Demonstrating the relevance of spatial-functional statistical analysis in marine ecological studies: the case of environmental variations in micronektonic layers

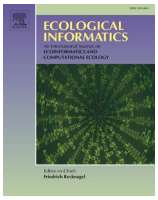
Yoba Kande, Ndague Diogoul, Patrice Brehmer, Sophie Dabo-Niang, Papa Ngom, Yannick Perrot

Ecological Informatics 81 (2024) 102547

Contents lists available at ScienceDirect

Ecological Informatics

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



Demonstrating the relevance of spatial-functional statistical analysis in marine ecological studies: The case of environmental variations in micronektonic layers

Yoba Kande [a,b,c,e,*], Ndague Diogoul [b,c], Patrice Brehmer [c], Sophie Dabo-Niang [d], Papa Ngom [e], Yannick Perrot [f]

[a] Université de Lille, CNRS, UMR 8524-Laboratoire Paul Painleve, Lille 59000, France
[b] Institut Sénégalais de Recherche Agricole, ISRA, Centre de Recherche Océanographique de Dakar Thiaroye, CRODT, Pôle de Recherche de Hann, Dakar 2241, Senegal
[c] IRD, Univ Brest, CNRS, Ifremer, LEMAR, CSRP, SRFC, Dakar 1386, Senegal
[d] Université de Lille, CNRS, UMR 8524-Laboratoire Paul Painleve, INRIA-MODAL, Lille 59000, France
[e] Université Cheikh Anta Diop de Dakar, UCAD, LMA, Dakar, Senegal
[f] IRD, Univ Brest, CNRS, Ifremer, Lemar, Délégation Régionale IRD France Ouest, Plouzané 29280, France

ARTICLE INFO

Keywords:
Clustering
Functional data analysis
Functional generalized spectral additive model
General additive model
Principal component analysis
Sound scattering layers
Spatial regression

ABSTRACT

In this study, we conducted an analysis of a multifrequency acoustics dataset acquired from scientific echosounders in the West African water. Our objective was to explore the spatial arrangement of marine organism aggregations. We investigated various attributes of these intricate biological entities, such as thickness, relative density, and depth, in relation to their surroundings. These environmental conditions were represented at a fine scale using a towed multiparameter system. This study is closely intertwined with two key domains: Fisheries acoustics techniques and functional data analysis. Fisheries acoustics techniques facilitate the collection of high-resolution spatial and temporal data concerning marine organisms at various depths and spatial scales, all without causing any disturbance. On the other hand, spatial-functional data analysis is a statistical approach for examining data characterised by functional attributes distributed across a spatial domain. This analysis encompasses dimension reduction techniques, as well as supervised and unsupervised methods, which take into consideration spatial dependencies within extensive datasets.

We began by applying multivariate statistical techniques and subsequently employed Functional Data Analysis (FDA). In the modeling section, we introduced the spatial dimension with the spatial coordinates as covariates in the General Additive Model (GAM) and Functional Generalized Spectral Additive Model (FGSAM) models, aiming to underscore its relevance in those contexts. In an exploratory phase, Multivariate Functional Principal Component Analysis provided detailed insights into the variations of parameters at different depths, a capability not offered by traditional Principal Component Analysis. When it came to regression tasks, we explored the interactions between descriptors of Sound Scattering Layers and key environmental variables, both with and without considering spatial dimensions. Our findings revealed significant distinctions between northern and southern Sound Scattering Layers, as well as between coastal and high-sea regions. The use of the spatial locations enhanced the performance of GAM and FGSAM, particularly in the case of salinity, reflecting the influence of water mixing and seawater temperature. The multifaceted effects of environmental variations on Sound Scattering Layers underscore the importance of spatial-functional statistical analysis in ecological studies involving complex, spatially functional objects. Beyond the scope of this specific case study, the application of functional data analysis shows promise for a wide array of ecological studies dealing with extensive spatial datasets.

* Corresponding author at: Université de Lille, CNRS, UMR 8524-Laboratoire Paul Painleve, Lille 59000, France.
E-mail address: yoba.kande.etu@univ-lille.fr (Y. Kande).

https://doi.org/10.1016/j.ecoinf.2024.102547
Received 7 August 2023; Received in revised form 26 February 2024; Accepted 26 February 2024
Available online 5 March 2024

1. Introduction

Advances in fisheries acoustics allow to observe various targets, over a wide range of depths and spatial scales (Brehmer et al., 2019; Simmonds and MacLennan, 2005). Among other underwater targets, the Sound Scattering Layers (SSL) represent a key biomass in the world ocean (Mair et al., 2005; Proud et al., 2019). These SSL are mainly composed of zooplankton and micronekton. Zooplanktonic and micronektonic species provide the main trophic link between primary producers and higher trophic levels. A large amount of energy passes through zooplankton and micronekton (Steele et al., 2007). One of the characteristics of macrozooplanktonic and micronekton organisms is their Diel Vertical Migrations (DVM) (Klevjer et al., 2016). This behavior is a globally observed phenomenon (Bianchi et al., 2013; Bianchi and Mislan, 2016) and is primarily attributed to the need to evade visual predators during the day while engaging in surface feeding at night (Bianchi et al., 2013; Haney, 1988; Lehodey et al., 2015). This vertical movement and associated metabolic activities significantly impact the carbon cycle (Bianchi et al., 2013), making zooplankton and micronekton indispensable actors in marine ecosystems. Furthermore, SSL have been correlated with an array of environmental factors that profoundly influence their spatial distribution. These factors encompass temperature, dissolved oxygen levels, primary production, light intensity, density, and wind-induced mixing, as demonstrated by Hays et al. (2005), Bianchi et al. (2013), Proud et al. (2015, 2017), Klevjer et al. (2016), and Aksnes et al. (2017). Considering the challenges posed by global change and the decline in fisheries resources, it becomes increasingly imperative to gain a comprehensive understanding of these organisms distribution in relation to environmental conditions.

Usually, the studies conducted on zooplanktonic communities did not take into account the spatial dimension of all the physical and biomass measurements collected with different sampling techniques i.e. fisheries acoustics (Diogoul et al., 2021) or using nets (Blanluet et al., 2019). The high dimensional structure of environmental variables (seawater temperature, salinity, fluorescence, and turbidity) combined with multifrequency fisheries acoustics measurements is a rich source of information. Analyzing such massive spatial data requires sophisticated techniques as proposed by Functional Data Analysis (FDA) (Ariza et al., 2022b), able to take into account the functional and spatial nature of the data. FDA transforms high dimensional data within a continuum into functional data, i.e. data objects, such as curves, shapes, images, or more complex mathematical objects, thought of as smooth realizations of a stochastic process mainly valued in a Hilbert space. Statistical methods for functional data have received a lot of attention from the scientific community (Ramsay and Silverman, 2005; Silverman and Ramsay, 2002) over the last decade. Li et al. (2022) and Koner and Staicu (2023) reviewed recently some of the fundamental concepts of the FDA, their recent advances, and their impact on practical cases. FDA is an alternative to the use of diverse and heterogeneous datasets (Pauthenet et al., 2017, 2019) of functional, shape or more complex structure. Here we studied the relationships between the SSL spatial structuration and the pelagic environment to take advantage of all the information available in the data. Previous work by Diogoul et al. (2020), among others, has been done on aggregated features from the original data. Classical multivariate biostatistical analyses such as Principal Component Analysis (PCA), clustering, and analysis of covariance (Zar, 2010) were performed on these aggregated data. In this context, our primary goal is to conduct a thorough and comprehensive analysis of the aggregated data. After initially conducting classical multivariate analysis, we adopt two distinct approaches. First, we incorporate the spatial dimension of the data into our statistical analysis, and second, we utilize both the spatial and functional characteristics of the data. These methods allow us to evaluate the significance of each approach and to determine the additional insights that can be obtained through functional data analysis.

There is an increasing interest in ecology concerning the application of FDA, as evident from several recent works. Henderson (2006) utilized FDA to explore trends in ecological variables of South Queensland Water's ambient dam water quality monitoring program. Embling et al. (2012) investigated the relationship between the behavior of North Sea prey species and a vulnerable surface-foraging predator, analyzing fine scale tidally driven changes in biophysical characteristics using Functional Principal Components Analysis (FPCA). FDA clustering techniques have been employed to classify temperature and salinity profiles in the ocean in various studies (Nerini et al., 2010; Pauthenet, 2018; Pauthenet et al., 2017; Pauthenet et al., 2019; Reyes et al., 2015). Bayle et al. (2015) applied a functional linear model to predict chlorophyll-a concentration profile from light data. Gong et al. (2015) used functional principal component analysis to study a surface water temperature dataset from Lake Victoria. Sierra et al. (2017) investigated FDA methods to examine particle-size distributions in a beach/shallow marine sedimentary environment in Gijón Bay (NW Spain). Acar-Denizli et al. (2018) applied a functional linear regression model to remote sensing data, enabling them to predict total suspended solids concentration in the coastal zone of the Guadalquivir estuary. Tarro-Saavedra et al. (2020) classified the seabed in coastal environments by employing FDA approaches on acoustic curves. Godard (2021) employed FDA to study high-frequency physical, biological, and behavioral data from electronic recorders deployed on marine predators, particularly southern elephant seals. Nonparametric functional spatial regression has been explored for mapping the presence of demersal coastal fish from Senegal in studies by Ndiaye et al. (2022, 2020). Yarger et al. (2022) utilized a spatio-temporal functional kriging methodology to predict temperature and salinity functions with depth from the Argo dataset at a fixed location, while Korte-Stapff et al., 2022 applied a multivariate functional-data mixture model to Argonaut oceanographic data in the Southern Ocean to predict oxygen concentration. Assunção et al. (2020) characterised the thermohaline structure in the tropical southwest Atlantic using hydrographic profiles, FPCA and functional hierarchical clustering. Ariza et al. (2022a) explored the variation of acoustic backscatter using functional PCA, and further applied FDA to classify the acoustic seascapes of the southwest tropical Atlantic into biogeographical regions. Ariza et al. (2022b) further investigated the variation of acoustic backscatter using functional PCA. In this paper, we assume that the environmental data correspond to functions in space (depending on the depth points), expressed by a basic system (bspline, Fourier, polynomial). By using FDA methods such as functional PCA, the dimension of the basic system is reduced and a better reconstruction of the variables is achieved. We also performed clustering based on the PCA scores to identify homogeneous profile groups. To the best of our knowledge, there is no functional methodology for functional multivariate spatial PCA and clustering in this framework. We apply these methods from the univariate spatial and multivariate non-spatial functional literature to our original dataset (Happ and Greven, 2018). By reconstructing the data using FDA techniques, we can effectively analyze, at a fine scale, the relationship between environmental data and marine organism aggregation. Specifically, we employ regression methods for functional data, such as functional generalized spectral additive models, to gain insights into the intricate ecological patterns and associations. Our findings convincingly illustrate the significance and advantages of FDA in our ecological studies when compared to classical biostatistics methods. Through this research, we aim to shed light on the potential of spatial-functional statistical analysis and emphasize its relevance in addressing ecological questions, particularly pertaining to the effects of environmental variation on micronektonic layers.

2. Materials and method

2.1. Materials

The Hydroacoustics AWA sea survey was carried out in the West African waters. The survey was conducted with the research vessel (R/

V) Thalassa (Ifremer) during the upwelling season (Tiedemann et al., 2017; Tiedemann and Brehmer, 2017) from February 24 to March 14, 2014.We used a remotely operated towed vehicle a Scanfish (Brown et al., 1996; Farrell et al., 2012) operated from the R/V Thalassa and hull-mounted multifrequency echosounders (Korneliussen and Ona, 2002). Three contrasted radials (ID1, 2 and 2′) were used in this study (Fig. 2). ID1 was carried out over the continental shelf of the Grande Côte in the north of Senegal (Balde et al., 2019) during the night, sunrise, and day. The radial 2 carry out in southern Senegal was divided into two parts: ID2' in the shallow continental shelf (0–100 m; sampled during the night, sunrise and day) and ID2 in the high sea (bottom depth > 500 m; sampled during sunset and night). The Scanfish was towed continuously from the R/V between the surface to 100 m depth, while the R/V insonified the water column continuously at four frequencies in continuous wave mode (CW) (Diogoul et al., 2020).

- Fisheries acoustics data were recorded continuously using multi-frequency echo sounders (operating at 38, 70, 120 and 200 kHz) at a pulse length of 1 ms and with a Time Varied Gain (TVG) function set in 20 log R (R the range in m). Two acoustic variables were measured: (i) the volume backscattering strength noted S_v (dB re 1 m^{-1}) (MacLennan et al., 2002) is a measure of the amount of sound reflected towards the echosounder from a target in the water column. It quantifies the scattering characteristics of biological organisms or other scatterers in the acoustic field. It provides information about the density and distribution of scatterers within the water column; (ii) Nautical Area backscattering coefficient s_A (m^2 nmi^{-2}) translates the acoustic energy received by the echosounder for one nautical mile, hereafter relative density.
- The S_v are represented in two dimensions: vertically, i.e., depths, and horizontally per Elementary Sampling Unit (ESU) in distance (here 0.1 nmi) travelled by the the R/V (Fig. 3). Matecho allows visual observation of the SSL and extraction of their characteristics (Perrot et al., 2018). We have used three SSL descriptors (depth in meter, thickness in meter (maximum depth (d_{max}) − minimum depth (d_{min})), and relative SSL density (mean s_A) using Matecho (Perrot et al., 2018). The bottom depth is obtained from the bottom line extraction at 38 kHz using Matecho.
- Environmental data, i.e., seawater temperature (Temp in °C), fluorescence (Fluo in ml l^{-1}; proxy of chlorophyll-a concentration), turbidity (TU in NTU) and salinity (Sal in psu), were acquired every second using the Scanfish along the path of the vessel (Fig. 3). These parameters are averaged for each echointegration cell (Perrot et al., 2018), allowing them to be matched at a fine scale (vs CTD probe collected during oceanographic surveys) with the SSL descriptors (Mouget et al., 2022). The fluorescence allows to measure the concentration of Chlorophyll-a, a proxy of marine primary productivity. The georeferenced positions (latitude (LatEsu) and longitude (LonEsu), in decimal degrees) were obtained from the onboard Global Positioning System. The environmental variables were matched with the acoustic variables using the matlab code available in Section 4 of the supplementary material. The data, which has a size of 2.61 GB, is stored in MATLAB format.

