinna A. Martini[1], Christopher R. Sherwood[1], Rachel Horwitz[1], Andree Ramsey[1], Frances S. Lightsom[1], Jessica R. Lacy[2] and Jingping Xu[2]

[1]USGS Woods Hole Science Center

[2]USGS Pacific Science Center

Overview

This section introduces the Hydratools toolbox and provides information about the history of the system's development. The USGS philosophy regarding data quality is discussed to provide an understating of the motivation for creating the system. General information about the following topics will also be discussed: hardware and software required for the system, basic processing steps, limitations of program usage, and unique features.

The Sediment Transport Instrumentation Group (STG) at the U.S. Geological Survey (USGS) Woods Hole Science Center has a long-standing commitment to providing scientists with high quality oceanographic data. To meet this commitment, STG personnel are vigilant in checking data as well as hardware for signs of instrument malfunction. STG data sets are accompanied by processing histories to detail data-processing procedures that may have modified the natural data signal while removing noise from the data. The history also allows the data to be reprocessed in the light of new insight into instrument function and moored conditions. This toolbox was compiled to meet these data quality commitments for data generated by Sontek Hydra systems using both ADV and PCADP probes.

In the mid 1990's, the USGS Coastal and Marine Program began frequent deployments of Sontek Hydra systems in support of projects in estuaries, coastal, and continental shelf regions nationwide. The structure of Hydra data sets are large and complex, and existing processing and editing tools consisted of fragments of MATLAB code written by USGS scientists to satisfy personal research needs. This code did not meet STG quality-control criteria. This toolbox permits engineers and scientists to monitor data quality by:

- processing data with interactive critical review;
- preserving data-quality indicators; and
- preserving minimally processed and partially processed versions of data sets.

STG usually deploys ADV and PCADP probes configured as downward looking, mounted on bottom tripods, with the objective of measuring high-resolution near-bed currents. The velocity profiles are recorded with minimal internal data processing. Also recorded are parameters such as temperature, conductivity, optical backscatter, light transmission, and high frequency pressure. Sampling consists of high-frequency bursts (1-10 hertz (Hz) of long duration (5-30 minutes) at regular and recurring intervals over a period of 1 to 6 months. The result is very large data files, often 500 MB per Hydra per deployment, in Sontek's compressed binary format.

Hardware & Software

The Hydratools toolbox consists of a series of M-files written in the Matlab language by Mathworks and is supported by most computer platforms. The programs have been specifically tested for Windows XP, however, it is most likely that the latest release will run on Unix, Windows 95, 98, 2000, NT and Macintosh OSX.

The current edition of the Hydratools is Version 0.0. This version has been tested with MATLAB version 7.0 service pack 2, release 14 and should work with MATLAB versions 7.0 and higher. This release of Hydratools has not been tested with MATLAB versions prior to 7.0, and currently we are unable to support this toolbox on earlier releases of MATLAB.

The input data file must conform to the format outlined in Sontek's ADVField/Hydra Operation Manual, appendix 2, Sept. 1, 2001. This format is binary and specifies the order in which the data appears.

After initial input, the data are converted to the NetCDF data format, and subsequent processed files are stored in the NetCDF format. Hydratools requires the NetCDF toolbox. Created for Matlab, the NetCDF toolbox was originally written by Dr. Charles Denham and was later maintained by John Evans. It can be downloaded from *http://mexcdf.sourceforge.net*

Data Terminology

The types of variables referenced throughout this document are defined below. They are required components of the input data file. The major data variables that are recorded for all three ADV and PCADP beams are:

- Velocity - velocity measured parallel to the acoustic transmit beam or acoustic receiver in mm/s
- Correlation – is an indication of the signal to noise ratio that is used as a measure of data quality
- Amplitude - a measure of the acoustic backscatter signal strength
- Heading, Pitch, Roll - the ADV or PCADP probe orientation

Ancillary data collected by Hydra systems include:

- External Sensors - any 0-5 volt output sensor, typically optical backscatter or transmissometers
- Pressure - piezoelectric pressure sensors or strain gauge sensors
- Temperature, internal - measured in the ADV or PCADP probe

- Temperature and Conductivity - to calculate salinity, measured by an external Sea-Bird Microcat

PCADPs produce a time series of vertical velocity profiles. Internal processing resolves the acoustic Doppler signal into a set of "cells", each of which represents a layer of water defined by the acoustic travel time from the head of the transducer. For those more familiar with acoustic Doppler profilers from other manufacturers, Sontek's use of the term "cell" is equivalent to other manufacturer's "bin".

A burst is a group of samples that have been recorded at a very high sampling rate, such as 2 to 10 Hz. Bursts are typically recorded at longer intervals such as once an hour.

The Best Basic Version (BBV) is the final version of the processed data, having known problems and noise removed.

Basic Processing

The first stage of the processing is to transfer the data from the instrument to the computer. In the Hydra, the data are recorded on PCMCIA Flash Memory cards. The data can be downloaded via the serial connection or copied from the PCMCIA card directly to the desktop using Sontek's Sonrecp program. The PCMCIA cards used by the Hydra are not formatted to be automatically recognized by Microsoft Windows.

It is important to establish a naming convention that will be carried throughout the data-processing steps. Using either download method, the files will be saved with the deployment file name given in the initial setup of the instrument. The USGS uses a naming convention that incorporates the mooring number, followed by a number indicating the instrument position on the mooring, and letters indicating the type of instrument. As of spring 2005, the USGS STG group indicates a Sontek ADV with the letters "adv" and the Sontek PCADP with the letters "pc". In most cases, the deployment name given to the Hydra will be in this convention. In the event that the filename is not in this form, the raw data file name should be changed to meet this convention. In each subsequent major processing step, a new file is created. Naming conventions used throughout the processing system are demonstrated by example in the section titled "Organizing the Data Files."

The Sontek binary files are translated to netCDF to allow ease of access using MATLAB tools. Editing and masking are performed on the netCDF formatted data, with final rotations to earth coordinated and scientific units being the last step in the process. Quality information is an important byproduct of the process. This toolbox, the quality information, script files, and metadata serve as a complete record of the processing steps performed and the specific quality of the data.

The fully processed data files meet the EPIC NetCDF standard developed by the NOAA Pacific Marine Environmental Laboratory (*http://www.pmel.noaa.gov/epic/*). These EPIC standards provide a universal vernacular allowing researchers from different organizations to share oceanographic data without needing translation. The final data file is called our Best Basic Version, indicating that it is clean of erroneous values, converted into earth coordinates, and compliant with the outlined standards.

Program Use

Several people have written the programs collected into the Hydratools toolbox, using slightly different approaches. The toolbox is designed to simplify the task of processing Hydra data, but it still requires a good understanding of the Sontek Hydra ADV and PCADP instruments, their capabilities, and limitations. This system was developed to be modular so that processing can be accomplished in a series of steps. Expert users can tailor some operations to their specific needs, and even modify programs to add processing capabilities. The result is a set of routines that can be run from a single script file in a series of steps, allowing overnight processing of the automatic and most time-consuming steps. A 500 MB PCADP data file can take as long as 8 hours to process on a machine with a 1.8 MHz single core processor.

Required Software

MATLAB® The hydratools software was written specifically for the MATLAB® computing environment. MATLAB® software is available from the Mathworks, Inc. This toolbox has been tested on Version 7.0.4.365 (R14) Service Pack 2 as of March 30th, 2005. It should be compatible with previous releases.

NetCDF The USGS Coastal and Marine Program, Woods Hole Field Center has decided to standardize on the netCDF data file format protocol. This format is a binary structure which allows up to a 2 GB file size and provides embedded data descriptors. NetCDF for all computing platforms and documentation can be obtained for free at: *http://my.unidata.ucar.edu/content/software/netcdf/index.html.*

Mexcdf This is the netCDF interface for MATLAB®. It can be obtained at: *http://mexcdf.sourceforge.net*

ncBrowse We suggest that you also obtain the ncBrowse utility for netCDF, which is located at *http://www.epic.noaa.gov/java/ncBrowse/*

The data generated by this toolbox is not only viewable with MATLAB® and ncBrowse, but with a host of applications available for free on the web. Data in netCDF format is portable to most computing platforms. The processing embeds metadata within each netCDF file, thus permanently attaching the descriptive contents of the data to each data file.

Limitations

The toolbox has been designed to detect many of the varieties of Hydra data automatically and process them. At the time of writing this document, the toolbox is able to process ADV from CPU firmware versions through 8.2 and PCADP data from CPU firmware versions through 17. However, there are some limitations:

- PCADP data with LISST or pressure series data cannot be read
- PCADP data with GPS or bottom track data can be read, but these data are ignored
- PCADP data with less than 401 profiles per burst may not have good ambiguity resolution, if ambiguity resolution is turned on

Some of these limitations may be overcome in the next major upgrade of this toolbox.

Installation

Unzip the toolbox files in a single directory called “hydratools.” Preserve the directory structure in the zip file. The ‘private’ directory is filled with special functions that may shadow or conflict with other functions on your MATLAB path, and in the private directory they are accessible only by hydratools. Edit your MATLAB® search path to include the hydratools directory. Any errors are often caused by path problems and conflicting function names. The “which” function is helpful to determine which function is being called by MATLAB®.

Quick Start

It is recommended that all users thoroughly review and understand this manual and how to operate the ADV and the PCADP. However, as a quick refresher to the seasoned user, the following procedure can be tried: Place your raw data file and the .drk pressure calibration file in the same directory. Edit the metadata example files metaPCADPexample.m and metaADVexample.m. Run them from the MATLAB® command prompt. The appendixes of this manual provide a full description of the metadata structure fields and program input settings.

General Procedure for Processing a Set of ADV and PCADP Data

Users have found that setting up the files and metadata, then doing a test run on a subset of the data (such as the first 500 bursts known to be in the water) is a great way to refresh the user on netCDF and toolbox procedures and flush out any metadata problems. It can take 4-8 hours of computing time to run an entire data set from start to finish, depending on the speed of the computer and the number of bursts in the data file.

Organizing the Data Files

Folder and File Naming Conventions

USGS establishes a unique number, called a mooring number, for each instrumentation platform it deploys. Using these mooring numbers and alpha-numeric descriptors for various instruments and sensors on each platform allow unique folder and file names to be defined for each time-series that originates from an experiment. An example file structure from a recent Eurostrataform deployment below illustrates how the USGS organizes its Hydra data for post processing with this toolbox. It may appear cryptic to the novice, but based on a long history from those working with the data daily, this file structure works well as a compromise between legibility and practicality. Mooring numbers are patterned as MMMN, MMM being the sequential number assigned to the mooring, and N being the sequential position for the instrument on the mooring. The numeral 1 is generally assigned to the uppermost instrument, but on heavily instrumented tripods this may be ambiguous, so the instruments on tripods are numbered according to significance to the project.

When data files or folders are named, the MMMN pattern is followed by an alphanumeric indicating the instrument type or data content, such as MMMNpc for a PCADP, MMMNadv for an ADV. Raw data files typically maintain the extension chosen by the instrument manufacturer. NetCDF data files have the extension .cdf for data files in work or .nc for processed data. What is described here is the method USGS has developed to track data files, and is designed for the convenience of USGS scientists and support personnel. The toolbox can be used with other file naming conventions that do not include mooring numbers. The following example applies to an ADV experiment with just one sampling scheme.

