

**GLOBAL DISTRIBUTION AND BIOMASS OF THE  
MESOPELAGIC MESOZOOPLANKTON AND MICRONEKTON  
COMMUNITY**

by

Yulia Egorova

B.Sc., Thompson Rivers University, 2012

M.Sc., University of British Columbia, 2016

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Oceanography)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

November 2023

© Yulia Egorova, 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Global Distribution and Biomass of the Mesopelagic Mesozooplankton and Micronekton Community

submitted by Yulia Egorova in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

in Oceanography

**Examining Committee:**

Dr. Evgeny A. Pakhomov, Professor, Department of Earth, Ocean, and Atmospheric Sciences and Institute for the Oceans and Fisheries, UBC

*Supervisor*

Dr. William W. L. Cheung, Professor, Institute for the Oceans and Fisheries, UBC

*Supervisory Committee Member*

Dr. Susan Allen, Professor, Department of Earth, Ocean, and Atmospheric Sciences, UBC

*University Examiner*

Dr. Scott Hinch, Professor, Department of Forest and Conservation Sciences, UBC

*University Examiner*

Dr. Les Watling, Professor, School of Life Sciences, University of Hawaii

*External Examiner*

**Additional Supervisory Committee Members:**

Dr. Marie Auger-Méthé, Associate Professor, Institute for the Oceans and Fisheries, Statistics, UBC

*Supervisory Committee Member*

Dr. Villy Christensen, Professor, Institute for the Oceans and Fisheries, UBC

*Supervisory Committee Member*

Dr. Brian P. V. Hunt, Associate Professor, Department of Earth, Ocean, and Atmospheric Sciences and Institute for the Oceans and Fisheries, UBC

*Supervisory Committee Member*

## Abstract

Despite the ecological importance of mesopelagic zooplankton, very little is known about their global distribution in the midwater realm, and even less information is available on global trends in their diversity. This study is the first attempt to explore the distribution, diversity gradients, and total biomass of mesopelagic mesozooplankton and micronekton in the global ocean. This study created the first quantitative Mesopelagic Mesozooplankton and Micronekton Database, organizing and standardizing a broad spectrum of literature sources for effective data analysis. Analyzing 861 mesozooplankton species using a range of species distribution models, the study found that the highest species richness was observed in the North Atlantic region, the west coast of India, and the Mediterranean Sea. Factors such as water temperature, euphotic zone depth, salinity, and dissolved nitrate concentration were essential for explaining their distribution. The results of this study indicate that NPP and POC are important factors that influence the biomass of mesopelagic mesozooplankton. The global mesozooplankton population in the mesopelagic zone was estimated using estimates of particulate organic carbon (POC) and primary productivity (NPP). Linear models fitted to predict mesopelagic mesozooplankton biomass using NPP and POC data were able to explain a moderate-to-substantial proportion of variance. The distribution patterns of mesopelagic mesozooplankton biomass showed enhanced values in certain regions, such as the Northern Hemisphere, west coasts of continents, and equatorial and 50°S bands. Global mesopelagic mesozooplankton biomass was estimated to range between 0.20-0.91 PgC, depending on the method used. Overall, this study provides an estimate of mesopelagic mesozooplankton biomass using the classical definition of the mesopelagic zone, and an estimate that considers the variable depth of this layer. The findings indicated that POC was a more reliable indicator of mesozooplankton biomass than NPP, particularly in more northerly latitudes. The research also showed that low mesopelagic mesozooplankton biomass is linked to low POC levels and low range-size rarity and species diversity, but higher biomass values are more often found in areas with moderate rarity values and high species diversity, combined with high POC.

## **Lay Summary**

This study aimed to understand the distribution and diversity of mesopelagic zooplankton, which play a crucial role in marine ecosystems. By analyzing data collected from various sources, I found that the highest species richness of these zooplankton was observed in the North Atlantic region, west coast of India, and Mediterranean Sea. Factors like temperature, depth, salinity, and nitrate concentration influenced their distribution. The study estimated the global population and biomass of mesopelagic zooplankton using particulate carbon and productivity data. The results showed that particulate carbon levels were a reliable indicator of biomass, especially in northern latitudes. Regions with moderate rarity and high species diversity had higher biomass values. The study provides valuable insights into the global dynamics of mesopelagic zooplankton and their relationship with environmental factors.

## Preface

This dissertation is an original work by Y. Egorova. In all chapters, I am the first author and personally led data collection and analyses, project development, and writing of all original manuscripts and chapters.

While dissertation guidelines often require the use of ‘I’ when describing one's own work, it is important for me to acknowledge the invaluable contributions of several individuals with whom I had the privilege of collaborating on various chapters of this thesis. Chapter 1 greatly benefited from the insightful edits provided by Dr. L. Kwong. In Chapter 2 Section 2.4, I had the opportunity to work alongside Nirupama Tamvada and Daniel Forrest during our FISH506 course. I was the project leader and was responsible for all major areas of concept formation, data preparation, and analysis, as well as for the majority of the manuscript composition. Additionally, Chapters 3 and 4 were written in collaboration with Dr. G. Reygondeau, and I am grateful for the valuable input and ideas shared by Dr. E. Pakhomov and Dr. W. Cheung. Their contributions significantly enhanced the quality and depth of the research presented in this thesis.

I am deeply grateful for the collaborative efforts of numerous individuals who have played a crucial role in making this work possible through generous data sharing. Specifically, I would like to express my sincere appreciation to Dr. E. Pakhomov and Dr. A. Yamaguchi for providing access to their valuable data. Furthermore, I extend my gratitude to Sutton et al. (2017) for sharing the mesopelagic boundaries data, Proud et al. (2017) for generously providing the mesopelagic provinces shapefiles, and Strömberg et al. (2009) for their kind contribution of the data on global epipelagic zooplankton biomass. Their willingness to share these resources has been instrumental in advancing our understanding of this field and has greatly contributed to the research presented in this work.

# Table of Contents

Abstract .....	iii
Lay Summary.....	iv
Preface.....	v
Table of Contents .....	vi
List of Tables .....	vii
List of Figures .....	viii
List of Abbreviations.....	xii
Acknowledgements .....	xiii
Dedication.....	xiv
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Mesopelagic ocean .....	1
1.2 Partitioning of the mesopelagic ocean .....	2
1.3 Mesopelagic mesozooplankton and micronekton community .....	9
1.4 Life of the mesopelagic.....	11
1.5 Gaps in knowledge.....	19
1.6 Project Objectives .....	22
<b>Chapter 2: Mesopelagic Mesozooplankton and Micronekton Database .....</b>	<b>23</b>
2.1 Introduction .....	23
2.2 Data & Methods.....	28
2.3 Database Standardization .....	40
2.4 Case Study:.....	48
2.5 Concluding remarks .....	50
<b>Chapter 3: Global Distribution and Diversity of Mesopelagic Mesozooplankton .....</b>	<b>53</b>
3.1 Introduction .....	53
3.2 Methods.....	55
3.3 Results.....	63
3.4 Discussion.....	73
3.5 Concluding remarks .....	84
<b>Chapter 4: Global Estimate of Mesopelagic Mesozooplankton Biomass .....</b>	<b>87</b>
4.1 Introduction .....	87
4.2 Methods.....	92
4.3 Results.....	97
4.4 Discussion.....	107
4.5 Concluding remarks .....	117
<b>Chapter 5: Synthesis and Conclusion.....</b>	<b>119</b>
<b>References .....</b>	<b>124</b>
<b>Appendices .....</b>	<b>136</b>
Appendix A - Supplementary information for Chapter 1 .....	137
Appendix B - Supplementary information for Chapter 2.....	180
Appendix C - Supplementary information for Section 2.4.....	210
Appendix D - Supplementary information for Chapter 3 .....	222
Appendix E - Supplementary information for Chapter 4.....	243

## List of Tables

Table 1.1: Summary of available pelagic ocean partitioning. ....	4
Table 2.1: Summary of available reviews/syntheses of mesopelagic organisms, aspects being studied, and region covered. ....	27
Table 2.2: Taxonomic resolution. ....	34
Table 2.3: Summary of number of unique species, genera, families, and orders for each phylum-class recorded in the Mesopelagic Mesozooplankton and Micronekton Database. ....	36
Table 2.4: Top 10 sampling methods in the Mesopelagic Mesozooplankton and Micronekton Database. ....	38
Table 2.5: Frequency of occurrence of data recorded in different abundance/ biomass units in the mesopelagic database. ....	39
Table 2.6: Distribution of various net classes in the Mesopelagic Mesozooplankton and Micronekton Database. ....	42
Table 3.1: Available databases of the environmental parameters used in this study. ....	57
Table 4.1: Global mesozooplankton biomass estimates in the upper pelagic ocean. ....	89
Table 4.2: Estimated global and 40°S-40°N means and standard errors of mesopelagic mesozooplankton biomass. ....	100
Table 4.3: Epipelagic to mesopelagic biomass ratio based on different estimates of the epipelagic and mesopelagic biomasses found in literature and calculated in this study. Mean and medial ratios are estimated for each column. ....	104
Table A1: Summary of recent expeditions in the mesopelagic realm and their major findings...	137
Table B1: Information collected from various published and unpublished sources for the mesopelagic mesozooplankton and micronekton databases. ....	180
Table B2: Number of entries, n, (with percentage, %) for 19 phyla from the Mesopelagic Mesozooplankton and Micronekton Database grouped by organism class and net size (Meso and Macro). ....	184
Table B3: Number of entries, n, (with percent, %) for 19 Phyla from the mesopelagic dataset grouped by organisms' order and net size (Meso and Macro) ....	185
Table C1: Variable descriptions of the subset of Calanoida species in the Southern Ocean. ....	212
Table C2: Comparison of Model Coefficients and Their Standard Errors. ....	218
Table C3: Inverse Gaussian GLM Model coefficients with and without the dispersion parameter ( $\pm$ standard error). ....	218
Table D1: Parameter settings of nine individual SDM in the biomod2 package in R and NPPEN and AQUAMEAN models. ....	239
Table D2: Evaluation metrics: receiver operating characteristic (ROC) values for all species distribution models included in ENSEMBLE. ....	241
Table D3: The difference in AUC ( $\Delta$ AUC) between pseudo-absence runs (PA1 vs PA2). ....	242
Table D4: The difference in AUC ( $\Delta$ AUC) between subsampling runs (RUN1 vs RUN2). ....	242
Table E1: Models Performance Metric for NPP and POC-derived models using classical definition of the mesopelagic zone (200-1,000 m) and variable depth model (mesopelagic zone depth is location-dependent). ....	243
Table E2: Models' Coefficients for NPP and POC-derived models using classical definition of the mesopelagic zone (200-1,000 m) and variable depth model (mesopelagic zone depth is location-dependent). ....	243

## List of Figures

Figure 1.1: The number of scientific publications related to the topic of vertical distribution of mesopelagic organisms retrieved from Web of Science.....	19
Figure 2.1: The depth distribution of OBIS records (kingdom animalia) in the mesopelagic zone (200-1,000 m). .....	24
Figure 2.2: Available OBIS records (kingdom Animalia) from the mesopelagic zone (200–1000 m) separated into eight strata (100 m bins).....	25
Figure 2.3: Sampling locations of A) compiled dataset B) existing databases (BioTIME, COPEPOD, JeDI) for depth bin 200-1,000 m. ....	31
Figure 2.4: Global distribution of records in the Mesopelagic Mesozooplankton and Micronekton Database. ....	32
Figure 2.5: Spatial distribution of maximum depth of the records. ....	33
Figure 2.6: Treemap of available phyla and classes in the Mesopelagic Mesozooplankton and Micronekton Database. ....	35
Figure 2.7: Phylogeny of life contained within the Mesopelagic Mesozooplankton and Micronekton Database for all taxa recorded at species levels. ....	37
Figure 2.8: Distribution of mesh sizes in the Mesopelagic Mesozooplankton and Micronekton Database. ....	38
Figure 2.9: Distribution of records in the Mesopelagic Mesozooplankton and Micronekton Database based on time of sampling.....	39
Figure 2.10: Histograms of mesh sizes (in $\mu\text{m}$ ) that are present in the Mesopelagic Mesozooplankton and Micronekton Database.....	42
Figure 2.11: Example of calibration procedure for order Euphausiacea colored by different mesh size (NetClass of 50, 150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	47
Figure 3.1: The Species Distribution Model workflow depicts the different analytical steps performed for each species. ....	60
Figure 3.2: Sum of habitat suitability indices for the mesopelagic mesozooplankton community (n=861) modelled by the ensemble (weighted mean) approach.....	64
Figure 3.3: Distributions of the area under the curve (AUC) scores of predictions from individual species distribution models and their ensemble. ....	66
Figure 3.4: Left column showing the difference in area under the curve scores ( $\Delta\text{AUC}$ ) as a function of number of occurrence points between two pseudo-absences runs (A) and between two subsampling runs (C).....	67
Figure 3.5: Mesopelagic mesozooplankton diversity maps.....	71
Figure 3.6: The natural logarithm of species richness as a function of water temperature ( $^{\circ}\text{C}$ ) in the mesopelagic zone and inverse thermal energy ( $1/kT$ , where k is Boltzmann's constant and T is temperature in Kelvin).....	72
Figure 4.1: Relationships between NPP (1 <sup>st</sup> column) or POC (2 <sup>nd</sup> column) and mesopelagic mesozooplankton biomass for Variable Depth Models (1 <sup>st</sup> row) and Classical Depth Models (2 <sup>nd</sup> row).....	97
Figure 4.2: NPP- (1 <sup>st</sup> column) and POC- (2 <sup>nd</sup> column) based estimates of mesopelagic mesozooplankton biomass in the global ocean using variable depth models (1 <sup>st</sup> row) and classical depth models (2 <sup>nd</sup> row).....	99



Figure 4.3: Global relationship between mesopelagic zooplankton biomass ( $\text{mgC}\cdot\text{m}^{-2}$ ) and potential species richness colored by the POC value for each cell.....	100
Figure 4.4: A) 2D map depicting the relationship between the mesopelagic mesozooplankton biomass derived from POC estimates and the potential species richness from Chapter 3. ....	101
Figure 4.5: Relationship between the mesopelagic mesozooplankton biomass and various diversity proxies.....	102
Figure 4.6: Global epipelagic to mesopelagic biomass ratio based on Stromberg et al. (2009) and Bogorov et al. (1968) epipelagic biomass maps. ....	105
Figure 4.7: Present-day mesopelagic biogeography derived by K-means clustering ( $k=10$ ).....	106
Figure 5.1: Graphical representation of the main components of this study, highlighting important connections between mesopelagic zooplankton diversity, food availability (POC concentrations), and mesopelagic biomass. ....	123
Figure B1: Histograms of logged densities (in $\text{indiv}\cdot\text{m}^{-3}$ ) of meso-(Meso) and macro-(Macro) zooplankton records from mesopelagic mesozooplankton and micronekton database grouped by phylum. ....	183
Figure B2: Example of calibration procedure for order Siphonophorae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	187
Figure B3: Example of calibration procedure for class Ostracoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	188
Figure B4: Example of calibration procedure for order Calanoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	189
Figure B5: Example of calibration procedure for order Cyclopoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	190
Figure B6: Example of calibration procedure for order Isopoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	191
Figure B7: Example of calibration procedure for order Amphipoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	192
Figure B8: Example of calibration procedure for class Polychaeta colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	193
Figure B9: Example of calibration procedure for order Phragmophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	194
Figure B10: Example of calibration procedure for order Aphragmophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ). ....	195
Figure B11: Example of calibration procedure for class Appendicularia colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	196
Figure B12: Example of calibration procedure for order Mormonilloida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	197
Figure B13: Example of calibration procedure for phylum Ctenophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	198
Figure B14: Example of calibration procedure for class Gastropoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	199
Figure B15: Example of calibration procedure for order Narcomedusae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	200

Figure B16: Example of calibration procedure for order Harpacticoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	201
Figure B17: Example of calibration procedure for class Bivalvia colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	202
Figure B18: Example of calibration procedure for order Decapoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	203
Figure B19: Example of calibration procedure for class Thaliacea colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	204
Figure B20: Example of calibration procedure for order Mysida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	205
Figure B21: Example of calibration procedure for class Cephalopoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	206
Figure B22: Example of calibration procedure for class Teleostei colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	207
Figure B23: Example of calibration procedure for phylum Echinodermata colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ). .....	208
Figure B24: Example of calibration procedure for order Trachymedusae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 $\mu\text{m}$ ).....	209
Figure C1: Histogram of zooplankton densities (in $\text{ind}\cdot\text{m}^{-3}$ ) in the dataset.....	212
Figure C2: Pairs plot of all model variables (both response and predictors).....	213
Figure C3: Temperature coefficient effects plots, with superimposed partial residuals. ....	217
Figure C4: Average depth coefficient effects plots, with superimposed partial residuals. ....	218
Figure D1: Number of missing biotic information collected in SealifeBase and FishBase for 1499 mesopelagic mesozooplankton and micronekton species.....	222
Figure D2: Pearson correlation coefficients in two sets of environmental variables .....	222
Figure D3: Distributions of the overall accuracy of different individual SDMs (all runs). ....	223
Figure D4: Time (in minutes) that took each species to run in relation to the number of presence data available. ....	223
Figure D5: Variable importance of nine environmental variables: .....	224
Figure D6: Comparison of habitat suitability index (HIS) between unrestricted and native range models for each mesopelagic province.....	225
Figure D7: Mesopelagic diversity of phylum Annelida. ....	226
Figure D8: Mesopelagic diversity of phylum Arthropoda.....	227
Figure D9: Mesopelagic diversity of phylum Chaetognatha. ....	228
Figure D10: Mesopelagic diversity of phylum Chordata. ....	229
Figure D11: Mesopelagic diversity of phylum Cnidaria. ....	230
Figure D12: Mesopelagic diversity of phylum Ctenophora. ....	231
Figure D13: Mesopelagic diversity of phylum Mollusca. ....	232
Figure D13: Average and total rarity plots using all species (n=861).....	233
Figure D14: Filtered total range-size rarity plots.....	234
Figure D15: Sampling effort.....	235
Figure D16: Variation in temperature in the mesopelagic zone. ....	236
Figure D17: Distribution of annual climatology data in the mesopelagic layer (MESO) .....	237

Figure D18: Distribution of annual climatology data in the combined epi- and mesopelagic layers (BOTH).....	238
Figure E1: Distribution of annual climatology data .....	244
Figure E2: Model Diagnostics for NPP (1 <sup>st</sup> column) and POC (2 <sup>nd</sup> column) models. ....	245
Figure E3: Density plots of log-transformed values of A) NPP and B) POC used in this study..	246
Figure E4: Average mesopelagic biomass in each mesopelagic province. ....	246
Figure E5: Digitized map of distribution of zooplankton biomass ( $\text{mg}\cdot\text{m}^{-3}$ wet weight) within the surficial layer of the world ocean based on map reported in (Bogorov et al., 1968)...	247
Figure E6: FAO (1972) biomass of zooplankton in the upper 100 m of the water column ( $\text{mg}\cdot\text{m}^{-3}$ wet weight). ....	248
Figure E7: 2D map depicting the relationship between the mesopelagic mesozooplankton biomass derived from POC estimates and FAO (1972) estimated total zooplankton biomass in a top 100 m of the water column. ....	248
Figure E8: K-means clustering ( $k = 10$ ) of gridded $\log_{10}$ -transformed biomass (derived from POC) and potential species richness (SR; estimated in Chapter 4). ....	249
Figure E9: Spatial distribution of mesopelagic mesozooplankton biomass (note $\log_{10}$ transformation of color scale) in the Mediterranean Sea. ....	249
Figure E10: Malaspina 2010 expedition cruise track showing the locations of the 51 stations sampled. ....	250

## List of Abbreviations

**ANN** – Artificial Neural Networks  
**AUC** – Area Under the Curve  
**CI** – Confidence Interval  
**CTA** – Classification Tree Analysis  
**DSL** – Deep-Scattering Layers  
**DVM** – Diel Vertical Migration  
**FDA** – Flexible Discriminant Analysis  
**GAIC** – Generalized Akaike Information Criterion  
**GAM** – Generalized Additive Models  
**GLM** – Generalized Linear Models  
**HNLC** – High-Nutrient Low-Chlorophyll  
**HSI** – Habitat Suitability Index  
**IQR** – Inter Quantile Range  
**LDG** – Latitudinal Diversity Gradient  
**MAXENT** – Maximum Entropy Modeling  
**MLD** – Mixed Layer Depth  
**MM** – Mesopelagic Meso zooplankton  
**MMM Database** – Mesopelagic Meso zooplankton and Micronekton Database  
**MOL** – Map of Life  
**NO<sub>3</sub>** – Nitrate Concentration  
**NPP** – Net Primary Production  
**NPPEN** – Non-Parametric Probabilistic Ecological Niche  
**O<sub>2</sub>** – Dissolved Oxygen Concentration  
**OBIS** – Ocean Biogeographic Information System  
**PO<sub>4</sub>** – Phosphate Concentration  
**PP** – Primary Productivity  
**RF** – Random Forests  
**ROC** – Receiver Operating Characteristic  
**Sal** – Salinity  
**SDM** – Species Distribution Model  
**SiO<sub>2</sub>** – Silicate Concentration  
**SSS** – Sea Surface Salinity  
**SST** – Sea Surface Temperature  
**SRE** – Surface Range Envelope (also known as BIOCLIM)  
**Temp** – Temperature  
**TSS** – True Skill Statistics  
**VIP** – Variable Importance Plots  
**WoRMS** – World Register of Marine Species  
**Zeu** – Euphotic Zone Depth

## **Acknowledgements**

This project would not have been possible without the support of many people. First and foremost, I am extremely grateful to my supervisor, Dr. Evgeny Pakhomov for his invaluable advice, continuous support, and patience during my PhD study. Evgeny's immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Thank you, Evgeny, for creating a close-knit and friendly atmosphere in our lab during my PhD. It felt like a small family. I would also like to thank Dr. Gabriel Reygondeau, without whom I would have graduated ten years later. Thank you, Gabriel, for explaining to me all the details about the modelling process and keeping up with my endless messages on Slack. I would like to extend my gratitude to our P-lab group for being my partners in crime during this crazy journey. I would like to thank each one of you for putting up with my constant stream of bad jokes, overly dramatic reactions, and my bad English. Finally, I would like to thank all the awesome staff at the coffee shops on the UBC campus for being faithful companions through countless 6 a.m. sessions.

## **Dedication**

This dissertation is dedicated to the memory of my beloved grandfather Vladimir I. Egorov. Although he was my inspiration for pursuing my doctoral degree, he was unable to see my graduation. This is for him.

I would also like to dedicate this work to my family and my partner, Ben Wang. Thank you for believing in me and encouraging me to pursue my dreams. This work is as much yours as it is mine.

# Chapter 1: Introduction

## 1.1 Mesopelagic ocean

The exploration of the global open ocean and deep seas (>200 m depth) holds paramount importance in the field of oceanography (Sutton *et al.*, 2017). These vast regions encompass a substantial portion of Earth's habitat, with a staggering volume exceeding 1.2 billion km<sup>3</sup> (Rogers, 2015) and a mean depth of 4.2 km. However, our understanding of the dynamics and ecological processes within the deep open ocean is severely limited. The scarcity of available information in these remote and challenging environments poses significant obstacles to comprehending the intricate workings of this vast realm. Bridging this knowledge gap is essential for unraveling the complexities of global oceanic systems, including biodiversity patterns, species distributions, trophic interactions, and the impacts of anthropogenic activities.

Comprising 71% of the Earth's surface (Grosberg *et al.*, 2012), oceans contain a significant number of unidentified species that remain to be explored and studied (Costello *et al.*, 2012). Angel (1993) and Gluchowska *et al.* (2017) showed that the composition of pelagic community across horizontal scale is 3 - 5 orders of magnitude coarser than the distribution across vertical scale (although fine vertical resolution is lost with depth). Therefore, the water column is usually partitioned into several zones based primarily on depth and its covariance with temperature and sunlight penetration (Ramirez-Llodra *et al.*, 2010; Angel, 2003).

The first biome is epipelagic (euphotic), located between the surface and a depth of ~200 m, and is characterized by substantial levels of sunlight penetrating the water column which allows primary production to occur. In addition to light availability, the rate of primary production is controlled by various abiotic factors such as nutrient availability and the depth of the mixing layer (Sutton, 2013; Charnock & Deacon, 1978).

The second biome is mesopelagic, also known as the twilight zone of the global ocean (Ramirez-Llodra *et al.*, 2010), and represents a part of the ocean where surface light is still detectable during the daytime, allowing organisms to distinguish between diurnal and nocturnal cycles (Sutton, 2013), but is at a level low enough that net primary production is no longer possible (Klevjer *et al.*, 2012; Proud *et al.*, 2017; Robinson *et al.*, 2010). Hedgpeth (1957) coined the term 'mesopelagic' as a noun, but the definition was unclear, except for its

positioning between the ‘epipelagic’ and ‘bathypelagic’ zones. No specific depth or oceanographic function boundaries were established, but according to Hedgpeth's original figures, the mesopelagic zone is situated approximately between 1.0 and 1.7 km. Recently, mesopelagic have been referred to as a zone in the ocean with a depth range between 200 and 1,000 m (Klevjer *et al.*, 2012; Sutton *et al.*, 2017; Gjosaeter & Kawaguchi, 1980; Proud *et al.*, 2017), and the choice of depth boundaries coincides with the maximum depth of seasonal variability in temperature found within this stratum (Ramirez-Llodra *et al.*, 2010; Sutton, 2013; Vereshchaka *et al.*, 2016; Herring, 2002). However, a recent study determined that the boundaries of the mesopelagic zone are not constant and vary tremendously across the globe (Reygondeau *et al.*, 2018).

Compared to the epipelagic zone, highly variable biophysical conditions are muted at deeper depths (Sutton, 2013). Nevertheless, the mesopelagic zone is not homogeneous and may contain strong environmental gradients at the border with the epipelagic zone (main thermocline) and within the mesopelagic zones, for example, oxygen minimum zones (Robison, 2003; Robinson *et al.*, 2010).

## **1.2 Partitioning of the mesopelagic ocean**

Unlike terrestrial ecosystems, where the distribution of organisms is usually two-dimensional, the distribution of life in the ocean is three-dimensional. In addition to the challenges of sampling 3D space, sampling of the deep ocean is labor- and cost-intensive (Berger, 2009). Consequently, sampling of the deep ocean is sparse and heterogeneous (Webb *et al.*, 2010). Thus, to fill in the gaps in knowledge regarding the horizontal and vertical distributions of mesopelagic zooplankton and micronekton, studies uncovering the spatial correlation between community composition and geographical location are required.

Biogeography helps define the distribution of ecosystems across space and time (Fay & McKinley, 2014; Cox & Moore, 2000). This is of interest because it allows the separation of the deep ocean into distinct provinces based on the parameters of interest (e.g., community composition or water chemistry). Such biogeographic partitioning helps to understand the spatial variability in ecosystem functioning (Proud *et al.*, 2017), explore the global distribution of different taxa (Fay & McKinley, 2014; Angel *et al.*, 2007), and explore the dynamics and



fate of different elemental fluxes (Boyd & Doney, 2003). In addition, the development of ecosystem-based management approaches and marine protected areas requires knowledge of the potential boundaries of ecologically important areas that correspond to specific taxa and/or community distributions.

It is critical to identify unique biogeographic areas of the mesopelagic zone to help policy planning and sustainable management, and to assess the impact of climate change and human-induced activities, as well as the effects of natural fluctuations of environmental parameters on mesopelagic community structure and composition. Biogeographic partitioning also helps to assess which spatial-scale deep-pelagic communities are similar in composition or abundance and expands our knowledge in understanding the biogeography of the deep ocean (Sutton *et al.*, 2017). Such knowledge will help 1) identify ecologically important areas/regions most impacted by global climate change, 2) assess the economic value of ecological provinces, and 3) determine the role each area plays in global food and environmental security. Specifically, identifying the biogeographical distributions of zooplankton and micronekton communities is required to determine trophic relationships, food web structures, and biogeochemical pathways in the global ocean (McIvor, 2011).


Several attempts have been made to partition the global mesopelagic ocean into distinct provinces. However, prior to presenting the synthesis of these efforts, it is imperative to delve into the central concept discussed by Spalding *et al.* (2012), which revolves around various levels of partitioning. The largest ocean partitioning is performed according to realms (or regions) that are defined by areas with similar biota at a higher taxonomic level (generic or family levels) established as a result of common evolutionary history. Biomes are finer partitions defined by common oceanographic features (e.g., ocean gyres or currents). Thus, this type of partitioning is more environmental-based, and as a result, it delineates ecologically similar ecosystems that share the same properties and functionalities but do not necessarily contain the same species. The global ocean can also be partitioned into distinct provinces, areas that contain a distinct species community that is established by prominent spatially and temporarily operating oceanographic drivers (upwellings, continental shelf and etc.).

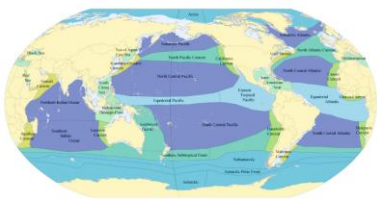
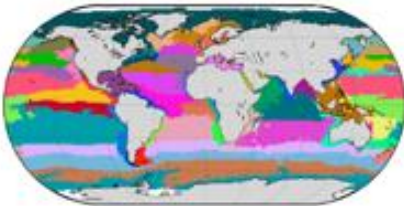
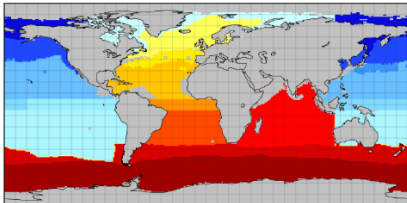
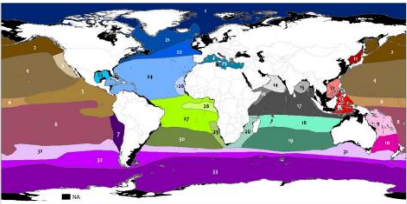
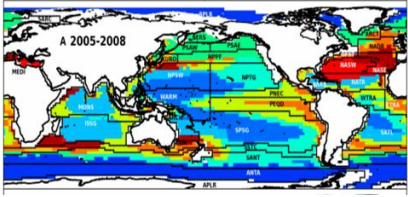
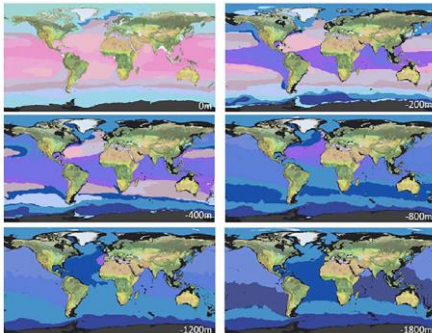
One of the most commonly known partitioning is the ecological provinces identified by Longhurst (2007). Technically, partitioning was not performed for the mesopelagic zone

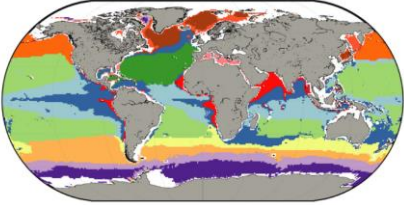
because surface environmental data, such as ocean currents, fronts, and chlorophyll concentrations, were employed in designing the ecological provinces. He proposed that the community of marine organisms within each province was less different than that of communities from other provinces. This epipelagic biological partitioning has been commonly used for comparison purposes in many mesopelagic research papers (Proud *et al.*, 2017; Oliver & Irwin, 2008; McIvor, 2011).

Spalding *et al.* (2012) synthesized all the available biogeography of the pelagic ocean present at that time and came up with their own partitioning of the off-shelf waters based on both known taxonomic and oceanographic biogeography. The authors also delineated the difference between different levels of global ocean classification (e.g., realm vs. biome vs. province). However, because of the scarcity of sampling in the deep ocean, they again focused on the epipelagic area and came up with seven biomes encompassing 37 pelagic provinces, acknowledging that at deeper depths, the difference between some provinces may be eliminated, resulting in a few unique provinces. However, this work has rarely been cited or referred to in biogeographic research (see Table 1.1). In addition, the authors confirmed a very tight linkage between taxonomic and non-taxonomic classifications, with the exception of several anomalies where the regions were split into smaller areas despite taxonomic integrity.

Table 1.1: Summary of available pelagic ocean partitioning.

Map	Authors	# of provinces	Parameters used
	Longhurst (2007)	56 biochemical provinces (32 excluding coastal areas)	sea surface temperature (SST), sea surface salinity (SSS), bathymetry, sea surface wind stress and height, chlorophyll concentration (Chl a), photic and mixed layer depth (MLD), primary productivity (PP) from satellite data

Map	Authors	# of provinces	Parameters used
	Spalding <i>et al.</i> (2012)	37 pelagic provinces	taxonomic and physiognomic patterns.
	Reygondeau <i>et al.</i> (2013)	56 biogeochemical provinces	SST, SSS, bathymetry, Chl a
	Fay & McKinley (2014)	17 global open-ocean (time-varying) biomes	Maximum MLD, spring/summer Chl a, SST, and sea ice fractional coverage
	Sutton <i>et al.</i> (2017)	33 ecoregions	Temperature extremes, salinity, dissolved oxygen (O <sub>2</sub> ) at 200, 500, 750 and 1000 m depths, surface water productivity
	Proud <i>et al.</i> (2017)	36 provinces are shown (best at 60 provinces)	surface PP, Temperature at deep scattering layer and wind stress
	Sayre <i>et al.</i> (2017)	37 distinct volumetric region units (ecological marine units) with 6 mesopelagic and 7 bathyal-pelagic	Temperature, Salinity, dissolved oxygen, nitrate, phosphate, and silicate at variable depth increments (5 m at the surface – 100 m at the bottom)

Map	Authors	# of provinces	Parameters used
	(Reygondeau <i>et al.</i> , 2018)	13 Mesopelagic BioGeoChemical Province (MBGCP)	Temperature, salinity, dissolved oxygen concentration, nutrient concentrations (NO <sub>3</sub> , SiO <sub>2</sub> , PO <sub>4</sub> ) and particulate organic carbon (POC) flux.

Proud *et al.* (2017) developed deep scattering layer (DSL) -based mesopelagic geography. The study confirmed a strong link between surface parameters (primary productivity or wind stress) and mesopelagic community biomass, but also showed the importance of conditions within the mesopelagic zone (water temperature at depth). This finding agrees with the work of Oliver & Irwin (2008), who also found that satellite data were sufficient to distinguish between provinces (although their work was primarily done for the epipelagic ocean). The biomass of the community was not estimated in the study because of the lack of information on community structure and size distribution; however, acoustic backscattering was used as a proxy for biomass.

The best model of Proud *et al.* (2017) revealed 22 classes, yielding 60 distinct provinces, with an increasing number of clusters producing finer-scale patterns linked to frontal features. The presence of fine-scale features in the mesopelagic community disagrees with the findings of Steinberg *et al.* (2008) and Stemmann *et al.* (2008b), who determined that mesozooplankton community distribution is structured on a larger global scale and does not necessarily reflect much finer scales such as oceanic frontal systems. The authors noted that the fine-scale pattern might be reflected only in the total biomass, rather than in the community structure. However, a 10-class model was discussed in Proud *et al.* (2017) instead of a 22-class one (corresponding to 36 provinces) for the convenience of comparison with provinces defined by Longhurst (2007). Such partitioning was very promising, as it revealed the fine-scale details of the global mesopelagic ocean zonation, despite being entirely based on the proxy of the biomass of the scattering layer rather than the community structure.

Partitioning of the mesopelagic ocean by Sutton *et al.* (2017) was based on ecoregions, for example, areas containing distinct faunal communities, rather than species range distributions.

Because of the sparsity and patchiness of environmental and biological data, the opinions of experts in different areas, such as physical oceanography, ecology, and deep-pelagic taxonomy, were used during the panel discussion to refine the mesopelagic boundaries. The distribution dynamics of gelatinous zooplankton, fishes, mollusks, crustaceans, and chaetognaths were considered in biogeographic partitioning along with various environmental parameters. Owing to a lack of available information, zonation of the mesopelagic zone was performed based on the entire water column (200-1000 m). Although such classification is a good starting point for mesopelagic biogeography, it is mostly skewed towards the upper portion of the mesopelagic zone as more information is available for the 200-400 m layer.

Another recent ocean partitioning was done by Sayre *et al.*(2017), where the entire global ocean was separated into distinct ‘Ecological Marine Units’ (EMU) entirely based on the ocean chemistry. Thirteen zones were identified within the mesopelagic zone: six EMU were found exclusively in the mesopelagic zone, while seven EMU were present throughout the water column (bathyal-pelagic EMU). As no information on marine organisms was included, the partitioning process did not assign geographically separated areas into different regions. As a result, there were only 13 unique EMU in the mesopelagic zone compared to the 33 unique provinces identified by Sutton *et al.*(2017), who considered faunal communities. However, Sayre *et al.*(2017) demonstrated that mesopelagic zonation varies with depth; therefore, mesopelagic partitioning should not be constant throughout the water column.

All researchers have acknowledged that the boundaries of the designed provinces are not static and vary with time. Reygondeau *et al.* (2013) explicitly showed how recreated provinces from Longhurst(2007) change with the season. Seasonal shifts in the boundaries of the ecological provinces were especially prevalent at high latitudes. However, the annual differences were not computed because the seasonal data were averaged for a 20-year period (1997-2007). Their model was able to capture longitudinal variation produced by the monsoon, upwelling, and other climatic events, but because of the rather coarse resolution of the data (i.e.,  $1^\circ \times 1^\circ$  grid), much finer scale features (eddies or currents) were not depicted by the model, unlike the findings of Proud *et al.* (2017).

Fay & McKinley (2014) explored large-scale patterns (scale of ocean gyres) of ocean circulation and identified 17 open-ocean biomes based only on a set of four environmental

drivers (SST, Chl a, ice fraction and maxMLD). The chosen environmental variables are better suited to characterizing the epipelagic ocean rather than mesopelagic. Their work excluded coastal areas and upwelling influenced regions. The authors acknowledged the dynamic nature of such biomes and computed the shift in the biomes' boundaries between 1998 and 2010. In addition, they illustrated the 'core biome map' with areas that remained unchanged during the 13-year period. Similar to the Reygondeau *et al.* (2013) findings, Fay & McKinley (2014) also found that high-latitude biomes have more dynamic boundaries with more than 15% of areas being excluded from the core biome map.

Most recently, Reygondeau *et al.* (2018) extended the work of Longhurst (2007) and statistically derived a new biogeochemical classification of the mesopelagic zone. The authors also updated the current definition of upper and lower boundaries of the mesopelagic realm and illustrated the difference in boundaries' depth around the globe when the vertical gradient of environmental parameters was considered.

Reygondeau *et al.* (2018) proposed to determine the upper mesopelagic boundary as the deepest layer where net primary production remains feasible using the critical depth definition from Sverdrup (1953). Sverdrup (1953) defined critical depth as the depth in the water column where the integrated net growth rate of phytoplankton (growth minus losses due to respiration and other factors) over the entire water column from the surface to that depth is zero. Owing to the absence of global climatology for vertical primary production profiles, the boundary depth has been quantified using the vertical limitation of photosynthetic activity as the shallowest depth between the euphotic depth and the mixed layer depth. The delineation between the mesopelagic and bathypelagic zones was established at the point where variations in the descent of POC become minor, assessed over successive 5-metre intervals. This particular depth is significant because it denotes a decrease in the flux of organic carbon resulting from biogeochemical processes and shows that shallower marine ecosystems have less of an impact on the deeper oceanic environment. It is noteworthy that the Reygondeau *et al.* analysis was performed at the 100 km scale using climatological data, which can smooth both local and seasonal patterns.

The authors demonstrated that epi-mesopelagic boundary could vary from 10 to 150 m with the deepest zones found in subtropical and polar regions while the shallowest boundary limits

were found in the tropics and areas of high productivity (Reygondeau *et al.*, 2018). The depths in the majority of cases were ~60m, which is less than a commonly accepted boundary depth of 200 m. The depth of the meso-bathypelagic boundary showed a latitudinal trend with depth decreasing polewards. The lower boundary of the mesopelagic zone varied between 180 and ~2,300 m (Reygondeau *et al.*, 2018). As a result, the total volume of the mesopelagic zone increased and reached ~31% of the global ocean volume, compared to previously suggested volume of ~19% calculated using the conventionally accepted 200-1,000 m boundary.

### **1.3 Mesopelagic mesozooplankton and micronekton community**

The mesopelagic encompasses at least 20% of the global ocean volume and plays an important role in biogeochemical cycling (Proud *et al.*, 2017; St. John *et al.*, 2016). The mesopelagic layer is important in both controlling marine productivity and sequestering anthropogenic carbon dioxide. Biological processes (e.g., plankton trophic dynamics, vertical migrations, decomposing rate) are highly significant to air-sea interactions (Siegel *et al.*, 2014; Bode *et al.*, 2018). The processes sustaining the twilight zone and its vulnerability to perturbations are, however, poorly known and far from being reliably quantified (Giering *et al.*, 2014; Choy *et al.*, 2015; Gloeckler *et al.*, 2018). Yet, more than 90% of organic matter sinking from upper layers decomposes within the mesopelagic zone (Fasham *et al.*, 2001). The escaping organic matter is transferred to the ocean interior and can be removed from the atmosphere for millennia. Thus, the organic matter removal efficiency is set in the mesopelagic realm. Despite extensive efforts focused on the construction of biochemical models in the upper ocean, biological removal of carbon from the surface and export to deeper waters, the ‘biological pump’, and its sensitivity to climate change and pollution have received limited attention (Steinberg & Landry, 2017). Various marine organisms inhabiting the mesopelagic zone contribute to the repackaging and reprocessing of sinking and suspended organic particles. In addition, many of the inhabitants of the zone perform diel vertical migration (DVM) travelling to the surface at dusk, thus actively transporting organic carbon (Robinson *et al.*, 2010).

During the past decade, we have learned that mesopelagic ecosystems may be very diverse and highly structured vertically and horizontally (Robinson *et al.*, 2010; Proud *et al.*, 2017; St. John *et al.*, 2016). Previous assumptions that the mesopelagic zone might not be affected by climate change and considered as a constant in models are no longer acceptable. This paradigm change

requires detailed knowledge of the structure and function of the mesopelagic ecosystem. Many mesopelagic organisms undertake extensive vertical migrations on either a daily or seasonal bases, occupying productive surface waters at night and descending to mesopelagic during the dawn to reduce predation risks or undertake diapause on seasonal scales. These migrations appear to contribute significantly to the rapid vertical transport of organic material from epipelagic to mesopelagic zones (Giering *et al.*, 2014). Models predict that active carbon transport may contribute up to half of the total respiration within the layers affected by migration (Davison *et al.*, 2015). While widespread in the ocean, diel vertical migrations of plankton and micronekton are generally not reflected in global geochemical models. This is prevented by inadequate taxonomic and quantitative knowledge of mesopelagic ecosystem dynamics. Absence of such understanding currently prevents well-founded parametrization of mesopelagic processes in global biochemical models and impairs our ability to predict accurately the processes driving the exchange in carbon dioxide and nutrients among the euphotic zone, the seafloor, and the ocean margins. Moreover, such knowledge is fundamental for assessing oceanic perturbations due to climate change and pollution, potential anthropogenic interventions (geo-engineering, e.g., iron fertilization), and growing interest in exploitation of mesopelagic (fish, squid and crustaceans) resources (Wishner *et al.*, 2013; Irigoien *et al.*, 2014).

This thesis focuses on two parts of the mesopelagic community: mesozooplankton and micronekton. Mesozooplankton is typically defined as zooplankton with a size range of 0.2-20 mm and micronekton (that also includes macrozooplankton) as organisms with a size range of 2-20 cm long. The mesozooplankton community is primarily made up of crustacean plankton (copepods, euphausiids, amphipods), meroplanktonic larva and small gelatinous zooplankton (Moriarty & O'Brien, 2013). These organisms are usually caught with nets with mesh sizes ranging from 200 to 333  $\mu\text{m}$  (Harris *et al.*, 2014; Moriarty & O'Brien, 2013). They are located near the bottom of the food web and feed on phytoplankton, microzooplankton, detritus, and sometimes small mesozooplankton (Buitenhuis *et al.*, 2006). They are important food sources for higher trophic levels (epipelagic and air-breathing organisms). For instance, krill, amphipods, and copepods are prey of mesopelagic fishes (Williams & Koslow, 1997; Williams *et al.*, 2001). In turn, krill and different mesopelagic fishes are important food sources for seabirds, different species of seals, cephalopods, and baleen and beaked whales (Hopkins



*et al.*, 1993; Sato & Benoit-Bird, 2017; Naito *et al.*, 2017). Mesopelagic fishes, major component of micronekton community, are among the major components of the global carbon cycle. New estimates suggest that mesopelagic fishes respire ~10% of the primary production in the deep pelagial region (Irigoien *et al.*, 2014) and can contribute up to 15% of the total oceanic carbonate production in their intestines (Wilson *et al.*, 2009).

Micronekton and mesozooplankton provide a critical trophic link between the upper and lower parts of the ocean foodweb and contribute significantly to the global carbon cycling (Ramirez-Llodra *et al.*, 2010; Klevjer *et al.*, 2012; Robinson *et al.*, 2010). Micronekton and zooplankton communities contribute to CO<sub>2</sub> removal from the atmosphere via DVM. Carbon is transferred between the depth and surface in the form of gut contents and bodies, e.g., active carbon transport (Bogorov *et al.*, 1968; Tseitlin, 1986; Koppelman & Frost, 2008). Although this carbon flux is widely acknowledged, the extent of this process remains poorly understood (Robinson *et al.*, 2010; Kwong, 2016). Despite their ecological significance and potentially considerable biomass, little is known about the mesopelagic micronekton and mesozooplankton composition, distribution, and standing stock (Béhagle *et al.*, 2015). Quantification of zooplankton over vertical and horizontal scales is essential for predicting vertical fluxes of organic matter to the deep ocean (Stemmann *et al.*, 2008a) because zooplankton are known to fragment, re-mineralize, and consolidate sinking particles through the water column (Turner, 2004; Stemmann *et al.*, 2008a).

## **1.4 Life of the mesopelagic**

### *1.4.1 Early explorations of the mesopelagic realm*

Over the course of history, the deep ocean has remained a mysterious and largely unexplored realm. In 1521, Ferdinand Magellan deployed a cannonball at a depth of approximately 750 m and failed to reach the bottom (Berger, 2009). This led to the belief that the ocean was immeasurably deep and lifeless because of a lack of currents, oxygen, and nutrients.

Vinogradov (1968) provided a thorough history of pelagic ocean exploration through the 19<sup>th</sup> century, while the physical environment was described by Thorpe (1996). Here, I describe some of the most important discoveries to date regarding deep pelagic organisms.

During the first half of the 19<sup>th</sup> century, most researchers denied the existence of life at greater depths, and some concluded that the degradation of organic matter at depths was not possible due to high pressures (Maury in Vinogradov, 1968). Despite this skepticism, Edward Forbes, a well-known marine zoologist, published a vertical zonation of Atlantic fauna in which life ceased to exist at depths greater than 550 m; the zone was therefore named 'azoic' (Forbes, 1859; Proud *et al.*, 2017; Berger, 2009). Although Forbes is most known for incorrectly defining this vertical stratum as a lifeless abyss, he contributed tremendously to the advancement of the marine biological research during the first half of the 19<sup>th</sup> century.

In the mid-19<sup>th</sup> century, the Samuel Morse's invention of the telegraph system sparked the idea of connecting Mediterranean countries with an underwater cable and, subsequently, a transatlantic cable between Europe and the USA (Vinogradov, 1968). The first cable was laid across the Straits of Dover in November 1851 (Thorpe, 1996), and in 1861, the broken cable connecting Sardinia and Algeria was brought to the surface from a depth of 2160 m for repair, revealing that it had been colonized by different forms of life. The discovery of sea lilies, originally thought to be extinct, further prompted scientists to question whether the deep ocean was truly lifeless and, rather, may be inhabited by many presumed extinct or unknown species (Vinogradov, 1968).

Despite its remote location, exploration of the Southern Ocean began early (Voronina, 1984). Captain Cook was the first person to report data on Antarctic plankton (1772-1775), while the first research paper, dedicated to the Antarctic phytoplankton, was published in 1844 by Christian Gottfried Ehrenberg based on the Ross expedition between 1839 and 1843 (Ehrenberg, 1844). From the mid-19<sup>th</sup> century until the start of World War I, approximately 30 expeditions to the Southern Ocean were undertaken (e.g., Challenger 1874, Discovery, Terra Nova, Aurora; described in Voronina, 1984), although most plankton studies were restricted to the top 200 m of the water column (Ward *et al.*, 2014).

In 1872, the first large-scale expedition of the British naval vessel, the Challenger, sailed around the globe, conducting numerous midwater trawls and bottom dredges, providing more evidence of life below 2,000 fathoms (3,658 m) (Murray, 1895; Russell, 1927). However, because open sampling nets were used, there was insufficient evidence to conclude that organisms were collected from the depth rather than from the surface during this voyage.

Further expeditions focused primarily on benthic fauna (Berger, 2009). During these expeditions, plankton were collected as a ‘contaminant,’ because the open net was brought to the surface. The lack of closing nets made it impossible to resolve the vertical distribution of this pelagic biota. The first open-closing nets were constructed in 1884 and proved that pelagic organisms were not restricted to surface layers but found throughout the water column. However, deep-water regions were not sampled, leaving the possibility of deep pelagic biota under question (Vinogradov, 1968).

Exploration of the pelagic community in the 19<sup>th</sup> century was largely qualitative, with expeditions providing limited information on the midwater zone. In 1887, Agassiz conducted one of the first investigations of the pelagic community and validated the presence of a unique pelagic community in the deep ocean (Agassiz, 1902). However, his findings showed an ‘empty’ midwater zone where no life was caught, whereas abundance increased at the surface and seafloor. These findings were a circumstance of the location sampled (west tropical Pacific), which is well known for its low productivity and subsequent low abundance of mesopelagic organisms. In addition, because of the small size of the nets, the low abundance of organisms in the midwater layer could not be effectively sampled (Murray & Lee, 1909). Nevertheless, this voyage sparked many expeditions dedicated to plankton investigation.

In 1888, Carl Chun conducted the first comprehensive vertical distribution study within the midwater realm of the Mediterranean Sea and described the seasonal vertical zooplankton migration (Chun, 1888). Already in 1889, during the Plankton Expedition, Victor Hensen conducted the first quantitative account of the midwater realm using stratified vertical plankton net (known as the Hensen net) tows during the National voyage, targeting 26 depth strata to a maximum depth of approximately 3,400 m. The expedition concluded that the volume of zooplankton is generally low, it is evenly distributed and varies little in the composition. Hensen introduced the term ‘plankton’ to encompass all biota carried by ocean currents (Dolan, 2021).

The discovery of plankton in the Arctic Ocean dated back to the Nansen’s *Fram* expedition (1897–1899), over a century later than the exploration of the Southern Ocean. However, during the first 80 years, zooplankton data collection was sporadic and patchy, with most sampling performed either on ships frozen in ice (*Fram* or Russian ice-breaker *Sedov*) or on surface

drifting ice platforms, which were highly dependent on surface currents and atmospheric processes and usually only sampled surface waters (Kosobokova & Hirche, 2009).

The majority of samples collected during the 19<sup>th</sup> century were used to study systematics or biogeography. During the 20<sup>th</sup> century, overfishing of commercial fish stocks resulted in a sharp decrease in fish populations, raising interest in research focused on quantitative measurements to assess areas of high productivity and prey abundance for commercial fish (Vinogradov, 1968). Such circumstances stimulated a great interest related to the assessment of the mesopelagic fishes standing stock.

In 1930, William Beebe, Otis Barton and Captain John J. Butler built a 'bathysphere,' a steel sphere that could go as deep as 500 m where Beebe observed glimpses of known and unknown organisms and vast bioluminescence sparks (Berger, 2009; Beebe, 1934). He also noted the high mobility of organisms that escaped the vehicle's light. Thus, marking the first observations of pelagic waters to 500 m, observations of the organisms observed during the dive were published in the National Geographic (Berger, 2009). Shortly thereafter, the first observations of water > 4000 m were obtained during the 1948 Swedish Albatross Expedition. A single trawl was deployed north of the West Indies between 7,600-7,900 m depth (Brunn *et al.*, 1956).

In the early to mid 20<sup>th</sup> century, new instruments and technology were developed during World War II and the following years. Echo-sounding of icebergs led to the development of new methods for studying oceanic biota (Berger, 2009). The newly developed echo-sounding technology revealed shallow seafloors, although these locations were known to be deeper than instruments suggested. As a result, subsequently it was referred to as a 'false bottom'. In some cases, the sound was reflected from this layer, so the physicists called it a 'deep scattering layer' (Berger, 2009), which exhibited diel variation, indicating that organisms within the DSL underwent daily vertical migration to shallower waters. This technique became increasingly popular for researching deep pelagic fishes (Balls, 1948; Kampa & Boden, 1954). Although scientists had been aware of the existence of a scattering layer in the upper layers of the ocean for several decades, it was not until the development of advanced sonar technology that they could detect and study the scattering layer in deeper waters.

The survey ship *Galathea 2* carried out several scientific deep-sea explorations from 1950 to 1952, extending to depths of more than 10,000 m for the first time (Brunn *et al.*, 1956). Later, the International Indian Ocean Expedition (IIOE, 1959-1965) conducted surveys covering the entire Indian Ocean basin, during which the Indian Ocean Standard Net for sampling zooplankton was adopted, allowing the comparison of the collected data across different ships.

Early studies in the Southern Ocean focused primarily on taxonomy, life cycle, and vertical distributions. Significant advances on the zooplankton distribution and abundance started during the Discovery Investigations carried out in the early to mid-20<sup>th</sup> century (Ward *et al.*, 2014).

Discrete-deep-water trawls have revealed more detailed patterns of fish distribution and their DVMs (Berger, 2009). The midwater trawl designed by John D. Isaacs and Lewis W. Kidd at the Scripps Oceanographic Institute revealed a layered distribution of plankton below the epipelagic zone, capturing organisms that had not been previously encountered. The use of opening-closing nets with a finer mesh and the Bongo Net became routine in the 1960s, allowing for depth-stratified sampling of zooplankton to a depth of several hundred meters (Berger, 2009).

However, such trawls cannot provide reliable information on more fragile gelatinous organisms, such as siphonophores. These organisms scatter sound efficiently and produce a signal picked up by echo sounders (Proud *et al.*, 2018) but are usually damaged or destroyed in nets. Therefore, their importance was acknowledged only after direct observations from diving vehicles (Mapstone, 2015). Further development of technologies such as manned submersibles or remotely operated vehicles has enabled the collection of additional data on the fragile parts of communities, such as gelatinous micronekton (Berger, 2009).

In the 1970-the 1990s, research shifted from research institutes such as the Woods Hole Oceanographic Institute (WHOI, U.S.A.), Institute of Oceanographic Sciences (IOS, U.K.), and Institut Français de Recherche Pour L'Exploitation de la Mer (IFREMER, France) to include university-based investigations (Sutton, 2013). In the 1970s, the exploration of the deep ocean became more popularized within universities and the military, with mesopelagic fishes being the focus of most of the research work. Gjosaeter & Kawaguchi (1980)

summarized the available resources and distribution of mesopelagic fishes as a result of the spiked interest by the Food and Agriculture Organization (FAO, Rome) in their potential source of unexplored commercial resources. This work provided the first estimate of the global biomass of mesopelagic fishes. This work inspired further research in the mesopelagic zone.

The 1970s-1980s were associated with the appearance of mesopelagic commercial fisheries in the Gulf of Oman (myctophids), southern Iceland (sternoptychid), and myctophids fished by the USSR in the South Atlantic/ Southwest Indian Ocean, myctophid fisheries in South Africa (until the mid-1980), and some other fisheries that are not labelled mesopelagic (redfish, blue whiting, and cephalopods) (Pauly *et al.*, 2021). However, no long-lasting fisheries in the mesopelagic region were established at that time. Numerous cruises were conducted by the USSR between 1982 and 1989 to explore the abundance of mesopelagic fishes. They found that the concentration of fish is not dense enough for open fisheries in certain areas or seasons (Poletaev *et al.*, 1991).

In the Arctic, a new phase of scientific discoveries began in the early 1980s, with the development of modern research ice breakers (Kosobokova & Hirche, 2009). Such vessels allowed for a more consistent sampling design, providing the ability to study seasonal patterns of pelagic communities and their links to environmental parameters. As a result, the idea of the Arctic ocean being a 'monotonous biological desert' (Vinogradov & Melnikov, 1980) was rejected as a much higher estimate of pelagic community biomass was obtained (Kosobokova & Hirche, 2009; Ingvaldsen *et al.*, 2023).

In the 1990s, mesopelagic fishes gained additional interest as a result of global ecosystem studies, such as the Joint Global Ocean Flux Study (JGOFS; <http://usjgofs.whoi.edu>) and Global Ocean Ecosystem Dynamics (GLOBEC; [www.globec.org](http://www.globec.org)). The 1990s was also the beginning of global collaborative work such as the Census of Marine Life, with several projects related to mesopelagic zooplankton [e.g., MAR-ECO ([www.mar-eco.no](http://www.mar-eco.no)), Census of Marine Zooplankton (CMarZ; [www.cmarz.org](http://www.cmarz.org)), and Arctic Ocean Diversity (ArcOD; <http://www.coml.org/projects/arctic-ocean-diversity-arcod>)] (Sutton, 2013).

### *1.4.2 The current state of knowledge (21<sup>st</sup> century)*

Recently, several novel approaches have been developed to explore the life of the mesopelagic. The first promising approach uses DNA barcoding to analyze species diversity. DNA barcodes are short DNA sequences used for species identification (Hebert *et al.*, 2003). This method is currently being implemented in several global projects, including the CMarZ (Machida *et al.*, 2009). Another approach is aimed at designing new machines to be able to visually monitor the deep-water layers and life in them. The Woods Hole Oceanographic Institution (WHOI), along with the USA's National Aeronautics and Space Administration space agency (NASA) and Norway's Institute of Marine Research, are in the process of designing new robotic machines (such as *Deep See*, *Mesobot*, and *Snowclips*; <http://www.whoi.edu/main/auvs>) that will allow researchers to obtain more extensive data on mesopelagic organisms in addition to data obtained by sonar technology.

Recently, many research cruises have focused on observations in the mesopelagic zone. However, I discuss only several major global ocean expeditions in the 21<sup>st</sup> century. The summary of the published research from the expeditions and other smaller-scale research work in the mesopelagic realm is listed in Table A1.

Over the past two decades, a number of expeditions have been conducted to investigate the physical, chemical, and biological properties of the world's oceans. One of the earliest expeditions was Japan's Blue Earth Global Expedition (BEAGLE) that took place from 2003 to 2004. During the BEAGLE cruise, researchers measured and analyzed a variety of environmental variables, including water temperature, salinity, dissolved oxygen, and nutrients at approximately 500 stations in the upper mesopelagic zone of the Southern Hemisphere. However, no biological information was collected from these depths.

In contrast, Denmark's *Galathea III* expedition, which took place from August 2006 to April 2007, covered a broad range of topics, including the biology of small organisms, such as bacteria, plankton, and algae. Over 60 different projects were conducted onboard the ship, covering topics such as geology, culture and history, climate, and the environment. The expedition travelled from the North Atlantic to the west coast of Africa, across the Indian

Ocean to Australia and the Solomon Islands, then back to Copenhagen via New Zealand, Antarctica, South America, and the Caribbean.

The *Sorcerer III* Global Ocean Sampling (GOS) expedition, inspired by the British Challenger Expedition, focused on the diversity, genetics, and biochemistry of the microbial community in various environments, ranging from surface waters to deep-sea thermal vents, high saline ponds, and polar ice. The data collected during the expedition contributed to the ‘microbial Earth catalogue’ in 2007 and 2008.

A series of the Tara Oceans Expeditions, conducted from 2009 to 2013, observed various marine organisms, from viruses to fish larvae, and collected a wide variety of environmental data from major oceanic regions between the surface and 1000 m depths. The expedition resulted in the creation of a comprehensive catalogue of genetic material available for use by other scientists, with more than 35,000 different species whose genomic content was previously unknown.

The Malaspina 2010, the Spanish Circumnavigation Expedition, aimed to assess the diversity of life in the global ocean with a focus on deep regions. The expedition’s sampling covered the whole ecosystem, encompassing physical, biological, and chemical processes from the atmosphere overlaying the ocean to a depth of 4,000 m. The extensive multidisciplinary approach has to date resulted in more than 70 publications in the high-profile scientific literature.

Other ocean expeditions have been conducted with specific focus. For example, the International Indian Ocean Expedition 2 aimed to study a variety of environments spanning from coastal to deep-sea realms. The French MyctoO-3D-MAP project aimed to assess myctophids in relation to oceanographic conditions by combining modelling, acoustic and predator data (<http://www.seapodym.eu/project/mycto3dmap-myctophid/>).

Overall, the above-mentioned ocean expeditions have provided invaluable insights into the physical, chemical, and biological properties of oceans worldwide. Through these expeditions, researchers have been able to better understand the distribution and diversity of marine organisms as well as the physical and chemical processes that shape the world's oceans biota. These insights are critical for understanding the impacts of climate change and other



environmental stressors on the world's oceans, and for informing policies and management practices to protect marine ecosystems.

## 1.5 Gaps in knowledge

The growing interest in the mesopelagic zone as a source of natural resources and ecosystem services, coupled with increasing discussions in scientific and public domains, reflects diverse perspectives on its current and future significance in supporting human life. Schadeberg *et al.* (2023) analyzed the main theme occurring in 2,226 scientific abstracts and 4,066 tweets about the mesopelagic realm. Two main ideas were the most frequently discussed on twitter and in scientific papers: the exploitation of fish resources and the mesopelagic zone's capacity as a carbon sink. The understanding of the ecosystem in the mesopelagic zone is effectively shaped by scientific research among both the public and policymakers due to the lack of comprehensive governance policies.

The Web of Science's search for 'mesopelagic zone' yields limited results (666 papers), with the bulk of the studies emerging after 2012. Additional keywords were added to obtain better results ('vertical distribution,' 'mesopelagic fish,' 'mesopelagic community,' 'midwater zooplankton' etc.). A search of the Web of Science produced ~18,000 research articles (Figure 1.1).

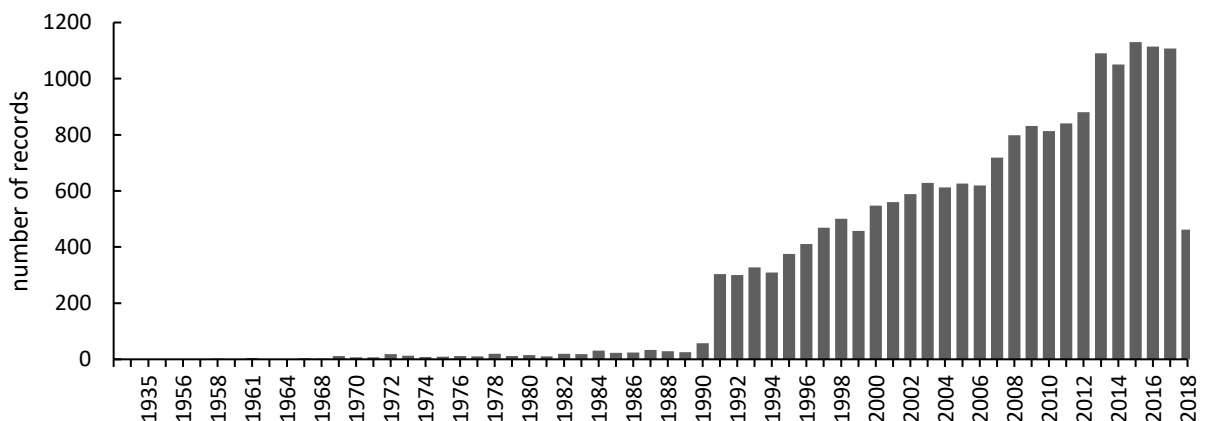


Figure 1.1: The number of scientific publications related to the topic of vertical distribution of mesopelagic organisms retrieved from Web of Science. Figure produced on June 8, 2018.

A review of ~400 published scientific research works related to mesopelagic mesozooplankton and micronekton communities with more than 170 research papers is synthesized in Table A1

(Appendix A) and represented in a table grouped by geographic area and organized in chronological order. However, research describing the life cycle or vertical distribution without any connection to abiotic or ecological significance is omitted from this synthesis as well as single-species/genera investigations. I primarily focus on the distribution of animals and their links to environmental conditions around the globe. In some instances, important findings on DSL or particulate organic carbon/dissolved organic carbon (POC/DOC) and other topics are also included in the table.

In summary, dozens of studies conducted on pelagic zooplankton have shed light on the community composition and vertical distribution, with a particular focus on the Mid-Atlantic Ridge (MAR) (Letessier *et al.*, 2011; Sutton *et al.*, 2008; Gaard *et al.*, 2008; Sweetman *et al.*, 2013; Vinogradov *et al.*, 2003). The unique community in this area is shaped by environmental factors like topography, proximity to the plume, and hydrothermal vents (Cook *et al.*, 2013; Vereshchaka & Vinogradov, 1999). Additionally, the global community structure of zooplankton is influenced by different water masses and resource partitioning, with gelatinous fauna playing a significant role (Youngbluth *et al.*, 2008; Vinogradov *et al.*, 1997).

Many studies have investigated DVMs as a prominent behavior of various mesopelagic organisms. The depths at which DVMs occur globally exhibit coherent large-scale patterns (Bianchi *et al.*, 2013). Firstly, migration depth is largely determined by seawater oxygen concentration. In areas where subsurface oxygen concentrations are high, migratory animals tend to descend to greater depths. Secondly, the presence of oxygen minimum zones (OMZs) also influences migration behavior. Animals undergoing DVM generally descend to depths close to the upper boundaries of low-oxygen waters associated with OMZs (Bianchi *et al.*, 2013). Certain species alter their DVM patterns in response to factors such as ontogeny, satiation, hunger, or external stimuli (Staby *et al.*, 2011). These dynamic behaviors further contribute to the complexity of the mesopelagic ecosystem. Preferred depth for migration was found to be related to the thermocline depth, while non-migrating species had a high tolerance to reduced oxygen concentrations and were found in oxygen minimum zones (OMZ) in high numbers (Peña *et al.*, 2014; Olivar *et al.*, 2017). OMZs, found in all oceans, are areas where dissolved oxygen concentration is very low, and their extent may expanded due to climate change (Cavan *et al.*, 2017). Certain biogeographical ecoregions are heavily influenced by

extreme OMZs (Stramma *et al.*, 2008; Stramma *et al.*, 2010). While OMZs can provide a refuge from predators, they also limit the vertical distribution of other fish species, resulting in low biomass of zooplankton and micronekton within OMZs (Koslow *et al.*, 2011).

Another well studied topic is related to understanding the mesopelagic community as a crucial component of the carbon cycle. Recent research estimated that respiration and carbon flux levels are higher than previously believed (Hernandez-Leon, 2004). The mesopelagic communities in the North Atlantic and Southern Oceans have been extensively studied, and the biomass of the DSL has been linked to primary productivity and surface environmental conditions (Proud *et al.*, 2017). At large scales, such as basin scales, the community structure appears to remain coherent, but this relationship breaks down at smaller scales, like oceanic fronts (Steinberg *et al.*, 2008). As a result, the composition of zooplankton can vary significantly at the meso-scale (approximately 180 km), with the number of taxa strongly correlated with surface productivity and depth (Andersen & Sardou, 1992; Cartes, 1998; Cartes *et al.*, 2013; Cartes *et al.*, 2009). Widely distributed zooplankton species have a broader tolerance ranges for key environmental variables and have a better adaption to a variable and changing conditions (Wiebe *et al.*, 2016).

Major gaps in our knowledge of the mesopelagic zooplankton and micronekton have also been identified. The lack of a global, unified, ecological (quantitative) database on mesopelagic biota, particularly zooplankton and micronekton, undoubtedly poses a challenge. The absence of comprehensive knowledge about the mesopelagic biota complicates geographical comparisons, which presents challenges in progressing the research and identifying specific areas that require immediate investigation. It is heavily impeded by the use of different sampling gears, employing custom designed sampling and often by a different focus on taxonomic groups, including identification resolutions. It is time to assess our mesopelagic sampling efforts and compile historical records. It is recognized that the mesopelagic realm acts as giant 'filter' attenuating the downward carbon flux yet supporting an unknown but conceivably large midwater community. Therefore, further investigation into a connection between the surface productivity and mesopelagic biomass, as well as the impact of the mesopelagic realm on global oceanic biochemistry is highly warranted. Furthermore, the

impact of climate change on the distribution and abundance of mesopelagic organisms also requires further investigation.

Addressing these knowledge gaps would require coordinated efforts from the scientific community, including standardizing sampling and data collection methods, developing new techniques for studying mesopelagic organisms, and increasing funding for research in this field. Ultimately, a better understanding of the mesopelagic community's role in the global ecosystem could help inform conservation and management efforts and contribute to our understanding of the impacts of climate change on the ocean.

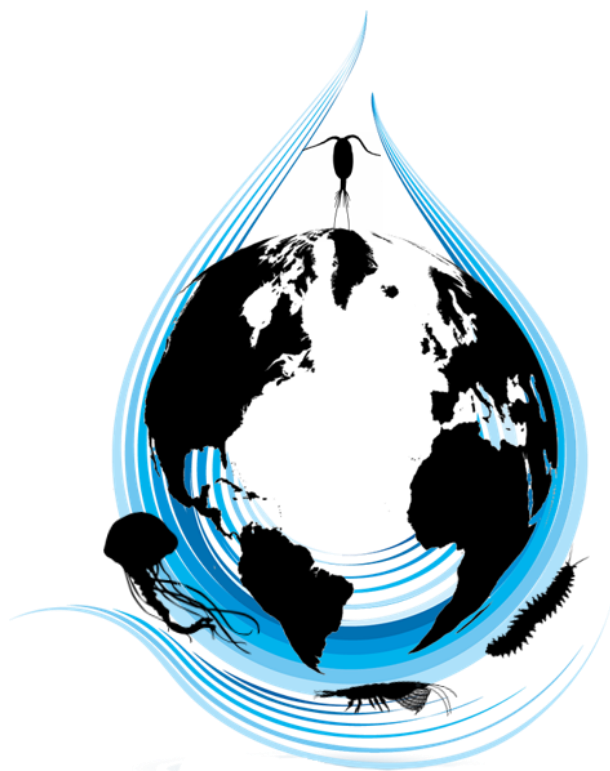
## **1.6 Project Objectives**

The main objective of this work was to develop a global baseline for the biogeography and diversity of the mesopelagic zone and to answer several fundamental macroecological and applied questions:

1. What is the current state of knowledge regarding the global distribution of mesopelagic mesozooplankton and micronekton community?
2. What is the global distribution and diversity of the mesopelagic mesozooplankton?
3. What is the global estimate of mesozooplankton biomass?

From the perspective of basic research, this project is envisaged to provide crucial insights into biological and biochemical processes within open-ocean ecosystems. In a changing world, with rising CO<sub>2</sub> levels, more acidic waters, and overexploitation of nearshore resources, an improved understanding of the ecological and biogeochemical interactions in the mesopelagic realm is required to predict the impact of climate change on ecological diversity and productivity of the ocean, which in turn affects global fish harvests. The major societal benefit of the proposed research will occur through an increased understanding of the diversity of the mesopelagic zone of the ocean and distribution of mesopelagic biomass throughout the ocean.

## Chapter 2: Mesopelagic Mesozooplankton and Micronekton Database



### 2.1 Introduction

Currently, the most complete database containing records of marine organisms is the Ocean Biogeographic Information System (OBIS; <http://www.iobis.org>), which contains more than 31.3 million records of more than 120,000 marine species collected throughout the water column (Fujioka *et al.*, 2012). However, less than 20% of OBIS records are from the mesopelagic zone, with more than 50% of the records coming from the epipelagic zone (Webb *et al.*, 2010). Due to the wide depth range of the mesopelagic zone, rigorous quantitative analyses of taxonomic and environmental data are problematic because records are spatially, horizontally and vertically, patchy, and/or are being collected in an inconsistent manner (Sutton *et al.*, 2017). Webb *et al.* (2010) demonstrated that the number of OBIS records decreases tenfold when at depths below 200 m. Within the mesopelagic zone, the number of records drops almost by another order of magnitude as the depths reach 1,000 m (Figure 2.1). It is clear that the mesopelagic zone lacks global consistent data and coverage (Proud *et al.*, 2017).

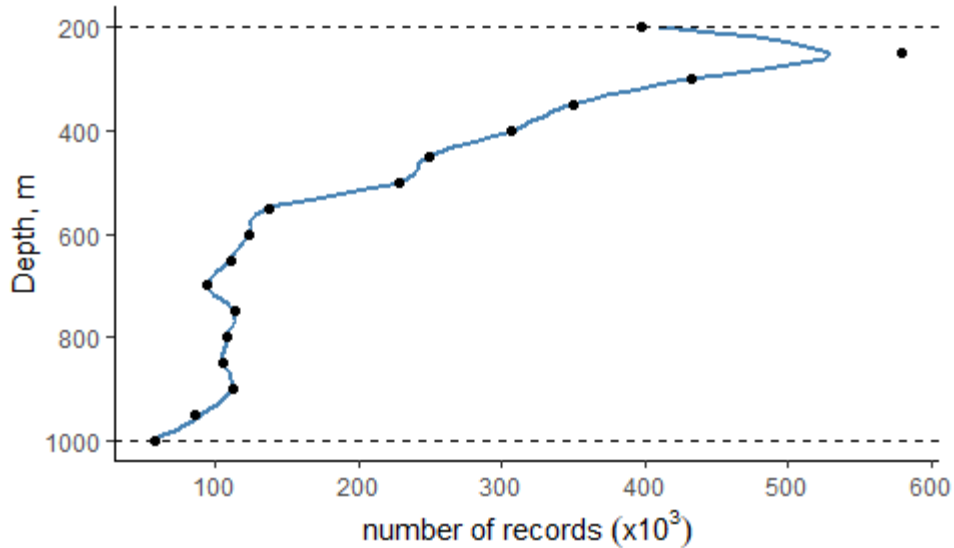


Figure 2.1: The depth distribution of OBIS records (kingdom animalia) in the mesopelagic zone (200-1,000 m). An average number of OBIS records (grouped by 50 m bins) against mesopelagic zone depth. The general trend is represented by lowess smooth (solid blue line). The dotted black lines represent the classical borders of the mesopelagic zone.

Webb *et al.* (2010) plotted the proportion of all OBIS records occurring within each of the five ocean regions (0-200 m, 200-1,000 m, 1-4 km, 4-6 km, >6 km) against the proportion of ocean area that region encompasses. The obtained point for the mesopelagic zone (200-1,000 m) lay above the 1:1 line, indicating that mesopelagic zone has proportionately more records than expected given their area (Webb *et al.*, 2010). However, this conclusion is misleading, because estimates are 2-dimensional and do not consider the 3-dimensional nature of the mesopelagic zone. If the horizontal distribution of available OBIS records is plotted against depth (Figure 2.2), it reveals an additional pattern of sampling bias, with almost no samples collected in the centers of the ocean basins, and this gap increases with depth. In addition, the sampling is skewed towards the North Atlantic area with small sampling efforts in the Indian, Arctic and Southeast Pacific Oceans. Indeed, such data gaps increase with depth.

Currently, mapping the global horizontal and vertical distributions of mesozooplankton and micronekton communities is challenging. Even the OBIS database, which contains information on the occurrence of certain species, lacks the full taxonomic classification of organisms, thus making it impossible, for instance, to filter the data based at the phylum level. In addition, the database shows the occurrence of species rather than biomass and abundance. Although the frequency of occurrence can be an indicator of the relative abundance of the species, it can be

biased by imbalanced sampling efforts, for example, if the species is targeted or seasonally more abundant.

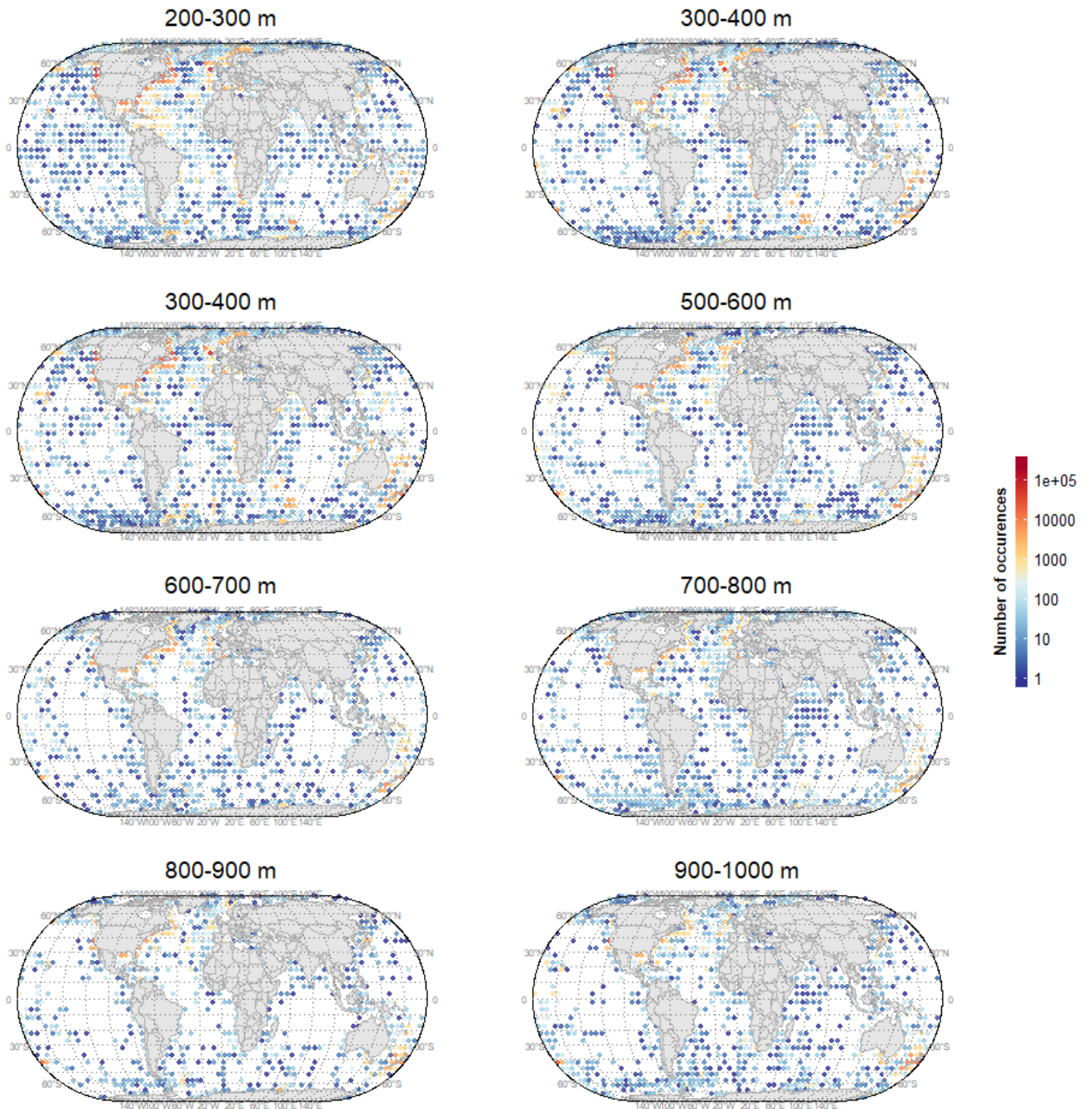


Figure 2.2: Available OBIS records (kingdom Animalia) from the mesopelagic zone (200–1000 m) separated into eight strata (100 m bins).

The COPEPOD database, on the other hand, contains the quantitative information of different species and communities of organisms, but has fewer records, especially in the mesopelagic

zone, and is incomplete, which will be demonstrated later in this chapter. The Marine Ecosystem Biomass Data (MAREDAT; [www.pangaea.de](http://www.pangaea.de)) is the first inventory of marine organisms based on counts of individual cells or organisms and contains many records of meso- and macro-zooplankton. However, data are scattered throughout the website and are hard to extract due to the different formats of each dataset with no option of filtering data based on the depth strata.

Acoustics provide very valuable information on mesopelagic fishes and other backscattering groups. However, the data are not species specific and only reported in the backscattering strength units (see MESOPP database; <http://www.mesopp.eu>), which is seldom converted into biomass or abundance units. The recently published BioTIME Database (Dornelas *et al.*, 2018), has many millions of records of different organisms, out of which only two categories are available in relation to the mesopelagic mesozooplankton and micronekton namely 'Fish' and 'Marine invertebrates.' When the database was sorted by these two categories, the search produced ~106 scientific studies or databases. While available datasets can offer estimates of biomass and abundance for various taxa, their lack of depth specificity complicates interpretation and comparison of these data. The majority of these estimates are indeed from the epipelagic zone.

To my current knowledge, no work has been done to attempt to synthesize the available global estimates of the mesozooplankton community to explore the current state of knowledge, or to perform global comparisons between geographical areas. Such syntheses were only attempted on total epipelagic mesozooplankton biomass and are in general outdated (Bogorov *et al.*, 1968; FAO, 1972). Micronekton, especially mesopelagic fishes, is a better studied group of organisms due to their potential high biomass and increased interest in potential harvesting for protein to replace depleted coastal fish stocks. Most of the work done on the mesopelagic is applicable to either ocean basins or even smaller features (seamounts, eddies and etc.; (Letessier *et al.*, 2017; Preciado *et al.*, 2017; Pusch *et al.*, 2004; Griffiths & Brandt, 1983) and are generally taxon-specific (Gibbons *et al.*, 1994; Vereshchaka, 1990; Cheney, 1985; Kruse *et al.*, 2010). Available global overviews of mesopelagic mesozooplankton and micronekton are listed in Table 2.1.



Table 2.1: Summary of available reviews/syntheses of mesopelagic organisms, aspects being studied, and region covered.

<b>Author</b>	<b>Organism studied</b>	<b>Aspect being investigated</b>	<b>Region</b>
Russell (1927)	Plankton plants and animals	Vertical distribution	Global
Banse (1964)	Holopelagic animals, mesopelagic larvae	Vertical distribution in relation to environmental factors	Global
Gjosaeter & Kawaguchi (1980)	Mesopelagic fishes	Distribution, ecology, and life history	Global
Kozlov (1995)	Mesopelagic fish (Myctophidae)	Trophic relationships	Southern Ocean
Hernandez-Leon & Ikeda (2004)	Mesozooplankton	Community respiration	Global
Brodeur & Yamamura (2005)	Mesopelagic micronekton (mesopelagic fishes, micronekton and cephalopods)	Biogeography	North Pacific
Kosobokova & Hirche (2009)	Zooplankton communities	The baseline estimates of major structural parameters and seasonal dynamics, abundance, and biomass	Arctic Ocean
McIvor (2011)	Mesozooplankton and micronekton	Distribution of biomass	9 Provinces
Atkinson & Ward (2012)	Zooplankton	Overview of the diversity and basic biology with an emphasis on abundance, distribution and feeding	Southern Ocean
Bednarsek et al. (2012)	Pteropods	Global distribution of carbon biomass, with a particular emphasis on temporal and spatial patterns	Global
Moriarty & O'Brien (2013)	Mesozooplankton	Distribution of biomass	Global
Hoving et al. (2014)	Deep-sea cephalopods	State of biogeographical knowledge	Global
Ward et al. (2014)	Mesozooplankton	Spatial and temporal patterns from Discovery Investigations	Southern Ocean
Vereshchaka et al. (2014)	mesopelagic shrimps (Sergestidae)	Global distribution	Global
Mapstone (2015)	Siphonophorae (Cnidaria: Hydrozoa)	History of discovery, species richness, a summary of worldwide distribution and some biological aspects	Global

Changing climate and rapid population growth with an increasing demand for protein lead to a decline in shelf and slope fish stocks. ‘New resources’ to unlock the economic potential of the global oceans can be found in the mesopelagic realm (St. John *et al.*, 2016) in the form of the large unexploited biomass of midwater fish, krill and cephalopods (Gjosaeter & Kawaguchi, 1980). Yet, knowledge of their abundance, biology and ecosystem role in the open-ocean system is inadequate to reliably assess the potential impacts of mesopelagic fish harvesting. The lack of such knowledge hinders the ability to assess the oceanic response and feedback in times of drastic climate changes and increased human activities (Sutton *et al.*, 2017).

The main goal of this chapter is to summarize the current state of knowledge on the global distribution of mesopelagic mesozooplankton and micronekton community through:

1. compiling all available data from literature (published and unpublished) and unpublished datasets on mesopelagic species distribution and abundance to create the first mesopelagic database;
2. systematizing and standardizing available information to account for differences in sampling depth, gear type used, day/night variations and seasonality of sampling;
3. building a unique, citable reference database for the compiled datasets.

## **2.2 Data & Methods**

### ***2.2.1 Mesopelagic Mesozooplankton and Micronekton Database Overview***

The Mesopelagic Mesozooplankton and Micronekton (MMM) Database was compiled using all available literature (published and unpublished) containing depth-stratified information on mesopelagic species distribution and abundance. Samples from epipelagic and/or bathypelagic are also included in the database if the sampling was done beyond the mesopelagic zone. Out of many entries collected for the mesopelagic, only data that had quantitative measurements of biomass and abundance were included in the database. Original columns of depth range and abundance/biomass estimate are preserved in the database. The database is in a single table with an accompanying table of column descriptions. The final database has 256,868 entries and 74 columns, collected from 262 different data sources. The coverage of the data, both

spatially and temporally, varies, but it extends from the year 1880 through to 2016. This database is a composite source of all the available data from various sources, so care is needed when using and interpreting these data due to the different sampling methods used. Although every effort was made to gather all available data, the database is likely incomplete as many data sets are not published. Therefore, it is envisaged that the database will encourage scientists around the world to submit their published, but not publicly available datasets.

### 2.2.2 Data sources

This work builds upon the research done by McIvor (2011), who compiled zooplankton distribution data from various sources to create vertical distribution profiles for organisms in the epi- and mesopelagic layers of nine pelagic ecological provinces. McIvor's database included 71 articles and 220 study locations and allowed for inter-area comparison during the summer period. This current work expands on McIvor's efforts by attempting to gather all available literature on mesopelagic zooplankton from various seasons.

A literature search was conducted using the keywords: 'mesopelagic,' 'vertical distribution,' 'mesopelagic zooplankton,' 'mesopelagic micronekton,' 'twilight zone,' and 'deepwater plankton' through Web of Science and Google Scholar search engines. Data were also augmented by searching the PANGAEA Data Publisher (<https://www.pangaea.de/>) using a query on oceans, and records with available geolocation and depths more than 200 meters for any available zooplankton and micronekton were included. This approach yielded 18 studies comprising 38,313 entries from PANGAEA. Additional sources included unpublished data from Dr. A. Yamaguchi on copepod data at various locations in the North Pacific, and Dr. Evgeny Pakhomov's unpublished data from the Southern Ocean.

### 2.2.3 Relationship to other datasets

The COPEPOD Database (<https://www.st.nmfs.noaa.gov/copepod/>) is a global plankton database that provides researchers with an integrated datasets of quality-reviewed plankton abundance, biomass, and composition data. Data were downloaded on January 7, 2019, including all subgroups, and contained 1,161,132 entries. To make comparisons easier, the data were then filtered based on depth (only entries between 200-1,000 m) and taxa (zooplankton only), and only quantitative records of biomass and abundance were included.

After applying these filters, the dataset contained 57,107 entries. However, the filtered dataset did not include any biomass or abundance estimates from the North Atlantic Ocean basin, despite significant sampling efforts in this area.

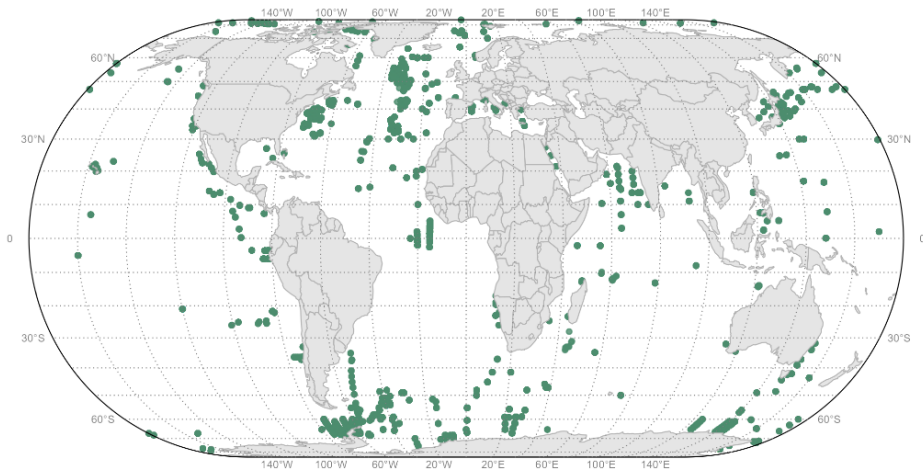
The BioTIME database (<http://biotime.st-andrews.ac.uk/>; downloaded on 2018/09/03) is a comprehensive collection of biodiversity time-series, comprising records of abundances and biomass for a range of marine and non-marine species. The marine realm alone contains approximately 5,664,196 records, represented by 152 studies across six categories (Benthos, Fish, Marine birds, Marine invertebrates, marine mammals, and Marine plants). However, extracting data based on depth or elevation is challenging, as no depth information was available in the downloaded file. To address this, a file with depth information was obtained through personal communication with the authors and linked with the main query using the STUDY\_ID as a common field. The resulting data were filtered for a depth range between 200 and 1,000 meters. However, some studies in the database were lacking depth information, resulting in a significant data reduction. In addition, only two taxa (fish and marine invertebrates) were applicable to the current study, resulting in only 6,720 entries from three studies. Fish records (n=719) were collected with 1.3 cm mesh and could not be classified as mesozooplankton or micronekton and were removed from the analysis, leaving 6,001 records for zooplankton. Unfortunately, information on mesh size or net type was not available for these studies.

The Jellyfish Initiative (JeDI) Database (<https://www.bco-dmo.org/dataset/526852>) is a comprehensive global database dedicated to gelatinous zooplankton (Cnidaria, Ctenophora and Thaliacea), aimed at defining a global baseline of gelatinous zooplankton populations (Condon *et al.*, 2012). The database consists of 476,000 quantitative, categorical, presence-absence, and presence-only records spanning three centuries (1790-2011), gathered from various published and unpublished sources. These records were collected globally, with the greatest concentration of data in the mid-latitudes of the Northern Hemisphere. After filtering for depth over 200 m and removing two density columns (density and density\_integrated) due to a lack of identified units, the data were further filtered for non-zero quantitative density of biomass values. This approach resulted in 3,284 entries remaining for analysis. Additionally,

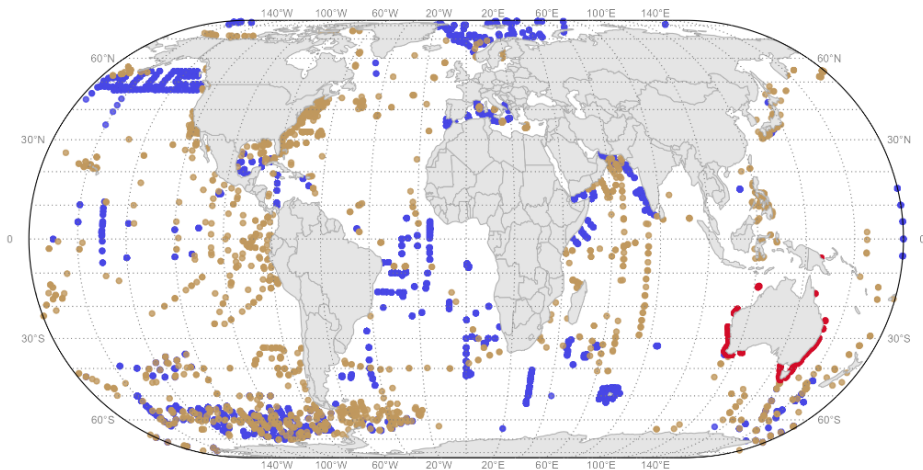
inconsistencies in units for mesh sizes reported in millimeters versus micrometers were corrected.