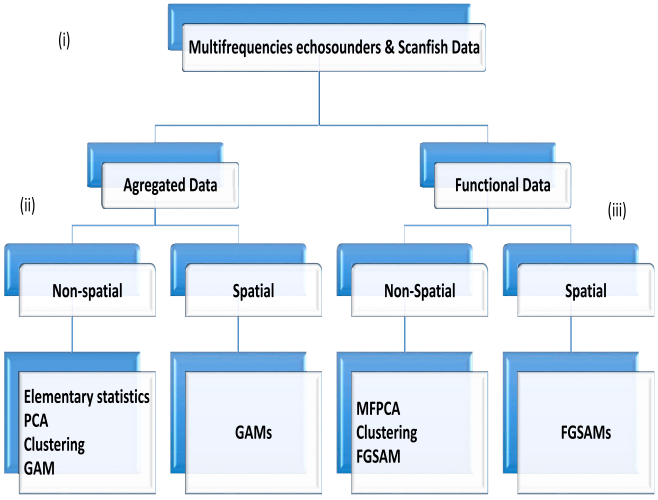


Fig. 1. Flow chart depicting the study methodology. The research unfolded in three key steps: (i) Raw data extraction from the echosounders and the Scan-Fish, followed by the computation of Sound Scattered Layers'SSL' characteristics and the integration of all data; (ii) Comparative analysis of SSL characteristics against diel periods, involving classical PCA (principal component analysis), clustering, and GAM (Generalized Additive Model) without spatial dimensions, then with spatial dimensions. This step showcases the relevance of Spatial-Functional Statistical Analysis in marine ecological studies, exemplified through environmental variations in SSL; (iii) Application of a Functional Analysis, encompassing Multivariate Functional Principal Component Analysis (MFPCA) and clustering. Subsequently, the relationship between SSL and physical parameters is modeled using FGSAM (Functional Generalized Spectral Analysis Model) initially without spatial dimensions and then incorporating spatial dimensions. The analysis and modeling explicitly explore the relationship between SSL descriptors (thickness, depth, density), treating them as observations of a real-valued spatial process.

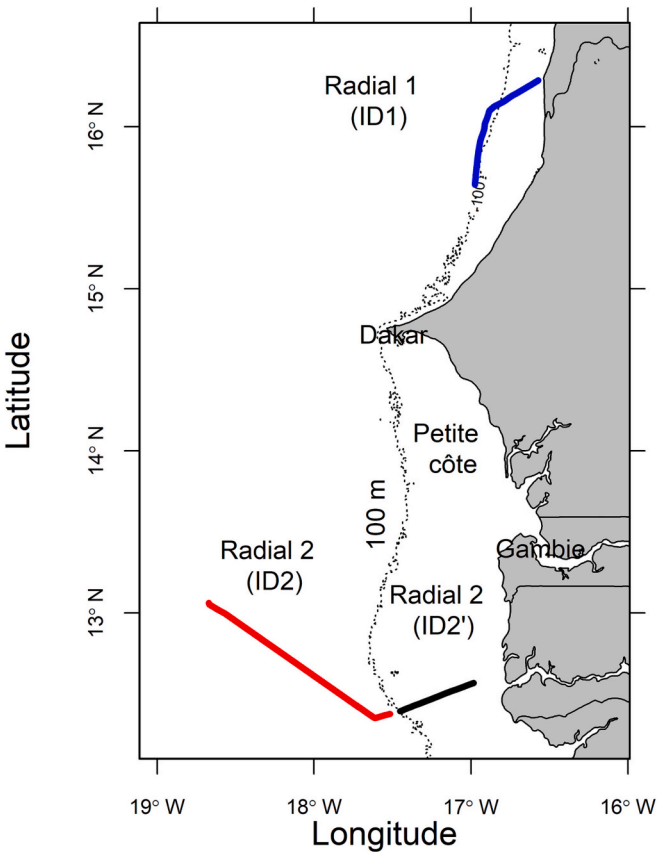


Fig. 2. Map with the 100 m isobath of Senegalese coastline (represented by a curve), (West Africa). The vessel has covered three radials collecting simultaneously acoustics and environmental variables at high resolution (sea survey AWA). The radial (dotted lines) 1 was carried out over the continental shelf in northern Senegal (in blue colour) and radial 2 in southern Senegal split into two parts, on-shore (continental shelf: ID2' in black colour) and off-shore (high sea: ID2 in red colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Methods

The research unfolded in three key steps:

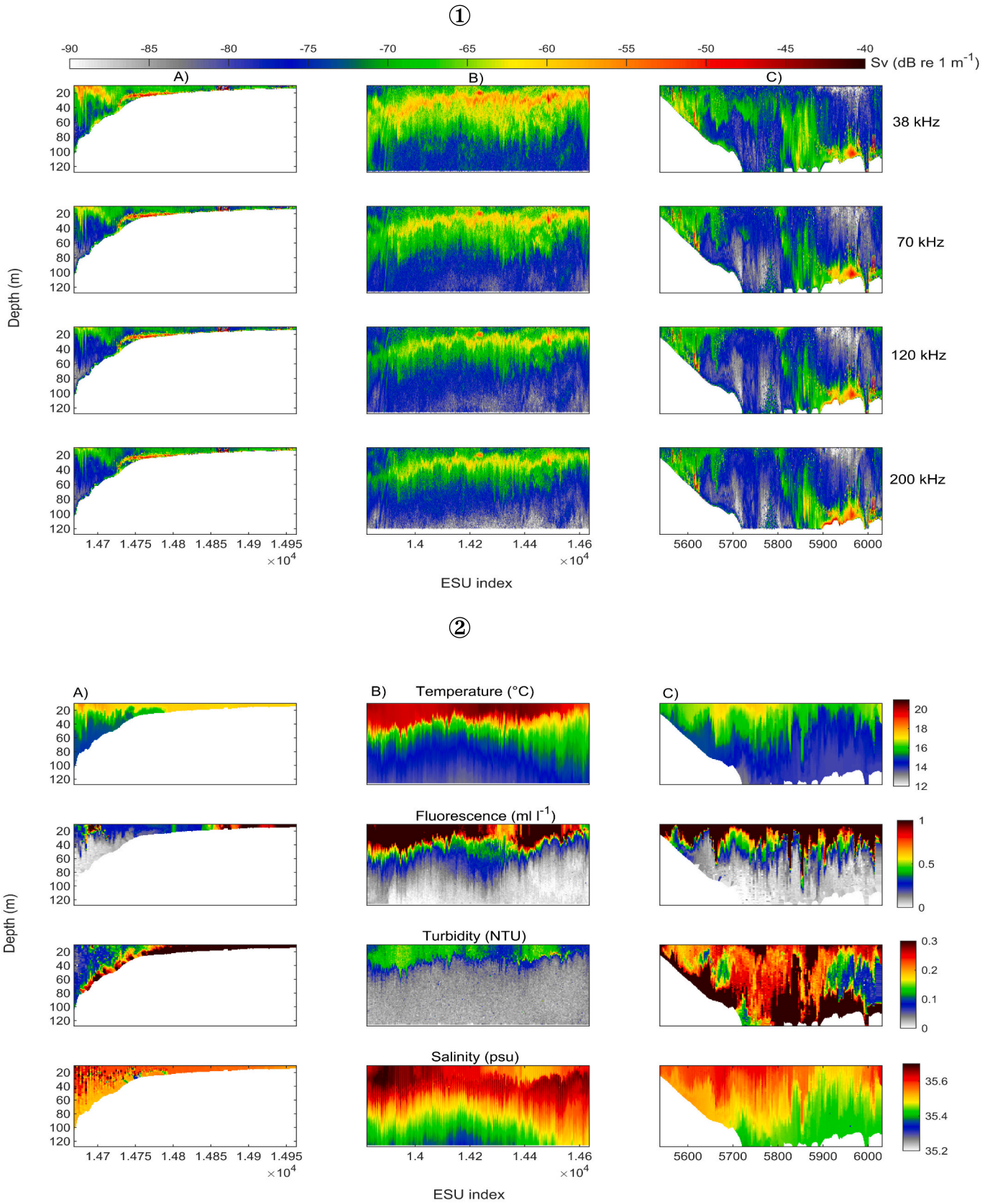


Fig. 3. (1) Echogram representing the acoustic intensities (in Sv backscatter coefficient) reverberated by the aggregated marine organisms structured in layer (i.e. here sound scattering layer) as detected by echosounders at four frequencies (38, 70, 120 and 200 kHz) in three dimensions (2D + time). (2) The contour plots of key physicochemical parameters of seawater temperature, salinity, turbidity and fluorescence over the three acoustic radials studied. A): southern continental shelf; B): southern high sea and C): northern continental shelf. Data obtained from AWA sea survey on-board FRV Thalassa of Senegal coastline.

(i) Raw data extraction from the echosounders and the ScanFish, followed by the computation of Sound Scattered Layers'SSL' characteristics and the integration of all data.
(ii) Comparative analysis of SSL characteristics against diel periods, involving classical PCA (principal component analysis), clustering, and GAM (Generalized Additive Model) without spatial dimensions, then with spatial dimensions. This step showcases the relevance of Spatial-Functional Statistical Analysis in marine ecological studies, exemplified through environmental variations in SSL.
(iii) Application of a Functional Analysis, encompassing Multivariate Functional Principal Component Analysis (MFPCA) and clustering. Subsequently, the relationship between SSL and physical parameters is modeled using FGSAM (Functional Generalized Spectral Analysis Model) initially without spatial dimensions and then incorporating spatial dimensions. The analysis and modeling explicitly explore the relationship between SSL descriptors (thickness, depth, density), treating them as observations of a real-valued spatial process.

Namely, we analyzed and modeled (see Fig. 1) the relationship between the SSL descriptors (thickness, depth, density) (considered as observations of a real-valued spatial process $\{Y_s\}_{s\in\mathbb{R}^2}$ at the sites $s_1,\ldots,s_n$) and the characteristics of the pelagic environment *i.e* the physical parameters (considered as observations of a functional spatial processes (see Fig. S2) $\{X_s(t),\ t\in[d_{s,min},d_{s,max}]\}_{s\in\mathbb{R}^2}$, where t is the depth and $s=(s^1,s^2)$ a spatial site where the data was collected). The peculiarity of the proposed methodology lies in the assumption that the raw measurements $x_{s,j,t}$ are noisy observations of a random variable $X=(X^1,\ldots,X^p)$ valued in a multivariate functional space. Namely $X^j\in L^2(\mathcal{T}_j), j=1,\ldots,p$ where $L^2(\mathcal{T}_j)$ is the space of squared integrable functions on $\mathcal{T}_j=[\mathcal{T}_{j,min},\mathcal{T}_{j,max}]$. In other words $x_{s,j,t}=\mu_j(t)+X_s^j(t)+\varepsilon_{s,j,t}$, with $\mu_j(.)$ the mean function of the parameter j. The functional variable X_s^j is centred and square integrable, while $\varepsilon_{s,j,t}$ are centred (real-valued) Gaussian i.i.d variables of finite variance.

2.2.1. Aggregated data modeling

To compare with the benchmark work Diogoul et al. (2020), we first modeled the layer descriptors by the environmental data, aggregating them through the mean. The aggregated variables are:

$$X_s^j=\frac{1}{D_s}\sum_{p=1}^{D_s}X_s^j(t_p), \tag{1}$$

where D_s is the number of depth where data are available and $s\in(s_1,\ldots,s_n)$.

2.2.1.1. Non-spatial analysis. We performed a principal component analysis (Jolliffe, 2005) followed by a hierarchical clustering (Rokach and Maimon, 2005) in order to identify homogeneous groups that are similar in terms of all environmental parameters. We were also interested in the implementation of the Generalized Additive Model (GAM) with the aggregated covariates X_s^j.

The GAM (Hastie and Tibshirani, 1990) postulates that

$$g(Y_s)=\sum_{j=1}^{p}f_j(X_s^j)+\varepsilon_s, \tag{2}$$

where g is the link function, f_j is in a family of functions (linear, splines, etc.), ε_s being the centered i.i.d (independent and identically distributed) error terms with finite variance. The estimation of f_j is based on splines nonparametric smoothing technique. This representation is chosen to minimize the root mean square error (RMSE) and mean absolute error (MAE).

2.2.1.2. Spatial analysis. We have introduced geographical coordinates as spatial covariates ($s=(s^1,s^2)$) in the GAM model:

$$g(Y_s)=\sum_{j=1}^{p}f_j(X_s^j)+g(s)+\varepsilon_s, \tag{3}$$

where $s=(s^1,s^2)$ is the vector of patial coordinates and g is like f_j in a family of functions (linear, splines,etc.).

