APPLICATIONS OF THE UNDERWATER VISION PROFILER FOR PARTICLE ANNOTATION IN THE OLIGOTROPHIC NORTH PACIFIC SUBTROPICAL GYRE

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

OCEANOGRAPHY

2024

AUTHORED BY:

REECE JAMES

THESIS COMMITTEE:

ANGELICQUE WHITE (CHAIRPERSON),

ERICA GOETZE, JIM POTEMRA

KEYWORDS: UNDERWATER VISION PROFILER, MACHINE LEARNING, ZOOPLANKTON, MARINE AGGREGATES, OPTICAL OCEANOGRAPHY

ACKNOWLEDGMENTS

The following thesis would not be possible if not for the support, advice, and gracious patience of those in the White and Goetze Labs at the University of Hawai‘i at Mānoa.

Special thanks are extended to Dr. Angelicque White, from whom many moments of scientific and personal counsel were called upon and answered with grace and without hesitation, and with whom many discussions were shared on the rich history of oceanography, and the future of the field. I would also like to thank Dr. Erica Goetze for sharing her wisdom on the topic of zooplankton identification, thesis writing, and seeing the forest for the trees. Further, thanks are owed to Jim Potemra for their guidance in communicating the science herein.

It would not be possible for any images to be annotated, nor particle data assessed, if not for the efforts of Robert “Tully” Rohrer, who navigated the arduous setup, configuration, and operation of the Underwater Vision Profiler aboard the Hawai‘i Ocean Time Series cruises and many other cruises thereafter.

For kindly listening to monologues on the topic of morphological taxonomy, and for the invaluable advice regarding optics, programming, and ecology, I have Fernanda Henderikx Freitas, James Ash, and Andrew Hirzel to thank.

Finally, I would like to acknowledge the many hours of image annotation and unparalleled endurance of Annotator 2: Natalie Summers.

ABSTRACT

Machine learning algorithms (MLAs) are increasingly applied to optical imaging datasets of oceanic plankton and marine aggregates to obtain improved image annotation efficiency while preserving high annotation consistency. However, this process relies on an ever-decreasing number of expert morphological taxonomists to annotate training sets and validate MLA outputs. While recent attention has focused on training annotators on how to *use* machine learning algorithms, there has been limited effort to educate new annotators on how to *annotate* the datasets needed to train and verify such algorithms. By teaching new annotators how to create regionally-relevant and accurate training sets for MLAs, one could better utilize instruments, such as the Underwater Vision Profiler 5 HD (UVP), that are the basis for growing databases of images collected over an expanding set of temporal and regional studies. The UVP is a high resolution *in situ* camera-based instrument that samples particles above 0.064 mm up to ~54 mm, producing images for those particles that are > 0.5 mm. The UVP images a size fraction of fragile plankton and marine aggregates known to play an important role in the Biological Carbon Pump (BCP) and can quantify their vertical distribution, changing morphological characteristics, and visual interactions from the sea surface to ~6000 db. In this study, the UVP has been used to assess particle distributions on 14 Hawai'i Ocean Time-series (HOT) cruises between 2020 to 2023 at Station ALOHA (22.75°N, 158.00 °W). Significant findings from this initial effort include - (1) seasonal changes in the slope of particle size distributions evidence summer (June – August) increases in the abundance of large particles, (2) subsurface peaks of large particles were frequently observed at the base of the euphotic zone between 100 to 150 db which we interpret to be the accumulation of sinking particles along isopycnals, and (3) in moving towards assessing organismal abundance, it became apparent that an annotation guide for the UVP was not available to the user community. To facilitate further research, we identified key classifiers for our region including 13 categories of organismal and detrital UVP images, including the genus *Trichodesmium*, Rhizaria, and marine aggregates. We then outlined a method and standards for development of a cooperatively annotated dataset with intra-annotator self-consistency. Individual annotations were made by two annotators and then compared to a cooperatively annotated dataset, displaying 87.6% and 88.1% agreement. Comparing the annotations made between annotator's individual datasets, the agreement was 85.2%. In comparison, predictions by a machine learning algorithm tailored to the UVP, the EcoTaxa random forest, had only 31% precision. With this manually annotated dataset, the temporal and spatial patterns of aggregates and organisms were then assessed. One pronounced pattern observed was that marine aggregates were found in concentrations more than double

that of organismal categories with peaks in concentration at the mixed layer boundary. Further research will investigate co-occurrence patterns and potential relationships with regional hydrodynamics and climate indices as the time-series of UVP data lengthens. Importantly, the standardized UVP annotation schema developed herein will allow increasingly large optical datasets collected by the HOT program to be annotated by multiple annotators. This will reduce the overall time spent manually annotating images and facilitate consistency across datasets. The classification guide and annotation pipeline described here maximizes the potential research questions that can be addressed with the large datasets generated by the UVP and outlines a path for new users in different regions to generate their own classification guides and annotation pipelines.

TABLE OF CONTENTS

ACKNOWLEDGMENTS 1

ABSTRACT 2

LIST OF TABLES 6

LIST OF FIGURES 6

1. INTRODUCTION 7

1.1 The Biological Carbon Pump 7
1.2 Methods of Measuring Particles 8
1.3 The Underwater Vision Profiler 10
1.4 Image Annotation and the Rise of Machine Learning 11
1.5 This Study 12

2. METHODS 13

2.1 Underwater Vision Profiler Instrument 13
2.2 Deployment 14
2.3 Particle Data Analysis 15
2.4 Building a Training Set 16
2.5 Classification Standards 18
2.5.1 Taxonomic Specificity 20
2.5.2 Reference Images 21
2.5.3 Labels 21
2.5.4 Descriptions 22
2.6 Process Study 26
2.6.1 Data Collection 27
2.6.2 Annotation Overview 27
2.6.3 Image Annotation Application 28
2.6.4 Image Status 29

3. RESULTS AND DISCUSSION 31

3.1 Large Particle Abundance, Distribution, and Seasonality 31
3.2 Image classification – Random Forest Performance 36
3.3 Process Study Efficacy 37
3.4 Dataset Annotation Comparisons 37
3.4.1 Annotation Counts 37
3.4.2 Annotation Size Thresholds 39
3.4.3 Comparing Annotator Classifications 40
3.5 Community Composition 45
3.6 Individual Object Insights 48

4. CONCLUSION 56

5. LITERATURE CITED..... 59

LIST OF TABLES

Table 1. Precision of EcoTaxa Random Forest Algorithm 36
Table 2. Subset Annotation Counts 38
Table 3. Subset Annotation Agreement 39

LIST OF FIGURES

Figure 1. UVP diagram 14
Figure 2. UVP image layout 19
Figure 3. Descriptor visualizations 25
Figure 4. Classification Standards Guide Example 26
Figure 5. Particle Abundance with Seasonality 32
Figure 6. HOT 345 Particle Size Distribution 33
Figure 7. Vertical Profile of Particle Size Distribution Slopes 34
Figure 8. Seasonal Change in Particle Size Distribution 35
Figure 9. Confusion matrix 1 mm ESD 42
Figure 10. Confusion matrix 2 mm ESD 43
Figure 11. Confusion matrix 3 mm ESD 44
Figure 12. Abundance of Validated Objects 45
Figure 13. Percent Abundance of Validated Objects 46
Figure 14. Percent Abundance of Validated Objects: Marine Aggregates Excluded 47
Figure 15. Size Range of Validated Objects 48
Figure 16. Trichodesmium Concentration 50
Figure 17. Marine Aggregate Concentration 51
Figure 18. Copepod Concentration 52
Figure 19. Size Range of Marine Aggregates with Depth 53
Figure 20. Size Range of Copepods with Depth 54
Figure 21. Size Range of Rhizaria with Depth 55

1. INTRODUCTION

1.1 The Biological Carbon Pump

Central to the aims of modern biological oceanography are the ability to predict primary productivity, catalog organismal diversity, and quantify biological abundances. Each of these aims is inherently reliant upon a comprehensive understanding of the organisms that produce, transform, and export particulate organic carbon (POC) and particulate inorganic carbon (PIC). The processes by which these particles are produced and transformed is referred to as the biological carbon pump (BCP). It is through vertical settling, transport, and remineralization of particles that energy-rich matter reaches the deep ocean and bioelements are redistributed throughout the water column(Karl et al., 2012; Volk & Hoffert, 1985). Determining the size, concentration, and character of particles in the open ocean that are produced and transformed within the BCP is therefore a central aim of many oceanographic programs.

Efforts to quantify these metrics in the open ocean have revealed that subtropical gyres could account for ~50% of marine organic carbon export (Emerson, 1997). In these remote open ocean settings, light penetration is high, nutrients are low, and the growth rates of phytoplankton appear tightly coupled to loss rates (Letelier, 1996). As such, the export of carbon is relatively low when compared to coastal environments. Current models predict that these open ocean regions may warm and further stratify in response to the pressures of climate change (Capotondi, 2012). While this may seem to imply reductions in the rates of primary productivity, direct measurements made at Station ALOHA (22.75 ^{o} N, 158.00 ^{o} W) by the Hawai‘i Ocean Time-series (HOT) over the past 30 years in the North Pacific Subtropical Gyre (NPSG) indicate an increasing trend (Karl et al., 2021).

Direct measurements made by net tows and sediment traps in the upper water column from the HOT time series provide crucial insight into the variability of zooplankton biomass as well as the flux of carbon at the sea surface. However, these classical methods rely on physical collection, altering the plankton and marine aggregates with which they interact. As such, these classical methods are limited in their ability to catalog the complete taxonomic diversity of the zooplankton community and the character of marine aggregates that contribute to either active or passive particle flux. As a result, there may be undiscovered patterns of plankton succession, vertical distribution, and community composition in response to stochastic phytoplankton blooms, summer export pulses, and decadal trends.

Marine aggregates play a significant role in all stages of the BCP, enhancing production by phytoplankton and bacteria (Alldredge et al., 1986; Alldredge & Youngbluth, 1985; Andrews et al., 1984; Blackburn et al., 1998; Prezelin & Alldredge, 1983), scavenging and depositing suspended PIC and POC to the deep ocean (Jackson, 1995; McCave, 1984), and contributing the majority of vertical carbon flux (Asper, 1985; Fowler & Knauer, 1986; Honjo, 1980; Jackson & Burd, 1998; McCave, 1984), with variations in magnitude based on abundance, composition, and size. Marine aggregates are generated through a set of processes shown so far to include extracellular exudation (Alldredge et al., 1998; Passow et al., 2001), cell lysis (Blackburn et al., 1998; Passow et al., 2001), particle compaction and egestion (Alldredge, 1976; Prezelin & Alldredge, 1983), and accumulation through physical particle-particle interactions (Jackson, 1995; McCave, 1984). However, the proportionate effect of these mechanisms on particle transport within the BCP is not always similar, with increasing importance of the biological components in higher productivity regimes (Guidi et al., 2008; Sheldon et al., 1972; Stemmann et al., 2008).

Further, large aggregate particles are primarily biogenic (Asper, 1985; Honjo, 1980; Silver & Alldredge, 1981), and globally distributed, making up the bulk of particles observed in sediment traps used to measure passive carbon flux across ocean basins (Asper, 1985; Honjo et al., 1984; Silver & Alldredge, 1981). Herein, large particles are defined as those >64 μm, with 'marine aggregates' defined conventionally as detrital particles > 500 μm (Alldredge & Silver, 1988; Alldredge & Youngbluth, 1985; Fowler & Knauer, 1986; Silver et al., 1978). These large particles, whether derived from biological or physical processes, remove PIC and POC from the water column, sequestering both high quality labile organics for deep sea consumption, as well as inorganic carbons, influencing atmospheric CO_2 exchange (Fellows et al., 1981; Karl et al., 2012).

1.2 Methods of Measuring Particles

Just as Hensen attempted to quantify changes in the productivity of the ocean through his invention of the quantitative net tow (Dolan, 2021; Hensen, 1891), so too has the modern oceanographic community invented productivity-oriented methods of particle capture and remote detection. Such methods of capture, from net tows to sediment traps (Asper, 1987; Fellows et al., 1981), and visualization, from scuba photography (Alldredge & Cohen, 1987; Shanks & Trent, 1979; Silver et al., 1978) to imaging flow cytobots (IFCBs) (Sosik, 2007), have been historically paired with morphological analysis of those particles (Alldredge & Cohen,

1987; Honjo et al., 1984; Orenstein et al., 2015). This has often provided insight into the method of production for those particles, such as in the case of densely packed fecal pellets (Alldredge, 1976) or the taxonomic identity of collected organisms (Remsen et al., 2004).

Methods of physical collection boast the invaluable benefit of compositional analysis, whereby particles may be broken down into their chemical and biological components for further insight into their production, fates, and contribution to the standing stock of nutrients. Additionally, physical collection of organisms can yield more detailed morphotaxonomic information than may be attained through imaging alone. However, physical collection of organisms and aggregates brings with it unique limitations, namely the alteration of the samples that they gather (Asper, 1987; Gibbs & Konwar, 1983; Honjo et al., 1984; Silver & Alldredge, 1981). For instance, gelatinous organisms often become damaged or fragmented during the process of towing the net through large volumes of seawater. Further, marine aggregates are indelibly altered by physical collection methods as they are passed through screens, siphoned through tubing, and aggregated in sediment trap media (Asper, 1987; Gibbs & Konwar, 1983). Imaging methods have the considerable advantage that the organisms and aggregates they observe are minimally altered, and thus more easily and accurately identified without damage to taxonomically important structures.

Optical and imaging methods range both in the size of their target populations and their methods of analysis. While some optical systems such as Sequoia's Laser In-Situ Scattering Transmissometer (LISST) use near-forward scattering of a laser beam to gather particle counts in the ~1-500 μm range, imaging systems such as the video plankton recorder (VPR) can create dark field images of particles in the 30 μm–5 cm size range (Davis, 1992). These approaches therefore fulfill fundamentally different purposes in how they enumerate particles and identify plankton and marine aggregates. Further, *in situ* optical methods and imaging, which do not concentrate particles, face statistical limitations surrounding imaging volume and are limited in their conclusions pertaining to the concentration of large and/or rare particles (Forest et al., 2012).

Thus, to investigate planktonic and marine aggregate distributions and their role within the BCP, the observing method must be tailored to the scientific question at hand. Here, we aim to enumerate large particles including plankton and marine aggregates while preserving fragile structures for accurate morphological identification throughout the water column of the NPSG. Therefore, there is need for a large volume high resolution *in situ* imaging system capable of full ocean depth profiling.

1.3 The Underwater Vision Profiler

This study used the Underwater Vision Profiler 5 HD (UVP), developed by Hydroptic Inc., to meet the above-described needs and investigate the concentration and character of large particles in the NPSG from sea surface to seafloor. The UVP acquires images of particles between 0.064 mm and ~ 54.0 mm from ~1-liter samples. The fraction larger than 0.500 mm is also captured as a vignette by an internally housed digital camera capable of capturing high resolution images with distinguishable morphological characteristics (Picheral et al., 2010). Objects in the UVP's field of view are illuminated by dual 625 nm flashing lights and sampled from the sea surface to a maximum of 6,000 db at an adjustable frequency between 6-20 Hz while attached to either an optics array or CTD rosette, lowered at an average of 1 m/s to prevent overlapping images.

Almost 9,000 profiles conducted in all ocean basins have been collected with standard or high-definition versions of the UVP since the instrument's development in 2008 (Kiko et al., 2022). Over this period, studies have explored the particle size spectrum and planktonic composition of nearshore and open ocean regimes in conjunction with traditional methods of enumeration, such as net tows, to compare with classically-derived abundances of copepods, appendicularia, chaetognatha, and protozoa (Forest et al. 2012). Studies that aimed to quantify the abundance of organisms that normally become damaged and less identifiable through net tows, such as Rhizaria, have been using the UVP with notable success (Barth & Stone, 2022; Biard & Ohman, 2020; Forest et al., 2012; Turner, 2015). For example, Biard and Ohman (2020) used the UVP to investigate vertical niche partitioning of Rhizaria. To classify morphologically distinct Rhizaria taxa, the authors defined several criteria for UVP images collected within the twilight zone of the California Current Ecosystem. This paper was one of the first to outline elements of a regional taxonomic guide for the UVP. Further, morphological criteria have also been developed for the UVP-based discrimination of marine aggregate morphotypes and their relationship to export flux (Guidi et al., 2008, 2012; Trudnowska et al., 2021).

Though large numbers of profiles have been collected using the UVP, many studies primarily or exclusively use the particle data from the UVP to reconstruct particle size distributions, and refrain from classifying images (Bisson et al., 2022; Kiko et al., 2022; Stemmann et al., 2008). This is, in part, due to the large time investment required to manually classify these images and the lack of standardized instructions for how to classify the images that are collected by new UVP users. As such, higher efficiency methods of image annotation and standardized guides for classifying UVP images are needed so that UVP users can address

questions about organismal diversity, seasonal community compositions, and basin-wide distributions of plankton and marine aggregates.

1.4 Image Annotation and the Rise of Machine Learning

Modern marine image collection has become increasingly semi-automated through attachment of optical instruments to floats and gliders (Ohman et al., 2019; Whitmore & Ohman, 2021). At the same time, the advent of complex machine learning algorithms (MLAs), such as convolutional neural networks (CNNs), has semi-automated the process of particle/organismal classification (Cheng et al., 2019; Lee et al., 2016; Luo et al., 2018; Py et al., 2016). Such MLAs extract features from annotated image datasets and apply learned patterns in image structure to the recognition of unannotated datasets to predict the identity of massive image libraries. These approaches aim to overcome some of the limitations of manual annotation of images collected by *in situ* imaging instruments. In particular, these MLA processes reduce the time-to-annotate for the large datasets collected by *in situ* imaging instruments which historically have been manually annotated.

Manual annotation of organismal and marine aggregate images also suffers from several human biases, including those associated with short term memory, recency, fatigue, positivity bias, and annotator confidence subjectivity (Culverhouse et al., 2003; González et al., 2017; Kenitz et al., 2023). While expert annotators reach annotation self-consistency when presented with replicate randomized samples of up to 99%, in some cases observed minimums can be as low as 68.2% (Culverhouse et al., 2014). Existing studies based on microscopy often vary in terms of the time-to-annotate the dataset, number of samples annotated, diversity of organisms in the target community, and relative level of expertise reported per annotator (Culverhouse et al., 2003; Culverhouse et al., 2014; Kenitz et al., 2023). Importantly, manual annotation is extraordinarily laborious, limiting the scope of image datasets that can be analyzed, a problem of particular importance to long-term ocean time series.

Although machine learning algorithms help to automate the image annotation process, they still require human oversight. MLAs require a training set of images that have been curated by expert human annotators. MLAs also require human validation of their outputs for verification of annotations and tailoring of MLA performance. The quality of the MLA outputs is reliant on the quality of the annotated datasets that are used in their training. Further, new training sets are needed for MLAs when used to characterize a novel region or time series, as borrowing non-contextualized training sets can lead to dataset shift (González et al., 2017). The need for human annotators is of particular concern for ocean time-series, which can outlive the

career of a single expert annotator, in addition to perhaps being prone to dataset shift as a function of their inherent establishment as a sentinel for changing environments. Collectively, the requirement for human expert annotators highlights the issue of a dwindling number of expert morphological taxonomists in the field of oceanography (Hebert et al., 2003).

