



**HAL**  
open science

# Analyse globale de la pompe à carbone biologique à partir de données en imagerie quantitative

Laetitia Drago

► **To cite this version:**

Laetitia Drago. Analyse globale de la pompe à carbone biologique à partir de données en imagerie quantitative. Biodiversité et Ecologie. Sorbonne Université, 2023. Français. NNT : 2023SORUS562 . tel-04483392

**HAL Id: tel-04483392**

**<https://theses.hal.science/tel-04483392>**

Submitted on 29 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

Ecole doctorale des Sciences et de l'Environnement d'Ile de France (ED 129)

Laboratoire d'Océanographie de Villefranche UMR 7093

---

# ANALYSE GLOBALE DE LA POMPE À CARBONE BIOLOGIQUE À PARTIR DE DONNÉES EN IMAGERIE QUANTITATIVE

---

Thèse présentée par

Laetitia Drago

pour l'obtention du grade de Docteur de Sorbonne Université  
Spécialité : Océanologie biologique et Écologie marine

Sous la supervision de Lars Stemmann et Rainer Kiko

Soutenue le 8 décembre 2023

Devant un jury composé de :

Stephanie Henson  
Barbara Niehoff  
Sakina-Dorothee Ayata  
Colleen Durkin  
Brivaëlla Moriceau  
Louis Legendre  
Lars Stemmann  
Rainer Kiko

Rapportrice  
Rapportrice  
Examinatrice  
Examinatrice  
Examinatrice  
Président  
Directeur de thèse  
Co-directeur de thèse

SORBONNE UNIVERSITÉ

École Doctorale des Sciences de l'Environnement d'Île de France  
Laboratoire d'Océanographie de Villefranche - UMR 7093

# Image-based global analysis of the biological carbon pump

**Laetitia Drago**

Under the supervision of  
**Lars Stemmann**  
**Rainer Kiko**

**PhD thesis in Biological Oceanology and Marine Ecology**

**Defended on 8<sup>th</sup> December 2023**

**The jury comprised:**

Stephanie Henson	Reviewer
Barbara Niehoff	Reviewer
Sakina-Dorothee Ayata	Examiner
Colleen Durkin	Examiner
Brivaëlla Moriceau	Examiner
Louis Legendre	President
Lars Stemmann	PhD Supervisor
Rainer Kiko	PhD Co-Supervisor



À ma famille.  
Que nous soyons proches  
ou séparés par une mer ou un océan,  
je vous ai toujours dans mon coeur.



*"Even if you never have the chance to see or touch the ocean, the ocean touches you with every breath you take, every drop of water you drink, every bite you consume. Everyone, everywhere is inextricably connected to and utterly dependent upon the existence of the sea."*

- Sylvia Earle, marine biologist and oceanographer,  
The World Is Blue: How Our Fate and the Ocean's Are One

*"The history of life on earth has been a history of interaction between living things and their surroundings."*

- Rachel Carson, Marine Biologist, Silent Spring

*"If you know you are on the right track, if you have this inner knowledge, then nobody can turn you off... no matter what they say."*

- Barbara McClintock, Cytogeneticist and winner of the 1983 Nobel Prize in Physiology.

*"Don't let anyone rob you of your imagination, your creativity, or your curiosity. It's your place in the world; it's your life. Go on and do all you can with it, and make it the life you want to live."*

- Mae Jemison, Engineer, Physician and Former NASA Astronaut





# Abstract

---

The biological carbon pump (BCP) plays a central role in the global ocean carbon cycle, transporting carbon from the surface to the deep ocean and sequestering it for long periods. This work aims to analyse two key players of the BCP: zooplankton and particles. To this end, we use *in situ* imaging data from the Underwater Vision Profiler (UVP5) to investigate two primary axes: 1) the global distribution of zooplankton biomass and 2) carbon export in the context of a North Atlantic spring bloom. Our objectives include a quantification of global zooplankton biomass, enhancing our comprehension of the BCP via morphological analysis of particles, and assessing and comparing the gravitational flux of detrital particles during a the North Atlantic spring bloom using high-resolution UVP5 data.

With the help of UVP5 imagery and machine learning through habitat models using boosted regression trees, we investigate the global distribution of zooplankton biomass and its ecological implications. The results show maximum zooplankton biomass values around 60°N and 55°S and minimum values within the oceanic gyres, with a global biomass dominated by crustaceans and rhizarians. By employing machine learning techniques on globally homogeneous data, this study provides taxonomical insights into the distribution of 19 large zooplankton groups (1-50 mm equivalent spherical diameter). This first protocol estimates global, spatially resolved zooplankton biomass and community composition from *in situ* imaging observations of individual organisms.

In addition, within the unique context of the EXPORTS 2021 campaign, we analyse UVP5 data obtained by deploying three instruments in a highly retentive eddy. After clustering the 1,720,914 images using Morphocluster, a semi-autonomous classification software, we delve into the characteristics of the marine particles, studying their morphology through an oblique framework that follows a plume of detrital particles between the surface and 800 m depth. The results of the plume following approach show that, contrary to expectations, aggregates become unexpectedly larger, denser, more circular and more complex with depth. In contrast, the evolution of fecal pellets is more heterogeneous and shaped by zooplankton activity. Such results challenge previous expectations and may require a reassessment of our view of sinking aggregates and fecal pellets.

We also studied concentration and carbon flux dynamics using a more traditional 1D framework where we explore the three key elements in flux estimation from *in situ* imaging data by comparing UVP5 and sediment trap flux estimates: size range covered, sinking rate and carbon content. According to the current literature, neutrally buoyant sediment traps (NBST) and surface-tethered traps (STT) usually cover a size range from 10  $\mu\text{m}$  to approximately 2 mm. In our study, we have found that by expanding the UVP size range to 10  $\mu\text{m}$  and limiting it to 2 mm, a more consistent comparison can be made between UVP5-generated flux and sediment trap fluxes (obtained by colleagues). However, it is worth noting that there remains a large flux contribution above this size threshold, necessitating further investigation of its implications through the use of complementary approaches such as the use of sediment traps with larger openings.

This manuscript not only advances our knowledge, but also addresses critical challenges in estimating zooplankton biomass and particle dynamics during export events. The findings of this study open up new avenues for future research on the biological carbon pump and deepen our understanding of marine ecosystems.

# Résumé

---

La pompe à carbone biologique (PCB) joue un rôle central dans le cycle global du carbone océanique, en transportant le carbone de la surface vers les profondeurs et en le séquestrant pendant de longues périodes. Ce travail vise à analyser deux acteurs clés de la PCB : le zooplancton et les particules. Pour cela, nous utilisons les données d'imagerie *in situ* de l'Underwater Vision Profiler (UVP5) pour étudier deux axes principaux : 1) la distribution globale de la biomasse du zooplancton et 2) l'exportation de carbone dans le contexte d'une efflorescence printanière dans l'Atlantique Nord.

À l'aide de l'UVP5 et de l'apprentissage automatique par le biais de modèles d'habitat utilisant des arbres de régression boostés, nous étudions la distribution mondiale de la biomasse du zooplancton et ses implications écologiques. Les résultats montrent des valeurs maximales de biomasse autour de 60°N et 55°S et des valeurs minimales au niveau des gyres océaniques, avec une biomasse globale dominée par les crustacés et les rhizaires. En utilisant des techniques d'apprentissage automatique sur des données globalement homogènes, cette étude fournit des informations sur la distribution de 19 grands groupes de zooplancton (1-50 mm de diamètre sphérique équivalent). Ce premier protocole permet d'estimer la biomasse du zooplancton et la composition de la communauté à l'échelle globale à partir d'observations d'imagerie *in situ* d'organismes individuels.

Dans le contexte unique de la campagne EXPORTS 2021, nous analysons les données UVP5 obtenues par le déploiement de trois instruments dans un tourbillon à forte rétention. Après avoir regroupé les 1 720 914 images à l'aide de Morphocluster, un logiciel de classification semi-autonome, nous nous intéressons aux caractéristiques des particules marines, en étudiant leur morphologie à travers un cadre oblique qui suit un panache de particules entre la surface et 800 m. Les résultats montrent que, contrairement aux attentes, les agrégats deviennent de manière inattendue plus grands, plus denses, plus circulaires et plus complexes avec la profondeur. En revanche, l'évolution des pelottes fécales est plus hétérogène et façonnée par l'activité du zooplancton. Ces résultats remettent en question les attentes antérieures et appellent à une réévaluation de notre vision des agrégats et des pelottes fécales.

Nous avons également étudié la dynamique des concentrations et des flux de carbone à l'aide d'un cadre 1D plus traditionnel dans lequel nous explorons les trois éléments clés de l'estimation des flux à partir d'imagerie *in situ* en comparant les estimations de l'UVP5 et des pièges à sédiments: la gamme de tailles couvertes, la vitesse de sédimentation et le contenu en carbone. Selon la littérature, les pièges à sédiments à flottabilité neutre (NBST) et les pièges attachés à la surface (STT) couvrent généralement une gamme de tailles allant de 10  $\mu\text{m}$  à environ 2 mm. Dans notre étude, nous avons constaté qu'en élargissant la gamme de tailles de l'UVP5 à 10  $\mu\text{m}$  et en la limitant à 2 mm, une comparaison plus consistante peut être faite entre le flux issu de l'UVP5 et celui des pièges à sédiments (obtenus par des collègues). Toutefois, il reste une contribution importante du flux au-dessus de ce seuil de taille qui nécessite une étude plus approfondie de ses implications par l'utilisation d'approches complémentaires telles que des pièges à sédiments avec des ouvertures plus grandes.

Ce manuscrit ne fait pas seulement progresser nos connaissances, mais il aborde également des défis critiques dans l'estimation de la biomasse du zooplancton et de la dynamique des particules pendant les événements d'export. Les résultats de cette étude ouvrent de nouvelles voies pour la recherche future sur la PCB et approfondissent notre compréhension des écosystèmes marins.

# Scientific communications

---

## Published papers

- Drago L, Panaiotis T, Irisson J-O, Babin M, Biard T, Carlotti F, Coppola L, Guidi L, Hauss H, Karp-Boss L, Lombard F, McDonnell AMP, Picheral M, Rogge A, Waite AM, Stemmann L and Kiko R (2022) "Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning". *Front. Mar. Sci.* 9:894372. doi: 10.3389/fmars.2022.894372 <https://www.frontiersin.org/articles/10.3389/fmars.2022.894372/full>

## Conference talks

- Drago L, Buesseler K, Karp-Boss L, McDonnell A. M. P, Schröder S, Siegel D. A., Stemmann L and Kiko R, "North Atlantic Biological Pump Dynamics in an Eddy Revealed via in Situ Imaging", Aquatic Sciences Meeting ASLO 2023, Palma, Spain
- Drago L, Panaiotis T, Irisson J-O, Babin M, Biard T, Carlotti F, Coppola L, Guidi L, Hauss H, Karp-Boss L, Lombard F, McDonnell AMP, Picheral M, Rogge A, Waite AM, Stemmann L and Kiko R, "Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning", TRIATLAS General Assembly 2022 (Virtual Meeting)
- Drago L, Panaiotis T, Irisson J-O, Babin M, Biard T, Carlotti F, Coppola L, Guidi L, Hauss H, Karp-Boss L, Lombard F, McDonnell AMP, Picheral M, Rogge A, Waite AM, Stemmann L and Kiko R, "Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning", Aquatic Sciences Meeting ASLO 2021, (Virtual Meeting)



# Acknowledgments

---

To whoever is reading this, thank you for opening this document. I have had the good fortune to encounter a very large number of wonderful people in my life that made it possible for me to do this PhD and write this manuscript, so this is going to be quite long. So, let's begin, shall we?

*A quiconque lit ces lignes, je tiens à vous remercier d'ouvrir ce document. J'ai eu la grande chance de rencontrer un nombre important de personnes merveilleuses tout au long de ma vie, qui ont rendu possible la réalisation de ce doctorat et la rédaction de ce manuscrit. Par conséquent, cette section de remerciements risque d'être assez longue. Alors, commençons, voulez-vous ?*

First of all, I would like to thank my PhD director Lars and co-director Rainer for trusting me with this project. Thank you for your guidance, expertise, and insights throughout my doctoral journey, as well as the opportunities you have made available to me. I am grateful that you allowed me to continue working on the fascinating creatures that are zooplanktonic organisms during these past 3 years.

*Tout d'abord, je tiens à remercier mon directeur de thèse, Lars, ainsi que mon co-directeur, Rainer, de m'avoir fait confiance pour ce projet. Merci pour vos conseils et votre expertise tout au long de mon parcours doctoral, ainsi que pour les opportunités que vous m'avez offertes. Je vous suis reconnaissante de m'avoir permis de poursuivre mes travaux sur les fascinantes créatures que sont les organismes zooplanctoniques au cours de ces trois dernières années.*

I would like to express my sincere gratitude to the members of the jury for accepting to review this manuscript and for their participation in the defense of my doctoral thesis. I am deeply honoured to present to you this manuscript.

*Je tiens à exprimer ma sincère gratitude aux membres du jury d'avoir accepté d'examiner ce manuscrit et pour leur participation à la soutenance de ma thèse doctorale. Je suis profondément honorée de vous présenter ce manuscrit.*

Je tiens à remercier chaleureusement les membres de mon comité de thèse : Fabio Benedetti, Marina Lévy et Anya Waite. Vos conseils ont été très importants dans le développement de mon projet de recherche, et je suis réellement reconnaissante pour les connaissances et l'expertise que vous avez partagées avec moi.

I would like to thank the doctoral school ED129 and the european TRIATLAS H2020 for financing this PhD. Though it is now much less prevalent, especially in oceanography, I have to acknowledge the privilege I had especially since there are still a lot of instances where PhD students do not have scholarships associated with their doctoral contract.

*Je tiens à remercier l'école doctorale ED129 ainsi que le projet européen TRIATLAS H2020 pour avoir financé ce doctorat. Même si c'est aujourd'hui beaucoup moins fréquent, notamment en océanographie, je dois reconnaître le privilège que j'ai eu d'autant plus qu'il existe encore de nombreux cas où les doctorants ne bénéficient pas de bourses associées à leur contrat doctoral.*

Au cours de ces 4 dernières années, j'ai eu la joie et le privilège de faire partie de l'équipe COMPLEX. Je suis très reconnaissante pour les bons moments passés à vos côtés et pour la belle ambiance de travail que vous avez instaurée. Vous avez été une source importante de soutien et d'inspiration (ainsi que d'excellents gâteaux) et je chérirai ces moments pendant de nombreuses années à venir.

Pendant cette thèse, j'ai eu le plaisir de travailler avec de nombreuses personnes que je tiens à saluer ici. Lors de mon stage de M2, j'ai été encadré par le trio formé par Lars, Jean-Olivier Irisson et Thelma Panaïotis. Vous avez toujours été très pédagogue et patient, me permettant de faire de gros progrès durant ces 6 mois qui semblent déjà si loin. Je tiens à réitérer mes remerciements à Raphaëlle Sauzède pour son aide avec la chlorophylle a, à Fabien Lombard pour l'étude des appendiculaires et à Tristan Biard pour l'étude des Rhizaria.

This manuscript presents the data acquired during the EXPORTS 2021 experiment that took place in the North Atlantic. I would like to sincerely thank all the people involved in this gigantic project. It really is impressive to see so many people from all around the globe get together to understand such important oceanic processes especially in the midst of a global pandemic! This campaign would not have been possible without the supportive Captains, Crew, Research Technicians and Administrative Staff of the three ships that brought us on site. I address a special thank you to the Sarmiento de Gamboa team who got us through 3 storms that stopped us for half of the days on site in our scientific endeavours. Thank you for the great moments spent with you from deployments to pinchos and discussions in spanish that have a special place in my heart. I was lucky to have worked with a great scientific team on board: Ken Buesseler, Heidi Sosik, Joel Llopiz, Alexi Shalapyonok, Michelle Cusolito, Marley Parker, John Soucie, Julia Cox, Justin Ossolinski, Kayla Gardner, Erin Frates, Jackson Sugar, Henry Holm, Helena McMonagle. Special thank you to my damas



favoritas: Elena Ceballos Romero, Cris Garcia Fernandez and Jessica Kozik who made this journey extra special. The work on the massive EXPORTS dataset needed for us to use a semi autonomous software to annotate the images. I would like to thank Martin Schroder for his patience and help in using Morphocluster. We then had to upload this tsunami of images to Ecotaxa which was not an easy task. Thank you to the Ecotaxa team, Béatrice Caraveo, Laurent Salinas and Julie Coustenoble, who kindly helped me through this. Thank you also to Emilia Trudnowska for your kindness and your guidance in using your method. The extension to smaller size classes would not have been possible without the help and guidance of Alberto Baudena, thank you!

Two chapters of this manuscript are dedicated to the EXPORTS 2021 experiment but I unfortunately did not have the opportunity to study the data that I helped acquire during a previous scientific campaign that also took place in early 2021 but this time in two fjords near Bergen, Norway. Thank you to all the participants who accompanied me for this first-time scientific campaign experience: Helena Hauss, Klas Möller, Daniel Blandfort, Jan Taucher, Andreas Neumann, Morten Iversen, Dong-gyun Kim, Christian Konrad and Saskia Rühl. Special thank you to Vanessa Stenvers for being such a great cabin mate and sharing your love for jellylike organisms with me.

Durant cette thèse, j'ai également eu l'opportunité d'enseigner à des étudiants en licence et en master sur des sujets allant de la modélisation à la taxonomie. Merci d'avoir fait de ces deux années d'enseignement une expérience aussi amusante et enrichissante ! Je suis vraiment reconnaissant pour le temps que j'ai pu passer avec vous. Merci à Marine, Thelma et Ophélie pour leur aide dans la préparation des nombreux cours pendant ces deux années. Merci à l'équipe pédagogique d'avoir rendu cette expérience aussi positive : Laure Mousseau, Véronique Gourbaud-Stevens, Katia Testard, Didier Jonas, Jean-Olivier Irisson, Fabien Lombard, Lars Stemmann et Rodolphe Lemée.

J'ai également scanné de nombreux échantillons TRIATLAS à l'aide du Zooscan et tiens à remercier l'équipe PIQv pour son aide. Un merci spécial à Laëtitia Jalabert, Amanda Elineau et Louis Caray-Counil pour leur patience et leurs conseils.

Ce parcours académique n'est pas seulement le résultat de ces dernières années. Il a commencé bien plus tôt avec des personnes qui étaient de grands supporters et qui ont fait tout ce qui était en leur pouvoir pour rendre ce chemin formidable. Les premiers furent M. Debarnot, Mme. Gasteau et M. Boffa. Je serai toujours reconnaissant envers vous trois.

Puis des années plus tard, j'ai fait ma licence en biologie des organismes et des écosystèmes à Nice où j'ai rencontré des gens formidables comme Florence, Alessia, Anaïs, Tiffany, Laura, Marion, Sébastien et Alix. Durant ces années, j'ai aussi eu la chance d'avoir de formidables professeurs et mentors parmi lesquels Patrice Francour, un professeur gentil et passionné. Vous avez laissé un impact durable sur moi et même si je suis triste de ne pas avoir l'opportunité de réaliser une thèse sous votre direction, vos photographies et vos cours sont toujours avec moi et je les ai utilisés pour construire certains des cours que j'ai donnés au cours de ces deux dernières années. Je tiens également à remercier John Dolan, Stéphane Noselli, Christian Ghiglione, Luisa Mangialajo et Paolo Guidetti pour leur soutien et pour avoir intensifié en moi le besoin d'en apprendre davantage sur les océans.

Merci à Rodolphe Lemée, directeur du LOV, pour son accueil au sein du laboratoire et merci à toutes les personnes qui participent à son bon fonctionnement.

Ce manuscrit ne serait pas ce qu'il est sans le soutien des grandes amitiés que j'ai eu l'honneur de nouer au cours de ces dernières années. Merci à mes amis du master Science de la Mer avec un spécial pour Anaïs R, Anaïs L, Manon, Julie, Louis, Théotime, Léonard et Alix.

Merci aux meilleurs collègues de bureau devenus de chers amis : Alexandre, Thelma et Florian. Ces amitiés se sont construites autour du sport (vélo, escalade sans oublier la pétanque!), du chocolat, du codage et des plantes (beaucoup de plantes même si Florian n'en était pas très content). Merci pour votre gentillesse et votre soutien à travers tout cela. Merci d'avoir veillé sur moi, surtout au cours de ces derniers mois.

Merci au meilleur trio de femmes que j'ai eu l'honneur d'avoir pendant cette thèse : Anaïs, Chloé et Aurélie. Nous avons vécu ensemble des moments très heureux et des moments difficiles. Votre amitié et votre soutien ont embelli ces dernières années et je suis convaincue que de nombreux souvenirs partagés se trouveront dans nos futurs.

Louise, on n'a eu l'occasion de passer beaucoup de temps ensemble, mais j'ai beaucoup apprécié ces quelques semaines à tes côtés.

J'aimerais également remercier les doctorants et post doctorants avec qui j'ai pu partager des bons moments au laboratoire, dans les jardins ou sur le ponton: Manon, Marine, Flavien, Chloé T., Julien, Loïc, Louis, Ana-Luiza, Zoé, Camille, Dodji, Mathilde, Miriam et Alberto.

Enfin, je tiens à exprimer ma plus profonde gratitude à ma famille pour son soutien indéfectible pendant ces dernières 28 années. Votre amour et votre encouragements ont été une source constante de force et de motivation. J'ai une pensée particulière pour Tonton et Papi qui sont, je l'espère, fiers du chemin que j'ai parcouru. Votre confiance en moi, même dans les moments les plus difficiles, a été cruciale pour en arriver jusque là. Votre amour et votre soutien ont fait toute la différence et je vous en serai éternellement reconnaissante.



# Contents

---

List of Figures [xxiii](#)

List of Tables [xxxv](#)

List of Acronyms [xxxvii](#)

## I General introduction

- 1 General introduction [3](#)
  - 1.1 The role of the Biological Gravitational Pump in the oceanic carbon cycle [3](#)
  - 1.2 The diverse nature of marine particles and their role in oceanic systems [3](#)
  - 1.3 Zooplankton: Key players in marine ecosystems and the Biological Carbon Pump [6](#)
  - 1.4 Shedding light on zooplankton and particles through *in situ* imaging [6](#)
  - 1.5 Machine learning in marine ecology [10](#)
  - 1.6 Objectives of this work [12](#)
    - 1.6.1 Global distribution of zooplankton biomass estimated by *in situ* imaging and machine learning [12](#)
    - 1.6.2 Export in the context of a North Atlantic bloom [13](#)

## II Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning

- 2 Global Zooplankton Biomass [17](#)
  - 2.1 Abstract [17](#)
  - 2.2 Introduction [18](#)
    - 2.2.1 Zooplankton [18](#)
    - 2.2.2 Spatial Distribution of zooplankton and its biomass [18](#)
    - 2.2.3 The study of zooplankton and its difficulties [19](#)
    - 2.2.4 Statistical habitat models [20](#)
  - 2.3 Material and methods [21](#)
    - 2.3.1 Plankton data collection and processing [21](#)
    - 2.3.2 Environmental data collection and processing [24](#)
    - 2.3.3 Habitat modeling [25](#)
  - 2.4 Results [30](#)
    - 2.4.1 Model comparison [30](#)
    - 2.4.2 Group-wise contribution to global zooplankton biomass [32](#)
    - 2.4.3 Spatial distribution patterns and occupied habitat [34](#)

2.4.4	<i>In situ</i> imaging compared to net based sampling	40
2.4.5	Global zooplankton biomass distribution	40
2.5	Discussion	41
2.5.1	Sensitivity of model prediction to partitioning	41
2.5.2	Group-wise contribution to global zooplankton biomass	42
2.5.3	Distribution patterns and occupied habitats	43
2.5.4	Comparison between net sampling and <i>in-situ</i> imaging	46
2.5.5	Global zooplankton biomass distribution	48
2.5.6	Conclusions and outlook	48
2.6	Supplementary Material	52
III North Atlantic biological pump dynamics in an anticyclonic eddy revealed via <i>in situ</i> imaging		
3	Follow the plume	65
3.1	Abstract	65
3.2	Introduction	66
3.2.1	The rain of detritus	66
3.2.2	An ever-changing morphological alteration	66
3.3	Material and Methods	71
3.3.1	Data collection and processing	71
3.3.2	Data selection according to the export plume	73
3.4	Results	76
3.4.1	Evolution of particle concentration and size distribution inside the plume	76
3.4.2	Evaluation of flux and vertical attenuation inside the plume	78
3.4.3	Evolution of the morphological properties of aggregates and fecal pellets inside the plume	80
3.5	Discussion	85
3.5.1	The plume approach	85
3.5.2	Major changes in the particle community inside the plume	86
3.5.3	Evolution of morphological variables inside the plume	90
3.6	Conclusion and perspective	92
3.7	Supplementary Material	95
IV Estimating carbon flux from UVP data: an EXPORTS perspective		
4	EXPORTS carbon flux	109
4.1	Abstract	109
4.2	Introduction	110
4.3	Material and Methods	111

4.3.1	Data collection	111
4.4	Results	115
4.4.1	Data composition	115
4.4.2	Sinking speed and flux estimations	117
4.4.3	Flux comparison between UVP5 and sediment traps	117
4.4.4	Contribution of different size range compartments to export flux	119
4.5	Discussion	121
4.5.1	The dominance of aggregates	121
4.5.2	The most appropriate parameters to estimate flux	121
4.5.3	Flux results	125
4.6	Conclusion and perspectives	127
4.7	Supplementary section	129
v	General discussion and Perspectives	
5	General discussion and Perspectives	133
5.1	General discussion	133
5.1.1	Zooplankton Biogeography and Ecological Impact	133
5.1.2	Particle characteristics and flux patterns: insights from a UVP-based analysis	137
5.1.3	Exploring the boundaries: challenges in studying zooplankton biomass and particle evolution morphologies	140
5.2	Conclusion	141
5.3	Perspectives	142
5.3.1	Navigating the challenges and opportunities of data-driven ecology.	142
5.3.2	Future of global zooplankton biomass research	143
5.3.3	Future prospects for studying particles via the follow the plume approach	144
vi	Résumé en français	
6	Résumé en français	149
6.1	Introduction	149
6.1.1	La Pompe Biologique Gravitationnelle	149
6.1.2	Les particules marines	149
6.1.3	Les rôles du zooplancton dans les écosystèmes marins	150
6.1.4	Etude du zooplancton et des particules grâce à l'imagerie in situ	152
6.1.5	L'apprentissage automatique en écologie marine	152
6.1.6	Objectifs de ma thèse	154

6.2	Distribution mondiale de la biomasse du zooplancton	158
6.2.1	Contexte et méthodes	158
6.2.2	Principales conclusions et perspectives	159
6.3	Dynamique de la pompe biologique dans un tourbillon anticyclonique	161
6.3.1	Contexte et méthodes	161
6.3.2	Principales conclusions et perspectives	163
6.4	Estimation des flux de carbone à partir des données de l'UVP	164
6.4.1	Contexte and méthodes	164
6.4.2	Principales conclusions et perspectives	165
6.5	Discussion	166
6.5.1	La biogéographie et l'impact écologique du zooplancton	166
6.5.2	Caractéristiques des particules et modèles de flux	167
6.5.3	Contraintes dans l'estimation de la biomasse du zooplancton et la morphologie des particules	168
6.6	Perspectives	169
6.6.1	Défis et opportunités des données en écologie	170
6.6.2	Le future de l'étude du zooplancton	170
6.6.3	Perspectives futures pour l'approche de suivi des particules	171
	Bibliography	173



# List of Figures

---

- Figure 1.1 Pathways of cycling and export of carbon by zooplankton in the ocean. Phytoplanktonic organisms take up CO<sub>2</sub> via photosynthesis in the euphotic layer. They are grazed upon by micro and mesozooplankton. The latter participates in the biological carbon pump through the production of sinking fecal pellets at the surface and at depth. Figure taken from Steinberg and Landry (2017) 5
- Figure 1.2 Four strategies to process raw plankton images identified by four circled numbers. Efficient interaction between instruments, computers, and human operators enables high throughput from laboratory- or field-based image acquisition to ecological exploitation. Figure taken from Irisson et al. (2022) 8
- Figure 1.3 Evolution of images uploaded on the Ecotaxa repository (Picheral et al., 2017) using instruments such as the Underwater Vision Profiler (UVP, Picheral et al. (2010, 2021), Imaging Flow-Cytobot (IFCB, Olson and Sosik (2007)), Flow-Cam (Sieracki et al., 1998), Zooscan (G. Gorsky et al., 2010), In Situ Ichthyoplankton Imaging System (ISIIS, Cowen and Guigand (2008)). Figure taken from Irisson et al. (2022) 9
- Figure 1.4 Steps taken by Morphocluster, a semi-autonomous software used to classify images. Images are projected to the feature space. Figure taken from Schröder et al. (2020) 11
- Figure 1.5 Graphical abstract summarising the main pathways of carbon in the water column taking into account 3 main actors: phytoplankton, zooplankton and bacteria. Solid arrows present a process resulting in particle formation or growth while dashed arrows correspond to processes resulting in degradation and/or fragmentation. Large circles represent the three chapters that compose this manuscript. 13

- Figure 2.1 Map of the UVP5 dataset used in this study. Transparency was used to illustrate the density of points on the map. 21
- Figure 2.2 Methodology followed from data selection to prediction of global biomass. 25
- Figure 2.3 Heatmap of the models'  $R^2$  between observed and predicted biomass for all zooplankton groups arranged from the most important in terms of biomass (Copepoda) to the least important (Limnacinidae) in the different depth layers. The regions correspond to: W for world (model run on all data) ; WL for world low (data between 40N and 40S from the world model) ; L for low latitude (model run between 40N and 40S) ; WH for world high (data outside of 40N and 40S from the world model) ; H for high latitude (model run outside of 40N and 40S). The stars indicate significant results (p-value < 0.05) obtained with the Pearson correlation test. 31
- Figure 2.4 Barplots showing the mean biomass predicted in PgC at 0-200 m and 200-500 m depth for each group ranked from highest to lowest biomass in 3 types of models : world, outside 40N-40S and inside 40N-40S. Error bars correspond to upper interval of the biomass estimation's standard deviation. The stars indicate a significant result (p-value < 0.05) obtained with the Pearson correlation test. 33
- Figure 2.5 Map of the mean biomass (color scale is log-transformed) of Copepoda as predicted by the model on 0-200 m (A) 200-500 m data (B) as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D). The color scale for the coefficient of variation has the same range for figures 5 to 9. 35
- Figure 2.6 Map of the mean biomass (color scale is log-transformed) of Eumalacostraca as predicted by the model on 0-200 m (A) 200-500 m data (B) as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D). 37

- Figure 2.7 Map of the mean biomass (color scale is log-transformed) of solitary Collodaria as predicted by the model on 0-200 m (A), as well as the coefficient of variation (B) as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D). 37
- Figure 2.8 Map of the mean biomass (color scale is log-transformed) of Phaeodaria as predicted by the model on 0-200 m (A), as well as the coefficient of variation for the 0-200 m model (B). In the map of predicted biomass, 12 cells in the California upwelling presented a value between 3 and 6  $mgC\ m^{-3}$  and were represented here in yellow to observe the distribution of this group on a global scale. The color scale for the coefficient of variation has the same range for all plots. 39
- Figure 2.9 Map of the mean biomass (color scale is log-transformed) of Acantharea as predicted by the model on 0-200 m (A), 200-500 m data (B), as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D). The color scale for the coefficient of variation has the same range for all plots. 39
- Figure 2.10 Comparison of the latitudinal distribution of biomass ( $mgC\ m^{-2}$ ) integrated over 0-500m depth between our models' estimation and the results from the Tara Ocean multinet (300  $\mu m$  mesh size), for Copepoda and Rhizaria. Trends were obtained by using Loess regression on: "BRT models" using the global model outputs for Copepoda or Rhizaria (summed across 0-200 m and 200-500 m depth) ; "UVP5" using the biomass as seen by the UVP5 between 0-500m ; "TARA Ocean net" using the sampling points between 0-500m. The shaded areas represent the 95% confidence interval of the Loess fit. 41
- Figure 2.11 Distribution map of the minimum global biomass between 0 and 500m using taxa which obtained a p-value  $< 0.05$  in Pearson test between the predicted and the biomass calculated from UVP5 data. 42

- Figure S2.1 Distribution of the conversion factors from biovolume ( $mm^3$  to biomass  $mgC$  in logarithmic scale) for the studied taxa according to their source in the literature. For Rhizaria, biovolume ( $mm^3$ ) to carbon ( $mgC$ ) conversions were done using factors from Biard et al., 2016, Mansour et al., 2021 and Marcolin et al., 2015. For other groups, the conversion from individual volume to individual wet weight assumed a density of  $1gcm^3$  (Kiørboe, 2013). Then the conversion from individual wet weight to individual biomass in carbon units ( $mgC$ ) was calculated using taxon-specific linear conversion factors from McConville et al. (2016) ; when several factors were available for a taxon, their median was used for each group). For solitary collodarians, the estimation of carbon ( $0.189 mgCmm^3$ ) by Mansour et al., 2021 applied as explained in the subsection 2.1.3 52
- Figure S2.2 Distribution maps of the environmental variables used in the model in the layer 0-200 m 53
- Figure S2.3 Distribution maps of the environmental variables used in the model in the layer 200-500 m 54
- Figure S2.4 Distribution maps of the environmental variables used in all models 55
- Figure S2.5 Distribution of UVP5 sampling points (symbolized by black dots) for which a match was found in the neighboring cell in the regions of the western coast of South America (A,B), in the Baffin Bay and Labrador Sea (C,D) and in the Mediterranean Sea (E,F). The colored cells represent the temperature (in  $^{\circ}C$ ) from the WOA dataset. In the epipelagic layer, 142 points have a neighboring match out of which 130 points (91%) are represented here (The map A contains 20 points, C contains 51 points and E contains 59 points). In the mesopelagic layer, 104 points have a neighboring match out of which 81 points (78%) are represented here (The map B contains 22 points, D contains 24 points and F contains 35 points). 56

- Figure S2.6 Distribution of sampling of environmental variables by UVP5 in red compared to global data from World Ocean Atlas in blue in the layer 0-200 m. 57
- Figure S2.7 Distribution of sampling of environmental variables by UVP5 in red compared to global data from World Ocean Atlas in blue in the layer 200-500 m. 57
- Figure S2.8 Distribution of the order of importance of variables in the model for Copepoda between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. 58
- Figure S2.9 Distribution of the order of importance of variables in the model for Eumalacostraca between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. For this group the model for the 200-500 m layer does not yield a significant correlation between model and data and results are therefore not shown. 59
- Figure S2.10 Distribution of the order of importance of variables in the model for solitary Collodaria between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. For this group the model for the 200-500 m layer does not yield a significant correlation between model and data and results are therefore not shown. 60
- Figure S2.11 Distribution of the order of importance of variables in the model for Phaeodaria between 0-200 m (A). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. 60

- Figure S2.12 Distribution of the order of importance of variables in the model for *Acantharea* between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. 61
- Figure 3.1 Main processes potentially affecting marine detrital particle morphology. Solid arrows present a process resulting in particle formation or growth while dashed arrows correspond to processes resulting in degradation and/or fragmentation. 67
- Figure 3.2 Sampling points inside ( $\leq 15$  km from the eddy center) and outside ( $> 15$  km) of the eddy core. The points represent the UVP5 vertical profiles done during the three epochs. The coloured background corresponds to the daily sea level anomaly (in m) for the middle day of each epoch: (A) May 7th for epoch 1, (B) May 16th for epoch 2 and (C) May 25th for epoch 3. The black star in each of the 3 figures corresponds to the centre of the eddy on the corresponding day. 72
- Figure 3.3 Interpolated field of A) aggregate concentration (in number per  $\text{m}^3$ ) and B) aggregate Equivalent Spherical Diameter (ESD in mm). Note that the colour bar is in logarithmic scale for the aggregate concentration. The 6 red lines correspond to the delimitation of the 5 masks with a  $50 \text{ m d}^{-1}$  slope. The continuous white line corresponds to the mean MLD. The black dashed line corresponds to the fit presented in Johnson et al. (2023) as the lower limit of the Surface Core Water based on the depth of the  $27.1 \text{ kg m}^{-3}$  isopycnal. 75
- Figure 3.4 Evolution of (A) MiP and (B) MaP abundance as well as the PSD slope. Each point corresponds to a profile's 20m mean value. Power law functions were fitted to the (A) MiP and (B) MaP abundances (in numbers  $\text{L}^{-1}$ ). For the PSD slope (C), the linear regression was performed on the SCW data. The black horizontal dashed line corresponds to the mean depth of the SCW inside the plume. 77

- Figure 3.5 Power law regressions fitted the mean values of flux in  $\text{mgC m}^{-2} \text{d}^{-1}$  for (A) MiPs, (B) MaPs and (C) all particles. Points in figures A through C represent the mean value per 20m bins. The jumps in the vertical profiles are due to averaging overtime in the succeeding depth bins. The carbon-specific degradation rate ( $D$ ,  $C_{\text{spec}}$  in  $\text{d}^{-1}$ ) is represented according to depth in figure D with circular points. The squared-shaped points represent data of carbon-specific remineralisation rate ( $C_{\text{remin}}$ ) from Belcher et al. (2016) and Collins et al. (2015) with the horizontal lines corresponding the standard deviation. The colours used to represent the masks in Figures A through D are presented in Figure D. 79
- Figure 3.6 Evolution of mean concentration (in numbers  $\text{m}^{-3}$ ) of aggregates (A) and fecal pellets (B). Each point corresponds to a 20 m mean value for each mask for which the colour legend is presented in Figure B. 80
- Figure 3.7 Aggregates images selected randomly inside the plume for masks 1, 3 and 5 between 0-100, 100-300, 300-500 and 500-800 m. ESD values correspond to the mean  $\pm$  standard deviation Equivalent Spherical Diameter. MGL values correspond to the mean  $\pm$  standard deviation Mean Grey Level. 81
- Figure 3.8 Comparison of aggregate and feces morphological variables. Each point corresponds to a 20 m mean value on all the profiles for either aggregate (in purple) or feces (in orange) among the five masks for figures A through G. Horizontal bars represent the 95% confidence interval. The ratio of aggregate over feces in the plume over depth is represented in Figure H. For each subplot, the black horizontal line corresponds to the mean depth of the Surface Core Waters inside the plume. 82

- Figure 3.9 Particle Size Distribution of imaged aggregates and feces for data between 0-200, 200-400, 400-600 and 600-800 m. The two axes are log-transformed. The colour of the points corresponds to the mean grey level of the corresponding combination of particle type, depth layer and size bin. Values for feces are linked with dashed lines and the ones for aggregates are linked with dotted lines. 84
- Figure S3.1 Sampling points inside ( $< 15$  km from the eddy center) and outside ( $> 15$  km) of the eddy core. The points represent the UVP5 vertical profiles done during each day of the EXPORTS 2021 campaign from May 4<sup>th</sup> to May 29<sup>th</sup> 2021. 95
- Figure S3.2 Evolution of (A) MiP and (B) MaP biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ). Each point corresponds to a profile's 20 m mean value. Power law functions were fitted to the (A) MiP biovolumes. The black horizontal dashed line corresponds to the mean depth of the SCW inside the plume. Two very high MaP biovolume values (417,054 and 476,708  $\text{mm}^3 \text{m}^{-3}$  obtained respectively at 5 on the 13<sup>th</sup> and at 345 m on the 21<sup>st</sup> were removed from this figure to make it more readable. 96
- Figure S3.3 Evolution of (A) MiP and (B) MaP abundance (in numbers  $\text{m}^{-3}$ ) per 100 m depth bins. The depth bins values in the legend corresponds to the centre of the depth bin. Each point corresponds to a mean daily value for the corresponding depth bin. 97
- Figure S3.4 Power law regressions fitted the mean values of flux (in  $\text{mgC m}^{-2} \text{d}^{-1}$ ) computed in a vertical framework for every day of the dataset from the 4<sup>th</sup> to the 29<sup>th</sup> May 2021. Points represent the mean value per 20 m bins. 98
- Figure S3.5 Evolution of aggregates morphological variables. Each point corresponds to a 20 m mean value for one profile. Linear regressions were fitted between 0 and SCW bottom limit and between this bottom limit and 800 m. They drawn in this figure if they had a significant slope and intercept. The corresponding slope and intercept values can be found in the Table S3.2 and S3.3 99



- Figure S3.6 Evolution of feces morphological variables. Each point corresponds to a 20 m mean value for one profile. 100
- Figure S3.7 Feces images selected randomly inside the plume for masks 1, 3 and 5 between 0-100, 100-300, 300-500 and 500-800 m. 101
- Figure S3.8 Concentration (in numbers  $\text{m}^{-3}$ ) of two zooplankton groups: Crustacea in orange and Rhizaria in purple. Horizontal bars represent the 95% confidence interval. 102
- Figure 4.1 Evolution of concentration (in number  $\text{m}^{-3}$ ) and biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ) of (A, C) aggregates and (B, D) feces for 0-100 m, 100-300 m, 300-500 m, 500-800 m. The mean concentration is represented by points and linked with a dashed line. The shaded areas represent the 95% confidence interval for the dashed line. The grey-coloured vertical bars correspond to the storms. 116
- Figure 4.2 Evolution of concentration (in number  $\text{m}^{-3}$ ) and biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ) of (A, C) aggregates and (B, D) feces for 0-100 m, 100-300 m, 300-500 m, 500-800 m. The mean concentration is represented by points and linked with a dashed line. The shaded areas represent the 95% confidence interval for the dashed line. 117
- Figure 4.3 Estimation of sinking speed (in  $\text{m d}^{-1}$ ) according to the Equivalent Spherical Diameter (ESD in mm). The circular dots represent the estimations of mean sinking speed done on UVP5 images for aggregates (in blue) and feces (in orange) with the 95% confident interval represented by the vertical error bars and on UVP5 counts (in black). The dashed blue line corresponds to the linear regression done on the imaged aggregates sinking speed. Continuous lines correspond to relationships found in the literature in Alldredge and Gotschalk (1988), Cael et al. (2021), Iversen et al. (2010), and Kriest (2002). 118
- Figure 4.4 Comparison sediment trap and UVP flux estimations (see next page). 120

- Figure 4.4 (on the previous page) Comparison between sediment trap and UVP flux estimations done using UVP5 counts between 10 $\mu$ m and 2 mm and published A and b couples: A) (Kiko et al. (2017) following Kriest (2002)), B) Clements et al. (2023), C) Guidi et al. (2008), D) Forest et al. (2013) and E) Fender et al. (2019). Figure F is a comparison between sediment trap and UVP flux estimations done using UVP5 counts between 0.128 and 16.4mm using the A and b couple from Kiko et al. (2017) following Kriest (2002). Figures G and H were obtained by adding flux computed on UVP5 counts between 100  $\mu$ m-1 mm to flux computed on UVP5 images of aggregates and feces. The flux computed on the images was done using Kriest (2002) Table 1 2a relationship for sinking speed and (G) Durkin et al. (2021) or (H) Alldredge (1998) for the carbon content. The dashed black line corresponds to a 1:1 ratio. The values presented in each subplot correspond to the coefficient of determination ( $R^2$ ), the root mean square error (RMSE), the mean bias error (MBE) and the slope. The slope and  $R^2$  were computed by doing a linear regression between the sediment trap and UVP fluxes. All slopes are statistically significant (p-value < 0.05), as denoted by an asterisk. 121
- Figure 4.5 Flux (in mgC m<sup>-2</sup> d<sup>-1</sup>) per size classes (in mm) computed on UVP particle counts for each matchup with sediment traps. Colours represent the method used to compute the flux. Values on top of each subplot represent the mean percentage  $\pm$  standard deviation of flux from particles between 10-100  $\mu$ m, 100  $\mu$ m-2 mm and 2-16.4 mm. 122
- Figure S4.1 Concentration of UVP counts normalised by the width of the ESD size bin [ m<sup>-3</sup> mm<sup>-1</sup>]. 129

- Figure S4.1 (on the previous page) The red dashed line corresponds to a size of 100  $\mu\text{m}$ . Concentrations were obtained from the UVP<sub>5</sub> particle counts per  $\text{m}^3$  between 0.102-16.4 mm (on the right of the red dashed line) and extrapolated to size classes between 0.010-0.102 mm. Both axes are on a logarithmic scale. Each subplot corresponds to the matchup between one sediment trap and the corresponding UVP profile as described in the Material and Method section. The vertical black lines correspond to the standard error of the standardised concentration. 130
- Figure 6.1 Voies du cycle et de l'exportation du carbone par le zooplancton dans l'océan. Les organismes phytoplanctoniques absorbent le  $\text{CO}_2$  par photosynthèse dans la couche euphotique. Ils sont broutés par le micro et le mésozooplancton qui participent à la pompe à carbone biologique en produisant des pelotes fécales qui coulent de la surface jusqu'en profondeur. Figure extraite de Steinberg and Landry (2017) 151
- Figure 6.2 Quatre stratégies de traitement de collections d'images brutes de plancton identifiées par quatre numéros encerclés. L'interaction efficace entre les instruments, les ordinateurs et les opérateurs humains permet un débit élevé depuis l'acquisition d'images en laboratoire ou sur le terrain jusqu'à l'exploitation écologique. Figure tirée de Irisson et al. (2022) 153
- Figure 6.3 Évolution du nombre d'images téléchargées sur le site Ecotaxa (Picheral et al., 2017) à l'aide d'instruments tels que l'Underwater Vision Profiler (UVP (Picheral et al., 2010, 2021)), FlowCytobot (IFCB, (Olson and Sosik, 2007)), FlowCam (Sieracki et al., 1998), Zooscan (G. Gorsky et al., 2010), In Situ Ichthyoplankton Imaging System (ISIIS, (Cowen and Guigand, 2008)). Figure tirée de Irisson et al. (2022) 154

- Figure 6.4 Résumé graphique résumant les principales voies du carbone dans la colonne d'eau en tenant compte de trois acteurs principaux : le phytoplancton, le zooplancton et les bactéries. Les flèches pleines présentent un processus aboutissant à la formation ou à la croissance de particules tandis que les flèches en pointillés correspondent à des processus aboutissant à la dégradation et/ou à la fragmentation. Les grands cercles représentent les trois chapitres qui composent ce manuscrit. 155
- Figure 6.5 Étapes suivies par Morphocluster, un logiciel semi-autonome utilisé pour classer les images. Figure extraite de Schröder et al. (2020) 158
- Figure 6.6 Carte de distribution de la biomasse globale minimale entre 0 et 500 mètres en utilisant les taxons qui ont obtenu une p-value  $< 0,05$  dans le test de Pearson entre la biomasse prédite et la biomasse calculée à partir des données UVP5. 160
- Figure 6.7 Champ interpolé de A) concentration d'agrégats ([nombre par  $m^3$ ]) et B) diamètre sphérique équivalent des agrégats (ESD [mm]). Veuillez noter que la barre de couleur est à l'échelle logarithmique pour la concentration d'agrégats. Les 6 lignes rouges correspondent à la délimitation des 5 masques avec une pente de  $50 \text{ m d}^{-1}$ . La ligne blanche continue correspond à la profondeur moyenne de la couche de mélange. La ligne noire en pointillés correspond à l'ajustement présenté dans Johnson et al. (2023) en tant que limite inférieure de la couche d'eau de base de surface définie en fonction de la profondeur de l'isopycne  $27,1 \text{ kg m}^{-3}$ . 163

# List of Tables

---

Table 2.1	Biovolume methods	23
Table 2.2	Comparison of global biomass estimates in the literature.	47
Table S2.1	List of groups used for the analysis	62
Table 3.1	Flux estimates (in $\text{mgC m}^{-2} \text{d}^{-1}$ ) from various studies done on the North Atlantic spring bloom. Buesseler et al. (1992) and Alkire et al. (2012) sampled at the JGOFS (Joint Global Ocean Flux Study) North Atlantic Bloom Experiment (NABE) at approximately $47^{\circ}\text{N } 20^{\circ}\text{W}$ . Bagniewski et al. (2011), Briggs et al. (2011), Omand et al. (2015) and Martin et al. (2011) sampled in the south of Iceland.	89
Table S3.1	Evolution of mean values of particles community parameters between 0 and 800 m inside the plume. The fluxes are in $\text{mgC m}^{-2} \text{d}^{-1}$ according to Kriest, 2002 and Guidi et al., 2008. The column Depth corresponds to the center of the 50 m depth bin.	102
Table S3.2	Linear regressions from Figure S3.5 for the SCW	103
Table S3.3	Linear regressions from Figure S3.5 between the lower limit of the SCW and 800 m	104
Table S3.4	Evolution of mean values of particles morphological parameters between 0 and 800 m inside the plume. The column depth corresponds to the center of the 50 m depth bin.	105
Table 4.1	A and b values to compute flux from previous studies	114
Table 4.2	Carbon content estimates from previous studies (Alldredge, 1998; Durkin et al., 2021). Alldredge (1998) uses biovolume in $\text{mm}^3$ while Durkin et al. (2021) uses biovolume in $\mu\text{m}^3$ . The sphere biovolume is $V_{\text{sphere}} = \frac{4}{3} \times \pi \times \left(\frac{\text{ESD}}{2}\right)^3$ and $V_{\text{cylinder}} = 1 \times \pi \times \left(\frac{w}{2}\right)^2$ with $w = \frac{553 \times \text{ESD}}{\text{ESD} + 996}$ and $l = \frac{\pi \left(\frac{\text{ESD}}{2}\right)^2}{w}$ .	115



# List of Acronyms

---

<b>BCP</b>	Biological Carbon Pump
<b>BGP</b>	Biological Gravitational Pump
<b>BRT</b>	Boosted Regression Trees
<b>CTD</b>	Conductivity, Temperature, Depth
<b>CV</b>	Coefficient of Variation
<b>ECW</b>	Eddy Core Waters
<b>ESD</b>	Equivalent Spherical Diameter
<b>EXPORTS</b>	EXport Processes in the Ocean from Remote Sensing
<b>GAM</b>	Generalised Additive Models
<b>GLM</b>	Generalised Linear Models
<b>MaP</b>	Large particles, 0.53-16.99 mm diameter
<b>MBE</b>	Mean Bias Error
<b>MiP</b>	Small particles, 0.14-0.53 mm diameter
<b>MLD</b>	Mixed Layer Depth
<b>MSC</b>	Marine Snow Catcher
<b>NBST</b>	Neutrally Buoyant Sediment Traps
<b>NOAA</b>	National Oceanic and Atmospheric Administration
<b>POC</b>	Particulate Organic Carbon
<b>PSD</b>	Particle Size Distribution
<b>RF</b>	Random Forest
<b>RMSE</b>	Root Mean Square Error
<b>SCW</b>	Surface Core Waters
<b>STT</b>	surface-tethered traps
<b>UVP</b>	Underwater Vision Profiler
<b>WOA</b>	World Ocean Atlas





## Part I

### General introduction



# General introduction

---

## 1.1 The role of the Biological Gravitational Pump in the oceanic carbon cycle

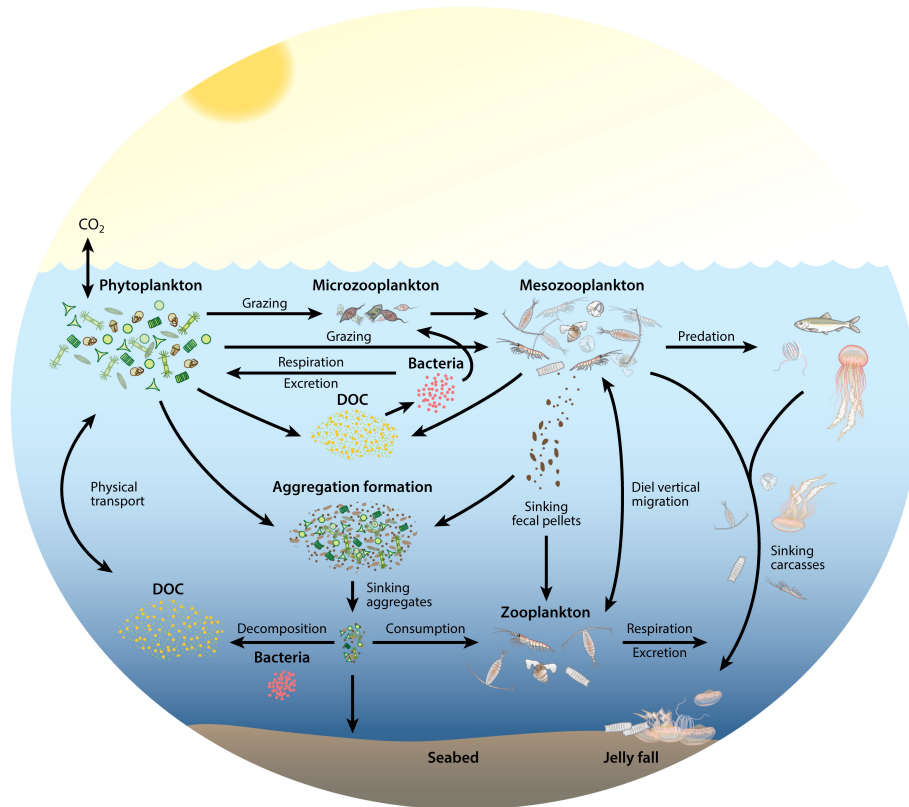
The biological carbon pump (BCP) designates the suite of biological processes that produce and mediate the transport of carbon from the upper ocean to depth (Fig. 1.1, (Honjo et al., 2008; Steinberg and Landry, 2017; Turner, 2015; Volk and Hoffert, 1985)). It is a fundamental mechanism responsible for the transport of particulate organic carbon (POC) from the ocean surface to its deep layers where the carbon can be sequestered for periods of decades to millennia (DeVries et al., 2012). The BCP is estimated to have a magnitude of 5 to 12 Pg C year<sup>-1</sup> (Boyd et al., 2019; DeVries and Weber, 2017; Henson et al., 2011; Nowicki et al., 2022; Siegel et al., 2014) and is composed of three main pathways: 1) physically driven subduction pumps (Boyd et al., 2019; K. O. Buesseler et al., 2020; Omand et al., 2015); 2) biologically driven pumps that depend on the diel vertical migration of zooplanktonic organisms to depth during the day and to the surface during the night (Bianchi et al., 2013; Kiko et al., 2020; Lampert, 1989) or seasonal and ontogenetic vertical migrations (Jónasdóttir et al., 2015) and 3) the biological gravitational pump (BGP) which refers to the passive sinking of POC as marine particles through the water column (Alldredge and Gotschalk, 1988; Boyd et al., 2019). The BGP dominates the BCP and is recognised as the major contributor to global net carbon export (Boyd et al., 2019), accounting for 56.1-70% (Nowicki et al., 2022; Stukel et al., 2022). While there are global estimations of carbon flux, the large uncertainty illustrates the lack of knowledge about particle dynamics.

## 1.2 The diverse nature of marine particles and their role in oceanic systems

Marine detrital particles are prevalent in the Earth's oceans and consist of aggregates (macroscopic clusters with a diameter greater than 500 µm, Alldredge and Silver (1988)), dead bodies and fecal pellets. Marine aggregates can result from the agglomeration of dead, but also alive phytoplankton, discarded larvacean houses, fecal matter, and other detritus (Alldredge et al., 1990; Alldredge and Silver, 1988). Several factors influence their sinking dynamics, including density (Bach et

al., 2019; Cael et al., 2021), size (Alldredge and Silver, 1988; Guidi et al., 2008; Iversen et al., 2010), composition (Laurenceau-Cornec et al., 2015) and morphology (Alldredge and Gotschalk, 1988; Trudnowska et al., 2021). Given the highly variable nature of marine particles, their contribution to the BGP exhibits significant spatio-temporal variability. Their distribution has been primarily driven by the estimate of their POC content and size distribution rather than by their individual characteristics (Clements et al., 2022, 2023; Guidi et al., 2008, 2015; Kiko et al., 2017). The majority of these particles are concentrated near the surface, where primary production takes place, and exhibit a rapid decrease in particle concentration as depth increases (Guidi et al., 2008, 2015; Iversen, 2023; Kiko et al., 2017; Stemmann et al., 2002; L. Stemmann et al., 2008).

Marine aggregates display a diverse range of morphologies, shaped by various physical and biological forces that govern their formation and modification. Physical processes can promote the aggregation (Alldredge, 2001; Kjørboe, 2001; Stemmann, Jackson, and Ianson, 2004) or disaggregation (Alldredge and Gotschalk, 1988; Song and Rau, 2022) of particles, thus altering their morphology. Similarly, various biological interactions may affect particle morphology, depending on the involved actors, such as bacteria, micro- and macrozooplankton. Microorganisms can enhance aggregation processes through swimming activities (Alldredge, 2001). However, their primary impact on particles is through their degradation, especially in the upper water column (Iversen, 2023; Turner, 2015). They modify particle structure (Stemmann, Jackson, and Ianson, 2004) to varying degrees depending on environmental factors (Amano et al., 2022; Marsay et al., 2015; Weber and Bianchi, 2020). In this context, it is crucial to clarify the role of the biological players (zooplankton and microorganisms) and to include them in the models in order to estimate the fluxes.




 Steinberg DK, Landry MR. 2017. *Annu. Rev. Mar. Sci.* 9:413–44

Figure 1.1: Pathways of cycling and export of carbon by zooplankton in the ocean. Phytoplanktonic organisms take up  $\text{CO}_2$  via photosynthesis in the euphotic layer. They are grazed upon by micro and mesozooplankton. The latter participates in the biological carbon pump through the production of sinking fecal pellets at the surface and at depth. Figure taken from Steinberg and Landry (2017)

### 1.3 Zooplankton: Key players in marine ecosystems and the Biological Carbon Pump

First defined as organisms adrift in the water by Hensen in 1887 (Smetacek, 1999), planktonic organisms are classified into two primary groups: phytoplanktonic primary producers that take up CO<sub>2</sub> via photosynthesis in the euphotic layer (see 1.1); and zooplankton that graze on them, serving as a major link between primary producers and higher trophic levels (Ikeda, 1985; Steinberg and Landry, 2017). Due to their position in marine food webs, zooplankton organisms are a crucial food source for various higher predators, including marine mammals, sea birds and fish (Chavez et al., 2008; Frederiksen et al., 2006; van der Lingen et al., 2009; Ware and Thomson, 2005). Zooplanktonic organisms exhibit an important diversity (de Vargas et al., 2015) and cover a large size range from micrometre to several meters (Lombard et al., 2019). Found in all of the world's oceans, the distribution of zooplankton is intricately linked with the environmental parameters that define the water masses they inhabit (Hays et al., 2005; Steinberg and Landry, 2017). Just as these parameters change greatly according to time and latitude, zooplankton biomass and diversity exhibit temporal and latitudinal variations. While biomass is high at higher latitudes and low at lower latitudes, diversity follows the opposite pattern (Ibarbalz et al., 2019; Ikeda, 1985; Moriarty et al., 2012; Rombouts et al., 2009).

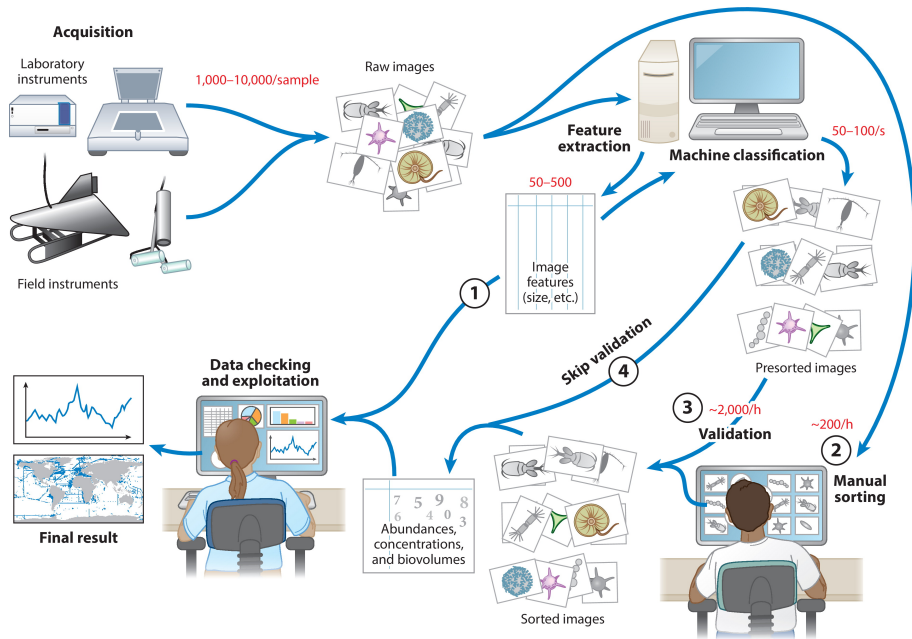
Defined as gatekeepers of POC flux (Jackson and Checkley, 2011), zooplankton organisms are important actors of the BCP (Steinberg and Landry, 2017; Turner, 2002, 2015). They have the important ability to repackage small particles into large, rapidly sinking fecal pellets (Atkinson et al., 2012; Turner, 2002, 2015). In addition to this, zooplankton can fragment sinking particles through their feeding activities (both active and passive, Alldredge et al. (1990), Huntley and Boyd (1984), and Steinberg et al. (2023)) and swimming capabilities (Dilling and Alldredge, 2000; Goldthwait et al., 2004). Understanding the significance of zooplankton in the BCP requires a thorough comprehension of their dynamics at a vertical scale. Upscaling to basin and global levels would aid in precisely quantifying their impact.

### 1.4 Shedding light on zooplankton and particles through *in situ* imaging

While zooplankton has traditionally been collected using nets or Niskin bottles, sediment traps have historically been used to collect particles. These two types of instruments have provided significant data on the distribution and biomass of zooplankton and particles (Moriarty and

O'Brien, 2013; Moriarty et al., 2012; Mouw et al., 2016), although they have some limitations. Indeed, they cannot determine the exact environmental context of the sampled specimens and their spatio-temporal definition is limited. Therefore, the data on biomass and concentration that is sampled is integrated over a large volume of water and over a long period. Zooplankton samples also require manual sorting, which is highly time-consuming. It should also be emphasised that homogenisation of data is difficult when using manual counts and net sampling. As a result, the production of comprehensive global distributions of zooplankton is hampered by the difficulty of collecting and standardising such diverse data sources. For example, comprehensive databases such as COPEPOD (O'Brien, 2005) exhibit notable biases towards regions with high sampling frequency despite covering 50 years of data (Moriarty and O'Brien, 2013). Moreover, nets have been shown to be skewed towards crustaceans that are more resistant than fragile organisms, namely rhizarians and gelatinous forms that end up being destroyed by the sampling (Biard et al., 2016; Lucas et al., 2014). The above statement holds for numerous particles, including larvacean houses and large aggregates that have a significant role in the BGP and possess a complicated and delicate structure that is vulnerable to destruction by conventional techniques (Alldredge and Silver, 1988; Remsen et al., 2004).

To address these limitations and gain a more comprehensive understanding, *in situ* methods to count (Herman, 1992; Reynolds et al., 2010), image (Cowen and Guigand, 2008; Picheral et al., 2010; Remsen et al., 2004) and record videos (Davis et al., 2005; Hoving et al., 2019) of zooplankton and particles are valuable tools for informing on their dynamics (Lombard et al., 2019). This is especially true for instruments which are widely commercialised and inter-calibrated, as is the case for the Underwater Vision Profiler 5 (Picheral et al., 2010). Together, these instruments allow the coverage of an important part of the planktonic size spectrum (Lombard et al., 2019).



Irison J-O, et al. 2022  
*Annu. Rev. Mar. Sci.* 14:277–301

Figure 1.2: Four strategies to process raw plankton images identified by four circled numbers. Efficient interaction between instruments, computers, and human operators enables high throughput from laboratory- or field-based image acquisition to ecological exploitation. Figure taken from Irison et al. (2022)



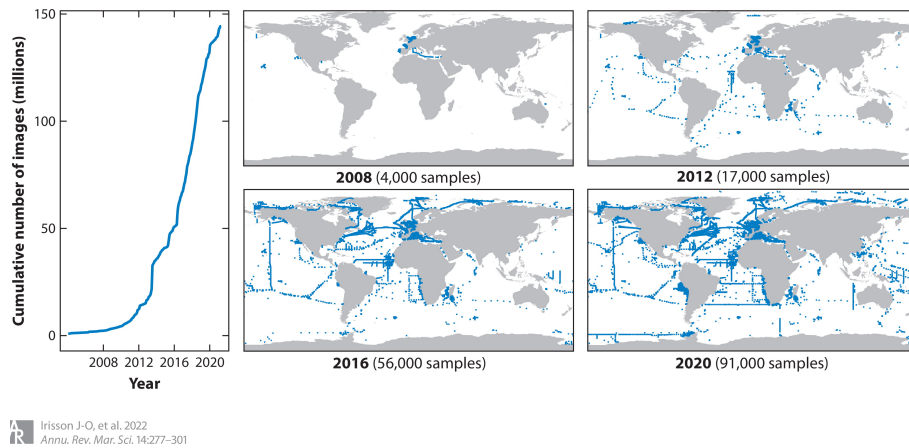


Figure 1.3: Evolution of images uploaded on the Ecotaxa repository (Picheral et al., 2017) using instruments such as the Underwater Vision Profiler (UVP, Picheral et al. (2010, 2021), Imaging FlowCytobot (IFCB, Olson and Sosik (2007)), FlowCam (Sieracki et al., 1998), Zooscan (G. Gorsky et al., 2010), In Situ Ichthyoplankton Imaging System (ISIIS, Cowen and Guigand (2008)). Figure taken from Irisson et al. (2022)

Notably, *in situ* imaging offers the advantage of high-frequency data acquisition and has emerged as a powerful tool for informing on zooplankton and particle distribution on both global and regional scales (Lombard et al., 2019). However, this advantage comes with a challenge: the high-frequency acquisition of data equates to a substantial volume of information to manage. To effectively address this issue, as demonstrated in Fig. 1.2 (Irison et al., 2022), innovative strategies and tools are required to streamline the handling and analysis of these image-rich datasets. Images and videos can provide information on the abundance of zooplankton and particles as well as measurements that can all be archived for further analysis (G. Gorsky et al., 2010; Lombard et al., 2019) as presented by strategy 1 (Fig. 1.2). While size-based parameters have been the primary focus of numerous studies (Cael et al., 2021; Guidi et al., 2008; Romagnan et al., 2016; Stemmann et al., 2002), few have looked at changes in other morphological traits (Giering et al., 2020; Trudnowska et al., 2021). Images can also undergo manual validation by experts (strategy 2, Fig. 1.2). However, this process can potentially introduce several cognitive biases, regardless of the level of taxonomic knowledge, as highlighted in Culverhouse (2007) and Culverhouse et al. (2014). The manual classification of such large amounts of data would require a significant workforce of highly taxonomically trained individuals, especially given the exponential rate of image acquisition that has occurred in recent years (Fig. 1.3). Therefore, the development and implementation of semi-automated (strategy 3,

Fig. 1.2) or fully automated (strategy 4, Fig. 1.2) tools for classification are now imperative in the field of *in situ* imaging.

## 1.5 Machine learning in marine ecology

Machine learning is a powerful computational approach that allows systems to learn from a training data set and make predictions or decisions on a test set without being explicitly programmed (Irisson et al., 2022). It encompasses different techniques that serve distinct purposes, such as classification and regression that have found utility in diverse applications in marine ecology (Rubbens et al., 2023). Classification entails assigning data points such as images into predetermined categories or classes, having learned from a learning set of manually annotated images (Luo et al., 2018). Regression focuses on predicting continuous numerical values (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000). Furthermore, machine learning can be either autonomous (Luo et al., 2018; Trudnowska et al., 2021) or semi-automatic (Biard et al., 2016; Panaiotis et al., 2023; Schröder et al., 2020; Vilgrain et al., 2021).

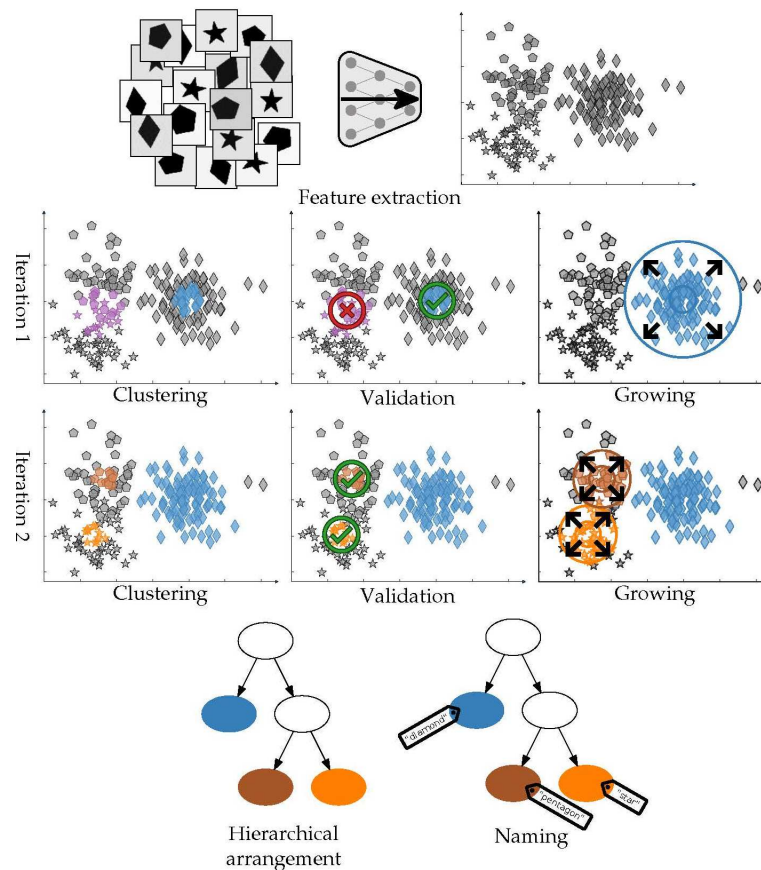


Figure 1.4: Steps taken by Morphocluster, a semi-autonomous software used to classify images. Images are projected to the feature space. Figure taken from Schröder et al. (2020)

Autonomous systems necessitate minimal human intervention, while semi-automatic systems incorporate human-guided processes for machine learning. As classification and regression predictions are based on a learning set, it is important to consider the potential for bias within this dataset and take measures to limit its impact (Orenstein et al., 2020).