2.2.2. Functional data modeling

2.2.2.1. Non-spatial analysis. Instead of aggregating the data, we use in this section models that are able to analyze the space-time dynamic of the data, the framework of Functional Data Analysis (Ramsay and Silverman, 1997). We first represent the discrete measures into functional data. For that smoothing an expansion in a basis function has to be done. After this preliminary FDA step, the second step is to resume the functional data by means of Functional Principal Component Analysis (FPCA). The theoretical basis of FPCA is the Karhunen-Loève theorem (Ash and Gardner, 1975). When dealing with several functional variables like in our case study, multivariate FPCA has been used thanks to Happ and Greven (2018). To identify groups with homogeneous characteristics on the basis of the multivariate functional data, we perform an hierarchical clustering based on the FPCA scores constructed using MFPCA. The principle of the adopted functional PCA method can be described by the following steps, setting $p=1$ (with p number of principal components of the PCA) to simplify the notations:

1. Dimension reduction by principal component analysis (PCA)
   The first step is to express raw data to functions using

$$X_s^j(t)=X_s(t)=\sum_{m=1}^{P}c_{s,j,m}b_m(t) \tag{4}$$

where $b_1(t),\ldots..,b_P(t)$ represents the collection of first P basis functions (Fourier, Spline, etc).

Let us recall some FDA notions useful for PCA. The empirical functional mean of $X_s^j(.)$ is:

$$\overline{X}_{j,n}(t)=\frac{1}{n}\sum_{s\in(s_1,\ldots,s_n)}X_s^j(t) \tag{5}$$

and the empirical covariance function:

$$\widehat{c}_{j,n}(t,u)=\frac{1}{n-1}\sum_{s\in(s_1,\ldots,s_n)}\left(X_s^j(t)-\overline{X}_{j,n}(t)\right)\left(X_s^j(u)-\overline{X}_{j,n}(u)\right). \tag{6}$$

Karhunen–Loeve's expansion (Ash and Gardner, 1975) is then used:

$$X_s^j(t)=\mu_j(t)+\sum_{k=1}^{K}\beta_{j,k,s}\phi_{j,k}(t) \tag{7}$$

where $\phi_{j,k}$ are the eigenfunctions (principal component functions; PCFs) associated with the variance-covariance operator (see Happ and Greven (2018) for more details) and $\lambda_{1j}>\lambda_{2j},\ldots..>\lambda_{Kj}$ are the eigenvalues, in theory $K=\infty$. The functional principal component scores $\beta_{j,k,s}=\int X_s^j(t)\phi_{j,k}(t)\,dt$ are assumed to be centered random variables. In practice, $\widehat{c}_{j,n}(.,.)$ is decomposed by:

$$\widehat{c}_n(t,u)=\sum_{k=1}^{K}\widehat{\lambda}_{k,j}\widehat{\phi}_{j,k}(t)\widehat{\phi}_{j,k}(u). \tag{8}$$

To determine eigenvalues and eigenfunctions, a number of works in the literature are proposed particularly in the context of univariate independent (Ramsay and Silverman, 2005) or multivariate independent (Happ and Greven, 2018) or univariate spatially as well as temporally

dependent (Hörmann et al., 2015; Kuenzer et al., 2021; Liu et al., 2017; Winzenborg, 2011).

We first use the estimated functional principal components (FPC) $\widehat{\phi}_{j,k}$ (by assuming that the data at the ESU are independent, Ramsay and Silverman (1997), Happ and Greven (2018)) to approximate X_s^j in a finite dimensional functional space:

$$X_s^j(t) \approx \overline{X}_{j,n}(t) + \sum_{k=1}^{K} \widehat{\beta}_{j,k,s}\widehat{\phi}_{j,k}(t) \tag{9}$$

When $p > 1$, univariate basis expansions for each element of the multivariate functional data $\mathbf{X}_s = \left(X_s^1, \ldots, X_s^p\right)$ are performed for each of the p functional components and the MPCA scores names $\widehat{\beta}_{k,s}$ are derived. (Happ and Greven, 2018; Happ-Kurz, 2020).

2. Clustering

   Whence the functional PCA is done, we conduct hierarchical clustering on the MPCA scores $\widehat{\beta}_{k,s}$.

3. Functional GAM (FGAM) models

   The functional PCA (R version 4.1.0; package MFPCA) carried out allowed the dimension reduction of the explanatory variables. We then used the functional generalized additive model (FGAM) McLean et al. (2014) as a more flexible alternative to the functional linear model (FLM) for regressing a scalar on functional predictors. The FLM first introduced by Ramsay and Dalzell (1991) and defined by:

$$Y_s = \beta_0 + \sum_{j=1}^{p} \int_{\mathcal{T}_j} \beta_j(t) X_s^j(t)\,\mathrm{d}t + \varepsilon \tag{10}$$

has been extensively studied (see Cardot et al. (1999, 2003), Crambes and Mas (2013), Yao et al. (2005)). Let $\mathbf{X}_s = \left(X_s^1, \ldots, X_s^p\right)$, we have the functional GAM expressed as follows:

$$\mathrm{E}(Y_s|\mathbf{X}_s) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} f_j\left(X_s^j\right)\right) \tag{11}$$

The estimation of f_j is based on the spectral decomposition of the covariance operator of $\mathbf{X}_s$ and from the scores of PCA of $\mathbf{X}_s$, namely $f_j\left(X_s^j\right) := \sum_{k=1}^{K} f_j^k\left(\xi_{j,s}^k\right)$, where $\xi_{j,s}^k$ are the PCA scores of X_s^j. This method is called Generalized Spectral Additive Model (GSAM) (Müller and Yao, 2008) and is more flexible because the independence of the scores for each functional covariate avoids problems of concurrency in the estimation of the partial functions associated with this covariate.

The Functional Generalized Spectral Additive Model (FGSAM), which aims to estimate the functional beta parameters using the R package fda.usc, selects the number of basis functions for the independent functional predictors and parameters that minimize the RMSE and MAE criteria. These criteria are used for model selection and help to ensure a better fit of the model to the data. The dataset was divided into training (75%) and test (25%) samples for the Generalized Additive Model (GAM) as well as for spatial Generalized Additive Model (GAMs), the Functional Generalized Spectral Additive Model (FGSAM) and the spatial Functional Generalized Spectral Additive Model (FGSAMs).

2.2.2.2. Spatial analysis. In addition to the functional covariate, we add the geographical coordinates as spatial co-variates in the FGSAM model:

$$\mathrm{E}(Y_s|\mathbf{X}_s) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} f_j\left(Z_s^j\right) + g(s)\right), \tag{12}$$

where $s = \left(s^1, s^2\right)$ is the vector of spatial coordinates.

3. Results

We first present the results of classical biostatistics methods and then those of functional methods.

3.1. Classical methods

3.1.1. Elementary statistics of SSL descriptors per frequencies

The analysis of SSL descriptors reveals interesting patterns in the thickness and depth of SSL across the different frequencies, with distinct variations according to the diel period.

Over the southern Senegalese continental shelf (Fig. 4), the SSL were thicker during the night while during the day, the SSL are almost thin. At all frequencies, SSL were deeper at night than at sunrise except at 200 kHz where we reported the opposite. In the high sea of southern Senegal, at 70, 120, and 200 kHz, the SSL were thicker at night than at sunset, in contrast to those detected at 38 kHz. The SSL were deeper at sunset than at night for all frequencies. In northern Senegal, the SSL were thicker during daytime and sunrise at 70, 120, and 200 kHz while at 38 kHz the SSL were thicker during the sunrise. In addition, the SSL were deeper during sunrise and day than during the night for all frequencies.

3.1.2. Principal components analysis (PCA) and clustering on environmental parameters

Clustering is performed on the principal components of the mean environmental parameters between the minimum and maximum depths of the SSL (Fig. 5, see Fig. S3 in the supplementary material, for more details), i.e., where the SSL were present in the water column. On the southern Senegalese continental shelf, at 38 kHz, we identified two classes. Class 1 is characterised by ESU sampled during nighttime and sunrise. They were deeper and more saline compared to the Class 2 sampled during the daytime, which were warmer, more turbid, and more fluorescent. We observed the same results at the three other frequencies. In the Senegalese high sea, the results were similar at 38, 70, and 120 kHz. Class 1 was characterised by night-sampled ESU that were shallower, warmer, less saline, less turbid, and less fluorescent compared to Class 2 sampled during sunset. Only at 200 kHz, the diel period did not influence the class distribution. Class 1 was characterised by shallower, less warm, less turbid, and fewer fluorescence ESU compared to Class 2. In the northern Senegalese shelf at 38 and 200 kHz, Class 1 was sampled during sunrise and daytime period and was deeper, less warm, less saline, and lower in fluorescence compared to Class 2. At 70 and 120 kHz, Class 1 is characterised by daybreak and daytime period that were deeper and more turbid compared to Class 2, which were warmer, more saline, and with higher fluorescence.

3.1.3. Spatial generalized additive model (GAMs)

GAMs results showed that on the continental shelf, environmental parameters influenced the thickness and depth of the Sound Scattering Layer (SSL) more during nighttime and sunrise compared to daytime, across all frequencies (Table 1, see Tables S1 and S2, Fig. S6 in the supplementary document for more detail). Factors such as local depth, sea temperature, salinity, fluorescence, turbidity, and geographical coordinates exhibited varying effects on SSL thickness and depth at different frequencies and periods. Similar patterns were observed in the southern Senegalese high sea, where bottom depth, temperature, salinity, fluorescence, turbidity, and geographical coordinates influenced SSL characteristics during sunset and nighttime. In the northern Senegal shelf, bottom depth, temperature, salinity, fluorescence, turbidity, and geographical coordinates played a role during sunrise and daytime. The analysis revealed both positive and negative effects of these environmental factors on SSL thickness and depth, underscoring the complexity of their interactions. However, the fit quality for SSL density (S_A) was found to be inadequate. Overall, the study highlights the significance of spatial statistical analysis in understanding the multiple effects of environmental variation on SSL in ecological studies. Overall, the GAMs analysis highlights the complex relationships between environmental variation and SSL, emphasizing the importance of spatial-functional statistical analysis in ecological studies.

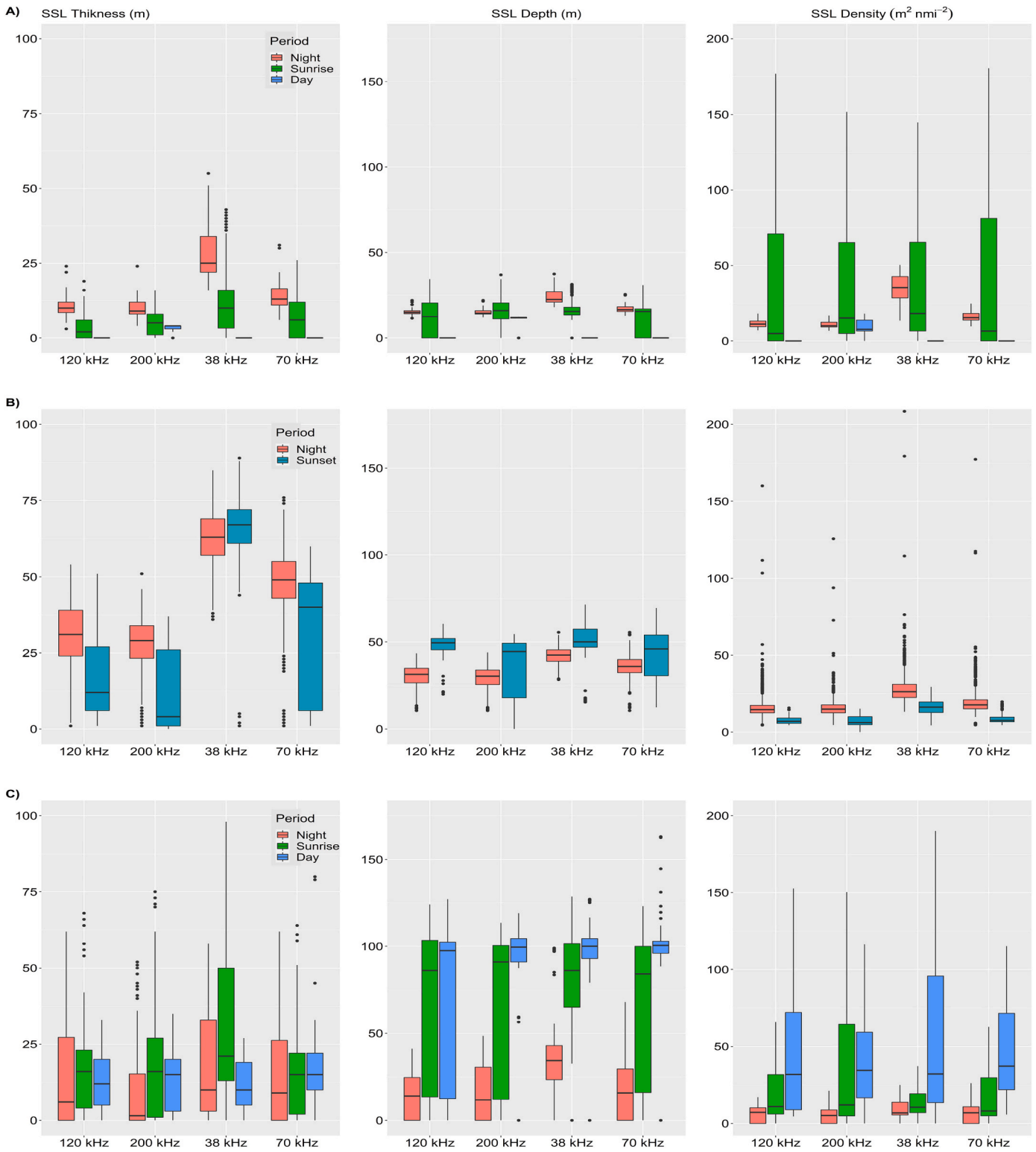


Fig. 4. Boxplot (minimum, maximum, and median) of sound scattering layer (SSL) descriptors (SSL thickness (m), SSL density (expressed in m^2 nmi^{-2}), and SSL depth (m)) grouped by diel period (Night, Sunrise, Day and Sunset) over Senegalese waters (A: southern continental shelf; B: southern high sea and C: northern continental shelf) as observed by the echosounders during the AWA sea survey at four different frequencies (kHz).

