

PyOPIA: A *Python Ocean Particle Image Analysis* toolbox

Emlyn Davies^a, Alex Nimmo-Smith^b, Raymond Nepstad^a, Tor Nordam^a, Ute Brönnert^a, Andreas Steinvik^a, Sari Giering^c, Mojtaba Masoudi^c, Zonghua Liu^c, Pierre Hélaouët^d, Kairan Cursons^{b,d}, Matthew Rau^e, Yixuan Song^f, Arsalan Mostaani^g, Karoline Barstein^g, Daniel Buscombe^h

^aSINTEF Ocean, Trondheim, Norway

^bUniversity of Plymouth, Plymouth, UK

^cNational Oceanography Centre, Southampton, UK

^dMarine Biological Association, Plymouth, UK

^eThe George Washington University, Washington, D.C., USA

^fUniversity of Rhode Island, Narragansett, RI, USA

^gNTNU, Trondheim, Norway

^hMarda Science LLC, Flagstaff, AZ, USA

Corresponding author: Emlyn Davies; emlyn.davies@sintef.no

Introduction

Research questions concerning particles suspended in the ocean present a constant challenge: what are they, how big are they and how do they move? These questions are difficult to address due to the intricate and diverse structures of these particles which span a wide range of sizes, concentrations, and movement characteristics. Whilst acoustic and optical scattering sensors provide some insight into the distributions of size and relative abundance, in-situ imaging systems can directly measure key characteristics of individual particles within their natural environment with minimal disruption and disturbance.

In recent years, an expanding array of in-situ particle imaging systems have been developed and routinely deployed within the ocean. These systems employ different methods to record focussed images with varying resolution within different-sized sample volumes, with advantages and disadvantages associated with each. Methods include macro-lens imaging of light-sheet illuminated sections (e.g. The Underwater Vision Profiler [Picheral et al., 2022]), back-illuminated telecentric imaging (e.g. SINTEF SilCam [Davies et al., 2017]), in-line digital holography (e.g. Sequoia LISST-HOLO2 derived from [Graham & Nimmo-Smith, 2010]), water pump onboard imaging (e.g. Plankton Imager [Culverhouse et al. 2015]), lab-based flow imaging microscopy (e.g. FlowCam [Jakobsen et al, 2011]) and its simplified version - PlanktoScope [<https://www.planktoscope.org/>]), in-situ underwater imaging flow cytometer (e.g. CytoSub [<https://www.cytobuoy.com/products/cytosub/>]). Whilst all these systems record digital images, they are in different formats (colours or grayscale; resolution; noise types/sources) that are then subject to different processing algorithms by different (sometimes proprietary, often closed-source) software prior to the extraction of the actual particle characteristics that may then be

further aggregated and analysed, and some specific software can only run on a certain operating systems or in-cloud. In addition to the differences in hardware configuration, this diversity within the processing software can lead to challenges in combining the outputs from different systems to better address the overarching research question. These range from simple differences in output data format and how particle data are presented, to fundamental discrepancies in how accurately complex particles are segmented and how specific particle metrics (such as size) are calculated. Furthermore, the exact image processing steps used by different software are not always well documented and logged as metadata, making it often challenging to reproduce results without significant user expertise. These differences make the incorporation of data from the disparate instruments difficult or impossible when for example applying automated classification to their outputs. Emerging challenges in the modern age of robotic adaptive sampling also require particle image analysis software capable of rapid integration with robotics control system software and needs the ability of running in real-time on offline (or edge) computers.

The challenges mentioned here culminate to an ever-growing volume and range of particle imaging datasets from around the world that are underutilised, both within the particle imaging community, and by other researchers and users wishing to make use of data produced. This underutilisation is due to one primary gap: the particle imaging community lacks common, open-source, analysis workflows that produce an interoperable data format. The acoustics community has experienced the same challenge, and promising initiatives such as EcoPype (<https://echopype.readthedocs.io>) [Lee et al., 2021] are demonstrating how such challenges can be overcome.