Current repositories of UVP particle data are increasingly large and are likely beyond the ability to annotate manually in their entirety (Kiko et al., 2022). In addition, MLAs continue to improve in the accuracy of their annotations, as previously discussed. To streamline the annotation process, accelerate group science, and scaffold time series annotations, an onramp for training new annotators and a set of inheritable classification standards is needed. This sentiment has been echoed by a community of experts who recommend the preservation of human annotation at all stages of automated image annotation to ensure high-quality ecological data outputs from MLAs (Axler, 2020; Kenitz et al., 2020; Kenitz et al., 2023).

1.5 This Study

The aims of this study were to 1) provide an onramp for new annotators in the creation of training sets and validation of MLA outputs based on classification standards established within the context of their study region; 2) translate morphological descriptions of taxa to those characteristics observable in UVP images; and 3) establish a standard methodology for training new annotation experts working with the UVP, based on a set of inheritable annotation criteria. To establish inheritable standards, criteria for the description of novel UVP objects are presented that draw from morphotaxonomic descriptions brought into the imaging context of the UVP. Here, I aim to increase the consistency of classifications made by human annotators in the establishment of high-quality training sets and validation of MLA products by creating a set of standards for annotation. A case study from Station ALOHA in the NPSG is presented to illustrate the scientific questions that can be addressed using these methods. For the case study, UVP data from August 2020 to October 2023 (HOT cruises 321 to 345 with noted gaps due to UVP use on conflicting cruises) and the Simons Collaboration on Ocean Processes and Ecology (SCOPE) Particles and Growth in the Oceanic Nutricline (PARAGON) 2021 and 2022 cruises were used to illustrate the described methods with open ocean data.

2. METHODS

2.1 Underwater Vision Profiler Instrument

The underwater vision profiler 5 (HD) was developed to examine the distribution, concentration, and classification of large (0.064 mm - ~54.0 mm) particles *in situ* and in combination with other optical and biogeochemical instrumentation for the exploration of ecological diversity and particle dynamics (Picheral et al., 2010; hereafter P10).

The UVP (shown in Figure 1) is comprised of an *in-situ* camera that collects images of particles illuminated by two flashing columnar red LEDs (625 nm) within a ~1 L volume of water, and with a depth rating of 6,000 db. With a weight of 30 kg, a vertical height of 1 m and a 35 cm base diameter, the UVP is small enough for attachment to a CTD rosette, or for solo casting inside a dedicated optics cage. The UVP uses an internal Sony CCD camera to acquire images at a rate between 6-20 Hz depending on user preset modes of collection and is optimized for a descent speed ~1 m s^{-1} in keeping with conventional CTD rosette descent speeds. The UVP can be connected to a Seabird CTD for measurements of conductivity, temperature and depth, facilitating comparison of images to common ecological parameters. The UVP houses an internal pressure sensor of its own in case connection to an independent CTD is not feasible.

Particles imaged by the UVP HD have a lower limit equivalent spherical diameter (ESD) of 0.064 mm and an upper maximum of ~54.0 mm. Imaged particles that are larger than 0.500 mm ESD are converted into vignettes that can be used for taxonomic classification. These images are collected in 4-megapixel greyscale and output by the UVP in units of $pixel^2$, to then be converted to mm^2 units through a method of minimization described in P10 (Picheral et al., 2010). The application of the Picheral equation is explained later in Section 2.2 on UVP particle data processing.

The UVP imaging volume is specific to each instrument and was determined via individual light submersion in an aquarium and analysis of the resulting light field (Picheral et al., 2010). The volume of the UVP used in this paper (serial number 222) was determined to be 1.22 L but will be referred to as 1 L for simplicity going forward. The particle counts for the UVP were calibrated through a side-by-side cast in the Mediterranean Sea using a 'gold standard' UVP maintained by Hydroptic Inc. as described in P10 (Picheral et al., 2010). In addition, the UVP uses an intelligent camera setup wherein the entire image volume is kept in focus to detect all particles that reflect the 625 nm red light emitted into the camera lens. The pixel size of imaged particles was also determined by Hydroptic Inc. using reference organisms of known dimensions and is specific to each instrument.

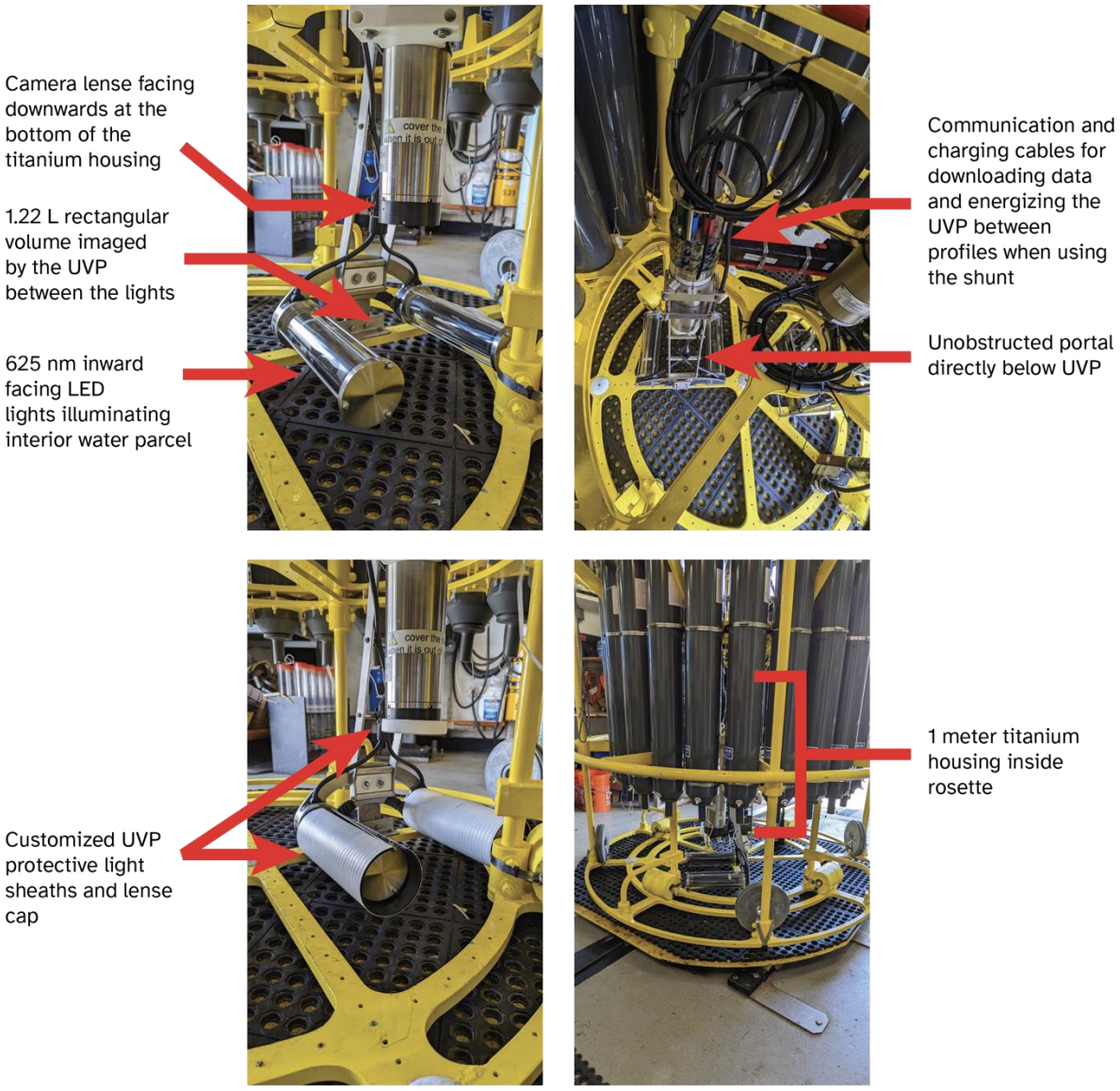


Figure 1. UVP diagram

Underwater Vision Profiler 5 HD (UVP) structures and attachment scheme as bound to a CTD rosette aboard R/V Kilo Moana during the Hawai‘i Ocean Time-series (HOT) 349.

2.2 Deployment

The UVP is traditionally attached to a CTD rosette or optics cage in a downward-facing orientation to collect images on the downcast, a practice that is intended to reduce possible perturbation of the imaged water volume by the frame on which the UVP is mounted. The start protocol that begins and terminates image collection can be accomplished by either the pressure protocol or the I/O shunt protocol. The pressure protocol uses the ascent and descent of the

UVP to start and stop image collection, whereas the I/O shunt protocol uses a shunt to start and stop the UVP much the same as flipping a switch, and this is done manually on deck prior to deployment. While the I/O shunt protocol is more straightforward in principle, we found that the pressure protocol was necessary to cool the UVP camera prior to all casts. Further details of the two protocols can be found in the UVP 5 User Manual.

During deployment the UVP was lowered at ~1 m s^{-1} to prevent the overlap of 1-liter images, which occurs at speeds < 0.3 m s^{-1} because of the 3.5 cm vertical dimension of the imaged volume and the 9 cm inter-image gap at an acquisition rate of 11 Hz (Picheral et al., 2010). However, due to ship movement, swells, and the necessary initial ascent in the pressure protocol, a constant rate of ~1 m s^{-1} in deployment was often unattainable and required post-processing to remove images from periods of descent below the 0.3 m s^{-1} threshold. Overlapping images collected at speeds below the threshold result in duplications of large particles and overestimation of organismal abundance and particle concentrations (Barth & Stone, 2022). Heave compensation was used on all cruises to smooth the descent of the UVP and reduce instances of image overlap.

2.3 Particle Data Analysis

Once the data were downloaded from the UVP following recovery, the vignettes were processed using Zooprocess (https://sites.google.com/view/piqv/softwares/uvp5) and uploaded to EcoTaxa following the guidelines outlined in Supplemental 1. Prior to discrimination of the image identities, all particles collected by the UVP were converted from $pixel^2$ to mm^2 to provide context for organismal size when manually annotating the images. To calculate the size of the images in mm^2 space (Sm) from $pixel^2$ space (Sp), equation 1 is used.

$$Sm = Aa \times Sp^{Exp} \quad (1)$$

This equation was derived in P10 (and adapted to this format in Kiko. et al, 2022) with Aa being the area of a single pixel in square millimeters, and Exp being a dimensionless adjustment factor (Kiko et al., 2022; Picheral et al., 2010). Both constants were determined experimentally by Hydroptic for this UVP 5 (serial number 222) via comparison to samples sourced from the Mediterranean by the Laboratoire d'Océanographie de Villefranche-sur-Mer (LOV). Further, using a log transformed minimization of subsampled particles and a subsequent jackknife procedure to estimate error, optimal values for Aa and B were determined. For additional information regarding the details of the two procedures, refer to P10 (Picheral et al., 2010). As a

result, by applying the Aa and Exp constants to the surface area in pixels^2 (Sp) for each particle such as is described in Eq. 1 above, the surface area of the particle in mm^2 (Sm) is found. While images captured by the UVP are not perfectly spherical, such as in the case of copepods, siphonophores, alciopids, and many others, the approximate ESD of these vignettes was used for comparison to conventional descriptions of UVP data. As such, ESD was found using the millimeter area of the imaged object following ESD = $\sqrt{4 \times Sm}$.

To evaluate the abundance of organisms in classified images the data were binned by depth, and the liters imaged within those bins were summed. The number of classified images was then divided by the liters imaged in those bin widths to provide a measurement of concentration in units of counts per liter (#/L). At this point it is essential that pauses, ascension periods, and duplicate images have been removed from the dataset to mitigate over- or underestimation of particles by double counting or over-dilution. The accuracy of the concentrations for the classified images is limited at the lower bound by the resolution of the UVP 5 (HD) camera, and at the upper bound by the maximum volume imaged, in this case, 1.22L. As such, it is important to consider the size of the target image category, for example siphonophores, which often exceed the dimensions of the imaged water volume, when considering the accuracy of the concentration determined by the UVP.

2.4 Building a Training Set

Training sets built for MLAs provide a set of images from which object features will be extracted and associated with pre-allocated or discovered image categories. For this reason, it is ideal that a training set be representative of the spatial and temporal ecological context to which the MLA will be applied (González et al., 2017). Furthermore, modern MLAs often require large quantities of image data, on the order of tens of thousands of images, to correctly define classes depending on the diversity of image categories in the dataset (Luo et al., 2018). This is, in part, due to the large range in orientations, sizes, and resolutions of the images provided to the MLA from an instrument such as the UVP. However, changes to the quality and consistency of the provided images could improve the performance of the MLA when quantity is lacking (Pei et al., 2021). Finally, the distribution of the image categories provided to the MLA should be similar to the distribution of those categories in the larger dataset to which the MLA will be applied. Considerations taken here, and suggested for future studies, concerning each of these aspects in an ideal training set are described.

Quantity vs. Quality

Conventionally, thousands of images per object category are required to optimize the performance of the modern MLA (Rangineni, 2023). This poses an issue for large particle annotation in the oligotrophic NPSG due to the natural rarity of most image categories. The most feasible solution was to improve the quality and consistency of the images provided in the training set. Higher quality images included representations of unique characteristics (such as copepod antennae), and therefore provided a consistent set of features from which to extract categories. To be clear, *images that are not perfect should be included* as they are representative of the true distribution of UVP images that the MLA will encounter. The distinction here is that images that the annotator is not confident belong to the category should not be included in the name of increased image quantity, as they may bias the training set feature associations. Further, by presenting consistent representations of images in their categories, feature extraction should be more consistent. It is therefore a priority that categories are morphologically homogenous, as categories that are too broad will introduce inconsistent structural information during training. Categories must also not be too narrowly defined, as this may not provide enough images for MLA training. Finding this balance involves evaluation of the MLA outputs and subsequent re-training with broader or narrower categories to increase MLA performance.

Distribution

Image categories provided in the training set should have a representative distribution in comparison to the larger datasets to which the trained MLA will be applied (González et al., 2017; Rangineni, 2023). This prevents minority categories and highly morphologically variable categories from disproportionately biasing feature extraction. Additionally, multiple orientations, sizes, and qualities of images that are still within the bounds of image annotation criteria should be provided. The MLA is likely to encounter the full range of image qualities within a category when processing the larger dataset, and as such, these should be represented appropriately in the training set.

Contextualized Data

The training set should include data that is contextually relevant to the larger dataset that the MLA will process. This prevents the phenomenon of dataset shift as described by Gonzalez (2017). Dataset shift describes how the regional or temporal morphotypes and relative distributions of the imaged populations in the training set can bias an MLA towards the population upon which it was trained, leading to poor performance within novel temporal and

regional settings. For this reason, training sets should be created using data from the region or temporal setting to which the MLA will be applied. In the context of a changing regime, such as a time series, it is recommended that the performance of the MLA is assessed periodically to determine when new training data needs to be curated and applied.

To ensure that training sets include a consistent quality of images that have been annotated within the context of the study region, it is necessary to create classification standards. By providing standards with contextualized criteria for image validation one can improve the quality of the training set, and thus potentially the performance of the MLA that is being trained. Including criteria for training set standardization also improves the efficiency of MLA product validation. Further, this keeps the validation process consistent with the annotations included in the training set. This is of particular concern for time series, which may have large periods of time between MLA trainings, or overturn in annotators who perform validations and create new training sets.

2.5 Classification Standards

To begin classifying the images that were collected by the UVP (example image shown in Figure 2), it is helpful to have a guide that describes the possible and important classifications for the region of study accompanied by images and descriptions for comparison to facilitate internal and inter-annotator consistency (González et al., 2017). This guide is not intended to provide a 'gold standard' to which all classified objects must be identical, but rather to highlight criteria that can be referred to for identifying the critical qualities of organisms in various orientations, sizes, and resolutions of collected images. The intention behind using geometric language (Figure 3) in morphological criteria is to allow easier feature extraction by MLA's and the ability to efficiently redesign classifications based on MLA performance. In addition, a guide-based schema, such as represented in Figure 4, is intended to increase project longevity by allowing annotators to hand off annotation criteria to future generations of annotators for extended image annotation projects, such as time series. This method also provides criteria for annotators to reference during classification rather than relying on subjective thresholds, such as the arbitrary 75% confidence for validation suggested for experts to internally regulate (Kenitz et al. 2023). Further, the detailing of classification standards allows communication in published manuscripts and to MLA developers, as to the morphological reasoning behind individual image validations.

Terminology

1. **Object** - The individual particle that is captured in the image by the UVP.
2. **Image** - The vignette of the object that was downloaded from the UVP after processing with Zooprocess.
3. **Category** - The identity group of the image as it is grouped taxonomically (e.g., copepod or marine aggregate).
4. **Classification** - The process of identifying the category to which an image belongs.

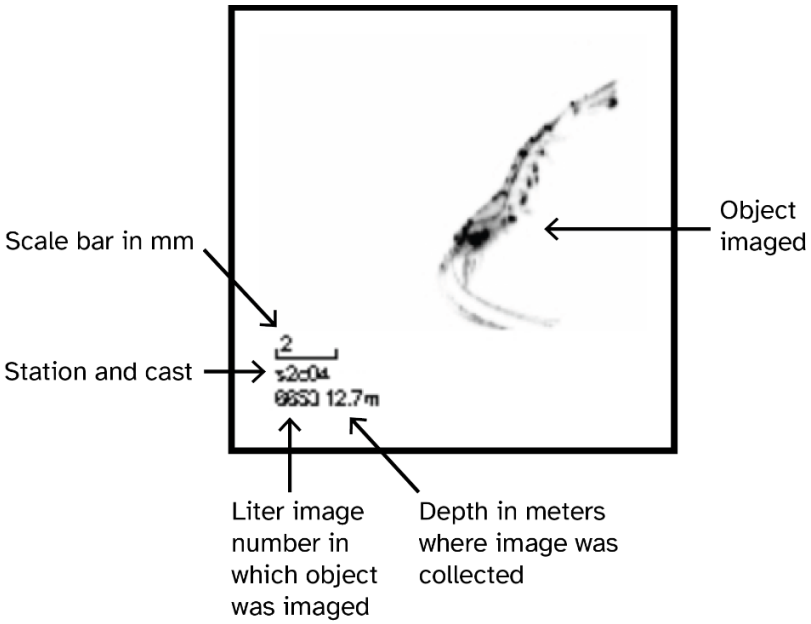


Figure 2. UVP image layout

Image from the UVP with description of the image attributes after processing through Zooprocess. Scale bar = 2; station and cast = s2c04; liter image = 6650; depth = 12.7 m.

Image pre-processing

When manually removing those duplicates that escape the post-processing of UVP ascension periods and pauses, it was decided that images that include the most complete and accurately identifiable criteria of the object imaged were to be included and used in the determination of the concentration for that object's occurrence. However, it must be noted that the most complete image of an object does not always contain the most identifiable characteristic of the object. An example being an image that contains most of a cnidarian but is imaged once more with less of the organism visible. In such a case, the image may now include a taxonomically unique feature that allows for a positive identification for which the first image

would not. Here, the less complete but higher accuracy image was used, prioritizing image classification accuracy over object completeness or size. The issue of duplication is common and described in detail in Barth et al., (2022).

2.5.1 Taxonomic Specificity

The taxonomic level to which images could be classified was largely determined by the resolution of the images collected by the UVP. As such, characteristic morphological features had to be unique to the classification category and describe a category of images numerically abundant enough to be scientifically relevant. In addition, objects quickly pass by the UVP and are imaged in various orientations. For this reason, criteria for classification had to be either universally apparent from all sides of the object imaged, or meet a combination of presented criteria, allowing certain identifiers to be absent. For example, copepods, primarily calanoids, are here described by the presence of a perpendicular antenna, a fusiform overall body shape, and a constriction separating the urosome from the prosome. While all three of these features are discernable in the ideal image collected by the UVP of a copepod, not all images were ideal. As such, it was determined that when two out of these three criteria were present, the image could be classified as a copepod while maintaining accuracy of the annotation. This is supported by the comparison of the copepod classifications by two annotators in the confusion matrix presented in Section 3.4.3.

