



URL: cdip.ucsd.edu

Contact info: Julie Thomas
9500 Gilman Drive La Jolla, CA 92093-0214
Email: jot@splash.ucsd.edu Tel: 858-534-3032 Fax: 858-455-5575

Following are examples of data acquisition , processing and dissemination guidelines. For further information check the Documents section of the CDIP URL.

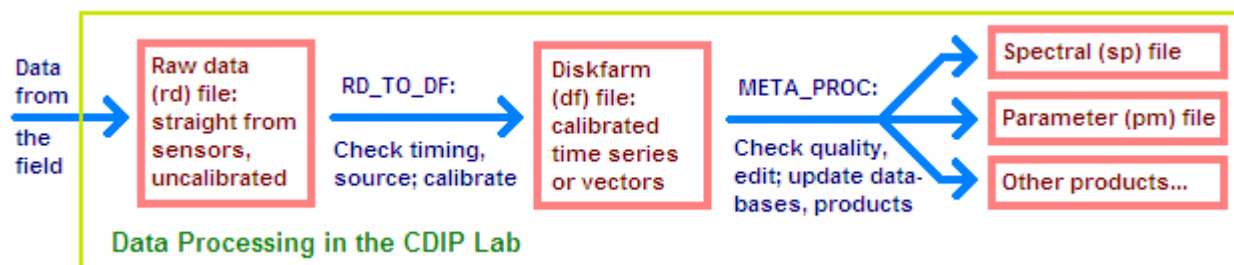
System Organization

Every hour of every day, the computers in the Lab, CDIP's central computing facility at SIO, contact all the active shore stations to collect their latest data. When these data arrive back in the Lab, they are sequentially passed, file by file, through CDIP's automated processing and distribution system. This system performs a wide range of analyses and data transformations, producing everything from error reports and diagnostic e-mail to condensed paramters and web tables.

When the data first arrive in the Lab, they are in the form of rd - raw data - files. This is data directly as read from the sensors in the field, without any significant modification or editing. The data in rd files have not been decoded or calibrated; they are effectively a byte-by-byte record of a sensor's output.

Turning this raw data into to all of the valuable products and information that are found on the CDIP website is basically a two-step process. First, after verifying the source and timing of the rd file, it is calibrated and used to produce a df - diskfarm - file. The df files constitute CDIP's core data set. An accurate record of the readings made by each sensor at each point in time - this is the essential foundation for all of the information that the program provides.

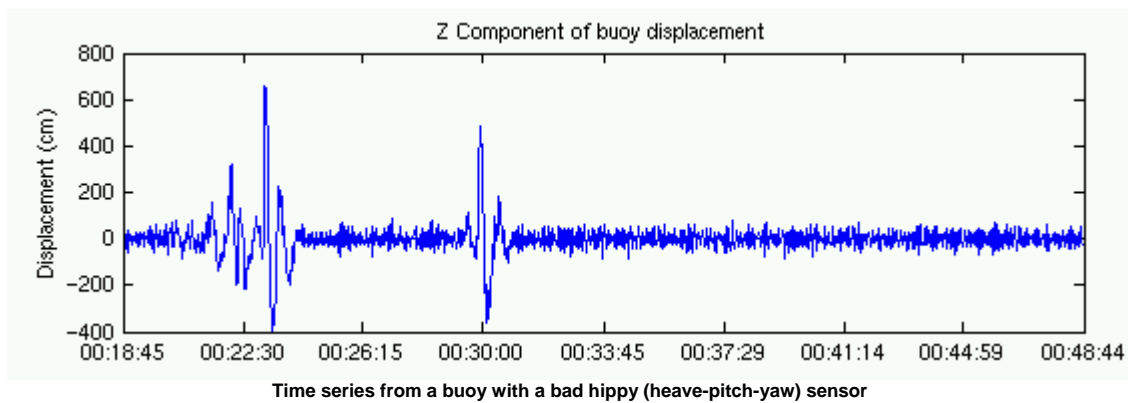
While the df files are accurate records of a sensor's readings, another major question remains: how well is the sensor measuring what it's supposed to? The second step in data processing is to address this very question. A range of quality control checks are performed on the data, to check if they are suitable for further processing. If so, a variety of calculations and transformations are performed on the data, and finally the results are distributed to all the appropriate products. (For a more detailed version of the image below, please see the [processing flowchart](#).)



Quality Control

CDIP needs to provide its users with data which are not only timely, but accurate as well; this is a responsibility that is taken very seriously. Rigorous quality controls are implemented at several stages in the processing, and catch the vast majority of problematic files.

