MBARI Power Buoy IFCB Data Stream Plan

Figure 1. *IFCB 161 sitting in a bucket before the deployment of the MBARI Power Buoy on April 26, 2022. Photo by Denis Klimov.*

First Draft: July 11, 2022
Second Draft: Dec. 18, 2022
Author: Patrick Daniel

Introduction

The University of California San Diego purchased an Imaging Flow Cytobot (s/n 161) in 2020 through Prop. 1 funding from the State of California as part of a project to expand the harmful algal bloom (HAB) monitoring capacity within state waters. In late 2021, the instrument was deployed on the MBARI Power Buoy, an experimental buoy that generates electricity from wave action around 11km southwest of Moss Landing, California in Monterey Bay.

The instrument is embedded in the foam disc of the MBARI Power Buoy (Figure 3) with an exposed intake screen that draws seawater up and into the intake at the top of the instrument. In this configuration, the IFCB is sampling within the upper meter of the ocean. Data are telemetered hourly to machines on shore at MBARI, where images are classified using an experimental convolutional neural network (CNN) classifier that is trained on images from the nearby Santa Cruz Wharf IFCB (IFCB s/n 104, aka “Tina”).

As of Decmeber 2022, the IFCB 161 has been deployed on the MBARI Power Buoy three times, generating over 13 million images and 250 gigabytes of data.

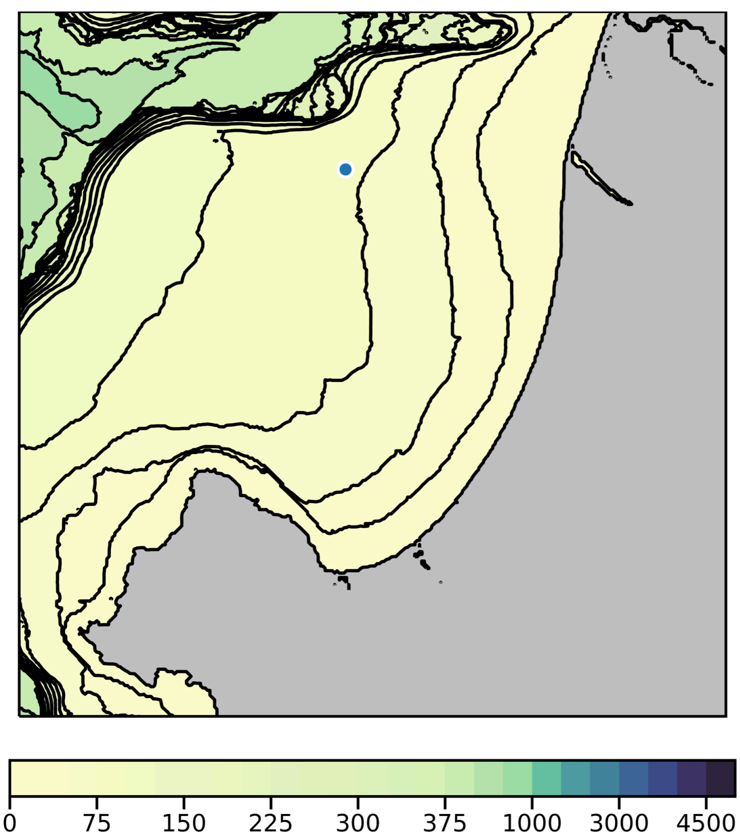


Figure 2*. Map of South Monterey Bay showing the deployment location of the MBARI Power Buoy (Blue Circle). The colored contours show the bathymetry*

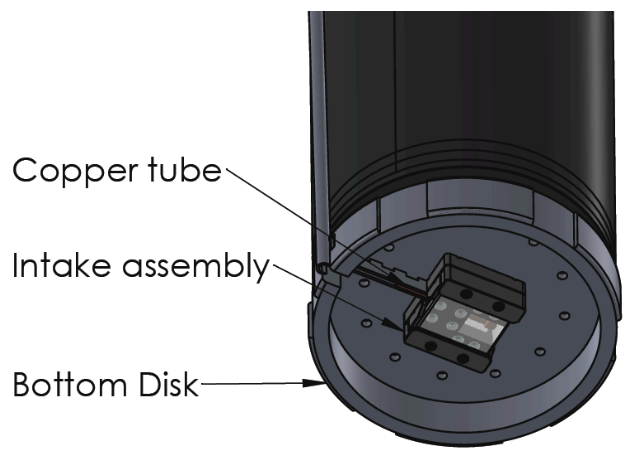
profile of this region. The instrument was deployed over the shelf where the depth is ~80 meters.

