Spatial environmental analyses using functional approaches: application to multifrequential fisheries acoustics data

Yoba Kande

UNIVERSITE CHEIKH ANTA DIOP DE DAKAR

Université de Lille

ECOLE DOCTORALE DE MATHEMATIQUES ET INFORMATIQUE (ED-MI)

ANNEE: 2023 N°d'ordre: 224

THESE DE DOCTORAT UNIQUE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ CHEIKH ANTA DIOP

MENTION: Mathématiques et Modélisation

SPÉCIALITÉ : Analyse, Statistiques et Applications

Et

DOCTEUR DE L'UNIVERSITÉ DE LILLE

MENTION: Mathématiques et leurs interactions

présentée par :

NOM: KANDE

PRENOM: Yoba

Title: Spatial environmental analyses using functional approaches: application to multifrequential fisheries acoustics data

---

Soutenue publiquement le 22 Décembre 2023

devant le jury composé de

Présidente du jury:	GUEYE DIAGNE Salimata	Professeure	Université Cheikh Anta Diop de Dakar
Rapporteurs:	DUPUY Jean-Francois	Professeur	INSA de Rennes
	KA DIONGUE Abdou	Professeur	Université Gaston Berger
	NERINI David	Maître de conférences	Université Aix-Marseille
Examinateur:	DIOP Aba	Maître de conférences	Université Alioune Diop de Bambey
Directeurs de thèse:	DABO-NIANG Sophie	Professeure	Université de Lille
	NGOM Papa	Professeur	Université Cheikh Anta DIOP de Dakar
	BREHMER Patrice	Chargé de recherche, HDR	IRD
Invités:	NDIAYE Fagueye	Maîtresse de conférences	Université Cheikh Anta Diop de Dakar
	DIOGOUL Ndague	Chercheur	ISRA/CRODT-IRD

Résumé en français

Au sein du domaine de la statistique fonctionnelle appliquée à l'environnement, l'accent de cette recherche est mis sur l'analyse de données présentées sous forme fonctionnelle. La statistique fonctionnelle explore un secteur de la statistique dédié à la manipulation de données fonctionnelles, fournissant des méthodes pour la réduction de dimension, l'apprentissage supervisé et non supervisé, tout en prenant en compte les dépendances temporelles et/ou spatiales inhérentes à ces données. L'essor des technologies modernes a rendu ces types de données de plus en plus accessibles, en particulier dans des domaines tels que les sciences de l'environnement. Un exemple concret d'application de la statistique fonctionnelle réside dans les techniques d'acoustique des pêches, qui permettent d'obtenir des échantillons spatiaux et temporels d'organismes marins à différentes profondeurs et échelles spatiales, sans nécessiter d'intrusion.

Au sein de cette recherche, un ensemble de données acoustiques multi-fréquences, extraites par des échosondeurs scientifiques halieutiques, a été minutieusement analysé pour explorer la structure spatiale des agrégations d'organismes marins micronectoniques, communément désignées sous le terme de couches diffusantes et en anglais de "Sound Scattering Layers" (SSL). L'examen des caractéristiques de ces entités biologiques complexes, telles que leur épaisseur, leur densité relative, et leur profondeur, a été réalisé en corrélation avec leur environnement pélagique. La représentation fine de cet environnement a été rendue possible grâce à l'utilisation d'un système multiparamétrique tracté derrière le navire (ScanFish). Dans cette démarche, nous avons initié l'analyse en recourant à des méthodes standards de statistique multivariée, pour ensuite exploiter des techniques de l'analyse de données fonctionnelles, avec ou sans prise en compte de la dimension spatiale.

Dans notre première analyse exploratoire, l'Analyse en Composantes Principales Fonctionnelle multivariée a fourni des informations précises sur la variation des paramètres le long des profondeurs, contrairement à l'Analyse en Composantes Principales classique. Dans le cadre des tâches de régression, nos analyses, qu'elles intègrent ou non la dimension spatiale, ont mis en évidence des interactions entre les caractéristiques des SSL et les variables environnementales clés à l'échelle spatiale. Des distinctions géographiques significatives ont été constatées entre les SSL de notre jeu de données, i.e. entre les zones septentrionales et méridionales, ainsi qu'entre ceux des zones côtières et hauturières au cours de cette étude. Ces conclusions demeurent pertinentes dans la seconde partie de la thèse, où une méthode d'apprentissage supervisé récente est employée, exploitant la notion de signatures extraites des données environnementales. La dernière contribution, plutôt méthodologique, introduit une nouvelle approche

d'analyse en composante principale pour les données fonctionnelles multivariées spatiales. Comparée à l'analyse en composante principale fonctionnelle multivariée classique, cette méthodologie est adaptée à l'exploration et à la réduction de la dimension des données spatiales dépendantes.

Les conclusions de la thèse mettent en lumière l'importance cruciale de l'analyse statistique spatiale-fonctionnelle dans les recherches écologiques portant sur des entités spatialement complexes. Ces résultats mettent en évidence la valeur ajoutée de la prise en compte de la dimension spatiale dans l'analyse de ces phénomènes biologiques complexes.

Au-delà de notre étude de cas spécifique, l'application de l'analyse de données fonctionnelles ouvre des perspectives prometteuses pour un large éventail d'études écologiques impliquant des données spatiales massives.

Abstract

In the field of functional statistics applied to the environment, the focus of this research lies in the analysis of data presented in functional form. Functional statistics explores a sector of statistics dedicated to handling functional data, providing methods for dimension reduction, supervised and unsupervised learning, while taking into account the temporal and/or spatial dependencies inherent in such data. The rise of modern technologies has made these types of data increasingly accessible, especially in fields such as environmental sciences. A concrete example of the application of functional statistics lies in fisheries acoustics techniques, which enable the collection of spatial and temporal samples of marine organisms at different depths and spatial scales, without requiring intrusion.

Within this research, a set of multi-frequency acoustic data, extracted by scientific fishery echosounders, has been meticulously analyzed to explore the spatial structure of aggregations of micronektonic marine organisms, commonly referred to as "Sound Scattering Layers" (SSL). The examination of the characteristics of these complex biological entities, such as their thickness, relative density, and depth, was conducted in correlation with their pelagic environment. The detailed representation of this environment was made possible through the use of a multiparametric system towed behind the vessel (ScanFish). In this approach, we initiated the analysis using standard multivariate statistical methods, and then exploited techniques of functional data analysis, with or without consideration of spatial dimension.

In our initial exploratory analysis, multivariate Functional Principal Component Analysis provided precise information about parameter variation along depths, unlike classical Principal Component Analysis. In the context of regression tasks, our analyses, whether integrating spatial dimension or not, revealed interactions between SSL characteristics and key environmental variables on a spatial scale. Significant geographical distinctions were observed among SSLs in our dataset, i.e., between northern and southern zones, as well as between coastal and offshore zones during this study. These conclusions remain relevant in the second part of the thesis, where a recent supervised learning method is used, leveraging the concept of signatures extracted from environmental data. The final, more methodological contribution introduces a new approach to principal component analysis for multivariate spatial functional data. Compared to multivariate functional principal component analysis, this methodology is suited for exploring and reducing the dimensionality of dependent spatial data.
The thesis findings highlight the crucial importance of spatial-functional statistical analysis in ecological research involving spatially complex entities. These results underscore the added value of considering

spatial dimension in the analysis of these complex biological phenomena.

Beyond our specific case study, the application of functional data analysis opens promising avenues for a wide range of ecological studies involving massive spatial data.

Dédicaces

C'est avec une immense gratitude que je dédie ce travail à mes parents, Saidou KANDE et Coumba BALDE, dont le soutien indéfectible a été une source constante d'inspiration tout au long de mes études. Je dédie ce travail également à:
Mes encadrants;
Mon merveilleux mari;
Mes frères et mes sœurs;
Mes oncles et mes tantes;
Mes amis et mes camarades de promotion.

Remerciements

Tout d'abord, mes sincères remerciements à Allah qui m'a donné la santé et la force nécessaires pour achever ce travail.
Je tiens à exprimer toute ma reconnaissance et ma plus profonde gratitude à ma directrice de thèse, Pr. Sophie DABO-NIANG. Je la remercie de m'avoir encadré, orienté, aidé et conseillé. Votre soutien, vos conseils et votre expertise ont été inestimables tout au long de mon parcours de recherche. Je tiens également à vous remercier pour votre disponibilité, même lorsque votre emploi du temps était chargé. Je suis honorée d'avoir eu la chance de travailler sous votre direction, et je suis enthousiaste à l'idée de poursuivre notre collaboration. Votre influence a façonné mon parcours académique de manière significative, et je suis convaincue que les enseignements que j'ai reçus resteront avec moi pour le reste de ma carrière. Je souhaite adresser mes vifs remerciements au Dr. Patrice BREHMER qui a co-dirigé cette thèse. J'ai eu le privilège de collaborer avec lui au cours de ces dernières années, une expérience qui a été enrichissante à bien des égards. Ce travail aurait été impossible sans son soutien constant et son enthousiasme infaillible. Sa disponibilité et sa bienveillance tout au long de la réalisation de cette thèse ont été inestimables. Je lui adresse mes remerciements les plus sincères pour sa patience, son inspiration, et le temps précieux qu'il a généreusement consacré à cette entreprise. Je tiens à exprimer ma gratitude envers mon directeur de thèse à l'université Cheikh Anta Diop de Dakar, le Pr. Papa NGOM, pour sa simplicité, sa disponibilité et les connaissances qu'il m'a transmises dans ce travail. Votre soutien restera gravé dans ma mémoire.

Je tiens également à exprimer ma gratitude envers les membres du jury qui ont eu la gentillesse d'évaluer ce travail avec attention: Pr. Salimata GUEYE DIAGNE pour avoir accepté de présider ce jury; Pr. Jean Francois DUPUY, Pr. Abdou KA DIONGUE et Dr. David NERINI pour avoir accepté d'être rapporteurs de cette thèse; Dr. Aba DIOP pour avoir accepté d'examiner cette thèse; Dr. Fagueye NDIAYE et Dr. Ndague DIOGOUL pour avoir accepté d'assister à la présentation de ce travail.

Je tiens à remercier également le Dr. Ndague DIOGOUL et le Dr. Yannick PERROT du LEMAR (Laboratoire des sciences de l'environnement marin) basé en France, pour leur précieuse aide dans la compréhension des données acoustiques et l'interprétation biologique. Je remercie les doctorants Idris SI-AHMED et Christelle Judith AGONKOUI pour leur collaboration précieuse et leur contribution

essentielle à la rédaction du cinquième chapitre.

Je tiens ainsi à exprimer ma reconnaissance envers les différentes structures d'accueil où j'ai séjourné pendant cette thèse, notamment le CRODT et le laboratoire Paul Painlevé. Je les remercie pour l'accueil chaleureux et les excellentes conditions de travail qu'ils m'ont offertes.

Je tiens à exprimer ma gratitude envers le Directeur du CRODT, le Dr. Ndiaga THIAM, ainsi qu'envers les chercheurs, doctorants et l'ensemble du personnel administratif du CRODT. Leur soutien, leurs encouragements et leurs conseils avisés ont été d'une grande importance pour moi.

Je remercie Campus France (Sénégal) et EMS-Simons for Africa pour m'avoir octroyé une bourse de séjour doctoral à l'université de Lille.

J'adresse mes sincères remerciements à tous les professeurs, intervenants et à toutes les personnes qui, par leurs paroles, leurs écrits, leurs conseils et leurs critiques, ont guidé mes réflexions. Je suis également reconnaissante à ceux qui ont accepté de me rencontrer et de répondre à mes questions durant mes recherches.

Je remercie du fond du cœur mes très chers parents, Saidou KANDE et Coumba BALDE, qui ont toujours été là pour moi. Vous avez tout sacrifié pour vos enfants, n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je suis redevable d'une éducation dont je suis fière. Je suis également reconnaissante envers mon très cher mari Abda BA, ma tante Mariama, mes frères Mamadou, Abdoulaye, Papis, Moussa et Ousmane, ainsi que mes sœurs Bouya et Nabou, pour leur encouragement et leur soutien.

Enfin, je tiens à remercier tous mes ami(e)s pour leur sincère amitié et leur confiance, à qui je dois ma reconnaissance et mon attachement.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

Contents

List of Acronyms and Abbreviations

AWA:	Ecosystem Approach to the management of fisheries and the marine environment in West African waters
AIC:	Akaike's Information Criterion
BIC:	Bayesian Information Criterion
CHA:	Cluster Hierarchical Analysis
CRODT:	Centre de Recherche Océanographique de Dakar-Thiaroye
CTD:	Conductivity, Temperature, Depth
dB:	Decibel
ESU:	Elementary Sampling Unit
FDA:	Functional Data Analysis
FGSAM:	Functionnal Generalized Spectral Additive Model
FGSAMs:	Spatial Functionnal Generalized Spectral Additive Model
FLM:	Functional Linear Model
FPCA:	Functional Principal Component Analysis
FRV:	Fisheries Research Vessel
FPLS:	Functional Partial Least Squares
GAM:	Generalized additive model
GAMs:	Spatial Generalized additive model
GCV:	Generalised Cross-Validation
GFLM:	Generalized Functional linear model
GPS:	Global Positioning System
GSAM:	Generalized Spectral Additive Model
IRD:	Institut de Recherche pour le Développement
ISRA:	Institut Sénégalais de Recherche Agricole
LRspca:	Linear Regression on spatial principal component scores
MAE :	Mean Absolute Error
MFPCA:	Multivariate Functional Principal Component Analysis
NMSE:	Normalized Mean Squared Error
NNs:	Spatial Neural Network

NN:	Neural Network
NNspca :	Neural Network on spatial principal component scores
NOAA:	National Oceanic and Atmospheric Administration
NTU:	Nephelometric Turbidity Unit
PCA:	Principal Component Analysis
PC1 and PC2:	First and Second Principal Component
PCR :	Principal Component Regression
PCRs:	Spatial Principal Component Regression
PSU:	Practical Salinity Unit
R:	Ridge
RMSE:	Root Mean Square Error
Rs:	Spatial Ridge
SHAP:	SHapley Additive exPlanations
SIM1 and SIM2:	First and Second simulated functional variable
sPCA:	Spatial Principal Component Analysis
SFPCA:	Spatial Functional Principal Component Analysis
SMFPCA:	Spatial Multivariate Functionnal Principal Component Analysis
SSL:	Sound Scattering Layer
SST:	Sea Surface Temperature
TMP:	Sea surface temperature variable
TS:	Target Strength

Notations

$\mathbb{N}$	set of natural numbers: $0, 1, 2 \ldots$
$\mathbb{N}^*$	set of non-zero natural numbers: $1, 2 \ldots$
$\mathbb{Z}$	set of integers: $\ldots, -1, 0, 1, \ldots$
$\mathbb{R}$	set of real numbers: $]-\infty, +\infty[$
$\mathbb{R}_+$	set of real positives numbers: $[0, +\infty[$
$\mathbb{R}^d$	euclidian space of dimension d
$\lfloor \cdot \rfloor$	integer part
$\lvert \cdot \rvert$	absolute value if the argument is number
	or determinant if the argument is matrix
$\lVert \cdot \rVert$	norm such that:
	if the argument is a vector $x \in \mathbb{R}^d$: $\lVert x \rVert = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$
	if the argument is a matrix A: $\lVert A \rVert = \sqrt{\sum \sum a_{ij}^2}$
	if the argument is a function f: $\lVert f \rVert = \sup \lvert f(x) \rvert$
$x^{'}$ or x^T	transpose of vector or matrix x
$\mathrm{tr}(\cdot)$	trace of matrix
$\otimes$	Kronecker product
$\varnothing$	empty set
$\mathrm{dist}(A, B)$	euclidian distance between A and B
$\mathbb{I}(\cdot)$ (or $\mathbb{I}_A(\cdot)$)	indicator function (of set A)
$L^2(\mathcal{T})$	space of square-integrable functions in interval $\mathcal{T}$
$\sigma(\ldots)$	σ-algebra generated by $(\ldots)$
$(\Omega, \mathcal{A}, P)$	probability space. Ω : nonempty set, $\mathcal{A}$:σ-algebra of subset of Ω
	and P: probability measure on $\mathcal{A}$
i.i.d	independent and identically distributed
$\mathcal{N}(0, 1)$	normal distribution
$u_n = O(v_n)$	a constant c exists such that $u_n \leq c v_n$
$u_n = o(v_n)$	$\frac{u_n}{v_n} \to 0$ as $n \to \infty$
S_v	Volume backscattering strength
S_A	Nautical Area Scattering Coefficient

List of Figures

List of Tables

Introduction

The oceans, constituting one of the main reservoirs of biodiversity in the world and occupying 71% of the earth's surface, are considered to be one of the most important natural resources for humans (Costanza, 1999). They make a major contribution to global climate regulation by absorbing about 25-30% of total anthropogenic carbon dioxide (CO_2) emissions (Le Quéré et al., 2018; Gruber et al., 2019; Sabine et al., 2004), while producing almost half of the world's oxygen. The complex nature of marine ecosystem functioning, combined with the inaccessibility and invisibility of the majority of their goods and services, calls for more appropriate and robust approaches to evaluate these resources. It should be noted that including consideration of fine-scale depth structure in biogeographic partitions and ecosystem models, allows to better understand the role of micronektonic communities in pelagic food webs (Proud et al., 2017) and their effect on climate change (Mouget et al., 2022). To establish models for exploiting marine resources and monitoring stocks, it is necessary to obtain fishing statistics over a wide spatial and temporal range, which is often difficult and time-consuming using conventional fishing methods (Brehmer et al., 2020). It is in this context that researchers have turned their attention to alternative methods, including the hydroacoustic methods, which are non-invasive and enable large volumes of water to be sampled rapidly (Simmonds and MacLennan, 2005; Stequert and Gerlotto, 1977; Guillard et al., 2023). Hydroacoustic methods have considerably developed in recent years and are increasingly used to acquire a wide variety of information on the aquatic ecosystem, providing a unique way of perce living in the marine environment (Brehmer et al., 2003; Guillard, 2016), and have become in recent years a major tool for monitoring the marine environment (Brehmer et al., 2019) and assessing stocks of marine species, allowing multiple types of applications such as transmission, measurement, detection, location of targets or obstacles, knowledge of predator-prey relationships etc. (Josso, 2010; Le Gall, 2015; Lurton, 2002; Marchal et al., 1996). They allow the collection of robust, high-resolution quantitative data on a large spatio-temporal scale (Brehmer et al., 2006), to assess the biomass and size distribution of detected biological targets (Burgos and Horne, 2008; Jean et al., 2011; Kim et al., 2018). Acoustic techniques allow not only to cover large geographical areas but also to study areas not suitable for conventional fisheries sampling, while avoiding harm to the marine environment. Underwater acoustics is a technique based on the properties of sound in water (Lurton, 2002). The principle of echolocation consists of emitting a sound pulse into the water at a certain frequency by using an antenna or transducer, which propagates, reflects off any obstacles encountered and returns to the emitting source. This impulse propagates in the form of waves at a speed that

depends on the physical properties of the environment such as pressure, salinity and temperature Guillard and Colon (1998) and creates local pressure variations which represent a flow of energy per unit time, thus indicating the intensity of the wave. These acoustic waves are subject to distortions, changes in amplitude and phase Mours (2017) and are reflected as echoes by any object encountered (bottom, plankton, fish, wreckage, etc.). The intensity of the echo depends particularly on the size of the obstacle, the intensity emitted and its power of reflection. The use of echosounders to locate schools of fish (Brehmer et al., 2002) has made it possible to detect planktonic organisms (Cushing and Richardson, 1956). These are much smaller organisms than fish, which are close together in the water column and whose echoes are produced as " sound scattering layers" (SSL) on the echograms Remond (2015). Recognising and interpreting the echoes of these micro-organisms has become essential for scientists and managers, as plankton form the basis of aquatic food chains and generally mark the boundaries of water bodies. The SSL, essentially composed of zooplanktons and micronectons (Kloser et al., 2002; Béhagle et al., 2017), organisms generally migrate to the surface at night in search of prey and descend to deeper waters during the day to avoid predators, a process known as diel vertical migration (DVM)(Aksnes et al., 2017; Behrenfeld et al., 2019; Longhurst and Harrison, 1988). Even if they are different in size, they are the first animal levels in the marine food web (macro-zooplankton from 2 mm to 2 cm, meso-zooplankton from 0.2 to 2 mm and micro-zooplankton from 20 μm to 0.2 mm and micronekton typically range in size from 2mm to 20 cm, but there is no strict size range that defines micronekton). Zooplanktons play a crucial role in aquatic ecosystems, transporting matter and energy to higher trophic levels and influencing nutrient dynamism. Zooplankton communities are highly sensitive to fluctuations in environmental parameters and are therefore an excellent indicator of change in aquatic systems (Molinero et al., 2005; Siokou-Frangou et al., 1998). Micronekton is a term used to describe a diverse group of small marine organisms that occupy the midwater or mesopelagic zone of the ocean. These organisms are typically larger than plankton but smaller than many larger fish and marine animals. As zooplankton, micronekton play a fundamental role in the marine food web by linking primary consumers to top predators, forming the prey of marine predators such as tuna, whose populations are heavily exploited by fisheries, as well as birds and marine mammals. The spatio-temporal variations of these abiotic factors can alter the abundance of zooplankton and in particular cause a reduction in fishery resources. It has been noted that a high biomass of fish is observed in areas with high plankton production. This is why, in the context of optimising resources, it is necessary to study the impact of environmental parameters on the organisms responsible for these diffusing layers (the SSL) at a fine spatial scale in order to gain a better understanding of how aquatic ecosystems function.

In this perspective, researchers from IRD (Senegal, France), along with various partners, collected acoustic and environmental data in the West African region from various spatial units, covering an ordered grid of depth points within a finite length range. To conduct a comprehensive analysis of this data, taking into account its spatial and functional nature, it is natural to employ spatio-functional statistical methods.

In recent years, Functional Data Analysis (FDA) has gained popularity within the international scientific community, as evidenced by numerous studies in this field (Clarkson, 2005; Genin and Ahmed, 2019; Giraldo et al., 2018; Bouzebda and Nemouchi, 2020; Diogoul et al., 2020),...

To our knowledge, the studies conducted in the Senegal region on sound scattering layers in relation to environmental parameters by Diogoul et al. (2020) are the only ones in the region. However, these studies did not take into account the spatial and functional aspects of the data.

This thesis contributes to the study of the effects of environmental parameters on the descriptors of acoustic layers in West Africa. It utilizes statistical methods of functional data analysis to assess the similarities among spatial units at sea in terms of environmental variables while examining their impact on sound scattering layers, taking spatial aspects into consideration. Few works have been conducted in this specific context.

The remainder of the document is organized as follows: Chapter 1 describes the geographical and biological context of the study area. Chapter 2 deals with fundamental concepts and provides a state-of-the-art overview of the statistical models and methods used. Chapter 3 presents an initial contribution to the study of the effects of environmental parameters on the descriptors of acoustic layers in West Africa, using methods and models such as Principal Component Analysis (PCA), clustering, and functional regression models, while comparing them to classical multivariate methods used in the study by Diogoul et al. (2020). Chapter 4 explores a novel regression method based on the theory of signatures to analyze functional data based on the data examined in Chapter 3. In Chapter 5, we propose a new principal component analysis method for multivariate spatial functional data and investigate its finite sample properties on space-time sea-surface temperature data.

Chapter 1

Biological framework of the study

1.1 Introduction

Senegal is a maritime country located in the extreme west of Africa with 700 km of coastline (Theodore, 1993; Thiao, 2009). Its maritime frontage of 240 nautical miles of latitude oriented in the N-S direction, is comprised between 12°30 and 16°20 (Rebert, 1982). This façade is separated by the Cap Vert peninsula into two sectors with different topographical characteristics (Roy, 1990).

The northern zone or Grande Côte, which extends over the regions of Dakar, Thies, Louga and Saint-Louis with 180 km long (Camara, 2008), constitutes the most unstable section of the coastline, particularly with the presence of a strong swell and a narrow and steep continental shelf (Niang, 2009). Generally oriented North-West, it generates a drift of the coastline towards the south (Niang Diop, 1995) because of the breaking waves arriving on the shore (Seck, 2014).

The southern zone or small coastline (Niang, 2009), between Dakar and the Sangomar point(Diaw et al., 1993), extends over nearly 110 km (Camara, 2008). With a wider continental shelf with a gentle slope (Teisson, 1983) that avoids the formation of large bars, the small coast is relatively more protected from the swell phenomenon than the large coast (Niang, 2009). It is also subject to a process of deep-water resurgence called upwelling, which is more frequent in this area.

Coastal upwelling is a physical phenomenon (Diankha et al., 2015), which results in the resurgence of deep waters, compensating under the action of the northward wind, for the drift of surface waters towards the high sea (Ndoye, 2016; Roy, 1991). This upwelling process of cold deep waters causes a supply of mineral salts (Croquette, 2007), thus favouring the development and maintenance of an important biological production in this area (Roy, 1991).

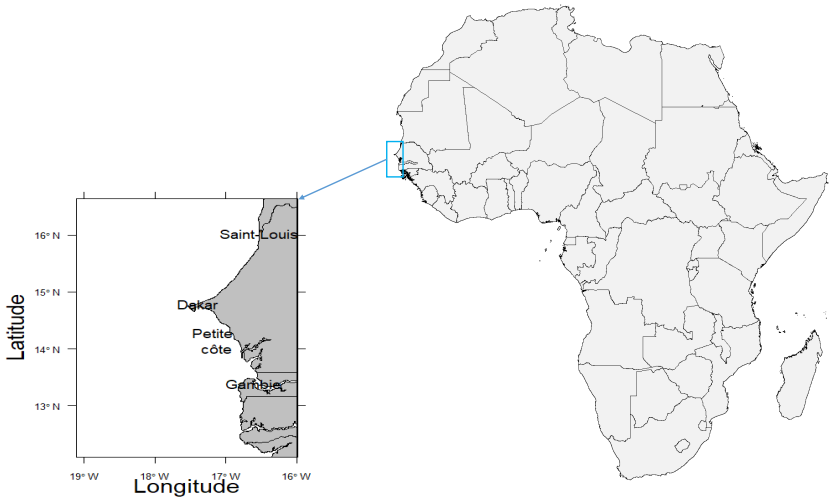


Figure 1.1: Map of the study area, Sénégal (West Africa).

1.2 General concept of underwater acoustics

The propagation of sound in water has always been widely used by many marine organisms to communicate, navigate, survive and locate prey in the ocean. The use of acoustic waves as a means of observing the oceans is relatively recent, even if the first attempts dating back to the end of the XV^{th} century when Leonardo da Vinci listened to the underwater noises of distant ships by plunging a tube into the water and sticking his ear to the end. Unlike electromagnetic waves, sound waves propagate very well in the marine environment and over long distances (Simmonds and MacLennan, 2005). For a better understanding of this study, only the key elementary notions of acoustics will be described to better understand the data we have used along the doctoral thesis. To do so we have mainly used the works of Lurton (2002); Simmonds and MacLennan (2005) for the basis of fisheries acoustics and also other works as Levenez et al. (2006); Leif et al. (2017) or Diner and Marchand (1995) a book on the applications of acoustics to maritime fishing for students.

1.2.1 Sound wave

A sound wave results from the propagation of a mechanical disturbance characterized by a vibration of the molecules of the material environment (air, water, gas or solid) around their equilibrium positions. It therefore corresponds to a succession of oscillations represented by phases of compression and expansion of the particles in this environment. It is described by several parameters that determine the characteristics of the sound produced (Lurton, 2002).

- **The amplitude A** is proportional to the maximum distance a vibrating particle can be moved from rest. It is transmitted from near to near to the particles in the environment in question. The greater the amplitude, the more intense the sound heard. It is the most commonly used parameter and is expressed in Pascal.
- **The period T** indicates the time taken for one of the phenomena of a wave to reproduce itself, i.e. the time elapsed between two similar positions or states of the source or fluid. The period

is expressed in seconds.

- **The frequency F** represents the number of oscillations or vibrations of the wave particles per second. It is expressed in Hertz (Hz) representing the number of cycles per second. The frequency is equivalent to the inverse of the period, defined as follows: $F = 1/T$. In fisheries acoustics, the frequency range used extends from 10 to 500 kHz.

- **The wavelength** λ represents the distance travelled by the wave during one period. In a given area, the wavelength is linked to the frequency and the sound speed by the following formula : $\lambda = c/f = c * T$. It is expressed in meters. The sound speed depends neither on the amplitude nor the frequency of the sound, but on the mechanical characteristics of the propagation area in particular sea temperature and salinity.

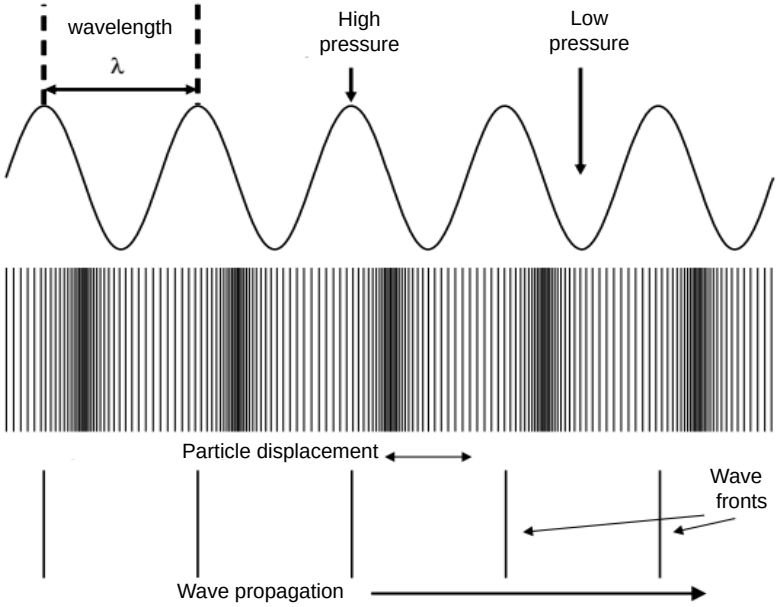


Figure 1.2: Sound waves. (Top): the pressure varies like a sine wave. (Middle): the displacement of the particles is out of phase with the pressure. (Bottom): the wave fronts corresponding to the maximum pressure (Simmonds and MacLennan, 2005).

1.2.2 Energy and sound intensity in acoustic signal

The propagation of an acoustic wave creates local variations in pressure in the environment, which corresponds to a flow of energy per unit of time, thus determining the acoustic intensity. Noted I, it is defined as the average quantity of energy per unit area and time. It is expressed in W/m^2 and is related to pressure by:

$$\mathrm{I} = \frac{\overline{P}^2}{\rho c} \tag{1.1}$$

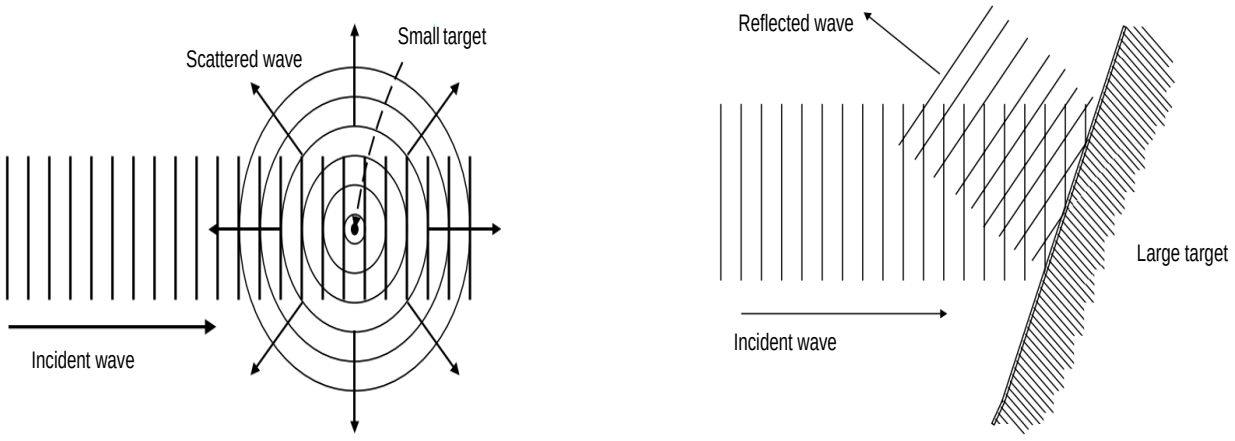
with $\overline{\mathrm{P}}^2$, the average acoustic pressure expressed in $\mu \mathrm{P}_a$; c is the celerity in m/s and ρ represents the density of water (in kg/m^3); the term ρc denotes the acoustic impedance of the environment which

characterises the resistance of the environment to the passage of this wave.
Due to their significant range (difference between the weakest and strongest values), the decibel (dB) is the unit used to logarithmically quantify acoustic quantities, enabling the representation, within a limited range of values, of intensity variations that span multiple orders of magnitude on the arithmetic scale. It is a dimensionless unit and corresponds to 10 times the base 10 logarithm of the ratio of the acoustic intensity to a reference intensity.

$$\mathrm{dB} = 10 * log10(\mathrm{I}/\mathrm{I}_0) \tag{1.2}$$

1.2.3 Acoustic backscatter

Using echosounder to detect and locate biological organisms involves insonifying a known volume of water, i.e. emitting sound waves that are transmitted and received in a well-defined beam. When a sound wave emitted by a transducer and propagating in an environment reaches a target with a density different from that of the propagation environment (e.g. wrecks, bottom, fish, plankton), it is scattered by the target in all directions in space and particular part of the energy is backscattered towards the emitting source, thus constituting the sound echo (Diner and Marchand, 1995). The intensity of this echo depends in particular on the size of the obstacle, the intensity emitted and its power of reflection. As a general rule, for an organism to reflect acoustic energy, its acoustic impedance must differs from that of the transmission medium, i.e. water. The higher the impedance, the more significant the backscattered energy.When the size of the target is considerably smaller than the wavelength ($L \ll \lambda$), the target is completely insonified and backscatter occurs in all directions. In this case, it is the volume of the target rather than its shape that determines the nature of the backscatter. On the other hand, if the size of the target is significantly larger than the wavelength ($L \gg \lambda$) only the part of the object within the insonified volume reflects the signal, in a privileged direction. (Doray, 2006).



A) B)

Figure 1.3: A) Backscatter from a small underwater target. The incident wave propagates in all directions; B) from a large target. The wave is reflected in a privileged direction (Simmonds and MacLennan, 2005).

Each target or reflecting object has an acoustic parameter called the reflection index or Target Strength (TS) expressed in decibels (unit of measurement used to express the relative intensity or power of the acoustic signal i.e. sound). This is a logarithmically quantified measurement, and corresponds to the proportion of incident energy backscattered by a target.

$$\mathrm{TS} = 10log(\mathrm{I}_r/\mathrm{I}_\mathrm{i}) \tag{1.3}$$

I_i and I_r correspond to the intensity of an incident sound wave (reaching one or more targets) and the intensity of the reflected sound wave (by one or more targets),respectively. The unit of the TS is the dBm^2. This reflection index has an arithmetic equivalent called the acoustic reflecting cross section or backscattering cross section, expressed in units of area (Bertrand, 1999).

$$\sigma_{bs} = R^2(\mathrm{I}_r/\mathrm{I}_i) \tag{1.4}$$

with R, the distance between the target and transducer (in meter).

The TS is an essential parameter for assessing the biomass of species, but unfortunately, it is one of the most difficult factors to estimate because it depends on several parameters such as the presence or absence of the species' swim bladder, its size, its orientation in the acoustic beam and the frequency used.We have not used it this work because of the lake of species identification by fishing operation. The backscattering of at least 90% of the acoustic energy of many marine species is mainly due to the presence of a sac located in the abdomen of their anatomy, called the swim bladder, which is full of nitrogen and oxygen, and whose usefulness is, among other things, to enable them to adjust their buoyancy (Foote, 1980; Ona, 1990; Salvetat et al., 1990). When a group of individual targets are in a volume of insonified water, their echoes combine to form a continuous signal that varies as a function of amplitude. It is then no longer possible to consider the energy backscattered by a target but rather an acoustic response per unit volume and surface area. Echo-integration techniques are used to estimate the acoustic densities of marine organisms recorded at a well-defined frequency. It consists of incorporating energy from targets contained in an echointegration cell by setting a threshold, below which echoes are not taken into account. The size of these echointegration cells (or Elementary Sampling Unit: ESU) is chosen by the user (Doray, 2006). Echo-integration techniques have made it possible to obtain two acoustic measurements.

- The volume backscattering coefficient, denoted s_v (m^{-1}), corresponds to the sum of the backscattering from all targets contained within the volume and is obtained by the formula

$$\mathrm{s}_v = \sum(\sigma_{bs}/V_0) \tag{1.5}$$

  with V_0, the sampled volume and σ_{bs} is the equivalent reflecting surface and the sum is obtained over all the targets contributing to the echoes of V_0. Its equivalent in logarithmic scale is expressed in dB re $1\ \mathrm{m}^{-1}$ and noted

$$\mathrm{S}_v = 10log(\mathrm{s}_v) \tag{1.6}$$

The principle of linearity translates into this equation, which states that the density of insonified organisms can be quantified as a function of acoustic intensity, which is proportional to the number of targets sampled in the acoustic beam (Doray, 2006).

- The area backscattering coefficient, denoted S_a, is defined as the integral of the s_v values over a depth range.

$$S_a = \int_{z_1}^{z_2} s_v \, dz \tag{1.7}$$

Converting S_a into nautical miles, it becomes the surface backscatter index per nautical mile (S_A) or Nautical Area Scattering Coefficient (NASC, $m^2\,nmi^{-2}$) (MacLennan et al., 2002).

$$S_A = 4\pi(1852)^2\, S_a \tag{1.8}$$

Acoustic methods can therefore be applied in ecology to estimate the density of marine species in order to assess the availability of resources and study the process of diel migrations and the distribution of these species on a spatial scale.

1.3 What is a Sound Scattering Layers (SSL)?

Sound Scattering Layers (SSL) have long been perceived as noise when interpreting acoustic data for the study of marine ecosystems (Ballón et al., 2011). Moreover, echoes from schools of fish have often been extracted or even filtered from acoustic backscatter data for better analysis (Cabreira et al., 2011; Korneliussen and Ona, 2002). SSL are defined as a group of organisms in the water column that appear on the echogram of an echo sounder and can extend horizontally over tens or even thousands of kilometres (Kloser et al., 2009). The first observations of these layers were made around the 20^{th} century with the detection, via naval echo sounders, of signals that were initially thought to be echoes from the seabed but which were in fact of biological origin (Brierley, 2014). The SSL are generally grouped into two groups: those detected from the surface to a depth of 200 m (epipelagic zone the uppermost layer of the ocean also known as the euphotic zone) and those found between 200 and 1000 m (mesopelagic zone, sunlight does not penetrate, so it is also known as the "twilight zone" because there is very little to no natural light) (Gjøsæter et al., 2020; Ariza et al., 2016) .We can note the existence of the bathypelagic Zone (1000-4000 meters), it is also known as the "midnight zone" because it is entirely devoid of sunlight, and the environment is pitch black (note considered in our work, out of range of the acoustic devices used as well as of sampling area considered). Containing a wide range of species with different sizes, the study of the specific composition of these SSL requires the acquisition of backscatter data at several frequencies (Lawson et al., 2004). It is difficult to determine the exact specific composition of the SSL, but the most dominant groups of species are zooplankton and micronekton. In our study area the copepod was the main species in biomass (Diogoul et al., 2020, 2021).

Chapter 2

General mathematical concepts

This chapter gives a general introduction and state of art on the main contributions of this thesis on the two fundamental concepts; Functional Data Analysis (FDA) and Spatial Statistics.

2.1 Functional data analysis (FDA)

Large and complex data with a dynamic component of space and/or time abound in many fields, in particular in the description of atmospheric, hydrological and oceanological systems, where the study of relationships between variables made up of high dimensional vectors and/or functional components is of prime importance for understanding the functioning of these natural systems. The application of classical multivariate analysis methods for the treatment of these types of data comes up against certain major problems such as the weakness of the sample size in relation to the dimension of the variable, in addition they do not take into account the continuity of the ordered character of the discretization (Zullo, 2016). Functional Data Analysis (FDA) transforms very high dimensional data into functional data, i.e. objects such as curves, shapes, images or any more complex mathematical object, thought as realisations of a stochastic process (Ramsay and Silverman, 1997).
FDA has become very popular and applied in many fields. For example, on measurements or images collected by satellites, methods from functional statistics have been applied. We can mention among

others Dabo-Niang et al. (2004, 2007) who are interested in classifying curves collected by the satellite at different locations in the Amazon, Cardot et al. (2003a) which study the evolution of vegetation using satellite data. In econometrics, FDA methods can model the volatility of financial markets (Müller et al., 2011; Alva et al., 2009), the evolution of online auction prices (Wang et al., 2008), the study of the relationship between unmet basic needs and gross domestic product (Pineda-Ríos et al., 2019). Functional modelling is very well suited to the study of climatological phenomena, as the individuals observed generally exhibit spatio-temporal variability. Fraiman et al. (2014) study trends in Antarctic temperature time series using FDA methods, Suhaila (2021) proves that the functional framework is more flexible in representing the climate phenomenon El Niño in its entirety. In the medical field, functional data analysis methods for modelling and forecasting data on a range of health and demographic issues have considerable advantages for better understanding the relationships between risk factors and the effectiveness of preventive measures (Hyndman and Ullah, 2007; Erbas et al., 2007). For more details and recent references on FDA, we refer to the monographs of (Ramsay and Silverman, 1997; Ferraty and Vieu, 2006a; Kokoszka and Reimherr, 2017; Bosq, 2000) and the review of Koner and Staicu (2023).

2.1.1 From discrete data to functions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a space of probability, $\mathcal{F}$ a space of functions (e.g. a separable Banach space).
A functional random variable is a variable

$$X = \{X(t), t \in \mathcal{T}\} : \Omega \to \mathcal{F}$$

taking values in $\mathcal{F}$ (of eventually infinite dimension).
A functional data is an observation of the functional random variable X.

If $\mathcal{T} \subseteq \mathbb{R}$ then X is a curve while an image may be considered as a functional data in the case where $\mathcal{T} \subseteq \mathbb{R}^2$. If $\mathcal{T} \subseteq \mathbb{R}^d$ $(d > 2)$, X has a more complex structure. The commonly used functional space is $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^P)$, $P \geq 1$; the space of P-dimensional vector-valued square-integrable functions on $\mathcal{T}$.
In practice, we only have discrete data at a finite number of points (Delsol, 2008) of the form:

$$x_i(t_l), \quad t_l \in \mathcal{T}, \quad i = 1, \ldots, n \quad l = 1 \ldots, L$$

Then, a first step is to transform the $x_i(t_l)$ as n observations $\{X_i(t), t \in \mathcal{T}\}$ of the functional random variable $\{X(t), t \in \mathcal{T}\}$, valued in a functional space (an Hilbert (e.g., $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^P)$) or a Banach space).

The most used approach is to express the curves in a space of finite dimension generated by a basis of functions. If we consider a functional random variable $X(t)$ evolving in an infinite-dimensional space with a countable basis $\phi_k(t)$, then we can find a sequence of real c_k such as:

$$X(t) = \sum_{k=0}^{\infty} c_k \phi_k(t), t \in \mathcal{T} \tag{2.1}$$

The functional object can be approximated by a linear combination of the first K elements of the

chosen basis:

$$X(t) \approx \sum_{k=1}^{K} c_k \phi_k(t), t \in \mathcal{T} \tag{2.2}$$

We thus obtain an approximation belonging to a finite dimensional subspace K of $\mathcal{F}$. The choice of ϕ_k depends on the assumptions about the variable to be studied. The function bases can be grouped into two groups:

- Fixed bases consisting of functions defined on a grid of nodes.
  – The trigonometric or Fourier basis is probably the best known of the Hilbertian bases, and the most appropriate for approximating random curves with periodic behaviour (Ramsay and Silverman, 2002a).
  – The basis of polynomial functions defined on sub-intervals of $\mathbb{R}$ called B-spline. A flexible base suitable for curve smoothing (Cardot et al., 2003b).
  – The wavelet basis (Antoniadis and Sapatinas, 2003) used for signal dimension reduction. Wavelets combine the advantages of the Fourier basis and splines (localised bases).

- The basis built from the information provided by the data.

  – The functional principal components which constitute a basis of functions constructed using sequential projections onto subspaces of $\mathcal{L}^2(\mathcal{T}, \mathbb{R}^p)$.

  – Functional Partial Least Squares (FPLS) (Cardot and Sarda, 2005).

2.1.2 Functional Principal Component Analysis (FPCA)

Functional principal component analysis is one of the first methods used for dimension reduction in FDA. It was introduced by Rao (1958) in order to compare growth curves by applying the Karhunen-Loève expansion (Karhunen, 1946) to these data. FPCA allows the covariance structure to be explored, revealing the main sources of variation in the dataset (Ramsay and Silverman, 2005b). A more complete mathematical framework for studying the asymptotic properties of these estimators has been developed by (Dauxois et al., 1982; Kleffe, 1973). Suppose we have a sample of n realizations i.i.d $(X_1, ..., X_n)$ of a functional random variable X valued in $\mathcal{L}^2(T, \mathbb{R}) = \mathcal{L}^2(T)$ with mean function $\mu(t) = E(X(t))$ and covariance function $\Gamma(s,t) = E\left[(X(s) - \mu(s))(X(t) - \mu(t))\right]$ $(s, t \in T$ which is an finite compact interval $\in \mathbb{R})$, $\Gamma(s,t)$ may admit a spectral decomposition of the shape:

$$\Gamma(s,t) = \sum_{k \geq 1} \lambda_k \phi_k(s) \phi_k(t), \tag{2.3}$$

where $\lambda_1 > \lambda_2 > .... > 0$, represent eigenvalues and ϕ_k are the orthogonal eigen-functions.

$$\lambda_k = \int_T \phi_k(s) \Gamma(s,t) \phi_k(t) \mathrm{d}s \, \mathrm{d}t \tag{2.4}$$

with $\int_T \phi_k^2(t)\,\mathrm{d}t = 1$ and $\int_T \phi_k\phi_i(t)\,\mathrm{d}t = 0$ $(i < k)$. Then the Karhunen-Loève expansion of X_i is written as:

$$X_i(t) = \mu(t) + \sum_{k\geq 1} \xi_{ik}\phi_k(t) \tag{2.5}$$

where $\xi_{ik} = \int_T (X_i(t) - \mu(t))\phi_k(t)dt$ refers to the functional principal component score with $E(\xi_{ik}) = 0$ and $Var(\xi_{ik}) = \lambda_k$ (ξ_{ik} are uncorrelated accross k and independant across i). From equation (2.5), note that:

$$\sup_t E[X_i(t) - \mu(t) - \sum_{k\geq 1}^{K} \xi_{ik}\phi_k(t)] \underset{K\to+\infty}{\longrightarrow} 0$$

The larger K, the better the approximation of X_i by

$$X_{i,K}(t) \approx \mu(t) + \sum_{k\geq 1}^{K} \xi_{ik}\phi_k(t) \tag{2.6}$$

makes it easier to reduce the dimension (see Ramsay and Silverman (2002a) and Shang (2014) for more details in FPCA).
In matrix form (Ramsay and Silverman, 2005b), taking $\mathbf{X} = (X_1, ..., X_n)^\top$as a vector valued function and the vector valued function $\phi = (\phi_1, ..., \phi_n)^\top$ the equation (2.6) may be rewritten:

$$\mathbf{X} = \xi\phi, \tag{2.7}$$

where ξ is an $n \times K$ coefficient matrix. In the literature, there are various methods for choosing K. we have the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Yao et al., 2005; Li et al., 2013), Hall and Vial (2006) uses bootstrap techniques. Scree plot or the fraction of variation explained (FVE) by first principal component is also used (Chen and Müller, 2012; Dai, 2018).

2.1.3 Principal Component Analysis for multivariate functional data (MFPCA)

Principal component analysis has been extended to the functional multivariate case (Ramsay and Silverman, 2005b; Berrendero et al., 2011). A significant portion of the current research in multivariate functional data analysis is centered on processes that are defined within the confines of a same one-dimensional domain (Di et al., 2009; Schmutz et al., 2020; Carroll et al., 2021; Zhang et al., 2019; Chiou et al., 2014; Jacques and Preda, 2014). Some studies focus on processes observed on different domains (Happ and Greven, 2018; Wong et al., 2019).

Let $\mathbf{X} = (X_1, \ldots, X_P)^\top$ be a vector-valued stochastic process with $P \geq 1$ an integer. For $1 \leq p \leq P$, let I_p be a compact set in $\mathbb{R}$, with finite (Lebesgue-) measure and such that $X_p : I_p \longrightarrow \mathbb{R}$ is assumed to belong to $\mathcal{L}^2(I_p, \mathbb{R}) = \mathcal{L}^2(I_p)$ (space of square integrable functions in I_p. We denote by $\mathbf{I} := I_1 \times \cdots \times I_P$, the P-Fold Cartesian product of I_p. So, $\mathbf{X}$ is a multivariate functional random variable indexed by $\mathbf{t} = (t_1, \cdots, t_P) \in \mathbf{I}$ and taking values in the P-Fold Cartesian product space $\mathcal{H} := \mathcal{L}^2(I_1) \times \cdots \times \mathcal{L}^2(I_P)$.

Let the inner product $\langle\langle \cdot, \cdot \rangle\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$,

$$\langle\langle f, g \rangle\rangle := \sum_{p=1}^{P} \langle f_p, g_p \rangle = \sum_{p=1}^{P} \int_{I_p} f_p(t_p) g_p(t_p) dt_p, \quad f, g \in \mathcal{H}.$$

Then, $\mathcal{H}$ is a Hilbert space with respect to the scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ (Happ and Greven, 2018).

2.1.3.1 Multivariate Karhunen-Loève representation

We assume that $\mathbb{E}[\mathbf{X}(\mathbf{t})] := (\mathbb{E}[X_1(t_1)], \ldots, \mathbb{E}[X_P(t_P)]) = \mathbf{0}$, $\forall\ \mathbf{t} \in \mathbf{I}$. Let C denote the $P \times P$ matrix-valued covariance function which, for $\mathbf{s}, \mathbf{t} \in \mathbf{I}$, is defined as

$$C(\mathbf{s}, \mathbf{t}) := \mathbb{E}[X(\mathbf{s}) X(\mathbf{t})^{\mathsf{T}}],$$

where the (p, q)th element of the matrix $C(\mathbf{s}, \mathbf{t})$, for $1 \leq p, q \leq P$, is the covariance function between the p-th and the qth components X:

$$C_{p,q}(s_p, t_q) := \mathbb{E}[X_p(s_p) X_q(t_q)] = \mathrm{Cov}(X_p(s_p), X_q(t_q)), \quad s_p \in I_p, t_q \in I_q.$$

In particular, $C_{p,q}(\cdot, \cdot)$ belongs to $\mathcal{L}^2(I_p \times I_q)$

Let $\Gamma : \mathcal{H} \to \mathcal{H}$ denotes the covariance operator of $\mathbf{X}$ on the Hilbert space $\mathcal{H}$, where for $f \in \mathcal{H}$ and $\mathbf{t} \in \mathbf{I}$, the qth component of $\Gamma f(\mathbf{t})$ is given by

$$(\Gamma f)^{(q)}(t_q) := \langle\langle C_{\cdot,q}(\cdot, t_q), f(\cdot) \rangle\rangle = \sum_{p=1}^{P} \int_{I_p} C_{p,q}(s_p, t_q) f_p(s_p) ds_p, \quad t_q \in I_q, f \in \mathcal{H}.$$

By the theory of Hilbert-Schmidt operators, there exists a complete orthonormal basis $\{\phi_j, j = 1, 2, \ldots\} \subset \mathcal{H}$ and a sequence of real numbers $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ such that

$$\Gamma \phi_j = \lambda_j \phi_j \quad \text{and} \quad \lambda_j \to 0 \quad \text{as} \quad j \to \infty.$$

The λ_j's are the eigenvalues of the covariance operator Γ and the ϕ_j's are the associated eigenfunctions. The multivariate version of the Karhunen-Loève's representation is:

$$X(\mathbf{t}) = \sum_{j=1}^{\infty} \xi_j \phi_j(\mathbf{t}), \quad \mathbf{t} \in \mathbf{I},$$

with zero mean random variables $\xi_j = \langle\langle X, \phi_j \rangle\rangle$ and $\mathrm{Cov}(\xi_j, \xi_l) = \lambda_l \mathbf{1}_{\{j=l\}}$.

Let $J \geq 1$ and assume that the first J eigenvalues are nonzero, i.e. $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_J \geq \lambda_{J+1}$. Up to a sign, the elements of the MFPCA basis are characterized by:

$$\begin{aligned}
\phi_1 &= \arg\max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle \quad \text{such that} \quad \phi = 1, \\
\phi_2 &= \arg\max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle \quad \text{such that} \quad \phi = 1 \quad \text{and} \quad \langle\langle \phi, \phi_1\rangle\rangle = 0, \\
&\vdots \\
\phi_{J+1} &= \arg\max_{\phi} \langle\langle \Gamma\phi, \phi\rangle\rangle \quad \text{such that} \quad \phi = 1 \text{ and } \langle\langle \phi, \phi_l\rangle\rangle = 0, \ \forall l \leq J.
\end{aligned}$$

Then, the truncated Karhunen-Loève expansion of the process $\mathbf{X}$ is

$$\mathbf{X}_{\lceil J \rceil}(\mathbf{t}) = \sum_{j=1}^{J} \xi_j \phi_j(\mathbf{t}), \quad \mathbf{t} \in \mathbf{I}, \quad J \geq 1; \tag{2.8}$$

and the truncated univariate Karhunen-Loève expansion of the pth component of $\mathbf{X}$ is

$$X_{p,\lceil J_p \rceil}(t_p) = \sum_{j=1}^{J_p} c_{p,j} \psi_{p,j}(t_p), \quad t_p \in I_p, \quad J_p \geq 1, \quad 1 \leq p \leq P, \tag{2.9}$$

where $\{\psi_{p,j}, j = 1, 2, \ldots\}$ is the univariate FPCA basis associated to the covariance operator Γ_p of X_p and the scores are $c_{p,j} = \langle X_p, \psi_{p,j}\rangle$. (Happ and Greven (2018), Prop.5, p.7) derived a direct relationship between the truncated representations (2.9) of the single elements X_p and the truncated representation (2.8) of the multivariate functional data $\mathbf{X}$.

The principal component elements are in general, not known and have to be estimated from a sample that are possibly observed on different sparse grid points. These elements are the eigenvalues $\{\lambda_j\}_{j\geq 1}$, the eigenfunctions $\{\phi_j\}_{j\geq 1}$ and the scores $\{\xi_j\}_{j\geq 1}$.
Given a sample of n spatial observations $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ of $\mathbf{X}$ at $s_1, ..., s_n$ stations, the estimation procedure for MFPCA consists:

- For each element X_p, estimate a univariate FPCA based on the observations $X_p^{(1)}, \ldots, X_p^{(n)}$ by an estimation of the variance-covariance function of X_p:

$$\hat{K}_p(s,t) = \frac{1}{n-1} \sum_{i=1}^{n} X_p^{(i)}(s) X_p^{(i)}(t)$$

  This results in univariate estimated eigenfunctions $\hat{\psi}_{p,j}$ and scores $\hat{c}_{p,j}^{(i)}$, $i = 1, \ldots, n, \ j = 1, \ldots, J_p$ for a given truncation integer J_p.

- Define the matrix $\boldsymbol{\Xi} \in \mathbb{R}^{n \times J_+}$ with $J_+ = \sum_{p=1}^{P} J_p$, where each row $(\hat{c}_{1,1}^{(i)}, \ldots, \hat{c}_{1,J_1}^{(i)}, \ldots, \hat{c}_{P,1}^{(i)}, \ldots, \hat{c}_{P,J_P}^{(i)})$ contains the estimated scores for the P components of the i-th observation. Let's consider the matrix $\mathbf{Z} \in \mathbb{R}^{J_+ \times J_+}$ consisting of blocks $\mathbf{Z}^{(pq)} \in \mathbb{R}^{J_p \times J_q}$ with

entries

$$Z_{jk}^{(pq)} = \text{Cov}(c_{p,j}, c_{q,k}), \quad j = 1, \ldots, J_p, \quad k = 1, \ldots, J_q, \quad p, q = 1, \ldots, P.$$

An estimate $\hat{\mathbf{Z}} \in \mathbb{R}^{J_+ \times J_+}$ of the matrix $\mathbf{Z}$ is given by

$$\hat{\mathbf{Z}} = \frac{1}{n-1} \mathbf{\Xi}^T \mathbf{\Xi}$$

- Perform a matrix eigen-analysis for $\hat{\mathbf{Z}}$ resulting in eigenvalues $\hat{\lambda}_j$ and orthonormal eigenvectors $\hat{\mathbf{v}}_\mathbf{j}$.
- Elements of the estimated multivariate eigenfunctions are given by

$$\hat{\phi}_{p,j}(t_p) = \sum_{k=1}^{J_p} [\hat{\mathbf{v}}_\mathbf{j}]_{p,k} \hat{\psi}_{p,k}(t_p), \quad t_p \in I_p, \quad j = 1, \ldots, J_+$$

and the corresponding multivariate scores can be calculated via

$$\hat{\xi}_j^{(n)} = \sum_{p=1}^{P} \sum_{k=1}^{J_p} [\hat{\mathbf{v}}_\mathbf{j}]_{p,k} \hat{c}_{p,k}^{(n)} = \mathbf{\Xi}_{n,\cdot} \hat{\mathbf{v}}_\mathbf{j}$$

The estimated eigen-values and functions are derived under the assumption of a finite sample size n and a finite Karhunen-Loève representation for each univariate function X_p. They are relevant in practice with an appropriate choice of the truncation orders.

2.1.4 Regression for functional data

Functional modelling is very often used to describe the relationship between a response variable (Y) and predictor variables (X) when at least one of the variables is functional (Ferraty and Vieu, 2006b; Horváth and Kokoszka, 2012; Cuevas, 2014).

- Y scalar and functional X: the functional linear model (FLM) is the most studied form of regression, it is introduced by Ramsay and Dalzell (1991) then described by Hastie and Mallows (1993) :

$$Y_i = \beta_0 + \int X_i(t)\beta(t)\mathrm{d}t + \epsilon_i, \; i = 1, ..., n \tag{2.10}$$

were here for simplicity we consider, $\beta \in L^2(T)$, $X_i \in L^2(T)$ are $i.i.d$ with $\mathrm{E}\|X_i\|^2 < \infty$, $E(\epsilon_i) = 0$ and X_i independent of ϵ_i.
Generally, the estimation of β is done using functional principal component analysis, Hilbert space approaches, penalised B-splines. Cardot et al. (1999) proposed an estimator of FLM using functional principal component analysis and gives some statistical inferences. For responses with a non-gaussian distribution, we have the Generalized FLM (GFLM) introduced by Marx and Eilers (1999):

$$g\{E(Y_i)\} = \beta_0 + \int X_i(t)\beta(t)\mathrm{d}t + \epsilon_i \tag{2.11}$$

Müller and Stadtmüller (2005) proposed an GFLM based on the reduction of the dimensions of functional predictors with a truncated Karhunen-Loève development.
McLean et al. (2014) extend the generalised additive model of Hastie and Tibshirani (1986) in the functional framework:

$$g\{E(Y_i)\} = \beta_0 + \int f\{X_i(t), t\}dt + \epsilon_i(t) \tag{2.12}$$

where β_0 is the intercept, g is a known link function (exponential family) and $f(x,t)$ is an unspecified smooth function to be estimated.

The functional linear model has been extended to non-linear or non-parametric models in order to take into account the case where the relationship between the response and predictor variable is not linear (Ferraty and Vieu, 2009). In the framework where the relationship between the scalar response and the functional predictor is non-linear Yao and Müller (2010) define a quadratic regression functional of the form:

$$Y_i = \beta_0 + \int X_i(t)\beta_1(t)\mathrm{d}t + \int\int \beta_2(s,t)X_i(s)X_i(t)\mathrm{d}s\,\mathrm{d}t + \epsilon_i,\ i = 1, ..., n, \tag{2.13}$$

where β_0 is an intercept. The linear part is the same as equation (2.10) and we have the quadratic part added. James and Silverman (2005) extend generalized linear models (GLM), generalized additive models (GAM) and projection pursuit regression (PPR) using the non-parametric version of FLM called functional adaptive model estimation (FAME):

$$g\{E(Y_i)\} = \beta_0 + \sum_{k=1}^{K} f_k\left\{\int X_i(t)\beta_k(t)\right\}\mathrm{d}t \tag{2.14}$$

The parameters $\beta_k(.)$, $X_i(.)$ and $f_k(.)$ were represented using cubic splines and the model has been extended to deal with multiple functional predictors. In the semi-parametric framework, we have the functional GAM between a real response Y_i and multivariate functional variable $\mathbf{X}_i$ expressed as follows:

$$\mathrm{E}(Y_i \mid \mathbf{X}_i) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} f_j(X_i^j)\right) \tag{2.15}$$

The estimation of f_j is based on the spectral decomposition of the covariance operator of the $\mathbf{X}_i$ and from the scores, namely $f_j(X_i^j) := \sum_{k=1}^{K} f_j^k(\xi_{j,i}^k)$, where $\xi_{j,i}^k$ are the PCA scores of X_i^j. This method is called Generalized Spectral Additive Model (GSAM) (Müller and Yao, 2008) due to the use of the spectral decomposition of the covariance operator of the process X.

- Y and X functionals and function-on-function regression (Ramsay and Silverman, 1997) : the relationship between the response variable and functional predictors is expressed by integrating the functional predictor weighted by an unknown bivariate coefficient function:

$$Y_i(t) = \beta_0(t) + \int_0^1 \beta(s,t)X_i(s)\mathrm{d}s + \epsilon_i(t),\ i = 1, ..., n, \tag{2.16}$$

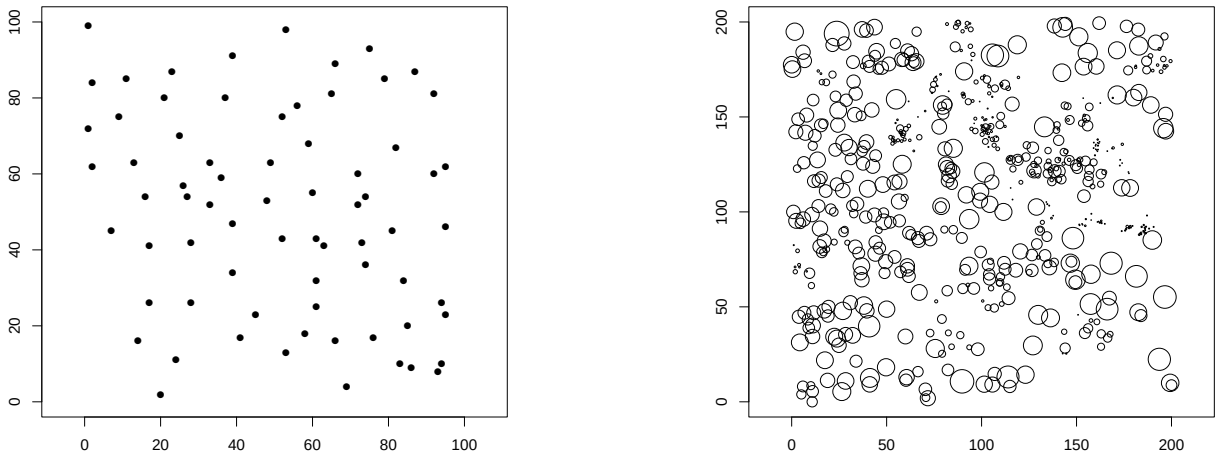
where $\beta(s,t)$ defines a Hilbert-Schmidt integral operator.

Extensions of these models to various directions have been proposed in the literature (Chowdhury and Chaudhuri, 2019; Zhou and Lin, 2016) and the recent review on FDA of Koner and Staicu (2023).

2.2 Spatial data analysis

Spatial statistics is concerned with the study of phenomena observed in a spatial domain $S \in \mathbb{R}^d, d > 1$. It is applied in many fields, notably in the environmental sciences, in econometrics (LeSage and Pace, 2008), in ecology, in epidemiology (Guttmann, 2014), in real estate studies (Srikhum, 2012), etc. Contrary to time series whose models are based on the notion of past and future, spatial statistics uses non-causal methods because there is no order in $\mathbb{R}^d$. Knowledge of the distribution of observations in space is essential in the study of spatial data. This distribution is designated by location information that allows the distance between observations to be determined, which is necessary to analyse spatial dependencies. Technological advances have allowed geo-located data to grow, then becoming more accessible and available in several formats (Moraga and Baker, 2022), leading the scientific community to develop new spatial analysis approaches to take into account the structure of spatial dependency. These methods include spatial regression, which takes account of spatial dependence (Anselin, 2009; Ward and Gleditsch, 2018; Chi and Zhu, 2008; Cherry, 1997; Ternynck, 2014). There are also spatial interpolation methods such as kriging, which makes it possible to predict the value of a spatial variable at unsampled sites on the basis of data observed at neighbor sites (Baillargeon, 2005; Matheron, 1969; Mateu and Giraldo, 2021; Giraldo et al., 2011).
Spatial data are modelled as observations of a family of random variables $X = \{X_s, s \in S\}$ indexed by S. The variable X_s, qualified as regionalized, describes a phenomenon that unfolds in space more or less regularly with a certain structure. Generally, the location of a point $s \in S$ is geographical ($S \in \mathbb{R}^2$ or $S \in \mathbb{R}^3$). Three main types of spatial data according to the nature of S are distinguished (Gaetan and Guyon, 2010; Cressie, 1993):

- **Point patterns**: the domain S is a point process belonging to $\mathbb{R}^d$, $d \geq 1$. In this case, it is the distribution of the set of points (or sites) s that is random. The objective of these processes is to know if this distribution is regular, random or presents aggregates. They can be generalised to marked point processes, if one associates one or more characteristics (called marks) to the points. These points can represent trees, disease status, robberies, crimes, birds' nests, etc.



A) B)

Figure 2.1: Examples of spatial point data: A) Locations of pine saplings in a Swedish forest represented as points; B) Locations and diameters of 584 Longleaf pine trees (the points have extra information called marks attached to them). Source: Swedish-pines and longleaf data set available on the spatstat.data package of R

- **Lattice data** : the process is only observed on a countable collection of zones (postal codes, regions, departments, countries,...) included in $S \in \mathbb{R}^2$. The location sites $s \in \mathbb{Z}^d$, generally represent geographical units arranged in a regular, discrete manner. To analyse this kind of data, it is usual to define the notion of neighbourhood and the influence exerted by an observation i on its neighbour j.