While many images did not contain sufficient morphological uniqueness to consistently discern phylogenetic level beyond phylum, others presented distinct features that could easily be discerned to the genus level, such as in the case of the colonial form of the diazotrophic genus *Trichodesmium*. However, occasionally there would be morphologically unique features that allowed for order or even genus level classification, but a paucity of organisms in those categories, requiring a grouping at a higher phylogenetic level after classification to draw scientific conclusions with statistical merit. Such was the case with the Rhizaria category, within which Phaeodaria could be distinguished, in addition to Collodaria and Acantharia, but for whom the abundances of these more specific categorizations were not numerous. They were given a more general morphological description that could be applied across the phylum Rhizaria.

Additionally, in some cases there was a morphological disparity between two or more categorizations that were numerically low yet shared higher-level taxonomy. For example, if one were to find low numbers of copepods, amphipods, and euphausiids it might be tempting to group them into the subphylum crustacea. However, these three organismal categories did not share appreciable similarities in their morphological appearance as seen by the UVP, and to

include them together for the purposes of training an MLA would have provided vastly different features under the same categorization. Thus, they were best grouped individually, yet at low abundance. Broadly, it was found that the level of specificity most useful for UVP training set classification was at the phylum or class level, with a few exceptions for notably abundant organisms (e.g., the genus *Trichodesmium*).

2.5.2 Reference Images

Within the UVP classification standards, reference images from the UVP, microscopy, and illustration were used to provide a point of comparison for the annotator. The purpose of using reference images in this guide was to establish criteria for various image orientations, resolution qualities, size, internal morphological dissimilarity, regional appearances, and potential identifiers.

Each category of image was listed with example UVP images collected using the UVP serial number 222 in the North Pacific. The images from the UVP were intentionally selected to vary in their rotation, degradation (such as in the case of broken or incomplete objects), and resolution. The UVP reference images included instances with clear appearance of the criteria for annotation, but also instances of images that were blurry, occluded, or absent of certain criteria, yet nevertheless accurate to the category. This inclusion of imperfect images alongside more exemplary images allowed for annotators to see the range of acceptable categorization, as well as learn what the thresholds were for an unacceptable image.

Secondarily, reference images from a higher order of clarity, in this case a light microscope and illustration, were included for reference. This allowed annotators to anticipate and identify features of objects which had not previously been imaged by the UVP, and thus could not be used for reference. In cases where no sample was able to be collected for imaging by light microscope, historical illustrations were used. The illustrated images that were included in the guide were sourced from *Marine Plankton* by Castellani, C., & Edwards, M. (2017). The references that were from light microscopy were collected via 333 μm net tow on HOT-345. Providing reference images from the same region as the UVP sample dataset serves the same function as creating contextualized training sets for MLAs intended to be used in a specific region (González et al., 2017).

2.5.3 Labels

To label each image descriptor such that a new annotator could learn the relative importance of the identifying criteria, three labels were created: critical criteria (CC), supplemental descriptors (SD), and possible confusions (PC). These labels helped to determine

which descriptors were being employed for positive validation of images and how to discern between categories that were visually similar.

The **critical criteria** include those identifiers that were necessary to include for all images of the described category, with certain exceptions where only a specified amount of the critical criteria was present at once but were interchangeable, such as in the previous example of the copepod identifiers in section 2.4. For each critical criteria the required number of CCs, presence of certain shape confirmations (such as colonial or solitary), and a brief description of the identifier was given.

For some images there were extra pieces of information, such as thoracic segmentation or seminal vesicles, that may be present in an image, but which were not critical to the identification of the category. Such descriptors were referred to as **supplemental descriptors** (SDs). SDs were intended to provide additional context to confirm the identity of an organism if the critical criteria were poorly resolved, unclear, but not absent. Additionally, these SDs helped to identify structural components that may have confused the annotator in their presence, but which were now noted as a possible feature that should not exclude the image from being categorized as the predicted identity.

Certain categories may have hard to distinguish descriptors that overlap with another category's descriptors. In such cases, explicitly denoting the delineations between the two categories was helpful. For this reason, the descriptions of categories that could be commonly confused with one another were denoted as **possible confusions** in the descriptions under PC's.

2.5.4 Descriptions

One of the principal tasks in creating annotation standards for the UVP was to translate the more detailed illustrations and descriptions based on primarily light microscopy into the more limited discernible features captured by the UVP. For example, while *Marine Plankton* by Castellani, C., & Edwards, M. (2017) describes the general morphology of chaetognatha using, in part, its tripartite body, its lateral fins, and eyes, not all of these descriptors may be useful for classifying UVP images. Based on our images, the tripartite body segments were consistently visible, and were thus included in the CC. The lateral fins were occasionally visible depending on image resolution and thus were labeled SD. Finally, the eyes were not resolved in any images and thus were not included in the standards for identification of chaetognatha by the UVP. These feature descriptions were reached through discussion of the most easily recognizable and consistently apparent qualities of the UVP images by the annotators and then compared to the necessary descriptors indicated by taxonomic reference material. The following are descriptions

of those parameters used here to convey to annotators common taxonomic guidelines for UVP-collected images.

Perimeter

The margins of an imaged object, whether interrupted or smooth, was helpful in communicating the transition between external and internal structures. For example, the delineation of the gelatinous sheath which encapsulates the irregular silica shells of the colonial radiolarian collodaria from the background of the image by the appearance of its smooth and complete surrounding perimeter. This helped to create a distinction between collodaria and a collection of small aggregates.

Gray level

Used as a proxy for the density or translucence described in taxonomic guides, descriptions of mean gray level helped to communicate the contrast between object structures. In addition, this provided a visible metric for the translucency of gelatinous organisms versus, say, the higher opacity of crustacean shells. An example of this variation in the mean grey level of dense and translucent structures is shown in the grey level box of Figure 3. The magnitude of an image's gray level also helped to identify the blurriness of an image by way of smooth gray level gradients, and the patterns created by repeating structures.

Blur

A quality that is readily apparent in many UVP images, and yet not included in classical criteria for identifying imaged objects was blur. Blur presented in images not as a binary quality, but as a gradient of image resolution and contrast which pushed certain images beyond the ability to recognize critical criteria, and thus made them unclassifiable. To describe blur as a quality in and of itself, and not as the absence of critical criteria it was thus described as such: the smooth transition from the external background gray level to a higher internal gray level leaving open the possibility for existing structures at the background gray level whether inside or outside the observed object perimeter.

Geometry

Geometric patterns and symmetry formed by object structures in their silhouettes were used to convey identifiable features of organisms and detritus. For example, the appearance of concentric 2-dimensional circles that denoted the ectoplasm and endoplasm of most Rhizaria

was simply communicated and replicated between annotators. In addition, the use of geometric language, such as perpendicularity or parallelism, in describing the relationship of object structures to one another within an image helped to denote the orientation of spines, tails, banding, and other such features. The use of geometric descriptions allowed annotators to identify structures that may move in 3-dimensional space, as the UVP does not constrain the orientation of the particles it images. This can provide, at times, limiting angles of observation. As such, the use of geometric language to establish a system of planar orientation which is specific to the organism and irrespective of its imaged orientation proved to be a more comprehensive method. This contrasts with descriptions built off the most ideal orientation of an organism, as is often the case with illustrated references. For example, establishing copepod antennae as being perpendicular to the *longitudinal axis* of the organism allows the description to hold true even when the copepod is imaged head-on, as the longitudinal axis does not change for the copepod, even though its *longest visible dimension* may change based on the orientation of the image.

Size

The length and width of object structures was a critical consideration for many images, not only in their total dimension, but also in the dimensions of structures relative to one another. For instance, the description of endoplasm or nucleus size relative to their ectoplasm size helped to distinguish the identity of different orders of Rhizaria.

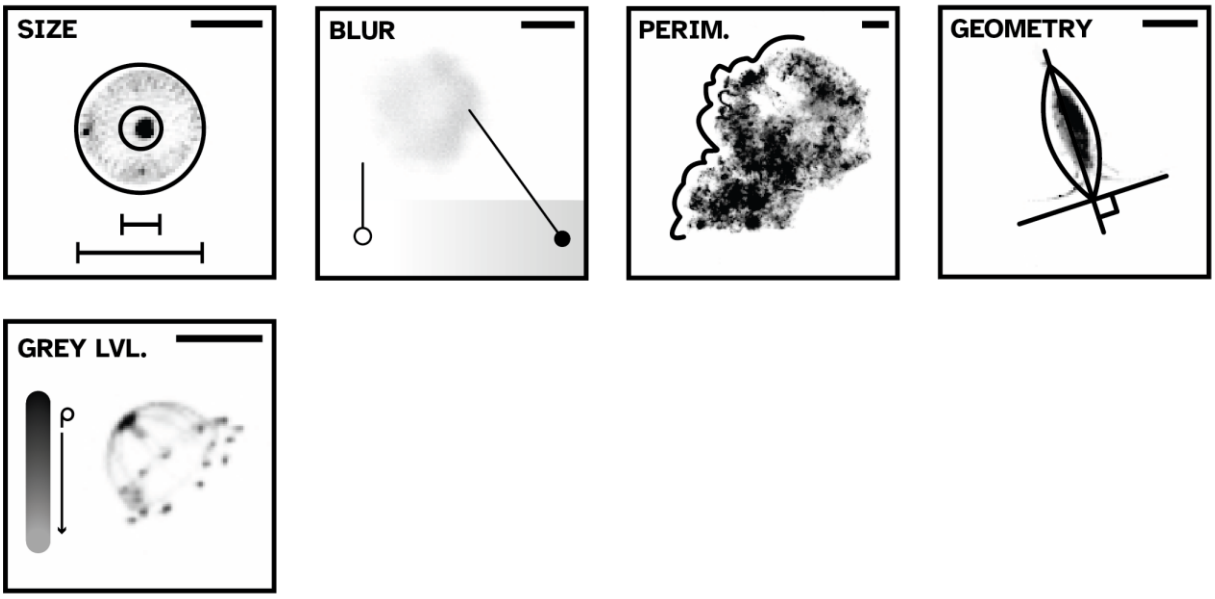
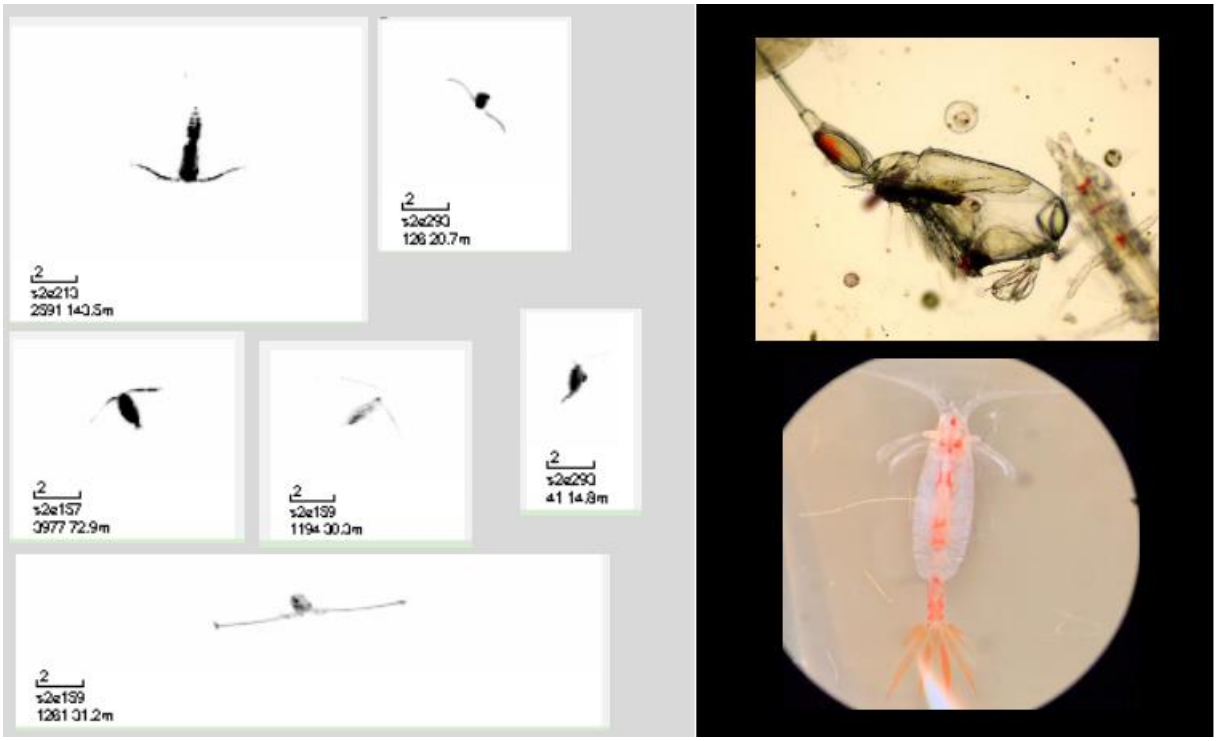


Figure 3. Descriptor visualizations

Visual representation of the parameters used to describe the appearance of objects imaged by the UVP in terms of relative sizes, blur, perimeter, geometry, and mean grey level. The scalebar in the top right corner of each image corresponds to 2 mm. The vertical gradient with the rho symbol in the 'grey lvl.' box indicates the decreasing progression of density with lighter mean grey level, which is used a density proxy, here shown with a cnidarian for presentation of its lighter mean grey level translucent bell, and its denser high mean grey level gonadal and oral organs.



Critical Descriptors: 2/3 required
Antennae pair - size and length variable but perpendicular to the longitudinal axis of the organism
Fusiform silhouette - visible from orientations which capture the longitidunal axis
Urosome - thinner section following a constriction located distally relative to the prosome but on the same longitudinal axis
Supplemental Descriptors:
periopods - rarely visible, possible to see from side view
eggs sacs - below or beside lower urosome / metasome
Possibly Confused with:
Ostracods - antennae shorter and not perpendicular to longitudinal axis, more spherical and absent of urosome.
Appendages shorter relative to overall size when visible

Figure 4. Classification Standards Guide Example

UVP classification standards guide section for the copepod category, including (from left to right) a section for the categorical name, UVP images, reference photos, and descriptors (CC, SD, and PC). DOI of classification guide used in this study: 10.5281/zenodo.10038893.

2.6 Process Study

The above standards, workflow, and terminology were designed to limit common biases and inconsistencies that currently exist in UVP image analysis (Culverhouse et al., 2003; Culverhouse et al., 2014; Kenitz et al., 2023). This new approach was intended to facilitate a higher inter- and intra-annotator classification consistency and better communicate categorization parameters during training and in publications. To provide evidence for the efficacy of the recommended approaches, a comparison of two annotators was conducted to evaluate their self-consistency and cooperative agreement. Two annotators cooperatively

annotated an entire dataset, referred to as the 'Main' set, from a single cruise (HOT-345) using the classification standards and workflow described above. They then returned to the dataset individually to re-annotate a randomized 'Subset' of the cruise data. The results were then compared by category with a confusion matrix, much like how one would assess the performance of an MLA product (Culverhouse et al., 2014), to visualize the agreement and discordance between annotators by image category.

2.6.1 Data Collection

The UVP images for the process study were collected using a UVP 5 (serial number 222) attached to a CTD rosette during the period of October 7th to October 11th of 2023 over 11 profiles aboard R/V Kilo Moana as part of HOT cruise 345 at Station ALOHA in the NPSG. The image data was collected via the pressure protocol deployment scheme and processed in Zooprocess according the UVP 5 Manual, and was subsequently uploaded to an image annotation application, in this case, EcoTaxa. UVP profiles collected on HOT-345 spanned depths from the surface to ~ 4000 m, although most profiles were in the upper 1000 m. From the 11 profiles conducted, 43,849 images of particles above 500 μm ESD were collected by the UVP. HOT 321-341 (excluding HOT 328-334 where the UVP was not onboard) images were used for random forest (RF) training and were collected following the same protocols as described above. Data for particle analysis also was collected during the 2021 and 2022 PARAGON phytoplankton bloom cruises. The deployment scheme and processing were identical to those performed on the HOT cruises, except an optical cage was used to mount the UVP during PARAGON 2021.

2.6.2 Annotation Overview

The Main set contained the full catalog of images from the HOT-345 cruise and was annotated in its entirety by both annotators. To determine the ability of the annotators to reproduce their annotations, and to compare those classifications to another annotator working on the same data set, a Subset of the main set was created. The Subset was created by random sampling of the Main set without replacement, selecting 1,000 images from each profile for an even distribution among casts. This provided 11,000 images, or ~25% of the Main set, for inclusion in the Subset. The 25% threshold was decided for two reasons. Firstly, the relative differences in the number of images collected by each profile limited the upper boundary of how many images could be sampled evenly. Secondarily, the fully annotated HOT-345 Main set included 28.83% of total images that could be validated positively to image categories. 25% was thus chosen to include an even distribution of sampled images from all casts.

Annotator expertise is an important factor in providing confident and accurate annotations for training sets and validation of MLA outputs (Culverhouse et al., 2003). The more experienced annotator, A1, created the classification standards. A1 then trained the second annotator, A2, on how to use the classification standards and image annotation application. A1 and A2 cooperated in annotating the Main set of UVP images collected from HOT-345. First, A2 reviewed the image categories and validated those images that matched the criteria from the classification standards set by A1. Secondarily, A1 reviewed all images, reassigning images where necessary. A1 and A2 then annotated their respective Subsets in isolation. The images in each Subset were identical and in the same order of appearance as presented in the image annotation application. Both annotators individually followed the classification standards for categorizing the images, as was done in the Main set.

2.6.3 Image Annotation Application

EcoTaxa was chosen for the annotation of the UVP datasets used here due to its use within the UVP community, its established documentation, and its compatibility with the native UVP data formatting (Barth & Stone, 2022). However, other such applications exist, such as Morphocluster (Schröder et al., 2020). Morphocluster may be used for the same image annotation purposes as EcoTaxa, but with differences in user interface, data formatting, and integrated classification algorithms. EcoTaxa was created and is operated by Hydroptic Inc. which manufactures the UVP and utilizes a RF algorithm for native image prediction.

Images uploaded to an EcoTaxa project are labeled with the initial status ‘unclassified’. These images have not been run through the random forest algorithm to predict their identity, nor have they been validated as their predicted classification. To assign a predicted classification to the image, the project must be run through the RF algorithm in EcoTaxa or uploaded to the EcoTaxa application after being processed through a non-native image classifier. Here, the native RF in EcoTaxa was used to bin images before manual review.

13 HOT cruises which had been previously annotated following these methods were used to train the EcoTaxa RF. During the feature selection phase, all feature options in EcoTaxa were used for prediction. Once run through the RF, the images were under the status ‘predicted’ and had been assigned to the image category which the RF estimated their highest likelihood of belonging. The RF was therefore limited to making predictions which align with the existing categories of images in its training set. These initial RF predictions were later compared to the manually validated Main set to ascertain the precision of the algorithm following equation 2.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

2

Here, the 'true positives' were defined as those images which the RF predicted as belonging to a category that was later validated as such in the Main set. 'False positives' were defined as those images which the RF predicted to belong in a category that the validated Main set had later assigned elsewhere. The precision of the RF thus communicated the proportion of correctly predicted images.

