Plankton community size structure in UK coastal waters: observations and modelling



Greg Macmillan

A thesis presented for the degree of
Doctor of Philosophy
September 2026

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Acknowledgements

My PhD experience was extremely positive thanks to the support of my peers, supervisors, friends, and family. Firstly, I'd like to thank my primary supervisor, Dr Bingzhang Chen, for his input, guidance, and support. Bingzhang always took time to think deeply about ideas and suggestions, and was my first port of call when I needed advice. My secondary supervisor, Dr Dafne Eerkes-Medrano, also deserves great credit for contributing a biological perspective to my study of zooplankton. Without Dafne's expertise, I would have struggled to make progress, particularly in the early stages of my PhD. Both Dafne and Bingzhang facilitated an open, constructive environment at weekly meetings from the very beginning of my PhD. I hope Dafne and Bingzhang take great satisfaction in the impact they have had on my research, and for providing me with such a fantastic experience.

I was fortunate to have the support of the Scottish Universities Partnership for Environmental Research (SUPER), who not only funded my research project but provided a range of opportunities. These included annual retreats, writing retreats, and a supportive community. I am also grateful to my two external supervisors, Prof Michael Burrows and Dr Rory O'Hara Murray, both of whom contributed ideas and a fresh perspective to this research.

In 2023, I took a short break from my PhD to work under the supervision of Dr Abdullah Gök on an interdisciplinary research project with the Hunter Centre for Entrepreneurship. I learned a great deal from Abdullah as a researcher, and about the process of planning, undertaking, and completing a research project. Abdullah was always understanding and flexible with deadlines which allowed me to prioritise my PhD when working on manuscripts after the initial research period. I am grateful to Abdullah, and to Bingzhang who facilitated this collaboration. Bingzhang also organised academic exchange visits to Ocean University of China and Hong Kong University of Science and Technology, both supported by SUPER.

My SUPER funding ended in 2025 and I entered the writing up phase of my PhD, working part-time as a Teaching Associate in Statistics. My work as a Teaching Associate contributed to the success of my PhD and inspired me to push through the final stages of writing. I must thank my line manager, Dr Ryan Stewart, who allowed me to work flexibly in my role, facilitating continued progress in the latter stages of my PhD. Through supervising and teaching, I hope I also improved my own writing and communication skills, for which I have my students to thank.

All that said, this research project would not have been possible without the work of an extended

network of researchers and crew members. There would be no plankton data to analyse without the efforts of those who collected the samples, for which I thank the crew of RV Plymouth Quest, MV Temora, Stella, and the various boats collecting samples in Loch Ewe. Thanks go to Claire Widdicombe, Dr Glen Tarran, Prof Angus Atkinson, Andrea McEvoy, Dr Eileen Bresnan, and Dr Dafne Eerkes-Medrano who provided personal guidance on the use of UK coastal plankton data; thanks also go to all other staff of Plymouth Marine Laboratory and Scottish Government Marine Directorate involved in the process of analysing plankton samples and providing organism size data. I would also like to acknowledge Dr Ricardo González-Gil and the late Dr James Cloern for contributing additional organism size data.

My viva took place in September 2026 and I was fortunate to receive valuable comments and feedback from my examiners, Dr Neil Banas and Dr Ben Ward. Both Neil and Ben are influential researchers in the field of plankton ecology whose modelling advancements continue to be an influence on my research. It was therefore a privilege to discuss my thesis with them.

Finally, I'd like to thank all those who continue to provide me with personal support. Mum, Dad, Paul, Elise, Lewis, and the rest of my family and friends are always on hand when I need them most. And, last but not least, I would like to thank my partner, Aislinn. The last few years have been the most enjoyable period of my life, largely thanks to the experiences we have shared.

Abstract

Plankton underpin marine food webs by facilitating the transfer of energy from primary producers to higher trophic levels. These organisms are incredibly diverse yet, among the many traits that influence ecological function, body size is regarded as a 'master trait' because physiological rates and trophic interactions can be described as an allometric function of size. While the seasonal cycles of biomass and taxonomic composition of the plankton community in UK coastal waters are well understood, comparatively little is known about size structure. This thesis addresses that gap by investigating the seasonal dynamics and drivers of plankton community size structure in UK coastal waters.

Using weekly observational data spanning more than 20 years from three observatories, we assessed plankton community size structure by estimating biomass-size distributions, and subsequently computing summary statistics such as mean size and size diversity. In line with expectations derived from our review of the literature, we found that phytoplankton mean size is generally smaller in winter, and larger during the productive period of summer. On the other hand, zooplankton mean size decreases during summer, due to contributions from unicellular microzooplankton. Size diversity also exhibits a strong seasonal signal, although this is more complex and is perhaps better explained by its parabolic relationship with biomass.

Having assessed seasonal patterns in plankton community size structure, we then scrutinised the validity of these results by assessing possible impacts of sampling bias on our estimates. Previous studies show that rare phytoplankton cells tend to be under-represented by traditional sampling methods and, since abundance decays exponentially with organism size widely in nature, concerns have been raised about the study of size structure using samples analysed by light microscopy. We addressed these concerns directly: firstly, by comparing the enumeration rates of cells by size between paired subsamples of two volumes; then, by conducting a simulated sampling experiment. We detected that bias significantly affects the estimation of biomass, mean size, and size diversity within the study of large phytoplankton. This effect varies seasonally, broadly tracking changes in biomass and size structure. However, our concerns do not extend to the overall study of community size structure in UK coastal waters because bias is most prevalent in winter, when cells tend to be smaller and are thus measured by flow cytometry. As such, winter size structure is largely driven by the extent in which small cells dominate the community, rather than by changes within the subset of larger cells.

Our focus then shifted to investigating the mechanisms driving changes in plankton community

size structure. To do so, we constructed a size-structured plankton ecosystem model and optimised unknown parameters using approximate Bayesian computation. Among a number of interesting results, a key finding is that light limitation through intracellular shading, which increases with cell size yet is often ignored in the modelling literature, plays a crucial role in driving the observed small mean size of phytoplankton in winter. In spring, larger phytoplankton cells are able to temporarily escape grazing, driving the increase in mean cell size. Furthermore, we investigated areas of poor performance of the model to construct hypotheses for future study. Firstly, the model's inability to reproduce the spring rise and subsequent summer depression in zooplankton mean size appears to be a result of deviations from the allometric relationship for optimal prey size by dinoflagellates and meroplankton larvae. Additionally, the omission of mixotrophy from the model appears to drive the erroneous reduction in phytoplankton mean size in summer, particularly given the prominence of mixotrophic taxa in observations at UK coastal sites in summer and early autumn. We outline possible extensions to the model to improve its accuracy in this regard and, based on an extensive assessment of the model, believe that the model is a suitable tool for future development.

Contents

Declaration i

Acknowledgements ii

Abstract iv

List of Figures x

List of Tables xvi

1 The role of organism size in plankton community ecology 1

1.1 Introduction to plankton physiology and ecology . . . 1

1.1.1 Phytoplankton . . . 3

1.1.2 Zooplankton . . . 7

1.1.3 Mixoplankton . . . 11

1.1.4 Allometry . . . 12

1.2 Plankton community ecology . . . 12

1.2.1 Measuring plankton . . . 13

1.2.2 Biomass . . . 15

1.2.3 Size structure . . . 18

1.3 Modelling approaches . . . 22

1.3.1 Basic NPZD model . . . 22

1.3.2 Size-structured models . . . 27

1.3.3 Photo-acclimation and variable stoichiometry . . . 32

1.3.4 Alternative food web representations . . . 35

1.4 Plankton in the UK coastal ecosystem . . . 36

1.4.1 Overview of the UK coastal environment . . . 36

1.4.2 Existing research . . . 41

1.5 Conclusion . . . 52

2 Estimation of plankton community size structure in UK coastal waters 55

2.1 Introduction . . . 55

2.1.1 UK coastal waters: a dynamic ecosystem . . . 55

2.1.2 Predictions about size structure from existing UK coastal research 56
2.1.3 Techniques for assessing size structure 57
2.1.4 Aims and objectives 58
2.2 Methods 58
2.2.1 Data 58
2.2.2 Estimation of plankton biomass-size distributions 63
2.2.3 Analysis of size structure 66
2.3 Results 70
2.3.1 Seasonal cycle of environmental conditions 70
2.3.2 Assessment of biomass distributions by season 76
2.3.3 Statistical modelling of the seasonality of plankton community size structure 83
2.3.4 Relationships between indicators of size structure 89
2.4 Discussion 94
2.4.1 Comparison to predictions from previous research at UK coastal sites . . . 94
2.4.2 Drivers of the seasonal cycle 96
2.4.3 Size structure and productivity 103
2.4.4 Long-term variability 105
2.4.5 Limitations and potential sources of bias 106
2.4.6 Summary of hypotheses 111
2.5 Conclusion 112

3 Impact of subsample volume on estimating phytoplankton community size structure 113
3.1 Introduction 113
3.1.1 Enumerating phytoplankton by inverted microscopy 113
3.1.2 Consequence for measuring taxonomic diversity 114
3.1.3 Statistical basis for large-cell undersampling 114
3.1.4 Aims and objectives 116
3.2 Methods 116
3.2.1 Investigating rates of successful identification across the size domain . . . 116
3.2.2 Simulating the impact of low-volume subsampling at the community level 119
3.3 Results 122
3.3.1 Relationship between enumeration rate and size at SH & LE 122
3.3.2 Community estimator bias from simulation experiments 126
3.4 Discussion 130
3.4.1 Scaling of size-related sampling bias from species level to community level 130
3.4.2 Inference for Chapter 2 results 131
3.4.3 Comparison with existing work 132
3.4.4 Limitations and future work 137
3.4.5 Recommendations 139
3.5 Conclusion 140

4 Construction, testing, and parameter optimisation of a size-structured plankton ecosystem model **141**
4.1 Introduction . . . 141
4.1.1 Modelling approaches . . . 141
4.1.2 Parameter space . . . 142
4.1.3 Aims and objectives . . . 143
4.2 Methods . . . 143
4.2.1 Model selection . . . 143
4.2.2 Model description . . . 144
4.2.3 Operationalisation . . . 151
4.2.4 Parameter optimisation . . . 153
4.2.5 Model stability . . . 158
4.3 Results . . . 160
4.3.1 Parameter optimisation . . . 160
4.3.2 Model outputs at PL . . . 165
4.3.3 Model outputs at SH & LE . . . 170
4.3.4 Model stability . . . 171
4.3.5 Sensitivity to parameters . . . 172
4.4 Discussion . . . 173
4.4.1 Choice of fixed model parameter values . . . 173
4.4.2 Scale and allometry of the initial slope of the photosynthesis-irradiance curve (α) . . . 177
4.4.3 Quadratic mortality . . . 181
4.4.4 Assessment of ABC approach & viable alternatives . . . 182
4.5 Conclusion . . . 184

5 Assessment of mechanistic hypotheses using the size-structured plankton ecosystem model **186**
5.1 Introduction . . . 186
5.1.1 Assessing hypotheses . . . 186
5.1.2 Aims and objectives . . . 189
5.2 Methods . . . 189
5.2.1 Extracting biological rates . . . 189
5.2.2 Decomposition of effects on photosynthesis . . . 191
5.2.3 Model reruns . . . 193
5.3 Results . . . 193
5.3.1 Initial assessment of gross growth and loss rates . . . 193
5.3.2 H2: Winter mean size at PL . . . 196
5.3.3 H3: Mean size of phytoplankton during bloom . . . 201
5.3.4 H6: Decrease of micro- and mesozooplankton mean size in summer . . . 208
5.3.5 H4: Large phytoplankton mean size in summer . . . 210

5.3.6 H5: Maximal phytoplankton size diversity in autumn . . . 214
5.3.7 H1: Lower winter phytoplankton biomass at SH compared to PL . . . 217
5.4 Discussion . . . 221
5.4.1 Key findings . . . 221
5.4.2 Future direction of modelling . . . 223
5.4.3 Summary of decisions on hypotheses . . . 227
5.5 Conclusion . . . 228

Future outlook **230**

Bibliography **234**

Appendices **268**
Appendix A Plankton organism size data . . . 269
Appendix B Diagnostics of fitted statistical models . . . 289
Appendix C Supplementary figures for interannual variability of plankton community size structure . . . 298
Appendix D Supplementary text for the sampling bias simulation experiment . . . 300
D.1 Decomposing sources of bias . . . 300
D.2 Comparison with Chapter 2 results . . . 301
D.3 Original model fitted with GAM . . . 303
Appendix E Supplementary figures for the size-structured plankton ecosystem model 304

List of Figures

2.1 Histogram of organism sizes at SH, LE, and PL from the 'phytoplankton' (inc. microzooplankton) and zooplankton microscopy datasets. Organism sizes can be found in Tables A.1−A.4. 59

2.2 (a) Location and depth of SH, LE, and PL around the coast of the UK (excluding Orkney, Shetland, and Northern Ireland); (b) summary of the available observational data at each site across the time series. SSM and WeStCOMS output are not included. 60

2.3 Typical biomass distributions of unicellular communities during winter, spring, summer, and autumn at SH, LE, and PL. Line colours represent different estimators of central tendency. Data excludes flow cytometry. 69

2.4 Surface nutrient concentrations of dissolved inorganic phosphate (DP), dissolved silicate (DSi), and dissolved nitrate + nitrite + ammonia (DN). LOD is marked by the black line; at SH & LE this is 0.01 μM, and at PL this is 0.02 μM. 70

2.5 Surface nutrient molar ratios, computed against dissolved inorganic phosphate (DIP). These values can be compared to the orange line at 16:1 (Redfield ratio). . 71

2.6 Temperature observed at surface and near-seabed depths. Surface samples at PL are taken at 2 m; at SH & LE they are taken at 1 m. Near-seabed samples are taken at 45 m, 35 m, and 50 m at SH, LE, and PL respectively. 72

2.7 Mixed layer depth estimated at SH, LE, and PL. Data from CTD casts at PL, WeStCOMS model output at LE, and SSM climatology at SH. 73

2.8 Mean density by depth in January and July estimated at SH, LE, and PL. Data from CTD casts at PL, WeStCOMS model output at LE, and the SSM climatology at SH. Shaded region indicates seabed. 74

2.9 Modelled PAR from observations at PL and sites near SH ($\sim$ 28 km) & LE ($\sim$ 31 km). GAMs fitted on spherical coordinates with an interactive term between time of day and day of year. Residual errors shown in Figure B.1. 75

2.10 Total seawater attenuation (k_d) and background seawater attenuation (inclusive of suspended matter; k_w). Concentration-dependent chlorophyll attenuation can be inferred from $k_d - k_w$, and was incorporated in the calculation of k_w. k_d calculated from Secchi depth at SH & LE, and from depth-resolved PAR observations from CTD casts at PL. 76

2.11 Typical biomass distributions of unicellular plankton recorded by microscopy and flow cytometry at SH and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$). 77
2.12 Typical biomass distributions of unicellular communities (excluding flow cytometry data) at SH, LE, and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$). For plots at PL, a zoomed window is provided for larger ESD values to show the biomass distributions over the interval $10 - 130\,\mu$m more clearly. 79
2.13 Typical biomass distributions of multicellular communities at SH, LE, and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$). 81
2.14 Phase portrait showing the typical seasonal cycle of mean size and size variance of phytoplankton measured by microscopy and flow cytometry at SH and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot ($\cdot$). 83
2.15 Phase portrait showing the typical seasonal cycle of mean size and size variance of micro- and mesozooplankton at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot ($\cdot$). 85
2.16 Phase portrait showing the typical seasonal cycle of mean size and size variance of 'microplankton' (diatoms, dinoflagellates, and ciliates) measured by microscopy at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot ($\cdot$). 86
2.17 Phase portrait showing the typical seasonal cycle of mean size and size variance of copepods at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot ($\cdot$). 87
2.18 Seasonal cycle of copepod biomass contribution at SH, LE, and PL by categorised copepodite stages. 88
2.19 Seasonal cycle of copepod biomass contribution at SH, LE, and PL by diet. . . . 89
2.20 Seasonal cycle of mean log-ESD of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined $(2000 - 2019)$ data. 90
2.21 Relationships between biomass and mean size, and biomass and size variance in phytoplankton and zooplankton communities at SH and PL. LE data excluded due to the absence of flow cytometry analysis of small cells. 91

2.22 Relationships of biomass, mean size and size variance between phytoplankton and zooplankton communities at SH and PL. $n = 9$ observations removed from plot in (b), but not model fitting. No significant difference between curves for SH & PL in (b). LE data excluded due to the absence of flow cytometry analysis of small cells. 91
2.23 Skewness-kurtosis pairs plot for biomass distributions at SH & PL. The green shaded elliptical area indicates the region of normality. The outermost dashed line indicates the boundary of the attainable domain for skewness-kurtosis pairs. The region between dashed lines indicates candidate region for bimodality, but only points in the blue shaded region are considered to be bimodal. Some extreme points omitted to allow appropriate plot scaling; the number of displayed points is annotated on each figure. LE data excluded due to the absence of flow cytometry analysis of small cells. 93

3.1 Biomass contributions of observed phytoplankton and microzooplankton taxa at SH and LE from the FOV dataset, and PL. Orange line bounding the shaded area shows the minimum recordable biomass corresponding to 1 mL subsamples. Purple line bounding the lower shaded area gives the corresponding values from a 50 mL subsample. 118
3.2 A comparison of candidate GAM and GAMLSS models assessed using simulation outputs from two subsample volumes. All models have a seasonal smooth (cyclic cubic splines or cubic splines on circular coordinates) fitted to location (expected value), while the GAMLSS model also has a seasonal term and a random intercept for the scale (variance) model. 121
3.3 Rate of enumeration of taxa at SH and LE from a 1 mL subsample, conditional on its presence in a 50 mL subsample or the 1 mL subsample itself, across phytoplankton cell size. 123
3.4 Seasonal variations in rate of enumeration of taxa at SH and LE from a 1 mL subsample, conditional on its presence in a 50 mL subsample or the 1 mL subsample itself, across phytoplankton cell size. 124
3.5 Frequency distributions of (a) taxa by size across FOV and P/A datasets at SH and LE, and (b) associated richness computed on each sample date. 125
3.6 Distribution of phytoplankton 'population' biomass, mean log-ESD, and log-ESD variance at PL. 126
3.7 Relationships between 'population' biomass and estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes. 127
3.8 Relationships between 'population' mean log-ESD and estimation bias of mean log-ESD, and 'population' log-transformed log-ESD variance and estimation bias of log-transformed log-ESD variance, at different subsample volumes. 128
3.9 Seasonal cycle of estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes. 129

4.1 Estimates of posterior distributions of free parameters following ABC optimisation. Acceptance threshold values of $\mathcal{D}$ are denoted by ε. A lower ε corresponds to better model fit. Density plotted using `geom_density()` of `ggplot`. 161
4.2 Accepted pairs of parameters for (a) α_a and α_b; (b) k_{Z1} and k_{Z2}. The grey shaded region shows the region of the uniform prior from which the parameters can be sampled. 163
4.3 Root mean square distance of nondimensionalised model outputs for paired high precision vs. reduced precision model runs across years 15 to 20 of the model. Each line represents one of 80 parameter sets sampled from the prior distribution. Due to the nondimensionalisation, RMSE values can be interpreted approximately as the proportional difference between the paired model runs. 164
4.4 Model outputs for each of the parameter sets accepted with the $\varepsilon_{0.001}$ are shown as grey lines. The best-fitting sample from the posterior which minimises $\mathcal{D}$ is shown in red. Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line. 166
4.5 Comparison between observations and the best-fitting sample from the posterior which minimises $\mathcal{D}$. Truncated mean is used, as in Chapter 2, to calculate the mean biomass density for observations within each size bin. The biomass contribution at ~ 4000 μm in June of plot (b) is driven by rare blooms of *Beroe spp.* (33000 μm). It is therefore included for completeness, but model fit should not be compared to these observations. 167
4.6 Model outputs resulting from the use of forcing from SH, LE, and PL. The parameter set which minimises $\mathcal{D}$ is used. 169
4.7 Evolution of distance between unperturbed and perturbed model trajectories from 60 randomly sampled parameter sets of the estimated posterior, across (a) the first 3 months and (b) final 5 years of the model runs. 171
4.8 Root mean square distance of nondimensionalised model outputs when parameters are adjusted by $\pm 50\%$, compared to the baseline set which minimises $\mathcal{D}$. Grey tiles indicate NAs, where parameters are outwith realistic ranges ($Q_N^{\min}$ cannot be greater than $Q_N^{\max}$; $Q_N^{\max}$ cannot be smaller than $Q_N^{\min}$; no valid solution for a reduction in $G_a^{\max}$ and an increase in $G_{b1}^{\max}$). 173
4.9 P-I curves for three species retrieved by Edwards *et al.* (2015): *Ceratium lineatum* (Chan, 1978); *Chaetoceros simplex* (Thompson *et al.*, 1991); *Skeletonema costatum* (Oh *et al.*, 2008). (a) shows the fit of Geider and Eilers-Peters curves for each species using these data. (b) adjusts the α^* for each curve by assuming $\alpha = 3.83 \times 10^{-7}$ (Ward *et al.*, 2012), and assuming carbon-to-chlorophyll ratio of 40 mg C (mg chl a)$^{-1}$. 178

5.1 Monthly-averaged daily gross growth and loss rates of phytoplankton, and zooplankton of the model using the optimised parameter set. Net growth / loss can be calculated as the difference between gross growth and gross loss. 194

5.2 Mean size of phytoplankton weighted by the distribution of gross growth, gross loss, and grazing-only loss. In (a) the weightings are by rates only, while in (b) rates are multiplied by biomass to yield a weighting by productivity. The rate-weighted log-ESD is the first distributional moment of Figure 5.1a. The productivity-weighted log-ESD is given by multiplying the rates (d^{-1}; Figure 5.1) by biomass of each size class (mmol C m^{-3}; Figure 4.5). 196
5.3 Model outputs for the optimal parameter set and an experiment with $\alpha_b = 0$. For the experiment, $\alpha_a = 0.45 \times 10^{-7}$. Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line. 197
5.4 Ratios of phytoplankton gross growth and loss rates compared to the dominant size of 5.7 μm, for (a) 1.8 μm cells, and (b) 27 μm cells. 199
5.5 Monthly-averaged grazing saturation for each zooplankton size class j of the model, given by $\frac{F_j^2}{k_{Z(1\text{ or }2)}^2 + F_j^2}$. 'Model' line is directly sourced from model, while 'Observations' are computed using observed biomasses of phytoplankton and zooplankton. Both cases used the same universal allometric scaling. 202
5.6 Monthly-average biomass distributions of phytoplankton and zooplankton of the model using the optimised parameter set with passive switching between herbivory and carnivory (blue), and without passive switching (red). 205
5.7 Monthly-averaged distribution of carnivory coefficients ($\varphi^{(Z)}$) by size. The herbivory can be yielded as $\varphi^{(P)} = 1 - \varphi^{(Z)}$. 'Model' line is directly sourced from model, while 'Observations' are computed using observed biomasses of phytoplankton and zooplankton. Both cases used the same universal allometric scaling. 207
5.8 Biomass-weighted mean size (E[s]) for micro- and mesozooplankton (excluding heterotrophic nanoeukaryotes) from GAM fitted to observations (blue points / line) and model outputs (red line). 208
5.9 Seasonal progression of nitrogen uptake rate, per unit of carbon biomass (V_N; mmol N (mmol C)$^{-1}$ d^{-1}) for three dominant size classes of the model. 212
5.10 Comparison of photosynthetic rate ($V_{C,i}$) by size, using different combination of forcing for light and temperature at SH and PL. Light forcing is inclusive of surface irradiance, background seawater attenuation, and MLD. 217
5.11 Comparison of modelled P:Z biomass ratios across November to January using forcing from SH, LE, and PL. Note a 360 day year is used by the model. 220

B.1 Residual errors against fitted values of model fitted to PAR SH, LE, and PL (Figure 2.9). 289

C.1 Seasonal cycle of biomass of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined $(2000 - 2019)$ data. 299
C.2 Seasonal cycle of log-ESD variance of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined (2000 − 2019) data. 299

D.1 Partial effect of population biomass on estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes. Here, the model is fitted with random intercepts for both location and scale (rather than just scale), which gives rise to the difference to Figure 3.7. 301
D.2 An adaptation of Figure 2.16. Phase portrait showing the typical seasonal cycle of mean size and size variance of 'microplankton' (diatoms, dinoflagellates, and ciliates) measured by microscopy at SH, LE, and PL. At PL, 1 mL simulated subsample is depicted by solid line and original data by transparent line, demonstrating the potential shift in estimation driven by bias arising from low-volume subsamples. 302
D.3 The equivalent of Figure 3.7 when a GAM is fitted rather than a GAMLSS model. The extent of model fitting bias is demonstrated at a seasonal scale in Figure 3.2. 303

E.1 Outputs from the neutral model for months 1, 4, and 7. All allometric exponents are set to zero, and all coefficients are equal across all size classes (e.g. $k_{Z1} = k_{Z2}$). To facilitate a neutral model, zooplankton grazing was herbivorous-only (since carnivorous grazing is only on some size classes), and all zooplankton size classes were subject to quadratic losses. This therefore replaced grazing losses for protozooplankton. 305
E.2 Seasonal cycle of Q_{Chl} for three dominant phytoplankton size classes of the model. These values can be transformed from their units of mg Chl a (mmol C)$^{-1}$ to mg C (mg Chl a)$^{-1}$ by computing $12/Q_{Chl}$. The equivalent values here range from 20 mg C (mg Chl a)$^{-1}$ to 52 mg C (mg Chl a)$^{-1}$. 305
E.3 Monthly-averaged nutrient limitation of photosynthesis (λ_N) by phytoplankton size. 306
E.4 Monthly-averaged proportion of gross losses contributed by size-independent sources across phytoplankton sizes. Size-independent losses arise from linear mortality and mixing, while size-dependent losses are from grazing. 306
E.5 Monthly-averaged biomass distributions of phytoplankton and zooplankton of the model using the parameter set which minimises $\mathcal{D}$ (and thus has $\alpha_b = -0.136$), and an experiment where $\alpha_b = 0$ with $\alpha_a = 0.45 \times 10^{-7}$. 307
E.6 Model outputs for the optimal parameter set with passive switching for herbivory and carnivory (black line), and an experiment with no switching (red line). Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line. 308
E.7 Seasonal cycle of realised photosynthetic rate (V_C) and saturating photosynthetic rate (V_C^{sat}) of three dominant size classes of the model. 308

List of Tables

1.1 Qualitative summary of allometric tradeoffs for key biological rates. Source: [a]Marañón *et al.* (2013), [b]Irwin *et al.* (2006), [c]Ward & Follows (2016), [d]Acevedo-Trejos *et al.* (2022), [e]Edwards *et al.* (2015), [f]Litchman *et al.* (2007), [g]Eppley *et al.* (1969), [h]Edwards *et al.* (2012), [i]Hansen *et al.* (1997), [j]Kiørboe (2008), [k]Hansen *et al.* (1994), [l]Banas (2011), [m]Finkel *et al.* (2004), [n]West *et al.* (1997), [o]Brown *et al.* (2004), [p]Chen & Liu (2010). 12

2.1 Resolution of microscopic measurement corresponding to ESD ranges. Entries above the dashed line show values extracted from Schartau et al. (2010); below this shows extrapolated values. 64
2.2 Parameter values of the carbon to volume relationship for phytoplankton (Menden-Deuer & Lessard, 2000) Adjusted parameter values due to discontinuity are marked by *. 65
2.3 Parameter values of the carbon to volume relationship for zooplankton (Harris, 1999). 66
2.4 Number of bootstraps, depending on the number of observed species in a subsample. 67

3.1 Probabilities of non-zero counts in subsamples as a result of varying population abundance, under the assumption that counts are a Poisson point process. 115
3.2 Decomposition of R^2_{CS} for selected GAMLSS models for the case $V = 0.001$. Here, the model for scale (σ_Y) is always saturated, and the model for location (μ_Y) is incrementally expanded from intercept-only to intercept and smooth. In the intercept-only model for location, all pseudo-R^2 is contributed by the scale model (σ_Y). All model terms have associated $p < 2 \times 10^{-16}$ (Table B.3). 126

4.1 Size-independent parameters of the model. Prior distributions for free parameters are presented in Table 4.3; posterior estimates in Table 4.4. Source: [a]Gan *et al.* (2010); [b]Dutkiewicz *et al.* (2012); [c]Marañón *et al.* (2013); [d]Eppley (1972); [e]Dam & Peterson (1988); [f]Geider *et al.* (1998); [g]Chen (pers. comm.); [h]Ward *et al.* (2012). *λ_Z is calculated under the assumption zooplankton N:C ratio is always balanced. 146

4.2 Size-dependent parameters of the model. All relationships are of the form $y = aV^b$. Prior distributions for free parameters are presented in Table 4.3; posterior estimates in Table 4.4. Source: [a]Marañón *et al.* (2013); [b]Menden-Deuer & Lessard (2000); [c]Litchman *et al.* (2007); [d]Hansen *et al.* (1997); [e]Hansen *et al.* (1994) via Banas (2011). *coefficient adjusted for continuity for model stability. **although Hansen *et al.* (1997) propose -0.16, the evidence is weaker. 147
4.3 Limits for uniform priors ($\pi(\cdot)$) of free parameters. *initially 1×10^{-6}, but reduced after poor fitting in a pilot test of 15,625 runs. 155
4.4 Minima, medians, maxima, and bounds of 95% Bayesian credible intervals for each of the free parameters for the $\varepsilon_{0.001}$ cutoff. The parameter set which minimises $\mathcal{D}$ is denoted $\mathcal{D}_{\min}$ 161
4.5 Correlation matrix for the estimated joint posterior distribution of free parameters when $\varepsilon_{0.001}$ is used as a cut-off. Diagonal values (always equal to 1) are replaced with '-' for ease of reading. 162

5.1 Photosynthetic parameters and performance across three dominant phytoplankton size classes, averaged across December and January when light is most limiting. The term $\frac{\alpha^*}{V_C^{\text{sat}}}$ governs photosynthetic saturation, for any given I. The values α, α^*, and $\frac{\alpha^*}{V_C^{\text{sat}}}$ are all given in units of 10^7 for brevity. 198
5.2 Fixed effects model estimates for $V_{C,12} \sim \lambda_{N,12} * Q_{Chl,12} + V_{N,12}$, decomposing effects of intracellular nutrient limitation, intracellular chlorophyll, and nutrient uptake (which incurs a penalty on carbon accumulation). $V_{C,12}$ (d^{-1}) is averaged $V_{C,i}$ for size class $i = 12$ ($27\,\mu\text{m}$) across July irradiance and temperature. The baseline level is where winter values of λ_N, Q_{Chl}, and V_N are used. 211
5.3 Fixed effects model estimates for $\bar{V}_C \sim I_0 * k_w * H$, decomposing effects of surface PAR, seawater attenuation (excluding that of chl *a*), and MLD. $\bar{V}_C$ (d^{-1}) is the unweighted average V_C across size class. $\bar{V}_C$ is then averaged across values December and January for 6-hourly resolved I_0, k_w, and H. The baseline level is where all irradiance forcing is used for PL, with temperature forcing from SH in all cases. 219

A.1 Sizes of phytoplankton and protozooplankton for PL, extracted directly from Widdicombe & Harbour (2021) and converted to ESD. Raw data included here for completeness; all dinoflagellates were considered either mixoplankton (both autotrophs and mixotrophs), or zooplankton in our analyses. The data are included under the Western Channel Observatory data policy, which should be consulted by all users. 269
A.2 Sizes of phytoplankton and protozooplankton for SH & LE, primarily compiled by González-Gil *et al.* (2022) and reduced for presentation in this thesis; the full table can be found in the associated supplementary material. Sources Cloern (2018) and Widdicombe & Harbour (2021) were added to the original table of González-Gil *et al.* (2022) as necessary. 275

A.3 Sizes of metazooplankton for SH & LE compiled by Scottish Government, Marine Directorate (2024) and reduced for presentation in this thesis. Converted to ESD from dry weight using the conversion in Table 2.3 (Harris, 1999). 279
A.4 Sizes of metazooplankton by season extracted directly from McEvoy *et al.* (2022). The data are included under the Western Channel Observatory data policy, which should be consulted by all users. 284

B.1 Summaries of GAMs fitted to various response variables, with a common explanatory cyclic smooth term of day of year. 290
B.2 Summaries of LMs fitted to indicators of plankton community size structure. Each model has $n = 568$ observations, those from PL spanning 2007-04-23 to 2007-12-18 and at SH 2015-01-19 to 2018-12-05. 295
B.3 Summaries of GAMLSS models fitted to estimates of bias. p-values for variances not shown; these were always $p < 2 \times 10^{-16}$. 296

Chapter 1

The role of organism size in plankton community ecology

1.1 Introduction to plankton physiology and ecology

Marine plankton are a set of taxonomically and functionally diverse organisms which drift with the ocean's currents (Louchart *et al.*, 2023; Acevedo-Trejos *et al.*, 2022; De Vargas *et al.*, 2015). The name is derived from the Greek word *planao* (to wander), and refers to their inability to actively move against the currents of the ocean (Brierley, 2017). Plankton cover a huge range of sizes − over 12 orders of magnitude in volume (V; Sieburth *et al.*, 1978; Beardall *et al.*, 2009; Finkel *et al.*, 2010) − and as such, are typically represented in equivalent spherical diameter (ESD, denoted x), which is given by:

$$x = \left(\frac{6V}{\pi}\right)^{\frac{1}{3}} \tag{1.1}$$

On this scale, plankton are typically classified as pico- ($0.2-2\,\mu$m ESD), nano- ($2-20\,\mu$m ESD), micro- ($20-200\,\mu$m ESD), meso- ($200-2000\,\mu$m ESD), or macroplankton ($2000-20000\,\mu$m ESD; Ward & Follows, 2016). For brevity, any measurement with units 'μm' is assumed to refer to ESD herein, unless otherwise stated.

Among their many services to the ecosystem, plankton form the lowest levels of the marine food web (Armengol *et al.*, 2019; Fenchel, 1988; Meddeb *et al.*, 2019; Di Pane *et al.*, 2024). In the first instance, phytoplankton (from Greek *phyton*, 'plant'; Pal & Choudhury, 2014) are defined in our case as autotrophic, unicellular plankton in the form of algae or bacteria. Marine phytoplankton are the dominant primary producers in world oceans, and are responsible for about 30−70 Gt C year^{-1} (Carr *et al.*, 2006; Friedrichs *et al.*, 2009; Carvalho *et al.*, 2017), equivalent to $\sim$ 50% of annual productivity, despite constituting less than 1% of the world's active biomass (Field *et al.*, 1998). This process contributed to the absorption of $\sim$ 30% of anthropogenic atmospheric CO_2 by oceans between 1850 and 1996 (Raven & Falkowski, 1999).

Zooplankton (from Greek *zoon*, 'animal'; Luo, 2013) are heterotrophic organisms which dominate the consumption of phytoplankton. The transfer of energy from phytoplankton to microzooplankton is estimated at ~ 30 Gt C year^{-1}, which corresponds to around 62% of global primary consumption (Schmoker *et al.*, 2013); a further ~ 5.5 Gt C year^{-1} of phytoplankton is consumed by mesozooplankton (Calbet, 2001). In turn, zooplankton provide nutrition for planktivorous organisms such as fish in the next trophic level and so a food web begins (Giraldo *et al.*, 2024; Mitra *et al.*, 2014; Zeldis & Décima, 2020).

Zooplankton can be uni- or multicellular and, although unicellular zooplankton (protozooplankton) are historically less studied than multicellular zooplankton (Liu *et al.*, 2021), they have an important role in marine food webs by primarily grazing on phytoplankton (Schmoker *et al.*, 2013; Buitenhuis *et al.*, 2010; Chen *et al.*, 2012). Multicellular zooplankton (metazooplankton) occupy the largest sizes in plankton (typically $> 200\,\mu$m in adult lifestages) and are morphologically diverse, ranging from 'shrimp-like' krill to jellyfish (Perhirin *et al.*, 2024).

Further complexity is added to the food web by mixotrophs, who are able to use both auto- and heterotrophy as a strategy to transfer energy (Stoecker *et al.*, 2017). These unicellular organisms are nano- and micro- sized, and contribute significant biomass in global ecosystems (Flynn *et al.*, 2013; Leles *et al.*, 2017).

Another key group of plankton is heterotrophic bacteria, which have a critical role in the process of recycling nitrogen and carbon by remineralising small organic particles (Wiebinga *et al.*, 1997; Mayor *et al.*, 2014; Buesseler *et al.*, 2007). Most of this activity is conducted at the lower boundary of the photic zone of the ocean, in the so-called twilight zone (Kong *et al.*, 2021). However, despite the fact these bacteria are heterotrophic plankton, they are not considered zooplankton because of their position in the food web.

The final group of plankton are viruses. These are extremely small (typically $0.002 - 0.2\,\mu$m; Suttle *et al.*, 1990), yet can have a significant influence on plankton population dynamics and nutrient cycling through plankton mortality (Fuhrman, 1999; Wilhelm & Suttle, 1999; Wommack & Colwell, 2000). Although they may infect all living organisms in the ocean from bacteria to whales (Suttle, 2005), they predominantly target small phytoplankton, particularly bacteria and prasinophytes (small green algae; Wommack & Colwell, 2000; Weynberg *et al.*, 2017). However, laboratory experiments show that adding viruses to cultures of larger cells, such as diatoms, can also lead to substantial mortality (Suttle *et al.*, 1990).

Within each of the outlined categories of plankton, great taxonomic diversity is to be found. Yet, much of an organism's interaction with the environment and food web can be explained by its size, which is considered a 'master trait' (Litchman *et al.*, 2007; Hillebrand *et al.*, 2022b). Key ecological characteristics, such as maximal growth rate and nutrient uptake half-saturation constants, can be described by an allometric relationship. A trait y is therefore a function of volume V, governed by the relationship:

$$y = aV^b \iff \log y = \log a + b \log V \tag{1.2}$$

where a is the scaling parameter (with the same units as y), and b is the allometric exponent. A trait will decrease with volume if $b < 0$ and increase for $b > 0$, making the allometric exponent essential in structuring marine ecosystems.

1.1.1 Phytoplankton

Marine phytoplankton consist of many taxonomic groups. At the smallest end of the size domain ($< 2\,\mu$m), many of the species are prokaryotic. Of particular importance globally are cyanobacteria of the genera *Prochlorococcus* ($\sim 0.5\,\mu$m) and *Synechococcus* ($\sim 1.5\,\mu$m). *Prochlorococcus* spp. are most common in warm oceanic regions, while *Synechococcus* spp. are more common in coastal and temperate regions (Zwirglmaier *et al.*, 2008; Armengol *et al.*, 2019). Picoeukaryotes are also important contributors of biomass in productive oceans, in particular small prasinophytes ($\sim 1\,\mu$m) which are chlorophytes (green algae).

Moving along the size domain to the small nanoplankton ($2 - 6\,\mu$m), phytoplankton biomass contributions become more dominated by eukaryotes. Nanoeukaryotes including, but not limited to larger chlorophytes, cryptophytes, chrysophytes (golden algae), and haptophytes (including coccolithophores) are important (Piwosz, 2019; Schulz *et al.*, 2017). Each of these have one or more flagella, making them motile on a local scale. A special case is the coccolithophores, most of which have a flagellated lifestage and a non-flagellated lifestage, during which the cells construct additional calcite plates (coccoliths). A reasonable assumption would be that this is a defence mechanism, however, copepod grazing appears unchanged on coccolithophores with coccoliths under laboratory conditions (Sikes & Wilbur, 1982). In this size interval, two other crucial groups become represented at the lower end of their size limit: diatoms and dinoflagellates.

Diatoms are perhaps the most recognisable and well-researched of the phytoplankton, in part due to their size (historically identifiable under the microscope), and because of their intricate and beautiful silica frustules. The frustule is composed of two overlapping thecae and serves as protective armour against predation from higher trophic levels (Xu *et al.*, 2024; Friedrichs *et al.*, 2013), and perhaps further to optimise photosynthesis under high-light conditions (Ellegaard *et al.*, 2016). Diatoms are described as centric (radially symmetric) or pennate (bilaterally symmetric), and possess the ability to form long chains as a further means to avoid predation (Toullec *et al.*, 2019). Most species exist in the $2 - 200\,\mu$m range, although some exceptionally large species exist (e.g. *Coscinodiscus wailesii*, $275\,\mu$m; Table A.1).

Dinoflagellates, which also primarily exist on $2 - 200\,\mu$m size range, are named after their unique helical swimming pattern which results from the positioning of their two flagella (Fenchel, 2001). Many of the flagellated taxa have complex nutritional strategies, employing both autotrophy and heterotrophy (further described in Section 1.1.3). Traditionally, most dinoflagellates are classified as either phytoplankton or zooplankton, but many species are best described as mixoplankton. In a functional ecology context, dinoflagellates require consideration on a species-by-species basis. Some dinoflagellates are thecate, possessing a wall of cellulose plates for protection, while others are athecate (naked).

1.1.1.1 Reproduction

Phytoplankton require both nutrients and light for cell growth and subsequent reproduction (Sigman & Hain, 2012; Charalampous *et al.*, 2018). Cells primarily reproduce asexually by dividing in two or more cells, depending on taxonomy (Zhao *et al.*, 2022; Kottmeier *et al.*, 2020). This process yields smaller cells than the original cell and as such, some cells are required to periodically reproduce sexually in order to restore cell size (Speijer *et al.*, 2015; Montresor *et al.*, 2016). Laboratory experiments on the diatom *Pseudo-nitzschia pungens* found that sexual reproduction rates were highest during the exponential growth phase ($\sim 7\%$ of cell pairs reproducing sexually), and acted as a way to restore cell size during population growth (Kim *et al.*, 2023). Conversely, dinoflagellates appear to use sexual reproduction as a growth strategy under nutrient-deficient and other growth-limiting conditions (Nagai *et al.*, 2004; Kalinina *et al.*, 2023). It follows that the range of conditions in which cells can grow varies between taxonomy and by cell size. In most cases, when nutrients and light are not limiting, diatoms emerge as the fastest growing cells, owing to their cell physiology (Richardson *et al.*, 1983). Maximal growth rates of diatoms are, for example, $\sim 1.9\times$ higher than other eukaryotes for a given size (Irwin *et al.*, 2006). Within phytoplankton functional groups, maximal growth rate is typically a decreasing function of size (Irwin *et al.*, 2006; displayed graphically in Ward *et al.*, 2012). However, due to the fact that different functional groups occupy different regions of the size domain, a unimodal curve centred around $\sim 6\mu$m can be used to describe maximal growth rate when averaging across different functional groups (Chen & Liu, 2010; Wirtz, 2011; Marañón *et al.*, 2013; Ward & Follows, 2016; Hillebrand *et al.*, 2022a).

1.1.1.2 Photosynthesis

Central to a phytoplankton cell's ability to grow is its ability to fix carbon through photosynthesis. Cells use the energy from solar radiation to convert environmental carbon dioxide and water into internal carbohydrates, lipids, and proteins, releasing oxygen in the process (Jónasdóttir, 2019). The photosynthesis-irradiance (P-I) response is primarily driven by photosynthetically active radiation (PAR). For an individual species, photosynthesis is typically an increasing function of PAR until an optimum value of PAR, and decreasing thereafter due to photoinhibition (Platt *et al.*, 1980; Eilers & Peeters, 1988; Edwards *et al.*, 2015). However, at a community level, the response to light is more complex to investigate and the relationship in natural communities has not been explicitly proven. In Edwards *et al.* (2015), the optimum irradiance covers almost two orders of magnitude for species with similar taxonomy, maximum growth rates, and initial slope of the P-I response. This implies that succession between (or within) species can allow a shift from low-light adapted to high-light adapted cells, such that the community-averaged P-I relationship is not necessarily subject to photoinhibition. The relationship between maximum growth rate and optimum irradiance is weak ($r = 0.2$) but apparently linear; if a species was assumed to only exist in a community under optimum irradiance, then photoinhibition would not be a community level effect, at least across the range of naturally occurring irradiance. Furthermore, the concept that within-species succession may negate a photoinhibition penalty is supported

explicitly in MacIntyre *et al.* (2002) for both cyanobacteria and eukaryotes. Intraspecific variation in *Synechococcus* sp. results in a saturating response to light, while the diatom *Skeletonema costatum* exhibits photoinhibition following adaptation to a low-light regime, but exhibits a saturating response when adapted to a high-light regime. Indeed, it is noted that many samples in the seminal paper Platt *et al.* (1980) are taken from deep water, where cells are more likely to be low-light adapted. As such, at a community level, the P-I curve is currently justifiably described as a saturating response (e.g., Geider *et al.*, 1997; Anderson *et al.*, 2015). Performance under low-light regimes varies widely both between and within functional groups, and across the size domain. The initial slope of the photosynthesis-irradiance response curve decreases with cell size as a result of intracellular chloroplast shading, such that small cells are likely to out-compete large cells in light-limited conditions, although observations of this relationship exhibited significant unexplained noise (Edwards *et al.*, 2015).

1.1.1.3 Nutrient assimilation

To balance the accumulation of carbon in the cell, phytoplankton must also assimilate nutrients from the environment. The four main elemental nutrients which limit phytoplankton growth are nitrogen, phosphorous, silicon, and iron (Bristow *et al.*, 2017). In particular, nitrogen is considered a key limiting nutrient in both open ocean and coastal environments (Ryther & Dunstan, 1971; Li *et al.*, 2015), making it a natural focus in observational and modelling works. Phytoplankton primarily assimilate N through environmental sources of nitrate (NO_3) and ammonium (NH_4; Gruber, 2008; Malerba *et al.*, 2015). High levels of NH_4 can inhibit phytoplankton growth (Dutkiewicz *et al.*, 2012), although NH_4 has a key role in the ocean's nitrogen cycle as it recycled to NO_2 and then to NO_3 by nitrifying bacteria. Phosphorous is required by phytoplankton, particularly for nucleic acids (both DNA and RNA), and thus plays a key role in genome duplication (Lin *et al.*, 2016). Iron is also required by phytoplankton although is typically only a limiting nutrient in oceanic regions close to or at the equator (Li *et al.*, 2015; Schartau *et al.*, 2010). Diatoms also require silicate (SiO_4) for cell growth to maintain the frustule, creating an additional dependency to be fulfilled before growth is exhibited.

Uptake of nutrients is typically expressed as a saturating Michaelis-Menten function of the form $V_N^{\max} \cdot N/(k_N + N)$, where k_N is the half-saturation constant for a nutrient N. As a result of the tradeoff between surface area-to-volume ratio with increasing cell size, which has an exponent of $-\frac{1}{3}$, k_N is an increasing function of cell size (Eppley *et al.*, 1969; Litchman *et al.*, 2007); as such, large cells have lower nutrient affinity than small cells. Additionally, the maximal (saturating) nutrient uptake ($V_N^{\max}$) rate scales negatively with cell size, although different magnitudes for the exponent are presented in the literature (Litchman *et al.*, 2007; Marañón *et al.*, 2013). This contrasts the maximal photosynthetic rate, which is unimodal.

1.1.1.4 Stoichiometry

The ratio of intracellular nutrient composition is an important factor which drives (or limits) nutrient uptake and photosynthesis (Droop, 1968; Flynn, 2008). Historically, the ratio of in-

tracellular C:N:P(:Si) was believed to be fixed in proportion, following a molar nutrient ratio of 106:16:1(:16) for plankton (Redfield *et al.*, 1963). This ratio can still be used as a rough approximation, but neglects the fact that significant fluctuations can be made from this ratio, and that the varying quota of internal nutrients influences phytoplankton growth (Flynn, 2008). The minimum and maximum limits of a cell's stoichiometry is related to its size, with greater flexibility possible for larger cells per unit of carbon mass (Edwards *et al.*, 2012; Marañón *et al.*, 2013). In small cells, a larger proportion of volume is required for essential components, leaving less space for storage of lipids and carbohydrates (Raven, 1994).

1.1.1.5 Respiration

Construction, maintenance and repair of the biological structures allowing reproduction, nutrient assimilation and photosynthesis incur a metabolic expense through respiration. Respiration in phytoplankton per unit body mass is comparatively high compared to other, larger organisms under the scaling of the Metabolic Theory of Ecology (MTE) which has allometric exponent -0.25 (West *et al.*, 1997; Brown *et al.*, 2004). However, there is some debate about whether the scaling exponent is true to -0.25 *within* the phytoplankton, with estimated exponents of -0.15 (López-Urrutia *et al.*, 2006), or even closer to isometry (Finkel *et al.*, 2004; Marañón *et al.*, 2013), indicating that the size-based tradeoff may not be as severe as proposed by the MTE.