One example of semi-autonomous classification is Morphocluster (Schröder et al., 2020), a program designed to annotate large data sets of images (Fig. 1.4). It extracts features from images which are subsequently projected in the features space. The clustering of images is then based on their position in this feature space. If a cluster is deemed good by a human validator, it is then grown under human supervision, and several iterations of the clustering and growing process are used to assign images to clusters. The clusters are then manually arranged hierarchically and named.

The combination of data exploration techniques and machine learning has the potential to yield valuable insights into marine ecosystems through the detection of hidden patterns and relationships within complex ecological datasets.

## 1.6 Objectives of this work

In order to accomplish an "image-based global analysis of the biological carbon pump", our study harnesses the capabilities of *in situ* imaging through UVP5 data. In doing so, we embark on a multifaceted investigation of the two primary components of the BCP for which the UVP5 instrument was purposefully designed to observe: zooplankton and particles (Fig. 1.5). In this framework, this manuscript is organised around two main topics: the global distribution of zooplankton biomass, as well as the detrital carbon export in the context of a North Atlantic bloom. It aims (1) to quantify the global zooplankton biomass; (2) to improve our understanding of the biological carbon pump through the study of particle morphology and its change with depth in relation to zooplankton distribution; (3) to compute and compare the flux of the North Atlantic spring bloom utilizing both sediment traps and high-resolution UVP5 data.

### 1.6.1 Global distribution of zooplankton biomass estimated by *in situ* imaging and machine learning

The first chapter focuses on the use of a global UVP5 dataset to examine the distribution of zooplankton biomass and addresses the following key research questions: (1) Can we accurately estimate the geographic distribution of large plankton groups across multiple depth layers using *in situ* imaging? (2) Is it feasible to derive regional estimates of zooplankton biomass using this methodology? (3) Do the obtained results align with previously established patterns?

To investigate these questions, we harnessed the power of prediction in Ecotaxa (Picheral et al., 2017) in collaboration with various experts who dedicated years to meticulous classification of this impressive dataset containing 466,872 images of zooplankton. We hypothesised that the distribution of organisms and their biomass depends on environmental factors. We utilised *in situ* imagery data from 3,549 UVP5 profiles worldwide and developed a novel technique for globally estimating macrozooplankton biomass based on these images and habitat models.

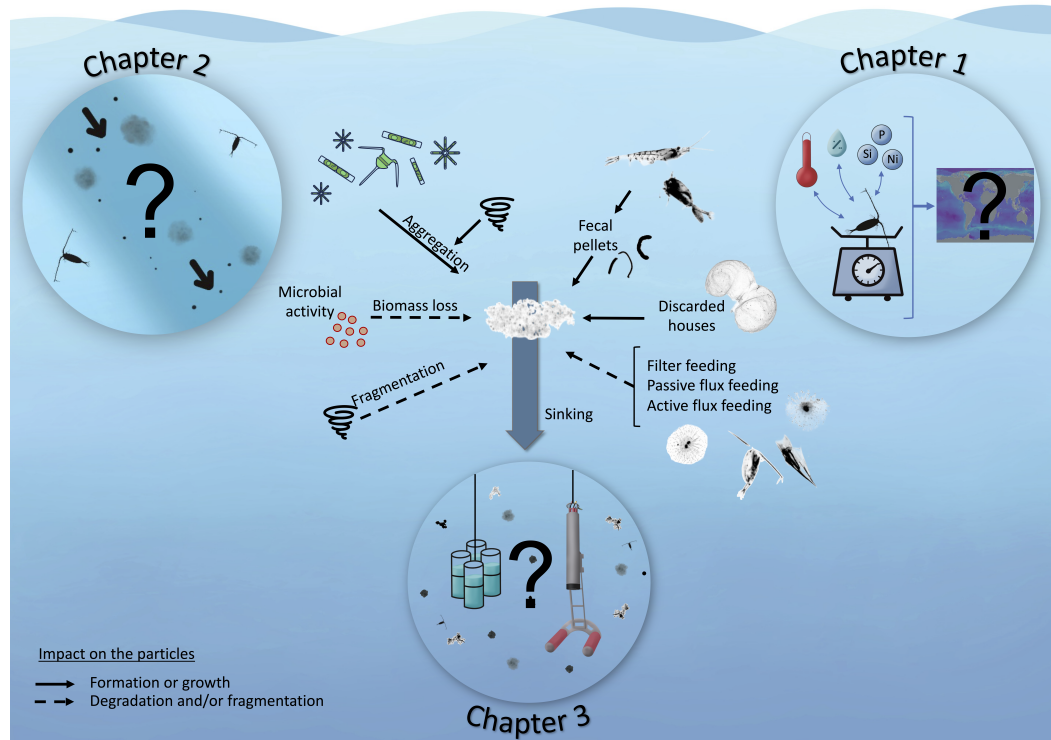


Figure 1.5: Graphical abstract summarising the main pathways of carbon in the water column taking into account 3 main actors: phytoplankton, zooplankton and bacteria. Solid arrows present a process resulting in particle formation or growth while dashed arrows correspond to processes resulting in degradation and/or fragmentation. Large circles represent the three chapters that compose this manuscript.

### 1.6.2 Export in the context of a North Atlantic bloom

The following two chapters focus on the EXPORTS 2021 campaign that took place in a North Atlantic eddy during the spring bloom (Johnson et al., 2023; Siegel et al., 2021). During this campaign three UVP5 units were deployed from three scientific vessels yielding 1,720,914 images. As we transition into the second and third chapters of this study, it becomes evident that the vast volume of images collected during the EXPORTS experiment necessitated a shift in approach. The sheer magnitude of data acquisition precluded the utilization of the Ecotaxa framework, prompting us to explore alternative methods. Moreover, we recognized the need for a more thorough examination of detritus, driving our pursuit of a deeper understanding in these subsequent chapters and the classification of the 1,720,914 images was done using Morphocluster, the semi-supervised method by Schröder et al. (2020) presented above.

In this context, Chapter 2 explores the export of imaged marine detritus using a new framework that employs an oblique plume following method. The chapter seeks to provide new perspectives on particle dynamics, morphology, flux, and attenuation within a North Atlantic eddy by answering the following questions: (1) How does the particle community evolve within this large export event? (2) Are zooplankton important relative to marine aggregates? (3) What is the attenuation rate of the exported material and how does it relate to biological activity? (4) How do aggregates and fecal pellets change in morphology with depth and time, and what processes are driving these changes? Due to the scarcity of research on particle morphology *in situ* with high vertical resolution across a wide depth range, our objective was to offer a more precise characterization of particle dynamics, focusing on aggregates and fecal pellets.

Chapter 3 uses a more traditional vertical approach to compute and compare the flux of the North Atlantic spring bloom using sediment traps and UVP5 data. It focuses on particle counts derived from UVP5 data as well as images of aggregates and fecal pellets (1 - 16.4 mm) and aims to answer the following questions: (1) What are the dominant particle types and their sinking speeds within the eddy core during the North Atlantic spring bloom? (2) How can the flux be effectively computed and compared using sediment traps and UVP5 data?

Finally, we conclude and discuss the main findings from each of these three chapters.

## Part II

### Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning

Laetitia Drago<sup>1</sup>, Thelma Panaïotis<sup>1</sup>, Jean-Olivier Irisson<sup>1</sup>, Marcel Babin<sup>2</sup>, Tristan Biard<sup>3</sup>, François Carlotti<sup>4, 5</sup>, Laurent Coppola<sup>1, 6</sup>, Lionel Guidi<sup>1</sup>, Helena Hauss<sup>7</sup>, Lee Karp-Boss<sup>8</sup>, Fabien Lombard<sup>1, 9</sup>, Andrew M. P McDonnell<sup>10</sup>, Marc Picheral<sup>1</sup>, Andreas Rogge<sup>11</sup>, Anya M. Waite<sup>12</sup>, Lars Stemann<sup>1†</sup> and Rainer Kiko<sup>1†</sup>

<sup>1\*</sup> Sorbonne Université, Laboratoire d'Océanographie de Villefranche-sur-mer, Villefranche-sur-mer, France,

<sup>2</sup> Takuvik International Research Laboratory, Québec Océan, Laval University (Canada) - Centre National de la Recherche Scientifique (CNRS), Département de biologie and Québec-Océan, Université Laval, QC, Canada,

<sup>3</sup> Laboratoire d'Océanologie et de Géosciences (LOG), Univ. Littoral Côte d'Opale, Univ. Lille, Centre National de la Recherche Scientifique (CNRS), UMR 8187, Wimereux, France,

<sup>4</sup> Département Ecologie Marine et Biodiversité (EMBIO), M.I.O. Institut Méditerranéen d'Océanologie Bâtiment Méditerranée, Marseille, France,

<sup>5</sup> Laboratoire d'Océanographie Physique et Biologique (LOPB), case 901 13288, Marseille, France,

<sup>6</sup> Sorbonne Université, Centre National de la Recherche Scientifique (CNRS), OSU STAMAR, Paris, France,

<sup>7</sup> Department Ocean Ecosystems Biology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany,

<sup>8</sup> School of Marine Sciences, University of Maine, Orono, ME, United States,

<sup>9</sup> Institut Universitaire de France (IUF), Paris, France,

<sup>10</sup> Oceanography Department, University of Alaska Fairbanks, Fairbanks, AK, United States,

<sup>11</sup> Section Benthopelagic Processes, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany,

<sup>12</sup> Ocean Frontier Institute and Oceanography Department, Dalhousie University, Halifax, NS, Canada

† These authors share last authorship

**Published:** 09 August 2022 in *Frontiers of Marine Sciences* <https://www.frontiersin.org/articles/10.3389/fmars.2022.894372/full>





# Global Zooplankton Biomass

---

## 2.1 Abstract

Zooplankton plays a major role in ocean food webs and biogeochemical cycles and provides major ecosystem services as a main driver of the biological carbon pump and a pivotal actor in sustaining fish communities. Zooplankton is also sensitive to its environment and reacts to its changes. To better understand the importance of zooplankton and to inform prognostic models that try to represent them, spatially-resolved biomass estimates of key plankton taxa are desirable. In this study, we for the first time predict the global biomass distribution of 19 zooplankton taxa (1-50 mm Equivalent Spherical Diameter) using observations by the Underwater Vision Profiler 5, a quantitative *in situ* imaging instrument. After computer-assisted classification of 466,872 organisms from more than 3,478 profiles (0-500 m) obtained between 2008 and 2019 throughout the globe, we estimated their individual bio-volume and converted it to biomass using taxa-specific factors. We then associated these biomass estimates with climatologies of environmental variables (temperature, salinity, oxygen, etc.), to build habitat models using boosted regression trees. The results reveal maximal zooplankton biomass values around 60°N and 55°S as well as minimal values around the oceanic gyres. An increased zooplankton biomass is predicted centered on the equator. Global integrated biomass (0-500 m) was estimated at 0.403 PgC. It was largely dominated by Copepoda (35.7%, mostly in polar regions), followed by Eumalacostraca (26.6%) Rhizaria (16.4%, mostly in inter tropical areas). The used machine learning approach is sensitive to the amount of training data and generates reliable predictions for abundant groups such as Copepoda ( $R^2 \approx 20-66\%$ ) but not for rare ones (Ctenophora, Cnidaria,  $R^2 < 5\%$ ). Still, this study offers a first protocol to estimate global, spatially resolved zooplankton biomass and community composition from *in situ* imaging observations of individual organisms. The underlying dataset was obtained within ten years, whereas similar approaches rely on data obtained using plankton nets gathered since about 1960. Increased use of digital imaging approaches should enable us to obtain zooplankton biomass distribution estimates at basin to global scales in shorter time frames in the future.

## 2.2 Introduction

### 2.2.1 Zooplankton

Present in all the oceans of the globe, zooplankton corresponds to organisms adrift in the water. They represent a great taxonomic diversity and sizes ranging from a few micrometers to several meters (de Vargas et al., 2015; Karsenti et al., 2011; Stemmann and Boss, 2012). Zooplankton play a central role in the biogeochemical carbon cycle as they contribute to the biological pump that drives the export of photosynthetically fixed organic carbon from the surface to the intermediate and deep oceans (Longhurst and Glen Harrison, 1989; Steinberg and Landry, 2017; Turner, 2002, 2015). As a major link between primary producers and higher trophic levels (Ikeda, 1985), zooplankton have central ecological as well as biogeochemical roles, with associated socio-economic interests. This socio-economic impact of plankton can be positive, such as their role as food source for fish (Lehodey et al., 2006; van der Lingen et al., 2006) or as an indicator of water quality (Suthers et al., 2019). It can also be negative, as e.g. harmful blooms of phytoplankton impact aquaculture and human health (Griffith and Gobler, 2020), jellyfish blooms can impact various human activities such as aquaculture and fishing (Richardson et al., 2009).

### 2.2.2 Spatial Distribution of zooplankton and its biomass

Zooplankton organisms are sensitive to environmental conditions and are thus considered sentinels of ocean changes. Their distribution is finely governed by the interactions between physical (i.e., temperature (Steinberg and Landry, 2017), currents, light (Hays et al., 2005), pressure) and chemical constraints (nutrients, oxygen (Steinberg and Landry, 2017)), but also by biological interactions (e.g. predator-prey, symbiosis, parasitism and commensalism). The dependence of zooplankton on environmental variables leads to very clear global scale patterns even at coarse taxonomic levels (Biard et al., 2016; Lucas et al., 2014). On a global scale, zooplankton diversity is higher at the equator and decreases towards the poles (Ibarbalz et al., 2019; Rombouts et al., 2009). Conversely, zooplankton biomass tends to increase with latitude and be low in the tropics, with large seasonal fluctuations in temperate and polar regions (Ikeda, 1985; Moriarty et al., 2012; Soviadan et al., 2022). Although global quantitative assessment of zooplankton biomass and functional groups is needed (e.g. to be incorporated in biogeochemical and ecological models), it is often hampered by the heterogeneity of measurement methods and the uneven distribution of observations,

causing high uncertainty in biomass estimates (Le Quéré et al., 2016; Moriarty and O'Brien, 2013; Moriarty et al., 2012).

### 2.2.3 The study of zooplankton and its difficulties

Assessments of the global distribution of zooplankton organisms are often based on regional datasets, obtained with heterogeneous sampling tools traditionally biased towards non-gelatinous taxa (Lucas et al., 2014), and combined using different standardization procedures (Buitenhuis et al., 2013; Moriarty and O'Brien, 2013; Moriarty et al., 2012). Consequently, the global distribution of only a few zooplankton groups that generally can be well sampled using plankton nets, e.g. crustaceans, have been well studied (Buitenhuis et al., 2013; Rombouts et al., 2009). Indeed, some zooplankton taxa are known to be fragile (cnidarians, ctenophores, rhizarians, etc.) and their destruction by traditional sampling methods (e.g., plankton nets) as well as their poor preservation in fixatives (Beers and Stewart, 1970) resulted in an underestimation of their biomass and their ecological role in marine ecosystems (Biard et al., 2016; Lucas et al., 2014). In this context, non-intrusive *in-situ* methods using imaging (Cowen and Guigand, 2008; Grossmann et al., 2015; Picheral et al., 2010; Remsen et al., 2004; Schulz et al., 2010; L. Stemmann et al., 2008; Sun et al., 2008) and video (Davis et al., 1992, 2005; Hoving et al., 2019) instruments have been developed in the past decades to allow the non destructive study of zooplankton (Lombard et al., 2019). Among the different systems, only the Underwater Vision Profiler (UVP) version 4 and 5 have been widely used for plankton on a global level which allowed comparisons of abundance patterns with the Longhurst (1995) provinces of the ocean (Biard et al., 2016; L. Stemmann et al., 2008). Since 2008, the creation and expansion of such a global dataset could be executed with the UVP5 thanks to numerous participating teams around the world and the wide commercialization of this *in situ* imaging tool. In this study, we used data from the UVP5, an *in-situ* imaging system designed to detect, measure and quantify the distribution of zooplankton organisms and marine particles (Picheral et al., 2010). This instrument, designed for the study of particle size spectra in the ocean (Guidi et al., 2009; Stemmann et al., 2002) was also previously used to obtain plankton data at a high spatial resolution (Forest et al., 2012) and to study fragile organisms (Biard and Ohman, 2020; Biard et al., 2016; Christiansen et al., 2018; Stukel et al., 2018). However, even with the progressive increase in the spatio-temporal density of observations allowed by the use of imaging instruments, the unevenness in the distribution of observations remains, preventing large scale biomass estimations. Such global observation could nevertheless serve as the basis for large scale

estimations through the use of interpolation or extrapolation methods, including statistical habitat models.

#### 2.2.4 Statistical habitat models

Habitat modeling is a machine learning tool to estimate the abundance of a taxon at an unobserved location: instead of interpolating between nearby observation points based on geographical distance, the environmental conditions (i.e. the habitat) are used to inform the estimation. Statistically, a regression analysis can be used to define the relationship between the abundance (or presence) of a taxon at observation sites and the environmental variables at those sites (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000). Then continuous maps of those environment variables can be used to predict continuous maps of the taxon's abundance (or presence), by applying the regression.

The objective of this work was the development of a method to estimate zooplankton biomass on a global scale and to study the spatial distribution of zooplankton in relation to its habitat. To obtain such a global view we used global data from the UVP5 *in-situ* imaging system. It is in most cases impossible to identify the imaged organisms to species level. We therefore applied the habitat modeling approach to broader taxonomic groups. We first estimated the individual biovolume and biomass of organisms classified in 25 broad taxonomic groups, within a global *in-situ* imaging dataset. We then applied the habitat model methodology to each taxonomic group and built models using different regional and vertical partitions of the data. E.g. we separated data of the epipelagic (0-200 m depth layer) from the upper mesopelagic (200 to 500 m depth layer), but also used a global partitioning to separate data from low latitudes between 40°S to 40°N from the remaining high latitude data. We hypothesize that these partitions should allow us to separate subgroups within those broad taxa, which occupy different horizontal and/or vertical habitats. Finally, we used the models' output to estimate the global marine zooplankton biomass distribution in the first 500 m of the water column.

*In situ* imaging observation with UVP5 have been widely used during the past decade to study zooplankton in the global ocean. Biard et al. (2016) used 694 stations from the UVP5 dataset to reveal that Rhizaria organisms were strongly underestimated in previous studies. We here use an updated version of this dataset, now including 3 529 stations to study the biomass distribution of Copepoda, Rhizaria and several other groups of planktonic organisms in the 1.02-50 mm size range. Based on current knowledge about zooplankton, we expect to obtain different distributions for different taxonomic groups. We hypothesize that the total biomass of zooplankton is distributed according to regional production characteristics, associated to climatic and hydrological patterns,

showing overall a high biomass in high latitudes and lower values in the subtropical gyres (Ikeda, 1985; Moriarty et al., 2012).

## 2.3 Material and methods

### 2.3.1 Plankton data collection and processing

#### 2.3.1.1 Global plankton imaging with the UVP5

UVP5 data (Fig. 2.1) was compiled from all oceans, covering a 10 year period (2008-2018). A detailed description of the operation of the UVP5 is given in Picheral et al. (2010). All particles large than  $\approx 100 \mu\text{m}$  in Equivalent Spherical Diameter (ESD) are measured and counted, but only images of particles (zooplankton and aggregates) larger than  $\approx 600 \mu\text{m}$  ESD are kept by the UVP5 for further processing because smaller objects contain too few pixels to be identifiable. Acquisition of metadata (geographic location, date, etc.) and processing of all 8.46 million images (95% being detritus) was carried out by the ZooProcess software which provided information on 42 morphological features associated with each object (area, major and minor axis, etc.). The result was imported into EcoTaxa (Picheral et al., 2017), an application which allows a taxonomic classification of images via supervised learning algorithms, followed by manual validation (Irison et al., 2022). As 61% of the profiles have a maximum depth  $\leq 500$  m, only images of organisms between 0-500 m were kept and the overall estimate of biomass were restricted to this depth range. To ensure that profiles were representative, a filter was also applied to only keep profiles that covered at least 80% of the layer of interest.

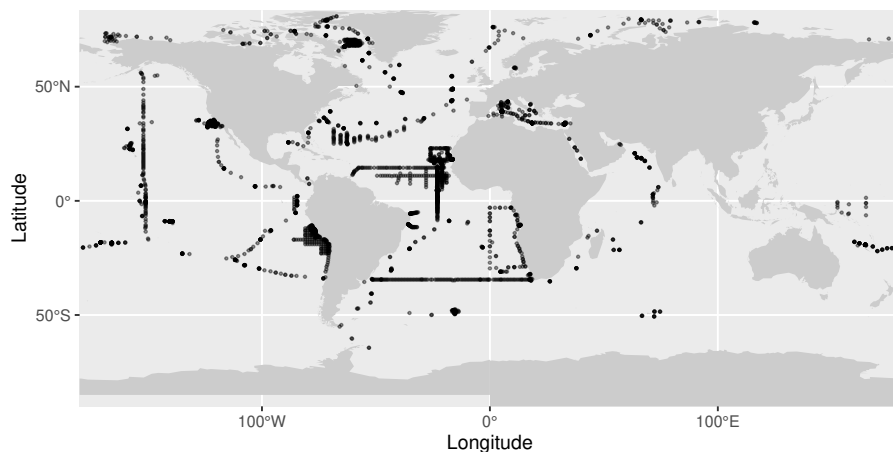


Figure 2.1: Map of the UVP5 dataset used in this study. Transparency was used to illustrate the density of points on the map.

### 2.3.1.2 Image classification and size range covered

Living organisms were separated from detritus (aggregates, fibers, fecal pellets) as well as artifacts (e.g. bubbles) and classified according to their taxonomic identity. Recognition and sorting of organisms can be a source of bias depending on the levels of perception and experience of the people who perform them. Several cognitive factors biases such as boredom, fatigue or a classification biased towards the most used groups have been presented by Culverhouse (2007) and Culverhouse et al. (2014). To reduce the risk of poor identification, a shared UVP5 taxonomic guide was used to homogenize image sorting into 119 taxonomic groups. The image data was thereafter grouped into 25 broader taxonomic groups (Table S1), and a subset of the resulting dataset was checked for homogeneity of sorting within these groups. A minimum of 51 images and a maximum of 10% of all images were extracted from each group and were independently checked after the assembly of the final data set. The maximum error or uncertainty rate per taxon was 9.8% and a vast majority of taxa were under 2.5%. We checked the classification and if accuracy was  $<95\%$ , we rechecked the categories to assure proper sorting. In addition, only fully validated profiles were used for this analysis. The resulting global data set consisted of 466 872 images from 3 529 stations. Under-sampled groups with less than 500 images in the dataset which could not be used for a global study were not included in the analysis.

We computed the organisms' size spectrum to detect the size range within which the UVP5 can be used to properly quantify their distribution. The concentration of objects in the ocean is expected to decrease with size; when this is computed as a normalized size spectrum, the relationship is expected to be linear (Forest et al., 2012). A peak in the size spectrum at the lower size range generally reflects the minimum size of objects efficiently recognized by *in situ* imaging while high variability in the large size range reflects the poor ability to detect rare large objects (Stemmann and Boss, 2012). With that in mind, a linear shape of the spectrum was observed between 1.02 and 50 mm with a peak found in many conditions at size smaller than the bin range 1.02-1.29 mm while stronger variability was observed at size larger than 50 mm ESD. We therefore excluded organisms outside of this range from the further analysis, as e.g. large mobile fauna (including large crustaceans) are likely to be undersampled and small zooplankton organisms close to the UVP5's threshold of detection are difficult to identify. This size range selection ensures that the data used in this study was properly quantified by the UVP5.

Method	Formula
Spheroid	$\frac{4}{3} \times \pi \times \left(\frac{ESD}{2}\right)^3$ with $ESD = 2 \times \sqrt{\frac{Area}{\pi}}$
Ellipsoid	$\frac{4}{3} \times \pi \times \frac{major}{2} \times \left(\frac{minor}{2}\right)^2$

Table 2.1: Methods of calculating individual biovolume with area ( $mm^2$ ); *ESD*, the equivalent spherical diameter equivalent (mm); *major*, the major axis (mm) of the best fit ellipse; *minor*, the minor axis (mm) of the most suitable ellipse.

### 2.3.1.3 Individual biomass estimation

To avoid errors due to incorrect ellipse fits (around appendages of organisms rather than their body, ellipse fitted to non-ellipsoidal organisms, etc.), we chose the spheroid method: it is based on the area (Table 2.1), which is more consistently measured by the image analysis performed in ZooProcess.

For Rhizaria, biovolume ( $mm^3$ ) to carbon ( $mgC$ ) conversions were done using factors from the literature (Fig. S2.1, Table S2). For other groups, the conversion from individual volume to individual wet weight assumed a density of  $1gcm^3$  (Kjørboe, 2013). Then the conversion from individual wet weight to individual biomass in carbon units ( $mgC$ ) was calculated using taxon-specific linear conversion factors from McConville et al. (2016); when several factors were available for a taxon, their median was used for each group). To take into account differences in density of some parts of the organisms, the Appendicularia group was actually split into Appendicularia\_body and Appendicularia\_house, whereby the "body" group contains images with only the animal and the "house" group contains the house and the animal. For the images labeled Appendicularia\_house, we used the relationship of house diameter (major axis) to Appendicularia trunk length from Lombard and Kjørboe (2010). We then converted this body size equivalent into carbon weight using the corresponding relationship from Lombard et al. (2009). For the images labeled Appendicularia\_body, we converted the biovolume of the organism into carbon weight using the corresponding relationship from Lombard et al. (2009). Two groups also have been created to separate the Collodaria into solitary Collodaria and colonial Collodaria. This choice was done based on the fact that solitary Collodaria are smaller than colonial ones and have a different vertical distribution (Faillettaz et al., 2016). For solitary collodarians with a dark central capsule (subgroup of solitary Collodaria) described in Biard et al. (2016), the estimation of carbon ( $0.189 mgC mm^{-3}$ ) by (Mansour et al., 2021) was done on the capsule of the organisms. As ZooProcess measures the area of the whole organism, we determined the ratio  $\frac{area\ whole\ organism}{area\ central\ capsule} = 0.713$  and applied this factor to avoid overestimation

of carbon biomass for this group. For the rest of the collodarians, the estimation of (Mansour et al., 2021) was directly applied.

### 2.3.2 Environmental data collection and processing

In order to develop relationships between regional characteristics of the environment (Fig. S2.2 to 2.4) and observed biomass, climatologies from the World Ocean Atlas (WOA) (Garcia et al., 2019) were used for temperature (in °C), salinity, oxygen (converted from  $\mu\text{mol kg}^{-1}$  to kPa for better physiological interpretation), and macronutrients (nitrate, phosphate and silicate in  $\mu\text{mol kg}^{-1}$ ). We selected the data sets defined on a  $1^\circ$  horizontal grid, over the 0-500 m range, and with a monthly temporal resolution. Temporal coverage was 2005 to 2017 for salinity and temperature and 1955 to 2017 for the other variables. We also used monthly averaged surface chlorophyll-a data (Chl *a* in  $\text{mg.m}^{-3}$ ) resolved to  $1/24^\circ$  from 2005 to 2017 from the Copernicus database (OCEANCOLOUR\_GLO\_CHL\_L4\_REP\_OBSERVATIONS\_009\_082) as well as bathymetry data from NOAA (Amante and Eakins, 2009) with a spatial resolution of 10 minutes; both were re-gridded to a  $1^\circ$  grid. Finally, distance to coast was computed by calculating the distance of all  $1^\circ \times 1^\circ$  cells to the closest cell associated to land using the raster package (Hijmans, 2021). To obtain annual climatologies, when relevant, each monthly variable was averaged over its time period of coverage.

This environmental data was then matched to the UVP5 data on the  $1^\circ \times 1^\circ$  grid. Since the  $1^\circ \times 1^\circ$  grid used by WOA does not necessarily follow the contour line of the coast perfectly, some UVP5 profiles could not be directly matched to the environmental grids. This is mostly the case where e.g. the coast is situated in a 45 degree angle to latitude or longitude, thereby creating triangle shaped areas that are not covered by the rectangular grid. For profiles that lie in such corners of the grid, we used the environmental values of the closest neighboring  $1^\circ \times 1^\circ$  WOA cell. In the epipelagic world model, 3 002 points have a direct match while 156 points do not have a direct match. Out of these 156 points, 14 are not in a neighboring  $1^\circ \times 1^\circ$  WOA cell and were removed from the model input. For the mesopelagic, 2 172 have a direct match, while 104 points have a match in a neighboring grid cell and 2 points do not and were removed from the model input. Maps that show the close vicinity of non-matching points to adjacent WOA cells are shown in supplementary figure 5.

To assess whether we are able to describe various environmental conditions with the UVP5 samples, we compared the distributions of each variable in the worldwide WOA dataset and in the subset matched to UVP5 profiles (Fig. S2.6 and S2.7). Although the geographical coverage is not homogeneous (Fig. 2.1), the coverage of environmental conditions is good and warrants the use of habitat models.



### 2.3.3 Habitat modeling

The steps of this process are summarized in figure 2.2.

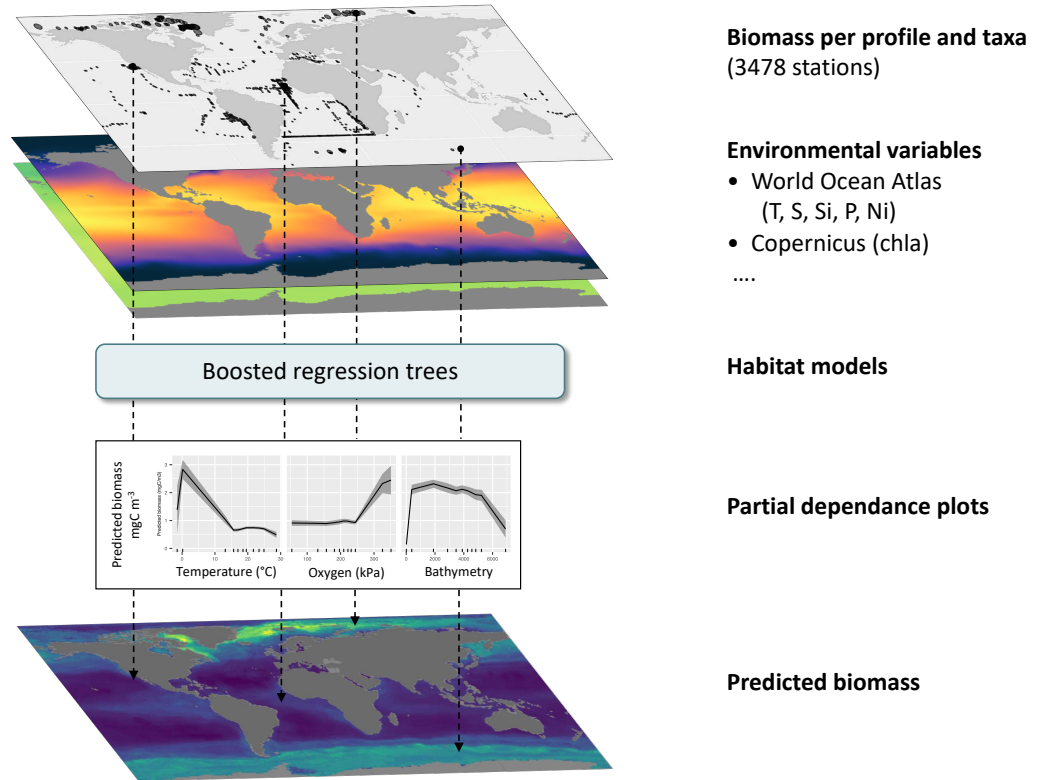


Figure 2.2: Methodology followed from data selection to prediction of global biomass.

#### 2.3.3.1 Modeling tools

In this work we used boosted regression trees (BRTs) to predict the biomass of different zooplankton groups as they show different advantages over other commonly used machine learning approaches for the nature of our dataset and intended application (Elith and Graham, 2009). This ensemble method uses regression trees, models that link a response (here biomass) to predictors (environmental variables) by successive dichotomous separations (Breiman et al., 1984; Hastie et al., 2001). Regression trees automatically select the relevant explanatory variables, can deal with categorical or continuous inputs, are not sensitive to the distribution of the continuous ones, can represent relations of arbitrary form and naturally include interactions among explanatory variables (Elith et al., 2006). With so-called surrogate splits, they can also deal with missing values in the explanatory variables. They

are therefore very comfortable to use, but their predictive power is often limited and they have difficulties to capture smooth relationships. Boosting is a way to overcome these drawbacks (Schapire, 2003). It is based on the fact that it is easier to find many rough rules of thumb than to find a single, highly accurate prediction rule (Schapire, 2003). BRTs combine many short regression trees in succession, each new tree being adjusted to consider the observations poorly predicted by the previous ones (Elith et al., 2008; Elith et al., 2006; Leathwick et al., 2006). This improves predictive performance and the smoothness of the prediction (Leathwick et al., 2006). In addition, only a random subset of the input data is used to fit each tree and this stochastic component reduces the variance of the final model ensemble (Friedman, 2002).

Boosted regression trees (BRTs) have an ability to handle a large number of variables and - other than Generalised Linear Models (GLMs, Nelder and Wedderburn (1972)) or Generalised Additive Models (GAMs, De'ath (2007), Elith et al. (2008), and Hastie and Tibshirani (1986)) - does not seek to fit one single model portraying the relationship of the response variable (here biomass) and its predictors (environmental variables). Various recent studies (Chen et al., 2020; González Carman et al., 2019; Hu et al., 2021) have compared BRTs results to other modeling tools such as GAMs, GLMs, RFs, Maximum Entropy modeling (Elith and Graham, 2009; S. J. Phillips et al., 2006) or neural networks and have obtained better predictive performance with BRTs. Other studies (Son et al., 2018; Zhang et al., 2018) used complementary GAMs and BRTs to study the effects of explanatory variables. However, BRTs could be slower than RFs (Chen et al., 2020) and training parameters need to be chosen carefully to avoid overfitting (Elith and Graham, 2009; Leathwick et al., 2006). BRTs were chosen over Random Forests (RFs) because of their capacity to reduce both the bias and the variance of model results (Hastie et al., 2001). BRTs are also less sensitive to the effect of extreme outliers and the inclusion of irrelevant predictors (Leathwick et al., 2006). This makes them suitable for plankton datasets, as sometimes very high plankton biomass values do occur during blooms (Brodeur et al., 2018; Pettitt-Wade et al., 2020). BRTs also have the ability to handle sharp discontinuities which is not the case of the GAMs (Elith et al., 2008) which is important when modeling taxa which can have a narrow habitat.

In addition, in regression trees, the loss function, used to choose which dichotomous split to perform, can be changed to be adapted to the distribution of residuals. Here we explored the classic mean squared error, which assumed a somewhat normal distribution of the residuals, as well as a Tweedie loss adapted to zero-inflated data (Zhou et al., 2019), and a Poissonian loss, which considered data as discrete counts, also including many zeros. To use the Poisson loss, the biomass was scaled so that the value of the 1% quantile was  $\geq 1$  and then

rounded to the nearest integer; the inverse scaling was performed after prediction. This later approach proved to produce the best fits and more robust models in a few test taxa and all models were therefore fitted with Poisson loss. The models and statistics were computed using the `xgboost` package (Chen et al., 2021) in R version 4.1.2 (R, 2021).

### 2.3.3.2 Spatial partitioning of the data

Individual biomass values derived from UVP5 images and environmental data measured at various layers were both averaged over a depth range of interest and matched geographically, on the  $1^\circ \times 1^\circ$  grid. Biomass values matched to the same  $1^\circ$  pixel, and therefore associated to exactly the same environmental data, were averaged.

We hypothesized that an association between biomass and environment done at a fine scale could be more efficiently learned by the model because it contains less noise, so we divided the data vertically between the epipelagic (0-200 m) and mesopelagic (200-500 m) zones and also tried a finer partition, into 100 m depth bins between 0 and 500 m. Learning separate models for each layer could allow to focus on finer subgroups within our quite coarse taxonomic units (some species being mostly present in one of the layers) and therefore define biomass-habitat relationships at a finer, more relevant biological level.

For the same reason, we also built models on subsets of data partitioned geographically. Indeed, polar copepods have a different thermal niche compared to tropical ones (McGinty et al., 2021; Rombouts et al., 2009). So, in addition to a model fitted on the whole world (`world`), we trained models on data from the region between  $40^\circ\text{S}$  and  $40^\circ\text{N}$  (`low latitude`) and from the data collected outside of this latitudinal band (`high latitude`). Out of the 3 529 profiles composing the UVP5 dataset, 2 837 are located between  $40^\circ\text{S}$  and  $40^\circ\text{N}$  and 712 were done outside of this latitudinal band.

### 2.3.3.3 Data splits for model training, assessment and evaluation

For each taxon in each spatial partition, the data was split to distribute 80% of it in a training and validation set, on which the model was fitted and assessed, and 20% to a test set, on which predictive performance was evaluated. This split was stratified according to the deciles of biomass in the data, to ensure that both the learning and test sets contained low and high biomass points.

To choose model hyperparameters (i.e. parameters of the model adjustment algorithm) and to evaluate the variability in the prediction due to the constitution of the training set, each 80% portion set was resampled through five-fold cross validation repeated 20 times (i.e. 100 resamples; Hastie et al. (2001)). For each cross-validation fold, the model was actually trained on four folds and validated on the last one.

The splits into the five folds were also stratified according to the deciles of biomass, for the same reason invoked above.

#### 2.3.3.4 Selection of hyperparameters and model evaluation

To extract as much information from the data, while avoiding overfitting, various combinations of hyperparameters were tested for each model (Elith et al., 2006). They included: 1) the learning rate per tree determining the contribution of each tree to the ensemble model (0.05, 0.08 and 0.1 were tested); 2) the maximum depth of a tree (2, 4 and 8 were tested); 3) the minimum number of elements per leaf (which also limits the depth of the trees; 1, 3 and 5 were used); 4) the number of trees used for the prediction (values up to 600 were tested). For each combination, the model was fitted on the training set and evaluated on the validation set of each of the 100 resamples; the loss was then averaged over the 100 resamples. The best set of hyperparameters is usually the one for which this average loss is minimal. The differences around that minimum are often small and not always meaningful; to be sure to avoid overfitting, we applied an early stopping criterion whereby the increase in the number of trees was stopped when the error did not decrease by more than 1% after adding 10 trees.

Once the best set of hyperparameters had been chosen, the relevance of the corresponding model was quantified by the Pearson correlation between the observed biomass data in the test set and the predicted biomass, where prediction is the average of the predictions of the 100 models fitted on the resamples. This metric captures the model's ability to correctly represent general trends and patterns in the data set and is one way to compute the  $R^2$ . The significance of this correlation can also be tested and quantified with a  $p$ -value. These metrics can be readily compared across the various spatial partitions of the data because they represent the skill of the models on an independent data set, not the quality of the fit on the training data (like the way the  $R^2$  is usually computed). To compare the worldwide and regional approaches fairly, it is important to focus on the same regional subset. To this effect, two additional  $R^2$  were computed for the global model: on the test data located inside the 40°S-40°N latitudinal band and on those outside of it (world low latitude and world high latitude).

#### 2.3.3.5 Effect of environmental variables

To identify which environmental variables drive the change of biomass in each specific model, the percentage of variance explained by each variable was calculated as the sum of the effects of the variable at each node of each tree where it was used. To describe the shape of the effect of each variable, univariate partial dependence plots were computed as the average  $\pm$  standard deviation marginal effect of the

variable in the 100 resamples. Practically, the variable of interest was set at a given value at all training points, the other variables were left at their original values, the average biomass predicted over all points was computed, for each resample; then the mean and standard deviation of those averages were computed across resamples. Finally, the variable was set to another value and so on. To describe the full range of each variable, the partial dependence was estimated at 10% quantile.

#### 2.3.3.6 Extrapolation to the globe

To obtain global maps of predicted biomass, the regression between UVP5 biomass data and environmental variables was applied to all points in the corresponding partition of the world, in depth and space. Because 100 models were fitted to the resamples of the training data, the standard deviation of biomass among the 100 predictions ( $\sigma_b$ ) can be computed in addition to the mean ( $m_b$ ), and the coefficient of variation (CV), defined as  $CV = \frac{\sigma_b}{m_b}$ , then gives an indication of the uncertainty of the model predictions.

To get a robust estimate of global zooplankton biomass in the 1.02 mm to 50 mm size range, we chose to be conservative (i.e. *ad minima*): only the taxonomic groups in the global partition for which the correlation between predicted and observed biomass was significant were used. The surface area of each  $1^\circ \times 1^\circ$  cell was computed using the following formula:

$$A = \frac{\Pi}{180} \times \mathcal{R}^2 \times (\sin(lat_S) - \sin(lat_N)) \times 10^6$$

with the area  $A$  in  $m^2$ , the south and north latitudinal limits of the cell in radians and  $\mathcal{R}$ , the earth radius (6378.137 km). For each group used, the biomass was integrated over the relevant layer in each  $1^\circ \times 1^\circ$  cell by the following calculation

$$\hat{b}_t = \hat{b} \times A \times l.$$

where  $\hat{b}$  is the estimated biomass in  $mgC.m^{-3}$ ,  $A$  in  $m^2$  is defined above,  $l$  is the layer thickness in  $m$  and therefore  $\hat{b}_t$  is the total biomass in  $mgC$ . Finally, the global *ad minima* zooplankton biomass estimate was computed by adding up the biomass for all selected groups and the 0-200 and 200-500 m depth layer.

## 2.4 Results

### 2.4.1 Model comparison

We estimated model performance on the worldwide UVP5 dataset and on a spatial partition of the dataset in low (inside 40°N and 40°S) and high latitudes (outside of the 40°N-40°S latitudinal band) as well as on different depth layers. We hypothesized that a finer data selection might enable the respective model to learn the regional or depth specific habitat more appropriately. Yet, this also meant fitting models on fewer data points. In the end, we find that no clear trend emerges from the relevant comparisons (Fig. 2.3) : global models are better in 13 comparisons and partitioned models are better in 14 comparisons, whereas for 11 comparisons no clear decision can be made. Comparisons can only be made within a given depth layer between the same regional partitions (e.g. world low latitude only containing the data predicted by the global model between 40°N-40°S vs low latitude ; world high latitude only containing data north of 40°N and south of 40°S from the global model vs high latitude).

In more detail, for some groups such as Annelida and some Mollusca, the high latitude model could not be computed (symbolized by a grey cell) either because they were considered as rare (<500 images in the layer modeled) or because the model could not learn the link between biomass and environment for this group. However, for others such as Copepoda, solitary Collodaria or Phaeodaria, high and low latitude models are generally better than the world model, as indicated by a higher  $R^2$  value (Fig. 2.3). In the epipelagic layer, for Copepoda, the  $R^2$  of world low latitude is 0.26 vs 0.37 in the low latitude model. For the mesopelagic, low latitude has an  $R^2$  of 0.07, lower than the one for world low latitude (0.62). For Appendicularia in the epipelagic layer, the best  $R^2$  values are obtained in the world low latitude (0.41) and world high latitude (0.24) models respectively compared to low latitude (0.01) and high latitude (0.19).

As for the vertical 100 m-bin layers partition, we obtained the best results overall with the global model. The finer vertical definition also gives better results for multiple other groups such as Appendicularia, Phaeodaria and Ostracoda between 0 and 300 m. In most cases, only the top 100m layer model worked for this 100 m vertical partition. Overall, the most consistently good choice, when considering all taxa, is a worldwide model fitted separately on the epipelagic (0-200 m) and mesopelagic (200-500 m) layers. This is therefore the configuration retained for the total, global biomass estimate. In figure 2.3, taxa are arranged in decreasing order of global biomass in the epipelagic layer.

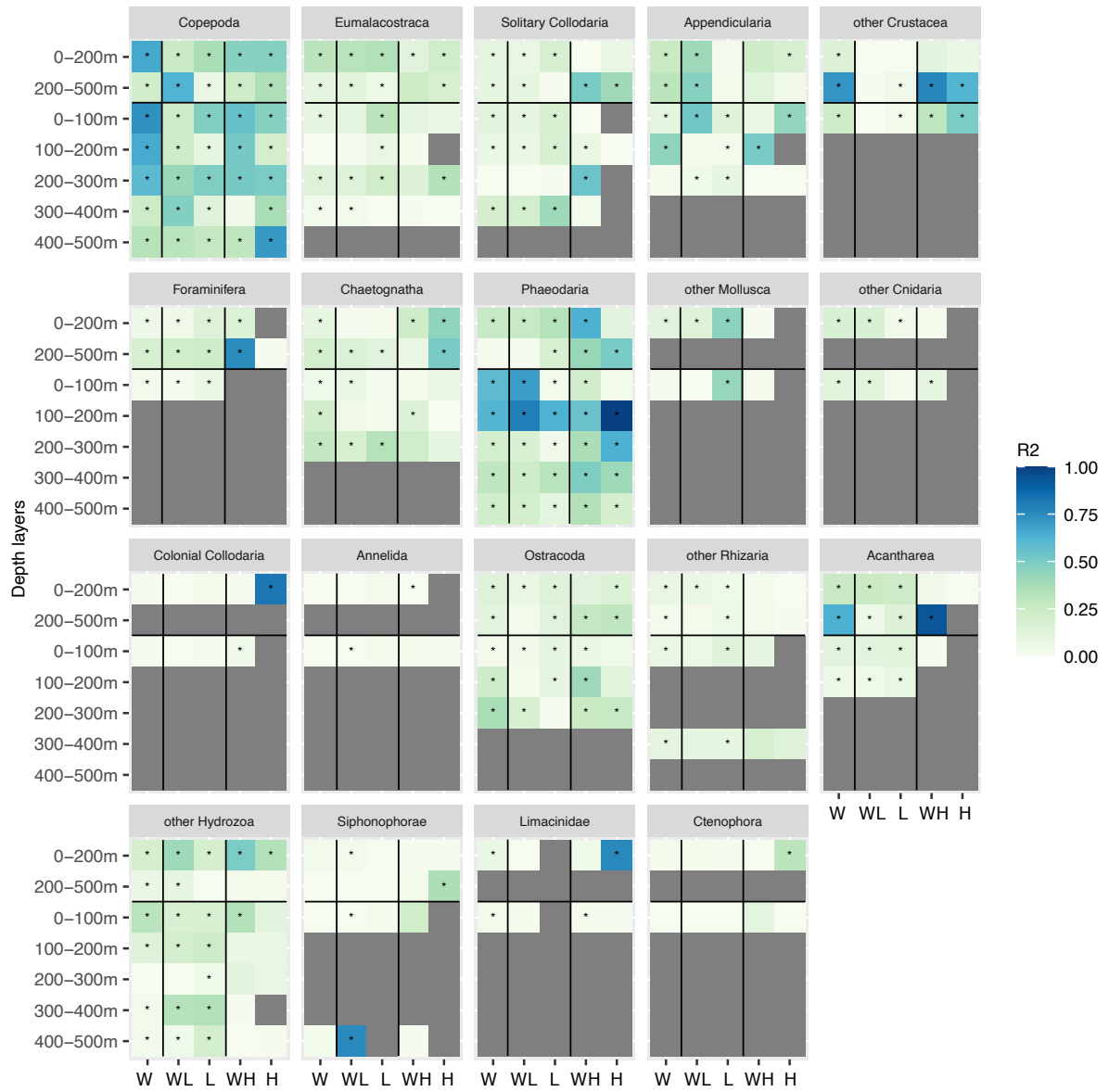


Figure 2.3: Heatmap of the models'  $R^2$  between observed and predicted biomass for all zooplankton groups arranged from the most important in terms of biomass (Copepoda) to the least important (Limacinidae) in the different depth layers. The regions correspond to: W for world (model run on all data) ; WL for world low (data between 40N and 40S from the world model) ; L for low latitude (model run between 40N and 40S) ; WH for world high (data outside of 40N and 40S from the world model) ; H for high latitude (model run outside of 40N and 40S). The stars indicate significant results ( $p$ -value  $< 0.05$ ) obtained with the Pearson correlation test.

For the top five (Copepoda ( $R^2 = 0.66$ ), Eumalacostraca ( $R^2 = 0.31$ ), solitary Collodaria ( $R^2 = 0.10$ ), Appendicularia ( $R^2 = 0.26$ )) other Crustacea ( $R^2 = 0.15$ ), the correlation between true and predicted biomass is significant ( $p$ -value  $< 0.05$ ) in the epipelagic worldwide model. In the mesopelagic layer the correlations for all five groups are also significant ( $p$ -value  $< 0.05$  with respective ( $R^2$  of 0.22, 0.10, 0.09, 0.30 and 0.72).

#### 2.4.2 Group-wise contribution to global zooplankton biomass

Figure 2.4 shows the biomass per group predicted for the three spatial partitions and divided into the epi- (0-200 m) and mesopelagic (200-500 m) layer. For the worldwide model, the dominant groups in terms of biomass in the epipelagic were Copepoda ( $0.083 \pm 0.020$  PgC), Eumalacostraca ( $0.058 \pm 0.017$  PgC) and solitary Collodaria ( $0.038 \pm 0.008$  PgC) (Fig. 2.4A). Among the groups displaying a significant correlation ( $p$ -value  $< 0.05$ ) between true and predicted biomass (and therefore retained for the global estimate), crustaceans (Copepoda, Eumalacostraca, other Crustacea and Ostracoda) represented 68.4% ( $0.157$  PgC) of the biomass in this layer; Rhizaria (solitary Collodaria, Foraminifera, Phaeodaria, other Rhizaria and Acantharea) made up 20.6% ( $0.047$  PgC); but the Cnidaria (other Cnidaria and other Hydrozoa) represented only 0.56% ( $0.0013$  PgC). In other words, Crustacea and Rhizaria together made up  $\sim 89.1\%$  of the biomass predicted in the epipelagic layer. In the deeper mesopelagic layer, Copepoda ( $0.061 \pm 0.016$  PgC) were still the dominant group in terms of biomass, followed by Eumalacostraca ( $0.049 \pm 0.014$  PgC) and other Crustaceans ( $0.017 \pm 0.001$  PgC) combined. Crustacea (Copepoda, Eumalacostraca, other Crustacea and Ostracoda) represented  $0.129$  PgC, equivalent to 74.4% of this layer's biomass, while Rhizaria (Foraminifera, solitary Collodaria, other Rhizaria and Acantharea) totaled  $0.014$  PgC, representing 10.1% equivalent to most of the remaining biomass in the layer. When combining the results from these two layers, Copepoda represented 44.4% of the global integrated biomass, followed by Eumalacostraca (15.6%), solitary Collodaria (13.1%) and other Crustacea (11.2%). And, more broadly, Crustacea (Copepoda, Eumalacostraca, other Crustacea and Ostracoda) represented  $0.222$  PgC or 71.3% of the biomass predicted over 0-500m, while Rhizaria (Foraminifera, solitary Collodaria, other Rhizaria and Acantharea) made up  $0.019$  PgC or 10.8% of biomass.

Copepoda were even more dominant in high latitudes especially in the epipelagic layer. In the low latitude model, solitary Collodaria were first in terms of biomass, followed by Eumalacostraca and Copepoda and Foraminifera in the epipelagic. In the mesopelagic layer, Eumalacostraca is dominant in low latitudes followed by Copepoda and Foraminifera.



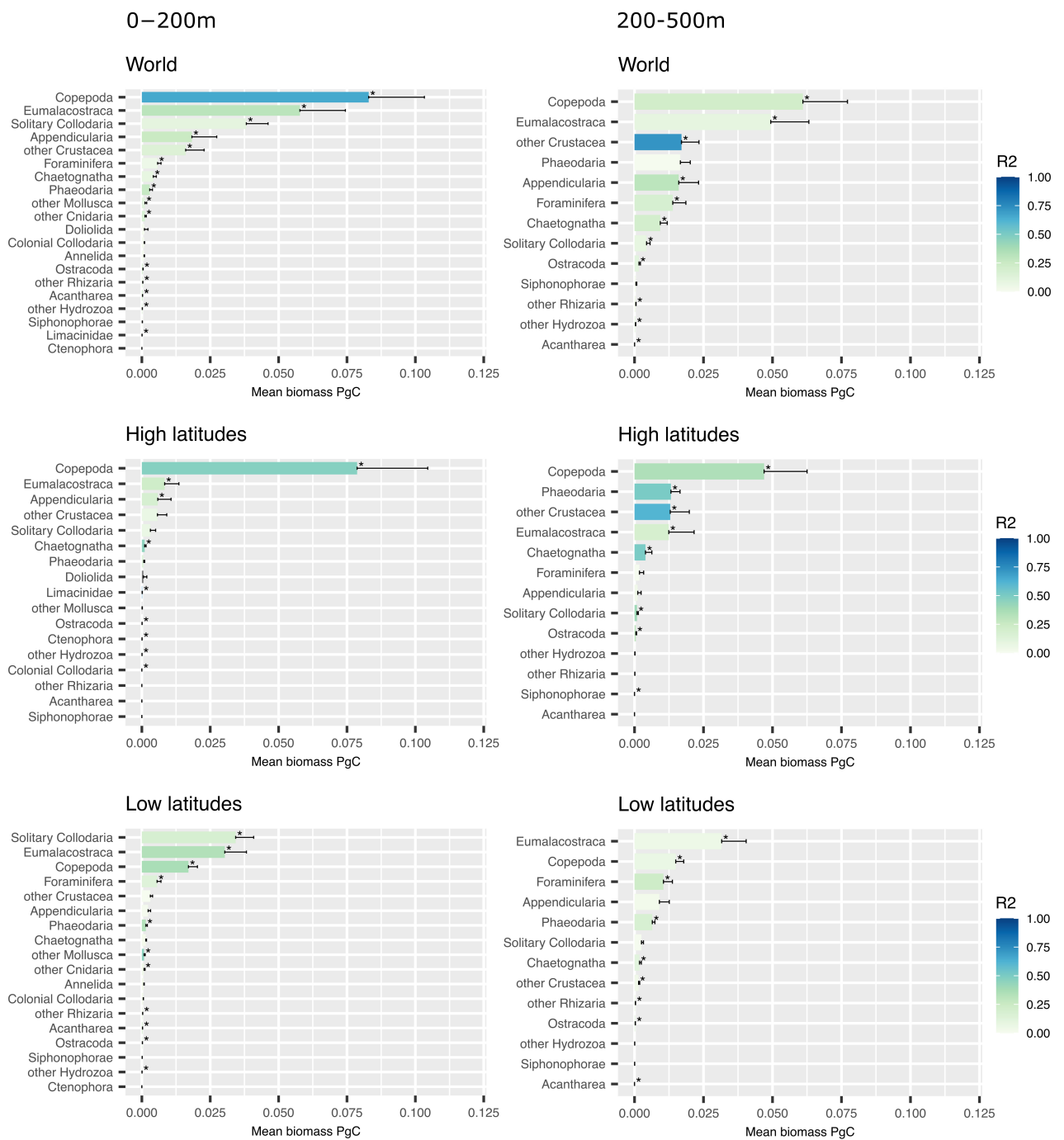


Figure 2.4: Barplots showing the mean biomass predicted in PgC at 0-200 m and 200-500 m depth for each group ranked from highest to lowest biomass in 3 types of models : world, outside 40N-40S and inside 40N-40S. Error bars correspond to upper interval of the biomass estimation's standard deviation. The stars indicate a significant result (p-value < 0.05) obtained with the Pearson correlation test.

### 2.4.3 Spatial distribution patterns and occupied habitat

Presenting the global distribution patterns of all zooplankton groups is beyond the scope of this paper. Instead, we focus on the results for the three groups contributing most to the total global biomass (Copepoda, Eumalacostraca and Solitary Collodaria) as well as on Phaeodaria and Acantharea, Rhizarians that were shown to be important contributors to zooplankton biomass that are underestimated by net-based sampling (Biard et al., 2016). The predicted fields for all modeled groups will be made available in PANGAEA upon publication of the article.

#### 2.4.3.1 Copepoda

Copepoda is one of the best predicted groups in the epipelagic  $R^2 = 0.66$ , likely because it is the most abundant. The structuring environmental variables were different for the epi- (Fig. S2.8A,B) and mesopelagic layers (Fig. S2.8C,D) : temperature (33%) and oxygen (19%) for the former and temperature (29%), bathymetry (19%) and chlorophyll *a* (15%) for the latter. Accordingly, in the 0-200 m depth layer (Fig. 2.5A), the highest predicted biomass can be found in high latitude water masses with low temperatures and high oxygen concentrations. In the mesopelagic layer (Fig. 2.5B), high biomass values were associated with shallow coastal and cold water masses. The patterns of distribution predicted by the global models were similar in both layers (Fig. 2.5A, B), with the highest predicted biomass values surrounding Greenland in the Baffin Bay, Labrador Sea and Greenland Sea as well as at the level of the Southern Ocean polar front. The lowest were predicted at oceanic gyres and in the Arctic and Antarctic Oceans. For both layers, the highest values of the coefficient of variation (Fig. 2.5C) were found in the Arctic Ocean, north of Canada and Greenland, as well as south of 60°S, especially for the epipelagic layer. These high values depict disagreement among the 100 models fitted on the data resamples and therefore inform on the uncertainty of the model in these zones. Caution is therefore advised regarding the interpretation of the very low values of biomass predicted in those regions. In the northern hemisphere, except for the Arctic ocean, the values of the coefficient of variation were rather low at locations where either low or high biomass values were predicted. In the southern hemisphere, model predictions varied relatively strongly at the level of the Antarctic polar front (Fig. 2.5C, D).

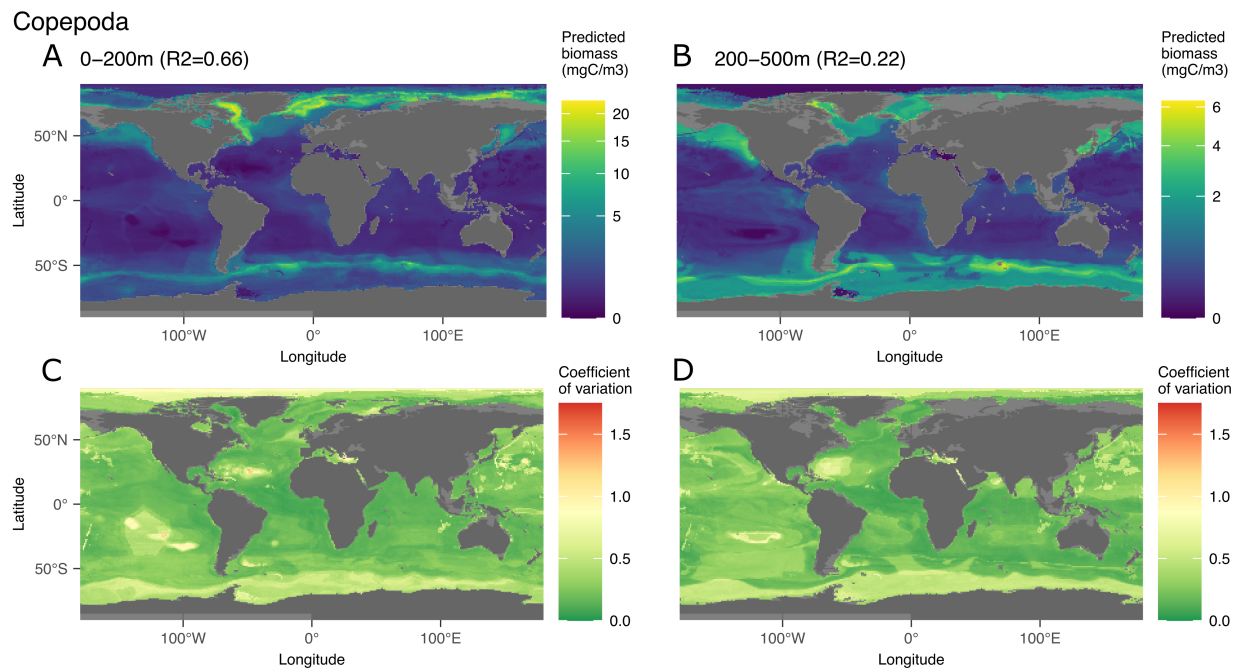


Figure 2.5: Map of the mean biomass (color scale is log-transformed) of Copepoda as predicted by the model on 0–200 m (A) 200–500 m data (B) as well as the coefficient of variation for the 0–200 m model (C) and 200–500 m one (D). The color scale for the coefficient of variation has the same range for figures 5 to 9.

#### 2.4.3.2 Eumalacostraca

Eumalacostraca contains mostly vignettes of euphausiids, amphipods and decapods. They were predicted globally with an  $R^2$  of 0.31 for the epi- and 0.1 for the mesopelagic layer, both with significant p-values (p-value < 0.05; Figure 2.3). In the epipelagic, high biomass of these organisms were explained by high concentrations of phosphate 22% and low concentrations of silicate (17%) (Fig. S9A,B). In the mesopelagic layer, the distribution of this group was associated with low concentrations of silicate (16%), bathymetry (15%) and high chlorophyll textita (15%) (Fig. S9C,D). In terms of spatial distribution, high biomass values are predicted along the eastern border of the continents, but especially at the level of the Peruvian and Californian upwellings (Fig. 2.6). Low biomass values are predicted in high latitudes and in the oceanic gyres, especially in the North Atlantic one. Similar patterns were predicted in the mesopelagic layer, but with lower biomass values. The model uncertainties are highest in the zones of low biomass (high latitudes and oceanic gyres).

#### 2.4.3.3 Solitary Collodaria

Solitary Collodaria were predicted globally with an  $R^2$  of 0.1 for the epi- and 0.09 for the mesopelagic layer, both with significant p-values (p-value < 0.05; Figure 2.3). In the epipelagic, the distribution of solitary Collodaria were mainly associated with low salinity (21%, between 35 and 37) and bathymetry (14%) (Fig. S10A,B). In the mesopelagic, high abundances of this group were explained by low distance to coast (18%) and high chlorophyll *a* (17%) (Fig. S10C,D). In this layer, 65% of the biomass was predicted at less than 1000 km distance from the coast. This groups' biomass was mainly located between 50°N and 50°S, in a rather diffuse manner (Fig. 2.7) with a maximum predicted at the equator. The highest biomass found in the intertropical range was found in productive areas such as the upwelling regions off the western coast of Africa (Cape Verde and Angola) and of the eastern boundary of the Pacific Ocean (Peru and California) in the epipelagic. The model also predicted high biomass values in the Mediterranean sea. The importance of the environmental variable "distance to coast" in the learning process created unusual patterns in the prediction map such as a hexagonal shape in the Pacific Ocean. North of 50°N and south of 50°S, environments characterized by low salinity (1<sup>st</sup> most structuring variable in the epipelagic) and high nitrate (4<sup>th</sup> variable), the predicted biomass was rather low especially in the epipelagic layer.

## Eumalacostraca

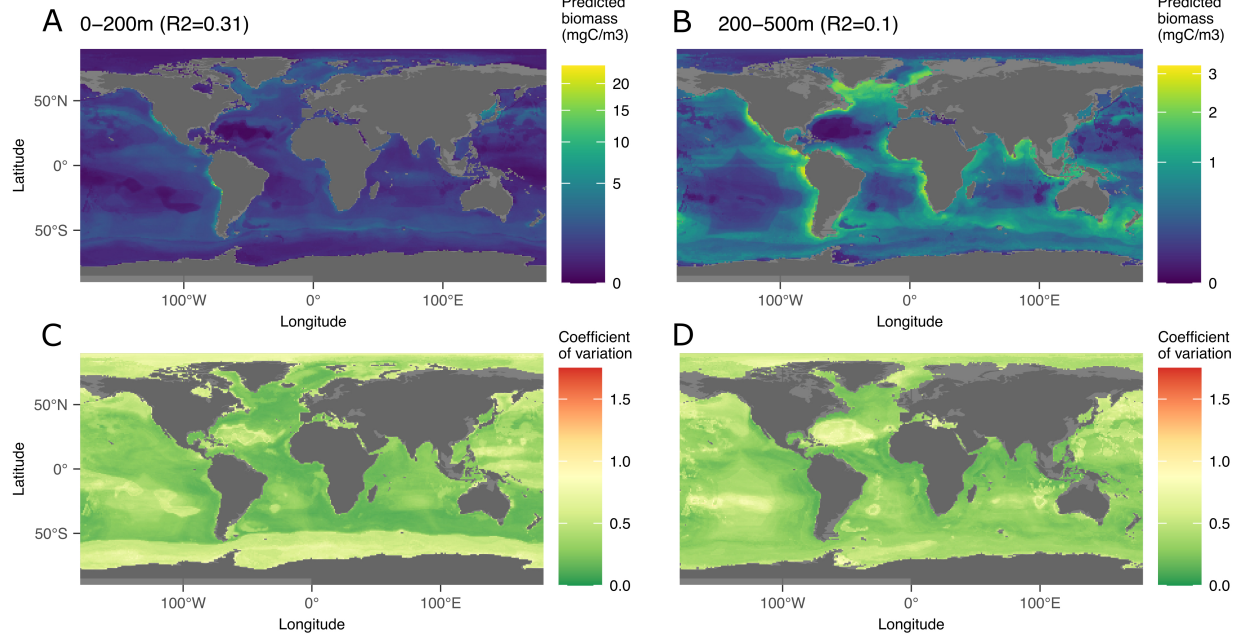


Figure 2.6: Map of the mean biomass (color scale is log-transformed) of Eumalacostraca as predicted by the model on 0-200 m (A) 200-500 m data (B) as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D).

## Solitary Collodaria

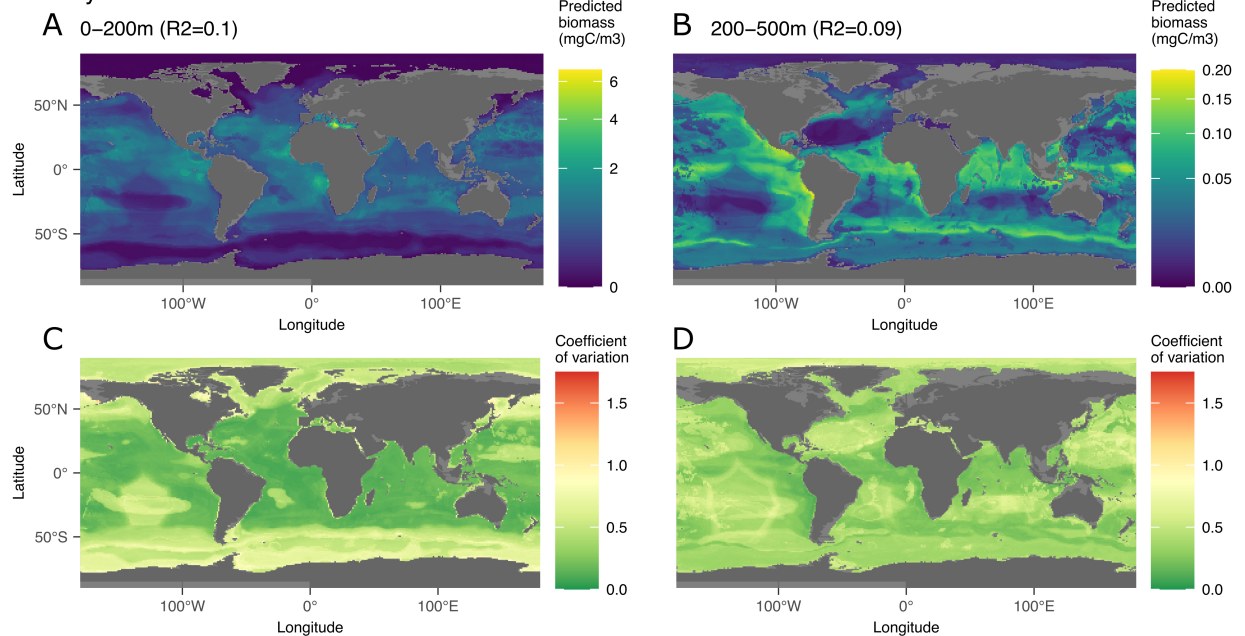


Figure 2.7: Map of the mean biomass (color scale is log-transformed) of solitary Collodaria as predicted by the model on 0-200 m (A), as well as the coefficient of variation (B) as well as the coefficient of variation for the 0-200 m model (C) and 200-500 m one (D).

#### 2.4.3.4 Phaeodaria

For this group, the worldwide epipelagic model was statistically significant ( $p$ -value  $< 0.05$ ; Figure 2.8) with an  $R^2$  of 0.27, but the mesopelagic model was not ( $p$ -value  $> 0.05$ ; Figure 2.3). Therefore, only the 0-200 m layer is displayed (Fig. 2.8). In this layer, Phaeodaria was one of the best predicted groups (Fig. 2.3) especially in the 100 m bin models between 0 and 200 m depth. The epipelagic distribution of Phaeodaria predictions is associated with low values of salinity (38%) followed by bathymetry (11%), surface chlorophyll *a* (10%) and oxygen and temperature (8% each) (Fig. S11A,B). This is visualised on the map of global prediction (Fig. 2.8A) on which high biomass was mainly predicted in the Californian upwelling (characterized by low salinity, cold and coastal waters), with lower biomass north of the upwelling up to the Gulf of Alaska, south of the upwelling. High biomass values were also predicted in the Bay of Bengal and Adaman Sea. The coefficient of variation in zones of high biomass is very low providing strong confidence in this pattern. The lowest predicted biomass for this group are found in oceanic gyres and high latitudes of the northern hemisphere.

#### 2.4.3.5 Acantharea

The group Acantharea was predicted with low total biomass (Fig. 2.4). This group was well predicted in the world model fitted in the epi- ( $R^2 = 0.26$ ) and mesopelagic ( $R^2 = 0.63$ ) layers (Fig. 2.9). In the epipelagic layer, nitrate (18%), salinity (15%) and phosphate (12%) were the main driving variables (Fig. S12A,B). In the mesopelagic layer, the link between biomass and environment (Fig. 2.9B) was defined by the influence of several variables: silicate (19%), phosphate (12%) followed by chlorophyll *a* (12%) (Fig. S12C,D). The highest epipelagic biomass (Fig. 2.9A) was predicted in the intertropical range, at the level of productive areas such as the upwellings off the West coasts of Africa (Cape Verde, Angola) and America (Peru and California). Those high biomass patches are associated with a salinity around 35 as the 2<sup>nd</sup> most structuring variable, as well as with high nitrate and phosphate concentrations (respectively 1<sup>st</sup> and 3<sup>rd</sup>). Intermediate biomass values were predicted mostly between 50°N and 50°S in a diffuse way, except in the oceanic gyres where the predicted biomass was lowest. The most important model uncertainty was present in the Southern and Arctic Oceans, Bering Sea and Gulf of Alaska where low biomass values were predicted (Fig. 2.9C). In the mesopelagic layer, biomass was predicted to be 16.7-times lower overall (Fig. 2.9B), with highest values found in the Gulf of Alaska and the Bering Sea. Intermediate biomass values were predicted for the upwelling regions and the Southern Ocean.