2.6.4 Image Status

In EcoTaxa, after prediction by its RF, there exist three image status categories: 'predicted', 'validated', and 'dubious'. A fourth status, 'unknown', has been manually included here through the creation of an image category named Unknown. Below is a description of each status as they were used in the current Process Study.

Predicted

Images with the 'predicted' status were assigned by the RF into categories for each annotator to assess. The annotator then reviewed the 'predicted' images, comparing each to the classification standards, and determined if the image belonged to one of the following statuses: validated, dubious, unknown, or predicted. If an image was found not to match any of the categorical criteria in the classification standards, it remained under the 'predicted' status.

Validated

An image was 'validated' when it met the criteria for classification in an image category as described in the classification standards. If the image was identifiable as a new category not yet included in the EcoTaxa predictions, then the new category was created, and the image was classified as 'validated' in that category. For this study, images were only validated if they belonged to an existing category in the classification standards.

Dubious

Dubiously labeled images were those which the annotator was unsure of in classification, requiring further review, but to which a category had been tentatively assigned and not

validated. These images often lacked appropriate resolution of critical criteria or contained multiple objects of interest and would be reassessed with more scrutiny later in the annotation process before being validated or left in the dubious category. Images left in the dubious category for this study were treated as 'predicted' since they were not validated.

Unknown

Images that contained unique features, such as long antennae, enormous eye stalks, or other such distinguishing features which were likely identifiable, but to which no classification standard matched in description would be categorized as 'unknown'. These images are distinct from the 'dubious' classification in that they are possibly identifiable with additional reference outside the described categories of images and are not tentatively or otherwise assignable to existing categories. The unknown status was thus used to hold images for future classification into novel categories. This is a notable designation as well in that it is not a native EcoTaxa status and was created by designation of a category called 'Unknown' within the EcoTaxa taxonomic list and was not to be used in RF training sets but only as a repository for novel images until their proper classification.

3. Results and Discussion

3.1 Large Particle Abundance, Distribution, and Seasonality

Particles larger than 53 μm are generally assumed to be 'sinking' particles (Buesseler et al., 1995), and thus contribute heavily to the flux measured by other *in situ* devices such as Particle Interceptor Traps (PITs). Shown in Figure 5 below is the vertical profile of large particle abundance (#/L) from 13 HOT cruises spanning August 2020 to October 2023. The highest particle counts were routinely measured in the summer (June, July, and August), and occasionally fall (September, October, November), cruises, as evidenced by particle peaks on the order of the 50 – 125 #/L present in the upper 75 db of the water column. Winter (December, January, February) cruises showed low particle counts throughout the water column, while spring (March, April, May) profiles showed occasional subsurface maxima around 100 to 150 db. Regardless of season, the presence of a particle peak was often associated with higher standard deviation values. This high variability of the large particle stock at Station ALOHA during the summer and fall indicates that events that produce or consume these particles are ephemeral. The high abundance and variability of large particles in the summer-fall are consistent with the known timing of phytoplankton blooms in the region (White, 2007). Utilizing UVP data as such could help connect time series observations of eddy fields and chlorophyll to the ongoing sediment trap flux measurements carried out by the HOT program by describing with fine scale resolution these pulses of fast sinking large particles as they are transformed through the water column.

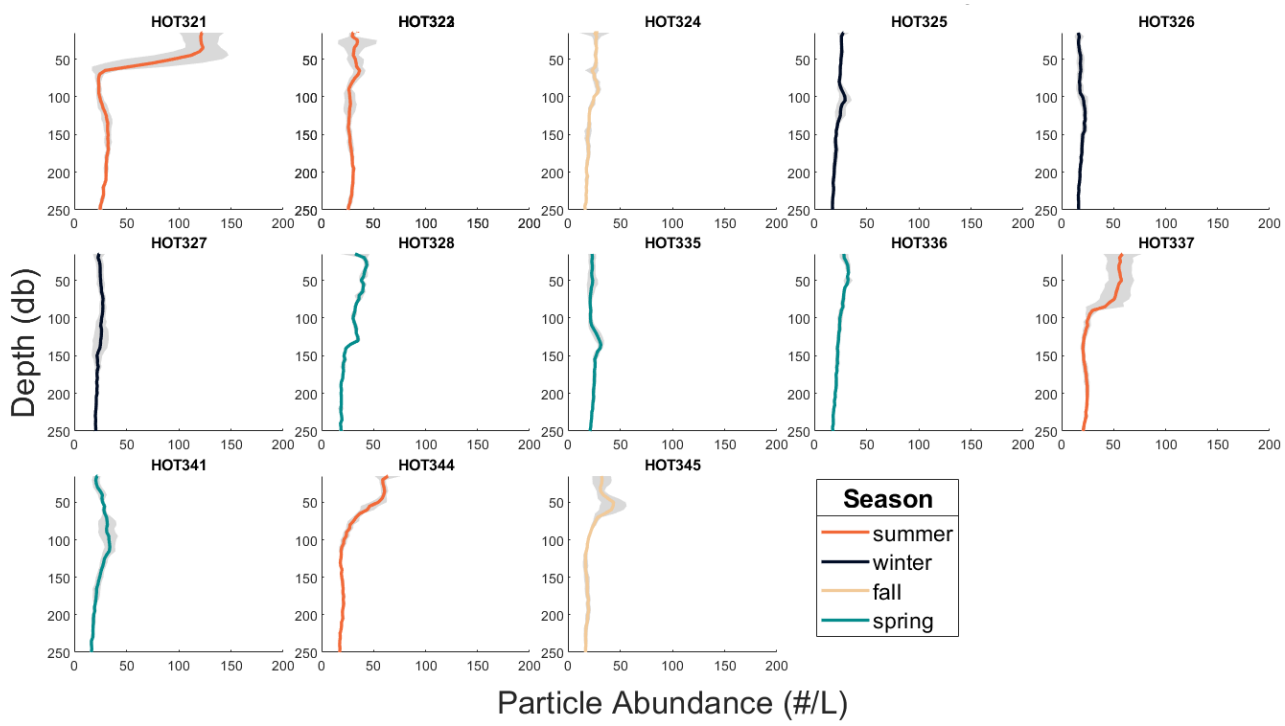


Figure 5. Particle Abundance with Seasonality

Profiles of particle concentration from sea surface to 250 db covering 13 HOT cruises conducted from 2020 to 2023. Concentration presented in units of particle counts per liters of water imaged by the UVP. Shaded portions of each graph represent 1 standard deviation from the average particle abundance plotted in each graph. Average abundances are colored by season.

The slope of the particle size distribution (PSD) can help describe perturbations to the relative concentrations of particles in specific size-classes, which have been tied to changes in the biological activity of the water column (Sheldon et al., 1972). Increasing the abundance of small particles increases the slope of the PSD whereas increased abundance of large particles decreases this slope. A PSD was derived using UVP data in the upper 150 db of Station ALOHA during HOT-345 and found to have slope of -3.33 ± 0.09 (Figure 6). This is roughly aligned with the global mean value of -3.57 ± 0.56 reported in Kiko et al. (2022) for the upper 200 db of casts exceeding 3000 db. However, as is shown in Figure 7, the slope of the HOT-345 PSD varies with depth, at times reaching values as steep as -3.76 in the twilight zone of the water column. Particle production, advective fluxes, aggregation and disaggregation may all serve to alter the observed slope of the PSD with depth (Asper, 1985; McCave et al., 1984; Sheldon et al., 1972). To track how upper water column PSD slopes change in response to season, the PSD of the upper

150 db was compiled (Figure 8). HOT 321 and HOT 337, both summer cruises, have PSD slopes closer in line to the PARAGON 2021 and 2022 phytoplankton blooms, also observed in the summer. The stochastic increases in PSD slope only being observed in summer shows the importance of this period at Station ALOHA for production of large particles, and perhaps contributions to the 'summer export pulse' observed by deep-moored traps (Karl et al., 2021). The remaining HOT cruises had PSD slopes between -3.3 and -3.6, with summer cruises having slightly shallower PSD slopes than winter cruises on average.

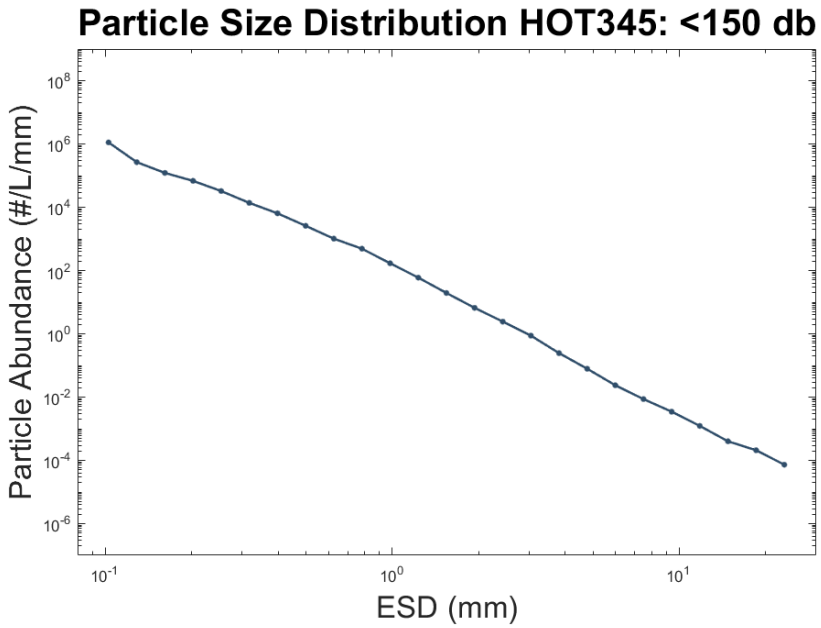


Figure 6. HOT 345 Particle Size Distribution

Distribution of large particles imaged by the UVP during HOT cruise 345 from sea surface to 150 db, plotted on a log-log scale. HOT cruise 345 conducted during October in 2023 at Station ALOHA. The slope of the PSD roughly equates to -3.33± 0.09.

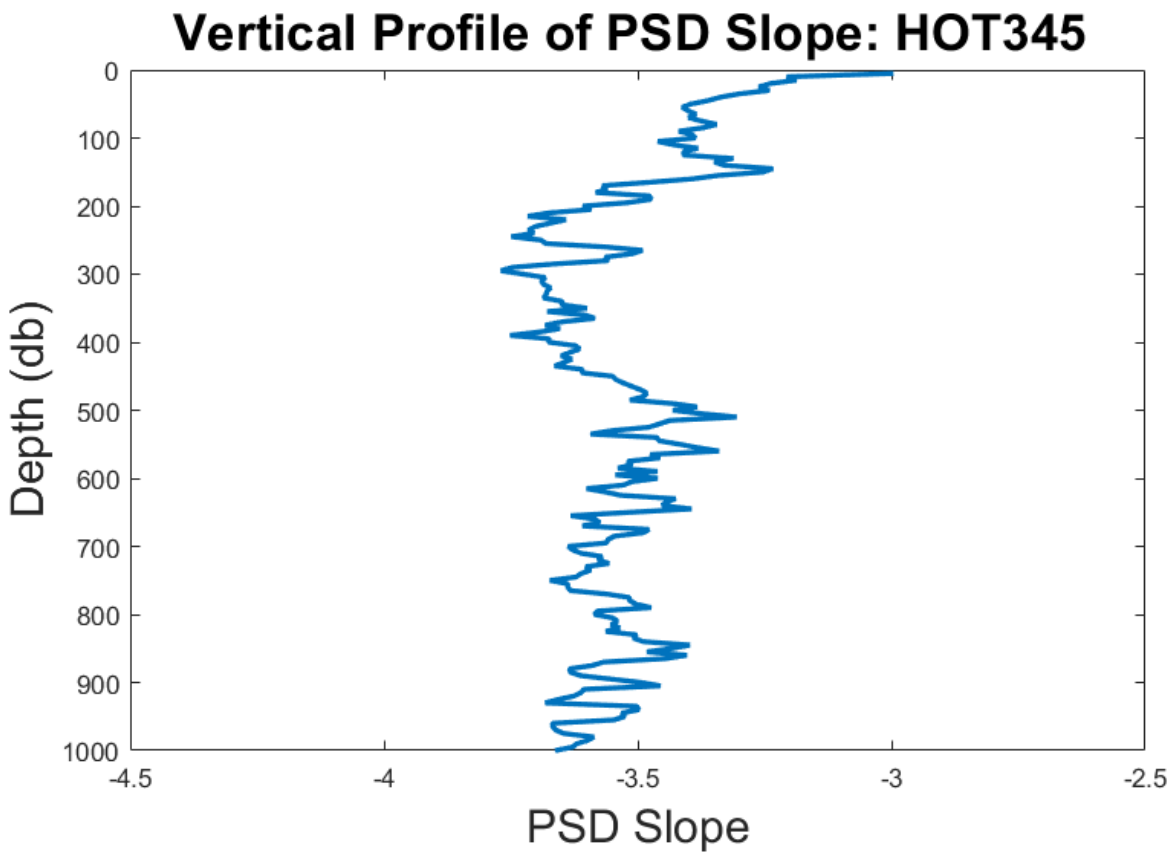


Figure 7. Vertical Profile of Particle Size Distribution Slopes

Particle size distribution slopes binned in 5 db intervals over the upper 150 db of the water column. Data was compiled from HOT cruise 345 during the month of October in 2023 at Station ALOHA. Variability in the slope of the PSD with depth indicates possible biological or physical perturbations of suspended and sinking particles.

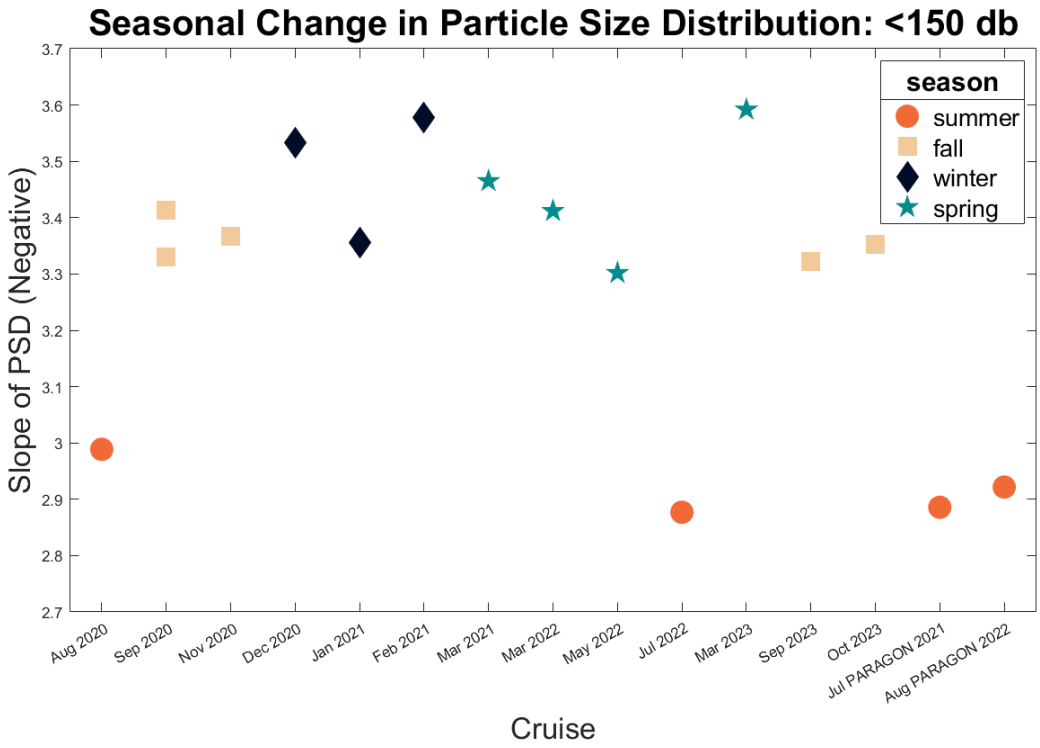


Figure 8. Seasonal Change in Particle Size Distribution

Particle size distribution slopes from 13 HOT cruises spanning August 2020 to October 2023, and PARAGON 2021 and 2022 phytoplankton blooms. Slopes here presented on the Y-axis as positive values derived from negative PSD slopes. Note that anomalous shallow PSDs occur exclusively during summer months, shown in orange circles. Winter cruises, as shown in dark blue diamonds, have steeper slopes (indicative of a greater influence from small particles) than summer cruises.

Perturbations to the particle size distribution, as shown above, have been tied to fluctuations in the biological activity of the ecosystem (Sheldon et al., 1972). Since the UVP creates images for those particles that are > 0.500 mm it was possible to describe changes to the populations of plankton and marine aggregates, and perhaps identify the changes to community composition that may perturb the PSD at the larger size range. To understand the changes that may be transforming the proportion of smaller particles, particularly those below 0.500 mm ESD, future studies should combine UVP particle analysis with other imaging systems that capture the < 0.500 mm size range, such as the IFCB and/or the Scripps Plankton Camera (SPC).

3.2 Image classification – Random Forest Performance

The EcoTaxa RF was evaluated through a comparison of its predictions to the eventual categorizations made in the Main set by both annotators. If the EcoTaxa RF was as able to discern organismal identity as confidently as manual annotations, one could theoretically use its predictions to curate training sets for more accurate and efficient MLAs down the line. The precision of the EcoTaxa RF was calculated using equation 2. Training the EcoTaxa RF with the images from 13 HOT cruises led to an overall precision of 0.3091, or ~ 31%, when applied to the HOT-345 Main set. The precision of individual object categories is shown in Table 1, alongside the image counts used in the training of the RF. While one would expect that an increased quantity of images would lead to increased precision, this is not always the case. *Trichodesmium* had the highest precision (64%) in terms of RF prediction and utilized 5000 training set images in feature extraction. Similarly, copepod prediction was also trained using 5000 images, and came in at 16% precision. This contrasts with the 459 images used to train the Eumalacostraca category, with a ~21% precision output. With such low confidence in object annotation, it would be inadvisable to rely on EcoTaxa's RF categorizations alone as training sets for MLAs. The RF would therefore serve a better purpose as a prescreening to roughly bin images before manual annotation and curation of a training set.

Table 1. Precision of EcoTaxa Random Forest Algorithm

The precision of the EcoTaxa random forest algorithm presented as a percentage and listed by the categories included in the Main set. Image counts per category are included to show the relationship between RF precision and abundance of images included in the training set per category.

Category	Precision (%)	Images in Training Set (#)
Chaetognatha	5.7	147
Cnidaria sp.	2.1	459
Eumalacostraca	21	456
Ostracoda	0.7	160
Rhizaria	2.6	5000
Siphonophorae	4.4	90
Trichodesmium	64	5000
Copepod	16	5000

Diatom Mat	13	62
Fecal Pellet	55	5000
Larval-Fishes	14	251
Marine Aggregate	35	5000
Phyllosoma	0	4

One may argue that the RF is being trained on inconsistent data annotations, and thus providing a variable categorization irrespective of image quantity. In such a case, the manual annotations between annotators, if based on inadequate classification standards or inconsistent annotator performance, should not be consistent. To ascertain such information, the performance of both annotators in their ability to replicate the annotations of the Main set, and comparison to one another, was evaluated.