1.1.1.6 Temperature effects

Each of these described biological processes have a maximal rate which is moderated by temperature. Within species temperature curves are often unimodal, such that a species will have an optimal temperature (thermal niche), with performance often reducing drastically when temperature exceeds the optimal value. Observations of optimal temperature scale with mean environmental temperature (Thomas *et al.*, 2012). However, when viewed through a community perspective, performance curves are better described by an exponential curve with a positive exponent (Eppley, 1972). In this sense, temperature can be considered a critical factor when monitoring and modelling individual species as it can tightly control their existence, whereas through a size-based lens, temperature more acts to moderate and adjust biological rates, with the fastest rate occurring at high temperatures.

1.1.1.7 Mortality

Phytoplankton growth is tightly coupled with the environment. Low light, nutrients, or temperature can diminish the community growth rate. Size-based trade-offs have a role to play in structuring the community, given that allometric exponents for saturating uptake and growth rates are negative (for $> 6\,\mu$m in the case of maximal photosynthesis); additionally, exponents for the initial slopes of uptake and growth are also negative, indicating that small cells are likely to outperform large cells under limiting conditions. Small cells appear to have higher gross growth rates under high-resource and low-resource conditions, so it follows that under these bottom-up controls on growth, a reasonable expectation would be that small cells (e.g., $< 10\mu$m) always

outcompete larger cells. However, the existence and prevalence of larger cells implies that they must have a competitive advantage under certain conditions (Finkel *et al.*, 2010). Therefore, the top-down control on the phytoplankton community is a key process in controlling biomass and structure (Rodríguez-Gálvez *et al.*, 2023). Among defence mechanisms such as a protective frustule or toxic chemicals, empirical observations show that size is positively related to the evasion of zooplankton grazing at a community level (Pancic & Kiørboe, 2018; Branco *et al.*, 2020), through increased individual cell size (Friedrichs *et al.*, 2013) or the ability to aggregate into colonies of many cells (Long *et al.*, 2007). In a broad sense, this pattern arises because larger phytoplankton are grazed by larger zooplankton (Hansen *et al.*, 1994), which have lower maximal grazing rates than small zooplankton (Hansen *et al.*, 1997). However, a laboratory study of the green algae species *Chlorella vulgaris* by Yoshida *et al.* (2004) found that the grazed community had evolved smaller size than the un-grazed community. This perhaps supports the fact that there is a complex relationship between predator diversity, prey diversity, and the evolution of size (García-Comas *et al.*, 2016).

1.1.2 Zooplankton

Zooplankton play a key role in the marine food web by facilitating the transfer of energy from phytoplankton to the higher trophic levels. Given that zooplankton are heterotrophic, constraints on zooplankton growth arise from the availability, capturing, and handling of food.

1.1.2.1 Grazing

In the first instance, grazing interactions are largely governed by size. Except in rare cases, such as grazing through fragmentation of aggregated detritus (Mayor *et al.*, 2014) or diatom chains (Toullec *et al.*, 2019), zooplankton predators are considered to be equal to or larger than their prey. The form of this relationship is debated: Kiørboe (2008) consider a fixed optimal predator-prey size ratio of 10:1 to prevail, whereas Hansen *et al.* (1994) argues that a positive relationship between optimal prey size and predator size exists (later parametrised by Banas, 2011), with the exception that large heterotrophic dinoflagellates eat comparatively larger prey, closer to a 1-to-1 size ratio (Hansen, 1992). Under both frameworks, pico- and small nanophytoplankton can generally be expected to be grazed by microzooplankton such as heterotrophic dinoflagellates and ciliates, and larger nano- and microplankton are grazed by mesozooplankton including (but not limited to) copepods, fish / bivalve larvae, and jellyfish. As such, zooplankton size structure is inherently influenced by the size structure of the prey distribution, since grazing efficiency by a zooplankton organism is expected to be optimised when its prey is optimally sized.

Zooplankton maximal grazing rates (per unit of biomass) are a decreasing function of organism size overall (Hansen *et al.*, 1997), but different strategies exhibited within neighbourhoods of the size domain mean that a trade-offs within each region of the size domain exist, with only a weak dependence between strategy and size (Kiørboe, 2011). Historically, strategy was believed to be driven by a tradeoff between the ability to first obtain resources, and then process them. This aligns with the scaling of k_Z and $G^{\max}$ (Table 1.1) which indicates that mesozooplankton typically

have lower half-saturation (and thus higher affinity), while nano- and small microzooplankton typically have higher maximal rates (Hansen *et al.*, 1997). However, more recently it is argued that zooplankton which excel at grazing in deplete prey conditions also excel in replete conditions (Kiørboe & Thomas, 2020). As such, the tradeoff is that these organisms also suffer greater predation risk (Verdolin, 2006).

1.1.2.2 Protozooplankton

The smallest component of the zooplankton size domain is dominated by heterotrophic nanoflagellates, dinoflagellates, and ciliates. These groups use a variety of ambush feeding strategies which vary between taxa (Hansen & Calado, 1999), and are specific to unicellular organisms. These are (i) direct engulfment (phagocytosis; e.g. Schnepf & Elbrachter, 1992), (ii) tube feeding, where prey matter is sucked through a feeding tube (myzocytosis; e.g. Schnepf & Deichgraber, 1984), and (iii) external ingestion (pallium feeding; e.g. Hansen, 1991). Across dinoflagellate taxa, thecate species generally use pallium feeding because the theca restricts the feasible prey size through phagotrophy to small bacteria, while naked cells can use phagotrophy to ingest large prey (Jacobson & Anderson, 1986), even up to their own size in the case of *Gyrodinium spirale* (Hansen, 1992). Consumption of organic food particles has been shown to facilitate biomass accumulation through both direct cell growth ($> 15\%$ increase in volume for *Oxyrrhis marina* and $\sim 7\%$ increase for *Gyrodinium dominans* over a 3-hour feeding period; Anderson & Menden-Deuer, 2017), and reproduction through asexual cell division and periodic sexual reproduction for both growth and survival (Gribble, 2006; similar to their autotrophic dinoflagellate counterparts). The result of these grazing methods and reproductive strategies is that protozooplankton can grow very fast. The maximal grazing rates among zooplankton are by nanoflagellates, in line with their small cell size; while dinoflagellates and ciliates share similar, larger cell size, ciliates have comparatively higher maximal grazing rates than dinoflagellates (Hansen *et al.*, 1997).

1.1.2.3 Metazooplankton

Generally speaking, on the domain of $100 - 200\,\mu$m, there is a shift from unicellular to multicellular organisms. This incurs significant changes in feeding modes, survival strategies, and reproduction. Morphologically, multicellular zooplankton can be non-gelatinous (primarily crustaceans), semi-gelatinous (such as chaetognaths and fish / bivalve larvae), or gelatinous (such as jellyfish and salps). Perhaps the most important tradeoff here is that non-gelatinous plankton are the most nutritionally dense among the multicellular zooplankton, while the gelatinous plankton have low nutrition (per unit size): approximately 9% of crustacean wet mass is carbon, compared to only 0.4% for gelatinous plankton (Harris, 1999). This key characteristic shapes the predatory and defensive mechanisms for key functional groups of multicellular zooplankton.

Of the crustacean plankton, copepods are considered a key taxonomic group in many marine environments – particularly in coastal (Peterson & Miller, 1977; Wells *et al.*, 2022) and temperate / subpolar oceanic waters (Melle *et al.*, 2014; Mackas & Tsuda, 1999). On a planktonic scale copepods are highly motile, allowing them to feed by one of the following methods: (i) rapidly

pounce on prey in the case of an ambush encounter; (ii) by generating a feeding current to suck in small prey, typically less than 10 μm (Berggreen *et al.*, 1988); or (iii) actively seek larger aggregates using a cruise strategy, albeit typically feeding on particles for a limited time, perhaps as a result of the increased mortality risk from larger organisms who may eat the particles whole while the copepod is feeding (speculated by Kiørboe, 2011). The speed of ambush predator-prey interactions is highlighted by Jiang & Paffenhöfer (2008), who showed that the copepod *Oithona plumifera* had an average attack speed of 55 mm s^{-1} on the ciliate *Strobilidium spp.*, despite only covering a distance of 0.8 mm, resulting in an average attack time of ~ 14 ms. On the basis that different feeding modes are optimised for differing prey types and sizes, it is reasonable to expect that ecological trade-offs between organism size, feeding mode, and prey availability are likely to impact the size-dependent grazing rate within the copepod community. Additionally, many copepods are omnivorous and can supplement their prey through the consumption of eggs and larvae (Landry, 1981; Bonnet & Carlotti, 2001), which may compensate for shortages in phytoplankton prey.

Each surviving copepod experiences six naupliar stages (N1 $-$ 6) and six copepodite stages (C1 $-$ 6), with sexual maturation and reproduction occurring in C6. Copepods grow throughout the lifecycle, with nauplii often $< 200\,\mu$m and adults $200 - 2000\,\mu$m, although this varies widely between taxa, and some larger species exist in polar regions. The entire life cycle lasts between a few weeks and a few years, depending on species (Banas *et al.*, 2016); for example, the small estuarine copepod *Eurytemora affinis* may reach sexual maturation in ~ 10 days at 20°C / ~ 23 days at 10°C (Ban, 1994), while the large arctic copepod *Calanus hyperboreus* may take $2 - 3$ years to mature (Brennan *et al.*, 2024). To maximise their survival rates, copepods are known to follow a daily cycle of vertical migration, moving to deeper waters during the day to avoid predation during the day from visually cruise feeding fish, and accessing phytoplankton prey at night in the euphotic zone (Brierley, 2014; Hobbs *et al.*, 2021). During periods of low prey availability such as in winter, many species enter diapause, where they migrate to deeper waters to effectively hibernate, using internal lipid stores for their reduced respiration rates (Hobbs, 2020).

Other crustacean zooplankton have an important role in global oceans. These include, but are not limited to amphipods and euphausiids (krill), which constitute the second-most and most consumed animals in polar regions, respectively (Dauby *et al.*, 2003). Amphipods are a diverse group of organisms with widespread importance, which can be herbivores or carnivores, but are predominantly detritivores or scavengers (Ritter & Bourne, 2024). Amphipods have both walking and swimming legs, allowing some meroplanktonic species to live in burrows on the seabed and in intertidal zones (Burd *et al.*, 2008). Additionally, many species can tolerate a wide range of salinity, making them widespread in estuarine and coastal environments (Koop & Grieshaber, 2000); here, their size is typically $1500 - 2000\,\mu$m (McEvoy *et al.*, 2022). Krill, on the other hand, are ubiquitous with the temperate and subpolar pelagic ocean, although some coastal species also exist. Primarily, krill are herbivorous, feeding on nano- and microplankton using filter feeding (Kawaguchi *et al.*, 1999), but can also exercise carnivory, including predation

on small copepods and larvae (Meyer *et al.*, 2009). Coastal krill nauplii may be around 500 μm, while adults are around 2000 μm (McEvoy *et al.*, 2022); Antarctic krill can approach 20000 μm (Candy & Kawaguchi, 2006).

Among the gelatinous plankton, the hydrozoans and scyphozoans (true jellyfish) are important taxonomic groups, with the latter contributing significant biomass in coastal regions, albeit with large interannual variation (Eriksen *et al.*, 2012) and spatial variation (Lynam *et al.*, 2005). Most jellyfish are meroplankton with a complex life cycle, with a planktonic medusa stage with sexual reproduction, and a benthic polyp stage with asexual reproduction. This strategy trades-off food availability with mortality risk: in the polyp stage, prey encounters can be rare but is associated with low mortality and respiration, while the opposite may be true in the medusa stage, under certain oceanographic conditions (Lucas, 2001). Planktonic dispersion also facilitates genetic exchange and colonisation (Bashevkin *et al.*, 2020). In both instances, jellyfish are considered to have a passive grazing approach, either through passive ambush or current feeding, on a wide range of particles (Kiørboe, 2011). Organisms including (but not limited to) phytoplankton, microzooplankton, copepod nauplii, and fish larvae (Pengpeng *et al.*, 2021) are ingested after becoming stuck to tentacles. Other gelatinous plankton, such as salps, play an important role in marine ecosystems. Like jellyfish, salps switch between asexual and sexual reproduction through their life cycle, but are planktonic in both stages. Salps are also generalist grazers which filter feed using a mucus net, even consuming particles as small as 1 μm bacteria (Pakhomov *et al.*, 2019).

The final, broad morphological category of plankton are the semi-gelatinous plankton. Much of the biomass in this category is contributed by larvae of fish and meroplankton. This represents an important link in the food web between the previously described zooplankton with the higher trophic levels, since fish and meroplankton larvae are planktonic as juveniles, before becoming benthic or nekton (for fish). Fish larvae, hatched from self-sustaining eggs, are largely carnivorous cruise feeders (Kiørboe, 2011), using their vision to feed selectively on microzooplankton and copepod nauplii (e.g., shown for *Maurolicus parvipinnis* by Landaeta *et al.*, 2011). This subsequently allows fish to grow and subsequently transfer energy up the food web. Based on fish larvae's dietary preferences of large micro- and mesoplankton organisms, it is unsurprising that fish larvae typically belong to the macroplankton with an approximate ESD of 3 mm (McEvoy *et al.*, 2022), although this varies widely between species. Meroplankton larvae are often $200 - 1000\,\mu$m and many taxa, such as bivalve larvae, employ filter feeding to extract phytoplankton from the water column (Lebour, 1933). Maximal grazing rates of meroplanktonic larvae are considerably lower than for copepods at equivalent size (Hansen *et al.*, 1997).

1.1.2.4 Excretion and evacuation

Grazing is necessary for zooplankton growth and survival, but is an inefficient process. Energy is lost in each trophic interaction due to: (i) ingestion inefficiencies, (ii) assimilation inefficiencies, and (iii) metabolic requirements. In the first instance, 'sloppy grazing' can arise when the predator is not able to ingest the entire prey due to fragmentation or damage (e.g., *Acartia*

tonsa exhibited a $\sim 15\%$ loss rate when grazing on small prey, versus $\sim 30\%$ loss on larger prey; Møller, 2007), or associated losses in the stomach may arise due to indigestible components of the prey, such as diatom frustules (Turner, 2002). This can account for a significant loss, and is exacerbated when prey are not optimally sized. Then, once digestible material is inside the stomach, losses through assimilation occur. As is the case for phytoplankton, zooplankton require to maintain a balance of nutrients which is driven primarily by the nutrition and abundance of the available prey (Conover, 1966). Finally, each organism has a basal metabolic rate to be satisfied, and also must use energy in the prey encounter, capture, and processing. As such, a classic baseline for the amount of energy typically obtained from food is around $50-70\%$ at each step of the food web (Conover, 1966), but only around $5-22\%$ is converted to actual biomass growth on average, due to metabolic requirements (Ikedai & Motoda, 1978).

1.1.3 Mixoplankton

1.1.3.1 Physiology

Some plankton are mixotrophic, meaning that they can use both autotrophy and heterotrophy to facilitate growth. This characteristic is widespread in nano- and microplankton across low-latitude, high-latitude, and coastal environments (Moorthi *et al.*, 2009; Edwards, 2019), and is almost entirely exercised by unicellular organisms such as cryptophytes, prymnesiophytes, nano- and dinoflagellates, and ciliates (graphical representation given in Figure 2 of Flynn *et al.*, 2013). How this is executed varies amongst taxa, but organisms can be categorised morphologically as either constitutive or non-constitutive mixotrophs (Mitra *et al.*, 2016). Constitutive mixotrophs produce and maintain both phagotrophic and photosynthetic apparatus in the cells, whereas non-constitutive cells use phagotrophy to obtain and host chloroplasts from other cells, in turn used for photosynthesis. Most nano- and dinoflagellates are constitutive mixotrophs ("phyto-plankton that eat") whereas ciliates tend to be non-constituitive mixotrophs ("zooplankton that are photosynthetic"; Stoecker *et al.*, 2017).

1.1.3.2 Strategy

From a strategic point of view, how much each mixotrophic organisms uses photosynthesis or phagotrophy varies widely between species. Obligate mixotrophs require both phototrophy and phagotrophy to grow and survive, while facultative mixotrophs can rely on one trophic strategy and complement it with the other. Facultative mixotrophs may be (i) obligate phototrophs with facultative phagotrophy, (ii) obligate phagotrophs with facultative phototrophy, or (iii) 'true' facultative mixotrophs, who can use either strategy flexibly (summarised in Table 1.1 of Schoonhoven, 2000). For example, *Mesodinium rubrum* is an obligate heterotroph which is non-constitutive of chloroplasts, as it can sustain photosynthetic growth for ~ 6 weeks and survival for ~ 11 weeks in the absence of their cryptophyte prey in the laboratory (Myung *et al.*, 2013), but beyond this point, the cell requires more food.

On the basis that strict auto- and heterotrophs exist and contribute significant biomass in many systems, it is clear that trade-offs are associated with the flexibility which is afforded

Table 1.1: Qualitative summary of allometric tradeoffs for key biological rates. Source: [a]Marañón *et al.* (2013), [b]Irwin *et al.* (2006), [c]Ward & Follows (2016), [d]Acevedo-Trejos *et al.* (2022), [e]Edwards *et al.* (2015), [f]Litchman *et al.* (2007), [g]Eppley *et al.* (1969), [h]Edwards *et al.* (2012), [i]Hansen *et al.* (1997), [j]Kiørboe (2008), [k]Hansen *et al.* (1994), [l]Banas (2011), [m]Finkel *et al.* (2004), [n]West *et al.* (1997), [o]Brown *et al.* (2004), [p]Chen & Liu (2010).

Tradeoff	Allometric exponent (b)
Max. photosynthetic rate	+ve for $V < 300\,\mu\text{m}^3$, −ve for $V > 40\,\mu\text{m}^3$ [abcdp]
Initial slope of P-I curve	−ve [e]
Max. N uptake rate	−ve [af]
Nutrient uptake half-sat. constant	+ve [fg]
Intracellular N:C quota range	+ve [ah]
Max. ingestion rate	−ve [i]
Grazing half-sat. constant	N/A, but lower for metazoans than protozoans [i]
Predator:prey optimal size ratio	0 [j] or +ve [kl]
Respiration	0 [am] or −ve [no]

by mixotrophy. Namely, cell space for apparatus is limited as a proportion of overall size, and so phagotrophic capability for a constitutive mixotroph may not be advantageous in a high-nutrient, high-light environment, since strict autotrophs can allocate more cell space to photosynthetic apparatus and nutrient uptake sites than mixotroph counterparts. This type of trade-off means that mixoplankton often have a generalist trophic strategy compared to the specialist strategies of phyto- and zooplankton (Edwards, 2019).

1.1.4 Allometry

The importance of size as an indicator of plankton ecosystem functioning is exhibited through growth rates, affinity, and by governing feeding interactions. In phytoplankton, allometric tradeoffs for biomass accumulation are complex: maximum nutrient uptake is a decreasing function of cell size, while the maximum photosynthetic rate per cell is unimodal. Affinity for nutrients and light, on the other hand, appears to be a decreasing function of cell size. However, zooplankton maximal grazing rates are a decreasing function of organism size under optimal prey size distributions. As such, predation pressure on smaller phytoplankton cells is expected to be greater than on larger cells. Yet, this may be further adjusted depending on the prey size distribution, with grazing becoming increasingly inefficient as predator-prey paris deviate from the optimal size ratio. A qualitative summary of allometry in plankton is given in Table 1.1.

1.2 Plankton community ecology

Here we describe patterns in community ecology, with their drivers. We consider these patterns at a global scale for context, but focus on temperate oceans specifically, to begin building a broad picture of the UK coastal ecosystem.

1.2.1 Measuring plankton

Given that each individual planktonic organism has a tiny impact on its ecosystem, community level output and structure are crucial indicators of the importance and functioning of plankton. To quantify how many plankton, and of what type are in the ocean, there are two main avenues of measurement: *in situ* and remote sensing.

1.2.1.1 Bottle sampling

In situ sampling and subsequent measurement is the simplest and most historically widespread method used to enumerate plankton. These methods revolve around a central concept of capturing organisms from the water and analysing them in a laboratory. The method of analysis depends on the size of the organisms in question. Phytoplankton and microzooplankton samples are normally preserved in Lugol's iodine, and subsequently analysed by a taxonomist using an inverted microscope. Cells are typically recorded by species so that species counts are measured along a transect or within a random field of view. Sizes are almost always allocated by taxon rather than by cell meaning that intraspecific size diversity is omitted (e.g. *Karenia brevis* cell size ranging $14-28\,\mu$m; Buskey & Hyatt, 2006). Cells larger than around 4μm can be enumerated efficiently using the method of Utermöhl (1931), however this is labour intensive and is typically only done for a small subsample ($1 - 50$mL), raising concerns about the ability of the method to accurately capture the whole community (Rodríguez-Ramos *et al.*, 2014), in particular large cells due to their inherently low abundance (Goldman, 1993).

A technological advancement which may help mitigate against these sampling biases is the use of imaging flow cytometers (e.g. FlowCAM). This method analyses cells by passing them individually through a flow cell. Images are taken so that size, fluorescence, and taxon may be estimated by software. Images can also be reviewed by the user, making the technology semi-automatic, although accuracy and precision of taxonomic estimation are fast improving through the implementation of artificial intelligence (Decrop *et al.*, 2025). FlowCAM allows for substantially faster measurement of cells, if cells are to be measured on an individual basis (Buskey & Hyatt, 2006); as mentioned above, cells are rarely measured individually by taxonomists under inverted microscopy due to the resource expense. Flow cytometry is also used to enumerate, and measure the individual size of, small cells ($< 5\,\mu$m) at the genus or functional group level.

Samples of smaller cells (around $2 - 10\,\mu$m) can also be analysed manually, using epifluorescence microscopy to identify auto- and heterotrophic cells, and subsequently enumerate them. Epifluorescence-based approaches emerged for enumerating plankton in the 1980s (Veldhuis & Kraay, 2000), but currently lack the capability to provide concise taxonomic information. Outputs from each of these methods yield cell counts which are converted to abundance by dividing by the subsample volume, and then subsequently converted to biomass using a carbon-per-cell ratio based on their size and taxonomy (e.g., Menden-Deuer & Lessard, 2000; Marañón *et al.*, 2013). Normally, individual cells are not measured through epifluorescence or light microscopy, so a fixed size must be allocated on a taxonomic basis, neglecting intraspecific variations in size and related traits (Rillo *et al.*, 2020; Orizar & Lewandowska, 2022; Hamer *et al.*, 2022; Hillebrand

et al., 2022b).

Measuring the chlorophyll concentration directly, as opposed to using it to identify individual cells, is also widely used as a technique to estimate phytoplankton biomass. Fluorometry is a widespread and accessible approach which can be used to accurately quantify the chlorophyll concentration (Holm-Hansen *et al.*, 1965), whereas high performance liquid chromatography (HPLC) can be used to extract individual pigments from the sample to gain understanding about the composition of the community (Mantoura & Llewellyn, 1983). HPLC allows for improved insight, but neither method provides exact estimation of cell sizes or taxonomy, although cells can be filtered incrementally (e.g., using 200 μm, 20 μm, and 2 μm filters) to obtain a size-fractionated estimate of chlorophyll *a*.

1.2.1.2 Remote sensing

Measurement of chlorophyll *a* is arguably most useful when executed remotely from satellites, which can estimate chlorophyll *a* remotely. The technology was pioneered by NASA Coastal Zone Color Scanner in 1978, and is now further developed and used for widespread sensing. This is a powerful tool since it facilitates global estimation of chlorophyll *a* at unprecedented large spatial scales. The main drawback of direct chlorophyll *a* measurement is that conversion to carbon relies on an assumed carbon-to-chlorophyll (C:Chl) ratio, since carbon biomass is a more useful measure of community size and energy in an ecological context. Variations in C:Chl are historically well-established in the literature; at least as early as Strickland (1960) who suggested that C:Chl of 30:1 should be used as an estimate in nutrient rich waters and 60:1 in nutrient-poor waters. Field observations across the global ocean (excluding polar and coastal sites, where values can be lower) have exhibited ratios between 31:1 to 358:1 (Graff *et al.*, 2015). Cells also adjust their C:Chl seasonally to maximise growth (Geider, 1987; Behrenfeld *et al.*, 2005). In the temperate coastal seas around Denmark, C:Chl increases from 20:1 in midwinter to 60:1 in midsummer (Jakobsen & Markager, 2016). As such, chlorophyll-based biomass estimates are easy to obtain in the case of fluoresence, and available on a wide scale in the case of remote sensing, but require some informed assumptions about the C:Chl ratio and lack insight into community size structure.

1.2.1.3 Net sampling

Zooplankton samples are often collected using net hauls. A common approach is for research vessels to make paired vertical hauls using nets of fixed mesh size. The most common size of mesh is 200 μm allowing capture of meso- and macroplankton, while 64 μm nets are also used commonly in order to capture large microzooplankton including protists and copepod nauplii, but can be too fine to be used for larger plankton as they can clog easily (Harris, 1999). Thus, a combination of both nets is required to obtain a full sample of the zooplankton size domain (albeit omitting the smallest microzooplankton $< 64\,\mu$m, which can be measured using the phytoplankton methodologies). Opportunistic data is collected by other vessels in the Continuous Plankton Recorder (CPR) Survey of the Marine Biological Association. CPR surveys use 200 μm

mesh nets and, as such, there exists a gap in the measured size structure for organisms $64-200\,\mu$m ESD. Once captured, subsamples are drawn from the sample so that organisms can be identified by taxa (and life stage, if possible) and enumerated. Traditionally, this is done by a taxonomist, although modern methods including ZooSCAN use machine learning methods to semi-automate the tanonomic identification. FlowCAM can also now be used similarly, despite being developed originally for microplankton by Sieracki *et al.* (1998). Typically, measurements of size / weight are then allocated to species based on literature values, or can be made directly (albeit time-consuming). These can be converted to biomass using conversion factors based on morphology (e.g., Harris, 1999).

1.2.2 Biomass

Biomass is a key measurement of the plankton community because it infers the energy available to higher trophic levels and export capability. Given that phytoplankton growth is constrained by light, nutrients, and temperature, patterns of phytoplankton biomass are broadly explained by these controls across global oceans, since these resources are regulated by local and global-scale oceanographic processes (Behrenfeld *et al.*, 2006). Broadly speaking, annually averaged biomass concentration is high in temperate, polar, and coastal upwelling regions ($> 20\,\mu$g C L^{-1}), moderate in equatorial ocean $\sim 10\,\mu$g C L^{-1}, and low in subtropic regions ($< 5\,\mu$g C L^{-1}; Ward *et al.*, 2012; Siegel *et al.*, 2014). In the latter, biomass is low and shows comparatively little variability, driven by year-round nutrient shortage due to high uptake pressure and low replenishment from nutrient-rich, deeper waters resulting from almost permanent stratification (McClain *et al.*, 2004). However, phytoplankton biomass is also regulated by mortality, particularly zooplankton grazing, and the coupling between predators and prey has significant consequence for food web dynamics. In fact, the equatorial seas are highly productive, despite having a lower standing stock of biomass compared to the temperate regions (Behrenfeld *et al.*, 2006; Siegel *et al.*, 2014), indicating effective primary consumption by zooplankton. This reflects the tight coupling and long food web exhibited in these regions, demonstrated by the higher average trophic position of metazooplankton (Stock & Dunne, 2010; Negrete-García *et al.*, 2022). Generally speaking, the zooplankton biomass is well-correlated with phytoplankton biomass in global oceans, but with slope less than 1 (Irigoien *et al.*, 2004). As such, the P:Z biomass ratio typically increases as community biomass increases (Gasol *et al.*, 1997). Remarkably, this pattern is observed widely in ecology, with a predator-to-prey biomass ratio exponent of 3/4 common across zooplankton, lions, hyenas, tigers, wolves, and their corresponding prey (Hatton *et al.*, 2015).

The biomass cycle in higher latitude oceans is somewhat different to those in the subtropic and equatorial oceans due to the strong seasonality in environmental conditions. Winter is often characterised by cold temperatures, low light, and storm-induced mixing (replenishing nutrients), meanwhile the summer sees higher temperatures, more light, and in some cases stratification through lack of mixing. As such, plankton biomass is lower in winter and higher in summer, showing greater variability than equatorial and subtropical systems (e.g., Siegel *et al.*, 2014; Anderson *et al.*, 2015; Acevedo-Trejos *et al.*, 2015). A key feature of this seasonal cycle which

pertains across these temperate regions is the spring bloom, whereby biomass accumulation is rapid and non-linear. The main constituents of the spring bloom are often larger cells, primarily small-to-mid size diatoms due to their high maximal growth rates (Mills *et al.*, 1994; Brunet *et al.*, 1996; Chen & Rynearson, 2016).

The driving mechanism of the spring bloom is still debated in the literature (Chiswell *et al.*, 2015). It originated with the critical depth hypothesis (CDH), which states that blooms are initiated when the mixed layer depth (MLD) becomes shallower than a critical value such that depth-averaged light facilitates overall phytoplankton growth (Sverdrup, 1953). However, this is problematic because biomass is seen to accumulate before the shoaling of the ML and incident radiation at its lowest during winter (Townsend *et al.*, 1992; Ellertsen, 1993; González-Gil *et al.*, 2022). In light of the findings of Townsend *et al.* (1992) and Ellertsen (1993), the critical turbulence hypothesis (CTH) was proposed by Huisman *et al.* (1999a). Using a plankton ecosystem model which was resolved in both time and vertical space (i.e., 1 dimension), it was shown that the extent of vertical turbulent diffusion could drive biomass accumulation, even if the ML was not shoaling. Huisman *et al.* (1999a) suggested that the CTH was a second possible mechanism to trigger blooms, alongside the CDH.

In opposition to both the CDH and CTH is the disturbance-recovery hypothesis (DRH), first introduced in Behrenfeld (2010) and later updated in Behrenfeld & Boss (2014). This is a top-down perspective, which states that deepening mixed layer depth in the winter facilitates net phytoplankton growth, since predation losses are minimised (Behrenfeld & Boss, 2014). However, it is unclear whether these mechanisms would allow the DRH to drive the bloom at comparatively shallow coastal sites, because winter mixing may occur through the entire water column (González-Gil *et al.*, 2022), and thus the predator concentration may not be sufficiently diluted. Indeed, Mignot *et al.* (2018) anticipated that net growth may not be exhibited during winter in coastal sites due to the comparative absence of dilution. González-Gil *et al.* (2022) used observational data at Stonehaven, one of the UK coastal sites studied in this thesis, to dispute this claim and show that the bloom initiates on 6th January on average, with low interannual variability. While the mechanisms driving the winter initiation of the bloom are not tested directly, González-Gil *et al.* (2022) propose that the DRH may still hold even in the absence of a defined MLD. The physical oceanography criteria underpinning the DRH may therefore be more relaxed, with temporal rate of change in PAR instead explaining around half of the observed variability in net growth rate. This version of the DRH is perhaps more bottom-up driven than the original mechanisms laid out of Behrenfeld & Boss (2014), although the linear relationship between net growth rate and the rate of change of light (and thus growth growth rate, since photosynthesis responds linearly to light at low irradiance; Edwards *et al.*, 2015) suggests that the top-down controls remain relaxed. Additionally, more research may be necessary to clarify the role of mixing and dilution in coastal sites, since the MLD is not directly investigated in González-Gil *et al.* (2022) due to the absence of vertically-resolved physical data, and instead the perceived lack of stratification in winter is determined by comparison of surface and seabed temperature.

The spring bloom is important as it is typically the first source of new organic material for higher levels of the marine food web, including zooplankton and their predators (p. 195, Simpson & Sharples, 2012; James *et al.*, 2003). A combination of nutrient exhaustion, sinking losses through aggregation, and zooplankton grazing is likely to control and cease the spring phytoplankton bloom (Banse, 1994; Behrenfeld & Boss, 2014). During the spring bloom, the grazer community is often initially dominated by fast-responding microzooplankton and later by copepods (Fransz & Gieskes, 1984; Haraguchi *et al.*, 2018), which in turn provide food for fish larvae who respond to the zooplankton prey availability (James *et al.*, 2003). In coastal communities, cirripede (barnacle) larvae also appear during the Spring bloom in large numbers (Eloire *et al.*, 2010; Highfield *et al.*, 2010) to feed on phytoplankton.

In open ocean temperate and polar regions, post-bloom phytoplankton biomass often declines to moderate levels. For example, in the open ocean of the North Atlantic ($45-50°$N, $25-35°$W), Behrenfeld (2010) show that 8-day rolling biomass during the peak bloom approaches $40\,\mu$g C L^{-1} in May, and decreases log-linearly thereafter to $20\,\mu$g C L^{-1} in August; the wintertime low is around $6\,\mu$g C L^{-1}. There is significant variability across this oceanic region presented by Behrenfeld (2010), although all sites display the same qualitative pattern to a large degree. In some coastal and upwelling regions, such as at Stonehaven in the Scottish coastal North Sea, the biomass can remain at or near bloom-levels of $100\,\mu$g C L^{-1} (González-Gil *et al.*, 2022). This is better reflected in carbon-based biomass estimates than chlorophyll, as reflected by the increasing C:Chl ratio from spring to summer (Jakobsen & Markager, 2016). In some estuarine communities such as the San Francisco Bay, however, the biomass is comparatively low after the bloom, and remains as such from May until December (Cloern, 1982; note these measurements are in chl but cover an order of magnitude and therefore deviate too significantly from likely C:Chl ratios to pose the argument that C biomass remains high). The retention of biomass post-bloom in some coastal systems is likely driven by enhanced nutrient availability. For example, the year-round availability of inorganic ammonium and seasonal availability of dissolved organic nitrogen forms has been shown to influence phytoplankton biomass in the Scottish sea loch, Loch Creran (Moschonas *et al.*, 2017). Here, the persistent concentration of $0.5\,\mu$N ammonium contributes little of the inorganic nitrogen forms in winter (which is dominated by nitrate), but dominates the inorganic nitrogen in summer when nitrate becomes exhausted. In shelf and coastal seas, ammonium is an important source of N during the summer (p. 197, Simpson & Sharples, 2012).

A species succession in the photosynthetic plankton is often observed in summer, with other cells such as dinoflagellates entering the fold in temperate coastal and open oceans (Raine, 2002; Widdicombe *et al.*, 2010). This likely arises because the flexible trophic ability of these cells is well-suited to the summer environment since acquisition of dilute micronutrients like N can be supported through heterotrophy (Ross & Sharples, 2007). Mixotrophic dinoflagellates exhibit significant control on some communities as predators (Ward & Follows, 2016; García-Oliva *et al.*, 2022). Some diatom taxa such as *Chaetoceros* spp. show much higher (+340%) ammonium assimilation than predicted by mass transfer theory (Olofsson *et al.*, 2019), which likely explains their summer / autumn dominance in Loch Creran observed by Moschonas *et al.* (2017). There

remains a persistent importance of microzooplankton (Widdicombe *et al.*, 2010) and copepods, however, the composition of the adult copepod community often shifts between species and lifestages, varying across different temperate oceanic environments (Uye & Liang, 1998; Madsen *et al.*, 2001; Eloire *et al.*, 2010). Additionally, summer often sees the introduction of gelatinous zooplankton (primarily cnidarians and appendicularians), gastropods, chaetognaths, and meroplankton larvae (Eloire *et al.*, 2010; Arima *et al.*, 2014). The extent of the jellyfish biomass accumulation is often inconsistent and exhibits to significant interannual variation (Lucas, 2001; Eriksen *et al.*, 2012). Fish egg spawning also occurs for species such as *Sardina pilchardus* (sardines) during summer and autumn (Coombs *et al.*, 2005, 2010).

During autumn in many temperate seas, the breakdown of stratification (Dorrell *et al.*, 2022) allows nutrients from deep waters to be entrained once again and dilutes grazer concentrations which, combined with sufficient light for photosynthesis, creates an autumn bloom in some systems (Findlay *et al.*, 2006). This is not as ubiquitous as the spring bloom across temperate oceans, and typically is of lesser magnitude than the spring bloom, with nutrients also not depleted to the same degree (Martinez *et al.*, 2011; Wihsgott *et al.*, 2019; Dalpadado *et al.*, 2020). This biomass accumulation is often composed of similar species to the summer community, namely nanoplankton, diatoms or dinoflagellate microplankton (Painter *et al.*, 2016; Oseji *et al.*, 2019). During this period, zooplankton communities may also shift on a local scale to be dominated by carnivorous plankton such as cnidarians, ctenophores, and chaetognaths (Bresnan *et al.*, 2015). Moving into winter, biomass concentration decreases in line with fading light, and also with deepening mixed layer and subsequent dilution.

1.2.3 Size structure

Given that size is identified as a master trait in plankton ecology which governs key biological processes, it is natural to question how abundance relates to size. This question spans beyond plankton ecology and into wider ecological literature and, as a result, different methods exist for assessing the relationship between abundance and body size. In the review of White *et al.* (2007), four methods for assessing the size-abundance relationship are identified; here, we discuss size spectra, defined by White *et al.* (2007) as "the pattern relating the number of individuals in a body-size class (irrespective of species) and the average size of that body-size class. More generally, this is a frequency distribution (or probability density) of body sizes of individuals in a community."

1.2.3.1 Size spectra

Normally, the size spectrum is summarised by the fit of a log-log linear regression. As such, the term 'size spectrum' is synonymous with the process of the fitted abundance-size relationship in the literature, as opposed to the raw frequency distribution described by White *et al.* (2007). The slope of the 'size spectrum' (exponent on the original scale) is the key unit of inference, as it describes the abundance tradeoff between small and large organisms.

Phytoplankton abundance decreases with cell size, often with exponent not significantly different

to -0.75 across oceanic regions (Cermeño *et al.*, 2006; Cermeño & Figueiras, 2008). Slopes of size spectra can be converted to units of biomass and are often almost size independent, since the allometric exponent of carbon content per cell is $0.8-0.9$ (Menden-Deuer & Lessard, 2000; Marañón *et al.*, 2013). This corresponds to a slope of -1 on the normalised biomass size spectrum (NBSS), which is computed over size linearly, rather than on a log scale.

Deviations from the averaged phytoplankton size-abundance slope of -0.75 presented in Cermeño *et al.* (2006) are estimated from -1 to -1.4 in oligotrophic regions (Cermeño & Figueiras, 2008; Cavender-Bares *et al.*, 2001). A steep negative slope indicates a tendency towards smaller cells. In these regions, phytoplankton community in these regions is often dominated by picoplankton including *Prochlorococcus*, *Synechococcus*, and picoeukaryotes (Armengol *et al.*, 2019), and the food webs are characterised by tight coupling between phytoplankton and microzooplankton (Chen *et al.*, 2009; Cáceres *et al.*, 2013). Conversely, the abundance-size relationship can have slope closer to -0.75 in eutrophic temperate, upwelling, and coastal regions, indicating that large cells contribute more biomass in these systems (Reul *et al.*, 2005; Marañón *et al.*, 2007; González-García *et al.*, 2023).

The exponent of the zooplankton size-abundance curve is approximately -1 when averaged globally, although this also varies regionally (San Martin *et al.*, 2006). The exponent of the zooplankton size spectrum (and subsequently, NBSS) has important consequences for higher trophic levels, including fish. Analysis of ecosystems globally indicated that an increasingly negative exponent of the entire plankton biomass spectrum, and the associated decrease of phytoplankton biomass is likely to have an escalated effect on fish; a 16% decrease in phytoplankton biomass may cause a 38% decrease in fish biomass (Atkinson *et al.*, 2024).

1.2.3.2 Size distribution

An alternative summary of the size spectrum (as defined by White *et al.*, 2007) is through the computation of distributional moments, such as mean size, size variance (a proxy for size diversity), skewness, and kurtosis.

Biogeographic patterns of phytoplankton community mean cell size are typically positively correlated with community biomass (Irigoien *et al.*, 2004; Hatton *et al.*, 2015; González-García *et al.*, 2023), although relationships presented inIrigoien *et al.* (2004) exhibit some site-dependent bifurcation at high biomass ($> 100\,\mu$g C L^{-1}), indicating that the relationship is monotonic at some sites and asymptotic at others. Nonetheless, the driving force of this large-scale pattern is likely driven primarily by a combination of two characteristics: (i) greater nutrient availability minimising the importance of surface area-to-volume constraints of nutrient uptake affinity, and (ii) strong top-down control on small cell biomass due to tight coupling of small phytoplankton and microzooplankton, meaning that small phytoplankton have a quickly-fulfilled carrying capacity for small-cell biomass. For example, the size-to-nutrient affinity tradeoff resulting from increasing half-saturation constant with size (Eppley *et al.*, 1969; Litchman *et al.*, 2007) is displayed at a community level in the IronEx II bloom experiment (Coale *et al.*, 1996). In this experiment, two patches of an open ocean site were analysed, one in its natural state and the

other enriched with iron, resulting in a bloom of large phytoplankton cells. Schartau *et al.* (2010) showed that the iron enrichment increased the biomass contribution of larger phytoplankton cells compared to the natural patch, such that biomass accumulations of 30 μm phytoplankton cells were supported approximately ten-fold, and cells $> 100\,\mu$m have similar biomass to 30 μm despite the fact that no biomass was recorded above 40 μm in the unfertilized patch. Perhaps more interestingly, the microzooplankton biomass increased three-fold in the $20 - 60\,\mu$m range despite no noticeable biomass increase in their most likely prey ($< 5\,\mu$m phytoplankton), suggesting that primary productivity may have increased on this domain but that it was too tightly controlled by microzooplankton for biomass to be accumulated.

It follows that the accumulation of larger phytoplankton cells is an important driver of the zooplankton biomass distribution. If phytoplankton production is confined to pico- and small nanophytoplankton, then much of the primary consumption is dominated by microzooplankton and, as a result, mesozooplankton such as copepods exist above the third (or higher) trophic level (Bode *et al.*, 2015; Negrete-García *et al.*, 2022). Due to the significant losses associated with the process of grazing at each step of the food web, biomass accumulations of mesozooplankton are severely limited in region such as the subtropical gyres. For example, a modelled 80% of primary production is consumed by microzooplankton in oligotrophic subtropic regions, leading to biomass dominance of microzooplankton; in a more productive at similar latitude, microzooplankton and mesozooplankton biomass is approximately equal, with comparatively more efficient energy transfer to mesozooplankton (Armengol *et al.*, 2019). These patterns are reflected in the latitudinal patterns of zooplankton mean size, which are lowest in the subtropical gyres and highest in the polar regions, with a linear increasing trend between these latitudes in line with broad patterns of productivity (San Martin *et al.*, 2006; Brandão *et al.*, 2021).

Based on the concept that biomass and size are positively correlated, and that small cells remain prominent in most ecosystems (albeit with a carrying capacity due to their tight grazing coupling), it would be reasonable to expect that phytoplankton diversity increases with biomass. Indeed, log-ESD variance (referred herein as size diversity) shows a strict positive correlation with both biomass and cell size (Acevedo-Trejos *et al.*, 2018). In support of this, modelled phytoplankton size variance at an individual temperate site is close to its annual maximum during the spring bloom, when biomass is at its maximum (decreasing thereafter through summer; Acevedo-Trejos *et al.*, 2015). From a taxonomic perspective, comparisons of species diversity and biomass in both phytoplankton and microzooplankton are perceived to be unimodal across oceanic regions, albeit with a large degree of unexplained variability ($R^2 = 0.21$; Irigoien *et al.*, 2004.

There are many potential drivers affecting plankton taxonomic diversity. Temperature is a common focus for studies, and often exhibits a strong relationship with species richness. In Righetti *et al.* (2019), a model fitted to observations of sea surface temperature and species richness explains much of the variability ($R^2 = 0.76$), but interestingly is nonlinear. The authors propose that in warmer waters ($> 19°$C), species richness may be explained linearly by the metabolic theory of ecology, but other mechanisms may act in regions where the water temperature is $< 19°$C,

reflected by a nonlinear (v-shaped) relationship. As a result, species richness in temperate and subpolar regions is lower than anticipated (Righetti *et al.*, 2019). These findings at a global level are corroborated by Ibarbalz *et al.* (2019), who further show that species evenness (through Shannon's index) is optimised at low latitudes. However, laboratory experiments often yield different results, with negative relationships observed between phytoplankton species richness and temperature (Lewandowska *et al.*, 2014; Schabhüttl *et al.*, 2013).

From a size structure perspective, on the basis that environmental conditions, biomass, and richness are seen to vary widely in temperate oceans, it is little surprise that size structure is similarly affected. However, there is not a common consensus of the cycle of plankton size in the literature. A common conception is that mean phytoplankton size is generally larger in winter and spring and smaller in summer, due to the stress exhibited on large cells by limiting nutrients (p. 197, Simpson & Sharples, 2012; Acevedo-Trejos *et al.*, 2015). Other works, from coastal regions, suggest that mean size is lower in winter and higher in summer (González-Gil *et al.*, 2022), or at the very least that larger taxonomic groups such as diatoms and dinoflagellates play a more important role in the spring and summer months (Widdicombe *et al.*, 2010). Across these works there is an agreement that size is (relatively) large in spring. Therefore, consideration is due to whether these differences in summer size reflect local-scale differences, or arise as an artefact of differing techniques. I outlined earlier that biomass is better sustained in coastal / upwelling regions in summer than at open ocean temperate sites; given the apparent link between biomass and mean size, it is a reasonable step to tie together these differences in mean size with the associated differences in summer biomass. Regarding the discrepancy in winter, then given that light is believed to limit phytoplankton biomass accumulations in winter, it follows that the light-limited growth of phytoplankton would disproportionately affect larger cells, who suffer from intracellular shading (Edwards *et al.*, 2015). However, this trade-off is not included in the model of Acevedo-Trejos *et al.* (2015), perhaps driving the pattern in winter cell size (nutrient-related trade-offs are included).

Zooplankton taxonomic diversity appears to increase in the summer months in temperate coastal seas based on the introduction of jellyfish, larvae, gastropods, chaetognaths, among other taxonomic groups. Work focused solely on copepods at Plymouth L4 indicates decreasing species diversity from winter to summer coinciding with the growth of other groups (Eloire *et al.*, 2010). The implications for size diversity are not as clear, although a positive relationship between zooplankton taxonomic and functional diversity (incorporating size) has been observed, again primarily for crustacea (Pomerleau *et al.*, 2015). On one hand, the reproduction copepods and subsequent pulse of small zooplankton such as copepod nauplii, and introduction of large meso- and macrozooplankton such as jellyfish, could increase the size diversity. Size diversity is positively correlated with the organism size-biomass slope along a depth gradient in open oceans, indicating that large organisms complement smaller ones (Ye *et al.*, 2013; Kim *et al.*, 2025). At present, there is no conclusive picture of the cycle of zooplankton size diversity in temperate oceans. However, increases in zooplankton size diversity are associated with increased trophic transfer efficiency, supported by both ecological theory and observations (Ye *et al.*, 2013); phy-

toplankton size diversity has the opposite effect (García-Comas *et al.*, 2016). Thus, the size diversity of the zooplankton community appears to be an important characteristic to investigate.

1.3 Modelling approaches

The wealth of observational data that are collected in the field from both sampling and remote sensing methods can be used to establish patterns in ecology, but provide a limited mechanistic understanding of the planktonic ecosystem. Laboratory experiments are able to control sufficient variables to establish a stimulus-response relationship from which an individual mechanism can be understood, but cannot be scaled to explain complex ecosystem-level dynamics. Conversely, microcosm experiments can allow a diverse set of organisms to coexist in a simulated environment but monitoring individual interactions is effectively impossible. Some field experiments (such as IronEx II) are able to establish large-scale community drivers, but these are not a sustainable option to continually implement and would be difficult to control when applied to other research questions. This necessitates a theoretical modelling approach to investigate the mechanisms control plankton community structure. A mechanistic model is composed of a system of differential equations, each composed of terms representing biological and physical processes. By working with this system, the modeller is able to understand the effect of each mechanism and thus its impact on the wider ecosystem. The planktonic ecosystem is well-suited for such a study, because: (i) individual-level interactions are relatively simple compared to those exhibited at higher trophic levels, (ii) a comparatively large amount of data is available for model validation (e.g., satellite chlorophyll *a*), (iii) most biological and physical rates are relatively well established through decades of laboratory experimental research. As a result, simple models work well, creating a solid foundation for exploring more complex ecological and biogeochemical questions.

