Progress in Oceanography 229 (2024) 103370

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Global characterization of modelled micronekton in biophysically defined provinces

S. Albernhe [a,b,*], T. Gorgues [b], P. Lehodey [c], C. Menkes [d], O. Titaud [a], S. Magon De La Giclais [a], A. Conchon [a]

[a] Collecte Localisation Satellites, 8-10 rue Hermès, Ramonville Sant Agne 31520, France
[b] Univ Brest, CNRS, Ifremer, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, F29280, Plouzané, France
[c] Mercator Ocean International, 2 Av. de l'Aérodrome de Montaudran, 31400, Toulouse, France
[d] ENTROPIE, IRD, Univ. de La Réunion, CNRS, Ifremer, Univ. de la Nouvelle-Calédonie, BP A5, 98848 Nouméa, New Caledonia

ARTICLE INFO

Keywords:
Micronekton
Mesopelagic
Modelling
Biomes
Acoustics
Clustering
Global scale

ABSTRACT

Micronekton are the mid-trophic level of the ecosystem and contribute to active carbon export to the deep ocean through diel vertical migrations. Better characterization of micronekton functional groups depending on relationships to environmental variables is useful for the management of marine resources, the conservation of biodiversity and a better understanding of climate change impacts. For this purpose, regionalization of global ocean into homogeneous provinces is an approach that is generating increasing interest. However, published regionalizations efforts (i) derived from environmental forcings, that do not specifically focus on micronekton and (ii) derived from acoustic backscatter, which do not allow direct estimates of micronekton biomass. Here, we propose to fill the gap between biophysical regionalizations and micronekton biomass. We notably defined biophysical biomes using global environmental variables known to affect micronekton: temperature of the epipelagic layer, temperature stratification, and net primary production (NPP). Six biophysical biomes were defined with a clustering method. A characterization of these biophysical biomes with simulated micronekton from the SEAPODYM-LMTL model displayed biome-specific relationships between biomass and the environmental variables used in the clustering (i.e. biomasses mostly structured by NPP and temperature). Biophysical biomes also displayed specific vertical structures suggested by modelled micronekton functional groups ratios. Then, a validation of biophysical biomes' boundaries was performed to identify potential vertical structure reorganization in acoustic backscattering response from adjacent biomes. The regionalization identified homogeneous areas in terms of acoustic vertical structure, which were also different between adjacent biomes. Finally, a comparison with another biomes' definition computed from micronekton biomasses suggested that environmental variables can account for only some of the variability of the micronekton structures.

1. Introduction

Micronekton are defined as organisms in a size range from 2 to 20 cm. They contain a wide diversity of taxa dominated by crustaceans, fish and molluscs. Some of the micronekton species are known to perform diel vertical migration (DVM), presumably to forage in the surface layer at night while minimizing predation risk and returning at depths to hide from predators during the day (Benoit-Bird et al., 2009). This daily vertical migration, which takes place all over the world ocean can be seen as the biggest animal daily migration on Earth and notably impacts the ocean carbon cycle by actively exporting carbon below the mixed layer (Bianchi et al., 2013; Gorgues et al., 2019; Pinti et al. 2021). Micronekton also constitute a key intermediate trophic level of the oceanic food web. They feed mostly on zooplankton and are the main prey of marine large predators, some of which are of crucial economic importance (e.g. tunas, see *Economic and Development Indicators and Statistics: Tuna Fisheries of the Western and Central Pacific Ocean;* Bell et al., 2015; McCluney et al., 2019). Despite their pivotal position, knowledge on micronekton remains fragmented with only a few rough estimates of biomasses (Irigoyen et al., 2014; Proud et al, 2019).

* Corresponding author at: 30 rue d'Assalit, 31500 Toulouse, France.
E-mail addresses: albernhesarah@gmail.com (S. Albernhe), thomas.gorgues@ird.fr (T. Gorgues), patrickl@spc.int (P. Lehodey), christophe.menkes@ird.fr (C. Menkes), otitaud@groupcls.com (O. Titaud), smagondelagiclais@groupcls.com (S. Magon De La Giclais), aconchon@groupcls.com (A. Conchon).

https://doi.org/10.1016/j.pocean.2024.103370
Received 19 April 2024; Received in revised form 10 October 2024; Accepted 21 October 2024
Available online 31 October 2024

Direct observations of the micronekton rely mainly on trawl sampling, which are costly and complex deep-water operations to perform (Pakhomov et al., 2010; Kwong et al., 2018). The method is known to be subject to biases due to avoidance and poor catchability of some species (Kaartvedt et al., 2012), and the destruction of the more fragile organisms, e.g., salps (Barbin et al., 2024). Net catchability also strongly depends on mesh sizes and trawling speed. Optical measurements, such as photos or videos acquired by a camera system often mounted on an acoustic probe, can provide a promising non-destructive alternative to trawl sampling, not impacted by net catchability (Graham et al., 2003; Allken et al., 2021). However, these measurements are challenging to conduct due to the lack of natural light below the epipelagic layer, and the impact of artificial lighting on micronekton behavior (Benoit-Bird et al., 2023). The analysis of top predators' diets can also serve as a method for observing micronekton and assessing trophic connections between mid- and high-trophic levels (Lansdell and Young 2007; Allain et al., 2012). However, since this approach involves examining stomach contents, identifying and quantifying micronekton organisms can be challenging, particularly depending on their state of digestion. As a result, the analysis is largely qualitative. Indirect observations through ship-borne acoustic echosounders are a non-destructive alternative approach which presents the advantage of covering large ocean areas (Haris et al., 2021; Ariza et al., 2022). Acoustic methods provide a proxy for micronekton density along the water column, with the single narrowband frequency of 38 kHz being commonly used (Kloser et al., 2009). However, interpreting backscattering response from this single frequency is challenging in ecosystems with a high diversity of taxa (Benoit-Bird and Lawson, 2016; Barbin et al., 2024). At 38 kHz, the scattering response is primarily influenced by organisms with gas-filled swim bladders, such as myctophids but not only. Additionally, the acoustic response of individuals varies with their size and orientation (McGehee et al., 1998; Scoulding et al., 2017). Although acoustic methods have biases in observing micronekton, detecting shifts or reorganizations in the backscatter intensity vertical structure is a useful tool for identifying ecosystem changes.

Machine learning techniques have significantly facilitated the use of acoustic datasets to define global oceanic acoustic micronekton biomes. For instance, Proud et al. (2017) and Ariza et al. (2022) defined, at global scales, biomes of similar acoustical characteristics related to a few environmental variables (e.g. temperature and (i) primary productivity and wind stress for Proud et al. (2017) or (ii) chlorophyll and subsurface oxygen for Ariza et al. (2022)). Yet, the approaches based on acoustic data present the drawbacks of not allowing direct estimates of micronekton biomasses or species composition without broad assumptions or additional ground truthing data (Davison et al., 2015; Benoit-Bird and Lawson, 2016; Barbin et al., 2024).

Thus, the observation and biogeographic description of mesopelagic provinces must be supplemented by statistical or mechanistic modeling. Environmental and acoustic data can be crossed in a statistical approach (Proud et al., 2017; Ariza et al., 2022), converted into biomass by theoretical scattering models (Benoit-Bird 2009; Jech et al., 2015; Barbin et al., 2024) or compared to ocean ecosystem models with more or less complex representations of the trophic levels, food web and energy transfer (e.g., Ecopath (Polovina and Marten, 1982; Walters et al., 1997); Lehodey et al., 2010; Apecosm (Maury, 2010; Dupont et al., 2023); Blanchard et al., 2017; Anderson et al., 2019; Petrik et al., 2019; Hatton et al., 2021; Hill Cruz et al., 2023).

Here, we thus propose a complementary approach to the biomes definitions and characterizations used in acoustic-based studies, focusing rather on biomasses of micronekton from the Spatial Ecosystem and Population Dynamics Model – Low and Mid Trophic Levels (SEAPODYM-LMTL: Lehodey et al., 2010; 2015; Conchon, 2016) model. The objective is to identify homogeneous micronekton functional patterns using only a simple and parsimonious set of biophysical variables, and to evaluate the biomes' micronekton biomasses characteristics. Differences with acoustic based studies are discussed. Moreover, biomes defined by a common, simple, set of environmental conditions and aligned with the structure of micronekton functional groups, help to estimate large scale biomass from sparse, unevenly distributed data. In this study, we are able to use biophysical data to extrapolate micronekton biomass, which, to our knowledge, is a novel contribution to the literature. Our approach also stands out for its dynamical aspect, with the production of biomes' time series that offer an accurate framework to observe the spatio-temporal variability of micronekton.

The present study is organized as follows. First, we compute biophysical biomes by clustering environmental variables known to have an impact on micronekton biomass and which are derived from SEAPODYM-LMTL forcings. We provide, at global scale, a time-averaged as well as a monthly definition of these biomes for the period considered: 1998–2019. Then, we characterize the biomes' micronekton biomasses using SEAPODYM-LMTL model outputs. We analyze the biomes-specific relations between the vertically integrated density of micronekton biomass and the environmental variables used in the clustering, micronekton vertical structure patterns per biome, quantitative biological indicators, and values of global and migrant micronekton biomass per biome. Then, we perform a validation of biophysical biomes' boundaries using acoustic backscatter data (observations). Finally, we compute a second clustering derived directly from micronekton biomasses using SEAPODYM-LMTL model outputs, to compare with the clustering from environmental variables and determine the potential differences introduced by the dynamics of the micronekton groups in the model.

2. Material and methods

2.1. Micronekton model

The numerical ecosystem model used in this study to compute micronekton biomass density is SEAPODYM-LMTL (Lehodey et al., 2010, 2015; Conchon, 2016). It is based on the energy transfer across the food web (Iverson, 1990) and a simplified view of the global marine ecosystem considering phytoplankton, zooplankton, micronekton and top predators. SEAPODYM-LMTL focuses on low and mid-trophic levels (mesozooplankton and micronekton respectively), modeling six micronekton functional groups defined according to their migratory behaviors, and one group of mesozooplankton.

SEAPODYM-LMTL uses the vertically integrated net primary production (NPP) as the source of energy for the functional groups of micronekton and mesozooplankton. Temperature and horizontal currents determine the population development (growth, mortality) and the spatial dynamics (Lehodey et al., 2010, 2015). The euphotic depth (Z_{eu}) is used to define three vertical layers inhabited by resident and migrant organisms: the epipelagic layer between surface and $1.5*Z_{eu}$ (~0–150 m), the upper mesopelagic layer between $1.5*Z_{eu}$ and $4.5*Z_{eu}$ (~150–400 m), and the lower mesopelagic layer between $4.5*Z_{eu}$ and $10.5*Z_{eu}$ (~400–1000 m) (hereafter referred to as L1, L2 and L3 respectively). Those definitions have been validated using a large dataset of acoustic data (Lehodey et al., 2015; Conchon et al., 2016). Temperature and currents are averaged over these three layers. In the following, we use the notation T1, T2, T3 for temperatures averaged in layer 1, 2 and 3 respectively. Unlike the resident functional groups, the migrant micronekton functional groups perform DVM and are exposed to different temperature and currents conditions in the vertical layers. Therefore, according to the layers inhabited in the daytime and nighttime (nomenclature: $Layer_{day}.Layer_{night}$), the model simulates three resident and three migrant micronekton functional groups: epipelagic (L1.L1), upper mesopelagic (L2.L2), lower mesopelagic (L3.L3), migrant upper mesopelagic (L2.L1), migrant lower mesopelagic (L3.L2) and highly migrant lower mesopelagic (L3.L1).

