MARINHA DO BRASIL
INSTITUTO DE ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA
UNIVERSIDADE FEDERAL FLUMINENSE
PROGRAMA ASSOCIADO DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA MARINHA

CAROLINA SIQUEIRA DOS REIS

ESTUDO DA DISTRIBUIÇÃO DO PLÂNCTON EM FINA ESCALA NA ÁREA NORTE DA PLATAFORMA SUDESTE BRASILEIRA

Arraial do Cabo

2025

CAROLINA SIQUEIRA DOS REIS

ESTUDO DA DISTRIBUIÇÃO DO PLÂNCTON EM FINA ESCALA NA ÁREA NORTE DA PLATAFORMA SUDESTE BRASILEIRA

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia Marinha do Instituto de Estudos do Mar Almirante Paulo Moreira (IEAPM) / Universidade Federal Fluminense (UFF) como requisito parcial para obtenção do título de Doutor em Biotecnologia Marinha.

Orientador: Prof. Dr. Lohengrin Dias de Almeida Fernandes
Coorientador: Dr. Lucas Nunes Teixeira

Arraial do Cabo

2025

Biblioteca do IEAPM

R375e Reis, Carolina Siqueira dos

Estudo da distribuição do plâncton em fina escala na área norte da plataforma sudeste brasileira / Carolina Siqueira dos Reis.-- Arraial do Cabo, RJ: IEAPM, 2025.

30 f.: il.; 30 cm.

Orientador: Lohengrin Dias de Almeida Fernandes

Coorientador: Lucas Nunes Teixeira

Tese (Doutorado em Biotecnologia Marinha) - Instituto de Almirante Paulo Moreira e Universidade Federal Fluminense - Associado de Pós-Graduação em Biotecnologia Marinha, Arraial do Cabo.

1. Aprendizado de máquinas. 2. FAIR. 3. Imageamento. 4. Mesozooplâncton I. Fernandes, Lohengrin Dias de Almeida II. Teixeira, Lucas Nunes III. Instituto de Estudos do Mar Almirante Paulo Moreira. IV. Universidade Federal Fluminense. V. Título.

CDD: Ed. 22 -- 660.6

CAROLINA SIQUEIRA DOS REIS

ESTUDO DA DISTRIBUIÇÃO DO PLÂNCTON EM FINA ESCALA NA ÁREA NORTE DA PLATAFORMA SUDESTE BRASILEIRA

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia Marinha do Instituto de Estudos do Mar Almirante Paulo Moreira (IEAPM)/Universidade Federal Fluminense (UFF) como requisito parcial para obtenção do título de Doutor em Biotecnologia Marinha. Linha de Pesquisa: Biotecnologia Ambiental.

BANCA EXAMINADORA

Dr. Lohengrin Dias de Almeida Fernandes
Instituto de Estudos do Mar Almirante Paulo Moreira
Professor Orientador - Presidente da Banca Examinadora

Dr. Fábio Contrera Xavier
Instituto de Estudos do Mar Almirante Paulo Moreira
(Examinador interno)

Dra. Louisi Souza de Oliveira
Instituto de Estudos do Mar Almirante Paulo Moreira
(Examinadora interna)

Dra. Catarina da Rocha Marcolin
Universidade Federal do Sul da Bahia
(Examinadora externa)

Dra. Érica Caroline Becker

Istituto Nazionale di Oceanografia e di Geofisica Sperimentale-OGS

(Examinadora Externa)

Dra. Andréa Green Koettker

Universidade Federal Santa Catarina

(Suplente Externo)

Arraial do Cabo

2025

AGRADECIMENTOS

Encerrar um doutorado aos 26 anos é um lembrete para a Carol do futuro de que com muita pouca idade já conquistamos muita coisa. E tudo foi possível graças à colaboração de muitas pessoas. Agradeço ao meu orientador as inúmeras oportunidades, confiança e zelo pela minha vida pessoal e profissional. Ao meu coorientador, pela paciência e cuidado ao me ensinar cada etapa das análises e programação. Ao Mission Atlantic, por me proporcionar viver incontáveis experiências internacionais ao lado de profissionais que tanto admiro. À CAPES, pelo financiamento durante meu período de doutorado na Bélgica e ao Flanders Marine Institute (VLIZ) - Marine Data Center (VMDC) - pelo suporte durante os 10 meses morando e trabalhando no coração da Europa. Agradeço a Marinha do Brasil por ceder laboratórios, navios, equipamentos e vários profissionais que me auxiliaram durante a coleta de dados e execução do meu trabalho. Aos meus amigos de uma vida inteira, pela compreensão e apoio durante esses últimos anos; e aos meus amigos do instituto, a rotina ao lado de vocês foi mais leve. Agradeço a minha família que sempre apoiou todos os meus sonhos e compreendeu minha ausência nos últimos anos. Por fim, um agradecimento especial ao nosso amigo Márcio (in memoriam), que deixou um imenso vazio em nossos corações. Nossos dias foram muito mais felizes ao seu lado. Que você sempre seja lembrado com boas risadas.

"Que nada nos defina,
que nada nos sujeite.
Que a liberdade seja a nossa própria substância,
já que viver é ser livre."

Simone de Beauvoir

RESUMO

Durante muitos anos, nosso conhecimento sobre organismos do plâncton foi obtido através de amostragens utilizando redes de arrasto. Embora tenha seus benefícios, analisar a diversidade de organismos coletados através desse método pode demandar um longo tempo de processamento. Para superar as limitações, inerentes às amostragens com rede, desde o início da década de 1990, câmeras *in situ* têm sido utilizadas para detectar, medir e identificar partículas marinhas do plâncton. Essas ferramentas comprovam a eficiência de uma amostragem em alta resolução, veloz e que permite estudos em escalas precisas da distribuição dos organismos planctônicos ao longo de diferentes coordenadas geográficas e profundidades. Visto a inovação do sistema de imageamento, esta tese é composta por dois capítulos, nos quais utilizamos imagens obtidas *in situ* com o sistema Lightframe On-sight Key species Investigation (LOKI), a fim de ampliar o conhecimento da distribuição do plâncton no norte da plataforma sudeste brasileira. No primeiro capítulo, estudamos diferentes ferramentas de gerenciamento de dados, focada no desenvolvimento de uma estrutura de trabalho que viabilize a inter-comparação entre diferentes formatos. Nossos resultados destacaram a importância da utilização de sistemas de harmonização de dados pré-definidos por iniciativas internacionais e amparados pelos princípios FAIR. No segundo capítulo, utilizamos o banco de dados padronizado no primeiro capítulo, para estudar a relação da estratificação dos oceanos com a distribuição vertical e horizontal do plâncton. Nossos resultados mostraram que a partir da combinação de sistemas de imagens e modelagem, temos potencial para prever a distribuição tridimensional de diferentes organismos, em escala surpreendentemente detalhada. Através desses dois capítulos demonstramos como a inter-comparação de diferentes informações pode ser utilizada para melhor compreender padrões ecológicos de diferentes escalas espaciais.

Palavras-chave: Aprendizado de máquinas, FAIR, Imageamento, Mesozooplâncton

ABSTRACT

For many years, our knowledge about plankton has been obtained through trawl sampling. Although this method has advantages, it can be time-consuming to analyze plancton diversity. To overcome the specific limitations of trawl sampling, *in situ* cameras have been used since the early 1990s to detect, measure, and identify plankton and marine particles. These tools demonstrate the efficiency of high-resolution and high-speed sampling, allowing for distribution studies on precise scales across different geographic regions and depths. Given the innovation of the imaging system, this thesis is composed of two chapters, in which we use images obtained *in situ* with the Lightframe On-sight Key Species Investigation (LOKI) system in order to expand knowledge of the distribution of plankton in the northern of South Brazilian Shelf. In the first chapter, we study different data management tools, focused on developing a framework that enables intercomparison between different formats. Our results highlight the importance of using data harmonization systems predefined by international initiatives and expanded by the FAIR principles. In the second chapter, we use the database standardized in the first chapter to study the relationship between ocean stratification and planning distribution. Our results show that by combining imaging and modeling systems, we have the potential to predict the three-dimensional distribution of different organisms at surprisingly small scales. Through these two chapters, we demonstrate how the intercomparison of different information can be used to better understand ecological patterns at different spatial scales.

Keywords: FAIR, Imaging system, Machine learning, Mezoplankton,

LISTA DE FIGURAS

LISTA DE TABELAS

ABREVIAÇÕES

CTD	Conductivity Temperature Density
CNN	Redes Neurais Convolucionais
QC	Quality Control
DIC	Carbono Inorgânico Dissolvido
DwC	Darwin Core
EBV	Essential Biodiversity Variable
EOV	Essential Ocean Variable
EMODnet	European Marine Observation and Data Network
EurOBIS	European Node of the Ocean Biodiversity Information System
FAIR	Findable Accessible Interoperable Reusable
GBIF	Global Biodiversity Information Facility
GOOS	Global Ocean Observing System
IA	Inteligência Artificial
IPT	Integrated Publishing Toolkit
LOKI	Lightframe On-sight key Species Investigation
LOPC	Laser Optical Particle Counter
MLD	Mixed Layer Depth
MSR	Mean Squared Residual
OBIS	Ocean Biogeographic Information System
POC	Carbono Orgânico Particulado
REA	Research Executive Agency

ROI	Region of Interest
SBS	South Brazilian Shelf
TDWG	Biodiversity Information Standards
VPR	Video Plankton Recorder

SUMÁRIO

1. INTRODUÇÃO GERAL

A utilização do mar sempre foi fortemente motivada pelas necessidades básicas dos seres humanos. Além da perspectiva econômica e histórica, os oceanos também viabilizam a vida no planeta por meio de diversos processos ecológicos sustentados, majoritariamente, por organismos presentes do plâncton. O plâncton é formado por organismos que derivam na água, sujeitos aos efeitos do vento, das ondas e correntes oceânicas. São onipresentes e dominam a vida no oceano em termos de abundância e biomassa (Bar-On and Milo, 2019). Em virtude da elevada diversidade dos grupos taxonômicos presentes neste compartimento, muitas classificações foram estabelecidas ao longo dos anos. A mais elementar (ie. fitoplâncton e zooplâncton) separa os organismos em produtores dos organismos consumidores, respectivamente. Outra divisão adotada ao longo dos anos é a escala de tamanho corporal dos organismos, abrangendo uma miríade de formas de vida que vão desde femtoplâncton (< 0,2µm, por exemplo, vírus marinhos) ao megaplâncton (> 20cm; águas-viva, Omori e Ikeda, 1992). Também, podem ser divididos por hábitos de vida (ie. holoplâncton e meroplâncton) dependendo do tempo de permanência no plâncton ao longo de seu ciclo de vida (Dipper, 2022).

Apesar de frequentemente serem quase imperceptíveis a olho nu, esses organismos compõem a base de toda a teia trófica marinha. São responsáveis pela transferência de energia e matéria orgânica, além da participação no ciclo global do carbono através do processo de bomba biológica (Litchman et al., 2015; Fenchel, 1988). A bomba biológica de carbono é um dos principais processos responsáveis pela regulação dos níveis de dióxido de carbono atmosférico e, consequentemente, o transporte desse dióxido de carbono para sedimentos marinhos. Esse processo contribui com aproximadamente dois terços do transporte vertical de carbono (Passow and Carlson, 2012; Sarmiento and Gruber, 2002). Por desempenharem papeis essenciais para a dinâmica da vida no planeta, são frequentemente utilizados como bioindicadores da saúde dos oceanos. Assim, a alta sensibilidade do plâncton às mudanças ambientais possibilita o aumento da compreensão sobre o impacto dessas mudanças na transferência de energia nos ecossistemas marinhos (Wadjikar *et al*., 2017; Hays et al., 2005; Hall J., et al. ,1976).

A dinâmica de distribuição do plâncton é resultado de múltiplas interações entre processos físicos, químicos e biológicos, como correntes oceânicas, estratificação térmica, interações ecológicas e disponibilidade de nutrientes (Krause et al., 2020;

Sailley et al., 2015 and therein). Horizontalmente, padrões globais de distribuição do plâncton refletem gradientes costa-oceano, de temperatura e frentes oceânicas, que em algumas regiões, atuam proporcionando condições ideais para o crescimento de organismos da base da teia trófica - o fitoplâncton (McManu et al., 2016). Enquanto verticalmente, organismos do zooplâncton são responsáveis por uma das maiores migrações do planeta Terra. A Migração Vertical do Zooplâncton (MVZ) é um fenômeno marcado pela migração ascendente de organismos em direção à superfície à noite, e um movimento descendente para águas mais profundas durante o dia (Forward, 1988). A MVZ é essencial para uso eficiente de recursos de diferentes hábitats (estratos de profundidade) e proporciona uma maior diversidade de espécies no oceano (Brierley, 2014).

Outro fator igualmente relevante e inter-relacionado que afeta a distribuição do plâncton é a estratificação na coluna d'água. A água do mar geralmente forma camadas estratificadas com águas mais leves perto da superfície e águas mais densas em maior profundidade. Esta configuração estável atua como uma barreira à mistura de água que afeta a eficiência das trocas verticais de calor, carbono, oxigênio e outros constituintes (Li et al., 2020). Quando afetada por eventos de turbulência (i.e. ventos, correntes), as águas se misturam alterando a disponibilidade de fornecimento de nutrientes e moldando as interações e a eficiência alimentar (Xue et al., 2022). Um modelo conceitual de controle físico e biológico na dinâmica do fitoplâncton (Lindemann and John, 2014) demonstra a multiplicidade dos fatores que afetam as variações de aclimatação fisiológica desses organismos. Além disso, esse modelo destaca o cenário da estratificação como mecanismo de eficiência nas interações, descrevendo que quanto maior a camada de mistura, menor o desempenho das interações ecológicas entre organismos, uma vez que os nutrientes estão mais dissolvidos e os organismos estão mais afastados. Em contrapartida, quando a camada de mistura apresenta uma amplitude menor, a eficácia das interações aumenta, juntamente com a concentração dos nutrientes.

As regiões de ressurgência costeira impulsionadas pelo vento são estão entre as áreas mais produtivas do oceano global devido à ressurgência de águas subterrâneas ricas em nutrientes na zona eufótica (Falkowski et al. 1998). Nessas regiões a estabilidade da coluna d'água diminui graças à ação dos ventos frequentes combinados a outros fatores como temperatura e das correntes de água. Considerando esse contexto, a área selecionada para este estudo foi o norte da plataforma sudeste brasileira que conta com uma alta produtividade em decorrência da região de Ressurgência de Cabo Frio (Gonzalez-Rodriguez et al., 1992). A mistura de águas superficiais e profundas ocorre devido à combinação da velocidade de distribuição vertical e da densidade da água (Coelho-Souza et al., 2012). Em períodos de ressurgência, a alta disponibilidade de nutrientes beneficia a floração do plâncton, que por sua vez favorece o aparecimento de uma rica fauna pelágica que inclui espécies economicamente..