Deployment Description

Coordinates: 36.744 N, 121.8845 W
Ocean Depth: ~80 meters
Deployment Length at Full Resolution: 6 months
Sampling Depth: 0.5 - 1 meter from surface
Auxiliary Measurements: Fixed CTD at 5 meters from surface (SBE37), co-located wire walker with CTD+ DO + Fl (RBR), co-located Spotter Wave Buoy

Power
The MBARI Power Buoy is capable of generating up to 1kW in stormy conditions but more typically generates 150 to 300 W, more than enough power for the ~30 W used by the IFCB during continual operation. Power is stored in a battery bank in the buoy hull. Power to the instrument can be controlled remotely through a relay.

Sampling
Samples are collected from the intake screen at the bottom of the instrument where the instrument is exposed to seawater. Calculation and testing showed that the volume of the water sample, when configured with a debubble routine, is enough to clear the intake line between samples. Given the location of the IFCB, samples are taken 0.5 to 1 meter from the sea surface. The sampling frequency is set to 24 minutes and typically sampled a total volume of 4 to 5 mL depending on the density of phytoplankton in the water. When under continuous operation, a biocide-bleach-beads (BBB) cycle is run after every 40 samples or about once every 21 hours. The volume of BBB reagents currently limits IFCB deployments to around six months at full operation.



Fouling
During the first deployment, copper tubing was used for plumbing seawater from the intake screen to the top of the instrument, where the intake valve is located. Over time, the copper tubing and the plating of the IFCB housing showed signs of corrosion, attributed to the sacrificial anode being on the top of the instrument, not submerged in seawater. For subsequent deployments, the copper tubing was replaced with Teflon tubing, and a sacrificial anode was

added to the bottom plate of the IFCB housing. Biofouling from hydroids was significant after the first deployment. Currently, there is no indication of the effect fouling has had on the phytoplankton communities or abundance observed by the IFCB, and will need to be explored further. For subsequent deployments, the bottom of the buoy disc was painted in an anti-fouling paint, and a copper mesh screen was added over the intake.

Buoy Effects

The disc hull of the MBARI Power Buoy experiences vertical acceleration and changes rapid changes in pitch from propagating surface waves. To test the effect of changes in pitch on the IFCB performance, a benchtop experiment was run by Denis Klimov, where the pitch of the instrument was incrementally altered while running bead samples. The initial results indicated that at extreme angles, the optic alignments were affected, resulting in a slight blurring of images. However, this effect has not been detected during *in situ* deployments. Hull orientation and vertical acceleration are measured by an accelerometer in the Power Buoy hull, and this data may be explored at a later stage.

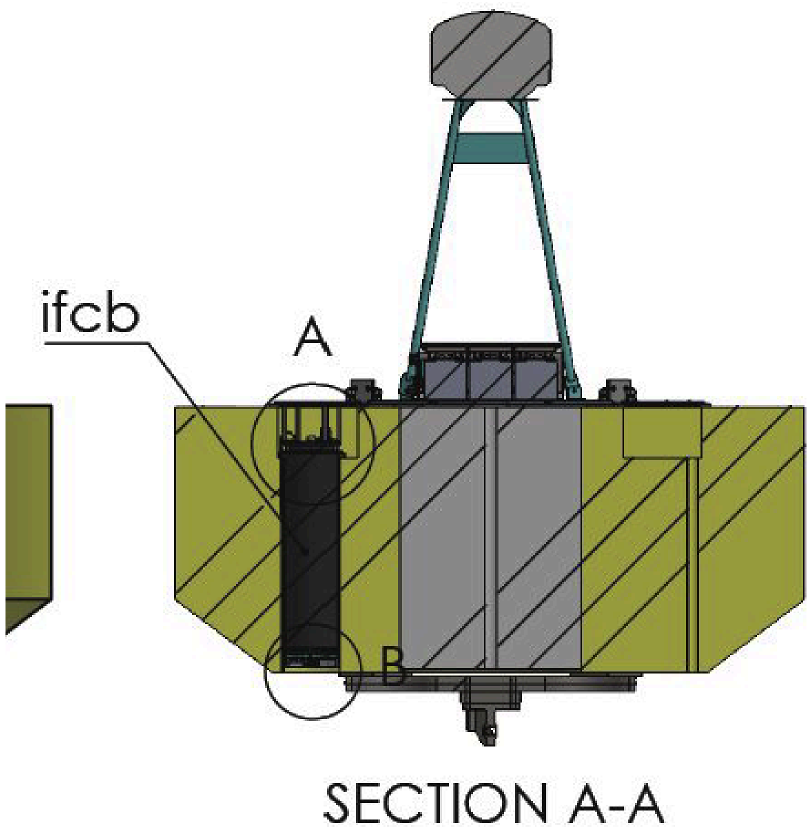


Figure 3 *Engineer drawing of a section of the MBARI Power Buoy. The IFCB is shown embedded in the foam hull of the buoy disc. Figure: Denis Klimov*

Telemetry

Data and communication with the instrument is achieved through a cellular modem using a 4G LTE MOXA cell modem. This modem creates a dynamic VPN tunnel to the MBARI network and allows communication and data transfer between the IFCB and a server at MBARI. Cellular service was generally abundant, with intermittent weather periods that led to periods of reduced bandwidth. During these periods, data were telemetered shoreside intermittently until bandwidth was restored and the full dataset could be updated.