The model domain is global with a monthly resolution and a spatial resolution of a quarter of a degree. The time series of the simulation extends from January 1998 (corresponding to the commissioning of

ocean color observation satellites) to December 2019. Prior to vertical averaging, 3D temperature and currents are taken from the FREEGLORYS ocean simulation, produced by Mercator Ocean International using the ocean general circulation model NEMO (Madec et al., 2008) in a configuration without data assimilation (Lellouche et al., 2021). NPP (mmolC/m^2/day) is computed with VGPM (Vertical Generalized Production Model, Behrenfeld and Falkowski, 1997), from satellite chlorophyll-a concentration. As there is no reliable satellite chlorophyll data for latitudes above ~ 60° due to sun glint, cloud coverage, and low light levels (e.g. Gregg & Casey, 2007), the dataset has been completed in high latitudes by a biogeochemical product using chlorophyll from PISCES biogeochemical model (Aumont et al., 2015) (See detailed information about the data in supplementary material, Table S1). This is to avoid gaps in forcing data.

2.2. Clustering analyses

In this section, we introduce the two clustering analyses conducted in our study: the biophysical clustering (derived from biophysical variables), and the biomass clustering (derived from modelled micronekton biomass). The flowchart in Fig. 1 illustrates the implementation of the methodology, including the two clustering processes described below.

2.2.1. Biophysical clustering

In order to identify potential micronekton homogeneous patterns using only a parsimonious set of environmental variables, we regionalize the global ocean in biomes based on environmental variables (hereafter referred to as "biophysical biomes"). Biophysical biomes are defined as contiguous areas with no gaps and no overlaps, displaying homogeneous biophysical characteristics. These biomes are built using the integrated NPP, the mean temperature in the epipelagic layer of SEAPODYM-LMTL (T1) and the temperature difference between the first and second pelagic layers (i.e. epi and upper *meso*-pelagic layers) used in SEAPODYM-LMTL, as an index of the stratification (hereafter referred to as "stratification"). These biophysical variables are derived from the ones that force SEAPODYM-LMTL, and are similarly computed from the FREEGLORYS ocean simulation (for T1 and stratification) and VGPM, supplemented by PISCES in high latitudes (for NPP) (Table S1). Note that (i) temperature is also used by Proud et al. (2017) and Ariza et al. (2022) to cluster acoustic data, (ii) NPP is also used in Proud et al. (2017), while Ariza et al. (2022) used chlorophyll concentration. Both latter variables (i.e. NPP and Chlorophyll) are highly correlated and present relatively similar patterns at global scale (Behrenfeld and Falkowski, 1997). Finally, the stratification used in this study is also related to the vertical mixing of the water column as are either the wind stress (Proud et al., 2017), the subsurface dissolved oxygen concentration (Ariza et al., 2022). Consequently, the use of the forcing variables of SEAPODYM-LMTL fits into a relatively similar philosophy underlying the choice of variables leading to regionalization, in our study as well as in those of Proud et al. (2017) and Ariza et al. (2022).

As the three environmental variables (i.e. epipelagic layer temperature, stratification and NPP) are partially correlated, a Principal Component Analysis (PCA) (Hotelling, 1933) is performed to reduce the dimension to independent variables while retaining the main modes of variability. Then, we perform a clustering (hereafter referred to as "biophysical clustering") on the PCA principal components that explain the most variance (see selected principal components' variances in section 3.1) to define the biophysical biomes.

We aim to create homogeneous biophysical biomes by identifying intrinsic structures or trends in the data without having any preconceived notions about clusters. To do so, we use the unsupervised machine learning clustering algorithm k-means (Lloyd, 1957; Pedregosa et al., 2011) that partitions the observations into k homogeneous clusters (k being a natural number less than the number of observations). Since the model is unsupervised, k-means requires that we define the optimal number of clusters k beforehand, to assign each individual to one of the k predefined clusters. Among the different existing methods to determine this optimal number of clusters k, we use the silhouette metrics (Rousseeuw et al., 1987; see Figure S1). The silhouette coefficient evaluates the quality of a dataset's partitioning in classification tasks. It measures both cohesion (how close data points are within a cluster) and separation (how distinct clusters are from each other). This

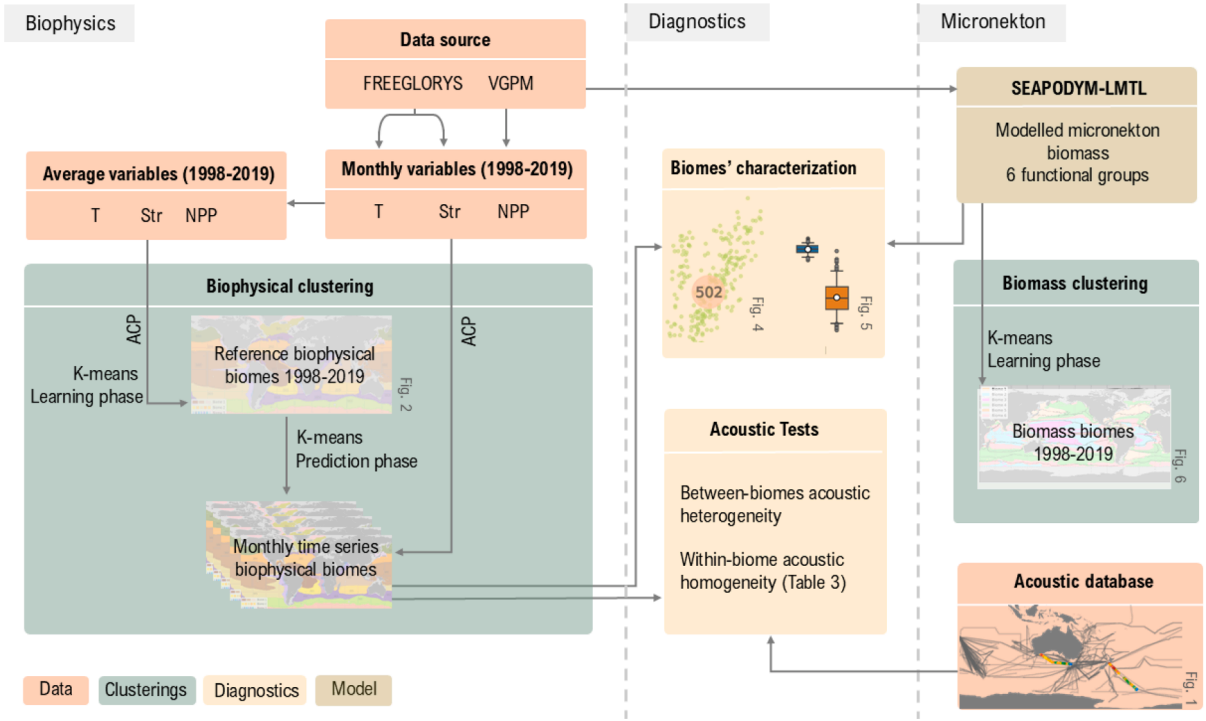

Fig. 1. Flowchart describing the implementation of the global methodology: data sources used, different methods and tests applied, and diagnostics. Each element of this figure is described in detail in the Material and Methods.

makes it a more comprehensive metric compared to others that may focus on only one aspect (for instance, the elbow method primarily considering cohesion, yielded inconclusive results when we tested it for determining *k*). Moreover, silhouette doesn't require labeled data, making it well-suited for unsupervised methods like k-means. We calculate the average silhouette coefficient for k-means clustering with different numbers of clusters. The optimal number of clusters *k* is the one that leads to the clustering with the higher average silhouette coefficient (see dashed line in Figure S1).

After having determined the clustering variables based on the PCA and set the optimal number of clusters, we apply the machine learning methodology. Machine learning methods involves two phases: (i) the training phase consists in the training of the machine learning model on input data to learn patterns from similarities in the data and (ii) the prediction phase consists in the application of the learned patterns to new data to make predictions.

Applying the training phase of the clustering model (i), we define static reference biophysical biomes representative of the entire period considered: 1998–2019, at global scale. The training phase of the clustering model is performed on the PCA principal components from time-averaged data (computed from monthly data spanning 1998 to 2019 with a ¼-degree spatial resolution), creating the reference clustering with *k* homogeneous clusters. After the training process, the model parameters are estimated, and we can use this model to make predictions on other data.

Hence, we apply the prediction phase of the model (ii) on monthly data over the same period 1998 to 2019 (monthly PCA principal components derived from monthly ¼ degree data). A monthly biophysical clustering is provided over the time period, i.e., monthly definition of biomes. This biophysical biomes' time series accounts for the seasonal to interannual variability. This dynamic approach facilitates the examination of micronekton's spatio-temporal variability and provides an accurate framework for gathering and analyzing field data, such as acoustic transects or trawl samples, within a unified spatio-temporal scale.

2.2.2. Biomass clustering

In order to understand how the biophysical regionalization compares to the micronekton biomasses, we run an alternative biomes' definition based on modelled micronekton biomasses.

A second clustering is thus performed using the same methodology as the biophysical clustering, but with modelled micronekton biomass instead of the biophysical variables. We refer to this as "biomass clustering". From the six functional groups of SEAPODYM-LMTL outputs, we derive three simplified ones, defined as:

- The epipelagic: resident epipelagic group (L1.L1).
- The mesopelagic: sum of resident mesopelagic groups (L2.L2 and L3.L3) and migrating between mesopelagic layers (L3.L2).
- The migrant: sum of mesopelagic groups migrating in the epipelagic layer (L2.L1 and L3.L1).

We consider two ratios of micronekton functional groups as clustering variables: respectively mesopelagic and migrant over the epipelagic simulated biomasses. The biomass clustering is performed following the same steps than the biophysical clustering (i.e. using k-means training phase, but no PCA). Six micronektonic homogeneous biomes are defined by clustering of these two variables, hereafter referred to as "biomass biomes".

2.3. Biophysical biomes, provinces and characterization

In this section, we first present a sensitivity analysis to assess the robustness of the biophysical clustering. Next, we explain how provinces are derived as subdivisions of the biophysical biomes. Finally, we introduce the methods used to characterize the biophysical biomes and provinces based on modeled micronekton biomass.

