Document author and contact info: Heidi Sosik (WHOI), hsosik@whoi.edu

Description of the instrument: Imaging of phytoplankton and other particles with Imaging FlowCytobot (IFCB; McLane Research Laboratories, Inc, Falmouth, MA). The IFCB is an imaging-in-flow cytometer. As such, it measures not only individual particle fluorescence and light scattering, but also captures a high resolution (~1 μm) image of each cell or chain in the size range ~5-150 μm width. Controlled flow and illumination conditions ensure a very high rate of images containing in focus, single targets aligned in the flow such that the largest cross-section is imaged. Images can be collected at up to ~15 Hz, depending on particle concentrations encountered. Images have a resolution of 2.77 pixels per micrometer.

Instrument calibration and maintenance: Main calibration issues are (1) ensuring sample volume is properly quantified (a function design criteria set during manufacture; user verification is good practice, but experience suggests this does not need to be repeated unless there are hardware changes in the instrument); and (2) determination of image scaling (micrometers per pixel; user determined with particles of interest).

Instrument settings that affect types and sizes of particles imaged: Images in this dataset were triggered by chlorophyll fluorescence, thus mainly representing phytoplankton but include herbivorous microzooplankton. IFCB trigger thresholds were set to image as wide a size range as possible, i.e., 5 to 150 micrometers, with quantitative observations in a narrower range. Images have resolution 2.77 pixels per micrometer.

Sample collection method: The cruise was on R/V Endeavor and identified as EN608, with the following Digital Object Identifier (DOI) https://doi.org/10.7284/908133. The IFCB was operated with chlorophyll fluorescence and scattering triggers enabled and it was configured to automatically sample 5 ml in 20-minutes intervals from the uncontaminated seawater flow (diaphragm pump source, pre-debubbler to ensure minimal damage to cells). IFCB was also used to analyze discrete samples from Niskin bottles (some with chlorophyll fluorescence triggering only). The samples were pre-filtered with a 150 μm nitex mesh to prevent system clogs.

Determining volume imaged per sample: The IFCB draws in 5 ml per sample but does not image the entire volume. The volume imaged per sample is provided in the online IFCB dashboard under Basic Info as Volume Analyzed.

Image processing method and version: Full resolution images are stored, though only the portion of the camera field that contains the target of interest (real time segmentation is done during acquisition). Images were processed with software for segmentation and feature extraction to determine size parameters per ROI (IFCB Features Version 4). Results from image processing are provided in the IFCB dashboard per sample as a features_v4.csv file. We selected a subset of 4 features to provide per ROI (Area, Biovolume, maxFeretDiameter, minFeretDiameter). The biovolume calculation method is described by Moberg and Sosik (2012) https://doi.org/10.4319/lom.2012.10.278. Image processing yielded no features for only 7 of the 144,281 ROIs in this dataset. For direct access to images replace the .html extension in the Level 1b associatedMedia with .jpg or .png. Processing code and wiki-based documentation is available at: https://github.com/WHOIGit/ifcb_classifier.

Methods for automated and/or manual classification and taxonomic assignment:

Identifications to morphological categories were done manually using annotation software with a database that also records the annotators and the number of times an annotation has been verified. We queried the database to export manual annotations for the geographic subset of IFCB102 samples and then further divided into subsets that include only those samples for which every ROI in the sample was verified by a

high-power annotator (to increase certainty in the manual identifications). In this version of this data product, each ROI corresponds to an occurrence of a single taxon (in future versions we may account for categories or tags for a small number of ROIs that represent multiple taxa). Most of the morphological category names could be resolved to accepted taxonomic names and machine-readable identifiers in the World Register of Marine Species (WoRMS). The level of taxonomic identification varies, but some distinctive taxa can be identified to species level. Some morphological categories could only be matched to WoRMS at a higher taxonomic level, for example mix_elongated is a morphological category of diatoms. Some morphological categories could not be matched to WoRMS but were matched to Eukaryota in AlgaeBase. We also resolved taxonomic names and identifiers for the full list of categories in the annotation database, to be able to indicate absence for those categories not observed in samples from this cruise (i.e., occurrenceStatus absent in a future version of this dataset). Several categories are not organisms thus taxonomic names are NotApplicable (e.g., bubble, detritus). Code and classification lookup tables are available in GitHub: https://github.com/klqi/EDI-NES-LTER-2019/tree/master/namespace_validation.

Summarization per sample: Summarization per sample is possible because we are only providing data for samples for which every ROI had a manual annotation. Code for the summarization from the Level 1b to the Level 2 data table is available in GitHub: https://github.com/klqi/EDI-NES-LTER-2019/blob/master/plot1. Concentrations per taxon per sample may be calculated by dividing the abundance or biovolume by the volume imaged. Abundance does not correspond necessarily to cell counts because chain- or colony-forming organisms may be imaged as a single ROI. Note concentrations will be underestimates for taxa that do not always trigger fluorescence.

Additional data cleaning and quality assurance: Additional data cleaning and metadata template assembly were performed with code available on GitHub: https://github.com/WHOIGit/nes-lter-ifcb-transect-winter-2018. We renamed or added attributes to enable harvesting of the level 1b data table as an occurrence table for the Ocean Biodiversity Information System (OBIS, e.g., occurrenceID, eventDate, decimalLongitude, decimalLatitude, occurrenceStatus, basisOfRecord). We assured that the geographic and temporal coverage and values for attributes were within expected ranges.

Key method references:

Olson, R. J., and H. M. Sosik. 2007. A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot. Limnol. Oceanogr. Methods 5: 195-203.

Sosik, H. M., and R. J. Olson. 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5: 204-216.

Sosik, H. M., J. Futrelle, E. F. Brownlee, E. Peacock, T. Crockford, and R. J. Olson. 2016. hsosik/ifcb-analysis: IFCB-Analysis software system, initial formal release at v2 feature stage [Data set]. Zenodo. http://doi.org/10.5281/zenodo.153978

Sosik, H.M., E. Peacock, and M. Santos. (2020). Abundance and biovolume of taxonomically-resolved phytoplankton and microzooplankton imaged continuously underway with an Imaging FlowCytobot along the NES-LTER Transect in winter 2018 Ver1. Environmental Data Initiative. https://doi.org/10.6073/pasta/74775c4af51c237f2a20e4a8c011bc53

Peacock, E.E., E. T. Crockford, and H.M. Sosik. 2018. IFCB at sea user guide. https://docs.google.com/document/d/14IfQBriV2AZs1akefM8JYirSAApnVFbDG2XQ74klIOI/