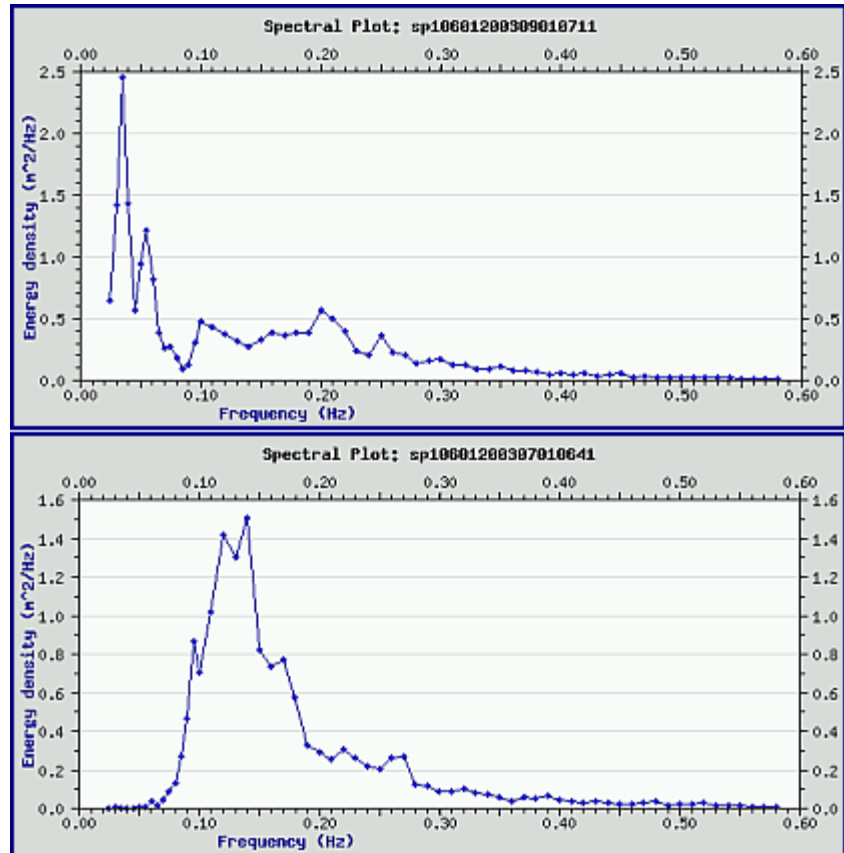
As described above, the first quality control checks ensure that each data file is properly attributed, with its full provenience - both time and place - accounted for. Then, as the data are processed by *meta_proc*, CDIP's full suite of QC algorithms and analyses is deployed. For time series data, a wide range of analyses are used, with different tests applied to different data types. For water column and vertical displacement time series - i.e. wave measurements - the checks include: extreme values test, spike test, mean shift test, flat episodes test, mean crossing test, equal peaks test, acceleration test, and period distribution test. Some of the tests edit the time series, cleaning up the data where possible; others simply flag it bad. Where multiple sensors are deployed in close proximity, the above tests are followed by a battery of comparison tests, to ensure that the sensors are in agreement. For a full description of the tests used and the data types to which they are applied, please refer to our [QC documentation](#).



Once the time series have been processed, the resulting values - condensed parameters and spectral information - also undergo QC checks. For Datawell buoys these checks are quite extensive, since the buoys perform their own time-series handling and spectral processing internally. For example, the time series above shows a problem - spikes in a large, long-period waveform - that CDIP's time series tests could easily identify. But since this data is from a Datawell buoy, the time series was processed internally, so no CDIP editing was applied.

Nonetheless, CDIP's post-processing checks correctly identify the problem with this file, since the resulting spectral distribution and parameter values (in this case T_p) are skewed. Here the long-period spikes result in a spectral shift to lower frequencies, and in an unnaturally high T_p value (approximately 28 seconds).

Two spectral plots. The upper plot shows the shifted distribution of the time series data above; the lower plot shows a more typical spectral distribution.



In addition to the checks outlined above, there is one more full stage of automated QC applied to CDIP data. While all the previous analyses are applied to a single file's data, the final QC tests address a station's data over longer time periods, dozens of files. Once per day, all of a station's recently acquired condensed parameters are compiled and compared. Once again, a range of tests check for spikes, unusual values, and the like, notifying CDIP staff via e-mail if anything seems amiss. (Please read our [post-processing documentation](#) for more details.) All of these automated measures, combined with periodic visual inspections, are very effective in preventing the distribution of erroneous data.