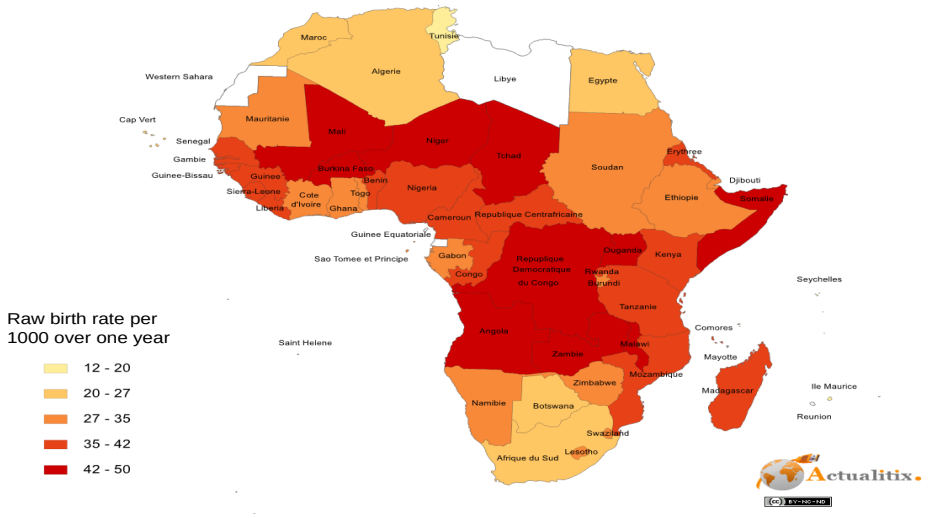


Figure 2.2: Example of lattice data; birth rate in Africa. Source: Actualitix

- **Geostatistical data**: the process is observed in n continuous points $(s_1, ..., s_n) \in S$ (continuous subspace of $\mathbb{R}^d$), deterministic (usually, latitude and longitude). The distribution of this points

can be either regular or not. These data can be found in the fields of the environmental (measurements of air pollution in various stations (Cameletti et al., 2013), heavy metal concentrations etc.) or in oceanography (Assunção et al., 2020).

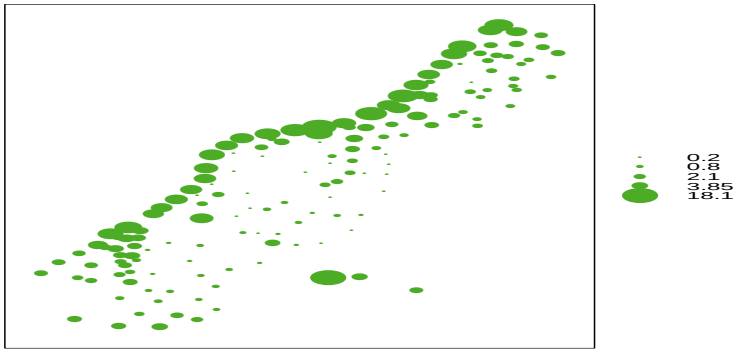


Figure 2.3: Example of geostatistical data. Topsoil lead concentrations at locations sampled in a flood plain of the river Meuse. Source: Meuse river data set available on the sp package of R

In the case of geostatistical processes, the locations in the area of interest can be chosen arbitrarily to obtain information, unlike point processes where these locations are given a priori.
In this thesis, we are interested in spatial geostatistical or lattice data.
Unlike inference in classical statistics, which assumes that observations are independent and identically distributed, the analysis of spatial data requires to take account a potential spatial dependency between observations and to assume that the data are spatially auto-correlated. Exploring spatial structure involves defining either spatial covariance (when the process is supposed weak stationary) or spatial weighting matrix noted $\mathbf{W}$ (in case of lattice data) that describes the auto-correlation or dependency between different locations.
Statistical methods and techniques for independent or time-series data have been extended to spatial data. A key spatial method of interest in our work describe in the following is spatial principal component analysis for areal of geo-statistical data.

2.3 Spatial Principal Component Analysis for areal real-valued data (sPCA)

Spatial PCA for areal data is mainly based on a weight matrix. There are various ways of constructing this matrix. (Griffith, 2020; Getis and Aldstadt, 2004).

2.3.1 Specification of the spatial weight matrix

The spatial weight matrix related to n locations $s_1, ..., s_n$ of observations of a spatial process is an $n * n$ positive, non stochastic, non-negative and symmetric matrix with the values w_{ij} at location s_i, s_j. $w_{i,j}$

represents the weights for each pair of locations. They are defined by rules which define the spatial relations among locations and, therefore, determine the spatial autocorrelation, by convention, $w_{ii} = 0$ (diagonal elements) (Zhou and Lin, 2008) i.e a location does not impact on itself directly:

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

Two main approaches are used to construct such weight matrix: weight based on boundary or contiguity and weight based on distance.

- **Contiguity spatial weight matrix**
  In this type of spatial weight, the allocation of weights depends on whether or not there are common boundaries or vertices between pairs of locations s_i and s_j. There are three types of spatial contiguity including common border (rook), common vertices adjacency (bishop) and either common border or vertices adjacency (queen). A typical specification of contiguity weight matrix is:
  $$w_{ij} = \begin{cases} 1, & \text{if } s_i \text{ and } s_j \text{ are contiguous} \\ 0, & \text{if } s_i \text{ and } s_j \text{ not contiguous} \end{cases}$$

- **Geographical distance spatial weight matrix**
  These weight matrices define spatial connections based on the geographical proximities between locations and the intensity of their interaction depends on the distance between their centroids (Le Gallo, 2002). There are various ways of specifying this matrix, but they are all based on the same intuition.

  - *k-Nearest Neighbour weights*
    $w_{ij} = \begin{cases} 1 & \text{if } s_j \in N_k(s_i), \\ 0 & \text{Otherwise,} \end{cases}$
    where $N_k(s_i)$ is the set of the k closest units or locations to s_i for $k \in \{1, ..., n-1\}$
  - *Inverse Distance*
    $w_{ij} = d_{ij}^{-\alpha}$; for $s_i \neq s_j$
    where $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ represent the euclidean distance calculated by using the geographic coordinates x and y of locations s_i and s_j; α represent the distance friction coefficient and generally $\alpha = 1$ or $\alpha = 2$.
  - *Radial distance*
    $w_{ij} = \begin{cases} 1 & \texttt{if } 0 \leq d_{ij} \leq \epsilon, \\ 0 & \texttt{if } d_{ij} > \epsilon, \end{cases}$
    where d_{ij} is the euclidean distance between units s_i and s_j, and ϵ is a critical distance (*threshold distance* or *bandwidth*) beyond which there is no direct spatial interaction between locations, and it should be able to guarantee that each location has at least one neighbour. This threshold is determined by a statistic derived from the distribution of

distances in the sample.

- *Exponential Distance Decay weights*
$$w_{ij} = \begin{cases} \exp(-\alpha d_{ij}) & \texttt{if } 0 \leq d_{ij} \leq \epsilon, \\ 0 & \texttt{if } d_{ij} > \epsilon, \end{cases}$$
- *Double-Power Distance weights*
$$w_{ij} = \begin{cases} [1 - (d_{ij}/\epsilon)]^k & \texttt{if } 0 \leq d_{ij} \leq \epsilon, \\ 0 & \texttt{if } d_{ij} > \epsilon \end{cases}$$

2.3.2 The Moran's index

One of the main methods of spatial data analysis is to determine whether there is spatial dependence between the values of a variable in different locations of the study area. Spatial autocorrelation measures have been developed, the most widely used in the literature is the Moran index. First introduced by Moran (1948, 1950) and developed by Cliff and Ord (1973, 1981), it generalises Pearson's correlation coefficient. The values of this index range from -1, indicating negative autocorrelation, to 1 indicating positive autocorrelation (neighbouring locations have similar values of the variable). When it is zero, there is no spatial autocorrelation. Consider n observations X_i at locations s_i, $i = 1, ..., n$ of the spatial process X. The Moran index is defined as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{2.17}$$

where w_{ij} derived from a spatial weight matrix W, describes the neighboring relation between two locations (s_i and s_j), $\bar{X}$ is the empirical mean of the X_i. In matrix notation, we have:

$$I(\mathbf{X}) = I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\mathbf{X}^T W \mathbf{X}}{\mathbf{X}^T \mathbf{X}}, \tag{2.18}$$

with $\mathbf{X} = (X_1, ..., X_n)^\top$.

2.3.3 Spatial principal component analysis using Moran index

As said before, principal Component Analysis (PCA) is one of the most widely used dimension reduction methods in the literature (Duby and Robin, 2006; Bro and Smilde, 2014; Kherif and Latypova, 2020). When calculating the scores of the principal components, classical PCA does not take account of spatial information. A spatial principal component analysis based on a Moran index was introduced by Jombart et al. (2008) and considered on this contribution. Consider the vector $\mathbf{X}$ of n observations X_i of a spatial process valued in $\mathbb{R}^p$. Let

$$V(\mathbf{X}) \ = \frac{1}{n}(\mathbf{X}^\top \mathbf{X}) I(\mathbf{X}) = \ \frac{1}{n}\, \mathbf{X}^\top W \mathbf{X}$$

It is highly positive when $\mathbf{X}$ has a large variance and shows a global spatial structure and is negative in a situation with high variance and gives a local structure.

The purpose of this spatial principal component (SPCA) proposed in Jombart et al. (2008) is based on scaled $\mathbb{R}^p$ vectors $\mathbf{u}$ (loadings) ($\|\mathbf{u}\|$=1) such that the n raw vectors $\chi = \mathbf{Xu}$ are scattered and spatially autocorrelated. In other words, the aim is to find the extreme values (Jombart et al., 2008) of

$$C(\mathbf{u}) = V(\mathbf{Xu}) = \frac{1}{n}\mathbf{u}^\top \mathbf{X}^\top W \mathbf{Xu} \tag{2.19}$$

The solutions (Jombart et al., 2008) are the eigenvectors $\mathbf{u}_k$ of $\frac{1}{2n}\mathbf{X}^\top(W + W^\top)\mathbf{X}$ associated with the largest and smallest eigenvalues $\alpha_k = var(\chi_k)\tilde{I}(\chi_k)$ (where $\chi_k = \mathbf{Xu}_k$, $var(\chi_k)$ the variance of χ_k). Note that some eigenvalues α_k may be negative since $\tilde{I}$ is not always positive.

Based on a positive covariance matrix, the classical PCA provides positive eigenvalues whereas the SPCA may give positive and/or negative eigenvalues, thus defining global and local spatial structures. Dray et al. (2008) proposes an equivalent approach, which is a multivariate spatial analysis (generalization of Multivariate Spatial Correlation Analysis of Wartenberg (1985)) based on the Moran index and finds a compromise between the relations among many variables (multivariate analysis) and their spatial structure (autocorrelation). Considering $\mathbf{D}$ as a scalar product of $\mathbb{R}^n$, and $\mathbf{Q}$ a scalar product of $\mathbb{R}^p$, Dray et al. (2008) intoduces the spatial weight matrix in the analysis of the triplet $(\mathbf{X}, \mathbf{Q}, \mathbf{D})$ associated to the PCA of $\mathbf{X}$. This approach seek to maximize the scalar product between a linear combination of original variables ($\mathbf{a}_1 = \mathbf{XQu}_1$) and a linear combination of lagged variables ($\tilde{\mathbf{a}}_1 = \mathbf{WXQu}_1$).
This spatial PCA for multivariate data has been extended to functional data by Ali Hassan (2021) to examine spatial auto-correlation of mortality rates for 28 European countries.
In our contribution, we have applied the functional spatial PCA (Ali Hassan, 2021).

2.4 Spatial Functional Principal Component Analysis (SF-PCA) for geostatistical data

Recent studies have extended the FPCA methods to take into account geo-spatially indexed functional data (Winzenborg, 2011; Li and Guan, 2014; Kuenzer et al., 2021; Liu et al., 2017a). Winzenborg (2011) extends the non-parametric version of FPCA to spatial data and applies it to real data from the medical field. Li and Guan (2014) and Liu et al. (2017a) proposed an FPCA for processes that are both spatially and temporally dependent. This thesis focuses on the work of Kuenzer et al. (2021), where a dimension reduction technique applicable to a univariate functional data, associated with spatial locations on a grid is introduced.

We consider that at n spatial units located on a region $\mathbf{D} \subset \mathbb{Z}^N$, $N > 1$, representing a rectangular grid, we observe a univariate spatial functional process $\{X_\mathbf{s}(.)\}$, where $\mathbf{s} \in \mathbf{D}$, $X_\mathbf{s} = \{X_\mathbf{s}(t), t \in \mathcal{T}\}$. Let $\mathcal{T}$ be a compact set in $\mathbb{R}$, with finite (Lebesgue-) measure and such that $X_\mathbf{s} : \mathcal{T} \longrightarrow \mathbb{C}$ is assumed to belong to $\mathcal{L}^2(\mathcal{T}, \mathbb{C})$, the space of complex square-integrable functions on $\mathcal{T}$. Let $\mathcal{L}^2(\mathcal{T}, \mathbb{C}) = \mathcal{L}^2(\mathcal{T})$.
Kuenzer et al. (2020) proposed a spatial principal component analysis, we briefly sum-up.
With $X_\mathbf{s}$ being a functional random variable taking values in $\mathcal{H} := \mathcal{L}^2(\mathcal{T})$, let the inner product

$\langle\langle\cdot,\cdot\rangle\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{R}$, for $f,g\in\mathcal{H}$:

$$\langle f,g\rangle = \int_{\mathcal{T}} f(t)g(t)dt$$

Then, $\mathcal{H}$ is a Hilbert space with respect to the scalar product $\langle\cdot,\cdot\rangle$

Let the observations of the spatial process on n sites $\mathbf{s}_1,...,\mathbf{s}_n\in\mathbf{D}$ denoted by $\mathbf{X}=\left(X_{\mathbf{s}_1},...,X_{\mathbf{s}_n}\right)$.

Let's suppose that each $\{X_{\mathbf{s}}\}$ is a weakly stationary functional process. Let the covariance operator $C := \mathbb{E}[(X_{\mathbf{s}}-\mu)\otimes(X_{\mathbf{s}}-\mu)]$ (where μ is the mean curve define by $\mu(t)=\mathbb{E}X_{\mathbf{s}}(t)$ with $t\in\mathcal{T}$) with kernel $c(t,s)=cov(X_{\mathbf{u}}(t),X_{\mathbf{u}}(s))$ $(t,s\in\mathcal{T})$. We have:
(i) $\mathbb{E}(X_{\mathbf{s}}(t))=\mathbb{E}(X_{\mathbf{0}}(t))=\mu(t)$, $t\in\mathcal{T}$ with $\mathbf{0}$ the null vector in $\mathbb{R}^N$
(ii) for all $\mathbf{s},\mathbf{h}\in\mathbf{D}$, and $t,s\in\mathcal{T}$; $c_{\mathbf{h}}(t,s) := Cov\left(X_{\mathbf{h}}(t),X_{\mathbf{0}}(s)\right)=Cov\left(X_{\mathbf{s}+\mathbf{h}}(t),X_{\mathbf{s}}(s)\right)$
The integral operator defined by the autocovariance kernel $c_{\mathbf{h}}$ is denoted $C_{\mathbf{h}}$ and defined by

$$(C_{\mathbf{h}}f)(t)=\int_{\mathcal{T}} c_{\mathbf{h}}(s,t)f(s)ds,\quad f\in\mathcal{L}^2(\mathcal{T}),\ t\in\mathcal{T}$$

Let us denote the spectral density operator of $X_{\mathbf{s}}$ by $\mathcal{F}_\theta$ with the following kernel:

$$f_\theta(t,s) := \frac{1}{(2\pi)^N}\sum_{\mathbf{h}\in\mathbb{Z}^N} c_{\mathbf{h}}(t,s)\exp(-i\mathbf{h}^\top\theta), \tag{2.20}$$

$$t,s\in\mathcal{T},\quad \theta\in[-\pi,\pi]^N,\quad i=\sqrt{-1},$$

where θ is the spatial frequency. We define $\mathcal{L}^2_U([-\pi,\pi]^N)$ as the space of measurable mappings $x:[-\pi,\pi]^N\to U$ satisfying $\int_{[-\pi,\pi]^N}\|x(\theta)\|^2d\theta<\infty$, with U the Hilbert space of all Hilbert–Schmidt operators from $\mathcal{L}^2(\mathcal{T})$ to $\mathcal{L}^2(\mathcal{T})$ (see Kuenzer et al. (2020) for further explanation). The operator $\mathcal{F}_\theta$ is understood as element of the space $\mathcal{L}^2_U([-\pi,\pi]^N)$ and is defined by

$$\left(\mathcal{F}_\theta G_\theta\right)(t)=\int_{\mathcal{T}_j} f_\theta(s,t)G_\theta(s)ds,$$

$$\text{with}\quad G_\theta\in\mathcal{L}^2_U([-\pi,\pi]^N)\quad\text{and}\quad t\in\mathcal{T}_j$$

Considering some assumptions (Kuenzer et al., 2020), $\mathcal{F}_\theta$ is a Hilbert-Schmidt operator (positive, self-adjoint) and admits a spectral decomposition that is used by these authors to build a spatial PCA. This spatial analysis has been extended in the last contribution to the multivariate case.

Chapter 3

Study of the impact of environmental variables on scattering layers using functional additive regression models

3.1 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 characterized 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.

3.2 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 (Haney, 1988; Bianchi et al., 2013; 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, SSLs 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. (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., 2020) 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, 2005a; 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 through 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; Reyes et al., 2015; Pauthenet et al., 2017; Pauthenet, 2018; Pauthenet et al., 2019). 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 con-

centration in the coastal zone of the Guadalquivir estuary. Tarrío-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 of 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) characterized 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.

3.3 Materials and method

3.3.1 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 (Figure 3.2). 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^2nmi^{-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 (Figure A2). 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 variables** , 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 (Figure A2). 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 (GPS). The environmental variables were matched with the acoustic variables using the matlab code available in Annexe 1. The data, which has a size of 2.61 GB, was stored in MATLAB format.

3.3.2 Method

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.
Namely, we analysed and modelled (see Figure 3.1) 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 $s_1, ..., s_n$) and the characteristics of the pelagic environment *i.e* the physical parameters (considered as observations of a functional spatial processes (see Figure A3)$\{X_s(t),\ t \in [d_{s,min}, d_{s,max}]\}_{s\in\mathbb{R}^2}$, 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 $X = (X^1, ..., X^p)$ valued in a multivariate functional space. Namely $X^j \in L^2(\mathcal{T}_j)$, $j = 1, ..., p$ 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 $x_{s,j,t} = \mu_j(t) + X_s^j(t) + \epsilon_{s,j,t}$, with $\mu_j(.)$ the mean function of the parameter j. The functional variable X_s^j is centred and square integrable, while $\epsilon_{s,j,t}$ are centred (real-valued) Gaussian i.i.d variables of finite variance.

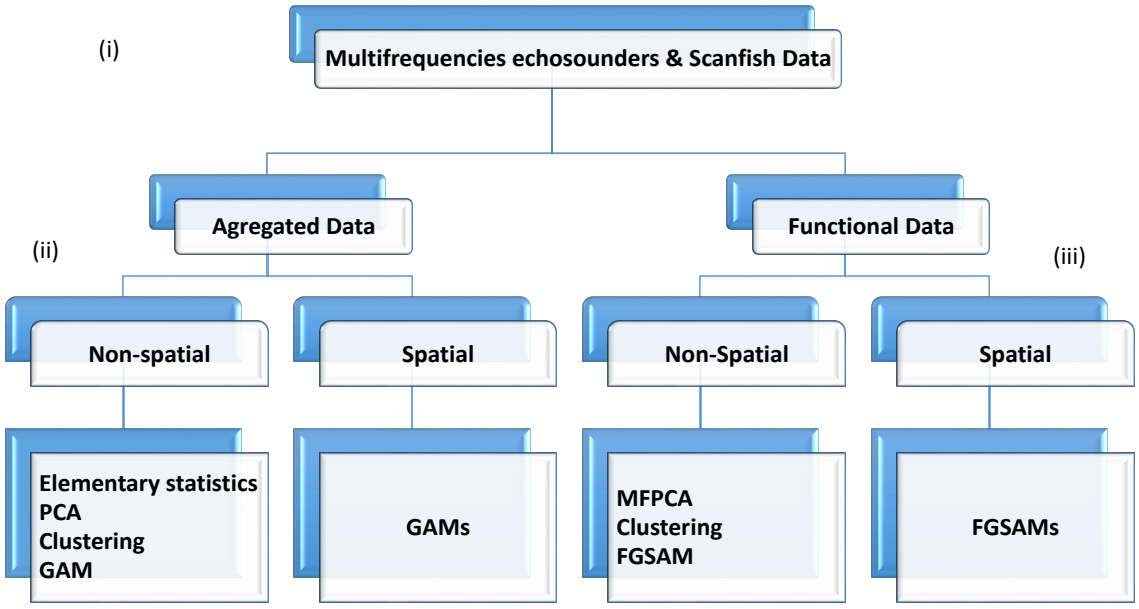


Figure 3.1: Flow chart depicting the study methodology. The research unfolded in three key steps: (i) Raw data extraction from the echosounders and the ScanFish, followed by the computation of Sound Scattered Layers 'SSL' characteristics and the integration of all data; (ii) Comparative analysis of SSL 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 Additive 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.

3.3.2.1 Aggregated data modelling

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

$$X_s^j = \frac{1}{D_s} \sum_{p=1}^{D_s} X_s^j(t_p), \tag{3.1}$$

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

3.3.2.1.1 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

$$g(Y_s) = \sum_{j=1}^{p} f_j(X_s^j) + \epsilon_s, \tag{3.2}$$

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 minimise the root mean square error (RMSE) and mean absolute error (MAE).

3.3.2.1.2 Spatial analysis

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

$$g(Y_s) = \sum_{j=1}^{p} f_j(X_s^j) + g(s) + \epsilon_s, \tag{3.3}$$

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.).

3.3.2.2 Functional data modeling

3.3.2.2.1 Non-Spatial analysis

Instead of aggregating the data, we use in this section models that are able to analyse 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

$$X_s^j(t) = X_s(t) = \sum_{m=1}^{P} c_{s,j,m} b_m(t) \tag{3.4}$$

   where $b_1(t), ..., b_P(t)$ 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:

$$\bar{X}_{j,n}(t) = \frac{1}{n} \sum_{s \in (s_1,...,s_n)} X_s^j(t) \tag{3.5}$$

   and the empirical covariance function:

$$\hat{c}_{j,n}(t,u) = \frac{1}{n-1} \sum_{s \in (s_1,...,s_n)} (X_s^j(t) - \bar{X}_{j,n}(t))(X_s^j(u) - \bar{X}_{j,n}(u)) \tag{3.6}$$

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

$$X_s^j(t) = \mu_j(t) + \sum_{k=1}^{K} \beta_{j,k,s} \phi_{j,k}(t) \tag{3.7}$$

   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 $\lambda_{1,j} > \lambda_{2,j}, ... > \lambda_{K,j}$ are the eigenvalues, in theory $K = \infty$. The functional principal component scores $\beta_{j,k,s} = \int X_s^j(t)\phi_{j,k}(t)\,\mathrm{d}t$ are assumed to be centered random variables. In practice, $\hat{c}_{j,n}(.,.)$ is decomposed by:

$$\hat{c}_{j,n}(t,u) = \sum_{k=1}^{K} \hat{\lambda}_{k,j} \hat{\phi}_{j,k}(t) \hat{\phi}_{j,k}(u). \tag{3.8}$$

   To determine eigenvalues and eigenfunctions, a number of works in the literature are proposed particularly in the context of univariate independent (Ramsay and Silverman, 2005a) or multivariate independent (Happ and Greven, 2018) or univariate spatially as well as temporally dependent (Hörmann et al., 2015; Winzenborg, 2011; Liu et al., 2017a; Kuenzer et al., 2021).
   We first use the estimated functional principal components (FPC) $\hat{\phi}_{j,k}$ (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:

$$X_s^j(t) \approx \bar{X}_{j,n}(t) + \sum_{k=1}^{K} \hat{\beta}_{j,k,s} \hat{\phi}_{j,k}(t) \tag{3.9}$$

When $p > 1$, univariate basis expansions for each element of the multivariate functional data $X_s = (X_s^1, ..., X_s^p)$ are performed for each of the p functional components and the MFPCA scores are derived (Happ and Greven, 2018; Happ-Kurz, 2020).

2. **Clustering**
When the functional PCA is done, we conduct hierarchical clustering on the MPCA scores $\hat{\beta}_{k,s}$.

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:

$$Y_s = \beta_0 + \sum_{j=1}^{p} \int_{\mathcal{T}_j} \beta_j(t)\, X_s^j(t)\, \mathrm{d}t + \epsilon \tag{3.10}$$

has been extensively studied (see Cardot et al. (1999, 2003b), Crambes and Mas (2013), Yao et al. (2005)). Let $X_s = (X_s^1, ..., X_s^p)$, we have the functional GAM expressed as follows:

$$\mathrm{E}(Y_s \mid X_s) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} f_j(X_s^j)\right) \tag{3.11}$$

The estimation of f_j is based on the spectral decomposition of the covariance operator of X_s and from the scores of PCA of X_s, namely $f_j(X_s^j) := \sum_{k=1}^{K} f_j^k(\xi_{j,s}^k)$, 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).

3.3.2.2.2 Spatial analysis

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

$$\mathrm{E}(Y_s \mid X_s) = g^{-1}\left(\beta_0 + \sum_{j=1}^{p} f_j(X_s^j) + g(s)\right), \tag{3.12}$$

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

3.4 Results

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

3.4.1 Classical methods

3.4.1.1 Elementary statistics of SSL descriptors per frequencies

The analysis of SSLs decriptors reveals interesting patterns in the thickness and depth of SSLs across the different frequencies, with distinct variations according to the diel period.
Over the southern Senegalese continental shelf (Figure 3.4), the SSLs were thicker during the night while during the day, the SSL are almost thin. At all frequencies, SSLs 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 SSLs were thicker at night than at sunset, in contrast to those detected at 38 kHz. The SSLs were deeper at sunset than at night for all frequencies. In northern Senegal, the SSLs were thicker during daytime and sunrise at 70, 120, and 200 kHz while at 38 kHz the SSLs were thicker during the sunrise. In addition, the SSLs were deeper during sunrise and day than during the night for all frequencies.

3.4.1.2 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 (Figure 3.7), 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 ESUs 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 ESUs 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 ESUs 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 sunrise and daytime period that were deeper and more turbid compared to Class 2, which were warmer, more saline, and with higher fluorescence.

3.4.1.3 Spatial Generalised 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 (Tables 3.1, 3.2; see Tables A1, A2 for more details). 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.

3.4.2 Functional methods

3.4.2.1 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 (Figure 3.10). It appears that the first component of each variable generally groups together ESUs with either high or low values at a particular depth range, while the second component groups together ESUs with high or low values at a different depth range (Table A3). This suggests that there are multiple factors influencing the distribution of these ESUs in the water column.

3.4.2.2 Clustering

Clustering was performed on the scores derived from the functional principal components (PC1 and PC2) (Figure 4.2). On the continental shelf of southern, at 38 kHz, class 1 was characterized 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 characterized 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 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.

3.4.2.3 Spatial Functional Generalized Spectral Additive Model (FGSAMs)

The results of the FGSAMs performed on the SSL descriptors provided estimated $\hat{\beta}$ parameters of the functional covariates that varied with depth (Figure 3.19). 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 3.4; see Table A4 for more details).

3.5 Discussion

3.5.1 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; Granato et al., 2018; Demšar et al., 2013). 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 spatial covariates showed overall the best performance (lower RMSE and MAE; Table 3.5). Spatial dependence was modelled 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.

3.5.2 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 characterized 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.

3.5.3 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 (Faye et al., 2015; Balde et al., 2019) 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 characterized 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 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 (Yoon et al., 2007; Diogoul et al., 2020). 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 (Dall'Olmo and Gitelson, 2005; Cui and Lv, 2014). 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 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 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 (Kloser et al., 2009; Blanluet et al., 2019). 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 remotely the marine organism acoustically and then better understand the underlying mechanisms driving the observed effects.

3.6 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.

Table 3.1: Results of Generalized additive model (GAM) between sound scattering layers (SSLs) and oceanographic conditions (sea temperature (Temp), salinity (Sal), turbidity (Turb), fluorescence (Fluo)), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) to predict (1) SSL thickness and (2) SSL Depth and (3) SSL density, spread over three geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf) as observed during the AWA sea survey. AIC (Akaike's Information Criterion); BIC (Bayesian Information Criterion); Log Likelihood (log-likelihood value of a model); Deviance (goodness-of-fit metric for statistical model); Deviance explained (proportion of the total deviance explained by the current model); R^2 (Adjusted R-Squared); GCV score (Generalised Cross-Validation score); Num.obs. (number of observations); Num. smooth terms (Number of smooth terms).

① (A)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**13.48*****	4.32	3.17	6.86	**14.33*****	5.45	4.27	**5.93****
	(1.23)	(3.35)	(6.27)	(8.78)	(2.24)	(3.44)	(3.43)	(2.16)
s(Bottom):Night	**3.15****	**2.51****	**4.86*****	**3.39*****	**3.86*****	**2.63***	**4.41*****	**5.09*****
	(3.27)	(2.95)	(5.34)	(3.85)	(3.98)	(3.09)	(4.91)	(5.46)
s(Bottom):Sunrise	**2.88*****	**5.76*****	**1.00*****	**4.04*****	**3.11*****	**3.90***	**3.28***	**4.14***
	(3.07)	(6.14)	(1.00)	(4.79)	(3.31)	(4.53)	(3.92)	(4.78)
s(Bottom):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Night	**3.84*****	**3.97****	**1.00***	**1.00***	**3.89*****	**4.74*****	**1.00****	**2.14*****
	(3.97)	(4.50)	(1.00)	(1.00)	(3.99)	(5.12)	(1.00)	(2.62)
s(Temp):Sunrise	2.18	**3.29****	**6.55*****	**2.78*****	1.00	**5.10*****	**1.00***	**1.00*****
	(2.63)	(4.07)	(7.40)	(3.44)	(1.00)	(6.07)	(1.00)	(1.00)
s(Temp):Day	**1.00*****	1.00	1.00	1.00	**1.00***	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	**3.52*****	2.00	2.66	2.41	**3.60*****	**2.54****	**3.69****	2.63
	(3.85)	(2.41)	(3.22)	(2.92)	(3.89)	(3.01)	(4.26)	(3.16)
s(Sal):Sunrise	1.12	2.29	**1.00***	1.00	1.00	**2.10****	1.00	1.00
	(1.24)	(2.89)	(1.00)	(1.00)	(1.00)	(2.66)	(1.00)	(1.00)
s(Sal):Day	1.00	1.00	1.00	1.00	**1.00***	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**1.78*****	**3.04*****	**3.20***	**2.29****	**1.80*****	**3.08*****	**4.31*****	**1.00****
	(1.98)	(3.54)	(3.82)	(2.74)	(2.00)	(3.57)	(4.91)	(1.00)
s(Fluo):Sunrise	1.22	**1.00****	**4.40*****	**4.14*****	**1.36***	2.31	2.10	1.00
	(1.40)	(1.00)	(4.95)	(4.65)	(1.58)	(2.76)	(2.54)	(1.00)
s(Fluo):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	1.65	1.82	1.00	**1.84***	**1.00*****	**1.95*****	1.00	**1.00*****
	(1.71)	(1.92)	(1.00)	(1.92)	(1.00)	(1.98)	(1.00)	(1.00)
s(Turb):Sunrise	**1.00****	**2.37****	1.00	**4.60*****	2.39	**3.65****	3.63	2.75
	(1.00)	(2.95)	(1.00)	(5.14)	(2.70)	(4.28)	(4.31)	(3.43)
s(Turb):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Lon,Lat)					**2.66*****	**24.90*****	**24.85*****	**25.48*****
					(2.85)	(27.28)	(27.32)	(27.82)
AIC	999.93	1001.16	909.55	1049.94	890.68	706.41	804.57	853.40
BIC	1105.99	1135.30	1034.71	1180.98	1007.53	949.49	1019.80	1055.34
Log Likelihood	−468.84	−461.21	−418.04	−486.51	−411.05	−281.86	−339.11	−367.43
Deviance	874.91	817.07	554.76	1025.20	521.04	163.56	273.33	352.35
Deviance explained	0.98	0.93	0.87	0.75	0.99	0.99	0.94	0.91
Dispersion	4.49	4.32	2.91	5.41	2.72	1.02	1.64	2.08
R^2	0.97	0.91	0.85	0.70	0.98	0.98	0.92	0.89
GCV score	494.03	492.90	452.39	513.97	441.06	385.69	421.92	444.04
Num. obs.	223	223	223	223	223	223	223	223
Num. smooth terms	15	15	15	15	16	16	16	16

(B)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**64.04*****	**49.68*****	**34.41*****	**31.70*****	**65.53*****	**48.90*****	**29.44*****	**28.75*****
	(0.55)	(0.88)	(0.61)	(0.52)	(1.00)	(0.96)	(0.98)	(0.71)
s(Bottom):Night	**8.57*****	**8.36*****	**8.76*****	**8.66*****	5.57	**4.98*****	1.01	1.00
	(8.93)	(8.85)	(8.96)	(8.94)	(6.55)	(6.00)	(1.01)	(1.00)
s(Bottom):Sunset	1.00	**1.85*****	**1.99*****	**1.94*****	**1.00*****	**1.95*****	**1.92****	**1.00*****
	(1.00)	(1.96)	(2.01)	(2.00)	(1.00)	(1.99)	(1.99)	(1.00)
s(Temp):Night	**1.00*****	**4.53*****	**1.96*****	**4.59***	**3.73*****	**5.32*****	**3.00*****	**6.95*****
	(1.01)	(5.46)	(2.56)	(5.72)	(4.63)	(6.25)	(3.93)	(8.00)
s(Temp):Sunset	**3.70*****	**4.00*****	**1.00****	**3.88*****	**2.69*****	**5.11*****	3.08	3.69
	(4.53)	(4.94)	(1.00)	(4.75)	(3.34)	(6.16)	(3.88)	(4.55)
s(Sal):Night	**3.63*****	**3.08*****	**2.04****	**6.77*****	**1.00*****	**1.00****	**1.00*****	4.78
	(4.51)	(3.88)	(2.64)	(7.63)	(1.00)	(1.00)	(1.00)	(5.86)
s(Sal):Sunset	1.49	2.22	3.87	1.00	3.18	**3.76***	1.10	1.00
	(1.82)	(2.84)	(4.71)	(1.00)	(3.93)	(4.65)	(1.18)	(1.00)
s(Fluo):Night	**6.40*****	**5.68*****	**7.51*****	**6.54*****	**3.63****	1.01	**6.16*****	**4.24****
	(7.37)	(6.57)	(8.48)	(7.68)	(4.59)	(1.02)	(7.47)	(5.36)
s(Fluo):Sunset	1.00	**5.16*****	**2.97*****	**5.08*****	**4.24***	4.51	**4.46*****	**4.16*****
	(1.00)	(6.15)	(3.71)	(6.05)	(5.09)	(5.39)	(5.44)	(5.04)
s(Turb):Night	**4.26*****	3.75	**7.15*****	3.04	**4.03****	**5.03****	**7.57*****	2.82
	(5.31)	(4.70)	(8.19)	(3.99)	(5.01)	(5.98)	(8.47)	(3.70)
s(Turb):Sunset	4.27	1.00	**1.00*****	1.01	**6.53*****	**4.65*****	**1.00***	1.00
	(5.35)	(1.00)	(1.00)	(1.01)	(7.41)	(5.60)	(1.00)	(1.00)
s(Lon,Lat)					**25.96*****	**25.31*****	**26.73*****	**26.99*****
					(27.96)	(27.63)	(28.54)	(28.61)
AIC	3811.84	3820.00	3822.70	3776.45	3471.41	3459.43	3479.46	3397.44
Deviance explained	0.79	0.86	0.83	0.82	0.89	0.93	0.91	0.91
R^2	0.77	0.85	0.82	0.81	0.88	0.92	0.90	0.90
Num. obs.	602	602	602	602	602	602	602	602
Num. smooth terms	10	10	10	10	11	11	11	11

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(C)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	−5.44	**22.09*****	**18.21*****	**13.32*****	**18.62***	**23.48*****	**23.50*****	4.30
	(8.22)	(6.15)	(2.38)	(3.70)	(8.66)	(4.10)	(4.55)	(6.23)
s(Bottom):Night	**4.48*****	**6.68*****	**6.38*****	**6.14*****	2.19	1.00	1.00	1.00
	(5.28)	(7.13)	(6.93)	(6.76)	(2.59)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	1.00	**3.70***	1.00	1.00	**4.57*****	1.00	1.00	**4.10*****
	(1.00)	(4.34)	(1.00)	(1.00)	(5.18)	(1.00)	(1.00)	(4.80)
s(Bottom):Day	1.00	**2.89*****	**3.12*****	1.12	1.00	**3.02*****	1.00	1.71
	(1.00)	(3.45)	(3.69)	(1.22)	(1.00)	(3.67)	(1.00)	(2.01)
s(Temp):Night	1.65	**4.43*****	**6.14*****	**4.82***	2.08	**4.20*****	**6.28*****	**4.75*****
	(1.98)	(5.23)	(6.90)	(5.65)	(2.57)	(5.09)	(7.02)	(5.72)
s(Temp):Sunrise	**7.39*****	**6.62*****	**5.96*****	**7.86*****	**2.12****	**7.64*****	**6.54*****	**8.57*****
	(7.78)	(7.58)	(6.91)	(8.51)	(2.60)	(8.35)	(7.38)	(8.90)
s(Temp):Day	1.41	1.00	1.00	1.00	1.00	1.00	2.21	1.00
	(1.64)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(2.76)	(1.00)
s(Sal):Night	2.78	1.00	1.00	1.00	3.19	1.00	1.44	1.00
	(3.41)	(1.00)	(1.00)	(1.00)	(3.86)	(1.00)	(1.73)	(1.00)
s(Sal):Sunrise	**4.91*****	**6.77*****	**6.67*****	**8.16*****	1.00	**6.51*****	**7.00*****	**8.41*****
	(5.91)	(7.60)	(7.35)	(8.58)	(1.00)	(7.43)	(7.71)	(8.84)
s(Sal):Day	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.01)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**1.00***	**6.07*****	**5.39***	2.98	**3.76*****	**4.45***	**4.33***	1.00
	(1.00)	(7.15)	(6.47)	(3.77)	(4.64)	(5.49)	(5.38)	(1.00)
s(Fluo):Sunrise	**6.63*****	**3.72***	**2.91*****	**7.16*****	**2.61*****	**5.49*****	**4.35****	1.00
	(7.01)	(4.57)	(3.62)	(7.86)	(3.16)	(6.47)	(5.23)	(1.00)
s(Fluo):Day	1.00	1.50	1.00	**1.99***	1.00	1.00	1.00	1.00
	(1.00)	(1.68)	(1.00)	(2.15)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	**4.33*****	2.56	2.08	2.96	**4.75***	**3.38****	1.00	1.67
	(5.29)	(3.13)	(2.56)	(3.67)	(5.75)	(4.07)	(1.00)	(2.08)
s(Turb):Sunrise	**5.50*****	**5.94*****	**5.67*****	**4.28*****	**4.51*****	**5.67*****	**4.35*****	**4.31*****
	(5.99)	(6.41)	(6.45)	(4.90)	(5.09)	(6.20)	(5.22)	(4.96)
s(Turb):Day	1.06	**1.00****	**1.00*****	**3.16*****	1.63	1.00	1.00	**3.15***
	(1.10)	(1.00)	(1.00)	(3.63)	(1.98)	(1.00)	(1.00)	(3.67)
s(Lon,Lat)					**26.08*****	**25.28*****	**23.23*****	**26.73*****
					(27.96)	(27.65)	(26.31)	(28.43)
AIC	2934.35	2537.17	2502.36	2527.95	2605.35	2373.69	2394.55	2315.45
Deviance explained	0.76	0.84	0.85	0.84	0.91	0.91	0.90	0.92
R^2	0.73	0.81	0.82	0.82	0.89	0.88	0.87	0.90
Num. obs.	371	371	371	371	371	371	371	371
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3.2: Results of Generalized additive model (GAM) and spatial Generalized additive model(GAMs) showing the effect (blue: Positive; red: Negative; gray : Not significant) of oceanographic conditions (sea temperature, salinity, turbidity and fluorescence), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) on sound-scattering layers (SSLs) (1) thickness and (2) depth, detected by echosounders (38, 70, 120 and 200 kHz), according to three Senegalese geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf).

① (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)

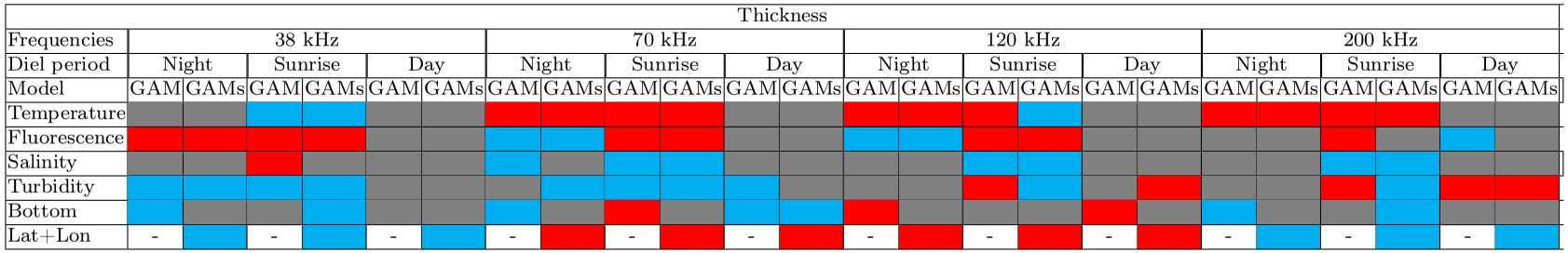


Table 3.3: Results of Multivariate Functional Principal Component Analysis (MFPCA) of oceanographic condition (sea Temperature (in °C), Salinity (in PSU), Fluorescence (in $ml1^{-1}$), Turbidity (in NTU)) along the depths) measured between the minimum and maximum depths of the layers at four frequencies. 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↗	25m↗
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↗

Table 3.4: 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 conditions (sea temperature, salinity, turbidity, fluorescence), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) on sound-scattering layers (SSLs) (1) thickness and (2) depth, 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).

① (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)

Frequencies	Thickness											
	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+	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)

Frequencies	Thickness																	
	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	-												-					

Table 3.5: Comparison of non-spatial (GAM and FGSAM) and spatial models (GAMs and FGSAMs) for (1) SSL thickness and (2) SSL depth, 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).

① (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

② (A)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.92	0.93	2.60	1.44	3.97	2.19
GAMs38	0.95	0.96	2.06	1.03	3.09	1.63
FGSAM38	0.92	0.93	2.53	1.73	4.09	2.43
FGSAMs38	0.95	0.96	1.88	1.22	3.05	1.86
GAM70	0.82	0.84	3.65	2.33	4.10	2.62
GAMs70	0.96	0.97	1.68	5.08	2.52	1.32
FGSAM70	0.84	0.87	3.35	2.42	3.55	2.66
FGSAMs70	0.96	0.97	1.60	1.07	2.37	1.67
GAM120	0.84	0.86	3.74	2.44	4.08	2.57
GAMs120	0.93	0.94	2.39	1.43	3.35	1.82
FGSAM120	0.90	0.92	2.86	1.89	3.37	2.26
FGSAMs120	0.93	0.94	2.46	1.52	3.20	1.89
GAM200	0.65	0.69	4.10	2.72	6.06	4.22
GAMs200	0.81	0.85	2.82	1.86	4.23	2.78
FGSAM200	0.70	0.75	3.68	2.77	5.25	3.74
FGSAMs200	0.79	0.82	3.10	2.07	4.53	2.85

(B)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.88	0.89	2.58	2.05	2.44	1.93
GAMs38	0.94	0.95	1.78	1.43	1.91	1.54
FGSAM38	0.83	0.86	2.86	2.17	3.40	2.47
FGSAMs38	0.92	0.93	1.96	1.51	2.33	1.84
GAM70	0.91	0.92	2.53	1.95	2.79	2.20
GAMs70	0.96	0.96	1.75	1.39	2.08	1.65
FGSAM70	0.84	0.86	3.35	2.41	3.79	2.58
FGSAMs70	0.86	0.89	3.01	2.16	3.50	2.47
GAM120	0.90	0.91	2.81	2.13	3.16	2.32
GAMs120	0.94	0.95	2.21	1.72	2.80	2.16
FGSAM120	0.89	0.91	2.88	2.06	3.70	2.66
FGSAMs120	0.91	0.93	2.54	1.86	3.48	2.50
GAM200	0.93	0.93	2.70	2.08	2.63	2.07
GAMs200	0.96	0.96	2.03	1.58	2.35	1.79
FGSAM200	0.70	0.75	5.18	3.30	6.47	4.19
FGSAMs200	0.73	0.77	4.96	3.22	6.23	4.02

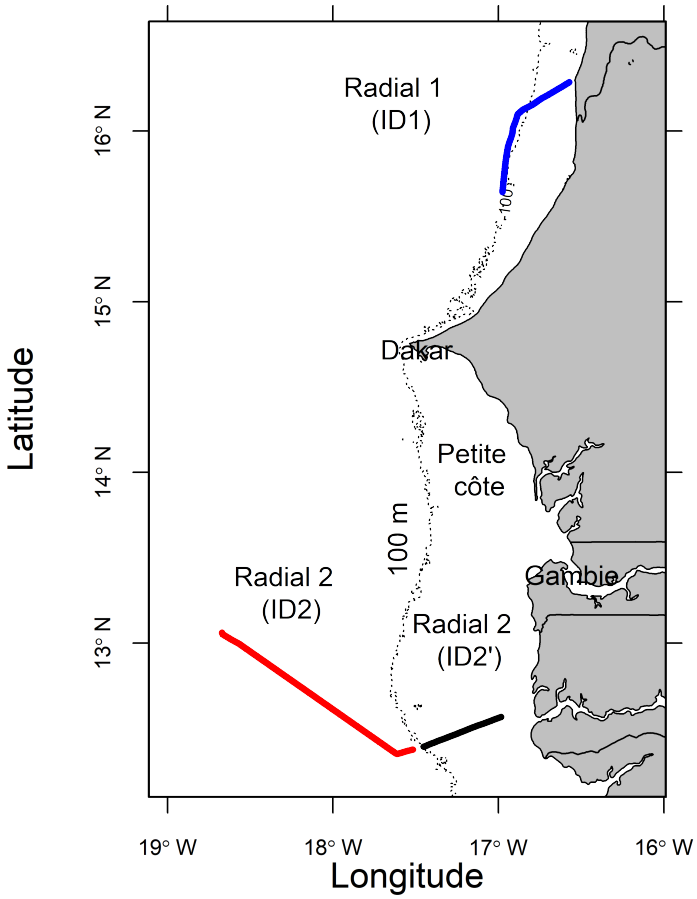


Figure 3.2: Map with the 100 m isobath of Senegalese coastline (represented by a curve), (West Africa). The vessel has covered three radials collecting simultaneously acoustics and environmental variables at high resolution (sea survey AWA). The radial (dotted lines) 1 was carried out over the continental shelf in northern Senegal (in blue color) and radial 2 in southern Senegal split into two parts, on-shore (continental shelf: ID2' in black color) and off-shore (high sea: ID2 in red color).

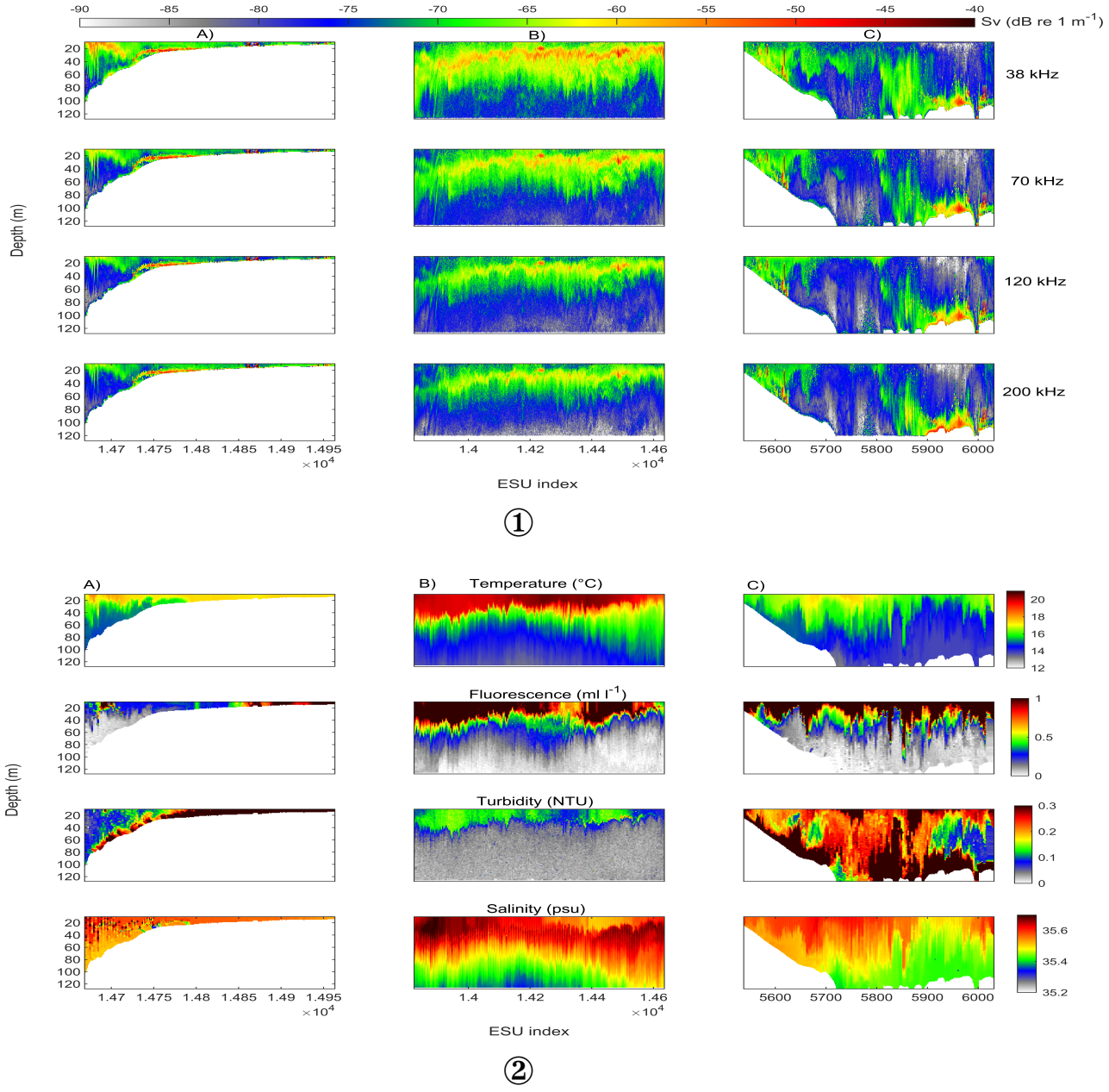


Figure 3.3: (1) Echogram representing the acoustic intensities (in Sv backscatter coefficient) reverberated by the aggregated marine organisms structured in layer (i.e. here sound scattering layer) as detected by echosounders at four frequencies (38, 70, 120 and 200 kHz) in three dimensions (2D + time). (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.

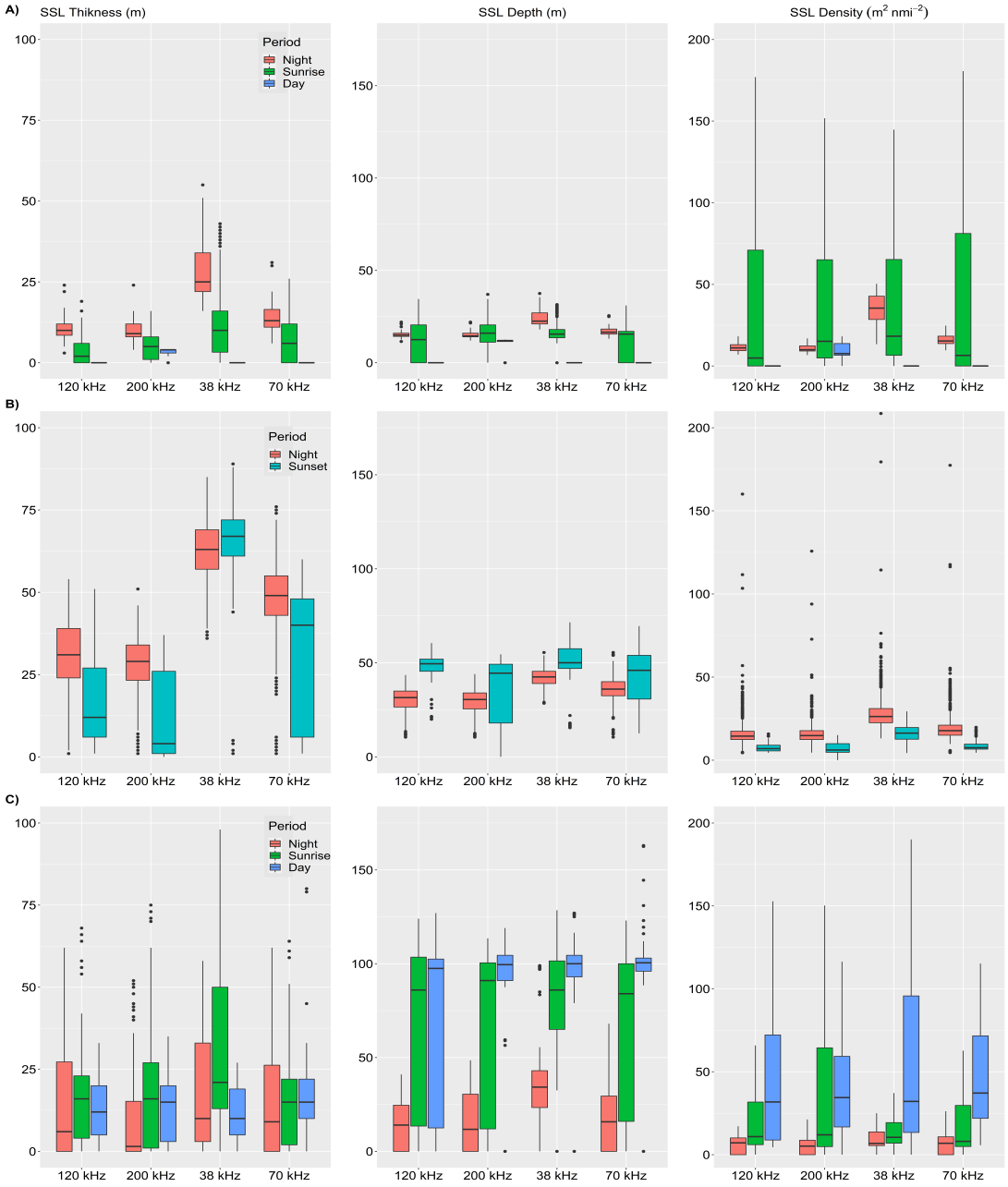
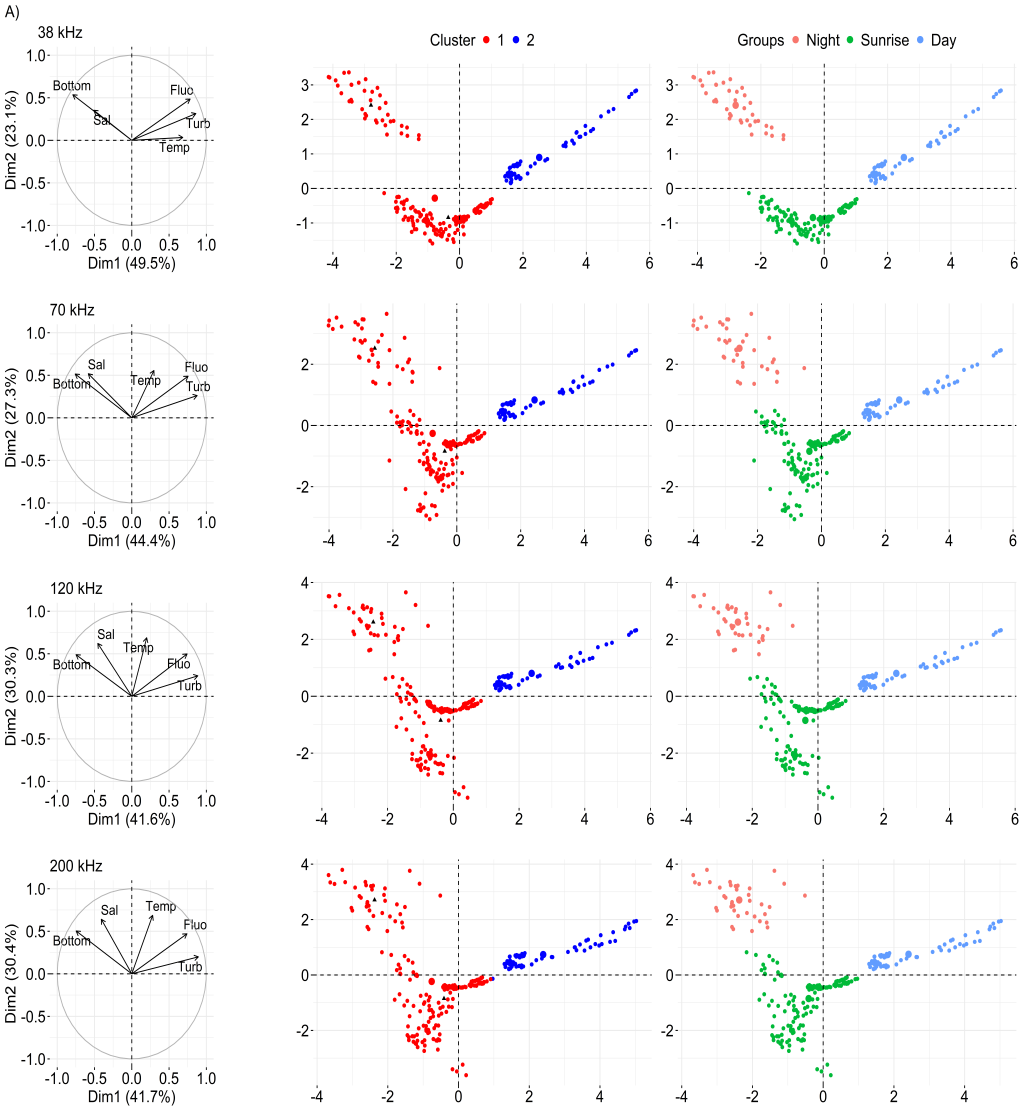
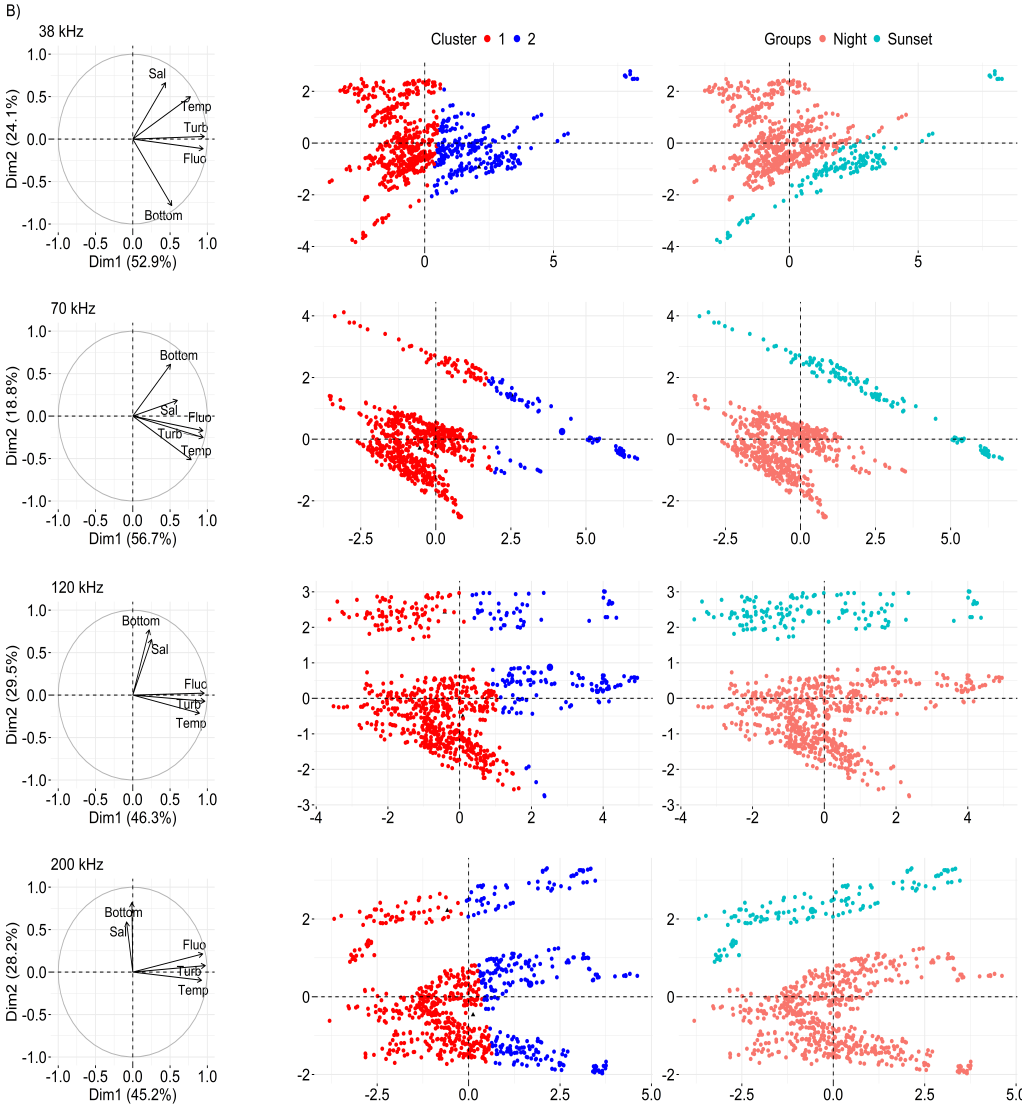


Figure 3.4: Boxplot (minimum, maximum, and median) of sound scattering layer (SSL) descriptors (SSL thickness (m), SSL density (expressed in $m^2\,nmi^{-2}$), and SSL depth (m)) grouped by diel period (Night, Sunrise, Day and Sunset) over Senegalese waters (A: southern continental shelf; B: southern high sea and C: northern continental shelf) as observed by the echosounders during the AWA sea survey at four different frequencies (kHz).

(C)

	R^2_ajusted	R^2	RMSE_Train	MAE_Train	RMSE_Test	MAE_Test
GAM38	0.95	0.96	7.59	5.53	9.09	6.40
GAMs38	0.97	0.98	5.24	3.90	6.42	4.87
FGSAM38	0.77	0.80	16.40	11.16	17.43	11.90
FGSAMs38	0.84	0.86	13.74	8.96	14.72	10.61
GAM70	0.81	0.84	6.22	4.30	50.66	35.05
GAMs70	0.98	0.99	4.93	3.24	11.30	5.65
FGSAM70	0.91	0.93	11.58	8.21	20.01	12.94
FGSAMs70	0.97	0.97	6.86	4.37	15.74	8.39
GAM120	0.98	0.98	5.38	3.13	18.38	6.64
GAMs120	0.99	0.99	4.38	2.44	8.06	4.09
FGSAM120	0.91	0.93	10.71	7.76	15.86	11.21
FGSAMs120	0.97	0.98	6.07	4.36	8.89	6.16
GAM200	0.97	0.98	6.52	4.08	9.97	5.76
GAMs200	0.99	0.99	4.29	2.72	8.66	4.78
FGSAM200	0.92	0.94	10.69	7.79	16.63	11.63
FGSAMs200	0.99	0.99	3.89	2.88	8.84	6.07




B)
38 kHz
Dim2 (24.1%)
Dim1 (52.9%)
Sal
Temp
Turb
Fluo
Bottom
Cluster 1 2
Groups Night Sunset
70 kHz
Dim2 (18.8%)
Dim1 (56.7%)
120 kHz
Dim2 (29.5%)
Dim1 (46.3%)
200 kHz
Dim2 (28.2%)
Dim1 (45.2%)

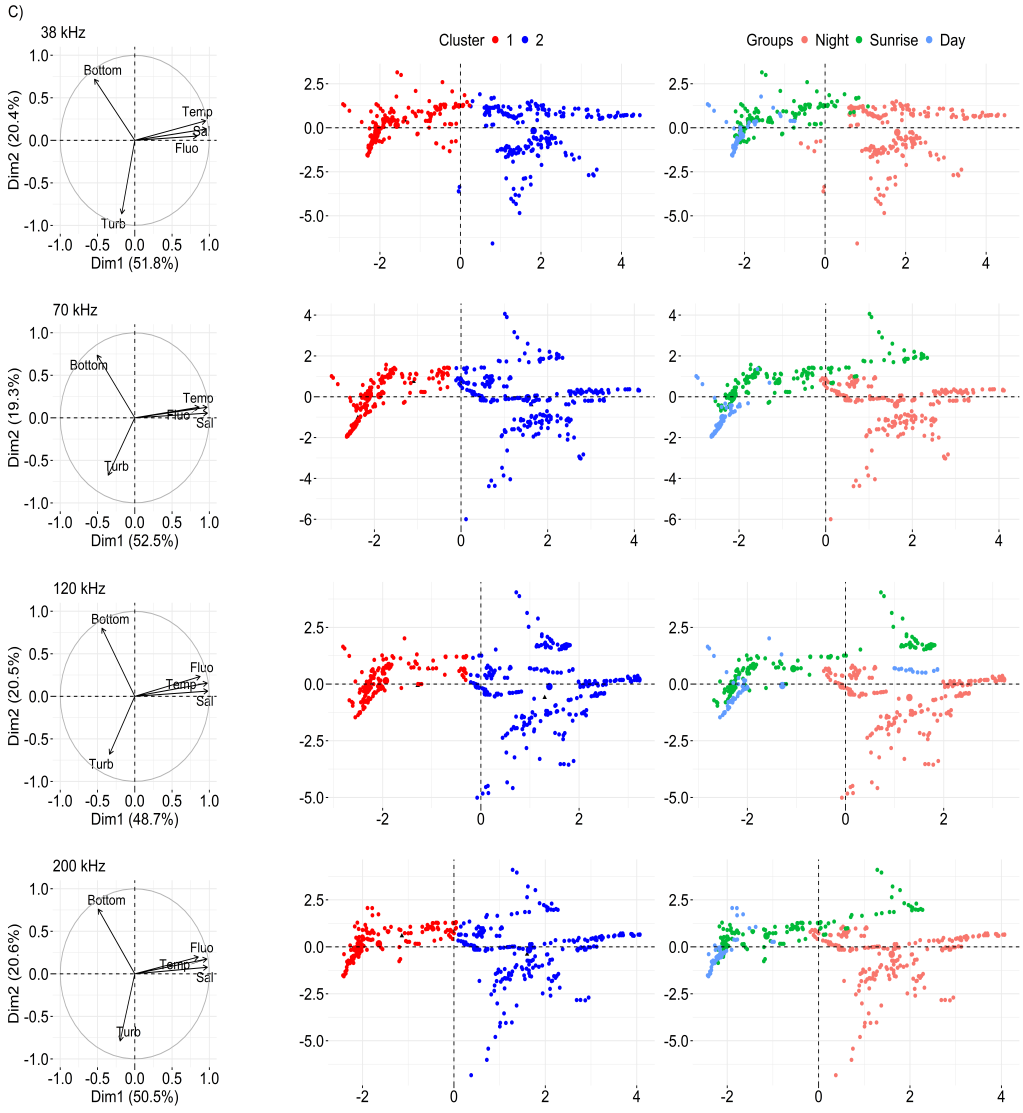


Figure 3.7: Principal Component Analysis (PCA: PC1 on x-axis and PC2 on y-axis) followed by classification (CAH) (red: Cluster 1; blue: Cluster 2) of the mean environmental parameters (sea temperature, fluorescence, turbidity, and salinity) measured within the minimum and maximum depths of the sound 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 (A: southern continental shelf; B: southern high sea and C: northern continental shelf) during the AWA sea survey.