Data from existing databases were used along with the data extracted in this study and were combined together for further analysis and standardization procedure. Compiled data provided an extensive coverage of mesopelagic zone; however, lack sampling center of ocean basins (Figure 2.3B).

A) Compiled Data



B) Global Datasets



- BioTIME
- JeDI
- COPEPOD

Figure 2.3: Sampling locations of A) compiled dataset B) existing databases (BioTIME, COPEPOD, JeDI) for depth bin 200-1,000 m. All maps in Eckert IV global equal area projection.

## 2.2.4 *Database Structure*

### **MMM Database overview**

From each article, raw density data for zooplankton taxa were collated based on biomass and/or density at discrete depth ranges within the mesopelagic layers (200 – 1,000 m). Depth strata were sampled using a range of nets and bottles conducted between 1880 and 2016 (Table B1). In addition to density data, vertical distribution and geographical location, information related to the net mesh size (or sieve mesh size for bottle data), taxa, and seasonality was also recorded. The data were collected in an excel file. The database fields and information that were collected are provided in Appendix B.

### **Geospatial coverage**

The 2D-spatial distribution of available record is extensive, covering all oceans (Figure 2.4). The least sampling was done in the center of ocean basins and Indian sector of Southern Ocean.

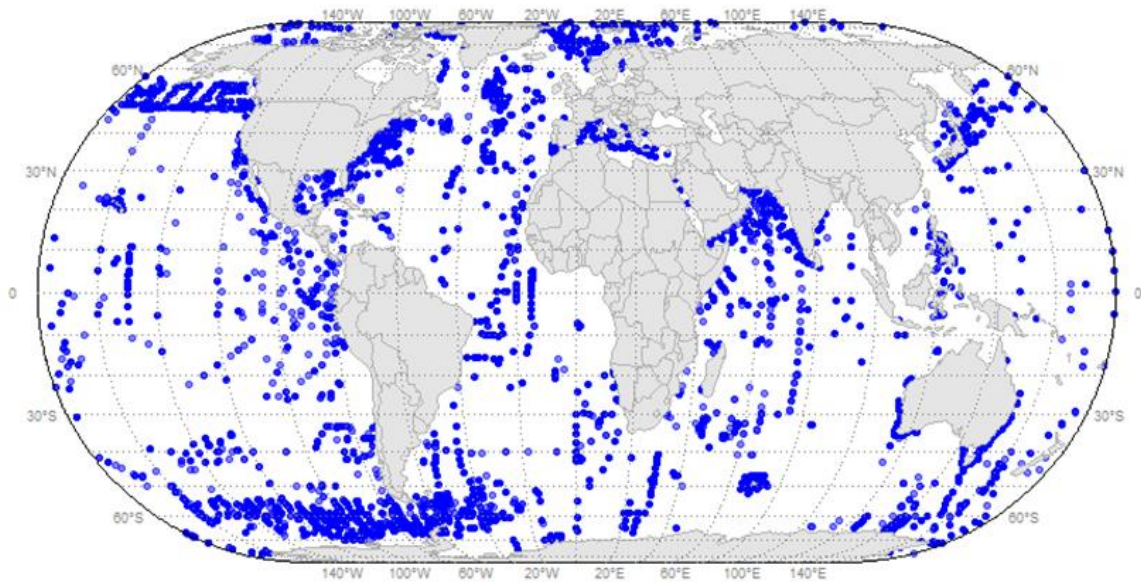


Figure 2.4: Global distribution of records in the Mesopelagic Mesozooplankton and Micronekton Database. All maps in Eckert IV global equal area projection.

Although the database shows extensive coverage, depth information is limited, and many records are missing depth information. Figure 2.5 shows distribution of maximum depth (filtered for depth > 200 m, excluding samples from epipelagic).

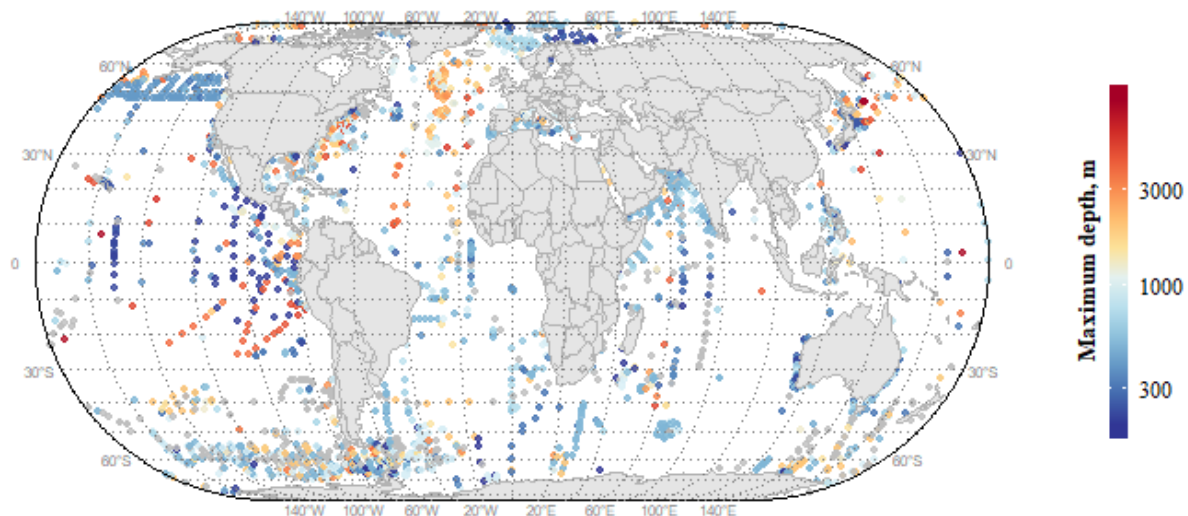


Figure 2.5: Spatial distribution of maximum depth of the records. Grey circles are points with no depth information or maximum depth < 200. Note the  $\log_{10}$  transformation of the color scale. All maps in Eckert IV global equal area projection.

### Taxonomic Resolution

The drawback of any existing database of species is the inability to filter the data given a specific taxonomic resolution. For instance, OBIS or COPEPODA databases have only one field that identifies the organism. However, data is usually entered at different taxonomic resolutions. Thus, searching the COPEPODA for Arthropoda will only return entries with the same taxonomic id, omitting species like *Acartia erythraea*, which belongs to this Phylum.

Additionally, depending on the date of collection/entry to the database, some taxonomic names can be entered as abbreviated (e.g., *S. gazellae*), have outdated names (e.g., *Phyllopus helgae* is unaccepted name for *Nullosetigera helgae*), or have been misspelled (e.g., Conchoeciinae instead of Conchoeciinae). To standardize the naming of different taxa, names were checked with World Register of Marine Species (WoRMS; WoRMS Editorial Board, 2022), an authoritative classification and catalogue of marine names. I used the *worms* package to match the species names to their accepted WoRMS name (Jan Holstein, 2018). The original names were retained, while the standardized values were put into *TAXA\_standartize* column. In addition, for each organism, Phylum, Class, Order, Family and Genus columns are provided to allow sorting at different taxonomic levels. The exceptions were paraphyletic terms such as Gammaridea or Natantia, where precise taxonomic classification was not available or when

several taxa were listed together (e.g., copepods, chaetognaths or Gelatinous zooplankton). In these cases, the rank ‘group’ was assigned to *TAXA\_standartize* and *rank* columns. In cases where the community of organisms were given (e.g., Mesozooplankton and Micronekton) it was assigned to rank ‘Zooplankton or Micronekton’. When organisms were recorded as unidentified, the rank and standardized name was assigned to ‘unidentified’ (see Table 2.2). five taxa were not found in WoRMS: Primnothorus, Cypris, Mesoniscus, *Conchoecia acuticosta*, *Bylgides pelagica* and *Pleuromamma abdonminalis abyssalis*. So, no standardized name or taxonomic resolution was included for these taxa. Records of terrestrial taxa (e.g., genus Mesoniscus) were removed from the database.

Table 2.2: Taxonomic resolution. Frequency of occurrence of different taxonomic units in the database. Rank Zooplankton or Micronekton is given for entries where density/biomass is given for the entire community of organisms. Group is assigned when species are reported in groups or when two different ranks are combined.

<b>Taxonomic rank</b>	<b>Number of entries</b>
Species	135821 (50.4%)
Genus	39127 (14.5%)
Order	25074 (9.3%)
Class	23813 (8.8%)
Zooplankton or Micronekton	13697 (5.1%)
Phylum	13493 (5.0%)
Family	8724 (3.2%)
group	3729 (1.4%)
Subphylum	1482 (0.6%)
Suborder	1133 (0.4%)
Subclass	846 (0.3%)
Subspecies	580 (0.2%)
Superclass	533 (0.2%)
unidentified	465 (0.2%)
Infraorder	329 (0.1%)
Superfamily	270 (0.1%)
Gigaclass	14 (<0.1%)
Infraclass	4 (<0.1%)
Infraphylum	4 (<0.1%)
Subfamily	3 (<0.1%)
Subkingdom	2 (<0.1%)
No rank	201 (<0.1%)

The majority of the data are available at low taxonomic resolution (50% of the data were recorded at species and 15% at genus levels). A large proportion of entries were also as a community entry (i.e., Zooplankton).



The MMM Database has 17 phyla (Figure 2.6, Table 2.3). Among these, the phylum Arthropoda accounts for the majority with 64% of the records. The distribution of other phyla is as follows: Cnidaria with 9.6%, Chordata with 9.4%, Chaetognatha with 8.9%, Mollusca with 4%, Annelida with 2.1%, Ctenophora with 0.6%, Echinodermata with 0.2%, and Nemertea with 0.1%. The remaining phyla collectively contribute to less than 1% of the total records. Additionally, there are eight phyla that have records available at a lower taxonomic resolution, such as species (Table 2.3). The number of species for each of the phyla is shown in the Figure 2.7.

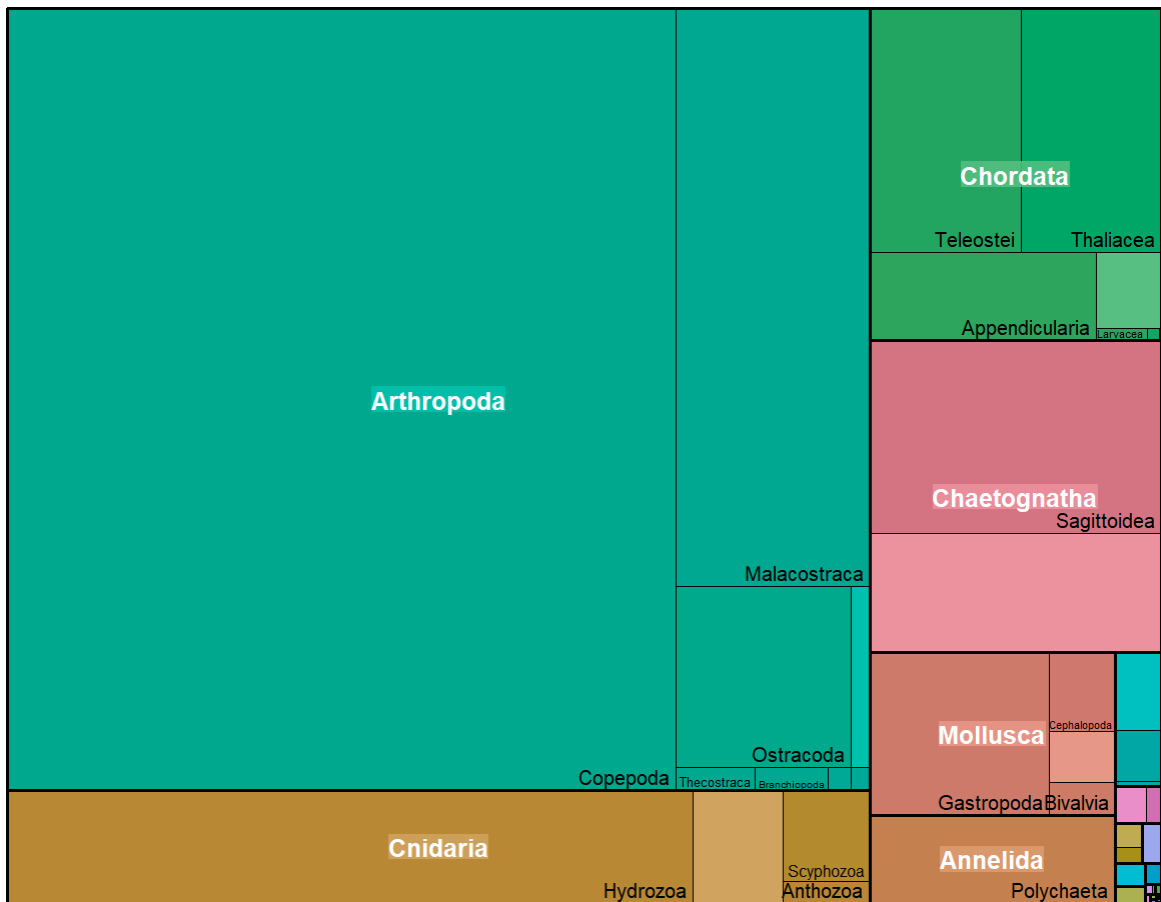


Figure 2.6: Treemap of available phyla and classes in the Mesopelagic Meso zooplankton and Micronekton Database. The area is proportional to the number of records in the database. Phylum names are shown in white, and class names are shown in black. Lighter shade of color corresponds to records recoded at the phylum-level.

Table 2.3: Summary of number of unique species, genera, families, and orders for each phylum-class recorded in the Mesopelagic Mesozooplankton and Micronekton Database.

<b>Phylum</b>	<b>Class</b>	<b>Number of orders</b>	<b>Number of families</b>	<b>Number of genera</b>	<b>Number of species</b>
<b>Annelida</b>	Polychaeta	2	7	11	12
<b>Arthropoda</b>	Arachnida	0	0	0	0
	Branchiopoda	4	4	6	4
	Copepoda	5	54	146	680
	Hexapoda	0	0	0	0
	Malacostraca	10	83	161	308
	Ostracoda	2	2	14	16
	Pycnogonida	0	0	0	0
	Thecostraca	1	2	0	0
<b>Chaetognatha</b>	Sagittoidea	2	5	14	37
<b>Chordata</b>	Appendicularia	1	2	3	9
	Ascidiacea	0	0	0	0
	Larvacea	0	0	0	0
	Leptocardii	0	1	1	0
	Teleostei	23	66	152	258
	Thaliacea	3	4	18	29
<b>Cnidaria</b>	Anthozoa	1	0	0	0
	Cubozoa	1	1	1	1
	Hydrozoa	7	47	119	197
	Scyphozoa	3	16	24	36
<b>Ctenophora</b>	Nuda	1	1	1	4
	Tentaculata	2	10	10	10
<b>Echinodermata</b>	Asteroidea	0	0	0	0
	Holothuroidea	0	0	0	0
<b>Hemichordata</b>	Enteropneusta	0	0	0	0
<b>Mollusca</b>	Bivalvia	1	1	1	0
	Cephalopoda	5	18	25	28
	Gastropoda	6	15	14	21
<b>Nematoda</b>	Chromadorea	1	1	1	0
<b>Nemertea</b>	Hoploneurtea	1	3	3	2
<b>Platyhelminthes</b>	Turbellaria	0	0	0	0
<b>Porifera</b>	Demospongiae	1	1	0	0

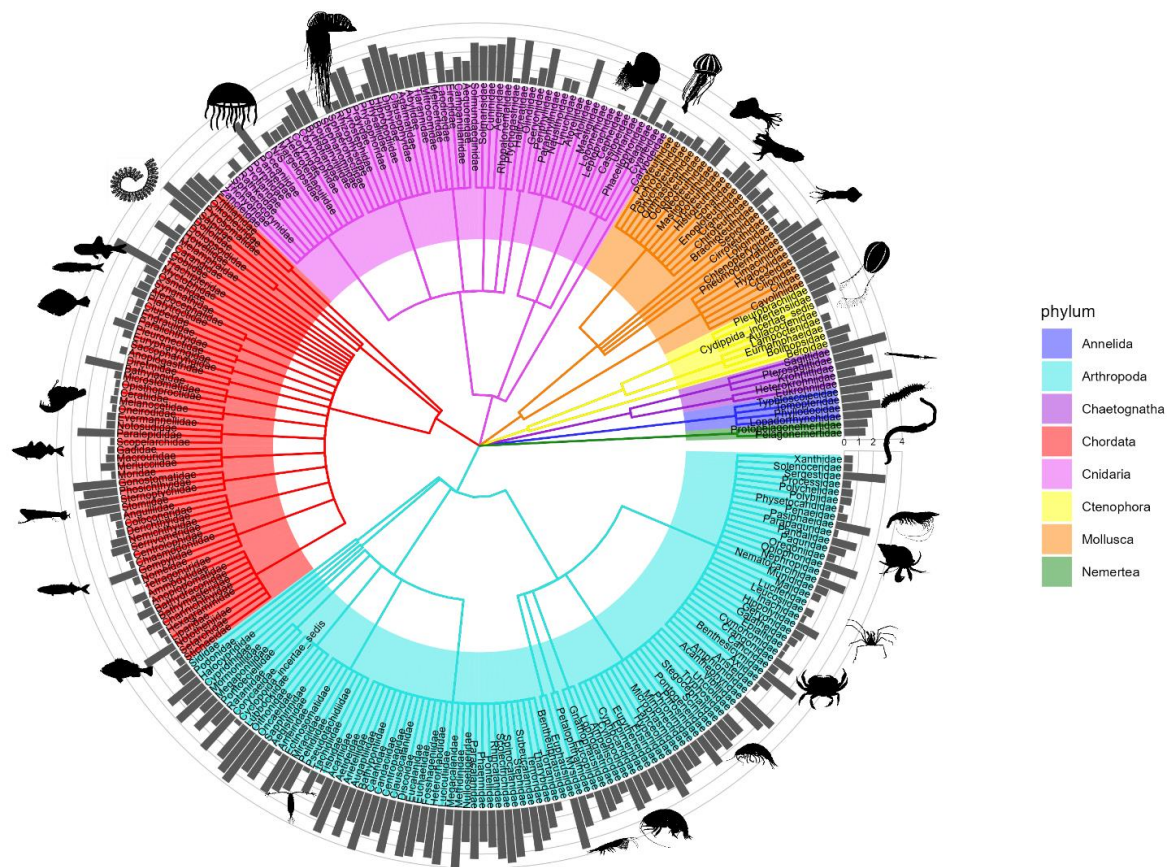


Figure 2.7: Phylogeny of life contained within the Mesopelagic Mesozooplankton and Micronekton Database for all taxa recorded at species levels. Here, the diversity of organisms is based on their taxonomic classification (phylum, class, order, family). Gray bars reflect the  $\log_{10}$ -transformed number of species recorded in each family. N/S in the diagram is for unidentified family for *Rhabdoon reesi*.

### Sampling Methods

Fifty-five different sampling methods were recorded in the MMM Database (Table 2.4). Juday net was the most common sampling gear in the database (21% of entries) followed by MOCNESS and RMT8 (16 and 7% of entries, respectively).

Table 2.4: Top 10 sampling methods in the Mesopelagic Mesozooplankton and Micronekton Database.

<b>Sampling Method</b>	<b>Number of entries (% of entries)</b>
Juday net	57,561 (21.4%)
MOCNESS (Multiple Opening and Closing Net System)	43,211 (16.0%)
RMT8 (rectangular midwater trawl system)	17,907 (6.6%)
Plankton net	15,948 (5.9%)
Bogorov – Ross net	12,262 (4.6%)
Multiple sampling techniques	9,855 (3.7%)
submersible observations	10,045 (3.7%)
MPN - HYDROBIOS (Multiple Plankton Net)	9,235 (3.4%)
DzhOM (Oceanic modification of the Juday) net	8,392 (3.1%)
BIONESS (Bedford Institute of Oceanography Net and Environmental Sampling System)	7,425 (2.8%)
Other catch methods	77,446 (28.8%)
No net information	58 (<0.1%)

Samples were collected by three tow types: vertical (44% of entries), oblique (26% of entries), horizontal (5% of entries) or combinations of the three (<1%). No tow information was recorded for 32% of entries. Samples were collected with variety of mesh sizes ranging from 30  $\mu\text{m}$  to 2.2 cm (Figure 2.8).

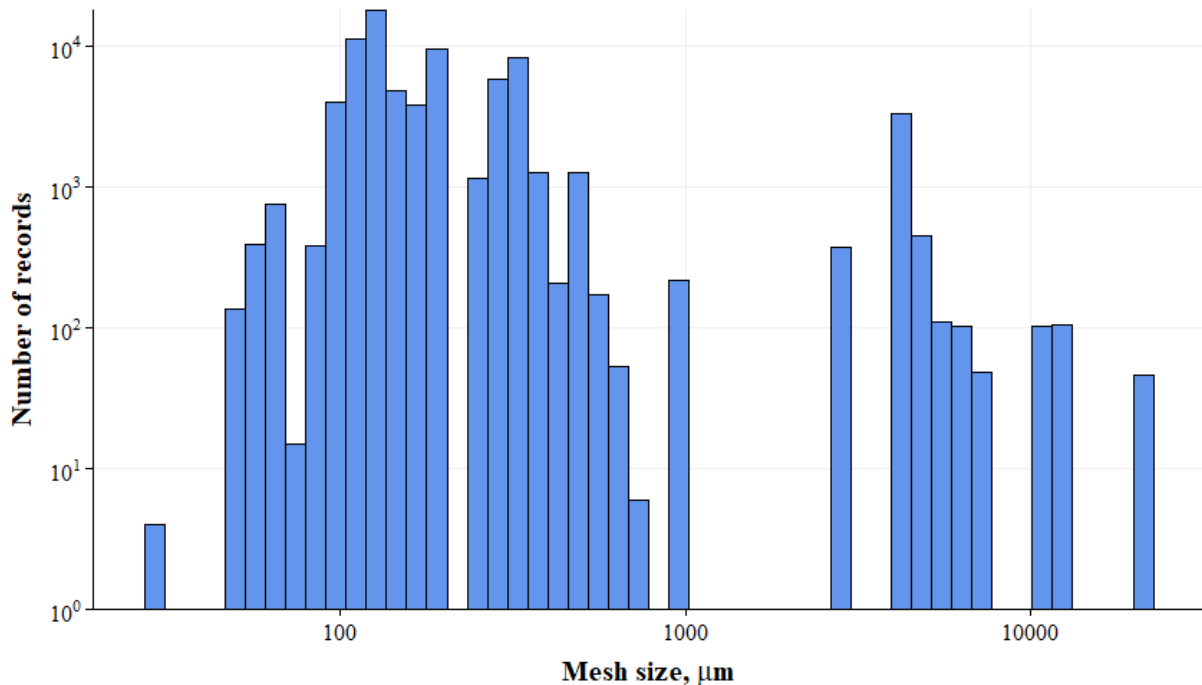


Figure 2.8: Distribution of mesh sizes in the Mesopelagic Mesozooplankton and Micronekton Database.

## Day & Night Observations

In total, 64% of entries had no record of the sampling time. Only 15 and 17% of records could be identified as day- and night-time collections, respectively (Figure 2.9). Finally, 3.4% of records contained pulled abundance/biomass estimates from both day and night.

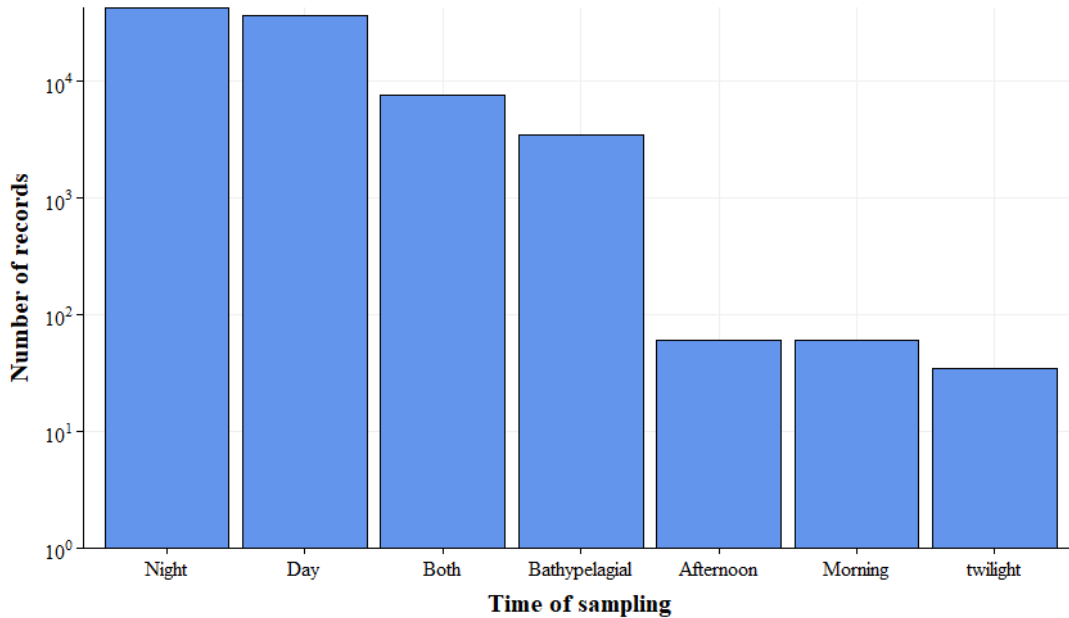


Figure 2.9: Distribution of records in the Mesopelagic Meso zooplankton and Micronekton Database based on time of sampling. Note that Bathypelagial was entered as time for sampling in the data obtained from MARine Ecosystem DATA (MAREDAT) initiative database.

## Abundance and biomass data

The most abundant quantitative data were present in terms of density ( $\text{ind}\cdot\text{m}^{-3}$ ) and areal abundance ( $\text{ind}\cdot\text{m}^{-2}$ ) which compromised 43% and 23%, respectively (Table 2.5).

Table 2.5: Frequency of occurrence of data recorded in different abundance/ biomass units in the mesopelagic database. Abbreviations: ind. - individuals; Wt. - weight; # - number; C-carbon.

<b>Biomass/Abundance Units</b>	<b>Number of entries (% of all entries)</b>
Abundance ( $\text{ind}\cdot\text{m}^{-3}$ )	117,026 (43.3%)
Abundance ( $\text{ind}\cdot\text{m}^{-2}$ )	62,927 (23.4%)
Catch rate ( $\text{ind}\cdot\text{haul}\cdot\text{h}^{-1}$ )	38,738 (14%)
Abundance (# ind.)	13,786 (5.0%)
Biomass ( $\text{g}\cdot\text{m}^{-3}$ )	5,453 (2.0%)

<b>Biomass/Abundance Units</b>	<b>Number of entries (% of all entries)</b>
Biomass (g wet Wt. ·m <sup>-3</sup> )	4,682 (1.7%)
% Abundance of total # ind.	4,622 (1.7%)
Biomass (mg C ·m <sup>-3</sup> )	4,078 (1.5%)
Biomass (g dry Wt. ·m <sup>-3</sup> )	3,773 (1.4%)
Biovolume (ml·m <sup>-3</sup> )	3,503 (1.3%)
Biomass composition (%)	2,409 (0.9%)
Biomass (mg C ·m <sup>-2</sup> )	1,854 (0.7%)
Total number of species (#)	1,805 (0.7%)
Biomass (g dry Wt.·m <sup>-2</sup> )	1,673 (0.6%)
Biomass (g wet Wt.·m <sup>-2</sup> )	1,124 (0.4%)
Catch rate (ind.·hour <sup>-1</sup> )	674 (0.2%)
Catch rate (kg·h <sup>-1</sup> )	564 (0.2%)
Biomass (µg dry Wt.·ind. <sup>-1</sup> )	240 (0.1%)
Biomass (g·m <sup>-2</sup> )	204 (0.1%)
Displacement volume (ml·m <sup>-2</sup> )	109 (<0.1%)
Biomass (g dry Wt.·ind. <sup>-1</sup> )	51 (<0.1%)
Biomass (µg C·ind. <sup>-1</sup> )	51 (<0.1%)
Catch rate (ml·haul <sup>-1</sup> )	14 (<0.1%)
Biovolume (ind. ·ml <sup>-1</sup> )	3 (<0.1%)
No Biomass/abundance value	6969 (2.5%)

## 2.3 Database Standardization

Although a large volume of quantitative information is available on the distribution of different mesopelagic mesozooplankton and micronekton species, this quantitative information is hardly comparable between the regions as the data are collected at different times of year, time of day, or by different sampling gears. All these factors, and many others such as net mouth area and type of organism sampled, complicate the comparison between the areas. This work attempts to standardize these data with adjusted density values.

The subset of the MMM Database used in this chapter consists of 112,000 entries (42.3% of all records; Table 2.5) for densities (the most common type of quantitative information in the MMM Database). Among these entries, 28,120 have zero densities, representing a substantial proportion of the database (25% of all densities). However, the presence of such a high number of zeros can potentially introduce bias into the linear model. To address this issue and ensure linearity in the relationships, a transformation technique was employed.

Given the multiplicative behavior of densities, logarithmic transformations are typically applied to linearize these relationships. Thus, when analyzing densities, a constant (e.g., 0.001) is added to each density value prior to performing the logarithmic calculation. However, the inclusion of this constant may pose challenges during the calibration.

In the calibration process, it is essential to avoid biasing the lower end of the density estimates by adding a constant. Since the focus of calibration is not on zero densities, these entries can be simply dropped from the database. By eliminating the zero entries, it becomes possible to calculate the logarithm without the need for a constant addition, ensuring that the calibration is not influenced by the added bias. By employing this methodology, the database is transformed and prepared for analysis, effectively addressing the issue of biased estimates caused by a high proportion of zero densities and allowing for accurate calibration without the need for constant addition.

Before proceeding, several columns needed standardization to reduce the number of groups (discussed in sections below).

### *2.3.1 Net size standardization*

The MMM Database has around 7,858 unique identifiable tows. Based on the distribution of available mesh sizes, the new classification can be made as a distinction between mesozooplankton (Meso) and micronekton (Macro), with a cutoff of 1,000  $\mu\text{m}$  (Figure 2.10). I used mesh size to classify the organisms into these groups as only 23% of records provide size range. Thus, my definition of mesozooplankton and micronekton differs from classically used one of 20 mm (as mentioned in Chapter 1.3).

The mesozooplankton mesh sizes were categorized into six net classes, namely '50', '150', '225', '300', '400', and '750', corresponding to mesh sizes  $\leq 100 \mu\text{m}$ , between 100 and 200  $\mu\text{m}$ , between 200 and 250  $\mu\text{m}$ , between 250 and 350  $\mu\text{m}$ , between 350 and 500  $\mu\text{m}$ , and between 500 and 1,000  $\mu\text{m}$ , respectively. The micronekton was categorized into three net classes based on their mesh sizes: '3000' for mesh sizes between 1000 and 3000  $\mu\text{m}$ , '4750' for mesh sizes between 4,500 and 5,000  $\mu\text{m}$ , and 'Big' for mesh sizes larger than 5,000  $\mu\text{m}$  (Table 2.6). Micronekton data were less frequent in the database and more scattered (Figure 2.10). Due to the very small proportion of micronekton, the standardization was only performed on

mesozooplankton groups with known net class ( $n = 73,825$ ). Mesozooplankton was standardized to 225 net class as this net class was the most frequent in a database.

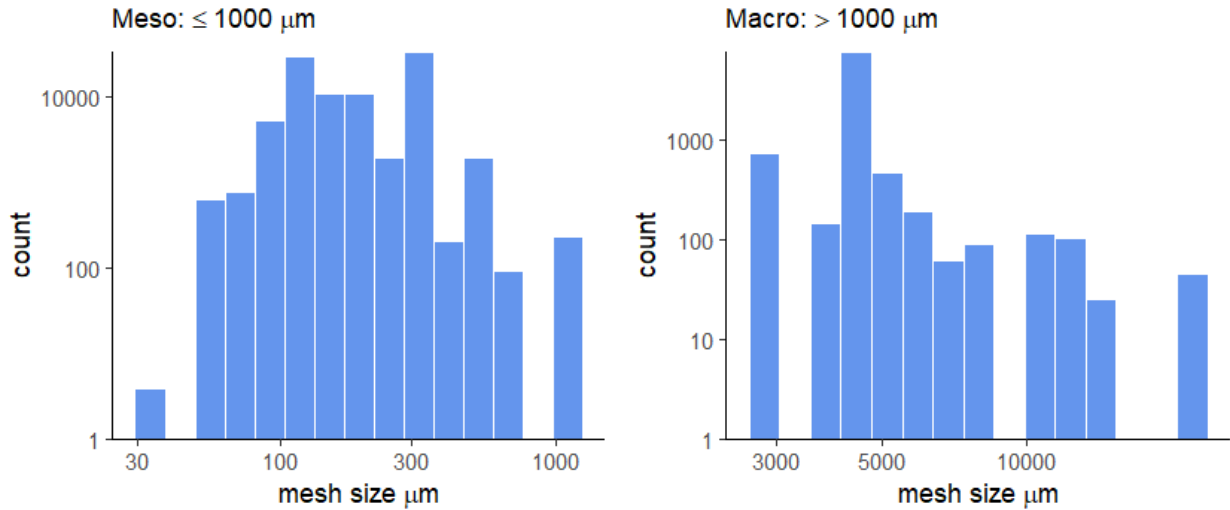


Figure 2.10: Histograms of mesh sizes (in  $\mu\text{m}$ ) that are present in the Mesopelagic Mesozooplankton and Micronekton Database. Note the  $\log_{10}$  transformation on both axes. Mesh sizes are separated into two net classes: mesozooplankton (Meso  $\leq 1,000 \mu\text{m}$ ) and micronekton (Macro, mesh  $> 1,000 \mu\text{m}$ ).

Table 2.6: Distribution of various net classes in the Mesopelagic Mesozooplankton and Micronekton Database.

<b>Class</b>	<b>Net Class, <math>\mu\text{m}</math></b>	<b>Mesh sizes, <math>\mu\text{m}</math></b>	<b>Number of records (%)</b>
<b>Meso</b> <b>(<math>\leq 1000 \mu\text{m}</math>)</b>	50	$\leq 100$	6549 (5.8%)
	150	100-200	40502 (35.9%)
	225 (reference group)	200-250	11407 (10.1%)
	300	250-350	33212 (29.5%)
	400	350-500	1660 (1.5%)
	750	500-1000	758 (0.7%)
<b>Macro</b> <b>(&gt;1000 <math>\mu\text{m}</math>)</b>	3000	1000-3000	897 (0.8%)
	4750	4500-5000	8236 (7.3%)
	Big	$\geq 5000$	639 (0.6%)
<b>No mesh size</b>	-	-	8838 (7.8%)

### 2.3.2 *Taxonomic standardization*

The phyla Porifera, Platyhelminthes, Phoronida, Nemertea, Nematoda, Hemichordata, Brachiopoda, and Bryozoa did not have enough data and were excluded from the calibration. The phyla Ctenophora and Echinodermata had very low counts, so were also left at the phylum level (Table B2).



Remaining phyla (Annelida, Chaetognatha, Mollusca, Arthropoda, Cnidaria and Chordata) were split further into taxonomic classes (Table B2).

- Phylum Annelida, except for four records, entirely consisted of a single class Polychaeta (Table B2). The majority of entries (85%) in Class Polychaeta did not specify the taxonomic order of the organisms, so this taxon was left at a class level (Table B3).
- Phylum Mollusca had low counts in general (Table B2), with 84% of the records from class Gastropoda, followed by Cephalopoda (2.3%), and Bivalvia (1.7%) and was modelled at these class-levels.

Remaining phyla (Chaetognatha, Arthropoda, Cnidaria and Chordata) were then further investigated at the order level (Table B2).

- Phylum Chaetognatha had a large number of records listed at a phylum-level (~38%; Table B2) but the rest belonged to class Sagittoidea (Table B2). Sagittoidea consisted of records of two orders: Phragmophora and APhragmophora (Table B2) and was modelled at these two order-levels.
- Phylum Arthropoda included seven classes (Table B2). Three classes, Copepoda, Malacostraca and Ostracoda, comprised 99% of all phylum records and were used in the calibration routine (Table B2). Class Ostracoda did not have enough records to resolve the taxa into a lower taxonomic resolution and was left off at a class level (Table B2). Class Malacostraca had nine orders represented in the database with 96% records belonging to Orders Euphausiacea, Amphipoda, Decapoda, Isopoda and Mysida (Table B3). Thus, the class Malacostraca was calibrated for all five orders. Furthermore, 94% of the class Copepoda was represented by four orders: Calanoida, Cyclopoida, Harpacticoida and Mormonilloida that all were used in the calibration exercise (Table B3).
- Phylum Cnidaria was represented by three classes: Anthozoa, Hydrozoa, and Scyphozoa, while 17% records remained at the phylum level (Table B2). Classes of Anthozoa and Schyphosoa were excluded from the analysis as those taxa did not have

enough records (Table B2). Class Hydrozoa primarily was represented by orders Siphonophorae (77%), Trachymedusae (8%) and Narcomedusae (3%) (Table B3). All three orders were chosen for calibration.

- Phylum Chordata was represented by six classes, with 90% of the records belonging to the classes Teleostei, Appendicularia and Thaliacea that were subsequently chosen for calibration (Table B2).

To summarize, the following 24 taxa were chosen for calibration: Ctenophora, Polychaeta, Ostracoda, Calanoida, Cyclopoida, Harpacticoida, Mormonilloida, Euphausiacea, Amphipoda, Decapoda, Isopoda, Mysida, Pragamphora, Aphragmophora, Actinipterygii, Appendicularia, Thaliacea, Siphonophorae, Trachymedusae, Narcomedusae, Coronotae, Bivalvia, Cephalopoda and Gastropoda. These taxa accounted for 90% of all records with non-zero mesozooplankton densities. The remaining less numerous groups (micronekton or other mesozooplankton groups) were not adjusted through calibration.

### *2.3.3 Space and Season*

In an effort to reduce spatio-temporal variability, I stratified locations both latitudinally and temporally by season, ensuring comprehensive sampling across all mesh sizes within these divisions. Latitudinal zones were defined as latitudinal bands: Sub-Antarctic (>50°S), Southern temperate (23.4-50°S), Tropical (23.4°S-23.4°N), Northern Temperate (23.4-50°N), Sub-Arctic (>50°N). Seasons were specified as follows: Summer was defined as December-February in the Southern Hemisphere and July-August in the Northern Hemisphere. Winter months were chosen as July-August in the Southern Hemisphere and December-February in the Northern Hemisphere. Fall was defined as March-May in the Southern Hemisphere and September-November in the Northern Hemisphere. Lastly, Spring was identified as September-November in the Southern Hemisphere and March-May in the Northern Hemisphere. The records that were located on 0° latitude were assigned to Seasons of the Southern Hemisphere. A limitation of this standardization is its failure to consider the varying expressions of seasons based on latitude.

### 2.3.4 Calibration steps

To ensure the calibration of as many entries as possible, my approach involved a sequential process due to incomplete information regarding mesh size, tow type, and season. I standardized all entries based on mesh size, followed by standardization of entries with tow information according to tow type. Lastly, entries with seasonal information were standardized based on the season.

First, I calibrated mesh size using latitudinal bins to account for and statistically remove the spatial variability. Mesozooplankton densities were standardized to the expected catch by 225  $\mu\text{m}$  net class (200-300  $\mu\text{m}$  mesh). The majority of nets used for mesozooplankton collections are equipped with mesh sizes that can generally be pooled into six net classes: 50, 150, 225, 300, 400 and 750  $\mu\text{m}$  (Table 2.6). The calibration estimate was derived from fitting a linear regression model:

$$\log_{10}(\text{density}) = \beta_0 + \beta_1 \text{Net Class} + \beta_2 \text{Latitudinal Band}$$

Where the 225  $\mu\text{m}$  net class formed the reference category. Subtracting the regression coefficient from the log-transformed densities of a sample ‘corrects’ the mean to the value expected in a 200-300  $\mu\text{m}$  net. Graphical examination of the raw and corrected values indicated that this effectively calibrated counts for a range of taxa (see panel B in Figure 2.11 and Figures B2-B24). Some taxa (Narcomedusae, Cephalopoda and Bivalvia) were not distributed widely enough to enable spatial strata to be included in the model, so these were calibrated using the simpler model of  $\log_{10}(\text{density}) = \beta_0 + \beta_1 \text{Net Class}$ .

The second stage of the calibration was to correct for tow-type. Correction for tow type was carried out by fitting the model:

$$\log_{10}(\text{density mesh corrected}) = \beta_0 + \beta_1 \text{tow type}$$

Where the reference level for tow type was the oblique haul. As before, subtracting the regression coefficient from the transformed data provided the calibration. An additional calibration term that was considered was net area, however graphical examination of the sample distributions indicated any correction effect was minimal, so the simpler model was used. Net-tow calibrated densities can be seen in panel C in Figure 2.11 and Figures B2-B24.

The last stage of calibration was to adjust the densities from step 3.

$$\log_{10}(\text{densities mesh-tow corrected}) = \beta_0 + \beta_1 \text{Season}$$

Where the reference level for season was summer. As before, subtracting the regression coefficient from the transformed data provided the calibration. Net-tow-season calibrated densities can be seen in panel D in Figure 2.11 and Figures B2-B24.

To illustrate the calibration process, a case of order Euphausiacea is presented (Figure 2.11). Euphausiacea was chosen as an example of taxa with lots of records in the MMM Database. In addition, this taxon was caught with different mesh sizes, tows and during different seasons. Distribution of raw densities (log-transformed) was variable. However, after standardization, the distribution of log-transformed densities looks more symmetrical and centered around the same mean (Figure 2.11).

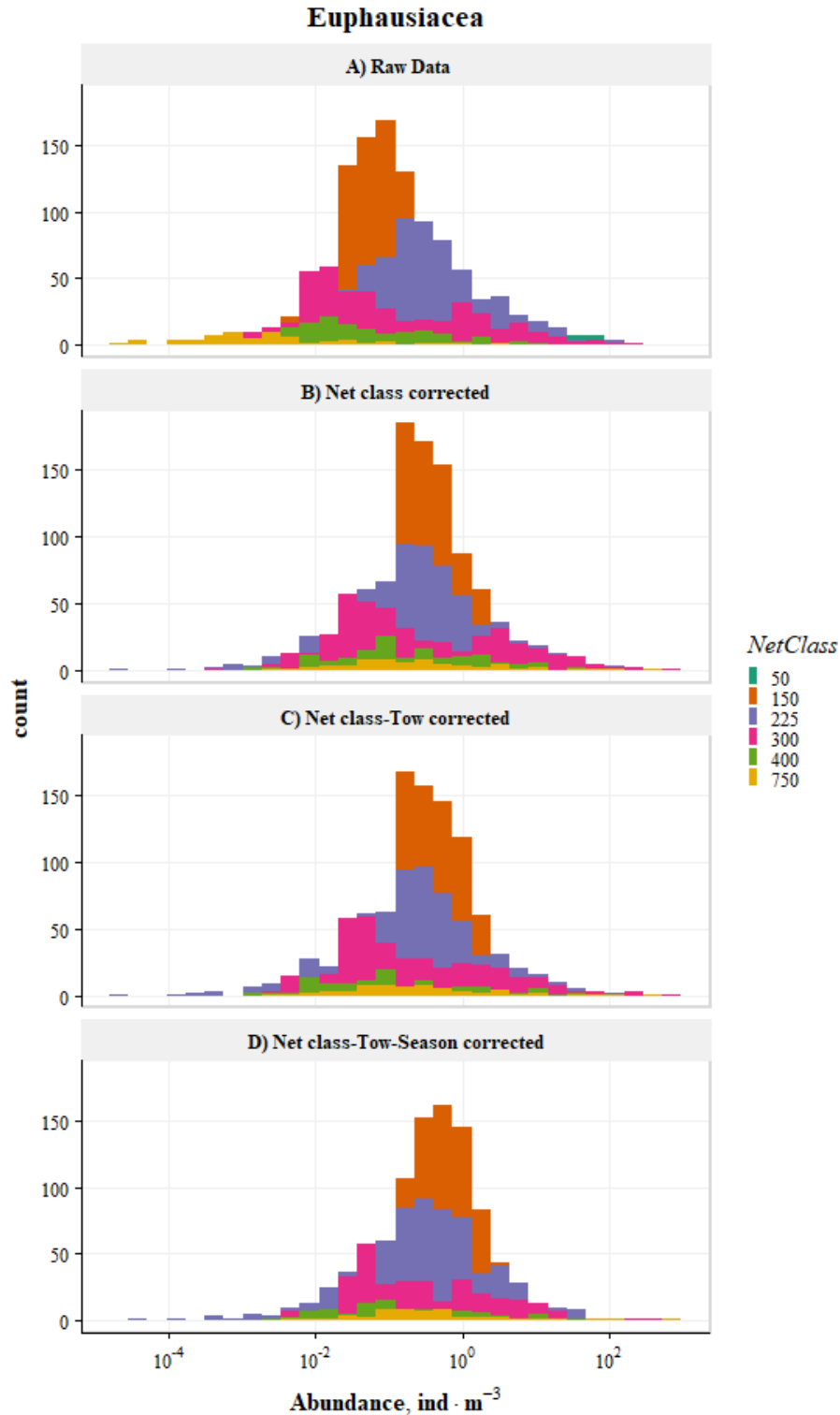


Figure 2.11: Example of calibration procedure for order Euphausiacea colored by different mesh size (NetClass of 50, 150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) densities from step B corrected for the tow type; D) densities from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

## **2.4 Case Study: A Comparison of Statistical Methods for Modelling Zooplankton Density**

The accuracy of population estimates can be significantly impacted by the variability in zooplankton sampling methods, seasons, day/night measurements, and years. Varying levels of selectivity can be caused by different sampling methods, resulting in certain species or life stages of zooplankton being collected more frequently than others. Skewed data and inaccurate population estimates can be caused by this.

One important aspect of variability that is often overlooked is the impact of different sampling depths. Zooplankton populations can be stratified based on depth, with certain species or life stages found at different depths. This stratification can induce substantial discrepancies in sampling rates, contingent upon the depth at which the sampling gear is deployed. However, this variability is rarely accounted for in zooplankton assessments, which can lead to inaccurate estimates of population size and density.

There are several statistical methods that can be used to address various sources of variability in zooplankton sampling. One commonly used approach is Generalized Linear Models (GLM), which can account for the effects of different sampling methods, seasons, and environmental factors on zooplankton density. Linear Regression Models and GLMs use regression analysis to explore the effects of one or more explanatory variables on a response variable of interest. In particular, ecologists often use these models to infer the effects of an organism's environment on said organism's distribution or density, or predict distribution and density from environmental variables (Oppel *et al.*, 2012).

Ecologists typically use methods such as the log-transformation of a skewed response variable, even when the response variable of interest is noticeably non-normal (Boldina & Beninger, 2016). In this case study, available data on mesozooplankton densities and related environmental variables, such as density and salinity, have been employed to conduct a comparative analysis of statistical modelling methodologies. The analysis focused on the densities of Calanoid species (Order Calanoida) in the Southern Ocean, with data collected in the year 1987. Two methodological questions were formulated and answered to provide a critical reflection on common practices by ecologists in statistical inference, and

recommendations were made based on the results. The primary objective was to evaluate whether a linear model with a log-transformed response variable (referred as Gaussian) demonstrates similar performance to GLMs. The secondary objective was to determine the effect of incorporating a dispersion parameter on model parameters and/or model performance. For a comprehensive understanding of the research, including detailed information about data preparation, model formulation, and the results, please refer to the Appendix C. This section will provide a short overview of the study.

In the comparative investigation between Gaussian model and GLMs, we consider three GLMs with the following error distributions: Log-normal, Gamma, and Inverse Gaussian. These distributions were chosen as a suitable approach for modelling continuous, positive, and skewed responses. The full model formulation can thus be written as follows:

$$y_i = \beta_0 + \beta_1 \cdot \text{temperature}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{temperature}_i \cdot \text{depth}_i + \epsilon_i \quad (1)$$

Here,  $y_i$  denotes the density (the response variable),  $\beta_0$  represents the intercept,  $\beta_1$  and  $\beta_2$  are coefficients corresponding to temperature and depth, respectively, and  $\beta_3$  is the coefficient for the interaction between these two variables. The term  $\epsilon_i$  represents the error term for each  $i^{\text{th}}$  observation.

Generalized Akaike Information Criteria (GAIC) was used for model comparisons (Symonds & Moussalli, 2011). To ensure that the GAIC of the linear model with the transformed response was comparable to the rest of GLM models, the transformed log-likelihood multiplied by the Jacobian was used (Akaike, 1978), and the GAIC was re-calculated manually for Gaussian model. The Inverse Gaussian distribution had the lowest GAIC, making them a more appropriate option compared to a transformation (Table C2). The Gaussian model and the Log-Normal GLM produced the same coefficients as expected (Table C2).

The estimates of density in the dataset have different precisions based on the depth range (i.e., bin width) sampled. A dispersion model was used to evaluate the effect of the dispersion parameter on model performance. To assess the impact of adding a dispersion parameter on model parameters, the best performing model was used (Inverse Gaussian) and the same model was run with the dispersion parameter. The coefficients of these models are presented in (Table C3). Interestingly, we see that while the interaction term between depth and temperature was

significant in the non-dispersion model, this was not the case when the dispersion parameter is included. In comparison to the dispersion model, we see that overall, when the dispersion is not modelled (i.e., estimated as a constant), lower p-values are estimated for the covariates. Such biases in p-values are a demonstration for why explicitly modelling the dispersion parameter is essential for the database: by accounting for the differential precisions of the density estimates, calculated estimates of coefficients and statistical significance are likely more reliable.

In summary, it was demonstrated by this case study that a better fit was provided by the use of a GLM with an Inverse Gaussian error distribution, compared to a traditional log-transformation for this dataset. Therefore, it is recommended to use a GLM framework (over standard linear models) to deal with non-normal/non-linear data distributions. The importance of explicitly modelling the dispersion parameter was also shown, indicating that failure to account for the different levels of precision of the estimated densities can result in potentially falsely identified significant variables.

## **2.5 Concluding remarks**

The first Mesopelagic Mesozooplankton and Micronekton (MMM) Database was compiled using a comprehensive literature search, including both published and unpublished sources to provide valuable information on the distribution and density of mesopelagic species. The data sources were diverse, ranging from research articles and databases to unpublished data from individual researchers.

The database includes quantitative measurements of biomass and density for mesopelagic species. The study builds upon previous research by McIvor (2011), who compiled zooplankton distribution data for the epi- and mesopelagic layers. The current work expands on McIvor's efforts by including data from various seasons and attempting to compile all freely available literature on the mesopelagic zooplankton. This work demonstrates the effort to create a more comprehensive database for the mesopelagic zone.

The database consists of 256,868 entries collected from 282 different data sources, spanning the years 1880 to 2016. However, the spatial and temporal coverage is variable, which should



be considered when using and interpreting the data. Additionally, different sampling methods were used in the original studies, which may introduce some variability in the database.

The MMM database was compared to other existing databases, such as the COPEPOD, the BioTIME, and the Jellyfish Database Initiative (JeDI). These comparisons highlighted the unique contribution of the new database and its potential for further analysis and standardization. Despite their usefulness, existing databases had limitations related to a lack of depth information in some records and at times restricted information for certain taxa or depth ranges. Nevertheless, while every effort was made to gather as comprehensive a database as possible, the MMM Database did not include all information currently available (especially for any data that became available after 2018 as data collection was stopped at that year). Therefore, it is envisaged that this database will serve as a starting platform to continue gathering mesopelagic zooplankton and micronekton datasets.

The database includes entries at various taxonomic levels, with the majority at the species and/or genus levels. However, there are cases of outdated or misspelled names, which were standardized using the WoRMS database. The availability of the taxonomic information allows for sorting and analysis at different taxonomic levels, facilitating further exploration of the database.

The geospatial coverage of the database is extensive, covering all oceans. However, the distribution of sampling points revealed poorly sampled areas, particularly in the ocean basin centers and in the Indian sector of the Southern Ocean. The lack of depth information for some records is also highlighted, indicating a limitation in the database that should be considered during analyses.

The database includes information on nets used for sampling, day and night observations, and quantitative measures of density and biomass. The results provide insights into the frequency of different density and biomass units present in the database, with areal density and volumetric density being the most common units.

For example, the spatial and temporal plankton density in published sources did not have the methodological uniformity and had to be standardized. Density values were chosen as these units are the most abundant in the MMM Database, comprising 42.3% of all records.

Mesozooplankton nets were equipped with mesh size ranging from 50 to 1,000  $\mu\text{m}$  and sampled water column with different profiles. Therefore, to make quantitative data comparable the best approach was to calibrate data based on mesh size, tow type and season. The compiled database was run through a number of translation and standardization routines to convert the values to common biomass units, unified mesh size, net area, and the taxonomic resolution. Linear regression models were applied as an initial technique as these models allow for easier statistical inference to examine how different measurement conditions are associated with organism densities.

To conclude, I explored the efficacy of various statistical methods to model the relationship of mesozooplankton density and several environmental covariates using a case study of five Calanoida species from Southern Ocean. The log-transformed response variable in a standard linear regression model performed worst among the tested methodologies and thus, though log-transformation of a right-skewed response variable is commonplace in ecology, it is likely not the best option for modelling this type of data distribution. Rather, the use of a Generalized Linear Model with an Inverse Gaussian error distribution provided a better model fit. However, additional work is needed to validate these results on different subsets of data. To account for variance inaccuracy of density measurements (bin width), a dispersion model was added to the best performing model. The Inverse Gaussian GLM with a dispersion model performed better than the model with only variable dispersion. Additionally, the inclusion of a dispersion model removed potentially false significance of some covariates.

# Chapter 3: Global Distribution and Diversity of Mesopelagic Mesozooplankton

## 3.1 Introduction

Covering over two-thirds of the Earth's surface, the global ocean is a vast and complex ecosystem, with a multitude of physical, chemical, and biological processes that influence its functioning and sustainability (Proud *et al.*, 2017; St. John *et al.*, 2016). The mesopelagic zone encompasses 20% of the global ocean volume and extends on average 200 to 1,000 meters below the ocean's surface. However, the actual depth range varies between ocean basins and latitudes (Reygondeau *et al.*, 2018). This area is also known as a twilight zone, layered between a sun-lit epipelagic zone and a bathypelagic zone, which receives no sunlight. However, the light reaching the mesopelagic is not sufficient for net photosynthesis to occur.

The mesopelagic food web is complex connecting the deep and the surface ocean. Mesozooplankton (organisms in the size range 0.2-20 mm) and micronekton (2-20 cm) communities contribute to CO<sub>2</sub> removal from the atmosphere (Guidi *et al.*, 2016) through a process named the 'biological pump'. It transfers organic carbon between the surface and deeper layers via two pathways: (a) sinking of particulate organic matter (passive carbon transport) and (b) zooplankton and micronekton diel vertical migrations after feeding near the surface at nighttime (active carbon transport) (Vinogradov, 1968; Tseitlin, 1986; Koppelman & Frost, 2008). Roughly 90% of the passively sinking organic matter is re-mineralized in the water column (Koppelman & Frost, 2008), reducing the amount of food available for deep-sea communities. Species from many taxonomic groups (e.g., copepods, euphausiids, cnidarians, fish or squid) perform extensive diel vertical migration (DVM), which is known as the largest movement of biomass on Earth (Hays, 2003). The active carbon transport thus has an important ecological role in linking surface and deep-sea (Ramirez-Llodra *et al.*, 2010) and is a crucial participant in the biochemical cycling (Klevjer *et al.*, 2012; Robinson *et al.*, 2010). Although this carbon flux is widely acknowledged, the extent of the process is still poorly understood due to insufficient sampling efforts in the mesopelagic realm (Robinson *et al.*, 2010; Kwong & Pakhomov, 2017; Steinberg & Landry, 2017).

Quantifying zooplankton vertical distributions, together with their biogeography, is fundamental in understanding pelagic food web structure and functioning (McIvor, 2011; Robison, 2009) and predicting vertical fluxes of organic matter to the deep ocean (Stemmann *et al.*, 2008a). Currently, the uncertainty in the active carbon transport contribution to the total downward particle flux is high, ranging from 5 to 90% (Steinberg & Landry, 2017). Despite plankton/micronekton ecological significance and considerable biomass, remarkably little is known about global mesopelagic diversity, composition and distribution (St. John *et al.*, 2016; Béhagle *et al.*, 2015). Unlike terrestrial ecosystems, where the distribution of organisms is usually two-dimensional, the distribution of life in the ocean is three-dimensional. In addition to the limitations of collecting patchy zooplankton in 3D space, sampling the deep ocean is logistically challenging, labor and cost-intensive (Berger, 2009). These constraints impede rigorous global quantitative analyses of taxonomic and environmental data (Sutton *et al.*, 2017). Webb *et al.* (2010) demonstrated that the number of OBIS records of marine organisms decreases tenfold when the depth reaches 200 m. Within the mesopelagic zone, the number of records further drops by another order of magnitude as the depth reaches 1,000 m. As a result, the twilight zone lacks consistent global data and hence a macro-scale and macroecological understanding (Proud *et al.*, 2017). The spatial correlation between community composition and geographical location is urgently needed to fill in the missing gaps in knowledge on the horizontal and vertical distribution of mesopelagic mesozooplankton and micronekton.

Species Distribution Models (SDM) or environmental niche models have been widely used to analyze species occurrence records and environmental data to predict their distributions across geographic space and time (Elith & Leathwick, 2009). These models can be used to understand how environmental conditions influence the distribution of species and are employed for predictive purposes (ecological forecasting; Elith & Leathwick, 2009). These approaches have been adapted for marine epipelagic species (Reygondeau & Beaugrand, 2011; Righetti *et al.*, 2019; Rombouts *et al.*, 2009; Benedetti *et al.*, 2021). However, no effort has been made to investigate the global diversity of mesozooplankton in the mesopelagic ocean.

This chapter aims to predict the global distribution of mesozooplankton species in the mesopelagic realm using species distribution models. In this study, I used a newly compiled MMM Database (see Chapter 2) to extract the list of all known extant species occurring in the

mesopelagic. A global species distribution model (SDM) for the mesozooplankton community was then created. The contemporary distribution of known mesopelagic mesozooplankton (MM) species and patterns of geographic rarity and species diversity are inferred. The most relevant environmental predictors of MM distributions are examined. Furthermore, the relative importance of different SDM processes in determining model performance is evaluated.

## 3.2 Methods

### 3.2.1 *Biotic Data*

I used records from the MMM Database to obtain a list of mesozooplankton and micronekton taxa occurring in the mesopelagic zone. All statistical analysis, data manipulations and visualizations were performed using the R Statistical Software (R Core Team, 2022).

For the analysis, I obtained a total of 1499 species from the MMM Database with their accepted taxonomic resolution from the World Register of Marine Species (WoRMS; WoRMS Editorial Board, 2022). To increase the number of available observations and to account for the dynamic nature of the taxonomic classification, a list of synonyms was compiled for each species through the WoRMS database. However, to avoid having the same synonym for different species, only unique synonyms were kept. Additionally, various traits of species habitat were collected from FishBase ([www.fishbase.org](http://www.fishbase.org); Froese & Pauly, 2022) and SeaLifeBase ([www.sealifebase.org](http://www.sealifebase.org); Palomares & Pauly, 2022) using the *rfishbase* (Boettiger *et al.*, 2012) package. Since the available MMM Database did not contain a comprehensive representation of mesopelagic fishes (only 221 species were found in the database), I focused only on the mesozooplankton community (1278 species).

For most mesozooplankton species, SeaLifeBase did not contain reliable information on depth range (53% of zooplankton species lack information on their depth range) and their association to epipelagic or mesopelagic habitat (34% missing habitat classifications). More details on missing data can be found in Figure D1. In addition, many planktonic species caught in the mesopelagic zone were classified as benthic. For some organisms, it can be explained by different habitats occupied by different life stages (i.e., juveniles are planktonic while their adult form transforms into a benthic organism). Since SeaLifeBase predominantly provided information on adult species, I applied an empiric classification to determine the habitat of the

zooplankton species. First, I updated the species' minimum and maximum depth range based on the depth range found in the MMM Database. Second, I classified each species based on whether they could be found in the mesopelagic zone. In many studies, the mesopelagic zone is defined as an ocean area between 200 and 1,000 m depths. However, Reygondeau *et al.* (2018) revealed that the vertical division of the zone is not constant over the global ocean but varies between ocean basins and latitude.

Thus, I used the dynamic depth range of mesopelagic provinces from Reygondeau *et al.* (2018) for each location to update the information on the species' habitat. Species were classified into four groups: EPI, MESO, BOTH and NEITHER. Species were classified as epipelagic ('EPI'; n=96) if they were found in the epipelagic depth range only. Species were classified as mesopelagic ('MESO'; n=314) if species' occurrences were found only in the mesopelagic zone and species were classified as both ('BOTH'; n=681) if they occurred in both epi- and mesopelagic zones. Species whose occurrences were only found below the mesopelagic zone were classified as 'NEITHER' (n=187). I focused on species that were part of the mesopelagic community, so species that were classified as NEITHER and EPI were excluded from the analysis.

### 3.2.2 Occurrence Records

Species distribution records were collected from the Ocean Biogeographic Information System (OBIS; <http://iobis.org>, accessed June 2022) using the *robis* (Provoost & Bosch, 2021) package. Occurrence records for each accepted species name were combined with occurrence records from their synonym names. Additional occurrence data were added from the MMM Database. Occurrences records with a coordinate of '0°, 0°' generally indicated missing information. Also, records found on land were considered as error. These records were excluded from the analysis. To eliminate duplicated records from different data sources and limit the number of records from repeated sampling events in the same area, occurrence data were gridded on Map of Life (MOL; <https://mol.org/>) equal-area grid with a per cell area of 3091 km<sup>2</sup>. Based on the occurrence data, the native range of each species was identified using the mesopelagic province developed by Sutton *et al.* (2017). If a province contained more than 1% of all the records, they were classified as part of the species native range.

### 3.2.3 *Environmental Data*

Nine environmental variables that are available globally were used to model species bioclimatic envelopes: water temperature (Temp), salinity (Sal), Mixed Layer Depth (MLD), dissolved oxygen concentration (O<sub>2</sub>), nitrate concentration (NO<sub>3</sub>), phosphate concentration (PO<sub>4</sub>), silicate concentration (SiO<sub>2</sub>), net primary production (NPP) and euphotic zone depth (Zeu). Apart from MLD and Zeu, data were obtained from the World Ocean Atlas 2018. Zeu estimates were taken from NASA MODIS database, and NPP was taken from GDFL ES2M model (Table 3.1). All variables were interpolated to an equal-area grid for further analysis. Three sets of environmental conditions were calculated: EPI (mean environmental condition in each cell in the epipelagic zone following Reygondeau *et al.* (2018), MESO (mean environmental condition in each cell in the mesopelagic zone) and BOTH (mean environmental condition in each cell for both epi-and mesopelagic zone). Since the goal of the study was to predict the MM distribution, only environmental conditions for species found in MESO and BOTH were used for further analysis (n=861). The distribution of environmental parameters is shown in Figures D17 and 18. NPP was removed from NPPEN models due to a problem of covariance in the matrix. To minimize the influence of coastal environmental dynamics on the open ocean, I excluded the data from shallow areas (<200 m) by using the Sutton *et al.* (2017) Mesopelagic Boundary shapefile to mask areas that are outside of the shapefile boundary.

Table 3.1: Available databases of the environmental parameters used in this study.

<b>Environmental variables</b>	<b>Unit</b>	<b>Data Source</b>	<b>Time period</b>	<b>Spatial resolution</b>
Ocean Temperature (Temp)	°C	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=t">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=t</a>	Long-term annual mean (1981-2010)	1°
Salinity (Sal)	psu	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=s">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=s</a>	Long-term annual mean (1981-2010)	1°

<b>Environmental variables</b>	<b>Unit</b>	<b>Data Source</b>	<b>Time period</b>	<b>Spatial resolution</b>
Mixed Layer Depth (MLD)	m	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=M">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18.pl?parameter=M</a>	Long-term annual mean (1981-2010)	1°
Dissolved Oxygen concentration (O <sub>2</sub> )	μmol·kg <sup>-1</sup>	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=o">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=o</a>	annual mean	1°
Nitrate concentration (NO <sub>3</sub> )	μmol·kg <sup>-1</sup>	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=n">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=n</a>	annual mean	1°
Silicate concentration (SiO <sub>2</sub> )	μmol·kg <sup>-1</sup>	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=i">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=i</a>	annual mean	1°
Phosphate concentration (PO <sub>4</sub> )	μmol·kg <sup>-1</sup>	World Ocean Atlas 2018 <a href="https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=p">https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/bin/woa18oxnu.pl?parameter=p</a>	annual mean	1°
Net Primary Production (NPP)	PgC·m <sup>-3</sup>	GDFL ESM2	Long-term annual mean (1996-2015)	9 km
Euphotic Depth (Z <sub>eu</sub> )	m	Aqua/MODIS Level-3 Mapped Euphotic Depth Data Version 2018 <a href="https://hermes.acri.fr/index.php?class=archive">https://hermes.acri.fr/index.php?class=archive</a>	Long-term annual mean (1997-2022)	4 km

I identified collinearity among the environmental parameters (Figure D2) using Pearson's correlation with  $|r| > 0.7$  as a threshold level for collinearity (Dormann *et al.*, 2013). For both MESO and BOTH sets of environmental variables, NO<sub>3</sub>, SiO<sub>2</sub> and PO<sub>4</sub> were highly correlated ( $r = 0.78-0.97$ ). Thus, I omitted SiO<sub>2</sub> and PO<sub>4</sub> before further analyses and modelling based on Variance Inflation Factor.



### 3.2.4 *Species Distribution Models*

Species distribution models (SDMs) were implemented with the R *biomod2* package development version 4.1 (Thuiller *et al.*, 2009; Thuiller *et al.*, 2022). I fitted and compared eight of the statistical algorithms: generalized linear models (GLM; McCullagh, 2019), generalized additive models (GAM; Hastie, 2017), random forests (RF; Breiman, 2001), artificial neural networks (ANN; Venables & Ripley, 1999), flexible discriminant analysis (FDA; Marmion *et al.*, 2009b), classification tree analysis (CTA; Breiman, 1984), surface range envelope (SRE or also known as BIOCLIM; Busby, 1991), maximum entropy modelling (MAXENT; Phillips *et al.*, 2006). In addition, I added two presence-only models, Non-Parametric Probabilistic Ecological Niche (NPPEN; Beaugrand *et al.*, 2011) and AQUAMAPS ([www.aquamaps.org](http://www.aquamaps.org); Kaschner *et al.*, 2019), to the analysis. NPPEN model was implemented using *nppen* package ([github.com/jiho/nppen](https://github.com/jiho/nppen)), and AQUAMAPS was modelled with the self-written package *aquamean* ([github.com/yuliaUU/aquamean](https://github.com/yuliaUU/aquamean)) and hence hereafter will be referred as AQUAMEAN. I only used species that had more than nine records for the SDM process. Selection criteria were determined based on default settings recommended in Thuiller *et al.* (2009), Thuiller *et al.* (2022), Jones & Cheung (2015), and can be found in Table D1. The overall modelling workflow for each species is depicted in Figure 3.1.

Models were calibrated using the area under the curve (AUC) of the receiver operating characteristic (ROC) and the true skill statistic (TSS). TSS and AUC were used to evaluate model performance using the *pROC* package (Robin *et al.*, 2011). TSS is defined as the sum of sensitivity (true positive rate) and specificity (true negative rate) minus one (Allouche *et al.*, 2006). The AUC is a threshold-independent measure representing the relationship between sensitivity and the corresponding proportion of false positives (1-specificity). AUC ranges between 0 and 1, with values above 0.9 indicating excellent prediction and values below 0.5 indicating a prediction no better than random AUC (Hajian-Tilaki, 2013).

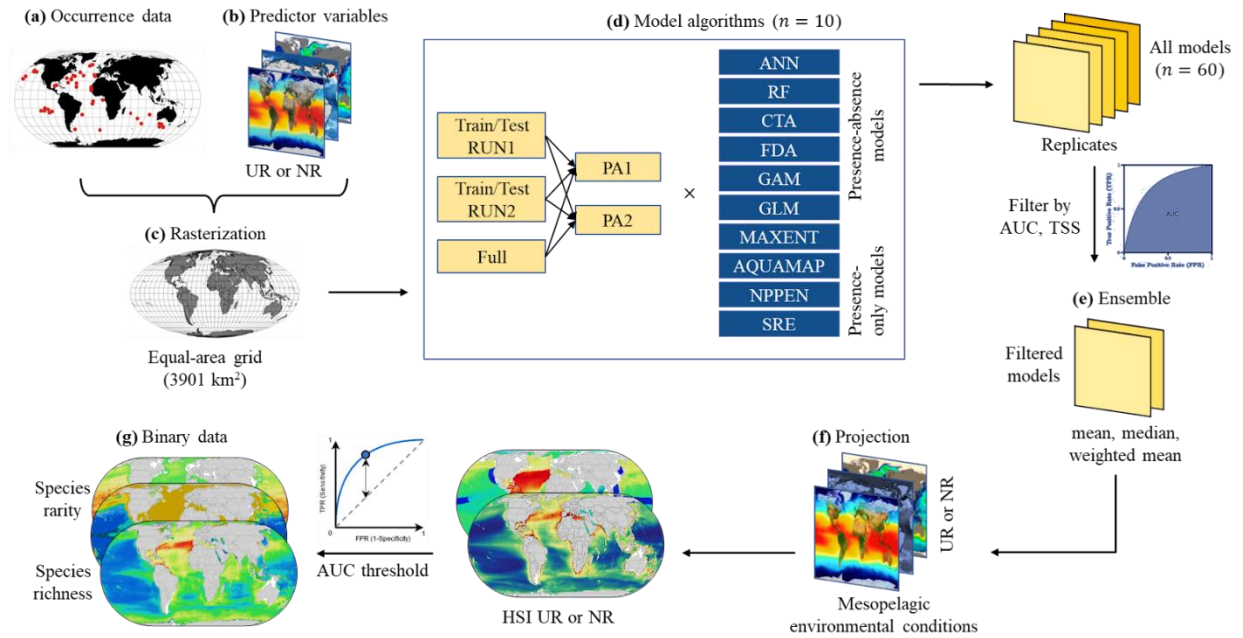


Figure 3.1: The Species Distribution Model workflow depicts the different analytical steps performed for each species. Depending on the sample size, the train/test split is different. Abbreviations used in the diagram: UR - species' unrestricted range; NR - species' native range; HSI - habitat suitability index; PA - sets of pseudo-absences; AUC- area under the receiver operating curve; TSS- true skill statistics.

Two model projections were selected: unrestricted range and native range models. The former was run using all the species occurrences, and the model was projected into the entire mesopelagic area. The latter used only occurrences from native range, and the projection was performed within native range habitat.

Several models integrated into the *biomod2* required both presence and absence data (except MAXENT, which uses background points or SRE which is a presence-only model). Since absence data for most marine species are unreliable due to the low catchability of sampling methods, I generated pseudo-absence (simulated absence; PA) data. The use of pseudo-absences in presence-absence models can have a significant effect on the results of the model (Grimmett *et al.*, 2020). The choice of pseudo-absences and the method used to generate them can impact the accuracy and reliability of the model predictions (Jones *et al.*, 2012). For example, if the pseudo-absences are not appropriately placed to reflect the full range of environmental conditions that the species could potentially inhabit, the model will be biased towards the presence of the species in only the specific type of habitat where it is known to

occur and will miss other potential habitats for the species. I used a random method to generate pseudo-absences to avoid this issue.

To improve the models' predictive accuracy (Barbet-Massin *et al.*, 2012), I set the prevalence (proportion of occupied locations relative to the number of pseudo-absence points) to 1:1, which generates the same number of pseudo-absences as there are presence data sent to the training set. Pseudo-absences were randomly generated in a predefined geographic area (native range and unrestricted range) except for the locations with presence data. For consistency, the same pseudo-absence points were selected as the background points in MAXENT (Thuiller *et al.*, 2022), and the same pseudo-absence points were used to compute the ROC curve for NPPEN and AQUAMEAN models. I assumed that pseudo-absences represented 'true' absence locations when I computed the AUC of the model predictions. Thus, predictions could be compared consistently across presence-only models using ROC curves and AUC scores.

### 3.2.5 Evaluation Measures

For every species and model algorithm, two replicate runs with different random subsamples (training and testing sets) were performed along with an additional full model run to evaluate model accuracy. The split ratio varied based on the number of observations. For species with a large number of occurrence points ( $n > 200$ ), I used 70% of the data to train (calibrate) the model and the remaining 30% to test (evaluate) model predictions (Fielding & Bell, 1997). For species with few occurrences between 44 and 200, I used an 80/20 split, and for species, with less than 44 observations, I used a 95/5 split. To explore the effect of a random choice of the pseudo-absences on model prediction and fit for each species, two pseudo-absence runs were run. For each species, 60 model runs were performed: ten model algorithms x two pseudo-absence runs x three runs (two subsamples and one full run).

The difference in subsampling and choice of pseudo-absences was determined by computing the difference in AUC scores ( $\Delta$ AUC) between two runs (same pseudo-absence) or between two pseudo-absence runs (same subsample), respectively. I consider an AUC of 0.1 to be an acceptable difference between the runs.

### 3.2.6 Ensemble Modelling

I modelled the species' climatic envelope to improve the robustness of forecast abilities using an ensemble forecasting approach (Araújo & New, 2007). Eight different modelling algorithms were combined into an ensemble model using three approaches: mean, weighted mean, and median (note that AQUAMEAN and NPPEN models were not included in the ensemble because they are designed for presence-only data). I only present here the weighted mean approach (a better model is given a stronger weight in the ensemble) because this method is significantly superior to other methods in predicting the accuracy of species distribution forecasts (Marmion *et al.*, 2009a). TSS and AUC evaluation statistics were used as a basis for weighting models (Hao *et al.*, 2019). Only robust replicate runs (individual models with an AUC score above 0.6) were included in ensemble models. The AUC score and 95% confidence interval (CI) were used to compare model performance to evaluate the best-performing model for each phylum or class. I report Variable Importance Plots (VIPs) for all the mesozooplankton groups and disaggregated per phylum. VIPs were constructed to explore environmental predictor variables' contribution to the ensemble weighted mean model output.

### 3.2.7 Global Species Distribution Projections

Each model was run for every species and used to predict the geographic location of the potential climatic niche. Then calibrated models were used to project species distributions using the set of climatic variables for the global mesopelagic zone (MESO set). Here I only present the output of the ensemble weighted mean model. The model output is a raster layer consisting of equal-area grid cells, spanning the global mesopelagic zone (for unrestricted range models) or the native range mesopelagic provinces (for native range models).

The habitat suitability index (HSI) was computed for each species using weighted averages, means and median. HSI has a scale of 0 to 1,000 (where 0 refers to low suitability and 1,000 to a high suitability area) for all *biomod2* models. Hence, the output of NPPEN and AQUAMEAN were converted to the same scale. Total HSI for the entire mesopelagic community was calculated as a sum of individual HSI for each species in each location. The total HSI score was normalized to a scale from zero and one to be able to compare unrestricted

range and native range models. To compare the output of unrestricted range and native range models, I plotted the correlation of HSI between the models for each ocean basin.

### 3.2.8 *Species Richness & Rarity*

Binary (modelled presence/absence) outputs were generated from the AUC threshold value of the ROC using the Youden index (Youden, 1950). This index maximizes the sum of sensitivity and specificity, and graphically can be represented as the maximum vertical distance between the ROC curve and the diagonal line. If the HSI was greater than or equal to the threshold, then the species were considered present. Species richness was subsequently computed by summing the number of species present in each grid cell. Additionally, I computed two metrics of rarity for each cell: total range-size and average range-size rarities. Total range-size rarity, also known as endemism richness (Kier & Barthlott, 2001) or weighted endemism (Crisp *et al.*, 2001), was calculated as sum of the inverse range-sizes for each cell (Pollock *et al.*, 2017). Average range-size rarity, also known as corrected weighted endemism (Crisp *et al.*, 2001), was calculated as total range-size rarity divided by the number of species in the cell. For each map of species richness and rarity, the latitudinal diversity gradient (LDG) was computed (along with a mean and maximum number of unique species per latitudinal degree). Global patterns of species diversity and endemism, represented by species richness and total range rarity, respectively, were mapped. In addition, I subsequently related natural logarithms of the mesopelagic species richness with average water temperature in the mesopelagic layer and the inverse of thermal energy. Inverse thermal energy was calculated using  $1/kT$  formula, where  $k$  is Boltzmann's constant and  $T$  is temperature in Kelvin.

## 3.3 Results

In both unrestricted range and native range, a total of 108,486 models were executed, covering 861 species, and utilizing ten algorithms, three ensembles, two subsampling runs, one full run, and two pseudo-absence set runs. Further analysis and model evaluation were done only on the unrestricted range models since native range model did not produce a realistic picture for HSI (see Figure 3.2B) due to potential sampling biases and lack of existing knowledge on mesopelagic zooplankton distribution (i.e., absence of range maps).

Out of all unrestricted range models, only 2.1% failed to run (2,247 models), due to having an AUC score of less than or equal to 0.5, making it impossible to determine the cutoff point. Most of the failed models were for species with a small number of occurrence points (interquartile range (IQR) is 16-34 observations). The distribution of TSS and AUC scores is shown in Figure D3, and sensitivity, specificity and threshold metrics are listed in Table D2. On average, a single species took 31 min to run (given personal computer configuration of Windows 11 operating system, 1TB SSD storage, Intel Core i9 processor, 32GB RAM), but the time varied from six min to over two hours. The time it took models to run grew logarithmically with the number of occurrences available (Figure D4).

### 3.3.1 Comparing HSI in Unrestricted and Native Range Models

For the unrestricted range model, areas with high HSI include the North Atlantic Ocean, coastal cells near the Mediterranean and Arabian Seas, and the Caribbean region (Figure 3.2A). Low HSI values were found in ocean basins, particularly in the Pacific Ocean. The Arctic Ocean generally has a medium HSI value, while the pattern of HSI in the Southern Ocean is more complex with lower values in circumglobal Subtropical Front and the sub-Antarctic waters.

For the native range model, only the North Atlantic basin was identified as an area with high HSI (Figure 3.2B). Such observation can be partially explained by high sampling bias with most of the observations found in the North Atlantic. Very low HSI values were observed in California Current, Equatorial Pacific, Guinea Basin and East Equatorial Atlantic, Peru Upwelling/Humboldt Current, Benguela Upwelling, Sea of Japan, South China Sea, and Coral Sea mesopelagic provinces.

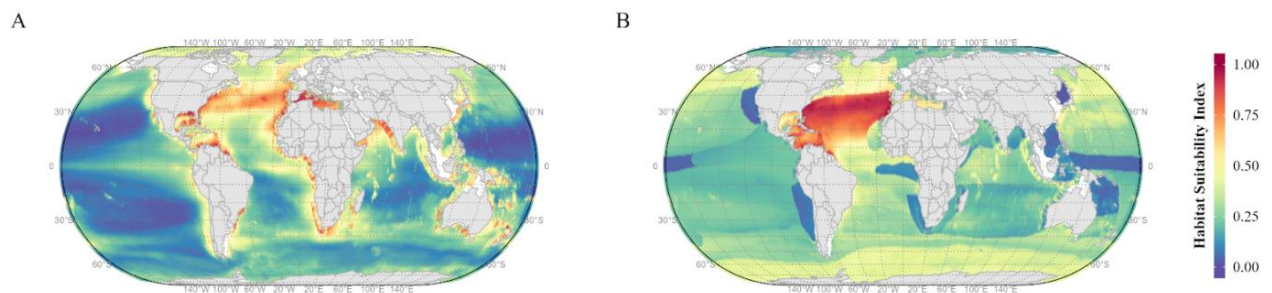


Figure 3.2: Sum of habitat suitability indices for the mesopelagic mesozooplankton community (n=861) modelled by the ensemble (weighted mean) approach using A) unrestricted range and

B) native range. Sum of the HSI is normalized to a maximum of 1. All maps in Eckert IV global equal area projection.

By comparing unrestricted and native range models, I can gain insights into the sufficiency of sampling. If the unrestricted range model consistently outperforms the native range model in predicting species distributions, it suggests that the sampling may not be sufficient. By considering occurrences outside the native range, the unrestricted range model suggests that the species may have a broader distribution than initially believed. In contrast, if the native range model performs equally well or even better than the unrestricted range model, it implies that the sampling is likely sufficient, and the species' distribution is adequately captured within its native range. When comparing the HSI index between unrestricted range and native range models, it becomes apparent that only certain mesopelagic provinces have sufficient sampling (Figure D6). Thus, Central North Atlantic has more than sufficient sampling effort, like the Northern Central Pacific, and Eastern Tropical Pacific and part of the Antarctic/Southern Ocean mesopelagic zones. The rest of the mesopelagic provinces showed severe under-sampling in the mesopelagic realm.

### 3.3.2 Models Evaluation

Many individual modelling algorithms showed good performance when applied to the test database (Table D2). The estimated threshold (cutoff point above which a presence prediction is made) was close to 500 for all models (range 479-509) except MAXENT and NPPEN whose thresholds were 344 and 173, respectively. Predictions from FDA and RF were the most accurate, with AUC scores of  $0.88\pm 0.16$  and  $0.94\pm 0.15$ , respectively. In contrast, predictions from AQUAMEAN and SRE had the lowest AUC scores of  $0.59\pm 0.12$  and  $0.68\pm 0.13$ , respectively. Predictions from CTA, GAM, and GLM had very similar AUC scores, with values ranging from 0.85 to 0.86. Predictions from MAXENT had an AUC score of  $0.83\pm 0.16$ , while the presence-only NPPEN model showed similar performance with an AUC score of  $0.84\pm 0.11$ . Finally, predictions from ANN had an AUC score of  $0.80\pm 0.18$ . AUC scores of the predictions of the model ensemble ( $0.94\pm 0.04$ ) were higher than all individual model predictions except for RF in which their AUC scores were similar. The ranking and AUC scores were consistent when AUC scores of model predictions by taxonomic groups were compared (Figure 3.3).

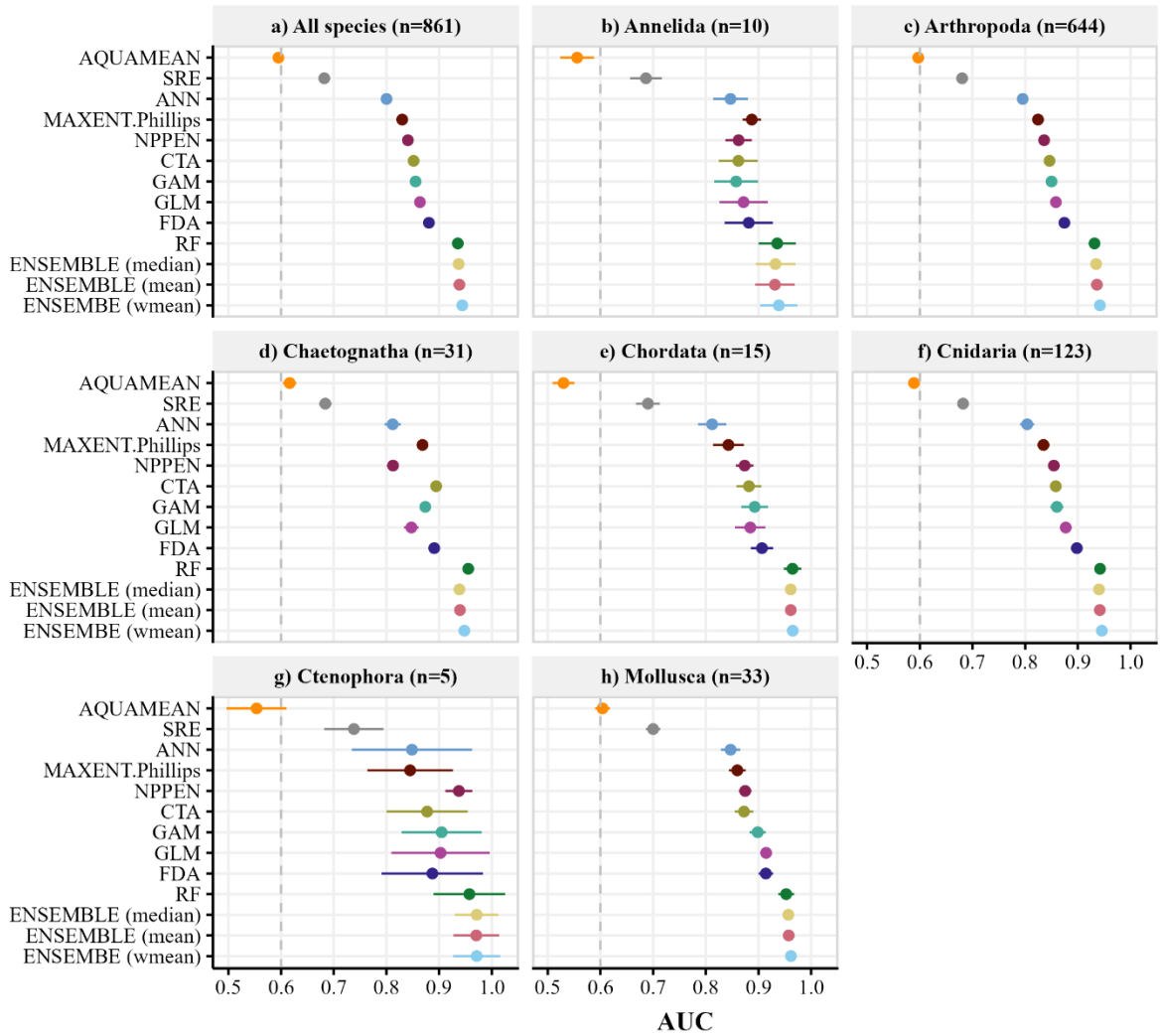


Figure 3.3: Distributions of the area under the curve (AUC) scores of predictions from individual species distribution models and their ensemble. Means and 95% confidence intervals are presented for a) all species combined (n=861) and by phylum: b) Annelida (n=10), c) Arthropoda (n=644), d) Chaetognatha (n=31), e) Chordata (n=15), f) Cnidaria (n=123), g) Ctenophora (n=5), h) Mollusca (n=33). The dotted vertical line represents the threshold (AUC score = 0.6) above which predictions from the models were included in the ensemble model.



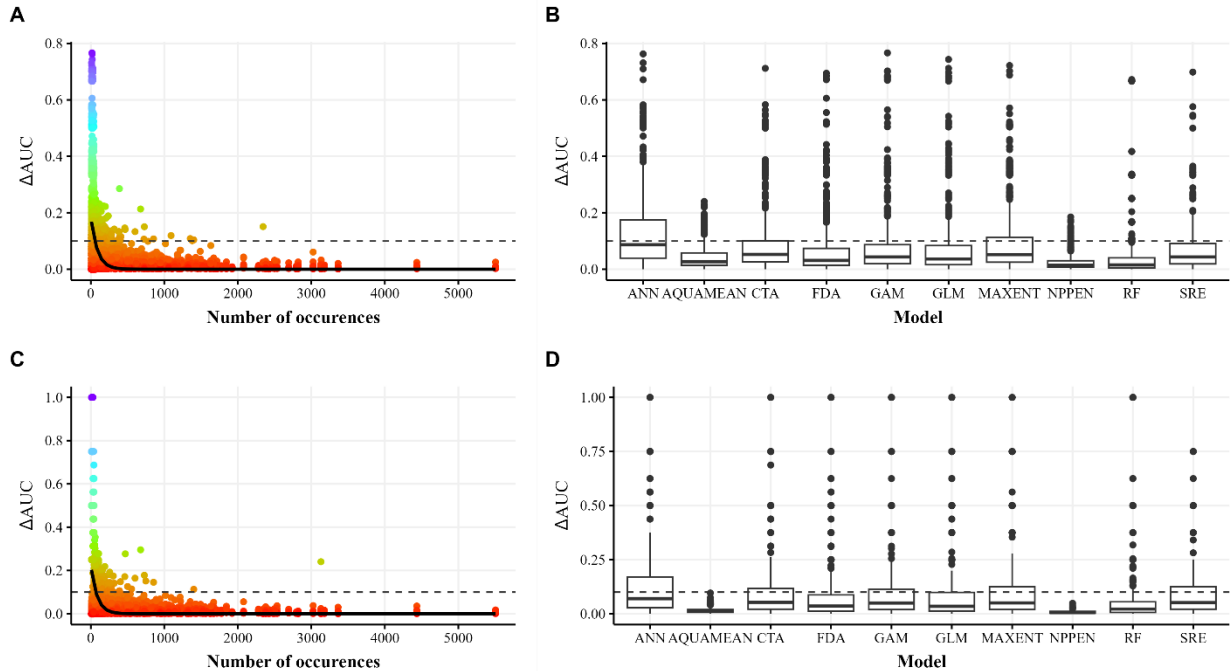


Figure 3.4: Left column showing the difference in area under the curve scores ( $\Delta AUC$ ) as a function of number of occurrence points between two pseudo-absences runs (A) and between two subsampling runs (C). Right column showing the difference in  $\Delta AUC$  disaggregated by modelling algorithm for two pseudo-absences runs (b) and between two subsampling runs (d). Box plots showing the estimated median (horizontal black lines), 25th–75th percentiles (boxes) and the 1.5 interquartile range (whiskers) of  $\Delta AUC$ , points are outliers. Solid lines in panels (a) and (c) were fitted with exponential functions : (a)  $\Delta AUC = 0.19e^{-0.01N}$ ,  $R^2 = 0.24$ ; (c)  $\Delta AUC = 0.23e^{-0.01N}$ ,  $R^2 = 0.19$ .

Different methods of generating pseudo-absences of species slightly affected the overall mean AUC scores of predictions although the sensitivity varies between models and species (Figure 3.4A, B). The mean difference in AUC between the two pseudo-absence runs was  $0.07 \pm 0.10$ . However, the effects of pseudo-absences on the predictions for 56 to 81 species were large with a difference in AUC score of  $\geq 0.25$  (Table D3). When excluding the presence-only models (SRE, AQUAMEAN and NPPEN), the RF model showed the smallest difference in performance, with a difference in AUC less than 0.1 for 88% of the species. FDA was the 2nd most accurate with 78% of species having  $\Delta AUC < 0.1$ . CTA and GAM models show  $< 0.1$  difference in AUC score in 74% of the species. MAXENT had slightly lower performance, with 70% of the species showing an acceptable difference in pseudo-absence runs. An acceptable difference between pseudo-absence runs was observed only in 39% of species in ANN models. Moreover, ANN had the largest number of species with at least one failed

pseudo-absence run (29% of all species). FDA and GAM have 3-5% of species where at least one of the runs failed, and GLM only had less than 1% of failed runs. The remaining models did not have any failed runs. The observed AUC difference between the pseudo-absence runs was correlated with the number of available presence records (Figure 3.4A). The difference between the pseudo-absence runs decreased exponentially with an increase in the available number of occurrence points (presence data).

Similar results were observed when comparing the difference between subsampling runs of the models (Table D4). All runs in AQUAMEAN and NPPEN had acceptable AUC differences ( $\Delta\text{AUC} < 0.1$ ) between resampling runs. RF also showed consistent results for 83% of species. The difference between the resampling runs was  $< 0.1$  only in 51% of the species in ANN models. The larger difference in AUC values was also attributed to species with a low number of presences (Figure 3.4C). ANN, FDA, GAM and GLM were the only models with failed runs (17.3%, 33.6%, 4.8% and 0.9%, respectively; Table D4)

### 3.3.3 Environmental Variable of Importance

The distribution predictions for the 861 species were primarily determined by salinity, dissolved nitrate, euphotic depth, and dissolved oxygen while NPP and mixed layer depth had the least effect on the predictions (Figure D5). The relative importance of each environmental variable in determining species' distribution differed between phyla. For instance, for Annelida (10 species) and Ctenophora (5 species), dissolved nitrate, euphotic depth, dissolved oxygen, and temperature were the most important variables while dissolved nitrate was the most important variable for Mollusca (33 species) and Chordata (15 species). For Cnidaria (123 species), salinity and euphotic depth were the most important variables while it was temperature for Chaetognatha (31 species). NPP was consistently the least important predictor for all phyla.