1.3.1 Basic NPZD model

1.3.1.1 Model equations

Mechanistic models were developed in works at least as early as Riley (1946), establishing some of the biological principles still used in today's approaches. These first models comprised only one differential equation for phytoplankton, but were able to accurately reproduce the spring bloom. Initially, models were constrained to be simple because they became computationally intractable as complexity was added. However, since the 1980s, models have increased in complexity with increasing computational power. The range of uses for models has expanded significantly, from end-to-end ecosystem modelling (e.g. Ito *et al.*, 2004; Heath *et al.*, 2021), quantifying global biogeochemical fluxes (e.g., Six & Maier-Reimer, 1996; Yool *et al.*, 2013), addressing complex size-structured food web interactions (e.g., Fuchs & Franks, 2010; Banas, 2011; Vallina *et al.*, 2014), and predicting climate change outcomes (e.g., Lovato *et al.*, 2022). Let us focus on the developments relevant to modelling size structure, building up from first principles.

The first task is to describe a basic framework for the ecosystem model, using a nutrient-

phytoplankton-zooplankton-detritus (NPZD) four-compartment configuration and simple slab physics. The following equations are drawn from EMPOWER-1.0 (Anderson *et al.*, 2015), but similar models have been developed much earlier in the literature (the physics are credited to Steele, 1956, while much of the biology is credited to Evans & Parslow, 1985; see Section 2 of Anderson *et al.*, 2015 for a thorough historical review). Note that I have made the following changes to EMPOWER-1.0:

- Exclude the density-dependent phytoplankton mortality term here, since density-dependent phytoplankton mortality is handled by zooplankton grazing, and it is favourable to minimise the number of unconstrained 'free' parameters. This therefore omits the explicit term representative of virus induced mortality, instead accounting for it implicitly (e.g., Ward *et al.*, 2012).
- Detritivorous grazing is excluded for simplicity and comparability with subsequently presented models, and to remove a further parameter.
- Temperature dependence is included for zooplankton grazing to reflect its importance in structuring the community (proven analytically by Norberg & DeAngelis, 1997 using a similar model). In fact, zooplankton normally show a greater sensitivity to temperature than phytoplankton (Eppley, 1972; Dam & Peterson, 1988; Rose & Caron, 2007; Chen *et al.*, 2012).

This style of model can be considered a basic model in the modern day, and its importance in this context is that it forms the foundational structure for advanced models (such as size-structured models). The model is spatially non-dimensional, but its slab physics simulate the changing of the mixed-layer depth (H; m) and resultant exchange between mixed layer and deep layer, in addition to constant cross-thermocline mixing rate (ω; m d^{-1}). The varying mixed layer, temperature (T), and surface irradiance (I_0) are externally forced using smoothed observations or an estimated seasonal curve. The deep layer is assumed to have $P = Z = D = 0$ and $N = N_0$. Each of the state variables are described as a concentration of nitrogen (μM N).

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \underbrace{\mu^{\max}(T) \cdot L_N(N) \cdot L_I(I) \cdot P}_{\text{P growth}} - \underbrace{G^{\max}(T) \cdot L_P(P) \cdot Z}_{\text{gross grazing loss}} - \underbrace{\left(m_P + \frac{\omega + H'}{H}\right) \cdot P}_{\text{mortality and mixing loss}} \tag{1.3}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \underbrace{\beta \cdot \lambda_Z \cdot G^{\max}(T) \cdot L_P(P) \cdot Z}_{\text{net grazing gain}} - \underbrace{\left(m_Z + m_{Z2} \cdot Z + \frac{\omega + H'}{H}\right) \cdot Z}_{\text{mortality and mixing loss}} \tag{1.4}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \underbrace{\beta \cdot (1 - \lambda_Z) \cdot G^{\max}(T) \cdot L_P(P) \cdot Z}_{\text{Z excretion}} + \underbrace{m_D \cdot D}_{\text{remineral.}} - \underbrace{\mu^{\max}(T) \cdot L_N(N) \cdot L_I(I) \cdot P}_{\text{P growth}} + \underbrace{\frac{\omega + H'}{H} \cdot (N_0 - N)}_{\text{mixing replenishment}} \tag{1.5}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \underbrace{(1 - \beta) \cdot G^{\max}(T) \cdot L_P(P) \cdot Z}_{\text{Z faeces and sloppy feeding}} + \underbrace{m_P \cdot P + m_Z \cdot Z}_{\text{mortality}} - \underbrace{\left(m_D + \frac{\omega + H' + v}{H}\right) \cdot D}_{\text{remineral., sinking and mixing loss}} \tag{1.6}$$

1.3.1.2 Biology

Here, phytoplankton biomass is assumed to be balanced with a fixed carbon-to-nitrogen ratio and, as such, carbon and nitrogen are accumulated at a fixed proportion. This allows the nutrient uptake to be a function of instantaneous photosynthesis and available nutrients. Maximal uptake of N ($\mu^{\max}$) is realised when nutrient and light limitation are saturated ($L_N(N) = L_I(I) = 1$), and is subsequently adjusted for temperature. At a reference temperature $T = T_{\text{ref}}$, $\mu^{\max}(T_{\text{ref}}) = \mu^{\max}_{\text{ref}}$ so that $\mu^{\max}(T) = \mu^{\max}_{\text{ref}} \cdot \lambda_T^{(P)}$. The temperature dependence $\lambda_T^{(P)}$ is normally used to moderate biological processes in the model and follows some form of exponential distribution. In the instance of Anderson *et al.* (2015), an Eppley curve is chosen (and can be rewritten) so that $\lambda_T^{(P)} = \exp\{R \cdot (T - T_{\text{ref}})\}$ with $R = 0.063$ (Eppley, 1972). This is a common choice in the literature, and corresponds to a Q_{10} value of approximately 1.895, although variations in temperature response across oceanic taxa may also require consideration (Chen & Laws, 2017). Almost ubiquitously in this fixed stoichiometry style of single-nutrient model, nutrient uptake saturation is modelled by a Michaelis-Menten / Monod response:

$$L_N(N) = \frac{N}{k_N + N} \tag{1.7}$$

where k_N is the half-saturation constant. N is assumed to be spatially uniform in the mixed layer and constant across the model timestep. The light limitation on phytoplankton growth shows a greater degree of variability within each timestep, since surface irradiance varies significantly

through the day and decays with depth. The light at depth $z \in [0, H]$ is given by $I(z) = I_0 \cdot \exp\{-(k_w + k_P \cdot P) \cdot z\}$, where k_w is the attenuation of seawater and k_P is the chlorophyll-induced shading by phytoplankton. At depth z and time t, the saturation of photosynthesis is given by:

$$L_I(I) = 1 - \exp\left\{-\frac{\alpha \cdot \theta_{\text{Chl}} \cdot I}{\mu^{\max}}\right\} \tag{1.8}$$

where α is the initial slope of the photosynthesis-irradiance response curve. It is noted that models of this style are often highly sensitive to α (Fasham *et al.*, 1990). Since light is highly variable across time of day and depth, it is necessary to average L_I over both domains. The daily cycle of irradiance can be externally forced using observational data, or using a trigonometric curve (see Appendix A in Anderson *et al.*, 2015). Models running a shorter timestep (e.g., hourly rather than daily) have the advantage in explicitly modelling the day-night cycle of light. Given the non-linear nature of the limitation from light, it may be advantageous to use this shorter timestep, although the effect of timestep size is not commonly discussed in the literature. Geider *et al.* (1997) discussed the lack of consensus regarding the importance of timestep in relation to variable θ_{Chl}.

Grazing is the key term controlling zooplankton growth (and subsequent phytoplankton losses). Grazing is maximised at high values of P (as $P \to \infty$, $L_P(P) \to 1$), with grazing saturation often modelled as an asymptotically saturating function (as per nutrient uptake), but with a low-density penalty to account for grazing inefficiency at low prey concentrations arising from heterogeneity of prey (Morozov, 2010) and to promote model stability (Vallina *et al.*, 2014). This is often modelled through a Holling III sigmoidal response, the use of a prey refuge term (e.g., Ward *et al.*, 2012), or a threshold approach with zero-grazing below a threshold value (Strom *et al.*, 2000). Anderson *et al.* (2015) use a Holling III response with an exponent of 2, which, with the omission of detritivory, is given as:

$$L_P(P) = \frac{P^2}{k_Z^2 + P^2} \tag{1.9}$$

where k_Z is the half-saturation constant of zooplankton grazing. In models where a prey-refuge term is used, this is implemented by modifying a Holling II response (which itself is identical in form to Michaelis-Menten / Monod) by multiplying the term $(1-\exp\{-\Lambda \cdot P\})$, which diminishes as $P \to 0$. In the model of Anderson *et al.* (2015), $G^{\max}$ is not temperature dependent, but it is in some other models (e.g., Ward *et al.*, 2012). In fact, zooplankton often show a stronger temperature dependence than phytoplankton, with a quoted Q_{10} value 2.2 for the clearance rate of the copepod *Temora longicornis* (which scales linearly with ingestion rate when the rate of change of gut contents is zero; Dam & Peterson, 1988), although more recent works suggest this may be overestimated under certain feeding conditions (Karaköylü & Franks, 2012). To represent the higher temperature sensitivity of zooplankton (Rose & Caron, 2007; Chen *et al.*, 2012), we set $\lambda_T^{(Z)} = \exp\{R \cdot (T - T_{\text{ref}})\}$ with $R = 0.8$ (Dam & Peterson, 1988), so that $G^{\max}(T) = G^{\max}_{\text{ref}} \cdot \lambda_T^{(Z)}$. Grazing is subject to inefficiencies, firstly through sloppy feeding and faecal waste (lost with proportion $1 - \beta$), and subsequently by excretion from the unassimilated nitrogen in the food

(lost with proportion $1 - \lambda_Z$). Often, the gross growth efficiency ($\beta \cdot \lambda_Z$) is around 0.3 (Straile, 1997).

An important feature in the model is the non-grazing mortality. Both phytoplankton and zooplankton are subject to a fixed linear mortality rate, representing losses primarily through cell / organism damage. This is fairly common in the literature, with rates freely selected normally around $1 - 2\%$ (e.g., Ward *et al.*, 2012; Anderson *et al.*, 2015) but also seen around $4 - 5\%$ (e.g., Fasham *et al.*, 1990; Acevedo-Trejos *et al.*, 2015). These losses are transferred to detritus, which as a whole is remineralized by bacteria at a fixed linear rate. Since higher trophic levels are not resolved in the model, the top-down control on zooplankton is normally inflicted using a density-dependent mortality term. The form of this term is almost always quadratic with fixed parameter (m_{Z2}) which is highly influential on the model dynamics (Steele & Henderson, 1992; Edwards, 2000). This biomass is removed from the modelled ecosystem at each timestep, reflecting the transfer of this biomass to higher trophic levels.

1.3.1.3 Physics

Each of the state variable is subject to loss / replenishment through exchange with the deep layer. First of all, a fixed cross-thermocline mixing rate ω exports N, P, Z, D and imports N_0. Although N_0 is not constrained to be larger than N, it invariably is since N is subject to uptake pressure from P, so that the result on N is net gain. Since the rate is constant, the proportion of the water column exchanged across the thermocline is higher when H is shallow. Detritus is also assumed to sink at rate v_D, with a similar consequence with respect to H as for ω. The deepening of the mixed layer ($H' > 0$) is also assumed to dilute the suspension of P, Z, D and further (net) replenish N. Conversely, shoaling of the mixed layer is not assumed to concentrate the community, since particles are assumed not to take evasive action to avoid entrainment to the deep layer. As such, $H' := 0$ when H is shoaling. A case can be made that Z are able to avoid entrainment into the deep layer (e.g. Fasham *et al.*, 1990; Acevedo-Trejos *et al.*, 2015; carnivorous zooplankton in Fasham, 1995). Some models also incorporate an overwintering switching function so that zooplankton enter a resting phase with no gains and minimal losses, where they are assumed to migrate to deep waters (Butenschön *et al.*, 2016). A consequence of exchange with the deep layer and export of density-dependent zooplankton mortality is that the model is not unconditionally mass conservative; that is, the total mass of $N + P + Z + D$ varies across the year.

Model ocean physics are an important aspect and the slab model described here is only one example of the range that exists. In even simpler models, the physics may be ignored entirely; this is particularly useful in models which seek to obtain an analytical solution, as it is easier to do so when the model is mass conservative. On the other end of the spectrum, models may become more complex by explicitly resolving the spatial distribution of plankton. One option is to model the depth distribution explicitly at a particular site (1D; e.g., Le Gland *et al.*, 2021). Alternatively, some models may consider the two spatial dimensions of depth and distance from some fixed point, such as a transect across a sea shelf (e.g., Baird *et al.*, 2004). A depth-resolved

model across a two-dimensional spatial domain is considered a 3D model (e.g., Ward *et al.*, 2012), and it is common for these models to be integrated across depths after simulation to show the distribution of depth-integrated biomass on a 2D surface such as latitude and longitude. Spatially-resolved models are often coupled to some physical ocean model to provide forcing for horizontal and vertical fluxes, which substantially increases both complexity and computational expense.

1.3.1.4 Parameterisation

Perhaps the ultimate challenge of model construction is constraining the parameters. These values are often selected from the literature but often require a degree of adjustment given the huge array of model formulations, experimental data for biological rates, and physical conditions of different ecosystems. A balance must be struck when optimising the parameters, firstly to ensure that inferential power is not lost from the model, and secondly because the process is computationally expensive. First of all, it is rare to adjust physiological parameters in the model such as growth rates to be adjusted without strong reasoning (e.g., Anderson *et al.*, 2015 reduced maximal photosynthetic rates due to iron limitation in some regions), as modellers wish to ensure that the biological aspect of the model remains realistic. Secondly, parameters are normally tested by repeated sampling from a grid of selected values. To take 3 values for each of the 18 different parameters described for this model, for example, would yield $3^{18} = 387,420,489$ combinations to run; reducing the number of adjusted parameters to 6 would yield $3^6 = 729$. In a model such as the one described, it is feasible and often necessary to experiment with a small number of free parameters (e.g., ω varying across sites due to local differences in oceanography), since many of the parameters are relatively well constrained. When the number of free parameters increases, alternative methods may be advantageous. Bayesian approaches give the advantage of producing an estimate of the full joint parameter space, allowing relationships between parameters to be understood in depth. Examples of Bayesian parameter optimisation are not common in the NPZD literature but can be found in similar applications, including the use of Metropolis-Hastings (MH, e.g. Heath *et al.*, 2021) or Approximate Bayesian Computation (ABC, e.g. Csilléry *et al.*, 2010).

1.3.2 Size-structured models

1.3.2.1 Discrete NPZD

A size structured NPZD model is most commonly constructed by splitting the P and Z compartments into bins representative of different sizes on the size domain. If there are say, $i = 1, ..., n_P$ phytoplankton sizes and $j = 1, ..., n_Z$ zooplankton sizes, then the same basic model equations as those presented in Equation 1.3−1.6 can be used, with a series of minor adjustments corresponding to the key allometric relationships discussed in Section 1.1.4.

First of all, for each phytoplankton size class i, growth rate will become size-dependent based on allometric relationships. Since the model is in units of nitrogen, the maximal growth rate may be chosen as monotonically decreasing (as is the case in Banas, 2011, since photosynthetic rate is not

explicitly resolved), while affinity will invariably decrease with size through increased nutrient uptake half-saturation constants and decreased initial slope of the P-I curve. From a bottom-up perspective, the smallest size class of phytoplankton will have the maximum N uptake advantage under all situations so that, without top-down control, this size class would therefore dominate the model. However, grazing rates are maximal at the small size classes of microzooplankton, so the evolution of fast growth and loss rates vs. slow rates is likely to control the community structure.

Grazing interactions are complex to resolve in a size-structured model because it must be determined which zooplankton graze on which phytoplankton. The simplest representation is to set a 1-to-1 grazing correspondence so that grazing on P_i by Z_j occurs for $i = j$ only. This could be a viable option if the purpose of the grazing is only to provide top-down control on phytoplankton. However, in models where inference is intended to be drawn from the biomass distribution of zooplankton, this application is unsuitable, because an unrealistic assumption is made that each Z_j would be consuming a different P_i, neglecting any direct competition between the zooplankton size classes. Instead, each zooplankton size class must be able to graze on several size classes of phytoplankton. Grazing is most efficient when the size of the optimal predator:prey size ratio is achieved, but quantifying the degree of preference for the other food particles adds a layer of complexity. Under the assumption that food particles becomes less attractive as they become smaller or larger than the optimal predator:prey size ratio, then a kernel density function can be used to create a feeding kernel governed by the equation:

$$\phi_{i,j} = \exp\left\{ -\left(\ln\left(\frac{\theta_{i,j}}{\theta_j^{\mathrm{opt}}} \right) \right)^2 \left(2\sigma^2\right)^{-1} \right\} \tag{1.10}$$

where $\theta_{i,j}$ is the predator:prey size ratio between predator j and prey i, θ_j^{opt} is the optimal predator:prey size ratio for predator j, and σ is the kernel width. This particular kernel is constructed assuming that predator-prey interactions are only herbivorous (Banas, 2011). Adding carnivorous interactions to the feeding interactions are constructed by using the same equation (Ward *et al.*, 2012). σ is chosen as a 'free' parameter arising only from modelling outputs, however, Banas (2011) provides a comparison of model responses across various levels of σ. The choice of θ_j^{opt} can be made either as a fixed ratio (Kiørboe, 2008) or increasing with predator size (Hansen *et al.*, 1994). Preferences can then be applied to grazing interactions to weight the amount of prey from each size class a predator will consume.

The predator-prey interactions described above assume that the composition of the zooplankton diet within each size class remains invariant as the prey field changes. The concept that individual predators exhibit switching behaviours when grazing (i.e., adjusted $[\phi_{i,j}]$ matrix) is founded in foraging theory, which predicts that predators should adapt to their prey field to maximise their fitness. There is some evidence of this type of switching exhibited in the larvae of *Acartia tonsa* (Kiørboe *et al.*, 1996) and *Paracartia grani*, although in the former case ingestion rates actually appear to decrease as a result of the switching response. This may relate to the complex interactions between prey-size availability and selectivity, the mechanisms of which are discussed

in Wirtz (2014). From a community modelling perspective, an argument can be made that the composition of zooplankton species within each of the modelled n_Z size classes is likely to change under different prey regimes, and the subsequent prey preference matrix may indeed change as a result. A switching preference matrix may then be described so that $\varphi_{i,j} \sim \phi_{i,j} \cdot P_i$. Implementation of this behaviour has been historically problematic because normalizing φ can lead to sub-optimal feeding (Gentleman *et al.*, 2003; Vallina *et al.*, 2014). A detailed discussion of the mathematics and ecological implications of this can be found in Vallina *et al.* (2014), who propose a theoretical optimal formulation for the active prey switching weighted on a pairwise $\{i, j\}$ basis, known as kill-the-winner (KTW), which is yet to be empirically validated.

One motivation for this style of grazing formulation is that competitive exclusion is a common problem affecting multi-compartment ecosystem models. Indeed, our mechanistic understanding of plankton ecology is apparently at odds with the existence of many plankton species in the ecosystem, as highlighted by the seminal work of Hutchinson (1961). 'The paradox of the plankton' highlights that phytoplankton remain diverse, even under limiting conditions where one species (or size) would be expected to dominate and exclude the others from existence. Therefore, the KTW approach described by Vallina *et al.* (2014) provides a potential solution to this quandary by providing active control of dominant species so that their biomass can never accumulate to the extent that they outcompete other species or sizes.

Switching can also be implemented in the model on a prey-type basis, as opposed to on an individual predator-prey interaction basis. For example in the simple NPZD model EMPOWER-1.0 (Anderson *et al.*, 2015), the grazing preference between phytoplankton and detritus is has a basic 2:1 P:D prey preference ratio, which is biomass adjusted. For example, if P biomass is double that of D, the instantaneous grazing preference becomes 4:1 P:D. Similarly, Ward *et al.* (2012) provide an example of grazing weighting for herbivory and carnivory which is related to the total biomass of phytoplankton and zooplankton prey respectively, and independent of the size distribution. This weighting provides sub-optimal grazing, as is proven below. Note that the grazing formulation here is an Ivlev formulation with prey refuge constant Λ. First we define the summed grazing by zooplankton j on phytoplankton i and zooplankton prey k, as G_j^{tot}. Then, the grazing formulation as given in Ward *et al.* (2012) is:

$$G_j^{\text{tot}} = G_j^{\max}(T) \cdot \left(\Phi_P \cdot \frac{\sum_i \phi_{i,j} \cdot P_i}{F_j^{\text{tot}} + k_Z} + \Phi_Z \cdot \frac{\sum_k \phi_{k,j} \cdot Z_k}{F_j^{\text{tot}} + k_Z} \right) \cdot \left(1 - \mathrm{e}^{-\Lambda \cdot F_j^{\text{tot}}} \right) \tag{1.11}$$

where $G_j^{\max}(T)$, k_Z, and $\phi_{(i \,\text{or}\, k),j}$ are defined as before. The total available food to the predator is given by:

$$F_j^{\text{tot}} = \sum_i \phi_{i,j} \cdot P_i + \sum_k \phi_{k,j} \cdot Z_k \tag{1.12}$$

and the herbivorous (Φ_P) and carnivorous (Φ_Z) grazing preferences are given by:

$$\Phi_P = \frac{\sum_i \phi_{i,j} \cdot P_i^2}{\sum_i \phi_{i,j} \cdot P_i^2 + \sum_k \phi_{k,j} \cdot Z_k^2} \tag{1.13}$$

$$\Phi_Z = \frac{\sum_k \phi_{k,j} \cdot Z_k^2}{\sum_i \phi_{i,j} \cdot P_i^2 + \sum_k \phi_{k,j} \cdot Z_k^2} \tag{1.14}$$

so that $\Phi_P + \Phi_Z = 1$. Upon inspection of Equation 1.11 we can see that, for any realistic case ($\exists i, j$ such that $P_i > 0$ and $Z_j > 0$):

$$\begin{aligned} \lim_{F_j^{\text{tot}} \to \infty} G_j^{\text{tot}} &= G_j^{\max} \cdot \left(\Phi_P \cdot \frac{\sum_i \phi_{i,j} \cdot P_i}{F_j^{\text{tot}}} + \Phi_Z \cdot \frac{\sum \phi_{k,j} \cdot Z_k}{F_j^{\text{tot}}} \right) \\ &< G_j^{\max} \cdot \left(\frac{\sum_i \phi_{i,j} \cdot P_i}{F_j^{\text{tot}}} + \frac{\sum \phi_{k,j} \cdot Z_k}{F_j^{\text{tot}}} \right) \end{aligned}$$

since $0 < \Phi_P < 1$ and $0 < \Phi_Z < 1$. Substituting Equation 1.12 for F_j^{tot} then we have that:

$$\lim_{F_j^{\text{tot}} \to \infty} G_j^{\text{tot}} < G_j^{\max}$$

and so the saturated grazing is sub-maximal, which is not a desirable property. In fact, the maximum possible grazing rate under saturating food conditions is minimised when $\Phi_P = \Phi_Z = 0.5$. Vallina *et al.* (2014) present a grazing formulation which is maximal (i.e., $G_j^{\text{tot}} = G_j^{\max}$ as $F_j^{\text{tot}} \to \infty$) and allows switching as above, which can be substituted in place of the expression above.

'The paradox of the plankton' has also inspired focus on the top-down control of the modelled zooplankton. The importance of the closure term in a simple NPZD model has already been highlighted (Steele & Henderson, 1992; Edwards, 2000), so a reasonable question arises in how to construct the density-dependent mortality term. At first thought, an obvious choice would be to set quadratic mortality as $m_{Z2} \cdot Z_j^2$. This however neglects the principles for the grazing kernel set out above, since predation on two or more (particular adjacent) size classes of Z_j are likely to be dependent to some degree. Applying the same theory to the top-down predation on zooplankton may suggest that a degree of community-weighted forcing should apply to the model. Record *et al.* (2014) present an alternative top-down control of the form:

$$j(\bar{Z}) \cdot Z_j = m_{Z2} \cdot \bar{Z}^{1-\nu} \cdot Z_j^{1+\nu} \tag{1.15}$$

where $0 < \nu < 1$ and $\bar{Z}$ is the mean zooplankton abundance / biomass across each of the n_Z size classes. The exponent ν is described on an open interval because when $\nu = 1$ the equation simplifies to a traditional quadratic mortality term, while when $\nu = 0$ the model is driven to extinction of all but one species. However, biodiversity appears to be maximised as $\nu \to 1$ (see Figure 1 of Record *et al.*, 2014; $0.5 < \nu < 1$ not shown in Figure 2), and so does not appear to be an effective mechanism to maximise species coexistence in an ecosystem model.

While the end goal would be to align modelled coexistence with observations, rather than to necessarily maximise it, the fact that Equation 1.15 appears to reduce coexistence compared to the standard quadratic mortality term is possibly the reason it has not gained significant traction in the modelling literature. Since models using the traditional quadratic mortality term generally struggle with facilitating observed coexistence, an updated modelling strategy would therefore aim to increase coexistence, whereas this formulation appears to reduce it. Nonetheless, Record *et al.* (2014) provide a viewpoint and discussion that should stimulate future studies, and it should be noted that formulation in Equation 1.15 is theoretically derived, rather than empirically.

Inclusion of size structure is crucial to understand biodiversity and ecosystem functioning in the planktonic ecosystem. However, as models become more complex, the number of decisions increases dramatically. Caution is required, in particular regarding the incorporation of multi-prey grazing (Gentleman *et al.*, 2003; Vallina *et al.*, 2014) and top-down zooplankton mortality (Record *et al.*, 2014). Additionally, computational expense is also much greater than for other models, since each of the modelled size classes must be in itself a state variable. The simple NPZD model described in Equation 1.3−1.6 has 4 state variables, whereas the one described here has $2 + n_P + n_Z$. For example, if $n_P = n_Z = 10$, which may be considered a conservative number of size classes, then there are 22 state variables.

1.3.2.2 Moment-based models

An elegant way to minimise the computational expense incurred by the discretised size structured NPZD model is to assume that the biomass distribution follows a fixed (say, log-normal) distribution which itself can be modelled. Since variations in the log-normal distribution can be described by its first two moments (mean and variance), the computational expense is drastically reduced. The mathematics supporting this approach are far more complex than those in the discrete NPZD model (these can be found in Wirtz & Eckhardt, 1996) which has perhaps limited its application in the literature, although there are examples of its use across various traits (Norberg *et al.*, 2001; Merico *et al.*, 2009; Acevedo-Trejos *et al.*, 2015). Much of the complexity is credited to the requirement of moment-closure correction terms in the model arising from the integration process. Its use in a size-structured NPZD model is demonstrated by Acevedo-Trejos *et al.* (2015) where phytoplankton cell size is the trait of interest and the model is constructed with 6 state variables: N, P, Z, and D masses, mean phytoplankton log-ESD, and phytoplankton log-ESD variance.

Originally, the primary drawback of this approach was that it was constrained for use with one trait. For example, in the Acevedo-Trejos *et al.* (2015) paper, the phytoplankton size is the only trait resolved (not zooplankton size). In this instance, grazing pressure and rate are considered a function of the instantaneous phytoplankton community size structure and zooplankton biomass, and a further free parameter describing the allometric exponent of the grazing preference is introduced to which the model is highly sensitive.

However, Le Gland *et al.* (2021) more recently presented a model which resolves two traits: nitrogen half-saturation constant and optimal temperature. The half-saturation constant can

be considered analogous to cell size, given the positive allometric exponent (Eppley *et al.*, 1969; Litchman *et al.*, 2007). While this is still an insufficient level of detail to investigate a wide range of allometric tradeoffs (Section 1.1.4), it appears that continual advancements in trait-based will push it to the forefront of modelling methodology in years to come.

1.3.3 Photo-acclimation and variable stoichiometry

The basic model described in Section 1.3.1 is a powerful tool and is shown to be relatively effective in reproducing many aspects of the seasonal cycle of phytoplankton. However, inference on the limitation of growth is itself limited, since the terms in the model are coarsely resolved. While the top-down control may be better resolved by incorporating a size-structured model, the bottom up control is still only partially described. For example: it was outlined above that the model is highly sensitive to the initial slope of the P-I curve (α; Fasham *et al.*, 1990), and review of Equation 1.8 shows that α is multiplied by θ_{Chl}. But we know that θ_{Chl} is highly variable in nature, for example changing threefold from winter to summer in a coastal community (Eriksen *et al.*, 2012). The term $\alpha{\cdot}\theta_{\mathrm{Chl}}$ should, therefore, show a threefold decrease from winter to summer, which is likely to significantly affect the seasonal limitation of photosynthesis in phytoplankton, given the sensitivity of the model to the initial slope of the P-I curve. The work of Geider *et al.* (1997) was key in incorporating variable C:Chl ratios through photo-acclimation to mechanistic models.

Another issue with the model basic model described above is that it relies on the unrealistic assumption that photosynthesis and subsequent carbon assimilation is limited by the instantaneous nitrogen availability. Given the ability of cells to perform 'luxury' uptake of nutrients (Elrifi & Turpin, 1985; Olofsson *et al.*, 2019), a quota-based approach to cell stoichiometry is necessary to accurately represent this, and was first presented mathematically by Droop (1968). In this original work, the N:C molar quota (Q_N) fulfilment controlled μ with a hyperbolic response, so that growth rate instead becomes:

$$\mu = \mu^{\max} \cdot \left(1 - \frac{Q_N^{\min}}{Q_N}\right). \tag{1.16}$$

where $Q_N^{\min}$ is the minimum possible N:C quota, and noting that as $Q_N \to \infty$, $\mu \to \mu^{\infty}$. Other hyperbolic forms of this response are well-described in the historical review in the Introduction of Flynn (2008), where it was shown that the optimal curve fitting the C:N quota growth-limitation is actually linear (first established by Geider *et al.*, 1998; as in Equation 1.20), although the authors show that is not the case for all nutrients. Notably, this hyperbolic form has an upper bound for Q_N, which is $Q_N^{\max}$, so that as $Q_N \to Q_N^{\max}$, $\mu \to \mu^{\max}$. The importance of the quota-based growth rate is clear from an ecological point of view, but its implementation is also important for biogeochemical applications, since C and N export are controlled by the composition of the plankton community.

The model of Geider *et al.* (1998) provides a modelling description of the C:N:Chl stoichiometry of a phytoplankton community. This expands on the foundational work of Geider *et al.* (1997),

where variable C:Chl was discussed and modelled. Therefore, the Geider *et al.* (1998) model is a perfect case study to demonstrate how phytoplankton nitrogen (P_N), carbon (P_C), and chlorophyll *a* (P_{Chl}) contents interact. The equations are outlined below (with symbols adjusted for continuity); note that this is not a full NPZD ecosystem model, but its implementation in an NPZD-style ecosystem model with predator-prey interactions and variable nutrient N is described in state-of-the-art models such as that of Ward *et al.* (2012).

$$\frac{\mathrm{d}P_C}{\mathrm{d}t} = \underbrace{V_C^{\max}(T) \cdot \lambda_N(Q_N) \cdot L_I(I, Q_{Chl}) \cdot P_C}_{\text{photosynthesis } (= V_C \cdot P_C)} - \underbrace{\zeta \cdot V_N \cdot P_C}_{\text{biosynthesis}} - \underbrace{R(T) \cdot P_C}_{\text{respiration}} \tag{1.17}$$

$$\frac{\mathrm{d}P_N}{\mathrm{d}t} = \underbrace{V_N^{\max}(T) \cdot Q_N^{\text{stat}}(Q_N) \cdot L_N(N) \cdot P_C}_{\text{N assimilation } (= V_N \cdot P_C)} - \underbrace{R(T) \cdot P_N}_{\text{respiration}} \tag{1.18}$$

$$\frac{\mathrm{d}P_{Chl}}{\mathrm{d}t} = \underbrace{\rho_{Chl} \cdot V_N \cdot P_C}_{\text{Chl synthesis}} - \underbrace{R(T) \cdot P_{Chl}}_{\text{respiration}} \tag{1.19}$$

Carbon accumulation rate through photosynthesis is expressed as a function of temperature-dependent maximum photosynthetic rate ($V_C^{\max}(T)$; d^{-1}) adjusted by N saturation ($\lambda_N(Q_N)$), and light limitation ($L_I(I, Q_{Chl})$). This is therefore very similar to Equation 1.3 in the simple NPZD model, with the two key differences that are desired: (i) nutrient-limitation on photosynthesis arises from internal quotas rather than environmental N concentrations, and (ii) light limitation is reflective of instantaneous, variable Chl:C quota. The equation for nutrient saturation is given as:

$$\lambda_N = \frac{Q_N - Q_N^{\min}}{Q_N^{\max} - Q_N^{\min}} \tag{1.20}$$

where $Q_N = P_N/P_C$, and $Q_N^{\min}$ and $Q_N^{\max}$ are the minimum and maximum possible quotas respectively. We also define the (light-) saturating maximal photosynthetic growth rate as $V_C^{\text{sat}} = V_C^{\max} \cdot \lambda_N$ so that the equation for light-limitation can also be obtained by setting $\theta_{\text{Chl}} = Q_{Chl}$ and $\mu^{\max} = V_C^{\text{sat}}$ in Equation 1.8.

The modelled phytoplankton must also assimilate nutrients to sustain balanced growth, again at a temperature-adjusted maximal rate ($V_N^{\max}(T)$; mol N (mol C)$^{-1}$ d^{-1}), which is moderated by the instantaneous nutrient availability ($L_N(N)$) and the existing quota fulfilment (Q_N^{stat}). $L_N(N)$ is described as in Equation 1.7, while the latter saturation term is necessary in to ensure that luxury nutrient uptake does not exceed $Q_N^{\max}$, and is given by:

$$Q_N^{\text{stat}} = \left(\frac{Q_N^{\max} - Q_N}{Q_N^{\max} - Q_N^{\min}} \right)^n \tag{1.21}$$

so that as $Q_N \to Q_N^{\min}, Q_N^{\text{stat}} \to 1$. There is a free exponent n controlling the shape of the uptake curve, which is assumed to be 1 in the absence of evidence to suggest otherwise (as used in Ward *et al.*, 2012).

Chlorophyll accumulation rate is expressed as a fraction of N uptake rate ($\rho_{Chl} \cdot V_N$; g chl *a* (mol C)$^{-1}$ d^{-1}). This is instantaneous and cannot be converted from internal quotas, and is inversely related to the degree of saturation in the P-I curve. The per unit of nitrogen uptake chlorophyll synthesis rate is given by:

$$\rho_{Chl} = \theta_N^{\max} \cdot \frac{V_C}{\alpha \cdot Q_{Chl} \cdot I} \tag{1.22}$$

where $\theta_N^{\max}$ is the maximum Chl:N ratio and noting that as $\alpha \cdot Q_{Chl} \cdot I \to \infty$, $\rho_{Chl} \to 0$ (and $V_C \to V_C^{\text{sat}}$). The asymptotic behaviour as $I \to 0$ requires L'Hôpital's Rule to determine the limit, so a brief proof is given as follows (noting that $Q_{Chl} > 0$ and $\alpha > 0$):

$$\text{Let } A(I) = V_C = V_C^{\text{sat}} \cdot \left(1 - \exp\left\{-\frac{\alpha \cdot Q_{Chl} \cdot I}{V_C^{\text{sat}}}\right\}\right) \text{ and } B(I) = \alpha \cdot Q_{Chl} \cdot I.$$

$$\text{Then } \lim_{I \to 0} \frac{A(I)}{B(I)} \text{ is unknown, since } \lim_{I \to 0} A(I) = 0 \text{ and } \lim_{I \to 0} B(I) = 0.$$

$$\text{Thus, we require } \lim_{I \to 0} \frac{A'(I)}{B'(I)} \text{ under L'Hôpital's Rule.}$$

$$A'(I) = \alpha \cdot Q_{Chl} \cdot \exp\left\{-\frac{\alpha \cdot Q_{Chl} \cdot I}{V_C^{\text{sat}}}\right\} \text{ and } B'(I) = \alpha \cdot Q_{Chl}$$

$$\lim_{I \to 0} \frac{A'(I)}{B'(I)} = \lim_{I \to 0} \exp\left\{-\frac{\alpha \cdot Q_{Chl} \cdot I}{V_C^{\text{sat}}}\right\} = 1 \qquad \lim_{I \to 0} \rho_{Chl} = \theta_N^{\max}$$

indicating that chlorophyll *a* synthesis is maximised when irradiance (and subsequent photosynthesis) is zero.

There are two respiratory costs included in the model: a temperature-dependent basal rate ($R(T)$; d^{-1}; which is set to zero in 3 of the 4 species model runs by Geider *et al.*, 1998), and the cost of biosynthesis at rate ζ. The latter is a proportional cost so that as N is assimilated, C is expended. Implementation of other mortality terms (such as linear mortality and grazing loss) would remove N, C, and Chl biomass in proportion to the existing quotas.

This style of model provides clear theoretical advantages when compared to the simple growth component of the model presented in Equation 1.3−1.6. The flexibility of the cells to store nutrients and use these to fuel photosynthesis is evidently important, and similarly for the cells to regulate their chlorophyll is likely to contribute to important functional differences when compared to a fixed C:Chl ratio. However, we still desire a model to resolve the functional diversity of the community, having established the importance that diversity has in structuring food web interactions and the transfer of energy to higher trophic level. That said, simplicity and tractability are favourable properties in any model to maximise usability and to have meaningful inference. Having established that biological rates and ecosystem interactions in plankton are well described by size, incorporating size structure into the model is a crucial step if the aim of the model is to understand the food web dynamics of a community.

1.3.4 Alternative food web representations

Conventional NPZD modelling approaches neglect a crucial aspect of food web dynamics: mixotrophy. Although efforts have been made to incorporate mixotrophy into ecosystem modelling approaches, its application remains limited in the literature because data quantifying the degree of auto- and heterotrophy across mixotrophic communities is limited (Flynn & Mitra, 2009). Modelling approaches to this point have either focused on the detailed processes within mixotrophs (Flynn & Mitra, 2009) or the place of mixotrophs within the food web using a more coarse representation of their biology (Ward & Follows, 2016; Serra-Pompei *et al.*, 2020).

Perhaps the simplest and most conservative method to construct an ecosystem model incorporating mixotrophy is to base it in roots of the NPZD model. Ward & Follows (2016) did so by including size-structured mixotrophs as a sum of same-sized phyto- and zooplankton processes described in their original ecosystem model described in Ward *et al.* (2012). What effectively arises here is a multi-compartment NPMZD model, for which the maximal uptake rates for the mixotroph (M) were multiplied by an adjusting exponent (0.5^{τ}) and the corresponding model outputs assessed. The mixotroph's rate of change is therefore analogous to 0.5^{τ} times phytoplankton gains and losses plus 0.5^{τ} times zooplankton gains and losses within each size class. When $\tau = 0$, the mixotroph has the full capabilities of both phytoplankton and zooplankton, thus excluding both and creating a mixotrophy-only model. Conversely, as $\tau \to \infty$, the model simplifies back down to the original NPZD-style model because the competitive ability of the mixotrophs diminishes. In the mixotrophy only model, the community is naturally structured by the size-structure of prey availability and the allometric nutrient uptake and affinity tradeoffs, so that trophic level broadly increases with size. The result is that the modelled global biomasses are lower in the pico- and nanoplankton, and higher in the micro-, meso-, and macroplankton in the mixotrophy-only model compared to the phytoplankton-zooplankton model.

The coexistence of mixotrophs alongside phytoplankton and zooplankton also propagates a positive shift in the size domain, with values $0.7 < \tau < 1.3$ approximately doubling the global mean plankton size compared to the original NPZD-style model, and contributing over half of the global plankton biomass. A further modelling experiment was conducted by adjusting the mixotroph's nutrient uptake and grazing affinities by 0.5^{τ}, in addition to the saturating rates. There is reasonable basis to incorporate this change; some of the nutrient uptake sites for example, located on cell's exterior, may have to be sacrificed to facilitate phagotrophy (and vice versa). The result here was a relative increase in mean size and biomass compared to the $\tau \to \infty$ case, but to a lesser extent than the saturation-only case. Overall, this work − while only theoretical − indicates the potential importance of mixotrophy in structuring the planktonic food web.

An alternative approach is that of the Nutrient-Unicellular-Multicellular framework (NUM; Serra-Pompei *et al.*, 2020). This model describes size structured state-variables for protists and copepods. The protists are able to exercise photosynthesis, nutrient uptake, and graze on other protists. Copepods grow through grazing on protists, other copepods, and (size-structured) detritus, and are also structured by their life history from nauplii to adults. This work places more of a focus on the emergent distribution of copepods and as such, lacks the variable stoi-

chiometry and photo-acclimation described by other models. Since most mixotrophs tend to be C-limited as opposed to N-limited (as is the case in for phytoplankton), inclusion of their intracellular nutrient concentrations are important to understand their ecosystem functioning (Flynn & Mitra, 2009). This has motivated the most recent advancements to the NUM model, where a fixed stoichiometry is resolved for C and N in generalist mixotrophs, as well as Si for an additional diatom class not resolved in the original model (Papapostolou *et al.*, 2026). Instantaneous uptake rates in this model are therefore driven by the need to maintain fixed stoichiometry.

At the other end of the spectrum, the model of Flynn & Mitra (2009) provides great detail on the functioning of mixotrophs, but is not size structured and does not resolve phyto- or zooplankton. In this model, there are 8 state variables: overall concentrations of C, N, P, and Chl, and respective food vacuole concentrations for each of these. This is highly advantageous as it allows the mixotrophs to exercise heterotrophy within realistic constraints. It would appear that this style of model is therefore more realistic, albeit by necessitating parameters for food vacuole size and digestive mechanisms. This resolution of mixotrophy is yet to be implemented in a full-ecosystem plankton model.

1.4 Plankton in the UK coastal ecosystem

1.4.1 Overview of the UK coastal environment

Long-term plankton time series data are available at three UK coastal sites: Plymouth L4, Stonehaven, and Loch Ewe. These sites represent three distinct regions of the UK coast: the western English Channel, the northwest North Sea, and a west coast sea loch in close proximity to land and influenced by the North Atlantic Ocean. The diversity among these monitoring stations should allow a holistic view of the UK coastal plankton community, since different regions are accounted for.

Plymouth L4 (PL; 50.25°N, 4.22°W) is an oceanographic monitoring station located in the Western English Channel with a depth of 58 m, 10 km offshore from the coastal city of Plymouth. In its proximity are two major rivers, River Tamar and River Plym (with a combined mean flow rate of $\sim$ 26 m^3 s^{-1}; NRFA, 2023), which exert a freshwater influence on the site particularly during periods of heavy rainfall (Rees *et al.*, 2009). Despite being classified as a constituent of the Greater North Sea subregion, the English Channel has its own distinct features with strong influence from the Atlantic Ocean and overall flow from west to east. The entire English Channel has a flushing rate of $\sim$ 500 days (Dauvin, 2019), but rates in the inner Plymouth coastal zone (depth 30 m) are typically 6.5 to 13.5 days (Uncles & Torres, 2013). Typical flushing rate may therefore be slightly greater than 14 days at Plymouth L4, but is likely to be far less than 500 days. Weekly zooplankton sampling began in 1988 and phytoplankton in 1992, and the station is normally visited between 40 and 45 weeks of the year (Harris, 2010).

The Scottish monitoring sites at Stonehaven and Loch Ewe are described by Bresnan *et al.* (2015) and the key details are listed here for completeness. The sampling site at Stonehaven (SH; 56.96°N, 2.10°W) is located $\sim$ 5 km off the east coast of Scotland in the northwest North

Sea with water depth 48 m. The River Dee and River Don are 24 km to the north and discharge an annual combined average of 68 m^3 s^{-1}, and 70 km south of the site, the River Tay discharges an average of 172 m^3 s^{-1} (NRFA, 2023). The residual current travels from north to south, and as such, the site is located to minimise the effects of freshwater input (Bresnan *et al.*, 2015). Weekly sampling commenced at SH in 1997 on a weather-permitting basis (O'Brien *et al.*, 2010).

Loch Ewe is a large sea loch (9th of 110 by volume; Edwards & Sharples, 1986) on the north-west coast of Scotland. There is strong tidal influence and, being a semi-enclosed sea loch, the flushing rate is 4 days on average (Bresnan *et al.*, 2015). Loch Ewe opens out to the North Minch, which, under the northward-travelling and fast-moving Scottish Coastal Current, is directly influenced by waters from the Irish Sea and the wider Atlantic Ocean (Simpson & Hill, 1986). The Loch Ewe sampling site (LE; 57.84°N, 5.61°W) is located $\sim$ 0.5 km offshore with a depth of 40 m (O'Brien *et al.*, 2010). Due to the proximity of the site to the coastline, there are significant freshwater influences (namely the River Ewe with flow rate $\sim$ 30 m^3 s^{-1}; NRFA, 2023). Data has been collected weekly (weather-permitting) since 2000 (Marine Scotland Science, 2018a).

1.4.1.1 Environment

Owing to a high latitude, the UK coastal seas experience strong seasonality in environmental conditions. Light is highly variable over the course of the year, with a daylight hours difference between the winter solstice and summer solstice of $\sim$ 8.5 hours on the southern coast of mainland England ($\sim$ 50°N) and $\sim$ 12 hours on the northern coast of mainland Scotland ($\sim$ 59°N). At Mountbatten near PL, an average of 1732 hours sunlight per year is observed, whereas at Dyce near SH 1447 hours and at Kinlochewe near LE only 926 hours per year are observed (Met Office, 2020). Despite the similar latitudes of the SH and LE sites, LE experiences significantly lower sunlight due to local weather influences. This is reflected by the number of days with more than 1 mm of rain: 148, 142, and 220 days at meteorology stations near PL, SH, and LE respectively (Met Office, 2020). Another consideration at this latitude is that the proportion of light reflected by the sea surface varies over the year due to the changing altitude of the sun. Based on global ocean estimates made by Kukla & Robinson (1980), the albedo of PL ($\sim$ 50° N) is at a maximum of 16.8% in February and a minimum of 6.1% in July; at SH & LE ($\sim$ 58° N) is 31.8% in March and 7% in August and September. The range is considerably higher for the Scottish sites reflecting the fact that at around 54°N the winter albedo starts to increase rapidly. Note that these are global ocean estimates and do not account for longitudinal differences in solar radiation and atmospheric conditions. Cloudiness reduces albedo due to an increased proportion of diffuse radiation, and similarly, the Albedo is maximised by calm seas (Payne, 1972). However, based on the fact that wave height is lower close to shore (Marine Scotland, 2020), then the empirically averaged wave penalty on albedo may be an over estimate when applied to the UK coastal sites. Also, the reflectivity of light varies by wavelength (Jin *et al.*, 2004).

Temperature is also variable across the seasons, but to a lesser degree than light, due to the moderating influence of the Atlantic meridional overturning circulation. Seasonality of temperature

is therefore more moderate than what may be expected when compared to other similar-latitude shelf-sea and coastal regions (McCarthy *et al.*, 2017). Temperatures range between $8 - 18°C$ on average at PL (Widdicombe *et al.*, 2010; McCluskey *et al.*, 2022), and $6 - 13°C$ at both SH (González-Gil *et al.*, 2022) and LE (Bresnan *et al.*, 2015). In line with many other oceanic regions, the UK shelf seas have seen significant warming over the last 50 years, with the highest rates observed on the north coast of Scotland (0.24°C per decade). Significant increases are also observed in both the northwestern coast and eastern coasts of the UK (0.2°C per decade), while increases on the Southern coast are observed ($0.12 - 0.16°C$ per decade) but not statistically significant (Tinker *et al.*, 2020).

A consequence of the seasonal effects of temperature and light is that some of the UK coastal seas are seasonally stratified in the summer, depending on the level of heat input, tidal mixing, and water depth (Sharples *et al.*, 2013). A clear thermocline emerges at PL (Widdicombe *et al.*, 2010; Dauvin, 2019), while at SH this is more transient; a strong thermocline is observed by (González-Gil *et al.*, 2022), but Bresnan *et al.* (2015) state that stratification is weak due to strong tidal influence. At LE stratification also occurs during summer, and salinity is seen to drop drastically at the surface on occasion due to freshwater influence (Bresnan *et al.*, 2015).