Data Processing

Data Transfer

Data is generated on the instrument every ~24 minutes and stored on the hard drive of the IFCB deployment. Each sample generates three files: a header file (.hdr extension), a Region of Interest file (.roi extension), and an instrument metadata file (.adc extension).

Data are retrieved from the IFCB using a python script that compares the files between the local server and the IFCB data folder (https://github.com/patcdaniel/ifcb-powerbuoy-retrieval). The python script is executed 30 minutes past the top of each hour and controlled using a cron script running on the *particle.shore* machine:

```
30 * * * * python3 /home/ifcb/bin/get-ifcb-data.py
```

New files are transferred using the secure copy (`scp`) subprocess[1]. Due to periods of limited bandwidth and sharing the cellular modem with other research groups, data transfer speeds are generally restricted to 1000 kbits/s. Files are transferred directly to a mount on the *particle.shore* machine that is configured for RAID 5. Files are organized into the scheme suggested by Axiom, "/CA-IFCB-161/YYYY/DYYYYMMdd", where YYYY represents the year, MM represents the month with a leading zero, and dd represents the day of the month with a leading zero. For example, files generated on April 30, 2022 would be organized into: /CA-IFCB-161/2022/D20220430/.

The bead sample data files are transferred to a directory on the *particle.shore* named "/beads". Data are transferred to Axiom using rsync to an FTP server. This command is made following the successful retrieval of data from the IFCB. A public key from the *particle.shore* machine was generated and shared with Axiom.

Bandwidth strain on the Axiom FTP server became an issue as the full network came online. To reduce bandwidth from the power buoy IFCB, the hourly rsync call is limited to the most recent four days of data through a rather complicated bash call:

```
rsync -qzme ssh --files-from=<(cd /opt/ifcb-data/power-buoy-deployment/CA-IFCB-161 && find . -type d \(
-name D$(date '+%Y%m%d')* -o -name D$(date '+%Y%m%d' --date='-1day')* -o -name D$(date '+%Y%m%d'
```

[1] Initially, *rsync* was used to compare between the local and remote directories and transfer files, but an unknown issue arose with *rsync* that caused the SSH tunnel between the powerbuoy network and shore.mbari network to crash.

```
--date='-2day')* -o -name D$(date '+%Y%m%d' --date='-3day')* \) ) --recursive -R
/opt/ifcb-data/power-buoy-deployment/./CA-IFCB-161 ifcb_mbari@data.axds.co:CA-IFCB-161
```

Data Storage and Redundancy

IFCB data during deployment is stored in three different locations. On the instrument, they exist for the deployment duration and should be considered temporary storage, as the data are removed as maintenance before each redeployment. Data are stored on the RAID mount of the *particle.shore* machine at MBARI. The mount is configured in RAID 5 and has ample storage capacity (~17 TB). The RAID configuration means that if one of the drives fails, it will be possible to restore the data. This machine is located on-site at MBARI in a protected server room and is only accessible while on the MBARI network. Finally, data are transferred to Axiom, acting as an off-site redundancy for the raw data. In the current network configuration, concerns about data loss are minimal.

Data Access Points

Data are primarily accessible through the Axiom IFCB Dashboard and the Dashboard API (see repo: https://github.com/WHOIGit/ifcbdb). Data on the instrument and MBARI machines is not publicly accessible and can only be accessed with credentials through the MBARI network.

Feature Extraction

As of September 2022, feature extraction for IFCB 161 is only done on-demand on local machines using the IFCB Matlab routines. Feature extraction is needed for biovolume calculations, so some routines must be implemented in the future. Blobs and Features are extracted using the IFCB Anaysis Matlab library (https://github.com/hsosik/ifcb-analysis). Version 2 features and blobs are generated by a workstation at UCSC and as of this time (2023-01-17), not being transfer to axiom. The long-term plans for Axiom Data Science to feature extraction and biovolume calculations for the entire California IFCB network. This practice will centralize and standardize feature extraction and data storage.

Classification

Currently, image classification for the power buoy is run every hour from an convolutional neural network (CNN) model on the *particle.shore* machine. For a detailed model training and architecture description (see the model card document). The result is a time-series point for each sample and the total images for each class. Because the sampling volume change depending on the number and size distribution of particles in the water, the sample volume is calculated for each syringe based on the difference between the run time and the inhibit time and an assumed flow rate of 0.25 ml per minute.

