PREPRINT OPEN ACCESS

A data-driven assessment of the global community structure of pelagic zooplankton biomass

Corentin Clerc[1,2] | Alexandre Schickele[2] | Fabio Benedetti[2,3] | Urs Hofmann Elizondo[2] | Servane Lunven[2] | Ruby Bader[2] | Sébastien Vivier[2] | Anthony J. Richardson[4,5] | Meike Vogt[2]

[1]Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques, LOCEAN/IPSL, F-75005 Paris, France | [2]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092, Zürich, Switzerland. | [3]Plant Ecology Group, Institute of Plant Sciences, University of Bern, 3013 Bern, Switzerland | [4]School of the Environment, The University of Queensland, St Lucia, 4072, Queensland, Australia [5]CSIRO Environment, Biosciences Precinct, St Lucia, 4067, Queensland, Australia |

Correspondence: Corentin Clerc (corentin.clerc@locean.ipsl.fr)

Received: Revised: Accepted:

Funding: Horizon 2020 Research and Innovation Programme under grant agreement no. 101094227 (Bluecloud2026), no. 101059915 (BIOcean5D) and no. 862923 (AtlantECO). Cnes-Tosca project SO-BIO-SAT

Keywords: Zooplankton, Biogeography, Biomass, Diversity, Global Ocean

ABSTRACT

Aim: Zooplankton often dominate pelagic metazoan biomass, making them a central component of marine ecosystems and biogeochemical cycles. In this study, we aimed to: (1) provide biomass distributions of zooplankton functional types to support marine ecosystem monitoring and biogeochemical model evaluation; and (2) investigate regional and seasonal patterns of biomass across marine zooplankton functional types and their relationships with species diversity.

Location: Global.

Time Period: 1926–2016.

Major Taxa Studied: Marine zooplankton (ten plankton functional types, PFTs).

Methods: We compiled zooplankton biomass measurements from net samples and developed habitat suitability models to generate monthly global epipelagic (0-200 m) and mesopelagic (200-500 m) biomass climatologies for ten PFTs. We then linked epipelagic biomass climatologies to recent empirical species richness estimates to assess biomass-richness relationships at both community and PFT scales.

Results: Global epipelagic zooplankton biomass was estimated at 2.9 Pg C and exhibited strong seasonality, with summer maxima at high latitudes and persistently elevated biomass in productive regions. Mesopelagic biomass was lower (0.9 Pg C) and showed winter maxima at high latitudes. Crustaceans dominated global zooplankton biomass (>50% of total biomass), but gelatinous and calcifying zooplankton contributions were substantial (>10% each). At the community scale, the biomass-richness relationship was non-monotonic: both species-poor and species-rich environments had high biomass. At the PFT scale, crustacean biomass decreased with richness, whereas gelatinous zooplankton biomass increased.

Main conclusion: Our analysis provides the first net-based monthly global biomass climatologies for ten zooplankton functional types across the epipelagic and mesopelagic zones. It shows that a few dominant crustacean groups capture much of the large-scale biomass structure but that resolving gelatinous taxa and calcifiers remains necessary because they contribute substantially to biomass and biogeochemical function. Together with the contrasting richness–biomass relationships among PFTs, these results indicate that biomass-based monitoring may be sufficient to track some crustacean-dominated patterns, whereas groups such as gelatinous zooplankton require metrics that also resolve diversity within PFTs.

1 | Introduction

Marine zooplankton modulate marine ecosystems by regulating energy transfer from phytoplankton to higher trophic levels and by contributing to carbon sequestration through influencing the efficiency and magnitude of the biological carbon pump [1]. These heterotrophic plankton also represent a major fraction of living biomass in marine pelagic environments [2]. Monitoring global zooplankton biomass distributions and constraining them in models are therefore essential for understanding marine ecosystem functioning, assessing ocean health, and quantifying the impacts of anthropogenic change on ocean biogeochemistry [3, 4]. This importance is reflected in the recognition of zooplankton biomass and diversity as Essential Ocean Variables (EOVs) [5], emphasizing their role in sustained ecosystem monitoring and climate-relevant observing systems [6]. Yet global zooplankton biomass remains difficult to constrain. Observationally, zooplankton biomass is highly patchy in space and time, sampling remains uneven across regions, depths, seasons, and taxa, and there is no global standard for measuring zooplankton biomass [7, 8, 9]. In marine biogeochemical models, zooplankton processes remain a major source of uncertainty, with grazing identified as the largest source of uncertainty in marine carbon-cycling in CMIP6 models [10]. Information on global zooplankton biomass is therefore required as both a fundamental target for ocean observation and a key benchmark for evaluating marine ecosystem and biogeochemical models.

A major step toward constraining global zooplankton biomass was the MAREDAT synthesis [7], which provided annual global epipelagic estimates for mesozooplankton, macrozooplankton, and calcifying zooplankton. MAREDAT also highlighted the limitations that still motivate current work: strong spatial heterogeneity driven by the patchy distribution of plankton and the lack of standardisation in collection procedures, uncertainty ranges often spanning orders of magnitude relative to mean biomass, limited information for different functional groups, and the absence of seasonal and depth-resolved data. Subsequent studies have progressively filled parts of this gap by resolving seasonal mesozooplankton and calcifier biomass [11, 12], depth-resolved mesozooplankton distributions [13, 14], and upper-ocean functional composition [15, 16]. Together, these products suggest that crustaceans dominate global zooplankton biomass, while gelatinous and calcifying groups are important contributors to ecosystem and biogeochemical functioning. However, no existing

product simultaneously resolves seasonality, depth structure, and multiple plankton functional types (PFTs) within a single global biomass framework.

This lack of detailed information on PFTs is important because zooplankton biomass is controlled by community composition. Grouping species into PFTs, an approach already used in MAREDAT and formalized in biogeochemical marine ecosystem models [17, 7], provides a simple way to represent plankton diversity. This approach leverages differences in feeding mode, body size, and vertical behavior that can substantially alter zooplankton biomass distributions and their ecological and biogeochemical impacts. For instance, nonselective filter-feeding groups such as salps and appendicularians may maintain relatively high biomass in oligotrophic environments [18, 19], whereas migration and life-history strategies such as diel vertical migration or diapause can control seasonal and depth-dependent biomass patterns [20, 12]. However, the coarse mesozooplankton and macrozooplankton categories used in early syntheses are no longer sufficient to match the growing interest and increasing need in explicitly representing distinct zooplankton groups in models, including salps [19], appendicularians [18], and jellyfish [21]. More detailed PFT-resolved zooplankton biomass products are therefore needed both to evaluate emerging model structure and to support monitoring frameworks that aim to identify which zooplankton groups are most informative for ecosystem monitoring.

Even the inclusion of many PFTs in models may not fully resolve zooplankton community structure. For example, dominant groups such as copepods contain multiple sub-PFTs, each with distinct ecological and biogeochemical roles [22]. The degree of diversity in the zooplankton community is a practical issue for both modeling and monitoring zooplankton biomass. In some communities, most biomass may be concentrated in a few dominant taxa, consistent with the mass-ratio hypothesis [23, 24], whereas in others it may reflect the contribution of many coexisting taxa through complementarity [25]. Biomass-richness relationships are informative in this context because they provide an empirical way to distinguish between these contrasting modes of biomass organization. Because these relationships vary among community types and phyla [26], different zooplankton communities may differ in how much diversity must be resolved to capture biomass patterns. This is directly relevant to model structure and EOV design because it determines whether zooplankton biomass can be represented by a few dominant groups or requires more detailed diversity-resolved modeling and monitoring.

Extrapolation of sparse quantitative observations to continuous fields using machine learning is increasingly used across ocean biogeochemistry and ecology to reconstruct variables such as carbonate-system properties [27] or species abundance distributions [28]. Biomass distribution models (BDMs) extend this logic to biomass data by relating observations to spatially complete environmental predictors to infer continuous biomass fields [15, 11, 12, 14]. This framework is particularly suited to zooplankton biomass, for which local observations are sparse, heterogeneous, and strongly patchy [7], but whose large-scale spatial and seasonal structure can still be resolved at climatological scales [12]. For zooplankton, BDMs have been used to derive global climatologies of mesozooplankton and calcifier biomass [11, 12] to infer mesopelagic biomass–richness relationships [14] and to map zooplankton functional composition from imaging observations [15].

Here we quantify the global spatial, seasonal, and functional organization of upper-ocean zooplankton biomass by deriving a monthly climatology of zooplankton carbon biomass for 10 zooplankton functional types in the epipelagic (0–200 m) and mesopelagic (200–500 m). We use an ensemble of BDMs trained on the AtlantECO-BASE dataset [29] and implemented within a recent predictive modeling framework [30]. We then examine three aspects of zooplankton biomass organization. First, we quantify the large-scale spatial, seasonal, and vertical structure of total zooplankton biomass across the upper 500 m. Second, we determine how total biomass is partitioned among functional groups and identify which PFTs account for most of the standing stock. Last, we assess how biomass, PFT composition, and species richness are related at the global scale to determine whether biomass is concentrated in a few or many taxa, and whether this differs among PFTs. These analyses provide an observation-based benchmark for marine ecosystem and biogeochemical models and quantitative guidance for zooplankton monitoring and EOV development.

2 | Methods

2.1 | Zooplankton functional types observations

TABLE 1 | Description of the zooplankton PFT dataset built from the AtlantECO-BASE database ([29]). Biomass values are given in mgC m^{-3}.

Dataset	Initial dataset					Epipelagic observations (0–200 m)			Mesopelagic observations (200–500 m)		
PFT	#Obs	% zeros	% species	% genus	Biomass	#Obs	% zeros	Biomass	#Obs	% zeros	Biomass
Amphipoda	25002	11%	10%	12%	30.38 (± 744.56)	11616	8%	5.96 (± 21.78)	329	4%	2.29 (± 10.22)
Appendicularia	27729	5%	5%	92%	0.07 (± 2.78)	13362	4%	0.03 (± 0.22)	214	7%	0.00 (± 0.03)
Chaetognatha	23457	6%	22%	38%	12.25 (± 212.72)	8026	7%	6.64 (± 28.85)	190	0.5%	9.84 (± 64.10)
Copepoda	569441	19%	66%	89%	3.53 (± 142.21)	20262	0.2%	49.48 (± 148.33)	846	0.0%	23.14 (± 42.33)
Euphausiacea	39667	15%	38%	51%	770.19 (± 9426.45)	12614	7%	89.50 (± 446.20)	408	4%	10.61 (± 28.89)
Foraminifera	268524	91%	95%	95%	0.01 (± 2.24)	9221	18%	0.07 (± 0.18)	1162	15%	0.01 (± 0.04)
Jellyfish	80176	42%	24%	44%	25.06 (± 362.14)	12650	14%	15.25 (± 55.78)	374	1%	8.57 (± 33.72)
Ostracoda	25369	9%	1%	4%	1.05 (± 15.18)	11279	10%	0.32 (± 1.26)	490	0.0%	0.38 (± 0.72)
Pteropoda	75466	60%	44%	69%	5.93 (± 128.94)	11280	11%	10.43 (± 40.71)	874	19%	1.85 (± 4.75)
Thaliacea	34910	43%	6%	42%	7.18 (± 180.52)	8322	15%	7.13 (± 35.82)	129	3%	0.47 (± 0.59)

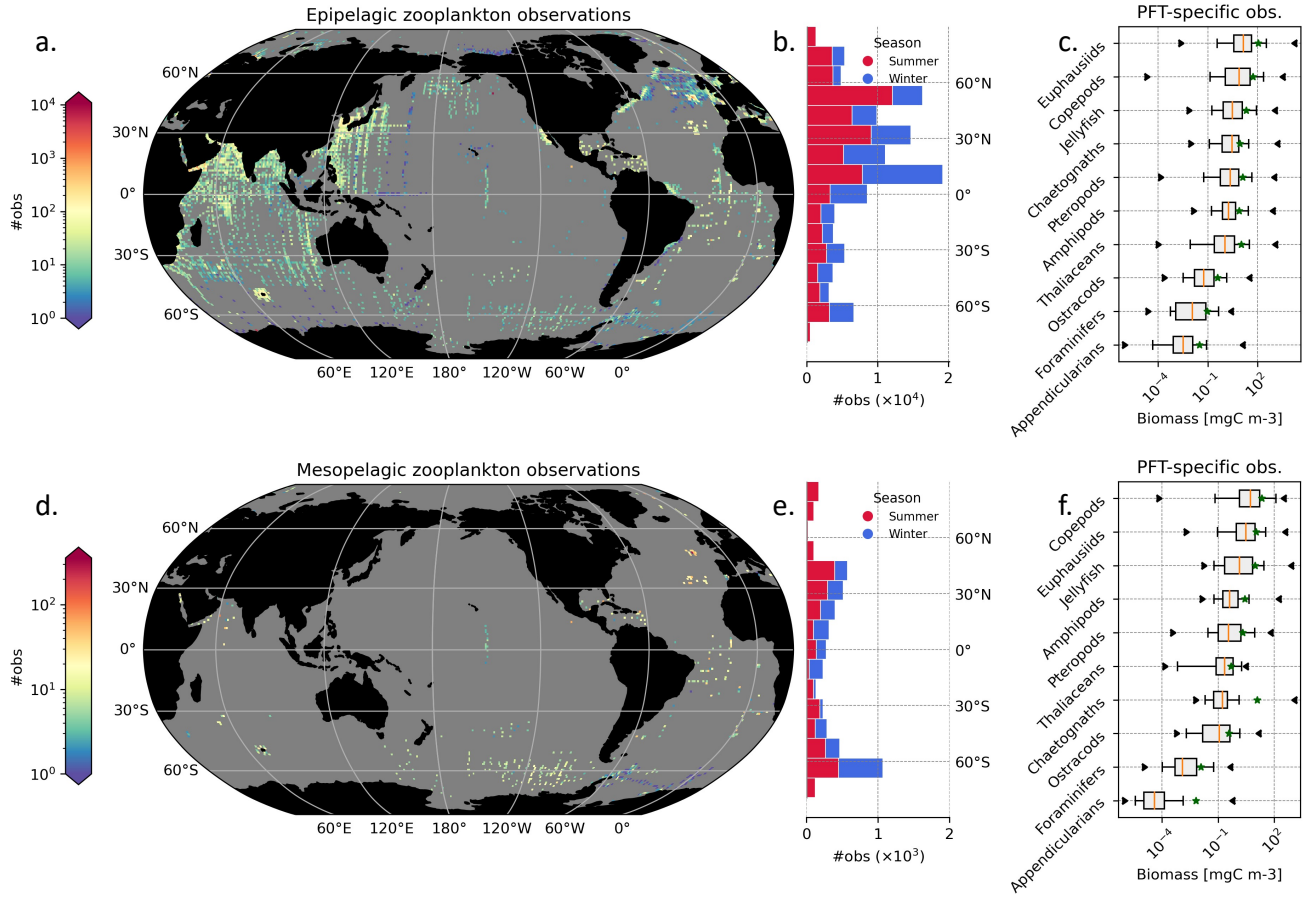

FIGURE 1 | Zooplankton observations distribution. Panels a-c: Epipelagic (0-200 m). a) Number of observations per 1° × 1° grid cell. b) Number of winter (April-September in the Southern Hemisphere; October-March in the Northern Hemisphere) and summer observations per 10° latitudinal band. c) Boxplots of biomass for each PFT, ordered by median (lowest to highest). Panels d-f: same as a-c for the mesopelagic (200-500 m).

Zooplankton abundance observations

We used the AtlantECO-BASE dataset (Table 1, [29], Supplementary Text S1) that compiled geo-referenced measurements of in-situ zooplankton abundances from global data archives to derive PFT-resolved monthly fields of zooplankton biomass (in mg C m^{-3}). The ten PFTs are Chaetognaths, Pteropods, Ostracods, Jellyfish, Thaliaceans, Foraminifers, Copepods, Amphipods, Euphausiids and Appendicularians ([31]. These PFTs were designed to group taxa that occupy similar trophic niches, reflecting shared feeding traits, resource use and predator-prey interactions [32, 31]. Small and large copepods were merged into a single functional type to retain observations recorded as "Copepod". We acknowledge that copepods comprise smaller distinct functional types based on traits such as body size and feeding strategies; this has been studied by [22]. Unfortunately, a fine functional resolution at lower taxonomic levels cannot be achieved consistently

across all PFTs in the current study. This synthesis of zooplankton abundances gathered 1,169,741 observations (Table 1), collected at a mean maximum sampling depth of 270±433 (±SD) m from 1926–2016.

Abundance to biomass conversions

A total of 61% of the zooplankton records were identified to the species level and 18% to the genus level (Table 1). We thus performed taxon-specific conversions from abundances (ind m^{-3}) to biomass concentrations (mg C m^{-3}). We used mean taxon-specific carbon weights (mg C ind^{-1}) based on group-specific length-mass or mass-mass linear and logistic regression equations compiled from different sources listed in Table S1. Not all observations had a species-specific carbon weight in the compilation because they were not identified to the species level (i.e., class-, order-, family- or genus-level observations). In these cases, we used the median carbon weight of the taxa composing the higher-level taxonomic group (e.g., the carbon weight assigned to Salpidae corresponded to the median carbon weight of all Salpidae species). All taxon-specific carbon-weight values used in this study are reported in [ZENODO Folder].