Example for a Eurostrataform project file folder tree:

EuroStrat	Root folder
702-ch10-flow	Folder with all mooring 702 data, Chienti 10 m site, flow tripod, contains folders with individual instrument data sets.
7022-pc30	Folder with all data from Sontek PCADP (pc), designated as instrument #2 on the tripod, serial number 30
C2PC1002.adp	Raw binary data downloaded from the instrument. Note that the name given in the field did not contain the mooring number. This is sometimes done, depending on the instrument.
7022pcb.cdf	Raw burst data in netCDF format
7022pcs.cdf	Raw statistical data in netCDF format
7022pcb-cal.nc	Processed burst data in netCDF format (b for burst, cal for calibrated data)
7022pcvp-cal.nc	Processed statistical data in netCDF format (v for velocity, p for pressure to identify the contents when this file is combined in directories with time series from other instruments)
meta7022.m	Script containing metadata and commands for processing this time series.
7022pcq.cdf	Quality information for editing data
T62493.DRK	Pressure sensor calibration information file
Ancillary	Folder containing logs of terminal sessions with this instrument.
7023-adv217	Folder with all data from Sontek ADV (adv), designated as instrument #3 on the tripod, serial number 217
C2AA1001.adr	Raw binary data downloaded from the instrument.
7023advb.cdf	Raw burst data in netCDF format
7023advs.cdf	Raw statistical data in netCDF format
7023advb1.cdf	Edited, un-calibrated burst data in netCDF format
7023advs1.cdf	Edited, un-calibrated statistic data in netCDF format
7023advb-cal.nc	Processed burst data in netCDF format (b for burst, cal for calibrated data)
7023advvp-cal.nc	Processed statistical data in netCDF format (v for velocity, p for pressure to identify the contents when this file is combined in directories with time series from other instruments)
Meta7023.m	Script containing metadata and commands for processing this time series
P82739.DRK	Pressure sensor calibration information file
7023advq.cdf	Quality information file
run20041014T165245	diary of MATLAB® processing session for this data
7024-adv220	Folder with all data from Sontek ADV (adv), designated as instrument #4 on the tripod, serial number 220

Variable Names in netCDF Files

A tremendous advantage of netCDF as a data-storage format is that it is self describing. For example, the netCDF files for oceanographic data can have a very different structure from those for image analysis. They are both in the same format on disk and can be read by many of the same generic tools. However, to be useful, consistent conventions have to be defined. USGS follows EPIC conventions for the final processed version of its data. This is how this toolbox typically defines variables within netCDF files. The EPIC conventions were established by the National Oceanographic and Atmospheric Administration Pacific Marine Environmental Lab (NOAA PMEL). More information about EPIC can be found here: *http://www.pmel.noaa.gov/epic/*

It should be noted that the EPIC conventions are not rules graven in stone, but conventions which evolved over time as new users of EPIC used new kinds of data. New EPIC definitions can be added by submitting requests to PMEL. Several new EPIC key names have been added for the USGS for other instrument applications, and the creation of this toolbox prompted the addition of a new Height definition, as there was no previous EPIC variable for the concept of height referenced to the ocean floor.

In the raw netCDF file, which has been simply converted from the instrument vendor's native raw binary to netCDF, EPIC names and units are not always available to define the data. For this reason, the names of the variables in the raw and in process version (.cdf) and the final version (.nc) are very different. The rule of thumb is this: the raw file (.cdf) contains variable names and information as close to the instrument vendor's original definitions as is practical. Units are generally in counts, engineering units or the units in which the instrument stored the data. Variable names in the processed file (.nc) are as close to the EPIC standard variables as is practical. Units are scientific units according to broadly accepted standards.

Gather the Metadata

For USGS data to be processed or archived, it must contain supporting information such as the depth and position of the instrument, the calibration information for the attached sensors, etc. Collectively, this information is known as the metadata. If the instruments were deployed by the USGS STG, such metadata are initially recorded in the mooring and instrumentation logs (field notebooks). Other organizations have other ways of recording and archiving metadata. However, all the metadata needs to be provided to the toolbox so that it can be written into the netCDF files. If not done correctly right from the beginning, metadata can often be the source of trouble with data sets later on.

Edit the Metadata Scripts

The functions that process the data have been written to accept metadata as a structure to allow automated scripting of the process of entering metadata, rather than prompting the user each time the program is run. MetaADVExample.m and metaPCADPExample.m are example script files that define metadata structure values and run the toolbox functions. See the appendix for a specific explanations explanation of each metadata element. **All fields are required unless otherwise indicated in the appendix.** Some of the toolbox functions take hours to execute, and others require user interaction. The example scripts have divided the procedures in to separate blocks that follow steps in this outline. The scripts should be used as examples of how to call the various functions that do the processing work. It is a good idea to use MATLAB's diary capability to record output from the toolbox m-files. This is set up in the metadata script examples.

Determine the Known Good Bursts

Before or after a deployment, instruments may sit out of the water collecting data. It is helpful to not include those bursts in the converted data. Use the manufacturer's software to determine which bursts were recorded when the instrument was out of the water, and limit the processing to data which was known to be recorded while in the water. Alternatively, one can convert all the bursts to a raw netCDF file, examine the data to see which bursts are bad, and then reprocess the file without the bad bursts.

Convert from Sontek Raw Binary to Raw netCDF and the Quality Information

This step converts the data from native raw Sontek format into netCDF. Run adr2cdf.m for ADV data or adp2cdf.m for PCADP data. These programs convert the raw Hydra binary files *.adr or *.adp to netCDF and generates the burst file *b.cdf with all the individual burst samples, the smaller statistics file *s.cdf with statistics for each burst, and the quality file *q.cdf which stores quality indicator statistics and results from editing and masking procedures for each burst.

At this step, no processing is performed. There are two exceptions: (1) rather than use counts, frequency pressure sensor data is expressed in hertz and (2) external sensor data is expressed in volts. With the data now in netCDF, the user can browse the data with several tools to find out more about the quality of the data and how much work may be involved in the next processing steps. NetCDF format allows access to all the data collected by the Hydra, including the analog channels. These are not necessarily accessible with Sontek's free software. **Be careful** when browsing the burst files; depending on how the experiment was set up, the files can be very big and can crash your system if you try to plot all bursts and all samples in each burst simultaneously. The tool browsehydraburst.m was written specifically to view burst data without loading too much data into a computer's memory at once.

Assess the Data and Quality Information and Use Automated Tools to Remove Noise

ADV data can be noisy at times. One particular problem that Sontek had with a faulty DSP chip can pollute some but not all of each burst. The toolbox has a method for examining the bursts and flagging all or part of a burst for elimination. One can skip to part V at this point if one is impatient to see rotated velocities, which may assist in evaluating the data, but eventually noise in the data will have to be addressed in order to write the Best Basic Version. So far, the PCADP has not had this problem.

A. Save a copy of the original raw data files by copying *.cdf to *1.cdf.
B. Cleanhydra.m will automatically remove outliers and other problems from ADV and PCADP data and recalculate statistics.
C. Use ncbrowser.m to look at the statistics data in MATLAB®.

D. Use browsehydraburst.m to look at burst data in MATLAB®. Don't use ncbrowser.m on burst files. They are simply too big, and one can inadvertently try to load all the samples into computer memory at once.
E. ncBrowse, a free java application available via UNIDATA can be used to browse either burst or statistics data files without using MATLAB®, and it will treat burst files better than ncbrowser.m, but it is still possible to hang your computer trying to display all the bursts and all the samples at once with this utility.
F. Examine your data for issues you may wish to address later during processing. Look for pressure spikes that might need to be masked out, strangeness in the heading, pitch and roll for which use of the median value will not compensate, etc. To correct these problems, there are overrides which can be defined when running adv2nc.m and pcadp2nc.m.

Interactive Flagging, Masking and Bad Data Removal

A. Run flagbadadv.m to interactively screen ADV data for noise with bursts. The displayed statistics are sets of five statistical time series, each calculated on one fifth of each burst. The standard deviation of the velocity for each subburst is displayed as a set of time series plots along with plots of correlation for the subburst. The user can set thresholds for the standard deviation of velocity and the correlation to include or exclude that subburst from the data. The user can also zoom in and click on the data to flag it, or type in [burst, subburst] to remove data. The user can preload flagbadadv.m with known bad burst and subburst flags. See the appendix "how to use flagbadadv.m". It is strongly suggested that the user repeat this process for each of the three ADV beams.
B. In the ADV data, remove bad data identified by flagbadadv.m and recalculate the statistics using fixbadadvvel.m.
C. In PCADP data, bad subbursts will have to be removed by hand; there is no equivalent tool to fixbadadvvel for the PCADP. Ambiguity correction is performed in the next step.

Apply Rotations and Write the Processed Data in Scientific Units

Run pcadp2nc.m or adv2nc.m to apply rotations, calibrations and output data in scientific units.

Burst Data vs Statistics Data

The Hydra is a powerful system that enables the user to collect very high resolution data in time and space from several different kinds of sensors. Statistics from these data can be useful to assess certain physical processes. Some of these statistics are provided by the instrument, both for the entire burst and even for each sample. The per-sample statistics represent the many pings that the PCADP and ADV average together to achieve the accuracy quoted in the instrument specifications. Examination of these per-sample statistics may reveal the source of bad data in certain situations. Other statistics are computed by this toolbox over an entire burst and are saved in the statistics (*s.cdf & *vp-cal.nc) files. The statistics files contain the average or median of the burst samples or profiles for the burst, as well as other statistics such as minima, maxima and standard deviation. These statistics allow the user to examine trends which occur over the course of a deployment. Individual, high resolution measurements in the ADV are referred to by Sontek and in this toolbox as 'samples'. For the PCADP, these are defined by Sontek and referred to in this toolbox as 'profiles'. The terms sample and profile are used in the netCDF burst files. Users should recognize the difference between per sample (for ADV data) and per profile (for PCADP data) statistics generated by the instrument and the statistics for entire bursts that are generated by the toolbox.

Quality Files

Each m-file that performs a processing step on the Hydra data generates information about the success of the processing and the quality of the data. This information is stored in a separate netCDF file called a quality file (*q.cdf). Adr2cdf.m and adp2cdf.m generate useful statistics for flagbadadv.m. Flagbadadv.m stores its data masks and overrides in the quality file. Cleanhydra also stores the results of operations such as deglitch, checkcorr, and thumbfinger in the quality file, information such as the number of points replaced, how each burst's statistics were changed, and the settings used to implement each of the operations.

Missing Sensors

If sensors are not installed, either by being set as "Not Installed" in the Hydra's set up menus or if a serial number is not provided by the user in the metadata structure, the data for that sensor is not written to the raw netCDF files. This is true for internal and external pressure, external analog sensors, compass and CTD. The compass serial number is not used in the metadata structure since the compass is an integral part of the PCADP and ADV probes, whose serial numbers are in the metadata structure. A strange oddity is that while an ADV may always have a temperature sensor, if the user sets 'Sensors' to 'Not Recorded' in the ADV, temperature will not be recorded as a burst parameter. It is in the header, however, and will in such a circumstance be written only to the statistics file by adr2cdf.m.

Cleaning Up the Data

The following strategies are employed to deal with typical problems in the following measured parameters:

Heading, pitch and roll:

Drift	Drift over the entire deployment cannot be corrected during processing. If such drift is detected, the heading, pitch, and roll overrides may be used to correct problems. Drift within a given burst is handled by using the median value as the orientation value for the burst.
Spikes	Wild points within a given burst are removed using the outlier removal tool called thumbfinger.m which can be executed using cleanhydra.m.

Pressure:

Drift	within a given burst is normal (the tides)
Spikes	within a given burst are an artifact of using a frequency pressure sensor while also running at sampling speeds of > 2 Hz. These are removed using an outlier removal tool called thumbfinger.m, which can be executed using cleanhydra.m. Often this problem shows up consistently in the first few samples of each burst of ADV data, and if this the case, a mask can be set up as one of the switches for adv2nc.m (switches.override_pmask). This has not been a problem with the PCADP.

Velocity:

Noise	in the ADV is removed automatically using cleanhydra.m or interactively using flagbadadv.m
Spikes	the tools for removing spikes are the same as for removing noise.

Ambiguity correction in the PCADP can be applied based on the resolution pulse velocity

External sensors are not examined for noise. Users may wish to use cleanhydra.m to clean up these data. This process is not automated in the toolbox because the editing of external sensor data can be very specific to each experiment.

About the noise filtering routines employed:

Thumbfinger	thumbfinger.m is a simple outlier-removal routine that removes points based on a threshold of a multiple of standard deviations from the mean, and then can replace these points with the mean (the default), median (recommended for heading, pitch and roll), or a user provided value.
Deglitch	deglitch1vector.m is a noise removal program that uses a filtering algorithm to detect bad points and replace them with interpolated points based on the nearest neighbors.
Checkcorr	checkcorr.m removes samples that have correlations below 65%.
Cleanhydra	cleanhydra.m automates the use of thumbfinger, checkcorr and deglitch. There are predefined ‘operations’ that one can apply to the data using known good default settings for thumbfinger, checkcorr, and deglitch. See the appendixes for details on how to use cleanhydra.m. Users with MATLAB programming experience can customize cleanhydra.m to apply their own cleanup operations to Hydra data. See the cleanhydra.m file for details on adding a new operation; use the existing operations as a template.
Flagbadadv	flagbadadv.m is an interactive GUI which allows the user to create a data mask using cutoff thresholds for correlation and the standard deviation of velocity, by burst, or by overriding the thresholds by interactively selecting individual bursts or parts of bursts with a mouse, or by entering a burst number. See the appendixes for details on how to use flagbadadv.m. Fixbadadvvel.m cleans up the data using the mask created by flagbadadv.m.