The **pyifcb (**https://github.com/joefutrelle/pyifcb**)** python library extracts image data from a .roi file. For classification, each image needs to be resampled to a shape of (224, 224, 3) and each image added to an array. For a sample with 1200 images, the array will be (1200, 224, 244, 3). As images are in greyscale, each pixel is represented by a float value from 0 to 1. This value is repeated for the three color channels, a requirement of the model architecture.

Model predictions are made for each image, and the output is an array of probabilities for each class. If the highest probability of all classes is below 0.75, the image is classified as "unclassified." The prediction probability threshold represents error within the model. For the power buoy, most higher concentrations of unclassified data result from organisms not well represented in the training set. As part of model development, this threshold is useful as a proxy for model accuracy over the deployment. The threshold of .75 was arbitrarily chosen.

These data are then aggregated into comma-separated values (.csv) text file where each row represents a single syringe sample and each column is the total number of images identified within a class. The last column is the syringe sample volume in mL.

Plotting

Plots of the cell concentration and relative abundance of time series are generated as .png files from the model output using a python script executed by a CRON job each hour. Images are copied to a CeNCOOS web-accessible folder and displayed on a page publicly accessible at https://www.cencoos.org/data/ifcb-161/.

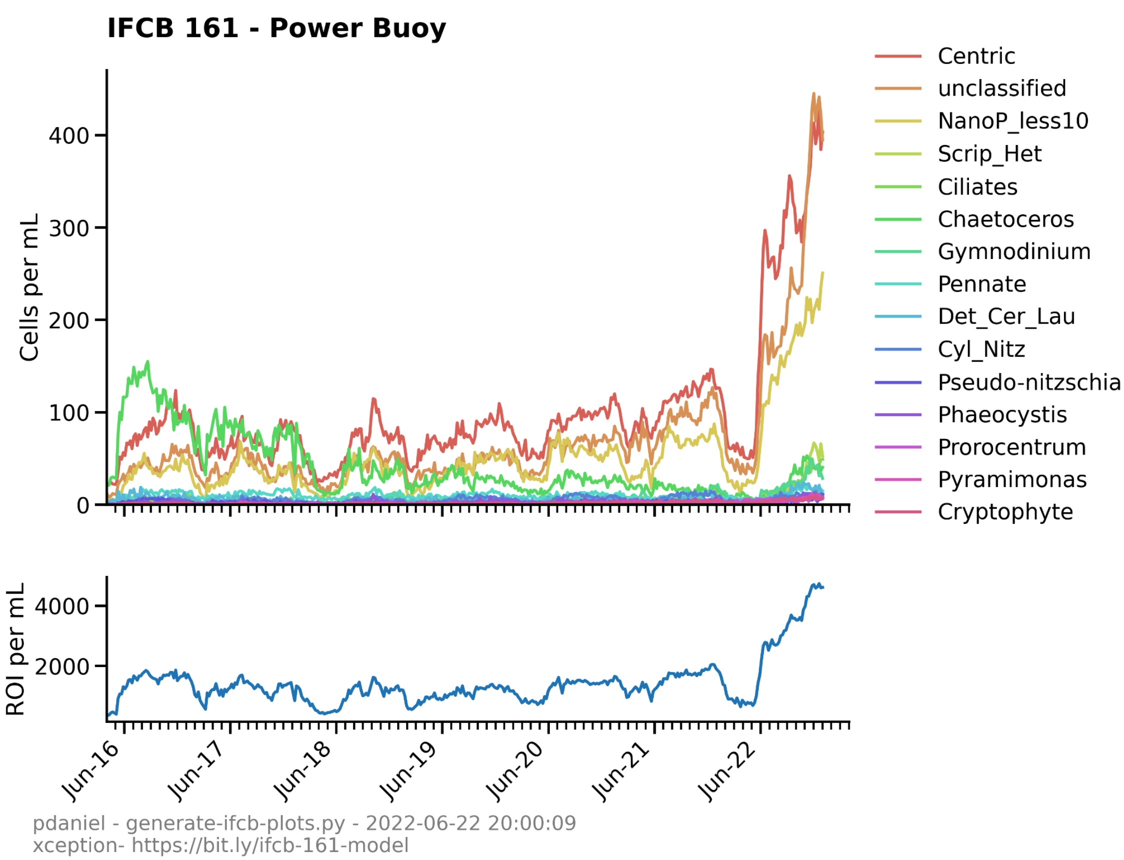


***Figure 4**. (Top panel) The classifier-identified time series of cell concentrations of the top classes. (Bottom panel) the total number of images taken per mL from each syringe.*