Na região de ressurgência de Cabo Frio, a comunidade planctônica é extensivamente documentada devido aos estudos de Jean-Louis Valentin (Valentin, 1984, 1999, 2001), bem como na plataforma sudeste brasileira, diversos estudos que incluem aspectos da variação espaço-temporal do plâncton, influência de fatores ambientais na distribuição e diversidade das comunidades (Carvalho et al., 2022; Pennick et al., 2023; Tosetto et al., 2021). Mais recentemente, observamos a aplicação de diferentes ferramentas para quantificar e identificaçaro plâncton nessa região, como uso de imagens de FlowCAM (Matos et al., 2024); sinais acústicos (Assunção et al., 2023) e as tradicionais coletas com rede (Garcia et al., 2021). No entanto, observamos uma lacuna tecnológica, pois apesar de uma bibliografia relevante sobre a região, poucos estudos abordam metodologias como técnicas de imagem *in situ*, tampouco ferramentas de modelagem de aprendizagem de máquina.

Das redes as telas

Durante muitos anos o estudo da distribuição do plâncton marinho foi realizado através de redes em arrastos horizontais, verticais ou oblíquos (Barnes, H., 1951). Apesar das redes possuírem benefícios como baixo custo, fácil manuseio e a possibilidade de realizar estudos moleculares com as amostras coletadas, o uso desse método integra e mistura organismos, danificando-os e gerando perdas de 12% a 87% de biomassa (Alcaraz et al., 2003). Além disso, a resolução espacial é limitada e as técnicas manuais para separação e identificação dos organismos são demoradas e dificultam a aplicação em larga escala (Eerola et al., 2024). Com o intuito de aprimorar o desempenho de estudos em uma escala mais refinada, o uso de sistemas ópticos para monitorar populações de plâncton vem sendo adotados continuamente nas últimas décadas (Le et al., 2022 and references therein).

Esses sistemas evoluíram das primeiras medições ópticas da água (*e.g.* absorção, espalhamento, atenuação e fluorescência) utilizadas desde 1950's para caracterizar propriedades em massa associadas a partículas e fitoplâncton, em particular.

Atualmente, além de propriedades como forma e intensidade da luz, os sistemas de imagem e/ou vídeo podem prover informações muito mais complexas sobre diferentes alvos. Isso é possível, pois esses sistemas alcançam organismos e partículas de diferentes tamanhos, atuando em diferentes resoluções e em escalas espaciais que alcançam até amostragens globais (Gardner et al., 2018).

A enorme variedade de equipamentos disponíveis abrange aplicações tecnológicas e ecológicas, como o *Lightframe On-sight Key species Investigation* (LOKI), Video Plankton Recorder (VPR), Laser Optical Particle Counter (LOPC), entre outros (Lombard et al., 2019; Orenstein et al., 2022; Irisson et al., 2022; Herman et al., 2004). Os estudos realizados com essas ferramentas comprovam a eficiência de uma amostragem em alta resolução e veloz. Além disso, permite o desenvolvimento de estudos em escalas precisas da distribuição dos organismos planctônicos ao longo de diferentes coordenadas geográficas e profundidades. Além de uma investigação biológica, horizontal e vertical, os sistemas de imageamento são, frequentemente, compostos por múltiplos sensores que possibilitam a observação dos parâmetros físicos e químicos da coluna d'água na mesma escala de observação.

O LOKI é um sistema *in situ* de imagem de plâncton e partículas menores ou iguais a 200 micrômetros. Foi utilizado pela primeira vez no Atlântico Sul em 2015, em áreas de alta turbidez devido a um desastre ambiental, demonstrando a eficácia do uso de sistemas de imagem em diferentes condições ambientais (Matos et al., 2024). Desde então, o Laboratório de Plâncton e Microbiologia Marinha do Instituto de Estudos do Mar Almirante Paulo Moreira (IEAPM), o único que dispõe desse equipamento no Brasil, atua no desenvolvimento e melhoramento de técnicas referentes à aquisição de imagens em múltiplos ambientes. Dentre os objetivos propostos ao implementar tal tecnologia, estavam a rápida avaliação das condições dos ambientes amostrados, investigação dos impactos ambientais e antropológicos nas comunidades pelágicas, análise da riqueza e diversidade dos organismos que compõem a base da teia trófica aquática e afins.

As imagens de plâncton e partículas podem ser processadas manualmente ou por meio de técnicas automatizadas, que permitem a análise da estrutura taxonômica de uma comunidade e a medição das características morfológicas de organismos individuais (Sosik e Olson, 2007; LeCun et al., 2015; Orenstein e Beijbom, 2017; Luo et al., 2018; Ellen et al., 2019), como "Machine learning", que consiste em um sistema de aprendizagem automática supervisionada de algoritmos que aprendem a categorizar dados novos a partir de um conjunto de exemplos de informação gerados por humanos (Irisson et al. 2022). O desenvolvimento de tecnologias de imagem de plâncton e partículas foi fortemente influenciado pelo desejo de reduzir o tempo de processamento de amostras. Entretanto, apesar dos incontáveis benefícios, as etapas de análise do conjunto de imagens coletadas, detecção e extração de regiões de interesse (ROIs) e, transformação dessas informações em *insights* ecológicos, ainda que com auxílio de algoritmos de aprendizagem, segue sendo uma atividade laboriosa.

O avanço de sistemas de imagens para estudo do plâncton resultou em alguns problemas, como a obtenção de diversos dados em formatos diferentes, dados fragmentados e desconectados, além da ampliação da quantidade de bases de dados não padronizadas. Esse efeito destacou a necessidade do gerenciamento de dados, e do estabelecimento de padrões e procedimentos de controle de qualidade que promovam a capacidade de tornar esses conjuntos de dados localizáveis, acessíveis, interoperáveis e reutilizáveis (Princípios FAIR, "*Findable, Accessible, Interoperable, Reusable*"; Wilkinson et al., 2016). A detecção, o acesso, a inter-comparação e a reutilização de dados é uma das principais maneiras de garantir que a comunidade acadêmica contribua para políticas públicas, ajudando os tomadores de decisão e incentivando uma ciência mais sustentável em relação ao uso de recursos financeiros e naturais (Ramos et al., 2022).

Para potencializar a intercomparação e reutilização dos dados de plâncton, foi utilizado o formato Darwin Core (Wieczorek et al. 2012), o mais atual padrão internacional estabelecido para publicação de dados de biodiversidade. Apesar de ainda não contemplar integralmente todos os aspectos do plâncton, especialmente partículas e atributos (Giering et al., 2020; Titocci et al, 2025), representa a prática mais adotada entre os trabalhos que estão alinhados com os princípios FAIR.

2. OBJETIVOS

O presente trabalho está vinculado ao Programa Mission Atlantic “Horizon 2020” [Grant Agreement No 862428], que visa ampliar a compreensão do funcionamento dos ecossistemas dos oceanos e dos fatores de mudança que afetam a biodiversidade, os fluxos tróficos e a produtividade de base da teia marinha no Atlântico sul.

Objetivo Geral:

Implementar novos padrões de estrutura de dados provenientes do sistema LOKI para, dessa forma, modelar a distribuição espacial de plâncton na Ressurgência de Cabo, uma das regiões mais importantes para uso humano no Brasil.

Objetivos Específicos:

- Estruturar e documentar os dados de imagem de plâncton provenientes do sistema LOKI.
- Seguir padrões como o Darwin Core (DwC) para garantir os princípios FAIR.
- Garantir a consistência e a interoperabilidade dos dados com repositórios globais.
- Combinar a aplicação de tecnologia de imagem com o uso de aprendizagem de máquina para modelar a distribuição do plâncton na região de Ressurgência de Cabo Frio.
- Avaliar e descrever a distribuição de micro- e mesoplâncton em diferentes condições de estratificação na Plataforma Sudeste Brasileira.

CAPÍTULO 1

Submetido à revista Biodiversity Data Journal

A framework to ensure data harmonization from *in situ* optical plankton recorder Lightframe On-sight Key species Investigation

Carolina Reis[1*], Lucas T. Nunes[1], Thiago da S. Matos[1], Ricardo Coutinho[1], Laura Marquez[2], Ruben Perez-Perez[2], Patricia Martin-Cabrera[3], Lennert Schepers[2], Lohengrin Fernandes[1]

[1] Marine Biotechnology, Instituto de Estudos do Mar Almirante Paulo Moreira, Arraial do Cabo - RJ, 28930-000, Brazil.

[2] Flanders Marine Institute (VLIZ), Jacobsenstraat 1, 8400, Belgium

[3] Intergovernmental Oceanographic Commission of UNESCO – Ocean Best Practice System, 7 Place de Fontenoy, 75007 Paris, France.

Abstract

We present the development and validation of a framework for harmonizing plankton image-based data obtained from the Lightframe On-Sight Key Species Investigation (LOKI) system, aligning it with the Darwin Core (DwC) standards. By using image processing, metadata integration and taxonomic mapping techniques, we aim to show interoperability and reusability in global biodiversity databases. The proposed workflow establishes guidelines for data collection, formatting and standardization. Results indicate an improvement in data accessibility for ecology and spatial modeling research, meeting the FAIR principles (*Findable, Accessible, Interoperable, Reusable*).

Key words: Imaging, *In situ,* Interoperability, Plankton, Harmonization, FAIR, DwC

Introduction

Plankton is listed as an Essential Ocean and Biodiversity variable (EBVs; Muller-Karger et al., 2018) by the Global Ocean Observing System (GOOS). It comprises a diverse array of life forms which are associated with multiple crucial processes in the marine ecosystems. These organisms regulate biogeochemical cycles shaping marine trophic chains, from local to global scales (Mitra et al., 2014). Due to these importances, plankton has been extensively studied through a variety of methodologies. However, it is widely acknowledged that traditional approaches in plankton ecology are often time-intensive and have constrained our ability to fully elucidate the complex factors and processes governing their abundance, distribution and diversity. Recent extensive development of automated image acquisition systems in plankton research has increased our understanding of these communities. Nevertheless, integrating different datasets, originating from different sources, became a major effort (Lombard et al., 2019; Martin-Cabrera et al., 2022).

The Lightframe On-sight Key species Investigation (LOKI) is an *in situ* plankton optical recorder, currently adopted for different purposes like marine ecosystem mapping (Snoeijs-Leijonmalm et al., 2022); risk assessment (Matos et al., 2024), trait observations (Maps et al., 2023) and biodiversity exploration (Wenzhöfer et al., 2024; Fong et al., 2024). Along with LOKI, there are several *in situ* instruments, collecting both images and videos, with similar purposes such as the Ichthyoplankton Imaging System (ISIIS; Cowen et Guigand, 2008), Video Plankton Recorder (VPR; Davis et al., 1992) and Underwater Vision Profile (UVP; Picheral et al., 2022), highlighting the importance and efficiency of using *in situ* imaging instruments in plankton ecology. Nevertheless, when it comes to these types of devices, the data volume, technical heterogeneity, missing semantic structure and availability are just a small sample of the wide array of challenges (Schoening et al., 2022).

Although the new optical technologies have made it possible to increase the spatiotemporal resolution of plankton research, most of them generate millions of images to be classified and *terabytes* of raw data to be stored (Irisson et al., 2022). Integrating diverse types of information from a wide array of observation tools—spanning spatial and temporal scales—in a manner that is both human-interpretable and machine-readable remains one of the most significant challenges for researchers. Nevertheless, it is also a crucial step to fill the gaps in marine biodiversity sciences. Additionally, collaborative data science benefits a wide range of stakeholders, including researchers, software and tool developers, funding agencies, and the broader data science community. It advances the integration, discovery, and accessibility of data, enriching global databases, and driving technological innovation (Titocci et al., 2025).

For achieving common standards and practices, principles regarding the findability, accessibility, interoperability and reusability of information (*FAIR principles*), started to be adopted by several international initiatives to ensure consistency and inter-comparability of biodiversity datasets (Wilkinson et al., 2016). The number of tools and workflows designed to standardize data from various optical recorder devices has grown substantially in recent years, especially for plankton. However, several devices remain unaddressed, highlighting an ongoing gap in the field. In order to improve the interoperability and reusability of plankton data derived from images, the aim of this work is to construct and validate a framework that harmonizes the output data from the Lightframe On-sight Key species Investigation (LOKI) system, ensuring the *FAIRification* process and making data available in global biodiversity databases.

Darwin Core Standards

The Darwin Core standard (Wieczorek et al. 2012) is the most used (meta)data standards to share biodiversity data in the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/) and the Ocean Biodiversity Information System (OBIS, https://obis.org/). Both are the largest biodiversity data repositories in the world, holding hundreds of millions of species occurrences. DwC was originally developed by the Biodiversity Information Standards (TDWG) community, offering a stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources. Thus, DwC plays a key role in the sharing, use and reuse of open access biodiversity data.

DwC archives are compact file packages containing interconnected text files, allowing scientists to share information using standard terminology. The package contains a varied number of standardized files that can be adjusted according to the needs of the data provider. For marine biodiversity data, OBIS recommends using the OBIS-ENV-DATA format (De Pooter, 2017). The OBIS-ENV-DATA format is a version of a DwC archive that contains up to three data files: The events file (1) contains all information regarding the sample and/or observation. In the occurrences file (2), all the information related to each organism in each event. Finally, in the measurements or facts file (3), information will be recorded relating to each individual occurrence (e.g. size, diameter, biomass, etc.); or relating to the event (e.g. sampling methodology, temperature, salinity, pH, among other physicochemical parameters); or extra information that scientists deems relevant to fully understand the content of the dataset.

After the data has been formatted according to DwC standards, the compressed package file undergoes quality control (QC) using the EMODnetBiocheck R package

(De Pooter and Perez-Perez, 2019) before being submitted to the Integrated Publishing Toolkit (IPT; Robertson et al. 2014). The IPT is a free, open-source software tool designed to publish and share biodiversity datasets through the GBIF network. This integration ensures that the data is widely accessible and available to the global research and conservation community. This standardization not only simplifies the process of publishing biodiversity datasets, but also facilitates the discovery, investigation, evaluation and comparison of datasets as they seek answers to today's data-intensive research and policy questions.

Here, we present all the steps involved in the creation of a framework to ensure the *FAIRification* of LOKI images, from generating a DwC file with the combination of different types of environmental and biological data, until the publication of standardized data in open repositories, passing through a validation process with real data sampled in the southwestern Atlantic.