### 3.3.4 Geographical Patterns of Species Richness

I utilized the HSI values from the unrestricted range model to investigate species diversity patterns, as the native range model did not perform well in many mesopelagic provinces due to the low sampling effort in those areas (Figure 3.2B; Figure D15C). The species richness map (unrestricted range) constructed using the weighted mean ensemble model (Figure 3.5A)

showed the highest diversity in the North Atlantic between 30 and 40°N band, in near-shore areas, and in the Mediterranean Sea. Equatorial areas had moderate species richness. Areas of low diversity spatially coincided with all basins of the Pacific, South Atlantic, Indian oceans and circumglobal Subtropical Front. Consistent with the spatial pattern of HSI (Figure 3.2), the Arctic Ocean showed uniformly moderate species richness values while those in the Southern Ocean varied between areas. The most extreme latitudes where species richness was observed extended up to 75.8°S in the southern hemisphere and 80.3°N in the northern hemisphere (Figure 3.5B).

Predicted ‘hotspots’ of species richness (high number of species per cell) were primarily found in cells between 40°S and 40°N with the highest values in species richness found around 30° latitudes (Figure 3.5B). At high latitudes, the maximum species richness per cell did not exceed 200 in the Southern Hemisphere and less than 300 species in Northern Hemisphere. However, species richness in areas between 30°S and 30°N was non-uniform and varied drastically depending on the cell location. In contrast, at high latitudes, the variability in species richness values was five orders of magnitude lower. As a result, although no species richness hotspots were found in polar areas (areas higher than 65.5° latitude), the number of species per cell was more uniform. Thus, when looking at mean species richness values per degree latitude (black line in Figure 3.5B), polar areas have on average higher mean species richness per cell than any equatorial or tropical areas; the highest mean species richness was found at approximately 40°N.

When all 861 species were used to calculate total and average- range size rarities, an unexpected pattern of high rarity values were found at the eastern side of the South Pacific gyre (Figure D13E and F). Closer examination of the area showed that the high rarity score (especially for average rarity) resulted from the presence of two crustacean species *Allosergestes pestifer* and *Echinomysis serratus*. *A. pestifer* was a formally known as *Sergestes pestifer* (Burkenroad, 1937) received its new taxonomic genus in 2008 (WoRMS, 2023), and its occurrence records are restricted to the eastern part of the South Pacific gyre and likely had a narrow environmental envelope due to a narrow environmental sampling. Since during the SDM process the distribution of species is determined by the set of environmental conditions (specifically mixed layer depth), the SDM could predict the occurrence of *A.*

*pestafer* only in a specified area. A similar effect of environmental variables on *E. serratus* prediction has also contributed to the observed anomaly. Based on the above, I decided to omit these species in calculations of rarity scores.

There were seven outlier cells with very high total range-size rarity scores ( $>0.97$ ) in the North Atlantic along the east coast of Canada (Figure 3.5D at  $\sim 40^\circ\text{N}$ ). It can be explained by the presence of *Unciola irrorata* only in those cells which made the rarity of those cells extremely high. To better see the patterns of total rarity and their LDG, I filtered out the total rarity greater than 0.97 to remove the points with high rarity. When outliers were removed, that pattern in total range-size rarity became more evident: high values were found in pelagic areas closest to coasts, in the North Atlantic and Mediterranean Sea (Figure D14). In general, total rarity was higher in polar areas of the Northern Hemisphere than in the Southern Hemisphere. Lowest rarity was observed in subtropical gyre regions of the Pacific, Indian and South Atlantic oceans. Total range-size rarity (Figure 3.5D, Figure D14) was generally higher towards the high latitudes of the Southern Hemisphere.

Total range-size rarity is correlated with species richness (Hobohm, 2014), so when corrected by a number of species in a cell, average range-size rarity showed a very biased outcome with a high rarity observed in the South Pacific gyre at  $30^\circ\text{S}$  (Figure 3.5E) that can be explained by a very small sampling effort in that area (only 5-20 species are predicted within that region). Overall, the average rarity remained constant between  $30^\circ\text{S}$  and  $30^\circ\text{N}$  latitudes, while it increased towards the poles in both hemispheres. Highest average scores were found in the North Pacific in areas close to the coast.

Overall, the North Atlantic, Arctic and the Southern Ocean were areas with high species richness and rarity (Figure 3.5G and H). Almost the entire Pacific and center of the Indian Ocean were areas with low rarity and low species richness, except for the eastern part of the South Pacific gyre where low species richness and high average rarity were observed (Figure 3.2H).

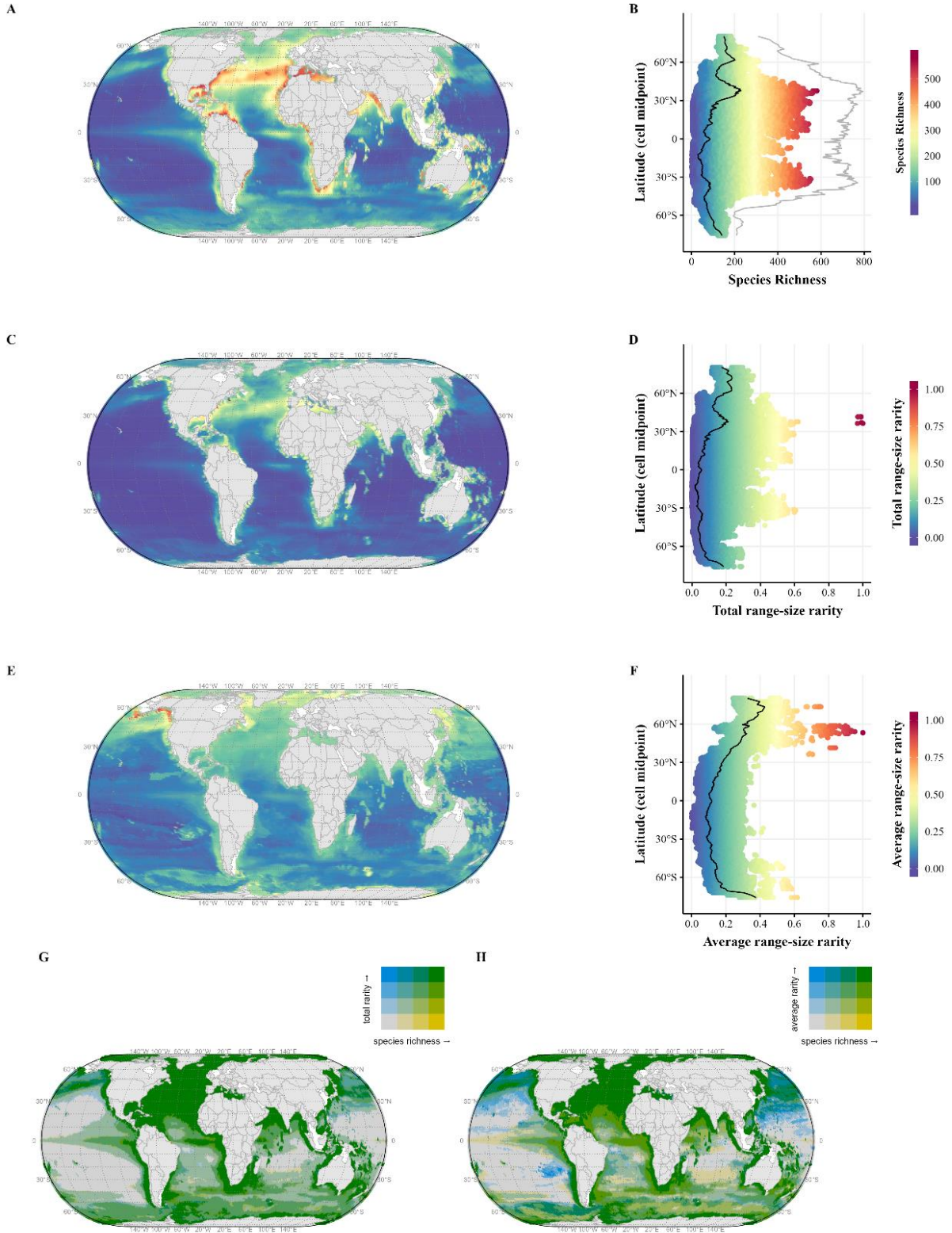


Figure 3.5: Mesopelagic mesozooplankton diversity maps. A) Mesopelagic mesozooplankton species richness map (n=861). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) Average range-size rarity map and E) Average range-size rarity map and F) Average range-size rarity map. G) Total range-size rarity map and H) Average range-size rarity map.

latitudinal gradient for the total range-size rarity. A solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. A solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 25th, 50th, and 75th percentiles. All maps in Eckert IV global equal area projection.

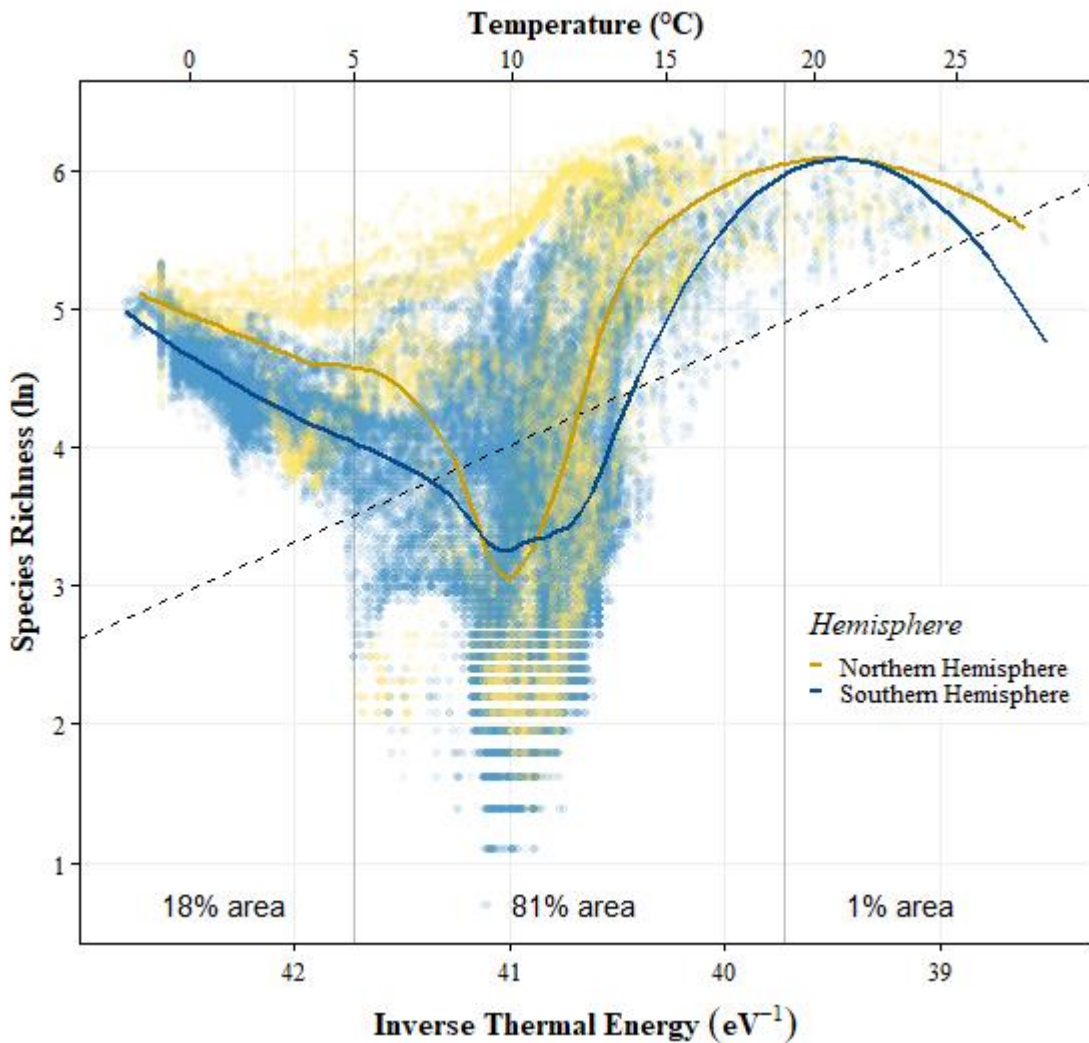


Figure 3.6: The natural logarithm of species richness as a function of water temperature ( $^{\circ}\text{C}$ ) in the mesopelagic zone and inverse thermal energy ( $1/kT$ , where  $k$  is Boltzmann's constant and  $T$  is temperature in Kelvin). Trendlines shown for each hemisphere (regressions with local polynomial fitting). Dashed black line represents the equation with slope  $-0.70$  from metabolic theory from Brown *et al.* (2004).

A common predictor of diversity gradient, especially for ectothermic species, is temperature. I subsequently related mesopelagic species richness with average water temperature in the mesopelagic layer and the inverse of thermal energy (Figure 3.6). 81% of the cells have a mean

temperature in mesopelagic zone between 5 and 19°C, 18% of cells had mean temperatures colder than 5°C, and only 1% of areas had temperatures higher than 19°C. Around 150 species per cell can be found at the coldest water temperatures, with species richness decreasing to its global minimum at around 10°C and then gradually increasing reaching maximum at approximately 20°C. However peak and decline at higher temperatures are not well established since very few data points were available at those temperature ranges. After closer inspection, I found that the dip at ~10°C was also characterized by very high variability in mean temperatures in those regions (Figure D16).

## 3.4 Discussion

### 3.4.1 *Unrestricted and Natural Ranges*

In various SDM models, it is a widespread practice to use species range maps to help establish the geographical boundaries of where the species can occur and avoid projecting species occurrences in the area where species cannot be found. It also helps with the selection of pseudo-absences, as they can be taken outside the species range. However, range maps are not widely available for zooplankton groups, except for a few species of plankton for whom printed range maps are published (van der Spoel & Heyman, 1983) but hardly numerically adapted. In this work, I have demonstrated that if I only model zooplankton into the native range area (which is determined by available species occurrence points), the produced HSI map is very patchy (Figure 3.4B), with apparent areas of the mesopelagic realm where almost no zooplankton species are projected (since few to no data are available in the area). The patchiness I see in native range HSI can be explained by uneven sampling effort (Figure D15C). The Pacific Ocean, South Atlantic, and the basin of Indian Ocean are severely under sampled, while the polar regions are characterized by comprehensive coverage and a moderate number of observed species (Figure D15C). As a result, many provinces lack consistent sampling or only have a few species sampled within them, and the environmental envelope that represents those areas are not adequately represented in SDM.

However, if the mesozooplankton species can occur in any part of the ocean as long as environmental conditions are optimal, the HSI map produces more consistent, smooth patterns of species occurrence (Figure 3.4A). The discrepancy between two approaches comes from

the severe undersampling in many mesopelagic provinces (Figure D6). Mesopelagic provinces in the Arctic, South Atlantic and Pacific Oceans lack consistent sampling, especially their central ocean basins.

However, one of the disadvantages of unrestricted range approach is that species can be projected in the areas where they do not occur. It was observed that 637 of the modeled species have occurrence records in both hemispheres at latitudes greater than 40°, thus the majority of species (74%) are already projected in both hemispheres. However, 26% of modelled species are projected into the hemisphere where they were not observed. For instance, species that can only be found in the Pacific Ocean, can be projected in Atlantic Ocean where they may or may not been recorded, or species that are only found in the Northern hemisphere can be projected into the Southern Hemisphere. Thus, projections of unrestricted range models can be less accurate than native range models since the majority of species are not found across the global ocean. However, it is plausible to assume that similar, ecologically equal, species (e.g., *Calanus finmarchicus* and *Calanus pacificus*) that have comparable niches could potentially be found in every basin with similar environmental settings. Therefore, the use of unrestricted range models to compare species richness maps and explore model performance can be reasonably justified.

I acknowledge that the produced species richness map is a simplified version of reality assuming that every species can be everywhere all at once. However, I rest on the plankton paradox (Hutchinson, 1961) to provide a close to reality picture of what the mesozooplankton diversity gradient should look like when relying on knowledge on the relationship between available occurrence and mesopelagic environment.

### 3.4.2 Model evaluation

The ensemble approach is considered the optimal way to model mesozooplankton distribution patterns (Figure 3.3). However, the process is very time-consuming (Figure D4). To make the process faster, RF models can be used as they have shown almost similar predictive performance (Figure 3.3) to the ensemble models, but computationally are substantially faster. In addition, RF models show consistent results with different choices of pseudo-absences or subsampling ratios (Figure 3.3B, D). However, there are several advantages to using the



ensemble technique: ensemble modelling combines the outputs of multiple machine learning models to produce more accurate and robust predictions (Polikar, 2006). By leveraging the strengths of different models, ensemble models can reduce overfitting, improve generalization, and be more flexible and robust to variations in the data. Also, the ensemble models allows one to quantify and map uncertainty between models showing areas of uncertainty in the projection. Having evaluated multiple performance metrics, I utilized the outcomes of the weighted mean ensemble models to investigate biodiversity patterns.

I explored two important factors that can affect the model performance: subsampling and the choice of pseudo-absence sets (Figure 3.4). Subsampling in machine learning involves selecting a smaller subset of data for training the model, and it can have both positive and negative effects on model performance. On the one hand, subsampling can help reduce overfitting, speed up training, and reduce computational resources. However, if the subsampling rate is too high, it may lead to bias, decreased accuracy, or loss of important patterns or relationships in the data due to niche truncation. As such, the effect of subsampling on model performance depends on various factors, including the dataset size, subsampling technique, and machine learning algorithm. The location of pseudo-absences is important for model performance, as they should be representative of true absences and located in areas where the target variable is unlikely to occur. If pseudo-absences are in areas too similar to presence locations, the model may struggle to differentiate between presence and absence. Such instances can lead to overfitting, a situation where the model becomes too specialized for the training data and does not generalize well to new, unseen data. On the other hand, if pseudo-absences are in areas that are too dissimilar to presence locations, the model may not learn the important factors that drive the target variable's distribution. Such a scenario can result in reduced predictive power because the model will not be able to accurately predict the target variable's presence in new locations. Thus, careful selection of pseudo-absences locations is crucial for effective modelling. Additionally, the number of pseudo-absences used in the model can also have an effect on the results. Using too few pseudo-absences can lead to overfitting, while using too many can lead to overgeneralization and decreased model performance. In such circumstances, employing range maps might be an effective strategy to mitigate this issue. Furthermore, characteristics of the species can typically guide the random sampling process and fine-tune the selection. In this study, I employed various data subsets to train the model

and found that the difference in AUC scores was generally less than 0.2 across most of the models (Figure 3.4D). However, when the number of occurrences was limited (Figure 3.4C), all models displayed AUC value differences exceeding 0.2. Notably, I utilized different splitting ratios based on occurrence numbers, thus the mean AUC differences can vary for each split. Very similar results were observed when different locations of pseudo-absences were chosen. For all models, the most consistent results are obtained for species with higher occurrence records (>500 observations; Figure 3.4A). In this case, the change in AUC was usually less than 0.1 for different pseudo-absences and subsampling runs (Table D3 and D4).

I would like to highlight the significance of carefully selecting the appropriate number of pseudo-absences for SDM modelling. Although Barbet-Massin *et al.* (2012) recommended using 10,000 pseudo-absence for modelling with GLM and GAM algorithms, I find such recommendations not applicable to MM species. I did not find this approach feasible as the amount of presence data for many plankton species is generally < 500, and in the absence of the expert range maps generally limited confidence in the location of pseudo-absences such as in Barbet-Massin *et al.* (2012). Selecting too many pseudo-absence points in random locations appeared to introduce additional errors in estimating species occurrences, by placing, for example, species absence in a location optimal for its occurrence. Therefore, I used the same number of pseudo-absences as presences, following a recommended ratio for CTA, BRT, RF models suggested by Barbet-Massin *et al.* (2012), and applied them for all modelling algorithms.

### 3.4.3 Environmental Variable of Importance

According to the study results, the distribution of 861 species of marine organisms was primarily driven by salinity, dissolved nitrate, euphotic depth, and dissolved oxygen, while NPP and mixed layer depth had the least impact (Figure D5). This pattern was particularly evident in the phylum Arthropoda, which made up 75% of the total species studied.

Salinity is important for mesopelagic zooplankton as it influences their distribution, physiology, and ecological interactions. These marine organisms must maintain a stable internal environment to survive, regulating their internal ion concentrations and water content in response to changes in salinity (osmoregulation) (McNamara *et al.*, 1983; Aguilar *et al.*,

1998). Variations in salinity create physical or chemical barriers that affect zooplankton movement and distribution, with different species having specific salinity preferences (Ojeda *et al.*, 2022). Indirectly, salinity influences mesopelagic zooplankton by affecting the distribution and productivity of phytoplankton, their primary prey, through changes in nutrient availability and water column stratification. Lastly, changes in salinity can alter the distribution and density of predators and competitors, affecting zooplankton survival and community dynamics.

Changes in nitrate concentrations can indirectly impact zooplankton distribution, density, and overall community structure. High nitrate concentrations can lead to the development of oxygen minimum zones (OMZs) in the water column (Noone *et al.*, 2013), which can negatively affect mesopelagic organisms. In these zones, the concentration of dissolved oxygen is low, which can limit the survival and growth of many species, particularly those that require high levels of oxygen (Bakun, 2022). It is also important to note that nitrate was highly correlated to silicate and phosphate. Thus, the response to species to changes in phosphate can also be linked to how species react to fluctuations in nitrate and silicate levels.

Dissolved oxygen is essential for the respiration and metabolism of zooplankton, and variations in dissolved oxygen concentrations can impact their survival, growth, and reproduction. High oxygen concentrations generally support a diverse and abundant zooplankton community, while low oxygen concentrations can create hypoxic or anoxic conditions, limiting the distribution of many zooplankton species and leading to reduced biodiversity (Soviadan *et al.*, 2022). Some species have adapted to tolerate low-oxygen environments (Vaquer-Sunyer & Duarte, 2008), but many are unable to survive under such conditions. Additionally, dissolved oxygen levels can influence the vertical distribution of zooplankton, as they often occupy specific depth ranges with suitable oxygen concentrations to meet their metabolic needs. Furthermore, variations in oxygen concentrations can affect the distribution and behavior of zooplankton predators and prey, ultimately shaping the structure and dynamics of mesopelagic food webs (Bakun, 2022).

Euphotic depth, the depth to which sufficient light is available for photosynthesis, can indirectly influence zooplankton distribution by altering the distribution and density of phytoplankton. Phytoplankton growth and biomass were also shown to be affected by thermal

stratification patterns, including modifications in vertical mixing processes that correspond to mixed layer depth (Henson *et al.*, 2021).

Water temperature, although was only found important for a couple of taxa (Figure D5) can affect the metabolic rates of zooplankton, with higher temperatures generally leading to increased metabolic activity, growth, and reproduction rates (Rohde, 1992). As a result, it plays a significant role in determining the distribution and density of different zooplankton species (Dipper, 2022), with each species occupying specific thermal niches in the water column. However, in the mesopelagic, the temperature gradient is smaller, and as a result temperature was not as important factor in this study as other environmental variables.

Across all phyla, NPP was not found to be an important predictor, which is consistent with findings from Gagné *et al.* (2020). NPP was an important variable in many studies especially in the epipelagic layer (Vallina *et al.*, 2014; Hernández-León *et al.*, 2020; Rombouts *et al.*, 2009). However, in the mesopelagic zone, the primary source of organic matter is not the direct result of photosynthesis, but rather, it is derived from the sinking of particulate organic carbon (POC) from the surface waters. Mesopelagic zooplankton, therefore, rely heavily on POC as a food source, and the availability and quality of this organic matter can be critical for their survival and growth (Guidi *et al.*, 2008).

However, the importance of each environmental variable changed when the species were disaggregated by other phyla. For instance, in Annelida and Ctenophora, salinity was one of the least important factors, along with mixed layer depth and NPP, while dissolved nitrate, euphotic depth, dissolved oxygen, and temperature were the most important variables. Dissolved nitrate was the most important predictor for Mollusca and Chordata, while salinity and euphotic depth were most important for Cnidaria, and temperature was the most important variable for Chaetognatha. These results highlight the importance of considering the specific ecological requirements of different taxa when predicting their distributions. However, I presented an average importance of each environmental variable aggregated by phylum. The results may differ when considering species-level distributions (Bosch *et al.*, 2018), as different species within the same phylum may have different ecological requirements due to specific evolutionary adaptation.

Furthermore, while the results suggest that certain environmental variables, such as salinity, dissolved nitrate, euphotic zone depth, and dissolved oxygen, have a major impact on the distribution of marine species, it is possible that other factors, such as biotic interactions, dispersal ability, and historical events, may also play a role in determining species distributions. Therefore, it is crucial to consider the results of this study as a first attempt to model species distributions and to acknowledge that potential influence of additional factors when making predictions about the distribution of individual species. Further research at the species level may provide more detailed and accurate insights into the environmental drivers of species distribution in marine environments.

#### 3.4.4 Geographical Patterns of Species Richness

Several studies have explored global patterns of marine species diversity (Tittensor *et al.*, 2010; Costello & Chaudhary, 2017; Righetti *et al.*, 2019; Chaudhary *et al.*, 2016). However, all these studies have focused either on the entire available diversity or only epipelagic species. This work, to my knowledge, is the first attempt to model individual species distributions in the mesopelagic realm using mesozooplankton fauna. Globally, approximately 7000 species of marine zooplankton, mostly holoplanktonic, have been identified and described to date (Boltovskoy *et al.*, 2002). However, the portion of organisms found in the mesopelagic realm is unknown. In this work, I am working with the smaller subset of 861 MM species representing only 12.3% of the total known zooplankton organisms. The chosen subset of species is unlikely to encompass all MM species because it is a challenge to classify a species as ‘mesopelagic,’ particularly in the absence of depth range information for many zooplankton species (Figure D1). Therefore, this work should be considered as a preliminary study that sets up new questions and encourages the ongoing enhancement of the database to support more comprehensive and detailed analyses in the future.

The majority of the modelled species were from the phylum Arthropoda (n=644) due to the large representation of copepods, and hence this pattern of observed species diversity and LDG was more representative of the arthropods than the entire mesopelagic community (Figures D7-14). The prevalence of arthropods in the database is not surprising as arthropods comprised 44 - 47% of total species diversity (0-5000 m depth range) in the Northwest Atlantic Ocean (Bucklin *et al.*, 2010). Members of this phylum also dominate the community

density and biomass values (Koplin, 2020; Jerez-Guerrero *et al.*, 2022; Drago *et al.*, 2022) in both epipelagic and mesopelagic layers.

Other taxa records are patchier and are absent from many regions (Figures D7-14) due to varying sampling effort. For instance, Chaetognatha records are sparse throughout the Pacific Ocean and South Atlantic (Figure D9). More sampling is needed for Chordata (Figure D10) and Mollusca (Figure D13) in the center of ocean basins especially given their importance for commercial activities. However, these taxa are generally more neritic with a lot of possibility of migration and movement making them harder to sample. In addition, very few species of mesopelagic jellyfish were available in MMM Database (<0.08% of the records are from the depth deeper than 200 m) due to the difficulty to sample these taxa with conventional methods without destroying them once brought to the surface, as reflected in the lack of data on jellyfish biomass and depth distribution in mesopelagic realm.

The construction of the species richness map using the weighted mean ensemble model revealed global patterns in zooplankton diversity (Figure 3.5A). The highest diversity was observed in the North Atlantic region, north of the South American continent, along the west coast of India, and in the Mediterranean Sea. Elevated species richness values were also observed in equatorial regions and Arctic and Southern Oceans. These results are consistent with the HSI map (Figure 3.2), which also showed moderate HSI values in the Arctic Ocean and variable species richness values in the Southern Ocean. In contrast, low diversity was observed in all basins of the Pacific, South Atlantic, Indian oceans, and the circumglobal subtropical front that are characterized as high-nutrient, low-chlorophyll (HNLC) regions.

The LDG analysis identified species richness ‘hotspots’ (high number of species per cell) between 40°S and 40°N, with the highest values found in cells around the 30° latitudes (Figure 3.5B) comparable to finding of Rogers *et al.* (2022) and Chaudhary *et al.* (2016) for all observed diversity. The absence of species richness hotspots in polar areas is noteworthy and perhaps counterintuitive, because the latitudinal mean species richness in polar regions appeared to be higher than in equatorial or tropical regions, despite localized hotspots. Although this pattern can arise from various factors such as temperature regime, productivity, currents, and mesoscale features, it appears that polar regions tend to exhibit simpler and less variable species communities. The harsh conditions prevalent in these regions likely support

only a limited and relatively constant number of specialized species. At the same time, warm and dynamic equatorial waters that likely favor a high species turnover rate inhabited by patchy communities of the low and high speciation. Lower mean species richness in equatorial waters is potentially a reflection of the domination of the low speciation patches. It may indeed also be an artifact of sampling efforts and thus should be further investigated.

Mean species richness peaked at approximately 40°N (Figure 3.5B), which may be attributed to several factors. This latitudinal band is a transition zone between colder, nutrient-rich, temperate, and warmer, nutrient-poor, subtropical waters allowing higher productivity that is able to support a wide range of zooplankton species. Furthermore, fronts are by definition mixing zones that harbor neighboring communities mechanistically increasing biodiversity. Marine LDG displays a bimodal distribution with peaks of species richness located around 30° of latitude in both hemispheres and tailing off towards the poles (Figure 3.5B). This result is also consistent with a previous review study that found bimodality in almost all global datasets of 65,000 recent and 50,000 fossil coastal marine species (Chaudhary *et al.*, 2016). However, LDG for global phytoplankton richness showed a unimodal pattern with a peak in the inner tropics (<5°N and S; Righetti *et al.*, 2019). Non-located species richness maxima between phytoplankton and zooplankton species were also recorded by Benedetti *et al.* (2021).

Previous research has suggested that a bimodal species richness pattern in zooplankton could be explained by several factors. One potential explanation is a sampling bias, where a higher number of samples are available in the Northern Hemisphere compared to the Southern Hemisphere (Flessa & Jablonski, 1995). Another possibility is the effect of continental shelf coverage, with equatorial areas having less continental shelf area. Temperature may also be a contributing factor, as suggested by studies such as Tittensor *et al.* (2010) and Beaugrand *et al.* (2020). While it may be true, Chaudhary *et al.* (2016) have demonstrated that even after correcting for the sampling bias, the bimodality persisted. In this analysis, I excluded areas of the continental shelf shallower than 200 m to avoid potential interference of the coastal environment with the mesopelagic community species richness analyses. In addition, temperature affect appeared to be unimportant in this analysis with euphotic depth, dissolved nitrate and salinity being primary drivers of the mesopelagic mesoplankton distribution (Figure D5).

This study found that total range-size rarity shows a similar pattern as species richness, with high values found in the offshore North Atlantic and in coastal areas across both hemispheres (Figure 3.5C). In general, total rarity was higher in polar areas with the highest values observed in the Northern Hemisphere. The lowest rarity was predicted in gyres of the Pacific Ocean. Notably, low rarity in gyres can indeed be explained by the sampling bias. Due to reduced sampling efforts, only a few species had been recorded from this area while well mixed and uniform conditions within gyres may not be properly represented in SDMs. The underrepresentation became even more evident when I explored an outlier of the high average rarity in the South Pacific gyre at 30°S (Figure 3.5E). After removing the outliers, the average size rarity showed a clearer pattern, with the lowest absolute rarity values observed in equatorial areas gradually increasing towards polar areas (Figure D13).

Similar to total range-size rarity, average rarity was higher in the Arctic Ocean compared to the Southern Ocean (Figure D13). The reason for the higher average rarity of zooplankton species in the Arctic compared to the Southern Ocean is not entirely clear and may be influenced by a range of factors. One possibility is that the Arctic has a more diverse range of habitats and niches than the Southern Ocean, which may support a greater diversity of rare and specialized species. Additionally, Arctic ecosystems are more fragmented and isolated than those in the Southern Ocean, which could promote the development of unique and highly specialized species (Hempel, 1985; Eastman, 1997).

After combining two metrics: rarity and species richness, it was found that the North Atlantic, Arctic, and Southern Ocean are areas of high species richness and high rarity (Figure 3.5G and H). The Pacific and Indian ocean basins are areas with low rarity and low species richness. These patterns can be explained by the proximity of the points to the coast, as well as the low resolution used in the study. One possibility is that ocean basins are typically characterized by low productivity and nutrient availability, which can limit the diversity and density of species. Additionally, ocean basins are often vast and homogeneous in terms of environmental conditions, which can limit the development of specialized and unique species.

The evolutionary rates hypothesis states that high temperatures and UV radiation can increase the mutation rate of land plants, resulting in a faster evolutionary rates and speciation (Rohde, 1978, 1992). Therefore, regions with high levels of energy may be characterized by a greater



diversity of species due to the accelerated evolutionary processes in those areas. Evans *et al.* (2005) demonstrated that the relationship between temperature and avifauna species richness typically follows a positive deaccelerating curve. Findings of this study showed a similar relationship when mesopelagic water temperature exceeded 10°C (Figure 3.6). However, I also found a species richness minimum at ~10°C with species richness values increasing slightly when temperatures are colder. Similar dip in species richness was also observed in the phytoplankton diversity (Righetti *et al.*, 2019). Similar to Righetti *et al.* (2023), three regimes across the mesopelagic ocean were identified: gradual decrease in species richness for temperatures below 10°C, steep increase in species richness for temperatures between 10 and 19°C, and more or less consistent high species richness for temperatures greater than 15°C, MM species richness reached maximum at water temperatures of ~20°C. Such warm mesopelagic temperatures found only in semi-enclosed mesopelagic basins (e.g., province 23, 14-16 in Figure D6) or areas closest to the coast in equatorial and tropical regions. However, I did not find good agreement with the metabolic theory of ecology that expect a linear relationship between the variables since the majority of points were located in areas with temperatures between 5 and 19°C, where a very pronounced V-shaped pattern is observed similar to Righetti *et al.* (2023). I speculate that such suppression of species richness in intermediate temperatures can be due to the high variability of temperature in the mesopelagic zone (Figure D6B). However, the underlying mechanism is not known and requires further investigation. There seems to be a similar relationship between species richness and temperature for both epipelagic and mesopelagic layer, indicating that there may be a unifying mechanism that established the species richness in the open pelagic ocean. In this work, I have identified salinity, dissolved nitrate, dissolved oxygen and euphotic depth to be important variables driving the distribution of species (Figure D5), so further work is needed to explore these drivers of species richness in mesopelagic ocean.

### 3.4.5 Limitations

This work provides the first attempt of modelling the species distribution of the MM. More work is indeed required to determine optimal tuning parameters for each species and broader taxonomic groups. Additional research should be done to explore the effect of subsampling ratio for species with low occurrences and a choice of the pseudo-absence numbers as well as

effects of different evaluation metrics (AUC vs TSS). The SDMs are machine learning algorithms that can project certain species into the area where they are not occurring even though the area has an optimal environmental niche. To avoid it, expert knowledge has to be applied in the post-processing step to create zooplankton species range maps (similar to the ones available for many fish species).

Another big limitation of this study is the lack of biological information on the habitat and depth range of many mesopelagic species. I acknowledge that my own classification of the organism into meso- or epipelagic groups is coarse and is limited by the amount of available data in MMM Database. For instance, if sampling were only done in the upper layer of the water column, the species would be classified as epipelagic even though it may be present at mesopelagic depths too. In addition, the boundary of the mesopelagic zone is dynamic and can change on seasonal and yearly timescales (Reygondeau *et al.*, 2018). Moreover, the spatial extent of each mesopelagic zone can vary with time. This study used a static position of the mesopelagic provinces that can both affect the classification of the organism into meso/epipelagic species but also affect the distribution of the organisms, especially when modelling the native range distributions. This work also has shown the important role of taxonomy and its influence on SR and rarity estimates. Finally, constant advances in updating the species taxonomy can lead to the overestimated rarity of the region/cell.

### **3.5 Concluding remarks**

This study found that the ensemble approach is the optimal way to model the distribution patterns of mesozooplankton, although it is a time-consuming process. To make the process faster, RF models can be used as they show similar predictive performance to the ensemble models but are computationally faster. However, the ensemble technique has several advantages, including reducing overfitting, improving generalization, and being more flexible and robust to variations in the data. I also explored two important factors that can affect model performance: subsampling and the choice of pseudo-absences. I found that subsampling can have both positive and negative effects on model performance and that careful selection of pseudo-absences is crucial for effective modelling.

The validity of global diversity patterns found in this study can be a result of the robust modelling approach employed. For example, the Atlantic Ocean northern and southern hemispheres were differentially sampled. However, discovered macroecological patterns exhibit symmetry with respect to the equator. This is a testimony to the modelling approach employed because discovered patterns appear to hold at vastly different levels of the data coverage. While further model tuning and data collection are required, employing this approach in regions with limited sampling would allow inferring true macroecological patterns in the world ocean.

This study has shown that the distribution of mesopelagic mesozooplankton organisms is primarily driven by salinity, dissolved nitrate, euphotic zone depth, and dissolved oxygen. However, the importance of these variables varies depends on the phylum and species studied. These findings highlight the importance of considering the ecological requirements of different species when predicting their distributions. Noteworthy, unlike the epipelagic layer, water temperature was not the most important variable for modelling species distribution. Although this study provides new insights into the environmental drivers of species distribution in mesopelagic, other factors such as biotic interactions and historical events may also play a role in determining species distributions. Therefore, further research at the species level is necessary to gain a more detailed and accurate understanding of these environmental drivers.

This study is the first attempt to model the MM biodiversity patterns in a global ocean, using a subset of 861 species. The results show that the majority of modelled species were from the phylum Arthropoda, reflecting the dominance of copepods in the mesopelagic community. Other taxa records, such as Chaetognatha, Chordata, and Mollusca are severely undersampled in many mesopelagic areas. The study revealed interesting patterns in the global diversity of MM, with the highest species richness observed in the North Atlantic region, north of the South American continent, along the west coast of India, and the Mediterranean Sea. The LDG analysis showed a bimodal distribution of species richness, with peaks around 30° latitude in both hemispheres and tailing off towards the poles. Such observations are similar to biodiversity patterns from the epipelagic zone indicating a tight link between surface and deep ocean. In addition, I found that biodiversity hotspots in the mesopelagic were only at lower latitudes while polar areas had fewer species. However, the distribution of species at polar

areas were more uniform resulting in high mean species richness values per latitude while areas between 30° latitudes had variable species richness values per cell, resulting in lower mean species richness values. The study also found a relationship between mesopelagic water temperature and species richness, with the highest values observed at water temperatures of 20°C, found only in semi-enclosed mesopelagic basins or areas closest to the coast in equatorial and tropical regions.

The total range-size rarity showed a similar pattern to species richness, with high values found in the offshore North Atlantic and coastal areas across both hemispheres. The results showed that total rarity was higher in polar areas with the highest values observed in the Northern Hemisphere. The lowest rarity was predicted in gyres of the Pacific Ocean. After combining the metrics of rarity and species richness, the North Atlantic, Arctic, and Southern Ocean were identified as areas of high rarity and high species richness, while the Pacific and Indian ocean basins were identified as areas with low rarity and low species richness. The study suggests that these patterns of rarity may be influenced by a range of factors, such as the diversity of habitats and niches, environmental conditions, and sampling effort.

Overall, this study serves as a valuable starting point for predicting the distribution of marine species, and its findings can be used to inform conservation and management efforts in marine ecosystems. However, more sampling is needed to fully understand the diversity and distribution of mesopelagic species and to address the discrepancies observed between the different phyla.

# Chapter 4: Global Estimate of Mesopelagic Mesozooplankton Biomass

## 4.1 Introduction

Mesozooplankton are a pivotal component of marine trophic webs, including organisms ranging in length from 0.2 to 20 mm (Sieburth *et al.*, 1978) and consisting primarily of crustacean plankton (copepods, amphipods and euphausiids), meroplanktonic larva, rhizaria, and smaller individual gelatinous zooplankton (Biard *et al.*, 2016; Moriarty & O'Brien, 2013). Mesozooplankton play a crucial role in pelagic ecosystems by serving as both consumers of primary productivity and prey for higher trophic levels, influencing energy flow, carbon sequestration, and nutrient cycling within pelagic ecosystems (Steinberg & Landry, 2017).

The mesopelagic zone, also known as the twilight zone, is a layer of ocean that extends from a depth of 200 to 1,000 m below the surface. This zone is characterized by low light levels and is home to a diverse community of organisms, including mesozooplankton. Despite the importance of mesopelagic zooplankton in oceanic food webs and the role in biochemical cycling, global estimates of its biomass are scarce. Most studies of this group have been spatially limited (Dornelas *et al.*, 2018; Weikert *et al.*, 2001; Nishikawa *et al.*, 2007). Although some estimates of ocean basin biomass of specific mesozooplankton groups have been obtained, they were either derived from a single transect (Vereshchaka *et al.*, 2016) or based on only a few studies (McIvor, 2011). Vereshchaka *et al.* (2016) estimated the standing stock (wet weight, WW) of the zooplankton community in the Atlantic Ocean to be 70 Mt, with the mesopelagic zone contributing between 13-16% or 9.1-11.2 Mt. The authors also discovered a strong relationship between surface chlorophyll concentrations and zooplankton biomass that is consistent with theoretical expectations of surface and deep-sea layer's connectivity (Vereshchaka *et al.*, 2016). The first available global ocean maps of zooplankton biomass are outdated and based on limited data, hand-drawn, and cover only the top 100 m of the epipelagic layer (Bogorov *et al.*, 1968; Reid, 1962). A recent study by Strömberg *et al.* (2009) addressed this issue by developing a model that relates the flow of energy from primary production to zooplankton biomass. The authors provided a map of the global distribution of modelled net zooplankton in the epipelagic zone (0-200 m) and report the mean global zooplankton biomass as  $5.52 \pm 8.94 \text{ mgC}\cdot\text{m}^{-3}$ .

There are indeed several recent mesozooplankton standing stock assessments that took advantage of summarized epipelagic biomass estimates across the globe (Moriarty & O'Brien, 2013; Buitenhuis *et al.*, 2013; Hatton *et al.*, 2021). However, these estimates are difficult to compare owing to differences in the reported units, variable sampling techniques, and gear deployed (see Table 4.1). The obtained values were standardized to PgC using either the conversion factor from Raymond (1984) or the estimated volume of the mesopelagic zone from Reygondeau *et al.* (2018). Moriarty & O'Brien (2013) estimated the global mesozooplankton biomass in the epipelagic zone to be 0.19 PgC, with the highest concentration in the Northern Hemisphere and a slight decrease in biomass from polar to temperate regions in both hemispheres. Buitenhuis *et al.* (2013) provided estimates of mesozooplankton biomass densities identical to those reported by Moriarty & O'Brien (2013), while Hatton *et al.* (2021) reported an estimate of 41 Gt of WW or roughly 0.49 PgC. None of the studies provided information on the uncertainty of the global estimates, making it impossible to compare these estimates accurately.

Several studies have attempted to investigate the mesozooplankton biomass in the mesopelagic zone (Buitenhuis *et al.*, 2013; Drago *et al.*, 2022; Hernández-León *et al.*, 2020). However, these studies generally covered mesopelagic depths down to 500 m (i.e., the upper mesopelagic zone). Drago *et al.* (2022) estimated the global biomass distribution of 19 zooplankton taxa at roughly 3,500 stations using *in situ* imaging observations and machine learning techniques. The results showed that zooplankton biomass was the highest in polar regions and at the equator, and the lowest in oceanic gyres. The global integrated biomass in the epipelagic and upper mesopelagic zone (0-500 m) was estimated to be 0.403 PgC, with Copepoda (35.7%) and Eumalacostraca (26.6%) being the most abundant groups, with the upper mesopelagic zone accounting for 0.173 PgC (Drago *et al.*, 2022). Buitenhuis *et al.* (2013) reported biomass values in the range 0.33-0.59  $\mu\text{gC}\cdot\text{L}^{-1}$  that was approximated to 0.14-0.25 PgC (Table 4.1).

Table 4.1: Global mesozooplankton biomass estimates in the upper pelagic ocean. Abbreviations: std - standard deviation; min - minimum value; max - maximum value; ww- wet weight.

Part of the Ocean	Depth Range (m)	Zooplankton Size class	Biomass Estimates	Reference
Epipelagic	0-200	150-650 $\mu\text{m}$ mesh: 0.11-0.49 mm <sup>a</sup>	0.19 PgC (based on median) mean 5.9 $\mu\text{gC}\cdot\text{L}^{-1}$ , median 2.7 $\mu\text{gC}\cdot\text{L}^{-1}$ std 10.6 $\mu\text{gC}\cdot\text{L}^{-1}$	(Moriarty & O'Brien, 2013)
		0.2-20 mm	0.19 PgC (based on median) mean 6 $\mu\text{gC}\cdot\text{L}^{-1}$ , median 2.7 $\mu\text{gC}\cdot\text{L}^{-1}$ std 10.6 $\mu\text{gC}\cdot\text{L}^{-1}$	(Buitenhuis <i>et al.</i> , 2013)
		0.2-20 mm	0.49 PgC <sup>c</sup> 4.1 Gt ww	(Hatton <i>et al.</i> , 2021)
		200, 300 and 500 $\mu\text{m}$ mesh: 0.15- 0.38 mm <sup>a</sup> < 5 mm (reported in the paper) <sup>b</sup>	0.48 PgC	(Hernández-León <i>et al.</i> , 2020)
		150-200 $\mu\text{m}$ Standardized to 330 $\mu\text{m}$ mesh: 0.11-0.25 mm <sup>a</sup>	Tuned to global (region): Mean 5.52 (7.29) $\text{mgC}\cdot\text{m}^{-3}$ Median 3.32 (4.97) $\text{mgC}\cdot\text{m}^{-3}$ Std 8.94 (9.90) $\text{mgC}\cdot\text{m}^{-3}$ 0.21 (0.32) PgC (based on median) <sup>d</sup> 0.42 PgC (calculated based on data) <sup>e</sup>	(Strömberg <i>et al.</i> , 2009)
	0-100	1-20 mm	Mean 100-200 $\text{mg m}^{-3}$ ww	(Bogorov <i>et al.</i> , 1968)

Epipelagic & Upper Mesopelagic	0-500	0.2-20 mm	0.05-0.09 PgC <sup>d</sup> min 0.33 $\mu\text{gC}\cdot\text{L}^{-1}$ , max 0.59 $\mu\text{gC}\cdot\text{L}^{-1}$	(Buitenhuis <i>et al.</i> , 2013)
		1.02-50 mm	0.403 PgC	(Drago <i>et al.</i> , 2022)
Upper Mesopelagic	200-500	1.02-50 mm	0.173 PgC	(Drago <i>et al.</i> , 2022)
Mesopelagic	200-1000	200, 300 and 500 $\mu\text{m}$ mesh: 0.15-0.38 mm <sup>a</sup> < 5 mm (reported in the paper) <sup>b</sup>	0.66 PgC	(Hernández-León <i>et al.</i> , 2020)
Full Water Depth	Integrated	0.2-20 mm	1.31 PgC <sup>c</sup> 10.9 Gt ww	(Hatton <i>et al.</i> , 2021)

<sup>a</sup> Zooplankton size was calculated using Nicholas and Thompson (1991)'s  $\frac{3}{4}$  law of mesh selection.

<sup>b</sup> Size reported in the paper excluded macrozooplankton and micronekton data from the analysis.

<sup>c</sup> Using 1 mg wet weight to 0.12 mg C conversion from Raymond (1984).

<sup>d</sup> Converted to PgC using the volume of the mesopelagic zone of 159,276,913 km<sup>3</sup> or epipelagic 63,710,765 km<sup>3</sup>.

<sup>e</sup> The original dataset was provided by the authors.



Hernández-León *et al.* (2020) was the only study that examined the distribution of mesozooplankton biomass in the mesopelagic zone (200-1,000 m; Figure E10). This study found a strong positive link between biomass in the dark ocean and the average epipelagic primary production. The authors estimated the total oceanic mesozooplankton biomass to be 1.4 Pg C, with 47% (or 0.66 PgC) located in the mesopelagic layer and 34% (or 0.48 PgC) in the epipelagic layer (Hernández-León *et al.*, 2020). This estimate assumes that the mesopelagic zone depth range is uniform across the globe. However, Reygondeau *et al.* (2018) proposed that the vertical division of the zone is not constant over the global ocean but varies between ocean basins and latitudes. Thus, each location requires the application of a dynamic depth range of mesopelagic provinces (*sensu* Reygondeau *et al.*, 2018) to obtain quantitative mesopelagic biomass values.

Zooplankton biomass can be influenced by a variety of environmental variables, including temperature (Dalpadado *et al.*, 2003; Drago *et al.*, 2022), primary productivity (Vereshchaka *et al.*, 2016; Hernández-León *et al.*, 2020), and predatory pressure (Lampert, 1962). In this study, I focused on Net Primary Production (NPP) as it is an important factor influencing zooplankton biomass. Net Primary Production (NPP) is an ecological term used to describe the amount of organic matter produced by photoautotrophs, excluding respiration costs. It measures the rate at which energy is converted from the sunlight into the chemical energy through photosynthesis and is typically expressed in units of mass per unit area per unit time (e.g., grams of carbon per square meter per day). NPP is an important measure of the capacity of an ecosystem to support biomass because it reflects the amount of organic matter available to support the growth and reproduction of heterotrophic organisms. As a result, NPP is often used to measure the productivity of the upper layers of the ocean or epipelagic zone.

However, not all NPP are available for export to mesopelagic zone. In central gyres, only a small fraction (1-10%) of NPP is exported from the euphotic zone, whereas in polar regions, a significantly larger portion (30-100%) of NPP may be exported (Buesseler, 1998). Additionally, the transport of particulate organic carbon (POC) to deeper depths is more significant in productive continental margins than in the central gyres. Another way to assess the amount of organic matter available in the mesopelagic region is to use POC standing stock. POC measures the amount of organic carbon present in particulate form and is typically measured in units of mass per unit volume (e.g., milligrams per cubic meter). POC can be produced from various sources, including

primary productivity, decomposition of organic matter, and sinking of organic matter from the ocean surface layers (Eppley & Peterson, 1979). POC is important because it can serve as a source of nutrition for other organisms, and it plays a role in carbon cycling throughout the ecosystem. As a result, the POC amount may provide a more direct proxy for the zooplankton biomass, because POC represents the amount of organic matter available for zooplankton consumption, whereas NPP represents the production of organic matter.

Investigating the relationship between mesozooplankton diversity and biomass in the mesopelagic zone is of critical importance. Despite its critical role for global climate regulation by sequestering carbon dioxide through the biological carbon pump, the mesopelagic zone remains a largely unexplored and under-studied part of our oceans (Robinson *et al.*, 2010; Proud *et al.*, 2017; St. John *et al.*, 2016). Understanding the correlation between mesozooplankton diversity and biomass in the mesopelagic zone could reveal important insights into how changes in species diversity affect carbon sequestration that can be used to improve global climate change models. Furthermore, the diversity-biomass relationship (Guo, 2007; Marrs *et al.*, 1996) is informative in understanding the ecosystem stability, resilience, and the overall health of the mesopelagic zone.

The objective of this study was to determine the global mesozooplankton biomass in the mesopelagic zone using POC and NPP estimates, and to compare these findings with previously published data. Furthermore, the study aimed to (a) examine the correlation between the biomass of both epipelagic and mesopelagic zooplankton, and (b) explore the relationship between mesozooplankton diversity and biomass in the mesopelagic zone.

## **4.2 Methods**

All statistical analyses, data manipulations, and visualizations were performed using R Statistical Software (R Core Team, 2022).

### ***4.2.1 Biomass data***

I used data from the 2010-2011 Malaspina cruise, as recorded by Hernández-León *et al.* (2020). Data were collected using a consistent sampling method across the entire sampling area to minimize variations in biomass estimates. The study covers a vast area ranging from 40°S to 30°N. The Malaspina cruise's zooplankton biomass (in  $\text{mgC}\cdot\text{m}^{-2}$ ) was downloaded from PANGAEA

(<https://doi.org/10.1594/PANGAEA.922974>). To allow for easier comparison of different areas on a global scale, biomass data were gridded on Map of Life (MOL; <https://mol.org/>) equal-area grid with a per cell area of 3091 km<sup>2</sup>. In addition, I used the mesopelagic depth boundary estimated by Reygondeau *et al.* (2018) and projected onto the same equal-area grid.

The original data contained three depth intervals (0-200, 200-1,000 and 1,000-2,000 m) where the 200-1,000 m interval was attributed to the mesopelagic realm. However, Reygondeau *et al.* (2018) demonstrated that mesopelagic depth varies globally, with the epi-mesopelagic boundary ranges between 3-150 m (mean 42 m) and meso-bathypelagic boundary ranging from 4 to 4197 m (mean 1358 m). Thus, the actual mesopelagic zone depth can overlap with all three sampling intervals from Hernández-León *et al.* (2020). To determine which of the sampling depth intervals were located within the mesopelagic zone, I calculated the overlap (in meters) between the sampling depth (0-200, 200-1,000 or 1,000-2,000 m) and the depth bin of the mesopelagic zone, according to Reygondeau *et al.* (2018).

#### 4.2.2 *Environmental Data*

NPP was downloaded from 1080 by 2160 Monthly HDF files from MODIS R2018 Data using the Standard Vertically Generalized Production Model (VGPM; Behrenfeld & Falkowski, 1997) for July 2002-December 2021 ([http://orca.science.oregonstate.edu/npp\\_products.php](http://orca.science.oregonstate.edu/npp_products.php)). The seasonal climatology data for POC were downloaded from NASA Earth Data (<https://oceancolor.gsfc.nasa.gov/l3/order/>) Standard AQUA-MODIS POC 9 km (mapped) for the period 2002-2022. Then the climatological means for NPP and POC were computed for each cell. The data were then projected onto an equal-area grid, similar to the gridding of biomass data. The distribution of environmental variables is shown in Figure E1.

#### 4.2.3 *Models*

Two types of models were constructed. A linear model (estimated using OLS) was fitted to predict the biomass using NPP or POC data (formula:  $\log(\text{Biomass}) \sim \log(\text{NPP})$  or  $\log(\text{Biomass}) \sim \log(\text{POC})$ ).

The mesopelagic zone depth can overlap with all three sampling intervals reported by Hernández-León *et al.* (2020), I only used biomass estimates from depth intervals, where 80% of the interval

overlapped with the dynamic mesopelagic zone depth. Overall, eight from 0-200 m, 42 from 200-1,000 m, and nine from 1,000-2,000 m intervals were used in this study. Since the depth of the mesopelagic zone in each cell differed from the depth bin from Hernández-León *et al.* data, a corrected biomass estimate was used to account for the variable depth of the mesopelagic zone. Thus, the original areal biomass units ( $\text{mg C}\cdot\text{m}^{-2}$ ) were first converted to volumetric units ( $\text{mg C}\cdot\text{m}^{-3}$ ) using the corresponding sampling depth bins (0-200, 200-1,000 or 1,000-2,000 m). The units were then converted back to areal estimates using the variable depth bin of the mesopelagic zone ( $\text{mg C}\cdot\text{m}^{-2}$  mesopelagic). Since these models use data from the variable depth of the mesopelagic zone based on latitude and longitude to model biomass, they will be referred to as ‘Variable Depth Models’ hereafter.

For comparison, I ran the same models but used the original biomass estimates for 200-1,000 m depth from Hernández-León *et al.* (2020) to represent global MM (further referred as MM) biomass using the classic definition of the mesopelagic zone. These models will be referred to as ‘Classical Depth Models.’ Assumptions of each linear model were checked with diagnostic plots provided in Figure E2. The 95% confidence intervals (CIs) and p-values were computed using a Wald t-distribution approximation (Barr, 2013).

#### 4.2.4 Projection into mesopelagic zone

I utilized the annual climatological data for POC and NPP to make predictions of mesozooplankton biomass on a global scale using the linear models I created. To ensure that the predictions corresponded only to the mesopelagic zone, coordinates were filtered using the mesopelagic zone boundaries established by Reygondeau *et al.* (2018). This allowed us to estimate zooplankton biomass within the mesopelagic zone based on a new set of environmental conditions. Global biomass estimates were expressed in PgC, with the standard error reported.

The global estimate of the mesozooplankton biomass ( $\hat{p}$ ) was calculated as the sum of the predicted biomass values in each cell:

$$\hat{p} = \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) = n\hat{\beta}_0 + S_x \hat{\beta}_1$$

where  $x_1, \dots, x_n$  are new set of values of the explanatory variable,  $\widehat{\beta}_0$  and  $\widehat{\beta}_1$  are coefficient estimates, and  $\widehat{y}_i$  is the predicted value for given  $x_i$ .  $S_x$  is a simplified notation for  $\sum_{i=1}^n x_i$ . The standard error of this prediction was calculated using the following formula:

$$Var(\hat{p}) = Var(n\widehat{\beta}_0) + Var(S_x\widehat{\beta}_1) + 2Cov(n\widehat{\beta}_0, S_x\widehat{\beta}_1)$$

Since Hernández-León et al. (2020) data only cover the area between 40°S and 40°N, I also calculated a global MM estimate for that particular band. However, biomass estimates for the entire global region can be considered adequate, as the range of environmental variables used to build the models is a representation of global ranges (Figure E3).

#### 4.2.5 Biomass and Diversity

To explore the relationship between MM biomass and diversity, I matched the predicted biomass estimates from the variable depth POC model with the potential species richness computed using species distribution models (see Chapter 3). I used 2D maps of MM biomass and species richness to explore the global patterns of species diversity and biomass along the corresponding latitudinal gradients.

The total and average range-size rarities per cell were calculated using the species richness data. The total range-size rarity, also referred to as weighted endemism (Hobohm, 2014), was calculated as the proportion of the distribution of a species found in a cell summed across all species (Pollock *et al.*, 2017). I then explored the relationship between the estimated biomass and total range-size rarity. The total range-size rarity is correlated with species richness; therefore, the average range-size rarity was calculated as a more appropriate measure of rarity (Crisp *et al.*, 2001). Average range-size rarity was determined by dividing the total range-size rarity by the number of species in each cell. I then examined the relationship between global and per-basin biomass and the average range-size rarity.

#### 4.2.6 Link Between Biomass in Epi- and Mesopelagic Zones

I used a global estimate of mesopelagic biomass reported from various sources (Table 4.1), three estimates of mesopelagic biomass (one reported by Hernández-León *et al.* (2020), and two estimates from a POC-based model from this study) to compute the global ratio between the epipelagic and mesopelagic biomass.

Two maps were utilized to study the relationship between MM biomass and total zooplankton biomass in the epipelagic layer. The first was the global map of zooplankton distribution by Bogorov *et al.* (1968), which was digitized as the original data were unavailable (Figure E6D). The second map (Figure E6A) was constructed based on data provided by Stormberg *et al.* (2009). Data were available through the NOAA portal: <https://www.st.nmfs.noaa.gov/plankton/biomass/>.

Originally, biomass estimates in the Bogorov *et al.* map were binned into six intervals: <25, 25-50, 51-100, 101-200, 201-500 and >500 mg·m<sup>-3</sup> (Figure E5), so I converted biomass bins to numeric estimates by taking a mid-value of the bin as an estimate of zooplankton biomass in that cell (for bin >500 mg·m<sup>-3</sup>, I used a 500-1,000 mg·m<sup>-3</sup> range and a mean value of 750 mg·m<sup>-3</sup>). Then, the values were converted to mgC using a conversion factor of 0.12 (Raymont, 1984) and were further multiplied by a depth bin of 100 m to convert them to areal estimates. For Strömberg *et al.* (2009), units were already given in carbon weight; therefore, only conversion to areal units was performed by multiplying the values by 200 depth bin. The values were then gridded on an equal-area grid. Both maps were matched with the mesozooplankton biomass estimated using the classical-depth POC model. Although the variable depth biomass estimates are better representation of mesopelagic biomass, we used biomass estimates for 200-1,000 m depth layer in order to avoid the overlap between epipelagic biomass estimated by Strömberg *et al.* (2009). The ratio between epipelagic and mesopelagic biomass was calculated for each cell. In addition, the global biomass of epipelagic was computed for each map.

The comparison of MM biomass to mesopelagic fish biomass can only be performed at a regional level, as there is limited information available on a global scale. To compare the distribution patterns of mesopelagic biomass of fish and mesozooplankton, I utilized the data from Clavel-Henry *et al.* (2020) for the Mediterranean Sea. Only visual comparisons were made. Additionally, the mesopelagic zone was divided into 10 clusters using k-means clustering based on the obtained biomass and species richness values, and was visually compared for consistency with Proud *et al.* (2017), who established mesopelagic provinces based on the depth and intensity of the deep scattering layer as a proxy for mesopelagic fish biomass.

## 4.3 Results

### 4.3.1 Models

I fitted a linear model (estimated using OLS) to predict MM biomass with NPP data (formula:  $\log_{10} Bio \sim \log_{10} NPP$ ) using the data from variable mesopelagic depth (Figure 4.1A). The model explained a moderate proportion of variance ( $R^2 = 0.24$ ). The effect of NPP (logged-transformed) was statistically significant and positive ( $\beta = 2.08$ , 95% CI [0.92, 3.24],  $t(42) = 3.61$ ,  $p < .001$ ). The same model was applied to data from a 200-1,000m depth (Figure 4.1C). The model explained a moderate proportion of variance ( $R^2 = 0.18$ ). The effect of NPP (logged-transformed) was statistically significant and positive ( $\beta = 1.62$ , 95% CI [0.53, 2.70],  $t(40) = 3.01$ ,  $p = 0.005$ ). Model coefficients and diagnostics are shown in Tables E1-2.

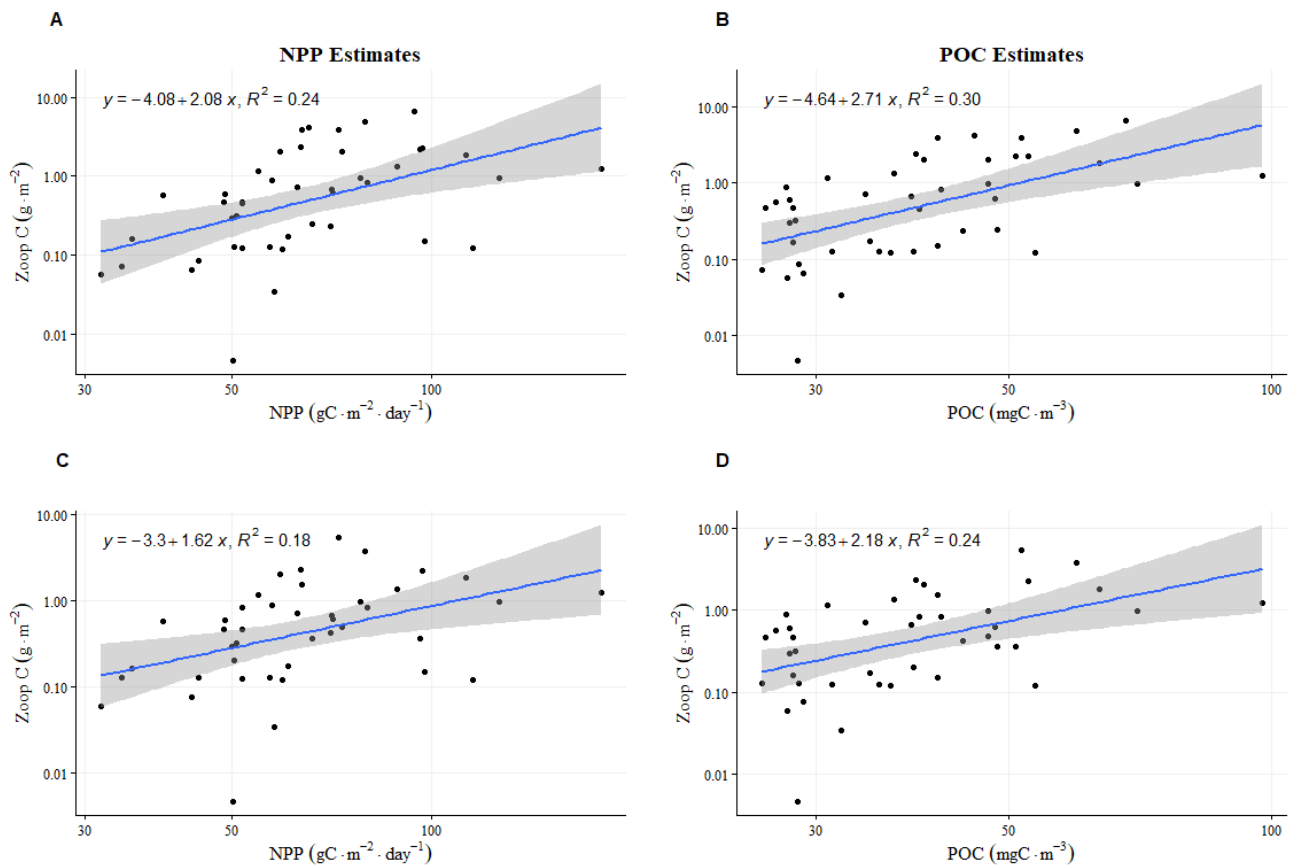


Figure 4.1: Relationships between NPP (1<sup>st</sup> column) or POC (2<sup>nd</sup> column) and mesopelagic mesozooplankton biomass for Variable Depth Models (1<sup>st</sup> row) and Classical Depth Models (2<sup>nd</sup> row). Solid blue lines represent regression lines, and grey bars are standard errors. Note that both axes are on a logarithmic scale.

I fitted a linear model (estimated using OLS) to predict MM biomass with POC (formula:  $\log_{10} Bio \sim \log_{10} POC$ ; Figure 4.1B). The model explained a substantial proportion of the variance ( $R^2 = 0.30$ ). Within this model, the effect of logged transformed POC (logged-transformed) was statistically significant and positive ( $\beta = 2.71$ , 95% CI [1.43, 3.99],  $t(42) = 4.27$ ,  $p < .001$ ). A similar formula was used, but for data within the 200-1,000 m depth range (Figure 4.1D), the model explained a moderate proportion of variance ( $R^2 = 0.24$ ), which was slightly lower than that of the model based on estimates from the variable mesopelagic depth bins. The effect of POC (log-transformed) was statistically significant and positive ( $\beta = 2.18$ , 95% CI [0.93, 3.43],  $t(40) = 3.52$ ,  $p = 0.001$ ). A summary of the performance of all the models is listed in Tables E1-2.

#### 4.3.2 Projection into the mesopelagic zone

Overall, all approaches showed similar distributional patterns of MM biomass (Figure 4.2) with enhanced values in the northern hemisphere, west coast of the continents, and equatorial and 50°S bands. Low biomass values were found in the centre of the ocean gyres. However, NPP-derived biomass values decreased at higher latitudes (Figure 4.2 A and E) while POC-derived values increased towards the higher latitudes in the Southern hemisphere and stayed constant in the Northern Hemisphere (Figure 4.2 C and G). Latitudinal gradients for NPP-derived biomass values shows a greater range of values with a sharp decrease in biomass values towards higher latitudes (Figure 4.2 B and F). POC and NPP-based values show a similar pattern between the 60°N-60°S with low biomass values at 30°S and 30°N with a slight gradual increase in the equatorial area. Biomass values gradually increase from 30° to the higher latitudes.

Estimates of global MM biomass in the mesopelagic zone ranged between 0.20-0.91 PgC depending on the method (Table 4.2). Using variable depth models, global biomass was estimated to be  $0.29 \pm 0.06$  PgC using NPP-derived estimates and three-fold higher values of  $0.91 \pm 0.08$  PgC using the POC-based estimates. Using the classical definition of the mesopelagic zone (200-1,000 m), NPP-based values were lower than the POC-based values ( $0.20 \pm 0.06$  PgC and  $0.45 \pm 0.08$  PgC, respectively). Majority of biomass concentrated between 40°S-40°N. NPP-derived biomass estimates between 40°S-40°N comprised 68-70% of the global biomass estimates, while POC-derived biomass estimates between 40°S-40°N comprised only 51-55% of the global biomass estimates.



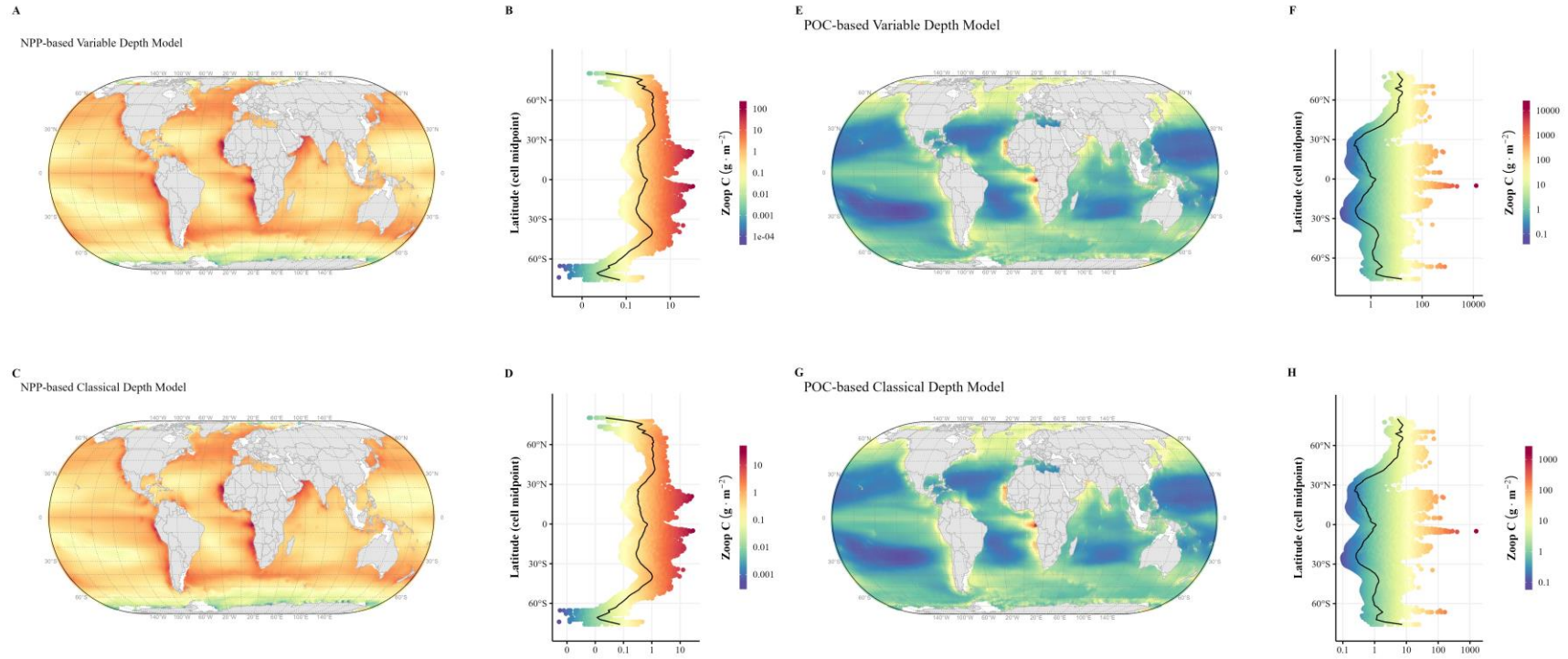


Figure 4.2: NPP- (1<sup>st</sup> column) and POC- (2<sup>nd</sup> column) based estimates of mesopelagic mesozooplankton biomass in the global ocean using variable depth models (1<sup>st</sup> row) and classical depth models (2<sup>nd</sup> row). Latitudinal biomass gradients for each model are shown in plots B, D, F and H. Solid black line in latitudinal graphs are the mean biomass value per latitude. All maps in Eckert IV global equal area projection.

Table 4.2: Estimated global and 40°S-40°N means and standard errors of mesopelagic mesozooplankton biomass in the pelagic ocean based on two modelling approaches (NPP and POC-based models).

Model	ENV	Global (PgC)	40°S-40°N (PgC)
Classical Depth	NPP	0.20 ± 0.06	0.14 ± 0.04
Variable Depth	NPP	0.29 ± 0.06	0.20 ± 0.05
Classical Depth	POC	0.45 ± 0.08	0.23 ± 0.04
Variable Depth	POC	0.91 ± 0.08	0.50 ± 0.04

### 4.3.3 *Biomass and Diversity*

I used biomass estimates derived from the variable depth POC model to explore the relationship between biomass and diversity. Average biomass values per mesopelagic province are shown in Figure E4.

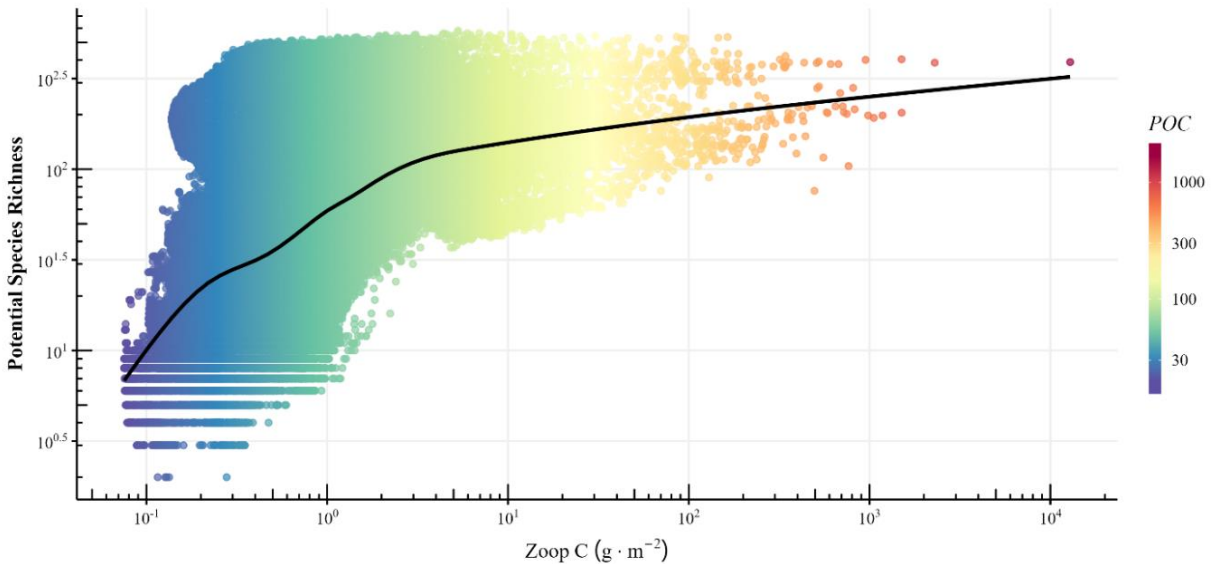


Figure 4.3: Global relationship between mesopelagic zooplankton biomass ( $\text{mgC} \cdot \text{m}^{-2}$ ) and potential species richness colored by the POC value for each cell. Note the log10 transformation on the x and y axes and colour scale. The black line is a fitted gam line (formula =  $y \sim s(x, bs = "cs")$ ).

Biomass was found to be related to the species richness (Figure 4.3). When both values are log<sub>10</sub>-transformed, the data follows the logarithmic trend: a rapid increase in species richness at the lower biomass values followed by a slower change in species richness when biomass values increase. Lower values of biomass and species richness were also associated with areas of low POC.

High biomass and species richness values were observed throughout the Arctic, North Pacific, coastal and equatorial areas, and Southern Ocean (Figure 4.4 A). Low biomass and low species richness regions were associated with centers of ocean basins in the North, South Pacific, and Indian Oceans. However, the center of the North Atlantic basin was associated with low biomass and high species richness values. Low biomass and high species richness were also recorded in the Mediterranean Sea and western boundaries of the subtropical gyres. Areas of high biomass but low species richness were in subtropical regions (both sides from the equatorial band), around the gyre edges, and some areas in the North Pacific and the 50°S band.

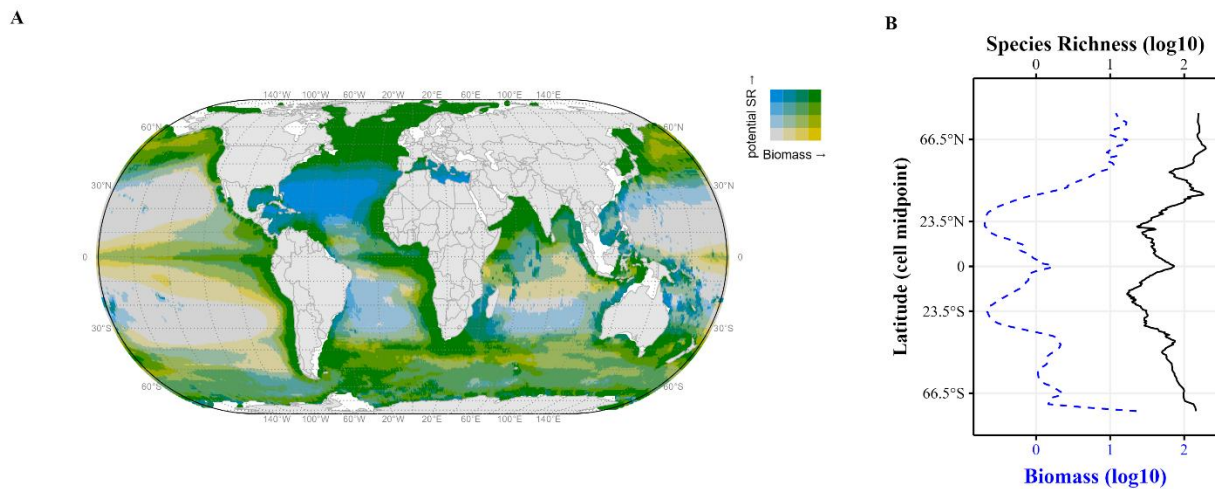


Figure 4.4: A) 2D map depicting the relationship between the mesopelagic mesozooplankton biomass derived from POC estimates and the potential species richness from Chapter 3. Quantile breaks in a legend were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. The map in Eckert IV global equal area projection. B) Latitudinal gradient of log10 transformed median biomass (blue dashed line) and median species richness (solid black line).

The latitudinal pattern of median MM biomass resembled the one for species richness (Figure 4.4 B) but with a larger magnitude of change. The lowest species richness and biomass values were found around 23.5°N and 23.5°S, gradually increasing towards the equator. In the Northern Hemisphere, the biomass gradually increased throughout the temperate region and stabilized at high values in the polar region. Species richness showed a similar increase in the northern temperate region, but the increase was not gradual, with some rises and declines around 45-50°N. In the Southern Hemisphere, biomass and species richness values also increased throughout the temperate region with a small peak around 30°S. However, biomass in the southern polar region

did not stabilize and kept increasing towards the pole, while species richness showed a moderate increase in polar areas.

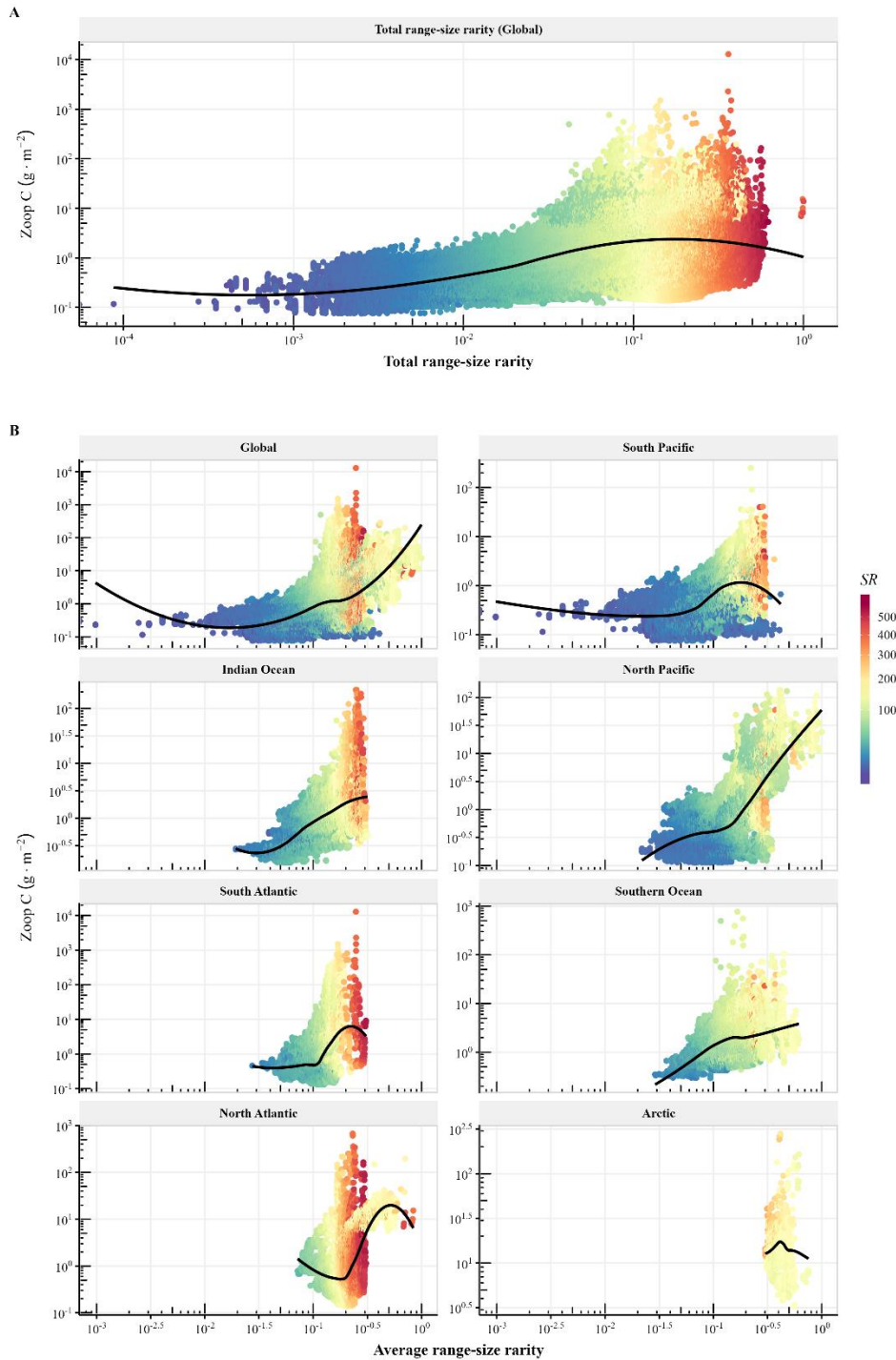


Figure 4.5: Relationship between the mesopelagic mesozooplankton biomass and various diversity proxies. A) The relationship between biomass and total range-size rarity colored by species richness. B) relationship between the biomass average range-size rarity colored by species richness

aggregated by Global and basin-based scales. Note that the color bar in A and B are square-root-transformed. The black line is fitted gam line (formula =  $y \sim s(x, bs = "cs")$ ).

A skewed hump-back relationship is observed between  $\log_{10}$ -transformed MM biomass and the total-range size rarity (Figure 4.5A). Biomass values remained low until the total rarity reached  $\sim 0.2$  value (two in ten species are rare). The biomass value peaked at about  $\sim 0.2$  rarity and then declined. As expected, species richness was highly and positively correlated to total rarity. The average rarity for each cell was calculated to remove the effect of species richness on the total rarity (Figure 4.5B). Biomass was positively associated with the average rarity, with biomass increasing as rarity increases. The variability of biomass estimates appeared to increase with the rise in average rarity scores. A peak in biomass values was observed at average rarity scores of 0.3 ( $10^{-0.5}$ ), which in turn was associated with the highest species richness values. Biomass, in general, was lower at high average rarity scores. The pattern between biomass and average rarity was evident in almost all ocean basins except for the Arctic Ocean, where only cells with high average rarity were found (Figure 4.5).

#### 4.3.4 Link Between Biomass in Epi- and Mesopelagic Zones

The global ratio of epipelagic to mesopelagic biomass varied greatly due to very different estimates of zooplankton biomass in the epipelagic layer (Table 4.3). The smallest estimate of epipelagic biomass of 0.14 PgC was calculated from a digitized Bogorov's map. However, the estimate was based only on the top 100 m layer. Other epipelagic estimates were higher (0.19-0.49 PgC) and were reported from the 0-200 m depth bin. Estimates of the mesopelagic biomass were also variable (0.45-0.91 PgC). First, the variability came from the different classifications of the mesopelagic depth. Two estimates were available using the classical definition of mesopelagic zone 0.66 PgC from Hernández-León *et al.* (2020) and the POC-based estimate of 0.45 PgC from this study. However, I also added a POC-based estimate of mesopelagic biomass using a variable depth, as I consider it a better representation of the biomass in the mesopelagic.

Using Hernández-León *et al.* (2020) estimates of mesopelagic biomass, the ratios ranged between 0.32-0.74 (mean  $0.53 \pm 0.18$  and median 0.48). I excluded the Bogorov *et al.* (1968) ratio from calculations since the authors only reported values from the top 100 m (but see below). When classical depth POC-model estimates were used, the ratios were slightly higher and ranged

between 0.47-1.01 (mean  $0.76 \pm 0.24$ , median 0.71). When variable depth estimates were used, the ratios were the lowest and ranged between 0.21 and 0.53 (mean  $0.38 \pm 0.13$ , median 0.35).

Table 4.3: Epipelagic to mesopelagic biomass ratio based on different estimates of the epipelagic and mesopelagic biomasses found in literature and calculated in this study. Mean and medial ratios are estimated for each column.

Study	Epipelagic Biomass (PgC) 0-200 m	Mesopelagic Biomass		
		200-1,000 m		Variable Depth
		0.66 PgC (Hernández- León et al., 2020)	0.45** PgC (this study)	0.91** PgC (this study)
(Hernández-León <i>et al.</i> , 2020)	0.48	0.72	1.01	0.52
(Hatton <i>et al.</i> , 2021)	0.49	0.74	1.01	0.53
(Moriarty & O'Brien, 2013) and (Buitenhuis <i>et al.</i> , 2013)	0.19	0.28	0.42	0.21
(Strömberg <i>et al.</i> , 2009)	0.21	0.32	0.47	0.23
- tuned to global	0.32	0.48	0.71	0.35
- tuned to region	0.42	0.64	0.93	0.46
- based on data				
(Bogorov <i>et al.</i> , 1968)	0.14*	0.21	0.31	0.15
	mean $\pm$ sd***	$0.53 \pm 0.18$	$0.76 \pm 0.24$	$0.38 \pm 0.13$
	median***	0.48	0.71	0.35

\* Estimate for 0-100 m

\*\* Based on the POC model

\*\*\* Excluding Bogorov *et al.* study

Epipelagic to mesopelagic ratios varied drastically depending on a cell location (Figure 4.6B and E). A very clear latitudinal pattern can be observed using the epipelagic estimates from Strömberg *et al.* (2009) (Figure 4.6B), and mean ratios were  $1.00 \pm 0.45$  (range 0.01-6.24, median 1.05). With the exception of several cells along the coast of Antarctica and the center of a subtropical gyre in the South Pacific where extremely high ratios were found ( $>3$ ), the high ratios ( $\sim 2.2$ ) were observed between  $30^\circ\text{S}$  and  $30^\circ\text{N}$  band. Polar and equatorial areas were characterized by lower ratios ( $<1$ ). The lowest epipelagic to mesopelagic ratio was observed in upwelling regions along the west coast of Africa. Based on Bogorov *et al.* (1968) estimates, epipelagic to mesopelagic ratios did not show a clear pattern due to the binned nature of original epipelagic values (Figure E7). The average ratio per cell was  $1.12 \pm 1.14$  (median 0.81), with ranges four times higher (range 0-21.49) than calculated based on estimates from Strömberg *et al.* (2009).

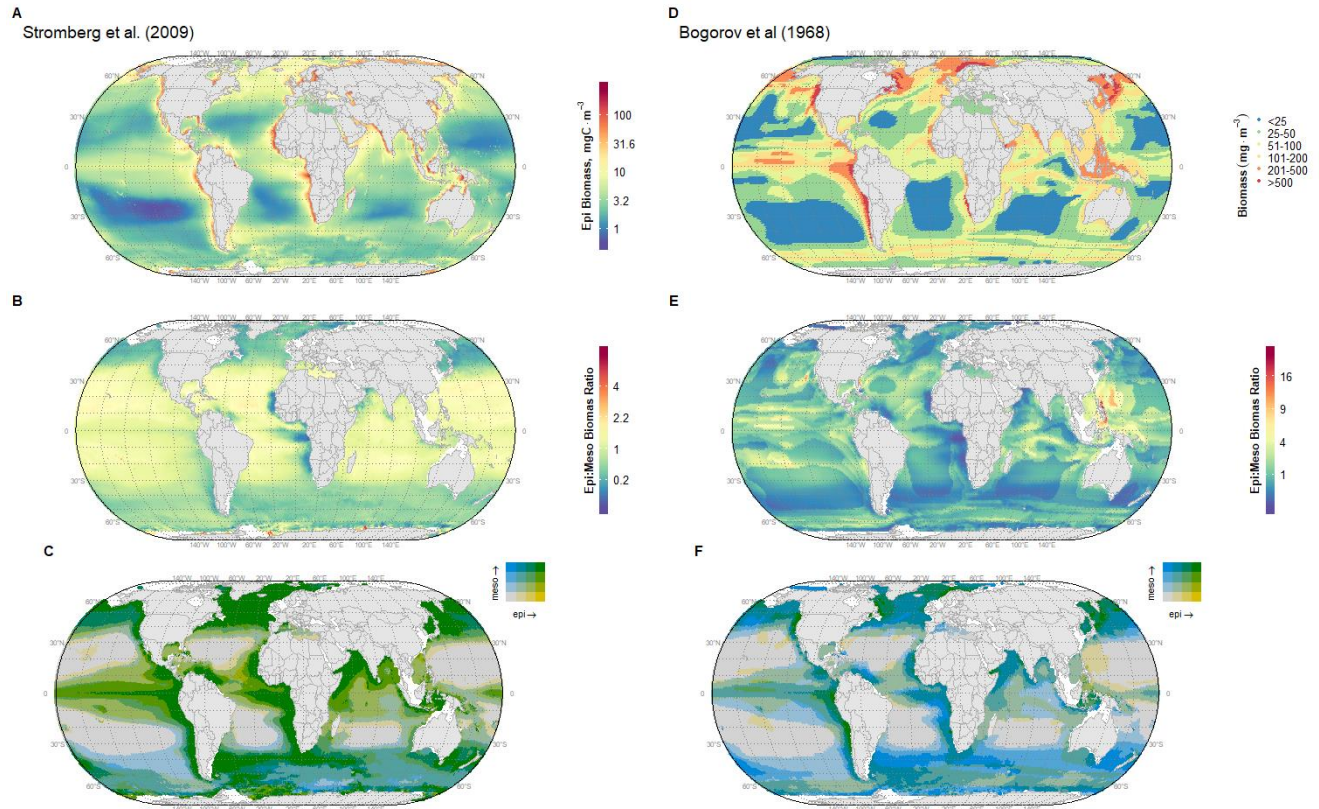


Figure 4.6: Global epipelagic to mesopelagic biomass ratio based on Stromberg *et al.* (2009) and Bogorov *et al.* (1968) epipelagic biomass maps. A) Epipelagic net zooplankton biomass (in  $\text{mgC}\cdot\text{m}^{-3}$ ) from Stromberg *et al.* (2009). Note the log-transformed color scale. D) Epipelagic net zooplankton biomass ( $\text{mg}\cdot\text{m}^{-3}$  wet weight) from Bogorov *et al.* (1968). B) Ratio between epipelagic and mesopelagic biomass using Stromberg *et al.* (2009) and epipelagic zooplankton and mesopelagic biomass from the classical-depth POC model. Note the square root of the scale for the color bar. E) Ratio between epipelagic and mesopelagic biomass using Bogorov *et al.* (1968) epipelagic zooplankton and mesopelagic biomass from the variable-depth POC model. Note the square root of the scale for the color bar. C) 2D map depicting the relationship between mesopelagic mesozooplankton biomass derived from POC estimates and Stromberg *et al.* (2009). Quantile breaks in the legend were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. F) 2D map depicting the relationship between the mesopelagic mesozooplankton biomass derived from POC estimates Bogorov *et al.* (1968) estimated the total zooplankton biomass in the top 100 m of the water column. Quantile breaks in the legend were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles for mesopelagic biomass and <50, 51-200, 201-500, >500  $\text{mg}\cdot\text{m}^{-3}$  wet weight intervals for Bogorov's biomass estimates. All maps in Eckert IV global equal area projection.