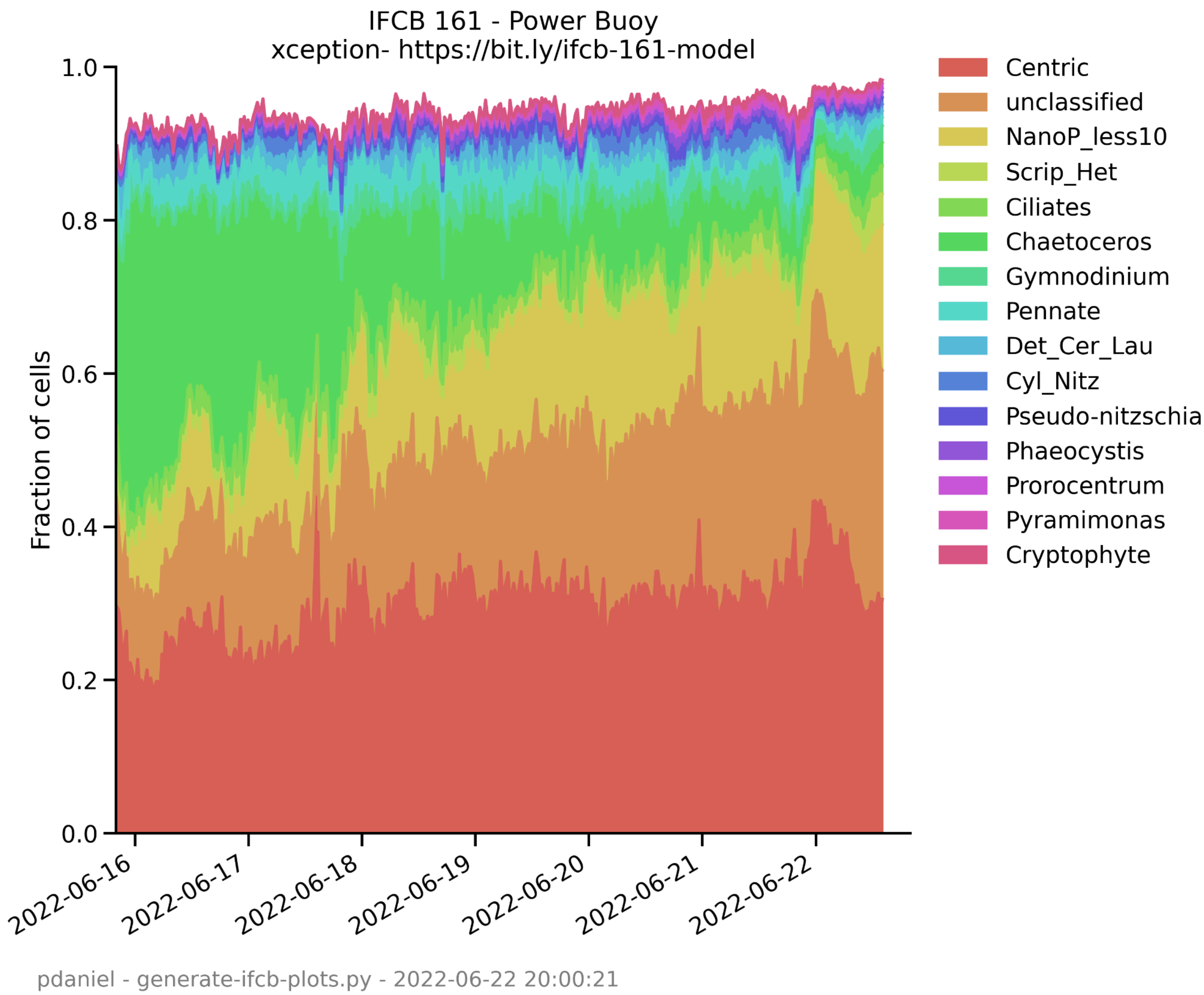


Figure 5. *Time series of relative abundance (i.e. fraction) of the top classes identified by the classifier. Data are shown in the full resolution.*

Data Use and Quality Control

Data Availability

The raw data are made publicly accessible through the IFCB Dashboard as soon as they are collected from the instrument. The dashboard includes some tools for downloading images, viewing metadata, and aggregating data. Currently, the classified data are not publicly available, apart from the plots generated and posted on the CeNCOOS website each hour.

Quality Control

No quality control is applied to the real-time data. If human-reviewed images are out of focus or cut off, they can be removed, but they are generally rare and mostly applies to when the instrument is being deployed or recovered.