Therefore, we have begun development of PyOPIA, a Python Ocean Particle Image Analysis (<https://github.com/SINTEF/PyOPIA>) toolbox, with the aim of providing the ocean research community with an open software toolbox for the analysis of particle images from diverse hardware systems. The design goals for PyOPIA are to:

1. be hardware agnostic (both from image acquisition and processing perspectives)
2. use free and open-source software to ensure accessibility and transparency
3. be robust, testable and clearly documented
4. be flexible and reconfigurable
5. generate interoperable data that are compatible with FAIR data principles
6. be usable and trusted by non-experts, with opportunity to contribute and commit code changes
7. support collaborative development by leading international experts
8. be capable of running offline and on edge hardware, with compatibility for Windows, Linux, macOS

Methods

PyOPIA aims to provide a pipeline-based workflow as a standard for analysis of images of particles in the ocean. This pipeline should be consistent across different instruments (hardware), and therefore has flexibility to adapt analysis steps to meet instrument-specific processing needs (e.g. holographic reconstruction), while maintaining a traceable workflow that is attached as metadata to a standard open-format output file that helps users follow FAIR data principles. While the initial development of PyOPIA has targeted in-situ instruments with open-path imaging, the aim is for this toolset to be widely applicable to in-situ and lab-based (e.g. PlanktoScope) imagers.

The most basic pipeline consists of steps that standardise tasks including: background correction, segmentation, calculating of particle geometries, output file formats, and a framework within which a particle classification algorithm can label particle types (e.g. using tensorflow). User setup of processing pipelines is also flexible, such that new steps can be added as required by specific hardware or uses (e.g. holographic reconstruction requires setup of the reconstruction kernel, and focussing steps). In addition to the processing pipeline philosophy, PyOPIA aims to provide a toolbox of high-level functions that can be used to perform specific analysis or visualisation tasks. Such a toolbox enables rapid prototyping and adaptation of data analysis tasks, while maintaining the ability to produce a consistent output data format.

With the repository hosted on GitHub, users are able to contribute their additions in the form of new help functions, or revisions to instrument- or use-specific pipelines. Changes to the source code are automatically tested, using GitHub Actions, with a set of fixed reference datasets. Such testing ensures that code changes do not introduce unwanted outputs as enhancements are made. In-line documentation also allows updates to PyOPIA to be rendered automatically at <https://pyopia.readthedocs.io>.

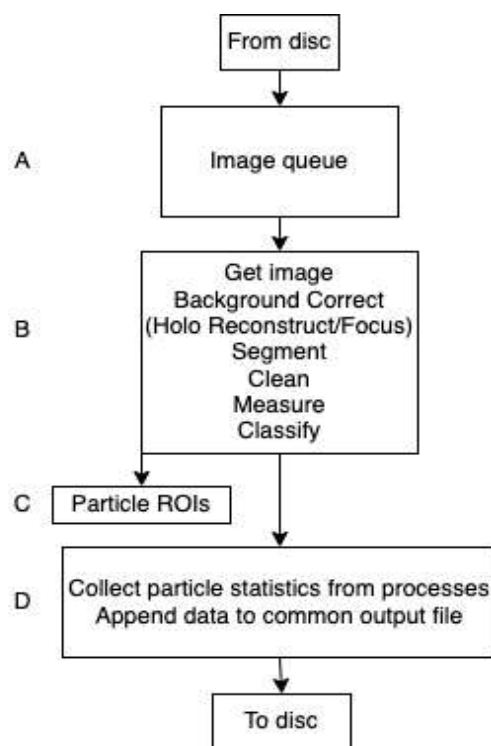


Figure 1: Schematic of the basic PyOPIA pipeline that after initialisation draws images sequentially from the image queue (A), applies background correction and particle processing steps (B), outputs particle ROIs (C) and then collates particle statistics (D) for saving and further analysis.

Applications and examples

Measuring bubbles under breaking waves:

In this case, the goal of the experiments was to measure bubble size distributions, as a function of depth and time, under breaking waves. The PyOPIA workflow here (Figure 2), takes a raw image, applies a background correction, segmentation (producing the binary image), and then regions of interest are extracted for each connected object in the binary image using `skimage.measure.regionprops` (<https://scikit-image.org/>).

