https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Open Access Earth System Science Data Discussions

A global consistent database of plankton and detritus from in situ imaging by the Underwater Vision Profiler 5

Ariadna C. Nocera[1,2], Lars Stemmann[2], Marcel Babin[3], Tristan Biard[4], Julie Coustenoble[2], François Carlotti[5], Laurent Coppola[2,6], Lucas Courchet[7], Laetitia Drago[2,a], Amanda Elineau[2], Lionel Guidi[2], Helena Hauss[8,9], Laëtitia Jalabert[7], Lee Karp-Boss[10], Rainer Kiko[2,8], Manon Laget[4,b], Fabien Lombard[2], Andrew McDonnell[11], Camille Merland[7], Solène Motreuil[7], Thelma Panaïotis[2,c], Marc Picheral[2], Andreas Rogge[12], Anya Waite[13], Jean-Olivier Irisson[2]

[1]Centro para el Estudio de Sistemas Marinos, CCT-CENPAT-CONICET, U9120ACD, Puerto Madryn, Chubut, Argentina
[2]Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, LOV, F-06230 Villefranche-sur-Mer, France
[3]Département de Biologie, Université Laval, Québec, Canada
[4]Laboratoire d'Océanologie et de Géosciences (LOG), Université du Littoral Côte d'Opale, Université Lille, CNRS, IRD, UMR 8187, Wimereux, France
[5]Mediterranean Institute of Oceanography, Aix-Marseille Université, Université de Toulon, CNRS, IRD, UMR 7294, Marseille, France
[6]Sorbonne Université, CNRS, OSU STAMAR, UAR2017, 4 Place Jussieu, 75252 Paris CEDEX 05, France
[7]Sorbonne Université, CNRS, Institut de la Mer de Villefranche, IMEV, F-06230 Villefranche-sur-Mer, France
[8]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[9]NORCE Norwegian Research Centre, Bergen, Norway
[10]School of Marine Sciences, University of Maine, Orono, ME, USA
[11]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, Alaska, USA
[12]Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
[13]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada

[a]current address: Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN-IPSL), Sorbonne Université, UMR 7159 CNRS-IRD-MNHN, Paris, France
[b]current address: Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
[c]current address: National Oceanography Centre, Southampton, UK

Correspondence to: Ariadna C. Nocera (anocera@cenpat-conicet.gob.ar) and Lars Stemmann (lars.stemmann@imev-mer.fr)

Abstract. Plankton and detritus are essential components of the Earth's oceans influencing biogeochemical cycles and carbon sequestration. Climate change impacts their composition and marine ecosystems as a whole. To improve our understanding of these changes, standardized observation methods and integrated global datasets are needed to enhance the accuracy of ecological and climate models. Here, we present a global dataset for plankton and detritus obtained by two versions of the Underwater Vision Profiler 5 (UVP5). This release contains the images classified in 33 homogenized categories, as well as the metadata associated with them, reaching 3,114 profiles and ca. 8 million objects acquired between 2008-2018 at global scale. The geographical distribution of the dataset is unbalanced, with the Equatorial region (30° S - 30° N) being the most represented, followed by the high latitudes in the northern hemisphere and lastly the high latitudes in the Southern Hemisphere. Detritus is the most abundant category in terms of concentration (90%) and biovolume (95%), although its classification in different morphotypes is still not well established. Copepoda was the most abundant taxa within the plankton, with *Trichodesmium* colonies being the second most abundant. The two versions of UVP5 (SD and HD) have different imagers, resulting in a different effective size range to analyse plankton and detritus from the images (HD objects >600 µm, SD objects >1 mm) and morphological

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

properties (grey levels, etc.) presenting similar patterns, although the ranges may differ. Therefore, recommendations are provided for the appropriate use of this data when conducting studies. A large number of images of plankton and detritus will be collected in the future by the UVP5, and the public availability of this dataset will help it being utilized as a training set for machine learning and being improved by the scientific community. This will reduce uncertainty by classifying previously unclassified objects and expand the classification categories, ultimately enhancing biodiversity quantification. The dataset that constitutes this first release is available at SEANOE.

Key words: plankton, images, Underwater Vision Profiler, Global Ocean

1 Introduction

Plankton and particulate matter play a crucial role in natural biogeochemical cycles and provide essential ecosystem services, including carbon sequestration, nutrient cycling, and primary food source for various marine organisms (Turner, 2015; Boyd et al., 2019; Stemmann and Boss, 2012; Steinberg and Laundry, 2017). These components of the marine environment are significantly influenced by climate change, which alters ocean temperature, acidity, oxygenation, and circulation patterns (Doney et al., 2012; Constable et al., 2014). Such changes may, in turn, impact plankton composition by affecting their growth and survival (Hays et al., 2005; Dam and Baumann, 2017; Yebra et al., 2022), which will impact the production and composition of their solid waste, hereafter referred as detritus, with potential consequences for elemental biogeochemical cycles and deep-sea ecosystems (Du Pontavice et al., 2020). Zooplankton biomass and diversity, as well as bulk particulate matter, were identified as essential ocean, biodiversity, and climate variables by the Global Ocean Observing System (Chiba et al., 2018; Batten et al., 2019).

The size range of plankton and detritus spans from pico- to macro-sizes, influencing the methods required for their observation. Different technologies, including microscopy, optical imaging, sediment traps, and remote sensing, are employed depending on the target size class (Karsenti et al., 2011; Davis et al., 2005; Stemmann et al., 2008; Möller et al., 2012; Siegel et al., 2024). Many planktonic organisms and detritus are fragile and their abundances decrease with their size, making their collection and analysis challenging (Stemmann et al., 2008; Soviadan et al., 2024; Atherden et al., 2024). Consequently, obtaining homogeneous global datasets remains difficult notably if different instruments that measure various properties are used (Moriarty and O'brien, 2013). Furthermore, plankton exhibits a patchy distribution in the ocean due to physical, chemical, and biological processes (Suthers et al., 2019). These constraints demand increased sampling efforts and the development of standardized methodologies to enhance comparability of global plankton and detritus observations. For plankton, quantitative imaging has been identified as the best means to do so (Lombard et al 2019). Addressing these challenges is critical for improving our understanding of oceanic carbon cycling and ecosystem responses to environmental change.

According to specific objectives, biogeochemists focus on biomass as a measure of carbon content, while ecologists are primarily interested in community composition and biodiversity. Conventional sampling methods such as sediment traps, nets, pumps, and bottles further contribute to these differences, as they may not provide data that is directly comparable across disciplines. Modelers, on the other hand, require integrative datasets that bridge these gaps,

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

enabling accurate simulations of biogeochemical cycles, ecosystem dynamics, and carbon fluxes (Ratnarajah et al., 2023). They are particularly interested in harmonized data formats, conversion factors between abundance and biomass, and standardized methodologies to improve the predictive power of ecological and climate models. Addressing these disparities is essential for advancing interdisciplinary research and improving our understanding of oceanic processes.

Predicting the future state of plankton biodiversity and/or biomass with habitat or biogeochemical models is difficult and requires data obtained synoptically and regularly. Models could be partially fed by Continuous Plankton Recorder surveys (Batten et al., 2019) and satellite programs, such as the recently launched PACE (Gorman et al., 2019), but in both cases data on plankton is limited to the surface layer. Deep collection by net tows followed by the identification with microscopes provides the classical taxonomic identification, but this sampling methodology destroys fragile organisms and requires dedicated ship time (Calbet, 2024; Soviadan et al., 2024; Giering et al., 2022). In addition, this classical classification is time-consuming and susceptible to human bias, which can hinder effective quality control (Goswami, 2004). Sensors mounted on conventional in situ platforms may provide better spatial resolution and have less impact on the observed target. In situ acoustics with broadband systems allow for broad taxonomic identification but are not adapted for detritus due to their poor scattering. In situ imaging methods are better suited to recognize plankton and detritus, but detect them in a smaller volume. Considering the advantages and limitations of each method, we can conclude that in situ imaging is particularly suited for observing plankton and detritus, using their shape to infer taxonomy or morphological attributes (Picheral et al., 2010; Stemmann et al., 2012; Lombard et al., 2019). However, there is a compromise between the observed volume and the resolution of the camera that defines a size range and taxonomic resolution for each type of sensor. Recent publications have suggested (Lombard et al., 2019, Kiko et al., 2023) and showcased (Drago et al. 2022, Clements et al. 2023, Laget et al. 2024) that cooperative observation of plankton with imaging systems can allow an upscaling of regional observations by independent observers to global scale. Thus, plans are now underway to measure these variables with imaging systems on large-scale observing programs, like the Bio-GO-SHIP (Clayton et al., 2022) or the BGC-Argo (Claustre et al., 2020; Picheral et al., 2022) programs.

The Underwater Vision Profiler camera system (UVP, version 5) is an optical imaging instrument that captures digital images of a defined volume of seawater, provisioned with LED lights. It was designed to automatically detect, size, and count biotic and abiotic particles in the marine environment as it profiles through the water column (Picheral et al., 2010). To date, the 25 units of the UVP5 have been used at 14,462 sites since 2008 providing 94.8 millions of images and ~1,000 new profiles every year. The UVP5 optimal size range to identify and measure morphological traits is in the order of 600 µm to a few millimeters (Picheral et al., 2010; Drago et al., 2022). The number of taxa that can be recognized is over 200, whereas robust data is available for the 30 most abundant taxa (Drago et al., 2022; Panaïotis et al., 2023). Semi-supervised classification (automatic prediction followed by human validation) of plankton images (Irisson et al., 2022) was performed at regional scales, because of the time required to analyse all profiles and the lack of dedicated work flow to deal with millions of images. Recently the development of the collaborative platform Ecopart (https://ecopart.obs-vlfr.fr/) for UVP metadata and data curation, and EcoTaxa

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

(https://ecotaxa.obs-vlfr.fr/) for supervised image classification facilitated data treatment for the users. The data, initially recorded to address regional questions, can then be aggregated in global datasets to investigate global distributions of counts, size and vertical flux of all detected objects (Kiko et al., 2022; Clements et al., 2022; Clements et al., 2023). Furthermore, these platforms allow global studies on macrozooplankton biomass (Drago et al., 2022), and community composition (Panaïotis et al., 2023), as well as on the community proportions of Rhizaria (Biard et al., 2016), their impact on carbon flux attenuation, and silicification (Laget et al., 2024). To date, these studies provided the following data products: global plankton biomass, counts and group specific carbon demand data. In addition, to facilitate data access for modelers, all UVP5 particle size profiles, which were analysed without image recognition, were released at a 1 degree spatial resolution (Kiko et al., 2022). A more recent database obtained by combining images from Imaging Flow CytoBot, Zooscan and UVP5 to provide data products on plankton and detritus biomass binned in 1 degree resolution was also released (Dugenne et al., 2024). An attempt was made to classify detritus images based on their morphology (Trudnowska et al., 2022) but given their high number and the lack of consistent shapes, their classification at global scale is still challenging.

Here, we present the global dataset comprising approximately 8 millions validated images of plankton and detritus, rigorously verified by expert taxonomists, associated with relevant metadata and their morphological measurements, that was not released in earlier studies. The dataset is of interest for marine ecologists interested in plankton biogeography and biogeochemical modeling and for computer scientists developing Artificial Intelligence methods to classify images. While the current plankton classification has been homogenized for 33 categories, there remains potential for further improvement to increase the list of taxa consistent at a global scale. Classification of detritus was not performed, as existing methodologies developed at regional scale must be improved for a global approach. By providing the raw images for plankton and detritus, we hope to foster the development of new algorithms to sort and analyze them. The manuscript is organized as follows: details about the UVP5, the inter-calibration and quality control procedures, as well as the dataset structure are provided in the Material and Methods section. Maps of dataset distribution, summarizing statistics regarding taxa composition, size spectrum, description of global detritus and plankton distribution in the Results section. Finally, recommendations for dataset use and potential future expansions of it are provided in the Discussion section.