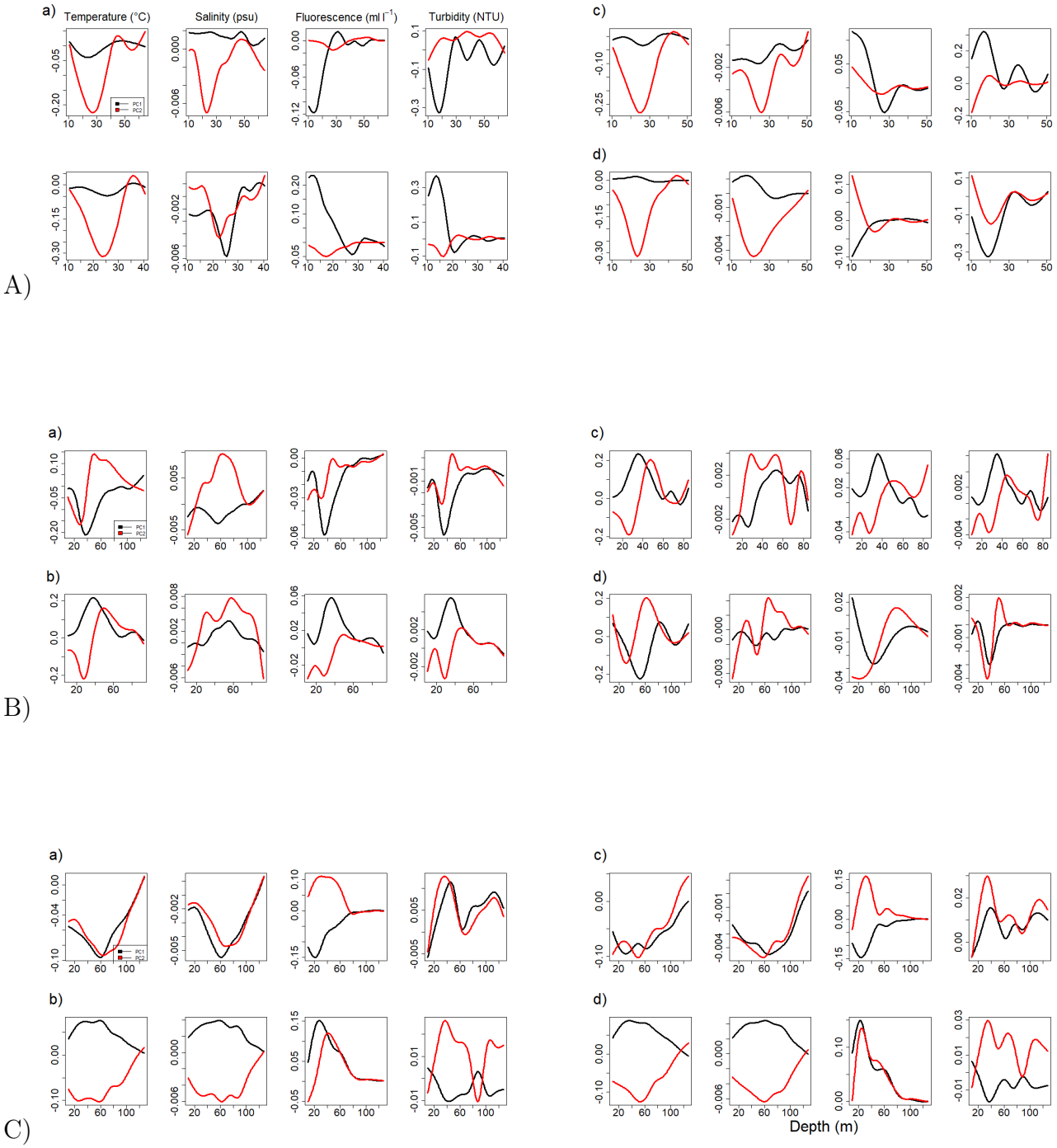
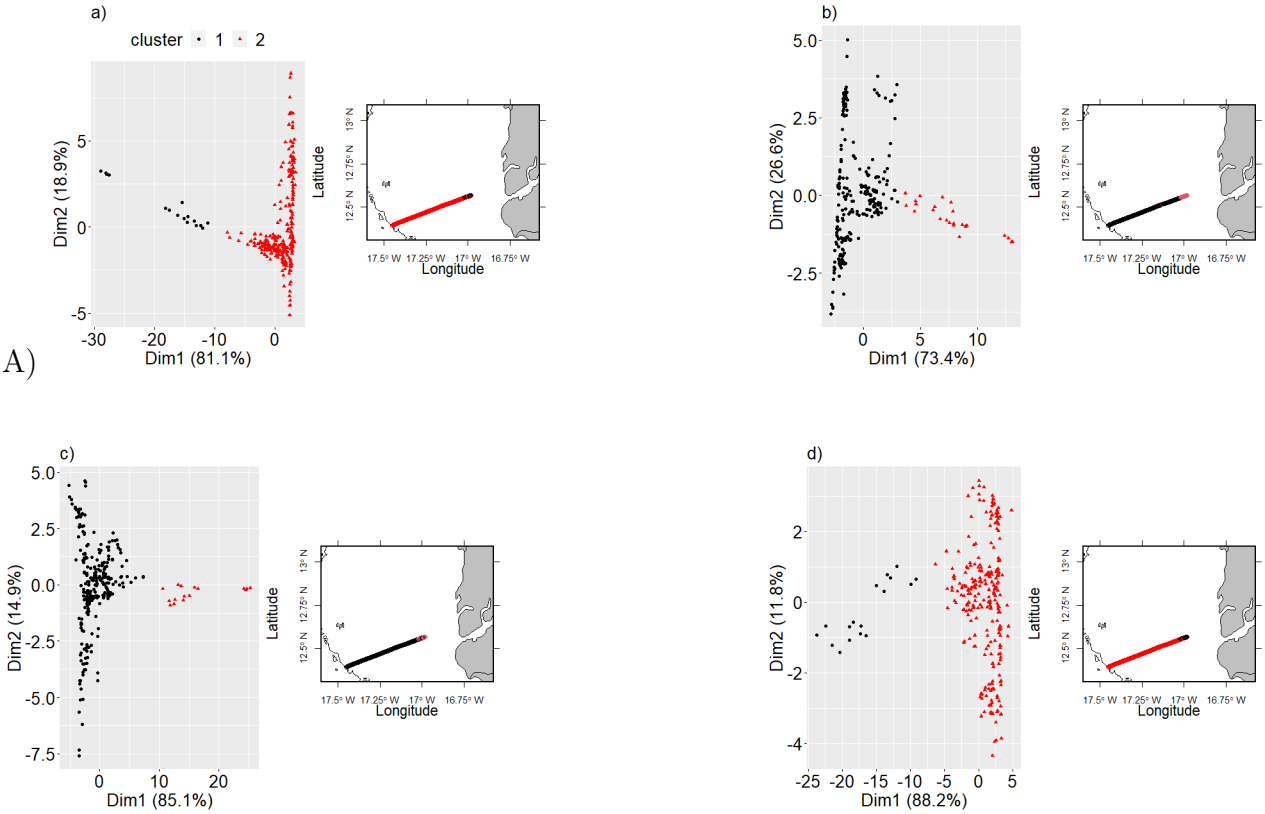
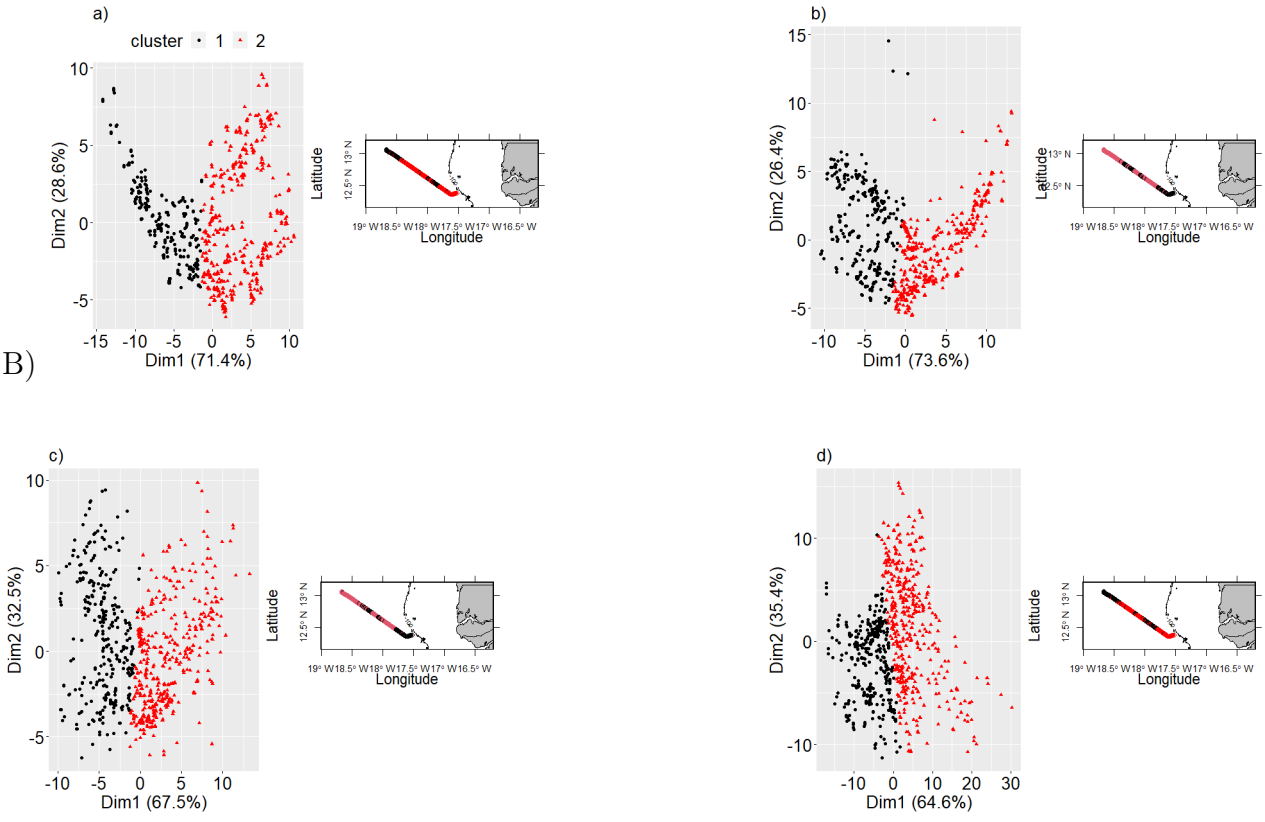


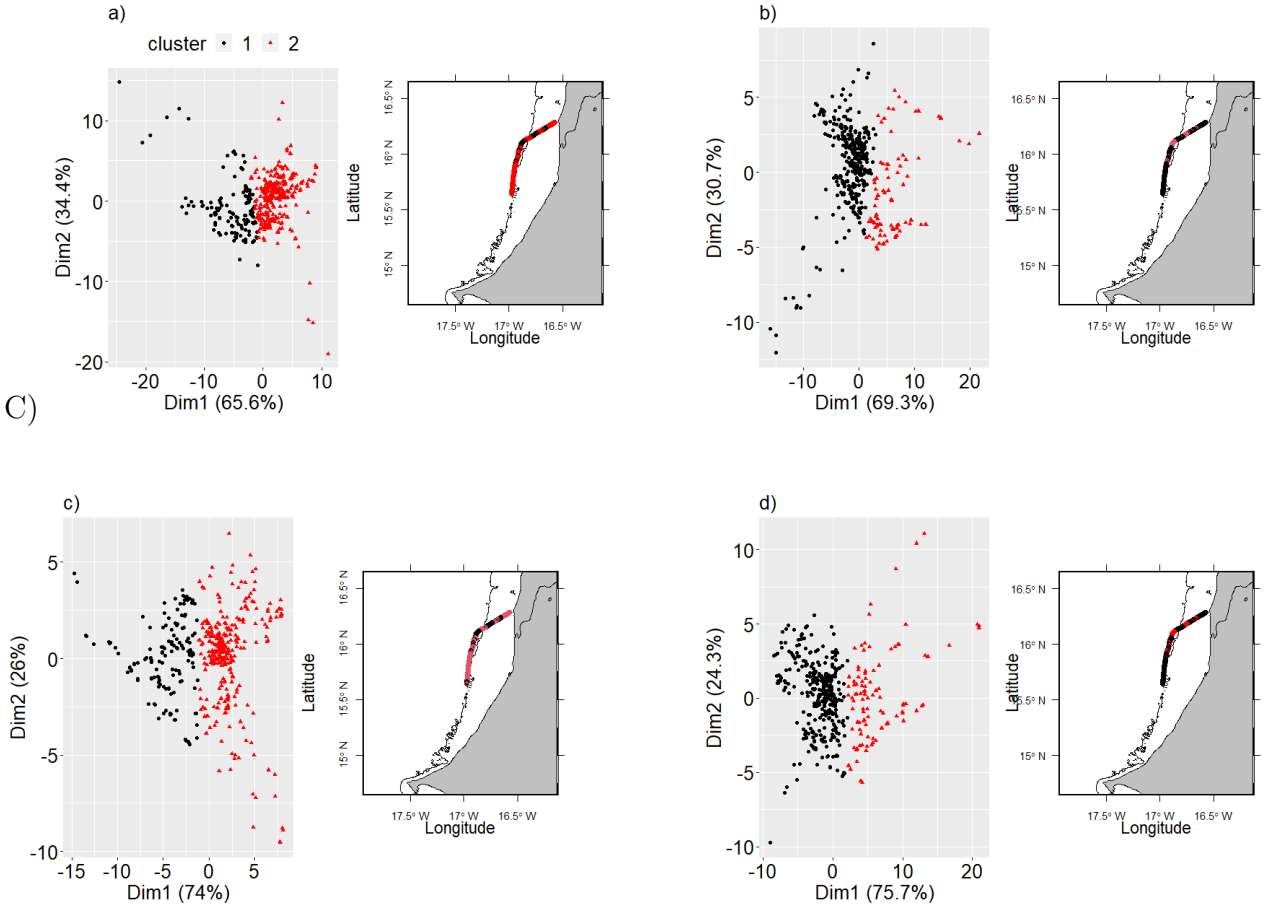
Figure 3.10: Multivariate Functional Principal Component Analysis (MFPCA) of oceanographic condition (sea temperature (in °C), salinity (in psu), fluorescence (in ml l^{-1}), turbidity (in NTU)) along the depths) measured between the minimum and maximum depths of the layers at four frequencies. Representation of the two main functional components (black : first component (PC1); red : second component (PC2)) over Senegalese waters (A: south continental shelf; B: south high sea and C: north continental shelf of the AWA sea survey), observed at various frequencies: (a) 38 , (b) 70 ,(c) 120 and (d) 200 kHz.

A)


a)
cluster 1 2
Dim2 (18.9%)
Dim1 (81.1%)
Latitude
Longitude
b)
Dim2 (26.6%)
Dim1 (73.4%)
c)
Dim2 (14.9%)
Dim1 (85.1%)
d)
Dim2 (11.8%)
Dim1 (88.2%)

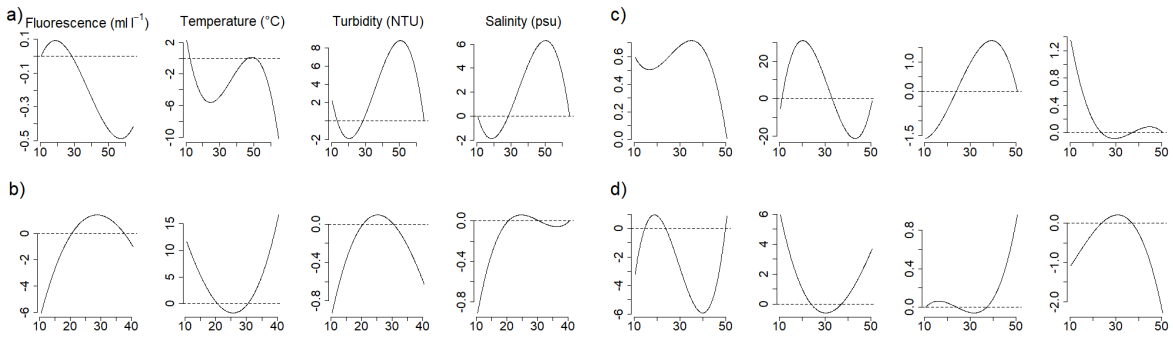
B)


a)
cluster 1 2
Dim2 (28.6%)
Dim1 (71.4%)
Latitude
Longitude
b)
Dim2 (26.4%)
Dim1 (73.6%)
c)
Dim2 (32.5%)
Dim1 (67.5%)
d)
Dim2 (35.4%)
Dim1 (64.6%)



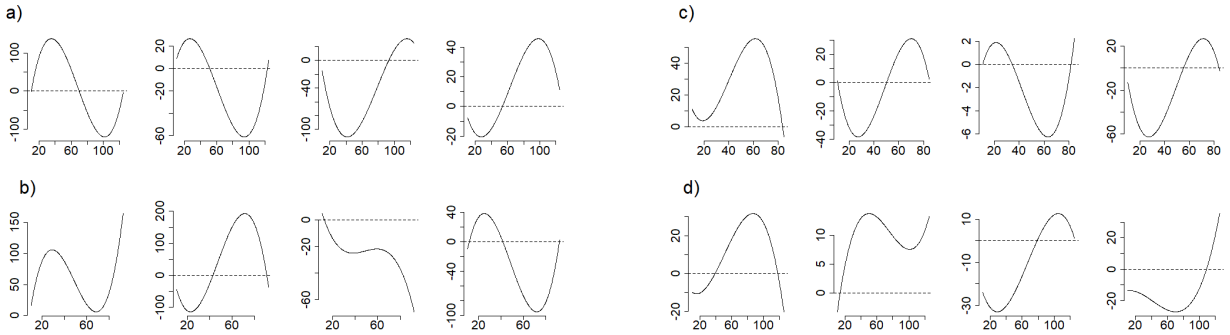
C)

Figure 3.13: Classification of functional principal component scores (derived from Multivariate functional principal component analysis MFPCA) (black: Cluster 1; red: Cluster 2) over Senegalese waters (A: south continental shelf; B: south high sea and C: north continental shelf of the AWA sea survey), observed at various frequencies: (a) 38, (b) 70, (c) 120 and (d) 200 kHz.

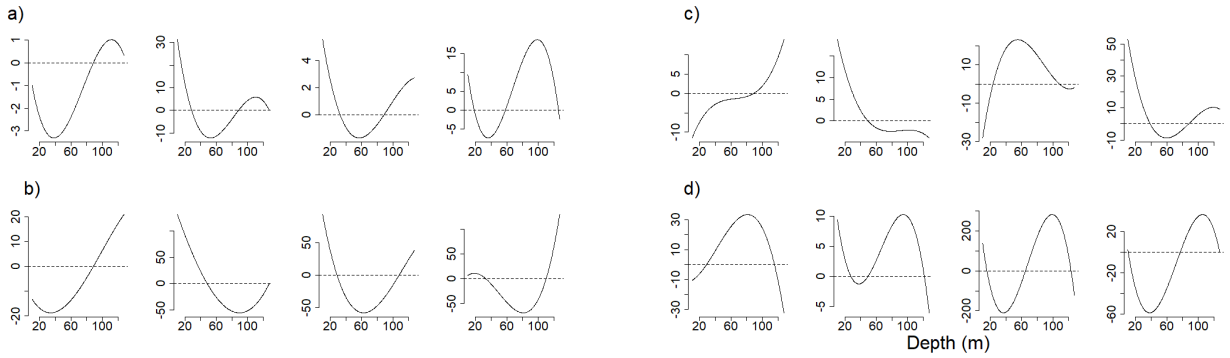


① A)

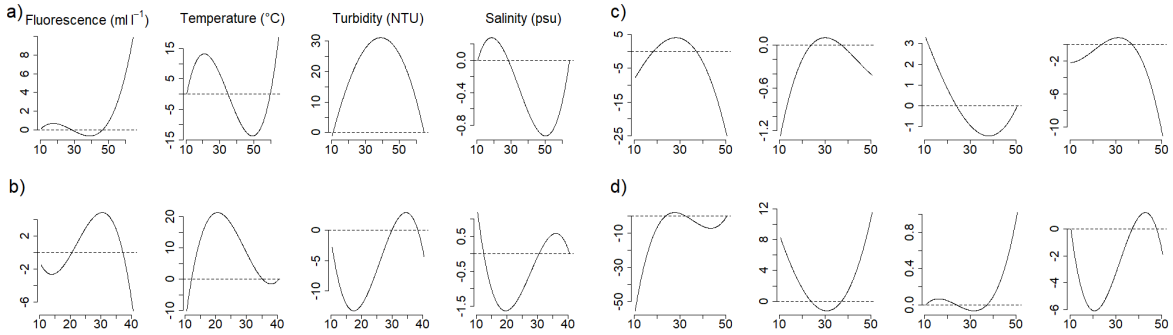
B)


a)
c)
b)
d)

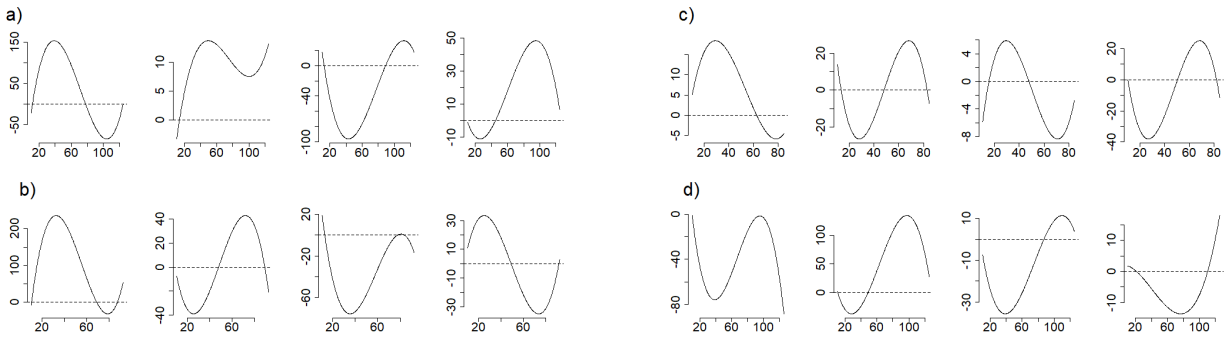
C)

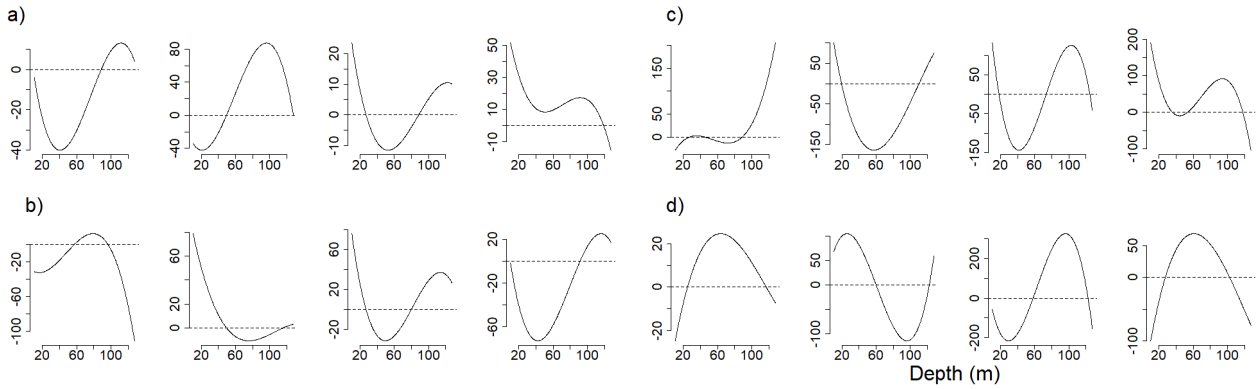

a)
c)
b)
d)
Depth (m)

② A)


a)
Fluorescence (ml l-1)
Temperature (°C)
Turbidity (NTU)
Salinity (psu)
c)
b)
d)

B)


a)
c)
b)
d)



C)

Figure 3.19: Estimated oceanographic condition parameters (sea temperature, fluorescence, salinity and turbidity) resulting from the functional model for (1) SSL thickness and (2) SSL depth with the spatial dimension (spatial Functionnal Generalized Spectral Additive Model (FGSAMs)) over three different areas (A: southern continental shelf of Senegal; B: southern high sea of Senegal; C: northern Senegal) and as expect at various echosounder frequencies (a) 38, (b) 70, (c) 120 and (d) 200 kHz.

Chapter 4

Study of the impact of environmental variables on scattering layers using signature-based regression

4.1 Abstract

Here we explore a new way of representing environmental data using the path signature (a collection of all the iterated integrals), which faithfully represents the form of variation of each environmental parameter. It offers understandable features and produces a vector of fixed length, regardless of the number of input points or their sampling depth. Furthermore, the product of two elements in a path signature can be represented as the summation of higher order terms that can encode a complex function of the parameter shape. The signatures were truncated to a finite order for the purposes of the analyses. A spatial principal component analysis was performed on these truncated signatures, revealing global spatial structures and strong spatial autocorrelation. The application of neural networks to the signature terms, taking into account the spatial aspect, gave good results in terms of mean square error.

4.2 Introduction

Oceanographic data collected over a wide spatio-temporal scale can be analysed using functional data analysis (FDA) methods. In recent years, this field has expanded considerably, as shown by the work

of Ramsay and Silverman (1997), Ferraty and Vieu (2006a), Yao et al. (2005), Mateu and Giraldo (2021), Wynne and Duncan (2022), Koner and Staicu (2023) among others.
The principle of FDA is to express discrete observations as a function using a set of basic functions such as Fourier basis, splines, wavelets etc and then derive modelling information from a collection of functional data by applying statistical concepts from multivariate data analysis.

In this work, we focus on a new approach to functional modelling based on the notion of signature. The principle consists in representing a functional data (here an environmental parameter) by its signature, defined as an infinite sequence of its iterated integrals (Fermanian, 2022; Lyons and Qian, 2002), a nonlinear transformation of paths. Signatures originate from the work of Chen (1958) in the 1960s and were further developed in the 1990s as key parameters of rough path theory. It uniquely characterises the path over a fixed interval up to temporal reparametrisation and tree equivalences (Hager, 2021). It is a non-parametric method for extracting characteristic features from data which are converted into a multi-dimensional path using various algorithms. In machine learning tasks, signature methods are increasingly used in several areas. These include (Morrill et al., 2019; Perez Arribas et al., 2018; Ghosh et al., 2021; Morrill et al., 2019) medicine and (Graham, 2013; Yang et al., 2015) character recognition. The aim of this chapter is to analyse the effect of environmental parameters represented using signatures methods on the characteristics of sound layers, which are essentially composed of micronekton and zooplankton. These parameters were sampled over a large spatial scale and measured at different depths.

4.3 Signature-based neural network regression

4.3.0.0.1 Concept of signature: This part introduces the concept of paths with finite variation and their signatures within a compact interval. The signature represents the integral of tensor products of differentials of the path, and it can be truncated to a desired order for analysis.

Let's consider a compact interval denoted as $\mathcal{T}$, and within this interval, we have a path represented as a p-dimensional function:

$$\begin{aligned} X : \quad & \mathcal{T} \to E \\ & t \to (X_t^{(1)}, ..., X_t^{(p)}) \end{aligned}$$

This means that for each point in time t within $\mathcal{T}$, our path X maps to a vector in E, with each component of the vector denoted as $X_t^{(i)}$ for i ranging from 1 to p.

We define a set of functions on $\mathcal{T}$ that have finite variation with respect to a parameter m, denoted as $BV_m(\mathcal{T}, E)$. This set consists of functions $X : \mathcal{T} \to E$ where the total variation of X with respect to m is finite.

The total variation of X with respect to m, denoted as $\|X\|_{TV,m}$, is calculated as the supremum (or the maximum value) over all finite partitions I of $\mathcal{T}$. Within each partition, we sum the mth power of the Euclidean norm of the differences between consecutive points, i.e., $\|X_{t_i} - X_{t_{i-1}}\|^m$. This entire expression is then raised to the power of $1/m$.

Now, we define a specific set, $\mathcal{C}_m(\mathcal{T}, E)$, which consists of continuous paths $X : \mathcal{T} \to E$ with finite m-variation. These paths are both continuous and have bounded variation.

Suppose we have a path X that belongs to $\mathcal{C}_m(\mathcal{T}, E)$. The signature of this path, denoted as $Sig(X)$, is represented as a sequence:

$$Sig(X) = (1, \mathbf{X}^1, \dots, \mathbf{X}^d, \dots)$$

Where the integral $\mathbf{X}^d$ is a meaningful expression and can be computed as follows:

$$\mathbf{X}^d = \int_{\substack{t_1 < \cdots < t_d \\ t_1, \dots, t_d \in \mathcal{T}}} \cdots \int \mathrm{d}X_{t_1} \otimes \ldots \otimes \mathrm{d}X_{t_d} \in E^{\otimes d}.$$

The truncated signature of order D is represented as $Sig^D(X) = (1, \mathbf{X}^1, \dots, \mathbf{X}^D)$, and this can be computed for any integer $D \geq 1$.

In the context we are considering, we assume that the vector space E is $\mathbb{R}^p$, and the path X belongs to the set of continuous paths with finite 1-variation on the compact interval $\mathcal{T}$. It's important to note that the assumption that X is in $\mathcal{C}_1(\mathcal{T}, \mathbb{R}^p)$ is actually less restrictive than requiring X to be in the L^2 space on $\mathcal{T}$ with values in $\mathbb{R}^p$. Furthermore, the space $\mathcal{C}_1(\mathcal{T}, \mathbb{R}^p)$ is equipped with a specific norm denoted as $||X||_{\mathcal{C}}$, which is defined as the sum of the total 1-variation of X and the maximum absolute value of X.

To work with these concepts effectively, let by convention $(\mathbb{R}^p)^{\otimes 0} = \mathbb{R}$, we define $T((\mathbb{R}^p))$ the tensor algebra space by

$$T((\mathbb{R}^p)) := \{(a_0, a_1, \dots, a_d, \dots) | \forall d \geq 0, a_d \in (\mathbb{R}^p)^{\otimes d}\}$$

and $T^d((\mathbb{R}^p))$ the d^{th} truncated tensor algebra space

$$T^d((\mathbb{R}^p)) := \bigoplus_{i=0}^{d} (\mathbb{R}^p)^{\otimes i}.$$

As a result, the signature $Sig(X)$ of the path X is an element of the tensor algebra space $T((\mathbb{R}^p))$, and the truncated signature of order D, denoted as $Sig^D(X)$, belongs to the D-th truncated tensor algebra space $T^D((\mathbb{R}^p))$.

Let the set denoted as A^*, which contains multi-indexes with entries in the range from 1 to p. For a multi-index J in A^* with a length of d, we can express J as a sequence of components $J = (i_1, \dots, i_d)$, where each i_j is an integer within the range from 1 to p, for every j in the range from 1 to d.

Let's consider the canonical orthonormal basis of the vector space E, which is $\mathbb{R}^p$. This basis is represented by $(e_i)_{i=1}^p$. For any positive integer d, the tensor product space $E^{\otimes d}$ is isomorphic to a free vector space generated by all possible words of length d using multi-indices from A^*. In this space, the basis is formed by vectors of the form $(e_{i_1} \otimes \cdots \otimes e_{i_d})$, where each component is selected from the set $\{1, 2, \dots, p\}$ for every i_j in the range from 1 to d.

Given this framework, we can express the signature of the path X (as defined in Levin et al. (2013)) in a different form. It can be rewritten as:

$$Sig(X) = 1 + \sum_{d=1}^{\infty} \sum_{(i_1, \dots, i_d)} \mathcal{S}_{(i_1, \dots, i_d)}(X) e_{i_1} \otimes \cdots \otimes e_{i_d} \in T((E)).$$

where

$$\mathcal{S}_{(i_1,\dots,i_d)}(X) = \int\limits_{\substack{t_1<\cdots<t_d \\ t_1,\dots,t_d\in\mathcal{T}}} \cdots \int \mathrm{d}X_{t_1}^{(i_1)} \dots \mathrm{d}X_{t_d}^{(i_d)} \in \mathbb{R}.$$

This representation allows us to express the signature of the path X in terms of its coefficients $\mathcal{S}_{(i_1,\dots,i_d)}(X)$ and the tensor product of the canonical orthonormal basis vectors $e_{i_1} \otimes \cdots \otimes e_{i_d}$, belonging to the tensor algebra space $T((E))$.

We define two vectors to represent the signature coefficients of the path X. The first vector, denoted as $\mathcal{S}(X)$, contains all the signature coefficients and is structured as follows:

$$\mathcal{S}(X) = (1, \mathcal{S}_{(1)}(X), \mathcal{S}_{(2)}(X), \dots, \mathcal{S}_{(p)}(X), \mathcal{S}_{(1,1)}(X), \mathcal{S}_{(1,2)}(X), \dots, \mathcal{S}_{(i_1,\dots,i_d)}(X), \dots)$$

Additionally, we define the truncated signature coefficients vector at order D for X, denoted as $\mathcal{S}^D(X)$. This vector includes the first D terms of the signature coefficients:

$$\mathcal{S}^D(X) = (1, \mathcal{S}_{(1)}(X), \mathcal{S}_{(2)}(X), \dots, \mathcal{S}_{(p,\dots,p)}(X))$$

As an aside, it's important to note that the dimension of the truncated signature coefficients vector at order D, denoted as $s_p(D)$, can be calculated as the sum of a geometric series:

$$s_p(D) = \sum_{d=0}^{D} p^d = \frac{p^{D+1} - 1}{p - 1}$$

This formula provides the dimension of the truncated signature coefficients vector when considering D terms.

In the present study, we do not observe a complete continuous path but a stream of environmental data sampled at different depth points. As in the previous chapter, we deal with four predictors (temperature, salinity, fluorescence and turbidity) leading to a $p = 4$ dimensional path. Considering $X_{\mathbf{s}}$, be the corresponding 4 dimensional random variable measured at different locations $\mathbf{s}$ and different depth points $u \in [T_1, T_2]$ then the truncated coefficients signature $\mathcal{S}^D(X_{\mathbf{s}})$ summarise the small-scale changes of $X_{\mathbf{s}}$ in $[T_1, T_2]$. By convention, the first term of the signature is always equal to 1. We have used iisignature (Reizenstein and Graham, 2018) which is a Python package that calculates iterated integral signatures at a fixed level $D = 2$, a high value of D increases the length of the signature (Reizenstein, 2017) and was very time consuming. We obtain a path signature of the form S=[1, (Temp), (Sal), (Fluo), (Turb), (Temp,Temp), (Temp,Sal), (Temp,Fluo), (Temp,Turb), (Sal,Temp), (Sal,Sal), (Sal,Fluo), (Sal,Turb), (Fluo,Temp), (Fluo,Sal), (Fluo,Fluo), (Fluo,Turb), (Turb,Temp), (Turb,Sal), (Turb,Fluo), (Tur,Turb)]. The 1^{st} order terms (for example (Temp)) correspond to the increments from the start to the end of the path temperature. The term (Temp, Fluo) corresponds to the case where the temperature increases quickly relative to the fluorescence.

We analyse the effect of predictor variables on layer descriptors using the following methods.

4.3.0.0.2 Spatial principal component Analysis and clustering:

We apply the areal spatial PCA (Jombart et al., 2008) on the path coefficients $\mathcal{S}^D(X_{\mathbf{s_i}})$, $i = 1, ..., n$ using Gabriel graph

weight matrix. Clustering has been done on the scores of the spatial PCA.

4.3.0.0.3 Principal Component Regression (PCR) (Jolliffe, 2002): A linear regression with covariate as PCA scores of the signature coefficients $\mathcal{S}^D(X_{\mathbf{s_i}})$, $i = 1, ..., n$ of dimension $n * D$; and response is the vector of the n descriptors $Y_{\mathbf{s_i}}$.
The number of PCA components has been chosen based on the explained variance (Jolliffe, 2002).

4.3.0.0.4 The ridge model (Hoerl and Kennard, 1970): . Let the linear model defined by:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{4.1}$$

where $\mathbf{X}$ is an matrix of the signature coefficients $\mathcal{S}^D(X_{\mathbf{s_i}})$, $i = 1, ..., n$ of dimension $n * (D + 1)$; $\mathbf{Y}$ is the vector of the n descriptors $Y_{\mathbf{s_i}}$; ϵ the vector of centered and i.i.d error terms (with variance σ^2) and $\beta = [\beta_1, ..., \beta D + 1]'$ the vector of parameters to be estimated. The ridge parameter estimator is

$$\hat{\beta}_{ridge} = (\mathbf{X'X} + \lambda I_p)^{-1}\mathbf{X'Y} \tag{4.2}$$

The case where $\lambda = 0$, the usual linear model has also been considered.

4.3.0.0.5 Multi-Layered neural networks (Shanmuganathan, 2016; Hastie et al., 2009): they are regression models (consisting of an input layer, one or multiple hidden layers and an output layer) inspired by the functioning of biological neurons in the human brain. The input layer, not performing any calculations, is made up of several neurons of dimension p with p the number of explanatory variables while the output layer contains the prediction and the hidden layers where the calculations are made link them together. A weight w_i is assigned to each input and the neuron's task is to multiply $X_{\mathbf{s}_i}$ by w_i. The output is as follows:

$$Y_{\mathbf{s}_i} = f\left(\sum_{i=1}^{p} w_i X_{\mathbf{s}_i} + b\right), i = 1, ..., n, \tag{4.3}$$

where f is the activation function (sigmoid, hyperbolic tangent, ReLU, softplus, identity etc).

We use the python library, SHAP (SHapley Additive exPlanations) to explain the outputs of the neural model (Lundberg and Lee, 2017). It calculates the impact of each predictor variable on the target variable.

4.4 Results

In the spatial PCA, we selected first 5 components that explain at least 90% of the total variance (see Figure 4.1). The results of the spatial PCA showed that there is only global structures and therefore positive autocorrelation. The Moran's I statistics calculated based on these principal components show significant spatial autocorrelation for the principal components reported in Table 4.2.

On the southern continental shelf of Senegal:

- At 38 kHz, principal axis 1 is positively correlated with the signatures coefficients related to (Sal, Turb), (Fluo, Temp),(Fluo, Fluo), (Fluo, Turb) and (Turb, Turb) and negatively correlated with the variables (Fluo), (Turb) and (Turb, Fluo); axis 2 is positively correlated with (Sal, Fluo) and negatively correlated with (Turb, Temp), (Fluo, Sal) and (Turb, Sal) ; axis 3 is positively correlated with the (Temp,Temp) and bottom variables and negatively correlated with the Temp variable; axis 4 is positively correlated with the (Sal) variable and negatively correlated with the (Temp, Fluo), (Temp,Turb) and (Sal,Sal), while axis 5 is positively correlated with the (Temp, Sal) and negatively correlated with the (Sal,Temp) variable.

- At 70 kHz, axis 1 is positively correlated with the variables (Fluo, Temp), (Fluo, Fluo), (Fluo, Turb), (Turb, Temp), (Turb, Fluo) and (Turb, Turb) and negatively correlated with the variables (Fluo), (Turb) and (Fluo, Sal); axis 2 is positively correlated with the (Sal) and (Sal, Sal) and negatively correlated with (Temp,Sal), (Turb, Sal) and bottom ; axis 3 is positively related to (Temp) and negatively related to the variables (Temp, Temp) and (Turb, Sal) while axis 4 is positively related to the variable (Sal, Temp) and negatively related to the variable (Temp, Turb); axis 5 has a negative correlation with (Temp, Fluo), (Sal, Fluo) and (Sal, Turb).

- At 120 kHz, axis 1 correlated positively with (Temp, Fluo), (Sal,Turb), (Fluo, Fluo) and (Fluo,Turb) and negatively with (Turb, Fluo), while axis 2 correlated positively with the variables (Sal,Temp), (Turb, Temp) and (Turb, Sal) and negatively with the variable (Temp, Sal); axis 3 is positively correlated with the (Fluo) and (Temp, Temp) and negatively correlated with the (Temp), (Fluo, Temp) variables, while axis 4 is positively correlated with the (Sal) and (Turb,Turb) and negatively correlated with the bottom, (Sal,Sal) and (Sal, Fluo); axis 5 is positively correlated with the (Temp, Turb) variable and negatively with (Turb) and (Fluo, Sal).

- 200 kHz, axis 1 is positively related to the variables (Fluo), (Sal,Sal) and (Turb,Fluo) and negatively related to the variables (Sal), (Sal,Turb), (Fluo,Fluo) and (Fluo,Turb) while axis 2 is positively related to the variables (Sal,Temp) and (Sal,Fluo) and negatively related to the variables (Turb,Temp) and (Turb,Sal) while axis 3 is positively related to the variables (Fluo,Temp) and (Fluo,Sal) ; axis 4 is positively linked to the variables (Temp,Temp), (Temp,Sal), (Temp,Turb) and (Turb,Turb) and negatively to the variables (Temp) and (Turb), while axis 5 is positively linked to the variable bottom and negatively to (Temp,Fluo). For more details, see Annexe 3.

Clustering of the spatial component scores showed more homogeneous classes on southern continental shelf of Senegal (Figure 4.2). For all frequencies, the 1^{st} component is more characteristic of the classes . At 38kHz Class 1 is characterised by low values of signatures coefficients related to (Sal,Turb), (Fluo, Temp),(Fluo, Fluo), (Fluo, Turb) and (Turb, Turb), while class 2 has the opposite effect. At 70 kHz Class 1 is characterised by low values of signatures coefficients related to (Fluo, Temp), (Fluo, Fluo), (Fluo,Turb), (Turb, Temp), (Turb, Fluo) and (Turb, Turb), while class 2 has the opposite effect.

At 120 kHz Class 1 is characterised by low values of signatures coefficients related to (Temp, Fluo), (Sal,Turb), (Fluo, Fluo) and (Fluo,Turb), while class 2 has the opposite effect. At 200 kHz, Class 2 is characterised by low values of (Temp, Fluo), (Sal,Turb), (Fluo, Fluo) and(Fluo,Turb), while Class 1 has the opposite effect (see Annexe3 for more details).

In the supervised analysis, we conducted PCR, linear, ridge and neural network regression models with and without spatial coordinates as covariates. We are interested in estimating a regression model between the four environmental parameters and SSL thickness and SSL depth in different part of the Senegalese region. Train and test samples have been used, mean square errors and coefficient of adjustment are computed on the two samples.
The neural network model with geographical coordinates based on layer descriptors produced better results overall than those obtained with the others (low RMSE values, see Table 4.1).

The results of the model on the thickness (Figure 4.3) showed that on the continental shelf of southern Senegal, at 38 and 70 kHz, the bottom depth is one of the most important variables with a positive effect on thickness, while the opposite effect is noted for the geographical coordinates. At 120 and 200 kHz, the geographical coordinates are among the most important in the model, and as they increase, the thickness decreases, as does (Temp, Sal), while the opposite effect is noted for the variable (Sal).
With regard to the SSL depth model, for all frequencies, geographical coordinates are among the most important variables, and as they increase, SSL depth decreases.

On the high sea of the southern Senegal, at 38 kHz, the variables (Temp, Sal) and (Sal,Turb) are among the most important in the model, as they increase, the thickness decreases, while the opposite effect is observed for the variables (Sal,Sal) and (Temp,Temp). At 70 kHz, the period is the most important variable, followed by (Fluo), whose increase leads to greater thickness, while the opposite effect is noted for (Turb). At 120 kHz, as the variables (Turb,Temp), longitude and (Temp) increase, so does the thickness, with the opposite effect noted for latitude. At 200 kHz, as (Fluo,Fluo) and (Temp,Temp) increase, so does the thickness, with the opposite effect noted for bottom depth.

In terms of the model for SSL depth, at 38 kHz, as the variables (Sal,Turb) and (Temp,Sal) increase, layer depth decreases. The same effect is observed at 70 kHz with the variables (Turb) and (Fluo,Fluo) and at 120 kHz with the variables (Sal,Sal) and (Fluo,Turb). At 200 kHz, as (Turb,Fluo) increases, the depth of the layers decreases, while the opposite effect is noted for (Fluo,Fluo).

In northern continental shelf, at 38 kHz, the period is the most important variable in the model for thickness, and as the variables (Sal,Temp) and (Fluo,Temp) increase, thickness decreases, while the opposite effect is noted for latitude. At 70 and 200 kHz, as longitude increases so does thickness, while the opposite effect is noted for latitude. At 120 kHz, as the variables (Temp,Temp) and (Temp,Turb) increase, so does the thickness, while the opposite effect is noted for (Fluo,Sal) and (Sal,Sal).

For the SSL depth, at 38 kHz, as the latitude increases the depth decreases while the opposite effect is noted for the variable (Fluo,Fluo). At 70 and 200 kHz, the geographical coordinates are the most important in the model, as they increase, the depth decreases, except for the longitude for the

depth of the 200 kHz SSL, where the opposite effect is noted. For the depth of the 120 kHz SSL, the more the longitude and the variable (Temp) increase, the more the depth decreases.

4.5 Discussion

Traditional principal component analysis is a dimension-reduction technique that explores the main sources of variation in a table of data (Jolliffe, 2002). However, it does not take spatial dependency into account when transforming multidimensional data. Applying spatial Principal Component Analysis (sPCA) shows the presence of spatial autocorrelation and reveals spatial global patterns i.e. the values of the environmental parameters observed in neighbouring ESUs tended to be similar. On the continental shelf of southern Senegal, we find a cluster of esus closer to the 100m isobath unlike another cluster which was shallower because it contains esus closer to the coast where SSLs were seldom as expected by Diogoul et al. (2020) who explains this by the turbulence in the water column generated by the upwelling phenomenon. This explains the higher values of certain environmental parameters such as temperature, fluorescence and turbidity measured in this class. The path signature was used to encode the interactions between the environmental parameters in order to predict the characteristics of the scattering layers. The regression coefficients acquired through the signature neural model possess a geometric interpretation, which is frequently beneficial for practical applications. In contrast to the coefficients in conventional functional linear models, these coefficients represent global measures of interaction between coordinates (Fermanian, 2022). Sugiura and Hosoda (2020) uses this method to encode non-linear interactions and functions in argo profiles in addition to the usual linear term before applying machine learning methods. On the continental shelf of southern Senegal the characteristics of the layers (depth and thickness) diminish at higher latitudes and longitudes, as they are closer to the coast. On the high sea, for the 38 kHz thickness model, values of the (Temp,Temp) signature terms correspond to the variation in temperature according to the depth and the SSLs are linked to strong vertical gradients of this parameter (Berge et al., 2014). For other frequencies, the signature terms (Fluo) and (Fluo,Fluo) also correspond to the variation in fluorescence (or chlorophyll-a concentration) according to the depth (last value minus initial value) and have an impact on SSl thickness because there are linked to the presence of a trophic relationship between the zooplankton and phytoplankton. In northern continental shelf, the value (Fluo,Temp) is the area under the curve (Fluorescence, Temperature). Its corresponding coefficient, therefore, contains information about the importance of the joint evolution of fluorescence and temperature to predict SSL features.

4.6 Conclusion

We examine the application of the signature method in a learning context on our environmental analysis. The use of signatures is based on the representation of discretely sampled data in the form of continuous paths. The signature method is a versatile approach for creating a set of features from sequential data, and is currently attracting a great deal of interest in the scientific community. Indeed, it produces results that are competitive with functional analysis methods, while being computationally efficient and capable of handling multidimensional series. This method captures the geometric proper-

ties of the process underlying the data and does not depend on a basis (fourier, bspline etc) expansion. The results for RMSE, as shown in Table 4.1, clearly demonstrate the significant enhancement in performance achieved by incorporating the spatial aspect.

Table 4.1: Comparison of non-spatial (NN (Neural Network), R (Ridge) and PCR (Principal Component Regression)) and spatial models (NNs (Spatial Neural Network), Rs (Spatial Ridge), PCRs (Spatial Principal Component Regression), NNspca (Neural Network on spatial principal component scores) and LRspca (Linear Regression on spatial principal component scores)) for (1) SSL thickness and (2) SSL depth, 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 (R-Squared); RMSE_Train (Root Mean Square Error in training set); RMSE_Test (Root Mean Square Error in test set).

① (A)

	RMSE_Train	R^2	RMSE_Test
NN38	1.943	0.979	3.855
NN70	1.991	0.926	2.412
NN120	0.674	0.980	1.196
NN200	0.004	1.000	2.188
R38	5.221	0.850	4.458
R70	3.779	0.735	4.103
R120	2.365	0.751	1.918
R200	3.272	0.469	3.298
PCR38	6.167	0.790	5.658
PCR70	4.474	0.628	3.974
PCR120	2.534	0.714	2.158
PCR200	3.462	0.405	3.118
NNs38	1.781	0.983	3.553
NNs70	1.673	0.948	1.986
NNs120	0.587	0.985	1.147
NNs200	0.005	1.000	2.015
Rs38	5.003	0.862	4.339
Rs70	3.019	0.831	5.378
Rs120	2.243	0.776	1.951
Rs200	2.834	0.602	3.034
PCRs38	5.512	0.833	5.132
PCRs70	4.019	0.700	3.695
PCRs120	2.472	0.728	2.122
PCRs200	3.299	0.460	3.051
NNspca38	0.005	1.0	3.539
NNspca70	0.006	1.0	2.525
NNspca120	0.008	1.0	2.836
NNspca200	0.004	1.0	2.842
LRspca38	9.490	0.504	7.414
LRspca70	5.452	0.448	5.115
LRspca120	2.726	0.669	2.498
LRspca200	3.513	0.388	3.201

(B)

	RMSE_Train	R^2	RMSE_Test
NN38	3.096	0.931	4.333
NN70	2.899	0.963	6.238
NN120	2.069	0.975	5.450
NN200	1.324	0.988	4.113
R38	9.059	0.410	7.605
R70	9.674	0.583	11.122
R120	8.825	0.549	9.392
R200	8.899	0.447	9.457
PCR38	9.441	0.359	8.393
PCR70	11.280	0.433	13.107
PCR120	10.618	0.347	11.864
PCR200	9.420	0.381	10.125
NNs38	2.357	0.960	3.731
NNs70	2.056	0.981	5.650
NNs120	1.316	0.990	5.166
NNs200	1.439	0.986	3.808
Rs38	8.951	0.424	7.345
Rs70	9.424	0.604	10.918
Rs120	8.379	0.594	8.771
Rs200	8.840	0.455	9.406
PCRs38	9.477	0.354	8.408
PCRs70	11.301	0.431	13.133
PCRs120	10.312	0.384	11.362
PCRs200	9.331	0.392	10.017
NNspca38	1.056	0.992	4.106
NNspca70	1.772	0.986	7.622
NNspca120	0.835	0.996	5.922
NNspca200	1.461	0.985	4.468
LRspca38	10.890	0.147	8.757
LRspca70	11.350	0.426	12.828
LRspca120	10.782	0.327	12.207
LRspca200	9.452	0.377	10.200

(C)

	RMSE_Train	R^2	RMSE_Test
NN38	1.574	0.996	6.214
NN70	1.658	0.989	4.780
NN120	1.246	0.993	2.853
NN200	1.240	0.994	5.079
R38	19.377	0.366	17.933
R70	11.778	0.441	10.602
R120	11.616	0.423	11.688
R200	12.032	0.441	12.671
PCR38	21.123	0.246	18.941
PCR70	12.698	0.351	11.334
PCR120	13.032	0.273	11.837
PCR200	13.687	0.276	12.558
NNs38	1.190	0.998	4.560
NNs70	0.672	0.998	4.415
NNs120	1.834	0.986	3.115
NNs200	2.584	0.974	4.957
Rs38	19.320	0.370	18.085
Rs70	11.272	0.488	10.517
Rs120	11.282	0.455	11.119
Rs2000	11.597	0.481	11.937
PCRs38	20.785	0.270	19.231
PCRs70	12.725	0.348	11.156
PCRs120	13.230	0.251	12.130
PCRs200	13.814	0.263	12.639
NNspca38	0.974	0.998	6.960
NNspca70	0.976	0.996	6.662
NNspca120	0.768	0.997	5.652
NNspca200	5.228	0.894	6.852
LRspca38	21.729	0.203	19.208
LRspca70	12.849	0.335	11.389
LRspca120	13.598	0.209	12.376
LRspca200	14.324	0.208	12.766

② (A)

	RMSE_Train	R^2	RMSE_Test
NN38	0.005	1.000	3.086
NN70	3.626	0.847	4.800
NN120	1.242	0.985	2.457
NN200	0.005	1.000	1.607
R38	3.950	0.842	3.246
R70	4.876	0.724	6.253
R120	5.564	0.694	4.780
R200	5.861	0.454	5.142
PCR38	4.259	0.816	4.180
PCR70	6.403	0.524	5.828
PCR120	6.450	0.588	5.561
PCR200	6.512	0.326	4.794
NNs38	0.006	1.000	1.529
NNs70	2.644	0.919	2.307
NNs120	0.006	1.0	1.426
NNs200	0.009	1.000	1.454
Rs38	3.822	0.852	2.983
Rs70	3.753	0.836	3.620
Rs120	3.982	0.843	3.906
Rs200	4.910	0.617	4.055
PCRs38	3.934	0.843	3.821
PCRs70	5.129	0.695	4.760
PCRs120	5.971	0.647	5.072
PCRs200	6.279	0.374	4.673
NNspca38	0.004	1.0	2.931
NNspca70	0.039	1.0	4.803
NNspca120	3.185	0.9	2.829
NNspca200	0.006	1.0	2.242
LRspca38	6.216	0.608	4.565
LRspca70	7.188	0.400	6.708
LRspca120	7.234	0.482	5.860
LRspca200	3.513	0.388	3.201

(B)

	RMSE_Train	R^2	RMSE_Test
NN38	0.190	0.999	2.683
NN70	1.853	0.960	3.575
NN120	1.167	0.985	2.562
NN200	0.892	0.993	4.597
R38	5.516	0.485	4.659
R70	6.473	0.508	6.507
R120	5.074	0.712	6.395
R200	9.244	0.239	7.453
PCR38	5.653	0.460	4.958
PCR70	7.607	0.321	7.450
PCR120	5.592	0.650	7.367
PCR200	9.955	0.118	8.466
NNs38	1.370	0.968	2.017
NNs70	1.384	0.978	3.483
NNs120	0.038	1.0	2.849
NNs200	0.813	0.994	4.334
Rs38	5.357	0.515	4.378
Rs70	6.207	0.548	6.283
Rs120	5.061	0.713	6.385
Rs200	9.218	0.244	7.391
PCRs38	5.653	0.460	4.875
PCRs70	7.576	0.327	7.428
PCRs120	5.631	0.645	7.406
PCRs200	9.956	0.118	8.464
NNspca38	0.686	0.992	2.452
NNspca70	1.673	0.967	3.960
NNspca120	1.609	0.971	3.153
NNspca200	0.899	0.993	5.344
LRspca38	6.278	0.334	5.002
LRspca70	7.694	0.305	7.568
LRspca120	5.654	0.642	7.329
LRspca200	9.971	0.115	8.403

(C)

	RMSE_Train	R^2	RMSE_Test
NN38	2.136	0.997	7.215
NN70	1.136	0.999	11.866
NN120	1.054	0.999	8.299
NN200	2.431	0.997	9.868
R38	29.311	0.360	29.596
R70	29.372	0.556	30.474
R120	29.617	0.525	30.546
R200	29.990	0.526	31.612
PCR38	30.999	0.284	32.074
PCR70	31.792	0.480	32.738
PCR120	34.079	0.371	33.917
PCR200	36.443	0.301	36.476
NNs38	3.678	0.99	7.05
NNs70	1.600	0.999	7.075
NNs120	1.188	0.999	5.461
NNs200	2.133	0.998	6.538
Rs38	22.453	0.624	21.597
Rs70	19.716	0.800	19.222
Rs120	22.428	0.727	20.359
Rs200	19.103	0.808	19.073
PCRs38	24.475	0.554	24.899
PCRs70	24.974	0.679	26.459
PCRs120	28.823	0.550	27.675
PCRs200	28.386	0.576	29.179
NNspca38	1.075	0.999	10.707
NNspca70	0.086	1.000	8.346
NNspca120	3.171	0.995	5.679
NNspca200	0.583	1.000	10.486
LRspca38	33.331	0.172	34.043
LRspca70	33.372	0.427	32.924
LRspca120	35.072	0.334	34.736
LRspca200	36.893	0.283	36.142

Table 4.2: Moran's test and explained variance on principal components scores using traditional principal component analysis (PCA) and spatial principal component analysis (SPCA), based on Gabriel graph weight matrices for environmental parameters signatures (sea temperature, fluorescence, turbidity, and salinity) 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)

(A)

	38 kHz		70 kHz		120 kHz		200 kHz	
	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)
PCA								
1^{st} score	0.9730***	40.65	0.9582***	51.17	0.9385***	42.66	0.9336***	42.41
2^{st} score	0.8477***	22.46	0.8287***	16.07	0.8291***	18.07	0.5832***	28.56
3^{st} score	0.7863***	15.54	0.7533***	9.84	0.8628***	14.09	0.8291***	10.37
4^{st} score	0.8806***	8.86	0.6600***	7.52	0.9143***	8.82	0.6987***	7.90
5^{st} score	0.8104***	6.09	0.7485***	6.24	0.7922***	6.84	0.8252***	3.37
SPCA								
1^{st} score	0.9744***	44.94	0.9596***	57.37	0.9433***	45.89	0.9361***	50.95
2^{st} score	0.8510***	21.65	0.8621***	15.90	0.8333***	17.16	0.6323***	22.20
3^{st} score	0.8078***	14.06	0.7496***	8.63	0.8658***	13.92	0.7027***	10.34
4^{st} score	0.8797***	8.86	0.7291***	6.10	0.9007***	9.16	0.7050***	7.13
5^{st} score	0.8041***	5.58	0.6691***	5.13	0.8016***	6.23	0.8358***	3.63

(B)

	38 kHz		70 kHz		120 kHz		200 kHz	
	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)
PCA								
1^{st} score	0.9827***	64.50	0.9726***	55.60	0.8431***	53.51	0.8888***	33.81
2^{st} score	0.9549***	16.59	0.8446***	19.28	0.8548***	15.51	0.9222***	30.09
3^{st} score	0.8807***	10.31	0.9677***	10.23	0.8609***	13.29	0.8686***	22.24
4^{st} score	0.9731***	3.10	0.8793***	8.00	0.9026***	7.89	0.9737***	8.26
5^{st} score	0.8527***	2.65	0.9492***	3.66	0.8894***	3.74	0.8996***	2.22
SPCA								
1^{st} score	0.9828***	65.95	0.9727***	57.79	0.8546***	53.74	0.9025***	33.66
2^{st} score	0.9559***	16.49	0.8514***	17.47	0.8482***	15.60	0.9195***	30.85
3^{st} score	0.8813***	9.45	0.9706***	10.60	0.8589***	13.52	0.8537***	21.28
4^{st} score	0.9722***	3.14	0.8647***	7.45	0.9065***	8.45	0.9724***	8.94
5^{st} score	0.8506***	2.35	0.9381***	3.69	0.8671***	3.89	0.9045***	2.23

	38 kHz		70 kHz		120 kHz		200 kHz	
	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)	Moran's I	variability (%)
PCA								
1^{st} score	0.9255***	41.89	0.9429***	50.81	0.9388***	54.88	0.8886***	45.13
2^{st} score	0.7555***	18.97	0.7189***	13.71	0.8176***	12.84	0.5499***	16.25
3^{st} score	0.7796***	12.87	0.7908***	11.19	0.8206***	11.50	0.8161***	14.12
4^{st} score	0.6772***	8.77	0.8645***	9.83	0.8343***	8.94	0.8029***	9.26
5^{st} score	0.8564***	7.98	0.7545***	4.68	0.9257***	3.78	0.7289***	4.88
SPCA								
1^{st} score	0.9282***	46.86	0.9443***	55.15	0.9391***	58.18	0.9046***	51.12
2^{st} score	0.7914***	17.73	0.7285***	11.40	0.8421***	12.04	0.8344***	15.08
3^{st} score	0.7750***	12.06	0.7935***	10.18	0.8008***	10.51	0.5880***	10.99
4^{st} score	0.8633***	8.34	0.8589***	9.77	0.8365***	8.43	0.7118***	9.12
5^{st} score	0.6697***	7.02	0.8321***	4.16	0.9293***	3.96	0.6845***	4.35

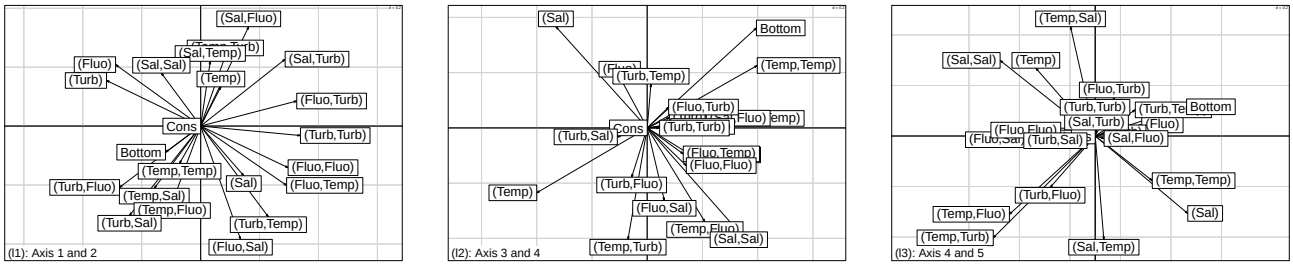
(C)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

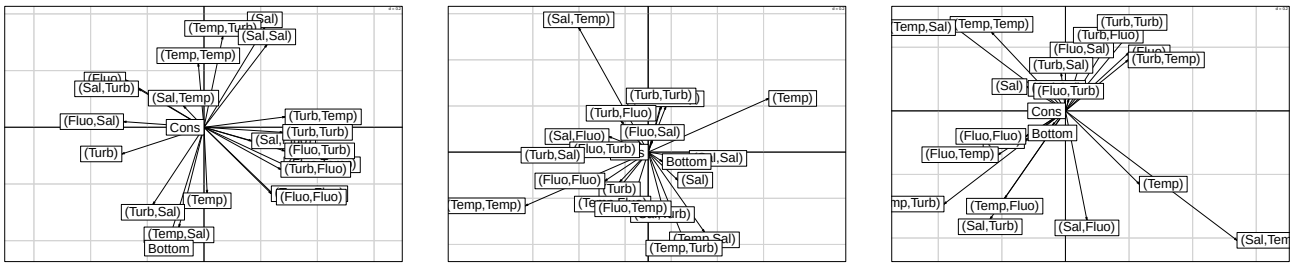
Figure 4.1: Spatial Principal Component Analysis ((l1): PC1 on x-axis and PC2 on y-axis,(l2): PC3 on x-axis and PC4 on y-axis, (l3): PC4 on x-axis and PC5 on y-axis) of signatures coefficients of the environmental parameters (sea temperature, fluorescence, turbidity, and salinity) at different frequencies; (a) 38, (b) 70, (c) 120, and (d) 200 kHz and bottom depth over Senegalese waters (A: southern continental shelf; B: southern high sea and C: northern continental shelf) during the AWA sea survey.