3.2. Functional methods

3.2.1. Multivariate functional principal component analysis (MFPCA)

The results of the multivariate functional principal component (MFPCA) analysis on the environmental parameters show the variation of the components along the depths (Fig. 6, see Fig. S4 for more detail). It appeared that the first component of each variable generally groups together ESU with either high or low values at a particular depth range, while the second component groups together ESU with high or low values at a different depth range (Table 2). This suggests that there are multiple factors influencing the distribution of these ESU in the water column.

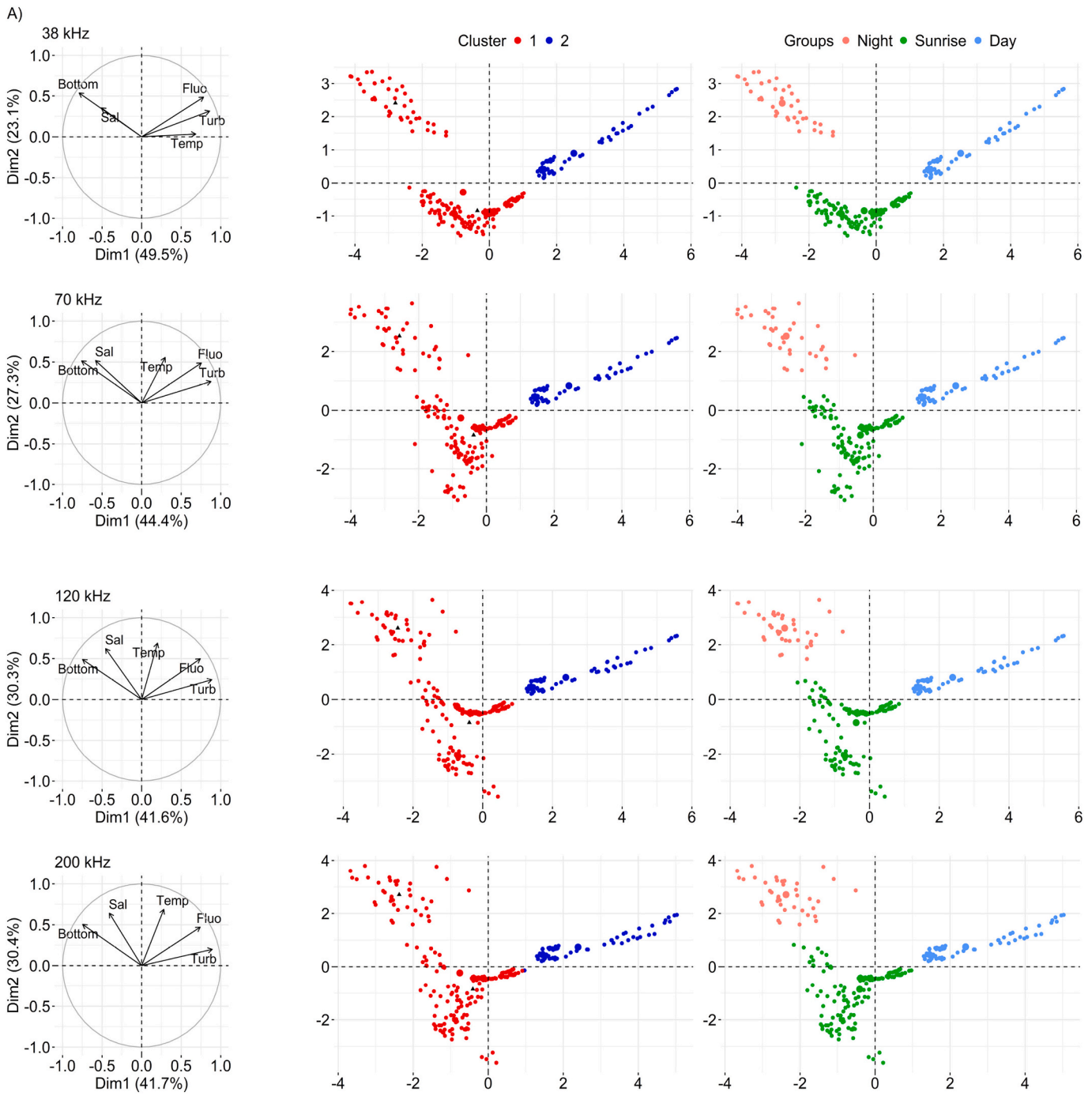


Fig. 5. Principal Component Analysis (PCA: PC1 on x-axis and PC2 on y-axis) followed by classification (CAH) (red: Cluster 1; blue: Cluster 2) of the mean environmental parameters (sea temperature, fluorescence, turbidity, and salinity) measured within the minimum and maximum depths of the sound scattering layer (SSL) at different frequencies; 38, 70, 120, and 200 kHz and bottom depth, grouped by diel period (day, sunset, night and sunrise) over Senegalese waters (southern continental shelf) during the AWA sea survey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.2. *Clustering*

Clustering was performed on the scores derived from the functional principal components (PC1 and PC2) (Fig. 7, see Fig. S5 for more detail). On the continental shelf of southern, at 38 kHz, class 1 was characterised by the second component while the first characterises Class 2. At 70 kHz, Class 2 was characterised by the first component, while Class 1 was characterised by the second component. At 120 kHz, class 1 contains low values of the first component, while Class 2 contains the opposite. At 200 kHz, Class 1 contains low values of the first component, while Class 2 contains the opposite. In the high sea southern Senegalese coast, at 38 kHz, Class 1 was characterised by low values of the first component, while Class 2 has the opposite. At 70 kHz, Class 2 was characterised by high values of the first component while Class 1 has the opposite. At 120 kHz, Class 1 was characterised by low values of the first component, while Class 2 has the opposite. At 200 kHz, Class 1 has low values of the first component while Class 2 has the opposite Over the northern Senegalese shelf, at 38 kHz, Class 1 was characterised by low values of the first component, while at Class 2 we observe the opposite. At 70 kHz, Class 1 was characterised by low values of the first component, while at Class 2 the opposite was observed. At 120 kHz, Class 2 was characterised

Table 1
Results of Generalized additive model (GAM) and spatial Generalized additive model (GAMs) showing the effects (blue: Positive; red: Negative; gray: Not significant) of oceanographic conditions (sea temperature, salinity, turbidity, fluorescence), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) on sound-scattering layers (SSLs) thickness, detected by echosounders at four frequency (38, 70, 120 and 200 kHz), per diel period (day, sunset, night and sunrise), and according to three Senegalese geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf).

(A)

Thickness																								
Frequencies	38 kHz						70 kHz						120 kHz						200 kHz					
Diel period	Night		Sunrise		Day		Night		Sunrise		Day		Night		Sunrise		Day		Night		Sunrise		Day	
Model	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs
Temperature																								
Fluorescence																								
Salinity																								
Turbidity																								
Bottom																								
Lat	-		-		-		-		-		-		-		-		-		-		-		-	
Lon	-		-		-		-		-		-		-		-		-		-		-		-	

(B)

Thickness																
Frequencies	38 kHz				70 kHz				120 kHz				200 kHz			
Diel period	Night		Sunset		Night		Sunset		Night		Sunset		Night		Sunset	
Model	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs
Temperature																
Fluorescence																
Salinity																
Turbidity																
Bottom																
Lat	-		-		-		-		-		-		-		-	
Lon	-		-		-		-		-		-		-		-	

(C)

Thickness																								
Frequencies	38 kHz						70 kHz						120 kHz						200 kHz					
Diel period	Night		Sunrise		Day		Night		Sunrise		Day		Night		Sunrise		Day		Night		Sunrise		Day	
Model	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs	GAM	GAMs
Temperature																								
Fluorescence																								
Salinity																								
Turbidity																								
Bottom																								
Lat+Lon	-		-		-		-		-		-		-		-		-		-		-		-	

by high values of the first component, while Class 1 was the opposite. At 200 kHz, class 1 contains low values of the first component, while Class 2 had the opposite.

3.2.3. Spatial functional generalized spectral additive model (FGSAMs)

The results of the FGSAMs performed on the SSL descriptors provided estimated $\widehat{\beta}$ parameters of the functional covariates that varied with depth (Fig. 8, see Fig. S7 in the supplementary material, for more details). On the southern continental shelf, only temperature has an overall negative effect on SSL thickness at 38 kHz while in the southern Senegal high sea, fluorescence has an overall positive effect on thickness at 70 and 120 kHz. For SSL depth, fluorescence has an overall negative effect in these two areas at 200 kHz. For the other parameters, positive and negative effects are recorded along the depths (Table 3 and see Table S3 in the supplementary material, for more details).

4. Discussion

4.1. Analytical approaches: Multivariate and functional biostatistics methods, with or without spatial considerations

PCA was used to visualize correlations between environmental variables and to identify homogeneous groups of ESU in a two-dimensional space (Abdi and Williams, 2010; Chatfield and Collins, 1980; Demšar et al., 2013; Granato et al., 2018). The drawback of such statistical analysis was that it did not take into account the shape of the variation of the data along the depth because it is performed on aggregated data. In our case study, the analyses were based on aggregated dataset where each ESU is associated with a single mean value of sea temperature, fluorescence, salinity, and turbidity. On the one hand, the functional multivariate PCA (Happ and Greven, 2018) was free of this aggregated aspect and showed in our case study, the similarities between the ESU through the high or low peaks of the parameters observed on given depth points. It allowed the study of the shape variation of the data along the water column. The GAM using continuous georeferenced positions (latitude and longitude) with a bivariate smoothing function as

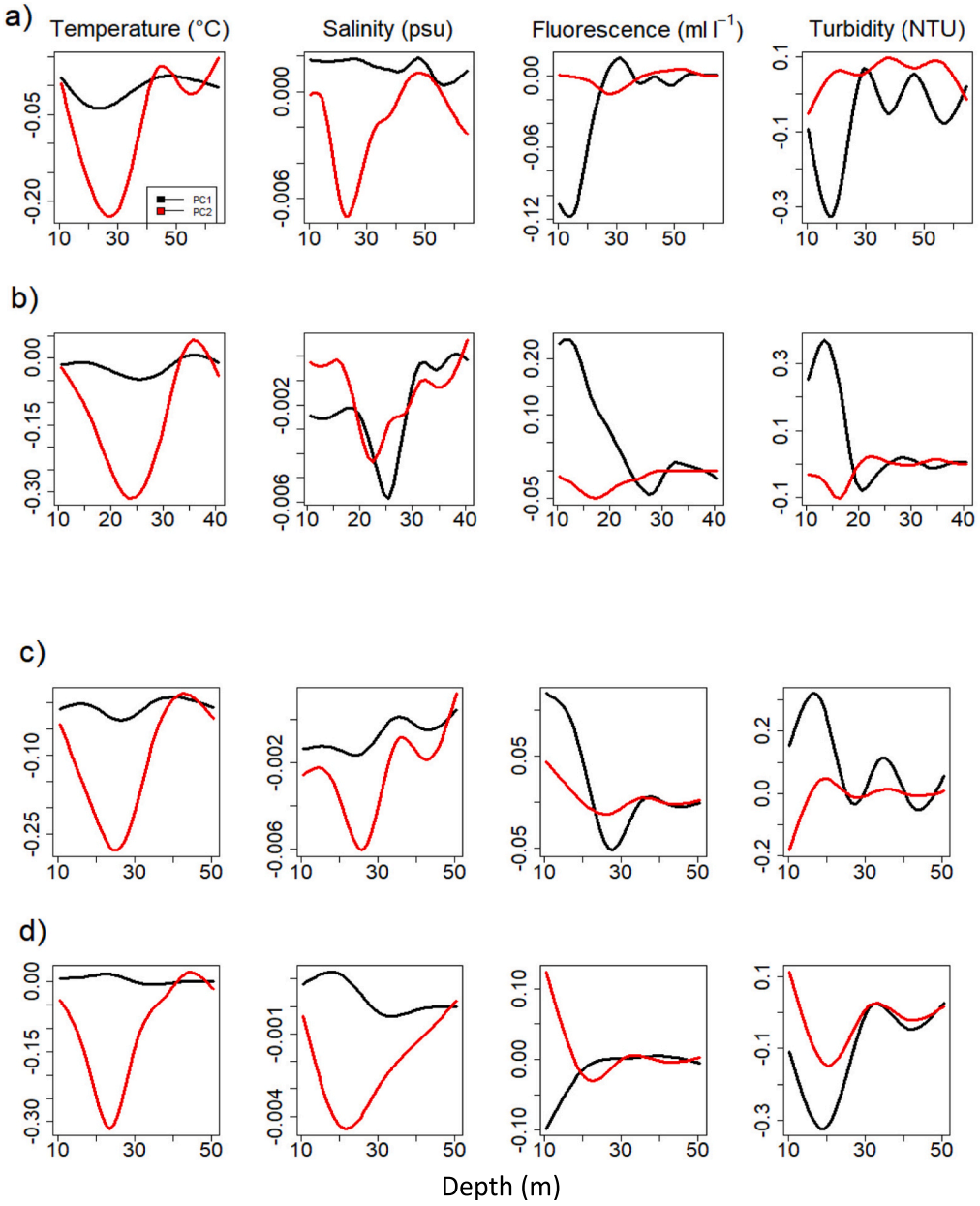


Fig. 6. Multivariate Functional Principal Component Analysis (MFPCA) of oceanographic conditions (sea temperature (in °C), salinity (in PSU), fluorescence (in ml l^{-1}), turbidity (in NTU)) along the depths) measured between the minimum and maximum depths of the layers at four frequencies. Representation of the two main functional components (black: first component (PC1); red: second component (PC2)) over Senegalese waters (south continental shelf), observed at various frequencies: (a) 38, (b) 70, (c) 120 and (d) 200 kHz. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spatial covariates showed overall the best performance (lower RMSE and MAE; Table 4; see Table S4 in the supplementary material, for more details). Spatial dependence was modeled in the systematic part of the model, which has improved the goodness of fit. The GAM were powerful exploratory tools because they offered great flexibility in data analysis by introducing nonparametric and/or parametric functions. Its main advantage is that it did not require making assumptions about the form of the relationship between the SSL descriptors and environmental parameters. Its application allowed to specify of the error model, to adjust the shape of the distribution of the data and thus has lower and more reliable *p*-values (Hastie and Tibshirani, 1990). The FGSAM offers also great flexibility and helps to avoid the curse of dimensionality (Febrero-Bande and González-Manteiga, 2013). The results obtained by the FGSAM showed that the effect of environmental parameters on the descriptors of SSL appeared complex and cannot be modeled by a functional linear model. Overall, multivariate biostatistics are suitable for analyzing datasets with a relatively small number of variables and can provide valuable insights into correlations (Diogoul et al., 2020). However, they may lack the ability to capture fine-scale variations and account for temporal or spatial dependencies in the data. In contrast, functional methods, whether spatialized or not, are specifically designed to handle functional data and excel in capturing complex patterns along a continuum as it is the case in our dataset. They offer more flexibility and are particularly useful when dealing with high-dimensional, spatial, or temporal datasets.