In this layer, the high biomass estimates correspond with low coefficient of variation values (Fig. 2.9D).

### Phaeodaria

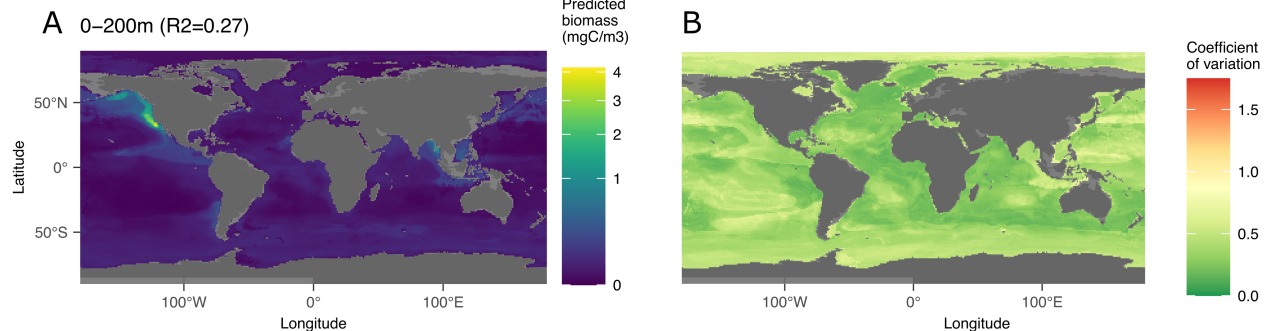


Figure 2.8: Map of the mean biomass (color scale is log-transformed) of Phaeodaria as predicted by the model on 0–200 m (A), as well as the coefficient of variation for the 0–200 m model (B). In the map of predicted biomass, 12 cells in the California upwelling presented a value between 3 and 6  $\text{mgC m}^{-3}$  and were represented here in yellow to observe the distribution of this group on a global scale. The color scale for the coefficient of variation has the same range for all plots.

### Acantharea

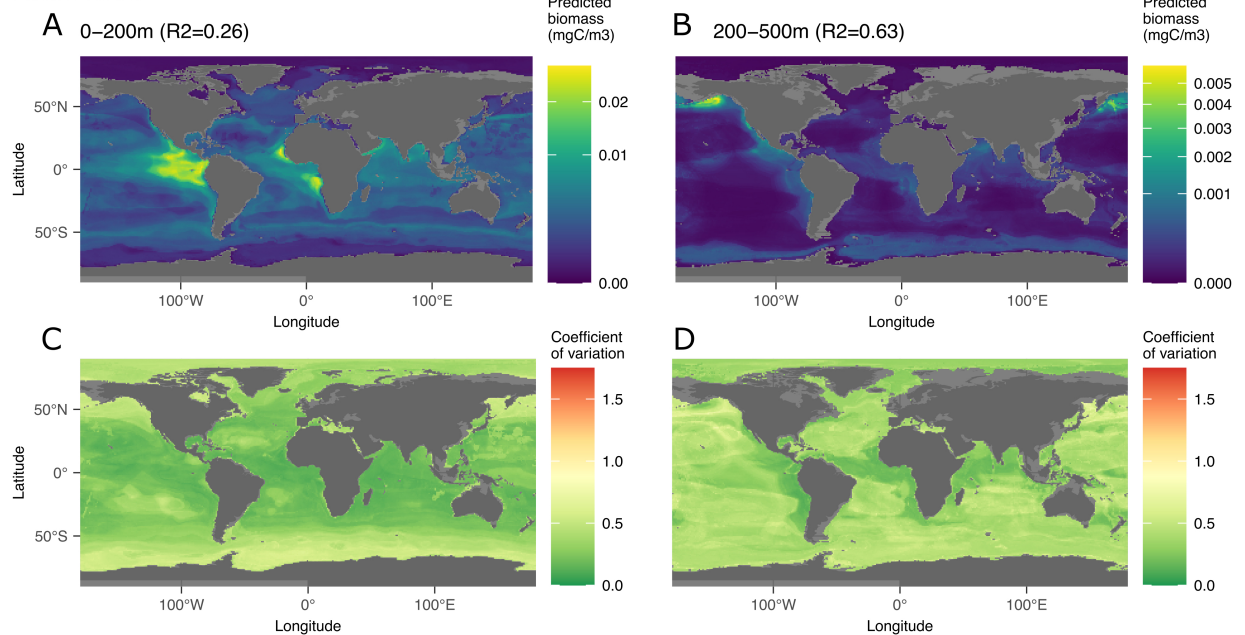


Figure 2.9: Map of the mean biomass (color scale is log-transformed) of Acantharea as predicted by the model on 0–200 m (A), 200–500 m data (B), as well as the coefficient of variation for the 0–200 m model (C) and 200–500 m one (D). The color scale for the coefficient of variation has the same range for all plots.

#### 2.4.4 *In situ* imaging compared to net based sampling

The latitudinal biomass distribution of Copepoda and Rhizaria obtained by combining the predictions of global models for the epi- and mesopelagic is shown in figure 2.10. It is compared against data (interpolated on 0-500 m) from the Tara Oceans mission (Pesant et al., 2015; Siviadan et al., 2022) acquired using 300  $\mu$ m multinet samples and ZooScan (G. Gorsky et al., 2010). To make the comparison meaningful, we only selected organisms in the ZooScan samples with an ESD  $>1$  mm. For Copepoda, the values observed by the UVP5 and the nets reveal a similar latitudinal pattern between 70°N and 60°S.

The trend computed on the output of the models shows lower biomass between 40°N and 40°S compared to Tara observations. For Rhizaria, the highest biomass was found in the UVP5 observations and models around the equator. Generally, almost no Rhizaria were observed in nets whereas they were consistently observed with the UVP5.

#### 2.4.5 Global zooplankton biomass distribution

The biomass integrated over 0-500 m was predicted to be maximal at around 60°N and 55°S, with values decreasing both north and south of these two latitudes (Fig. 2.11). The lowest values of biomass were predicted in the Arctic Ocean and the Weddell Sea as well as in the oceanic gyres (especially in the southern hemisphere). We also observed an increase of the predicted biomass around the equator. The highest biomass values were predicted between 50 and 80°N, in coastal cells in the Labrador Sea and Baffin Bay, as well as in the Greenland Sea. Relatively high biomass was predicted around these points as well as in the Gulf of Alaska, Bering Sea and Sea of Okhotsk. A band of high biomass was predicted between 40 and 50°S around the Arctic polar front. Finally, by summing the predictions of models for which the predictions significantly correlate with observations, we can get to a first robust, conservative, global biomass estimate of zooplankton biomass based on UVP5 *in situ* imaging. As not all groups could be included in this computation, we refer to the following numbers as biomass *ad minima*. With that in mind, the zooplankton biomass estimated by the models was 0.229 PgC for the epipelagic, and 0.173 PgC for the mesopelagic. This sums to 0.403 PgC between 0-500 m.



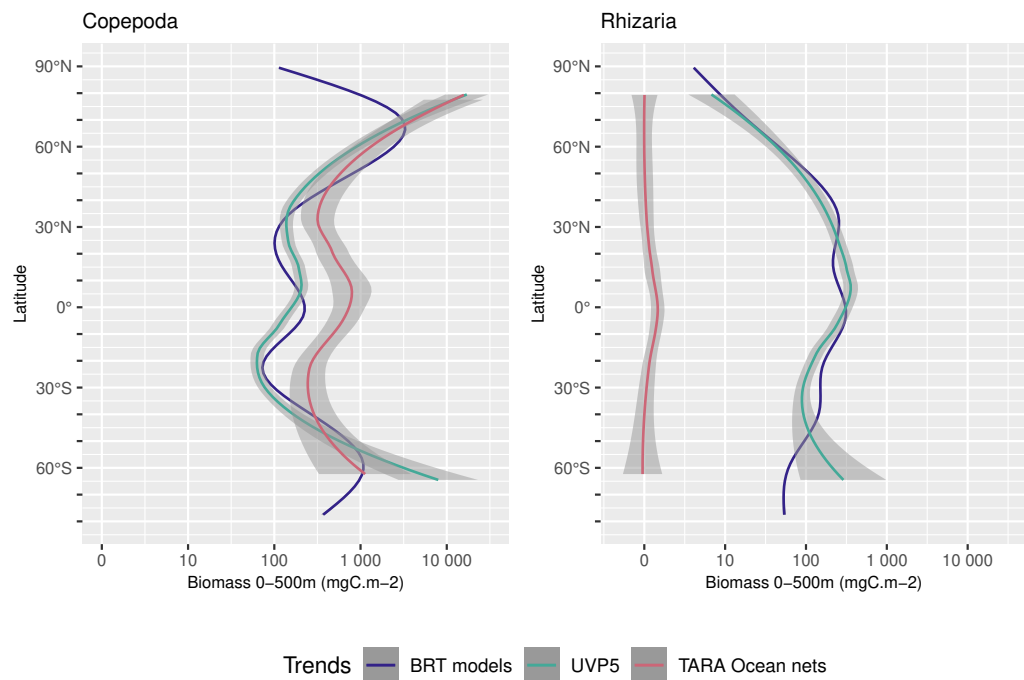


Figure 2.10: Comparison of the latitudinal distribution of biomass ( $mgC\ m^{-2}$ ) integrated over 0-500m depth between our models' estimation and the results from the Tara Ocean multinet (300  $\mu m$  mesh size), for Copepoda and Rhizaria. Trends were obtained by using Loess regression on: "BRT models" using the global model outputs for Copepoda or Rhizaria (summed across 0-200 m and 200-500 m depth) ; "UVP5" using the biomass as seen by the UVP5 between 0-500m ; "TARA Ocean net" using the sampling points between 0-500m. The shaded areas represent the 95% confidence interval of the Loess fit.

## 2.5 Discussion

### 2.5.1 Sensitivity of model prediction to partitioning

In this study, we explored whether a partitioning approach would improve model performance through the use of different horizontal and vertical divisions of our dataset. The aim of using partitioned models was to test if we could model local organism subgroups that would be mixed within the coarse taxonomic definition imposed by the dataset. The  $R^2$  computed on the models' output show a high variability across groups, layers and regional combinations. Overall, when comparing each partitioned model to the same zone in the global one, neither the global nor the partitioned approach is overall better suited. The reduction in dataset size might be the explanation why in many cases global models perform better than the smaller partitioned

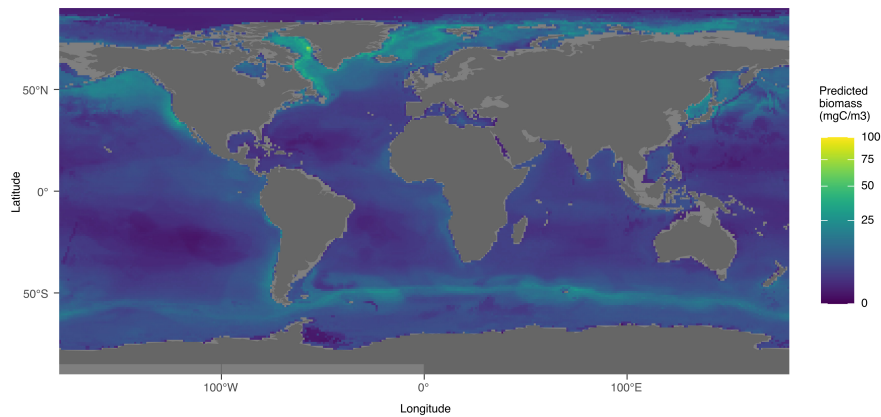


Figure 2.11: Distribution map of the minimum global biomass between 0 and 500m using taxa which obtained a  $p$ -value  $< 0.05$  in Pearson test between the predicted and the biomass calculated from UVP5 data.

models. The high latitude dataset contains 712 UVP5 profiles, the low latitude 2 837 and the world 3 549 data points. Another drawback of the partitioned models could be that some groups might have an environmental habitat associated with regions on both sides of the limits of the two models (here  $40^{\circ}\text{N}$  or  $40^{\circ}\text{S}$ ). A vertical resolution of 0-200 (epipelagic) and 200-500 m (mesopelagic) depth gave the best results compared to a finer depth separation. The reduction of data per model with a finer depth layer resolution probably made it impossible for some models to learn the association between a group's biomass distribution and the associated habitat properties, either because the model could not learn this association or because the group was considered rare ( $< 500$  images). If enough data is available, however, a finer vertical model might perform better, because it better delimits the vertical habitat structure. This seems to be the case e.g. for the Phaeodaria for which models with 100 m resolution obtained higher  $R^2$  results, especially for those between 0 and 300 m depth.

### 2.5.2 Group-wise contribution to global zooplankton biomass

Globally, in the 1.02 - 50 mm size range, we observed up to four zooplankton groups dominating each region and layer (Fig. 2.4), mainly including Crustacea (Copepoda, Eumalacostraca, other Crustacea) and Rhizaria (solitary Collodaria, Phaeodaria, Foraminifera). The dominance by copepods was expected: they are known to be a central trophic link in marine ecosystems (Steinberg and Landry, 2017) and their dominance was already shown in several studies (Dai et al., 2016; Forest et al., 2012; Turner, 2004). Rhizaria were also presented as substantial participants in the global zooplankton biomass by Biard et al. (2016)

with Phaeodaria and Collodaria being the most important contributors to rhizarian biomass. In addition, Rhizaria were previously shown to play an important role in the biological carbon pump by intercepting (Stukel et al., 2018, 2019) but also generating particle flux (Lampitt et al., 2009). In contrast, gelatinous predators such as Chaetognatha and other Cnidaria (other Cnidaria, other Hydrozoa, Siphonophorae) can be well predicted but their predicted biomass is low. This might be due to different reasons, ranging from their low carbon content (McConville et al., 2016), their size range which can exceed the specific range of the UVP5 (1.02 - 50 mm), their lower abundance reducing the probability of observation in the rather small volume of the UVP5 and the reduced capacity of the UVP5 to image them due to their transparency. Other instruments, such as the pelagic *in-situ* observation system (PELAGIOS, Hoving et al. (2019)), the Zooglider (Ohman, 2019) or the *In-Situ* Ichthyoplankton Imaging System (ISIIS, Cowen and Guigand (2008)) might be more adapted to study these organisms, thanks to their larger sampling volumes or different image approach.

### 2.5.3 Distribution patterns and occupied habitats

#### 2.5.3.1 Copepoda

Copepoda biomass was predicted to be highest in high latitudes in both epi- and mesopelagic layers of the global models. The lowest values were predicted at the gyres and an increase of biomass was observed centered at the equator. In the global models, temperature always appeared within the top three environmental factors explaining the distribution of copepods (except for 0-100 m model where it appeared 4<sup>th</sup>), which is in agreement with previous work suggesting that surface temperature and thermal tolerance of marine ectotherms, including copepods, are important constraints for their distribution and abundance (Beaugrand et al., 2009; Sunday et al., 2012). We also predict significant Copepoda biomass centered at 50°S in the Southern Ocean, at the location of the strongest horizontal gradient of temperature within the epipelagic layer. This geographic pattern is in agreement with earlier observations of high Copepoda occurrence along the Polar front (Pinkerton et al., 2020). Hence, despite a low number of UVP5 profiles in this latitudinal band, the model is able to retrieve this fundamental pattern. Higher values of the coefficient of variation (Fig. 2.5C) are found in the Arctic Ocean, as well as south of 60°S. More data from these regions could help to further reduce the uncertainty of our models.

### 2.5.3.2 Eumalacostraca

The distribution of the predicted Eumalacostraca biomass presented high values in coastal areas mainly on the eastern border of the Atlantic and Pacific Oceans and low values at high latitudes and at the locations of the oceanic gyres. Due to the low image resolution, a finer taxonomic resolution than Eumalacostraca (mostly euphausiids, decapods and amphipods) is not possible for UVP5 vignettes and especially a distinction of euphausiids and decapods is rarely possible. Euphausiids are well known for their ability to avoid capture by standard oceanographic plankton nets (Brinton, 1967; Sameoto et al., 1993; Wiebe et al., 1982) and even low noise gliders (Guihen et al., 2022). This behavior might also be dependent on the species and stage development while the UVP5 mostly detects small Eumalacostraca ( $\leq 50$  mm) for which taxonomic identification is not possible. Nevertheless, as Euphausiids are the second most abundant crustacean taxon after copepods (Castellanos et al., 2009) they may compose a large fraction of the biomass in this group. They are described as widely distributed in high numbers in the world ocean between 0-300 m with the exception of the eastern Canadian Arctic and the Arctic Ocean (Castellanos et al., 2009). This coincides with our predictions of higher biomass in the epipelagic (0.058 PgC) than the mesopelagic (0.049 PgC), and low values predicted for the Arctic Ocean. The high Eumalacostraca biomass predicted in the North Atlantic also coincides with high abundances of krill observed by (Edwards et al., 2021). *Euphausia superba* and *Euphausia mucronata* have been described as keystone species of the Antarctic and the Humboldt Current System, respectively (Antezana, 2010). The comparatively low values of biomass predicted in the Antarctic in the epipelagic layer (Fig. 2.6A) might be too low, as *Euphausia superba* is known to show a patchy distribution (Siegel, 2005, 2016). As we only have very few samples from the Antarctic Ocean, we probably under-sampled this region and specifically krill. The high coefficient of variation in this region seems to reflect this problem. Overall, we need to state that we might underestimate the abundance of Euphausiids and of Eumalacostraca in general due to avoidance of the CTD/UVP5 system, the comparatively small sampling volume of the system and the low sample abundance in the Southern Ocean.

### 2.5.3.3 Solitary Collodaria

Global models in epi- and mesopelagic layers predicted a widespread distribution of solitary Collodarians between 50°N and 50°S, from oligotrophic to eutrophic zones. Their distribution can be explained by the selective advantage of their mixotrophy, since all collodarian species live in symbiosis with photosynthetic microalgae (Biard et al., 2016; Suzuki and Not, 2015). Consistently with the models' prediction

of solitary Collodaria as the third most important group in terms of global biomass in 0-200 m, it has been shown by Biard et al. (2016) that Collodaria contribute most to the biomass of the Rhizaria between 0-100 m.

#### 2.5.3.4 Phaeodaria

The distribution of Phaeodaria shows a latitudinal pattern with three peaks, one at 50°N (with high biomass values at the level of the subarctic gyres), one at 5°N and at 60°S. These three peaks were not observed by Biard et al. (2016). The highest values being predicted in the subarctic gyre are consistent with D. K. Steinberg et al. (2008) who estimated their mean biomass there as 5.5% (range 2.7–13%) of the metazoan biomass sampled using a MOCNESS (Wiebe et al., 1985). The distribution of this group in the epipelagic (high biomass in coastal regions especially around the Californian upwelling and low biomass in gyres conditions) could be related to food availability which might not be abundant enough in the open ocean. In the models' output, this group only contributes to ~1.2% of the global biomass in the epipelagic. This is consistent with previous work describing these organisms as being distributed in water below 150-200 m (Biard and Ohman, 2020; Boltovskoy et al., 2017; L. Stemann et al., 2008; Suzuki and Not, 2015). The high ( $R^2 = 0.50$ ) and low latitude ( $R^2 = 0.39$ ) models done on the mesopelagic layer reveals similar patterns as the ones shown for the epipelagic layer in Figure 2.8. This pattern of high biomass predicted in the North Pacific can be put in perspective with a previous study (Ikenoue et al., 2019) which highlighted Phaeodaria in the Western North Pacific as one of the major carriers of carbon in the twilight zone, defined by K. O. Buesseler and Boyd (2009) between 200-1000 m depth, with an organic carbon standing stock reaching its highest value at depths between 200-500 m. A maximum in abundance of Phaeodaria was observed in the lower epipelagic or mesopelagic zone in the Sea of Japan by Nakamura et al. (2013) as well as in the Antarctic beneath the sea ice with similar abundances as the North Atlantic and Pacific (Morley and Stepien, 1984). In the regional mesopelagic predictions, the mean biomass in the Sea of Japan is not particularly high, but it reached higher values in the Southern Ocean.

#### 2.5.3.5 Acantharea

Here, we present results on large Acantharea only, but it should be kept in mind that most species are smaller than 600  $\mu\text{m}$  (Biard et al., 2016). Most Acantharea species are associated with symbiotic algae (Michaels, 1991) which could explain the rapid observed biomass decline with depth. Indeed, the biomass predicted is 16.7-times lower in the mesopelagic ( $1.36 \cdot 10^{-5}$  PgC) compared to the epipelagic layer

( $2.27 \cdot 10^{-4}$  PgC). These mixotrophs are present throughout the world oceans (Suzuki and Not, 2015) and commonly distributed in intertropical latitudes (Bottazzi and Andreoli, 1982) mostly in the surface with an abundance rapidly declining below 20-50 m depth (Michaels, 1988). The model confirmed this biomass diminution in the epi- and mesopelagic layers (Fig. 2.9). We also observed latitudinal distributions which were maximal in intertropical areas consistent with these previous studies. The highest biomass of Acantharea predicted in the mesopelagic global model in the Gulf of Alaska coincides with a large number of organisms imaged by the UVP5. This is surprising knowing the above described distribution patterns. More observations from this region are required to clarify whether this was a temporally limited occurrence or whether this represents a region of permanent abundance maxima. The predicted biomass in the Antarctic waters in this depth layer is also surprising. Acantharea are marine planktonic unicellular eukaryotes in the Rhizaria group and produce a mineral skeleton made of strontium sulfate (Decelle and Not, 2015; Michaels, 1991). The surprisingly high abundance at high latitudes might be important for studies done on the strontium biogeochemical cycle (Bernstein et al., 1987; Decelle et al., 2013).

#### 2.5.4 Comparison between net sampling and *in-situ* imaging

The integrated global predicted biomass is dominated by Copepoda (35.7%), Eumalacostraca (26.6%) and Rhizaria (16.4%). Because of their important contribution to the predicted global biomass, the distribution map of total biomass *ad minima* (Fig. 2.11) reflects in part the major distribution patterns of these three groups: polar waters are dominated by Copepoda and intertropical waters are dominated by mixotrophic Rhizaria. Eumalacostraca follows the predicted distribution of zooplankton with 3 peaks of biomass at 60°N (55°N for zooplankton), at the equator and at 45°S (55°S for zooplankton). The comparison of the models' output with data from the Tara Ocean expedition, obtained with a 300  $\mu$ m mesh size multinet (Pesant et al., 2015; Soviadan et al., 2022) shows a good agreement for the latitudinal patterns of Copepod biomass. Net data is estimated to be higher than biomass estimated from UVP5 data in the intertropical latitude range for this group. Results in the high latitudes regions with strong seasonality and sea ice cover should be taken with care as no data was available in the UVP5 dataset in winter for these latitudes. For Rhizaria, we observe that at most locations the biomass estimated by the nets is zero, whereas the UVP5 yields considerable biomass in this group (Fig. 2.10). In the TARA Ocean multinet samples, only Acantharea, Foraminifera and Phaeodaria are sometimes detected, while Collodaria are consistently absent from these samples. Indeed, Collodaria and

Study	Size range	Depth	Global estimates
Moriarty et al., 2012	$\geq 2$ mm	0-350 m	0.02 PgC
Moriarty and O'Brien, 2013	$\geq 200 \mu m$	0-200	0.19 PgC
Buitenhuis et al., 2013	$\geq 200 \mu m$	Integrated	0.33-0.59 PgC
Buitenhuis et al., 2013	$\geq 2$ mm	0-500 m	0.22-1.52 PgC
Hatton et al., 2021	$\geq 200 \mu m$	0-200 m	0.53-31.57 PgC
Hatton et al., 2021	$\geq 2$ mm	0-200 m	0.02-2.64 PgC
This study	$\geq 765 \mu m$ -37.5 mm	0-200 m	0.229 PgC

Table 2.2: Comparison of global biomass estimates in the literature. Please note that we have converted the size range we cover with the UVP5 (1.02-50 mm ESD) to meshsize using the empirical Nichols and Thompson, 1991's 3/4 law of mesh selection.

Acantharea are poorly sampled by nets and are not well preserved in plankton samples fixed with regular fixatives such as formaldehyde (Suzuki and Not, 2015). Yet, solitary Collodaria are predicted as the 3<sup>rd</sup> most important group in terms of biomass in the upper 200 m of the global model. We here show that *in situ* imaging is far more suitable for the study of these and all other fragile plankton groups. As described above, several important zooplankton groups are generally well modeled, allowing us to combine the taxon-specific models to yield a global estimate of zooplankton biomass in the 1.02 to 50 mm size range. Previous studies (Table 2.2) have computed such global zooplankton biomass obtained largely (Hatton et al., 2021) or completely (Buitenhuis et al., 2013; Moriarty and O'Brien, 2013; Moriarty et al., 2012) from net collected organisms. These studies also used a proportionality method by estimating the global biomass presented in Table 2.2 by multiplying the median value of biomass with the surface of the ocean and the studied depth. Our predictions are within the same order of magnitude — but at the lower limit — of these compilations if one combines their meso- and macrozooplankton biomass estimates. We refrain from a more detailed comparison due to the difference in size studied (here 1.02 - 50 mm ESD — equivalent to 765  $\mu m$  to 37.5 mm meshsize according to Nichols and Thompson (1991)'s 3/4 law of mesh selection — compared to  $\geq 200 \mu m$  for the cited meso- and macrozooplankton studies), sampling methods and depth covered (Buitenhuis et al., 2013). Contrary to the complementary use of nets and Zooscan, such as with the TARA dataset, these previous studies are based on data obtained through methods which do not allow to split the organisms based on fixed criteria (size, area of the organism or taxonomy). One would expect a large contribution to biomass in the 200 to 765  $\mu m$  mesh size range (Gallienne, 2001; Hwang et al., 2007).

### 2.5.5 Global zooplankton biomass distribution

The distribution of the global integrated biomass (0-500 m) *ad minima* follows the patterns described by Ikeda (1985), Moriarty et al. (2012) and Hatton et al. (2021) which correspond to a latitudinal distribution of the biomass with high values north of 55°N and south of 55°S, low values in between increase around the equator between 15°N and 15°S. The benefit of our work and of compiled datasets such as the ones used in Moriarty et al. (2012), Moriarty and O'Brien (2013), Buitenhuis et al. (2013) and Hatton et al. (2021) is that they bring together numerous single transects and allow to have an integrated view of global zooplankton distribution. The results depicted in Figure 2.11 in the Southern Ocean are consistent with a recent study done with BRTs (Pinkerton et al., 2020) showing that the highest environmental suitability for zooplankton was located between the Subantarctic Front and the southern limit of the Antarctic Circumpolar Current with a lower suitability north and south of this band. The spatial distribution of plankton biomass thus shows the importance of oceanographic hydrodynamics leading to oligotrophy in central gyres and mesotrophy in areas of high latitudes and equatorial and coastal upwellings. Zooplankton plays a crucial role in fisheries in these environments e.g. in the Humboldt Current System which harbors the largest fishery in the world and most economically important fish species, supported by the upwelling of Peru (Chavez et al., 2008). Peruvian anchovies and sardines obtain most of their energy from zooplankton (van der Lingen et al., 2009).

### 2.5.6 Conclusions and outlook

In summary, our results show, for the first time, that spatial patterns and global biomass of key zooplankton groups can be calculated using a machine learning method (BRT) to extrapolate individual zooplankton biomass estimates from sparse UVP5 observation. They also highlight the important contribution of Rhizaria (predicted mainly in the intertropical range) and Copepoda (predicted mainly in high latitudes) to the global estimate of zooplankton biomass. Within the size range covered, Copepoda contributes 35.7%, Eumalacostraca 26.6% and Rhizaria 16.4% to global zooplankton biomass. This suggests that it is especially crucial to extend work on the fragile Rhizaria, which are comparatively little studied. As a biogeographical study, our aim was not to represent proximal mechanisms that drive the distribution of zooplankton, or to describe seasonal or transient (e.g. mesoscale) features, but rather to represent the global distribution patterns of biomass according to general properties of the water masses. This method worked well in general as seen in Figure 3 for at least 3 of the combinations of regions



and depths. It made it possible to model 19 groups of zooplankton and obtain corresponding maps with the relative importance of the environmental variables used for the model. The WOA climatologies used in this study compile data of salinity and temperature (2005-2017) and other variables (1955-2017). The temporal coverage of the latter being much rarer, we hope to use more constrained nutrient dataset for our future work once a narrower time frame will be available.

The zooplankton biomass predictions based on UVP5 datasets presented here are important for global biogeochemical modeling of pelagic ecosystems because they usually lack zooplankton observations to constrain their development (Buitenhuis et al., 2013; Séférian et al., 2020; Stemmann and Boss, 2012). A current trend is to add more realism in plankton representation to better predict future ecosystem states and ocean conditions to inform sustainable management strategies for climate mitigation at global scale (Séférian et al., 2020). The UVP5, the newly developed UVP6 (Picheral et al., 2021) and other commercialized *in situ* systems, provided that they are inter-calibrated (Lombard et al., 2019), will continue to be used in the foreseeable future increasing data availability. Still, the bottleneck lies in the classification of the massive amount of images which still require human validation, but new algorithms to recognise plankton types and traits are expected (Irisson et al., 2022). The further anticipated expansion of image datasets will enable the quantitative assessment of rare groups that were not well predicted here. In addition, the deployment of the UVP6 on autonomous platforms will also help to sample certain areas with very specific environmental characteristics, difficult to access during a period of the year such as polar regions in winter. The large dataset used in this study was acquired in 10 years which can be compared to the COPEPOD database collected since about 1960. The possibilities given by imaging systems could hence help to reach a useful amount of data in a much smaller time frame. It would be interesting to use other imaging system's datasets such as the ones presented by Lombard et al. (2019) to reconstruct the wider size spectrum of these groups in terms of biomass. To have a better understanding of the vertical habitat of zooplanktonic groups, we highly recommend that UVP5 and 6 profiles should be done to at least 1000 m when the bathymetry allows it. Long term inter annual data acquisition is also highly recommended. This will enable us to monitor global zooplankton biomass changes at pace with the speed of global change.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author Contributions

LD, RK, LS, JOI and TP developed the study's concept; RK, LS, LD, FC, FL, MB, TB, LC, LG, HH, LKB, AMD, MP, AR, AW, contributed to data acquisition; RK, LS, LD, JOI, TP contributed substantially to the data analysis; LD led the code development with major assistance by JO and guidance by TP and RK; LD created all figures and drafted the manuscript; All authors contributed substantially to drafting the manuscript; All authors approved the final submitted manuscript.

## Funding

LD, RK and LS received support by the European Union project TRIATLAS (European Union Horizon 2020 program, grant agreement 817578) and the Sorbonne Université through the Ecole doctorale 129. LS was supported in the initial phase of the development by the CNRS/-Sorbonne University Chair VISION. R.K. furthermore acknowledges support via a Make Our Planet Great Again grant from the French National Research Agency (ANR) within the Programme d'Investissements d'Avenir (reference ANR-19-MPGA-0012).

## Acknowledgments

The sampling effort on which this paper is based was enabled by the dedicated cruise leaders and participants who helped in the creation of the UVP5 dataset. We thank them as well as all the people who have participated in the classification of the huge amount of UVP5 images. Furthermore, we would like to thank our colleagues from the Laboratoire d'Océanographie de Villefranche (LOV) as well as Marina Levy and Fabio Benedetti for their support and precious guidance. We also thank Yawouvi Dodji Soviadan for his help with the Tara Oceans mission data.

## Data Availability Statement

The inputs and outputs of the world models for 0-200 m and 200-500 m were uploaded to the github repository: [https://github.com/dlaetitia/Global\\_zooplankton\\_biomass\\_distribution](https://github.com/dlaetitia/Global_zooplankton_biomass_distribution)

The dataset of environmental data from World Ocean Atlas is available at <https://www.ncei.noaa.gov/products/world-ocean-atlas>.

The surface chlorophyll *a* data is available on the Copernicus website at <https://resources.marine.copernicus.eu/product-detail/OCEANCOLOUR/T1/L40982/>

## Code Availability

The code used for the models and the post treatment of their outputs was made available on Github: [https://github.com/dlaetitia/Global\\_zooplankton\\_biomass\\_distribution](https://github.com/dlaetitia/Global_zooplankton_biomass_distribution)

## 2.6 Supplementary Material

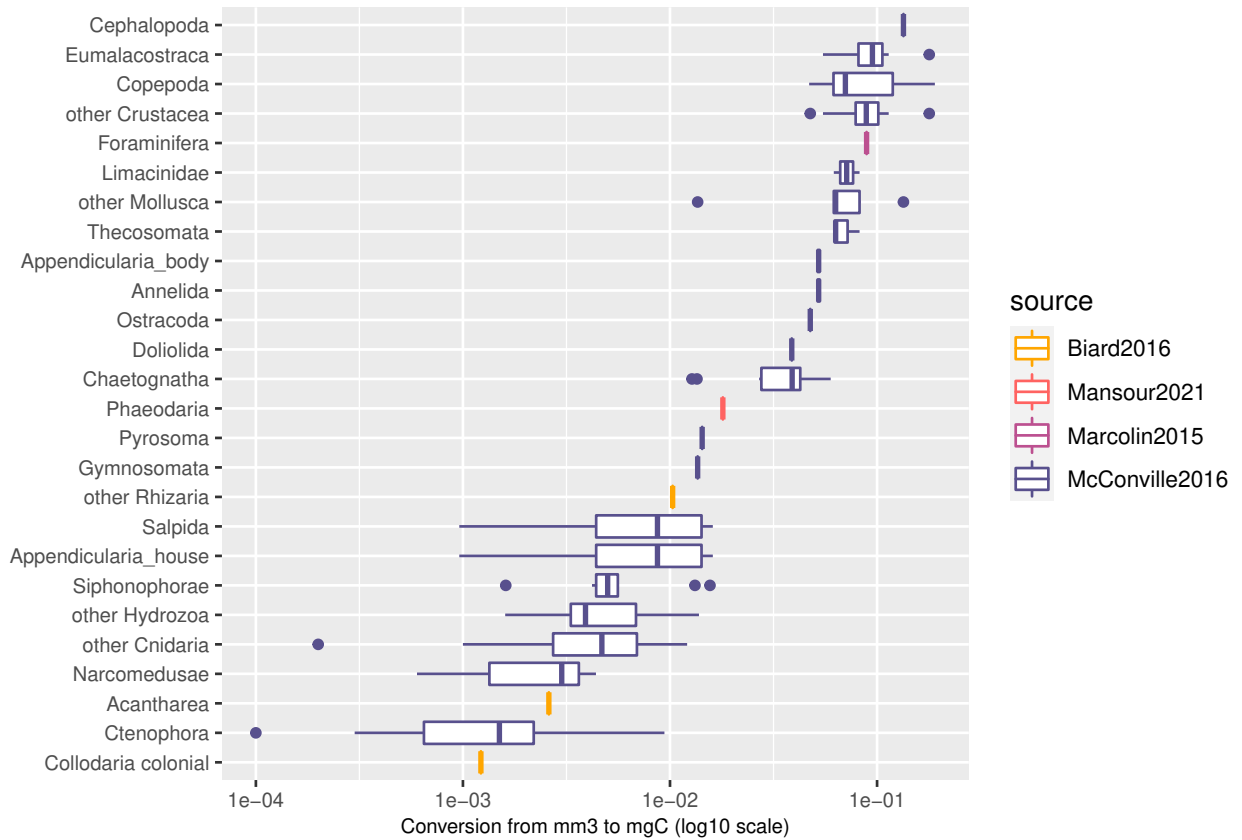


Figure S2.1: Distribution of the conversion factors from biovolume ( $mm^3$  to biomass  $mgC$  in logarithmic scale) for the studied taxa according to their source in the literature. For Rhizaria, biovolume ( $mm^3$ ) to carbon ( $mgC$ ) conversions were done using factors from Biard et al., 2016, Mansour et al., 2021 and Marcolin et al., 2015. For other groups, the conversion from individual volume to individual wet weight assumed a density of  $1gcm^3$  (Kjørboe, 2013). Then the conversion from individual wet weight to individual biomass in carbon units ( $mgC$ ) was calculated using taxon-specific linear conversion factors from McConville et al. (2016) ; when several factors were available for a taxon, their median was used for each group). For solitary collodarians, the estimation of carbon ( $0.189 mgCmm^3$ ) by Mansour et al., 2021 applied as explained in the subsection 2.1.3

0–200m

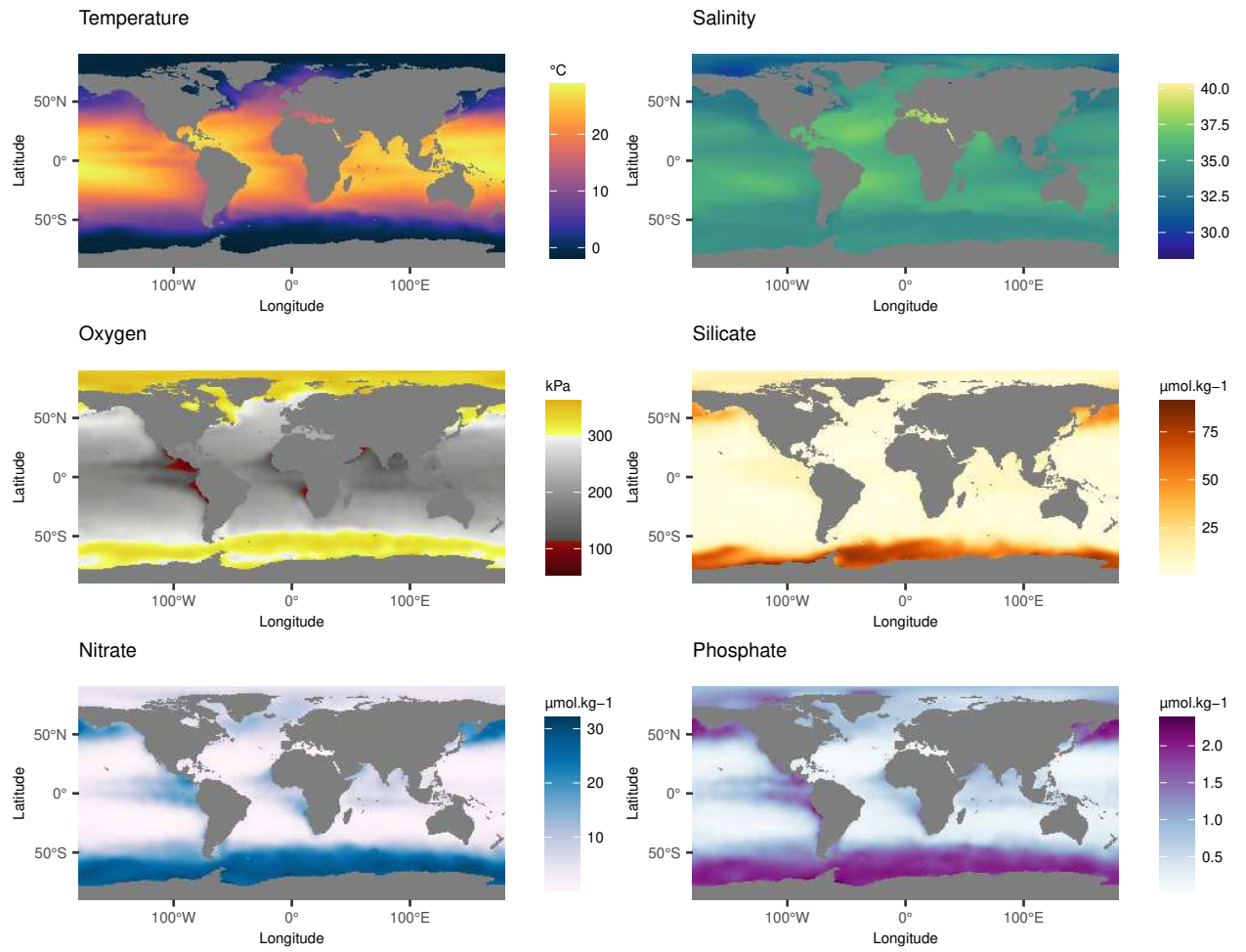


Figure S2.2: Distribution maps of the environmental variables used in the model in the layer 0-200 m

200–500m

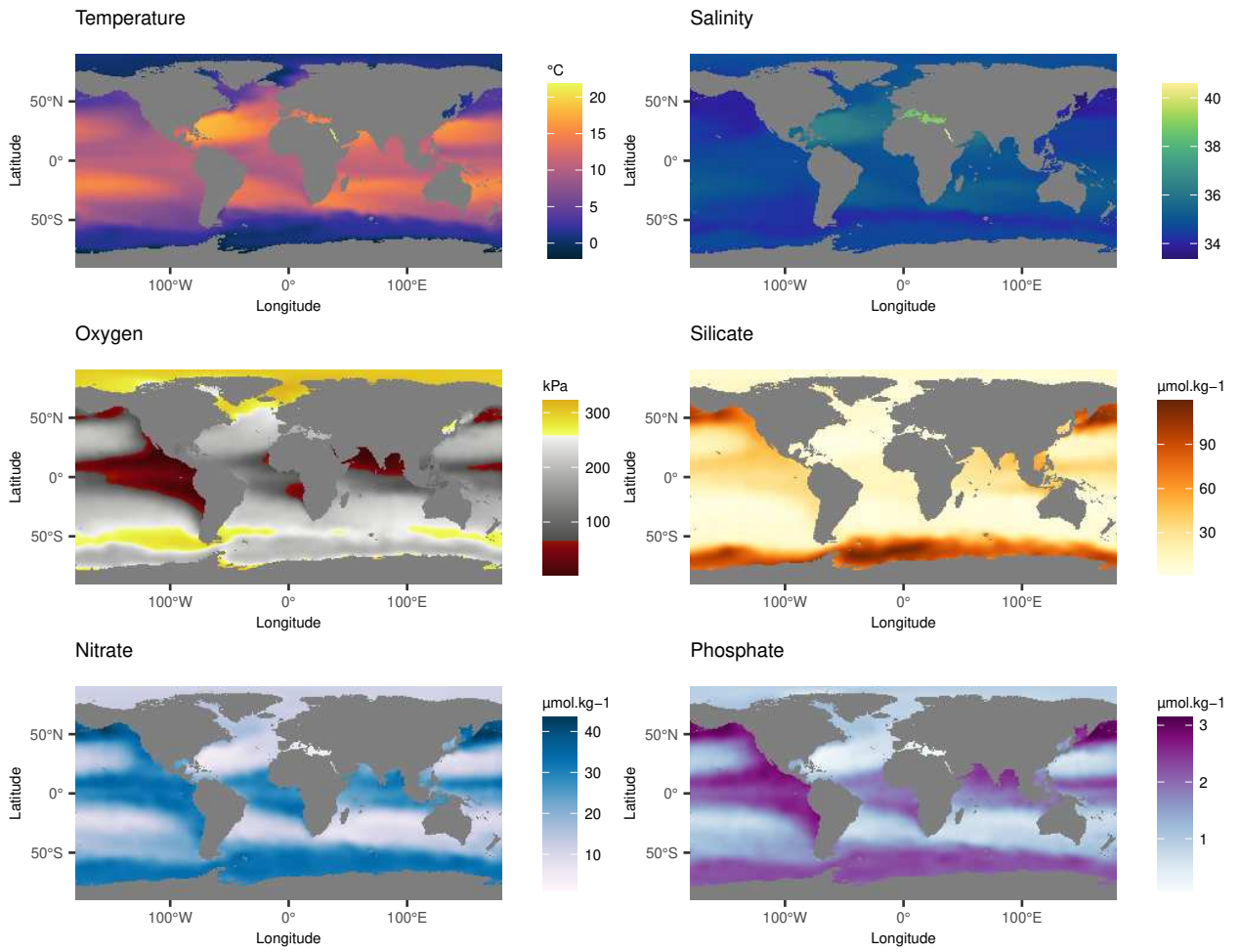


Figure S2.3: Distribution maps of the environmental variables used in the model in the layer 200-500 m

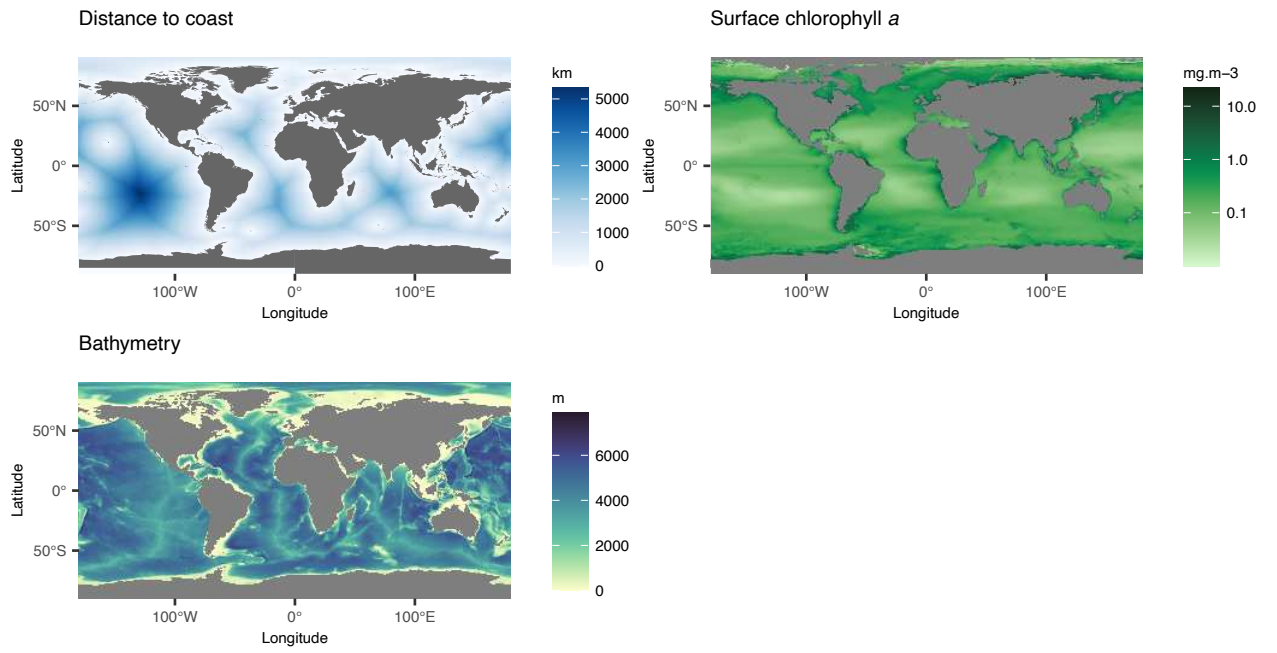


Figure S2.4: Distribution maps of the environmental variables used in all models

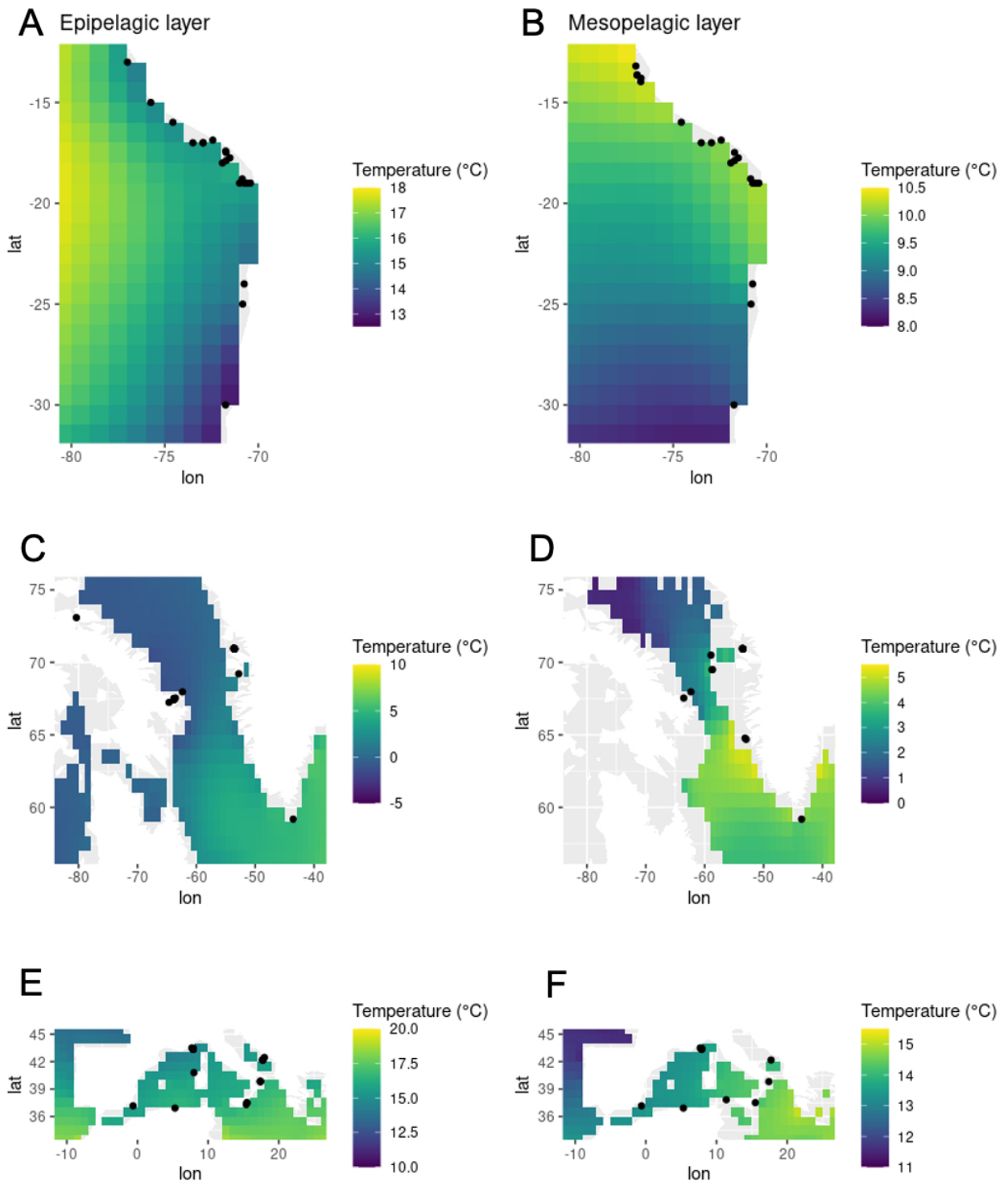


Figure S2.5: Distribution of UVP5 sampling points (symbolized by black dots) for which a match was found in the neighboring cell in the regions of the western coast of South America (A,B), in the Baffin Bay and Labrador Sea (C,D) and in the Mediterranean Sea (E,F). The colored cells represent the temperature (in °C) from the WOA dataset. In the epipelagic layer, 142 points have a neighboring match out of which 130 points (91%) are represented here (The map A contains 20 points, C contains 51 points and E contains 59 points). In the mesopelagic layer, 104 points have a neighboring match out of which 81 points (78%) are represented here (The map B contains 22 points, D contains 24 points and F contains 35 points).



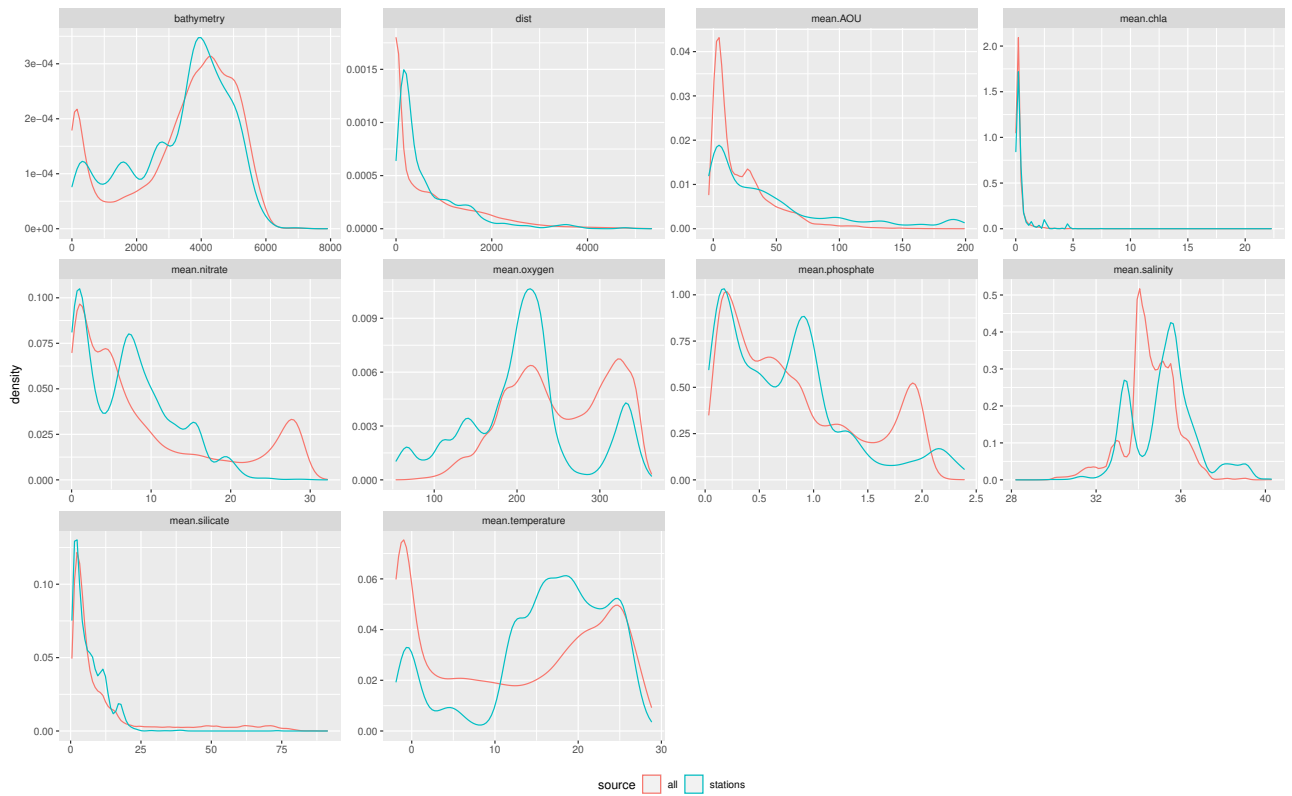


Figure S2.6: Distribution of sampling of environmental variables by UVP5 in red compared to global data from World Ocean Atlas in blue in the layer 0-200 m.

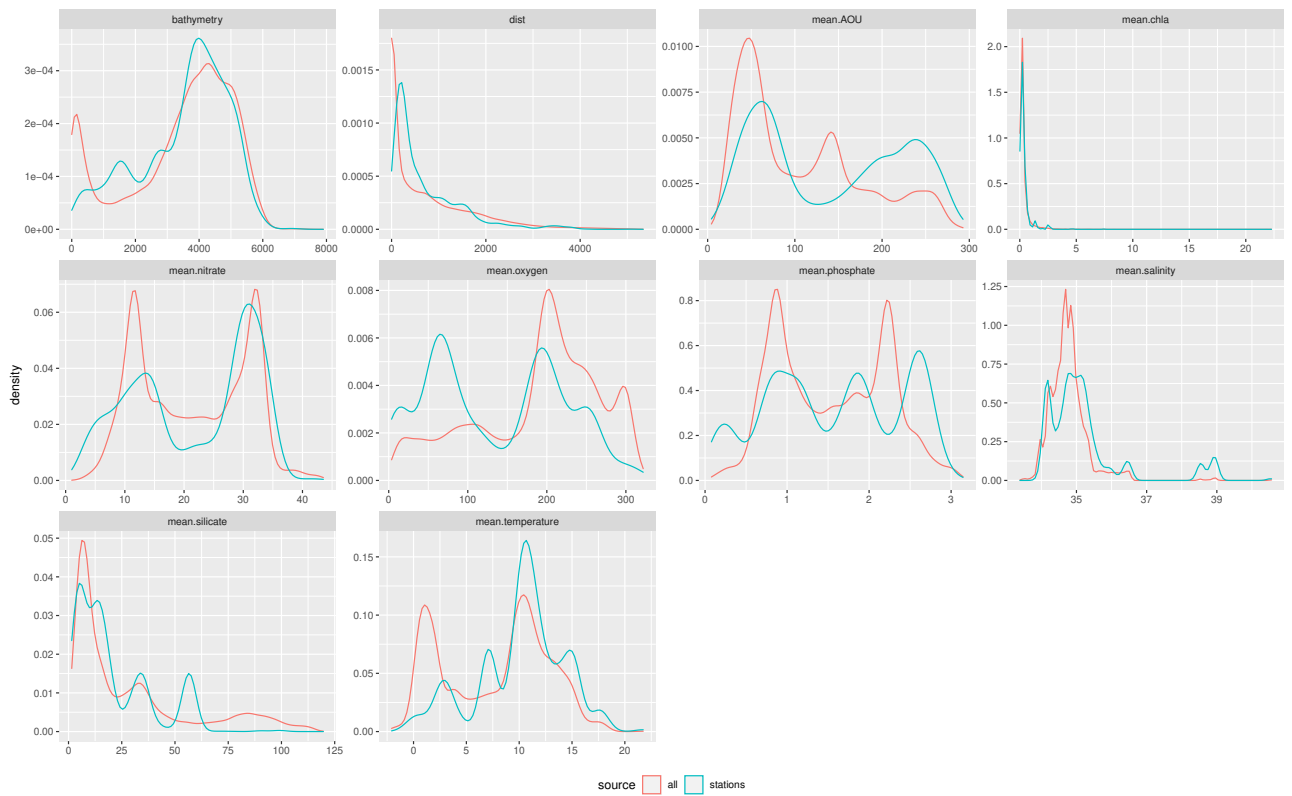


Figure S2.7: Distribution of sampling of environmental variables by UVP5 in red compared to global data from World Ocean Atlas in blue in the layer 200-500 m.

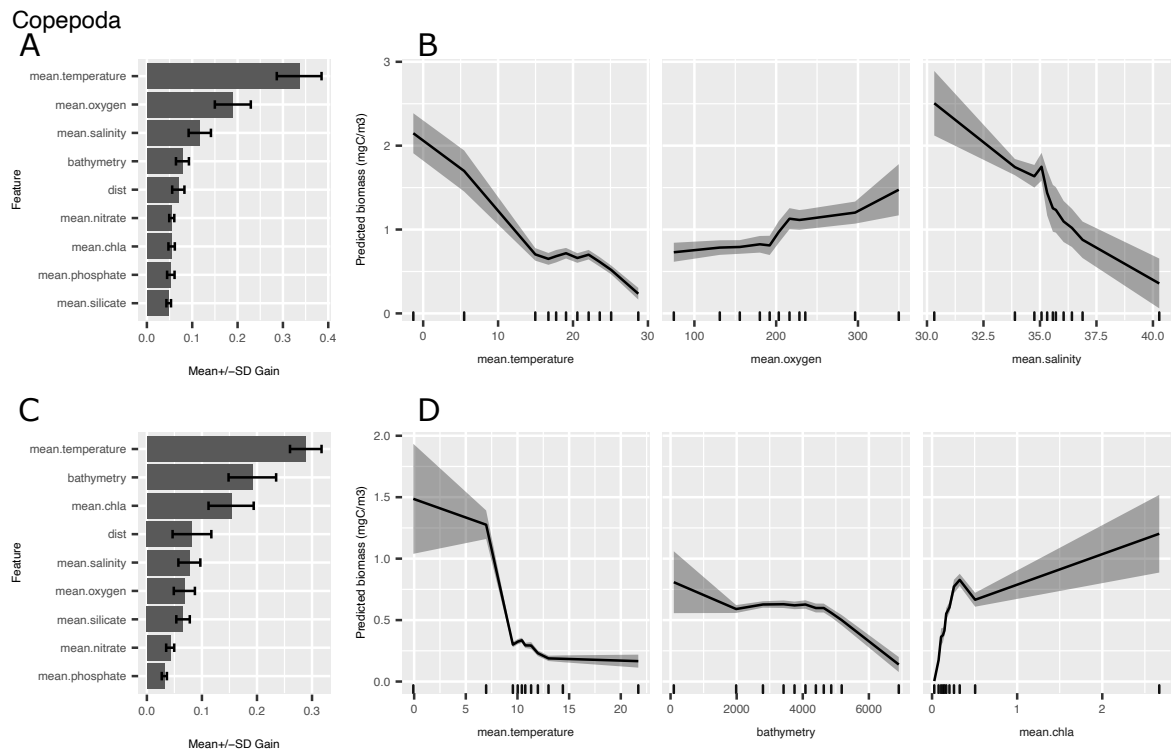


Figure S2.8: Distribution of the order of importance of variables in the model for Copepoda between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks.

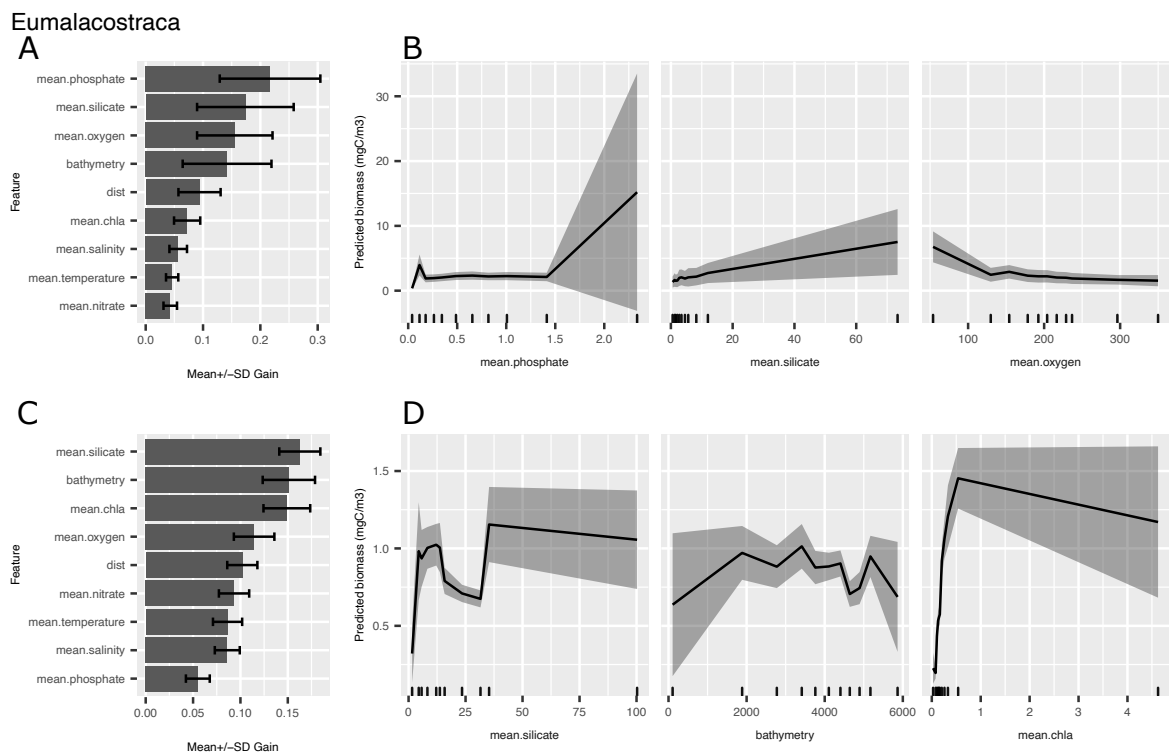


Figure S2.9: Distribution of the order of importance of variables in the model for Eumalacostraca between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. For this group the model for the 200-500 m layer does not yield a significant correlation between model and data and results are therefore not shown.

Solitary Collodaria

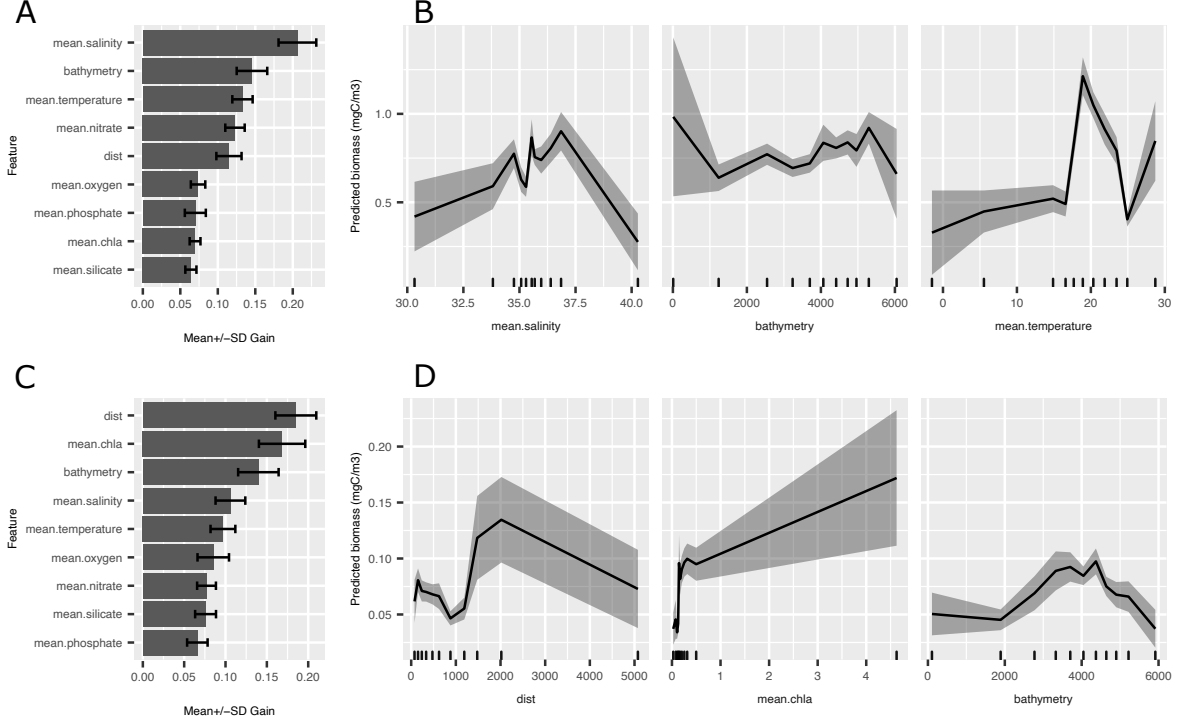


Figure S2.10: Distribution of the order of importance of variables in the model for solitary Collodaria between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks. For this group the model for the 200-500 m layer does not yield a significant correlation between model and data and results are therefore not shown.

Phaeodaria

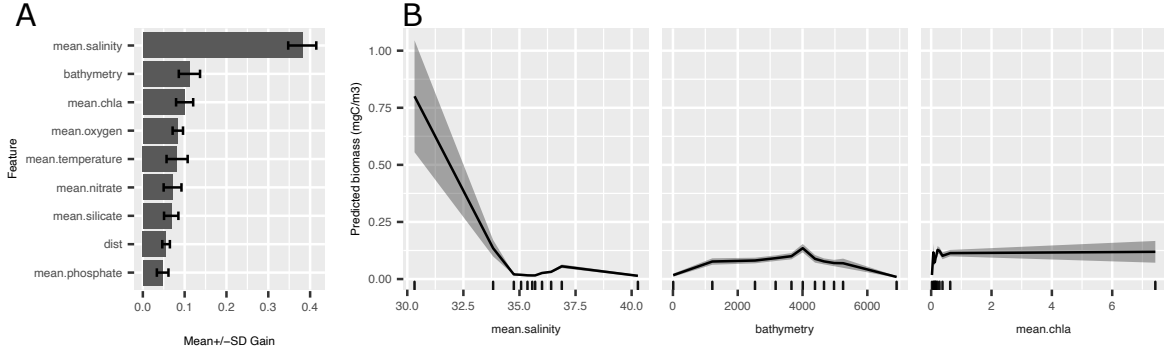


Figure S2.11: Distribution of the order of importance of variables in the model for Phaeodaria between 0-200 m (A). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks.

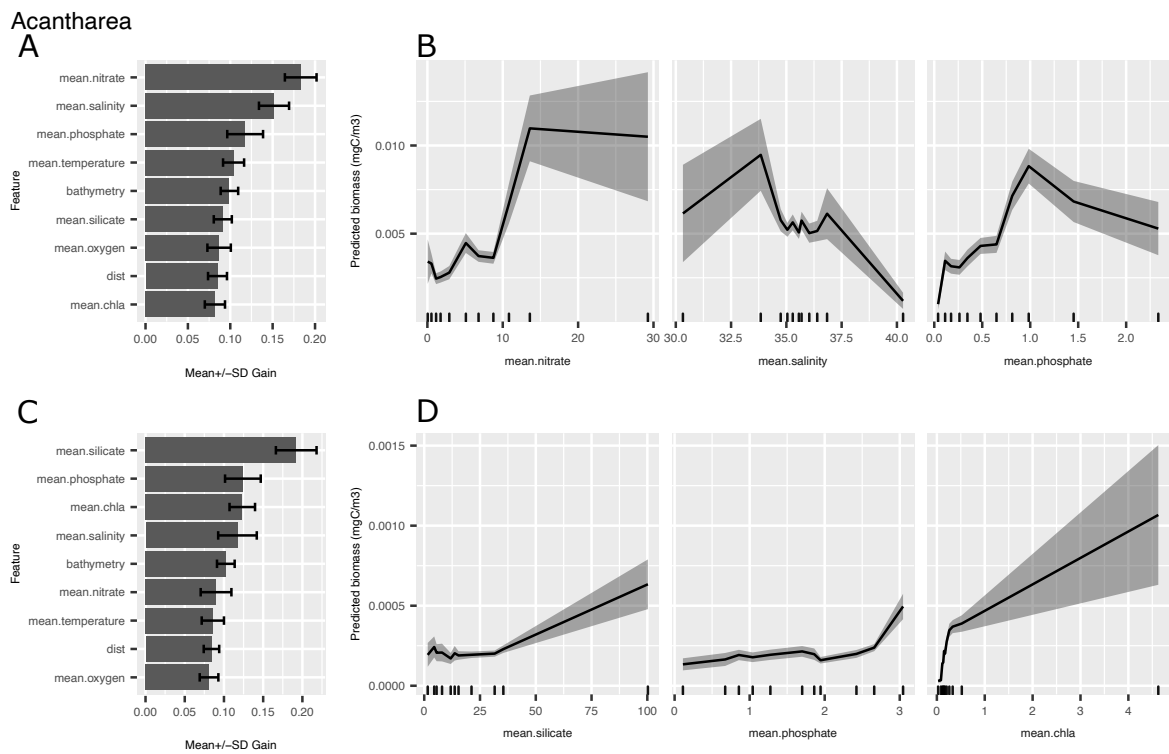


Figure S2.12: Distribution of the order of importance of variables in the model for Acantharea between 0-200 m (A) and 200-500 m (C). Partial dependence plots of the 3 most important variables in the model for 0-200 m (B) and 200-500 m (D). The ticks on the x axis inform on the probability of the predicted data. There is 10% of the prediction between 2 ticks.