Epi- and mesopelagic zooplankton biomass datasets

For each PFT, zooplankton biomasses from single-net samples were summed when necessary (e.g., when multiple taxa from the same PFT were sorted within a unique sample) to represent a PFT-level point measurement. Observations with missing minimum or maximum tow depths were excluded. We divided the dataset into two depth layers: the epipelagic (0–200 m) and the mesopelagic (200–500 m). We excluded deeper samples because of the very low sampling coverage below 500 m depth (see Fig. S1). Observations were then filtered to include only samples collected with mesh sizes between 100 and 500 μm (85% of the dataset), so that all carbon biomass values could then converted to their 200 μm-equivalent using the conversion factors from [33]. This conversion reduces bias in total zooplankton biomass estimates associated with mesh size, but it does not account for potential differences in species-specific selectivity among mesh sizes. For each dataset, biomass observations deviating from the mean by more than two times the standard deviation were discarded [11]. This led to the exclusion of 0.5% of the dataset, which shows that the data distribution is highly skewed. These could be artificial outliers (faulty measurement or error in reporting an otherwise correct measurement) or natural outliers (e.g., a massive bloom outside the climatological conditions we are trying to capture here).

The resulting epipelagic dataset (Table 1, Fig. 1) contained 118,632 observations of PFT biomass, collected at a mean maximum sampling depth of 98±63 m from 1938-2015. Sampling coverage was strongly spatially biased, with pronounced data gaps in the Arctic Ocean and in the oligotrophic open-ocean gyres of both the Pacific and Atlantic basins. Observed PFT biomasses ranged from 0 to 12,143 mg C m^{-3} (24 ± 180 mg C m^{-3}), with 8% zero values. The mesopelagic dataset (Table 1) contained 5,016 individual observations of PFT biomass, collected at a mean maximum sampling depth of 427±90 m from 1938-2009. Mesopelagic sampling was even more sparse (Fig. 1), with limited coverage in Arctic, Pacific and Indian oceans. Observed PFT biomasses ranged from 0 to 928 mg C m^{-3} (6 ± 29 mg C m^{-3}), with 8% zero values. These datasets were used to fit separate BDMs for the epipelagic and mesopelagic zones (Section 2.2).

2.2 | Biomass distribution models

Environmental predictors

We used 27 environmental monthly climatologies (Table S2) selected as potentially relevant for modeling zooplankton biomass [31, 11, 12, 30]. Where necessary, these environmental predictor fields were averaged and re-gridded to monthly climatologies at a 1°×1° resolution. To reduce collinearity among predictors, we performed two separate principal component analyses (PCAs), one for the epipelagic and one for the mesopelagic zone. Satellite-based predictors were identical in both PCAs, whereas depth-resolved predictors from the World Ocean Atlas (WOA) were selected to match the considered depth layer (epipelagic vs. mesopelagic). In both cases, the first four principal components (PC1–PC4) were retained and used as model inputs, thereby retaining the dominant environmental gradients (70% of the variance explained).

For the epipelagic PCA (Table S3, Fig. S4), the variance explained by PC1–PC4 was 38%, 21%, 9%, and 6%, respectively. PC1 primarily captured temperature and nutrient gradients (e.g., SST: r =0.95; nitrate: r =-0.87), PC2 reflected primary production conditions (chl-a: r =-0.80), PC3 was associated with oxygen availability (r =0.83), and PC4 represented hydrodynamical conditions (mixed layer depth: r =0.61, current intensity: r =0.54). For the mesopelagic PCA (Table S4, Fig. S5), the variance explained by PC1–PC4 was 31%, 22%, 17%, and 6%, respectively. PC1 primarily captured temperature

(r =0.89), PC2 was associated with oxygen availability (r =0.86), PC3 reflected nutrient gradients (nitrate: r =0.62), and PC4 represented hydrodynamical conditions (mixed layer depth: r =0.55, current intensity: r =0.65).

The analysis assumes that the selected set of environmental predictors is sufficiently diverse to represent the range of conditions sampled by the observations. Consequently, we focus in the results section on predictive performance and estimated PFT biomass distributions, and treat the retained PCs primarily as a compact representation of the environmental space rather than as evidence of causal drivers.

Habitat model pipeline

Biomass distribution models were run using the Comprehensive Ensemble Pipeline for Habitat modeling Across Large-scale Ocean Pelagic Observation Datasets (CEPHALOPOD; 30), a framework specifically tailored to extrapolate scarce and biased marine biological observations. We applied the six algorithms implemented in CEPHALOPOD: generalized additive models (GAM), generalized linear models (GLM), random forests (RF), boosted regression trees (BRT), multilayer perceptrons (MLP), and support vector machines (SVM). The full modeling workflow is described in [30]; here we report only the elements most relevant to the present study.

Model evaluation relied on the three CEPHALOPOD quality criteria relevant to this study. First, environmental predictors had to explain >5% more variance than randomly permuted predictors [QC#1: pre-VIP $>$ 0.05; 30], ensuring that fitted relationships were more informative than a null expectation. Second, predictive skill was evaluated from spatial block cross-validation using the coefficient of determination [QC#2: R^2; 30]. In this study, we adopted a less conservative acceptance threshold of $R^2 \geq 0.1$ because PFT-resolved biomass observations are highly patchy and methodologically heterogeneous, and similar thresholds have been used in previous biomass distribution studies [15, 11]. Last, prediction uncertainty estimated from bootstrap replicates had to remain moderate [QC#3: normalized standard deviation, NSD < 0.5; 30]. In addition, to homogenize the distribution of observations across environmental conditions before model fitting, we applied a thinning procedure in the four-dimensional environmental space defined by the first four principal components (PC1–PC4): each axis was discretized into 15 bins, and observations within each occupied environmental cell were aggregated by their median biomass. We also computed a Multivariate Environmental Similarity Surfaces (MESS) [34] to identify geographic areas where the algorithm extrapolates beyond the range of observed environmental conditions, and thus should be interpreted with caution.

2.3 | Experimental design and analyses

CEPHALOPOD experiments

For each PFT, CEPHALOPOD was applied separately to the epipelagic and mesopelagic datasets. Quality checks were evaluated independently for each group, and algorithms were retained if they satisfied all selected criteria or, to preserve a complete PFT representation in the integrated analyses, if one QC metric fell within 30% of the limit defined for each metric. Final PFT climatologies were computed as the ensemble mean of the retained algorithms and produced at $1° \times 1°$ spatial resolution with a 12-month temporal coverage.

Richness-biomass relationship

To compare our PFT-specific zooplankton biomass estimates to PFT-specific species-richness estimates, we used the global species richness maps of [31], derived from compilations of plankton species occurrences and an ensemble species distribution modeling framework [35]. Richness patterns were available for all PFTs considered here except Ostracoda, yielding 9 PFT-specific diversity-biomass analyses.

To describe the global (all-PFT) non-linear association between total richness R and total biomass B, we fitted a GLM with a Gamma error distribution and a log link to biomass data: $\log \mathbb{E}[B] = \beta_0 + \beta_1 R + \beta_2 R^2$. We compared the quadratic model to a linear alternative using Akaike's Information Criterion (AIC), and retained the quadratic form with the lower AIC value (ΔAIC = 1266.53). We computed the turning point, $R_0 = -\beta_1/(2\beta_2)$, to partition observations into low- and high-richness subsets ($R < R_0$ and $R \geq R_0$). To focus on high-biomass conditions and allow for distinct regimes, we defined high-biomass grid cells as those exceeding the top 20% of $\log_{10}(B)$ and classified them according to whether they fell on the low- or high-richness side of R_0. We additionally examined the spatial structure of these regimes using the biomass-to-richness ratio B/R.

To analyse PFT-specific richness-biomass relationships, we fitted GLMs with a Gamma error distribution and log link (`biomass ~ richness + PFT + richness:PFT`), allowing slopes to vary among PFTs via the richness×PFT interaction. Note that we used a linear richness effect within each PFT so that the effect direction and magnitude are directly comparable across functional types as a constant per-unit change. Effect sizes were therefore expressed as the PFT-specific percent change in expected biomass per +1 richness unit ($100[\exp(\beta_{R,\mathrm{PFT}})-1]$, where $\beta_{R,\mathrm{PFT}}$ denotes the fitted PFT-specific richness coefficient on the log scale).

3 | Results

3.1 | Model evaluation

TABLE 2 | Quality criteria (QC) values for the ten PFTs. Detailed QC from each algorithm are given in table S4. R^2 : Univariate coefficient of determination [QC $R^2 \geq 0.1$]. NSD : Normalised standard deviation (NSD) [QC: NSD < 0.5]. Pre-VIP difference in proportion of variance explained by the pre-selected environmental predictors with the proportion of variance explained by a set of randomly permuted (NULL) predictors [QC: preVIP > 0.05].

	Epipelagic (0-200 m)				Mesopelagic (200-500 m)			
PFT	R^2	**NSD**	**preVIP**	**models**	R^2	**NSD**	**preVIP**	**models**
Amphipods	0.18	0.19	0.14	MLP+GAM+SVM+RF	0.16	0.25	0.16	GLM+MLP+GAM+SVM+RF
Appendicularians	0.13	0.23	0.11	RF				
Chaetognaths	0.21	0.23	0.14	MLP+SVM+RF				
Copepods	0.18	0.40	0.05	MLP+SVM+RF	0.23	0.35	0.19	GLM+MLP+GAM+SVM+RF
Euphausiids	0.14	0.39	0.08	BRT+SVM+RF	0.21	0.38	0.14	GLM+MLP+SVM+RF
Foraminifers	0.34	0.19	0.23	GLM+MLP+GAM+SVM+RF	0.15	0.36	0.08	SVM+RF
Jellyfishes	0.15	0.48	0.08	MLP+SVM+RF	0.40	0.35	0.27	SVM+RF
Ostracods	0.14	0.20	0.09	RF	0.25	0.32	0.24	GLM+MLP+GAM+SVM+RF
Pteropods	0.15	0.31	0.03	SVM+RF	0.25	0.31	0.22	GLM+MLP+SVM+RF
Thaliaceans	0.15	0.30	0.15	MLP+SVM+RF	0.28	0.42	0.26	GLM+MLP+BRT+GAM+SVM+RF

Overall, 42% of the BDMs and 53% of the BDMs passed all quality criteria for the epipelagic and mesopelagic layers, respectively (Table S5). Most of the BDMs (i.e., two of three) that did not pass these criteria did so because of insufficient predictive performance. When multiple algorithms passed, they generally produced similar spatial patterns (Fig. S10). For epipelagic Pteropoda, one algorithm narrowly missed the pre-VIP threshold (0.03) and was retained, so that all epipelagic PFTs and 8 of 10 mesopelagic PFTs had at least one selected algorithm (Table 2). In contrast, mesopelagic Chaetognatha and Appendicularia models showed low predictive skill across algorithms ($R^2 \leq 0.01$), indicating that the selected predictors explain little of the observed variance for these groups. Because observed medians for these groups were two orders of magnitude lower than total mesopelagic zooplankton (Table 1, Fig. 1), excluding them is unlikely to influence the total mesopelagic biomass patterns presented in the next section.

The epipelagic MESS analysis indicated that environmental conditions across >95% of the mapped area fell within the multidimensional envelope defined by the training observations, suggesting limited extrapolation (Fig. 2a). In contrast, the mesopelagic MESS identified extrapolation over 30% of the mapped area, primarily in the Arctic and in oligotrophic gyres, which coincide with regions lacking biomass observations (Fig. 2d). The environmental predictors for each PFT model of biomass distribution varied in their relative importance (Fig. S6,S7,S8 and S9).

3.2 | Global zooplankton biomass distribution

The global annual mean epipelagic zooplankton biomass was estimated at 2.9 ± 0.6 PgC (Table 3). The spatial distribution of epipelagic zooplankton biomass followed productivity gradients (Fig. 2(a)), with the highest biomass (>70 mg C m^{-3}) found at high latitudes, in coastal areas, and in low latitude upwelling systems. Lowest biomass (< 20 mg C m^{-3}) was found in oligotrophic gyres. Biomass patterns showed small (< 20%) seasonal variations within tropical bands ($\pm 40°$) (Fig. 2(a)). However, a clear seasonal cycle emerged poleward of 40° latitude, with biomass reaching up to twice as high in summer as in winter (Fig. 2(b)). The Arctic Ocean exhibited the strongest seasonal variability in biomass, although this result should be interpreted with caution due to limited model coverage in regions where sea-ice conditions resulted in missing predictors.

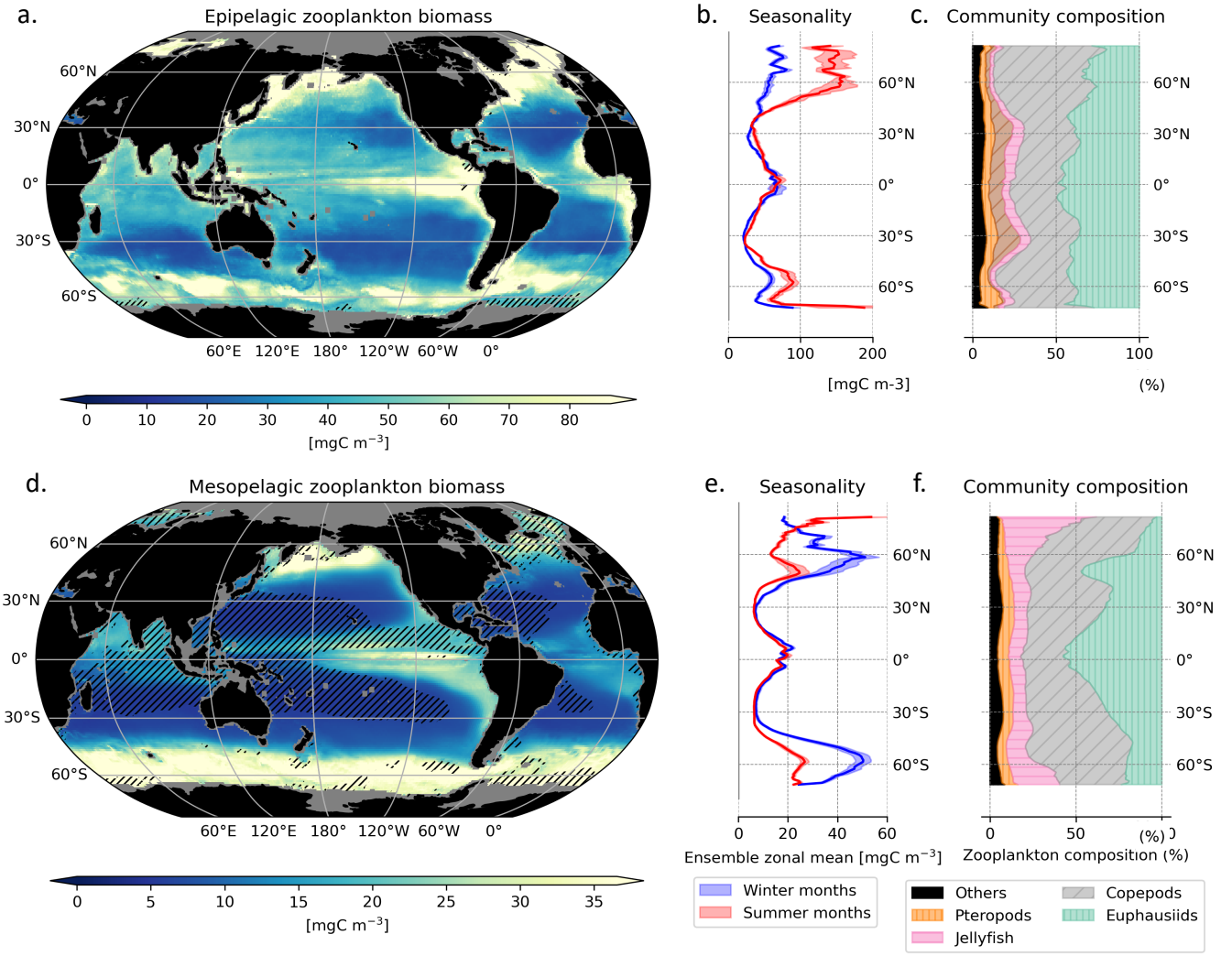

FIGURE 2 | Zooplankton biomass distribution and composition. Panels a-c: Epipelagic (0-200 m). a) Zooplankton biomass distribution obtained by summing the distribution of plankton functional types simulated by biomass distribution models. Hatching indicates areas where model predictions fall outside the training conditions (MESS < 0) for at least one PFT in at least three months. b) Winter (June-August in the Southern Hemisphere; December-February in the Northern Hemisphere) and summer zonal profile of total zooplankton biomass distribution. Uncertainty corresponds to the standard deviation of the ten boostraps produced by CEPHALOPOD c) Zonal profile of the relative contributions of each PFT to the total zooplankton biomass. Panels d-f: same as a-c for the mesopelagic (200-500 m).