Metadata Profile

Currently, there is no standardized metadata profile for raw ifcb data. “Standards and Practices for Reporting Plankton and Other Particle Observations from Images” (doi: 10.1575/1912/27377) outlines how categorized classified data using a readiness scale:

Level 0: Raw images collected by the imaging-in-flow cytometer
Level 1a: Automated classification by an algorithm (automated annotation) and/or manual annotation
Level 1b: Individual level counts with automatic (including interpretation of class scores or probabilities) and manual classifications, and biovolume and size parameters for each ROI
Level 2: Summary data for sample e.g., taxonomic groupings

Most of relevant the instrument level data is stored in the .adc and .hdr files for each sample, but other important data including coordinates and depth are not by default integrated into those files.

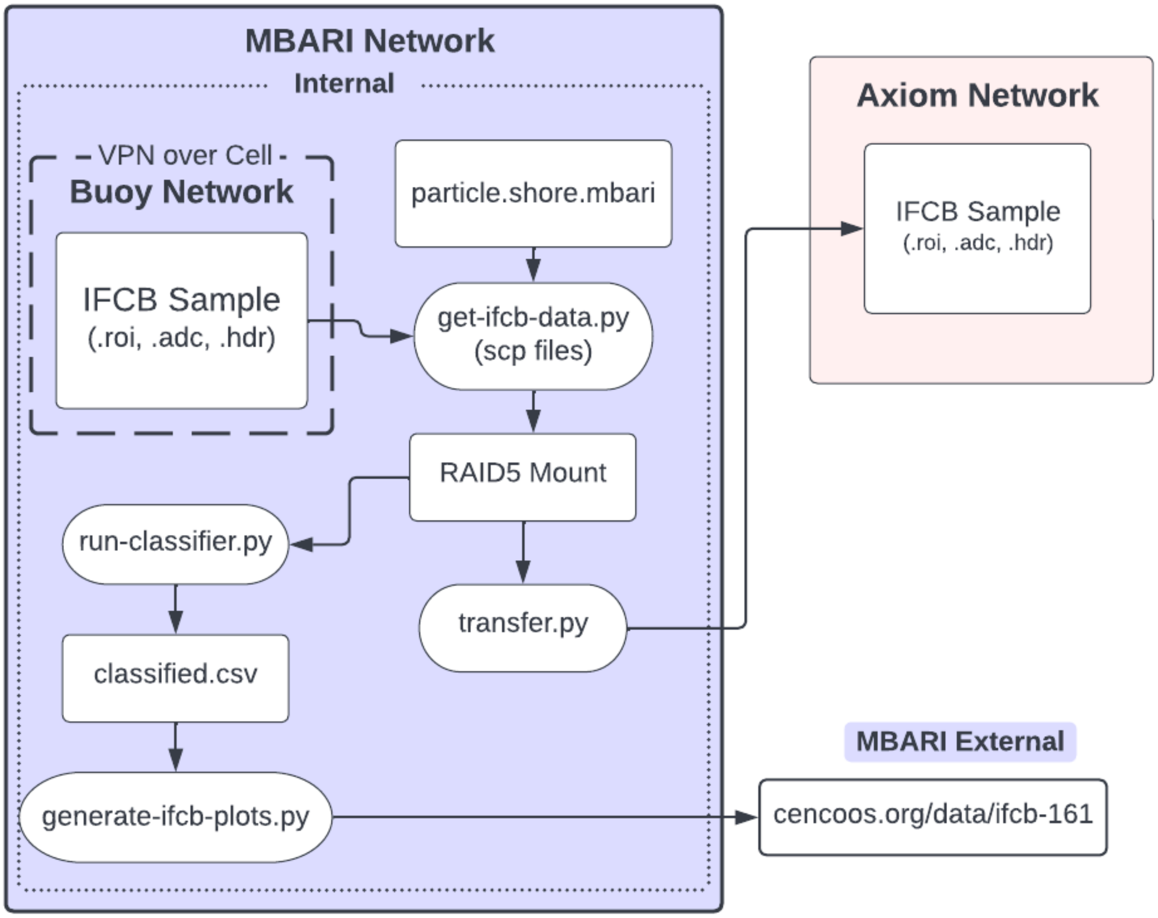


Figure 6, Wire Diagram of data flow from instrument to Axiom Dashboard as of 2022-12-18. Purple shading indicates processes that occur within the MBARI internal nextwork.

Are there ethical restrictions to data sharing?
No

Who holds intellectual property rights (IPR) to the data?
Monterey Bay Aquarium Research Institute (MBARI) and CeNCOOS

Describe any effect of IPR on data access.
None

Datas Source and Quality Control

Indicate the data source type (i.e. Federal, Non-Federal, University, Stage Agency, Local Municipality, Military Establishment (branch), private industry, NGO, non-Profit, Citizen Science, Private individual)
non-Profit

Indicate the data reporting type (e.g. real-time, historical)
Real-time

If real-time, list the QARTOD procedures that are currently applied.
Data currently falls outside of the scope of QARTOD. The data are a set of photos and are inspected by the IFCB operators on an ad hoc basis. The ML model derived identifications are addressed in the Model Card reporting.

