

cBathy Data Test Bed

Rob Holman¹, Greg Wilson¹ and Kara Burch²
July 29, 2021

Background:

The 2013 cBathy publication [Holman *et al.*, 2013] tested the original algorithm against a set of 16 CRAB (Coastal Research Amphibious Buggy) bathymetry surveys from 2010-12. Since then, the cBathy algorithm has continued to evolve and is now being used by a large international group of researchers. With further algorithm developments, there has become a need for a more up-to-date test data set that can be used as a standard for verifying improvements or for anyone who wishes to study cBathy or other remote sensing algorithms.

This has led to the development of a new data test bed that can become a standard for Argus research. The new test data set will continue to be based on data from the Army Corps of Engineers Field Research Facility (FRF) at Duck, NC, but has been modernized to use more mature data collections that began with the installation of the next generation Argus Station on 02/21/15, called 'argus02b'. The goal is to create a data repository that has high quality survey data, sufficient Argus data coverage (not judging image or geometry quality), and support environmental data. Surveys with incomplete coverage or limited Argus or environmental data have been omitted but are available in separate directories.

The selection process included:

1. downloading all available FRF bathymetry survey data (CRAB or LARC (Lighter Amphibious Resupply Cargo)),
2. accepting only surveys with full coverage from the dune to $x=800$ m (cross-shore), $y = 0$ to 1000 m (along-shore),
3. searching all Argus cBathy collections for the period beginning three days prior to the survey through the survey date and accepting only survey dates with at least 90 of a possible 96 cBathy data sets,
4. downloading wave and tide information for each cBathy collection time.

Usage:

The goal is to provide easy access to a verified, complete set of bathymetry surveys and cBathy data collections. The key file is *bathyTestDataSet.mat* (included in this GitHub repository) that contains a record for each survey as described below. This includes the bathy survey data, a listing of all of the cBathy input data files associated with each survey, and a listing of environmental conditions for each cBathy data run. One likely implementation would be to load each gridded cBathy input file and run some version of cBathy, and then compare answers to the survey data. Note the need to pay attention to errors in both the survey and the cBathy estimates.

¹ College of Earth, Ocean and Atmospheric Sciences, Oregon State University

² Innovative Imaging and Research, Corp., Stennis Space Center, MS

The cBathy data files are accessed using FTP to the Oregon State University CIL (Coastal Imaging Laboratory). The data are organized and stored by day and year in `ftp://cil-ftp.ceoas.oregonstate.edu/pub/argus02b/`. A separate MATLAB script (*downloadcBathyTestFiles.m*, included in this GitHub repository) is provided that illustrates how to access and download cBathy input files from the CIL FTP site.

bathyTestDataSet Structure:

- The included mat file, *bathyTestDataSet.mat*, stores all of the relevant information in a structure, 'bathyTestDataSet' which includes fields
 - surveyDateStr – survey mm/dd/yy, expressed as a string for plot labelling,
 - surveyfn – source filename of the gridded survey,
 - survey – bathymetry survey ground truth (structure described below),
 - ArguscBathyFiles – list of the related argus cBathy input data files,
 - env – environmental metadata (described below).
- The mat file also includes a variable, 'ftpServer', which identifies the ftp server where the cBathy input files are stored.

Bathymetry Data:

- Survey data were downloaded from the FRF thredds survey server in transect format, vector lists of x, y, z, t survey points along transects (done using `ncfiles_scraper.sh` from Greg Wilson, supplemented by recent files downloaded by John Stanley).
- The transect data were gridded using `loessInterp` (using m-file `saveGriddedCRABData2015On`), with smoothing of $L_x=30$, $L_y=75$ over the region $x = [80:10:800]$, $y = [0:25:1000]$. Results are stored at the CIL under `griddedCRABData` in files labelled `CRAByyyymmdd.mat`, but are also captured in 'bathyTestDataSet'.
- Data are stored as a 'survey' structure with fields:
 - dnBar – matlab datenum of mean of time field (mid-survey time)
 - dnBarStrShort – easy to read string, mm/dd/yy for labeling
 - platform – survey vehicle (string)
 - raw – transect data with fields x, y, z, t
 - gridded – loess-interpolated data with fields xm, ym, zi (grid) and rmse (grid)

Argus Data:

- cBathy data filenames were found for the survey day plus the three preceding days (total of four days of half-hourly cBathy data sets per survey). There was occasional down time due to equipment issues, so we required that there are at least 90 of the possible 96 cBathy collections to be acceptable.
- cBathy data are saved in 'mBW' files (derived from matrix `BeachWizard`) and include standard merged cBathy time stack output including XYZ (Mx3 matrix), T (1xN vector), RAW (NxM matrix) and CAM (Mx1 vector) where N is the number of time samples and M is the number of pixels. These data are sufficient to run cBathy.

Environmental Data:

- Tide data plus wave data from the 8 m array and the 26 m waverider were downloaded from thredds servers, then interpolated for the start time for each cBathy data run.

- They are saved in a structure called ‘env’ with fields: epoch (start of cBathy run epoch time), predicted tide (tidePred) and measured tide (tideMeas), then wave height, period, and direction, H_s , T_m and waveDir for both the 8 and 26 m gauges. Finally, the daily-averaged H_s , T_m and waveDir for each instrument and each day are listed specifically. Descriptions of variables are included in the original netcdf data files. Units are meters, seconds or degrees, appropriate to the variable.

