


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](#)
LATEST NEWS: Data updates will be affected by a planned power outage Wednesday 20 March 7am - 10am

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

Data Management

One great benefit of CDIP's longevity is the fact the program has been able to generate numerous long-term data sets. For many locations, wave information has been collected that spans three decades. All of these data - from the first observations made in 1975 to those being made at this very moment - are archived in a number of standardized, easily accessible formats. These archives form an invaluable resource for researchers and engineers, providing a unique record of wave climatology along our nation's coasts.

Stations and Sets

At the broadest level, CDIP data are organized by station. But what, exactly, is a station? The term is used freely on the website and throughout this documentation without definition, largely because the implied meaning is generally correct: A station is location where CDIP maintains sensors and collects wave and climatological data. Thus stations are named according to their geographic locales - Dana Point, Waimea Bay, etc.

CDIP's station definition

Defining the term station more precisely is quite difficult. For the first couple of decades of CDIP's existence, data were organized strictly according to shore station, i.e. according to the site at which the data were initially recorded and transmitted. So if data from sensors miles apart - say inside and outside of a harbor, as at Noyo (Station 030) - were recorded at the same site, they would fall under the same station. Whereas sensors which had their data recorded at different sites would always be separate stations, even if the sensors were located just a few hundred meters apart. Essentially it was the logistics of setting up the shore station hardware that determined how sensors were grouped in stations, rather than any direct consideration of the wave climates or geographic characteristics of the locations involved.



Mission Bay Harbor:
Five sensors as one station (015)



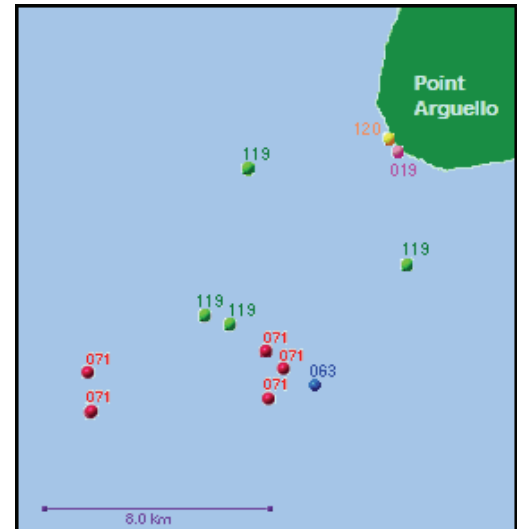
Oceanside Harbor:
Three sensors as three stations (068, 069, 070)

When CDIP started relying more heavily on buoys in the 1990's, the situation became even more

muddled. Unlike pressure sensors and arrays, buoys require frequent maintenance and redeployment. And redeployment sometimes leads to changes in location: a mile further offshore or inshore, atop a different bank or depth contour. When buoys are in continuous use for 10 years or more, it is not uncommon for them to be deployed at half a dozen distinct sites, maybe even more. Thus the use of wave buoys has made stations even more dynamic, and harder to get a handle on.

It is therefore important to remember that while each of CDIP's stations does correspond to a general location, they do not correspond to precise locations or wave environments. Over the program's history a wide range of influences - from funding sources to the logistics of laying cable - have affected how sensors are assigned to stations. As a result, it is possible that a single station covers several varying wave environments, and that a single wave environment contains several different stations.

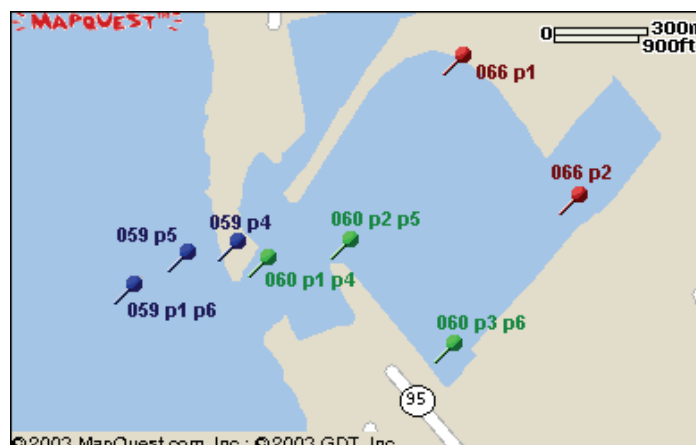
The waters off of Point Arguello and Point Conception have been heavily studied, with sensor data recorded under several different station numbers.



Data sets

When a single station's sensors have collected data from two or more distinct wave environments, it is clearly essential that data from these differing locales be kept distinct. To this end, station data can be organized into separate data sets. Within the processing archive - CDIP's extensive database of processing information - there are detailed instructions on how to organize the data from each station. Which records from which sensors on which dates should be placed in which data sets? The processing archive has the answers.

Most data sets are set up to distinguish distinct wave environments, situations where a station's sensors cover a range of geographic locales. Data sets also have other uses, however, especially for keeping the results of different processing regimes distinct. For instance, pressure sensor data can undergo three types of energy processing: standard wave processing, surge processing, and basin processing. In some locations, data from a single sensor is run in more than one of these regimes. In these instances, the results from surge processing will be kept in a SURGE data set, the results from basin processing will be kept in a BASIN data set, etc. Each data set is assigned a name which indicates either the geographical locale to which it corresponds, or the processing regime which was applied.



