National Park Service
U.S. Department of the Interior

Inventory and Monitoring Program

Southeast Alaska Network

SEAN Oceanography Cast Data User Guide

Abstract

Basic oceanographic observations are maintained by SEAN in a consistent database. This paper summarizes the scope of these data. Methods for accessing data are explained. References are given for finding definitions of the collected attributes.

Table of Contents

Revision history

Revision Date	Base Version	Author	Nature of Change
4/4/2013	1/4/2012	Bill Johnson	Revised for new web interface
1/4/2012	2/12/2011	Bill Johnson	Legacy conversion documentation citation added
2/12/2011	9/22/2010	Bill Johnson	Additional data limitations explained
9/22/2010	3/2/2010	Bill Johnson	Post legacy data conversion project
3/2/2010	1/25/2010	Bill Johnson	New sampling location map
1/25/2010	12/28/2009	Bill Johnson	Documented legacy data access
12/28/2009	2/23/2009	Bill Johnson	Revised for protocol OC-2009.1
2/23/2009	2/17/2009	Bill Johnson	Temporal coverage added
2/17/2009	1/30/2009	Bill Johnson	Known data limitations added
1/30/2009	-new-	Bill Johnson	Initial version

1   The Origin of These Cast Data

Oceanographic data have been collected in the Glacier Bay area since 1993. Data points are obtained by lowering a CTD device through the water column and recording sensor readings at specific intervals. The set of observations collected at a particular station and time is referred to as a cast. Cast data are further stratified by depth. A CTD is fitted with detectors that measure such environmental factors as temperature, conductivity, turbidity, dissolved oxygen, and chlorophyll *a* proxy. At various times different CTD instruments have been employed. Not all factors have been recorded at all times. For example, OBS turbidity was only measured after 1998.

The output of a CTD is captured as "raw data." To be useful, the raw data are normalized against calibration factors unique to the CTD. The sensor outputs are converted into engineering units such as degrees Celsius. Various corrections are applied for items such as physical location of a sensor on the instrument, response delay of a type of sensor, etc. Data are then allocated to 1-meter bins, with the first bin typically labeled depth 0. Finally, quality control measures are taken to flag or disqualify faulty observations and to correct for extraordinary circumstances.

2   Availability of Detailed Data

A major revision to the program was made when protocol OC-2010.1 was adopted in summer 2010. The protocol document provides extensive details regarding the data collected and derived. Each data row retrieved by the database browser tool represents values collected at a particular date/station/depth.

Data from the years preceding 2009 were originally taken and processed under several different protocols. A project in summer 2010 was executed that recast these data under the tighter OC-2010.1 protocol. All data in the database are consistent with the OC-2010.1 detailed definitions. It should be kept in mind, however, that earlier protocols may have had slightly different operating procedures from those detailed in OC-2010.1. For example, in the early 1990s a shorter "soak time" was often employed. Details of the reconstruction process are explained in Johnson and Stachura, available on the SEAN web site.

Where particular data points were not collected (as, say, when a specific detector was not in service) those values are left null. A single data structure holds all cast data, and it may be queried in a consistent manner regardless of particular season, equipment used, or project staffing during data collection.

SEAN's data are in the public domain.

3   Known Data Limitations

HEX files are original instrument data. The only adjustment ever made to HEX files is removal of obviously damaged rows caused by hardware errors. Resulting CNV files are generated from HEX data using a standard, consistent process. The only adjustment ever made to CNV files are "bad flags" applied by the processing software to individual sensor values and/or bins during file creation.

The cumulative database contains SEAN's best available data, which may not always be identical to its underlying source CNVs. For example, occasionally the standard CNV data, which is based on downcast readings, may be revised in the database with upcast values when the downcast exhibits significant data loss. Similarly, misleading CNV readings that wildly exceed expected values are not copied to the database.

Due to variability of field techniques over the years, the first recorded depth bin may not always be at the surface. As a possible convenience to those using the data, if the zero depth bin is missing but a bin is recorded within the first five meters, data are imputed by duplicating the first valid bin up through depth zero. If no valid data occur within the first five meters, no attempt is made to fill in the missing data.