When the MM biomass was linked to the surface biomass of the zooplankton using 2D maps, a very good agreement was found between the two methods (Figure 4.6C and F). The following patterns were observed: both epi- and mesopelagic biomass values were low in the center of the ocean gyres. Both maps show consistently high biomass values in both layers at high latitudes (particularly in the Northern Hemisphere), equatorial areas, and the west coasts of continents. Low

surface biomass and high mesopelagic biomass values were found around the gyre regions and the 50°S and 50°N bands.

I also found good agreement between the distribution of MM and mesopelagic fishes in the Mediterranean Sea. The Mediterranean Sea showed a longitudinal pattern in the biomass distribution of mesopelagic zooplankton (Figure E9), with higher values found in the western subregion, medium values in the Ionian subregion, and the lowest biomass values in the eastern sub-basin. In general, areas closer to the shore are associated with higher biomass values. A similar distribution can be seen in mesopelagic fishes, as shown by Clavel-Henry *et al.* (2020).

I investigated the correlation between the mesopelagic zone provinces derived from the depth and echo intensity of acoustic deep scattering layers in Proud *et al.* (2017) and the clustering of biomass and species richness in this study. My results revealed that the provinces were consistent at low latitudes, with distinct regions in the central gyres and equatorial areas. Additionally, the partitioning of the Mediterranean zone was similar between the two approaches. However, findings show a different pattern at higher latitudes than those reported by Proud *et al.* (2017). In the latter case, there was a continuous province in the Southern Ocean (Figure 4.7A). However, this study did not show a clear latitudinal band (Figure 4.7B).

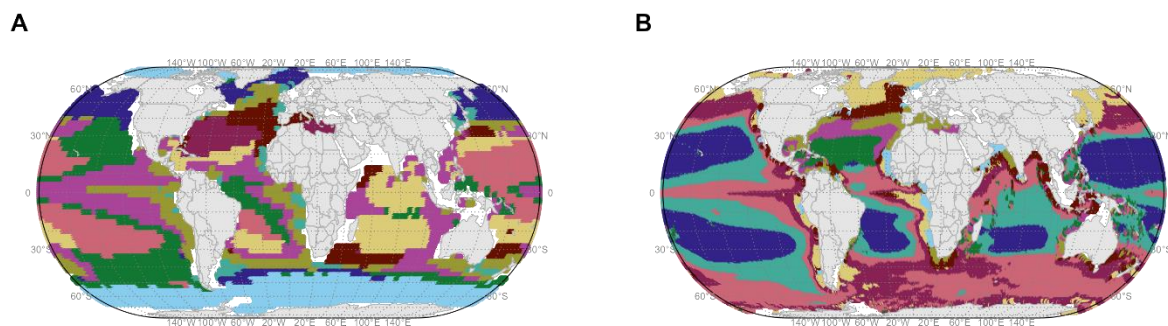


Figure 4.7: Present-day mesopelagic biogeography derived by K-means clustering (k=10) of A) Proud *et al.* (2017) gridded mean annual primary production (gCm<sup>-2</sup>·day<sup>-1</sup>) and temperature at the principal depth of the deep scattering layer and B) gridded biomass (derived from POC) and potential species richness (estimated in Chapter 5). The cluster number in A represents increasing backscatter (a proxy for mesopelagic biomass). The clusters in B are shown in Figure E8. All maps in Eckert IV global equal area projection.



## 4.4 Discussion

### 4.4.1 *Models*

The study used ordinary least squares (OLS) regression to examine the relationship between MM biomass and environmental conditions using POC or NPP data. Despite model simplicity, the models could explain 18-30% of the variance in the biomass estimates depending on the method used. Simple OLS regression used in this study performed better (higher  $R^2$  values) than more complex models (boosted regression trees) built on a more complex set of environmental variables (Drago *et al.*, 2022). In this study, NPP-based models showed lower  $R^2$  values (18-24%) than POC-based models (24-30%). However, the  $R^2$  values of the NPP-based models were lower than those reported for a similar model by Hernández-León *et al.* (2020). One potential reason for these lower values could be the use of climatological data, which may not have accurately represented the conditions in 2011, when the empirical biomass data were collected. In 2011, La Niña and the Pacific Decadal Oscillation cooled parts of the Pacific Ocean, but unusually warm temperatures predominated elsewhere (Xue *et al.*, 2012). The 2011 global average ocean temperature was 0.40°C above the 20<sup>th</sup> century average and ranked as the 11<sup>th</sup> warmest on record (NOAA National Centers for Environmental Information, 2012).

### 4.4.2 *Biomass estimates*

When modelling MM distribution (presence/absence), NPP was shown to be one of the least important factors (Figure D5). However, this work demonstrated that NPP is a very important factor for modelling mesozooplankton biomass and can explain up to 24% of the variation in MM biomass. POC was an even better predictor of zooplankton biomass in the mesopelagic zone, with up to 30% of the variance explained by POC concentrations (Table 4.2).

Regardless of the chosen method (variable or classical depth models), all models produced similar spatial patterns of the MM biomass distribution (Figure 4.2A, E, C, and G). Overall, MM biomass showed enhanced values in the Northern Hemisphere, along the west coasts of continents, in the equatorial region, and at the 50°S band (Arctic Polar Front). Low biomass values were observed at the centers of ocean gyres. Such distributional patterns are consistent with previously reported patterns (Drago *et al.*, 2022). A similar latitudinal distribution of biomass between the 30°N and 30°S bands was also evident when the lowest values observed at 30°N and 30°S gradually

increased towards the equatorial region (Figure 4.2). This pattern is consistent with the latitudinal distribution of the Copepoda biomass reported by Drago *et al.* (2022). Indeed, crustaceans comprise the majority (~74% of biomass) of the total mesopelagic biomass (Drago *et al.*, 2022) explaining similar spatial patterns. Overall, the distribution of the NPP-derived biomass corresponded with the latitudinal distribution of the Copepoda biomass estimated using boosted regression trees (BRTs) (Drago *et al.*, 2022). However, the POC-derived biomass pattern more closely resembled the spatial patterns reported by the underwater vision profiler's samples (Drago *et al.*, 2022).

The major difference between the two projections appeared to be biomass patterns at higher latitudes where NPP-derived biomass values decreased at higher latitudes while POC-derived biomass values showed the opposite trend. Decreasing trends of NPP model were mirrored by increasing (or stable) patterns observed in POC-based models. Furthermore, in this study NPP-based models showed a peak in biomass around 40-50° latitudes, while it was recorded at 60° latitudes in the Drago *et al.* (2022) study (BRT model). The discrepancy in biomass patterns at higher latitudes could be explained by an overall decrease in the NPP estimates in the polar regions (Figure E1A). Alternatively, the POC values at higher latitudes remained relatively high, resulting in higher biomass levels (Figure E1B). The choice between NPP and POC and the discrepancy between estimates could be related to the omnivory proportion mesozooplankton, particularly in the midwater realm (Zeldis & Décima, 2020; Kumar, 2005). Indeed, the proportion of omnivorous and carnivorous mesozooplankton overwhelmingly outweighed the herbivorous zooplankton in the mesopelagic zone (Halfter *et al.*, 2020). The major delivery mode of organic matter within the mesopelagic realm is through sinking marine snow and detritus (Omand *et al.*, 2020). Hence, the POC-based biomass estimates would be, in my view, more realistic. Another explanation of the difference in biomass estimates between NPP and POC model can be attributed to the POC values that are calculated using e-ratios (the ratio between POC flux and total productivity) (Xie *et al.*, 2021).

Despite spatial coherence of biomass hotspots, models employed in this study showed differences between overall biomass estimates. The NPP-based variable-depth and classical depth model estimates were 68% and 56% lower than their POC-based counterparts, respectively. This highlights the importance of considering the dynamic depth of the mesopelagic zone in biomass

calculations because, in many instances, the variable depth range of the mesopelagic zone was larger than the classically accepted (since classical definition of mesopelagic zone underestimates the total volume of the mesopelagic by 40%).

The global mesozooplankton biomass in the mesopelagic zone varied between 0.20 and 0.91 PgC, depending on the method employed (Table 4.2). The classical depth NPP model showed a similar estimate (0.20 PgC) to that (0.173 PgC) reported by Drago *et al.* (2022). Although the Drago *et al.* (2022) assessment was confined to the upper mesopelagic layer, the similarity in estimates suggests that the majority of the mesozooplankton biomass is located in the upper mesopelagic layer of the water column (Martin *et al.*, 1987; Puelles *et al.*, 2019). All methods, except the variable depth POC model, showed lower estimates than 0.66 PgC reported by Hernández-León *et al.* (2020). For a similar depth range (200-1,000 m), Hernández-León *et al.* values were 0.21 PgC higher than POC-derived estimates and 0.46 PgC higher than NPP-derived estimates. Lower biomass estimates can be explained by the exclusion in the analysis of highly productive coastal areas. However, the estimate using the variable depth POC model yielded the highest to date mesozooplankton standing stock, e.g.,  $0.91 \pm 0.08$  PgC. This again highlights that biomass of the mesopelagic zone is usually underestimated by the classical definition of mesopelagic zone. Unfortunately, Hernández-León *et al.* (2020) did not provide a level of variability in the estimated global biomass value, and currently, it is not clear how different the estimates are.

Regarding latitudinal distribution, biomass of mesopelagic mesozooplankton within the 40°S–40°N belt accounted for 68-70% of the global biomass estimates using NPP models and 51-55% using POC-based models (Table 4.2). As mentioned above, it is plausible that mesopelagic biomass at higher latitudes can be underestimated by NPP-based models. Furthermore, POC-based models demonstrated that the average biomass values at higher latitudes were an order of magnitude higher than those at lower latitudes (Figure 4.3F and G), which was consistent with a previous study by Puelles *et al.* (2019), who found more biomass in high latitudes than in the tropics. Higher biomass at polar areas can be explained by the fact that polar regions are characterized by extreme seasonality in primary productivity values compared to the tropics (Murphy *et al.*, 2016). This regime produces an intense productivity period for several months during the year. Thus, for most of the year, the pelagic community relies on the microbial community (Arrigo, 2014). I concur that POC-based models appear to be a more effective method

for determining mesozooplankton biomass in the mesopelagic zone, as POC provides a more precise estimate of the organic matter standing stock that is available for zooplankton consumption in midwater layers, particularly at higher latitudes (assuming that POC values are estimated correctly). Furthermore, enormous biomass of overwintering (diapausing) zooplankton contributes significantly to the mesopelagic zooplankton biomass seasonally (Boyd *et al.*, 2019; Kobari *et al.*, 2013; Kobari *et al.*, 2008).

#### 4.4.3 Biomass and Diversity

The results of this study showed a positive relationship between zooplankton biomass and species richness in different marine environments (Figure 4.3). The log-transformed species richness values showed a sharp increase as the log-transformed biomass values increased. However, this increase was significant only until the biomass value reached  $1 \text{ mgC}\cdot\text{m}^{-2}$ . The subsequent increase was, however, small, even as biomass values continued to increase. The species richness plateaued at approximately 300 species per cell. Using POC-based estimates, I found that regions with high biomass also tended to have high species richness (Figure 4.4A). This pattern was particularly evident in the Arctic, North Pacific, coastal, equatorial areas, and the Southern Ocean. In contrast, low biomass regions were often associated with the centers of ocean basins in the North/South Pacific and Indian Oceans. However, an exception to this pattern was observed in the center of the North Atlantic Basin, where low biomass was paired with high species richness values. Additionally, low biomass and high species richness values were observed in the Mediterranean Sea and along the western boundaries of subtropical gyres. Finally, areas of high biomass but low species richness were found in subtropical regions around the edges of gyres and in some areas of the North Pacific and Southern Hemisphere.

Hence, while high biomass can be a good indicator of high species diversity, it is not the only factor at play. Other environmental conditions such as temperature, nutrient availability, and ocean circulation patterns may also shape the distribution of mesozooplankton species in the mesopelagic zone. Future studies should uncover these factors to better understand the relationship between biomass and diversity in this important ecosystem. The species richness used in this study was estimated based on SDM models and thus may not represent the actual species richness in the cell because the accuracy of SDM models is highly dependent on the occurrence data for each species. Thus, rare species or areas with low sampling efforts may underestimate species richness values.

The latitudinal pattern of the median MM biomass mirrored the pattern observed in the species richness. Both biomass and species richness showed a gradual increase towards the equator, with the lowest values found around 23.5°N and °S. In the Northern Hemisphere, the biomass gradually increased throughout the temperate region and stabilized at high values in the polar region. Species richness showed a similar increase in the northern temperate region, with some increases and declines around 45-50°N. In the Southern Hemisphere, biomass and species richness values increased throughout the temperate region, with a small peak at approximately 30°S. However, while biomass in the southern polar region continued to increase towards the pole, species richness showed only a moderate increase in polar areas. These observations were consistent with the findings reported by Drago *et al.* (2022), who observed high values north of 55°N and south of 55°S, and relatively higher values of biomass were predicted around the equator (15°N-15°S). The main difference was observed in the Southern Ocean area, where Drago *et al.* (2022) reported that high biomass values were located between the Subantarctic Front and the southern limit of the Antarctic Circumpolar Current, with lower biomass values north and south of the band. This discrepancy can be explained by the choice of environmental parameters, particularly prey availability proxies (NPP- (Figure 4.2A and E) vs. POC-based (Figure 4.2C and G) calculations) used in the model.

I found high mean values for both biomass and species richness at higher latitudes compared to the equatorial region (Figure 4.3B). The cold, dynamic, highly seasonal environment, and long periods of darkness in the polar regions result in slower speciation and may favor fewer but highly specialized species. The enhanced average range-size rarity supports this suggestion (Figure 3.5). In addition to creating a suitable environment for specialized species, the area appears to support high zooplankton biomass despite polar systems being characterized by a ‘pulsing’ (highly seasonal) productivity. In spite of the primary productivity varying drastically seasonally, POC values, on the other hand, may reflect the accumulation of organic matter that occurs over longer time scales and is less influenced by seasonal fluctuations in NPP. Additionally, polar regions export approximately twice as much of their production compared to tropical regions (Antia *et al.*, 2001), leading to higher POC values. Such high values can support a higher biomass of organisms at high latitudes.

High biomass levels in polar regions are model projections built on models' biomass values collected at lower latitudes. Even though the range of the environmental variables used to build the models was comparable to values used to predict the biomass values on a global scale, I assumed that the relationship between the POC or NPP levels remained the same for both low and high latitudes. However, the assumption of a consistent relationship between POC or NPP levels in low and high latitudes may not necessarily be true, as the mode of primary productivity varies drastically between areas with high seasonal and year-round primary productivity in polar areas and at lower latitudes. In addition, I discuss only the mean climatology values of POC and NPP, which do not reflect annual or depth variability. Hence, obtained biomass data are only a snapshot of the mean climatological conditions and could differ from biomass values at seasonal or annual scales or at variable depths.

In some cases, the high biomass of a single dominant species can lead to a reduced diversity because the dominant species could outcompete other species for resources. Similarly, low biomass may not always result in low diversity, as some species may survive and thrive even in low-biomass environments. For instance, the shape of the productivity–diversity relationship for marine phytoplankton appears to be unimodal with diversity peaking at intermediate productivity levels (Vallina *et al.*, 2014). A negative correlation between the biomass and species diversity of marine zooplankton has been found in regional zooplankton studies, e.g. in the Indian Ocean (Ghilarov & Timonin, 1972).

The hump-back relationship between log-transformed total range-size rarity and biomass refers to a pattern observed in biodiversity studies, where the highest biomass values occurred at intermediate biodiversity levels. In this study, I found a skewed humpback relationship, where the biomass repeatedly peaked at approximately 2/10 of the total range-size rarity (two rare species out of 10 species present) (Figure 4.5A). The species richness and total range-size rarity were generally low at low biomass levels because fewer resources were available for organisms to survive and reproduce (Tseitlin, 1986). Low biomass values were associated with ocean gyres (Figure 4.2A). The water in the gyres tends to be warm and relatively stable, with little mixing between the surface and deeper waters. The lack of mixing means less exchange of nutrients between the surface and deeper waters, which can limit phytoplankton growth and enhance recycling. Thus, food availability can be a limiting factor in keeping the MM biomass low in these

areas. The biomass peak was found at  $\sim 2/10$  rarity values, indicative of areas with few rare species. Such areas are also characterized by intermediate species richness values (Figure 4.5A). Biomass is generally high at intermediate levels of species richness because of the way different species interact with each other and with their environment. There is a balance between competition and resource availability at the intermediate species richness levels. The diversity of species present means that organisms can exploit different resources, leading to a more efficient use of available resources (Finke & Snyder, 2008; Ives *et al.*, 2005; Casula *et al.*, 2006). Additionally, a greater number of 'redundant' species can lead to a more stable ecosystem, as different species can buffer each other against environmental fluctuations (Ives *et al.*, 2000). At high species richness levels, competition for resources can intensify, limiting the growth and reproduction of individual organisms, and ultimately decreasing the overall biomass (Huston, 1979; Grime, 1973).

An interesting pattern was observed between the mesozooplankton biomass, average rarity, and species richness (Figure 4.5B). The average rarity of  $3/10$  seems to be an important threshold associated with a peak in biomass and species richness. The biomass decreased rapidly as the average rarity decreased from  $3/10$ . This pattern was observed on both the global and ocean-basin scales (Figure 4.5B). The following implies that the peak in species richness can be established only in areas of low average rarity (one rare species out of 10). Such areas were primarily found in pelagic areas along the coastlines of the continents,  $50\text{-}60^\circ\text{N}$  band in the North Pacific and North Atlantic.

Based on the results of this study, I can conclude that MM biomass was positively associated with POC values. In areas where POC is abundant, the mesozooplankton biomass tends to be high, whereas in areas where POC is scarce, the biomass levels are low. Regions with low biomass are often located in the center of ocean gyres and are characterized by low species rarity and/or low species richness. In contrast, areas with high biomass are found in locations where both the species richness and rarity are at a certain level. Specifically, high biomass is associated with high species richness and is equal to or lower than three rare species in every ten species. Therefore, to achieve a high mesozooplankton biomass, both high species richness and mid-level rarity are necessary conditions.

#### 4.4.4 Link Between Biomass in Epi- and Mesopelagic Zones

The ratio representing the proportion of epipelagic biomass in relation to the mesopelagic zone was presented in Table 4.3. Using the classical depth POC model, I found that the mean ratio was 0.76, suggesting that 76% of the epipelagic biomass was located within the mesopelagic zone. This estimate was very similar to the ratio reported by Hernández-León *et al.* (2020), who calculated a ratio of 0.72. However, when I considered the variable depth of the mesopelagic zone, the mean ratio decreased to 0.38, indicating that only 38% of the epipelagic biomass was found in the mesopelagic zone.

These findings suggest that the zooplankton biomass in the mesopelagic ocean may be much larger than previously assumed. However, these ratios are highly dependent on the total estimates of epipelagic and mesopelagic biomass. I found that even the well-studied epipelagic zone has variable estimates of the global zooplankton biomass. In the mesopelagic zone, only one study has attempted to estimate the biomass, and even that was based on the classical definition of the mesopelagic zone. This variability in the estimates can also be explained by the lack of global coverage of zooplankton biomass in the upper layers of the ocean. I also acknowledge that the ratio can additionally be influenced by the conversion between units. For example, I used a single factor of 0.12 to convert wet weight to carbon mass (Table 4.1). However, the mesopelagic community composition is not spatially and temporally uniform, and can be composed of variable taxa in different regions, and taxa-specific carbon content can vary between 0.003 and 0.15 (Kiørboe, 2013).

The link between the surface and mesopelagic biomass cannot be easily determined, as very few studies have provided a continuous global estimate of zooplankton biomass. I first attempted to calculate the global distribution of the epipelagic : mesopelagic ratio using zooplankton biomass maps from Bogorov *et al.* (1968) and Strömberg *et al.* (2009). This study demonstrated that epipelagic-to-mesopelagic biomass ratios were spatially heterogeneous (Figure 4.7B and E). The digitized map of Bogorov *et al.* (1968), due to the binned nature of the map, limited the ability to look at the global distribution of epipelagic to mesopelagic biomass ratios. Nevertheless, using Strömberg *et al.* (2009) epipelagic zooplankton biomass data, I concluded that the global ocean epipelagic/mesopelagic biomass ratio was not uniform with ~ 50% of cells had ratios < 1, with the majority falling below 0.5. The mean cell ratio was 1.00, indicating that the mesopelagic zone



biomass was same as in the epipelagic zone. At the same time, the remaining ~ 50% cell ratios exceeded 1, indicating that the epipelagic biomass in those areas was greater than that in the mesopelagic realm. These patterns were very closely related to the depth of the mesopelagic-bathypelagic boundary, and areas with high epipelagic/mesopelagic ratios coincided with the deepest mesopelagic layers (Reygondeau *et al.*, 2018; Reygondeau & Dunn, 2019). Such areas were found in the center of the ocean basins and were characterized by low primary productivity (Figure E1). In contrast, areas with low epipelagic/mesopelagic ratios were found north of 30° in both hemispheres and some equatorial upwelling areas. These regions are characterized by strong seasonality and water mixing that support high seasonal zooplankton biomass in the epipelagic zone. As a result, there is a possibility that a larger proportion of this productivity ending in the mesopelagic zone, and POC values are among the highest in these regions, supporting the larger biomass of mesopelagic zooplankton. In addition, several temperate and subpolar large calanoids of the genera *Calanus*, *Eucalanus*, and *Neocalanus*, after accumulating carbon in lipids during a few summer months, undergo seasonal vertical migration, thus reallocating a large biomass between the epipelagic and mesopelagic realms and supporting the carnivorous midwater food web (Darnis & Fortier, 2014).

Overall, MM biomass can be linked to surface zooplankton and mesopelagic fish biomass (Figure 4.7C and D). When I examined maps of biomass patterns, both epipelagic and MM biomass were low in all oceanic gyres. The same was also evident when zooplankton data compiled by the Food and Agriculture Organization (FAO) were used, although the surface biomass values in these areas were higher. Furthermore, biomass values for both (epi- and mesopelagic) layers were consistently high at high latitudes, particularly in the Northern Hemisphere, equatorial region, and along the west coasts of continents. In contrast, low surface biomass and high mesopelagic biomass values were found in areas around 50°S and 50°N near the edges of the ocean gyres.

By clustering mesozooplankton biomass and species richness values, I observed good agreement between mesopelagic provinces identified by Proud *et al.* (2017) as a proxy for mesopelagic fish biomass. The clusters were similar in almost all areas, except for the Southern Ocean. In contrast to the findings of Proud *et al.* (2017) and Reygondeau *et al.* (2018) this study observed patchy patterns in the Atlantic sector of the Southern Ocean, rather than a continuous circumpolar province, within the defined mesopelagic provinces. This discrepancy can be explained by the

POC patterns (Figure E1) in the region, which had a defining influence on mesopelagic biomass estimates.

Good agreement was found between the distribution of mesopelagic fish biomass reported by Clavel-Henry *et al.* (2020) and the zooplankton biomass modelled in this study (Figure E9). The distribution of MM and mesopelagic fishes in the Mediterranean Sea showed a longitudinal pattern, with higher biomass values found in the western subregion and near-shore areas and lower biomass values in the eastern sub-basin. Such findings can be an indicator of a tight link between the upper and lower layers of the pelagic ocean as well as between mesozooplankton and their direct predators (mesopelagic fishes).

It is important to address the presence of diel vertical migration (DVMs) between epipelagic and upper mesopelagic layers (Freer & Hobbs, 2020). Indeed, a certain portion of the estimated mesozooplankton biomass in the mesopelagic realm is represented by diel migrants (Kelly *et al.*, 2019). However, because global biomass in both layers has not been previously calculated, the proportion of migrant biomass is not known. Modelling approaches estimated that the zooplankton DVM is an important part of the biological pump, increasing the global export flux by 14% (Archibald *et al.*, 2019; Boyd, 2015; Boyd *et al.*, 2019). To validate this estimate, I also compiled the day/night epipelagic mesozooplankton biomass ratios from various regions of the ocean (Valencia *et al.*, 2018; Landry & Swalethorp, 2022; Head, 1999; Wishner *et al.*, 1998; Le Borgne, 2003; Kitamura *et al.*, 2016; Lenz *et al.*, 1993). The ratios varied between 0.04 and 0.39 with a mean value of 0.20 (median 0.17). Thus, I applied a median ratio of 17% to calculate the portion of estimated mesopelagic biomass that can be attributed to migrating organisms from the epipelagic layer, producing a migrant biomass of 0.15 PgC (when using the variable depth model) or 0.08 PgC (using the classical definition of the mesopelagic zone).

One of the major complications in estimating mesozooplankton biomass is the lack of community biomass estimates resulting from community undersampling and sparsity of sampling efforts. True species richness values are unknown for many locations, as only a few species/taxa have been reported. Consequently, many areas do not have complete community composition and biomass estimates. In this study, I used the biomass estimates from the Malaspina Expedition, which sampled mainly tropical and subtropical regions; hence, the models built may not adequately represent the biomass at higher latitudes (Vinogradov *et al.*, 1999). In addition, net catches are

prone to under-sampling of the gelatinous zooplankton. Hence, estimates of mesopelagic biomass may be conservative. Indeed, mesopelagic depth boundaries are dynamic and vary with the season and time. I used a static representation of the mesopelagic zone and used a constant biomass value within the mesopelagic layer because biomass data were only available for the entire 200-1,000 m depth range (Hernández-León *et al.*, 2020). However, it has been shown that zooplankton density declined with depth across the three oceans (Puelles *et al.*, 2019).

## 4.5 Concluding remarks

This study examined the relationship between MM biomass and environmental conditions by using OLS regression with either POC or NPP data. The models explained 18-30% of the variance in biomass estimates, with POC-based models performing better than NPP-based models. The results showed that NPP is an important factor for modelling mesozooplankton biomass, whereas POC is an even better predictor. All models showed similar patterns of MM biomass, with enhanced values in the Northern Hemisphere, along the west coasts of continents, equatorial region, and 50°S band. The global MM biomass was estimated to range between 0.20 and 0.91 PgC, depending on the method and model used. I found that NPP-based models may underestimate mesozooplankton biomass at higher latitudes compared with POC-based estimates, which showed that the average biomass values at higher latitudes were an order of magnitude higher than those at lower latitudes. Thus, POC-based models may be more effective for determining mesozooplankton biomass in deeper ocean layers and at higher latitudes. The POC-based variable-depth model resulted in the highest estimate of  $0.91 \pm 0.08$  PgC.

This study found that the MM biomass was lower in areas with low POC values and was associated with low species richness and/or low range-size rarity. However, high biomass values were more commonly observed in regions with intermediate rarity (rarity of 0.3) values and high species richness, although both conditions should be met to ensure high biomass values. These results suggest that high species richness and intermediate rarity, coupled with high POC, are necessary for high mesozooplankton biomass.

This study found a connection between the surface and mesopelagic biomass, linking the MM biomass to the surface zooplankton and mesopelagic fish biomass. Both layers show low biomass values in the central regions of ocean gyres and consistently high values in the high latitudes and

equatorial regions. This study revealed that the proportion of epipelagic biomass in relation to the mesopelagic zone varies greatly depending on the depth of the mesopelagic zone. Based on the classical depth approach, the global epipelagic mesozooplankton biomass accounted for  $\sim 3/4$  of the mesopelagic biomass. However, if the variable depth approach was deployed, the global epipelagic mesozooplankton biomass could only constitute just over  $\sim 1/3$  of the mesopelagic biomass. These findings highlight higher than previously believed mesopelagic biomass when using variable depths to estimate the biomass of zooplankton in the mesopelagic ocean. However, the variability in estimates is also influenced by the lack of global coverage of zooplankton biomass in the upper layers of the ocean and the diversity of the mesopelagic community composition. Further research is necessary to obtain a more accurate estimation of the zooplankton biomass in the mesopelagic zone.

In addition, we attempted to estimate the global spatial distribution of the epipelagic to mesopelagic biomass ratio using maps from previous studies. The results showed that the ratios varied significantly depending on location, with over 50% of cells having ratios below 1, indicating a higher mesopelagic biomass compared to the epipelagic zone. These areas with low ratios are found in temperate and polar regions and characterized by shallower mesopelagic layer and 'pulsing' (seasonal) mode of productivity. Such areas support high epipelagic biomass, and even larger portions of productivity in these regions end up in the mesopelagic zone, supporting an even larger biomass of mesopelagic zooplankton. In addition, low ratios were found in equatorial and upwelling areas. On the other hand, areas with high ratios are found in oligotrophic areas with low primary productivity. As these areas primarily rely on regenerative productivity, higher biomass values were found in the epipelagic zone than in the mesopelagic zone.

The study also found agreement between mesozooplankton biomass and species richness, and mesopelagic provinces defined by previous studies, with a patchy pattern observed in the Southern Ocean due to complicated POC patterns. These findings indicate a tight link between the upper and lower layers of the pelagic ocean and between mesozooplankton and their direct predators, mesopelagic fishes.

## Chapter 5: Synthesis and Conclusion

The mesopelagic zone of the world's oceans and the organisms therein are grossly understudied (Behagle et al., 2015), yet they play critical roles in carbon and nutrient cycling in the ocean. Mesozooplankton (zooplankton with a size range of 0.2-20 mm) are small heterotrophic organisms that feed on phytoplankton, detritus, and small heterotrophic animals. They also perform DVMs via vertical migration in the water column, travelling closer to the surface to feed at night and back to deeper water during the day (Sutton, 2013). This feeding cycle brings carbon and nutrients from productive surface waters to less-productive deeper waters (Stemmann et al., 2008). Some of this carbon remains in deep waters for 100-1,000s of years, while some can ultimately settle on the seabed, where it can remain sequestered for millions of years or come back to the surface (Roy et al., 2000). Thus, mesozooplankton populations play a direct role in mitigating climate change. Owing to the scarcity of measurements of mesopelagic mesozooplankton density and their importance with regard to oceanic nutrient cycling and climate change, scientists wish to estimate global mesozooplankton density for use in models of ocean biomass or biodiversity, earth system processes, and policy and management contexts (Hidalgo & Browman, 2019). Despite the ecological importance of mesozooplankton, little is known about their global distribution and diversity in the mesopelagic realm. This study aimed to understand the spatial distribution and diversity of mesopelagic mesozooplankton in the global ocean.

The primary aim of this study was to establish a comprehensive global foundation for understanding the biogeography and diversity of mesopelagic zones. This study seeks to address various fundamental macroecological and applied inquiries, including three key questions:

1. What is the current level of understanding of the global distribution of mesopelagic mesozooplankton and micronekton communities? By examining existing knowledge, this study aimed to assess the current state of information regarding the spatial distribution of these organisms in the mesopelagic zone worldwide.
2. What is the extent of global distribution and diversity observed in mesopelagic mesozooplankton? This study aimed to unravel the patterns and variations in the distribution and diversity of mesopelagic mesozooplankton species across different regions of the world's oceans.

3. What is the estimated global mesozooplankton biomass? By analyzing available data, this study aimed to provide an estimate of the total biomass of mesozooplankton in the global mesopelagic zone, shedding light on their ecological significance and contribution to marine ecosystems.

To explore the current state of knowledge, an extensive review of the literature was conducted in Chapter 1 to explore what is known about mesopelagic organisms and to identify important gaps in knowledge. Over the years, our understanding of mesopelagic zones has advanced significantly through a combination of expeditions and technological breakthroughs. Early explorations in the 19<sup>th</sup> century defied the prevailing belief that the deep ocean was devoid of life. Expeditions such as *Challenger* and *Galathea* provided groundbreaking evidence of life below the ocean's surface, unveiling the presence of a scattering layer and diel vertical migrations. However, these studies primarily focused on taxonomy and benthic fauna, leaving the midwater zone relatively unexplored. In the 20<sup>th</sup> century, progress was made in unraveling the distribution and diversity of mesopelagic organisms. Echo-sounding technology has revolutionized our understanding by revealing the intricate patterns of DVM and the existence of a deep scattering layer. The 21<sup>st</sup> century has witnessed the emergence of novel approaches that further propelled our understanding. DNA barcoding techniques have enabled species identification, while the development of robotic machines for deep-sea visual monitoring has opened new avenues for data collection. Global ocean expeditions, including BEAGLE, Galathea III, Tara Oceans, and Malaspina, have significantly contributed to our knowledge of the physical, chemical, and biological properties of the oceans. Despite these advancements, substantial gaps in our understanding have persisted.

One of the major gaps is the lack of quantitative data on the zooplankton density and biomass in the mesopelagic ocean, especially in center of ocean basins or at depths deeper than 500 m. Available data are very variable in sampling methods, taxonomic resolution, and temporal and spatial coverage. Standardizing sampling methods across different regions remains a challenge, hindering comprehensive comparisons. Uncovering the intricate connection between surface productivity and mesopelagic biomass is crucial to understand the dynamics of this zone. Additionally, a comprehensive assessment of mesopelagic zooplankton biomass is needed to provide valuable insights into their role in carbon cycling and ecosystem functioning. However, the impact of climate change on mesopelagic organisms requires further investigation. Enhancing

our understanding of mesopelagic zones is vital for effective conservation and management strategies and is critical for comprehending the profound impacts of climate change on the intricate dynamics of the world's oceans.

Chapter 2 presented a comprehensive assessment of the global distribution and density of mesopelagic mesozooplankton and micronekton communities. The first quantitative Mesopelagic Mesozooplankton and Micronekton Database was established from a comprehensive collection of literature, including published and unpublished sources. An attempt was made to systematically organize and standardize the gathered information to account for variations in sampling depth, gear used, sampling location and season, thus enabling effective analysis and comparison of the data. This chapter also focuses on exploring the effectiveness of different statistical methods for modelling the relationship between mesozooplankton density and environmental covariates. I found that the commonly used approach of log-transforming the response variable in a standard linear regression model yielded the poorest results for this type of data distribution. Instead, the Generalized Linear Model (GLM) with an Inverse Gaussian error distribution provided a better fit for the data.

In Chapter 3, I gathered 861 mesozooplankton species from the mesopelagic zooplankton and micronekton databases and analyzed them using an ensemble of species distribution models. I used 10 modelling algorithms and three multi-model ensemble approaches to determine the best method for predicting mesopelagic species distribution and to quantify model uncertainty. In addition, the relative importance of different environmental conditions was quantified and compared to explain the distribution of mesopelagic species. Overall, the highest species richness of mesozooplankton was observed in the North Atlantic region along the west coast of India and Mediterranean Sea. Similar to previous studies that examined epipelagic zooplankton diversity, I found that latitudinal zonal species richness of mesozooplankton peaked at around 30° N and 30° S. Moreover, hotspots (areas with high species richness) were largely found at 40°S and 40°N, whereas such hotspots were absent in polar areas. In contrast, the mean species richness per area in polar areas was higher and more evenly distributed across polar biomes than in equatorial or tropical areas. The study also found a relationship between water temperature and species richness in the mesopelagic zone, with the highest species richness observed at a water temperature of approximately 25°C. The species richness decreased sharply at temperatures below 10°C. The total range-size rarity showed

a similar pattern as species richness, with high values found in the offshore North Atlantic and in coastal areas across both hemispheres, while the Pacific and Indian ocean basins were identified as areas with low rarity and low species richness. The euphotic zone depth, salinity, and dissolved nitrate concentration were estimated as the most important variables for explaining the distribution of mesopelagic mesozooplankton. The results from the ensemble modelling were robust in areas with abundant observational records, but uncertainty was high in data-limited regions, thus highlighting the need to increase the sampling effort in these regions.

Chapter 4 investigated the relationship between mesopelagic mesozooplankton biomass and environmental conditions using regression models based on either particulate organic carbon (POC) or net primary productivity (NPP) data. The models explained a significant portion of the variance (18-30%) in biomass estimates, with POC-based models performing better than NPP-based models. High biomass values were observed in the Northern Hemisphere along the west coast, equatorial region, and 50°S band. The results showed that POC produced a higher estimate of mesozooplankton biomass than NPP, particularly at higher latitudes. The study also revealed that low mesopelagic mesozooplankton biomass is associated with low POC values and low range-size rarity and species richness, but high biomass values are more commonly found in regions with intermediate rarity values and high species richness, coupled with high POC (Figure 5.1).

This study highlighted the connection between surface and mesopelagic biomass, as well as the variability in estimating zooplankton biomass in the mesopelagic zone. The proportion of epipelagic biomass relative to the mesopelagic zone varied depending on the mesopelagic zone depth. The global spatial distribution of the epipelagic to mesopelagic biomass ratio indicated a higher mesopelagic biomass compared to the epipelagic zone, particularly in temperate, polar, equatorial, and upwelling areas. The findings also demonstrated agreement between mesozooplankton biomass, species richness, and mesopelagic provinces, revealing a tight link between the upper and lower layers of the pelagic ocean and between mesozooplankton and their predators, mesopelagic fishes.



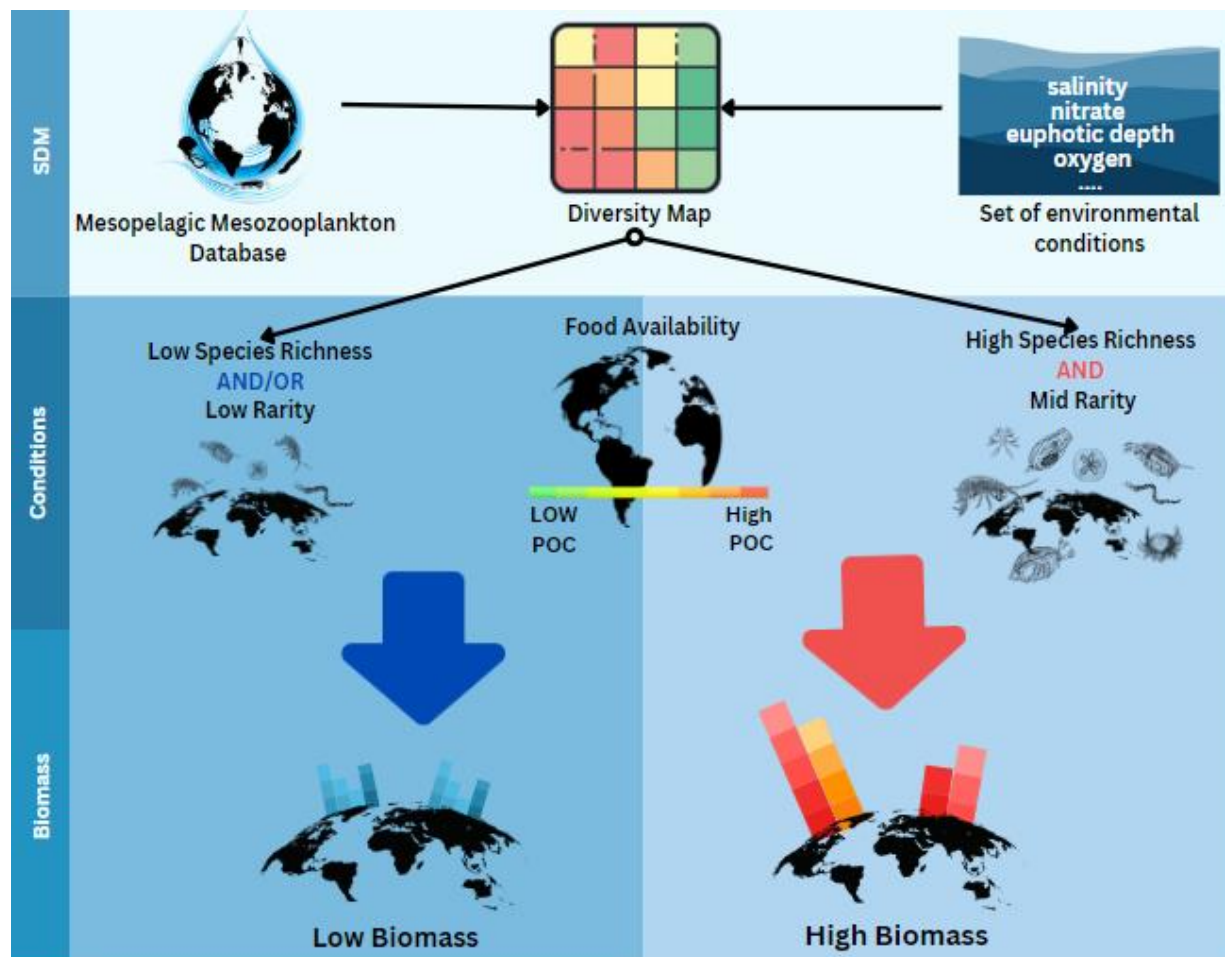


Figure 5.1: Graphical representation of the main components of this study, highlighting important connections between mesopelagic zooplankton diversity, food availability (POC concentrations), and mesopelagic biomass.

Overall, this study made significant contributions by establishing a comprehensive global database for mesopelagic mesozooplankton and micronekton and uncovering their distribution and diversity patterns. It also highlighted the importance of particulate organic carbon (POC) in predicting mesozooplankton biomass and revealed the spatial patterns of higher biomass in specific regions. These findings have implications for biodiversity, ecosystem functioning, and climate change research.

## References

- Agassiz, A. (1902) *Reports on the scientific results of the expedition to the Tropical Pacific. I. Preliminary Report.*
- Aguilar, M., Diaz, F. & Buckle, L.F. (1998) The effect of salinity on oxygen consumption and osmoregulation of *Macrobrachium tenellum*. *Marine and Freshwater Behaviour and Physiology*, **31**, 105–113.
- Akaike, H. (1978) On the Likelihood of a Time Series Model. *The Statistician*, **27**, 217.
- Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.
- Andersen, V. & Sardou, J. (1992) The diel migrations and vertical distributions of zooplankton and micronekton in the Northwestern Mediterranean Sea. 1. Euphausiids, mysids, decapods and fishes. *Journal of Plankton Research*, **14**, 1129–1154.
- Angel, M.V. (1993) Biodiversity of the Pelagic Ocean. *Conservation Biology*, **7**, 760–772.
- Angel, M.V., ed. (2003) *Ecosystems of the Deep Oceans*. Elsevier.
- Angel, M.V., Blachowiak-Samolyk, K., Drapun, I. & Castillo, R. (2007) Changes in the composition of planktonic ostracod populations across a range of latitudes in the North-east Atlantic. *Progress in Oceanography*, **73**, 60–78.
- Antia, A.N., Koeve, W., Fischer, G., Blanz, T., Schulz-Bull, D., Schölten, J., Neuer, S., Kremling, K., Kuss, J., Peinert, R., Hebbeln, D., Bathmann, U., Conte, M., Fehner, U. & Zeitzschel, B. (2001) Basin-wide particulate carbon flux in the Atlantic Ocean: Regional export patterns and potential for atmospheric CO<sub>2</sub> sequestration. *Global Biogeochemical Cycles*, **15**, 845–862.
- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in ecology & evolution*, **22**, 42–47.
- Archibald, K.M., Siegel, D.A. & Doney, S.C. (2019) Modeling the Impact of Zooplankton Diel Vertical Migration on the Carbon Export Flux of the Biological Pump. *Global Biogeochemical Cycles*, **33**, 181–199.
- Arrigo, K.R. (2014) Sea ice ecosystems. *Annual review of marine science*, **6**, 439–467.
- Bakun, A. (2022) Adjusting intuitions as to the role of oxygen constraints in shaping the ecology and dynamics of ocean predator–prey systems. *Environmental Biology of Fishes*, **105**, 1287–1299.
- Balls, R. (1948) Herring fishing with echometer. *Journal du Conseil Permanent International pour l'Exploration de la Mer*, **15**, 193–206.
- Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.
- Barr, D.J. (2013) Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, **4**, 328.
- Beaugrand, G., Kirby, R. & Goberville, E. (2020) The mathematical influence on global patterns of biodiversity. *Ecology and evolution*, **10**, 6494–6511.
- Beaugrand, G., Lenoir, S., Ibañez, F. & Manté, C. (2011) A new model to assess the probability of occurrence of a species, based on presence-only data. *Marine Ecology Progress Series*, **424**, 175–190.
- Beebe, W. (1934) *Half Mile Down*. New York Zoological Society, New York: Harcourt, Brace and Company.
- Béhagle, N., Roudaut, G., Josse, E., Brehmer, P., Cotte, C., Duhamel, G., Gauthier, O., Cherel, Y. & Ryan, T.E., eds. (2015) *Acoustics characterization of micronekton spatial distribution in Indian Ocean using scientific surveys and integrated marine observing system database: Acoustics characterization of micronekton*. IEEE.
- Behrenfeld, M.J. & Falkowski, P.G. (1997) Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, **42**, 1–20.

- Benedetti, F., Vogt, M., Elizondo, U.H., Righetti, D., Zimmermann, N.E. & Gruber, N. (2021) Major restructuring of marine plankton assemblages under global warming (vol 12, 5226, 2021). *Nature communications*, **12**.
- Berger, W.H. (2009) *Ocean: reflections on a century of exploration*. Univ of California Press.
- Bianchi, D., Galbraith, E.D., Carozza, D.A., Mislán, K.A. S. & Stock, C.A. (2013) Intensification of open-ocean oxygen depletion by vertically migrating animals. *Nature Geoscience*, **6**, 545–548.
- Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R. & Not, F. (2016) In situ imaging reveals the biomass of giant protists in the global ocean. <https://www.nature.com/articles/nature17652#citeas>.
- Bode, M., Koppelman, R., Teuber, L., Hagen, W. & Auel, H. (2018) Carbon Budgets of Mesozooplankton Copepod Communities in the Eastern Atlantic Ocean-Regional and Vertical Patterns Between 24°N and 21°S. *Global Biogeochemical Cycles*, **32**, 840–857.
- Boettiger, C., Lang, D.T. & Wainwright, P. (2012) rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of fish biology*.
- Bogorov, V. G., Vinogradov, M.D., Varonina, N.M., Kanaeva, I.P. & Suetova, I.A. (1968) Distribution of zooplankton biomass within the surficial layer of the world ocean. *Dokl. Akad. Nauk. USSR*, 1205–1207.
- Boldina, I. & Beninger, P.G. (2016) Strengthening statistical usage in marine ecology: Linear regression. *Journal of Experimental Marine Biology and Ecology*, **474**, 81–91.
- Boltovskoy, D., Correa, N. & Boltovskoy, A. (2002) Marine zooplanktonic diversity: a view from the South Atlantic. *Oceanologica Acta*, **25**, 271–278.
- Bosch, S., Tyberghein, L., Deneudt, K., Hernandez, F. & Clerck, O. de (2018) In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. *Diversity and Distributions*, **24**, 144–157.
- Boyd, P.W. (2015) Toward quantifying the response of the oceans' biological pump to climate change. *Frontiers in Marine Science*, **2**.
- Boyd, P.W., Claustre, H., Levy, M., Siegel, D.A. & Weber, T. (2019) Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, **568**, 327–335.
- Boyd, P.W. & Doney, S.C., eds. (2003) *Impact of climate change and feedback processes on the ocean carbon cycle*. in Ocean Biogeochemistry, Springer, New York. pp. 157–193.
- Breiman, L. (1984) *Classification and Regression Trees*. Routledge.
- Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5–32.
- Brown, H.J., Gillooly, J., Allen, P.A., van Savage, M. & West, B.G. (2004) Toward a metabolic theory of ecology, **85**, 1771–1789.
- Brunn, A.F., Greve, S.V., Mielchie, H. & Sparck, R., eds. (1956) *The Galathea Deep Sea Expedition, 1950-1952, described by members of the expedition*. Macmillan, 1956., London.
- Bucklin, A., Ortman, B.D., Jennings, R.M., Nigro, L.M., Sweetman, C.J., Copley, N.J., Sutton, T. & Wiebe, P.H. (2010) A “Rosetta Stone” for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 2234–2247.
- Buesseler, K.O. (1998) The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, **12**, 297–310.
- Buitenhuis, E.T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S.C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R. & Swan, C. (2013) MAREDAT: towards a world atlas of MARine Ecosystem DATA. *Earth System Science Data*, **5**, 227–239.
- Burkenroad, M.D. (1937) The Templeton Crocker Expedition. XII. Sergestidae (Crustacea Decapoda) from the Lower Californian region, with descriptions of two new species and some remarks on the Organs of Pesta in Sergestes. *Zoologica, New York.*, **22**, 315–329.
- Busby, J.R. (1991) BIOCLIM - a bioclimate analysis and prediction system. *Plant Protection Quarterly*, **6**, 8–9.

- Cartes, J.E. (1998) Dynamics of the bathyal Benthic Boundary Layer in the northwestern Mediterranean: depth and temporal variations in macrofaunal– megafaunal communities and their possible connections within deep-sea trophic webs. *Progress in Oceanography*, **41**, 111–139.
- Cartes, J.E., Fanelli, E., López-Pérez, C. & Lebrato, M. (2013) Deep-sea macroplankton distribution (at 400 to 2300m) in the northwestern Mediterranean in relation to environmental factors. *Journal of Marine Systems*, **113-114**, 75–87.
- Cartes, J.E., Maynou, F., Fanelli, E., Romano, C., Mamouridis, V. & Papiol, V. (2009) The distribution of megabenthic, invertebrate epifauna in the Balearic Basin (western Mediterranean) between 400 and 2300 m: Environmental gradients influencing assemblages composition and biomass trends. Environmental gradients influencing assemblages composition and biomass trends. *Journal of Sea Research*, **61**, 244–257.
- Casula, P., Wilby, A. & Thomas, M.B. (2006) Understanding biodiversity effects on prey in multi-enemy systems. *Ecology letters*, **9**, 995–1004.
- Cavan, E.L., Trimmer, M., Shelley, F. & Sanders, R. (2017) Remineralization of particulate organic carbon in an ocean oxygen minimum zone. *Nature communications*, **8**, 14847.
- Charnock, H. & Deacon, G., eds. (1978) *Advances in Oceanography*. Springer US, Boston, MA.
- Chaudhary, C., Saeedi, H. & Costello, M.J. (2016) Bimodality of Latitudinal Gradients in Marine Species Richness. *Trends in ecology & evolution*, **31**, 670–676.
- Cheney, J. (1985) Spatial and temporal abundance patterns of oceanic chaetognaths in the western North Atlantic—II. Vertical distributions and migrations. *Deep Sea Research Part A. Oceanographic Research Papers*, **32**, 1061–1075.
- Choy, C.A., Popp, B.N., Hannides, C.C. S. & Drazen, J.C. (2015) Trophic structure and food resources of epipelagic and mesopelagic fishes in the North Pacific Subtropical Gyre ecosystem inferred from nitrogen isotopic compositions. *Limnology and Oceanography*, **60**, 1156–1171.
- Chun, C. (1888) Die pelagische Tierwelt in grossen Meerestiefen und ihre Beziehungen zu der Oberflachenfauna. *Bibliogr. Zool*, **1**.
- Clavel-Henry, M., Piroddi, C., Quattrocchi, F., Macias, D. & Christensen, V. (2020) Spatial Distribution and Abundance of Mesopelagic Fish Biomass in the Mediterranean Sea. *Frontiers in Marine Science*, **7**, 1136.
- Condon, R.H., Graham, W.M., Duarte, C.M., Pitt, K.A., Lucas, C.H., Haddock, S.H., Sutherland, K.R., Robinson, K.L., Dawson, M.N., Decker, M.B., Mills, C.E., Purcell, J.E., Malej, A., Mianzan, H., Uye, S., Gelcich, S. & Madin, L.P. (2012) Questioning the Rise of Gelatinous Zooplankton in the World's Oceans. *BioScience*, **62**, 160–169.
- Cook, A.B., Sutton, T.T., Galbraith, J.K. & Vecchione, M. (2013) Deep-pelagic (0–3000m) fish assemblage structure over the Mid-Atlantic Ridge in the area of the Charlie-Gibbs Fracture Zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, **98**, 279–291.
- Costello, M.J. & Chaudhary, C. (2017) Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current biology : CB*, **27**, R511-R527.
- Costello, M.J., Wilson, S. & Houlding, B. (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic biology*, **61**, 871–883.
- Cox, C.B. & Moore, P.D., eds. (2000) *Biogeography: an ecological and evolutionary approach. Sixth edition*.
- Crisp, M.D., Laffan, S., Linder, H.P. & Monro, A. (2001) Endemism in the Australian flora. *Journal of Biogeography*, **28**, 183–198.
- Dalpadado, P., Ingvaldsen, R. & Hassel, A. (2003) Zooplankton biomass variation in relation to climatic conditions in the Barents Sea. *Polar Biology*, **26**, 233–241.
- Darnis, G. & Fortier, L. (2014) Temperature, food and the seasonal vertical migration of key arctic copepods in the thermally stratified Amundsen Gulf (Beaufort Sea, Arctic Ocean). *Journal of Plankton Research*, **36**, 1092–1108.

- Davison, P., Lara-Lopez, A. & Anthony Koslow, J. (2015) Mesopelagic fish biomass in the southern California current ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, **112**, 129–142.
- Dipper, F. (2022) The seawater environment and ecological adaptations. *Elements of Marine Ecology*. Elsevier, pp. 37–151.
- Dolan, J.R. (2021) Pioneers of plankton research: Victor Hensen (1835–1924). *Journal of Plankton Research*, **43**, 507–510.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R. G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- Dornelas, M., Antão, L.H., Moyes, F., et al. (2018) BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, **27**, 760–786.
- 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, A.M. P., Picheral, M., Rogge, A., Waite, A.M., Stemmann, L. & Kiko, R. (2022) Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning. *Frontiers in Marine Science*, **9**.
- Eastman, J.T. (1997) Comparison of the Antarctic and Arctic fish faunas. *Cybiurn (Paris)*, **21**, 335–352.
- Ehrenberg, C.G. (1844) *Einige vorläufige Resultate der Untersuchungen der von der Sudpolarreise des Capitain Ross, so wie in den Herrn Schayer und Darwin zugekommenen Materialien*. Monasberg: Preuss. Akad. Wiss.
- Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- Eppley, R.W. & Peterson, B.J. (1979) Particulate organic matter flux and planktonic new production in the deep ocean. *Nature*, **282**, 677–680.
- Evans, K.L., Greenwood, J.J. D. & Gaston, K.J. (2005) Dissecting the species-energy relationship. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2155–2163.
- FAO (1972) *Atlas of the Living Resources of the Seas*.
- Fasham, M.J., Balino, B.M., Bowles, M.C., et al. (2001) A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS). *AMBIO*, 4–31.
- Fay, A.R. & McKinley, G.A. (2014) Global open-ocean biomes: mean and temporal variability. Mean and temporal variability. *Earth System Science Data*, **6**, 273–284.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Finke, D.L. & Snyder, W.E. (2008) Niche partitioning increases resource exploitation by diverse communities. *Science (New York, N.Y.)*, **321**, 1488–1490.
- Flessa, K.W. & Jablonski, D. (1995) Biogeography of recent marine bivalve molluscs and its implications for paleobiogeography and the geography of extinction: A progress report. *Historical Biology*, **10**, 25–47.
- Forbes, E., ed. (1859) *The natural history of the European seas.*, London.
- Freer, J.J. & Hobbs, L. (2020) DVM: The World’s Biggest Game of Hide-and-Seek. *Frontiers for Young Minds*, **8**.
- Froese, R. & Pauly, D. (2022) FishBase. [www.fishbase.org](http://www.fishbase.org), ( 06/2022 ).
- Fujioka, E., Berghe, E.V., Donnelly, B., Castillo, J., Cleary, J., Holmes, C., McKnight, S. & Halpin, P. (2012) Advancing Global Marine Biogeography Research with Open-source GIS Software and Cloud Computing. *Transactions in GIS*, **16**, 143–160.
- Gaard, E., Gislason, A., Falkenhaus, T., Sjøiland, H., Musaeva, E., Vereshchaka, A. & Vinogradov, G. (2008) Horizontal and vertical copepod distribution and abundance on the Mid-Atlantic Ridge in June 2004. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 59–71.

- Gagné, T.O., Reygondeau, G., Jenkins, C.N., Sexton, J.O., Bograd, S.J., Hazen, E.L. & van Houtan, K.S. (2020) Towards a global understanding of the drivers of marine and terrestrial biodiversity. *PloS one*, **15**, e0228065.
- Ghilarov, A.M. & Timonin, A.G. (1972) Relations between Biomass and Species Diversity in Marine and Freshwater Zooplankton Communities. *Oikos*, **23**, 190.
- Gibbons, M.J., Macpherson, E. & Barange, M. (1994) Some observations on the pelagic decapod *Pasiphaea semispinosa* Holthuis 1951 in the Benguela upwelling system. *South African Journal of Marine Science*, **14**, 59–67.
- Giering, S.L. C., Sanders, R., Lampitt, R.S., Anderson, T.R., Tamburini, C., Boutrif, M., Zubkov, M.V., Marsay, C.M., Henson, S.A., Saw, K., Cook, K. & Mayor, D.J. (2014) Reconciliation of the carbon budget in the ocean's twilight zone. *Nature*, **507**, 480–483.
- Gjosaeter, J. & Kawaguchi, K. (1980) *A review of the world resources of mesopelagic fish*, FAO Fisheries Technical Paper 193. FAO, Rome.
- Gloeckler, K., Choy, C.A., Hannides, C.C. S., Close, H.G., Goetze, E., Popp, B.N. & Drazen, J.C. (2018) Stable isotope analysis of micronekton around Hawaii reveals suspended particles are an important nutritional source in the lower mesopelagic and upper bathypelagic zones. *Limnology and Oceanography*, **63**, 1168–1180.
- Gluchowska, M., Trudnowska, E., Goszczko, I., Kubiszyn, A.M., Blachowiak-Samolyk, K., Walczowski, W. & Kwasniewski, S. (2017) Variations in the structural and functional diversity of zooplankton over vertical and horizontal environmental gradients en route to the Arctic Ocean through the Fram Strait. *PloS one*, **12**, e0171715.
- Griffiths, F.B. & Brandt, B.S. (1983) Distribution of mesopelagic decapod Crustacea in and around a warm-core eddy in the Tasman Sea. *Marine Ecology Progress Series*, **12**, 175–184.
- Grime, J.P. (1973) Competitive Exclusion in Herbaceous Vegetation. *Nature*, **242**, 344–347.
- Grimmett, L., Whitsed, R. & Horta, A. (2020) Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. *Ecological Modelling*, **431**, 109194.
- Grosberg, R.K., Vermeij, J.G. & Wainwright, C.P. (2012) Biodiversity in water and on land., **22**, R900-R903.
- Guidi, L., Chaffron, S., Bittner, L., et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, **532**, 465-+.
- Guidi, L., Jackson, G.A., Stemann, L., Miquel, J.C., Picheral, M. & Gorsky, G. (2008) Relationship between particle size distribution and flux in the mesopelagic zone. *Deep Sea Research Part I: Oceanographic Research Papers*, **55**, 1364–1374.
- Guo, Q. (2007) The diversity–biomass–productivity relationships in grassland management and restoration. *Basic and Applied Ecology*, **8**, 199–208.
- Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, **4**, 627–635.
- Halfter, S., Cavan, E.L., Swadling, K.M., Eriksen, R.S. & Boyd, P.W. (2020) The Role of Zooplankton in Establishing Carbon Export Regimes in the Southern Ocean – A Comparison of Two Representative Case Studies in the Subantarctic Region. *Frontiers in Marine Science*, **7**.
- Hao, T., Elith, J., Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2019) A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, **25**, 839–852.
- Hastie, T.J., ed. (2017) *Statistical Models in S*. Routledge.
- Hatton, I.A., Heneghan, R.F., Bar-On, Y.M. & Galbraith, E.D. (2021) The global ocean size spectrum from bacteria to whales. *Science advances*, **7**, eabh3732.
- Hays, G.C. (2003) A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia*, **503**, 163–170.
- Head, R.N. (1999) A comparative study of size-fractionated mesozooplankton biomass and grazing in the North East Atlantic. *Journal of Plankton Research*, **21**, 2285–2308.

- Hedgpeth, J.W. (1957) Chapter 2: Classification of Marine Environments. 67VI. Geological Society of America, pp. 17–28.
- Hempel, G. (1985) On the biology of polar seas, particularly the Southern Ocean. *Marine biology of polar regions and effects of stress on marine organisms* (J S Gray, M E Christiansen, eds ) Wiley, Chichester.
- Henson, S.A., Cael, B.B., Allen, S.R. & Dutkiewicz, S. (2021) Future phytoplankton diversity in a changing climate. *Nature communications*, **12**, 5372.
- Hernandez-Leon, S. (2004) A global assessment of mesozooplankton respiration in the ocean. *Journal of Plankton Research*, **27**, 153–158.
- Hernández-León, S., Koppelman, R., Fraile-Nuez, E., Bode, A., Mompeán, C., Irigoien, X., Olivar, M.P., Echevarría, F., Fernández de Puellas, M.L., González-Gordillo, J.I., Cózar, A., Acuña, J.L., Agustí, S. & Duarte, C.M. (2020) Large deep-sea zooplankton biomass mirrors primary production in the global ocean. *Nature communications*, **11**, 6048.
- Herring, P., ed. (2002) *The biology of the deep ocean*.
- Hobohm, C. (2014) *Endemism in vascular plants*. Springer, Dordrecht.
- Huston, M. (1979) A General Hypothesis of Species Diversity. *The American Naturalist*, **113**, 81–101.
- Hutchinson, G.E. (1961) The paradox of the plankton, **95**, 137–145.
- Ingvaldsen, R.B., Eriksen, E., Gjøsæter, H., Engås, A., Schuppe, B.K., Assmann, K.M., Cannaby, H., Dalpadado, P. & Bluhm, B.A. (2023) Under-ice observations by trawls and multi-frequency acoustics in the Central Arctic Ocean reveals abundance and composition of pelagic fauna. *Scientific reports*, **13**, 1000.
- Irigoien, X., Klevjer, T.A., Røstad, A., Martinez, U., Boyra, G., Acuña, J.L., Bode, A., Echevarria, F., Gonzalez-Gordillo, J.I., Hernandez-Leon, S., Agustí, S., Aksnes, D.L., Duarte, C.M. & Kaartvedt, S. (2014) Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nature communications*, **5**, 3271.
- Ives, A.R., Cardinale, B.J. & Snyder, W.E. (2005) A synthesis of subdisciplines: predator-prey interactions, and biodiversity and ecosystem functioning. *Ecology letters*, **8**, 102–116.
- Ives, A.R., Klug, J.L. & Gross, K. (2000) Stability and species richness in complex communities. *Ecology letters*, **3**, 399–411.
- Jan Holstein (2018) worms: Retriving Aphia Information from World Register of Marine Species. <https://CRAN.R-project.org/package=worms>.
- Jerez-Guerrero, M., Giraldo, A. & Criales-Hernandez, M.I. (2022) Temporal changes in the epipelagic copepod assemblage at Gorgona Island, Colombian Eastern Tropical Pacific Ocean. *REGIONAL STUDIES IN MARINE SCIENCE*, **52**.
- Jones, M.C. & Cheung, W.W. L. (2015) Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES Journal of Marine Science*, **72**, 741–752.
- Jones, M.C., Dye, S.R., Pinnegar, J.K., Warren, R. & Cheung, W.W. (2012) Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling*, **225**, 133–145.
- Kampa, E.M. & Boden, b. (1954) Submarine illumination and the twilight movements of a sonic scattering layer. *Nature*, **174**, 869–871.
- Kaschner, K., Kesner-Reyes, K., Garilao, C., Segschneider, J., Rius-Barile, J.T. & Froese, R. (2019) *AquaMaps: Predicted range maps for aquatic species*.
- Kelly, T.B., Davison, P.C., Goericke, R., Landry, M.R., Ohman, M.D. & Stukel, M.R. (2019) The Importance of Mesozooplankton Diel Vertical Migration for Sustaining a Mesopelagic Food Web. *Frontiers in Marine Science*, **6**, 2083.
- Kier, G. & Barthlott, W. (2001) Measuring and mapping endemism and species richness: a new methodological approach and its application on the flora of Africa. *Biodiversity and Conservation*, **10**, 1513–1529.
- Kjørboe, T. (2013) Zooplankton body composition. *Limnology and Oceanography*, **58**, 1843–1850.

- Kitamura, M., Kobari, T., Honda, M.C., Matsumoto, K., Sasaoka, K., Nakamura, R. & Tanabe, K. (2016) Seasonal changes in the mesozooplankton biomass and community structure in subarctic and subtropical time-series stations in the western North Pacific. *Journal of Oceanography*, **72**, 387–402.
- Klevjer, T.A., Torres, D.J. & Kaartvedt, S. (2012) Distribution and diel vertical movements of mesopelagic scattering layers in the Red Sea. *Marine Biology*, **159**, 1833–1841.
- Kobari, T., Kitamura, M., Minowa, M., Isami, H., Akamatsu, H., Kawakami, H., Matsumoto, K., Wakita, M. & Honda, M.C. (2013) Impacts of the wintertime mesozooplankton community to downward carbon flux in the subarctic and subtropical Pacific Oceans. *Deep Sea Research Part I: Oceanographic Research Papers*, **81**, 78–88.
- Kobari, T., Steinberg, D.K., Ueda, A., Tsuda, A., Silver, M.W. & Kitamura, M. (2008) Impacts of ontogenetically migrating copepods on downward carbon flux in the western subarctic Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 1648–1660.
- Koplin, J. (2020) *Community composition of epipelagic zooplankton in the Eurasian Basin 2017 determined by ZooScan image analysis*. Thesis, University of Hamburg.
- Koppelman, R. & Frost, J. (2008) *The ecological role of zooplankton in the twilight and dark zones of the ocean*. Chapter 2. Nova Science; Lancaster: Gazelle [distributor], Hauppauge, N.Y.
- Koslow, J.A., Goericke, R., Lara-Lopez, A. & Watson, W. (2011) Impact of declining intermediate-water oxygen on deepwater fishes in the California Current. *Marine Ecology Progress Series*, **436**, 207–218.
- Kosobokova, K. & Hirche, H.-J. (2009) Biomass of zooplankton in the eastern Arctic Ocean – A base line study. *Progress in Oceanography*, **82**, 265–280.
- Kruse, S., Brey, T. & Bathmann, U. (2010) Role of midwater chaetognaths in Southern Ocean pelagic energy flow. *Marine Ecology Progress Series*, **416**, 105–113.
- Kumar, R. (2005) Feeding Modes and Associated Mechanisms in Zooplankton. *Ecology of Plankton*.
- Kwong, L. (2016) *A novel approach to estimate active carbon flux using the micronekton biomass spectra*. MSc, University of British Columbia, Vancouver, Canada.
- Kwong, L.E. & Pakhomov, E.A. (2017) Assessment of active vertical carbon transport: new methodology. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki*, **159**, 492–509.
- Lampert, W. (1962) The Relationship between Zooplankton Biomass and Grazing: A Review. *Limnologica* (ed. by S.E.B. Abella, D. Barthelmes, A. Benndorf, et al.). De Gruyter, Berlin, Boston, pp. 11–20.
- Landry, M.R. & Swalethorp, R. (2022) Mesozooplankton biomass, grazing and trophic structure in the bluefin tuna spawning area of the oceanic Gulf of Mexico. *Journal of Plankton Research*, **44**, 677–691.
- Le Borgne, R. (2003) Mesozooplankton biomass and composition in the equatorial Pacific along 180°. *Journal of Geophysical Research*, **108**.
- Lenz, J., Morales, A. & Gunkel, J. (1993) Mesozooplankton standing stock during the North Atlantic spring bloom study in 1989 and its potential grazing pressure on phytoplankton: a comparison between low, medium and high latitudes. *Deep Sea Research Part II: Topical Studies in Oceanography*, **40**, 559–572.
- Letessier, T.B., Falkenhaus, T., Debes, H., Bergstad, O.A. & Brierley, A.S. (2011) Abundance patterns and species assemblages of euphausiids associated with the Mid-Atlantic Ridge, North Atlantic. *Journal of Plankton Research*, **33**, 1510–1525.
- Letessier, T.B., Grave, S. de, Boersch-Supan, P.H., Kemp, K.M., Brierley, A.S. & Rogers, A.D. (2017) Seamount influences on mid-water shrimps (Decapoda) and gnathophausiids (Lophogastridea) of the South-West Indian Ridge. *Deep Sea Research Part II: Topical Studies in Oceanography*, **136**, 85–97.
- Longhurst, A.R. (2007) *Ecological geography of the sea*, 2nd edn. Academic Press, Burlington, MA.
- Machida, R.J., Hashiguchi, Y., Nishida, M. & Nishida, S. (2009) Zooplankton diversity analysis through single-gene sequencing of a community sample. *BMC genomics*, **10**, 438.
- Mapstone, G. (2015) Correction: Global diversity and review of Siphonophorae (Cnidaria: Hydrozoa). *PLoS one*, **10**, e0118381.
- Marmion, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009a) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, **220**, 3512–3520.



- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009b) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- Marrs, R.H., Grace, J.B. & Gough, L. (1996) On the Relationship between Plant Species Diversity and Biomass: A Comment on a Paper by Gough, Grace and Taylor. *Oikos*, **75**, 323.
- Martin, J.H., Knauer, G.A., Karl, D.M. & Broenkow, W.W. (1987) VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, **34**, 267–285.
- McCullagh, P. (2019) *Generalized linear models*. Routledge, London.
- McIvor, I. (2011) *A comparison of the geographical and vertical distribution of mesozooplankton communities from 0 to 1000 m at a global scale*. MSc, University of British Columbia, Vancouver, Canada.
- McNamara, J.C., Moreira, G.S. & Moreira, P.S. (1983) The effect of salinity on respiratory metabolism, survival and moulting in the first zoea of *Macrobrachium amazonicum* (Heller) (Crustacea, Palaemonidae). *Hydrobiologia*, **101**, 239–242.
- Moriarty, R. & O'Brien, T.D. (2013) Distribution of mesozooplankton biomass in the global ocean. *Earth System Science Data*, **5**, 45–55.
- Murphy, E.J., Cavanagh, R.D., Drinkwater, K.F., Grant, S.M., Heymans, J.J., Hofmann, E.E., Hunt, G.L. & Johnston, N.M. (2016) Understanding the structure and functioning of polar pelagic ecosystems to predict the impacts of change. *Proceedings. Biological sciences*, **283**.
- Murray, J. (1895) *Summary of the scientific results, obtained at the sounding, dredging and trawling stations of "Challenger"*.
- Murray, J. & Lee, G.V. (1909) Reports on the scientific results of the expedition of the "Albatross" to the tropical Pacific, 1899-1900. *XII. Memoirs of the Museum of Comparative Zoology at Harvard College*, **38**.
- NOAA National Centers for Environmental Information (2012) *Monthly Global Climate Report for Annual 2011*, published online.
- Noone, K.J., Diaz, R.J. & Sumaila, U.R. (2013) Managing ocean environments in a changing climate. <https://www.sei.org/publications/?pid=2064>.
- Ojeda, V., Serra, B., Lagares, C., Rojo-Francàs, E., Sellés, M., Marco-Herrero, E., García, E., Farré, M., Arenas, C., Abelló, P. & Mestres, F. (2022) Interannual fluctuations in connectivity among crab populations (*Liocarcinus depurator*) along the Atlantic-Mediterranean transition. *Scientific reports*, **12**, 9797.
- Olivar, M.P., Hulley, P.A., Castellón, A., Emelianov, M., López, C., Tuset, V.M., Contreras, T. & Molí, B. (2017) Mesopelagic fishes across the tropical and equatorial Atlantic: Biogeographical and vertical patterns. *Progress in Oceanography*, **151**, 116–137.
- Oliver, M.J. & Irwin, A.J. (2008) Objective global ocean biogeographic provinces. *Geophysical Research Letters*, **35**, 60.
- Omand, M.M., Govindarajan, R., He, J. & Mahadevan, A. (2020) Sinking flux of particulate organic matter in the oceans: Sensitivity to particle characteristics. *Scientific reports*, **10**, 5582.
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I. & Louzao, M. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, **156**, 94–104.
- Palomares, M. & Pauly, D. (2022) SeaLifeBase. [www.sealifebase.org](http://www.sealifebase.org), version (04/2022).
- Pauly, D., Piroddi, C., Hood, L., Bailly, N., Chu, E., Lam, V., Pakhomov, E.A., Pshenichnov, L.K., Radchenko, V.I. & Palomares, M.L. D. (2021) The Biology of Mesopelagic Fishes and Their Catches (1950–2018) by Commercial and Experimental Fisheries. *Journal of Marine Science and Engineering*, **9**, 1057.
- Peña, M., Olivar, M.P., Balbín, R., López-Jurado, J.L., Iglesias, M. & Miquel, J. (2014) Acoustic detection of mesopelagic fishes in scattering layers of the Balearic Sea (western Mediterranean). *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 1186–1197.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

- Poletaev, V.A., Gorchinsky, K.V., Flinn, A.A. & Chumakov, A.K. (1991) *Preliminary results of studies on lanternfishes from the Atlantic and south Ocean*. PINRO, Murmansk.
- Polikar, R. (2006) Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, **6**, 21–45.
- Pollock, L.J., Thuiller, W. & Jetz, W. (2017) Large conservation gains possible for global biodiversity facets. *Nature*, **546**, 141–144.
- Preciado, I., Cartes, J.E., Punzón, A., Frutos, I., López-López, L. & Serrano, A. (2017) Food web functioning of the benthopelagic community in a deep-sea seamount based on diet and stable isotope analyses. *Deep Sea Research Part II: Topical Studies in Oceanography*, **137**, 56–68.
- Proud, R., Cox, M.J. & Brierley, A.S. (2017) Biogeography of the Global Ocean's Mesopelagic Zone. *Current biology : CB*, **27**, 113–119.
- Proud, R., Cox, M.J., Le Guen, C. & Brierley, A.S. (2018) Fine-scale depth structure of pelagic communities throughout the global ocean based on acoustic sound scattering layers. *Marine Ecology Progress Series*, **598**, 35–48.
- Provoost, P. & Bosch, S. (2021) robis: Ocean Biodiversity Information System (OBIS) Client. <https://CRAN.R-project.org/package=robis>.
- Puelles, F.d., Gazá, Cabanellas-Reboredo, Santandreu, Irigoien, González-Gordillo & Duarte (2019) Zooplankton Abundance and Diversity in the Tropical and Subtropical Ocean. *Diversity*, **11**, 203.
- Pusch, C., Hulley, P.A. & Kock, K.-H. (2004) Community structure and feeding ecology of mesopelagic fishes in the slope waters of King George Island (South Shetland Islands, Antarctica). *Deep Sea Research Part I: Oceanographic Research Papers*, **51**, 1685–1708.
- R Foundation for Statistical Computing, Vienna, Austria (2022) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramirez-Llodra, E., Brandt, A., Danovaro, R., Mol, B. de, Escobar, E., German, C.R., Levin, L.A., Martinez Arbizu, P., Menot, L., Buhl-Mortensen, P., Narayanaswamy, B.E., Smith, C.R., Tittensor, D.P., Tyler, P.A., Vanreusel, A. & Vecchione, M. (2010) Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. *Biogeosciences*, **7**, 2851–2899.
- Raymont, J.E. G. (1984) Plankton and Productivity in the Oceans. *Journal of the Marine Biological Association of the United Kingdom*, **64**, 971.
- Reid, J.L. (1962) On circulation, phosphate-phosphorus content, and zooplankton volumes in the upper part of the Pacific Ocean. *Limnology and Oceanography*, **7**, 287–306.
- Reygondeau, G. & Beaugrand, G. (2011) Future climate-driven shifts in distribution of *Calanus finmarchicus*. *Global Change Biology*, **17**, 756–766.
- Reygondeau, G. & Dunn, D. (2019) Pelagic Biogeography. *Encyclopedia of Ocean Sciences* (ed. by J.K. Cochran, H.J. Bokuniewicz and P.L. Yager). Elsevier Science & Technology, San Diego, pp. 588–598.
- Reygondeau, G., Guidi, L., Beaugrand, G., Henson, S.A., Koubbi, P., MacKenzie, B.R., Sutton, T.T., Fioroni, M. & Maury, O. (2018) Global biogeochemical provinces of the mesopelagic zone. *Journal of Biogeography*, **45**, 1–15.
- Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D. & Maury, O. (2013) Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles*, **27**, 1046–1058.
- Righetti, D., Vogt, M., Gruber, N., Psomas, A. & Zimmermann, N.E. (2019) Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science advances*, **5**, eaau6253.
- Righetti, D., Vogt, M., Gruber, N. & Zimmermann, N.E. (2023) *Mapping global marine biodiversity under sparse data conditions*.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. & Müller, M. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**, 77.
- Robinson, C., Steinberg, D.K., Anderson, T.R., Arístegui, J., Carlson, C.A., Frost, J.R., Ghiglione, J.-F., Hernández-León, S., Jackson, G.A., Koppelman, R., Quéguiner, B., Ragueneau, O., Rassoulzadegan, F., Robison, B.H., Tamburini, C., Tanaka, T., Wishner, K.F. & Zhang, J. (2010) Mesopelagic zone

- ecology and biogeochemistry – a synthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 1504–1518.
- Robison, B.H. (2003) What drives the diel vertical migrations of Antarctic midwater fish? *Journal of the Marine Biological Association of the UK*, **83**, 639–642.
- Robison, B.H. (2009) Conservation of deep pelagic biodiversity. *Conservation biology: the journal of the Society for Conservation Biology*, **23**, 847–858.
- Rogers, A.D. (2015) Environmental Change in the Deep Ocean. *Annual Review of Environment and Resources*, **40**, 1–38.
- Rogers, A.D., Appeltans, W., Assis, J., et al. (2022) Discovering marine biodiversity in the 21st century. *Advances in marine biology*, **93**, 23–115.
- Rohde, K. (1978) Latitudinal gradients in species diversity and their causes. II. Marine parasitological evidence for a time hypothesis. *Biol. Zbl.*, **97**, 405–418.
- Rohde, K. (1992) Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. *Oikos*, **65**, 514.
- Rombouts, I., Beaugrand, G., Ibanez, F., Gasparini, S., Chiba, S. & Legendre, L. (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings. Biological sciences*, **276**, 3053–3062.
- Russell, F.S. (1927) The vertical distribution of plankton in the sea. *Biological Reviews*, **2**, 213–262.
- Sayre, R., Wright, D., Breyer, S., Butler, K., van Graafeiland, K., Costello, M., Harris, P., Goodin, K., Guinotte, J., Basher, Z., Kavanaugh, M., Halpin, P., Monaco, M., Cressie, N., Aniello, P., Frye, C. & Stephens, D. (2017) A Three-Dimensional Mapping of the Ocean Based on Environmental Data. *Oceanography*, **30**, 90–103.
- Schadeberg, A., Kraan, M., Groeneveld, R., Trilling, D. & Bush, S. (2023) Science governs the future of the mesopelagic zone. <https://www.nature.com/articles/s44183-023-00008-8#citeas>.
- Sieburth, J.M., Smetacek, V. & Lenz, J. (1978) Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. *Limnology and Oceanography*, **23**, 1256–1263.
- Siegel, D.A., Buesseler, K.O., Doney, S.C., Sailley, S.F., Behrenfeld, M.J. & Boyd, P.W. (2014) Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochemical Cycles*, **28**, 181–196.
- Soviadan, Y.D., Benedetti, F., Brandão, M.C., Ayata, S.-D., Irisson, J.-O., Jamet, J.L., Kiko, R., Lombard, F., Gnanidi, K. & Stemmann, L. (2022) Patterns of mesozooplankton community composition and vertical fluxes in the global ocean. *Progress in Oceanography*, **200**, 102717.
- Spalding, M.D., Agostini, V.N., Rice, J. & Grant, S.M. (2012) Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. A biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management*, **60**, 19–30.
- St. John, M.A., Borja, A., Chust, G., Heath, M., Grigorov, I., Mariani, P., Martin, A.P. & Santos, R.S. (2016) A Dark Hole in Our Understanding of Marine Ecosystems and Their Services: Perspectives from the Mesopelagic Community. *Frontiers in Marine Science*, **3**.
- Staby, A., Røstad, A. & Kaartvedt, S. (2011) Long-term acoustical observations of the mesopelagic fish *Maurollicus muelleri* reveal novel and varied vertical migration patterns. *Marine Ecology Progress Series*, **441**, 241–255.
- Steinberg, D.K., Cope, J.S., Wilson, S.E. & Kobari, T. (2008) A comparison of mesopelagic mesozooplankton community structure in the subtropical and subarctic North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 1615–1635.
- Steinberg, D.K. & Landry, M.R. (2017) Zooplankton and the Ocean Carbon Cycle. *Annual review of marine science*, **9**, 413–444.
- Stemmann, L., Hosia, A., Youngbluth, M.J., Sjøiland, H., Picheral, M. & Gorsky, G. (2008a) Vertical distribution (0–1000m) of macrozooplankton, estimated using the Underwater Video Profiler, in different hydrographic regimes along the northern portion of the Mid-Atlantic Ridge. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 94–105.

- Stemmann, L., Youngbluth, M., Robert, K., Hosia, A., Picheral, M., Paterson, H., Ibanez, F., Guidi, L., Lombard, F. & Gorsky, G. (2008b) Global zoogeography of fragile macrozooplankton in the upper 100–1000 m inferred from the underwater video profiler. *ICES Journal of Marine Science*, **65**, 433–442.
- Stramma, L., Brandt, P., Schafstall, J., Schott, F., Fischer, J. & Körtzinger, A. (2008) Oxygen minimum zone in the North Atlantic south and east of the Cape Verde Islands. *Journal of Geophysical Research*, **113**, 829.
- Stramma, L., Schmidtko, S., Levin, L.A. & Johnson, G.C. (2010) Ocean oxygen minima expansions and their biological impacts. *Deep Sea Research Part I: Oceanographic Research Papers*, **57**, 587–595.
- Strömberg, K.P., Smyth, T.J., Allen, J.I., Pitois, S. & O'Brien, T.D. (2009) Estimation of global zooplankton biomass from satellite ocean colour. *Journal of Marine Systems*, **78**, 18–27.
- Sutton, T.T. (2013) Vertical ecology of the pelagic ocean: classical patterns and new perspectives. Classical patterns and new perspectives. *Journal of fish biology*, **83**, 1508–1527.
- Sutton, T.T., Clark, M.R., Dunn, D.C., et al. (2017) A global biogeographic classification of the mesopelagic zone. *Deep Sea Research Part I: Oceanographic Research Papers*, **126**, 85–102.
- Sutton, T.T., Porteiro, F.M., Heino, M., Byrkjedal, I., Langhelle, G., Anderson, C., Horne, J., Sjøiland, H., Falkenhaug, T., Godø, O.R. & Bergstad, O.A. (2008) Vertical structure, biomass and topographic association of deep-pelagic fishes in relation to a mid-ocean ridge system. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 161–184.
- Sverdrup, H.U. (1953) On Conditions for the Vernal Blooming of Phytoplankton. *ICES Journal of Marine Science*, **18**, 287–295.
- Sweetman, C.J., Sutton, T.T., Vecchione, M. & Latour, R.J. (2013) Distribution of the biomass-dominant pelagic fish, *Bathylagus euryops* (Argentiniformes: Microstomatidae), along the northern Mid-Atlantic Ridge. *Deep Sea Research Part I: Oceanographic Research Papers*, **78**, 16–23.
- Symonds, M.R. E. & Moussalli, A. (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, **65**, 13–21.
- Thorpe, S.A. (1996) *Oceanography. an illustrated guide*.
- Thuiller, W., Georges, D., Gueguen, M., Robin, E., Breiner, F. & Lafourcade Bruno (2022) biomod2: Ensemble Platform for Species Distribution Modeling.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V. & Worm, B. (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466**, 1098–1101.
- Tseitlin, V. (1986) Energetics of Deep-Water Communities. *Nauka*.
- Turner, J.T. (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs, **43**, 255–266.
- Valencia, B., Décima, M. & Landry, M.R. (2018) Environmental Effects on Mesozooplankton Size Structure and Export Flux at Station ALOHA, North Pacific Subtropical Gyre. *Global Biogeochemical Cycles*, **32**, 289–305.
- Vallina, S.M., Follows, M.J., Dutkiewicz, S., Montoya, J.M., Cermeno, P. & Loreau, M. (2014) Global relationship between phytoplankton diversity and productivity in the ocean. *Nature communications*, **5**, 4299.
- van der Spoel, S. & Heyman, R.P. (1983) *A Comparative Atlas of Zooplankton*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Vaquer-Sunyer, R. & Duarte, C.M. (2008) Thresholds of hypoxia for marine biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15452–15457.
- Venables, W.N. & Ripley, B.D. (1999) *Modern Applied Statistics with S-PLUS*. Springer New York, New York, NY.
- Vereshchaka, A., Abyzova, G., Lunina, A., Musaeva, E. & Sutton, T. (2016) A novel approach reveals high zooplankton standing stock deep in the sea. *Biogeosciences*, **13**, 6261–6271.

- Vereshchaka, A.L. (1990) The vertical distribution of euphausiids, pelagic decapods and mysids in the benthic layer in the Western Indian Ocean. *Oceanology*, **30**, 94–98.
- Vereshchaka, A.L. & Vinogradov, G.M. (1999) Visual observations of the vertical distribution of plankton throughout the water column above Broken Spur vent field, Mid-Atlantic Ridge. *Deep Sea Research Part I: Oceanographic Research Papers*, **46**, 1615–1632.
- Vinogradov, G.M., Vereshchaka, A.L. & Aleinik, D.L. (2003) Zooplankton Distribution over Hydrothermal Fields of the Mid-Atlantic Ridge. *Oceanology*, **43**, 656–669.
- Vinogradov, G.M., Vereshchaka, A.L., Shushkina, E.A., Arnautov, G.N. & D'yakonov, V. (1997) Vertical Distribution of zooplankton over the Broken Spur Hydrothermal field in the North Atlantic Gyre (29°N, 34°W Water). *Oceanology*, **37**, 502–512.
- Vinogradov, M.E. (1968) *Vertical distribution of the oceanic zooplankton*. NAUKA, Moscow.
- Vinogradov, M.E. & Melnikov, I.A. (1980) *The study of the pelagic ecosystem of the central Arctic basin*. Nauka (science) Press, Moscow.
- Vinogradov, M.E., Vereshchaka, A.L., Shushkina, E.A. & Arnautov, G.N. (1999) Structure of zooplankton communities of the frontal zone between the Gulf Stream and the Labrador Current. *Oceanology*, **39**, 504–514.
- Voronina, N.M. (1984) *Pelagic ecosystems of Southern Ocean*. NAUKA, Moscow, Russia.
- Ward, P., Tarling, G.A. & Thorpe, S.E. (2014) Mesozooplankton in the Southern Ocean: Spatial and temporal patterns from Discovery Investigations. *Progress in Oceanography*, **120**, 305–319.
- Webb, T.J., Vanden Berghe, E. & O'Dor, R. (2010) Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PloS one*, **5**, e10223.
- Wiebe, P.H., Bucklin, A., Kaartvedt, S., Røstad, A. & Blanco-Bercial, L. (2016) Vertical distribution and migration of euphausiid species in the Red Sea. *Journal of Plankton Research*, **38**, 888–903.
- Wishner, K.F., Gowing, M.M. & Gelfman, C. (1998) Mesozooplankton biomass in the upper 1000m in the Arabian Sea: overall seasonal and geographic patterns, and relationship to oxygen gradients. *Deep Sea Research Part II: Topical Studies in Oceanography*, **45**, 2405–2432.
- Wishner, K.F., Outram, D.M., Seibel, B.A., Daly, K.L. & Williams, R.L. (2013) Zooplankton in the eastern tropical north Pacific: Boundary effects of oxygen minimum zone expansion. *Deep Sea Research Part I: Oceanographic Research Papers*, **79**, 122–140.
- WoRMS (2023) *Ilosergestes pestafer* (Burkenroad, 1937). Accessed at: <https://www.marinespecies.org/aphia.php?p=taxdetails&id=514113> on 2023-01-21.
- WoRMS Editorial Board (2022) *World Register of Marine Species*. Available from <https://www.marinespecies.org> at VLIZ. Accessed 2023-06-07. VLIZ.
- Xie, F.T., Zhou, X., Tao, Z., Lv, T.T., Wang, J. & Li, R.X. (2021) MODIS-based Monthly Dataset of Particulate Organic Carbon Flux in the Bottom of the Global Ocean Euphotic Layer (2003-2018). *Global Change Data & Discovery*, **5**, 11–18.
- Xue, Y., Hu, Z., Kumar, A., Banzon, V., Smith, T.M. & Rayner, N.A. (2012) [Global oceans] Sea surface temperatures [in “State of the Climate in 2011”]. *Bulletin of the American Meteorological Society*, **93**.
- Youden, W.J. (1950) Index for rating diagnostic tests, **3**, 32–35.
- Youngbluth, M., Sørnes, T., Hosia, A. & Stemmann, L. (2008) Vertical distribution and relative abundance of gelatinous zooplankton, in situ observations near the Mid-Atlantic Ridge. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 119–125.
- Zeldis, J.R. & Décima, M. (2020) Mesozooplankton connect the microbial food web to higher trophic levels and vertical export in the New Zealand Subtropical Convergence Zone. *Deep Sea Research Part I: Oceanographic Research Papers*, **155**, 103146.