2 Material and methods

2.1 Sampling sites

Data from the Underwater Vision Profiler 5 (UVP5), an in-situ imaging system designed to detect, measure, and quantify the distribution of zooplanktonic organisms and marine particles (Picheral et al., 2010), were used. The UVP5 dataset (Figure 2) was compiled from observations across all oceans over a 10-year period (2008–2018) through a collaboration of international partners. It includes 3,114 profiles from 62 EcoTaxa projects conducted during research cruises by different institutions (Table A1), each aimed at addressing specific local or basin-scale research questions.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

2.2 Image acquisition with UVP5

2.2.1 UVP5 description

The UVP5 was commercialized in 2010 and produced until 2021. The standard definition (SD) version with a 1.3 Megapixel greyscale camera was produced between 2008 and 2016 (serial numbers 000 to 011) and the high-definition (HD) version with a 4 Megapixel greyscale camera was produced between 2016 and 2021 (serial numbers 200 to 223). In the standard definition setting, the UVP5 images a volume of about 1 L at a frequency of 5 to 20 Hz and can be deployed down to 6000 m depth. The UVP5 is mostly integrated in the CTD-Rosette and therefore its results (particle and plankton counts) can be related to environmental data obtained with other sensors. All particles larger than appr. 100 µm Equivalent Spherical Diameter (ESD) are sized and counted but their images are not stored because the low number of pixels precludes any recognition of the particles. Recognized objects larger than 30 pixels (UVP5SD) or 80 pixels (UVP5HD) are automatically cropped and the resulting vignettes stored and further analyzed.

2.2.2 Image analysis by zooprocess

Vignettes contain plankton organisms, detritus and artefacts larger than approximately 1 mm ESD for UVP5SD and 600 µm for UVP5HD. Pixel size-to-millimeter conversions for UVP5SD and UVP5HD are derived from objects of 30 and 80 pixels, respectively. For simplicity, we converted the surface area in pixels to its Equivalent Spherical Diameter. In addition, the released dataset contains 42 morphological features characterizing each object. Metadata collection (geographic location, date, etc.) and processing of all 8.46 million images was carried out using the ZooProcess software. The processing included the segmentation of the object from the obtained raw vignette and the measurement of 42 morphological features associated with each object (area, major and minor axis, grey level etc.; Picheral and Mériguet, 2025; Table A2). For user convenience, the images were inverted to show dark objects on a white background and a scale bar was added to each vignette. Images and metadata were imported into EcoTaxa (Picheral et al., 2017), an application which allows a taxonomic classification of images via supervised learning algorithms, followed by manual validation (Irisson et al., 2022).

The two versions of the UVP5 have been inter-calibrated based on the size spectrum of all particles measured concurrently in natural conditions (Kiko et al., 2022). Therefore, size is a conservative property between the instruments and plankton community composition can be inter-compared. While all parameters, except for shape-based ones, differ between the SD and HD versions due to the distinct dynamics of their imagers, other morphological properties, such as opacity, brightness, etc., may not be conservative, requiring caution in their joint analysis. For the HD version, measurements are taken from images with a dark field, which are then converted to a light field comparable to those in the SD version. Additionally, only images from a continuous descent of the instrument were retained.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

2.2.3 Semi-automatic image identification

Using EcoTaxa (https://ecotaxa.obs-vlfr.fr/), we applied semi-automated recognition in two steps following the third strategy path defined in Figure 2 by Irisson et al. (2022). First, features obtained using Zooprocess software or features re-calculated by a convolutional neural network (CNN) are considered to automatically predict object identities using a Random Forest algorithm. Therefore, various image learning sets were used independently on a project-by-project basis by different users. In this way, plankton organisms can be successfully classified into a few broad categories from the overdominant category of detritus with reasonable success. In a second step, manual validation and further sorting ensured correct classification of all plankton categories and detritus. In general the sorting was performed by various users and institutions. To reduce the risk of wrong identification, a shared UVP5 taxonomic guide was used by all annotators to homogenize image sorting (EcoTaxaGuide application). In addition, annotations were rigorously quality checked by the authors. The classification of all images was strictly based on the morphological features of the objects in most cases. However, few annotators have used contextual information provided by other means (e.g., bottle or net collection, depth, GPS) to classify objects with inadequate morphological attributes into taxa.

2.3 Dataset

The UVP5 image dataset comprises vertical profiles validated to more than 99%, with only near-completely validated profiles retained for subsequent analysis. Currently, the different annotators of the 62 EcoTaxa projects chosen have sorted the image datasets in more than 250 categories from species level to phyla or by adding morphological attributes and life stages. Such detailed classification is not possible over the whole dataset with high accuracy because recognition and sorting of organisms can be a source of bias depending on the levels of perception and experience of the people who perform them. Several cognitive biases exist, such as boredom, fatigue or a classification biased towards the most used groups (Culverhouse, 2007; Culverhouse et al., 2014). Currently 71 different annotators have classified the images, potentially introducing substantial classification variability and errors (Table A3).

In the present dataset, to reduce these errors and ensure classification homogeneity, the images were thereafter grouped into only 33 broader taxonomic groups for living organisms by combining all vignettes within the children categories (Table 1), while in EcoTaxa the original categories are maintained. The broad taxonomy list was established following recent published works using UVP5 data (Drago et al., 2022; Panaïotis et al., 2023; Laget et al., 2024) to ensure a minimum of 50 images per category and ecological patterns at regional or global scales for specific groups or community composition. For example, rhizarians were well studied at local (Biard and Ohman, 2020) and global scales (Biard et al., 2016; Laget et al., 2024) while polychetes were mostly studied in the tropical Atlantic (Christiansen, 2016). Community composition analyses have been conducted at local scales by Forest et al. (2012) and Barth et al. (2020), and at a global scale by Panaïotis et al. (2022). Global biomass was estimated in Drago et al. (2022). Inter-comparison with net data at global scale (Soviadan et al., 2024) allowed to set confidence in the abundance assessments of many of these groups.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Images of unidentifiable objects exhibiting probable biological characteristics (e.g., symmetry, appendages, or tentacles) were classified into the "plankton-like" category. Thus, the term -like in this dataset refers to all objects that have a structured morphology (appendages, symmetry) that make them resemble an organism but that can't be classified as such. This labeling includes issues due to bad focus, low image resolution, incomplete organism capture, or the unknown nature of the organisms to the annotator that hinder the accurate taxonomic identification. For example, copepoda-like contains all vignettes of copepods for which the antenna are not visible while the body shape resembles typical prosome and urosome. If many identical objects were common but not evenly distributed and not yet classified, new temporary categories were created and labeled temporary ('tmpXXX'). When contextual information to identify them was used, "contextual" was added to their names. All images of non-living organisms, including poor-quality images (blurred, low grey level, bubbles, artificial nature, badly segmented images) or particles (aggregates, fibers, pellets) were sorted into several categories without consistency criteria across the consortium. Low-quality images (mostly objects not completely in focus) were set in a category named "artefact" that includes bad focus particles, plankton and real artefact, while particles were grouped into a single detritus category.

To ensure data consistency across projects, a subset of the dataset was reviewed to verify the accuracy of taxonomic sorting within each group. A total of 200 images were extracted from each category, or all available images if fewer than 200 were present. For detritus, 10 % of the total images were selected and independently reviewed after the final dataset was assembled. The error or uncertainty rate was under 8 % for taxa classification. Among the detritus, less than 0.1 % were plankton organisms. The resulting global dataset consisted of 7.05 million detritus images and 734 thousands plankton images from 3,114 profiles.

Due to the large number of plots required to display the results for all categories, results are shown for a subset of the original categories. From the 33 initial categories derived from taxonomy (Table 1), five taxonomic groups were formed for further analysis based on feeding strategies and/or their ecology. These groups can be highly useful as input for biogeochemical or community ecology models. The groups include Crustacea (Copepoda and Malacostraca), gelatinous filter-feeders (Appendicularia, Salpida), gelatinous carnivores (Chaetognatha, Narcomedusae, Siphonophorae, Ctenophora), Rhizaria (Collodaria, colonial Collodaria, other Rhizaria, Foraminifera, Phaeodaria, Acantharea), and *Trichodesmium*. From these groups, biovolumes were calculated along with concentrations at different depth layers and for the analysis of the average grey level.

2.3.1 Data output format

We provide four tables with numerical data (in .tsv format), as well as all images (with .jpg extension) while their full description is available in the information documents (UVP5_dataset_organization.txt) during the download. Briefly, 'objects.tsv' is the data table which contains each object information. The three first columns refer to the unique object ID given by Ecotaxa, original object classification in Ecotaxa and object classification according to Table 1. The following 42 columns contain all the object features obtained from the image (see Table A2). The 'samples.tsv' file is the table with the metadata corresponding to all the coordinates (latitude, longitude, date and time) and acquisition details (pixel size and UVP model). The 'samples_volume.tsv' is a table that contains the volume of

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

248 water in which the object was extracted and that is used to calculate concentrations (middle depth bin and volume
249 imaged). Finally, 'properties_per_bin.tsv' is a table with concentrations, biovolume, average size (ESD) of the 33
250 consistent categories per depth bin and per sample. The native size bin of the dataset is 5 m, but to reduce the size of
251 the file and avoid too many zeros, we provide 25 m depth bins in the surface ocean (up to 200 m), 50 m depth bins
252 (up to 500 m), 100 m bins (up to 1000 m) to 250 m below 1000 m. We provide the code in R to recalculate the table
253 with a depth bin of 5 m. The two files 'ODV_biovolumes.txt' and 'ODV_concentrations.txt'' contains the same
254 information as 'properties_per_bin.tsv' but in the ODV format.

255 **Table 1: List of categories that are quantitatively consistent in this work and names of children categories not homogeneous**
256 **among the projects. Categories also defined by their morphotypes or life stages are in bold. The number of images in each**
257 **category is given in parenthesis. The published dataset contains both the broad and the children categories. When available**
258 **publications using the dataset at regional scales are listed. Exponent numbers denote the association with an image example**
259 **in Figure 1.**

Broad categories selected in this work (number of vignettes)	If more than 50 vignettes, children categories (number of vignettes) proposed by data owners but not yet homogenized	Reference
Acantharia (12,616)[1]		
Actinopterygii (482)[2]		
Annelida (3,261)[3]	Alciopidae (74), Phyllodocidae (53), Tomopteridae (237), Acrocirridae (218), Flota (185), Poeobius (358)	Christiansen et al., 2022; Barth and Stone, 2022
Appendicularia (9,354)[4]	*body (540), house (836)*	Forest et al., 2012
Bacillariophyta-contextual (44,637)[5]	All objects classified as such using contextual data	
Chaetognatha (6,793)[6]	**tail**<*Chaetognatha (64),* **head**<*Chaetognatha (62)*	Forest et al., 2012; Barth and Stone, 2022
other Collodaria (7,644)[7]	*Solitaryblack (5,905), solitaryglobule (1,738)*	Biard and Ohman, 2020
colonial Collodaria (2,329)[8]		Biard and Ohman, 2020
Copepoda (173,909)[9]	like<Copepoda (72,284), Eucalanidae (782), like<Temoridae (280), *ovigerous (117), ovigerous_side (3,484), ovigerous_top-bottom (287)*, Pseudocalanus (99), Harpacticoida (65)	Forest et al., 2012; Vilgrain et al., 2021; Barth and Stone, 2022
Ctenophora (2,743)[10]	Beroida (238), Cestidae (66), Lobata (111), **tentacle**<*Ctenophora (65)*	Forest et al., 2012
Doliolida (1,304)[11]		
Eumalacostraca (11,160)[12]	Amphipoda (434), Pleuroncodes (258), Euphausiace (531), Munnopsis (70)	Forest et al., 2012; Barth and Stone, 2022
Foraminifera (3,014)[13]		
Gasteropoda (1,500)[14]	Gymnosomata (87), Limacinidae (1015), Creseidae (120), other Thecosomata (198)	

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Narcomedusae (2,938)[15]	*Solmundella bitentaculata* (1,791), Solmaris (203)	
Nostocales (1,369)[16]	Aphanizomenon (689), Dolichospermum (179), Nodularia (501)	
Ostracoda (8,427)[17]		Barth and Stone, 2022
Phaeodaria (65,030)[18]	Castanellidae (2,405), Circoporidae (271), Tuscaroridae (85), Aulacanthidae (236), Aulacantha (18,572), Aulographis (827), Coelodendrum (283), Coelographis (2,474), ***leg***<*Coelodendridae* (1,513), Medusettidae (459), Aulosphaeridae (31,616), Aulatractus (217), **colonial**<*Aulosphaeridae (906),* like<*colonial* (1,381), Cannosphaeridae (1,416), **colonial**<*Cannosphaeridae* (84)	Biard and Ohman, 2020
Pyrosoma (334)[19]		
Salpida (592)[20]		
Siphonophorae (2,557)[21]	**part**<Siphonophorae (380), **tentacle**<Siphonophorae (1,381)	
Trachymedusae (462)[22]	Halicreatidae (184), *Aglantha digitale* (100)	
Trichodesmium (88,676)[23]	Puff (41,936), Tuff (46,740)	Guidi et al., 2012; Sandel et al., 2015
Trichodesmium-contextual (98,956)[24]	all objects were classified as such by using contextual data, 95% are from the same cruise	Dupouy et al., 2018
other Cnidaria (1,994)[25]	Ceriantharia (74)	
other Crustacea (4,569)[26]	**larvae**<*Crustacea (58),* **tail**<*Crustacea (122)*	
other Hydrozoa (8,659)[27]	Leptothecata (96)	
other Mollusca (3,833)[28]	veliger (3,092)	
other Rhizaria (9,911)[29]	contains plankton resembling to Rhizaria	
tentacle of Cnidaria (8,008)[30]		
plankton-like (74,241)[31]	contains all structured objects most probably living organisms, which are in multiple categories to check (35,083), possibly plankton (18,289), to resort (15,535), other living (5,868) and to rename (2,942)	
detritus (7,055,138)[32]	contains particles such as aggregates, filaments, mucus net which are not yet sorted	Trudnowska et al., 2021
Artefact (49,636)[33]	bubble (8,332), turbid (9,507)	