Workflow

In this framework we present the steps from data collection to data integration and publication in open repositories (Fig. 1). We, thus, highlight the necessity of guidelines for steps after the annotation (see below). Although there is a considerable diversity of methods and softwares available for LOKI and other imaging devices—such as ZooProcess (Picheral., 2010), LOKI Browser (Schulz et al., 2010) and equivalents—all of them focus on the images curation prior to *FAIRification*. The initial step is to select and combine the environmental data and the images generated for each sampled depth. The images are automatically cropped as part of the LOKI system's factory operation and configuration, however, image enhancement techniques such as contrast and brightness adjustments are applied using ImageJ (Image J version 2.3.0, Schneider et al., 2012), favoring the identification of the photographed organisms in later stages.

After this initial processing, the images are uploaded to EcoTaxa, where taxonomic information is added—a process known as “annotation”. The EcoTaxa is an interactive and collaborative platform for uploading and annotating images with their respective metadata (Picheral et al., 2017; Irisson et al., 2022). In this platform, we distinguished living organisms from inorganic materials, detritus, and artifacts (e.g., bubbles, dust), which were not relevant to our study. This process is supported by partial assistance provided by machine learning algorithms from the platform, in which users can train models based on previous identifications in the database to suggest labels for newly uploaded images (Irisson et al., 2022). Once the process of classifying and validating the images has been completed, the data is exported along with metadata, essential information from the dataset, such as title, geographical coordinates of sampled sites, the instrument used to acquire the images and environmental data.

After the annotation, the data is exported from EcoTaxa and goes through a processing and harmonization stage following the established standards, available on "GitHub" (https://github.com/missionatlantic/Plankton_IEAPM_Brazil/tree/master). After processing, the harmonized data are transferred to global open access databases and information systems on marine biodiversity for science, conservation and sustainable development (GBIF, OBIS and EMODNet).

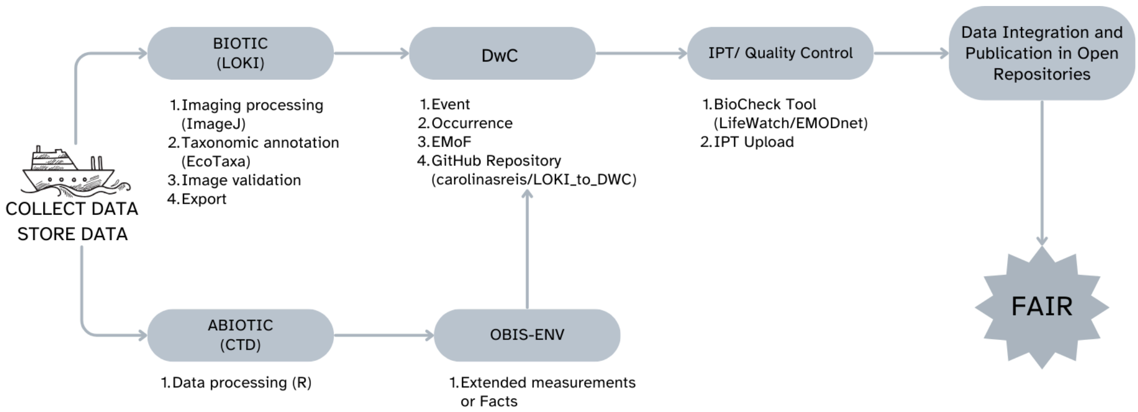

Fig. 1 Workflow showing the steps from data acquisition to results availability in an interoperable and reusable format in accordance with the Darwin Core standards and meeting the FAIR principles.

Study case

Sampling occurred during summer in the South Hemisphere, on the Oceanographic Commission "Ressurgência III", on board the Brazilian Navy's Hydroceanographic ship H39 *'Vital de Oliveira*', through vertical hauls (heave speed ~0.5 m.s^{-1}) from 140 m deep to surface at six different stations. The LOKI system fits a plankton concentration net (mesh size 200 µm, mouth opening 0.28 m^2) combined with a computer and a CTD equipped with environmental sensors (temperature, salinity, pressure, dissolved oxygen and fluorescence). The camera is a Prosilica GC 1280H (AVT-Allied Vision Technologies, Canada) with the Pensax 2514-M lens, able to acquire images with a final resolution of 23 µm.pixel^{-1}. It has a high-power LED unit, synchronized with the camera's exposure-shooting signal, which allows a fast shut-off time (55 µs) avoiding motion blurring that causes image distortion. In combination, it has an image channel 4 mm high (length = 31.3 mm, width = 20.75 mm, volume = 2.6 cm^3), causing all the particles of the image to stay in focus. In total, 514,736 images were registered (https://ecotaxa.obs-vlfr.fr/prj/5415), of which 92,839 were manually validated by a taxonomist according to the highest taxonomic level achieved, given the resolution of the image (image annotation).

Discussion

Expanding efforts to develop standardized and upgradable systems benefits and encourages collaboration among several stakeholders at different levels, from those who collect the data, those who utilize it for research and to those who use the results of the research to make decisions (Muscuugwa et al., 2021). However, imaging data is a quickly evolving field, presenting ongoing challenges, where data collected from the marine environment are increasing in heterogeneity, preventing objective comparison (Schoening et al., 2022). To tackle a portion of this challenge, the framework developed in this work was designed specifically with imaging outputs from the LOKI system, to ensure that the acquired metadata and data reach the highest level of Findability, Accessibility, Interoperability, and Reusability *(FAIR Principles*) as possible.

Our framework covered fine-scale (~1 m) depth intervals and different taxonomies of big data, albeit its complexity. Combining environmental and biological data requires tackling challenges such as the interpolation of different data formats, aggregating depth intervals, selecting appropriate depth ranges, and managing the time and expertise required. In this framework output, we prioritize the smallest possible depth intervals combined to the highest amount of information we had to maximize the possibilities of future investigations. This level of granularity facilitates a more detailed comparison. Additionally, different decisions related to depth intervals lead to distinct understanding of several aspects regarding ocean structure and functioning, its stratification, species distribution and multiple analysis of climate variables and organisms interactions (Easson et al., 2019; Zwaan et al., 2019; Vereshchaka et al., 2017).

However, when we tried to compare with other datasets, there is a lack of imaging data repositories, while the existing data is scattered and not easily accessible. Only five repositories/online platforms related to biodiversity and/or plankton imaging data were located based on the keyword ‘Lightframe On-sight Key species Investigation’ (**Table 1**).

Table. 1 Platforms and/or data collections related to the Light frame On Sight Key Species Investigation (LOKI).

Repository/Platform	Type	LOKI data available?	Workflow stage	FAIR Principles adherence	Notes
EcoTaxa	Annotation tool	Yes (67 projects)	Pre-FAIR	Partial (not findable in open repositories)	Distribution of data through an e-mail request
PANGAEA	Repository	Yes, only physical parameters (6)	Post-sampling	Low (without taxonomy available)	Schulz, J (2008) Physical oceanography and technical parameters
OBIS	Repository	Yes (1 use case - Reis et al., 2024)	Post-FAIR	High	Darwin Core format
EurOBIS	Repository	Yes (1 use case - Reis et al., 2024)	Post-FAIR	High	Darwin Core format
EMODnet	Repository	Yes (1 use case - Reis et al., 2024)	Post-FAIR	High	Darwin Core format
ZooProcess	Pre-processing software	Yes	Pre-FAIR	Without direct integration to FAIR principles	Tool to extract features and segmentation
LOKI Browser	Visualization Software	Yes	Pre-FAIR	None	Images inspection tool

Through a review of open-access databases and platforms related to biodiversity and plankton, we identified 67 projects on EcoTaxa. Although EcoTaxa is not a repository, it is clear that this is the most widely used platform to annotate and consequently store the taxonomy information about plankton images. By accessing other repositories, we found out that 6 projects were available at PANGAEA, however, all of them contained solely environmental sensor data. No taxonomic annotation or images originating from the LOKI system was easily found. In the meantime, the use case presented in this paper was the only project (Reis et al., 2024) utilizing the LOKI system while addressing all the current requirements of FAIR principle, easily findable, accessible, following the interoperable formats and prepared to be reusable by everyone.

Considering this, we underscore the need for fostering a culture of data sharing in the scientific community. The main challenges in moving marine biodiversity data management toward the FAIR principles were already reviewed (Tanhua et al., 2019) and we acknowledge the wide diversity of planktonic organisms, as well as the multitude of disparate data management structures, the increased volume of generated data, the widely used formats that are not always universally applicable. Nevertheless, we firmly uphold the benefits of sharing data following the FAIR recommendations and, we also ensure that the widespread use of centralized databases will greatly facilitate the reanalysis of existing data (Maps et al., 2023), enhancing the fine scale monitoring of marine ecosystems.

Conclusions

Imaging technology is a quickly advancing field, so even with all the required infrastructure in place, there is a growing demand to keep datasets FAIR (Schoening et al., 2022). Moreover, it is essential to remain prepared by continuously developing or updating protocols, best practices, recommendations and methodologies. Generating and implementing methods that are highly adaptable to emerging technologies acts as a way to ensure that data management practices keep pace with these innovations. We believe our framework is beneficial for LOKI users and other devices that have not been addressed with this type of framework yet, since it is easily modifiable. We strongly encourage all scientists to publish their data, reusing and extending internationally adopted and advanced metadata standards for image- and video-based data, ensuring that their data meets the open science standards.

Funding

This project received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 862428 (MISSION ATLANTIC). This output reflects the author's view and the Research Executive Agency (REA) cannot be held responsible for any use that may be made of the information contained therein.

Acknowledgments

We acknowledge the VLIZ Marine Data Center (VMDC) for the first insights about data sharing and FAIR principles.

Data availability

The data used in this study is openly available, following the FAIR standards, and can be accessed at EMODnet (https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/6d617269-6e65-696e-666f-000000008234), OBIS (https://obis.org/dataset/a17af320-138c-40a7-a9a3-0dde2983b8d1) and EurOBIS (https://www.eurobis.org/toolbox/en/download/selection/1679274bc1f608) under Creative Commons Attribution 4.0 International License.

References

Cowen RK, Guigand CM, et al. (2008) In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. Limnology and Oceanography: Methods 6 (2): 126‐132.

Davis CS, Gallager SM, Berman MS, et al. (1992) The video plankton recorder (VPR): design and initial results. Archiv für Hydrobiologie 36: 67‐81.

De Pooter D, Perez-Perez R (2019) EMODnetBiocheck: LifeWatch & EMODnet Biology QC Tool. URL: https://github.com/EMODnet/EMODnetBiocheck

Easson CG, Lopez JV, et al. (2019) Depth-dependent environmental drivers of microbial plankton community structure in the northern Gulf of Mexico. Frontiers in Microbiology 9: 3175. https://doi.org/10.3389/fmicb.2018.03175

Fong AA, Hoppe CJ, Aberle N, et al. (2024) Overview of the MOSAiC expedition. Elementa: Science of the Anthropocene 12 (1): 00135. https://doi.org/10.1525/elementa.2023.00135

GBIF (2020) gde: GBIF Dataset Explorer. CRAN: Contributed Packages https://doi.org/10.32614/cran.package.gde

Irisson JO, Ayata SD, Lindsay DJ, et al. (2022) Machine learning for the study of plankton and marine snow from images. Annual Review of Marine Science 14 (1): 277‑301. https:// doi.org/10.1146/annurev-marine-041921-013023

Lombard F, Boss E, Waite AM, et al. (2019) Globally consistent quantitative observations of planktonic ecosystems. Frontiers in Marine Science 6: 196. https://doi.org/10.3389/ fmars.2019.00196

Maps F, et al. (2024) Automatic estimation of lipid content from in situ images of Arctic copepods using machine learning. Journal of Plankton Research 46 (1): 41‑47. https://doi.org/10.1093/plankt/fbad048

Martin-Cabrera P, Perez Perez R, Irrison J, et al. (2022) Establishing plankton imagery dataflows towards international biodiversity data aggregators. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.94196

Matos TS, Reis CS, et al. (2024) Effect of turbidity on semi-automatic analysis of copepod size and abundance distribution in the water column. Advances in Bioscience and Biotechnology 15 (6): 380‑395. https://doi.org/10.4236/abb.2024.156023

Mitra A, Flynn KJ, Burkholder JM, et al. (2014) The role of mixotrophic protists in the biological carbon pump. Biogeosciences 11: 995‑1005. https://doi.org/10.5194/bg-11-995-2014

Muller-Karger FE, Miloslavich P, et al. (2018) Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks. Frontiers in Marine Science 5: 211.https://doi.org/10.3389/fmars.2018.00211

OBIS (2025) Ocean Biodiversity Information System. Intergovernmental Oceanographic Commission of UNESCO. https://obis.org

Picheral M (2010) ZooProcess manual. Laboratoire d'Océanologie de Villefranche-sur Mer (updated for version 6.16), p.106.

Picheral M, Colin S, Irisson J (2017) EcoTaxa, a tool for the taxonomic classification of images. http://ecotaxa.obs-vlfr.fr.

Picheral M, Catalano C, Brousseau D, et al. (2022) The Underwater Vision Profiler 6: an imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms. Limnology and Oceanography: Methods 20 (2): 115‑129.

Pooter D, Appeltans W, Bailly N, et al. (2017) Toward a new data standard for combined marine biological and environmental datasets-expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: 10989. https://doi.org/10.3897/bdj.5.e10989

Reis C, et al. (2024) Institute of Sea Studies Admiral Paulo Moreira (IEAPM), Brazilian Navy; Department of Marine Biotechnology, Brazil: Spatial distribution of plankton in the Cabo Frio Upwelling System in the South Brazilian Shelf. URL: https://doi.org/ 10.14284/708.

Robertson T, et al. (2014) The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the internet. PLOS One 9 (8): 102623. https://doi.org/ 10.1371/journal.pone.0102623

Schneider CA, Rasband WS, Eliceiri KW, et al. (2012) NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9 (7): 671‑675. https://doi.org/10.1038/nmeth.2089

Schoening T, et al. (2022) Making marine image data FAIR. Scientific data 9 (1): 414. https://doi.org/10.1038/s41597-022-01491-3

Schulz J (2008) Physical oceanography and technical parameters at LOKI station MSM07/3_102-6 [dataset. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research https://doi.org/10.1594/PANGAEA.708193

Schulz J, Barz K, Ayon P, et al. (2010) Imaging of plankton specimens with the lightframe on-sight keyspecies investigation (LOKI) system. Journal of the European Optical Society-Rapid Publications 5: 10017. https://doi.org/10.2971/jeos.2010.10017s

Snoeijs-Leijonmalm P, Flores H, Sakinan S (2022) Unexpected fish and squid in the central Arctic deep scattering layer. Science Advances 8 (7): 7536. https://doi.org/10.1126/sciadv.abj7536

Tanhua T, Pouliquen S, Hausman J (2019) Ocean FAIR data services. Frontiers in Marine Science 6: 440. https://doi.org/10.3389/fmars.2019.00440

Titocci J, Pata PR, et al. (2025) Pathways for converting zooplankton traits to ecological insights are paved with findable, accessible, interoperable, and reusable (FAIR) data practices. ICES Journal of Marine Science 82 (2): 017.