Table 2
Results of Multivariate Functional Principal Component Analysis (MFPCA) of oceanographic condition (sea Temperature (in °C), Salinity (in PSU), Fluorescence (in ml l^{-1}), Turbidity (in NTU)) along the depths) measured between the minimum and maximum depths of the layers at four frequencies. PC1 represents the first component and PC2 the second component; values in intervals (m) represent ranges of variation where we have low and high peak (↗:high peak; ↘: low peak). The results are presented for three different Senegalese areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf).

(A)

Frequencies (in kHz)	38		70		120		200	
Components	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Temperature	20–30↘;40–50↗	20–30↘;40–50↗	20–30↘;35–40↗	20–30↘;35–40↗	25–30↘;40↗	25↘;40–45↗	20–25↗;35↘	20–25↘
Fluorescence	10–20↘;30↗	20–30↘	10–15↗;25–30↘	15–20↘	25–30↘	25↘	40↗	20–25↘
Salinity	40–50↗;50–60↘	20–30↘;40–50↗	25↘	20–25↘	25↘;35↗	25↘	15–20↗;30–35↘	20–25↘
Turbidity	10–20↘;30↗	30–40↗	10–15↗;20↘	15–20↘;20–25↗	15–20↗;40–45↘	20↗	20↘;20–35↗	20↘

(B)

Frequencies (in kHz)	38		70		120		200	
Components	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Temperature	25–50↘	50↗	25–50↗	50↗;25–37↘	30–40↗	40–50↗	50↘; 75–87↗	25–37↘;62↗
Fluorescence	25–50↘	50↗	37↗	50↗	30–40↗	20–30↘	37–50↘	75↗
Salinity	25↗; 50–62↘	50–75↗	50–60↗	50–62↗	20–30↘	20–30↗	25↗	25–37↗
Turbidity	25–50↘	50↗	37↗	25–37↘;50↗	30–40↗	20–30↘	37↘	25–37↘; 50↗

(C)

Frequencies (in kHz)	38		70		120		200	
Components	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Temperature	10–62↘	10–62↘	25–75↗	50–62↗	25–37↘	50↘	37↗	50–62↘
Fluorescence	12–25↘	25–50↗	25–37↗	37–50↗	25↘	25–37↗	12–25↗	25 m↗
Salinity	50–75↘	75↘	50–62↗	62↘	62–75↘	50–62↘	62↗	50–62↘
Turbidity	37–50↗	25–50↗	87↗	87↘	37↗	25–37↗	37↘	37↗

4.2. Diel vertical migration: Varied SSL patterns in Senegalese waters

The analysis of SSL descriptors demonstrates the significant impact of the diel period on SSL thickness and depth, revealing distinct patterns influenced by the diel cycle. Over the southern Senegalese continental shelf, nighttime SSL exhibit greater thickness and depth compared to daytime SSL can be attributed to the widely recognized diel vertical migration (DVM) patterns displayed by numerous marine species. DVM is a behavioral mechanism generally characterised by an ascent towards the surface of zooplanktonic and micronektonic organisms during the night to feed and a descent during the day to avoid predation by visual predators. This type of migration is known as DVM type I (Bianchi et al., 2013; Lehodey et al., 2015). In northern Senegal, a different set of patterns emerged. SSL were thicker during the daytime and sunrise, consistently exhibiting greater depth during both sunrise and the day compared to the night. This SSL thickness and depth pattern reflects an inverse DVM. Indeed, some zooplanktonic organisms perform an inverse DVM (type II), ascending in the morning and descending in the evening, which is the reverse pattern generally observed with vertically migrating animals (Cushing, 1951; Ohman et al., 1983). In the Southern Senegalese continental shelf, the sampling was mainly achieved during the daytime, which may have biased the observed DVM type II. Nevertheless, Tiedemann and Brehmer (2017) also reported such inverse DVM for ichthyoplankton, which confirms the trophic relationship between ichthyoplankton and copepod, as Diogoul et al. (2022) showed that the SSL are mainly composed of copepods.

4.3. Effect of the pelagic environment on the spatial structure of the micronektonic layer (SSL)

PCA analysis allowed the physicochemical characterization of the water masses. The dataset was organized into two distinct ESU classes. Through PCA, we identified patterns and similarities within these classes, enabling a statistical distinction between shallow, cold areas and deeper, warmer regions. The application of PCA allowed for a nuanced exploration of the physicochemical attributes associated with each ESU class. The analysis not only facilitated the identification of common features but also provided a statistical foundation for characterizing the differences between the two classes. In particular, the distinction between shallow and cold conditions versus deeper and warmer environments became evident through the derived statistical insights. This approach enhances our understanding of water mass characteristics, and is in line with current knowledge (Ndoye et al. (2014); Tiedemann and Brehmer (2017)). On the southern Senegalese continental shelf, the 38 and 70 kHz observations give similar results on SSL due to their somewhat similar acoustic responses (Guillard et al., 2004). The Class 2 ESU were shallower closest to the coast, where SSL were seldom as expected by Diogoul et al. (2020) who explain this by the turbulence in the water column generated by the upwelling phenomenon occurring at the time of the survey. The upwelling phenomenon occurs on the Senegalese coast between November and May (Balde et al., 2019; Faye et al., 2015) and is mainly determined by favourable meridional wind, topography, and density stratification (Estrade et al., 2008). Coastal upwelling is an oceanographic process that brings cold and nutrient-rich waters to the ocean surface from depth (Ndoye et al., 2017). Upwelled coastal waters are cold and rich in nutrients, which is favourable to increase primary production and thus chlorophyll-a concentration (Auger et al., 2016), while the offshore zone is characterised by warm and less nutrient-rich waters (Tiedemann and Brehmer, 2017). The Class 2 ESU water were warmer and the SSL were more at the surface during daytime (due to low bottom depth in Class 2 ESU). The Class 1 ESU were deeper because they were closer to the continental slope than the Class 2 ESU, which were onshore and characterised by waters that are more turbid. In the southern Senegalese high sea, at 38, 70, and 120 kHz, the results were similar. Class 2 ESU were warmer, more turbid, deeper, and richer in chlorophyll-a concentration. In the northern Senegalese shelf, the 38

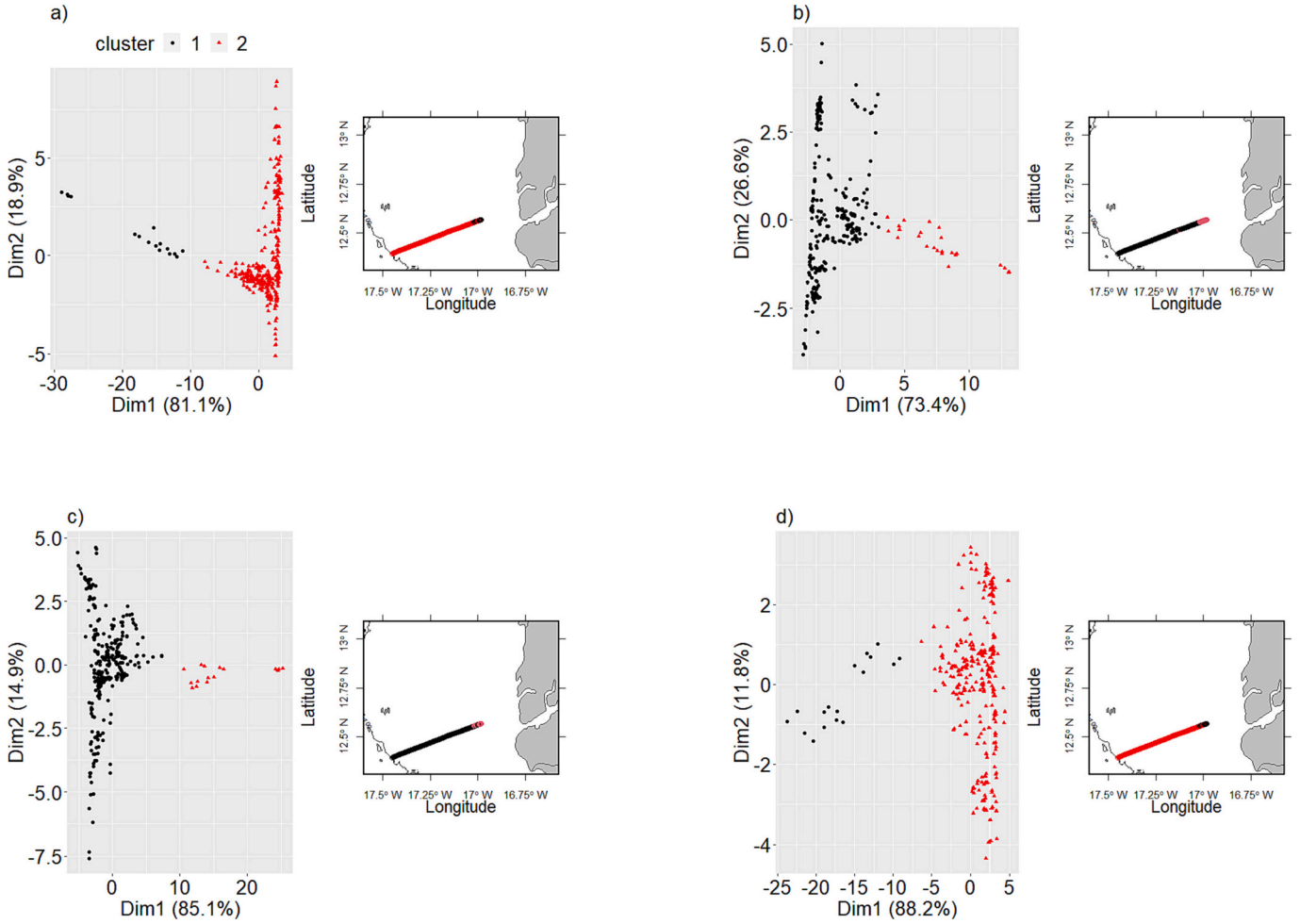


Fig. 7. Classification of functional principal component scores (derived from Multivariate functional principal component analysis MFPCA) (black: Cluster 1; red: Cluster 2) over Senegalese waters (south continental shelf), observed at various frequencies: (a) 38, (b) 70, (c) 120 and (d) 200 kHz. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and 200 kHz Class 2 SSL during the night were found upper in the water column and more onshore. The Class 2 surface waters were also warmer due to solar radiation and were more turbid and richer in chlorophyll-a concentration i.e. due to the effect of the Senegal River waters. At 70 and 120 kHz, the Class 2 ESU sampled at night were warmer than Class 1 because their SSL were near the surface. The utilization of GAM has enabled the recognition of factors that affect SSL vertical distribution (thikness and depth), highlighting the importance of variables like temperature, chlorophyll-a concentration, turbidity. In the GAMs results, the low values of RMSE obtained with the introduction of geographical positions could be due to the spatial autocorrelation between the sample points. The vertical distribution of the SSL is influenced by the geographical position (longitude and latitude), which, in turn, aligns with the local bathymetry. The coastal SSL (both North and South) exhibited shallower depths compared to those in the open sea. On the southern Senegalese continental shelf, the SSL thickness increased at low latitude (south) and longitude (offshore) but on the southern Senegalese high sea and northern shelf the direction of variation was irregular. On the southern Senegalese continental shelf, the GAMs showed that the SSL thickness increases with bottom depth during the night and sunrise. The bottom depth and diel period were the main parameters influencing the SSL thickness in Senegal (North and South) shelf and high sea. Bathymetry is described as one of the physical factors that control the SSL depth in the water column in the shelf north and south and high sea. In Senegal, Diogoul et al. (2020) also reported a strong effect of bottom depth and diel period on both SSL thickness and depth. The distribution of SSL is often linked to the hydrographic structure of the water column (Berge et al., 2014). The vertical distribution of SSLs and their thickness are linked to strong vertical gradients of temperature, chlorophyll-a concentration, and salinity. In the daytime ESU, the SSL had a small thickness or is sometimes non-existent due to the unstable conditions of the water column (Diogoul et al., 2020), and in this southern zone, the gradients are less marked, which can explain the no significance of the parameters during the day on the thickness. The depth of the SSL frequently showed associations the thermocline (Diogoul et al., 2020; Yoon et al., 2007). We showed a significant effect on SSL depth attributable to sea temperature and chlorophyll-a concentration during the night, as well as turbidity during both night and sunrise. This reveals the complexity of the dynamics of SSL responses subjected to the influence of different parameters, emphasizing the importance of considering the diel period. The SSL are located in areas of strong vertical gradients and at chlorophyll-a concentration peaks i.e. a high phytoplankton production. The relationship between chlorophyll-a concentration and SSL thickness (mainly zooplankton (Diogoul et al., 2020)) at night is explained by the presence of a trophic relationship between phytoplankton and zooplankton which appeared to expend SSL size in vertical. Turbidity also affects the SSL, it is a parameter that was often linked to chlorophyll-a concentration in the high sea (Cui and Lv, 2014; Dall'Olmo and Gitelson, 2005). However, in Senegal, we note a negative correlation between water turbidity and chlorophyll-a concentration in the North, while observing a positive correlation in the South. The presence of the Casamance River is believed to contribute to