2.3.1. Sensitivity analysis

The robustness of the biophysical clustering obtained with the reference dataset, i.e., FREEGLORYS for the physical variables and VGMP for the biological variable (see section 2.2.1., and Table S1), is tested by computing other biophysical clusterings derived from alternative environmental datasets. These alternative datasets include physical data from ARMOR3D (Guinehut et al., 2012; Mulet et al., 2012) and biological data from Eppley-VGPM (Eppley 1972; Morel 1991; Behrenfeld and Falkowski 1997) or from the biogeochemical model PISCES (See detailed information about the data sources in supplementary material, Table S1).

We use the F1-score, a widely used metric in machine learning and information retrieval, as the evaluation metric to compare the biophysical clusterings derived from the different datasets (Hastie et al., 2009). F1-score is the harmonic mean of precision (ratio of true positives to the sum of true positives and false positives, measuring the accuracy of positive predictions) and recall (ratio of true positives to the sum of true positives and false negatives, measuring the model's ability to capture all actual positives). F1-score ranges from 0 to 1; respectively the worst and best degree of prediction. This metric is calculated between the reference biophysical clustering using FREEGLORYS – VGPM and every alternative biophysical clustering derived from the other datasets. The following combinations of data sources for alternative clusterings are considered for comparison with the reference dataset FREEGLORYS-VGPM: FREEGLORYS-Eppley-VGPM and FREEGLORYS-PISCES for sensitivity to biogeochemical forcings, and AMOR3D-VGPM for sensitivity to physical forcings.

The results of the biophysical clustering analysis is consistent regardless of the physical data source used: substituting ARMOR3D data to FREEGLORYS generates 93 % of clustering accurate prediction compared to the reference clustering (Table 1). On the other hand, the clustering is more sensitive to the biogeochemical data source used. Indeed, substituting the biological data source VGPM with VGPM-EPPLEY generates 76 % of accurate prediction and this decreases to 69 % when substituting with the data derived from biogeochemical model PISCES.

A sensitivity analysis is also conducted using varying spatial resolutions, demonstrating that degrading the resolution from 0.25 to 1 degree does not affect the clustering results. The reference biophysical clustering considered hereafter is based on the same datasets as those used to force SEAPODYM-LMTL (i.e., FREEGLORYS and VGPM) with a spatial resolution of 0.25 degrees.

2.3.2. From biophysical biomes to provinces

The biophysical clustering leads to the definition of large homogeneous biophysical biomes, characterizing environmental regimes at global scale. From these large biomes, we define provinces as subregions split between different ocean basins and hemispheres. As similar oceanographic regimes are repeated in multiple locations, biophysical biomes are extended over different ocean basins. However, large-scale communities with similar structural and functional compositions may nonetheless consist of distinct species assemblages due to

Table 1
Sensitivity analysis results. F1-scores calculated to compare the reference biophysical clustering (VGPM-FREEGLORYS) with alternative biophysical clusterings from different datasets. F1-score ranges from 0 to 1 (respectively the worst and best degree of prediction).

	VGPM – FREEGLORYS	VGPM – ARMOR3D	Eppley VGPM – FREEGLORYS	PISCES – FREEGLORYS
VGPM-FREEGLORYS	1	0.93	0.76	0.69

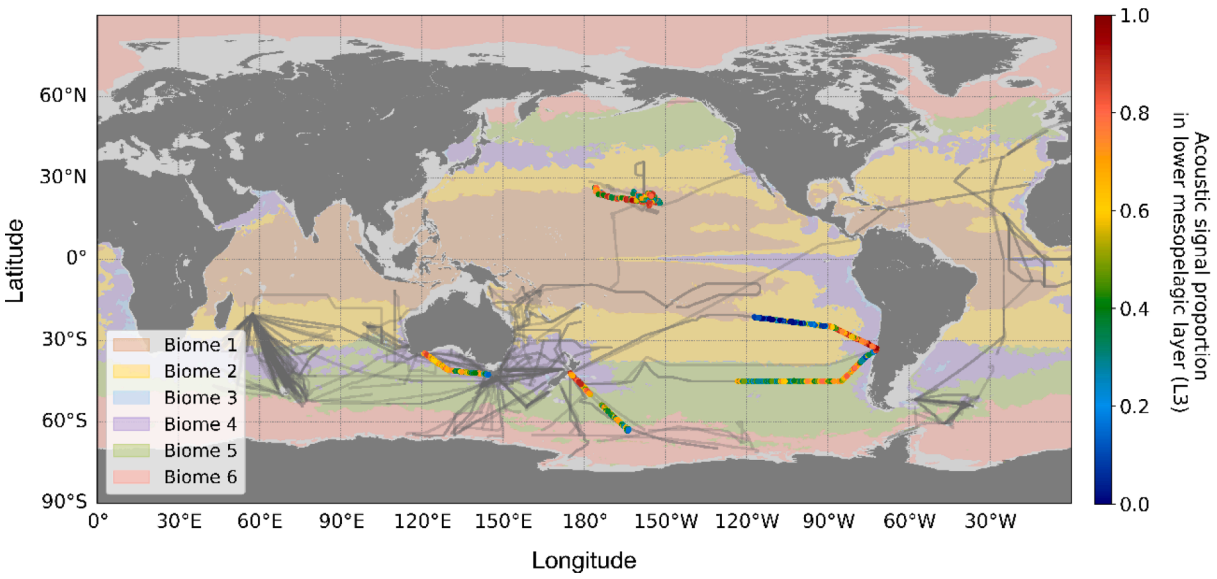

Fig. 2. Map of the acoustic database available for this study from 2006 to 2019, transects are plotted in grey. The acoustic transects plotted in color are the ones registered in November 2017. The colors represent the proportion of backscattering in layer 3 of SEAPODYM-LMTL out of the total acoustic backscatter in the water column (from blue to red, respectively low and high proportions). The background map displays the biophysical biomes in November 2017 (extracted from the monthly biomes time series). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

spatially separated co-evolutions between different hemispheres or ocean basins (Spalding et al., 2012; Sutton et al., 2017). Here we defined provinces as biophysical biomes' sub-divisions at the scale of ocean basin and hemisphere characterized both by areas of stable biophysical drivers and potential taxonomic identity. A map of these provinces is provided in the supplementary material, also identifying the biome they belong to (Figure S2).

2.3.3. *Characterization with modelled micronekton biomass*

With the purpose of demonstrating the relationship between our biophysical clustering and micronekton biomass, we use modeled micronekton biomass to characterize the biophysical biomes and provinces.

We restrict the spatial domain of the biophysical clustering to depths greater than 10.5*Zeu (i.e. roughly 1000 m deep, the bottom depth for L3), ensuring the existence of the three pelagic layers of SEAPODYM-LMTL. This restriction allows the biomes' characterization with the micronekton biomass modeled by SEAPODYM-LMTL for all functional groups. This excludes the shallow coastal areas (see grey areas in Fig. 3). From the six functional groups of SEAPODYM-LMTL, we derive three simplified ones, epipelagic, mesopelagic and migrant (see detailed composition of the simplified groups section 2.2.2.).

Spatially averaged biomasses of these three functional groups are computed from SEAPODYM-LMTL outputs to characterize the biophysical biomes and provinces in terms of micronekton biomass and vertical structure of the ecosystem. For each biome or province, we compute the mean micronekton biomass (wet weight in g/m2) for each functional group, examine the relationships between total biomass, epipelagic layer temperature, and NPP, and analyze the biomass ratios of mesopelagic/epipelagic and migrant/epipelagic to understand the ecosystem's vertical structure.

This way, we are able to compare micronekton biomass and vertical structure between the different biophysical biomes, but also between the provinces belonging to the same biome.

Biome's characterization by quantitative biological indicators and values of global and migrant micronekton biomass can also be useful tools for future mid-trophic biological studies. We compute two biological indicators for each biome: the generation time (t_G) and maximum lifespan (t_{max}) of organisms, which both depend on temperature (Huntley and Lopez, 1992; Gillooly et al., 2002; Lehodey et al., 2001; 2010) (Table S2). The generation time is the age at maturity for the micronekton population considered. The maximum lifespan is defined as in Lehodey at al., 2010 as the time necessary to see the population reduced to a specific level (e.g., x = 5 % here). As the relationships are exponentially decreasing functions with temperature, the colder the biome, the longer the generation time and maximum lifespan (Lehodey at al., 2010). Here we consider a monthly time series of temperatures from 1998 to 2019 for each pelagic layer of SEAPODYM-LMTL, spatially averaged for each biome. We compute the values of t_G and t_{max} from the mean temperature of each vertical layer per biome (Table S2). These values can be compared to the biological characteristics of resident species in each corresponding layer of the biophysical biomes. For migrant species, these parameters values would need to be weighted by the time spent in each layer based on daytime and nighttime durations, which vary by date and latitude.

2.4. *Acoustic database and biophysical biomes' boundaries validation*

Acoustic transects are used to validate the biophysical biomes' boundaries. Here we use in *situ* acoustic backscattering response from echosounders at the frequency of 38 kHz. Despite the biases associated with acoustic methods in micronekton observation (described in Introduction section), acoustics provide a proxy of the organisms' density throughout the water column, allowing for the observation of their vertical and spatial distribution. These data are used to investigate (i) the heterogeneity of acoustic backscatter profiles between different biophysical biomes, and (ii) the homogeneity inside each biome.

2.4.1. *Acoustic database*

We use a database of 394 echointegrated acoustic transects collected between 2006 and 2019 (Fig. 2) and compiled from the sources listed in Table 2.

More detailed information on acoustic data, such as calibration acquisition methods, data processing software or transducer specifications is available in supplementary material (Table S3). Our acoustic database is similar to the one used in Proud et al., 2017 (e.g., IMOS and BAS transects), but we are missing some from the Pelagic Ecology Research Group (University of St Andrews, UK) and the SMILES project (Plymouth University, UK). However, we've added several transects to our database, including NOAA's transects, and additional IMOS and BAS transects.

Table 2
Description of the acoustic database sources, number of transects (N = 394) and references.