Vereshchaka A, Abyzova G, Lunina A, et al. (2017) The deep-sea zooplankton of the North, Central, and South Atlantic: Biomass, abundance, diversity. Deep Sea Research Part II: Topical Studies in Oceanography 137: 89‐101. https://doi.org/10.1016/j.dsr2.2016.06.017

Wenzhöfer F (2024) Expedition Programme PS143/1. Expeditions Programm Polarstern

Wieczorek J, Bloom D, Guralnick R, Blum S, et al. (2012) Darwin Core: An evolving community-developed biodiversity data standard. PLOS One 7 (1). https://doi.org/10.1371/journal.pone.0029715

Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1): 1‐9.

Zwaan GJ, Jorissen FJ, Stigter HC, et al. (1990) The depth dependency of planktonic/benthic foraminiferal ratios: constraints and applications. Marine Geology 95 (1): 1‐16. https://doi.org/10.1016/0025-3227(90)90016-D

CAPÍTULO 2

Submetido à revista Scientific Reports

Mapping marine plankton abundance along ecological gradients in the South Brazilian Shelf

Carolina Reis[1*], Lucas T. Nunes[1], Thiago da S. Matos[1], Guillem Chust[2], Ricardo Coutinho[1], Patrizio Mariani[3], Marinez E. G. Scherer[4], Lohengrin Fernandes[1]

[1] Marine Biotechnology, Instituto de Estudos do Mar Almirante Paulo Moreira, Arraial do Cabo - RJ, 28930-000, Brazil.

[2] AZTI Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi Ugartea z/g, 48395 Sukarrieta, Spain

[3] National Institute of Aquatic Resources, Section for Oceans and Arctic, Technical University of Denmark, Henrik Dams All´e, 201, 148, 2800 Kgs. Lyngby

[4] Laboratory of Integrated Coastal Management (LAGECI), Department of Geosciences, Federal University of Santa Catarina, Florianópolis, Brazil

* Corresponding author: carolinasreis@gmail.com

Abstract

Plankton communities form the foundation of marine food webs, determining the ecosystem carrying capacity and the services they provide. Their spatial distribution patterns are highly dynamic, influenced by ocean stratification, mixed layer depth, and nutrient availability. In this study, we combined advanced imaging systems with machine learning models to explore how environmental factors shape zooplankton distribution on the South Brazilian Shelf. Our three-dimensional predictions highlight depth as a key driver for all taxonomic groups, while revealing taxon-specific responses to water column stratification. Gelatinous are notably confined to the mixed layer and dominate oligotrophic regions, irrespective of coastal nutrient availability. In contrast, copepods exhibit a heterogeneous and continuous distribution, with higher abundances near nutrient-rich coastal areas. These contrasting patterns underscore the intricate dynamics of plankton communities and reinforce the need for detailed investigations to better understand their ecological roles and responses to environmental changes.

Keywords: Environmental gradients, Machine learning, Mixed layer, Stratification

Introduction

Plankton is an Essential Ocean Variable (EOV) with a vital role in regulating fish populations, seabirds, and marine mammals (Miloslavitch et al. 2018). With remarkable diversity spanning several orders of magnitude in size, plankton is extensively studied across spatial and temporal scales (Lombard et al., 2019; Bi et al., 2022). Furthermore, plankton facilitates active and passive carbon export to the deep ocean and drives essential biogeochemical cycles, including those of iron, oxygen, nitrogen, and phosphorus (Araujo et al., 2022; Kwon et al., 2009; Parekh et al., 2006). Thus, as a key mediator of energy transfer between different trophic levels, plankton research enhances our understanding of ocean food web linkages.

Plankton distribution varies vertically and horizontally from metres to kilometres across vertical (euphotic to aphotic) and horizontal (coast-ocean or equatorial-polar) gradients (Irisson et al., 2022). Studies have described that, along with environmental gradients, other anthropogenic aspects such as climate change and offshore wind farms also affect the balance between the mixing and stratification in the water column (Holt et al., 2018; Dorrell et al., 2022). These turbulent events mix the water and change the nutrient supply availability, shaping the interactions and the fish feeding efficiency (Xue et al., 2022).

Tracking and predicting how marine plankton responds to ocean stratification and environmental gradients will empower the global community to envision climate change impacts on ocean primary productivity and biodiversity (Fragoso et al., 2022; Zampollo et al., 2022). However, performing a detailed investigation of those organisms remains a challenging task. Traditional methods are time-consuming even at low spatial-temporal resolution; they normally integrate plankton data over some vertical and/or horizontal distance, and sometimes, the time lag between sampling and data analysis makes it difficult to provide feedback to policymakers in a timely manner (Remsen et al, 2004; Benfield et al., 2007).

Information addressing the fine-scale distribution of marine planktonic organisms and its relationship with environmental factors in the South Brazilian Shelf (SBS; Fig. 1) is scarce in the literature (Matos et al., 2024). Therefore, we combined the use of a non-invasive *in situ* imaging system and machine learning modelling in order to understand plankton vertical and horizontal distribution as a response to environmental characteristics on the South Brazilian Shelf. This combination has grown into a powerful tool to improve the monitoring of marine organisms and the investigation of the marine ecosystems at multiple scales. We hypothesize that planktonic distribution is shifted by water column stratification gradients along the SBS. We expect that different populations of marine planktonic organisms are likely to respond to vertical changes in the mixed layer depth, thermocline and mixing events.

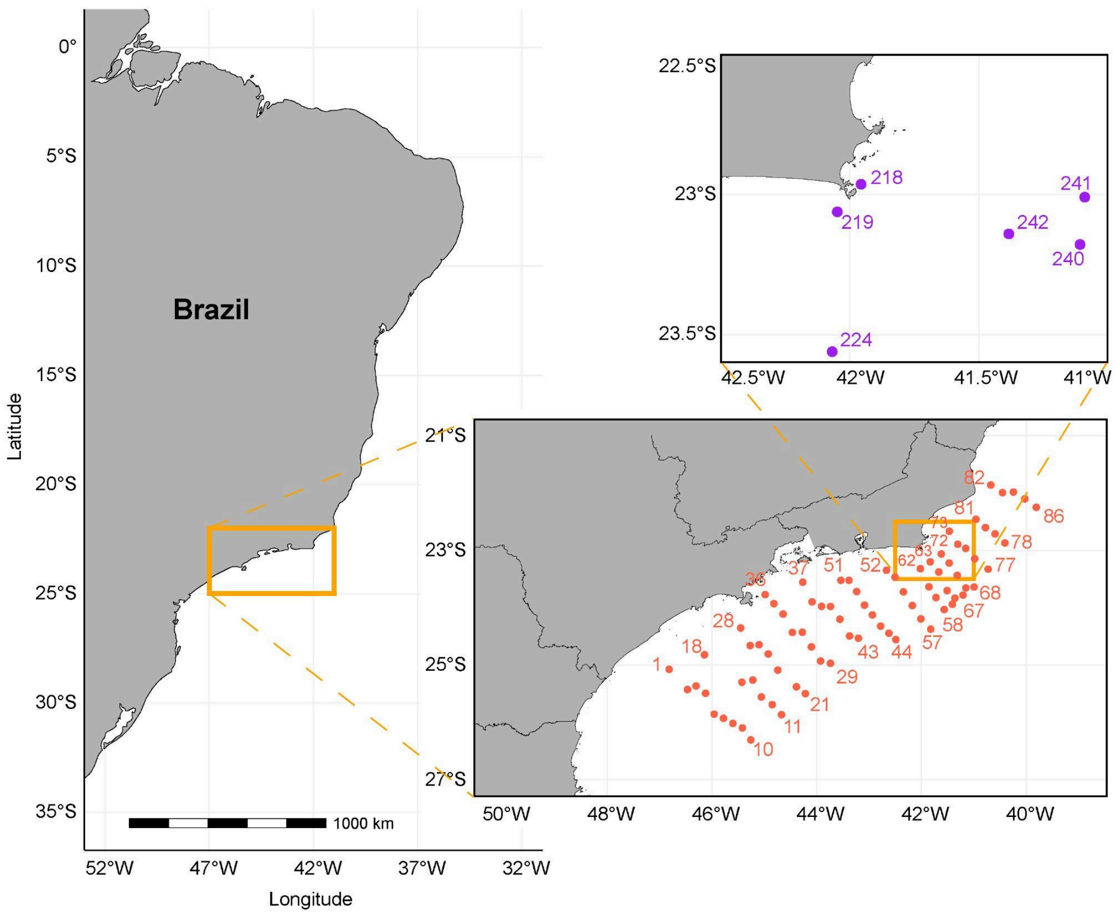

Figure 1. Location of sampled stations on the South Brazilian Shelf (SBS). Orange points indicate oceanographic stations where environmental data were collected at 1-meter intervals along a 140-meter water column. Purple points indicate oceanographic stations in the Cabo Frio Upwelling System (CFUS), where both plankton imaging and environmental data were collected.

Results

We analyzed the vertical gradients of environmental variables along the 92 stations with depths up to 140 m (Fig. S1). The temperature profiles were used as a proxy to define the stratification, divided into: mixed layer depth (MLD), thermocline and below thermocline. By defining the stratification, we observed that most of the stations presented a well-defined thermocline and, on the other hand, stations with a non-existing thermocline (St 57). The minimum depth range for MLD was 5 m but there were some stations in which this layer reached 50 m depth (e.g., ST 58, 59, 66).

Within the six modelled stations (Fig. 1), Rhizaria was the most abundant taxa on the MLD and thermocline. Depth was highly important to explain this pattern, while temperature, salinity and fluorescence seem to have a similar, but less important, effect (Fig. S2). Diatoms were most abundant on the MLD (Fig. 2), also most influenced by depth, while other variables had a lower importance (Fig. S3). Copepoda presented a heterogeneous and continuous distribution with an apparent small increase in surface layers (MLD, Fig. 2). This taxon was most influenced by environmental temperature along with depth (Fig. S4). Gelatinous dominated the MLD layer, being almost absent on the layer below the thermocline and most influenced by salinity and depth (Fig. S5).

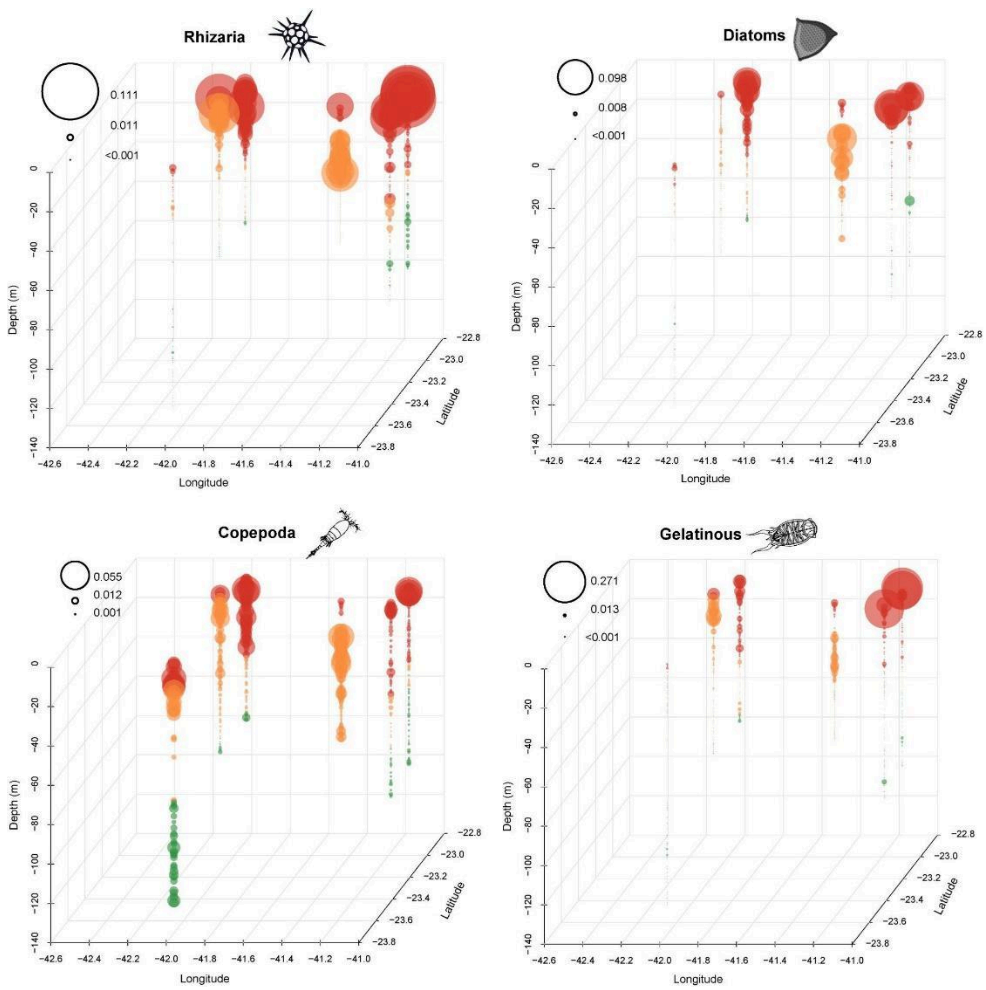

Figure 2. Tridimensional map of Rhizaria, Diatoms, Copepoda and Gelatinous, relative abundance along the 6 oceanographic stations. Colors indicate stratification layers MLD (red), Thermocline (orange), Deep layer (green).

Regarding the environmental conditions (ie. temperature, fluorescence, and depth), three groups (Rhizaria, Diatom, and Gelatinous) were positively influenced by temperature and negatively by depth. Fluorescence was surprisingly far less indicative of organisms' abundance than temperature. Rhizaria and diatoms, for instance, increased in temperatures between 16 °C and 24 °C. Fluorescence was a significant proxy of abundance of phytoplankton, but only under specific MLD conditions. When the thermocline is close to the surface (<20 m) and phyto- tends to be concentrated facilitating its predation by zooplankton. This scenario can be seen in the stations 30,31,32 in which MLD was restricted to the upper 20 meters and fluorescence was high (Fig. 3, yellow circles). Copepods presented a high variability in all the environmental conditions. Its relative abundance varied along the entire water column regardless of the temperature and fluorescence. Gelatinous adapted well to temperatures ranging between 19°C and 21°C.