Table S2.1: List of groups used for the analysis

Groups	Organisms classified as such
Acantharea	Acantharea
Annelida	Annelida
Appendicularia	Body or house of Appendicularia
Cephalopoda	Cephalopoda
Chaetognatha	Chaetognatha
other Cnidaria	Cnidaria with the exception of Hydrozoa
Colonial collodaria	Colonial collodaria
Copepoda	Copepoda or Copepoda_like
other Crustacea	Crustacea with the exception of Copepoda, Ostracoda and Eumalacostraca
Ctenophora	Ctenophora
Doliolida	Doliolida
Eumalacostraca	Eumalacostraca
Foraminifera	Foraminifera
Gymnosomata	Gymnosomata
other Hydrozoa	Hydrozoa with the exception of Narcomedusae and Siphonophorae
Limacinidae	Limacinidae
other Mollusca	Mollusca with the exception of Gymnosomata, Limacinidae, Cavolina and Creseis
Narcomedusae	Narcomedusae
Ostracoda	Ostracoda
Phaeodaria	Phaeodaria
Pyrosoma	Pyrosoma
other Rhizaria	Rhizaria with the exception of Acantharea, Collodaria, Foraminifera and Phaeodaria
Salpida	Salpida
Siphonophorae	Siphonophorae
Solitary Collodaria	Collodaria classified as Collodaria, solitary collodarians with dark central capsule, solitary collodarians with a fuzzy central capsule, solitary collodarians with a grey central capsule, solitary collodarians with a globule-like appearance
Thecosomata	Cavolina or Creseis

## Part III

### North Atlantic biological pump dynamics in an anticyclonic eddy revealed via in situ imaging

Laetitia Drago<sup>1</sup>, Ken Busseler<sup>2</sup>, Lee Karp-Boss<sup>3</sup>, Andrew M. P. McDonnell<sup>4</sup>, Simon-Martin Schröder<sup>5</sup>, David A. Siegel<sup>6</sup>, Lars Stemmann<sup>1†</sup> and Rainer Kiko<sup>1,7†</sup>

<sup>1</sup> Sorbonne Université, Laboratoire d'Océanographie de Villefranche-sur-mer, Villefranche-sur-mer, France,

<sup>2</sup> Woods Hole Oceanographic Institution, Woods Hole, MA, USA.

<sup>3</sup> School of Marine Sciences, University of Maine, Orono, ME, United States,

<sup>4</sup> Oceanography Department, University of Alaska Fairbanks, Fairbanks, AK, United States,

<sup>5</sup> MIP, Computer Science, Kiel University, Kiel, Germany

<sup>6</sup> Earth Research Institute, University of California, Santa Barbara, California, USA

<sup>7</sup> GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstr. 1-3, 24148 Kiel, Germany

† These authors share last authorship





## Follow the plume

---

### 3.1 Abstract

Understanding the dynamics of particle flux and morphology in oceanic biogeochemical cycles is critical for unravelling the complexity of marine ecosystems and the global carbon cycle. In this study, we used a Lagrangian plume approach by deploying three Underwater Vision Profiler 5 (UVP5) units in a North Atlantic eddy during the EXPORTS 2021 experiment. Over 26 days, we tracked the plume of sinking detritus during a large export event.

Our investigation revealed major changes in the particle community within the plume. A power-based decrease in the abundance of MiPs (small particles, 0.14-0.53 mm diameter) and MaPs (large particles, 0.53-16.88 mm diameter) suggested degradation, aggregation, and potential microbial remineralisation contributions. We also investigated the slope dynamics of the particle size distribution (PSD), which revealed a two-step pattern with a flattening of the slope below the surface core waters (SCW). Our approach reliably estimated the carbon export flux compared to published estimates of the North Atlantic spring bloom.

Characterisation of the morphology of aggregates and fecal pellets within the plume using *in situ* imaging and semi-supervised clustering showed that aggregates followed an almost linear evolution, becoming darker, more circular and denser with depth. Conversely, fecal pellets exhibited distinct dynamics, shaped by their production by zooplankton and degradation with depth. Our results challenge previous assumptions and reveal an increase in size with depth for some aggregates and fecal pellets, likely caused by zooplankton compaction and differential settlement.

To understand flux attenuation, we investigated the carbon-specific degradation rate ( $C_{\text{spec}}$ ) within the plume. Building on previous research, our analysis suggests that microbial remineralisation activity was dominant in the mid to lower mesopelagic region (300-800 m). However, this activity did not fully explain the attenuation observed in the upper mesopelagic zone (50-300m), suggesting that zooplanktonic organisms may have contributed to the remainder of this attenuation in this layer.

This study provides new perspectives on particle dynamics, sinking particle morphology, flux and attenuation within a North Atlantic eddy. The results emphasise the need to consider morphological features in

future flux calculations and advocate for a deeper understanding of particle complexity in marine ecosystems. These findings open new avenues for research into particle dynamics and biogeochemical processes in the ocean.

## 3.2 Introduction

### 3.2.1 The rain of detritus

Marine detritus particles have long captivated scientists (Silver, 2015). They comprise aggregates, fecal pellets and dead bodies. Aggregates are found throughout the world's oceans and designate macroscopic clusters ( $>500\ \mu\text{m}$ , Alldredge and Silver (1988)) comprising aggregated phytoplankton, abandoned larvacean houses, fecal matter, and other detrital material (Alldredge et al., 1990; Alldredge and Silver, 1988). Detrital particles play a significant role in the biological carbon pump. As they sink, they contribute to the biological gravitational pump, which involves the settling of particles (Alldredge and Gotschalk, 1988) from the ocean's surface (Boyd et al., 2019). In conjunction with the diurnal vertical migration-based pump (Steinberg and Landry, 2017), the gravitational pump represents one of the two fundamental mechanisms that drive the carbon storage in the global ocean (Boyd et al., 2019; Nowicki et al., 2022).

Particles present a variety of shapes and sizes resulting from multiple competing physical and biological processes leading to aggregation and fragmentation (Alldredge, 2001; Alldredge and Gotschalk, 1988). As a result, marine aggregates can take various forms during their sinking journeys such as spheroids, compact spheres, strands or comet-shaped (Alldredge, 1998; Alldredge and Silver, 1988). The structure of newly-created particles is determined by a variety of characteristics, most of which depend on the concentration and nature of the phytoplankton that compose them.

Until recently, observing the change in aggregate morphologies was not possible *in situ* over a large depth range, necessitating reliance on ecological insights stemming from *in situ* collection and laboratory experiments.

### 3.2.2 An ever-changing morphological alteration

Marine aggregates can be affected by numerous processes (Fig. 3.1) which can be classified as either physical or biological acting either towards particle formation and growth or towards particle degradation and fragmentation. In this context, the main biological actors of

aggregate morphological alteration are zooplankton (micro- and macro-zooplankton) and bacteria.

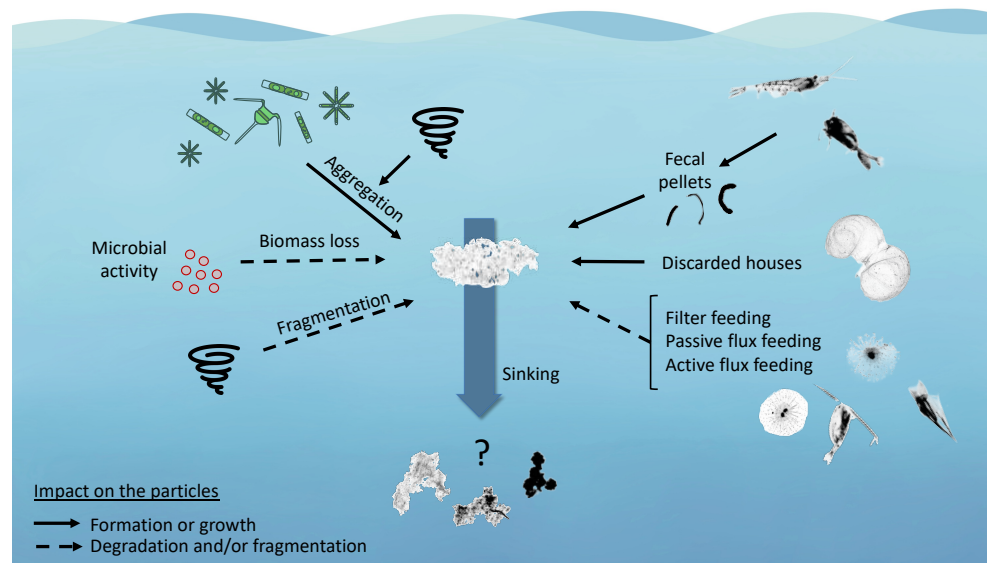


Figure 3.1: Main processes potentially affecting marine detrital particle morphology. Solid arrows present a process resulting in particle formation or growth while dashed arrows correspond to processes resulting in degradation and/or fragmentation.

### 3.2.2.1 Particle formation and growth

Physical processes can induce aggregation, or coagulation, in which two particles combine to form a larger one (Stemmann, Jackson, and Ianson, 2004). Brownian motion, shear, differential settlement, surface coagulation, and diffusive capture can promote this process (Alldredge, 2001; Kiørboe, 2001). If aggregation occurs in the presence of minerals that hinder larger particle formation, resulting aggregates may exhibit smaller, spherical, and compact characteristics (Laurenceau-Cornec et al., 2015).

Alongside physical processes, biology significantly influences aggregate morphology. Zooplankton's feeding strategies impact the nature and abundance of aggregates within a specific space and time. They package marine snow material, producing fast-sinking compact fecal pellets (Atkinson et al., 2012; Turner, 2002, 2015). Fecal pellet characteristics like shape, density and size (Durkin et al., 2021) vary with the nature and concentration of food sources (Dagg and Walser, 1986; Feinberg and Dam, 1998). Moreover, producer species and individual size affect fecal pellet attributes (Atkinson et al., 2012): salps generate large flakes, euphausiids form long, thin cylindrical ones (Atkinson

et al., 2012; Gleiber et al., 2012), and copepods produce ellipsoid or ovoid pellets (Gleiber et al., 2012; Koster et al., 2011). Fecal pellets are highly attenuated in the upper water column (Alldredge et al., 1987; Durkin et al., 2021) and can be integrated into aggregates through differential settlement.

Zooplankton impact aggregation through filter feeding, as seen in appendicularians using houses (Alldredge et al., 1990; Lombard and Kiørboe, 2010; Robison et al., 2005), salps (Steinberg et al., 2023) and crustaceans (Huntley and Boyd, 1984). This strategy lowers particle concentration, and filtration rates depend on particle concentration, saturating at high concentrations (Harbison et al., 1986; Lombard et al., 2009).

Small copepods (Toullec et al., 2019) and microorganisms (Alldredge, 2001) can foster the aggregation of free-floating cells or small aggregates through their swimming activities. Moreover, biological influence on aggregation can arise at the phytoplankton level, where sufficiently large and sticky cells can enhance encounter likelihood and coagulation efficiency (Laurenceau-Cornec et al., 2015).

### 3.2.2.2 Degradation and fragmentation of particles

Aggregates, as per laboratory studies, exhibit reduced porosity and increased density as their size decreases (Laurenceau-Cornec et al., 2015). This size decrease (Alldredge et al., 1990) has been associated with the deformation of sinking aggregates due to fluid flow around them (Alldredge and Gotschalk, 1988). In rolling tanks, Song and Rau (2022) observed aggregates elongate under shear until they became too fragile and split into separate particles. Appendicularian houses were found to become denser over time, which also resulted in lower fractal dimensions and higher sinking speed (Lombard and Kiørboe, 2010).

Filter feeding reduces particle abundance and volume across size spectra (Gillard et al., 2022). Flux feeding, on the other hand, decreases total biovolume (Burd and Jackson, 2009; Gillard et al., 2022; Stemmann, Jackson, and Ianson, 2004). It can be passive (Christiansen et al., 2018; Gilmer and Harbison, 1986; Gowing, 1989; Kiørboe, 2011) or active (Dilling, 2004; Dilling and Alldredge, 2000; Frost, 1972; Jackson and Kiørboe, 2004; Kiørboe and Jackson, 2001). Zooplankton may consume particles partially (Steinberg and Landry, 2017; Stemmann, Jackson, and Ianson, 2004), altering concentration and biovolume. Particle fragmentation, caused by organisms like euphausiids (Bach et al., 2019; Dilling and Alldredge, 2000) and copepods (Toullec et al., 2019), disrupts aggregates and can create smaller slower-sinking particles, increasing their residence time and remineralization potential (Alldredge et al., 1990; Bach et al., 2019; Dilling and Alldredge, 2000).

Bacterial degradation of marine snow (particles and fecal pellets) primarily occurs in the upper water column (Iversen, 2023; Turner, 2015), facilitated by bacterial colonization and growth (Kjørboe, 2001; Kjørboe et al., 2002). This process causes a rapid attenuation of sinking particulate organic carbon (POC) (Iversen, 2023), diminishing marine snow volume without loss in particle abundance (Gillard et al., 2022; Iversen and Ploug, 2010). Microbial consumption alters the geometry of a given particle, either hollowing it out or shrinking it (Stemmann, Jackson, and Ianson, 2004), leading to increased porosity and a reduced fractal level. Bacterial remineralization declines with depth as aggregates are colonized by surface-derived bacteria that sink into the deep sea (Iversen, 2023). Their activity then depends on environmental factors; high activity occurs in warmer waters (E. Cavan and Boyd, 2018; Marsay et al., 2015), impacting microbial physiology. In contrast, low activity was measured in oxygen minimum zones (Weber and Bianchi, 2020) and at high pressure (Amano et al., 2022). Particle composition similarly influences their activity (Laurenceau-Cornec et al., 2015). All these processes exert an influence on particle morphology, an aspect for which our current understanding remains limited.

### 3.2.2.3 A difficult study of particles morphology

Due to the constraints associated with aggregate research, the *in situ* continuous change in particle morphology has been seldom explored in the past (Trudnowska et al., 2021). Over recent decades, marine snow studies have mainly concentrated on quantifying particle abundance, sinking speed, and flux (Cael et al., 2021; Guidi et al., 2008; Iversen and Ploug, 2010), degradation rates (Kjørboe, 2000; McDonnell et al., 2015) or modelling various processes impacting their abundance or flux (Burd and Jackson, 2009; Stemmann, Jackson, and Ianson, 2004). While particle size has been the focus of various studies (Cael et al. (2021), Guidi et al. (2008), and Stemmann et al. (2002)), few have explored changes in other morphological traits (Giering et al., 2020; Trudnowska et al., 2021). Individual particles can be sampled using instruments such as the Marine Snow Catcher (Lampitt et al., 1993) or gel traps (Durkin et al., 2021), as well as by examination in laboratory-controlled settings like flow chambers (Ploug and Jørgensen, 1999), settling columns or roller tanks (Ionescu et al., 2015; Shanks and Edmondson, 1989; Song and Rau, 2022). Some studies even generate particles artificially to study their temporal evolution (Laurenceau-Cornec et al., 2020; H. Ploug et al., 1999; H. Ploug and Grossart, 2000). While these methods aid aggregate analysis and chemical study, they overlook various *in situ* parameters influencing particle morphology in the water column (Alldredge and Gotschalk, 1988; Cael et al., 2021). In particular, the diverse marine organisms and their impacts on particle morphology

can hardly be studied with the above-mentioned focus on individual particles studied *ex situ*. Enhanced understanding and modelling of processes shaping particle morphology while sinking necessitate more *in situ* data acquisition and analysis.

This limitation can be addressed by employing non-intrusive *in situ* imaging technologies like the Underwater Vision Profiler 5 (UVP5, Picheral et al. (2010)). The UVP5 was successfully used to study particle abundance and distribution with high temporal and vertical resolution (Guidi et al., 2008; Kiko et al., 2017; Kiko et al., 2022; Stemmann et al., 2002). These *in situ* sensors were predominantly used to assess particle size (Guidi et al., 2009), but a recent study has also for the first time explored particle morphology (Trudnowska et al., 2021). This research used UVP5 images to investigate aggregate morphology evolution in Baffin Bay and in Fram Strait using k-mean clustering. This unsupervised technique enabled the exploration of Arctic ice-associated phytoplankton bloom dynamics at a larger scale through changes in marine snow morphology over space and time.

Our study focuses on an anticyclonic retentive eddy observed in May 2021 in the North Atlantic near the Porcupine Abyssal Plain (PAP) station as part of the EXPORT Processes in the Ocean from RemoTe Sensing (EXPORTS) program (Siegel et al., 2016, 2021). The EXPORTS initiative employs satellite remote sensing, field sampling, and modelling to enhance understanding of the biological carbon pump's mechanisms (Siegel et al., 2016, 2021). We aimed to explore the North Atlantic spring bloom (Omand et al., 2015) within a drifting eddy situated between 14.9 and 14.7°W and 49.1 and 48.7°N (Erickson et al., 2022). To gain insights into open ocean particle distribution and dynamics, we deployed three Underwater Vision Profiler 5 units (UVP5, Picheral et al. (2010)) on a CTD rosette aboard the three research vessels RRS Discovery DY130, RRS James Cook JC214, and R/V Sarmiento de Gamboa SG2105. In this article, we present a series of 26-day deployments using the UVP5. We introduce a plume-based approach employing UVP5 counts and images to enhance the understanding of *in situ* particle morphology during an export event. By tracking an export plume, our study aims to achieve the following objectives: 1) Quantify carbon flux and its attenuation during export by following the particle plume; 2) Characterize the evolving community of imaged particles; 3) Investigate potential changes in the morphological features of aggregates and feces within this plume over time and with depth.

### 3.3 Material and Methods

#### 3.3.1 Data collection and processing

##### 3.3.1.1 In situ imaging in the EXPORTS 2021 experiment

The UVP5 was developed to detect, measure, and quantify the vertical distribution of zooplankton and marine particles (Kiko et al., 2022; Picheral et al., 2010). All particles with an Equivalent Spherical Diameter (ESD)  $\geq 100 \mu\text{m}$  were measured and counted by the UVP5 during its descent. Additionally, vignettes of zooplankton and aggregates with an ESD  $\geq 600 \mu\text{m}$  were recorded by the UVP5. A thorough description of the UVP5's operation is provided in Picheral et al. (2010). For further information on raw data treatment, see also Kiko et al. (2022).

Criteria for detecting and monitoring eddies were established months before the field campaign began. This was done to find an eddy that would remain coherent and retentive until the completion of the field campaign (Erickson et al., 2022). Through the use of altimetry data, numerical modelling and the deployment of ocean gliders and SeaGliders, the eddy-tracking team ended up selecting a small anticyclonic eddy (Erickson et al., 2022) with a core region diameter of approximately 30 km (Johnson et al., 2023).

In total, the dataset includes 173 vertical profiles done by the three UVP5s intercalibrated before the cruise (Fig 3.2). We downloaded the corresponding Ecopart file from Ecotaxa (Picheral et al., 2017) containing the metadata and particle counts.

##### 3.3.1.2 Inside the eddy core

The eddy core waters demonstrated high retention, and as a result, offer an opportunity to investigate the connection between the evolution of particle morphology and the associated sinking fluxes. The Surface Core Water (SCW) was defined by Johnson et al. (2023) as the lower limit of the SCW based on the depth of the  $27.1 \text{ kg m}^{-3}$  isopycnal. Below this SCW is the Eddy Core Waters (ECW), defined by this same context paper as being between the lower boundary of the SCW and around 600 m depth. Although many profiles were conducted at various distances from the eddy's centre, we will focus in this article on the eddy core as defined by Johnson et al. (2023) ( $<15\text{km}$  from the eddy centre). We computed the distance between each profile and the eddy centre at the sampling time and selected for further analysis the 99 profiles located less than 15 km away from the eddy centre.

As only 6% of the profiles covered the 800-1000 m layer, only images between 0-800 m were kept for further analysis. Once the artefacts -

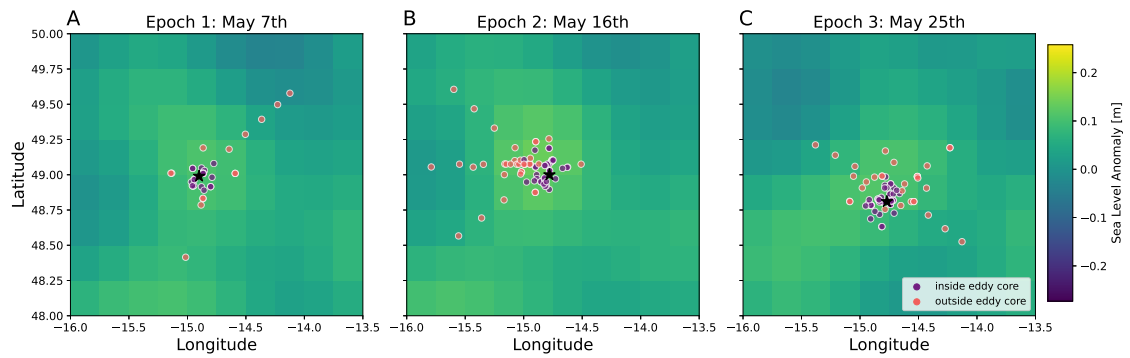


Figure 3.2: Sampling points inside ( $\leq 15$  km from the eddy center) and outside ( $> 15$  km) of the eddy core. The points represent the UVP5 vertical profiles done during the three epochs. The coloured background corresponds to the daily sea level anomaly (in m) for the middle day of each epoch: (A) May 7th for epoch 1, (B) May 16th for epoch 2 and (C) May 25th for epoch 3. The black star in each of the 3 figures corresponds to the centre of the eddy on the corresponding day.

mainly due to bubbles and unfocused images - were removed, the data set contained 1,720,914 images.

### 3.3.1.3 Image sorting and classification

The Zooprocess software was used to process all 1,720,914 images and log metadata (geographical position, date, etc.). To deal with this large data set, the images were grouped using Morphocluster, a program designed to annotate large data sets of images (Schröder et al., 2020). A more extensive description of the Morphocluster steps can be found in Schröder et al. (2020). The 1,720,914 images were clustered in the 155 clusters leaving 61,374 images non-clustered.

The Ecotaxa program (Picheral et al., 2017) was then used to manually validate the clustered images (Irisson et al., 2022), reclassifying by hand the wrongly classified ones. In order to do that, we treated each annotated class separately. We arranged the images with 100 images per page and sorted them by decreasing size. If less than five objects were relocated from three consecutive pages for each annotated class, sorting was halted. Once this was done, we used the Ecotaxa embedded random forest classifier to learn on the clean clusters and classify the 61,374 images that were not clustered by Morphocluster.



### 3.3.1.4 Morphology of imaged objects

The Zooprocess software also provided information on 42 morphological features associated with each object. Porosity is not computed by Zooprocess but is described by Alldredge and Gotschalk (1988) as the fraction of an aggregate not occupied by solid matter. It was computed following this formula:

$$porosity = \frac{V_{empty\ space}}{V_{total}}$$

with

$$V_{empty\ space} = \frac{4}{3}\pi \left( \frac{\text{area empty space}}{2} \right)^3$$

$$V_{total} = \frac{4}{3}\pi \left( \frac{\text{area empty space}}{2} \right)^3$$

The circularity computed by Zooprocess does not have a unit and ranges from 0 (elongated polygon) to 1 (perfect circle). The mean grey level corresponds to the average grey value within the object (G. Gorsky et al., 2010). It is a sum of the grey values of all pixels in the object divided by the number of pixels and its value can range from 0 (black) to 255 (white) (G. Gorsky et al., 2010). The fractal variable corresponds to the fractal dimension of the object boundary and is computed following Bérubé and Jébrak (1999).

## 3.3.2 Data selection according to the export plume

### 3.3.2.1 Sinking speed estimation of the export plume

Following a similar method to the one described by Briggs et al. (2020) in which the evolution of particulate optical backscattering (bbp, a proxy for particulate mass concentrations (Reynolds et al., 2016)) of small and large aggregates was studied for different export events, we applied a Gaussian fit to the daily median concentration of different aggregate size classes per 100 m bins. This method was applied both to the concentration of imaged particles (either aggregates or feces) and also to the UVP<sub>5</sub> particle counts. For each size class, a linear regression was done on the coordinates in depth and time of each of the Gaussian fits' maximums. The value of the slope was stored as an indicator of the sinking speed if the Gaussian fit was successful for at least half the 100 m bin. We obtained a sinking speed of  $41.7 \pm 5.7$  m d<sup>-1</sup> for

aggregates, of  $36.4 \pm 0.3 \text{ m d}^{-1}$  for feces and of  $44.3 \pm 7.3 \text{ m d}^{-1}$  for particle counts. Nevertheless, this method did not work for all the size classes because of the low signal-to-noise ratio at both limits of the size range. A comparison with sinking speed-to-size linear relationships from the literature was done with the obtained slope values. The relationship presented in Kriest (2002) (Reference 9, Table 2) seemed to be closest to both estimation based on counts and images, we computed the MiPs (small particles with 0.14-0.53 mm diameter, Kiko et al. (2017)), MaPs (large particles with 0.53-16.88 mm diameter, Kiko et al. (2017)) and total flux in  $\text{mgC m}^{-2} \text{ d}^{-1}$  based on it.

To investigate the plume depicted in Figure 3.3, we assumed a homogeneous sinking speed for the particle community at  $50 \text{ m d}^{-1}$ , selecting a value that was slightly higher than the upper boundary of previously computed values for different types of particles that did not yield significantly different estimates. This makes it possible to keep track of rapidly sinking particles. Six equally-spaced lines were traced starting at the surface from the 11<sup>th</sup> to the 21<sup>st</sup> of May with an interval of 2 days and a  $50 \text{ m d}^{-1}$  slope creating five different masks (Fig. 3.3A and B). These masks allow a study of the plume through time from mask 1 to mask 5. We then computed the mean value of the different parameters and morphological variables presented in Figure 3.8 per 20 m bins of every profile in each of the five masks. Note that the first mask (the one furthest to the left) covers the whole depth range from 0 to 800 m but this coverage is reduced as we go from mask 1 through 5. A power-based Martin fit was then computed on each of the masks' flux estimations following Martin et al. (1987). The mean mixed layer depth (MLD) at 42.7 m was used as the depth of reference.

### 3.3.2.2 Computation of the particle community metrics inside the plume

The morphological variables (ESD, mean grey level, circularity, fractal, porosity, elongation) were computed using the images' Zooprocess features while the parameters (MiP and MaP abundance) were computed on the particle count. Concentration and biovolume were computed per 5 m bins along each profile using water volume data from the UVP5 in  $\text{m}^3$ . The Particle Size Distributions (PSD) were computed on both particle counts and image aggregates. The PSD is computed by dividing the abundance of particles within a given size bin by the width of the ESD bin (Jouandet et al., 2011). The PSD slope is a useful tool for analysing the mechanisms that determine particle distributions and relative proportion based on their diameter (Clements et al., 2022; McCave, 1984; Stemmann, Jackson, and Ianson, 2004)). It was computed following Kiko et al. (2022).

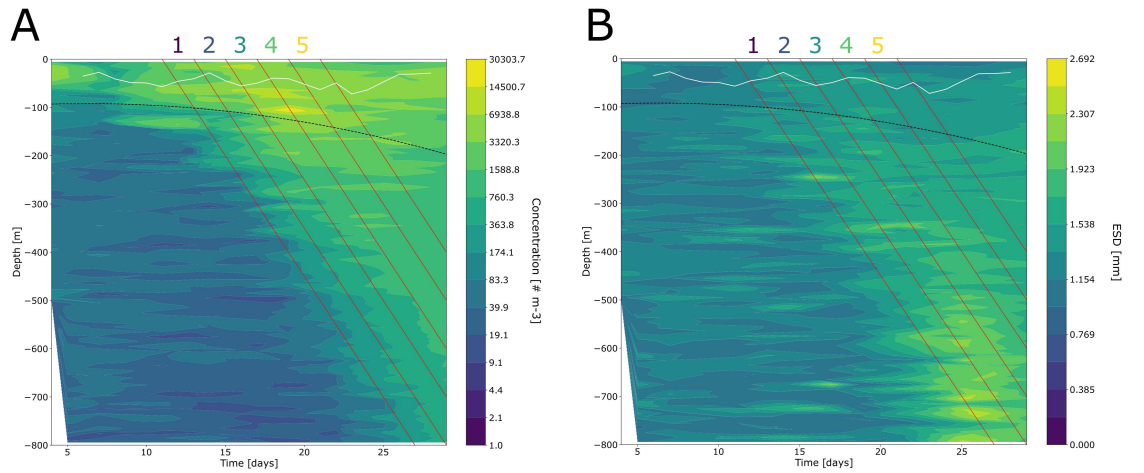


Figure 3.3: Interpolated field of A) aggregate concentration (in number per  $\text{m}^3$ ) and B) aggregate Equivalent Spherical Diameter (ESD in mm). Note that the colour bar is in logarithmic scale for the aggregate concentration. The 6 red lines correspond to the delimitation of the 5 masks with a  $50 \text{ m d}^{-1}$  slope. The continuous white line corresponds to the mean MLD. The black dashed line corresponds to the fit presented in Johnson et al. (2023) as the lower limit of the Surface Core Water based on the depth of the  $27.1 \text{ kg m}^{-3}$  isopycnal.

### 3.3.2.3 Computation of the POC loss inside the plume

The carbon-specific degradation rate ( $C_{\text{spec}}$ ) was computed following Iversen (2023):

$$C_{\text{spec}} = \frac{POC_{\text{loss}}}{POC_{\text{conc}}}$$

with  $POC_{\text{loss}}$  in  $\text{gC m}^{-3} \text{ d}^{-1}$  and  $POC_{\text{conc}}$  in  $\text{g m}^{-3}$

$$POC_{\text{loss}} = 1 - \frac{\Delta F}{\Delta z}$$

where  $\Delta F$  and  $\Delta z$  correspond to the differences in POC flux ( $\text{gC m}^{-2} \text{ d}^{-1}$ ) and in depth (m) inside the layer of interest.

$$POC_{\text{conc}} = \frac{F_{\text{top}}}{w_{\text{av}}}$$

where  $w_{\text{av}}$  corresponds to the assumed constant average particle settling velocity for particles at all depths, and  $F_{\text{top}}$ , the POC flux ( $\text{gC m}^{-2} \text{ d}^{-1}$ ) at the top of each layer of interest. We here choose the mean value of sinking speed obtained for aggregates ( $41.7 \text{ d}^{-1}$ ).  $C_{\text{spec}}$  was computed on the following depth definition: 50-100, 100-300, 300-500, 500-800 m.

## 3.4 Results

### 3.4.1 Evolution of particle concentration and size distribution inside the plume

The composition of the plume particles is largely dominated by fluffy particles with 92% of total images (particles with an ESD > 600 $\mu$ m) inside the plume between 0-50 m. This percentage decreases inside the plume and reaches its lowest value between 650-700 m with 81%. In the whole plume data, fluffy aggregates dominate (87.8%) followed by dense aggregates (10%), fecal pellets (1.1%), living organisms (0.94%) and fibers (0.2%).

The UVP5 counts of particles with a size between 100-600  $\mu$ m show a power law-based decrease in the MiP abundance at the surface to much lower values below 400m depth (Fig. 3.4A). The rapid decrease observed in the first 100 m for the MiP abundance is also observed, in a lesser intensity, for the MaPs (Fig. 3.4B). In terms of values, the mean MiP abundance is divided by 22.3 from 0-50 m to 750-800 m, while the MaP abundance is divided by 10 (Table S1). Furthermore, we may investigate the evolution of particle community metrics over time using the five masks, such as the exponent of the power law fit, which is a measure of export efficiency across depth. The MiP exponent is higher than the MaP exponent for the masks 1 through 3. They have the same exponent for mask 4 and the MaP exponent is higher for mask 5.

While representing the mean MiP and MaP abundance per 100 m depth bins throughout time inside the plume all masks combined, we observe different dynamics for the two sizes. The MiPs exhibit a general decrease in the top 100 m depth layer with two maximums around the 17<sup>th</sup> and 21<sup>st</sup> of May. The second layer also exhibits the second maximum on the 21<sup>st</sup> but presents low values (28-106 MiPs m<sup>-3</sup>) compared to the first (175-430 MiPs m<sup>-3</sup>). From the third layer (200-300 m) to the last (700-800 m) we observe very low values (< 35 MiPs m<sup>-3</sup>). The MaPs tell a different story with an increase in mean abundance in all layers from the surface to the 700-800 m layer.

The slope of the particle size distribution spectra (PSD slope) is negative throughout the whole plume (Fig. 3.4C). Its mean value is at -3.6 between 0-50 m and follows a linear decrease until it reaches the base of the SCW. Below this depth, its value remains relatively constant between 2.5 and 2.7 (Table S1), but with lower values and large variation in the first two masks (representing the beginning of the export) and higher values with very much reduced variation in masks 3 to 5 (representing the massive export). The slope value becomes flatter in the ECW from mask 1 to 5.

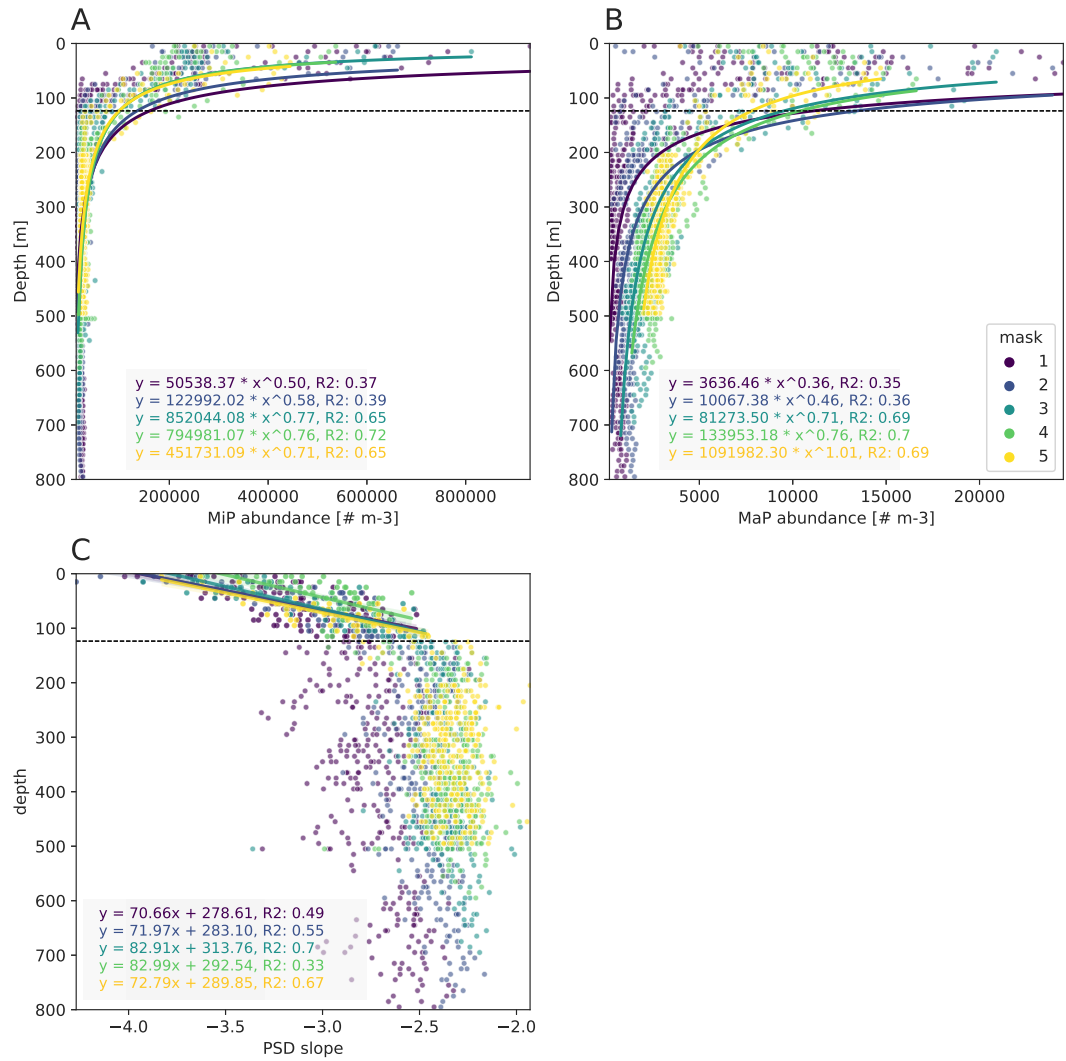


Figure 3.4: Evolution of (A) MiP and (B) MaP abundance as well as the PSD slope. Each point corresponds to a profile's 20m mean value. Power law functions were fitted to the (A) MiP and (B) MaP abundances (in numbers  $L^{-1}$ ). For the PSD slope (C), the linear regression was performed on the SCW data. The black horizontal dashed line corresponds to the mean depth of the SCW inside the plume.

### 3.4.2 Evaluation of flux and vertical attenuation inside the plume

The  $b$  values obtained using the power law regression fit are a reflection of the attenuation rate and the biological carbon pump efficiency in the different masks (Fig 3.5). For the global flux, the  $b$  value for mask 5 (0.259) was two-fold lower than the one for mask 1 (0.474). Higher values of  $b$  were obtained for the MiPs flux (Fig. 3.5A) compared to the MaPs (Fig. 3.5B) and global flux (Fig. 3.5C). In contrast, the values of MiPs flux reached much lower values (maximum around 200 mgC m<sup>-2</sup> d<sup>-1</sup>) compared to the MaPs flux which reached up to around 700 mgC m<sup>-2</sup> d<sup>-1</sup> for the 2nd mask in dark blue.

The carbon-specific degradation rates ( $C_{\text{spec}}$ , fig. 3.5D) obtained following Iversen (2023)'s methods presented two main messages. The first one was that  $C_{\text{spec}}$  values decrease rapidly in the first 300 m of the water column. The second was that these values were always distributed the same way when comparing the mask with a decrease from mask 1 to mask 5. A comparison is made in Figure 3.5D with other studies made in the North Atlantic using aggregates from trap (Collins et al., 2015) or from Marine Snow Catcher samples (Belcher et al., 2016).

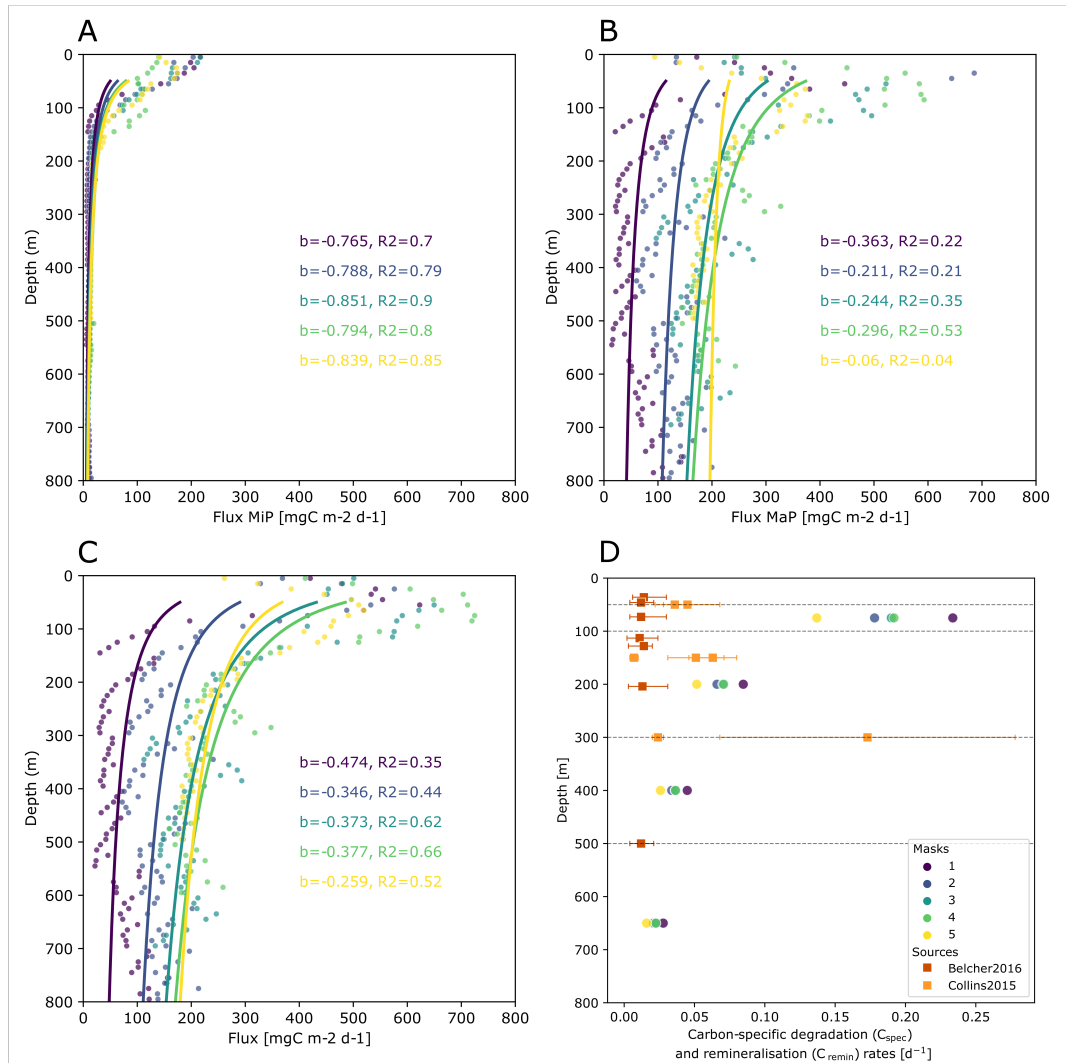


Figure 3.5: Power law regressions fitted the mean values of flux in  $\text{mgC m}^{-2} \text{d}^{-1}$  for (A) MiPs, (B) MaPs and (C) all particles. Points in figures A through C represent the mean value per 20m bins. The jumps in the vertical profiles are due to averaging overtime in the succeeding depth bins. The carbon-specific degradation rate (D,  $C_{\text{spec}}$  in  $\text{d}^{-1}$ ) is represented according to depth in figure D with circular points. The squared-shaped points represent data of carbon-specific remineralisation rate ( $C_{\text{remin}}$ ) from Belcher et al. (2016) and Collins et al. (2015) with the horizontal lines corresponding the standard deviation. The colours used to represent the masks in Figures A through D are presented in Figure D.

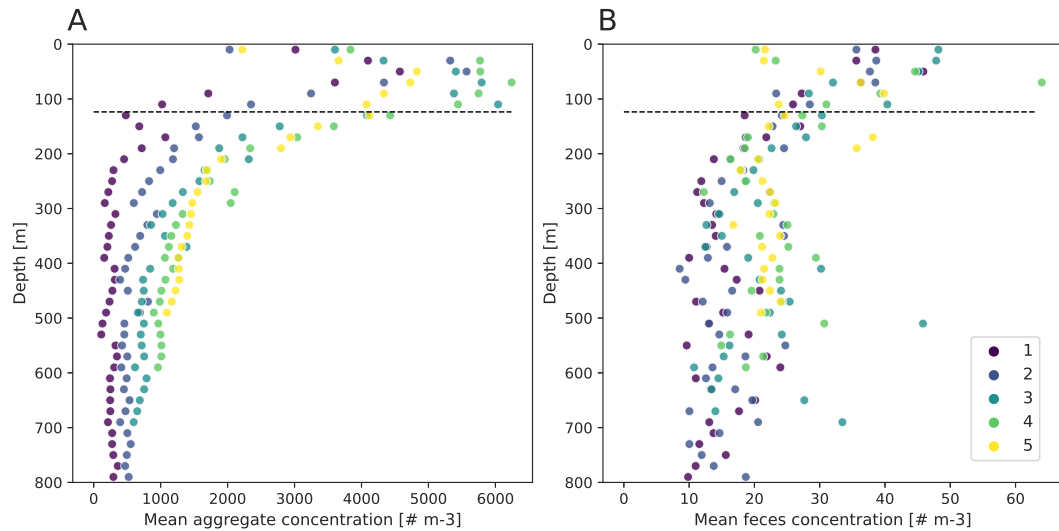


Figure 3.6: Evolution of mean concentration (in numbers  $\text{m}^{-3}$ ) of aggregates (A) and fecal pellets (B). Each point corresponds to a 20 m mean value for each mask for which the colour legend is presented in Figure B.

### 3.4.3 Evolution of the morphological properties of aggregates and fecal pellets inside the plume

The previous sections showcased results concerning particle size distribution, without differentiation by type; we now shift our focus to an in-depth analysis of aggregates and feces sorted using Morphocluster.

Aggregates have ESD values ranging from 0.89 to 27.7 mm, whereas feces have sizes ranging from 0.89 to 5.22 mm. Aggregates present an important decrease in concentration (Fig. 3.6A), especially in the first 200 m of the water column. Concentration decreases less sharply from mask 1 to 5. In terms of feces (Fig. 3.6B), we observe two dynamics: a reduction in mean concentration between 0 and 250 m, and almost constant values below that. The masks show a larger feces concentration in mask 5, as it did for the aggregates in Figure 3.6A, showing a higher aggregate concentration in the plume later in the experiment. It is worth noting that the aggregates/feces ratio decreases with depth, starting at the surface with 132 aggregates for one fecal pellet and reaching 17 at 700 m. It then increases again to 29 at 800 m (Fig. 3.8H).



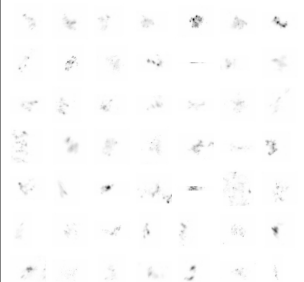
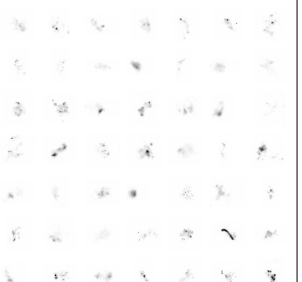
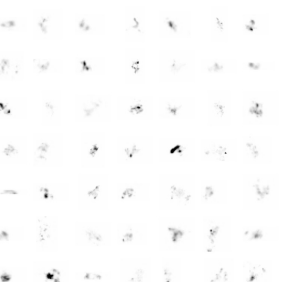
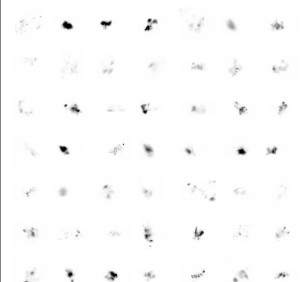
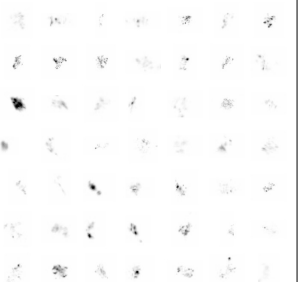
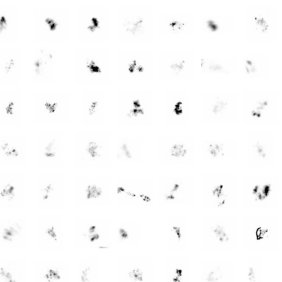
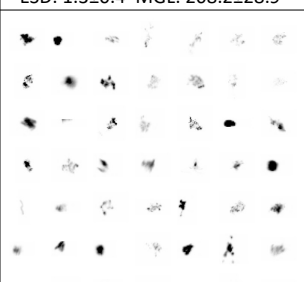
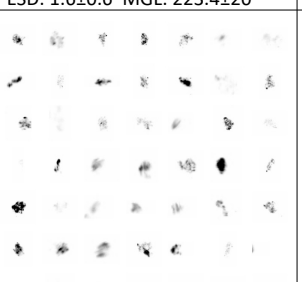
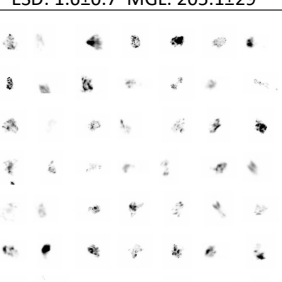
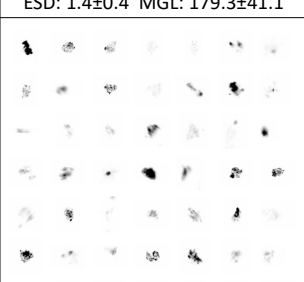
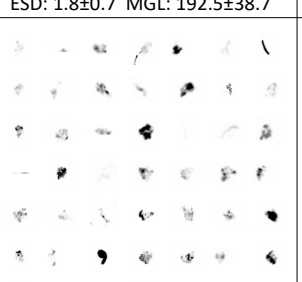
	Mask 1	Mask 3	Mask 5
0 - 100 m	 ESD: $1.3 \pm 0.3$ MGL: $229.4 \pm 14.3$	 ESD: $1.4 \pm 0.4$ MGL: $225.4 \pm 15$	 ESD: $1.2 \pm 0.3$ MGL: $221.5 \pm 17.6$
100 - 300 m	 ESD: $1.3 \pm 0.4$ MGL: $208.2 \pm 28.9$	 ESD: $1.6 \pm 0.6$ MGL: $223.4 \pm 20$	 ESD: $1.6 \pm 0.7$ MGL: $205.1 \pm 29$
300 - 500 m	 ESD: $1.4 \pm 0.4$ MGL: $179.3 \pm 41.1$	 ESD: $1.8 \pm 0.7$ MGL: $192.5 \pm 38.7$	 ESD: $1.7 \pm 0.7$ MGL: $186.2 \pm 38.7$
500 - 800 m	 ESD: $2.1 \pm 1$ MGL: $193 \pm 42.9$	 ESD: $1.8 \pm 1$ MGL: $182.1 \pm 44.2$	

Figure 3.7: Aggregates images selected randomly inside the plume for masks 1, 3 and 5 between 0-100, 100-300, 300-500 and 500-800 m. ESD values correspond to the mean  $\pm$  standard deviation Equivalent Spherical Diameter. MGL values correspond to the mean  $\pm$  standard deviation Mean Grey Level.

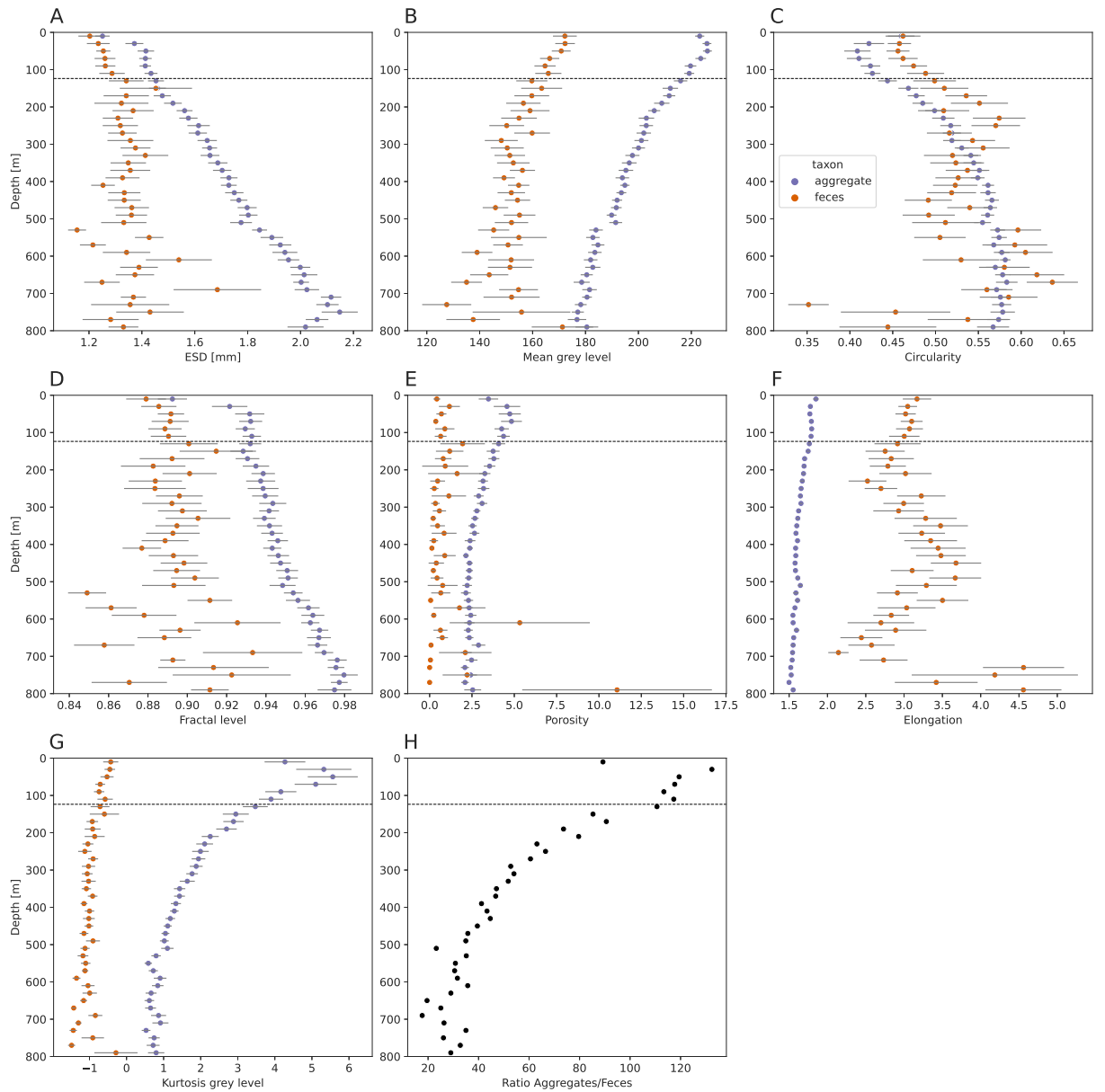


Figure 3.8: Comparison of aggregate and feces morphological variables. Each point corresponds to a 20 m mean value on all the profiles for either aggregate (in purple) or feces (in orange) among the five masks for figures A through G. Horizontal bars represent the 95% confidence interval. The ratio of aggregate over feces in the plume over depth is represented in Figure H. For each subplot, the black horizontal line corresponds to the mean depth of the Surface Core Waters inside the plume.

Regarding their morphology, aggregates exhibit a different dynamic in the SCW and below its lower bound. These two dynamics were represented through two linear regressions in two different layers (Fig. S3.5): the first fitted between 0 and the depth of the mean SCW lower bound, and the second fitted below this limit. All linear regressions are significant (Kruskall-Wallis,  $p - value < 0.05$ ). They demonstrate variances in dynamics at various levels. First, we observed differences comparing variables, e.g. the ESD (Fig S3.5A) has a trend of increase for both layers while other variables show more heterogeneity. Secondly, we also noted differences in the trends for one variable inside one of the layers, e.g. for the mean grey level, circularity and fractal level (Fig S3.5B, C, D), the linear regressions did not have the same trend of increase or decrease within the same layer of water. On average inside the plume, aggregates became bigger (ESD decreased from 1.3 mm between 0-50 m to 2.1 mm between 750-800 m, Tables S3.2, S3.3), darker (mean grey decreased from 225 to 177), more circular (from 0.4 to 0.6) and less elongated (from 1.8 to 1.5). They also presented a higher fractal dimension (from 0.91 to 0.98) and a lower porosity (from 4.2 to 2.1). Figure 3.7 illustrates the evolution of the aggregates by presenting 49 random images from masks 1, 3 and 5 for the depth layers 0-100, 100-300, 300-500, 500-800 m. The ESD increased as the mean grey level decreased across all depth layers and masks from 1 to 5, indicating that the particle grew darker. We also noticed the increasing prevalence of black components in fluffy particles, as well as a rise in the fraction of dense particles. Another indicator of aggregate structure is the kurtosis of grey level which indicates a peaked (positive values) or flat distribution (negative values) of grey values in a particle (Legendre and Legendre, 2012). Aggregates kurtosis level (Fig 3.8G) depicts an increasingly more peaked distribution than normal distribution in the first 60 m translating homogeneous aggregates with more observed values around the mean grey level. It then decreases below this depth to reach less than 1 at 800 m depth indicating more heterogeneous aggregates with a more spread out distribution of observed grey level values. The linear regressions throughout the five masks from mask 1 to 5 can also be an indication of the evolution of the morphology in time. A trend of increased slope in aggregates' circularity, porosity and elongation can also be noted. In contrast, the ESD slopes appear more stable throughout the masks.

In regards to the SCW limit, some parameters behaved differently above and below it e.g. particle mean grey level increased above it and decreased below the SCW. Porosity and elongation also seemed to be affected by the same trend of increase in the SCW and decrease below it.

No clear pattern emerges from the feces results compared to the aggregates' (Figure S3.6) with more spread out values than compared

with figure S3.5. In comparison to aggregates, fecal pellets were smaller, and darker (mean grey level between 127-172 for feces compared to 176-226 for aggregates), with a lower porosity and more elongated (between 2.1 and 4.5 compared to 1.5 to 1.8 for the aggregates, Fig. 3.8). The mean kurtosis grey level is negative and takes values approximately between -2 and 0 throughout the water column. This corresponds to a flatter distribution than the normal distribution of grey level values, indicating heterogeneity in the images' grey level. The 95% confidence intervals were wider for feces compared to aggregates. Note that the SCW limit seemed to have an effect on the ESD as well as the fractal level (values almost constant for both aggregates and feces above this limit and increasing for aggregates below this limit while feces have a more heterogeneous distribution). These patterns can be observed in Figure S3.7 which contrasts with Figure 3.7 by a lack of visible pattern from surface to 800 m and from mask 1 to 5.

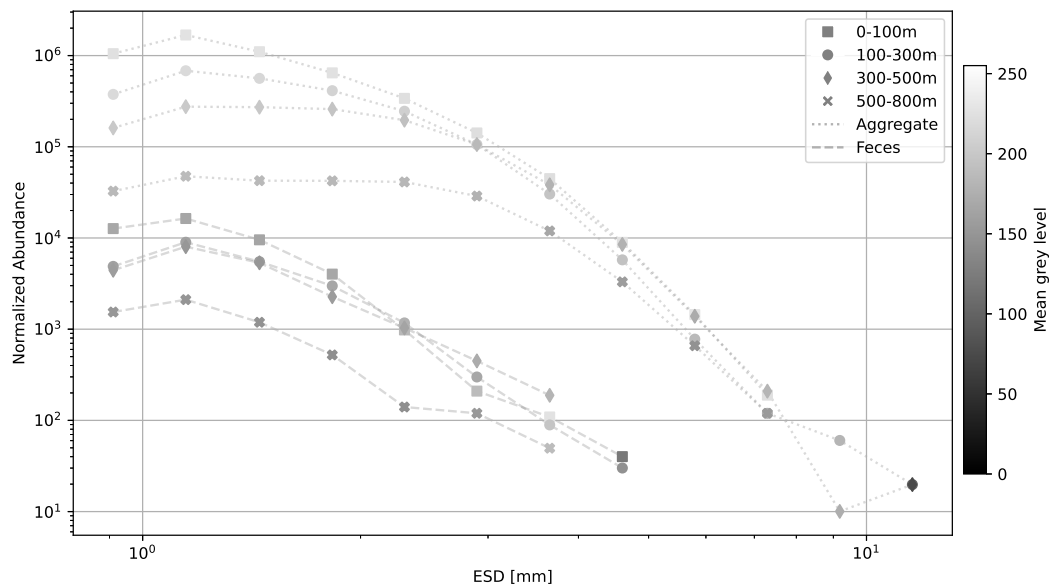


Figure 3.9: Particle Size Distribution of imaged aggregates and feces for data between 0-200, 200-400, 400-600 and 600-800 m. The two axes are log-transformed. The colour of the points corresponds to the mean grey level of the corresponding combination of particle type, depth layer and size bin. Values for feces are linked with dashed lines and the ones for aggregates are linked with dotted lines.

The development of imaged aggregates and feces normalised abundance with their respective mean grey level throughout binned depth layers are summarised in Figure 3.9). It illustrates the decrease through

depth and size classes of aggregates and feces abundance. It also highlights that aggregates darkened from surface to depth and from small to large size classes. In terms of size classes, aggregates populate a larger range than feces that goes beyond 5 mm. Aggregates spectrum exhibits a flattening of the curve from surface to depth for the size classes  $<2$  mm with a similar slope above 600 m for sizes  $>2$  mm. Note an important flattening of the slope for 600-800 m. For feces, we also see a difference between 0-600 m and 600-800 m, the lowest curve showing a more important reduction in abundance. The abundance of large fecal pellets bigger than  $>3$  mm does not decrease with depth and presents an increased abundance in the 300-500 m layer compared to the shallower depths.

Overall, these results show that aggregates presented a clear morphological change with depth. Changes in fecal pellets are more dynamic and heterogeneous.

### 3.5 Discussion

Using three UVP5 instruments within a retentive eddy in the North Atlantic, approximately 170 km east of the PAP site (Johnson et al., 2023), we gathered a significant 26-day dataset during the EXPORTS 2021 experiment and tracked the plume of settling detritus during a large export event. Our study quantified for the first time the development of an extensive export plume within an eddy core in terms of composition, flux, and attenuation across both time and space using a Lagrangian approach. We aimed to elucidate the decrease in estimated flux through the evolving morphology of aggregates and feces. We now discuss the observed changes in the entire particle community and their relationship to flux attenuation within the plume, followed by an exploration of the evolving morphological characteristics of aggregates and feces within the plume.

#### 3.5.1 The plume approach

In distinction from the conventional vertical approach that attempts to link layers with distinct but unrelated particle dynamics, the plume approach employs a Lagrangian framework to follow a particle population, resulting in a more coherent method. We postulated that employing the Martin approach within the extracted plume is more suitable than cutting vertical data bands, as our dataset allows direct plume monitoring as it sinks. Our study took advantage of the initial scarcity of particles in the eddy, providing a clear insight into the particle community through depth within the plume, and time through the applied

masks. Furthermore, we calculated the sinking speed for various particle types (aggregates and fecal pellets) in this plume. Our approach's results align with the previously observed range of 10 to 100 m d<sup>-1</sup> for 0.81 to 3.25 mm sized particles (Riley et al., 2012; Villa-Alfageme et al., 2014). Notably, it revealed a significant relationship for aggregates between size and sinking speed in the narrow observed size range.

### 3.5.2 Major changes in the particle community inside the plume

#### 3.5.2.1 Power-based decline of MiPs and MaPs abundance

We observed a power-based reduction in the abundance of MiPs and MaPs within the plume. This decline could be influenced by degradation and aggregation processes. If fragmentation were the dominant factor, we would anticipate an increase in MiP abundance following a decrease in MaP abundance, as MaPs break down into smaller particles. However, we do not observe this trend, even though MiP abundances significantly outnumber MaP abundances. Another possibility is that particles are fragmenting into <100 µm aggregates than what the UVP5 detects.

Another potential contributor to biovolume reduction (Fig. S3.2) is microbial activity associated with particle remineralization. This process entails a decrease in overall volume without necessarily leading to a commensurate decrease in total particle abundance (Gillard et al., 2022). This mechanism has been demonstrated to be more effective in the upper 300 m, with up to 50% of carbon being remineralised (Martin et al., 1987). Remineralisation shrinks particles without obliterating them (Iversen and Ploug, 2010). However, in our observation, the decrease in total biovolume, particularly for MiPs, aligns with a decline in total particle abundance. This suggests that remineralization might not be the predominant process at play, especially in the upper water layers. For MaPs, we observe a different trend with a decline in biovolume vertically but an increase over time from mask 1 to 5 (Fig. S3.2 B).

Examining the mean abundance of MiPs and MaPs within fixed-depth layers in the plume revealed distinct behaviours for the two size categories (Figures S3.2, S3.3). The observed increase in MaPs at these layers could suggest a higher influx of larger sinking aggregates over time, possibly due to elevated particle production and export in the surface layer. Furthermore, the smaller exponent for the MiPs power fit compared to the one obtained for the MaPs indicates a less efficient export of MiPs within the water column. This disparity in exponents might imply that the slower sinking rate of MiPs, attributed to their

size, makes them more susceptible to processes like remineralization and differential settling with denser particles such as fecal pellets.

The reduction in the abundance of larger particles, along with a decline in their biovolume (Fig. S3.2), may indicate the influence of passive flux feeders such as pteropods and certain copepods (Gillard et al., 2022; Gilmer and Harbison, 1986; Kiørboe, 2011), as well as filter feeders like larvaceans, salps, and crustaceans (Alldredge et al., 2005; Huntley and Boyd, 1984; Robison et al., 2005). These feeding strategies intercept particles, leading to an overall decrease in the concentration and biovolume of MiPs and MaPs that the UVP5 detects. While large pteropods were observed in net samples above 500 m during the EXPORTS experiment, their contribution to fecal pellet production was relatively small compared to other zooplankton taxa (Steinberg et al., 2022). Though we observed a reduction in large particles through depth, it is much less pronounced than the reduction in small particles hinting that flux feeding is not a preponderant process at play here (Stemmann, Jackson, and Ianson, 2004). Moreover, zooplankton was in a highly reproductive phase with a high number of eggs and juveniles observed during the cruise (A. Maas, personal communication, August 24, 2023).

### 3.5.2.2 Two PSD slope dynamics

The PSD slope calculated for particle counts is a valuable indicator of particle dynamics (Clements et al., 2022; Stemmann and Boss, 2012). A flatter slope indicates a higher proportion of larger particles or a lower proportion of smaller particles (Clements et al., 2022; Jouandet et al., 2011). The observed reduction in MiPs and MaPs abundance leads to a flatter PSD slope with increasing depth. However, this vertical decrease follows a two-step pattern rather than a continuous change. Initially, as we descend in the SCW, the decrease in MiPs and the influx of larger particles from the surface causes the slope to flatten. Below the SCW, the slope stabilizes and continues to flatten over time from mask 1 to 5. In a previous study (Clements et al., 2022) using the UVP5 dataset (Kiko et al., 2022), the projected PSD slope values for May were observed to range from 3.65 to 3.7 at the base of the euphotic zone in the North Atlantic region. This layer was estimated to be around 40-50 m deep in our experiment, corresponding to PSD slopes of 2.9-3.5. These estimates are in line with those of Clements et al. (2022), albeit closer to their lower limit, indicating that our approach yields reasonable outcomes.

### 3.5.2.3 Flux and attenuation

Our ability to effectively track the evolution of aggregates and feces morphology and flux within the export event was made possible by

the utilization of the plume-following technique. Previous studies that used isotope disequilibrium techniques, modelling, bbp signal and lagrangian floats measurements did yield a range of carbon export flux values for the North Atlantic spring bloom (Table 3.1). POC flux estimations varied between 170-620 mgC m<sup>-2</sup> d<sup>-1</sup> at 100 m depth (Bagniewski et al., 2011; Briggs et al., 2011; Martin et al., 2011; Omand et al., 2015) which is compatible with our results ranging from 128.8 mgC m<sup>-2</sup> d<sup>-1</sup> for mask 1 to 373.7 mgC m<sup>-2</sup> d<sup>-1</sup> in mask 4. At 200 m, Briggs et al. (2011) computed a value of 270 mgC m<sup>-2</sup> d<sup>-1</sup> from bbp signal and (Bagniewski et al., 2011) estimated values based on lagrangian float data between 74.4-332.4 mgC m<sup>-2</sup> d<sup>-1</sup> which is also compatible with our results ranging from 92.7 mgC m<sup>-2</sup> d<sup>-1</sup> for mask 1 and 287.7 mgC m<sup>-2</sup> d<sup>-1</sup> for mask 4.

Our POC flux results reveal two notable patterns: a power-based decrease and an increase in carbon flux within the plume, particularly pronounced in the initial three masks. The last three masks present much closer values characterising the export at its highest strength in the UVP5 dataset. The comparison made with previous studies indicates that our method not only offers improved suitability for tracking the morphology of a sinking detritus plume but also yields flux values that are comparable to those obtained in traditional approaches.

The *b* values obtained in our study indicate a low attenuation rate of the computed POC flux within the export plume and a high efficiency of the biological carbon pump, which increases over time from mask 1 to 5. Notably, our estimated *b* values of 0.259 to 0.474 are considerably lower than those reported for the PAP site in previous studies (0.71 in Belcher et al. (2016), 0.7 in Marsay et al. (2015), and 0.57 in Cavan et al. (2017)). Previous global studies conducted in a vertical context by Guidi et al. (2015) and Henson et al. (2012) demonstrated substantial seasonal and regional variability in this *b* exponent. Their findings showed values ranging from 0.24 to 1.75, with higher regionalized *b* values of 0.92 in Guidi et al. (2015) and 0.69 in Henson et al. (2012) for this section of the global ocean. Remarkably, our results fall at the lower end of their modelled ranges, suggesting that *b* is significantly lower during an export event compared to steady-state situations, as typically encountered in studies not following plumes of settling particles. The *b* values derived from the traditional method (Figure S3.4) exhibited notably higher values at the plume's inception. This discrepancy is likely attributed to the substantial differences in particle distribution between the densely populated surface and the sparsely populated depth. Conversely, *b* values for the experiment's later days, when concentration was high throughout the water column, aligned closely with those obtained through the plume technique. Opting for



Reference	Source	Depth (m)	Flux estimate
Buesseler et al. (1992)	$^{234}\text{Th}$ disequilibria	35	60-492
Bagniewski et al. (2011)	Modeling	100	476.4-568.8
Bagniewski et al. (2011)	Modeling	200	74.4-332.4
Martin et al. (2011)	$^{234}\text{Th}$ disequilibria	100	369-620
Martin et al. (2011)	$^{234}\text{Th}$ disequilibria	320	76.2
Martin et al. (2011)	$^{234}\text{Th}$ disequilibria	600	164
Martin et al. (2011)	$^{234}\text{Th}$ disequilibria	750	154
Briggs et al. (2011)	bbp signal	100	514
Briggs et al. (2011)	bbp signal	200	270
Alkire et al. (2012)	Lagrangian float	57	432-684
Omand et al. (2015)	Modeling	100	170
This study	UVP5	100	128.8-373.7
This study	UVP5	200	92.7-287.7

Table 3.1: Flux estimates (in  $\text{mgC m}^{-2} \text{d}^{-1}$ ) from various studies done on the North Atlantic spring bloom. Buesseler et al. (1992) and Alkire et al. (2012) sampled at the JGOFS (Joint Global Ocean Flux Study) North Atlantic Bloom Experiment (NABE) at approximately  $47^\circ\text{N } 20^\circ\text{W}$ . Bagniewski et al. (2011), Briggs et al. (2011), Omand et al. (2015) and Martin et al. (2011) sampled in the south of Iceland.

the plume-following strategy facilitated a more comprehensive and specific understanding of the particles' progression across both time and depth. Given our goal of investigating the export process, the decision to track the plume diagonally was the most suitable approach.