Mesopelagic zooplankton biomass was lower (0.9 ± 0.2 PgC, Table 3) than epipelagic biomass. The mesopelagic zooplankton biomass distribution showed a clear latitudinal pattern (Fig. 2(d)): within tropical bands ($\pm 40°$), mean biomass varied between 10 and 20 mg C m^{-3}, with the minimum located in oligotrophic gyres. However, biomass increased sharply around 50° N and 50° S to values above 40 mg C m^{-3}. This latitudinal structure persisted across seasons, but high-latitude mesopelagic biomass was enhanced in winter, where winter biomass was $\approx 2\times$ the summer biomass poleward of 40° in both hemispheres (Fig. 2(e), S3).

3.3 | Community composition

Within each depth layer, modeled and observed PFT biomasses showed the same rank order among PFTs (Tables 1, 3; Fig. 2c,f). Community composition was strongly dominated by crustaceans in both layers: euphausiids and copepods represented $\sim$ 41% and $\sim$ 36% of the epipelagic biomass, respectively, and $\sim$ 33% and $\sim$ 43% of the mesopelagic biomass. In the epipelagic, secondary contributors included chaetognaths ($\sim$ 7%), jellyfish ($\sim$ 5%), and pteropods ($\sim$ 5%), followed by amphipods ($\sim$ 4%) and thaliaceans ($\sim$ 2%), whereas ostracods, foraminifers, and appendicularians each contributed $<$ 1%. In the mesopelagic, jellyfish ($\sim$ 12%) and pteropods ($\sim$ 6%) represented the main non-crustacean components,

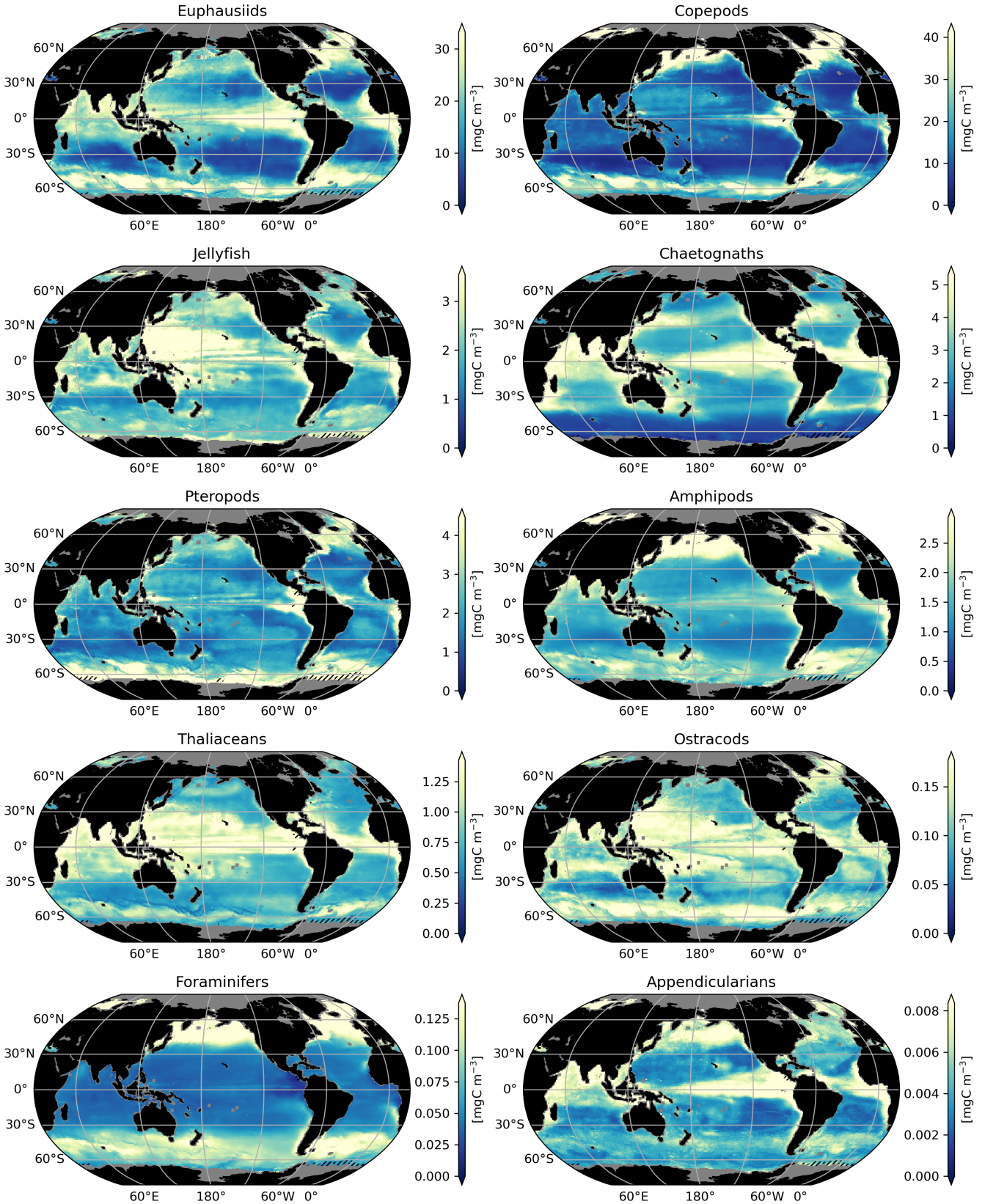

FIGURE 3 | PFT-specific epipelagic (0-200 m) biomass distribution simulated by the biomass distribution models. Hatching indicates areas where model predictions fall outside the training conditions (MESS < 0) for at least three months. Maps are ranked by total biomass.

while amphipods, thaliaceans, ostracods, and foraminifers each accounted for only a few percent or less; the modeled biomass for chaetognaths and appendicularians were not retained for this layer (Table 3).

The dominant epipelagic crustacean PFTs exhibited contrasting spatial patterns (Fig. 3). Amphipods and copepods showed broadly similar distributions, with highest biomasses at high latitudes and in major upwelling systems (amphipods $>$ 3 mgC m^{-3}; copepods $>$ 40 mgC m^{-3}) and markedly lower biomasses in oligotrophic subtropical gyres (amphipods $<$ 10 mgC m^{-3}; copepods $\sim$ 2 mgC m^{-3}). Ostracods also peaked at high latitudes ($>$ 0.2 mgC m^{-3}), but, in contrast to amphipods and copepods, they displayed local maxima in tropical oligotrophic gyres ($>$ 0.2 mgC m^{-3}). Euphausiids showed a distinct pattern: they dominated the tropical and subtropical community, reaching up to $\sim$ 40% of total zooplankton biomass, but their relative contribution declined poleward as copepods become dominant. Euphausiid biomass was minimal in tropical oligotrophic gyres ($<$ 10 mgC m^{-3}).

Non-crustacean epipelagic PFTs also displayed substantial spatial variability (Fig. 3). Gelatinous zooplankton had gradients broadly opposite to those of copepods, with higher biomass near the equator (jellyfish $>$ 3 mgC m^{-3}; chaetognaths $>$ 4 mgC m^{-3}; thaliaceans $>$ 1.5 mgC m^{-3}) and lower biomass toward higher latitudes (jellyfish $<$ 2 mgC m^{-3}; chaetognaths $<$ 1 mgC m^{-3}; thaliaceans $<$ 1 mgC m^{-3}). Appendicularians showed a stronger decline in oligotrophic gyres (a decrease of $>$ 80% relative to their maximum values) compared to the other gelatinous PFTs. Calcifying zooplankton exhibited contrasting patterns: pteropods broadly followed the copepod distribution, with elevated biomasses at high latitudes and in upwelling regions ($>$ 3 mgC m^{-3}) and reduced biomass at low latitudes ($<$ 2 mgC m^{-3}; $\sim$ 2 mgC m^{-3}). Foraminifers also peaked at high latitudes ($>$ 0.1 mgC m^{-3}) but decreased more uniformly toward the equator ($<$ 0.03 mgC m^{-3}) without a peak in upwelling areas.

Compared to the epipelagic, spatial patterns for mesopelagic biomass were more similar across PFTs (Fig. 3; Fig. S2). At high latitudes, mean biomass was high (exceeding the median) for all PFTs. All mesopelagic groups also exhibited a pronounced biomass minimum in oligotrophic subtropical gyres. These features showed that some groups, particularly gelatinous PFTs (thaliaceans, jellyfish, and ostracods), had latitudinal gradients that were the reverse of their epipelagic distributions, and thus resembling the pattern observed for mesopelagic copepods. In contrast, other groups (euphausiids, pteropods, amphipods, and foraminifers) showed elevated biomass both at high latitudes and in upwelling regions. These depth-dependent differences in spatial patterns were reflected in community composition. In the epipelagic, gelatinous zooplankton contributions peaked in the tropics (up to 15%; Fig. 2c), whereas in the mesopelagic they peaked at high latitudes (up to 20%; Fig. 2f). Conversely, euphausiids exhibited a stronger tropical maximum in the mesopelagic, strengthening the contrast in community structure between depth layers (Figs. 2c,f).

3.4 | Zooplankton biomass-richness relationship

At the global scale, the relationship between the biomass of all zooplankton PFTs combined ($\log_{10}$(biomass)) and species richness was U-shaped (Fig. 4b). A quadratic model explained 23% of the variance in $\log_{10}$(biomass) ($R^2 = 0.23$) and yielded a turning point at a richness of $x_0 = 180$ (Fig. 4a). To interpret this pattern spatially, we focused on high-biomass conditions by selecting the upper 20% of biomass values ($B \geq 62$ mgC m^{-3}) and then partitioned these grid cells into low- and high-richness subsets using x_0 (Fig. 4a). The high-richness/high-biomass subset was concentrated in low latitude upwelling systems (*80%* of this subset lies within |lat| $<$ 30°), whereas the low-richness/high-biomass subset was predominantly associated with high latitudes (*97%* of this subset lies within |lat| $\geq$ 50°). This indicates the coexistence of at least two geographically distinct high-biomass regimes. Consistent with this spatial segregation, the biomass-to-richness ratio was higher in the high-latitude high-biomass regime ($>$ 1.4 mgC m^{-3} species^{-1}) than in the low-latitude high-biomass regime, indicating that fewer taxa accounted for a comparatively large biomass at high latitudes (Fig. 4a).

Our analysis revealed how PFT composition contributed to the biomass-richness relationship (Fig. 4c). PFTs that contributed most to total biomass also tended to be the most species-rich ($\rho_{spearman} = 0.7$, $p = 0.03$, computed on the PFT centroids shown on Fig. 4c). Using a Gamma GLM with a log link, we found that the biomass-richness scaling differed among PFTs. Biomass-richness relationships were strongly PFT-dependent, as indicated by the significant richness $\times$ PFT interaction terms (all $p < 0.001$). All crustacean and calcifying PFTs showed negative biomass-richness scaling (e.g., foraminifers −7% per added species; amphipods −4%; pteropods −4%; euphausiids −2%; copepods −1%, Fig. S11), whereas gelatinous PFTs exhibited positive scaling (e.g., appendicularians +25% per added species; chaetognaths +9.0%; thaliaceans +11%; jellyfish +0.5%, Fig. S11).

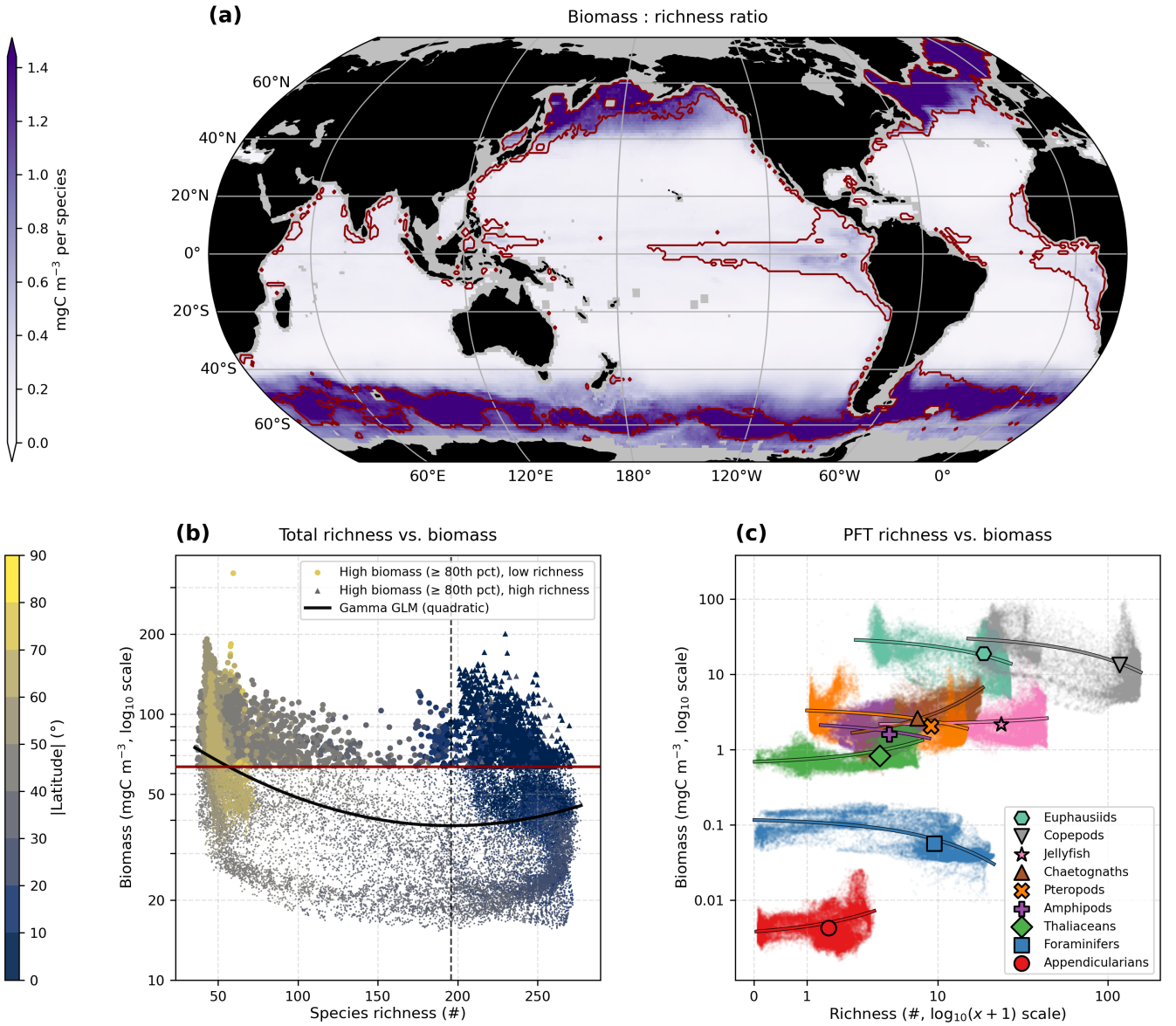

FIGURE 4 | **Global zooplankton biomass-richness relationship(s).** (a) Relationship between $\log_{10}$(biomass) and species richness across the 1° global grid, with points coloured by absolute latitude; the black curve shows the fitted quadratic Gamma GLM (log link) for expected biomass and the dashed vertical line marks the turning point (R_0). High-biomass grid cells (above the 80th percentile of $\log_{10}$(biomass); dark red horizontal line) are highlighted and separate into low- and high-richness regimes. The 95% confidence band for the fitted GLM mean is not visually apparent because it is very narrow (median width ~2.0%). (b) Global map of the biomass:richness ratio (biomass per species), with the dark red contour delineating grid cells where zooplankton biomass exceeds the 80th percentile. (c) PFT-specific zooplankton biomass-richness relationship. One point corresponds to a 1° × 1° grid cell for one PFT, colours indicate PFTs and symbols indicate the centroid for each PFT. Solid lines show PFT-specific Gamma GLM (log link) fits, plotted over the observed richness range of each PFT (shaded bands indicate 95% confidence intervals)

4 | Discussion

To our knowledge, this study provides the first global monthly climatology of zooplankton carbon biomass partitioned among 10 PFTs and resolved separately for the epipelagic and mesopelagic (Fig. 2). The resulting maps (Fig. 3) show that global zooplankton biomass is strongly structured by a few dominant functional groups: crustaceans dominate the standing stock in both layers, with copepods and euphausiids accounting for most of the biomass. These groups capture much of the large-scale biomass structure and may therefore provide a useful first-order description for monitoring and model evaluation [3]. However, a fuller representation of ecosystem structure also requires resolving other PFTs, notably gelatinous taxa and calcifiers, because they contribute importantly to biomass (Table 3) and biogeochemical function [19, 37].

TABLE 3 | Comparison of zooplankton biomass estimates (Mt C) in the epipelagic (0–200 m) and mesopelagic (200–500 m) layers between this study and published global estimates.

