Open Access

Earth System
Science
Data

Metazoan zooplankton in the Bay of Biscay: a 16-year record of individual sizes and abundances obtained using the ZooScan and ZooCAM imaging systems

Nina Grandremy[1],★, Paul Bourriau[1], Edwin Daché[2], Marie-Madeleine Danielou[3], Mathieu Doray[1], Christine Dupuy[4], Bertrand Forest[5], Laetitia Jalabert[6], Martin Huret[7], Sophie Le Mestre[7], Antoine Nowaczyk[8], Pierre Petitgas[9], Philippe Pineau[4], Justin Rouxel[10], Morgan Tardivel[10], and Jean-Baptiste Romagnan[1],★

[1]DECOD (Ecosystem Dynamics and Sustainability), IFREMER, INRAE, Institut Agro, Nantes, Centre Atlantique, Rue de l'Ile d'Yeu, BP 21105, 44311 Nantes CEDEX 03, France
[2]Unité Biologie et Ecologie des Ecosystèmes marins Profonds, Laboratoire Environnement Profond, Ifremer Centre Bretagne, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France
[3]Unité DYNECO-PELAGOS, Laboratoire d'Ecologie Pélagique, Ifremer Centre Bretagne, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France
[4]BIOFEEL, UMRi LIENSs, La Rochelle University/CNRS, 2, rue Olympe de Gouges, 17000 La Rochelle, France
[5]Laboratoire Hydrodynamique Marine, Unité RDT, Ifremer Centre Bretagne, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France
[6]Sorbonne Université, Institut de la Mer de Villefranche, 06230 Villefranche-sur-mer, France
[7]DECOD (Ecosystem Dynamics and Sustainability), IFREMER, INRAE, Institut Agro, Centre Bretagne, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France
[8]UMR CNRS 5805 EPOC–OASU, Station Marine d'Arcachon, Université de Bordeaux, 2 Rue du Professeur Jolyet, 33120 Arcachon, France
[9]Departement Ressources Biologiques et Environnement, Ifremer Centre Atlantique, Rue de l'Ile d'Yeu, BP 21105, 44311 Nantes CEDEX 03, France
[10]Laboratoire Détection, Capteurs et Mesures, Unité RDT, Ifremer Centre Bretagne, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France
★These authors contributed equally to this work.

Correspondence: Nina Grandremy (grandremy.n@gmail.com) and Jean-Baptiste Romagnan (jean.baptiste.romagnan@ifremer.fr)

Abstract. This paper presents two metazoan zooplankton datasets obtained by imaging samples collected on the Bay of Biscay continental shelf in spring during the PELGAS (PELagique GAScogne) integrated surveys over the 2004–2019 period. The samples were collected at night with a 200 µm mesh-size WP2 net fitted with a Hydrobios (back-run stop) mechanical flowmeter and hauled vertically from the sea floor to the surface, with the maximum depth set at 100 m when the bathymetry was deeper than this. The first dataset originates from samples collected from 2004 to 2016 and imaged on land with the ZooScan and is composed of 1 153 507 imaged and measured objects. The second dataset originates from samples collected from 2016 to 2019 and imaged onboard the R/V *Thalassa* with the ZooCAM and is composed of 702 111 imaged and measured objects. The imaged objects are composed of zooplankton individuals, zooplankton pieces, non-living particles and imaging artefacts ranging from 300 µm to 3.39 mm in equivalent spherical diameter which were individually imaged,

measured and identified. Each imaged object is geolocated and associated with a station, a survey, a year and other metadata. Each object is described by a set of morphological and grey-level-based features (8 bit encoding, 0 = black, 255 = white), including size, that were automatically extracted from each individual image. Each object was taxonomically identified using the web-based application Ecotaxa with built-in random-forest and CNN-based semi-automatic sorting tools, which was followed by expert validation or correction. The objects were sorted into 172 taxonomic and morphological groups. Each dataset features a table combining metadata and data at individual-object granularity from which one can easily derive quantitative population and community descriptors such as abundances, mean sizes, biovolumes, biomasses and size structure. Each object's individual image is provided along with the data. These two datasets can be used in combination for ecological studies, as the two instruments are interoperable, or they can be used as training sets for ZooScan and ZooCAM users. The data presented here are available at the SEANOE dataportal: https://doi.org/10.17882/94052 (ZooScan dataset, Grandremy et al., 2023c) and https://doi.org/10.17882/94040 (ZooCAM dataset, Grandremy et al., 2023d).

1   Introduction

Metazoan planktonic organisms, hereafter referred to as zooplankton, encompass an immense diversity of life forms which have successfully colonized the entire ocean, from eutrophic estuarine shallow areas to the oligotrophic open ocean and from the sunlit ocean to hadal depths. Their body sizes span 5 to 6 orders of magnitude in length: from µm to tens of metres (Sieburth and Smetacek, 1978). Zooplankton plays a pivotal role in marine ecosystems (Banse, 1995). It transfers the organic matter produced in the epipelagic domain by photosynthesis to the deeper layers of the ocean (Siegel et al., 2016) by producing fast-sinking aggregates (Turner, 2015) and by diel vertical migration (Steinberg et al., 2000; Ohman and Romagnan, 2016). Zooplankton therefore participates in mitigating the anthropogenic carbon dioxide buildup in the atmosphere that is responsible for climate change. Moreover, zooplankton is an exclusive trophic resource for commercially important fish during their larval stage, so a shift in zooplankton species or phenology can have dramatic effects on recruitment (e.g. for North Sea cod; Beaugrand et al., 2003). In addition, it is a major trophic resource for adult small planktivorous pelagic fish known as forage fish (van der Lingen, 2006). Recent studies suggest that zooplankton dynamics may have a significant effect on small pelagic fish population dynamics and individual body condition (Brosset et al., 2016; Menu et al., 2023) and therefore impact wasp-waist ecosystem-based fisheries and socio-ecosystems that are dependent on those fisheries worldwide (Cury et al., 2000).

Despite zooplankton being of such global importance in both climate change effects on ecosystems and the management of fisheries (Chiba et al., 2018; Lombard et al., 2019), it is still technically difficult to monitor compared to other marine ecological compartments. Zooplankton biomass, diversity and spatio-temporal distributions cannot be estimated from spaceborne sensors, unlike those of phytoplankton (Uitz et al., 2010), commercial exploitation data of zooplankton do not exist yet, unlike the corresponding data for fish. One noticeable exception is the CPR Survey network, which enables zooplankton data generation at spatio-temporal scales that are fine enough to study climate change and diversity-related zooplanktonic processes (Batten et al., 2019). Yet, generating zooplankton data often requires dedicated surveys at sea, specific sampling instruments and trained taxonomic analysts. Moreover, besides actual observation, modelling zooplankton remains a challenging task due to the diversity of traits, such as life forms, life cycles, body sizes, and physiological processes, exhibited by zooplankton (Mitra and Davis, 2010; Mitra et al., 2014). However, over the past 2 decades, the development of imaging and associated machine-learning semi-automatic identification tools (Irisson et al., 2022) has greatly improved the capability of scientists to analyse long (Feuilloley et al., 2022), high-frequency (Romagnan et al., 2016) or spatially resolved (Grandremy et al., 2023a) zooplankton time series as well as trait-based data (Orenstein et al., 2022). Imaging and machine learning have particularly enabled the increased development of combined size and taxonomy zooplankton ecological studies (e.g. Vandromme et al., 2014; Romagnan et al., 2016; Benedetti et al., 2019). Yet, the use of these machine-learning tools is not trivial because they require abundant, scientifically qualified, sensor-specific training image data (i.e. a learning set and test set; Irisson et al., 2022) and complex hardware and software setups (Panaïotis et al., 2022). One good example of such an image dataset is the ZooScan-Net dataset (Elineau et al., 2018), which features an extensive ZooScan (Gorsky et al., 2010) imaging dataset usable as a training set for ecologists as well as for imaging and machine-learning scientists.

The objective of this paper is to present two freely available zooplankton imaging datasets originating from two different instruments, the ZooScan (Gorsky et al., 2010) and the ZooCAM (Colas et al., 2018). These datasets originate from the PELGAS (PELagique GAScogne) integrated survey in the Bay of Biscay (Doray et al., 2018a), a continental shelf ecosystem supporting major European fisheries (ICES, 2021). Combined, these datasets make up a 16-year

time series of sized and taxonomically resolved zooplankton, along with context metadata allowing the calculation of quantitative data, covering the whole Bay of Biscay continental shelf from the French coast to the continental slope and from the Basque Country to southern Brittany in spring. These datasets can be used for ecological studies (Grandremy et al., 2023a), machine-learning studies and modelling studies.

2 Methods

2.1 Sampling

Zooplankton samples were collected during successive PELGAS integrated surveys carried out over the Bay of Biscay (BoB) French continental shelf every year in spring from 2004 to 2019 onboard the R/V *Thalassa*. The aim of this survey is to assess small pelagic fish biomass and monitor the pelagic ecosystem to inform ecosystem-based fisheries management. Fish data, hydrology, phyto- and zooplankton samples, and megafauna sightings (marine mammals and seabirds) are concomitantly collected to build long-term spatially resolved time series of the BoB pelagic ecosystem. The PELGAS sampling protocols combine daytime en-route data collection (small pelagic fish and megafauna) with night-time depth-integrated hydrology and plankton sampling at fixed points. Detailed PELGAS survey protocols can be found in Doray et al. (2018a, 2021). The PELGAS survey datasets providing hydrological, primary producers, fish and megafauna data are available as gridded data in the SEANOE data portal (Doray et al., 2018b) at the following link: https://doi.org/10.17882/53389.

The number of zooplankton samples obtained per year varied between 41 (2005) and 65 (2019) due to adjustments in the sampling strategy and the weather conditions, with 889 zooplankton samples collected in total. From 2004 to 2006, samples were collected in the southern Bay of Biscay up to the Loire estuary only (Fig. 1). Sampling was carried out in vertical tows performed during the night using a 200 µm mesh-size WP2 net, generally from 100 m depth (or 5 m above the seabed) to the surface. In 2004 and 2005, the targeted maximum sampling depth was 200 m. In 2004, 15 samples were collected deeper than 100 m, among which 11 were deeper than 120 m; in 2005, 20 samples were collected deeper than 100 m, among which 13 were deeper than 120 m. Before 2014, the sampled water volume was estimated by multiplying the cable length by the net opening surface ($0.25 \, \text{m}^2$), whereas the net has been equipped with a Hydrobios back-run stop flowmeter since 2014. The samples originating from the 2004 to 2016 surveys were preserved in 4 % formaldehyde (final concentration) and analysed on land in the laboratory with the ZooScan, while they have been analysed live onboard with the ZooCAM since 2016.