The change in flux with depth offers insights into the carbon-specific degradation rate ( $C_{\text{spec}}$ , Iversen (2023)), corresponding to the sum of processes leading to particle flux attenuation (zooplankton feeding and swimming activity, bacterial remineralisation). The obtained  $C_{\text{spec}}$  values in our study exhibited a decline with both depth and time from mask 1 to 5. The rapid decrease observed in the upper layer indicates heightened biological activity in the uppermost water column, aligning with existing literature (Iversen, 2023; Martin et al., 1987; Stemmann, Jackson, and Ianson, 2004). However, the reduction in  $C_{\text{spec}}$  values was more gradual in the 3 deeper layers (100-300, 300-500, and 500-800 m). This suggests either lower biological activity or that biological factors had limited influence during the export event.

Previous research by Belcher et al. (2016) derived bacterial respiration rates within a small range of average values ( $0.011\text{-}0.014 \text{d}^{-1}$ ) using Marine Snow Catcher samples from 36 to 500 m depth at the PAP site. One additional study on trap aggregates was conducted in the

North Atlantic Ocean (Collins et al., 2015), aligning with Belcher et al. (2016) for the shallowest depth layer (50-100m). Between 100 and 300m, the values reported by Collins et al. (2015) are higher than those proposed by Belcher et al. (2016), and these values are much closer to our estimations. The variance between our results and these studies in the shallowest layer can probably be attributed to other processes than bacterial respiration such as zooplankton-related activities (flux and filter feeding, as well as fragmentation through swimming behaviour). These processes could significantly contribute to the pronounced flux attenuation observed near the surface. Nevertheless, given that the bacterial respiration rate estimates cited above generally align with the lower end of our  $C_{\text{spec}}$  values between 100 and 800 m, it is reasonable to conjecture that the attenuation of the export plume in the mesopelagic is mainly driven by bacterial respiration. This is in agreement with prior studies, which conclude that zooplankton and bacterial activity are both accountable for the large flux attenuation in the upper mesopelagic (50-300 m) (Iversen et al., 2010; Stemmann, Jackson, and Ianson, 2004), and that the dominant attenuation process below these depths is bacterial activity (Stemmann, Jackson, and Ianson, 2004).

The combined insights from the  $b$  and  $C_{\text{spec}}$  values offer a comprehensive understanding of flux attenuation. Globally, the attenuation is low (indicated by low  $b$  values). However, when examining various depths, the most significant attenuation occurs near the surface (evident by high  $C_{\text{spec}}$  values), in line with existing literature (Iversen, 2023; Martin et al., 1987; Stemmann, Jackson, and Ianson, 2004). It is worth noting that over time, the attenuation rates lessen, as indicated by declining  $b$  and  $C_{\text{spec}}$  values from mask 1 to 5, suggesting a potential slowdown in the export process. The biogeochemical perspective does not allow for a complete understanding of sinking particles, suggesting the need to delve into morphology for a more complete understanding.

### 3.5.3 Evolution of morphological variables inside the plume

For the first time, we characterise the distribution of detritus (fecal pellets, dense and fluffy aggregates) within an export plume and offer a morphological description using in situ imaging of aggregates and feces, achieved through a semi-supervised and ecologically relevant clustering technique.

#### 3.5.3.1 While aggregates' morphology is altered through depth...

Prior studies have primarily examined aggregate evolution in controlled laboratory settings, as described earlier. However, our study presents a novel approach by describing the continuous evolution of aggregate

features in an in situ context over a considerable depth range. We confirm the expected trends of denser and less porous aggregates with increasing depth, consistent with prior research (Laurenceau-Cornec et al., 2015). However, contrary to previous assumptions, we find that the ESD of the aggregate community increases along the export plume. This size increase of aggregates with depth is likely influenced by various factors, including zooplankton-induced particle compaction (Turner, 2002, 2015) and differential settlement (Alldredge, 2001; Burd and Jackson, 2009). The hypothesis of aggregation through differential settlement gains support from our observation of a reduced abundance of small aggregates with depth in the particle size distribution, along with the decrease in kurtosis grey level from 60 to 800 m. These findings suggest increased heterogeneity in aggregates further down the plume, likely enriched with elements like fecal pellets.

Deep within the plume, we observed a trend towards more circular and less elongated aggregates which could be linked to the fragmentation of elongated aggregates (Song and Rau, 2022). It's plausible that biological processes are also contributing, as the outer, more accessible parts of aggregates could undergo remineralization or shedding (Kiko et al., 2017).

#### 3.5.3.2 ...feces exhibit a different dynamic.

Fecal pellets come in various shapes, influenced by factors such as the organism producing them (Atkinson et al., 2012; Gleiber et al., 2012; Koster et al., 2011), food source concentration and type (Dagg and Walser, 1986; Feinberg and Dam, 1998), and the processes that might have affected their structure (Iversen, 2023; Lampitt et al., 1990; Svensen et al., 2012; Turner, 2002). This inherent variability, coupled with the relatively limited number of fecal pellet images compared to aggregates, could account for the heterogeneous signal observed in our study.

The highest elongation and minimal circularity values were observed in the SCW and between 400 and 500 m, where potential producers of fecal pellets, such as crustaceans and rhizarians, were found at higher levels. Zooplankton act as POC flux gatekeepers (Jackson and Checkley, 2011) and their contribution to fecal pellet production varies both spatially and temporally (Turner, 2015). We hypothesise that crustaceans are present at these depths likely through diel vertical migration (Steinberg and Landry, 2017; Steinberg et al., 2022), which has been linked to gut flux (Kiko et al., 2020; R. Lampitt et al., 2010), enhancing the efficiency of carbon export (Gorgues et al., 2019) through the production of fecal pellets which can be incorporated in marine snow through differential settlement.

This is supported by the maximal fecal pellet concentration observed in the SCW and between 250-600 m (particularly for masks 3-5). Among the zooplankton, Rhizaria have been demonstrated to influence the biological carbon pump by intercepting particles (Stukel et al., 2018, 2019) and generating particle flux (Lampitt et al., 2009). In our dataset, Rhizaria concentrations increased between 300 and 500 m (Fig. S3.8), coinciding with the elevated mean feces concentration, especially in masks 3-5. In this layer, rhizarians were primarily represented by Phaeodaria but also by Foraminifera, consistent with prior observations of their vertical distribution (Biard and Ohman, 2020). Phaeodarians, often found below 150-200 m (Biard and Ohman, 2020; Boltovskoy et al., 2017; Nakamura and Suzuki, 2015; L. Stemann et al., 2008), are known flux feeders (Gowing, 1989) and significant contributors to the POC flux (Ikenoue et al., 2019). These organisms have been associated with the formation of fast-sinking gelatinous aggregates in the region (Lampitt et al., 2009; Riemann, 1989), suggesting that their presence in our data in the mesopelagic layer influences both POC flux attenuation and generation through mini pellet production (Ikenoue et al., 2019).

Fecal pellets present both at the surface and at depth showcase various morphologies ranging from stick-like to curled-up forms across the plume (Fig. S3.7). Below 500 m, the size and fractal level distribution appear heterogeneous, likely due to the presence of fresh, stick-like fecal pellets generated by zooplankton organisms, alongside curled-up fecal pellets that have undergone bacterial (Svensen et al., 2012; Turner, 2015) and zooplankton-mediated degradation (Iversen, 2023; Lampitt et al., 1990).

### 3.6 Conclusion and perspective

Our study uncovered distinct dynamics across the water column, likely influenced by a combination of physical and biological processes that impact particle flux and its attenuation. Parameters such as porosity (Bach et al., 2019; Cael et al., 2021), composition (Laurenceau-Cornec et al., 2015), structure (Johnson et al., 1996), and density have been previously recognized as important for sinking aggregates, but our study uniquely provides insights in a lagrangian and in situ context, utilizing a plume-following technique. The evolution of aggregates and fecal pellets morphology observed in this study is a strong indication that UVP5 particle counts should be used keeping in mind that these counts do not encompass the complete detritus narrative. When computing flux, a more comprehensive understanding of particle morphology can be crucial. The impact of phytoplankton community composition on the morphological characteristics of aggregates and feces may be of particular significance.

Similar to the application of trait-based approaches to planktonic organisms (Buitenhuis et al., 2013; Le Quéré et al., 2016) to explore the functioning of phytoplankton (Guidi et al., 2009) and zooplankton (Brun et al., 2016; Orenstein et al., 2022; Vilgrain et al., 2021) communities and their impact on biogeochemical cycles, we propose that the traits identified in this study should be considered to enhance our comprehension of the life cycle of marine snow from the ocean's surface to its depths. Building upon the work of Trudnowska et al. (2021), who employed unsupervised clustering to study marine snow particles, our approach benefits from the intrinsic meaningfulness of Morphoclust's classifications which enabled the extraction of well-recognized categories such as aggregates and feces. Our results clearly show that multiple morphological characteristics (porosity, density, circularity, elongation, etc.) show an unexpected evolution during the settling of the plume. Consequently, this study advocates for the development of tools and methodologies that can incorporate these factors into future flux calculations.

Much like the attention given to zooplankton, which has led to the creation of various classification tools based on morphological traits rather than just size (Irison et al., 2022), this work underscores the importance of classifying marine particles with greater attention (Schröder et al., 2020). By acknowledging the complexity of particle morphology and its dynamic evolution, we can unlock a more comprehensive understanding of the intricate processes driving particle flux and biogeochemical cycling in the ocean.

## Funding

The EXPORTS program is funded by the NASA Ocean Biology and Biogeochemistry program with contributions from the U.S. National Science Foundation. LD, RK and LS received support by the European Union project TRIATLAS (European Union Horizon 2020 program, grant agreement 817578) and the Sorbonne Université through the Ecole doctorale 129. RK acknowledges support via a Make Our Planet Great Again grant from the French National Research Agency (ANR) within the Programme d'Investissements d'Avenir ANR-19-MPGA-0012 and funding from the Heisenberg Programme of the German Science Foundation KI 1387/5-1.

## Acknowledgments

We are grateful to the EXPORTS scientific team both on board of the three ships and on land for the organisation and realisation of this

one of the kind oceanic experiment that took years of planning. We sincerely acknowledge the Captains, Crews, Research Technicians and Administrative Staff of the three ships James Cook (JC214), RRS Discovery (DY130 & DY131) and the R/V Sarmiento de Gamboa without whom this experiment would never have taken place. Thank you for making this campaign possible and for your support especially during the three storms that hit us during the fieldwork. We also thank Florian Ricour for his help with the interpolated fields.

## 3.7 Supplementary Material

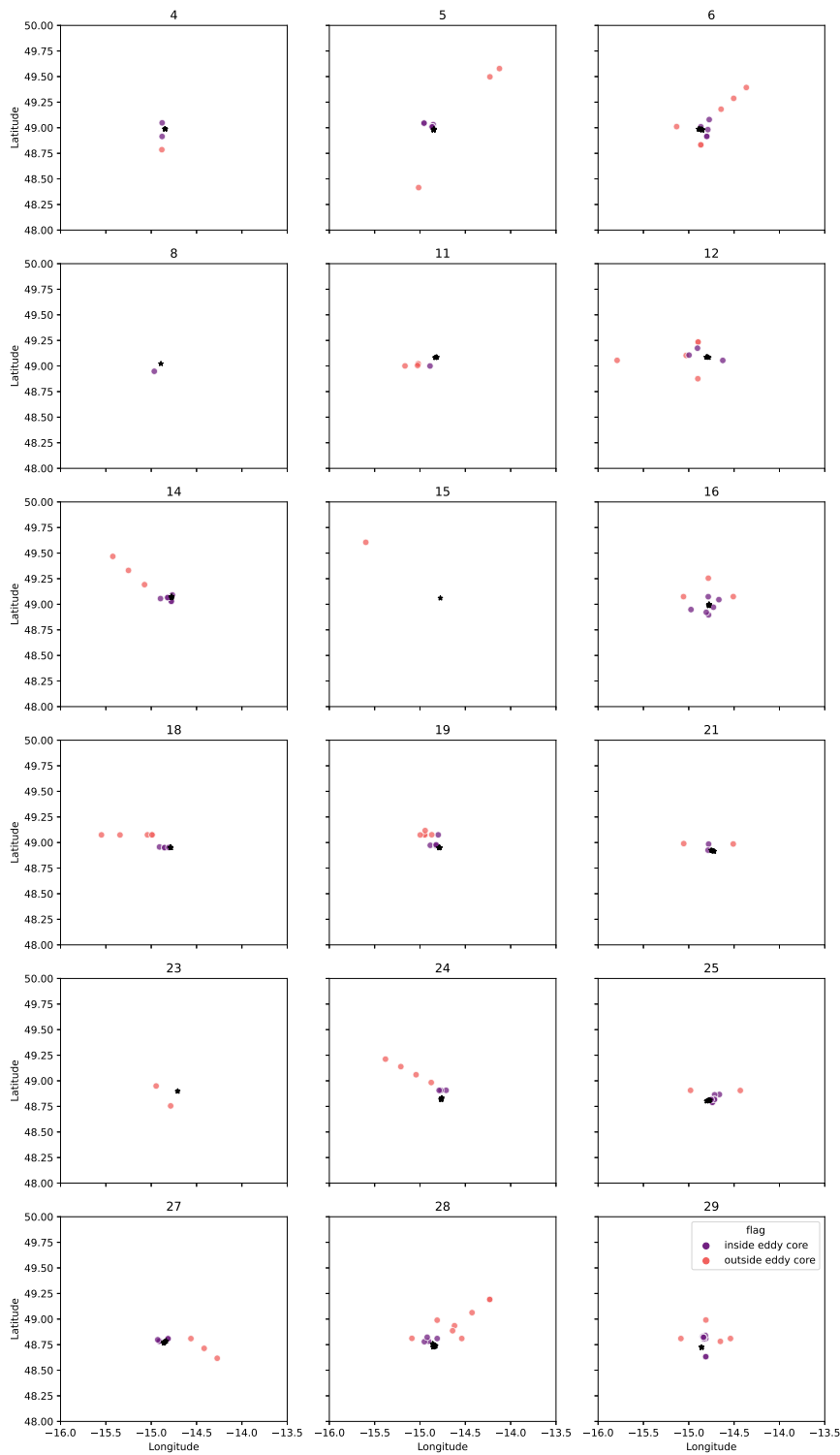


Figure S3.1: Sampling points inside ( $< 15$  km from the eddy center) and outside ( $> 15$  km) of the eddy core. The points represent the UVP5 vertical profiles done during each day of the EXPORTS 2021 campaign from May 4<sup>th</sup> to May 29<sup>th</sup> 2021.

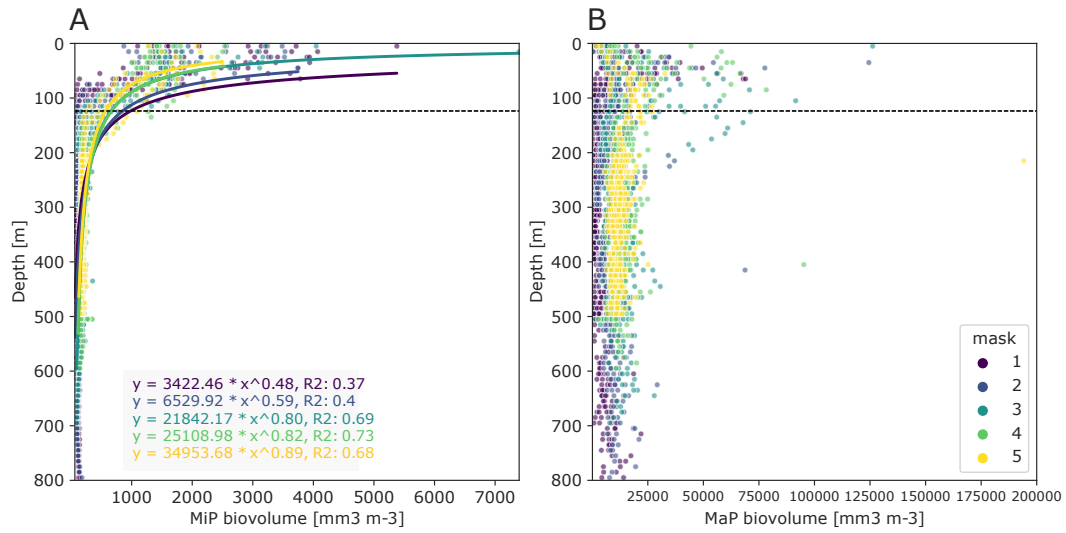


Figure S3.2: Evolution of (A) MiP and (B) MaP biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ). Each point corresponds to a profile's 20 m mean value. Power law functions were fitted to the (A) MiP biovolumes. The black horizontal dashed line corresponds to the mean depth of the SCW inside the plume. Two very high MaP biovolume values ( $417,054$  and  $476,708 \text{ mm}^3 \text{m}^{-3}$  obtained respectively at 5 on the 13<sup>th</sup> and at 345 m on the 21<sup>st</sup> were removed from this figure to make it more readable.



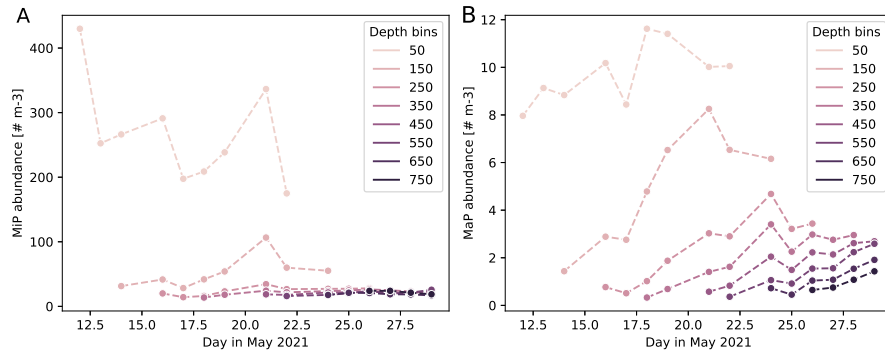


Figure S3.3: Evolution of (A) MiP and (B) MaP abundance (in numbers  $m^{-3}$ ) per 100 m depth bins. The depth bins values in the legend corresponds to the centre of the depth bin. Each point corresponds to a mean daily value for the corresponding depth bin.

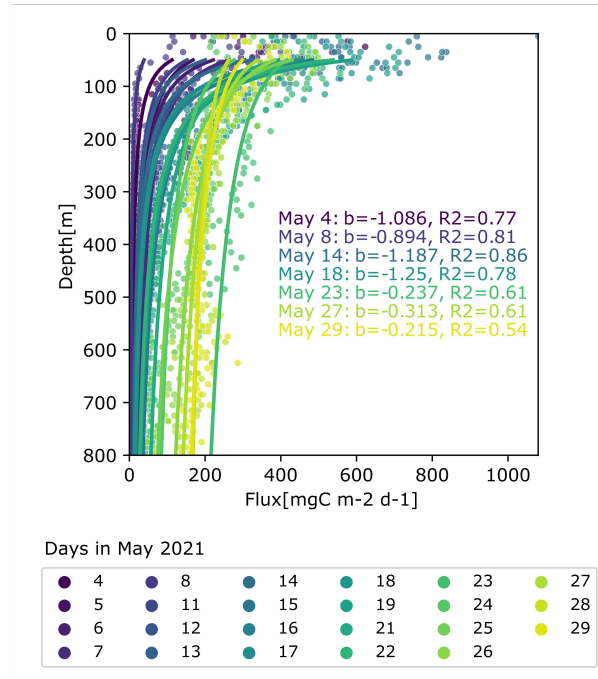


Figure S3.4: Power law regressions fitted the mean values of flux (in  $\text{mgC m}^{-2} \text{d}^{-1}$ ) computed in a vertical framework for every day of the dataset from the 4<sup>th</sup> to the 29<sup>th</sup> May 2021. Points represent the mean value per 20 m bins.

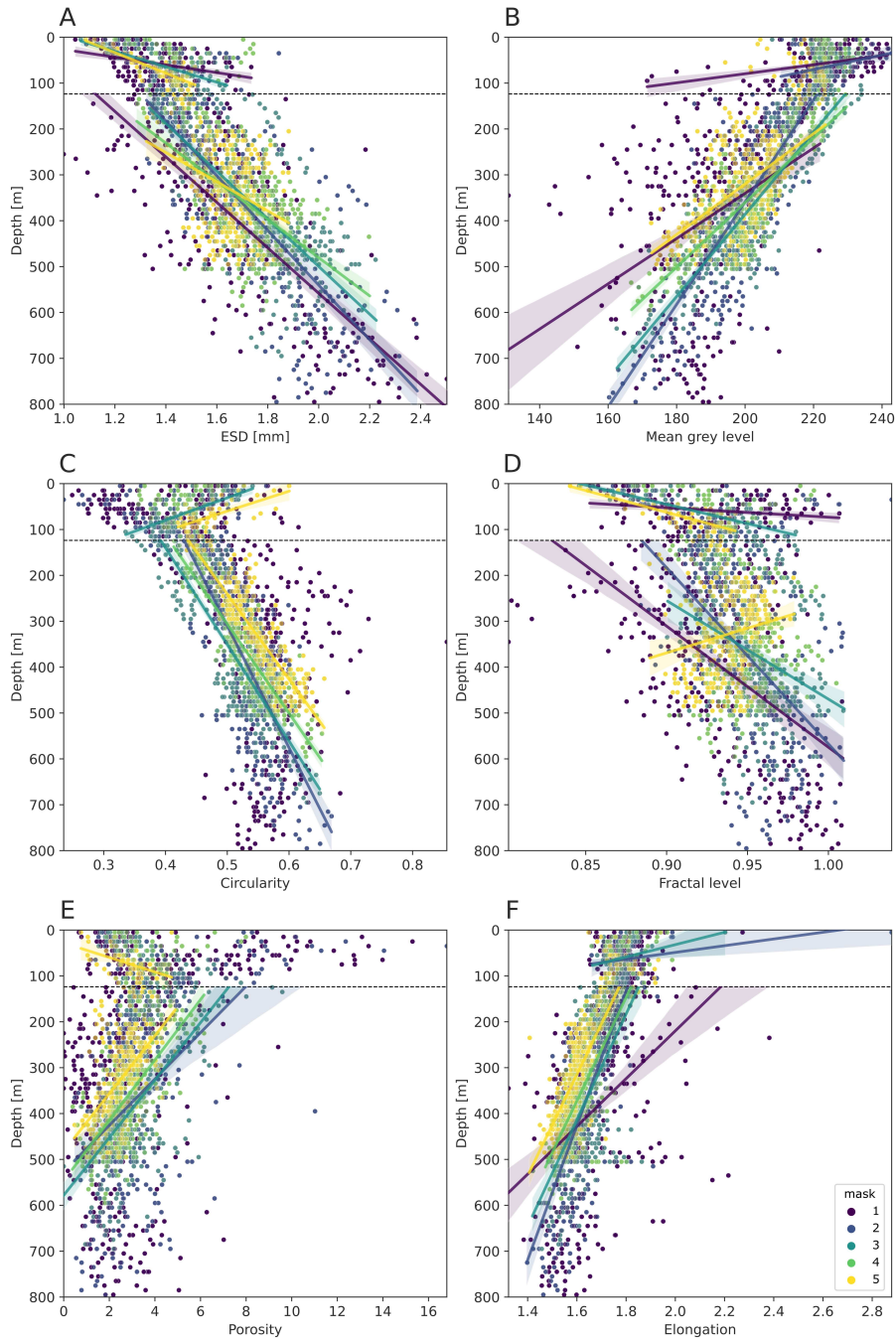


Figure S3.5: Evolution of aggregates morphological variables. Each point corresponds to a 20 m mean value for one profile. Linear regressions were fitted between 0 and SCW bottom limit and between this bottom limit and 800 m. They drawn in this figure if they had a significant slope and intercept. The corresponding slope and intercept values can be found in the Table [S3.2](#) and [S3.3](#)

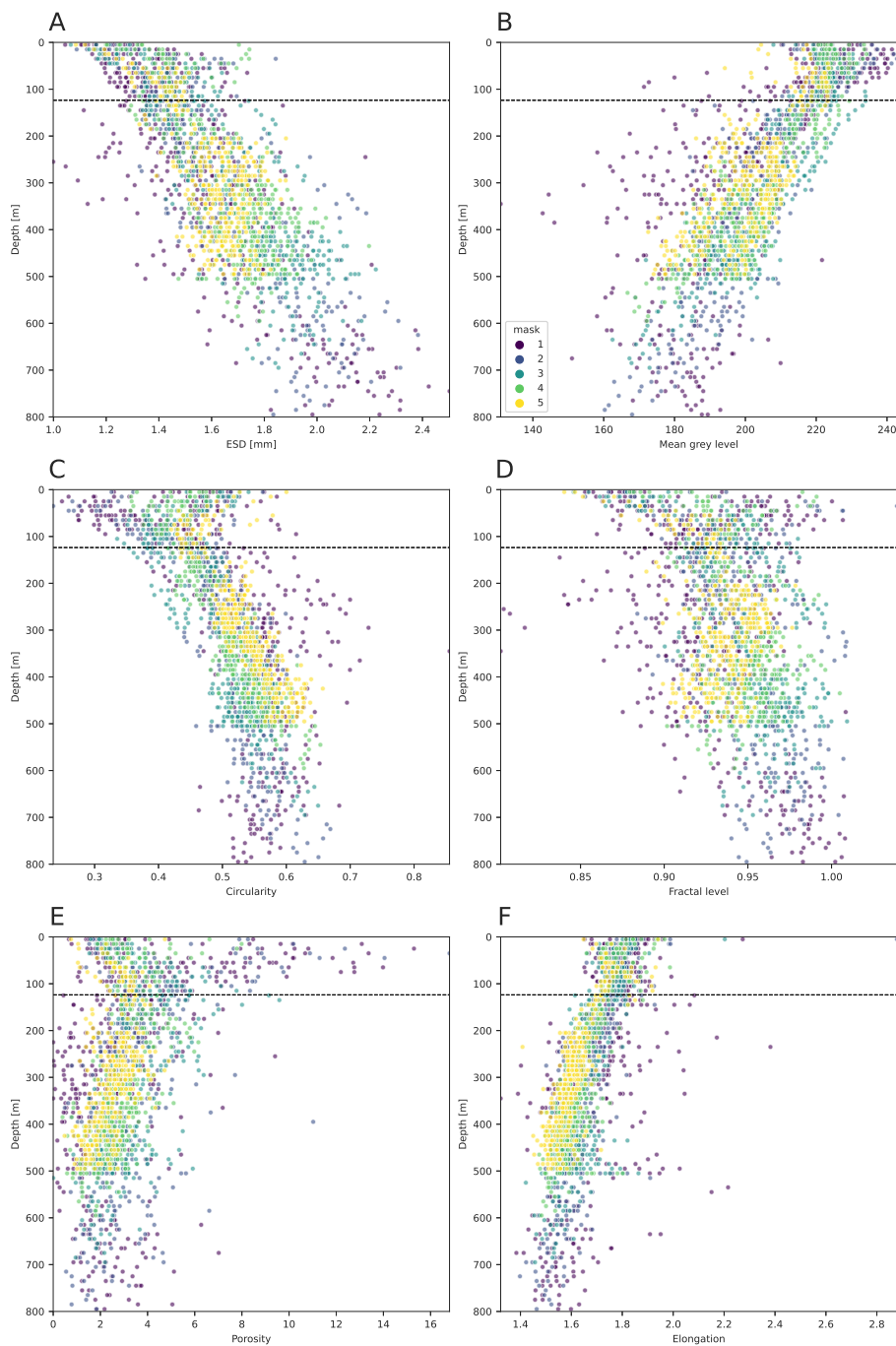


Figure S3.6: Evolution of feces morphological variables. Each point corresponds to a 20 m mean value for one profile.


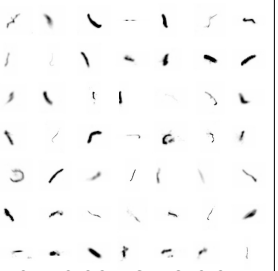

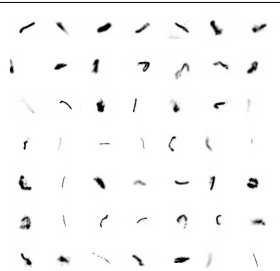
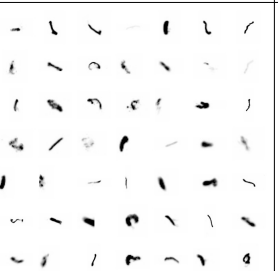
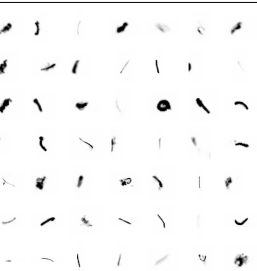
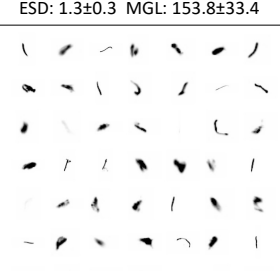
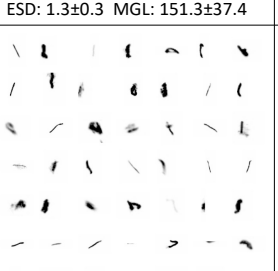
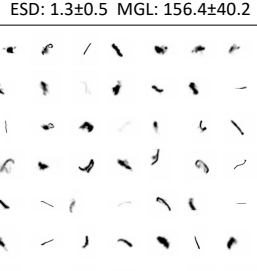
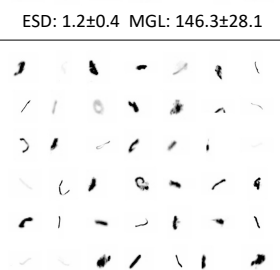
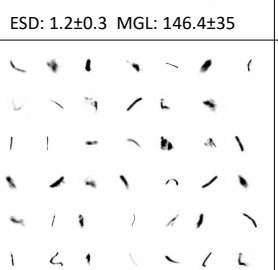
	Mask 1	Mask 3	Mask 5
0 – 100 m	 ESD: $1.2 \pm 0.3$ MGL: $169.9 \pm 37.7$	 ESD: $1.2 \pm 0.3$ MGL: $167.9 \pm 34.5$	 ESD: $1.2 \pm 0.3$ MGL: $167.9 \pm 35.6$
100 – 300 m	 ESD: $1.3 \pm 0.3$ MGL: $153.8 \pm 33.4$	 ESD: $1.3 \pm 0.3$ MGL: $151.3 \pm 37.4$	 ESD: $1.3 \pm 0.5$ MGL: $156.4 \pm 40.2$
300 – 500 m	 ESD: $1.2 \pm 0.4$ MGL: $146.3 \pm 28.1$	 ESD: $1.2 \pm 0.3$ MGL: $146.4 \pm 35$	 ESD: $1.3 \pm 0.4$ MGL: $149.7 \pm 39$
500 – 800 m	 ESD: $1.3 \pm 0.5$ MGL: $144 \pm 38.3$	 ESD: $1.2 \pm 0.4$ MGL: $150.4 \pm 33.8$	

Figure S3.7: Feces images selected randomly inside the plume for masks 1, 3 and 5 between 0-100, 100-300, 300-500 and 500-800 m.

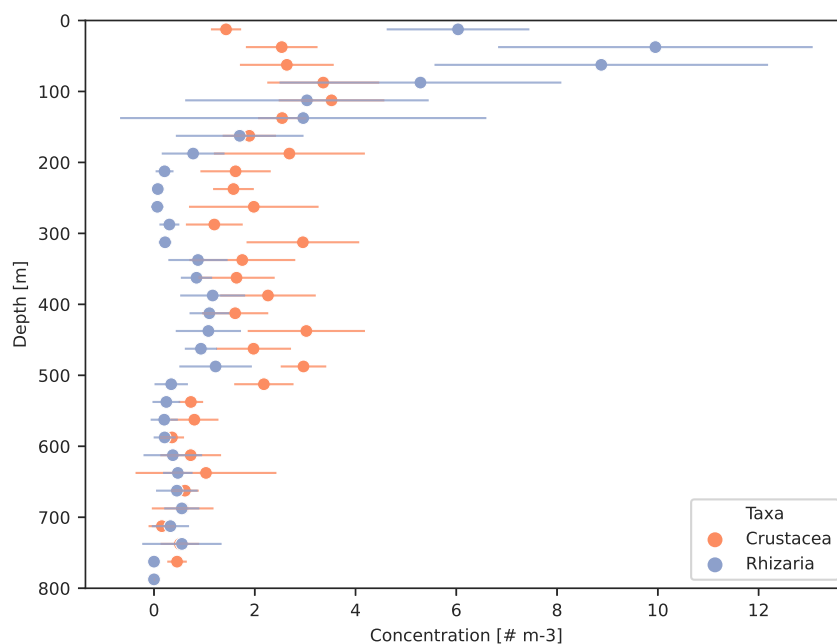


Figure S3.8: Concentration (in numbers  $\text{m}^{-3}$ ) of two zooplankton groups: Crustacea in orange and Rhizaria in purple. Horizontal bars represent the 95% confidence interval.

Table S3.1: Evolution of mean values of particles community parameters between 0 and 800 m inside the plume. The fluxes are in  $\text{mgC m}^{-2} \text{d}^{-1}$  according to Kriest, 2002 and Guidi et al., 2008. The column Depth corresponds to the center of the 50 m depth bin.

Depth	MiP abun	MaP abun	Flux Kriest2002	Flux Guidi2008	PSD slope	%fluffy
25	491.9	10.0	505.6	474.8	-3.6	0.92
75	237.1	8.9	417.7	403.2	-3.2	0.91
125	87.6	5.7	260.1	254.8	-2.9	0.90
175	43.6	3.6	175.3	172.6	-2.6	0.89
225	33.5	2.4	137.2	135.1	-2.7	0.86
275	32.0	2.3	136.8	134.9	-2.6	0.86
325	31.1	2.1	132.1	130.3	-2.6	0.86
375	29.8	2.0	135.9	134.2	-2.6	0.86
425	30.1	2.1	142.0	140.2	-2.5	0.86
475	27.7	1.7	127.3	125.6	-2.5	0.85
525	28.0	1.5	112.5	110.9	-2.6	0.83
575	22.7	1.3	112.7	111.3	-2.5	0.83
625	23.0	1.1	106.9	105.5	-2.6	0.82
675	24.8	0.9	88.9	87.3	-2.7	0.81
725	25.4	0.9	100.1	98.5	-2.6	0.84
775	22.3	1.0	109.8	108.5	-2.6	0.84

Table S3.2: Linear regressions from Figure S3.5 for the SCW

Parameter	Mask	Slope	Intercept	R <sup>2</sup>	p-value slope	p-value intercept
ESD	1	83.94	-56.71	0.13	0.00	0.02
ESD	2	52.10	-10.82	0.04	0.10	0.80
ESD	3	167.95	-170.22	0.44	0.00	0.00
ESD	4	53.35	-31.24	0.05	0.07	0.46
ESD	5	223.96	-234.71	0.63	0.00	0.00
Mean grey level	1	-0.98	276.71	0.14	0.00	0.00
Mean grey level	2	-1.58	417.28	0.10	0.01	0.00
Mean grey level	3	0.92	-142.43	0.01	0.28	0.45
Mean grey level	4	-1.43	368.20	0.03	0.18	0.12
Mean grey level	5	1.52	-259.86	0.12	0.02	0.06
Circularity	1	11.42	52.03	0.00	0.72	0.00
Circularity	2	3.32	61.22	0.00	0.95	0.01
Circularity	3	-486.97	273.09	0.46	0.00	0.00
Circularity	4	-162.87	115.11	0.03	0.18	0.03
Circularity	5	-432.86	274.94	0.41	0.00	0.00
Fractal level	1	206.64	-133.69	0.06	0.00	0.03
Fractal level	2	151.10	-78.45	0.02	0.20	0.47
Fractal level	3	839.14	-710.45	0.56	0.00	0.00
Fractal level	4	305.07	-238.36	0.06	0.05	0.09
Fractal level	5	949.30	-792.20	0.64	0.00	0.00
Porosity	1	0.07	56.38	0.00	0.92	0.00
Porosity	2	-0.57	65.75	0.00	0.67	0.00
Porosity	3	16.36	2.29	0.43	0.00	0.77
Porosity	4	0.08	45.66	0.00	0.98	0.00
Porosity	5	15.79	28.63	0.19	0.00	0.04
Elongation	1	-4.44	64.73	0.00	0.90	0.29
Elongation	2	-73.18	195.55	0.08	0.02	0.00
Elongation	3	-134.48	301.21	0.06	0.02	0.00
Elongation	4	-79.10	188.72	0.03	0.17	0.07
Elongation	5	70.90	-55.30	0.03	0.27	0.62

Table S3.3: Linear regressions from Figure S3.5 between the lower limit of the SCW and 800 m

Parameter	Mask	Slope	Intercept	R <sup>2</sup>	p-value slope	p-value intercept
ESD	1	495.75	-433.08	0.58	0.00	0.00
ESD	2	595.83	-650.92	0.63	0.00	0.00
ESD	3	519.39	-538.35	0.51	0.00	0.00
ESD	4	416.35	-351.53	0.34	0.00	0.00
ESD	5	318.11	-193.33	0.11	0.00	0.02
Mean grey level	1	-4.94	1328.42	0.17	0.00	0.00
Mean grey level	2	-11.19	2595.80	0.67	0.00	0.00
Mean grey level	3	-8.98	2180.39	0.73	0.00	0.00
Mean grey level	4	-7.06	1772.00	0.63	0.00	0.00
Mean grey level	5	-5.52	1422.63	0.37	0.00	0.00
Circularity	1	804.17	-55.07	0.07	0.00	0.56
Circularity	2	2669.00	-1025.48	0.56	0.00	0.00
Circularity	3	2112.56	-706.04	0.65	0.00	0.00
Circularity	4	1948.47	-668.37	0.76	0.00	0.00
Circularity	5	1859.72	-689.30	0.72	0.00	0.00
Fractal level	1	2647.15	-2072.23	0.29	0.00	0.00
Fractal level	2	3851.68	-3283.33	0.27	0.00	0.00
Fractal level	3	2153.71	-1683.87	0.10	0.00	0.00
Fractal level	4	782.18	-379.68	0.02	0.02	0.24
Fractal level	5	-1078.02	1337.74	0.03	0.00	0.00
Porosity	1	-11.36	420.69	0.01	0.11	0.00
Porosity	2	-50.67	528.20	0.14	0.00	0.00
Porosity	3	-62.88	579.70	0.36	0.00	0.00
Porosity	4	-65.99	548.17	0.41	0.00	0.00
Porosity	5	-63.16	483.99	0.29	0.00	0.00
Elongation	1	-521.94	1264.21	0.18	0.00	0.00
Elongation	2	-1472.19	2783.57	0.52	0.00	0.00
Elongation	3	-1160.28	2272.84	0.39	0.00	0.00
Elongation	4	-1129.95	2182.43	0.47	0.00	0.00
Elongation	5	-1107.26	2088.19	0.49	0.00	0.00



Table S3.4: Evolution of mean values of particles morphological parameters between 0 and 800 m inside the plume. The column depth corresponds to the center of the 50 m depth bin.

Type	Depth	ESD	mean_grey	circ	fractal	porosity	elongation
aggregate	25	1.3	225.1	0.4	0.9	4.2	1.8
aggregate	75	1.4	222.3	0.4	0.9	4.6	1.8
aggregate	125	1.4	216.3	0.4	0.9	4.1	1.8
aggregate	175	1.5	210.8	0.5	0.9	3.7	1.7
aggregate	225	1.6	203.9	0.5	0.9	3.2	1.7
aggregate	275	1.6	202.2	0.5	0.9	3.1	1.7
aggregate	325	1.7	198.2	0.5	0.9	2.7	1.6
aggregate	375	1.7	195.3	0.5	0.9	2.5	1.6
aggregate	425	1.7	193.9	0.6	0.9	2.3	1.6
aggregate	475	1.8	190.8	0.6	0.9	2.4	1.6
aggregate	525	1.8	188.1	0.6	0.9	2.2	1.6
aggregate	575	1.9	183.9	0.6	1.0	2.4	1.6
aggregate	625	2.0	181.4	0.6	1.0	2.3	1.6
aggregate	675	2.0	180.9	0.6	1.0	2.5	1.5
aggregate	725	2.1	179.3	0.6	1.0	2.4	1.5
aggregate	775	2.1	177.7	0.6	1.0	2.1	1.5
feces	25	1.2	172.6	0.5	0.9	0.9	3.1
feces	75	1.3	166.0	0.5	0.9	0.6	3.1
feces	125	1.4	162.4	0.5	0.9	1.3	2.9
feces	175	1.3	159.9	0.5	0.9	1.0	2.8
feces	225	1.3	155.9	0.5	0.9	0.9	2.8
feces	275	1.3	152.8	0.5	0.9	0.6	3.0
feces	325	1.4	150.5	0.5	0.9	0.5	3.2
feces	375	1.3	153.4	0.5	0.9	0.5	3.3
feces	425	1.3	153.8	0.5	0.9	0.5	3.5
feces	475	1.4	150.9	0.5	0.9	0.3	3.4
feces	525	1.3	152.1	0.5	0.9	0.7	3.3
feces	575	1.3	143.8	0.6	0.9	0.8	2.9
feces	625	1.4	150.3	0.6	0.9	2.7	2.8
feces	675	1.5	146.0	0.6	0.9	1.2	2.3
feces	725	1.4	157.2	0.5	0.9	0.9	3.8
feces	775	1.3	139.7	0.5	0.9	3.0	3.4



## Part IV

### Estimating carbon flux from UVP data: an EXPORTS perspective

Laetitia Drago<sup>1</sup>, Ken Busseler<sup>2</sup>, Elena Ceballos Romero<sup>2</sup>, Lee Karp-Boss<sup>3</sup>, Andrew M. P McDonnell<sup>4</sup>, Simon-Martin Schröder<sup>5</sup>, David A. Siegel<sup>6</sup>, Lars Stemann<sup>1†</sup> and Rainer Kiko<sup>1,7†</sup>

<sup>1</sup> Sorbonne Université, Laboratoire d'Océanographie de Villefranche-sur-mer, Villefranche-sur-mer, France,

<sup>2</sup> Woods Hole Oceanographic Institution, Woods Hole, MA, USA.

<sup>3</sup> School of Marine Sciences, University of Maine, Orono, ME, United States,

<sup>4</sup> Oceanography Department, University of Alaska Fairbanks, Fairbanks, AK, United States,

<sup>5</sup> MIP, Computer Science, Kiel University, Kiel, Germany

<sup>6</sup> Earth Research Institute, University of California, Santa Barbara, California, USA

<sup>7</sup> GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstr. 1-3, 24148 Kiel, Germany

† These authors share last authorship



## EXPORTS carbon flux

---

### 4.1 Abstract

The EXPORTS experiment conducted in the North Atlantic aimed to compute and compare the flux resulting from the North Atlantic spring bloom utilizing both sediment traps and high-resolution Underwater Vision Profiler 5 (UVP5) units. This study focused on the composition of particles in the eddy core, revealing a significant predominance of aggregates, particularly in surface waters. The data highlighted the large predominance of aggregates (97.8% of the images between the surface and 1000 m depth) during this North Atlantic bloom event.

Particle sinking speed was computed by analysing the propagation of the export plume to depth. The sinking speed of both aggregates and fecal pellets exhibited a positive size-sinking speed relationship, consistent with previous research. Differences between these two categories were not significant. Specifically, aggregates between 0.81 and 3.25 mm displayed increasing sinking speeds, with values ranging from 33.9 to 45.91 m d<sup>-1</sup>. This relationship supported the notion that larger particles tend to sink more rapidly in marine environments. However, challenges arose when dealing with smaller particles, as limitations in image analysis and UVP counts hindered precise sinking speed estimation for this size fraction.

Computing particle fluxes using previously published relationships emphasized the need for context-specific methods. Previous research showed that the neutrally buoyant sediment traps (NBST) and surface tethered sediment traps (STT) deployed at EXPORTS quantitatively sample flux in the size range 10  $\mu$ m to 2 mm. We find the best fit between sediment trap data (obtained by colleagues) and UVP estimates when the UVP size range is extended to 10  $\mu$ m, constrained to 2 mm. A hybrid approach using both the UVP counts and the aggregates and feces images to calculate the flux was tested in this study, but it did not give any better results than the one carried out on the UVP counts alone. Our results underscore the importance of matching size ranges observed by UVP and sediment traps for an accurate comparison of UVP-derived flux data with sediment trap measurements. In future research, extending sinking speed estimation to smaller particles presents an avenue for enhancing our understanding of particle dynamics. In addition, we found that 30-46% of the flux derived from the UVP origi-

nated from particles between 2 and 16 mm in size and may be missed by the NBST and STT traps.

## 4.2 Introduction

The biological carbon pump is the process by which particulate organic carbon (POC) is exported from the surface to depth, primarily by the biological gravitational pump (Boyd et al., 2019; Nowicki et al., 2022). This pathway corresponds to the passive sinking of POC as marine detritus consisting of aggregates and fecal pellets (Alldredge and Gotschalk, 1988; Boyd et al., 2019; Turner, 2015) and is recognised as the major contributor to net carbon export globally (Boyd et al., 2019; Nowicki et al., 2022), accounting for 56.1-70% (Nowicki et al., 2022; Stukel et al., 2022). The study of aggregates and fecal pellets is complex as they can be altered by a variety of physical (Alldredge, 2001; Kiørboe, 2001; Stemmann, Jackson, and Ianson, 2004) and biological (Gillard et al., 2022; Iversen, 2023; Steinberg and Landry, 2017; Stemmann, Jackson, and Ianson, 2004) processes. They have been extensively examined using various instruments, such as sediment traps (Bach et al., 2019; Collins et al., 2015; Lampitt et al., 1993) and marine snow catchers (Belcher et al., 2016; Durkin et al., 2021). Though these instruments can sample aggregates and allow for their laboratory-controlled study (Ionescu et al., 2015; Ploug and Jørgensen, 1999; Song and Rau, 2022), they are limited in their spatiotemporal coverage and their ability to collect without breaking the fragile marine particles.

More recently, *in situ* imaging instruments have been deployed around the world's oceans to study particles across a wide size range (Lombard et al., 2019). Among these tools is the Underwater Vision Profiler (UVP, Picheral et al. (2010, 2021)) which was specifically designed for the study of the distribution of marine particles. Previous studies have used this instrument to inform on particle distribution at global (Clements et al., 2022, 2023; Guidi et al., 2008, 2015), and regional scales (e.g. Kiko et al. (2017)) without any attempt to identify the nature of particles. A recent publication presented the extensive and unique data set gathered over the years using UVP5 instruments through an international collaboration (Kiko et al., 2022). In another recent study where particle images were sorted, the usefulness of particle classification was demonstrated by the observed change in particle type during an Arctic bloom (Trudnowska et al., 2021).

In May 2021, our study investigated an anticyclonic retentive eddy located in the North Atlantic close to the Porcupine Abyssal Plain (PAP) station within the framework of the EXport Processes in the Ocean from

RemoTe Sensing (EXPORTS) initiative (Siegel et al., 2016, 2021). This unique experiment was conducted in an international collaboration to investigate the dynamics of the biological carbon pump during the North Atlantic spring bloom (Omand et al., 2015) using instruments deployed from three research vessels (RRS Discovery DY130, RRS James Cook JC214, and R/V Sarmiento de Gamboa SG2105), alongside satellite remote sensing and modelling efforts (Johnson et al., 2023). 4.3 million images were sorted to determine which type of particle dominated the flux and to assess whether the nature of the particle community affected the flux. Based on the 26-day UVP5 deployments done during this scientific campaign we investigate in more detail how gravitational flux estimates can be calculated based on UVP5 data that match observed sediment trap measurements of gravitational flux, acknowledging the advantages of the UVP5 in providing higher spatial and temporal resolution.

We first provide a spatiotemporal analysis of the composition of the particle community within the top 1,000 m of the water column. We then propose a comparison method between UVP5 and sediment trap flux, using an adapted UVP5 size spectrum and various previously published relationships. This comparison relies on UVP5 counts and a hybrid method using both counts as well as aggregates and fecal pellet images clustered using a semi-autonomous approach (Schröder et al., 2020). Finally, we discuss the optimal parameter combination and the impact of the various methods applied to obtain particle flux estimations.

## 4.3 Material and Methods

### 4.3.1 Data collection

The UVP5 was designed to detect, measure, and quantify the vertical distribution of zooplankton and marine particles (Picheral et al., 2010). It counts particles with a size of 100  $\mu\text{m}$  or larger during its descent and records images of zooplankton and aggregates larger than 600  $\mu\text{m}$ . Detailed operational information can be found in Picheral et al. (2010), and data processing details are in Kiko et al. (2022). During the EXPORTS experiment, 173 vertical profiles were obtained from three intercalibrated UVP5 instruments deployed each on a CTD rosette from the research vessels RRS Discovery DY130, RRS James Cook JC214, and R/V Sarmiento de Gamboa SG2105. Metadata and particle counts were collected from the Ecopart file on Ecotaxa, as described by Picheral et al. (2017).

The EXPORTS experiment targeted the spring bloom inside a North Atlantic eddy chosen according to criteria established well in advance of the field campaign to select a stable eddy (Erickson et al., 2022). This small anticyclonic eddy with a core diameter of approximately 30 km was identified using altimetry data, numerical modelling, ocean gliders, and SeaGliders (Johnson et al., 2023). It was highly retentive, enabling the study of carbon export flux within the eddy core that was characterized as the central 15 km of the eddy.

#### 4.3.1.1 Data processing

Profiles done in the eddy core were kept for further analysis. As only 6% of the profiles covered the 800-1000 m, they were only kept for data set composition analysis and not for flux calculations, which were done down to 800 m. These images and metadata (geographical position, date, etc.) were first processed with Zooprocess and then clustered with Morphocluster (Schröder et al., 2020), a program designed to annotate large image datasets. The clustered images were then manually validated or reclassified using the Ecotaxa program (Picheral et al., 2017). After removing the artefacts, mainly caused by bubbles and blurred images, the dataset consisted of 1,720,914 images. Concentration in numbers (numbers  $\text{m}^{-3}$ ) and biovolume ( $\text{mm}^3 \text{m}^{-3}$ ) using both UVP5 particle counts and imaged particles were computed per 5 m bins for each profile with UVP5 water volume data in  $\text{m}^3$ .

To study the evolution of imaged aggregates and feces concentration (numbers  $\text{m}^{-3}$ ) and biovolume ( $\text{mm}^3 \text{m}^{-3}$ ), we grouped the images per layer using the following vertical definition: 0-100, 100-300, 300-500, 500-800, 800-1000 m. Percentages in aggregates and feces over the whole dataset were computed using the same vertical definition.

#### 4.3.1.2 Sinking speeds

To estimate the sinking speed of imaged particles, aggregates and feces we employed a similar method as the one described by Briggs et al. (2020). We used a Gaussian curve fitting technique to analyze the distribution of the daily median concentration profiles of different aggregate size classes within 100 m depth bins and to obtain a robust estimate of the date at which the concentration peaked at a given depth. This same methodology was applied to the particle counts obtained from UVP5 data. For each size class, we then conducted a linear regression analysis on the depth and time coordinates of the Gaussian fits' maximum values. The resulting slope from this regression analysis as well as its 95% confidence interval served as an indicator of the sinking speed, provided that the Gaussian fit was successful for at least half of the 100 m depth bins. When possible, a linear regression was performed on the sinking speed values according to ESD. These



sinking speed-to-size estimates were then compared to relationships found in the literature (Fig. 4.3).

Our estimations were not successful for all size classes but show that relationship 9 from Table 2 of Kriest (2002) was the most appropriate for our dataset. This method did not work for small classes or for categories of objects that were too rare.

#### 4.3.1.3 Flux estimation

In general, two methods were employed to estimate flux in different size ranges using the UVP5 data: 1) from particle counts without consideration of particle identity, and 2) hybrid method from particle counts between 10  $\mu\text{m}$  and 1 mm as well as from imaged aggregates and feces in the 1 to 16.4 mm size range. UVP5 particle counts are generated for the size range 102  $\mu\text{m}$  to 16.4 mm. However, the neutrally buoyant sediment traps (NBST, 12cm diameter opening) and surface-tethered traps (STT, 7.5cm diameter opening) used during the EXPORTS experiment covered a size range of approximately 10  $\mu\text{m}$ -2 mm (Durkin et al. (2021); personal communication Durkin, 2023; sediment trap data available on SeaBASS, see Data Availability section). To compare the UVP5 and sediment traps flux estimates as precisely as possible, we therefore also estimated particle abundances in the 10 to 100  $\mu\text{m}$  size range by applying a linear regression on UVP5 counts between 0.102-3.25 mm following Baudena et al. (in review) (Fig. S4.1). Several metrics were computed on the comparison between sediment trap flux (obtained by colleagues) and UVP flux: coefficient of determination ( $R^2$ ), the root mean square error (RMSE), the mean bias error (MBE) and the slope. The slope and  $R^2$  were computed by doing a linear regression between the sediment trap and UVP fluxes. The MBE was computed as the mean difference between UVP5 and sediment trap flux. It has a positive value if UVP5 overestimate the sediment trap flux on average and a negative value if it underestimates it on average.

We first computed POC flux on particle counts following previously published relationships between flux and size. Two types of approaches were present in the literature: 1) methods separating the size-carbon content and/or size-sinking speed relationships (such as Alldredge (1998), Durkin et al. (2021), and Kriest (2002)); 2) methods that defined size-flux relationship based on a minimization procedure between UVP and sediment trap data (such Clements et al. (2023), Fender et al. (2019), Forest et al. (2013), and Guidi et al. (2008)). Table 4.1 presents A and b couples published in these previous studies.

The sinking flux of each size class was computed as follows:  $A \times \text{ESD}^b$  ( $\text{mgC m d}^{-1}$ ) with ESD in mm for all the previous studies except for Kriest (2002) for which ESD was in cm. While A represents the mass

Source	A	b
Kriest (2002)	2.865	2.24
Guidi et al. (2008)	12.5	3.81
Forest et al. (2013)	28.1	2
Fender et al. (2019)	15.4	1.05
Clements et al. (2023)	18	2.63

Table 4.1: Values of A and b used to compute flux from previous studies (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008; Kriest, 2002). Values labelled from Kriest (2002) were obtained following the method from Kiko et al. (2017).

flux of particles of the smallest size, b is related to the fractal dimension of aggregates (D) with  $D = (B + 1)/2$  (Guidi et al., 2008).

POC flux ( $\text{mgC m}^{-2} \text{d}^{-1}$ ) was obtained by multiplying the sinking flux with the particle concentration in each size class (in particles  $\text{m}^{-3}$ ) and by integrating over all size classes. This computation was done for five size ranges: all size classes from 10  $\mu\text{m}$  - 16.4 mm, between 10 - 100 $\mu\text{m}$ , 10 $\mu\text{m}$  - 1 mm, 10  $\mu\text{m}$  - 2 mm and between 2 - 16.4 mm. These size ranges were chosen to allow a quantitative comparison of UVP flux and sediment trap flux in the size range covered by the sediment traps (10  $\mu\text{m}$  - 2 mm), to enable the combination of particle count and image-derived flux (10  $\mu\text{m}$  - 1 mm), and to diagnose the contribution of the flux not "seen" by the UVP5 (10 - 100 $\mu\text{m}$ ) and the sediment traps (2 - 16.4 mm).

POC flux was also computed for imaged aggregates and feces > 1 mm ESD. Individual aggregate and feces carbon content were estimated using Durkin et al. (2021) and Alldredge (1998) while the sinking speed for aggregates or feces was estimated using the Kriest (2002) (reference 9 of Table 2) size to sinking speed relationship, as this showed highest consistency with our estimates (Fig. 4.3).

The individual carbon content and particle sinking speeds were then multiplied to obtain the individual flux. The flux per 5 m depth bins for each UVP5 profile was obtained by adding up the individual particle flux values in the corresponding depth bins.

#### 4.3.1.4 UVP5 and sediment trap match

To compare flux estimates obtained from UVP5 and sediment traps, a spatio-temporal match-up was done between the deployments of these two instruments. For each sediment trap deployment in the eddy core, we computed the mean UVP5 flux in a 20 m layer above the sediment

Source	Category	Volume	A	b
Allredge1998	Aggregate	$8 \times 10^{-4} \times \text{ESD}^{-2.3}$	0.99	0.52
Allredge1998	Fecal Pellets	$8 \times 10^{-4} \times \text{ESD}^{-2.3}$	1.05	0.51
Durkin2021	Aggregate	Sphere	$0.1 \times 10^{-9}$	0.8
Durkin2021	Dense detritus	Sphere	$0.1 \times 10^{-9}$	0.83
Durkin2021	Large loose fecal pellet	Cylinder	$0.1 \times 10^{-9}$	0.83
Durkin2021	Long cylindrical fecal pellet	Cylinder	$0.1 \times 10^{-9}$	1

Table 4.2: Carbon content estimates from previous studies (Allredge, 1998; Durkin et al., 2021). Allredge (1998) uses biovolume in  $\text{mm}^3$  while Durkin et al. (2021) uses biovolume in  $\mu\text{m}^3$ . The sphere biovolume is  $V_{\text{sphere}} = \frac{4}{3} \times \pi \times \left(\frac{\text{ESD}}{2}\right)^3$  and  $V_{\text{cylinder}} = l \times \pi \times \left(\frac{w}{2}\right)^2$  with  $w = \frac{553 \times \text{ESD}}{\text{ESD} + 996}$  and  $l = \frac{\pi \left(\frac{\text{ESD}}{2}\right)^2}{w}$ .

trap deployment depth for all eddy core UVP5 profiles conducted between the opening and closing of the respective trap.

This allowed for a comparison between sediment traps and the two types of UVP5 flux obtained from either particle counts (10  $\mu\text{m}$ -2 mm) or images (1-2 mm) supplemented by particle counts (10  $\mu\text{m}$ -1 mm).

## 4.4 Results

### 4.4.1 Data composition

The detrital particle community inside the eddy core is largely dominated by detritus comprising 99.27% and only 0.73% living organisms. For the detritus, 97.8% of the images were labelled as aggregates (88% fluffy and 9.8% dense) and 1% as fecal pellets.

Looking at the aggregates which make up most of the data, we observe an increase in their concentration (Fig. 4.1A) in the 0-100 m layer from 796 aggregates per  $\text{m}^3$  on May 5<sup>th</sup> to 6027 aggregates per  $\text{m}^3$  on May 18<sup>th</sup>, which represents a 7.6-fold increase. This pattern of increased concentration is reflected progressively in deeper layers as the campaign progresses, with less intensity as we move down the water column. Aggregate concentrations go down after reaching a peak on May 18<sup>th</sup> for 0-100 m and on May 21<sup>st</sup> for 100-300 m. The concentration is reduced by a factor of 2.1 between the two peaks (from 6027 to 2844 aggregates per  $\text{m}^3$ )

A similar pattern is observed for the biovolume (Fig. 4.1C) even though the values obtained by the end of the cruise are much more similar than those observed for the concentration.

Fecal pellets also present a pattern of an important increase in concentration (Fig. 4.1B), especially after the first storm when their number increases by a factor of 3, going from 11-13 feces per  $\text{m}^3$  between May 4<sup>th</sup> to 7<sup>th</sup> to 34 feces per  $\text{m}^3$  on May 11<sup>th</sup>. A maximal value is observed for all three layers between 100 and 500 m between May 27<sup>th</sup>. The pattern observed for the concentration is also mirrored in the biovolume for fecal pellets (Fig. 4.1D).

The large domination of aggregates in the dataset is striking (Figure 4.2A), as all layers are found to present a percentage consistently superior to 80%. This predominance becomes even more evident following the initial storm in the uppermost 100 metres, continuing in the lower layers to nearly 100% in all of them except the lowest one that reaches 90%.

Fecal pellets, however, make up a small percentage of the global dataset. This contribution decreases over time.

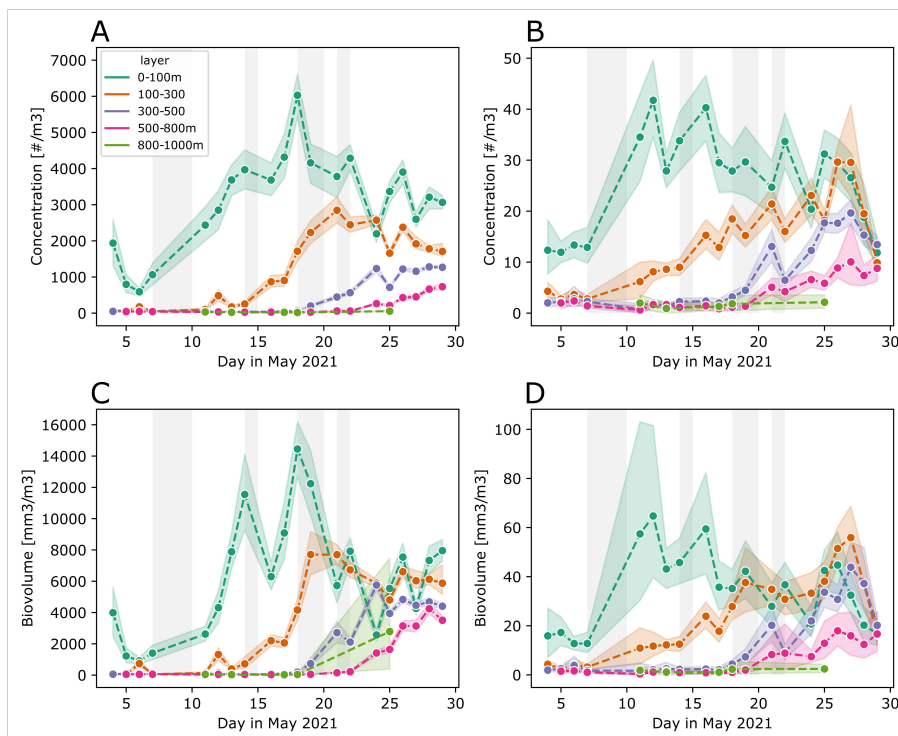


Figure 4.1: Evolution of concentration (in number  $\text{m}^{-3}$ ) and biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ) of (A, C) aggregates and (B, D) feces for 0-100 m, 100-300 m, 300-500 m, 500-800 m. The mean concentration is represented by points and linked with a dashed line. The shaded areas represent the 95% confidence interval for the dashed line. The grey-coloured vertical bars correspond to the storms.

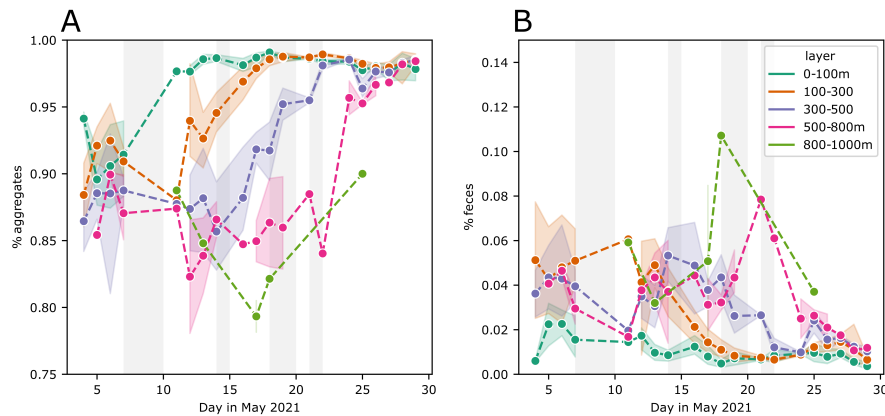


Figure 4.2: Evolution of concentration (in number  $\text{m}^{-3}$ ) and biovolume (in  $\text{mm}^3 \text{m}^{-3}$ ) of (A, C) aggregates and (B, D) feces for 0-100 m, 100-300 m, 300-500 m, 500-800 m. The mean concentration is represented by points and linked with a dashed line. The shaded areas represent the 95% confidence interval for the dashed line.

#### 4.4.2 Sinking speed and flux estimations

Aggregates between 0.81 and 3.25 mm exhibited an increasing sinking speed according to ESD from 33.9 to 45.91  $\text{m d}^{-1}$  with a significant linear regression following the equation  $y = 7.94x + 29.31$  (Fig. 4.3). For fecal pellets, our method worked for two sizes classes between 1.02-1.29 mm as well as 1.29-1.63 mm, yielding a mean sinking speed of 36.13 and 36.71  $\text{m d}^{-1}$  respectively. The size-to-sinking speed relationship Kriest (2002) (reference 9 Table 2) shows a steeper slope than the relationship we established, but is the only published relationship that coincides (within uncertainties) with the aggregate, feces and particle sinking speeds we could establish. The relationship from Alldredge and Gotschalk (1988) has a similar slope as our fitted regression line for aggregates, but generally overestimates the sinking speeds, particularly at small particle sizes, whereas Kriest (2002) (reference 8 Table 2) generally predicts too low sinking speeds. Compared to our results, Iversen et al. (2010) and Cael et al. (2021) predict too high sinking speeds, whereas Guidi et al. (2008) predicts a too steep slope of the size to sinking speed relationship.

#### 4.4.3 Flux comparison between UVP5 and sediment traps

As mentioned above, the sediment traps used during EXPORTS 2021 covered a size range from 10  $\mu\text{m}$  to 2 mm. We hence used UVP5 flux estimations extended to a lower bound of 10  $\mu\text{m}$  and restricted to an upper bound of 2 mm to compare UVP5-based flux estimates to

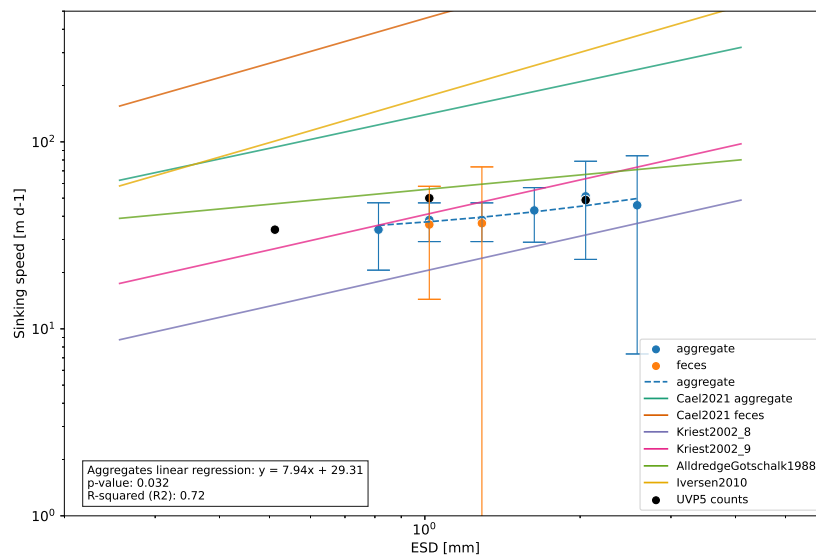


Figure 4.3: Estimation of sinking speed (in  $\text{m d}^{-1}$ ) according to the Equivalent Spherical Diameter (ESD in mm). The circular dots represent the estimations of mean sinking speed done on UVP5 images for aggregates (in blue) and feces (in orange) with the 95% confident interval represented by the vertical error bars and on UVP5 counts (in black). The dashed blue line corresponds to the linear regression done on the imaged aggregates sinking speed. Continuous lines correspond to relationships found in the literature in Alldredge and Gotschalk (1988), Cael et al. (2021), Iversen et al. (2010), and Krist (2002).

sediment trap flux estimates. The flux estimation from UVP5's first method produced the findings displayed in Figure 4A through F. We use four factors to identify the best-suited fit for our data, namely the slope of the linear regression that should be as close to 1 as feasible, the highest possible  $R^2$ , as well as a small Root Mean Square Error (RMSE) and a small Mean Bias Error (MBE).

Apart from Fender et al. (2019) for which the comparison shows a very bad matchup to sediment trap flux (as shown by the high RMSE and MBE), the results show  $R^2$  values between 0.52 for Forest et al. (2013) and 0.73 for Guidi et al. (2008). Slope values are between 0.43 for Guidi et al. (2008) and 0.76 for Forest et al. (2013). The results obtained using Guidi et al. (2008)'s A and b show much lower flux estimates for a group of UVP-sediment trap matchups (Fig. 4.4C). These are located in the lower range of flux estimates from the sediment traps as illustrated by the negative MBE. Using Forest et al. (2013) A and b values leads to

an overestimation of the flux for the UVP compared to the sediment trap with an RMSE of 219.03 and an MBE of -111.24 (Fig. 4.4D). Using Kriest (2002) and Clements et al. (2023) leads to the best results with a slope value of 0.69 and 0.57, an  $R^2$  of 0.64 and 0.71 respectively and low values of MBE (8.93 and 1.03 respectively). With the obtained metrics, Kriest (2002) seems to be the most appropriate algorithm to convert particle size information to flux for our data.

Using the Kriest (2002) method, the flux computed on the size spectrum covered by the sediment traps (10  $\mu\text{m}$ -2 mm) produces superior results compared to the same method applied across the entire size spectrum (0.128-16.4 mm). Although a higher  $R^2$  is obtained for the whole spectrum (0.71 compared to 0.64), this is merely a reflection of the linear regression between UVP flux and sediment trap. The three other metrics offer more valuable insight into the optimal correlation between UVP and sediment flux. They are more suitable for the size spectrum that the sediment trap covers than for the entire size spectrum. We obtained a lower RMSE of 100.72 versus 161.96, a lower MBE of 8.93 versus 47.79 and a steeper slope of 0.69 versus 0.5.