PFT / group	Layer	This study	Buitenhuis	Knecht	Hatton	Drago	Lombard (%)	Hernández-León	Egorova	Lucas
Appendicularians	Epipelagic	0.3 (<0.1%)	–	–	–	~22 (~9.6%)	–	–	–	–
	Mesopelagic	–	–	–	–	~14 (~8.1%)	–	–	–	–
Foraminifers	Epipelagic	4 ± 1 (0.1%)	0.9–2	1–3	–	~7 (~3.1%)	–	–	–	–
	Mesopelagic	0.9 ± 0.1 (0.1%)	–	–	–	~12 (~6.9%)	–	–	–	–
Ostracods	Epipelagic	8 (0.3%)	–	–	–	~1.5 (~0.7%)	–	–	–	–
	Mesopelagic	11 ± 2 (1.2%)	–	–	–	~2 (~1.2%)	–	–	–	–
Thaliaceans	Epipelagic	60 ± 9 (2.1%)	–	–	–	–	–	–	–	6.5 (0.3–133.5)[§]
	Mesopelagic	10 ± 3 (1.1%)	–	–	–	–	–	–	–	–
Amphipods	Epipelagic	104 ± 15 (3.6%)	–	–	–	–	–	–	–	–
	Mesopelagic	28 ± 4 (3.0%)	–	–	–	–	–	–	–	–
Jellyfish	Epipelagic	148 ± 22 (5.1%)	–	–	–	–	–	–	–	402.5 (49.9–4228.3)[§]
	Mesopelagic	110 ± 18 (11.7%)	–	–	–	–	–	–	–	–
Pteropods	Epipelagic	150 ± 26 (5.2%)	26–670	45–155	–	–	–	–	–	–
	Mesopelagic	59 ± 11 (6.3%)	–	–	–	–	–	–	–	–
Chaetognaths	Epipelagic	198 ± 23 (6.8%)	–	–	–	~6 (~2.6%)	~4%[†]	–	–	–
	Mesopelagic	–	–	–	–	~9 (~5.2%)	~4%[†]	–	–	–
Euphausiids	Epipelagic	1041 ± 258 (35.8%)	–	–	–	–	–	–	–	–
	Mesopelagic	408 ± 102 (43.4%)	–	–	–	–	–	–	–	–
Copepods	Epipelagic	1194 ± 199 (41.1%)	–	–	–	83 (36.2%)	~35%[†]	–	–	–
	Mesopelagic	314 ± 55 (33.4%)	–	–	–	61 (35.3%)	~35%[†]	–	–	–
Crustaceans[a]	Epipelagic	2347 (80.7%)	–	–	–	157 (68.4%)	~50%[†]	–	–	–
	Mesopelagic	761 (80.9%)	–	–	–	129 (74.4%)	~50%[†]	–	–	–
Large crustaceans[b]	Epipelagic	1145 (39.4%)	–	–	–	58 (25.3%)	~15%[†]	–	–	–
	Mesopelagic	436 (46.3%)	–	–	–	49 (28.3%)	~15%[†]	–	–	–
Gelatinous taxa[c]	Epipelagic	406 (14.0%)	–	–	–	–	~10%[†]	–	–	409.0 (50.2–4361.8)[§]
	Mesopelagic	120 (12.8%)	–	–	–	–	~10%[†]	–	–	–
Total zooplankton[d]	Epipelagic	2907 ± 552 (100%)	550–2110	–	55–3407	229 (100%)	100%[†]	480	–	–
	Mesopelagic	941 ± 194 (100%)	–	–	–	173 (100%)	100%[†]	660	200–910[‡]	–

Values from this study are biomass ± inter-algorithm SD (when > 1 algorithm passed QC), with percentage contribution to total layer biomass in parentheses; ~ indicates visual approximation from published figures. Comparison values come from [7, 11, 2, 15, 16, 13, 14, 36]. [†] 0–500 m integrated, not layer-specific. [‡] 200–1000 m mesopelagic definition. [§] Converted epipelagic comparators not fully matched to our PFT definitions. [a] Crustacea = copepods + euphausiids + amphipods + ostracods here; copepods + large crustaceans in [16]. [b] Large crustaceans = amphipods + euphausiids here; closest to Eumalacostraca in [15]. [c] Gelatinous taxa = thaliaceans + appendicularians + chaetognaths + jellyfish here; gelatinous filter feeders + chaetognaths + gelatinous carnivores in [16]; Thaliacea + Cnidaria + Ctenophora in [36]. [d] [7] and [2] totals are broader mesozooplankton + macrozooplankton aggregates.

4.1 | Global zooplankton biomass and its seasonality

Our epipelagic annual-mean zooplankton biomass estimate of 2.9 ± 0.6 PgC falls within the upper range of previous net-based estimates (0.2-3.4 PgC, Table 3), while the mesopelagic estimate (0.9 ± 0.2 PgC) is also within the envelope of recent observation-based products (0.2-0.9 PgC, Table 3). Direct observation-based estimates are, however, uncertain (Table 3). Such variability likely reflects both real spatial heterogeneity in zooplankton biomass, driven by small-scale physical processes (e.g., eddies; 38) and biological processes (e.g., swarm dynamics; 19) [7], and methodological differences among studies, including sparse data and differing biomass modelling approaches [8]. By using BDMs, we extract a climatological signal focused on large-scale variation, which is more comparable to the scales typically resolved by global biogeochemical models [8, 39, 10].

Spatially, mesopelagic biomass broadly mirrors the epipelagic pattern, with high values at high latitudes and in upwelling systems and low values in oligotrophic gyres, consistent with epipelagic–mesopelagic coupling through vertical migration and particle export [1]. Seasonality differs, however, between layers: epipelagic biomass peaks during summer at high latitudes, whereas mesopelagic biomass is elevated during winter. We hypothesize that this reversed seasonal pattern reflects seasonal vertical migration and winter diapause in high-latitude taxa such as copepods [20, 40, 41], suggesting that diapause may be an important process to represent in zooplankton-enabled biogeochemical models [42].

4.2 | Zooplankton community composition and PFT-specific patterns

Among the 10 PFTs, crustaceans dominate pelagic zooplankton biomass in both layers, consistent with previous studies (Table 3). Copepods show high-latitude and upwelling-enhanced epipelagic biomass distribution that matches independent observation and model-based estimates [12, 14, 43, 15, 10]. Euphausiids provide a notable contrast: their predicted contribution peaks in tropical-subtropical regions, particularly in the upwelling systems, consistent with the distribution reported by [15] and the reported importance of krill in these regions [44, 45, 46]. Their low predicted contribution at high latitudes, despite well-documented high Southern Ocean krill biomasses [47, 48], most likely reflects sampling biases in the underlying dataset, which was compiled primarily from observations collected with fine mesh nets that are generally not towed at speeds optimized to capture large, fast-swimming euphausiids. This limitation primarily affects Southern Ocean euphausiids and should therefore be interpreted as an underestimate rather than a failure of the biomass

distribution model. Amphipod biomass is concentrated in high-latitude and upwelling systems, whereas ostracods display elevated values in oligotrophic gyres. Despite the sparse observations available [49, 50], these climatologies provide large-scale constraints for two groups that remain essentially absent from other global biomass products.

Beyond crustaceans that dominate zooplankton biomass, our PFT climatologies also provide new large-scale constraints for gelatinous taxa and calcifiers (Table 3). Gelatinous zooplankton contribute six times less carbon biomass than crustaceans, but their high volume-to-dry weight ratios imply a larger contribution to biovolume [51, 16]. Appendicularian biomass is likely to be underestimated because mucous houses are typically not sampled [52]. Even so, thaliaceans, jellyfish, and chaetognaths reach relatively high biomass at low latitudes, consistent with previous studies suggesting a competitive advantage of gelatinous groups in low- to mid-nutrient environments [53, 54, 55]. This improved large-scale constraint is especially relevant because gelatinous taxa may become increasingly favored under climate change [56, 57, 58, 52]. Our pteropod and foraminifer estimates lie near the upper end of recent values (Table 3) and provide valuable constraints for quantifying the carbonate pump [37], a process through which calcifying zooplankton may strongly influence the future ocean carbon cycle [59]. Together, these PFT-specific climatologies provide useful first-order benchmarks for monitoring [5] and for mechanistic models that aim to move beyond bulk zooplankton compartments [10].

4.3 | Richness-biomass relationship(s)

These differences in the zooplankton community structure raises the question of whether zooplankton biomass can be further refined by diversity metrics, especially given recent evidence that even within dominant groups such as copepods, substantial functional sub-structure exists, with multiple sub-PFTs playing distinct ecological and biogeochemical roles [22]. At the global scale, epipelagic zooplankton biomass exhibits a U-shaped relationship with richness, reflecting two geographically distinct high-biomass regimes (Fig. 4). At high latitudes, high biomass-to-richness ratios suggest that community biomass is driven by a few dominant taxa, consistent with the mass-ratio hypothesis [23] (e.g., 24). In contrast, the low-latitude high-biomass regime is associated with comparatively lower biomass per species, consistent with niche complementarity in diverse communities [25]. We note, however, that this pattern may also be partly idiosyncratic: biomass hotspots may primarily reflect high resource availability [60], whereas richness may be structured by temperature [61], such that their covariance generates a U-shape without a direct causal effect of richness on biomass.

From a monitoring perspective, this total-community pattern suggests that high-latitude environments may be comparatively easier to characterize because a large fraction of biomass is associated with relatively few dominant taxa (Fig. 4). This is especially relevant for the dominant crustacean groups, copepods and euphausiids, which represent most of zooplankton biomass and species richness at the global scale and show negative biomass-richness relationships in our analysis (Fig. 4). Consequently, observations targeting these groups may provide a particularly efficient description of high-latitude zooplankton communities. However, this simplification does not hold uniformly across functional types and regions. Within our analysis, biomass–richness relationships differ among PFTs: we detect negative relationships at the global scale for crustacean PFTs (consistent with [24]), but positive relationships for gelatinous zooplankton PFTs (Fig. 4). By contrast, the positive biomass-richness relationships reported by [26] for copepods in mid-latitude coastal systems suggest that these relationships may also vary across regions. For monitoring and EOV development, this implies that priority groups and metrics should be defined by both PFT and region, rather than assuming that biomass provides a universal proxy for diversity.

4.4 | Limitations and perspectives

Although biomass is harder to model than occurrence because it also reflects trophic and aggregation processes [28], model selection, cross-validation, and uncertainty analyses within CEPHALOPOD [30] show that the main PFTs respond to predictors consistently enough to recover coherent large-scale spatial and seasonal patterns. However, there are several sources of uncertainty. First, most observations provide net-based abundance rather than in situ carbon biomass, so biomass must be inferred indirectly [8]. Second, net samples are not an unbiased measure of the in situ community: mesh-size effects can lead to extrusion of small individuals, avoidance by larger individuals can reduce the sampled abundance of larger taxa, and samples are often skewed toward larger, frequently adult, stages, all of which affect abundance-to-biomass conversions [8]. We addressed part of this problem by applying a first-order correction for mesh-size effects and, when direct biomass measurements were unavailable, converting abundance to carbon biomass with taxon-specific factors rather than generic bulk relationships. However, these conversions are often based on relatively few samples in

particular areas [62]. Applying these steps are the most robust way for estimating zooplankton biomass [53], but they do not remove all sampling biases. Third, abundance of gelatinous taxa such as jellyfish, appendicularians, thaliaceans, and chaetognaths are underestimated in net samples [54, 52]. Last, the epipelagic and mesopelagic climatologies should be interpreted as time-averaged layer occupancy rather than strictly distinct communities, because diel and seasonal migrants can contribute to both layers over the annual cycle [20].

Taken together, our products should therefore be interpreted as an observation-constrained description of large-scale zooplankton biomass organization rather than as a precise estimate of absolute biomass for every region and taxon. Even so, by identifying where biomass is concentrated, which PFTs dominate, and how biomass–richness relationships differ among PFTs, it helps define which components of the zooplankton community are most informative for ecosystem modeling and monitoring. In terms of modeling the data products not only provides a basis for model evaluation, but also for model development that can help transform sparse observations into consistent zooplankton monitoring products. Future progress will depend on several key developments: improved observations of fragile taxa (e.g., appendicularians and jellyfish); and better sampling of poorly targetted groups (e.g., amphipods and ostracods); systematic comparison across complementary observing systems, including nets, imaging, omics, and acoustics [9]; and encouragement of field observationalists to deposit standardized, well-documented datasets in open-access repositories [3].

Author Contributions

C.C., A.S., F.B., M.V. were responsible for the conceptualization of this study; Data curation was conducted by C.C., A.S., F.B., U.H.E., S.L., R. B. and S. V. ; Analysis and Investigation was led by C.C., with contribution by A.S., F.B., M.V. and A.R.; Funding was acquired by M.V.; Writing by C.C., A.S., F.B. and M.V. , and the feedback of all other authors.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 101094227 (Bluecloud2026), no. 101059915 (BIOcean5D) and no. 862923 (AtlantECO). This output reflects only the author's view, and the European Union cannot be held responsible for any use that may be made of the information contained therein. The authors want to thank all the people involved in the AtlantECO database initiative and related scientific cruises for making their data publicly available. MV acknowledges members of the MAREDAT team for inspiration and the first data collection and thoughts on biomass conversion, upon which the current one is built. CC acknowledges the Cnes-Tosca project SO-BIO-SAT. AR acknowledges the Australian Research Council Centre of Excellence for Our Future Oceans (CE260100001).

Financial Disclosure

None reported.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Deborah K Steinberg, and Michael R Landry. "Zooplankton and the ocean carbon cycle." *Annual review of marine science*9 (2017): 413–444.
2. Ian A Hatton, Ryan F Heneghan, Yinon M Bar-On,, and Eric D Galbraith. "The global ocean size spectrum from bacteria to whales." *Science advances* 7, no. 46 (2021): eabh3732.
3. Lavenia Ratnarajah, Rana Abu-Alhaija, Angus Atkinson, Sonia Batten, Nicholas J Bax, Kim S Bernard, Gabrielle Canonico, Astrid Cornils, Jason D Everett, Maria Grigoratou, et al. "Monitoring and modelling marine zooplankton in a changing climate." *Nature Communications* 14, no. 1 (2023): 564.
4. Camille Richon, Charlotte Wagner, Elsie M Sunderland, Sakina-Dorothée Ayata,, and Alessandro Tagliabue. "A global biogeography analysis reveals vulnerability of surface marine zooplankton to anthropogenic stressors." *One Earth* 7, no. 1 (2024): 146–160.
5. Patricia Miloslavich, Nicholas J Bax, Samantha E Simmons, Eduardo Klein, Ward Appeltans, Octavio Aburto-Oropeza, Melissa Andersen Garcia, Sonia D Batten, Lisandro Benedetti-Cecchi, David M Checkley Jr, et al. "Essential ocean variables for global sustained observations of biodiversity and ecosystem changes." *Global change biology* 24, no. 6 (2018): 2416–2433.
6. Belén Martín Míguez, Emma Heslop, Narissa Bax, Lisandro Benedetti-Cecchi, Gabrielle Canonico, Kim Currie, Karen Evans, Albert S Fischer, Véronique Garçon, Maria Hood, et al. "GOOS Essential Ocean Variables: the backbone of a sustained and evolving global ocean observing system." *Frontiers in Marine Science*13 (2026): 1737002.