# Appendices

## Appendix A - Supplementary information for Chapter 1

Table A1: Summary of recent expeditions in the mesopelagic realm and their major findings

Region	Source of data	Time period	Depth samples	Important findings	Ref.
<b>Global</b>					
Global	The Multiple Unit Large Volume in-situ Filtration System (MULVFS)	1973-2005	1000	This study presents a comprehensive global dataset consisting of 62 open ocean profiles, encompassing measurements of particulate organic carbon (POC), CaCO <sub>3</sub> , and opal concentrations. The investigation reveals that the strength and efficiency of the biological pump exhibit dynamic characteristics, which can be broadly categorized into two regimes. The first regime is characterized by a consistent and high transfer efficiency, particularly observed at low to moderate surface POC levels. In contrast, the second regime, referred to as the bloom regime, exhibits a weak deep biological pump and low transfer efficiency, with the height of the bloom being a defining feature. The variability in POC attenuation can be attributed to the decoupling between the strength of the shallow biological pump and the strength of the deep biological pump. Notably, large diatoms that undergo blooming events display a low mesopelagic transfer efficiency and contribute minimally to deep-sea exportation, whereas smaller cells and those with moderately high CaCO <sub>3</sub> content exhibit a higher mesopelagic transfer efficiency.	(Lam <i>et al.</i> , 2011)
Global	389 cruises conducted from the US National Oceanic Data Center Joint Archive for Shipboard ADCP and from the British Oceanographic Data Center	1990 - 2011	n/a	The depths of these DVMs determined globally follow coherent large-scale patterns: 1) seawater oxygen is a single predictor in migration depth with migration depth is greater where subsurface oxygen concentrations are high. 2) where OMZs are present, migratory animals generally descend as far as the upper margins of the low-oxygen waters. Changing climate will alter the extent of OMZ resulting in a change in the migratory depths of marine organisms. Such changes may influence marine biogeochemistry, food webs, and fisheries.	(Bianchi <i>et al.</i> , 2013)
Global	WOA2005 (World Ocean Atlas, 2005) data	n/a	n/a	Using updated environmental conditions, the four known tropical OMZs have been described (the Eastern South Pacific, Eastern Tropical North Pacific, the Arabian Sea, and the Bay of Bengal). In addition, three new OMZs were described: one permanent one in the Eastern Sub-Tropical North Pacific and two seasonal OMZs in West Bering Sea and the Gulf of Alaska. The total surface of permanent OMZ was calculated to be 30.4 million of km <sup>2</sup> (8% of the total surface area of the ocean). The	(Paulmier & Ruiz-Pino, 2009)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				volume of the OMZ cores was 17 times higher than the previous estimate and was 10.3 million of km <sup>3</sup> . The OMZs, being 3 times larger than the associated denitrification zone, indicates that there is a great imbalance of nitrogen cycle on the global scale.	
Global (global tropical and subtropical oceans)	Hydrobase QC Dataset, CALCOFI Dataset, NOAA Ship <i>Ka'imimoana</i> Dataset, Meteor 28/1, 47/1, 55 and 68/2; NACP West Coast Cruise Reville 322, Ron H. Brown 2005 Sonne 89, Sonne 102 and 128, Thalassa 2000, WOCE P06e RP		200–700-dbar depth	dissolved oxygen differences within the OMZ for two periods (1960–1974 and 1990–2008) were investigated. Lack of dissolved oxygen in coastal ecosystems can result in acute ecosystem perturbations. In most regions of the tropical Pacific, Atlantic, and Indian Oceans the oxygen content in the 200–700-dbar layer has declined. tropical low oxygen zones have expanded in both horizontal and vertical directions in all three tropical oceans between the time periods 1960–1974 and 1990–2008. The area at 200dbar unsuitable for large mobile macro-organisms increased by 4.5 million km <sup>2</sup> . Subsurface oxygen has decreased offshore of most continental shelves. However, there are large areas (subtropical gyres at the depths analyzed) in the subtropics with weak increases in oxygen has increased. Habitat compression and loss of biodiversity of various hypoxia-intolerant taxa is predicted in many regions. The changes in oxygen concentration could accelerate the shifts in animal distributions. major shifts in nutrient cycling and trophic structure triggered by the expansion of tropical OMZs are expected.	(Stramma <i>et al.</i> , 2010), (Stramma <i>et al.</i> , 2008)
Global (The Mediterranean Sea, North Atlantic shelves, Mid-Atlantic Ridge, tropical Pacific Ocean, eastern Indian Ocean, and Subantarctic Ocean)	OIM2006, Barnned, BIOSOPE, KEOPS, Maine2002, MARECO, MARECO2002,POMME 3, SS052006	Sep 2002, Nov 2002, Mar-Jun 2003, Nov 2004, Jun-Jul 2004, Jan-Feb 2005, May 2006, Sep 2006, Oct-Nov 2001	1000	Mesopelagic gelatinous zooplankton fauna diversity, distribution, and abundance were investigated using the underwater video profiler. Community composition was similar within each region and differ significantly between the regions, suggesting that spatial structure of the community occurs on large scales (e.g., basin scales) but not necessarily on smaller scales (e.g., oceanic front).	(Steinberg <i>et al.</i> , 2008)
Global	numerous research and fishing vessels	2005-2008	1,000	Compiled a database of DSL characteristics globally. We show that DSL depth and acoustic backscattering intensity (a measure of biomass) can be modeled using just surface primary productivity, temperature, and wind stress. By 2100 the study predicts the homogenization of	(Proud <i>et al.</i> , 2017)



Region	Source of data	Time period	Depth samples	Important findings	Ref.
				mesopelagic communities and that increase in mesopelagic biomass by approximately 17%.	
Global	Literature review	n/a	n/a	Respiration by pelagic zooplankton is considerably higher than previous estimates at the global scale. As a result, inability to account for these organisms as major participants in a carbon cycle can result in larger uncertainty in the estimates of primary production	(Hernandez-Leon, 2004)
Global	Malaspina 2010, the Spanish Circumnavigation Expedition	December 2010–July 2011	1000	Global mesopelagic fish's biomass is about one order of magnitude higher than previous estimates of 1,000 million tons	(Irigoin <i>et al.</i> , 2014)
Global	Malaspina cruise	December 2010–July 2011	1000	Mesopelagic SLs were observed in all areas sampled areas. strong spatial gradients in mesopelagic DVM patterns. DVM was the deepest in the southern Indian Ocean (with only ~20% of organisms performing migrations) and shallowest at the oxygen minimum zone in the eastern Pacific (migrating proportions was ~90% in the Eastern Pacific).	(Klevjer <i>et al.</i> , 2016)
Global	Malaspina 2010 Expedition on R/V <i>Hespérides</i>	December 2010–July 2011	200	Spatial community structure ( $\beta$ -diversity) for several planktonic and nektonic taxa was investigated. A significant negative relationship, stronger than environmental differences, between $\beta$ -diversity and surface ocean transit time was observed. Estimated dispersal scales for different groups were negatively correlated. Global spatial patterns of diversity are established through the local abundance of the organisms that determine their dispersal scale which it turns scales with body size.	(Villarino <i>et al.</i> , 2018)
Global (Western Tropical Pacific, Eastern Equatorial Pacific, and NW Atlantic)	ROV <i>Hercules</i> , M/Y <i>Alucia</i> , the E/V <i>Nautilus</i> (leg NA064), and the R/V <i>Endeavor</i> (leg EN564)	Sep 2013, Jun/Jul 2015, Nov 2014	2500	A novel lowlight, high-resolution bioluminescent imaging methods (a scientific complementary-metal-oxide-semiconductor microscopy camera) was applied to study marine bioluminescence in the deep sea.	(Phillips <i>et al.</i> , 2016)
<b>Mediterranean and Red Seas</b>					
Northwestern Mediterranean Sea.	Institute of Oceanographic Sciences two different scientific projects (CICYT MAR90/757 and CICYT AMB93/0283);	1990, 1985–1992, four annual seasons between	2300	Vertical distributions and diel migrations of the main species (euphausiids, mysid, decapod, and fishes) of micronekton were investigated. Different behavior patterns were observed in different species: 1) some species performed clear DVMs, 2) others show diurnal symmetry of the migratory cycles, 3) repetitive bimodal day distribution with only part of its population migrating and 4) non-migratory species. Changes in the composition and biomass distribution of deep-living zooplankton over wide gradients of depth and longitude were	(Andersen & Sardou, 1992), (Cartes, 1998)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
	ANTROMARE0710 cruise, on board R/V <i>García del Cid</i> (CSIC);  Cruise BT1; IDEA and BIOMARE projects;  the National project ANTROMARE. BATIMAR from. Fishes were collected onboard R/V <i>García del Cid</i>	1991–1995 Jul 2010 Mar 1994  2002–2004, 2007–2008  1987-1989 2010-2014		investigated. The composition of zooplankton changed significantly at the chosen mesoscale (~180 km). There was an increase in biomass of the near-bottom zooplankton coinciding with the lowest temperatures and maximum O <sub>2</sub> concentration.  bathyal mysid, <i>Boromycin arctica</i> , had a low trophic impact of this bathyal species in spite of its numerical dominance in the suprabenthic assemblage of the Catalan Sea. Epibenthic assemblages of the Mediterranean was organized into five groups as a function of increasing depth.  Species richness of western Mediterranean macroplankton (excluding gelatinous taxa) was greater in the period 1991/1992 than in 2007, due to higher values of the North Atlantic Oscillation index in 1991/1992. the degree of stratification and homogenization of the water and longitude was a primary factor influencing species composition.  The contributions of food falls of both natural and anthropogenic origin to the diets of large fish were analyzed. large food falls contribute only a little to fish diets by weight. However, the local importance of food falls (e.g., cetacean blubber and carcharhinid shark remains) was very high near canyons. Remains, originated from human activity, may locally alter the food webs.	(Cartes <i>et al.</i> , 2013) (Cartes & Maynou, 2001) (Cartes <i>et al.</i> , 2010) (Cartes <i>et al.</i> , 2009) (Cartes <i>et al.</i> , 2016)
Balearic Sea (western Mediterranean)	R/V <i>Sarmiento de Gamboa</i> (La Unidad de Tecnología Marina – Consejo Superior de Investigaciones Científicas, UTM–CSIC)	Nov - Dec 2009 and Jul 2010	900	The distribution of micronekton layers was investigated. Acoustic estimates were compared with the sampled abundances. The link between environmental variables and the vertical distribution and migration pattern was analyzed. Preferred depth for migration was found to be related to the thermocline depth, while non-migrating species were found in OMZ.	(Peña <i>et al.</i> , 2014)
Mediterranean	n/a	May 1975; Apr and Jun 1976; Jul and Oct 1977; Feb 1979.	3000	The vertical zonation and community structure of copepods were investigated for different depth intervals and periods of the year. Three major copepod communities were identified primarily based on species composition rather than depth. Temporal changes in community structure were only observed in the surface waters with no change in species composition in deep layers. the low zooplankton biomass in deep layers was due to low standing crops at the surface. the efficiency of the deep Mediterranean subsystem was similar to the open ocean system despite the absence of a true bathypelagic fauna.	(Di Carlo <i>et al.</i> , 1984)
The Mediterranean Sea	Mediterranean Targeted Project II, Mass Transfer	Jul 1997- Mar 1998,	~2400	The first study to investigate the change in deep-sea zooplankton over one annual cycle. It was found that zooplankton assemblages i) change in their abundance and structure over time, ii) have different dominant	(Dano varo <i>et</i>

Region	Source of data	Time period	Depth samples	Important findings	Ref.
	and Ecosystem Response (MTP II MATER)	Apr-May 1998, Sep 1997 - Aug 1998		taxa in different basins, and iii) display clear taxonomic segregation between shallow and near-bottom waters. Biodiversity decreases with depth while the equitability increases. The difference in zooplankton abundance and assemblage structure are likely due to variable trophic condition characterizing the basins.	<i>al.</i> , 2017)
Tyrrhenian Sea (Mediterranean Sea)	'TM3' oceanographic cruise on board the RV 'Universitatis'	Apr 2007	2000	vertical distribution and diel variability of the mesozooplankton carbon requirement was investigated. A noticeable variation between day and night carbon demand was present in top 300m with the higher amounts during nighttime. 13% of carbon loss was due to the mesozooplankton in mesopelagic zone.	(Minutoli & Guglielmo, 2012)
Western Mediterranean	the R/V <i>Sarmiento de Gamboa</i> , and the F/V <i>Punta des Vent</i>	Nov - Dec 2009	850	first study to analyze the vertical migrations of the developmental stages of the pelagic shrimps. Mesopelagic early larvae perform reverse DVM in the non-stratified conditions to escape from the predators that use tactile stimuli prey detection. Deep-spawning for the penaeids was confirmed.	(Torres <i>et al.</i> , 2018)
NW Mediterranean	(1) the MEDITS surveys, (2) the IDEADOS surveys	summer 2007-2016; Dec 2009 -Jul 2010	800	Several most abundant elasmobranch and cephalopod species were analyzed. Ontogenetic shifts in the diet (from pelagic to benthic) were observed in rays, sharks, octopuses, and squids. A clear food partitioning among species was discovered with the mesopelagic prey being an important food resource for deep-sea elasmobranchs and cephalopods. Cephalopods are very important participants in the benthopelagic coupling, while demersal elasmobranchs are main contributors to a one-way flux transferring energy to the deep-sea.	(Valls <i>et al.</i> , 2017)
Aegean Sea	n/a	Mar-Apr 1997 and Sep 1997, Mar 1997	2200	first study to describe vertical distribution of mesozooplankton through the mesopelagic zone during the periods of mixing and stratification in the study area. Biomass and abundance decreased with depth. Due to specific depth preferences by various species, distinct communities were found in different zones (including the north-south gradient of oligotrophy in the epipelagic and deep zones). Seasonal and horizontal differences were less significant in the mesopelagic zone.	(Siokou <i>et al.</i> , 2013)
The southern Adriatic Sea	VECTOR-AM1 and VECTOR-AM3 oceanographic cruises on board the RV 'Universitatis.'	Nov 2006 and Apr 2007	1000	This study compared the zooplankton carbon requirement with organic carbon vertical fluxes. Difference between day and night zooplankton carbon requirement was present at both sampling locations. The difference was not attributed to the vertical distribution pattern of euphausiids but rather to the number of non-living specimens and/or the ratio between gelatinous and crustacean taxa. The mean metabolic requirement was lower in November than in April. The higher remineralization percentages detected in April is linked to seasonal	(Minutoli <i>et al.</i> , 2014)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				differences (higher primary production, zooplankton biomass, vertical amount of sinking POC flux and etc.).	
Levantine Basin of the Eastern Mediterranean	two <i>Meteor</i> cruises (M5/1 and M25/2 )	Jan 1987, Jun 1993	4000	Mesozooplankton samples showed a strong temporal variability in abundance and composition that have not been known previously. Unlike 1987, in 1993 biomass did not decrease exponentially with depth, but rather stayed constant. In addition, the dominance of two interzonal calanoid copepods at bathypelagic and abyssopelagic depths was observed in 1993 that resulted into a substantial faunal change. The encountered difference was linked to the several hydrographical factors associated with Eastern Mediterranean Transient	(Weikert <i>et al.</i> , 2001)
Red Sea	R/V <i>Thuwal</i>	Jan 2014	600	The effect of extreme environmental conditions on vertical distribution and vertical migration of five euphausiid species was investigated. Widely distributed zooplankton species have a broader tolerance range for key environmental variables and have a better adaptation to a variable and changing conditions.	(Wiebe <i>et al.</i> , 2016)
<b>Atlantic Ocean</b>					
North Atlantic	MAR-ECO cruise	Jun 2004	3000	Species composition of the chaetognaths between the Sub-Arctic Province and the Cold Temperate Province and a frontal region were investigated. Chaetognatha can be an indicator of changes in water masses and temperature. There was an inverse relationship between biomass and number of species. No east-west division was found in the Chaetognatha on either side of the Mid-Atlantic Ridge, but chaetognath species assemblages were divided in accordance with the biogeographic provinces by Longhurst (2007).	(Pierrot-Bults, 2008)
North, Central, and South Atlantic	36th cruise of the R/V <i>Akademik Sergey Vavilov</i> and 37th cruise of the R/V <i>Akademik Sergey Vavilov</i>	Oct–Nov 2012, Sep–Oct 2013	3000	First deep-sea quantitative survey of the zooplankton in the area was conducted to analyze the interaction between environment and zooplankton abundance and biomass. Over 300 plankton taxa were identified with >80% belonging to Copepoda. Biomass and abundance were depth-dependent for most of the taxa. The number of taxa was depended on surface productivity; community diversity was strongly correlated with depth. Each of depth strata was characterized by unique copepod assemblages.	(Vereshchaka <i>et al.</i> , 2017)
Central, South, and North Atlantic	36th and 37th cruises of the R/V <i>Akademik Sergey Vavilov</i> and the 34th, 37th, 39th, 42nd, 46th, 47th, 49th, and 50th cruises of	1996-2012	3000	a novel approach to estimate deep-pelagic zooplankton biomass on an ocean basin scale is presented. The wet biomass of the different plankton taxa is affected primarily by chlorophyll concentrations. Zooplankton biomass in the upper bathypelagic was higher than previously thought. macroplanktonic shrimps compromised the majority	(Vereshchaka <i>et al.</i> , 2016)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
	the R/V <i>Akademik Mstislav Keldysh</i>			of this biomass. The contribution of the deep-ocean zooplankton for biogeochemical cycles may be more important than previously thought.	
Atlantic	research cruises Geotraces III, Moca, and Medea I and II.	Mar 2010, Oct 2010, Oct 2011, Jun/Jul 2012	5000	Changes in the phylogenetic composition with depth in relation to changes in the bacterial and archaeal transporter proteins were investigated. extensive metaproteomic and a metagenomic dataset of microbial communities was compiled. Below the euphotic zone, heterotrophic microbes rely largely on the solubilized particulate organic matter as a carbon and energy source.	(Bergauer <i>et al.</i> , 2018)
A latitudinal transect from 24°N to 21°S in the eastern Atlantic Ocean	research cruise ANT-XXIX/1 on board of R/V <i>Polarstern</i>	Nov 2012	2000	Vertical distribution, community structure, and diversity of calanoid copepods were studied. Highest abundance was found in the top 100 m with exponential decrease with depth. In northern- and southernmost stations, a maximum of biomass was observed at upper mesopelagic (200-400 m). Depth was a stronger factor establishing different community structure in different depth zones. Biodiversity maxima were observed in the mesopelagic zone because of various factors such as physical stability of the environment, predator avoidance, or small population sizes.	(Bode <i>et al.</i> , 2018)
North Atlantic gyre	deep-sea manned vessel <i>Mir</i>	Nov 1998, Sep-Nov 1999, Nov-Dec 1998	3000	vertical distribution of <i>meso</i> - and macroplankton at the north and south borders of the gyre was investigated. macroplankton biomass peaked twice through the water column in the main pycnocline layer and at 1000-1200 m depth. Distribution of the deep-sea zooplankton and species composition were affected by the current of intermediate waters.	(Vinogradov <i>et al.</i> , 2000)
northeastern Atlantic (Northwest African coast, the west of Spain)		April-May, 1977-1978	4500	Regardless of geographical location, there were exponential decreases in biomass with depth. The difference in standing stocks was explained by the presence of larger number of gelatinous zooplankton in Northern latitudes. Micronekton standing crops were about a third of the plankton.	(Angel & Baker, 1982)
Northeastern North Atlantic	Marine Institute fisheries science deepwater trawl surveys on board of RV <i>Celtic Explorer</i>	Sep 2006 and Dec 2009.	1800	Mid-water and benthopelagic-feeding demersal fishes are important participants in carbon cycle in the ocean. Carbon is transferred to deep, long-term storage, bypassing the detrital particle flux with >50% of demersal fish biomass is supported by biological flux processes. There was an increase in biomass and diversity of fishes at mid-slope depth explained by a decrease in demersal fish predators. It was estimated that benthopelagic fishes capture and store over 1 million tons of CO <sub>2</sub> annually.	(Trueman <i>et al.</i> , 2014)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
NE Atlantic The Galicia Bank seamount	LIFE+INDEMARES project	2009, 2010 and 2011	1800	Trophic interactions in the deep-sea fish community of the seamount were investigated. Vertically migrating macrozooplankton and shrimps were a major pelagic prey for the deep-sea fishes. high degree of resource partitioning was found in the study area. However, benthopelagic feeders had a high competition. When calculating trophic levels using both the SCA (snapshot of diet) and SIA (assimilated food in previous months) approaches is required due to a number of the discrepancies between the two approaches.	(Preciado <i>et al.</i> , 2017)
NE Atlantic	cruises P384 on RV <i>Poseidon</i> and M83/2, M79/3 on RV <i>Meteor</i>	Sep/Oct 2009, May 2009, Nov/ Dec 2010	4000	the trophic structure of zooplankton and micronekton above and around two shallow seamounts was investigated. Zooplankton and micronekton usually occupied the 1st–3rd trophic level. $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values were lower in zooplankton and micronekton in the subtropical waters than the tropical region, as a result of the different nutrient availability and phytoplankton communities. A linear food chain based on a single energy source from primary production was established. Seamount communities rely to a large extent on advected food sources. Trophic structure and the main prey of benthopelagic fishes were also investigated. Benthopelagic fishes feed on mainly zooplanktivores and mixed feeders, but also benthivores, piscivores, and predator-scavengers. Benthopelagic fishes occupied 2 <sup>nd</sup> -4 <sup>th</sup> trophic positions. A resource partitioning among the benthopelagic fishes was observed.	(Denda <i>et al.</i> , 2017a), (Denda <i>et al.</i> , 2017b)
Masfjorden (Norway), NE Atlantic	RV <i>Håkon Mosby</i> (University of Bergen and Institute of Marine Research).	Nov 2007	500	The global estimate of mesopelagic fish that is commonly referred to in the scientific literature adds up to $948 \times 10^6$ t wet weight (Gjøsæter & Kawaguchi 1980) and was slightly revised to $999 \times 10^6$ t (Lam & Pauly 2005)	(Kaartvedt <i>et al.</i> , 2012)
Mas-fjorden, Norway,	RVs <i>H. Mosby</i> and <i>T. Braarud</i>	Jul 2007 and Oct 2008,	392	The temporal patterns of vertical distribution of <i>Maurolicus muelleri</i> scattering layers (SL) were investigated. Apart from observing “normal” patterns of DVMs, midnight sinking between dusk and dawn, and periods without migrations, new behaviors were discovered such as early morning ascents, reverse diel vertical migrations, and interrupted ascents in the evening. Such findings suggest that <i>M. muelleri</i> can change its behavior in response to several factors (ontogeny, satiation, and hunger or other external stimuli).	(Staby <i>et al.</i> , 2011)
NE Atlantic (Canary Islands)	R/V <i>Cornide de Saavedra</i>	Apr 2012	2000	Here we combine results from acoustic observations at 18 and 38 kHz with limited net sampling to unveil the origin of acoustic phenomena around the Canary Islands, subtropical northeast Atlantic Ocean. Trawling data revealed a high diversity of fishes, decapods, and	(Ariza <i>et al.</i> , 2016)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				cephalopods (152 species), although few dominant species likely were responsible for most of the sound scattering in the region. We identified four different acoustic scattering layers in the mesopelagic realm: (1) at 400–500 m depth, a swimbladder resonance phenomenon at 18 kHz produced by gas-bearing migrant fish such as <i>Vinciguerria</i> spp. and <i>Lobianchia dofleini</i> , (2) at 500–600 m depth, a dense 38 kHz layer resulting primarily from the gas-bearing and non-migrant fish <i>Cyclothone braueri</i> , and to a lesser extent, from fluid-like migrant fauna also inhabiting these depths, (3) between 600 and 800 m depth, a weak signal at both 18 and 38 kHz ascribed either to migrant fish or decapods, and (4) below 800 m depth, a weak non-migrant layer at 18 kHz which was not sampled. All the diel migrating layers reached the epipelagic zone at night, with the shorter-range migrations moving at $4.6 \pm 2.6 \text{ cm s}^{-1}$ and the long-range ones at $11.5 \pm 3.8 \text{ cm s}^{-1}$ . This work reduces uncertainties interpreting standard frequencies in mesopelagic studies, while enhances the potential of acoustics for future research and monitoring of the deep pelagic fauna in the Canary Islands.	
Subtropical North East Atlantic Ocean	museum collections (namely Museu de História Natural do Funchal, MMF), field expeditions (in MMF RV with baited-traps) and literature research	n/a	5000	The first study that investigated the shape of bathymetric diversity gradients and faunal turnover of decapod crustaceans in an island context. all bathymetric gradients of pelagic diversity (order and suborder levels) displayed parabolic trends. The same taxonomic group in one geographic region can have different bathymetric patterns of diversity, depending on the taxonomic level and on the group's life strategies. Study supports the species-energy hypothesis where the interplay between temperature and food availability is the predictor of diversity variation.	(Rosa <i>et al.</i> , 2012)
Saint Peter and Saint Paul Archipelago (the Equator in the Atlantic Ocean)	n/a	May 2008	Subsurface samples	There was a decline in biodiversity and production with increasing distance from the archipelago, meaning, that even small features like the SPSPA can affect the copepod community in tropical oligotrophic oceanic areas.	(Melo <i>et al.</i> , 2014)
Equator and 33uN in the North-Atlantic Ocean	R/V <i>Pelagia</i> and NIOZ-MTM	One and a half years (exact years are	1600	Association of DVM with latitudinal and seasonal daylight variation was investigated. DVM was observed to follow seasonal and latitudinal trends in day length and latitude at all depths. The deepest plankton was consistently the first to migrate upwards, suggesting that another	(van Haren & Compt

Region	Source of data	Time period	Depth samples	Important findings	Ref.
		not specified)		mechanism (rather than light) could maintain the solar diurnal and seasonal rhythms.	on, 2013)
Tropical and equatorial Atlantic	n/a	April 2015	800	Changes in fish species occurrences, relative abundances, and vertical distribution across the tropical and equatorial Atlantic were investigated. Species richness in study areas was higher than in other parts of the Atlantic. Despite the differences in hydrographic features across the tropical and equatorial Atlantic, DVM was a common feature of the mesopelagic zone. Migratory fishes could cross the OMZ during their DVM without any challenges while non-migratory species had a high tolerance to reduce oxygen concentrations and were present n OMZ in high numbers.	(Olivar <i>et al.</i> , 2017)
Eastern transect of North Atlantic	'SOND' cruise of RRS Discovery	1968 - 1974	2000	latitudinal and bathymetric changes in species composition and diversity of assemblages of 117 species of halocyprid ostracods were investigated. Differences between the day and night profiles were due to DVM, local-scale heterogeneity, and to a degree of net avoidance. There was a gradient of increasing species richness and diversity from high to low latitudes. No relationship was found between diversity, abundance, and productivity. Changes in community structure are consistent with Longhurst's biogeochemical provinces.	(Angel <i>et al.</i> , 2007)
Frontal Zone between the Gulf Stream and the Labrador Current	<i>Mir</i> submersible.	Sep 1998,1995	3000	Vertical distribution of meso- and macroplankton was investigated in the area with most sharply pronounced climatic frontal zone. In the frontal zone, 2 different communities (the North Atlantic Subtropical Gyre and arctic-boreal communities) co-occur. The community of the North Atlantic Subtropical Oyre was found to be more mature in terms of succession with pelagic shrimps forming the "living network" feeding on the organisms from the arctic-boreal community. Thus, the biomass of the shrimps was higher than the biomass of their preys.	(Vinogradov <i>et al.</i> , 1999)
Gulf Stream 65w 40 N	RV <i>Knorr</i> cruise 65 and 71, part of a multidisciplinary study of Gulf Stream cold core rings	Apr 1977	1000	Net avoidance of euphausiid crustacean <i>N. megalops</i> by two different nets (1 and 10 m <sup>2</sup> mouth opening) was studied in relation to its sampling distribution in day and night catches. No difference in the vertical pattern of distribution, catch rate, size selectivity was observed between the nets. Nets also gave good agreement in nighttime abundance estimates, but do not in daytime estimates. Night catches exceed day catches especially for the smaller net as a result of greater avoidance during the day. Barkley's avoidance theory (less avoidance of large nets with greater mouth area) was not supported by the field data due to individuals reacting to the approaching net at a greater distance. Thus, size-frequency distributions were not affected by avoidance.	(Wiebe <i>et al.</i> , 1982)



Region	Source of data	Time period	Depth samples	Important findings	Ref.
Western Atlantic	<i>Johnson-Sea-Link</i> submersibles	Oct. and Nov 1984, Oct 1988, and Nov 1989	900	<i>in situ</i> observations on midwater gelatinous zooplankton. Medusa was very abundant at the mesopelagic depth (>500m) and was found to be an important prey source for decapods and cnidarians. However, the guts of mesopelagic medusae were usually empty.	(Larson <i>et al.</i> , 1991)
Warm-core eddy located in The Northwest Atlantic	Northern Atlantic crossing (CE15007) cruise	Apr-May 2015	600	Amount of microplastics in several species of mesopelagic fish was investigated. 73% of all fish had plastics in their guts that were higher than previously reported. The amount of plastic was not affected by fullness, species and the depth at which fish were caught. However, plastics were similar to those found at the surface water.	(Wieczorek <i>et al.</i> , 2018)
Georges Bank NW Atlantic	NOAA research vessel ' <i>Delaware II</i> ' as part of the Census of Marine Life Gulf of Maine Area field project	May 2004	670	Results of a pelagic nekton survey along a transition area between coastal and deep-sea environments were presented. Macrocrustacea was a dominant taxon by numbers and biomass in total nekton community. Gut content of <i>biomass dominant mesopelagic fish</i> revealed a discriminatory feeding behavior with preference on large euphausiid and decapod prey. A major trophic pathway in this area was found to be similar to 'oceanic rim' ecosystems.	(Feagans-Bartow & Sutton, 2014)
the eastern area of the Sargasso Sea on the Corner Rising seamounts	26 scouting and research surveys and commercial vessel cruises were carried out by the USSR and Russia	1976 - 1995	n/a	Overview of the fishery on deepwater fish by the Soviet Union/Russia since 1976 is described. The total catch taken during the period was > 19 000 tons, although limited stock sizes were observed. The description of biological observations on the main species, including spawning and feeding habits, distribution, and formation of deepwater aggregation, and hydrographic conditions is provided.	(Vinnichenko, 1997)
the south-western Atlantic Ocean	n/a	Nov 1993	500	influence of oceanic fronts on mesozooplankton distribution and grazing activity was investigated in the south-west Atlantic. abundance and biomass were 10-fold higher in the Subtropical Confluence Zone and in the Antarctic Polar Front than in subtropical and subantarctic areas outside of those frontal systems. mesozooplankton was found to exert a significant grazing impact on phytoplankton (22-40% of the total chlorophyll stock) over much of the frontal areas studied.	(Lopes <i>et al.</i> , 2016)
Sub-Polar Front and the Charlie-Gibbs Fracture Zone	ECOMAR project, RSS 2007 <i>James Cook</i> cruises JC011 and JC037	Jun-Aug 2007, Aug-Sep 2009	800	A total of 56 species of midwater fishes were collected including several new species not previously sampled. 4 species contributed 88% of all specimens collected, disagreeing with net-based sampling in the same area. Abundance estimates of dominant species were very similar with previous studies regardless differences in gear. Sub-Polar Front was found to be a taxon-specific biogeographic boundary: 1) being semi-permeable in one direction, 2) a strong boundary for species in another direction, or 3) having no effect on deeper-living fish.	(Sutton <i>et al.</i> , 2013)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Charlie-Gibbs Fracture Zone (North Atlantic)	Observations of plankton from the manned submersible "Mir-1"	Jun 2003	4361	Direct observations of the plankton vertical distribution were performed. Pelagic shrimps, chaetognaths, and gelatinous animals were the most abundant taxa with a peak of densities corresponding to the main pycnocline. The near-bottom peak of appendicularian abundance is a common feature in the entire Central Atlantic, and it is not related to the different bottom biotopes.	(Vinogradov, 2005)
The Atlantis Seamount (34°09' N; 30°15' W) and Great Meteor Seamount	cruise with RV "Meteor" and a cruise with RV "Heincke"	Sep 1998, Oct 2000	800	The effects of decreasing water depth on vertically migrating mesopelagic fish were investigated. There were lower mesopelagic fish densities, species numbers and diversity Above the seamounts than in the surrounding oceanic deep water. no evidence of unique seamount-associated mesopelagic community was observed. Mesopelagic fish assemblages were described as a "thinned-out oceanic community": with lack of mesopelagic species on the plateaus of seamounts. Such gap can be explained by shoaling of the vertical migration range and enhanced predation by benthopelagic species.	(Pusch <i>et al.</i> , 2004)
Gulf of Mexico	Natural Resource Damage Assessment conducted by the National Oceanic and Atmospheric Administration (NOAA) aboard the NOAA ship FSV <i>Pisces</i> and M/V <i>Meg Skansi</i>	Jun 2011	1000	Variations in abundance and taxonomic composition of mesopelagic organisms were investigated. Based on acoustics, 4 taxonomic categories were assigned to mesopelagic sound scattering layers (SSLs): 1) crustacean and small non-swimbladdered fish, 2) large non-swimbladdered fish, 3) swimbladdered fish, and 4) unclassified. At dusk, part of the acoustic energy moved from the mesopelagic into the upper epipelagic (but still below the thermocline depth). Different taxonomic categories performed migration to different depth while other taxa seem to stay in mesopelagic habitats, resulting in certain habitat separation.	(D'Elia <i>et al.</i> , 2016)
The Gulf of Mexico.	R.V. "Columbus Iselin" and the R.V. "Suncoaster"	Summers 1975-1977, 1990	1000	Euphausiids of the mesopelagic zone were feeding on marine snow generated in the epipelagic zone.	(Kinsey & Hopkins, 1994)
The eastern Gulf of Mexico	RV <i>Columbus Iselin</i> and RV <i>Suncoaste</i>	summers of 1975-1977 and 1984-1985	1000	Species composition, vertical distribution, and food habits of the 10 species of sergestid shrimp was investigated. Species composition was very similar to the ones in subtropical western North Atlantic. Most of the groups performed DVM. During the daytime, the peak of the distribution of the mature stages below 200m. resource partitioning was observed among the sergestids.	(Flock & Hopkins, 1992)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Eastern Gulf of Mexico	University of South Florida	1996-1999	2500	the bathypelagic zone contains a distinct pelagic community, biologically and ecologically different from that of the mesopelagic zone.	(Burghart <i>et al.</i> , 2007)
Bahamas	R/V 'Sharp'	Jul 2015.		The spatial dynamics of DSLs were explored in 2 locations that are important foraging habitats for a deep-diving predator, Blainville's beaked whales. Diffuse layers were made up of larger animals with greater migration distance and biomass, indicating that more energy is being transferred between energy-rich surface waters and the deep sea. Presence of fine-scale features (intense layers dominated by thinner layers) may increase foraging efficiency by layer predators.	(Sato & Benoit-Bird, 2017)
Benguela upwelling system	RS <i>Africana</i> , RRS <i>Discovery</i> and RV <i>Maria S. Merian</i>	Dec 2009, Sep/Oct 2010, Feb 2011	1000	ecophysiology of pelagic decapods was investigated. Species-specific regional distribution limits were observed for various species. Decapods play a more prominent role in the study area than previously thought, exerting considerable predation impact on calanoid copepods (up to 13% of standing stock).	(Schukat <i>et al.</i> , 2013)
<b>Mid-Atlantic Ridge</b>					
Mid-Atlantic Ridge (MAR) between Iceland and the Azores.	research vessels <i>Arni Fridriksson</i> , <i>G.O. Sars</i> , <i>Henry Bigelow</i> and RSS <i>James Cook</i>	2003, 2004, 2007 and 2009	3000	New baseline information was collected on biogeography, abundance, and distribution of euphausiids along a section of MAR. overall, MAR weekly affected the distribution and abundance of the euphausiids. 18 species were recorded in the study area with 4 major species assemblages. The abundance of euphausiids was highly variable but decreased logarithmically with depth.	(Letessier <i>et al.</i> , 2011)
Mid-Atlantic Ridge (MAR; Iceland to the Azores)	2004 R/V <i>G.O. Sars</i> , MAR-ECO expedition	2004	4300	The assemblage structure and vertical distribution of deep-pelagic fishes relative to a mid-ocean ridge system are described up to 4300 m. unknown topographic association of a bathypelagic fish assemblage with a mid-ocean ridge system was discovered, opposing the general "open-ocean" paradigm that biomass decreases exponentially with depth. bathypelagic assemblage over the northern MAR that consistent along the ridge.	(Sutton <i>et al.</i> , 2008)
The Mid-Atlantic Ridge, along with a transect extending from Iceland to the Azores	Leg 1 of the MAR-ECO cruise	Jun 2004	2500	Copepod composition, abundance, and distribution were investigated. In total, 68 genera and 117 species of copepods were identified. The number of genera increased southwards. The total copepod number was variable with no clear north-south trend. 3 distinct assemblages were identified primarily based on copepod community structure rather than water masses. Depth distribution varied between the species and was affected by hydrographic conditions.	(Gaard <i>et al.</i> , 2008)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Mid-Atlantic Ridge (MAR), from Iceland to the Azores	MAR-ECO expeditions. Leg 1 of the R/V <i>G.O. Sars</i> , NOAA FSV <i>Henry B. Bigelow</i>	Jun–Jul 2004, Jun-Jul 2009	2300	Distribution of <i>B. euryops</i> (the biomass-dominant pelagic species) in relation to physical and biological variables was investigated. Several important factors such as ridge section, depth zone, and prey abundance were structuring the distribution along the MAR. Larger fish were found deeper in the water column, due to an ontogenetic migration to depth. Mean fish size increased with depth.	(Sweetman <i>et al.</i> , 2013)
Hydrothermal vents over Mid-Atlantic Ocean ridge	<i>Mir</i> submersibles, Cruises 34 and 39 of R/V <i>Akademik Mstislav Keldysh</i>	Sep 1996, Jun-Jul 2002,	3100	The different productivity of the upper layers resulted in significant difference in plankton distribution above southern abyssal and northern abyssal fields. Majority of the taxonomic groups (except gelatinous organisms) showed no increase in concentration neither near-bottom nor near the plumes. This finding supports the hypothesis of plankton enrichment due to the production of hydrothermal fields. In close proximity to the “black smokers,” common planktonic organisms were replaced by the community of hydrothermal fauna. Gelatinous animals were highly abundant at the plume boundaries. Overall, the hydrothermal community is quasi-closed 3D structure.	(Vinogradov <i>et al.</i> , 2003) (Vinogradov <i>et al.</i> , 1997)
Broken Spur vent field, Mid-Atlantic Ridge	39th cruise of R/V <i>Akademik Mstislav Keldysh</i> with 2 deep-sea manned submersibles <i>Mirai</i> aboard	Sep 1997	1000	Biomass of various mesopelagic taxa is investigated in relation to depth. Plankton distribution shows 2 aggregations: one within the main pycnocline and the other near the plume. Both aggregations are dominated by gelatinous animals and radiolarians. Large aggregations of organisms occurred at the upper and lower borders of the plume, while the plume core has low biomass.	(Vereshchaka & Vinogradov, 1999)
Lost City (Atlantis underwater massif) and the Broken Spur hydrothermal fields	cruise 50 of R/V <i>Akademik Mstislav Keldysh</i>	Aug 2005	1000	there was no significant increase in the plankton concentration above the seamount. the horizontal heterogeneity of the plankton distribution was observed above the Lost City. The near-bottom aggregations of euphausiids and amphipods are the artifact due to the attraction of the animals by the submersible’s headlights rather than a natural phenomenon.	(Vinogradov & Vereshchaka, 2006)
North Atlantic (TAG and Broken Spur hydrothermal fields)	DSRV <i>Mir</i>	n/a	1000	vertical distribution of zooplankton over the hydrothermal fields was investigated. No local enrichment in the pelagic zooplankton was observed. the bottom community was dominated by benthopelagic shrimps. Overall, net pelagic plankton was almost absent, and bathometer plankton was scarce, with a high proportion of dead organisms. the hydrothermal ecosystems are bioenergetically closed.	(Vinogradov <i>et al.</i> , 1996)
Mid-Atlantic Ridge	ROVs <i>Aglantha</i> and <i>Bathysaurus</i>	Jun 2004	2335	The major part of the fauna was comprised of medusae, ctenophores, appendicularians, siphonophores, and tunicates. All of these animals, except the tunicates, occurred throughout the water column with the	(Youngbluth

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Charlie-Gibbs Fracture Zone				highest abundance of gelatinous taxa were observed from 350 to 750m. The majority of these organisms were between 1 and 10 cm. Detritivores species were a prominent feature of the epibenthic macrozooplankton community.	<i>et al.</i> , 2008)
the northern Mid-Atlantic Ridge Charlie-Gibbs Fracture Zone	R/V <i>Henry B. Bigelow</i> MAR-ECO Expedition	Jun–Jul 2009	3000	Structure and distribution of Pelagic faunal assemblage were investigated. first study to characterize a DVM of pelagic nekton in the study area. Peak in diversity and biomass was observed between depths of 700–1900 m. The latitudinal gradient in sea-surface temperature and water masses was reflected in a similar gradient in ichthyofaunal diversity. Eight deep-pelagic fish assemblages were identified with depth being a primary factor. Species composition and abundance estimate did not differ from previous expeditions.	(Cook <i>et al.</i> , 2013)
<b>Japan</b>					
Sagami Bay, Japan,	remotely operated vehicle (ROV) <i>Dolphin 3K</i> , crewed submersible <i>Shinkai 2000</i>	May-June 1997.	1450	Method to systematically analyze the videotapes from submersible dive to ensure the more efficient data collection	(Hunt & Lindsay, 1999)
Toyama Bay, Japan Sea  Sagami Bay, north-western Pacific	crewed submersible ' <i>Shinkai 2000</i> '	Jul 1999 and 2000		Species composition differed remarkably between the two bays. Biodiversity of cnidarians and ctenophores was lower in Toyama Bay than in Sagami Bay. Horizontal patchiness of the gelatinous forms should be considered for accurate calculations of biodiversity indices. Authors also point to the necessity of multiple dives in a single area and survey period. Highest richness was observed in 400–600 m depth layer in Sagami Bay due to vertical migration and predation.	(Lindsay & Hunt, 2005)
cold-current Oyashio and the warm-current Kuroshio in the NW Pacific	research vessels <i>Wakataka-maru</i> and <i>Hokko-maru</i> , research project DEEP (Deep-sea Ecosystem and Exploitation Program)	May 2004 to Mar 2006	1500	seasonal and regional change in vertical distribution and DVM patterns of four euphausiid species were investigated. There was a regional change in vertical distribution between studied species. Species were found in wide temperature ranges, indicating that they were able to adapt to different temperatures in different regions. the dominant species in the area has a trade-off of long migrations and a warmer environment that accelerates metabolism, in return for obtaining a food-rich environment.	(Soga <i>et al.</i> , 2016)
Oyashio current	FRV <i>Tankai Maru</i> and <i>Hokko-Maru</i>	Mar-Oct 2000	2000	Studied species of copepods habitats were due to the season, vertical distribution, and food resource partitioning.	(Tsuda <i>et al.</i> , 2014)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Oyashio region, western North Pacific Ocean	T/S 'Oshoro-Marui' and T/S 'Hokusei-Marui'	Sep 1996 - Oct 1997	2000	Diel and seasonal vertical distribution, life cycle and body allometry of 2 copepod species ( <i>P. scutullata</i> and <i>H. tanneri</i> ) were investigated. The observed differences are discussed in relation to the differences in food habits and the production cycle of their prey.	(Yamaguchi, 2000)
<b>Arctic Ocean</b>					
Arctic Ocean	numerous Russian expeditions	1929 – 1993	3000	Fauna of pelagic ostracods of Arctic Ocean is very poor and is entirely isolated from the fauna of the Pacific Ocean. Ostracods are most abundant in Atlantic layer from 250–300 to 750–900 m. The average size of the ostracods decreases with an increase in depth due to food deficiency.	(Bashmanov & Chavtur, 2008), (Bashmanov & Chavtur, 2009)
The Arctic Ocean and adjacent Barents, Kara, Laptev, and East-Siberian seas	four RV <i>Polarstern</i> during ARK cruises	1993-1998	3000	Baseline assessments of the zooplankton biomass, its regional variability, and factors affecting the biomass distribution were investigated. Regional variability was correlated with the circulation pattern. Biomass was the highest in the core of the Atlantic inflow and steadily decreased towards the shelves and basins. The vertical distribution did not differ much between the stations was similar at all stations in the upper 200m.	(Kosobokova & Hirche, 2009)
Arctic	Remotely Operated Vehicle (ROV) <i>Global Explorer</i> (Deep Sea Systems)	Jun - Jul 2005	~3900	a baseline study to understand biodiversity and distribution of gelatinous zooplankton in the Arctic Ocean. >50 different gelatinous taxa were observed, several new species were discovered. Major range extensions for known species were observed. Vertical and horizontal distribution of species were found to be linked to water masses, bottom topography, and geographic location.	(Raskoff <i>et al.</i> , 2010)
Arctic Ocean Canada basin, Arctic's central basins	US Coastguard Cutter <i>Healy</i> . two expeditions of the Russian drift stations North Pole (NP-22 and NP-23), six expeditions of the research icebreaker <i>Polarstern</i> , and one	Jun– Jul 2005, 1975 - 2007	3000	The composition and biomass of the zooplankton community within the water column was investigated. Most of the species were common to the Arctic waters, while several Pacific copepod species were encountered, suggesting no zoogeographical barrier between the Canadian and Eurasian basins throughout the entire depth range. Overall species composition did not change over the past 50 years and was similar to the Eurasian Basins. Although the 50 % of the biomass was concentrated in the top 100m, the majority of species diversity occurred below 100m.	(Kosobokova & Hopcroft, 2010) (Kosobokov

Region	Source of data	Time period	Depth samples	Important findings	Ref.
	expedition of the U.S. Coastguard <i>Cutter Healy</i>			comparison with the historical data revealed good documentation of the zooplankton in epi- and mesopelagic layers, while bathypelagic layers were poorly described. The difference in species composition between the upper and midwater layers are related to the occurrence of Atlantic and Pacific expatriates	a <i>et al.</i> , 2011)
Beaufort Sea slope	US–Canada Transboundary Fish and Lower Trophic Levels project	Aug and Sep 2012–2014	1000	the vertical structure of zooplankton communities was investigated. 95 holoplanktonic categories (88 species, 5 genera, 1 order and 1 phylum) were documented. Distinct communities were associated with the main water masses (community structure was correlated with salinity and depth). Species composition was similar to the interior basins but had higher biomass due to elevated coastal production.	(Smoot & Hopcroft, 2017)
Arctic Ocean	Pan-Archive ADCP (Acoustic Doppler current profiler) Moorings	>50 years	n/a	DVM of zooplankton occurs in a high-latitude marine environment when light levels are extremely low. Migration of zooplankton over the entire Arctic ocean environments are driven by lunar illumination. Lunar vertical migration (LVM) produces monthly pulses of carbon remineralization every ~30 days coinciding with the moon cycle. Presence of LVM suggests that this behavior of zooplankton should be considered as a “baseline” in the Arctic Ocean.	(Last <i>et al.</i> , 2016)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Arctic Baffin Bay	n/a	Aug 1983	1043	The vertical distributions of chaetognaths are described. Younger chaetognaths prefer a shallower depth compare to the older more mature stages. No DVM was observed in any of the described species. Main food source for chaetognaths were copepods. The maximum concentration of chaetognaths was found at a depth below the maximum concentration of copepods.	(Same moto, 1987)
<i>en route</i> to the Arctic Ocean through the Fram Strait.	IO PAN annual Arctic summer cruise	2012	1000	Depth along with large horizontal spatial gradients (latitudinal and longitudinal) was the most important factors structuring zooplankton community.	(Gluchowska <i>et al.</i> , 2017)
Subarctic Pacific	Collections as part of the joint U.S.-Canada High Seas survey.	1956-1962 and 1980-1989	150	The intensity of the winter wind is positively correlated with the subsequent summer zooplankton biomass in the subarctic gyre	(Brodeur & Ware, 1992)
<b>Pacific Ocean</b>					
North Pacific	PICES Working Group based on multiple micronekton sampling programs from Japan (SNFRI(JFA), HU, ORI, TNFRI, HNFRI), Russia (TINRO), China (YSFRI), and USA (NMFSC/ SWFSC/ AFSC, Oregon State University )	Various, 1987-2005		Synthesis of all the information on taxonomy, distribution, and trophic relationships of micronekton in the North Pacific Ocean. The difference in DVM patterns and resource partitioning commonly results in segregation of micronektonic fish community.	(Brodeur & Yamaura, 2005)
Pacific trenches (Central Pacific) Hadal trenches	free-vehicle baited trap	Apr-May 2014, Nov – Dec 2014	6945	First study where the life history of the deepest-living fishes, hadal liparids, were investigated. The maximum age of the liparid family was expanding up to 2 years. Thermal history analysis revealed a change of >5 °C in habitat temperature through ontogeny, suggesting that pelagic larval stage for the hadal liparids are found 1000 m and descend to a	(Gerringer <i>et al.</i> , 2018)



Region	Source of data	Time period	Depth samples	Important findings	Ref.
				greater depth when they mature. the potential role of trophic ecology in structuring fish communities at the abyssal-hadal boundary was investigated. Amphipods was a primary prey source with a minor contribution from decapod crustaceans, polychaetes, and remains of carrion (squid and fish). Suction-feeding hadal liparids are descending into the trench to get access to a higher number of amphipods. More generalist feeders and scavengers rely on carrion, do not benefit at hadal depths.	(Gerri nger <i>et al.</i> , 2017)
Northeast Pacific	CFAV <i>Endeavour</i>	Jul 1991-1994	3000	medusae make up a larger proportion of total zooplankton abundance and biomass from mesopelagic to bathypelagic layers in the immediate vicinity of the hydrothermal vents.	(Burd & Thoms on, 2000)
PAP site NE Atlantic	RRS <i>Discovery</i>	Jul - Aug 2009	1000	Reconciliation of the carbon budget in the mesopelagic realm was presented. prokaryotes are responsible for 70-92 % of the estimated remineralization, although much of the organic carbon is exported in the form of large, fast-sinking particles. zooplankton fragment and ingest 50% of these particles and >30% of which is suspended and slowly sinking further, feeding the deep-ocean microbial loop.	(Giering <i>et al.</i> , 2014)
North Pacific (Kurile-Kamchatka region)	submersibles <i>Mir-1</i> and <i>Mir-2</i> diving from the support ship R/V <i>Akademik Mstislav Keldysh</i>	Jul-Oct 2010	6000	The vertical distribution of several medusa species was investigated. A faunistic border was present at a depth of 3000 m. gelatinous organisms are very important participants in transportation of organic matter and energy from the productive surface zone into the oceanic depths. gelatinous zooplankton compromise a significant part of the predator in the epipelagial, meso- and bathypelagial zones.	(Vinogradov & Shushkina, 2002)
Subtropical western North Pacific	three cruises of the R/V <i>Mirai</i> (MR10-01, R10-06, and MR11-03)	Jan 2010, Oct 2010, Nov 2010, Apr 2011	200	temperature elevation and the addition of dissolved organic carbon can enhance heterotrophic prokaryotic production and prokaryotic respiration in mesopelagic zone.	(Uchi miya <i>et al.</i> , 2016)
Western North Pacific Ocean	R/V <i>Mirai</i>	January - February 2010	1000	Mesozooplankton community facilitates the downward carbon flux. The magnitude of the transported carbon flux might be larger during winter than the flux of sinking fecal pellets.	(Kobari <i>et al.</i> , 2013)
Western North Pacific Ocean	As part of the research program “WEST-COSMIC” (Western Pacific Environment Study on CO <sub>2</sub> Ocean Sequestration for	Aug 1998	5000	vertical distribution patterns of community structure and size spectra of plankton organisms divided into four major groups (bacteria, phytoplankton, protozooplankton, and mesozooplankton) were studied at three sites. For all the groups, biomass decreased exponentially with depth. A close link was determined between bacteria and protozooplankton groups. This bacteria–protozooplankton link is	(Yamaguchi <i>et al.</i> , 2002a)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
	Mitigation of Climate Change)			discussed in relation to the 'biological pump' in the western North Pacific Ocean.	
Western North Pacific	As part of the research program WEST-COSMIC Phase I	1997–2001	5800	vertical profiles of the 4 plankton groups (heterotrophic bacteria, phytoplankton, protozooplankton and metazooplankton) were investigated in subarctic, transitional, and subtropical regions. The biomass of all four groups decreased with depths with the various magnitude among the groups. As plankton community structure, metazooplankton biomass or/and bacterial biomass occupied >50% of the total biomass. Integrated metazooplankton biomass exhibited a clear latitudinal pattern (high north and low south. A new link ("sinking POC-DOC-heterotrophic bacteria- heterotrophic microflagellates - heterotrophic dinoflagellates" ) was proposed as a microbial food chain operative in the deep layer of the western North Pacific. implications to the findings to the 'biological pump' are discussed.	(Yamaguchi <i>et al.</i> , 2004b)
Western North Pacific Ocean	As part of the research program Western Pacific Environment Study on CO <sub>2</sub> Ocean Sequestration for Mitigation of Climate Change (WEST-COSMIC)	Aug 1998	5800	plankton distribution down to 5800 m was investigated in four stations in the western North Pacific Ocean. Mass of the plankton was higher at high latitudes. At all locations, mass decreased exponentially with depth. The dominance of detritus in samples below 3000 m was found, and the overall effect of detritus in measuring the chemical composition of plankton samples was discussed.	(Yamaguchi <i>et al.</i> , 2005)
Northwest Pacific	R/V <i>Wakataka-Maru</i> of the Tohoku National Fisheries Research Institute, Fisheries Research Agency (TNFRI, FRA) and the R/V <i>Hokkoumaru</i> of the Hokkaido National Fisheries Research Institute (HNFRI, FRA)	May 2005 and Mar 2006	1500	jellyfish composition and abundance in the North Pacific Intermediate Water differed remarkably between the shallower layer and the deeper layer (300–500 m boundary). Although species abundance was low in deeper layers, the diversity was the highest and consisted of 27 taxa common in all the regions throughout the year. In the Oyashio waters, jellyfish abundance was higher than in the Transition waters but with low diversity. High biomass of jellyfish in the midwater zone was due to the occurrence of larger species. Carbon-based jellyfish biomass exceeded that of other organism groups.	(Morita <i>et al.</i> , 2017)
Western subarctic Pacific Ocean	VERTical Transport In the Global Ocean (VERTIGO) program.	Jul–16 Aug 2005	1000	the impacts of ontogenetically (seasonally) migrating copepods on carbon transport to the mesopelagic zone was investigated. Copepod ingestion rate was ~26–37% of the primary production. But phytoplankton was only 37–59% of the ingested carbon. Thus copepod relies on detritus and microzooplankton for their nutrition. Transport of fecal pellets by copepods and active transport by diel and ontogenetic migration are very important during the summer season dominated by	(Kobari <i>et al.</i> , 2008)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				small phytoplankton. ontogenetic migrants in the subarctic Pacific Ocean are more important participants in active carbon flux than in other open-ocean regions.	
Subarctic Pacific	T/S <i>Oshoro-Mar</i> and T/S <i>Hokusei-Mar</i>	Jun 2009, Sep 1996 - Oct 1997	3000 2000	longitudinal changes in population structure and vertical distribution of the dominant copepod species were investigated. Within the studied species, stage-specific distribution depths remained constant with the season. Ontogenetic vertical migration was present in studied with early and late copepodite stages occurring in deeper and middle copepodite stages occurring at shallower depths. This ontogenetic vertical migration pattern along with vertical separation within the same genus is hypothesized to be related to the body density and reduced predatory mortality in the deep layer.	(Yamaguchi <i>et al.</i> , 2013)  (Yamaguchi & Ikeda, 2002)
western subarctic Pacific	As part of the research program “WEST-COSMIC” (Western Pacific Environment Assessment Study on CO <sub>2</sub> Ocean Sequestration for Mitigation of Climate Change)	Aug 1998	4000	day/night vertical distribution patterns of copepodite stages of 6 epipelagic, 2 mesopelagic, and 6 bathypelagic copepods were investigated. All epipelagic species performed ontogenetic vertical migration (OVM; descent with the progression of development) with only one species exhibiting DVM at later stages. No OVM was observed in mesopelagic species, but DVM was a common feature at later stages. Bathypelagic species underwent reverse OVM (ascent with the progression of development) with no DVM. According to predation pressure hypothesis, these differences can be interpreted as results of life history traits toward reducing predation mortality. The absence of OVM in mesopelagic copepods is a life history trait that falls somewhere between these 2 extremes. This is supported by the observation that the fecundity of these copepods decreases with increasing depth.	(Yamaguchi <i>et al.</i> , 2004a)
western subarctic Pacific	As part of the research program Western Pacific Environment Study on CO <sub>2</sub> Ocean Sequestration for Mitigation of Climate Change (WEST-COSMIC)	Aug 1998	4000	vertical distribution and community structure of 98 species of copepods were investigated. Both abundance and biomass of copepods were greatest in the near-surface layer and decreased exponentially with depth. Copepod carcasses occurred throughout the water column with the highest carcasses to living specimens ratio in the deepest layer. Based on the species similarity indices, copepod community could be classified into epipelagic, mesopelagic, and bathypelagic communities, while based on the feeding strategies, copepods could be classified as suspension feeders, suspension feeders in diapause, detritivores, and	(Yamaguchi <i>et al.</i> , 2002b)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				carnivores. on average, 32% of the particulate carbon flux is consumed by copepods in the water column.	
western tropical North Pacific	n/a	Aug to Oct 2014	3000	the vertical distribution of the zooplankton abundance, biovolume, and size spectra down to 3000 m were investigated. zooplankton abundance and biovolume decreased with increasing depth but described by different models. the zooplankton communities were characterized by low productivity and high energy transfer efficiency. 3 distinct zooplankton communities were identified, which was consistent with the vertical zonation of the water column.	(Dai <i>et al.</i> , 2017)
eastern tropical North Pacific	Eastern Tropical Pacific Project cruises on the R/V <i>Seward Johnson</i> and the R/V <i>Knorr</i>	Oct– Nov 2007, Dec 2008–Jan 2009	1000	Structural and functional features of pelagic OMZ zooplankton communities were investigated. Peak in zooplankton biomass occurred at the thermocline regardless of OMZ extent. The secondary peak in zooplankton biomass along with unique community was associated with the lower oxycline at a specific level of oxygen concentration. another secondary peak due to DVM was observed in upper oxycline or OMZ core at the same depth and temperature regardless of different oxygen concentrations.	(Wishner <i>et al.</i> , 2013)
Eastern Tropical North Pacific (OMZ)	RRS <i>James Cook</i>	Dec 2013 -Feb 2014	500	The remineralization length scale is very high in OMZ. 70% of POC remineralization is due to microbial respiration in the study area as a result of lower particle fragmentation by zooplankton. Zooplankton's sensitivity to the low oxygen concentration in OMZ and further expansion of this zone can affect atmospheric carbon sequestration and have negative feedback on climate change.	(Cavan <i>et al.</i> , 2017)
off the coast of Valparaiso	n/a	Jul 1994 - Sep 1995	900	The bathymetric distribution of chaetognaths was investigated. The greatest density of chaetognaths was found in the epipelagic layer with the species diversity decreased gradually throughout the mesopelagic zone. The vertical distribution of the chaetognaths showed a strong association with the water masses present.	(Ulloa <i>et al.</i> , 2000)
<b>California Current System</b>					
central California within the California Current	Monterey Bay Aquarium Research Institute	1991–2016.	4000	Overview of interactions within the 'jelly web,' highlighting the importance of medusae, ctenophores, and siphonophores as key predators. Importance of these organisms is comparable to large fish and squid species in deep pelagic food webs.	(Choy <i>et al.</i> , 2017)
California Current off southern California and	RV " <i>New Horizon</i> " and one cruise of the National Oceanic and Atmospheric Administration (NOAA)	2009 and 2010	n/a	ontogenetic changes in swimbladder inflation and body density in 71 mesopelagic fish were investigated.	(Davis <i>et al.</i> , 2011)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
North Pacific Subtropical Gyre	FSV “ <i>Bell Shimada</i> ” in the North Pacific				
Southern California Current Ecosystem	CalCOFI cruises	2010-2012	1500	Estimated biomass of mesopelagic fishes of southern California (25-37 g·m <sup>2</sup> ) is comparable to the density of inshore epipelagic zooplanktivorous fishes. Thus, mesopelagic fishes are likely to play a major role in food webs of this region.	(Davis <i>et al.</i> , 2015)
California current system	California Cooperative Oceanic Fisheries Investigations (CalCOFI) program,	1951-2008	525	the impact of declining oxygen on midwater fishes within the OMZ was investigated. The abundance of mesopelagic fishes was 63% lower in low oxygen concentrations compared to periods of high oxygen concentrations. the shoaling of the hypoxic boundary layer during periods of reduced oxygen result in higher vulnerability of the mesopelagic organisms to visually orienting predators.	(Koslow <i>et al.</i> , 2011)
Humboldt current system	RV METEOR	Feb 2009	500	Some of the mesopelagic zooplankton (e.g., <i>C. chilensis</i> ) are found within oxygen minimum zone. Although being highly abundant, only part of the species biomass is available to the predators (anchovy and other important fish species) which can only tolerate waters with higher oxygen concentrations.	(Hirche <i>et al.</i> , 2014)
Monterey Bay, California	Monterey Accelerated Research System (MARS)	Feb 2009 - Aug 2010	875	the responses of animals in DSL to oceanic variability at seasonal and sub-seasonal timescales were investigated. Pelagic animals were found higher in the water column and were lower in abundance during spring upwelling as a result of a shoaling of oxycline and advection offshore. Seasonal changes were most apparent in the non-migrating DSL. Correlations of acoustic backscatter with oceanographic variability also differed with depth. Deeper water had similar correlations as shallower layers but with the increased lag that was consistent with sinking speeds of marine snow. Variability in backscatter was correlated with sea-surface height at the sub-seasonal level, indicating the importance of passive physical transport.	(Urmy & Horne, 2016)
Monterey Bay, California	the Monterey Bay Aquarium Research Institute (MBARI), ROV <i>Ventana</i>	1993–1996	1000	The vertical distribution and seasonal abundance of several small calycophoran siphonophores were investigated. The vertical distribution varied seasonally, coupled to the upwelling in the bay with the abundance peaking after upwelling. Species perform DVM, possibly with two separately migrating groups. No difference in abundance or distribution was observed between years except in 1993 due to El Niño event in the previous year. The abundance of these siphonophores was negatively correlated with competing species of a physonect siphonophore.	(Silguro, 2000)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
In the Catalina Basin, off the eastern coast of Catalina Island, California	autonomous echosounder system (split-beam Simrad EK60s)	Sep 2013	600	The first study to show that the formation of layers of life in the sea is a highly organized and the internal layer structure of DSL is made up of many topologically scaled, mono-specific aggregations, or “schools” rather than an unordered mix of sizes and species. the formation of layers is partially driven by biotic pressures for cohesion.	(Benoit-Bird <i>et al.</i> , 2017)
Pacific California	jaw motion recorders mounted on 14 seals	2011, 2012, 2013	1200	The study assessed the feeding rate and dive cycle time of 14 female northern elephant seals in the OMZ. OMZ plays an important role in the survival and evolution of elephant seals. Mid-mesopelagic was a common hunting grounds for all seals. Mesopelagic fishes in OMZ had reduced mobility and seemed to have a reduced escape response from the seals, enhancing the efficiency of foraging of the seals.	(Naito <i>et al.</i> , 2017)
<b>Hawaii</b>					
ALOHA	the Hawaii Ocean Time-series station	2011	1000	A major fraction of the food source for deepwater zooplankton is surface derived. Carbon demand of zooplankton is way larger than the flux of particle remineralization in the mesopelagic zone. So, the additional source of carbon can be food consumed at the surface and the consequent carnivorous consumption of migrant zooplankton at depth.	(Hannides <i>et al.</i> , 2013), (Koppelman, 2003), (Steinberg <i>et al.</i> , 2008)
Station ALOHA (2245’N 15800’W)	Hawaii Ocean Time-series (HOT) cruises 236, 238–240, 242–243, and 247, Research and Education (C-MORE) cruise Hawaii Ocean Experiment	Nov 2011 and Oct 2012	1000	Measurements of the temporal and depth variability of microbial community respiration (MCR) . upper mesopelagic had a higher MCR than in lower mesopelagic with the peak occurring at 600-650 m. Seasonal variability in respiration rates was found in the mesopelagic zone at Station ALOHA where export flux controls respiration in the mesopelagic zone.	(Martínez-García, 2017)
Hawaii Ocean Time-series station ALOHA in the subtropical Pacific and the Japanese times-series site K2	RV Kilo Moana and the RV Roger Revelle	Jun- Jul 2004, Jul - Aug 2005	1000	This study compares losses of sinking POC with bacteria and zooplankton metabolic requirements. mesopelagic bacterial C demand was larger than that of zooplankton at ALOHA but was similar at K2. At both locations, sinking POC flux was insufficient to meet metabolic demands. this additional C demand must be met through two processes ultimately supplying organic C to mesopelagic bacteria: 1) by DVM of zooplankton feeding at the surface and 2) by carnivory at depth.	(Steinberg <i>et al.</i> , 2000)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
along the Kona coast of the island of Hawaii	NOAA IEA 2013 cruise, NOAA IEA 2014 cruise and UH/SOI KONA14 cruise	Jun 2013, Mar 2014, Feb 2014	800	numerical density and size of marine animals of DSLs were estimated from active acoustic technique using a DIDSON sonar. numerical densities estimated in this work were orders of magnitude higher than those estimated from trawls, and average sizes of animals were also larger. Numerical density and length of animals varied by month, with numerical density also a function of depth.	(Giorli <i>et al.</i> , 2018)
<b>Indian Ocean</b>					
the eastern coast of Oman the western Indian Ocean	<i>Discovery</i> Cruise 209	August 1994	~3500	The highest biomass of plankton and micronekton were in the top 100m due to very high oxygen levels at these depths. large populations of myctophid , photichthyid fishes, and decapod crustaceans were present below the oxycline by day. Most of these organisms migrated to the surface at night. During the daytime, multiple fine layers were observed that were correlated with salinity differences. At the base of the OMZ, there was a large increase in biomass.	(Herring <i>et al.</i> , 1998)
South-West Indian Ocean	(i) research surveys within the framework of the MyctO-3D-MAP project and (ii) fishing trips from the IMOS-BASOOP	2010-2014	800	Changes of the micronektonic vertical structure were investigated. the surface layer acoustic density and thickness decrease in a southward direction. The intermediate layer was generally absent, except for the area between 30°S and 40°S. The deep layer acoustic density increased in a northward direction, but thickness stayed unchanged. a positive correlation was found between the vertical acoustic organization and oceanographic fronts' position.	(Béhgale <i>et al.</i> , 2015)
the southwest Indian Ocean	two research vessels: RV <i>Dr. Fridtjof Nansen</i> and RSS <i>James Cook</i>	Nov 2009 - Apr 2012	1000	vertical and geographical distribution of Shallow and Deep SLs were investigated. Structurally distinct SL regimes were found across the study area. A close relationship between sea surface temperature and mean volume backscatter, with higher levels of backscatter found in the subtropical convergence zone. Biogeographic zonation of scattering layer biota was observed in the survey area and is likely to have a significant influence on its predators and carbon cycling.	(Boersch-Supan <i>et al.</i> , 2017)
South-Western Indian Ocean	fishing trips carried out within the Australian Integrated Marine Observing System-Bio-Acoustic Ship Of Opportunity Program (IMOS-BASOOP); research surveys carried out within MyctO-3D-MAP programme.	2010 - 2013 2013 - 2014	800	Large-scale distribution of micronekton over a latitudinal gradient (was investigated for two contrasted seasons. DVM and vertical distribution of micronekton in three distinct layers were two consistent features in a study area. the significant correlation observed between abundance and distribution of acoustic backscatter and position relative to front and water masses. No significant seasonal pattern was found. Northern winter shift of the fronts resulted in the northward latitudinal shift of the peak in abundance.	(Béhgale <i>et al.</i> , 2016)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Bay of Bengal, Indian Ocean	Marine Research on Living Resources program	Nov 2005-Jan 2006	1000	Presented a novel method using abundance-weighted tolerance index to predict the species, capable of performing DVM. Only two species of the total twenty-five species exhibited significant DVM which opposed the general concept of DVM in tropical environments. Highly variable physicochemical parameters acted as an invisible barrier for zooplankton to perform DVM that could significantly impact biological pump in the system.	(Tv, 2017)
in the western Bay of Bengal in.	Bay of Bengal Process Studies (BOBPS) on board ORV <i>Sagar Kanya</i> cruise 182	Sep to Oct 2002	1000	Variation in mesozooplankton biomass, abundance, and copepod assemblages between the cold-core eddy and non-eddy regions was investigated. Moderate oligotrophy in study area resulted in the high copepod diversity throughout upper 1,000 m. 93 copepod species were recorded for the first time. Cold-core eddies seem to play a vital role in sustaining zooplankton in nutrient-limiting regions.	(Fernandes & Ramai ah, 2013)
seamounts along the South West Indian Ridge and on a seamount off the Madagascar Ridge	R/V <i>Fridtjof Nansen</i> during Cruise 2009–410	2009	1100	31 micronektonic crustaceans were caught in the study area, but no latitude-related pattern of species richness or abundance was observed. Species richness was the highest in the vicinity (~1 km) of seamounts compared with over the abyssal plains and ridge slopes. Enhanced biomass and species richness are due to “habitat enrichment” that seamounts serve as favorable habitats for crustacean organisms. Thus, ‘oasis effect’ of seamounts conventionally associated with higher trophic levels was also observed for micronektonic crustaceans at lower trophic levels.	(Letessier <i>et al.</i> , 2017)
<b>Arabian and Adaman Seas</b>					
Arabian Sea	USJGOFS cruise	1995	1100	Mesozooplankton and micronekton biomass are drastically reduced in the OMZ and increasing when oxygen start to increase at both boundaries of the zone.	(Gowing <i>et al.</i> , 2003)
Andaman Sea	FORV <i>Sagar Sampada</i> Cruise #220		1000	Majority of myctophids and shrimps occurred in abundance in OMZ	(Karup pasamy <i>et al.</i> , 2011)
<b>Southern Ocean</b>					
Southern Ocean	<i>Discovery Investigations</i>	Dec 1933, Mar, Sep, Oct, and Nov 1934	2500	median zooplankton abundance was ~22 times greater in the epipelagic than at 1000 m. variability in abundance in the epipelagic was 3–4-fold while in mesopelagic variability was <1-fold. Depth was the primary factor separating epipelagic, upper, and lower mesopelagic communities. 3 different depth-integrated communities could be	(Ward <i>et al.</i> , 2014)



Region	Source of data	Time period	Depth samples	Important findings	Ref.
				identified in accordance with different water masses, with such grouping weakening when deeper horizons (500–1000 m) are considered and seasonal migrants are removed. The seasonal difference becomes muted with depth. Depth was a primary factor influencing the sample diversity. However, water mass regime and month also have a noticeable contribution. atmospheric and ocean warming in the last 80 years lead to increase in species richness may increase in the Antarctic water masses with sub-Antarctic species shifting southward.	
Croker Passage, vic. Antarctic Peninsula  southern Scotia and western Weddell Seas	R/V <i>Hero</i>	March and April 1983	1000	A micronektonic and macrozooplanktonic community was investigated. the study area (Croker Passage ) was a faunal transition zone between oceanic and nearshore communities. Krill density was highly variable with time and space while pelagic fishes showed less of the change. The observed finding could be an artifact of the fishing method as the fishes are more vulnerable to sampling gear than krill. Most species had broad vertical ranges with no apparent DVM patterns ( with few exceptions). Biomass was higher than in Pacific subarctic waters. Pelagic fish biomass was greater than the biomass of birds , seals, and whales, suggesting that mesopelagic fish are the major krill predators in the Antarctic oceanic system.	(Lancraft <i>et al.</i> , 2004), (Lancraft <i>et al.</i> , 1989)
Lazarev Sea (Southern Ocean)	German-funded Lazarev Sea KRILL Study (LAKRIS)	2004-2008	3000	The ice-covered surface layer is an important functional node of the pelagic ecosystem.	(Flores <i>et al.</i> , 2014)
Southern Ocean	<i>Discovery Investigations.</i>	Dec 1933, Mar, Sep, Oct, and Nov 1934	1000	Mesozooplankton samples during the productive austral season were investigated. Median zooplankton abundance in the epipelagic zone was 22 greater than at 100m depth. A 3–4-fold variability of abundance was observed within the top 250 m, while at greater depth the variability dropped and was <1-fold. Depth was the primary factor separating distinct communities of epipelagic and upper and lower mesopelagic horizons. Integrated over all depth horizons, resulted in 3 different communities' in accordance with water-mass regimes. This groups were less apparent at greater depths (500–1000 m) and excluding seasonal migrants. Seasonal signals across all data became less distinct with depth. diversity increased with depth. species richness may increase in the Antarctic water masses as sub-Antarctic species move into southward regions.	(Ward <i>et al.</i> , 2014)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Southern Ocean	<i>Discovery Investigations.</i>	Oct 1932, Dec 1933 -Nov 1934, Apr 1938 – Mar 1939, Sep 1936. Jun 1936, 1951	1000	regional comparisons of standing stock in the Southern Ocean was investigated. There was a gradient in zooplankton biomass from low to high latitudes. During summer, the majority of zooplankton is found at the surface during both day and night, while in winter, plankton migrate to the deeper waters. within an entire area, little circumpolar variation is observed in the standing zooplankton crop. Zooplankton standing crop on the Antarctic is 4 times higher of the crop in the tropics. Little seasonal variation in standing crop was observed (over the entire water column)	(Foxton, 1956)
Southern Ocean (island of South Georgia)	n/a	austral summer 1981/1982 and again in winter 1983	250	The mesoscale features of zooplankton distribution were investigated. There was a noticeable difference between the plankton of the shelf and the surrounding ocean. within the oceanic part, warmer waters were inhabited by a fauna characteristic of warmer latitudes to the north. However, not all of the characteristic warm- and cold-water species were correlated with water origin. on a smaller scale, intrusions of warmer water may result in advection of new organisms into the area.	(Atkinson <i>et al.</i> , 1990)
Southern Ocean (Pacific sector of the Antarctic)	cruises of the USNS <i>Eltanin</i>	1963 to 1965	2000	Zooplankton standing crop was estimated. Copepods, chaetognaths, and euphausiids made up most of the catch in the Antarctic, subantarctic, and convergence zone waters. Total biomass in the upper 1000 meters averaged ~2 g dry wt/m <sup>2</sup> . Seasonal vertical migration was present in the spring and summer, the majority of the biomass was located in the top 250m, decreasing through late fall, winter, and early spring months.	(Hopkins, 1971)
Southern Ocean (Crocker Passage, Antarctic Peninsula)	RV <i>Hero</i>	Mar-Apr 1983	1000	Zooplankton species composition, abundance and vertical distribution were investigated. biomass distribution was polymodal with the majority of the zooplankton were less than 1 mm. All the principal species had broad vertical distributions during both day and night and were concentrated below 200 m.	(Hopkins, 1985)
Southern Ocean (west central Weddell Sea)	U.S. National Science Foundation sponsored AMERIEZ program on board USCG <i>Glacier</i> and RV <i>Melville</i>	Mar 1986	1000	The zooplankton community in the vicinity of the ice edge was investigated. Species diversity in the epipelagic zone was moderate with metazoan microzooplankton (< 1 mm) being most abundant. Species composition was similar in open areas and in the vicinity to the ice, but differ in vertical patterns: sparse populations were observed in the upper 50 m under the ice. Diversity increased with depth. Biomass of zooplankton in the top 1000m was low compared to other Southern Ocean areas due to the circulation patterns that resulted in low annual primary production in the central Weddell Sea.	(Hopkins & Torres, 1988)
Southern Ocean (Scotia-	Antarctic Marine Ecosystem Research in the	spring 1983	1000	The structure of the food web in open waters near the marginal ice zone was investigated. Most of the zooplankton were omnivorous with	(Hopkins <i>et</i>

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Weddell confluence region)	Ice Edge Zone (AMERIEZ) program from the RV <i>Melville</i>			phytoplankton, protozoans, and small copepods being dominant prey items. Only two copepod species were found to be exclusively herbivorous. Micronekton were carnivores and primarily feed on copepods. The mesopelagic fish ( <i>E. antarctica</i> ), cephalopods, midwater decapod shrimps and carrion were the dominant prey items in the majority of the examined seabird species. over half the species examined had similar diets in spring, winter, and fall. The significant intraspecific shifts in diet that occurred were primarily due to regional, seasonal, and interannual effects.	<i>al.</i> , 1993)
Southern Ocean, Scotia Sea	<i>Discovery Expeditions</i>	1920s to 1950s	1000	Ecology of small and mesopelagic copepod species were investigated. However, little is known of some basic aspects of their ecology. All mesopelagic taxa except for two warmer-water species <i>Metridia lucens</i> were found throughout the entire study area. here was no evidence that the Polar Front was a major biogeographic boundary to copepod distribution. Total copepod abundance was thus higher in the vicinity of the Polar Front than in any other region.	(Atkinson & Sinclair, 2000)
Southern Ocean	<i>Discovery Investigations</i>	Dec 2009	1000	the catching ability of a Bongo net and a reconstructed version of an N70 net was compared. Intercalibration of the catches from 1926/1927 was performed and the catches were compared with Bongo net hauls made post-1995.	(Ward <i>et al.</i> , 2012b)
Southern Ocean	cruises ANT-28/4 and ANT 28/5	Mar-Apr 2012	2000	a strong influence of isolated water masses such as the basin of the Bransfield Strait on the composition of bacterial communities in the dark ocean.	(Milic <i>et al.</i> , 2017)
South Georgia, South Atlantic	RRS <i>James Clark Ross</i> . <i>Discovery Investigations</i>	Oct/Nov 1997  Dec-Jan 1927	1000	Mesozooplankton were sampled at a shelf and an oceanic station. Onshelf zooplankton biomass was very high exciding those from the oceanic stations. At the oceanic stations, large calanoids composed ~50% of the standing stock while Antarctic krill occurrence was low. Based on the percentage similarity index across different cruises did not reveal any systematic differences in species composition between 1926/27 and the present (2005)	(Ward & Shreeve, 1999) (Ward <i>et al.</i> , 2008)
Southern Ocean (the Marguerite Bay region of the Western Antarctic Peninsula )	onboard the R.V. <i>Lawrence M. Gould</i> , as part of the Southern Ocean Global Ocean Ecosystems Dynamics Program (SO GLOBEC)	Apr-May 2001, 2002	800	the vertical distribution of zooplankton and habitat partitioning in the deeper water column was investigated. The vertical patterns of copepods, euphausiids, amphipods, and mysids were not correlated with either the distributions of pigments, temperature, salinity, nor density. observed vertical distributions were the result of different behavior patterns, including seasonal vertical migration to deeper water for	(Marrari <i>et al.</i> , 2011)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				overwintering or reduce the risk of predation and vertical habitat partitioning to reduce competition.	
Southern Ocean (two deep basins beneath the Croker Passage, west of the Antarctic Peninsula)	RV 'Hero'	Feb 1982 and Mar-Apr 1983	1000	Many Antarctic mesopelagic fish perform DVMs in the absence of two of the three advantages attributed to this behavior: no metabolic bonus as the water is almost isothermal or the prey source does not perform the vertical migration. The only benefit and the principal driver of diel vertical migrations for midwater fish at lower latitudes is avoidance of visual predators in the upper water column.	(Robison, 2003)
Southern Ocean	RRS "Discovery" Cruise 100, RRS 'James Clark Ross' Cruise JR06	Jan-Mar 1979, Feb-Mar 1994	2000	In Weddel gyre, cnidarians contributed >50% of biovolume of the samples (although the range was high). Many species of siphonophores and medusae were present in the mesopelagic zone. Although many species were widespread in that geographical area, some species distribution was controlled by environmental parameters. In the Antarctic Polar Front Zone, gelatinous plankton also compromised >60% of biovolume. Some species had limited vertical distribution due to the trophic interaction between ostracods, amphipods, and cnidarians. Several predators (birds and fish) of medusae and ctenophores were identified.	(Pages <i>et al.</i> , 1994) (Pages <i>et al.</i> , 1996)
The Scotia Sea Southern Ocean	RRS James Clark Ross (cruise JR161)	Oct-Dec 2006	1000	the study investigated spatial and temporal patterns in <i>Electrona antarctica</i> along with abundance, population structure, and diet. Depth distribution varied with season, size-related sexual dimorphism, and size-specific vertical stratification for both species. For <i>E. antarctica</i> , latitudinal trends in sex ratio and female body size were observed. <i>E. carlsbergi</i> didn't recruit in the Scotia Sea.	(Saunders <i>et al.</i> , 2014)
Scotia Sea	The British Antarctic Survey's	Jan 1991	1000	first study to examine/describe the summer nekton community in the Scotia Sea. The most abundant groups were mesopelagic fish, salps, coelenterates, decapod crustaceans, and amphipods. Species diversity increased with depth. The nekton community of the Scotia Sea is an important alternative to the more conventional Antarctic food chain (where krill is consumed by top predators directly) where nekton serves as an intermediate link between the zooplankton and top-predators.	(Piatkowski <i>et al.</i> , 1994)
Scotia Sea	Cruise JR161, Cruise JR177, and Cruise JR200.	Oct-Dec 2006, Jan-Feb 2008, Mar-Apr 2009	400	Mesozooplankton distribution and community structure were investigated. Small copepods dominated numerically across all nets. season; changes in zooplankton biomass were recorded. two main station groups in all 3 seasons were established: 1) stations lying within the seasonal sea-ice zone (with low zooplankton abundance and	(Ward <i>et al.</i> , 2012a)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				biomass), and 2) stations north of the Southern Antarctic Circumpolar Current Front (with higher abundance and biomass). The difference between the groups was most noticeable in the population rather than at the taxonomic level. sea-ice, temperature, and chlorophyll biomass may have influenced the development and seasonal succession of zooplankton populations.	
South Georgia in the northern Scotia Sea	RRS <i>James Clark Ross</i> cruise 100	Mar and Apr 2004	1000	high level of depth stratification was observed within the myctophid community, with only some fishes were performing the DVM.	(Collins <i>et al.</i> , 2008)
northern Scotia Sea NW of South Georgia	RRS ' <i>James Clark Ross</i> ' Cruise 100	Mar and Apr 2004	1000	The diets of 9 species of myctophid fishes were investigated. Resource partitioning due to dietary specialization among study species was observed with 3 distinct feeding groups were identified among myctophid community. As fish size increased, the diet shifted from copepods to euphausiids and amphipods. Only larger myctophids consumed Antarctic krill were only consumed by the larger fish, forming a krill-independent link between secondary production and higher trophic levels. Also, predation of myctophid on copepod production had a very small impact contrary to the predation impact on macrozooplankton.	(Shreeve <i>et al.</i> , 2009)
Southern Ocean	British Antarctic Survey's POETS and SCOOBIES projects within the Ecosystems program, on board of RRS <i>James Clark Ross</i>	Oct-Dec 2006, Jan-Feb 2008, and Mar-Apr 2009	1000	This study investigates the applicability of Bergmann's rule (body size increases with decreasing temperature and increasing latitude) to the mesopelagic fish community in the Southern Ocean. Bergmann's rule was met in 72% of biomass-dominant myctophid species (for both intraspecific and interspecific levels)	(Saunders & Tarling, 2018)
Weddel sea	The "Ice Station POLarstern" (ISPOL) expedition	Dec 2004 and Jan 2005	1000	vertical distribution, species composition, abundance, and developmental structure of the mesozooplankton community was investigated with a focus on calanoid copepods. Copepods were the most abundant taxon (94% of the total mesozooplankton), and 66 calanoid copepod species were identified (but the dominance of only a few species). The community structure was uniform between stations.	(Schnack-Schiel <i>et al.</i> , 2008)
East Antarctica Indian sector of the Southern Ocean	The Collaborative East Antarctic Marine Census (CEAMARC) on board of Training and Research Vessel (TR/V) <i>Umitaka-maru</i>	Jan-Feb 2008, Jan 2005, and Jan 2006	2000	Spatial distribution of pelagic cnidarians was investigated. Three major communities were identified: (1) an epipelagic group in the oceanic zone with low abundance and species diversity; (2) a meso- and bathypelagic group characterized by high abundance and species diversity; and (3) a neritic group. In the epipelagic zone, cinadarian communities were separated in accordance with hydrographic structures,	(Toda <i>et al.</i> , 2014) (Toda <i>et al.</i> , 2010)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
				while in the meso- and bathy-pelagic zones due to relatively uniform deep water, communities were more stable and higher diverse. Reduced in abundance/biomass In epipelagic and lower meso- and bathypelagic communities are likely due to bottom-up control, while the upper mesopelagic community remained stable.	
<b>Tasman Sea</b>					
Tasman Sea	Various fishing vessels	austral winter from 2004 to 2007	1200	Acoustic mapping can be efficiently done on a large scale (basin scales) using commercial fishing vessels. Wet-weight biomass estimates of the micronekton fish in the study area vary considerably by a factor of 5–58 between acoustics, nets, and large spatial-scale, ecological models.	(Kloser <i>et al.</i> , 2009)
The continental slope of south Tasmania	n/a	April 1993	~1000	Comparison of estimates of biomass and community structure between acoustics and trawl-derived estimates. The acoustic biomass estimate of biomass was 7-fold higher than the trawl based that were consistent with regional estimates of primary production and trophodynamic calculations. TS distributions from the 4 pelagic communities indicated an increase in body size with depth.	(Koslow <i>et al.</i> , 1997)
the southern coast of Tasmania	four seasonal cruises by the CSIRO Division of Marine Research vessel, FRV <i>Southern Surveyor</i>	Feb 1992, Nov 1992, Apr 1993, and Jul 1993.	900	Vertical distribution of the zooplankton was investigated. 20 micronekton species made up 80% of the total biomass. Presence of DVM resulted in the noticeable day/night shift in the distribution of biomass; species that do not perform DVM or have limited vertical migration was also found in the region. Three distinct assemblages were identified: epipelagic, lower, and upper mesopelagic. The vertical distribution of these assemblages was correlated with the primary water masses. Advection of mesopelagic prey in Antarctic intermediate water can support the populations of predators of the micronekton in a study area.	(Williams & Koslow, 1997)
the southern coast of Tasmania	four seasonal cruises by the CSIRO Marine Research vessel, F.R.V. <i>Southern Surveyor</i>	Feb 1992, Nov 1992, Apr 1993, and Jul 1993	1400	the feeding ecology of characteristic micronekton species was investigated. Euphausiids and calanoid copepods were the main prey source of myctophids. For larger fish species copepod were less important compared to the euphausiids. When euphausiids biomass peaked in fall, all the fish shifted its diet to the euphausiids. Differences in the timing and duration of feeding were explained by the differences in their spatio-temporal overlap with prey sources. The vertical flux of near-surface plankton production to the mesopelagic zone is based primarily on diel feeding in the top 500 m.	(Williams <i>et al.</i> , 2001)

Region	Source of data	Time period	Depth samples	Important findings	Ref.
Eastern Tasmania	CSIRO FRV <i>Southern Sur-Veyor</i>	May-Jun 1992-1994	350	The link between zooplankton distribution and presence of bluefin tuna in the study area was investigated. the biomass in the convergence zone was not different from the warm E.Australian current or cool subantarctic waters. shelf waters had the greatest biomass of zooplankton and micronekton. nutrient-rich subantarctic and west coast waters along with the upwelling along the shelf break result in higher biomass over the shelf region. In particular, certain species of zooplankton is a figure largely in the diet of jack mackerel, an important prey tuna. Offshore, the process of energy transfer between primary production and top predators is less understood. However, gelatinous zooplankton may be an important player linking primary production and higher trophic level predators.	(Young <i>et al.</i> , 1996)
<b>Vancouver Area/ Canada</b>					
Strait of Georgia and inlets on the west coast of Vancouver Island	submersible <i>PISCES IV</i> .	winter and spring 1980-1983	733	The vertical distribution and abundance of various zooplankton taxa were systematically recorded. Species composition and distribution were stable within the study area. The depth of 175 m was the cut-off point between the migratory and non-migratory shift in behavior and was taken as the demarcation point between the meso- and bathypelagic zones.	(Mackie, 1985)
Atlantic Canadian continental shelves and slopes. Gully submarine canyon (44°N 59°W)	Surveys by Canada's Department of Fisheries and Oceans (DFO)	2006 – 2009	2,300	DNA barcode sequences were obtained from 557 mesopelagic and upper bathypelagic teleost organisms for inclusion in the Barcode of Life Database (BOLD). Out of these, 24 specimens belonged to species that did not have a reference barcode available. The analysis revealed the presence of cryptic speciation, indicating genetic divergence among populations in various ocean basins. Interestingly, despite the genetic differences, the morphology of these organisms remained conserved. This can be attributed to the absence of pronounced horizontal environmental gradients in the deep ocean, which potentially limits the impact of natural selection on morphological variation.	(Kenchington <i>et al.</i> , 2017)
Nova Scotian Slope	n/a	May 1977 and Apr 1977, Jun 1978	~430	The investigation focused on examining the daily consumption of zooplankton by euphausiids and myctophids in two distinct water masses. It was found that both taxa played significant roles as major consumers of zooplankton during the spring season. Importantly, the study revealed that the food chain was not predominantly controlled by tunicates, chaetognaths, or gelatinous organisms, suggesting that other factors and organisms were more influential in shaping the dynamics of the ecosystem.	(Sameoto, 1982)

## References

- Andersen, V. & Sardou, J. (1992) The diel migrations and vertical distributions of zooplankton and micronekton in the Northwestern Mediterranean Sea. 1. Euphausiids, mysids, decapods and fishes. *Journal of Plankton Research*, **14**, 1129–1154.
- Angel, M.V. & Baker, d.C. A. (1982) Vertical Distribution of the Standing Crop of Plankton and Micronekton at Three Stations in the Northeast Atlantic. *Biological Oceanography*, **2**, 1–30.
- Angel, M.V., Blachowiak-Samolyk, K., Drapun, I. & Castillo, R. (2007) Changes in the composition of planktonic ostracod populations across a range of latitudes in the North-east Atlantic. *Progress in Oceanography*, **73**, 60–78.
- Ariza, A., Landeira, J.M., Escánez, A., Wienerroither, R., Aguilar de Soto, N., Røstad, A., Kaartvedt, S. & Hernández-León, S. (2016) Vertical distribution, composition and migratory patterns of acoustic scattering layers in the Canary Islands. *Journal of Marine Systems*, **157**, 82–91.
- Atkinson, A. & Sinclair, J.D. (2000) Zonal distribution and seasonal vertical migration of copepod assemblages in the Scotia Sea. *Polar Biology*, **23**, 46–58.
- Atkinson, A., Ward, P., Peck, J.M. & Murray, A.W. (1990) Mesoscale distribution of zooplankton around South Georgia. *Deep Sea Research Part A. Oceanographic Research Papers*, **37**, 1213–1227.
- Bashmanov, A.G. & Chavtur, V.G. (2008) Distribution of *Boroecia borealis* (Ostracoda: Halocyprididae) in the Arctic Ocean and Adjacent Waters of the Atlantic. *Planktology*, **34**, 341–350.
- Bashmanov, A.G. & Chavtur, V.G. (2009) Structure and distribution of pelagic ostracods (Ostracoda: Myodocopa) in the Arctic Ocean. *Russian Journal of Marine Biology*, **35**, 359–373.
- Béhagle, N., Cotté, C., Ryan, T.E., Gauthier, O., Roudaut, G., Brehmer, P., Josse, E. & Cherel, Y. (2016) Acoustic micronektonic distribution is structured by macroscale oceanographic processes across 20–50°S latitudes in the South-Western Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, **110**, 20–32.
- Béhagle, N., Roudaut, G., Josse, E., Brehmer, P., Cotte, C., Duhamel, G., Gauthier, O., Cherel, Y. & Ryan, T.E., eds. (2015) *Acoustics characterization of micronekton spatial distribution in Indian Ocean using scientific surveys and integrated marine observing system database: Acoustics characterization of micronekton*. IEEE.
- Benoit-Bird, K.J., Moline, M.A. & Southall, B.L. (2017) Prey in oceanic sound scattering layers organize to get a little help from their friends. *Limnology and Oceanography*, **62**, 2788–2798.
- Bergauer, K., Fernandez-Guerra, A., Garcia, J.A. L., Sprenger, R.R., Stepanauskas, R., Pachiadaki, M.G., Jensen, O.N. & Herndl, G.J. (2018) Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E400–E408.
- Bianchi, D., Galbraith, E.D., Carozza, D.A., Mislán, K.A. S. & Stock, C.A. (2013) Intensification of open-ocean oxygen depletion by vertically migrating animals. *Nature Geoscience*, **6**, 545–548.
- Bode, M., Hagen, W., Cornils, A., Kaiser, P. & Auel, H. (2018) Copepod distribution and biodiversity patterns from the surface to the deep sea along a latitudinal transect in the eastern Atlantic Ocean (24°N to 21°S). *Progress in Oceanography*, **161**, 66–77.
- Boersch-Supan, P.H., Rogers, A.D. & Brierley, A.S. (2017) The distribution of pelagic sound scattering layers across the southwest Indian Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **136**, 108–121.
- Brodeur, R.D. & Ware, D.M. (1992) Long-term variability in zooplankton biomass in the subarctic Pacific Ocean. *Fisheries Oceanography*, **1**, 32–38.
- Brodeur, R.D. & Yamamura, O. (2005) *Micronekton of the North Pacific*.
- Burd, B.J. & Thomson, R.E. (2000) Distribution and relative importance of jellyfish in a region of hydrothermal venting. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**, 1703–1721.
- Burghart, S.E., Hopkins, T.L. & Torres, J.J. (2007) The bathypelagic Decapoda, Lophogastrida, and Mysida of the eastern Gulf of Mexico. *Marine Biology*, **152**, 315–327.