The hybrid method, which involves utilising both UVP<sub>5</sub> counts (10  $\mu\text{m}$  to 1 mm) and images (1 mm and larger), results in reasonable results using Kriest (2002) for the size-sinking speed relationship and either Durkin et al. (2021) (Fig. 4.4G) or Alldredge (1998) (Fig. 4.4H) for the size-carbon relationship. They both present a high  $R^2$  (0.67 and 0.73) but the three other metrics indicate that the use of Durkin et al. (2021)'s size-to-carbon content relationships yielded the best results. Indeed, it gets a lower RMSE of 6.01 versus 209.82 and a steeper slope of 0.62 versus 0.51. The MBE indicates that this carbon-size relationship also leads to an underestimation of the flux compared to the sediment trap with a value of -42.12 while using Alldredge (1998) leads to an overestimation of the flux with an MBE of 120.5.

#### 4.4.4 Contribution of different size range compartments to export flux

Depending on the method used to compute the flux between 10  $\mu\text{m}$ -16.4 mm, the variability in terms of representation by the 3 size ranges 10-100  $\mu\text{m}$  (small), 100  $\mu\text{m}$ -2 mm (medium) and 2-16.4 mm (large) varies a lot. Figure 4.5 illustrates this with a median distribution of the flux being mostly in the small and medium size classes using Kriest (2002) and Forest et al. (2013) while it is shared between large and medium for Clements et al. (2023) and almost completely in the large one for Guidi et al. (2008). The contributions to total flux by the extrapolated 10 to 100  $\mu\text{m}$  size range in the case of Kriest (2002) and Clements et al. (2023) are

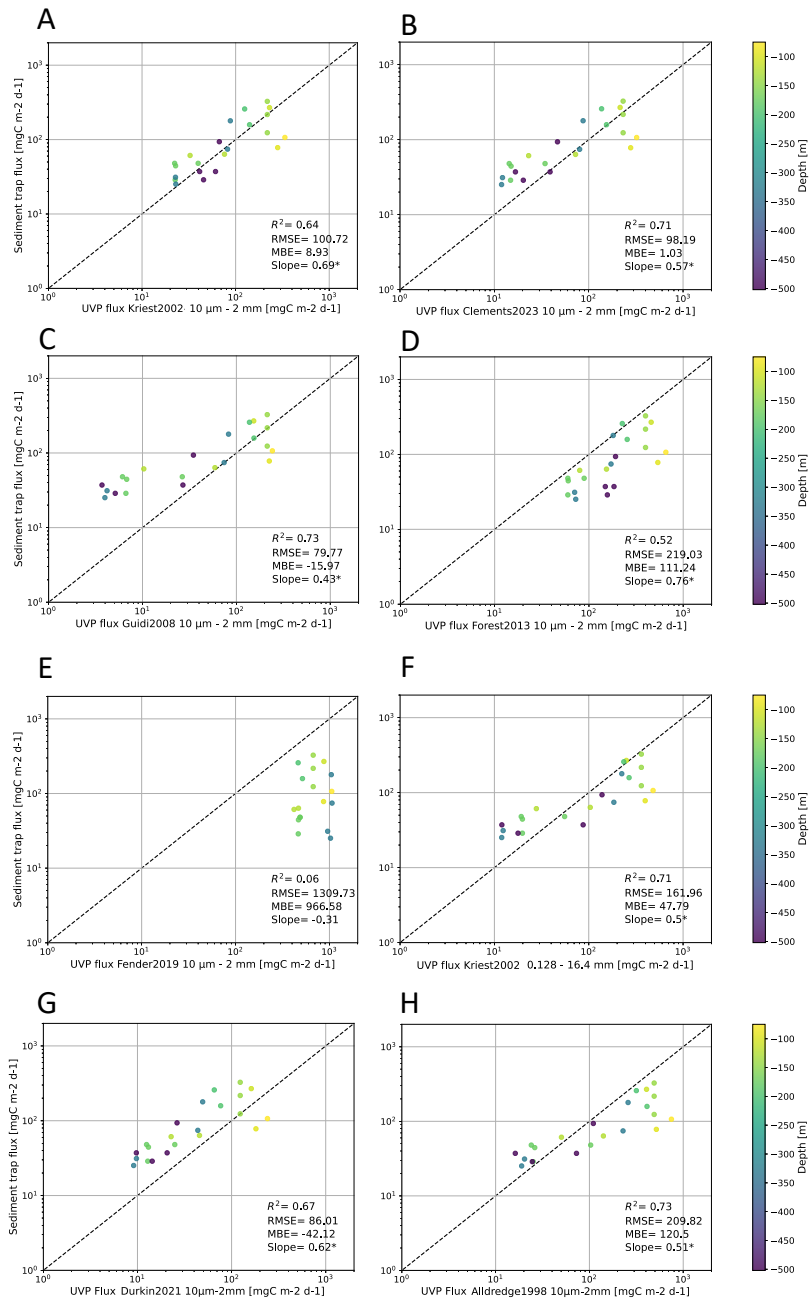


Figure 4.4: Comparison sediment trap and UVP flux estimations (see next page).

comparatively low (maximum 18.6%), indicating that the extrapolation does not add large biases to flux estimation in these cases.



Figure 4.4: (on the previous page) Comparison between sediment trap and UVP flux estimations done using UVP5 counts between 10 $\mu$ m and 2 mm and published A and b couples: A) (Kiko et al. (2017) following Kriest (2002)), B) Clements et al. (2023), C) Guidi et al. (2008), D) Forest et al. (2013) and E) Fender et al. (2019). Figure F is a comparison between sediment trap and UVP flux estimations done using UVP5 counts between 0.128 and 16.4mm using the A and b couple from Kiko et al. (2017) following Kriest (2002). Figures G and H were obtained by adding flux computed on UVP5 counts between 100  $\mu$ m-1 mm to flux computed on UVP5 images of aggregates and feces. The flux computed on the images was done using Kriest (2002) Table 1 2a relationship for sinking speed and (G) Durkin et al. (2021) or (H) Alldredge (1998) for the carbon content. The dashed black line corresponds to a 1:1 ratio. The values presented in each subplot correspond to the coefficient of determination ( $R^2$ ), the root mean square error (RMSE), the mean bias error (MBE) and the slope. The slope and  $R^2$  were computed by doing a linear regression between the sediment trap and UVP fluxes. All slopes are statistically significant (p-value < 0.05), as denoted by an asterisk.

## 4.5 Discussion

### 4.5.1 The dominance of aggregates

The dataset composition indicates a significant prevalence of aggregates. As expected, there are more aggregates in the uppermost layer, with aggregate formation surging at the surface during the start of the cruise. These aggregates sink to greater depth afterwards. The observed large dominance of these aggregates over zooplanktonic organisms is in agreement with previous studies (Stemmann and Boss, 2012; Trudnowska et al., 2021). The North Atlantic bloom observed during the EXPORTS experiment is associated with significant surges of particles (Lampitt, 1985), probably stemming from the formation of phytoplankton aggregates. The dominance observed in the images leads us to believe that this dominance persists in the size classes smaller than 1 mm suggesting that UVP particle counts could offer insights into aggregate dynamics and enable us to calculate gravitational flux from particle counts.

### 4.5.2 The most appropriate parameters to estimate flux

Three elements are crucial in flux computation: particle sinking speed, particle carbon content and size range observed. We first examine the suitability of published size-to-sinking speed relationships for our flux calculations, before undertaking a sensitivity analysis of different ap-

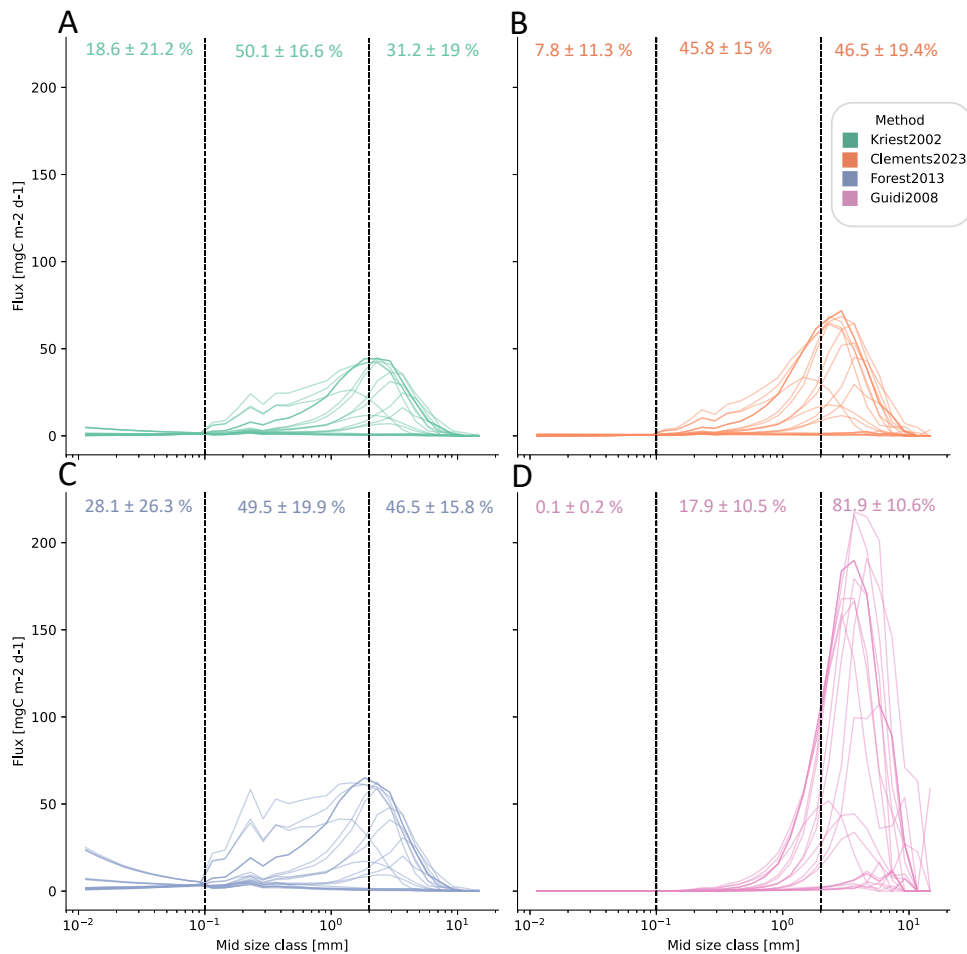


Figure 4.5: Flux (in  $\text{mgC m}^{-2} \text{d}^{-1}$ ) per size classes (in mm) computed on UVP particle counts for each matchup with sediment traps. Colours represent the method used to compute the flux. Values on top of each subplot represent the mean percentage  $\pm$  standard deviation of flux from particles between  $10\text{-}100 \mu\text{m}$ ,  $100 \mu\text{m}\text{-}2 \text{mm}$  and  $2\text{-}16.4 \text{mm}$ .

proaches to combine these with size-to-carbon relationships to calculate flux, and of approaches to estimate flux using relationships obtained using statistical approaches. We use sediment trap data as ground truth for flux calculations and, importantly, we restrict the size range we use to calculate flux to the size range that is efficiently captured by the sediment traps that were used during EXPORTS 2021.

### Size-sinking speed relationship

Sinking speed has been previously estimated to range from a few  $\text{m d}^{-1}$  for small particles to a few thousand  $\text{m d}^{-1}$  for dense salp fecal pellets (Phillips et al., 2009; Steinberg et al., 2023; Turner, 2002). In this study,

we estimated a significant positive size-sinking speed relationship for aggregates between 0.81 and 3.25 mm. Though we obtained enough data points for aggregates, it was more complicated for fecal pellets whose concentrations were much lower compared to the aggregates. The sinking speed of fecal pellets between 1 and 1.6 mm exhibits a higher variability than aggregates. While our approach demonstrated effectiveness in estimating sinking speeds within the size range of 0.81 to 3.25 mm, we encountered limitations when it came to smaller particles, both in terms of image analysis and UVP counts. Smaller particles typically exhibit reduced sinking velocities (Alldredge and Gotschalk, 1988; Cael et al., 2021; Iversen and Ploug, 2013; Kriest, 2002), which, in turn, may lead to an insufficient amount of available data across multiple depth layers for the accurate determination of the Gaussian maxima and subsequent sinking speed estimations.

In the past, the relationship between particle size and sinking speed in the marine environment has been considered to be a positive relationship (Alldredge and Gotschalk, 1988; Iversen et al., 2010), indicating that as particle size increases, their sinking speed also increases. When comparing our estimates to the parameterizations provided in (Alldredge and Gotschalk, 1988; Iversen et al., 2010), both overestimated sinking speed. However, this view has come under scrutiny by recent studies that suggested that the relationship should be considered in a more nuanced manner (Williams and Giering, 2022), taking into account factors such as particle porosity and density (Iversen and Lampitt, 2020) or more context-dependent factors (Cael et al., 2021). While it's worth mentioning that the study by Iversen and Lampitt (2020) examined a limited number of in situ particles with a unique setup, Cael et al. (2021) supports the idea that the relationship between particle size and sinking speed can vary in different local contexts. However, we currently cannot estimate porosity based on UVP5 counts or images and when using context-dependent factors for aggregates ( $\log(\alpha) = 2.04$  and  $\beta = 0.61$ ) and fecal pellets ( $\log(\alpha) = 2.52$  and  $\beta = 0.79$ ) from Cael et al. (2021), we obtain much higher sinking speeds for aggregates and fecal pellets than those we observed. In their compilation, most of the data used to estimate  $a$  and  $b$  came from laboratory experiments where aggregates were selected from the available community (from MSC or grown in the laboratory). This process may have influenced the results towards faster-sinking aggregates. Comparable studies which utilized the sinking speed from a plume (Briggs et al., 2011; Stemmann et al., 2002; Trudnowska et al., 2021) observed values similar to those obtained here. This suggests that the in situ population of particles may not necessarily sink at the rate given by experiments.

Overall, our results suggest that there is no strong deviation of the size-to-sinking speed relationship for the aggregates and fecal pellets observed during the EXPORTS experiment to Kriest (2002)'s equations

8 and 9 with the latter showing a better match. While equation 9 from Kriest (2002) produces the most favourable outcomes for this dataset, it should not be assumed to be applicable in all other situations.

### Size-carbon relationship

Particle carbon content can vary greatly depending on their composition. Because carbon content for the in situ imaged particles could not be obtained, we relied on prior studies (Alldredge, 1998; Durkin et al., 2021) that determined size-carbon content relationships to, in combination with the size-sinking speed relationship from Kriest (2002) to obtain particle flux. By utilising the identical size-sinking speed relationship for both UVP counts and the hybrid method based on UVP counts (for elements with an ESD 10-1000  $\mu\text{m}$ ) and images (for elements  $> 1000 \mu\text{m}$  ESD), we can compare the two approaches to estimate carbon content from size. In addition, we also examined other previously published relationships (Clements et al., 2023; Forest et al., 2013; Guidi et al., 2008) that utilize statistical approaches to determine the A and b values of the size-to-flux relationship without distinguishing between the size-carbon and size-sinking speed relationships. In general, these empirical approaches used minimization approaches to optimize the A and b estimates for a close matchup between UVP5-derived flux estimates and sediment trap data. To ground-truth our estimates of carbon flux from the diverse methods used, we utilized carbon flux observations obtained using NBSTs and STTs during the EXPORTS campaign.

### Covered size range

While the selection of the size-to-sinking speed, size-to-carbon, and/or size-to-flux relationships have long been debated (Iversen and Lampitt, 2020; Laurenceau-Cornec et al., 2020; Williams and Giering, 2022), there remains no agreement regarding size restrictions for comparing UVP5 and sediment trap data. Previous studies have implemented various size ranges that cover part or all of the size spectra encompassed by the UVP5 counts (100  $\mu\text{m}$  to larger than 26 mm ESD). While Guidi et al. (2008) focuses on 0.25-1.5 mm, Kiko et al. (2017) and Kiko et al. (2020), both based on Kriest (2002) implement a size range between 0.14-16.4 mm. Forest et al. (2013) uses classes between 0.08 - 4.2 mm and Clements et al. (2023) extends the size range at the lower limit via extrapolation and then uses 0.035-5 mm. However, in these efforts, it was not considered that sediment traps have an effective size limit. Durkin et al. (2021) showed that with NBSTs and STTs, with diameters of 12 cm and 7.5 cm, significant flux was observed in the 10  $\mu\text{m}$  to 2 mm size range. To compare sediment trap and UVP flux, we, therefore, extrapolated the concentration of particle counts in the UVP data between 10 and 100  $\mu\text{m}$ , similar to the approach used by Clements et al. (2023) and

following Baudena et al. (in review). Supplementary Figure S4.1 shows that these extrapolated values show a good alignment with the larger size categories. After this extension, we could constrain the size range to 10  $\mu\text{m}$  to 2 mm to calculate flux using different parameterizations and to enable the comparison to the sediment trap data.

#### 4.5.3 Flux results

##### Using A and b couples on UVP5 counts

When evaluating the suitability of the different models used to calculate flux, it is important to consider several criteria. An ideal fit between sediment trap data and UVP-derived flux estimates should demonstrate a slope close to 1, minimal mean bias error (MBE) and root mean square error (RMSE), with the coefficient of determination ( $R^2$ ) being of minor importance in this context. The best fit between UVP flux estimates for the 10  $\mu\text{m}$  to 2 mm size range and sediment trap flux was achieved using the Kriest (2002) (MBE = 8.93 and slope = 0.69) and Clements et al. (2023) (MBE = 1.03 and slope = 0.57) approaches. In our study, we found that using the A and b couple from Guidi et al. (2008) led to an underestimation of the flux with an MBE of -15.97 and a comparably flat slope of 0.43. In contrast, the use of the couple from Fender et al. (2019) led to a large overestimation of the flux compared to sediment traps with an MBE of 966.58 which was also the case for the Forest et al. (2013) A and b couple with an MBE of 111.24 and a slope of 0.76. The low values of parameter b in the pairs generating the optimal fit imply that the particles in our dataset possess a relatively high porosity, suggesting that they are mostly aggregates. This hypothesis is reinforced by the prevalence of aggregates (in particular, the fluffy variety) seen in this dataset. This further highlights the need to tailor UVP flux estimates to specific occasions, both in terms of region and biogeochemical context (Fender et al., 2019; Iversen et al., 2010). The methods used to obtain these results are determined by the factor A and the exponent b, with the first representing the mass flux of particles of the smallest size and the latter being linked to the fractal dimension of particles (Guidi et al., 2008). In a prior investigation of the export plume observed in this dataset, we discovered an increase in the fractal level of aggregates within the plume, which comprises a significant proportion of the dataset. To date, we have not been able to integrate this factor into our flux estimates due to limitations in the available data. Nonetheless, it would be advantageous to explore the development of a tool capable of accounting for the fractal level using UVP5 images in future research.

In the past, it was commonly assumed that small particles either do not sink at all or sink at such a slow rate that they were metabolized

within the upper mesopelagic zone, participating in the flux mainly through aggregation and disaggregation processes (Giering et al., 2014; Riley et al., 2012; Stemmann, Jackson, and Gorsky, 2004). According to the A and b couple utilised to calculate the flux, the impact of these small particles can vary greatly, mainly due to the exponent b, which will give greater importance to large particles compared to small ones if b is high (e.g. as is the case for (Guidi et al., 2008)). When we conducted flux estimations within the original UVP size range of 0.128 to 16.4 mm ESD using the Kriest (2002) parameterization, the linear regression metrics between these results and sediment trap flux exhibited lower performance, compared to the 10  $\mu\text{m}$  to 2 mm size range. Apart from the coefficient of determination ( $R^2$ ) that increased from 0.64 to 0.71 and is a reflection of the linear regression done between the UVP and sediment trap flux, the root mean square error (RMSE) increased from 100.72 to 161.96, the mean bias error (MBE) increased from 8.93 to 47.79 and the slope flattened from 0.69 to 0.5. This comparison highlights that the adjustment of the size spectrum used to match the UVP flux estimates with sediment trap data is a useful approach.

Having established that we can match UVP-derived fluxes with sediment trap fluxes using an extended (down to 10  $\mu\text{m}$ ) and constrained (upper limit 2 mm) particle size distribution, we can now inquire about the fraction of flux that occurs in the larger size range and is therefore not captured by the sediment traps used as well as the fraction that occurs in the smaller size range (10  $\mu\text{m}$ -128  $\mu\text{m}$ ). Our findings, obtained using the Kriest (2002) method, unveiled a potential  $31.2 \pm 19\%$  underestimation of carbon flux by sediment traps in the size range of 2 to 16.4 mm. Using the A and b couples from Clements et al. (2023) which also gave a satisfactory result estimates the flux in this fraction to be higher with  $46.5 \pm 19.4\%$ . In the small fraction, the flux estimated using this method accounts for  $18.6 \pm 21.2\%$  of the overall flux. These results underscore the critical importance of considering the specific size range that the sediment trap covers when attempting to compare UVP-derived flux data to take into account the part of the particle flux that is captured by both instruments. Future investigations could further validate these findings by examining additional datasets not only from this campaign but also from others, including the utilization of thorium data, which could help test these results.

#### Using the hybrid method

The hybrid method was performed by summing flux estimates made on UVP counts for size classes 10  $\mu\text{m}$ -1 mm using the Kriest (2002) method and flux estimates made on UVP images of aggregates and feces 1-2 mm size-range using the size-sinking speed relationship from Kriest (2002) and the size-carbon relationship from either Durkin et al. (2021) or Alldredge (1998).

So far, in almost all studies that used UVP data to estimate flux, no distinction was made between detrital particles, artefacts and zooplankton (Clements et al., 2022, 2023; Guidi et al., 2008; Kiko et al., 2017). This is primarily because particle images are only available for a portion of the size distribution (roughly 1 mm and up), but also because it is time-consuming to classify thousands or millions of images. Morphocluster allowed for the separation of aggregates and feces from zooplanktonic organisms and artefacts that represent respectively 0.7% and 17.4% of the complete dataset. With that in mind, this hybrid method didn't give the expected better results than those obtained using A and B couples from Kriest (2002) and Clements et al. (2023). Specifically, the hybrid approach tends to underestimate flux when using the size-carbon relationship from Durkin et al. (2021) and overestimate flux when relying on Alldredge (1998). It should be noted that the current dataset is primarily comprised of aggregates (80.8% of the complete dataset); however, other datasets may have more prevalent zooplanktonic organisms and fecal pellets (Turner (2002), Table 3). Given that the majority of the images (1 mm up) were composed of aggregates, we can hypothesise that a similar distribution may have occurred in the UVP5 counts. This could account for why the hybrid method did not produce significantly superior results compared to the other method. On this note, it is important to bear in mind that the UVP undersamples rare elements due to its low sampling volume, such as large fecal pellets, which can potentially transport significant amounts of carbon within the water column (Stamieszkin et al., 2021). Similarly, we show that sediment traps undersample the flux. While the analysis of the EXPORTS dataset reveals an abundance of aggregates, the applicability of the hybrid method may become more noticeable in different environmental contexts characterized by a substantial presence of zooplankton or artefacts.

## 4.6 Conclusion and perspectives

In summary, our study offers valuable insights and potential avenues for future research. Firstly, our analysis of the EXPORTS experiment has shed light on the preponderance of aggregates in particle abundance, highlighting the possibility of using UVP counts in the study of aggregate behaviour in this experiment. Secondly, our investigation has established that on average in this North Atlantic bloom, aggregates and fecal pellets manifested sinking speeds in alignment with earlier observations and modelling endeavours. Our study also underscores the need for better knowledge of the aggregates and fecal pellets' sinking speed and carbon content in order to obtain better matches between UVP and sediment trap using the hybrid approach.

We suggest that this method might be particularly useful in scenarios characterized by a more substantial contribution of living organisms to particle abundance. In the future, we advise that the deployment of UVP5 units in conjunction with sediment traps may help facilitate the estimation of particle carbon content imaged by the UVP5. Both gel-filled sediment traps (to get the size spectrum sampled by the sediment traps) and standard traps (to get the carbon flux) should be used in this configuration. As demonstrated in this study and previous research, small particles sinking slowly can significantly contribute to the carbon flux in the ocean (Durkin et al., 2015). Therefore, we recommend that future studies calculate the flux on the same size range as the sediment traps before comparing the UVP5 flux with them to facilitate a more accurate comparison. Additionally, our study highlighted the potential underestimation of the mean flux by sediment traps to be between 30% (using Kriest (2002)) and 46.5% (using Clements et al. (2023)), which could partially mitigate the missing carbon supply (Burd et al., 2010).

Lastly, our findings advocate for the necessity of comprehensive coverage across the entire size range sampled by both sediment traps and UVP before conducting the flux comparison between the two instruments. In this regard, the use of in situ instruments designed to look at organisms from 10 to 200  $\mu\text{m}$  could play a pivotal role in enhancing our understanding of microplankton dynamics. Furthermore, applying the sinking speed estimation method employed in this study to smaller particles has the potential to provide additional data points for comparing size and sinking speed. Collectively, these insights pave the way for enhanced comprehension of particle dynamics within marine ecosystems and point to promising avenues for further exploration.

## Data availability

UVP5 and sediment trap datasets are available on SeaBASS: <https://seabass.gsfc.nasa.gov/cruise/EXPORTSNA>



## 4.7 Supplementary section

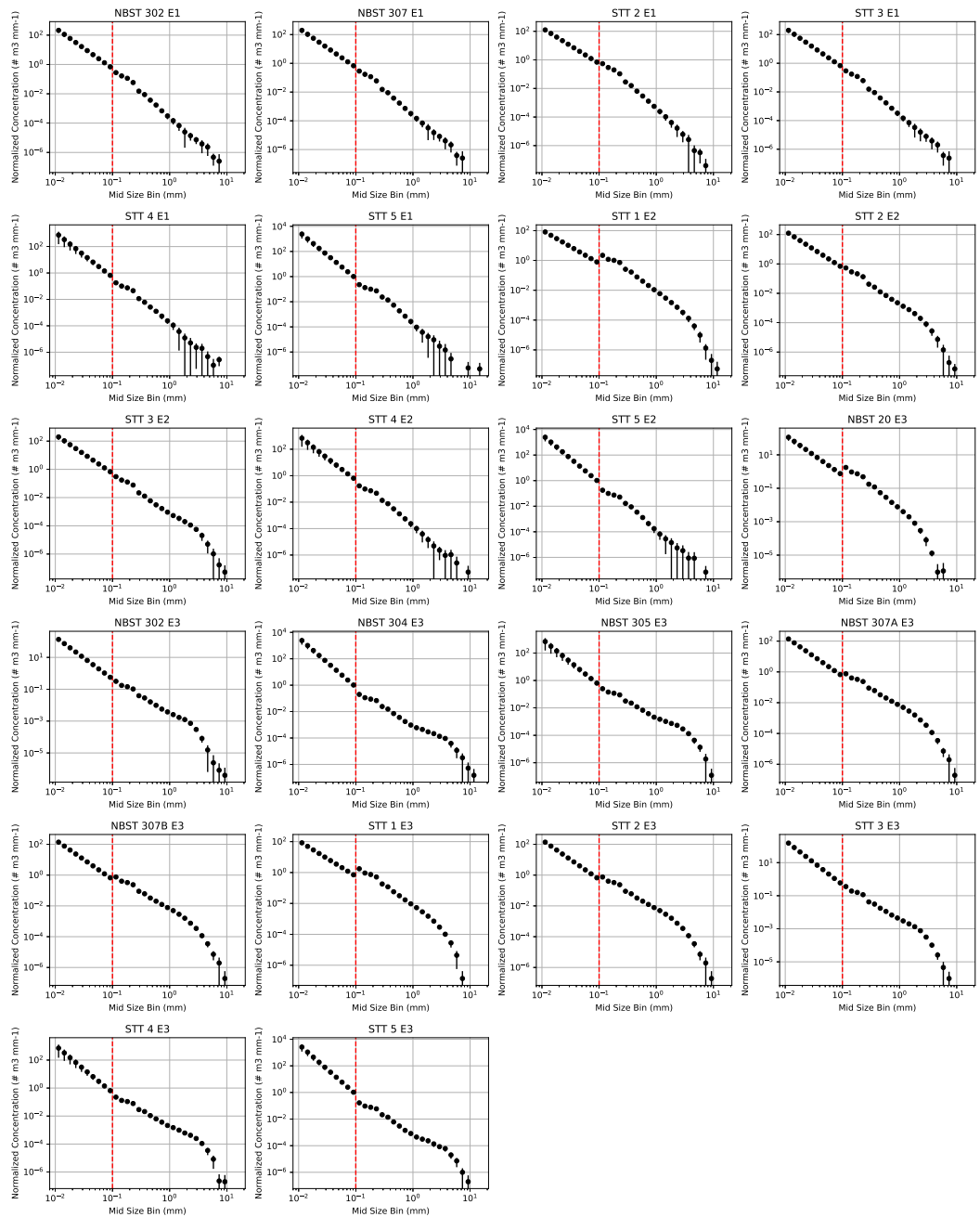


Figure S4.1: Concentration of UVP counts normalised by the width of the ESD size bin [ m<sup>-3</sup> mm<sup>-1</sup>].

Figure S4.1: (on the previous page) The red dashed line corresponds to a size of 100  $\mu\text{m}$ . Concentrations were obtained from the UVP5 particle counts per  $\text{m}^3$  between 0.102-16.4 mm (on the right of the red dashed line) and extrapolated to size classes between 0.010-0.102 mm. Both axes are on a logarithmic scale. Each subplot corresponds to the matchup between one sediment trap and the corresponding UVP profile as described in the Material and Method section. The vertical black lines correspond to the standard error of the standardised concentration.

## Part V

### General discussion and Perspectives



## General discussion and Perspectives

---

At the heart of our investigation into the Biological Carbon Pump (BCP) using *in situ* imaging lie two fundamental components: zooplankton and particles. Zooplankton was the focus of our comprehensive global examination (Chapter 1) and was also considered in a regional (Chapter 2) study, while particles were the focus of the regional study in the context of a North Atlantic eddy (Chapters 2 and 3).

### 5.1 General discussion

#### 5.1.1 Zooplankton Biogeography and Ecological Impact

Prior to the publication of Chapter 1, the biogeography of zooplankton biomass was primarily derived from punctual studies and heterogeneous datasets characterised by differences in data collection, tools and analytical methods. Previous studies heavily (Hatton et al., 2021) or completely (Buitenhuis et al., 2013; Moriarty and O'Brien, 2013; Moriarty et al., 2012) relied on net-based datasets that contained data sampled using different net types and measurement methods, reflecting the distribution of zooplankton larger than 200  $\mu\text{m}$ . However, net sampling is biased towards crustaceans (Lucas et al., 2014). Furthermore, estimates of global zooplankton biomass were calculated in most of these studies using a proportionality method by multiplying the mean observed zooplankton biomass by the area of the world ocean and the depth studied (Buitenhuis et al., 2013; Moriarty and O'Brien, 2013; Moriarty et al., 2012). It's important to note that this methodology may introduce a bias, as it relies on data that is distributed heterogeneously and does not cover large parts of the ocean, potentially leading to skewed results. The sampling coverage from biomass net samples is often patchy and skewed towards high latitudes and upwelling areas (Moriarty et al., 2012), undersampling open ocean regions (Buitenhuis et al., 2013; Moriarty and O'Brien, 2013; Moriarty et al., 2012). The estimation of zooplankton biomass is not a simple task, but an important one due to zooplankton's position as a crucial link between primary producers and higher trophic levels (Ikeda, 1985; Steinberg and Landry, 2017). Through their feeding activities (Alldredge et al., 1990; Huntley and Boyd, 1984; Steinberg et al., 2023), swimming capabilities (Dilling and Alldredge, 2000; Goldthwait et al., 2004) and ability to repackage

small particles into large, rapidly sinking fecal pellets (Atkinson et al., 2012; Turner, 2002, 2015), zooplankton organisms are important actors of the BCP (Steinberg and Landry, 2017; Turner, 2002, 2015). In order to accurately quantify the potential impact of zooplankton on the BCP, it is essential to have a clearer understanding of their distribution.

In this context, the work published in Drago et al. (2022) for the first time used boosted regression trees (a machine learning technique) on the global UVP5 in situ imaging dataset. The aim of this study was to develop a methodology to estimate zooplankton biomass on a global scale while examining how zooplankton's spatial distribution related with its habitat. It targeted large zooplankton between 1 and 50 mm using 3,549 profiles distributed all around the globe. Although the geographic coverage of the data points was not homogenous, we assessed that environmental conditions were adequately covered, enabling the use of habitat models. We employed a broad taxonomic definition, as imposed by the dataset, in order to homogenise the taxonomic definition across the scientific campaigns. Using this approach, we made a notable contribution to biological carbon pump research by producing continuous zooplankton biomass distribution maps for the 0-200, 200-500 and 0-500 m depth layers, using both global and regional perspectives and a homogeneously acquired dataset. With this approach, we also produced a more robust estimation of zooplankton biomass for depth ranges of 0-200, 200-500, and 0-500 m. We estimated the global integrated biomass (0-500 m) at 0.403 PgC. This estimation was in the same order of magnitude as previous studies (Buitenhuis et al., 2013; Hatton et al., 2021; Moriarty and O'Brien, 2013; Moriarty et al., 2012) though we refrained from a more detailed comparison due to the variations in sampling instruments, data analysis, and scaling techniques.

Predictions placed Copepoda as the most important contributor to biomass (35.7%) corroborating with previous studies (Dai et al., 2016; Forest et al., 2012; Turner, 2004). Their participation was especially high in the polar regions and their distribution was mostly driven by temperature, in agreement with previous studies (Beaugrand et al., 2009; Pinkerton et al., 2020; Sunday et al., 2012). This group was followed by Eumalacostraca (26.6%) and Rhizaria (16.4%). The vertical and horizontal distribution (mostly in the intertropical convergence zone) of the Rhizaria groups followed the description of previous studies (Biard et al., 2016; Michaels, 1988; Nakamura et al., 2013; Nakamura and Suzuki, 2015; Suzuki and Not, 2015).

We also attempted a more regional approach by partitioning the dataset in both latitude and depth. Unfortunately, this resulted in less satisfactory outcomes than the global methodology, primarily due to the diminished size of the learning datasets. The need for big enough

datasets is also what prompted us to use large zooplankton groups. In the future it may be possible to achieve a more detailed taxonomic and vertical definition after the collection of more data. More detailed models could potentially achieve better performance by providing a more accurate delineation of the vertical habitat structure of finer taxonomic group definitions. This appears to be the case for Phaeodaria, where 100 m resolution models gave higher  $R^2$  results, particularly in the depth range between 0 and 300 m.

A prior study conducted on a more restricted UVP5 dataset indicated that Rhizaria contributed 5.2% to the carbon standing stock in the 0-200 meter range (Biard et al., 2016). Our findings validate the significance of this group on a greater depth range, revealing Rhizaria's contribution of 16.4% to the global biomass integrated over 0-500 meters, suggesting a more important role of Rhizaria, in particular in the upper mesopelagic. These findings emphasise that traditional net-based sampling methods are unsuitable for the study of Rhizaria. The comparison I conducted using the Tara Oceans dataset (Chapter 1, Figure 10) provides further evidence for this conclusion and advocates for the deployment of multiple instruments during zooplankton sampling (Lombard et al., 2019) in order to comprehensively study these fragile organisms alongside other large zooplankton groups.

Our study not only contributed with global maps of biomass for the main zooplankton groups but also presented the main variables influencing the observed distribution of zooplankton biomass. It showed the evolution of zooplankton biomass within the ranges of these variables using partial dependence plots, although these partial dependence plots sometimes proved difficult to interpret. A recent study conducted by Soviadan et al. (2023) analysed the normalised biovolume size spectrum of major zooplanktonic groups in tropical, temperate and polar regions using a combination of UVP5 and multinet data. The study found that, in the polar region, crustacean organisms dominated, while Rhizaria dominated in the tropics, with Collodaria showing high contribution in the 0-200 m layer and Phaeodaria below 200 m, aligning with our findings.

In Chapter 2, our investigation of zooplankton centred on a North Atlantic eddy during the spring bloom. This study was the first to use Morphocluster (Schröder et al., 2020) to classify UVP5 images to determine the relative contribution of zooplankton to total particles in the water column. Our objective was to elucidate the complex depth scales and temporal dynamics occurring within this export event. In this context, we sought to investigate the potential role of zooplankton in influencing the BCP. Comparing our attenuation results with earlier

research conducted on bacterial respiration rates during a comparable season in the region (Belcher et al., 2016; Collins et al., 2015) suggested that zooplankton were not the main gatekeepers of the BCP. This was particularly the case below 300 m where bacterial activity was suggested to be the main driver of the observed attenuation. Surface and mesopelagic zooplankton populations were rather rare, casting doubt on their ability to significantly influence carbon flux through fecal pellet production in this particular dataset. However, their feeding activity via filter (Alldredge et al., 1990; Harbison et al., 1986; Huntley and Boyd, 1984) and flux feeding (Burd and Jackson, 2009; Gillard et al., 2022; Stemmann, Jackson, and Ianson, 2004), coupled with fragmentation through swimming behaviour (Dilling and Alldredge, 2000), may account for some of the high attenuation observed in the layer between 50 and 100 m, which cannot be attributed to bacterial activity alone (Belcher et al., 2016; Collins et al., 2015). The observed increased abundance of zooplankton noted in this layer with a maximum of rhizarians observed at 50 m and crustaceans at 125 m provides supporting evidence for this hypothesis. During the EXPORTS experiment, zooplankton organisms were in a highly reproductive phase, with a substantial number of eggs and juveniles observed both in net samples and through experiments conducted onboard (A. Maas, personal communication, August 24, 2023). It may have proven beneficial to plan for a longer sampling period to monitor the response of zooplankton productivity to increased productivity and export. Doing so might have revealed changes in zooplankton composition and abundance.

However, our observations uncovered changes in the morphology of fecal pellets which became more elongated and less circular within the mesopelagic layer, which coincided with an increased abundance of zooplankton, as confirmed by acoustic (M. Sato, EXPORTS meeting, September 17, 2021) and net data (A. Maas, personal communication, August 24, 2023). These observations prompted us to consider the potential influence of zooplankton on particle attenuation and dynamics within the mesopelagic layer through various processes.

One important process of interest is the differential settling of fast-sinking fecal pellets compared to relatively slow-sinking aggregates (Alldredge, 2001; Burd and Jackson, 2009; Kiørboe, 2001; Kriest and Oschlies, 2008). This difference in settling speed can lead to the integration of fecal pellets within aggregates, as they might collide and interact during their descent. Ultimately, this could result in larger and denser aggregates, which is consistent with some of the observations made. Notably, the increasing density and size of aggregates can have a significant effect on sinking rates and carbon content, potentially leading to a greater export of carbon from the surface ocean to deeper



layers. This process is essential for comprehending carbon fluxes in the marine environment, as it is linked to the efficiency of the BCP.

To advance our understanding in this area, it is essential to broaden the scope of our investigations. In this regard, the extensive EXPORTS 2021 dataset could offer further understanding of zooplankton impacts by using datasets such as those collected by nets and various on-board experiments. Coupled with the findings obtained from chapters 2 and 3, they have the potential to elucidate further aspects of the zooplankton ecology. Investigating different oceanic environments beyond this North Atlantic eddy can present a more comprehensive outlook on the influence of zooplankton on carbon flux, resulting in a more global overview of these processes.

#### 5.1.2 Particle characteristics and flux patterns: insights from a UVP-based analysis

The EXPORT 2021 experiment (Johnson et al., 2023; Siegel, 2016; Siegel et al., 2021) comprised three ships and a vast array of assets from in situ instruments to satellites, which were used to gather data before and during the on-site study. The experiment aimed to describe a selected eddy (Erickson et al., 2022) during a North Atlantic bloom event using instrumentation to examine the interplay between physics, biology, and ecology through a holistic view. Prior to the EXPORTS experiment, no previous studies involving the massive deployment of three UVP5 units within a geographically restricted area had been conducted. Operating within the unique framework of a highly retentive eddy allowed for a precious opportunity to conduct in-situ observations, providing novel insights into particle characteristics and flux patterns.

Previous in situ imaging studies using the UVP5 have predominantly focused on zooplankton (Biard et al., 2016; Panaiotis et al., 2023; L. Stemmann et al., 2008; Vilgrain et al., 2021) or particle counts (Clements et al., 2022, 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008), with aggregates and fecal pellets often being grouped under the classification of "detritus" on the Ecotaxa repository (personal observation from work on chapter 1). However, this classification fails to acknowledge the considerable diversity among these particles and their potential ecological significance. The investigation of particle morphology and the factors that affect it presented significant difficulties resulting in a lack of knowledge regarding the evolution of both aggregate and fecal pellet morphology at a high temporal and spatial scale that Chapter 2 aimed to address. With the deployment of the UVP5 system in a lagrangian framework, we gained the capability to track and describe the morphology of particles within an exported

plume, offering a unique 2D perspective. Inspired by the method used in Briggs et al. (2020) to study the particulate optical backscattering (bbp) which is a proxy for particulate mass concentrations (Reynolds et al., 2016), we implemented, for the first time, an oblique framework to follow the plume of imaged particles through five parallel masks in depth and time. This allowed us to better understand the dynamic nature of the particle flux within the plume. Our observations revealed distinct patterns for aggregates and fecal pellets: aggregates displayed a nearly linear change with depth, while fecal pellets exhibited more dynamic and heterogeneous alterations. Contrary to previous observations (Alldredge et al., 1990; Laurenceau-Cornec et al., 2015), our findings indicate that on average aggregates become larger as well as denser as depth increases. This could be explained by the differential settlement hypothesis by which fast sinking particles collide with smaller ones (Alldredge, 2001; Kiørboe, 2001; Kriest and Oschlies, 2008). This assumption is supported by a lower mean kurtosis value for aggregates in the lower plume, suggesting more heterogeneous aggregates. In the pictures shown in Chapter 2, aggregates appear to have incorporated more elements during their descent. A previous study of export events in Baffin Bay and Fram Strait also found that aggregates became denser, as their observations showed that most of the aggregates ended up in their dark morphotype associated with mostly compact objects, especially in Baffin Bay (Trudnowska et al., 2021). However, this dark morphotype was mostly associated with the smallest observed objects, which is not consistent with our observations in the North Atlantic eddy. However, Trudnowska et al. (2021) presented results showing that the k-means clustering method predominantly relied on size-related parameters for clustering images in their study. In contrast, before extracting features from the images, Morphocluster scales them to the same size (Schröder et al., 2020), removing the risk that images would be mainly clustered based on their size. The algorithm then conducts unsupervised clustering of the images while the operator supervises the next phases of validation and growth. While the unsupervised k-means clustering method is efficient and relatively easy to apply to a dataset, it may not provide enough information about the content of each cluster, as was possible with Morphocluster. The latter method, though much more time consuming, successfully distinguished between fecal pellets, various zooplankton and aggregate types within the dataset thanks to the semi-supervised approach.

The UVP5 provides particle counts that have historically been used to compute particle flux and attenuation in a vertical 1D framework (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008), disregarding the dynamic characteristics of particle communities at varying depths. That is one of the aspects that we addressed through

the plume following approach. We hypothesized that the characteristics and composition of particles at depth may not correspond to those present in shallower layers on a given day since most particles do not sink at a sufficiently rapid rate to reach depth in a single day. The previous studies (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008) utilized UVP5 counts across different size ranges to estimate carbon flux. The procedures used in these studies involved implementing a minimization process on a specific part of the UVP5 dataset and comparing it to the sediment trap estimates. However, they did not attempt to establish a definitive agreement regarding the size range employed. To achieve the most accurate flux estimates possible using the EXPORTS dataset, in Chapter 3 different flux calculation methods (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008; Kriest, 2002) were evaluated and compared with sediment trap flux estimates by varying the three main components of flux estimation: the size range considered, the size-carbon relationship and the size-sinking speed relationship. This comparison revealed that the best agreement between UVP5 and sediment trap flux estimates was obtained by extending and constraining the size spectrum for the UVP5 data to 10  $\mu\text{m}$ -2 mm, corresponding to the size range covered by the used sediment trap, as well as using parameters from Kriest (2002) and Clements et al. (2023), with a better result for the former for flux estimation. This study also highlighted that due to their limitation to the 10  $\mu\text{m}$ -2 mm size range, the used sediment traps could potentially underestimate carbon flux by 31.2-46.5%. This revelation suggests that our current methods of assessing carbon flux in marine ecosystems may be incomplete. Sediment traps, while valuable tools, could potentially fail to capture a substantial portion of carbon flux in the system. This result may have profound impacts on our understanding of carbon dynamics and resolve some of the imbalances found for mesopelagic carbon budgets between BGP export fluxes and those derived from biogeochemical tracers (Burd et al., 2010). On a global scale, this could potentially have significant implications for assessing the global performance of the biological carbon pump and marine carbon sequestration. This work compared UVP5 flux estimates with neutrally buoyant sediment traps (NBST, 12cm diameter opening) and surface-tethered traps (STT, 7.5cm diameter opening). Sediment traps with a wider opening such as the Kiel trap (Zeitzschel et al., 1978) could prove helpful in trying to cover a larger size range, but also in this case the potential for undersampling flux in the upper end of the size spectrum needs to be considered. While the use of images was anticipated to enhance the outcome of the hybrid flux computation method, it is worth noting that it did not produce satisfactory results. Given that the sinking speed estimates using the Kriest (2002) method appear to be fitting, it raises questions about the appropriateness of the conversion factors utilized

for estimating the carbon content of aggregates and fecal pellets (Alldredge, 1998; Durkin et al., 2021) within the context of this dataset. Our results have important implications for particle dynamics research and encourages the testing of this method on datasets from different biogeochemical contexts to see if this also leads to more consistent and comparable results across studies.

### 5.1.3 Exploring the boundaries: challenges in studying zooplankton biomass and particle evolution morphologies

As discussed above in chapter 1, the estimation of global zooplankton biomass has brought significant advances to the field, in particular as it presents the first global assessment of zooplankton biomass using a completely homogeneous sampling and data analysis approach. However, we faced some limitations during our journey that prompted critical reflection. An essential challenge was the carbon content values available in the existing literature, which often proved unsatisfactory, especially for the enigmatic Rhizaria group. Since the publication of Drago et al. (2022), a significant advancement has been published by Laget et al. (in review) where the carbon content of Rhizaria organisms collected in the California Current Ecosystem and in the Bay of Villefranche-sur-Mer, France was directly measured. The authors determined conversion factors from volume to carbon content for the Phaeodaria, Collodaria, Acantharia, and Foraminifera groups that were utilised in the habitat models outlined in Chapter 1. These factors exhibited a much lower carbon density than what was used in Drago et al. (2022). A revision of these models was subsequently carried out using even more UVP5 profiles down to 1,000 m, which should provide a more accurate global biomass estimate with greater depth coverage (Laget et al., in review).

In addition, the datasets at our disposal presented challenges as they did not allow us to construct reliable models for various taxonomic groups. This limitation stemmed from either an unclear relationship between biomass and environmental factors or an insufficient volume of data to establish meaningful associations.

With 19 groups of zooplankton, seven layers of depth, and five geographical definitions, there was a total of 665 models. Manually tuning all these models was not realistically feasible. To address this, we automated the tuning of hyperparameters for each model, aiming to avoid overfitting and to minimize average loss. This automated approach enabled us to find the best hyperparameter combinations for each model. Acquiring additional data for each combination could further enhance the model's ability to capture complex relationships

between biomass and environmental features, leading to improved generalization across the global ocean.

Our investigation of particle evolution provided compelling insights into morphological changes, concentration and biovolume dynamics. To achieve a more comprehensive understanding, an ecologically relevant clustering technique (Schröder et al., 2020) was implemented to reveal concealed patterns within the observed particle community during the EXPORTS experiment. A previous study used a k-mean clustering method on UVP5 detritus images, which yielded 5 clusters of detritus (Trudnowska et al., 2021). Although this gave an insight into the composition of the exported detritus through five ecologically relevant clusters, it did not allow the aggregates to be separated from the fecal pellets to be studied individually, as was done in our study. Although sorting detritus images is a laborious task, our endeavours allowed us to describe the progression of these two groups' morphology and hypothesise the processes that shape it. This database is now a valuable learning dataset for prospective investigations, providing a fundamental basis for the construction of classification models.

Finally, our investigation presented a new framework for studying UVP data in the context of an export event. Tracking particle community dynamics through this framework enabled us to gain a deeper understanding of the intricate processes that govern particle dynamics during this observed export event. One notable limitation of this approach is the requirement for an almost empty water column before the flux event and the necessity for high-frequency sampling within a Lagrangian framework. However, potential solutions to these limitations can be explored, such as the use of UVP6 mounted on ARGO floats to enhance sampling and capture a more comprehensive view of particle dynamics in diverse oceanic conditions.

## 5.2 Conclusion

In summary, our comprehensive study of the distribution of zooplankton at both global and regional scales, as presented in Chapter 1, has greatly improved our understanding of the biogeography of zooplankton biomass. We hope that the dataset will be of use for modelers studying the global ocean's biogeochemical cycles and the distribution of species feeding on zooplanktonic organisms. Additionally, our investigation of an export event in a North Atlantic eddy during the spring bloom in Chapter 2 has provided insight into the potential role of zooplankton in shaping particle dynamics and carbon fluxes in the top 500 m of the water column during a large export event.

We introduced a new framework for studying UVP5 data in the context of an export event through a plume following method. This approach allowed us to track particle community dynamics and gain a deeper understanding of the processes shaping particle flux within an export plume. Our investigation into flux calculation methods has shown that harmonizing the size range used in UVP5 data analysis with that of sediment traps can improve the accuracy of flux estimates and that sediment traps could miss a large part of this carbon flux. The findings presented in Chapters 2 and 3 on the EXPORTS experiment challenged existing assumptions on the evolution of particles morphology down the water column and brought contributions to the understanding of particle characteristics, attenuation patterns, and flux dynamics.

To conclude, this collection of studies not only pushes the boundaries of knowledge but also addresses critical challenges in estimating zooplankton biomass and deciphering the evolution of particle morphologies and flux in the context of an export event. These developments open up new avenues for future research that will be needed to address the challenges and unanswered questions that have emerged from this work.

### 5.3 Perspectives

In this section, we broaden the scope of our discussion, going beyond the immediate findings of this thesis to explore broader issues in data-driven research. While our primary focus remains on the biological pump and imaging aspects, we also recognize the significance of these discussions in the context of advancing ocean science.

#### 5.3.1 Navigating the challenges and opportunities of data-driven ecology.

The field of data-driven ecology has opened up a new era of scientific inquiry that is redefining our understanding of the natural world. Data-driven science necessitates a substantial investment in resources, both human and technological. The enormous amount of data produced, gathered and examined in present-day ecological investigations (Irisson et al., 2022) can be overwhelming and is qualified as big data (Guidi et al., 2020). Big data and machine learning tools have assisted studies related to ecosystem management, including marine protected areas (Benedetti et al., 2019; Muñoz et al., 2017) and the identification of zooplankton indicators (Uusitalo et al., 2016). Despite the potential of this data to transform our understanding of ecosystems, it also presents challenges in implementing good data management practices following

the FAIR data principles outlined by Wilkinson et al. (2016) that aim to facilitate the options to find, access, interoperate and reuse data.

In addition, the use of big data raises significant concerns regarding resource management due to the increased need for storage capacity and computing power to analyse the data. This increased need for resources raises crucial inquiries regarding the balance between sustainability, affordability, and equitable access to cutting-edge scientific tools and knowledge. In light of these challenges, innovative frugal tools have been developed such as the PlanktoScope, a compact modular imaging platform (Pollina et al., 2022) or the Foldscope, a low-cost origami-based microscope (Cybulski et al., 2014).

As data-driven ecology advances, researchers and institutions must proactively address these issues. Reconciling the enormous potential of big data with today's resource constraints and ethical considerations is crucial to ensuring that data-driven ecology continues to be a powerful and inclusive tool for scientific exploration and environmental conservation.

### 5.3.2 Future of global zooplankton biomass research

As we explore the future of understanding zooplankton biomass distribution, several exciting avenues await exploration. One promising avenue involves the continued deployment of the UVP5 during research cruises and the autonomous deployment of the UVP6 (Picheral et al., 2021) on floats, gliders, moorings and other vectors. The expansion of the UVP datasets is expected to contribute to improved annual biomass predictions and perhaps even enable seasonal modelling of zooplankton biomass at global levels. The deployment of UVP6 units on BGC Argo floats, gliders and moorings will offer long-term series that enhance our comprehension of local dynamics. In chapter 1, we saw that 61% of the profiles had a maximum depth of 500 m. Consequently, systematically extending vertical profiles when possible to cover the entire 0-1000 m depth range has the potential to provide a more extensive understanding of the distribution of zooplankton throughout a larger water column coverage.

Furthermore, it is worth noting that ongoing efforts (Dugenne et al., 2023) in the exploration of zooplankton size distribution using a multi-parameter learning approach could represent a critical advancement. The structure of the plankton and its particle size has been used as a means to study complex ecological processes (Sheldon et al., 1972; Siviadan et al., 2023; Stemmann and Boss, 2012). This method has the potential to improve our comprehension of zooplankton functions in marine ecosystems and to explore more thoroughly the subtle con-

nections and ecological importance between different size groups of zooplankton.

Additionally, as we venture into the realm of data analysis, it is essential to consider the continuous evolution of deep learning, a sub-field of machine learning (LeCun et al., 2015). These rapidly advancing technologies offer exciting prospects for managing and extracting meaningful insights (Luo et al., 2018; Orenstein et al., 2022; Schröder et al., 2020) from the ever-expanding volumes of data at our disposal (Irisson et al., 2022). They have already found various applications in plankton studies as recently presented by Goodwin et al. (2022). By embracing these advancements, we can enrich our understanding of zooplankton in marine ecosystems and their intricate dynamics.

### 5.3.3 Future prospects for studying particles via the follow the plume approach

In the course of studying carbon export using the plume following approach, several important considerations have emerged that point to promising avenues for future research. First and foremost, the plume following approach has the potential to be applied to various other export events. Applying this methodology to different scenarios would not only help to generalise our findings but also shed light on the variability of export processes across different contexts. Achieving this may require deploying UVP6 units on moorings or mounted on BGC Argo floats in various biogeochemical environments. These expanded data sets would allow for a more comprehensive view of carbon export dynamics, providing better insights into variations over time and under different scenarios. The learning set of images of particles created using MorphoCluster on the EXPORTS dataset could be used and extended in such research efforts. Recent research was conducted by colleagues (Ricour, 2023) to implement an embedded recognition algorithm in UVP6 units mounted on BGC Argo floats to classify imaged particles. At present, aggregates and fecal pellets are tagged as "detritus" in accordance with the classification usually done on Ecotaxa that was discussed earlier. Although this detection algorithm was limited in the number of classes that could be used, it would be worthwhile to try using Morphocluster on UVP6 detritus images in order to differentiate between aggregates and fecal pellets and to determine if the embedded algorithm can achieve satisfactory results for their identification.

An interesting question arises regarding the determination of the appropriate flux attenuation factor - whether to follow the plume or to use a standard approach. While we have suggested that the former is the most appropriate for our data set, it would be important to in-



investigate this aspect further. The validation of the attenuation factor, possibly through comparative studies, would provide valuable insights into the robustness and applicability of our approach. Moreover, we must emphasise that our research primarily concentrated on observations rather than detailed modelling of the plume event, including the calculation of disaggregation and aggregation rates. Modeling these rates and their impact could provide a deeper understanding of particle dynamics within the plume. Expanding the scope of our research to cover additional export events could also help to address these elements. Furthermore, we do not know whether the patterns identified in this dataset for both aggregates and fecal pellets represent a normal signature or are more of an exception. Do aggregates generally become larger and denser with depth, as our results suggest? Do they generally also become more complex and circular with depth? Applying this method to other export events may help answer these questions and enhance the comprehensiveness of our findings, contributing to a broader understanding of carbon export dynamics

Moreover, while implementing the plume methodology, limitations have been identified concerning data availability for sinking speed estimations, especially for smaller particle sizes. Therefore, collecting more data for sinking velocity estimation covering a wider range of particle sizes could be beneficial. An important future goal should be to establish a direct relationship between the sinking rate and the morphology and carbon content of aggregates. This link would provide insight into the factors that affect in situ sinking rates of diverse particle types and their potential for carbon export, thus advancing our understanding of this key process. Deploying UVP5 and sediment traps simultaneously could provide comprehensive datasets, complementing insights into carbon export dynamics. This integrated approach would significantly increase the accuracy and reliability of the data collected.

Overall, this manuscript not only advances our knowledge, but also addresses critical challenges in estimating zooplankton biomass and particle dynamics during export events. The findings of this study open up new avenues for future research on the biological carbon pump and deepen our understanding of marine ecosystems.



## Part VI

### Résumé en français



## Résumé en français

---

### 6.1 Introduction

#### 6.1.1 La Pompe Biologique Gravitationnelle

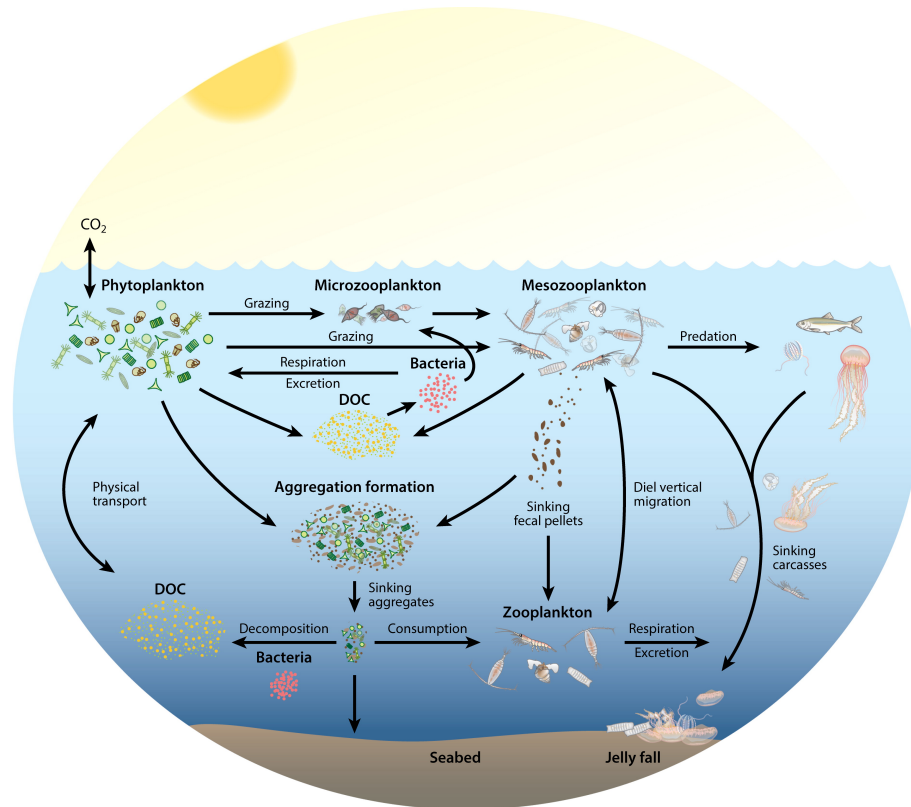
La pompe à carbone biologique transporte le carbone de la surface de l'océan vers ses profondeurs (Fig. 6.1, Honjo et al. (2008) and Volk and Hoffert (1985)), séquestrant le carbone organique particulaire (POC) pendant de longues périodes (DeVries et al., 2012). Elle comprend des mécanismes physiques (Boyd et al., 2019; K. O. Buesseler et al., 2020; Omand et al., 2015) et biologiques (Bianchi et al., 2013; Jónasdóttir et al., 2015; Lampert, 1989), la pompe gravitationnelle biologique (BGP) étant le principal facteur de l'exportation nette de carbone à l'échelle mondiale (Boyd et al., 2019; Nowicki et al., 2022). Bien qu'il existe des estimations du flux de carbone mondial, des incertitudes subsistent en raison des lacunes dans notre compréhension de la dynamique des particules.

#### 6.1.2 Les particules marines

Les particules marines, y compris les agrégats, les organismes morts et les pelotes fécales, sont très répandues dans les océans (Alldredge et al., 1990; Alldredge and Silver, 1988). Leur dynamique de chute dépend de leur densité (Bach et al., 2019; Cael et al., 2021), de leur taille (Guidi et al., 2008; Iversen et al., 2010), de leur composition (Laurenceau-Cornec et al., 2015) et de leur morphologie (Trudnowska et al., 2021), la majorité d'entre elles étant concentrées près de la surface où se produit la production primaire (Guidi et al., 2015; Kiko et al., 2017; Stemmann et al., 2002). Ces particules présentent diverses morphologies façonnées par des processus physiques (Alldredge, 2001; Kiørboe, 2001; Stemmann, Jackson, and Ianson, 2004) et biologiques (Alldredge, 2001; Marsay et al., 2015; Turner, 2015), ce qui nécessite la prise en compte des microorganismes et du zooplancton dans les modèles d'estimation des flux.

### 6.1.3 Les rôles du zooplancton dans les écosystèmes marins

Définis pour la première fois comme des organismes à la dérive dans l'eau par Hensen en 1887 (Smetacek, 1999), les organismes planctoniques sont classés en deux groupes principaux : les producteurs primaires phytoplanctoniques qui absorbent le CO<sub>2</sub> par photosynthèse dans la couche euphotique (voir Fig. 6.1) ; et le zooplancton qui les broute, servant de lien majeur entre les producteurs primaires et les niveaux trophiques supérieurs (Ikeda, 1985; Steinberg and Landry, 2017). Le zooplancton est un groupe diversifié dont la taille varie du micromètre aux mètres (Lombard et al., 2019). Ces organismes présentent des variations temporelles et latitudinales de biomasse et de diversité liées aux paramètres environnementaux (Ibarbalz et al., 2019; Ikeda, 1985; Moriarty et al., 2012; Rombouts et al., 2009). Le zooplancton contribue de manière significative à la pompe biologique à carbone en convertissant les petites particules en pelotes fécales qui coulent rapidement (Atkinson et al., 2012; Turner, 2002, 2015) et en fragmentant les particules qui coulent par le biais de diverses activités de nutrition (Alldredge et al., 1990; Huntley and Boyd, 1984; Steinberg et al., 2023) et de nage (Dilling and Alldredge, 2000; Goldthwait et al., 2004). Il est essentiel de comprendre leur dynamique à différentes échelles pour quantifier avec précision leur impact sur la pompe biologique à carbone.




 Steinberg DK, Landry MR. 2017. *Annu. Rev. Mar. Sci.* 9:413–44

Figure 6.1: Voies du cycle et de l'exportation du carbone par le zooplancton dans l'océan. Les organismes phytoplanctoniques absorbent le  $\text{CO}_2$  par photosynthèse dans la couche euphotique. Ils sont broutés par le micro et le mésozooplancton qui participent à la pompe à carbone biologique en produisant des pelotes fécales qui coulent de la surface jusqu'en profondeur. Figure extraite de Steinberg and Landry (2017)

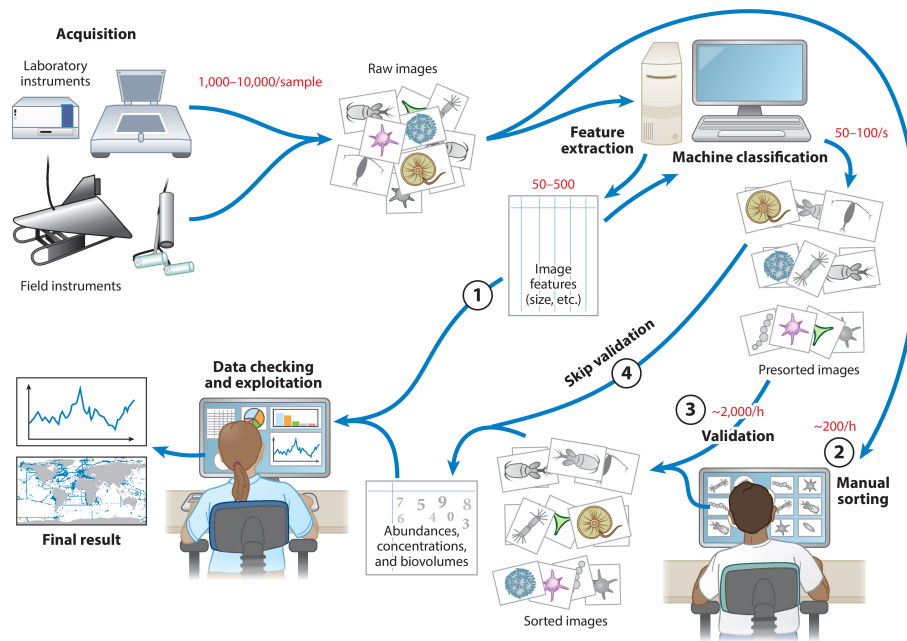
#### 6.1.4 Etude du zooplancton et des particules grâce à l'imagerie in situ

Alors que les paramètres basés sur la taille ont fait l'objet de nombreuses études (Cael et al., 2021; Guidi et al., 2008; Romagnan et al., 2016; Stemmann et al., 2002), peu d'entre elles ont examiné les changements d'autres traits morphologiques (Giering et al., 2020; Trudnowska et al., 2021). En outre, les images peuvent également faire l'objet d'une validation manuelle par des experts (stratégie 2, Fig. 6.2). Toutefois, ce processus peut potentiellement introduire plusieurs biais cognitifs, quel que soit le niveau de connaissance taxonomique des individus, comme le soulignent les études de Culverhouse (2007) and Culverhouse et al. (2014). La classification manuelle de ces grandes quantités d'images nécessiterait un nombre important de personnes hautement qualifiées en taxonomie, en particulier compte tenu du taux exponentiel d'acquisition d'images observé ces dernières années (Fig. 6.3). Toutefois, comme le montre la Figure 6.2, la gestion de jeux de données riches en images nécessite des stratégies innovantes (Irisson et al., 2022), notamment des méthodes de classification semi-automatiques (stratégie 3, figure 6.2) ou entièrement automatisées (stratégie 4, figure 6.2) afin de simplifier le traitement et l'analyse de ces jeux de données riches en images.

#### 6.1.5 L'apprentissage automatique en écologie marine

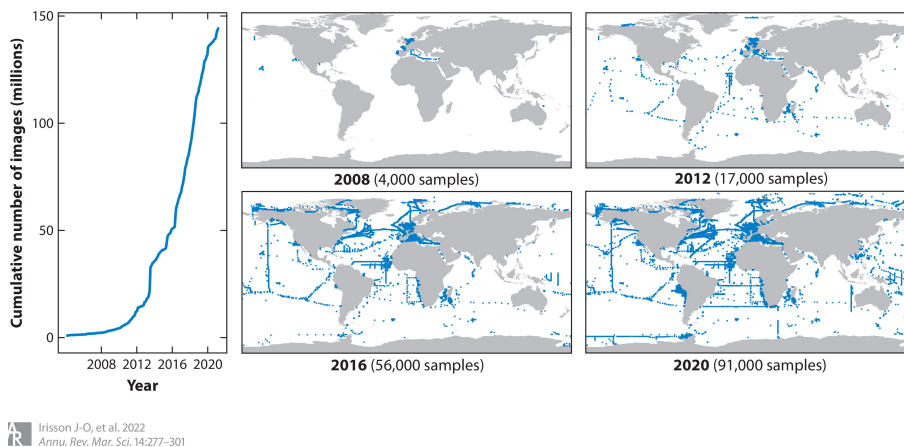
L'apprentissage automatique est une approche informatique puissante qui a trouvé des applications en écologie marine (Rubbens et al., 2023). Il comprend des techniques de classification (Luo et al., 2018; Orenstein et al., 2020) et de régression (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000) des données. Ces méthodes peuvent être automatiques (Luo et al., 2018; Trudnowska et al., 2021) ou semi-automatiques (Biard et al., 2016; Panaïotis et al., 2023; Schröder et al., 2020; Vilgrain et al., 2021), avec la nécessité de traiter les biais potentiels présents dans les données d'entraînement. La combinaison de l'exploration des données et de l'apprentissage automatique peut révéler des patrons et des relations cachées dans des données écologiques complexes, ce qui permet de mieux comprendre les écosystèmes marins.





Irison J-O, et al. 2022  
*Annu. Rev. Mar. Sci.* 14:277-301

Figure 6.2: Quatre stratégies de traitement de collections d'images brutes de plancton identifiées par quatre numéros encadrés. L'interaction efficace entre les instruments, les ordinateurs et les opérateurs humains permet un débit élevé depuis l'acquisition d'images en laboratoire ou sur le terrain jusqu'à l'exploitation écologique. Figure tirée de Irison et al. (2022)



Irissou J-O, et al. 2022  
*Annu. Rev. Mar. Sci.* 14:277–301

Figure 6.3: Évolution du nombre d'images téléchargées sur le site Ecotaxa (Picheral et al., 2017) à l'aide d'instruments tels que l'Underwater Vision Profiler (UVP (Picheral et al., 2010, 2021)), FlowCytobot (IFCB, (Olson and Sosik, 2007)), FlowCam (Sieracki et al., 1998), Zooscan (G. Gorsky et al., 2010), In Situ Ichthyoplankton Imaging System (ISIIS, (Cowen and Guigand, 2008)). Figure tirée de Irissou et al. (2022)

#### 6.1.6 Objectifs de ma thèse

Ce travail vise à réaliser une analyse globale de la distribution du zooplancton et de l'export vertical de particules basée sur l'imagerie in situ à partir des données de l'UVP5 (Fig. 6.4). L'étude s'articule autour de deux thèmes principaux : la distribution mondiale de la biomasse du zooplancton et l'exportation de carbone détritique dans le contexte d'une efflorescence de l'Atlantique Nord. Elle vise (1) à quantifier la biomasse mondiale de zooplancton (2) à améliorer notre compréhension de la pompe à carbone biologique par l'étude de la morphologie des particules et de son changement avec la profondeur en relation avec la distribution du zooplancton (3) à calculer et à comparer le flux issu de la production lors de l'efflorescence printanière de l'Atlantique Nord en utilisant à la fois les pièges à sédiments et les données UVP5 à haute résolution.