Data	Description	Number of transects	Source
IMOS BA-SOOP (Integrated Marine Observing System, Bio-Acoustic Ships of Opportunity sub-facility)	Australian data maintained by the CSIRO. These data are mainly situated in the South Indian Ocean and the South Pacific Ocean.	298	Available through Open Access to Ocean Data aodn.org.au
British Antarctic Survey (BAS)	Data gathered around South Georgia and between the UK and South Georgia.	82	Available through bas.ac.uk
Mycto-3D-MAP and PIRATA cruises	A set of transects from French IRD-CNRS institutes collected during the research cruises Mycto-3D-MAP in the Southeast Indian Ocean and PIRATA in the equatorial Atlantic Ocean.	Mycto-3D-MAP: 10 PIRATA: 2	Mycto-3D-MAP: Béhagle et al., 2016 Yves Cherel, personal communication PIRATA: Habasque et al., 2024 https://doi.org/10.17600/15001800 https://doi.org/10.17600/16002300
NOAA's TZCF Oceanographic Survey (SE0902L1, EK60) and (SE1102L2, EK60), North Pacific Ocean	Cruises in 2009 and 2011 in the North Pacific Subtropical Frontal Zone.	2	Réka Domokos, personal communication doi:10.7289/V57S7KQ9 (2009) doi:10.7289/V53776PJ (2011)

2.4.2. Between-biomes heterogeneity test

To assess the heterogeneity of the acoustic backscatter profiles between biophysical biomes, acoustic transects crossing a boundary between two biomes are segmented at the point where they cross the biomes' boundary, in two distinct sections. Then, the nonparametric Mann Whitney statistical test (Wilcoxon, 1945; Mann and Whitney 1947) is used to compare the data distributions between the two sections, i.e., the distribution of the acoustic backscatter data on each side of the biome border. The Mann Whitney *U* test is used to test the hypothesis that the distributions of two groups of data are similar. Among the numerous tests existing to compare data distributions, it is probably one of the most relevant one in our case, being non-parametric (our data are not normally distributed) and considering only two populations of independent samples. It computes a p-value as a metric to reject or not the null hypothesis (H_0: "the two distributions are identical"). If the p-value is less than the significance level $\alpha = 0.05$, the null hypothesis can be rejected with a confidence of 95 %, which we consider here as a validation of the biomes' heterogeneity. This test is hereafter referred to as the between-biomes heterogeneity test.

2.4.3. Within-biome homogeneity test

In conjunction with the between-biomes heterogeneity test, the acoustic data variance is used as a comparison metric to assess the homogeneity of the backscattering response within each biome. For each transect straddling two biophysical biomes, we compare the variance of the acoustic data within a biome to the one computed over the entire transect. A threshold of 5 % is considered to enforce the significance of that difference in the variances. This test is hereafter referred to as the within-biome homogeneity test.

2.4.4. Data treated by the heterogeneity and homogeneity tests

In order to preserve the information on the vertical structure and DVM given by the acoustic backscatter profiles, daytime and nighttime periods are distinguished based on solar elevation angle (daytime is classically taken when the local solar elevation is greater than 18°). The between-biomes heterogeneity test and within-biome homogeneity test described above are performed independently on daytime and nighttime data. For the same vertical structure information considerations, we also apply the statistical tests distinctively on the acoustic data from each of the three pelagic layers of SEAPODYM-LMTL. More precisely, as the transects from different acoustic campaigns are not intercalibrated, we do not consider backscattering signal directly, but the proportion of backscattering signal received in each pelagic layer of SEAPODYM-LMTL out of the total acoustic backscatter in the water column, on each measurement point of the transects (see acoustic transects plotted in color and associated color bar Fig. 2).

Therefore, the between-biomes heterogeneity test provides six p-values per transect crossing a boundary, compares the acoustic data between two biophysical biomes for (i) the epipelagic layer during daytime and (ii) nighttime, (iii) the upper mesopelagic during daytime and (iv) nighttime, (v) the lower mesopelagic during daytime and (vi) nighttime. The within-biome homogeneity test is also applied on the six sub-sections of each transect crossing a boundary between two biophysical biomes.

We restrict the analysis to transects with at least one hundred measurement points within each biome, ensuring a reliable representation of the acoustic response. As a result, transects that consist in very few measurement points sampling a biome (around ten) are excluded.

The 394 acoustic transects of our database are used (Fig. 2). 60 % of the boundaries between monthly biomes are crossed by at least one transect of the database. This database enables the study of 380 boundary crossings between 2006 and 2019. We can only propose a global validation of the biophysical clustering using acoustics, not a validation per specific boundaries because of a lack of transects in some less documented areas.

3. Results

3.1. Biophysical biomes' definition

To define the homogeneous biophysical biomes, we perform a clustering on principal components generated by the PCA performed on the three environmental variables (i.e. epipelagic layer temperature, stratification and NPP). We selected the two principal components that explain the most variance, accounting for 97,9% of the variance (72,4% and 25,5% for the first and second PCA respectively).

Using the two principal components as new variables for the biophysical clustering, the optimal number of clusters is then determined to be six using the silhouette metric (Figure S1) (see section 2.2.1.). We refer to these six spatial clusters computed from averaged data over the period 1998–2019 as "reference biophysical biomes" (Fig. 3).

A video animation displaying the monthly time series of provinces is available in the Supplementary Material (Video S1).

Focusing on the biophysical conditions for each biome, we consider the data distribution for averaged epipelagic layer temperature, stratification, and NPP. Fig. 4 shows monthly values of these three variables

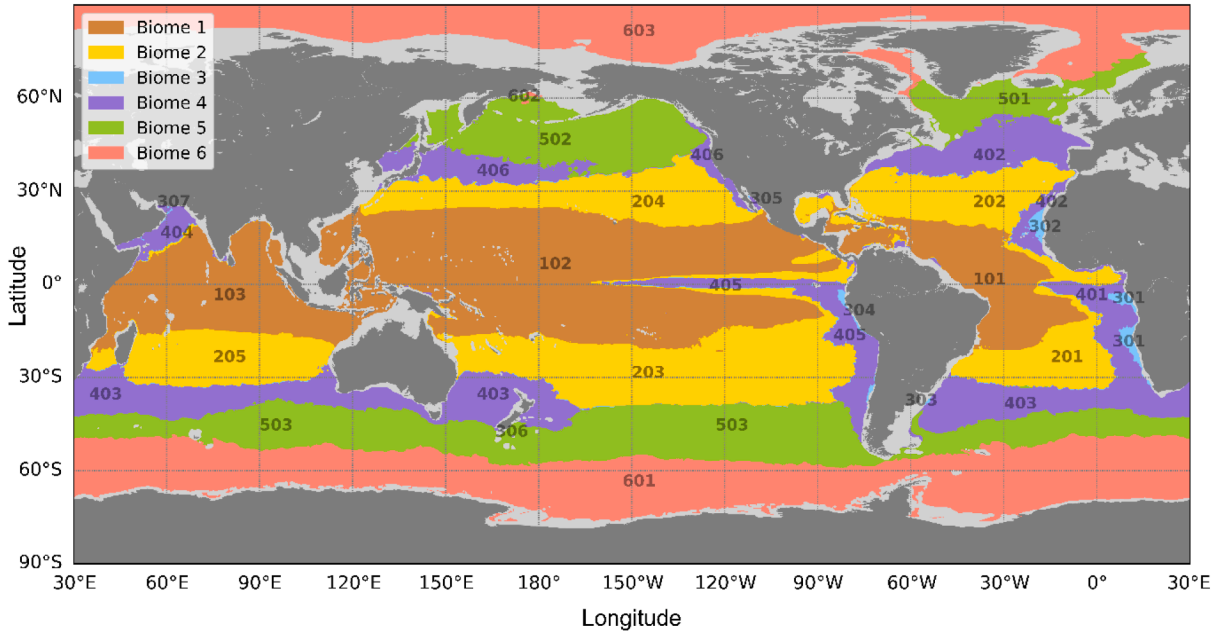

Fig. 3. Map of reference biophysical biomes obtained by PCA principal component clustering from averaged epipelagic layer temperature, stratification, and NPP over the 1998–2019 time period. The 6 biophysical biomes are represented by a color defined in the legend. Geographical separation between different areas of the same biome defines 27 associated provinces. One label is attributed to each province with the hundreds' digit corresponding to the biome in which they belong. Grey areas delimitate the domain where the depth of the water column is not sufficient to ensure the existence of the three pelagic layers of SEAPODYM-LMTL (see section 2.3.2.).

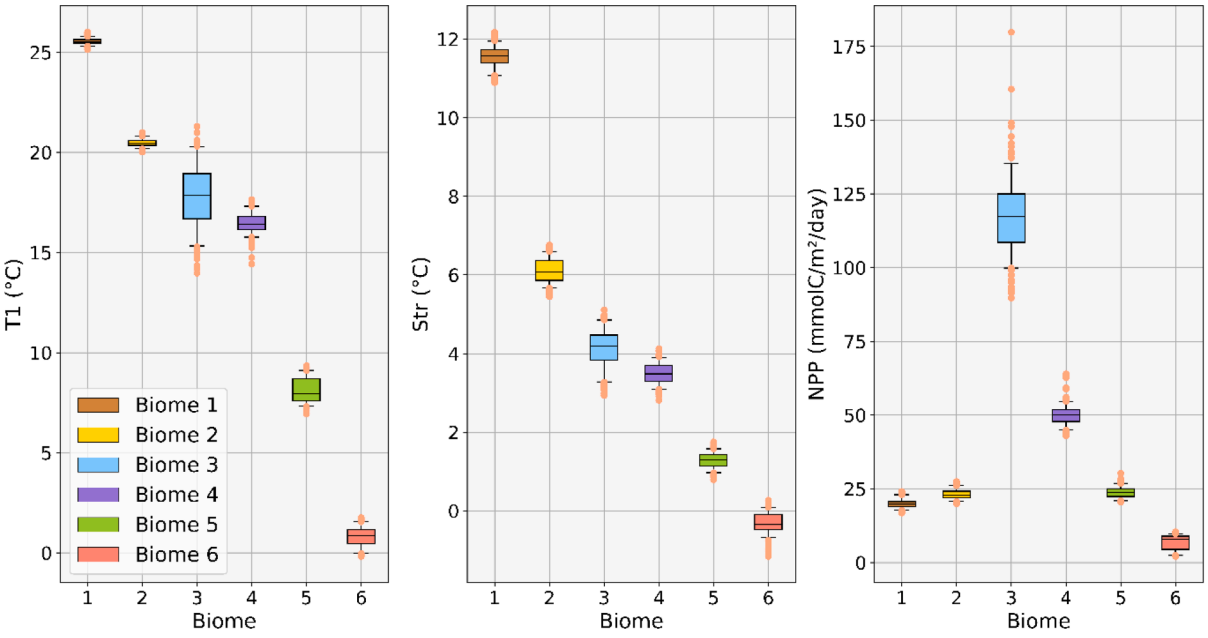

Fig. 4. Biophysical biomes characterization with monthly environmental forcings: epipelagic layer temperature (T1(°C)), stratification (Str(°C)), NPP (mmolC/m^2/day), from 1998 to 2019. The data considered are monthly values of T1, Str and NPP spatially averaged over each biome (i.e. one value of each environmental variable per month per biome). The boxplots show the median of data distribution in the rectangles' centers, top and bottom of the rectangles represent first and third quartiles, segments' ends represent percentiles 5 and 95, and the orange dots represent the outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from 1998 to 2019, spatially averaged for each biome. The spatial distribution of the reference biophysical biomes shows mostly a latitudinal structure with the addition of areas influenced notably by upwellings (either eastern boundary or equatorial, see Biomes 3 and 4) (Fig. 3). The six biophysical biomes can be categorized as follow:

- **Biome 1.** ***The tropical biome*** extends roughly between latitudes 20°N and 20°S in the three oceans. It is characterized by the warmest and most stratified waters associated with relatively low biological production (below 25 mmolC/m^2/day).
- **Biome 2.** ***The subtropical biome*** is mostly centered around 30°N and 30°S in all basins. It is characterized by warm water temperatures

combined with relatively high stratification (yet almost half the one associated with Biome 1) and production comparatively as low as that of Biome 1.