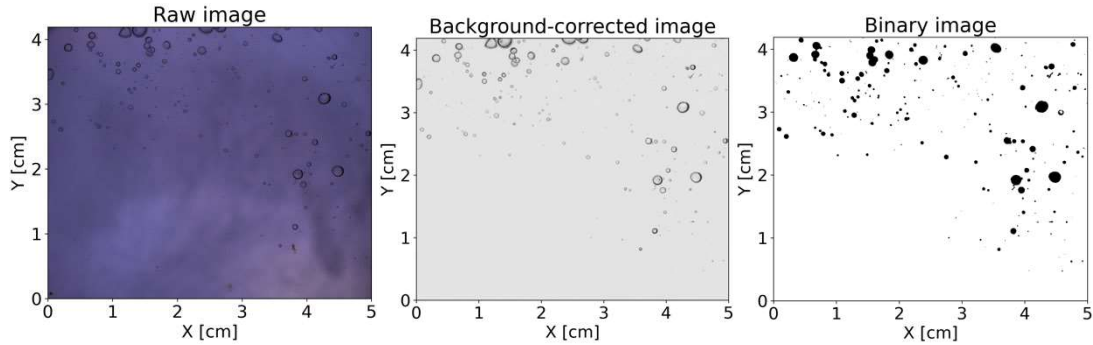


Figure 2: Processing steps taking raw images (left), correcting to their background (middle), and segmenting them (right).

Given that the pixel size is known (in this case $20\ \mu\text{m}$), the size of the detected objects can be found. The example in Figure 3 shows all detected particles with an equivalent diameter larger than $100\ \mu\text{m}$ outlined in red, with a fitted ellipse. Following this, bubbles size measurements provided data on the time evolution of the particle counts, d50, and volume concentration during breaking wave experiments.

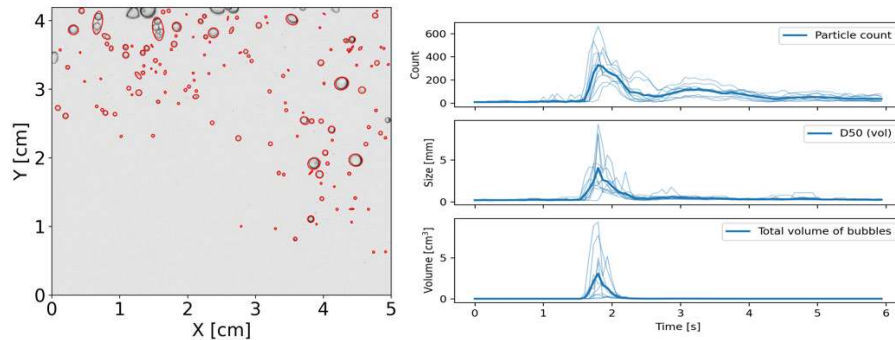


Figure 3: Fitted ellipses around bubbles following a basic PyOPIA analysis (left), and the time-series evolution of particle statistics under breaking waves derived from these data (right).

Generalised particle measurements, classification, and visualisation capabilities:

In executing a standard pipeline (in this case for SilCam analysis), PyOPIA is able to take a folder of raw images from the instrument, and produce particle data that can be utilised by the available plotting and classification functions to produce a summary of a dataset (in this case, following removal of bubbles near the surface), as the one shown in Figure 4.

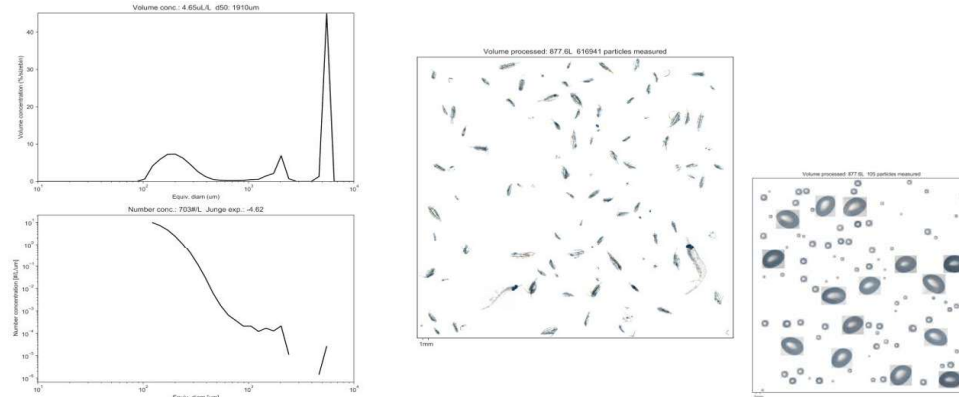


Figure 4. Example summary of particle images and statistics from a SilCam during profiling in Billefjorden, Svalbard, Jan 2020. The volume distribution and number distribution of the measured particles after removing detected bubbles, is shown in the upper and lower left panels, respectively. The montage of particle images shown in the middle panel is auto-generated from material larger than 500 μm , using a packaging algorithm in the 'pyopia.statistics' module that attempts to represent the size distribution of particles to correspond with what was measured, but in doing so, concentration (or separation between particles) is not represented. The small montage to the right shows detected bubbles that were removed from the statistics presented.