Rivers further influence the marine environment by transporting nutrients and suspended sediment to the coastal seas. For example, during an extreme rainfall event in Plymouth in 2007, the dissolved inorganic nitrate concentration was 24× higher than the weeks immediately before and after at PL, and 5× higher than the long-term summertime mean (Rees *et al.*, 2009). The coastal North Sea is also known to be affected by riverine inputs, with the rivers Tay, Dee, and Don providing an average input rate of 15.4 mol s^{-1} nitrate and 0.2 mol s^{-1} phosphate, at a rate $2 - 3\times$ higher in winter than in summer (Lyons *et al.*, 1993). The estuary of the Don is particularly nutrient rich because of intensive agriculture and (relatively) low freshwater input (Balls, 1994). Although no known work has established a clear pattern of direct riverine nutrient loading from River Ewe, it is worth noting that the catchment area is subject to limited agricultural and anthropogenic activity when compared to PL and SH. Furthermore, highland rivers are known to be lower in nitrate and phosphate than their lowland counterparts, often with lower average concentrations than the coastal seas (Balls, 1994), and likely due to differences in soil composition (Lyons *et al.*, 1993). There is, however, no direct evidence that this is the case in the River Ewe.

External loading of nutrients, sediment, and organic matter is not solely driven by input from rivers. Tides and tidally-averaged currents advect matter horizontally from neighbouring waters, and in some cases short-term variability in phytoplankton biomass may be driven by horizontal exchange (Blauw *et al.*, 2018). Such an effect has been shown at PL, with counts of phytoplankton and gelatinous zooplankton almost doubling at the same depth, but at different times in a tidal cycle. This is primarily driven by observed advective speeds of up to 0.5 m s^{-1} at the surface, equivalent to 1.8 km h^{-1} (compared to 0.2 m s^{-1} at the seabed; Cross *et al.*, 2015). On a longer timescale, in the northwest North Sea of which SH is a constituent, southward advection carries *Calanus finmarchicus* from overwintering sites. This process accounts for an estimated 41% of

C. finmarchicus biomass in the region, and the community is thus not self-sustaining and instead reliant on inflow (Gao *et al.*, 2021). Thus, unlike the exchange with the deep layer, where water is assumed to be nutrient-rich and without plankton (Section 1.3.1), the horizontal exchange of water generally has its own plankton community to be considered.

Storms are commonplace in UK coastal seas, most commonly but not exclusively occurring in winter. This is partially reflected by the overall increased wind speeds from on-land sites near each of the coastal monitoring sites. January and July wind average wind speeds are as follows: 6.25 m s^{-1} vs 4.53 m s^{-1} at Mountbatten near PL, 5.10 m s^{-1} vs. 3.75 m s^{-1} at Dyce near SH, and 6.15 m s^{-1} vs 3.99 m s^{-1} at Aultbea near LE (Met Office, 2020). Moreover, storm systems generated and travelling across the North Atlantic Ocean and within the North Sea considerably enhance swell over the winter months. Average wave heights vary from 2 m in summer to 5 m in winter in offshore seas on the West coast of Scotland, and 1.5 m to 3 m on the East coast (Marine Scotland, 2020). These events enhance the mixing in the water column, replenishing nutrients but suspending sediments (p. 199, Simpson & Sharples, 2012). As a result, seawater attenuation is often between 0.2 m^{-1} and 0.3 m^{-1} at SH, at its highest in January and February (González-Gil *et al.*, 2022), and with similar values recorded in the Irish Sea (Bowers *et al.*, 1998). In the Irish Sea, attenuation is driven primarily by suspended solids rather than chlorophyll *a*, and is much higher than what may be expected at an open ocean site − particularly in winter, when the mass of suspended solids is 2.7× higher than in summer (Bowers *et al.*, 1998). For example, if using the standard equation for irradiance decay with depth (near Equation 1.8), with $k_w = 0.04$ m^{-1} (Dutkiewicz *et al.*, 2012) and $k_{Chl} = 0.03$ m^{-1} (μg chl)$^{-1}$ (Moore *et al.*, 2001) as per Ward *et al.* (2012), then attenuation would be expected to be 0.046 m^{-1} in January and 0.143 m^{-1} in May, using 0.2 μg chl *a* L^{-1} and 3 μg chl *a* L^{-1} respectively from data at Stonehaven presented in O'Brien *et al.* (2010). Assuming the chlorophyll-specific attenuation rate of Moore *et al.* (2001) to be constant across taxa, and using attenuation data presented in González-Gil *et al.* (2022), then a brief calculation will find that, at SH, $k_w = 0.28$ m^{-1} in January and $k_w = 0.16$ m^{-1} in May, showing qualitative similarities between SH and the Irish Sea. The fact that observed attenuation is actually higher in winter despite the fact that chl *a* is lower reflects the significant seasonal variability in the sediment-driven seawater attenuation, which is rarely implemented in open ocean models (e.g., Anderson *et al.*, 2015) but appears essential in a coastal application. As such, light is further penalised in winter through the water column, in addition to the short days, low intensity peak irradiance at the sea surface, and increased proportion of reflected light.

Tides also play an important role in driving mixing in coastal seas, where frictional forces are greater, with about 75% of global tidal energy being dissipated in shelf seas (p. 46, Simpson & Sharples, 2012). The result is that the shallowest coastal sites tend to be permanently mixed, while stratification is generally only exhibited in offshore regions of coastal seas with greater bottom depth (p. 180, Simpson & Sharples, 2012). In such offshore regions are SH & PL, which are both still substantially influenced by tidal mixing, thus leading to transient stratification (Widdicombe *et al.*, 2010; Bresnan *et al.*, 2015). Mixing also has the secondary effect of cycling the phytoplankton through the mixed layer, so that cells experience periods of diminished light

due to attenuation.

1.4.1.2 Ecosystem

The UK coastal seas are home to a number of commercially and environmentally important species of fish. The UK Sea Fisheries Statistics 2023 report (UK Marine Management Organisation, 2024) states that in 2023, UK fisheries landed 719,000 tonnes of fish, valued at £1.1 billion. 62% of landings were pelagic, of which mackerel, herring, and blue whiting were the most valuable species. 20% of landings were demersal fish (primarily haddock, cod, and angler-/ monkfish), and 18% were shellfish (primarily lobsters, scallops, and crabs). In the North Sea, fisheries are believed to exert a significant impact on the top-down control on the planktonic ecosystem (Reid, 2000).

Plankton are a crucial food source for several pelagic fish. Mackerel, during their planktonic larval life stage, feed initially on phytoplankton, copepod nauplii, and copepod eggs (Patel *et al.*, 2023); particularly 80 μm *Acartia* spp. eggs and 160 μm *Calanus* spp. eggs (Conway *et al.*, 1999). Feeding shifts to copepodite copepods more predominantly with increasing body length ($>$ 6 mm), as well as piscivory and cannibalism facilitated by the presence of teeth when $>$ 5 mm in length (Conway *et al.*, 1999). In their adult stage, mackerel move up the food web to feed on smaller fish such as herring (Dahl & Kirkegaard, 1986), which themselves are planktivorous, feeding primarily on copepods ($>$ 80% in summer and autumn; Patel *et al.*, 2023) with a size-selective prey preference for stage C4−C5 copepods approximately $300 - 500\,\mu$m (Arrhenius, 1996; assuming prolate spheroid shape with 0.3 width-length ratio). Adult blue whiting predominantly eat euphausiids, amphipods and decapod larvae, with little grazing on copepods except in juveniles (Dolgov *et al.*, 2010; Cabral & Murta, 2002).

Many other fish rely on plankton for direct prey in UK coastal waters, such as sandeels and sprat. These fish have less commercial interest as other species mentioned above due to the vulnerability of their stocks, but are important components of the ecosystem (Fulton & Sainsbury, 2024). Sandeels feed preferentially on small copepods, but recent declines in small copepod abundances in the North Sea and changes in timing of copepod biomass accumulation have driven the decrease in overall size and nutritional quality in most (but not all) regions of the western North Sea (Olin *et al.*, 2022, 2025). There are significant implications for higher trophic levels, including fish species such as whiting and Atlantic salmon, which appeared to stop feeding on sandeels in 2004 due to a lack of prey availability (Utne *et al.*, 2022); seabirds including kittiwakes, puffins, and guillemots, which show strong positive correlation between breeding success and sandeel prey availability and abundance (Searle *et al.*, 2023; Régnier *et al.*, 2024); and marine mammals including grey seals, harbour porpoises, and, to a lesser degree, harbour seals (Ransijn, 2023). Sprat, which are of similar ecosystem importance, have a partially overlapping diet with herring at both larval and adult stages (Arrhenius, 1996).

As is the case within the plankton, predator-prey relationships in fish are well described by size, which is demonstrated through the predator-prey interactions between plankton and herring, mackerel, and sandeels at various lifestages described above. This is proven directly by Pinnegar

et al. (2003) and, moreover, growth rates are maximised when prey is optimally sized, highlighting the importance of the planktonic prey size distribution.

1.4.2 Existing research

The plankton in UK coastal seas are diverse, with significant representation from phytoplankton groups such as cyanobacteria (primarily *Synechococcus*), phaeocystis, autotrophic nanoflagellates, coccolithophores, diatoms, and auto- / mixotrophic dinoflagellates (Rodríguez *et al.*, 2000; Widdicombe *et al.*, 2010; Bresnan *et al.*, 2009, 2015). Zooplankton groups are similarly diverse with representation from groups such as nanozooplankton, heterotrophic dinoflagellates, ciliates, copepods, other crustacea (including cladocerans, amphipods and euphausiids), jellyfish, salps, chaetognaths, gastropods, and meroplankton larvae (Bresnan *et al.*, 2015; Wells *et al.*, 2022; Eloire *et al.*, 2010; González-Gil *et al.*, 2022). On the basis that many of the numerically abundant ciliates are non-constitutive mixotrophs, they are included as zooplankton. A notable and worthwhile exception is of *Mesodinium rubrum*, which can use the chloroplasts for much longer, because they inherit the nucleus and other apparatus of their prey cells, as opposed to only the chloroplasts (Myung *et al.*, 2013).

1.4.2.1 Plymouth L4 (PL)

Extensive data are available at PL, which is arguably the most-researched site in UK coastal coastal waters. A full suite of environmental data are available ($2000-2023$), in addition to flow cytometry ($2008-2023$), phytoplankton microscopy ($1994-2023$), microplankton by FlowCAM ($63\,\mu$m mesh; 2015-2020), and mesozooplankton by vertical net haul ($1989-2023$), chlorophyll by fluorometer ($1995-2023$), and chlorophyll by HPLC ($1999-2015$; McEvoy *et al.*, 2023).[1]

Winter

Phytoplankton chlorophyll concentration at PL is at the annual minimum in December-January ($\sim 0.5\,\mu$g chl *a* L^{-1};Rodríguez *et al.*, 2000; McEvoy *et al.*, 2023). During this period, the biomass is dominated by small phytoplankton groups. Autotrophic nanoflagellates (classified as $2\,\mu$m or $5\,\mu$m), measured by fluorescence microscopy, contribute typically around $\sim 7\,\mu$g C L^{-1} biomass (Widdicombe *et al.*, 2010; Rodríguez *et al.*, 2000; McEvoy *et al.*, 2023). The annually-averaged abundance proportion of these flagellates in the microscopy dataset is 87% (Widdicombe *et al.*, 2010). Diatoms appear to contribute negligible biomass during the period, but are reported to be composed of typically larger species in winter than in summer (Widdicombe *et al.*, 2010). The remaining majority of the biomass ($\sim 25\,\mu$g C L^{-1}) is measured by flow cytometry, and contributed by *Synechococcus* ($1.72\,\mu$m), picoeukaryotes ($1.83\,\mu$m), nanoeukaryotes ($5.4\,\mu$m), coccolithophores ($7.68\,\mu$m), and cryptophytes ($5.48\,\mu$m) in roughly similar proportions (McEvoy

[1]The data presented in Rodríguez *et al.* (2000) and McEvoy *et al.* (2023) are reported in units of biomass, whereas the work of Widdicombe *et al.* (2010) and Eloire *et al.* (2010) are reported in abundance. From a community size structure perspective, biomass units are more informative, so a focus was made on the former papers. Rodríguez *et al.* (2000) used data only from October 1992 to January 1994 and so was only used for qualitative validation. All sizes are retrieved from WCO metadata (Tarran & May, 2019; Widdicombe & Harbour, 2021; McEvoy *et al.*, 2022). Phytoplankton sizes are fixed, while zooplankton sizes (by taxon) vary seasonally based on typical measurements. These are found in Tables A.1 & A.4.

et al., 2023). Heterotrophic bacteria contribute a further $\sim 10\,\mu$g C L^{-1} in this period (McEvoy *et al.*, 2023), but are not considered in our analyses. The summed phytoplankton biomass in units of carbon is therefore $\sim 32\,\mu$g C L^{-1}, at a mean ratio of ~ 64 g C (g chl a)$^{-1}$. This is surprisingly large given the time of year when compared to a similar latitude coastal site in Denmark where 20 g C (g chl a)$^{-1}$ is observed (Jakobsen & Markager, 2016). This may be an artefact of the differing sample depths (0 m for chl a, 10 m for microscopy / flow cytometry). Alternatively, the ratio may result from the composition of the nanoflagellate community, which can include mixo- and heterotrophic species. Broadly, the phytoplankton microscopy samples show little variation between November and March on average (Widdicombe *et al.*, 2010), indicating a stable community.

During January the zooplankton community experiences annual minimum biomass with a total $4\,\mu$g C L^{-1} contributed by small copepods, copepod nauplii, and semi-gelatinous predators (McEvoy *et al.*, 2023). Numerically, *Oncaea* spp. (40%; $270.4\,\mu$m), *Pseudocalanus elongatus* (15%; 357.6μm) *Oithona* spp. (12%; $196.2\,\mu$m), and *Paracalanus* spp. (10%; $317.5\,\mu$m) dominate the mesozooplankton abundance in this period. Chaetognaths ($807 - 2244\,\mu$m) and polychaete larvae ($276\,\mu$m) only contribute 2% of abundance but contribute considerably more biomass proportion owing to their larger size. The *Calanus* egg production rate is also at its minimum in December−January, at less than 5 eggs per female on average (McEvoy *et al.*, 2023).

Spring

In Spring, the phytoplankton community changes considerably. Biomass increases to an average of $85\,\mu$g C L^{-1} in April ($40\,\mu$g C L^{-1} recorded by microscopy), reflected by an increase to $\sim 2.5\,\mu$g chl a L^{-1} (Smyth *et al.*, 2010; McEvoy *et al.*, 2023), now expressing an average carbon-to-chlorophyll ratio of ~ 34 g C (g chl a)$^{-1}$. However, inspection of chlorophyll seasonal trends show that the accumulation during this period is highly variable, with an upper bound of $10\,\mu$g chl a L^{-1} observed until 2010 (Smyth *et al.*, 2010). Additionally, average chlorophyll maximum is higher at 50 m (approx $5.5\,\mu$g chl a L^{-1}) than at the surface; this is the only time of year where 50 m chl a is greater than surface chl a (McEvoy *et al.*, 2023). Observations at 10 m below the surface show that biomass is accumulated by nanoeukaryotes, diatoms, *Phaeocystis*, and to a lesser degree, dinoflagellates. *Synechococcus* biomass diminishes (Rodríguez *et al.*, 2000; McEvoy *et al.*, 2023). Within the diatoms, seasonal variability is not explicitly described by Widdicombe *et al.* (2010), although the month of maximum abundance is presented for each taxon. Furthermore, annually averaged numerical abundance is dominated by *Pseudo-nitzschia delicatissima* (19.6%; $4.6\,\mu$m), *Chaetoceros* spp. ($\geq 30.7\%$; $3.4 - 12.6\,\mu$m), and *Leptocylindrus* spp. ($\geq 16.91\%$; $3.3 - 7.2\,\mu$m; Widdicombe *et al.*, 2010). In April−May, *Chaetoceros socialis* ($5.8\,\mu$m), *Skeletonema costatum* ($4.3\,\mu$m), and *Guinardia delicatula* ($22.7\,\mu$m) have the maximum abundance of their seasonal cycle; so do *Phaeocystis* ($6\,\mu$m) and coccolithophores ($5 - 26\,\mu$m; Widdicombe *et al.*, 2010). Furthermore, large diatoms caught in the 63μm mesh net increase in biomass from $1\,\mu$g C L^{-1} to $4\,\mu$g C L^{-1} from March to May (McEvoy *et al.*, 2023). Based on the data presented, it is difficult to draw conclusions on the biomass-dependent size structure, although the maximal abundance of *Guinardia delicatula* and increasing biomass of large diatoms

may drive an increase in mean size (and size variance) when compared to winter.

The zooplankton community is also subject to change into Spring. At the start of Spring, in March, a strong pulse of meroplanktonic cirripede larvae enters the community (contributing 46% of abundance; $510 - 574\,\mu$m), which remains prominent in April (31%) and negligible thereafter (Eloire *et al.*, 2010). This is reflected in biomass contributions of meroplankton ($\sim 4\,\mu$g C L^{-1}; McEvoy *et al.*, 2023). Similarly, fish eggs and larvae (1.5%; $1730 - 2660\,\mu$m) pulse in March (Eloire *et al.*, 2010). Zooplankton community biomass rises to about $30\,\mu$g C L^{-1} by May with significant contributions from ciliates ($\sim 7\,\mu$g C L^{-1}) and small copepods ($\sim 8\,\mu$g C L^{-1}), although all key groups make some biomass contribution, in particular as large copepods and other crustacean plankton (McEvoy *et al.*, 2023; Rodríguez *et al.*, 2000). This is reflected by the changing proportional abundances, with mesozooplankton contributions by *Temora longicornis* (12%; 556.3 μm), the cladoceran *Evadne nordmanni* (16%; 432.6 μm), *Acartia clausi* (7%; 397.4 μm), and increasing abundance of the large copepod *Calanus helgolandicus* (3%; 961.6 μm; Eloire *et al.*, 2010). *Pseudocalanus elongatus* (18%; 389.3 μm) and *Oithona* spp. (16%; 215.0 μm) remain prominent, while *Oncaea* spp. (261.7 μm) is diminished until the autumn (Eloire *et al.*, 2010). This likely has important effects for the biomass-weighted size structure, particularly considering the fact that, per-individual, *Calanus helgolandicus* is expected to have $\sim 15\times$ greater carbon mass than *Pseudocalanus elongatus* and other similar sized copepods. In May and June, the biomass proportion of copepod nauplii is at its annual minimum, despite *Calanus* egg production rate being at its highest (McEvoy *et al.*, 2023).

Summer

In summer, dinoflagellate contributions become very significant with about $50\,\mu$g C L^{-1} in July on average (McEvoy *et al.*, 2023), although with high variability displayed by the measurement of $640\,\mu$g C L^{-1} in July 1993 (Rodríguez *et al.*, 2000). However, it is not clear which component of these are constituted by auto- / mixo- / heterotrophic cells. Although the abundance of heterotrophic dinoflagellates in summer is of the order 10^3 cells L^{-1} and the autotrophic dinoflagellates of the order 10^4 cells L^{-1} (Widdicombe *et al.*, 2010), mean size is typically larger in heterotrophic taxa (Ismael, 2003; Sherr & Sherr, 2007), thus contributing higher biomass per individual. In the instance of July 1993, autotrophic dinoflagellates contributed $500\,\mu$g C L^{-1} while heterotrophs contributed $140\,\mu$g C L^{-1} (Rodríguez *et al.*, 2000), displaying the ability of taxa of both trophic modes to contribute large biomass. However, plastid-containing mixotrophic taxa were likely to be absorbed into the autotrophic dinoflagellates (not discussed in Rodríguez *et al.*, 2000) and, as such, the total heterotrophic activity would extend further than the quoted biomass estimates suggest. Similarly, the taxa *Scrippsiella trochoidea* and *Prorocentrum micans* are described as autotrophic in Widdicombe *et al.* (2010), but are described as mixotrophic in other works such as Jeong *et al.* (2005). In our review of the literature below, it is assumed that dinoflagellates described as 'autotrophic' in the original works are plastid-containing, and *may* possess the ability to exercise heterotrophy. Heterotrophs, on the other hand, are assumed not to contain plastids.

Within the dinoflagellates, average auto- / mixotrophic cell concentrations are 8.2×10^4 cells

L^{-1} and the heterotrophic counterparts are 4.6×10^3 cells L^{-1}, with both groups showing similar degree of seasonal variability (Eloire *et al.*, 2010). In the auto- / mixotrophic portion, annually-averaged numerically dominant taxa are small *Prorocentrum* spp. and *Heterocapsa* spp. (80.1% of counts; $5.4 - 8.9\,\mu$m), and mid-sized *Karenia mikimotoi* and *Scrippsiella trochoidea* (17.6%; $17.8 - 20.0\,\mu$m; Widdicombe *et al.*, 2010). Of the other listed auto- / mixotrophic dinoflagellates, *Prorocentrum micans* (0.4%; $28.2\,\mu$m) contributes to the larger end of the size domain (Widdicombe *et al.*, 2010). With $\sim 28\times$ more carbon per cell than the small *Prorocentrum* spp. (based on sizes presented in Table A.1 converted using Menden-Deuer & Lessard, 2000), this a significant biomass contribution (around one-seventh of the small *Prorocentrum* / *Heterocapsa* spp. biomass). Of the heterotrophic dinoflagellates, small taxa such as *Gymnodinium* spp. and *Katodinium* spp. are the most dominant (36.0% of counts; $7.6 - 14.4\,\mu$m), but are closely followed by the large *Gyrodinium* spp. (30.71%; $32.7\,\mu$m; Widdicombe *et al.*, 2010). The remaining third of proportional abundance appears to be contributed by a wide range of taxa (most of which are not listed; Widdicombe *et al.*, 2010).

In the non-dinoflagellate phytoplankton community, summer biomass is high with around $95\,\mu$g C L^{-1} in July and $105\,\mu$g C L^{-1} in August, on average (McEvoy *et al.*, 2023). In both months, the biomass contributed by micrsocopy is $\sim 30\,\mu$g C L^{-1} exhibiting the dominance of nanophytoplankton and*Synechococcus*, which contribute much of the remainder (McEvoy *et al.*, 2023). Into summer, diatoms remain important ($15 - 20\,\mu$g C L^{-1}; McEvoy *et al.*, 2023) although there is some suggestion of species succession, with three of the four numerically dominant *Chaetoceros* spp. ($3.4 - 12.6\,\mu$m) and *Leptocyndrus danicus* ($7.2\,\mu$m) experiencing annual maximum abundance (Widdicombe *et al.*, 2010). Reducing large diatom biomass captured by the $63\,\mu$m mesh net combined with the emergence of small diatom species suggests that overall diatom mean size may decrease during this period. Similarly, nanoflagellates show little variability from spring to summer, thus contributing significant biomass ($15\,\mu$g C L^{-1}; McEvoy *et al.*, 2023). In the absence of dinoflagellates, there is an indication that phytoplankton cell size is likely to decrease. However, the auto- / mixotrophic dinoflagellates *Karenia mikimotoi* ($20.0\,\mu$m), *Scrippsiella trochoidea* ($17.8\,\mu$m) and *Prorocentrum micans* ($28.2\,\mu$m) experience their annual abundance maxima between June and August (Widdicombe *et al.*, 2010), indicating a potential replacements of diatoms as the larger cells in the phytoplankton community. As a consequence, it is difficult to speculate the progression of the phytoplankton community structure. In August, chl *a* concentrations are approximately $2.3\,\mu$g chl *a* L^{-1} (Eloire *et al.*, 2010; McEvoy *et al.*, 2023) so that C:Chl is between 45 g C (g chl *a*)$^{-1}$ and 67 g C (g chl *a*)$^{-1}$, depending on the proportion of dinoflagellates which are autotrophic (and neglecting kleptoplasts in non-constitutive ciliates / dinoflagellates). This marks a likely increase from spring, regardless of the degree of autotrophy in dinoflagellates.

The zooplankton community shifts in summer with changing contributions from key functional groups. The overall biomass is around $33\,\mu$ g C L^{-1} in the absence of heterotrophic dinoflagellates (McEvoy *et al.*, 2023). Ciliates remain key contributors ($10\,\mu$g C L^{-1}), as do small ($5\,\mu$g C L^{-1}), large ($2\,\mu$g C L^{-1}), and naupliar copepods ($2\,\mu$g C L^{-1}; McEvoy *et al.*, 2023). However,

the contribution of small copepods is reduced compared to spring, with increasing biomass of meroplankton ($4\,\mu$g C L^{-1}), semi-gelatinous ($2\,\mu$g C L^{-1}), and gelatinous predators ($3\,\mu$g C L^{-1}), and other holoplankton groups ($4\,\mu$g C L^{-1}; McEvoy *et al.*, 2023). On the basis of biomass evenness across groups, it appears that the zooplankton community is at its most diverse in summer. The maximum species-specific mesozooplankton abundance is of the copepod *Temora longicornis* (17%; 403.9 μm), while *Pseudocalanus elongatus* (12%; 389.8 μm) and *Oithona* spp. (10%; 207 μm) remain prominent numerically (Eloire *et al.*, 2010). *Calanus helgolandicus* approaches its annual maximum contribution (4%; 857 μm), while *Evadne nordmanni* (8%; 407.3 μm), appendicularians (4%; 1248.0 μm), siphonophores (3%; $811.2 - 1914.0\,\mu$m), echinoderm larvae (8%; 314.6 μm), gastropods (2.5%; $301.9 - 634.9\,\mu$m), and chaetognaths / polychaete larvae (2.5%; $484.4 - 2244.3\,\mu$m) contribute significant abundance (Eloire *et al.*, 2010). Thus, within the mesozooplankton, typical organism size appears to increase. However, the introduction of heterotrophic dinoflagellates − in large biomasses − may reduce the overall zooplankton size.

Autumn

From September to November, phytoplankton biomass reduces considerably from around 85 μg C L^{-1} to 42 μg C L^{-1} (McEvoy *et al.*, 2023). Abundances of most phytoplankton functional groups decreases, although the autotrophic nanoflagellates maintain a greater proportion of biomass by decreasing from around 10 μg C L^{-1} to 8 μg C L^{-1} over the autumn (McEvoy *et al.*, 2023). This reflects the shift back towards the winter regime, which is dominated by small taxa recorded by flow cytometry and nanoflagellates under microscopy. Despite this, a small contribution of large diatoms ($1.5\,\mu$g C L^{-1}) are present throughout the autumn, recorded by FlowCAM (McEvoy *et al.*, 2023). The biomass decrease is reflected by decreasing chlorophyll *a*, from about 2 μg chl *a* L^{-1} to 0.8 μg chl *a* L^{-1} (Eloire *et al.*, 2010; McEvoy *et al.*, 2023; Smyth *et al.*, 2010). None of the numerically dominant diatoms exhibit their annual maximum abundance during this period (Widdicombe *et al.*, 2010). Depending on the proportion of autotrophy in dinoflagellates, the C:Chl ratio is between 42.5 g C (g chl *a*)$^{-1}$ and 55 g C (g chl *a*)$^{-1}$ in September, and around 54 g C (g chl *a*)$^{-1}$ in November.

Dinoflagellates exhibit decreasing biomass during the autumn ($26\,\mu$g C L^{-1} to $2\,\mu$g C L^{-1}; McEvoy *et al.*, 2023), but this period is important for some of the numerically dominant species. Three species of the genus *Prorocentrum* ($6.2 - 9.9\,\mu$m), which contribute 71.4% of the annual autotrophic dinoflagellate abundance combined, exhibit their maximum numerical contribution in September−October (Widdicombe *et al.*, 2010). Similarly, within the heterotrophs, the numerically dominant *Gyrodinium* spp. (32 μm) are at their maximum in September (Widdicombe *et al.*, 2010). The contribution of larger dinoflagellates measured by FlowCAM is $1\,\mu$g C L^{-1} from September−November, thus becoming a significant portion of the biomass by the end of autumn (McEvoy *et al.*, 2023). In the phytoplankton, there appears to be a shift to smaller cells in autumn, but the remaining abundance of larger cells measured by FlowCAM may indicate that the (proportionally-weighted) size structure could still be close to its annual maximum mean and variance.

Zooplankton biomass decreases through the autumn from 18 μg C L^{-1} to 10 μg C L^{-1} (McEvoy

et al., 2023), with a relatively linear progression in community structure towards the winter community. With the exception of small copepods and copepod nauplii, most uni- and multicellular functional groups exhibit steady proportional decreases in biomass from September to November. Within the small copepod community, much of the herbivorous population previously contributed by *Pseudocalanus elongatus* (now 5% of mesozooplankton; 308.1 μm) and *Temora longicornis* (5%; 457.8 μm) appears to be replaced by *Paracalanus parvus* (25% of mesozooplankton; 320.4 μm; Eloire *et al.*, 2010). Similarly, the omnivorous taxon *Ditrichocorycaeus anglicus* (6%; 386.0 μm) emerges in coincidence with the decrease of omnivorous *Acartia clausi* (falling from 10% to 2%; 315.8 μm); meanwhile, *Oncaea* spp. re-emerges over the autumn (from 10% to 35%; 280.0 μm; Eloire *et al.*, 2010). Outwith the copepod community, siphonophores (3%; $1227.4 - 1667.1\,\mu$m) and chaetognaths (1.5%; $618.9 - 1318.1\,\mu$m) remain as important contributors in the zooplankton, while bivalve larvae (3.5%; 303.3 μm) reach their maximum proportional contribution (Eloire *et al.*, 2010). The maintenance of some large (semi-)gelatinous species, combined with the drastic reduction in ciliate and dinoflagellate biomass, presents the possibility that mean size may be higher during this period but with diminished size diversity. A notable anomaly in the transition to the described winter community is that, in December, semi-gelatinous predators appear to have their annual maximum biomass (about 2 μ g C L^{-1}; McEvoy *et al.*, 2023). However, from Eloire *et al.* (2010) it is not clear which taxa contribute this biomass.

1.4.2.2 Stonehaven (SH) & Loch Ewe (LE)

Two long-term plankton observatories are maintained in Scotland, at SH (since 1997) and LE (since 2000). Both sites are well-researched, with publications discussing phyto- and zooplankton community composition on seasonal and interannual scales. However, there is less data for pico- and small nanoplankton than at PL; only a short time series of flow cytometry data is presented at SH ($2015 - 2017$), and none is available at LE. Furthermore, while separate analysis is made at $\times 400$ microscopy to identify periods in which nanoflagellates are more numerically dominant, difficulties associated with this method mean that the quantitative extent of the nanoflagellates community is not well known (Bresnan *et al.*, 2009). Larger nano- and microplankton are enumerated by inverted microscopy, but only a small subsample is recorded (typically ~ 1 mL) with a focus on numerically dominant species[2]. Zooplankton are enumerated from 200 μm mesh net samples; although a 63 μm mesh net haul is taken, the samples are not yet analysed. It is therefore likely that the zooplankton community is subject to under-representation in the $63 - 200\,\mu$m size interval. Observed community composition is similar at both sites, although biomass is typically slightly higher at LE through the year, and particularly during the autumn bloom is which is a more prominent feature of the seasonal cycle (Bresnan *et al.*, 2015). As such, the SH

[2]Phytoplankton sizes are first retrieved from Widdicombe & Harbour (2021), and subsequently from the supplementary material of González-Gil *et al.* (2022). Sizes in the latter were obtained from the literature, so it was decided that measurements made at PL were likely to be more applicable. Conversely, for zooplankton, SCObs metadata retrieved from the literature were used (transformation between units made using Harris, 1999 where necessary). This is because the style of data reporting is different at SCObs compared to WCO, in particular with respect to copepods which are recorded by lifestage. Phytoplankton size data are found in Table A.2, while zooplankton data are found in Table A.3.

and LE are discussed in conjunction here, with key differences highlighted.

Winter

In winter, biomass at both SH & LE is low ($\sim 0.2\,\mu$g chl a L^{-1}; Bresnan *et al.*, 2015). At SH, estimated phytoplankton biomass is around $8\,\mu$g C L^{-1} (inclusive of flow cytometry) with a reported dominance of small cells (González-Gil *et al.*, 2022). This is validated by the reported low concentrations of diatoms and dinoflagellates at both sites ($< 10^4$ cells L^{-1}; Bresnan *et al.*, 2015) combined with the reported carbon and chlorophyll a concentrations. Within each of these functional groups, the abundances are small such that little is reported about the species / size composition of diatoms and dinoflagellates in winter. The resultant ratio is 40 g C (g chl a)$^{-1}$ at SH. Meanwhile, the zooplankton biomass in winter is also composed of low biomass (Bresnan *et al.*, 2015).

Spring

In spring, the plankton community exhibits biomass accumulation and structural changes. At SH, the maximum phytoplankton biomass accumulation rate of around 0.05 d^{-1} typically occurs at the start of April, during which period carbon biomass is estimated at approx. $50\,\mu$g C L^{-1} (González-Gil *et al.*, 2022). At SH, chlorophyll concentration is around $2.2\,\mu$g chl a L^{-1}, whereas it is significantly higher at LE ($3.3\,\mu$g chl a L^{-1}; Bresnan *et al.*, 2015). This reflects the earlier accumulation of biomass at LE to the annual chlorophyll maximum in April; at SH, the peak chlorophyll is normally in May, and can remain until early June (with median of $3\,\mu$g chl a L^{-1}; Bresnan *et al.*, 2015). By this point, the seasonal average carbon biomass at SH approaches the annual maximum of $130\,\mu$g C L^{-1}, although observations in excess of $300\,\mu$g C L^{-1} are common. The accumulation of biomass is reflected by increasing concentrations of diatoms and, to a lesser numerical degree, dinoflagellates at both Scottish sites (Bresnan *et al.*, 2015). In the $2015-2017$ findings presented by González-Gil *et al.* (2022), the peak seasonal carbon biomass is expected in May at SH, in line with reported chlorophyll a observations by Bresnan *et al.* (2015). Cell abundance plots presented for SH in Bresnan *et al.* (2009) indicate a shift from typical *Skeletonema* spp. ($0.5-1\times10^5$ cells L^{-1}; $9.0\,\mu$m) and *Thalassiosira* spp. ($0.5-5\times10^4$ cells L^{-1}; $2.5-34.6\,\mu$m) dominance in March-April to *Chaetoceros* spp. ($0.5-1\times10^6$ cells L^{-1}; $2.5-24.8\,\mu$m) in May across the period $1997-2006$. However, May $1998-2001$ exhibited much higher *Chaetoceros* spp. abundances (and overall diatom abundances) than May in $2002-2006$, indicating qualitative differences in the biomass and species succession during spring across different time periods (Bresnan *et al.*, 2009). Temporal variability on the longer-scale means that results from Bresnan *et al.* (2009), Bresnan *et al.* (2015), and González-Gil *et al.* (2022) should be compared with caution. Indeed, across the $2003-2012$ period at SH, the diatoms reduce in May on average, from 1.6×10^5 cells L^{-1} in March / April to 0.5×10^5 cells L^{-1} (Bresnan *et al.*, 2015). Regardless, the shift from *Skeletonema* spp. and *Thalassiosira* spp. to *Chaetoceros* spp. may indeed drive an increase in the mean size of phytoplankton from March to May, on average, depending on the composition of *Chaetoceros* spp. Furthermore, at SH, the introduction of mixotrophic dinoflagellates of the genus *Alexandrium* in April ($1-5\times10^3$ cells L^{-1}; $24.7-44.6\,\mu$m ESD; Bresnan *et al.*, 2009) are likely to drive an increase in mean size and size variance, although

the timing and size of this aggregation is variable across years. Dinoflagellates also accumulate in similar number at LE during April, although less detail is published about the composition of the species and community structure. At SH, the C:Chl ratio is approximately 50 g C (g chl $a)^{-1}$ in spring.

In line with the phytoplankton community, the mesozooplankton community shows increase in biomass at both SH and LE in spring (Bresnan *et al.*, 2015). However, the LE biomass is typically significantly higher than at SH until July; community dry weight (DW) reported in Bresnan *et al.* (2015) are 60 μg DW L^{-1} vs 20 μg DW L^{-1} in April and 85 μg DW L^{-1} vs 25 μg DW L^{-1} in May, for LE and SH respectively. Depending on the proportion of gelatinous plankton in the community, this corresponds to $8.5 - 38$ μg C L^{-1} at LE in May (using conversion from Harris, 1999). The community is dominated by crustacean species (Bresnan *et al.*, 2015), so estimates are likely to be much closer to the upper boundary.

The seasonal succession of the zooplankton species at SH & LE is not explicitly presented in Bresnan *et al.* (2015), but the overall properties of the community are described in detail. Key species in the community at both SH & LE are *Pseudocalanus* spp. (484.2 μm at C6F), which are the most numerically dominant, as well as *Calanus helgolandicus* (1207.3 μm at C6F) and *Calanus finmarchicus* (1465.3 μm at C6F) in lower abundances (Bresnan *et al.*, 2015). At SH, *Calanus finmarchicus* are known to exit their overwintering sites in the Faroe-Shetland Channel and arrive at feeding sites, such as SH, in spring (Heath *et al.*, 1999). However, in their review and discussion, Bresnan *et al.* (2015) found that less is known about the behaviour of *Calanus finmarchicus* overwintering strategies in the west coast of Scotland and therefore at LE; similarly, the overwintering behaviour of *Calanus helgolandicus* is not well understood. At LE, *Acartia clausi* (449.0 μm at C6F), *Temora longicornis* (669.9 μm at C6F), *Oithona* spp. (359.8 μm at C6F), and *Paracalanus parvus* (416.7 μm at C6F) have also been studied with respect to overall declines in abundance from 2011 to 2017 (Wells *et al.*, 2022). Other important phytoplankton consumers in the community at both sites include appendicularians (1323 μm), bivalve larvae (392.7 μm), and cirripede nauplii (351.6 μm); at LE, cladocerans ($286.4 - 1419.7$ μm) also contribute significantly to the community (Bresnan *et al.*, 2015). By July, the maximum median grazer biomass is achieved at both sites ($\sim$ 90 μg DW L^{-1}), however, biomass is retained for longer after this point at LE, while at SH it diminishes approximately linearly between August and October (Bresnan *et al.*, 2015).

Summer

In summer, the chlorophyll *a* biomass is reduced at both sites. In July, the concentration is around 1.4 μg chl *a* L^{-1} at both sites based on $2003 - 2012$ data (Bresnan *et al.*, 2015). Over the $2015 - 2017$ period at SH, the chlorophyll *a* is also reduced compared to spring, although the fitted value is closer to 2.1 μg chl *a* L^{-1} (González-Gil *et al.*, 2022). However, it appears from the fitted GAM and its residuals that the chlorophyll *a* data are skewed and, as such, the chlorophyll *a* concentration is likely to be overestimated by the model. Regardless, over this period, the carbon biomass (which is $\log_{10}$ transformed, and appears less likely to violate model assumptions) remains high through summer, only dropping below 100 μg C L^{-1} in mid-August

on average (González-Gil *et al.*, 2022). At SH, the C:Chl ratio remains at approximately 50 g C (g chl a)$^{-1}$, based on the GAM fitted values in González-Gil *et al.* (2022).

A succession is observed at both SH & LE, with dinoflagellates becoming increasingly important in the community.[3] At both sites, dinoflagellate abundances rise to a median of approximately 1×10^4 cells L^{-1} between June and August. Cells of the genera *Tripos* (formerly *Ceratium*; 29.7−47.4 μm), *Protoperidinium* (17.3−95.3 μm), and *Dinophysis* (17.32−59.6 μm) are abundant, while occasional large blooms of *Karenia mikimotoi* (29.1 μm) occur at LE (Bresnan *et al.*, 2015). Specific plots from 1997-2006 show that *Protoperidinium* spp. tend to have a more variable seasonal cycle at SH, with annual peaks observed between May and September, whereas *Tripos* consistently peaks in July-August (Bresnan *et al.*, 2015). Dinoflagellates were not separated by trophic mode in Bresnan *et al.* (2015), but *Tripos* spp. are mixotrophic (Smalley *et al.*, 2003) and *Protoperidinium* spp. are considered microzooplankton (Table A.1). Overall diatom abundances remain significant, with July−August abundances around 1.6×10^5 at LE and 1.3×10^5 at SH (Bresnan *et al.*, 2015). At SH, *Skeletonema* spp. made no significant contribution in summer between $1997 - 2005$ (very small accumulation in 2006), whereas *Chaetoceros* spp. and *Thalassiosira* spp. made some sporadic contributions (Bresnan *et al.*, 2009).

Autumn

Autumn is perhaps where the LE and SH communities show the greatest degree of difference. At LE, an autumn diatom bloom is common in late August / September, with median chlorophyll a observations $\sim 2.2\,\mu$g chl a L^{-1} (Bresnan *et al.*, 2015). This is reflected by the pulse of diatoms to a median abundance of 2.2×10^5 cells L^{-1} alongside decreasing dinoflagellate abundance (Bresnan *et al.*, 2015). Although the autumn bloom is not a common feature of the cycle at SH, there is some evidence of rare occurrence in González-Gil *et al.* (2022). In most years, when there is no positive accumulation of biomass, much of the summer biomass is maintained in September. Then at both sites, from October onwards, biomass is drastically reduced, to $\sim 1\,\mu$g chl a L^{-1} (Bresnan *et al.*, 2015). At SH, over the $2015 - 2017$ period, this is equivalent to around 40 μg C L^{-1} in October and 12 μg C L^{-1} by the end of November (with around 0.3 μg chl a L^{-1}; González-Gil *et al.*, 2022). By this point, diatoms and dinoflagellates have very low abundances Bresnan *et al.*, 2015, indicating that the majority of the biomass is contributed by small plankton (either bacteria and pico- / nanoeukaryotes recordable by flow cytometry, or nanoflagellates recordable by fluorescence microscopy / ×400 inverted microscopy). A decrease in mean size and size variance is therefore anticipated.

At both SH & LE the zooplankton community is seen to change in late summer and into autumn, with increasing proportional biomass of carnivorous predators in the community (Bresnan *et al.*,

[3]The published results appear to show some disagreement about the extent of the dinoflagellate abundance. For example, dinoflagellate concentrations $\geq 1 \times 10^4$ are observed in July − September 2006 in Bresnan *et al.* (2009), yet in the box-whisker plot of Bresnan *et al.* (2015), these values are not seen or plotted as outliers. However, based on the approximate subsample volume of 1 mL used in analysis and a subsequent minimum recordable abundance of 1×10^3 cells L^{-1}, the Bresnan *et al.* (2015) results are one order of magnitude smaller in this instance than this limit would dictate. We therefore work with a ×10 correction with the Bresnan *et al.* (2015) results to reconcile the information with Bresnan *et al.* (2009).

2015). At both sites, in September, the herbivorous grazer community begins to decrease in biomass (median 65 μg DW L^{-1} at LE, 40 μg DW L^{-1} at SH) while the carnivore community increase to their annual maximum (10 μg DW L^{-1} at LE, 8 μg DW L^{-1} at SH; Bresnan *et al.*, 2015). The most important taxa driving the increase in carnivore biomass are hydrozoan and anthozoan cnidarians ($775.2 - 2855.8\,\mu$m), ctenophores (2855.8 μm), and chaetognaths (996.1 μm; Bresnan *et al.*, 2015). Most of these taxa are therefore large mesozooplankton or macrozooplankton. Combined with the importance of the microzooplankton such as *Protoperidinium* at this time (Bresnan *et al.*, 2009), it is likely that predator size diversity is close to or at its maximum. By November at SH and December at LE, the overall biomass of the zooplankton community has returned to its minimum (Bresnan *et al.*, 2015).

1.4.2.3 Summary

Each of the sites are characterised by low phytoplankton biomass in winter, a spring bloom, followed by lower (but elevated) biomass in summer. A late-summer / autumn bloom is also observed at PL (Eloire *et al.*, 2010; Rodríguez *et al.*, 2000) and LE, but not normally at SH (Bresnan *et al.*, 2015). Phytoplankton community biomass changes are broadly tracked by changes in the zooplankton community; at PL, the zooplankton abundance responds quickly to phytoplankton abundance, emulating spring and autumn blooms (Eloire *et al.*, 2010), while biomass accumulation of zooplankton is more sinusoidal at SH and LE, with annual maximum in July (Bresnan *et al.*, 2015). In winter, the P:Z biomass ratios are typically at their highest, suggesting that top-down controls are not limiting in the community in line with the DRH (Behrenfeld, 2010; Behrenfeld & Boss, 2014). This increases through spring and into summer, where top-down control is more likely to control biomass and influence phytoplankton size structure.

Similar patterns for the cycle of plankton size are reported across sites. González-Gil *et al.* (2022) report that cells $< 10\,\mu$m dominate the phytoplankton in winter at SH, which is validated at SH & LE with the very low observed diatom / dinoflagellate abundances (Bresnan *et al.*, 2009, 2015). Similarly, at PL, pico- and small nanoplankton dominate the community with minimal diatom and dinoflagellates contributions (Widdicombe *et al.*, 2010; McEvoy *et al.*, 2023). Therefore, the phytoplankton mean size and size variance is likely to be low in winter at all UK sites. The biomass of the small pico- and nanoplankton recorded at PL appears to be much larger than at SH, reflected by both carbon biomass and chlorophyll *a* measurements. Although flow cytometry is not available at LE, the winter diatom / dinoflagellate community is similar to that at SH; combined with the fact that chlorophyll *a* concentrations are almost equal, it is likely that the community structure is similar.

During the winter period, the zooplankton biomass is very low (and at its lowest proportional biomass of the total plankton assemblage). It is reported that this biomass is primarily contributed by crustacean herbivores at all sites. Small copepods of the genus *Pseudocalanus* are mentioned at all three coastal sites, and copepods of *Oncaea* spp. are also reported as important at PL. Although there are negligible contributions by heterotrophic dinoflagellates and ciliates, which are considerably smaller than *Paracalanus* spp. and *Oncaea* spp., the overall mean size is

likely to be low. Additionally, the size variance is likely to be low due to the lack of microzooplankton and mesozooplankton biomass contributions.

The spring bloom at all three sites is prominent feature of the seasonal cycle. At Scottish sites, diatoms appear to accumulate significant abundance by the end of March ($\sim 1.3 \times 10^5$ cells L^{-1}; Bresnan *et al.*, 2015), whereas at PL the accumulation is normally in mid-late April ($\sim 4 \times 10^5$ cells L^{-1}; Widdicombe *et al.*, 2010). At all sites, relatively small diatoms of the genera *Skeletonema* and *Chaetoceros* are abundant. Additionally, there is some evidence of increasing large diatom contributions at PL, in the form of *Guinardia delicatula* and large diatoms captured by a 63 μm mesh net. At PL, phaeocystis and coccolithophores also contribute their annual maximum biomass (McEvoy *et al.*, 2023). These are not explicitly reported at Scottish sites, despite flow cytometry data being available at SH. Overall, at all sites, the mean size is expected to be increasing due to increased proportional biomass of diatoms and other larger taxa, but not yet at its maximum. Size variance is also likely to be higher, since there is evidence (at PL) that the pico- and small nanoplankton community remain important, and as such the introductions of larger taxa is complementary. It is possible that this is the period of maximum size variance, since the proportional contributions across the size domain are relatively even.

The zooplankton community also accumulates significant biomass in spring. At PL, this is dominated by a large pulse of cirripede larvae / nauplii, which are notably larger on average than the dominant crustacean taxa of the winter. Although not explicitly displayed, it is reported that the seasonal cycle of meroplankton is similar at Scottish sites to PL (Bresnan *et al.*, 2015). The large copepod *Calanus helgolandicus* (and *Calanus finmarchicus* at SH & LE) also enter the community from overwintering at all three sites. By the end of Spring at PL, copepod nauplii have their annual minimum biomass, however the protozooplankton community accumulates significant biomass through ciliates. At SH & LE, ciliates are rarely mentioned and no abundances are reported. With the effective absence of heterotrophic dinoflagellates until summer at Scottish sites, it is unlikely that the microzooplankton are important grazers in spring at SH & LE. At all sites, the mean size may be expected to increase due to the influence of meroplankton larvae. At PL, the increase in size variance may be driven by ciliate accumulations, whereas it would appear that size variance is unlikely to change significantly at Scottish sites.

In summer, dinoflagellates are important at all UK coastal sites. There is strong evidence that obligate phototrophic (including facultative heterotrophic) species and obligate heterotrophic species are significant contributors, however, dinoflagellate abundance / biomass is primarily reported as one group. On the assumption that both trophic classes of dinoflagellates contribute biomass through the summer and early autumn (supported by observations), then the phytoplankton community is likely to experience an increase in mean size. At PL, the dinoflagellates appear to be more dominant than at SH & LE from the presented results, and thus may force a decrease in phytoplankton size variance.