- Cartes, J.E. (1998) Dynamics of the bathyal Benthic Boundary Layer in the northwestern Mediterranean: depth and temporal variations in macrofaunal– megafaunal communities and their possible connections within deep-sea trophic webs. *Progress in Oceanography*, **41**, 111–139.
- Cartes, J.E., Fanelli, E., López-Pérez, C. & Lebrato, M. (2013) Deep-sea macroplankton distribution (at 400 to 2300m) in the northwestern Mediterranean in relation to environmental factors. *Journal of Marine Systems*, **113-114**, 75–87.
- Cartes, J.E., Fanelli, E., Papiol, V. & Zucca, L. (2010) Distribution and diversity of open-ocean, near-bottom macroplankton in the western Mediterranean: Analysis at different spatio-temporal scales. Analysis at different spatio-temporal scales. *Deep Sea Research Part I: Oceanographic Research Papers*, **57**, 1485–1498.
- Cartes, J.E. & Maynou, F. (2001) Trophodynamics of the deep-water suprabenthic mysid *Boreomysis arctica* in the Catalan Sea (western Mediterranean). *Marine Ecology Progress Series*, **211**, 225–234.
- Cartes, J.E., Maynou, F., Fanelli, E., Romano, C., Mamouridis, V. & Papiol, V. (2009) The distribution of megabenthic, invertebrate epifauna in the Balearic Basin (western Mediterranean) between 400 and 2300 m: Environmental gradients influencing assemblages composition and biomass trends. Environmental gradients influencing assemblages composition and biomass trends. *Journal of Sea Research*, **61**, 244–257.
- Cartes, J.E., Soler-Membrives, A., Stefanescu, C., Lombarte, A. & Carrassón, M. (2016) Contributions of allochthonous inputs of food to the diets of benthopelagic fish over the northwest Mediterranean slope (to 2300 m). *Deep Sea Research Part I: Oceanographic Research Papers*, **109**, 123–136.
- Cavan, E.L., Trimmer, M., Shelley, F. & Sanders, R. (2017) Remineralization of particulate organic carbon in an ocean oxygen minimum zone. *Nature communications*, **8**, 14847.
- Choy, C.A., Haddock, S.H. D. & Robison, B.H. (2017) Deep pelagic food web structure as revealed by in situ feeding observations. *Proceedings of the Royal Society B: Biological Sciences*, **284**, 1–10.
- Collins, M.A., Xavier, J.C., Johnston, N.M., North, A.W., Enderlein, P., Tarling, G.A., Waluda, C.M., Hawker, E.J. & Cunningham, N.J. (2008) Patterns in the distribution of myctophid fish in the northern Scotia Sea ecosystem. *Polar Biology*, **31**, 837–851.
- Cook, A.B., Sutton, T.T., Galbraith, J.K. & Vecchione, M. (2013) Deep-pelagic (0–3000m) fish assemblage structure over the Mid-Atlantic Ridge in the area of the Charlie-Gibbs Fracture Zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, **98**, 279–291.
- Dai, L., Li, C., Tao, Z., Yang, G., Wang, X. & Zhu, M. (2017) Zooplankton abundance, biovolume and size spectra down to 3000 m depth in the western tropical North Pacific during autumn 2014. *Deep Sea Research Part I: Oceanographic Research Papers*, **121**, 1–13.
- Danovaro, R., Carugati, L., Boldrin, A., Calafat, A., Canals, M., Fabres, J., Finlay, K., Heussner, S., Miserocchi, S. & Sanchez-Vidal, A. (2017) Deep-water zooplankton in the Mediterranean Sea: Results from a continuous, synchronous sampling over different regions using sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers*, **126**, 103–114.
- Davison, P. (2011) The specific gravity of mesopelagic fish from the northeastern Pacific Ocean and its implications for acoustic backscatter. *ICES Journal of Marine Science*, **68**, 2064–2074.
- Davison, P., Lara-Lopez, A. & Anthony Koslow, J. (2015) Mesopelagic fish biomass in the southern California current ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, **112**, 129–142.
- D'Elia, M., Warren, J.D., Rodriguez-Pinto, I., Sutton, T.T., Cook, A. & Boswell, K.M. (2016) Diel variation in the vertical distribution of deep-water scattering layers in the Gulf of Mexico. *Deep Sea Research Part I: Oceanographic Research Papers*, **115**, 91–102.
- Denda, A., Stefanowitsch, B. & Christiansen, B. (2017a) From the epipelagic zone to the abyss: Trophic structure at two seamounts in the subtropical and tropical Eastern Atlantic - Part I zooplankton and micronekton. *Deep Sea Research Part I: Oceanographic Research Papers*, **130**, 63–77.
- Denda, A., Stefanowitsch, B. & Christiansen, B. (2017b) From the epipelagic zone to the abyss: Trophic structure at two seamounts in the subtropical and tropical Eastern Atlantic - Part II Benthopelagic fishes. *Deep Sea Research Part I: Oceanographic Research Papers*, **130**, 78–92.

- Di Carlo, B., Ianora, A., Fresi, E. & Hure, J. (1984) Vertical zonation patterns for Mediterranean copepods from the surface to 3000 m at a fixed station in the Tyrrhenian Sea. *Journal of Plankton Research*, **6**, 1031–1056.
- Feagans-Bartow, J.N. & Sutton, T.T. (2014) Ecology of the oceanic rim: pelagic eels as key ecosystem components. Pelagic eels as key ecosystem components. *Marine Ecology Progress Series*, **502**, 257–266.
- Fernandes, V. & Ramaiah, N. (2013) Mesozooplankton community structure in the upper 1,000 m along the western Bay of Bengal during the 2002 fall intermonsoon. *Zoological Studies*, **52**, 31.
- Flock, M.E. & Hopkins, T.L. (1992) Species Composition, Vertical Distribution, and Food Habits of the Sergestid Shrimp Assemblage in the Eastern Gulf of Mexico. *Journal of Crustacean Biology*, **12**, 210–223.
- Flores, H., Hunt, B.P., Kruse, S., Pakhomov, E.A., Siegel, V., van Franeker, J.A., Strass, V., van de Putte, A.P., Meesters, E.H. & Bathmann, U. (2014) Seasonal changes in the vertical distribution and community structure of Antarctic macrozooplankton and micronekton. *Deep Sea Research Part I: Oceanographic Research Papers*, **84**, 127–141.
- Foxton, P. (1956) The distribution of the standing crop of the zooplankton in the Southern Ocean. *Discovery Reports*, **XXVIII**, 191–236.
- Gaard, E., Gislason, A., Falkenhaus, T., Sjøiland, H., Musaeva, E., Vereshchaka, A. & Vinogradov, G. (2008) Horizontal and vertical copepod distribution and abundance on the Mid-Atlantic Ridge in June 2004. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 59–71.
- Gerringer, M.E., Andrews, A.H., Huss, G.R., Nagashima, K., Popp, B.N., Linley, T.D., Gallo, N.D., Clark, M.R., Jamieson, A.J. & Drazen, J.C. (2018) Life history of abyssal and hadal fishes from otolith growth zones and oxygen isotopic compositions. *Deep Sea Research Part I: Oceanographic Research Papers*, **132**, 37–50.
- Gerringer, M.E., Popp, B.N., Linley, T.D., Jamieson, A.J. & Drazen, J.C. (2017) Comparative feeding ecology of abyssal and hadal fishes through stomach content and amino acid isotope analysis. *Deep Sea Research Part I: Oceanographic Research Papers*, **121**, 110–120.
- Giering, S.L. C., Sanders, R., Lampitt, R.S., Anderson, T.R., Tamburini, C., Boutrif, M., Zubkov, M.V., Marsay, C.M., Henson, S.A., Saw, K., Cook, K. & Mayor, D.J. (2014) Reconciliation of the carbon budget in the ocean's twilight zone. *Nature*, **507**, 480–483.
- Giorli, G., Drazen, J.C., Neuheimer, A.B., Copeland, A. & Au, W.W. (2018) Deep sea animal density and size estimated using a Dual-frequency IDentification SONar (DIDSON) offshore the island of Hawaii. *Progress in Oceanography*, **160**, 155–166.
- Gluchowska, M., Trudnowska, E., Goszczko, I., Kubiszyn, A.M., Blachowiak-Samolyk, K., Walczowski, W. & Kwasniewski, S. (2017) Variations in the structural and functional diversity of zooplankton over vertical and horizontal environmental gradients en route to the Arctic Ocean through the Fram Strait. *PloS one*, **12**, e0171715.
- Gowing, M.M., Garrison, D.L., Wishner, K.F. & Gelfman, C. (2003) Mesopelagic microplankton of the Arabian Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, **50**, 1205–1234.
- Hannides, C.C. S., Popp, B.N., Choy, C.A. & Drazen, J.C. (2013) Midwater zooplankton and suspended particle dynamics in the North Pacific Subtropical Gyre: A stable isotope perspective. A stable isotope perspective. *Limnology and Oceanography*, **58**, 1931–1946.
- Hernandez-Leon, S. (2004) A global assessment of mesozooplankton respiration in the ocean. *Journal of Plankton Research*, **27**, 153–158.
- Herring, P.J., Fasham, M., Weeks, A.R., Hemmings, J., Roe, H., Pugh, P.R., Holley, S., Crisp, N.A. & Angel, M.V. (1998) Across-slope relations between the biological populations, the euphotic zone and the oxygen minimum layer off the coast of Oman during the southwest monsoon (August, 1994). *Progress in Oceanography*, **41**, 69–109.
- Hirche, H.J., Barz, K., Ayon, P. & Schulz, J. (2014) High resolution vertical distribution of the copepod *Calanus chilensis* in relation to the shallow oxygen minimum zone off northern Peru using LOKI, a new plankton imaging system. *Deep Sea Research Part I: Oceanographic Research Papers*, **88**, 63–73.

- Hopkins, T.L. (1971) Zooplankton Standing Crop in the Pacific Sector of the Antarctic. *Biology of the Antarctic seas IV* (ed. by G.A. Llano and I.E. Wallen). American Geophysical Union of the National Academy of Sciences-National Research Council, Washington, pp. 347–362.
- Hopkins, T.L. (1985) The Zooplankton community of Croker Passage, Antarctic Peninsula. *Polar Biology*, **4**, 161–170.
- Hopkins, T.L., Ainley, D.G., Torres, J.J. & Lancraft, T.M. (1993) Trophic structure in open waters of the marginal ice zone in the Scotia-Weddell confluence region during spring (1983). *Polar Biology*, **13**, 389–397.
- Hopkins, T.L. & Torres, J.J. (1988) The zooplankton community in the vicinity of the ice edge, western Weddell Sea, March 1986. *Polar Biology*, **9**, 79–87.
- Hunt, J.C. & Lindsay, D.J. (1999) Hunt & Lindsay 1999. *Plankton Biol. Ecol.*, **46**, 75–87.
- Irigoiien, X., Klevjer, T.A., Røstad, A., Martinez, U., Boyra, G., Acuña, J.L., Bode, A., Echevarria, F., Gonzalez-Gordillo, J.I., Hernandez-Leon, S., Agustí, S., Aksnes, D.L., Duarte, C.M. & Kaartvedt, S. (2014) Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nature communications*, **5**, 3271.
- Kaartvedt, S., Staby, A. & Aksnes, D.L. (2012) Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass. *Marine Ecology Progress Series*, **456**, 1–6.
- Karuppasamy, P.K., Raj, C., Muraleedharan, K.R. & Nair, M. (2011) Myctophid and pelagic shrimp assemblages in the oxygen minimum zone Andaman Sea during the winter monsoon. *Indian journal of Geo-Marine Sciences*, **40**, 535–541.
- Kenchington, E.L., Baillie, S.M., Kenchington, T.J. & Bentzen, P. (2017) Barcoding Atlantic Canada's mesopelagic and upper bathypelagic marine fishes. *PLoS one*, **12**, e0185173.
- Kinsey, S.T. & Hopkins, T.L. (1994) Trophic strategies of euphausiids in a low-latitude ecosystem. *Marine Biology*, **118**, 651–661.
- Klevjer, T.A., Irigoien, X., Røstad, A., Fraile-Nuez, E., Benítez-Barrios, V.M. & Kaartvedt, S. (2016) Large scale patterns in vertical distribution and behaviour of mesopelagic scattering layers. *Scientific reports*, **6**, 19873.
- Kloser, R.J., Ryan, T.E., Young, J.W. & Lewis, M.E. (2009) Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges. *ICES Journal of Marine Science*, **66**, 998–1006.
- Kobari, T., Kitamura, M., Minowa, M., Isami, H., Akamatsu, H., Kawakami, H., Matsumoto, K., Wakita, M. & Honda, M.C. (2013) Impacts of the wintertime mesozooplankton community to downward carbon flux in the subarctic and subtropical Pacific Oceans. *Deep Sea Research Part I: Oceanographic Research Papers*, **81**, 78–88.
- Kobari, T., Steinberg, D.K., Ueda, A., Tsuda, A., Silver, M.W. & Kitamura, M. (2008) Impacts of ontogenetically migrating copepods on downward carbon flux in the western subarctic Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 1648–1660.
- Koppelman, R. (2003) Vertical distribution of mesozooplankton and its  $\delta^{15}\text{N}$  signature at a deep-sea site in the Levantine Sea (eastern Mediterranean) in April 1999. *Journal of Geophysical Research*, **108**, 69.
- Koslow, J.A., Goericke, R., Lara-Lopez, A. & Watson, W. (2011) Impact of declining intermediate-water oxygen on deepwater fishes in the California Current. *Marine Ecology Progress Series*, **436**, 207–218.
- Koslow, J.A., Kloser, R.J. & Williams, A. (1997) Pelagic biomass and community structure over the mid-continental slope off southern Australia based upon acoustic and midwater trawl sampling. *Marine Ecology Progress Series*, **146**, 21–35.
- Kosobokova, K. & Hirche, H.-J. (2009) Biomass of zooplankton in the eastern Arctic Ocean – A base line study. *Progress in Oceanography*, **82**, 265–280.
- Kosobokova, K.N. & Hopcroft, R.R. (2010) Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 96–110.
- Kosobokova, K.N., Hopcroft, R.R. & Hirche, H.-J. (2011) Patterns of zooplankton diversity through the depths of the Arctic's central basins. *Marine Biodiversity*, **41**, 29–50.
- Lam, P.J., Doney, S.C. & Bishop, J.K. B. (2011) The dynamic ocean biological pump: Insights from a global compilation of particulate organic carbon,  $\text{CaCO}_3$ , and opal concentration profiles from the

- mesopelagic. Insights from a global compilation of particulate organic carbon, CaCO<sub>3</sub>, and opal concentration profiles from the mesopelagic. *Global Biogeochemical Cycles*, **25**, n/a-n/a.
- Lancraft, T.M., Reisenbichler, K.R., Robison, B.H., Hopkins, T.L. & Torres, J.J. (2004) A krill-dominated micronekton and macrozooplankton community in Croker Passage, Antarctica with an estimate of fish predation. *Deep Sea Research Part II: Topical Studies in Oceanography*, **51**, 2247–2260.
- Lancraft, T.M., Torres, J.J. & Hopkins, T.L. (1989) Micronekton and macrozooplankton in the open waters near Antarctic ice edge zones (AMERIEZ 1983 and 1986). *Polar Biology*, **9**, 225–233.
- Larson, R.J., Mills, C.E. & Harbison, G.R. (1991) Western Atlantic midwater hydrozoan and scyphozoan medusae: *in situ* studies using manned submersibles. *Hydrobiologia*, **216/217**.
- Last, K.S., Hobbs, L., Berge, J., Brierley, A.S. & Cottier, F. (2016) Moonlight Drives Ocean-Scale Mass Vertical Migration of Zooplankton during the Arctic Winter. *Current biology : CB*, **26**, 244–251.
- Letessier, T.B., Falkenhaus, T., Debes, H., Bergstad, O.A. & Brierley, A.S. (2011) Abundance patterns and species assemblages of euphausiids associated with the Mid-Atlantic Ridge, North Atlantic. *Journal of Plankton Research*, **33**, 1510–1525.
- Letessier, T.B., Grave, S. de, Boersch-Supan, P.H., Kemp, K.M., Brierley, A.S. & Rogers, A.D. (2017) Seamount influences on mid-water shrimps (Decapoda) and gnathophausiids (Lophogastridea) of the South-West Indian Ridge. *Deep Sea Research Part II: Topical Studies in Oceanography*, **136**, 85–97.
- Lindsay, D.J. & Hunt, J.C. (2005) Biodiversity in midwater cnidarians and ctenophores: submersible-based results from deep-water bays in the Japan Sea and north-western Pacific. Submersible-based results from deep-water bays in the Japan Sea and north-western Pacific. *Journal of the Marine Biological Association of the UK*, **85**, 503–517.
- Longhurst, A.R. (2007) *Ecological geography of the sea*, 2nd edn. Academic Press, Burlington, MA.
- Lopes, R.M., Marcolin, C.R. & Brandini, F.P. (2016) Influence of oceanic fronts on mesozooplankton abundance and grazing during spring in the south-western Atlantic. *Marine and Freshwater Research*, **67**, 626.
- Mackie, G. (1985) Midwater macroplankton of British Columbia studied by submersible PISCES IV. *Journal of Plankton Research*, **7**, 753–777.
- Marrari, M., Daly, K.L., Timonin, A. & Semenova, T. (2011) The zooplankton of Marguerite Bay, western Antarctic Peninsula—Part II: Vertical distributions and habitat partitioning. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58**, 1614–1629.
- Martínez-García, S. (2017) Microbial respiration in the mesopelagic zone at Station ALOHA. *Limnology and Oceanography*, **62**, 320–333.
- Melo, P.A. M. C., Melo Júnior, M. de, Macêdo, S.J. de, Araujo, M. & Neumann-Leitão, S. (2014) Copepod distribution and production in a Mid-Atlantic Ridge archipelago. *Anais da Academia Brasileira de Ciências*, **86**, 1719–1733.
- Milici, M., Vital, M., Tomasch, J., Badewien, T.H., Giebel, H.-A., Plumeier, I., Wang, H., Pieper, D.H., Wagner-Döbler, I. & Simon, M. (2017) Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: Evidence of dispersal limitation in the Bransfield Strait. *Limnology and Oceanography*, **62**, 1080–1095.
- Minutoli, R., Granata, A., Brugnano, C., Zagami, G. & Guglielmo, L. (2014) Mesozooplankton carbon requirement in the southern Adriatic Sea: vertical distribution, diel and seasonal variability, relation to particle flux. Vertical distribution, diel and seasonal variability, relation to particle flux. *Marine Ecology Progress Series*, **495**, 91–104.
- Minutoli, R. & Guglielmo, L. (2012) Mesozooplankton carbon requirement in the Tyrrhenian Sea: its vertical distribution, diel variability and relation to particle flux. Its vertical distribution, diel variability and relation to particle flux. *Marine Ecology Progress Series*, **446**, 91–105.
- Morita, H., Toyokawa, M., Hidaka, K., Nishimoto, A., Sugisaki, H. & Kikuchi, T. (2017) Spatio-temporal structure of the jellyfish community in the transition zone of cold and warm currents in the northwest Pacific. *Plankton and Benthos Research*, **12**, 266–284.
- Naito, Y., Costa, D.P., Adachi, T., Robinson, P.W., Peterson, S.H., Mitani, Y. & Takahashi, A. (2017) Oxygen minimum zone: An important oceanographic habitat for deep-diving northern elephant seals,

- Mirounga angustirostris. An important oceanographic habitat for deep-diving northern elephant seals, *Mirounga angustirostris*. *Ecology and evolution*, **7**, 6259–6270.
- Olivar, M.P., Hulley, P.A., Castellón, A., Emelianov, M., López, C., Tuset, V.M., Contreras, T. & Molí, B. (2017) Mesopelagic fishes across the tropical and equatorial Atlantic: Biogeographical and vertical patterns. *Biogeographical and vertical patterns*. *Progress in Oceanography*, **151**, 116–137.
- Pages, F., Pugh, P.R. & Gili, J.-M. (1994) Macro- and megaplanktonic cnidarians collected in the eastern part of the Weddell Gyre during summer 1979. *J. mar. biol. Ass. U.K.*, **74**, 873–894.
- Pages, F., White, M.G. & Rodhouse, P.G. (1996) Abundance of gelatinous carnivores in nekton community of the antarctic Polar Frontal Zone in summer 1994. *Marine Ecology Progress Series*, **141**, 139–147.
- Paulmier, A. & Ruiz-Pino, D. (2009) Oxygen minimum zones (OMZs) in the modern ocean. *Progress in Oceanography*, **80**, 113–128.
- Peña, M., Olivar, M.P., Balbín, R., López-Jurado, J.L., Iglesias, M. & Miquel, J. (2014) Acoustic detection of mesopelagic fishes in scattering layers of the Balearic Sea (western Mediterranean). *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 1186–1197.
- Phillips, B.T., Gruber, D.F., Vasan, G., Roman, C.N., Pieribone, V.A. & Sparks, J.S. (2016) Observations of in situ deep-sea marine bioluminescence with a high-speed, high-resolution sCMOS camera. *Deep Sea Research Part I: Oceanographic Research Papers*, **111**, 102–109.
- Piatkowski, U., Rodhouse, P.G., White, M.G., Bone, D.G. & Symon, C. (1994) Nekton community of the Scotia Sea as sampled by the RMT 25 during austral summer. *Marine Ecology Progress Series*, **112**, 13–28.
- Pierrot-Bults, A.C. (2008) A short note on the biogeographic patterns of the Chaetognatha fauna in the North Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 137–141.
- Preciado, I., Cartes, J.E., Punzón, A., Frutos, I., López-López, L. & Serrano, A. (2017) Food web functioning of the benthopelagic community in a deep-sea seamount based on diet and stable isotope analyses. *Deep Sea Research Part II: Topical Studies in Oceanography*, **137**, 56–68.
- Proud, R., Cox, M.J. & Brierley, A.S. (2017) Biogeography of the Global Ocean's Mesopelagic Zone. *Current biology : CB*, **27**, 113–119.
- Pusch, C., Beckmann, A., Porteiro, F.M. & Westernhagen, H. von (2004) The influence of seamounts on mesopelagic fish communities. *Arch. Fish. Mar. Res.*, **51**, 165–186.
- Raskoff, K.A., Hopcroft, R.R., Kosobokova, K.N., Purcell, J.E. & Youngbluth, M. (2010) Jellies under ice: ROV observations from the Arctic 2005 hidden ocean expedition. ROV observations from the Arctic 2005 hidden ocean expedition. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 111–126.
- Robison, B.H. (2003) What drives the diel vertical migrations of Antarctic midwater fish? *Journal of the Marine Biological Association of the UK*, **83**, 639–642.
- Rosa, R., Boavida-Portugal, J., Trübenbach, K., Baptista, M., Araújo, R. & Calado, R. (2012) Descending into the abyss: Bathymetric patterns of diversity in decapod crustaceans shift with taxonomic level and life strategies. Bathymetric patterns of diversity in decapod crustaceans shift with taxonomic level and life strategies. *Deep Sea Research Part I: Oceanographic Research Papers*, **64**, 9–21.
- Samemoto, D.D. (1987) Vertical distribution and ecological significance of chaetognaths in the Arctic environment of Baffin Bay. *Polar Biology*, **7**, 317–328.
- Sameoto, D.D. (1982) Zooplankton and Micronekton Abundance in Acoustic Scattering Layers on the Nova Scotian Slope. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**, 760–777.
- Sato, M. & Benoit-Bird, K.J. (2017) Spatial variability of deep scattering layers shapes the Bahamian mesopelagic ecosystem. *Marine Ecology Progress Series*, **580**, 69–82.
- Saunders, R.A., Collins, M.A., Foster, E., Shreeve, R., Stowasser, G., Ward, P. & Tarling, G.A. (2014) The trophodynamics of Southern Ocean Electrona (Myctophidae) in the Scotia Sea. *Polar Biology*, **37**, 789–807.
- Saunders, R.A. & Tarling, G.A. (2018) Southern Ocean Mesopelagic Fish Comply with Bergmann's Rule. *The American Naturalist*, **191**, 343–351.

- Schnack-Schiel, S.B., Michels, J., Mizdalski, E., Schodlok, M.P. & Schröder, M. (2008) Composition and community structure of zooplankton in the sea ice-covered western Weddell Sea in spring 2004—with emphasis on calanoid copepods. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 1040–1055.
- Schukat, A., Bode, M., Auel, H., Carballo, R., Martin, B., Koppelman, R. & Hagen, W. (2013) Pelagic decapods in the northern Benguela upwelling system: Distribution, ecophysiology and contribution to active carbon flux. Distribution, ecophysiology and contribution to active carbon flux. *Deep Sea Research Part I: Oceanographic Research Papers*, **75**, 146–156.
- Shreeve, R.S., Collins, M.A., Tarling, G.A., Main, C.E., Ward, P. & Johnston, N.M. (2009) Feeding ecology of myctophid fishes in the northern Scotia Sea. *Marine Ecology Progress Series*, **386**, 221–236.
- Silguero, J. (2000) Seasonal abundance and vertical distribution of mesopelagic calycophoran siphonophores in Monterey Bay, CA. *Journal of Plankton Research*, **22**, 1139–1153.
- Siokou, I., Zervoudaki, S. & Christou, E.D. (2013) Meso-zooplankton community distribution down to 1000 m along a gradient of oligotrophy in the Eastern Mediterranean Sea (Aegean Sea). *Journal of Plankton Research*, **35**, 1313–1330.
- Smoot, C.A. & Hopcroft, R.R. (2017) Depth-stratified community structure of Beaufort Sea slope zooplankton and its relations to water masses. *Journal of Plankton Research*, **39**, 79–91.
- Sogawa, S., Sugisaki, H., Saito, H., Okazaki, Y., Ono, T., Shimode, S. & Kikuchi, T. (2016) Seasonal and regional change in vertical distribution and diel vertical migration of four euphausiid species (*Euphausia pacifica*, *Thysanoessa inspinata*, *T. longipes*, and *Tessarabrachion oculatum*) in the northwestern Pacific. *Deep Sea Research Part I: Oceanographic Research Papers*, **109**, 1–9.
- Staby, A., Røstad, A. & Kaartvedt, S. (2011) Long-term acoustical observations of the mesopelagic fish *Maurollicus muelleri* reveal novel and varied vertical migration patterns. *Marine Ecology Progress Series*, **441**, 241–255.
- Steinberg, D.K., Carlson, C.A., Bates, N.R., Goldthwait, S.A., Madin, L.P. & Michaels, A.F. (2000) Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**, 137–158.
- Steinberg, D.K., Cope, J.S., Wilson, S.E. & Kobari, T. (2008) A comparison of mesopelagic mesozooplankton community structure in the subtropical and subarctic North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 1615–1635.
- Stramma, L., Brandt, P., Schafstall, J., Schott, F., Fischer, J. & Körtzinger, A. (2008) Oxygen minimum zone in the North Atlantic south and east of the Cape Verde Islands. *Journal of Geophysical Research*, **113**, 829.
- Stramma, L., Schmidtko, S., Levin, L.A. & Johnson, G.C. (2010) Ocean oxygen minima expansions and their biological impacts. *Deep Sea Research Part I: Oceanographic Research Papers*, **57**, 587–595.
- Sutton, T.T., Letessier, T.B. & Bardarson, B. (2013) Midwater fishes collected in the vicinity of the Sub-Polar Front, Mid-North Atlantic Ocean, during ECOMAR pelagic sampling. *Deep Sea Research Part II: Topical Studies in Oceanography*, **98**, 292–300.
- Sutton, T.T., Porteiro, F.M., Heino, M., Byrkjedal, I., Langhelle, G., Anderson, C., Horne, J., Sjøiland, H., Falkenhaus, T., Godø, O.R. & Bergstad, O.A. (2008) Vertical structure, biomass and topographic association of deep-pelagic fishes in relation to a mid-ocean ridge system. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 161–184.
- Sweetman, C.J., Sutton, T.T., Vecchione, M. & Latour, R.J. (2013) Distribution of the biomass-dominant pelagic fish, *Bathylagus euryops* (Argentiniformes: Microstomatidae), along the northern Mid-Atlantic Ridge. *Deep Sea Research Part I: Oceanographic Research Papers*, **78**, 16–23.
- Toda, R., Lindsay, D.J., Fuentes, V.L. & Moteki, M. (2014) Community structure of pelagic cnidarians off Adélie Land, East Antarctica, during austral summer 2008. *Polar Biology*, **37**, 269–289.
- Toda, R., Moteki, M., Ono, A., Horimoto, N., Tanaka, Y. & Ishimaru, T. (2010) Structure of the pelagic cnidarian community in Lützw-Holm Bay in the Indian sector of the Southern Ocean. *Polar Science*, **4**, 387–404.

- Torres, A.P., Reglero, P., Hidalgo, M., Abelló, P., Simão, D.S., Alemany, F., Massutí, E. & Dos Santos, A. (2018) Contrasting patterns in the vertical distribution of decapod crustaceans throughout ontogeny. *Hydrobiologia*, **808**, 137–152.
- Trueman, C.N., Johnston, G., O’Hea, B. & MacKenzie, K.M. (2014) Trophic interactions of fish communities at midwater depths enhance long-term carbon storage and benthic production on continental slopes. *Proceedings. Biological sciences*, **281**.
- Tsuda, A., Saito, H. & Kasai, H. (2014) Vertical distributions of large ontogenetically migrating copepods in the Oyashio region during their growing season. *Journal of Oceanography*, **70**, 123–132.
- Tv, R. (2017) Plankton and the Invisible Barriers in the Tropical Ocean. *Journal of Aquaculture & Marine Biology*, **6**.
- Uchimiya, M., Ogawa, H. & Nagata, T. (2016) Effects of temperature elevation and glucose addition on prokaryotic production and respiration in the mesopelagic layer of the western North Pacific. *Journal of Oceanography*, **72**, 419–426.
- Ulloa, R., Palma, S. & Silva, N. (2000) PII: S0967-0637(00)00020-0. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**, 2009–2027.
- Urmy, S.S. & Horne, J.K. (2016) Multi-scale responses of scattering layers to environmental variability in Monterey Bay, California. *Deep Sea Research Part I: Oceanographic Research Papers*, **113**, 22–32.
- Valls, M., Rueda, L. & Quetglas, A. (2017) Feeding strategies and resource partitioning among elasmobranchs and cephalopods in Mediterranean deep-sea ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, **128**, 28–41.
- van Haren, H. & Compton, T.J. (2013) Diel vertical migration in deep sea plankton is finely tuned to latitudinal and seasonal day length. *PLoS one*, **8**, e64435.
- Vereshchaka, A., Abyzova, G., Lunina, A. & Musaeva, E. (2017) The deep-sea zooplankton of the North, Central, and South Atlantic: Biomass, abundance, diversity. Biomass, abundance, diversity. *Deep Sea Research Part II: Topical Studies in Oceanography*, **137**, 89–101.
- Vereshchaka, A., Abyzova, G., Lunina, A., Musaeva, E. & Sutton, T. (2016) A novel approach reveals high zooplankton standing stock deep in the sea. *Biogeosciences*, **13**, 6261–6271.
- Vereshchaka, A.L. & Vinogradov, G.M. (1999) Visual observations of the vertical distribution of plankton throughout the water column above Broken Spur vent field, Mid-Atlantic Ridge. *Deep Sea Research Part I: Oceanographic Research Papers*, **46**, 1615–1632.
- Villarino, E., Watson, J.R., Jönsson, B., et al. (2018) Large-scale ocean connectivity and planktonic body size. *Nature communications*, **9**, 142.
- Vinnichenko, V.I. (1997) Russian Investigations and Deep Water Fishery on the Corner Seamount in Subarea 6. *NAFO Sci. Council Studies*, **30**, 41–49.
- Vinogradov, G.M. (2005) Vertical distribution of macroplankton at the Charlie-Gibbs Fracture Zone (North Atlantic), as observed from the manned submersible ?Mir-1? Mir-1? *Marine Biology*, **146**, 325–331.
- Vinogradov, G.M. & Vereshchaka, A.L. (2006) Zooplankton distribution above the Lost City (Atlantic massif) and broken spur hydrothermal fields in the North Atlantic according to the data of visual observations. *Oceanology*, **46**, 217–227.
- Vinogradov, G.M., Vereshchaka, A.L. & Aleinik, D.L. (2003) Zooplankton Distribution over Hydrothermal Fields of the Mid-Atlantic Ridge. *Oceanology*, **43**, 656–669.
- Vinogradov, G.M., Vereshchaka, A.L., Shushkina, E.A., Arnautov, G.N. & D'yakonov, V. (1997) Vertical Distribution of zooplankton over the Broken Spur Hydrothermal field in the North Atlantic Gyre (29°N, 34°W Water). *Oceanology*, **37**, 502–512.
- Vinogradov, M.E. & Shushkina, E.A. (2002) Vertical Distribution of Gelatinous Macroplankton in the North Pacific Observed by Manned Submersibles *Mir-1* and *Mir-2*. *Journal of Oceanography*, **58**, 295–303.
- Vinogradov, M.E., Vereshchaka, A.L. & Shushkina, E.A. (1996) The vertical structure of zooplankton in the Oligotrophic areas of the North Atlantic and the effect of hydrothermal fields. *Oceanology*, **36**, 64–71.

- Vinogradov, M.E., Vereshchaka, A.L., Shushkina, E.A. & Arnautov, G.N. (1999) Structure of zooplankton communities of the frontal zone between the Gulf Stream and the Labrador Current. *Oceanology*, **39**, 504–514.
- Vinogradov, M.E., Vereshchaka, A.L., Vinogradov, G.M. & Musaeva, E. (2000) Vertical distribution of Zooplankton at the Periphery of the North Atlantic Subtropical Gyre. *Biology Bulletin*, **27**, 417–429.
- Ward, P., Atkinson, A. & Tarling, G. (2012a) Mesozooplankton community structure and variability in the Scotia Sea: A seasonal comparison. A seasonal comparison. *Deep Sea Research Part II: Topical Studies in Oceanography*, **59-60**, 78–92.
- Ward, P., Meredith, M.P., Whitehouse, M.J. & Rothery, P. (2008) The summertime plankton community at South Georgia (Southern Ocean): Comparing the historical (1926/1927) and modern (post 1995) records. *Progress in Oceanography*, **78**, 241–256.
- Ward, P. & Shreeve, R.S. (1999) The spring mesozooplankton community at South Georgia: a comparison of shelf and oceanic sites. *Polar Biology*, **22**, 289–301.
- Ward, P., Tarling, G.A., Coombs, S.H. & Enderlein, P. (2012b) Comparing Bongo net and N70 mesozooplankton catches: using a reconstruction of an original net to quantify historical plankton catch data. *Polar Biology*, **35**, 1179–1186.
- Ward, P., Tarling, G.A. & Thorpe, S.E. (2014) Mesozooplankton in the Southern Ocean: Spatial and temporal patterns from Discovery Investigations. *Progress in Oceanography*, **120**, 305–319.
- Weikert, H., Koppelman, R. & Wiegatz, S. (2001) Evidence of episodic changes in deep-sea mesozooplankton abundance and composition in the Levantine Sea (Eastern Mediterranean). *Journal of Marine Systems*, **30**, 221–239.
- Wiebe, P.H., Boyd, S.H., Davis, B.M. & Cox, J.L. (1982) Avoidance of towed nets by the euphausiid *Nematoscelis megalops*. *Fishery Bulletin*, **80**, 75–91.
- Wiebe, P.H., Bucklin, A., Kaartvedt, S., Røstad, A. & Blanco-Bercial, L. (2016) Vertical distribution and migration of euphausiid species in the Red Sea. *Journal of Plankton Research*, **38**, 888–903.
- Wieczorek, A.M., Morrison, L., Croot, P.L., Allcock, A.L., MacLoughlin, E., Savard, O., Brownlow, H. & Doyle, T.K. (2018) Frequency of Microplastics in Mesopelagic Fishes from the Northwest Atlantic. *Frontiers in Marine Science*, **5**, 1985.
- Williams, A., J., K., A., T. & K., H. (2001) Feeding ecology of five fishes from the mid-slope micronekton community off southern Tasmania, Australia. *Marine Biology*, **139**, 1177–1192.
- Williams, A. & Koslow, J.A. (1997) Species composition, biomass and vertical distribution of micronekton over the mid-slope region off southern Tasmania, Australia. *Marine Biology*, **130**, 259–276.
- Wishner, K.F., Outram, D.M., Seibel, B.A., Daly, K.L. & Williams, R.L. (2013) Zooplankton in the eastern tropical north Pacific: Boundary effects of oxygen minimum zone expansion. Boundary effects of oxygen minimum zone expansion. *Deep Sea Research Part I: Oceanographic Research Papers*, **79**, 122–140.
- Yamaguchi, A. (2000) Vertical distribution, life cycle and body allometry of two oceanic calanoid copepods (*Pleuromamma scutullata* and *Heterorhabdus tanneri*) in the Oyashio region, western North Pacific Ocean. *Journal of Plankton Research*, **22**, 29–46.
- Yamaguchi, A., Homma, T., Saito, R., Matsuno, K., Ueno, H., Hirawake, T. & Imai, I. (2013) East-west differences in population structure and vertical distribution of copepods along 47°N in the subarctic Pacific in June 2009. *Plankton and Benthos Research*, **8**, 116–123.
- Yamaguchi, A. & Ikeda, T. (2002) Vertical Distribution Patterns of Three Mesopelagic Paraeuchaeta Species (Copepoda: Calanoida) in the Oyashio Region, Western Subarctic Pacific Ocean. *Bull. Fish. Sci. Hokkaido Univ.*, **53**, 1–10.
- Yamaguchi, A., Ikeda, T., Watanabe, Y. & Ishizaka, J. (2004a) Vertical distribution patterns of pelagic copepods as viewed from the predation pressure hypothesis. *Zoological Studies*, **43**, 475–485.
- Yamaguchi, A., Watanabe, Y., Ishida, H., Harimoto, T., Furusawa, K., Suzuki, S., Ishizaka, J., Ikeda, T. & Mac Takahashi, M. (2002a) Structure and size distribution of plankton communities down to the greater depths in the western North Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **49**, 5513–5529.



- Yamaguchi, A., Watanabe, Y., Ishida, H., Harimoto, T., Furusawa, K., Suzuki, S., Ishizaka, J., Ikeda, T. & Mac Takahashi, M. (2004b) Latitudinal Differences in the Planktonic Biomass and Community Structure Down to the Greater Depths in the Western North Pacific. *Journal of Oceanography*, **60**, 773–787.
- Yamaguchi, A., Watanabe, Y., Ishida, H., Harimoto, T., Furusawa, K., Suzuki, S., Ishizaka, J., Ikeda, T. & Takahashi, M.M. (2002b) Community and trophic structures of pelagic copepods down to greater depths in the western subarctic Pacific (WEST-COSMIC). *Deep Sea Research Part I: Oceanographic Research Papers*, **49**, 1007–1025.
- Yamaguchi, A., Watanabe, Y., Ishida, H., Harimoto, T., Maeda, M., Ishizaka, J., Ikeda, T. & Mac Takahashi, M. (2005) Biomass and chemical composition of net-plankton down to greater depths (0–5800m) in the western North Pacific Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, **52**, 341–353.
- Young, J.W., Bradford, R.W., Lamb, T.D. & Lyne, V.D. (1996) Biomass of zooplankton and micronekton in the southern bluefin tuna fishing grounds off eastern Tasmania, Australia. *Marine Ecology Progress Series*, **138**, 1–14.
- Youngbluth, M., Sørnes, T., Hosia, A. & Stemmann, L. (2008) Vertical distribution and relative abundance of gelatinous zooplankton, in situ observations near the Mid-Atlantic Ridge. *Deep Sea Research Part II: Topical Studies in Oceanography*, **55**, 119–125.

## Appendix B - Supplementary information for Chapter 2

Table B1: Information collected from various published and unpublished sources for the mesopelagic mesozooplankton and micronekton databases.

Database field	Units/Format	Field description
ID	-	Unique ID number for each data entry
Data Source	Author et al. (year)	The lead author and year of the peer reviewed article examined.
Region	-	Generic geographic identification of the general area of study
Approx. Lower limit size	cm	The estimated lower size limit of zooplankton based on statements in the article in question (not always discernable)
Approx. Upper limit size	cm	The estimated upper size limit of zooplankton based on statements in the article in question (not always discernable)
Mean Size	cm	Mean size of the organism if given
Organism type	Nekton/Plankton/Both	Whether the animals sampled were nekton, plankton or both nekton and plankton
Ship	-	The name of the research vessel from which sampling was conducted
Time/ Time (gmt) RANGE	hh:mm-hh:mm	The Greenwich time period (range) of sampling
Local Time /Local Time RANGE	hh:mm-hh:mm	Local time period at the station of concern
Season	Summer/ Fall/ Spring/ Winter	Season when sampling took place: Summer, Fall, Spring, Winter
Time of Day	Day/ Night/Twilight	Whether it was day, night, or twilight. Twilight was defined as being within 1 h before/after sundown/sunrise.
Date/ Date RANGE	dd/mm/yyyy-dd/mm/yyyy	The date/date range over which sampling occurred.
Latitude/ Longitude	(N) (W) (E) (S)	the geographical coordinates of the stations taken directly from the literature source in the original forma (dd mm ss and in decimal degrees). NOTE, in some cases the geographical locations for each station were not provided in which case an estimated 'average' lat/long is provided.
Voyage/Cruise/Region	-	Where applicable the name of the voyage or cruise is included
Station name	-	Where applicable the name of the station where sampling occurred
Substation name	-	If any subdivision of study area or station is given

Database field	Units/Format	Field description
Depth range	m	Given depth range at which sampling was done
Median depth	m	Median depth at which sampling was done
Type of net haul	horizontal / vertical / oblique	kind of haul was recorded. N/S was used if type of haul was not given
Net type (OR OBS.)	-	Type of net. Or in some cases where vertical distribution is based on submersible observations obs. Is indicated.
mesh size	µm	Diameter of net mesh
Net Mouth Area	m <sup>2</sup>	Area of net mouth
Duration of haul	min	Where applicable the length of time in which the haul was conducted
Towing speed	cm·s <sup>-1</sup> knots	Where applicable, the speed of the haul
Volume of water Filtered	m <sup>3</sup>	Where applicable the volume of water filtered
Lower limit	m	In case of vertical tows, the lowest depth of the tow was recorded
Upper limit	m	In case of vertical tows, the shallowest depth of the tow was recorded
Abundance	specimens ·m <sup>-3</sup> ind. ·m <sup>-3</sup> ind. ·m <sup>-2</sup> # ind. ·litre <sup>-1</sup> # ind. % of total # ind.	Abundance of the organism or group of organisms. Different columns were used for different units. Abbreviations: individuals (ind.), number(#), and present (%)
Biomass	mg C ·litre <sup>-1</sup> mg C · m <sup>-3</sup> mg C · m <sup>-2</sup> g wet. wt. g wet. wt. · m <sup>-3</sup> g wet. wt. ·m <sup>-2</sup> g wet. wt. ·haul <sup>-1</sup> g dry. wt. ·m <sup>-3</sup> g dry. wt. ·m <sup>-2</sup> g · m <sup>-2</sup> g · m <sup>-3</sup> % of a total	Biomass of the organism or group of organisms. Different columns were used for different units. Abbreviations: carbon (C), wet weight (wet. wt), dry weight (dry. wt.), and present (%)
Biovolume	mL · m <sup>-3</sup>	

Database field	Units/Format	Field description
displacement volume	mL · m <sup>-2</sup>	
Catch rate	kg · h <sup>-1</sup> # · hour <sup>-1</sup> ind. · haul <sup>-1</sup>	
Total abundance	# ind. ind. · m <sup>-2</sup> ind. · m <sup>-3</sup>	If the total abundance of the community is given or the abundance of the entire water column
Total biomass	g dry. wt. · m <sup>-2</sup> g wet. wt. · m <sup>-2</sup> g dry. wt. · m <sup>-3</sup> g wet. wt. · m <sup>-3</sup> g · m <sup>-3</sup> mg C · m <sup>-3</sup>	If the total biomass estimate of the community is given or the abundance of the entire water column. Abbreviations: carbon (C), wet weight (wet. wt), dry weight (dry. wt.)
Total number of species	#	Number of species in a group, if given
sex/life stage/size/Other classification	-	Further information about species / taxa of concern
Species / Group names	-	Species of concern or in some cases broader classification at a broader level (this is the original grouping given in the paper)
Class/Order/Family/Phylum/Genus	-	Additional information about organism classification
Taxa	-	Taxa used for calibration of densities
NetCalib	ind · m <sup>-3</sup>	Mesh-adjusted densities for taxa listed in Taxa column
NetTowCalib	ind · m <sup>-3</sup>	Mesh and tow-adjusted densities for taxa listed in Taxa column
NetTowSeasonCalib	ind · m <sup>-3</sup>	Mesh-Tow-Season-adjusted densities for taxa listed in Taxa column
lowest taxonomic level	-	summary of the lowest taxonomic resolution available for each entry in the database
Reference	-	Full reference for peer reviewed article
NOTES	-	Any information which was important to elucidate but which could not be included in any of the previously mentioned categories

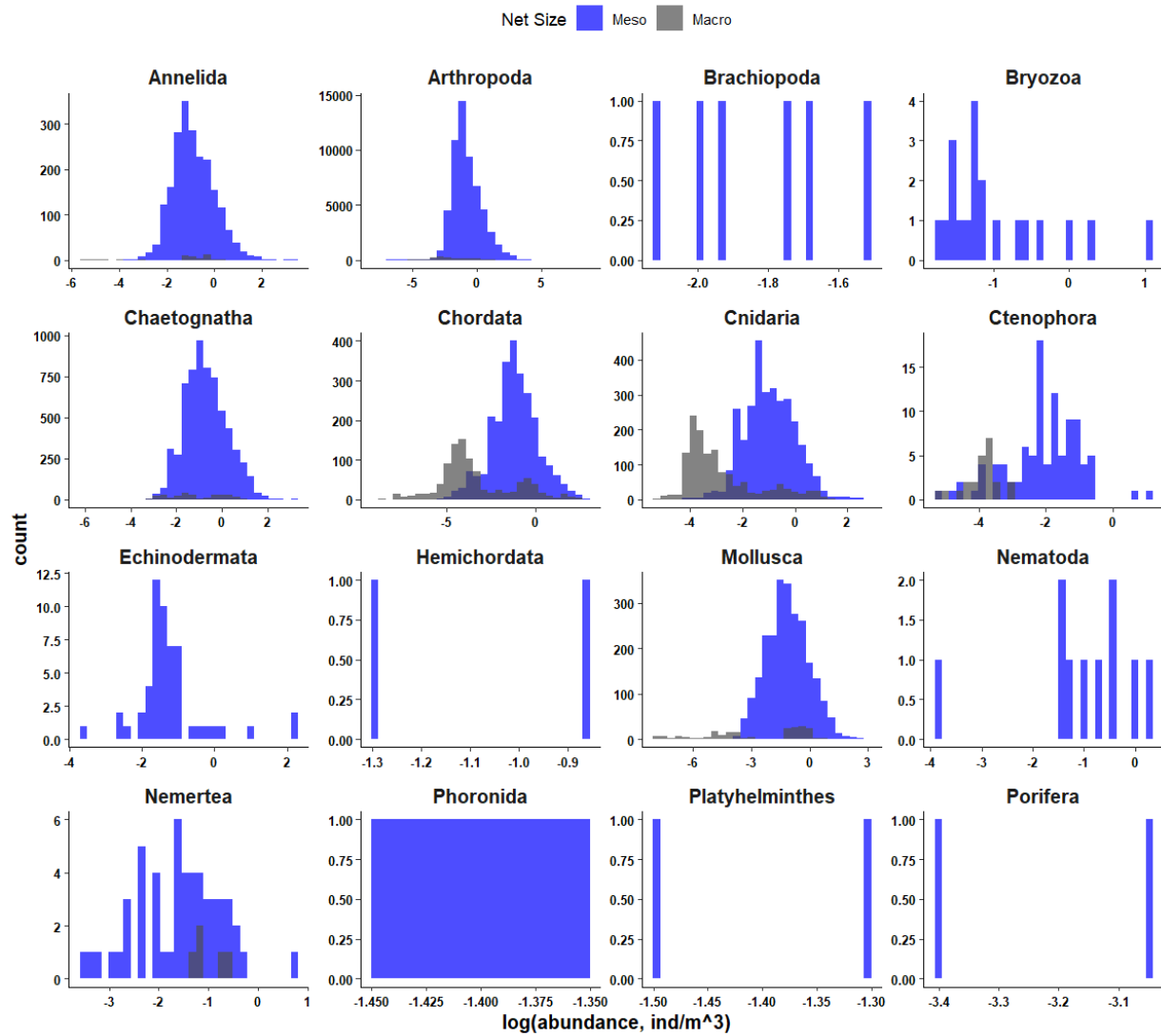


Figure B1: Histograms of logged densities (in  $\text{ind}\cdot\text{m}^{-3}$ ) of meso-(Meso) and macro-(Macro) zooplankton records from mesopelagic mesozooplankton and micronekton database grouped by phylum. Zero abundances were dropped.

Table B2: Number of entries, n, (with percentage, %) for 19 phyla from the Mesopelagic Mesozooplankton and Micronekton Database grouped by organism class and net size (Meso and Macro).

Phylum	Class	Meso		Macro	
		N	%	N	%
<b>Annelida</b>	phylum-level	12	0.5	-	-
	Polychaeta	2,315	99.5	135	100.0
<b>Arthropoda</b>	phylum-level	197	0.3	14	0.5
	Arachnida	51	0.1	-	-
	Branchiopoda	131	0.2	-	-
	Copepoda	61,856	84.6	338	13.0
	Hexapoda	65	0.1	-	-
	Malacostraca	7,349	10.1	2,228	85.4
	Ostracoda	3,258	4.5	29	1.1
	Pycnogonida	3	0.0	-	-
	Thecostraca	212	0.3	-	-
<b>Chaetognatha</b>	phylum-level	2,499	34.8	142	34.3
	Sagittoidea	4,691	65.2	272	65.7
<b>Chordata</b>	phylum-level	347	10.6	10	0.6
	Appendicularia	1,065	32.7	-	-
	Ascidiacea	3	0.1	-	-
	Larvacea	36	1.1	-	-
	Leptocardii	17	0.5	-	-
	Teleostei	737	22.6	1,454	86.2
	Thaliacea	1,054	32.3	222	13.2
<b>Cnidaria</b>	phylum-level	290	7.9	166	3.7
	Anthozoa	228	6.2	-	-
	Hydrozoa	3,056	83.4	3,821	85.5
	Scyphozoa	92	2.5	482	10.8
<b>Ctenophora</b>	phylum-level	195	78.9	55	100.0
	Nuda	12	4.9	-	-
	Tentaculata	40	16.2	-	-
<b>Echinodermata</b>	phylum-level	116	63.7	-	-
	Asteroidea	65	35.7	-	-
	Holothuroidea	1	0.5	-	-
<b>Mollusca</b>	phylum-level	365	9.5	-	-
	Bivalvia	135	3.5	-	-
	Cephalopoda	223	5.8	210	56.9
	Gastropoda	3,136	81.3	159	43.1

Table B3: Number of entries, n, (with percent, %) for 19 Phyla from the mesopelagic dataset grouped by organisms' order and net size (Meso and Macro)

<b>Phylum</b>	<b>Class</b>	<b>Order</b>	<b>N</b>	<b>%</b>	
<b>Annelida</b>	<b>Polychaeta</b>	Class-level	1,957	84.5	
		Phyllodocida	355	15.3	
		Spionida	3	0.1	
<b>Arthropoda</b>	<b>Arachnida</b>	Class-level	51	100.0	
	<b>Branchiopoda</b>	Class-level	14	10.7	
		Anomopoda	2	1.5	
		Cladocera	90	68.7	
		Ctenopoda	3	2.3	
		Onychopoda	22	16.8	
		<b>Copepoda</b>	Class-level	3,011	4.9
	Calanoida		46,766	75.6	
	Cyclopoida		9,688	15.7	
	Harpacticoida		1,504	2.4	
	Mormonilloida		658	1.1	
	Siphonostomatoida		229	0.4	
	<b>Hexapoda</b>	Class-level	65	100.0	
	<b>Malacostraca</b>	Class-level	103	1.4	
		Amphipoda	2,038	27.7	
		Cumacea	21	0.3	
		Decapoda	1,157	15.7	
		Euphausiacea	3,177	43.2	
		Isopoda	433	5.9	
		Leptostraca	11	0.1	
		Lophogastrida	65	0.9	
		Mysida	270	3.7	
		Mysidacea	67	0.9	
		Stomatopoda	7	0.1	
		<b>Ostracoda</b>	Class-level	2,909	89.3
			Halocyprida	349	10.7
	<b>Pycnogonida</b>	Class-level	3	100.0	
<b>Thecostraca</b>	Class-level	200	94.3		
	Balanomorpha	12	5.7		
<b>Chaetognatha</b>	<b>Sagittoidea</b>	Aphragmophora	2,749	58.6	
		Phragmophora	1,942	41.4	
<b>Chordata</b>	<b>Appendicularia</b>	Copelata	1,065	100.0	
	<b>Asciacea</b>	Class-level	3	100.0	
	<b>Larvacea</b>	Class-level	36	100.0	
	<b>Leptocardii</b>	Class-level	17	100.0	
	<b>Teleostei</b>	Anguilliformes	8	1.1	
		Argentiniiformes	107	14.5	
		Aulopiformes	25	3.4	
		Gadiformes	6	0.8	
		Holocentriformes	10	1.4	
		Myctophiformes	300	40.7	
Perciformes	15	2.0			
Scombriformes	2	0.3			

<b>Phylum</b>	<b>Class</b>	<b>Order</b>	<b>N</b>	<b>%</b>
		Stomiiformes	264	35.8
	<b>Thaliacea</b>	Class-level	498	47.2
		Doliolida	173	16.4
		Pyrosomatida	21	2.0
		Salpida	362	34.3
<b>Cnidaria</b>	<b>Anthozoa</b>	Class-level	188	82.5
		Actiniaria	40	17.5
	<b>Hydrozoa</b>	Class-level	287	9.4
		Anthoathecata	46	1.5
		Leptothecata	13	0.4
		Narcomedusae	109	3.6
		Siphonophorae	2,304	75.4
		Trachymedusae	297	9.7
	<b>Scyphozoa</b>	Class-level	65	70.7
		Coronatae	21	22.8
		Semaeostomeae	6	6.5
<b>Mollusca</b>	<b>Bivalvia</b>	Class-level	135	100.0
	<b>Cephalopoda</b>	Class-level	214	96.0
		Octopoda	3	1.3
		Oegopsida	6	2.7
	<b>Gastropoda</b>	Class-level	446	14.2
		[unassigned] Caenogastropoda	14	0.4
		Littorinimorpha	136	4.3
		Neogastropoda	12	0.4
		Pleurotomariida	12	0.4
		Pteropoda	1,954	62.3
		Thecosomata	562	17.9



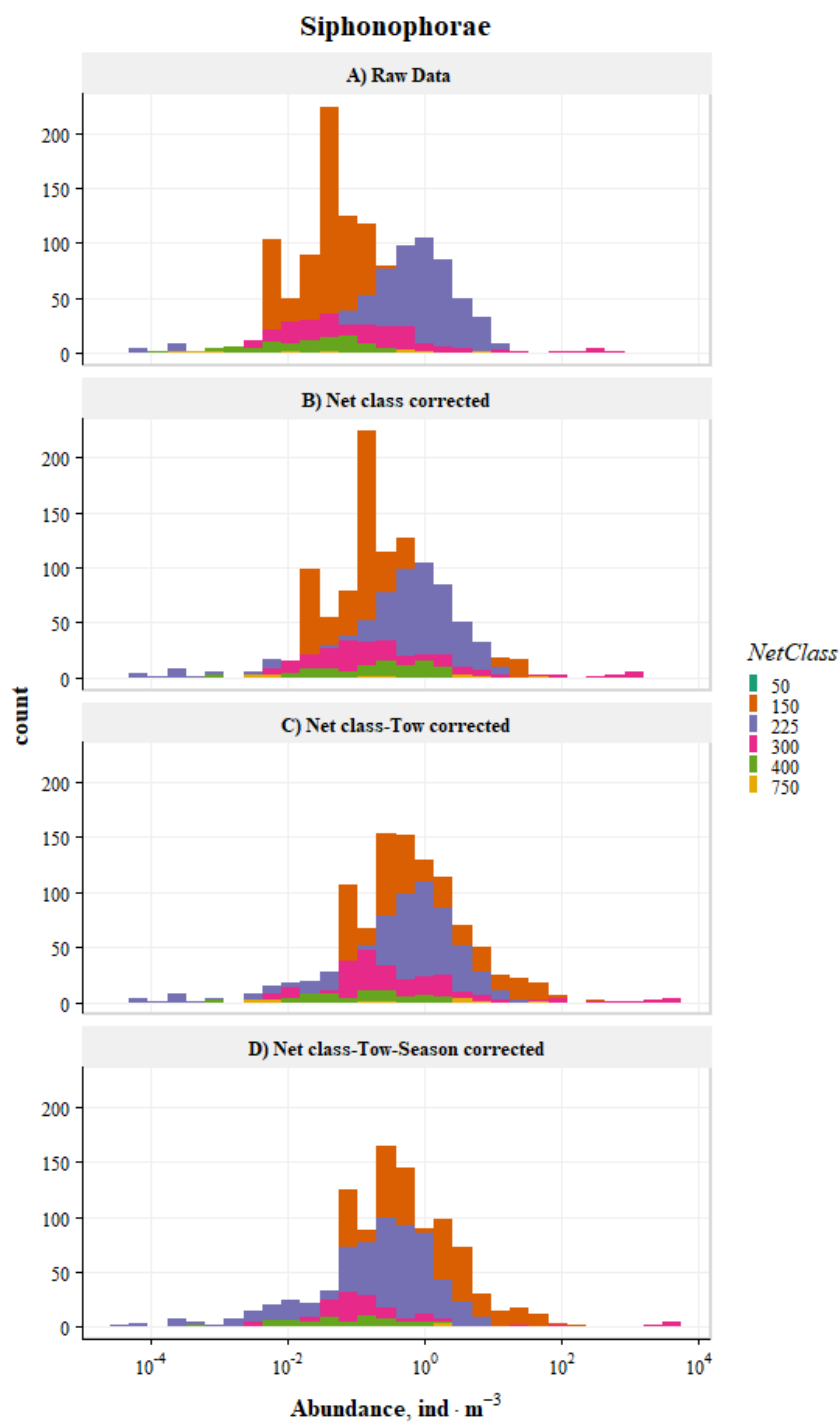


Figure B2: Example of calibration procedure for order Siphonophorae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the log<sub>10</sub> transformation on *x* axis.

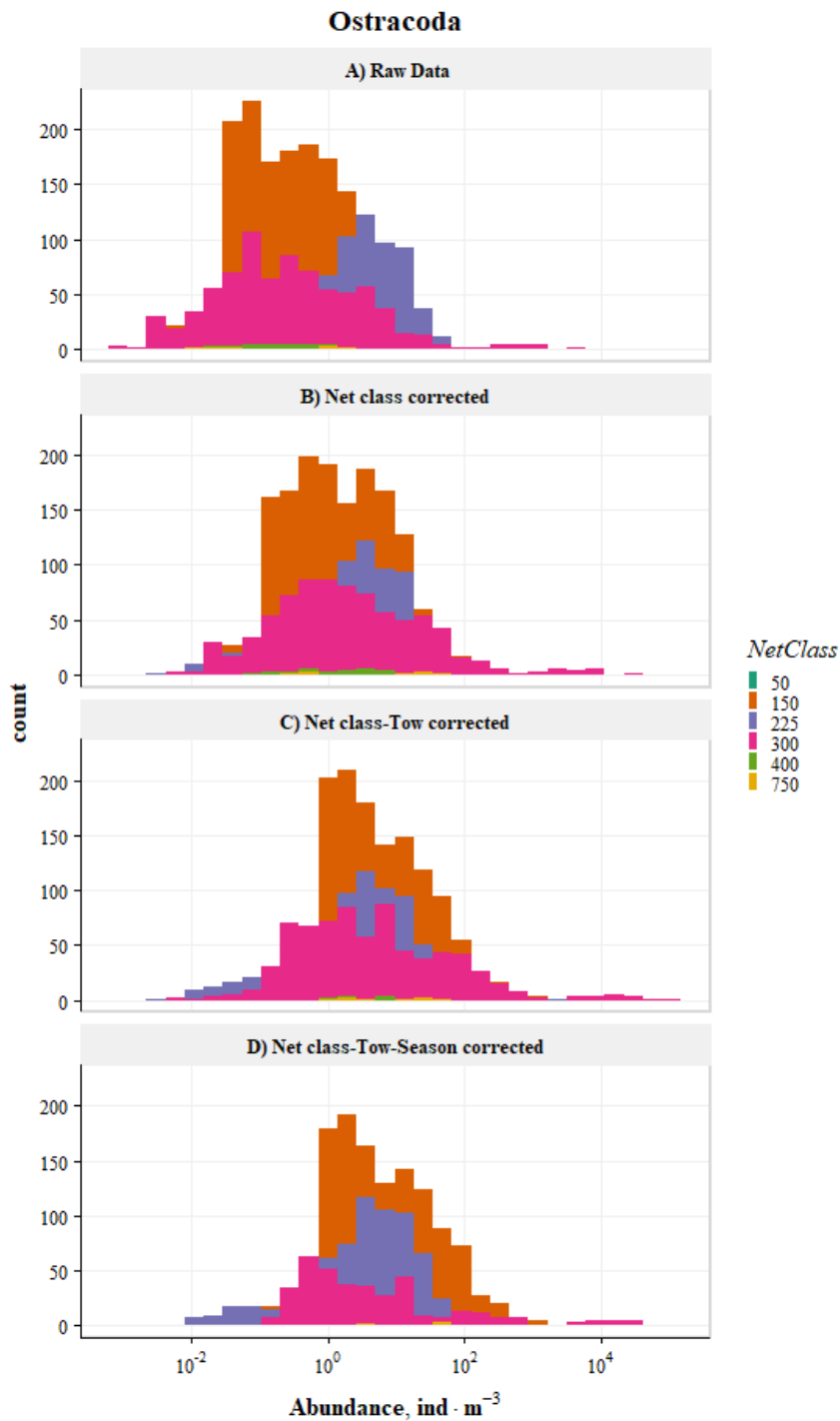


Figure B3: Example of calibration procedure for class Ostracoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

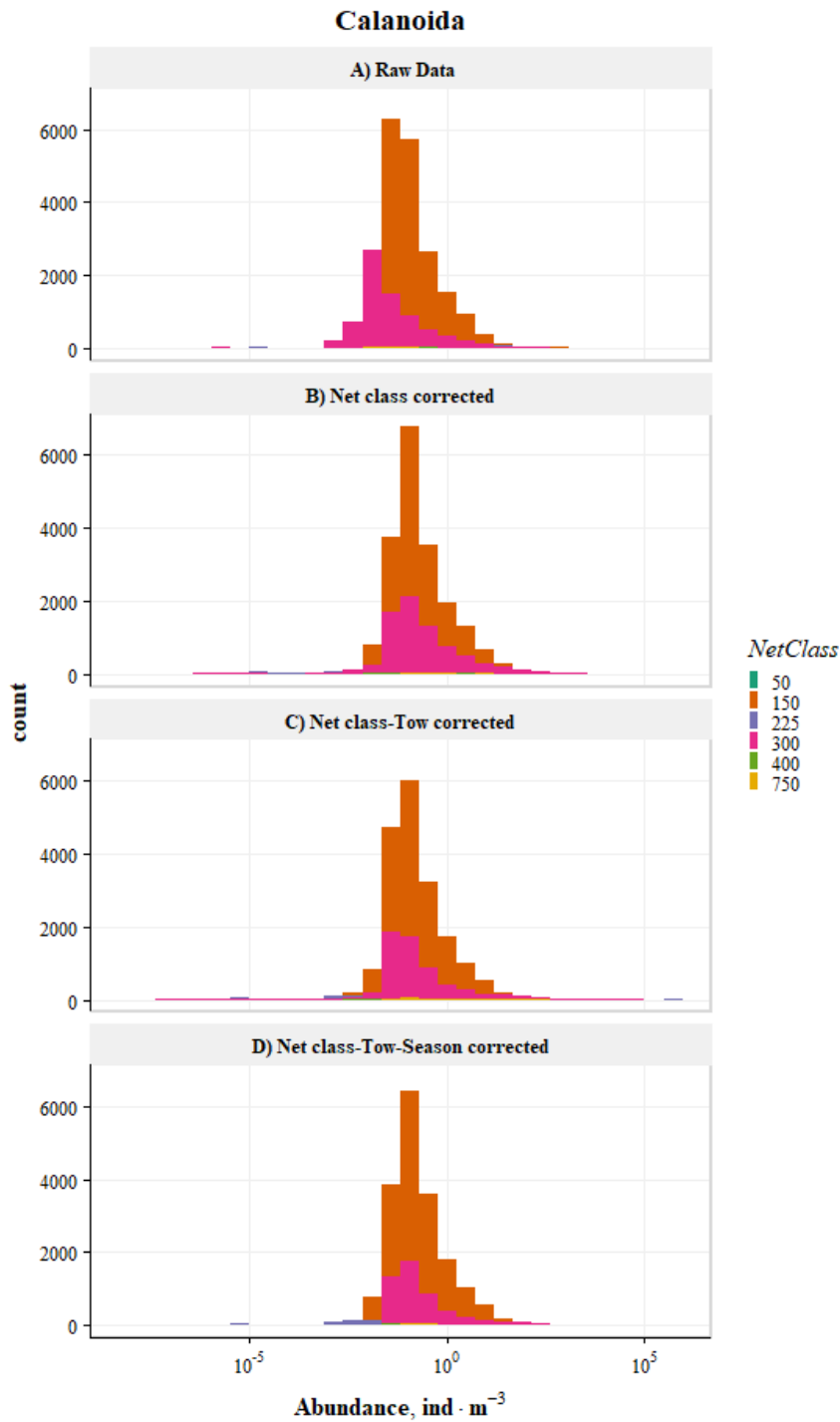


Figure B4: Example of calibration procedure for order Calanoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the log<sub>10</sub> transformation on *x* axis.

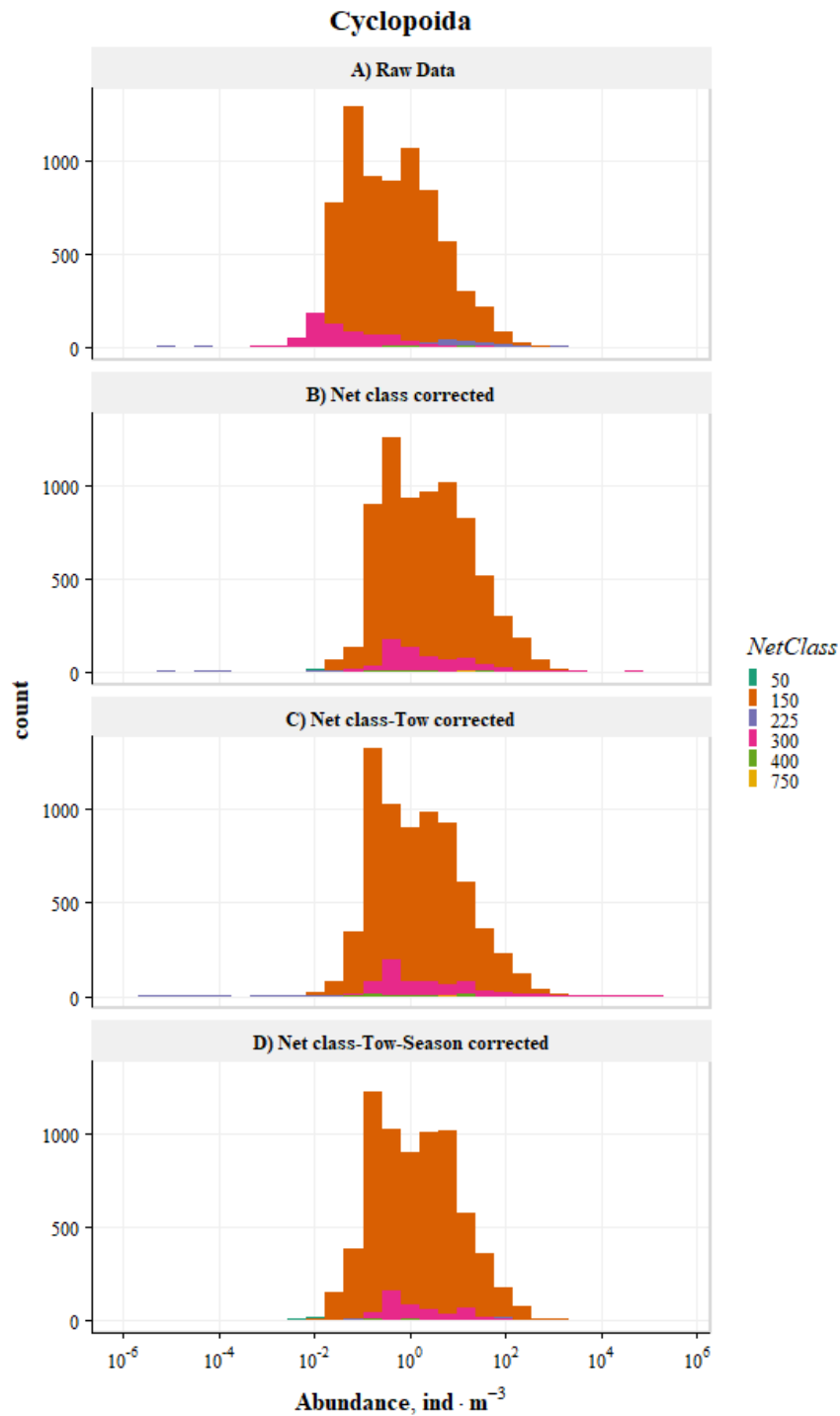


Figure B5: Example of calibration procedure for order Cyclopoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

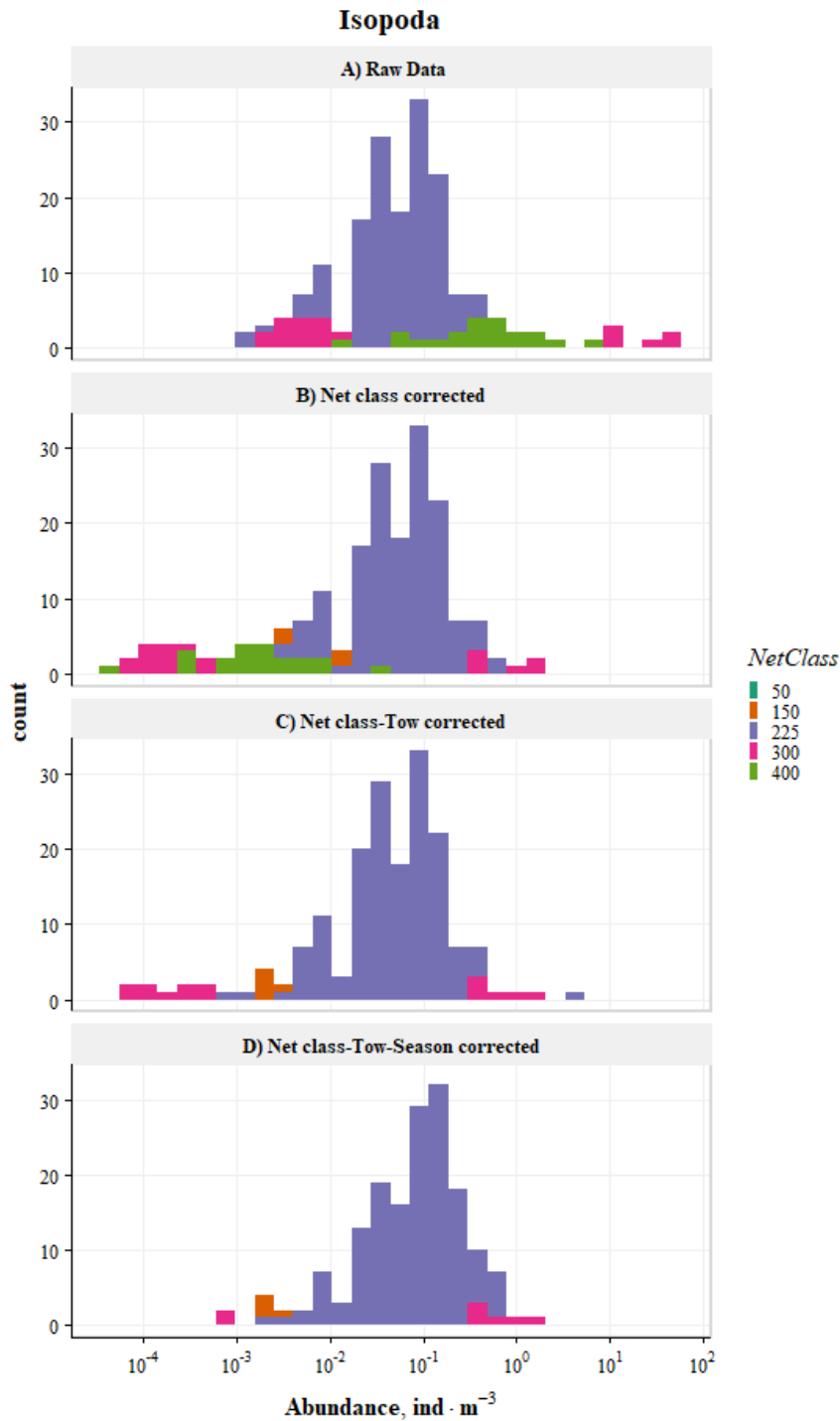


Figure B6: Example of calibration procedure for order Isopoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

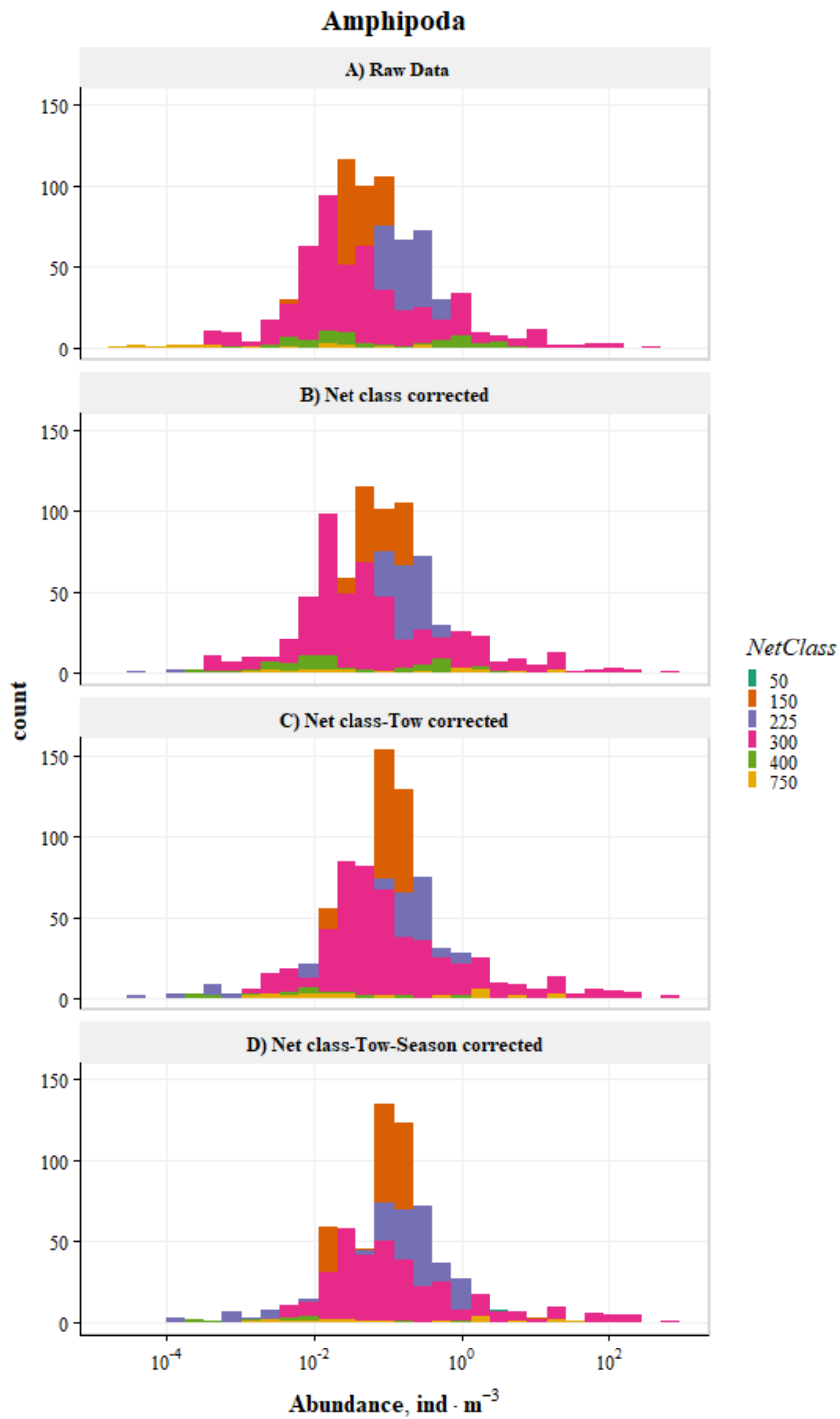


Figure B7: Example of calibration procedure for order Amphipoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

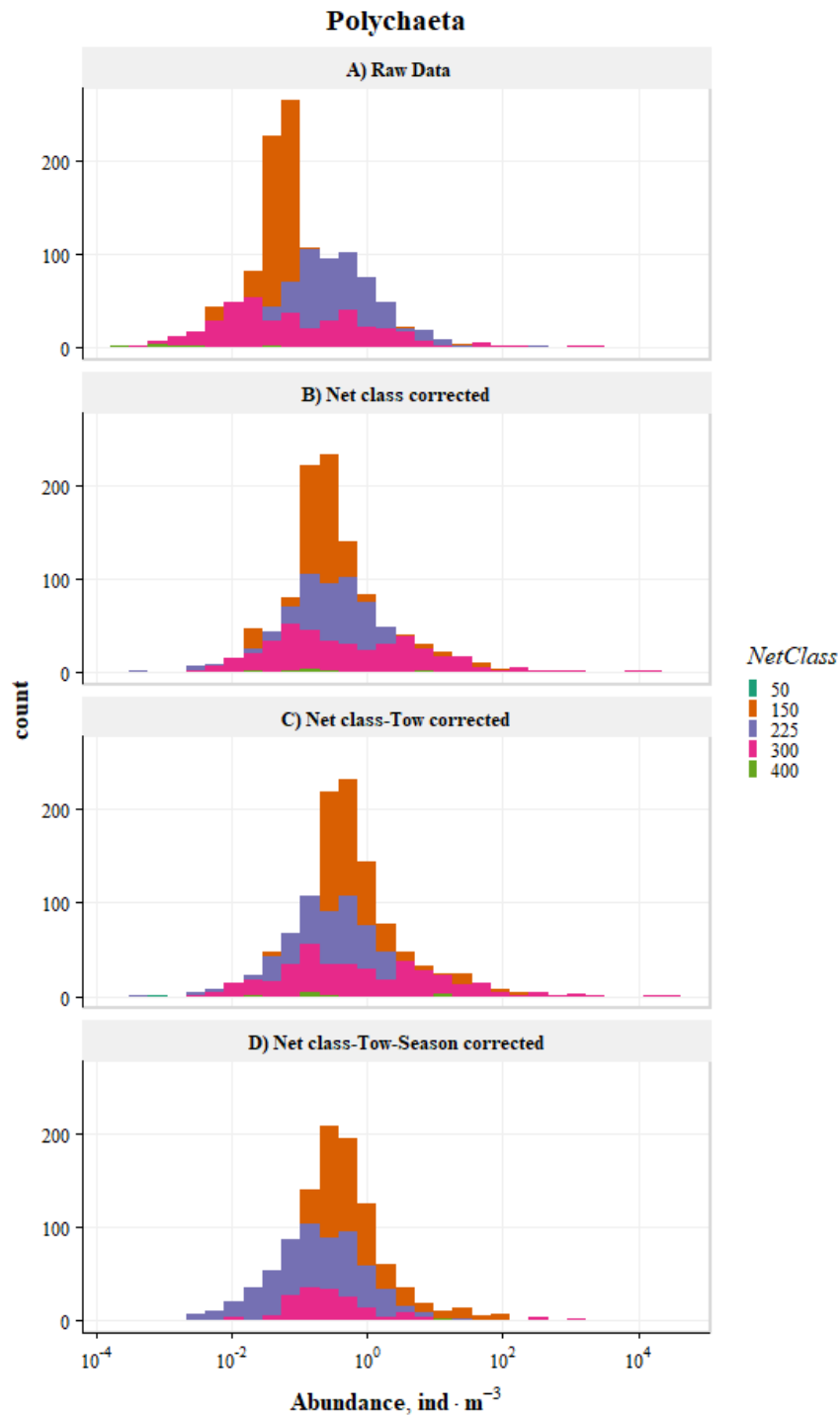


Figure B8: Example of calibration procedure for class Polychaeta colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the log<sub>10</sub> transformation on *x* axis.

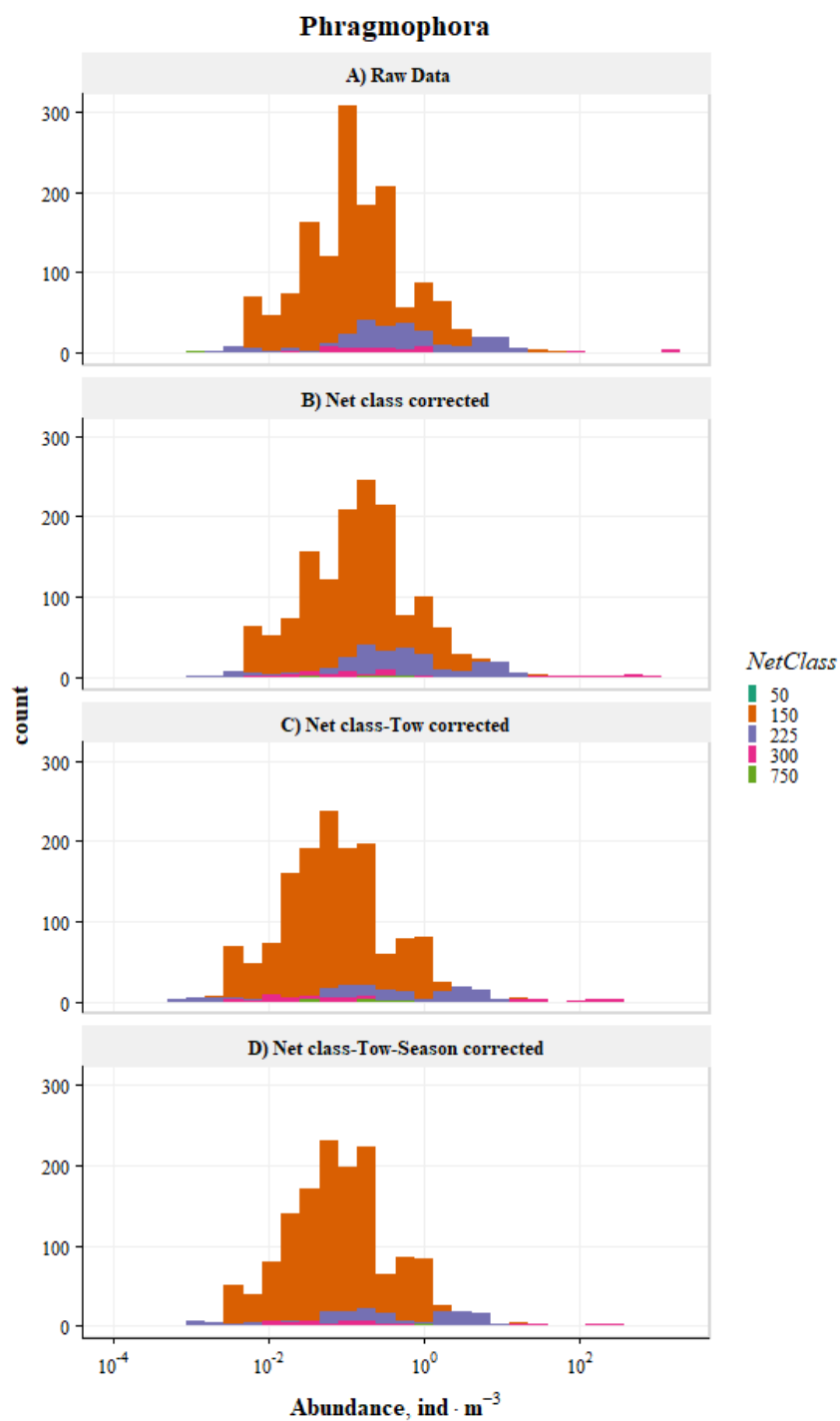


Figure B9: Example of calibration procedure for order Phragmophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.