Barbers Point Data Sets:

059 - Barbers Point

- p1: Array
- p4: Single Point Onshore
- p5: Single Point Offshore
- p6: Surge

060 - Barbers Point Harbor

- p1: Entrance Basin
- p2: Midpoint Basin
- p3: SW Corner Basin
- p4: Entrance Surge
- p5: Midpoint Surge
- p6: SW Corner Surge

066 - Barbers Pt East Harbor

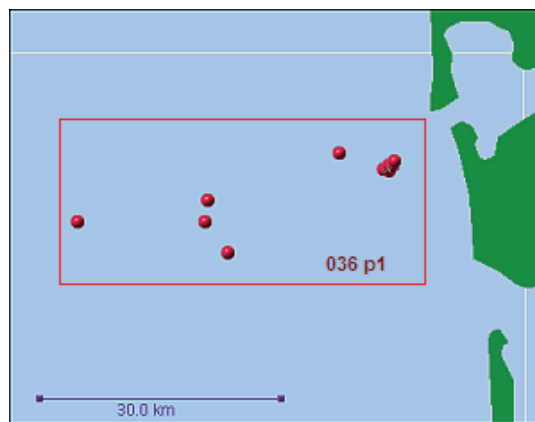
- p1: Northeast Basin
- p2: Southeast Basin

The 12 data sets from Barbers Point are distinguished by location and/or processing type

Setting up data sets based on processing differences is straightforward - either surge processing was applied, or it was not; there is no middle ground. Data sets based on differences in location are more problematic, however. This is especially true in the case of buoys, with their frequent changes in position. How far can a sensor move before it requires a distinct data set? What, precisely, defines a wave environment, and what distinguished one wave environment from another?

Grays Harbor, Station 036:

Data from dozens of buoy deployments - spanning nearly 25nm - are all grouped into one data set, providing a long-term (20+ year) climatological record.



Because CDIP has been involved in a wide range of research projects over the years, collecting data for numerous uses, there is no definitive answer to these questions. In some instances, sensors very close together have been assigned to separate data sets, so that researchers can investigate the subtle differences between them; in other instances, locations miles apart have been grouped together in a single data set to provide researchers with a single, long-term climatological record. Of course, the manner in which CDIP has set up its stations and data sets will not suit all users. But this is never an insurmountable problem, since all the data and products supplied by CDIP can be traced back to their precise origin in space and time, as will be explained below.

Data Storage and Files

As discussed in the Data Processing documentation, CDIP's core data archive is the diskfarm. The diskfarm is composed of millions of separate files, each containing the output of a single sensor over a standard sampling period, generally close to 30 minutes or one hour. These files are named and placed in a directory structure on the station, sensor, and date. In this way, it is easy to obtain the data for any given sensor or station over a specified period of time.

All of CDIP's products are stored in a similar manner. High-volume products - such as time series files and spectral files - generate a new file for each sampling period. Low-volume products - like condensed parameters and nine-band summaries - are stored in monthly files. In all cases, the products and files are easily queried according to date and station. Understanding how the files are named and how the directory structures are organized allows users to ascertain all the characteristics of any data in question.

CDIP's standard files are all named according to a strict format. For single-sample files, the filenames are 19 characters long; for monthly products, they are 14 characters.

Three sample filenames, their components color-coded

df03601198310261022	sp083p2199611091246	pm121p1200401
19-character names		14-char name

Character layout: df - 2-char file type; df=diskfarm, sp=spectral, etc.

036 - 3-digit station number

01 - 2-character stream label

198310261022 - UTC time: YYYYMMDDhhmm or YYYYMM

198310261022 - UTC time; YYYYMMDDhhmm or YYYYMM

The first two characters of the filename specify the file type; see the next section, **File and Data Formats**, for a complete listing of types. Next comes the three-digit station number; leading zeroes are used with any number less than 100. Next comes the two-character stream specifier; the stream concept is explained in the next paragraph. And last is the UTC time of the files. For single-sample files, this is the start time of the data, given to the nearest minute. For monthly files only the year and month are specified, resulting in a shorter filename.

Of the four components of a filename, the stream identifier is probably the only one that requires more explanation. In CDIP jargon, a stream is a sensor and processing specifier. For products created without any special processing instructions, the stream is simply the sensor number. Thus the file df03601198310261022 above comes directly from sensor 1 at station 036; precise details about its location, serial number, or the like can be obtained from the CDIP sensor archive. For more complex products, the stream is alphanumeric - p1, p2, p3 - and refers to a set of handling instructions in the processing archive. This for the spectral file sp083p2199611091246 above, the p2 points us to the processing archive, where we can see that is a directional wave processing stream that drew its 1996 data from sensors 03, 04, and 05 at station 083. Also note that web data sets always require special processing - instructions as to exactly which sensors from which times should be included in the data set - so they always have alphanumeric stream specifiers, as in the parameter file pm121p1200401 above.

For details on the content of CDIP's different data files, please refer to the next section in the documentation, **CDIP Products**.

[Back to top](#)

CDIP's major funding contributors are the **US Army Corps of Engineers**
and the **California Department of Boating and Waterways**.



Official Web Page of the University of California, San Diego