PCADP Ambiguity Correction

The PCADP high resolution profiling head available with the Sontek Hydra allows the use of pulse coherent mode to greatly improve measurement accuracy at fine spatial resolution. However, there is a tradeoff between cell size and maximum resolvable horizontal velocity. The PCADP emits a resolution ping to measure velocity in a single cell at the beginning of each set of profiling pings. The velocity measured by this ping can be used to correct the profile velocities. This is known as ambiguity correction. This toolbox uses the ambiguity correction outlined by Lacy and Sherwood (2004). This method assumes that the velocity shear between cells will be less than the ambiguity velocity. Ambiguity correction may be turned on in pcadp2nc.m by setting “switches.resolve_ambiguity” to 1, or off by setting the value to 0.

Orientation Issues

Orientation statistics can have two origins, the instrument or as calculated by this toolbox. This toolbox uses the median value of heading, pitch and roll in any statistical output. When statistical output is selected by the user in the ADV, then the actual in-situ mean and standard deviation values are transferred to the raw netCDF file. Such raw netCDF data files of ADV data will contain both mean and median values.

Depth, Height and All That

There are various measurements of depth and height above boundary in the PCADP and ADV data. Because there is a need to position the sensors and tripod in time and space for the deployment, there is the need to have a time series of how the sensors' distance to the bottom over time. At the beginning of every USGS deployment, a record is made of each sensor's nominal height above the bed by measuring the height of the sensors from the deck and recording this in the USGS mooring log (field notebook). Terms used in the netCDF files are defined as follows:

In all the files: The global attribute "WATER_DEPTH" is the nominal best known water depth of the deployment site. It is assumed that this depth does not change during the deployment. "WATER_DEPTH" is taken from the field "water_depth" in the structure "metadata" that is provided to adr2cdf.m and adp2cdf.m at run time. The variable attribute "sensor_depth" is the measured sensor height recorded in the mooring log subtracted from the "WATER_DEPTH".

In the raw (*.cdf) files depths, heights and/or distances to boundaries are expressed as the instrument measures them. In the PCADP burst file, Range1, Range2, and Range3 are the raw ranges to the bed for each beam recorded by the PCADP for each profile. This measurement can change dramatically throughout the deployment if the tripod shifts or sinks into the sediment. In the PCADP statistics file Range1, Range2, and Range3 are the average range to the bed for each beam for each burst. Brange, in the ADV statistics file, is the range from the instrument probe to the boundary (usually the bottom). Vrange, in the ADV statistics file, is the range from the measurement volume to the boundary. Brange - vrange is the ADV sample volume offset distance. This value has been seen to change for some deployments and to remain constant for most deployments. According to Sontek, variability in brange-vrange is a function of variability in the speed of sound.

In the .nc file, depth and height measurements are expressed in a way that would be most useful for analysis. The variable "depth" contains the depth of the primary measurement. For the ADV, this is the depth below the surface of the measurement volume, calculated by the formula WATER_DEPTH - initial_sensor_height + mean range to velocity measurement. For the PCADP, this formula is WATER_DEPTH - initial_sensor_height + blanking distance + (cell #)* (cell size). For the PCADP, Range1-3, the ranges to boundary for each beam, are averaged and stored as the variable brange. The user may ask pcadp2nc.m to correct the range from each PCADP beam for pitch and roll using switches.range_correct. For the ADV, brange and vrange are the same as for the raw file. It is left to the user doing the analysis of the ADV data to decide which parameter is best used to express the measurement distance to the boundary of interest.

Processing Transmissometers and Optical Backscatter Sensors

Transmissometers and optical backscatter sensors (OBS) are frequently used with Sontek Hydra systems. Voltage output for these sensors (generally attached as external sensors #1 and #2) is always written to the output files. However, this toolbox includes some basic processing for these sensors for output to the .nc files by adv2nc and pcadp2nc. The correct metadata must be supplied, otherwise only the sensor's voltage output will be written. The function which handles these sensors is dohydraoptic.m. The toolbox is currently tailored specifically for the most generic methods used by USGS scientists to process transmissometer and OBS data. However, the toolbox is designed to accommodate customized methods by adjusting the cals.coef metadata.

For the transmissometer: if pre and post calibration information and path length are provided in the metadata structure, then attenuation will be computed and saved in the .nc file by adv2nc and pcadp2nc as the variable "ATTN#_55". The calibration parameters are blocked and unblocked voltage output readings before and after the deployment, with the lenses as clean as possible. Transmissometers are connected to the Hydra using specially designed conversion cables.

For the OBS, processing is tailored to the D&A OBS-3 because this is the sensor the Hydra is specifically designed to support. No Seapoint sensor support is provided at this time. The OBS processing is less rigid and depends on how the sensor is calibrated. The metadata field gain is suggested in the example scripts; however it is not used in processing. This is the sensor gain discussed in the D&A OBS-3 manual as being approximately 1.0, 0.3 and 0.08 for mud, silt, and sand, respectively, in volts per mg/l. To provide flexibility, the user may apply a calibration equation, or simply apply the gain, by setting the metadata.ext#.cals.coef field in the metadata structure appropriately. "Coef = [# # #... n]" are the coefficients for an Nth degree polynomial. The accompanying descriptive fields are "units" and "equation" which document the units and the equation described by the coefficients. For example:

- To omit these OBS calculations, set metadata.ext#.cals.coef to [], the empty matrix, or omit the field entirely.
- To multiply the voltage output by a gain of 0.33, use metadata.ext#.cals.coef = [0 0.33], for a first degree polynomial.
- To apply the following formula, a+bv+cv2, which may have been developed as a calibration fit, use metadata.ext#.cals.coef = [a b c], for a second degree polynomial. Units might therefore be metadata.ext#.cals.units = 'kg/m-3'; and equation might be described as metadata.ext#.cals.equation = 'Conc [kg/m-3] = coef(1)+coef(2)*V+coef(3)*V^2';

The resulting calculation is written to the .nc file by adv2nc and pcadp2nc and the variable "Sed#_981".

If some other kind of sensor is connected to the Hydra's external sensor channels, the voltage output is always written to the output files and can be used to process these sensors using one's own procedures separate from this toolbox. The best course of action is to process the data generically without calibrations applied, then use the netCDF toolbox to access and process the data directly using MATLAB scripts of your own. Learning to use direct netCDF calls can be very useful to process external sensor data and to add forgotten metadata and fix other issues that can arise when processing data.

Acknowledgments

The routines for ADV and PCADP rotations and ambiguity correction came from Jessica Lacy's code. The method of reading the ADV binary files originated from work done by Jingping Xu. Some of the filtering and statistics code has come from both Chris Sherwood and Rich Signell. The initial work of Jessica Lacy, Chris Sherwood and Jingping Xu on ADV and PCADP processing for MATLAB® made it possible for this software to be written. Their work has been integrated with netCDF. Andree Ramsey started the ADV software and worked on it until Keara Ramsey came along. Frances Lightsom has been the Data Steward for archived oceanographic data at the USGS Woods Hole Science Center and patiently kept the methodology for the PCADP and ADV consistent with how the USGS handles the rest of their data. Rachel Horwitz has had the dubious honor of being the first real user of this software, and therefore has been instrumental in helping to root out the bugs. She got to do this without the benefit of any documentation. Thanks to the netCDF people at Unidata who have coordinated the maintenance of an excellent tool for oceanographic data. Thanks to Donald Denbo for writing the platform independent ncBrowse application to go with the platform independent netCDF. And finally, thanks to Jonathan Borden and Steve Ruane who did a lot of field preparation work so that Marinna Martini had time to bring the toolbox together. The netCDF toolbox for MATLAB®, without which none of this would be possible, was originally created by Charles Denham. Updated interfaces to netCDF for MATLAB® have been maintained by John Evans at sourceforge.net. The reviewers of this toolbox and documentation were John Warner and Charlene Sullivan.

References

J. R. Lacy, Sherwood, C. R., 2004. "Accuracy of a Pulse-Coherent Acoustic Doppler Profiler in a Wave-Dominated Flow," Journal of Atmospheric and Oceanic Technology, vol. 21, pp. 1448-1461.

SonTek/YSI ADVField/Hydra Acoustic Velocimeter (Field) technical Documentation, Sept. 1, 2001, Sontek/YIS Inc.

SonTek/YSI ADVField Software Manual (for use with software version 7.7), Sept. 1, 2001, Sontek/YSI Inc.

SonTek/YSI Acoustic Doppler Velocimeter Principles of Operation, Sept. 1, 2001, Sontek/YSI Inc.

Using Pressure Frequency Sensors with SonTek/YSI Systems, Dec. 2001, Sontek/YSI Inc.

ADP (Acoustic Doppler Profiler) Operation Manual Firmware Version 5.3, Oct. 1998, Sontek Inc.

ADP (Acoustic Doppler Profiler) Software Manual Firmware Version 5.4, Oct. 1998, Sontek Inc.

Acoustic Doppler Profiler (ADP) Principals of Operation, Sontek Technical Notes, Oct. 1998, Sontek Inc.

PCADP Read Me First, 11/15/2004, Sontek/YSI Inc.

PCADP: Principles of Operation, Sontek Technical Notes, Feb. 2001, Sontek Inc.

PCADP: Frequently Asked Questions, Sontek/YSI Technical Notes, May 2002, Sontek/YSI Inc.

Appendixes

1. Help
2. The Metadata Structure
3. The Settings, Inputs and Outputs for the Major Toolbox Functions
4. How to use Flagbadadv.m
5. How to use Browsehydraburst
6. Quality file variables
7. ADV Raw Burst File (*b.cdf) Variables
8. ADV Raw Statistics File (*s.cdf) Variables
9. PCADP Raw Burst File (*b.cdf) Variables
10. PCADP Raw Statistics File (*s.cdf) Variables
11. ADV & PCADP Best Basic Version Statistics File (*vp-cal.nc) Variables
12. ADV &PCADP Best Basic Version Burst File (*b.cal.nc) Variables
13. Attributes common to all Hydra files
14. Attributes specific to ADV data
15. Attributes specific to PCADP data

Appendix 1. Help

Typing help hydratools will result in a listing of the functions and their descriptions. They will appear as hyperlinks which you can use to access the help section for each individual function. This will result in the most up-to-date information on the functions while this beta version is still being debugged.

```
>> help hydratools
adp2cdf - read a binary Sontek ADP file and generate netcdf files of the raw data
 adr2cdf - Converts Sontek ADV files to netCDF format
 adv2nc - process adv burst data from a netcdf raw file
 browsehydraburst - browse ADV burst file data by burst (works for PCADP too)
 checkcorr - Identify and remove points with bad correlation
 cleanhydra - apply basic editing tools to hydra data
 deglitch1vector - Identify and remove suspicious points from ADV data
 fixbadadvvel - use the results from flagbadadv.m to remove bad blocks of data
 fixhydrarange - use acoustic range information from the ADV and PCADP to
 flagbadadv - interactively flag bad ADV data
 getburst - quick grab of burst data from a netCDF file
 metaADVExample - an example script for processing a single ADV data file.
 metaPCADPExample - an example script for processing a single PCADP data file.
 pcadp2nc - is a script that reads pcadp data from a netcdf raw file into matlab, applies rotations
 thumbfinger - Identify and remove points N stds beyond the mean
 xferOverridesExample - how to transfer flagbadadv overrides between beams
```

Appendix 2. The Metadata Structure

The raw binary to netCDF conversion programs adr2cdf and adp2cdf require an extensive amount of metadata to process and archive Hydra data. A quick start is to simple copy and paste the relevant information from metaPCADPexample.m and metaADVexample.m. **All fields are required unless stated otherwise below**. To make sure your information is correct, see the definitions below:

Structure Field Name	Definition	Example
mooring_number	the four digit mooring number as NNNL, where NNN is the mooring and L is the position	'7282'
metadata.deployment_start	ideally, the time the instrument entered the water, NOT the time logging commences, 'dd-Mmm-yyyy hh:mm:ss'	'23-Oct-2003 13:27:00'
metadata.deployment_end	ideally, the time the instrument leaves the water, NOT the time logging ends, 'dd-Mmm-yyyy hh:mm:ss'	'21-jan-2004 21:09:00'
metadata.lon	nominal site location, longitude, decimal degrees, West = negative	-78.77912
metadata.lat	nominal site location, latitude, decimal degrees, South = negative	33.72247
metadata.declination	local declination, magnetic variation, degrees, West is negative	-8.22
metadata.water_depth	nominal water depth at the deployment site, corrected for tidal variation, meters	10.6
metadata.system.serial	Hydra data logger electronics serial number	‘G136'
metadata.system.synch_to	if synching is used, the serial number of the system the unit is slaved to, otherwise ‘nothing’	‘G136'
metadata.pcadp.serial	serial number of PCADP probe, PCADP metafiles only	‘H30'
metadata.adv.serial	serial number of ADV probe, ADV metafiles only	‘217'
metadata.pcadp.height	nominal height above the bed of the PCADP probe, m	1.10
metadata.adv.height	nominal height above the bed of the ADV probe, m	1.10
metadata.adv.sample_volume_offset	nominal distance of the sample volume from the ADV’s emitter, 18 cm for ADVO, 10 cm for ADVF 10 MHz, 5 cm for ADVF 16 MHz	18
metadata.extpress.serial	serial number of the external pressure sensor, use the druck file name without the extension, use the null character, ‘’, if there is no external pressure sensor	‘T60005'
metadata.extpress.height	nominal height above the bed of the external pressure sensor, m. This is the same as the probe height if the sensor is in the ADV or PCADP probe	2.5

Structure Field Name	Definition	Example
metadata.ext#.serial	serial number of external sensor 1 or 2 where # = 1 | 2, the serial number is expressed as OBS### | TRANS### | XXX####, for OBS and transmissometers, the software looks for the leading OBS and TRANS in order to store the data in the correct units and compute attenuation. Use the null character, ‘’, if there is no sensor.	‘OBS2181'
metadata.ext#.height	nominal height above the bed of external sensor 1 or 2 where # = 1 | 2, m.	0.6
metadata.ext#.cals.XXX	Anything listed as a field under metadata.ext#.cals will get written to the netCDF files as an attribute to the variable containing the external sensor data, where # = 1 | 2	
metadata.ext#.cals.path_length	for TRANS only, focal length of the transmissometer, 5 | 25 | 100 cm, required for attenuation calculations, where # = 1 | 2	25
metadata.ext#.cals.pre	for TRANS only, pre calibration of the transmissometer, voltage output for air reading: [blocked unblocked], required for attenuation calculations, where # = 1 | 2	[5.0 0]
metadata.ext#.cals.post	for TRANS only, post calibration of the transmissometer, voltage output for air reading: [blocked unblocked], required for attenuation calculations, where # = 1 | 2	[4.0 0]
metadata.ext#.cals.gain	for OBS only, gain in OBS g/l per volt, where # = 1 | 2	0.33
metadata.ext#.cals.coef	for OBS only, calibration coefficients, use [0 1 0] if unknown, or [] to suppress calculation of sediment concentration, where # = 1 | 2	[0 1 0] | []
metadata.ext#.cals.units	for OBS, recommended to explain the coefficients, normally grams/liter where # = 1 | 2	‘g/l’
metadata.ext#.cals.equation	for OBS, recommended to explain how the coefficients are applied as coded in dohydraoptic, where # = 1 | 2	'Conc [kg/m-3] = coeff(1)+coeff(2)*V+coeff(3)*V^2'
metadata.CTD.height	nominal height of the CTD above the bed, m	2.28
metadata.CTD.serial	CTD serial number, Use the null character, ‘’, if there is no sensor. SBE37's record in real units and to not need calibration information.	‘SBE372924'
metadata.metafile	the name of the metadata file, supplied by MATLAB®, not written to the netCDF file	mfilename('fullpath')
metadata.metafile_date	the date of the metadat file, supplied by MATLAB®, not written to the netCDF file	datestr(now)
metadata.metafile_version	the version number of the metadata file, not written to the netCDF file	‘1.0'

Structure Field Name	Definition	Example
metadata.metafile_author	name of author of the metadata file, not written to the netCDF file	'MM'
metadata.origin	entity that collected the data, can be used to identify collaborators' data, such as USGS/WHFC, USC, NOAA, etc. If not provided, 'unknown' will be written to the netCDF file.	'USGS/WHFC'
metadata.experiment	experiment name, if not provided, 'unknown' will be written to the netCDF file.	'Boston Longterm'
metadata.project	project name, and/or funding source, if not provided, 'unknown' will be written to the netCDF file	'WHFC'
metadata.goodbursts	it is useful to omit bursts at the beginning and the end that were not recorded, when the instrument was in the water, use [1 Inf] to process all data	[1 100]
metadata.whichscheme	for the ADV only, where multiple sampling schemes were set up, which scheme is being processed.	1
metadata.####.cals.XXX	For any sensor #### which can be extpress | pcadp | adv | ext#, anything listed as field XXX will get written to the netCDF files as an attribute to the variable containing sensor data	

Appendix 3. Settings, Inputs and Outputs for Major Toolbox Functions

The following settings and overrides allow the toolbox to be tailored for specific problems in processing PCADP and ADV data:

For adr2cdf

For adr2cdf: adr2cdf(adrFile, cdfFile, [writeburstfile],[writestatsfile],[diagnostics], metadata)

adrFile	raw Sontek ADV binary file *.adr
cdfFile	file name to output the data in netCDF format
writeburstfile	optional burst file output selector, [startburst endburst] for specific bursts, [1 Inf] for all bursts, [] to prevent a burst file from being written, if writeburstfile is omitted, then all data are written
writestatsfile	optional statistics file output selector, [startburst endburst] for specific bursts, [1 Inf] for all bursts, [] to prevent a burst file from being written, if writestatsfile is omitted, then all data are written
diagnostics.pcountout	= 1: output the raw pressure count in the burst file, use to check pressure data
diagnostics.fixpfreq	= 1: apply a correction algorithm if wrong pressure offset was used. The number of points fixed per burst is returned in Qa.pfixed
metadata.*	see metadata structure table description

For adp2cdf

For adp2cdf adp2cdf(adpFile,cdfFile,[writeburstfile],[writestatsfile],[diagnostics], metadata)

adpFile	raw Sontek PCADP binary file *.adp
cdfFile	file name to output the data in netCDF format
writeburstfile	optional burst file output selector, [startburst endburst] for specific bursts, [1 Inf] for all bursts, [] to prevent a burst file from being written, if writeburstfile is omitted, then all data are written
writestatsfile	optional statistics file output selector, [startburst endburst] for specific bursts, [1 Inf] for all bursts, [] to prevent a burst file from being written, if writestatsfile is omitted, then all data are written
diagnostics.verbose	= 1: display hex start and end points in .adp file for each record, for file debugging
diagnostics.pcountout	= 1: output the raw pressure count in the burst file, use to check pressure data
diagnostics.fixpfreq	= 1: apply a correction algorithm if wrong pressure offset was used. The number of points fixed per burst is returned in Qa.pfixed
diagnostics.rotationSTDs	= 1: output the standard deviation of heading, pitch, and roll for each profile
diagnostics.temperatureSTDs	= 1: output the standard deviation of temperature for each profile
metadata.*	see metadata structure table description

For cleanhydra

For cleanhydra: function cleanhydra(cdfbFile, cdfsFile, cdfqFile, operation, ['variables'], [{vnames}], ['settings'], [settings.n], ['burstrange'], [1:Inf])
Results are saved to the quality file. See cleanhydra.m, deglitch1vector.m, checkcorr.m and thumbfinger.m code to see how to change the default settings for clean hydra, or how to apply your own cleaning algorithms, expert users only.

cdfbFile	raw ADV or PCADP netCDF burst data converted by adr2cdf.m or adp2cdf.m
cdfsFile	raw ADV or PCADP netCDF statistics data converted by adr2cdf.m or adp2cdf.m
cdfqFile	file to save the quality info returned by the cleanup operation
variables	cell array of variable names from the burst file to operate on
settings	a structure of control settings for each operation below
burstrange	range of bursts or individual bursts to work on, using the Sontek burst number as a vector of individual bursts [1 5 9 10 ...] or as a vector of burst ranges [1:10 20:30 ...]
operation	which cleanup operation to perform on the data variable to apply operation to, only one operation per run
operation = autoclean	there are no settings for this, they are automatic, applies thumbfinger & deglitch1vector to selected variables
operation = deglitch1vector	uses the deglitch1vector function with default settings
operation = checkcorr	operates only on velocity, other variables ignored, uses default settings
operation = thumbfinger	operates on heading, pitch, roll, temperature, velocity unless variable names are provided

For flagbadadv

For flagbadadv:	flagbadadv('QualityFile', cdfqFile,'settings', settings)
cdfqFile	neCDF quality file generated by adr2cdf
settings.currentBeam	the beam operated on
settings.new_bad	{[]; []; []; [1 2 3]; []} a 5 cell matrix of arrays of burst numbers of known bad data to add to mask override
settings.new_good	{[]; [101]; []; [1 2 3]; [35]} a 5 cell matrix of arrays of burst numbers of known good data to ass to mask override
settings.new_scutoff	5 element array with a new velocity standard deviation threshold
settings.new_ccutoff	5 element array with a new correlation threshold

For adv2nc

For adv2nc:	Adv = adv2nc(cdfbFile, cdfsFile, ncFile, switches, bursts);
cdfbFile	raw ADV data netCDF file of burst data *b.cdf
cdfsFile	raw ADV netCDF file of stats data *s.cdf
ncFile	file name for the netCDF output file, to which b.nc is appended for the burst data, which is not EPIC compliant, shaped [time, sample, depth]; and *vp-cal.nc for the statistics, in EPIC compliant format
switches	a structure of settings that will affect how data are processed, if they are omitted, defaults shown below will be used, switches = [] invokes all defaults
switches.ecorr_showplots	= 0, only set to 1 to get plots if processing < 10 bursts.
switches.override_heading	a vector of heading values which can be used instead of instrument measured heading. Vector length must match the number of bursts this function is asked to process, the default is the null vector []. If the instrument did not have a compass installed, and no overrides are provided, all heading, pitch and roll values will be assumed to be zero.
switches.override_pitch	like for heading, but overrides pitch
switches.override_roll	like for heading, but overrides roll
switches.override_pmask	for external frequency pressure sensors, mask sets of wild points. override_pmask is the same length as nsamples per burst, and set to zero where spiky data are to be omitted in each burst. This masking method works best for spikes at the beginning of a pressure burst. Example: switches.override_pmask = [zeros(5,1); ones(4500-5,1)];
bursts	bursts in the *b.cdf file to process, as [startburst endburst], as Sontek burst number
Adv	data from the last burst read, for diagnostics or extracting one burst of data

For pcadp2nc

For pcadp2nc:	[Qa, Pcadp] = pcadp2nc(cdfbFile, cdfsFile, ncFile, switches, bursts)
cdfbFile	raw PCADP data netCDF input file of burst data *b.cdf
cdfsFile	raw PCADP data netCDF input file of statistics data *s.cdf
ncFile	file name for the netCDF output file, to which b.nc is appended for the burst data, which is not EPIC compliant, shaped [time, profile, depth]; and *vp-cal.nc for the statistics, in EPIC compliant format
switches	structure of controls as to how data are processed
switches.ecorr_showplots	= 0, only set to 1 to get plots if processing < 10 bursts
switches.resolve_ambiguity	= 0 do not resolve, 1 resolve ambiguity
switches.amb.showplots	= 0 suppress plots, 1 show plots (this is useful for 1 or two bursts that you are trying to clean up)
switches.amb.timecorrect	= 1 use time correction, 0 suppress (use this if your profiles are more than 5 sec apart in time)
switches.amb.cut_factor	= 0.7, change at your own risk...;
switches.amb	= []; use all defaults for ambiguity correction. The following can be a per burst vector omitted entirely. Vector length must match the number of bursts this function is asked to process
switches.range_correct	= 1; adjust distance to bottom measured by beam range for pitch and roll
switches.override_heading	see adv2nc
switches.override_pitch	see adv2nc
switches.override_roll	see adv2nc
bursts	Bursts in the *b.cdf file to process, as [startburst endburst], as Sontek burst number
Pcadp	output structure of data from the last burst read

Appendix 4. How to Use Flagbadadv

Flagbadadv is an interactive screening program to identify bad ADV data for later removal. It operates on data from one ADV axis at a time. It presents beam correlation, the standard deviation of current velocity and pressure standard deviation as measures of ADV data quality. Each burst is divided into five sub-bursts to give the user the option of preserving parts of bursts. Sub-burst #1 consists of the first fifth of a burst; sub-burst #2 consists of the second fifth, and so on. A fifth is calculated by the number of samples per burst divided by five. The statistics are calculated using each sub-burst (one fifth) of the burst data, then the statistics for each sub-burst are assembled as a time series. These time series of statistics of the sub-bursts are calculated by adr2cdf.m and saved in the quality file, and subsequently displayed in flagbadadv's plots. The user may select data for elimination by using threshold settings of correlation and current velocity standard deviation, by pointing and clicking on individual fifths of bursts or by specifying burst and sub-burst pairs.