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

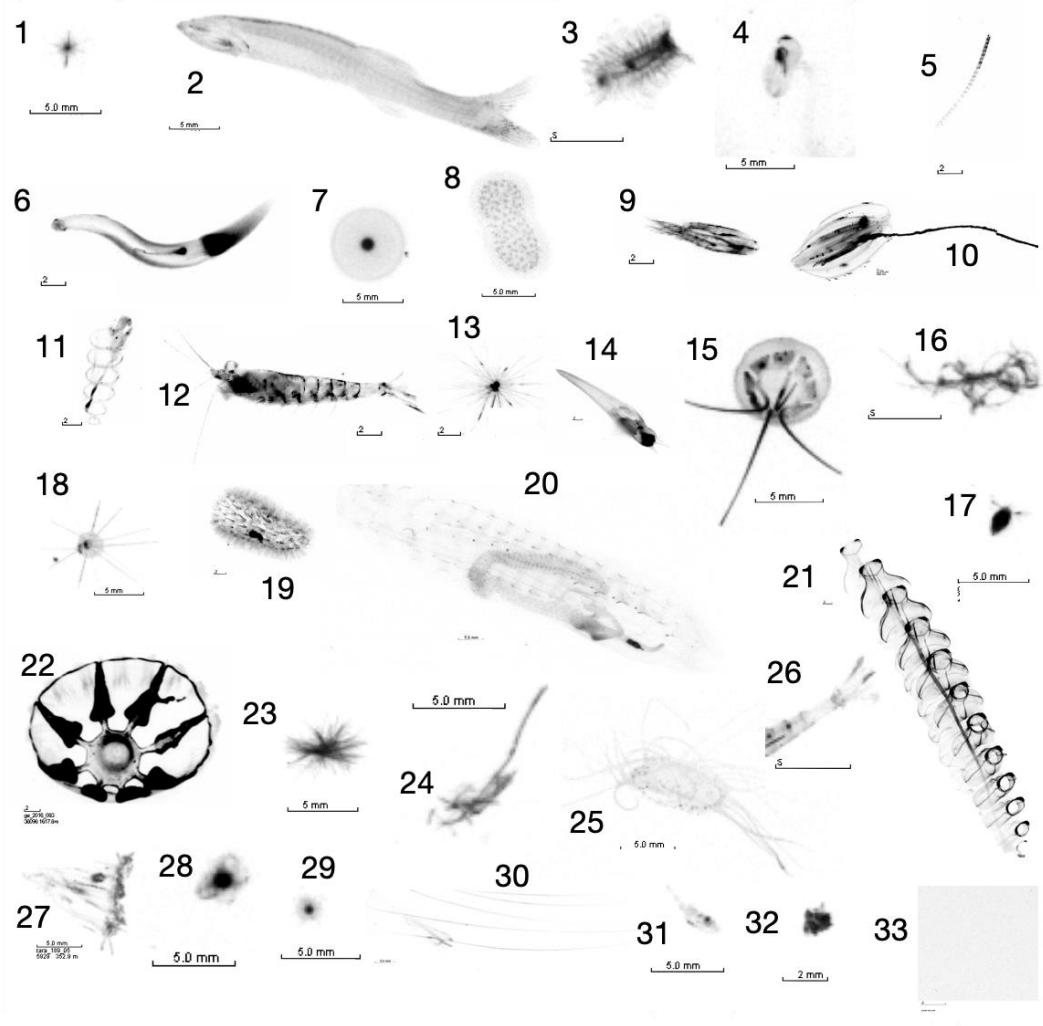

Figure 1: Examples of images for each category (Table 1) present in the dataset.

3 Results

3.1 Data coverage in time and space

The dataset presented here reveals a heterogeneous global distribution. The map in Figure 2 illustrates the profiles collected using two different versions of the UVP5 (SD and HD) during deployments across the global ocean between 2008 and 2018. While sampling efforts cover the global ocean, UVP5 deployments predominantly occurred in the northern hemisphere (66.3% of profiles), especially in tropical and temperate latitudes where seasonal variations are well observed. In contrast, the southern hemisphere has fewer observations (33.7% of profiles), resulting in a less

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

comprehensive dataset. The equatorial region (30º S-30º N) represents the highest proportion of UVP5 profiles (47.25%). The vertical profiles distribution reveals a predominant concentration of measurements in the upper 1000 m of the water column, representing 40.6% and 46.8% for the SD and HD models respectively, with a peak occurring within the upper 500 m. This indicates that most deployments targeted shallower depths. The SD version of the UVP5 demonstrates the most extensive coverage, with profiles spanning the entire latitudinal range. Meanwhile, the HD version UVP5 profiles are more prevalent in the Northern Hemisphere. Importantly, the two versions were rarely co-located in space and time, preventing meaningful inter-comparison of zooplankton community composition. These patterns provide valuable insights into the sampling strategies and data representation of UVP5 profiles across the global ocean. They also highlight the seasonal variability observed primarily in the northern hemisphere and the under-sampling of the southern hemisphere and the deep sea.

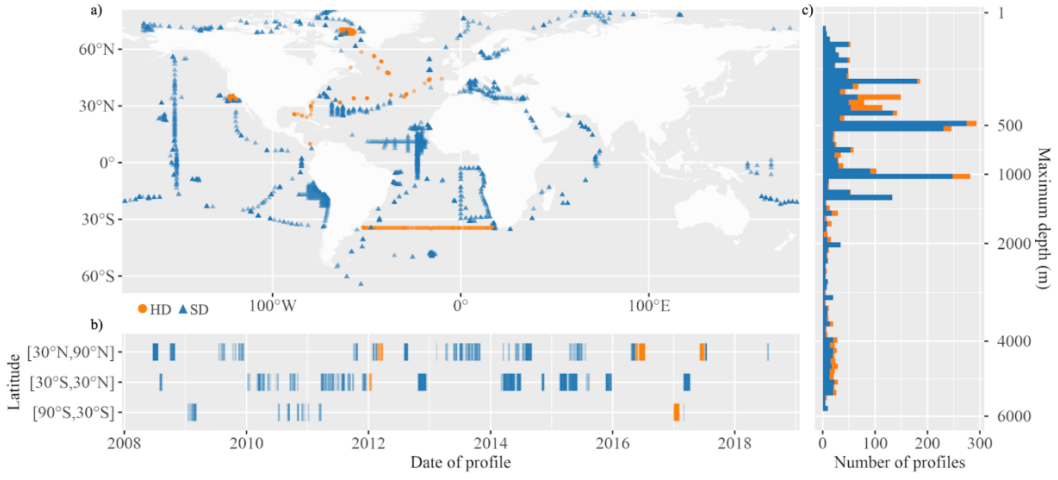

Figure 2: a) Global sampling effort for the two versions of UVP5 (SD and HD, blue and orange, respectively) between 2008-2018 in different oceans. b) Indicates the latitudinal distribution of UVP5 profiles and c) shows the maximum vertical extent of the profiles.

3.2 Detritus and plankton composition in the whole dataset

Figure 3 presents the relative abundance of dataset categories, with the six most prevalent in decreasing order being: detritus, Copepoda, *Trichodesmium*-contextual, *Trichodesmium*, plankton-like, and Phaeodaria. Detritus is the most abundant category by far (90.5% of total images), indicating a substantial presence of non-living particulate organic material. Copepoda dominates the zooplankton and represents a significant fraction of the dataset (2.2%), followed by *Trichodesmium* (1.3%) and the *Trichodesmium*-contextual (1.1%) categories. The category plankton-like reflecting some level of taxonomic uncertainty in the dataset, is the fourth category in terms of images within the dataset (1%). The artefact category, including bubbles and artefact, occupied the sixth position (0.9%). This category is followed by Phaeodaria (0.8%) and Bacillariophyta-contextual (0.6%). Each of the remaining groups represent less than 0.16% of the total number of objects.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

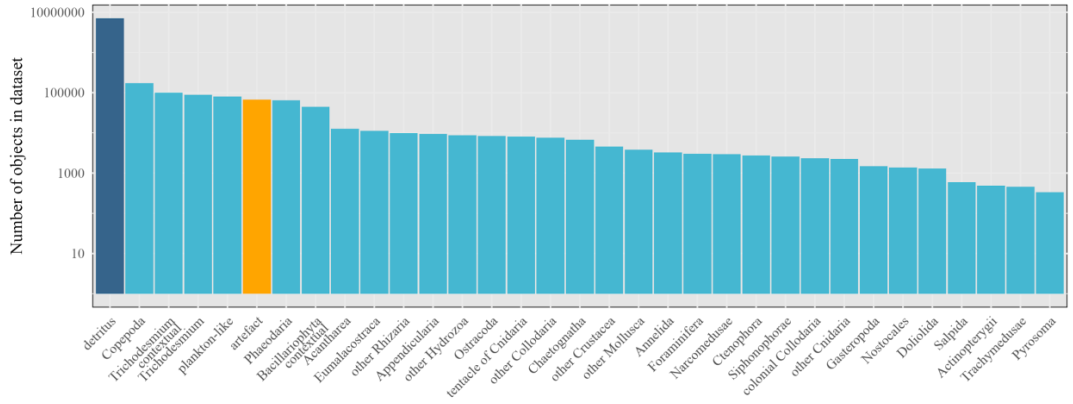

Figure 3: Barplot with total number of objects in logarithmic scale available in the dataset per category (detritus and plankton) for the 33 consistent categories. The categories are ordered by descending count, with specific categories ("detritus" and "artefact") highlighted in distinct colors. Note that the artefact category represented here contains all vignettes including turbid, bubbles, badfocus (see section 2.3.1).

3.3 Global distribution of plankton and detritus

The spatial patterns of plankton and detritus concentrations in the global ocean vary significantly with depth and location. Figure 4 depicts the global distribution of abundances in three depth layers. Across all depth layers, plankton and detritus abundances are increased (up to 5 plankton/m^3 and 60 detritus/m^3 in the 0-100 m depth layer and up 2 plankton/m^3 and 20 detritus/m^3 in the 500-1000 m depth layer) in eastern boundary current systems (especially in the Californian and Senegal upwelling systems) or in coastal seas. Minimum values (up to 2 plankton/m^3 and 20 detritus/m^3 in the 0-100 m depth layer and up 1 plankton/m^3 and 1 detritus/m^3 in the 500-1000 m depth layer) are usually found in the center of large ocean gyres (notably in the South Pacific and the Indian Ocean).