(A)

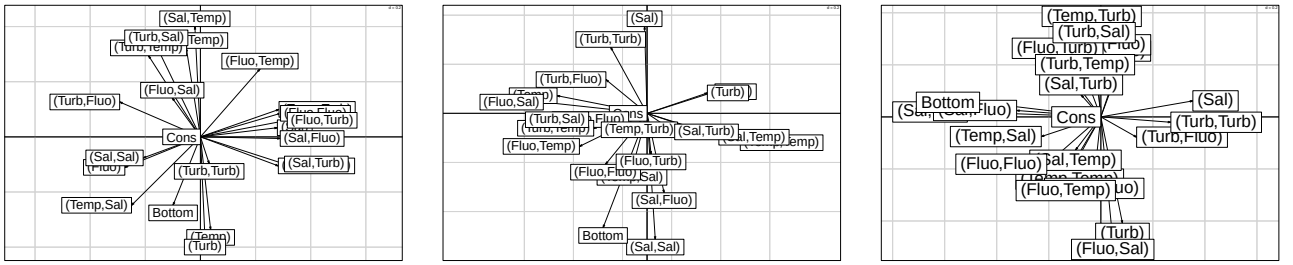
a)



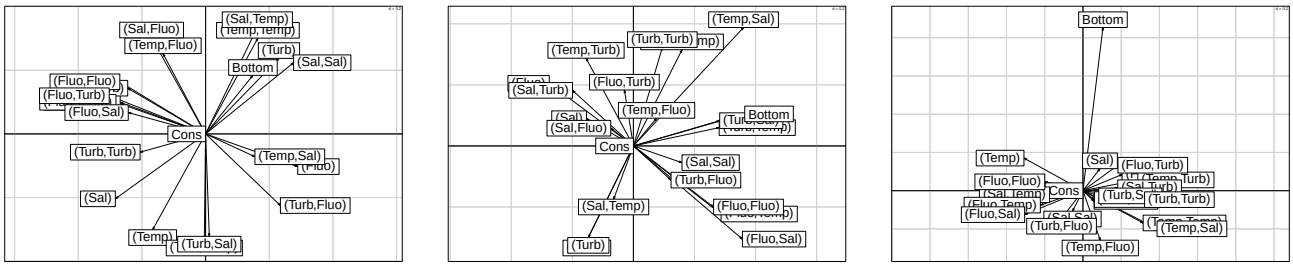
b)



c)

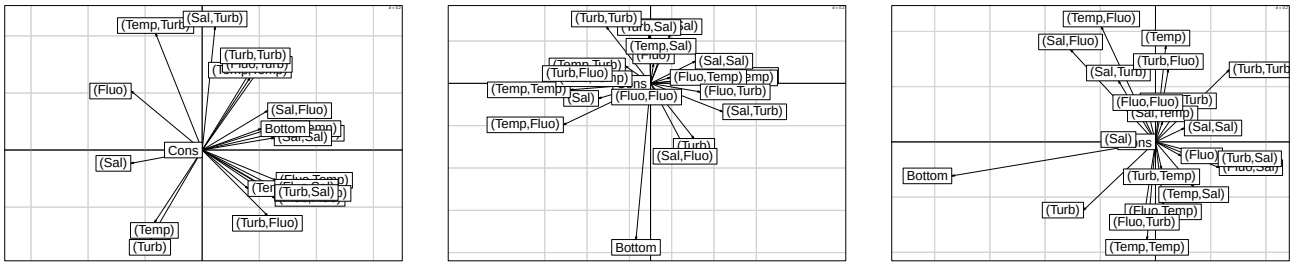


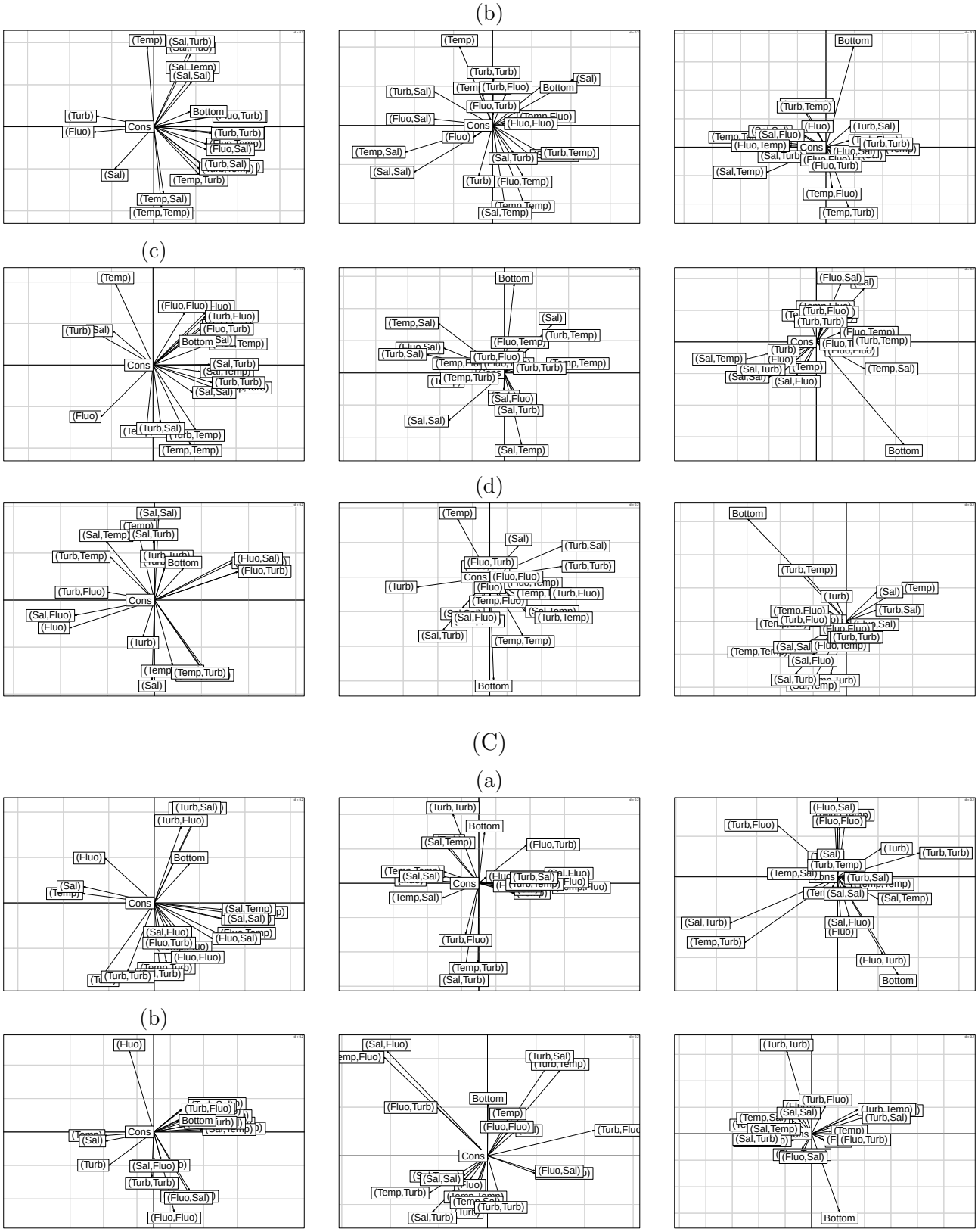
d)



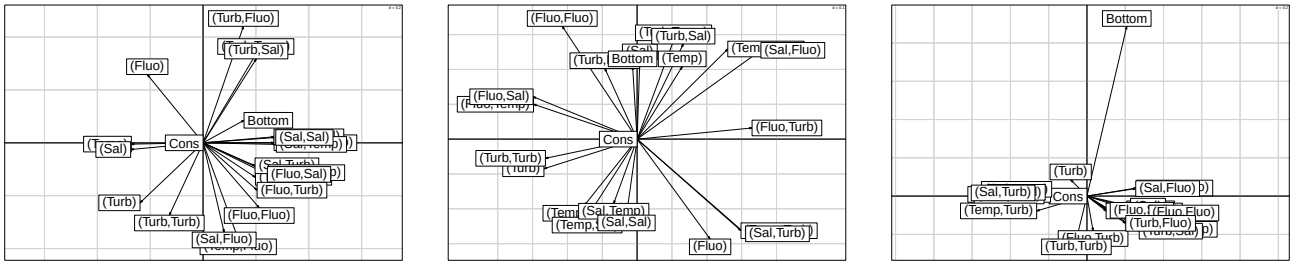
(B)

(a)




(b)
(Temp)
(Sal,Turb)
(Sal,Temp)
(Sal,Sal)
(Turb)
(Fluo)
Cons
Bottom
(Turb,Turb)
(Fluo,Sal)
(Turb,Sal)
(Sal)
(Temp,Turb)
(Temp,Sal)
(Temp,Temp)
(Turb,Turb)
(Turb,Sal)
(Fluo,Turb)
(Fluo,Fluo)
(Temp,Sal)
(Sal,Sal)
(Sal,Turb)
(Turb,Temp)
(Turb)
(Fluo,Temp)
(Sal,Temp)
Bottom
(Turb,Temp)
(Fluo)
(Turb,Sal)
(Fluo,Temp)
Cons
(Turb,Turb)
(Sal,Turb)
(Fluo,Turb)
(Sal,Temp)
(Temp,Fluo)
(Temp,Turb)
(c)
(Temp)
(Fluo,Fluo)
(Turb,Fluo)
(Turb,Sal)
(Fluo,Turb)
Bottom
(Temp)
Cons
(Sal,Turb)
(Sal,Temp)
(Turb,Turb)
(Sal,Sal)
(Fluo)
(Turb,Sal)
(Turb,Temp)
Bottom
(Temp,Sal)
(Sal)
(Turb,Sal)
(Fluo,Temp)
(Turb,Temp)
(Turb,Sal)
(Turb,Fluo)
(Temp,Temp)
(Temp,Turb)
(Sal,Fluo)
(Sal,Turb)
(Sal,Sal)
(Sal,Temp)
(Fluo,Sal)
(Sal)
(Turb,Turb)
(Turb)
(Fluo)
(Sal,Temp)
(Sal,Turb)
(Temp,Sal)
(Sal,Fluo)
Bottom
(d)
(Sal,Sal)
(Temp)
(Sal,Temp)
(Sal,Turb)
(Turb,Temp)
(Turb,Turb)
Bottom
(Fluo,Sal)
(Fluo,Turb)
(Turb,Fluo)
Cons
(Sal,Fluo)
(Fluo)
(Turb)
(Temp,Turb)
(Sal)
(Temp)
(Sal)
(Turb,Sal)
(Fluo,Turb)
(Turb,Sal)
Cons
(Fluo,Fluo)
(Turb)
(Fluo)
(Temp,Fluo)
(Turb,Fluo)
(Temp,Fluo)
(Sal,Fluo)
(Turb,Temp)
(Sal,Turb)
(Temp,Temp)
Bottom
Bottom
(Turb,Temp)
(Turb)
(Sal)
(Temp)
(Temp,Fluo)
(Turb,Fluo)
(Turb,Sal)
(Sal,Sal)
(Fluo,Temp)
(Turb,Turb)
(Sal,Temp)
(Sal,Fluo)
(Sal,Turb)
(Turb)
(C)
(a)
(Turb,Sal)
(Turb,Fluo)
(Fluo)
Bottom
(Sal)
(Temp)
Cons
(Sal,Temp)
(Sal,Sal)
(Sal,Fluo)
(Fluo,Sal)
(Fluo,Turb)
(Fluo,Fluo)
(Temp,Turb)
(Turb,Turb)
(Turb)
(Turb,Turb)
Bottom
(Sal,Temp)
(Fluo,Turb)
Cons
(Temp,Sal)
(Turb,Fluo)
(Temp,Turb)
(Sal,Turb)
(Fluo,Sal)
(Fluo,Fluo)
(Turb,Fluo)
(Sal)
(Turb)
(Turb,Turb)
(Turb,Temp)
(Temp,Sal)
(Sal,Turb)
(Sal,Temp)
(Sal,Fluo)
(Fluo)
(Temp,Turb)
(Fluo,Turb)
Bottom
(b)
(Fluo)
(Turb,Fluo)
Cons
(Sal)
(Turb)
(Sal,Fluo)
(Turb,Turb)
(Fluo,Sal)
(Fluo,Fluo)
(Sal,Fluo)
(Turb,Sal)
(Turb,Temp)
Bottom
(Fluo,Turb)
(Temp)
(Fluo,Fluo)
(Turb,Fluo
Cons
(Fluo,Sal)
(Temp,Turb)
(Turb,Turb)
(Sal,Turb)
(Turb,Turb)
(Turb,Fluo)
(Sal,Sal)
(Turb,Temp)
(Turb,Sal)
(Temp)
(Sal,Temp)
(Sal,Turb)
(Fluo,Turb)
(Fluo,Sal)
Bottom

(c)


(Turb,Fluo)
(Fluo)
(Turb,Sal)
Bottom
(Sal)
Cons
(Sal,Sal)
(Fluo,Sal)
(Fluo,Turb)
(Turb)
(Turb,Turb)
(Fluo,Fluo)
(Sal,Fluo)
(Fluo,Fluo)
(Turb,Sal)
(Turb,
Bottom
(Temp)
(Sal,Fluo)
(Fluo,Sal)
(Fluo,Turb)
Cons
(Turb,Turb)
(Sal,Temp)
(Temp,
(Sal,Sal)
(Fluo)
(Sal,Turb)
Bottom
(Turb)
(Sal,Turb)
(Sal,Fluo)
Cons
(Temp,Turb)
(Fluo,Fluo)
(Turb,Fluo)
(Turb,Turb)

(d)

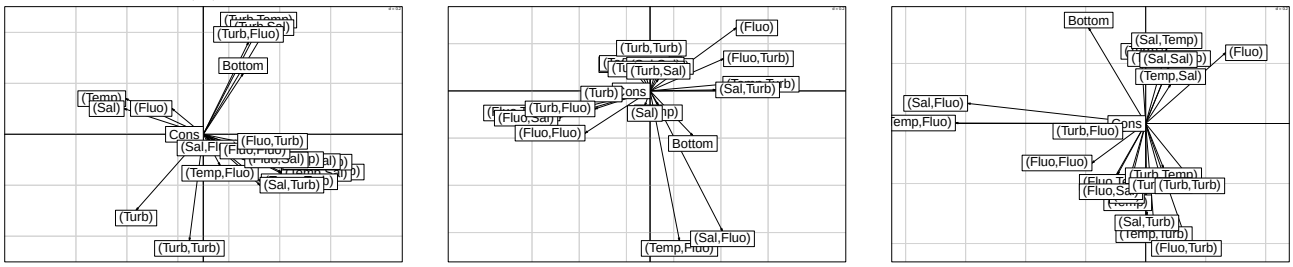
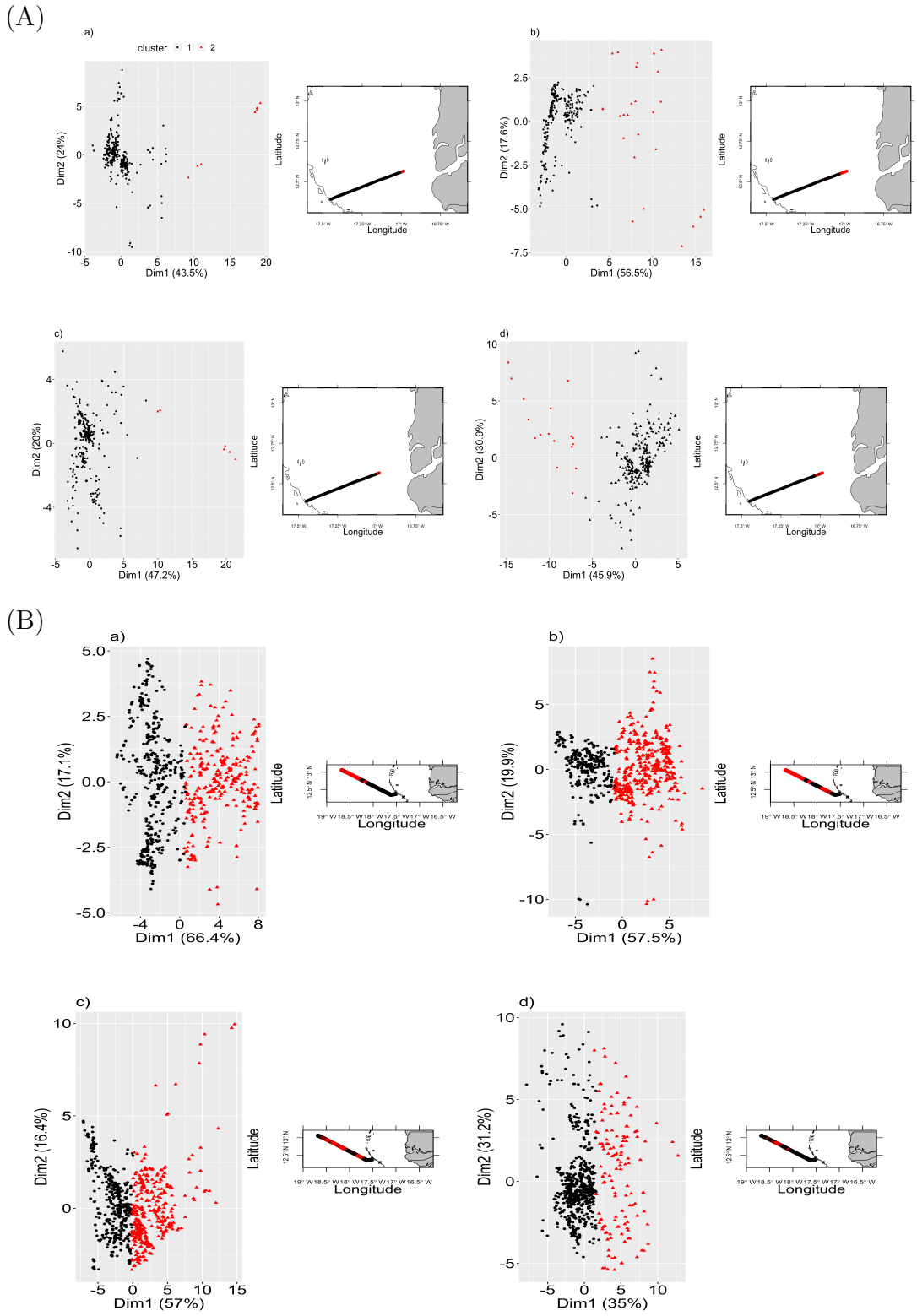

(Turb,Fluo)
Bottom
(Temp)
(Sal)
(Fluo)
Cons
(Fluo,Turb)
(Temp,Fluo)
(Sal,Turb)
(Turb)
(Turb,Turb)
(Fluo)
(Turb,Turb)
(Turb,Sal)
(Fluo,Turb)
(Turb)
Cons
(Sal,Turb)
(Turb,Fluo)
(Sal)
(Fluo,Fluo)
Bottom
(Sal,Fluo)
Bottom
(Sal,Temp)
(Fluo)
(Temp,Sal)
(Sal,Fluo)
(Turb,Fluo)
Cons
(Fluo,Fluo)
(Sal,Turb)
(Fluo,Turb)

Figure 4.2: Classification of the esus based on the spatial functional principal component (SPCA) scores (black: Cluster 1; red: Cluster 2) over Senegalese waters (A: south continental shelf; B: south high sea and C: north continental shelf of the AWA sea survey), observed at various frequencies: (A) 38, (B) 70, (C) 120 and (D) 200 kHz.



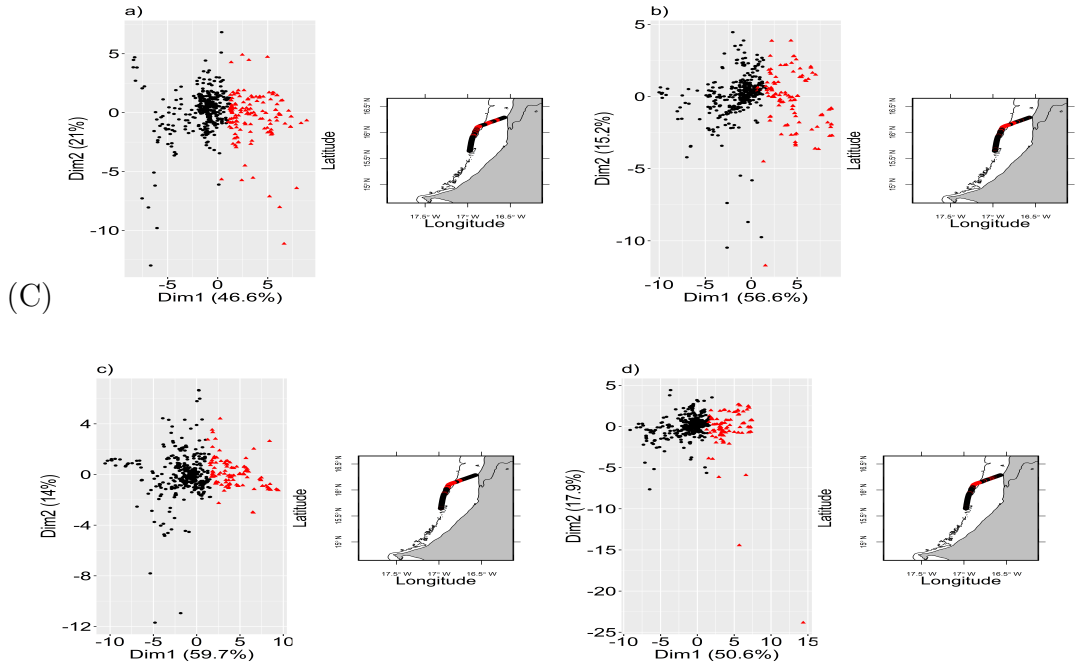
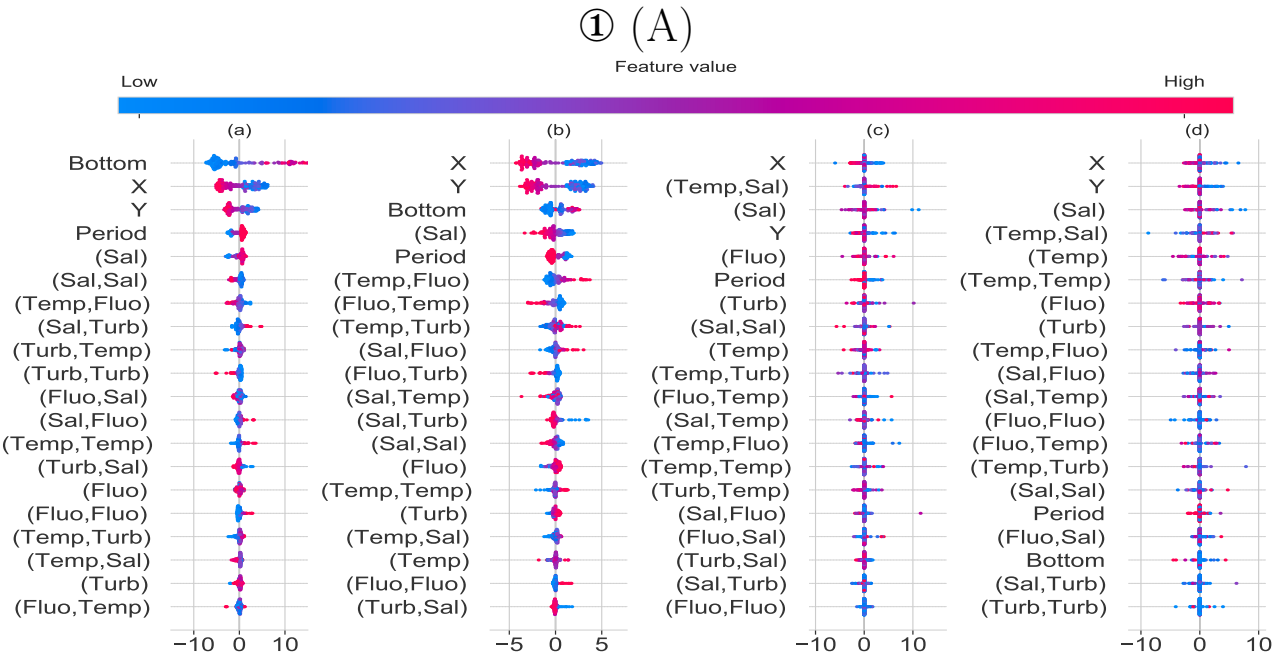
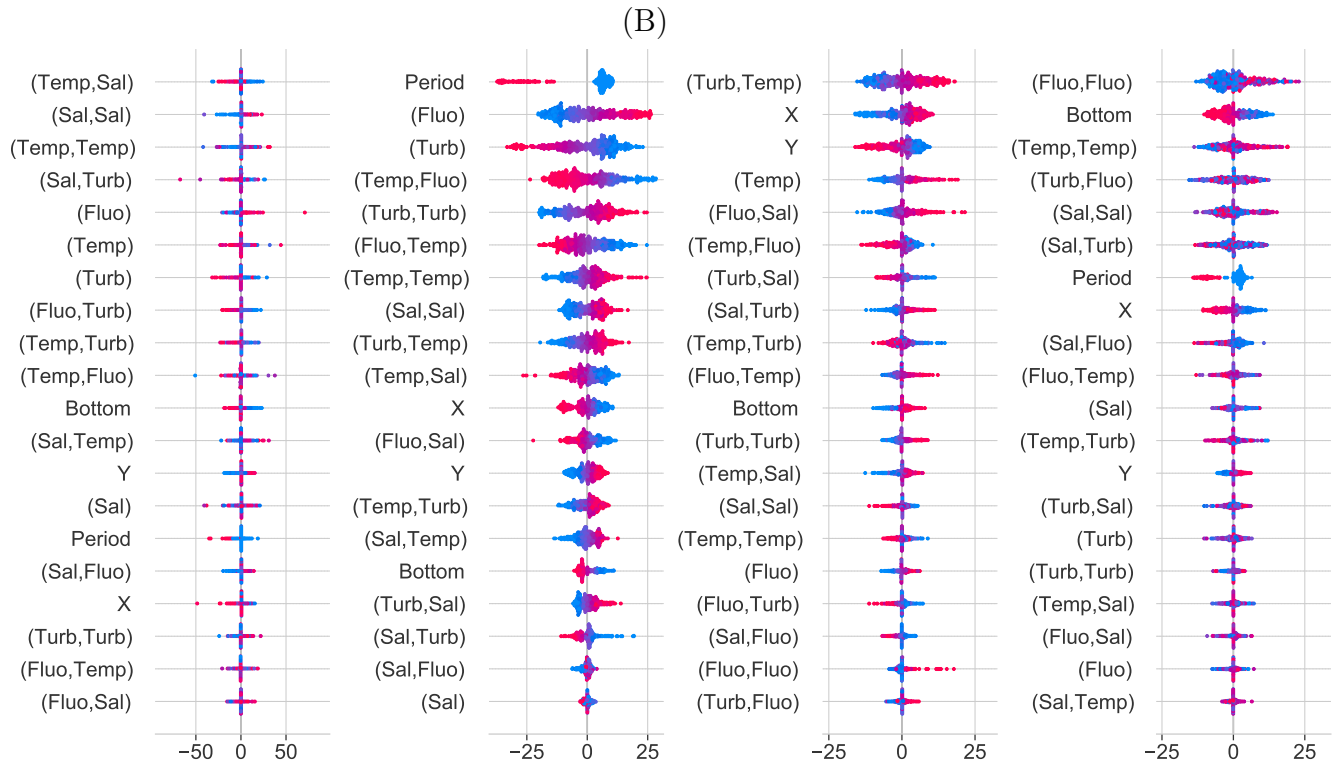
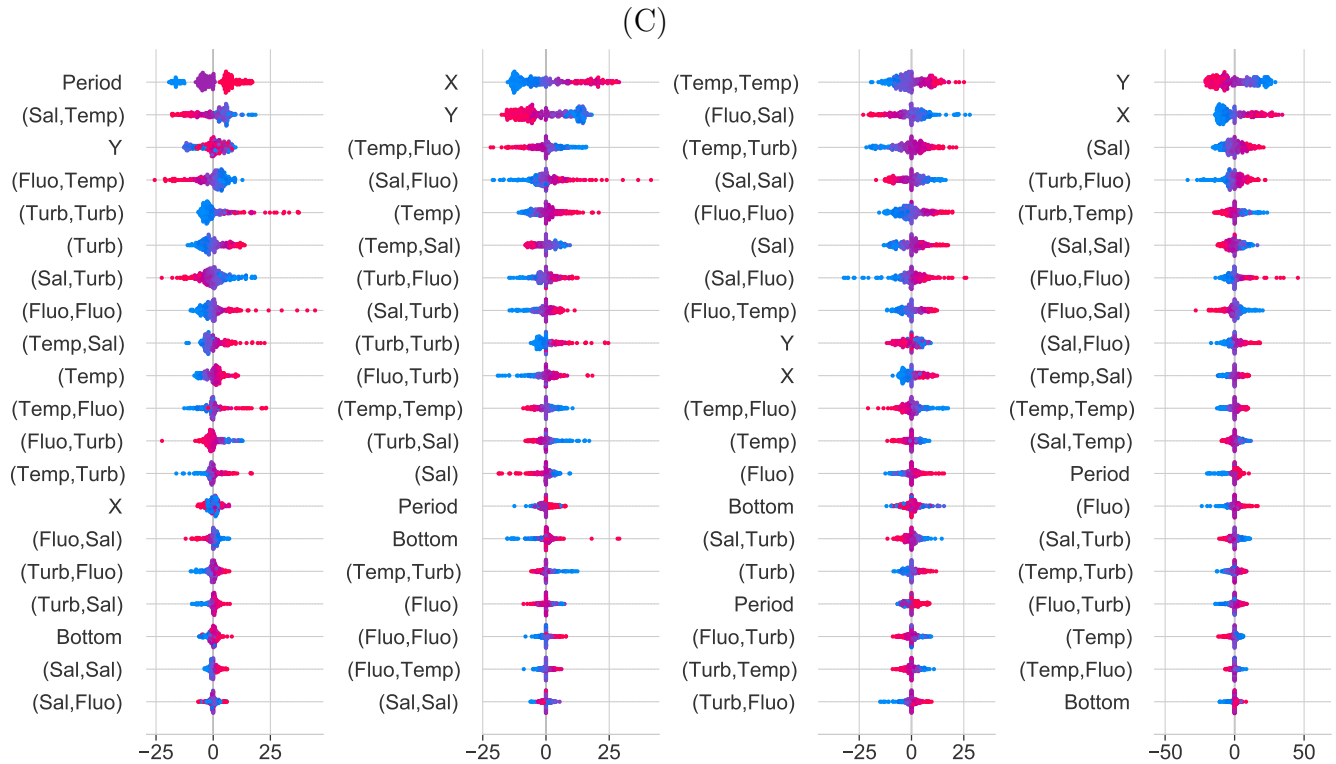
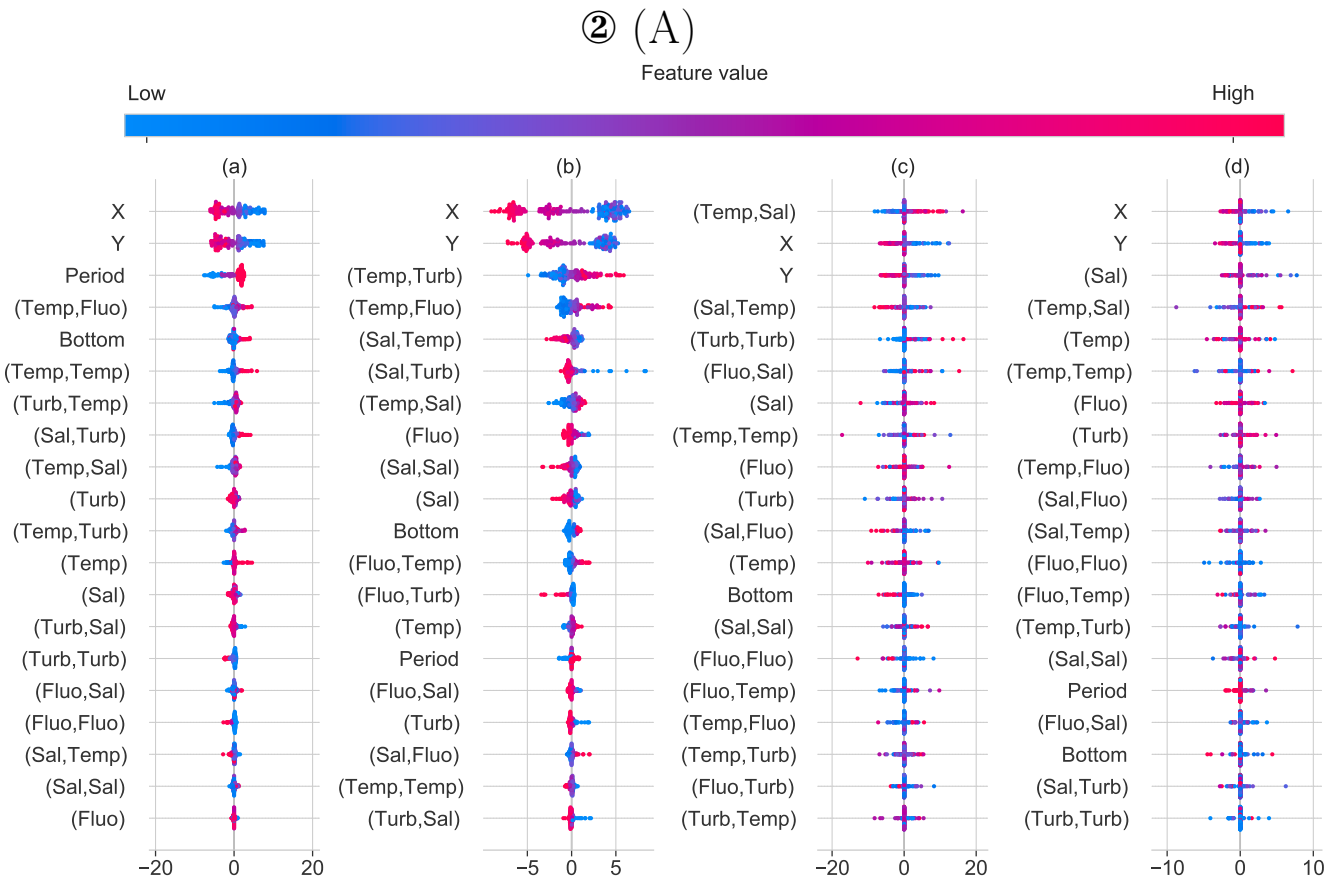


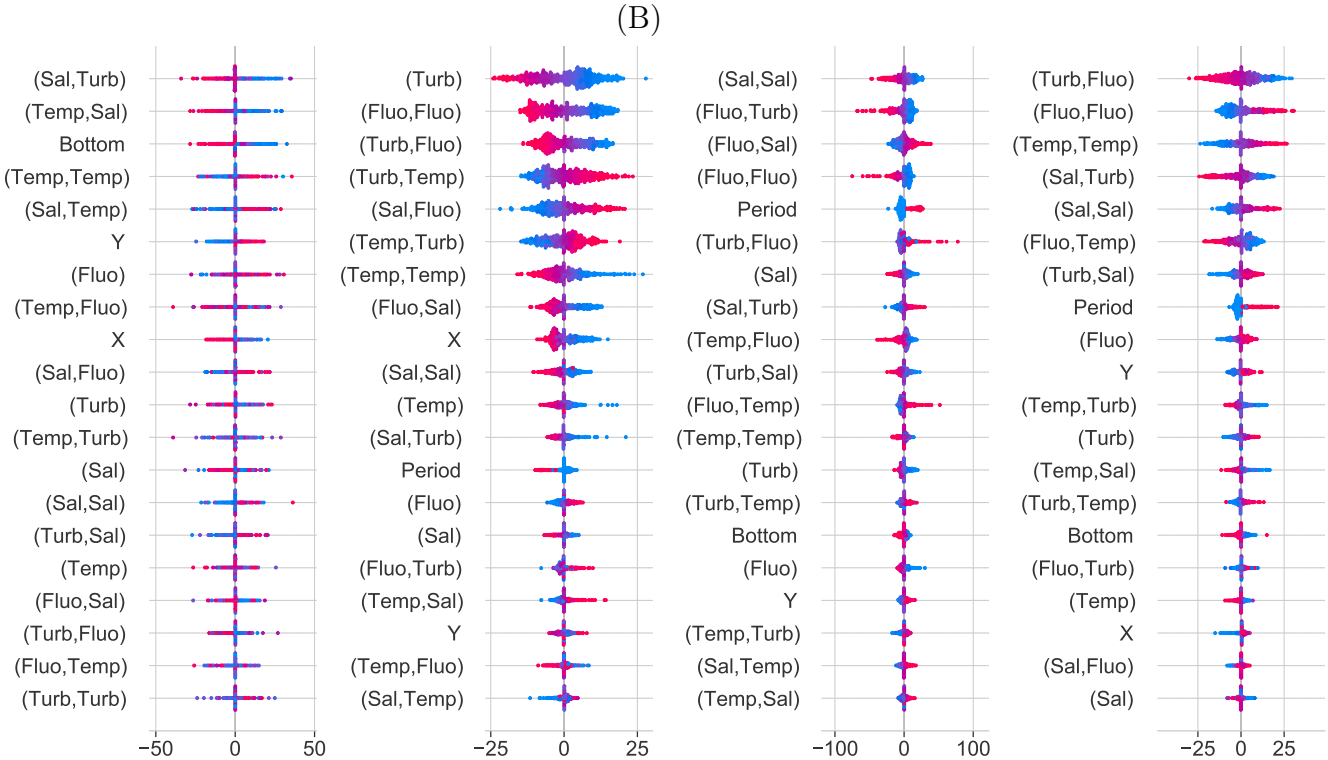
Figure 4.3: SHAP summary graph in which the contribution of each variable to the neural model ((1): Thickness and (2): Depth) is displayed taking into account all the values of each of the environmental parameters signatures (sea temperature, fluorescence, turbidity and salinity) and geographical coordinates (Longitude (X) and latitude (Y)). The color indicates the SHAP value (importance) of the variable (red is high, blue is low). The x-axis indicates that the samples are sorted by their SHAP values and the y-axis indicates that the variables are sorted by their mean absolute SHAP values over Senegalese waters (A: south continental shelf; B: south high sea and C: north continental shelf of the AWA sea survey), observed at various frequencies: (a) 38, (b) 70, (c) 120 and (d) 200 kHz.



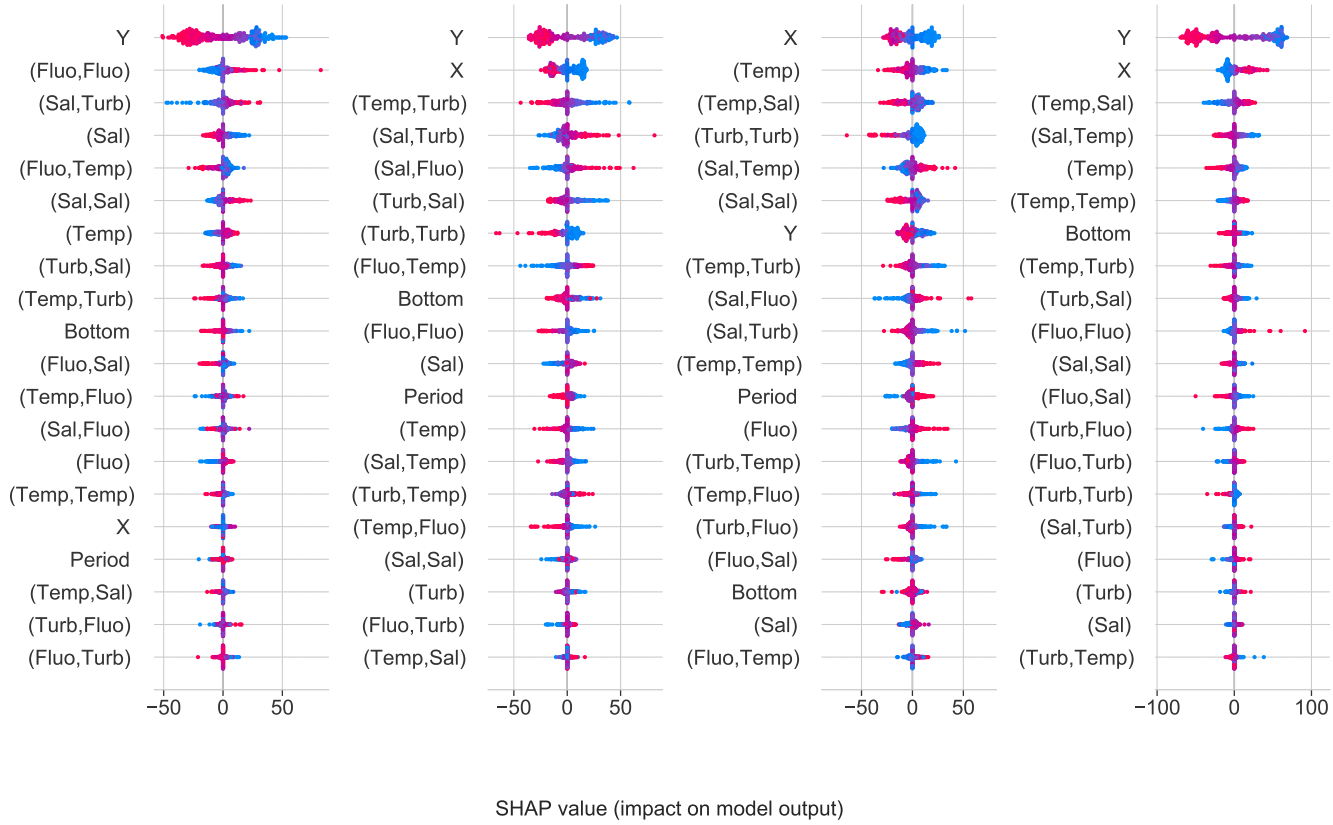

(B)
(Temp,Sal)
(Sal,Sal)
(Temp,Temp)
(Sal,Turb)
(Fluo)
(Temp)
(Turb)
(Fluo,Turb)
(Temp,Turb)
(Temp,Fluo)
Bottom
(Sal,Temp)
Y
(Sal)
Period
(Sal,Fluo)
X
(Turb,Turb)
(Fluo,Temp)
(Fluo,Sal)
−50 0 50
Period
(Fluo)
(Turb)
(Temp,Fluo)
(Turb,Turb)
(Fluo,Temp)
(Temp,Temp)
(Sal,Sal)
(Turb,Temp)
(Temp,Sal)
X
(Fluo,Sal)
Y
(Temp,Turb)
(Sal,Temp)
Bottom
(Turb,Sal)
(Sal,Turb)
(Sal,Fluo)
(Sal)
−25 0 25
(Turb,Temp)
X
Y
(Temp)
(Fluo,Sal)
(Temp,Fluo)
(Turb,Sal)
(Sal,Turb)
(Temp,Turb)
(Fluo,Temp)
Bottom
(Turb,Turb)
(Temp,Sal)
(Sal,Sal)
(Temp,Temp)
(Fluo)
(Fluo,Turb)
(Sal,Fluo)
(Fluo,Fluo)
(Turb,Fluo)
−25 0 25
(Fluo,Fluo)
Bottom
(Temp,Temp)
(Turb,Fluo)
(Sal,Sal)
(Sal,Turb)
Period
X
(Sal,Fluo)
(Fluo,Temp)
(Sal)
(Temp,Turb)
Y
(Turb,Sal)
(Turb)
(Turb,Turb)
(Temp,Sal)
(Fluo,Sal)
(Fluo)
(Sal,Temp)
−25 0 25


(C)
Period
(Sal,Temp)
Y
(Fluo,Temp)
(Turb,Turb)
(Turb)
(Sal,Turb)
(Fluo,Fluo)
(Temp,Sal)
(Temp)
(Temp,Fluo)
(Fluo,Turb)
(Temp,Turb)
X
(Fluo,Sal)
(Turb,Fluo)
(Turb,Sal)
Bottom
(Sal,Sal)
(Sal,Fluo)
−25 0 25
X
Y
(Temp,Fluo)
(Sal,Fluo)
(Temp)
(Temp,Sal)
(Turb,Fluo)
(Sal,Turb)
(Turb,Turb)
(Fluo,Turb)
(Temp,Temp)
(Turb,Sal)
(Sal)
Period
Bottom
(Temp,Turb)
(Fluo)
(Fluo,Fluo)
(Fluo,Temp)
(Sal,Sal)
−25 0 25
(Temp,Temp)
(Fluo,Sal)
(Temp,Turb)
(Sal,Sal)
(Fluo,Fluo)
(Sal)
(Sal,Fluo)
(Fluo,Temp)
Y
X
(Temp,Fluo)
(Temp)
(Fluo)
Bottom
(Sal,Turb)
(Turb)
Period
(Fluo,Turb)
(Turb,Temp)
(Turb,Fluo)
−25 0 25
Y
X
(Sal)
(Turb,Fluo)
(Turb,Temp)
(Sal,Sal)
(Fluo,Fluo)
(Fluo,Sal)
(Sal,Fluo)
(Temp,Sal)
(Temp,Temp)
(Sal,Temp)
Period
(Fluo)
(Sal,Turb)
(Temp,Turb)
(Fluo,Turb)
(Temp)
(Temp,Fluo)
Bottom
−50 0 50
SHAP value (impact on model output)


② (A)
Feature value
Low
High
(a)
X
Y
Period
(Temp,Fluo)
Bottom
(Temp,Temp)
(Turb,Temp)
(Sal,Turb)
(Temp,Sal)
(Turb)
(Temp,Turb)
(Temp)
(Sal)
(Turb,Sal)
(Turb,Turb)
(Fluo,Sal)
(Fluo,Fluo)
(Sal,Temp)
(Sal,Sal)
(Fluo)
−20 0 20
(b)
X
Y
(Temp,Turb)
(Temp,Fluo)
(Sal,Temp)
(Sal,Turb)
(Temp,Sal)
(Fluo)
(Sal,Sal)
(Sal)
Bottom
(Fluo,Temp)
(Fluo,Turb)
(Temp)
Period
(Fluo,Sal)
(Turb)
(Sal,Fluo)
(Temp,Temp)
(Turb,Sal)
−5 0 5
(c)
(Temp,Sal)
X
Y
(Sal,Temp)
(Turb,Turb)
(Fluo,Sal)
(Sal)
(Temp,Temp)
(Fluo)
(Turb)
(Sal,Fluo)
(Temp)
Bottom
(Sal,Sal)
(Fluo,Fluo)
(Fluo,Temp)
(Temp,Fluo)
(Temp,Turb)
(Fluo,Turb)
(Turb,Temp)
−20 0 20
(d)
X
Y
(Sal)
(Temp,Sal)
(Temp)
(Temp,Temp)
(Fluo)
(Turb)
(Temp,Fluo)
(Sal,Fluo)
(Sal,Temp)
(Fluo,Fluo)
(Fluo,Temp)
(Temp,Turb)
(Sal,Sal)
Period
(Fluo,Sal)
Bottom
(Sal,Turb)
(Turb,Turb)
−10 0 10


(B)
(Sal,Turb)
(Temp,Sal)
Bottom
(Temp,Temp)
(Sal,Temp)
Y
(Fluo)
(Temp,Fluo)
X
(Sal,Fluo)
(Turb)
(Temp,Turb)
(Sal)
(Sal,Sal)
(Turb,Sal)
(Temp)
(Fluo,Sal)
(Turb,Fluo)
(Fluo,Temp)
(Turb,Turb)
−50 0 50
(Turb)
(Fluo,Fluo)
(Turb,Fluo)
(Turb,Temp)
(Sal,Fluo)
(Temp,Turb)
(Temp,Temp)
(Fluo,Sal)
X
(Sal,Sal)
(Temp)
(Sal,Turb)
Period
(Fluo)
(Sal)
(Fluo,Turb)
(Temp,Sal)
Y
(Temp,Fluo)
(Sal,Temp)
−25 0 25
(Sal,Sal)
(Fluo,Turb)
(Fluo,Sal)
(Fluo,Fluo)
Period
(Turb,Fluo)
(Sal)
(Sal,Turb)
(Temp,Fluo)
(Turb,Sal)
(Fluo,Temp)
(Temp,Temp)
(Turb)
(Turb,Temp)
Bottom
(Fluo)
Y
(Temp,Turb)
(Sal,Temp)
(Temp,Sal)
−100 0 100
(Turb,Fluo)
(Fluo,Fluo)
(Temp,Temp)
(Sal,Turb)
(Sal,Sal)
(Fluo,Temp)
(Turb,Sal)
Period
(Fluo)
Y
(Temp,Turb)
(Turb)
(Temp,Sal)
(Turb,Temp)
Bottom
(Fluo,Turb)
(Temp)
X
(Sal,Fluo)
(Sal)
−25 0 25

(C)


Y
(Fluo,Fluo)
(Sal,Turb)
(Sal)
(Fluo,Temp)
(Sal,Sal)
(Temp)
(Turb,Sal)
(Temp,Turb)
Bottom
(Fluo,Sal)
(Temp,Fluo)
(Sal,Fluo)
(Fluo)
(Temp,Temp)
X
Period
(Temp,Sal)
(Turb,Fluo)
(Fluo,Turb)
−50 0 50
Y
X
(Temp,Turb)
(Sal,Turb)
(Sal,Fluo)
(Turb,Sal)
(Turb,Turb)
(Fluo,Temp)
Bottom
(Fluo,Fluo)
(Sal)
Period
(Temp)
(Sal,Temp)
(Turb,Temp)
(Temp,Fluo)
(Sal,Sal)
(Turb)
(Fluo,Turb)
(Temp,Sal)
−50 0 50
X
(Temp)
(Temp,Sal)
(Turb,Turb)
(Sal,Temp)
(Sal,Sal)
Y
(Temp,Turb)
(Sal,Fluo)
(Sal,Turb)
(Temp,Temp)
Period
(Fluo)
(Turb,Temp)
(Temp,Fluo)
(Turb,Fluo)
(Fluo,Sal)
Bottom
(Sal)
(Fluo,Temp)
−50 0 50
Y
X
(Temp,Sal)
(Sal,Temp)
(Temp)
(Temp,Temp)
Bottom
(Temp,Turb)
(Turb,Sal)
(Fluo,Fluo)
(Sal,Sal)
(Fluo,Sal)
(Turb,Fluo)
(Fluo,Turb)
(Turb,Turb)
(Sal,Turb)
(Fluo)
(Turb)
(Sal)
(Turb,Temp)
−100 0 100
SHAP value (impact on model output)

Chapter 5

Principal Component Analysis of Multivariate geo-spatial Functional Data

5.1 Abstract

This chapter is dedicated to dimension reduction techniques for multivariate geo-spatial functional data. We introduce an innovative method named Spatial Multivariate Funtional Principal Component Analysis (SMFPCA), which stands for principal component analysis for multivariate spatial data. Unlike the conventional Multivariate Karhunen-Loève approach, SMFPCA excels at effectively capturing spatial dependencies among multiples functions. SMFPCA conducts spectral functional component analysis on multivariate spatial data, encompassing data points located within a regular grid. The methodological framework and algorithm for SMFPCA have been developed to address the challenges posed by the lack of suitable methods for handling such data. The efficiency of the proposed methodology has been substantiated through comprehensive assessments of its performance using and simulated datasets and sea-surface temperature of interest in fisheries analysis, providing valuable insights into the properties of multivariate spatial functional data within a finite sample.

5.2 Introduction

Analysis of complex data structures, such as multivariate and spatially indexed functional data, has become more prevalent in recent years. Spatial functional data constitute a data type in which each observation located a spatial location is represented by a function rather than a conventional vector of values. These data are eventually of infinite dimension, as it is defined across the entire continuum of points.

Principal component analysis (PCA) is a technique used for exploratory data analysis. It is commonly employed to reduce the multidimensional space, where the information is not easily interpretable, into a space of reduced dimension. Furthermore, PCA allows us to derive more informative attributes from the data, identify hidden patterns in a dataset, and discover correlations among variables.

Within the realm of spatially indexed functional data analysis, the data are measured over space. This could include, for example, temperature measurements taken at different locations in a geographic region or hyperspectral image analysis. One of the most common uses of principal component analysis on functional spatial data (SFPCA) is in studies where the functional spatial data consists of spatially distributed objects, such as environmental measurement sampling sites, watersheds, administrative districts, etc. The value of spatial analysis is to understand and explore the interrelationship between the spatial positioning of objects, phenomena, and their characteristics (Feuillet et al., 2018). The analysis of spatially indexed functional data serves two purposes. Firstly, it enables the identification of spatial patterns inherent in the data. These patterns provide valuable insights into the underlying spatial structure and dynamics of the phenomena being studied. Secondly, the analysis facilitates the development of models that can be utilized for making predictions or drawing conclusions about the spatial distribution of the data. In the literature, Delicado et al. (2010) provided a useful introduction to selected methods for geostatistical functional data; Aristizabal et al. (2019) developed extending ANOVA techniques to analyze spatially correlated functional data. Gromenko et al. (2012) and Gromenko et al. (2017) used a spatially indexed functional data framework to solve space physics problems. Aston et al. (2017) and Constantinou et al. (2017) developed spatial-temporal separability tests for functional data. Liu et al. (2017b) considered tests of anisotropy using Karhunen-Loève expansion but with a semiparametric estimation procedure custom-developed for their objectives. The work of Bernardi et al. (2017) is an example of a substantive application of spatial functional modeling to a problem of practical importance. Kuenzer et al. (2020) proposed a dimension reduction technique suitable for functional data indexed by spatial locations on a regular grid. They developed the mathematical foundations for the spectral analysis of such data, including spectral theory for linear spatial filters. In our work, employing the spectral analysis approach, our focus has been directed towards situations where multiple variables can be investigated, all of which are defined within the same spatial-temporal domain. For example, to study the co-variability of temperature within a particular region over a specific time period, we have positioned ourselves within this framework to seek a viable solution. The objective of this study is to introduce a novel approach that enables spectral functional PCA on multivariate spatial data, defined within both the same and different domains, all structured on a regular grid. We refer to this approach as SMFPCA. Its aim is to address limitations identified in previous works related to MFPCA Happ (2017) and SFPCA Kuenzer et al. (2020). To

our knowledge, no previous work has employed multivariate spectral FPCA on spatial and temporal variables within the same and different domain. We apply spectral spatial univariate FPCA, and the results can be used to estimate multivariate functional principal components, eigenvalues, and scores based on their univariate equivalents. Spectral decomposition is a powerful method for analyzing functional spatial data. It is based on the spectral density operator, which describes the spatial dependence structure of the data by separating it into distinct frequencies. The advantage of this approach is that it allows the identification of dominant frequencies that drive the variation in the spatial dependence structure of the data. These dominant frequencies can then be used to define the functional principal components (FPCs), providing insights into the underlying stochastic processes contributing to the data's variation. Thus, spectral decomposition is an excellent tool for a better understanding of the complex spatial dependence structure. We conducted a test of our method on multivariate and spatially indexed environmental data. SMFPCA facilitates the integration of this type of data within a unified analytical framework. This approach effectively captures and elucidates the variations between variables as well as the inherent spatial variations present in the data. The SMFPCA technique also possesses the capability to extract latent information from functional multivariate spatial data, uncovering previously unobserved latent dimensions. This, in turn, explains variations in the data and reveals spatial-temporal patterns. As a result, SMFPCA provides valuable insights into the underlying mechanisms driving the observed data.

This part is structured as follows: In Section 3, we describe the SMFPCA methodology; in Section 4, we present Numerical experiments. Section 5 give the conclusion and some perspectives.

5.3 Spectral principal component analysis of multivariate spatial functional data

This section describes the proposed methodology and the mathematical model for conducting a multivariate spectral PCA on spatial-functional data. The methodology is based on the principles of Karhunen-Loève theory, incorporating spectral analysis and Functional Principal Component Analysis (FPCA). Additionally, key properties and concepts are derived from the works of Kuenzer et al. (2020), Ramsay and Silverman (2005c), Ramsay and Silverman (2002b), and Happ (2017). The approach specifically targets a multivariate spatial-functional phenomenon observed on a regular grid and defined within both the same and different domains.

In section 5.3.1, we delve into the mathematical model employed in our approach. We provide definitions for various components, including the spatial multivariate functional data model, covariance operator, spectral density operator, eigenfunctions (representing functions that characterize significant variations or dominant modes within the functional data), eigenvalues (corresponding to the weights assigned to each eigenfunction and quantifying the total variance explained), and scores (representing the projections of the original data onto the basis functions). Once these elements are clearly established, we proceed to outline the subsequent steps of SMFPCA in section 5.3.3.

5.3.1 Multivariate Spatial Functional Data

We consider that at n spatial units located on a region $\mathbf{D} \subset \mathbb{Z}^N$, $N > 1$, representing a rectangular grid, we observe a multivariate spatial functional process $\{\mathbf{X_s}(.) = (X_{\mathbf{s}}^{(1)}(.), ..., X_{\mathbf{s}}^{(p)}(.))^\top\}$, $p \geq 1$, where $\mathbf{s} \in \mathbf{D}$, $X_{\mathbf{s}}^{(j)} = \{X_{\mathbf{s}}^{(j)}(t_j), t_j \in \mathcal{T}_j\}$. For $1 \leq j \leq p$, let $\mathcal{T}_j$ be a compact set in $\mathbb{R}$, with finite (Lebesgue-) measure and such that $X_{\mathbf{s}}^{(j)} : \mathcal{T}_j \longrightarrow \mathbb{C}$ is assumed to belong to $\mathcal{L}^2(\mathcal{T}_j, \mathbb{C})$, the space of complex square-integrable functions on $\mathcal{T}_j$. In the following let $\mathcal{L}^2(\mathcal{T}_j, \mathbb{C}) = \mathcal{L}^2(\mathcal{T}_j)$. Note that the special case $p = 1$ corresponds to the univariate spatial-functional case Kuenzer et al. (2020).
We denote by $\mathcal{T} := \mathcal{T}_1 \times \cdots \times \mathcal{T}_p$, the p-Fold Cartesian product of the $\mathcal{T}_j$. So, $\mathbf{X_s}$ is a multivariate functional random variable indexed by $\mathbf{t} = (t_1, \cdots, t_p) \in \mathcal{T}$ and taking values in the p-Fold Cartesian product space $\mathcal{H} := \mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_p)$. Let the inner product $\langle\langle \cdot, \cdot \rangle\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$, for $f, g \in \mathcal{H}$:

$$\langle\langle f, g \rangle\rangle := \sum_{j=1}^{p} \langle f_j, g_j \rangle = \sum_{j=1}^{p} \int_{\mathcal{T}_j} f_j(t_j) g_j(t_j) dt_j$$

Then, $\mathcal{H}$ is a Hilbert space with respect to the scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ (Happ, 2017).
For each component $X_{\mathbf{s}}^{(j)}$, let's define the functions observed on n sites $\mathbf{s}_1, ..., \mathbf{s}_n \in \mathbf{D}$ by:

$$\mathbf{X}^{(j)} = \left(X_{\mathbf{s}_1}^{(j)}, ..., X_{\mathbf{s}_n}^{(j)} \right), \quad j = 1, \ldots, p$$

5.3.2 Univariate spatial functional PCA

We independently consider each of the spatial functional univariate sample $X^{(j)}$, to compute a univariate SFPCA. To achieve this, we apply the univariate spatial FPCA Kuenzer et al. (2020). Let $X^{(j)} \in \mathcal{L}^2(\mathcal{T}_j)$ possesses a covariance operator $C_j := \mathbb{E}[(X^{(j)} - \mu^j) \otimes (X^{(j)} - \mu^j)]$ (where μ^j is the mean curve define by $\mu^j(t) = \mathbb{E}X^{(j)}(t)$ with $t \in \mathcal{T}_j$) with kernel $c_j(t, s) = cov(X^{(j)}(t), X^{(j)}(s))$ $(t, s \in \mathcal{T}_j)$. Then, the integral operator C_j is defined by

$$(C_j f)(t) = \int_{\mathcal{T}_j} c_j(s, t) f(s) ds, \quad f \in \mathcal{L}^2(\mathcal{T}_j),\ t \in \mathcal{T}_j.$$

Let's suppose that each $\{X_{\mathbf{s}}^{(j)}\}$ is a weakly stationary functional process. We have:
(i) $\mathbb{E}(X_{\mathbf{s}}^{(j)}(t)) = \mathbb{E}(X_{\mathbf{0}}^{(j)}(t)) = \mu^j(t)$, $t \in \mathcal{T}_j$ with $\mathbf{0}$ the null vector in $\mathbb{R}^N$
(ii) for all $\mathbf{s}, \mathbf{h} \in \mathbf{D}$, and $t, s \in \mathcal{T}_j$; $c_{j,\mathbf{h}}(t, s) := Cov\left(X_{\mathbf{h}}^j(t), X_{\mathbf{0}}^j(s) \right) = Cov\left(X_{\mathbf{s}+\mathbf{h}}^j(t), X_{\mathbf{s}}^j(s) \right)$

The integral operator defined by the autocovariance kernel $c_{j,\mathbf{h}}$ is denoted $C_{j,\mathbf{h}}$ and defined by

$$(C_{j,\mathbf{h}} f)(t) = \int_{\mathcal{T}_j} c_{j,\mathbf{h}}(s, t) f(s) ds, \quad f \in \mathcal{L}^2(\mathcal{T}_j),\ t \in \mathcal{T}_j.$$

The following assumptions have been made about the process $X_{\mathbf{s}}^{(j)}$ i.e. it is weakly stationary and has

mean zero. We suppose the autocovariance operators are absolutely summable:

$$\sum_{\mathbf{h}\in D}\|C_{j,\mathbf{h}}\| < \infty. \tag{5.1}$$

Let us denote the spectral density operator of $X_{\mathbf{s}}^{(j)}$ by $\mathcal{F}_{\theta}^{X^{(j)}}$ with the following kernel:

$$f_{\theta}^{X^{(j)}}(t,s) := \frac{1}{(2\pi)^N}\sum_{\mathbf{h}\in\mathbb{Z}^N} c_{j,\mathbf{h}}(t,s)\exp(-i\mathbf{h}^{\top}\theta) \tag{5.2}$$

$$t,s\in\mathcal{T}_j, \quad \theta\in[-\pi,\pi]^N, \quad i=\sqrt{-1},$$

where θ is the spatial frequency. We define $\mathcal{L}_U^2([-\pi,\pi]^N)$ as the space of measurable mappings $x : [-\pi,\pi]^N \to U$ satisfying $\int_{[-\pi,\pi]^N}\|x(\theta)\|^2 d\theta < \infty$, with U the Hilbert space of all Hilbert–Schmidt operators from $\mathcal{L}^2(\mathcal{T}_j)$ to $\mathcal{L}^2(\mathcal{T}_j)$ (see Kuenzer et al. (2020) for further explanation). The operator $\mathcal{F}_{\theta}^{X^{(j)}}$ is understood as element of the space $\mathcal{L}_U^2([-\pi,\pi]^N)$ and is defined by

$$\left(\mathcal{F}_{\theta}^{X^{(j)}}G_{\theta}\right)(t) = \int_{\mathcal{T}_j} f_{\theta}^{X^{(j)}}(s,t)G_{\theta}(s)ds,$$

$$\text{with} \quad G_{\theta}\in\mathcal{L}_U^2([-\pi,\pi]^N) \quad \text{and} \quad t\in\mathcal{T}_j.$$

Considering the condition (5.1) and the assumptions previously defined for the process weakly stationary $X_{\mathbf{s}}^{(j)}$, $\mathcal{F}_{\theta}^{X^{(j)}}$ is a Hilbert-Schmidt operator (positive, self-adjoint) and admits a decomposition Kuenzer et al. (2020):

$$\mathcal{F}_{\theta}^{X^{(j)}} = \sum_{m\geq 1}\lambda_{j,m}(\theta)\varphi_{j,m}(\theta)\otimes\varphi_{j,m}(\theta), \tag{5.3}$$

where $\lambda_{j,m}(\theta) \geq \lambda_{j,m}(\theta) \geq \ldots \geq 0$ are eigenvalues (continuous functions of θ), and $\varphi_{j,m}(\theta)$ are associated eigenfunctions. Let $\varphi_{j,m}(t|\theta)$ be the value of the eigenfunction $\varphi_{j,m}(\theta)$ at $t\in\mathcal{T}_j$. The Fourier coefficients are

$$\phi_{m,\mathbf{l}}^{(j)}(t) := \frac{1}{(2\pi)^N}\int_{[-\pi,\pi]^N}\varphi_{j,m}(t|\theta)\exp(-i\mathbf{l}^{\top}\theta)d\theta, \tag{5.4}$$

$t\in\mathcal{T}_j$ and the corresponding expansion of $\varphi_{j,m}(t|\theta)$ is

$$\varphi_{j,m}(t|\theta) = \sum_{\mathbf{l}\in\mathbb{Z}^N}\phi_{m,\mathbf{l}}^{(j)}(t)\exp(-i\mathbf{l}^{\top}\theta). \tag{5.5}$$

With the property (5.1) and the assumptions previously defined for the process weakly stationary $\{X_{\mathbf{s}}^{(j)}\}$, we define the mth spatial functional principal component (SFPC) score by:

$$\xi_{m,\mathbf{s}}^{(j)} := \sum_{\mathbf{l}\in\mathbf{D}}\langle X_{\mathbf{s}-\mathbf{l}}^{(j)}, \phi_{m,\mathbf{l}}^{(j)}\rangle \tag{5.6}$$

where $\phi_{m,\mathbf{l}}^{(j)}$ is defined by (5.4). The corresponding SFPC filter are $(\phi_{m,\mathbf{l}}^{(j)})_{\mathbf{l}\in\mathbb{Z}^N}$.
We deduce that

- $\xi_{m,\mathbf{s}}^{(j)}$ converges in mean square with :

$$\mathbb{E}[\xi_{m,\mathbf{s}}^{(j)}] = 0, \quad \mathbb{E}[(\xi_{m,\mathbf{s}}^{(j)})^2] = \sum_{\mathbf{l}\in\mathbb{Z}^N}\sum_{\mathbf{k}\in\mathbb{Z}^N} \langle C_{\mathbf{l}-\mathbf{k}}^{X^{(j)}} \phi_{m,\mathbf{l}}^{(j)}, \phi_{m,\mathbf{k}}^{(j)} \rangle$$

  .

- If $X_{\mathbf{s}}^{(j)}$ is real then $\phi_{m,\mathbf{l}}^{(j)}$ and $\xi_{m,\mathbf{s}}^{(j)}$ are also real.
- if $C_{\mathbf{h}}^{X^{(j)}} = 0$ then $\forall\ \mathbf{h} = \mathbf{0}$, $\xi_{m,\mathbf{s}}^{(j)}$ coincides with the scores of the FPCA.
- $\forall\ m \neq m'$ and $\mathbf{s} \neq \mathbf{s}' \in \mathbf{D}$ the SFPCA scores $\xi_{m,\mathbf{s}}^{(j)}$ and $\xi_{m',\mathbf{s}'}^{(j)}$ are uncorrelated.