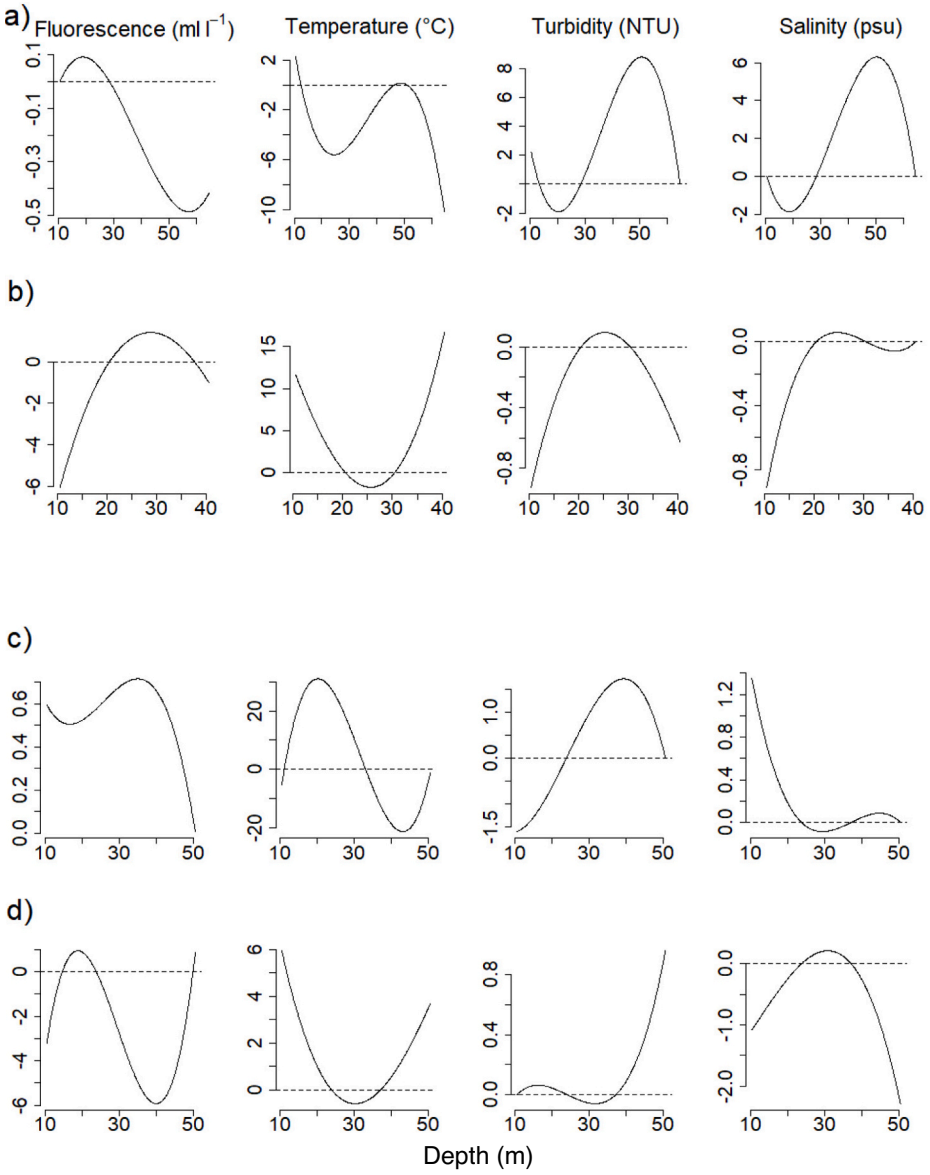


Fig. 8. Estimated oceanographic conditions parameters (sea temperature, fluorescence, salinity and turbidity) resulting from the functional model for SSL thickness with the spatial dimension (spatial Functionnal Generalized Spectral Additive Model (FGSAMs)) over Senegalese waters (south continental shelf) and as expect at various echosounder frequencies (a) 38, (b) 70, (c) 120 and (d) 200 kHz.

this contrasting relationship. On the high sea of southern Senegal, the bottom depth affects the SSL depth and thickness. During sunset, the SSLs tend to be deeper and thinner (except at the 38 kHz). In contrast, during the night, SSLs become shallower and thicker (except at the 38 kHz) as organisms migrate towards the surface for feeding. In the North of Senegal, the nighttime ESU closer to the coast compared to the daytime and sunrise ones, have thicker and deeper SSL, which can bias the interpretation and call for additional data before to conclude.

The clustering performed on the MFPCA scores provided also homogeneous groups similarly on the southern Senegalese sea. At 38 kHz, SSL were thicker and found at deeper locations (Class 2), while few SSL was found in coastal domain (Class 1) due to more turbulent oceanographic conditions (Diogoul et al., 2020). High chlorophyll-a concentration and turbidity are observed in coastal areas. The depth and location of SSL vary, with Class 2 typically found closer to the 100 m isobath at 38 and 200 kHz and closer to the coast at 70 and 120 kHz, where SSL are shallower. On the high sea of southern Senegal on the 38 and 200 kHz frequency. On the 70 and 120 kHz frequency, Class 2 is found further offshore where the observed SSL were deeper and physical parameters were measured to a certain depth in contrast to Class 1. In the North of Senegal, homogeneous classes were not found, which reflects a more homogeneous environment than in the two southern study areas, where the Casamance river influence the shallow part of the shelf (Thiam and Singh, 2002).

In the FGSAM results, complex effects of sea temperature on SSL depth and thickness with positive and negative effects were observed at different depth and frequency (Table 2, S3). This relationship has been observed in other regions (Diogoul et al., 2020; Kang et al., 2021). Warmer sea temperatures often lead to increased thermal stratification, reducing vertical mixing (Somavilla et al., 2017) and potentially causing a decrease in SSL thickness (Diogoul et al., 2020; Proud et al., 2017). Phytoplankton biomass (Chlorophyll-a concentration), exhibits effects on SSL thickness and depth, varying with depth and frequency used (Receveur et al., 2020; Song et al., 2022). In some depth ranges, higher phytoplankton biomass was associated with a positive effect on SSL thickness, suggesting increased biological activity and biomass. However, using FDA we observed that beyond certain depth ranges, a negative effect on SSL thickness may be observed, possibly indicating

Table 3
Results of Functional Generalized Additive Spectral Model (FGSAM) and spatial Functional Generalized Additive Spectral Model (FGSAMs) showing the effect (blue: Positive; red: Negative; gray: Not significant; values in intervals (m) represent ranges of variation; + represents beyond this value (m) and Night represents the reference modality (by default, coefficients are calculated compared to this modality)) of oceanographic condition (temperature, salinity, turbidity, fluorescence, diel period, and bottom depth) on sound-scattering layers (SSLs) thickness, detected by echosounders (38, 70, 120 and 200 kHz), according to three geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf).

(A)

Thickness								
Frequencies	38 kHz		70 kHz		120 kHz		200 kHz	
Model	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs
Temperature				10-20 ; 20-30 ; 30+	10-35 ; 35+	10-35 ; 35+	10-30 ; 30-55 ; 55+	10-25 ; 25-40 ; 40+
Fluorescence	10-30 ; 30+	10-30 ; 30+	10-20 ; 20-35 ; 35+	10-20 ; 20-35 ; 35+		10-20 ; 20+	10-30 ; 30+	<15 ; 15-25 ; 25+
Salinity	10-30 ; 30+	10-30 ; 30+	10-20 ; 20-30 ; 30+	10-20 ; 20-30 ; 30+	10-25 ; 25-35 ; 35+	10-25 ; 25-35 ; 35+	10-30 ; 30-50 ; 50+	10-25 ; 25-40 ; 40+
Turbidity	10-30 ; 30+	10-30 ; 30+	10-20 ; 20-30 ; 30+	10-20 ; 20-30 ; 30+	10-25 ; 25+	10-25 ; 25+	10-30 ; 30-55 ; 55+	10-25 ; 25-35 ; 35+
Bottom								
Night								
Sunrise								
Day								
Lat	-		-		-		-	
Lon	-		-		-		-	

(B)

Thickness								
Frequencies	38 kHz		70 kHz		120 kHz		200 kHz	
Model	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs
Temperature	10-45 ; 45+	10-50 ; 50+	10-42 ; 42+	10-40 ; 40+	<10 ; 10-45 ; 45+	10-50 ; 50+	<20 ; 20-60 ; 60-118 ; 118+	<10 ; 10+
Fluorescence	10-45 ; 45+	10-70 ; 70+			10-80 ; 80+		10-40 ; 40-120 ; 120+	10-40 ; 40-120 ; 120+
Salinity		10-50 ; 50+	10-40 ; 40+	10-40 ; 40+	10-45 ; 45+	10-50 ; 50+	10-115 ; 115+	10-110 ; 110+
Turbidity	10-80 ; 80+	10-90 ; 90+	<25 ; 25-80 ; 80+	10+	10-80 ; 80+	10-37 ; 37-80 ; 80+	10-80 ; 80+	10-80 ; 80+
Bottom								
Night								
Sunset								
Lat	-		-		-		-	
Lon	-		-		-		-	

(C)

Thickness								
Frequencies	38 kHz		70 kHz		120 kHz		200 kHz	
Model	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs
Temperature	10-50 ; 50-80 ; 80+	10-30 ; 30-90 ; 90 +	10-40 ; 40+	10-50 ; 50+	10-50 ; 50+	10-50 ; 50+	10-50 ; 50+	10-30 ; 30-50 ; 50-120 ; 120+
Fluorescence	10-90 ; 90+	10-80 ; 80 +	10-80 ; 80+	10-90 ; 90 +	10-80 ; 80-110 ; 110+	10-90 ; 90 +	10-80 ; 80-110 ; 110+	10-30 ; 30-120 ; 120+
Salinity	10-20 ; 20-90 ; 90+	<20 ; 20-60 ; 60 +	20-40 ; 40-110 ; 110+	10-30 ; 30-110 ; 110 +	10-90 ; 90 +	10-40 ; 40-90 ; 90 +		10-80 ; 80+
Turbidity	20-80 ; 80+	10-30 ; 30-90 ; 90 +	10-100 ; 100+	10-30 ; 30-110 ; 110+	20-60 ; 60+	10-20 ; 20-110 ; 110+	10-60 ; 60-120	10-20 ; 20-62 ; 62-120 ; 120 +
Bottom								
Night								
Sunrise								
Day								
Lat	-		-		-		-	
Lon	-		-		-		-	

changes in scattering components and the presence of different marine organisms. Species identification through biological sampling has been highlighted as an important step in understanding the composition of the SSL (Blanluet et al., 2019; Kloser et al., 2009). Regarding SSL depth, at lower frequencies a positive effect of phytoplankton biomass was often shown, indicating the presence of organisms with trophic relations on phytoplankton, such as macrozooplankton. However, at higher frequencies, a negative effect might be observed, suggesting the presence of different organisms or scattering components not strongly linked to phytoplankton, such as diatoms (Strom et al., 2001). That underlines both interest in the multifrequency approach and the application of the FDA for fine-scale observations. The spatial FDA has allowed investigation of the fine-scale effects of turbidity on SSL thickness and depth. Our findings suggest that turbidity generally has a negative effect on SSL thickness within certain depth ranges, indicating reduced scattering and potentially lower biomass or decreased presence of scattering organisms. Klevjer et al. (2016) reported that under high dissolved oxygen conditions, weighed mean depth of SSL decreased with increasing turbidity. However, beyond those depth ranges, turbidity can have a positive effect on SSL thickness, indicating increased scattering and potentially higher biomass or the presence of different types of scattering organisms. Similar to the influence of phytoplankton biomass and turbidity, the effect of salinity on SSL thickness and depth depend also on frequency and depth. As with seawater temperature, Salinity provides useful indications of water mixing (Ndoye, 2016). The effects can be both negative and positive within specific depth ranges. Further study must be encouraged on linking SSL behavior mechanistically to salinity with so low amplitude of variation. In the Arctic Ocean, salinity is one of the factors that affect the SSL distribution (La et al., 2015). The effects of environmental factors on SSL descriptors also showed regional variations, as evidenced by the FDA in the three Senegalese studied areas. These variations indicate that the relationships between environmental factors and SSL descriptors were influenced by local conditions and ecosystem dynamics. Assunção et al. (2020) using FPCA also confirmed the interest of the FDA to discriminated regional variation as Ariza et al. (2022b) Further research and validation are necessary to identify

Table 4
Comparison of non-spatial (GAM and FGSAM) and spatial models (GAMs and FGSAMs) for SSL thickness, as detected using four different echosounder frequencies (in kHz; denoted as suffix after statistical model abbreviations) and spread over three Senegalese geographical areas (A) southern continental shelf, (B): southern high sea and (C): northern continental shelf (during the AWA fisheries acoustics sea survey). R^2_ajusted (Adjusted R-Squared); R^2 (R-Squared); RMSE_Train (Root Mean Square Error in training set); MAE_Train (Mean Absolute Error in training set); RMSE_Test (Root Mean Square Error in test set); MAE_Test (Mean Absolute Error in test set).