- **Biome 3.** ***The eastern boundary coastal upwelling systems*** display lower epipelagic layer temperature than Biomes 1 and 2, moderate stratification, and is the most productive biome. It is also the smallest in surface.
- **Biome 4.** ***The oceanic mesotrophic systems*** display an average epipelagic layer temperature as well as a stratification close to Biome 3 but only half of its biological productivity. This second most productive biome has a much bigger surface than Biome 3, as it includes equatorial upwelling zones, large oceanic extensions of the eastern boundary upwelling systems between equator and latitudes 50°, and temperate mid-latitudes regions centered around latitudes 50° in the Northern and Southern Hemispheres.
- **Biome 5.** ***The sub-polar biome*** is weakly stratified and productive. It covers the cold waters of the Arctic and Southern Oceans, roughly between 40° and 60° in latitudes, and the seasonally Baffin Sea.
- **Biome 6.** ***The polar biome*** displays the weakest stratification, and the lowest epipelagic layer temperature than any other biomes and extends from latitude 60° to the poles. The NPP is also the lowest of all biomes.

From these six biophysical biomes, five will be analyzed in term of the SEAPODYM-LMTL predicted biomass. Indeed, the NPP used within the polar biome (i.e. #6) to force SEAPODYM-LMTL is not entirely consistent with the one used in other biomes as it relies on chlorophyll biogeochemical model outputs (used at latitudes higher than 60°) rather than from satellite chlorophyll. This geographical discrepancy in the NPP and the known biases of the biogeochemical model in high latitudes region (see Aumont et al., 2015) do not allow a meaningful analysis of this specific polar biome in terms of the SEAPODYM-LMTL computed biomasses.

3.2. Characterization of biophysical biomes and provinces

The biophysical clustering, with the different environmental characteristics that are described in the previous section, leads to a contrasted predicted density of total micronekton biomasses averaged over each biome. In this section, we characterize the biophysical biomes and provinces by (i) biomes-specific relations between density of total micronekton biomasses and the environmental variables used in the biophysical clustering (Fig. 5), (ii) comparison of the characterization between provinces belonging to a same biome (Fig. 5), (iii) micronekton vertical structure patterns from functional groups biomass ratios (Fig. 6), (iv) quantitative biological indicators and values of global and migrant micronekton biomass per biome and province (Tables S2 and S4 respectively).

The results shown in Fig. 5 and Fig. 6 are derived from the monthly biomes definitions, with the aim of exploring variability in micronekton biomass and vertical structure patterns within biomes and provinces.

Fig. 5 shows that the relations between the density of total micronekton biomasses and the environmental variables used in the biophysical clustering (NPP and epipelagic layer temperature respectively) differ markedly between biomes (i). Biophysical biomes are firstly structured following the NPP (see the background scatterplots in Fig. 5, that can be easily divided along the x-axis which depicts NPP). Clearly by itself, the NPP is sufficient to discriminate the productive Biomes 3 and 4. These biomes are characterized by specific ranges of NPP (respectively 40–60 mmolC/m^2/day and $>$ 80 mmolC/m^2/day), and display a wide range of total micronekton biomass density values. On the other hand, Biomes 1, 2 and 5 share a common range of NPP (about 20 mmolC/m^2/day) but display specific ranges of total micronekton biomass density. For these biomes, total micronekton biomass density relates to the epipelagic layer temperature (see the foreground large circles' colors in Fig. 5, showing increasing temperature alongside y-axis which stands for micronekton density).

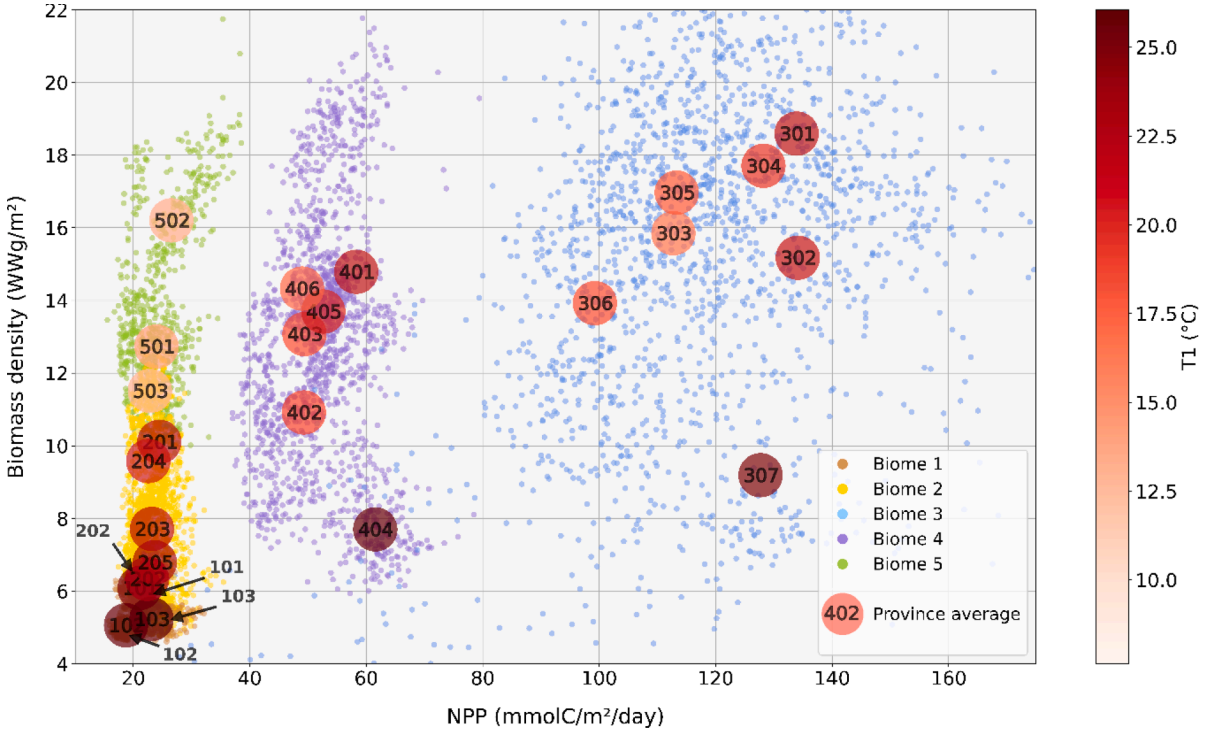

Fig. 5. Micronekton total biomass density (WWg/m^2) as a function of NPP (mmolC/m^2/day) and epipelagic layer temperature (T1 (°C)). The background scatter plot represents the monthly data for each biome (1 point per month on 1998–2019 per biome). The colors of this background scatter plot are associated with the color of each biome in Fig. 3 (see legend bottom right corner). The foreground large circles represent temporal average data on 1998–2019 for each province, indicating provinces' labels. The colors of these foreground large circles represent the mean epipelagic layer temperature of the province (see T1 color bar on the right). Modeled biomass density characterizing the biophysical biomes and provinces is extracted from SEAPODYM-LMTL.

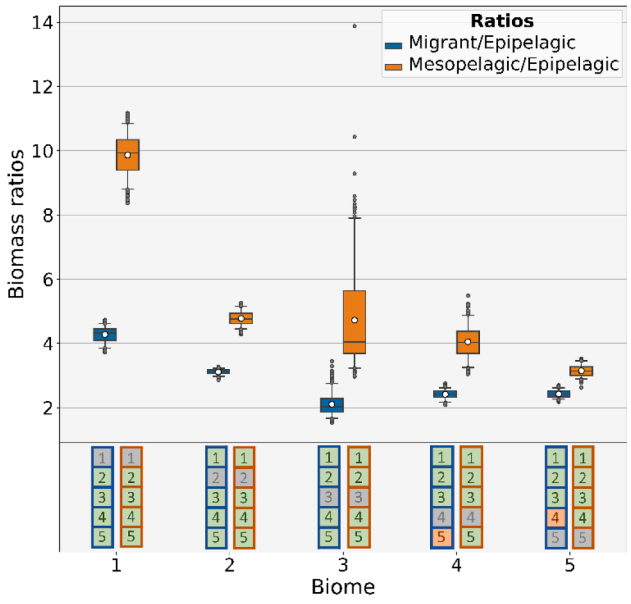

Fig. 6. Boxplots of biomass density ratios of migrant/epipelagic (blue) and mesopelagic/epipelagic (orange) micronekton groups, monthly ratios per biome for 1998–2019. The boxplots show the median of ratios distribution displayed by the black horizontal line inside le boxes (to be distinguished from the white circles that represent the mean), top and bottom of the rectangles represent first and third quartiles, white dots are the mean values, segments' ends represent the percentiles 5 and 95. The gray dots are the outliers. The lower part of the figure indicates, for each boxplot, from which other biomes' boxplots data distributions show a statistically significant difference. Each column indicates with which other biome (with its label from 1 to 5, self-label in grey) the data distribution is significantly different (light green) or not (light red). For instance, the migrant/epipelagic ratio of Biome 4 and Biome 5 are not significantly different. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

These large circles in the foreground in Fig. 5 allow the comparison of the characterization between provinces belonging to the same biome (ii). The circles' colors, corresponding to the mean epipelagic layer temperature of each province, highlight the temperature homogeneity in the different provinces belonging to a same biome, for Biomes 1, 2 and 5 respectively. *A contrario*, they display the important temperature variability in the different provinces of Biomes 3 and 4, i.e., the most productive biomes, also adding variability in total micronekton biomass density.

Variability in micronekton vertical structure patterns between biophysical biomes (iii) is also explored using migrant/epipelagic and mesopelagic/epipelagic biomass density ratios from SEAPODYM-LMTL modeled biomass (Fig. 6). This diagnostic suggests specific vertical structure patterns in the different biophysical biomes for both ratios, with distinct data distributions and relatively low dispersion. To assess the significance of these differences between biophysical biomes, the data distributions used for each boxplot are compared in pairs using the Mann Whitney statistical tests (for both ratios distinctively). Every pair of biophysical biomes displays statistically significant differences in data distribution for both ratios, except between Biomes 4 and 5 for the migrant/epipelagic ratio (p_{value} = 0.43) (see lower part of Fig. 6 displaying orange boxes between biomes without statistically significant difference in data distribution). Elsewhere, both ratios increase as the latitude of the biome approaches the equator (i.e. increases with temperature in the epipelagic layer). The mesopelagic/epipelagic ratio of Biome 1 is particularly high, around 10. Since this biome is highly stratified (Fig. 4), the relationship between generation time and temperature (Gillooly et al., 2002; Lehodey et al., 2010) drives a strong mortality gradient between surface and depth. In addition to the difference in temperature (and thus mortality rates), the simulated biomass is also controlled by the energy transfer coefficient (associated to each functional group) from the NPP the SEAPODYM-LMTL biomasses. Ratio values for the provinces in Biome 2 (subtropical) show very low dispersion compared to other biomes. It is the opposite for Biomes 3 and 4 that also present the widest range of epipelagic layer temperature (Fig. 4).