2.2 Sample processing and analyses
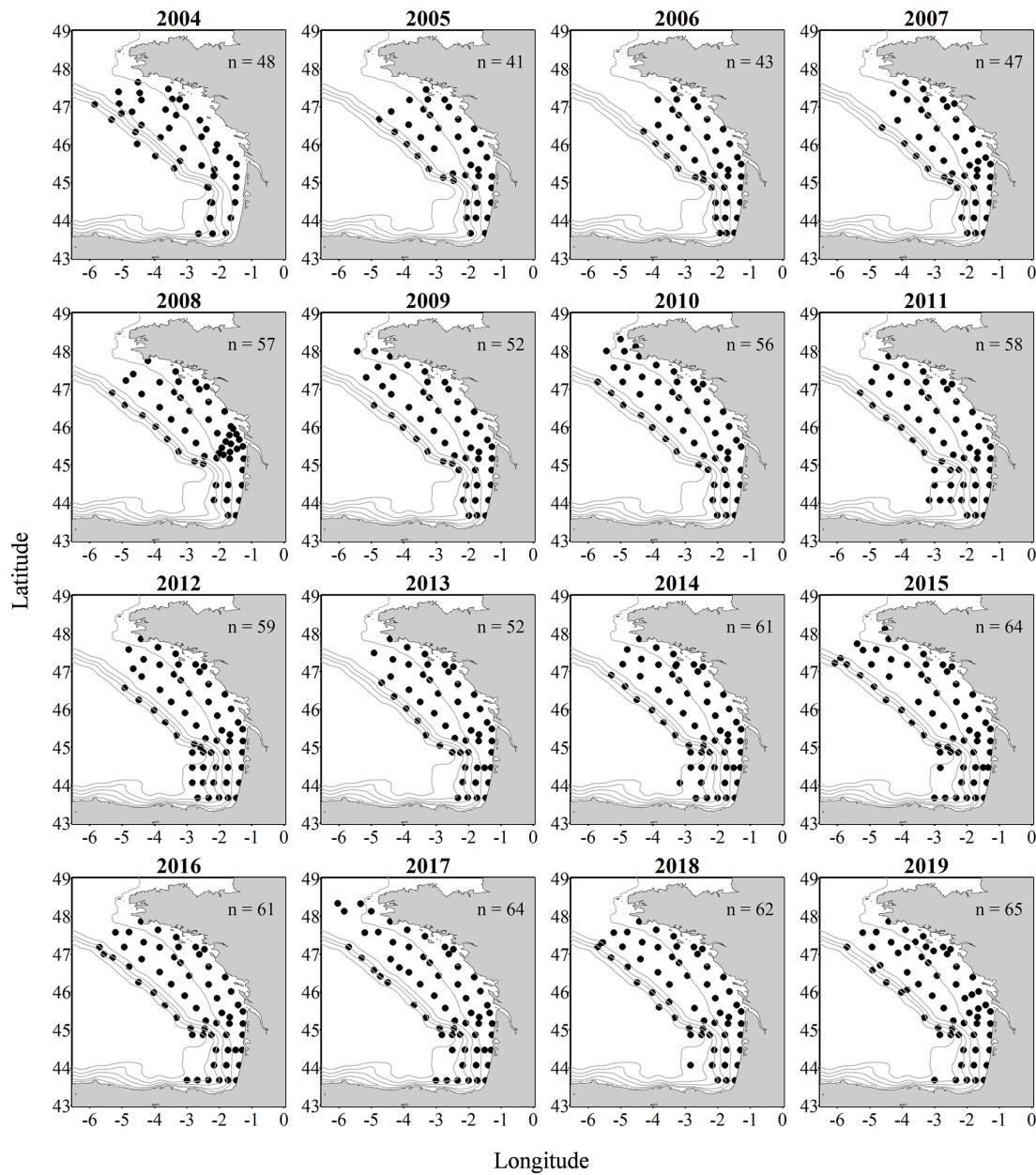
2.2.1 Digitization with the ZooScan

Preserved samples were digitized with the ZooScan (Gorsky et al., 2010), a flatbed scanner generating 16 bit grey-level high-resolution images (2400 dpi, pixel size: 10.56 µm, image size: 15×24 cm equivalent to $14\,200 \times 22\,700$ pixels). It is well suited for the imaging of preserved organisms ranging in size from 300 µm to several centimetres. The ZooScan is run by the custom-made, ImageJ-based, ZooProcess software, which generates a single large image for each scan that contains up to 2000 organisms, depending on the size of the imaged organisms.

Prior to digitization, the seawater and formaldehyde solution was filtered through a 180 µm mesh sieve into a trash tank under a fume hood. The organisms were then gently but thoroughly rinsed with freshwater over the tank in the sieve. They were then size fractionated with a 1 mm sieve into organisms larger and smaller than 1 mm. This size-splitting step is recommended when using the ZooScan as it addresses the possible bias due to the underrepresentation of large objects caused by the necessary subsampling. Each size fraction was subsampled separately with a Motoda splitter to obtain two subsamples containing 500–1000 objects for the large-organism size fraction and 1000–2000 objects for the small-organism size fraction. To mitigate the number of overlapping objects, each subsample was imaged after the manual separation of objects on the scanning tray, as recommended in Vandromme et al. (2012). Overall, 699 samples were digitized following this protocol, corresponding to 1397 scans (one sample was not divided into size fractions as it did not contain organisms larger than 1 mm).

2.2.2 Digitization with the ZooCAM

The ZooCAM is an in-flow imaging instrument designed to digitize preserved as well as live zooplankton samples onboard, immediately after net collection (Colas et al., 2018). The ZooCAM features a cylindrical transparent tank in which the zooplankton sample is mixed with filtered seawater. Depending on the richness of the sample and the subsampling (if necessary), the volume of seawater can be adjusted between 2–7 L. The organisms were pumped at $1 \, \text{L} \, \text{min}^{-1}$ from the tank to a flow cell inserted between a CCD camera (pixel size: 10.3 µm) and a red LED flashing device, where they were imaged at 16 fps. Given the flow cell volume, the size of the field of view, the imaging frequency and the flow rate, all the seawater volume containing the organisms was imaged (Colas et al., 2018). Before all the initial volume was imaged, the tank and the tubing were carefully and thoroughly rinsed with filtered seawater to ensure the imaging of all the organisms poured into the tank. For each sample, the ZooCAM generates a stack of small-size ($\sim 1 \, \text{MB}$) raw images that are subsequently analysed with the ZooCAM software. Depending on the initial water content of the tank and

Figure 1. Metazoan zooplankton sampling locations during the PELGAS cruises in the Bay of Biscay from 2004 to 2019. The years with the poorest coverage are 2005 and 2006, with 41 and 43 sampling stations respectively, and the years with the best coverage are 2015, 2017 and 2019, with 64, 64 and 65 sampling stations respectively.

the rinsing, a ZooCAM run can generate up to 10 000 raw images from which the individual organism vignettes will be extracted. A ZooCAM run on a live sample often generates up to 5000–10 000 vignettes of individual organisms. It is very important to subsample the initial samples with a dichotomic splitter (a Motoda splitter was used here) to ensure that the object concentration in subsamples is low enough to reduce the risk of imaging overlapping objects and avoid any dependency on the water volume imaged when reconstructing quantitative estimates of zooplankton, as the initial and

rinsing volumes are variable. Overall, 190 samples were digitized live onboard with the ZooCAM.

2.3 Image processing

Both instruments generate grey-level working images (8 bit encoding, 0 = black, 255 = white). In both cases, image processing consisted of (i) a "physical" background homogenization in which an empty background image was subtracted from each sample image (one for ZooScan and as

many as there were raw images for ZooCAM), (ii) a thresholding of each raw image (the threshold value was 243 for ZooScan and 240 for ZooCAM), and (iii) the segmentation of each object imaged. The ZooProcess software was set to detect and segment objects with an area equal to or larger than 631 pixels, whereas the ZooCAM software was set to detect objects with an area equal to or larger than 667 pixels, which in both cases equals 300 µm equivalent spherical diameter (ESD), or a biovolume of 0.014 mm^3 (using a spherical biovolume model; Vandromme et al., 2012).

Morphological features were then extracted for each detected object. Features generated by the ZooScan are defined in Gorsky et al. (2010) and those generated by the ZooCAM are defined in Colas et al. (2018). ZooScan images were processed with the ZooProcess v7.39 (4 October 2020) open-source software. ZooCAM images were processed with the proprietary ZooCAM custom-made software, which uses the MIL (Matrox Imaging Library, Dorval, Québec, Canada) as the individual-object processing kernel. Each detected object was finally cropped from the working sample images and saved as a unique labelled vignette in a sample-specific folder, along with a sample-specific single text file containing the object features arranged as a table with objects arranged in lines and features in columns.

2.4 Touching objects

ZooProcess features a tool that enables the digital separation of possibly touching objects in the final image dataset for each sample. As touching objects may impair abundance and size-structure estimations (Vandromme et al., 2012), remaining touching objects were searched for in the individual vignettes from the ZooScan and were digitally manually separated with the ZooProcess separation tool to improve the quality of further identifications, counts and the size structure of zooplankton. The ZooCAM software does not offer such a tool.

2.5 Taxonomic identification of individual images

All individual vignettes from both instruments were sorted into two instrument-specific separated sets and identified with the help of the online application Ecotaxa (Picheral et al., 2017). Ecotaxa features a random forest algorithm (Breiman, 2001) and a series of instrument-specific tuned spatially sparse convolutional neural networks (Graham, 2014) that were used in a combined approach to predict identifications of unidentified objects. First, an automatic classification of non-identified individual vignettes into coarse zooplankton and non-zooplankton categories was carried out. In both cases (ZooScan and ZooCAM), Ecotaxa hosted instrument-specific image datasets, previously curated and freely available, that were used as initial learning sets. These initial classifications were then visually inspected, manually validated or corrected when necessary, and taxonomically re-

fined when possible. After a few thousand images had been validated in each project, they were used as dataset-specific learning sets to improve the initial coarse automatic identifications. This process was iterated until all the individual vignettes were classified up to their maximum reachable level of taxonomical detail. A subsequent quality check of automatic taxonomic identifications was realized in a two-step process: first, a complete review (validation and/or correction) of all individual automatic identifications was done by Nina Grandremy and Jean-Baptiste Romagnan; then, trained experts reviewed and curated the ZooScan and the ZooCAM datasets (Laeticia Jalabert handled the ZooScan dataset and Antoine Nowaczyk handled the ZooCAM dataset) at the individual level. Although some identification errors may still remain in the datasets, we consider this double check process to be sufficient to provide taxonomically qualified data.