Similarly, the zooplankton community becomes more complex at UK coastal sites. The community appears to become more diverse, with accumulations of all locally important functional groups. Copepods appear to remain important, but to a lesser degree than before. *Calanus*

finmarchicus is reported at its highest proportional abundance in summer at PL; combined with contributions from large (semi)gelatinous predators, and copepod nauplii, dinoflagellates, and ciliates in the microzooplankton, an increase in size variance is expected. However, it is unclear how mean size may be affected. A multimodal biomass distribution is expected in the zooplankton.

During late summer / early autumn, a second chlorophyll *a* biomass peak is exhibited at PL & LE, but not normally at SH. At LE, it is reported that the autumn bloom is dominated by larger diatoms, indicating a possible increase in mean size. The effect on size variance is unknown since flow cytometry data are not available. At PL, the second biomass peak is normally equal to or larger than the spring bloom. On the basis that plankton measured by microscopy are declining after July and those by flow cytometry are increasing until September, mean size at PL may be expected to decrease. However, since dinoflagellates are still the most dominant of the large nano- and microplankton, and not explicitly separated into phyto- and zooplankton, it is difficult to predict how phytoplankton mean size and size variance will respond in this period. Although a bloom is not exhibited at SH, biomass only drops significantly in October and thereafter. However, in September the diatom abundance is typically lower in SH than in August, whereas at LE it is higher in September. In early autumn it appears that nuances between sites are therefore most profound; dinoflagellates dominate the microplankton at PL, diatoms at LE, and across-the-board decreases at SH. After this point, diatoms and dinoflagellate both diminish through autumn at all UK coastal sites, indicating a likely decrease in mean size and size variance as the community returns effectively linearly to its winter state of low mean size and low size variance.

The zooplankton community declines in biomass through autumn, with evidence that the protozooplankton (which are important at the start of autumn in early September) are the first group to show significant loss of biomass. This may be linked to the increase in carnivore biomass observed at both SH & LE; it may also be true at PL with gelatinous predators maintaining biomass until the end of October. Changes are presented in the small copepod community in PL, although there does not appear to be a clear influence on size structure. The proportional abundance of large copepods decreases. A reduction in dinoflagellates at all sites, and further of ciliates at PL, may lead to an increase in mean size until the large carnivorous predators also diminish in biomass, at which point the mean size may reduce to its winter level.

1.5 Conclusion

Plankton are a crucial component of marine ecosystems globally due to their contribution to primary productivity, facilitation of energy transfer through the food web, and carbon export. food web dynamics are particularly important in regions of economic significance, such as the UK coastal seas, where fisheries are a major industry. Despite a long-running history of data collection and research at three distinct UK coastal sites, the functional ecology of the planktonic ecosystem is not fully understood.

Until now, most studies in UK coastal sites have focused on the seasonality of important groups, e.g., diatoms and copepods, and their relationship with the environment. However, to better understand the community dynamics and factors limiting production and energy transfer, a size-based perspective is required. Size is described as a 'master trait' in plankton ecology as it governs key ecological processes such as maximal growth rates, resource affinity, and selectivity in predator-prey interactions. By identifying and understanding the properties of the seasonal cycle that control plankton size, understanding of the UK coastal ecosystem can be improved. This information can also be used to progress research efforts at other coastal sites or on a regional / global scale.

A reasonable degree of qualitative knowledge exists on the seasonal cycle of plankton community size structure in the UK coastal waters. The phytoplankton community is dominated by small cells ($< 5\,\mu$m) in winter, while larger cells such as diatoms and dinoflagellates contribute significant biomass from roughly March−September. These larger cells provide food for larger, important species such as *Calanus helgolandicus*, but also concur with the growth of non-crustacean zooplankton including ciliates, cnidarians, chaetognaths, and larval meroplankton. However, as yet, there has been no quantitative assessment of the size structure of the plankton community at any of the UK coastal sites. Since extensive observational data are available for plankton, often recorded to species level, it is feasible to construct a typical seasonal cycle for plankton community size structure. This can be used to improve the understand the ecosystem itself, but can also be used to guide theoretical modelling efforts, which can be used to identify the important processes controlling the ecosystem (e.g., light and prey availability).

Other important features of the functional ecology of the UK plankton community may also be better understood with further research. For example, most of the reporting at all UK coastal sites are provided in units of abundance. From the perspective of carbon export potential and energy transfer, a biomass-centric approach is preferable (particularly in combination with information on size structure). Furthermore, key species reported on an abundance basis are inherently subject to bias, since large organisms are generally less abundant than small organisms, even if they contribute significant biomass. A conclusive picture of the biomass cycle of phyto- and zooplankton is therefore an opportunity to improve understanding, and make direct comparison between UK coastal sites within a single framework. Species can be described by biomass contribution to the community, as opposed to numerical contribution.

A further aspect of existing reporting at UK coastal sites is that the dinoflagellates − which are one of the most important functional groups within the plankton − are often reported as a single group. However, dinoflagellates themselves are functionally diverse, with auto-, mixo-, and heterotrophy exhibited by different taxa within the dinoflagellates. Although some numerically dominant species of 'phyto-' and 'zooplankton' are reported in the literature, there is limited understanding of how representation of different trophic modes within the dinoflagellates evolves over the course of the season. Furthermore, this makes estimation of biomass of phyto- and zooplankton more challenging, since the proportion of primarily auto- / heterotrophic dinoflagellates is not known. Thus, when constructing biomass estimates, dinoflagellates may best be split

by trophic mode. In the context of a size-based study, these are particularly important features since the majority of predator-prey interactions in plankton are governed by size.

Perhaps a reasonable argument would be that, under the outlined circumstances regarding biomass, a chlorophyll *a*-based approach to understanding biomass may be favourable. However, this itself is problematic since the carbon-to-chlorophyll biomass ratio is known to vary across spatial and seasonal scales. A quantitative and mechanistic analysis of the C:Chl ratio at UK coastal sites may help to improve understanding of how the phytoplankton community respond to light, which is highly variable across the year. Additionally, the intracellular shading effect that occurs within large phytoplankton means there is likely a complex tradeoff that occurs between chlorophyll, light availability, and plankton size. Therefore, a size-based perspective which acknowledges variations in cellular stoichiometry and the complexity of the food web would provide a valuable insight into ecosystems of the UK coastal seas.

Chapter 2

Estimation of plankton community size structure in UK coastal waters

2.1 Introduction

In Chapter 1, we outlined the importance of plankton organism size (Litchman *et al.*, 2007; Hillebrand *et al.*, 2022a) and discussed the existing wealth of research undertaken in UK coastal seas. By doing so, we identified a key knowledge gap in the current understanding of this ecosystem: no known work has explicitly investigated plankton community size structure in UK coastal waters. Filling this knowledge gap may have potential implications at a range of scales. Firstly, presenting and investigating the cycle of plankton community size structure in a temperate coastal ecosystem can contribute to existing research in the field of plankton ecology, since mechanisms driving size structure at a community level are still debated. Secondly, the composition of the plankton community affects the prey available to fish and other planktivores, which may then have implications at a wider level, including for fisheries − a crucial component of the UK economy (UK Marine Management Organisation, 2024) − and in the context of climate change.

The UK coastal region has a wealth of data available for the purpose of investigating plankton community size structure. Three long-term monitoring sites (PL, SH, and LE), each with over 20 years of weekly plankton samples and environmental data are available. Furthermore, oceanographic models such as the Scottish Shelf Model (De Dominics *et al.*, 2018) and West of Scotland Coastal Ocean Modelling System (Aleynik *et al.*, 2016) may be used to complement these data, or help to fill in gaps in observations. In this chapter, we make use of the available data to establish the patterns of plankton community size structure in UK coastal waters.

2.1.1 UK coastal waters: a dynamic ecosystem

A key feature of the UK coastal ecosystem is the strong seasonal changes in environmental conditions such as light and wind. In association with seasonal changes in the environment, the

plankton community biomass shows strong variability (Bresnan *et al.*, 2015; González-Gil *et al.*, 2022; McEvoy *et al.*, 2023), indicating that the processes which control the plankton community vary over the course of the year. Furthermore, the key functional groups, which occupy different regions of the organism size domain, show substantial seasonal variability (McEvoy *et al.*, 2023). Tradeoffs are therefore dynamics in this system, which is reflected by seasonal variability in phytoplankton cell size (González-Gil *et al.*, 2022).

The UK coastal plankton observatories are situated such that there are notable differences between each site. PL & SH have higher sunshine hours than LE, as well as less rain (Met Office, 2020). On the other hand, temperatures are generally warmer at PL (Widdicombe *et al.*, 2010; McCluskey *et al.*, 2022) than at SH & LE (Bresnan *et al.*, 2015). Indeed, one interesting difference that arises between sites is the greater chlorophyll *a* concentration at PL in winter (McEvoy *et al.*, 2023), which is more than double that of SH & LE (Bresnan *et al.*, 2015); yet, this difference subsides in spring. As such, we anticipate that the mechanism which controls this difference in winter biomass may no longer control the community in spring, or is counteracted by some other effect.

Based on the significant seasonal variability in the environment and plankton biomass (González-Gil *et al.*, 2022; McEvoy *et al.*, 2023), we focus on the seasonal cycle of plankton community size structure, and associated differences between sites. Understanding the patterns and subsequently the mechanisms which control plankton in UK coastal sites at a seasonal scale is the natural first step and creates a baseline. Future works may then wish to use this baseline to assess interannual or long-term variability.

2.1.2 Predictions about size structure from existing UK coastal research

Based on the existing research undertaken at UK coastal sites (presented in Section 1.4.2), the following pattern predictions are made about the seasonal cycle of the plankton community in UK coastal waters.

P1. As a result of dominance of small cells in the phytoplankton community in winter (González-Gil *et al.*, 2022; McEvoy *et al.*, 2023):

   (i) Mean cell size will be small

   (ii) Size variance will be small

P2. Since there are low biomass contributions by microzooplankton at PL in winter (McEvoy *et al.*, 2023):

   (i) Mean organism size will be large

   (ii) Size variance will be small

P3. The introduction of larger phytoplankton taxa such as diatoms in spring (Bresnan *et al.*, 2009; Widdicombe *et al.*, 2010; Bresnan *et al.*, 2015), accompanied by the continued contributions of pico- and small nanophytoplankton (González-Gil *et al.*, 2022; McEvoy *et al.*, 2023) mean that:

(i) Mean cell size will increase

(ii) Size variance will increase

P4. The immigration of meroplankton larvae and *Calanus* spp. in spring (Eloire *et al.*, 2010; Bresnan *et al.*, 2015; McEvoy *et al.*, 2022), accompanied by the increasing contributions of microzooplankton, means that:

(i) Mean organism size is likely to stay constant

(ii) Size variance will increase

P5. As a result of the general shift from diatoms to dinoflagellates in the phytoplankton community over summer (Rodríguez *et al.*, 2000; Bresnan *et al.*, 2009; Widdicombe *et al.*, 2010; Bresnan *et al.*, 2015; McEvoy *et al.*, 2022):

(i) Mean cell size is likely to stay constant

(ii) Size variance may increase associated with an apparent increase in taxonomic diversity

These pattern predictions have the intention of being easily testable using observational data. Later in this chapter, we subsequently generate mechanistic hypotheses (labelled H) which aim to understand the drivers of these (and other) observed patterns. The mechanistic hypotheses are not directly testable in the same way, instead relying on investigation through mechanistic modelling, and subsequent discussion.

2.1.3 Techniques for assessing size structure

The various techniques available for estimating the relationship between size and abundance / biomass are discussed in Section 1.2.3. The assessment of size structure through the fitting of statistical models to the abundance-size relationship is insightful at a broad-scale. It is particularly powerful in the context of the wider ecosystem, since the size domain can be extended to larger organisms, and slopes can be meaningfully compared. However, this approach lacks the precision to detect more subtle change in size structure, or changes in distributional shape. These may be changes that can be identified by visual assessment of the distribution of abundance / biomass by size, or through statistical summaries of the distribution. In particular, the size variance – which can be used as a proxy for size diversity, and thus compared to other diversity indices – is not discernable from traditional analyses of size spectra.

Given that the focus of our work is a detailed assessment of plankton community size structure, we utilise an approach which communicates changes to size structure directly and clearly. The method of Schartau *et al.* (2010) is therefore an optimal method to do so, as it provides an approach to construct abundance density estimates by plankton organism size. These abundance density estimates can then be converted to biomass density estimates by using a carbon conversion (e.g. Menden-Deuer & Lessard, 2000). Then, there are two uses of these biomass distributions: they can be assessed directly (e.g., García-Oliva *et al.*, 2022), or summarised to descriptive statistics such as mean size and size variance (e.g., Acevedo-Trejos *et al.*, 2015), or even higher distributional moments such as skewness and kurtosis (e.g. García-Oliva *et al.*, 2022).

The advantage of computing summaries is that they describe the community in a low number of dimensions, and subsequently can be used in the fitting of statistical models which directly describe the community traits of interest.

Once the seasonal cycle of plankton community size structure is established, a central objective of this thesis is to investigate the mechanisms which drive size structure. The first step of this process is to produce hypotheses based on the observations and some theoretical tradeoff. For example, it is known that the phytoplankton biomass is lower at SH & LE than PL in winter, in line with lower temperatures. Since maximal growth rates are affected by temperature, a reasonable hypothesis may be that low temperatures restrict growth, and subsequently biomass, at SH & LE. However, other mechanisms (e.g. top-down) may drive this and, as such, we aim to produce as many hypotheses as possible, which can be supported by observations and some theoretical tradeoff. Such an approach aligns with the fundamental ethos of 'strong inference' (Platt, 1964).

2.1.4 Aims and objectives

Using long-term observational data at three UK coastal sites, we aim to:

1. Assess the environment, with a focus on potential influences on the plankton community (e.g., temperature, light).
2. Establish the seasonal cycle of plankton community size structure by:
   (a) Presenting estimates of biomass distributions as a function of size.
   (b) Computing estimates of summary statistics such as biomass, mean size, and size diversity, and subsequently fitting statistical models to these data.
3. Reflect on predictions derived from previous works in UK coastal sites, outlined in Section 2.1.2
4. Construct hypotheses about the mechanisms driving size structure based on observations of plankton community size structure and theoretical tradeoffs.

2.2 Methods

2.2.1 Data

2.2.1.1 Overview of data sources

Data were downloaded from online repositories. At PL, data were sourced through the Western Channel Observatory (WCO) website and downloaded from the British Oceanographic Data Centre (BODC) website. The following data were used: phytoplankton (and protozooplankton) abundance (Widdicombe & Harbour, 2021), zooplankton abundance (McEvoy *et al.*, 2022), flow cytometry (Tarran & May, 2019), CTD (Fishwick, 2018), surface irradiance (Western Channel Observatory, 2020), nutrients (Woodward & Harris, 2023), and chlorophyll *a* (Wilkinson *et al.*,

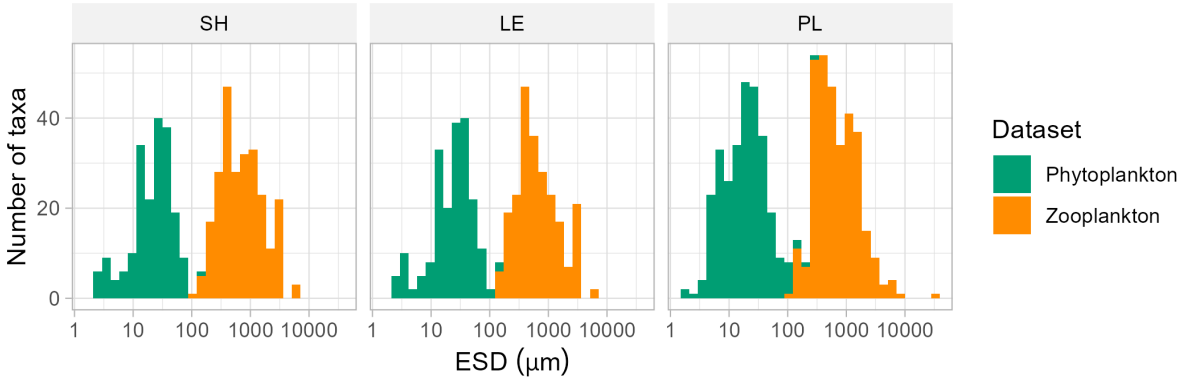


Figure 2.1: Histogram of organism sizes at SH, LE, and PL from the 'phytoplankton' (inc. microzooplankton) and zooplankton microscopy datasets. Organism sizes can be found in Tables A.1−A.4.

2023). Size data for PL were supplied alongside abundances for both phytoplankton (Table A.1; Widdicombe & Harbour, 2021) and zooplankton (Table A.4; McEvoy *et al.*, 2022). Both of these size datasets were based on typical sizes for each taxa using measurements of a subset of organisms of each taxon at PL. For zooplankton, the allocated organism sizes were also resolved by metrological season.

At SH & LE, data were accessed through the Scottish Coastal Observatory (SCObs) website. Flow cytometry (at SH only), phytoplankton abundance, zooplankton abundance, chlorophyll *a* & nutrients, sea surface temperature, and Secchi depth data were directly downloaded for SH (Marine Scotland Science, 2018b) & LE (Marine Scotland Science, 2018a). Zooplankton sizes at SH & LE were allocated using a compiled trait list from a combination of literature sources and measurements of sampled taxa (Table A.3; Scottish Government, Marine Directorate, 2024). These were provided in units of dry weight and were converted to ESD through using conversion factors by functional group (Table 2.3; Harris, 1999). Although the sizes were not resolved seasonally, identification of organisms was to a higher precision, often including lifestage. Phytoplankton sizes were retrieved from the literature by González-Gil *et al.* (2022); this list of sizes was extended using sizes from Widdicombe & Harbour (2021) and Cloern (2018; pers. comm.) where necessary (Table A.2).

Frequency distributions of taxa by size, are shown for each site in Figure 2.1. A clear pattern across all sites is the trough in observations around 100 μm. This region of the size domain is likely undersampled, because the cellular abundances are too low for the 'phytoplankton' sampling method, but the organisms are too small to be captured by the standard 200 μm net in the 'zooplankton' dataset.

Surface irradiance was not directly measured at SH or LE, so measurements from Dyce Airfield (57.21°N 2.21°W; ∼ 28 km from SH; Station 00161) and Kinlochewe (57.61°N 5.31°W; ∼ 31 km from LE; Station 00066) were used as a nearest estimate (Met Office, 2023). Furthermore, temperature, salinity, and density data at SH & LE were recorded at near-surface and near-bed

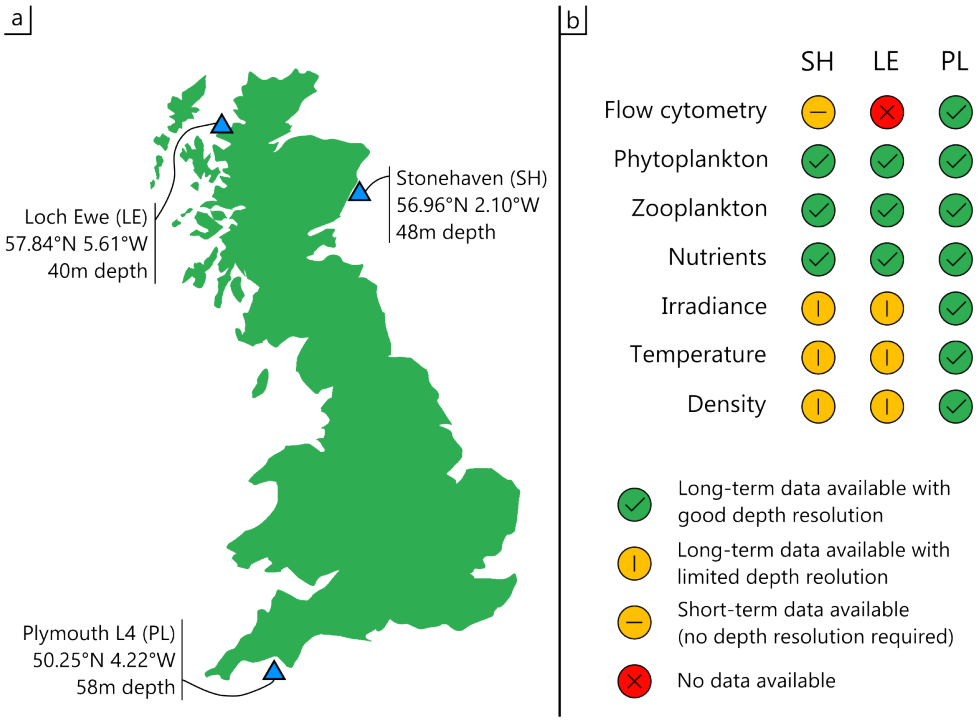


Figure 2.2: (a) Location and depth of SH, LE, and PL around the coast of the UK (excluding Orkney, Shetland, and Northern Ireland); (b) summary of the available observational data at each site across the time series. SSM and WeStCOMS output are not included.

depth. In order to avoid excessive interpolation, output from the Scottish Shelf Model climatology (SSM; De Dominics *et al.*, 2018) was used to estimate MLD at SH, while West of Scotland Coastal Ocean Modelling System (WeStCOMS; Aleynik *et al.*, 2016) output aggregated across 2020−2024 was used at LE, since the SSM does not sufficiently resolve the complex bathymetry of LE. SSM was validated using CTD observations of temperature and salinity between May 2008 and March 2009 (Wolf *et al.*, 2016). WeStCOMS outputs were evaluated using temperature observations collected on tagged flapper skates, which showed that early versions of the model overestimated temperature by 0.53°C on average, but this was reduced almost entirely in the version used here (Lavender *et al.*, 2022).

All data processing and analysis was undertaken in R v4.2.2 using the `tidyverse` package (Wickham *et al.*, 2019). Brief summaries of data availability and the location of sampling sites are provided in Figure 2.2.

2.2.1.2 Nutrients

At each site, nutrient data (nitrate, nitrite, ammonium, dissolved inorganic phosphate, and dissolved silicate) were available consistently through the time series. On occasion, nutrients were observed at the limit of detection (LOD); in these cases, the computed value was taken as half of the LOD. For brevity, nitrogen sources were pooled as dissolved nitrogen (DN) encompassing nitrate, nitrite, and ammonium. This decision was made in the absence of a clear size-based

tradeoff for growth with respect to the composition of nitrogen sources.

2.2.1.3 Calculation of MLD

MLD was calculated using a threshold approach, whereby density at each depth below 5 m were compared to the density at 5 m. The first depth with a $+0.125$ g L^{-1} difference to the density at 5 m was determined as the MLD (Jiao et al. 2021). At PL, density data was collected through CTD cast at 0.5−1 m depth intervals, allowing a precise estimate of the MLD for each sample. At SH & LE, SSM & WeStCOMS outputs were used respectively, with files read using the `ncdf4` package (Pierce, 2010), and a selection of points selected around the sampling site. At SH, a set of 4 points within a 0.01° radius around the observatory were used in aggregation. At LE, the same condition was used with an additional constraint that points between 35 and 45 m be used (due to the complex bathymetry of the seabed); in total, 3 points satisfied the criteria. Pressure, salinity, and temperature were extracted from the SSM climatology output for each of the 20 sigma layers at SH, and for the 10 sigma layers of WeStCOMS at LE. Using the package `gsw` (Kelley *et al.*, 2015), pressure, salinity, and temperature were used to calculate density:

1. Conversion from practical salinity to absolute salinity was made using `gsw_SA_from_SP()`.
2. Subsequent conversion from temperature to conservative temperature, dependent on practical salinity, was made using `gsw_CT_from_t()`.
3. An estimate of density at each sigma layer was made using `gsw_rho()`. The depth of each sigma layer was determined as the midpoint of its lower and upper boundaries.

Plots were produced to show density profiles at each site, and the estimate of MLD using the aforementioned threshold method. These were produced using `ggplot2` (Wickham *et al.*, 2019), with `ggpubr` (Kassambara, 2016) used to construct multi-plot figures.

2.2.1.4 Calculation of PAR & attenuation coefficients

The observed irradiance was transformed to PAR by using a fixed conversion factor of ~ 2.1 μEin m^{-2} s^{-1} (W m^{-2})$^{-1}$, which is equivalent to PAR constituting 45% of total solar radiation (Akitsu *et al.*, 2022). The conversion is made under the assumption that PAR constituted the $400 - 700$ nm range of wavelength.

The seawater attenuation coefficient (k_w) at PL was estimated by solving the equation for light decay:

$$I(z) = I_0 \exp\{-z \cdot (k_w + k_{Chl} P_{Chl})\} \tag{2.1}$$

where $I(z)$ is the irradiance (W m^{-2}) at depth z (m), P_{Chl} is the phytoplankton chlorophyll a concentration (mg chl a m^{-3}), k_w is the attenuation rate of seawater (m^{-1}), and $k_{Chl} = 0.025$ (mg chl a)$^{-1}$ m^2 is the attenuation rate of planktonic chlorophyll a (Gan *et al.*, 2010). For each sample, Equation 2.1 was solved for k_w, using measurements of P_{Chl} and $I(z)$. At SH & LE, where depth resolved $I(z)$ was not available, Secchi depth (z_S; m) and P_{Chl} measurements were

used so that k_w was obtained from the equation given in Castillo-Ramírez *et al.* (2020):

$$k_d = k_w + k_{Chl} P_{Chl} = \frac{1.37}{z_S}. \tag{2.2}$$

The value of 1.37 can be implemented in Equation 2.1 to show that the Secchi depth is the point at which $\sim 1/4$ of the light remains in the water column. For consistency between the three sites, the depth at which measured PAR was 1/4 of its value at the surface was used when solving Equation 2.1 at PL.

2.2.1.5 Phytoplankton

At all sites, abundances of phytoplankton taxa were estimated by extrapolating subsample cell counts to abundance concentration (cells L^{-1}). Three enumeration methods were used across the dataset to estimate phytoplankton abundance, depending on size: (i) traditional inverted microscopy for cells > 6 μm, (ii) epifluorescence microscopy for cell $3 - 8$ μm, and (iii) flow cytometry for cells $< 5\mu$m. Methods (i)$-$(iii) were used at PL; (i) & (iii) at SH; and only (i) at LE. At PL, a 50 mL sample volume is used when abundances are low (primarily in winter), however this can drop to below 20 mL during periods of high abundance. At SH & LE, a fixed sample volume of 1 mL is used throughout the year. Observations of taxa below 1000 cells L^{-1} are frequent at PL, so it is likely that rare species are systematically underestimated at SH & LE.

2.2.1.6 Zooplankton

Zooplankton were collected through vertical net hauls at each site. The organisms were then enumerated by taxon and abundance concentrations were estimated by extrapolating organism counts. Sizes were allocated not only depending on taxa, but also by recorded lifestage at SH & LE, or by season at PL. Samples were drawn with a 200 μm net; hauls from 63 μm mesh nets analysed by FlowCAM at PL were not available during the data collection period (2022), while the 63 μm net hauls at SH & LE are stored but were not analysed as of 2022.

2.2.1.7 Plankton size unit conversions

Size data were given in various forms but were transformed to a common unit of measurement: equivalent spherical diameter (ESD; μm). Typically, phytoplankton size data were supplied either in cell volumes or in ESD itself. Volume was converted to ESD using the equation of a sphere (Equation 1.1). Zooplankton size data were typically provided as dry mass. These values were converted to volume by first estimating the wet mass using a conversion factor and assuming that the density of the organism is equal to that of seawater (~ 1.025g cm^{-3}). From this point, volume was converted to ESD as described in Equation 1.1. In some instances, and in particular with respect to allometry, ln(ESD) is used to represent size.

2.2.2 Estimation of plankton biomass-size distributions

Apportioning a fixed size by taxon does not reflect true distribution of sizes within species, since plankton can vary significantly in size within species (Rillo *et al.*, 2020; Hamer *et al.*, 2022; Orizar & Lewandowska, 2022; Hillebrand *et al.*, 2022a). As such, estimates of size diversity are likely to be severely underestimated in the absence of intraspecific size variance. We addressed this issue by adapting the method of Schartau *et al.* (2010) to construct density estimates of plankton size structure. Two levels of stochasticity were applied during the process: firstly, a randomness reflecting the assumption of spatial heterogeneity of organisms in the water column, and secondly, a randomness associated with organism size by taxon.

2.2.2.1 Plankton functional groups

Plankton data were partitioned into key functional groups: diatoms, other autotrophic phytoplankton, mixotrophic dinoflagellates and ciliates, microzooplankton, copepods, other non-gelatinous zooplankton, semi-gelatinous zooplankton, meroplankton and fish larvae, and gelatinous plankton. This reflects an interest in community-level characteristics. Since it is only possible to apply one carbon conversion equation to each composite abundance-size distribution, the data were partitioned such that all organisms considered within a density estimate were governed by the same carbon conversion.

2.2.2.2 Resampling

Under the assumption of spatial heterogeneity, a large degree of variability is expected in the recorded species' estimated abundances from one subsample to the next, due to their tiny volume when compared to the water body from which they were sampled. To reflect this uncertainty, we resampled organism counts within each subsample several times, allowing an estimate of the uncertainty of the abundance estimate. To do so, the abundances of each taxon i (A_i; continuous) were converted back to expected counts (λ_i; discrete) by introducing a volume multiplying factor r (with units L^{-1}). For species $i = \{1, 2, ..., n\}$, counts were modelled as $C_i \sim \mathrm{Poi}(\lambda_i)$ distribution, where $\lambda_i = A_i/r$. The Poisson distribution is a natural choice since the mean and variance scale with direct proportion for any choice of $r(\neq 0)$. r was calculated to minimise computational expense, and so was set as the maximum value for which all A_i in a sample partition are divisible.

2.2.2.3 Applying measurement error

To account for the natural variation of size *within* species, we assumed that each organism with typical size μ_x has a corresponding size variance σ_x which scales proportionally with its typical size. We applied this measurement error to only one dimension of size (ESD) since in general the shape of organisms or the length and width measurements were unknown. Values were retrieved from Schartau *et al.* (2010) and extrapolated to include larger organisms (meso- / macrozooplankton). In doing so, the qualitative assumption was made that as typical organism size increases, so too does its variance. Equation 2.3 was fitted empirically from lines $1 - 6$ of

Table 2.1: Resolution of microscopic measurement corresponding to ESD ranges. Entries above the dashed line show values extracted from Schartau et al. (2010); below this shows extrapolated values.

Resolution of microscopic measurement ($\sigma_x; \mu$**m**)	**Size range for ESD** ($\mu_{\bar{x}}; \mu$**m**)
0.2	$0-2$
1	$2-6$
2	$6-40$
5	$40-80$
10	$80-180$
20	$180-300$
40	$300-500$
75	$500-800$
125	$800-1300$
200	$1300-2000$
300	$2000-3000$
450	$3000-4300$
650	$4300-6000$

Table 2.1, where σ_x denotes resolution of microscopic measurement and $\mu_{\bar{x}}$ is the geometric mean of the size range of ESD (note that values $\sqrt{\mu_{\bar{x}}}$ are equally spaced

$$\sqrt{\sigma_x} = -0.30475 + 0.35817\sqrt{\mu_{\bar{x}}}. \tag{2.3}$$

Lines $7-13$ in Table 2.1 determined by this equation.

For any organism of typical size μ_x and size variance σ_x, we set $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $S = \ln X$. However, for estimates to be realistic, we require that $x > 0$. Since cases where $x \leq 0$ were rare ($< 0.1\%$ of cases), we constrained x to be positive by taking $|x|$.

2.2.2.4 Construction of kernels

Abundance density estimates were produced between $0.5-300\,\mu$m for unicellular plankton, and $100-6000\,\mu$m or multicellular plankton. Implementation in R 4.2.2 required the ESD values to be discretised; we did so using 1,500 evenly log-spaced sizes for each domain. For $i = 1, ..., n$ recorded taxa in a sample partition, each with $j = 1, ..., C_i$ counts, this abundance density estimate given by:

$$\hat{A} = r \sum_{i=1}^{n} \sum_{j=1}^{C_i} \frac{1}{h\sqrt{2\pi}} \cdot \exp\left[-\frac{(s_{ij} - \ln(\mathrm{esd}))^2}{2h^2} \right] \tag{2.4}$$

where h is the estimated bandwidth, given by:

$$h = 1.06 \cdot \min\left\{ \hat{\sigma}_S^{tot}, \frac{\mathrm{IQR}}{1.34} \right\} \cdot \left(r \cdot \sum_{i=1}^{n} C_i \right)^{-1/5} \tag{2.5}$$

where $\hat{\sigma}_S^{tot}$ is the estimated standard deviation of the s in the total sample partition and IQR

Table 2.2: Parameter values of the carbon to volume relationship for phytoplankton (Menden-Deuer & Lessard, 2000) Adjusted parameter values due to discontinuity are marked by *.

Class	**Size interval (μm)**	**$\log_{10}a$**	b
Diatom	$0.5 - 15.5$	-0.541	0.811
Diatom*	$15.5 - 20.4$	0.743	0.421
Diatom	$20.4 - 300$	-0.933	0.881
Dinoflagellate	$0.5 - 300$	-0.353	0.864
Ciliate	$0.5 - 300$	-0.168	0.841
Protist	$0.5 - 15.5$	-0.583	0.860
Protist*	$15.5 - 20.4$	-2.515	1.447
Protist	$20.4 - 300$	-0.685	0.939

is the estimated inter-quartile range of s from the sample partition (Schartau *et al.*, 2010). The formulation of h in Equation 2.5 is robust to multimodality and seeks to optimise the bias-variance tradeoff by accounting for differences between IQR and standard deviation arising from skewness / multimodality (Schartau *et al.*, 2010).

2.2.2.5 Conversion to biomass density

Abundance density estimates were converted to units of biomass (μg C L^{-1}). There are differences in the typical carbon concentrations per unit of volume between plankton groups (Menden-Deuer & Lessard, 2000; Harris, 1999), and so the abundance-to-biomass transformation varies accordingly. In the broadest sense, the biomass density estimate ($\hat{B}$) is given by:

$$\hat{B} = \hat{A} \cdot m(\text{esd}), \tag{2.6}$$

with size-dependent carbon cellular content given by m(esd). The form and parametrisation of m(esd) varies between groups and is described below.

Phytoplankton

The carbon to volume relationship for phytoplankton is governed by an allometric relationship, described in detail in Menden-Deuer & Lessard (2000), which takes the form:

$$m = a \cdot V^b \iff \log_{10}(m) = \log_{10}(a) + b \cdot \log_{10}(V) \tag{2.7}$$

for a cell of volume V (μm^3) and corresponding carbon mass m (pg C). We adapted the parameters a and b over a small size interval, since the original parameters produce a discontinuity point at $3,000\,\mu m^3$ (Table 2.2).

Zooplankton

Carbon conversion values provided by Harris (1999) were used to transform zooplankton size from wet mass to carbon mass. Zooplankton sizes were converted from to wet mass under the assumption that organisms have the same density as seawater ($\rho = 1.025$g cm^{-3}), then wet mass

Table 2.3: Parameter values of the carbon to volume relationship for zooplankton (Harris, 1999).

Composition	p_D	p_C
Gelatinous	0.04	0.1
Semi-gelatinous	0.1	0.3
Non-gelatinous	0.2	0.45

was converted to carbon mass using the following equation:

$$m = \rho \cdot V \cdot p_D \cdot p_C \tag{2.8}$$

where p_D is the proportion of wet mass constituted by dry mass, and p_C is the proportion of dry mass constituted by carbon. The respective parameter values are given in Table 2.3 and are size-independent.

2.2.2.6 Bootstrapping

The biomass density estimation method incorporates two degrees of randomness by resampling counts and imposing intraspecific size variance. Each computed density estimate was produced from several bootstrap runs in order to obtain an accurate biomass distribution estimate. We estimated the average density for each value of ESD, and a corresponding standard error for the calculation of each point on the composite distributional curve across all the bootstrapped runs.

Through experimentation, we found that in order to obtain a smooth composite curve, the required number of bootstraps correlated with the number of species recorded in the sample. The number of bootstraps required with respect to the species richness within a sample partition is shown in Table 2.4. Each sample partition was then amalgamated as the mean density within each of the discretised size bins.

2.2.3 Analysis of size structure

2.2.3.1 Exploratory analysis

An initial exploratory analysis was undertaken by plotting and inspecting composite density estimates of biomass distributions. Each plot aggregates data across the time series, for each season, at each site. This was an important step to establish the accumulation of biomass at different regions of the size domain, and identify the contributions of each group. The composite estimates were constructed by computing the truncated mean density within each size bin for each metrological season across the entire sampling period for each site. A similar method is used when estimating diatom frustule size in Finkel *et al.* (2005). The reason for using a truncated mean was to mitigate against the deficiencies of the median and mean densities under a highly asymmetric distribution of biomass density values within each size bin. Because of the nature of biomass data, which are often right-skewed, the mean is susceptible to overestimation of the central tendency, since rare, individual observations of high biomass can have disproportionally large leverage on the estimate. This is highlighted, for example, at LE in winter, where a total

Table 2.4: Number of bootstraps, depending on the number of observed species in a subsample.

Range of species richness	**Number of bootstraps**
1	5000
2	1500
$3-5$	500
$6-10$	200
>10	100

count of three *Conscinodicus* spp. ($123.7\,\mu$m) cells over the 20-year sampling period combine to contribute proportionally high average biomass (Figure 2.3). These occurrences are exceptionally rare and would not appear indicative of the 'typical' distribution for cell sizes in winter.

Conversely, the biomass density values within individual size bins are frequently zero or infinitesimal, such that the median is likely to under-represent the biomass accumulations from some regions of the size domain (particularly large cells, which are numerically less abundant and therefore more likely to be under-sampled). As a result, the median is also susceptible to bias arising from differences in sample volume, since, for example, a species at abundance of 100 cells L^{-1} may be regularly sampled with a 50 mL subsample volume (providing a non-zero biomass density), but rarely noticed by a 1 mL subsample (falsely reporting a zero-biomass median density).

This creates a qualitative deficiency in the understanding of the community, since important cells, which may only be measured, for example, 30% of the time (despite a true population presence $>50\%$ of the time) are not represented by the median. Figures 2.11, 2.12, and 2.13 therefore provide insight into the important, typical biomass contributions, and are therefore intended to be used for qualitative purposes. The compromise is that give a biased quantitative estimate of overall community biomass. The composite curves cannot be used to estimate size variance or mean size since they average across communities over a long time period.

Cut-off points for the truncated mean were tested and compared, as shown in Figure 2.3. The presented Figures 2.12, 2.13, 2.11 are computed as mean densities from the central $5-95\%$ of values. A $10-90\%$ interval is also plotted in Figure 2.3, but a $5-95\%$ was chosen in favour with the intention of including as much data as possible and minimising bias. The extent of underestimation by the median and overestimation by the mean is also demonstrated.

2.2.3.2 Summary statistics

To summarise plankton community biomass and size structure, we calculated summary statistics directly from estimates of $\hat{B}$. Primarily, we took interest in the mean and variance of log-ESD ($\mathrm{E}[s]$ and $\mathrm{Var}[s]$ respectively), as well as skewness ($\mathcal{S}$) and kurtosis ($\mathcal{K}$). These were calculated as distributional moments of $\hat{B}^* = \hat{B}/\int_s \hat{B}$ (i.e., the normalised estimate of $\hat{B}$ with combined mass of 1). Log-ESD variance is a proxy for size diversity, and so it can be compared to other diversity indices (e.g. richness or Shannon). Initially, these size structure indicators were calculated on a per-sample basis, but estimates of mean log-ESD were also calculated by averaging across a

week-of-year basis in an effort to increase the effective sample volume at SH & LE; both results are presented. It was not appropriate to average higher order distributional moments such as variance across years, since varying degrees of inter-sample variability in the size structure across seasons would affect these estimates.

The skewness-kurtosis pairs can be used to assess the shape of the biomass distribution (e.g., García-Oliva *et al.*, 2022). There are two regions of interest for skewness-kurtosis pairs we focused on. Firstly, whether the distribution is approximately normal, defined here in approximate terms using the equation:

$$4\mathcal{S}^2 + (\mathcal{K} - 3)^2 \leq 1, \qquad -0.5 \leq \mathcal{S} \leq 0.5 \quad \text{and} \quad 2 < \mathcal{K} < 4$$

which is an ellipse on the domain $-0.5 \leq \mathcal{S} \leq 0.5$ and $2 \leq \mathcal{K} \leq 4$. This is not a formal statistical test (e.g. Jarque-Bera test), but rather is a informal test based around the classical 'rule of thumb' definitions for approximate normality based on skewness and kurtosis. Secondly, the bimodality region is defined such that

$$\mathcal{K} \leq \frac{9}{5}(\mathcal{S}^2 + 1), \qquad -1 \leq \mathcal{S} \leq 1$$

This corresponds to Sarle's bimodality coefficient which, again, is an informal test. It is important to note that, under Pearson's inequality, $\mathcal{K} \geq \mathcal{S}^2 + 1$ is always true, and so the bimodality region is bounded below for values of $\mathcal{K}$. The constraint on skewness is used so that two modes are sufficiently similar in size.

2.2.3.3 Statistical modelling

For each of the ecological and environmental indicators, the typical seasonal cycle was determined by fitting generalized additive models (GAMs) with cyclic cubic (CC) regression splines using the `mgcv` package (Wood, 2011). An identity link was used with transformation of the response where necessary. Fitting of the response with a log-link function was investigated, but did not improve the fit of the model. Back-transformation of positively skewed variables (such as biomass) was used to make inference on the fitted median in original units of the response.

Some indicators did not require transformation as response variables, and so were modelled in their original units: temperature, MLD, PAR, and mean log-ESD. Meanwhile, the following were log-transformed and modelled: DN, DSi, DP, DN:DP, DSi:DP, k_d, k_w, $k_{Chl} \cdot Chl$, and all biomasses. log-ESD variance was square-root transformed, in effect generating a model for log-ESD standard deviation. For all seasonal models except that fitted to PAR, day of year was the explanatory variable. CC splines were chosen due to the constraint that the fitted response must be continuous when crossing from the last day of one year to the first day of the next year. Surface PAR was resolved hourly at each site and, as such, the GAM was fitted on spherical co-ordinates so that the model could account for variability on the hourly basis, daily basis, and the tensor product (interaction) between the two (e.g., days in July are longer than in January).

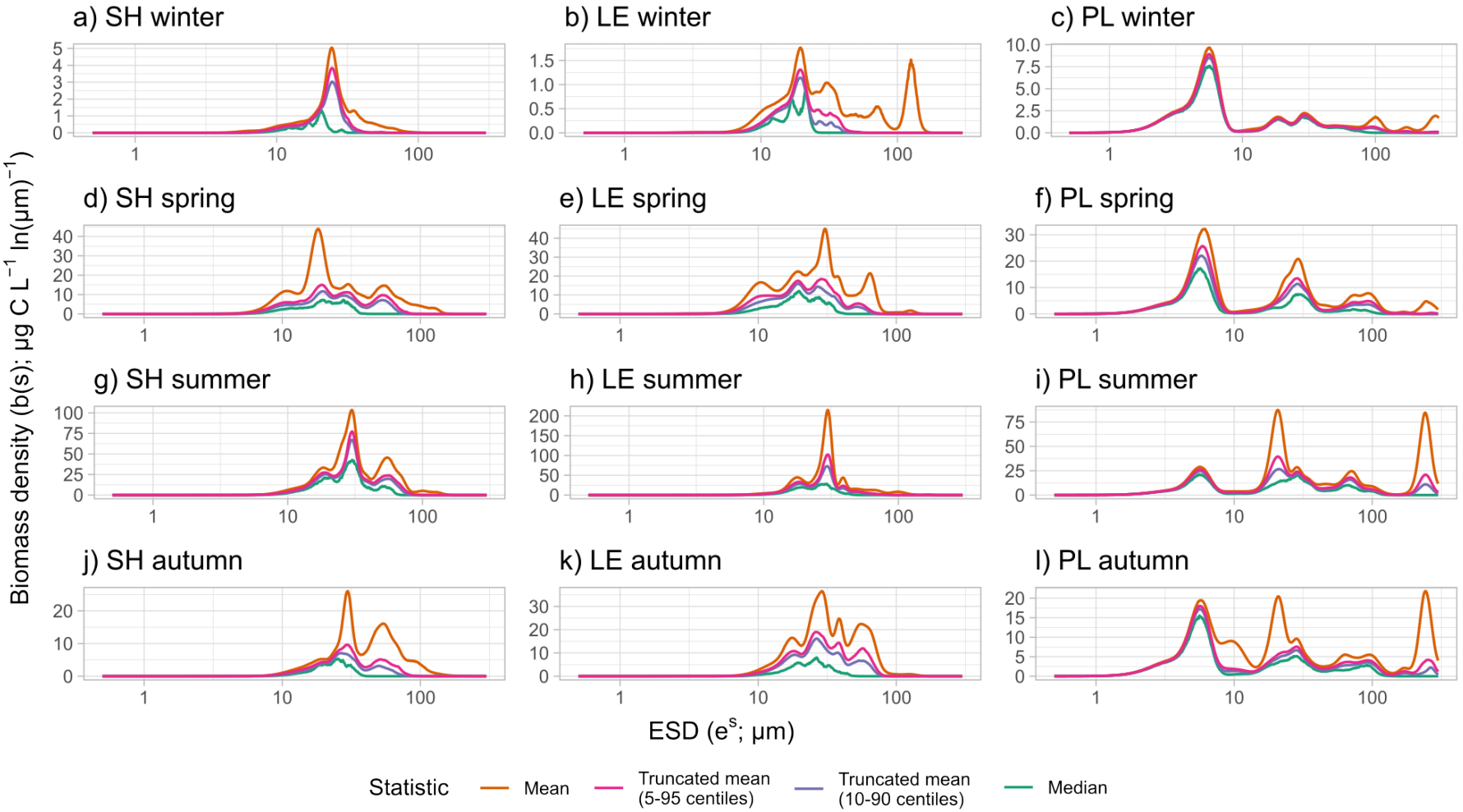


Figure 2.3: Typical biomass distributions of unicellular communities during winter, spring, summer, and autumn at SH, LE, and PL. Line colours represent different estimators of central tendency. Data excludes flow cytometry.

Each of the above models was summarised by the p value of the smooth term and the adjusted R^2. Furthermore, the residuals were assessed by interpretation of residual error plots (which were plotted against fitted values), and Q-Q plots. A full report of p values, R^2 values, the number of observations n, the time period of the observations, and notes on the model diagnostics is described in Table B.1 for seasonal models. In some cases, overdispersion and/or heteroscedasticity was present in the residual error distribution of the fitted GAMs. In these instances, the p-value was not considered as robust (White, 1980; Freedman, 2006) and was therefore omitted from analysis.

To analyse the relationships between indicators, linear models (LMs) were fitted, with polynomial explanatory terms included in the model where necessary. All polynomial LMs were fitted under the principle of marginality so that a higher order polynomial term could only be fitted if its corresponding lower order terms were included. However, it should be noted that polynomial explanatory terms were orthogonalised by default by the `poly()` function nested within `lm()`. GAMs were avoided in this process to allow the interpretation of model coefficients on an order-by-order basis.

2.3 Results

2.3.1 Seasonal cycle of environmental conditions

Environmental conditions are predicted to have an important role in controlling bottom-up effects on the size structure of the phytoplankton community through the allometric tradeoffs for nutrient uptake and photosynthesis (Section 1.1.4). Temperature may also have an indirect effect on size structure, since zooplankton are more sensitive to temperature than phytoplankton (Rose & Caron, 2007; Chen *et al.*, 2012), which may in turn affect maximal grazing rates of zooplankton (which decrease with size; Hansen *et al.*, 1997). Finally, variations in the MLD impact a number of processes controlling biomass accumulation and size structure. Originally, the DRH stipulated that the spring bloom occurs due to the dilution of zooplankton in the mixed

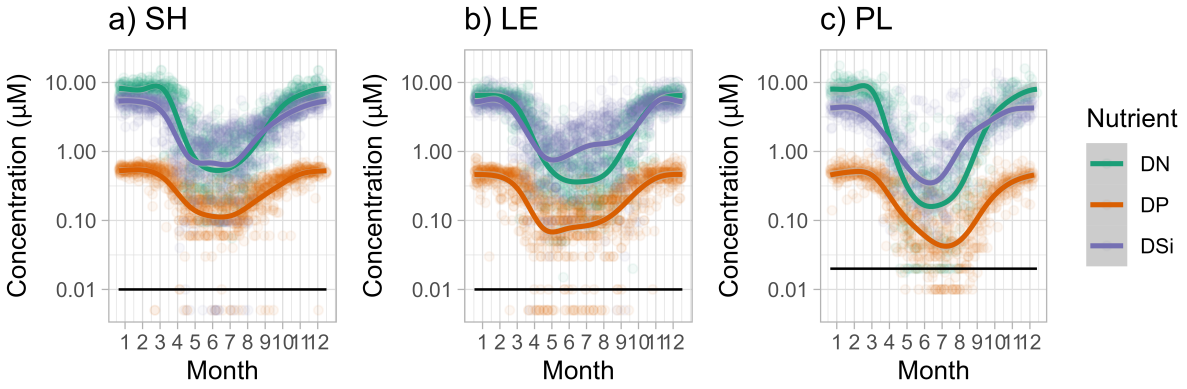


Figure 2.4: Surface nutrient concentrations of dissolved inorganic phosphate (DP), dissolved silicate (DSi), and dissolved nitrate + nitrite + ammonia (DN). LOD is marked by the black line; at SH & LE this is 0.01 μM, and at PL this is 0.02 μM.

layer, thus reducing gross loss rates for phytoplankton (Behrenfeld & Boss, 2014). However, this may not necessarily be true in coastal sites with the disturbance to equilibrium being caused by the temporal gradient of irradiance (González-Gil *et al.*, 2022). Secondly, depth-averaged photosynthesis across the mixed-layer is also likely to reduce with deepening MLD, due to the increasing combined attenuation of seawater with depth. Finally, deepening of the MLD will facilitate nutrient replenishment. As such, we investigated the seasonal cycle of environmental conditions, with findings presented below.