The spatial Karhunen–Loève expansion of $X_{\mathbf{s}}^{(j)}$ is given by

$$X_{\mathbf{s}}^{(j)}(t) = \sum_{m=1}^{\infty} X_{m,\mathbf{s}}^{(j)}(t) \quad t \in \mathcal{T}_j, \text{with} \tag{5.7}$$

$$X_{m,\mathbf{s}}^{(j)}(t) := \sum_{\mathbf{l}\in\mathbb{Z}^N} \xi_{m,\mathbf{s}+\mathbf{l}}^{(j)} \phi_{m,\mathbf{l}}^{(j)}(t).$$

In practice, the spectral density operator is unknown and has to be estimated using the sample $\mathbf{X}^{(j)}$ observed on the grid $\mathbf{D} = \{\mathbf{s} = (s_1, ..., s_N),\ 1 \leq s_i \leq n_i, i = 1, ..., N\}$, the sample size is then $n = \sum_{i=1}^{N} n_i$, and we use the notation $\mathbf{n} = (n_1, n_2, \ldots, n_N)$.
The spectral density operator is estimated by:

$$\widehat{\mathcal{F}}_{\boldsymbol{\theta}}^{X^{(j)}} := \frac{1}{(2\pi)^N} \sum_{|\mathbf{h}|\leq\mathbf{q}} w(\mathbf{h}/\mathbf{q}) \widehat{C}_{j,\mathbf{h}} e^{-i\mathbf{h}^\top\boldsymbol{\theta}} \tag{5.8}$$

where w represents a weight function and the vector $\mathbf{q} = (q_1, q_2, ..., q_N)$ consists of positive coordinates, the sample autocovariance operators are estimated as follows:

$$\widehat{C}_{j,\mathbf{h}} := \frac{1}{n} \sum_{\mathbf{s}\in M_{\mathbf{h},\mathbf{n}}} \left(X_{\mathbf{s}+\mathbf{h}}^{(j)} - \bar{X}^{(j)}\right) \otimes \left(X_{\mathbf{s}}^{(j)} - \bar{X}^{(j)}\right) \tag{5.9}$$

with $M_{\mathbf{h},\mathbf{n}} = \left\{\mathbf{s} \in \mathbb{Z}^N : 1 \leq s_i, s_i + h_i \leq n_i\, \forall 1 \leq i \leq N\right\}$. If the set $M_{\mathbf{h},\mathbf{n}}$ is empty, we set $\widehat{C}_{j,\mathbf{h}} = 0$.

5.3.3 Spatial Multivariate Funtional Principal Component Analysis Methodology

In this subsection, we present the methodology for computing SMFPCA. The methodology is divided into two parts.

The first part relies on the univariate SFPCA of Kuenzer et al. (2020). This involves considering a spectral analysis on $\mathbf{X}^{(j)}$, following the subsequent steps:

1. Compute the spectral covariance operator.

2. Decompose the spectral covariance operator to obtain the estimated eigenfunctions $\hat{\varphi}_{j,m}(\theta)$, involving SFPC filters $(\hat{\phi}^{(j)}_{m,\mathbf{s}})_{\mathbf{s}\in\mathbf{D}}$, and $\hat{\lambda}_{j,m}(\theta)$, the estimated eigenvalues associated with the spectral variability, where $m = 1, \ldots, M_j$ for suitably chosen truncation lags M_j.
   The estimator of the filter function $\phi^{(j)}_{m,\mathbf{s}}$ is given by

$$\hat{\phi}^{(j)}_{m,\mathbf{s}}(t) := \frac{1}{(2\pi)^N} \int_{[-\pi,\pi]^N} \hat{\varphi}_{j,m}(t|\theta) \exp(-i\mathbf{s}^\top\theta) d\theta,$$

   where the functions $\hat{\varphi}_{j,m}(t|\theta)$ are the eigenfunctions of the spectral density operator estimator $\widehat{\mathcal{F}}^{X^{(j)}}_{\boldsymbol{\theta}}$.

3. Finally, the $\mathbf{X}^{(j)}$ are projected onto the spectral eigenfunctions, yielding the estimated scores $\hat{\xi}^{(j)}_{m,\mathbf{s}}$, defined by:

$$\hat{\xi}^{(j)}_{m,\mathbf{s}} := \sum_{\|\mathbf{l}\|_\infty \le L} \langle X^{(j)}_{\mathbf{s}-\mathbf{l}}, \hat{\phi}^{(j)}_{m,\mathbf{l}} \rangle,$$

   assuming that $1 + L \le s_i \le n_i - L$ for all $1 \le i \le N$, where L is an integer-valued truncation parameter.

In the second part, the scores and SFPC filters of $\mathbf{X}^{(j)}$ are utilized to compute the multivariate eigen elements, following the steps:

5. Define the matrix $\mathbf{E} \in \mathbb{R}^{n\times M_+}$ of rows $(\hat{\xi}^{(1)}_{1,\mathbf{s}}, \ldots, \hat{\xi}^{(1)}_{M_1,\mathbf{s}}, \ldots, \hat{\xi}^{(p)}_{1,\mathbf{s}}, \ldots, \hat{\xi}^{(p)}_{M_p,\mathbf{s}}), \mathbf{s} \in \mathbf{D}$ consists of all the scores estimated from the Spectral PCA on each $\mathbf{X}^{(j)}$, with $M_+ = M_1 + \ldots + M_p$.
   Let's consider the matrix $\mathbf{Z} \in \mathbb{R}^{M_+\times M_+}$ (Prop.5., p.7 (Happ and Greven, 2018)) consisting of blocks $\mathbf{Z}^{(jk)} \in \mathbb{R}^{M_j\times M_k}$ with entries

$$Z^{(jk)}_{ml} = \mathrm{Cov}(\hat{\xi}^{(j)}_{m,\mathbf{s}}, \hat{\xi}^{(k)}_{l,\mathbf{s}}),$$

$$m = 1, \ldots, M_j, \quad l = 1, \ldots, M_k, \quad j, k = 1, \ldots, p.$$

   An estimate $\hat{\mathbf{Z}} \in \mathbb{R}^{M_+\times M_+}$ of the matrix $\mathbf{Z}$ is given by $\hat{\mathbf{Z}} = (n-1)^{-1}\mathbf{E}^T\mathbf{E}$.

6. Perform a matrix eigen-analysis for $\hat{\mathbf{Z}}$ resulting in eigenvalues $\hat{\nu}_m$ and orthonormal eigenvectors $\hat{c}_m$.

7. The multivariate eigenfunctions applied for each operator of each variable are obtained as follows:

$$\hat{\psi}^{(j)}_{m,\mathbf{s}}(t_j) \approx \sum_{l=1}^{M_j} [\hat{c}_m]^{(j)}_l \, \hat{\phi}^{(j)}_{l,\mathbf{s}}(t_j), \tag{5.10}$$

$$t_j \in \mathcal{T}_j, \mathbf{s} \in \mathbf{D}, \ m = 1, ..., M_+$$

   where $[\hat{c}_m]^{(j)} \in \mathbb{R}^{M_j}$ denotes the j-th block of the (orthonormal) eigenvector $\hat{c}_m$ of $\hat{\mathbf{Z}}$.

Furthermore, multivariate scores are calculated as:

$$\hat{\rho}_{m,\mathbf{s}} = \sum_{j=1}^{p} \sum_{l=1}^{M_j} [\hat{c}_m]_l^{(j)} \, \hat{\xi}_{l,\mathbf{s}}^{(j)}. \tag{5.11}$$

We can deduce a sample version of the spatial Karhunen–Loève expansion for each univariate component:

$$X_{\mathbf{s}}^{(j)}(t_j) \approx \sum_{m=1}^{M_j} \hat{X}_{m,\mathbf{s}}^{(j)}(t_j), \quad t_j \in \mathcal{T}_j,$$
$$\text{with} \quad \hat{X}_{m,\mathbf{s}}^{(j)}(t_j) := \sum_{\|\mathbf{l}\|_\infty \leq L} \hat{\xi}_{m,\mathbf{s}+\mathbf{l}}^{(j)} \hat{\phi}_{m,\mathbf{l}}^{(j)}.$$

assuming $1 + 2L \leq s_i \leq n_i - 2L$ for $1 \leq i \leq N$.

5.4 Numerical experiments

In this section, we applied the SMFPCA procedure to simulated and real data. We compare the results of the proposed PCA methodology with MFPCA results. The experimentation for this study was conducted using the RStudio on a machine with the following specifications: a 64-bit Intel Core i5 processor and 12 GB of RAM. The principal libraries employed in our approach include fsd and fda. These libraries were instrumental in implementing the necessary functionalities for our analysis.

5.4.1 Simulation Study

In this subsection, we extend the simulation context of Kuenzer et al. (2020) to a multivariate case.

We consider $N = 2$ and simulate SFARMA process $\{X_{s,t}^j(.)\}$ defined by:

$$\begin{aligned} X_{s,t}^j =& A_{10} X_{s-1,t}^j + A_{01} X_{s,t-1}^j + \varepsilon_{s,t} + B_{10}\varepsilon_{s-1,t} \\ &+ B_{01}\varepsilon_{s,t-1} + B_{11}\varepsilon_{s-1,t-1}, \; t = 1, 2, \ldots 50, \end{aligned} \tag{5.12}$$

where $(\mathrm{A}_{kl})_{k,l\in P}$ and $(\mathrm{B}_{kl})_{k,l\in Q}$ are Hilbert–Schmidt operators, P and Q two finite index sets valued in $\mathbb{Z}^N$; the errors $\epsilon_{t,s}$ are i.i.d gaussian variables. To assess the effectiveness of integrating the spatial and multivariate aspects through SMFPCA, it is common to reconstruct the original functional data from the already computed scores and filters. For each instance, we employ both the novel SMFPCA and the conventional MFPCA. Figure 5.2 presents the SFPCA filters and the functional principal components in the case of a single simulated sample. We assess the effectiveness of dimension reduction using the metric known as the normalized mean squared error (NMSE) for each univariate component, as defined by:

$$\text{NMSE}(M_j) = \frac{\sum_{\mathbf{s}\in\mathbf{D_n}} \left\| X_{\mathbf{s}}^{(j)} - \sum_{m=1}^{M_j} \widehat{X}_{m,\mathbf{s}}^{(j)} \right\|^2}{\sum_{\mathbf{s}\in\mathbf{D_n}} \left\| X_{\mathbf{s}}^{(j)} \right\|^2}, \tag{5.13}$$

$\mathbf{D_n}$ represents a region where the average is calculated ($\mathbf{D_n} = \{\mathbf{s} \in \mathbb{Z}^N : 1 \leq \mathbf{s}_i \leq n_i \, 1 \leq i \leq n\}$)

Another alternative is to compute the error defined in equation (5.13) using the component eigenvalues $\hat{\lambda}_{j,m}$ of the spectral density operator $\widehat{\mathcal{F}}_\theta^X$:

$$\text{NMSE}^*_{\text{spat}}\,(M_j) = 1 - \frac{\sum_{m\leq M_j} \int_{[-\pi,\pi]^N} \hat{\lambda}_{j,m}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\sum_{m\geq 1} \int_{[-\pi,\pi]^N} \hat{\lambda}_{j,m}(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{5.14}$$

This measure assesses the quality of the approximation without being influenced by grid's boundary effects unlike the NMSE, where the approximation of $X_{\mathbf{s}}^{(j)}$ is less accurate at the boundary.

The experiments are conducted in 4 setting:

1. The first setting involves generating within the same domain $\mathcal{T}_j \in [0,1]$ with $j = 1,2$, two simulated functional variables following (5.12), denoted SIM_1 and SIM_2 with 10 and 14 Fourier basis functions respectively.

2. In setting two, we simulate variables defined in the first configuration and introduce errors following a normal distribution.

3. The third setting involves generating two SFARMA variables (denoted SIM_1 and SIM_2) in distinct domains, with $\mathcal{T}_1 \in [0,1]$ and $\mathcal{T}_2 \in [2,4]$. We use a set of 14 and 10 Fourier basis functions for SIM_1 and SIM_2 respectively.

4. In the fourth setting, we simulate two functional variables as in the third configuration and introduce errors following a normal distribution.

After obtaining the spatial functional data, a centering step is performed, followed by the application of univariate SFPCA for each variable. In the configuration utilized for each setting, we need to define two tuning parameters q and L.

These two parameters are selected with the automatic routine function defined by Kuenzer et al. (2020), we designate for the two components $q = (18, 17)$. The parameter L is 12, 10, 12, and 10 for the four settings respectively.

Then we proceed with describing steps in procedure 5.3.3, and compute 4 multivariate principal components to capture more than 80% of variability.

After performing SMFPCA, we conducted a comparative assessment of the reconstructed functional data. We applied equations (5.13) and (5.14), and the corresponding results are shown in Tables 5.1, 5.2, 5.3, and 5.4. We can observe the performance of SMFPCA compare to the conventional MFPCA approach, which does not account for spatial considerations. Using the first configuration as an illustrative example, as depicted in Table 5.1, the results emphasize that the enhancement in NMSE and NMSE* quality is linked to the number of functional principal components. This phenomenon is attributed to the fact that when the first three FPCAs are considered, they collectively account

for approximately 80% of the total variance. The measure NMSE*, which describes the quality of the approximation with no boundary effects, shows better result compare to NMSE. Regarding the outcomes depicted in Tables 5.2, 5.3, and 5.4, even when considering data defined in various domains and adding errors measurements, a similar trend can be observed in the results.

Table 5.1: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables SIM_1 and SIM_2 in different domain [0,1] and [2,4].

Cumulative PCA	PC1		PC2		PC3	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE SIM_1	0.5473	0.5877	0.3882	0.5807	0.2214	0.5659
NMSE* SIM_1	0.4366	0.5811	0.2350	0.3924	0.1369	0.2425
NMSE SIM_2	0.5448	0.9078	0.3798	0.6036	0.3621	0.3712
NMSE* SIM_2	0.4253	0.6039	0.2336	0.3710	0.1303	0.2419

Table 5.2: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables SIM_1 and SIM_2 in different domain [0,1] and [2,4] with introduced errors.

Cumulative PCA	PC1		PC2		PC3	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE SIM_1	0.4500	0.4591	0.3359	0.4589	0.2006	0.4503
NMSE* SIM_1	0.3617	0.4590	0.2007	0.2927	0.1193	0.1859
NMSE SIM_2	0.5764	0.9882	0.3305	0.5005	0.3123	0.3137
NMSE* SIM_2	0.3527	0.5005	0.1929	0.3136	0.1099	0.1941

Table 5.3: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables SIM_1 and SIM_2 in domain [0,1].

Cumulative PCA	PC1		PC2		PC3		PC4	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE SIM_1	0.6347	0.7171	0.4298	0.6040	0.4294	0.6034	0.2593	0.3840
NMSE* SIM_1	0.4729	0.6040	0.2705	0.3844	0.1542	0.2481	0.0904	0.1631
NMSE SIM_2	0.5566	0.7071	0.4465	0.6257	0.2645	0.4002	0.2643	0.3993
NMSE* SIM_2	0.4702	0.6258	0.2340	0.4002	0.1330	0.2532	0.0788	0.1644

Table 5.4: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables SIM_1 and SIM_2 in domain [0,1] with introduced errors.

Cumulative PCA	PC1		PC2		PC3		PC4	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE SIM_1	0.5753	0.8565	0.4199	0.6079	0.4174	0.5540	0.2526	0.3707
NMSE* SIM_1	0.4612	0.6080	0.2618	0.3710	0.1517	0.2433	0.0896	0.1577
NMSE SIM_2	0.5922	0.6404	0.4383	0.6234	0.2674	0.3932	0.2671	0.3897
NMSE* SIM_2	0.4655	0.6247	0.2375	0.3903	0.1348	0.2456	0.0793	0.1629

5.4.2 Application to real data

Following the application of the SMFPCA method to simulated data, we proceed to assess its performance on real data. The data we use in this part concerns sea surface temperature (SST) data (Kuenzer et al., 2020), from the NOAA Optimum Interpolation Sea Surface Temperature dataset. The dataset exhibiting typical spatial dependence characterized by an exponential decay. The SST dataset is obtained through the aggregation of satellite observations and in-situ measurements from ships and buoys (Kuenzer et al., 2020). It provides daily observations on a global grid with a resolution of 0.25, covering the entire sea area. The temperature data has been recorded since 1982, spanning a period of 33 years. The dataset exhibits annual quasi-periodicity and represents a spatially indexed functional random field. A subset extracted from the Indian Ocean area (60 to $93E$ longitude and 15 to $44S$ latitude) is chosen for its homogeneity and lack of significant oceanic currents. To address strong correlation among nearby observations, the grid resolution is reduced to 0.75, reducing computational load and supporting condition (5.1) where slight differences at the 0.25 grid lead to slow spatial autocorrelation decay. Figure 5.1 illustrates a snapshot of this extensive dataset.

For our analysis, we use two representative sea surface temperature variables, namely $TMP-2000$ and $TMP-2001$, which correspond to the years 2000 and 2001 respectively, which represent a multivariate aspect. To verify the efficiency of our methodology, we aim to conduct a comprehensive evaluation and analysis with three variables of $TMP-1996$, $TMP-1998$ and $TMP-1999$, which represent the sea surface temperature data for the respective years 1996, 1998, and 1999.

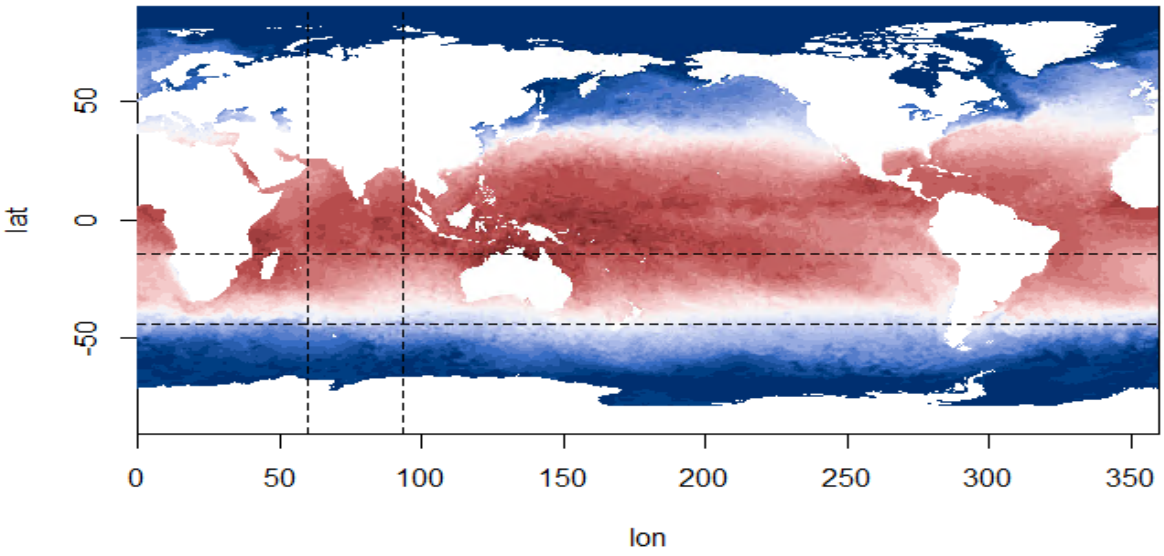


Figure 5.1: Indian Ocean ranging approximately from 60 to $93E$ longitude and 15 to $44S$ latitude.

We apply the SMFPCA and ordinary MFPCA on NOAA data, and illustrate the performance estimation procedure in two setting:

1. Application SMFPCA and ordinary MFPCA on two NOAA variables $TMP-2000$ and $TMP-2001$.

2. Application SMFPCA and ordinary MFPCA on three NOAA variables $TMP-1996$, $TMP-1998$ and $TMP-1999$.

We begin with the first configuration. The starting step of the SMFPCA algorithm involves computing the functional data. We transform the $TMP-2000$ and $TMP-2001$ into spatial function data using 15 Fourier basis functions, this projection yields functions denoted as $X_{s,u}^{(j)}(.)$, where the indices s, u, and j represent longitude, latitude, and year, respectively. The time domain is $\mathcal{T}_j$, it represents the intra-year time and is scaled to the unit interval $[0,1]$. Then we perform univariate SFPCA on these variables with a spatial parameter $L=22$ and $q=(18,17)$ for the $TMP-2000$ variable, and $L=22$ and $q=(19,19)$ for the $TMP-2001$ variable. We estimate the first 15 principal components to capture the total variation present in the functional data. Figures 5.2 and 5.3 show the first three principal components, which account for more than 80% of the total data variation. Notice that the results corresponding to position 0 represent an ordinary FPCA without considering the spatial parameter.

At this step, we have applied the initial part of our methodology, which involves conducting separate functional spectral analyses on the $TMP-2000$ and $TMP-2001$. Once the functional spatial filters have been computed, we obtain the resulting scores and spatial filter operators as shown in Figures 5.2 and 5.3. Subsequently, we proceed to the second part of the methodology, wherein the multivariate aspect is considered, following the guidelines specified in procedure 5.3.3. We have chosen to compute 4 multivariate spatial FPCA to capture more than 80% of the total variation.

Table 5.5 displays the percentage of variance explained by the principal components obtained through the application of SMFPCA to the variables $TMP-2000$ and $TMP-2001$. Additionally, Figures 5.4 and 5.5 show the filter operators presented in Table 5.5.

Table 5.5: Cumulative percentage (CP) of explained variance of principal components obtained by SFPCA considering the variables $TMP-2000$ and $TMP-2001$.

CP	PC1	PC2	PC3
$TMP-2000$	56.43	74.02	83.34
$TMP-2001$	49.38	72.89	83.20

In a manner similar to the simulated section, we assess the efficacy of dimensionality reduction through the evaluation of the Normalized Mean Square Error (NMSE) for each univariate component. We apply equation (5.13), and the outcomes are detailed in Table 3. Additionally, we employ the NMSE* metric as defined in equation (5.14), which characterizes the quality of approximation while accounting for boundary effects.

Table 5.7 shows the explained cumulative variance of the FPCA after application of the SMFPCA to the variables $TMP-2000$ and $TMP-2001$. The SFPCA integrate functional, multivariate, and spatial information, effectively capturing variations between $TMP-2000$ and $TMP-2001$ variables on marine surfaces and spatial models. We can observe that the first FPCA accounts for the largest variation in the data, which represents 58.80%. Furthermore, the 4 first components capture a significant amount of the overall variation, which represents 87.33% of the total variation.

The results presented in Table 5.6 provide strong evidence that the incorporation of the spatial aspect consistently leads to improved performance in the context of Normalized Mean Squared Error

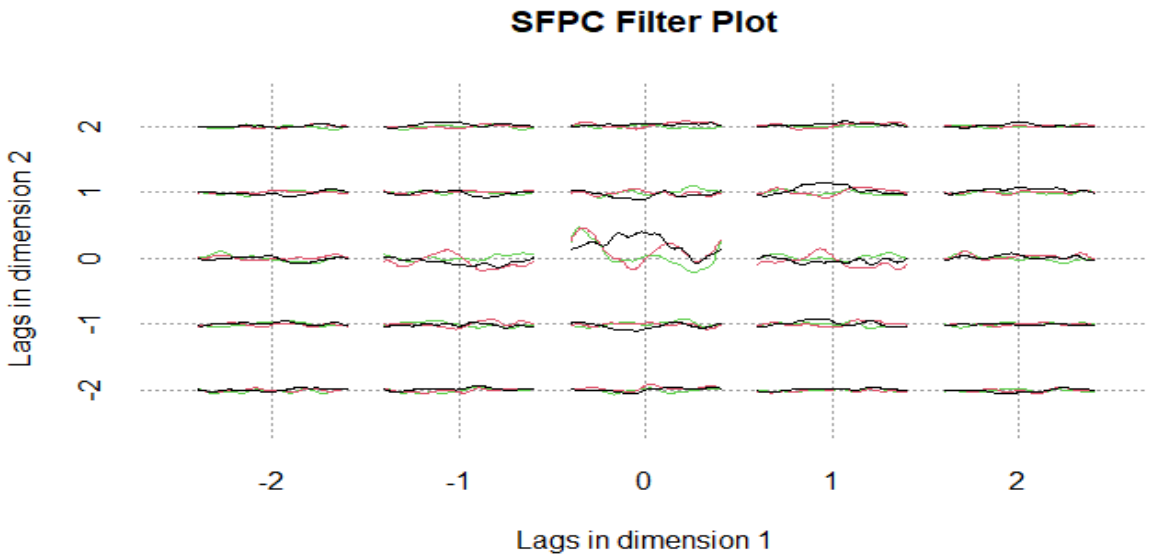


Figure 5.2: Evolution of three functional spatial filters of $TMP-2000$ variable over the lag.

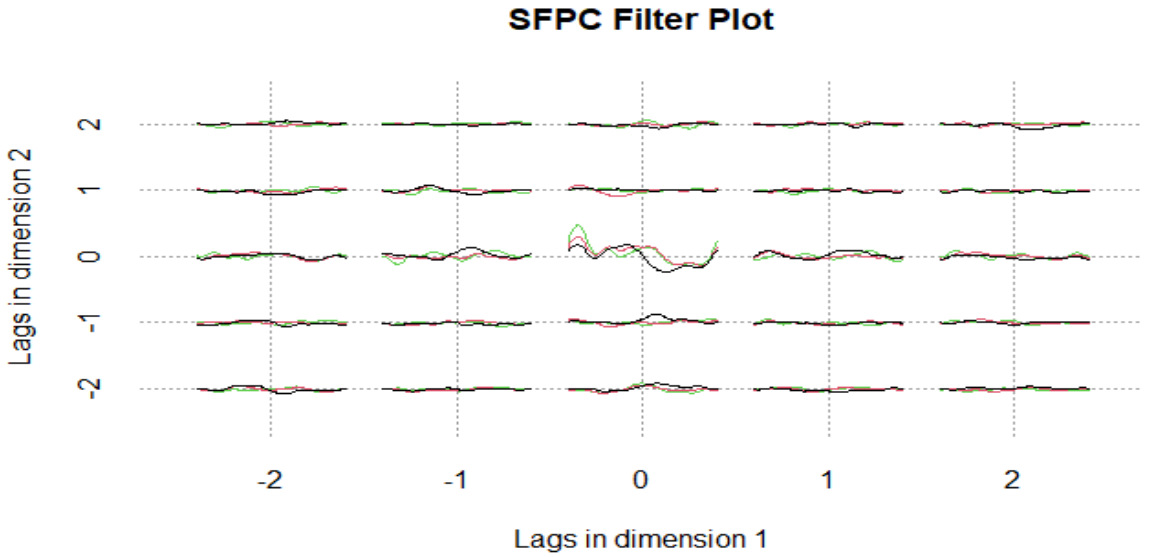


Figure 5.3: Evolution of three functional spatial filters of $TMP-2001$ variable over the lag.

(NMSE), we observe also that with NMSE*. The NMSE and NMSE* results exhibit a decreasing trend as the cumulative number of principal components increases, which is consistent with the cumulative variance explained by the FPCA. In other words, as more variance is accounted for by the FPCA, the NMSE and $\text{NMSE}^*_{\text{spat}}$ values decrease.

In setting 2, supplementary tests were conducted using sea surface temperature data to examine the performance of the SMFPCA algorithm on three variables. We randomly selected the NOAA variables representing the years 1996, 1998 and 1999, considering both the spatial and non-spatial aspects. We projected the data onto 15 Fourier basis functions, and we performed univariate SFPCA with the spatial parameters $L = 22$ and q defined as:

Table 5.6: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables $TMP-2000$ and $TMP-2001$.

Cumulative PCA	PC1		PC2		PC3	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE 2000	0.4796	0.5416	0.3396	0.5147	0.2103	0.3749
NMSE* 2000	0.4356	0.5156	0.2596	0.3342	0.1664	0.2695
NMSE 2001	0.5178	0.6016	0.3665	0.4121	0.3578	0.3627
NMSE* 2001	0.5061	0.6021	0.2709	0.3788	0.1678	0.2686

Table 5.7: Cumulative percentage (CP) of explained variance of principal components obtained by SMFPCA and MFPCA considering the variables $TMP-2000$ and $TMP-2001$.

CP	PC1	PC2	PC3	PC4
SMFPCA	58.80	69.97	79.72	87.33
MFPCA	37.59	52.11	63.78	68.63

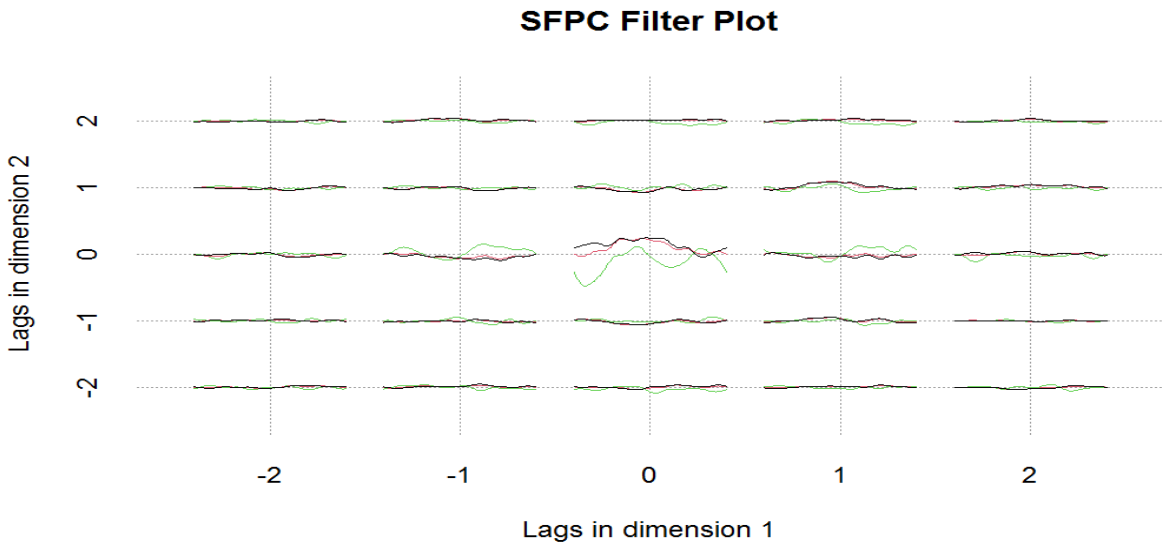


Figure 5.4: Evolution of three functional spatial filters of $TMP-2000$ variable over the lag after SMFPCA application.

- 1996 : $q=(15,14)$
- 1998 : $q=(22,22)$
- 1999 : $q=(19,18)$

Then we proceed with describing steps in procedure 5.3.3 for both settings, and compute the first 4 multivariate principal components to capture more than 80% of variability.

After computing SMFPCA, we performed a comparative assessment of the reconstructed functional data with ordinary MFPCA, and the corresponding results are shown in Table 5.8. The NMSE and NMSE* results exhibit a decreasing trend as the cumulative number of principal components increases, which is consistent with the cumulative variance explained by the FPCA. In other words, as more

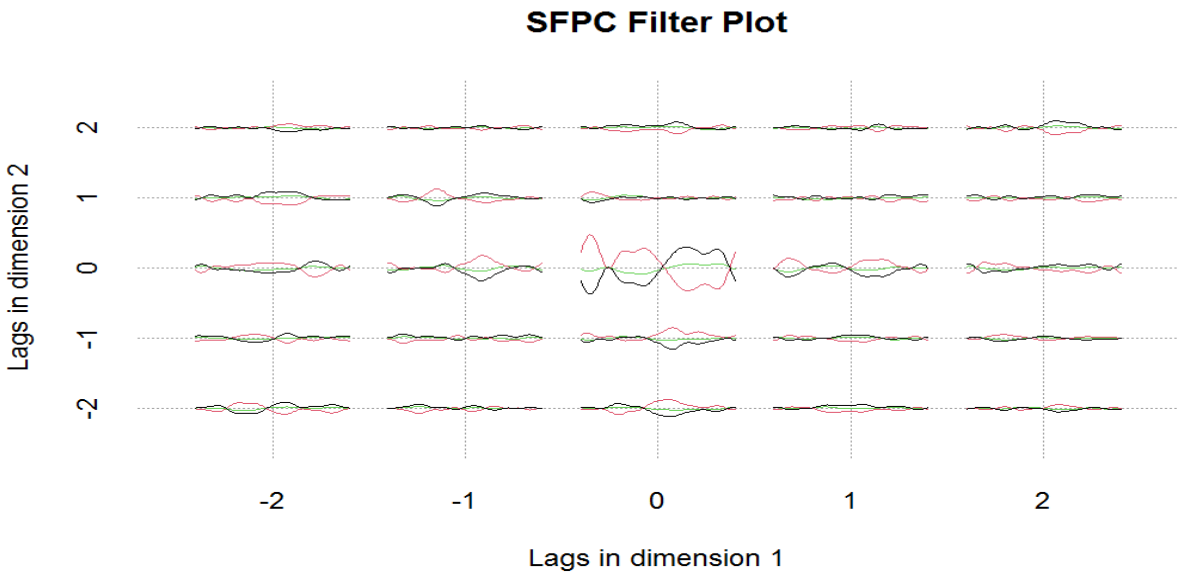


Figure 5.5: Evolution of three functional spatial filters of $TMP-2001$ variable over the lag after SMFPCA application.

variance is accounted for by the FPCA, the NMSE and NMSE* values decrease. This test confirms favorable outcomes, even when considering three or more variables.

Dimensionality reduction of multivariate functional spatial data shown in Figures 5.4 and 5.5, is achieved by identifying the principal directions of variation among the functional variables. Consequently, the data is succinctly represented through a restricted number of FPCA, facilitating simplified analysis and interpretation of two or more functional spatial variables. Additionally, this approach enhances the visualization and comprehension of results by enabling the identification of spatial patterns, structures, and trends within the data.

Table 5.8: NMSE and NMSE* results obtained by SMFPCA and MFPCA considering the variables $TMP-1996$, $TMP-1998$, and $TMP-1999$.

Cumulative PCA	PC1		PC2		PC3	
Spatial consideration	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
NMSE 1996	0.5090	0.6364	0.5069	0.5215	0.3223	0.5029
NMSE* 1996	0.4523	0.5358	0.2786	0.3772	0.1794	0.2851
NMSE 1998	0.6980	0.7111	0.3418	0.5812	0.3026	0.5069
NMSE* 1998	0.4476	0.5791	0.2624	0.3855	0.1640	0.2837
NMSE 1999	0.4377	0.4762	0.3053	0.3941	0.2758	0.3237
NMSE* 1999	0.4254	0.4744	0.2739	0.3520	0.1889	0.2778

5.5 Conclusion

In this work, we are interested to dimension reduction for multivariate spatial functional data. Within this domain, we have introduced a novel method known as SMFPCA (Spatial Multivariate Functional Principal Component Analysis). In the literature, the majority of existing methods are primarily based on Karhunen–Loève methodologies. In the existing literature, there is an absence of methods for processing multivariate spatial data using spectral functional analysis. In this context, we have developed methodologies and mathematical properties for conducting spectral FPCA on multivariate functional spatial data sampled on a regular grid, both within the same domain or different domains.

We initially conducted tests on simulated variables, implementing our methodology without considering the spatial dimension. This allowed us to make a comparison with the results obtained from the data that had spatial indexing. We evaluated our approach by using the Normalized Mean Squared Error (NMSE), which involves reconstructing data based on the calculated scores and filters. Furthermore, we assessed the quality of the approximation without any boundary effects using the NMSE* measure. The results for NMSE and NMSE* presented in Tables 5.6, 5.8, and 5.1, clearly demonstrating the substantial performance improvement achieved through the inclusion of the spatial aspect. In addition, we conducted supplementary tests on NOAA data, this data showed a low spatial dependency. We selected sea surface temperature variables for the years 2000 and 2001 as representative examples of the spatially indexed multivariate aspect. Furthermore, we performed supplementary tests on data from three NOAA variables taken from different years.
SMFPCA enhances data visualization by revealing spatial patterns and trends. This approach unveils spatial-temporal patterns, and provides valuable insights.

Conclusion and perspectives

The research conducted in this thesis represents a crucial and fundamental step towards a better understanding of the functioning of the pelagic ecosystem in West Africa. This thesis allowed for the examination of the impact of environmental factors on the characteristics of scattering layers. The conducted studies have highlighted an organization of the vertical and horizontal distributions of scattering layers, influenced by environmental conditions. The presented modeling framework uses FDA to analyzing acoustics data gathered periodically from various locations over time. FDA offers a range of tools to examine spatio-temporal patterns at a fine scale like in the data of interest. Combining FDA with marine data (multifrequential fish acoustics and environmental data) collected continuously from the different location surveyed over diurnal cycles has proven highly effective for indentifying fine-scale mechanisms and has wide-ranging applicability for coupling marine physical variables with "Sound Scattering Layers " caracteristics (thikness, depth and density).

In the first part of this thesis, we utilize traditional multivariate statistical unsupervised and supervised methods followed by FDA methods with or without spatial considerations. During the exploratory analysis, multivariate Functional Principal Component Analysis provides detailed insights into the variation of parameters along depths, in contrast to the conventional Principal Component Analysis. For regression tasks, we examine the interactions between Sound Scattering Layers descriptors and crucial environmental variables, both with and without spatial dimensions. We identify significant differences between Sound Scattering Layers in the northern and southern regions, as well as in coastal areas compared to open sea regions. The inclusion of spatial variables enhances the performance of the General Additive model, Functional Generalized Spectral model and signature-based PCR, linear, ridge and neural network models. The last contribution is proposing a multivariate functional principal component analysis for functional geostatistical data. The primary motivation for developing such methods relies on its capacity to take into account spatial dependency in the analysis.

Regarding the first contribution, the spatial dimension has been taken into account by using the spatial coordinates are covariates in the regression modeling. We may consider in the following more spatial modeling like geographical regression models. The use of weight matrices in the spatial principal component analysis of the second contribution may be improved by the geostatistical PCA as the one developed in the last contribution. The choice of weight matrices is also a hot point, weight matrices based on ecological distances would be relevant. The principal component analysis proposed in the last contribution has not been applied to the acoustic data. Using this methodology compare

to the PCA based on Moran index would be considered in the future. Furthermore, we have plans to develop a software package on the last contribution in both the R and Python programming languages. This would significantly expand the applicability and usability of the methodology, enabling more comprehensive analyses across a wider range of data structures.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2, 433–459.

Acar-Denizli, N., Delicado, P., Başarır, G., Caballero, I., 2018. Functional regression on remote sensing data in oceanography. Environmental and ecological statistics 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. Science Advances 3, e1602468.

Ali Hassan, A., 2021. Analyse des données spatiales: applications à la santé des populations. Ph.D. thesis. Université de Lille (2018-2021).

Alva, K., Romo, J., Ruiz Ortega, E., 2009. Modelling intra-daily volatility by functional data analysis: an empirical application to the spanish stock market. Statistics and Econometrics Series, vol 9. Universidad Carlos III de Madrid .

Anselin, L., 2009. Spatial regression. The SAGE handbook of spatial analysis 1, 255–276.

Antoniadis, A., Sapatinas, T., 2003. Wavelet methods for continuous-time prediction using hilbert−valued autoregressive processes. Journal of Multivariate Analysis 87, 133–158.

Aristizabal, J., Giraldo, R., Mateu, J., 2019. Analysis of variance for spatially correlated functional data: application to brain data. Spatial Statistics 32, 100381.

Ariza, A., Landeira, J., Escánez, A., Wienerroither, R., de Soto, N.A., Røstad, A., Kaartvedt, S., Hernández-León, S., 2016. Vertical distribution, composition and migratory patterns of acoustic scattering layers in the canary islands. Journal of Marine Systems 157, 82–91.

Ariza, A., Lebourges-Dhaussy, A., Nerini, D., Pauthenet, E., Roudaut, G., Assunção, R., Tosetto, E., Bertrand, A., 2022a. Acoustic seascape partitioning through functional data analysis. Journal of Biogeography .

Ariza, A., Lengaigne, M., Menkes, C., Lebourges-Dhaussy, A., Receveur, A., Gorgues, T., Habasque, J., Gutiérrez, M., Maury, O., Bertrand, A., 2022b. Global decline of pelagic fauna in a warmer ocean. Nature Climate Change 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. Progress in Oceanography 187, 102399.

Aston, J.A., Pigoli, D., Tavakoli, S., 2017. Tests for separability in nonparametric covariance operators of random surfaces. The Annals of Statistics 45, 1431–1461.

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.

Baillargeon, S., 2005. Le krigeage: revue de la théorie et application à l'interpolation spatiale de données de précipitations. Ph.D. thesis. Université Laval Laval.

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. Regional Studies in Marine Science 31, 100801.

Ballón, M., Bertrand, A., Lebourges-Dhaussy, A., Gutiérrez, M., Ayón, P., Grados, D., Gerlotto, F., 2011. Is there enough zooplankton to feed forage fish populations off peru? an acoustic (positive) answer. Progress in Oceanography 91, 360–381.

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. Progress in Oceanography 134, 221–231.

Béhagle, N., Cotté, C., Lebourges-Dhaussy, A., Roudaut, G., Duhamel, G., Brehmer, P., Josse, E., Cherel, Y., 2017. Acoustic distribution of discriminated micronektonic organisms from a bi-frequency processing: The case study of eastern kerguelen oceanic waters. Progress in Oceanography 156, 276–289.

Behrenfeld, M.J., Gaube, P., Della Penna, A., O'Malley, R.T., Burt, W.J., Hu, Y., Bontempi, P.S., Steinberg, D.K., Boss, E.S., Siegel, D.A., et al., 2019. Global satellite-observed daily vertical migrations of ocean animals. Nature 576, 257–261.

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. Journal of Plankton Research 36, 1279–1297.

Bernardi, M.S., Mazza, G., Ramsay, J.O., Sangalli, L.M., 2017. A penalized regression model for spatial functional data with application to the analysis of the production of waste in venice province. Stochastic Environmental Research and Risk Assessment 31, 23–38.

Berrendero, J., Justel, A., Svarc, M., 2011. Principal components for multivariate functional data. Computational statistics & data analysis 55, 2619–2634.

Bertrand, A., 1999. Le système {thon-environnement} en Polynésie française: caractérisation de l'habitat pélagique, étude de la distribution et de la capturabilité des thons, par méthodes acoustiques et halieutiques. Ph.D. thesis. École nationale supérieure agronomique de Rennes (1961-2004).

Bianchi, D., Mislan, K., 2016. Global patterns of diel vertical migration times and velocities from acoustic data. Limnology and Oceanography 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. Global Biogeochemical 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.

Bosq, D., 2000. Linear processes in function spaces: theory and applications. volume 149. Springer Science & Business Media.

Bouzebda, S., Nemouchi, B., 2020. Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional u-statistics involving functional data. Journal of Nonparametric Statistics 32, 452–509.

Brehmer, P., Balde, B., Mouget, A., 2020. Projet AGD-pélagique (CSRP-IRD): Appui à la Gestion Durable des pélagiques: rapport technique IRD 2020. Ph.D. thesis. IRD.

Brehmer, P., Gerlotto, F., Guillard, J., Sanguinède, F., Guénnegan, Y., Buestel, D., 2003. New applications of hydroacoustic methods for monitoring shallow water aquatic ecosystems: the case of mussel culture grounds. Aquatic Living Resources 16, 333–338.

Brehmer, P., Gerlotto, F., Rouault, A., 2002. In situ inter-standardization of acoustics data: an integrated database for fish school behaviour studies. Acta Acustica united with Acustica 88, 730–734.

Brehmer, P., Lafont, T., Georgakarakos, S., Josse, E., Gerlotto, F., Collet, C., 2006. Omnidirectional multibeam sonar monitoring: applications in fisheries science. FISH and Fisheries 7, 165–179.

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: An International Journal of Marine Sciences 35, 177–189. doi:10.1007/s41208-018-0107-9.

Brierley, A.S., 2014. Diel vertical migration. Current Biology 24, R1074–R1076. URL: https://www.sciencedirect.com/science/article/pii/S0960982214010677, doi:https://doi.org/10.1016/j.cub.2014.08.054.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Analytical methods 6, 2812–2831.

Brown, J., Fernand, L., Hill, A.E., 1996. Scanfish: high performance towed undulator. Sea Technology 37, 23–28.

Burgos, J.M., Horne, J.K., 2008. Characterization and classification of acoustically detected fish spatial distributions. ICES Journal of Marine Science 65, 1235–1247.

Cabreira, A., Madirolas, A., Brunetti, N., 2011. Acoustic characterization of the argentinean short-fin squid aggregations. Fisheries Research 108, 95–99.

Camara, M.M.B., 2008. Quelle gestion des pêches artisanales en Afrique de l'Ouest? Etude de la complexité de l'espace halieutique en zone littorale sénégalaise. Ph.D. thesis. Université Cheikh Anta Diop de Dakar.

Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the spde approach. AStA Advances in Statistical Analysis 97, 109–131.

Cardot, H., Faivre, R., Goulard, M., 2003a. Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. Journal of Applied Statistics 30, 1185–1199.

Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. Statistics & Probability Letters 45, 11–22.

Cardot, H., Ferraty, F., Sarda, P., 2003b. Spline estimators for the functional linear model. Statistica Sinica , 571–591.

Cardot, H., Sarda, P., 2005. Estimation in generalized linear models for functional data via penalized likelihood. Journal of Multivariate Analysis 92, 24–41.

Carroll, C., Müller, H.G., Kneip, A., 2021. Cross-component registration for multivariate functional data, with application to growth curves. Biometrics 77, 839–851.

Chatfield, C., Collins, A.J., 1980. Principal component analysis. Springer US, Boston, MA. volume 13. pp. 57–81. doi:10.1007/978-1-4899-3184-9_4.

Chen, K., Müller, H.G., 2012. Modeling repeated functional observations. Journal of the American Statistical Association 107, 1599–1609.

Chen, K.T., 1958. Integration of paths–a faithful representation of paths by noncommutative formal power series. Transactions of the American Mathematical Society 89, 395–407.

Cherry, S., 1997. Non-parametric estimation of the sill in geostatistics. Environmetrics: The official journal of the International Environmetrics Society 8, 13–27.

Chi, G., Zhu, J., 2008. Spatial regression models for demographic analysis. Population Research and Policy Review 27, 17–42.

Chiou, J.M., Chen, Y.T., Yang, Y.F., 2014. Multivariate functional principal component analysis: A normalization approach. Statistica Sinica , 1571–1596.

Chowdhury, J., Chaudhuri, P., 2019. Nonparametric depth and quantile regression for functional data. Bernoulli 25, 395–423.

Clarkson, D.B., 2005. S+ functional data analysis : user's manual for Windows®. 1st ed. 2005. ed., Springer New York, New York.

Cliff, A.D., Ord, J.K., 1973. Spatial autocorrelation. London: Pion Ltd .

Cliff, A.D., Ord, J.K., 1981. Spatial processes: models & applications. Taylor & Francis .

Constantinou, P., Kokoszka, P., Reimherr, M., 2017. Testing separability of space-time functional processes. Biometrika 104, 425–437.

Costanza, R., 1999. The ecological, economic, and social importance of the oceans. Ecological economics 31, 199–213.

Crambes, C., Mas, A., 2013. Asymptotics of prediction in functional linear regression with functional outputs. Bernoulli 19 (5B) , 2627–2651.

Cressie, N., 1993. Statistics for spatial data. Wiley series in probability and mathematical statistics Applied probability and statistics. revised edition ed., John Wiley & Sons, Inc, New York.

Croquette, M., 2007. Contribution à l'étude de l'upwelling du Pérou Chili. Ph.D. thesis. Université Paul Sabatier - Toulouse III.

Cuevas, A., 2014. A partial overview of the theory of statistics with functional data. Journal of Statistical Planning and Inference 147, 1–23.

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. Biological reviews 26, 158–192.

Cushing, D., Richardson, I., 1956. A record of plankton on the echo-sounder. Journal of the Marine Biological Association of the United Kingdom 35, 231–240.

Dabo-Niang, S., Ferraty, F., Vieu, P., 2004. Nonparametric unsupervised classification of satellite wave altimeter forms. Proceedings in Computational Statistics, Ed. J.Antoch, Physica-Verlag, Heidelberg New-York , 879–887.

Dabo-Niang, S., Ferraty, F., Vieu, P., 2007. On the using of modal curves for radar waveforms classification. Computational Statistics Data Analysis 51, 4878–4890.

Dai, X., 2018. Principal Component Analysis for Riemannian Functional Data and Bayes Classification. University of California, Davis.

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. Applied optics 44, 412–422.

Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. Journal of multivariate analysis 12, 136–154.

Delicado, P., Giraldo, R., Comas, C., Mateu, J., 2010. Statistics for spatial functional data: some recent contributions. Environmetrics 21, 224–239.

Delsol, L., 2008. Régression sur variable fonctionnelle: Estimation, tests de structure et Applications. Ph.D. thesis. Université Paul Sabatier-Toulouse III.

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., McLoone, S., 2013. Principal component analysis on spatial data: an overview. Annals of the Association of American Geographers 103, 106–128.

Di, C.Z., Crainiceanu, C.M., Caffo, B.S., Punjabi, N.M., 2009. Multilevel functional principal component analysis. The annals of applied statistics 3, 458.

Diankha, O., Thiaw, M., Sow, B.A., Brochier, T., GAyE, A.T., Brehmer, P., 2015. Round sardinella (sardinella aurita) and anchovy (engraulis encrasicolus) abundance as related to temperature in the senegalese waters. Thalassas 31, 9–17.

Diaw, A.T., Bâ, A., Bouland, P., Diouf, P.S., Lake, L.A., Mbow, M.A., Ndiaye, P., Thiam, M.D., 1993. Gestion des ressources côtières et littorales du Sénégal: Actes de l'Atelier de Gorée 27-29 JuiIIet 1992. UICN, Gland, Suisse. URL: https://portals.iucn.org/library/sites/library/files/documents/WTL-009.pdf.

Diner, N., Marchand, P., 1995. Acoustique et pêche maritime. Editions Quae.

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. Scientific reports 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 ICES 2022. Working Group of Fisheries Acoustics, Science and Technology (WGFAST) Ed. Michael Jech. ICES Scientific Reports 4:54.

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 Science 16, 65–81.

Niang Diop, I., 1995. L'érosion côtière sur la petite côte du Sénégal à partir de l'ensemble de Rufisque : passé, présent et futur. Ph.D. thesis. Université D'Angers.

Doray, M., 2006. L'agrégation de thons de sub-surface au sein du système [DCP ancré-macronecton-environnement-pêche] en Martinique: étude hiérarchique par méthodes acoustiques, optiques et halieutiques. Ph.D. thesis. Ecole Nationale Supérieure Agronomique de Rennes.

Dray, S., Saïd, S., Débias, F., 2008. Spatial ordination of vegetation data using a generalization of wartenberg's multivariate spatial correlation. Journal of vegetation science 19, 45–56.

Duby, C., Robin, S., 2006. Analyse en composantes principales. Institut National Agronomique, Paris-Grignon 80, 53.

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. Journal of Applied Ecology 49, 481–492.

Erbas, B., Hyndman, R.J., Gertig, D.M., 2007. Forecasting age-specific breast cancer mortality using functional data models. Statistics in Medicine 26, 458–470.

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. Journal of marine research 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. doi: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. Frontiers in Physics 3, 76.

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

Fermanian, A., 2022. Functional linear regression with truncated signatures. Journal of Multivariate Analysis , 192:105031.

Ferraty, F., Vieu, P., 2006a. Nonparametric functional data analysis. Springer.

Ferraty, F., Vieu, P., 2006b. Nonparametric functional data analysis: theory and practice. Springer Science Business Media.

Ferraty, F., Vieu, P., 2009. Additive prediction and boosting for functional data. Computational Statistics Data Analysis 53, 1400–1413.

Feuillet, T., Loonis, V., Bellefon, M.P., 2018. Manuel d'analyse spatiale. Théorie et mise en œuvre pratique avec R. Number 131 in Insee Méthodes, Insee, Eurostat.

Foote, K.G., 1980. Importance of the swimbladder in acoustic scattering by fish: a comparison of gadoid and mackerel target strengths. The Journal of the Acoustical Society of America 67, 2084–2089.

Fraiman, R., Justel, A., Liu, R., Llop, P., 2014. Detecting trends in time series of functional data: A study of antarctic climate change. Canadian Journal of Statistics 42, 597–609.

Gaetan, C., Guyon, X., 2010. Spatial statistics and modeling. Springer Series in Statistics.

Genin, M., Ahmed, M.S., 2019. A functional-model-adjusted spatial scan statistic. arXiv preprint arXiv:1903.01130 .

Getis, A., Aldstadt, J., 2004. Constructing the spatial weights matrix using a local statistic. Geographical analysis 36, 90–104.

Ghosh, S., Pahuja, K., Jiji, J.M., Puthussery, A., Shukla, S., Unal, A., 2021. Classifying bipolar personality disorder (bpd) using long short-term memory (lstm), in: Data Science and Security: Proceedings of IDSCS 2020, Springer. pp. 169–176.

Giraldo, R., Dabo-Niang, S., Martinez, S., 2018. Statistical modeling of spatial big data: An approach from a functional data analysis perspective. Statistics & Probability Letters 136, 126–129.

Giraldo, R., Delicado, P., Mateu, J., 2011. Ordinary kriging for function-valued spatial data. Environmental and ecological statistics 18, 411–426.

Gjøsæter, H., Ingvaldsen, R., Christiansen, J.S., 2020. Acoustic scattering layers reveal a faunal connection across the fram strait. Progress in Oceanography 185, 102348.

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 Environmental Sciences 26, 127–130.

Graham, B., 2013. Sparse arrays of signatures for online character recognition. arXiv preprint arXiv:1308.0371 .

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 in Food Science & Technology 72, 83–90. doi:10.1016/j.tifs.2017.12.006.

Griffith, D.A., 2020. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models, in: Practical handbook of spatial statistics. CRC press, pp. 65–82.

Gromenko, O., Kokoszka, P., Zhu, L., Sojka, J., 2012. Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. The Annals of Applied Statistics 6, 669–696.

Gromenko, O., Kokoszka, P., Zhu, L., Sojka, J., 2017. Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. The Annals of Applied Statistics 11, 898–918.

Gruber, N., Clement, D., Carter, B.R., Feely, R.A., Van Heuven, S., Hoppema, M., Ishii, M., Key, R.M., Kozyr, A., Lauvset, S.K., et al., 2019. The oceanic sink for anthropogenic co2 from 1994 to 2007. Science 363, 1193–1199.

Guillard, J., 2016. L'acoustique au service de l'écologie halieutique. Traitement du signal 33, 113–130.

Guillard, J., Baudoin, J.M., Goulon, C., 2023. Optimisation d'une méthode non intrusive de l'étude des populations de poissons en lac : l'hydro-acoustique. Sciences eaux & territoires la revue du Cemagref , 49–54.

Guillard, J., Colon, B., 1998. Estimation hydroacoustique du nombre de poissons migrateurs franchissant l'écluse de beaucaire-vallabrègues (rhône). Bulletin Français de la Pêche et de la Pisciculture , 79–90.

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 Journal of Marine Science 61, 267–273.

Guttmann, A., 2014. Evaluation des méthodes statistiques en épidémiologie spatiale: cas des méthodes locales de détection d'agrégats. Ph.D. thesis. Université d Auvergne.

Hager, P.P., 2021. Rough analysis with application in markets and related fields. Technische Universitaet Berlin, Germany.

Hall, P., Vial, C., 2006. Assessing the finite dimensionality of functional data. Journal of the Royal Statistical Society Series B: Statistical Methodology 68, 689–705.

Haney, J.F., 1988. Diel patterns of zooplankton behavior. Bulletin of marine science 43, 583–603.

Happ, C., Greven, S., 2018. Multivariate functional principal component analysis for data observed on different (dimensional) domains. Journal of the American Statistical Association 113, 649–659.

Happ, C.M., 2017. Statistical methods for data with different dimensions. Ph.d. thesis. Ludwig-Maximilians-Universität München.

Happ-Kurz, C., 2020. Object-oriented software for functional data. Journal of Statistical Software 93. doi:10.18637/jss.v093.i05.

Hastie, T., Mallows, C., 1993. [a statistical view of some chemometrics regression tools]: Discussion. Technometrics 35, 140–143.

Hastie, T., Tibshirani, R., 1986. Generalized additive models. Statistical science 1, 297–310.

Hastie, T., Tibshirani, R., 1990. Generalized additive models london chapman and hall. Inc .

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. volume 2. Springer.

Hays, G.C., Richardson, A.J., Robinson, C., 2005. Climate change and marine plankton. Trends in ecology & evolution 20, 337–344.

Henderson, B., 2006. Exploring between site differences in water quality trends: a functional data analysis approach. Environmetrics: The official journal of the International Environmetrics Society 17, 65–80.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Hörmann, S., Kidziński, Ł., Hallin, M., 2015. Dynamic functional principal components. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77, 319–348.

Horváth, L., Kokoszka, P., 2012. Inference for functional data with applications. volume 200. Springer Science Business Media.

Hyndman, R.J., Ullah, M.S., 2007. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis 51, 4942–4956.

Jacques, J., Preda, C., 2014. Model-based clustering for multivariate functional data. Computational Statistics & Data Analysis 71, 92–106.

James, G.M., Silverman, B.W., 2005. Functional adaptive model estimation. Journal of the American Statistical Association 100, 565–576.

Jean, G., Paul, F., Laloë, T., Brehmer, P., 2011. Three-dimensional internal spatial structure of young-of-the-year pelagic freshwater fish provides evidence for the identification of fish school species. Limnology and Oceanography: Methods 9, 322–328.

Jolliffe, I., 2005. Principal component analysis. Encyclopedia of statistics in behavioral science (eds B.S. Everitt and D.C. Howell). doi:10.1111/jbi.14534.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer Series in Statistics. second edition. ed., Springer, New York.

Jombart, T., Devillard, S., Dufour, A.B., Pontier, D., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101, 92–103.

Josso, N., 2010. Caractérisation des milieux sous marins en utilisant des sources mobiles d'opportunité. Ph.D. thesis. Université de Grenoble.

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. Journal of Marine Science and Technology 29, 9.

Karhunen, K., 1946. Zur spektraltheorie stochastischer prozesse. Ann. Acad. Sci. Fennicae, AI 34.

Kherif, F., Latypova, A., 2020. Principal component analysis, in: Machine Learning. Elsevier, pp. 209–225.

Kim, H., Kang, D., Cho, S., Kim, M., Park, J., Kim, K., 2018. Acoustic target strength measurements for biomass estimation of aquaculture fish, redlip mullet (chelon haematocheilus). Applied Sciences 8, 1536.

Kleffe, J., 1973. Principal components of random variables with values in a seperable hilbert space. Mathematische Operationsforschung und Statistik 4, 391–406.

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

Kloser, R., Ryan, T., Sakov, P., Williams, A., Koslow, J., 2002. Species identification in deep water using multiple acoustic frequencies. Canadian Journal of Fisheries and Aquatic Sciences 59, 1065–1077.

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, 998–1006.

Kokoszka, P., Reimherr, M., 2017. Introduction to functional data analysis. CRC press.

Koner, S., Staicu, A.M., 2023. Second-generation functional data. Annual Review of Statistics and Its Application 10, 547–572.

Korneliussen, R.J., Ona, E., 2002. An operational system for processing and visualizing multi-frequency acoustic data. ICES Journal of Marine Science 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. arXiv preprint arXiv:2211.04012 .

Kuenzer, T., Hörmann, S., Kokoszka, P., 2021. Principal component analysis of spatially indexed functions. Journal of the American Statistical Association 116, 1444–1456.

Kuenzer, T., Hörmann, S., Kokoszka, P., 2020. Principal component analysis of spatially indexed functions. Journal of the American Statistical Association 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 Research Part II: Topical Studies in Oceanography 120, 114–123.

Lawson, G.L., Wiebe, P.H., Ashjian, C.J., Gallager, S.M., Davis, C.S., Warren, J.D., 2004. Acoustically-inferred zooplankton distribution in relation to hydrography west of the antarctic peninsula. Deep Sea Research Part II: Topical Studies in Oceanography 51, 2041–2072.

Le Gall, Y., 2015. Problèmes inverses en acoustique sous–marine: prédiction de performances et localisation de source en environnement incertain. Ph.D. thesis. Université de Bretagne Occidentale.

Le Gallo, J., 2002. Econométrie spatiale: l'autocorrélation spatiale dans les modèles de régression linéaire. Economie prevision 155, 139–157.

Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P.A., Korsbakken, J.I., Peters, G.P., Canadell, J.G., Arneth, A., Arora, V.K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L.P., Ciais, P., Doney, S.C., Gkritzalis, T., Goll, D.S., Harris, I., Haverd, V., Hoffman, F.M., Hoppema, M., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Johannessen, T., Jones, C.D., Kato, E., Keeling, R.F., Goldewijk, K.K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F.N., van der Laan-Luijkx, I.T., van der Werf, G.R., Viovy, N., Walker, A.P., Wiltshire, A.J., Wright, R., Zaehle, S., Zheng, B., 2018. Global carbon budget 2018. Earth System Science Data 10, 2141–2194. URL: https://www.earth-syst-sci-data.net/10/2141/2018/, doi:10.5194/essd-10-2141-2018.

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 Journal of Marine Science 72, 1399–1412.

Leif, B., H., N.I.T., David, B., Bjørnø, L., 2017. Applied underwater acoustics. Elsevier Science, San Diego, CA.

LeSage, J.P., Pace, R.K., 2008. Spatial econometric modeling of origin-destination flows. Journal of Regional Science 48, 941–967.

Levenez, J.J., Lebourges-Dhaussy, A., Josse, E., 2006. Notions d'acoustique appliquée à l'halieutique et l'écologie aquatique. URL: https://docplayer.fr/35277785-Notions-d-acoustique-appliquee-a-l-halieutique-et-l-ecologie-aquatique-jean-jacques-html.

Levin, D., Lyons, T., Ni, H., 2013. Learning from the past, predicting the statistics for the future, learning an evolving system. arXiv preprint arXiv:1309.0260 .

Li, Y., Guan, Y., 2014. Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. Journal of the American Statistical Association 109, 1205–1215.

Li, Y., Qiu, Y., Xu, Y., 2022. From multivariate to functional data analysis: Fundamentals, recent developments, and emerging areas. Journal of Multivariate Analysis 188, 104806.

Li, Y., Wang, N., Carroll, R.J., 2013. Selecting the number of principal components in functional data. Journal of the American Statistical Association 108, 1284–1294.

Liu, C., Ray, S., Hooker, G., 2017a. Functional principal component analysis of spatially correlated data. Statistics and Computing 27, 1639–1654.

Liu, C., Ray, S., Hooker, G., 2017b. Functional principal component analysis of spatially correlated data. Statistics and Computing 27, 1639–1654.

Longhurst, A.R., Harrison, W.G., 1988. Vertical nitrogen flux from the oceanic photic zone by diel migrant zooplankton and nekton. Deep Sea Research Part A. Oceanographic Research Papers 35, 881–889.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30, 4765—-4774.

Lurton, X., 2002. An introduction to underwater acoustics: principles and applications. Springer Science & Business Media.

Lyons, T., Qian, Z., 2002. System control and rough paths. Oxford University Press.

MacLennan, D.N., Fernandes, P.G., Dalen, J., 2002. A consistent approach to definitions and symbols in fisheries acoustics. ICES Journal of Marine Science 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. Journal of Plankton Research 27, 623–633.

Marchal, E., Josse, E., Lebourges Dhaussy, A., 1996. Prédateurs et proies: une approche acoustique. Océanis 22, 117–132.

Marx, B.D., Eilers, P.H., 1999. Generalized linear regression on sampled signals and curves: a p-spline approach. Technometrics 41, 1–13.

Mateu, J., Giraldo, R., 2021. Geostatistical functional data analysis. volume 46. John Wiley & Sons.

Matheron, G., 1969. Le krigeage universel. volume 1. École nationale supérieure des mines de Paris Paris.

McLean, M.W., Hooker, G., Staicu, A.M., Scheipl, F., Ruppert, D., 2014. Functional generalized additive models. Journal of Computational and Graphical Statistics 23, 249–269.

Molinero, J.C., Ibanez, F., Nival, P., Buecher, E., Souissi, S., 2005. Climat nord-atlantique et variabilité du plancton nord-ouest méditerranéen. Limnologie et Océanographie 50, 1213–1220.

Moraga, P., Baker, L., 2022. rspatialdata: a collection of data sources and tutorials on downloading and visualising spatial data using r. F1000Research 11.

Moran, P.A.P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B, Methodological 10, 243–251.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Morrill, J., Kormilitzin, A., Nevado-Holgado, A., Swaminathan, S., Howison, S., Lyons, T., 2019. The signature-based model for early detection of sepsis from electronic health records in the intensive care unit, in: 2019 Computing in Cardiology (CinC), IEEE. pp. Page–1.

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. doi:10.3390/fishes7020086.

Mours, A., 2017. Localisation de cible en sonar actif. Ph.D. thesis. Université Grenoble Alpes.

Müller, H.G., Stadtmüller, U., 2005. Generalized functional linear models. Annals of Statistics 33, 774–805.

Müller, H.G., Yao, F., 2008. Functional additive models. Journal of the American Statistical Association 103, 1534–1544.

Müller, H.G., Sen, R., Stadtmüller, U., 2011. Functional data analysis for volatility. Journal of Econometrics 165, 233–245. URL: https://www.sciencedirect.com/science/article/pii/S0304407611001606, doi:https://doi.org/10.1016/j.jeconom.2011.08.002.

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

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. Mathematical Modeling of Random and Deterministic Phenomena , 31–51.

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. Université Pierre et Marie Curie Paris VI.

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. Journal of Geophysical Research: 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. Geophysical Research Letters 44, 5034–5043.

Nerini, D., Monestiez, P., Manté, C., 2010. Cokriging for spatial functional data. Journal of Multivariate Analysis 101, 409–418.

Niang, N.A., 2009. Dynamique socio-environnementale et développement local des régions côtières du Sénégal: l'exemple de la pêche artisanale. Ph.D. thesis. Université de Rouen.

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

Ona, E., 1990. Physiological factors causing natural variations in acoustic target strength of fish. Journal of the Marine Biological Association of the United Kingdom 70, 107–127.

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. Journal of Physical Oceanography 47, 29–47.

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

Perez Arribas, I., Goodwin, G.M., Geddes, J.R., Lyons, T., Saunders, K.E., 2018. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Translational psychiatry 8, 274.

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.

Pineda-Ríos, W., Giraldo, R., Porcu, E., 2019. Functional sar models: With application to spatial econometrics. Spatial statistics 29, 145–159.

Proud, R., Cox, M.J., Brierley, A.S., 2017. Biogeography of the global ocean's mesopelagic zone. Current Biology 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 Journal of Marine Science 76, 718–733.

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

Ramsay, J., Silverman, B.W., 1997. Functional data analysis (Springer series in statistics). Springer Series in Statistics.

Ramsay, J.O., Dalzell, C., 1991. Some tools for functional data analysis. Journal of the Royal Statistical Society: Series B (Methodological) 53, 539–561.

Ramsay, J.O., Silverman, B.W., 2002a. Applied functional data analysis: methods and case studies. volume 77. Springer.

Ramsay, J.O., Silverman, B.W. (Eds.), 2002b. Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics.

Ramsay, J.O., Silverman, B.W., 2005b. Functional Data Analysis. 2nd ed., Springer Series in Statistics. Springer, New York.