2.6 Intercalibration of the two instruments

The two datasets are usable separately. However, considered together, they form a 16-year-long spatio-temporal time series. A comparison study was done to ensure these datasets are homogeneous and can thus be combined for ecological studies (Grandremy et al., 2023b). All the zooplankton samples from year 2016 (61 sampling stations over the whole BoB continental shelf) were imaged with both instruments. In brief, all non-zooplankton and touching-object images were removed from the initial datasets. Then, the interoperable size range was determined with an assessment based on a comparison of the normalized biovolume–size spectra (NB-SSs) for the instruments. This size interval ranges between 0.3–3.39 mm ESD. Finally, the zooplankton communities as seen by the ZooScan and the ZooCAM were compared by taxa and by station using 27 taxonomic groups. Poorly represented taxa as well as non-taxonomically identified objects were not taken into account in the computation of zooplankton variables and in community structure analyses. Both instruments showed similar NB-SS slopes for 58 out of 61 stations and depicted equivalent abundances, biovolumes and mean organism sizes as well as similar community compositions for a majority of the sampling stations. They also estimated similar spatial patterns of the zooplankton community at the scale of the Bay of Biscay. However, some taxonomic groups showed discrepancies between instruments, which originate from differences in the sample preparation protocols before image acquisition, the imaging techniques and quality, and whether the samples were imaged live or fixed. For example, the mineralized protists (Rhizaria here) dissolve in formalin and are considered to be underestimated in preserved seawater samples (Biard et al., 2016). Also, the random orientation of objects in the ZooCAM flow cell leads to a loss of taxonomic identification accuracy due to the difficulty in spotting the specific features needed for the identification (Colas et al., 2018; Grandremy et al., 2023b). This is particularly acute for copepods, for

which the ZooScan seems to provide better identification capabilities for experts, as the organisms are imaged in a lateral view most of the time, whereas the ZooCAM often images them in a non-lateral, randomly oriented view, preventing the visualization of specific features. A detailed discussion about how to explain the discrepancies between the ZooScan and the ZooCAM can be found in Grandremy et al. (2023b). We assume that the two presented datasets form a single 16-year-long spatio-temporal time series of abundances (Fig. 2) and sizes of zooplanktonic organisms (Fig. 3) from which biovolumes, biomasses, the Shannon index (Fig. 4) and the zooplankton community size structure can be derived (Vandromme et al., 2012).

3 Datasets

3.1 Taxonomic groups and operational morphological groups

The ZooScan dataset is composed of 1 153 507 zooplankton individuals, zooplankton parts, non-living particles and imaging artefacts individually imaged and measured with the ZooScan and ZooProcess (Gorsky et al., 2010) and sorted into 127 taxonomic and morphological groups. The ZooCAM dataset is composed of 702 111 zooplankton individuals, zooplankton parts, non-living particles and imaging artefacts individually imaged and measured with the ZooCAM (Colas et al., 2018) and sorted into 127 taxonomic and morphological or life-stage groups. The total number of different groups identified with both instruments combined is 170, among which 84 are common to both datasets (Table 1), 43 belong to the ZooScan dataset only and 43 others belong to the ZooCAM dataset only (Table 2). The identified groups were divided into actual taxa and operational morphological groups (OMGs). Typically, OMGs are either non-adult life stages of taxa, aggregated morphological groups, or non-living groups (see Tables 1 and 2). Among the groups common to both instruments, 45 are actual taxa and 39 are OMGs (Table 1). Among the ZooScan-only groups, 22 are taxa and 21 are OMGs, and among the ZooCAM-only groups, 18 are taxa and 25 are OMGs (Table 2).

The differences in identified groups, in the taxa/OMGs ratio and in the associated counts arose from several aspects of the data generation. Firstly, the two imaging methods differ in their technical set-up. The main difference is that, on the one hand, fixed organisms are laid down and arranged manually on the imaging sensor and digitized in a lab (i.e. steady 2-D) set-up when using the ZooScan, whereas organisms are imaged live in a moving fluid in a 3-D environment (the flow cell) onboard when digitized with the ZooCAM. Their position in front of the camera may not enable an identification as precise as that achieved when they are laid on the scanner tray (Grandremy et al., 2023b; Colas et al., 2018). Secondly, the datasets are sequential in time: the ZooCAM dataset follows the ZooScan's. Zooplankton communities in the Bay of

Biscay may have changed over time, even if their biomass as aggregated groups shows remarkable space-time stability (Grandremy et al., 2023a). Thirdly, we cannot guaranty that there is no adverse effect on taxonomic identification, as validation involved several experts (Culverhouse, 2007). Although we paid great attention to homogenizing the final detailed datasets, we recommend that taxa and OMGs should be aggregated and the biological resolution should be reduced for ecological studies (Grandremy et al., 2023a, b). Additionally, numerous identified and sorted taxa and OMGs do not belong to the metazoan zooplankton or they are non-adult life stages or parts of organisms. Those were included in the presented datasets because they are always found in natural samples. They need to be separated from entire organisms to ensure that abundance estimations are as accurate as possible, and they must be taken into account to ensure accurate biovolume or biomass estimations. A good example is the siphonophore issue: numerous swimming bells of degraded siphonophore individuals can be found and imaged in a sample. Determining an accurate siphonophore abundance may not be easy, but this could be overcome by considering the biovolume or biomass of siphonophores by adding up the biovolumes or biomasses of the numerous parts of the organisms imaged.

OMG names mainly take the form of two words separated by "_" (underscore). Although we tried to name them as explicitly as possible, a few potentially needed clarifications can be found in Table 3.

3.2 Data and images

3.2.1 Data
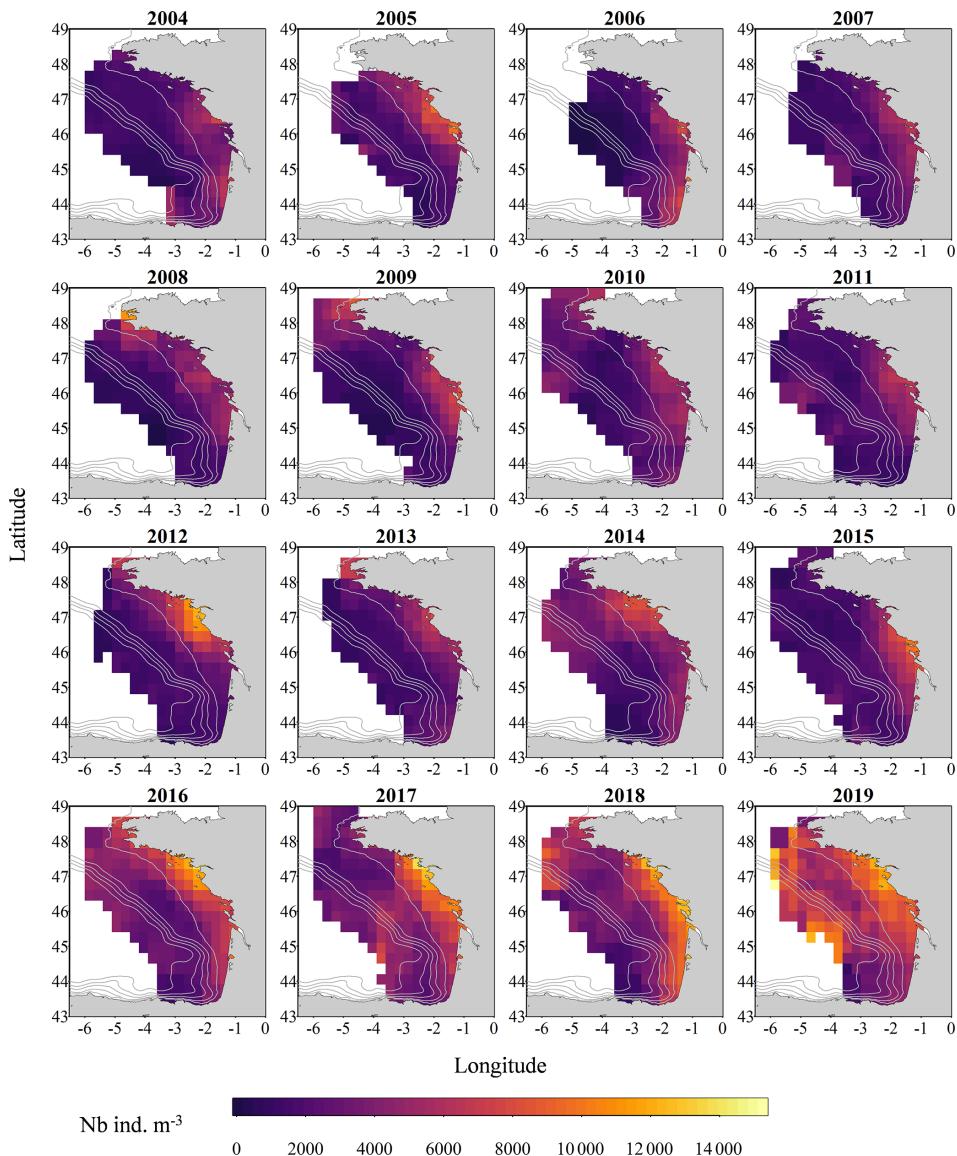
The data are divided into two datasets available as tab-separated files, one for each instrument. Within each dataset, the data is organized as a table containing text data as well as numerical data. Each dataset contains a combination of actual data and metadata at individual-object granularity. For each object, the user can find descriptors originating from the image processing (i.e. features), sampling metadata (i.e. latitude and longitude of the sampling station, date and time of sampling, sampling device, etc.), and sample processing metadata (i.e. subsampling factor, seawater sampled volume, pixel size) in columns and individual objects in lines. The column headers are defined in Tables A1 and A2 for the ZooCAM and ZooScan datasets respectively. The following prefixes enable the segregation of types of data and metadata: (i) "object_", which identifies variables assigned to each object individually; (ii) "sample_", which identifies variables assigned to each sample; (iii) "acq_", which identifies variables assigned to each data acquisition for the same sample (note that this type of variable is only found in the ZooScan dataset, as ZooScan samples were split into two size fractions corresponding to two acquisitions); and (iv) "process_", which identifies variables describing key image processing

Table 1. Taxa and operational morphological groups (OMGs) common to the ZooCAM and ZooScan datasets. Taxa are listed on the left of the table and OMGs are listed on the right of the table. OMG names are spelled as they appear in the datasets. The numbers next to each taxon and OMG are the count and the percentage (%) for each category for each instrument in the whole dataset. Non-zooplanktonic OMGs are highlighted in bold, and genera and species are formatted in italics.

