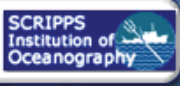


**The Coastal Data Information Program**
Integrative Oceanography Division

**SCRIPPS**
Institution of
Oceanography

**DBW**

**USACE**

Themes Recent Historic **Documents** Station ID

news | contact us | home [index](#) [faq](#)

Documentation

FAQs & Summaries Glossary Publications

Introduction

History and Funding
Program Goals

Wave Measurement

Wave Generation
Wave Dynamics
Irregular Waves
Spectral Analysis
Gauging Waves
Hurricane Events
Tsunami Events

Instrumentation

Underwater Sensors
Surface Buoys
Meteorological

Data Acquisition

System Organization
Hardware
Software

Data Processing

System Organization
Software
Quality Control

Data Management

Stations and Sets
Files and Storage

CDIP Products

Data Formats
Web Products
COOS Integration
QARTOD
Wave Eval Tool
Metadata
Custom Products
NDBC XML/NWS Format
NDBC Dial-A-Buoy
Access Instructions

Related Links

.docs/processing/data_QC.txt

Last Modified: 08/02/2012

Data QC - data checks and editing

This document describes the quality control measures that are incorporated into CDIP's basic data handling programs, outlining the methodology for data checks and editing.

All data are objectively and automatically edited before analysis. They are subjected to a rigorous battery of verification and inspection algorithms.

Pre-processing QC - RD_TO_DF

The first data assessment and QC occurs in the program rd_to_df. Rd_to_df reads raw data (rd) files and converts them into the diskfarm (df) files which are permanently archived by CDIP. The QC performed by rd_to_df does not concern the actual data values received; rather, it checks that the rd file has been properly and completely transmitted back to SIO and that accurate times can be assigned to the data.

Two formats of data are received:

1. Time series data
2. Datawell buoy vectors, from Datawell directional buoys.

Different QC is performed on each data format.

Time series:

CDIP time series data are recorded along with synchronizing time tags, placed together at 60-second intervals. These tags are checked by rd_to_df. When gaps or timing problems are found, the data are either rejected entirely - meaning that no df files are created - or edited.

Currently rd time series are rejected entirely if

- 1) There are more than five gaps in the data;
- 2) There is a single gap of two minutes or more;
- 3) The data are more than 11 minutes older than expected.

If the time series passes these tests but still has gaps, the gaps will be eliminated by concatenating the data together. The resulting df file does not reflect the fact that the original data had gaps; it will appear to be a continuous time series.

Datawell vectors:

Unlike time series rd files, Datawell rd files are always converted into df files. This is true because every vector of data includes an error byte which can be set to indicate the presence of the sorts of problems for which time series files are rejected.

The Datawell vectors include counters and sync words. These values are checked by rd_to_df. When necessary the vectors are edited (i.e. the error byte is reset) to note the following:

- 1) that there is missing data, a gap in the vectors; and
- 2) that there are vectors for which the time is not precisely known.

Refer to .docs/processing/directional_buoys/df_format.txt for more details on the format of Datawell vectors and the error codes used when editing the data.

Datawell iridium and logger files:

Iridium and logger files include checksums and filetype ids in the header. If the filetype is not properly set or the checksums do not match, the file is flagged bad and not processed.

Processing QC - META_PROC

When df files are processed to produce CDIP's various products, additional QC is performed. This QC primarily concerns the data values in the df files. If these values are unreasonable or inconsistent, meta_proc will either edit the values or reject the data. Once again, the details of this QC depend upon the data format.

Datawell vectors

There are two main products created from Datawell buoy df files: xy (displacement) files and sp (spectral) files. Both xy and sp files contain only vectors with error codes indicating that they are error-free. For the xy files, no further QC is done; any displacement value is acceptable if the code indicates that no errors are present.

For the spectral files, a few basic variables are checked to insure that the values are reasonable. The following are the acceptable variable ranges:

0.1 m <= Hs <= 16.0 m
1.7 s <= Tp <= 30.0 s
0 deg <= Dp <= 360 deg
0.0 C <= SST <= 35.0 C

If any of these variables falls outside the acceptable range, the entire spectral transmission is rejected; no sp file is created. (Although SST is not a spectral value, it is measured once per half hour, in correspondence with the spectral data.)

Two additional tests generate errors and warnings, although they do not automatically cause the rejection of the data. One is a check on the magnetic field inclination measured by the buoy; if it is more than three degrees off the expected value for its location, a warning message is sent. Second, the check factors of the spectral processing's frequency bands are inspected; if more than 25% exceed 2.0, a warning is issued.

Note that no editing is performed on Datawell vectors by meta_proc; the data are either accepted as are or rejected.

Time Series

Time series data can be edited or rejected for a wide range of reasons; an extensive range of tests is run on this data set. Except when processing surge data, meta_proc uses the most recent 2048 seconds of the time series, or 1024 seconds if 2048 seconds are not available. For surge data, generally sampled at 0.125 Hz instead of 1 Hz, the processing uses 16384 seconds of data, or 8192 seconds where necessary.

Unlike the Datawell buoys, there is no on-board processing or any internal QC. The specifics of the QC depend on data type - temperature, wind speed, water pressure, etc. - being analyzed.

TEMPERATURE:

The following checks are performed on temperature time series. If any of these tests are not passed, the data are rejected; no editing is done.

Max value - the maximum value must not exceed 33 C.

Min value - the minimum value must not fall below 3 C.

Delta - the delta - the difference between any two consecutive points - in the series must never exceed 2.0 C. (Files processed prior to 11/20/2002 were checked against a limit of 10 C.)