Figures 1 and 2 illustrate how flagbadadv works. Figure 1 is a screen shot of flagbadadv applied to raw quality statistics for ADV velocity data for the x axis. The top five plots are the correlation (left axis) and velocity standard deviation (right axis) for 1 fifth of each burst, plotted as the first fifth in the top plot to the last fifth in the plot fifth from the top. There are over 35000 bursts in this experiment, and all are shown in the flagbadadv window. Subburst 1 is shown zoomed in to the region of burst #35000 to illustrate detail. Zooming is achieved using MATLAB®'s native plotting functions accessed by the menus and buttons at the top of the MATLAB® figure window. The bottom plot shows pressure standard deviation for the entire burst; a spike is visible as these are raw data. Green dashed lines correspond to the velocity standard deviation cutoff settings in the edit boxes to the right of the plots. Nearly all the bursts in this data set possess good correlation, thus the threshold for correlation is not visible in the plots since it is out of range of the axis limits (set to reject samples with correlations below 70%). These cutoff values can be changed by entering new values in the edit boxes to the right of each plot. In the velocity plots, the x axis shows the burst number, in the pressure standard deviation plot, the x axis shows the corresponding date. Small red dots show data automatically selected for elimination by the cutoff settings. These dots may be toggled on and off by either clicking on the plot or entering the burst number and subburst in the edit box in the lower right hand part of the figure. Clicking on a point already marked by a red dot adds this data (burst, subburst) to the override as good list, and marks it with a little red circle. Clicking on it again adds the data to the override as bad list, and marks it with a little red asterisk. Similarly, clicking on data that was not marked bad puts it in the override as bad list. The final data mask is constructed from the cutoffs and the override lists and saved in the quality file. Points may be selected and deselected using the cursor in the MATLAB window as described, or burst and subbursts may be entered directly in an edit box in the lower right corner of the flagbadadv display. The override lists may be reset using the two nearby buttons.

Examples marked by circles A, B and C in Figure 1 correspond to plots A, B and C in Figure 2. In Figure 1, example A is an area where several subbursts have exceeded the set threshold. Example A is in a region of relatively high standard deviation of velocity, which is probably due to a storm, as there is correspondingly high pressure standard deviation for the same bursts (bottom plot). The individual samples are all good, and there is nothing to eliminate here. Examination of the burst using browsehydraburst (figure 2) shows this to be the case. This indicates that the velocity standard deviation cutoff threshold should probably be raised.

Example B shows a case where an individual spike is still good data and should be marked using override as good, where raising the threshold may not be the best solution. Example C is clearly a bad spike, and if most of the rest of the data in the experiment are good, setting this point to override as bad (marked with an asterisk) would allow high cutoff thresholds and simplify the editing process.

The reset zoom button is a fast way to reset all the plots to display all the data if you have used the MATLAB zoom tool to inspect the detail of one of the time series.

This interactive tool was developed initially to save data polluted by the noise from a DSP chip problem which did not respond well to standard methods of eliminating outliers. Flagbadadv also turned out to be useful for general screening. A good rule of thumb is that any data with correlation values less than 70% are assumed to be bad. When in doubt, use browsehydraburst to verify the burst data directly.

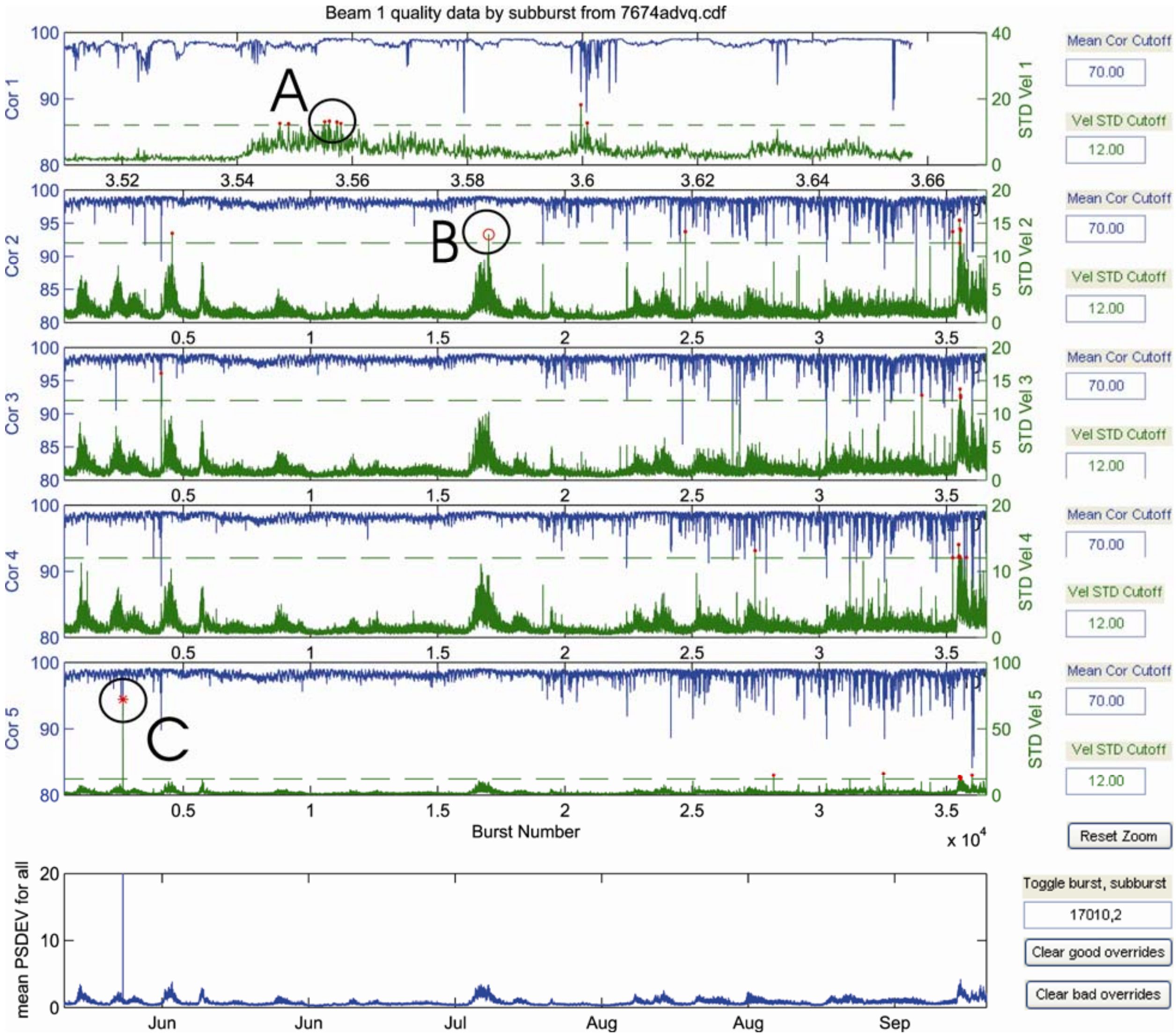

Figure 1. Screen shot of flagbadadv.m GUI with example areas A, B & C circled.

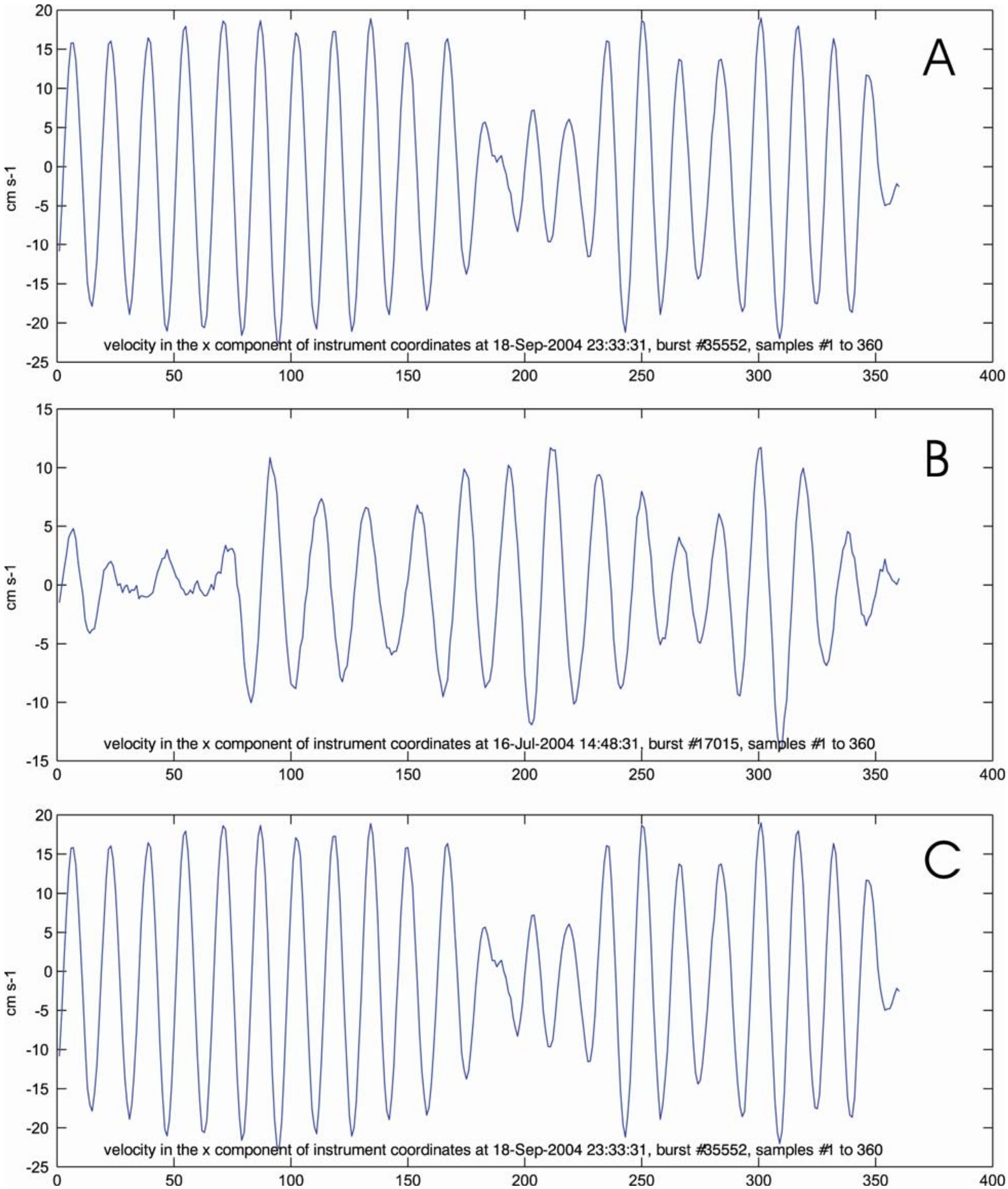

Figure 2. Detail of examples A, B & C for flagbadadv.m

Appendix 5. How to Use Browsehydraburst

Browsehydraburst provides the user with the means to browse burst data displaying only one burst at a time, in detail. Burst files are extremely large. Are typical USGS four month deployment will generate a burst file that is often larger than 1 GB. Fortunately, netCDF allows rapid indexed access to netCDF files. Browsehydraburst capitalizes on this by loading only the data desired.

Figure 3 shows screen shots of ADV burst data displayed by browsehydraburst. The top panel shows processed ADV correlation. The bottom panel shows pressure. One can step through a burst file quickly by clicking on the middle slider, and can set zooming with the top and bottom sliders. All MATLAB® figure manipulation tools are available. Bursts may be accessed directly by entering the burst number in the edit box. Data are selected using the drop down list box in the lower right hand corner. The plots in Figure 2 were captured using browsehydraburst.