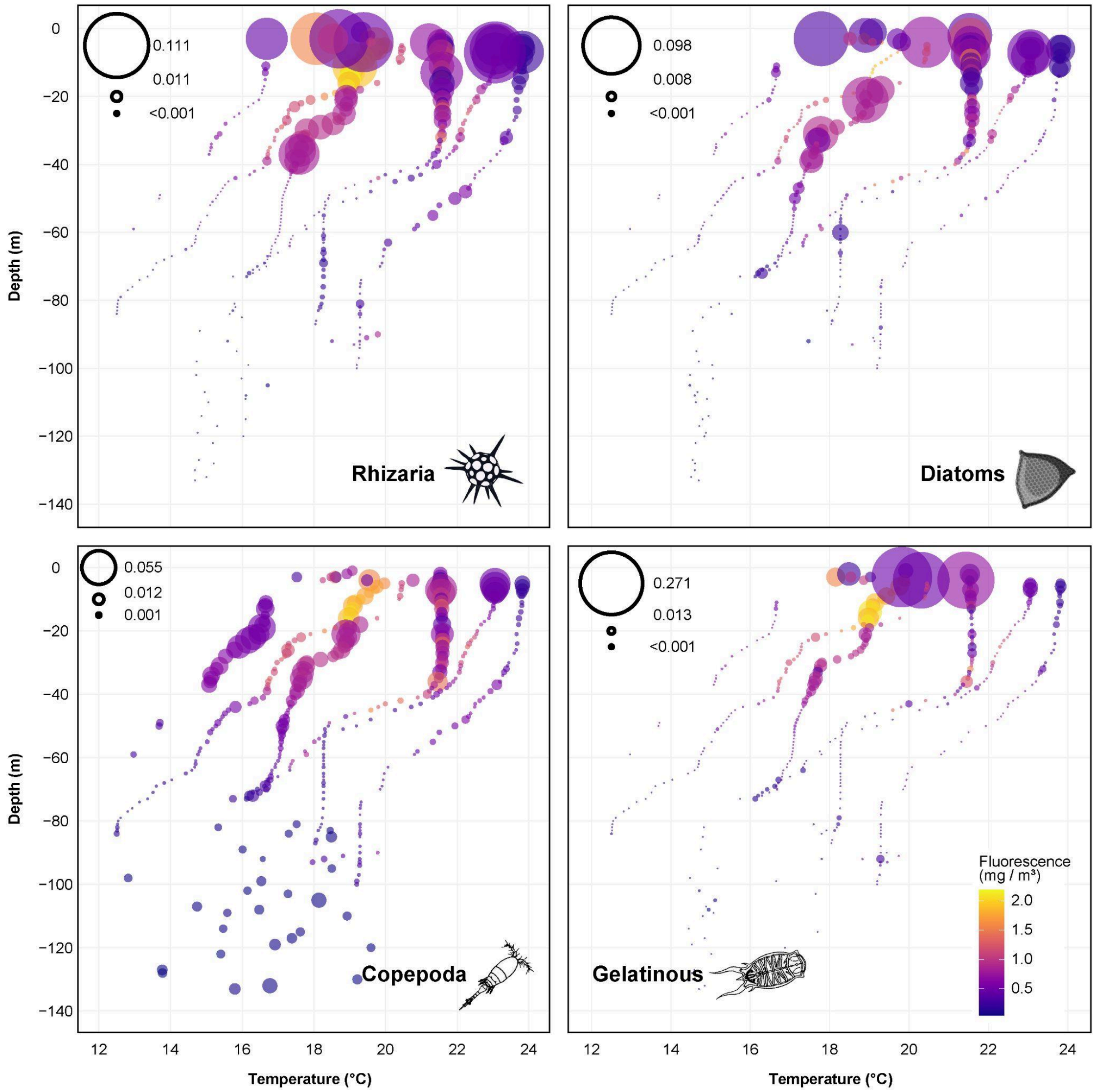

Figure 3. Patterns of relative abundance of Rhizaria, Diatoms, Copepoda and Gelatinous, following temperature, fluorescence and depth measurements in a vertical profile.

Our model (Table 1) accounted for about 20% of the variability in plankton abundance distribution, which is substantial considering the multiple factors that might influence their dynamic in ocean ecosystems. Apart from the difficulty in distinguishing correlation from causation in such complex, dynamic and highly coupled systems, our model highlighted the most explained variables.

Table1. Model performance and validation metrics for the four taxonomic groups. *Range*: mean of minimum–maximum relative abundance; *MSR*: Mean Squared Residual; *VarExpl*: Mean Variance Explained (%); *RMSE*: Root Mean Squared Error; *MAE*: Mean Absolute Error. Values are shown as means ± standard deviation.

Plankton	Range	MSR	VarExpl	RMSE	MAE
Rhizaria	0 - 0.16	< 0.0001 ± < 0.001	26.6 ± 5.8	0.02 ± 0.004	0.01 ± 0.002
Diatoms	0 - 0.17	< 0.0001 ± < 0.001	9.6 ± 4.8	0.02 ± 0.005	0.01 ± 0.002
Copepoda	0 - 0.09	< 0.0001 ± < 0.001	21.4 ± 5.5	0.01 ± 0.002	0.009 ± 0.001
Gelatinous	0 - 0.41	0.001 ± < 0.001	27.4 ± 15.9	0.03 ± 0.007	0.02 ± 0.003

The spatial outputs of the model (Fig. 4) predict higher likelihood of Rhizaria occurrence mostly near the coast, and with a slight presence in oligotrophic regions, but also spreading over the shelf. The prediction probability increases between latitude 42°W and 40°W in all the three used strata. For Diatoms, the predictions showed that it predominated in coastal regions which evidently decrease along the oligotrophic area. The potential occurrence of Copepoda is clearly elevated near the coasts, coinciding with both Rhizaria and Diatoms. However, it spreads along the ocean gradient and copepods are estimated in lower concentrations. Our spatial prediction for Gelatinous suggested higher relative abundance in oligotrophic regions, exhibiting a distribution pattern different from that of other groups analyzed in this study.

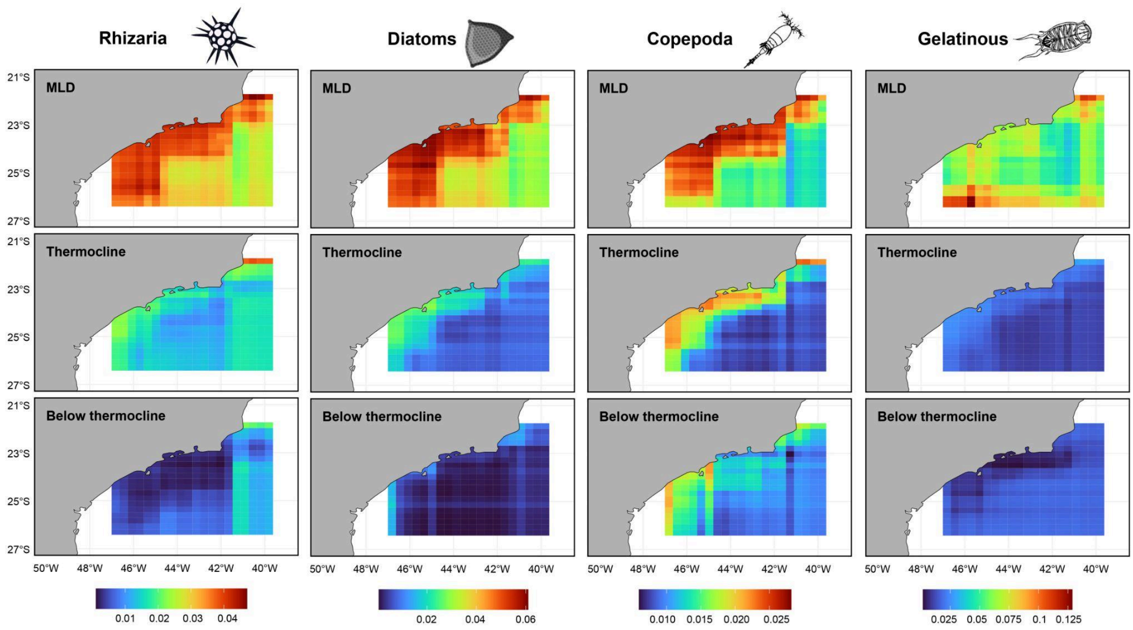

Figure 4. Relative abundance percentage patterns predicted for Rhizaria, Diatoms, Copepoda and Gelatinous along the SBS and for three different stratification layers: Mixed Layer Depth (MLD), Thermocline and Below thermocline.

Our predictions pointed to the predominance of higher abundances of plankton communities in shallow and coastal waters with a gradual decrease as the distance from the coast extends toward the open seas, except for the gelatinous group. Our result is consistent with analysis in the tropical Atlantic that describe large-bodied planktonic organisms dominating offshore while smaller organisms, mostly Copepods, dominate at the shelf (De Figueiredo et al., 2025). Indeed, coastal regions frequently experience a higher input of nutrients either from anthropogenic activities, riverine runoffs, bottom resuspension, or local upwelling (Ferreira et al., 2020), especially in shallow waters where frequent mixing in the water column (the mixed layer) affects the environment functioning and the community dynamics. However, particular taxons presented a notable exception in the modelled spatial distribution pattern. Gelatinous organisms were found to be spread along the oligotrophic regions over the coast-shelf gradient, highlighting the advantages of their low metabolic costs and living strategies like the ability of ctenophores to 'de-grow' when deprived of food (Steele et al., 2001). Also, in all the three strata, defined in this study, Rhizaria was predicted to thrive in oligotrophic regions. One possible explanation for this pattern stems from the presence of Collodaria-symbiont association described in photic zones and oligotrophic environments globally (Biard et al., 2016, 2017).

Depth and temperature were the most important explanatory variables affecting the distribution of all plankton groups. Rhizaria predominated on the upper portion of the water column (MLD); such a pattern was already observed for the North Atlantic (Barth et al., 2024). The phytoplankton, herein exemplified by the Diatoms, prevailed in the MLD, in line with the photic gradient, high sunlight and warm waters. Despite the continuous distribution of copepods along the water column, higher abundances are likely to happen in the upper layers (MLD and thermocline). Evidence was found for Calanoida suspension-feeders in the upper layers (Dias et al., 2024), while the detritivorous Ergasilida had a slightly deeper below the thermocline but wider distribution. Interestingly, gelatinous estimations revealed that this taxon was remarkably restricted to the MLD with very low occurrence probability in the thermocline and the layer below it. The depth shapes most of the conditions that change plankton distribution, including the photic gradients, the stratification, and the depth-dependency of some relationships (Bressan et al., 1993; Van der Zwaan et al., 1990).

The stratification also plays a key role in multiple processes, such as the formation of the surface mixed layer, the regulation of light supply for photosynthesis, nutrient supply from the subsurface ocean, and the phytoplankton communities regulation (Chen et al., 2021; Mena et al., 2019). Additionally, studies have pointed that the vertical stratification can hinder the migration of small zooplankton populations and indicate different grazing pressures (Mitra and Flynn, 2005 and Long et al., 2021). Other studies suggest that phytoplankton growth and mortality are equally affected by the stratification, limiting nutrients supply and affecting the N:P ratio (Gupta et al., 2020).

Although we expected higher abundances of plankton especially in regions close to the Cabo Frio Upwelling Systems (CFUS), extensively documented in literature due to the high input of nutrient-rich cold waters (Valentin, 2001; Coelho-Souza et al., 2012; Gonazalez-Rodriguez et al., 1992), no evidences of higher abundances was find in this region. Except for Gelatinous that predominated in oligotrophic regions, the higher abundances close to the coast did not follow the CFUS delimitation. Our suggestion is that CFUS is a highly dynamic region, so there's a regime shift in the plankton assemblages and sometimes large phyto- and zooplankton groups can be used as early indicators of productivity shifts in upwelling ecosystems. This regime has been described for diatoms and copepods, highlighting the link between temporal changes in the ecosystem conditions and the composition of the assemblage (Matos et al., 2024).

We validated our model using the importance of environmental variables (temperature, fluorescence and depth) and its consequent influence on the regulation of the marine plankton distribution. We noticed that for Rhizaria, diatoms, and gelatinous organisms, the explained variance was very promising, considering that the majority of the occurrences were predicted following a specific pattern of distribution. Distinct from the rest, Copepoda was proved to be a challenging endeavor, given their heterogeneous distribution. Although our model successfully estimated the vertical and horizontal distribution of plankton, we emphasize its simplicity, as it assumes a direct relationship between plankton abundance and environmental variables (temperature, salinity, fluorescence). Depth emerged as the most influential predictor, likely due to its integration of multiple underlying unmeasured variables. Despite this simplicity, we believe our models provide robust predictions, supported by high explained variance and low residuals, and offer valuable insights into plankton distribution patterns across the South Brazilian Shelf.

Once the Marine Spatial Planning (MSP) pilot-project in Brazil has been developed in the SBS area, these results could inform MSP trade-offs and use of allocations. It can be used to inform decision-makers on potential impacts of human maritime uses, especially the ones that could change water temperature and fluorescence or even affect the water column in different depths. Furthermore, Climate Change effects on water temperature are likely to affect plankton distribution as well and must be taken into account in managerial initiatives. Future refinements could incorporate additional biotic and abiotic factors to further enhance predictive accuracy. Also, the outputs generated during this study point towards new perspectives to understand plankton distribution on the SBS

Methods

Study area and sampling design

The South Brazilian Shelf (SBS) in the South Atlantic Ocean spans from latitude 22°S to 34°S, encompassing the continental shelf (depth <200 m) from the southernmost point of Brazil (Chuí) to the Cabo Frio Upwelling System (CFUS; Scherer et al., 2024). Within the SBS, the CFUS is recognized as a highly productive region where wind-driven upwelling events enhance primary production, supporting high biodiversity (Matos et al., 2024). In this region, we conducted a fine-scale sampling using the Light frame On-sight Key-species Investigation (LOKI) in vertical hauls (heave speed ~0.5 m s^{-1}) from 140 m deep to the surface at six oceanographic stations (Fig. 1). This system fits a plankton concentration net (200 µm mesh size, mouth opening 0.28 m^2) combined with a computer and a CTD equipped with environmental sensors (temperature, salinity, pressure, dissolved oxygen, and fluorescence). The camera is a Prosilica GC 1280H (AVT-Allied Vision Technologies, Canada) with the Pensax 2514-M lens, able to acquire images with a final resolution of 23 µm pixel $^{-1}$. It has a high-power LED unit, synchronized with the camera's exposure-shooting signal, which allows a fast shut-off time (55 µs) avoiding motion blurring that causes image distortion. In combination, it has an image channel 4 mm high (length = 31.3 mm, width = 20.75 mm, volume = 2.6 cm^3), causing all the particles of the image to stay in focus.