Hardware-agnostic analysis:

Using the same analytical approach on multiple hardware configurations allows side-by-side comparisons, and eventual merging, of multiple instruments' measurements. In the case shown in Figure 5, below, both a SINTEF SilCam and UVP-6 were deployed alongside each other on a profiling frame in June 2022 on the Norwegian shelf. A comparison of both instruments' vertical particle volume distributions is demonstrated, together with depth-integrated montages that visualise the particle populations imaged by the two instruments. It is also possible to pinpoint locations of each unique particle, where, for example the two copepods expanded are from a denser region of zooplankton in the surface region indicated by the blue ovals.

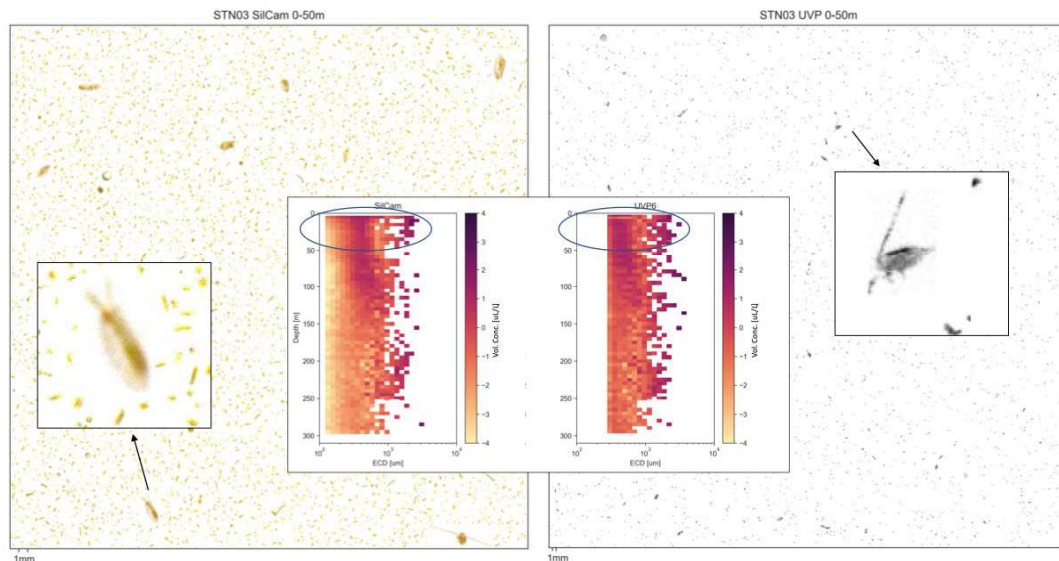


Figure 5: Example depth-integrated montages and particle volume distribution profiles from SilCam and UVP-6 (inverted intensities) using PyOPIA analysis functions. ECD is Equivalent Circular Diameter.

Holographic imaging reconstruction and analysis – LISST-HOLO2:

The example in Figure 6 shows a particle montage generated from the PyOPIA toolbox. As set out in Figure 1B, additional processing steps are activated to reconstruct, clean (de-noise) and then localise (focus) the particles within the 3D sample volume using functions within the *pyopia.instrument.holo* subpackage. Additional instrument-specific parameters for the LISST-HOLO2, such as laser wavelength and sample-volume dimensions are used during the pipeline initialisation to generate the required reconstruction kernel. All identified particles are displayed after performing segmentation, from which statistics are generated as above.