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

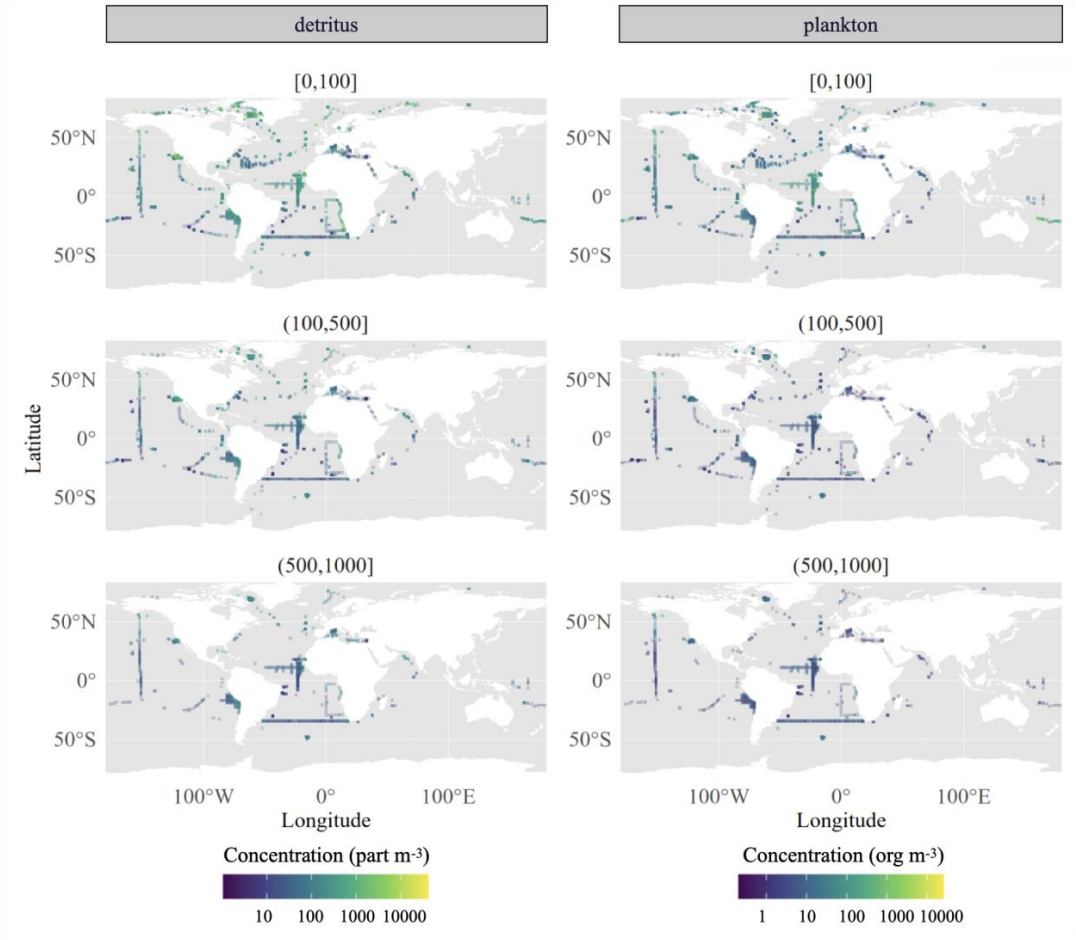

Figure 4: Map of the global detritus and plankton concentration (log particles/organism m^{-3}) in three layers (0-100, 100-500 and 500-1000 m).

3.4 Average size distribution of SD and HD versions

The distribution of detritus, the most abundant category, follows a dome-like pattern (criteria from Drago et al., 2022), with UVP5 showing a lower detection threshold for particles below 1 mm for UVP5 SD and 600 µm for UVP5 HD. In contrast, taxonomic categories exhibit a much flatter normalized biomass/biovolume size spectra (NBSS) distribution compared to detritus. Detritus dominates the smaller size range while the contribution of plankton increases relative to detritus in the larger size range. Except for Rhizaria, smaller planktonic organisms were more abundant in the upper 100 m in the UVP5hd dataset. This analysis emphasizes the differing detection limits of the two camera versions in the small size range for both plankton and detritus (Figure 5). Consequently, we recommend a standardized size cutoff of 1.02 mm for ecological studies focusing on abundance and biomass integrating both data sets, while studies utilizing solely UVP5 HD should employ a 600 µm size threshold.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

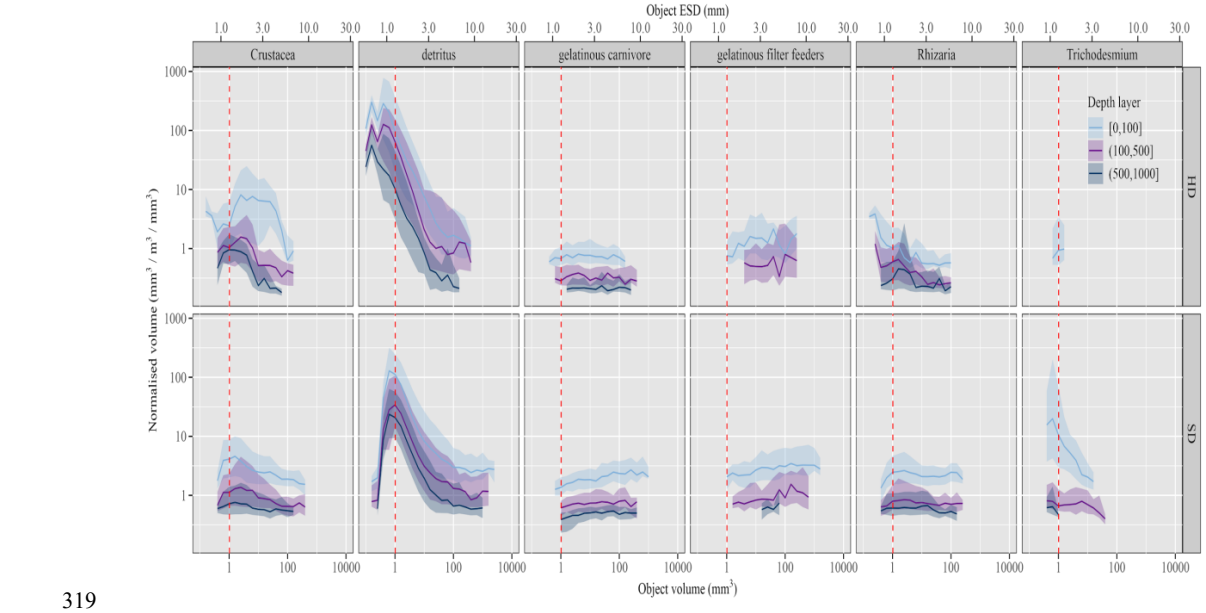

Figure 5: Average size distribution of detritus and plankton categories obtained by the SD and HD version of the UVP5 in three depth layers. Red dashed line indicates the threshold value (1.02 mm) below which UVP5 detection of organisms is difficult to quantify properly. Crustacea (Copepoda and Malacostraca), gelatinous filter-feeders (Appendicularia, Salpida), gelatinous carnivores (Chaetognatha, Narcomedusae, Siphonophorae, Ctenophora), Rhizaria (Collodaria, colonial Collodaria, other Rhizaria, Foraminifera, Phaeodaria, Acantharea), and *Trichodesmium*.

3.5 Plankton to detritus ratio as a function of size

For both UVPs, the proportion of plankton images on all images (detritus and plankton) increases with size from 0.01 at 500µm to about 0.8 for objects larger than 3 mm in ESD. The two instruments display a similar pattern with a plateau in this ratio starting at about 3 mm. In general, the ratio does not vary with depth and remains at 0.8 for objects larger than 3mm.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

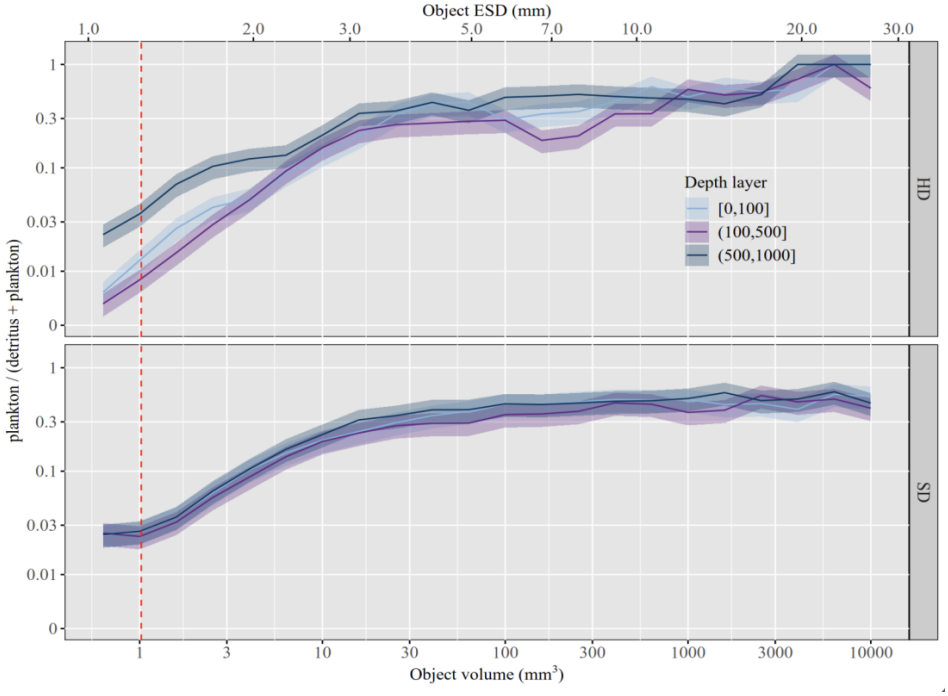

Figure 6: Abundance ratio of plankton (without plankton-like) over all images as a function of size for the two different UVP5 versions. Red dashed line indicates the threshold value (1.02 mm) below which a robust detection of organisms is limited.

3.6 Vertical profiles of detritus and plankton concentrations

When objects are analyzed across different vertical layers (Figure 7), detritus exhibits increased concentrations, primarily within the uppermost 100 m of the water column. The biovolume distribution mirrors the concentration pattern, with detritus exhibiting substantially higher values—several orders of magnitude greater—compared to living plankton organisms. Trichodesmium (not including Trichodesmium-contextual) was the most abundant group in terms of concentration within the upper 100 meters, with its presence decreasing with depth. However, in terms of biovolume, due to their smaller size (Figure 4), it represented a relatively small proportion compared to other groups analyzed. Crustacea, on the other hand, were present in similar proportions in both concentration and biovolume, showing a high abundance in the surface layer and a decline with depth. In the case of gelatinous plankton, both with a filter-feeding (Appendicularia, Salpida) and carnivores strategies (Chaetognatha, Narcomedusae, Siphonophorae, Ctenophora), their concentrations were very low in all analyzed depths, while their biovolume became significantly important, particularly in the upper 100 m of the water column. Both parameters showed a decrease with depth for these groups, similar to the patterns observed for the aforementioned groups. Rhizaria, in turn, showed peak concentrations in the upper 100 m, which gradually decreased to 1000 m. Their biovolume values exhibited patterns similar to those observed for concentration.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

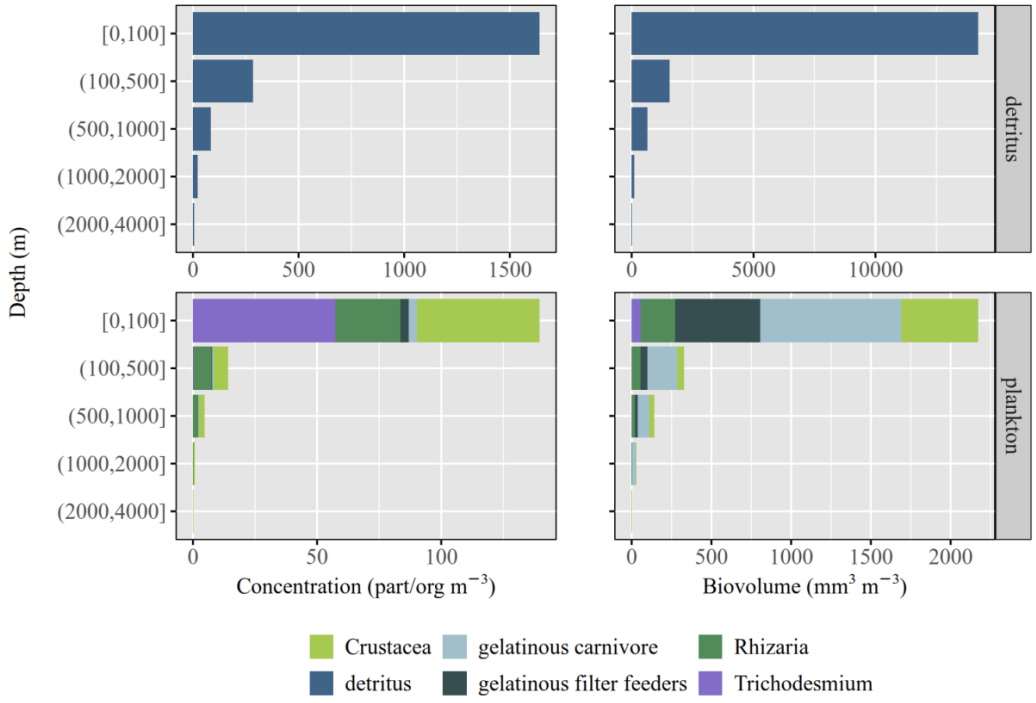

Figure 7: Global vertical profiles of concentration and biovolume for detritus and the five plankton functional groups.part: particles of detritus; org: organisms from different taxa categories. Crustacea (Copepoda and Malacostraca), gelatinous filter-feeders (Appendicularia, Salpida), gelatinous carnivores (Chaetognatha, Narcomedusae, Siphonophorae, Ctenophora), Rhizaria (Collodaria, colonial Collodaria, other Rhizaria, Foraminifera, Phaeodaria, Acantharea), and *Trichodesmium*.