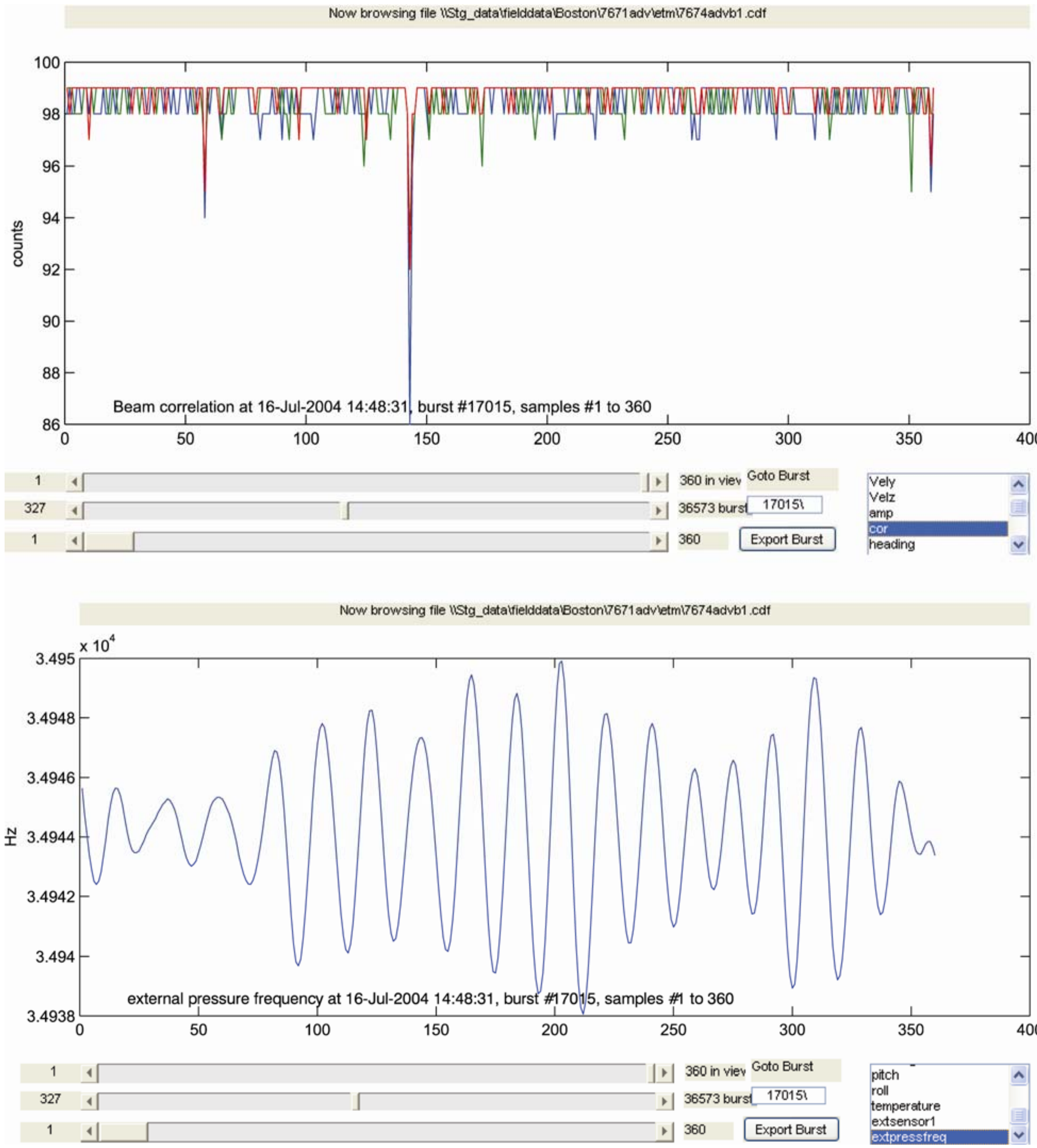

Figure 3. Screen capture of browsehydraburst GUI.

Appendix 6. Quality File (*q.cdf) Variables

The names of the variables stored in the quality file are constructed according to the m-file which generated or uses the data in the variable, followed by the names of the parameter. Global attributes are similarly named, and contain the settings used by the m-file to clean the data. Flagbadadv related data has an additional dimension called subbursts, which are subsets of samples in each burst. These are used to allow interactive removal of part of a noisy burst without sacrificing all the samples in the burst.

Quality File (*q.cdf) Variables and Attributes

Variable or Attribute name	Description	Origin
flagbadadv_stdv	Standard deviation of the velocity samples in a burst, [burst, axis, subburst]	calculated by adv2cdf.m, updated by cleanhydra.m
flagbadadv_scutoff	threshold to remove bad data based on flagbadadv_stdv, [axis, subburst]	initially calculated by adv2cdf.m, updated by flagbadadv based on user interaction
flagbadadv_mcor	Mean correlation for the velocity samples in a burst, [burst, axis, subburst]	calculated by adv2cdf.m, updated by cleanhydra.m
flagbadadv_ccutoff	threshold to remove bad data based on flagbadadv_mcor, [axis, subburst]	initially calculated by adv2cdf.m, updated by flagbadadv based on user interaction
flagbadadv_stdp	Standard deviation of the pressure samples in a burst, [burst, axis] to help differentiate sensor noise from wave processes	calculated by adv2cdf.m, updated by cleanhydra.m
flagbadadv_mask	mask marking good (0) and bad (1) ADV velocity data, [burst, axis, subburst]	recorded by flagbadadv.m, used by fixbadadvvel.m
flagbadadv_override_as_bad	record of data forced as bad interactively by the user, [burst, axis, subburst]	recorded by flagbadadv.m
flagbadadv_override_as_good	record of data forced as good interactively by the user, [burst, axis, subburst]	recorded by flagbadadv.m
checkcorr_locorr	low correlation cutoff level (default = 65%)	recorded by cleanhydra.m
checkcorr_rvalue	replacement value (default = NaN)	recorded by cleanhydra.m
checkcorr_Velx_nbad, checkcorr_Vely_nbad, checkcorr_Velz_nbad	record of the number of points removed in each burst by checkcorr, [burst]	recorded by cleanhydra.m when applying checkcorr.m
thumbfinger_nsd	number of standard deviations outside of which defines an outlier (default = 2)	recorded by cleanhydra.m
thumbfinger_rvalue	replacement value, 'mean', 'median' or a numeric value (default = 'mean')	recorded by cleanhydra.m
thumbfinger_nbad	record of the number of points removed in each burst by thumbfinger, [burst]	recorded by cleanhydra.m when applying thumbfinger.m
thumbfinger_delmean	ratio of the mean before and after points were replaced, [burst]	recorded by cleanhydra.m when applying thumbfinger.m

Quality File (*q.cdf) Variables and Attributes		
Variable or Attribute name	Description	Origin
thumbfinger_delvar	ratio of the variance before and after points were replaced, [burst]	recorded by cleanhydra.m when applying thumbfinger.m
thumbfinger_stdr	standard deviation of the residuals, [burst]	recorded by cleanhydra.m when applying thumbfinger.m
deglitch_nsd	number of standard deviations outside of which defines an outlier (strongly recommended default = 2.8)	recorded by cleanhydra.m
deglitch_samplerate	sample rate for the filtering algorithm = 1/(time between samples), taken from the file	recorded by cleanhydra.m
deglitch_ndt	cutoff frequency (strongly recommended default = 6)	recorded by cleanhydra.m
deglitch_verbose	show statistical output, 1 is on (strongly recommended default = 0 or it will slow processing)	recorded by cleanhydra.m
deglitch_nbad	record of the number of points removed in each burst by deglitch, [burst]	recorded by cleanhydra.m when applying deglitch.m
deglitch_delmean	ratio of the mean before and after points were replaced, [burst]	recorded by cleanhydra.m when applying deglitch.m
deglitch_delvar	ratio of the variance before and after points were replaced, [burst]	recorded by cleanhydra.m when applying deglitch.m
deglitch_stdr	standard deviation of the residuals, [burst]	recorded by cleanhydra.m when applying deglitch.m
deglitch_nfixed_dg	number of points fixed by deglitch, [burst]	recorded by cleanhydra.m when applying deglitch.m
deglitch_nfixed_filt	number of points fixed by filtering, [burst]	recorded by cleanhydra.m when applying deglitch.m

What's in the Raw netCDF files

Appendix 7. ADV Raw Burst File (*b.cdf) Variables

[Data recorded for each sample]

Variable	Description, [dimension]	Origin
burst	burst sequence number	Hydra, ADV Burst Header
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000, [burst, sample]	calculated from the Hydra, ADV Burst Header at the beginning of burst
time2	msec since 0:00 GMT, [burst, sample]	calculated from the above
Velx	velocity in the X direction, in instrument coordinates, [burst, sample], cm/s	ADV, ADV data structure
Vely	velocity in the Y direction, in instrument coordinates, [burst, sample], cm/s	ADV, ADV data structure
Velz	velocity in the Z direction, in instrument coordinates, [burst, sample], cm/s	ADV, ADV data structure
amp	beam amplitude, [burst, sample, beam], counts	ADV, ADV data structure
corr	beam correlation, [burst, sample, beam], counts	ADV, ADV data structure
heading	heading from compass, [burst, sample], degrees magnetic	ADV, compass data structure
pitch	rotation about Y axis, [burst, sample], degrees	ADV, compass data structure
roll	rotation about X axis, [burst, sample], degrees	ADV, compass data structure
temperature	ADV Probe temperature, [burst, sample], degrees C	ADV, sensor data structure
pressure	ADV Probe strain gauge pressure, [burst, sample], counts	ADV, sensor data structure
extsensor1	external sensor #1 voltage, [burst, sample], volts	external sensor data structure
extsensor2	external sensor #2 voltage, [burst, sample], volts	external sensor data structure
extpress	external pressure count, if installed and requested for diagnostics, [burst, sample], counts	external pressure data structure
extpressfreq	external pressure frequency, [burst, sample], Hz	Calculated

What's in the Raw netCDF files

Appendix 8. ADV Raw Statistics File (*s.cdf) Variables

[Data read or calculated for every burst]

Variable	Description, [dimension]	Origin
burst	burst sequence number	Hydra, ADV Burst Header
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000	Hydra, ADV Burst Header, beginning of burst
time2	msec since 0:00 GMT	calculated from the above
MeanVelx	mean velocity in the X direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
StdVelx	Std. Dev. velocity in the X direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
MeanVely	mean velocity in the Y direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
StdVely	Std. Dev. velocity in the Y direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
MeanVelz	mean velocity in the Z direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
StdVelz	Std. Dev. velocity in the Z direction, in instrument coordinates, [burst], cm/s	calculated from samples in burst
MeanAmp	mean beam amplitude, [burst, beam], counts	Hydra, ADV Time Series Statistics
StdAmp	Std. Dev. beam amplitude, [burst, beam], counts	Hydra, ADV Time Series Statistics
MeanCor	mean beam correlation, [burst, beam], counts	Hydra, ADV Time Series Statistics
StdCor	Std. Dev. beam correlation, [burst, beam], counts	Hydra, ADV Time Series Statistics
MeanHeading	mean heading from compass, [burst], degrees magnetic	Hydra, ADV Time Series Statistics
StdHeading	Std. Dev. heading from compass, [burst], degrees magnetic	Hydra, ADV Time Series Statistics
MeanPitch	mean rotation about Y axis, [burst], degrees	Hydra, ADV Time Series Statistics
StdPitch	Std. Dev. rotation about Y axis, [burst], degrees	Hydra, ADV Time Series Statistics
MeanRoll	mean rotation about X axis, [burst], degrees	Hydra, ADV Time Series Statistics
StdRoll	Std. Dev. rotation about X axis, [burst], degrees	Hydra, ADV Time Series Statistics
MeanTemperature	mean ADV Probe temperature, [burst], degrees C	Hydra, ADV Time Series Statistics
StdTemperature	Std. Dev. ADV Probe temperature, [burst], degrees C	Hydra, ADV Time Series Statistics

What's in the Raw netCDF files

Appendix 8. ADV Raw Statistics File (*s.cdf) Variables

[Data read or calculated for every burst]