	ZooCAM		ZooScan			ZooCAM		ZooScan	
Taxon	Counts	%	Counts	%	OMG	Counts	%	Counts	%
Calanoida	137 536	19.58	149 956	13.00	**detritus**	105 751	15.06	219 541	19.03
Oithonidae	112 977	16.09	110 510	9.58	*diatoma*	36 842	5.25	1084	0.09
Acartiidae	30 403	4.33	66 353	5.75	**bubble**	32 563	4.64	1112	0.10
Temoridae	13 520	1.93	31 335	2.72	**Noctiluca_Noctilucaceae**	22 165	3.16	20 784	1.80
Oncaeidae	11 843	1.69	34 651	3.00	other_living	15 029	2.14	5861	0.51
Calanidae	9578	1.36	91 513	7.93	dead_copepoda	13 383	1.91	17 151	1.49
Limacinidae	8966	1.28	6423	0.56	**fiber_detritus**	13 379	1.91	25 124	2.18
Appendicularia	6724	0.96	34 027	2.95	nauplii_cirripedia	6766	0.96	6008	0.52
Cladocera	5590	0.80	18 213	1.58	gonophore_diphyidae	4395	0.63	1462	0.13
Centropagidae	4592	0.65	14 651	1.27	multiple_copepoda	3740	0.53	961	0.08
Neoceratium	2984	0.43	4830	0.42	nauplii_crustacea	3422	0.49	10 747	0.93
Euchaetidae	2643	0.38	12 957	1.12	**artefact**	2643	0.38	60 718	5.26
Metridinidae	2333	0.33	15 081	1.31	multiple_other	1928	0.27	10 303	0.89
Corycaeidae	2021	0.29	4720	0.41	pluteus_echinodermata	1623	0.23	1441	0.12
Euterpina	1043	0.15	2870	0.25	calyptopsis_euphausiacea	1396	0.20	3246	0.28
Euphausiacea	889	0.13	1195	0.10	bivalvia_mollusca	1324	0.19	3766	0.33
Calocalanus	820	0.12	1196	0.10	bract_diphyidae	1315	0.19	386	0.03
Chaetognatha	624	0.09	7274	0.63	cypris	862	0.12	2363	0.20
Harpacticoida	481	0.07	1697	0.15	nectophore_diphyidae	839	0.12	14 389	1.25
Obelia	459	0.07	1016	0.09	**egg_actinopterygii**	768	0.11	3596	0.31
Annelida	256	0.04	2434	0.21	tail_appendicularia	753	0.11	11 349	0.98
Decapoda	173	0.02	471	0.04	cyphonaute	684	0.10	2218	0.19
Microsetella	116	0.02	1169	0.10	eudoxie_diphyidae	501	0.07	69	0.01
Phoronida	90	0.01	163	0.01	larvae_echinodermata	483	0.07	2200	0.19
Actinopterygii	85	0.01	2113	0.18	part_siphonophorae	279	0.04	12 976	1.12
Candaciidae	70	0.01	2773	0.24	larvae_annelida	244	0.03	708	0.06
Amphipoda	68	0.01	853	0.07	**egg sac_egg**	152	0.02	394	0.03
Tomopteridae	58	0.01	618	0.05	zoea_decapoda	151	0.02	1405	0.12
Ostracoda	55	0.01	341	0.03	cnidaria_metazoa	148	0.02	4974	0.43
Doliolida	26	< 0.01	128	0.01	larvae_porcellanidae	127	0.02	2838	0.25
Echinodermata	24	< 0.01	253	0.02	nectophore_physonectae	106	0.02	696	0.06
Aetideidae	15	< 0.01	75	0.01	ctenophora_metazoa	94	0.01	126	0.01
Branchiostoma	15	< 0.01	210	0.02	**egg unkn temp_Engraulidae temp**	61	0.01	192	0.02
Thecosomata	15	< 0.01	59	0.01	part_ctenophora	30	< 0.01	319	0.03
Heterorhabdidae	8	< 0.01	205	0.02	tornaria larvae	21	< 0.01	83	0.01
Pontellidae	6	< 0.01	299	0.03	**egg_other**	17	< 0.01	2281	0.20
Cumacea	4	< 0.01	180	0.02	megalopa	6	< 0.01	460	0.04
Mysida	3	< 0.01	885	0.08	**scale**	2	< 0.01	53	< 0.01
Eucalanidae	2	< 0.01	839	0.07	siphonula	1	< 0.01	20	< 0.01
Insecta	2	< 0.01	3	< 0.01					
Foraminifera	1	< 0.01	384	0.03					
Haloptilus	1	< 0.01	5	< 0.01					
Isopoda	1	< 0.01	123	0.01					
Rhincalanidae	1	< 0.01	127	0.01					
Sapphirinidae	1	< 0.01	21	< 0.01					

Table 2. Taxa and operational morphological groups (OMGs) present in either the ZooCAM dataset or the ZooScan dataset but not both. Taxa and OMGs appearing exclusively in the ZooCAM dataset are listed on the left; those appearing exclusively in the ZooScan dataset are listed on the right. OMG names are spelled as they appear in the dataset. The numbers next to each taxon and OMG are the count and the percentage (%) for each category for each instrument in the whole dataset. Non-zooplanktonic taxa and OMGs are highlighted in bold, and genera and species are formatted in italics.

| Taxon/OMG	ZooCAM Counts	%	Taxon/OMG	ZooScan Counts	%
light_detritus	38 126	5.43	**badfocus_artefact**	34 507	2.99
Rhizaria	13 347	1.90	badfocus_Copepoda	11 656	1.01
Copepoda X	6727	0.96	Eumalacostraca	9815	0.85
fluffy_detritus	3589	0.51	part_Crustacea	7530	0.65
Evadne	1889	0.27	Fritillariidae	3635	0.32
Hydrozoa	1674	0.24	trunk_appendicularia	1210	0.10
Poecilostomatoida	1094	0.16	*Aglaura*	1113	0.10
Rhizaria X	857	0.12	*Pleuromamma*	695	0.06
Rhizosolenids	761	0.11	part_Cnidaria	692	0.06
dead_harpacticoida	528	0.08	zoea_galatheidae	660	0.06
gelatinous	348	0.05	pluteus_ophiuroidea	640	0.06
Trichodesmium	265	0.04	Salpida	470	0.04
aggregata	253	0.04	Harosa	374	0.03
feces	227	0.03	tail_chaetognatha	251	0.02
Halosphaera	193	0.03	*Euchirella*	239	0.02
Podon	162	0.02	protozoea_mysida	229	0.02
Diphyidae	144	0.02	*Solmundella bitentaculata*	178	0.02
larvae_gastropoda	116	0.02	Peltidiidae	133	0.01
chainlarge	114	0.02	*Liriope tetraphylla*	121	0.01
veliger	113	0.02	part_Annelida	121	0.01
egg 1 temp_Sardina temp	100	0.01	larvae_crustacea	114	0.01
egg 1 temp_Engraulidae temp	65	0.01	larvae_mysida	73	0.01
Isias	51	0.01	ephyra_scyphozoa	64	0.01
egg 2 3 temp_Sardina temp	49	0.01	actinula_hydrozoa	49	< 0.01
Calycophorae	30	< 0.01	part_thaliacea	44	< 0.01
egg 9 11 temp_Sardina temp	26	< 0.01	*Atlanta*	43	< 0.01
egg unkn temp_Sardina temp	23	< 0.01	like_laomediidae	36	< 0.01
Calocalanus tenuis	17	< 0.01	Nemertea	31	< 0.01
egg 4 6 temp_Sardina temp	15	< 0.01	protozoea_penaeidae	28	< 0.01
egg 9 11 temp_Engraulidae temp	14	< 0.01	Cavoliniidae	21	< 0.01
egg 7 8 temp_Engraulidae temp	13	< 0.01	Actiniaria	13	< 0.01
Enteropneusta_Hemichordata	12	< 0.01	pilidium_nemertea	12	< 0.01
Chaetoceros sp.	9	< 0.01	protozoea_sergestidae	12	< 0.01
head_crustacea	9	< 0.01	phyllosoma	8	< 0.01
Centropages hamatus	8	< 0.01	Creseidae	7	< 0.01
Thaliacea	7	< 0.01	Penaeoidea	7	< 0.01
egg 4 6 temp_Engraulidae temp	6	< 0.01	Paguridae	4	< 0.01
Sphaeronectidae	4	< 0.01	larvae_squillidae	4	< 0.01
Thalassionema	4	< 0.01	Cephalopoda	3	< 0.01
egg 2 3 temp_Engraulidae temp	3	< 0.01	*Cymbulia peroni*	3	< 0.01
Jaxea	2	< 0.01	Nannosquillidae	2	< 0.01
Pyrosoma	1	< 0.01	*Lubbockia*	1	< 0.01
larvae_ascidiacea	1	< 0.01	Monstrilloida	1	< 0.01
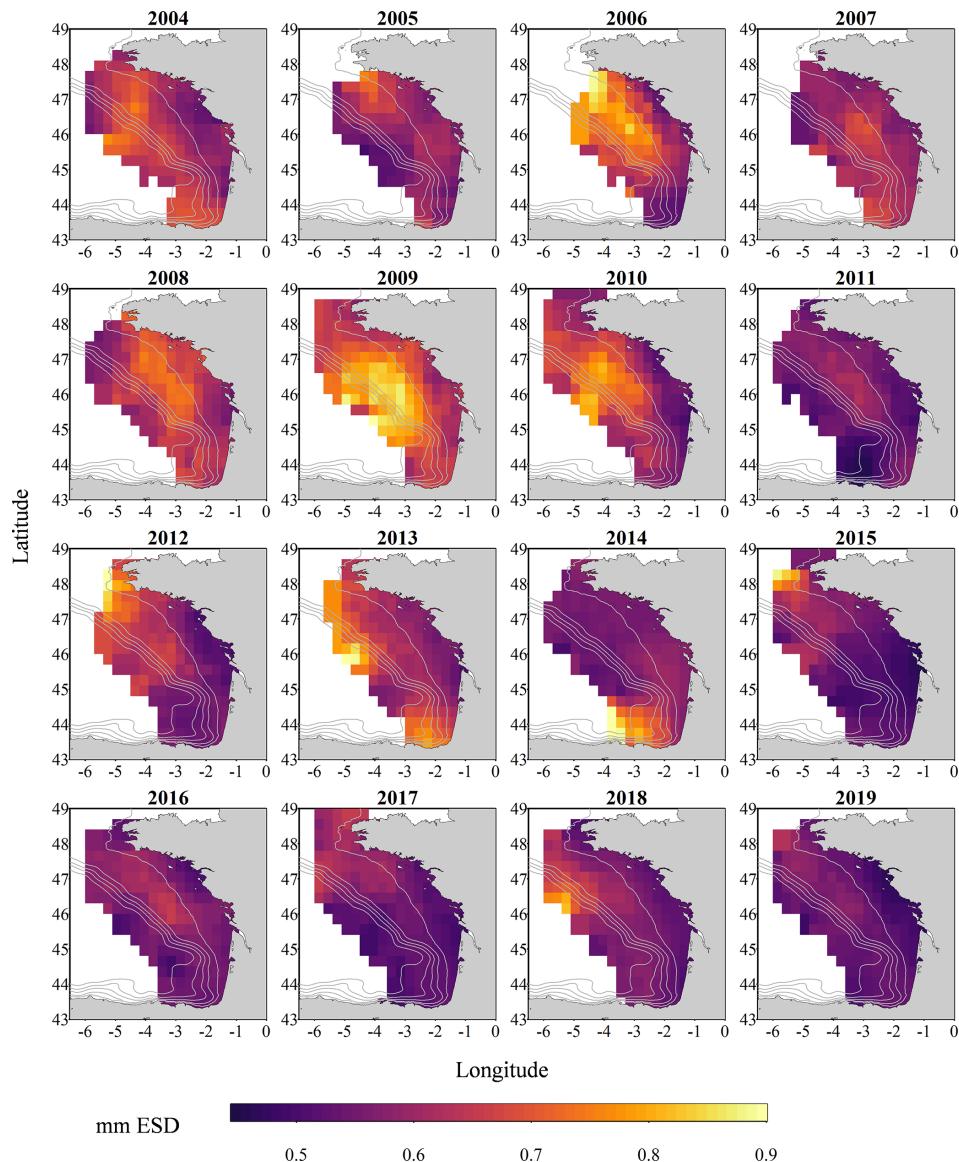
Figure 2. Gridded maps of total zooplankton abundances (expressed in individuals per cubic metre of sampled seawater) during the PELGAS cruises in the Bay of Biscay from 2004 to 2019. The abundances are well within the range of zooplankton abundances seen over other temperate continental shelves. They exhibit a marked coastal to offshore gradient, with abundances being higher at the coast. Abundances also show an overall increase over the years. The gridding procedure is presented in Petitgas et al. (2009, 2014). See also Doray et al. (2018c) and Grandremy et al. (2023a) for application examples.