3.7 Morphological properties

The key morphological properties for detritus and the five taxonomic groups presented similar patterns between the two UVP5 versions (SD and HD). In some cases, the range differed between versions for the groups depending on the property examined (Figure 8). Gelatinous groups were the biggest categories in terms of area measured, while they presented the lowest circularity, compared to the remaining groups. Crustacea was the darkest group, whilst gelatinous were the lightest ones.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

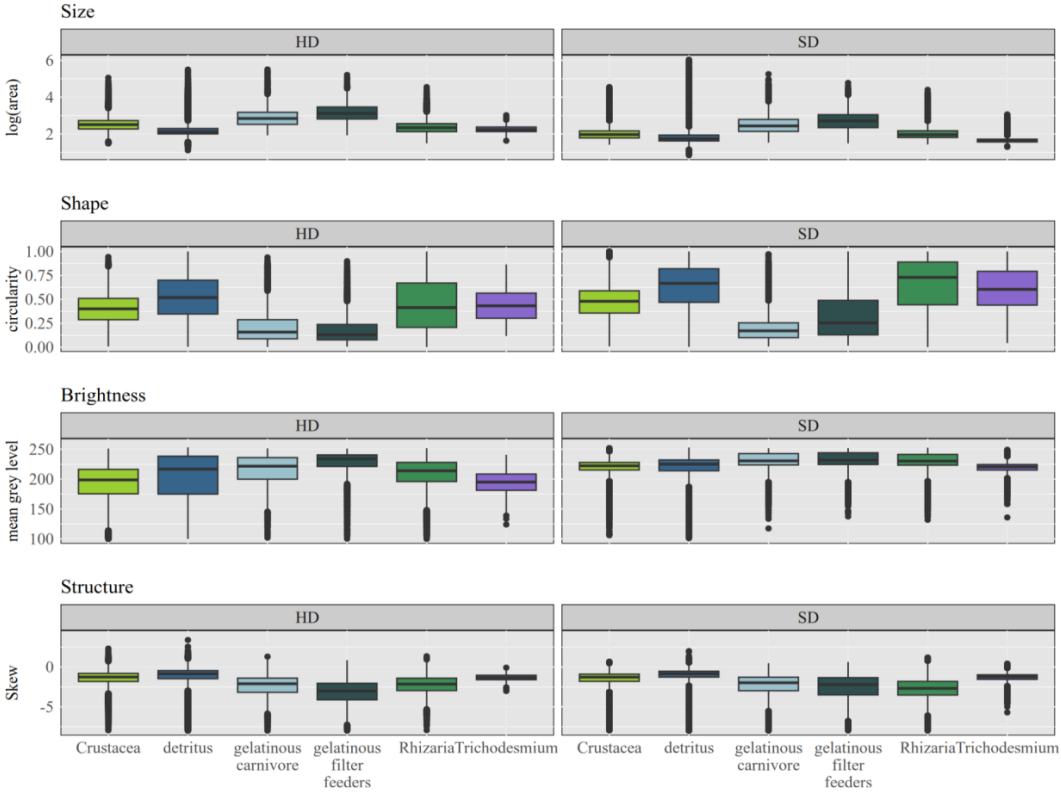

Figure 8: Box plots of key exemplary morphological properties (area, circularity, mean grey level, skew) for detritus and the five taxonomic groups (Crustacea, gelatinous carnivore, gelatinous filter feeders, Rhizaria and *Trichodesmium*) for the two UVP5 versions (HD and SD). Crustacea (Copepoda and Malacostraca), gelatinous filter-feeders (Appendicularia, Salpida), gelatinous carnivores (Chaetognatha, Narcomedusae, Siphonophorae, Ctenophora), Rhizaria (Collodaria, colonial Collodaria, other Rhizaria, Foraminifera, Phaeodaria, Acantharea), and *Trichodesmium*.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

4 Discussion

The availability and accessibility of both the images and their associated metadata present new challenges, encouraging the scientific community to leverage and refine them for future research. With this objective in mind, the dataset presented here serves as a starting point for further exploration of the current database (10,000 profiles not fully validated) and future data incorporated in EcoTaxa (approximately 500 profiles annually on average over the last three years for UVP5 versions only), covering both plankton and detritus. In this context, we discuss the strengths and limitations of the dataset, as well as its potential applications and future directions for the UVP database.

4.1 Consideration related to spatial coverage and instruments performances

The dataset covers all major ocean basins, though the southern hemisphere and the deep sea remains largely undersampled. The list of profiles included are primarily confined to the mid- to high-latitudes of the northern hemisphere (Figure 2). Some specific regions are well sampled as they belong to monitoring or time-series programs developed during the last decades, such as the California Current, the Western Mediterranean, and the Equatorial region, through initiatives like the CCELTER program (e.g., Biard and Ohman, 2020), ongoing observational efforts of MOOSE (e.g., Llopis Monferrer et al., 2022), and long-term studies in the Equatorial region (e.g., Kiko et al. 2017, Fernández-Carrera et al., 2023). The even distribution across ocean basins is mostly the result of the Tara Ocean Expeditions (2009-2013). Altogether, this dataset represents global coverage over a decade, with homogeneous and intercomparable methodologies across UVP5 versions. Despite this extensive coverage, there are some potential issues regarding data quality. For instance, the small volume sampled by UVP5s can lead to challenges in detecting rare groups, and the quality of the images may affect the accuracy of object identification. Additionally, "zeros" might not always reflect the actual absence of objects or natural patchiness of plankton and particles, but rather a result of low volume sampled or detection limits of the instrument.

At larger size ranges, as organisms become rare, statistical analysis may be challenging and required to aggregate counts across thick ocean layers. At the surface, abundant taxa may permit 10 m binning intervals (Dupouy et al., 2018; Biard and Ohman., 2020), though 20 m bins are typically more appropriate (Picheral et al., 2010). For rare taxonomic groups, aggregation across several hundred meters may be necessary to achieve statistically meaningful analyses (Hauss et al., 2016; Panaïotis et al., 2023). In this study, we employed a depth-layered approach to mitigate variability in rare taxonomic groups, aggregating data into five progressively wider depth layers (0-100, 100-500, 500-1000, 1000-2000 and 2000-4000 m). While previous publications have chosen 0-200 m and 200-500 m (Drago et al., 2022) or a dynamic one, for the first layer, set as the deepest value among the mixed layer depth and the euphotic depth (median 88 m, Panaïotis et al., 2023). Such a choice depends on the scientific question.

The two versions of the UVP5 are inter-calibrated based on the size spectrum of all objects obtained from repeated joint profiles at the same location (Picheral et al., 2010; Kiko et al., 2022). A peak in the size spectrum at

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

the lower size range generally reflects the minimum size of efficient detection by in situ imaging, while high variability in the large size range due to the poor ability to detect rare large objects results in constant (e.g., flatter) distribution (Stemmann and Boss, 2012). In between these two size ends, a steep decrease is generally observed which translates to a quasi-linear decrease on a log scale. The detritus size spectra (Figure 5) clearly show that the HD version, presenting a smaller pixel size, has a better detection than the SD version for detritus smaller than 1 mm, but that above a threshold size of 1 mm both versions are efficient up to a few mm as the slopes get flatter.

4.2 Considerations related to classification

Variations in taxonomic classification depth among annotators necessitated the merging of certain groups, resulting in a final list of 33 taxa with a lower common denominator (Table 1). To validate classification accuracy, a random subset of images (200 from each taxon and 10,000 for detritus) was independently reviewed by a single operator. Error rates were consistently below 10% across all groups, with most groups exhibiting error rates under 2.5%. An additional source of error in the case of detritus is that many annotators had moved un-focused particles into the category artefact. This category includes various sources and types of images that may result from factors such as a poorly tied rope, which was recorded during the image capture, as well as seals or cables. On the other hand, it may also consist of effects like bubbles caused by the descent through the water column and the air intrusion, as well as artifacts resulting from image blurriness where no actual object or non-recognizable object is present (real artifact). When analyzing the artefact category which contains some subjectivity, considering for example that bad-focused could be very transparent objects, we found that it consists of 85% of real detritus shaped objects, 5% of plankton and 10% of artefact. Sorting them in the EcoTaxa database system would be too time-consuming, so users must decide whether to include or exclude them during the analysis. Here, we merged them with detritus as they would become a minority (less than few percents). However, it is important to note that objects classified as artefacts (not shown in Figure 7) represent less than 1% of the total dataset when the concentration is calculated, but contribute to more than 200% when volume is considered. Their impact is high mostly on large objects because many artefacts are observed in very turbid water where particles are badly segmented. This highlights the significance of their removal.

Plankton-like

We have gathered in plankton-like all objects that are most probably plankton. Inside this category, ~60% of the objects in the possibly plankton group may be Rhizaria, but the image definition impairs their recognition. It is important to keep this category apart from the detritus category because they represent a significant fraction of the total plankton count and probably living biomass. Although a significant portion indicates uncertain classification, it is also in this category that most new findings can be performed.

Plankton contextual

The term -contextual in this dataset refers to objects that were classified as a taxonomic category but do not have the morphology of the taxa. Scientists on oceanographic surveys who had other contextual information

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

have created the category *Trichodesmium*-contextual and Bacillariophyta-contextual for objects that aren't formally identified as Trichodesmium puff and tuff or as diatom chains. In this context, the classification is supported by many other optical observations (for example see Dupouy et al., 2018 for *Trichodesmium*). Due to the risk of mis-interpreting ecological results or impairing the quality of the learning set, it is better to distinguish them from objects whose images are typical of plankton organisms. In the case of the two categories, *Trichodesmium* and *Trichodesmium*-contextual, they contain nearly the same number of objects, so including both would double the amount of data for this group. On the other hand, Bacillariophyta-contextual represents the sole category for this genus within the dataset, where its inclusion or not could lead to its complete exclusion.

4.3 Scientific relevance

Overall, this dataset provides a unique and comprehensive tool for the scientific community, enabling researchers to examine the intricate relationships between plankton, detritus, and ecosystem function. The detailed classification of plankton and detritus provides a valuable resource for understanding ecological dynamics at global (Drago et al., 2022; Panaïotis et al., 2023; Laget et al., 2024) and local scales (see references in Table 1) within marine systems as already shown in several studies. By analyzing the abundance, diversity, and distribution of planktonic organisms, researchers can gain insights into food web structures, energy flow, and the interactions between different trophic levels (e.g., D'Alelio et al., 2016; Perhirin et al., 2025). The data allows ecologists the monitoring of plankton communities, which are essential indicators of environmental health and can reflect changes in water quality, nutrient levels, and ecosystem productivity (Muller-Karger et al., 2018). The dataset is equally valuable in biogeochemical studies, so researchers can better predict how changes in plankton community composition and detritus abundance influence nutrient cycling and the overall functioning of marine ecosystems. The complex relationships between these components provide critical insights into primary productivity, nutrient fluxes, and the potential impacts of environmental stressors such as climate change and ocean acidification. Additionally, biogeochemical models that integrate this dataset can help refine predictions of carbon sequestration in the ocean, an essential process for regulating atmospheric CO_2 levels and mitigating global warming. As plankton and detritus play pivotal roles in the ocean's carbon pump, understanding their dynamics is crucial for assessing the resilience of marine ecosystems and their ability to mitigate or exacerbate climate change effects (e.g., Le Quéré et al., 2016; Rohr et al., 2023). Through these integrated models, the dataset offers a robust framework for exploring the complex feedback mechanisms between marine biological communities and biogeochemical processes, advancing our ability to predict future changes in marine ecosystems under varying environmental scenarios.