Ramsay, J.O., Silverman, B.W., 2005c. Functional data analysis, in: Springer Series in Statistics, Springer New York. pp. 327–348.

Rao, C.R., 1958. Some statistical methods for comparison of growth curves. Biometrics 14, 1–17. URL: http://www.jstor.org/stable/2527726.

Rebert, J.P., 1982. Hydrologie et dynamique des eaux du plateau continental sénégalais. Documents scientifiques du Centre de Recherche Océanographique Dakar Thiaroye, Dakar, Senegal .

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 Research Part II: Topical Studies in Oceanography 175, 104655.

Reizenstein, J., 2017. Calculation of iterated-integral signatures and log signatures. arXiv preprint arXiv:1712.02757 .

Reizenstein, J., Graham, B., 2018. The iisignature library: efficient calculation of iterated-integral signatures and log signatures. arXiv preprint arXiv:1802.08252 .

Remond, B., 2015. Les couches diffusantes du golfe de Gascogne: caractérisation acoustique, composition spécifique et distribution spatiale. Ph.D. thesis. Université Pierre et Marie Curie-Paris VI.

Reyes, A., Giraldo, R., Mateu, J., 2015. Residual kriging for functional spatial prediction of salinity curves. Communications in Statistics-Theory and Methods 44, 798–809.

Rokach, L., Maimon, O., 2005. Clustering methods, in: Data mining and knowledge discovery handbook. Springer, pp. 321–352.

Roy, C., 1990. Réponses des stocks de poissons pélagiques à la dynamique des upwellings en Afrique de l'Ouest : analyse et modélisation. Ph.D. thesis. Université de Bretagne Occidentale, Brest, France. URL: http://www.documentation.ird.fr/hor/fdi:34170.

Roy, C., 1991. Les upwellings: le cadre physique des pêcheries côtières ouest-africaines. Pêcheries Ouest Africaines: Variabilité, Instabilité et Changement.(Eds P. Cury and C. Roy.) pp , 38–66.

Sabine, C.L., Feely, R.A., Gruber, N., Key, R.M., Lee, K., Bullister, J.L., Wanninkhof, R., Wong, C., Wallace, D.W., Tilbrook, B., et al., 2004. The oceanic sink for anthropogenic co2. science 305, 367–371.

Salvetat, R., Garandel, Y., Aragon, B., GUILHOT, J., TOURENQ, J., 1990. Développement d'une instrumentation d'évaluation en biomasse de stocks piscicoles. Le Journal de Physique Colloques 51, C2–907.

Schmutz, A., Jacques, J., Bouveyron, C., Cheze, L., Martin, P., 2020. Clustering multivariate functional data in group-specific functional subspaces. Computational Statistics 35, 1101–1131.

Seck, A., 2014. Les pêcheurs migrants de Guet-Ndar (Saint-Louis du Sénégal): analyse d'une territorialité diverse entre espaces de conflits et espaces de gestion. Ph.D. thesis. Université de Liège, Belgique.

Shang, H.L., 2014. A survey of functional principal component analysis. AStA Advances in Statistical Analysis 98, 121–142.

Shanmuganathan, S., 2016. Artificial neural network modelling: An introduction. Springer.

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

Silverman, B., Ramsay, J., 2002. Applied functional data analysis: methods and case studies. Springer, New York.

Simmonds, J., MacLennan, D.N., 2005. Fisheries acoustics: theory and practice. Theory and Practice, 2nd edn. Blackwell, Oxford.

Siokou-Frangou, l., Papathanassiou, E., Lepretre, A., Frontier, S., 1998. Zooplankton assemblages and influence of environmental parameters on them in a mediterranean coastal area. Journal of Plankton Research 20, 847–870.

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? Journal of Geophysical Research: 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. Frontiers in Marine Science 9, 782032.

Srikhum, P., 2012. Statistiques spatiales et étude immobilière. Ph.D. thesis. Université Paris Dauphine-Paris IX. France.

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. Progress in Oceanography 74, 423–448.

Stequert, B., Gerlotto, F., 1977. Une méthode acoustique rapide d'évaluation des stocks de poissons pélagiques côtiers: l'écho-intégration. Pêche Marit. 56, 160–165.

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

Sugiura, N., Hosoda, S., 2020. Machine learning technique using the signature method for automated quality control of argo profiles. Earth and Space Science 7, e2019EA001019.

Suhaila, J., 2021. Functional data visualization and outlier detection on the anomaly of el niño southern oscillation. Climate 9, 118.

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

Teisson, C., 1983. Le phénomène d'upwelling le long des côtes du sénégal: caractéristiques physiques et modélisation. Documents scientifiques du Centre de Recherche Océanographique Dakar Thiaroye, Dakar, Senegal .

Ternynck, C., 2014. Contributions à la modélisation de données spatiales et fonctionnelles: applications. Ph.D. thesis. Université Charles de Gaulle-Lille III.

Theodore, D., 1993. Conflicts in coastal fisheries in cameroon, in: Workshop on Conflicts in Coastal Fisheries in West Africa. Cotonou, Benin, pp. 24–26.

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.

Thiao, D., 2009. Un système d'indicateurs de durabilité des pêcheries côtières comme outil de gestion intégrée des ressources halieutiques sénégalaises. Ph.D. thesis. Université de Versailles Saint-Quentin en Yvelines.

Tiedemann, M., Brehmer, P., 2017. Larval fish assemblages across an upwelling front: Indication for active and passive retention. Estuarine, Coastal and Shelf Science 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. Fisheries Oceanography 26, 655–667.

Wang, S., Jank, W., Shmueli, G., 2008. Explaining and forecasting online auction prices and their dynamics using functional data analysis. Journal of Business & Economic Statistics 26, 144–160.

Ward, M.D., Gleditsch, K.S., 2018. Spatial regression models. volume 155. Sage Publications.

Wartenberg, D., 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. Geographical analysis 17, 263–283.

Winzenborg, I., 2011. Spatial functional principal component analysis and its application in diagnostics. Ph.D. thesis. Universität Ulm.

Wong, R.K., Li, Y., Zhu, Z., 2019. Partially linear functional additive models for multivariate functional data. Journal of the American Statistical Association 114, 406–418.

Wynne, G., Duncan, A.B., 2022. A kernel two-sample test for functional data. Journal of Machine Learning Research 23, 1–51.

Yang, W., Jin, L., Liu, M., 2015. Character-level chinese writer identification using path signature feature, dropstroke and deep cnn. arXiv preprint arXiv:1505.04922 .

Yao, F., Müller, H.G., 2010. Functional quadratic regression. Biometrika 97, 49–64.

Yao, F., Müller, H.G., Wang, J.L., 2005. Functional linear regression analysis for longitudinal data .

Yarger, D., Stoev, S., Hsing, T., 2022. A functional-data approach to the argo data. The Annals of Applied Statistics 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.

Zhang, J., Siegle, G.J., D'Andrea, W., Krafty, R.T., 2019. Interpretable principal components analysis for multilevel multivariate functional data, with application to eeg experiments. arXiv preprint arXiv:1909.08024 .

Zhou, X., Lin, H., 2008. Spatial Weights Matrix. Springer US, Boston, MA. pp. 1113–1113. URL: https://doi.org/10.1007/978-0-387-35973-1_1307, doi:10.1007/978-0-387-35973-1_1307.

Zhou, Z., Lin, Z., 2016. Asymptotic normality of locally modelled regression estimator for functional data. Journal of Nonparametric Statistics 28, 116–131.

Zullo, A., 2016. Analyse de données fonctionnelles en télédétection hyperspectrale: application à l'étude des paysages agri-forestiers. Ph.D. thesis. Université de Toulouse, Université Toulouse III-Paul Sabatier.

Appendix A

Appendix

A.1 Appendix 1: matlatb code; association of Scanfish Data with echosounder echointegration cell

Matlatb code for averaging in for each echointegration cell (allowing them to be matched with the SSL descriptors), the environmental variables, i.e., seawater temperature (Temp in °C), fluorescence (Fluo in $ml\,l^{-1}$), turbidity (Turb in NTU), and salinity (Sal in psu; obtained from conductivity measurement), acquired every second using the Scanfish along the path of the vessel.

```
Idfreq=2;
load('DataScanFishEcho_RadialXX.mat');
IdEsu=dataek(Idfreq).EIIndexEsu;
IdDepth=dataek(Idfreq).EIIndexDepth;
ei=matfile('Echointegration');
EI_sf=ei.Sv_surface(IdDepth, IdEsu,Idfreq);
MeanEsuDuration=mean(diff(dataek(Idfreq).EITime));
MeanPingDuration=(diff(dataek(Idfreq).Time));
%Temperature
PingTime=dataek(Idfreq).Time; EsuTime=dataek(Idfreq).EITime;
   MeanEsuDuration=mean(diff(EsuTime));
PingDepth=dataek(Idfreq).Depth; EsuDepth=dataek(Idfreq).EIDepth;
   EIHeight=mean(diff(EsuDepth));
for kesu=1:length(EsuTime)
    idesu=find(PingTime>=EsuTime(kesu)-MeanEsuDuration/2 &
       PingTime<=EsuTime(kesu)+MeanEsuDuration/2);
    for kdep=1:length(EsuDepth)
        iddep=find(PingDepth>=EsuDepth(kdep)-EIHeight/2 &
           PingDepth<EsuDepth(kdep)+EIHeight/2);
        TemperatureforEI(kdep,kesu)=nanmean(nanmean(Temperature(
           iddep,idesu)));
    end
end
%Fluorescence
```

```
PingTime=dataek(Idfreq).Time; EsuTime=dataek(Idfreq).EITime;
   MeanEsuDuration=mean(diff(EsuTime));
PingDepth=dataek(Idfreq).Depth; EsuDepth=dataek(Idfreq).EIDepth;
   EIHeight=mean(diff(EsuDepth));
for kesu=1:length(EsuTime)
    idesu=find(PingTime>=EsuTime(kesu)-MeanEsuDuration/2 &
       PingTime<=EsuTime(kesu)+MeanEsuDuration/2);
    for kdep=1:length(EsuDepth)
        iddep=find(PingDepth>=EsuDepth(kdep)-EIHeight/2 &
           PingDepth<EsuDepth(kdep)+EIHeight/2);
        FluorescenceforEI(kdep,kesu)=nanmean(nanmean(Fluorescence
           (iddep,idesu)));
    end
end
%Turbidity
PingTime=dataek(Idfreq).Time; EsuTime=dataek(Idfreq).EITime;
   MeanEsuDuration=mean(diff(EsuTime));
PingDepth=dataek(Idfreq).Depth; EsuDepth=dataek(Idfreq).EIDepth;
   EIHeight=mean(diff(EsuDepth));
for kesu=1:length(EsuTime)
    idesu=find(PingTime>=EsuTime(kesu)-MeanEsuDuration/2 &
       PingTime<=EsuTime(kesu)+MeanEsuDuration/2);
    for kdep=1:length(EsuDepth)
        iddep=find(PingDepth>=EsuDepth(kdep)-EIHeight/2 &
           PingDepth<EsuDepth(kdep)+EIHeight/2);
        TurbiditeforEI(kdep,kesu)=nanmean(nanmean(Turbidity(iddep
           ,idesu)));
    end
end

%Salinity
PingTime=dataek(Idfreq).Time; EsuTime=dataek(Idfreq).EITime;
   MeanEsuDuration=mean(diff(EsuTime));
PingDepth=dataek(Idfreq).Depth; EsuDepth=dataek(Idfreq).EIDepth;
   EIHeight=mean(diff(EsuDepth));
for kesu=1:length(EsuTime)
    idesu=find(PingTime>=EsuTime(kesu)-MeanEsuDuration/2 &
       PingTime<=EsuTime(kesu)+MeanEsuDuration/2);
    for kdep=1:length(EsuDepth)
        iddep=find(PingDepth>=EsuDepth(kdep)-EIHeight/2 &
           PingDepth<EsuDepth(kdep)+EIHeight/2);
        SaliniteforEI(kdep,kesu)=nanmean(nanmean(Salinity(iddep,
           idesu)));
    end
end
```

Figure A1: A flowchart outlining the main research process

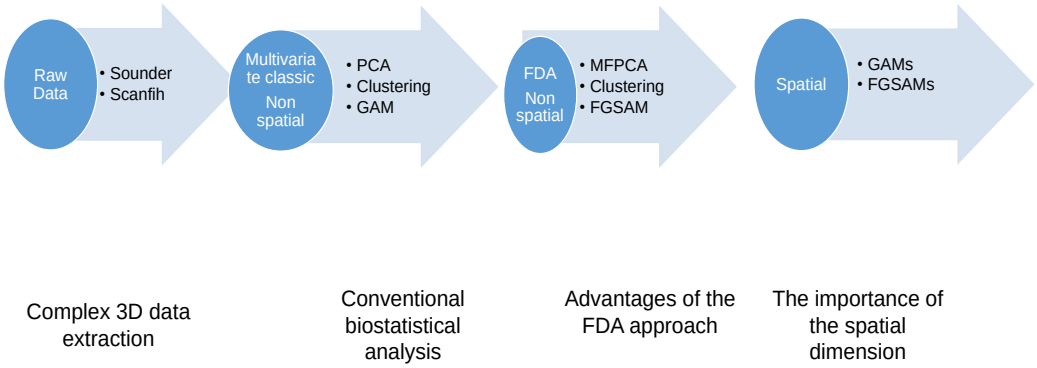
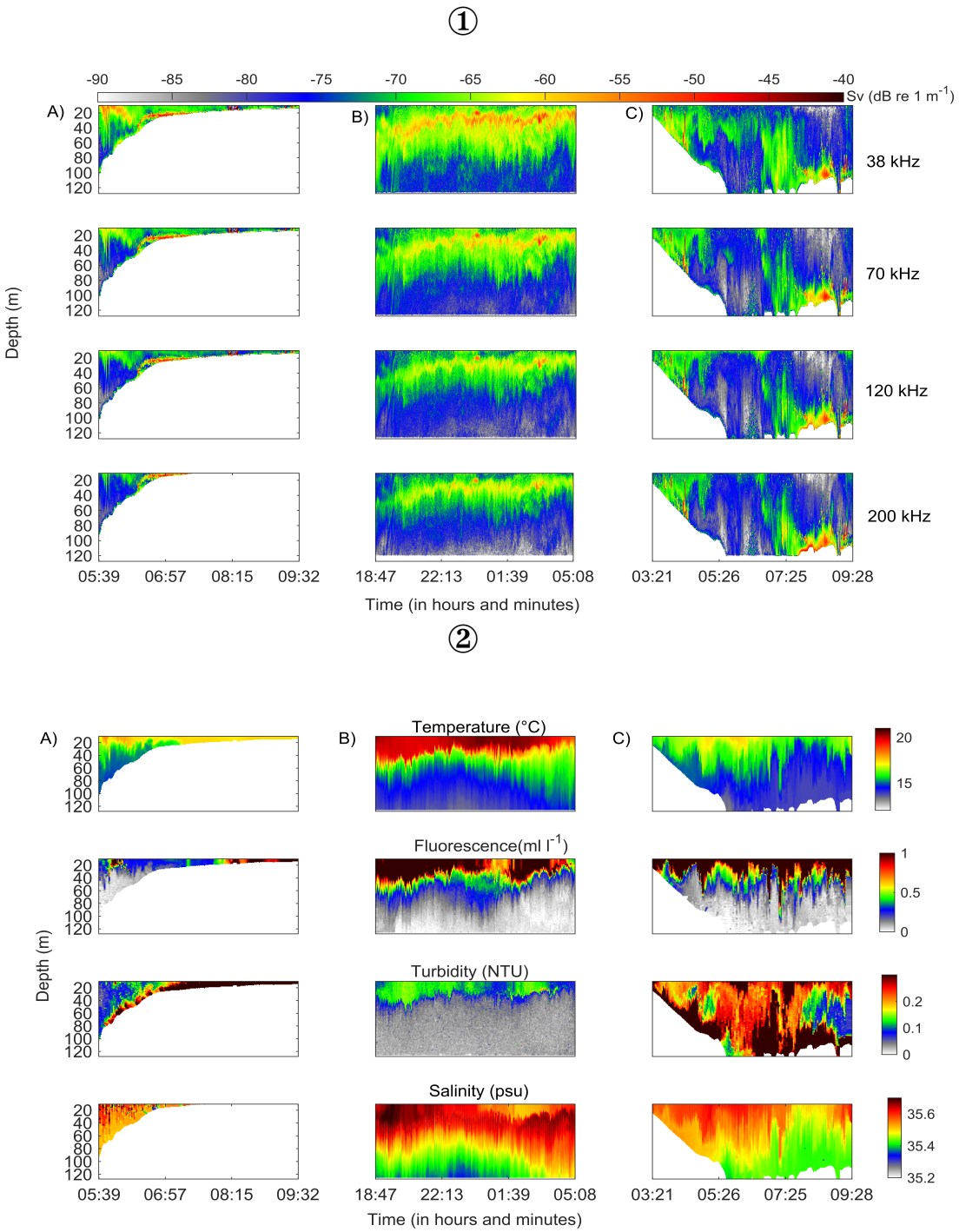


Figure A2: (1) Echogram representing the acoustic intensities (in Sv backscatter coefficient) reverberated by the aggregated marine organisms structured in layer (i.e. here sound scattering layer) as detected by echosounders at four frequencies (38, 70, 120 and 200 kHz) in three dimensions (2D + time). (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.



A.2 Appendix 2: Study of the impact of environmental variables on scattering layers using functional additive regression models

A.2.1 Spatial Generalized additive model (GAMs)

A.2.1.1 Descriptive results on the continental shelf delivered by the GAMs

On the continental shelf, the spatial GAM results (Tables A1 and A2) showed that the environmental parameters influence the SSL descriptors (thickness and depth) more during nighttime and sunrise than during daytime for all frequencies.

Local depth affects SSL thickness at night (negative effect on SSL thickness at 70 kHz and positive effect on SSL thickness of the other three frequencies) and at sunrise (positive effect for SSL thickness at 38 kHz and negative effect for SSL thickness for the other three frequencies) at all frequencies. The sea temperature also has an effect on thickness at nighttime (negative effect on SSL thickness for all frequencies) and sunrise for all frequencies except at 38 kHz where it was not significant during sunrise (negative effect on SSL thickness at 120 and 200 kHz and positive effect for SSL thickness at 70 kHz), salinity affects thickness at night except at 200 kHz (positive effect on SSL thickness at 38 and 70 kHz and negative effect at 120 kHz), during sunrise it only affects the 70 kHz SSL thickness (positive effect), and salinity during the day only affects thickness at 38 kHz (negative effect). Fluorescence affects thickness at nighttime at all frequencies (negative effect) and during sunrise, fluorescence only affects SSL thickness at 38 kHz (positive effect). Turbidity affects thickness at night for all frequencies except 120 kHz (negative effect on SSL thickness at 70 kHz and positive for the SSL thickness of the two others), while sunrise turbidity only affects 70 kHz thickness (positive effect). Geographical coordinates have a negative effect on SSL thickness for all frequencies.

For SSL depth, bottom depth had an effect at 38 and 70 kHz during the night (negative effect), while during sunrise an effect is observed only the 120 and 200 kHz (negative effect). Temperature affects SSL depth during sunrise at all frequencies except 38 kHz (positive effect on SSL depth at 120 kHz and negative effect for the two others), while nighttime sea temperature only affects the SSL depth at 70 and 38 kHz (negative effect) and daytime temperature only at 38 kHz (negative effect). Salinity only affects depth during sunrise at 120 (negative effect) and 200 kHz (positive effect). Turbidity affects the SSL depth at all frequencies except at 200 kHz during sunrise (positive effect). Geographical coordinates have a negative effect on SSL depth for all frequencies. For SSL density (Sa), the goodness of fit obtained is not satisfactory.

A.2.1.2 Descriptive results in the southern Senegalese high sea delivered by the GAMs

In the southern Senegalese high sea, bottom depth influences the SSL thickness at all frequencies during sunset (positive effect for SSL thickness at 38 and 200 kHz and negative effect for the SSL thickness of the two others) while at night SSL thickness is only influenced at 70 kHz (negative effect). Temperature influences SSL thickness during nighttime at all frequencies (positive effect only in SSL thickness at 200 kHz), whereas during sunset it only affects SSL thickness at 38 and 70 kHz (negative

effect). Salinity affects SSL thickness for all frequencies at night except 200 kHz (negative effect) and during sunset, it only affects the thickness at 70 kHz (positive effect). Fluorescence affects thickness during sunset (negative effect on SSL thickness at 120 kHz and positive effect on SSL thickness at 38 and 200 kHz) and night (negative effect on SSL thickness at 200 kHz and positive effect on SSL thickness at 38 and 120 kHz) for all frequencies except 70 kHz while turbidity has no effect at 200 kHz (positive effect on SSL thickness at the three frequencies during sunset and only positive effect at 70 kHz during night). Geographical coordinates have an effect on SSL thickness for all frequencies (positive effect of latitude on SSL thickness and positive effect of longitude on only SSL thickness at 120 kHz).

For the SSL depths, bottom depth affects 38 and 70 kHz at night (negative effect), while during sunset bottom depth affects the SSL depth at 120 and 200 kHz (positive effect). The sea temperature is highly significant at all frequencies (positive effect only on SSL depth at 200 kHz during sunset and negative for the three others) and also fluorescence is significant for all frequencies (positive effect on SSL depth during the night for all frequencies except at 120 kHz where we have a negative effect. During sunset, negative effects were recorded on SSL depth for all frequencies) while salinity has no effect at 70 kHz at night (negative effects were recorded on SSL depth for the three frequencies during the night and negative effect only at SSL depth of 120 kHz during sunset). Turbidity has an effect on all SSL depths during these two period except at 200 kHz during sunset (positive effect on SSL depth at 38, 70, and 120 kHz during sunset and negative effect on SSL depth for all frequencies during the night). Latitude has a positive effect on SSL depth for all frequencies while longitude has a positive effect only at SSL depth at 120kHz.

A.2.1.3 Descriptive results in the the northern Senegal delivered by the GAMs

In the northern Senegal shelf, bottom depth affects the SSL thickness at 38 and 200 kHz during sunrise (positive effect) while the daytime bottom depth increase only the thickness at 70 kHz. Temperature is significant for all frequencies during sunrise (positive effects were recorded on SSL thickness at 38 and 120 kHz while negative effects were recorded on SSL thickness at 70 and 200 kHz), while at night it does not affect the SSL thickness at 38 kHz (negative effect). Salinity only affects the SSL thickness at 70, 120, and 200 kHz during sunrise (positive effect). Fluorescence affects the 38, 70, and 120 kHz SSL thickness at night (negative effect only on SSL thickness at 38 kHz contrary to the two others) and sunrise (negative effect on SSL thickness for these three frequencies), while turbidity at sunrise has significant effects on all frequencies (positive effect), and turbidity at night affects only the 38 and 70 kHz SSL thickness (positive effect). Geographic coordinates have positive effects on the thickness at 38 and 200 kHz and negative effects on the thickness at the other frequencies.

Bottom depth affects the SSL depth at all frequencies during the day, except 200 kHz (positive effect) and sunrise (positive effect). Temperature affects SSL depth at all frequencies during sunrise (positive effect only on SSL depth at 120 kHz compared to the others) while at nighttime temperature did not affect the 70 and 120 kHz depth SSL (negative effect on SSL depth at 38 kHz and positive effect on SSL depth at 200 kHz) and daytime temperature affects only SSL depth at 200 kHz (positive effect). Salinity affects the SSL depth at 38, 70, and 120 kHz during sunrise. During daytime sea salinity affects SSL depth at 38 and 70 kHz (positive effect), and salinity only affects the 38 and 120 kHz SSL

depth during nighttime (positive effect on SSL depth at 38 kHz and negative effect on SSL depth at 120 kHz). Turbidity affects SSL depth during sunrise for all frequencies (positive effect) and fluorescence affects SSL depth during sunrise for all frequencies (negative effect). Turbidity affects only SSL depth at 38 kHz during the night (positive effect). Fluorescence did not affect the 200 kHz SSL depth during daytime (negative effect on SSL depth at 38, 70, and 120 kHz) while at nighttime fluorescence does not affect the 120 kHz SSL depth (positive effects were recorded only on SSL depth at 70 kHz compared to the 38 and 200 kHz). Geographic coordinates have negative effects on the SSL depth for all frequencies. For the SSL density (Sa), their fit qualities were low.

A.2.2 Multivariate Functional Principal Component analysis (MFPCA)

A.2.2.1 MFPCA results on the continental shelf of southern Senegal

On the continental shelf of southern Senegal (Figure 3.10), sea temperature data obtained at 38 kHz revealed a low peak between 20 and 30 meters, and a high peak between 40 and 50 meters in the first component. Similar observations were recorded for the second component. The salinity data showed a high peak between 40 and 50 meters in the first component, followed by a low peak between 50 and 60 meters. The second component exhibited a low peak between 20 and 30 meters, followed by a high peak between 40 and 50 meters. Fluorescence data at 38 kHz revealed a low peak between 10 and 20 meters in the first component, followed by a high peak at 30 meters. The second component displayed a low peak between 20 and 30 meters. Turbidity data at 38 kHz showed a weak peak between 10 and 20 meters in the first component, followed by a high peak at 30 meters. The second component was characterized by a high peak between 30 and 40 meters.

At 70 kHz frequency, the first component of sea temperature showed a low peak between 20 and 30 meters, and a high peak between 35 and 40 meters. The second component exhibited similar observations. The salinity data displayed a weak peak at 25 meters in the first component, and a weak peak between 20 and 25 meters in the second component. Fluorescence data showed a high peak between 10 and 15 meters in the first component, followed by a low peak between 25 and 30 meters. The second component displayed a low peak between 15 and 20 meters. Turbidity data in the first component was characterized by a high peak between 10 and 15 meters, followed by a low peak at 20 meters. The second component exhibited a low peak between 15 and 20 meters, followed by a high peak between 20 and 25 meters.

At 120 kHz frequency, sea temperature data displayed a low peak at 25-30 meters in the first component and a high peak at 40 meters. The second component exhibited a low peak at 25 meters, and a high peak at 40-45 meters. The salinity data showed a low peak at 25 meters in the first component and a high peak at 35 meters. The second component displayed a low peak at 25 meters. Fluorescence data revealed a low peak between 25 and 30 meters in the first component, and a low peak at 25 meters in the second component. Turbidity data exhibited a high peak between 15 and 20 meters in the first component, followed by a low peak between 40 and 45 meters. The second component displayed a high peak at 20 meters.

At 200 kHz frequency, sea temperature data displayed a high peak between 20 and 25 meters in the first

component, and a low peak at 35 meters. The second component showed a low peak between 20 and 25 meters. The salinity data exhibited a high peak between 15 and 20 meters in the first component, followed by a low peak between 30 and 35 meters. The second component displayed a low peak between 20 and 25 meters. Fluorescence data showed a high peak at 40 meters in the first component, and a low peak between 20 and 25 meters in the second component. Turbidity data displayed a low peak at 20 meters in the first component, followed by a high peak between 30 and 35 meters. The second component exhibited a low peak at 20 meters.

A.2.2.2 MFPCA results on the southern Senegalese high sea

The paragraph describes how different variables are grouped at different depth levels on the southern Senegalese high sea, using different frequencies. At 38 kHz, temperature ESUs with low values between 25 and 50 m are grouped in the first component, while those with high values at 50 m are grouped in the second component. Salinity ESUs with higher values at 25 m are grouped in the first component, while those with higher values between 50 and 62 m are grouped in the second component. Fluorescence ESUs with low values between 25 and 50 m are grouped in the first component, while those with higher values at 50 m are grouped in the second component. Turbidity showed similar variations.
At 70 kHz, temperature ESUs with higher values between 25 and 50 m are grouped in the first component, while those with high values at 50 m and low values between 25 and 37 m are grouped in the second component. Salinity ESUs with higher values between 50 and 62 m are grouped in the first component, with a similar peak in the second component. Fluorescence ESUs with higher values at 37 m are grouped in the first component, while those with higher values at 50 m are grouped in the second component. Turbidity ESUs with higher values at 37 m are grouped in the first component, while those with lower values between 25 and 37 m and higher values at 50 m are grouped in the second component.
At 120 kHz, temperature ESUs with higher values between 30 and 40 m are grouped in the first component, while those with higher values between 40 and 50 m are grouped in the second component. Salinity ESUs with low values between 20 and 30 m are grouped in the first component, while those with high values in the same range are grouped in the second component. Fluorescence ESUs with higher values between 30 and 40 m are grouped in the first component, while those with lower values between 20 and 30 m are grouped in the second component. Turbidity shows similar observations.
At 200 kHz, temperature ESUs with lower values at 50 m and higher values between 75 and 87 m are grouped in the first component, while those with low values between 25 and 37 m and high values at 62 m are grouped in the second component. Salinity ESUs with high values at 25 m are grouped in the first component, while those with higher values between 25 and 37 m are grouped in the second component. Fluorescence ESUs with low values between 37 and 50 m are grouped in the first component, while those with high values at 75 m are grouped in the second component. Turbidity ESUs with low values at 37 m are grouped in the first component, while those with low values between 25 and 37 m and high values at 50 m are grouped in the second component.

A.2.2.3 MFPCA results in the northern Senegalese shelf

The study conducted over the northern Senegalese shelf (Figure S2), revealed distinct patterns in the acoustic response of the ecosystem at different frequencies. At 38 kHz, the first component of temperature grouped ESUs with low values between 50 and 62 m, with similar observations made on the second component. The first component of salinity grouped ESUs with low values between 50 and 75 m, while the second component grouped those with low values at 75 m. The first component of fluorescence grouped ESUs with low values between 12 and 25 m, while the second component grouped those with high values between 25 and 50 m. The first component of turbidity grouped ESUs with high values between 37 and 50 m, while the second component grouped those with high values between 25 and 50 m.

At 70 kHz, the first component of temperature grouped ESUs with high values between 25 and 75 m, while the second component grouped those with low values between 50 and 62 m. The first component of salinity grouped ESUs with high values between 50 and 62 m, while the second component grouped those with low values at 62 m. The first component of fluorescence grouped ESUs with high values between 25 and 37 m, while the second component grouped those with high values between 37 and 50 m. The first component of turbidity grouped ESUs with high values at 87 m, while the second component grouped those with low values at this depth.

At 120 kHz, the first component of temperature grouped ESUs with low values between 25 and 37 m, while the second component grouped those with low values at 50 m. For salinity, the first component grouped ESUs with low values between 62 and 75 m, while the second component grouped those with low values between 50 and 62 m. The first component of fluorescence grouped ESUs with low values at 25 m, while the second component grouped those with high values between 25 and 37 m. The first component of turbidity grouped ESUs with high values at 37 m, while the second component grouped those with high values between 25 and 37 m.

Finally, at 200 kHz, the first component of temperature grouped ESUs with high values at 37 m, while the second component grouped those with low values between 50 and 62 m. For salinity, the first component grouped ESUs with high values at 62 m, while the second component grouped those with low values between 50 and 62 m. The first component of fluorescence grouped ESUs with high values between 12 and 25 m, while the second component grouped those with high values at 25 m. The first component of turbidity grouped ESUs with low values at 37 m, while the second component grouped those with high values at this depth.

A.2.3 Spatial functional generalized spectral additive model (FGSAMs)

A.2.3.1 Descriptive results on the southern continental shelf delivered by the FGSAMs

On the southern continental shelf (Figure 3.19 and Table A4), at 38 kHz, fluorescence has a positive effect on SSL thickness between 10 and 30m and a negative effect beyond. Temperature has an overall negative effect while turbidity has a negative effect between 10 and 30 and a positive effect beyond that. The same observations are done for salinity.

For SSL thickness, at 70 kHz fluorescence has a negative effect between 10 and 20, positive between

20 and 35, and negative beyond. The temperature has a positive effect between 10 and 20m, negative between 20 and 30, and positive beyond, while turbidity has a negative effect between 10 and 20m, positive between 20 and 30, and negative beyond. Salinity has a negative effect between 10 and 20 m, a positive between 20 and 30, and a negative beyond.
On the 120 kHz SSL thickness, fluorescence has a negative effect between 10 and 20m and positive beyond, while temperature has a positive effect between 10 and 35 and negative beyond. Turbidity has a negative effect between 10 and 25 and a positive effect beyond. Salinity has a positive effect between 10 and 25 and negative between 25 and 35 and a positive one beyond.
Fluorescence has a negative effect on the 200 kHz SSL thickness at depths below 15 and a positive effect between 15 and 25 and a negative effect beyond. The temperature has a positive effect between 10 and 25, a negative between 25 and 40, and a positive beyond. Turbidity has a positive effect between 10 and 25, a negative between 25 and 35, and a positive beyond. Salinity has a negative effect between 10 and 25, a positive between 25 and 40, and a negative beyond.
For SSL depth, on the 38 kHz frequency, fluorescence has a positive effect between 10 and 30, negative between 30 and 45, and positive beyond. The temperature has a positive effect between 10 and 35, a negative between 35 and 55, and a positive beyond, while turbidity has an overall positive effect. Salinity has a positive effect between 10 and 30 and a negative effect beyond.
For the 70 kHz depth, fluorescence has a negative effect between 10 and 20, a positive between 20 and 35, and a negative beyond. The temperature has a negative effect below 12m, positive between 12 and 35, and negative beyond. Turbidity has a negative effect between 10 and 30m, positive between 30 and 37m, and negative beyond. Salinity has a positive effect at depths below 12m, negative between 12 and 30, and positive beyond.
At 120 kHz, fluorescence has a negative effect between 10 and 20 m, positive between 20 and 40, and negative beyond. The temperature has a negative effect between 10 and 25, a positive between 25 and 40, and a negative beyond. Turbidity has a positive effect between 10 and 25 and a negative beyond. Salinity has a negative effect between 10 and 25, a positive between 25 and 40, and a negative beyond. Fluorescence has an overall negative effect on the depth of the 200 kHz SSL. The temperature has a positive effect at depths below 20m, a negative effect between 20 and 35, and a positive effect beyond. Turbidity has a positive effect between 10 and 25, a negative between 25 and 35, and a positive beyond. Salinity has a negative effect between 10 and 35, a positive between 35 and 47, and a negative beyond.

A.2.3.2 Descriptive results in the southern Senegal high sea delivered by the FGSAMs

In the southern Senegal high sea, at 38 kHz, fluorescence has a positive effect on SSL thickness, between 10 and 70 m and negative beyond. The Sea temperature has a positive effect between 10 and 50 m and a negative beyond. The water turbidity has a negative effect between 10 and 90 m and positive beyond while salinity has a negative effect between 10 and 50 m and positive beyond.
At 70 kHz, fluorescence has a positive effect, temperature has a negative effect between 10 and 40 m and positive beyond, turbidity has a negative effect beyond 10 m while salinity has a positive effect between 10 and 40 m and negative beyond.
At 120 kHz, fluorescence has an overall positive effect on thickness, the temperature has a negative

effect between 10 and 50 m and a positive effect beyond, turbidity has a positive effect between 10 and about 37 m and then a negative effect between 37 and 80 m and a positive effect beyond while salinity has a negative effect between 10 and 50 m and a positive effect beyond.
At 200 kHz, fluorescence has a negative effect between 10 and 40 m, positive between 40 and 120 m and negative beyond, the temperature has a negative effect at 10 m and positive beyond, turbidity has a negative effect between 10 and 80 m and positive beyond and salinity has a negative effect between 10 and 110 m and positive beyond.
For the 38 kHz fluorescence has a positive effect between 10 and 80 m on SSL depth and negative thereafter, temperature has a negative effect between 10 and 20 m and positive between 20 and 80 m and negative thereafter, turbidity has a negative effect between 10 and 90 m and positive thereafter while salinity has a negative effect between 10 and 50 m and positive thereafter.
On the 70 kHz, fluorescence has a positive effect on SSL depth between 10 and 70 m then slightly negative between 70 and 90 m then positive beyond, temperature has a negative effect between 10 and 50 m then positive between 50 and 90 m, and negative beyond, turbidity has an overall negative effect while salinity has a positive effect between 10 and 50 m and negative beyond.
At 120 kHz, fluorescence has a positive effect between 10 and 60 m and negative thereafter, temperature has a negative effect between 10 and 50 m, positive between 50 and 80 m, and negative thereafter, turbidity has a negative effect at depths below 20 m, positive between 20 and 50 m and negative thereafter, while salinity has a negative effect between 10 and 50 m and positive to 80 m and negative thereafter.
At 200 kHz, fluorescence has an overall negative effect on depth, the temperature has a negative effect between 10 and 50 m and a positive effect beyond that, turbidity has a negative effect between 10 and 80 m and a positive effect beyond that, while salinity has a negative effect between 20 and 110 m and a positive effect beyond that.

A.2.3.3 Descriptive results in northern Senegalese shelf sea delivered by the FGSAMs

In northern Senegalese shelf, at 38 kHz, fluorescence has a negative effect on SSL thickness between 10 and about 80 m and positive beyond, the temperature has a positive effect between 10 and 30 m, negative between 30 and 90 m and positive beyond, turbidity has a positive effect between 10 and 30 m, negative between 30 and 90 m and positive beyond, while salinity has a positive effect at depths less than 20 m, negative between 20 and 60 m and positive beyond.
At 70 kHz, fluorescence has a negative effect between 10 and 90 m and a positive effect beyond, temperature has a positive effect between 10 and 50 m and a negative effect beyond, turbidity has a positive effect between 10 and 30 m, a negative effect between 30 and 110 m and a positive effect beyond while salinity has a positive effect between 10 and 30 m, a negative effect between 30 and 110 m and a positive effect beyond.
At 120 kHz, fluorescence has a negative effect between 10 and 90 m and a positive effect beyond, temperature has a positive effect between 10 and 50 m and a negative effect beyond, turbidity has a negative effect between 10 and 20 m, a positive effect between 20 and 110 m and a negative effect beyond and salinity has a positive effect between 10 and 40 m, a negative effect between 40 and 90 m and a positive effect beyond.

At 200 kHz, fluorescence has a negative effect between 10 and 30 m, positive between 30 and about 120 m, and negative beyond, temperature has a positive effect between 10 and 30 m, negative between 30 and 50 m, positive between 50 and 120 m and negative beyond, turbidity has a positive effect between 10 and 20 m, negative between 20 and about 60 m, positive between 60 and 120 m and negative beyond while salinity has a negative effect between 10 and 80 m and positive beyond.
For SSL depth at 38 kHz, fluorescence has a negative effect between 10 and 90 m and a positive effect beyond, temperature has a negative effect between 10 and 50 m and a positive effect beyond, turbidity has a positive effect between 10 and 30 m, a negative effect between 30 and 90 m and a positive effect beyond and salinity has a positive effect between 10 and 120 m and a negative effect beyond.
At 70 kHz, fluorescence has a negative effect between 10 and 60 m, positive between 60 and 100 m and negative thereafter, temperature has a positive effect between 10 and 50 m, negative between 50 and 110 m and positive thereafter, turbidity has a positive effect between 10 and 30 m, negative between 30 and 80 m and positive thereafter while salinity has a negative effect between 10 and 90 m and positive thereafter.
At 120 kHz, fluorescence has negative and zero effects between 10 and 90 m and then positive beyond, the temperature has a positive effect between 10 and 20 m, negative between 20 and 110 m and then positive beyond, turbidity has a positive effect between 10 and 20 m, negative between 20 and 70 m, positive between 70 and 120 m and then negative beyond while salinity has a positive effect between 10 and about 37 m, negative between 37 and 57 m, positive between 57 m and 120 m and then negative beyond.
At 200 kHz, fluorescence has a negative effect between 10 and about 22 m, positive between 22 and 117 m and negative thereafter, temperature has a positive effect between 10 and 60 m, negative between 60 and 120 m and positive thereafter, turbidity has a negative effect between 10 and 60 m, positive between 60 and 120 m and negative thereafter, while salinity has a negative effect between 10 and 22 m, positive between 22 and 100 m and negative thereafter.

Table A1: Results of Generalized additive model (GAM) between sound scattering layers (SSLs) and oceanographic conditions (sea temperature (Temp), salinity (Sal), turbidity (Turb), fluorescence (Fluo)), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) to predict (1) SSL thickness and (2) SSL Depth and (3) SSL density, spread over three geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf) as observed during the AWA sea survey. AIC (Akaike's Information Criterion); BIC (Bayesian Information Criterion); Log Likelihood (log-likelihood value of a model); Deviance (goodness-of-fit metric for statistical model); Deviance explained (proportion of the total deviance explained by the current model); R^2 (Adjusted R-Squared); GCV score (Generalised Cross-Validation score); Num.obs. (number of observations); Num. smooth terms (Number of smooth terms).

① (A)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**13.48*****	4.32	3.17	6.86	**14.33*****	5.45	4.27	**5.93****
	(1.23)	(3.35)	(6.27)	(8.78)	(2.24)	(3.44)	(3.43)	(2.16)
s(Bottom):Night	**3.15****	**2.51****	**4.86*****	**3.39*****	**3.86*****	**2.63***	**4.41*****	**5.09*****
	(3.27)	(2.95)	(5.34)	(3.85)	(3.98)	(3.09)	(4.91)	(5.46)
s(Bottom):Sunrise	**2.88*****	**5.76*****	**1.00*****	**4.04*****	**3.11*****	**3.90***	**3.28***	**4.14***
	(3.07)	(6.14)	(1.00)	(4.79)	(3.31)	(4.53)	(3.92)	(4.78)
s(Bottom):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Night	**3.84*****	**3.97****	**1.00***	**1.00***	**3.89*****	**4.74*****	**1.00****	**2.14*****
	(3.97)	(4.50)	(1.00)	(1.00)	(3.99)	(5.12)	(1.00)	(2.62)
s(Temp):Sunrise	2.18	**3.29****	**6.55*****	**2.78*****	1.00	**5.10*****	**1.00***	**1.00*****
	(2.63)	(4.07)	(7.40)	(3.44)	(1.00)	(6.07)	(1.00)	(1.00)
s(Temp):Day	**1.00*****	1.00	1.00	1.00	**1.00***	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	**3.52*****	2.00	2.66	2.41	**3.60*****	**2.54****	**3.69****	2.63
	(3.85)	(2.41)	(3.22)	(2.92)	(3.89)	(3.01)	(4.26)	(3.16)
s(Sal):Sunrise	1.12	2.29	**1.00***	1.00	1.00	**2.10****	1.00	1.00
	(1.24)	(2.89)	(1.00)	(1.00)	(1.00)	(2.66)	(1.00)	(1.00)
s(Sal):Day	1.00	1.00	1.00	1.00	**1.00***	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**1.78*****	**3.04*****	**3.20***	**2.29****	**1.80*****	**3.08*****	**4.31*****	**1.00****
	(1.98)	(3.54)	(3.82)	(2.74)	(2.00)	(3.57)	(4.91)	(1.00)
s(Fluo):Sunrise	1.22	**1.00****	**4.40*****	**4.14*****	**1.36***	2.31	2.10	1.00
	(1.40)	(1.00)	(4.95)	(4.65)	(1.58)	(2.76)	(2.54)	(1.00)
s(Fluo):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	1.65	1.82	1.00	**1.84***	**1.00*****	**1.95*****	1.00	**1.00*****
	(1.71)	(1.92)	(1.00)	(1.92)	(1.00)	(1.98)	(1.00)	(1.00)
s(Turb):Sunrise	**1.00****	**2.37****	1.00	**4.60*****	2.39	**3.65****	3.63	2.75
	(1.00)	(2.95)	(1.00)	(5.14)	(2.70)	(4.28)	(4.31)	(3.43)
s(Turb):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Lon,Lat)					**2.66*****	**24.90*****	**24.85*****	**25.48*****
					(2.85)	(27.28)	(27.32)	(27.82)
AIC	999.93	1001.16	909.55	1049.94	890.68	706.41	804.57	853.40
BIC	1105.99	1135.30	1034.71	1180.98	1007.53	949.49	1019.80	1055.34
Log Likelihood	−468.84	−461.21	−418.04	−486.51	−411.05	−281.86	−339.11	−367.43
Deviance	874.91	817.07	554.76	1025.20	521.04	163.56	273.33	352.35
Deviance explained	0.98	0.93	0.87	0.75	0.99	0.99	0.94	0.91
Dispersion	4.49	4.32	2.91	5.41	2.72	1.02	1.64	2.08
R^2	0.97	0.91	0.85	0.70	0.98	0.98	0.92	0.89
GCV score	494.03	492.90	452.39	513.97	441.06	385.69	421.92	444.04
Num. obs.	223	223	223	223	223	223	223	223
Num. smooth terms	15	15	15	15	16	16	16	16

(B)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**64.04*****	**49.68*****	**34.41*****	**31.70*****	**65.53*****	**48.90*****	**29.44*****	**28.75*****
	(0.55)	(0.88)	(0.61)	(0.52)	(1.00)	(0.96)	(0.98)	(0.71)
s(Bottom):Night	**8.57*****	**8.36*****	**8.76*****	**8.66*****	5.57	**4.98*****	1.01	1.00
	(8.93)	(8.85)	(8.96)	(8.94)	(6.55)	(6.00)	(1.01)	(1.00)
s(Bottom):Sunset	1.00	**1.85*****	**1.99*****	**1.94*****	**1.00*****	**1.95*****	**1.92****	**1.00*****
	(1.00)	(1.96)	(2.01)	(2.00)	(1.00)	(1.99)	(1.99)	(1.00)
s(Temp):Night	**1.00*****	**4.53*****	**1.96*****	**4.59***	**3.73*****	**5.32*****	**3.00*****	**6.95*****
	(1.01)	(5.46)	(2.56)	(5.72)	(4.63)	(6.25)	(3.93)	(8.00)
s(Temp):Sunset	**3.70*****	**4.00*****	**1.00****	**3.88*****	**2.69*****	**5.11*****	3.08	3.69
	(4.53)	(4.94)	(1.00)	(4.75)	(3.34)	(6.16)	(3.88)	(4.55)
s(Sal):Night	**3.63*****	**3.08*****	**2.04****	**6.77*****	**1.00*****	**1.00****	**1.00*****	4.78
	(4.51)	(3.88)	(2.64)	(7.63)	(1.00)	(1.00)	(1.00)	(5.86)
s(Sal):Sunset	1.49	2.22	3.87	1.00	3.18	**3.76***	1.10	1.00
	(1.82)	(2.84)	(4.71)	(1.00)	(3.93)	(4.65)	(1.18)	(1.00)
s(Fluo):Night	**6.40*****	**5.68*****	**7.51*****	**6.54*****	**3.63****	1.01	**6.16*****	**4.24****
	(7.37)	(6.57)	(8.48)	(7.68)	(4.59)	(1.02)	(7.47)	(5.36)
s(Fluo):Sunset	1.00	**5.16*****	**2.97*****	**5.08*****	**4.24***	4.51	**4.46*****	**4.16*****
	(1.00)	(6.15)	(3.71)	(6.05)	(5.09)	(5.39)	(5.44)	(5.04)
s(Turb):Night	**4.26*****	3.75	**7.15*****	3.04	**4.03****	**5.03****	**7.57*****	2.82
	(5.31)	(4.70)	(8.19)	(3.99)	(5.01)	(5.98)	(8.47)	(3.70)
s(Turb):Sunset	4.27	1.00	**1.00*****	1.01	**6.53*****	**4.65*****	**1.00***	1.00
	(5.35)	(1.00)	(1.00)	(1.01)	(7.41)	(5.60)	(1.00)	(1.00)
s(Lon,Lat)					**25.96*****	**25.31*****	**26.73*****	**26.99*****
					(27.96)	(27.63)	(28.54)	(28.61)
AIC	3811.84	3820.00	3822.70	3776.45	3471.41	3459.43	3479.46	3397.44
Deviance explained	0.79	0.86	0.83	0.82	0.89	0.93	0.91	0.91
R^2	0.77	0.85	0.82	0.81	0.88	0.92	0.90	0.90
Num. obs.	602	602	602	602	602	602	602	602
Num. smooth terms	10	10	10	10	11	11	11	11

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(C)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	−5.44	**22.09*****	**18.21*****	**13.32*****	**18.62***	**23.48*****	**23.50*****	4.30
	(8.22)	(6.15)	(2.38)	(3.70)	(8.66)	(4.10)	(4.55)	(6.23)
s(Bottom):Night	**4.48*****	**6.68*****	**6.38*****	**6.14*****	2.19	1.00	1.00	1.00
	(5.28)	(7.13)	(6.93)	(6.76)	(2.59)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	1.00	**3.70***	1.00	1.00	**4.57*****	1.00	1.00	**4.10*****
	(1.00)	(4.34)	(1.00)	(1.00)	(5.18)	(1.00)	(1.00)	(4.80)
s(Bottom):Day	1.00	**2.89*****	**3.12*****	1.12	1.00	**3.02*****	1.00	1.71
	(1.00)	(3.45)	(3.69)	(1.22)	(1.00)	(3.67)	(1.00)	(2.01)
s(Temp):Night	1.65	**4.43*****	**6.14*****	**4.82***	2.08	**4.20*****	**6.28*****	**4.75*****
	(1.98)	(5.23)	(6.90)	(5.65)	(2.57)	(5.09)	(7.02)	(5.72)
s(Temp):Sunrise	**7.39*****	**6.62*****	**5.96*****	**7.86*****	**2.12****	**7.64*****	**6.54*****	**8.57*****
	(7.78)	(7.58)	(6.91)	(8.51)	(2.60)	(8.35)	(7.38)	(8.90)
s(Temp):Day	1.41	1.00	1.00	1.00	1.00	1.00	2.21	1.00
	(1.64)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(2.76)	(1.00)
s(Sal):Night	2.78	1.00	1.00	1.00	3.19	1.00	1.44	1.00
	(3.41)	(1.00)	(1.00)	(1.00)	(3.86)	(1.00)	(1.73)	(1.00)
s(Sal):Sunrise	**4.91*****	**6.77*****	**6.67*****	**8.16*****	1.00	**6.51*****	**7.00*****	**8.41*****
	(5.91)	(7.60)	(7.35)	(8.58)	(1.00)	(7.43)	(7.71)	(8.84)
s(Sal):Day	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.01)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**1.00***	**6.07*****	**5.39***	2.98	**3.76*****	**4.45***	**4.33***	1.00
	(1.00)	(7.15)	(6.47)	(3.77)	(4.64)	(5.49)	(5.38)	(1.00)
s(Fluo):Sunrise	**6.63*****	**3.72***	**2.91*****	**7.16*****	**2.61*****	**5.49*****	**4.35****	1.00
	(7.01)	(4.57)	(3.62)	(7.86)	(3.16)	(6.47)	(5.23)	(1.00)
s(Fluo):Day	1.00	1.50	1.00	**1.99***	1.00	1.00	1.00	1.00
	(1.00)	(1.68)	(1.00)	(2.15)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	**4.33*****	2.56	2.08	2.96	**4.75***	**3.38****	1.00	1.67
	(5.29)	(3.13)	(2.56)	(3.67)	(5.75)	(4.07)	(1.00)	(2.08)
s(Turb):Sunrise	**5.50*****	**5.94*****	**5.67*****	**4.28*****	**4.51*****	**5.67*****	**4.35*****	**4.31*****
	(5.99)	(6.41)	(6.45)	(4.90)	(5.09)	(6.20)	(5.22)	(4.96)
s(Turb):Day	1.06	**1.00****	**1.00*****	**3.16*****	1.63	1.00	1.00	**3.15***
	(1.10)	(1.00)	(1.00)	(3.63)	(1.98)	(1.00)	(1.00)	(3.67)
s(Lon,Lat)					**26.08*****	**25.28*****	**23.23*****	**26.73*****
					(27.96)	(27.65)	(26.31)	(28.43)
AIC	2934.35	2537.17	2502.36	2527.95	2605.35	2373.69	2394.55	2315.45
Deviance explained	0.76	0.84	0.85	0.84	0.91	0.91	0.90	0.92
R^2	0.73	0.81	0.82	0.82	0.89	0.88	0.87	0.90
Num. obs.	371	371	371	371	371	371	371	371
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

② (A)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**20.91****	9.64	12.11	22.52	**25.55****	**19.14****	**20.50***	**21.94*****
	(6.97)	(12.70)	(11.64)	(13.38)	(7.88)	(6.16)	(9.09)	(6.15)
s(Bottom):Night	2.28	1.00	1.00	1.00	**1.00***	**1.00***	1.00	1.00
	(2.67)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	**3.07*****	**2.85*****	**3.80*****	1.00	1.00	2.34	**3.65***	**4.04***
	(3.82)	(3.57)	(4.54)	(1.00)	(1.01)	(2.82)	(4.20)	(4.58)
s(Bottom):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.39
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.55)
s(Temp):Night	**3.06***	1.00	1.00	1.00	**2.69****	**1.00***	1.00	1.00
	(3.72)	(1.00)	(1.00)	(1.00)	(3.28)	(1.00)	(1.00)	(1.00)
s(Temp):Sunrise	1.00	**1.00*****	**7.41*****	**4.57*****	1.00	**3.68*****	**6.87*****	**4.22*****
	(1.00)	(1.00)	(7.90)	(5.49)	(1.00)	(4.53)	(7.44)	(5.01)
s(Temp):Day	**1.34***	1.00	1.00	1.00	**1.00****	1.00	1.00	1.00
	(1.54)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	1.77	1.00	1.00	1.00	1.67	1.47	1.00	1.00
	(2.20)	(1.00)	(1.00)	(1.00)	(2.07)	(1.79)	(1.00)	(1.00)
s(Sal):Sunrise	1.00	1.00	**1.00****	**3.80****	1.00	1.00	**1.08****	**3.80***
	(1.00)	(1.00)	(1.00)	(4.70)	(1.00)	(1.00)	(1.15)	(4.67)
s(Sal):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**2.02****	1.73	1.26	1.27	1.00	1.00	1.00	1.00
	(2.42)	(2.11)	(1.47)	(1.48)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Sunrise	1.00	**4.14*****	**3.42****	**4.11*****	1.00	1.00	1.28	1.00
	(1.00)	(4.62)	(3.98)	(4.70)	(1.00)	(1.00)	(1.45)	(1.00)
s(Fluo):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	1.18	1.00	1.00	1.00	1.71	1.65	1.00	1.00
	(1.30)	(1.00)	(1.00)	(1.00)	(1.90)	(1.86)	(1.00)	(1.00)
s(Turb):Sunrise	**4.65*****	**4.11****	**4.02***	1.00	**4.53*****	**4.12*****	**4.33****	3.74
	(5.24)	(4.71)	(4.68)	(1.00)	(5.13)	(4.72)	(4.91)	(4.40)
s(Turb):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Lon,Lat)					**20.04*****	**23.72*****	**20.60*****	**21.66*****
					(24.37)	(26.87)	(24.53)	(25.31)
AIC	1122.71	1266.56	1290.21	1320.44	1054.94	972.39	1132.36	1209.56
BIC	1231.44	1361.97	1407.97	1420.49	1224.48	1158.40	1322.09	1405.56
Log Likelihood	−529.45	−605.28	−610.54	−630.85	−477.71	−431.60	−510.49	−547.25
Deviance	1506.72	2974.34	3118.14	3741.18	947.34	626.49	1271.20	1767.63
Deviance explained	0.93	0.84	0.86	0.69	0.96	0.97	0.94	0.85
Dispersion	7.70	15.01	16.23	18.97	5.25	3.58	7.30	10.21
R^2	0.92	0.82	0.84	0.65	0.95	0.96	0.93	0.81
GCV score	539.71	607.98	628.73	633.21	511.06	484.72	559.92	592.63
Num. obs.	223	223	223	223	223	223	223	223
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(B)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**42.89*****	**36.84*****	**30.92*****	**31.64*****	**43.53*****	**36.20*****	**30.94*****	**32.09*****
	(0.27)	(0.47)	(0.28)	(0.27)	(0.48)	(0.47)	(0.63)	(0.56)
s(Bottom):Night	**8.62*****	**8.49*****	**8.11*****	**8.40*****	**5.09***	**6.46****	4.84	5.62
	(8.94)	(8.90)	(8.68)	(8.85)	(6.14)	(7.31)	(5.82)	(6.63)
s(Bottom):Sunset	**1.00*****	**1.90*****	**1.29*****	**1.89*****	1.00	1.00	**1.86*****	**1.89***
	(1.00)	(1.97)	(1.48)	(1.99)	(1.00)	(1.00)	(1.96)	(1.97)
s(Temp):Night	**1.00*****	**4.82*****	**2.14*****	**5.52*****	**3.11*****	**5.98*****	**4.83*****	**6.41*****
	(1.00)	(5.76)	(2.80)	(6.69)	(3.91)	(6.82)	(5.97)	(7.53)
s(Temp):Sunset	**5.27*****	**5.12****	**8.00*****	**6.45*****	**4.77*****	**6.05*****	**7.70*****	**6.11*****
	(6.27)	(6.28)	(8.66)	(7.35)	(5.73)	(7.16)	(8.45)	(7.06)
s(Sal):Night	**3.64*****	**3.30*****	**3.94*****	**6.10*****	**1.00*****	1.00	**3.74*****	**1.01***
	(4.53)	(4.14)	(4.87)	(7.11)	(1.00)	(1.00)	(4.66)	(1.01)
s(Sal):Sunset	**4.63*****	**4.27*****	**3.73*****	**8.03*****	**4.18****	**5.22*****	**3.68*****	**8.38*****
	(5.57)	(5.20)	(4.58)	(8.54)	(5.04)	(6.21)	(4.51)	(8.77)
s(Fluo):Night	**6.52*****	**5.73*****	**6.45*****	**6.96*****	**4.42*****	**3.97***	**5.74*****	**4.70*****
	(7.48)	(6.60)	(7.64)	(8.03)	(5.48)	(4.90)	(7.01)	(5.86)
s(Fluo):Sunset	**5.05*****	**6.67*****	2.29	1.00	**5.70*****	**6.54*****	**3.50***	**2.65*****
	(5.98)	(7.44)	(2.89)	(1.00)	(6.49)	(7.34)	(4.33)	(3.27)
s(Turb):Night	**3.75*****	**4.71***	**5.00***	4.01	**3.09****	**3.89****	**6.11*****	**3.23*****
	(4.76)	(5.67)	(6.20)	(5.15)	(3.94)	(4.83)	(7.29)	(4.18)
s(Turb):Sunset	**1.00****	**6.26*****	**4.71****	**7.58*****	**5.09*****	**6.27*****	**4.85****	5.04
	(1.00)	(7.16)	(5.68)	(8.38)	(6.06)	(7.16)	(5.85)	(6.11)
s(Lon,Lat)					**26.58*****	**25.69*****	**25.51*****	**26.57*****
					(28.23)	(27.86)	(27.79)	(28.29)
AIC	2941.41	2945.36	3065.03	3030.08	2548.67	2542.51	2830.61	2723.37
Deviance explained	0.89	0.92	0.91	0.93	0.95	0.96	0.95	0.96
R^2	0.88	0.91	0.90	0.93	0.94	0.96	0.94	0.96
Num. obs.	602	602	602	602	602	602	602	602
Num. smooth terms	10	10	10	10	11	11	11	11

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(C)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**63.15****	**19.35***	**25.49*****	14.29	**92.97*****	**43.44*****	**26.50*****	**39.40*****
	(21.65)	(9.68)	(7.47)	(10.01)	(20.62)	(8.98)	(7.37)	(10.33)
s(Bottom):Night	**4.40*****	**5.84*****	**5.81*****	**6.12*****	2.82	1.00	1.00	1.00
	(5.21)	(6.55)	(6.50)	(6.74)	(3.30)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	**5.08*****	**4.99*****	**5.38*****	**5.19*****	**4.58****	**4.27*****	**5.06*****	**4.19*****
	(5.50)	(5.53)	(5.89)	(5.64)	(5.05)	(4.88)	(5.62)	(4.76)
s(Bottom):Day	**1.44***	**1.00*****	**2.49*****	**3.35***	**1.00****	**1.00*****	**1.00****	3.10
	(1.73)	(1.00)	(2.94)	(3.90)	(1.00)	(1.00)	(1.00)	(3.62)
s(Temp):Night	**6.11*****	**1.00*****	2.61	**4.20***	**6.43*****	3.97	2.02	**4.72*****
	(6.73)	(1.00)	(3.21)	(4.94)	(7.00)	(4.83)	(2.54)	(5.54)
s(Temp):Sunrise	**7.68*****	**6.99*****	**7.92*****	**6.03****	**7.10*****	**8.09*****	**7.69*****	**8.05*****
	(7.94)	(7.88)	(8.60)	(6.95)	(7.56)	(8.62)	(8.38)	(8.54)
s(Temp):Day	1.00	1.00	1.00	2.94	1.00	1.00	1.00	**3.62*****
	(1.00)	(1.00)	(1.00)	(3.34)	(1.00)	(1.00)	(1.00)	(3.89)
s(Sal):Night	**5.40*****	**4.21****	**5.09***	2.18	**5.03*****	1.00	**5.26*****	2.86
	(6.05)	(5.00)	(5.91)	(2.72)	(5.74)	(1.00)	(6.13)	(3.56)
s(Sal):Sunrise	**5.08*****	**6.85*****	**8.60*****	**6.95*****	**4.93****	**7.40*****	**8.53*****	4.87
	(6.08)	(7.70)	(8.89)	(7.54)	(5.87)	(8.16)	(8.84)	(5.79)
s(Sal):Day	2.58	1.00	1.32	1.00	**2.89****	**2.66*****	1.00	1.00
	(2.98)	(1.00)	(1.50)	(1.00)	(3.31)	(2.96)	(1.00)	(1.00)
s(Fluo):Night	**4.87*****	**2.22*****	3.88	3.92	**5.29*****	**5.13*****	1.00	**6.19****
	(5.91)	(2.80)	(4.82)	(4.88)	(6.36)	(6.21)	(1.00)	(7.17)
s(Fluo):Sunrise	**6.43*****	**4.70*****	**8.40*****	**6.64*****	**6.27*****	**8.32*****	**7.44*****	**7.42*****
	(6.77)	(5.68)	(8.84)	(7.43)	(6.64)	(8.74)	(8.20)	(8.07)
s(Fluo):Day	**1.00*****	**2.16*****	**5.00*****	1.00	**1.00****	**1.00***	**5.11*****	1.00
	(1.00)	(2.36)	(5.56)	(1.00)	(1.00)	(1.00)	(5.67)	(1.00)
s(Turb):Night	**1.56***	2.27	1.00	1.00	**1.00***	1.56	1.00	1.20
	(1.92)	(2.79)	(1.00)	(1.00)	(1.01)	(1.92)	(1.00)	(1.36)
s(Turb):Sunrise	**3.83*****	**6.00*****	**6.23*****	**5.74*****	**6.87*****	**6.34*****	**3.34*****	**6.35*****
	(4.50)	(6.44)	(6.82)	(6.19)	(7.17)	(6.74)	(4.19)	(6.77)
s(Turb):Day	1.00	1.00	1.47	2.43	1.00	1.00	1.79	2.82
	(1.00)	(1.00)	(1.76)	(2.96)	(1.00)	(1.00)	(2.19)	(3.30)
s(Lon,Lat)					**24.04*****	**24.83*****	**25.14*****	**22.28*****
					(26.66)	(27.38)	(27.49)	(25.59)
AIC	2689.73	2673.54	2451.26	2580.09	2465.45	2412.89	2323.17	2319.48
Deviance explained	0.96	0.97	0.98	0.98	0.98	0.99	0.99	0.99
R^2	0.95	0.96	0.98	0.97	0.97	0.98	0.99	0.99
Num. obs.	371	371	371	371	371	371	371	371
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

③ (A)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	−355.57	62.86	−12.24	−773.82	330.45	694.83	886.00	380.21
	(1215.52)	(1112.02)	(951.35)	(1108.95)	(1257.77)	(1383.81)	(971.35)	(1086.23)
(Bottom):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	1.01	4.13	**4.43***	**3.86***	1.00	2.27	**2.15***	**1.97***
	(1.01)	(4.91)	(5.21)	(4.67)	(1.00)	(2.75)	(2.77)	(2.54)
s(Bottom):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Sunrise	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Sunrise	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Sunrise	**4.53*****	**4.86*****	**4.83*****	**4.52*****	**4.50*****	**4.58*****	**4.83*****	**4.54*****
	(4.85)	(5.18)	(5.16)	(4.96)	(4.83)	(4.90)	(5.19)	(5.01)
s(Fluo):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Sunrise	**6.42*****	**5.14*****	**5.48*****	**5.40*****	**6.41*****	**4.80*****	**5.31*****	**5.05*****
	(6.67)	(5.48)	(5.82)	(5.84)	(6.67)	(5.20)	(5.67)	(5.53)
s(Turb):Day	1.00	1.00	1.00	1.25	1.00	1.00	1.00	1.58
	(1.00)	(1.00)	(1.00)	(1.46)	(1.00)	(1.00)	(1.00)	(1.96)
s(Lon,Lat)					2.00	10.12	**2.00****	**2.00*****
					(2.00)	(13.90)	(2.00)	(2.01)
AIC	3053.96	3138.70	3022.89	3132.03	3052.99	3133.92	3021.19	3124.12
BIC	3144.37	3239.50	3125.73	3233.99	3150.10	3272.77	3122.16	3226.50
Log Likelihood	−1500.44	−1539.77	−1481.26	−1536.09	−1498.00	−1526.20	−1480.96	−1532.01
Deviance	9121706.49	12979044.80	7680023.36	12558234.92	8923868.89	11492746.05	7659412.69	12107392.14
Deviance explained	0.66	0.52	0.56	0.43	0.66	0.58	0.56	0.46
Dispersion	46058.95	66266.42	39333.02	64087.13	45510.36	61056.24	39138.43	61820.40
R^2	0.61	0.46	0.50	0.36	0.62	0.50	0.51	0.38
GCV score	1448.39	1485.52	1432.51	1479.85	1431.56	1463.47	1412.05	1456.46
Num. obs.	223	223	223	223	223	223	223	223
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(B)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	**28.90*****	**19.25*****	**15.57*****	**15.00*****	**27.99*****	**18.00*****	**14.58*****	**14.39*****
	(0.81)	(0.68)	(0.69)	(0.50)	(1.69)	(1.21)	(1.15)	(0.80)
s(Bottom):Night	**8.50*****	**8.35*****	**7.83*****	**7.62*****	**1.00***	**1.57****	1.00	1.00
	(8.91)	(8.85)	(8.54)	(8.46)	(1.00)	(1.95)	(1.00)	(1.00)
s(Bottom):Sunset	**1.00*****	**1.00****	**1.15*****	**1.62*****	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.27)	(1.81)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Night	1.00	**1.84***	1.00	1.01	2.16	1.12	1.00	1.00
	(1.00)	(2.33)	(1.00)	(1.01)	(2.76)	(1.24)	(1.00)	(1.00)
s(Temp):Sunset	1.00	1.00	1.00	1.92	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(2.38)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.01)	(1.00)	(1.01)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Sunset	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.01)	(1.01)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	**1.00***	**1.00****	**2.17****	**1.04*****	1.96	1.00	1.00	2.60
	(1.00)	(1.00)	(2.75)	(1.07)	(2.55)	(1.00)	(1.00)	(3.29)
s(Fluo):Sunset	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(2.48)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Night	**1.51***	**1.28****	**4.29*****	**4.98*****	1.00	1.00	1.32	3.05
	(1.89)	(1.50)	(5.30)	(6.09)	(1.00)	(1.00)	(1.57)	(3.85)
s(Turb):Sunset	1.00	1.92	1.00	1.00	1.89	1.00	1.00	1.00
	(1.00)	(2.39)	(1.00)	(1.00)	(2.40)	(1.00)	(1.00)	(1.00)
s(Lon,Lat)					**25.50*****	**26.21*****	**25.53*****	**24.50*****
					(28.05)	(28.35)	(28.10)	(27.53)
AIC	4453.14	4311.72	4235.92	3972.48	4354.37	4192.21	4132.82	3887.44
Deviance explained	0.39	0.36	0.33	0.40	0.52	0.51	0.46	0.51
R^2	0.38	0.34	0.31	0.38	0.49	0.47	0.43	0.48
Num. obs.	602	602	602	602	602	602	602	602
Num. smooth terms	10	10	10	10	11	11	11	11

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

(C)

	Non-Spatial				Spatial			
	38 kHz	70 kHz	120 kHz	200 kHz	38 kHz	70 kHz	120 kHz	200 kHz
(Intercept)	20.58	26.64	13.41	**40.78*****	29.02	346.35	520.74	**59.44*****
	(11.23)	(82.38)	(96.24)	(11.80)	(16.44)	(229.78)	(279.30)	(17.66)
s(Bottom):Night	2.72	1.00	1.01	**4.99*****	1.00	1.00	**1.00***	1.00
	(3.30)	(1.00)	(1.02)	(5.88)	(1.00)	(1.00)	(1.00)	(1.00)
s(Bottom):Sunrise	1.54	1.00	1.00	1.52	1.00	1.00	1.00	1.00
	(1.89)	(1.00)	(1.00)	(1.84)	(1.00)	(1.00)	(1.00)	(1.00)
s(Bottom):Day	**1.44*****	1.00	1.00	1.00	**1.98****	1.00	1.00	1.00
	(1.75)	(1.00)	(1.00)	(1.00)	(2.39)	(1.00)	(1.00)	(1.00)
s(Temp):Night	**1.00***	1.00	1.00	1.00	1.92	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(2.44)	(1.00)	(1.00)	(1.00)
s(Temp):Sunrise	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Temp):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Night	1.00	**7.95*****	**8.82*****	1.00	1.00	**7.95*****	**8.83*****	1.00
	(1.00)	(8.05)	(8.97)	(1.00)	(1.00)	(8.05)	(8.97)	(1.00)
s(Sal):Sunrise	1.00	1.00	1.00	**4.19****	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(5.12)	(1.00)	(1.00)	(1.00)	(1.00)
s(Sal):Day	**1.00***	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Night	1.04	1.00	1.00	3.85	1.00	1.00	1.00	4.19
	(1.08)	(1.00)	(1.00)	(4.75)	(1.00)	(1.00)	(1.00)	(5.12)
s(Fluo):Sunrise	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Fluo):Day	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.97***
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(2.20)
s(Turb):Night	**4.59****	1.00	1.00	**2.74***	**4.71****	1.00	1.00	2.67
	(5.60)	(1.00)	(1.00)	(3.38)	(5.73)	(1.00)	(1.00)	(3.27)
s(Turb):Sunrise	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
s(Turb):Day	2.33	1.00	1.00	2.94	2.91	1.00	1.00	3.43
	(2.85)	(1.00)	(1.00)	(3.51)	(3.51)	(1.00)	(1.00)	(3.88)
s(Lon,Lat)					5.17	2.00	2.00	**16.70*****
					(6.93)	(2.00)	(2.00)	(20.98)
AIC	4189.26	5667.89	5779.16	4136.79	4183.99	5669.16	5778.41	4106.43
Deviance explained	0.30	0.39	0.43	0.34	0.33	0.39	0.44	0.43
R^2	0.25	0.35	0.40	0.28	0.27	0.35	0.40	0.36
Num. obs.	371	371	371	371	371	371	371	371
Num. smooth terms	15	15	15	15	16	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A2: Results of Generalized additive model (GAM) and spatial Generalized additive model(GAMs) showing the effect (blue: Positive; red: Negative; gray : Not significant) of oceanographic conditions (sea temperature, salinity, turbidity and fluorescence), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) on sound-scattering layers (SSLs) (1) thickness and (2) depth, detected by echosounders (38, 70, 120 and 200 kHz), according to three Senegalese geographical areas (A: southern continental shelf; B: southern high sea and C: northern continental shelf).