features (i.e. pixel size). Those prefixes originate from the use of the Ecotaxa web application to sort and identify the images (Picheral et al., 2017), which promotes this specific formatting. The ZooCAM dataset is arranged as a matrix with 72 columns (variables) × 702 111 rows (individual imaged objects), and the ZooScan dataset is arranged as a matrix with 71 columns (variables) × 1 153 507 rows (individual imaged objects).

Among the 70+ variables, it is worth noting the following ones:

i. objid: this is a unique individual object numerical identifier that enables a single data line to be linked to a corresponding single image in the image dataset.

ii. taxon: this is the taxonomic or OMG identification of the imaged objects, written as they appear in Tables 1 and 2.

iii. lineage: this is the full taxonomic lineage of the taxon. Lineage may be used to aggregate taxa at higher taxonomic levels (respecting taxonomic lineages).

iv. classif_id: This is a unique numerical taxon identifier.

Figure 3. Gridded maps of total zooplankton mean size (expressed in mm equivalent spherical diameter) during the PELGAS cruise in the Bay of Biscay from 2004 to 2019. They exhibit a coastal to offshore gradient as well as a north–south gradient. Mean body sizes are smaller at the coast and usually smaller in the south. In general, mean body sizes show an overall decrease over the years. The gridding procedure is presented in Petitgas et al. (2009, 2014). See also Doray et al. (2018c) and Grandremy et al. (2023a) for application examples.

v. sample_sub_part/acq_sub_part: these are the subsampling ratios for ZooCAM and ZooScan respectively, which are needed to reconstruct the quantitative estimates of the sample abundances.

vi. sample_fishingvolume/sample_tot_vol: these are the total sampled seawater volumes for ZooCAM and ZooScan respectively, which are needed to normalize the sample concentrations by the seawater volume.

One can therefore calculate quantitative abundance estimates for a taxon in a sample as follows:

$$\text{ZooCAM: } Ab_{\text{taxon}} = \frac{n_{\text{taxon}} \times \text{sample\_sub\_part}}{\text{sample\_fishingvolume}} \quad (1)$$

$$\text{ZooScan: } Ab_{\text{taxon}} =$$

$$\frac{\left(n_{\text{taxon}_{\text{acq1}}} \times \text{acq\_sub\_part}_{\text{acq1}}\right) + \left(n_{\text{taxon}_{\text{acq2}}} \times \text{acq\_sub\_part}_{\text{acq2}}\right)}{\text{sample\_tot\_vol}}, \quad (2)$$

where Ab is the abundance in ind m^{-3} and n is the number of individuals for "taxon".

Table 3. Non-exhaustive list of OMG prefixes, their types (morphological, developmental stage, taxonomic, non-living, or imaging artefact) and their contents.

Prefix	Type	Content of category
bract	morphological	single siphonophore bracts
eudoxie	morphological	single siphonophore eudoxid zooids
gonophore	morphological	single siphonophore gonozooids
nectophore	morphological	single siphonophore swimming bells
trunk	morphological	single appendicularian trunks detached from their tails
tail	morphological	tail-shaped part of an appendicularian's or chaetognath's body
head	morphological	individual organisms' heads detached from the body
part	morphological	unidentified body part
egg sac	morphological	detached copepod egg sacs
like	morphological	look alike, but identification is not absolutely certain
multiple	morphological	two or more objects touching each other in the same vignette
other	morphological	non-identified living object
actinula	developmental stage	undefined hydrozoan actinula larval stage
calyptopsis	developmental stage	euphausiacean calyptopsis larval stage
egg	developmental stage	egg larval stage
ephyra	developmental stage	hydrozoan ephyra larval stage
larvae	developmental stage	undefined larval stage
nauplii	developmental stage	crustacean nauplius larval stage
pilidium	developmental stage	free-swimming larvae of a nemertean worm
protozoea	developmental stage	crustacean protozoea larval stage
pluteus	developmental stage	echinoderm pluteus larval stage
zoea	developmental stage	crustacean zoea larval stage
egg 1 temp	developmental stage	clupeid fish embryo developmental stage 1*
egg 2 3 temp	developmental stage	clupeid fish embryo developmental stages 2 and 3 (aggregated)*
egg 4 6 temp	developmental stage	clupeid fish embryo developmental stages 4 to 6 (aggregated)*
egg 7 8 temp	developmental stage	clupeid fish embryo developmental stages 7 and 8 (aggregated)*
egg 9 11 temp	developmental stage	clupeid fish embryo developmental stages 9 to 11 (aggregated)*
egg unknown	developmental stage	unidentified clupeid fish embryo developmental stage*
Bivalvia	taxonomic	small bivalve larvae of unidentified mollusc
dead	non-living	exuvia, carcass, or part of dead body of a copepod
fiber	non-living	fibre-like detritus
fluffy	non-living	very porous detritic particles
light	non-living	very transparent detritic particles
badfocus	imaging artefact	out-of-focus objects

* Clupeid fish embryo developmental stages according to Ahlstrom (1943) and Moser and Ahlstrom (1985).

3.2.2 Images
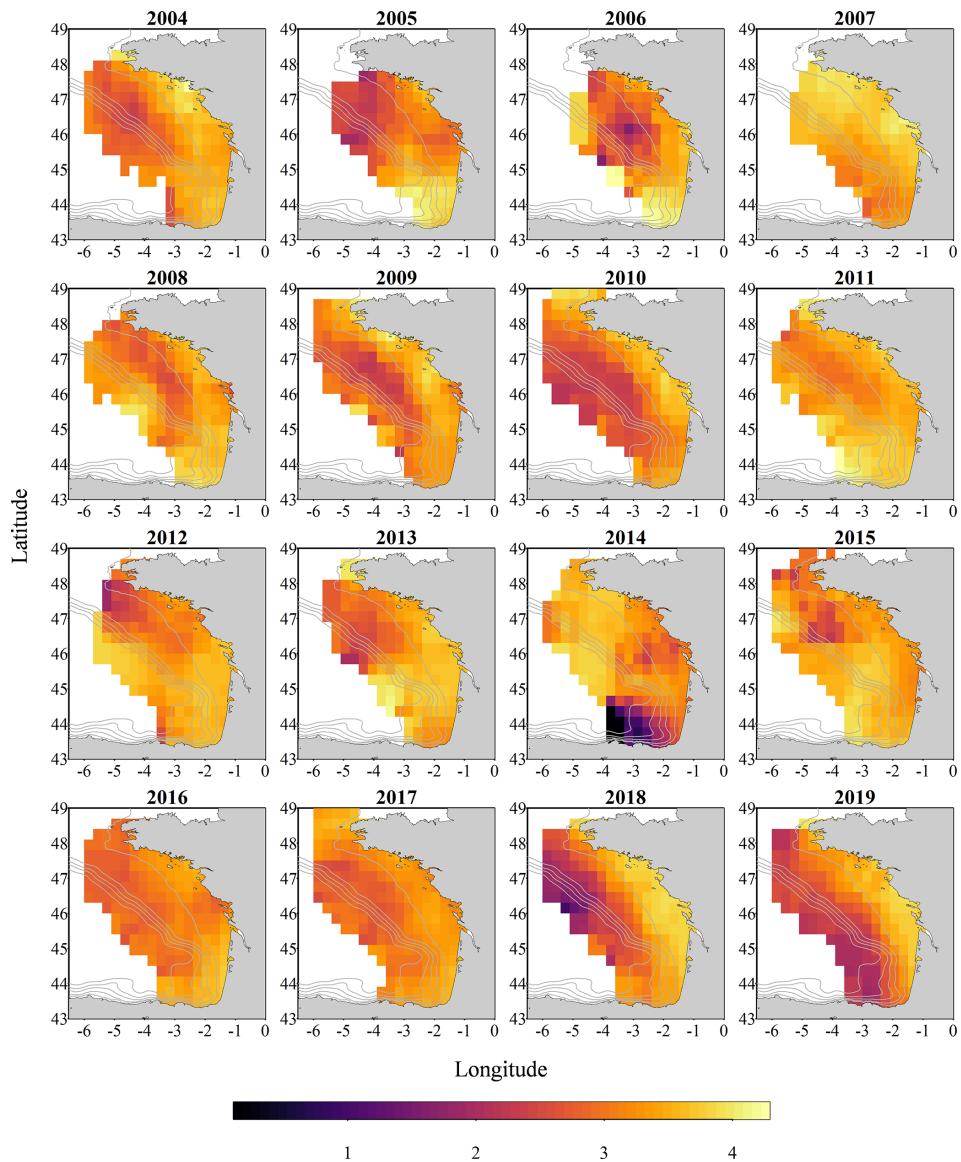
There are two sets of individual images sorted into folders by category (Tables 1 and 2) in each dataset. For the ZooCAM only, the images from years 2016 and 2017 contain printed region of interest (ROI) bounding-box limits and text at the bottom of each image and a non-homogenized background within and around the ROI bounding box; images from year 2018 contain a non-homogenized background within the ROI bounding box only; and images from 2019 have a completely homogeneous and thresholded background around the object. These differences arose from successive ZooCAM software updates that did not modify the calculation of an object's features. The ZooScan images all have a completely homogeneous and thresholded background around the object, with no bounding-box limits nor text printed in the im-

ages. All images for the two instruments' datasets have a 1 mm scale bar printed at the bottom left corner.