4.4 Moving forward: future development

The published dataset (3,114 profiles, 7,000,000 images) is publicly available, while only validated images are visible from the remaining profiles in EcoTaxa (6,000 profiles, 83,000,000 images), although not yet publicly accessible to be downloaded and/or re-processed. Therefore, future users should contact the data owners through the Ecopart and EcoTaxa platforms to request access to the complete datasets.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Despite the efforts of the data owners, many images in the EcoTaxa database for UVP5 were not recognized in any taxonomic group, so were categorized under a plankton-like category. The challenge of further sorting plankton-like organisms is significant due to its great abundance, but it also provides an opportunity to increase the biodiversity within the dataset and refine the taxonomy.

Future classification should focus its efforts on sorting the detritus category, as a recent study highlights that sinking speed estimated for various particle categories varies with particle morphology (Trudnowska et al., 2022; Soviadan et al., 2025) or that mesoscale activities mix them them vertically according to their morphology (Accardo et al., 2025). In addition, they represent a very high proportion in abundance and probably of organic biomass compared to the plankton. Additionally, the specific classification of detritus particles, such as fecal pellets and their associated zooplankton communities, will enhance our understanding of their dynamics, improve the accuracy of carbon flux calculations, and refine their representation in biogeochemical models (Perhirin et al., 2025). Therefore, classifying particles based on their traits could provide valuable new insights into their origin and role in the carbon cycle through the gravitational pump (Boyd et al., 2019). When it comes to zooplankton, recent studies revealed that analyzing their traits, such as body size, morphology, and metabolic rates, which significantly influence their ecological roles and interactions within marine ecosystems, provides a framework for linking organismal characteristics to ecosystem functions, offering deeper insights into zooplankton behavior, community structure, and their contributions to biogeochemical cycles (Titocci et al., 2025).

Transfer learning has shown promising potential in leveraging existing datasets to enhance training for improved recognition tasks. By utilizing the present dataset as a model training for machine learning and artificial intelligence (AI) techniques can be employed to better classify and extract relevant features (Lumini and Nanni, 2019; Irisson et al., 2022). Specifically, AI could play a key role in sorting and categorizing the 7,000,000 detritus images in the dataset, enhancing the processing of the remaining 5,598 UVP5 profiles and their corresponding 70 million images. Furthermore, integrating new profiles, such as the anticipated 500 annually for UVP5 and 5,000 annually for the new version of the UVP (UVP6lp and hd; Picheral et al., 2022), into the observation system will enhance the robustness of the database, ensuring its adaptability and scalability for future data handling and feature extraction. This observation effort is supported by GOOS (Dexter and Summerhayes, 2010) and already in operation in large-scale observing programs, like the Bio-GO-SHIP (Clayton et al., 2022) or possibly by the BGC-Argo (Claustre et al., 2020) program. Having better embedded classifiers is particularly important for BGC Argo for which the floats and the cameras are not recovered. Given the versatility of the new UVP6, which can be attached or associated with different platforms, it has allowed for widespread distribution and, consequently, a significant deployment of profiles (~23 thousands on EcoTaxa), reaching 87 millions of images in the last 4 years since its commercialization in 2020. These advancements will ultimately contribute to more efficient data analysis and support the continuous improvement of environmental monitoring systems.

Data availability

The dataset is available on SEANOE at https://doi.org/10.17882/107583 (Nocera et al., 2025).

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Acknowledgements

We are grateful for the ship time provided by the respective institutions and programmes and we would like to thank all scientists, officers and crew of the various R/V which took part in UVP5 data collection. We acknowledge the MOOSE programme (Mediterranean Ocean Observing System for the Environment) coordinated by CNRS-- INSU and the Research Infrastructure ILICO (CNRS--IFREMER). JOI acknowledges support by the Belmont Forum grant #ANR--18--BELM--0003--01. RK, LS and LD received support from the European Union project TRIATLAS (European Union Horizon 2020 Programme, grant agreement 817578). We also thank the National Science Foundation grants OCE-10-26607, OCE-1637632 and OCE-1614359 which supported the CCE-LTER site.

RK furthermore acknowledges support via a Make Our Planet Great Again grant from the French National Research Agency (ANR) within the Programme d'Investissements d'Avenir #ANR--19--MOPGA--0012 and funding from the Heisenberg Programme of the German Science Foundation #KI 1387/5--1. AR was funded by the PACES II (Polar Regions and Coasts in a Changing Earth System) programme of the Helmholtz Association and the INSPIRES programme of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. LS was supported by the CNRS/Sorbonne University Chair VISION to initiate the global observation and the Chair PLACARDO at the Institut de France. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement #862923. ACN was supported by a postdoctoral fellowship from the National Scientific and Technical Research Council of Argentina. LD received funding from Sorbonne Université through the Ecole doctorale 129 as well as Horizon Europe RIA under Grant Number 101081273 (NECCTON project) for her post-doctoral research funding.

Author contribution method

ACN, JOI and LS formulated the goals for dataset presentation, quality control and the publication of a global UVP5 dataset and their respective images. LS and MP led the community effort supported by all UVP managers and quality control endeavours, supported by all co-authors. ACN and LS conceived and drafted the article. All authors participated in writing the article.

Conflict of interest statement

The authors declare no conflicts of interest.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Open Access Earth System Science Data Discussions

529 **Appendix A**

530 **Table A1. List of all projects included in the present study (N = 62), with their corresponding project identification (pprojid),**
531 **title (ptitle), data owner and license registered on EcoTaxa (https://ecotaxa.obs-vlfr.fr/).**

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

pprojid	ptitle	data_owner	projid	title	license
1	uvp5_sn003_jerico_2017	lars.stemmann@obs-vlfr.fr	578	UVP5 JERICO 2017	
2	uvp5_sn000_boum2008	lars.stemmann@obs-vlfr.fr	122	UVP5 BOUM 2008	
3	uvp5_sn000_ccelter_2012	tristan.biard@univ-littoral.fr	45	UVP5 CCELTER (Iter2008 ccelter_2012 ccelter_2014)	CC BY-NC 4.0
4	uvp5_sn000_lohafex2009	lars.stemmann@obs-vlfr.fr	109	UVP5 LOHAFEX 2009	CC0 1.0
5	uvp5_sn000_lter2008	tristan.biard@univ-littoral.fr	45	UVP5 CCELTER (Iter2008 ccelter_2012 ccelter_2014)	CC BY-NC 4.0
6	uvp5_sn000_tara2009	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
7	uvp5_sn000_tara2010	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
9	uvp5_sn000_tara2012	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
10	uvp5_sn002_moose_dyf_2013	lars.stemmann@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
11	uvp5_sn002_moose_dyf_2014	lars.stemmann@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
12	uvp5_sn002_moose_dyf_2015	lars.stemmann@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
13	uvp5_sn002_moose_dyf_2016	lguidi@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
14	uvp5_sn002_moose_dyf_2017	coppola@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
15	uvp5_sn002_moose_ge_2013	lars.stemmann@obs-vlfr.fr	30	UVP5 MooseGE 2012 2013 2014 2015 2016	CC BY-NC 4.0
16	uvp5_sn000_malina2009	lars.stemmann@obs-vlfr.fr	28	UVP5 MALINA 2009	CC BY 4.0
17	uvp5_sn002_moose_ge_2014	lars.stemmann@obs-vlfr.fr	30	UVP5 MooseGE 2012 2013 2014 2015 2016	CC BY-NC 4.0
20	uvp5_sn000_operex2008	lars.stemmann@obs-vlfr.fr	110	UVP5 OPEREX 2008	CC BY 4.0
21	uvp5_sn001_2012_moose_ge	stemmann@obs-vlfr.fr	30	UVP5 MooseGE 2012 2013 2014 2015 2016	CC BY-NC 4.0
22	uvp5_sn001_2012_msm22	rkiko@geomar.de	593	UVP5 Geomar 2012 msm22	
23	uvp5_sn001_2012_msm23	rkiko@geomar.de	584	UVP5 Geomar 2012 msm23	
31	uvp5_sn002_moose_ge_2015_filtered	lars.stemmann@obs-vlfr.fr	30	UVP5 MooseGE 2012 2013 2014 2015 2016	CC BY-NC 4.0
32	uvp5_sn002_somba_ge_2014	lars.stemmann@obs-vlfr.fr	36	UVP5 SOMBA 2014	CC BY 4.0

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

41	uvp5_sn003_cassiopee_2015	lars.stemmann@obs-vlfr.fr	38	UVP5 CASSIOPEE 2015	CC BY 4.0
42	uvp5_sn003_ccelter_2014	tristan.biard@univ-littoral.fr	45	UVP5 CCELTER (Iter2008 ccelter_2012 ccelter_2014)	CC BY-NC 4.0
43	uvp5_sn003_ccelter_2016	tristan.biard@univ-littoral.fr	156	UVP5 CCELTER 2016	CC BY-NC 4.0
44	uvp5_sn003_dewex_spring_2013	lars.stemmann@obs-vlfr.fr	4	UVP5 DEWEX 2013 (spring)	CC BY 4.0
46	uvp5_sn003_outpace_2015	lguidi@obs-vlfr.fr	37	UVP5 OUTPACE 2015	CC BY 4.0
47	uvp5_sn003_sargasso_a	lombard@obs-vlfr.fr	22	UVP5 Sargasso 2014 (sargasso_a sargasso_b)	CC BY 4.0
48	uvp5_sn003_sargasso_b	lombard@obs-vlfr.fr	22	UVP5 Sargasso 2014 (sargasso_a sargasso_b)	CC BY 4.0
49	uvp5_sn003_tara2013	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
51	uvp5_sn003zp_tara2012	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
53	uvp5_sn005_dewex_2013_winter	lars.stemmann@obs-vlfr.fr	3	UVP5 DEWEX 2013 (winter)	CC BY 4.0
54	uvp5_sn005_dy032_2015_filtered	francois.carlotti@mio.osupytheas.fr	40	UVP5 DY032 205	CC BY 4.0
55	uvp5_sn005_moose_ge_2013	lars.stemmann@obs-vlfr.fr	30	UVP5 MooseGE 2012 2013 2014 2015 2016	CC BY-NC 4.0
56	uvp5_sn008_an1304	marcel.babin@takuvik.ulaval.ca	41	UVP5 AN Arctique (an1304 an1405 an1407) part A to validate	
57	uvp5_sn008_an1405	marcel.babin@takuvik.ulaval.ca	41	UVP5 AN Arctique (an1304 an1405 an1407) part A to validate	
58	uvp5_sn008_an1406	marcel.babin@takuvik.ulaval.ca	618	UVP5 AN 1406	CC BY-NC 4.0
59	uvp5_sn008_an1407	marcel.babin@takuvik.ulaval.ca	41	UVP5 AN Arctique (an1304 an1405 an1407) part A to validate	
60	uvp5_sn008_green_2015_icecamp	marcel.babin@takuvik.ulaval.ca	42	UVP5 GREEN EDGE Ice Camp 2015	CC BY-NC 4.0
61	uvp5_sn008_green_2016_icecamp	marcel.babin@takuvik.ulaval.ca	315	UVP5 GREEN EDGE Ice Camp 2016	CC BY-NC 4.0
62	uvp5_sn008_subice_2014	stemmann@obs-vlfr.fr	107	UVP5 Subice 2014	
64	uvp5_sn009_2015_p16n	amcdonnell@alaska.edu	43	uvp5_sn009_2015_p16n	CC BY 4.0
65	uvp5_sn009_2015_p16n_goa	amcdonnell@alaska.edu	43	uvp5_sn009_2015_p16n	CC BY 4.0
68	uvp5_sn010_2014_m105	rkiko@geomar.de	602	UVP5 Geomar 2014 m105	
69	uvp5_sn010_2014_m106	rkiko@geomar.de	595	UVP5 Geomar 2014 m106	