(A)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.97	0.98	1.98	1.36	3.37	2.26
GAMs38	0.98	0.99	1.53	1.01	2.77	1.61
FGSAM38	0.98	0.99	1.46	1.00	2.46	1.58
FGSAMs38	0.99	0.99	1.01	0.69	1.82	1.19
GAM70	0.91	0.93	1.91	1.25	2.41	1.59
GAMs70	0.98	0.99	0.86	0.54	1.88	1.08
FGSAM70	0.88	0.89	2.31	1.62	2.74	1.86
FGSAMs70	0.97	0.97	1.17	0.77	1.89	1.20
GAM120	0.85	0.87	1.58	1.03	2.41	1.53
GAMs120	0.92	0.94	1.11	0.68	1.64	0.99
FGSAM120	0.90	0.92	1.28	0.77	2.27	1.31
FGSAMs120	0.90	0.92	1.25	0.75	2.24	1.28
GAM200	0.70	0.75	2.14	1.55	3.10	2.28
GAMs200	0.89	0.91	1.26	0.94	1.75	1.24
FGSAM200	0.72	0.75	2.12	1.52	3.30	2.33
FGSAMs200	0.82	0.85	1.64	1.11	2.67	1.72

(B)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.77	0.79	5.34	4.20	5.50	4.33
GAMs38	0.88	0.89	3.83	3.03	4.21	3.39
FGSAM38	0.70	0.74	5.92	4.39	6.20	4.68
FGSAMs38	0.88	0.90	3.63	2.88	4.19	3.24
GAM70	0.85	0.86	5.34	4.25	5.51	4.35
GAMs70	0.92	0.93	3.79	2.95	4.39	3.47
FGSAM 70	0.81	0.84	5.77	4.19	7.21	5.22
FGSAMs70	0.85	0.88	5.11	3.83	6.77	5.03
GAM120	0.82	0.83	5.37	4.29	5.33	4.09
GAMs120	0.90	0.91	3.90	3.11	4.53	3.57
FGSAM120	0.87	0.89	4.36	3.38	5.47	4.20
FGSAMs120	0.89	0.91	3.92	3.09	4.97	3.70
GAM200	0.81	0.82	5.13	3.97	4.86	3.81
GAMs200	0.90	0.91	3.64	2.75	4.12	3.03
FGSAM200	0.85	0.87	4.37	3.30	5.08	3.96
FGSAMs200	0.90	0.92	3.53	2.65	4.43	3.33

(C)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.73	0.76	10.96	7.80	13.41	9.19
GAMs38	0.89	0.91	6.67	4.62	7.16	5.02
FGSAM38	0.80	0.84	8.99	6.56	12.12	8.99
FGSAMs38	0.90	0.92	6.44	4.65	8.23	6.05
GAM70	0.81	0.84	6.22	4.30	9.96	6.02
GAMs70	0.88	0.91	4.75	3.42	8.24	4.97
FGSAM70	0.83	0.86	5.81	3.80	11.90	7.30
FGSAMs70	0.84	0.87	5.50	3.78	10.38	6.45
GAM120	0.82	0.85	6.02	3.93	6.94	4.11
GAMs120	0.87	0.90	4.95	3.12	5.70	3.35
FGSAM120	0.85	0.87	5.52	3.93	7.62	5.16
FGSAMs120	0.91	0.93	4.02	2.81	5.51	3.38
GAM200	0.82	0.84	6.17	4.03	11.25	6.29
GAMs200	0.90	0.92	4.44	2.98	6.25	3.96
FGSAM200	0.79	0.83	6.49	4.52	7.33	5.48
FGSAMs200	0.90	0.93	4.24	2.72	4.93	3.60

remotely the marine organism acoustically and then better understand the underlying mechanisms driving the observed effects.

5. Conclusion

In this study, we analyzed a multifrequency acoustics dataset acquired from scientific echosounders in the West African waters using FDA, both with and without the incorporation of spatial dimensions. Additionally, we compared the outcomes of this analysis with multivariate methods. The FDA methodology provides valuable insights into the variations of parameters at various depths, insights that are not accessible through traditional multivariate methods.

The combination of data obtained from multifrequency echosounders with the Scanfish remotely operated towed vehicle enabled to establish finely detailed relationships between aggregating marine organisms within Sound Scattering Layers (SSL) and their pelagic

environment. These complex relationships, which are not observable through conventional multivariate statistical methods, emphasize the frequency-dependent aspects of micronektonic SSL detection. We underscore the importance of transitioning towards broadband data acquisition, as recommended by Blanluet et al. (2019).

Our study highlights the significant advantage of employing FDA in comparison to traditional multivariate statistical methods when investigating the impact of environmental variations on SSL. Leveraging our original biogeographical three-dimensional datasets, we have demonstrated the capacity of FDA to capture nuanced parameter variations and enhance the spatial characterization of SSL by incorporating geographic variables. Through the application of FDA techniques, we have showed intricate patterns and variations across different variables at various depths and frequencies, providing insights that are beyond the reach of multivariate statistics.

This approach has enabled us to categorize variables into ecologically meaningful components, yielding valuable insights into the multifactorial impacts of both biotic and abiotic factors on SSL. The substantial enhancements in model representations achieved through the FDA mark a pivotal methodological advancement for ecological studies. This discovery encourages further applications of the FDA in examining the influence of climate change and seasonal fluctuations on marine organisms. Understanding these dynamics is essential for comprehending the catchability of exploited fish resources and the role of SSL in the ocean's biological pump, which aids in carbon sequestration.

Beyond the scope of our specific case study, the adoption of FDA holds promise for a broad spectrum of ecological studies involving extensive spatial data. Embracing this approach can lead to a deeper comprehension of complex ecological processes and can inform more effective conservation and management strategies aimed at preserving the delicate balance of marine ecosystems.

Our research strongly advocates for FDA as a powerful tool for unraveling ecological intricacies and promoting sustainable practices in marine conservation efforts and beyond. By embracing innovative statistical methodologies, we can better address pressing environmental challenges and safeguard the biodiversity and health of our planet's marine resources.

CRediT authorship contribution statement

Yoba Kande: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Ndague Diogoul:** Writing – review & editing, Investigation, Formal analysis. **Patrice Brehmer:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis. **Sophie Dabo-Niang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis. **Papa Ngom:** Supervision, Methodology, Investigation, Formal analysis. **Yannick Perrot:** Software, Data curation.

Data availability

The authors do not have permission to share data.

Acknowledgements

The work received support from Campus France (Senegal) and EMS-Simons for Africa. The early stages of this study were made possible by the AWA project (IRD-BMBF, grant 01DG12073E; “ecosystem approach to the management of fisheries and the marine environment in West African waters”), which was implemented by the Sub Regional Fisheries Commission (SRFC). The authors extend their heartfelt appreciation to the crew of FRV Thalassa (Ifremer) for their support during the AWA survey (doi: 10.17600/14001400). Special thanks also go to the staff of ISRA and CRODT for their support during Yoba Kande's stay in Senegal. Additionally, we acknowledge the assistance provided by the Triatlas and nextGEMS projects, funded through the European Union's Horizon 2020 Research and Innovation Program under grant agreement number 817578 and 101003470.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102547.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. In: Wiley Interdisciplinary Reviews: Computational Statistics, 2, pp. 433–459.
Acar-Denizli, N., Delicado, P., Başarr, G., Caballero, I., 2018. Functional regression on remote sensing data in oceanography. Environ. Ecol. Stat. 25, 277–304.
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. Sci. Adv. 3, e1602468.
Ariza, A., Lebourges-Dhaussy, A., Nerini, D., Pauthenet, E., Roudaut, G., Assunção, R., Tosetto, E., Bertrand, A., 2022a. a. Acoustic seascape partitioning through functional data analysis. J. Biogeogr. 1–15. https://doi.org/10.1111/jbi.14534.
Ariza, A., Lengaigne, M., Menkes, C., Lebourges-Dhaussy, A., Receveur, A., Gorgues, T., Habasque, J., Gutiérrez, M., Maury, O., Bertrand, A., 2022b. b. Global decline of pelagic fauna in a warmer ocean. Nat. Clim. Chang. 12, 928–934.
Ash, R.B., Gardner, M.F., 1975. Topics in Stochastic Processes. Academic Press, London.
Assunção, R.V., Silva, A.C., Roy, A., Bourles, B., Silva, C.H.S., Ternon, J.F., Araujo, M., Bertrand, A., 2020. 3d characterisation of the thermohaline structure in the southwestern tropical Atlantic derived from functional data analysis of in situ profiles. Prog. Oceanogr. 187, 102399.
Auger, P.A., Gorgues, T., Machu, E., Aumont, O., Brehmer, P., 2016. What drives the spatial variability of primary productivity and matter fluxes in the north-west african upwelling system? A modelling approach. Biogeosciences 13, 6419–6440.
Balde, B.S., Fall, M., Kantoussan, J., Sow, F.N., Diouf, M., Brehmer, P., 2019. Fish-length based indicators for improved management of the sardinella fisheries in Senegal. Reg. Stud. Mar. Sci. 31, 100801.
Bayle, S., Monestiez, P., Guinet, C., Nerini, D., 2015. Moving toward finer scales in oceanography: predictive linear functional model of chlorophyll a profile from light data. Prog. Oceanogr. 134, 221–231.
Berge, J., Cottier, F., Varpe, Ø., Renaud, P.E., Falk-Petersen, S., Kwasniewski, S., Griffiths, C., Søreide, J.E., Johnsen, G., Aubert, A., et al., 2014. Arctic complexity: a case study on diel vertical migration of zooplankton. J. Plankton Res. 36, 1279–1297.
Bianchi, D., Mislan, K., 2016. Global patterns of diel vertical migration times and velocities from acoustic data. Limnol. Oceanogr. 61, 353–364.
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. Glob. Biogeochem. Cycles 27, 478–491.
Blanluet, A., Doray, M., Berger, L., Romagnan, J.B., Le Bouffant, N., Lehuta, S., Petitgas, P., 2019. Characterization of sound scattering layers in the bay of Biscay using broadband acoustics, nets and video. PLoS One 14, e0223618.
Brehmer, P., Sancho, G., Trygonis, V., Itano, D., Dalen, J., Fuchs, A., Faraj, A., Taquet, M., 2019. Towards an autonomous pelagic observatory: experiences from monitoring fish communities around drifting fads. Thalassas Int. J. Mar. Sci. 35, 177–189. https://doi.org/10.1007/s41208-018-0107-9.
Brown, J., Fernand, L., Hill, A.E., 1996. Scanfish: high performance towed undulator. Seal. Technol. 37, 23–28.
Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. Stat. Prob. Lett. 45, 11–22.
Cardot, H., Ferraty, F., Sarda, P., 2003. Spline estimators for the functional linear model. Stat. Sin. 571–591.
Chatfield, C., Collins, A.J., 1980. Principal Component Analysis, 13. Springer US, Boston, MA, pp. 57–81. https://doi.org/10.1007/978-1-4899-3184-9_4.
Crambes, C., Mas, A., 2013. Asymptotics of prediction in functional linear regression with functional outputs. Bernoulli 19 (5B), 2627–2651.
Cui, J.S., Lv, P.Y., 2014. Turbidity effect on the fluorescence determination of chlorophyll-a in water. In: Applied Mechanics and Materials, Trans Tech Publ., pp. 60–63.
Cushing, D., 1951. The vertical migration of planktonic crustacea. Biol. Rev. 26, 158–192.
Dall’Olmo, G., Gitelson, A.A., 2005. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. Appl. Opt. 44, 412–422.
Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., McLoone, S., 2013. Principal component analysis on spatial data: an overview. Ann. Assoc. Am. Geogr. 103, 106–128.
Diogoul, N., Brehmer, P., Perrot, Y., Tiedemann, M., Thiam, A., El Ayoubi, S., Mouget, A., Migayrou, C., Sadio, O., Sarré, A., 2020. Fine-scale vertical structure of sound-scattering layers over an east border upwelling system and its relationship to pelagic habitat characteristics. Ocean Sci. 16, 65–81.
Diogoul, N., Brehmer, P., Demarcq, H., El Ayoubi, S., Thiam, A., Sarre, A., Mouget, A., Perrot, Y., 2021. On the robustness of an eastern boundary upwelling ecosystem exposed to multiple stressors. Sci. Rep. 11, 1908.

Diogoul, N., Brehmer, P., Perrot, Y., Rodrigues, E., Thiam, A., El Ayoubi, S., Mouget, A., Sarré, A., Lebourges-Dhaussy, A., 2022. A bi-frequency discrimination method of copepods in the senegalese coast. In: Jech, Michael (Ed.), ICES 2022. Working Group of Fisheries Acoustics, Science and Technology (WGFAST), 4. ICES Scientific Reports, p. 54.

Embling, C.B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K.C., Scott, B.E., 2012. Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. J. Appl. Ecol. 49, 481–492.