Test Data Set:

In all there are 39 full survey data sets spanning from 05/20/15 to 04/17/19. Table 1 lists all survey dates included in the test data set. Half-hourly cBathy data is provided for survey date and three preceding days, giving a total of 3,735 cBathy data collections. Figure 1 shows the histograms of daily averaged environmental conditions for the 26 m (left) and 8 m (right data). The data set represents a fairly wide range of conditions although not much with high energy.

Table 1. List of test data set survey dates

Survey Dates	
5/20/2015	7/7/2017
8/8/2015	9/5/2017
9/15/2015	11/21/2017
11/16/2015	12/11/2017
12/21/2015	2/26/2018
1/14/2016	3/15/2018
3/10/2016	4/18/2018
4/28/2016	5/23/2018
5/27/2016	6/18/2018
6/28/2016	7/20/2018
7/26/2016	8/14/2018
8/17/2016	9/19/2018
11/16/2016	10/17/2018
1/5/2017	11/20/2018
1/27/2017	12/13/2018
4/11/2017	2/1/2019
5/18/2017	2/15/2019
6/6/2017	3/25/2019
6/13/2017	4/17/2019
6/21/2017	

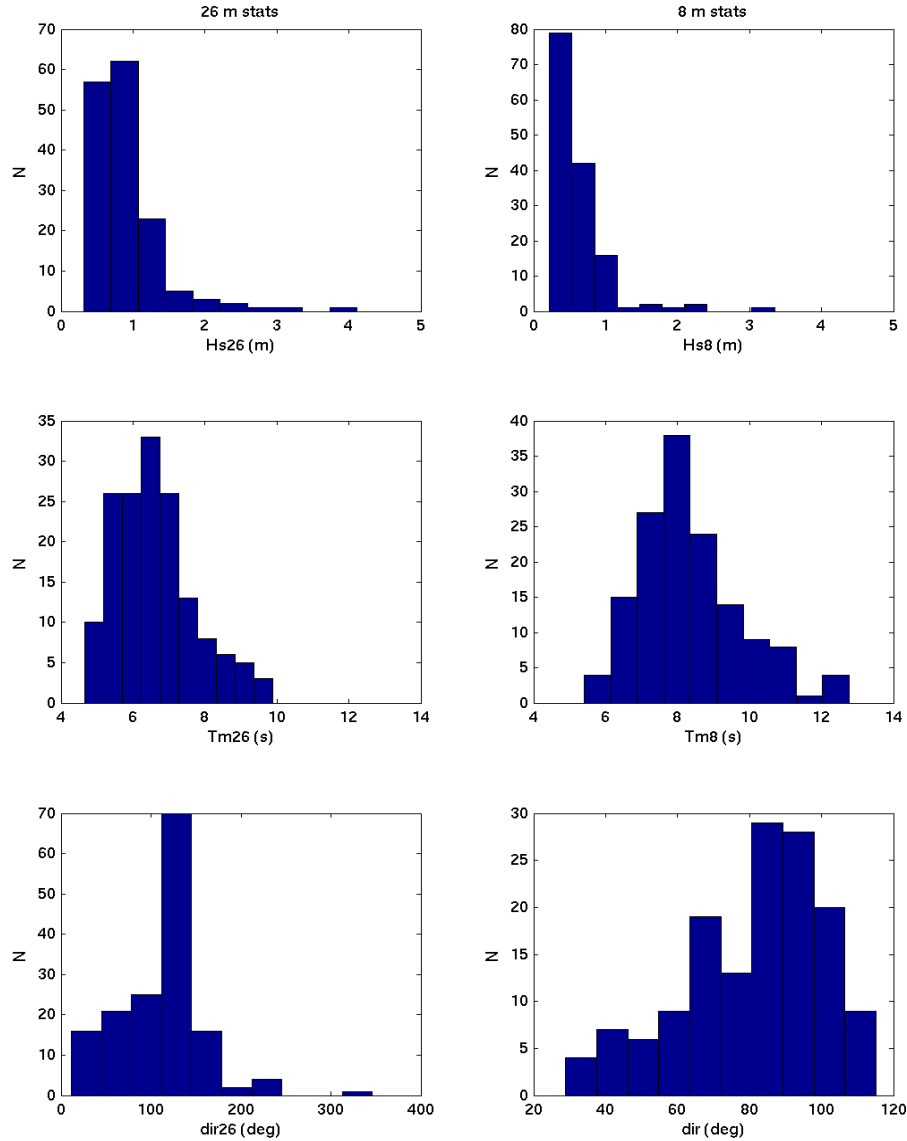


Figure 1. Histograms of environmental variables (H_s (top), T_m (middle) and dir (bottom)) for the bathy test bed data set for the 26 m data (left) and the 8 m data (right). Note different scale on wave direction panels (bottom).

Acknowledgments:

This work was funded by the US Army Corps of Engineers through an SBIR awarded to Innovative Imaging and Research, Corp (contract numbers W912HZ18C0035 and W912HZ20C0014), with subcontract to Oregon State University.

References:

Holman, R. A., N. G. Plant, and K. T. Holland (2013), cBathy: A robust algorithm for estimating nearshore bathymetry, *Journal of Geophysical Research*, 118, 1-15, doi:10.1002/jgrc.20199.