Any item in the database that has been revised from its original CNV value has this reflected in its DATA_QUALITY, DATA_QUALITY_COMMENT, and/or COMMENT fields. Data quality levels are defined in Appendix F of the OC-2010.1 protocol.

Data collected prior to 2009 were run through the rigorous validation required by protocol OC-2010.1. Corrections and adjustments were made as a result of that process. Legacy copies of data no longer exactly match the corrected dataset. The nature of specific changes is documented in:

> Johnson, W. F. and M. M. Stachura. 2011. Creation of Glacier Bay oceanographic information products defined in protocol OC-2010.1 using legacy data covering 1993 through 2008: methods, quality levels, and exceptions. Natural Resource Data Series NPS/SEAN/NRDS—2011/130. National Park Service, Fort Collins, Colorado.

4   Mechanics of Accessing Data

Authoritative data are available through the SEAN web site under the data tab of http://science.nature.nps.gov/im/units/sean/OC_main.aspx. A purpose of this web page is to provide access to the basic cast data stored in the cumulative database. Results are available as both web grids and file downloads. There is a button on the web page to generate each. Grids are populated with results and may be copied and pasted into other applications using standard Windows facilities. The grid is generally limited to 500 rows. This limit protects users from overloading their computers; there are several hundred thousand rows available, each with many columns. In actual operation, it is usually convenient to adjust the filtering parameters until the grid contents confirm that desired data are being delivered. Then data sets of any size may be

SEAN Oceanography Cast Data User Guide

saved on one's workstation in comma separated value files by pressing the "Save Results to a CSV File" button.

Filtering may be done by specifying values in the appropriate fields on the page. Filter parameters restrict the result set. If filter parameters are omitted, then all values are included for that attribute.

A particular year may be selected, or all years. If a set of, say, three or four specific years are desired, then a separate download must be made for each. Alternatively, all years may be selected which is then narrowed down by specifying a range of dates that cross a span of multiple years.
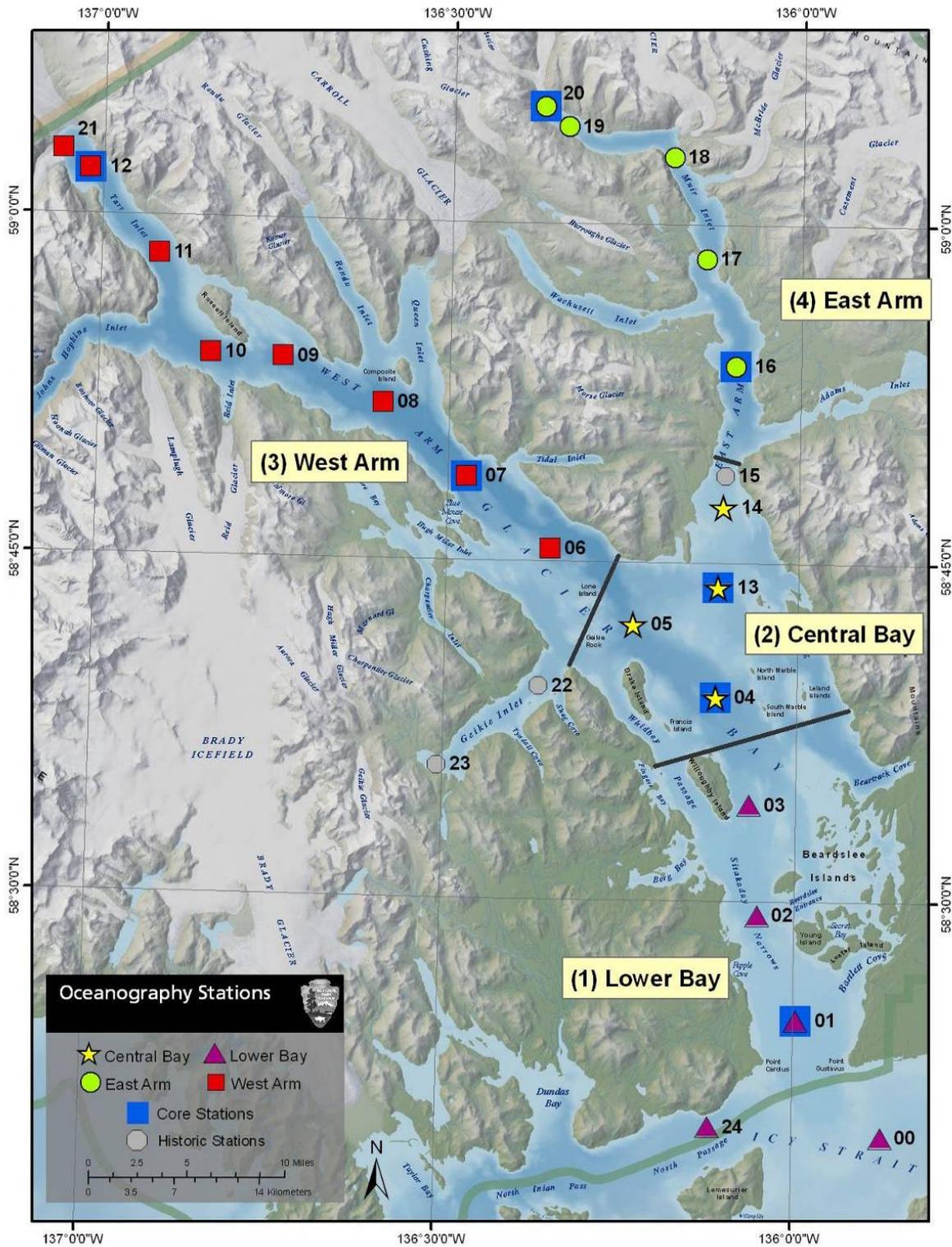
A sampling date range may be specified. If earliest date is omitted, the first date ever sampled is used. If latest date is omitted, then today's date is used. Be aware that specifying a date range lying outside the chosen year will result in finding zero rows.