7. Erik T Buitenhuis, Meike Vogt, Róisín Moriarty, Nina Bednaršek, Scott C Doney, Karine Leblanc, Corinne Le Quéré, Y-W Luo, Colleen O'Brien, Todd O'Brien, et al. "MAREDAT: towards a world atlas of MARine Ecosystem DATa." *Earth System Science Data* 5, no. 2 (2013): 227–239.
8. Jason D Everett, Mark E Baird, Pearse Buchanan, Cathy Bulman, Claire Davies, Ryan Downie, Chris Griffiths, Ryan Heneghan, Rudy J Kloser, Leonardo Laiolo, et al. "Modeling what we sample and sampling what we model: challenges for zooplankton model assessment." *Frontiers in Marine Science*4 (2017): 77.
9. Fabien Lombard, Emmanuel Boss, Anya M Waite, Meike Vogt, Julia Uitz, Lars Stemmann, Heidi M Sosik, Jan Schulz, Jean-Baptiste Romagnan, Marc Picheral, et al. "Globally consistent quantitative observations of planktonic ecosystems." *Frontiers in Marine Science*6 (2019): 196.
10. Tyler Rohr, Anthony J Richardson, Andrew Lenton, Matthew A Chamberlain,, and Elizabeth H Shadwick. "Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models." *Communications Earth & Environment* 4, no. 1 (2023): 212.
11. Nielja S Knecht, Fabio Benedetti, Urs Hofmann Elizondo, Nina Bednaršek, Sonia Chaabane, Catharina de Weerd, Katja TCA Peijnenburg, Ralf Schiebel,, and Meike Vogt. "The impact of zooplankton calcifiers on the marine carbon cycle." *Global Biogeochemical Cycles* 37, no. 6 (2023): e2022GB007685.
12. Corentin Clerc, Laurent Bopp, Fabio Benedetti, Nielja Knecht, Meike Vogt,, and Olivier Aumont. "Effects of mesozooplankton growth and reproduction on plankton and organic carbon dynamics in a marine biogeochemical model." *Global Biogeochemical Cycles* 38, no. 9 (2024): e2024GB008153.
13. Santiago Hernández-León, Rolf Koppelmann, Eugenio Fraile-Nuez, Antonio Bode, Carmen Mompeán, Xabier Irigoien, M Pilar Olivar, Fidel Echevarría, María Luz Fernández de Puelles, J Ignacio González-Gordillo, et al. "Large deep-sea zooplankton biomass mirrors primary production in the global ocean." *Nature communications* 11, no. 1 (2020): 6048.
14. Yulia Egorova, Gabriel Reygondeau, William WL Cheung,, and Evgeny A Pakhomov. "Global estimate of mesopelagic mesozooplankton biomass." *Scientific Reports* 15, no. 1 (2025): 22100.
15. Laetitia Drago, Thelma Panaïotis, Jean-Olivier Irisson, Marcel Babin, Tristan Biard, François Carlotti, Laurent Coppola, Lionel Guidi, Helena Hauss, Lee Karp-Boss, et al. "Global distribution of zooplankton biomass estimated by in situ imaging and machine learning." *Frontiers in Marine Science*9.
16. Fabien Lombard, Lionel Guidi, Manoela C Brandão, Pedro Coelho Luis, Sébastien Colin, John Richard Dolan, Amanda Elineau, Nicolas Henry, Laetitia Jalabert, Michel Loreau, et al. "Ubiquity of inverted'gelatinous' ecosystem pyramids in the global ocean." . *bioRxiv*.
17. Corinne Le Quéré, Sandy P Harrison, I Colin Prentice, Erik T Buitenhuis, Olivier Aumont, Laurent Bopp, Hervé Claustre, Leticia Cotrim Da Cunha, Richard Geider, Xavier Giraud, et al. "Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models." *Global Change Biology* 11, no. 11 (2005): 2016–2040.
18. Jessica Y Luo, Charles A Stock, Natasha Henschke, John P Dunne,, and Todd D O'Brien. "Global ecological and biogeochemical impacts of pelagic tunicates." . *Progress in Oceanography* (2022): 102822.
19. Corentin Clerc, Laurent Bopp, Fabio Benedetti, Meike Vogt,, and Olivier Aumont. "Including filter-feeding gelatinous macrozooplankton in a global marine biogeochemical model: model–data comparison and impact on the ocean carbon cycle." *Biogeosciences* 20, no. 4 (2023): 869–895.
20. Mark F Baumgartner, and Ann M Tarrant. "The physiology and ecology of diapause in marine copepods." *Annual review of marine science*9 (2017): 387–411.
21. Rebecca M Wright, Corinne Le Quéré, Erik Buitenhuis, Sophie Pitois,, and Mark J Gibbons. "Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model." *Biogeosciences* 18, no. 4 (2021): 1291–1320.
22. Fabio Benedetti, Jonas Wydler,, and Meike Vogt. "Copepod functional traits and groups show divergent biogeographies in the global ocean." *Journal of Biogeography* 50, no. 1 (2023): 8–22.
23. JP Grime. "Benefits of plant diversity to ecosystems: immediate, filter and founder effects." *Journal of Ecology* 86, no. 6 (1998): 902–910.
24. Fabio Benedetti, Jonas Wydler, Corentin Clerc, Nielja Knecht,, and Meike Vogt. "Emergent relationships between the functional diversity of marine planktonic copepods and ecosystem functioning in the global ocean." *Global Change Biology* 31, no. 3 (2025): e70094.
25. Angelos Amyntas, Emilio Berti, Benoit Gauzens, Georg Albert, Wentao Yu, Alexandra S Werner, Nico Eisenhauer,, and Ulrich Brose. "Niche complementarity among plants and animals can alter the biodiversity–ecosystem functioning relationship." *Functional Ecology* 37, no. 10 (2023): 2652–2665.
26. Alex L Pigot, Laura E Dee, Anthony J Richardson, Declan LM Cooper, Nico Eisenhauer, Richard D Gregory, Simon L Lewis, Callum J Macgregor, Dario Massimino, Daniel S Maynard, et al. "Macroecological rules predict how biomass scales with species richness in nature." *Science* 387, no. 6740 (2025): 1272–1276.
27. Luke Gregor, and Nicolas Gruber. "OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification." *Earth System Science Data* 13, no. 2 (2021): 777–808.
28. Conor Waldock, Rick D Stuart-Smith, Camille Albouy, William WL Cheung, Graham J Edgar, David Mouillot, Jerry Tjiputra,, and Loïc Pellissier. "A quantitative review of abundance-based species distribution models." *Ecography* 2022, no. 1.
29. Meike Vogt, Fabio Benedetti, Hugo Sarmento, Paula Huber, Clara Arboleda-Baena, Ruby Rose Bader, Dominic Eriksson, Nielja Knecht, Noémy Chénier, Olivier Jaillon, et al. "AtlantECO deliverable 2.1: AtlantECO-BASE1." .
30. Alexandre Schickele, Corentin Clerc, Fabio Benedetti, Daniele De Angelis, Urs Hofmann Elizondo, Matthias Münnich, Jean-Olivier Irisson,, and Meike Vogt. "CEPHALOPOD, a package to standardize marine habitat-modelling practices and

enhance inter-comparability across biological observations." *Methods in Ecology and Evolution* 16, no. 6 (2025): 1126–1135.
31. Fabio Benedetti, Nicolas Gruber,, and Meike Vogt. "Global gradients in species richness of marine plankton functional groups." *Journal of Plankton Research* 45, no. 6 (2023): 832–852.
32. Thomas Kiørboe. "How zooplankton feed: Mechanisms, traits and trade-offs." *Biological reviews* 86, no. 2 (2011): 311–339.
33. Roisin Moriarty, and TD O'Brien. "Distribution of mesozooplankton biomass in the global ocean." *Earth System Science Data* 5, no. 1 (2013): 45–55.
34. Jane Elith, Michael Kearney,, and Steven Phillips. "The art of modelling range-shifting species." *Methods in ecology and evolution* 1, no. 4 (2010): 330–342.
35. Fabio Benedetti, Meike Vogt, Urs Hofmann Elizondo, Damiano Righetti, Niklaus E Zimmermann,, and Nicolas Gruber. "Major restructuring of marine plankton assemblages under global warming." *Nature communications* 12, no. 1 (2021): 5226.
36. Cathy H Lucas, Daniel OB Jones, Catherine J Hollyhead, Robert H Condon, Carlos M Duarte, William M Graham, Kelly L Robinson, Kylie A Pitt, Mark Schildhauer,, and Jim Regetz. "Gelatinous zooplankton biomass in the global oceans: geographic variation and environmental drivers." *Global Ecology and Biogeography* 23, no. 7 (2014): 701–714.
37. Alban Planchat, Laurent Bopp, Lester Kwiatkowski,, and Olivier Torres. "The carbonate pump feedback on alkalinity and the carbon cycle in the 21st century and beyond." *Earth System Dynamics* 15, no. 3 (2024): 565–588.
38. Marina Lévy, and Adrian P Martin. "The influence of mesoscale and submesoscale heterogeneity on ocean biogeochemical reactions." *Global Biogeochemical Cycles* 27, no. 4 (2013): 1139–1150.
39. O. Aumont, C. Ethé, A. Tagliabue, L. Bopp,, and M. Gehlen. "PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies." *Geoscientific Model Development* 8, no. 8 (2015): 2465–2513.
40. Kanchana Bandara, Øystein Varpe, Lishani Wijewardene, Vigdis Tverberg,, and Ketil Eiane. "Two hundred years of zooplankton vertical migration research." *Biological Reviews* 96, no. 4 (2021): 1547–1589.
41. Guang Yang, Angus Atkinson, Evgeny A Pakhomov, Katrin Schmidt, Weilei Wang, Jennifer J Freer,, and Geraint A Tarling. "Seasonally migrating zooplankton strongly enhance Southern Ocean carbon sequestration." *Limnology and Oceanography* 70, no. 8 (2025): 2208–2222.
42. Sigrún Huld Jónasdóttir, André W Visser, Katherine Richardson,, and Michael R Heath. "Seasonal copepod lipid pump promotes carbon sequestration in the deep North Atlantic." *Proceedings of the National Academy of Sciences* 112, no. 39 (2015): 12122–12126.
43. Kailin Liu, Zhimeng Xu, Xin Liu, Bangqin Huang, Hongbin Liu,, and Bingzhang Chen. "Modelling global mesozooplankton biomass using machine learning." *Progress in Oceanography*229 (2024): 103371.
44. Edward Brinton 1967. *Distributional atlas of Euphausiacea (Crustacea) in the California current region*, State of California Marine Research Committee.
45. Rachel L Kaplan, Kim S Bernard, Solène Derville, Jennifer L Fisher, Elizabeth M Phillips, Elizabeth A Daly, Joseph D Warren,, and LG Torres. "Krill swarm biomass, energetic density, and species composition drive humpback whale distribution in the Northern California Current." *ICES Journal of Marine Science* 82, no. 1 (2025): fsaf005.
46. Macarena Díaz-Astudillo, Ramiro Riquelme-Bugueño, Kim S Bernard, Gonzalo S Saldías, Reinaldo Rivera,, and Jaime Letelier. "Disentangling species-specific krill responses to local oceanography and predator's biomass: The case of the Humboldt krill and the Peruvian anchovy." *Frontiers in Marine Science*9 (2022): 979984.
47. Volker Siegel, and Jonathan L Watkins. 2016. "Distribution, biomass and demography of Antarctic krill, Euphausia superba." In *Biology and ecology of Antarctic krill*, 21–100. Springer.
48. Nadine M Johnston, Eugene J Murphy, Angus Atkinson, Andrew J Constable, Cédric Cotté, Martin Cox, Kendra L Daly, Ryan Driscoll, Hauke Flores, Svenja Halfter, et al. "Status, change, and futures of zooplankton in the Southern Ocean." *Frontiers in Ecology and Evolution*9 (2022): 624692.
49. Tammy Horton, Claude De Broyer, Denise Bellan-Santini, Charles Oliver Coleman, Denis Copilaș-Ciocianu, Laure Corbari, Mikhail E Daneliya, Jean Claude Dauvin, Wim Decock, Lucia Fanini, et al. "The World Amphipoda Database: history and progress." .
50. Joanna Mierkiewicz, Emilia Trudnowska, Katarzyna Błachowiak-Samołyk,, and Luiza Bielecka. "Distinctive zonation of planktonic ostracods assemblages in the dynamic Southeastern Atlantic." *Journal of Sea Research*187 (2022): 102256.
51. Julie Lemoine, Sakina-dorothée Ayata, Cornelia Jaspers,, and Fabien Lombard. "Biomass-to-volume ratio as a central continuous functional trait for marine zooplankton." *Limnology and Oceanography* 70, no. 9 (2025): 2673–2687.
52. Cornelia Jaspers, Russell R Hopcroft, Thomas Kiørboe, Fabien Lombard, Ángel López-Urrutia, Jason D Everett,, and Anthony J Richardson. "Gelatinous larvacean zooplankton can enhance trophic transfer and carbon sequestration." *Trends in Ecology & Evolution* 38, no. 10 (2023): 980–993.
53. Ryan F Heneghan, Jason D Everett, Patrick Sykes, Sonia D Batten, Martin Edwards, Kunio Takahashi, Iain M Suthers, Julia L Blanchard,, and Anthony J Richardson. "A functional size-spectrum model of the global marine ecosystem that resolves zooplankton composition." *Ecological Modelling*435 (2020): 109265.
54. Natasha Henschke, Jason D Everett, Anthony J Richardson,, and Iain M Suthers. "Rethinking the role of salps in the ocean." *Trends in ecology & evolution* 31, no. 9 (2016): 720–733.
55. Laura E Lilly, Iain M Suthers, Jason D Everett,, and Anthony J Richardson. "A global review of pyrosomes: Shedding light on the ocean's elusive gelatinous "fire-bodies"." *Limnology and Oceanography Letters* 8, no. 6 (2023): 812–829.
56. Robert H Condon, William M Graham, Carlos M Duarte, Kylie A Pitt, Cathy H Lucas, Steven HD Haddock, Kelly R Sutherland, Kelly L Robinson, Michael N Dawson, Mary Beth Decker, et al. "Questioning the rise of gelatinous zooplankton in the world's oceans." *Bioscience* 62, no. 2 (2012): 160–169.

57. Corentin Clerc, Olivier Aumont,, and Laurent Bopp. "Filter-feeding gelatinous macrozooplankton response to climate change and implications for benthic food supply and global carbon cycle." . *Global Change Biology*.
58. Ryan F Heneghan, Jason D Everett, Julia L Blanchard, Patrick Sykes,, and Anthony J Richardson. "Climate-driven zooplankton shifts cause large-scale declines in food quality for fish." *Nature Climate Change* 13, no. 5 (2023): 470–477.
59. Lester Kwiatkowski, Alban Planchat, Marc Pyolle, Olivier Torres, Nathaelle Bouttes, Adrien Comte,, and Laurent Bopp. "Declining coral calcification to enhance twenty-first-century ocean carbon uptake by gigatonnes." *Proceedings of the National Academy of Sciences* 122, no. 23 (2025): e2501562122.
60. KH Patrik Strömberg, Timothy J Smyth, J Icarus Allen, Sophie Pitois,, and Todd D O'Brien. "Estimation of global zooplankton biomass from satellite ocean colour." *Journal of Marine Systems* 78, no. 1 (2009): 18–27.
61. Chhaya Chaudhary, Anthony J Richardson, David S Schoeman,, and Mark J Costello. "Global warming is causing a more pronounced dip in marine species richness around the equator." *Proceedings of the National Academy of Sciences* 118, no. 15 (2021): e2015094118.
62. Peter H Wiebe. "Functional regression equations for zooplankton displacement volume, wet weight, dry weight, and carbon: a correction." *Fish. Bull.*86 (1988): 833–835.

Supporting Information

All model outputs and R scripts used for data preparation and analysis have been deposited in a Zenodo repository (DOI: 10.5281/zenodo.20560560). The repository is currently restricted during peer review and will be made publicly available upon publication.

Supporting information OPEN ACCESS

Supporting information for : A data-driven assessment of the global community structure of pelagic zooplankton biomass

Corentin Clerc[1,2,1] | Alexandre Schickele[2] | Fabio Benedetti[2,4] | Urs Hofmann Elizondo[2] | Servane Lunven[2] | Ruby Bader[2] | Sébastien Vivier[2] | Anthony J. Richardson[3] | Meike Vogt[2]

[1]Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques, LOCEAN/IPSL, F-75005 Paris, France | [2]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092, Zürich, Switzerland. | [3]School of the Environment, The University of Queensland, St Lucia, 4072, Queensland, Australia | [4]Institute of Plant Sciences, University of Bern, Bern, Switzerland.

Correspondence: Corentin Clerc (corentin.clerc@locean.ipsl.fr)

Received: Revised: Accepted:

15 # FIGURES

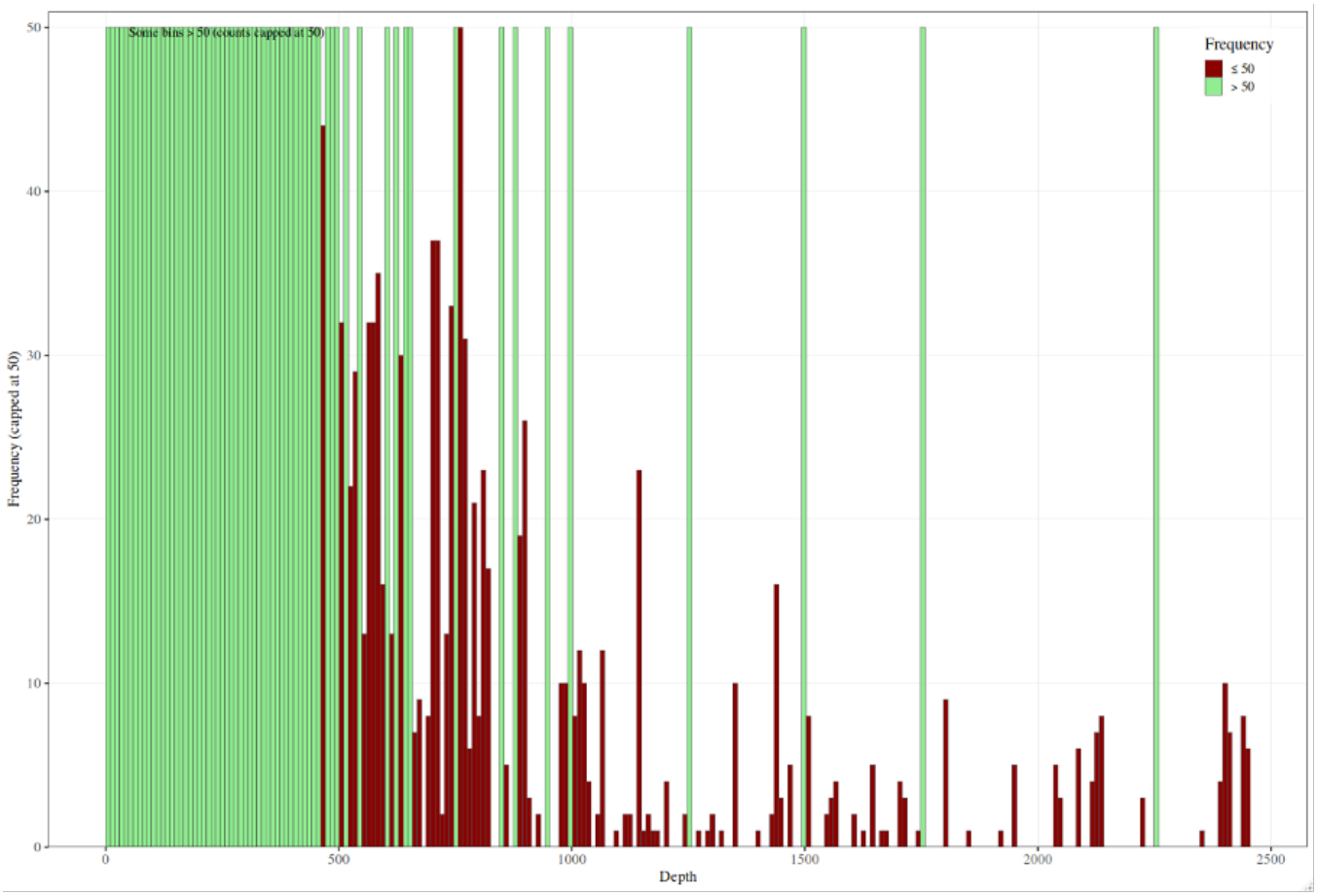

FIGURE S1 | Histogram of the number of observations across 10m depth bins. Red color indicates bins in which less than 50 observations are available.