① (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	-		-		-		-		-		-		-		-		-		-		-		-	

② (A)

Depth																								
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)

Depth																
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)

Depth																								
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	-		-		-		-		-		-		-		-		-		-		-		-	

Table A3: 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↗	25m↗
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↗

Table A4: 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 conditions (sea temperature, salinity, turbidity, fluorescence), diel period (day, sunset, night and sunrise), bottom depth (Bottom) and geographical positions (latitude (Lat) and longitude (Lon)) on sound-scattering layers (SSLs) (1) thickness and (2) depth, 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).

① (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	-		-		-		-	

② (A)

Depth										
Frequencies	38 kHz		70 kHz		120 kHz		200 kHz	
Model	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs	FGSAM	FGSAMs
Temperature	10-45 ; 45+	10-35 ; 35-55 ; 55+	10-20 ; 20-30 ; 30+	<12 ; 12-35 ; 35+	10-30 ; 30+	10-25 ; 25-40 ; 40+		<20 ; 20-35 ; 35+
Fluorescence	10-30 ; 30-45 ; 45+	10-30 ; 30-45 ; 45+		10-20 ; 20-35 ; 35+		10-20 ; 20-40 ; 40+		
Salinity	10-30 ; 30+	10-30 ; 30+	10-30 ; 30+	<12 ; 12-30 ; 30+	10-35 ; 35+	10-25 ; 25-40 ; 40+	10-25 ; 25+	10-35 ; 35-47 ; 47+
Turbidity			10-30 ; 30+	10-30 ; 30-37 ; 37+	10-25 ; 25+	10-25 ; 25+	10-25 ; 25-40 ; 40+	10-25 ; 25-35 ; 35+
Bottom								
Night								
Sunrise								
Day								
Lat	-		-		-		-	
Lon	-		-		-		-	

(B)

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

(C)

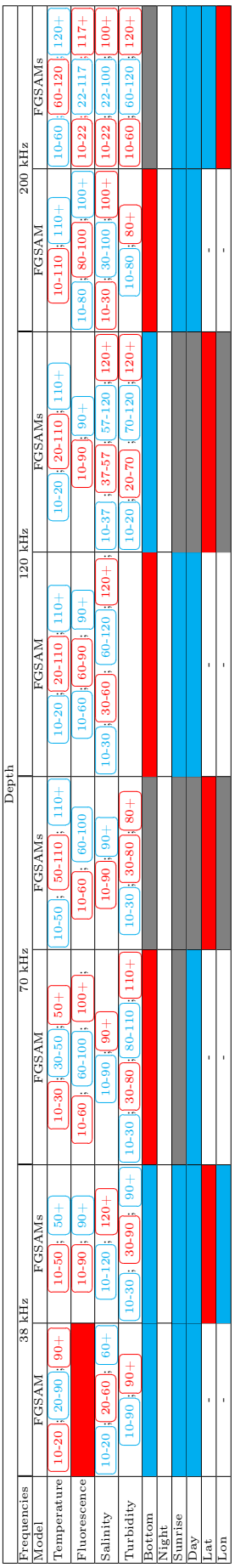

Depth
Frequencies
38 kHz
70 kHz
120 kHz
200 kHz
Model
FGSAM
FGSAMs
FGSAM
FGSAMs
FGSAM
FGSAMs
FGSAM
FGSAMs
Temperature
10-20 ; 20-90 ; 90+
10-50 ; 50+
10-30 ; 30-50 ; 50+
10-50 ; 50-110 ; 110+
10-20 ; 20-110 ; 110+
10-20 ; 20-110 ; 110+
10-110 ; 110+
10-60 ; 60-120 ; 120+
Fluorescence
10-90 ; 90+
10-60 ; 60-100 ; 100+
10-60 ; 60-100
10-60 ; 60-90 ; 90+
10-90 ; 90+
10-80 ; 80-100 ; 100+
10-22 ; 22-117 ; 117+
Salinity
10-20 ; 20-60 ; 60+
10-120 ; 120+
10-90 ; 90+
10-90 ; 90+
10-30 ; 30-60 ; 60-120 ; 120+
10-37 ; 37-57 ; 57-120 ; 120+
10-30 ; 30-100 ; 100+
10-22 ; 22-100 ; 100+
Turbidity
10-90 ; 90+
10-30 ; 30-90 ; 90+
10-30 ; 30-80 ; 80-110 ; 110+
10-30 ; 30-80 ; 80+
10-20 ; 20-70 ; 70-120 ; 120+
10-80 ; 80+
10-60 ; 60-120 ; 120+
Bottom
Night
Sunrise
Day
-
-
Lat
-
-
Lon
-
-

Figure A3: Variation of physical parameters along depths over three contrasted Senegalese areas: (A) southern continental shelf; (B) southern high sea and (C) northern continental shelf) during the AWA sea survey.

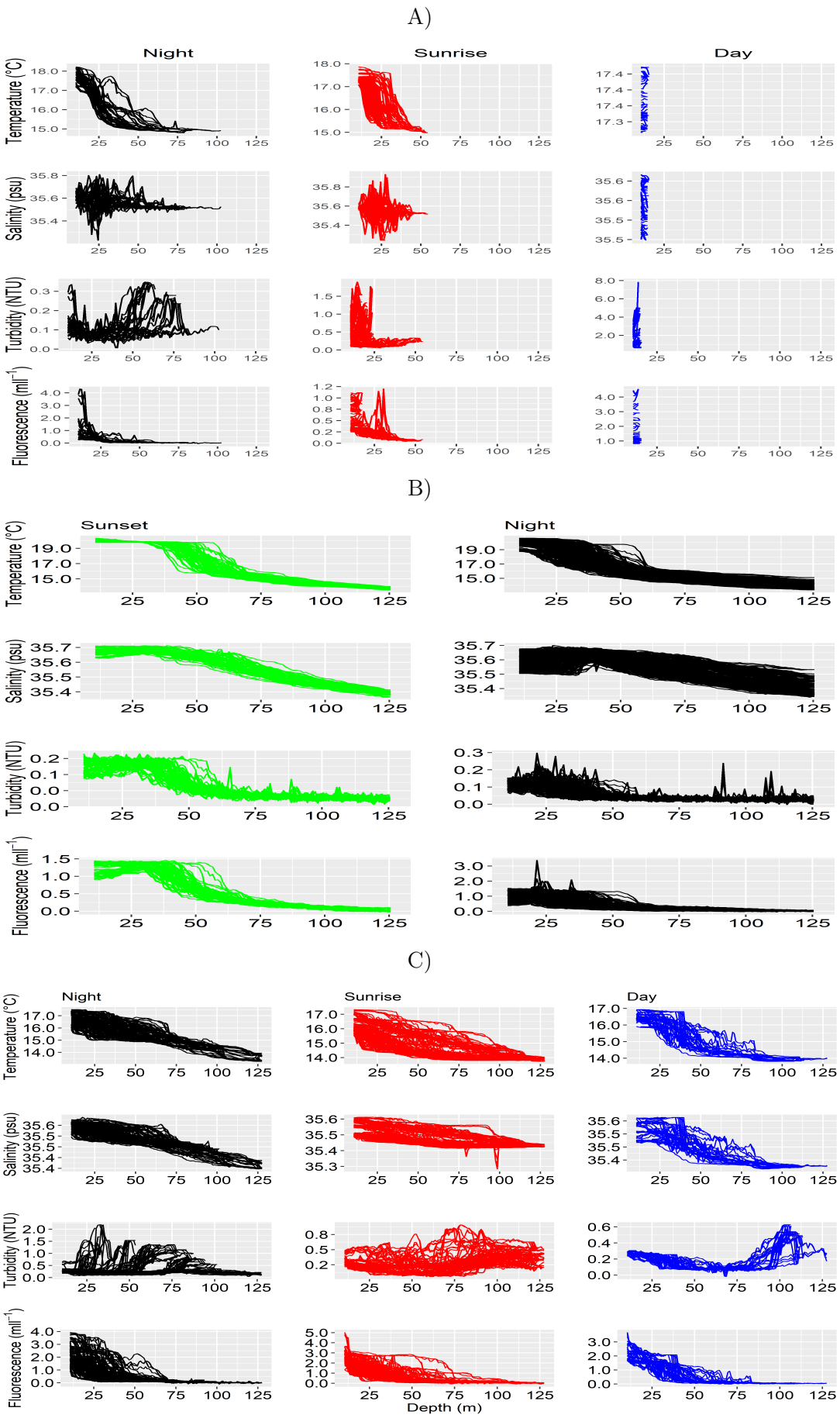
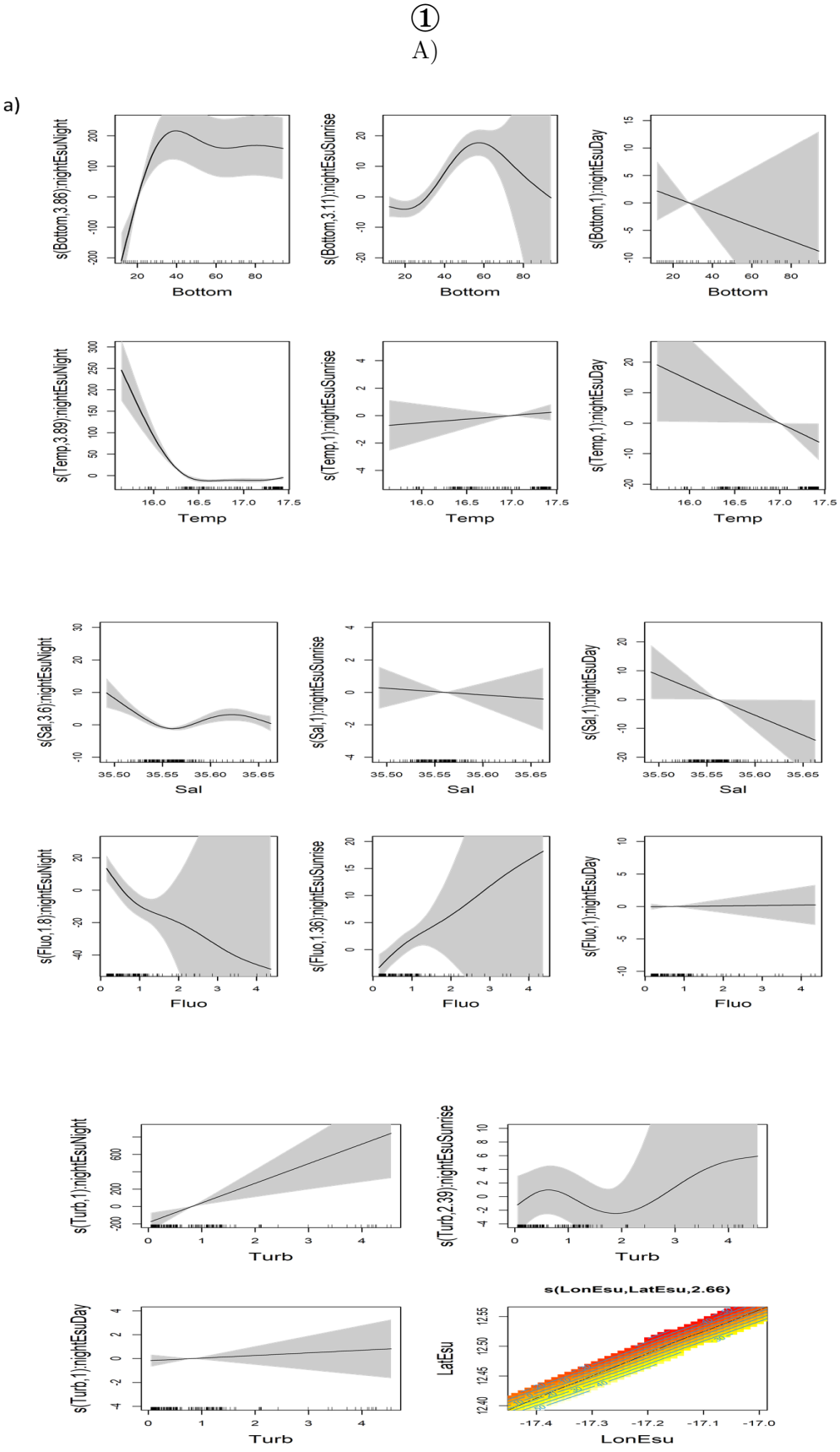
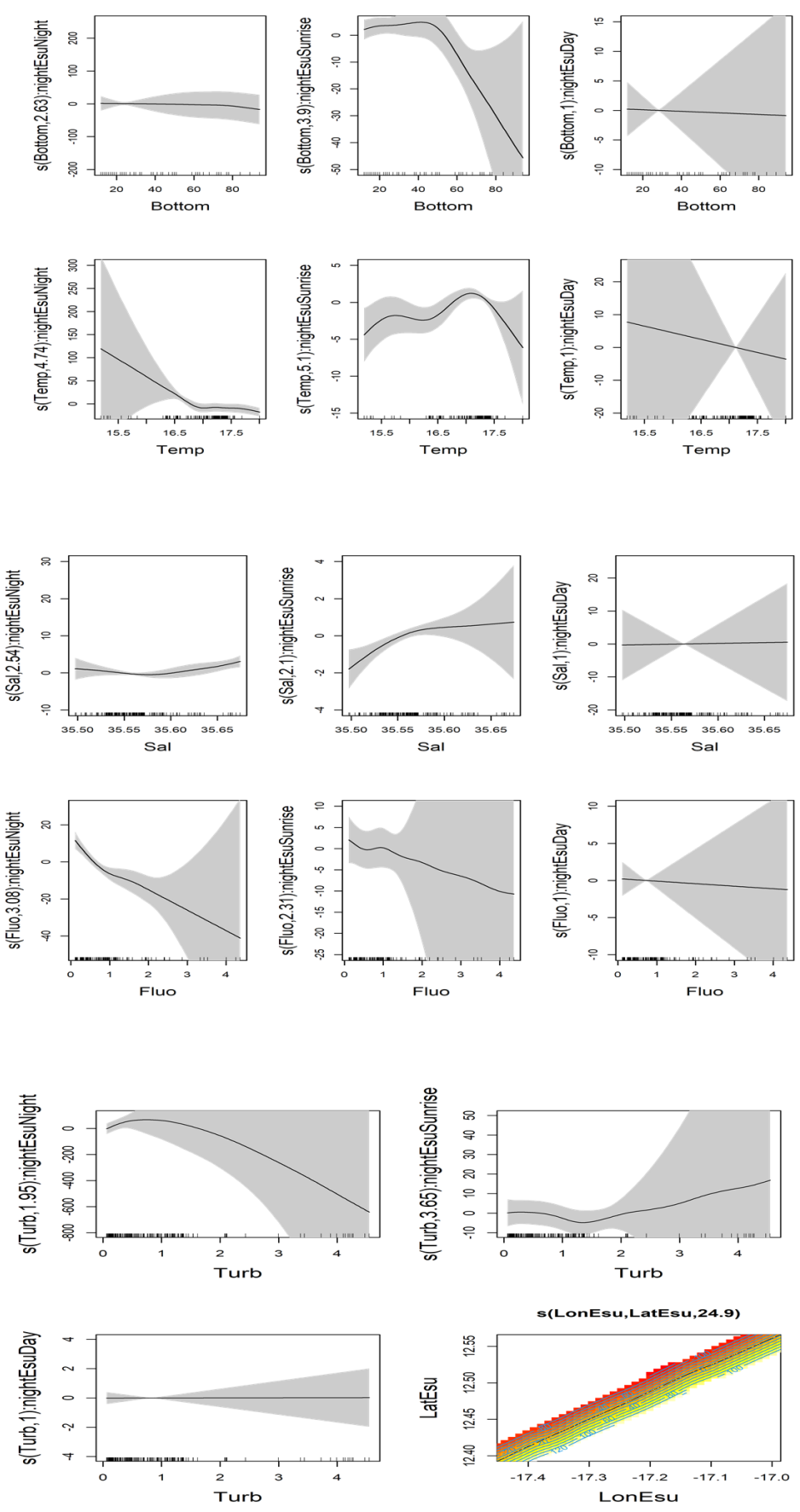


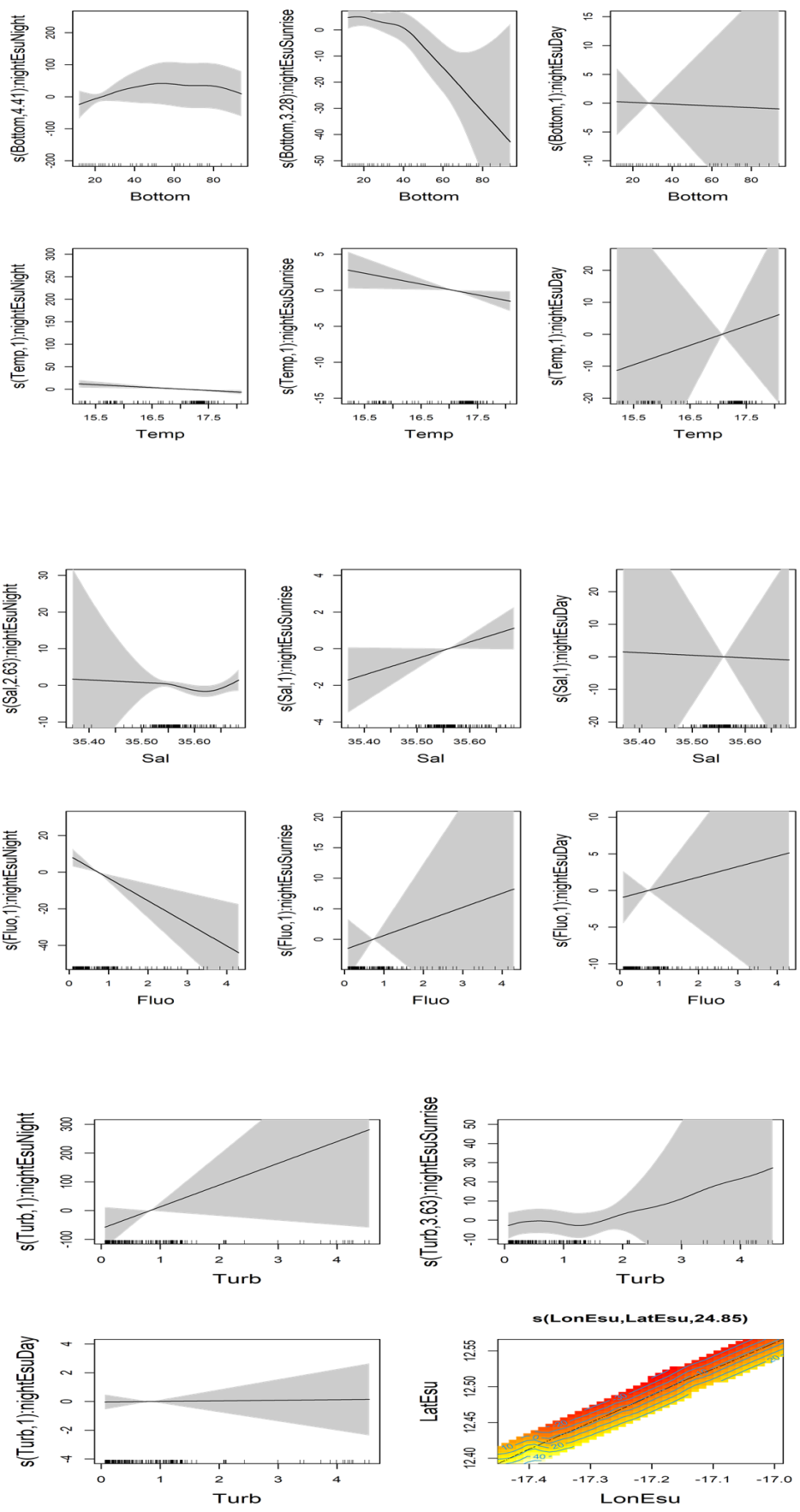
Figure A4: The smooth terms resulting from the spatial GAM (GAMs) model between oceanographic condition parameters (sea temperature, fluorescence, salinity and turbidity) grouped by diel period (the variable is called nightEsu in the model) (day, sunset, night and sunrise) and SSL descriptors: (1) SSL thickness and (2) SSL depth over three different areas (A: southern continental shelf of Senegal; B: southern high sea of Senegal; C: northern Senegal).



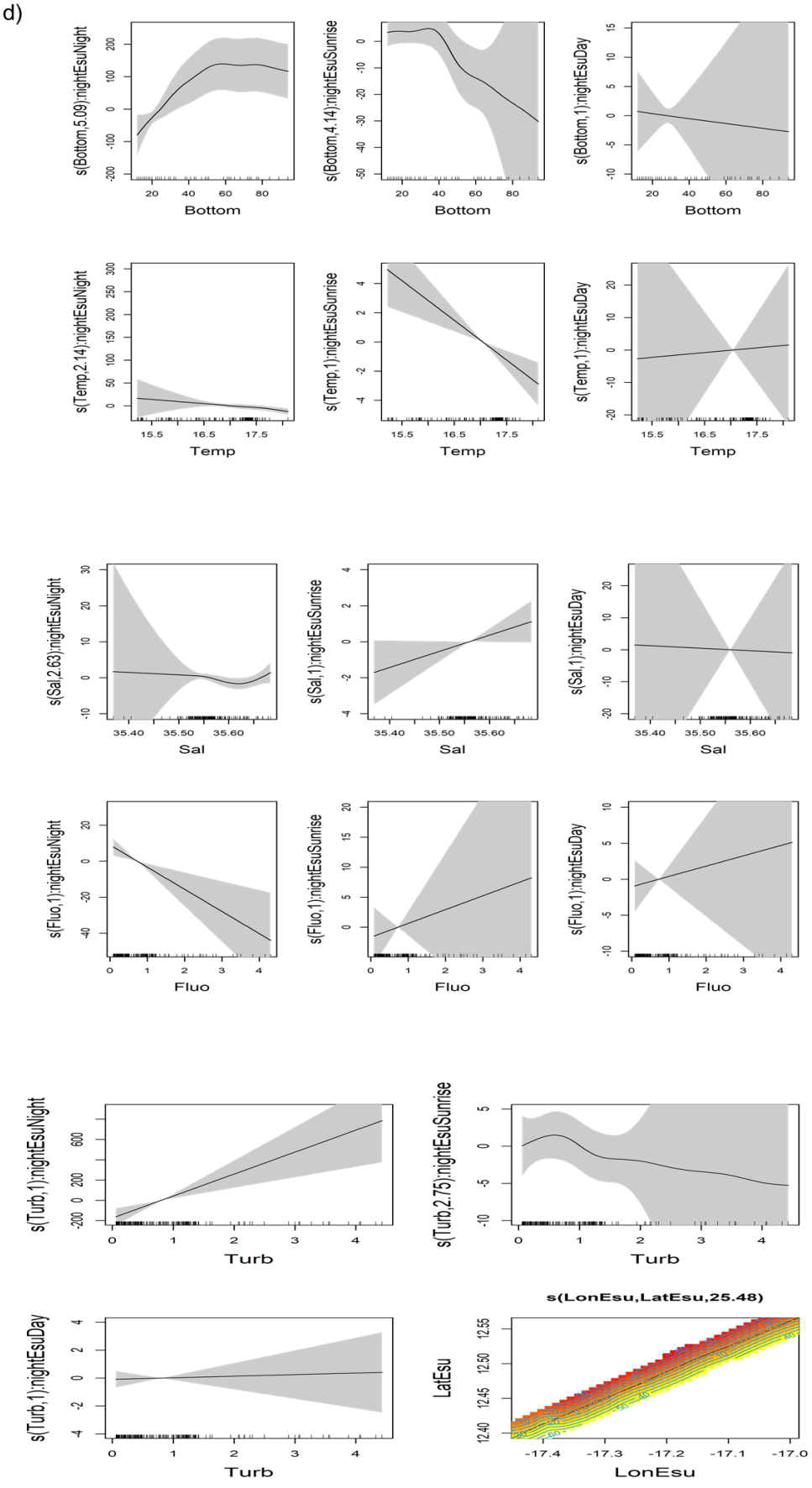
b)


s(Bottom,2.63):nightEsuNight
Bottom
s(Bottom,3.9):nightEsuSunrise
Bottom
s(Bottom,1):nightEsuDay
Bottom
s(Temp,4.74):nightEsuNight
Temp
s(Temp,5.1):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,2.54):nightEsuNight
Sal
s(Sal,2.1):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,3.08):nightEsuNight
Fluo
s(Fluo,2.31):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1.95):nightEsuNight
Turb
s(Turb,3.65):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,24.9)
LatEsu
LonEsu

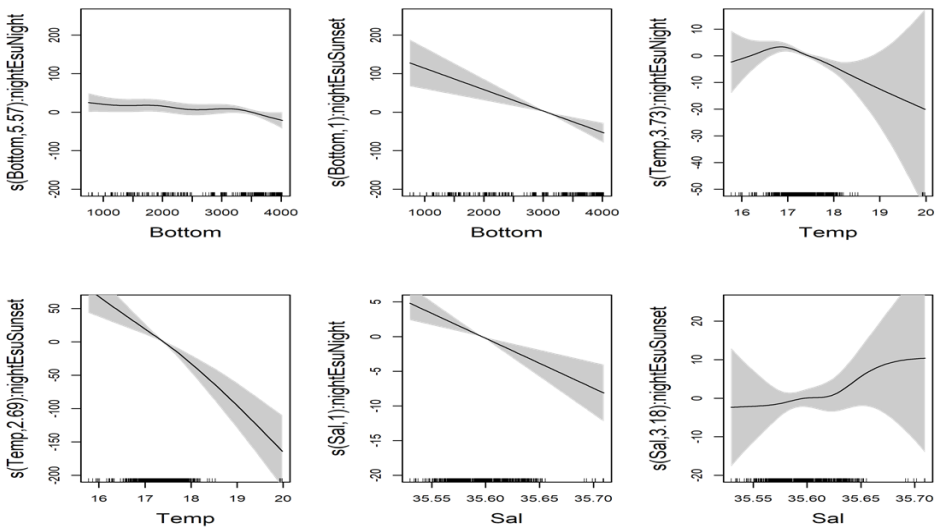
c)


s(Bottom,4.41):nightEsuNight
Bottom
s(Bottom,3.28):nightEsuSunrise
Bottom
s(Bottom,1):nightEsuDay
Bottom
s(Temp,1):nightEsuNight
Temp
s(Temp,1):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,2.63):nightEsuNight
Sal
s(Sal,1):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
Fluo
s(Fluo,1):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
Turb
s(Turb,3.63):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,24.85)
LatEsu
LonEsu

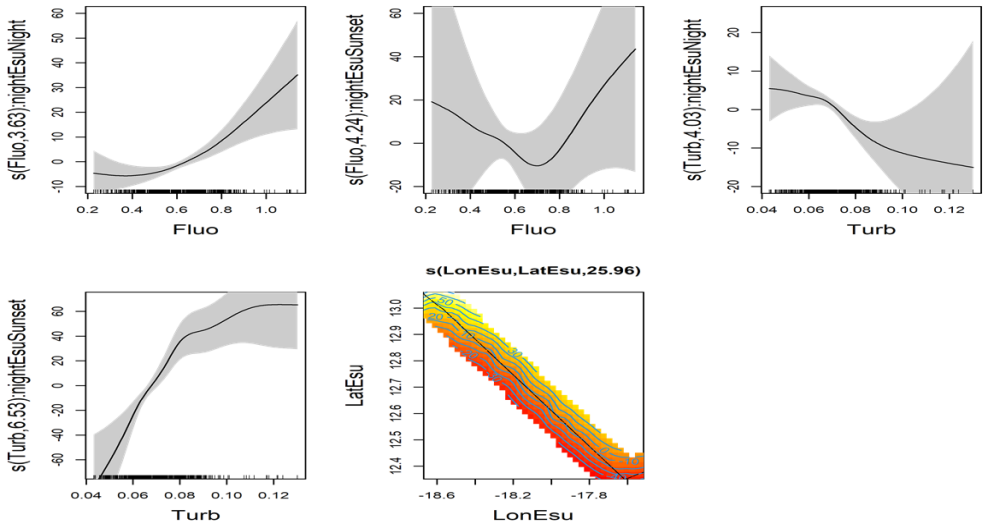
d)


s(Bottom,5.09):nightEsuNight
s(Bottom,4.14):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,2.14):nightEsuNight
s(Temp,1):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp
s(Sal,2.63):nightEsuNight
s(Sal,1):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
s(Fluo,1):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
s(Turb,2.75):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,25.48)
LatEsu
LonEsu

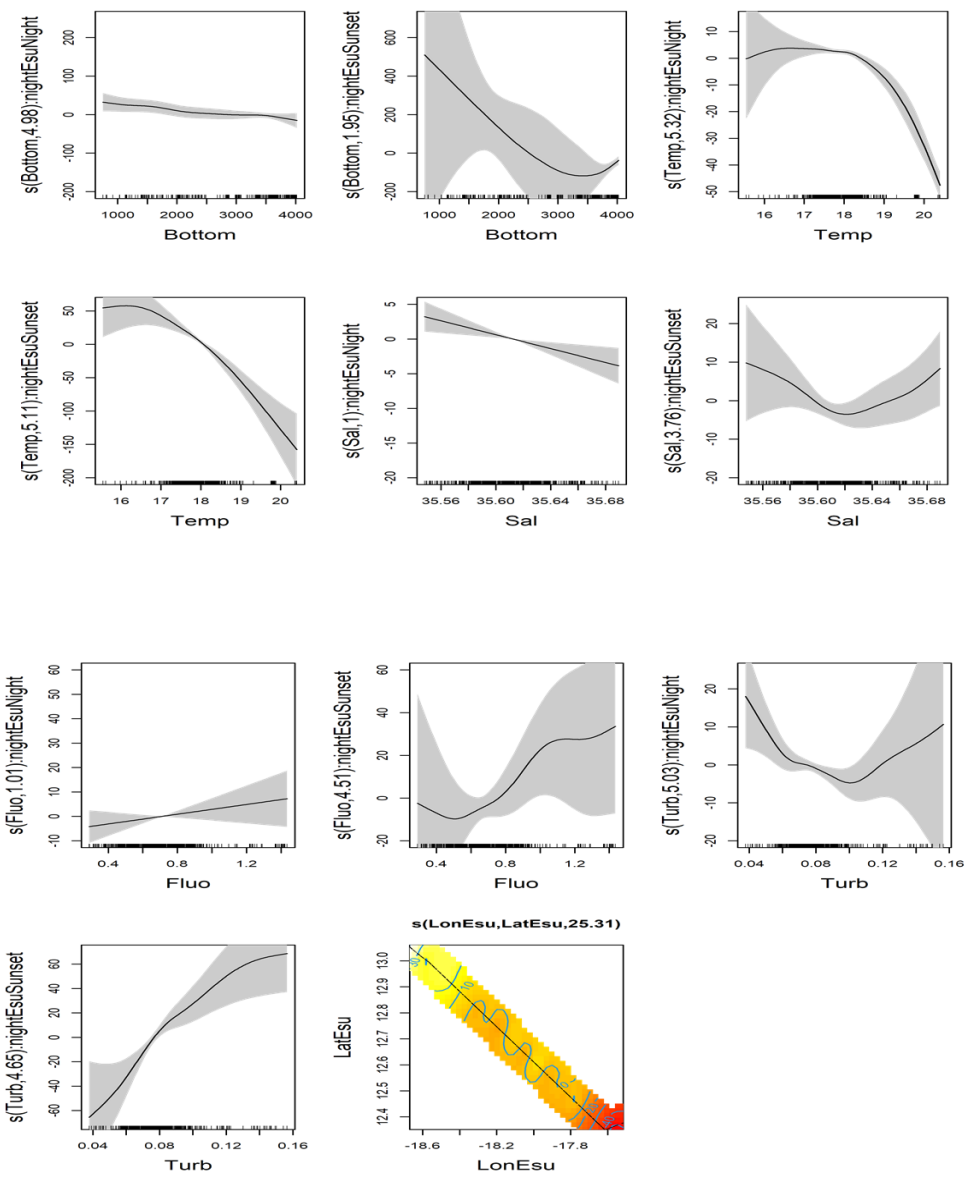
a)


s(Bottom,5.57):nightEsuNight
Bottom
s(Bottom,1):nightEsuSunset
Bottom
s(Temp,3.73):nightEsuNight
Temp
s(Temp,2.69):nightEsuSunset
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,3.18):nightEsuSunset
Sal

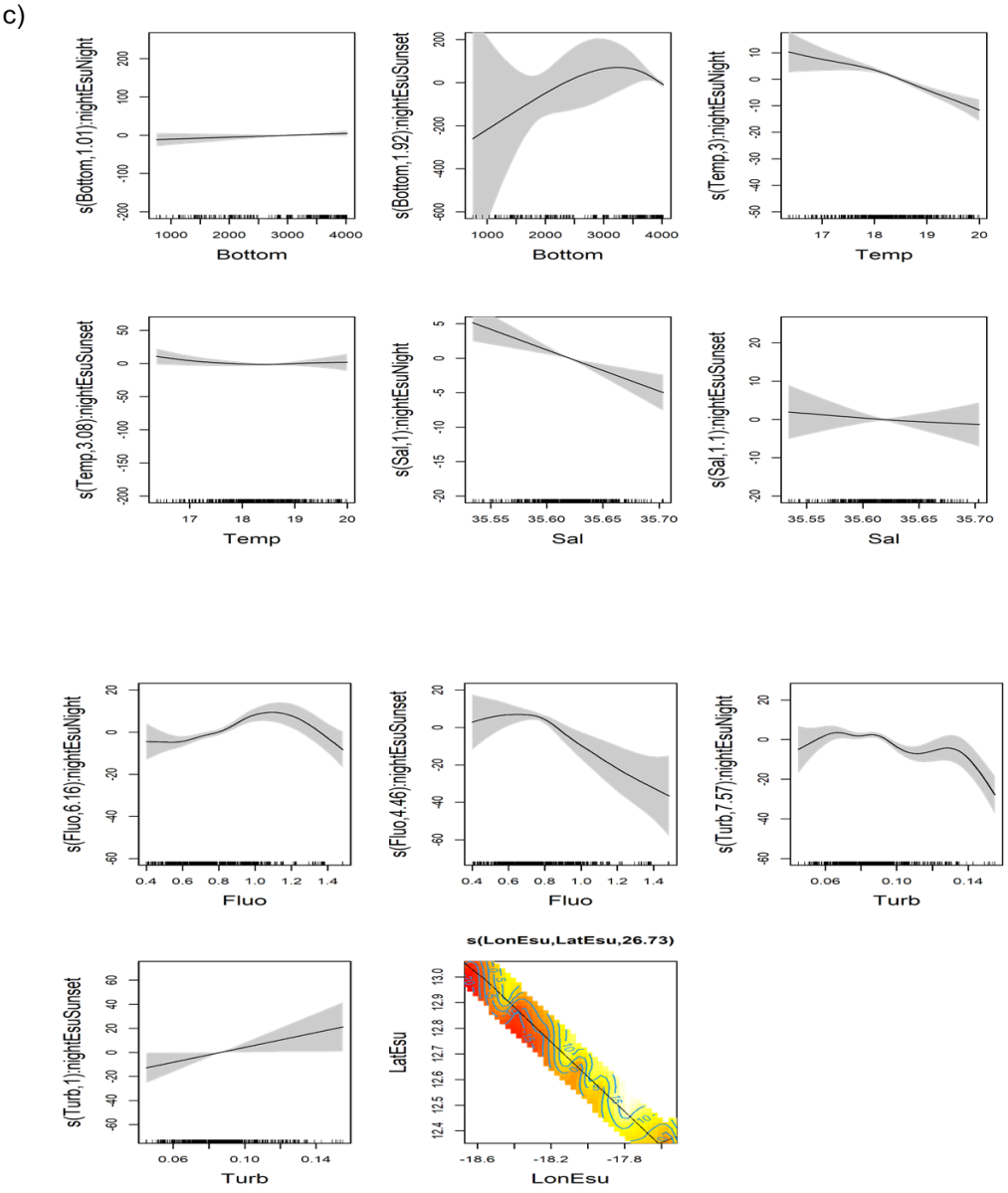
B)


s(Fluo,3.63):nightEsuNight
Fluo
s(Fluo,4.24):nightEsuSunset
Fluo
s(Turb,4.03):nightEsuNight
Turb
s(Turb,6.53):nightEsuSunset
Turb
s(LonEsu,LatEsu,25.96)
LatEsu
LonEsu

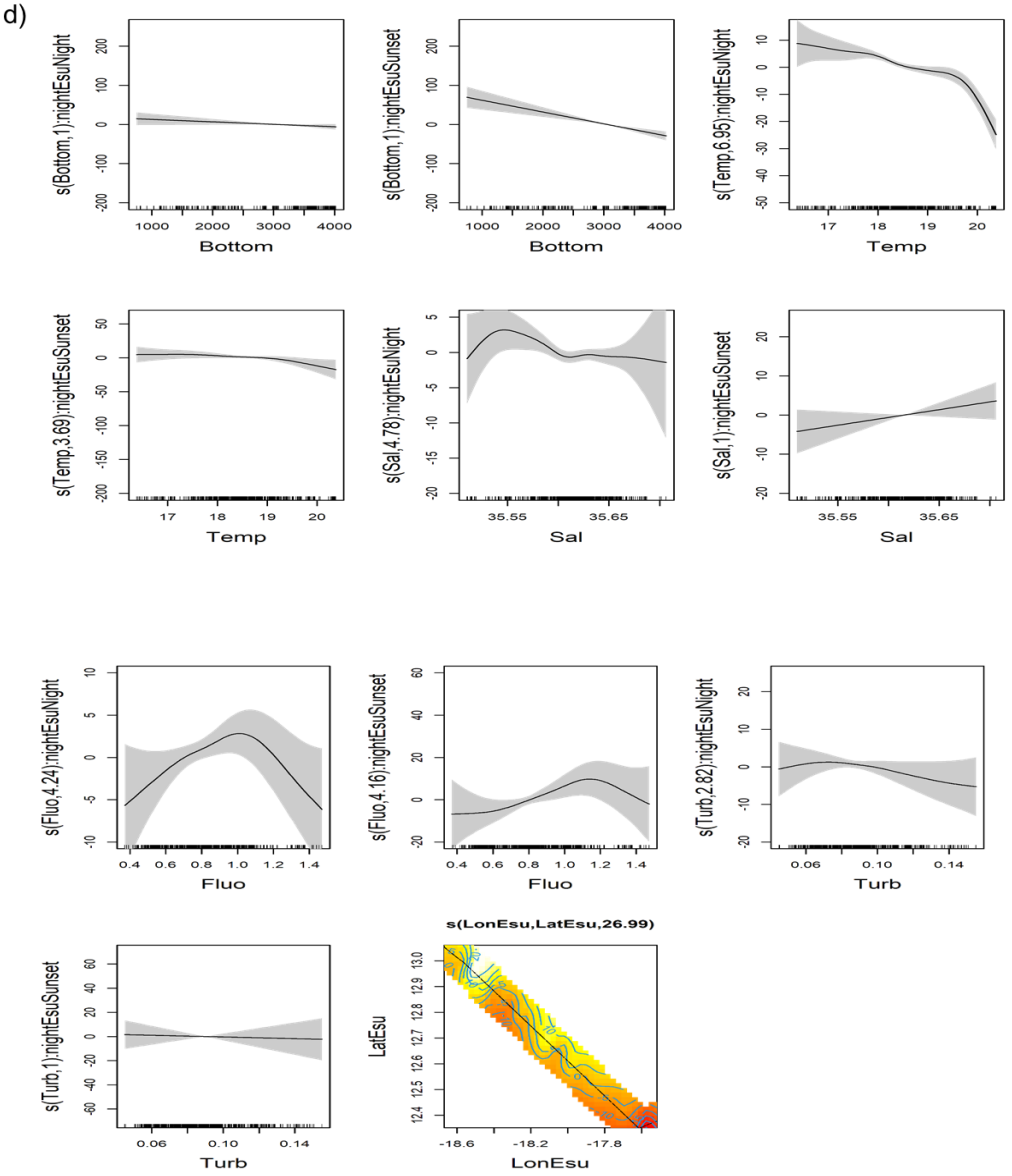
b)


s(Bottom,4.98):nightEsuNight
Bottom
s(Bottom,1.95):nightEsuSunset
Bottom
s(Temp,5.32):nightEsuNight
Temp
s(Temp,5.11):nightEsuSunset
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,3.76):nightEsuSunset
Sal
s(Fluo,1.01):nightEsuNight
Fluo
s(Fluo,4.51):nightEsuSunset
Fluo
s(Turb,5.03):nightEsuNight
Turb
s(Turb,4.65):nightEsuSunset
Turb
s(LonEsu,LatEsu,25.31)
LatEsu
LonEsu

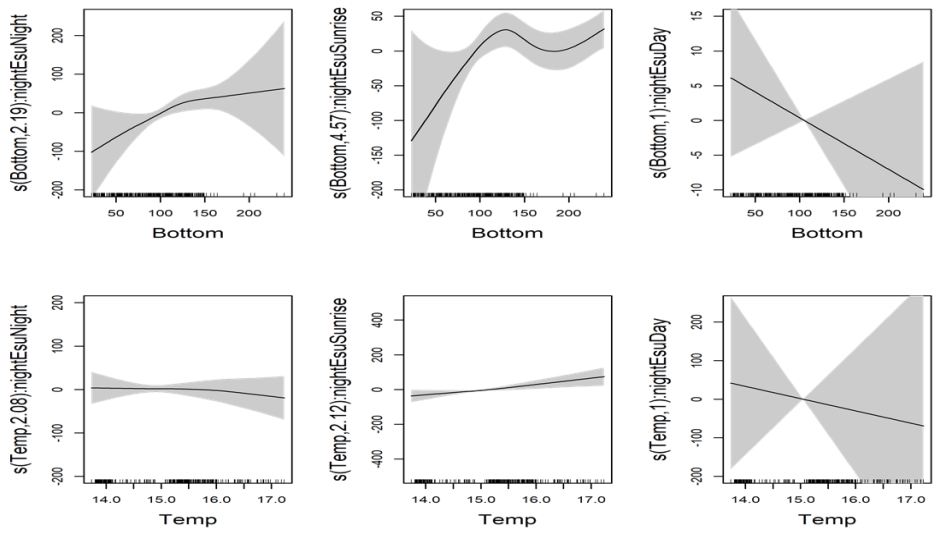
c)


s(Bottom,1.01):nightEsuNight
Bottom
s(Bottom,1.92):nightEsuSunset
Bottom
s(Temp,3):nightEsuNight
Temp
s(Temp,3.08):nightEsuSunset
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,1.1):nightEsuSunset
Sal
s(Fluo,6.16):nightEsuNight
Fluo
s(Fluo,4.46):nightEsuSunset
Fluo
s(Turb,7.57):nightEsuNight
Turb
s(Turb,1):nightEsuSunset
Turb
s(LonEsu,LatEsu,26.73)
LatEsu
LonEsu

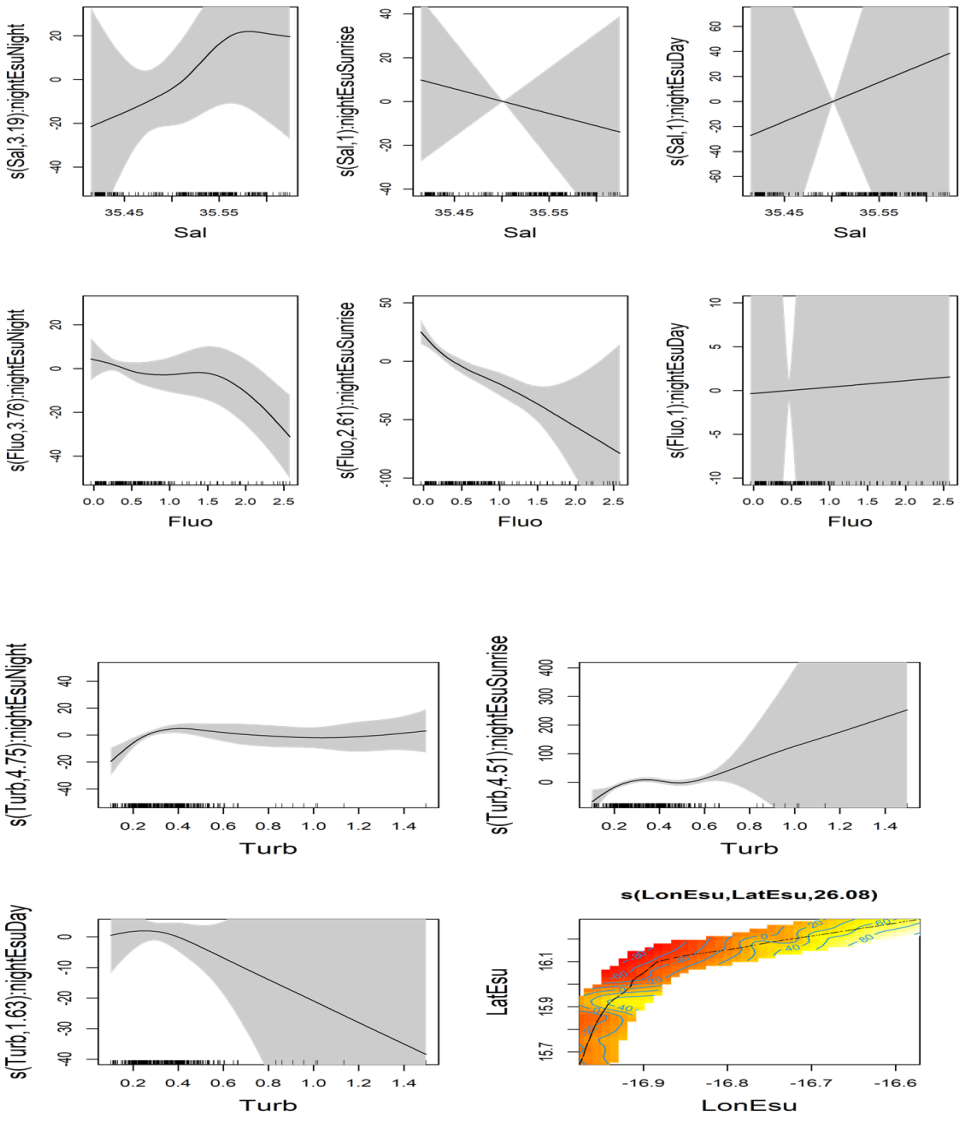
d)


s(Bottom,1):nightEsuNight
Bottom
s(Bottom,1):nightEsuSunset
Bottom
s(Temp,6.95):nightEsuNight
Temp
s(Temp,3.69):nightEsuSunset
Temp
s(Sal,4.78):nightEsuNight
Sal
s(Sal,1):nightEsuSunset
Sal
s(Fluo,4.24):nightEsuNight
Fluo
s(Fluo,4.16):nightEsuSunset
Fluo
s(Turb,2.82):nightEsuNight
Turb
s(Turb,1):nightEsuSunset
Turb
s(LonEsu,LatEsu,26.99)
LatEsu
LonEsu

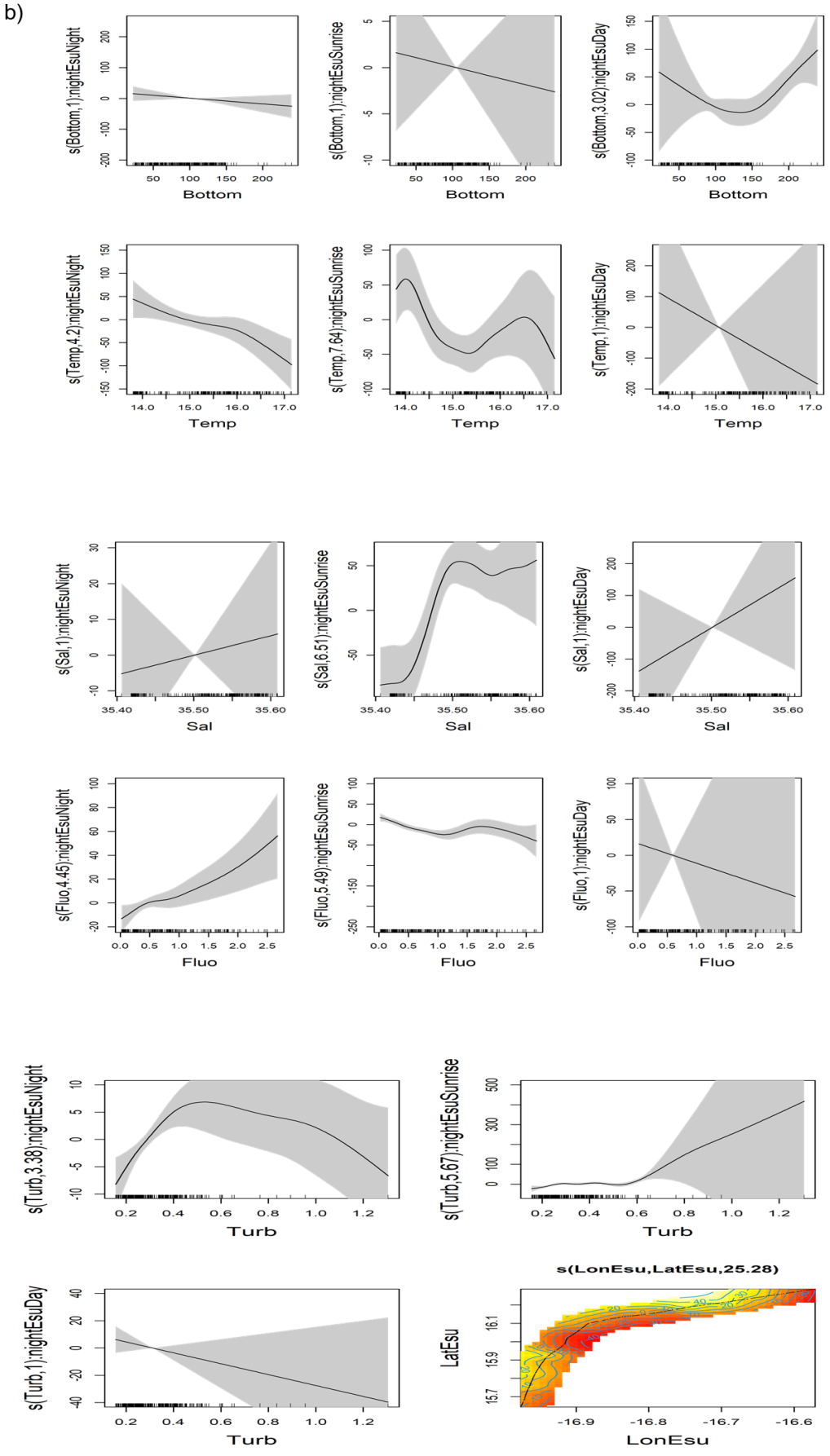
a)


s(Bottom,2.19):nightEsuNight
s(Bottom,4.57):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,2.08):nightEsuNight
s(Temp,2.12):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp

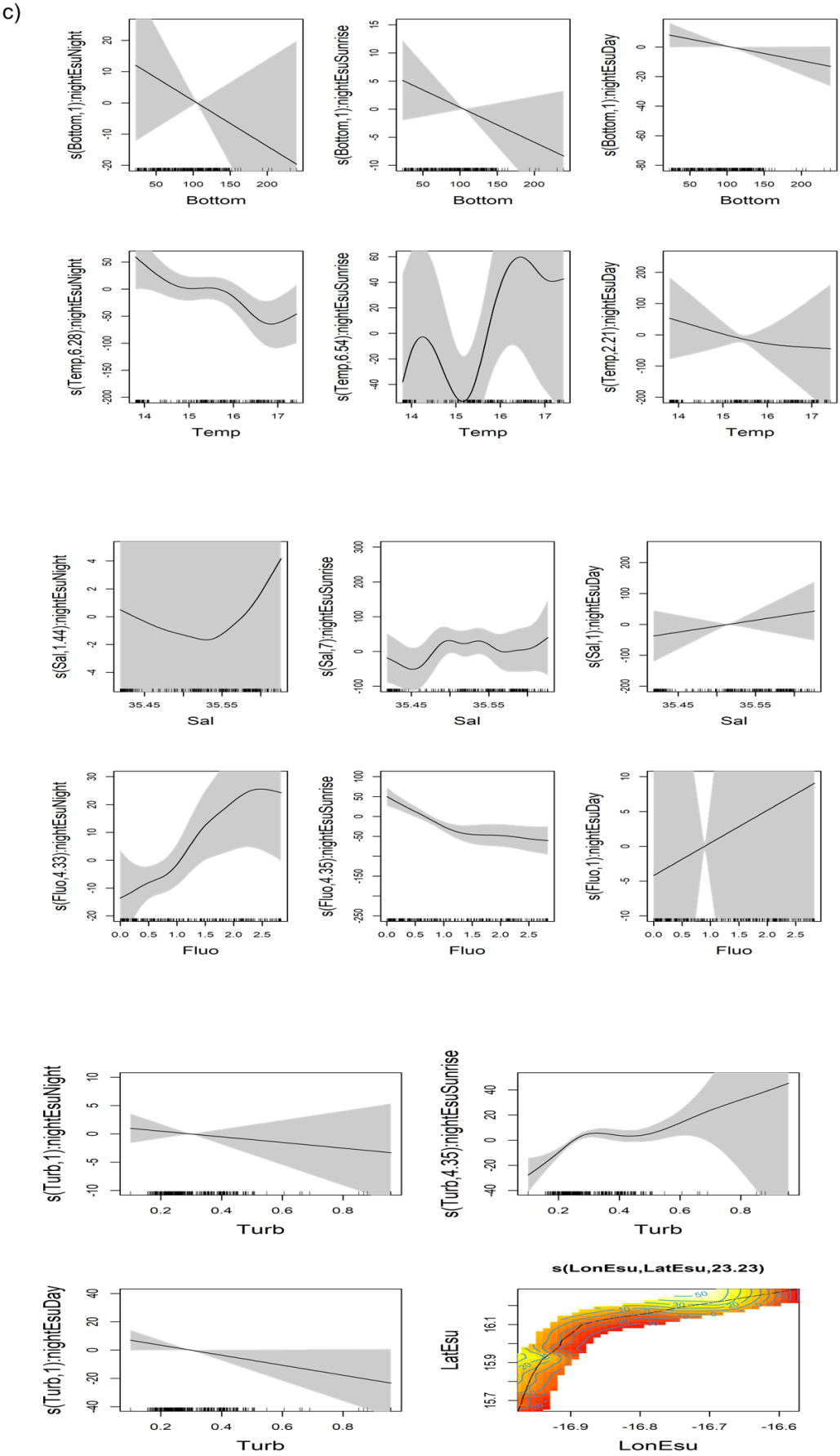
C)


s(Sal,3.19):nightEsuNight
s(Sal,1):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,3.76):nightEsuNight
s(Fluo,2.61):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo
s(Turb,4.75):nightEsuNight
s(Turb,4.51):nightEsuSunrise
Turb
s(Turb,1.63):nightEsuDay
s(LonEsu,LatEsu,26.08)
LatEsu
LonEsu

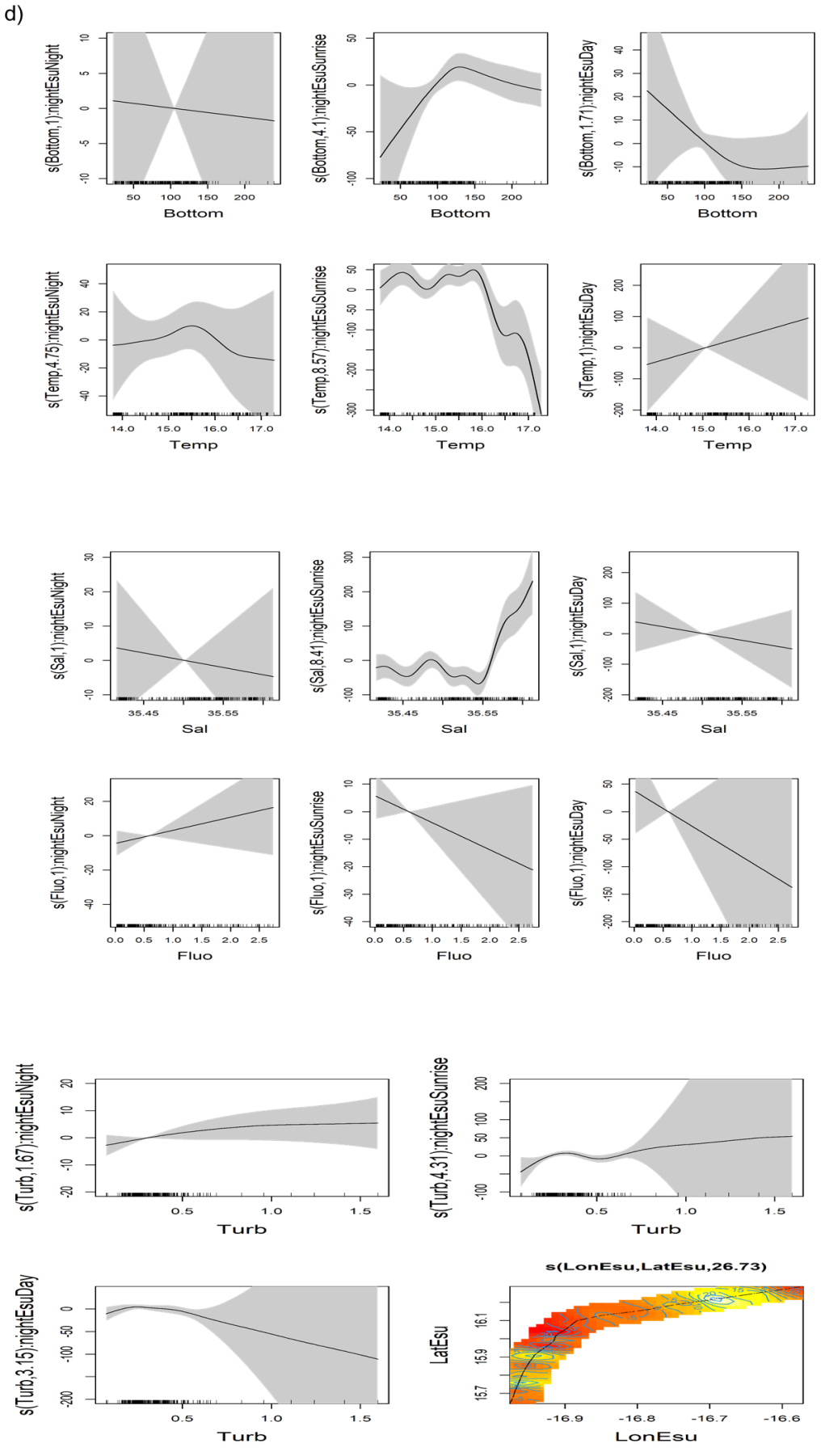
b)


s(Bottom,1):nightEsuNight
Bottom
s(Bottom,1):nightEsuSunrise
Bottom
s(Bottom,3.02):nightEsuDay
Bottom
s(Temp,4.2):nightEsuNight
Temp
s(Temp,7.64):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,6.51):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,4.45):nightEsuNight
Fluo
s(Fluo,5.49):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,3.38):nightEsuNight
Turb
s(Turb,5.67):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,25.28)
LatEsu
LonEsu

c)

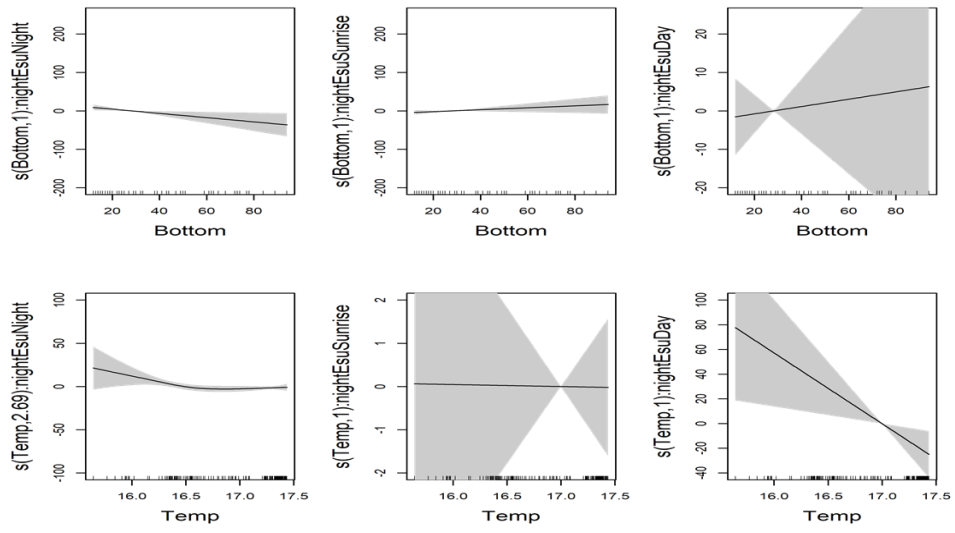

s(Bottom,1):nightEsuNight
s(Bottom,1):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,6.28):nightEsuNight
s(Temp,6.54):nightEsuSunrise
s(Temp,2.21):nightEsuDay
Temp
s(Sal,1.44):nightEsuNight
s(Sal,7):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,4.33):nightEsuNight
s(Fluo,4.35):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
s(Turb,4.35):nightEsuSunrise
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,23.23)
LatEsu
LonEsu

d)