The SEAPODYM-LMTL model with the parametrization we use here (Lehodey et al., 2010) previously estimated the global micronekton biomass to be 2.82 Gt (WW), including 0.96 Gt (WW) of migrant mesopelagic organisms (sum of functional groups 2.1 and 3.1). Here, we refine those by providing biomass estimates for each biome and province (see Table S4 supplementary material), which can be a valuable input for regional studies.

3.3. Biophysical biomes' boundaries validation with acoustic data

As shown above, biophysical biomes show different characteristics of predicted biomass of SEAPODYM-LMTL functional groups (Fig. 6). We use acoustic data to validate if such transitions are also observed in the acoustic backscattering response from adjacent biomes. Considering acoustic transects that cross at least two biophysical biomes, we propose two metrics measuring (i) the heterogeneity of the acoustic backscatter profiles between two neighboring biomes and (ii) the homogeneity of the acoustic backscatter profiles inside each biome.

The two metrics are applied to each of the three vertical layers of SEAPODYM-LMTL and distinguishing daytime and nighttime periods (see section 2.4.4.). This makes the diagnostic more accurate for the description of vertical structure of the acoustic signal and DVM.

Acoustic validation of biophysical biome boundaries uses the monthly time series of these biomes. To detect transitions in acoustic backscattering between adjacent biomes, precise biomes' borders aligned with the acoustic recording period are crucial. The monthly timeseries displays a significant seasonality in biomes position and expansion, notably in poleward shifts, making monthly resolution more relevant than a static approach for comparison with observations.

3.3.1. Between-biomes acoustic heterogeneity

To assess the heterogeneity of the acoustic backscatter profiles between two neighboring biophysical biomes, we apply the between-biomes acoustic heterogeneity test. This test involves applying Mann-Whitney U tests to the six sub-sections (three pelagic layers, during the daytime and nighttime) of each transect crossing a boundary between two biophysical biomes. The null hypothesis (H_0) 'the two distributions are identical' (see section 2.4.2.), allows for the comparison of data distributions between the two biomes crossed.

In 93,8% of the cases (N = 1983), the null hypothesis H_0 of the Mann Whitney test is rejected with a confidence of 95 %. Rejection of the null hypothesis indicates that the acoustic data distributions are significantly different between the two biophysical biomes crossed by the acoustic transect.

3.3.2. Within-biome acoustic homogeneity

To assess the homogeneity of the acoustic backscatter profiles within each biophysical biome, we apply the within-biome acoustic homogeneity test (see section 2.4.3.). Our results show that the variance of acoustic data within a biome is significantly smaller than the variance over the whole transect in 61.76 % of all the cases (N = 3639) (Table 3).

The lower mesopelagic layer (L3) shows the greatest homogeneity in backscattering signal (Table 3). The lower mesopelagic layer (L3) coincides with the daytime habitat of the organisms of the Deep Scattering Layer (DSL), characterized by a strong acoustic response and the diel vertical migration (DVM) associated to light penetration (Aksnes et al., 2017). Some of the consistency observed in L3 may also be due to the influence of the scattering physics. The epipelagic layer (L1) is also strongly impacted by the DVM, representing the main zooplankton

Table 3
Results of within-biome acoustic homogeneity tests on the acoustic database, distinguishing daytime and nighttime data, in the three pelagic layers of SEAPODYM LMTL. Values are the proportion of within-biome acoustic homogeneity test showing that acoustic data variance within a biome is significantly smaller than the variance over the whole transect (Var_{biome} < 95 % $Var_{transect}$). Each test considers one transect, with Var_{biome} the variance of acoustic data within a biome, and $Var_{transect}$ the variance of acoustic data within the whole transect.

Daytime/Nighttime	Daytime acoustic data				Nighttime acoustic data			
Layer	L1	L2	L3		L1	L2	L3	
Var_{biome} < 95 % $Var_{transect}$ (% of within-biome acoustic homogeneity tests)	61,35 %	57,40 %	63,65 %		63,64 %	58,84 %	65,46 %	
		60,80 %				62,72 %		
				61,76 %				

vertical habitat. Migratory behavioral patterns are thus strongly reflected by the acoustic backscattering in these two layers. As they display the greatest homogeneity in the backscattering signal both for daytime and nighttime data (Table 3), it appears that these patterns are identified in our regionalization. The upper mesopelagic layer (L2) has the lowest homogeneity inside biomes, in all likelihood because it is a transition layer between the epipelagic and lower mesopelagic layers inhabited by migrant organisms of the DSL.

3.4. Biomass biomes' definition

For comparison purposes, here we define biomass biomes by performing an alternative clustering derived from modelled micronekton biomass (previously referred to as biomass clustering, see section 2.2.2.). The variables used are two ratios of modelled micronekton functional groups from SEAPODYM-LMTL outputs (i.e. respectively mesopelagic and migrant over the epipelagic biomass). From the clustering of these ratios, six homogeneous biomass biomes are defined (Fig. 7).

One identifies the Pacific equatorial upwelling region as a specific cluster, highlighting a distinct vertical structure of modelled micronekton in this latitudinal band. A second biome gathers dynamic regions around the globe: Kuroshio and Oyashio currents, Gulf Stream and North Atlantic current, Brazil and South Atlantic currents, Tasman Sea and Mauritania upwelling. Finally, other biomes follow tropical band patterns, with an east–west structuration identifying the subtropical gyres (see the Pacific and Atlantic Ocean poleward of 5° and equatorward of 20°, as well as the Indian ocean). These biomass biomes defined by micronekton vertical structures resembles the biophysical biomes (Fig. 3) yet with some notable differences (Fig. 7, Figure S3).

Overall, there is a good agreement between the two methods in latitudes poleward of 30° (e.g. North Pacific and Atlantic, the Kuroshio and Gulf stream regions, south-eastern Australia, and south America), and, to some extent, in equatorial and eastern boundary upwelling systems. Yet, some differences arise. The biomass clustering displays less latitudinally structured patterns than the biophysical clustering within the tropical band patterns, with an east–west structuration that better represents the subtropical gyres. Interestingly, Eastern Boundary Upwelling Systems (EBUS) are less visible in the biomass clustering in the Pacific.

4. Discussion

Our clustering derived from biophysical variables (i.e. temperature of the epipelagic layer, NPP and stratification) subdivided global ocean into specific biogeographic regions that represent our "biophysical biomes". Other studies (e.g. Proud et al., 2017; Sutton et al., 2017, Ariza et al. 2022) employed a comparable methodology, also using

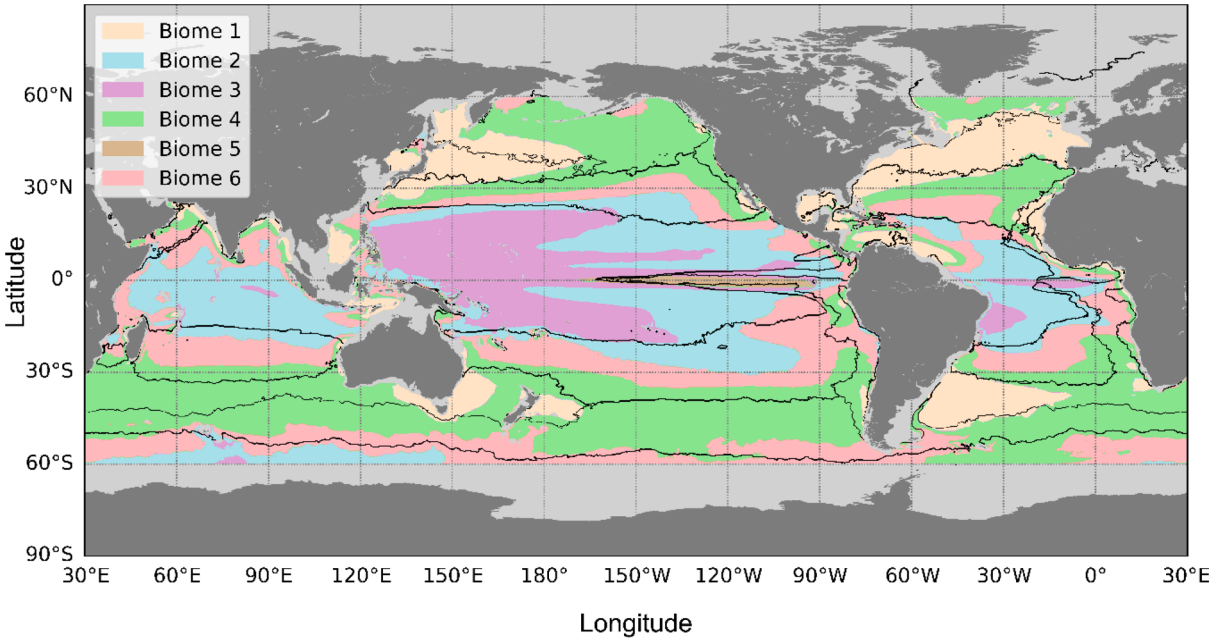

Fig. 7. Map of the biomass biomes obtained by clustering of two modelled micronekton functional groups ratios, respectively mesopelagic and migrant over the epipelagic simulated biomasses. These ratios are computed from SEAPODYM-LMTL outputs averaged on the 1998–2019 time period. The 6 biomass biomes are represented by a color defined in the legend. The foreground black lines represent biophysical biomes' contours (Fig. 3). Grey areas delimitate the domain where the depth of the water column is not sufficient to ensure the existence of the three pelagic layers of SEAPODYM-LMTL (see section 2.3.2.).

biophysical variables to derive biogeographic regions. In Sutton et al. (2017), the authors derived their classification from environmental drivers, complemented with some expert knowledge. In Proud et al. (2017), the authors similarly derived their classification from environmental drivers (e.g. surface primary production, temperature, and wind stress) using a clustering method, with the aim to model DSL characteristics (depth and backscattering intensity). In Ariza et al. (2022), the authors performed a clustering using acoustic data as a proxy of micronektonic biomasses, which they reconstructed from biophysical data (satellite-derived chlorophyll concentration, sea surface temperature and subsurface dissolved oxygen).