3.3 Process Study Efficacy

Should manual annotation prove a better method of classifying images for training sets and MLA outputs than reliance upon the EcoTaxa RF alone, then the agreement of A1 and A2 with the Main set should be higher than the precision of the RF. Further, if time series or other extended projects were to make use of such methods such that annotations might be inherited, then the agreement of A1 and A2 should be commensurate with intra-annotator consistency. Both assertions are here addressed through the application of classification standards as described above and evaluated for performance below. Classifications were compared based on intra-annotator consistency, as described by agreement of A1 and A2 Subsets with the Main set of cooperatively annotated images, and inter-annotator consistency, as described by the agreement of classifications between annotator A1 and A2 Subsets. The categories included cover 13 classifications, as listed in Table 2 below. During the annotation process there appeared several images of organisms and detritus for which annotators could identify unique structures but for which no annotation standard yet existed, and thus were not included in the results of the annotations as ‘predicted’ or as ‘validated’ objects and were labeled as ‘unknown’. These images were not included in the counts presented in Table 2.

3.4 Dataset Annotation Comparisons

3.4.1 Annotation Counts

Within the randomized Subset marine aggregate, *Trichodesmium,* Rhizaria, fecal pellets, and copepods dominated the abundance of image counts, with marine aggregates reporting the

largest (Main - 1728 : A1 - 1586 : A2 - 2400) count by an order of magnitude. Except for those images counted in the *Trichodesmium* category, A2 classified higher counts for each category of image. With the context of the A2 being a more novice annotator, one could interpret this as overconfidence in the ability to annotate certain categories of images. By looking at the mean of the three subsets, the over-annotation by A2 can be assessed, as for all categories except marine aggregate (standard deviation = 434.8), *Trichodesmium* (standard deviation = 51.4) and fecal pellets (standard deviation = 53.1) the counts of A2 were within 1 standard deviation of the mean. The most likely explanation is that A2, in their relative inexperience, annotated images that were below the level of confidence applied cooperatively in the Main set and by A1 in their Subset. This is evidenced by the lower number of images left in the 'predicted' category by A2 in their Subset than in the Main set and A1 Subset.

Table 2. Subset Annotation Counts

Abundance of subset HOT-345 UVP images by category as classified by both annotators cooperatively (Main set) and individually by Annotator 1 (A1 Subset) and Annotator 2 (A2 Subset). Counts presented here from the Main set are those which correspond to the images randomly selected for annotation in the individual A1 and A2 Subsets, and do not include all annotated image counts from the Main set.

Category	**Main**	**A1**	**A2**
Chaetognatha	3	2	2
Cnidaria sp.	9	7	11
Eumalocastraca	17	12	15
Ostracoda	2	1	1
Rhizaria	144	122	154
Siphonophorae	2	2	2
Trichodesmium	882	922	820
Copepod	115	73	100
Diatom mat	1	1	1
Fecal pellets	486	488	579
Larval fishes	5	6	6
Marine agg.	1728	1586	2400

Predicted	6708	6880	6011

3.4.2 Annotation Size Thresholds

The Subsets' consistency was analyzed by percentage of images with the same classification, including all image categories, as shown in Table 3. The Subset contained 11,000 images that were selected randomly from the Main set. The total Subset images above 1 mm ESD sums to 10,103 images. 1 mm was chosen as a threshold for image analysis based on Barth & Stones 2022 findings that objects below this ESD were not reliably identifiable (Barth & Stone, 2022). This was consistent with the observations of both annotators during the classification of images. If one compares the overall agreement for those images above 1 mm ESD, both the inter-annotator and intra-annotator comparisons fall above 80% similarity. Perhaps intuitively, as the size of images is increased to >2 mm ESD and >3 mm ESD fractions, the similarity of classified images increased to > 90% and > 95 % respectively. As images become larger, it becomes easier to identify the structures of the objects being imaged as a function of resolution, and therefore the agreement of annotation standards being applied properly to the classification of objects increases.

Table 3. Subset Annotation Agreement

Comparison via percent agreement in Subset image classifications as shown through size fractions of annotated images in equivalent spherical diameter (ESD). Images here classified by both annotators cooperatively (Main set) and individually by Annotator 1 (A1 Subset) and Annotator 2 (A2 Subset).

Subset Comparisons	**>1 mm ESD (%)**	**>2 mm ESD (%)**	**>3 mm ESD (%)**
A1 / Main	87.6	93.5	97.8
A2 / Main	88.1	95.4	96.7
A1 / A2	85.2	92.8	96.5
Total Subset Images	10103	2858	1121

Shown by the total counts in Table 2, the filtering of those images above 2 mm ESD and 3 mm ESD decreases the total count of images being annotated by ~78% and ~89% respectively, in comparison to the >1 mm ESD dataset. As such, only including images above a 2 mm ESD or

3 mm ESD threshold for the purposes of creating a higher quality training set would decrease the quantity of images, and thus orientations, morphotypes, and taxonomic diversity of images from which a potential MLA would extract features. Given the large image quantities needed for many MLAs, this would greatly increase the number of datasets for which human annotation is needed such that thousands of images in each category could be provided to a training set. This translates to increased time, and thus cost, of creating a training set. Additionally, it is possible that this bulk, and rather rough, analysis of agreement could be biased towards a specific category of images, and thus, a more discreet level of analysis is required which looks at the classification of individual images, as is explored below.

3.4.3 Comparing Annotator Classifications

To analyze the categorical classification of images by A1 and A2 a confusion matrix (Figures 9 - 11) was created for which the results were presented as percentages normalized by the counts from A2. Looking at the confusion matrix for annotated images >1 mm ESD there was a majority consensus on most image categories between annotators. Those categories with 100% agreement consist primarily of categories where <10 images were categorized, such as in the case of the chaetognatha category which contains 2 images annotated. Generally, the confusion matrix shows that more of A1's classifications were left in the 'predicted' category across the classification spectrum. In the case of *Trichodesmium* 15% of objects were left in the predicted category for A1 when classified elsewise by A2; this difference was 36% for Cnidaria. The previous bulk percentage agreement in Table 2 showed that there was likely over-annotation by A2. The confusion matrix in Fig. 9 shows that this over-annotation was not of a single category, but rather a general trend across categories. As such, there is not a singular problem category in need of re-description, but possibly a realignment of general classification standard applications between A1 and A2. However, in the case of a cooperative annotation process, over-annotation by a more novice annotator who carries out a first pass of the dataset is preferable to under-annotation.

By looking at the categories from which the most dominant classifications were attributed, a wider spread in agreement can be seen. Rhizaria (75%), copepods (63%), fecal pellets (77%), and marine aggregates (64%) all fall below 80% in inter-annotator classification agreement. However, the *Trichodesmium* category, which is also amongst the dominant taxa has an 85% agreement between annotators. By identifying the categories that low percentage (<80%) classifications were confused with, the underlying morphological structure overlap or ambiguity in morphological description can be inferred. For example, the fecal pellet category was most frequently confused with *Trichodesmium* (3%) and marine aggregate (3%). During

HOT-345 there was an observation of larger *Trichodesmium* colonies than are usually observed, and as such, the tuft shape of *Trichodesmium* had a larger overlap with the visual structure of the elongated fecal pellets than on other cruises. This observation was made by both annotators. As such, clarification in the 'possible confusion' section of the classification standards for *Trichodesmium* was made to help distinguish the two categories. Further, since fecal pellets and marine aggregates are both detritus, it is not surprising that the two were also confused, as they share a certain amorphous quality which creates categorization overlap when marine aggregates are elongated in silhouette. Understanding where these confusions lie helps to re-define categories in classification standards and assess overall annotator inter-comparison and confidences, as well as being an important diagnostic tool regularly employed in the validation of MLA products.

When the confusion matrix is resolved by size, in >2 mm ESD and >3 mm ESD images, we can see a broad trend of increasing percent agreement across classification categories. Most notable is the jump in percent agreement in the marine aggregates category, from 66.54% in the >1 mm ESD matrix to 80.26% in the >2 mm ESD fraction and 85.71% in the >3 mm ESD fraction. While this is caveated by the large decreases in image quantities per group, it speaks to the consensus in which categorizing marine aggregates is dependent on the size of the object imaged. As such, the expertise associated with annotating UVP images may not rely solely on the experience of the annotator, or the clarity of the taxonomic description, but also largely upon the resolved size of the target object in question within UVP datasets.

The level of expertise established in previous studies (Culverhouse et al., 2014), which review the threshold for annotator proficiency and the biases associated with morphological annotation, typically do not include images of the quality produced by the UVP, especially with respect to representation of marine aggregates. Further studies including a wider range of annotators, both in number and level of experience, may be helpful to the UVP community to set instrument specific expertise thresholds. It is likely from the results presented here, when compared to higher resolution optical imaging systems and the physical collection of certain organisms, that the UVP will have a lower threshold for what is considered expert annotation. This may set an upper limit on the accuracy of annotations feasible for certain categories of image. An upper limit on taxonomic confidence should thus be considered when designing scientific questions that include MLA training on datasets created with UVP images like those included here. However, to utilize the unique benefits described above for UVP image datasets in training automated image classifiers and validating MLA products, it is imperative that UVP specific limitations, reference standards, and morphological descriptors be considered.

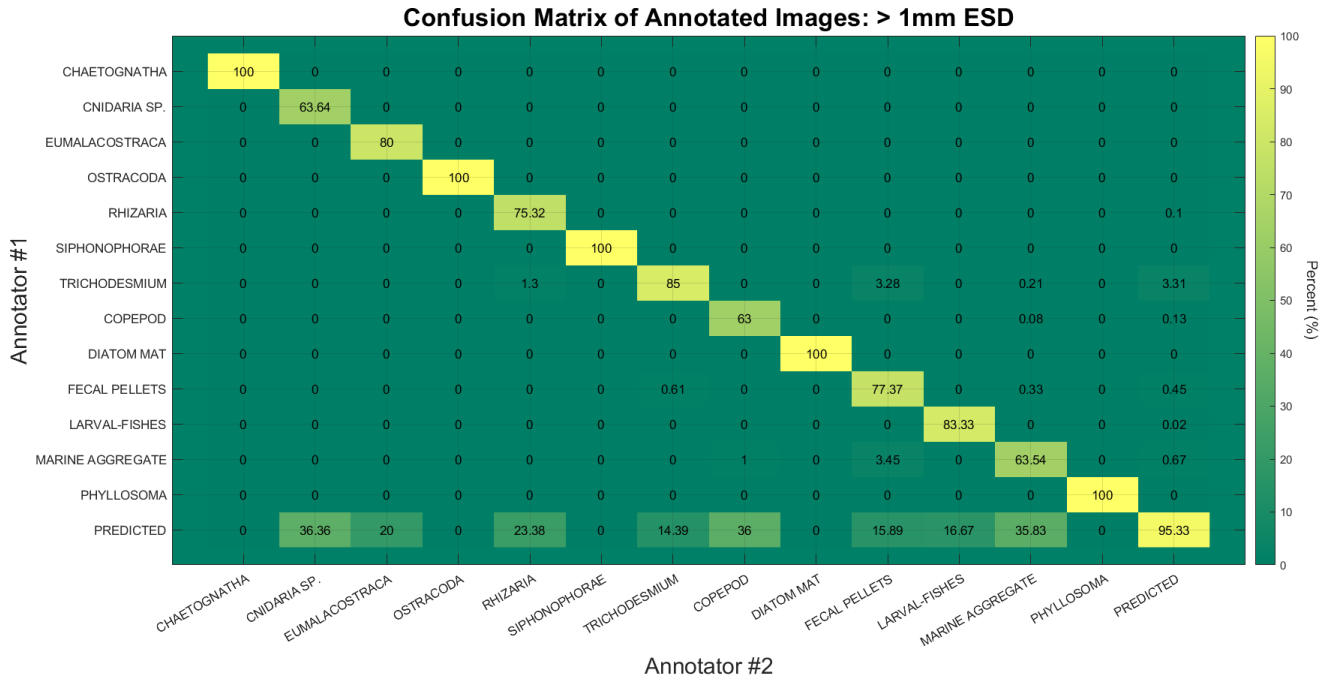


Figure 9. Confusion matrix 1 mm ESD

Confusion matrix of UVP images comparing the subset annotations by category of A1 and A2 in terms of A2 normalized percent agreement. The images categorized here include those which were above 1 mm ESD.

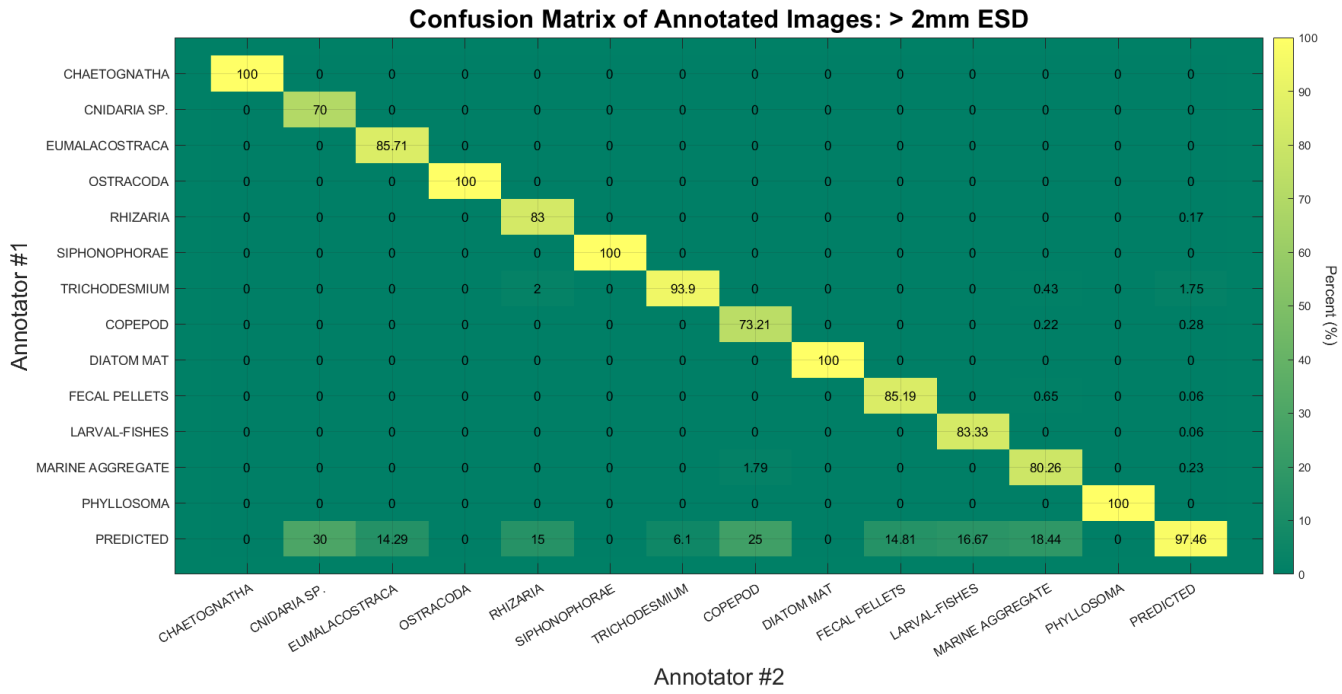


Figure 10. Confusion matrix 2 mm ESD

Confusion matrix of UVP images comparing the subset annotations by category of A1 and A2 in terms of A2 normalized percent agreement. The images categorized here include those which were above 2mm ESD.

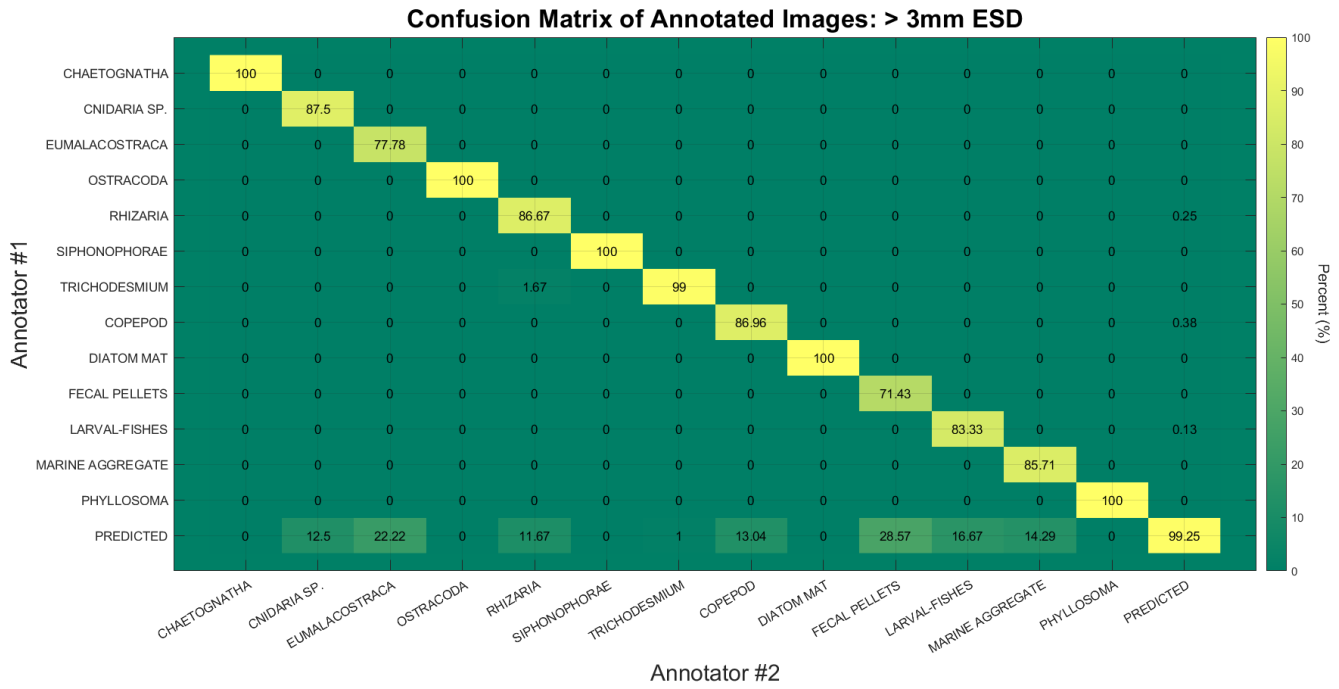


Figure 11. Confusion matrix 3 mm ESD

Confusion matrix of UVP images comparing the subset annotations by category of A1 and A2 in terms of A2 normalized percent agreement. The images categorized here include those which were above 3mm ESD.

3.5 Community Composition

Investigation of the Main set validated images reveals that most identifiable images were in the upper 100 db of the water column with a steep drop off in the count of validated images below this depth (Fig. 12). This is likely due to the exponential decline in abundance of larger organisms with depth in the NPSG. UVP-collected images were predominantly from surface populations that were imaged more frequently than those from deeper in the water column and this trend in the data may be considered a bias when evaluating MLA performance.