WIND SPEED:

The following checks are performed on wind speed time series. If either of these tests is not passed, the data are rejected; no editing is done.

Max value - the maximum value must not exceed 50 m/s (100 kn).

Min value - the minimum value must not fall below 0 m/s.

WIND DIRECTION:

The following checks are performed on wind direction time series. If either of these tests is not passed, the data are rejected; no editing is done.

Max value - the maximum value must not equal or exceed 360 deg.

Min value - the minimum value must not fall below 0 deg.

AIR PRESSURE:

The following checks are performed on air pressure time series. If either of these tests is not passed, the data are rejected.

Max value - the maximum value must not exceed 1050 mB.

Min value - the minimum value must not fall below 970 mB.

Spike editing is also performed on air pressure data. When a point differs by more than 10 mB from the previous point, it is set to the average of its value and the previous point. If less than one percent of the points are identified as spikes, and they can be removed with five or fewer loops through the time series, the edited data will be accepted and processed; otherwise the data are rejected.

WATER PRESSURE:

CDIP's non-buoy wave measurement is done with water pressure data. The pressure time series undergo the most rigorous QC of any data type. The specifics of the QC depend on the sort of processing and analysis for which the time series is intended - standard, energy basin, or surge.

The tests and editing are done as follows, in the order given.

STANDARD -

Max wave height test - the data are rejected if the wave height (calculated as 4 times the series standard deviation) is greater than the max allowable value.

Flat episodes test - the data are rejected if there are five or more sections in the series with unchanging (or very slowly changing) values.

Spike edit - spikes in the time series - defined as data points > 4 times the series standard deviation from the previous point - are edited by setting them equal to their average with the previous point. If these spikes represent less than 1% of the series and can be eliminated with five or fewer passes through the time series, the data are accepted; otherwise it is rejected.

Max value - after spike editing, the max value must not exceed 2 times the sensor depth.

Min value - after spike editing, the min value must not fall below 0.

Mean shift test - if the mean of consecutive sections of the time series varies by more than 10% of the wave height, the data are rejected. The time series is divided into sections of 256 points for this test.

Equal peaks test - rejects data where the series peaks (or troughs) frequently exhibit the exact same values. (This test is skipped if the time series was acquired using a Paros sensor.)

Acceleration test - rejects the data if the values indicate that the ocean surface was experiencing an acceleration greater than $(1/3)g$ ($g = 9.8 \text{ m/s}^2$) more than three times in the series. (Files processed prior to 11/20/2002 were tested against a limit of g , not $g/3$.)

Mean crossing test - the data are rejected if the values do not consistently cross the mean value in each 1024-point section of the time series. If more than 15% of a section passes without a mean crossing, it is considered a failure.

Period distribution test - if more than 20% of the wave periods fall into a bin with period greater than 22 seconds, the series is rejected.

ENERGY BASIN - Processing used for instruments deployed in low energy areas, i.e. harbors, rivers and protected inlets.

Detrend - the time series is first detrended, removing the tidal component.

Max wave height test - (as above)

Spike edit - (as above)

Mean shift test - (as above)

Acceleration test - (as above)

SURGE - Data collection and processing used for instruments deployed in low energy areas, i.e. harbors, rivers and protected inlets. Initially the sample rates of pressure sensors intended to detect surge were set to 0.125Hz (1 sample every 8 seconds) due to the limited capability to store data. As data storage became more affordable, sample rates changed to 1 Hz. The surge data sets cover longer time (8192-16384 seconds or ~2.3-4.6 hours).

Surge spike edit - surge spikes, defined as deltas of greater than 40 cm, are edited by setting the 'spikey' value equal to the previous value. If spikes represent more than 1% of the data, the series is rejected.

Detrend - (same as energy basin)
 Max wave height test - (as above)
 Spike edit - (as above)
 Mean shift test - (as above)
 Equal peaks test - (as above)
 Acceleration test - (as above)

(For all the details on any of the tests mentioned above, please refer to the code in .f90/editor.f.)

Note that the handling of some stations' water pressure data deviates from the procedures outlined above. The differences are as follows:

Stations 083, 082, 085 -
 - skip the flat episode test if the Hs is less than 50;
 - skip the mean crossing test;
 - skip the period crossing test.

VERTICAL DISPLACEMENT:

Non-directional buoys produce displacement time series. The tests and editing performed on these time series are quite similar to the standard energy QC, as indicated below.

Buoy mean test - checks that the mean of the time series falls within the specifications of the non-directional buoy.
 All standard energy tests as above, except for the min value test, max value test, and acceleration test.

Additional time series QC: ARRAY PROCESSING

CDIP performs directional wave processing on the time series returned by arrays of pressure sensors. Since these time series are synchronized, a number of additional comparison tests can be performed. After each individual time series passes the tests above, the whole group is subjected to the following agreement tests. (For each test, if there is a failure, the outlying time series in the group is discarded, and then the test is repeated on the remaining series.)

Uncorrected for depth energy test - the variance of the time series of the individual sensor must agree to within 20%. This test is only run when the estimated wave height is greater than 30cm. (Note that the estimated wave height is calculated without detrending the time series, so that tidal shifts may sometimes push the estimated wave height over 30cm even when the calculated Hs is very low.)

Depth test - the mean of the time series must agree to within 60 cm.

Correlation test - the correlation coefficient between time series must be at least 0.85.

Corrected energy test - the depth corrected variance of the time series must agree to within 15%. This test is only run when the estimated wave height is greater than 30cm.

One additional type of QC is performed during directional processing as the spectral file is being produced. For each spectral band with a period of greater than eight seconds, meta_proc checks to ensure that the calculated direction is indicative of an incident wave. If not, the direction for that spectral band is discarded.