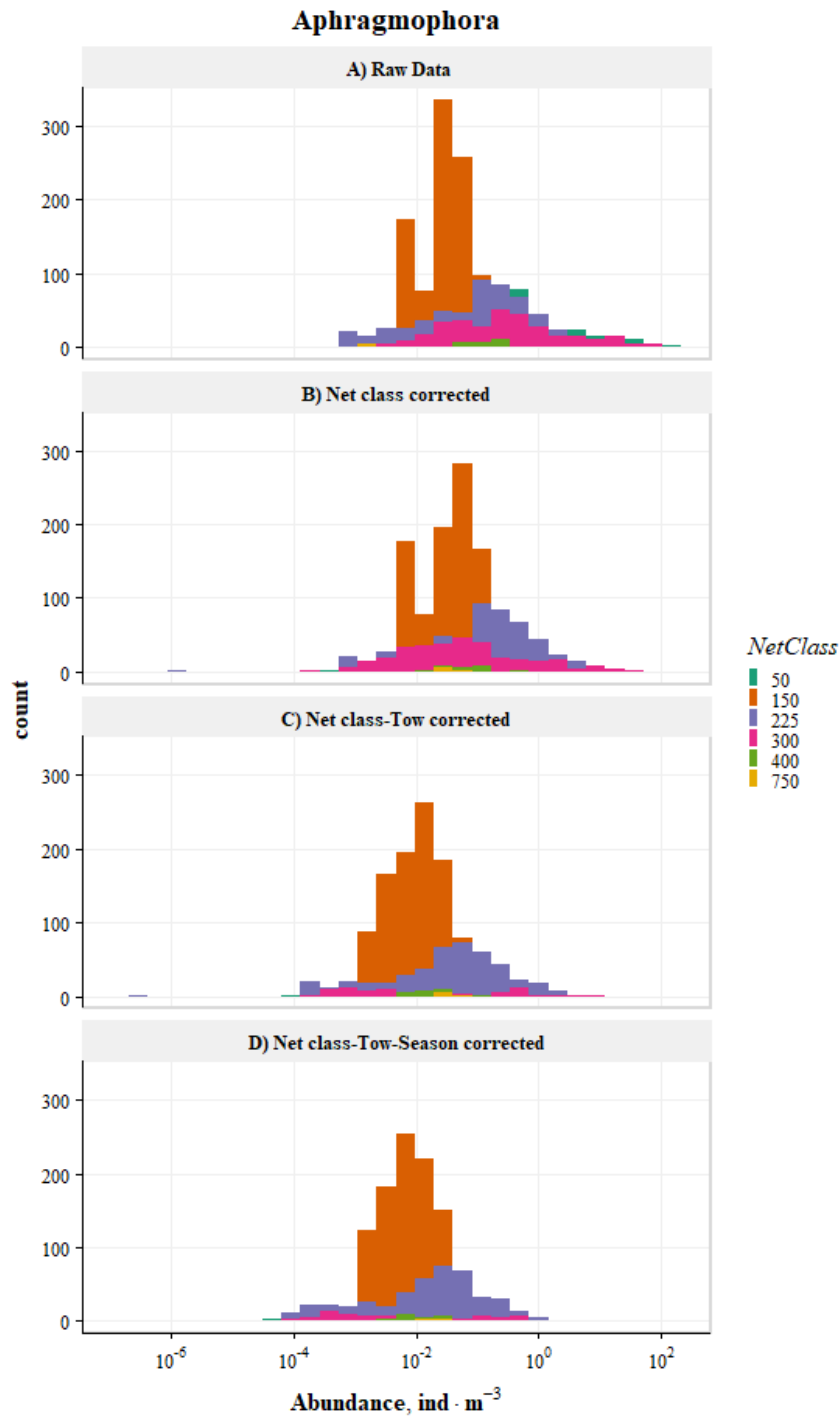


Figure B10: Example of calibration procedure for order Aphragmophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

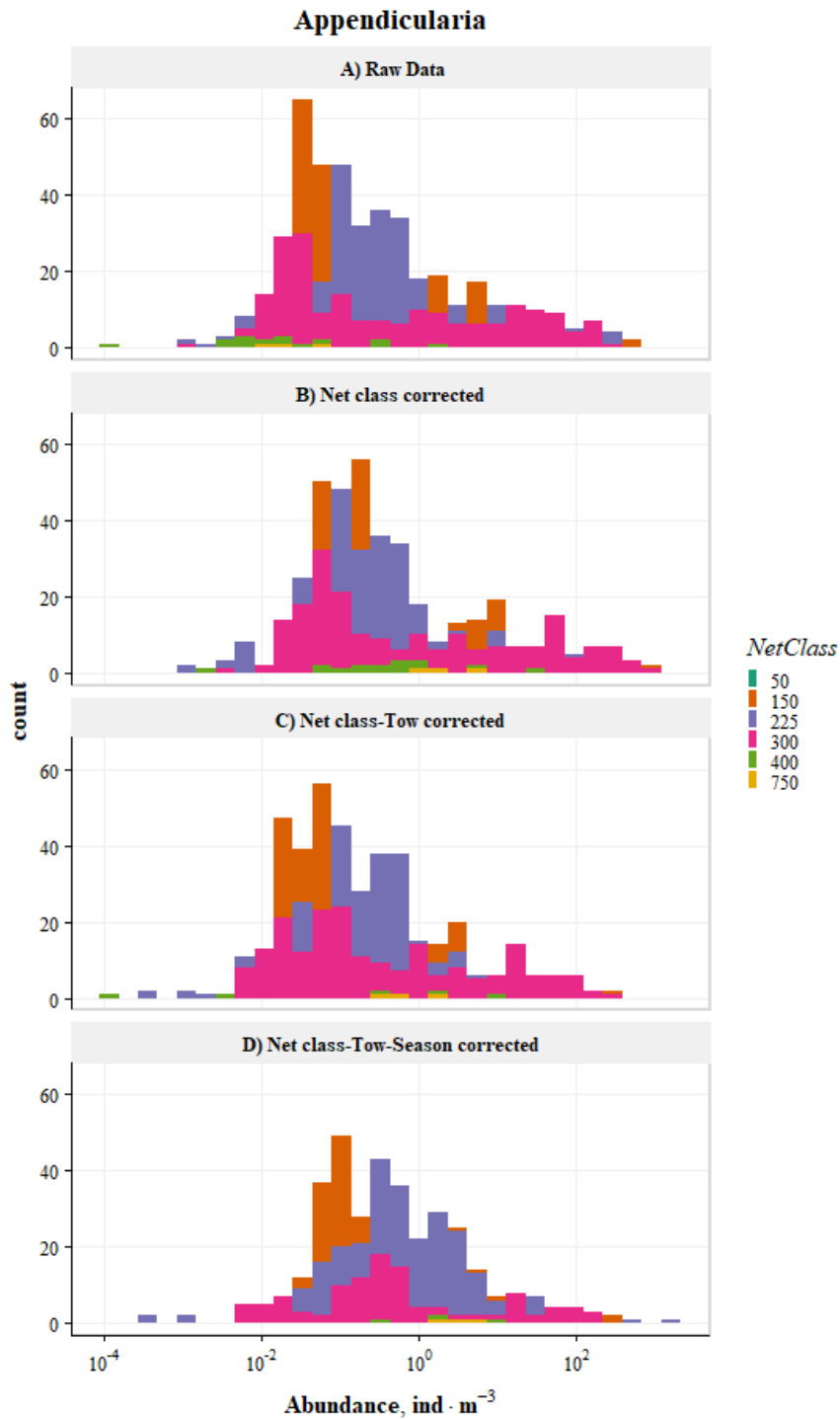


Figure B11: Example of calibration procedure for class Appendicularia colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

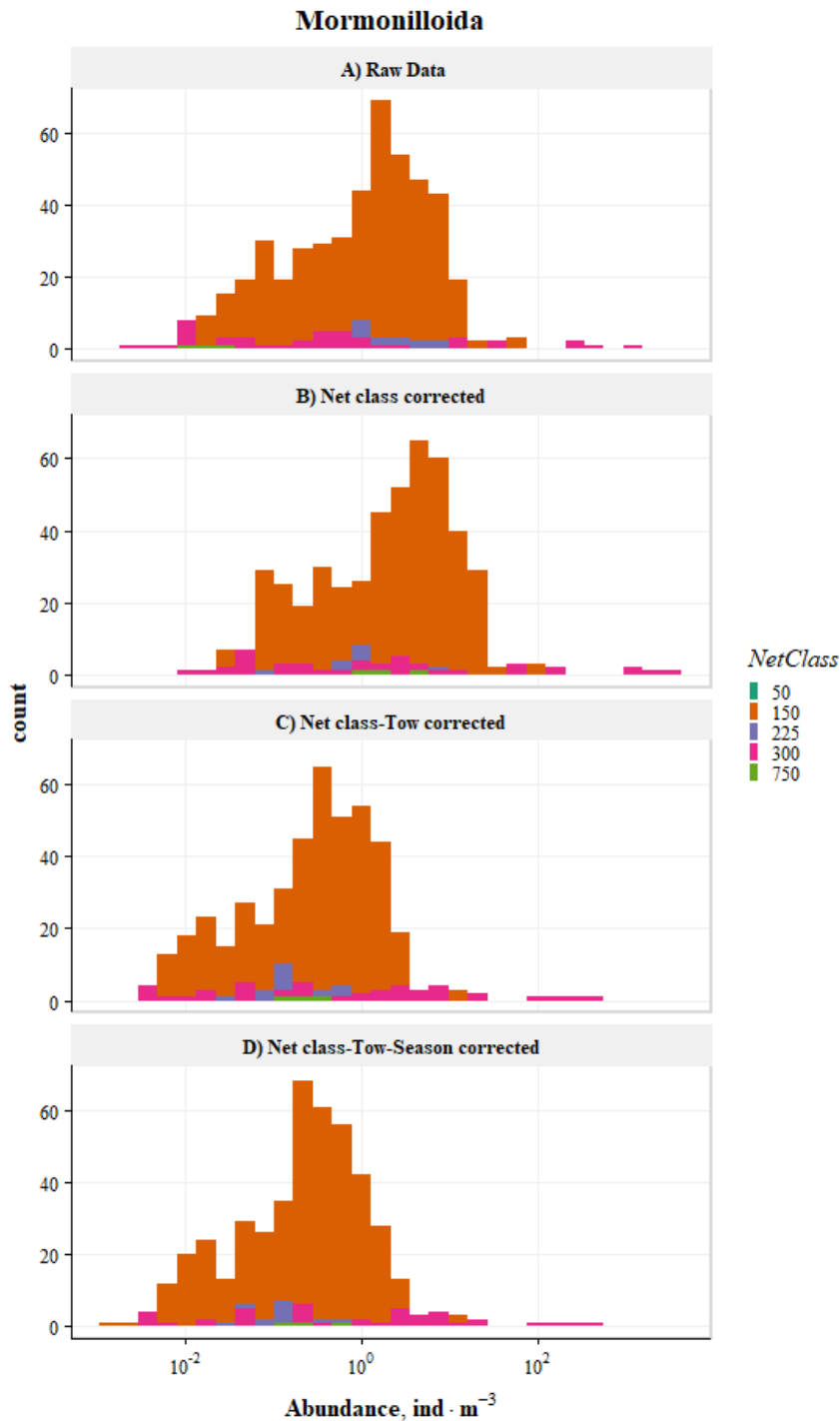


Figure B12: Example of calibration procedure for order Mormonilloida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

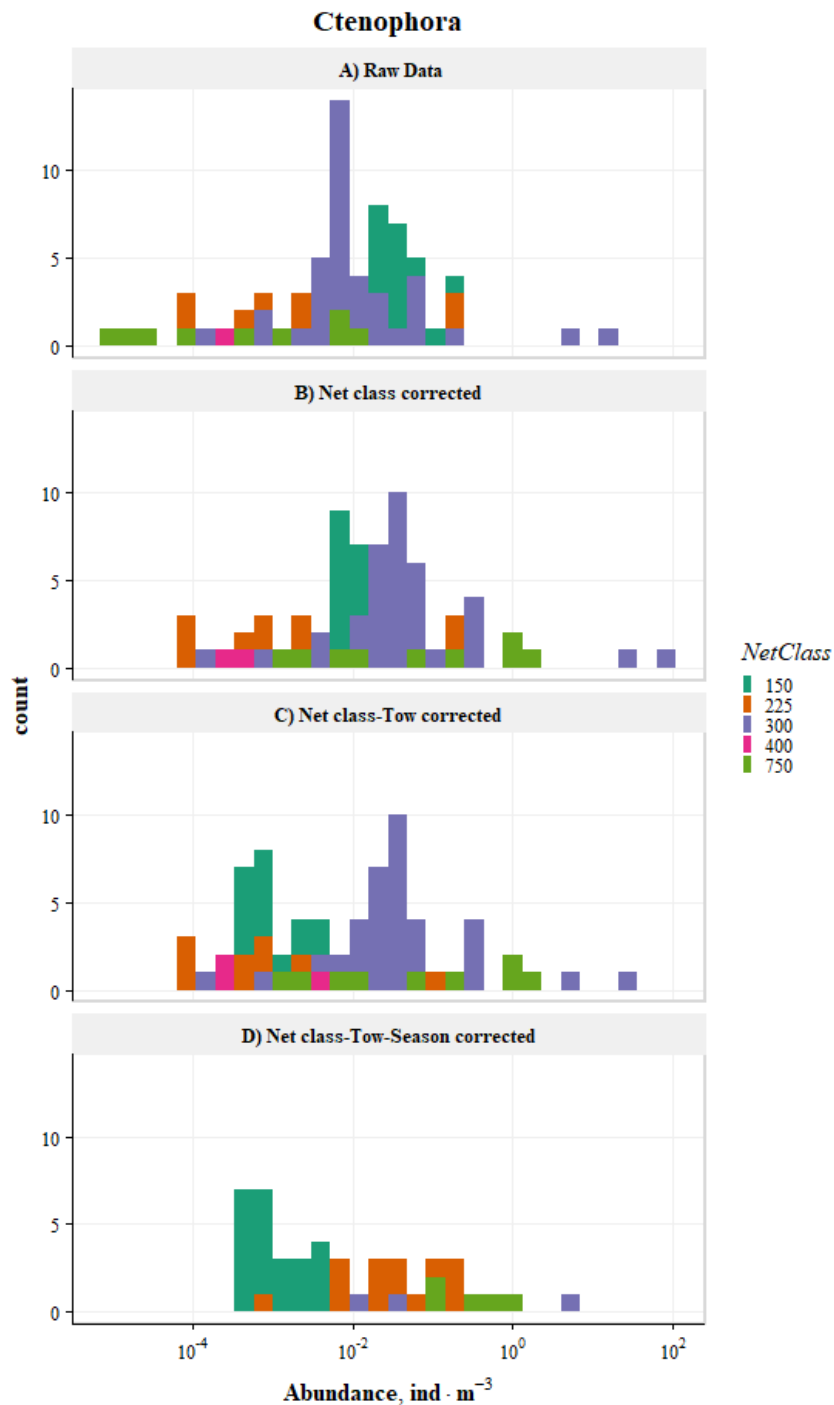


Figure B13: Example of calibration procedure for phylum Ctenophora colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

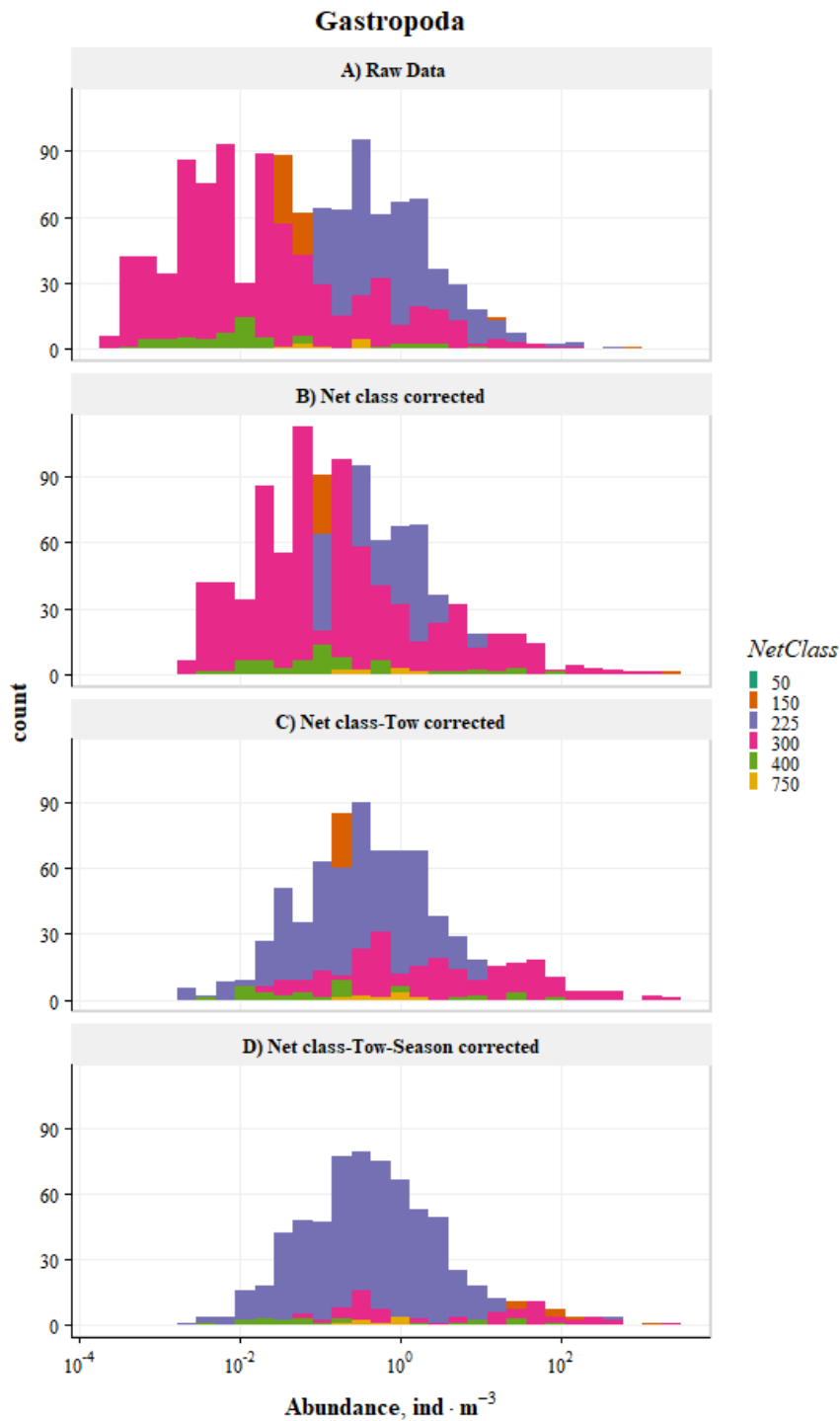


Figure B14: Example of calibration procedure for class Gastropoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the log<sub>10</sub> transformation on *x* axis.

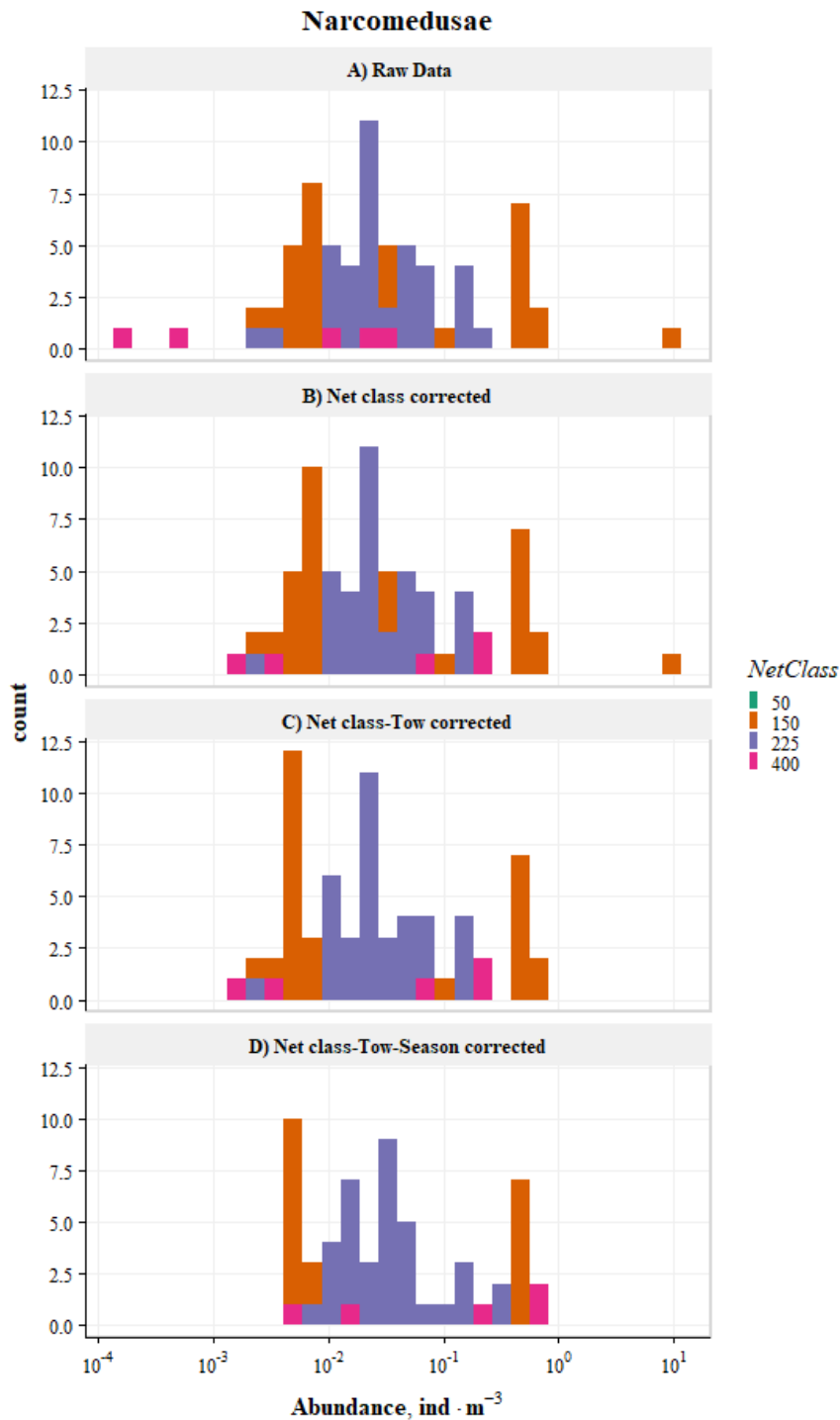


Figure B15: Example of calibration procedure for order Narcomedusae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

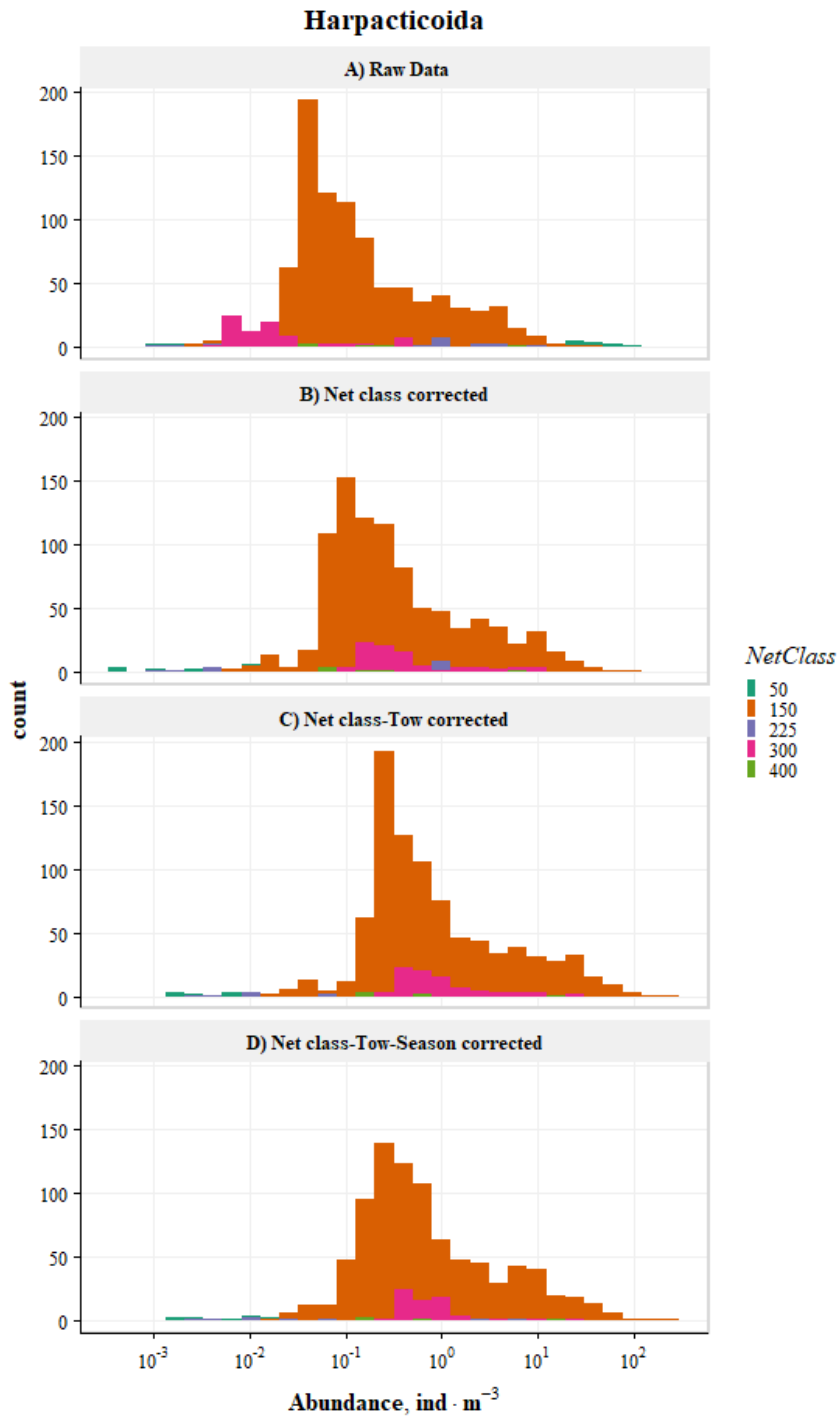


Figure B16: Example of calibration procedure for order Harpacticoida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

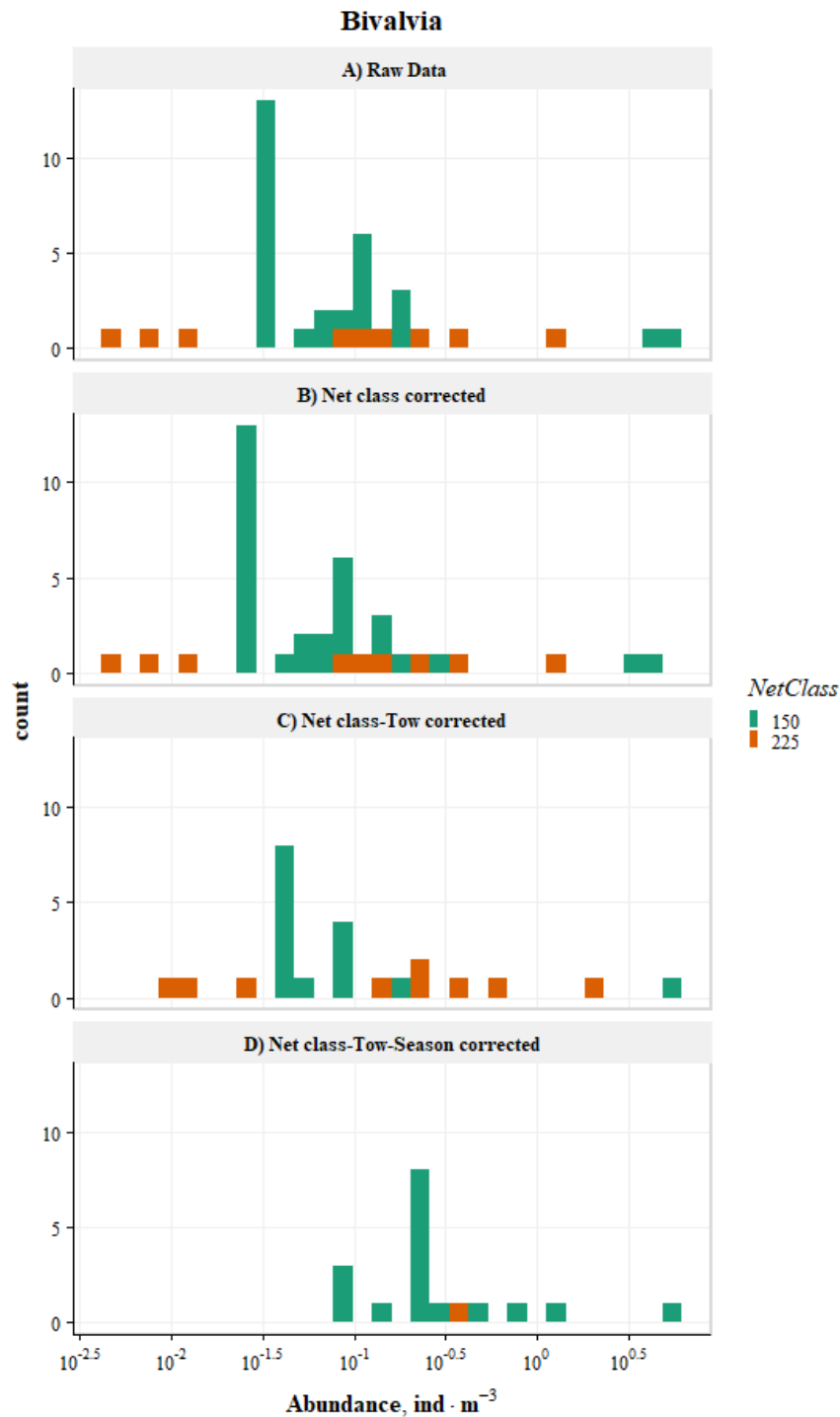


Figure B17: Example of calibration procedure for class *Bivalvia* colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.



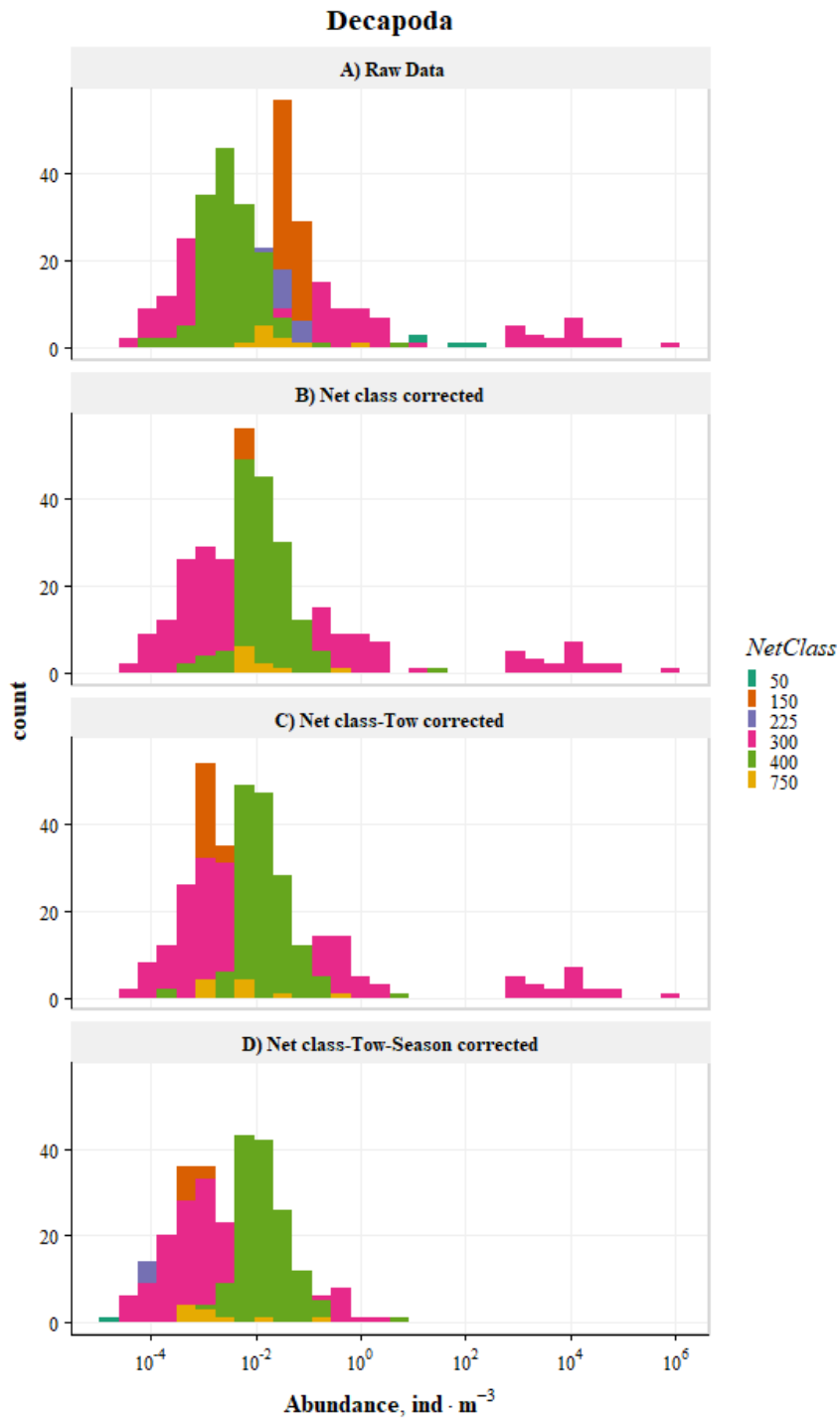


Figure B18: Example of calibration procedure for order Decapoda colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

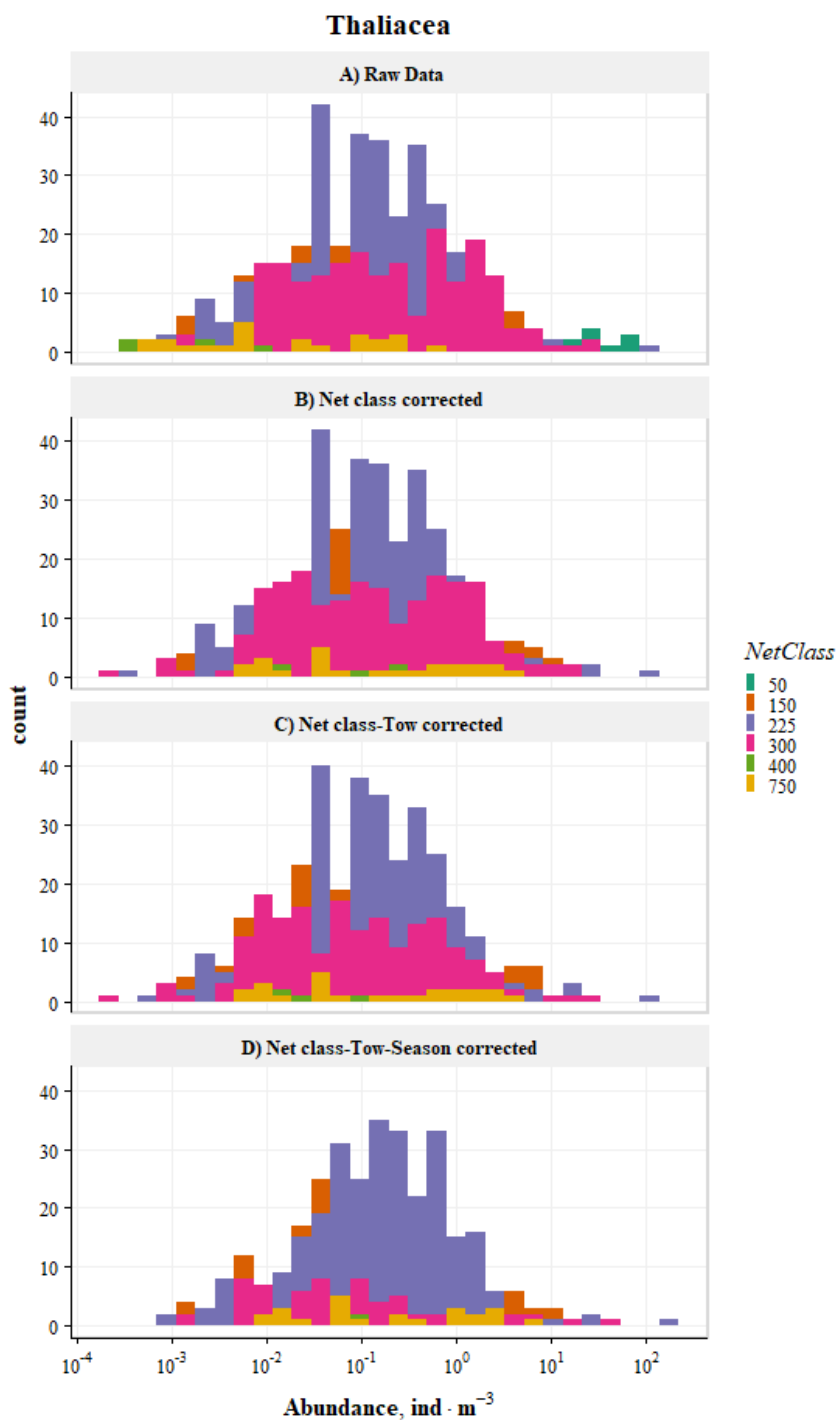


Figure B19: Example of calibration procedure for class Thaliacea colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

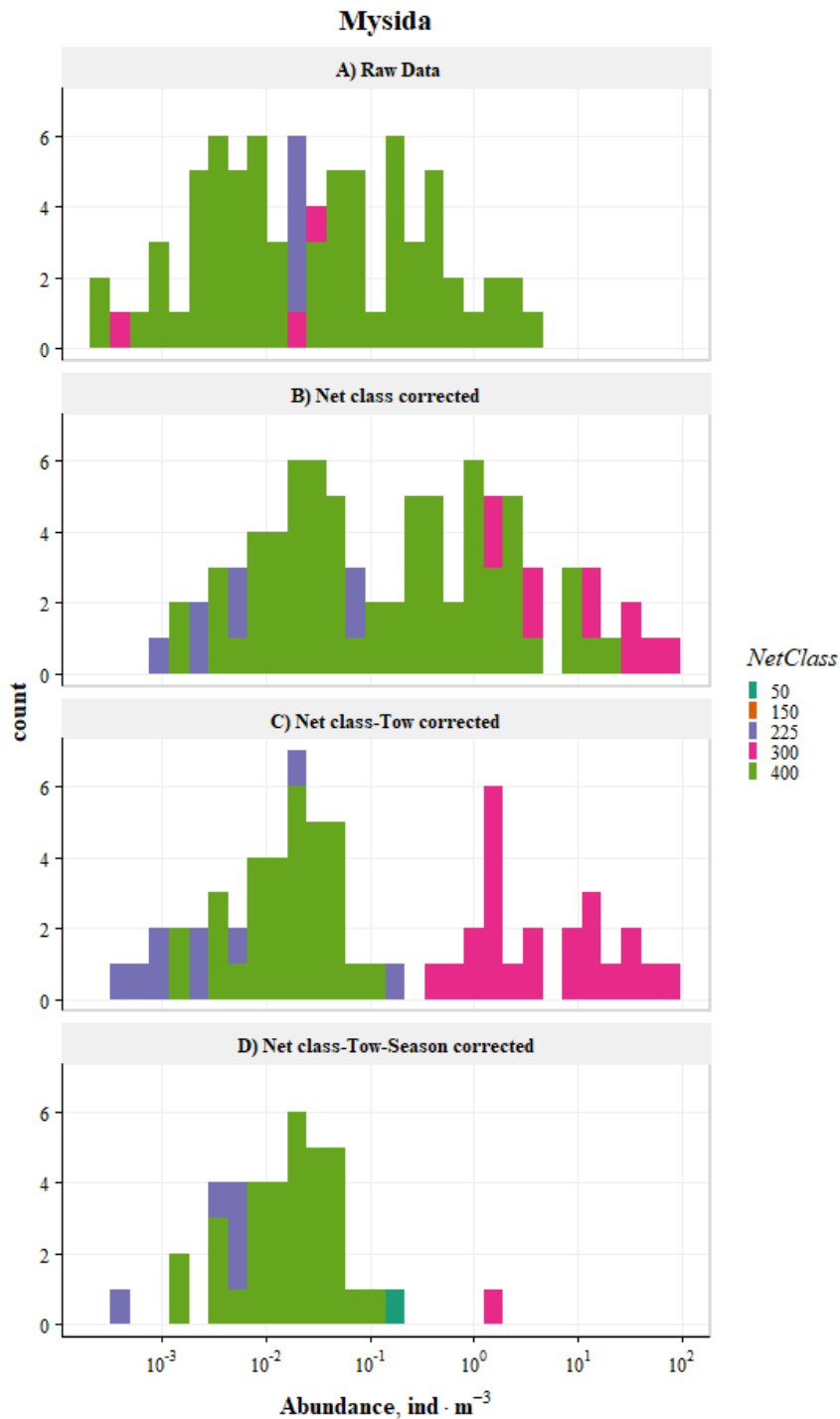


Figure B20: Example of calibration procedure for order Mysida colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

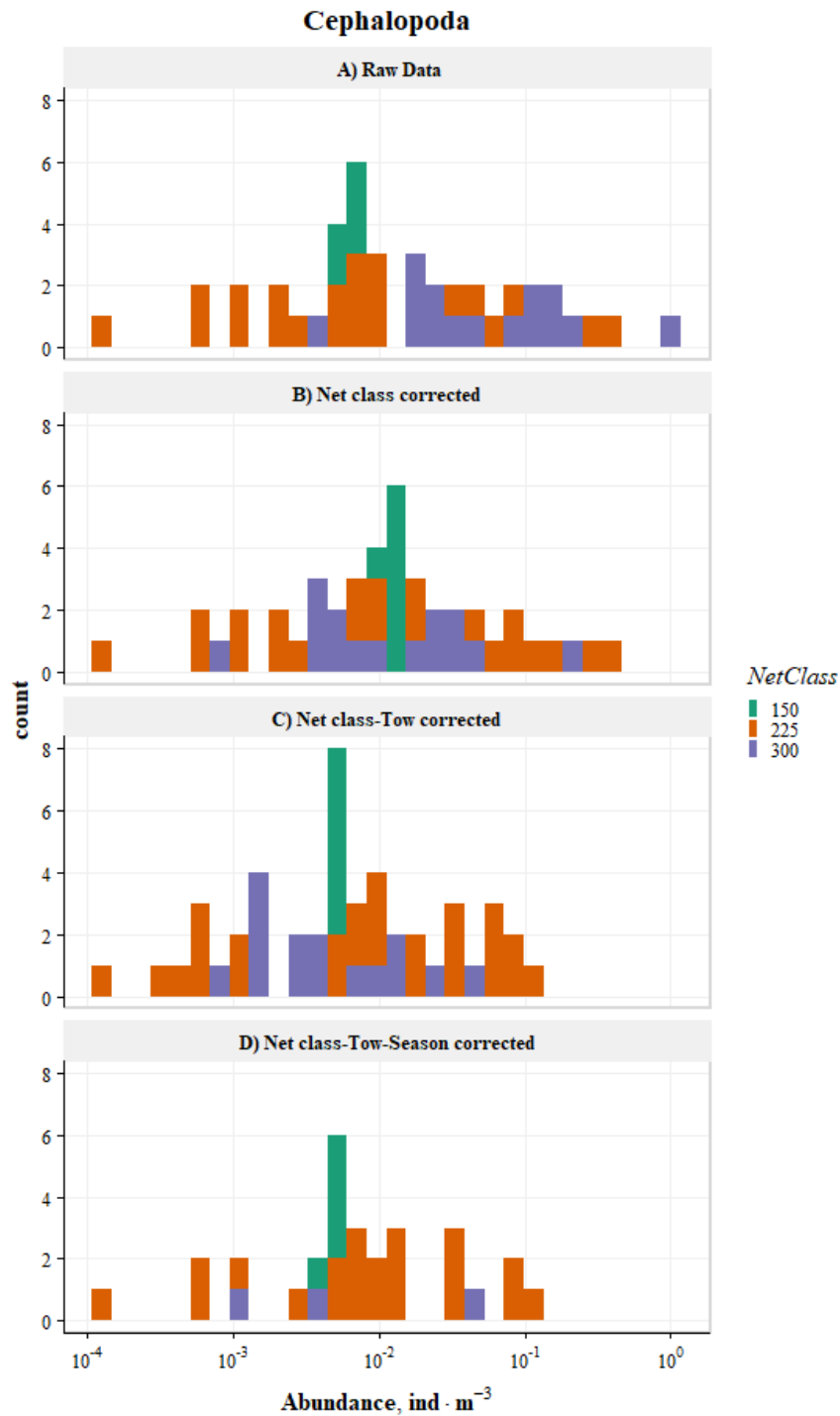


Figure B21: Example of calibration procedure for class Cephalopoda colored by different mesh size (NetClass of 50, 150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

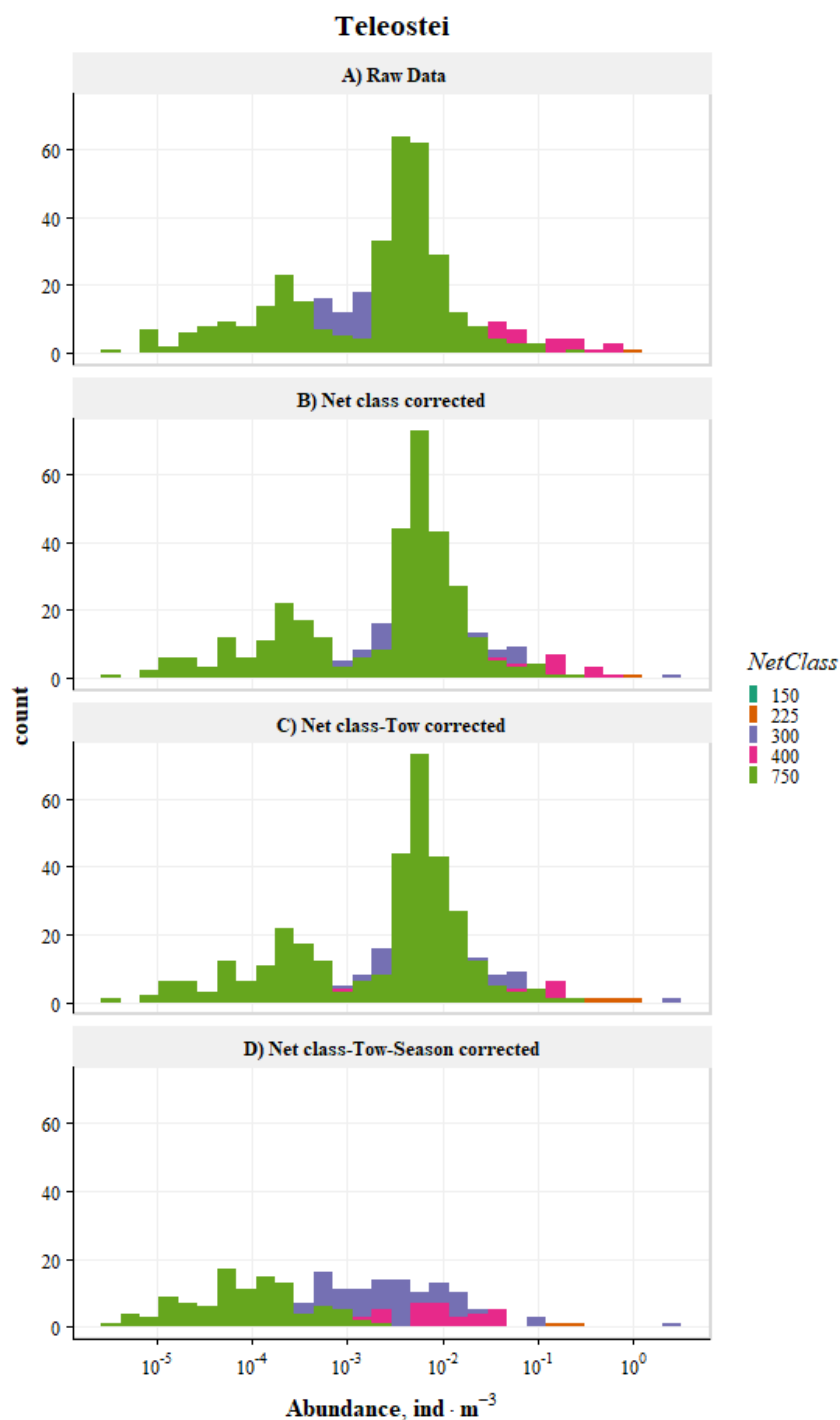


Figure B22: Example of calibration procedure for class Teleostei colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

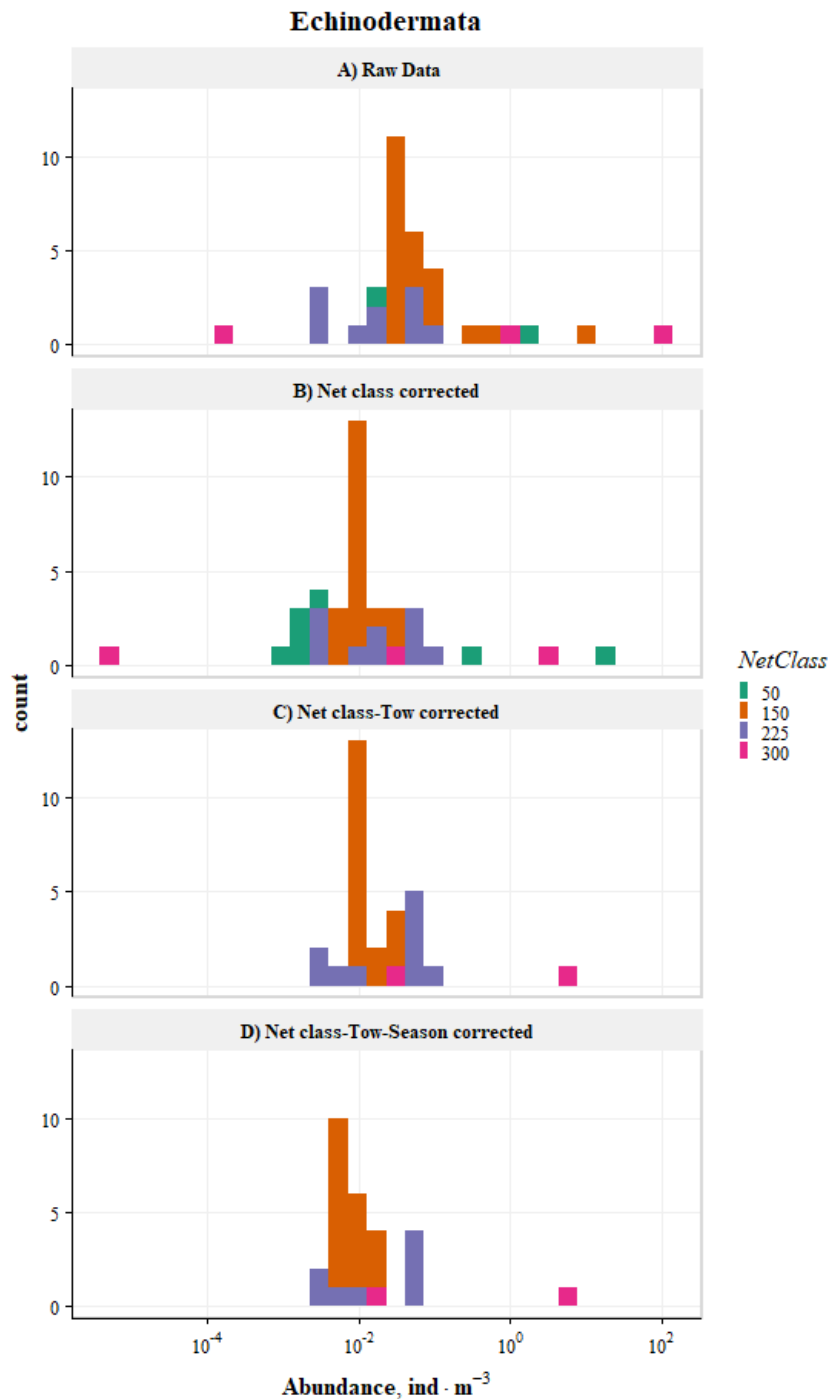


Figure B23: Example of calibration procedure for phylum Echinodermata colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750  $\mu\text{m}$ ). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the  $\log_{10}$  transformation on  $x$  axis.

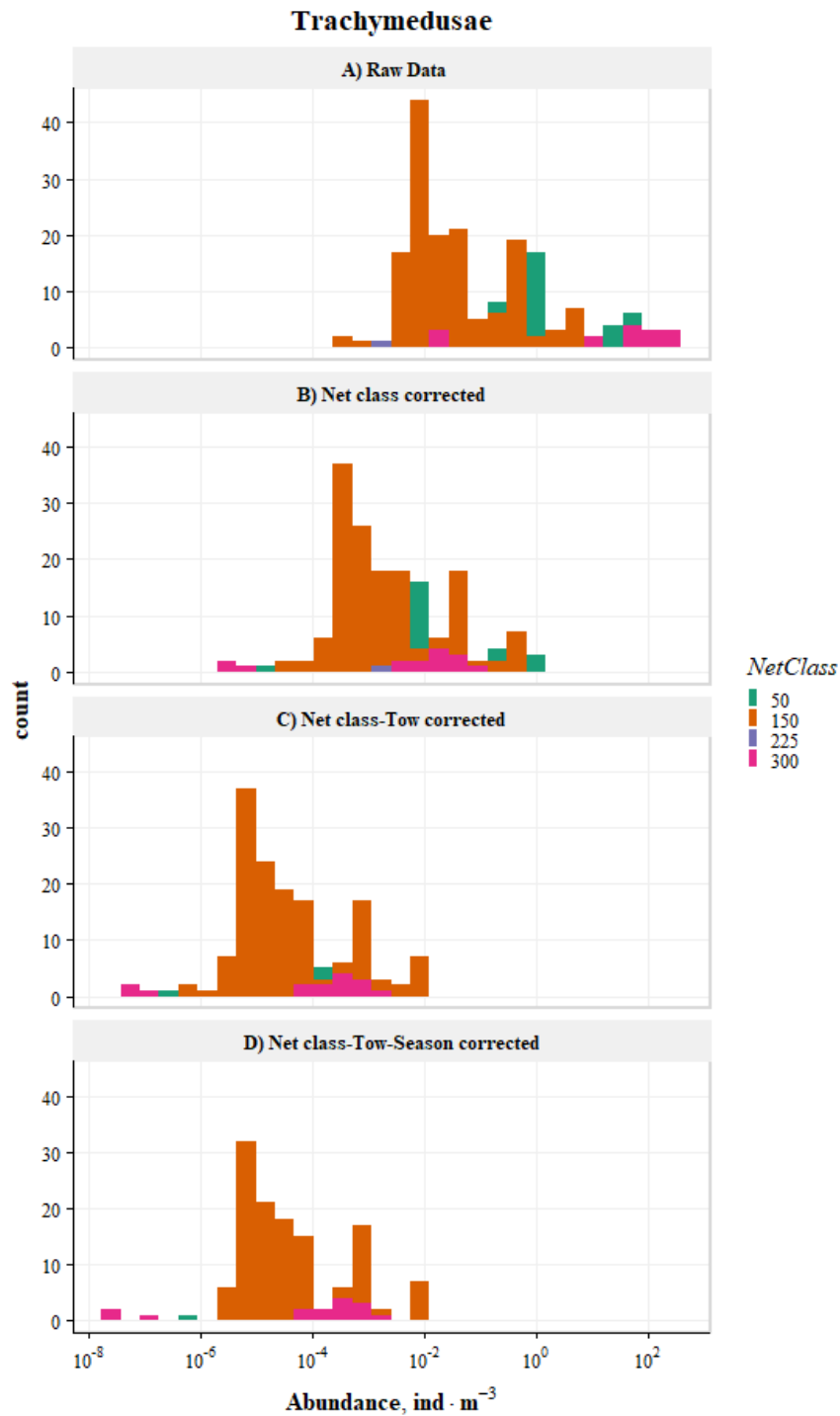


Figure B24: Example of calibration procedure for order Trachymedusae colored by different mesh size (NetClass of 50,150, 225, 300, 400 and 750 µm). A) raw densities, B) densities after correction for different net class and latitudinal class; C) abundance values from step B corrected for the tow type; D) abundance values from step C corrected for the season. Note the log<sub>10</sub> transformation on *x* axis.

## Appendix C - Supplementary information for Section 2.4

### A Comparison of Statistical Methods for the Analysis of Environmental Drivers of Zooplankton Abundance in the Southern Ocean

#### Introduction

Playing a crucial role as secondary producers in marine ecosystems, zooplankton serve as a bridge between primary productivity and higher trophic levels (Turner, 2004). They also significantly contribute to the marine biological pump and foster vertical carbon transport (Steinberg & Landry, 2017). As a result, alterations in zooplankton community composition can have far-reaching impacts on aquatic ecosystems' biogeochemical cycles and energy pathways (Mitra *et al.*, 2014). Moreover, zooplankton are valuable environmental change indicators, responding to variations brought about by climate change or pollution (Hays, 2003). Therefore, it is important to be able to robustly estimate zooplankton populations.

The accuracy of population estimates can be significantly impacted by the variability in zooplankton sampling methods, seasons, day/night measurements, and years (Doubek *et al.*, 2020; Mack *et al.*, 2012). Varying levels of selectivity can be caused by different sampling methods, resulting in certain species or life stages of zooplankton being collected more frequently than others. One important aspect of variability that is often overlooked is the impact of different sampling depths. Zooplankton populations can be stratified based on depth, with certain species or life stages found at different depths. This can lead to significant differences in sampling rates depending on the depth of the sampling gear. However, this variability is rarely accounted for in zooplankton assessments, which can lead to inaccurate estimates of population size and abundance.

Ecologists typically use linear models with the log-transformation of a skewed response variable, even when the response variable of interest is noticeably non-normal (Boldina & Beninger, 2016; Hernández-León *et al.*, 2020; Vereshchaka *et al.*, 2016). This method serves to induce an assumption of normality, thereby facilitating more robust statistical analyses (Menge *et al.*, 2018). To our knowledge, the variability of population estimates introduced by different sampling bins has never been accounted for. As a result, the precision of the abundance estimates derived from different bin sizes is not uniform. For instance, though they have identical mean depths, each of two samples with bin sizes of 400-600 m (200 m range) and 300-700 m (400 m range) surveys a



distinct volume of water. It is anticipated that abundance estimates from smaller bin sizes will exhibit higher precision, because they survey a narrower depth range and can therefore attribute the abundance of species to a more specific range.

There are several statistical methods that can be used to address sources of variability in zooplankton sampling. The emergence of generalized linear models (GLMs) has expanded our statistical toolbox, affording the capacity to craft a model for the mean of a response variable not constrained by normal distribution. Linear Regression Models and GLMs use regression analysis to explore the effects of one or more explanatory variables on a response variable of interest (Oppel *et al.*, 2012). This contrasts with models that merely allow for a normally distributed transformed response. This development eradicates the necessity for transformation assumptions, thereby providing a more flexible and direct approach for analyzing data that adheres to non-normal distribution patterns.

This study utilized existing data on zooplankton densities and related environmental factors to perform a comparative analysis of statistical modelling approaches. The study focused on two specific methodological objectives. The first objective aimed to assess whether a linear model with a log-transformed response variable (referred to as Gaussian) exhibits comparable performance to GLMs. The second objective aimed to investigate the impact of incorporating a dispersion parameter on model parameters and/or performance. The study also includes reflections on prevalent practices in statistical inference among ecologists, even if they may not be optimal. Ultimately, the findings of this research inform recommendations for future studies in this field.

### **Data Processing**

The analysis focused on the density of Calanoid species (Order Calanoida) in the Southern Ocean. The dataset contains five most abundant species in the area: *Rhincalanus gigas*, *Metridia lucens*, *Calanus simillimus*, *Pleuromamma robusta*, and *Ctenocalanus vanus*. To mitigate potential confounding effects across the years, the data from the single year (1987) were used. Water temperature, salinity, and dissolved oxygen concentration were used to model species bioclimatic envelope. Environmental data were obtained from the World Ocean Atlas 2018 (<https://www.ncei.noaa.gov/access/world-ocean-atlas-2018>). Average sampling depth was also included in the study as zooplankton abundance was shown to decline with depth across the oceans

(Puelles et al., 2019). All statistical analyses, data manipulations and visualizations were performed using the R Statistical Software (R Core Team, 2022).

The covariates were median-centered prior to modelling to account for differential ranges and avoid fitting issues for the interaction terms. Overall, this study uses a focused approach to analyze the relationships between environmental variables, depth, and the abundance of zooplankton species in the Southern Ocean, ensuring data quality and avoiding potential confounding effects.

Table C1: Variable descriptions of the subset of Calanoida species in the Southern Ocean.

Variable	Description
Abundance bins	Abundance of the organism at a certain depth ( $\text{ind}\cdot\text{m}^{-3}$ )
Average Depth	Depth bin (m): difference between max and min depth of sampling
temperature	Average depth of sampling (m)
salinity	Ocean Temperature ( $^{\circ}\text{C}$ )
oxygen	Salinity (ppm)
	Oxygen concentration ( $\text{ml}\cdot\text{l}^{-1}$ )

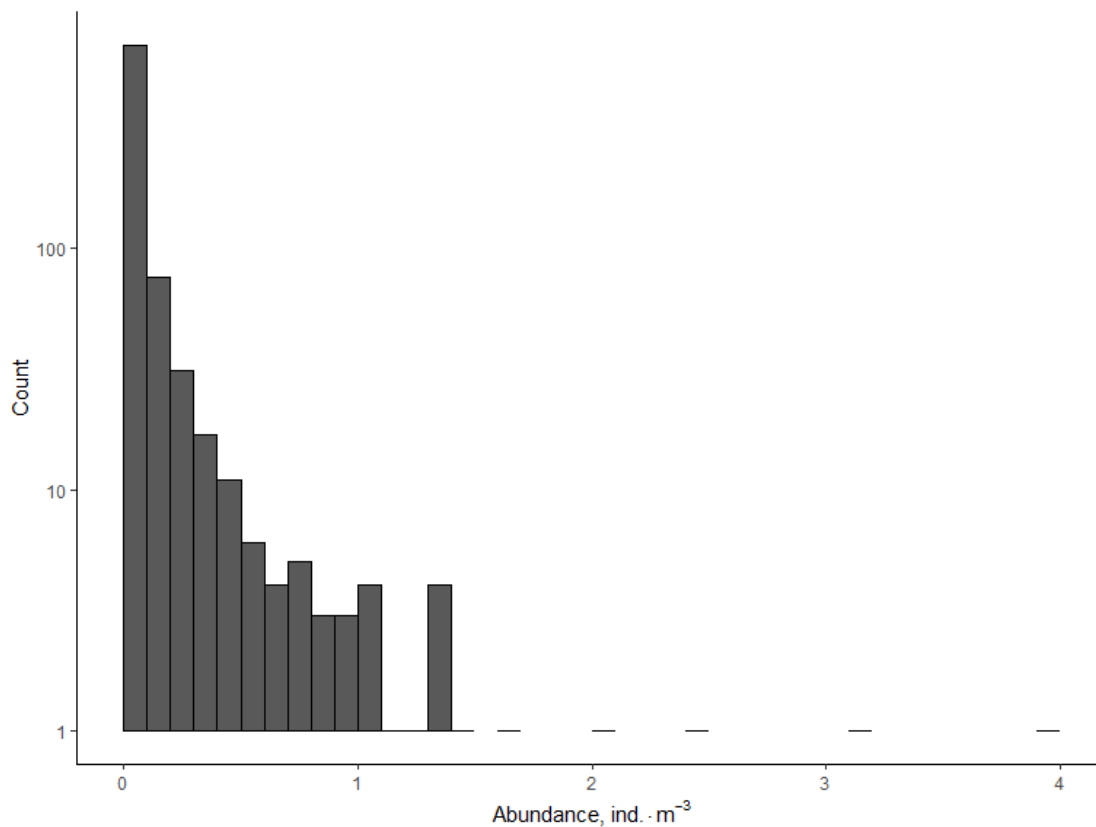


Figure C1: Histogram of zooplankton densities (in  $\text{ind}\cdot\text{m}^{-3}$ ) in the dataset. Note the log

transformation on y-axis.

The abundance of zooplankton (in  $\text{ind}\cdot\text{m}^{-3}$ ) was found to exhibit a non-zero, positive, and right-skewed distribution (Figure C1). Most of the abundance values ranged between 0 and 1  $\text{ind}\cdot\text{m}^{-3}$ , with a few extreme values up to 4  $\text{ind}\cdot\text{m}^{-3}$ . To identify the covariates of interest, a pairwise comparison was conducted between all the covariates and the response variable (Figure C2). It was observed that salinity and oxygen showed strong correlations with several other variables (Figure C2). To avoid issues with multicollinearity, salinity and oxygen were not included in the model. Ultimately, the study focused on the variables of temperature and average depth as the covariates of interest.

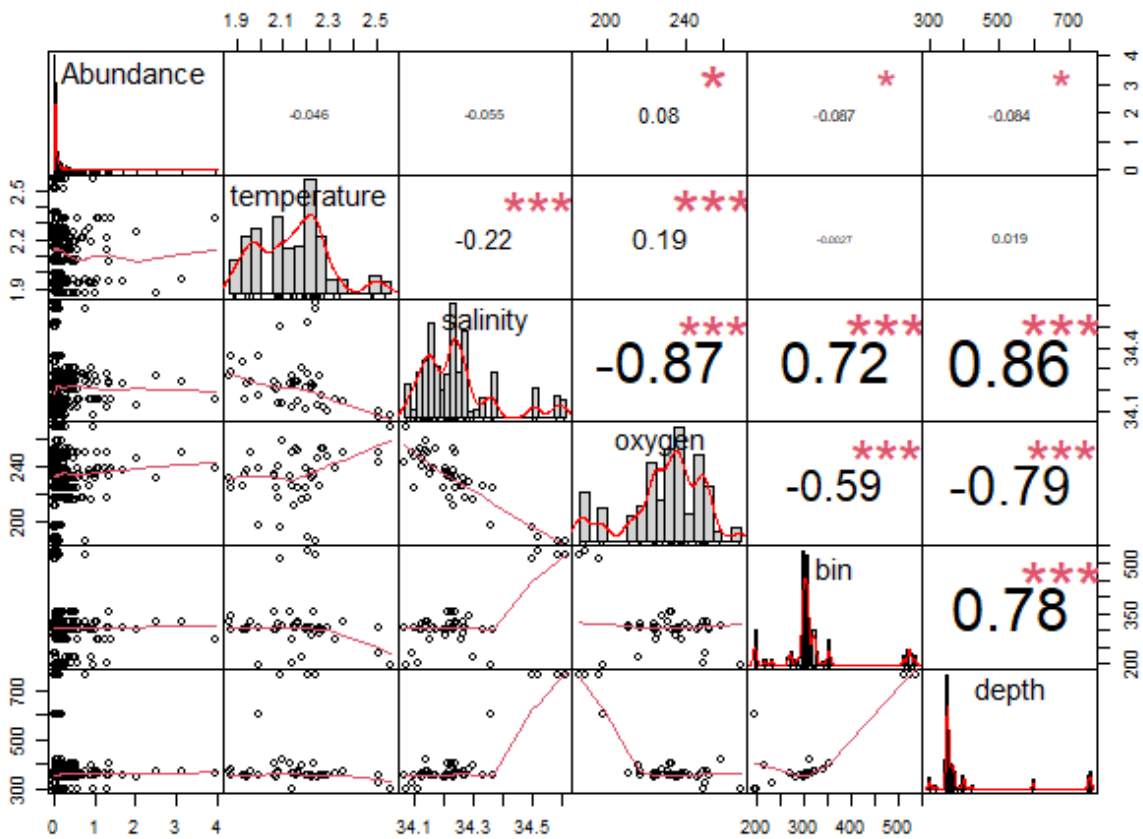


Figure C2: Pairs plot of all model variables (both response and predictors). Asterisks indicate level of significance (\*\*\*) 0.001, \*\* 0.01, \* 0.05) of correlations.

## Model Formulation and Selection

Linear models were formulated with the abundance of zooplankton as the response variable, and depth and temperature as fixed effects. Additionally, an interaction term between temperature and average depth was included. This interaction accounts for the observed pattern where the effect of temperature on abundance varies in relation to depth, a trend noticeably depicted in the exploratory pairs plot (Figure C2). The full model formulation can thus be written as follows:

$$y_i = \beta_0 + \beta_1 \cdot \text{temperature}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{temperature}_i \cdot \text{depth}_i + \epsilon_i \quad (1)$$

Here,  $y_i$  denotes the abundance (the response variable),  $\beta_0$  represents the intercept,  $\beta_1$  and  $\beta_2$  are coefficients corresponding to temperature and depth, respectively, and  $\beta_3$  is the coefficient for the interaction between these two variables. The term  $\epsilon_i$  represents the error term for each observation  $i$ .

In ecological studies, it is a widely accepted practice to handle such skewed distributions by applying a Gaussian distribution to a log-transformed response variable. The distribution of the response in the common ecological model is as follows:

$$\begin{aligned} \log(y_i) &\sim N(\mu_i, \sigma^2) \\ E[\log(y_i)] &= \beta_0 + \beta_1 \cdot \text{temperature}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{temperature}_i \cdot \text{depth}_i \end{aligned} \quad (2)$$

where  $E$  is the expected value of  $\log(y_i)$ , and  $\sigma^2$  denotes the variance.

As previously discussed, the distribution of our response variable (abundance) is highly skewed, and restricted to positive, non-zero values. However, the advent of generalized linear models (hereupon referred to as GLMs) has enabled the ability to specify a model for the mean of a non-normally distributed response, rather than the mean of a transformed response assumed to be normally distributed. In the comparative investigation between Generalized Linear Models (GLMs) and the transformation-based approach, three error distributions are taken into consideration: Log-Normal, Gamma, and Inverse Gaussian.

$$\begin{aligned} y_i &\sim \text{Lognormal}(\mu_i, \sigma^2) \\ \log(E[y_i]) &= \beta_0 + \beta_1 \cdot \text{temperature}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{temperature}_i \cdot \text{depth}_i \end{aligned} \quad (3)$$

A member of the Tweedie family, the Gamma distribution is a flexible distribution that is commonly used to model continuous, positive, and skewed responses. A log-link is used here in

place of the canonical link (the inverse) for more interpretable comparisons. The following parameterization is considered:

$$y_i \sim \Gamma(\mu_i, \phi\mu_i^2)$$

$$\log(\mu_i) = \beta_0 + \beta_1 \cdot \text{tempearture}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{tempearture}_i \cdot \text{depth}_i \quad (4)$$

where the dispersion parameter  $\phi = 1/\alpha$ .

Lastly, consideration is given to an Inverse Gaussian distribution, another member of the Tweedie family, parameterized with the mean and variance as outlined below. A log-link is also employed in this instance to map the mean response to the linear predictor. Although both the Gamma and Inverse Gaussian are part of the Tweedie family, the Inverse Gaussian holds the capability to model more extreme positive values, along with supporting heavier tails, characteristics that could prove beneficial for this specific response.

$$y_i \sim \text{INV}(\mu_i, \lambda)$$

$$\log(\mu_i) = \beta_0 + \beta_1 \cdot \text{tempearture}_i + \beta_2 \cdot \text{depth}_i + \beta_3 \cdot \text{tempearture}_i \cdot \text{depth}_i \quad (5)$$

where the dispersion parameter  $\phi = 1/\alpha$ . The following parameterizations were used as part of the *gamlss* package, which was used to fit these models (Rigby & Stasinopoulos, 2005).

### Model selection

To compare fit between the models with different error distributions, the Generalized Akaike Information Criterion (hereafter GAIC) was used. The GAIC can be represented as follows:

$$GAIC(k) = -2\log(\text{Likelihood}) + (k \times df) \quad (6)$$

where  $df$  denotes the total model effective degrees of freedom and  $k$  is the fixed penalty term (Rigby & Stasinopoulos, 2005). The AIC is a special case of the GAIC when  $k = 2$ . Importantly, the *GAIC()* function implemented in *gamlss* does not exclude the constant terms when calculating the log-likelihood, which ensures that different models are comparable and is particularly essential for a dispersion model. To ensure that the GAIC of the linear model with the transformed response was comparable, the transformed log-likelihood multiplied by the Jacobian was used (Akaike, 1978), and the GAIC was re-calculated manually. To facilitate comparison across different distributions, the full models were compared.

### Dispersion Model

The depth ranges (bin sizes) used to measure abundance displayed substantial variation. The data contained 35 unique sampling depth intervals with bin size ranging between 195-537 m. In typical linear models, the mean parameter of the distribution is modelled, but it is also possible to fit covariates to the variance parameter to model its change with a variable in the dataset, rather than estimating it as a constant. This type of model is referred to as a dispersion model. The ideal scenario involves associating a smaller variance with a smaller bin size and a larger variance with a larger bin size. To achieve this, the normalized bin size (individual bin sizes divided by the sum) is fit as a covariate to the variance parameter using a log link. The dispersion model can be expressed for any distribution as follows:

$$y_i \sim N(\eta_{\mu i}, e^{(\eta_{\sigma i})}) \quad (7)$$

Here,  $\eta_{\mu i}$  refers to the linear predictor mapped to the mean response, and  $\eta_{\sigma i}$  refers to the linear predictor mapped to the variance parameter with the log-link and is expressed as an exponent to reflect that the variance only takes positive values. The linear predictor for the variance here can be expressed as:

$$\eta_{\sigma i} = \alpha_0 + \alpha_1 \text{bin size}_i \quad (8)$$

Where  $\alpha_0$  represents the intercept, and  $\alpha_1$  the coefficient estimated for the normalized bin size. Overall, we are interested in seeing how p-values, coefficients, and GAIC are affected by the use of different error distributions, different model specifications and the modelling the dispersion.

## Results

To visualize the impacts of the predictors on the response variable and obtain an initial picture of model fit, coefficient effects plots were constructed, with superimposed partial residuals. Notably, a positive effect of temperature on abundance was observed across all models (Figure C3). Likewise, all four models predicted that average depth has a negative relationship with abundance (Figure C4). As anticipated, the Gaussian and log-normal GLM yielded identical values for all coefficients (Table C2).

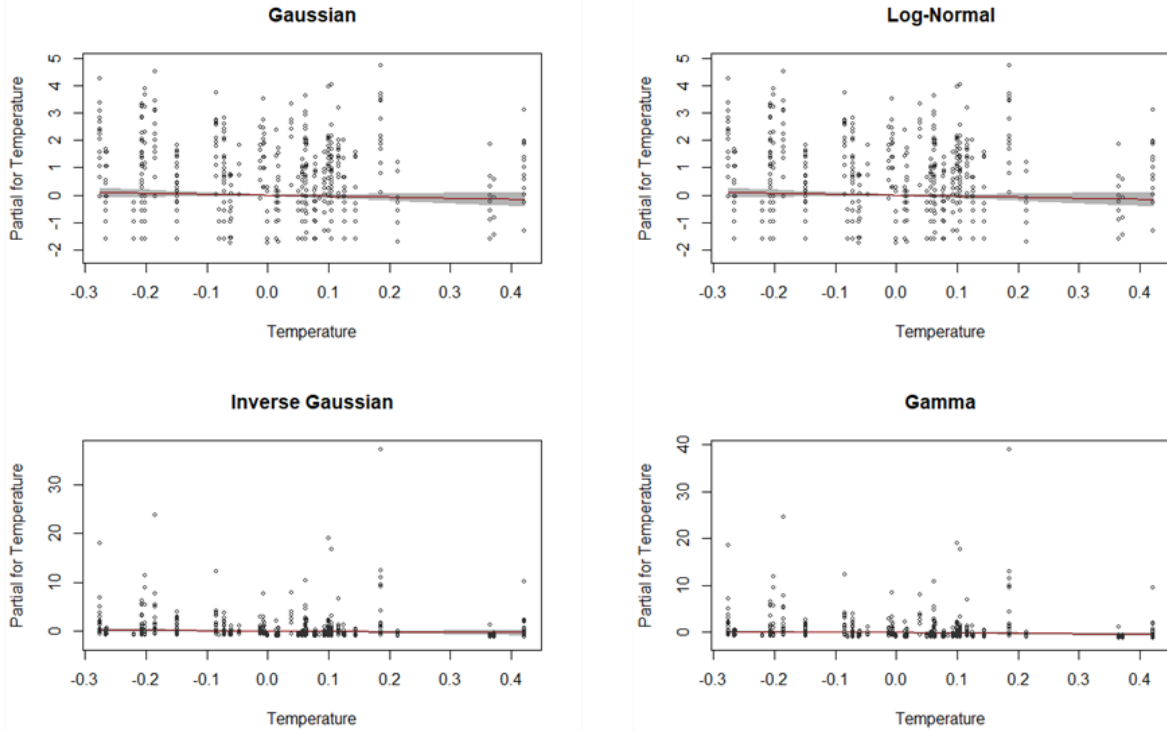


Figure C3: Temperature coefficient effects plots, with superimposed partial residuals. Residuals are shown as black dots. Grey shaded areas are 95% confidence intervals. It is to be noted that the axes of the four figures are on different scales.

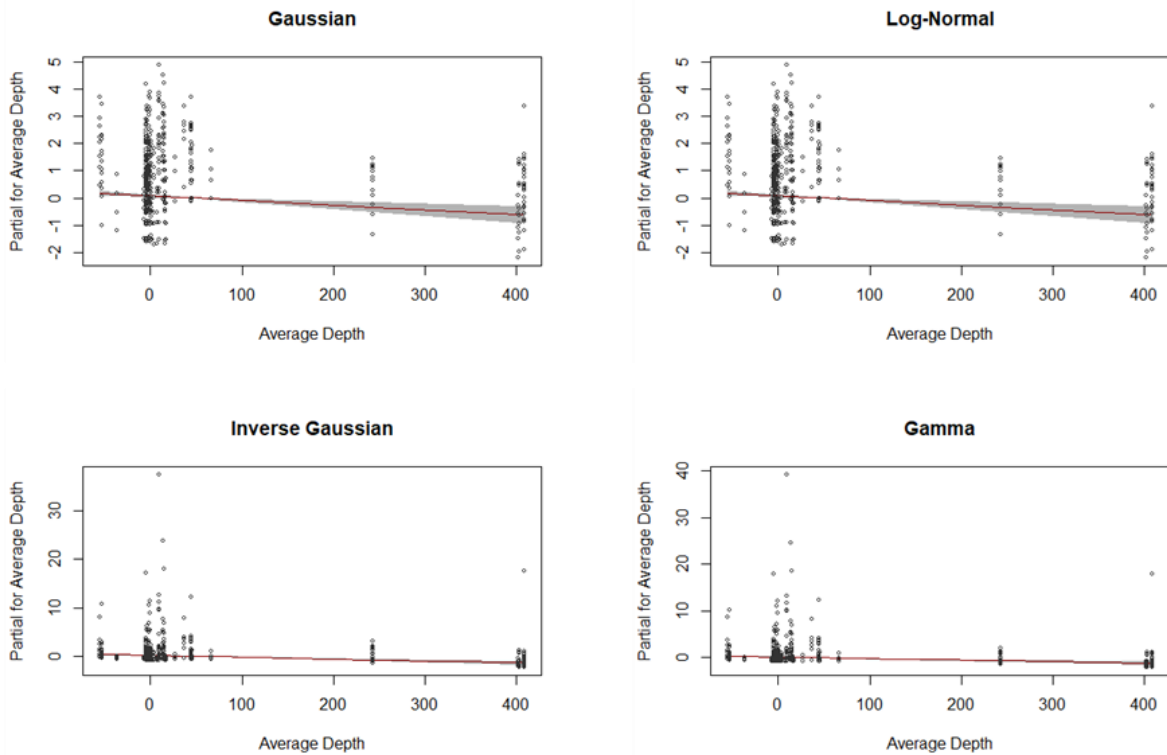


Figure C4: Average depth coefficient effects plots, with superimposed partial residuals. Grey shaded areas are 95% confidence intervals. It is to be noted that the axes of the four figures are on different scales.

Table C2: Comparison of Model Coefficients and Their Standard Errors Coefficients are presented on the log-link scale. Models compared include Gaussian, Log-Normal, Inverse Gaussian, and Gamma. Significance level notation: \*\*\*0.001, \*\*0.01, \*0.05.

Covariates	Gaussian	Log-normal	Inverse Gaussian	Gamma
<b>Mean Coefficients</b>				
Temperature	-0.3607 (0.2908)	-0.3607 (0.2908)	-0.7297 (0.4405)	-0.6853 (0.267)*
Depth	-0.0017 (4e-04)***	-0.0017(4e-04)***	-0.0038 (4e-04)***	-0.0032 (4e-04)***
Temperature :Depth	-0.0038 (0.004)	-0.0038 (0.004)	0.0118 (0.0038)**	0.0066 (0.004)
<b>Model Quality</b>				
GAIC	-2951.27	-2949.31	-3068.25	-2523.36

When the natural logarithm of a positive random variable exhibits a normal distribution, this variable follows a log-normal distribution. Therefore, one would expect identical results from a log-normal distribution in a generalized linear modelling framework and a Gaussian distribution with a log-transformed response variable in a linear modelling framework. As expected, log-normal distribution and Gaussian distribution had identical fits, with the same estimated coefficients, significant covariates, and GAIC (Table C2), With the different distributions, we also saw a change in the significance of the covariates, with temperature and the interaction being significant with the Gaussian and Log-normal distributions, temperature and depth with the Gamma and only depth with the Inverse Gaussian. In terms of GAIC, we saw that the Inverse Gaussian had the lowest score. Hence this model were chosen for future investigation.

Table C3: Inverse Gaussian GLM Model coefficients with and without the dispersion parameter ( $\pm$  standard error). Significance level notation: \*\*\*<0.001, \*\*<0.01, \*<0.05.

	GLM w/o dispersion model		GLM w dispersion model	
	Coefficient	p-value	Coefficient	p-value
Temperature	-0.7297 (0.4405)	0.09796	-0.5475 (0.4295)	0.2027
Depth	-0.0038 (4e-04)***	< 2e-16	-0.0035 (5e-04)***	1.51e-12
Temperature :Depth	0.0118 (0.0038)**	0.00229	0.0079 (0.0043)	0.0655
<b>Variance Coefficients</b>				
Bin Size			395.5786 (97.5937)***	5.51e-05
<b>Model Quality</b>				
GAIC	-3068.25		-3084.61	

To evaluate the impact of the dispersion parameter, it was incorporated into the Inverse Gaussian model (selected as the best model according to GAIC) to compare model fit. In the absence of the



dispersion parameter, depth and interaction between depth and temperature were significant predictors (Table C3). However, upon incorporating the dispersion parameter, it was found that only the depth variable exhibited statistical significance (Table C3). Moreover, the p-values were higher in magnitude for model with dispersion parameter, indicating decreased significance when the dispersion parameter was included. Additionally, a decrease in the GAIC was also observed when the dispersion parameter was added to the model.

## **Discussion**

In dealing with such skewed ecological data, it has become common practice to apply a log transformation to the data. The purpose of this transformation is to improve the linearity of the response to fit a normal distribution and to meet the homogeneity of the variance assumption (O'Hara & Kotze, 2010). However, use of GLM offers several advantages over using ordinary least square regression with log-transformed data. GLMs allow for modelling the response variable using a distribution appropriate for the data, which may better capture the inherent characteristics such as skewness or non-normality of the response variable. GLMs also allow for the use of various link functions to model the relationship between the predictors and the response variable. This flexibility allows for capturing a wide range of relationships, including non-linear associations. GLMs provide interpretable model coefficients that can be directly linked to the response variable's scale, even if the predictors are on different scales or transformations are applied. In linear regression with log-transformation, the interpretation of coefficients becomes less intuitive due to the log-transformed response variable. Lastly, GLMs can handle heteroscedasticity, where the variability of the response may change across different levels of the predictors. In linear regression with log-transformation, heteroscedasticity can still pose challenges, especially if the transformation does not fully address the issue.

In addition, GLMs offer the advantage of implicitly defining the relationship between the mean and the variance, which offers a flexible method of dealing with heterogeneity in variance. Additionally, these types of models allow the explicit specification of a model for the response that reflects an understanding of the underlying data-generation process and distribution (O'Hara & Kotze, 2010).

Our results showed that coefficients and significance levels changed between the models fit with different distributions (Table C2). As expected, the log-transformed Gaussian model and the Log-

Normal GLM did estimate the same coefficients. The results of this study showed that the Inverse Gaussian distribution had the lowest GAIC and was able to appropriately capture both the extremely high values of abundance between 0 and 1, as well as the high values present in the tails.

As previously mentioned, the estimates of abundance in the dataset have different precisions based on the depth range (i.e., bin width) sampled. In short, a higher level of precision is associated with abundances that are estimated from smaller bin sizes. To account for this, a dispersion model was used, where the variance parameter was modelled with bin size in addition to modelling the mean. Based on GAIC values, a model with the dispersion parameter performed better than the one with no dispersion (Table C3). Interestingly, while interaction term between depth and temperature was significant in the non-dispersion model, this was not the case when the dispersion parameter was added (Table C3). In comparison to the dispersion model, when the dispersion was not modelled, lower p-values were obtained for our covariates. This finding demonstrates the critical importance of explicitly modelling the dispersion parameter in our dataset. By incorporating the consideration of differential precisions in abundance estimates, we can ensure that our assessments of coefficients and statistical significance are more robust and reliable. Finally, the model fit was assessed through diagnostic plots. Evident from the partial effects plots, there was an observable inherent grouping in the residuals that the model failed to account for (Figures C4 and C4).

### **Conclusions and recommendations**

In the conducted study, the use of a GLM with an Inverse Gaussian error distribution was found to provide a better fit compared to a traditional linear model with log-transformed predictor variable. This suggests a preference for a GLM framework over standard linear models when dealing with data distributions that are non-normal or non-linear. A significant revelation of this study was the importance of explicitly modelling the dispersion parameter, given its critical influence on the precision of estimated abundances. Overlooking this variability in precision could lead to the misidentification of significant variables and an overestimation of p-values.

Moreover, stepwise model selection was employed to select the final model, which includes temperature and average depth as main effects and models the dispersion to vary explicitly with bin size. Overall, the final model exhibits the lowest GAIC when compared to all other models fitted in this project. Hence, model selection is recommended to remove extraneous variables and

improve model fit. However, preliminary diagnostic plots indicate that there is still significant room for improvement in the final model.

## References

- Akaike, H. (1978) On the Likelihood of a Time Series Model. *The Statistician*, **27**, 217.
- Boldina, I. & Beninger, P.G. (2016) Strengthening statistical usage in marine ecology: Linear regression. *Journal of Experimental Marine Biology and Ecology*, **474**, 81–91.
- Doubek, J.P., Goldfarb, S.K. & Stockwell, J.D. (2020) Should we be sampling zooplankton at night? *Limnology and Oceanography Letters*, **5**, 313–321.
- Hays, G.C. (2003) A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia*, **503**, 163–170.
- Hernández-León, S., Koppelman, R., Fraile-Nuez, E., Bode, A., Mompeán, C., Irigoien, X., Olivar, M.P., Echevarría, F., Fernández de Puellas, M.L., González-Gordillo, J.I., Cózar, A., Acuña, J.L., Agustí, S. & Duarte, C.M. (2020) Large deep-sea zooplankton biomass mirrors primary production in the global ocean. *Nature communications*, **11**, 6048.
- Mack, H.R., Conroy, J.D., Blocksom, K.A., Stein, R.A. & Ludsin, S.A. (2012) A comparative analysis of zooplankton field collection and sample enumeration methods. *Limnology and Oceanography: Methods*, **10**, 41–53.
- Menge, D.N. L., MacPherson, A.C., Bytnerowicz, T.A., Quebbeman, A.W., Schwartz, N.B., Taylor, B.N. & Wolf, A.A. (2018) Logarithmic scales in ecological data presentation may cause misinterpretation. *Nature ecology & evolution*, **2**, 1393–1402.
- Mitra, A., Castellani, C., Gentleman, W.C., Jónasdóttir, S.H., Flynn, K.J., Bode, A., Halsband, C., Kuhn, P., Licandro, P., Agersted, M.D., Calbet, A., Lindeque, P.K., Koppelman, R., Møller, E.F., Gislason, A., Nielsen, T.G. & St. John, M. (2014) Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. *Progress in Oceanography*, **129**, 176–199.
- O’Hara, R.B. & Kotze, D.J. (2010) Do not log-transform count data. *Methods in Ecology and Evolution*, **1**, 118–122.
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O’Connell, A.F., Miller, P.I. & Louzao, M. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, **156**, 94–104.
- Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507–554.
- Steinberg, D.K. & Landry, M.R. (2017) Zooplankton and the Ocean Carbon Cycle. *Annual review of marine science*, **9**, 413–444.
- Turner, J.T. (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs, **43**, 255–266.
- Vereshchaka, A., Abyzova, G., Lunina, A., Musaeva, E. & Sutton, T. (2016) A novel approach reveals high zooplankton standing stock deep in the sea. *Biogeosciences*, **13**, 6261–6271.

## Appendix D - Supplementary information for Chapter 3

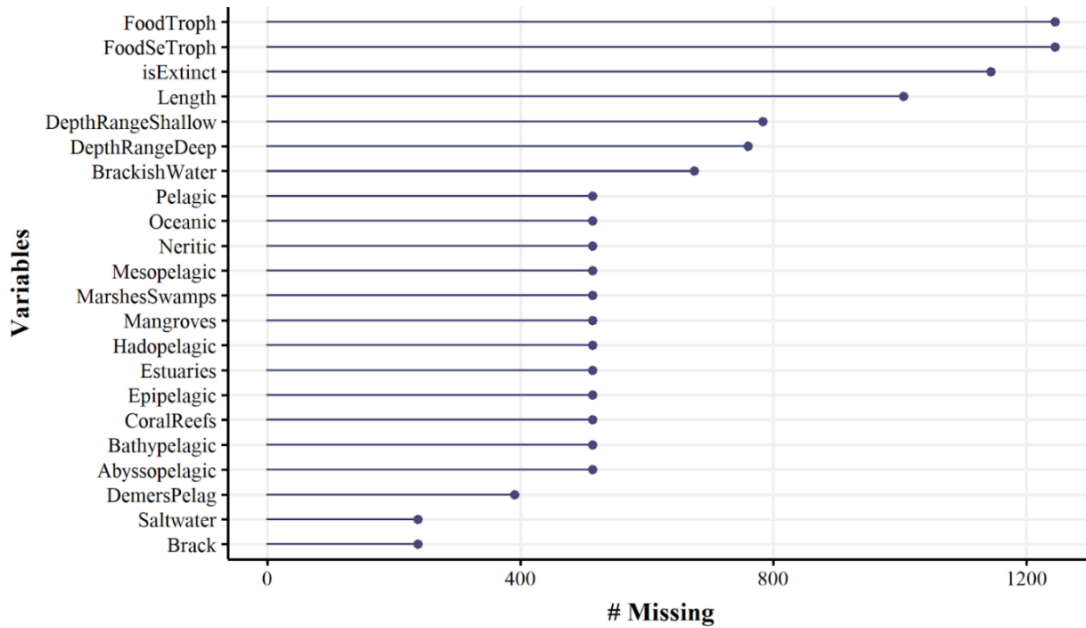
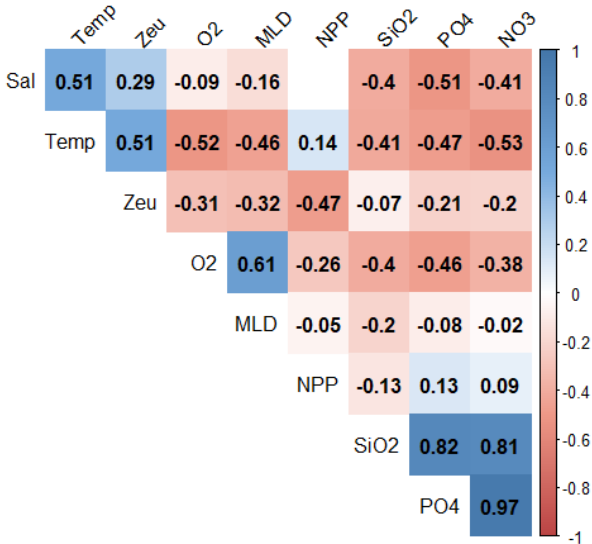


Figure D1: Number of missing biotic information collected in SealifeBase and FishBase for 1499 mesopelagic mesozooplankton and micronekton species.

MESO



BOTH

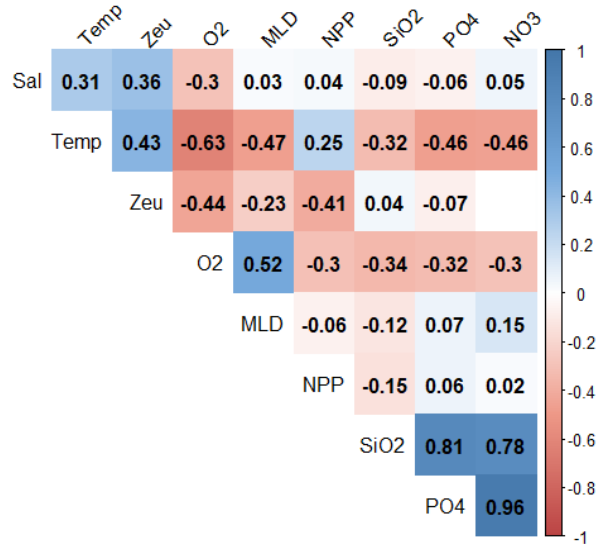


Figure D2: Pearson correlation coefficients in two sets of environmental variables (mesopelagic – MESO; and epi-mesopelagic - BOTH): salinity (Sal), water temperature (Temp), euphotic zone depth (Zeu), dissolved oxygen concentration (O<sub>2</sub>), mixed layer depth (MLD), net primary production (NPP), silicate (SiO<sub>2</sub>), phosphate (PO<sub>4</sub>) and nitrate (NO<sub>3</sub>).

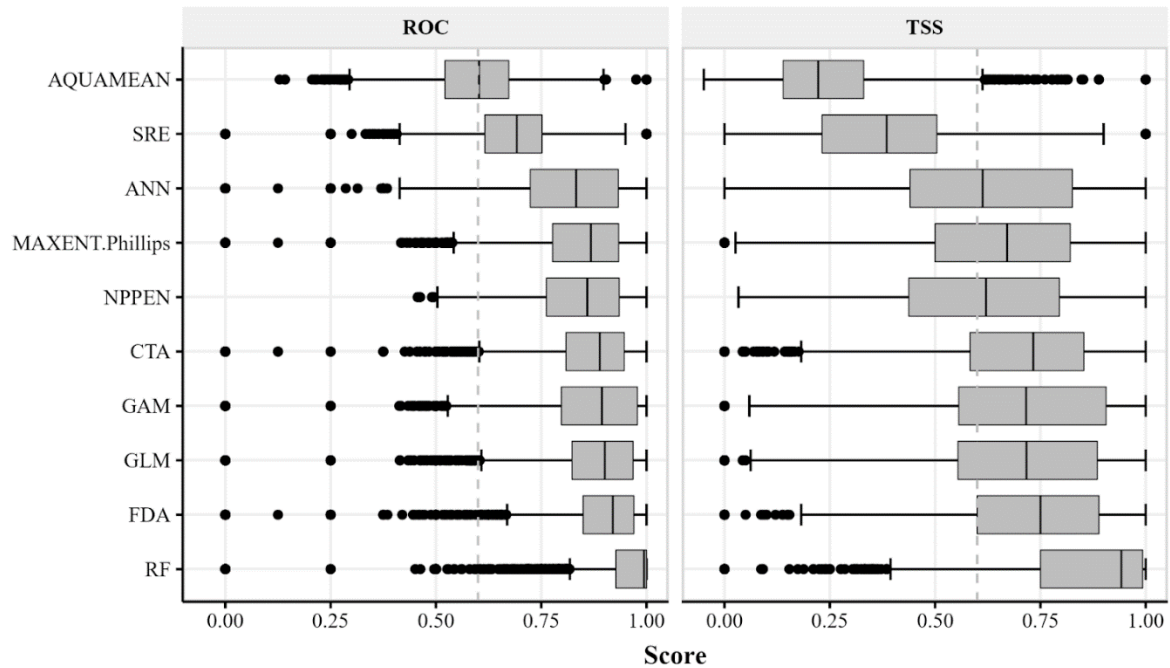


Figure D3: Distributions of the overall accuracy of different individual SDMs (all runs). Box plots showing the estimated median (horizontal black lines), 25th–75th percentiles (grey boxes) and the  $1.5 \times \text{IQR}$  (whiskers) of ROC (AUC score) and TSS. The dashed horizontal grey line represents the selecting criterion (AUC = 0.6, TSS = 0.6) of individual models for inclusion in the ensemble model.

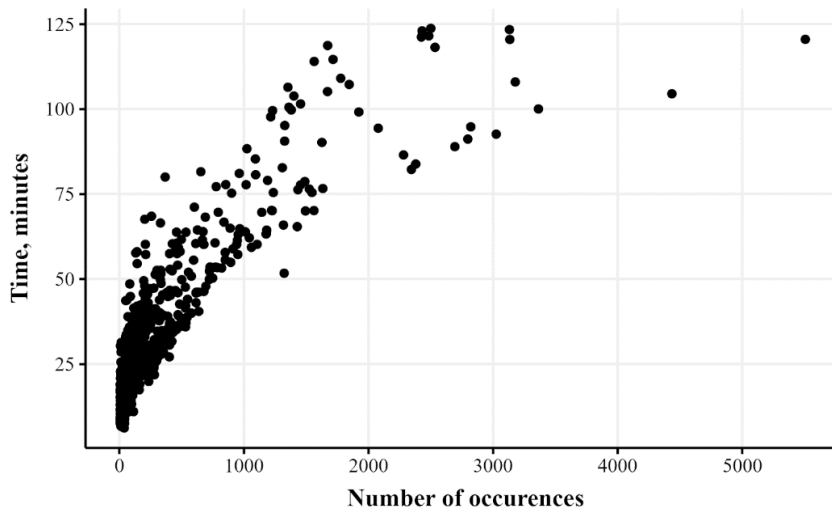


Figure D4: Time (in minutes) that took each species to run in relation to the number of presence data available.

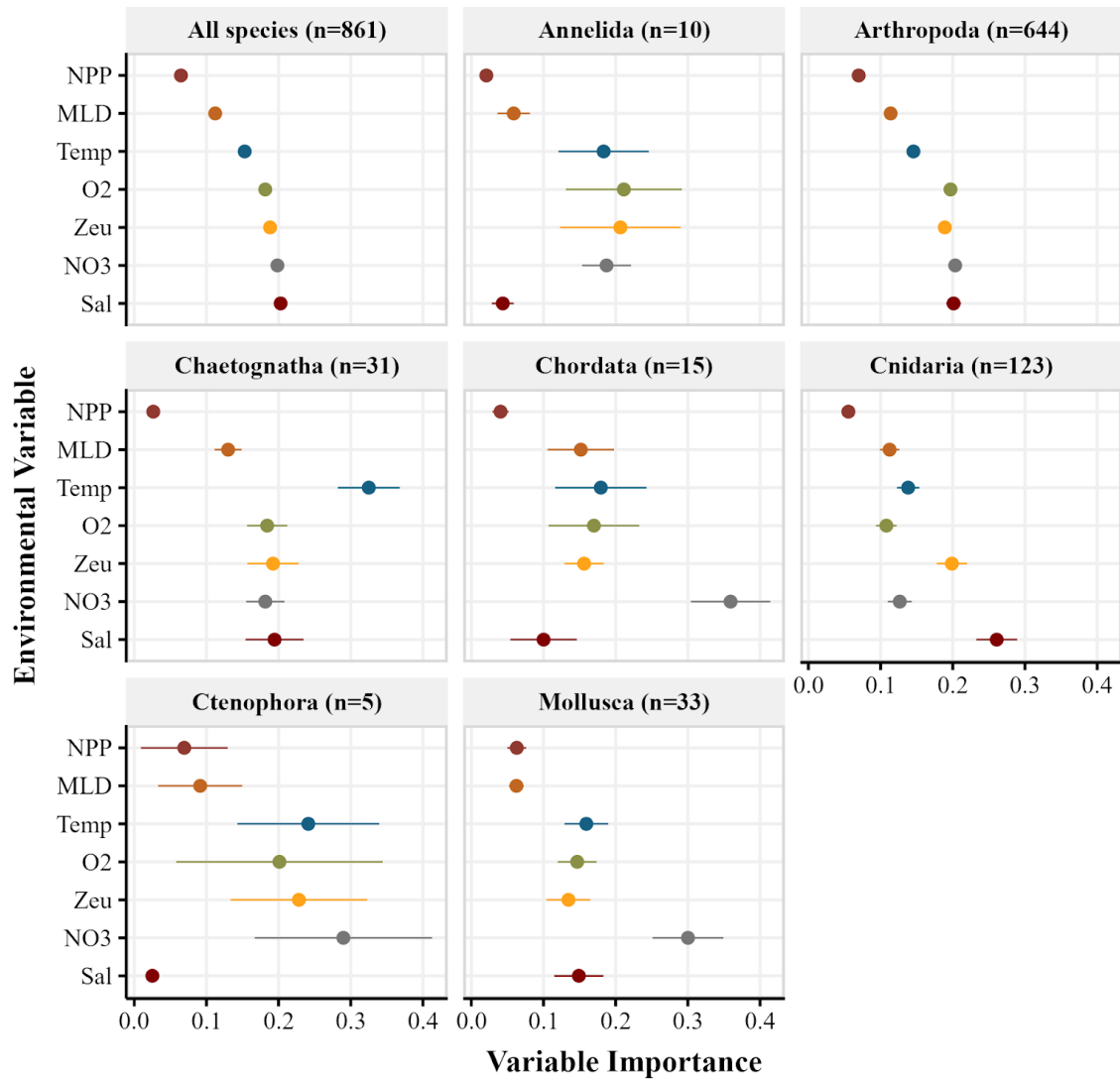
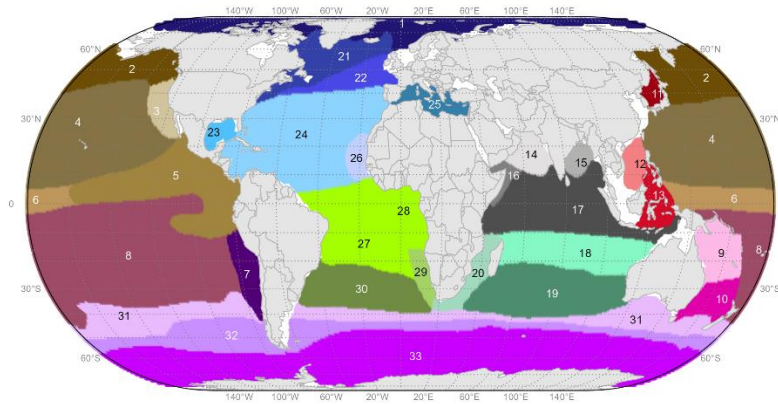


Figure D5: Variable importance of nine environmental variables: water temperature (Temp), salinity (Sal), mixed layer depth (MLD), dissolved oxygen concentration (O2), nitrate (NO3), phosphate (PO4) and silicate (SiO2), net primary production (NPP) and euphotic zone depth (Zeu). Means and 95% confidence intervals are shown per phylum and for all the species combined.