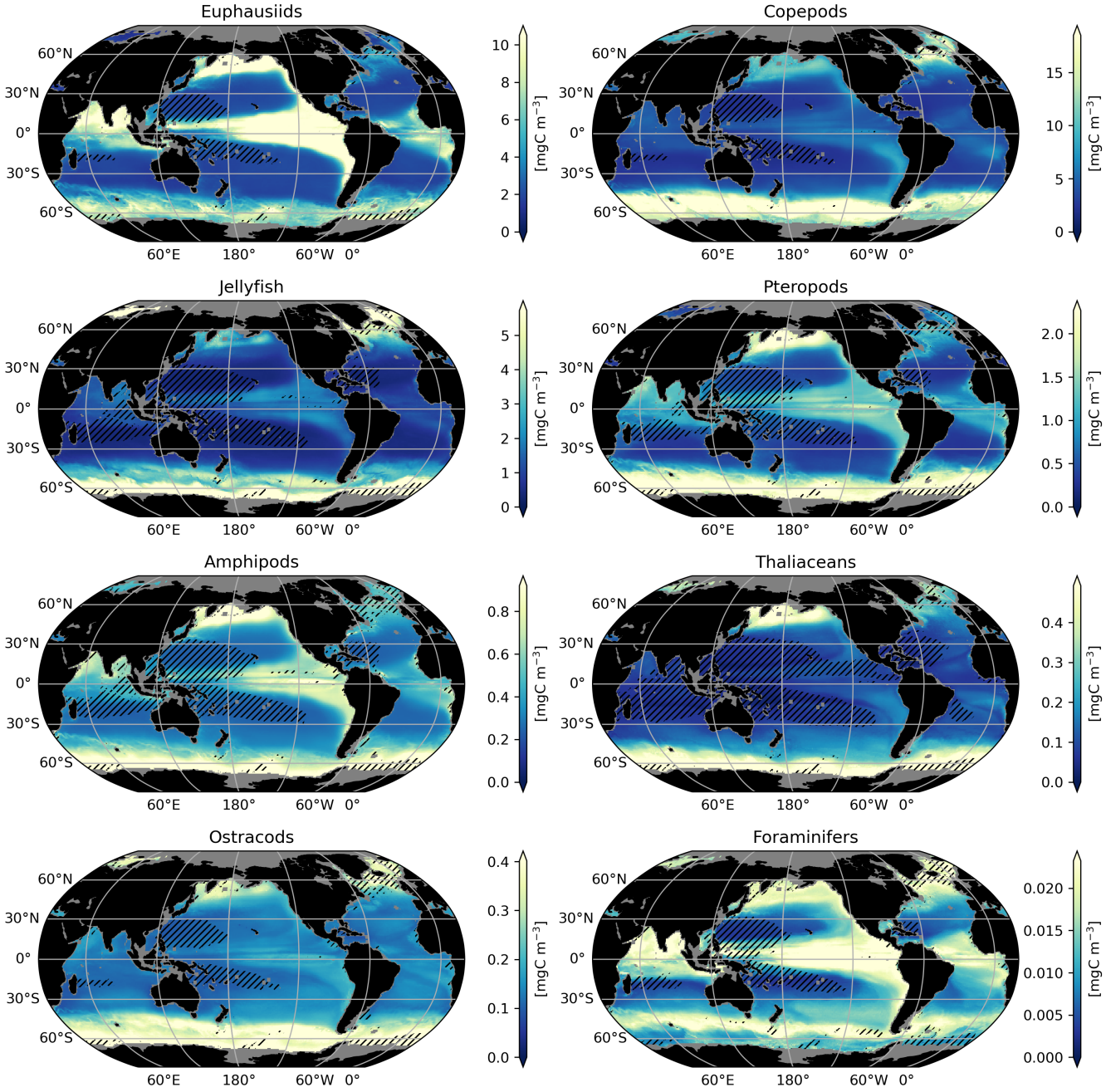

FIGURE S2 | PFT-specific mesopelagic (200-500 m) biomass distribution simulated by the biomass distribution models. Hatching indicates areas where model predictions fall outside the training conditions (MESS < 0) for at least one PFT in at least three months.

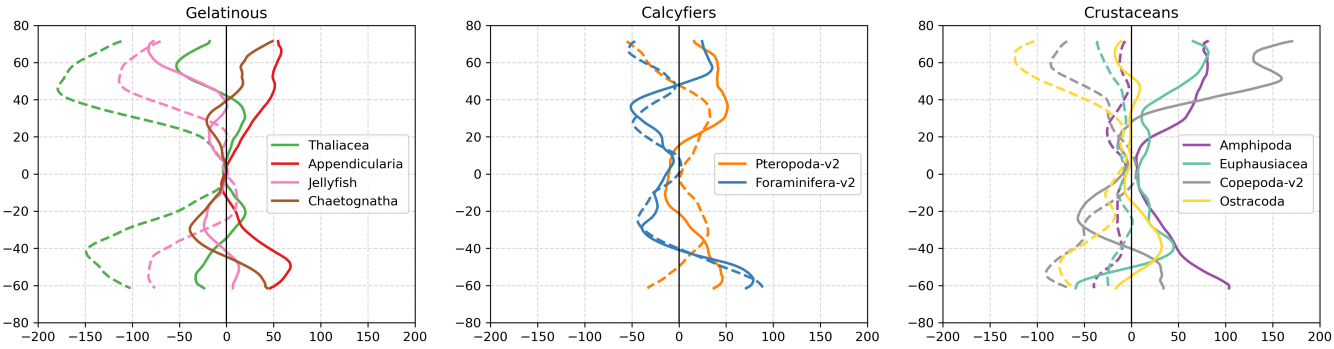

FIGURE S3 | Change in biomass from summer to winter (in %). Dotted lines indicates mesopelagic zonal profiles. Plain lines indicates epipelagic zonal profiles. Profiles are smoothed by applying a 5° running mean across latitudes.

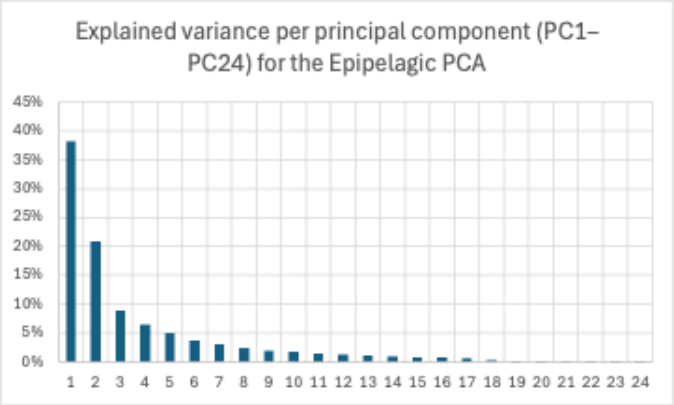

FIGURE S4 | Epipelagic PCA : Explained variance per principal component (PC1–PC24).

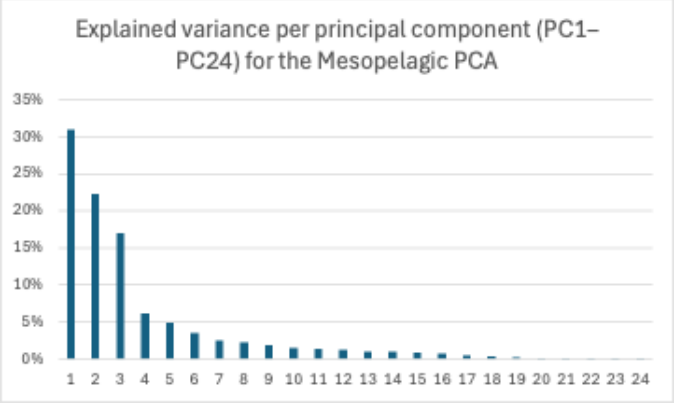

FIGURE S5 | Mesopelagic PCA : Explained variance per principal component (PC1–PC24).

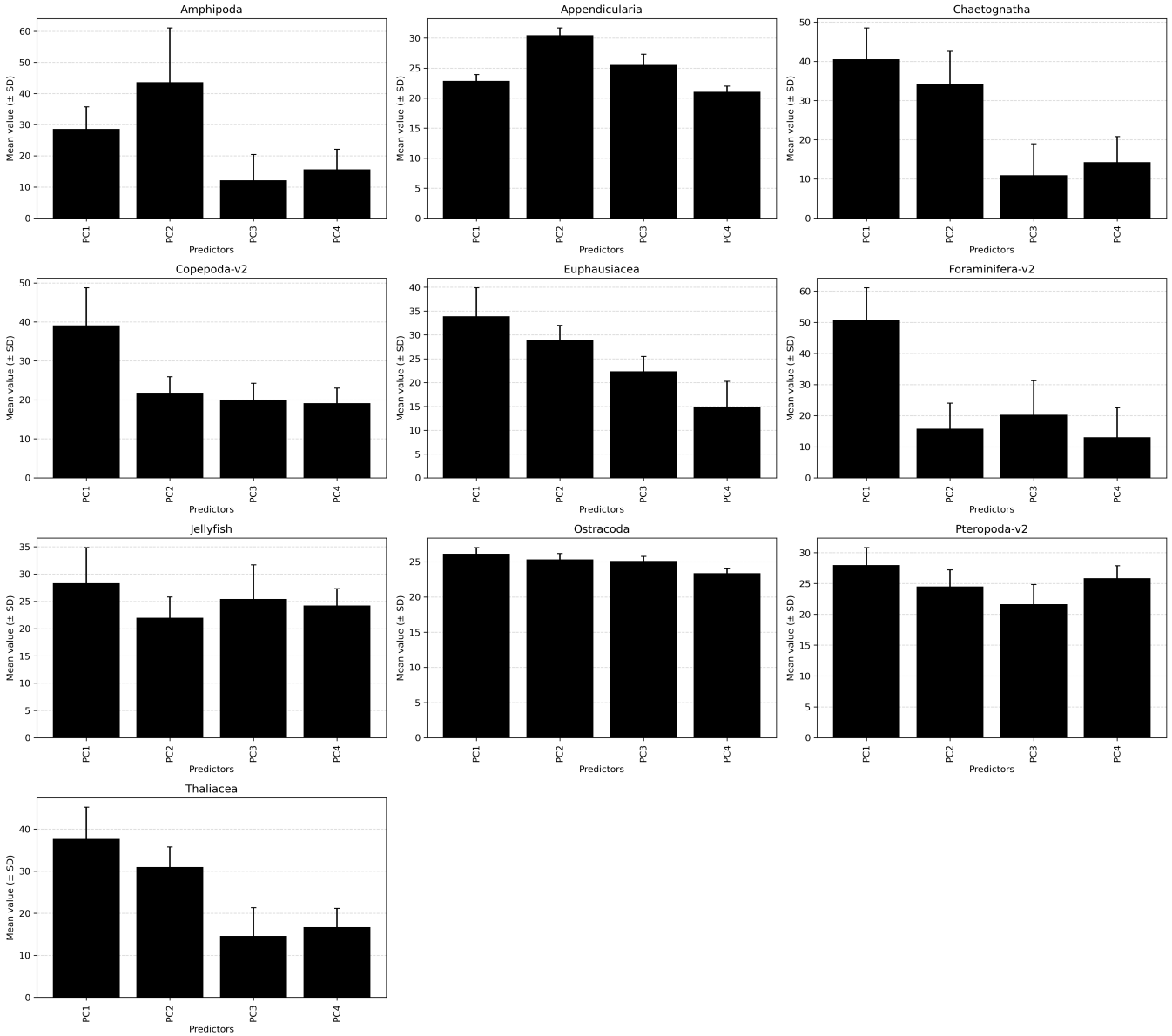

FIGURE S6 | Environmental predictors selected by the algorithms for the epipelagic distribution.

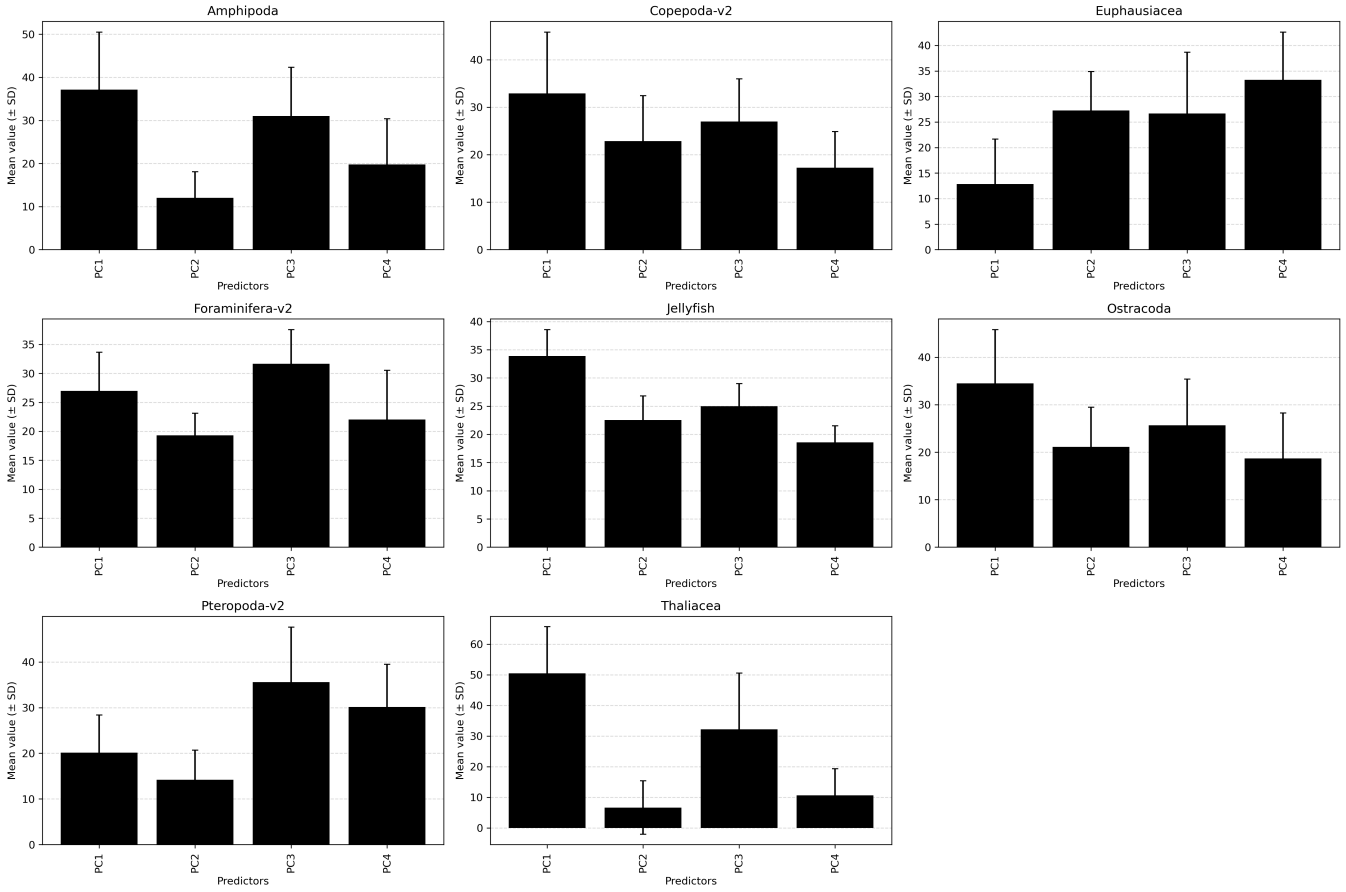

FIGURE S7 | Environmental predictors selected by the algorithms for the mesopelagic distribution.