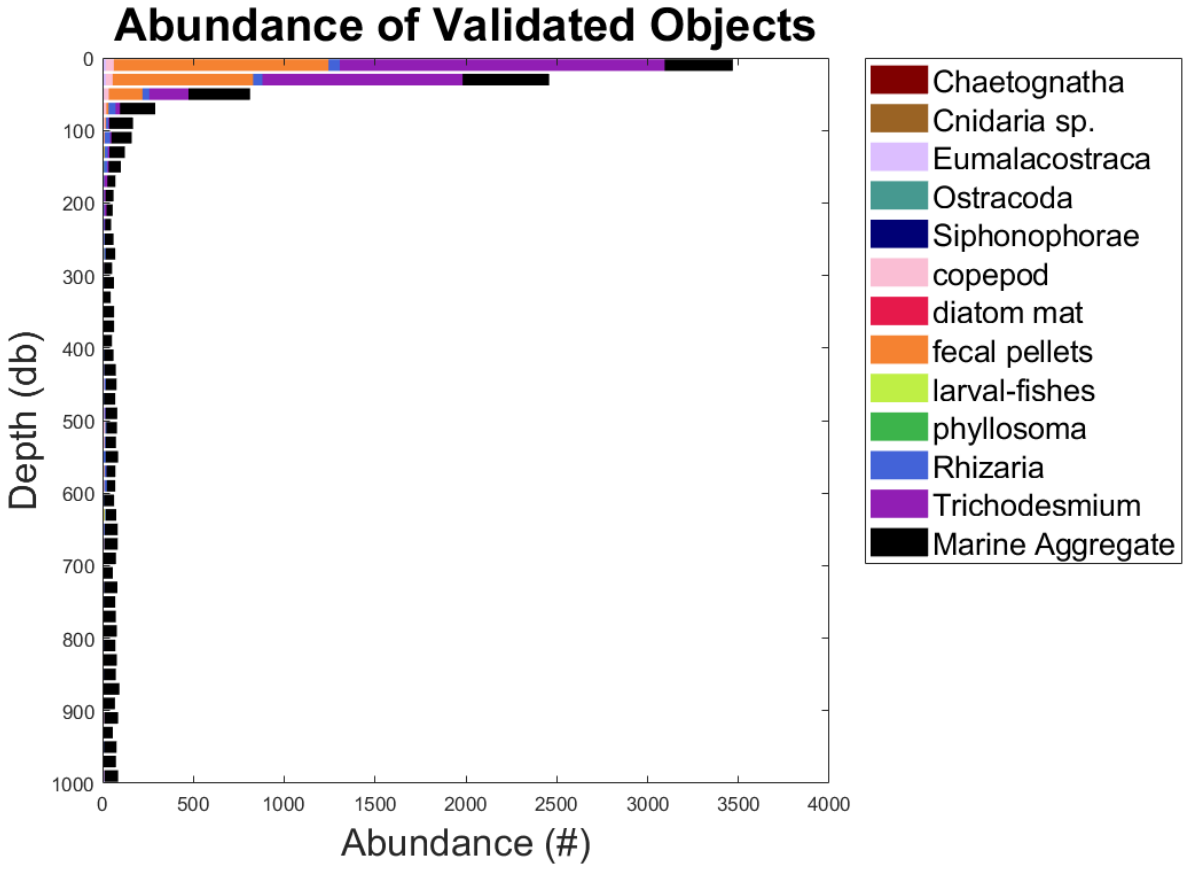


Figure 12. Abundance of Validated Objects

Abundance of Main set validated images shown over the top 1000 db of the water column, binned in 20 db intervals and collected during HOT-345 during October of 2023 at Station ALOHA. Images are separated by taxonomic category, with the majority of images validated in the upper 100 db.

Overall, most classified images belong to the marine aggregate category (Fig. 13). While the upper 50 db is comprised primarily of fecal pellets and the colony-forming cyanobacteria

Trichodesmium, all other depth intervals are dominated by marine aggregates. An increase in copepods relative to surface peaks in *Trichodesmium* and fecal pellets is evident in the twilight zone (Figure 14). Because the data here were not segregated by day and night cycle, it is possible that diel vertical migration (DVM) behavior could lead to these distributions; further partitioning of the datasets by photoperiod would need to be explored to address this pattern. The increase in the relative abundance of Rhizaria between 75-275 db could also be related to migratory behavior of Foraminifera, or perhaps be indicative of niche portioning of Rhizarian subgroups as was described in Biard & Ohman, (2020). Future UVP studies may correlate particle distributions, diel patterns, and higher phylogenetic specificity of Rhizarian populations than is described here to investigate their temporal variability.

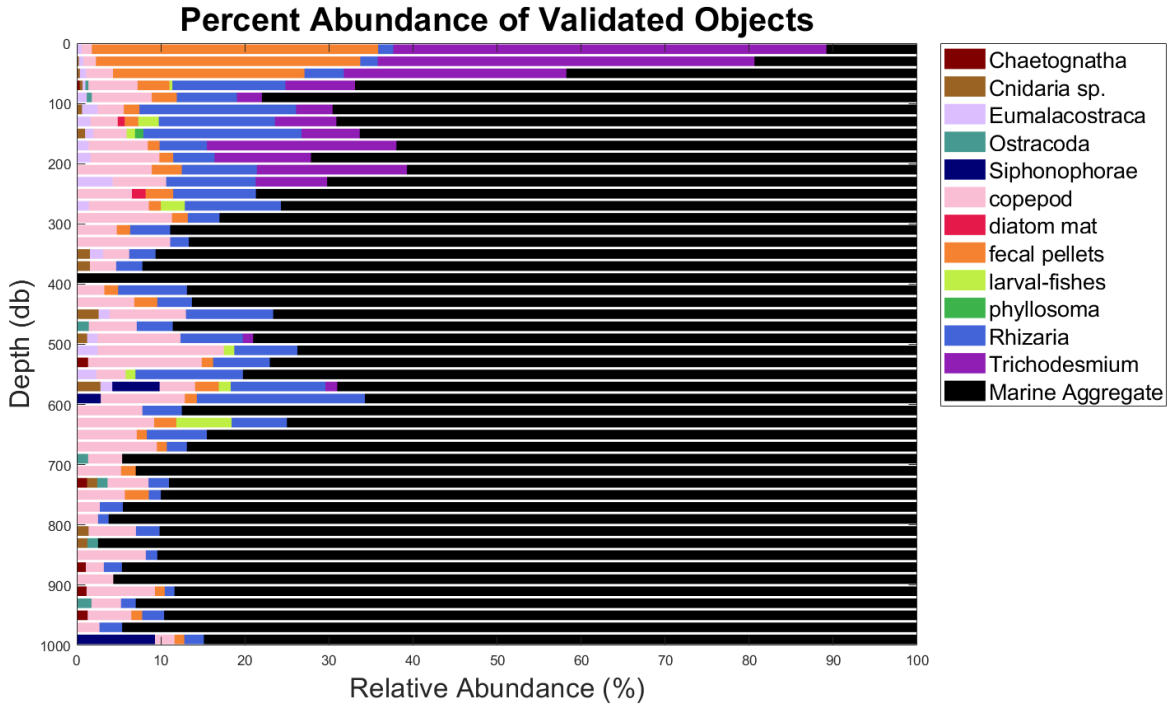


Figure 13. Percent Abundance of Validated Objects

Percent abundance of Main set validated images shown over the top 1000 db of the water column, binned in 20 db intervals and collected during HOT-345 during October of 2023 at Station ALOHA. Images are separated by taxonomic category, with marine aggregates dominating abundances below 100 db.

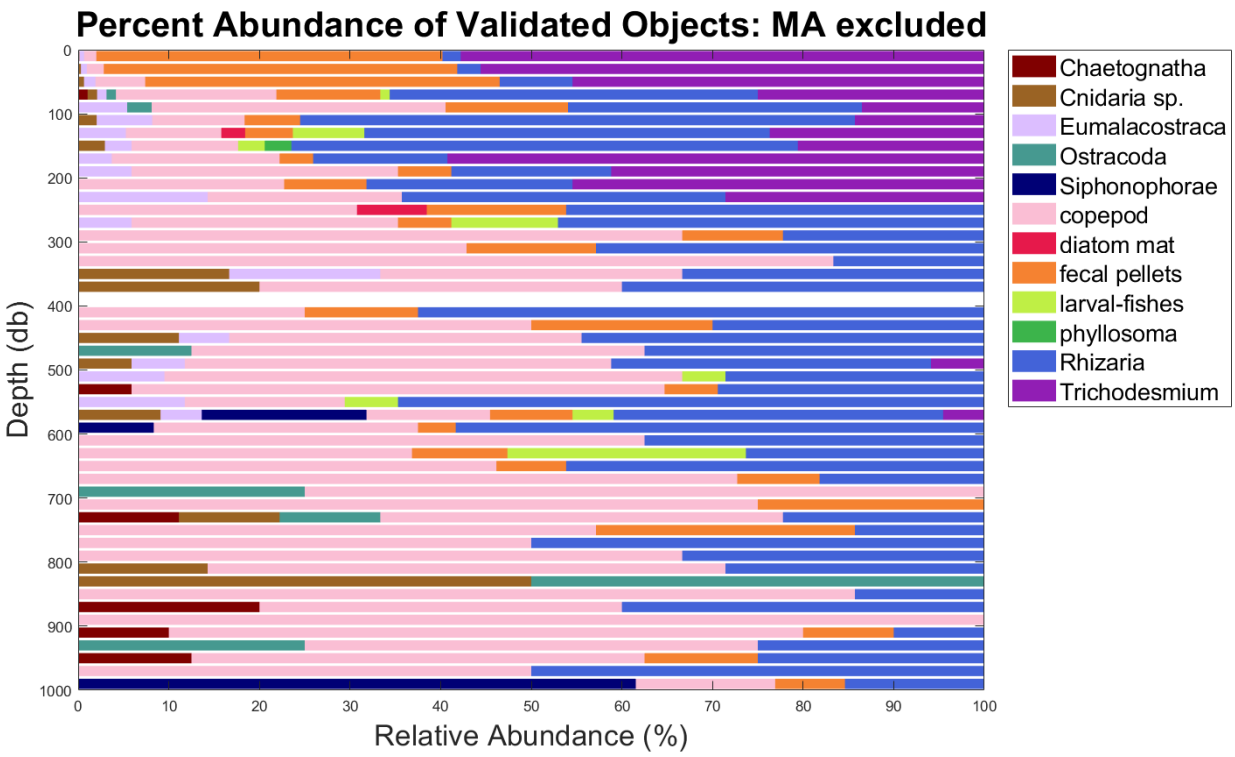


Figure 14. Percent Abundance of Validated Objects: Marine Aggregates Excluded

Abundance of Main set validated images, excluding marine aggregates, shown over the top 1000 db of the water column. The data were collected during HOT-345 during October of 2023 at Station ALOHA. Images are separated by taxonomic category and binned in 20 db intervals. While fecal pellets and *Trichodesmium* dominate the upper 100 db of the water column, a transition is then made with depth to an increased percent abundance of copepods and Rhizaria.

The ESD size ranges of the validated categories were plotted to visualize the categories that trend towards larger or smaller ESD sizes or that show a wide range of observed sizes (Figure 15). While detrital categories (fecal pellets and marine aggregates) were shifted towards a median of 1 mm, other objects such as Eumalacostraca and Larval-Fishes were spread more broadly into the 5 and 10 mm ESD sizes. However, almost all categories had median values below 5 mm with outliers in the 5 to 42 mm ESD range. Such an analysis could help to illuminate phenological size changes within each category across season, depth or environmental gradients.

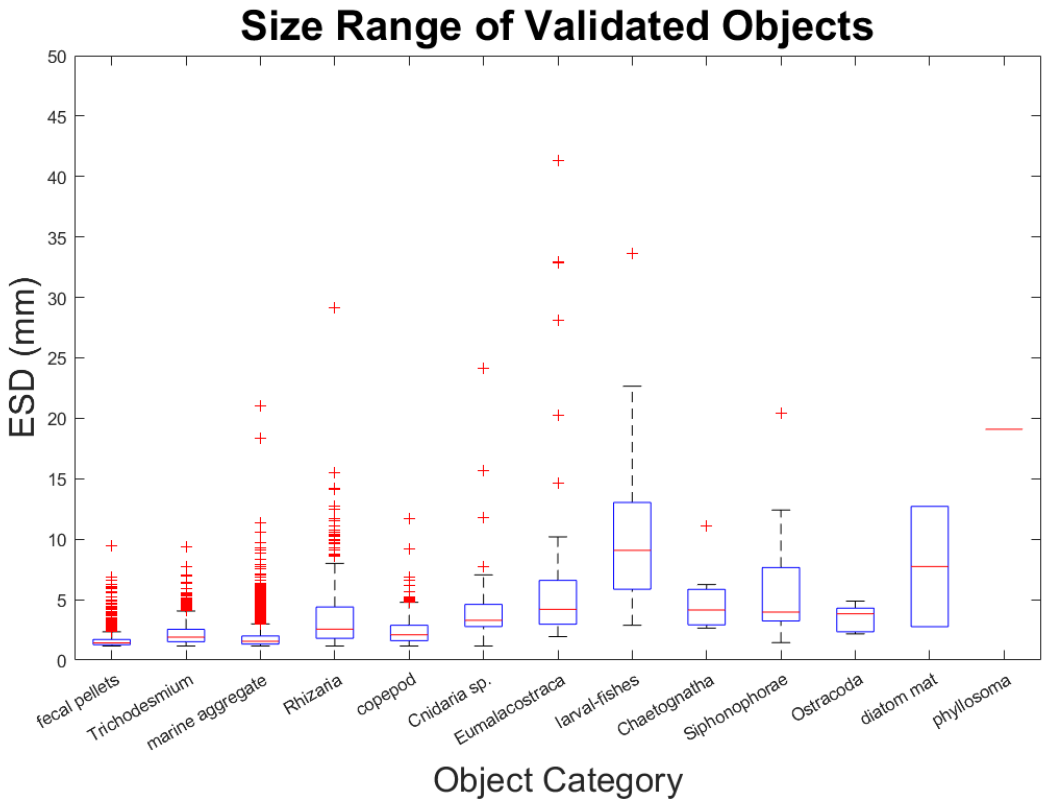


Figure 15. Size Range of Validated Objects

Box plot showing the size ranges of Main set validated objects collected during HOT-345 during October of 2023 at Station ALOHA. Most categories show median values below 5 mm ESD, with exceptions in the case of Larval-Fishes, Phyllosoma, and Diatom Mats. These larger categories found in the oligotrophic NPSG were also lower in abundance.

3.6 Individual Object Insights

To understand the vertical niches of major categories their abundances across all casts were enumerated and normalized by the volume of water which was collected in 15 db bins. This normalization helped to control for depths that were imaged more frequently. These values were presented as concentrations in units of counts per liter (#/L) (Figs. 16 – 18). Certain photosynthetic organisms, such as *Trichodesmium* (Fig. 16), were distributed as expected throughout the euphotic zone (at Station ALOHA top of nutricline ~91 db in winter, ~117 db in summer) with a sharp peak here seen within the mixed layer (<50 db) (Letelier et al., 2004). Some images of *Trichodesmium* were found much deeper, at times in the 200 db range. As the UVP cannot distinguish between living and non-living material, these organisms may be migrating, dead and sinking as clumps or have been subducted rapidly from the surface by

downwelling (White et al., 2006). Marine aggregates (Fig. 17) showed highest concentrations in the euphotic zone as well but peaked at the mixed layer boundary (~50 db). Peak concentrations of marine aggregates were more than double those for copepods and dominated the signal in the upper water column across all profiles, as is consistent with findings elsewhere (Trudnowska et al., 2021; Turner, 2015). As aggregates showed a peak at the mixed layer boundary, it is possible that aggregates accumulate here during settling, providing a region of nutrient-rich material for organismal particle trophy or a site for particle colonization by bacteria. Below this marine aggregate peak there was a notable depression in aggregate concentrations (Fig. 16), which is consistent with particle remineralization and disaggregation with increasing depth. However, the subsequent uptick in aggregate concentrations at the 350 db depth interval is currently unexplained but could be explained by a density feature. In this case the aggregates may become suspended on account of their neutral buoyancy at this depth. This may reveal a more complicated and changing story for 'sinking' vs. 'suspended' particle discussions based on size and warrants further investigation.

Lastly, the copepod vertical profile of abundance is shown in Figure 18. The low total counts of copepods below 50 db likely reflects real declines in animal abundance with depth. While it is unlikely that the UVP can be used to probe questions regarding copepod niche partitioning alone, on account of its taxonomic resolution, this instrument may be used in tandem with methods of physical collection that concentrate large volumes and reveal more morphotaxonomic detail. By comparing the high-resolution *in-situ* analysis of the UVP with more taxonomically specific methods of analysis, such as by net tow, VPR, or Zooscan image analyses, the true diversity underneath this vertical profile may be better interrogated.

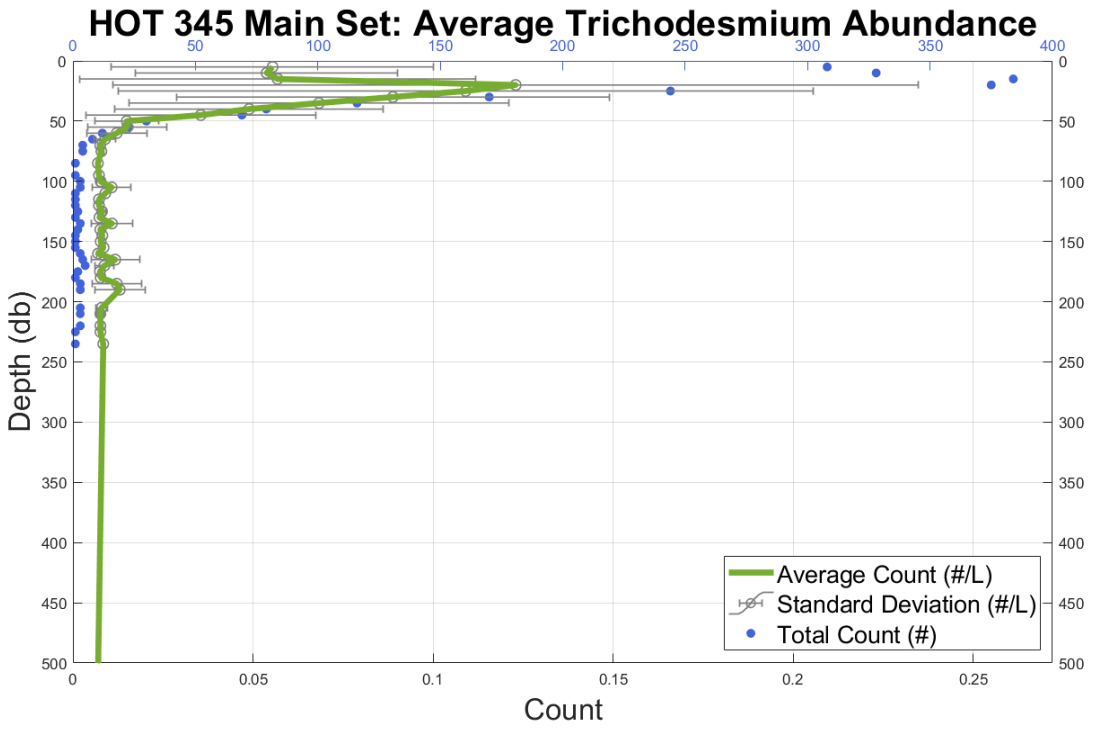


Figure 16. *Trichodesmium* Concentration

Mean concentration of *Trichodesmium* (green) and total *Trichodesmium* (blue) imaged by the UVP over 11 profiles, binned at 5m depth intervals down to 500db within the water column. Error bars (grey) represent 1 standard deviation from the mean. Data collected during HOT 345 during October of 2023 at Station ALOHA

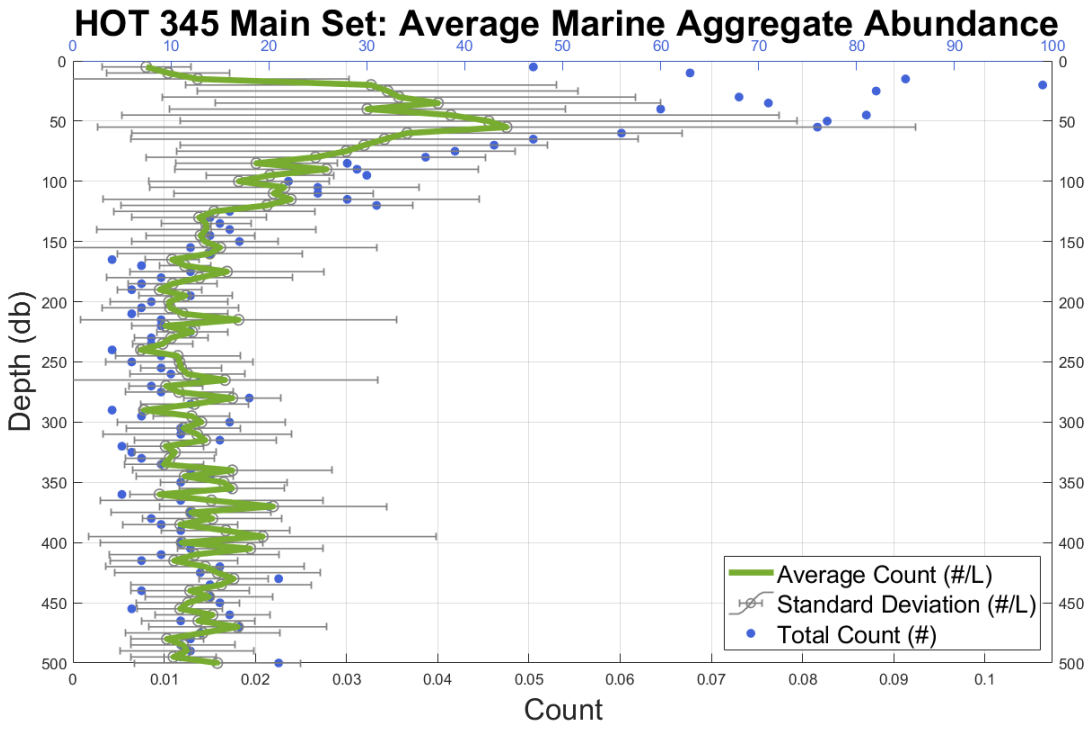


Figure 17. Marine Aggregate Concentration

Mean concentration of marine aggregates (green) and total marine aggregates (blue) imaged by the UVP over 11 profiles, binned at 5 m depth intervals down to 500 db within the water column. Error bars (grey) represent 1 standard deviation from the mean. Data collected during HOT 345 during October of 2023 at Station ALOHA.