Despite the discrepancies that can be found in these three previously mentioned approaches from the existing literature, the resulting biogeographical regions closely mirror our biophysical biomes, most probably because the biophysical variables that they used, include the same fundamental information about temperature, biological productivity, and the degree of the water column mixing, than the variables used in our study. Indeed, in addition to the latitudinal banding that is particularly visible in the austral Ocean, more complex structures appear within these different regionalizations in the North Atlantic, in midlatitudes frontal zones (except in the south Pacific sector) and in upwelling regions (see Fig. 4 in Sutton et al., 2017, Fig. 3A in Proud et al., 2017, Fig. 2a in Ariza et al., 2022 and our Fig. 3).

But do these biogeographical regions computed from biophysical variables correspond to different characteristics of the ecosystem, as they are supposed to? Existing regionalizations based on environmental forcings such as Sutton et al. (2017) do not specifically focus on mid trophic levels. On the other hand, our biophysical clustering specifically targets micronekton as it is derived from environmental variables selected for this purpose. Proud et al. (2017) and Ariza et al. (2022) took different approaches: Ariza et al. (2022) applied clustering directly to acoustic data, while Proud et al. (2017) performed clustering on environmental drivers and then established the connection to DSL features. However, both studies converge in their common focus on acoustics. In the end, they have defined acoustic regions that display different clusters of mesopelagic biomasses (which comprises micronekton). Yet, biases in micronekton estimations from acoustic methods, together with the multilinear relationship used to compute the backscatter (proxy of the mesopelagic biomass) from the biophysical variables reduces the meaning of this result. To avoid the limitations of acoustic clusterings in representing micronekton biomass, we demonstrate the relationship between our biophysical clustering and micronekton biomass by (i) using modeled micronekton biomass to characterize the biophysical biomes, and (ii) comparing the biophysical clustering with the biomass clustering. This approach allows us to use biophysical data to extrapolate micronekton biomass.

Addressing point (i), our biophysical biomes are characterized by different modelled micronektonic biomasses (Fig. 5). Yet, this structuration cannot be as directly linked to the biophysical variables used for the clustering, as micronektonic biomasses are computed from SEAPODYM-LMTL which is not a linear model with respect to the biophysical forcings (Lehodey et al., 2010). In addition to the micronektonic biomasses, the vertical structure of the ecosystem displays differences between our biophysical biomes (Fig. 6). The differences and homogeneity between our biophysical biomes in terms of biomasses and vertical structures are significant and are mechanistically interpretable (e.g. higher vertical biomass gradient in highly stratified regions). Finally, a validation of the biophysical biomes' borders using acoustic data shows that our biomes' borders, which are computed from biophysical variables, correspond to a shift in the backscatter vertical structure. This advocates for good correspondence between biophysically derived biomes and observed ecoregions delineated by backscatter reorganization along the water column.

Addressing point (ii), another way to test the informative potential of the biophysical clustering is to use a clustering from the biomasses and the vertical structure of epi-, *meso*-pelagic and migrant micronektonic communities. Fig. 7 compares such biomass clustering from SEAPODYM-LMTL outputs to the biophysical biomes computed from the biophysical variables. Overall, the two methods show good agreement; however, some differences emerge, particularly in the representation of subtropical gyres and Eastern Boundary Upwelling Systems (EBUS).

Despite the differences in the clustering methods, a tentative comparison of Ariza et al. (2022) regionalization (Fig. 2a in Ariza et al., 2022), our biomasses clustering here (Fig. 7), and our biophysical clustering here as well (Fig. 3) can bring information on the degree of explainability of the micronektonic biomasses due to the biophysical variables, as well as some indication of the SEAPODYM-LMTL model performances. Indeed, some identified patterns are common to the three clustering methods as seen poleward of 30° (e.g. the Kuroshio and the Gulf Stream regions) or the equatorial upwelling regions. In those regions, specific characteristics of the micronekton ecosystem (from the data and the model) can thus be mostly explained by the biophysical variables. Yet in the tropical band, Ariza et al. (2022)'s clustering and our biomass clustering seem to identify the subtropical gyres while the biophysical clustering is more latitudinally structured. This suggests that, in those regions, the biophysical clustering misses the east–west differences in the vertical structuration of the micronektonic biomasses, which are present both in our SEAPODYM-LMTL outputs and in Ariza et al. (2022)'s acoustic dataset. Finally, some regions such as the offshore extension of the Pacific EBUS and the north Indian Ocean (Bay of Benghal and Arabian sea) are only present in Ariza et al. (2022)'s clustering. It suggests that neither the biophysical variables nor the biomasses from SEAPODYM-LMTL are able to capture this region-specific vertical structure of the micronekton. Indeed, those latter geographical patterns resemble those of well-known oxygen minimum zones (OMZs). Yet oxygen cannot really be accounted for by our biophysical variables nor by SEAPODYM-LMTL, which does not use oxygen to simulate micronektonic biomasses.

Lastly, our approach also stands out for its dynamical aspect. We produced a monthly time series of provinces that are consistent with micronekton biomass, representing a significant advancement in the field. This allows for the study of the spatio-temporal variability of micronekton and offers an accurate framework for gathering and analyzing field data, such as acoustic transects or trawl samples, within a coherent temporal scale.

5. Conclusion

The ambition to identify micronekton homogeneous functioning patterns leads to a global definition of 6 biophysical biomes (tropical, subtropical, eastern boundary coastal upwelling systems, oceanic mesotrophic systems, sub-polar biomes, polar). Three environmental variables widely used in mid-trophic ecosystem study are used in the biophysical biomes' definition: NPP, epipelagic layer temperature, and stratification of the mesopelagic ocean. We investigate the differences between these biophysical biomes in terms of simulated micronekton biomass using the SEAPODYM-LMTL model. This biophysical biomes' characterization displays biomes-specific relations between modelled biomass and the environmental variables used in the clustering. Indeed, we showed that biophysical biomes display a first structuration following the NPP, that is related with micronekton biomass according to biomes-specific relationships for the most productive biomes. NPP mostly discriminates the eastern boundary coastal upwelling systems and the oceanic mesotrophic systems in terms of total micronektonic density. Elsewhere, biophysical biomes display a second structuration following the epipelagic layer temperature, that is also related with micronekton biomass according to biomes-specific relationships but only for the less productive biomes (NPP values of about 20 mmolC/m^2/day). Temperature of the epipelagic layer discriminates tropical, subtropical and sub-polar biomes in terms of total micronektonic biomass. Epipelagic layer temperature also adds internal variability in productive biomes, displaying more discrepancies between the related provinces.

For the less productive biomes, conversely, related provinces show very similar characteristics in terms of modeled micronektonic biomass. This observation challenges the relevance of biophysical biomes' subdivision into provinces. Characterizing the biomes with micronekton functional groups ratios from modeled biomass, we observed that most stratified biophysical biomes show greater proportions of mesopelagic and migrant micronekton than epipelagic micronekton. Furthermore, the validation of biophysical biomes' boundaries with acoustic data demonstrates that the regionalization outlines more homogeneous areas in terms of acoustic backscatter vertical structure, diverging with the adjacent biome.

Finally, comparisons between biomes issued from different clustering methods allows us to discuss the rationale supporting the regionalization from biophysical variables (this kind of regionalization being often used in studies focusing on mid-trophic level). Indeed, the biophysical clustering displays an overall relatively good agreement with the biomass clustering, yet with patterns more latitudinally distributed. As we established the relationship between our biophysical biomes and micronekton biomass with this clusterings comparisons, together with the biophysical biomes' characterization with modeled micronekton biomass, we are able to use biophysical data to extrapolate micronekton biomass. Moreover, with the notable exception in the OMZs, Ariza et al. (2022)'s clustering from the acoustic vertical structure agrees relatively well with our biomass biomes computed from modelled micronektonic biomasses, which may increase confidence in the model producing vertical structure of the micronekton. However, the authors would suggest that, at a minimum, oxygen should be considered in any further clustering from biophysical variables and might also be a meaningful addition to the SEAPODYM-LMTL.

CRediT authorship contribution statement

S. Albernhe: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **T. Gorgues:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **P. Lehodey:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **C. Menkes:** Writing – review & editing, Methodology, Conceptualization. **O. Titaud:** Writing – review & editing, Validation, Software, Resources, Data curation, Conceptualization. **S. Magon De La Giclais:** Writing – review & editing, Visualization, Validation, Software, Resources, Formal analysis. **A. Conchon:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by CLS (Collecte Localisation Satellites) and the NECCTON project, which received support from Horizon Europe RIA under grant agreement No. 101081273. The authors wish to thank the CLS Fisheries Expertise team and the Datalab for their valuable contributions in sharing relevant and innovative knowledge and methodologies. We thank the providers of the acoustic data, which enabled us to conduct a comprehensive study of global ocean, and who have offered prompt support for any questions regarding their data. We also extend our thanks to the two anonymous reviewers and the editors of Progress in Oceanography for their insightful feedback on an earlier version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pocean.2024.103370.

Data availability

Data will be made available on request.