2.3.1.1 Nutrients

Each UK coastal site displays strong seasonality in surface nutrient concentrations. The seasonal cycle consists of replete concentrations in the winter, which decline through spring, to depletion in summer, and recovery through autumn (Figure 2.4). The fastest rate of depletion appears to occur between March and April, for DN, DSi, and DP at all three sites, although the decreasing nutrient concentrations appear to run well into July at PL, resulting in lower average concentrations than at SH & LE (despite a higher LOD). Between April and August at all three sites, concentrations of DN, DP, and DSi typically fall below 1 μM. During September and October, the gradient of nutrient replenishment appears to be steepest.

The molar ratios of nutrients was also fitted by season. In winter, when nutrients are replete, DN:DP measures approximately 16:1 in line with typical Redfield ratios (Figure 2.5). However, through spring, DN diminishes faster than DIP to the extent that, during summer when DN concentrations are below 1μM at all sites, DN:DP ratios are typically below 5:1. DSi:DP ratios exhibit their strongest relative depletion in May (a process starting in February / March). At SH, DSi remains close to $\sim 6:1$ until August, whereas at LE a recovery appears to commence in May, and at PL in June. From August onwards, DSi:DP shows a strong recovery during the late summer to early autumn period at PL, to $\sim 15:1$. Both the DN:DP and DSi:DP ratios show an earlier commencement of full recovery at PL than at SH & LE, where the ratios show a relatively greater degree of stationarity though the summer.

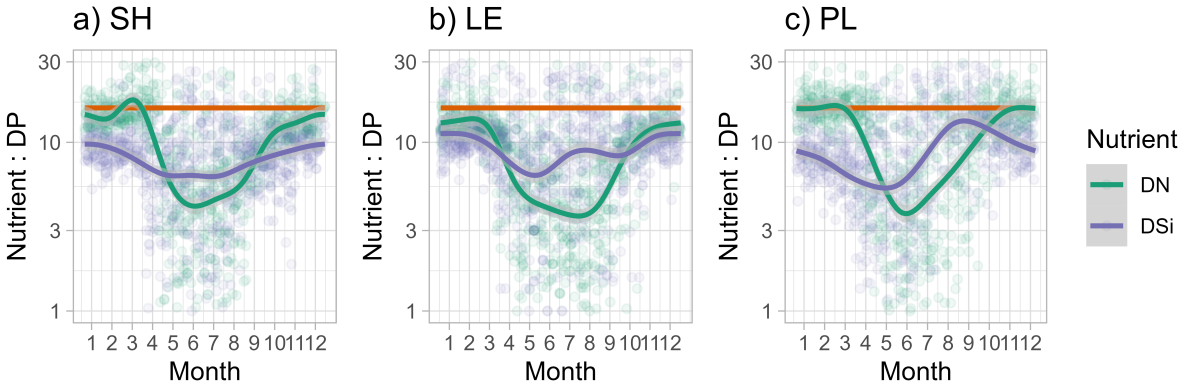


Figure 2.5: Surface nutrient molar ratios, computed against dissolved inorganic phosphate (DIP). These values can be compared to the orange line at 16:1 (Redfield ratio).

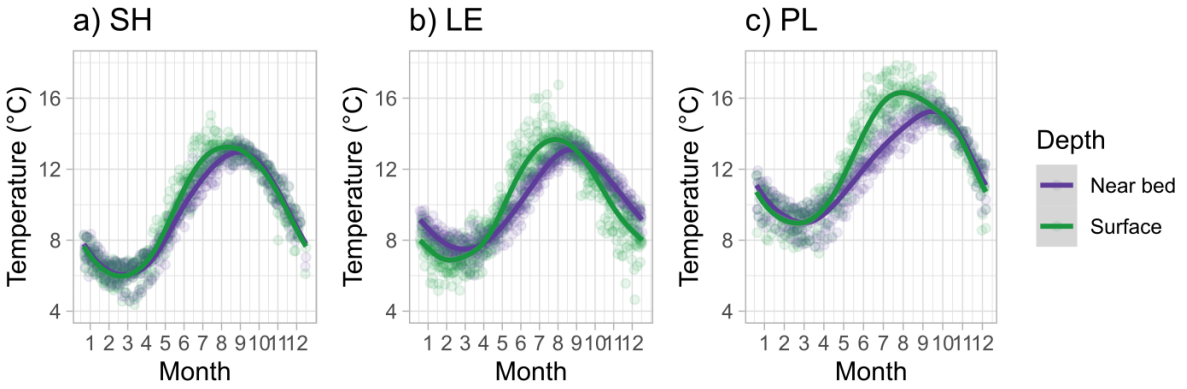


Figure 2.6: Temperature observed at surface and near-seabed depths. Surface samples at PL are taken at 2 m; at SH & LE they are taken at 1 m. Near-seabed samples are taken at 45 m, 35 m, and 50 m at SH, LE, and PL respectively.

The seasonal GAMs broadly represent the data well, with R^2 ranging between $0.451 - 0.747$ across ln(DN), ln(DSi), and ln(DP) across each of the coastal sites. The models fitted to molar ratios explained a lower proportion of variance, with the DN:DP models having a higher explanatory power ($R^2 = 0.371 - 0.427$) at all sites than the corresponding DSi:DP models ($R^2 = 0.0548 - 0.347$). In both molar ratio instances, the R^2 score is higher at PL and lower at SH & LE. However, a decrease in residual error variance with increasing fitted value is observed for all nutrient concentration and molar ratio models, and so too is moderate overdispersion of residual errors. Because of this, inference is not made in relation to p-values. The relationship between variability of residual errors and fitted values indicates that, on the log-scale, nutrient concentrations are more erratic during the periods of depletion than during period of repletion. A log-transformation of the response may therefore not be best suited to the response; however, GAMs fitted on the original units of the response underestimated the winter concentration. As such, the a transformation was preferred so that the model fit was unbiased.

2.3.1.2 Temperature

Seasonal variations in sea surface temperature vary by around ±3.5°C around annual mean at the surface at all sites; at PL this occurs around the mean of 12.6°C, while at LE (10.1°C) and SH (9.6°C) the water is considerably colder at the surface. At all sites, the lowest SST is typically recorded in March, and the highest in August. Differences in temperature between the surface and deep layer also tend to appear in summer at all sites, and also in winter at LE (Figure 2.6). In summer, the extent of the temperature difference between the surface and the bottom layer during summer is more pronounced at PL than at SH & LE. The strength of this temperature gradient is also reflected in the density gradient in July at PL (Figure 2.8). Similarly, the thermocline that exists in winter at LE is supported by the property that density is typically higher in summer than in winter (in contrast to SH & PL; Figure 2.8), and is likely driven by the proximity of the Loch to freshwater sources. These sources are anticipated to be more influential in winter as a result of enhanced precipitation. All seasonal GAM smooths

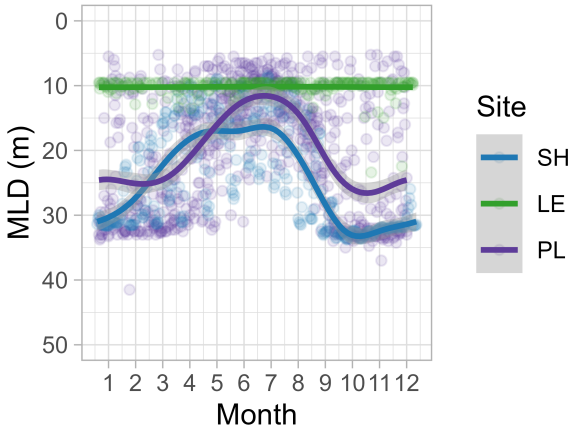


Figure 2.7: Mixed layer depth estimated at SH, LE, and PL. Data from CTD casts at PL, WeStCOMS model output at LE, and SSM climatology at SH.

fitted to SST and near-bed temperature are highly significant ($p < 2 \times 10^{-16}$), and much of the variability is explained by these models ($R^2 = 0.884 - 0.957$).

2.3.1.3 Mixed layer depth

Significant differences in the seasonal cycle of estimated MLD are observed between UK coastal sites (Figure 2.7). The MLD appears to seasonally vary at SH & PL ($p < 2 \times 10^{-16}$), while at LE the estimated MLD is shallow through the course of the year arising from the strong density gradient (Figure 2.8), with no evidence of a seasonal cycle ($p = 0.131$). It is possible that other methods of analysis (such as a temperature-based approach) may yield more insight to the mixing at LE; on the other hand, it is unsurprising that stratification does not appear to occur given the short tidal flushing rate of the sea loch (Bresnan *et al.*, 2015) and proximity to freshwater sources.

At SH & PL, the MLD shoals during the summer and is at its deepest during winter, although differences in timing and magnitude are observed (Figure 2.7). At SH, the ML appears to shoal from late January until late April, covering a range of $31 - 17$ m depth. The ML then deepens from late July until early October, and remains approximately constant thereafter. At PL, the deepest ML occurs between October and March ($\sim$ 25 m), and shoaling of the ML does not appear to commence until late March, and approaches its shallowest by early July (12 m). Yet, by the end of July, it appears to deepen once again until October. The fitted models therefore suggest that the ML has a shorter period of stationarity at PL in the summer. Despite this, this density gradient appears stronger in July at PL than at SH (Figure 2.8) and, as such, the onset of stratification itself appears to be stronger.

Unsurprisingly, the data at PL are subject to a greater degree of daily stochasticity ($R^2 = 0.343$) compared to SH ($R^2 = 0.632$), since they are derived from observations rather than averaged ocean model outputs. Analysis of model residuals at PL illustrates that winter residuals are slightly negatively skewed. As such, the estimate of MLD in winter from the fitted GAM is likely

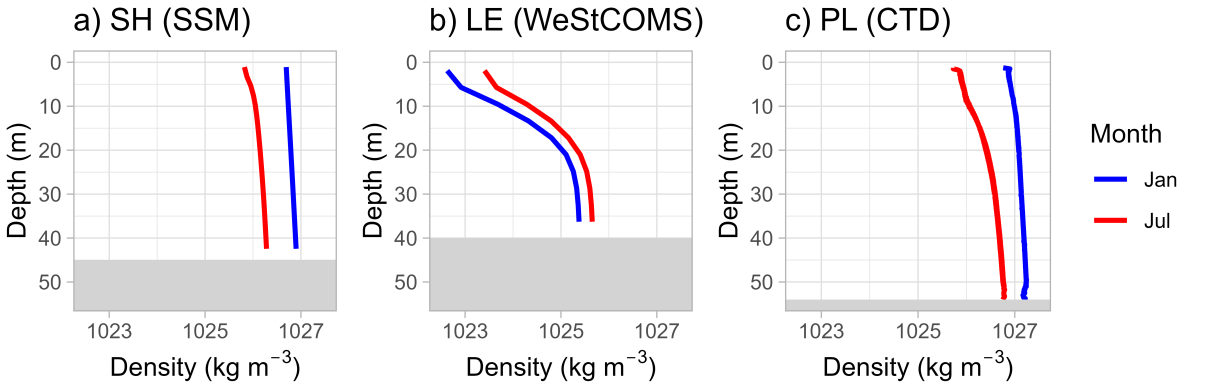


Figure 2.8: Mean density by depth in January and July estimated at SH, LE, and PL. Data from CTD casts at PL, WeStCOMS model output at LE, and the SSM climatology at SH. Shaded region indicates seabed.

slightly underestimated since most observations are clustered around 30 m (Figure 2.7).

At SH & PL, the density is greater in January than in July throughout the water column (Figure 2.8). This pattern is inverted at LE, where density is greater in July than in January throughout the water column. This is in alignment with the temperature plots of Figure 2.6, where temperature is colder at the surface than at near-bed depth; at SH & PL, no significant thermocline is observed in winter. A reasonable inference is that, at LE, freshwater input has a greater effect on the water column temperature and density dynamics than heating through solar radiation. This is further supported by the strong density gradient observed in Figure 2.8, despite the fast flushing rate of the loch.

It is interesting that the MLD appears to be established year-round at SH, in contrast to previous expectations that the full depth of the water column is mixed in winter (González-Gil *et al.*, 2022). This would suggest that the support for the DRH could still be a result of dilution through deepening of the ML. Since our estimates are conditional on the accuracy of the SSM climatology, and the subsequent estimation method of MLD, further work is necessary to study the role of dilution in the original, MLD-dependent DRH (Behrenfeld & Boss, 2014) at SH and similar coastal sites since the exact mechanisms remain unclear.

2.3.1.4 Light & attenuation

Surface irradiance

The UK coastal sites display a strong seasonal cycle of PAR, which is characterised by long, bright days in the summer and short, dark days in the winter. Throughout the year, daily maximum PAR is consistently highest at PL, and lowest at LE, despite the overall day length being slightly longer at Scottish coastal sites in summer.

At each site, the daily maximum PAR is typically at its lowest during January. During this period, Figure 2.9 shows that PAR at PL ($\sim$ 280 μmol m^{-2} s^{-1}) is almost double that at LE ($\sim$ 150 μmol m^{-2} s^{-1}), and also is significantly greater than at SH ($\sim$ 180 μmol m^{-2} s^{-1}). The

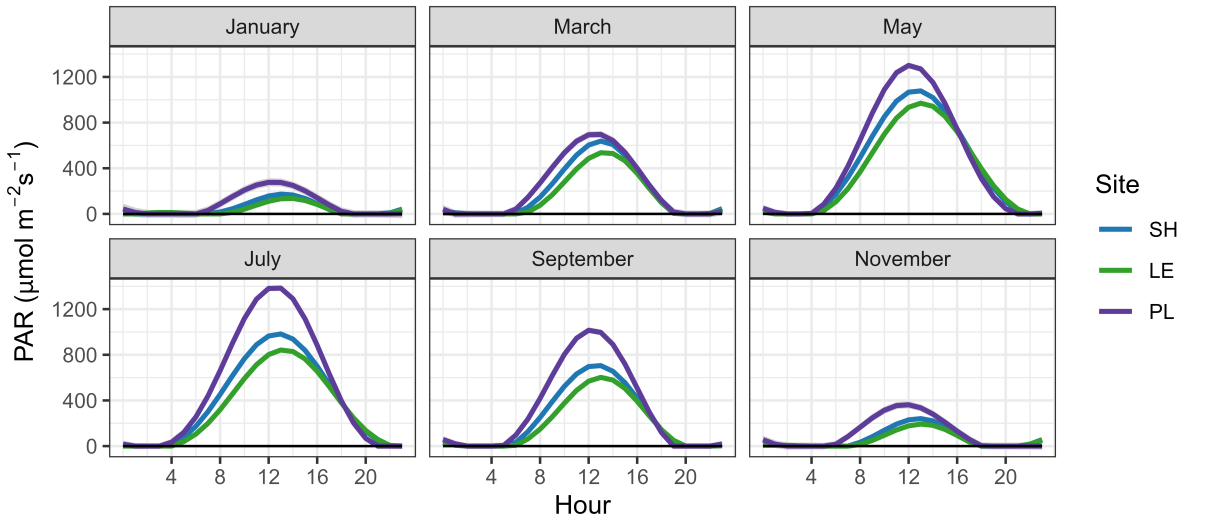


Figure 2.9: Modelled PAR from observations at PL and sites near SH ($\sim$ 28 km) & LE ($\sim$ 31 km). GAMs fitted on spherical coordinates with an interactive term between time of day and day of year. Residual errors shown in Figure B.1.

highest daily maximum PAR occurs in May/June at the Scottish coastal sites ($\sim$ 1000 μmol m^{-2} s^{-1}), and during July at PL ($\sim$ 1400 μmol m^{-2} s^{-1}).

The difference in timing of the annual maximum PAR is almost certainly a result of differences in local weather. Met Office (2020) weather averages data indicate that the number of sunshine hours drop only slightly from May to June at PL (227 h to 209 h), whereas at LE reduce drastically (158 h to 108 h) over this period. Thus, while the daily PAR cycle is similar between PL and the Scottish sites in March-May, PL receives much more light from mid summer to early autumn.

Because the PAR data have minimum and maximum values which are constrained by the time of day, the residual errors are highly structured (Figure B.1). There is a strong positive relationship between the variance of the residual error and the fitted value, and a strong overdispersion of errors. Despite this, there is symmetry in the distribution of errors and, as such, the fit of the model does not appear to be biased. Furthermore, a log-transformation of the response or log-link function is inappropriate since the night observations are zero-dominated. A hierarchical model with an initial day / night step would therefore be necessary to construct a model which adequately resolves the error variance.

Turbidity & attenuation

Total attenuation (k_d) in the water column follows a strong seasonal cycle at SH & LE ($p < 2 \times 10^{-16}$; $R^2 = 0.144 - 0.18$). k_d is highest in January–February at SH & LE ($\sim$ 0.25 m^{-1}) and decreases through spring (Figure 2.10). Although the annual minimum at SH & LE ($\sim$ 0.17 m^{-1}) typically occurs in August, attenuation is below 0.2 m^{-1} between May and October at both sites. At PL, little seasonal variation is visible around the mean of 0.16 m^{-1}, despite statistical significance ($p = 3.01 \times 10^{-3}$). Despite the minimal amount of variation explained by the seasonal

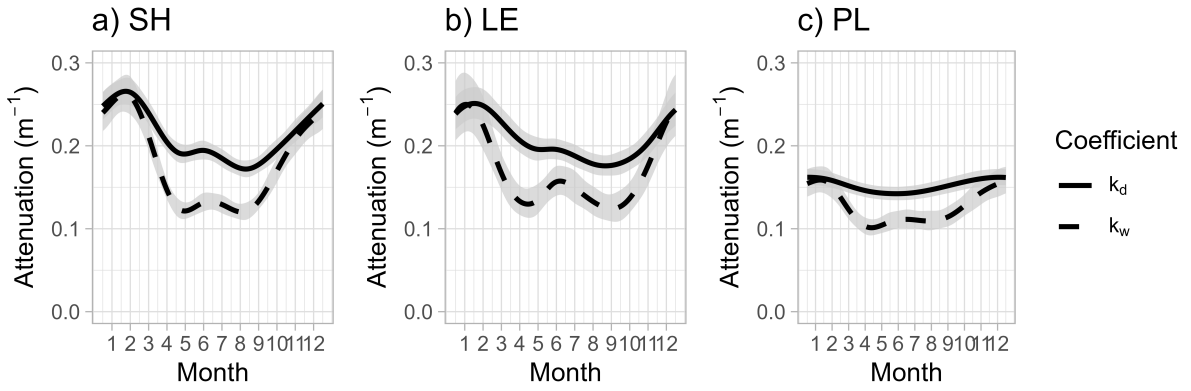


Figure 2.10: Total seawater attenuation (k_d) and background seawater attenuation (inclusive of suspended matter; k_w). Concentration-dependent chlorophyll attenuation can be inferred from $k_d - k_w$, and was incorporated in the calculation of k_w. k_d calculated from Secchi depth at SH & LE, and from depth-resolved PAR observations from CTD casts at PL.

GAM fitted to k_d at PL ($R^2 = 0.024$), models fitted to k_w and $k_{Chl} \cdot Chl$ explain a larger degree of variability, albeit still relatively low ($R^2 = 0.124$ and $R^2 = 0.28$, respectively).

The contribution of suspended sediment (k_w) dominates attenuation year-round. In winter, the attenuation is driven almost entirely by sediment since chlorophyll contributions are minimal. In summer, chlorophyll-dependent attenuation increases, but the attenuation by sediment decreases: at PL, these changes effectively balance eachother out, and so there is little seasonality in the combined seawater attenuation (k_d); at SH & LE, the reductions in k_w are substantial enough to reduce k_d, despite the fact that chlorophyll-based attenuation increases. Importantly, k_w is often considered to be constant in ocean models (e.g., Ward *et al.*, 2012; Anderson *et al.*, 2015; Acevedo-Trejos *et al.*, 2015), and so the fact that significant seasonal variability is observed suggests that a seasonally varying k_w may be necessary in modelling efforts at UK coastal sites.

2.3.2 Assessment of biomass distributions by season

2.3.2.1 Unicellular plankton: flow cytometry + microscopy (SH & PL only)

There is strong seasonality in the size structure of unicellular plankton at all three UK coastal sites. The most complete overview of the phytoplankton biomass distribution is given at SH & PL, where flow cytometry data is used in complement of microscopy data. At these sites, a year-round characteristic of the biomass distribution is that contributions are made by nanoeukaryotes ($\sim 4\,\mu$m) and, to a lesser (but still important) extent, picoeukaryotes and *Synechococcus* (1.5 μm; Figure 2.11). At SH, only observations of autotrophic nanoeukaryotes are made, whereas at PL, a modest proportion of heterotrophic nanoplankton are observed in each season. The biggest changes in the community, proportionally speaking, come from the changes in the biomass composition of larger cells ($> 10\mu$m) which have much greater biomass in summer than in winter.

The winter community is dominated by small cells at both sites, but the dominance of cells $< 10\,\mu$m is not absolute at SH. A biomass peak is observed around 24 μm which consists almost

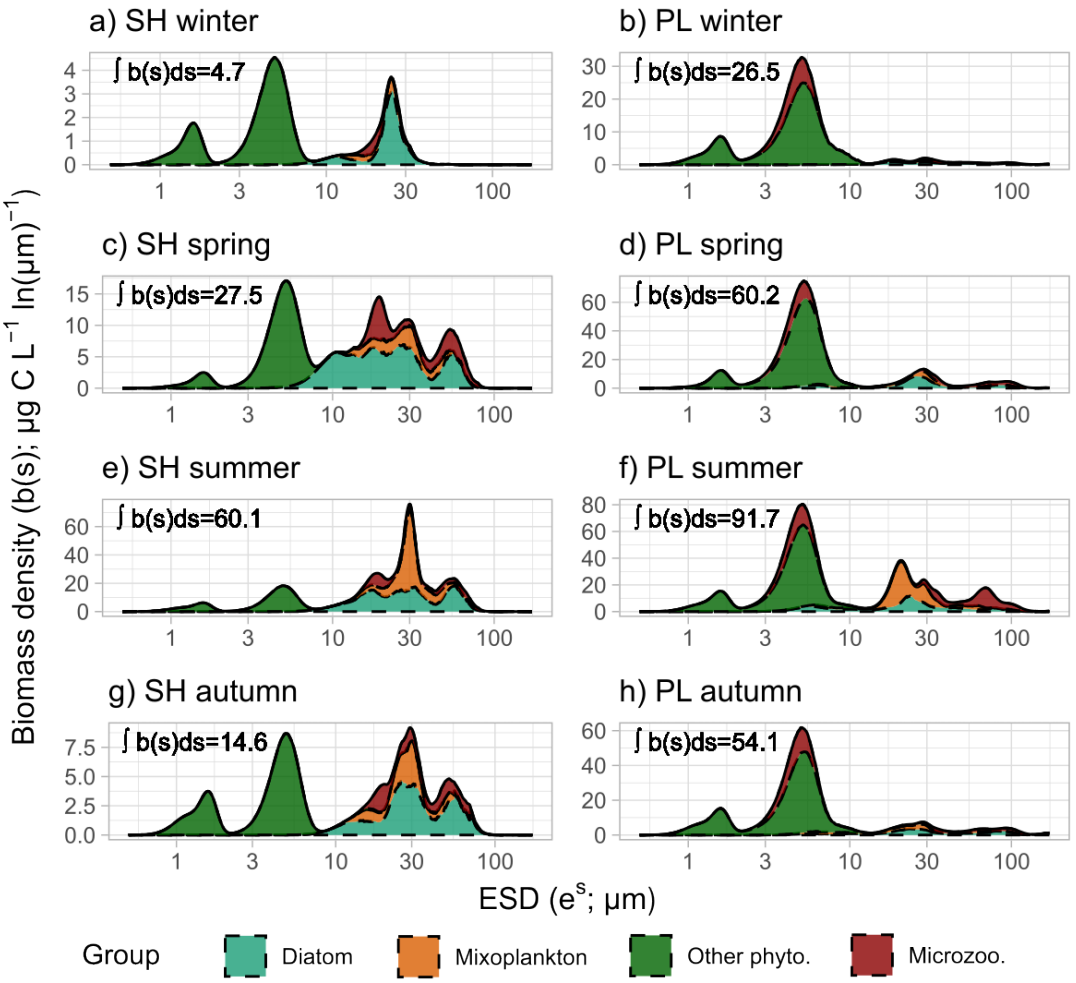


Figure 2.11: Typical biomass distributions of unicellular plankton recorded by microscopy and flow cytometry at SH and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$).

entirely of diatoms (Figure 2.11a). While a similar absolute biomass contribution is made are by cells sized $20-30\,\mu$m at PL, this is proportionally a much smaller component of the community, because it is heavily outweighed by the smaller cells (Figure 2.11b).

In Spring, biomass increases drastically at both sites. At PL, the proportional contribution of cells $> 10\,\mu$m appears to increase slightly, with a second biomass contribution centred around $\sim 27\,\mu$m (Figure 2.11d). The proportional contribution of picoplankton is also observed to decrease (despite increasing slightly in units of biomass), so that small nanoplankton remain dominant in the community. At SH, the biomass contribution of nanoplankton increases, but reduces as a proportion of the community, as contributions between $10-70\,\mu$m become more important (Figure 2.11c). A substantial portion of this biomass consists of diatoms, while mixotrophic dinoflagellates and microzooplankton also accumulate.

When averaged across seasons, biomass increases through to summer, with cells $> 10\,\mu$m ac-

cumulating further biomass, while the pico- and small nanoplankton remain relatively invariant. As such, the proportional biomass contributions become increasingly weighted towards the large cells. At SH, the large cells dominate the summer community, largely thanks to a strong biomass peak of mixotrophic dinoflagellates around $30\,\mu$m (Figure 2.11e). Conversely, at PL, the nanoplankton remain dominant, despite increasing contribution from dinoflagellates (around $20\,\mu$m) and large microzooplankton ($\sim 60\,\mu$m; Figure 2.11f).

Autumn, in many ways, looks similar to spring, despite a lower seasonally-averaged biomass. At both SH & PL, the contribution of cells $> 10\,\mu$m reduces in both absolute and proportional terms. At SH, the biomass reduction of small nanoplankton is also relatively severe (by around half; Figure 2.11g), whereas the reduction is less prominent at PL (Figure 2.11h). Although the year-round absolute biomass of small cells at PL is much higher than at SH, this is most notable in autumn and winter. Even when excluding the contributions made by cells measured by epifluorescence microscopy (which is used at PL but not SH), the peak winter biomass density at $4\,\mu$m is $\sim 25\,\mu$g C L^{-1} ln(μm)$^{-1}$, which is much higher than at SH ($4.5\,\mu$g C L^{-1} ln$(\mu m)^{-1}$). As such, it appears that there is a significant qualitative difference in the size structure between the sites on the broad scale of $0.5 - 200\,\mu$m.

2.3.2.2 Unicellular plankton: microscopy only

Despite the biomass dominance of cells $< 10\,\mu$m (particularly at PL), it is still necessary to understand the seasonal progression of the community of cells larger than $10\,\mu$m. This results from the fact that the ecosystem functioning itself evolves along the size domain. Even if, say, biomass of large diatoms in consistently dwarfed by nanoeukaryotes (as is the case at PL), the ecosystem contribution of these diatoms remain important. For example, copepods and other mesozooplankton species require food in this size domain for optimal grazing (Hansen *et al.*, 1994). Truncated mean biomass densities are shown in Figure 2.12 for unicells measured by microscopy, and are thus displayed for all three coastal sites.

Perhaps the most notable feature of the biomass distribution in Figure 2.12 is that the autotrophic nanoflagellates ($\sim 5\,\mu$m) observed at PL contribute significant biomass, and their omission at SH & LE may impact inference. As such, distributions on the subset of the size domain$> 10\,\mu$m are plotted separately at PL for comparison. On this interval, the seasonal cycle of unicellular plankton community size structure is relatively similar across sites − in particular SH & LE. Biomass is low in winter, with contributions at SH & LE centred around $20 - 25\,\mu$m. At SH, these are dominated by diatoms (Figure 2.12a), whereas at LE, contributions are made relatively equally by diatoms, dinoflagellates, and microzooplankton (when averaging across the sampling period, and therefore not necessarily in coexistence; Figure 2.12b). At PL, similar contributions are made on the $10 - 40\,\mu$m range, with further contributions of diatoms and microzooplankton made between $40 - 100\,\mu$m (Figure 2.12c). As such, the main difference is in the presence of large cells in the community.

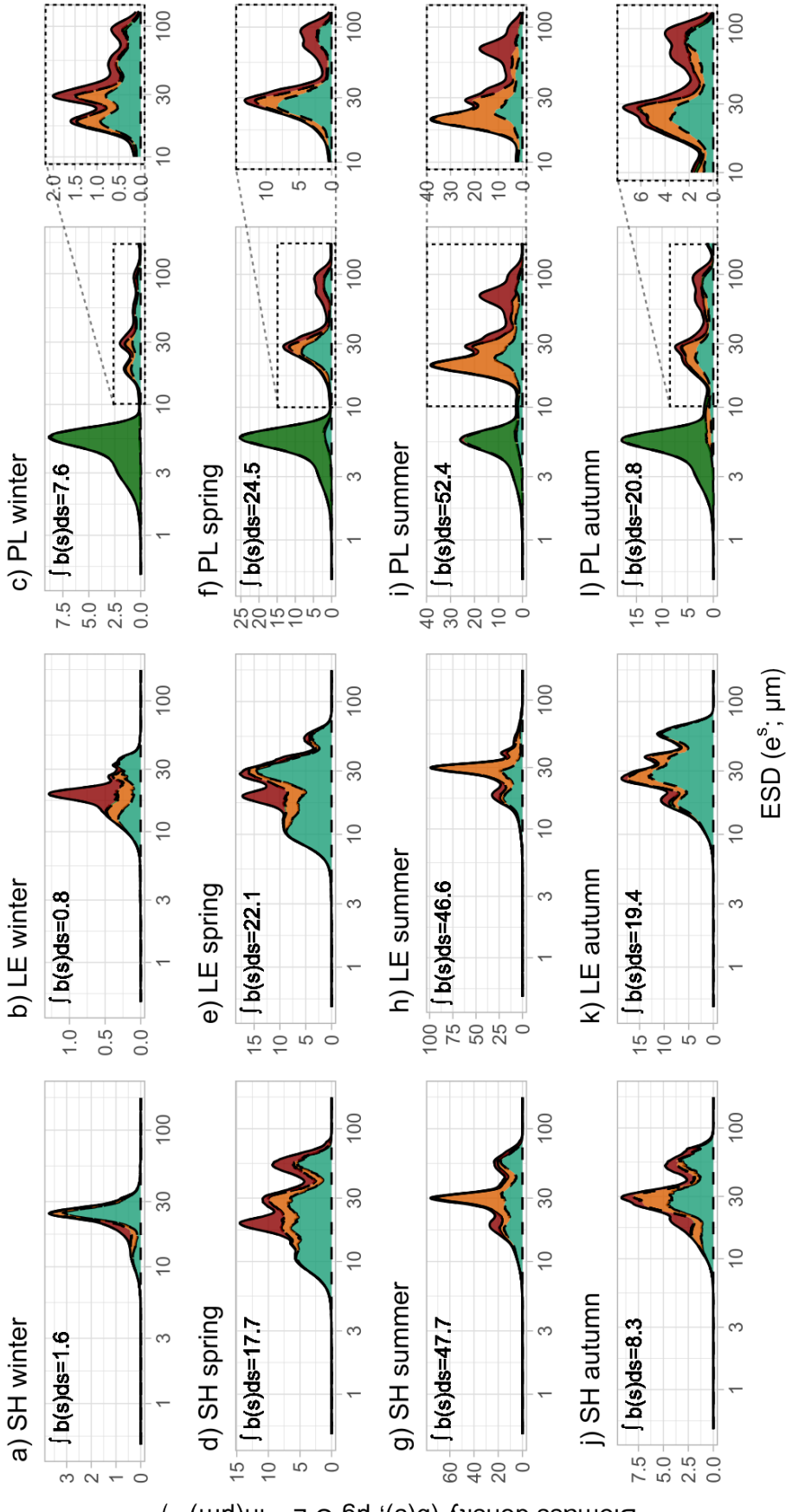


Figure 2.12: Typical biomass distributions of unicellular communities (excluding flow cytometry data) at SH, LE, and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$). For plots at PL, a zoomed window is provided for larger ESD values to show the biomass distributions over the interval $10 - 130$ μm more clearly.

When biomass accumulates in spring, the biomass distributions at SH & LE appear to grow outwards, indicating that accumulations of large nano- and microplankton cover a greater range of sizes. Diatoms contribute biomass relatively evenly across the size domain at SH between $10 - 60\,\mu$m (albeit with a dip at $40\,\mu$m; Figure 2.12d). These contributions are primarily made by *Chaetoceros* spp. ($17.4\,\mu$m), *Skeletonema* spp. ($9.0\,\mu$m), with *Ditylum brightwelii* ($51.9\,\mu$m), and *Guinardia Delicatula* ($22.5\,\mu$m) contributing biomass at larger sizes. Microzooplankton biomass in contributed by ciliates (unidentified; $18.4\,\mu$m). At LE, the overall picture is similar, but composition of the diatom community more dominated by cells sized $30\,\mu$m due to the dominance of *Thalassiosira* spp., while *Chaetoceros spp.* contribute less biomass (Figure 2.12e). At PL, diatoms accumulations are focused around $28\,\mu$m through contribution by *Guinardia delicatula* ($22.5\,\mu$m), *Cerataulina pelagica* ($27.0\,\mu$m), *Lauderia annulata* ($30.0\,\mu$m) , with a second smaller peak around $80\,\mu$m (*Guinardia flaccida*; Figure 2.12f).

In summer, at each of the UK coastal sites, the community attains its (seasonally-averaged) maximum biomass, and shifts towards a biomass dominance of mixoplankton. At SH & LE, a dominant peak is centred around $30\,\mu$m (Figure 2.12g & h) due to contributions by *Tripos lineatus* ($29.8\,\mu$m) and *Karenia mikimotoi* ($29.1\,\mu$). At PL, the mixoplankton peak is also driven by accumulations of *Karenia mikimotoi*, which dominate biomass contributions by an order of magnitude (Figure 2.12i). A second, smaller peak is also contributed by microzooplankton at $\sim 70\,\mu$m by *Gyrodinium spirale* ($66.0\,\mu$m) and *Polykrikos schwarzii* ($72.3\,\mu$m). Diatoms play a secondary role in summer in a proportional sense, having maintained much of their biomass and distributional shape at all three coastal sites. As such, they do not appear to be entirely excluded from the community by mixoplankton, but do appear to be outcompeted.

In autumn, the dinoflagellate surge has diminished. At PL, biomass contributions are balanced across the key groups of diatoms, mixoplankton, and microzooplankton (Figure 2.12l). Around $20\,\mu$m, contributions by diatoms and mixoplankton are approximately equal, while on the $40 - 80\,\mu$m range, diatoms and microzooplankton have similar accumulations. In terms of balance across functional groups, and their contributions, a similar story is true at SH. At $30\,\mu$m mixoplankton and diatoms contribute, and at $50\,\mu$m diatoms and microzooplankton contribute (Figure 2.12j). At LE, mixo- and microzooplankton accumulations are similar to SH in absolute terms, but are less substantial proportionally due to the much greater diatom biomass observed between $20 - 60\,\mu$m at LE in autumn (Figure 2.12k).

2.3.2.3 Multicellular plankton

The multicellular plankton community is characterised primarily by biomass contributions by copepods and fish / meroplankton larvae at all sites. At PL, the biomass of large copepods is dominated by *Calanus helgolandicus* ($857 - 1024\,\mu$m) year-round, whereas the small copepod ($300 - 600\,\mu$) biomass is dominated by a wider range of copepod taxa (see Eloire *et al.*, 2010 for more detail). The contributions at SH & by large copepods are shared by *Calanus helgolandicus* and *Calanus finmarchicus* ($1374 - 1465\,\mu$m), and a diverse community of smaller copepods is also observed.

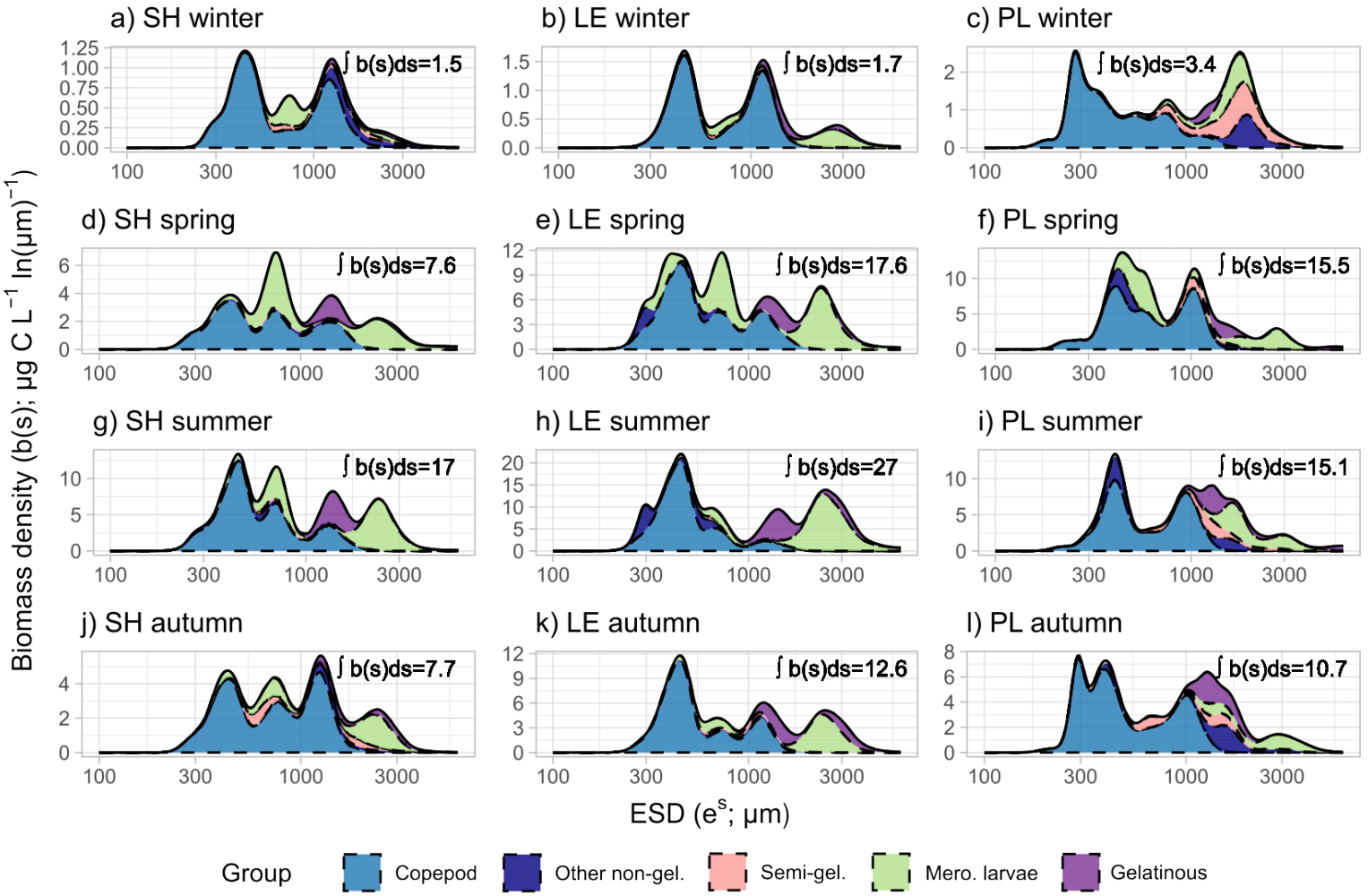


Figure 2.13: Typical biomass distributions of multicellular communities at SH, LE, and PL in winter, spring, summer, and autumn. Annotations on each figure show integrated area beneath each curve, equivalent to total community biomass concentrations (μg C L $^{-1}$).

At all sites, two peaks of fish / meroplankton larvae occur: those of cirripede larvae (514 μm), particularly in spring, and those of fish larvae (1800 − 3000 μm). Gelatinous plankton also contribute biomass, primarily through *Appendicularia* spp. ($\sim$ 1300 μm), while the contributions of semi-gelatinous plankton (*Ploychaete* larvae; $\sim$ 1000 μm) are mainly restricted to PL. The size domain is subject to significant seasonal variation in biomass density, although the overall structure of the community does not show a clear, identifiable seasonal progression in the same way that the unicellular community does (when flow cytometry data is included). Despite this, the biomass distributions do show changes from season-to-season, and proportional contributions from different functional groups within regions of the size domain are also variable.

At SH & LE, the winter community is dominated by copepods with peaks at 450 μm and 1100 μm (Figure 2.13a&b). Small copepods are also important at PL (300 μm) and, while the proportional contribution is lower, copepods up to 800 μm make important biomass contributions (Figure 2.13c). Beyond this, at PL, biomass contributions are made by all other non-larval plankton groups, resulting in a second peak around 1800 μm. Conversely, at SH & LE a smattering of non-copepod contributions are made along the size domain.

During the spring biomass accumulation at SH & LE, copepods remain important over the 300 − 1300 μm range, while cirripede larvae become prominent around 600 μm. At SH, the modal size is driven by these biomass accumulations (Figure 2.13d). At LE, where biomass is consistently higher than at SH, the peak at $\sim$ 600 μm is complemented by larger contributions of small copepods at $\sim$ 400 μm (Figure 2.13e). Furthermore, a third, smaller peak is contributed by fish larvae at 2000 μm. At PL, the biomass distribution becomes narrower, with peaks at $\sim$ 500 μm and 1000 μm contributed primarily by copepods and meroplankton larvae (Figure 2.13f).

In summer, the PL biomass distribution is similar, albeit with less contribution of meroplankton at 500 μm, and the evolution of peak at 1000 μm to plateau at 1000 − 1500 μm due to diversification of biomass contributions, with contributions from all key groups (Figure 2.13i). At SH & LE, the biomass distribution are also similar between spring and summer, although the seasonally-averaged biomass appears to increase significantly, unlike at PL (Figure 2.13g&h). While the meroplankton larvae decline in biomass, fish larvae continue to accumulate biomass. Notably, the large copepods (1100 μm) show a decrease in relative proportion of biomass and a slight reduction in absolute terms at LE, with the community becoming increasingly dominated by smaller copepods.

In autumn, the biomass is reduced across all sites; at PL, this appears to mark a transition to the winter community, with the biomass distribution once again widening out. Small copepods produce bimodal peaks at 300 μm and 400 μm which are complemented by the proportional persistence of the diverse broad peak at 1000−1500 μm (Figure 2.13l). At SH, the main change in the distribution of biomass is the increase (in absolute and proportional terms) of large copepods at 1100 μm. Gelatinous plankton contributions appear to cease (Figure 2.13j). At LE, a similar pattern is seen, although the gelatinous plankton persist to some degree, and the 450 μm copepods are able to maintain a higher biomass than at SH & PL (Figure 2.13k).

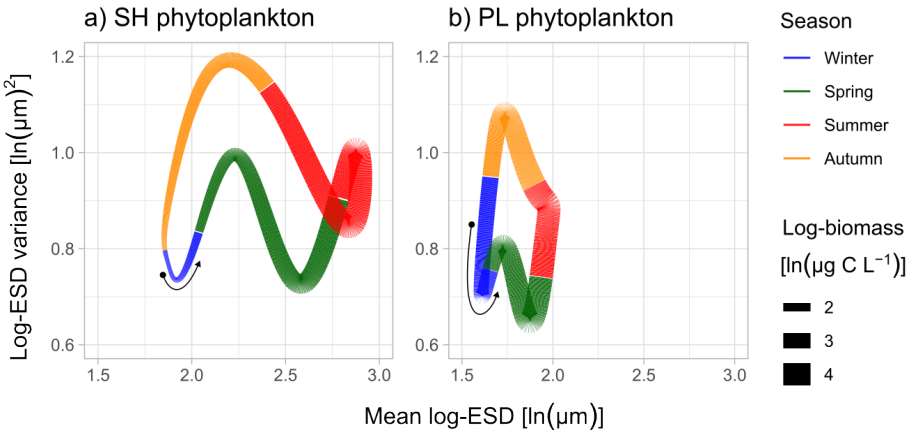


Figure 2.14: Phase portrait showing the typical seasonal cycle of mean size and size variance of phytoplankton measured by microscopy and flow cytometry at SH and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot (·).

An important feature to consider in the context of the zooplankton community is the component of the unicellular zooplankton (protozooplankton). For example, at PL in summer, the truncated mean biomass density for 75 μm microzooplankton is 15 μg C L^{-1} ln(μm)$^{-1}$ (Figure 2.12i), which is comparable to the peak observed in the multicellular community during the same period (13 μg C L^{-1} ln(μm)$^{-1}$ at 400 μm; Figure 2.13i). We favour representing the microzooplankton data on the unicellular plots, rather than multicellular plots, to avoid any complications arising from different sampling methodologies. Additionally, the protozooplankton – while above the phytoplankton in the food web – are still grazed by much of the metazooplankton, due to their size.

2.3.3 Statistical modelling of the seasonality of plankton community size structure

2.3.3.1 Phytoplankton (inc. flow cytometry)

Phytoplankton community size structure shows significant seasonal variation at SH & PL ($p \leq 0.0238$), with a clearly defined cycle of biomass, mean size, and size variance. In the first instance, modelled median biomass is at its minimum in winter (4.1 μg C L^{-1} at SH, 19.8 μg C L^{-1} at PL) and at its maximum in summer (60.8 μg C L^{-1} SH, 79.4 μg C L^{-1} PL). Notably, the biomass is lower at SH, particularly in winter when considered on its modelled logarithmic scale. For reference, the difference in modelled log-biomass between SH and PL is 1.57 in winter, but only 0.27 in summer. When combined with Figure 2.11, the difference in winter biomass appears to be a result of the much larger biomass contribution of small plankton, which also explains the smaller mean size at PL in winter (Figure 2.14).

As biomass transitions between these two states, so too does mean log-ESD, from winter minimum to summer maximum (1.85 – 2.90 ln(μm) at SH, 1.61 – 1.98 ln(μm) at PL). Caution is required when back-transforming modelled mean log-ESD to its original units, since we do not

assume normality of the biomass distribution across log-ESD. However, for context, $e^{2.90} = 18.2$ while $e^{1.98} = 7.2$. This result can be used to emphasise that mean size is lower at PL, particularly in summer. The difference in mean size in summer is driven largely by the growth of larger cells ($> 10\,\mu$m) which, at Scottish sites, is much greater than at PL (Figure 2.11).

The increase of mean log-ESD in spring is not necessarily accompanied by a drastic change in log-ESD variance. In fact, the minimum log-ESD variance is achieved in late-April ($0.73\ \ln(\mu\text{m})^2$ at SH, $0.65\ \ln(\mu\text{m})^2$ at PL), when the mean size is increasing. Comparison with Figure 2.11 cannot give insight to the driver of this change (having been amalgamated over several years), but there is an indication that the proportional decrease of picoplankton may be an important feature (aligning with the increase in mean log-ESD). Log-ESD variance reaches its maximum in autumn ($1.19\ \ln(\mu\text{m})^2$ at SH, $1.09\ \ln(\mu\text{m})^2$ at PL), while biomass and mean size are both decreasing.