Variable	Description, [dimension]	Origin
MeanPressure	mean ADV Probe strain guage pressure, [burst], counts	Hydra, ADV Time Series Statistics
StdPressure	Std. Dev. ADV Probe strain guage pressure, [burst], counts	Hydra, ADV Time Series Statistics
MeanSoundspd	Sound speed, presumably calculated from the other statistics, [burst], m/s	Hydra, ADV Burst Header
brange	range to the boundary, [burst], mm	Hydra, ADV Burst Header
vrange	range to the volume boundary, [burst], mm	Hydra, ADV Burst Header
MeanExtpress	mean external pressure count, if requested for diagnostics, [burst, sample], counts	calculated from samples in burst
StdExtpress	Std. Dev. external pressure count, if requested for diagnostics, [burst], counts	calculated from samples in burst
MeanExtpressfreq	mean external pressure frequency, [burst], Hz	calculated from samples in burst
StdExtpressfreq	Std. Dev. external pressure frequency, [burst], Hz	calculated from samples in burst
MeanExtsensor1	mean external sensor #1 voltage, [burst], volts	calculated from samples in burst
StdExtsensor1	Std. Dev. external sensor #1 voltage, [burst], volts	calculated from samples in burst
MeanExtsensor2	mean external sensor #2 voltage, [burst], volts	calculated from samples in burst
StdExtsensor2	Std. Dev. external sensor #2 voltage, [burst], volts	calculated from samples in burst
CTD_temp	temperature from external CTD, single sample per burst, [burst], degrees C	Hydra, CTD data structure
CTD_cond	conductivity from external CTD, single sample per burst, [burst], S/m	Hydra, CTD data structure
CTD_press	pressure from external CTD, single sample per burst, [burst], dbar	Hydra, CTD data structure
CTD_sal	salinity from external CTD, single sample per burst, [burst], PSU	Hydra, CTD data structure

What's in the Raw netCDF files

Appendix 9. PCADP Raw Burst File (*b.cdf) Variables

[Data read for each profile]

Variable	Description	Origin
burst	burst sequence number	Calculated
profilenum	profile sequence number	ADP profile header
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000, [burst, profile]	ADP profile header, there is a unique time stamp for every profile
time2	msec since 0:00 GMT, [burst, profile]	calculated from the above
Vel1, 2,3	velocity for beam 1-3, [burst, profile, cell], cm/s	ADP profile header
Cor1,2,3 -or- StdVel1,2,3	percent correlation for beam 1-3 when in PC mode, otherwise Std.Dev. (Cm/s) of beam 1-3 velocity, [burst, profile, cell]	ADP profile header
Amp1,2,3	amplitude for beam 1-3, [burst, profile, cell], counts	ADP profile header
Range1,2,3	range to boundary for beam 1-3 when in PC mode, [burst, profile, cell], mm	ADP pulse coherent mode record
Ures1,2,3	resolution pulse velocity for beam 1-3 when in PC mode, [burst, profile], mm/s	ADP pulse coherent mode record
Cres1,2,3	resolution pulse correlation for beam 1-3 when in PC mode, [burst, profile], percent	ADP pulse coherent mode record
heading	heading from compass, [burst, profile], degrees magnetic	ADP profile header
StdHeading	Std. Dev. heading from compass, [burst, profile], degrees magnetic, output for diagnostics only	ADP profile header
pitch	rotation about Y axis, [burst, profile], degrees	ADP profile header
StdPitch	Std. Dev. rotation about Y axis, [burst, profile], degrees, output for diagnostics only	ADP profile header
roll	rotation about X axis, [burst, profile], degrees	ADP profile header
StdRoll	Std. Dev. rotation about X axis, [burst, profile], degrees, output for diagnostics only	ADP profile header
temperature	ADP Probe temperature, [burst, profile], degrees C	ADP profile header
StdTemperature	Std. Dev. ADP Probe temperature, [burst, profile], degrees C, output for diagnostics only	ADP profile header
pressure	ADP Probe strain gauge pressure, [burst, profile], counts	ADP profile header
StdPressure	Std. Dev. ADP Probe strain gauge pressure, [burst, profile], counts	ADP profile header

What's in the Raw netCDF files

Appendix 9. PCADP Raw Burst File (*b.cdf) Variables

[Data read for each profile]

Variable	Description	Origin
extsensor1	external sensor #1 voltage, [burst, profile]	external sensor data structure
StdExtsensor1	Std. Dev. external sensor #1 voltage, [burst, profile]	external sensor data structure
extsensor2	external sensor #2 voltage, [burst, profile]	external sensor data structure
StdExtsensor2	Std. Dev. external sensor #2 voltage, [burst, profile]	external sensor data structure
SoundSpeed	speed of sound in m/s	ADP profile header
extpress	external pressure count, if installed and requested for diagnostics, [burst, profile], counts	ADP profile header
pressuremsb	most significant byte of pressure data, for external sensor, output for diagnostic purposes only, count	ADP profile header
extpressfreq	external pressure frequency, if installed, [burst, profile], Hz	Calculated
ResLag	ambiguity resolution pulse lag, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
ResUa	ambiguity velocity for the resolution lag, mm/s, when in PC mode, [burst, profile]	ADP pulse coherent mode record
ResStart	blanking distance to the resolution cell, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
ResLength	size of the resolution cell, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
PrfLag	profile pulse lag, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
PrfUa	ambiguity velocity for the profile, mm/s, when in PC mode, [burst, profile]	ADP pulse coherent mode record
PrfStart	profile blanking distance, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
PrfLength	length of profile range, mm, when in PC mode, [burst, profile]	ADP pulse coherent mode record
Npings	# of pings averaged in the profile, [burst, profile]	ADP profile header

What's in the Raw netCDF files

Appendix 10. PCADP Raw Statistics File (*s.cdf) Variables

[Data read or calculated for each burst]

Variable	Description	Origin
burst	burst sequence number, [burst]	Calculated
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000, [burst]	ADP profile header, from the first profile in the burst
time2	msec since 0:00 GMT, [burst]	calculated from the above
depth	depth of measurement, [cell], m	Calculated
MeanVel1, 2,3	mean velocity for beam 1-3, [burst, cell], cm/s	calculated from profiles in burst
StdVel1, 2,3	velocity for beam 1-3, [burst, cell], cm/s	calculated from profiles in burst
MeanCor1,2,3 -or- MeanStdVel1, 2,3	mean correlation, percent, for beam 1-3 when in PC mode, otherwise mean of Std. Dev., cm/s, of beam 1-3 velocity, [burst, cell]	calculated from profiles in burst
StdCor1,2,3	Std. Dev. correlation for beam 1-3 only when in PC mode, otherwise not used, [burst, cell], percent	calculated from profiles in burst
MeanAmp1,2,3	mean amplitude for beam 1-3, [burst, cell], counts	calculated from profiles in burst
StdAmp1,2,3	Std. Dev. amplitude for beam 1-3, [burst, cell], counts	calculated from profiles in burst
MeanRange1,2,3	mean range to boundary for beam 1-3, [burst], mm	calculated from profiles in burst
StdRange1,2,3	Std. Dev. range to boundary for beam 1-3, [burst], mm	calculated from profiles in burst
MeanUres1,2,3	mean resolution pulse velocity for beam 1-3, [burst], mm/s	calculated from profiles in burst
StdUres1,2,3	Std. Dev. resolution pulse velocity for beam 1-3, [burst], mm/s	calculated from profiles in burst
MeanCres1,2,3	mean resolution pulse correlation for beam 1-3, [burst], percent	calculated from profiles in burst
StdCres1,2,3	Std. Dev. resolution pulse correlation for beam 1-3, [burst], percent	calculated from profiles in burst
MedianHeading	median heading from compass, [burst], degrees magnetic	calculated from profiles in burst
StdHeading	Std. Dev. heading from compass, [burst], degrees	calculated from profiles in burst
MedianPitch	median rotation about Y axis, [burst], degrees	calculated from profiles in burst
StdPitch	Std. Dev. rotation about Y axis, [burst], degrees	calculated from profiles in burst
MedianRoll	median rotation about X axis, [burst], degrees	calculated from profiles in burst
StdRoll	Std. Dev. rotation about X axis, [burst], degrees	calculated from profiles in burst

What’s in the Raw netCDF files

Appendix 10. PCADP Raw Statistics File (*s.cdf) Variables

[Data read or calculated for each burst]

Variable	Description	Origin
MeanTempera-ture	mean ADP Probe temperature, [burst], degrees Celsius	calculated from profiles in burst
StdTemperature	Std. Dev. ADP Probe temperature, [burst], degrees Celcius	calculated from profiles in burst
MeanPressure	mean ADP Probe strain guage pressure, [burst], counts	calculated from profiles in burst
StdPressure	Std. Dev. ADP Probe strain guage pressure, [burst], counts	calculated from profiles in burst
MeanExtsensor1	mean external sensor #1 voltage, [burst]	calculated from profiles in burst
StdExtsensor1	Std. Dev. external sensor #1 voltage, [burst]	calculated from profiles in burst
MeanExtsensor2	mean external sensor #2 voltage, [burst]	calculated from profiles in burst
StdExtsensor2	Std. Dev. external sensor #2 voltage, [burst]	calculated from profiles in burst
MeanExtpress	mean external pressure count, if installed and requested for diagnostics, [burst]	calculated from profiles in burst
StdExtpress	Std. Dev. external pressure count, if installed and requested for diagnostics, [burst]	calculated from profiles in burst
MeanExtpress-freq	mean external pressure frequency, [burst], Hz	calculated from profiles in burst
StdExtpressfreq	Std. Dev. external pressure frequency, [burst], Hz	calculated from profiles in burst
MeanSoundspd	mean speed of sound, [burst], m/s	calculated from profiles in burst
MeanResLag	mean ambiguity resolution pulse lag, mm, when in PC mode, [burst, profile]	calculated from profiles in burst
MeanResUa	mean ambiguity velocity for the resolution lag, mm/s, when in PC mode, [burst]	calculated from profiles in burst
MeanResStart	mean blanking distance to the resolution cell, mm, when in PC mode, [burst]	calculated from profiles in burst
MeanResLength	mean size of the resolution cell, mm, when in PC mode, [burst]	calculated from profiles in burst
MeanPrfLag	mean profile pulse lag, mm, when in PC mode, [burst]	calculated from profiles in burst
MeanPrfUa	mean ambiguity velocity for the profile, mm/s, when in PC mode, [burst]	calculated from profiles in burst
MeanPrfStart	mean profile blanking distance, mm, when in PC mode, [burst]	calculated from profiles in burst
MeanPrfLength	mean length of profile range, mm, when in PC mode, [burst]	calculated from profiles in burst

What's in the Raw netCDF files

Appendix 10. PCADP Raw Statistics File (*s.cdf) Variables

[Data read or calculated for each burst]

Variable	Description	Origin
MeanNpings	mean # of pings averaged in this profile, [burst]	calculated from profiles in burst
StdNpings	Std. Dev. # of pings averaged in this profile, [burst]	calculated from profiles in burst
MeanBattery	mean of battery voltage	Calculated from profiles in burst

What's in Best Basic Version (*.nc) files

For the archived version, or Best Basic Version of its time series data, the USGS tries to stay as close to the EPIC format as possible. EPIC defines certain oceanographic parameters in standard ways using an EPIC key table, which can be viewed at *ftp://www.epic.noaa.gov/epic/epic.key*. Sticking to standardized definitions allows USGS to combine similar parameters from very different instrumentation during data analysis, and access to great tools such as Don Denbo's ncbrowse, a multi platform data browser written in java. At this point in the process, if velocity from a PCADP has the same EPIC name as velocity from an ADV, then it is simply velocity, in earth coordinates, and most importantly, in the same units, and it does not matter anymore if this velocity was measured by a PCADP or an ADV. Most, but not all variables are EPIC compliant.