4 Data availability

The ZooScan dataset can be found as the PELGAS Bay of Biscay ZooScan zooplankton Dataset (2004–2016) at the SEANOE data portal at https://doi.org/10.17882/94052; Grandremy et al., 2023c). Individual object images can be freely viewed and explored by anyone using the Ecotaxa web application (https://ecotaxa.obs-vlfr.fr/, Picheral et al., 2017; no registration): search for the project name "PELGAS Bay of Biscay ZooScan zooplankton Dataset (2004–2016)" under the tab "explore images".

The ZooCAM dataset can be found as the PELGAS Bay of Biscay ZooCAM zooplankton Dataset (2016–2019) at

Figure 4. Gridded maps of total zooplankton Shannon index (calculated on spherical biovolumes) during the PELGAS cruise in the Bay of Biscay from 2004 to 2019. The Shannon index exhibits a coastal to offshore gradient as well as a north–south gradient. The Shannon index is larger at the coast and in the south except in 2014, where it is smaller offshore and in the south. The gridding procedure is presented in Petitgas et al. (2009, 2014). See also Doray et al. (2018c) and Grandremy et al. (2023a) for application examples.

the SEANOE data portal (https://doi.org/10.17882/94040; Grandremy et al., 2023d). Individual object images can be freely viewed and explored by anyone using the Ecotaxa web application (https://ecotaxa.obs-vlfr.fr/; no registration): search for the project name "PELGAS Bay of Biscay ZooCAM zooplankton Dataset (2016–2019)" under the tab "explore images".

Each dataset comes as a .zip archive that contains

– one tab-separated file containing all data and metadata associated with each imaged and identified object

– one comma-separated file containing the name, type, definition and unit of each field (column)

– one comma-separated file containing the taxonomic list of the dataset with counts and the nature of the content of the category

– a directory named *individual_images* containing images of individual objects that are named according to the object id *objid* and sorted into subdirectories according to taxonomic identification, year and sampling station.

5 Concluding remarks

Recent studies showed that small pelagic fish (SPF) communities have suffered from a drastic loss of condition in the Mediterranean Sea and in the Bay of Biscay (Van Beveren et al., 2014; Doray et al., 2018d; Saraux et al., 2019) over the last 20 years. This loss of condition is expressed in particular as a constant decrease in SPF size- and weight-at-age (Doray et al., 2018d; Véron et al., 2020) and is possibly explained by a change in SPF trophic resource composition, size and quality (Brosset et al., 2016; Queiros et al., 2019; Menu et al., 2023). Identifying and measuring zooplankton at appropriate temporal and spatial scales is not an easy task, but it can be addressed with imaging. These datasets were assembled in an effort to make possible the exploration of the relationship between the observed dynamics of SPF in the Bay of Biscay and the dynamics of their main food resource, metazoan zooplankton. This zooplankton imaging data series is a significant output of Nina Grandremy's PhD (2019–2023), and is currently being exploited (Grandremy et al., 2023a), and the intention is to continue this series and update it on a yearly basis in the framework of the PELGAS program to better understand the underlying processes presiding over the long-term SPF dynamics. Moreover, those two zooplankton datasets can be associated with the PELGAS survey datasets published in 2018 (which can also be found at the SEANOE data portal), which feature hydrological, primary producers, fish and megafauna data arranged as gridded data (Doray et al., 2018b). Together, all these datasets allow all of the pelagic ecosystem compartments to be studied simultaneously with a coherent spatial domain (the Bay of Biscay continental shelf), resolution and time series. Nevertheless, spatial gridding of the data is highly recommended (as presented in Figs. 2, 3 and 4), since the spatial coverage of the sampling protocols can vary between years (Fig. 1) within and between each pelagic ecosystem compartment. A procedure for such batch data spatial smoothing is presented in, for example, Petitgas et al. (2009, 2014). See also Doray et al. (2018c) and Grandremy et al. (2023a) for application examples. As several descriptors of the spring zooplankton community (abundances, sizes, biovolumes, biomass) can be derived from this 16-year-long spatially resolved time series at several taxonomic levels, these datasets are intended to be used in various ecological studies that include the zooplankton compartment, especially in modelling studies, where zooplankton is usually underrepresented (Mitra, 2010; Mitra et al., 2014). Finally, these datasets could also be applied as learning datasets when machine learning is used in plankton studies.

Appendix A

Table A1. ZooCAM dataset column headers – definitions of data and metadata fields.

Column name	Definition
object_id	Name of object and associated image
objid	Unique Ecotaxa internal object identifier
object_lat	Latitude of sampling
object_lon	Longitude of sampling
object_date	Date of sampling
object_time	Time of sampling
object_depth_min	Minimum sampling depth
object_depth_max	Maximum sampling depth
object_taxon	Taxonomic name
object_lineage	Full taxonomic lineage corresponding to the taxon
classif_id	Unique Ecotaxa internal taxon identifier
object_area	Object's surface
object_area_exc	Object surface excluding white pixels
object_%area	Proportion of the image corresponding to the object
object_area_based_diameter	Object's area-based diameter: $2\times (object\_area/pi)^{(1/2)}$
object_meangreyimage	Mean image grey level
object_meangreyobjet	Mean object grey level
object_modegreyobjet	Modal object grey level
object_sigmagrey	Object grey-level standard deviation
object_mingrey	Minimum object grey level
object_maxgrey	Maximum object grey level
object_sumgrey	Object grey-level integrated density: $object\_mean \times object\_area$
object_breadth	Breadth of the object along the best-fitting ellipsoid's minor axis
object_length	Breadth of the object along the best-fitting ellipsoid's major axis
object_elongation	Elongation index: object_length/object_breadth
object_perim	Object's perimeter
object_minferetdiam	Minimum object's Feret diameter
object_maxferetdiam	Maximum object's Feret diameter
object_meanferetdiam	Average object's Feret diameter
object_feretelongation	Elongation index: object_maxferetdiam/object_minferetdiam
object_compactness	Isoperimetric quotient: the ratio of the object's area to the area of a circle with the same perimeter
object_intercept0	The number of times that a transition from background to foreground occurs at an angle of 0° for the entire object
object_intercept45	The number of times that a transition from background to foreground occurs at an angle of 45° for the entire object
object_intercept90	The number of times that a transition from background to foreground occurs at an angle of 90° for the entire object
object_intercept135	The number of times that a transition from background to foreground occurs at an angle of 135° for the entire object
object_convexhullarea	Area of the convex hull of the object
object_convexhullfillratio	ratio object_area/convexhullarea
object_convexperimeter	Perimeter of the convex hull of the object
object_n_number_of_runs	Number of horizontal strings of consecutive foreground pixels in the object
object_n_chained_pixels	Number of chained pixels in the object
object_n_convex_hull_points	Number of summits of the object's convex hull polygon
object_n_number_of_holes	Number of holes (closed white pixel area) in the object
object_transparence	Ratio object_sumgrey/object_area
object_roughness	Measure of small-scale variations of amplitude in the object's grey levels
object_rectangularity	Ratio of the object's area to the area of its best bounding rectangle
object_skewness	Skewness of the object's grey-level distribution
object_kurtosis	Kurtosis of the object's grey-level distribution
object_fractal_box	Fractal dimension of the object's perimeter
object_hist25	Grey level value at quantile 0.25 of the normalized cumulative histogram of the object's grey level values
object_hist50	Grey level value at quantile 0.5 of the normalized cumulative histogram of the object's grey level values
object_hist75	Grey level value at quantile 0.75 of the normalized cumulative histogram of the object's grey level values
object_valhist25	Sum of grey levels at the 0.25 quantile of the normalized cumulative histogram of the object's grey level values
object_valhist50	Sum of grey levels at the 0.5 quantile of the normalized cumulative histogram of the object's grey level values
object_valhist75	Sum of grey levels at the 0.75 quantile of the normalized cumulative histogram of the object's grey level values
object_nobj25	Number of objects after thresholding at grey level object_valhist25
object_nobj50	Number of objects after thresholding at grey level object_valhist50
object_nobj75	Number of objects after thresholding at grey level object_valhist75
object_symetrieh	Index of horizontal symmetry
object_symetriev	Index of vertical symmetry
object_thick_r	Maximum object thickness/mean object thickness
object_cdist	Distance between the mass and the grey-level object's centroids
object_bord	Tag for an object touching the frame edge
sample_id	Name of the sample from which the object originates
sample_ship	Name of the ship used to collect the samples
sample_campaign	Name of the cruise during which the samples were collected
sample_station	Name of the station where the samples were collected
sample_depth	Bottom depth at station
sample_device	Net used to collect the sample
sample_fishingvolume	Seawater volume sampled
sample_sub_part	Subsampling elevation factor
process_id	Name of software/software version used to analyse the digitized sample images
process_resolution_camera_micron_par_pixel	Pixel size

Table A2. ZooScan dataset column headers – definitions of data and metadata fields.