A



B

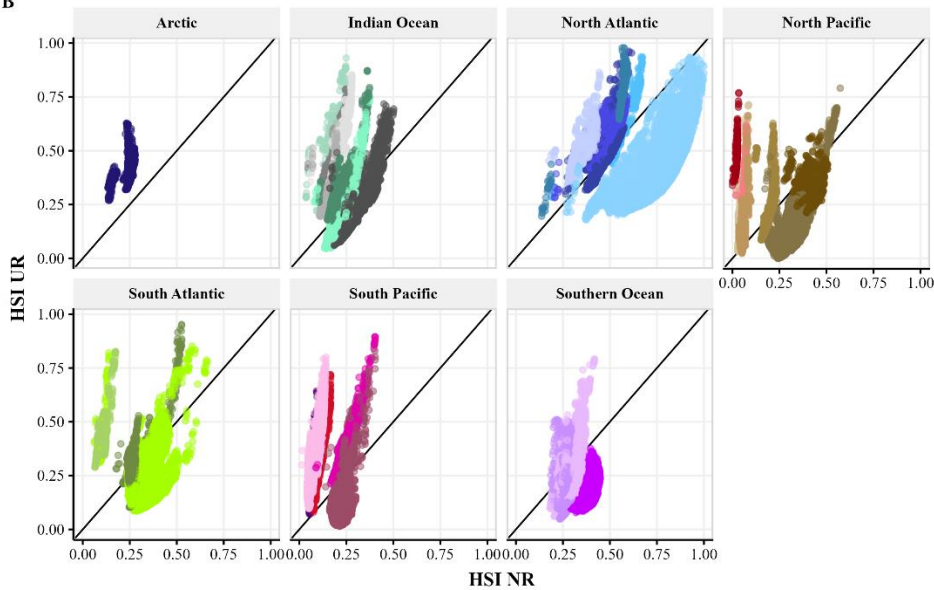


Figure D6: Comparison of habitat suitability index (HSI) between unrestricted and native range models for each mesopelagic province. A) map of Sutton et al. (2017) mesopelagic provinces. Areas with depths less than 200 m are shaded in white. Numbers correspond to the following mesopelagic provinces' names: 1-Arctic, 2-Subarctic Pacific, 3-California Current, 4-Northern Central Pacific, 5-Eastern Tropical Pacific, 6-Equatorial Pacific, 7-Peru Upwelling/Humboldt Current, 8-Southern Central Pacific, 9-Coral Sea, 10-Tasman Sea, 11-Sea of Japan, 12-South China Sea, 13-Southeast Asian Pocket basins,14-Arabian Sea,15-Bay of Bengal,16-Somali Current,17-Northern Indian Ocean, 18-Mid-Indian Ocean, 19-Southern Indian Ocean, 20-Agulhas Current, 21-Northwest Atlantic Subarctic,22-North Atlantic Drift, 23-Gulf of Mexico, 24-Central North Atlantic, 25-Mediterranean, 26-Mauritania/Cape Verde, 27-Tropical and West Equatorial Atlantic, 28-Guinea Basin and East Equatorial Atlantic, 29-Benguela Upwelling, 30-South Atlantic, 31-Circumglobal Subtropical Front, 32-Subantarctic waters, 33-Antarctic/Southern Ocean. B) Relationship between the HSI in unrestricted and native ranges for each ocean basin. The solid black line is a 1:1 line. Each mesopelagic province is colored based on the scheme shown in A).

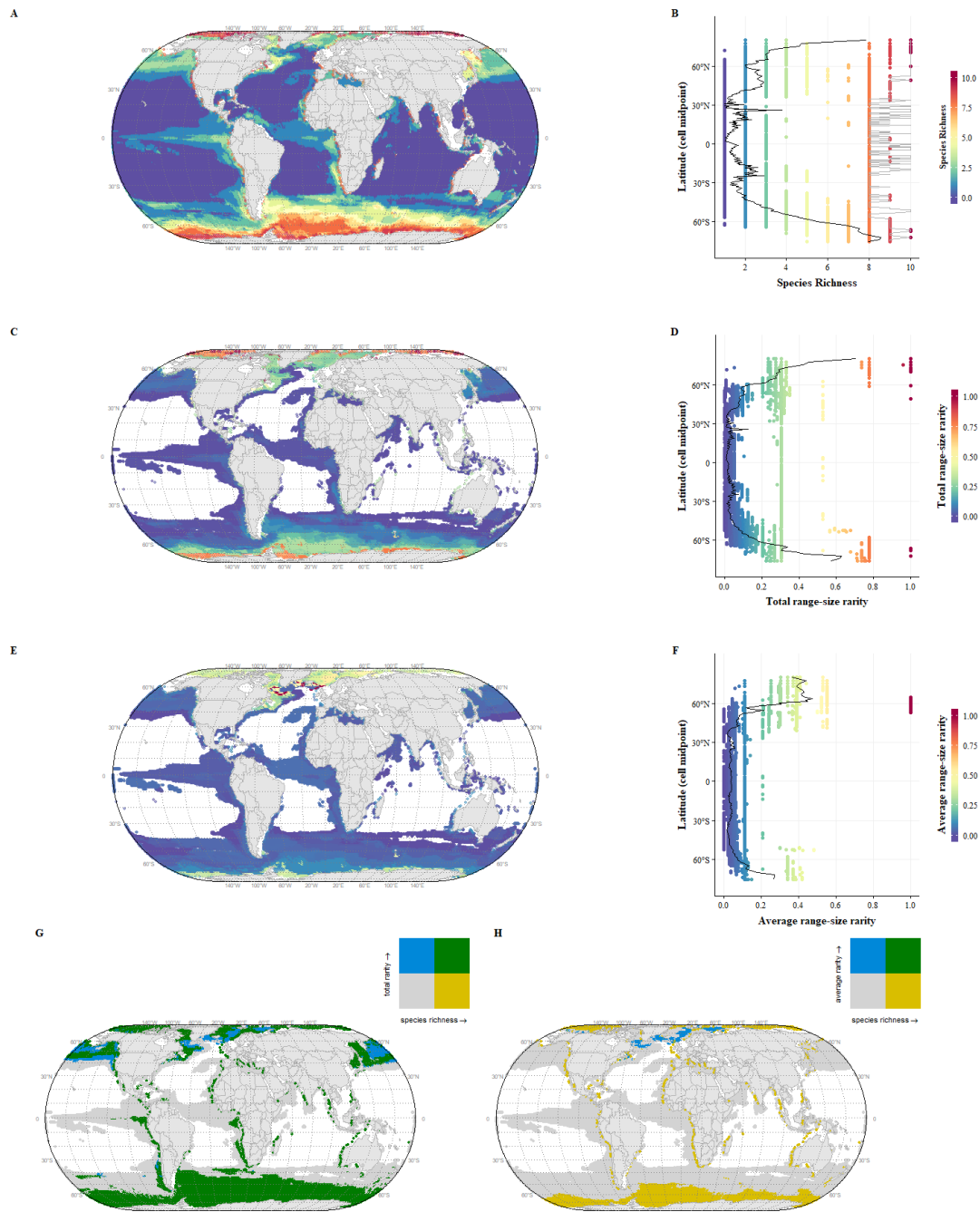


Figure D7: Mesopelagic diversity of phylum Annelida. A) Mesopelagic Annelida species richness map (n=10). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 50<sup>th</sup> percentile. All maps in Eckert IV global equal area projection.



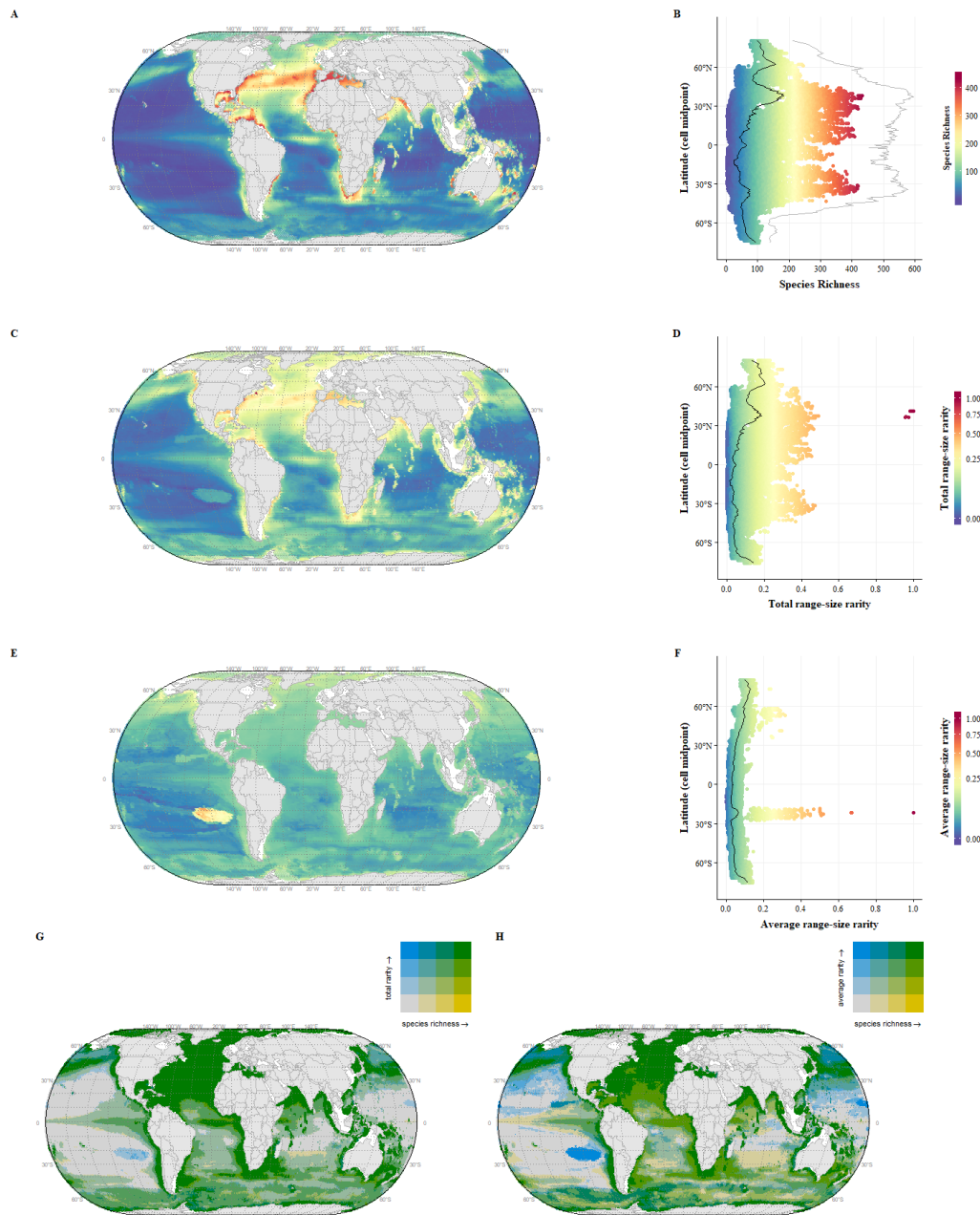


Figure D8: Mesopelagic diversity of phylum Arthropoda. A) Mesopelagic Arthropoda species richness map (n=644). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. Note, that square root transformation was applied to color scale in plots C-F for better interpretation. All maps in Eckert IV global equal area projection.

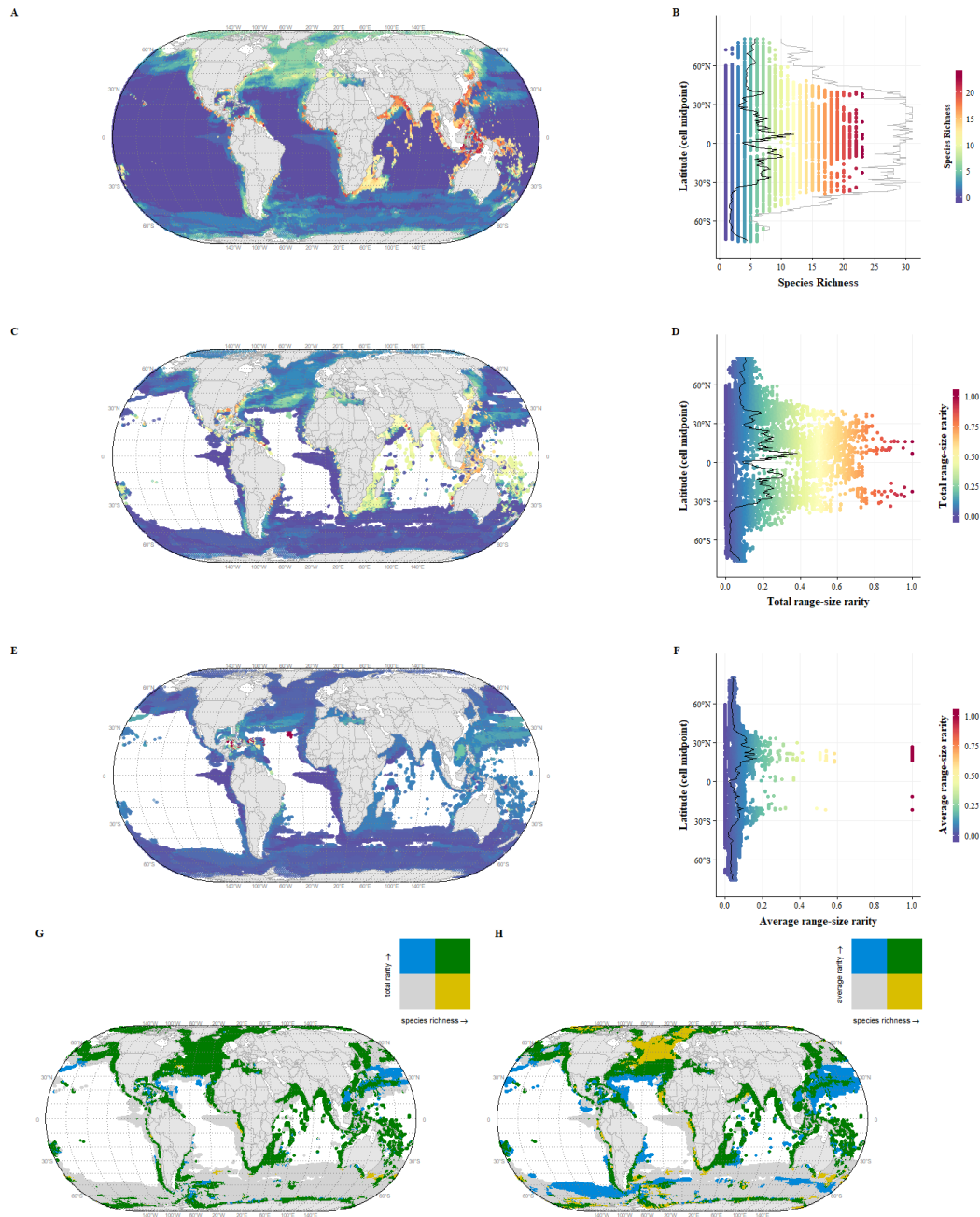


Figure D9: Mesopelagic diversity of phylum Chaetognatha. A) Mesopelagic Chaetognatha species richness map (n=31). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 50<sup>th</sup> percentile. All maps in Eckert IV global equal area projection.

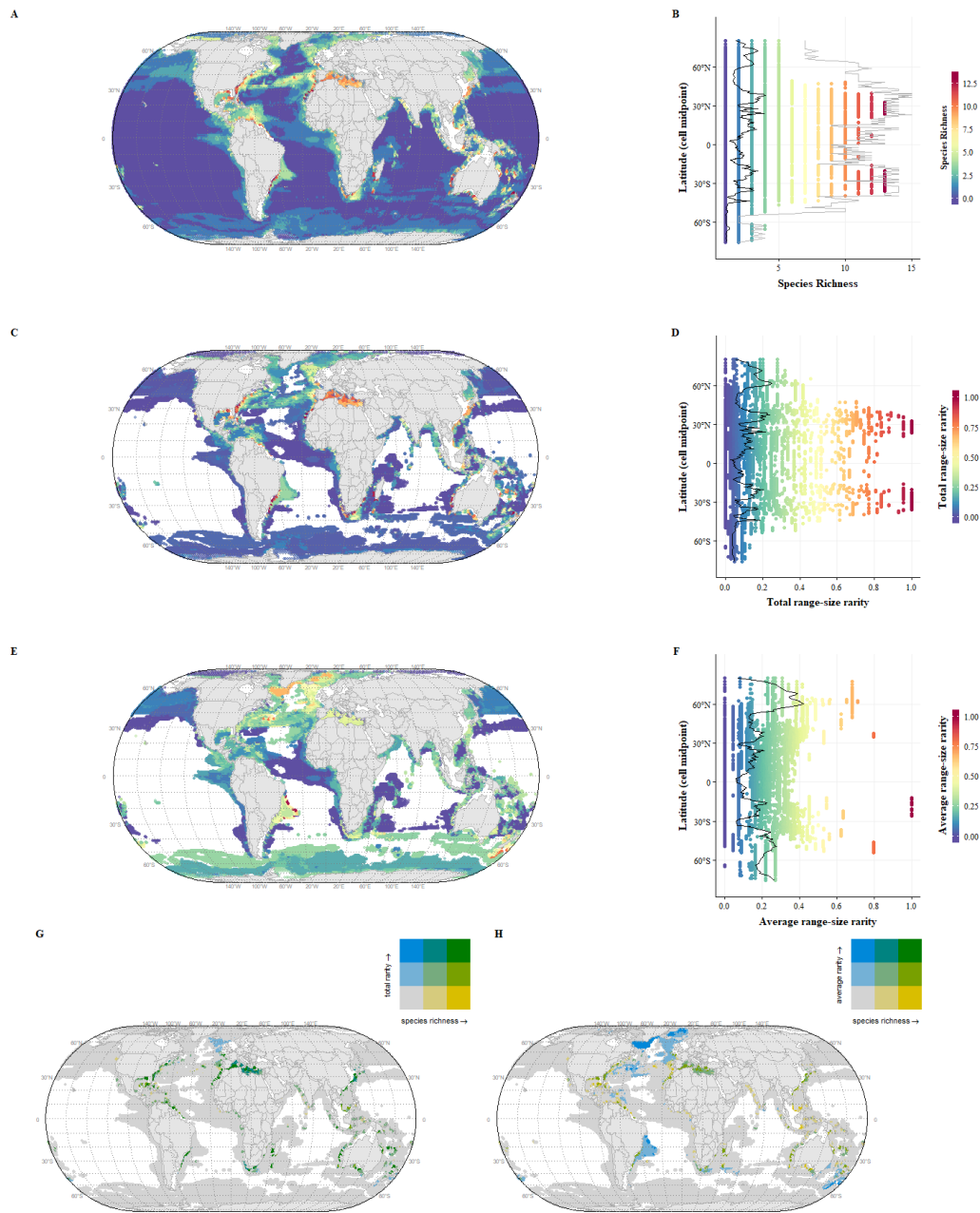


Figure D10: Mesopelagic diversity of phylum Chordata. A) Mesopelagic Chordata species richness map (n=15). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, breaks were created by dividing mapping data into 3 equal intervals. All maps in Eckert IV global equal area projection.

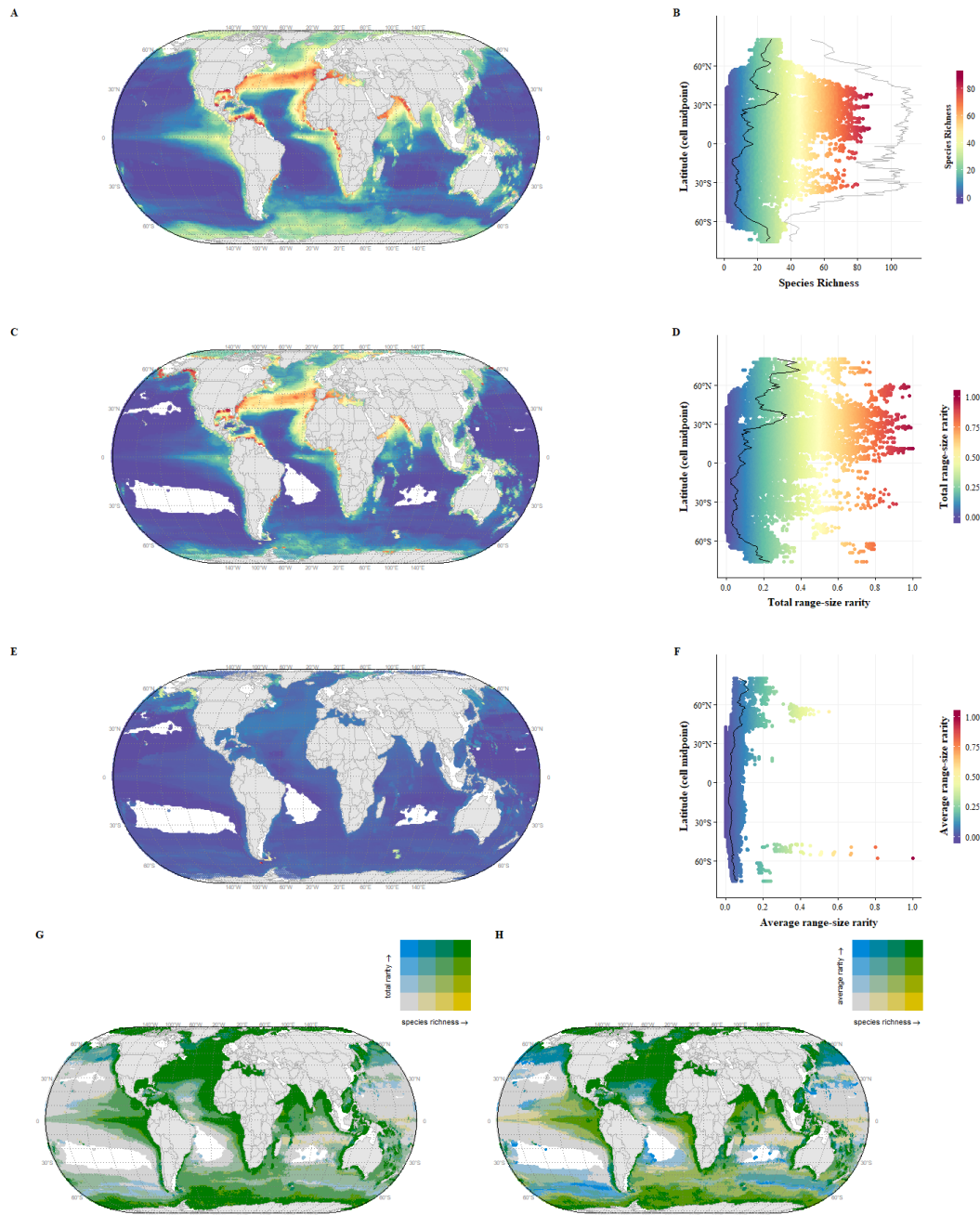


Figure D11: Mesopelagic diversity of phylum Cnidaria. A) Mesopelagic Cnidaria species richness map (n=123). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. All maps in Eckert IV global equal area projection.

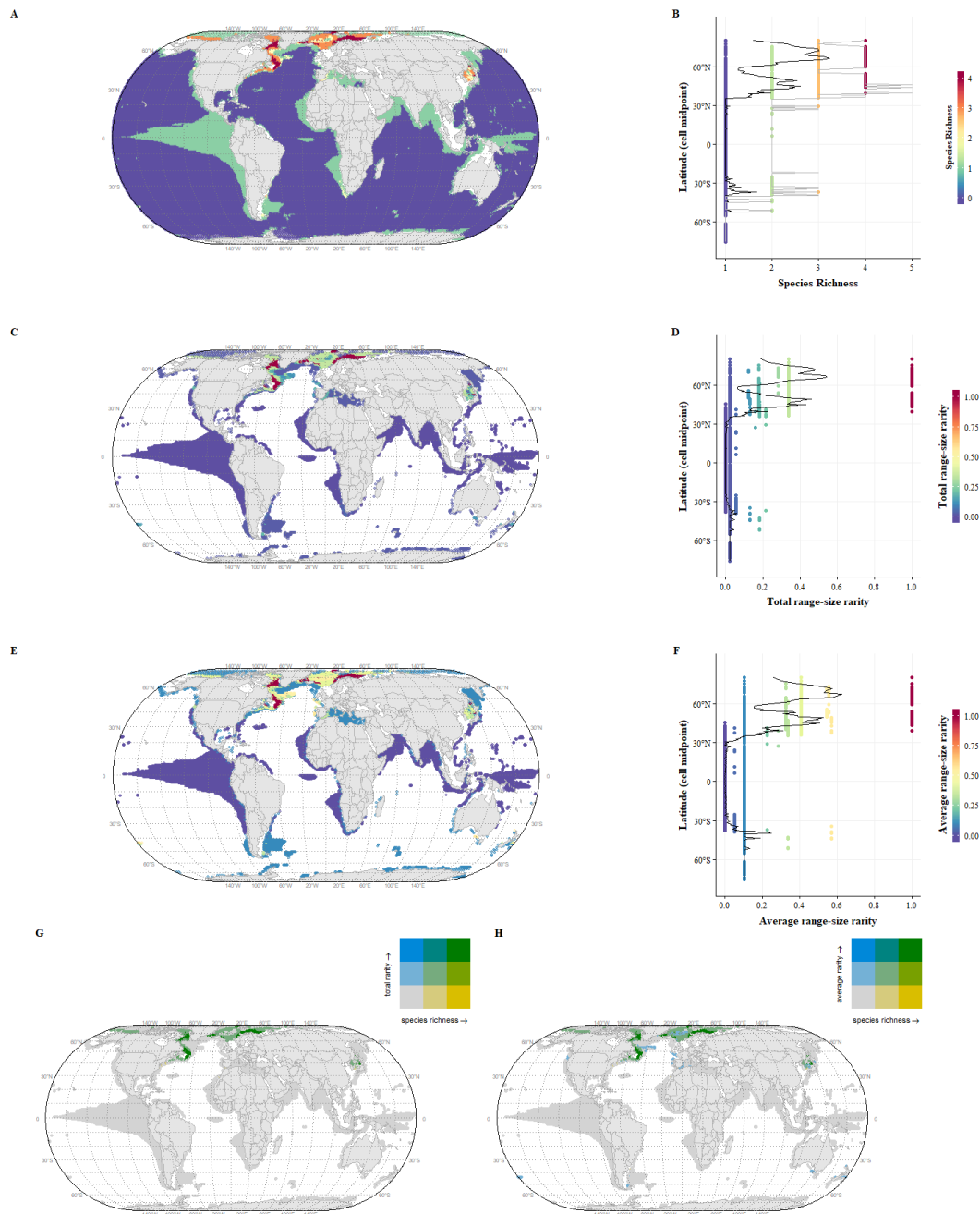
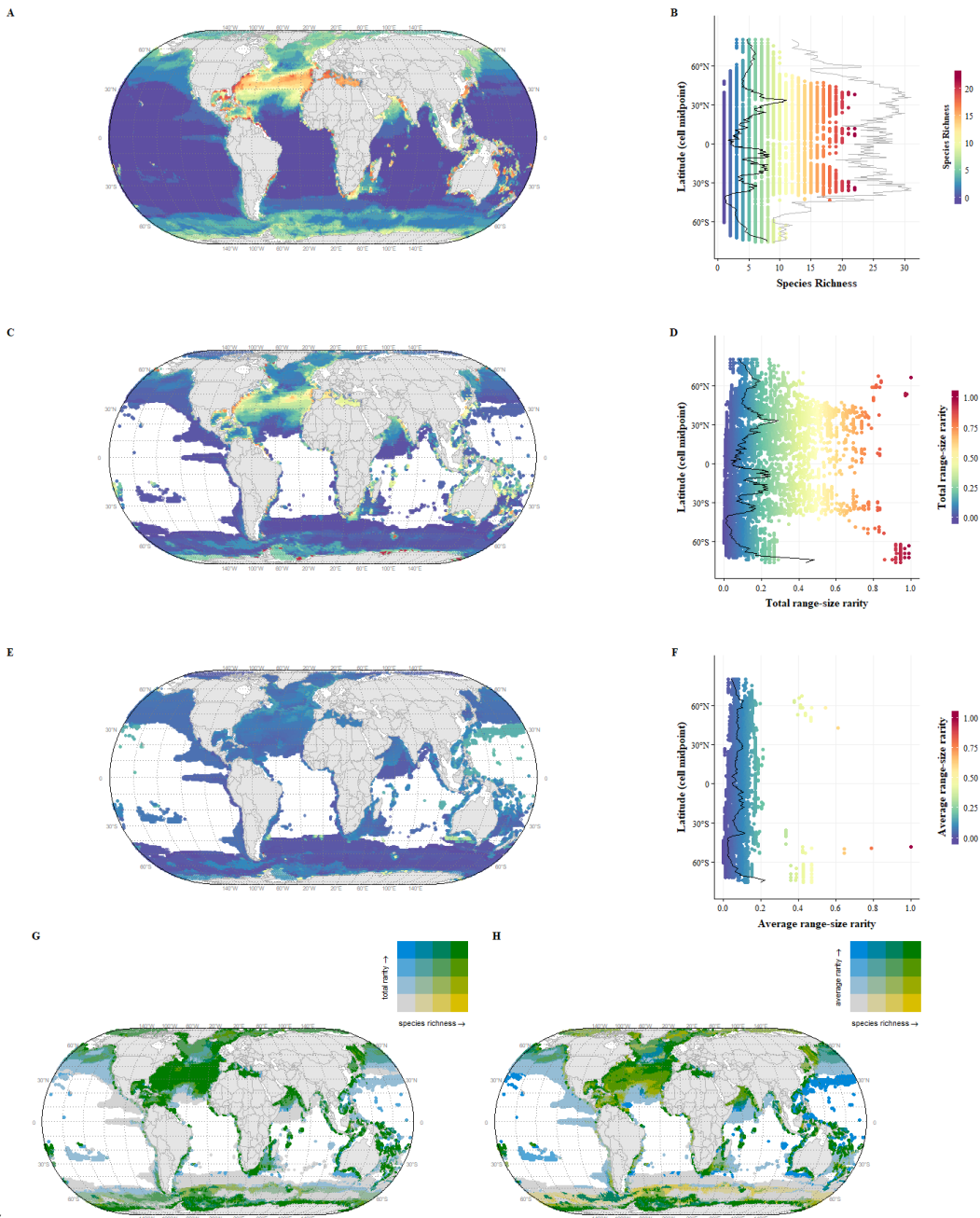


Figure D12: Mesopelagic diversity of phylum Ctenophora. A) Mesopelagic Ctenophora species richness map (n=8). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created by dividing mapping data into 3 equal intervals. All maps in Eckert IV global equal area projection.



Z

Figure D13: Mesopelagic diversity of phylum Mollusca. A) Mesopelagic Mollusca species richness map (n=33). B) Latitudinal gradient showing mean (solid black line) and a number of unique species (grey solid line) per latitude. C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. Solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. Solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. All maps in Eckert IV global equal area projection.

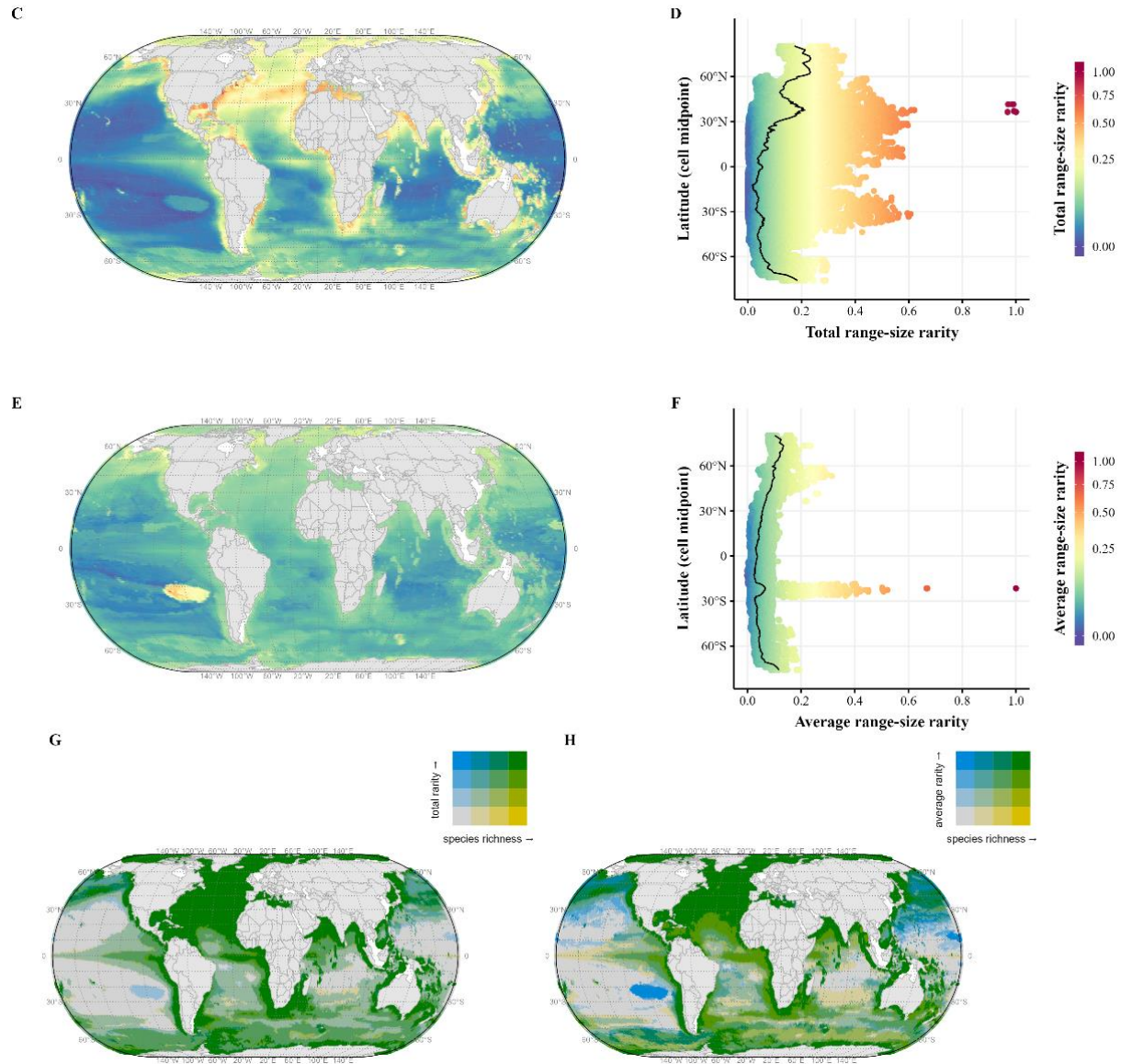


Figure D13: Average and total rarity plots using all species ( $n=861$ ). C) Total range-size rarity map and D) latitudinal gradient for the total range-size rarity. A solid black line is a mean total range-size rarity per latitude. E) Average range-size rarity map and F) latitudinal gradient for the average range-size rarity. A solid black line is a mean total range-size rarity per latitude. G) 2D map depicting the relationship between the species richness and total range-size rarity. H) 2D map depicting the relationship between the species richness and average range-size rarity. For both maps G and F, quantile breaks were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. All maps in Eckert IV global equal area projection.

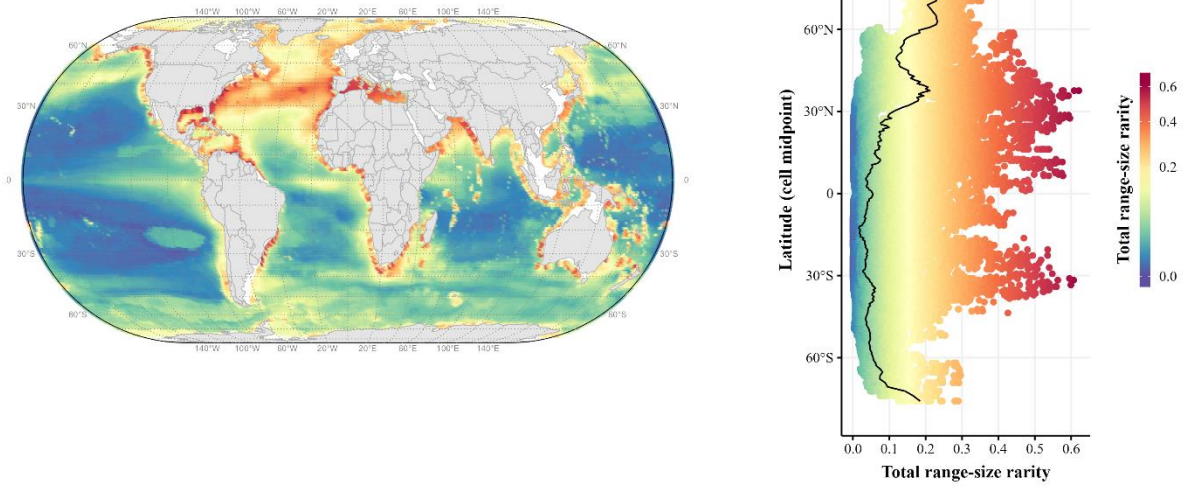


Figure D14: Filtered total range-size rarity plots. Filtered total range-size rarity map (rarity values  $>0.97$  were removed; left) and latitudinal gradient for the total range-size rarity (right). A solid black line is a mean total range-size rarity per latitude. The map in Eckert IV global equal area projection.



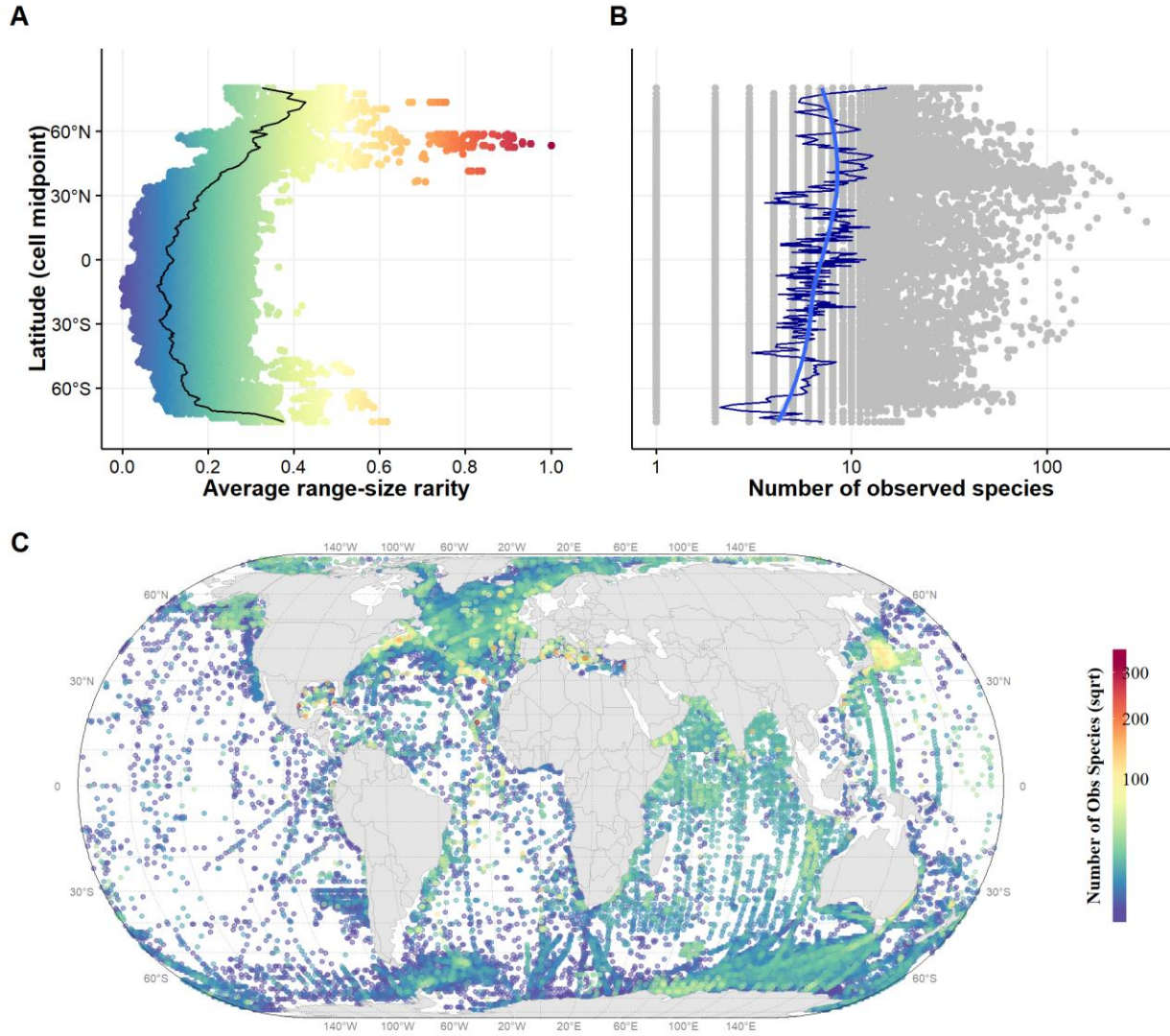


Figure D15: Sampling effort. A) latitudinal gradient of average range-size rarity in relation to the distribution of B) number of observed species. Black line is a mean value per latitude, blue line is a fitted loess line. C) Sampling effort showing number of species recorded in each cell. The map in Eckert IV global equal area projection.

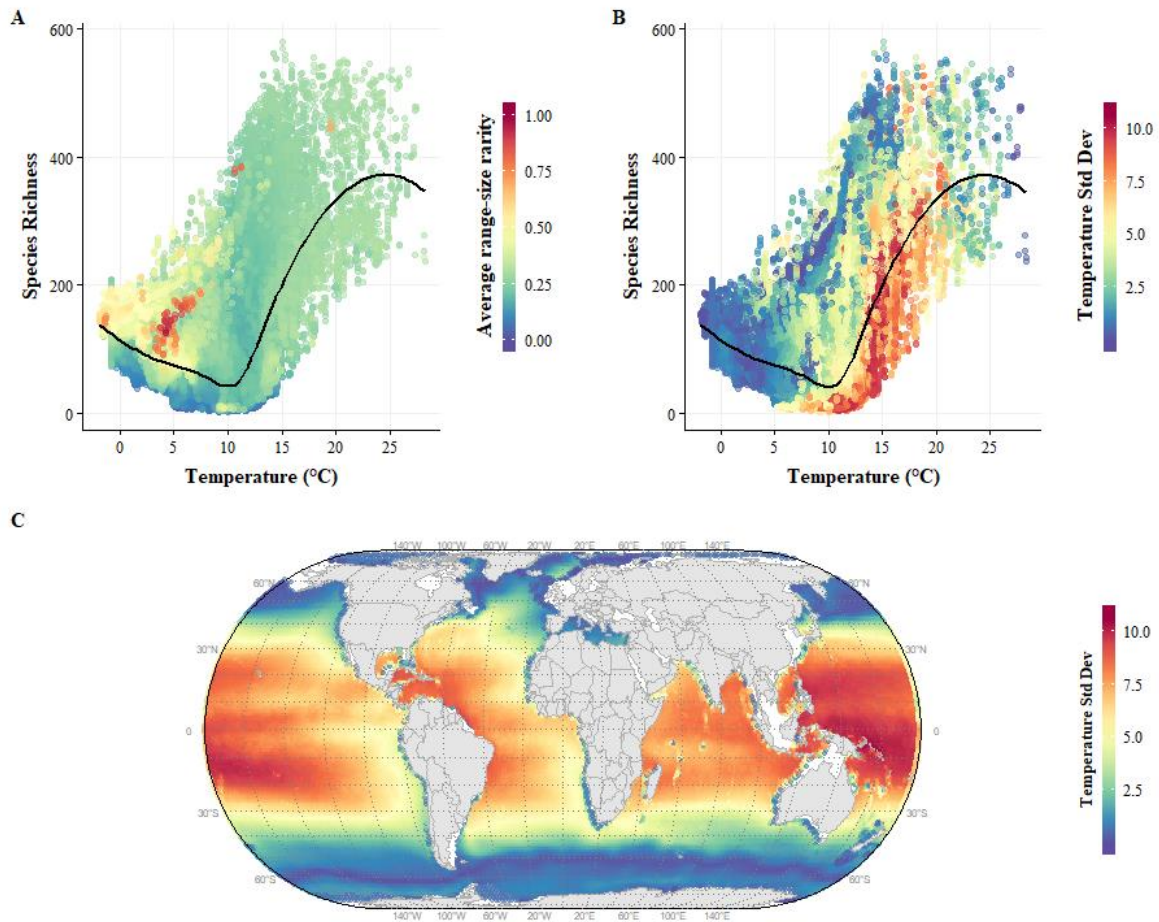


Figure D16: Variation in temperature in the mesopelagic zone. A) Relationship between the calculated species richness with the seawater temperature of the mesopelagic layer colored by average rarity. Black line is loess-fitted line. B) Relationship between the calculated species richness with the seawater temperature of colored by standard deviation of climatological mean temperature (°C) in the mesopelagic layer. Black line is loess-fitted line. C) Map showing the distribution of standard deviation of climatological mean temperature (°C) in the mesopelagic layer in the mesopelagic zone. The map in Eckert IV global equal area projection.

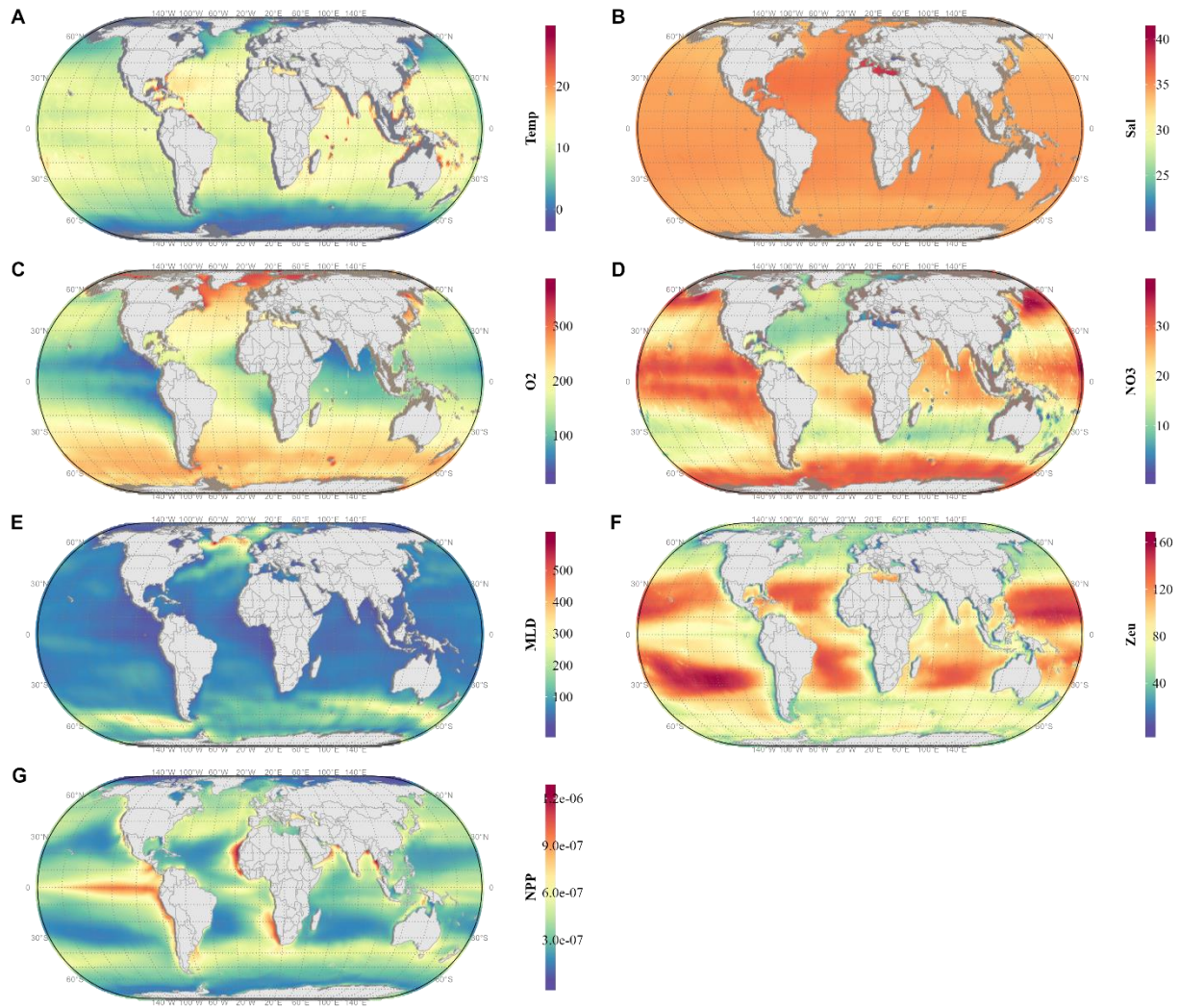


Figure D17: Distribution of annual climatology data in the mesopelagic layer (MESO) for A) Temperature (Temp, °C), B) Salinity (Sal, ‰), C) dissolved oxygen ( $O_2$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), D) dissolved nitrate ( $\text{NO}_3$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), E) dissolved phosphate ( $\text{PO}_4$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), F) euphotic zone depth (Zeu, m), and G) Net primary production (NPP,  $\text{PgC}\cdot\text{m}^{-3}$ ). All maps in Eckert IV global equal area projection.

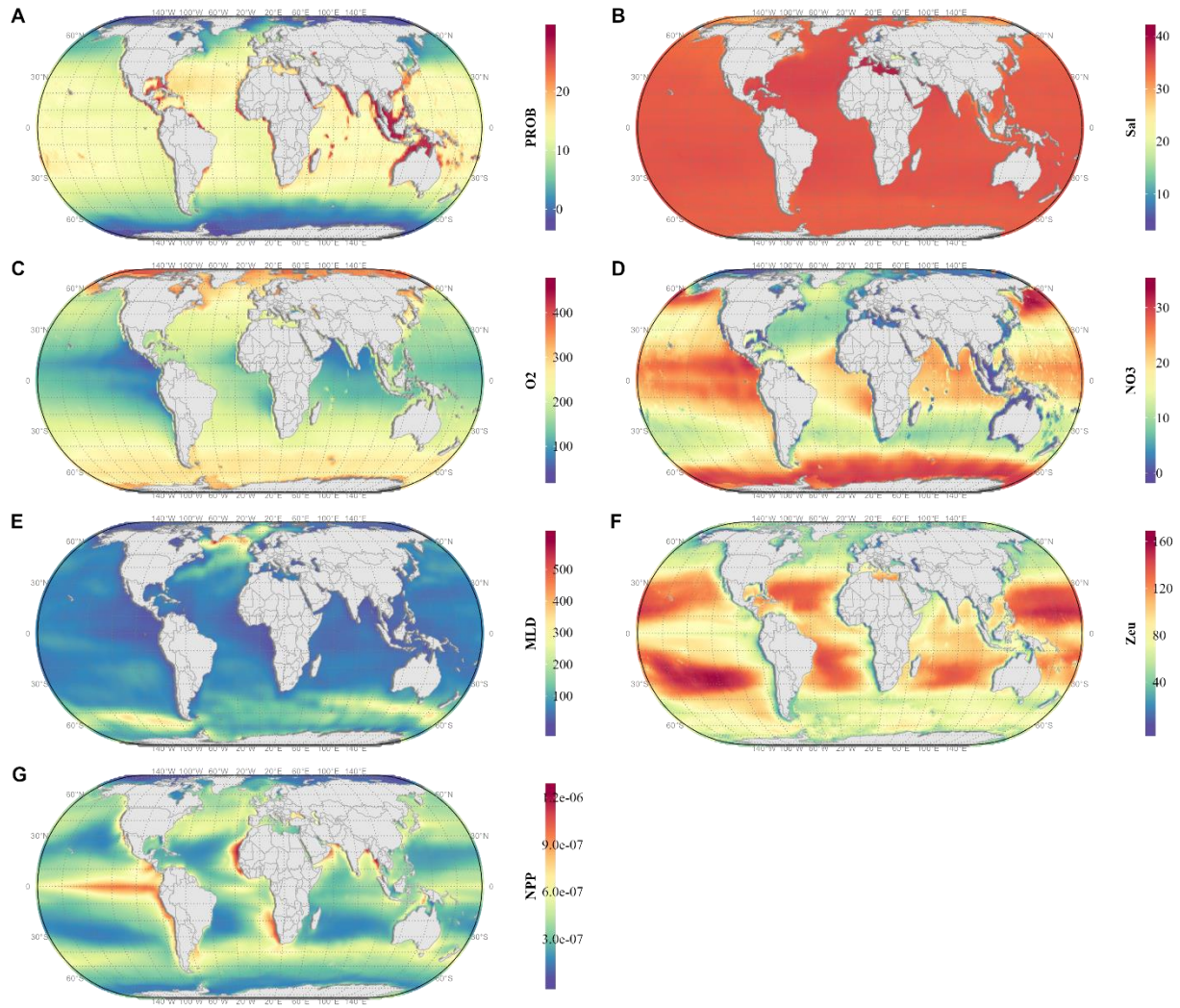


Figure D18: Distribution of annual climatology data in the combined epi- and mesopelagic layers (BOTH) for A) Temperature (Temp, °C), B) Salinity (Sal, ‰), C) dissolved oxygen ( $O_2$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), D) dissolved nitrate ( $NO_3$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), E) dissolved phosphate ( $PO_4$ ,  $\mu\text{mol}\cdot\text{kg}^{-1}$ ), F) euphotic zone depth (Zeu, m), and G) Net primary production (NPP,  $\text{PgC}\cdot\text{m}^{-3}$ ). All maps in Eckert IV global equal area projection.

Table D1: Parameter settings of nine individual SDM in the *biomod2* package in R and NPPEN and AQUAMEAN models. Selection criteria are defined by Jones & Cheung, 2011 and Albouy et al, 2011).

<b>Algorithm</b>	<b>Model Parameters</b>	<b>Parameter Value</b>
Maximum Entropy model (MaxEnt)	Weighted response weights	0.5
	Maximum number of iterations	200
	Linear/quadratic/product/ threshold/hinge feature thresholds	TRUE
	Product Feature thresholds	80
	Quadratic feature threshold	10
	Hinge feature threshold	15
	Regularization options: Threshold feature regularization/Categorical feature regularization/Linear/quadratic/product feature regularization/Hinge feature regularization/Regularization multiplier	-1
	Prevalence	0.5
Artificial Neural Network (ANN)	Weighted response weights	NULL
	NbCV (number of cross-validation repetitions)	5
	maximum number of iterations	300
Generalized Linear Model (GLM)	positive convergence tolerance	1e-08
	maximal number of IWLS iterations	1000
	output produced for each iteration	FALSE
	Type	quadratic
	Interaction level	0
	Test	BIC
Generalized Additive Model (GAM)	Algorithm	GAM_gam (n<44) GAM_mgcv (n≥44)
	Weighted response weights	0.5
	Interaction level	0
	irls.reg	0.0
	Epsilon	1e-07
	Max MLE iterations	200

<b>Algorithm</b>	<b>Model Parameters</b>	<b>Parameter Value</b>
	Convergence tolerance	1e-07
	Number of halvings	15
Flexible Discriminant Analysis (FDA)	Weighted response weights	0.5
	Resampling Method	0 mars
Surface Range Envelope (SRE)	Weighted response weights	0.5
	Resampling Quantile	0 0.025
Classification Tree Analysis (CTA)	Weighted response weights	0.5
	Cross-validation folds	10
	Minimum bucket	1
	Minimum split	20
	Complexity parameter	0.01
	Maximum depth	30
Random Forest (RF)	Number of trees to grow	500
	Node size	5
Non-Parametric Probabilistic Ecological Niche (NPPEN)	Single total covariance matrix is used	TRUE
AQUAMEAN	quantiles	.01, 0.25, 0.75, 0.99

Table D2: Evaluation metrics: receiver operating characteristic (ROC) values for all species distribution models included in ENSEMBLE. The mean value of all models runs (standard deviation) is given. Note that the metrics were calculated only for non-failed models (where AUC>0.5).

Model/ Metrics	ANN	AQUA-MEAN	CTA	FDA	GAM	GLM	MAXENT	NPPEN	RF	SRE	ENSEMBLE		
											wmean	mean	median
<b>Threshold-dependent measures</b>													
Sensitivity	88.01 (15.47)	63.53 (24.01)	88.36 (11.92)	90.03 (10.8)	91.76 (9.76)	89.62 (11.46)	82.48 (16.03)	82.66 (13.91)	94.4 (9.07)	62.67 (19.38)	91.88 (6.46)	91.34 (6.84)	90.84 (7.09)
Specificity	72.33 (27.62)	61.58 (25.69)	80.44 (22.75)	82.67 (19.32)	78.56 (21.72)	80.43 (21.06)	81.2 (25.6)	78.35 (15.39)	90.65 (16.36)	74.92 (30.46)	86.46 (7.37)	85.94 (7.62)	86.31 (7.31)
Threshold	462.33 (182.17)	819.26 (133.51)	480.9 (162.93)	493.02 (215.39)	532.45 (231.62)	490.32 (144.64)	344.57 (207)	173.15 (140.07)	486.52 (123.91)	500 (0)	577.64 (99.49)	540.2 (87.82)	592.61 (131.73)
<b>Threshold-independent measures</b>													
AUC	0.8 (0.18)	0.59 (0.12)	0.85 (0.15)	0.88 (0.16)	0.86 (0.17)	0.86 (0.17)	0.83 (0.16)	0.84 (0.11)	0.94 (0.15)	0.68 (0.13)	0.94 (0.04)	0.94 (0.04)	0.94 (0.04)

Table D3: The difference in AUC ( $\Delta$ AUC) between pseudo-absence runs (PA1 vs PA2). Numbers in brackets represent the number of models (species).

<b>Model</b>	<b><math>\Delta</math>AUC &lt; .1</b>	<b><math>\Delta</math>AUC &lt; .25</b>	<b><math>\Delta</math>AUC <math>\geq</math> .25</b>	<b>failed to run</b>
ANN	39.4% (339)	21.5% (185)	9.5% (82)	29.6% (255)
AQUAMEAN	90.8% (782)	9.2% (79)	0.0% (0)	0.0% (0)
CTA	74.6% (642)	17.0% (146)	8.5% (73)	0.0% (0)
FDA	78.2% (673)	10.8% (93)	7.4% (64)	3.6% (31)
GAM	74.8% (644)	11.8% (102)	8.5% (73)	4.9% (42)
GLM	78.4% (675)	11.3% (97)	9.4% (81)	0.9% (8)
MAXENT	70.3% (605)	21.6% (186)	8.1% (70)	0.0% (0)
NPPEN	96.1% (827)	3.9% (34)	0.0% (0)	0.0% (0)
RF	88.2% (759)	3.1% (27)	8.7% (75)	0.0% (0)
SRE	77.1% (664)	16.4% (141)	6.5% (56)	0.0% (0)

Table D4: The difference in AUC ( $\Delta$ AUC) between subsampling runs (RUN1 vs RUN2). Numbers in brackets represent the number of models (species).

<b>Model</b>	<b><math>\Delta</math>AUC &lt; .1</b>	<b><math>\Delta</math>AUC &lt; .25</b>	<b><math>\Delta</math>AUC <math>\geq</math> .25</b>	<b>failed to run</b>
ANN	51.1% (440)	15.8% (136)	15.8% (136)	17.3% (149)
AQUAMEAN	100.0% (861)	0.0% (0)	0.0% (0)	0.0% (0)
CTA	71.5% (616)	13.6% (117)	14.9% (128)	0.0% (0)
FDA	74.2% (639)	9.3% (80)	12.9% (111)	3.6% (31)
GAM	68.9% (593)	12.2% (105)	14.2% (122)	4.8% (41)
GLM	75.1% (647)	9.2% (79)	14.8% (127)	0.9% (8)
MAXENT	69.6% (599)	14.6% (126)	15.8% (136)	0.0% (0)
NPPEN	100.0% (861)	0.0% (0)	0.0% (0)	0.0% (0)
RF	83.0% (715)	6.5% (56)	10.5% (90)	0.0% (0)
SRE	68.8% (592)	13.6% (117)	17.7% (152)	0.0% (0)



## Appendix E - Supplementary information for Chapter 4

Table E1: Models Performance Metric for NPP and POC-derived models using classical definition of the mesopelagic zone (200-1,000 m) and variable depth model (mesopelagic zone depth is location-dependent).

Estimate Model		R <sup>2</sup>	adj. R <sup>2</sup>	$\sigma$	statistic	p.value	df	logLik	N
NPP	Variable Depth	0.24	0.22	1.36	13.03	<0.01	1	-74.82	44
	Classical Depth	0.18	0.16	1.25	9.06	<0.01	1	-67.90	42
POC	Variable Depth	0.30	0.29	1.30	18.21	<0.01	1	-72.84	44
	Classical Depth	0.24	0.22	1.21	12.42	<0.01	1	-66.51	42

Table E2: Models' Coefficients for NPP and POC-derived models using classical definition of the mesopelagic zone (200-1,000 m) and variable depth model (mesopelagic zone depth is location-dependent).

Estimate	Model	term	estimate	std.error	statistic	p.value
NPP	Variable Depth	Intercept	-23.748	6.375	-3.725	<0.01
		Slope	2.078	0.576	3.610	<0.01
	Classical Depth	Intercept	-18.776	5.947	-3.157	<0.01
		Slope	1.618	0.538	3.011	<0.01
POC	Variable Depth	Intercept	-23.748	6.375	-3.725	<0.01
		Slope	2.078	0.576	3.610	<0.01
	Classical Depth	Intercept	-18.776	5.947	-3.157	<0.01
		Slope	1.618	0.538	3.011	<0.01

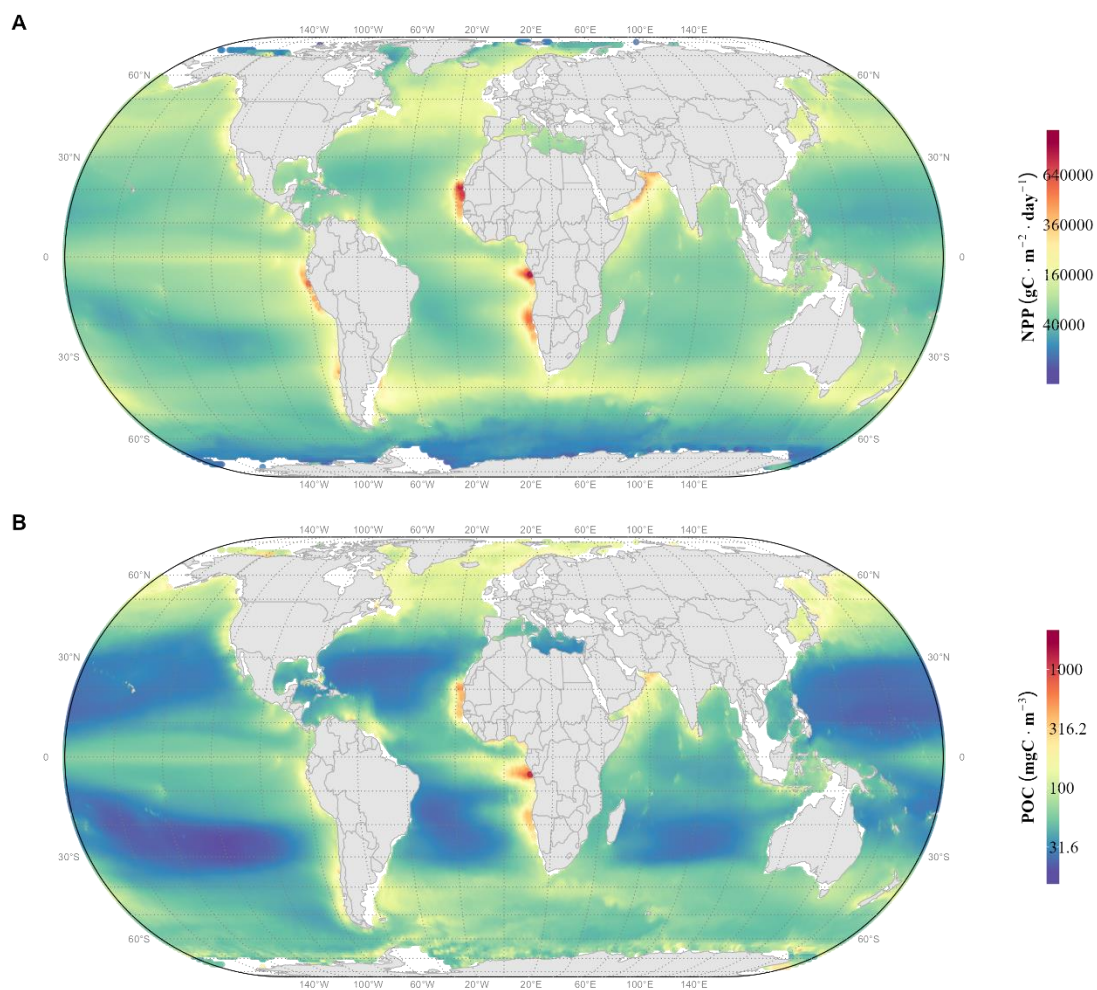


Figure E1: Distribution of annual climatology data for A) NPP (not a square-root- transformed scale) and B) POC (not a log- transformed scale) used in this study. All maps in Eckert IV global equal area projection.

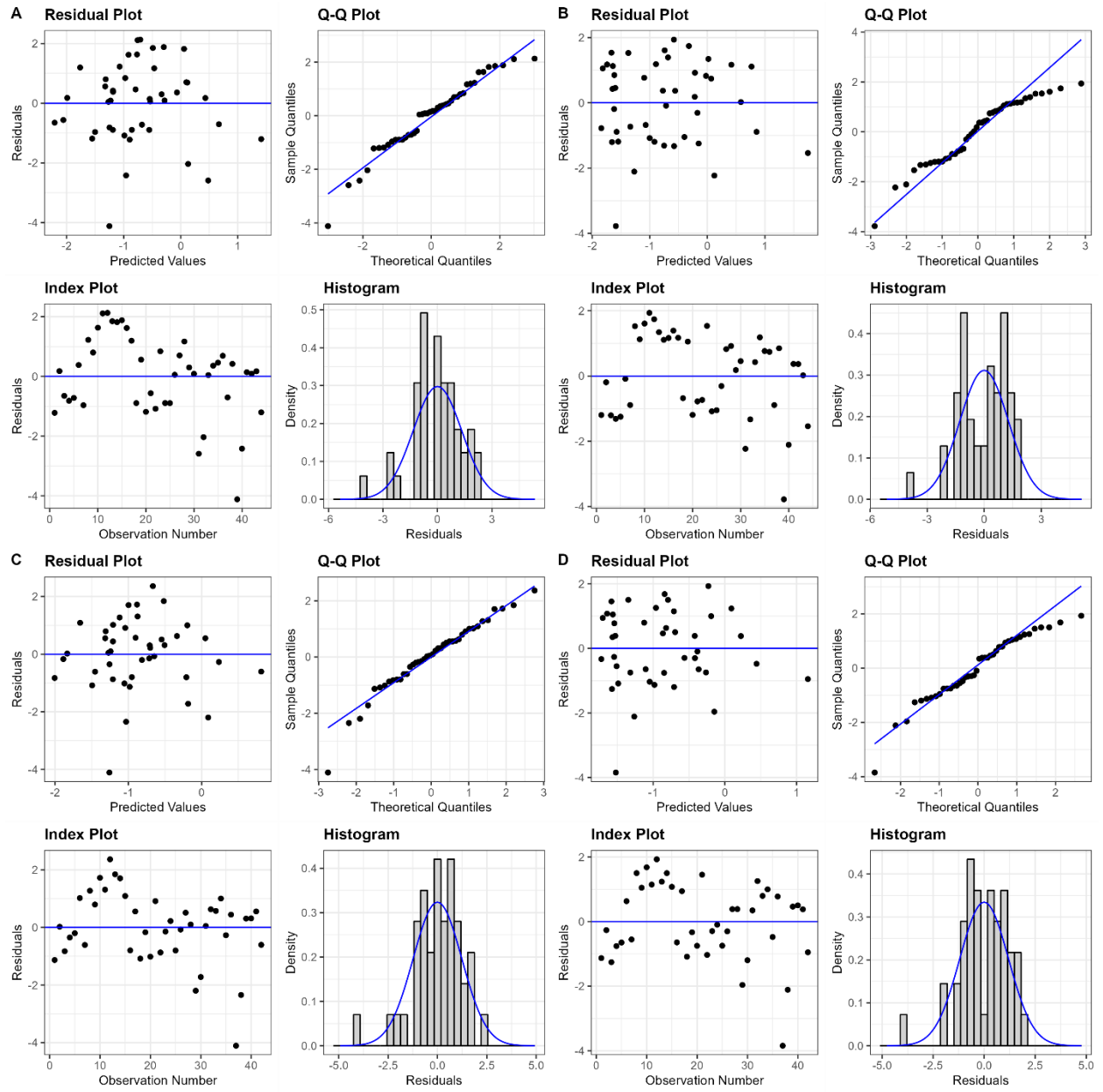


Figure E2: Model Diagnostics for NPP (1<sup>st</sup> column) and POC (2<sup>nd</sup> column) models. Models that use variable mesopelagic depth are shown in row 1 and 200-100 m depth bin estimates are in 2<sup>nd</sup> row.

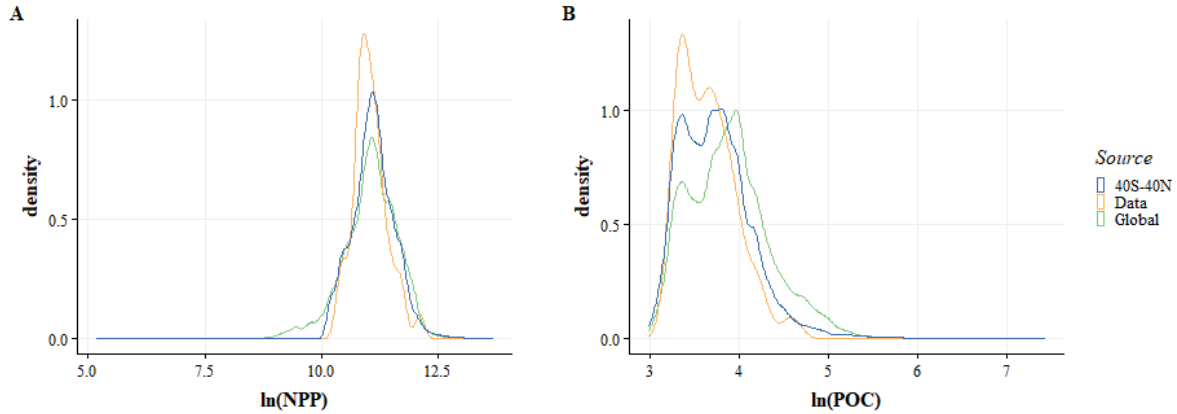


Figure E3: Density plots of log-transformed values of A) NPP and B) POC used in this study. Distribution of global values are shown in green (Global), distribution of values for the biomass values for MALASPINA dataset is shown in orange (Data) and the distribution of values for environmental variables between 40°S-40°N are shown in blue (40S-40N).

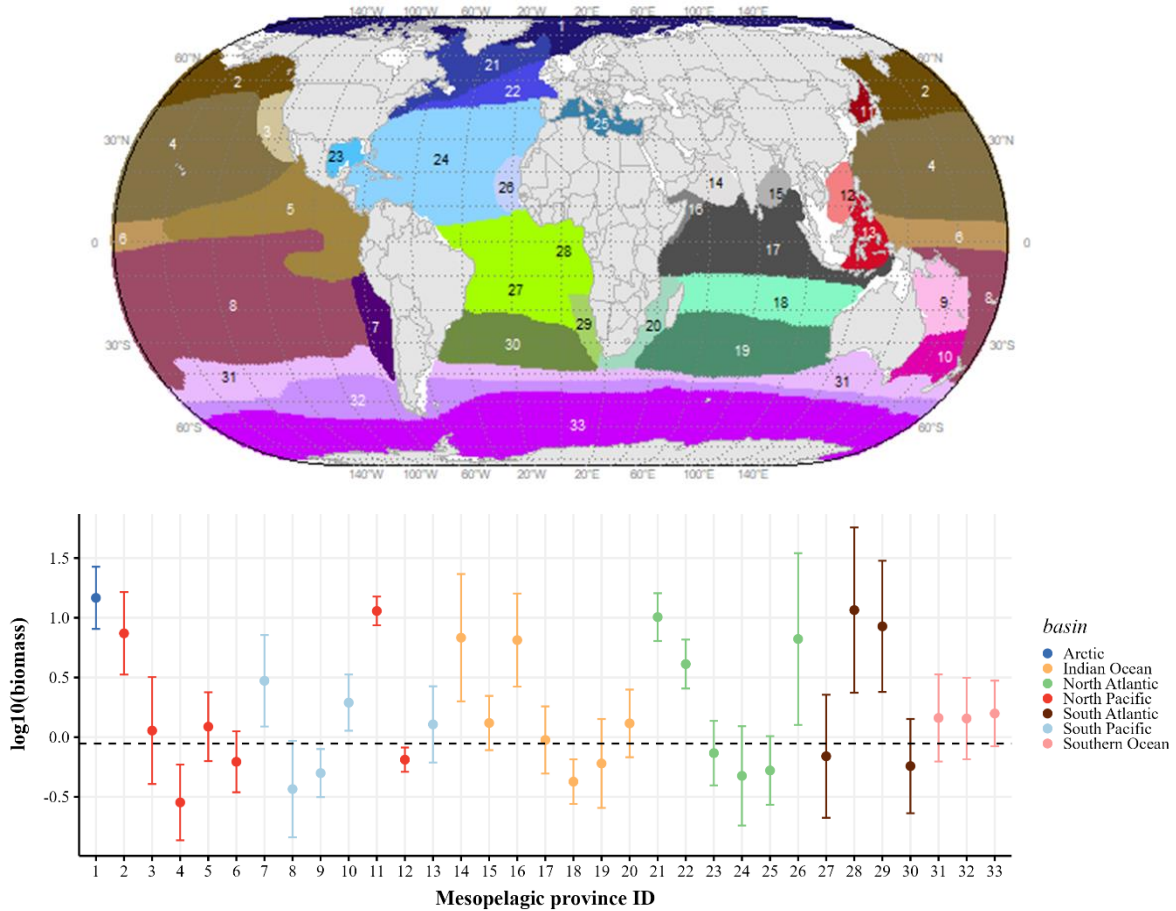


Figure E4: Average mesopelagic biomass in each mesopelagic province. A) map of Sutton et al. (2017) mesopelagic provinces. Areas with depths less than 200 m are shaded in white. The map in Eckert IV global equal area projection. B) Average biomass values per mesopelagic province

defined in Reygondeau *et al.* (2018). Numbers correspond to the following mesopelagic provinces' names: 1-Arctic, 2-Subarctic Pacific, 3-California Current, 4-Northern Central Pacific, 5-Eastern Tropical Pacific, 6-Equatorial Pacific, 7-Peru Upwelling/Humboldt Current, 8-Southern Central Pacific, 9-Coral Sea, 10-Tasman Sea, 11-Sea of Japan, 12-South China Sea, 13-Southeast Asian Pocket basins,14-Arabian Sea,15-Bay of Bengal,16-Somali Current,17-Northern Indian Ocean, 18-Mid-Indian Ocean, 19-Southern Indian Ocean, 20-Agulhas Current, 21-Northwest Atlantic Subarctic,22-North Atlantic Drift, 23-Gulf of Mexico, 24-Central North Atlantic, 25-Mediterranean, 26-Mauritania/Cape Verde, 27-Tropical and West Equatorial Atlantic, 28-Guinea Basin and East Equatorial Atlantic, 29-Benguela Upwelling, 30-South Atlantic, 31-Circumglobal Subtropical Front, 32-Subantarctic waters, 33-Antarctic/Southern Ocean.

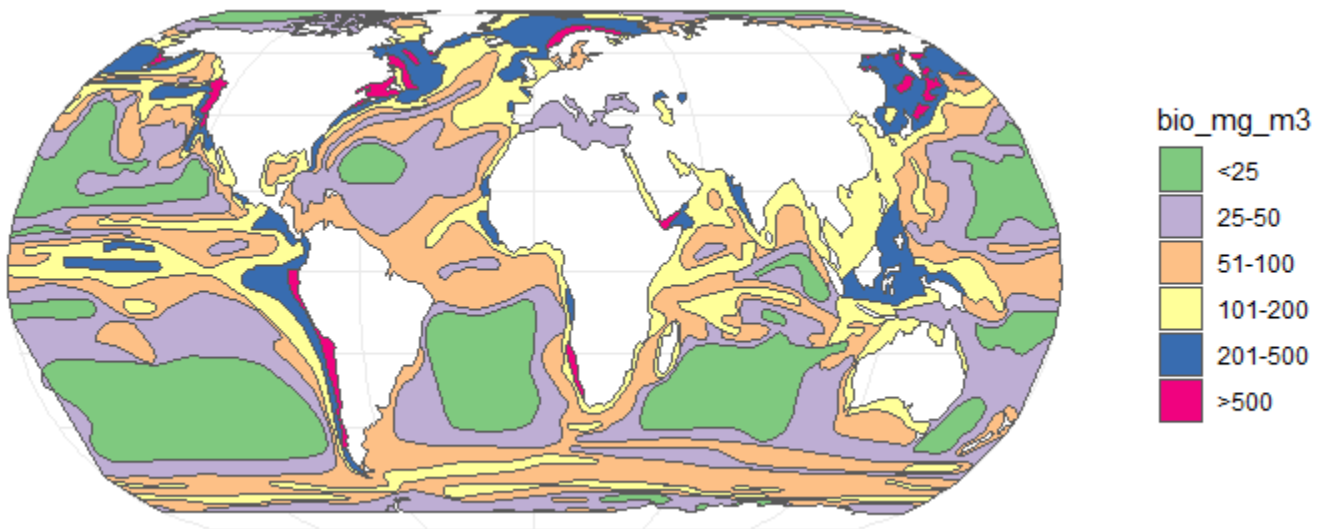


Figure E5: Digitized map of distribution of zooplankton biomass ( $\text{mg}\cdot\text{m}^{-3}$  wet weight) within the surficial layer of the world ocean based on map reported in (Bogorov *et al.*, 1968). The map in Eckert IV global equal area projection.

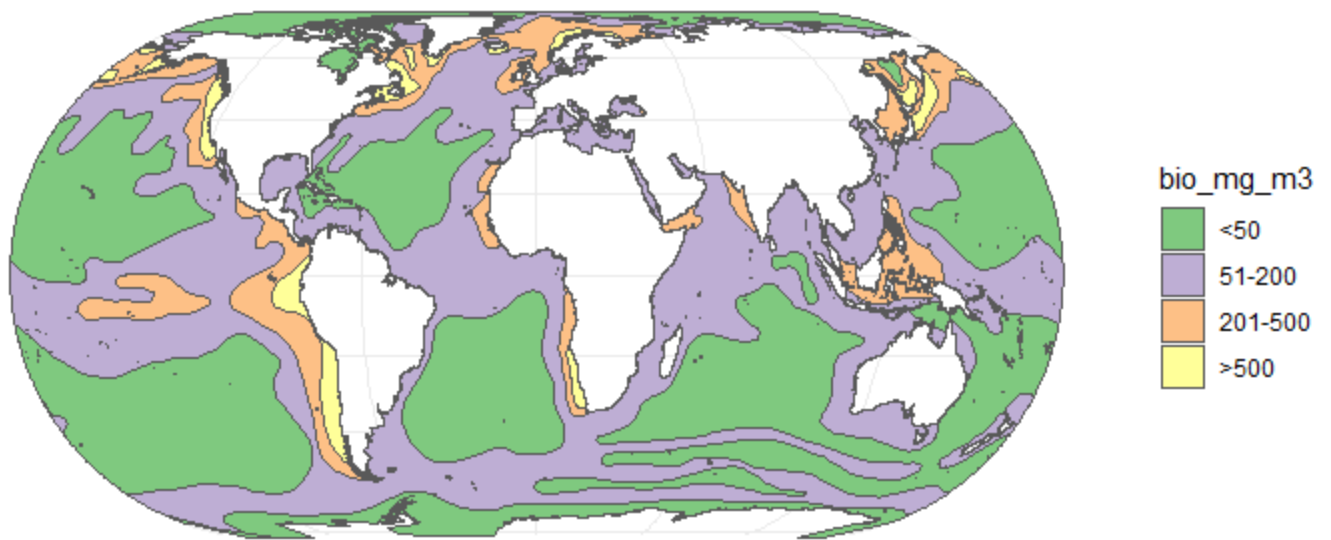


Figure E6: FAO (1972) biomass of zooplankton in the upper 100 m of the water column ( $\text{mg}\cdot\text{m}^{-3}$  wet weight). This map is an updated version of the map created by Bogorov *et al.* (1968). The map in Eckert IV global equal area projection.

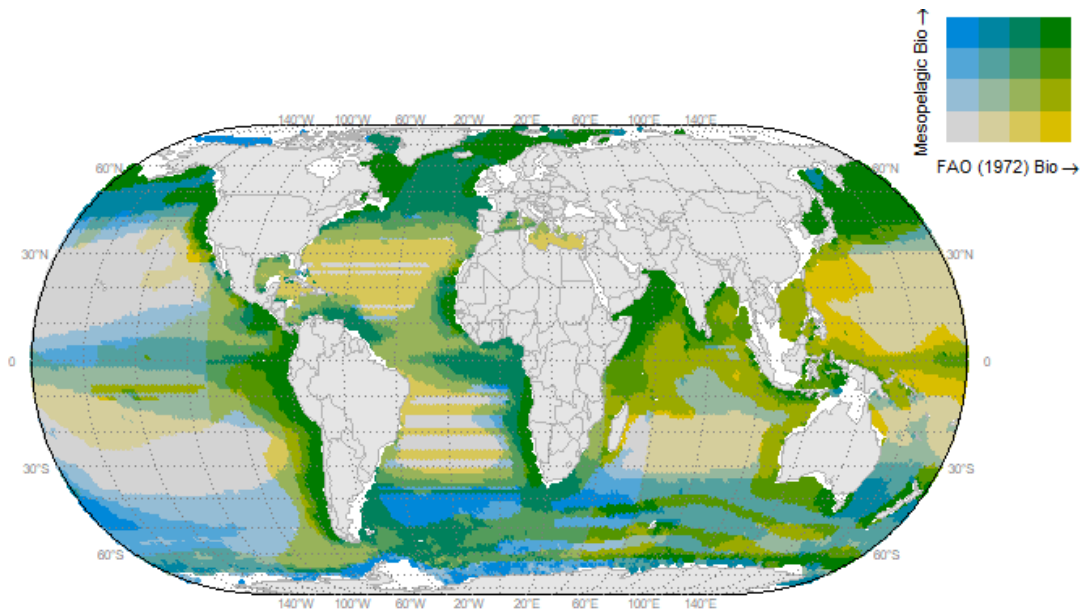


Figure E7: 2D map depicting the relationship between the mesopelagic mesozooplankton biomass derived from POC estimates and FAO (1972) estimated total zooplankton biomass in a top 100 m of the water column. Quantile breaks in a legend were created at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles for mesopelagic biomass (Mesopelagic Bio) and at a <50, 51-200, 201-500, >500  $\text{mg}\cdot\text{m}^{-3}$  wet weight. The map in Eckert IV global equal area projection.

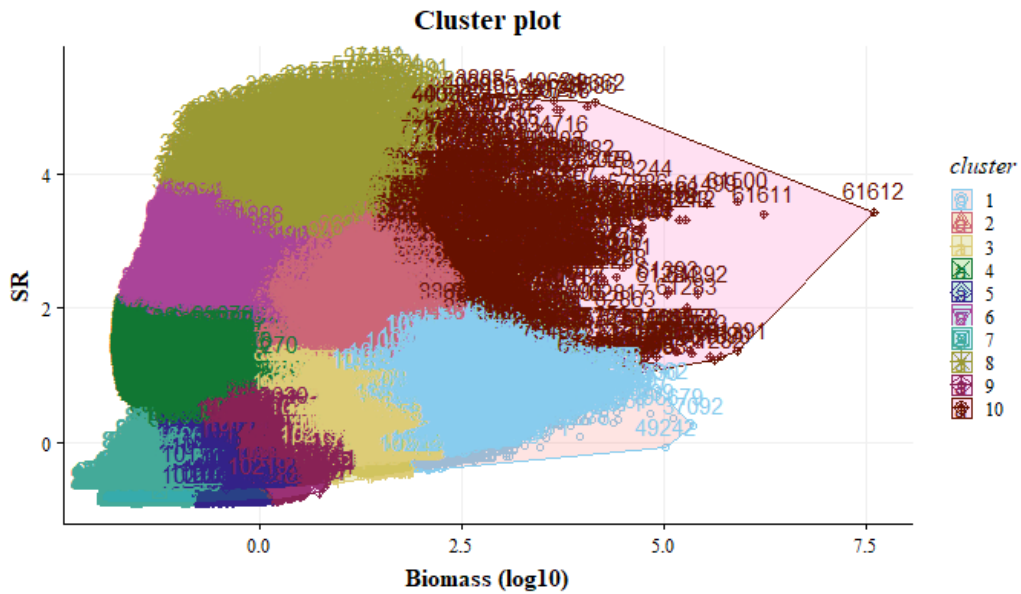


Figure E8: K-means clustering ( $k = 10$ ) of gridded  $\log_{10}$ -transformed biomass (derived from POC) and potential species richness (SR; estimated in Chapter 4).

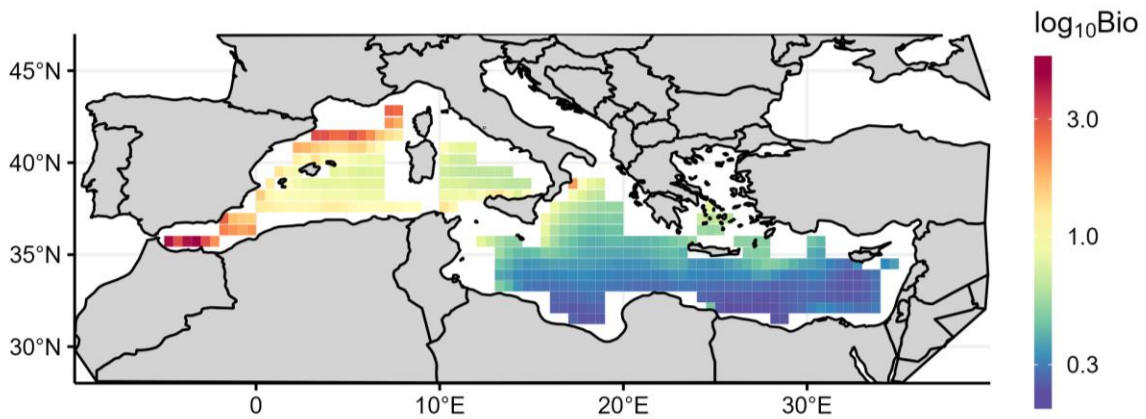


Figure E9: Spatial distribution of mesopelagic mesozooplankton biomass (note  $\log_{10}$  transformation of color scale) in the Mediterranean Sea.

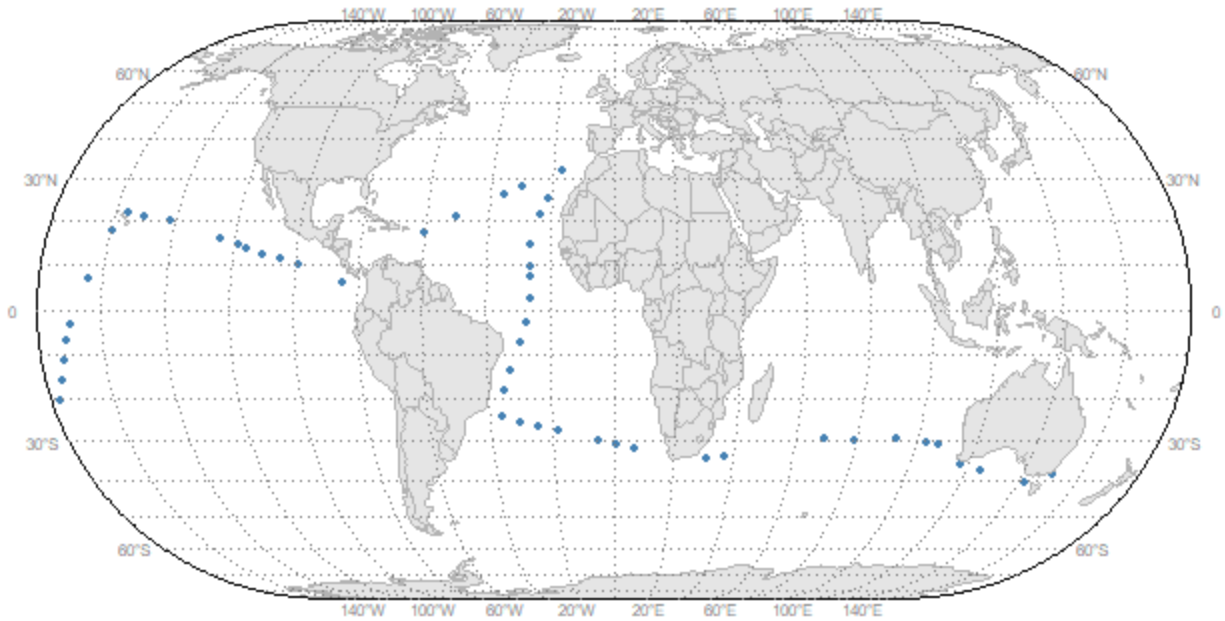


Figure E10: Malaspina 2010 expedition cruise track showing the locations of the 51 stations sampled. The map in Eckert IV global equal area projection.