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

70	uvp5_sn010_2014_m107	rkiko@geomar.de	559	UVP5 Geomar 2014 m107	
71	uvp5_sn010_2014_m108	rkiko@geomar.de	586	UVP5 Geomar 2014 m108	
72	uvp5_sn010_2014_ps88b	rkiko@geomar.de	558	UVP5 Geomar 2014 ps88b	
73	uvp5_sn010_2015_m116	rkiko@geomar.de	589	UVP5 Geomar 2015 m116	
76	uvp5_sn010_2015_m121	rkiko@geomar.de	560	UVP5 Geomar 2015 m121	
81	uvp5_sn201_2016_naames_02	lee.karp-boss@maine.edu	146	UVP5hd NAAMES02	CC BY-NC 4.0
83	uvp5_sn201_ccelter_2017	tristan.biard@univ-littoral.fr	627	UVP5hd CCELTER 2017	CC BY-NC 4.0
86	uvp5_sn203_greenedge_2016	lars.stemmann@obs-vlfr.fr	149	UVP5hd GreenEdge 2016	CC BY-NC 4.0
87	uvp5_sn203_greenedge_2016_1b	lars.stemmann@obs-vlfr.fr	149	UVP5hd GreenEdge 2016	CC BY-NC 4.0
92	uvp5_sn201_2017_naames_03	lee.karp-boss@maine.edu	637	UVP5hd NAAMES03	CC BY-NC 4.0
95	uvp5_sn010_2017_m135	rkiko@geomar.de	548	UVP5 Geomar 2017 m135	
99	uvp5_sn000_tara2011	lars.stemmann@obs-vlfr.fr	579	UVP5 Tara Oceans (tara2009, tara2010, tara2011, tara2012, tara2013)	CC BY 4.0
105	uvp5_sn008_ips_amundsen_2018	marcel.babin@takuvik.ulaval.ca	1165	UVP5 IPS Amundsen 2018	CC BY 4.0
109	uvp5_sn010_2014_eddy	rkiko@geomar.de	881	UVP5 Geomar 2014 eddy	
166	uvp5_sn002_moose_dyf_2018	coppola@obs-vlfr.fr	35	UVP5 Dyfamed 2013 2014 2015 2016 2017 2018	
231	uvp5_sn202_msm060_filtered	andreas.rogge@awi.de, anya.waite@dal.ca	351	UVP5hd msm060 (2017)	Copyright
236	uvp5_sn201_2018_naames_04_filtered	lee.karp-boss@maine.edu	1252	UVP5hd Naames 04	CC BY-NC 4.0

532

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

533 **Table A2. Morphological features computed for the object (not the background). Areas and/or length units are expressed**
534 **in pixels.**

Feature	Description
objid	unique object identifier in EcoTaxa (integer number)
area	surface area of the object (integer number)
mean	average grey value within the object; sum of the grey values of all pixels in the object divided by the number of pixels
stddev	standard deviation of the grey value used to generate the mean grey value
mode	modal grey value within the object
min	minimum grey value within the object (0 = black)
max	maximum grey value within the object (255 = white)
perim	the length of the outside boundary of the object
width	width of the smallest rectangle enclosing the object
height	height of the smallest rectangle enclosing the object
major	primary axis of the best fitting ellipse for the object
minor	secondary axis of the best fitting ellipse for the object
angle	angle between the primary axis and a line parallel to the x-axis of the image
circ	circularity = (4 * Pi * Area) / Perim2) ; a value of 1 indicates a perfect circle, a value approaching 0 indicates an increasingly elongated polygon
feret	maximum feret diameter, i.e., the longest distance between any two points along the object boundary
intden	integrated density, corresponding to the sum of the grey values of the pixels within the object (i.e. = Area*Mean)
median	median grey value within the object
skew	skewness of the histogram of grey level values
kurt	kurtosis of the histogram of grey level values
%area	percentage of object's surface area that is comprised of holes, defined as the background grey level
area_exc	surface area of the holes in the object, in square pixels (=Area*(1-(%area/100))
fractal	fractal dimension of object boundary (Berube and Jebrak, 1999)
skelarea	surface area of skeleton in pixels. In a binary image, skeleton is obtained by repeatedly removing pixels from the edges of objects until they are reduced to the width of a single pixel
slope	slope of the grey level normalized cumulative histogram
histcum 1, 2, 3	grey level value at the first, second and third quartile of the normalized cumulative histogram of grey levels
nb1, nb2, nb3	number of remaining objects in the image after thresholding on level histcum1, 2 and 3
symetrieh	bilateral horizontal symmetry index
symetriev	bilateral vertical symmetry index
symetriehc	symmetry of the largest remaining object in relation to the horizontal axis after thresholding at the grey level histcum1 value
symetrievc	symmetry of the largest remaining object in relation to the vertical axis after thresholding at the grey level histcum1 value
convperim	the perimeter of the smallest polygon within which all points in the object fit
convarea	the area of the smallest polygon within which all points in the object fit
fcons	measure of contrast based in the texture feature descriptor (Amadasun and King, 1989)
thickr	thickness ratio: relation between the maximum thickness of an object and the average thickness of the object excluding the maximum

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

elongation	major/minor
range	grey max - grey min
meanpos	(mean-max)/(mean-min)
cv	100*(stddev/mean)
sr	100*(stddev/(max-min))
perimareaexc	perim/(sqrt(area_exc))
feretareaexc	feret/(sqrt(area_exc))
perimferet	perim/feret
perimmajor	perim/major
circex	(4*PI*area_exc)/(pow(perim,2))
kurt_mean	mean kurtosis of the histogram of grey level values
skew_mean	mean Skewness of the histogram of grey level values
convperim_perim	perimeter of the smallest polygon within which all points in the object fit
convarea_area	area of the smallest polygon within which all points in the object fit
symetrieh_area	symetrieh/area
symetriev_area	symetriev/area
nb1, nb2, nb3_area	nb1, nb2, nb3/area
nb1, nb2, nb3_range	nb1, nb2, nb3/range
median_mean	median/grey mean
median_mean_range	(median-grey mean)/range
skeleton_area	skelarea/area

535

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Open Access Earth System Science Data Discussions

536 **Table A3. List of annotators involved in the classification of objects within the current UVP5 dataset.**

Annotator name	Annotator name (*cont.*)	Annotator name (*cont.*)
Alikacem Yasmine	Gawinski Christine	Panchal Aakash
Andersson Bjoern	Gremion Gwenaelle	Panaïotis Thelma
Anhaus Philipp	Guérin Sébastien	Pelletier Noemie
Arboit Genevieve	Hasnain Sarah	Perhirin Margaux
Aubry Cyril	Hauss Helena	Perron Christophe
Bansept Marc-Antoine	Hintringer Moritz	Picheral Marc
Benoît-Gagné Maxime	Irisson Jean-Olivier	Pretty Jess
Berline Léo	Jalabert Laetitia	Prondzinsky Alannah Paulina
Berrenger Hugo	Karp Boss Lee	Reimer Jody
Biard Tristan	Kiko Rainer	Renaut Sophie
Blanc Benjamin	Lafond Augustin	Rogge Andreas
Blunck Frauke	Leguen Guillaume	Singh Akanksha
Bourdin Guillaume	Lekanoff Rachel	Singh Arvind
Caray-Counil Louis	Leroux Riwan	Spahic Susanne
Chawarski Julek	Levesque-Desrosiers Felix	Stemmann Lars
Christiansen Svenja	Li Juan	Suwaki Caroline
Courchet Lucas	Lombard Fabien	Toullec Jordan
Cram Jacob	Lopes Rubens	Trudnowska Emilia
Denoso Ferez Katty	Maps Frédéric	Vilgrain Laure
Donggyun Kim	Marec Claudie	Wauthy Maxime
Drago Laetitia	Mcdonnell Andrew M. P.	

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

Dullaert Emma	Migotto Alvaro
Elineau Amanda	Motreuil Solène
Faustmann Jannik	Nagata Renato
Fowler Victoria	Oliveira Bruna

537

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

538 **References**

539 Accardo, A., Laxenaire, R., Baudena, A., Speich, S., Kiko, R., & Stemmann, L. (2025). Intense and localized export
540 of selected marine snow types at eddy edges in the South Atlantic Ocean. *Biogeosciences*, *22*(5), 1183-1201.

541 Atherden, F., Slomska, A., & Manno, C. (2024). Sediment trap illustrates taxon-specific seasonal signals in Southern
542 Ocean zooplankton. Marine Biology, 171(9), 173.

543 Batten, S. D., Abu-Alhaija, R., Chiba, S., Edwards, M., Graham, G., Jyothibabu, R., ... & Wilson, W. (2019). A global
544 plankton diversity monitoring program. *Frontiers in Marine Science*, *6*, 321.

545 Barth, A., & Stone, J. (2022). Comparison of an in situ imaging device and net-based method to study
546 mesozooplankton communities in an oligotrophic system. *Frontiers in Marine Science*, *9*, 898057.

547 Barth, A., Walter, R. K., Robbins, I., & Pasulka, A. (2020). Seasonal and interannual variability of phytoplankton
548 abundance and community composition on the Central Coast of California. *Marine Ecology Progress Series*,
549 *637*, 29-43.

550 Biard, T., & Ohman, M. D. (2020). Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone
551 revealed by in situ imaging. *Limnology and Oceanography*, *65*(11), 2583-2602.

552 Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R. & Not,
553 F. (2016). In situ imaging reveals the biomass of giant protists in the global ocean, *Nature* 532(7600), 504–
554 507.

555 Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., & Weber, T. (2019). Multi-faceted particle pumps drive carbon
556 sequestration in the ocean. *Nature*, *568*(7752), 327-335.

557 Calbet, A. (2024). Sampling Plankton. In *The Wonders of Marine Plankton* (pp. 131-136). Cham: Springer Nature
558 Switzerland.

559 Chiba, S., Batten, S., Martin, C. S., Ivory, S., Miloslavich, P., & Weatherdon, L. V. (2018). Zooplankton monitoring
560 to contribute towards addressing global biodiversity conservation challenges. *Journal of plankton research*,
561 *40*(5), 509-518.

562 Christiansen, S. (2016). Distribution, eddy association and biogeochemical importance of the pelagic polychaete
563 *Poeobius sp.* in the tropical Atlantic (Doctoral dissertation, Christian-Albrechts-Universität Kiel).

564 Clayton, S., Alexander, H., Graff, J. R., Poulton, N. J., Thompson, L. R., Benway, H., ... & Martiny, A. (2022). Bio-
565 GO-SHIP: the time is right to establish global repeat sections of ocean biology. *Frontiers in Marine Science*,
566 *8*, 767443.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

567 Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the global ocean with biogeochemical-Argo. *Annual*
568 *review of marine science*, *12*(1), 23-48.

569 Clements, D. J., Yang, S., Weber, T., Mcdonnell, A. M. P., Kiko, R., Stemmann, L. & Bianchi, D. (2022). Constraining
570 the Particle Size Distribution of Large Marine Particles in the Global Ocean With In Situ Optical
571 Observations and Supervised Learning, *Global Biogeochemical Cycles* 36(5), e2021GB007276.

572 Clements, D. J., Yang, S., Weber, T., Mcdonnell, A. M. P., Kiko, R., Stemmann, L. & Bianchi, D. (2023). New
573 Estimate of Organic Carbon Export From Optical Measurements Reveals the Role of Particle Size
574 Distribution and Export Horizon, *Global Biogeochemical Cycles* 37(3), e2022GB007633.

575 Constable, A. J., Melbourne-Thomas, J., Corney, S. P., Arrigo, K. R., Barbraud, C., Barnes, D. K., ... & Ziegler, P.
576 (2014). Climate change and Southern Ocean ecosystems I: how changes in physical habitats directly affect
577 marine biota. *Global change biology*, *20*(10), 3004-3025.

578 Culverhouse, P. F. (2007). Natural object categorization: man versus machine. *Automated Taxon Identification in*
579 *Systematics: Theory, Approaches and Applications; MacLeod, N., Ed*, 25-46.

580 Culverhouse, P. F., Macleod, N., Williams, R., Benfield, M. C., Lopes, R. M., & Picheral, M. (2014). An empirical
581 assessment of the consistency of taxonomic identifications. *Marine Biology Research*, *10*(1), 73-84.

582 Dam, H. G., & Baumann, H. (2017). Climate change, zooplankton and fisheries. *Climate change impacts on fisheries*
583 *and aquaculture: A global analysis*, *2*, 851-874.

584 Davis, C. S., Thwaites, F. T., Gallager, S. M., & Hu, Q. (2005). A three-axis fast-tow digital Video Plankton Recorder
585 for rapid surveys of plankton taxa and hydrography. *Limnology and Oceanography: Methods*, *3*(2), 59-74.

586 D'Alelio, D., Libralato, S., Wyatt, T., & d'Alcala, M. R. (2016). *Ecological-network models link diversity, structure*
587 *and function in the plankton food-web. Sci. Rep. 6, 21806*.

588 Dexter, P., and Summerhayes, C.P., 2010, Ocean Observations - the Global Ocean Observing System (GOOS).
589 Chapter 11 in Pugh, D., and Holland, G., (eds.), Troubled Waters: Ocean Science and Governance. CUP,
590 Cambridge.161-178

591 Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., ... & Talley, L. D. (2012).
592 Climate change impacts on marine ecosystems. *Annual review of marine science*, *4*(1), 11-37.