Column name	Definition
object_id	Name of object and associated image
objid	Unique Ecotaxa internal object identifier
object_lat	Latitude of sampling
object_lon	Longitude of sampling
object_date	Date of sampling
object_time	Time of sampling
object_depth_min	Minimum sampling depth
object_depth_max	Maximum sampling depth
object_taxon	Taxonomic name
object_lineage	Full taxonomic lineage corresponding to the taxon
classif_id	Unique Ecotaxa internal taxon identifier
object_area	Object's surface
object_mean	Mean object grey level
object_stddev	Object grey-level standard deviation
object_mode	Modal object grey level
object_min	Minimum object grey level
object_max	Maximum object grey level
object_perim.	Object's perimeter
object_major	Length of major axis of best-fitting ellipse
object_minor	Length of minor axis of best-fitting ellipse
object_circ.	Circularity: $4 \times$ pi(object_area/object_perim.^{2})
object_feret	Maximum Feret diameter
object_intden	Object grey-level integrated density: object_mean $\times$ object_area
object_median	Median object grey level
object_skew	Skewness of the object's grey-level distribution
object_kurt	Kurtosis of the object's grey-level distribution
object_ %area	Proportion of the image corresponding to the object
object_area_exc	Object surface excluding white pixels
object_fractal	Fractal dimension of the object's perimeter
object_skelarea	Surface of the one-pixel-wide skeleton of the object
object_slope	Slope of the cumulated histogram of the object's grey levels
object_histcum1	Grey level value at the 0.25 quantile of the histogram of the object's grey levels
object_histcum2	Grey level at the 0.5 quantile of the histogram of the object's grey levels
object_histcum3	Grey level at the 0.75 quantile of the histogram of the object's grey levels
object_nb1	Number of objects after thresholding at object_histcum1 grey level
object_nb2	Number of objects after thresholding at the grey level object_histcum2
object_symetrieh	Index of horizontal symmetry
object_symetriev	Index of vertical symmetry
object_symetriehc	Index of horizontal symmetry after thresholding at object_ histcum1 grey level
object_symetrievc	Index of vertical symmetry after thresholding at object_ histcum1 grey level
object_convperim	Perimeter of the convex hull of the object
object_convarea	Area of the convex hull of the object
object_fcons	Object's contrast
object_thickr	Maximum object thickness/mean object thickness
object_esd	Object's equivalent spherical diameter: $2 \times$ (object_area/pi)$^{(1/2)}$
object_elongation	Elongation index: major/minor
object_range	Range of greys: max$-$min
object_meanpos	Relative position of the mean grey: (max$-$min)/range
object_centroids	Distance between the mass and the grey-level object's centroids
object_cv	Coefficient of variation of greys: $100 \times$ (stddev/mean)
object_sr	Index of variation of greys: $100 \times$ (stddev/range)
object_perimareaexc	Index of the relative complexity of the perimeter: object_ perim/object_area_exc
object_feretareaexc	Another elongation index: object_feret/object_area_exc
object_perimferet	Index of the relative complexity of the perimeter: object_perim/object_feret
object_perimmajor	Index of the relative complexity of the perimeter: object_perim/object_major
object_circex	Circularity of the object excluding white pixels: $4 \times$ pi(object_area_exc/object_perim.^{2})
object_cdexc	Distance between the mass and the grey-level object's centroids calculated with object_area_exc
sample_id	Name of the sample from which the object originates
sample_ship	Name of the ship used to collect the samples
sample_program	Name of the cruise from which the samples were collected
sample_stationid	Name of the station where the samples were collected
sample_bottomdepth	Bottom depth at station
sample_net_type	Net used to collect the sample
sample_tot_vol	Seawater volume sampled
sample_comment	Comments associated with sampling/sample treatment
process_id	Name of software/software version used to analyse digitized sample images
process_particle_pixel_size_mm	Pixel size
acq_id	Name of subsample (if any)
acq_min_mesh	Minimum sieve size of subsample
acq_max_mesh	Maximum sieve size of subsample
acq_sub_part	Subsampling elevation factor

References

Ahlstrom, E. H.: Studies on the Pacific Pilchard Or Sardine (Sardinops Caerulea): Influence of Temperature on the Rate of Development of Pilchard Eggs in Nature, U.S. Fish and Wildlife Service, 206 pp., 1943.

Banse, K.: Zooplankton: Pivotal role in the control of ocean production: I. Biomass and production, ICES J. Mar. Sci., 52, 265–277, https://doi.org/10.1016/1054-3139(95)80043-3, 1995.

Batten, S. D., Abu-Alhaija, R., Chiba, S., Edwards, M., Graham, G., Jyothibabu, R., Kitchener, J. A., Koubbi, P., McQuatters-Gollop, A., Muxagata, E., Ostle, C., Richardson, A. J., Robinson, K. V., Takahashi, K. T., Verheye, H. M., and Wilson, W.: A Global Plankton Diversity Monitoring Program, Front. Mar. Sci., 6, 321, https://doi.org/10.3389/fmars.2019.00321, 2019.

Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S., and Reid, P. C.: Plankton effect on cod recruitment in the North Sea, Nature, 426, 661–664, https://doi.org/10.1038/nature02164, 2003.

Benedetti, F., Jalabert, L., Sourisseau, M., Becker, B., Cailliau, C., Desnos, C., Elineau, A., Irisson, J.-O., Lombard, F., Picheral, M., Stemmann, L., and Pouline, P.: The Seasonal and Inter-Annual Fluctuations of Plankton Abundance and Community Structure in a North Atlantic Marine Protected Area, Front. Mar. Sci., 6, 214, https://doi.org/10.3389/fmars.2019.00214, 2019.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., and Not, F.: In situ imaging reveals the biomass of giant protists in the global ocean, Nature, 532, 504–507, https://doi.org/10.1038/nature17652, 2016.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brosset, P., Bourg, B. L., Costalago, D., Bănaru, D., Beveren, E. V., Bourdeix, J.-H., Fromentin, J.-M., Ménard, F., and Saraux, C.: Linking small pelagic dietary shifts with ecosystem changes in the Gulf of Lions, Mar. Ecol. Prog. Ser., 554, 157–171, https://doi.org/10.3354/meps11796, 2016.

Chiba, S., Batten, S., Martin, C. S., Ivory, S., Miloslavich, P., and Weatherdon, L. V.: Zooplankton monitoring to contribute towards addressing global biodiversity conservation challenges, J. Plankton Res., 40, 509–518, https://doi.org/10.1093/plankt/fby030, 2018.

Colas, F., Tardivel, M., Perchoc, J., Lunven, M., Forest, B., Guyader, G., Danielou, M. M., Le Mestre, S., Bourriau, P., Antajan, E., Sourisseau, M., Huret, M., Petitgas, P., and Romagnan, J. B.: The ZooCAM, a new in-flow imaging system for fast onboard counting, sizing and classification of fish eggs and metazooplankton, Prog. Oceanogr., 166, 54–65, https://doi.org/10.1016/j.pocean.2017.10.014, 2018.

Culverhouse, P. F.: Human and machine factors in algae monitoring performance, Ecol. Inform., 2, 361–366, https://doi.org/10.1016/j.ecoinf.2007.07.001, 2007.

Cury, P., Bakun, A., Crawford, R. J. M., Jarre, A., Quiñones, R. A., Shannon, L. J., and Verheye, H. M.: Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems, ICES J. Mar. Sci., 57, 603–618, https://doi.org/10.1006/jmsc.2000.0712, 2000.

Doray, M., Huret, M., Authier, M., Duhamel, E., Romagnan, J.-B., Dupuy, C., Spitz, J., Sanchez, F., Berger, L., Dorémus, G., Bourriau, P., Grellier, P., Pennors, L., Masse, J., and Petitgas, P.: Gridded maps of pelagic ecosystem parameters collected in the Bay of Biscay during the PELGAS integrated survey, SEANOE [data set], https://doi.org/10.17882/53389, 2018a.

Doray, M., Petitgas, P., Huret, M., Duhamel, E., Romagnan, J. B., Authier, M., Dupuy, C., and Spitz, J.: Monitoring small pelagic fish in the Bay of Biscay ecosystem, using indicators from an integrated survey, Prog. Oceanogr., 166, 168–188, https://doi.org/10.1016/j.pocean.2017.12.004, 2018b.

Doray, M., Hervy, C., Huret, M., and Petitgas, P.: Spring habitats of small pelagic fish communities in the Bay of Biscay, Prog. Oceanogr., 166, 88–108, https://doi.org/10.1016/j.pocean.2017.11.003, 2018c.

Doray, M., Petitgas, P., Romagnan, J. B., Huret, M., Duhamel, E., Dupuy, C., Spitz, J., Authier, M., Sanchez, F., Berger, L., Dorémus, G., Bourriau, P., Grellier, P., and Massé, J.: The PELGAS survey: Ship-based integrated monitoring of the Bay of Biscay pelagic ecosystem, Prog. Oceanogr., 166, 15–29, https://doi.org/10.1016/j.pocean.2017.09.015, 2018d.

Doray, M., Boyra, G., and van der Kooij, J.: ICES Survey Protocols – Manual for acoustic surveys coordinated under ICES Working Group on Acoustic and Egg Surveys for Small Pelagic Fish (WGACEGG), https://doi.org/10.17895/ICES.PUB.7462, 2021.

Elineau, A., Desnos, C., Jalabert, L., Olivier, M., Romagnan, J.-B., Costa Brandao, M., Lombard, F., Llopis, N., Courboulès, J., Caray-Counil, L., Serranito, B., Irisson, J.-O., Picheral, M., Gorsky, G., and Stemmann, L.: ZooScanNet: plankton images captured with the ZooScan, SEANOE [data set], https://doi.org/10.17882/55741, 2018.

Feuilloley, G., Fromentin, J.-M., Saraux, C., Irisson, J.-O., Jalabert, L., and Stemmann, L.: Temporal fluctuations in zooplankton size, abundance, and taxonomic composition since 1995 in the North Western Mediterranean Sea, ICES J. Mar. Sci., 79, 882–900, https://doi.org/10.1093/icesjms/fsab190, 2022.

Gorsky, G., Ohman, M. D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.-B., Cawood, A., Pesant, S., García-Comas, C., and Prejger, F.: Digital zooplankton image analysis using the ZooScan integrated system, J. Plankton Res., 32, 285–303, https://doi.org/10.1093/plankt/fbp124, 2010.

Graham, B.: Spatially-sparse convolutional neural networks, arXiv [preprint], https://doi.org/10.48550/arXiv.1409.6070, 2014.

Grandremy, N., Romagnan, J.-B., Dupuy, C., Doray, M., Huret, M., and Petitgas, P.: Hydrology and small pelagic fish drive the spatio–temporal dynamics of springtime zooplankton assemblages over the Bay of Biscay continental shelf, Prog. Oceanogr., 210, 102949, https://doi.org/10.1016/j.pocean.2022.102949, 2023a.