##### 6.1.6.1 Distribution mondiale de la biomasse du zooplancton

Le premier chapitre se concentre sur l'utilisation d'un ensemble de données UVP5 mondiales pour examiner la distribution de la biomasse du zooplancton et aborde les questions de recherche clés suivantes : 1) Pouvons-nous estimer avec précision la distribution géographique des grands groupes de plancton à travers plusieurs couches de profondeur en utilisant l'imagerie in situ ? 2) Est-il possible d'obtenir des estimations régionales de la biomasse du zooplancton à l'aide de cette

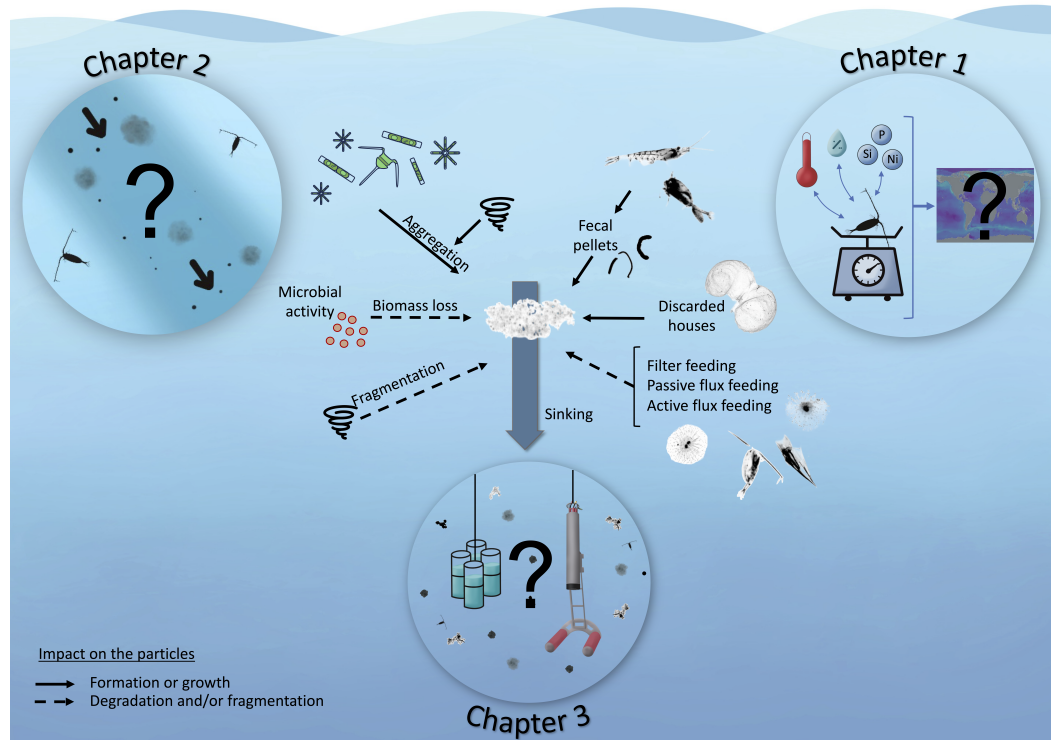


Figure 6.4: Résumé graphique résumant les principales voies du carbone dans la colonne d'eau en tenant compte de trois acteurs principaux : le phytoplancton, le zooplancton et les bactéries. Les flèches pleines présentent un processus aboutissant à la formation ou à la croissance de particules tandis que les flèches en pointillés correspondent à des processus aboutissant à la dégradation et/ou à la fragmentation. Les grands cercles représentent les trois chapitres qui composent ce manuscrit.

méthodologie ? 3) Les résultats obtenus s'alignent-ils sur les données existantes et sur les modèles précédemment établis ?

Pour étudier ces questions, nous avons exploité la capacité de prédiction d'Ecotaxa (Picheral et al., 2017) en collaboration avec divers experts qui ont consacré des années à la classification méticuleuse de cet impressionnant ensemble de données contenant 466 872 images de zooplancton. Nous avons émis l'hypothèse que la distribution des organismes et leur biomasse dépendent de facteurs environnementaux. Nous avons utilisé ces données d'imagerie in situ provenant de 3 549 profils UVP5 dans le monde entier et avons développé une nouvelle technique pour estimer globalement la biomasse du macrozooplancton sur la base de ces images et de modèles d'habitat.

### 6.1.6.2 Export durant une efflorescence en Atlantique Nord

Les deux chapitres suivants se concentrent sur la campagne EXPORTS 2021 qui s'est déroulée dans un tourbillon de l'Atlantique Nord pendant l'efflorescence printanière (Johnson et al., 2023). Au cours de cette campagne, trois unités UVP5 ont été déployées à partir de trois navires scientifiques, ce qui a permis d'obtenir 1 720 914 images. Alors que nous abordons les deuxième et troisième chapitres de cette étude, il devient évident que le grand volume d'images collectées au cours de l'expérience EXPORTS a nécessité un changement d'approche. L'ampleur de l'acquisition des données a empêché l'utilisation du cadre Ecotaxa, ce qui nous a incités à explorer d'autres méthodes. En outre, nous avons reconnu la nécessité d'un examen plus approfondi des débris, ce qui a motivé notre recherche d'une compréhension plus profonde dans les chapitres suivants et la classification des 1 720 914 images à l'aide de Morphocluster, une classification semi-supervisée conçue pour annoter de grands ensembles de données d'images (Fig. 6.5, Schröder et al. (2020)). Ce programme extrait des traits d'images qui sont ensuite projetés dans l'espace des traits. Le regroupement des images est ensuite basé sur leur position dans cet espace de traits. Si un groupe est jugé satisfaisant par un validateur humain, il est ensuite agrandi sous supervision humaine, et plusieurs itérations du processus de regroupement et d'agrandissement sont utilisées pour affecter les images aux groupes. Les groupes sont ensuite organisés manuellement de manière hiérarchique et nommés.

Dans ce contexte, le chapitre 2 explore l'exportation de débris marins imagés en utilisant un nouveau cadre qui emploie une méthode de suivi du panache d'export oblique. Ce chapitre vise à fournir de nouvelles perspectives sur la dynamique, la morphologie, le flux et l'atténuation des particules dans un tourbillon de l'Atlantique Nord en répondant aux questions suivantes : 1) Comment la communauté de particules évolue-t-elle au cours de ce grand événement d'exportation ? 2) Le zooplancton est-il important par rapport aux agrégats marins ? 3) Quel est le taux d'atténuation du matériel exporté et comment est-il lié à l'activité biologique ? 4) Comment la morphologie des agrégats et des pelotes fécales évolue-t-elle en fonction de la profondeur et du temps, et quels sont les processus à l'origine de ces changements ? En raison de la rareté des recherches sur la morphologie des particules in situ avec une haute résolution verticale sur une large gamme de profondeur, notre objectif était d'offrir une caractérisation plus précise de la dynamique des particules, en se concentrant sur les agrégats et les pelotes fécales.

Le chapitre 3 utilise une approche verticale plus traditionnelle pour calculer et comparer le flux de l'efflorescence printanière de l'Atlantique Nord en utilisant des pièges à sédiments et des données UVP<sub>5</sub>. Il se concentre sur les comptages de particules dérivés des données UVP<sub>5</sub> ainsi que sur les images d'agrégats et de pelotes fécales (1 - 16,4 mm) et vise à répondre aux questions suivantes : 1) Quels sont les types de particules dominants et leurs vitesses de chute dans le cœur du tourbillon pendant l'efflorescence printanière de l'Atlantique Nord ? 2) Comment le flux peut-il être calculé et comparé efficacement en utilisant les pièges à sédiments et les données de l'UVP<sub>5</sub> ?

Enfin, nous concluons et discutons les principaux résultats de chacun de ces trois chapitres.

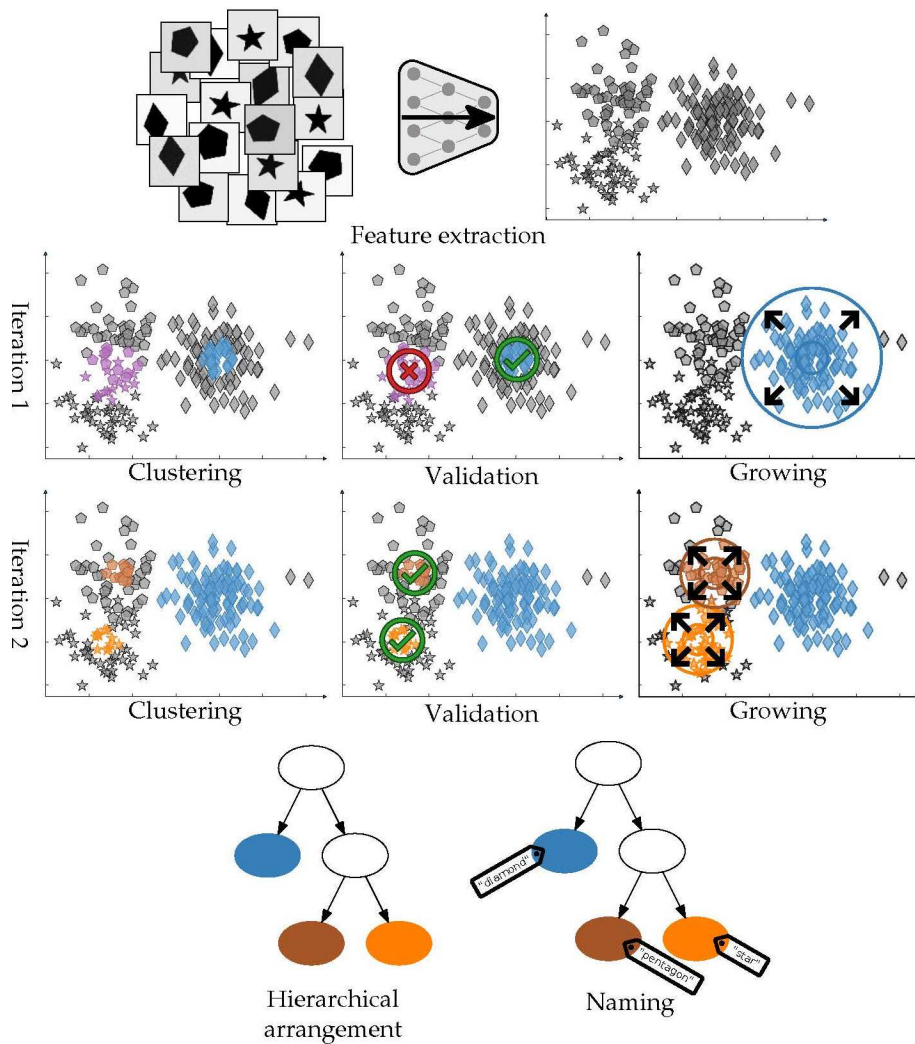


Figure 6.5: Étapes suivies par Morphocluster, un logiciel semi-autonome utilisé pour classer les images. Figure extraite de Schröder et al. (2020)

## 6.2 Distribution mondiale de la biomasse du zooplancton

### 6.2.1 Contexte et méthodes

Comme décrit précédemment, le zooplancton correspond aux organismes dérivant avec les courants. Il est très diversifié (de Vargas et al., 2015; Karsenti et al., 2011) et fait partie intégrante du cycle biogéochimique du carbone et de l'équilibre écologique de l'océan (Longhurst and Glen Harrison, 1989; Steinberg and Landry, 2017; Turner, 2002, 2015). Le zooplancton est sensible aux facteurs environnementaux, ce qui se traduit par des schémas mondiaux distincts en termes de diversité et

de biomasse (Ibarbalz et al., 2019; Rombouts et al., 2009). Alors que la biomasse est élevée aux hautes latitudes et faible aux basses latitudes, la diversité suit le schéma inverse (Ibarbalz et al., 2019; Ikeda, 1985; Moriarty et al., 2012; Rombouts et al., 2009). Ces schémas, déterminés par la température, les courants, les nutriments et les interactions biologiques, permettent de mieux comprendre les changements océaniques. Cependant, les évaluations de la biomasse sont entravées par l'hétérogénéité des données et des observations limitées (Le Quéré et al., 2016; Moriarty and O'Brien, 2013; Moriarty et al., 2012).

Les méthodes d'échantillonnage traditionnelles, biaisées en faveur des taxons non gélatineux, sous-estiment la biomasse et le rôle écologique d'organismes fragiles comme les cnidaires et les rhizaires (Biard et al., 2016; Lucas et al., 2014). Les outils d'imagerie in situ non intrusifs comme l'UVP5 (Underwater Vision Profiler 5, (Picheral et al., 2010)) ont révolutionné l'étude du zooplancton, mais les données restent inégalement réparties. Après une classification assistée par ordinateur de 466 872 organismes provenant de plus de 3 478 profils (0-500 m) obtenus entre 2008 et 2019 à travers le monde, nous avons estimé le biovolume individuel et l'avons converti en biomasse en utilisant des facteurs spécifiques à chaque taxon pour 25 grands groupes taxonomiques. Nous avons ensuite associé ces estimations de biomasse aux climatologies des variables environnementales (température, salinité, oxygène, etc.), afin d'élaborer une nouvelle technique pour estimer globalement la biomasse du macrozooplancton grâce à des modèles d'habitat utilisant des arbres de régression boostés (Elith and Graham, 2009; Hastie et al., 2001). Pour cela, nous avons utilisé différentes partitions des données basées sur la profondeur et la latitude, afin de capturer les différentes préférences en matière d'habitat.

### 6.2.2 Principales conclusions et perspectives

Les résultats révèlent des valeurs maximales de biomasse de zooplancton autour de 60°N et 55°S ainsi que des valeurs minimales autour des gyres océaniques (Fig. 6.6). Une augmentation de la biomasse de zooplancton est prédite autour de l'équateur. La biomasse intégrée globale (0-500 m) a été estimée à 0,403 PgC. Elle était largement dominée par les copépodes (35,7 %, principalement dans les régions polaires), suivis par les Eumalacostraca (26,6 %) et les Rhizaria (16,4 %, principalement dans les zones intertropicales). L'approche d'apprentissage automatique était sensible à la quantité de données d'entraînement et a généré des prédictions fiables pour les groupes abondants tels que Copepoda ( $R^2$  20-66%) mais pas pour les groupes rares (Ctenophora, Cnidaria,  $R^2 < 5\%$ ). Néanmoins, cette étude offre le premier protocole permettant d'estimer la biomasse globale et spatialement résolue du zooplancton

et la composition de la communauté à partir d'observations d'imagerie in situ d'organismes individuels. L'ensemble des données sous-jacentes a été obtenu en l'espace de dix ans, alors que des approches similaires s'appuient sur des données obtenues à l'aide de filets à plancton collectés depuis 1960 environ. L'utilisation accrue des méthodes d'imagerie numérique devrait nous permettre à l'avenir d'obtenir des estimations de la distribution de la biomasse du zooplancton à l'échelle du bassin et à l'échelle mondiale dans des délais plus courts.

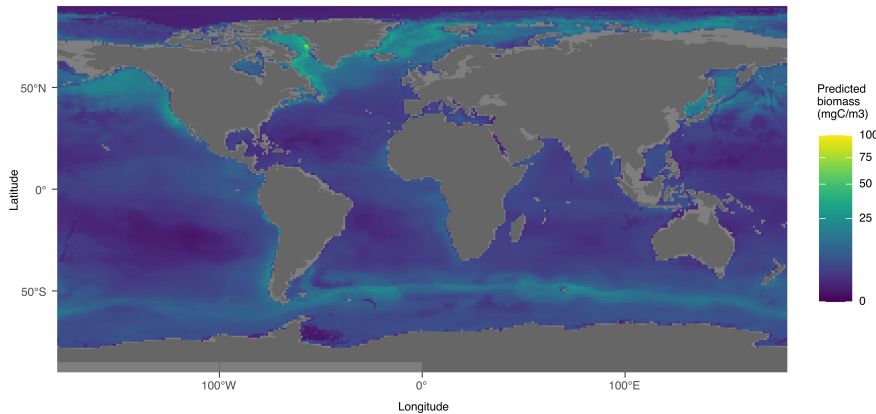


Figure 6.6: Carte de distribution de la biomasse globale minimale entre 0 et 500 mètres en utilisant les taxons qui ont obtenu une p-value  $< 0,05$  dans le test de Pearson entre la biomasse prédite et la biomasse calculée à partir des données UVP5.

Les prédictions ont placé les Copépodes comme les contributeurs les plus importants à la biomasse (35,7 %), ce qui corrobore avec des études précédentes (Dai et al., 2016; Forest et al., 2012; Turner, 2004). Leur participation était particulièrement élevée dans les régions polaires, et leur distribution était principalement influencée par la température, en accord avec des études antérieures (Beaugrand et al., 2009; Pinkerton et al., 2020; Sunday et al., 2012). Ce groupe était suivi par les Eumalacostracés (26,6 %) et les Rhizaires (16,4 %). La distribution verticale et horizontale de ces groupes de Rhizaires suivait la description des études antérieures (Biard et al., 2016; Michaels, 1988; Nakamura et al., 2013; Nakamura and Suzuki, 2015; Suzuki and Not, 2015).

Nous avons également tenté une approche plus régionale en partitionnant le jeu de données à la fois par latitude et par profondeur. Malheureusement, cela a donné des résultats moins satisfaisants que la méthodologie globale, principalement en raison de la diminution de la taille des jeux de données d'apprentissage. Le besoin de disposer de jeux de données suffisamment importants est également ce qui nous a poussés à utiliser une définition large des groupes de zooplancton. À l'avenir, il pourrait être possible d'obtenir une définition taxonomique



et verticale plus fine grâce à la collecte de plus de données. Les modèles plus détaillés pourraient potentiellement obtenir de meilleures performances en fournissant une délimitation plus précise de la structure de l'habitat vertical des groupes taxonomiques. Cela semble être le cas pour les Phaeodaria, où les modèles à résolution de 100 mètres ont donné des résultats de  $R^2$  plus élevés, en particulier dans la plage de profondeur entre 0 et 300 mètres.

Dans le contexte de la modélisation biogéochimique mondiale, les données sur la biomasse du zooplancton générées par cette étude comblent une lacune importante, car ces modèles manquent souvent d'observations sur le zooplancton. À l'avenir, alors que la technologie continue de progresser, le déploiement de systèmes d'imagerie tels que l'UVP6 (Picheral et al., 2021) sur des plateformes autonomes (profileurs BGC-Argo et planeurs) devrait améliorer notre capacité à échantillonner et à surveiller les populations de zooplancton dans une gamme plus large de profondeurs et de régions océaniques, ce qui nous permettra de suivre plus efficacement les changements globaux de la biomasse de zooplancton.

## 6.3 Dynamique de la pompe biologique dans un tourbillon anticyclonique

### 6.3.1 Contexte et méthodes

Les détritiques marins, y compris les agrégats et les pelotes fécales, présentent un grand intérêt pour les scientifiques en raison de leur importance dans la pompe à carbone biologique (Alldredge and Gotschalk, 1988; Boyd et al., 2019). Les agrégats, qui consistent en des amas macroscopiques ( $>500 \mu\text{m}$ , Alldredge and Silver (1988)), sont composés de divers matériaux tels que le phytoplancton, les logettes d'appendiculaires et la matière fécale (Alldredge et al., 1990; Alldredge and Silver, 1988). Ces particules détritiques jouent un rôle crucial dans le transfert du carbone de la surface de l'océan vers ses profondeurs, contribuant ainsi à la pompe gravitationnelle biologique (Boyd et al., 2019; Nowicki et al., 2022). Les agrégats marins présentent des formes diverses (Alldredge, 1998; Alldredge and Silver, 1988) qui résultent de processus physiques (Alldredge, 2001; Kiørboe, 2001; Stemmann, Jackson, and Ianson, 2004) et biologiques complexes (Dilling and Alldredge, 2000; Steinberg and Landry, 2017; Stemmann, Jackson, and Ianson, 2004).

Les processus biologiques affectent de manière significative la morphologie des agrégats. Le zooplancton et les bactéries jouent un rôle

important dans la modification de la structure des agrégats. Le zooplancton contribue à la formation des particules grâce à diverses stratégies d'alimentation et à la production de pelotes fécales compactes (Atkinson et al., 2012; Gleiber et al., 2012). Les agrégats peuvent subir des processus de dégradation et de fragmentation qui peuvent impacter leur porosité et leur taille (Alldredge et al., 1990; Laurenceau-Cornec et al., 2015). Les populations de zooplancton dans les couches de surface et mésopélagique étaient peu nombreuses, jetant des doutes sur leur capacité à influencer de manière significative le flux de carbone par la production de pelotes fécales dans ce jeu de données particulier. Cependant, leur activité alimentaire via la nutrition par filtration (Alldredge et al., 1990; Harbison et al., 1986; Huntley and Boyd, 1984) et la consommation directe des particules (Burd and Jackson, 2009; Gillard et al., 2022; Stemmann, Jackson, and Ianson, 2004), couplée à la fragmentation due à leurs comportements de nage (Dilling and Alldredge, 2000), pourrait expliquer une partie de l'atténuation élevée observée dans la couche entre 50 et 100 mètres, qui ne peut pas être attribuée uniquement à l'activité bactérienne (Belcher et al., 2016; Collins et al., 2015). L'augmentation observée de l'abondance du zooplancton dans cette couche, avec un maximum d'observations de rhizaires à 50 mètres et de crustacés à 125 mètres, constitue une preuve à l'appui de cette hypothèse.

Les études antérieures se sont principalement concentrées sur la quantification de l'abondance, de la vitesse de chute et du flux de carbone (Cael et al., 2021; Guidi et al., 2008; Iversen and Ploug, 2010). Peu d'entre elles ont étudié les changements d'autres traits morphologiques (Trudnowska et al., 2021; Williams and Giering, 2022). Notre étude se concentre sur un tourbillon rétentif anticyclonique dans l'Atlantique Nord dans le cadre du programme EXPORTS (Johnson et al., 2023; Siegel, 2016). Nous avons déployé trois unités UVP5 à partir de navires de recherche pour explorer l'efflorescence printanière de l'Atlantique Nord (Omand et al., 2015) en mai 2021 et avons obtenu 1 720 914 images. Pour traiter ce grand ensemble de données, les images ont été regroupées à l'aide de Morphocluster, un programme conçu pour annoter de grands ensembles d'images présenté dans l'introduction (Schröder et al., 2020). Nos objectifs comprennent la quantification du flux de carbone, la caractérisation de la communauté de particules et l'étude des changements dans la morphologie des agrégats et des pelotes fécales au sein d'un panache d'export en fonction du temps et de la profondeur. Cette étude vise à obtenir une meilleure compréhension de la dynamique et de la distribution des particules en haute mer. Pour cela, nous avons utilisé une approche de sélection du panache d'export de manière oblique (Fig. 6.7) en définissant 5 masques qui

nous ont donc permis de suivre l'export des particules et des pelotes fécales en profondeur et au cours du temps.

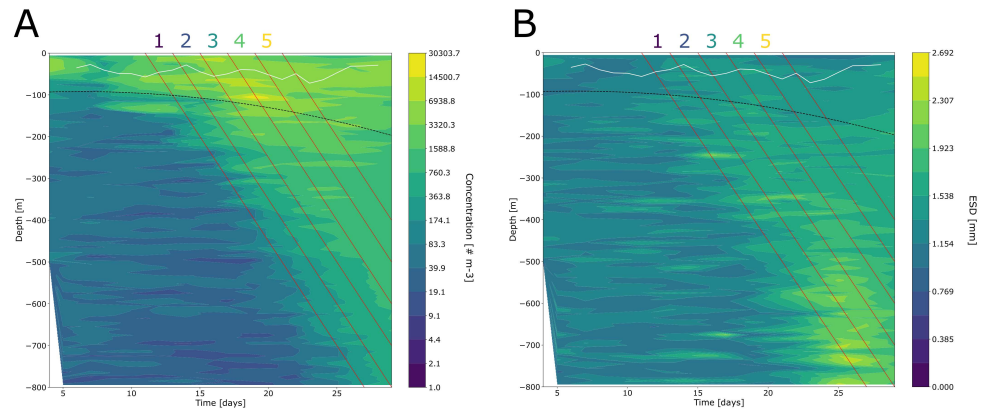


Figure 6.7: Champ interpolé de A) concentration d'agrégats ([nombre par  $m^3$ ]) et B) diamètre sphérique équivalent des agrégats (ESD [mm]). Veuillez noter que la barre de couleur est à l'échelle logarithmique pour la concentration d'agrégats. Les 6 lignes rouges correspondent à la délimitation des 5 masques avec une pente de  $50 \text{ m d}^{-1}$ . La ligne blanche continue correspond à la profondeur moyenne de la couche de mélange. La ligne noire en pointillés correspond à l'ajustement présenté dans Johnson et al. (2023) en tant que limite inférieure de la couche d'eau de base de surface définie en fonction de la profondeur de l'isopycne  $27,1 \text{ kg m}^{-3}$ .

### 6.3.2 Principales conclusions et perspectives

Notre étude a révélé des changements significatifs dans la communauté des particules du panache, indiquant des processus tels que la dégradation, l'agrégation et la reminéralisation microbienne potentielle pour les petites (MiPs, 0,14-0,53 mm de diamètre) et les grandes (MaPs, 0,53-16,88 mm de diamètre) particules. Les spectres de taille des particules (PSD) ont montré un schéma en deux étapes, s'aplatissant sous les eaux centrales de surface du tourbillon (SCW). Notre approche a permis d'estimer de manière fiable le flux d'exportation de carbone, en accord avec les estimations publiées sur l'efflorescence printanière dans l'Atlantique Nord.

L'analyse morphologique des agrégats et des pelotes fécales à l'intérieur du panache par imagerie *in situ* et regroupement a révélé des schémas distincts. Les agrégats sont devenus plus sombres, plus circulaires et plus denses avec la profondeur, suivant une évolution linéaire. Inversement, les pelotes fécales présentaient une dynamique influencée par la production et la dégradation du zooplancton.

Pour mieux comprendre l'atténuation du flux de carbone, nous avons examiné le taux de dégradation spécifique du carbone ( $C_{\text{spec}}$ , Iversen (2023)) correspondant à la somme des processus conduisant à l'atténuation du flux de particules (nutrition et activité natatoire du zooplancton, reminéralisation bactérienne) ainsi que les valeurs du facteur  $b$  issu des régressions utilisant un modèle puissance effectuées sur les valeurs de flux estimées à l'intérieur du panache. Les valeurs  $b$  obtenues dans notre étude indiquent un faible taux d'atténuation du flux de carbone particulaire calculé à l'intérieur du panache d'export et une efficacité élevée de la pompe à carbone biologique qui augmente avec le temps. Sur la base des valeurs de  $C_{\text{spec}}$  précédemment publiées pour la zone (Belcher et al., 2016; Collins et al., 2015) et sur la distribution verticale du zooplancton dans la zone étudiée, nous avons émis l'hypothèse que la reminéralisation microbienne dominait dans la région mésopélagique moyenne à inférieure (300-800 m), tandis que les organismes zooplanctoniques contribuaient probablement à l'atténuation dans la zone mésopélagique supérieure (50-300 m).

Cette étude offre de nouvelles perspectives sur l'étude de la dynamique et de la morphologie des particules qui sédimentent, le flux de carbone et son atténuation au sein d'un tourbillon de l'Atlantique Nord. Les résultats soulignent la nécessité de prendre en compte les caractéristiques morphologiques des particules dans les futurs calculs de flux de carbone et plaident en faveur d'une meilleure connaissance de la complexité des particules dans les écosystèmes marins. Ces résultats ouvrent de nouvelles voies de recherche sur la dynamique des particules et les processus biogéochimiques dans l'océan global.

## 6.4 Estimation des flux de carbone à partir des données de l'UVP

### 6.4.1 Contexte and méthodes

Dans le prolongement du deuxième chapitre, nous utilisons le jeu de données UVP<sub>5</sub> provenant de la campagne scientifique EXPORTS 2021. L'utilisation de cet instrument répond aux limites des instruments traditionnels tels que les pièges à sédiments et les collecteurs de neige marine, qui offrent une couverture spatio-temporelle limitée et risquent d'endommager les particules fragiles (Durkin et al., 2021; Lampitt et al., 1993; Ploug and Jørgensen, 1999). En analysant 4,3 millions d'images, nous avons évalué le type de particule dominant qui affecte le flux et exploré le potentiel de l'UVP<sub>5</sub> pour estimer le flux gravitationnel avec précision par rapport aux mesures des pièges à sédiments. Nous avons sélectionné et étendu la gamme de taille observée par l'UVP<sub>5</sub>

pour correspondre à celle des pièges à sédiments entre 10  $\mu\text{m}$  et 2 mm. Afin d'obtenir la meilleure correspondance entre ces deux instruments, nous avons joué sur les deux autres paramètres qui influencent les estimations de flux de carbone: les relations entre taille de particules et leur contenu en carbone mais également celle entre la taille et la vitesse de sédimentation de ces particules. Pour cela, nous avons estimé le flux de carbone en utilisant diverses relations publiées (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008; Kriest, 2002) et avons exploré une méthode hybride combinant les comptages de l'UVP<sub>5</sub> et l'analyse d'images.

#### 6.4.2 Principales conclusions et perspectives

Cette étude s'est concentrée sur les particules dans le centre du tourbillon, révélant une prédominance des agrégats, notamment aux niveaux de la surface (97,8% des images entre la surface et une profondeur de 1000 mètres ont été annotées comme étant des particules). Notre étude a trouvé une relation positive entre la taille des agrégats et leur vitesse de sédimentation entre 0,81 et 3,25 mm. Cependant, nous n'avons pas pu estimer la vitesse de sédimentation de plus petites particules dû à un manque de données. En comparant nos résultats à des relations précédemment publiées sur le lien entre taille et vitesse de sédimentation (Alldredge and Gotschalk, 1988; Cael et al., 2021; Iversen et al., 2010; Kriest, 2002), il a été avéré que la relation la plus proche de nos estimations correspondait à la relation 9 publiée dans l'article de Kriest (2002).

Pour la méthode hybride utilisant les images et les comptages de particules, nous avons également pris en compte les relations entre la taille et la teneur en carbone des particules des agrégats et pelotes fécales des images en nous appuyant sur des études antérieures (Alldredge, 1998; Durkin et al., 2021). Cela nous a permis de comparer différentes approches pour calculer le flux, en utilisant à la fois les relations entre la taille et la vitesse de sédimentation et les relations entre la taille et la teneur en carbone.

L'extension de la gamme de taille de l'UVP à 10  $\mu\text{m}$ , limitée à 2 mm, a donné la meilleure adéquation entre les données des pièges à sédiments et les estimations de l'UVP<sub>5</sub>. Bien qu'une approche hybride combinant à la fois les comptages de l'UVP<sub>5</sub> et les images d'agrégats et de pelotes fécales ait été testée, elle n'a pas produit de meilleurs résultats par rapport à l'utilisation des seuls comptages de l'UVP<sub>5</sub>. De plus, une partie importante (30-46 %) du flux dérivé de l'UVP<sub>5</sub>, provenant de particules comprises entre 2 et 16 mm, a été estimée comme potentiellement manquée par les pièges à sédiments qui ont

observé des particules entre 10  $\mu\text{m}$  et 2 mm pendant cette campagne. Notre étude discute des combinaisons de paramètres optimales et des implications de ces méthodes sur les estimations du flux de particules.

## 6.5 Discussion

Au cœur de nos travaux sur la Pompe à Carbone Biologique à l'aide de l'imagerie in situ se trouvent deux composants fondamentaux : le zooplancton et les particules. Le zooplancton était au centre de notre étude globale (Chapitre 1) et a également été pris en compte dans une étude régionale (Chapitre 2), tandis que les particules étaient au centre du contexte régional d'un tourbillon de l'Atlantique Nord pendant un bloom printannier (Chapitres 2 et 3).

### 6.5.1 La biogéographie et l'impact écologique du zooplancton

Les études précédentes sur la biogéographie du zooplancton se sont appuyées sur des ensembles de données basées sur des filets (Buitenhuis et al., 2013; Hatton et al., 2021; Moriarty and O'Brien, 2013; Moriarty et al., 2012). La plupart se sont également basées sur des méthodes de proportionnalité, ce qui a également entraîné des biais et des limitations. Notre travail dans le chapitre 1 a proposé l'utilisation d'une technique d'apprentissage automatique pour analyser un ensemble de données d'imagerie in situ mondial UVP5, ce qui a abouti à des cartes de distribution continue de la biomasse de grands groupes du zooplancton pour différentes couches de profondeur. Nous avons fourni une estimation plus robuste de la biomasse du zooplancton et identifié l'importance des rhizaires dans la biomasse mondiale, soulignant la nécessité de méthodes d'imagerie in situ pour l'étude de ces organismes. De plus, nous avons exploré les principales variables influençant la distribution du zooplancton et leurs relations grâce aux figures de dépendance partielle.

Le chapitre 2 s'est concentré sur un tourbillon de l'Atlantique Nord pendant la floraison du printemps, en utilisant Morphocluster pour classifier les images de l'UVP5. Nous avons étudié différentes échelles de profondeur et les dynamiques temporelles au sein de ce phénomène. Nous avons également évalué le rôle potentiel du zooplancton dans l'influence sur la pompe biologique à carbone. Notre étude souligne la complexité du rôle du zooplancton dans le cycle du carbone de l'océan. Alors que notre analyse globale a fourni des informations précieuses sur la biogéographie du zooplancton, l'étude locale a révélé des dynamiques complexes au sein de cet écosystème spécifique.

En conclusion, notre travail contribue de manière significative à notre connaissance de la distribution du zooplancton et de leur impact potentiel sur l'écosystème marin, soulignant l'importance de l'utilisation de techniques avancées et de méthodes d'échantillonnage exhaustives pour étudier efficacement ces organismes.

### 6.5.2 Caractéristiques des particules et modèles de flux

Avant l'expérience EXPORTS (Johnson et al., 2023; Siegel, 2016), il n'y avait eu aucune recherche basée sur l'UVP5 sur un ensemble de données aussi important obtenu en déployant trois unités UVP5 dans une zone géographiquement aussi restreinte. Ce cadre unique au sein d'un tourbillon hautement rétenteur a offert une opportunité exceptionnelle pour des observations in situ de haute qualité, apportant de nouvelles informations sur les caractéristiques des particules et les schémas de flux.

Les précédentes études d'imagerie in situ utilisant l'UVP5 se sont principalement concentrées sur le zooplancton (Biard et al., 2016; L. Stemann et al., 2008; Vilgrain et al., 2021) ou les comptages de particules (Clements et al., 2022, 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008), regroupant souvent les agrégats et les pelotes fécales sous la classification de "détritus". Cependant, cette classification néglige la diversité existant parmi ces particules et leur importance écologique. L'étude de la morphologie des particules et des facteurs qui l'influencent a été confrontée à des défis en raison de la fragilité des agrégats et des limitations dans leur manipulation. Le chapitre 2 visait à combler cette lacune en suivant et en décrivant la morphologie des particules au sein d'un panache exporté, offrant une perspective unique en 2D. Un nouveau cadre oblique a été utilisé pour suivre le panache de particules imagées en profondeur et dans le temps, révélant des patrons distincts pour les agrégats et les pelotes fécales. Contrairement aux observations précédentes, nos résultats indiquent que les agrégats avaient tendance à devenir plus gros mais également plus denses avec la profondeur, peut-être en raison de la chute différentielle (Alldredge, 2001; Burd and Jackson, 2009). Ce processus d'intérêt important est dû à une différence de vitesse de sédimentation entre les pelotes fécales et les agrégats. Il peut entraîner l'intégration des pelotes fécales au sein des agrégats qui peuvent entrer en collision et interagir pendant leur chute. Cela pourrait entraîner la formation d'agrégats plus gros et plus denses, ce qui est cohérent avec certaines des observations faites dans cette étude. Notamment, l'augmentation de la densité et de la taille des agrégats peut avoir un effet significatif sur les taux de sédimentation et la teneur en carbone, potentiellement entraînant une plus grande exportation de carbone de la surface de l'océan vers les couches plus

profondes. Ce processus est donc essentiel à la compréhension des flux de carbone dans l'environnement marin car il est lié à l'efficacité de la pompe biologique à carbone.

L'UVP5 fournit des comptages de particules qui ont historiquement été utilisés pour calculer le flux de particules et l'atténuation dans un cadre vertical unidimensionnel (Clements et al., 2023; Fender et al., 2019; Forest et al., 2013; Guidi et al., 2008), négligeant les caractéristiques dynamiques des communautés de particules à différentes profondeurs. Le chapitre 3 a abordé cette question en évaluant différentes méthodes de calcul du flux et en les comparant aux estimations de flux des pièges à sédiments. La comparaison a révélé des sous-estimations potentielles du flux de carbone allant jusqu'à environ 40% dans certaines gammes de tailles, suggérant que nos méthodes actuelles d'évaluation du flux de carbone dans les écosystèmes marins pourraient être incomplètes. Cela pourrait avoir des implications significatives pour la compréhension de la dynamique du carbone et la résolution des problèmes liés au budget du carbone mésopélagique à l'échelle mondiale (Burd et al., 2010).

De plus, l'évaluation de l'atténuation, un aspect crucial de la dynamique des particules, a produit des résultats significatifs. Le chapitre 2 a montré des niveaux d'atténuation faibles dans la région du panache, principalement attribués aux zooplancton et aux micro-organismes de la zone mésopélagique supérieure, avec une transition vers une prédominance des micro-organismes dans les profondeurs mésopélagiques plus profondes.

### 6.5.3 Contraintes dans l'estimation de la biomasse du zooplancton et la morphologie des particules

L'estimation de la biomasse mondiale du zooplancton dans le chapitre 1 a marqué une avancée significative dans le domaine. Cependant, nous avons rencontré plusieurs défis qui ont conduit à des réflexions importantes. Un défi clé était la disponibilité des données sur la teneur en carbone des différents groupes zooplanctoniques, en particulier pour le groupe des Rhizaria.

Les limitations liées au jeu de données ont également entravé notre capacité à créer des modèles fiables pour divers groupes taxonomiques en raison de signaux peu clairs entre biomasse zooplanctonique et environnement ou même en raison de données pas suffisamment nombreuses. Une étude récente de Laget et al. (in review) a mesuré directement la teneur en carbone des rhizaires, conduisant à l'amélioration des facteurs de conversion pour les modèles d'habitat. Ces modèles mis à jour, basés sur un plus grand nombre de profils UVP5 jusqu'à 1



000 mètres de profondeur, promettent une estimation plus précise de la biomasse mondiale avec une couverture en profondeur plus large (Laget, en révision).

Notre étude sur l'évolution des particules a fourni des informations précieuses sur les changements morphologiques ainsi que sur l'évolution de leur concentration et de leur biovolume. Nous avons utilisé des techniques de regroupement écologiquement pertinentes pour révéler des patrons au sein de la communauté de particules lors de l'expérience EXPORTS. Nos efforts pour décrypter la morphologie des agrégats, bien que laborieux, ont établi un ensemble de données précieux pour les futures études d'images de l'UVP<sub>5</sub> et pourront représenter un jeu de données d'apprentissage pour la classification de ces images.

Dans nos efforts pour calculer le flux à partir de données UVP<sub>5</sub> et le comparer aux pièges à sédiments, nous avons fait une découverte significative. Il a été déterminé que la correspondance la plus précise entre les données UVP et les données des pièges à sédiments était obtenue en utilisant la même gamme de taille que celle couverte par les pièges à sédiments (10 µm-2 mm). Cette harmonisation de la gamme de taille améliore la fiabilité de nos calculs de flux, renforçant ainsi l'utilité de la technologie UVP<sub>5</sub> dans l'étude de la dynamique des particules.

Enfin, notre enquête a présenté un nouveau cadre pour l'étude des données UVP dans le contexte d'un événement d'exportation. Le suivi de la dynamique de la communauté des particules à travers ce cadre oblique nous a permis d'acquérir une compréhension plus profonde des processus complexes qui régissent la dynamique des particules lors des événements d'export.

En conclusion, cette collection d'études repousse non seulement les limites de la connaissance, mais elle aborde également les défis cruciaux liés à l'estimation de la biomasse du zooplancton et à la compréhension de l'évolution des morphologies et des flux de particules dans le contexte d'un événement d'export. Ces développements ouvrent de nouvelles voies pour la recherche future et ouvrent la voie à une compréhension plus profonde des écosystèmes marins.

## 6.6 Perspectives

Dans cette section, nous élargissons la portée de notre discussion, allant au-delà des conclusions immédiates de cette thèse pour explorer des questions plus larges liées à la recherche basée sur les données. Bien

que notre accent principal reste sur la pompe biologique et ses aspects liés à l'imagerie, nous reconnaissons également la pertinence de ces discussions dans le contexte de l'avancement de la science dans le domaine de l'écologie marine.

#### 6.6.1 Défis et opportunités des données en écologie

L'écologie axée sur les données révolutionne la recherche scientifique mais exige d'importantes ressources. La quantité colossale de données générées et analysées, appelée "big data" (Guidi et al., 2020), pose des défis en matière de gestion des données, conformément aux principes FAIR (findable, accessible, interoperable, reusable c'est à dire facilement trouvable, accessible, interopérable et réutilisable) Wilkinson et al. (2016). La gestion du "big data" nécessite une augmentation des ressources liées à leur stockage et à la puissance de calcul nécessaire à leur étude, suscitant des préoccupations en matière de durabilité et d'accès équitable. Ce besoin accru en ressources soulève des questions cruciales concernant l'équilibre entre la durabilité, l'accessibilité économique et l'accès équitable aux outils scientifiques de pointe et aux connaissances. Face à ces défis, des outils innovants et relativement faiblement coûteux comparés aux instruments utilisés actuellement ont été développés, tels que le PlanktoScope, une plateforme d'imagerie compacte et modulaire (Pollina et al., 2022), ou le Foldscope, un microscope bas coût basé sur les principes d'origami (Cybulski et al., 2014). Trouver un équilibre entre le potentiel du "big data", les contraintes de ressources limitées et les considérations éthiques est essentiel pour le succès de l'écologie axée sur les données et son inclusivité dans l'exploration scientifique et la conservation de l'environnement.

#### 6.6.2 Le future de l'étude du zooplancton

L'exploration de l'avenir de la distribution de la biomasse du zooplancton ouvre des perspectives passionnantes. Le déploiement en cours d'unités d'UVP5 ainsi que les déploiements autonomes d'UVP6 (Picheral et al., 2021) sur des flotteurs et des mouillages promet d'élargir les jeux de données, ce qui pourrait améliorer les prédictions de biomasse annuelles et potentiellement permettre leur modélisation saisonnière. Le déploiement d'unités d'UVP6 sur des mouillages a pour but d'obtenir des séries temporelles de données à long terme, fournissant des informations sur les dynamiques locales au fil du temps.

Quant à l'utilisation de l'UVP5, nous encourageons l'expansion systématique, si possible, des profils verticaux pour couvrir la plage de profondeur de 0 à 1000 mètres. Ceci pourrait permettre d'obtenir une

compréhension plus complète de la distribution du zooplancton dans l'ensemble de la colonne d'eau.

### 6.6.3 Perspectives futures pour l'approche de suivi des particules

L'étude de l'export du carbone grâce à l'approche de suivi du panache de particules ouvre plusieurs pistes de recherche pour l'avenir. Des jeux de données plus longs comprenant plusieurs événements d'export peuvent être acquis en déployant des unités UVP6 sur des mouillages ou des flotteurs BGC Argo dans différents environnements biogéochimiques. Ces jeux de données pourraient permettre d'obtenir des informations sur la dynamique de l'export du carbone au fil du temps et dans des conditions variables en appliquant notre méthode de suivi du panache.

Établir un lien direct entre les vitesses de sédimentation, la morphologie et la teneur en carbone des agrégats pourrait également faire progresser notre compréhension des processus d'export du carbone. Pour ce faire, le déploiement simultané de l'UVP5 et de pièges à sédiments pourrait fournir des jeux de données importants pour améliorer la précision des relations entre la taille et le contenu en carbone, améliorant ainsi notre compréhension de leur rôle dans l'exportation du carbone.



# Bibliography

---

- Alkire, M. B. et al. (2012). "Estimates of Net Community Production and Export Using High-Resolution, Lagrangian Measurements of O<sub>2</sub>, NO<sub>3</sub><sup>-</sup>, and POC through the Evolution of a Spring Diatom Bloom in the North Atlantic." In: *Deep Sea Research Part I: Oceanographic Research Papers* 64, pp. 157–174. ISSN: 09670637. DOI: [10.1016/j.dsr.2012.01.012](https://doi.org/10.1016/j.dsr.2012.01.012).
- Allredge (1998). "The Carbon, Nitrogen and Mass Content of Marine Snow as a Function of Aggregate Size." In: *Deep Sea Research Part I: Oceanographic Research Papers* 45.4-5, pp. 529–541. ISSN: 09670637. DOI: [10.1016/S0967-0637\(97\)00048-4](https://doi.org/10.1016/S0967-0637(97)00048-4).
- (2001). "Particle Aggregation Dynamics." In: *Encyclopedia of Ocean Sciences*. Elsevier, pp. 2090–2097. ISBN: 978-0-12-227430-5. DOI: [10.1006/rwos.2001.0468](https://doi.org/10.1006/rwos.2001.0468).
- Allredge, G. Gorsky, M. Youngbluth, and D. Deibel (2005). "The Contribution of Discarded Appendicularian Houses to the Flux of Particulate Organic Carbon from Oceanic Surface Waters." In: *Response of marine ecosystems to global change: Ecological impact of appendicularians*, pp. 309–326.
- Allredge and C. Gotschalk (1988). "In Situ Settling Behavior of Marine Snow<sub>1</sub>: Sinking Rates of Marine Snow." In: *Limnol. Oceanogr.* 33.3, pp. 339–351. ISSN: 00243590. DOI: [10.4319/lo.1988.33.3.0339](https://doi.org/10.4319/lo.1988.33.3.0339).
- Allredge, C. C. Gotschalk, and S. MacIntyre (1987). "Evidence for Sustained Residence of Macrocrustacean Fecal Pellets in Surface Waters off Southern California." In: *Deep Sea Research Part A. Oceanographic Research Papers* 34.9, pp. 1641–1652. ISSN: 01980149. DOI: [10.1016/0198-0149\(87\)90113-0](https://doi.org/10.1016/0198-0149(87)90113-0).
- Allredge, T. C. Granata, C. C. Gotschalk, and T. D. Dickey (1990). "The Physical Strength of Marine Snow and Its Implications for Particle Disaggregation in the Ocean." In: *Limnol. Oceanogr.* 35.7, pp. 1415–1428. ISSN: 00243590. DOI: [10.4319/lo.1990.35.7.1415](https://doi.org/10.4319/lo.1990.35.7.1415).
- Allredge and Silver (1988). "Characteristics, Dynamics and Significance of Marine Snow." In: *Progress in Oceanography* 20.1, pp. 41–82. ISSN: 00796611. DOI: [10.1016/0079-6611\(88\)90053-5](https://doi.org/10.1016/0079-6611(88)90053-5).
- Amano, C., Z. Zhao, E. Sintes, T. Reinthaler, J. Stefanschitz, M. Kisadur, M. Utsumi, and G. J. Herndl (2022). "Limited Carbon Cycling Due to High-Pressure Effects on the Deep-Sea Microbiome." In: *Nat. Geosci.* 15.12, pp. 1041–1047. ISSN: 1752-0894, 1752-0908. DOI: [10.1038/s41561-022-01081-3](https://doi.org/10.1038/s41561-022-01081-3).

- Amante, C. and B. Eakins (2009). *ETOPO1 1 Arc-Minute Global Relief: Procedures, Data Sources and Analysis*. NOAA Technical Memorandum NESDIS NGDC-24.
- Antezana, T. (2010). "Euphausia Mucronata: A Keystone Herbivore and Prey of the Humboldt Current System." In: *Deep Sea Research Part II: Topical Studies in Oceanography* 57.7-8, pp. 652–662. ISSN: 09670645. DOI: [10.1016/j.dsr2.2009.10.014](https://doi.org/10.1016/j.dsr2.2009.10.014).
- Atkinson, A., K. Schmidt, S. Fielding, S. Kawaguchi, and P. Geissler (2012). "Variable Food Absorption by Antarctic Krill: Relationships between Diet, Egestion Rate and the Composition and Sinking Rates of Their Fecal Pellets." In: *Deep Sea Research Part II: Topical Studies in Oceanography* 59–60, pp. 147–158. ISSN: 09670645. DOI: [10.1016/j.dsr2.2011.06.008](https://doi.org/10.1016/j.dsr2.2011.06.008).
- Bach, L. T., P. Stange, J. Taucher, E. P. Achterberg, M. Algueró-Muñiz, H. Horn, M. Esposito, and U. Riebesell (2019). "The Influence of Plankton Community Structure on Sinking Velocity and Remineralization Rate of Marine Aggregates." In: *Global Biogeochem. Cycles* 33.8, pp. 971–994. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2019GB006256](https://doi.org/10.1029/2019GB006256).
- Bagniewski, W., K. Fennel, M. J. Perry, and E. A. D'Asaro (2011). "Optimizing Models of the North Atlantic Spring Bloom Using Physical, Chemical and Bio-Optical Observations from a Lagrangian Float." In: *Biogeosciences* 8.5, pp. 1291–1307. ISSN: 1726-4189. DOI: [10.5194/bg-8-1291-2011](https://doi.org/10.5194/bg-8-1291-2011).
- Baudena, A. et al. (in review). "Five Months in the Life of an Oceanic Eddy: Carbon Flux and Oxygen Demand." In.
- Beaugrand, G., C. Luczak, and M. Edwards (2009). "Rapid Biogeographical Plankton Shifts in the North Atlantic Ocean." In: *Global Change Biology* 15.7, pp. 1790–1803. ISSN: 13541013, 13652486. DOI: [10.1111/j.1365-2486.2009.01848.x](https://doi.org/10.1111/j.1365-2486.2009.01848.x).
- Beers, J. R. and G. L. Stewart (1970). "The Preservation of Acantharians in Fixed Plankton Samples." In: *Limnol. Oceanogr.* 15.5, pp. 825–827. ISSN: 00243590. DOI: [10.4319/lo.1970.15.5.0825](https://doi.org/10.4319/lo.1970.15.5.0825).
- Belcher, Iversen, Giering, V. Riou, S. A. Henson, L. Berline, L. Guilloux, and R. Sanders (2016). "Depth-Resolved Particle-Associated Microbial Respiration in the northeast Atlantic." In: *Biogeosciences* 13.17, pp. 4927–4943. ISSN: 1726-4189. DOI: [10.5194/bg-13-4927-2016](https://doi.org/10.5194/bg-13-4927-2016).
- Benedetti, F. et al. (2019). "The Seasonal and Inter-Annual Fluctuations of Plankton Abundance and Community Structure in a North Atlantic Marine Protected Area." In: *Front. Mar. Sci.* 6, p. 214. ISSN: 2296-7745. DOI: [10.3389/fmars.2019.00214](https://doi.org/10.3389/fmars.2019.00214).
- Bernstein, R. E., P. R. Betzer, R. A. Feely, R. H. Byrne, M. F. Lamb, and A. F. Michaels (1987). "Acantharian Fluxes and Strontium to Chlorinity Ratios in the North Pacific Ocean." In: *Science* 237.4821, pp. 1490–1494. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.237.4821.1490](https://doi.org/10.1126/science.237.4821.1490).

- Bérubé, D. and M. Jébrak (1999). *High Precision Boundary Fractal Analysis for Shape Characterization*. <https://reader.elsevier.com/reader/sd/pii/S0098300499000679?token=west-1&originCreation=20230410171809>. DOI: [10.1016/S0098-3004\(99\)00067-9](https://doi.org/10.1016/S0098-3004(99)00067-9).
- Bianchi, D., C. Stock, E. D. Galbraith, and J. L. Sarmiento (2013). "Diel Vertical Migration: Ecological Controls and Impacts on the Biological Pump in a One-Dimensional Ocean Model: DIEL VERTICAL MIGRATION IMPACTS." In: *Global Biogeochem. Cycles* 27.2, pp. 478–491. ISSN: 08866236. DOI: [10.1002/gbc.20031](https://doi.org/10.1002/gbc.20031).
- Biard, T. and M. D. Ohman (2020). "Vertical Niche Definition of Test-bearing Protists (Rhizaria) into the Twilight Zone Revealed by in Situ Imaging." In: *Limnol Oceanogr* 65.11, pp. 2583–2602. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.11472](https://doi.org/10.1002/lno.11472).
- Biard, T., L. Stemmann, M. Picheral, N. Mayot, P. Vandromme, H. Hauss, G. Gorsky, L. Guidi, R. Kiko, and F. Not (2016). "In Situ Imaging Reveals the Biomass of Giant Protists in the Global Ocean." In: *Nature* 532.7600, pp. 504–507. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature17652](https://doi.org/10.1038/nature17652).
- Boltovskoy, D., O. R. Anderson, and N. M. Correa (2017). "Radiolaria and Phaeodaria." In: *Handbook of the Protists*. Ed. by J. M. Archibald, A. G. Simpson, and C. H. Slamovits. Cham: Springer International Publishing, pp. 731–763. ISBN: 978-3-319-28147-6 978-3-319-28149-0. DOI: [10.1007/978-3-319-28149-0\\_19](https://doi.org/10.1007/978-3-319-28149-0_19).
- Bottazzi, E. M. and M. G. Andreoli (1982). "Distribution of Adult and Juvenile Acantharia (Protozoa Sarcodina) in the Atlantic Ocean." In: p. 22.
- Boyd, H. Claustre, M. Levy, D. A. Siegel, and T. Weber (2019). "Multi-Faceted Particle Pumps Drive Carbon Sequestration in the Ocean." In: *Nature* 568.7752, pp. 327–335. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/s41586-019-1098-2](https://doi.org/10.1038/s41586-019-1098-2).
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984). "Classification and Regression Trees. Wadsworth & Brooks." In: *Cole Statistics/Probability Series*.
- Briggs, N., G. Dall'Olmo, and H. Claustre (2020). "Major Role of Particle Fragmentation in Regulating Biological Sequestration of CO<sub>2</sub> by the Oceans." In: *Science* 367.6479, pp. 791–793. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.aay1790](https://doi.org/10.1126/science.aay1790).
- Briggs, N., M. J. Perry, I. Cetinić, C. Lee, E. D'Asaro, A. M. Gray, and E. Rehm (2011). "High-Resolution Observations of Aggregate Flux during a Sub-Polar North Atlantic Spring Bloom." In: *Deep Sea Research Part I: Oceanographic Research Papers* 58.10, pp. 1031–1039. ISSN: 09670637. DOI: [10.1016/j.dsr.2011.07.007](https://doi.org/10.1016/j.dsr.2011.07.007).
- Brinton, E. (1967). "Vertical Migration and Avoidance Capability of Euphausiids in the California Current." In: *Limnology and Oceanography* 12.3, pp. 451–483. ISSN: 1939-5590. DOI: [10.4319/lo.1967.12.3.0451](https://doi.org/10.4319/lo.1967.12.3.0451).

- Brodeur, R., I. Perry, J. Boldt, L. Flostrand, M. Galbraith, J. King, J. Murphy, K. Sakuma, and A. Thompson (2018). "An Unusual Gelatinous Plankton Event in the NE Pacific: The Great Pyrosome Bloom of 2017." In: 26.1, p. 6.
- Brun, P., M. R. Payne, and T. Kiørboe (2016). "Trait Biogeography of Marine Copepods - an Analysis across Scales." In: *Ecol Lett* 19.12. Ed. by P. Jeyasingh, pp. 1403–1413. ISSN: 1461023X. DOI: [10.1111/ele.12688](https://doi.org/10.1111/ele.12688).
- Buesseler, M. P. Bacon, J. Kirk Cochran, and H. D. Livingston (1992). "Carbon and Nitrogen Export during the JGOFS North Atlantic Bloom Experiment Estimated from  $^{234}\text{Th}$ : $^{238}\text{U}$  Disequilibria." In: *Deep-Sea Research* 39.7/8, pp. 1115–1137.
- Buesseler, K. O. and Boyd (2009). "Shedding Light on Processes That Control Particle Export and Flux Attenuation in the Twilight Zone of the Open Ocean." In: *Limnology and Oceanography* 54.4, pp. 1210–1232. ISSN: 1939-5590. DOI: [10.4319/lo.2009.54.4.1210](https://doi.org/10.4319/lo.2009.54.4.1210).
- Buesseler, K. O., P. W. Boyd, E. E. Black, and D. A. Siegel (2020). "Metrics That Matter for Assessing the Ocean Biological Carbon Pump." In: *Proc. Natl. Acad. Sci. U.S.A.* 117.18, pp. 9679–9687. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1918114117](https://doi.org/10.1073/pnas.1918114117).
- Buitenhuis, E. T. et al. (2013). "MAREDAT: Towards a World Atlas of MARine Ecosystem DATA." In: *Earth Syst. Sci. Data* 5.2, pp. 227–239. ISSN: 1866-3516. DOI: [10.5194/essd-5-227-2013](https://doi.org/10.5194/essd-5-227-2013).
- Burd, A. B. and Jackson (2009). "Particle Aggregation." In: *Annu. Rev. Mar. Sci.* 1.1, pp. 65–90. ISSN: 1941-1405, 1941-0611. DOI: [10.1146/annurev.marine.010908.163904](https://doi.org/10.1146/annurev.marine.010908.163904).
- Burd, A. B. et al. (2010). "Assessing the Apparent Imbalance between Geochemical and Biochemical Indicators of Meso- and Bathypelagic Biological Activity: What the @\$#! Is Wrong with Present Calculations of Carbon Budgets?" In: *Deep Sea Research Part II: Topical Studies in Oceanography* 57.16, pp. 1557–1571. ISSN: 09670645. DOI: [10.1016/j.dsr2.2010.02.022](https://doi.org/10.1016/j.dsr2.2010.02.022).
- Cael, B. B., E. L. Cavan, and G. L. Britten (2021). "Reconciling the Size-Dependence of Marine Particle Sinking Speed." In: *Geophysical Research Letters* 48.5. ISSN: 0094-8276, 1944-8007. DOI: [10.1029/2020GL091771](https://doi.org/10.1029/2020GL091771).
- Castellanos, I., E. Suárez-Morales, and Á. Álvaro Morales-Ramírez (2009). "Euphausiids." In: *Marine Biodiversity of Costa Rica, Central America*. Ed. by I. S. Wehrmann and J. Cortés. Vol. 86. Dordrecht: Springer Netherlands, pp. 199–207. ISBN: 978-1-4020-8277-1 978-1-4020-8278-8. DOI: [10.1007/978-1-4020-8278-8\\_18](https://doi.org/10.1007/978-1-4020-8278-8_18).
- Cavan, E. and Boyd (2018). "Effect of Anthropogenic Warming on Microbial Respiration and Particulate Organic Carbon Export Rates in the Sub-Antarctic Southern Ocean." In: *Aquat. Microb. Ecol.* 82.2, pp. 111–127. ISSN: 0948-3055, 1616-1564. DOI: [10.3354/ame01889](https://doi.org/10.3354/ame01889).



- Cavan, S. A. Henson, A. Belcher, and R. Sanders (2017). "Role of Zooplankton in Determining the Efficiency of the Biological Carbon Pump." In: *Biogeosciences* 14.1, pp. 177–186. ISSN: 1726-4189. DOI: [10.5194/bg-14-177-2017](https://doi.org/10.5194/bg-14-177-2017).
- Chavez, F. P., A. Bertrand, R. Guevara-Carrasco, P. Soler, and J. Csirke (2008). "The Northern Humboldt Current System: Brief History, Present Status and a View towards the Future." In: *Progress in Oceanography* 79.2-4, pp. 95–105. ISSN: 00796611. DOI: [10.1016/j.pocean.2008.10.012](https://doi.org/10.1016/j.pocean.2008.10.012).
- Chen, H. Liu, W. Xiao, L. Wang, and B. Huang (2020). "A Machine-Learning Approach to Modeling Picophytoplankton Abundances in the South China Sea." In: *Progress in Oceanography* 189, p. 102456. ISSN: 00796611. DOI: [10.1016/j.pocean.2020.102456](https://doi.org/10.1016/j.pocean.2020.102456).
- Chen et al. (2021). *Xgboost: Extreme Gradient Boosting*.
- Christiansen, S. et al. (2018). "Particulate Matter Flux Interception in Oceanic Mesoscale Eddies by the Polychaete *Poecobius* Sp." In: *Limnol. Oceanogr.* 63.5, pp. 2093–2109. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.10926](https://doi.org/10.1002/lno.10926).
- Clements, D. J., S. Yang, T. Weber, A. M. P. McDonnell, R. Kiko, L. Stemmann, and D. Bianchi (2022). "Constraining the Particle Size Distribution of Large Marine Particles in the Global Ocean With *In Situ* Optical Observations and Supervised Learning." In: *Global Biogeochemical Cycles* 36.5. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2021GB007276](https://doi.org/10.1029/2021GB007276).
- (2023). "New Estimate of Organic Carbon Export From Optical Measurements Reveals the Role of Particle Size Distribution and Export Horizon." In: *Global Biogeochemical Cycles* 37.3, e2022GB007633. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2022GB007633](https://doi.org/10.1029/2022GB007633).
- Collins, J. R., B. R. Edwards, K. Thamatrakoln, J. E. Ossolinski, G. R. DiTullio, K. D. Bidle, S. C. Doney, and B. A. S. Van Mooy (2015). "The Multiple Fates of Sinking Particles in the North Atlantic Ocean." In: *Global Biogeochem. Cycles* 29.9, pp. 1471–1494. ISSN: 0886-6236, 1944-9224. DOI: [10.1002/2014GB005037](https://doi.org/10.1002/2014GB005037).
- Cowen, R. K. and C. M. Guigand (2008). "In Situ Ichthyoplankton Imaging System ( I SIIS): System Design and Preliminary Results: In Situ Ichthyoplankton Imaging System." In: *Limnol. Oceanogr. Methods* 6.2, pp. 126–132. ISSN: 15415856. DOI: [10.4319/lom.2008.6.126](https://doi.org/10.4319/lom.2008.6.126).
- Culverhouse, P. F. (2007). "Human and Machine Factors in Algae Monitoring Performance." In: *Ecological Informatics* 2.4, pp. 361–366. ISSN: 15749541. DOI: [10.1016/j.ecoinf.2007.07.001](https://doi.org/10.1016/j.ecoinf.2007.07.001).
- Culverhouse, P. F., N. Macleod, R. Williams, M. C. Benfield, R. M. Lopes, and M. Picheral (2014). "An Empirical Assessment of the Consistency of Taxonomic Identifications." In: *Marine Biology Research* 10.1, pp. 73–84. ISSN: 1745-1000, 1745-1019. DOI: [10.1080/17451000.2013.810762](https://doi.org/10.1080/17451000.2013.810762).

- Cybulski, J. S., J. Clements, and M. Prakash (2014). "Foldscope: Origami-Based Paper Microscope." In: *PLoS ONE* 9.6. Ed. by L. Martens, e98781. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0098781](https://doi.org/10.1371/journal.pone.0098781).
- Dagg, M. J. and W. E. Walser (1986). "The Effect of Food Concentration on Fecal Pellet Size in Marine Copepods<sub>1</sub>: Fecal Pellet Size." In: *Limnol. Oceanogr.* 31.5, pp. 1066–1071. ISSN: 00243590. DOI: [10.4319/lo.1986.31.5.1066](https://doi.org/10.4319/lo.1986.31.5.1066).
- Dai, L., C. Li, G. Yang, and X. Sun (2016). "Zooplankton Abundance, Biovolume and Size Spectra at Western Boundary Currents in the Subtropical North Pacific during Winter 2012." In: *Journal of Marine Systems* 155, pp. 73–83. ISSN: 09247963. DOI: [10.1016/j.jmarsys.2015.11.004](https://doi.org/10.1016/j.jmarsys.2015.11.004).
- Davis, C. S., S. M. Gallagher, and A. R. Solow (1992). "Microaggregations of Oceanic Plankton Observed by Towed Video Microscopy." In: *Science* 257.5067, pp. 230–232. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.257.5067.230](https://doi.org/10.1126/science.257.5067.230).
- Davis, C. S., F. T. Thwaites, S. M. Gallagher, and Q. Hu (2005). "A Three-Axis Fast-Tow Digital Video Plankton Recorder for Rapid Surveys of Plankton Taxa and Hydrography: New Video Plankton Recorder." In: *Limnol. Oceanogr. Methods* 3.2, pp. 59–74. ISSN: 15415856. DOI: [10.4319/lom.2005.3.59](https://doi.org/10.4319/lom.2005.3.59).
- de Vargas, C. et al. (2015). "Eukaryotic Plankton Diversity in the Sunlit Ocean." In: *Science* 348.6237, pp. 1261605–1261605. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1261605](https://doi.org/10.1126/science.1261605).
- De'ath, G. (2007). "Boosted Trees for Ecological Modeling and Prediction." In: *Ecology* 88.1, pp. 243–251. ISSN: 0012-9658. DOI: [10.1890/0012-9658\(2007\)88\[243:BTFEMA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2).
- Decelle, J., P. Martin, K. Paborstava, D. W. Pond, G. Tarling, F. Mahé, C. de Vargas, R. Lampitt, and F. Not (2013). "Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans." In: *PLoS ONE* 8.1. Ed. by L. J. Stal, e53598. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0053598](https://doi.org/10.1371/journal.pone.0053598).
- Decelle, J. and F. Not (2015). "Acantharia." In: *eLS*. Ed. by John Wiley & Sons, Ltd. 1st ed. Wiley, pp. 1–10. ISBN: 978-0-470-01617-6 978-0-470-01590-2. DOI: [10.1002/9780470015902.a0002102.pub2](https://doi.org/10.1002/9780470015902.a0002102.pub2).
- DeVries, T., F. Primeau, and C. Deutsch (2012). "The Sequestration Efficiency of the Biological Pump: BIOLOGICAL PUMP SEQUESTRATION EFFICIENCY." In: *Geophys. Res. Lett.* 39.13, n/a–n/a. ISSN: 00948276. DOI: [10.1029/2012GL051963](https://doi.org/10.1029/2012GL051963).
- DeVries, T. and T. Weber (2017). "The Export and Fate of Organic Matter in the Ocean: New Constraints from Combining Satellite and Oceanographic Tracer Observations: EXPORT AND FATE OF MARINE ORGANIC MATTER." In: *Global Biogeochem. Cycles* 31.3, pp. 535–555. ISSN: 08866236. DOI: [10.1002/2016GB005551](https://doi.org/10.1002/2016GB005551).

- Dilling, L. (2004). "Quantifying Marine Snow as a Food Choice for Zooplankton Using Stable Silicon Isotope Tracers." In: *Journal of Plankton Research* 26.9, pp. 1105–1114. ISSN: 1464-3774. DOI: [10.1093/plankt/fbh103](https://doi.org/10.1093/plankt/fbh103).
- Dilling, L. and Alldredge (2000). "Fragmentation of Marine Snow by Swimming Macrozooplankton: A New Process Impacting Carbon Cycling in the Sea." In: *Deep Sea Research Part I: Oceanographic Research Papers* 47.7, pp. 1227–1245. ISSN: 09670637. DOI: [10.1016/S0967-0637\(99\)00105-3](https://doi.org/10.1016/S0967-0637(99)00105-3).
- Drago, L. et al. (2022). "Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning." In: *Front. Mar. Sci.* 9, p. 894372. ISSN: 2296-7745. DOI: [10.3389/fmars.2022.894372](https://doi.org/10.3389/fmars.2022.894372).
- Dugenne, M., M. Corrales, L. Stemmann, C. Stock, T. O'Brien, F. Lombard, J. O. Irisson, R. Kiko, and J. Y. Luo (2023). *Zooplankton Size Distribution in the Atlantic Ocean: Comparison with Recent Estimates from the Global Pelagic Size Structure Database and Investigation of Environmental Links*.
- Durkin, C. A., K. O. Buesseler, I. Cetinić, M. L. Estapa, R. P. Kelly, and M. Omand (2021). "A Visual Tour of Carbon Export by Sinking Particles." In: *Global Biogeochem Cycles* 35.10. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2021GB006985](https://doi.org/10.1029/2021GB006985).
- Durkin, C. A., M. L. Estapa, and K. O. Buesseler (2015). "Observations of Carbon Export by Small Sinking Particles in the Upper Mesopelagic." In: *Marine Chemistry* 175, pp. 72–81. ISSN: 03044203. DOI: [10.1016/j.marchem.2015.02.011](https://doi.org/10.1016/j.marchem.2015.02.011).
- Edwards, M., P. Hélaouët, E. Goberville, A. Lindley, G. A. Tarling, M. T. Burrows, and A. Atkinson (2021). "North Atlantic Warming over Six Decades Drives Decreases in Krill Abundance with No Associated Range Shift." In: *Commun Biol* 4.1, p. 644. ISSN: 2399-3642. DOI: [10.1038/s42003-021-02159-1](https://doi.org/10.1038/s42003-021-02159-1).
- Elith, J., J. R. Leathwick, and T. Hastie (2008). "A Working Guide to Boosted Regression Trees." In: *J Anim Ecology* 77.4, pp. 802–813. ISSN: 0021-8790, 1365-2656. DOI: [10.1111/j.1365-2656.2008.01390.x](https://doi.org/10.1111/j.1365-2656.2008.01390.x).
- Elith, J. and C. H. Graham (2009). "Do They? How Do They? WHY Do They Differ? On Finding Reasons for Differing Performances of Species Distribution Models." In: *Ecography* 32.1, pp. 66–77. ISSN: 09067590, 16000587. DOI: [10.1111/j.1600-0587.2008.05505.x](https://doi.org/10.1111/j.1600-0587.2008.05505.x).
- Elith, J. and Leathwick (2009). "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." In: *Annu. Rev. Ecol. Evol. Syst.* 40.1, pp. 677–697. ISSN: 1543-592X, 1545-2069. DOI: [10.1146/annurev.ecolsys.110308.120159](https://doi.org/10.1146/annurev.ecolsys.110308.120159).
- Elith, J. et al. (2006). "Novel Methods Improve Prediction of Species' Distributions from Occurrence Data." In: *Ecography* 29.2, pp. 129–151. ISSN: 09067590. DOI: [10.1111/j.2006.0906-7590.04596.x](https://doi.org/10.1111/j.2006.0906-7590.04596.x).