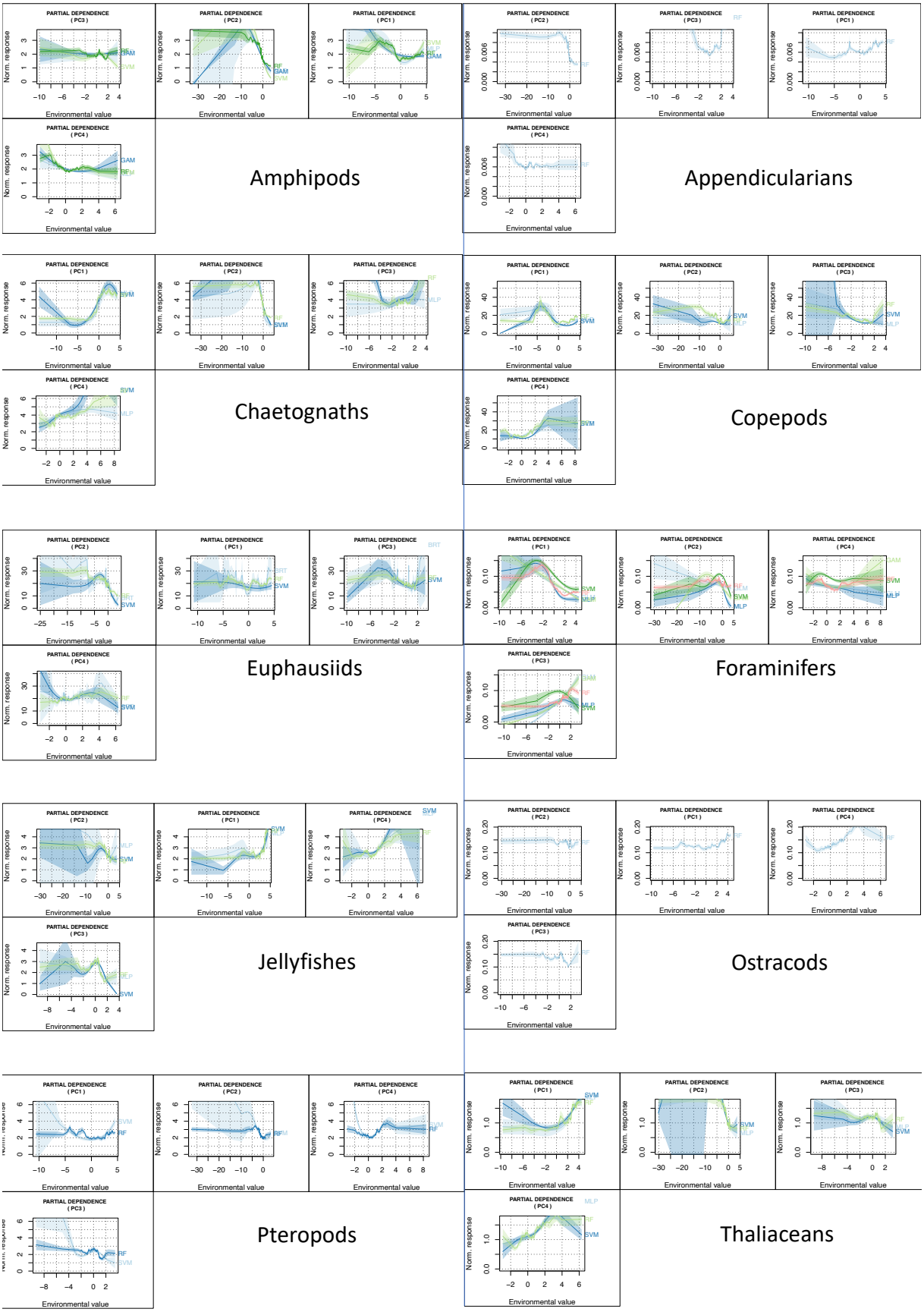

FIGURE S8 | Partial dependence plots (PDP) for the environmental predictors (PC1, PC2, PC3 and PC4) in the epipelagic zooplankton PFT models. The curves indicate the relations learned by the different algorithms.

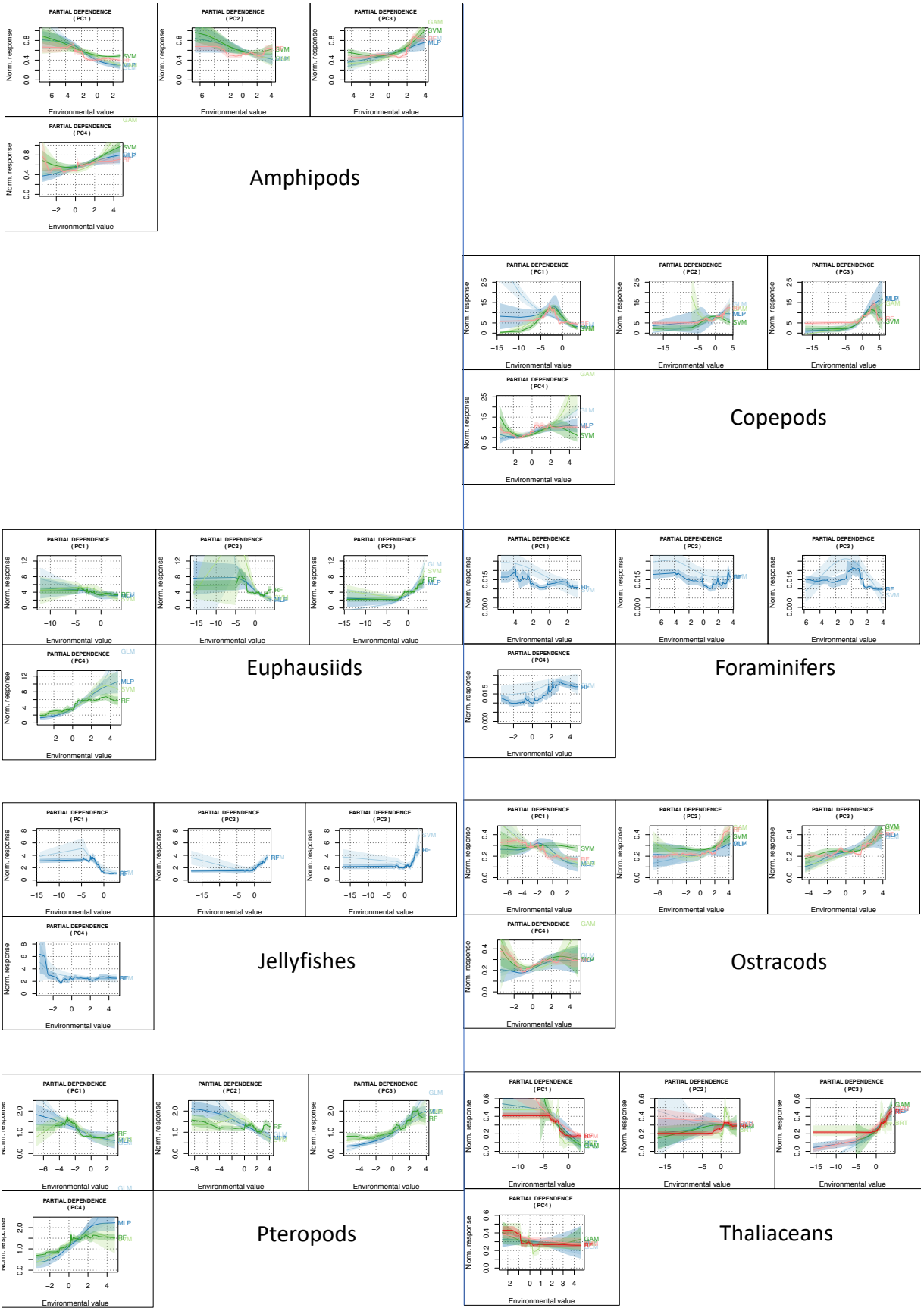

FIGURE S9 | Partial dependence plots (PDP) for the environmental predictors (PC1, PC2, PC3 and PC4) in the mesopelagic zooplankton PFT models. The curves indicate the relations learned by the different algorithms.

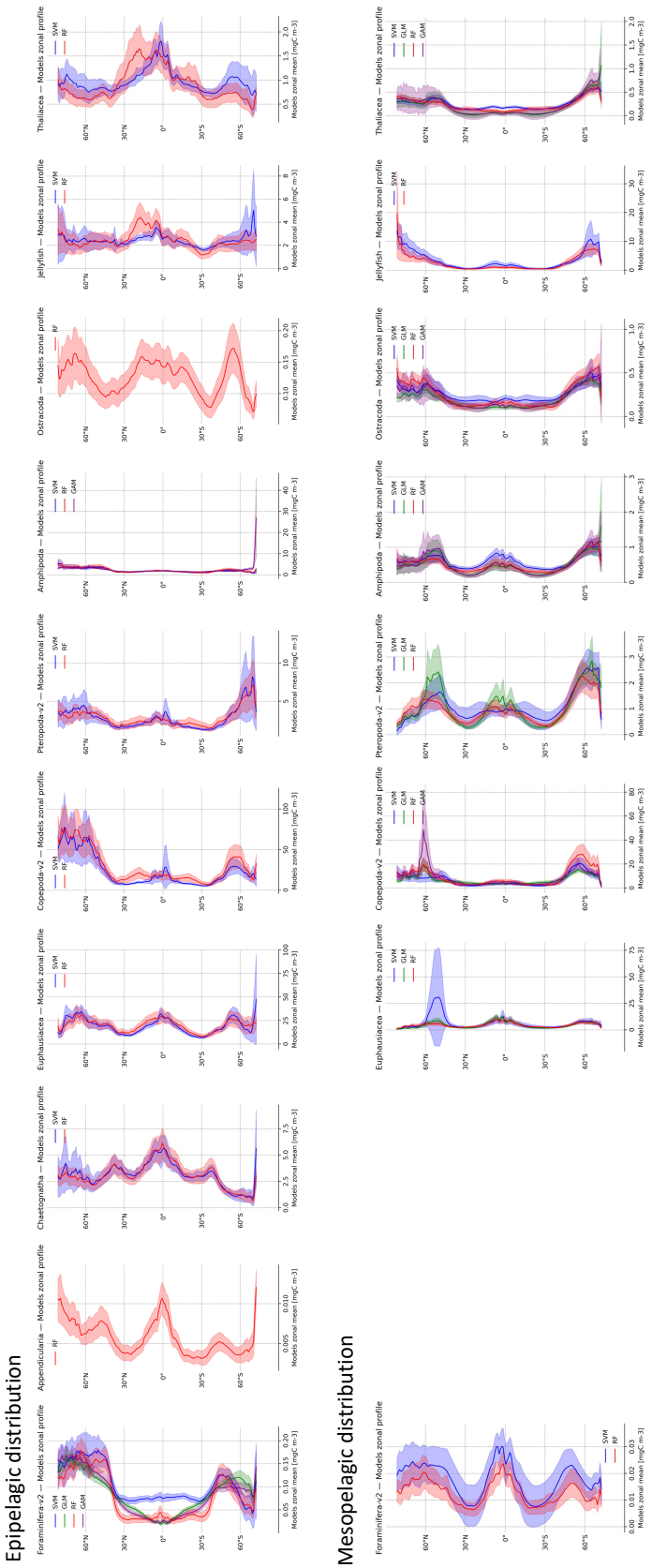

FIGURE S10 | PFT-specific zonal profiles of the different CEPHALOPOD algorithms that passed the quality checks. SVM = Single Vector Machine, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model.

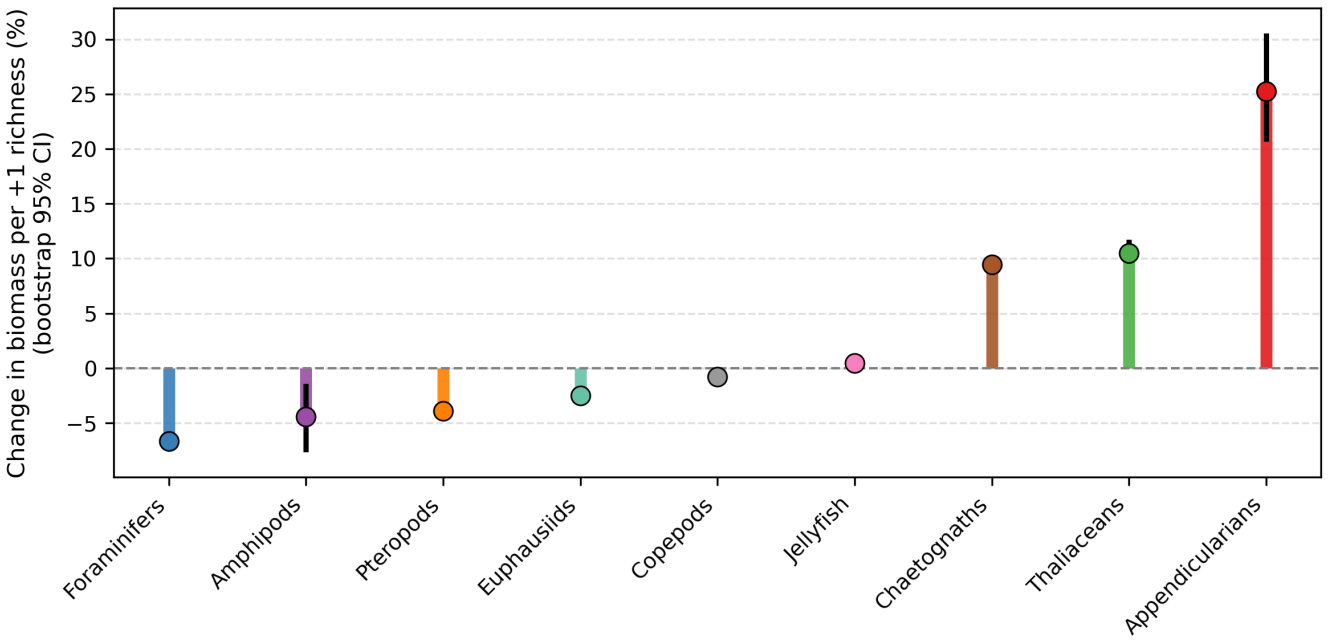

FIGURE S11 | Ranked PFT-specific percent change in biomass per +1 richness (derived from the GLM richness×PFT interaction), with bootstrap 95% confidence intervals computed after coarsening fields to 10°×10° and using 100 bootstrap replicates. Axes are log-scaled in the left panel. Uncertainty was quantified using a non-parametric spatial block bootstrap: richness and biomass fields were coarsened to 10° × 10° (block means), coarsened grid cells were resampled with replacement, and the same Gamma GLM was refitted for each replicate; 95% confidence intervals were obtained from the 2.5th–97.5th percentiles.

TABLES

TABLE S1 | References used to derive abundance-to-carbon biomass conversion factors for each pelagic functional type (PFT). Conversion factors for thaliaceans were taken from [1] and references therein, and those for foraminifera and pteropods from [2] and references therein. Conversion factors for the remaining eight PFTs are reported in this study and are based on a compilation originally assembled for AtlantECO-MAPS [3]. The underlying references used for these eight PFTs were [4], COPEPEDIA (https://www.st.nmfs.noaa.gov/copepedia/taxa/T4000000/html/biometricframe.html), [5], and [6], with [7] additionally used for euphausiids.

PFT	Conversion factor reported in	Underlying reference(s)	Notes
Thaliaceans	[1]	[1] and references therein	Published synthesis
Foraminifera	[2]	[2] and references therein	Published synthesis
Pteropods	[2]	[2] and references therein	Published synthesis
Copepods	This study	[8]; COPEPEDIA; [5]; [6]	
Euphausiids	This study	[7]; [5]; [6]	—
Amphipods	This study	[5]; [6]	—
Ostracods	This study	[5]; [6]	—
Chaetognaths	This study	[5]; [6]	—
Cnidarians	This study	[5]; [6]	—
Ctenophores	This study	[5]; [6]	—
Appendicularians	This study	[5]; [6]	—

TABLE S2 ¦ Collection of observation-based, monthly-resolved environmental features at the global scale derived from WOA18 [9], Ocean Soda [10], GMIS, CMEMS [11] and AVISO. A full description of these files is available at https://data.up.ethz.ch/doku.php?id=biocean:readme.