References

Aksnes, D.L., Røstad, A., Kaartvedt, S., Martinez, U., Duarte, C.M., Irigoien, X., 2017. Light penetration structures the deep acoustic scattering layers in the global ocean. Science Advances 3 (5), e1602468.
Allain, V., Fernandez, E., Hoyle, S.D., Caillot, S., Jurado-Molina, J., Andrefouet, S., Nicol, S.J., 2012. Interaction between coastal and oceanic ecosystems of the Western and Central Pacific Ocean through predator-prey relationship studies. PLoS One 7 (5), e36701.
Allken, V., Rosen, S., Handegard, N.O., Malde, K., 2021. A deep learning-based method to identify and count pelagic and mesopelagic fishes from trawl camera images. ICES Journal of Marine Science 78 (10), 3780–3792.
Anderson, T.R., Martin, A.P., Lampitt, R.S., Trueman, C.N., Henson, S.A., Mayor, D.J., 2019. Quantifying carbon fluxes from primary production to mesopelagic fish using a simple food web model. ICES Journal of Marine Science 76 (3), 690–701.
Ariza, A., Lengaigne, M., Menkes, C., Lebourges-Dhaussy, A., Receveur, A., Gorgues, T., Bertrand, A., 2022. Global decline of pelagic fauna in a warmer ocean. Nature Climate Change 12 (10), 928–934.
Aumont, O., Éthé, C., Tagliabue, A., Bopp, L., Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. Geoscientific Model Development Discussions 8 (2), 1375–1509.
Barbin, L., Lebourges-Dhaussy, A., Allain, V., Receveur, A., Lehodey, P., Habasque, J., Menkes, C., 2024. Comparative analysis of day and night micronekton abundance estimates in west Pacific between acoustic and trawl surveys. Deep Sea Research Part i: Oceanographic Research Papers 204, 104221.
Béhagle, N., Cotté, C., Ryan, T.E., Gauthier, O., Roudaut, G., Brehmer, P., Cherel, Y., 2016. Acoustic micronektonic distribution is structured by macroscale oceanographic processes across 20–50 S latitudes in the South-Western Indian Ocean. Deep Sea Research Part i: Oceanographic Research Papers 110, 20–32.
Behrenfeld, M.J., Falkowski, P.G., 1997. A consumer's guide to phytoplankton primary productivity models. Limnology and Oceanography 42 (7), 1479–1491.
Bell, J.D., Allain, V., Allison, E.H., Andréfouët, S., Andrew, N.L., Batty, M.J., Williams, P., 2015. Diversifying the use of tuna to improve food security and public health in Pacific Island countries and territories. Marine Policy 51, 584–591.
Benoit-Bird, K.J., 2009. The effects of scattering-layer composition, animal size, and numerical density on the frequency response of volume backscatter. ICES Journal of Marine Science 66 (3), 582–593.
Benoit-Bird, K.J., Au, W.W., Wisdoma, D.W., 2009. Nocturnal light and lunar cycle effects on diel migration of micronekton. Limnology and Oceanography 54 (5), 1789–1800.
Benoit-Bird, K.J., Lawson, G.L., 2016. Ecological insights from pelagic habitats acquired using active acoustic techniques. Annual Review of Marine Science 8, 463–490.
Benoit-Bird, K.J., Waluk, C.M., Martin, E.J., Reisenbichler, K.R., Sherlock, R.E., McGill, P.R., Robison, B.H., 2023. Schrödinger's fish: Examining the robotic observer effect on pelagic animals. Limnology and Oceanography: Methods 21 (9), 563–580.
Bianchi, D., Stock, C., Galbraith, E.D., Sarmiento, J.L., 2013. Diel vertical migration: Ecological controls and impacts on the biological pump in a one-dimensional ocean model. Global Biogeochemical Cycles 27 (2), 478–491.
Blanchard, J.L., Heneghan, R.F., Everett, J.D., Trebilco, R., Richardson, A.J., 2017. From bacteria to whales: using functional size spectra to model marine ecosystems. Trends in Ecology & Evolution 32 (3), 174–186.
Conchon, A., 2016. Modélisation du zooplancton et du micronecton marins. Université de La Rochelle. Doctoral dissertation.
Davison, P.C., Koslow, J.A., Kloser, R.J., 2015. Acoustic biomass estimation of mesopelagic fish: backscattering from individuals, populations, and communities. ICES Journal of Marine Science 72 (5), 1413–1424.
De Robertis, A., Higginbottom, I., 2007. A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise. ICES Journal of Marine Science 64 (6), 1282–1291.
Dupont, L., Le Mézo, P., Aumont, O., Bopp, L., Clerc, C., Éthé, C., Maury, O., 2023. High trophic level feedbacks on global ocean carbon uptake and marine ecosystem dynamics under climate change. Global Change Biology 29 (6), 1545–1556.
Eppley, R.W., 1972. Temperature and phytoplankton growth in the sea. Fishery Bulletin 70 (4), 1063.
Gillooly, J.F., Charnov, E.L., West, G.B., Savage, V.M., Brown, J.H., 2002. Effects of size and temperature on developmental time. Nature 417 (6884), 70–73.
Gorgues, T., Aumont, O., Memery, L., 2019. Simulated changes in the particulate carbon export efficiency due to diel vertical migration of zooplankton in the North Atlantic. Geophysical Research Letters 46 (10), 5387–5395.
Graham, W.M., Martin, D.L., Martin, J.C., 2003. In situ quantification and analysis of large jellyfish using a novel video profiler. Marine Ecology Progress Series 254, 129–140.
Gregg, W.W., Casey, N.W., 2007. Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. Remote Sensing of Environment 111 (1), 25–35.

Guinehut, S., Dhomps, A.L., Larnicol, G., Le Traon, P.Y., 2012. High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. Ocean Science 8 (5), 845–857.
Habasque, J., Bourlès, B., Bertrand, A., Lebourges-Dhaussy, A., Grelet, J., Rousselot, P., 2024. *French PIRATA cruises: acoustic data*.
Haris, K., Kloser, R.J., Ryan, T.E., Downie, R.A., Keith, G., Nau, A.W., 2021. Sounding out life in the deep using acoustic data from ships of opportunity. Scientific Data 8 (1), 23.
Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. springer, New York, pp. 1–758.
Hatton, I.A., Heneghan, R.F., Bar-On, Y.M., Galbraith, E.D., 2021. The global ocean size spectrum from bacteria to whales. Science Advances 7 (46), eabh3732.
Hill Cruz, M., Kriest, I., Getzlaff, J., 2023. Diving deeper: Mesopelagic fish biomass estimates comparison using two different models. Frontiers in Marine Science 10, 1121569.
Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24 (6), 417.
Huntley, M.E., Lopez, M.D., 1992. Temperature-dependent production of marine copepods: a global synthesis. The American Naturalist 140 (2), 201–242.
Irigoien, X., Klevjer, T.A., Røstad, A., Martinez, U., Boyra, G., Acuña, J.L., Kaartvedt, S., 2014. Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nature Communications 5 (1), 3271.
Iverson, R.L., 1990. Control of marine fish production. Limnology and Oceanography 35 (7), 1593–1604.
Jech, J.M., Horne, J.K., Chu, D., Demer, D.A., Francis, D.T., Gorska, N., Sawada, K., 2015. Comparisons among ten models of acoustic backscattering used in aquatic ecosystem research. The Journal of the Acoustical Society of America 138 (6), 3742–3764.
Kaartvedt, S., Staby, A., Aksnes, D.L., 2012. Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass. Marine Ecology Progress Series 456, 1–6.
Kloser, R.J., Ryan, T.E., Young, J.W., Lewis, M.E., 2009. Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges. ICES Journal of Marine Science 66 (6), 998–1006.
Kwong, L.E., Pakhomov, E.A., Suntsov, A.V., Seki, M.P., Brodeur, R.D., Pakhomova, L.G., Domokos, R., 2018. An intercomparison of the taxonomic and size composition of tropical macrozooplankton and micronekton collected using three sampling gears. Deep Sea Research Part i: Oceanographic Research Papers 135, 34–45.
Lansdell, M., Young, J., 2007. Pelagic cephalopods from eastern Australia: species composition, horizontal and vertical distribution determined from the diets of pelagic fishes. Reviews in Fish Biology and Fisheries 17, 125–138.
Lehodey, P., 2001. The pelagic ecosystem of the tropical Pacific Ocean: dynamic spatial modelling and biological consequences of ENSO. Progress in Oceanography 49 (1–4), 439–468.
Lehodey, P., Murtugudde, R., Senina, I., 2010. Bridging the gap from ocean models to population dynamics of large marine predators: a model of mid-trophic functional groups. Progress in Oceanography 84 (1–2), 69–84.
Lehodey, P., Conchon, A., Senina, I., Domokos, R., Calmettes, B., Jouanno, J., Kloser, R., 2015. Optimization of a micronekton model with acoustic data. ICES Journal of Marine Science 72 (5), 1399–1412.
Lellouche, J.M., Bourdalle-Badie, R., Greiner, E., Garric, G., Melet, A., Bricaud, C., Drevillon, M., 2021. The Copernicus global 1/12° oceanic and sea ice reanalysis. In *EGU General Assembly Conference Abstracts*.
Lloyd, S., 1957. Least squares quantization in PCM. *IEEE transactions on information theory* 28 (2), 129–137, 1982.
Madec, G., & the NEMO team (2008). NEMO ocean engine. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL). (No. 27). France.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statistics 18, 50–60.
Maury, O., 2010. An overview of APECOSM, a spatialized mass balanced "Apex Predators ECOSystem Model" to study physiologically structured tuna population dynamics in their ecosystem. Progress in Oceanography 84 (1–2), 113–117.
McCluney, J.K., Anderson, C.M., Anderson, J.L., 2019. The fishery performance indicators for global tuna fisheries. Nature Communications 10 (1), 1641.
McGehee, D.E., O'Driscoll, R.L., Traykovski, L.M., 1998. Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz. Deep Sea Research Part II: Topical Studies in Oceanography 45 (7), 1273–1294.
Morel, A., 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. Progress in Oceanography 26 (3), 263–306.
Mulet, S., Rio, M.H., Mignot, A., Guinehut, S., Morrow, R., 2012. A new estimate of the global 3D geostrophic ocean circulation based on satellite data and in-situ measurements. Deep Sea Research Part II: Topical Studies in Oceanography 77, 70–81.
Pakhomov, E. A., Suntsov, A. V., Seki, M. P., Brodeur, R. D., Domokos, R., Pakhomova, L. G., & Owen, K. R. (2010). 2 First Micronekton Inter-calibration Experiment, MIE-1. *PICES SCIENTIFIC REPORT No. 38 2010*, 3.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É., 2011. Scikit-learn: machine learning in python. The Journal of Machine Learning Research 12, 2825–2830.
Petrik, C.M., Stock, C.A., Andersen, K.H., van Denderen, P.D., Watson, J.R., 2019. Bottom-up drivers of global patterns of demersal, forage, and pelagic fishes. Progress in Oceanography 176, 102124.
Pinti, J., DeVries, T., Norin, T., Serra-Pompei, C., Proud, R., Siegel, D.A., Visser, A.W., 2021. Metazoans, migrations, and the ocean's biological carbon pump. BioRxiv.
Polovina, J.J., Marten, G.G., 1982. A comparative study of fish yields from various tropical ecosystems,. In: Pauly, D., Murphy, G.I. (Eds.), Theory and Management of Tropical Fisheries. ICLARM Conf. Proc. 9, 255–286.
Proud, R., Cox, M.J., Brierley, A.S., 2017. Biogeography of the global ocean's mesopelagic zone. Current Biology 27 (1), 113–119.
Proud, R., Handegard, N.O., Kloser, R.J., Cox, M.J., Brierley, A.S., 2019. From siphonophores to deep scattering layers: uncertainty ranges for the estimation of global mesopelagic fish biomass. ICES Journal of Marine Science 76 (3), 718–733.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.
Scoulding, B., Gastauer, S., MacLennan, D.N., Fässler, S.M., Copland, P., Fernandes, P.G., 2017. Effects of variable mean target strength on estimates of abundance: the case of Atlantic mackerel (Scomber scombrus). ICES Journal of Marine Science 74 (3), 822–831.
Spalding, M.D., Agostini, V.N., Rice, J., Grant, S.M., 2012. Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. Ocean & Coastal Management 60, 19–30.
Sutton, T.T., Clark, M.R., Dunn, D.C., Halpin, P.N., Rogers, A.D., Guinotte, J., Heino, M., 2017. A global biogeographic classification of the mesopelagic zone. *Deep Sea Research Part I:* Oceanographic Research Papers 126, 85–102.
Walters, C., Christensen, V., Pauly, D., 1997. Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. Rev. Fish Biol. Fish. 7, 139–172.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometries 1, 80–83.