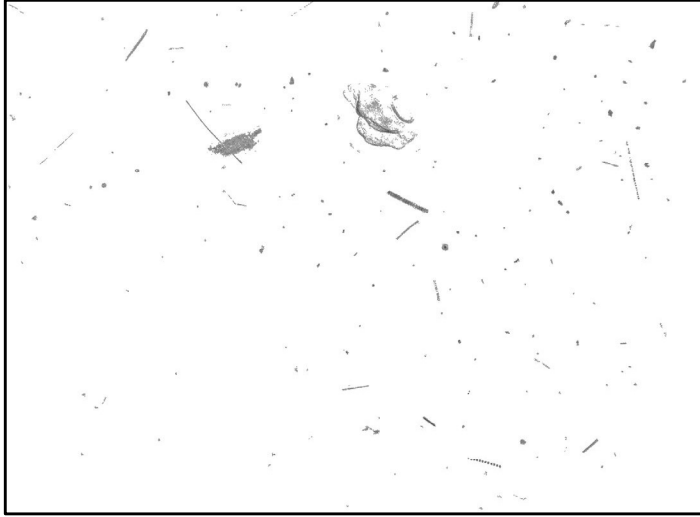


Figure 6: Example PyOPIA-generated montage from a LISST-HOLO2

Holographic imaging reconstruction and analysis – iCPRholo:

As a further example of the application of PyOPIA to holographic images, Figure 7 shows a particle montage and volume distribution collected using an iCPRholo – a bespoke in-line holographic camera with different resolution and sample volume size to the LISST-HOLO –

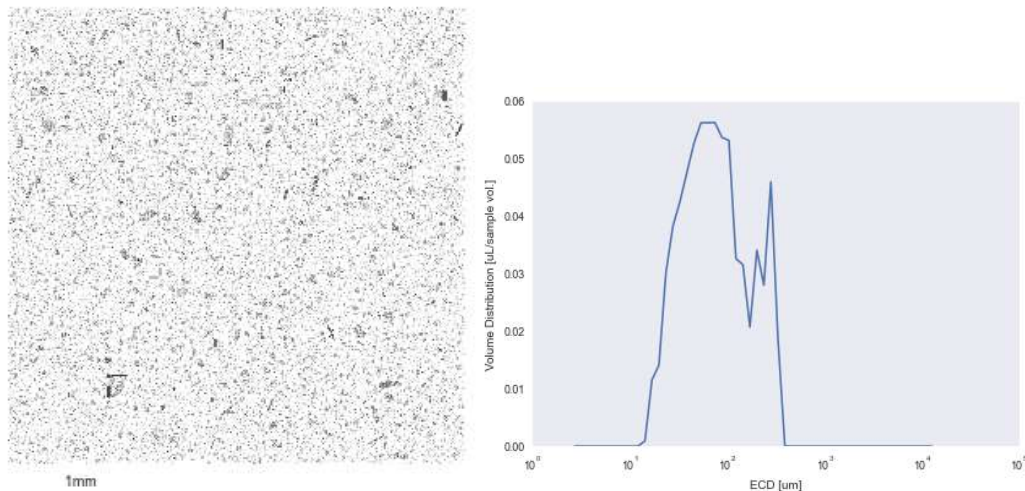


Figure 7: Example PyOPIA-generated montage and volume distribution of all particles (23,435) within a 1 nautical mile segment of an 85 nautical mile long iCPRholo tow across the Irish Sea, November 2022.

mounted on a Continuous Plankton Recorder (CPR) tow-body on a routine 85 nautical mile tow across the Irish Sea in November 2022. Here, all particles from a 1 nautical-mile segment are shown, with a prevalence of small ($\sim 90\mu\text{m}$ equivalent circular diameter, ECD) flocs.

Flexibility in operational data pipelines:

Edge computing is becoming a vital component in operational monitoring and adaptive sampling [Fossum et al., 2019]. In the context of operational ocean monitoring, edge computing would mean the data processing (classification of particles into different types) could happen onboard a glider or AUV. This both reduces the amount of data that needs to be transmitted back to the surface, and allows for real-time or near real-time processing of data. The real-time results (as outlined in Figure 8) can directly be reused for either data-driven sampling based on previous results ('follow the particles') or in an immediate and targeted response to the particles to minimise their impact on the environment, should they be pollutants such as oil, gas, or plastic.