File Name	Unit	Short Name	Resolution	Source	Sensor (if relevant)	Ref	obs_time_extent
climatology_A_0_50.nc	umol kg^{-1}	Apparent Oxygen Utilization (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_A_200_300.nc	umol kg^{-1}	Apparent Oxygen Utilization (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_dic_SODA.nc	umol kg^{-1}	Dissolved inorganic carbon	360x180x12	OceanSodaETHZ2023		[10]	1982–2022
climatology_A_CHLA_regridded.nc	mg Chl m^{-3}	Chlorophyll-a	360x180x12	GMIS	MODIS-AQUA	a	2002–2017
climatology_fsle_aviso_2001_2020.nc	d^{-1}	FSLE based on maximum eigenvalue	360x180x12	AVISO	–	b	2000–2021
climatology_i_0_50.nc	umol kg^{-1}	Silicate concentration (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_i_200_300.nc	umol kg^{-1}	Silicate concentration (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_M_0_0.nc	m	Ocean mixed layer thickness (0 m)	360x180x12	WOA18	–	[9]	1981–2010
climatology_n_0_50.nc	umol kg^{-1}	Nitrate concentration (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_n_200_300.nc	umol kg^{-1}	Nitrate concentration (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_o_0_50.nc	umol kg^{-1}	Dissolved oxygen concentration (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_O_0_50.nc	%	Oxygen saturation (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_o_200_300.nc	umol kg^{-1}	Dissolved oxygen concentration (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_O_200_300.nc	%	Oxygen saturation (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_p_0_50.nc	umol kg^{-1}	Phosphate concentration (0–50 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_p_200_300.nc	umol kg^{-1}	Phosphate concentration (200–300 m)	360x180x12	WOA18	–	[9]	1900–2017
climatology_P_SST_regridded.nc	°C	Temperature	360x180x12	GMIS	PATHFINDER	a	1981–2009
climatology_ph_total_SODA.nc	-log($[H^+]$)	pH total	360x180x12	OceanSodaETHZ2023		[10]	1982–2022
climatology_S_PP_regridded.nc	mg C m^{-2} d^{-1}	Primary production	360x180x12	GMIS	SEAWIFS	a	1997–2010
climatology_t_0_50.nc	°C	Sea water temperature (0–50 m)	360x180x12	WOA18	–	[9]	1955–2017
climatology_t_200_300.nc	°C	Sea water temperature (200–300 m)	360x180x12	WOA18	–	[9]	1955–2017
climatology_CHL_1998_2020_CMEMS.nc	mg Chl m^{-3}	Chlorophyll-a	360x180x12	CMEMS	MODISA, VIIRSN, OLCIa, VIIRSJ1, OLCIb	[12, 13, 11]	1998–2010
climatology_TOT_POC_CMEMS.nc	mg C m^{-3}	Total POC climatology	360x180x12	CMEMS	MOD, VIR	[14]	1998–2010
climatology_total_currents.nc	m s^{-1}	Geostrophic currents velocity	360x180x12	Giacomo Poli	–	b	1993–2019
climatology_S_CHLA_regridded.nc	mg Chl m^{-3}	Chlorophyll-a	360x180x12	GMIS	SEAWIFS	a	2002–2017
climatology_V_CHLA_regridded.nc	mg Chl m^{-3}	Chlorophyll-a	360x180x12	GMIS	VIIRS	a	2002–2017
climatology_T_SST_regridded.nc	°C	Temperature	360x180x12	GMIS	MODIS-TERRA	a	1981–2009
climatology_mld_soda.nc	m	Mixed layer depth	360x180x12	Giacomo Poli	–	[15]	–

[a] GMIS product information: https://data.jrc.ec.europa.eu/collection/gmis#description; see also https://joint-research-centre.ec.europa.eu/scientific-tools-and-databases/global-marine-information-system-gmis_en.
[b] AVISO product information: https://www.aviso.altimetry.fr.

TABLE S3 | PCA for modeling the epipelagic layer. Correlations between environmental predictors and retained principal components (PCs) Pearson correlation coefficients (r) between each environmental predictor (rows) and each retained principal component (columns). Cell colours indicate the sign and magnitude of r, ranging from blue (r=-1) to red (r=1). Predictor definitions and data sources are provided in Table S2.

Epipelagic PCA	PC1	PC2	PC3	PC4
climatology_t_0_50	0.95	-0.17	-0.12	0.07
climatology_P_SST_regridded	0.95	-0.19	-0.13	0.07
climatology_T_SST_regridded	0.95	-0.19	-0.14	0.05
climatology_O_0_50	0.39	-0.04	0.83	-0.15
climatology_total_currents	0.17	-0.12	-0.20	0.54
climatology_sla	0.15	-0.01	0.02	-0.29
climatology_fsle_aviso_2000_2020	0.01	0.07	-0.38	-0.61
climatology_S_PP_regridded	-0.23	-0.78	-0.09	0.10
climatology_TOT_POC_CMEMS	-0.38	-0.76	0.02	0.10
climatology_V_CHLA_regridded	-0.40	-0.72	0.01	0.01
climatology_M_CHLA_regridded	-0.42	-0.80	0.02	0.01
climatology_S_CHLA_regridded	-0.43	-0.78	-0.01	0.01
climatology_M_0_0	-0.43	0.32	0.18	0.61
climatology_A_CHLA_regridded	-0.44	-0.77	0.00	0.01
climatology_A_0_50	-0.46	0.06	-0.82	0.06
climatology_mld_soda	-0.48	0.43	0.00	0.49
climatology_ph_total_SODA	-0.50	-0.10	0.50	-0.16
climatology_CHL_1998_2020_CMEMS	-0.51	-0.80	0.02	0.01
climatology_i_0_50	-0.72	0.21	-0.33	-0.27
climatology_dic_SODA	-0.82	0.32	0.15	-0.02
climatology_n_0_50	-0.87	0.29	-0.14	0.00
climatology_p_0_50	-0.87	0.25	-0.18	-0.01
climatology_o_0_50	-0.91	0.16	0.24	-0.14

TABLE S4 | PCA for modeling the mesopelagic layer. Correlations between environmental predictors and retained principal components (PCs) Pearson correlation coefficients (r) between each environmental predictor (rows) and each retained principal component (columns). Cell colours indicate the sign and magnitude of r, ranging from blue (r=-1) to red (r=1). Predictor definitions and data sources are provided in Table S2.

Mesopelagic PCA	PC1	PC2	PC3	PC4
climatology_T_SST_regridded	0.89	-0.30	-0.28	0.05
climatology_P_SST_regridded	0.89	-0.30	-0.28	0.06
climatology_t_200_300	0.84	0.01	-0.42	0.05
climatology_sla	0.14	-0.06	-0.02	-0.28
climatology_total_currents	0.13	-0.27	0.00	0.65
climatology_fsle_aviso_2000_2020	0.07	-0.22	0.33	-0.56
climatology_A_200_300	0.07	-0.85	0.48	0.02
climatology_O_200_300	-0.20	0.86	-0.41	-0.06
climatology_S_PP_regridded	-0.38	-0.54	-0.49	0.11
climatology_M_0_0	-0.39	0.46	0.13	0.55
climatology_mld_soda	-0.39	0.43	0.32	0.45
climatology_p_200_300	-0.43	-0.59	0.63	0.08
climatology_o_200_300	-0.47	0.80	-0.25	-0.05
climatology_n_200_300	-0.50	-0.53	0.62	0.09
climatology_ph_total_SODA	-0.54	0.41	-0.30	-0.26
climatology_TOT_POC_CMEMS	-0.54	-0.41	-0.50	0.09
climatology_V_CHLA_regridded	-0.55	-0.36	-0.46	-0.02
climatology_i_200_300	-0.56	-0.32	0.60	-0.18
climatology_S_CHLA_regridded	-0.60	-0.41	-0.51	-0.01
climatology_A_CHLA_regridded	-0.60	-0.39	-0.50	-0.02
climatology_M_CHLA_regridded	-0.60	-0.44	-0.50	-0.01
climatology_CHL_1998_2020_CMEMS	-0.70	-0.44	-0.47	-0.01
climatology_dic_SODA	-0.77	0.40	0.26	-0.06

TABLE S5 | Quality check values for the different algorithms and different zooplankton PFT modeled, in the mesopelagic (up) and epipelagic (down) layers. Only r^2 and NSD are presented as pre-VIP does not differ across algorithms and is provided in Table 2 of the manuscript. Green dots indicate algorithms that passed the QC, orange dot algorithms that had QC with values up to 30 % lower than the passing threshold, and red dots none passing algorithms.

	MESOPELAGIC MODELS						
	PFT	**GLM**	**MLP**	**BRT**	**GAM**	**SVM**	**RF**
NSD	*Amphipods*	0.24	0.34	0.99	0.31	0.16	0.18
	Appendicularians	0.41	0.61	1.72	0.51	0.30	0.40
	Chaetognaths	0.24	0.71	0.44	0.24	0.19	0.22
	Copepods	0.17	0.67	1.19	0.31	0.32	0.29
	Euphausiids	0.21	0.45	0.97	3.16	0.56	0.28
	Foraminifers	0.27	0.72	0.88	0.42	0.40	0.32
	Jellyfishes	3.16	0.78	3.16	3.29	0.38	0.32
	Ostracods	0.18	0.65	0.90	0.31	0.23	0.23
	Pteropods	0.31	0.33	1.69	1.93	0.34	0.26
	Thaliaceans	0.35	0.58	0.60	0.57	0.14	0.25
	PFT	**GLM**	**MLP**	**BRT**	**GAM**	**SVM**	**RF**
R^2	*Amphipods*	0.15	0.17	0.16	0.15	0.14	0.22
	Appendicularians	0.02	0.04	0.09	0.03	0.00	0.02
	Chaetognaths	0.00	0.02	0.02	0.00	0.01	0.03
	Copepods	0.14	0.22	0.25	0.15	0.26	0.36
	Euphausiids	0.22	0.22	0.17	0.22	0.20	0.23
	Foraminifers	0.07	0.15	0.10	0.09	0.18	0.17
	Jellyfishes	0.34	0.38	0.41	0.36	0.44	0.47
	Ostracods	0.19	0.22	0.29	0.18	0.26	0.34
	Pteropods	0.25	0.24	0.21	0.25	0.23	0.30
	Thaliaceans	0.24	0.27	0.30	0.27	0.29	0.29

	EPIPELAGIC MODELS						
	PFT	**GLM**	**MLP**	**BRT**	**GAM**	**SVM**	**RF**
NSD	*Amphipods*	4.81	0.28	1.70	0.10	0.17	0.20
	Appendicularians	0.07	0.41	0.80	0.10	0.06	0.23
	Chaetognaths	0.11	0.26	0.74	3.16	0.21	0.23
	Copepods	1.81	0.53	1.18	3.16	0.34	0.34
	Euphausiids	0.10	0.33	0.68	0.08	0.22	0.26
	Foraminifers	0.10	0.39	0.78	0.11	0.14	0.23
	Jellyfishes	0.07	0.49	3.24	0.10	0.65	0.30
	Ostracods	0.05	0.26	0.58	0.08	0.11	0.20
	Pteropods	3.16	1.61	3.18	1.02	0.34	0.29
	Thaliaceans	0.08	0.40	2.72	0.10	0.22	0.27
	PFT	**GLM**	**MLP**	**BRT**	**GAM**	**SVM**	**RF**
R^2	*Amphipods*	0.08	0.17	0.17	0.10	0.21	0.25
	Appendicularians	0.04	0.10	0.09	0.07	0.07	0.13
	Chaetognaths	0.10	0.18	0.23	0.12	0.21	0.29
	Copepods	0.02	0.12	0.17	0.04	0.15	0.29
	Euphausiids	0.02	0.10	0.12	0.03	0.11	0.19
	Foraminifers	0.21	0.37	0.34	0.26	0.39	0.46
	Jellyfishes	0.02	0.13	0.13	0.03	0.12	0.23
	Ostracods	0.01	0.06	0.06	0.02	0.07	0.14
	Pteropods	0.01	0.09	0.11	0.02	0.13	0.20
	Thaliaceans	0.08	0.14	0.12	0.09	0.14	0.18

SUPPLEMENTARY TEXT S1

To compile the AtlantECO-BASE dataset, six main data sources (i.e., previously published data compilations and more recent scientific cruises) were retrieved: NOAA's Coastal and Oceanic Plankton Ecology, Production, and Observation Database (COPEPOD; O'Brien, 2014); the Jellyfish Database Initiative (JeDI; Lucas et al., 2014); KRILLBASE (Atkinson et al., 2017), the MARine Ecosystem DATa initiative (MAREDAT; Buitenhuis et al., 2013), the British Oceanographic Data Centre (BODC), and Tara Oceans (Brandão et al., 2021). This compilation synthesized planetary-scale plankton concentration measurements collected through a broad variety of sampling devices over the last 100 years (1926-2017). Since the precision of the taxonomic identification was variable and not always up to date, we curated and harmonized the scientific names and the taxonomic classification of each plankton observation based on the classification and accepted list of species names of the World Register of Marine Species (WoRMS; WoRMS Editorial Board, 2023), using the "worms" R package version 0.2.2 (Holstein, 2018). All datasets provided concentrations in individuals per cubic meter (ind m^{-3}), except for KRILLBASE, which provided densities in ind m^2, which we converted to ind m^3 based on the maximum sampling depth of the corresponding net tows.

The AtlantECO database is freely available at the following link : link.

References

1. Corentin Clerc, Laurent Bopp, Fabio Benedetti, Meike Vogt,, and Olivier Aumont. "Including filter-feeding gelatinous macrozooplankton in a global marine biogeochemical model: model–data comparison and impact on the ocean carbon cycle." *Biogeosciences* 20, no. 4 (2023): 869–895.
2. Nielja S Knecht, Fabio Benedetti, Urs Hofmann Elizondo, Nina Bednaršek, Sonia Chaabane, Catharina de Weerd, Katja TCA Peijnenburg, Ralf Schiebel,, and Meike Vogt. "The impact of zooplankton calcifiers on the marine carbon cycle." *Global Biogeochemical Cycles* 37, no. 6 (2023): e2022GB007685.
3. Meike Vogt, Hugo Sarmento, Fabio Benedetti, Théophile Sanchez, Dominic Eriksson, Alexandre SCHICKELE, Corentin Clerc, Paula Huber, Roniel Freitas-Oliveira, Daniele De Angelis, Luigi Maiorano, Nielja Sofia Knecht, Giacomo Poli, Xinhang Li, Olivier Jaillon, Paul Frémont, Margaux Crédeville, Roy El Hourany, Marinna GAUDIN, Damien Eveillard, Samuel Chaffron, Marion Gehlen, Germain Bénard, Olivier Torres, Fabien Lombard, Lionel Guidi, Laetitia Drago, lars stemmann, Erik van Sebille, Tim Collart, Thomas Frölicher,, and Stéphane PESANT. 2026, April). "AtlantECO Deliverable - D02.5 - AtlantECO-MAPS (version 2)." , .
4. Philipp Brun, Thomas Kiørboe, Priscilla Licandro,, and Mark R Payne. "The predictive skill of species distribution models for plankton in a changing climate." *Global change biology* 22, no. 9 (2016): 3170–3181.
5. Roisin Moriarty, and TD O'Brien. "Distribution of mesozooplankton biomass in the global ocean." *Earth System Science Data* 5, no. 1 (2013): 45–55.
6. Thomas Kiørboe. "Zooplankton body composition." *Limnology and Oceanography* 58, no. 5 (2013): 1843–1850.
7. Patti Virtue, Bettina Meyer, Ulrich Freier, Peter D Nichols, Zhongnan Jia, Rob King, Jacob Virtue, Kerrie M Swadling, Klaus M Meiners,, and So Kawaguchi. "Condition of larval (furcilia VI) and one year old juvenile Euphausia superba during the winter–spring transition in East Antarctica." *Deep Sea Research Part II: Topical Studies in Oceanography*131 (2016): 182–188.
8. Philipp Brun, Mark R Payne,, and Thomas Kiørboe. "Trait biogeography of marine copepods–an analysis across scales." *Ecology letters* 19, no. 12 (2016): 1403–1413.
9. Hernan E Garcia, Timothy P Boyer, Olga K Baranova, Ricardo A Locarnini, Alexey V Mishonov, A ea Grodsky, Christopher R Paver, Katharine W Weathers, Igor V Smolyar, James R Reagan, et al. "World ocean atlas 2018: Product documentation." .
10. Luke Gregor, and Nicolas Gruber. "OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification." *Earth System Science Data* 13, no. 2 (2021): 777–808.
11. Hongyan Xi, Svetlana N Losa, Antoine Mangin, Philippe Garnesson, Marine Bretagnon, Julien Demaria, Mariana A Soppa, Odile Hembise Fanton d'Andon,, and Astrid Bracher. "Global chlorophyll a concentrations of phytoplankton functional types with detailed uncertainty assessment using multisensor ocean color and sea surface temperature satellite products." *Journal of Geophysical Research: Oceans* 126, no. 5 (2021): e2020JC017127.
12. F Gohin, JN Druon,, and L Lampert. "A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters." *International journal of remote sensing* 23, no. 8 (2002): 1639–1661.
13. Chuanmin Hu, Zhongping Lee,, and Bryan Franz. "Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference." *Journal of Geophysical Research: Oceans* 117, no. C1.
14. Dariusz Stramski, Rick A Reynolds, Marcel Babin, S Kaczmarek, Marlon R Lewis, R Röttgers, Antoine Sciandra, M Stramska, MS Twardowski, BA Franz, et al. "Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans." *Biogeosciences* 5, no. 1 (2008): 171–201.
15. James A Carton, Gennady A Chepurin,, and Ligang Chen. "SODA3: A new ocean climate reanalysis." *Journal of Climate* 31, no. 17 (2018): 6967–6983.