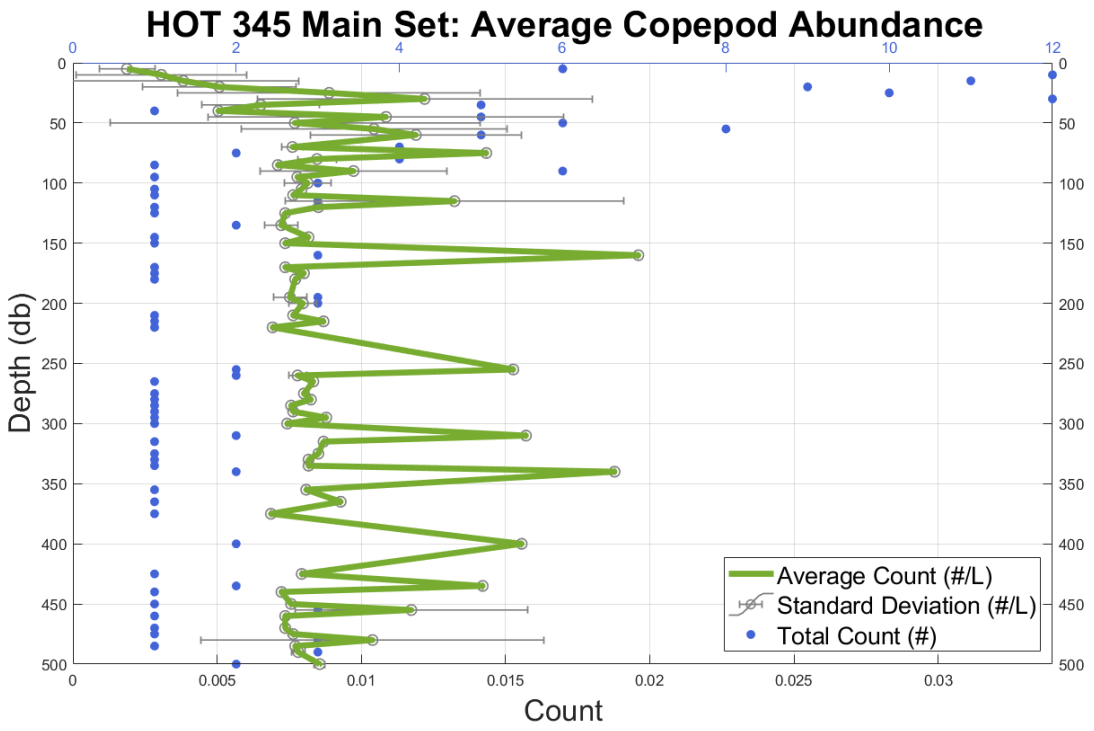


Figure 18. Copepod Concentration

Mean concentration of copepods (green) and total copepods (blue) imaged by the UVP over 11 profiles, binned at 5 m depth intervals down to 500 db within the water column. Error bars (grey) represent 1 standard deviation from the mean. Data collected during HOT 345 during October of 2023 at Station ALOHA

Within categories, mean object size can also change with depth. While Figure 15 shows the overall size range of object categories, Figures 19-21 extend this analysis with vertical resolution for individual objects. Marine aggregates largely do not deviate from the smaller 1-2 mm median size with depth (Fig. 19). However, marine aggregate outliers in the 3 – 10 mm range are more common in the euphotic zone, and within the mixed layer. This is likely a signature of large particle production in surface waters, and subsequent remineralization or disaggregation with depth. In the case of copepods (Fig. 20), there is a slight subsurface (~160 - 360 db) peak in organismal sizes. Additionally, Rhizaria (Fig. 21) show a slightly bimodal size profile with larger organisms being found in the surface and again around 280 db. The Rhizarian diversity includes large (a few micrometers to a maximum of three meters) light-dependent populations (e.g., collodaria), which may explain upper ocean size ranges (Biard et

al., 2017). Further, the secondary peak could be explained by Foraminifera, Phaeodaria, or other such larger migratory Rhizaria that have been shown to venture into the twilight zone (Biard & Ohman, 2020).

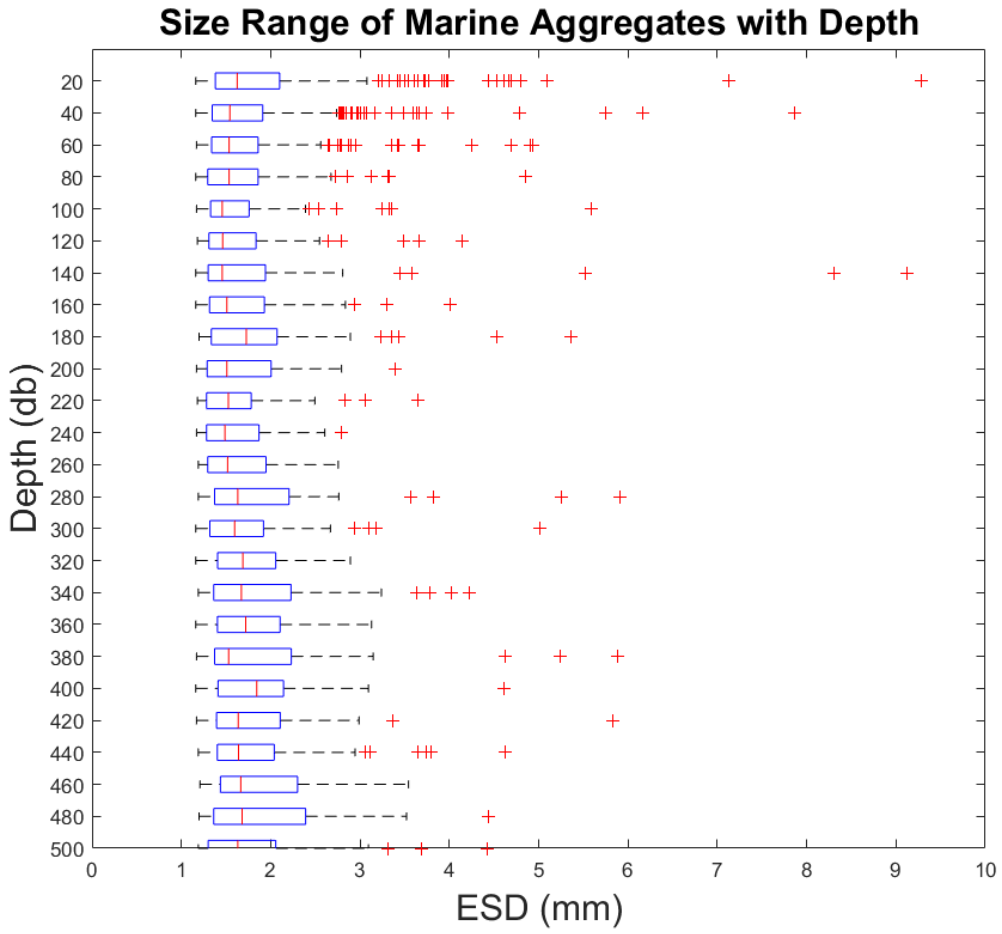


Figure 19. Size Range of Marine Aggregates with Depth

Box plot of Marine Aggregates validated in the Main set over the upper 500 db of the water column. The median size for MA remains relatively constant, but outliers in the upper maxima of reported ESDs are more common above 100 db. Data was collected during HOT 345 during October of 2023 at Station ALOHA.

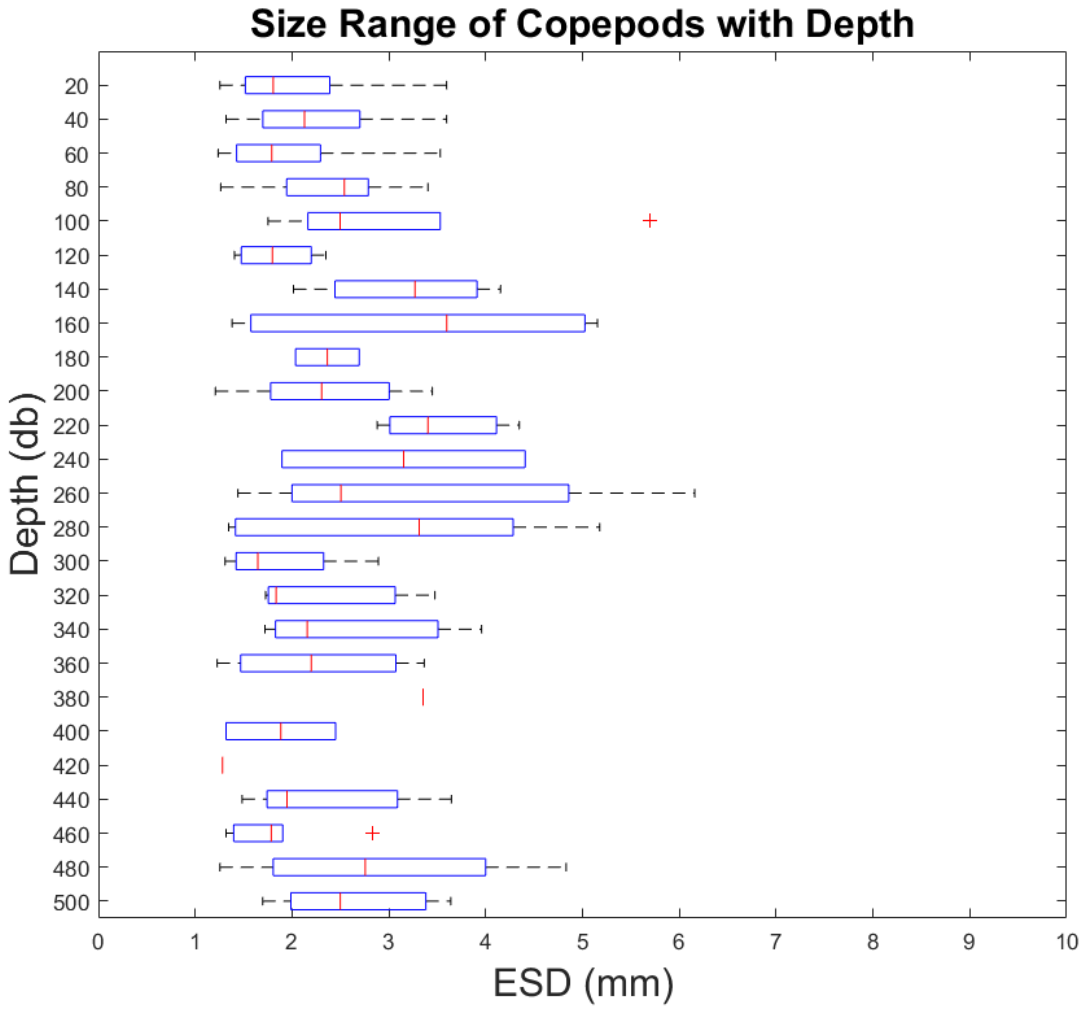


Figure 20. Size Range of Copepods with Depth

Box plot of Copepods validated in the Main set over the upper 500 db of the water column. The median size for copepods reaches a maximum in the twilight zone of the water column. Data was collected during HOT 345 during October of 2023 at Station ALOHA.

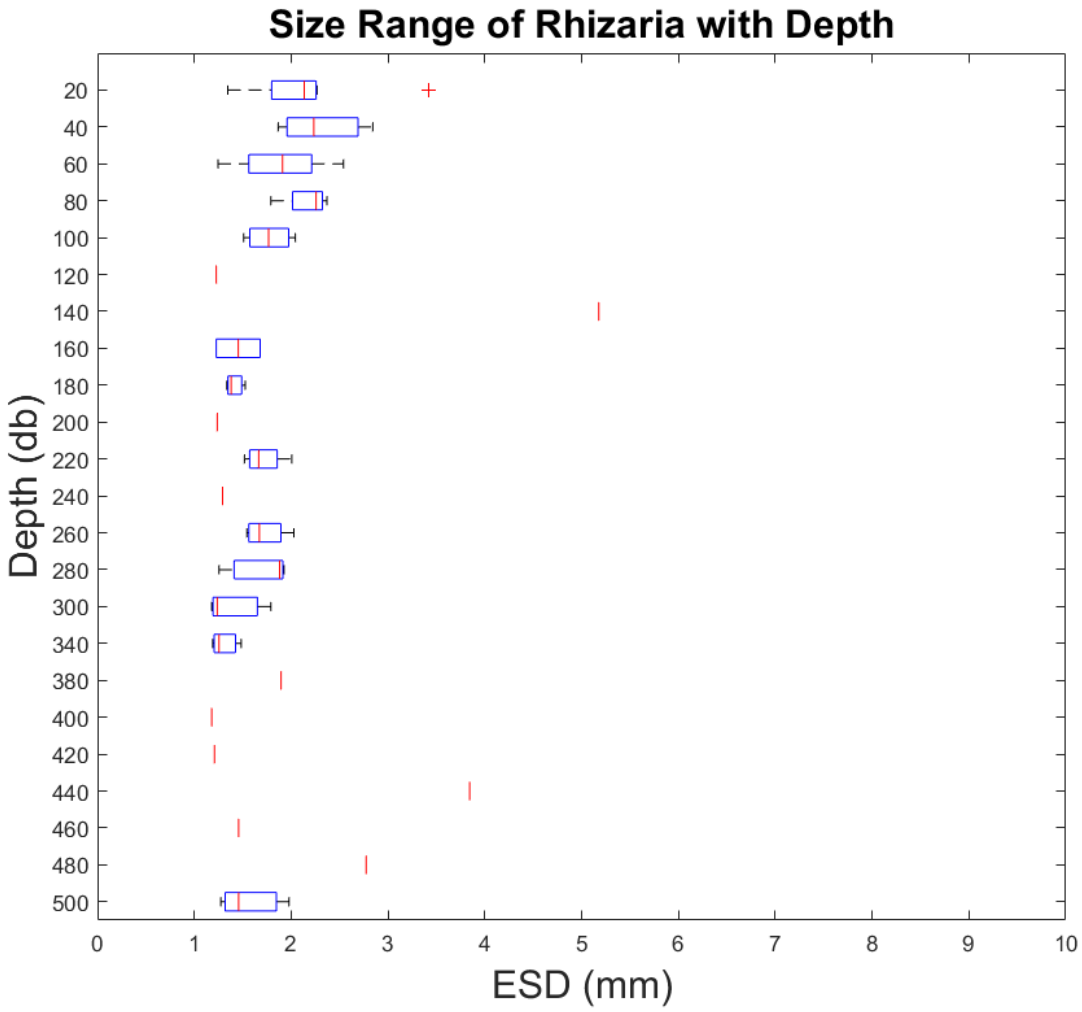


Figure 21. Size Range of Rhizaria with Depth

Box plot of Rhizaria validated in the Main set over the upper 500 db of the water column. The median size shows bimodal peaks around 60 db and 260 db and is perhaps indicative of underlying taxonomic complexities. Data was collected during HOT 345 during October of 2023 at Station ALOHA.

4. Conclusion

The Underwater Vision Profiler for Particle Imaging

Through analysis of images collected throughout the water column in the oligotrophic NPSG it is clear that the UVP 5 is proficient in the acquisition of images for populations of particles, particularly marine aggregates, that would otherwise be altered or undercounted by other methods of observation (Asper, 1987; Gibbs & Konwar, 1983; Honjo et al., 1984; Silver & Alldredge, 1981). By imaging a larger volume than that of other optical systems, performing such data collection *in situ*, and acquiring images with high depth resolution, the UVP is arguably the state of the art for morphological identification and enumeration of particles > 500 μm (Forest et al., 2012). However, due to the limitations of its imaging resolution, the non-orientation of particles, and the lack of compositional analysis inherent to a non-destructive method of collection, the UVP would benefit from comparisons to other measurements. Measurements from methods of physical collection, such as net tows, and higher resolution imaging, such as from the VPR and IFCB, would supplement the accuracy of morphological classification possible with the UVP. These instruments could also provide chlorophyll fluorescence, such as for the identification of diatom mats, and higher phylogenetic specificity, such as through DNA barcoding of collected organisms. Further, the ability to provide a unique perspective on the size and structure of large fragile particles and *in situ* enumeration of such populations makes the UVP an ideal instrument with much benefit to wholistic studies of organismal and marine aggregate size, diversity, and spatial distribution.

Findings from the Hawai'i Ocean Time-series

Particle data presented here show an increase in the abundance and variability of large particles during the summer-fall of 2020 to 2023, and a PSD slope of -3.33 ± 0.09 for HOT-345. These findings are in line with previously described timing of phytoplankton blooms in the NPSG and provide further evidence supporting the impact of the summer export pulse at Station ALOHA from the high-resolution perspective of the large particle size spectrum. Shallow PSD slopes (~ -2.80) seen in exclusively summer cruises further support the trend of increasing large particle stocks during the summer. Comparing the vertical profile of marine aggregate sizes described here at Station ALOHA to biogeochemical measurements routinely made during HOT cruises could provide further insight into the nature of large particle export. Particularly, the presence of subsurface spikes in large particle abundance, which may hint at an evolving definition of 'sinking' and 'suspended' particles dependent on density features and

disaggregation. Additionally, the changes to plankton community composition with depth, particularly increases in Rhizaria, here supports the need for deeper (>200 db) casts to illuminate open ocean organismal dynamics. Further, observations of deep water *Trichodesmium* may indicate migrating populations that could be investigated further through analysis of photoperiods, physical collection and speciation, and colony shape identification.