From the results of a cruise performed during summer in the South Hemisphere, we obtained around 500,000 images at all six stations (Fig. 1). In the laboratory, we process those raw images by running an automatic recognition of the region of interest (ROI), cropping and removing the background. Each organism in the images was identified with the help of the online tool EcoTaxa (Picheral et al., 2017). The EcoTaxa considerably reduces the sample processing time due to the combination of a random forest algorithm and convolutional neural networks used to predict identifications of unidentified objects (Breiman, 2001; Graham, 2014; Grandremy et al., 2024). The initial classifications were carefully manually validated, amounting to 90,000 images and 30,000 organisms were classified according to their taxonomic identity (https://ecotaxa.obs-vlfr.fr/prj/5415). We characterized the studied organisms with a size range of 200 to 20000 µm by four groups: Rhizaria, Diatoms, Copepoda and Gelatinous.

Additionally to the biological data, we used the environmental variables of a cruise performed during spring in the South Hemisphere, such as salinity, temperature (°C), fluorescence, and depth (m) measurements from 92 vertical oceanographic stations along the SBS (Fig. 1). The environmental variables were measured using a conductivity-temperature-depth (CTD) system (Seabird SBE 911 plus) at ~1- meter depth intervals, spanning from 140 m to surface. Temperature profiles from each of the 92 hauls were used to delineate the water column into the mixed layer depth (MLD), thermocline and below the thermocline layers, following the 0.2°C threshold criterion outlined by de Boyer Montegut et al. (2004) (see Supplementary Material).

Modelling plankton vertical distribution

We used relative abundance to analyse population structural variations caused by differences in the numerical contributions of individual species. For each taxon, we applied a normal distribution to exclude abundance values outside the 99% confidence interval. Missing values from this procedure were replaced with the mean relative abundance, and the response variable was subsequently log-transformed. Missing data for environmental variables (*ie. max of 67 values for fluorescence from a total of 375 samples*) were estimated as a function of the response variable using the "rfImpute" function from the “randomForest” package (Liaw & Wiener 2002), which estimates values based on weighted averages derived from sample similarities across 999 trees. Model performance was assessed using the mean squared residual (MSR) and the proportion of explained variance (%Var), where lower MSR and higher %Var indicated better predictive accuracy.

To model the vertical distribution of the four plankton groups at the six oceanographic stations, we applied Supervised Random Forest Regression models using the “randomForest” function from the *randomForest* package. To account for the inherent variability in the randomization process, we ran 1000 iterations for each zooplankton group, each with a random split of the dataset into 80% training and 20% testing subsets. The training data were used to fit the Random Forest model with 999 trees, using depth, temperature, fluorescence, and salinity as predictors. This "bootstrapped" approach aims to evaluate the model performance across multiple randomized splits, serving as an internal validation mechanism to ensure robustness and reduce the potential for overfitting. Additionally, at each iteration, the model predicted the vertical relative abundance of plankton groups at 86 oceanographic stations within the SBS region, using environmental variables obtained at these stations.

Mapping zooplankton horizontal distribution

To visualize the vertical distribution of the plankton groups across the oceanographic stations, we performed an interpolation to estimate the abundance values at locations between the stations. A grid of points was generated over the study area, with latitudes and longitudes sampled at regular intervals. The analysis was conducted separately for each vertical stratum (upper, middle, and below). For each layer, we trained a Random Forest model to predict plankton abundance based on the latitude and longitude coordinates of the stations. The observed abundance values at the stations were used as the response variable. After training the models, we used them to predict the abundance at all points on the grid, creating a continuous surface of plankton distribution for each depth layer. All analyses and graphics were performed using R software version 4.2.2 (R Core Team 2022).

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 862428 (MISSION ATLANTIC). This output reflects only the author's view and the Research Executive Agency (REA) cannot be held responsible for any use that may be made of the information contained therein. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Data availability

The data used in this study is openly available, following the FAIR standards, and can be accessed at EMODnet (https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/metadata/6d617269-6e65-696e-666f-000000008234), OBIS (https://obis.org/dataset/a17af320-138c-40a7-a9a3-0dde2983b8d1) and EurOBIS (https://www.eurobis.org/toolbox/en/download/selection/1679274bc1f608) under Creative Commons Attribution 4.0 International License.

References

Araujo, G., Pacheco, D., Cotas, J. & Pereira, L. *et al.* Plankton: Environmental and economic importance for a sustainable future. In *Plankton Communities* (2022).

Barth, A., Blanco-Bercial, L., Johnson, R. & Stone, J. Rhizaria in the oligotrophic ocean exhibit clear temporal and vertical variability. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 212, 104371 (2024).

Batchelder, H. P. *et al.* Climate impacts on zooplankton population dynamics in coastal marine ecosystems. *Oceanography* 26, 34–51 (2013).

Bi, H. *et al.* Temporal characteristics of plankton indicators in coastal waters: High-frequency data from PlanktonScope. *J. Sea Res.* 189, 102283 (2022).

Boltovskoy, D. *et al.* Marine zooplanktonic diversity: A view from the South Atlantic. *Oceanol. Acta* 25, 271–278 (2003).

Bressan, P., Ganis, G. & Vallortigara, G. The role of depth stratification in the solution of the aperture problem. *Perception* 22, 215–228 (1993).

Chen, Z., Sun, J., Gu, T., Zhang, G. & Wei, Y. Nutrient ratios driven by vertical stratification regulate phytoplankton community structure in the oligotrophic western Pacific Ocean. *Ocean Sci.* 17, 1775–1789 (2021).

Coelho-Souza, S. A., López, M. S., Guimarães, J. R. D., Coutinho, R. & Candella, R. N. Biophysical interactions in the Cabo Frio upwelling system, Southeastern Brazil. *Braz. J. Oceanogr.* 60, 353–365 (2012).

Couret, M. *et al.* Mesozooplankton size structure in the Canary Current System. *Mar. Environ. Res.* 188, 105976 (2023).

De Boyer Montegut, C., Madec, G., Fischer, A. S., Lazar, A. & Iudicone, D. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J. Geophys. Res.* 109, C12003 (2004).

De Figueiredo, G. G. A. A. *et al.* Zooplankton abundance and biovolume size-spectra in the western tropical Atlantic—From the shelf towards complex oceanic current systems. *Mar. Environ. Res.* 204, 106906 (2025).

Dorrell, R. M. *et al.* Anthropogenic mixing in seasonally stratified shelf seas by offshore wind farm infrastructure. *Front. Mar. Sci.* 9, 830927 (2022).

Ferreira, A. *et al.* Phytoplankton response to nutrient pulses in an upwelling system assessed through a microcosm experiment (Algarrobo Bay, Chile). *Ocean Coast. Manag.* 190, 105167 (2020).

Gonzalez-Rodriguez, E., Valentin, J. L., André, D. L. & Jacob, S. A. Upwelling and downwelling at Cabo Frio (Brazil): comparison of biomass and primary production responses. *J. Plankton Res.* 14, 289–306 (1992).

Gupta, A. S. *et al.* Drivers and impacts of the most extreme marine heatwaves events. *Sci. Rep.* 10, 19359 (2020).

Holt, J. *et al.* Climate‐driven change in the north Atlantic and arctic oceans can greatly reduce the circulation of the North Sea. *Geophys. Res. Lett.* 45, 11–827 (2018).

Horn, H. G., van Rijswijk, P., Soetaert, K. & van Oevelen, D. Drivers of spatial and temporal micro-and mesozooplankton dynamics in an estuary under strong anthropogenic influences (The Eastern Scheldt, Netherlands). *J. Sea Res.* 192, 102357 (2023).

Teixeira, I. G., Crespo, B. G., Nielsen, T. G. & Figueiras, F. G. Stratification-mixing cycles and plankton dynamics in a shallow estuary (Limfjord, Denmark). *J. Plankton Res.* 36, 475–489 (2014).

Li, Y. *et al.* Bottom-up and top-down effects on phytoplankton communities in two freshwater lakes. *PLoS One* 15, e0231357 (2020).

Litzow, M. A. & Ciannelli, L. Oscillating trophic control induces community reorganization in a marine ecosystem. *Ecol. Lett.* 10, 1124–1134 (2007).

Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* 2, 18–22 (2002).

Long, Y. *et al.* Western pacific zooplankton community along latitudinal and equatorial transects in autumn 2017 (northern hemisphere). *Diversity* 13, 58 (2021).

Lynam, C. P. *et al.* Interaction between top-down and bottom-up control in marine food webs. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1952–1957 (2017).

Matos, T. D. S. *et al.* Early warning indicators of decadal shifts in the planktonic assemblage of the Cabo Frio upwelling ecosystem. *Ecol. Indic.* 167, 112674 (2024).

Mena, C. *et al.* Phytoplankton community structure is driven by stratification in the oligotrophic Mediterranean Sea. *Front. Microbiol.* 10, 1698 (2019).

Miloslavich, P. *et al.* Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Glob. Change Biol.* 24, 2416–2433 (2018).

Mitra, A. & Flynn, K. J. Predator-prey interactions: is 'ecological stoichiometry' sufficient when good food goes bad? *J. Plankton Res.* 27, 393–399 (2005).

Mojica, K. D. *et al.* Phytoplankton community structure in relation to vertical stratification along a north–south gradient in the Northeast Atlantic Ocean. *Limnol. Oceanogr.* 60, 1498–1521 (2015).

Picco, P., Schiano, M. E., Pensieri, S. & Bozzano, R. Time-frequency analysis of migrating zooplankton in the Terra Nova Bay polynya (Ross Sea, Antarctica). *J. Mar. Syst.* 166, 172–183 (2017).

R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria (2022).

Scherer, M. E. *et al.* Under pressure: an integrated assessment of human activities and their potential impact on the ecosystem components of the Southern Brazilian continental shelf. *npj Ocean Sustain.* 3, 9 (2024).

Steele, J. H., Thorpe, S. A. & Turekian, K. K. *Encyclopedia of Ocean Sciences.* (Academic Press, 2001).

Sutor, M. M., Cowles, T. J., Peterson, W. T. & Pierce, S. D. Acoustic observations of finescale zooplankton distributions in the Oregon upwelling region. *Deep Sea Res. Part II* 52, 109–121 (2005).

Temperoni, B., Isla Naveira, R., Turina, Y. & Massa, A. E. Seasonal and spatial variations in macrozooplankton quality as food for fish in a Southwest Atlantic Ocean gulf – the role of lipids and fatty acids. *Mar. Biol.* 171, 159 (2024).

Valentin, J. L. The Cabo Frio upwelling system, Brazil. In *Coastal Marine Ecosystems of Latin America* 97–105 (Springer, Berlin, Heidelberg, 2001).

Van der Zwaan, G. J., Jorissen, F. J. & De Stigter, H. C. The depth dependency of planktonic/benthic foraminiferal ratios: constraints and applications. *Mar. Geol.* 95, 1–16 (1990).

Xue, T. *et al.* Mixed layer depth promotes trophic amplification on a seasonal scale. *Geophys. Res. Lett.* 49, e2022GL098720 (2022).

Zhao, W. *et al.* Characteristics of zooplankton community structure and its relationship with environmental factors in the South Yellow Sea. *Mar. Pollut. Bull.* 176, 113471 (2022).

Supplementary material

Mapping marine plankton abundance along ecological gradients

Carolina Reis[1*], Lucas T. Nunes[1], Thiago da S. Matos[1], Guillem Chust[2] , Ricardo Coutinho[1], Patrizio Mariani[3], Marinez E. G. Scherer[4], Lohengrin Fernandes[1]

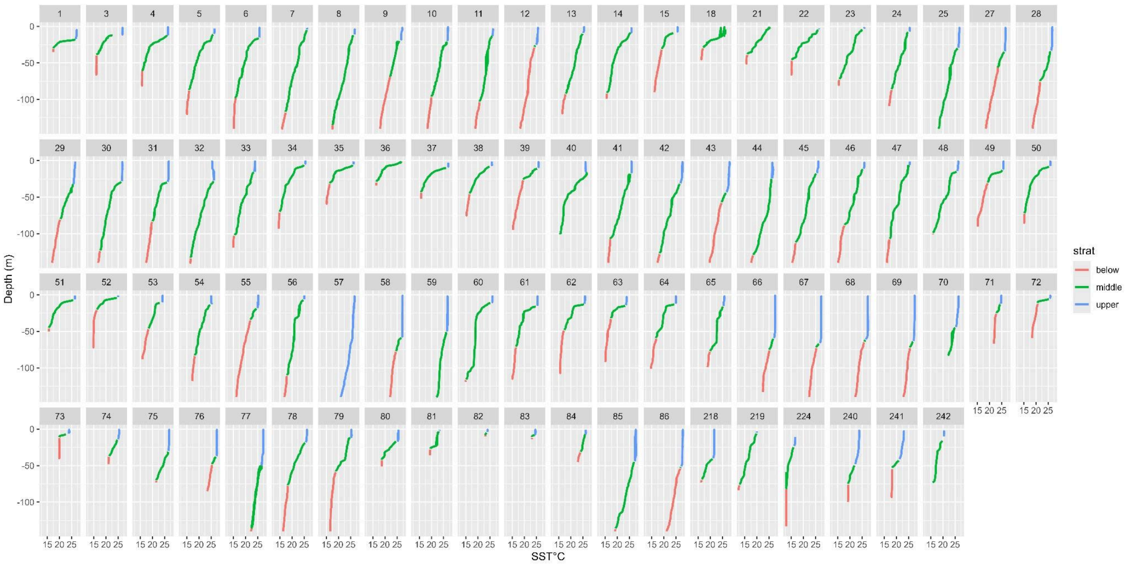

Figure S1. Stratification profiles for the 92 oceanographic stations. Temperature data was used to define the following stratification: Mixed Layer Depth (MLD; Blue), Thermocline (Green) and Below thermocline (red).