What is the status of the reported data? (e.g. raw, some QC, incomplete, delayed mode processed but not QC'd)
Both raw data and ML modeled identifications are presented. QC of the model output is discussed in the model card.

Describe the data control procedures that were applied by the originator.
The originator monitors the output images to make sure that the images are clear and that the instrument is operating properly and behaving in the specified manner.

Describe the data control procedures that were applied by CeNCOOS.
Axiom, the CeNCOOS data partner, applies the ML algorithm to the data and generates the output subject to all the caveats described in the model card.

Provide a line to any documented procedures
https://www.cencoos.org/data/ifcb-doc-proto/index.html

List the procedures taken for data that could not be QC'd as directed.
N/A

Stewardship and Preservation Policies
Who is responsible for long-term data archiving?
Long-term archiving is currently the responsibility of the operating institution with Axiom Data Science as a backup. Future plans are to use Ocean Vision AI with NCEI for archive of images.

Which long-term data storage facility will be used for preservation?
Near real-time images are planned to be stored through Ocean Vision AI through the NCEI connection.

Describe any transformations necessary for data preservation.
Data will be bundled into tar files for transfer.

List the metadata or other documentation that will be archived with the data.
N/A

Additional Images and Figures

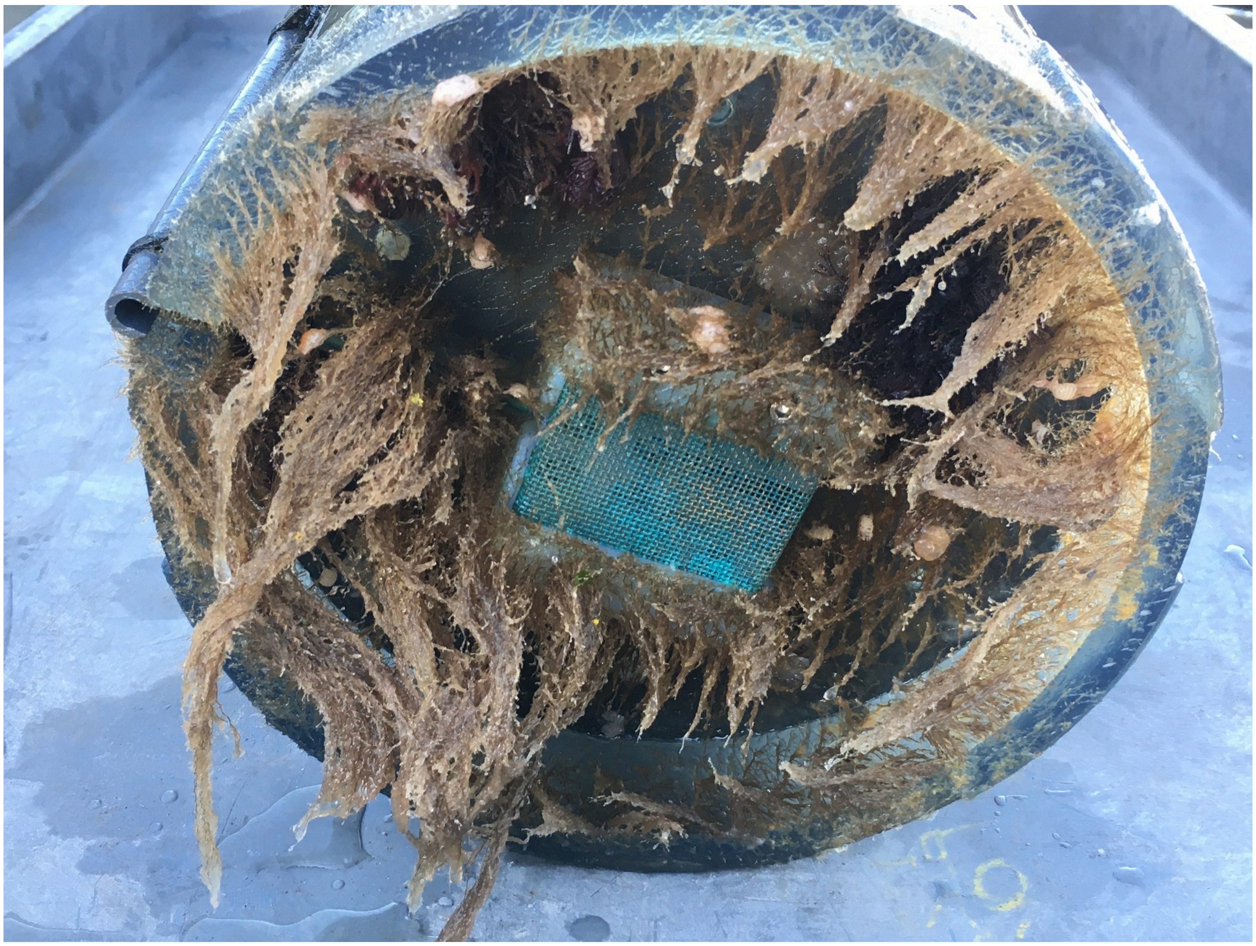


Figure 7. Intake screen post recovery of the initial deployment. Corrosion lead to holes in the copper intake tubing. Minor corrosion occurred along the bottom plate, due in part to the sacrificial anode being located on the top of the instrument and not submerged.