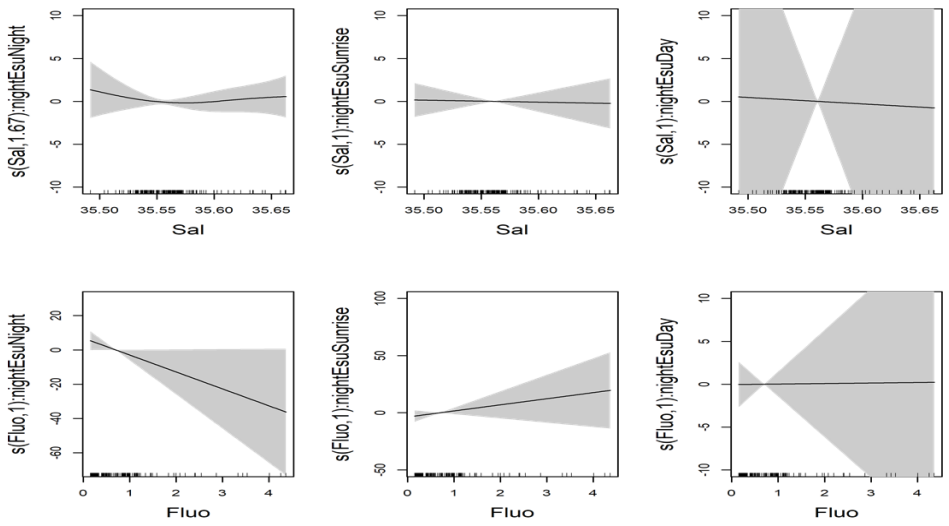

s(Bottom,1):nightEsuNight
Bottom
s(Bottom,4.1):nightEsuSunrise
Bottom
s(Bottom,1.71):nightEsuDay
Bottom
s(Temp,4.75):nightEsuNight
Temp
s(Temp,8.57):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,8.41):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
Fluo
s(Fluo,1):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1.67):nightEsuNight
Turb
s(Turb,4.31):nightEsuSunrise
Turb
s(Turb,3.15):nightEsuDay
Turb
s(LonEsu,LatEsu,26.73)
LatEsu
LonEsu

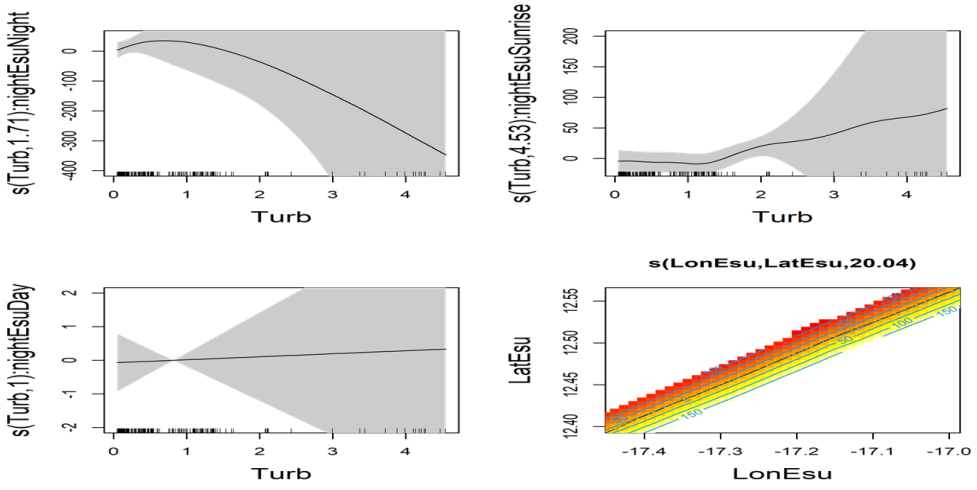
②

a)

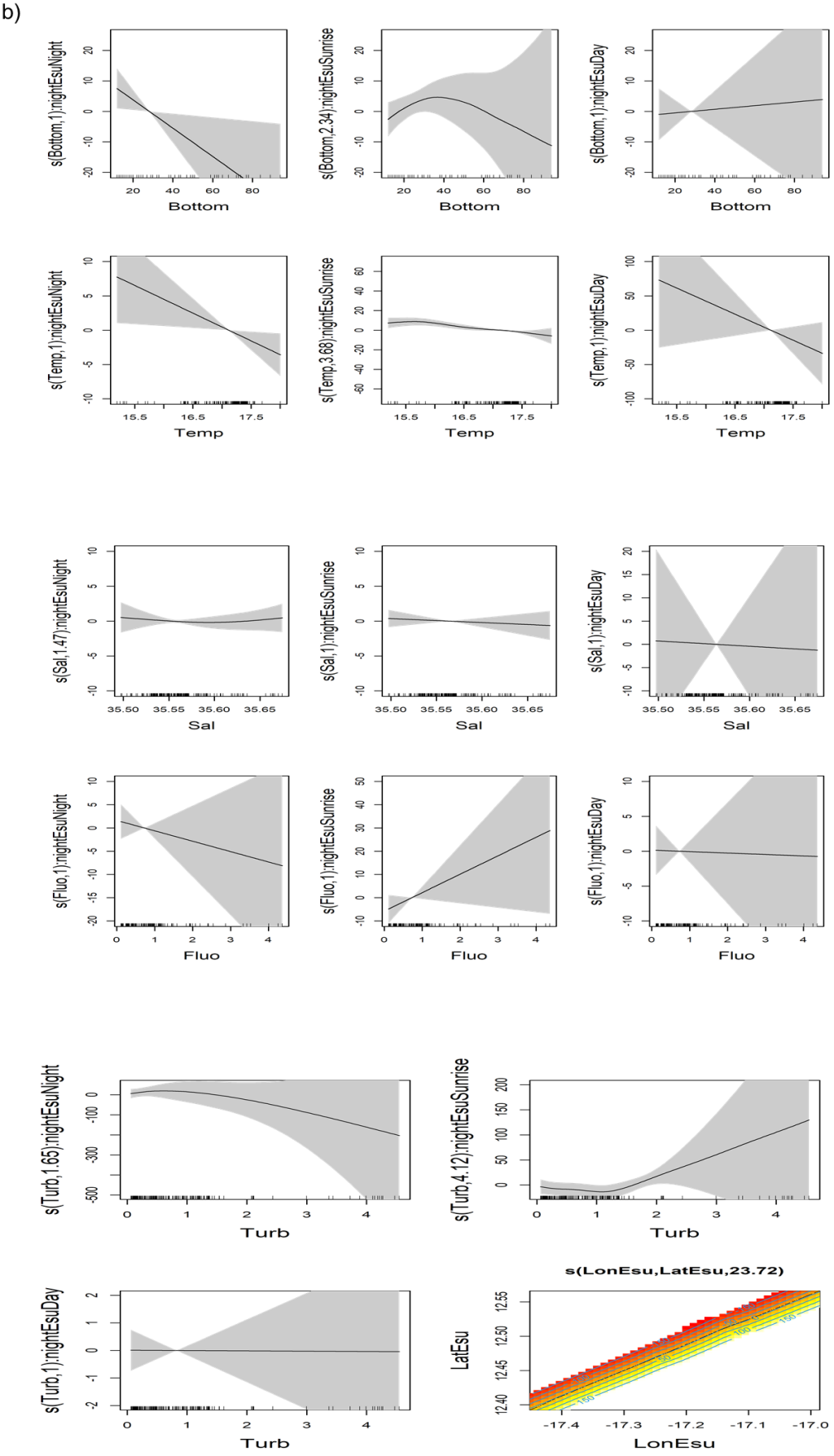

s(Bottom,1):nightEsuNight
s(Bottom,1):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,2.69):nightEsuNight
s(Temp,1):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp

A)

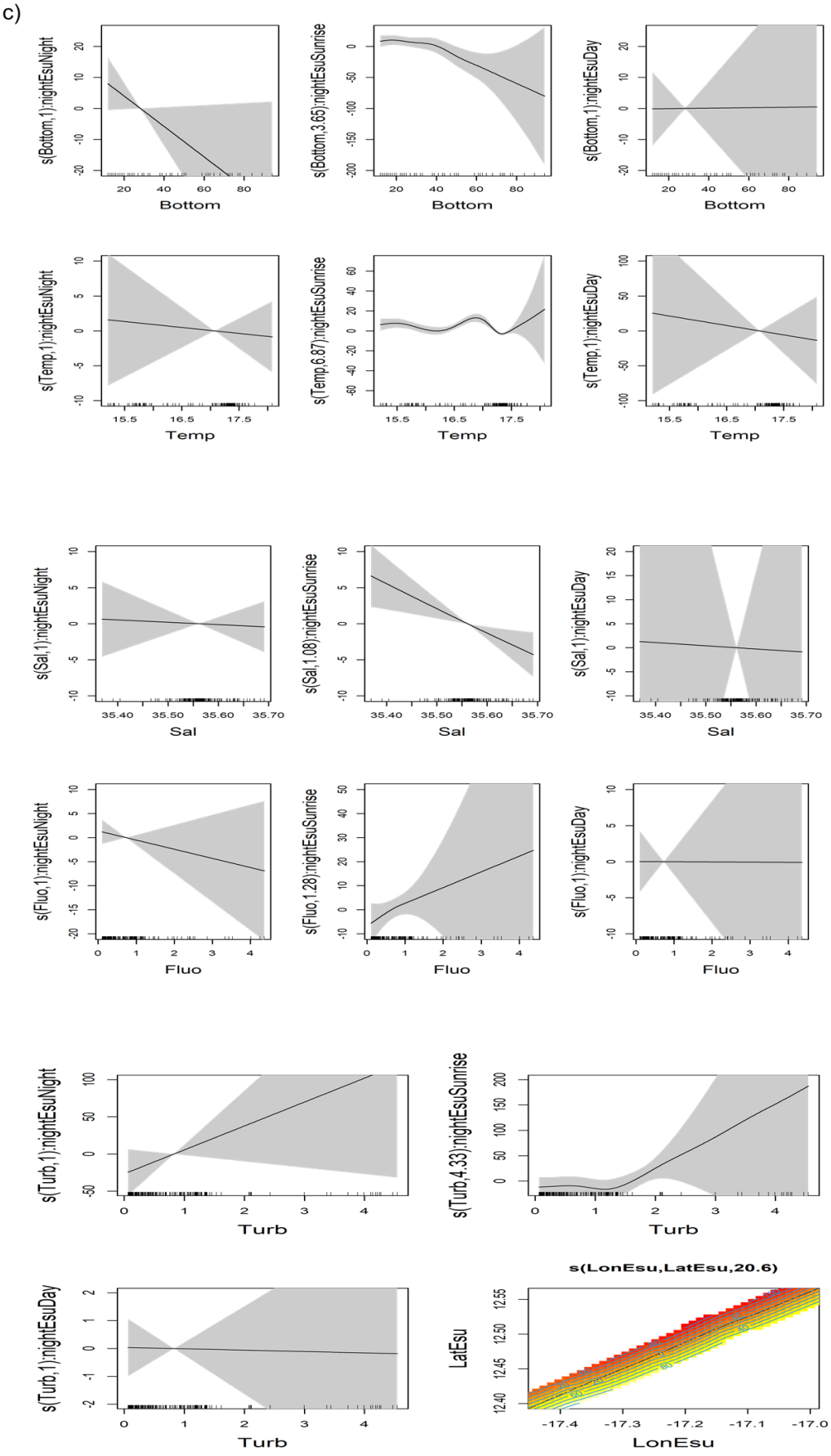

s(Sal,1.67):nightEsuNight
s(Sal,1):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
s(Fluo,1):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo


s(Turb,1.71):nightEsuNight
s(Turb,4.53):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
s(LonEsu,LatEsu,20.04)
LatEsu
LonEsu

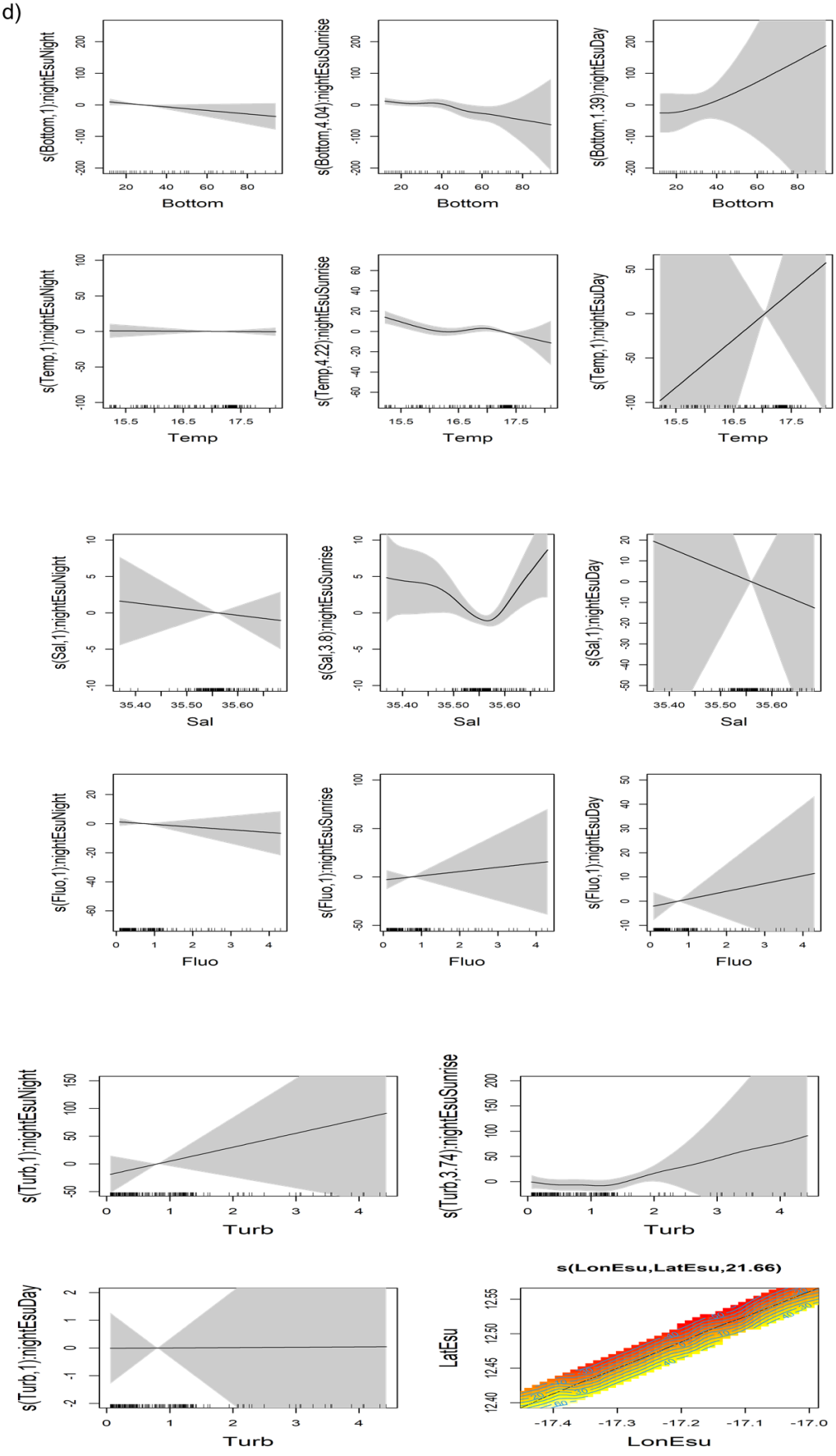
b)


s(Bottom,1):nightEsuNight
s(Bottom,2.34):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,1):nightEsuNight
s(Temp,3.68):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp
s(Sal,1.47):nightEsuNight
s(Sal,1):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
s(Fluo,1):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1.65):nightEsuNight
s(Turb,4.12):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
s(LonEsu,LatEsu,23.72)
LatEsu
LonEsu

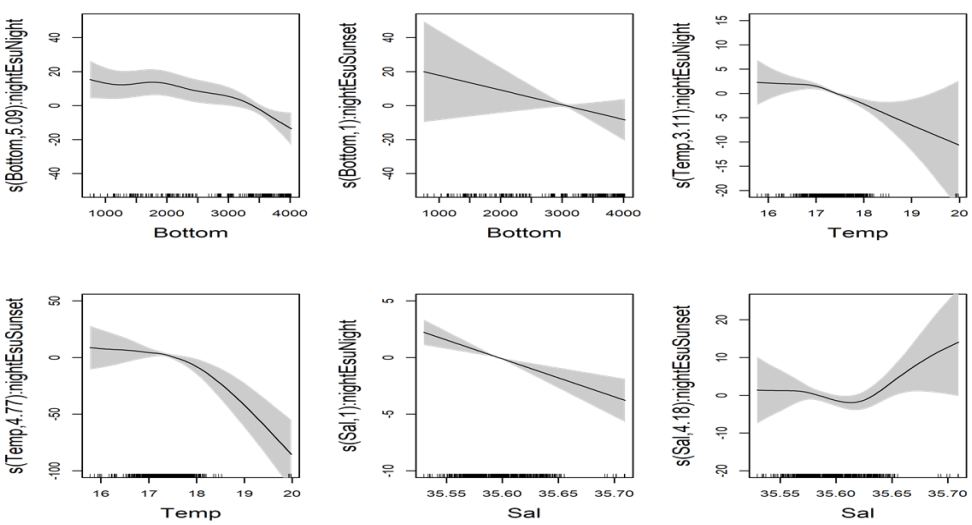
c)


s(Bottom,1):nightEsuNight
Bottom
s(Bottom,3.65):nightEsuSunrise
Bottom
s(Bottom,1):nightEsuDay
Bottom
s(Temp,1):nightEsuNight
Temp
s(Temp,6.87):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,1.08):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
Fluo
s(Fluo,1.28):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
Turb
s(Turb,4.33):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,20.6)
LatEsu
LonEsu

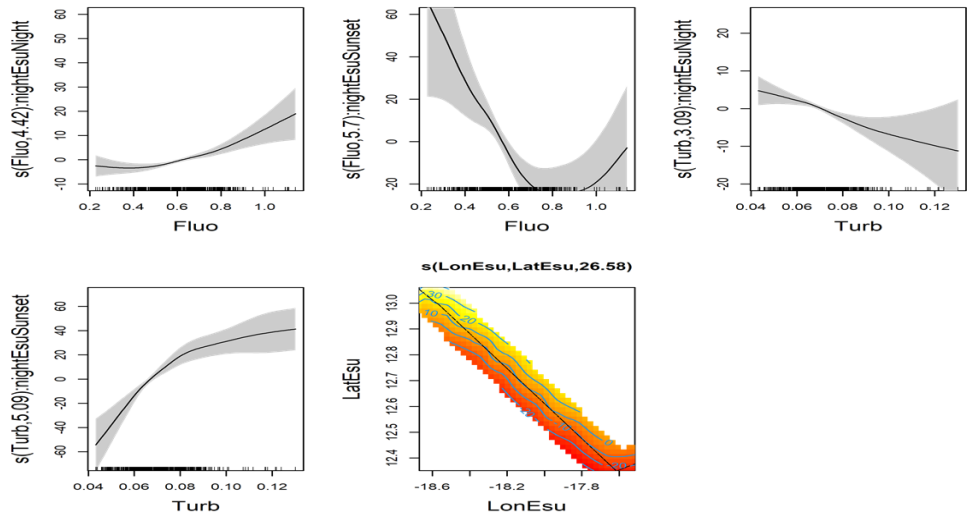
d)


s(Bottom,1):nightEsuNight
s(Bottom,4.04):nightEsuSunrise
s(Bottom,1.39):nightEsuDay
Bottom
s(Temp,1):nightEsuNight
s(Temp,4.22):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp
s(Sal,1):nightEsuNight
s(Sal,3.8):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
s(Fluo,1):nightEsuSunrise
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
s(Turb,3.74):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
s(LonEsu,LatEsu,21.66)
LatEsu
LonEsu

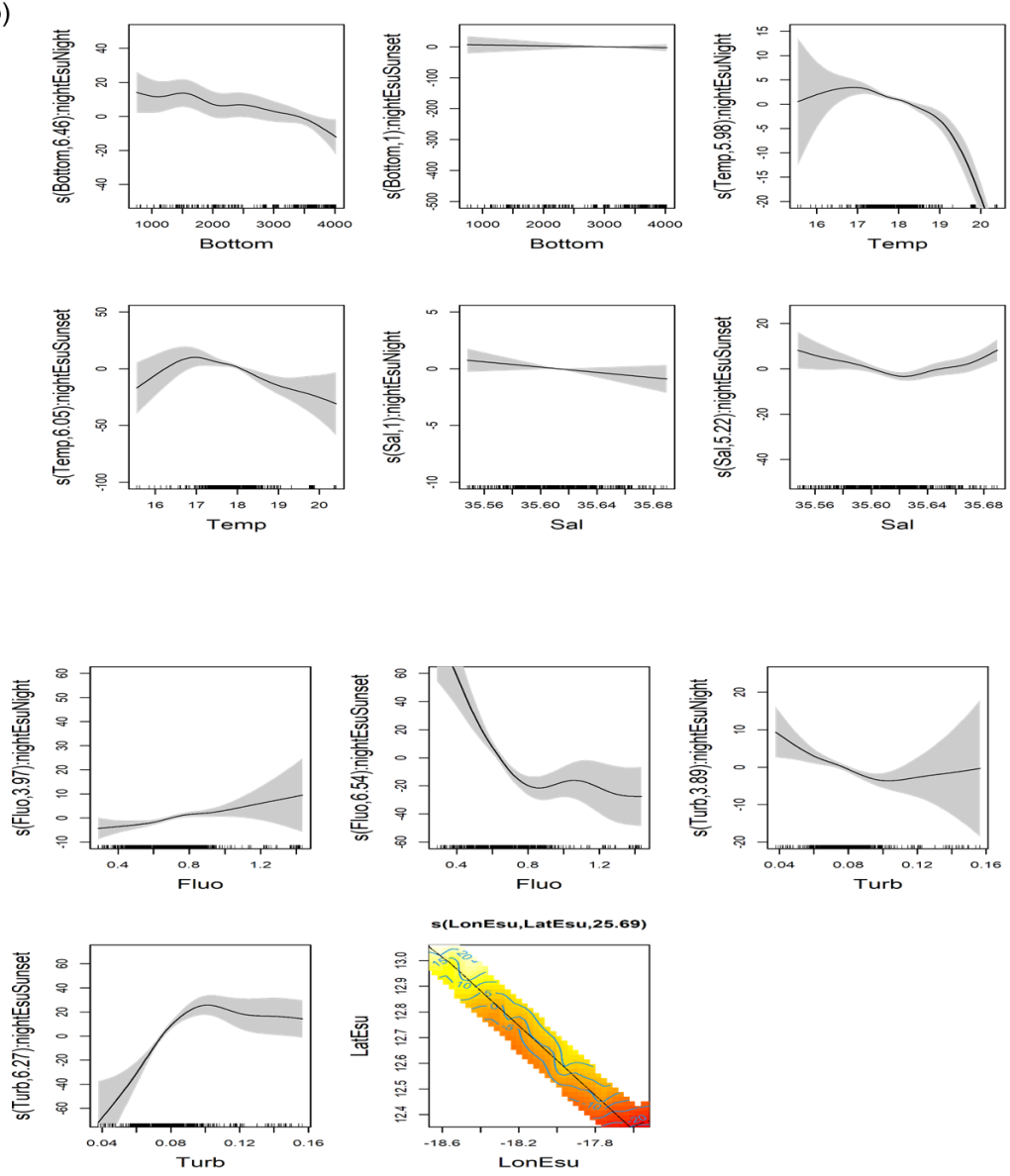
a)


s(Bottom,5.09):nightEsuNight
Bottom
s(Bottom,1):nightEsuSunset
Bottom
s(Temp,3.11):nightEsuNight
Temp
s(Temp,4.77):nightEsuSunset
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,4.18):nightEsuSunset
Sal

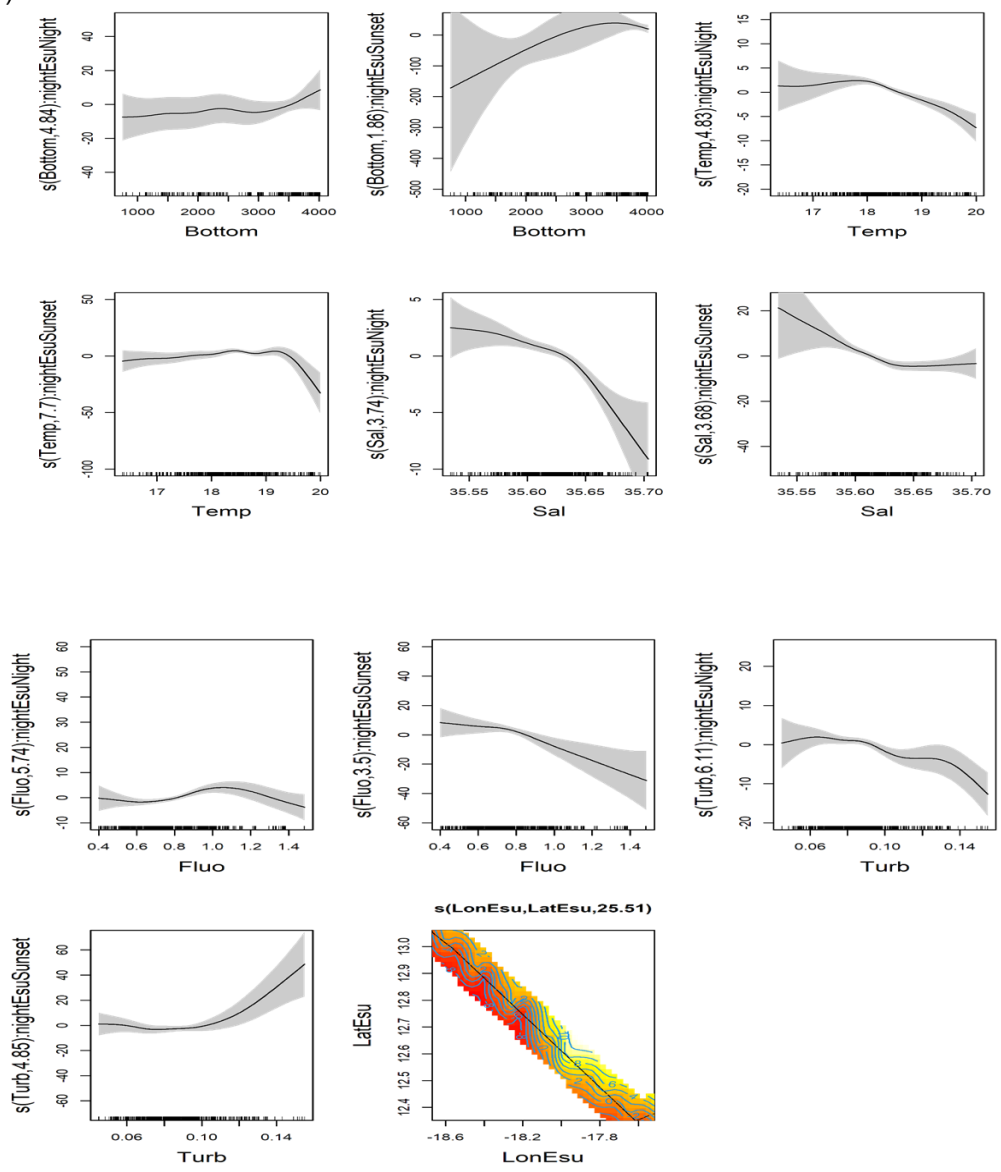
B)


s(Fluo,4.42):nightEsuNight
Fluo
s(Fluo,5.7):nightEsuSunset
Fluo
s(Turb,3.09):nightEsuNight
Turb
s(Turb,5.09):nightEsuSunset
Turb
s(LonEsu,LatEsu,26.58)
LatEsu
LonEsu

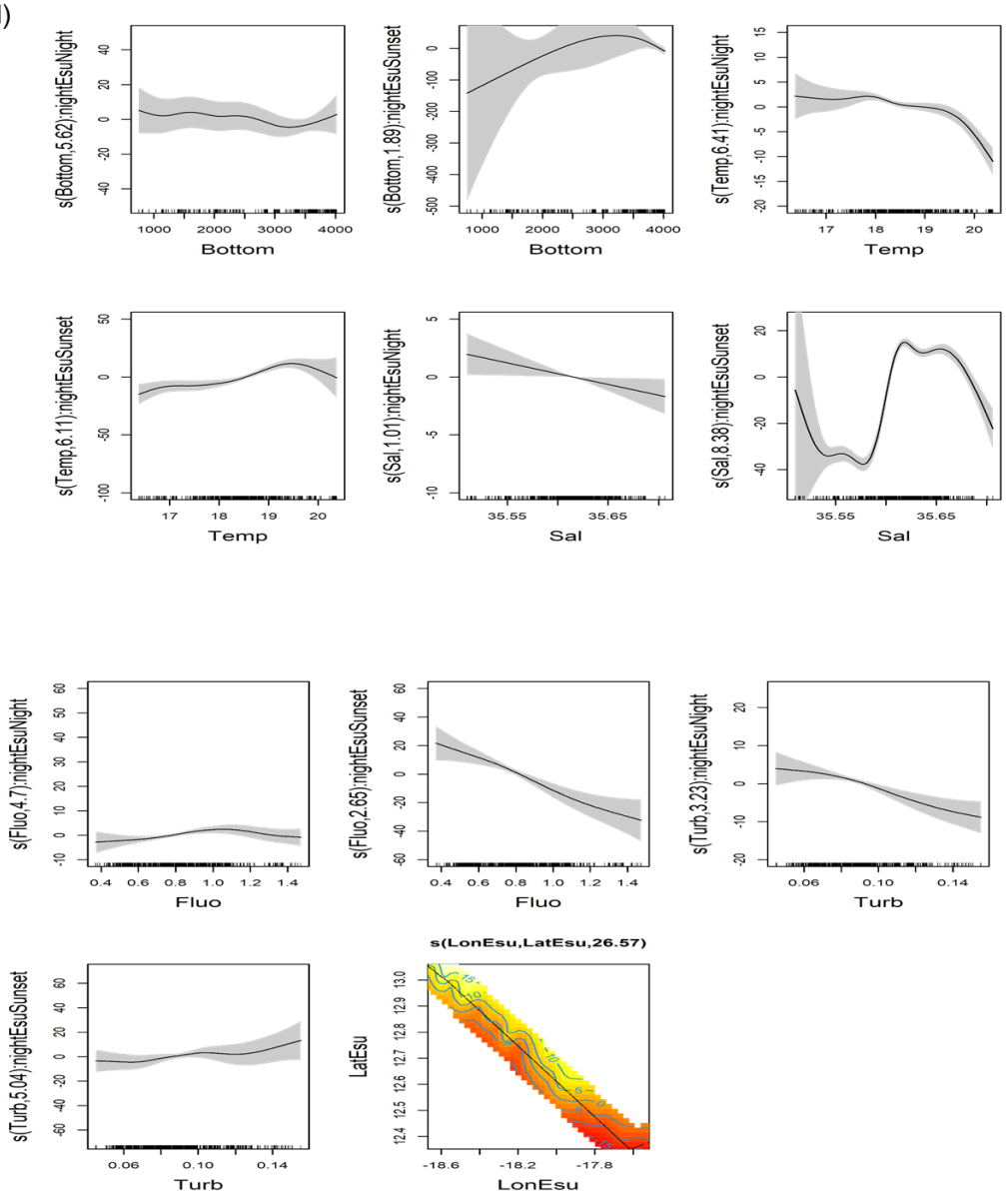
b)


s(Bottom,6.46):nightEsuNight
Bottom
s(Bottom,1):nightEsuSunset
Bottom
s(Temp,5.98):nightEsuNight
Temp
s(Temp,6.05):nightEsuSunset
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,5.22):nightEsuSunset
Sal
s(Fluo,3.97):nightEsuNight
Fluo
s(Fluo,6.54):nightEsuSunset
Fluo
s(Turb,3.89):nightEsuNight
Turb
s(Turb,6.27):nightEsuSunset
Turb
s(LonEsu,LatEsu,25.69)
LatEsu
LonEsu

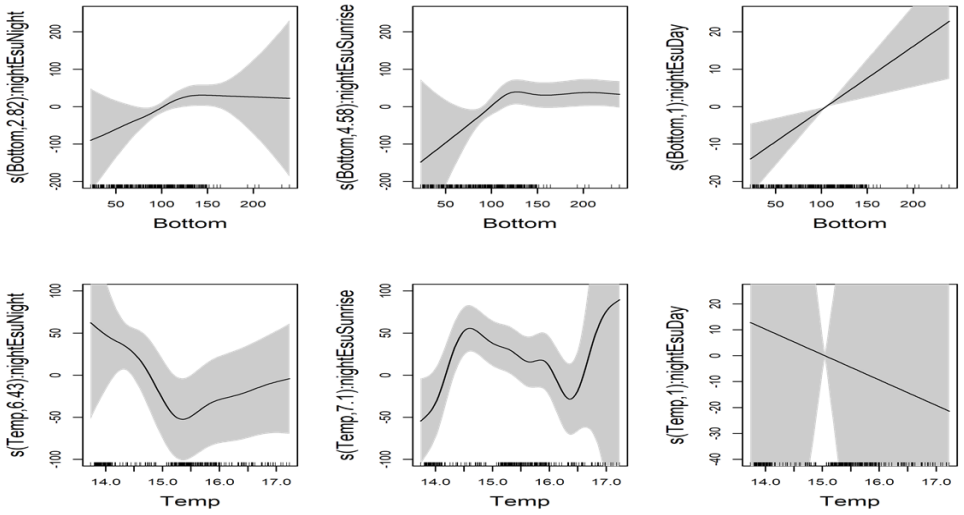
c)


s(Bottom,4.84):nightEsuNight
Bottom
s(Bottom,1.86):nightEsuSunset
Bottom
s(Temp,4.83):nightEsuNight
Temp
s(Temp,7.7):nightEsuSunset
Temp
s(Sal,3.74):nightEsuNight
Sal
s(Sal,3.68):nightEsuSunset
Sal
s(Fluo,5.74):nightEsuNight
Fluo
s(Fluo,3.5):nightEsuSunset
Fluo
s(Turb,6.11):nightEsuNight
Turb
s(Turb,4.85):nightEsuSunset
Turb
s(LonEsu,LatEsu,25.51)
LatEsu
LonEsu

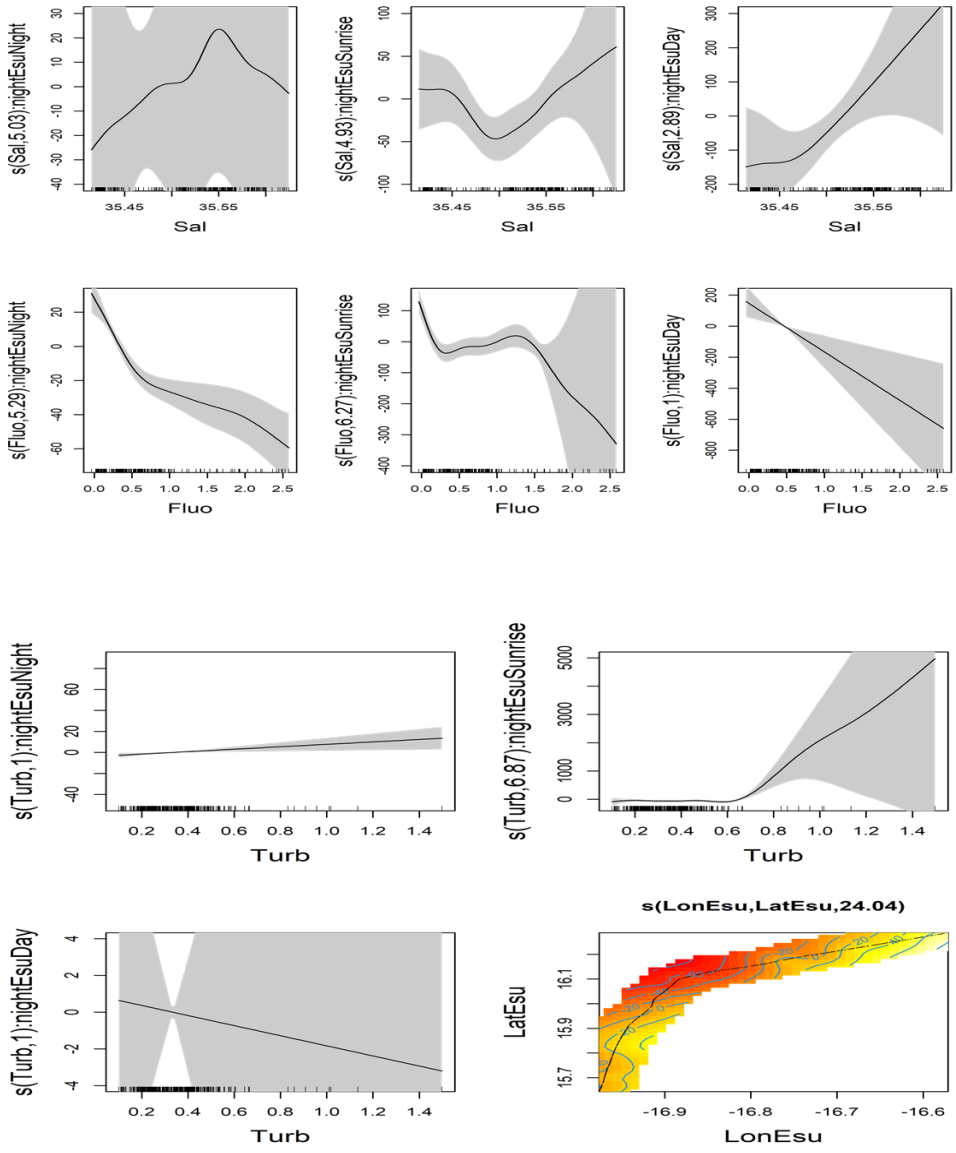
d)


s(Bottom,5.62):nightEsuNight
Bottom
s(Bottom,1.89):nightEsuSunset
Bottom
s(Temp,6.41):nightEsuNight
Temp
s(Temp,6.11):nightEsuSunset
Temp
s(Sal,1.01):nightEsuNight
Sal
s(Sal,8.38):nightEsuSunset
Sal
s(Fluo,4.7):nightEsuNight
Fluo
s(Fluo,2.65):nightEsuSunset
Fluo
s(Turb,3.23):nightEsuNight
Turb
s(Turb,5.04):nightEsuSunset
Turb
s(LonEsu,LatEsu,26.57)
LatEsu
LonEsu

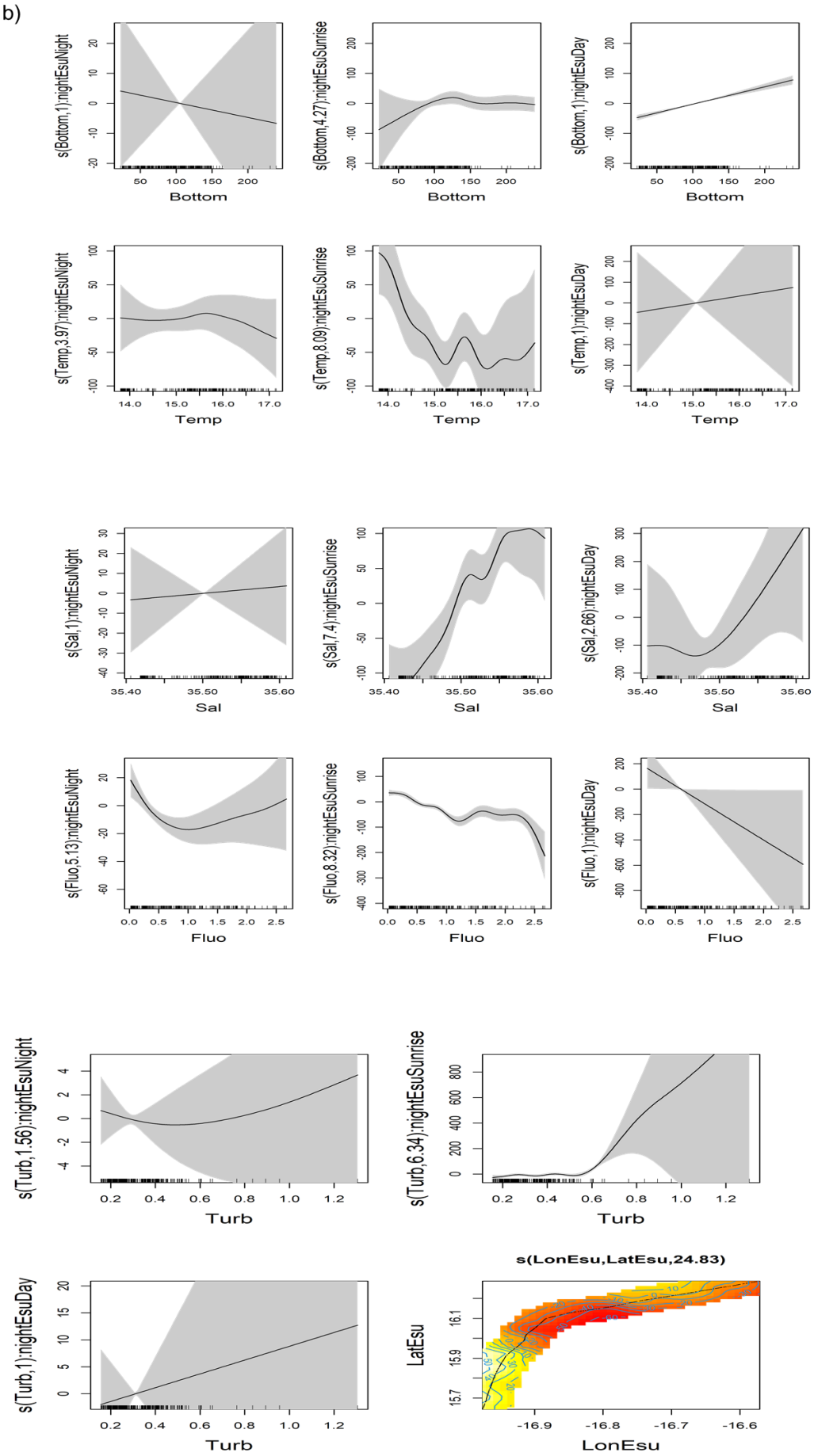
a)


s(Bottom,2.82):nightEsuNight
s(Bottom,4.58):nightEsuSunrise
s(Bottom,1):nightEsuDay
s(Temp,6.43):nightEsuNight
s(Temp,7.1):nightEsuSunrise
s(Temp,1):nightEsuDay
Bottom
Temp

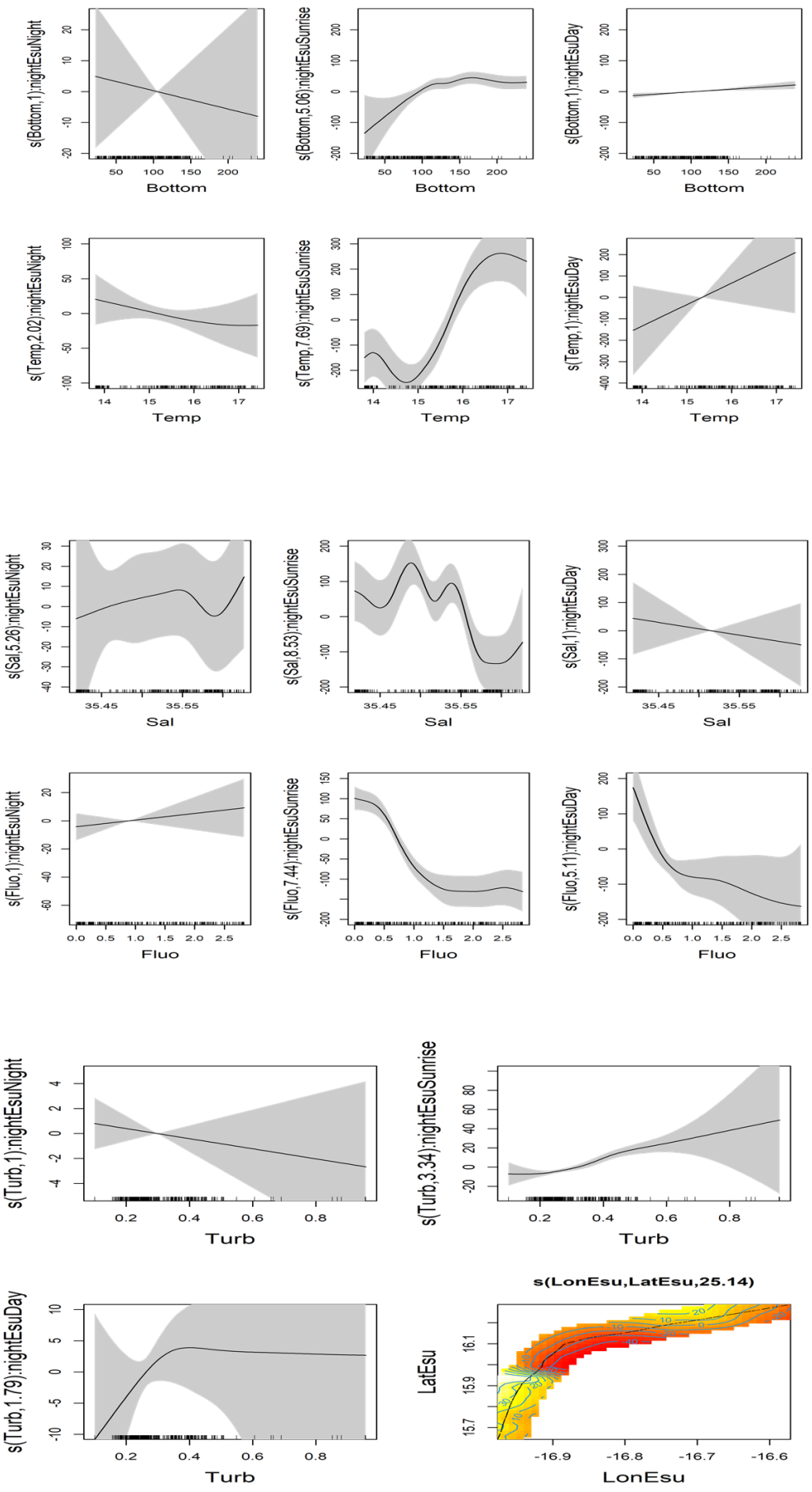
C)


s(Sal,5.03):nightEsuNight
s(Sal,4.93):nightEsuSunrise
s(Sal,2.89):nightEsuDay
s(Fluo,5.29):nightEsuNight
s(Fluo,6.27):nightEsuSunrise
s(Fluo,1):nightEsuDay
s(Turb,1):nightEsuNight
s(Turb,6.87):nightEsuSunrise
s(Turb,1):nightEsuDay
s(LonEsu,LatEsu,24.04)
Sal
Fluo
Turb
LatEsu
LonEsu

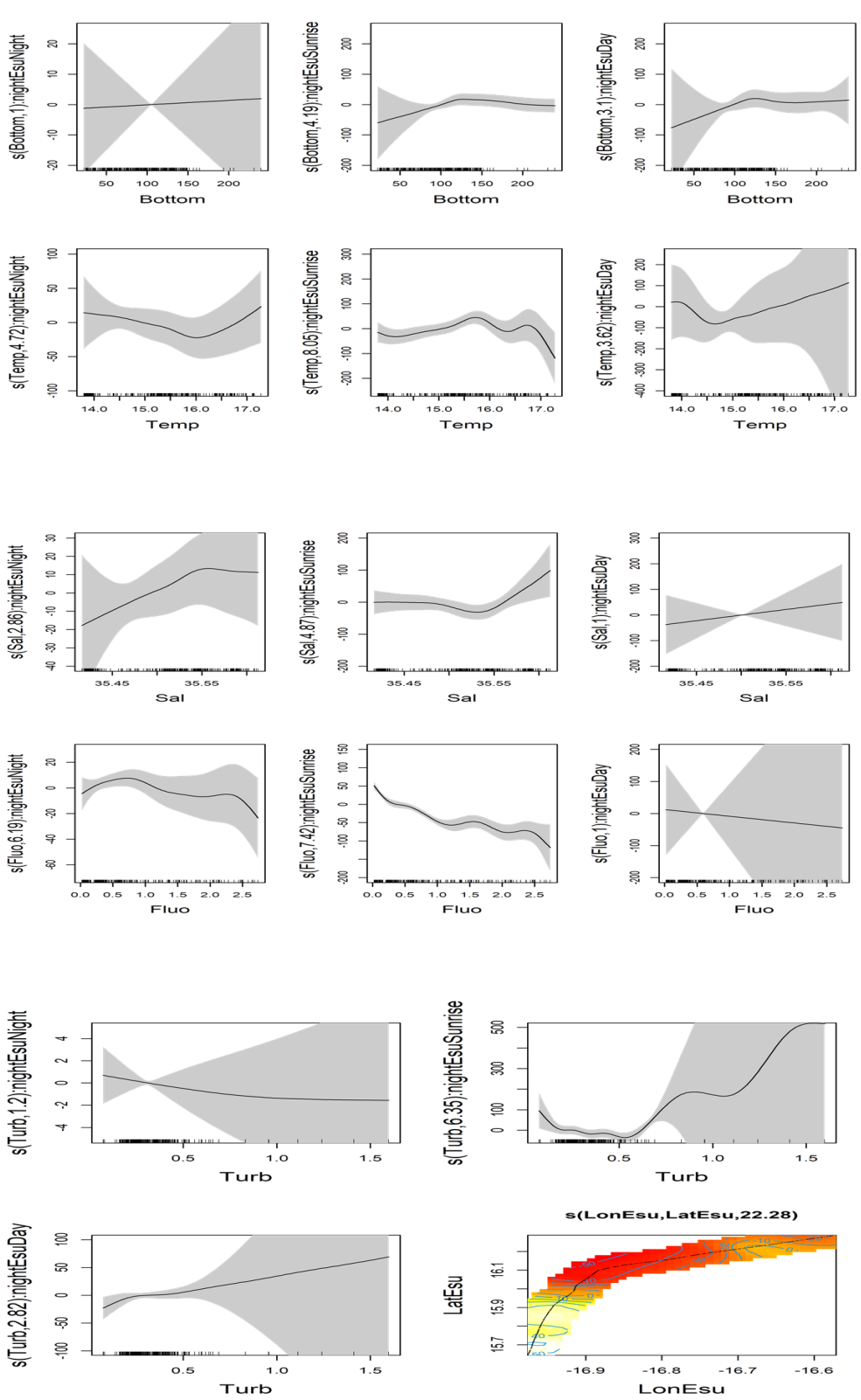
b)


s(Bottom,1):nightEsuNight
Bottom
s(Bottom,4.27):nightEsuSunrise
Bottom
s(Bottom,1):nightEsuDay
Bottom
s(Temp,3.97):nightEsuNight
Temp
s(Temp,8.09):nightEsuSunrise
Temp
s(Temp,1):nightEsuDay
Temp
s(Sal,1):nightEsuNight
Sal
s(Sal,7.4):nightEsuSunrise
Sal
s(Sal,2.66):nightEsuDay
Sal
s(Fluo,5.13):nightEsuNight
Fluo
s(Fluo,8.32):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1.56):nightEsuNight
Turb
s(Turb,6.34):nightEsuSunrise
Turb
s(Turb,1):nightEsuDay
Turb
s(LonEsu,LatEsu,24.83)
LatEsu
LonEsu

c)


s(Bottom,1):nightEsuNight
s(Bottom,5.06):nightEsuSunrise
s(Bottom,1):nightEsuDay
Bottom
s(Temp,2.02):nightEsuNight
s(Temp,7.69):nightEsuSunrise
s(Temp,1):nightEsuDay
Temp
s(Sal,5.26):nightEsuNight
s(Sal,8.53):nightEsuSunrise
s(Sal,1):nightEsuDay
Sal
s(Fluo,1):nightEsuNight
s(Fluo,7.44):nightEsuSunrise
s(Fluo,5.11):nightEsuDay
Fluo
s(Turb,1):nightEsuNight
s(Turb,3.34):nightEsuSunrise
Turb
s(Turb,1.79):nightEsuDay
s(LonEsu,LatEsu,25.14)
LatEsu
LonEsu

d)


s(Bottom,1):nightEsuNight
Bottom
s(Bottom,4.19):nightEsuSunrise
Bottom
s(Bottom,3.1):nightEsuDay
Bottom
s(Temp,4.72):nightEsuNight
Temp
s(Temp,8.05):nightEsuSunrise
Temp
s(Temp,3.62):nightEsuDay
Temp
s(Sal,2.86):nightEsuNight
Sal
s(Sal,4.87):nightEsuSunrise
Sal
s(Sal,1):nightEsuDay
Sal
s(Fluo,6.19):nightEsuNight
Fluo
s(Fluo,7.42):nightEsuSunrise
Fluo
s(Fluo,1):nightEsuDay
Fluo
s(Turb,1.2):nightEsuNight
Turb
s(Turb,6.35):nightEsuSunrise
Turb
s(Turb,2.82):nightEsuDay
Turb
s(LonEsu,LatEsu,22.28)
LatEsu
LonEsu

A.3 Appendix 3: Study of the impact of environmental variables on scattering layers using signature-based regression

A.3.1 Spatial PCA results

On high sea of southern Senegal

- At 38, kHz axis 1 is positively correlated with the variables (Temp, Sal), (Sal,Sal), (Fluo, Temp), (Fluo, Sal), (Fluo, Fluo), (Turb,Temp) and (Turb,Sal) and negatively correlated with the variables (Sal) and (Fluo) while axis 2 is positively correlated with the variables (Temp,Turb), (Sal,Turb), (Fluo,Turb) and (Turb,Turb) and negatively correlated with the variable (Turb) ; axis 3 is positively correlated with the (Sal,Temp) and (Temp) variables and negatively correlated with the (Temp,Temp) and (Temp,Fluo) variables, while axis 4 is negatively correlated with the bottom variable; axis 5 is positively correlated with the (Turb,Fluo), (Sal,Fluo) and (Turb,Fluo) variables.

- At 70 kHz, axis 1 is positively correlated with the (Fluo,Fluo), (Fluo,Turb), (Turb,Fluo) and (Turb,Turb) variables and negatively correlated with the (Fluo) and (Turb) variables, while axis 2 is positively correlated with the (Sal,Fluo) and (Sal,Turb) variables; axis 3 is positively correlated with the (Sal) and (Turb,Temp) variables and negatively correlated with the (Temp,Sal), (Sal,Sal), (Fluo,Sal) and (Turb,Sal) variables, while axis 4 is positively correlated with the (Temp) variable and negatively correlated with the (Temp,Temp), (Sal,Temp) and (Fluo,Temp) variables; axis 5 is positively correlated with the bottom variable and negatively correlated with the (Temp,Fluo) and (Temp,Turb) variables.

- At 120 kHz, axis 1 is positively correlated with the variables (Temp,Turb), (Sal,Fluo), (Sal,Turb), (Fluo,Temp), (Fluo,Turb), (Temp,Fluo), (Turb,Fluo) and (Turb,Turb) and negatively correlated with the variable (Turb), while axis 2 is positively correlated with the variables (Temp), (Temp,Fluo), (Fluo,Fluo) and (Turb,Temp) and negatively correlated with the variables (Fluo) and (Temp,Temp); axis 3 is negatively correlated with the (Temp,Sal), (Sal,Sal), (Fluo,Sal) and (Turb,Sal) variables, while axis 4 is negatively correlated with the (Sal,Temp) and axis 5 is negatively correlated with bottom and positively with (Sal) .

- At 200 kHz, axis 1 is positively correlated with the variables (Fluo,Temp), (Fluo,Sal), (Fluo,Fluo) and (Fluo,Turb) and negatively correlated with the variables (Fluo) and (Sal,Fluo) while axis 2 is positively correlated with the variable (Sal,Sal) and negatively correlated with the variables (Sal), (Temp,Sal), (Temp,Fluo) and positively correlated with the variable (Sal,Sal) ; axis 3 is positively correlated with the variables (Turb,Sal), (Turb,Fluo) and (Turb,Turb) and negatively correlated with the variable (Turb), while axis 4 is positively correlated with the variable (Temp) and negatively correlated with the variable (Temp,Temp); axis 5 is positively correlated with the variables (Turb,Temp) and bottom and negatively correlated with the variables (Sal,Temp), (Temp,Turb) and (Sal,Turb).

In northern Senegal

- At 38 kHz , axis 1 is positively correlated with the variables (Temp,Temp), (Temp,Sal), (Sal,Temp), (Sal,Sal), (Fluo,Temp) and negatively correlated with the variables (Temp) and (Sal), while axis 2 is positively correlated with the variables (Turb,Temp), (Turb,Sal) and (Turb,Fluo) and negatively correlated with the variable (Turb) ; axis 3 is positively correlated with the (Temp,Fluo), (Sal,Fluo) and (Fluo,Fluo) variables and negatively correlated with the (Fluo) variable, while axis 4 is positively correlated with the (Turb,Turb) variable and negatively correlated with the (Temp,Turb) and (Sal,Turb) variables; axis 5 is negatively correlated with the (Fluo,Turb) and bottom variables and positively with (Fluo,Sal).

- At 70 kHz, axis 1 is positively correlated with the (Temp,Temp), (Temp,Sal), (Temp,Turb), (Sal,Temp), (Sal,Sal) and (Sal,Turb) variables and negatively correlated with the (Temp), (Sal) and (Turb) variables, while axis 2 is positively correlated with the (Fluo) variable and negatively correlated with the (Fluo,Temp), (Fluo,Sal) and (Fluo,Fluo) variables; axis 3 is positively correlated with the variable (Turb,Fluo) and negatively correlated with the variables (Temp,Fluo) and (Fluo,Turb), while axis 4 is positively correlated with the variables (Sal,Fluo), (Turb,Temp) and (Turb,Sal); axis 5 is positively correlated with the variable (Turb,Turb) and negatively correlated with bottom.

- At 120 kHz, axis 1 is positively correlated with the variables (Temp,Temp), (Temp,Sal), (Sal,Temp) and (Sal,Sal) and negatively correlated with the variables (Temp) and (Sal), while axis 2 is positively correlated with the variables (Turb,Temp), (Turb,Sal) and (Turb,Fluo) and negatively correlated with the variable (Temp,Fluo) ; axis 3 is positively correlated with the variables (Temp,Turb), (Sal,Fluo), (Sal,Turb) and (Fluo,Turb) and negatively correlated with the variables (Turb), (Fluo,Temp), (Fluo,Sal) and (Turb,Turb), while axis 4 is negatively correlated with the variable (Fluo) and positively correlated with the variable (Fluo,Fluo); axis 5 is positively correlated with the variable bottom.

- At 200 kHz , axis 1 correlates positively with the variables (Temp,Temp), (Temp,Sal), (Sal,Temp) and (Sal, Sal) and negatively with the variables (Temp) and (Sal), while axis 2 correlates positively with the variables (Turb,Temp), (Turb,Sal) and (Turb,Fluo) and negatively with the variables (Turb) and (Turb,Turb); axis 3 is correlated with the (Fluo) and (Temp,Turb) variables and negatively with the (Fluo,Temp), (Fluo,Sal) and (Fluo,Fluo) variables, while axis 4 is negatively related to the (Temp,Fluo) and (Sal,Fluo) variables; axis 5 is positively related to the bottom variable and negatively related to the (Sal,Turb) and (Fluo,Turb) variables.

A.3.2 Clustering results

- On high sea of southern Senegal
  For all frequencies also, the 1^{st} component is more characteristic of the classes. At 38kHz, Class 1 is characterised by low values of signatures coefficients related to (Temp, Sal), (Sal,Sal), (Fluo, Temp),(Fluo, Sal), (Fluo, Fluo), (Turb,Temp) and (Turb,Sal), while Class 2 has the opposite effect. At 70 kHz Class 1 is characterised by low values of signatures coefficients related to

(Fluo,Fluo), (Fluo,Turb), (Turb,Fluo) and(Turb,Turb), while Class 2 has the opposite effect. At 120 kHz Class 1 is characterised by low values of signatures coefficients related to (Temp,Turb), (Sal,Fluo), (Sal,Turb),(Fluo,Temp), (Fluo,Turb), (Temp,Fluo), (Turb,Fluo) and (Turb,Turb), while Class 2 has the opposite effect. At 200 kHz, Class 1 is characterised by low values of (Fluo,Temp), (Fluo,Sal), (Fluo,Fluo)and (Fluo,Turb), while Class 2 has the opposite effect.

- In northern Senegal
  For all frequencies also, the 1^{st} component is more characteristic of the classes. At 38kHz, Class 1 is characterised by low values of signatures coefficients related to (Temp,Temp), (Temp,Sal), (Sal,Temp),(Sal,Sal), (Fluo,Temp), while Class 2 has the opposite effect. At 70 kHz Class 1 is characterised by low values of signatures coefficients related to Temp,Temp), (Temp,Sal), (Temp,Turb),(Sal,Temp), (Sal,Sal) and (Sal,Turb), while Class 2 has the opposite effect. At 120 kHz Class 1 is characterised by low values of signatures coefficients related to (Temp,Temp), (Temp,Sal), (Sal,Temp)and (Sal,Sal), while Class 2 has the opposite effect. At 200 kHz, Class 1 is characterised by low values of (Temp,Temp), (Temp,Sal), (Sal,Temp)and (Sal, Sal), while Class 2 has the opposite effect.

A.4 Appendix 4: Papers and oral communications

A.4.1 Articles

- Yoba Kande, Ndague Diogoul, Patrice Brehmer, Sophie Dabo-Niang, Papa Ngom, Yannick Perrot. (2024). Demonstrating the relevance of spatial-functional statistical analysis in marine ecological studies: The case of environmental variations in micronektonic layers. Ecological Informatics, 81, 102547. https://doi.org/10.1016/j.ecoinf.2024.102547

- Idris Si-Ahmed, Leila Hamdad, Christelle Judith Agonkoui, Yoba Kande, Sophie Dabo-Niang. (2023). Principal Component Analysis of Multivariate Spatial Functional Data. (submitted at Big Data Research)

- Amidou Sonko, Copin Dalyal, Patrice Brehmer, Cheikh Diop, Guillaume Constantin De Magny, Mamadou Fall, Yoba Kande, Patricia Moulin, Naby Souleymane Faye, Saliou Faye, Isabelle Linossier Gaël Le Pennec. (2023). Assessment of the global toxicity of marine sediments from the Dakar peninsula (Senegal, West Africa). Environmental Monitoring Assessment 195, 185. https://doi.org/10.1007/s10661-022-10635-2

- Amidou Sonko, Patrice Brehmer, Guillaume Constantin de Magny, Gaël Le Pennec, Bissoume Sambe Ba, Ousmane Diankha, Mamadou Fall, Isabelle Linossier, Maryvonne Henry, Issa NDiaye, Saliou Faye, Yoba Kande, François Galgani. (2022). Pollution assessment around a big city in West Africa reveals high concentrations of microplastics and microbiologic contamination. Regional Studies in Marine Science, 59, 102755. https://doi.org/10.1016/j.rsma.2022.102755

A.4.2 Oral communications

- *Oral*; Amidou Sonko, Patrice Brehmer, Guillaume Constantin de Magny, Gaël Le Pennec, Bissoume Sambe Ba, Ousmane Diankha, Mamadou Fall, Isabelle Linossier, Maryvonne Henry, Issa N'Diaye, Saliou Faye, Yoba Kande, and François Galgani . (2023). Pollution assessment around a big city in West Africa reveals high concentrations of microplastics and microbiologic contamination. Symposium international conservation de la biodiversité marine et côtière dans la sous-région nord-ouest africaine; 08-10/05/2023, Hôtel Fleur de Lys, Dakar, Sénégal

- *Poster*; Amidou Sonko, Copin Dalyal, Brehmer Patrice, Cheikh Diop, Guillaume Constantin De Magny, Mamadou Fall, Yoba Kande, Patricia Moulin, Naby Souleymane Faye, Saliou Faye, Isabelle Linossier and Gaël Le Pennec. (2023). Assessment of the global toxicity of marine sediments from the Dakar peninsula (Senegal, West Africa). Symposium international conservation de la biodiversité marine et côtière dans la sous-région nord-ouest africaine; 08-10/05/2023, Hôtel Fleur de Lys, Dakar, Sénégal

- *Oral*; Yoba Kande, Sophie Dabo-Niang, Patrice Brehmer. (2022). Spatial functionnal analysis of acoustic fisheries data. 20th Edition of the International Workshop in Spatial Econometrics and Statistics, Lille, France. https://hal.science/hal-04490551

- *Poster*; Yoba Kande, Sophie Dabo-Niang, Ndague Diogoul, and Patrice Brehmer. (2022).Spatial functionnal analysis application on fisheries acoustics data coupled with fine scale environmental data. In : Jech M. (ed.).. Working group of fisheries acoustics, science and technology (WGFAST). ICES Scientifics Reports, 4 (54), p. 52. ICES. Working Group of Fisheries Acoustics, Science and Technology (WGFAST), Somone (SEN), 2022/04/25-28. ISSN 2618-1371. https://hal.science/hal-04460161

- *Poster*; Yoba Kande, Sophie Dabo-Niang, Patrice Brehmer. (2021). Spatial functionnal analysis of acoustic fisheries data. Méthodes aléatoires pour les sciences de la santé et de l'environnement (MASSE'21). UCAD, CNRS, Univ Lille, INRIAE, IRD. Dakar, Sénégal. https://stat-sciences-sante.fr/wp-content/uploads/2021/12/Resume_KANDE.pdf

- *Poster*; Amidou Sonko, Patrice Brehmer, Mariline Diara, Guillaume Constantin de Magny, Amy Gassama Sow, Massal Fall, Luc Finot, Maryvonne Henry, Yoba Kande, Hawa Diallo, Ousmane Diankha, Cheikh Diop, Ibrahima Cissé, François Galgani. 2017. Toxicité Globale des Sédiments, de la qualité des eaux et des microplastiques. United Nations Water: World water day, Journée mondiale de l'eau. 22 mars, Thiès, Senegal

- *Poster*; Yoba Kande, Patrice Brehmer, Guillaume Constantin De Magny, Amy Gassama, Amidou Sonko, Aba Diop, Luc Finot, Ibrahima Cissé, Massal Fall, Oumar Sadio, Mariline Diara, François Galgani, and Ousmane Diankha. (2016). Study of the spatial variability of marine pollution around the peninsula of Cape Verde. Third International Conference AWA "ICAWA", 13-15 December 2016, Dakar, Senegal. Session 3 ICAWA, WP3_S3_86, pp. 81-82. https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers17-12/010071832.pdf hal-04488074v1

- *Oral*; Amidou Sonko, Patrice Brehmer, Guillaume Constantin De Magny, Amy Gassama, Cheikh Diop, Ibrahima Cisse, Massal Fall, Luc Finot, Maryvonne Henry, Yoba Kande, Mariline Diara, and François Galgani. (2016). Study of the overall toxicity, water quality and microplastics of the peninsula of Cape Verde, Senegal. Third International Conference AWA "ICAWA", 13-15 December 2016, Dakar, Senegal, pp. 79-80. https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers18-01/010072126.pdf