Despite differences in the scale of biomass and mean log-ESD at both sites and small differences in the timing of changes, the seasonal cycle of phytoplankton size structure is qualitatively similar at SH & PL. This would indicate that common drivers are forcing the evolution of the size structure. The main difference between the phytoplankton community between the sites is that the picoplankton and small nanoplankton are far more abundant at PL than at SH. While some of this may be accounted for with the absence of nanoflagellate enumeration at SH, it would appear highly unlikely that nanoflagellates contribute such a large degree of biomass, given that the nanoflagellates only contribute a (relatively small) proportion of the small nanoplankton biomass at PL compared to the nanoeukaryotes observed by flow cytometry (via comparison of Figures 2.11 and 2.12).

Much of the cycle of log-biomass is explained by the seasonal smooth ($R^2 = 0.576 - 0.739$), with a very high level of statistical significance ($p < 2 \times 10^{-16}$). The statistical significance of the mean log-ESD is equally as strong, although a lesser degree of variation is explained by the seasonal smooth ($R^2 = 0.126 - 0.310$). In both instances, R^2 is higher at SH. Log-ESD variance shows a lesser degree of statistical significance at SH ($p = 0.0238$), and similar R^2 across both sites ($R^2 \simeq 0.07$). Based on subjective interpretation of the data, phase portraits, and R^2 values, it appears likely that the larger p-value at SH arises from a smaller collection of data when flow cytometry was collected ($n = 144$, spanning $2015 - 2018$ inclusive). In all other instances – where flow cytometry data is not necessary – $n \geq 820$, giving much greater statistical power.

2.3.3.2 Zooplankton (inc. microzooplankton)

Zooplankton, inclusive of microzooplankton (but not heterotrophic nanoeukaryotes measured by flow cytometry at PL), also exhibit a statistically significant seasonal cycle of community size structure ($p \leq 9.72 \times 10^{-6}$). However, the cycle itself is somewhat complex, with more nuance than that demonstrated by the phytoplankton. In the first instance, median biomass behaves as expected, with minimum occurring in winter ($\sim 1.2\,\mu$g C L^{-1} at SH & LE, $3.2\,\mu$g C L^{-1} at PL) and the maximum in summer ($30.7 - 34.9\,\mu$g C L^{-1}). Again, the biomass is proportionally higher at PL in winter, while summer values at broadly similar across sites.

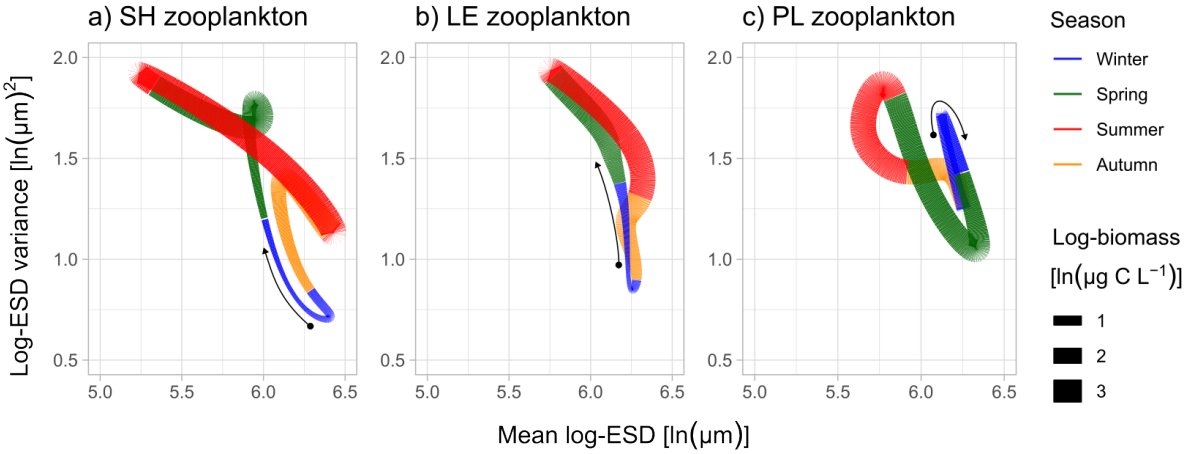


Figure 2.15: Phase portrait showing the typical seasonal cycle of mean size and size variance of micro- and mesozooplankton at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot (·).

The winter community is defined by large mean log-ESD ($\sim$ 6.25 ln(μm)) and, at SH & LE the annual minimum log-ESD variance (0.70 and 0.84 ln(μm)^2, respectively). At PL, the size variance is slightly elevated but soon approaches its minimum in early spring (1.04 ln(μm)^2). At all three sites, while biomass accumulates through spring, the mean size decreases and size variance increases. At SH, the decrease is approximately -50% on the original scale of ESD, while at LE & PL is -40%. The size approaches 5.25 ln(μm) at SH ($e^{5.25} = 190$) indicating that the contribution of the microzooplankton community is very important by late spring at SH & LE, and early summer at PL. At this point, size variance is also at its maximum at all three UK coastal sites ($\sim$ 1.9 ln(μm)^2). Thereafter, the mean size decreases with increasing mean size. In autumn, when biomass begins to diminish, the size variance decreases; at SH & LE, there is little change in mean log-ESD, while at PL there is a small increase.

Statistical models fitted to the zooplankton community at SH & PL yielded highly significant seasonal terms, while the p-values were unreliable at LE due to the residual error distribution. This may help to explain the particularly low proportion of variability explained at LE ($R^2 = 0.0343$) when compared to SH & PL ($R^2 = 0.145$ & $R^2 = 0.129$ respectively; Table B.1). A reasonable explanation for this is that the model did not capture the high-biomass occurrences of microzooplankton particularly well (or that these occurrences were not mirrored by the large mesozooplankton community). The R^2 values were lower at LE than SH in all occurrences across the fitted zooplankton models, indicating that there may be a more complex set of drivers for size structure at LE, or a greater degree of randomness.

With the exception of winter log-ESD variance at PL, the seasonality of zooplankton community size structure at UK coastal sites can be reconciled to a relatively common theme – perhaps more so than first glance at Figure 2.15 would suggest. The mean size is driven down with increasing biomass by the accumulation of microzooplankton until late spring / early summer. Thereafter, mean size then increases with further slightly increasing biomass, as the mesozoo-

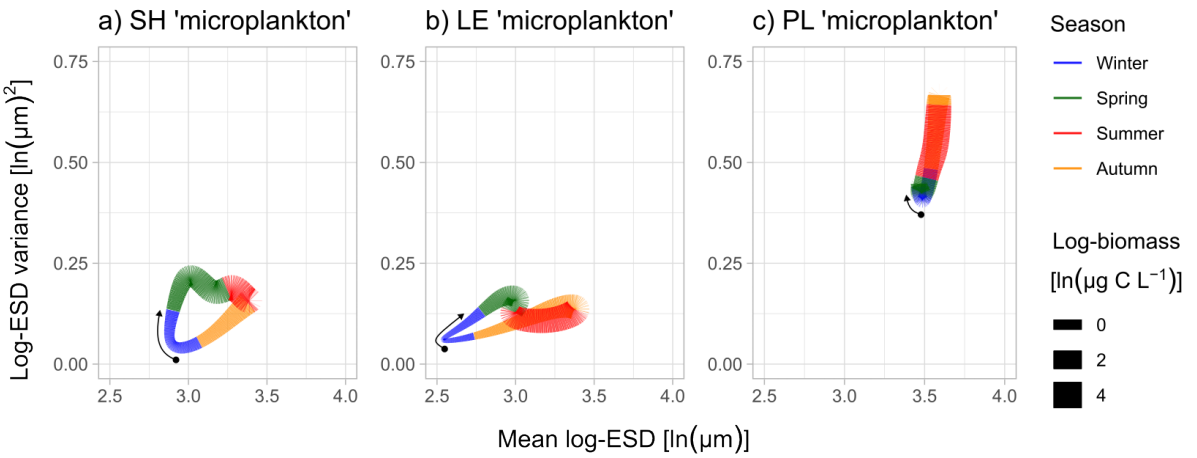


Figure 2.16: Phase portrait showing the typical seasonal cycle of mean size and size variance of 'microplankton' (diatoms, dinoflagellates, and ciliates) measured by microscopy at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot (·).

plankton contribute a greater proportion of the community. In autumn, a transition is seen to the winter state, more dominated by large organisms, with low biomass.

2.3.3.3 Microplankton

Although the term 'microplankton' is a misnomer (as it incorporates large nanoplankton on the range $10 - 20\,\mu$m), it is often used to refer to diatoms, dinoflagellates, and ciliates which form the prey of copepods and other important mesozooplankton groups. For example, Banas (2011) predicts that zooplankton on the $150 - 470\,\mu$m range would feed on $10 - 20\,\mu$m plankton (using data from Hansen *et al.*, 1994). This incorporates the small copepods which make a crucial contribution to the zooplankton community (Figure 2.13). Moreover, this aspect of the unicellular plankton has been studied for a longer period, and at all three UK coastal sites. As such, there is clear value in understanding the size structure of this portion of the community.

The seasonal cycle of microplankton community size structure is relatively similar between SH & LE, while PL shows less seasonal variability. At each of the UK coastal sites, the median microplankton biomass is lowest in winter ($0.2\,\mu$g C L^{-1} at LE, $0.9\,\mu$g C L^{-1} at SH, $2.2\,\mu$g C L^{-1} at PL), and highest in summer ($\sim 54\,\mu$g C L^{-1} at all sites). Particular note should be paid to LE in winter, where only a few cells were counted in an average subsample, meaning that estimates are likely to be erratic due to a high degree of extrapolation.

The cycle of mean log-ESD is much more pronounced at the Scottish sites, while PL shows little (but significant; $p = 0.0017$) seasonal variability. In mid-January, the fitted mean log-ESD is 2.55 ln(μm) at SH, 2.85 ln(μm) at LE, and 3.5 ln(μm) at PL (approximately $12.8, 17.3, 33.1\,\mu$m on original units). Yet, the mean log-ESD is similar across UK coastal sites when it approaches its maximum at the end of summer (~ 3.5 ln(μm)). At Scottish sites, the transition between the size minimum in winter and size maximum in summer are relatively consistent and sustained.

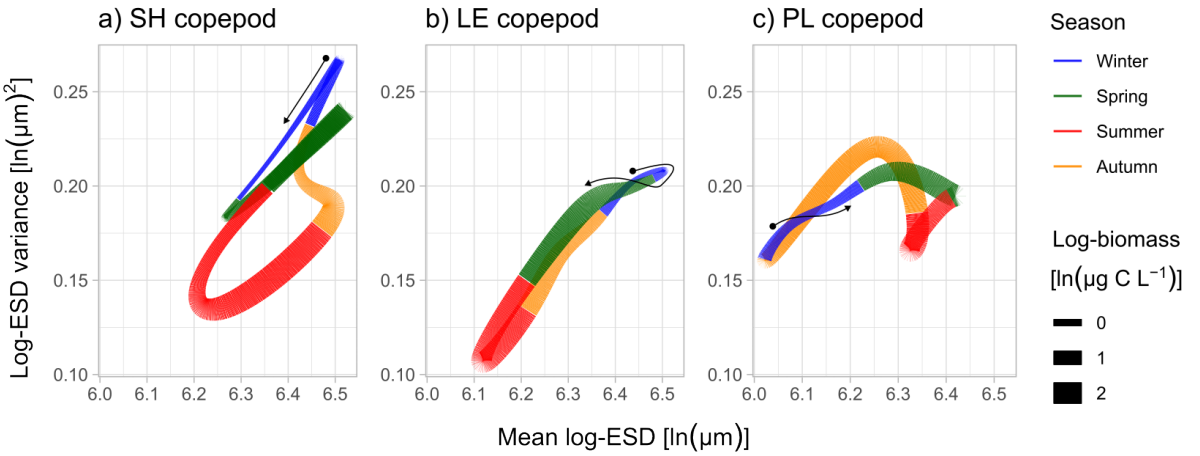


Figure 2.17: Phase portrait showing the typical seasonal cycle of mean size and size variance of copepods at SH, LE, and PL. Day of year progresses in the direction of annotated arrows, with the 1st of January marked by the circular dot (·).

During the period of biomass accumulation in spring, the size variance increases to its maximum at SH & LE (0.22 $\ln(\mu\text{m})^2$ and 0.17 $\ln(\mu\text{m})^2$, respectively) which reduces slightly before being sustained through the summer. At PL, it is not until the summer that the size variance increases. However, the most noticeable feature is that the size variance is much higher at PL ($0.40 - 0.67\ \ln(\mu\text{m})^2$) than at SH & LE ($0.04 - 0.22\ \ln(\mu\text{m})^2$). This most likely points towards the difference of subsample volume used between the Scottish sites and PL, which means that low-abundance cells of different taxa are more likely to be missed.

With the exception of mean log-ESD at PL ($p = 0.00172$), all models have $p < 2 \times 10^{-16}$. As is the case with all other seasonal GAMs fitted in this section, the predictive power for the model is strong for biomass ($R^2 > 0.53$), but is lower when fitted to mean log-ESD and log-ESD variance. This suggests that less variability in mean size and size variance is explained purely by seasonal changes, and other factors may affect these community traits to a greater degree than biomass.

2.3.3.4 Copepods

The observed cycle of size structure in the copepod community differs between Scottish sites and PL. At SH & LE, mean log-ESD is higher in midwinter and lower in midsummer, with log-ESD variance similarly decreasing over the period. Conversely, at PL, the copepod mean size is lower in winter and higher in summer. This is most likely taxa riven, although it may also be caused by differences in data collection methods. At SH & LE, taxa are measured with associated copepodite stage. At PL, taxa are measured and their size varies by season, based on the typical size per taxon by season. The transition between winter and summer at SH seems more erratic than at LE, which is more linear. At PL, the cycle of size appears smooth, and the size variances reaches its maximum during the transitional periods of spring and autumn.

With the exception of a few summertime observations where biomass is lower than expected, models fitted well and yielded significant seasonal terms ($p < 2 \times 10^{-16}$).

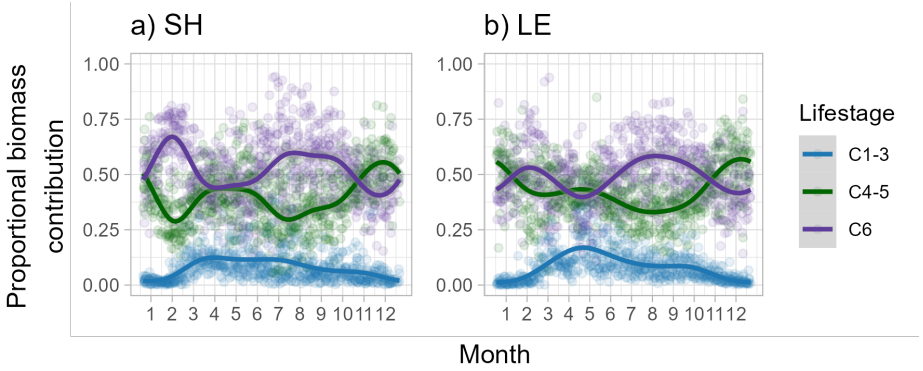


Figure 2.18: Seasonal cycle of copepod biomass contribution at SH, LE, and PL by categorised copepodite stages.

At SH & LE, there is evidence of changing age structure in the copepod community, and likely plays a role in driving the size structure of the community. A greater proportion of juveniles ($C1 - 5$) are present in the community during spring and summer (Figure 2.18). For any given species, a taxon's typical size is strictly increasing with age. It follows that increased biomass contribution from the early copepodite stages likely drives the mean size downwards through spring; however, the (approximate) maximum proportion of C6 copepods is achieved in July / August, when mean size is at its minimum. This suggests that a taxonomic driver must force the copepod mean size to be small in summer, rather than age structure, at SH & LE.

The diet of copepods also shows seasonal variability, as well as differences between sites. At PL, a much greater proportion of biomass is contributed by herbivores year-round. At SH & LE, herbivorous taxa and carnivorous taxa contribute a similar level of biomass (with herbivores slightly more on average). At SH, herbivores are proportionally dominant between September and May ($\sim$ 65% of biomass contributed), with carnivores rising to a peak in July (70%) and decreasing smoothly thereafter. At LE, the state between July and February is approximately evenly split between the herbivores and omnivores, then herbivores peak in April / May (75%). At PL, the omnivores reach a maximum in October (35%). The food web dynamics across the UK coastal sites appear to have differences, yet the phytoplankton community appears to show many similarities across UK coastal sites.

2.3.3.5 Long-term variability of seasonality

The plankton community at each of the UK coastal sites show relatively consistent seasonal cycles when partitioned into 5-year periods for the unicellular and multicellular communities (flow cytometry omitted due to data availability). Biomass – which is the most predictable of each of the indicators in a seasonal context, reflected by its R^2 – shows very little long-term variability in seasonality. However, an important and notable exception is the multicellular plankton between $2015 - 2017$ at LE (data do not extend to 2019; Figure C.1e). The observed biomass is notably lower during this period and concurs with the findings of Wells *et al.* (2022), who showed this in an abundance context. Despite this, no major differences are observed in

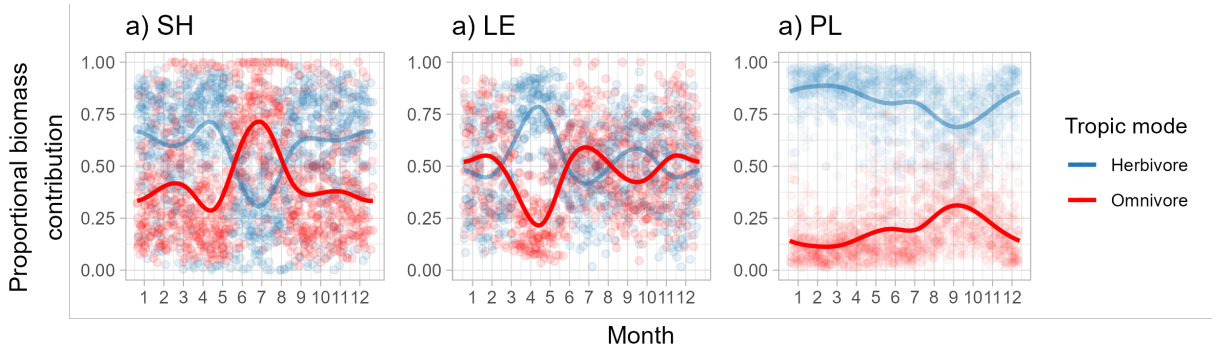


Figure 2.19: Seasonal cycle of copepod biomass contribution at SH, LE, and PL by diet.

mean log-ESD or log-ESD variance in the uni- or multicellular community at SH & LE (plots a,b,d,e of Figures 2.20 & C.2). Despite this, a greater degree of variation between each 5-year time period is observed, which likely contributes to the decrease in R^2 compared to the fitted models for biomass across the complete timescale.

At PL, there is some evidence of change in the size structure across $2000-2019$. Mean log-ESD appears to increase consistently from the $2000-2004$ period to the $2015-2019$ period in both the unicellular and multicellular communities. The maximum log-ESD – which occurs in the uni- and multicellular communities increases by around 0.75 $\ln(\mu\text{m})$ in both instances from the $2000-2004$ to the $2015-2019$ period. This is an approximate doubling on the original units of ESD. In the multicellular community, this difference is also observed in winter (with spring and autumn showing less long-term variability).

2.3.4 Relationships between indicators of size structure

2.3.4.1 Relationships between indicators within phyto- and zooplankton communities

Significant relationships were observed between mean size and biomass, and size variance and biomass within the phyto- and zooplankton communities at SH & PL. In each instance, a test of a null hypothesis of ubiquitously zero slope coefficients was rejected, highlighting that the biomass was related to the mean size and size variance ($p \leq 1.16 \times 10^{-6}$; Table B.2). The scale of these relationships varied across sites, albeit with many common themes (Figure 2.21).

In the first instance, a linearly increasing relationship was observed between log-biomass and mean log-ESD in phytoplankton communities at PL & SH (Figure 2.21a). Unsurprisingly, the fitted value of mean log-ESD was higher at SH for any value of log-biomass on the range of observations (following from Figure 2.14). However, the slope at SH was also significantly greater ($p = 1.69 \times 10^{-7}$; Table B.2), estimated as approximately double that of PL. This likely relates to the year-round dominance of the small nanoplankton observed at PL (Figure 2.11), whereas the dominance appears to shift from small nanoplankton to large nanoplankton and microplankton at SH during high-biomass periods at SH. Within the zooplankton, a negative relationship is observed between mean size and biomass (Figure 2.21b). The form of this relationship is linear at

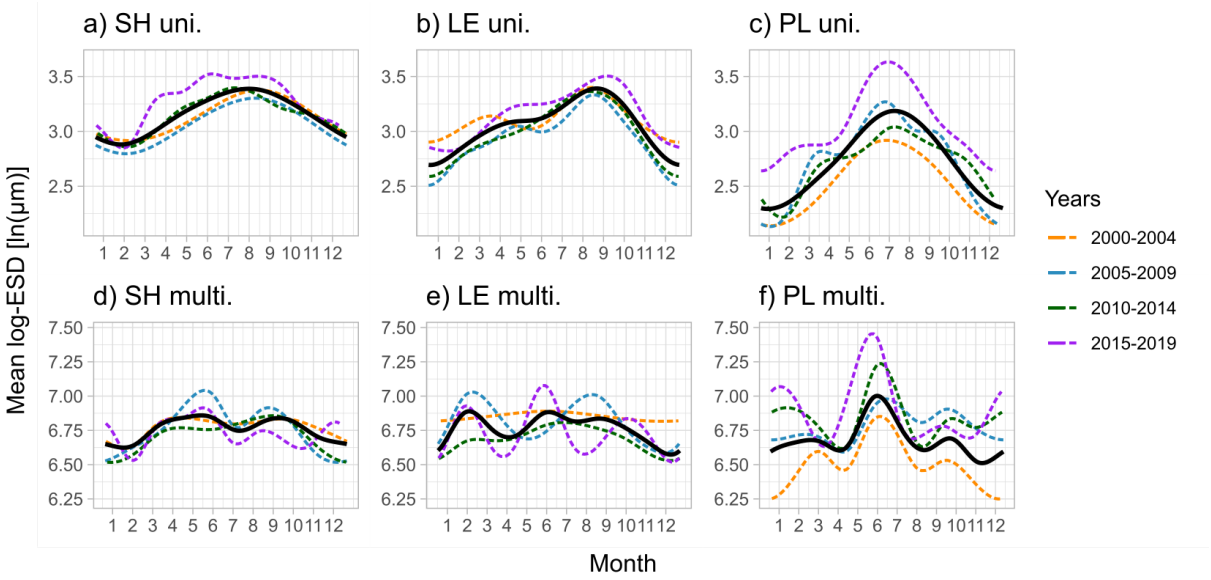


Figure 2.20: Seasonal cycle of mean log-ESD of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined (2000 – 2019) data.

SH, and quadratic at PL. At SH, the second order polynomial term for the log-biomass variable was included to improve the residual error distribution, but was not statistically significant ($p = 0.617$). The first order term was significant ($p = 1.87 \times 10^{-4}$), leading to the conclusion that the mean size is negatively related to biomass in the zooplankton (but that the model cannot adequately conclude a linear relationship). At PL, the inclusion of the second order polynomial was significant ($p = 0.0363$) and reflected the apparent lack of association between size and biomass for biomass approximately $< 10\,\mu$g C L^{-1}, and negative relationship thereafter. In both instances, the relationships is almost certainly driven by the increasing proportional contribution of unicellular microzooplankton as biomass increases.

Phytoplankton size variance exhibits a significant unimodal relationship with biomass at both sites. At SH, the second order polynomial term is significant ($p = 7.05 \times 10^{-4}$), while at PL, the interactive second order polynomial term is also significant, with a negative estimate indicating a greater degree of curvature ($p = 5.62 \times 10^{-3}$). Within the zooplankton, the results are somewhat mixed. Inclusion of all data demonstrates that zooplankton size diversity exhibits a unimodal relationship with biomass at SH (second order polynomial term $p = 1.77 \times 10^{-8}$). However, when the observations with no microzooplankton recorded are excluded, the relationship is negative linear ($p = 2.79 \times 10^{-11}$, with no significant evidence of a quadratic fit). It is indeed possible that no microzooplankton are present in the water column when zooplankton biomass is low, although may also be an artefact of the small subsample volume used for microscopy analysis of unicellular plankton at SH. As such, both results are presented. At PL, no significant relationship is observed between zooplankton size variance and biomass. Although when fitted as an effect in the overall LM PL exhibits significant differences to SH, the p-values of the first and second-order polynomials have $p > 0.05$ when fitted in their own right. As such, the relationship differs between PL and SH because no significant relationship is observed at PL.

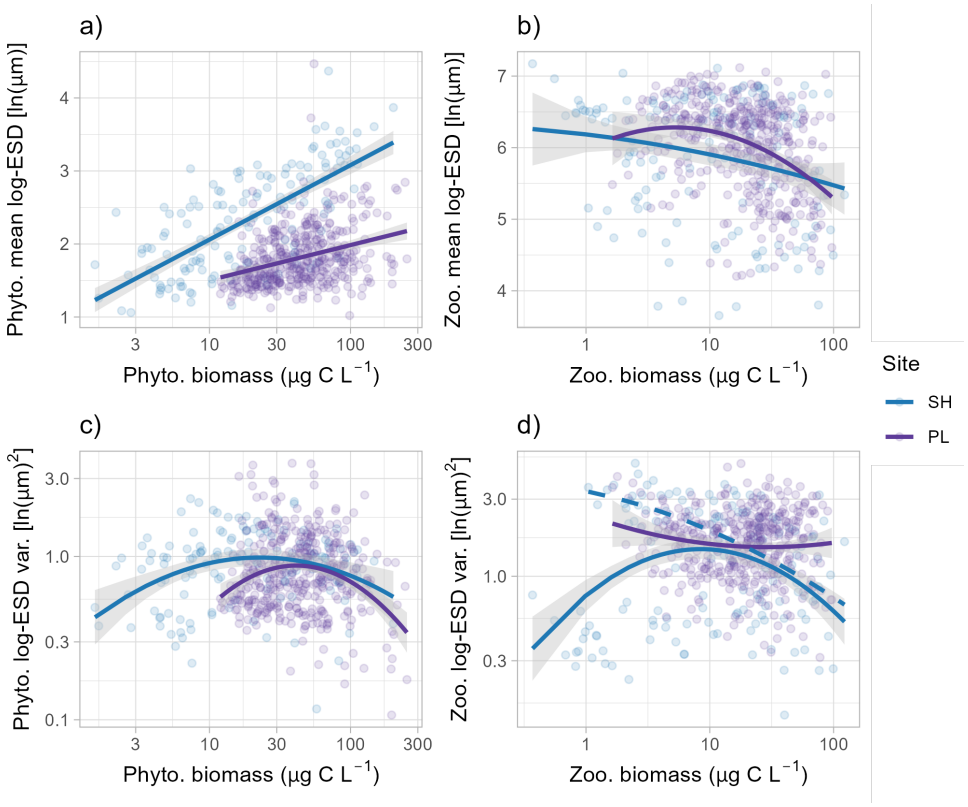


Figure 2.21: Relationships between biomass and mean size, and biomass and size variance in phytoplankton and zooplankton communities at SH and PL. LE data excluded due to the absence of flow cytometry analysis of small cells.

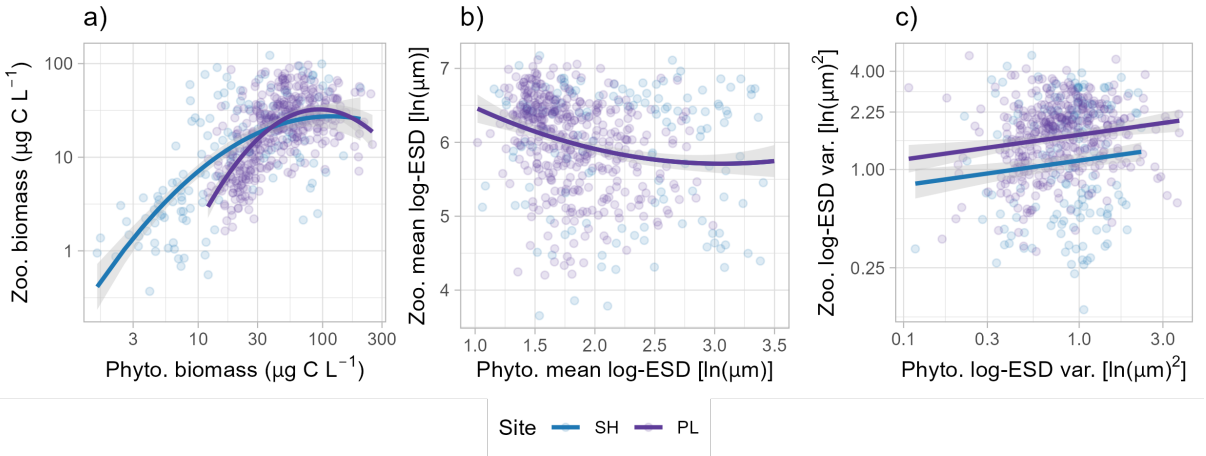


Figure 2.22: Relationships of biomass, mean size and size variance between phytoplankton and zooplankton communities at SH and PL. $n = 9$ observations removed from plot in (b), but not model fitting. No significant difference between curves for SH & PL in (b). LE data excluded due to the absence of flow cytometry analysis of small cells.

2.3.4.2 Relationships between indicators across uni- and multicellular communities

Significant relationships are observed between phytoplankton and zooplankton across biomass, mean size, and size variance (overall model fit $p \leq 1.59 \times 10^{-10}$; Figure 2.22). The qualitative patterns are consistent across sites, and only for biomass is there is a significant difference in the quantitative estimates of slope parameters ($p = 9.62 \times 10^{-16}$ for the first order term, and $p = 1.98 \times 10^{-4}$ for the second order term).

The relationship between phyto- and zooplankton biomass is positive across much of its range, but the zooplankton biomass shows evidence of saturation for values of phytoplankton biomass approximately $> 100\,\mu$g C L^{-1}. Although at PL the slope becomes negative after this point, there are too few observations to support such a conclusion. Furthermore, fit of the first and second order polynomial is inherently driven by *all* observations under ordinary least squares estimation. As such the extent of curvature may be driven by observations on the domain of, for example, $< 30\,\mu$g C L^{-1} phytoplankton biomass as opposed to observations $> 100\,\mu$g C L^{-1}. A GAM may be more robust for exploring the relationship past this point, although wide confidence limits would still be expected $> 100\,\mu$g C L^{-1} due to the lack of observations. A conservative conclusion is that zooplankton biomass increases with phytoplankton biomass, qualitatively in alignment with the 3/4 scaling law of predator-prey biomass (Hatton *et al.*, 2015), but shows evidence of saturation (or potential decrease) at around $100\,\mu$g C L^{-1} phytoplankton biomass at both sites.

Zooplankton mean size exhibits a negative relationship with phytoplankton mean size over the range for which most of the observations are made ($p < 2.2 \times 10^{-16}$), although the model has poor explanatory power ($R^2 = 0.06374$). There is no evidence of differences in intercept or either slope coefficient between SH and PL ($p \geq 0.0878$ upon t-testing of individual terms). Although the second order polynomial term is significant ($p = 6.62 \times 10^{-3}$), this only acts to reduce the magnitude of the slope with increasing phytoplankton mean size. There is insufficient data to support a positive relationship between phyto- and zooplankton mean size on any of the observed domain. As such the inference made is that there is a negative relationship between phyto- and zooplankton mean size over the range where much of the data are collected, and that there is some evidence that this relationship levels off with increasing phytoplankton mean size.

A positive linear relationship is observed between zooplankton size variance and phytoplankton size variance at both sites (Figure 2.22c; note log-log scale necessary for residual structure in this model but not those of log-ESD variance in Figure 2.21b&d). The slope is significant at both sites ($p = 1.55 \times 10^{-3}$), although no significant difference is observed in the slope between SH & PL ($p = 0.9581$). However, the intercept is greater at PL indicating that the zooplankton community is more diverse than at SH, on average, along the observed range of phytoplankton size diversity ($p = 7.23 \times 10^{-10}$).

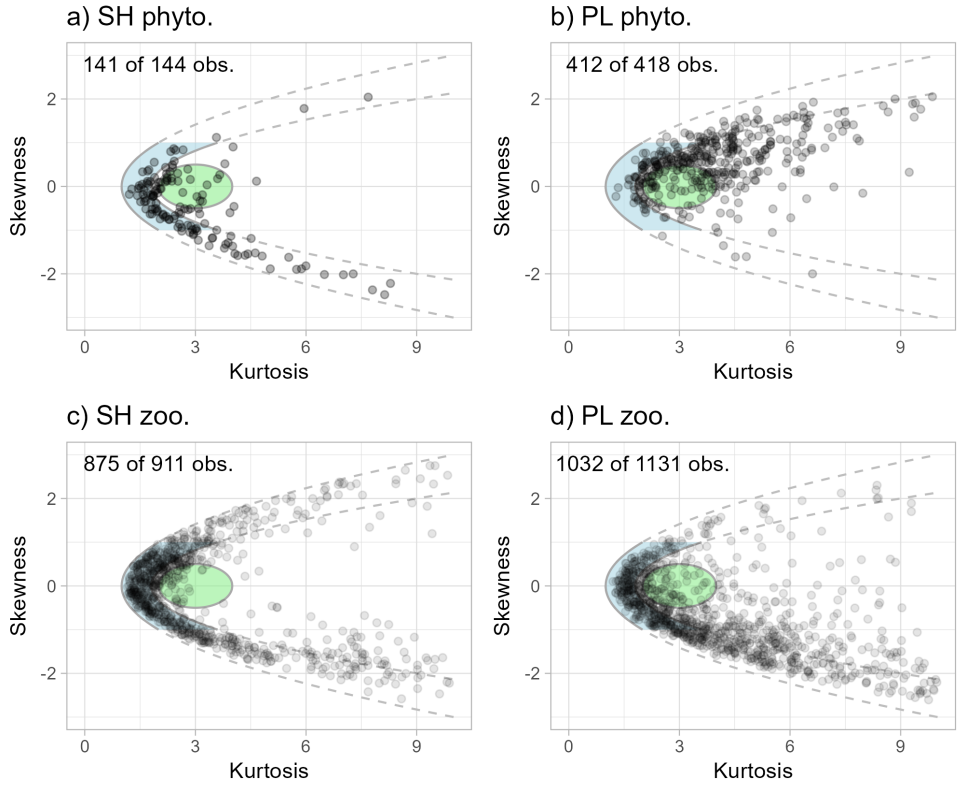


Figure 2.23: Skewness-kurtosis pairs plot for biomass distributions at SH & PL. The green shaded elliptical area indicates the region of normality. The outermost dashed line indicates the boundary of the attainable domain for skewness-kurtosis pairs. The region between dashed lines indicates candidate region for bimodality, but only points in the blue shaded region are considered to be bimodal. Some extreme points omitted to allow appropriate plot scaling; the number of displayed points is annotated on each figure. LE data excluded due to the absence of flow cytometry analysis of small cells.

2.3.4.3 Shape of the biomass-size distribution

Skewness and kurtosis were calculated for individual biomass distributions to demonstrate the distributional tendencies of the community at each site. These vary between phyto- and zooplankton and between sites.

At SH, phytoplankton biomass distributions are typically clustered in the bimodal (41.0%) and strongly negative skewed regions (23.6%). This is reflective of two of the main distributional shapes observed at SH, the first of which is dominated by larger phytoplankton on $> 10\,\mu$m with notable (but less substantial) contributions $< 6\,\mu$m, and the second of which is composed of approximately equal parts biomass contributed by phytoplankton from these size classes. At PL, positive skew is far more common in the phytoplankton community, and is driven by the dominance of small cells. Furthermore, although the prevalence of strong positive skew is of similar to that of the strong negative skew at SH (22.5%), bimodality is less prevalent (11.7%). Variations of a positively skewed or symmetric unimodal distribution are therefore more representative of the PL phytoplankton community, as opposed to a bimodal distribution.

These results align qualitatively with Figure 2.11 and support its validity in representing the communities, despite amalgamation across a multi-year scale.

Of the four plots shown in Figure 2.23, SH zooplankton show the greatest degree of symmetry in the distribution of skewness estimates (mean skewness of -0.16). Coupled with the variability exhibited in the skewness-kurtosis pairs, it would appear that a fixed characteristic cannot be assigned to the zooplankton distribution at SH, other than the fact that the distributions themselves are highly variable. However, bimodality is common (43.3%), and unimodal biomass distributions are rare. At PL, there is a tendency towards negative skewed zooplankton biomass distributions, with 40.8% of observations exhibiting strong negative skew. This is most likely related to the greater contributions of microzooplankton at PL than at SH, demonstrated in Figure 2.12. Nonetheless, bimodality is also prevalent, albeit to a lesser extent than at SH (26.1%). In terms of distributional shape, the zooplankton appear to show more variability than the phytoplankton, reflected by a larger variability in skewness and kurtosis than their phytoplankton counterparts.

2.4 Discussion

2.4.1 Comparison to predictions from previous research at UK coastal sites

2.4.1.1 Winter size structure

Our results align with predictions about phytoplankton winter mean size (P1i) and size variance (P1ii) at SH & PL, where flow cytometry data are collected. The mean phytoplankton size is at its minimum at both sites in winter, which is driven by a dominance of small nanoplankton ($\sim 4\,\mu$m) and, to a lesser degree, *Synechococcus* ($1.5\,\mu$m). At all three coastal sites (including LE), the biomass of larger phytoplankton, measured by light microscopy is low. The minimal contributions by larger nano- and microplankton drives the size variance ot be low, as expected, in winter at SH & PL. However, this is not necessarily the only annual minimum. Similar, or smaller, values of phytoplankton size variances occur during the spring bloom at both sites.

The zooplankton size structure somewhat aligns with predictions of low mean size (P2i) and size variance (P2ii), which was predicted due to the relative absence of microzooplankton from the community at all three coastal sites. However, the patterns are not as clear as for phytoplankton, and are perhaps more nuanced. Firstly, the zooplankton community is at, or near, its maximum mean size in winter at all three UK coastal sites. In this period, the biomass distribution of metazoan plankton does not clearly appear to have a larger mean size, and so the relative absence of microzooplankton appears to be a reasonable explanation for this predicted pattern. At SH & LE, the zooplankton size variance is also at its annual minimum size variance. However, at PL, the size variance in winter is actually relatively high at PL, in contrast to expectations. This appears to be a result of the parting of the two biomass modes of metazooplankton at PL. In winter, these are $\sim 300\,\mu$m and $2000\,\mu$m, whereas in spring these are $\sim 450\,\mu$m and $1000\,\mu$m. Note also that we excluded heterotrophic nanoplankton from the computation of mean size and

size variance at PL, but that these also contribute significant biomass.

2.4.1.2 Spring size structure

The introduction of larger phytoplankton in spring, particularly diatoms, drives the phytoplankton mean size to be larger, as predicted (P3i) at SH & PL. However, the size variance does not necessarily increase at either site, as anticipated (H2ii). Interestingly, the mean log-ESD at SH is fitted as 2.5 $\ln(\mu\text{m})$ during this period (approximately equivalent to 12.2 μm), meaning that the reduction in size variance is not driven purely by dominance of small nanoplankton. As such, a proportional reduction in picoplankton may be an important driver for the reduction in size diversity in spring. At PL, however, the same may not be true: biomass distributions indicate that the small nanoplankton remain dominant (supported by fitted mean size approximately equivalent to 6.7 μm). It is therefore possible that the low size variance is driven by dominance at PL, rather than reduced proportional picoplankton biomass.

In spring, we observed increasing biomass of microzooplankton, copepods, and meroplankton in line with previous studies (Bresnan *et al.*, 2015; McEvoy *et al.*, 2023). We predicted that mean size decreases (P4i) and size variance increases (P4ii) as a result of these changes. At SH & LE, the decrease in mean size is originally relatively small in spring, in relative alignment with this statement. This concurs with the timing of the cirripede larvae pulse, which generally occurs in March or April. However, the size structure of the copepod community evolved differently at each site. At SH, a pulse of increasing mean size − likely driven by immigration of *Calanus finmarchicus* − is observed. At LE, copepod mean size is decreasing uniformly through spring, where at PL it is uniformly increasing. It is therefore difficult to dissect the effects of the copepod community on mean size. Nonetheless, at all sites in late spring − after the main cirripede larvae pule − the mean size decreases. Given that size variance also increases during this period (in line with P4ii), the increasing microzooplankton community appears to drive these changes. Fish eggs & larvae also contribute substantial biomass and are likely to increase the size variance, given their large size.

2.4.1.3 Summer size structure

Over summer, we predicted that phytoplankton mean size would remain relatively constant (P5i), and size variance would increase (P5ii) as a result of the succession from diatom to dinoflagellate dominance of the community. At PL, the mean size shows little variability throughout summer, with a slight increase following be a slight decrease, in line with P5i. However, at SH, the mean size decreases substantially in mid-to-late summer. This perhaps reflects the transition from the summer to autumn states, with the former dominated by microplankton, and the latter a more balanced community. Such a result would also align with the observed increase in size diversity at SH (P5ii). At PL, size diversity is also increasing but apparently for the opposite reason: the nanoplankton no longer dominate, and the microplankton contributions support a more balanced distribution. As such, increases in size diversity do not appear to be driven by the emergence of dinoflagellate,s but rather variations in biomass on the larger scale of the size domain.

2.4.2 Drivers of the seasonal cycle

2.4.2.1 Difference of phytoplankton biomass at SH & PL

In the first instance, phytoplankton biomass follows an expected strong seasonal cycle at each of the UK coastal sites, as has already been established (Smyth *et al.*, 2010; McEvoy *et al.*, 2023; Rodríguez *et al.*, 2000; Bresnan *et al.*, 2009, 2015; González-Gil *et al.*, 2022). In the fitted GAMs, the log-biomass minimum occurs in winter, and the maximum in early summer. It should be noted that this does not necessarily align with the time at which the annual observed maximum biomass observation occurs. This is shown in González-Gil *et al.* (2022) at SH, where the timing of the biomass maximum at SH is observed in early May, yet the modelled maximum is in early June. Indeed, the fitted value in early May is exceeded until late July. There are two reasons for this: firstly, due to smoothing by GAMs; secondly, time is treated as fixed (rather than variable), despite the face that *timing* of the bloom is almost certainly a variable in itself, influenced by both bottom-up and top-down factors. As such, the approach undertaken in our work is therefore reflective of overall seasonal trends, but poorly resolves the timing of individual events. It is also worth noting that there are some late-summer / autumn biomass accumulations observed at SH, however, these are less substantial and far less frequent than in Spring (González-Gil *et al.*, 2022), and indeed compared to the late-summer / autumn blooms at LE (Bresnan *et al.*, 2015) & PL (Smyth *et al.*, 2010; McEvoy *et al.*, 2023).

A key result in this chapter is the quantitative difference in phytoplankton biomass between SH & PL, which is largely driven by the difference in nanoplankton biomass (and thus reflected by smaller mean size; Figure 2.14). Year-round, the fitted median phytoplankton biomass is typically higher at PL by $15 - 20\,\mu$g C L^{-1}, but the ratio of fitted biomass at PL compared to SH is much higher in winter (4.83 : 1) than in summer (1.31 : 1). This is largely driven by a background community of small cells at PL which is consistently larger than at SH. In the apparent absence of substantial microzooplankton biomass at SH in winter to control the small cell phytoplankton community, a reasonable hypothesis would be that the difference in wintertime biomass is driven by some bottom-up control. A brief assessment of the nutrient status (Figure 2.4) would indicate that nutrients are not limiting and, as such, the most possible candidates to limit the phytoplankton biomass are (i) light, or (ii) temperature.

Light is more restricted at SH year-round than at PL, and the limitation on photosynthesis is predicted to be more severe in when light is low (in winter) than when light is high (in summer; Equation 1.8). Light is widely accepted as an important prerequisite for the accumulation and retention of phytoplankton biomass (e.g., Huisman *et al.*, 1999b; Sommer & Lengfellner, 2008), even among hypotheses focused on mixing and top-down control. Indeed, Behrenfeld (2010) states that, "From [December−February] forward to the end of bloom, grazing pressure increases (because decoupling by dilution has ended) but is countered by light driven increases, allowing [growth rate] to remain high (even increase) well into the spring". As such, it follows that differences in light between SH & PL are likely to impact biomass accumulation. This may be further enhanced by the fact that the attenuation at SH is much greater in winter than at

PL, reducing the proportion of PAR penetrating the water column, which is known to impact dynamics of biomass accumulation (Tian *et al.*, 2009). Finally, a possible third effect relates to the ocean albedo, which is generally greater at higher latitudes, particularly in winter (29% at 57°N vs 14.2% at 50°N averaged from December−February; Kukla & Robinson, 1980). However, *in situ* observations would be necessary to investigate albedo, due to influences such as swell and atmospheric conditions which vary longitudinally at any given latitude (Payne, 1972). As such, in the absence of location-specific observations for ocean albedo, PAR and attenuation are considered here as potential drivers for the difference in biomass. These effects may be additive or synergistic, due to the shape of the P-I curve (Equation 1.8). This gives rise to the first testable hypothesis: light limitation drives the difference in phytoplankton biomass between PL and SH in winter (H1i).

An alternative explanation may relate to the temperature difference between SH and PL, since both phyto- and zooplankton biological rates are affected by temperature. Typically, the community-level temperature response is described as monotonically increasing by an Eppley curve (Eppley, 1972). With an approximately constant difference of 3°C between SH & PL across the course of the year, the Eppley curve predicts that the phytoplankton growth is approximately 13% higher at PL then SH year-round. Yet, zooplankton tend to be more sensitive to temperature than phytoplankton (Rose & Caron, 2007; Chen *et al.*, 2012), and would have a predicted 27% difference in rates between SH & PL (Dam & Peterson, 1988). As such, there is not a clear reason to predict that would lead directly to a varying biomass difference in winter, particularly given that the minimum temperature is attained in late February and remains low when biomass accumulates in Spring. However, it is notable that the plankton community at PL - which is warmer - is composed of smaller cells year-round than at SH, on average. This appears to align with the temperature-size rule, and so, bottom-up temperature effects may be driven indirectly through the fact that smaller cells - which have comparatively high maximal rates, and high affinity for light (Table 1.1.4) have their maximal growth rate moderated by temperature. Thus, the difference in winter phytoplankton biomass between SH & PL may be driven by the fact that small cells dominate the community, and that these are bottom-up limited by temperature, since light is expected to be less limiting for these cells under the tradeoff for low-light photosynthesis (H1ii). *A priori*, this hypotheses appears to have weaker support than H1i, but is presented in line with the ethos of the 'strong inference' framework.

2.4.2.2 Seasonality of mean phytoplankton size

The observed seasonal cycle of phytoplankton mean size aligns with existing anecdotal reports. At SH, previous work describes that the winter is dominated by smaller cells and the summer by larger cells (González-Gil *et al.*, 2022), while at PL biomass contribution plots by functional group demonstrated a similar theme (McEvoy *et al.*, 2023). We confirm that the mean phytoplankton size is lowest in winter and largest in summer, and shows steady progression between these extrema through spring and autumn. Furthermore, it was unsurprising to observe that the microplankton community was composed of broadly similar biomass contributions along the

size domain at each of the three UK coastal sites, based on existing knowledge of taxonomic abundances (Bresnan *et al.*, 2009, 2015; Widdicombe *et al.*, 2010).

A bottom-up driver of increasing phytoplankton mean size from winter to summer may be difficult to support when considering conventional beliefs about phytoplankton cell biology. Most modelling works focus on nutrients as the primary bottom-up driver of phytoplankton cell size, and in some cases this is reflected by a decrease in mean size from winter to summer (Simpson & Sharples, 2012; Acevedo-Trejos *et al.*, 2015). This follows primarily from the increasing half-saturation constant of nutrient uptake with cell size (Eppley *et al.*, 1969; Litchman *et al.*, 2007), so that the observed diminishing concentration of nutrients through spring, to depletion in summer, would suggest a favouring of small cells. Despite this, light may be an important factor in controlling phytoplankton cell size. Many previous modelling works assume an independence of limitation on photosynthesis and size (e.g., Ward *et al.*, 2012; Ward & Follows, 2016; Acevedo-Trejos *et al.*, 2015). However, this appears to contrast both theory and observations, which show that the initial slope of photosynthesis decreases with size as a result of intracellular shading (Edwards *et al.*, 2015). Due to the saturating nature of the P-I curve, the greatest difference in photosynthetic rate between small and large cells (per unit of biomass) is predicted to be largest when light is low, in winter. As such, we hypothesise that the phytoplankton mean size is limited by light in winter (H2i).