Classification Standards for Building and Validating Training Sets

With increasingly large UVP image data sets and the growing popularity of automated image classification it is necessary that there be a standardized onramp for new annotators to produce MLA training sets and validate MLA products. To harness the improving accuracy of contemporary MLAs and image classification efficiency that is beyond human annotation, the current schema of MLAs requires human annotators to establish contextualized training sets and validate MLA products. It has been acknowledged in recent works that more attention has been given to the education of annotators on how to use MLAs, while the process of curating the training sets and validating classifications which MLAs are reliant upon has been neglected. As such, with dwindling numbers of expert morphological taxonomists who could properly curate and validate datasets, there will become an increasingly tight bottleneck at the point of human annotation.

I present here a novel method that uses classification standards to build contextualized and consistent training sets, while providing criteria for the validation of MLA outputs. Classification standards provided increased inter- and intra-annotator consistency and laid a foundation for better communication of image validation criteria for annotator training, manuscripts, and MLA performance troubleshooting. When compared to the native EcoTaxa RF, annotators using the classifications standards were better able to reproduce the categorizations of objects made in the Main set. As such it is recommended that manual annotation with classification standards, following RF pre-binning, be used to provide the highest quality of training sets and most accurate validations of MLA outputs. While annotation agreement for certain categories fell below previously reported levels of expertise (~80%) as established by net tow and Zooscan studies, the argument is made that the resolution of UVP images may require re-assessment of such thresholds within the context of individual instrumentation (Culverhouse, 2003). Further, while inter- and intra-annotator consistency has been investigated in previous studies, though not for the UVP, it is here extended to cooperatively annotated datasets. Findings here imply that annotations made across multiple annotators can be comparable to self-consistency when using classification standards. This

provides both feasibility and a framework for long-term UVP image annotation projects such as by time series.

Applications to Future Studies

While stated previously above, it bears repeating in service of future efforts that a high resolution *in situ* particle imaging instrument such as the UVP is particularly suited to investigate the concentrations and characteristics of large particles in the oligotrophic NPSG, as shown here through various applications of UVP imaging data. However, the large volume of imaging data required to properly categorize the broad diversity, distribution and variability of plankton and marine aggregates in the open ocean requires an automated imaging scheme beyond that which is possible by human annotators alone. To make use of machine learning algorithms for this purpose, it is shown here how a new UVP user may develop contextualized training sets, validate images, and apply their results to oceanographic questions. Through use of classification standards future studies may implement time series observations of plankton and marine aggregates using the UVP, and more accurately and consistently describe the organisms and aggregates which constitute the biological carbon pump.

5. Literature Cited

Alldredge, A. L. (1976). Discarded appendicularian houses as sources of food, surface habitats, and particulate organic matter in planktonic environments. *Limnology and Oceanography*, *21*(1), 14–24. https://doi.org/10.4319/lo.1976.21.1.0014

Alldredge, A. L., & Cohen, Y. (1987). Can microscale chemical patches persist in the sea? Microelectrode study of marine snow, fecal pellets. *Science*, *235*, 689–691. https://doi.org/10.1126/science.235.4789.689

Alldredge, A. L., & Silver, M. W. (1988). Characteristics, dynamics and significance of marine snow. *Progress in Oceanography*, *20*(1), 41–82. https://doi.org/10.1016/0079-6611(88)90053-5

Alldredge, A. L., & Youngbluth, M. J. (1985). The significance of macroscopic aggregates (marine snow) as sites for heterotrophic bacterial production in the mesopelagic zone of the subtropical Atlantic. *Deep Sea Research Part A, Oceanographic Research Papers*, *32*(12), 1445–1456. https://doi.org/10.1016/0198-0149(85)90096-2

Alldredge, A. L., Cole, J. J., & Caron, D. A. (1986). Production of heterotrophic bacteria inhabiting macroscopic organic aggregates (marine snow) from surface waters. *Limnology and Oceanography*, *31*(1), 68–78. https://doi.org/10.4319/lo.1986.31.1.0068

Alldredge, A. L., Passow, U., & Haddock, S. H. D. (1998). The characteristics and transparent exopolymer particle (TEP) content of marine snow formed from thecate dinoflagellates. *Journal of Plankton Research*, *20*(3), 393–406. https://doi.org/10.1093/plankt/20.3.393

Andrews, C. C., Karl, D. M., Small, L. F., & Fowler, S. W. (1984). Metabolic activity and bioluminescence of oceanic faecal pellets and sediment trap particles. *Nature*, *307*, 539–541. https://doi.org/10.1038/307539a0

Asper, V. L. (1985). Accelerated settling of particulate matter by 'marine snow' aggregates [Doctoral thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution]. Woods Hole Open Access Server. https://doi.org/10.1575/1912/3367

Asper, V. L. (1987). Measuring the flux and sinking speed of marine snow aggregates. *Deep Sea Research Part A, Oceanographic Research Papers*, *34*(1), 1–17. https://doi.org/10.1016/0198-0149(87)90117-8

Axler, K. E. (2020). Fine-scale larval fish distributions and predator prey dynamics in a coastal river-dominated ecosystem. *Marine Ecology Progress Series*, (650), 37–61.

Barth, A., & Stone, J. (2022). Comparison of an In Situ Imaging Device and Net-Based Method to Study Mesozooplankton Communities in an Oligotrophic System. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.898057

Biard, T., & Ohman, M. D. (2020). Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone revealed by in situ imaging. *Limnology and Oceanography*, *65*(11), 2583–2602. https://doi.org/10.1002/lno.11472

Biard, T., Bigeard, E., Audic, S., Poulain, J., Gutierrez-Rodriguez, A., Pesant, S., et al. (2017). Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. *ISME Journal*, *11*(6), 1331–1344. https://doi.org/10.1038/ismej.2017.12

Bisson, K. M., Kiko, R., Siegel, D., Guidi, L., Picheral, M., & Boss, E. (2022). Sampling uncertainties of particle size distributions and derived fluxes. *Global Biogeochemical Cycles*, *20*(12), 754–767. https://doi.org/10.1002/essoar.10508460.1

Blackburn, N., Fenchel, T., & Mitchell, J. (1998). Microscale nutrient patches in planktonic habitats shown by chemotactic bacteria. *Science*, *282*(5397), 2254–2256. https://doi.org/10.1126/science.282.5397.2254

Buesseler, K. O., Andrews, J. A., Hartman, M. C., Belastock, R., & Chai, F. (1995). Regional estimates of the export flux of particulate organic carbon derived from thorium-234 during the JGOFS EqPac program. *Deep-Sea Research Part II*, *42*(2–3), 777–804. https://doi.org/10.1016/0967-0645(95)00043-P

Cheng, K., Cheng, X., Wang, Y., Bi, H., & Benfield, M. C. (2019). Enhanced convolutional neural network for plankton identification and enumeration. *PLoS ONE*, *14*(7): e0219570. https://doi.org/10.1371/journal.pone.0219570

Capotondi, A., Alexander, M. A., Bond, N. A., Curchitser, E. N., & Scott, J. D. (2012). Enhanced upper ocean stratification with climate change in the CMIP3 models. *Journal of Geophysical Research: Oceans*, *117*(4): C04031. https://doi.org/10.1029/2011JC007409

Culverhouse, Philip F., Williams, R., Reguera, B., Herry, V., & González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, *247*, 17–25. https://doi.org/10.3354/meps247017

Culverhouse, Philip F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M., & Picheral, M. (2014). An empirical assessment of the consistency of taxonomic identifications. *Marine Biology Research*, *10*(1), 73–84. https://doi.org/10.1080/17451000.2013.810762

Davis, C. S., Gallager, S. M., Berman, M. S., & Haury, L. R. and Strickler, J. R. (1992). The Video Plankton Recorder (VPR): Design and initial results. *Archiv fur Hydrobiologie, Beihefte, Ergebnisse der Limnologie*, *36* 67–81. https://www.researchgate.net/publication/284686405

Dolan, J. R. (2021). Pioneers of plankton research: Victor Hensen (1835-1924). *Journal of Plankton Research*, *43*(4), 507–510. https://doi.org/10.1093/plankt/fbab045

Emerson, S., Quay, P., Karl, D., Winn, C., Tupas, L., & Landry, M. (1997). Experimental determination of the organic carbon flux from open-ocean surface waters. *Nature*, *389*(6654), 951–954. https://doi.org/10.1038/40111

Fellows, D. A., Karl, D. M., & Knauer, G. A. (1981). Large particle fluxes and the vertical transport of living carbon in the upper 1500 m of the northeast Pacific Ocean. *Deep Sea Research Part A, Oceanographic Research Papers*, *28*(9), 921–936. https://doi.org/10.1016/0198-0149(81)90010-8

Forest, A., Stemmann, L., Picheral, M., Burdorf, L., Robert, D., Fortier, L., & Babin, M. (2012). Size distribution of particles and zooplankton across the shelf-basin system in southeast Beaufort Sea: Combined results from an Underwater Vision Profiler and vertical net tows. *Biogeosciences*, *9*(4), 1301–1320. https://doi.org/10.5194/bg-9-1301-2012

Fowler, S. W., & Knauer, G. A. (1986). Role of large particles in the transport of elements and organic compounds through the oceanic water column. *Progress in Oceanography*, *16*(3), 147–194. https://doi.org/10.1016/0079-6611(86)90032-7

Gibbs, R. J., & Konwar, L. N. (1983). Sampling of Mineral Floes Using Niskin Bottles. *Environmental Science and Technology*, *17*(6), 374–375. https://doi.org/10.1021/es00112a014

González, P., Álvarez, E., Díez, J., López-Urrutia, Á., & del Coz, J. J. (2017). Validation methods for plankton image classification systems. *Limnology and Oceanography: Methods*, *15*(3), 221–237. https://doi.org/10.1002/lom3.10151

Guidi, L., Jackson, G. A., Stemmann, L., Miquel, J. C., Picheral, M., & Gorsky, G. (2008). Relationship between particle size distribution and flux in the mesopelagic zone. *Deep-Sea Research Part I: Oceanographic Research Papers*, *55*(10), 1364–1374. https://doi.org/10.1016/j.dsr.2008.05.014

Guidi, L., Calil, P. H. R., Duhamel, S., Björkman, K. M., Doney, S. C., Jackson, G. A., et al. (2012). Does eddy-eddy interaction control surface phytoplankton distribution and carbon export in the North Pacific Subtropical Gyre? *Journal of Geophysical Research: Biogeosciences*, *117*(2). https://doi.org/10.1029/2012JG001984

Hebert, P. D. N., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321.

Hensen, V. (1891). Die Plankton-Expedition und Haeckel's Darwinismus: ueber einige Aufgaben und Ziele der beschreibenden Naturwissenschaften. Lipsius & Tischer.

Honjo, S., Doherty, K. W., Agrawal, Y. C., & Asper, V. L. (1984). Direct optical assessment of large amorphous aggregates (marine snow) in the deep ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, *31*(1), 67–76. https://doi.org/10.1016/0198-0149(84)90073-6

Honjo, Susumu. (1980). Material fluxes and modes of sedimentation in the mesopelagic and bathy pelagic zones. *Journal of Marine Research*, *38*(1).

Jackson, G. A. (1995). Coagulation of Marine Algae. In *Aquatic Chemistry* (Vol. 244, pp. 203–217). American Chemical Society. https://doi.org/10.1021/ba-1995-0244.ch010

Jackson, G. A., & Burd, A. B. (1998). Aggregation in the marine environment. *Environmental Science and Technology*, *32*(19), 2805–2814. https://doi.org/10.1021/es980251w

Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M., & Mahaffey, C. (2012). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(6), 1842–1849. https://doi.org/10.1073/pnas.1120312109

Karl, D. M., Letelier, R. M., Bidigare, R. R., Björkman, K. M., Church, M. J., Dore, J. E., & White, A. E. (2021). Seasonal-to-decadal scale variability in primary production and particulate matter export at Station ALOHA. *Progress in Oceanography*, *195*. https://doi.org/10.1016/j.pocean.2021.102563

Kenitz, K., Orenstein, E. (2020). Environmental drivers of population variability in colony-forming marine diatoms. *Limnology and Oceanography*, *65*(10), 2515–2528.

Kenitz, K., Orenstein, E., Anderson, C. R., Barth, A. J., Briseño-Avena, C., Caron, D. A., Carter, M. L., Eggleston, E., Franks, P. J. S., Fumo, J. T., Jaffe, J. S., McBeain, K. A., Odell, A., Seech, K., Shipe, R., Smith, J., Taniguchi, D. A. A., Venrick, E. L., & Barton, A. D. (2023). Convening Expert Taxonomists to Build Image Libraries for Training Automated Classifiers. *Limnology and Oceanography Bulletin*, *32*(3), 89–97. https://doi.org/10.1002/lob.10584

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., et al. (2022). A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5. *Earth System Science Data*, *14*(9), 4315–4337. https://doi.org/10.5194/essd-14-4315-2022

Lee, H., Park, M., & Kim, J. (2016). Plankton classification on imbalanced large-scale database via convolutional neural networks with transfer learning. *2016 IEEE International Conference on Image Processing (ICIP),* Pheonix, AZ, USA, pp. 3713-3717. doi: 10.1109/ICIP.2016.7533053

Letelier, R. M., Karl, D. M., Abbott, M. R., & Bidigare, R. R. (2004). Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. *Limnology and Oceanography*, *49*(2), 508–519. https://doi.org/10.4319/lo.2004.49.2.0508

Letelier, R. M., Dore, J. E., Winn, C. D., & Karl, D. M. (1996). Seasonal and interannual variations in photosynthetic carbon assimilation at station ALOHA. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *43*(2–3), 467–490. https://doi.org/10.1016/0967-0645(96)00006-9

Luo, J. Y., Irisson, J. O., Graham, B., Guigand, C., Sarafraz, A., Mader, C., & Cowen, R. K. (2018). Automated plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, *16*(12), 814–827. https://doi.org/10.1002/lom3.10285

McCave, I. N. (1984). Size spectra and aggregation of suspended particles in the deep ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, *31*(4), 329–352. https://doi.org/10.1016/0198-0149(84)90088-8

Ohman, M. D., Davis, R. E., Sherman, J. T., Grindley, K. R., Whitmore, B. M., Nickels, C. F., & Ellen, J. S. (2019). Zooglider: An autonomous vehicle for optical and acoustic sensing of zooplankton. *Limnology and Oceanography: Methods*, *17*(1), 69–86. https://doi.org/10.1002/lom3.10301

Passow, U., Shipe, R. F., Murray, A., Pak, D. K., Brzezinski, M. A., & Alldredge, A. L. (2001). The origin of transparent exopolymer particles (TEP) and their role in the sedimentation of particulate matter. *Continental Shelf Research*, *21*(4), 327–346. https://doi.org/10.1016/S0278-4343(00)00101-1

Pei, Y., Huang, Y., Zou, Q., Zhang, X., & Wang, S. (2021). Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(4), 1239–1253. https://doi.org/10.1109/TPAMI.2019.2950923

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(SEPT), 462–473. https://doi.org/10.4319/lom.2010.8.462

Prezelin, B. B., & Alldredge, A. L. (1983). Primary production of marine snow during and after an upwelling event. *Limnology and Oceanography*, *28*(6), 1156–1167. https://doi.org/10.4319/lo.1983.28.6.1156

Py, O., Hong, H., & Zhongzhi, S. (2016). Plankton classification with deep convolutional neural networks. In *Proceedings of 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2016*. https://doi.org/10.1109/ITNEC.2016.7560334

Rangineni, S. (2023). An Analysis of Data Quality Requirements for Machine Learning Development Pipelines Frameworks. *International Journal of Computer Trends and Technology*, *71*(8), 16–27. https://doi.org/10.14445/22312803/ijctt-v71i8p103

Remsen, A., Hopkins, T. L., & Samson, S. (2004). What you see is not what you catch: A comparison of concurrently collected net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder data from the northeast Gulf of Mexico. *Deep-Sea Research Part I: Oceanographic Research Papers*, *51*(1), 129–151. https://doi.org/10.1016/j.dsr.2003.09.008

Schröder, S. M., Kiko, R., & Koch, R. (2020). Morphocluster: Efficient annotation of Plankton images by clustering. *Sensors (Switzerland)*, *20*(11), 3060. https://doi.org/10.3390/s20113060

Shanks, A. L., & Trent, J. D. (1979). Marine snow: Microscale nutrient patches. *Limnology and Oceanography*, *24*(5), 850–854. https://doi.org/10.4319/lo.1979.24.5.0850

Sheldon, R. W., Prakash, A., & Sutcliffe, W. H. (1972). The size distribution of particles in the ocean. *Limnology and Oceanography*, *17*(3), 327–340. https://doi.org/https://doi.org/10.4319/lo.1972.17.3.0327

Silver, M. W., & Alldredge, A. L. (1981). Bathypelagic marine snow: deep-sea algal and detrital community. *Journal of Marine Research*, *39*(3), 501–530. https://elischolar.library.yale.edu/journal_of_marine_research/1556

Silver, M. W., Shanks, A. L., & Trent, J. D. (1978). Marine snow: Microplankton habitat and source of small-scale patchiness in pelagic populations. *Science*, *201*(4353), 371–373. https://doi.org/10.1126/science.201.4353.371

Sosik, H. M., Olson, R. J. (2007). A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot. *Limnology and Oceanography: Methods*, *5*(6), 195–203. https://www.researchgate.net/publication/280686102

Stemmann, L., Eloire, D., Sciandra, A., Jackson, G. A., Guidi, L., Picheral, M., & Gorsky, G. (2008). Volume distribution for particles between 3.5 to 2000 μm in the upper 200 m region of the South Pacific Gyre. *Biogeosciences*, *5*(2), 299–310. https://doi.org/10.5194/bg-5-299-2008

Trudnowska, E., Lacour, L., Ardyna, M., Rogge, A., Irisson, J. O., Waite, A. M., et al. (2021). Marine snow morphology illuminates the evolution of phytoplankton blooms and determines their subsequent vertical export. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22994-4

Turner, J. (2015). Investigating Marine Particle Distributions and Processes Using In Situ Optical Imaging In The Gulf Of Alaska [Doctoral thesis, University of Alaska, Fairbanks]. http://hdl.handle.net/11122/6406

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. *The Carbon Cycle and Atmospheric CO*, *32*(1), 99–110. https://doi.org/10.1029/gm032p0099

White, A. E., Spitz, Y. H., & Letelier, R. M. (2006). Modeling carbohydrate ballasting by Trichodesmium spp. *Marine Ecology Progress Series*, *323*, 35–45. https://doi.org/10.3354/meps323035

White, A. E., Spitz, Y. H., & Letelier, R. M. (2007). What factors are driving summer phytoplankton blooms in the North Pacific Subtropical Gyre? *Journal of Geophysical Research: Oceans*, *112*(12): C12006. https://doi.org/10.1029/2007JC004129

Whitmore, B. M., & Ohman, M. D. (2021). Zooglider-measured association of zooplankton with the fine-scale vertical prey field. *Limnology and Oceanography*, *66*(10), 3811–3827. https://doi.org/10.1002/lno.11920