593 Drago, L., Panaïotis, T., Irisson, J. O., Babin, M., Biard, T., Carlotti, F., ... & Kiko, R. (2022). Global distribution of
594 zooplankton biomass estimated by in situ imaging and machine learning. *Frontiers in Marine Science*, *9*,
595 894372.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

596 Dugenne, M., Corrales-Ugalde, M., Luo, J. Y., Kiko, R., O'Brien, T. D., Irisson, J. O., ... & Vilain, M. (2024). First
597 release of the Pelagic Size Structure database: global datasets of marine size spectra obtained from plankton
598 imaging devices. *Earth System Science Data*, *16*(6), 2971-2999.

599 Du Pontavice, H., Gascuel, D., Reygondeau, G., Maureaud, A., & Cheung, W. W. (2020). Climate change undermines
600 the global functioning of marine food webs. *Global Change Biology*, *26*(3), 1306-1318.

601 Dupouy, C., Frouin, R., Tedetti, M., Maillard, M., Rodier, M., Lombard, F., ... & Sempéré, R. (2018). Diazotrophic
602 Trichodesmium influence on ocean color and pigment composition in the South West tropical Pacific.
603 *Biogeosci. Discuss*, 1-43.

604 Fernández-Carrera, A., Kiko, R., Hauss, H., Hamilton, D. S., Achterberg, E. P., Montoya, J. P., ... & Subramaniam,
605 A. (2023). Nitrogen fixation rates in the Guinea Dome and the equatorial upwelling regions in the Atlantic
606 Ocean. *Biogeochemistry*, *166*(3), 191-210.

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

610 Giering, S. L., Culverhouse, P. F., Johns, D. G., McQuatters-Gollop, A., & Pitois, S. G. (2022). Are plankton nets a
611 thing of the past? An assessment of in situ imaging of zooplankton for large-scale ecosystem assessment and
612 policy decision-making. *Frontiers in Marine Science*, *9*, 986206.

613 Gorman, E. T., Kubalak, D. A., Patel, D., Mott, D. B., Meister, G., & Werdell, P. J. (2019, October). The NASA
614 Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission: an emerging era of global, hyperspectral Earth
615 system remote sensing. In *Sensors, systems, and next-generation satellites XXIII* (Vol. 11151, pp. 78-84).
616 SPIE.

617 Goswami, S. C. (2004). Zooplankton methodology, collection & identification. A field manual. *National Institute of*
618 *Oceanography,* 26.

619 Guidi, L., Calil, P. H., Duhamel, S., Björkman, K. M., Doney, S. C., Jackson, G. A., ... & Karl, D. M. (2012). Does
620 eddy-eddy interaction control surface phytoplankton distribution and carbon export in the North Pacific
621 Subtropical Gyre?. *Journal of Geophysical Research: Biogeosciences*, *117*(G2).

622 Hays, G. C., Richardson, A. J., & Robinson, C. (2005). Climate change and marine plankton. *Trends in ecology &*
623 *evolution*, *20*(6), 337-344.

624 Irisson, J. O., Ayata, S. D., Lindsay, D. J., Karp-Boss, L., & Stemmann, L. (2022). Machine learning for the study of
625 plankton and marine snow from images. *Annual Review of Marine Science*, *14*(1), 277-301.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

626 Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., De Vargas, C., Raes, J., ... & Tara Oceans Consortium. (2011). A
627 holistic approach to marine eco-systems biology. *PLoS biology*, *9*(10), e1001177.

628 Kiko R, Biastoch A, Brandt P, Cravatte S, Hauss H, Hummels R, Kriest I, Marin F, McDonnell AMP, Oschlies A,
629 Picheral M, Schwarzkopf FU, Thurnherr AM, Stemmann L (2017) Biological and physical influences on
630 marine snowfall at the equator. Nature Geoscience 10:852.Kiko, R., Picheral, M., Antoine, D., Babin, M.,
631 Berline, L., Biard, T., ... & Stemmann, L. (2022). A global marine particle size distribution dataset obtained
632 with the Underwater Vision Profiler 5. *Earth System Science Data Discussions*, *2022*, 1-37.

633 Kiko R, Lopes RM, Soviadan YD, Stemmann L (2023) Towards a distributed and operational pelagic imaging
634 network. Ocean Coast Res 71:e23058.

635 Laget, M., Drago, L., Panaïotis, T., Kiko, R., Stemmann, L., Rogge, A., ... & Biard, T. (2024). Global census of the
636 significance of giant mesopelagic protists to the marine carbon and silicon cycles. *Nature Communications*,
637 *15*(1), 3341.

638 Le Quéré, C., Buitenhuis, E. T., Moriarty, R., Alvain, S., Aumont, O., Bopp, L., ... & Vallina, S. M. (2016). Role of
639 zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles.
640 *Biogeosciences*, *13*(14), 4111-4133.

641 Llopis Monferrer, N., Biard, T., Sandin, M. M., Lombard, F., Picheral, M., Elineau, A., ... & Not, F. (2022). Siliceous
642 Rhizaria abundances and diversity in the Mediterranean Sea assessed by combined imaging and
643 metabarcoding approaches. *Frontiers in Marine Science*, *9*, 895995.

644 Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., ... & Appeltans, W. (2019). Globally consistent
645 quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, *6*, 196.

646 Lumini, A., & Nanni, L. (2019). Deep learning and transfer learning features for plankton classification. *Ecological*
647 *informatics*, *51*, 33-43.

648 Moriarty, R., & O'brien, T. D. (2013). Distribution of mesozooplankton biomass in the global ocean. *Earth System*
649 *Science Data*, *5*(1), 45-55.

650 Möller, K. O., John, M. S., Temming, A., Floeter, J., Sell, A. F., Herrmann, J. P., & Möllmann, C. (2012). Marine
651 snow, zooplankton and thin layers: indications of a trophic link from small-scale sampling with the Video
652 Plankton Recorder. *Marine Ecology Progress Series*, *468*, 57-69.

653 Muller-Karger, F. E., Miloslavich, P., Bax, N. J., Simmons, S., Costello, M. J., Sousa Pinto, I., ... & Geller, G. (2018).
654 Advancing marine biological observations and data requirements of the complementary essential ocean
655 variables (EOVs) and essential biodiversity variables (EBVs) frameworks. *Frontiers in Marine Science*, *5*,
656 211.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

657 Nocera A., Stemmann L., Babin M., Biard T., Coustenoble J., Carlotti F., Coppola L., Courchet L., Drago L., Elineau
658 A., Guidi L., Hauss H., Jalabert L., Karp-Boss L., Kiko R., Laget M., Lombard F., McDonnell A., Merland
659 C., Motreuil S., Panaïotis T., Picheral M., Rogge A., Waite A., Irisson J.O. (2025). A global consistent
660 database of plankton and detritus from in situ imaging by the Underwater Vision Profiler 5. SEANOE.
661 https://doi.org/10.17882/107583.

662 Panaïotis, T., Babin, M., Biard, T., Carlotti, F., Coppola, L., Guidi, L., ... & Irisson, J. O. (2023). Three major
663 mesoplanktonic communities resolved by in situ imaging in the upper 500 m of the global ocean. *Global*
664 *Ecology and Biogeography*, *32*(11), 1991-2005.

665 Perhirin, M., Vilgrain, L., Perrin, G., Lalande, C., Picheral, M., Maps, F., & Ayata, S. D. (2025). Identifying
666 zooplankton fecal pellets from in situ images. *Journal of Plankton Research*, *47*(1), fbae078.

667 Picheral, M., & Mériguet, Z. (2025). Description of the metadata and data issued from the Zooprocess and UVPapp
668 applications and imported or exported from the Ecotaxa application. Zenodo.
669 https://doi.org/10.5281/zenodo.14704251

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

673 Picheral, M., Colin, S., & J.-O., Irisson. (2017). *EcoTaxa, a tool for the taxonomic classification of images*.
674 http://ecotaxa.obs-vlfr.fr

675 Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The Underwater Vision Profiler
676 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton.
677 *Limnology and Oceanography: Methods*, *8*(9), 462-473.

678 Ratnarajah, L., Abu-Alhaija, R., Atkinson, A., Batten, S., Bax, N. J., Bernard, K. S., ... & Yebra, L. (2023). Monitoring
679 and modelling marine zooplankton in a changing climate. *Nature Communications*, *14*(1), 564.

680 Rohr, T., Richardson, A. J., Lenton, A., Chamberlain, M. A., & Shadwick, E. H. (2023). Zooplankton grazing is the
681 largest source of uncertainty for marine carbon cycling in CMIP6 models. *Communications Earth &*
682 *Environment*, *4*(1), 212.

683 Sandel, V., Kiko, R., Brandt, P., Dengler, M., Stemmann, L., Vandromme, P., ... & Hauss, H. (2015). Nitrogen fuelling
684 of the pelagic food web of the tropical Atlantic. *PLoS One*, *10*(6), e0131258.

685 Siegel, D., Burd, A., Estapa, M., Fields, E., Johnson, L., Romanelli, E., ... & Steinberg, D. (2024). Dynamics of
686 aggregates and sinking carbon fluxes in a turbulent ocean. In *PNAS Proceedings of the National Academy of*
687 *Sciences of the United States of America*. National Academy of Sciences.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

688 Soviadan, Y. D., Beck, M., Habib, J., Baudena, A., Drago, L., Accardo, A., Laxenaire, R., Speich, S., Brandt, P., Kiko,
689 R., and Stemmann, L.: Marine snow morphology drives sinking and attenuation in the ocean interior,
690 EGUsphere (accepted), https://doi.org/10.5194/egusphere-2024-3302, 2025.

691 Soviadan, Y. D., Dugenne, M., Drago, L., Biard, T., Trudnowska, E., Lombard, F., ... & Stemmann, L. (2024).
692 Combining in situ and ex situ plankton image data to reconstruct zooplankton (> 1 mm) volume and mass
693 distribution in the global ocean. *Journal of Plankton Research*, *46*(5), 461-474.

694 Soviadan, Y. D., Benedetti, F., Brandão, M. C., Ayata, S. D., Irisson, J. O., Jamet, J. L., ... & Stemmann, L. (2022).
695 Patterns of mesozooplankton community composition and vertical fluxes in the global ocean. *Progress in*
696 *Oceanography*, *200*, 102717.

697 Steinberg, D. K., & Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Annual review of marine science*,
698 *9*(1), 413-444.

699 Stemmann, L., Youngbluth, M., Robert, K., Hosia, A., Picheral, M., Paterson, H., ... & Gorsky, G. (2008). Global
700 zoogeography of fragile macrozooplankton in the upper 100–1000 m inferred from the underwater video
701 profiler. *ICES Journal of Marine Science*, *65*(3), 433-442.

702 Stemmann, L., & Boss, E. (2012). Plankton and particle size and packaging: from determining optical properties to
703 driving the biological pump. *Annual Review of Marine Science*, *4*(1), 263-290.

704 Suthers, I. M., Redden, A. M., Bowling, L., Kobayashi, T., & Rissik, D. (2019). Plankton processes and the
705 environment. *Plankton. A guide to their ecology and monitoring for water quality*, *1*, 21-35.

706 Titocci, J., Pata, P. R., Durazzano, T., Ayata, S. D., Clerc, C., Cornils, A., ... & Hunt, B. P. (2025). Pathways for
707 converting zooplankton traits to ecological insights are paved with findable, accessible, interoperable, and
708 reusable (FAIR) data practices. *ICES Journal of Marine Science*, *82*(2), fsaf017.

709 Trudnowska, E., Dragańska-Deja, K., Sagan, S., & Błachowiak-Samołyk, K. (2022). Cells of matter and life–towards
710 understanding the structuring of particles and plankton patchiness in the Arctic fjords. *Frontiers in Marine*
711 *Science*, *9*, 909457.

712 Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress*
713 *in Oceanography*, *130*, 205-248.

714 Vilgrain, L., Maps, F., Picheral, M., Babin, M., Aubry, C., Irisson, J. O., & Ayata, S. D. (2021). Trait-based approach
715 using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone.
716 *Limnology and Oceanography*, *66*(4), 1155-1167.

https://doi.org/10.5194/essd-2025-522
Preprint. Discussion started: 29 August 2025

717 Yebra, L., Puerto, M., Valcárcel-Pérez, N., Putzeys, S., Gómez-Jakobsen, F., García-Gómez, C., & Mercado, J. M.
718 (2022). Spatio-temporal variability of the zooplankton community in the SW Mediterranean 1992–2020:
719 Linkages with environmental drivers. *Progress in Oceanography*, *203*, 102782.