Grandremy, N., Dupuy, C., Petitgas, P., Mestre, S. L., Bourriau, P., Nowaczyk, A., Forest, B., and Romagnan, J.-B.: The ZooScan and the ZooCAM zooplankton imaging systems are intercomparable: A benchmark on the Bay of Biscay zooplankton, Limnol. Oceanogr.-Meth., 21, 718–733, https://doi.org/10.1002/lom3.10577, 2023b.

Grandremy, N., Bourriau, P., Daché, E., Danielou, M. M., Doray, M., Dupuy, C., Huret, M., Jalabert, L., Le Mestre, S., Nowaczyk, A., Petitgas, P., Pineau, P., Raphalen, E., and Romagnan, J.-B.: PELGAS Bay of Biscay ZooScan zooplankton Dataset (2004–2016), SEANOE [data set], https://doi.org/10.17882/94052, 2023c.

Grandremy, N., Bourriau, P., Danielou, M. M., Doray, M., Dupuy, C., Forest, B., Huret, M., Le Mestre, S., Nowaczyk, A., Petitgas, P., Pineau, P., Rouxel, J., Tardivel, M., and Romagnan, J.-B.: PELGAS Bay of Biscay ZooCAM zooplankton Dataset (2016–2019), SEANOE [data set], https://doi.org/10.17882/94040, 2023d.

ICES: Bay of Biscay and Iberian Coast ecoregion – Fisheries overview, ICES Advice: Fisheries Overviews, https://doi.org/10.17895/ices.advice.9100, 2021.

Irisson, J.-O., Ayata, S.-D., Lindsay, D. J., Karp-Boss, L., and Stemmann, L.: Machine Learning for the Study of Plankton and Marine Snow from Images, Annu. Rev. Mar. Sci., 14, 277–301, https://doi.org/10.1146/annurev-marine-041921-013023, 2022.

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., Sosik, H. M., Schulz, J., Romagnan, J.-B., Picheral, M., Pearlman, J., Ohman, M. D., Niehoff, B., Möller, K. O., Miloslavich, P., Lara-Lpez, A., Kudela, R., Lopes, R. M., Kiko, R., Karp-Boss, L., Jaffe, J. S., Iversen, M. H., Irisson, J.-O., Fennel, K., Hauss, H., Guidi, L., Gorsky, G., Giering, S. L. C., Gaube, P., Gallager, S., Dubelaar, G., Cowen, R. K., Carlotti, F., Briseño-Avena, C., Berline, L., Benoit-Bird, K., Bax, N., Batten, S., Ayata, S. D., Artigas, L. F., and Appeltans, W.: Globally Consistent Quantitative Observations of Planktonic Ecosystems, Front. Mar. Sci., 6, 196, https://doi.org/10.3389/fmars.2019.00196, 2019.

Menu, C., Pecquerie, L., Bacher, C., Doray, M., Hattab, T., van der Kooij, J., and Huret, M.: Testing the bottom-up hypothesis for the decline in size of anchovy and sardine across European waters through a bioenergetic modeling approach, Prog. Oceanogr., 210, 102943, https://doi.org/10.1016/j.pocean.2022.102943, 2023.

Mitra, A. and Davis, C.: Defining the "to" in end-to-end models, Prog. Oceanogr., 84, 39–42, https://doi.org/10.1016/j.pocean.2009.09.004, 2010.

Mitra, A., Castellani, C., Gentleman, W. C., Jónasdóttir, S. H., Flynn, K. J., Bode, A., Halsband, C., Kuhn, P., Licandro, P., Agersted, M. D., Calbet, A., Lindeque, P. K., Koppelmann, R., Møller, E. F., Gislason, A., Nielsen, T. G., and St. John, M.: Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link, Prog. Oceanogr., 129, 176–199, https://doi.org/10.1016/j.pocean.2014.04.025, 2014.

Moser, H. G. and Ahlstrom, E. H.: Staging anchovy eggs, National Marine Fisheries Service, Southwest Fisheries Center, NOM,

PO. Box 271, La Jolla, CA 92038, NOAA Tech. Rep. NMFS, 36, 37-41, 1985.

Ohman, M. D. and Romagnan, J.-B.: Nonlinear effects of body size and optical attenuation on Diel Vertical Migration by zooplankton, Limnol. Oceanogr., 61, 765–770, https://doi.org/10.1002/lno.10251, 2016.

Orenstein, E. C., Ayata, S., Maps, F., Becker, É. C., Benedetti, F., Biard, T., de Garidel-Thoron, T., Ellen, J. S., Ferrario, F., Giering, S. L. C., Guy-Haim, T., Hoebeke, L., Iversen, M. H., Kiørboe, T., Lalonde, J., Lana, A., Laviale, M., Lombard, F., Lorimer, T., Martini, S., Meyer, A., Möller, K. O., Niehoff, B., Ohman, M. D., Pradalier, C., Romagnan, J., Schröder, S., Sonnet, V., Sosik, H. M., Stemmann, L. S., Stock, M., Terbiyik-Kurt, T., Valcárcel-Pérez, N., Vilgrain, L., Wacquet, G., Waite, A. M., and Irisson, J.: Machine learning techniques to characterize functional traits of plankton from image data, Limnol. Oceanogr., 67, 1647–1669, https://doi.org/10.1002/lno.12101, 2022.

Panaïotis, T., Caray–Counil, L., Woodward, B., Schmid, M. S., Daprano, D., Tsai, S. T., Sullivan, C. M., Cowen, R. K., and Irisson, J.-O.: Content-Aware Segmentation of Objects Spanning a Large Size Range: Application to Plankton Images, Front. Mar. Sci., 9, 870005, https://doi.org/10.3389/fmars.2022.870005, 2022.

Petitgas, P., Goarant, A., Massé, J., and Bourriau, P.: Combining acoustic and CUFES data for the quality control of fish-stock survey estimates, ICES J. Mar. Sci., 66, 1384–1390, https://doi.org/10.1093/icesjms/fsp007, 2009.

Petitgas, P., Doray, M., Huret, M., Massé, J., and Woillez, M.: Modelling the variability in fish spatial distributions over time with empirical orthogonal functions: anchovy in the Bay of Biscay, ICES J. Mar. Sci., 71, 2379–2389, https://doi.org/10.1093/icesjms/fsu111, 2014.

Picheral, M., Colin, S., and Irisson, J.O.: EcoTaxa, a tool for the taxonomic classification of images, https://ecotaxa.obs-vlfr.fr/ (last access: 1 June 2023), 2017.

Queiros, Q., Fromentin, J.-M., Gasset, E., Dutto, G., Huiban, C., Metral, L., Leclerc, L., Schull, Q., McKenzie, D. J., and Saraux, C.: Food in the Sea: Size Also Matters for Pelagic Fish, Front. Mar. Sci., 6, 385, https://doi.org/10.3389/fmars.2019.00385, 2019.

Romagnan, J. B., Aldamman, L., Gasparini, S., Nival, P., Aubert, A., Jamet, J. L., and Stemmann, L.: High frequency meso-zooplankton monitoring: Can imaging systems and automated sample analysis help us describe and interpret changes in zooplankton community composition and size structure – An example from a coastal site, J. Marine Syst., 162, 18–28, https://doi.org/10.1016/j.jmarsys.2016.03.013, 2016.

Saraux, C., Van Beveren, E., Brosset, P., Queiros, Q., Bourdeix, J.-H., Dutto, G., Gasset, E., Jac, C., Bonhommeau, S., and Fromentin, J.-M.: Small pelagic fish dynamics: A review of mechanisms in the Gulf of Lions, Deep-Sea Res. Pt. II, 159, 52–61, https://doi.org/10.1016/j.dsr2.2018.02.010, 2019.

Sieburth, J. McN., Smetacek, V., and Lenz, J.: Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1, Limnol. Oceanogr., 23, 1256–1263, https://doi.org/10.4319/lo.1978.23.6.1256, 1978.

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., Burd, A., Carlson, C. A., D'Asaro, E. A., Doney, S. C., Perry, M. J., Stanley, R. H. R., and Steinberg, D. K.: Prediction of the Export and Fate of Global Ocean Net Primary Production: The EXPORTS Science Plan, Front. Mar. Sci., 3, 22, https://doi.org/10.3389/fmars.2016.00022, 2016.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., and Michaels, A. F.: Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea, Deep-Sea Res. Pt. I, 47, 137–158, https://doi.org/10.1016/S0967-0637(99)00052-7, 2000.

Turner, J. T.: Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump, Prog. Oceanogr., 130, 205–248, https://doi.org/10.1016/j.pocean.2014.08.005, 2015.

Uitz, J., Claustre, H., Gentili, B., and Stramski, D.: Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations, Global Biogeochem. Cy., 24, GB3016, https://doi.org/10.1029/2009GB003680, 2010.

Van Beveren, E., Bonhommeau, S., Fromentin, J.-M., Bigot, J.-L., Bourdeix, J.-H., Brosset, P., Roos, D., and Saraux, C.: Rapid changes in growth, condition, size and age of small pelagic fish in the Mediterranean, Mar. Biol., 161, 1809–1822, https://doi.org/10.1007/s00227-014-2463-1, 2014.

van der Lingen, C., Hutchings, L., and Field, J.: Comparative trophodynamics of anchovy Engraulis encrasicolus and sardine Sardinops sagax in the southern Benguela: are species alternations between small pelagic fish trophodynamically mediated?, Afr. J. Mar. Sci., 28, 465–477, https://doi.org/10.2989/18142320609504199, 2006.

Vandromme, P., Stemmann, L., Garcìa-Comas, C., Berline, L., Sun, X., and Gorsky, G.: Assessing biases in computing size spectra of automatically classified zooplankton from imaging systems: A case study with the ZooScan integrated system, Meth. Oceanogr., 1–2, 3–21, https://doi.org/10.1016/j.mio.2012.06.001, 2012.

Vandromme, P., Nogueira, E., Huret, M., Lopez-Urrutia, Á., González-Nuevo González, G., Sourisseau, M., and Petitgas, P.: Springtime zooplankton size structure over the continental shelf of the Bay of Biscay, Ocean Sci., 10, 821–835, https://doi.org/10.5194/os-10-821-2014, 2014.

Véron, M., Duhamel, E., Bertignac, M., Pawlowski, L., and Huret, M.: Major changes in sardine growth and body condition in the Bay of Biscay between 2003 and 2016: Temporal trends and drivers, Prog. Oceanogr., 182, 102274, https://doi.org/10.1016/j.pocean.2020.102274, 2020.