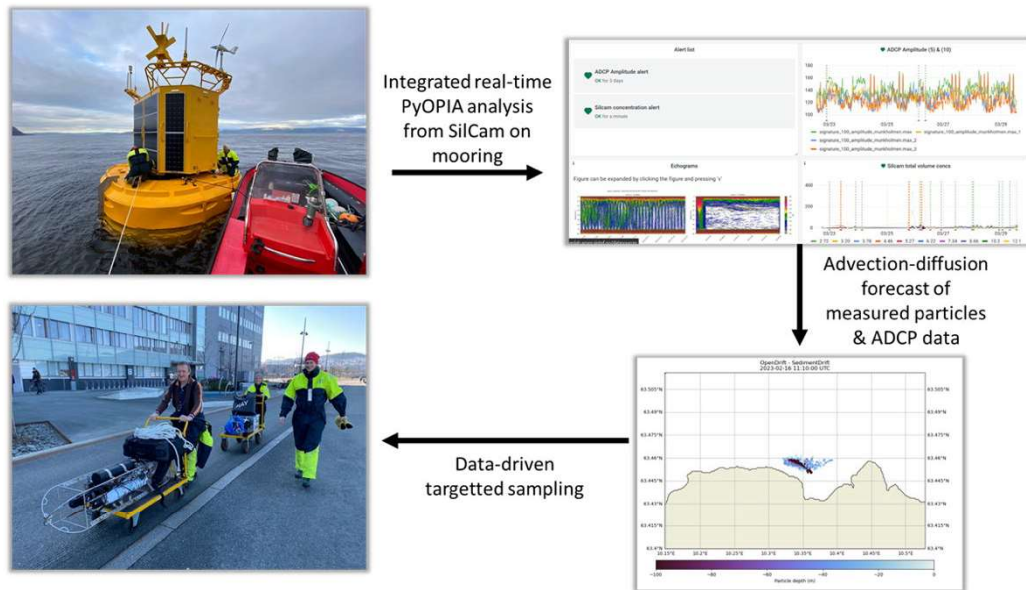


Figure 8: Example operational use of PyOPIA for data-driven adaptive sampling of zooplankton in OceanLab, Trondheim.

With the advent of digital twins of the ocean, that aim to create a digital replica of the physical ocean which can be used to model, analyse, and predict the behaviour in an ocean area, operational modelling with flexible data pipelines (Figure 8) can become an integral part of the twin for observations, model validation as well as event detection and responses [Br  nner et al., 2023].

Outlook

PyOPIA represents the start of a standardised open framework for particle image processing in the ocean community. For development to continue, wider community adoption, feedback and input is needed. We encourage researchers in the community to start using this toolbox to

analyse their own unique research images and to contribute refinements, new code developments, and new pipeline suggestions as they see fit to reach broad applicability. In particular, refinements of processing methods for new instruments and challenging imaging cases (e.g. high particle densities, particle aggregates, highly transparent particles, etc.) are encouraged. It is our aim to implement these varied processing options as different pipeline steps or user-defined input choices to maximise functionality and flexibility while retaining standardised and documented data processing standards.

References

Brönnert U, Sonnewald M, Visbeck M: Digital Twins of the Ocean can foster a sustainable blue economy in a protected marine environment. *International Hydrographic Review* (29(1)), 26-40 (2023). <https://doi.org/10.58440/ihr-29-a04>

Culverhouse, P. F., et al. "An instrument for rapid mesozooplankton monitoring at ocean basin scale." (2015).

Davies EJ, P.J. Brandvik, F. Leirvik, R. Nepstad, The use of wide-band transmittance imaging to size and classify suspended particulate matter in seawater, In *Marine Pollution Bulletin*, Volume 115, Issues 1–2, 2017, Pages 105-114, ISSN 0025-326X, doi:10.1016/j.marpolbul.2016.11.063

Fossum TO et al., “Toward adaptive robotic sampling of phytoplankton in the coastal ocean,” *Science Robotics*, vol. i, pp. 1–12, 2019. DOI:[10.1126/scirobotics.aav3041](https://doi.org/10.1126/scirobotics.aav3041)

Graham GW & Nimmo Smith WAM (2010) 'The application of holography to the analysis of size and settling velocity of suspended cohesive sediments' *Limnology and Oceanography: Methods* 8, (JAN) 1-15 , DOI: 10.4319/lom.2010.8.1

Jakobsen, Hans Henrik and Jacob Carstensen. “FlowCAM: Sizing cells and understanding the impact of size distributions on biovolume of planktonic community structure.” *Aquatic Microbial Ecology* 65 (2011): 75-87.

Lee, W., Mayorga, E., Setiawan, L., Majeed, I., Nguyen, K., & Staneva, V. (2021). Echotype: A Python library for interoperable and scalable processing of water column sonar data for biological information. *arXiv preprint arXiv:2111.00187*

Picheral M, Catalano C, Brousseau D, Claustre H, Coppola L, Leymarie E, et al. The Underwater Vision Profiler 6: an imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms. *Limnology and Oceanography: Methods* 2022;20(2):115–129.