Sampling depth is recorded as zero for the surface and positive numbers for meters below the surface. If no shallowest depth is specified, 0 is used. If no greatest depth is specified, a number large enough to include the greatest depth possible is used. As of this writing, the greatest depth reached by any cast is 411 meters.

If no region is specified, all are returned. Check boxes may be used to restrict results to certain regions. Each station exists in only one region. The following map shows the relation between stations and regions.

A check box is available for specifying whether or not to include comment fields in the extract. Because comments on each row can be lengthy it is sometimes more convenient to retrieve data without them.

5 Spatial Coverage

A station is a single unique geographical location. It is possible that new stations may be added

to the program and old ones dropped over the years. In such cases a previously used the station identifier is never reused. Original data are retained and available under the proper station identifier, regardless of whether or not a particular station is still being actively surveyed.

6 Temporal Coverage

Below is a snapshot of data coverage as of this writing. The most up-to-date version of this table is available on the web site under the Original Source Data tab at "Data Availability".

7 Authority

The data browser pulls rows directly from SEAN's vital sign database. These data reflect the authoritative source. Final analyses should always be based on data pulled from the browser. Other sources, such as spreadsheet files, paper tables typed back into computers, or email communications cannot be relied on to be authoritative.

SEAN also submits cumulative data files periodically to the National Ocean Data Center and NPS Datastore. It should be noted that the maintenance of data copied to these repositories is out of the control of SEAN. If any data discrepancy is seen among the sources, the SEAN website should be considered the authority.

8   Analysis Facilities

The data browsing feature gives everyone the ability to extract data of interest onto their local computers. From there they are free to use whatever tools they choose to perform whatever analyses they are interested in. The date browser does not provide embedded analytical tools.

9   Further Information

Additional information, as well as current contact persons, may be found in the oceanography section of http://science.nature.nps.gov/im/units/sean/ .

10 Description of Data Fields

Formal metadata, including brief field descriptions, are available on the SEAN web site at http://science.nature.nps.gov/im/units/sean/OC_Main.aspx . More detailed explanations are provided in the protocol document itself, available on SEAN's main oceanography web page.