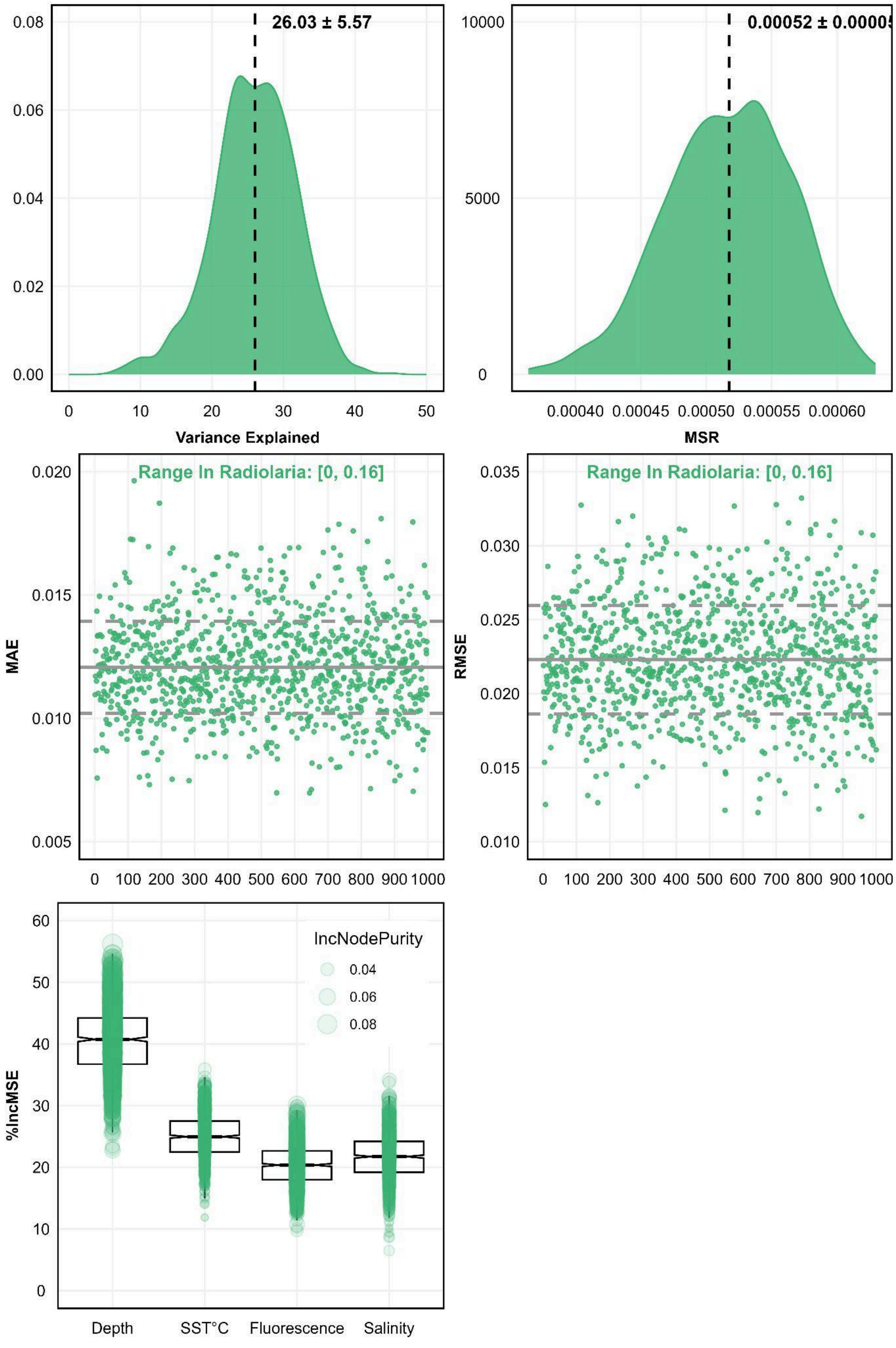

Figure S2. Rhizaria model's performance metrics, including Variance Explained, Mean Squared Residual, validation metrics (MAE and RMSE), and variable importance indices (%IncMSE and IncNodePurity). Vertical grey dashed lines show the robust mean ± robust standard deviation for Variance Explained and Mean Squared Residual, calculated via Huber's estimator.

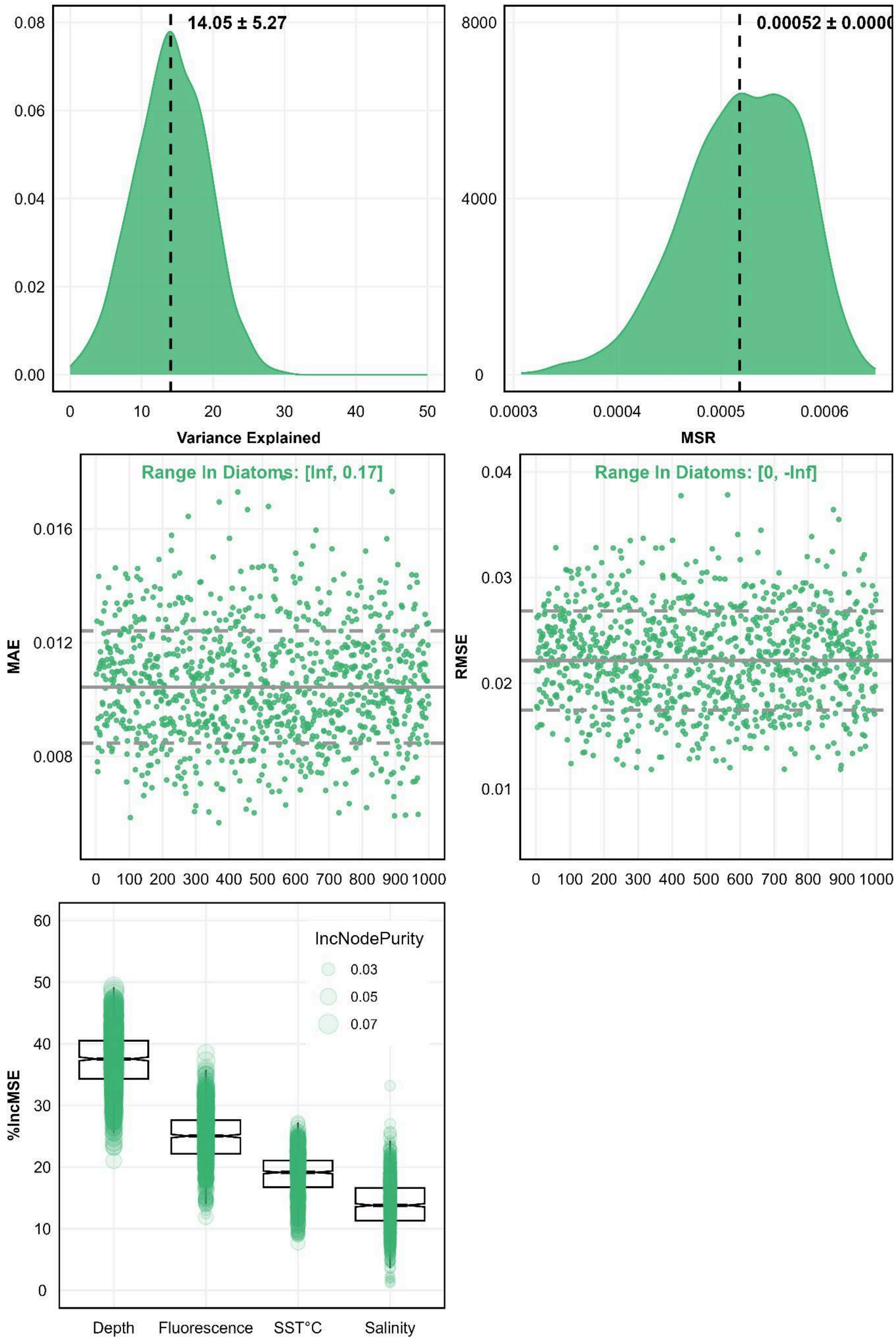

Figure S3. Diatoms model's performance metrics, including Variance Explained, Mean Squared Residual, validation metrics (MAE and RMSE), and variable importance indices (%IncMSE and IncNodePurity). Vertical grey dashed lines show the robust mean ± robust standard deviation for Variance Explained and Mean Squared Residual, calculated via Huber's estimator.

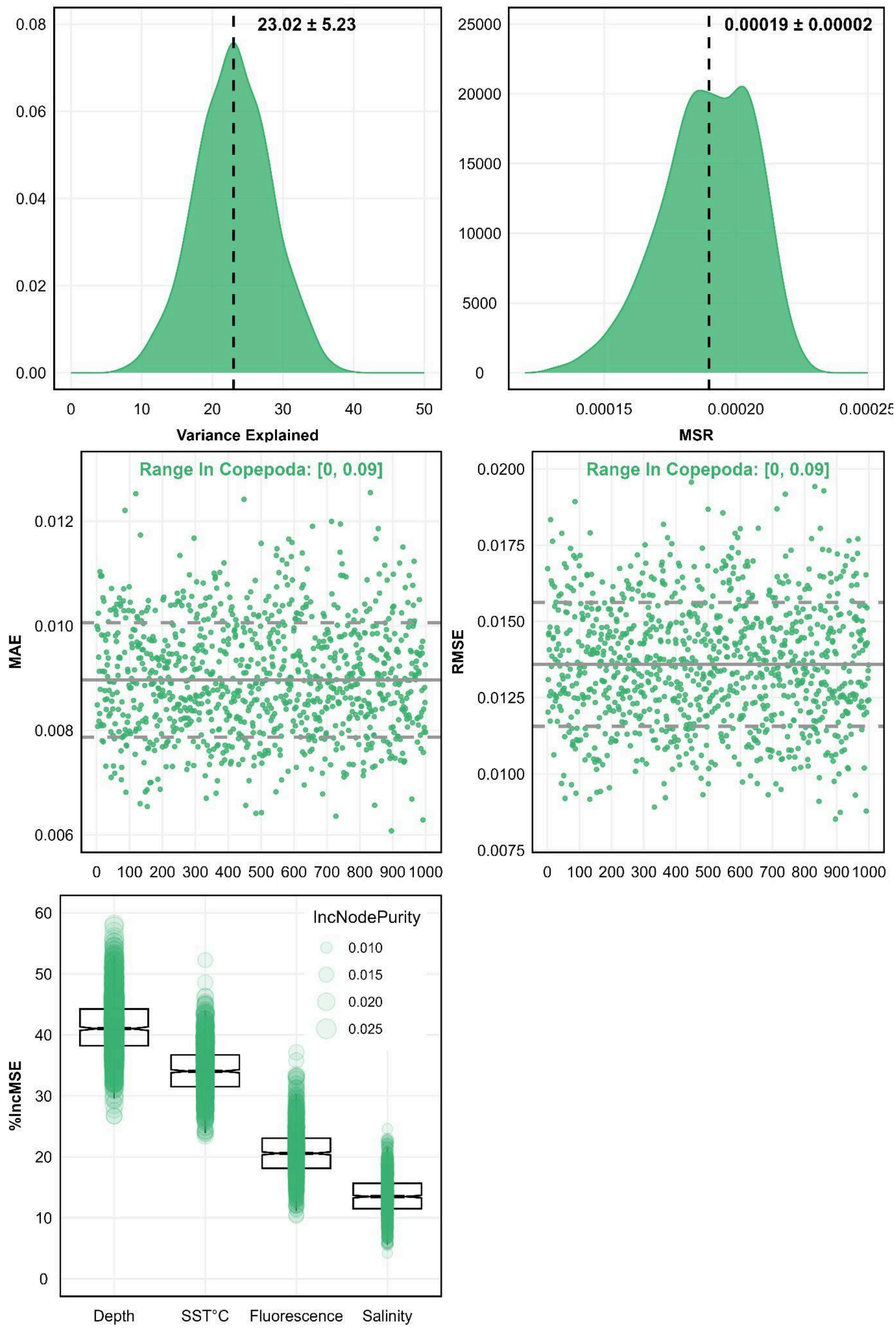

Figure S4. Copepoda model's performance metrics including Variance Explained, Mean Squared Residual, validation metrics (MAE and RMSE), and variable importance indices (%IncMSE and IncNodePurity). Vertical grey dashed lines show the robust mean ± robust standard deviation for Variance Explained and Mean Squared Residual, calculated via Huber's estimator.

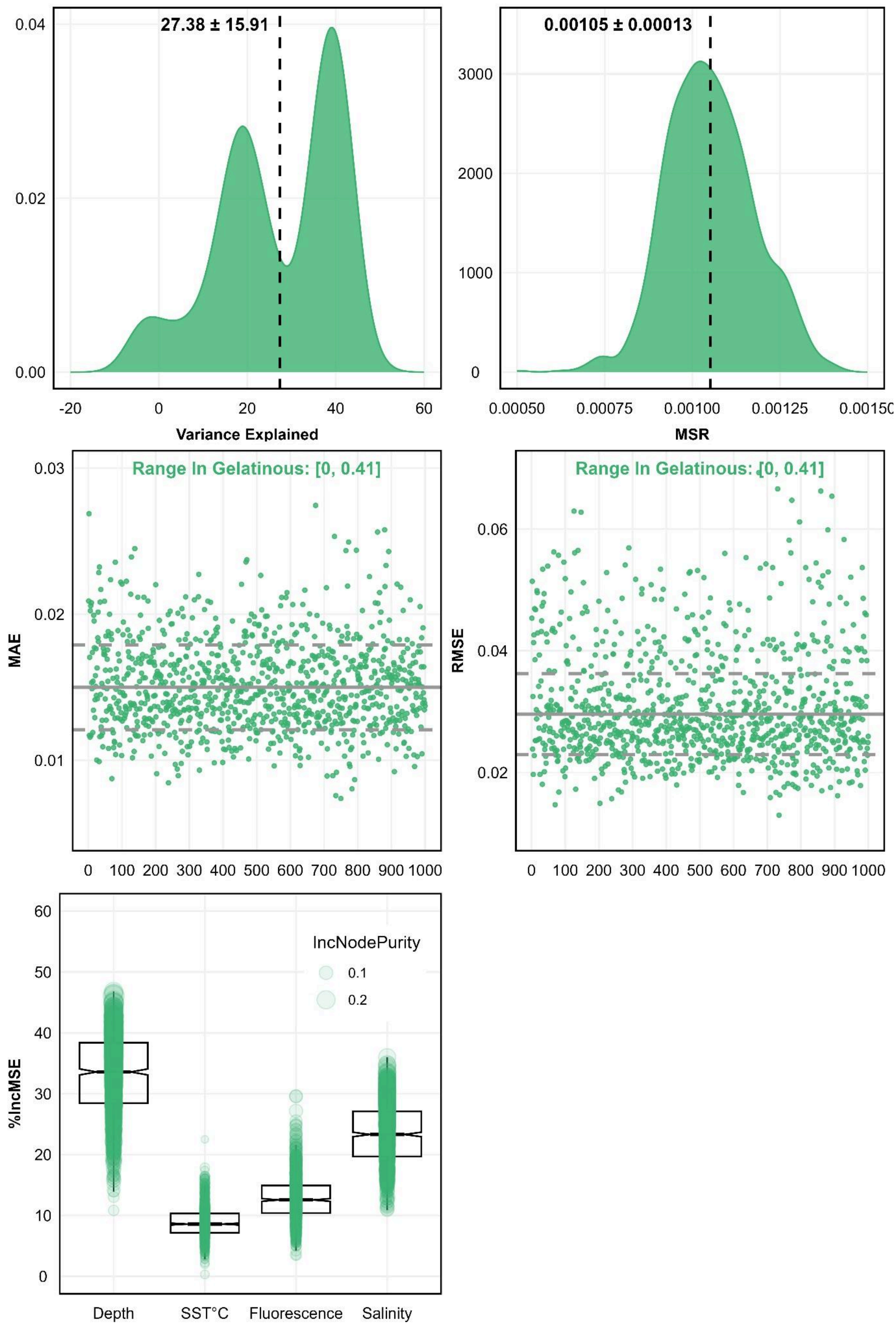

Figure S5. Gelatinous model's performance metrics including Variance Explained, Mean Squared Residual, validation metrics (MAE and RMSE), and variable importance indices (%IncMSE and IncNodePurity). Vertical grey dashed lines show the robust mean ± robust standard deviation for Variance Explained and Mean Squared Residual, calculated via Huber's estimator.