Estrade, P., Marchesiello, P., De Verdière, A.C., Roy, C., 2008. Cross-shelf structure of coastal upwelling: a two–dimensional extension of ekman's theory and a mechanism for inner shelf upwelling shut down. J. Mar. Res. 66, 589–616.

Farrell, H., Gentien, P., Fernand, L., Lunven, M., Reguera, B., González-Gil, S., Raine, R., 2012. Scales characterising a high density thin layer of *Dinophysis acuta* Ehrenberg and its transport within a coastal jet, 15, 36–46. https://doi.org/10.1016/j.hal.2011.11.003.

Faye, S., Lazar, A., Sow, B.A., Gaye, A.T., 2015. A model study of the seasonality of sea surface temperature and circulation in the Atlantic North-eastern tropical upwelling system. Front. Phys. 3, 76.

Febrero-Bande, M., González-Manteiga, W., 2013. Generalized additive models for functional data. Test 22, 278–292.

Godard, M., 2021. Caractérisation du comportement alimentaire de l'éléphant de mer Mirounga leonina et liens avec les structures physiques sub-mésoéchelles (1-10km) dans l'océan Austral: une approche par analyse de données fonctionnelles. Ph.D. thesis. Aix-Marseille.

Gong, M., Miller, C., Scott, M., 2015. Functional pca for remotely sensed lake surface water temperature data. Procedia Environ. Sci. 26, 127–130.

Granato, D., Santos, J.S., Escher, G.B., Ferreira, B.L., Maggio, R.M., 2018. Use of principal component analysis (pca) and hierarchical cluster analysis (hca) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective. Trends Food Sci. Technol. 72, 83–90. https://doi.org/10.1016/j.tifs.2017.12.006.

Guillard, J., Lebourges-Dhaussy, A., Brehmer, P., 2004. Simultaneous sv and ts measurements on young-of-the-year (yoy) freshwater fish using three frequencies. ICES J. Mar. Sci. 61, 267–273.

Haney, J.F., 1988. Diel patterns of zooplankton behavior. Bull. Mar. Sci. 43, 583–603.

Happ, C., Greven, S., 2018. Multivariate functional principal component analysis for data observed on different (dimensional) domains. J. Am. Stat. Assoc. 113, 649–659.

Happ-Kurz, C., 2020. Object-oriented software for functional data. J. Stat. Softw. 93 https://doi.org/10.18637/jss.v093.i05.

Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. London Chapman and Hall Inc.

Hays, G.C., Richardson, A.J., Robinson, C., 2005. Climate change and marine plankton. Trends Ecol. Evol. 20, 337–344.

Henderson, B., 2006. Exploring between site differences in water quality trends: a functional data analysis approach. Environmetrics 17, 65–80.

Hörmann, S., Kidziński, Ł., Hallin, M., 2015. Dynamic functional principal components. J. R. Stat. Soc. Ser. B Stat Methodol. 77, 319–348.

Jolliffe, I., 2005. Principal component analysis. In: Encyclopedia of Statistics in Behavioral Science (eds B.S. Everitt and D.C. Howell). https://doi.org/10.1111/jbi.14534.

Kang, M., Fajaryanti, R., Lee, K., Yoon, E.A., Oh, W.S., Min, E., Shin, Y.J., Choi, Y.S., Yi, B.H., Zhang, H., et al., 2021. Geospatial and acoustic application in an artificial reef site of South Korea. J. Mar. Sci. Technol. 29, 9.

Klevjer, T.A., Irigoien, X., Røstad, A., Fraile-Nuez, E., Benítez-Barrios, V.M., Kaartvedt, S., 2016. Large scale patterns in vertical distribution and behaviour of mesopelagic scattering layers, 6, p. 19873. https://doi.org/10.1038/srep19873.

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 J. Mar. Sci. 66, 998–1006.

Koner, S., Staicu, A.M., 2023. Second-generation functional data. Ann. Rev. Stat. Appl. 10, 547–572.

Korneliussen, R.J., Ona, E., 2002. An operational system for processing and visualizing multi-frequency acoustic data. ICES J. Mar. Sci. 59, 293–313.

Korte-Stapff, M., Yarger, D., Stoev, S., Hsing, T., 2022. A multivariate functional-data mixture model for spatio-temporal data: inference and cokriging. eprint arXiv: 2211.04012. doi:10.48550/arXiv.2211.04012.

Kuenzer, T., Hörmann, S., Kokoszka, P., 2021. Principal component analysis of spatially indexed functions. J. Am. Stat. Assoc. 116, 1444–1456.

La, H.S., Kang, M., Dahms, H.U., Ha, H.K., Yang, E.J., Lee, H., Kim, Y.N., Chung, K.H., Kang, S.H., 2015. Characteristics of mesozooplankton sound-scattering layer in the pacific summer water, arctic ocean. Deep-Sea Res. II Top. Stud. Oceanogr. 120, 114–123.

Lehodey, P., Conchon, A., Senina, I., Domokos, R., Calmettes, B., Jouanno, J., Hernandez, O., Kloser, R., 2015. Optimization of a micronekton model with acoustic data. ICES J. Mar. Sci. 72, 1399–1412.

Li, Y., Qiu, Y., Xu, Y., 2022. From multivariate to functional data analysis: fundamentals, recent developments, and emerging areas. J. Multivar. Anal. 188, 104806.

Liu, C., Ray, S., Hooker, G., 2017. Functional principal component analysis of spatially correlated data. Stat. Comput. 27, 1639–1654.

MacLennan, D.N., Fernandes, P.G., Dalen, J., 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES J. Mar. Sci. 59, 365–369.

Mair, A.M., Fernandes, P.G., Lebourges-Dhaussy, A., Brierley, A.S., 2005. An investigation into the zooplankton composition of a prominent 38-khz scattering layer in the north sea. J. Plankton Res. 27, 623–633.

McLean, M.W., Hooker, G., Staicu, A.M., Scheipl, F., Ruppert, D., 2014. Functional generalized additive models. J. Comput. Graph. Stat. 23, 249–269.

Mouget, A., Brehmer, P., Perrot, Y., Uanivi, U., Diogoul, N., El Ayoubi, S., Jeyid, M.A., Sarré, A., Béhagle, N., Kouassi, A.M., et al., 2022. Applying acoustic scattering layer descriptors to depict mid-trophic pelagic organisation: the case of Atlantic african large marine ecosystems continental shelf. Fishes 7, 86. https://doi.org/10.3390/fishes7020086.

Müller, H.G., Yao, F., 2008. Functional additive models. J. Am. Stat. Assoc. 103, 1534–1544.

Ndiaye, M., Dabo-Niang, S., Ngom, P., Thiam, N., Fall, M., Brehmer, P., 2020. Nonparametric prediction for spatial dependent functional data: application to demersal coastal fish off Senegal. In: Mathematical Modeling of Random and Deterministic Phenomena, pp. 31–51.

Ndiaye, M., Dabo-Niang, S., Ngom, P., 2022. Nonparametric prediction for spatial dependent functional data under fixed sampling design. Revista Colombiana de Estadstica 45, 391–428.

Ndoye, S., 2016. Fonctionnement dynamique du centre d'upwelling sud-sénégalais: approche par la modélisation réaliste et l'analyse d'observations satellite de température de surface de la mer. Ph.D. thesis. In: Université Pierre et Marie Curie-Paris VI; Université Cheikh Anta Diop (Dakar).

Ndoye, S., Capet, X., Estrade, P., Sow, B., Dagorne, D., Lazar, A., Gaye, A., Brehmer, P., 2014. Sst patterns and dynamics of the southern Senegal-Gambia upwelling center. J. Geophys. Res. Oceans 119, 8315–8335.

Ndoye, S., Capet, X., Estrade, P., Sow, B., Machu, E., Brochier, T., Döring, J., Brehmer, P., 2017. Dynamics of a "low-enrichment high-retention" upwelling center over the southern Senegal shelf. Geophys. Res. Lett. 44, 5034–5043.

Nerini, D., Monestiez, P., Manté, C., 2010. Cokriging for spatial functional data. J. Multivar. Anal. 101, 409–418.

Ohman, M.D., Frost, B.W., Cohen, E.B., 1983. Reverse diel vertical migration: an escape from invertebrate predators. Science 220, 1404–1407.

Pauthenet, E., 2018. Unraveling the Thermohaline Structure of the Southern Ocean Using Functional Data Analysis. Ph.D. thesis. Department of Meteorology, Stockholm University.

Pauthenet, E., Roquet, F., Madec, G., Nerini, D., 2017. A linear decomposition of the southern ocean thermohaline structure. J. Phys. Oceanogr. 47, 29–47.

Pauthenet, E., Roquet, F., Madec, G., Sallée, J.B., Nerini, D., 2019. The thermohaline modes of the global ocean. J. Phys. Oceanogr. 49, 2535–2552.

Perrot, Y., Brehmer, P., Habasque, J., Roudaut, G., Behagle, N., Sarré, A., Lebourges-Dhaussy, A., 2018. Matecho: an open-source tool for processing fisheries acoustics data. Acoustics Australia 46, 241–248.

Proud, R., Cox, M.J., Wotherspoon, S., Brierley, A.S., 2015. A method for identifying sound scattering layers and extracting key characteristics. Methods Ecol. Evol. 6, 1190–1198.

Proud, R., Cox, M.J., Brierley, A.S., 2017. Biogeography of the global ocean's mesopelagic zone. Curr. Biol. 27, 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 J. Mar. Sci. 76, 718–733.

Ramsay, J.O., Dalzell, C.J., 1991. Some tools for functional data analysis. J. R. Stat. Soc. B. Methodol. 53, 539–561.

Ramsay, J., Silverman, B.W., 1997. Functional Data Analysis. Springer, Berlin Heidelberg New York.

Ramsay, J., Silverman, B., 2005. Functional data analysis. In: Springer Series in Statistics, Second edition. ed. Springer New York Springer e-books, New York, NY.

Receveur, A., Menkes, C., Allain, V., Lebourges-Dhaussy, A., Nerini, D., Mangeas, M., Ménard, F., 2020. Seasonal and spatial variability in the vertical distribution of pelagic forage fauna in the southwest pacific. Deep-Sea Res. II Top. Stud. Oceanogr. 175, 104655.

Reyes, A., Giraldo, R., Mateu, J., 2015. Residual kriging for functional spatial prediction of salinity curves. Commun. Stat. Theory Methods 44, 798–809.

Rokach, L., Maimon, O., 2005. Clustering methods. In: Data Mining and Knowledge Discovery Handbook. Springer, pp. 321–352.

Sierra, C., Flor-Blanco, G., Ordoñez, C., Flor, G., Gallego, J.R., 2017. Analyzing coastal environments by means of functional data analysis. Sediment. Geol. 357, 99–108.

Silverman, B., Ramsay, J., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer, New York, NY.

Simmonds, J., MacLennan, D.N., 2005. Fisheries Acoustics: Theory and Practice. Theory and Practice, 2nd edn. Blackwell, Oxford, UK.

Somavilla, R., González-Pola, C., Fernández-Diaz, J., 2017. The warmer the ocean surface, the shallower the mixed layer. H ow much of this is true? J. Geophys. Res. Oceans 122, 7698–7716.

Song, Y., Wang, C., Sun, D., 2022. Both dissolved oxygen and chlorophyll explain the large-scale longitudinal variation of deep scattering layers in the tropical pacific ocean. Front. Mar. Sci. 9, 782032.

Steele, J.H., Collie, J.S., Bisagni, J.J., Gifford, D.J., Fogarty, M.J., Link, J.S., Sullivan, B., Sieracki, M.E., Beet, A.R., Mountain, D.G., et al., 2007. Balancing end-to-end budgets of the georges bank ecosystem. Prog. Oceanogr. 74, 423–448.

Strom, S., Brainard, M., Holmes, J., Olson, M., 2001. Phytoplankton blooms are strongly impacted by microzooplankton grazing in coastal north pacific waters. Mar. Biol. 138, 355–368.

Tarro-Saavedra, J., Sánchez-Carnero, N., Prieto, A., 2020. Comparative study of fda and time series approaches for seabed classification from acoustic curves. Math. Geosci. 52, 669–692.

Thiam, E.H.I., Singh, V., 2002. Space-time-frequency analysis of rainfall, runoff and temperature in the casamance river basin, Southern Senegal, West Africa. Water SA 28, 259–270.

Tiedemann, M., Brehmer, P., 2017. Larval fish assemblages across an upwelling front: indication for active and passive retention. Estuar. Coast. Shelf Sci. 187, 118–133.
Tiedemann, M., Fock, H.O., Brehmer, P., Döring, J., Möllmann, C., 2017. Does upwelling intensity determine larval fish habitats in upwelling ecosystems? The case of Senegal and Mauritania. Fish. Oceanogr. 26, 655–667.
Winzenborg, I., 2011. Spatial Functional Principal Component Analysis and its Application in Diagnostics. Ph.D. thesis. Universität Ulm.
Yao, F., Müller, H.G., Wang, J.L., 2005. Functional linear regression analysis for longitudinal data. Ann. Stat. 33, 2873–2903.
Yarger, D., Stoev, S., Hsing, T., 2022. A functional-data approach to the argo data. Ann. Appl. Stat. 16, 216–246.
Yoon, W., Nival, P., Choe, S., Picheral, M., Gorsky, G., 2007. Vertical distribution and nutritional behaviour of *cyclothone braueri*, nematoscelis megalops, meganyctiphanes norvegica and *salpa fusiformis* in the nw mediterranean mesopelagic zone. ICES CM F 3, 1–28.
Zar, J.H., 2010. Biostatistical Analysis, 5th edition ed. Prentice-Hall/Pearson.