- Erickson, Z. et al. (2022). *EXPORTS North Atlantic Eddy Tracking*. Tech. rep. NASA/TM-20220009705. NASA Technical Memorandum.
- Faillietaz, R., M. Picheral, J. Y. Luo, C. Guigand, R. K. Cowen, and J.-O. Irisson (2016). "Imperfect Automatic Image Classification Successfully Describes Plankton Distribution Patterns." In: *Methods in Oceanography* 15–16, pp. 60–77. ISSN: 22111220. DOI: [10.1016/j.mio.2016.04.003](https://doi.org/10.1016/j.mio.2016.04.003).
- Feinberg, L. and H. Dam (1998). "Effects of Diet on Dimensions, Density and Sinking Rates of Fecal Pellets of the Copepod *Acartia Tonsa*." In: *Mar. Ecol. Prog. Ser.* 175, pp. 87–96. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps175087](https://doi.org/10.3354/meps175087).
- Fender, C. K., T. B. Kelly, L. Guidi, M. D. Ohman, M. C. Smith, and M. R. Stukel (2019). "Investigating Particle Size-Flux Relationships and the Biological Pump Across a Range of Plankton Ecosystem States From Coastal to Oligotrophic." In: *Front. Mar. Sci.* 6, p. 603. ISSN: 2296-7745. DOI: [10.3389/fmars.2019.00603](https://doi.org/10.3389/fmars.2019.00603).
- Forest, A., L. Stemmann, M. Picheral, L. Burdorf, D. Robert, L. Fortier, and M. Babin (2012). "Size Distribution of Particles and Zooplankton across the Shelf-Basin System in Southeast Beaufort Sea: Combined Results from an Underwater Vision Profiler and Vertical Net Tows." In: *Biogeosciences* 9.4, pp. 1301–1320. ISSN: 1726-4189. DOI: [10.5194/bg-9-1301-2012](https://doi.org/10.5194/bg-9-1301-2012).
- Forest, A. et al. (2013). "Ecosystem Function and Particle Flux Dynamics across the Mackenzie Shelf (Beaufort Sea, Arctic Ocean): An Integrative Analysis of Spatial Variability and Biophysical Forcings." In: *Biogeosciences* 10.5, pp. 2833–2866. ISSN: 1726-4189. DOI: [10.5194/bg-10-2833-2013](https://doi.org/10.5194/bg-10-2833-2013).
- Frederiksen, M., M. Edwards, A. J. Richardson, N. C. Halliday, and S. Wanless (2006). "From Plankton to Top Predators: Bottom-up Control of a Marine Food Web across Four Trophic Levels." In: *J Anim Ecology* 75.6, pp. 1259–1268. ISSN: 0021-8790, 1365-2656. DOI: [10.1111/j.1365-2656.2006.01148.x](https://doi.org/10.1111/j.1365-2656.2006.01148.x).
- Friedman, J. H. (2002). "Stochastic Gradient Boosting." In: *Computational Statistics & Data Analysis* 38.4, pp. 367–378. ISSN: 01679473. DOI: [10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- Frost, B. W. (1972). "Effects of Size and Concentration of Food Particles on the Feeding Behavior of the Marine Planktonic Copepod *Calanus Pacificus* 1:Feeding Behavior of *Calanus*." In: *Limnol. Oceanogr.* 17.6, pp. 805–815. ISSN: 00243590. DOI: [10.4319/lo.1972.17.6.0805](https://doi.org/10.4319/lo.1972.17.6.0805).
- Gallienne, C. P. (2001). "Is Oithona the Most Important Copepod in the World's Oceans?" In: *Journal of Plankton Research* 23.12, pp. 1421–1432. ISSN: 14643774. DOI: [10.1093/plankt/23.12.1421](https://doi.org/10.1093/plankt/23.12.1421).
- Garcia, H. et al. (2019). *World Ocean Atlas 2018: Product Documentation*.
- Giering, S. L. C., B. Hosking, N. Briggs, and M. H. Iversen (2020). "The Interpretation of Particle Size, Shape, and Carbon Flux of Marine

- Particle Images Is Strongly Affected by the Choice of Particle Detection Algorithm." In: *Front. Mar. Sci.* 7, p. 564. ISSN: 2296-7745. DOI: [10.3389/fmars.2020.00564](https://doi.org/10.3389/fmars.2020.00564).
- Giering, S. L. C. et al. (2014). "Reconciliation of the Carbon Budget in the Ocean's Twilight Zone." In: *Nature* 507.7493, pp. 480–483. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature13123](https://doi.org/10.1038/nature13123).
- Gillard, B., R. P. Harbour, N. Nowald, L. Thomsen, and M. H. Iversen (2022). "Vertical Distribution of Particulate Matter in the Clarion Clipperton Zone (German Sector)—Potential Impacts From Deep-Sea Mining Discharge in the Water Column." In: *Front. Mar. Sci.* 9, p. 820947. ISSN: 2296-7745. DOI: [10.3389/fmars.2022.820947](https://doi.org/10.3389/fmars.2022.820947).
- Gilmer, R. W. and G. R. Harbison (1986). "Morphology and Field Behavior of Pteropod Molluscs: Feeding Methods in the Families Cavoliniidae, Limacinidae and Peraclididae (Gastropoda: Thecosomata)." In: *Mar. Biol.* 91.1, pp. 47–57. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00397570](https://doi.org/10.1007/BF00397570).
- Gleiber, M., D. Steinberg, and H. Ducklow (2012). "Time Series of Vertical Flux of Zooplankton Fecal Pellets on the Continental Shelf of the Western Antarctic Peninsula." In: *Mar. Ecol. Prog. Ser.* 471, pp. 23–36. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps10021](https://doi.org/10.3354/meps10021).
- Goldthwait, S., J. Yen, J. Brown, and A. Alldredge (2004). "Quantification of Marine Snow Fragmentation by Swimming Euphausiids." In: *Limnol. Oceanogr.* 49.4, pp. 940–952. ISSN: 00243590. DOI: [10.4319/lo.2004.49.4.0940](https://doi.org/10.4319/lo.2004.49.4.0940).
- González Carman, V., A. Piola, T. D. O'Brien, D. D. Tormosov, and E. M. Acha (2019). "Circumpolar Frontal Systems as Potential Feeding Grounds of Southern Right Whales." In: *Progress in Oceanography* 176, p. 102123. ISSN: 00796611. DOI: [10.1016/j.pocean.2019.102123](https://doi.org/10.1016/j.pocean.2019.102123).
- Goodwin, M., K. T. Halvorsen, L. Jiao, K. M. Knausgård, A. H. Martin, M. Moyano, R. A. Oomen, J. H. Rasmussen, T. K. Sørvalen, and S. H. Thorbjørnsen (2022). "Unlocking the Potential of Deep Learning for Marine Ecology: Overview, Applications, and Outlook." In: *ICES Journal of Marine Science* 79.2. Ed. by D. Demer, pp. 319–336. ISSN: 1054-3139, 1095-9289. DOI: [10.1093/icesjms/fsab255](https://doi.org/10.1093/icesjms/fsab255).
- Gorgues, T., O. Aumont, and L. Memery (2019). "Simulated Changes in the Particulate Carbon Export Efficiency Due to Diel Vertical Migration of Zooplankton in the North Atlantic." In: *Geophysical Research Letters* 46.10, pp. 5387–5395. ISSN: 1944-8007. DOI: [10.1029/2018GL081748](https://doi.org/10.1029/2018GL081748).
- Gorsky, G., M. D. Ohman, M. Picheral, S. Gasparini, L. Stemann, J.-B. Romagnan, A. Cawood, S. Pesant, C. Garcia-Comas, and F. Prejger (2010). "Digital Zooplankton Image Analysis Using the ZooScan Integrated System." In: *Journal of Plankton Research* 32.3, pp. 285–303. ISSN: 0142-7873, 1464-3774. DOI: [10.1093/plankt/fbp124](https://doi.org/10.1093/plankt/fbp124).

- Gowing, M. M. (1989). "Abundance and Feeding Ecology of Antarctic Phaeodarian Radiolarians." In: *Marine Biology* 103.1, pp. 107–118. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00391069](https://doi.org/10.1007/BF00391069).
- Griffith, A. W. and C. J. Gobler (2020). "Harmful Algal Blooms: A Climate Change Co-Stressor in Marine and Freshwater Ecosystems." In: *Harmful Algae* 91, p. 101590. ISSN: 15689883. DOI: [10.1016/j.hal.2019.03.008](https://doi.org/10.1016/j.hal.2019.03.008).
- Grossmann, M. M., S. M. Gallager, and S. Mitarai (2015). "Continuous Monitoring of Near-Bottom Mesoplankton Communities in the East China Sea during a Series of Typhoons." In: *J Oceanogr* 71.1, pp. 115–124. ISSN: 0916-8370, 1573-868X. DOI: [10.1007/s10872-014-0268-y](https://doi.org/10.1007/s10872-014-0268-y).
- Guidi, L., Jackson, L. Stemman, J.-C. Miquel, M. Picheral, and G. Gorsky (2008). "Relationship between Particle Size Distribution and Flux in the Mesopelagic Zone." In: *Deep Sea Research Part I: Oceanographic Research Papers* 55.10, pp. 1364–1374. ISSN: 09670637. DOI: [10.1016/j.dsr.2008.05.014](https://doi.org/10.1016/j.dsr.2008.05.014).
- Guidi, L., L. Legendre, G. Reygondeau, J. Uitz, L. Stemmann, and S. A. Henson (2015). "A New Look at Ocean Carbon Remineralization for Estimating Deepwater Sequestration." In: *Global Biogeochem. Cycles* 29.7, pp. 1044–1059. ISSN: 08866236. DOI: [10.1002/2014GB005063](https://doi.org/10.1002/2014GB005063).
- Guidi, L., L. Stemmann, G. A. Jackson, F. Ibanez, H. Claustre, L. Legendre, M. Picheral, and G. Gorsky (2009). "Effects of Phytoplankton Community on Production, Size, and Export of Large Aggregates: A World-Ocean Analysis." In: *Limnol. Oceanogr.* 54.6, pp. 1951–1963. ISSN: 00243590. DOI: [10.4319/lo.2009.54.6.1951](https://doi.org/10.4319/lo.2009.54.6.1951).
- Guidi, L. et al. (2020). *Big Data in Marine Science*. Tech. rep. Zenodo. DOI: [10.5281/ZENODO.3755793](https://doi.org/10.5281/ZENODO.3755793).
- Guihen, D., J. A. Brearley, and S. Fielding (2022). "Antarctic Krill Likely Avoid Underwater Gliders." In: *Deep Sea Research Part I: Oceanographic Research Papers* 179, p. 103680. ISSN: 09670637. DOI: [10.1016/j.dsr.2021.103680](https://doi.org/10.1016/j.dsr.2021.103680).
- Guisan, A. and N. E. Zimmermann (2000). "Predictive Habitat Distribution Models in Ecology." In: *Ecological Modelling* 135.2-3, pp. 147–186. ISSN: 03043800. DOI: [10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9).
- Harbison, G. R., V. L. McAlister, and R. W. Gilmer (1986). "The Response of the Salp, *Pegea Confoederata*, to High Levels of Particulate Material: Starvation in the Midst of Plenty: Pegea and Particles." In: *Limnol. Oceanogr.* 31.2, pp. 371–382. ISSN: 00243590. DOI: [10.4319/lo.1986.31.2.0371](https://doi.org/10.4319/lo.1986.31.2.0371).
- Hastie, T. and R. Tibshirani (1986). "Generalized Additive Models." In: 1.3, p. 14.
- Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Second Edition. Springer.

- Hatton, I. A., R. F. Heneghan, Y. M. Bar-On, and E. D. Galbraith (2021). *The Global Ocean Size-Spectrum from Bacteria to Whales*. Preprint. Ecology. DOI: [10.1101/2021.04.03.438320](https://doi.org/10.1101/2021.04.03.438320).
- Hays, G., A. Richardson, and C. Robinson (2005). "Climate Change and Marine Plankton." In: *Trends in Ecology & Evolution* 20.6, pp. 337–344. ISSN: 01695347. DOI: [10.1016/j.tree.2005.03.004](https://doi.org/10.1016/j.tree.2005.03.004).
- Henson, S. A., R. Sanders, and E. Madsen (2012). "Global Patterns in Efficiency of Particulate Organic Carbon Export and Transfer to the Deep Ocean: EXPORT AND TRANSFER EFFICIENCY." In: *Global Biogeochem. Cycles* 26.1, n/a–n/a. ISSN: 08866236. DOI: [10.1029/2011GB004099](https://doi.org/10.1029/2011GB004099).
- Henson, S. A., R. Sanders, E. Madsen, P. J. Morris, F. Le Moigne, and G. D. Quartly (2011). "A Reduced Estimate of the Strength of the Ocean's Biological Carbon Pump: BIOLOGICAL CARBON PUMP STRENGTH." In: *Geophys. Res. Lett.* 38.4, n/a–n/a. ISSN: 00948276. DOI: [10.1029/2011GL046735](https://doi.org/10.1029/2011GL046735).
- Herman, A. W. (1992). "Design and Calibration of a New Optical Plankton Counter Capable of Sizing Small Zooplankton." In: *Deep Sea Research Part A. Oceanographic Research Papers* 39.3-4, pp. 395–415. ISSN: 01980149. DOI: [10.1016/0198-0149\(92\)90080-D](https://doi.org/10.1016/0198-0149(92)90080-D).
- Hijmans, R. J. (2021). *Raster: Geographic Data Analysis and Modeling*.
- Honjo, S., S. J. Manganini, R. A. Krishfield, and R. Francois (2008). "Particulate Organic Carbon Fluxes to the Ocean Interior and Factors Controlling the Biological Pump: A Synthesis of Global Sediment Trap Programs since 1983." In: *Progress in Oceanography* 76.3, pp. 217–285. ISSN: 00796611. DOI: [10.1016/j.pocean.2007.11.003](https://doi.org/10.1016/j.pocean.2007.11.003).
- Hoving, H.-J., S. Christiansen, E. Fabrizius, H. Hauss, R. Kiko, P. Linke, P. Neitzel, U. Piatkowski, and A. Körtzinger (2019). "The Pelagic In Situ Observation System (PELAGIOS) to Reveal Biodiversity, Behavior, and Ecology of Elusive Oceanic Fauna." In: *Ocean Sci.* 15.5, pp. 1327–1340. ISSN: 1812-0792. DOI: [10.5194/os-15-1327-2019](https://doi.org/10.5194/os-15-1327-2019).
- Hu, C., J. Xu, X. Li, Z. Shi, and R. Li (2021). "Environmental Regulations on Bacterial Abundance in the South China Sea Inferred from Regression Models." In: *Science of The Total Environment* 774, p. 146315. ISSN: 00489697. DOI: [10.1016/j.scitotenv.2021.146315](https://doi.org/10.1016/j.scitotenv.2021.146315).
- Huntley, M. and Boyd (1984). "Food-Limited Growth of Marine Zooplankton." In: *The American Naturalist* 124.4, pp. 455–478. ISSN: 0003-0147, 1537-5323. DOI: [10.1086/284288](https://doi.org/10.1086/284288).
- Hwang, J.-S., R. Kumar, H.-U. Dahms, L.-C. Tseng, and Q.-C. Chen (2007). "Mesh Size Affects Abundance Estimates of Oithona Spp. (Copepoda, Cyclopoida)." In: *Crustac* 80.7, pp. 827–837. ISSN: 0011-216X, 1568-5403. DOI: [10.1163/156854007781363169](https://doi.org/10.1163/156854007781363169).
- Ibarbalz, F. M. et al. (2019). "Global Trends in Marine Plankton Diversity across Kingdoms of Life." In: *Cell* 179.5, pp. 1084–1097. ISSN: 00928674. DOI: [10.1016/j.cell.2019.10.008](https://doi.org/10.1016/j.cell.2019.10.008).

- Ikeda, T. (1985). "Metabolic Rates of Epipelagic Marine Zooplankton as a Function of Body Mass and Temperature." In: *Marine Biology* 85, pages 1–11.
- Ikenoue, T., K. Kimoto, Y. Okazaki, M. Sato, M. C. Honda, K. Takahashi, N. Harada, and T. Fujiki (2019). "Phaeodaria: An Important Carrier of Particulate Organic Carbon in the Mesopelagic Twilight Zone of the North Pacific Ocean." In: *Global Biogeochem. Cycles* 33.8, pp. 1146–1160. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2019GB006258](https://doi.org/10.1029/2019GB006258).
- Ionescu, D., M. Bizic-Ionescu, A. Khalili, R. Malekmohammadi, M. R. Morad, D. De Beer, and H.-P. Grossart (2015). "A New Tool for Long-Term Studies of POM-bacteria Interactions: Overcoming the Century-Old Bottle Effect." In: *Sci Rep* 5.1, p. 14706. ISSN: 2045-2322. DOI: [10.1038/srep14706](https://doi.org/10.1038/srep14706).
- Irisson, J.-O., S.-D. Ayata, D. J. Lindsay, L. Karp-Boss, and L. Stemmann (2022). "Machine Learning for the Study of Plankton and Marine Snow from Images." In: *Annu. Rev. Mar. Sci.* 14.1, annurev-marine-041921-013023. ISSN: 1941-1405, 1941-0611. DOI: [10.1146/annurev-marine-041921-013023](https://doi.org/10.1146/annurev-marine-041921-013023).
- Iversen (2023). "Carbon Export in the Ocean: A Biologist's Perspective." In: *Annu. Rev. Mar. Sci.* 15.1, annurev-marine-032122-035153. ISSN: 1941-1405, 1941-0611. DOI: [10.1146/annurev-marine-032122-035153](https://doi.org/10.1146/annurev-marine-032122-035153).
- Iversen and Lampitt (2020). "Size Does Not Matter after All: No Evidence for a Size-Sinking Relationship for Marine Snow." In: *Progress in Oceanography* 189, p. 102445. ISSN: 00796611. DOI: [10.1016/j.pocean.2020.102445](https://doi.org/10.1016/j.pocean.2020.102445).
- Iversen, N. Nowald, H. Ploug, G. A. Jackson, and G. Fischer (2010). "High Resolution Profiles of Vertical Particulate Organic Matter Export off Cape Blanc, Mauritania: Degradation Processes and Ballasting Effects." In: *Deep Sea Research Part I: Oceanographic Research Papers* 57.6, pp. 771–784. ISSN: 09670637. DOI: [10.1016/j.dsr.2010.03.007](https://doi.org/10.1016/j.dsr.2010.03.007).
- Iversen and Ploug (2010). "Ballast Minerals and the Sinking Carbon Flux in the Ocean: Carbon-Specific Respiration Rates and Sinking Velocity of Marine Snow Aggregates." In: *Biogeosciences* 7.9, pp. 2613–2624. ISSN: 1726-4189. DOI: [10.5194/bg-7-2613-2010](https://doi.org/10.5194/bg-7-2613-2010).
- (2013). "Temperature Effects on Carbon-Specific Respiration Rate and Sinking Velocity of Diatom Aggregates – Potential Implications for Deep Ocean Export Processes." In: *Biogeosciences* 10.6, pp. 4073–4085. ISSN: 1726-4189. DOI: [10.5194/bg-10-4073-2013](https://doi.org/10.5194/bg-10-4073-2013).
- Jackson and D. M. Checkley (2011). "Particle Size Distributions in the Upper 100m Water Column and Their Implications for Animal Feeding in the Plankton." In: *Deep Sea Research Part I: Oceanographic Research Papers* 58.3, pp. 283–297. ISSN: 09670637. DOI: [10.1016/j.dsr.2010.12.008](https://doi.org/10.1016/j.dsr.2010.12.008).



- Jackson and T. Kiørboe (2004). "Zooplankton Use of Chemodetection to Find and Eat Particles." In: *Mar. Ecol. Prog. Ser.* 269, pp. 153–162. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps269153](https://doi.org/10.3354/meps269153).
- Johnson, X. Li, and B. E. Logan (1996). "Settling Velocities of Fractal Aggregates." In: *Environ. Sci. Technol.* 30.6, pp. 1911–1918. ISSN: 0013-936X, 1520-5851. DOI: [10.1021/es950604g](https://doi.org/10.1021/es950604g).
- Johnson et al. (2023). *Assessment of Oceanographic Conditions during the North Atlantic EXport Processes in the Ocean from RemoTe Sensing (EXPORTS) Field Campaign*. Preprint. Biogeochemistry. DOI: [10.31223/X5C097](https://doi.org/10.31223/X5C097).
- Jónasdóttir, S. H., A. W. Visser, K. Richardson, and M. R. Heath (2015). "Seasonal Copepod Lipid Pump Promotes Carbon Sequestration in the Deep North Atlantic." In: *Proc. Natl. Acad. Sci. U.S.A.* 112.39, pp. 12122–12126. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1512110112](https://doi.org/10.1073/pnas.1512110112).
- Jouandet, M.-P., T. W. Trull, L. Guidi, M. Picheral, F. Ebersbach, L. Stemmann, and S. Blain (2011). "Optical Imaging of Mesopelagic Particles Indicates Deep Carbon Flux beneath a Natural Iron-Fertilized Bloom in the Southern Ocean." In: *Limnol. Oceanogr.* 56.3, pp. 1130–1140. ISSN: 00243590. DOI: [10.4319/lo.2011.56.3.1130](https://doi.org/10.4319/lo.2011.56.3.1130).
- Karsenti, E. et al. (2011). "A Holistic Approach to Marine Eco-Systems Biology." In: *PLoS Biol* 9.10, e1001177 1–5. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.1001177](https://doi.org/10.1371/journal.pbio.1001177).
- Kiko, R. et al. (2017). "Biological and Physical Influences on Marine Snowfall at the Equator." In: *Nature Geosci* 10.11, pp. 852–858. ISSN: 1752-0894, 1752-0908. DOI: [10.1038/ngeo3042](https://doi.org/10.1038/ngeo3042).
- Kiko, R., P. Brandt, S. Christiansen, J. Faustmann, I. Kriest, E. Rodrigues, F. Schütte, and H. Hauss (2020). "Zooplankton-Mediated Fluxes in the Eastern Tropical North Atlantic." In: *Front. Mar. Sci.* 7, p. 358. ISSN: 2296-7745. DOI: [10.3389/fmars.2020.00358](https://doi.org/10.3389/fmars.2020.00358).
- Kiko, R. et al. (2022). "A Global Marine Particle Size Distribution Dataset Obtained with the Underwater Vision Profiler 5." In: *Earth Syst. Sci. Data* 14.9, pp. 4315–4337. ISSN: 1866-3516. DOI: [10.5194/essd-14-4315-2022](https://doi.org/10.5194/essd-14-4315-2022).
- Kiørboe, T. (2000). "Colonization of Marine Snow Aggregates by Invertebrate Zooplankton: Abundance, Scaling, and Possible Role." In: *Limnol. Oceanogr.* 45.2, pp. 479–484. ISSN: 00243590. DOI: [10.4319/lo.2000.45.2.0479](https://doi.org/10.4319/lo.2000.45.2.0479).
- (2001). "Formation and Fate of Marine Snow: Small-Scale Processes with Large-Scale Implications." In: *Sci. Mar.* 65.S2, pp. 57–71. ISSN: 1886-8134, 0214-8358. DOI: [10.3989/scimar.2001.65s257](https://doi.org/10.3989/scimar.2001.65s257).
- (2011). "How Zooplankton Feed: Mechanisms, Traits and Trade-Offs." In: *Biological Reviews* 86.2, pp. 311–339. ISSN: 14647931. DOI: [10.1111/j.1469-185X.2010.00148.x](https://doi.org/10.1111/j.1469-185X.2010.00148.x).

- Kjørboe, T. (2013). "Zooplankton Body Composition." In: *Limnol. Oceanogr.* 58.5, pp. 1843–1850. ISSN: 00243590. DOI: [10.4319/lo.2013.58.5.1843](https://doi.org/10.4319/lo.2013.58.5.1843).
- Kjørboe, T., H.-P. Grossart, H. Ploug, and K. Tang (2002). "Mechanisms and Rates of Bacterial Colonization of Sinking Aggregates." In: *Appl Environ Microbiol* 68.8, pp. 3996–4006. ISSN: 0099-2240, 1098-5336. DOI: [10.1128/AEM.68.8.3996-4006.2002](https://doi.org/10.1128/AEM.68.8.3996-4006.2002).
- Kjørboe, T. and Jackson (2001). "Marine Snow, Organic Solute Plumes, and Optimal Chemosensory Behavior of Bacteria." In: *Limnol. Oceanogr.* 46.6, pp. 1309–1318. ISSN: 00243590. DOI: [10.4319/lo.2001.46.6.1309](https://doi.org/10.4319/lo.2001.46.6.1309).
- Koster, M., R. Sietmann, A. Meuche, and G.-A. Paffenhofer (2011). "The Ultrastructure of a Doliolid and a Copepod Fecal Pellet." In: *Journal of Plankton Research* 33.10, pp. 1538–1549. ISSN: 0142-7873, 1464-3774. DOI: [10.1093/plankt/fbr053](https://doi.org/10.1093/plankt/fbr053).
- Kriest, I. and A. Oschlies (2008). "On the Treatment of Particulate Organic Matter Sinking in Large-Scale Models of Marine Biogeochemical Cycles." In:
- Kriest, I. (2002). "Different Parameterizations of Marine Snow in a 1D-model and Their Influence on Representation of Marine Snow, Nitrogen Budget and Sedimentation." In: *Deep Sea Research Part I: Oceanographic Research Papers* 49.12, pp. 2133–2162. ISSN: 0967-0637. DOI: [10.1016/S0967-0637\(02\)00127-9](https://doi.org/10.1016/S0967-0637(02)00127-9).
- Laget, M., L. Drago, T. Panaïotis, R. Kiko, L. Stemmann, A. Rogge, N. Llopis Monferrer, A. Leynaert, J. O. Irisson, and T. Biard (in review). "A Global Census of the Significance of Giant Mesopelagic Protists to the Biogeochemical Cycles of Carbon and Silicon." In: *Nature Communications*.
- Lampert, W. (1989). "The Adaptive Significance of Diel Vertical Migration of Zooplankton." In: *Functional Ecology* 3.1, p. 21. ISSN: 02698463. DOI: [10.2307/2389671](https://doi.org/10.2307/2389671). JSTOR: 2389671.
- Lampitt (1985). "Evidence for the Seasonal Deposition of Detritus to the Deep-Sea Floor and Its Subsequent Resuspension." In: *Deep Sea Research Part A. Oceanographic Research Papers* 32.8, pp. 885–897. ISSN: 01980149. DOI: [10.1016/0198-0149\(85\)90034-2](https://doi.org/10.1016/0198-0149(85)90034-2).
- Lampitt, T. Noji, and B. Von Bodungen (1990). "What Happens to Zooplankton Faecal Pellets? Implications for Material Flux." In: *Mar. Biol.* 104.1, pp. 15–23. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF01313152](https://doi.org/10.1007/BF01313152).
- Lampitt, R., D. Billett, and A. Martin (2010). "The Sustained Observatory over the Porcupine Abyssal Plain (PAP): Insights from Time Series Observations and Process Studies." In: *Deep Sea Research Part II: Topical Studies in Oceanography* 57.15, pp. 1267–1271. ISSN: 09670645. DOI: [10.1016/j.dsr2.2010.01.003](https://doi.org/10.1016/j.dsr2.2010.01.003).

- Lampitt, I. Salter, and D. Johns (2009). "Radiolaria: Major Exporters of Organic Carbon to the Deep Ocean." In: *Global Biogeochemical Cycles* 23.1. ISSN: 1944-9224. DOI: [10.1029/2008GB003221](https://doi.org/10.1029/2008GB003221).
- Lampitt, K. F. Wishner, C. M. Turley, and M. V. Angel (1993). "Marine Snow Studies in the Northeast Atlantic Ocean: Distribution, Composition and Role as a Food Source for Migrating Plankton." In: *Marine Biology* 116.4, pp. 689–702. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00355486](https://doi.org/10.1007/BF00355486).
- Laurenceau-Cornec, F. A. C. Le Moigne, M. Gallinari, B. Moriceau, J. Toullec, M. H. Iversen, A. Engel, and C. L. De La Rocha (2020). "New Guidelines for the Application of Stokes' Models to the Sinking Velocity of Marine Aggregates." In: *Limnol Oceanogr* 65.6, pp. 1264–1285. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.11388](https://doi.org/10.1002/lno.11388).
- Laurenceau-Cornec, T. Trull, D. Davies, C. De La Rocha, and S. Blain (2015). "Phytoplankton Morphology Controls on Marine Snow Sinking Velocity." In: *Mar. Ecol. Prog. Ser.* 520, pp. 35–56. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps11116](https://doi.org/10.3354/meps11116).
- Le Quéré, C. et al. (2016). "Role of Zooplankton Dynamics for Southern Ocean Phytoplankton Biomass Andglobal Biogeochemical Cycles." In: *Biogeosciences* 13.14, pp. 4111–4133. ISSN: 1726-4189. DOI: [10.5194/bg-13-4111-2016](https://doi.org/10.5194/bg-13-4111-2016).
- Leathwick, J. Elith, M. Francis, T. Hastie, and P. Taylor (2006). "Variation in Demersal Fish Species Richness in the Oceans Surrounding New Zealand: An Analysis Using Boosted Regression Trees." In: *Mar. Ecol. Prog. Ser.* 321, pp. 267–281. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps321267](https://doi.org/10.3354/meps321267).
- LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep Learning." In: *Nature* 521.7553, pp. 436–444. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- Legendre and Legendre (2012). *Numerical Ecology*. Third English edition. Developments in Environmental Modelling 24. Amsterdam: Elsevier. ISBN: 978-0-444-53868-0.
- Lehodey, P. et al. (2006). "Climate Variability, Fish, and Fisheries." In: *J. Climate* 19.20, pp. 5009–5030. ISSN: 0894-8755, 1520-0442. DOI: [10.1175/JCLI3898.1](https://doi.org/10.1175/JCLI3898.1).
- Lombard, F. and T. Kiørboe (2010). "Marine Snow Originating from Appendicularian Houses: Age-dependent Settling Characteristics." In: *Deep Sea Research Part I: Oceanographic Research Papers* 57.10, pp. 1304–1313. ISSN: 09670637. DOI: [10.1016/j.dsr.2010.06.008](https://doi.org/10.1016/j.dsr.2010.06.008).
- Lombard, F., F. Renaud, C. Sainsbury, A. Sciandra, and G. Gorsky (2009). "Appendicularian Ecophysiology I." In: *Journal of Marine Systems* 78.4, pp. 606–616. ISSN: 09247963. DOI: [10.1016/j.jmarsys.2009.01.004](https://doi.org/10.1016/j.jmarsys.2009.01.004).

- Lombard, F. et al. (2019). "Globally Consistent Quantitative Observations of Planktonic Ecosystems." In: *Front. Mar. Sci.* 6. ISSN: 2296-7745. DOI: [10.3389/fmars.2019.00196](https://doi.org/10.3389/fmars.2019.00196).
- Longhurst, A. and W. Glen Harrison (1989). "The Biological Pump: Profiles of Plankton Production and Consumption in the Upper Ocean." In: *Progress in Oceanography* 22.1, pp. 47–123. ISSN: 00796611. DOI: [10.1016/0079-6611\(89\)90010-4](https://doi.org/10.1016/0079-6611(89)90010-4).
- Longhurst, A. (1995). "Seasonal Cycles of Pelagic Production and Consumption." In: *Progress in Oceanography* 36.2, pp. 77–167. ISSN: 00796611. DOI: [10.1016/0079-6611\(95\)00015-1](https://doi.org/10.1016/0079-6611(95)00015-1).
- Lucas, C. H., D. O. B. Jones, C. J. Hollyhead, R. H. Condon, C. M. Duarte, W. M. Graham, K. L. Robinson, K. A. Pitt, M. Schildhauer, and J. Regetz (2014). "Gelatinous Zooplankton Biomass in the Global Oceans: Geographic Variation and Environmental Drivers: Global Gelatinous Biomass." In: *Global Ecology and Biogeography* 23.7, pp. 701–714. ISSN: 1466822X. DOI: [10.1111/geb.12169](https://doi.org/10.1111/geb.12169).
- Luo, J. Y., J.-O. Irisson, B. Graham, C. Guigand, A. Sarafranz, C. Mader, and R. K. Cowen (2018). "Automated Plankton Image Analysis Using Convolutional Neural Networks." In: *Limnology & Ocean Methods* 16.12, pp. 814–827. ISSN: 1541-5856, 1541-5856. DOI: [10.1002/lom3.10285](https://doi.org/10.1002/lom3.10285).
- Mansour, J. S., A. Norlin, N. Llopis Monferrer, S. L'Helguen, and F. Not (2021). "Carbon and Nitrogen Content to Biovolume Relationships for Marine Protist of the Rhizaria Lineage (Radiolaria and Phaeodaria)." In: *Limnol Oceanogr* 66.5, pp. 1703–1717. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.11714](https://doi.org/10.1002/lno.11714).
- Marsay, C. M., R. J. Sanders, S. A. Henson, K. Pabortsava, E. P. Achterberg, and R. S. Lampitt (2015). "Attenuation of Sinking Particulate Organic Carbon Flux through the Mesopelagic Ocean." In: *Proc. Natl. Acad. Sci. U.S.A.* 112.4, pp. 1089–1094. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1415311112](https://doi.org/10.1073/pnas.1415311112).
- Martin, G. A. Knauer, D. M. Karl, and W. W. Broenkow (1987). "VERTEX: Carbon Cycling in the Northeast Pacific." In: *Deep Sea Research Part A. Oceanographic Research Papers* 34.2, pp. 267–285. ISSN: 01980149. DOI: [10.1016/0198-0149\(87\)90086-0](https://doi.org/10.1016/0198-0149(87)90086-0).
- Martin, R. S. Lampitt, M. Jane Perry, R. Sanders, C. Lee, and E. D'Asaro (2011). "Export and Mesopelagic Particle Flux during a North Atlantic Spring Diatom Bloom." In: *Deep Sea Research Part I: Oceanographic Research Papers* 58.4, pp. 338–349. ISSN: 09670637. DOI: [10.1016/j.dsr.2011.01.006](https://doi.org/10.1016/j.dsr.2011.01.006).
- McCave, I. (1984). "Size Spectra and Aggregation of Suspended Particles in the Deep Ocean." In: *Deep Sea Research Part A. Oceanographic Research Papers* 31.4, pp. 329–352. ISSN: 01980149. DOI: [10.1016/0198-0149\(84\)90088-8](https://doi.org/10.1016/0198-0149(84)90088-8).

- McConville, K., A. Atkinson, E. S. Fileman, J. I. Spicer, and A. G. Hirst (2016). "Disentangling the Counteracting Effects of Water Content and Carbon Mass on Zooplankton Growth." In: *J. Plankton Res.*, pp. 246–256. ISSN: 0142-7873, 1464-3774. DOI: [10.1093/plankt/fbw094](https://doi.org/10.1093/plankt/fbw094).
- McDonnell, A. M. et al. (2015). "The Oceanographic Toolbox for the Collection of Sinking and Suspended Marine Particles." In: *Progress in Oceanography* 133, pp. 17–31. ISSN: 00796611. DOI: [10.1016/j.pocean.2015.01.007](https://doi.org/10.1016/j.pocean.2015.01.007).
- McGinty, N., A. D. Barton, N. R. Record, Z. V. Finkel, D. G. Johns, C. A. Stock, and A. J. Irwin (2021). "Anthropogenic Climate Change Impacts on Copepod Trait Biogeography." In: *Glob. Change Biol.* 27.7, pp. 1431–1442. ISSN: 1354-1013, 1365-2486. DOI: [10.1111/gcb.15499](https://doi.org/10.1111/gcb.15499).
- Michaels, A. F. (1988). "Vertical Distribution and Abundance of Acantharia and Their Symbionts." In: *Marine Biology* 97.4, pp. 559–569. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00391052](https://doi.org/10.1007/BF00391052).
- Michaels, A. F. (1991). "Acantharian Abundance and Symbiont Productivity at the VERTEX Seasonal Station." In: *J Plankton Res* 13.2, pp. 399–418. ISSN: 0142-7873, 1464-3774. DOI: [10.1093/plankt/13.2.399](https://doi.org/10.1093/plankt/13.2.399).
- Moriarty, R. and T. D. O'Brien (2013). "Distribution of Mesozooplankton Biomass in the Global Ocean." In: *Earth Syst. Sci. Data* 5, pp. 45–55. DOI: [doi:10.5194/essd-5-45-2013](https://doi.org/10.5194/essd-5-45-2013).
- Moriarty, R., E. T. Buitenhuis, C. Le Quéré, and M.-P. Gosselin (2012). "Distribution of Known Macrozooplankton Abundance and Biomass in the Global Ocean." In: *Earth Syst. Sci. Data Discuss.* 5.1, pp. 187–220. ISSN: 1866-3591. DOI: [10.5194/essdd-5-187-2012](https://doi.org/10.5194/essdd-5-187-2012).
- Morley, J. and J. Stepien (1984). "Siliceous Microfauna in Waters beneath Antarctic Sea Ice." In: *Mar. Ecol. Prog. Ser.* 19, pp. 207–210. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps019207](https://doi.org/10.3354/meps019207).
- Mouw, C. B., A. Barnett, G. A. McKinley, L. Gloege, and D. Pilcher (2016). "Global Ocean Particulate Organic Carbon Flux Merged with Satellite Parameters." In: *Earth System Science Data* 8.2, pp. 531–541. ISSN: 1866-3508. DOI: [10.5194/essd-8-531-2016](https://doi.org/10.5194/essd-8-531-2016).
- Muñoz, M., A. Reul, M. Vargas-Yáñez, F. Plaza, B. Bautista, M. García-Martínez, F. Moya, M.-L. Gómez-Moreno, J. Fernandes, and V. Rodríguez (2017). "Fertilization and Connectivity in the Garrucha Canyon (SE-Spain) Implications for Marine Spatial Planning." In: *Marine Environmental Research* 126, pp. 45–68. ISSN: 01411136. DOI: [10.1016/j.marenvres.2017.02.007](https://doi.org/10.1016/j.marenvres.2017.02.007).
- Nakamura, Y., I. Imai, A. Yamaguchi, A. Tuji, and N. Suzuki (2013). "Aulographis Japonica Sp. Nov. (Phaeodaria, Aulacanthida, Aulacanthidae), an Abundant Zooplankton in the Deep Sea of the Sea of Japan." In: *Plankton Benthos Res* 8.3, pp. 107–115. ISSN: 1880-8247, 1882-627X. DOI: [10.3800/pbr.8.107](https://doi.org/10.3800/pbr.8.107).

- Nakamura, Y. and N. Suzuki (2015). "Phaeodaria: Diverse Marine Cercozoans of World-Wide Distribution." In: *Marine Protists*. Ed. by S. Ohtsuka, T. Suzaki, T. Horiguchi, N. Suzuki, and F. Not. Tokyo: Springer Japan, pp. 223–249. ISBN: 978-4-431-55129-4 978-4-431-55130-0. DOI: [10.1007/978-4-431-55130-0\\_9](https://doi.org/10.1007/978-4-431-55130-0_9).
- Nelder, J. A. and R. W. M. Wedderburn (1972). "Generalized Linear Models." In: p. 16.
- Nichols, J. and A. Thompson (1991). "Mesh Selection of Copepodite and Nauplius Stages of Four Calanoid Copepod Species." In: *J Plankton Res* 13.3, pp. 661–671. ISSN: 0142-7873, 1464-3774. DOI: [10.1093/plankt/13.3.661](https://doi.org/10.1093/plankt/13.3.661).
- Nowicki, M., T. DeVries, and D. A. Siegel (2022). "Quantifying the Carbon Export and Sequestration Pathways of the Ocean's Biological Carbon Pump." In: *Global Biogeochemical Cycles*, p. 22.
- O'Brien, T. D. (2005). *COPEPOD: A Global Plankton Database*.
- Ohman, M. D. (2019). "A Sea of Tentacles: Optically Discernible Traits Resolved from Planktonic Organisms in Situ." In: *ICES Journal of Marine Science* 76.7. Ed. by H. Browman, pp. 1959–1972. ISSN: 1054-3139, 1095-9289. DOI: [10.1093/icesjms/fsz184](https://doi.org/10.1093/icesjms/fsz184).
- Olson, R. J. and H. M. Sosik (2007). "A Submersible Imaging-in-Flow Instrument to Analyze Nano-and Microplankton: Imaging FlowCytobot: In Situ Imaging of Nano- and Microplankton." In: *Limnol. Oceanogr. Methods* 5.6, pp. 195–203. ISSN: 15415856. DOI: [10.4319/lom.2007.5.195](https://doi.org/10.4319/lom.2007.5.195).
- Omand, M. M., E. A. D'Asaro, C. M. Lee, M. J. Perry, N. Briggs, I. Cetinić, and A. Mahadevan (2015). "Eddy-Driven Subduction Exports Particulate Organic Carbon from the Spring Bloom." In: *Science* 348.6231, pp. 222–225. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1260062](https://doi.org/10.1126/science.1260062).
- Orenstein, E. C., K. M. Kenitz, P. L. Roberts, P. J. Franks, J. S. Jaffe, and A. D. Barton (2020). "Semi- and Fully Supervised Quantification Techniques to Improve Population Estimates from Machine Classifiers." In: *Limnology & Ocean Methods* 18.12, pp. 739–753. ISSN: 1541-5856, 1541-5856. DOI: [10.1002/lom3.10399](https://doi.org/10.1002/lom3.10399).
- Orenstein, E. C. et al. (2022). "Machine Learning Techniques to Characterize Functional Traits of Plankton from Image Data." In: *Limnology & Oceanography* 67.8, pp. 1647–1669. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.12101](https://doi.org/10.1002/lno.12101).
- Panaïotis, T. et al. (2023). "Three Major Mesoplanktonic Communities Resolved by in Situ Imaging in the Upper 500 m of the Global Ocean." In: *Global Ecol Biogeogr*, geb.13741. ISSN: 1466-822X, 1466-8238. DOI: [10.1111/geb.13741](https://doi.org/10.1111/geb.13741).
- Pesant, S. et al. (2015). "Open Science Resources for the Discovery and Analysis of Tara Oceans Data." In: *Sci Data* 2.1, p. 150023. ISSN: 2052-4463. DOI: [10.1038/sdata.2015.23](https://doi.org/10.1038/sdata.2015.23).

- Pettitt-Wade, H., T. Pearce, D. Kuptana, C. P. Gallagher, K. Scharfenberg, E. V. Lea, N. E. Hussey, and L. L. Loseto (2020). "Inuit Observations of a Tunicata Bloom Unusual for the Amundsen Gulf, Western Canadian Arctic." In: *Arctic Science* 6.3, pp. 340–351. ISSN: 2368-7460. DOI: [10.1139/as-2020-0018](https://doi.org/10.1139/as-2020-0018).
- Phillips, P. Kremer, and L. P. Madin (2009). "Defecation by *Salpa Thompsoni* and Its Contribution to Vertical Flux in the Southern Ocean." In: *Mar Biol* 156.3, pp. 455–467. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/s00227-008-1099-4](https://doi.org/10.1007/s00227-008-1099-4).
- Phillips, S. J., R. P. Anderson, and R. E. Schapire (2006). "Maximum Entropy Modeling of Species Geographic Distributions." In: *Ecological Modelling* 190.3-4, pp. 231–259. ISSN: 03043800. DOI: [10.1016/j.ecolmodel.2005.03.026](https://doi.org/10.1016/j.ecolmodel.2005.03.026).
- Picheral, M., S. Colin, and J.-O. Irisson (2017). *EcoTaxa, a Tool for the Taxonomic Classification of Images*.
- Picheral, M., L. Guidi, L. Stemann, D. M. Karl, G. Iddaoud, and G. Gorsky (2010). "The Underwater Vision Profiler 5: An Advanced Instrument for High Spatial Resolution Studies of Particle Size Spectra and Zooplankton: Underwater Vision Profiler." In: *Limnol. Oceanogr. Methods* 8.9, pp. 462–473. ISSN: 15415856. DOI: [10.4319/lom.2010.8.462](https://doi.org/10.4319/lom.2010.8.462).
- Picheral, M. et al. (2021). "The Underwater Vision Profiler 6: An Imaging Sensor of Particle Size Spectra and Plankton, for Autonomous and Cabled Platforms." In: *Limnology & Ocean Methods*, lom3.10475. ISSN: 1541-5856, 1541-5856. DOI: [10.1002/lom3.10475](https://doi.org/10.1002/lom3.10475).
- Pinkerton, M. H., M. Décima, J. A. Kitchener, K. T. Takahashi, K. V. Robinson, R. Stewart, and G. W. Hosie (2020). "Zooplankton in the Southern Ocean from the Continuous Plankton Recorder: Distributions and Long-Term Change." In: *Deep Sea Research Part I: Oceanographic Research Papers* 162, p. 103303. ISSN: 09670637. DOI: [10.1016/j.dsr.2020.103303](https://doi.org/10.1016/j.dsr.2020.103303).
- Ploug, H., H. Grossart, F. Azam, and B. Jørgensen (1999). "Photosynthesis, Respiration, and Carbon Turnover in Sinking Marine Snow from Surface Waters of Southern California Bight: Implications for the Carbon Cycle in the Ocean." In: *Mar. Ecol. Prog. Ser.* 179, pp. 1–11. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps179001](https://doi.org/10.3354/meps179001).
- Ploug, H. and H.-P. Grossart (2000). "Bacterial Growth and Grazing on Diatom Aggregates: Respiratory Carbon Turnover as a Function of Aggregate Size and Sinking Velocity." In: *Limnol. Oceanogr* 45.7, pp. 1467–1475. ISSN: 00243590. DOI: [10.4319/lo.2000.45.7.1467](https://doi.org/10.4319/lo.2000.45.7.1467).
- Ploug and B. Jørgensen (1999). "A Net-Jet Flow System for Mass Transfer and Microsensor Studies of Sinking Aggregates." In: *Mar. Ecol. Prog. Ser.* 176, pp. 279–290. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps176279](https://doi.org/10.3354/meps176279).

- Pollina, T., A. G. Larson, F. Lombard, H. Li, D. Le Guen, S. Colin, C. De Vargas, and M. Prakash (2022). "PlanktoScope: Affordable Modular Quantitative Imaging Platform for Citizen Oceanography." In: *Front. Mar. Sci.* 9, p. 949428. ISSN: 2296-7745. DOI: [10.3389/fmars.2022.949428](https://doi.org/10.3389/fmars.2022.949428).
- R (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Remsen, A., T. L. Hopkins, and S. Samson (2004). "What You See Is Not What You Catch: A Comparison of Concurrently Collected Net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder Data from the Northeast Gulf of Mexico." In: *Deep Sea Research Part I: Oceanographic Research Papers* 51.1, pp. 129–151. ISSN: 09670637. DOI: [10.1016/j.dsr.2003.09.008](https://doi.org/10.1016/j.dsr.2003.09.008).
- Reynolds, R. A., D. Stramski, V. M. Wright, and S. B. Woźniak (2010). "Measurements and Characterization of Particle Size Distributions in Coastal Waters." In: *J. Geophys. Res.* 115.C8, p. C08024. ISSN: 0148-0227. DOI: [10.1029/2009JC005930](https://doi.org/10.1029/2009JC005930).
- Reynolds, R. A., D. Stramski, and G. Neukermans (2016). "Optical Backscattering by Particles in Arctic Seawater and Relationships to Particle Mass Concentration, Size Distribution, and Bulk Composition." In: *Limnology & Oceanography* 61.5, pp. 1869–1890. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.10341](https://doi.org/10.1002/lno.10341).
- Richardson, A. J., A. Bakun, G. C. Hays, and M. J. Gibbons (2009). "The Jellyfish Joyride: Causes, Consequences and Management Responses to a More Gelatinous Future." In: *Trends in Ecology & Evolution* 24.6, pp. 312–322. ISSN: 01695347. DOI: [10.1016/j.tree.2009.01.010](https://doi.org/10.1016/j.tree.2009.01.010).
- Ricour, F. (2023). "Towards a New Insight of the Carbon Transport in the Global Ocean." PhD thesis. Sorbonne Université ; Université de Liège.
- Riemann, F. (1989). "Gelatinous Phytoplankton Detritus Aggregates on the Atlantic Deep-Sea Bed: Structure and Mode of Formation." In: *Mar. Biol.* 100.4, pp. 533–539. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00394830](https://doi.org/10.1007/BF00394830).
- Riley, J. S., R. Sanders, C. Marsay, F. A. C. Le Moigne, E. P. Achterberg, and A. J. Poulton (2012). "The Relative Contribution of Fast and Slow Sinking Particles to Ocean Carbon Export." In: *Global Biogeochem. Cycles* 26.1, n/a–n/a. ISSN: 08866236. DOI: [10.1029/2011GB004085](https://doi.org/10.1029/2011GB004085).
- Robison, B. H., K. R. Reisenbichler, and R. E. Sherlock (2005). "Giant Larvacean Houses: Rapid Carbon Transport to the Deep Sea Floor." In: *Science* 308.5728, pp. 1609–1611. DOI: [10.1126/science.1109104](https://doi.org/10.1126/science.1109104).
- Romagnan, J. B., L. Aldamman, S. Gasparini, P. Nival, A. Aubert, J. L. Jamet, and L. Stemmann (2016). "High Frequency Mesozooplankton Monitoring: Can Imaging Systems and Automated Sample Analysis Help Us Describe and Interpret Changes in Zooplankton Community Composition and Size Structure — An Example from a Coastal Site."



- In: *Journal of Marine Systems* 162, pp. 18–28. ISSN: 09247963. DOI: [10.1016/j.jmarsys.2016.03.013](https://doi.org/10.1016/j.jmarsys.2016.03.013).
- Rombouts, I., G. Beaugrand, F. Ibañez, S. Gasparini, S. Chiba, and L. Legendre (2009). “Global Latitudinal Variations in Marine Copepod Diversity and Environmental Factors.” In: *Proc. R. Soc. B* 276.1670, pp. 3053–3062. ISSN: 0962-8452, 1471-2954. DOI: [10.1098/rspb.2009.0742](https://doi.org/10.1098/rspb.2009.0742).
- Rubbens, P. et al. (2023). “Machine Learning in Marine Ecology: An Overview of Techniques and Applications.” In: *ICES Journal of Marine Science* 80.7. Ed. by C. Whidden, pp. 1829–1853. ISSN: 1054-3139, 1095-9289. DOI: [10.1093/icesjms/fsad100](https://doi.org/10.1093/icesjms/fsad100).
- Sameoto, D., N. Cochrane, and A. Herman (1993). “Convergence of Acoustic, Optical, and Net-Catch Estimates of Euphausiid Abundance: Use of Artificial Light to Reduce Net.” In: *Canadian Journal of Fisheries and Aquatic Science* 50, pp. 334–346.
- Schapire, R. E. (2003). “The Boosting Approach to Machine Learning: An Overview.” In: *Nonlinear Estimation and Classification*. Ed. by D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, and B. Yu. Lecture Notes in Statistics. New York, NY: Springer, pp. 149–171. ISBN: 978-0-387-21579-2. DOI: [10.1007/978-0-387-21579-2\\_9](https://doi.org/10.1007/978-0-387-21579-2_9).
- Schröder, S.-M., R. Kiko, and R. Koch (2020). “MorphoCluster: Efficient Annotation of Plankton Images by Clustering.” In: *Sensors* 20.11, p. 3060. ISSN: 1424-8220. DOI: [10.3390/s20113060](https://doi.org/10.3390/s20113060).
- Schulz, J., K. Barz, P. Ayon, A. Ludtke, O. Zielinski, D. Mengedoht, and H.-J. Hirche (2010). “Imaging of Plankton Specimens with the Lightframe On-Sight Keyspecies Investigation (LOKI) System.” In: *JEOS:RP* 5, 10017s. ISSN: 1990-2573. DOI: [10.2971/jeos.2010.10017s](https://doi.org/10.2971/jeos.2010.10017s).
- Séférian, R., S. Berthet, A. Yool, J. Palmieri, L. Bopp, A. Tagliabue, L. Kwiatkowski, O. Aumont, J. Christian, J. Dunne, et al. (2020). “Tracking Improvement in Simulated Marine Biogeochemistry between CMIP5 and CMIP6.” In: *Current Climate Change Reports* 6.3, pp. 95–119.
- Shanks, A. L. and E. W. Edmondson (1989). “Laboratory-Made Artificial Marine Snow: A Biological Model of the Real Thing.” In: *Mar. Biol.* 101.4, pp. 463–470. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00541648](https://doi.org/10.1007/BF00541648).
- Sheldon, R. W., A. Prakash, and W. H. Sutcliffe (1972). “The Size Distribution of Particles in the Ocean.” In: *Limnol. Oceanogr.* 17.3, pp. 327–340. ISSN: 00243590. DOI: [10.4319/lo.1972.17.3.0327](https://doi.org/10.4319/lo.1972.17.3.0327).
- Siegel (2005). “Distribution and Population Dynamics of Euphausia Superba: Summary of Recent Findings.” In: *Polar Biol* 29.1, pp. 1–22. ISSN: 0722-4060, 1432-2056. DOI: [10.1007/s00300-005-0058-5](https://doi.org/10.1007/s00300-005-0058-5).
- ed. (2016). *Biology and Ecology of Antarctic Krill*. Advances in Polar Ecology. Cham: Springer International Publishing. ISBN: 978-3-319-29277-9 978-3-319-29279-3. DOI: [10.1007/978-3-319-29279-3](https://doi.org/10.1007/978-3-319-29279-3).

- Siegel, K. O. Buesseler, S. C. Doney, S. F. Sailley, M. J. Behrenfeld, and P. W. Boyd (2014). "Global Assessment of Ocean Carbon Export by Combining Satellite Observations and Food-Web Models." In: *Global Biogeochem. Cycles* 28.3, pp. 181–196. ISSN: 08866236. DOI: [10.1002/2013GB004743](https://doi.org/10.1002/2013GB004743).
- Siegel et al. (2016). "Prediction of the Export and Fate of Global Ocean Net Primary Production: The EXPORTS Science Plan." In: *Front. Mar. Sci.* 3. ISSN: 2296-7745. DOI: [10.3389/fmars.2016.00022](https://doi.org/10.3389/fmars.2016.00022).
- Siegel et al. (2021). "An Operational Overview of the EXport Processes in the Ocean from RemoTe Sensing (EXPORTS) Northeast Pacific Field Deployment." In: *Elementa: Science of the Anthropocene* 9.1, p. 00107. ISSN: 2325-1026. DOI: [10.1525/elementa.2020.00107](https://doi.org/10.1525/elementa.2020.00107).
- Sieracki, C., M. Sieracki, and C. Yentsch (1998). "An Imaging-in-Flow System for Automated Analysis of Marine Microplankton." In: *Mar. Ecol. Prog. Ser.* 168, pp. 285–296. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps168285](https://doi.org/10.3354/meps168285).
- Silver (2015). "Marine Snow: A Brief Historical Sketch." In: *Limnology and Oceanography Bulletin* 24.1, pp. 5–10. ISSN: 1539607X. DOI: [10.1002/lob.10005](https://doi.org/10.1002/lob.10005).
- Smetacek, V. (1999). "Revolution in the Ocean." In: *Nature* 401.6754, pp. 647–647. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/44281](https://doi.org/10.1038/44281).
- Son, D., H. Cho, and E. J. Lee (2018). "Determining Factors for the Occurrence and Richness of Submerged Macrophytes in Major Korean Rivers." In: *Aquatic Botany* 150, pp. 82–88. ISSN: 03043770. DOI: [10.1016/j.aquabot.2018.07.003](https://doi.org/10.1016/j.aquabot.2018.07.003).
- Song, Y. and M. J. Rau (2022). "A Novel Method to Study the Fragmentation Behavior of Marine Snow Aggregates in Controlled Shear Flow." In: *Limnology & Ocean Methods* 20.10, pp. 618–632. ISSN: 1541-5856, 1541-5856. DOI: [10.1002/lom3.10509](https://doi.org/10.1002/lom3.10509).
- Soviadan, Y. D., F. Benedetti, M. C. Brandão, S.-D. Ayata, J.-O. Irisson, J. L. Jamet, R. Kiko, F. Lombard, K. Gnandi, and L. Stemmann (2022). "Patterns of Mesozooplankton Community Composition and Vertical Fluxes in the Global Ocean." In: *Progress in Oceanography* 200, p. 102717. ISSN: 00796611. DOI: [10.1016/j.pocean.2021.102717](https://doi.org/10.1016/j.pocean.2021.102717).
- Soviadan, Y. D. et al. (2023). *Complete Zooplankton Size Spectra Re-Constructed from « in Situ » Imaging and Multinet Data in the Global Ocean*. Preprint. Ecology. DOI: [10.1101/2023.06.29.547051](https://doi.org/10.1101/2023.06.29.547051).
- Stamieszkin, K., D. K. Steinberg, and A. E. Maas (2021). "Fecal Pellet Production by Mesozooplankton in the Subarctic Northeast Pacific Ocean." In: *Limnology & Oceanography* 66.7, pp. 2585–2597. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.11774](https://doi.org/10.1002/lno.11774).
- Steinberg, D. K., J. S. Cope, S. E. Wilson, and T. Kobari (2008). "A Comparison of Mesopelagic Mesozooplankton Community Structure in the Subtropical and Subarctic North Pacific Ocean." In: *Deep Sea*

- Research Part II: Topical Studies in Oceanography* 55.14-15, pp. 1615–1635. ISSN: 09670645. DOI: [10.1016/j.dsr2.2008.04.025](https://doi.org/10.1016/j.dsr2.2008.04.025).
- Steinberg and M. R. Landry (2017). “Zooplankton and the Ocean Carbon Cycle.” In: *Annu. Rev. Mar. Sci.* 9.1, pp. 413–444. ISSN: 1941-1405, 1941-0611. DOI: [10.1146/annurev-marine-010814-015924](https://doi.org/10.1146/annurev-marine-010814-015924).
- Steinberg, A. Maas, K. Sharpe, and K. Stamieszkin (2022). *Ocean Sciences Meeting 2022 - Large, Gelatinous Pteropods (Pelagic Snails) as Agents of Export in the Subarctic Northeast Atlantic Ocean*.
- Steinberg et al. (2023). “The Outsized Role of Salps in Carbon Export in the Subarctic Northeast Pacific Ocean.” In: *Global Biogeochemical Cycles* 37.1. ISSN: 0886-6236, 1944-9224. DOI: [10.1029/2022GB007523](https://doi.org/10.1029/2022GB007523).
- Stemmann and E. Boss (2012). “Plankton and Particle Size and Packaging: From Determining Optical Properties to Driving the Biological Pump.” In: *Annu. Rev. Mar. Sci.* 4.1, pp. 263–290. ISSN: 1941-1405, 1941-0611. DOI: [10.1146/annurev-marine-120710-100853](https://doi.org/10.1146/annurev-marine-120710-100853).
- Stemmann, G. Gorsky, J.-C. Marty, M. Picheral, and J.-C. Miquel (2002). “Four-Year Study of Large-Particle Vertical Distribution (0–1000 m) in the NW Mediterranean in Relation to Hydrology, Phytoplankton, and Vertical Flux.” In: *Deep Sea Research Part II: Topical Studies in Oceanography* 49.11, pp. 2143–2162.
- Stemmann, Jackson, and Gorsky (2004). “A Vertical Model of Particle Size Distributions and Fluxes in the Midwater Column That Includes Biological and Physical Processes—Part II: Application to a Three Year Survey in the NW Mediterranean Sea.” In: *Deep Sea Research Part I: Oceanographic Research Papers* 51.7, pp. 885–908. ISSN: 09670637. DOI: [10.1016/j.dsr.2004.03.002](https://doi.org/10.1016/j.dsr.2004.03.002).
- Stemmann, Jackson, and Ianson (2004). “A Vertical Model of Particle Size Distributions and Fluxes in the Midwater Column That Includes Biological and Physical Processes—Part I: Model Formulation.” In: *Deep Sea Research Part I: Oceanographic Research Papers* 51.7, pp. 865–884. ISSN: 09670637. DOI: [10.1016/j.dsr.2004.03.001](https://doi.org/10.1016/j.dsr.2004.03.001).
- Stemmann, L., M. Youngbluth, K. Robert, A. Hosia, M. Picheral, H. Paterson, F. Ibanez, L. Guidi, F. Lombard, and G. Gorsky (2008). “Global Zoogeography of Fragile Macrozooplankton in the Upper 100–1000 m Inferred from the Underwater Video Profiler.” In: *ICES J Mar Sci* 65.3, pp. 433–442. ISSN: 1054-3139. DOI: [10.1093/icesjms/fsn010](https://doi.org/10.1093/icesjms/fsn010).
- Stukel, M. R., T. Biard, J. Krause, and M. D. Ohman (2018). “Large Phaeodaria in the Twilight Zone: Their Role in the Carbon Cycle: Phaeodarian Ecology in the Twilight Zone.” In: *Limnol. Oceanogr.* 63.6, pp. 2579–2594. ISSN: 00243590. DOI: [10.1002/lno.10961](https://doi.org/10.1002/lno.10961).
- Stukel, M. R., M. Décima, and M. R. Landry (2022). “Quantifying Biological Carbon Pump Pathways with a Data-Constrained Mechanistic Model Ensemble Approach.” In: *Biogeosciences* 19.15, pp. 3595–3624. ISSN: 1726-4189. DOI: [10.5194/bg-19-3595-2022](https://doi.org/10.5194/bg-19-3595-2022).

- Stukel, M. R., M. D. Ohman, T. B. Kelly, and T. Biard (2019). "The Roles of Suspension-Feeding and Flux-Feeding Zooplankton as Gatekeepers of Particle Flux Into the Mesopelagic Ocean in the Northeast Pacific." In: *Frontiers in Marine Science* 6. ISSN: 2296-7745.
- Sun, H., P. Benzie, N. Burns, D. Hendry, M. Player, and J. Watson (2008). "Underwater Digital Holography for Studies of Marine Plankton." In: *Phil. Trans. R. Soc. A*. 366.1871, pp. 1789–1806. ISSN: 1364-503X, 1471-2962. DOI: [10.1098/rsta.2007.2187](https://doi.org/10.1098/rsta.2007.2187).
- Sunday, J. M., A. E. Bates, and N. K. Dulvy (2012). "Thermal Tolerance and the Global Redistribution of Animals." In: *Nature Clim Change* 2.9, pp. 686–690. ISSN: 1758-678X, 1758-6798. DOI: [10.1038/nclimate1539](https://doi.org/10.1038/nclimate1539).
- Suthers, I., D. Rissik, and A. Richardson (2019). *Plankton: A Guide to Their Ecology and Monitoring for Water Quality*. CSIRO Publishing. ISBN: 978-1-4863-0880-4.
- Suzuki, N. and F. Not (2015). "Biology and Ecology of Radiolaria." In: *Marine Protists*. Ed. by S. Ohtsuka, T. Suzaki, T. Horiguchi, N. Suzuki, and F. Not. Tokyo: Springer Japan, pp. 179–222. ISBN: 978-4-431-55129-4 978-4-431-55130-0. DOI: [10.1007/978-4-431-55130-0\\_8](https://doi.org/10.1007/978-4-431-55130-0_8).
- Svensen, C., C. Wexels Riser, M. Reigstad, and L. Seuthe (2012). "Degradation of Copepod Faecal Pellets in the Upper Layer: Role of Microbial Community and *Calanus finmarchicus*." In: *Mar. Ecol. Prog. Ser.* 462, pp. 39–49. ISSN: 0171-8630, 1616-1599. DOI: [10.3354/meps09808](https://doi.org/10.3354/meps09808).
- Toullec, J., D. Vincent, L. Frohn, P. Miner, M. Le Goff, J. Devesa, and B. Moriceau (2019). "Copepod Grazing Influences Diatom Aggregation and Particle Dynamics." In: *Front. Mar. Sci.* 6, p. 751. ISSN: 2296-7745. DOI: [10.3389/fmars.2019.00751](https://doi.org/10.3389/fmars.2019.00751).
- Trudnowska, E., L. Lacour, M. Ardyna, A. Rogge, J. O. Irisson, A. M. Waite, M. Babin, and L. Stemann (2021). "Marine Snow Morphology Illuminates the Evolution of Phytoplankton Blooms and Determines Their Subsequent Vertical Export." In: *Nat Commun* 12.1, p. 2816. ISSN: 2041-1723. DOI: [10.1038/s41467-021-22994-4](https://doi.org/10.1038/s41467-021-22994-4).
- Turner, J. T. (2004). "The Importance of Small Planktonic Copepods and Their Roles in Pelagic Marine Food Webs." In: *Zoological Studies* 43, pp. 255–266.
- (2002). "Zooplankton Fecal Pellets, Marine Snow and Sinking Phytoplankton Blooms." In: *Aquat. Microb. Ecol.* 27, pp. 57–102. ISSN: 0948-3055, 1616-1564. DOI: [10.3354/ame027057](https://doi.org/10.3354/ame027057).
- (2015). "Zooplankton Fecal Pellets, Marine Snow, Phytodetritus and the Ocean's Biological Pump." In: *Progress in Oceanography* 130, pp. 205–248. ISSN: 00796611. DOI: [10.1016/j.pocean.2014.08.005](https://doi.org/10.1016/j.pocean.2014.08.005).
- Uusitalo, L., J. A. Fernandes, E. Bachiller, S. Tasala, and M. Lehtiniemi (2016). "Semi-Automated Classification Method Addressing Marine Strategy Framework Directive (MSFD) Zooplankton Indicators." In: *Ecological Indicators* 71, pp. 398–405. ISSN: 1470160X. DOI: [10.1016/j.ecolind.2016.05.036](https://doi.org/10.1016/j.ecolind.2016.05.036).

- van der Lingen, C. et al. (2009). "Trophic Dynamics of Small Pelagic Fish." In: pp. 333–403.
- van der Lingen, C., L. Hutchings, and J. Field (2006). "Comparative Trophodynamics of Anchovy *Engraulis Encrasicolus* and Sardine *Sardinops Sagax* in the Southern Benguela: Are Species Alternations between Small Pelagic Fish Trophodynamically Mediated?" In: *African Journal of Marine Science* 28.3-4, pp. 465–477. ISSN: 1814-232X, 1814-2338. DOI: [10.2989/18142320609504199](https://doi.org/10.2989/18142320609504199).
- Vilgrain, L., F. Maps, M. Picheral, M. Babin, C. Aubry, J.-O. Irisson, and S.-D. Ayata (2021). "Trait-based Approach Using in Situ Copepod Images Reveals Contrasting Ecological Patterns across an Arctic Ice Melt Zone." In: *Limnol Oceanogr* 66.4, pp. 1155–1167. ISSN: 0024-3590, 1939-5590. DOI: [10.1002/lno.11672](https://doi.org/10.1002/lno.11672).
- Villa-Alfageme, M., F. de Soto, F. A. C. Le Moigne, S. L. C. Giering, R. Sanders, and R. García-Tenorio (2014). "Observations and Modeling of Slow-Sinking Particles in the Twilight Zone: SLOW SINKING PARTICLES IN TWILIGHT ZONE." In: *Global Biogeochem. Cycles* 28.11, pp. 1327–1342. ISSN: 08866236. DOI: [10.1002/2014GB004981](https://doi.org/10.1002/2014GB004981).
- Volk, T. and M. I. Hoffert (1985). "Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO<sub>2</sub> Changes." In: *Geophysical Monograph Series*. Ed. by E. Sundquist and W. Broecker. Washington, D. C.: American Geophysical Union, pp. 99–110. ISBN: 978-1-118-66432-2 978-0-87590-060-5. DOI: [10.1029/GM032p0099](https://doi.org/10.1029/GM032p0099).
- Ware, D. M. and R. E. Thomson (2005). "Bottom-Up Ecosystem Trophic Dynamics Determine Fish Production in the Northeast Pacific." In: *Science* 308.5726, pp. 1280–1284. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1109049](https://doi.org/10.1126/science.1109049).
- Weber, T. and D. Bianchi (2020). "Efficient Particle Transfer to Depth in Oxygen Minimum Zones of the Pacific and Indian Oceans." In: *Front. Earth Sci.* 8, p. 376. ISSN: 2296-6463. DOI: [10.3389/feart.2020.00376](https://doi.org/10.3389/feart.2020.00376).
- Wiebe, P. H., S. H. Boyd, M. Davis, and J. L. Cox (1982). "Avoidance of Towed Nets by the Euphausiid *Nematoscelis Megalops*." In: *Fishery Bulletin* 80, pp. 75–91.
- Wiebe, P. H., A. W. Morton, A. M. Bradley, R. H. Backus, J. E. Craddock, V. Barber, T. J. Cowles, and G. R. Flierl (1985). "New Development in the MOCNESS, an Apparatus for Sampling Zooplankton and Micronekton." In: *Mar. Biol.* 87.3, pp. 313–323. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00397811](https://doi.org/10.1007/BF00397811).
- Wilkinson, M. D. et al. (2016). "The FAIR Guiding Principles for Scientific Data Management and Stewardship." In: *Sci Data* 3.1, p. 160018. ISSN: 2052-4463. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- Williams, J. R. and S. L. C. Giering (2022). "In Situ Particle Measurements Deemphasize the Role of Size in Governing the Sinking Veloc-

- ity of Marine Particles." In: *Geophysical Research Letters*. ISSN: 0094-8276, 1944-8007. DOI: [10.1029/2022GL099563](https://doi.org/10.1029/2022GL099563).
- Zeitzschel, B., P. Diekmann, and L. Uhlmann (1978). "A New Multisample Sediment Trap." In: *Mar. Biol.* 45.4, pp. 285–288. ISSN: 0025-3162, 1432-1793. DOI: [10.1007/BF00391814](https://doi.org/10.1007/BF00391814).
- Zhang, M., D. Straile, F. Chen, X. Shi, Z. Yang, Y. Cai, J. Yu, and F. Kong (2018). "Dynamics and Drivers of Phytoplankton Richness and Composition along Productivity Gradient." In: *Science of The Total Environment* 625, pp. 275–284. ISSN: 00489697. DOI: [10.1016/j.scitotenv.2017.12.288](https://doi.org/10.1016/j.scitotenv.2017.12.288).
- Zhou, H., Y. Yang, and W. Qian (2019). "Tweedie Gradient Boosting for Extremely Unbalanced Zero-inflated Data." In: *arXiv:1811.10192 [stat]*. arXiv: [1811.10192 \[stat\]](https://arxiv.org/abs/1811.10192).

This document was typeset with L<sup>A</sup>T<sub>E</sub>X using a modified version of the classicthesis developed by André Miede and Ivo Pletikosić.