Appendix 11. ADV & PCADP Best Basic Version Statistics File (*vp-cal.nc) Variables

Variable	Description, [dimension]	Origin
burst	burst sequence number, [time]	Hydra, ADV Burst Header, PCADP it is calculated
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000, [time, sample | profile]	Hydra, ADV Burst Header, beginning of burst
time2	msec since 0:00 GMT, [time, sample | profile]	calculated from the above
depth	depth of measurements, [depth], m	calculated from metadata supplied by user and in instrument
lat	latitude of measurement, degree east, [lat]	metadata supplied by user
lon	longitude of measurement, degree north, [lat]	metadata supplied by user
u_1205	mean Eastward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
USTD_4097	Std. Dev. of East component of velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
v_1206	mean Northward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
VSTD_4098	Std. Dev. of North component of velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
w_1204	mean Vertical velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
USTD_4099	Std. Dev. of Vertical component of velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
u_1205min	minimum Eastward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
v_1206min	minimum Northward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
w_1204min	minimum Vertical velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
u_1205max	maximum Eastward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data

What's in Best Basic Version (*.nc) files

Appendix 11. ADV & PCADP Best Basic Version Statistics File (*vp-cal.nc) Variables

Variable	Description, [dimension]	Origin
v_1206max	maximum Northward velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
w_1204max	maximum Vertical velocity, cm/s, [time, depth, lat, lon]	calculated from burst data
CS_300	current speed, cm/s, [time, depth, lat, lon]	calculated from u and v
CD_310	current direction, true, degrees, [time, depth, lat, lon]	calculated from u and v
ResU	Mean resolution velocity, Eastward, cm/s, [time, lat, lon]	calculated from burst data
ResV	Mean resolution velocity, Northward, cm/s, [time, lat, lon]	calculated from burst data
ResW	Mean resolution velocity, Vertical, cm/s, [time, lat, lon]	calculated from burst data
adv_nbad	count of bad velocity points in burst that were omitted for statistical calculations, [time, depth, lat, lon]	calculated from burst data
u_dmean	ratio of means for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
v_dmean	ratio of means for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
w_dmean	ratio of means for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
u_dvar	ratio of variance for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
v_dvar	ratio of variance for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
w_dvar	ratio of variance for burst before and after points were edited by deglitch, [time, depth, lat, lon]	calculated from burst data
Hdg_1215	median heading from compass, degrees, [time, depth, lat, lon]	calculated from burst data
HSD_1218	Std. Dev. heading from compass, degrees, [time, depth, lat, lon]	calculated from burst data
Ptch_1216	median rotation about Y axis, degrees, [time, depth, lat, lon]	calculated from burst data
PSD_1219	Std. Dev. rotation about Y axis, degrees, [time, depth, lat, lon]	calculated from burst data
Roll_1217	median rotation about X axis, degrees, [time, depth, lat, lon]	calculated from burst data
RSD_1220	Std. Dev. rotation about X axis, degrees, [time, depth, lat, lon]	calculated from burst data

What’s in Best Basic Version (*.nc) files

Appendix 11. ADV & PCADP Best Basic Version Statistics File (*vp-cal.nc) Variables

Variable	Description, [dimension]	Origin
Tx_1211	mean transducer temperature, C, [time, depth, lat, lon]	calculated from burst data
P_4023	average burst pressure, mbar, [time, lat, lon]	calculated from burst data
SDP_850	Std. Dev. Pressure, mbar, [time, depth, lat, lon]	calculated from burst data
brange	mean range to the boundary, cm, [time, depth, lat, lon]	ADV Burst Header, PCADP pulse coherent mode record
vrange	mean range to the volume boundary, cm, [time, depth, lat, lon], (ADV data only)	Hydra, ADV Burst Header
Sed*_981	mean sediment concentration, g/l, [time, depth, lat, lon], where * is the external sensor number, 1 or 2	calculated from burst data
NEP*_56	mean backscatter intensity, volts, [time, depth, lat, lon], where * is the external sensor number, 1 or 2	calculated from burst data
ATTN*_55	mean attenuation, 1/m, [time, depth, lat, lon], where * is the external sensor number, 1 or 2	calculated from burst data
tran*_4010	mean transmission, volts, [time, depth, lat, lon], where * is the external sensor number, 1 or 2	calculated from burst data
CTDTMP_4211	temperature from external CTD, deg. C, ITS-90, [time, depth, lat, lon]	Hydra, CTD data structure
CTDCON_4218	conductivity from external CTD, S/m, [time, depth, lat, lon]	Hydra, CTD data structure
CTDPRS_4203	pressure from external CTD, dbar, [time, depth, lat, lon]	Hydra, CTD data structure
CTDSAL_4214	salinity from external CTD, sal, PSS-78, [time, depth, lat, lon]	Hydra, CTD data structure

What's in Best Basic Version (*.nc) files

Appendix 12. ADV & PCADP Best Basic Version Burst File (*b-cal.nc) Variables

Variable	Description, [dimension]	Origin
burst	burst sequence number, [time]	Hydra, ADV Burst Header, PCADP it is calculated
time	UT Julian days using USGS convention where 1968-05-23 00:00:00 UT = 2440000, [time, sample | profile]	Hydra, ADV Burst Header, beginning of burst
time2	msec since 0:00 GMT, [time, sample | profile]	calculated from the above
depth	depth of measurements, [depth], m	calculated from metadata supplied by user and in instrument
lat	latitude of measurement, degree east, [lat]	metadata supplied by user
lon	longitude of measurement, degree north, [lon]	metadata supplied by user
u_1205	mean Eastward velocity, cm/s, [time, sample | profile]	calculated from burst data
v_1206	mean Northward velocity, cm/s, [time, sample | profile]	calculated from burst data
w_1204	mean Vertical velocity, cm/s, [time, sample | profile]	calculated from burst data
P_4022	burst pressure, mbar, [time, sample | profile]	calculated from burst data
Sed*_981	sediment concentration, g/l, [time, sample | profile], where * is the external sensor number, 1 or 2	calculated from burst data
NEP*_56	backscatter intensity, volts, [time, sample | profile], where * is the external sensor number, 1 or 2	calculated from burst data
ATTN*_55	attenuation, 1/m, [time, sample | profile], where * is the external sensor number, 1 or 2	calculated from burst data
tran*_4010	transmission, volts, [time, sample | profile], where * is the external sensor number, 1 or 2	calculated from burst data

Appendix 13. Attributes Common to All Hydra Files

<table>
<tr><th>Attribute name</th><th>Meaning</th><th>Source</th></tr>
<tr><td>MOORING</td><td>USGS mooring number</td><td>user provided metadata</td></tr>
<tr><td>WATER_DEPTH</td><td>depth at station</td><td>user provided metadata</td></tr>
<tr><td>latitude</td><td>latitude where tripod was deployed</td><td>user provided metadata</td></tr>
<tr><td>longitude</td><td>longitude where tripod was deployed</td><td>user provided metadata</td></tr>
<tr><td>magnetic_variation</td><td>local magnetic variation where tripod was deployed</td><td>user provided metadata</td></tr>
<tr><td>INST_TYPE</td><td>Sontek ADV | PCADP</td><td>defined by adp2cdf | adr2cdf</td></tr>
<tr><td>DATA_ORIGIN</td><td>organization which collected the data</td><td>user provided metadata</td></tr>
<tr><td>EXPERIMENT</td><td>the name of the experiment</td><td></td></tr>
<tr><td>PROJECT</td><td>the name of the project</td><td></td></tr>
<tr><td>DATA_SUBTYPE</td><td>not used</td><td>n/a</td></tr>
<tr><td>WATER_MASS</td><td>not used</td><td>n/a</td></tr>
<tr><td>COMPOSITE</td><td>1 if this is data concatenated from several other files</td><td>defined as 0 by adp2cdf | adr2cdf</td></tr>
<tr><td>POS_CONST</td><td>1 if position is not consistent</td><td>defined as 0 by adp2cdf | adr2cdf</td></tr>
<tr><td>FILL_FLAG</td><td>1 if data has fill values</td><td>defined as 1 by adp2cdf | adr2cdf</td></tr>
<tr><td>DEPTH_CONST</td><td>1 if depth is not consistent</td><td>defined as 1 by adp2cdf | adr2cdf</td></tr>
<tr><td>DATA_CMT</td><td>not used</td><td>n/a</td></tr>
<tr><td>VAR_FILL</td><td>fill value</td><td>set to 1E35 by adp2cdf | adr2cdf</td></tr>
<tr><td>Deployment_date</td><td>date & time into the water</td><td>user provided metadata</td></tr>
<tr><td>Recovery_date</td><td>date & time out of the water</td><td>user provided metadata</td></tr>
<tr><td>CREATION_DATE</td><td>date the data file was written</td><td>defined by processing software</td></tr>
<tr><td>DESCRIPT</td><td>description of data, such as: Sontek ADV calibrated data statistics file</td><td>defined by processing software</td></tr>
<tr><td>COORD_SYSTEM</td><td>geographical for statistics files, geographical + burst | sample for burst files</td><td>defined by processing software</td></tr>
<tr><td>DATA_TYPE</td><td>time or time+sample or time+profile</td><td>defined by processing software</td></tr>
</table>

Appendix 13. Attributes Common to All Hydra Files

Attribute name	Meaning	Source
history	what was done to the data by which programs	defined by processing software
DELTA_T	interval of variable called 'time'	defined by processing software
VAR_DESC	list of variable short names from variables in the file	defined by processing software
start_time	earliest time value in the file	defined by processing software
stop_time	latest time value in the file	defined by processing software
ADVCalcSampleVolume-Offset	measured distance from ADV emitter to sample volume, calculated from brange-vrange, m	defined by processing software
ADVSampleVolumeOffset	nominal distance from ADV emitter to sample volume, cm	user provided metadata
ADVProbeHeight, PCADPProbeHeight	ADV | PCADP probe height above the boundary, m	user provided metadata
cpuSoftWareVerNum, dspSoftWareVerNum	CPU and DSP chips software version	PCADP sensor configuration, ADV system configuration
ADVType	ADV probe type, 10MHz_5cm | 10MHz_10cm | OCEAN	ADV System configuration
SensorOrientation	up, down or side looking sensor	ADV system configuration, PCADP sensor configuration
CompassInstalled	presence or absence of compass data	ADV system configuration, PCADP sensor configuration
RecorderInstalled	presence or absence of internal recorder	ADV system configuration, PCADP sensor configuration
TempInstalled	presence or absence of thermistor data	ADV system configuration, PCADP sensor configuration
PressInstalled	presence or absence of internal pressure sensor data	ADV system configuration, PCADP sensor configuration
PressScale	strain gauge pressure scale adjustment	ADV system configuration, PCADP sensor configuration
PressScale_2	strain gauge pressure scale adjustment	ADV system configuration, PCADP sensor configuration
PressOffset	strain gauge pressure scale adjustment	ADV system configuration, PCADP sensor configuration
CompassOffset	compass adjustment	ADV system configuration, PCADP sensor configuration

Appendix 13. Attributes Common to All Hydra Files

Attribute name	Meaning	Source
PressFreqOffset	adjustment for frequency pressure sensors	ADV system configuration, PCADP sensor configuration
ExtSensorInstalled	presence or absence of external voltage output sensor data	ADV system configuration, PCADP sensor configuration
ExtPressInstalled	presence or absence of external pressure sensor data	ADV system configuration, PCADP sensor configuration
CTDInstalled	presence or absence of external CTD data	ADV system configuration, PCADP sensor configuration

Appendix 14. Attributes Specific to ADV Data

Attribute name	Meaning	Source
ADVProbeConfig	type of ADV probe	ADV Probe Configuration
ADVProbeSerial	serial number of ADV probe	ADV Probe Configuration
ADVProbeNbeams	number of beams	ADV Probe Configuration
ADVProbeSampInterval	sample interval, RAM cycles between samples, count	ADV Probe Configuration
ADVProbeSamplingVolumeOffset	distance to sampling volume, cm	ADV Probe Configuration
ADVDeploymentSetupSyncMode	master or slave to another hydra	ADV Deployment Setup
ADVDeploymentSetupSampleRate	rate of hydra sampling, Hz	ADV Deployment Setup
ADVDeploymentSetupBurstInterval	burst interval of hydra sampling, sec	ADV Deployment Setup
ADVDeploymentSetupSamplesPerBurst	samples per burst of hydra sampling	ADV Deployment Setup

Appendix 15. Attributes Specific to PCADP Data

Attribute name	Meaning	Source
PCADPProbeConfig	type of PCADP probe	PCADP Probe Configuration
PCADPProbeSerial	serial number of PCADP probe	PCADP Probe Configuration
PCADPProbeFrequency	transmit frequency, hz	PCADP Probe Configuration
PCADPProbeNbeams	number of acoustic beams in the PCADP probe	PCADP Probe Configuration
PCADPProbeSlantAngle	angle of beams, degrees	PCADP Probe Configuration
PCADPProbeProfilesPerBurst	number of profiles in each burst	PCADP Probe Configuration
PCADPProbeBlankDist	distance to center of first cell, m	PCADP Probe Configuration
PCADPUserSetupNcells	number of cells	PCADP User Setup
PCADPUserSetupCellSize	size of each cell, m	PCADP User Setup
PCADPUserSetupAvgInterval	sampling time per profile, sec	PCADP User Setup
PCADPUserSetupProfileInterval	interval between profiles, sec	PCADP User Setup
PCADPUserSetupPingInterval	interval between pings that make up a profile, sec	PCADP User Setup
PCADPUserSetupBurstMode	turn on burst sampling mode	PCADP User Setup
PCADPUserSetupBurstInterval	time between bursts, sec	PCADP User Setup
PCADPUserSetupProfilesPerBurst	number of profiles sampled in each burst	PCADP User Setup
PCADPUserSetupPCMode	pulse coherent mode in use	PCADP User Setup