CONCLUSÕES

Em um mundo cada vez mais dinâmico, o volume de informações sobre a biodiversidade marinha se multiplica diariamente, devido a uma ciência cada vez mais tecnológica. Do fundo do mar até a órbita terrestre, diferentes ferramentas nos permitem explorar, mapear e monitorar a saúde dos oceanos a partir de dados como a diversidade e a distribuição do plâncton, base da teia trófica marinha. A partir dessas informações, somos capazes de estimar estoques pesqueiros, gerar prognósticos da saúde dos oceanos, mitigar impactos ambientais que refletem nos oceanos e, possivelmente, amortecer os impactos das mudanças climáticas globais sobre a biodiversidade.

Embora o volume de dados gerados a cada segundo seja imensurável, boa parte desses dados não são encontrados facilmente. Não estão disponíveis em plataformas abertas, tampouco formatados seguindo o padrão internacional de dados de biodiversidade. Ou seja, não estão em conformidade com os princípios FAIR - *Findable, Accessible, Interoperable, Reusable.* Estes princípios ajudam no fortalecimento de uma ciência colaborativa, garantindo que as informações sejam robustas, comparáveis, reprodutíveis e justas. Superam desafios na integração, no uso e na análise de dados referentes aos ambientes marinhos, sustentando tomadas de decisões baseadas em evidências.

Nessa perspectiva, no Capítulo 1, nossos resultados destacaram a importância da utilização de sistemas de harmonização de dados pré-definidos por iniciativas internacionais e amparados pelos princípios FAIR. O desenvolvimento de uma estrutura especializada para dados gerados pelo sistema LOKI é o primeiro passo para que haja uma intercomparação entre dados de imagem plâncton. Bem como dados obtidos por qualquer outra ferramenta que faça uso das mesmas recomendações, garantindo a encontrabilidade, acessibilidade, interoperabilidade e reutilização de informações. Ao desenvolver um fluxo de trabalho que garante práticas padronizadas e a utilização de plataformas colaborativas para gerenciamento de dados dentro do contexto da biodiversidade marinha, tornamos mais fácil para os pesquisadores acessarem dados e se beneficiarem da experiência diversificada e complementar de toda comunidade científica.

Nessa linha de raciocínio, no Capítulo 2, utilizamos o banco de dados, padronizados no meu primeiro capítulo, para estudar a relação da estratificação dos oceanos com a distribuição do plâncton. Nesse capítulo, modelamos a distribuição tridimensional, que destacou a profundidade como um fator-chave para todos os grupos taxonômicos abordados no trabalho, ao mesmo tempo que revelam respostas específicas dos grupos à estratificação da coluna de água. Os gelatinosos estão notavelmente confinados à camada mista e dominam as regiões oligotróficas, independentemente da disponibilidade de nutrientes costeiros. Em contraste, os copépodes exibem uma distribuição heterogênea e contínua, com maior abundância perto de áreas costeiras ricas em nutrientes. Estes padrões contrastantes sublinham a intrincada dinâmica que governa as comunidades planctônicas e reforçam a necessidade de investigações detalhadas para compreender melhor os seus papéis ecológicos e respostas às mudanças ambientais.

Futuramente, pretende-se expandir esta pesquisa integrando novas abordagens para explorar em condições ainda mais detalhadas, táxons que foram amostrados e curiosamente ainda não possuem literatura extensa na região da costa brasileira. Além disso, ampliar as áreas de estudo utilizando a aplicação de modelagem preditiva para avaliar possíveis mudanças na distribuição do plâncton em diferentes cenários.

REFERÊNCIAS BIBLIOGRÁFICAS

Assunção, R., Lebourges-Dhaussy, A., da Silva, A.C., Roudaut, G., Ariza, A., Eduardo, L.N., Queiroz, S. and Bertrand, A., 2023. Fine-scale vertical relationships between environmental conditions and sound scattering layers in the Southwestern Tropical Atlantic. Plos one, 18(8), p.e0284953.

Alcaraz, Miquel, Enric Saiz, Albert Calbet, Isabel Trepat, and Elisabetta Broglio. "Estimating zooplankton biomass through image analysis." *Marine Biology* 143 (2003): 307-315.

Barton, A.D., Pershing, A.J., Litchman, E., Record, N.R., Edwards, K.F., Finkel, Z.V., Kiørboe, T. and Ward, B.A., 2013. The biogeography of marine plankton traits. *Ecology letters*, *16*(4), pp.522-534.

Bar-On, Y.M., and R. Milo. 2019. The biomass composition of the oceans: A blueprint of our blue planet. Cell 179(7):1,451–1,454, https://doi.org/10.1016/j.cell.2019.11.018.

Barnes, H., 1951. Horizontal plankton hauls. ICES Journal of Marine Science, 17(2), pp.133-139.

Brierley, A.S., 2014. Diel vertical migration. Current biology, 24(22), pp.R1074-R1076.

Carvalho, A., Kerr, R., Tavano, V.M. et al. The Southwestern South Atlantic continental shelf biogeochemical divide. Biogeochemistry 159, 139–158 (2022). https://doi.org/10.1007/s10533-022-00918-8

Coelho-Souza, S.A., López, M.S., Guimarães, J.R.D., Coutinho, R. and Candella, R.N., 2012. Biophysical interactions in the Cabo Frio upwelling system, Southeastern Brazil. Brazilian Journal of Oceanography, 60, pp.353-365.

Dipper F., 2022. Elements of marine ecology. Butterworth-Heinemann.

Eerola, T., Batrakhanov, D., Barazandeh, N.V., Kraft, K., Haraguchi, L., Lensu, L., Suikkanen, S., Seppälä, J., Tamminen, T. and Kälviäinen, H., 2024. Survey of automatic plankton image recognition: challenges, existing solutions and future perspectives. *Artificial Intelligence Review*, *57*(5), p.114.

Ellen, J.S., Graff, C.A. and Ohman, M.D., 2019. Improving plankton image classification using context metadata. Limnology and Oceanography: Methods, 17(8), pp.439-461.

Falkowski, P.G.,R.T. Barber, and V. Smetacek, 1998: Biogeochemical controls and feedback on ocean primary production. Science, 281, 200–206, https://doi.org/10.1126/science.281. 5374.200.

Fenchel, T., 1988. Marine plankton food chains. Annual Review of Ecology and Systematics, pp.19-38.

Forward, R.B., 1988. Diel vertical migration: zooplankton photobiology and behaviour. Oceanogr. Mar. Biol. Annu. Rev, 26(36), pp.1-393.

Garcia, T.M., Santos, N.M., Campos, C.C., Costa, G.A., Belmonte, G., Rossi, S. and Soares, M.O., 2021. Plankton net mesh size influences the resultant diversity and abundance estimates of copepods in tropical oligotrophic ecosystems. Estuarine, Coastal and Shelf Science, 249, p.107083.

Gardner, W.D., Mishonov, A.V. and Richardson, M.J., 2018. Decadal comparisons of particulate matter in repeat transects in the Atlantic, Pacific, and Indian Ocean basins. *Geophysical Research Letters*, *45*(1), pp.277-286.

Giering, S.L.C., Cavan, E.L., Basedow, S.L., Briggs, N., Burd, A.B., Darroch, L.J., Guidi, L., Irisson, J.O., Iversen, M.H., Kiko, R. and Lindsay, D., 2020. Sinking organic particles in the ocean—flux estimates from in situ optical devices. *Frontiers in Marine Science*, *6*, p.834.

Gonzalez-Rodriguez, E., Valentin, J.L., André, D.L. and Jacob, S.A., 1992. Upwelling and downwelling at Cabo Frio (Brazil): comparison of biomass and primary production responses. Journal of Plankton Research, 14(2), pp.289-306.

Hall, D.J., Threlkeld, S.T., Burns, C.W. and Crowley, P.H., 1976. The size-efficiency hypothesis and the size structure of zooplankton communities. *Annual Review of Ecology and Systematics*, pp.177-208.

Hays, G.C., Richardson, A.J. and Robinson, C., 2005. Climate change and marine plankton. *Trends in ecology & evolution*, *20*(6), pp.337-344.

Herman, A.W., Beanlands, B. and Phillips, E.F., 2004. The next generation of optical plankton counter: the laser-OPC. *Journal of Plankton Research*, *26*(10), pp.1135-1145.

Irisson, J.O., Ayata, S.D., Lindsay, D.J., Karp-Boss, L. and Stemmann, L., 2022. Machine learning for the study of plankton and marine snow from images. Annual Review of Marine Science, 14(1), pp.277-301.

Krause, J.W., Brzezinski, M.A., Largier, J.L., McNair, H.M., Maniscalco, M., Bidle, K.D., Allen, A.E. and Thamatrakoln, K., 2020. The interaction of physical and biological factors drives phytoplankton spatial distribution in the northern California Current. Limnology and oceanography, 65(9), pp.1974-1989.

Kwon, E.Y., Primeau, F. and Sarmiento, J.L., 2009. The impact of remineralization depth on the air–sea carbon balance. *Nature Geoscience*, *2*(9), pp.630-635.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

Li, G., Cheng, L., Zhu, J., Trenberth, K.E., Mann, M.E. and Abraham, J.P., 2020. Increasing ocean stratification over the past half-century. Nature Climate Change, 10(12), pp.1116-1123.

Lindemann, C., Visser, A. and Mariani, P., 2017. Dynamics of phytoplankton blooms in turbulent vortex cells. *Journal of the Royal Society Interface*, *14*(136), p.20170453.

Litchman, E., de Tezanos Pinto, P., Edwards, K.F., Klausmeier, C.A., Kremer, C.T. and Thomas, M.K., 2015. Global biogeochemical impacts of phytoplankton: a trait‐based perspective. Journal of ecology, 103(6), pp.1384-1396.

Lombard, F., Boss, E., Waite, A.M., Vogt, M., Uitz, J., Stemmann, L., Sosik, H.M., Schulz, J., Romagnan, J.B., Picheral, M. and Pearlman, J., 2019. Globally consistent quantitative observations of planktonic ecosystems. Frontiers in Marine Science, 6, p.196.

Luo, J.Y., Irisson, J.O., Graham, B., Guigand, C., Sarafraz, A., Mader, C. and Cowen, R.K., 2018. Automated plankton image analysis using convolutional neural networks. Limnology and Oceanography: methods, 16(12), pp.814-827.

Matos, T. D. S. *et al.* Early warning indicators of decadal shifts in the planktonic assemblage of the Cabo Frio upwelling ecosystem. *Ecol. Indic.* 167, 112674 (2024).

Martin, J.H., Knauer, G.A., Karl, D.M. and Broenkow, W.W., 1987. VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, *34*(2), pp.267-285

McManus, M.A. and Woodson, C.B., 2012. Plankton distribution and ocean dispersal. Journal of Experimental Biology, 215(6), pp.1008-1016.

Omori, M. and Ikeda, T., 1984. Methods in marine zooplankton ecology. John-Wiely and Sone. *New York*.

Orenstein, E.C. and Beijbom, O., 2017, March. Transfer learning and deep feature extraction for planktonic image data sets. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1082-1088). IEEE.

Orenstein, E.C., Ayata, S.D., Maps, F., Becker, É.C., Benedetti, F., Biard, T., de Garidel‐Thoron, T., Ellen, J.S., Ferrario, F., Giering, S.L. and Guy‐Haim, T., 2022. Machine learning techniques to characterize functional traits of plankton from image data. Limnology and oceanography, 67(8), pp.1647-1669.
Passow, U. and Carlson, C.A., 2012. The biological pump in a high CO2 world. Marine Ecology Progress Series, 470, pp.249-271.

Penninck S. B., Rubens M Lopes, Short-term variability in plankton abundance on the inner shelf off Ubatuba, Brazil, Journal of Plankton Research, Volume 45, Issue 3, May/June 2023, Pages 485–508, https://doi.org/10.1093/plankt/fbad009

Sailley, S.F., Polimene, L., Mitra, A., Atkinson, A. and Allen, J.I., 2015. Impact of zooplankton food selectivity on plankton dynamics and nutrient cycling. Journal of Plankton Research, 37(3), pp.519-529.

Sarmiento, J.L. and Gruber, N., 2002. Sinks for anthropogenic carbon. Physics today, 55(8), pp.30-36.

Titocci J, Pata PR, et al. (2025) Pathways for converting zooplankton traits to ecological insights are paved with findable, accessible, interoperable, and reusable (FAIR) data practices. ICES Journal of Marine Science 82 (2): 017.

Tosetto, Arnaud Bertrand, Sigrid Neumann-Leitão, Alex Costa da Silva, Miodeli Nogueira Júnior, Spatial patterns in planktonic cnidarian distribution in the western boundary current system of the tropical South Atlantic Ocean, Journal of Plankton Research, Volume 43, Issue 2, March/April 2021, Pages 270–287, https://doi.org/10.1093/plankt/fbaa066

Valentin, J.L., 1984. Spatial structure of the zooplankton community in the Cabo Frio region (Brazil) influenced by coastal upwelling. Tropical Zooplankton, pp.183-199.

Valentin, J.L., 1999. Numerical modelling of phytoplankton bloom in the upwelling ecosystem of Cabo Frio (Brazil). Ecological Modelling, 116(2-3), pp.135-148.

Valentin, J.L., 2001. The cabo frio upwelling system, Brazil. In Coastal marine ecosystems of Latin America (pp. 97-105). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wadjikar, B., Masram, A., Rai, M.M. and Darvekar, A., 2017. Monitoring Water Quality Using Zooplankton Organisms As Bioindicators In The Well Water Of Nagpur City (MH) India. *International Journal of Researches In Biosciences, Agriculture and Technology*, *5*(13).

Wieczorek J, Bloom D, Guralnick R, Blum S, et al. (2012) Darwin Core: An evolving community-developed biodiversity data standard. PLOS One 7 (1). https://doi.org/ 10.1371/journal.pone.0029715

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), pp.1-9.

Xue, T. *et al.* Mixed layer depth promotes trophic amplification on a seasonal scale. *Geophys. Res. Lett.* 49, e2022GL098720 (2022).