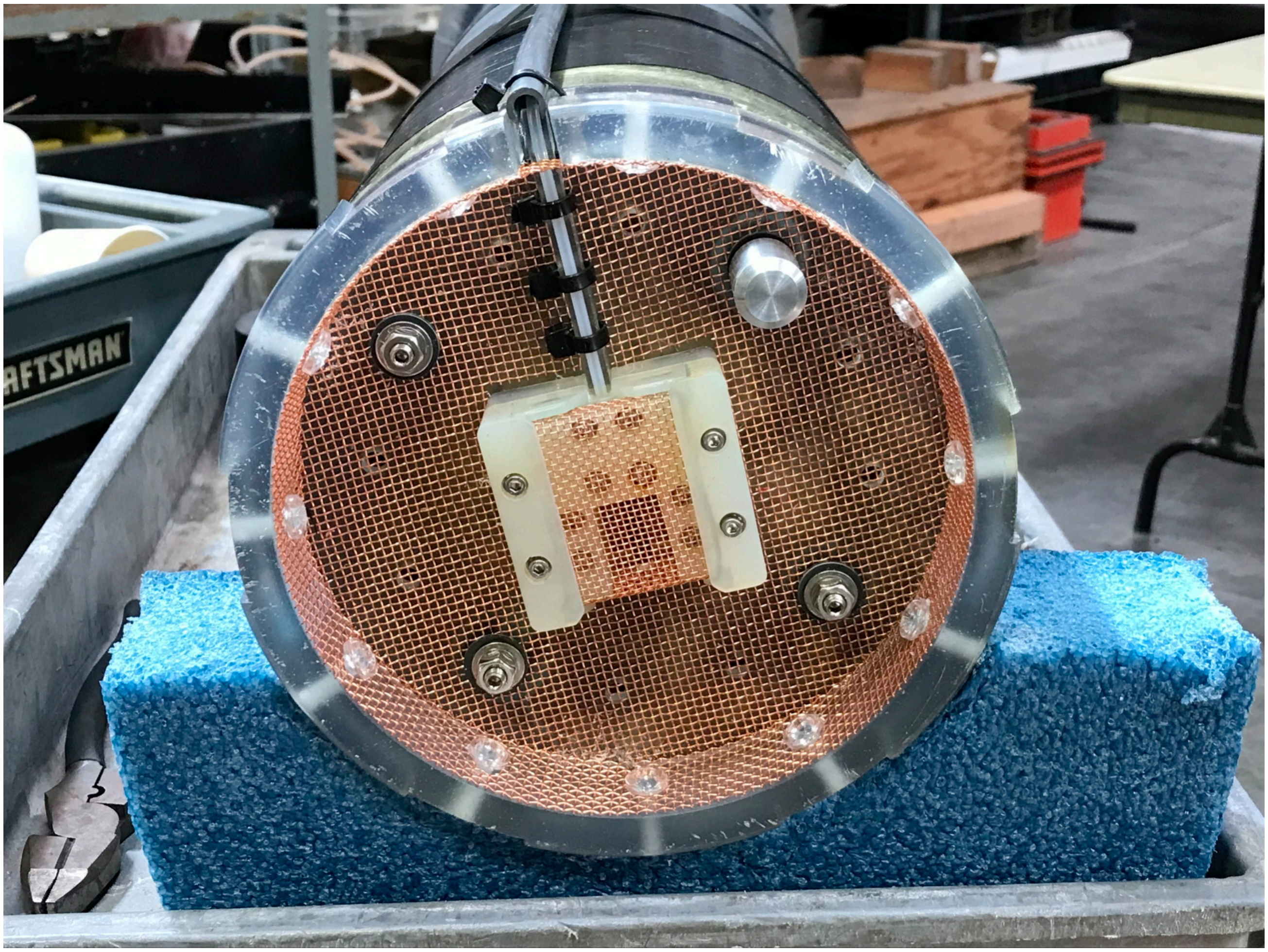


Figure 8. Intake screen preceding the third deployment. Additional copper screening was added around the bottom plate to prevent growth. A scafracfical zinc anode was attached to the bottom plate to minimize corrosiion on the housing.