CDIP Quality Control Measures

The tests applied are determined by data/processing type, as summarized in the table below. Note that CDIP's time series products are for the most part unedited; other than checking for gaps in the data and incorrect times, the time series values are not quality-controlled. It is only when the time series are used for spectral and parameter processing that they are submitted to CDIP's full set of quality control routines.

- **AR = Directional Array of Pressure Sensors**
- **BS = Basin Energy Processing, Pressure Sensor**
- **DW = Datawell Directional Buoy**
- **NB = Non-directional Buoy**
- **PR = Single Point Pressure Sensor**
- **SG = Surge Processing, Pressure Sensor**

TIME SERIES VALUES						
TEST: description (click name for more details)	AR	BS	DW	NB	PR	SG
GAP TEST: edit or discard time series with missing values.	✓	✓		✓	✓	✓
DW STATUS/TRANSMISSION BYTE: discard imperfectly received data.			✓			
MAX WAVE HEIGHT: check wave height against max in archive.	✓	✓		✓	✓	✓
FLAT EPISODES TEST: test if time series is changing too slowly.	✓			✓	✓	
SPIKE EDIT: replace spikes with average of point and previous.	✓	✓		✓	✓	✓
MAX/MIN VALUE TEST: for water column, check if values are sensible	✓				✓	
MEAN SHIFT TEST: check for a shift in the mean in the series.	✓	✓		✓	✓	✓
EQUAL PEAKS TEST: check for successive peaks with the same value.	✓				✓	✓
ACCELERATION TEST: check if acceleration is greater than 1/3 g	✓	✓			✓	✓
MEAN CROSSING TEST: check if there are too few mean crossings.	✓			✓	✓	
PERIOD DISTRIBUTION TEST: check for excessive long-period waves.	✓			✓	✓	
UNCORRECTED ENERGY COMPARISON: compare time series variance.	✓					
DEPTH COMPARISON: compare time series means.	✓					
SERIES COMPARISON: check correlation of time series.	✓					
CORRECTED ENERGY COMPARISON: compare depth-corrected variance.	✓					
SPECTRAL VALUES						
TEST: description (click name for more details)	AR	BS	DW	NB	PR	SG
INCIDENT WAVE TEST: check that directions are incident to shore.	✓					
DW STATUS/TRANSMISSION BYTE: discard imperfectly transmitted data.			✓			
BAND DIRECTION MAX/MIN: check Dmean for each band.			✓			
DW CHECK FACTORS: issue warning for bad check factors.			✓			
SUMMARY VALUE MAX/MIN: reject Hs, Tp values outside limits.			✓			
PARAMETER VALUES						
TEST: description (click name for more details)	AR	BS	DW	NB	PR	SG
SPIKE TEST: check for abrupt changes in values between records.	✓	✓	✓	✓	✓	✓
RANGE TEST: apply max/min tests to all parameters.	✓	✓	✓	✓	✓	✓
CONSECUTIVE VALUE TEST: check for excessively constant values.	✓	✓	✓	✓	✓	✓
Dp SPREAD TEST: check if Dp values are too spread out.	✓		✓			
Dp RANDOM TEST: check if Dp values are too variable and random.	✓		✓			
VISUAL INSPECTION: review monthly Hs, Tp, Dp plots.	✓	✓	✓	✓	✓	✓

CDIP Station 071, Set p1 - HARVEST, CA BUOY

Metadata:

- Identification_Information
- Data_Quality_Information
- Entity_and_Attribute_Information
- Distribution_Information
- Metadata_Reference_Information

Identification_Information:

Citation:

Citation_Information:

Originator: The Coastal Data Information Program (CDIP)

Publication_Date: Unpublished material

Title: CDIP Station 071, Set p1 - HARVEST, CA BUOY

Online_Linkage:

<<http://cdip.ucsd.edu/?sub=data&nav=historic&pub=public&stn=071&stream=p1>>

Description:

Abstract:

CDIP's station 071 data sets consist of coastal environment measurements taken in the vicinity of Harvest, Ca. Wave measurements were made in water from 182.88 to 548.64 meters deep. The station is operated by CDIP, and funded by USACE/CDBW.

Purpose:

These data complement CDIP's core mission: measuring, analyzing, archiving, and disseminating coastal environment data for use by coastal engineers, planners, and managers, as well as scientists and mariners.

Time_Period_of_Content:

Time_Period_Information:

Range_of_Dates/Times:

Beginning_Date: 19951201

Ending_Date: Present

Currentness_Reference: Ground Condition

Status:

Progress: In work

Maintenance_and_Update_Frequency: Continually

Spatial_Domain:

Bounding_Coordinates:

West_Bounding_Coordinate: -120.782501