An alternative explanation is that the phytoplankton mean size is top-down driven, since the zooplankton mean size is near its annual maximum at all sites, due to the (relative) absence of microzooplankton. It follows that the grazing on large phytoplankton cells would be comparatively higher than at other times of year, given the tradeoff for optimal predator-prey size ratio (Table 1.1.4). As such, we hypothesise that phytoplankton mean size in winter is small due to the large mean size of zooplankton organisms, which would graze optimally on large phytoplankton cells and thus limit their biomass (H2ii).

However, H2i or H2ii would explain the increase in mean size in summer, indicating that the mechanisms driving size structure may change over this period. In the absence of any limitation from nutrient or light, maximum nutrient uptake (per unit of biomass) is log-linearly decreasing (Marañón *et al.*, 2013; Litchman *et al.*, 2007), and maximum photosynthetic rate is also decreasing after 6 μm (Marañón *et al.*, 2013; Irwin *et al.*, 2006; Ward & Follows, 2016; Acevedo-Trejos *et al.*, 2022; Chen & Liu, 2010). Therefore in the absence of any predation, small cells $\leq 6\,\mu$m would be predicted to exclude large cells year-round. Given that observations show an increasing contribution of biomass from winter to summer, and that there is no theory to support a stronger bottom-up limitation on the small cells than the large cells, then a top-down control must have some influence on the phytoplankton size structure.

An important consideration in the structure of the phytoplankton community is the varying top-down pressure being exerted by the micro- and mesozooplankton. Small microzooplankton have the fastest maximal grazing rates (Hansen *et al.*, 1997), allowing fast growth in response to available prey, and, as a result, small phytoplankton are predicted to be closely controlled by their predators (Banse, 1994). Indeed, this is already believed as the reason that nanophytoplankton

exhibit comparatively little seasonality at PL compared to larger cells (Holligan & Harbour, 1977; Stelfox-Widdicombe *et al.*, 2004). Mesozooplankton, on the other hand, typically have lower maximal grazing rates (Hansen *et al.*, 1997). Let us state that the grazing rate of a zooplankton size class j can be defined at a general level as

$$G_j^{\text{tot}} = G_j^{\max} \cdot L_F(F_j) \cdot Z_j \tag{2.9}$$

where $L_F(F_j)$ is the saturation of the grazing response (e.g., Holling II, Holling III, Ivlev, etc.). Then, for a given prey availability F_j, the total grazing is dependent on Z_j and $G_j^{\max}$. Yet, the cycle of observed zooplankton mean size at all sites is at its minimum in summer, meaning that Z_j is larger for smaller size classes, and comparatively smaller for larger size classes. Given that the prey field appears to favour larger zooplankton, since phytoplankton mean size approaches its maximum, but that Z_j is smaller for large zooplankton, it follows that the tradeoff $G_j^{\max}$ − which is decreasing with size − must have an important effect on structuring the phytoplankton community. However, since $G_j^{\max}$ directly affects Z_j (and vice versa), the mechanism that drives Z_j to be comparatively small for larger size classes must also be understood.

Therefore, we hypothesise that, through spring to summer, the large phytoplankton can exploit 'unfulfilled' growth which the small phytoplankton are unable to attain, due to the tight coupling with their predators (H4i[1]). This hypothesis aligns with the fact the modelled phytoplankton mean size approaches its maximum while the zooplankton mean size approaches its minimum. A further consideration is that incorporating a lag time and detailed analysis of biomass distributions would provide insight to the relationship between phytoplankton and their grazers using a similar strategy to (García-Oliva *et al.*, 2022), since prey availability controls growth rate but not necessarily instantaneous biomass.

An interesting pattern observed at SH & LE is that the microplankton mean size also increases from winter to summer. Furthermore, omission of microzooplankton from the microplankton cohort retains statistical significance with qualitatively similar features at SH & LE (minimum mean size in winter, maximum in later summer). This indicates that some change is observed within the large phytoplankton. A potential driver for this is the increasing contribution of dinoflagellate biomass through the summer, which moves away from the diatom domination of biomass in the spring. Since diatoms typically have the highest maximal growth rates at any given size in the phytoplankton (Irwin *et al.*, 2006), it is possible that the increase in 'phytoplankton' size arises from a change in strategy from pure autotrophy to constitutive mixotrophy. This would be supported by the fact that pico- / small nanoplankton increase in overall biomass from spring to summer, providing a potential resource for growth of larger plankton, despite the fact nutrients are depleted in summer. Thus, we hypothesise that mixotrophy accounts for the increase in mean size of photosynthetic plankton from spring to summer (H4ii).

[1]H3 is described in Section 2.4.2.3, as we decided to list the hypotheses in chronological order.

2.4.2.3 Seasonality of phytoplankton size diversity

Phytoplankton log-ESD variance, used here as a proxy for phytoplankton size diversity, also shows a strong seasonal pattern at SH & PL. This is similar between both sites, both quantitatively and qualitatively, with minima being fitted first in winter and second in spring. The winter minimum is likely due to the proportional absence of large phytoplankton at PL; however, the result is slightly surprising at SH, on the basis of the biomass distributions exhibited in Figure 2.11. The fact that time-averaged biomass distributions can show apparent high diversity through a multimodal distribution, yet have low diversity on a sample-to sample basis, indicates that the observed 'diversity' in the time-averaged plot is from between-sample variability as opposed to within sample diversity. This is likely to arise from the small subsample volume used for phytoplankton measured by microscopy at SH (& LE), and is indicated by the proportionally vast difference between the median and mean biomass density around $20-30\,\mu$m at Scottish sites (Figure 2.3). Observations along much the large-cell subset of the size domain ($> 6\,\mu$m), when biomass is low in winter, are dominated by zeros such that picoplankton / small nanoplankton would dominate in most instances (resulting in a similar biomass distribution seen at PL in Figure 2.11). Then, on the rare occasion that a large cell is recorded, its biomass may be comparatively large ($\sim 1\,\mu$g C L^{-1} for a $25\,\mu$m cell at the minimum recordable abundance). The end result, over many years, may be that the time-averaged mean biomass density is true to the typical value. However, because GAMs were fitted to log-ESD variance with equal weights on a sample-by-sample basis, the reported size variance is likely to be small as a result of the fact that many of the samples at SH did not record, or underestimated abundance of large phytoplankton cells (following the predictions of Goldman, 1993). It should also be noted that, since subsample volume is fixed at SH, the probability of 'missing' cells is likely to diminish at SH when abundance increases (i.e., from spring to early autumn), and so this issue likely affects winter to a greater degree than summer (investigated in Chapter 3). It is therefore possible that, in reality, the winter community at SH is diverse, driven by contributions across the pico-, nano- and microphytoplankton (as suggested in Figure 2.12). However, in the absence of large-subsample data, this hypothesis is difficult to robustly test.

In spring, the phytoplankton community typically exhibits the only other observed minimum in size diversity at SH & PL. This coincides with the sharp increase in biomass and indicates that the dominance of a small number of similar cell sizes is associated with the sudden biomass accumulation. This aligns with taxonomic perspectives of spring blooms at UK coastal sites (Bresnan *et al.*, 2009; Napoléon *et al.*, 2014), although it should be noted that a species succession may often occur (e.g., from diatoms to *Phaeocystis* at SH; Bresnan *et al.*, 2009). This pattern is also supported by modelling works which predict a small number of species should dominate in a coastal Spring bloom (Huisman *et al.*, 1999b), although the model of Acevedo-Trejos *et al.* (2015) fitted in the temperate open ocean predicts high size diversity during the Spring bloom. Notably, the dominance of cells $\sim 6\,\mu$m at PL aligns with the peak of the allometric maximal photosynthetic rate curve (Marañón *et al.*, 2013; Irwin *et al.*, 2006; Ward & Follows, 2016; Acevedo-Trejos *et al.*, 2022), supporting the hypothesis that the bloom of small nanoplankton

cells at PL is driven by the optimised maximal growth rate, and results in the near-competitive exclusion of other cell sizes (H3). At SH, the fitted phytoplankton mean size is larger than at PL during the period that size diversity is at its minimum, indicating that the bloom is constituted by larger cells, on average, but exhibits an equal degree of competitive exclusion. This is an interesting feature − with no obvious bottom-up explanation. Perhaps the top-down pressure on small phytoplankton is already strong by April; Figure 2.12 indicates that microzooplankton do constitute an important proportion of the microplankton biomass, with a greater biomass density at $\sim 20\,\mu$m than at PL.

After this point, size diversity increases until early autumn (albeit with an intermediate small trough in mid-summer at SH, likely due to the strong dinoflagellate blooms at $\sim 30\mu$m). Interestingly, mean size is decreasing through this period, and so too is biomass by autumn. This indicates that the change in the size structure is likely an expansion 'leftwards' of the biomass distribution, with increasing proportional contributions by picoplankton. A potential driver for this is due to the replenishment of nutrients. Small cells have the fastest uptake of nutrients under saturating conditions per unit of biomass (Litchman *et al.*, 2007; Marañón *et al.*, 2013) and, as such, may outcompete larger cells for the uptake of nutrients, which were previously deplete (H5). While light is decreasing at its fasted rate over September, it is still at a similar level to March, during which time the mean size is able to increase. It therefore appears unlikely that the tradeoff for low-light photosynthesis is driving the community to a smaller mean size during this period.

2.4.2.4 Zooplankton size structure

The main takeaway from models fitted to zooplankton size structure is that the mean size follows a similar cycle across UK coastal sites. In general, the mean size is larger in winter and smaller in early summer, driven by increasing proportional biomass of microzooplankton, and coinciding with the fitted maximum phytoplankton biomass. At all sites, the transition in spring is relatively consistent (albeit the onset is slightly delayed at PL), which would appear most likely driven by the observed increase in prey availability. The fact that zooplankton mean size is decreasing through spring aligns with allometric theory that dictates microzooplankton should respond quickest to the increase in prey availability (in the absence of other limitation on growth / enhanced losses; Hansen *et al.*, 1997). In conjunction with this, the optimal predator size based on mean phytoplankton log-ESD at the end of spring is 4.4 $\ln(\mu\text{m})$ at PL and 5.5 $\ln(\mu\text{m})$ at SH (based on Banas, 2011 fit of Hansen *et al.*, 1994 data). At SH in particular, the mean zooplankton size (5.25 $\ln(\mu\text{m})$) is relatively similar to the prey-dependent optimal zooplankton size. At PL, there is a larger difference between optimal and fitted mean size (5.75 $\ln(\mu\text{m})$), although it should be noted much of the small nanoplankton biomass present during this period is already present in winter, and in the absence of any substantial microzooplankton biomass. However, it is possible that the mesozooplankton graze on small nanophytoplankton in winter at PL, rather than microzooplankton, as observed in similar regions (Vargas & González, 2004; Motwani & Gorokhova, 2013). Furthermore, the fitted relationship between optimal prey size

and predator size is undoubtedly subject to some degree of random variability at both the species and organism levels. As such, its use should be interpreted tentatively, particular in the absence of any confidence interval estimate.

Based on these factors, there are two possible non-mutually exclusive hypotheses: zooplankton mean size decreases in spring because of the allometric maximum grazing rate relationship (H6i); or zooplankton mean size decreases because the small zooplankton − particularly dinoflagellates, which graze on comparatively larger cells (Hansen *et al.*, 1994) − are most optimally sized to respond to the increasing prey availability (H6ii). During the period of decreasing zooplankton mean size, the size diversity also increases to its fitted maximum at all UK coastal sites, indicating that the mesozooplankton still play an important role in the community through spring to summer.

Zooplankton mean size increases thereafter, perhaps in response to the increasing phytoplankton mean size and microzooplankton. It is possible that the microzooplankton and nanophytoplankton reach a steady state due to their tight coupling, or that mesozooplankton directly control the microzooplankton. Under H4ii, the unexploited phytoplankton growth potential which is fulfilled by larger phytoplankton would provide prey for mesozooplankton (28 μm phytoplankton optimally grazed by 833 μm zooplankton; Hansen *et al.*, 1994; Banas, 2011).

2.4.2.5 Difference in copepod mean size between PL & Scottish sites

The standout difference in copepod size structure is that mean size increases from winter to early summer at PL, whereas it is broadly decreasing at SH & LE (at SH, there is a pulse in mean size in mid spring, which is likely driven by the influx of *Calanus finmarchicus*; Heath *et al.*, 1999). The difference in the seasonal cycle may relate to the differing diet preference, which varies across sites (Figure 2.19), and likely varies in conjunction with the different prey availability at each site (in particular the much higher biomass of small cells at PL). Herbivores dominate at PL year-round, and particularly in winter; concurrently, there remains substantial phytoplankton prey at PL, unlike at SH. It is therefore possible that the winter copepod mean size is small to facilitate grazing on small phytoplankton. However, despite the fact that some small copepods are known to feed on pico- and small nanophytoplankton prey (Vargas & González, 2004; Motwani & Gorokhova, 2013), it should be noted be noted that this marks a large deviation from the mean optimal prey size ratio (Hansen *et al.*, 1994; Banas, 2011). The copepod community at PL in winter, with fitted mean size of 6 ln(μm), have an average optimal prey size of 2.9 ln(μm) which suggests they would likely feed suboptimally in a prey pool dominated by small cells with mean size 1.5 ln(μm). Yet, the fitted microplankton mean size is $\sim$ 3.5 ln(μm) year-round at PL, so there does appear to be some circumstantial support for small copepods grazing on nanophytoplankton, given that the copepods would be assumed to align with their optimal prey size if they were only feeding on microplankton and in the absence of a significant age structure or top down control. At PL, where copepods are primarily herbivorous, we hypothesise that the winter copepod community is sustained by nanophytoplankton, and that increasing biomass of large nano- and microplankton sustains an increase in mean copepod size. At SH & LE, where

omnivory and herbivory have a varying share which is approximately equal averaged across the year, we hypothesise that the shift from herbivores to omnivores drives the decrease in mean size from mid-spring to mid-summer. A supporting observation is that the decrease in mean size at SH & LE coincides with the increasing proportion of juvenile $C1-3$ and $C4-5$ copepods in each site through spring, yet the fitted minimum mean size occurring in mid-summer actually coincides with a maximum in the proportion of C6 copepods ($\sim 60\%$ at both sites). Evidently, if changes in mean size were purely a function of age structure, this observation would not be made, and so the mean size cycle must relate at least partially to diet and / or taxonomy.

2.4.3 Size structure and productivity

The observed positive relationship between phytoplankton biomass and mean size is in agreement with previous findings in the literature (Irigoien *et al.*, 2004; Hatton *et al.*, 2015; González-García *et al.*, 2023; Cloern, 2018; it should be noted that the findings in Irigoien *et al.*, 2004 do not overlap with the presented data). The slope of this relationship is greater at SH than at PL, despite the fact that the phytoplankton biomass covers a wider range of values at SH. This indicates that the size structure of phytoplankton communities at SH & PL diverges as the biomass increases. Surprisingly, although the low biomass state occurring in winter exhibits big differences in the observed biomass between SH & PL, the mean size is more similar then than during the summer (when biomasses are similar). However, in light of concerns related to the subsample size at SH, it should be noted that the steepness of the slope may be influenced by undersampling of low-abundance large cells (Goldman, 1993).

The phytoplankton size diversity displays a unimodal relationship with log-biomass. While there are no known observational studies assessing this relationship, a unimodal relationship is predicted in modelling work (Banas, 2011). It is also worth noting that Irigoien *et al.* (2004) and model outputs of Goebel *et al.* (2013) show a similar relationship between biomass and taxonomic diversity (although a positive linear relationship is observed in García-Comas *et al.* (2016) and Acevedo-Trejos *et al.* (2018)). Our findings suggest that diversity is maximised at intermediate levels of biomass. This aligns with H4, whereby high levels of biomass during blooms are driven by allometric growth rate selection. The fitted maximum size diversity occurs at different values of biomass at SH & PL. However, it scales similarly across the nondimensionalised range of biomass at each site (if dividing by log-biomass range for each site). This suggests that size diversity behaves similarly at SH & PL once the differences in each site's range of biomasses are accounted for. However, concerns about the subsample volume for the portion of the phytoplankton community analysed by microscopy at SH mean that the inference is tentative at this site, since small subsample are known to underestimate taxonomic diversity (Rodríguez-Ramos *et al.*, 2014).

The relationship between biomass and size diversity for zooplankton is perhaps less coherent. If it is assumed that microzooplankton always constitute a proportion of the zooplankton community − even if very small − then a negative relationship between zooplankton biomass and size diversity is observed at SH. However, if all points are included (including those where no

microzooplankton are recorded) the observed pattern is unimodal. Because the subsample volume is small for measuring microzooplankton at SH, it is difficult to know whether the lack of recorded microzooplankton is because they are truly absent (or in low abundance), or whether they have just been missed by the sampling. Thus, inference can only be tentatively made about size diversity when biomass is low.

Zooplankton mean size decreases with biomass across both sites (albeit it after an initial plateau at PL), most likely driven by increasing microzooplankton contributions, and perhaps further by an increased proportion of juvenile copepodite copepods or changes in copepod taxonomic structure.

The zooplankton-to-phytoplankton biomass ratio decreases with community plankton biomass, in agreement with Gasol *et al.* (1997), Irigoien *et al.* (2004), and Hatton *et al.* (2015). At both SH & PL this is indicated by a saturating relationship across (almost all) of the range of observed biomass values. The fitted curve at PL attains a negative gradient at $\sim 200\,\mu$g C L^{-1} phytoplankton biomass, but this is not supported by a sufficient number of observations and so inference is not made on this domain (particularly since the shape of a polynomial linear model is weighted across the entire domain of the exploratory variable, as opposed to a piecewise modelling approach like a GAM). However, it should be noted that the small number of observations do align with the model fit. The fact that proportional rate of zooplankton biomass accumulation decreases as the phytoplankton biomass increases aligns with theory. Under both Type-II and Type-III responses, zooplankton prey ingestion saturates as the prey biomass increases. An alternative explanation may be that the food web itself becomes more complex and therefore less efficient when phytoplankton biomass is large due to microzooplankton grazing activity and subsequent grazing on microzooplankton by mesozooplankton (Nejstgaard *et al.*, 1997; Olson *et al.*, 2006; Schnetzer & Caron, 2005). This is supported by the elevated proportion of omnivorous copepods in summer (when fitted seasonal biomass attains its maximum) compared to spring, at all sites.

A positive linear relationship between phytoplankton and zooplankton size variance was observed. This is analogous to the patterns in taxonomic diversity oh phytoplankton and zooplankton observed in Irigoien *et al.* (2004). More widely in ecology, relationships between taxonomic diversity of predator and prey may be positive, including between mammals (Sandom *et al.*, 2013). However, this appears to directly contrast the findings of García-Comas *et al.* (2016) who observe a negative relationship. In this work, a positive association between trophic transfer efficiency (TTE; for which Z:P is a proxy) and predator size diversity is observed, and a negative association between prey size diversity and TTE is observed. These results are supported by ecological theory: dietary niche partitioning would suggest that a diverse set of predators will deal with a monotypic set of prey efficiently, but a monotypic set of predators will graze inefficiently on a diverse set of prey, since some of the prey will not be optimally sized (Ye *et al.*, 2013). However, importantly, microzooplankton and carnivorous mesozooplankton are omitted from zooplankton in the work of García-Comas *et al.* (2016). Although exclusion of carnivorous mesozooplankton is necessary to surmise a relationship between zooplankton predators and phytoplankton prey,

the contribution of microzooplankton to the zooplankton community is evidently important in UK coastal waters (Holligan & Harbour, 1977; Stelfox-Widdicombe *et al.*, 2004). Ultimately, microzooplankton have a complex contribution to the ecosystem since they act as predators of phytoplankton and prey of mesozooplankton. It is therefore understandable that they must be excluded from the zooplankton category. Since the reported results here cover several trophic levels, with a potential complex transfer of phytoplankton → microzooplankton → herbivorous mesozooplankton → carnivorous mesozooplankton, comparison is intractable. However, the fact that the inclusion of microzooplankton and carnivorous mesozooplankton yields a positive relationship between phyto- and zooplankton size diversity provides motivation to investigate if complex food web interactions could have a positive association with size diversity in future research.

2.4.4 Long-term variability

Broadly speaking, PL exhibits a greater degree of multi-year variability in plankton size structure than Scottish sites. This is seen most clearly in models fitted to mean log-ESD across 5-year periods, which exhibit a notable increase in both unicellular and multicellular mean sizes in $2015-2019$ compared to the overall $2000-2019$ period. In particular, the unicellular community exhibits a smaller degree variability between $2000-2014$, and then a shift to consistently larger cells in $2015-2019$. Existing work focused on taxonomy were published in 2010 and, as such, further study is required to reveal the shift in taxonomy which is responsible for the change in mean size. However, it is notable that significant interannual variability has been observed during the Spring bloom at PL, in terms of timing, intensity, and taxonomic composition (Widdicombe *et al.*, 2010). However, the reported taxonomic variations are often confined to the genus of *Chaetoceros* (with the exception of the *Skeletonema costatum* bloom in 2005). It would appear that a more significant shift has occurred during the $2015-2019$ period, and that this is not confined to the spring bloom given the year-round elevated mean cell size.

At SH, there is also some evidence of an increasing mean cell size in the unicellular community in $2015-2019$, albeit less strong that at PL. Again, existing research was published in 2009 & 2015 and so future research may be motivated by this apparent change in size structure. Unlike at PL, the observed difference is largely during the high-biomass period between March and September, with little interannual variability observed in the autumn and winter period. The fact that the divergence over the $2015-2019$ period occurs during the bloom period is interesting. In Bresnan *et al.* (2009), which studied the $1997-2006$ time period, the bloom dynamics at SH appeared to change around 2002, with a shift from *Chaetoceros spp.* dominated blooms to *Skeletonema spp.* dominated blooms, with *Thalassiosira spp.* making important bloom contributions throughout the time series. It is possible that a regime shift occurred in the phytoplankton community during this period at both SH & PL. However, since the presented results incorporate unicellular plankton (inclusive of microzooplankton), the resultant shift at both SH & PL may also be due to food web dynamics.

Less biomass is observed in multicellular plankton at LE in the $2015-2017$ period, in line with

decreasing abundances (Wells *et al.*, 2022). Although no clear change in multicellular plankton size structure or its seasonality is observed, the August−September size diversity approaches a local minimum, and is notably lower than previous years. The mean size is also slightly lower in this period. Interestingly, little difference appears to occur in the phytoplankton community. The reduced biomass of mesozooplankton is believed to be related to a reduction in salinity (Wells *et al.*, 2022). Further investigation into interannual variability and long term trends of mesozooplankton size structure should be considered for future research.

2.4.5 Limitations and potential sources of bias

2.4.5.1 Seasonal amalgamation of biomass distributions

In recognition of the fact that the observed distributions of biomass distributions were often complex / multimodal, and that contributions by biomass varied by functional / taxonomic group, it became impossible to discuss the distributional properties of the biomass distribution solely through summary statistics. As such, it was necessary to plot seasonally-averaged biomass distributions at each site. However, significant challenges were associated with averaging these data across a long time period. A truncated mean approach was chosen for constructing amalgamated seasonal biomass distributions, due to the tendency of the median to underestimate biomass density (particularly when biomass was approaching the minimum recordable abundance), and the excessive leverage permitted by the mean for extreme observations. Truncated means have been previously used when estimating diatom frustule size (Finkel *et al.*, 2005) indicating the appropriate use in this context, as well as other applications such as measuring CO_2 flux during marine heatwaves (Mignot *et al.*, 2022). However, two issues are associated with the truncated mean method: (i) it produces a statistically biased estimate of the mean when the distribution is asymmetric, and (ii) a subjective element is introduced in the choice of cut-off point. As such, the development of an improved method which is robust across different sites, and in particular different ecosystems where the distribution of biomass contributions may be more or less skewed, is desirable in future research.

2.4.5.2 Overdispersion and heteroscedasticity in fitted models

Each presented LM and GAM assumes a normal distribution of errors with constant variance. However, there are some cases where the assumption of constant variance is violated (heteroscedasticity), or errors are overdispersed. This limits the inference that can be made about the coefficient estimates, since the standard errors of the coefficient estimates and subsequent p-values are inaccurate.

An extreme case of overdispersion and non-constant variance in the errors is the GAM fitted to PAR. Inherently, PAR has a non-constant variance structure since its minimum is zero yet its maximum is constrained by astronomy. The fitted mean and error variance therefore exhibit an extremely strong relationship. To take the two most extreme cases, the variance of PAR in the middle of the night in winter is negligible (if not zero), yet the PAR at midday in midsummer exhibits very large variance (driven by variable atmospheric conditions).

The result of overdispersion is that the standard error estimates of model parameters, and subsequently the associated p-values are unreliable (White, 1980). As such, the parameter estimates / model fit should be considered a descriptive statistic rather than an inferential statistic in the absence of a robust variance estimate (Freedman, 2006). However, in the instance of PAR, a p-value was not deemed necessary to infer that light exhibits strong daily and seasonal variability. Importantly, the model predictions are still unbiased. Therefore, such a brazen approach to neglect biased standard errors and p-values is only appropriate where it is beyond any reasonable doubt that the effect is significant (e.g. variation of light across the day). This approach was not applied to models such as those comparing biomass and mean size / size variance, since the effective null hypothesis was that no relationship existed. There were cases where slight overdispersion was observed in models, meaning that test statistics and subsequent p-value were likely to be slightly inflated and deflated respectively. However, in each of these instances, the reported p-value was $< 2.2 \times 10^{-16}$ and the overdispersion was minimal. On this basis, we concluded that it was highly unlikely that final inference was affected by these issues.

Furthermore, regardless of any overdispersed / non-constant variance of the error structure in the fitted model, we required that the residuals had an approximately symmetric distribution across the range of fitted values. There were two reasons for this: (i) to ensure linearity in the response, and (ii) to minimise potential over-influence by outliers. Inadequate shape-fitting generally only applies to LMs, which have predetermined shape governed by the explanatory variables. In some cases, higher order polynomial terms were included despite the lack of statistical significance to meet this criterion. The issue of linearity was not an issue for GAMs because, due to their inherent flexibility, they are robust to nonlinearity. In some cases, a small degree of positive or negative skew was observed in the residuals, but was not deemed sufficient to have a significant effect on the fitted mean of the model.

Nonetheless, it should still be noted that the models with heteroscedasticity could be improved, particularly if being used for predictive purposes. In the case of PAR, a logarithmic or reciprocal power transformation of the response may seem initially attractive to stabilise to the mean-variance relationship. However, the data are constituted by a large proportion of zeros; particularly in winter when days are short. As such, implementation of this transformation would need to be incorporated as part of a hurdle model, which first determines the probability of observing a non-zero value, before subsequently modelling. However, even after separating the zero from non-zero values, the existing symmetric distribution of residuals would likely be impeded by this transformation. Perhaps a more elegant solution would be to separately model the variance of the residuals. For the applications discussed here, this would appear to add an unnecessary degree of complexity.

2.4.5.3 Ordinary least squares vs. standardised major axis regression

When fitting linear models to biomass, mean size, and size variance, ordinary least squares (OLS) were used. This neglects the variability in the explanatory variables which, in this instance, are themselves subject to a large degree of random variability (being observations of a dynamic

ecosystem, from small samples, and sampled from a typically spatially heterogeneous domain). Standardised major axis (SMA) regression, which accounts for the variability in both the response and the explanatory variables, could be a viable alternative. However, the tendency of SMA regression to over-estimate the slope coefficients in a linear model, in particular when compared to OLS regression, increases the chance of false positive inference. This is reflected in the work of Edwards *et al.* (2015) who experimented with SMA regression when exploring the allometry of photosynthetic performance, and found much larger slope estimates for SMA compared to OLS (approximately 3× greater). As such, OLS regression was favoured for a conservative approach to inference. Furthermore, no known feasible method for implemeting SMA-style regression to GAMs was found in the literature. Given that the timing of biomass gain / loss (e.g., the spring bloom) may be considered a random variable itself (as shown by González-Gil *et al.*, 2022), the models should not be considered appropriate for predicting the timing of individual events such as the spring bloom. This is particularly relevant at UK coastal sites where blooms exhibit strong interannual variability (Bresnan *et al.*, 2009; Widdicombe *et al.*, 2010)

2.4.5.4 Phytoplankton microscopy subsampling volume

We believe that there is potential for significant bias in the analysis of unicellular plankton by microscopy at SH & LE. The unicellular microscopy data (e.g., auto- / mixo- / heterotrophic plankton cells approx. $> 6\,\mu$m) has varying subsample size by site. At PL, the size of the subsample is 50 mL at most, although this drops to < 20 mL to remain within practical constraints when abundance is high. At SH & LE, the subsample is 1 mL year-round. This means that the minium recordable abundance (MRA) at PL is 20 cells L^{-1} and at SH & LE is 1000 cells L^{-1}.

At all sites, cell counts are made along one or more transects of the subsample chamber, with the aim of mitigating any within-subsample bias. These are referred to as fields of view (FOV) analyses. At SH & LE, two alternative datasets are also available: presence / absence (P/A) data based on a larger (50 mL) subsample, and first 400 cell counts (F400) with no fixed subsample size. The F400 has the desirable property that it will always have the same resolution when recording taxonomic diversity, since each measured cell contributes 1/400 of the community abundance. This is in contrast to the FOV data, where the winter resolution is likely to be lower, since the abundances are extrapolated from a smaller number of counts, almost acting as a binary operator at times when abundance are low ($<$ 1000 cells L^{-1}). However, F400 would appear likely to miss important aspects of biodiversity when one species is dominant, and in particular in a size structure context. This potential problem arises because the abundance generally decreases linearly with size on the log-log scale, and so a bloom of small cells may mask the contribution of a larger species, despite the fact that the larger species contributes far greater biomass than the small cell on a per-cell basis. Conversely, if equal biomass was contributed by a slightly larger species with lesser abundance, the large cells would not be missed. Therefore, if biomass is the measure of quantity, as is the case here, then the F400 method is likely to be biased due to abundance-size and the cellular size-carbon relationships. In an ideal case, a fixed subsample of carbon may be measured, but this is practically impossible under the existing

framework.

A potential mitigating factor for this proposed bias is that, geometrically, in a chamber with two cells of distinct size, the larger cell is more likely to be detected than the small cell. Per cell, larger cells would take up more 2D space in the settled chamber than the small cell. Under this logic, shape would also be a factor, as the 2D silhouette varies for a given ESD (e.g., an oblate spheroid would be more likely to be detected than a prolate spheroid). A detailed experimental approach would be necessary to establish the probability of different cells being recorded, within a given subsample.

Although the FOV method appears to be more robust for mitigating bias within the subsample based on the presented theory, there are concerns on whether a small subsample itself can accurately and consistently capture the community. Consider the two following cases, defined as equally size-diverse by our definition (noting that a 28 μm cell has approximately 42$\times$ the carbon mass as the 6 μm cell; Marañón *et al.*, 2013):

1. 6 μm cells @10,000 cells L^{-1} and 28 μm cells @ 2,000 cells L^{-1}
2. 6 μm cells @2,000 cells L^{-1} and 28 μm cells @ 400 cells L^{-1}

At SH & LE, where the MRA is 1,000 cell L^{-1}, it is evident that the measurement of the larger cell is likely to be highly erratic in Case 2, with individual subsamples only able to under- or overestimate abundance and biomass (as 0, or $1000, 2000, 3000, ...$ cells L^{-1}). Conversely, we anticipate that Case 1 will present a more stable report of the community, since abundances of both cell sizes are above the minimum recordable abundance. At PL, no issues would be expected since both abundances are well above the MRA threshold of 20 cells L^{-1}. Since a clear seasonal pattern is acknowledged for abundance (Bresnan *et al.*, 2009, 2015) and biomass (González-Gil *et al.*, 2022) at SH, it is possible that the extent of the sampling inconsistency would be greater when abundance is low − in winter. In this instance, mean size and size diversity would potentially be under-estimated, since large cells were missed proportionally more often than small cells. These would follow from the fact that large diatoms are undersampled using conventional methods (Goldman, 1993), and furthermore, that taxonomic diversity is too (Rodríguez-Ramos *et al.*, 2014). Despite this, no previous work has directly investigated the effect of low-volume subsampling on observed size structure. On the basis of the available data, we hypothesise that:

- Large cells are omitted more often than small cells in low-volume phytoplankton subsamples (H7i), and the effect is more prevalent in winter since overall abundances are lower (H7ii).
- A reduced subsample volume results in a decrease in observed phytoplankton biomass (H8i), mean size (H8ii) and size diversity (H8iii).

Although entirely circumstantial, it is interesting to observe the difference in microplankton size structure seasonality at PL when compared to SH & LE. In the first instance − as would be expected through a larger subsample − observed phytoplankton size diversity is consistently higher at PL than Scottish sites. So too is mean size, although the mean size increases drastically

with biomass. Given that this aligns with the above hypotheses, we make tentative inference on the observed size structure of the microplankton.

2.4.5.5 Intraspecific size diversity

An inherent issue with most plankton data is the lack of species measurements on a sample-by-sample basis. Phytoplankton species are known to exhibit intraspecific variance in size and related traits (Rillo *et al.*, 2020; Orizar & Lewandowska, 2022; Hamer *et al.*, 2022; Hillebrand *et al.*, 2022b). Additionally, variations in size may relate to reproduction (Speijer *et al.*, 2015; Montresor *et al.*, 2016) or in response to grazing (Branco *et al.*, 2020). Although we used the method of Schartau *et al.* (2010) to impose some intraspecific size variance around the taxonomic mean cell size, this does not account for the seasonally variable reproductive (Kim *et al.*, 2023; Nagai *et al.*, 2004; Kalinina *et al.*, 2023) and defensive behaviours exhibited by phytoplankton (Friedrichs *et al.*, 2013; Long *et al.*, 2007). Furthermore, choosing one size means consistency across the dataset; but, for common species, the lack of a verified measurement at SH & LE means that systematic bias could affect the results (note that species cell sizes are explicitly measured for a small number of cells at PL).

It is also possible to get the raw flow cytometry files to get a better indication of size, but we just used fixed sizes following McEvoy *et al.* (2023). This reflects the focus on the overall community size structure of this work. Extracting individual cell sizes form flow cytometry was not feasible in the given time period. For a detailed analysis of the pico- and small nanoplankton this improvement could be made in future work.

2.4.5.6 Missing data

One consistent theme throughout this chapter has been the importance of the pico- and small nanoplankton. At PL, a huge biomass contribution is made year-round by these plankton, the majority of which recorded by flow cytometer and the remainder by autotrophic nanoflagellates. At SH, flow cytometry is used and this has a significant influence on the perceived size structure. However, auto- and heterotrophic nanoflagellates are not reliably enumerated despite their prevalence (Bresnan *et al.*, 2015) and, as such, the small plankton are likely underestimated at SH. The same is true at LE, where flow cytometry is also not employed, meaning that the pico / small nanophytoplankton cannot be enumerated. Given then differences observed between PL and SH in terms of flow cytometry contribution, it would be very interesting to compare the small plankton community at SH and LE. Many characteristics of the plankton community are surprisingly similar at SH & LE despite clear differences in oceanography (offshore site vs. enclosed sealoch) and bloom dynamics (Bresnan *et al.*, 2015). To establish if this is true across the whole phytoplankton community, flow cytometry would be invaluable.

2.4.6 Summary of hypotheses

H1. Phytoplankton biomass is lower at SH than PL in winter because:

i. Differences in light arising from latitude and attenuation limit growth

ii. Biomass accumulation, which is largely governed by small cells in this period, is limited by temperature

H2. Phytoplankton mean size is limited in winter by:

i. Intracellular shading, affecting low-light photosynethtic performance of large cells

ii. Top-down pressure from zooplankton, which are near their annual maximum mean size

H3. The bloom of small nanoplankton cells at PL is driven by the optimised maximal growth rate since neither nutrients or light are limiting, and results in the competitive exclusion of other cell sizes

H4. Through spring to summer, the phytoplankton mean cell size increases because:

i. Large phytoplankton can exploit 'unfulfilled' growth which the small phytoplankton are unable to attain, due to the tight coupling between small phytoplankton their predators resulting in focused high intensity grazing

ii. Mixotrophy accounts for some of the increase in mean size within large nano- and microphytoplankton from spring to summer at SH & LE

H5. The increase in phytoplankton size diversity in late summer / early autumn, associated with a decrease in mean size, is due to the allometric tradeoff for maximal nutrient uptake, since the nutrients replenish with the deepening ML during this period.

H6. Zooplankton mean size decreases in spring because, one, or either of:

i. the allometric maximum grazing rate relationship

ii. small zooplankton are most optimally sized to respond to the prey availability

H7. Lower volume subsamples affect the monitoring accuracy of large cells by:

i. omitting them more often than small cells in low-volume phytoplankton subsamples

ii. showing seasonal variability in identification rates, with the lowest values in winter

H8. A reduced subsample volume results in a decrease in observed phytoplankton:

i. biomass

ii. mean size

iii. size variance

2.5 Conclusion

Plankton community size structure experiences a strong seasonal cycle in UK coastal waters. A key feature of the seasonal cycle is the consistent increase in phytoplankton mean size from winter to summer. It is hypothesised that light limits cell size in winter due to intracellular shading, and that grazing pressure also promotes increasing mean cell size after the initial spring bloom, or mixotrophy by photosynthetic plankton. Then, phytoplankton size diversity maximises in early autumn, which is hypothesised to result from the replenishment of nutrients allowing picoplankton to contribute a larger proportion of biomass. The size diversity minimum during the spring bloom is hypothesised to be driven by competitive exclusion relating to allometric maximal growth rate trade-offs. Zooplankton size structure is strongly affected by microzooplankton accumulation, confirming their importance at these sites. Interestingly, mean size of zooplankton broadly decreases when phytoplankton mean size increasing, perhaps due to the larger maximal grazing rates of microzooplankton. Although some differences are observed between SH, LE, and PL, many of the qualitative features are similar between sites. However, the main difference between PL and SH is the large background community of small nanophytoplankton which exist at PL year-round; the prevalence of these cells is more seasonal at SH. From a methodological perspective, are some concerns that the phytoplankton analysed by microscopy at Scottish sites are affected by bias resulting from low subsample volume. The hypotheses generated here guide the remaining work undertaken in this thesis.

Chapter 3

Impact of subsample volume on estimating phytoplankton community size structure

3.1 Introduction

3.1.1 Enumerating phytoplankton by inverted microscopy

Traditional sampling methods are generally only able to provide a glimpse of a phytoplankton community, since the volume of water used in subsamples is tiny compared to the oceanic region inhabited by the community. Bottle sampling and analysis by inverted microscopy remains a popular method for phyto- / microzooplankton enumeration, likely due to its comparative lack of specialist equipment. The most common method of inverted microscopy is that of Utermöhl (1931), where cells in a small water subsample are left for $24-48$ hours to settle on the base of a sedimentation chamber. These cells are counted by an inverted microscope, and typically identified taxonomically. The result is a set of observations of taxa with corresponding abundances computed by extrapolating the cell counts from the sample to units of cells L^{-1}.

The main problem facing taxonomists when enumerating a plankton sample is the time associated with identifying and counting cells. Typical subsample volumes vary, since overall sampling effort is a product of sampling volume and the number of cells to be counted. In some cases, such as at SH & LE, a fixed subsample volume (1 mL) is used year-round to provide consistency. Conversely, at PL, a large subsample volume is used when possible (50 mL), but this drops to 25 mL during times of high abundance, or lower during vast blooms. Examples can be found elsewhere of varying subsample volumes: $2-50$ mL in Cloern (2018), $2-5$ mL in García-Comas *et al.* (2016), 3.25 mL in Jacobs *et al.* (2024a), and $50-100$ mL in González-García *et al.* (2023). Evidently, a balance must be found such that enough cells are counted to provide an accurate description of the community, whilst not being too resource intensive so as to be an unsustainable or impossible task.

3.1.2 Consequence for measuring taxonomic diversity

Because enumeration is effort intensive, an established problem is the ability of the inverted microscopy technique to accurately estimate taxonomic diversity. Even incremental increases in subsample volume beyond 10 L identify taxa that were not present in the previous subsample (Cermeño *et al.*, 2014). However, richness returns are diminishing with increasing subsample volume: estimated species richness typically reaches 90% of its asymptotic value after $0.25 - 1$ L of subsample is analysed (Cermeño *et al.*, 2014). Similar results are also demonstrated in paired observational and simulated work of Rodríguez-Ramos *et al.* (2014).

What is not clear is the consequence of the this volume-dependent estimator bias in a functional ecology context. Beyond 10 L, for example, a phytoplankton cell will be so rare that it would have a negligible impact on proportionally-weighted community indicators like mean cell size or size variance. Conversely, a cell identified in a 2 mL water sample which was not identified in a 1 mL water sample may significantly affect the parameter estimation. Ultimately, the theoretical probability of successfully sampling a randomly distributed[1] taxon is increasing with both the taxon's population abundance and the sampling volume.

3.1.3 Statistical basis for large-cell undersampling

The theoretical result that sampling probability is a function of population abundance and subsampling volume is problematic in the study of size structure. Abundance (A) is acknowledged as size-dependent at a general level, governed by the equation $A = cV^b$ for a cell of volume V, where c is a scaling constant and $b = -0.75$ (Cermeño *et al.*, 2006; Cermeño & Figueiras, 2008). Values $b \leq -1$ may be observed in oligotrophic regions (Cavender-Bares *et al.*, 2001) making large cells proportionally even more rare. Such a result means that, for a given subsample volume, large cells are less likely to be identified. This issue is raised previously (Goldman, 1993; Rodríguez-Ramos *et al.*, 2014), but has not yet been directly addressed.

To display this result at a quantitative level, let us assume cells are spatially randomly distributed[1] as a Poisson point process. For population specific abundance A_i, expected number of counts per subsample of volume V is given by $\lambda_i = A_i V$. Then, the probability of counting at least one cell is given by $P(C_i \geq 1) = 1 - P(C_i = 0)$ with probabilities governed by the Poisson probability mass function. Associated probabilities for $V = 1$ mL and $V = 50$ mL can be found in Table 3.1 for abundances between 10 and 10,000 cells L^{-1}. Evidently, as abundance decreases, the probability of successful identification also decreases.

The most relevant results arise when abundances exist below the minimum recordable abundance ($A_{\min}$), i.e. when $\lambda_i < 1$. Take, for example, the taxon in Table 3.1 with $A_i = 300$ cells L^{-1} which is subject to a sampling regime with $V = 1$ mL. Its probability of enumeration ($P(C \geq 1) \approx 1/4$) indicates that, on average, 3 subsamples will omit it and only one will identify its presence. In the instance that it is identified as present, then number of counts C_i have associated probabilities

[1] A discussion of spatial randomness, patchiness, and clustering within the sedimentation chamber is provided in Section 3.4.4.2.

Table 3.1: Probabilities of non-zero counts in subsamples as a result of varying population abundance, under the assumption that counts are a Poisson point process.

	$V =$ 1mL		$V =$ 50mL	
Abundance (A_i; cells L^{-1})	λ_i	P($C_i \geq 1$)	λ_i	P($C_i \geq 1$)
10	0.01	0.01	0.5	0.39
30	0.03	0.03	1.5	0.78
100	0.1	0.09	5	0.99
300	0.3	0.26	15	1.00
1000	1	0.63	50	1.00
3000	3	0.95	150	1.00
10000	10	1.00	500	1.00

$P(C_i = 1|C_i \geq 1) = 0.86$ and $P(C_i > 1|C_i \geq 1) = 0.14$. A typical set of observations over 4 subsamples may therefore be $\hat{A}_i = \{0, 0, 0, 1000\}$ cells L^{-1}, none of which are actually close to $A_i = 300\,\text{cells}\,L^{-1}$. The estimator may not be asymptotically biased in a classical sense (over the long run, $E[\hat{A}_i] \to A_i$), yet it would not be appropriate to deal with individual samples since any $\hat{A}_i$ cannot possibly reflect A_i, given that $A_{\min} = 1000$ cells L^{-1} and $A_i = 300$ cells L^{-1}. This is problematic in the context of long-term monitoring since data are likely to be zero-dominated with sporadic observations of extreme values.

Furthermore, while $A_{\min}$ is constant, the associated minimum recordable biomass (MRB; $B_{\min}$) is a function of size. Indeed, $B_{\min}(s) = A_{\min}m(s)$ where $m(s)$ is the cellular carbon biomass (μg C L^{-1}) which is monotonically increasing for a cell of size s (e.g. using Menden-Deuer & Lessard, 2000). This means that larger cells must contribute increasing carbon to be reliably observable. Given that biomass may be considered a more informative measure of community size, productivity, and subsequent ecosystem functioning, $A_{\min}$ must be sufficiently low (by having sufficiently large V) to minimise any bias.

Dealing with potential bias at or below $A_{\min}$ is difficult. A natural approach would be to deal with plankton abundance estimates using a similar method to those employed for handling concentration data, such as nitrate, at the limit of detection (LOD). Common examples include dividing the LOD by 2 and, while basic methods such as these have their own issues, more robust statistical methods exist (Wood *et al.*, 2011). However, there is a fundamental difference in how nutrient data are measured compared to plankton counts. The signal measured in inverted microscopy is fundamentally discrete: the raw data, cell counts, are non-negative integers. Conversely, nitrate is measured as a continuous signal. For example, at PL, a segmented flow analyser is used to measure nitrate concentration via colorimetric spectrophotometry. The sample reacts with chemical reagents to form a coloured compound, and the instrument measures its absorbance to determine nitrate concentration. As a result, LODs are comparatively extremely low (0.02 μM at PL; Woodward & Harris, 2023), and observations below the LOD are often identifiable. Since the observations below the LOD are not identifiable for discrete counts, sampling bias associated with low-volume subsampling when estimating plankton size structure must be established so that limitations on inference are well understood, and a specific method

to handle bias can be developed. Such results would extend the work of Cermeño *et al.* (2014) and Rodríguez-Ramos *et al.* (2014) from a taxonomic diversity to functional diversity context.

3.1.4 Aims and objectives

The aims and objectives set here are a direct result of the hypotheses generated in Sections 2.4.6. Firstly, we use *in situ* data from SH & LE to test the hypothesis that large cells are omitted more often than small cells in low-volume phytoplankton subsamples (H7i), and that this is more prevalent in winter due to the overall lower abundances (H7ii). Then, a simulation-based approach is implemented to test the hypothesis that a reduced subsample volume results in a decrease in observed phytoplankton biomass (H8i), mean size (H8ii) and size diversity (H8iii).

3.2 Methods

3.2.1 Investigating rates of successful identification across the size domain

3.2.1.1 Data

Two datasets were used at SH & LE: the Fields of View (FOV) data, where cell counts are made by species from a 1 mL subsample and extrapolated to cells L^{-1}, and the presence / absence (P/A) of each species recorded from a 50 mL subsample. Both datasets were retrieved as in Chapter 2 (Marine Scotland Science, 2018a,b). Abundances of phyto- / microzooplankton recorded at PL by microscopy are also presented for comparison to observations at SH & LE (Widdicombe & Harbour, 2021); in this dataset, 50 mL subsamples were primarily enumerated, although this dropped to $\leq$ 25 mL during period of high abundance. Here, cells measured by fluorescence microscopy were omitted to allow appropriate comparison with SH & LE.

To test for the presence of a small-cell bias, only the P/A and FOV data at SCObs sites were used. These observations are paired since they are analysed subsamples from the same sample, and so it is appropriate to compare the observations on a sample-by-sample basis. To facilitate this analysis, the FOV and P/A were joined by date and taxon using the `tidyverse` package in R (Wickham *et al.*, 2019). Three cases arose for each observed taxon by date: present in both P/A and FOV datasets, present in P/A but not FOV, or present in FOV but not P/A. Sizes were allocated on a taxonomic basis. A list of sizes can be found in Table A.2.

3.2.1.2 Exploratory analysis

Firstly, we investigated the distribution of specific biomass contributions across the size domain. These biomass contributions were compared to the MRB for both 1 mL and 50 mL subsamples to demonstrate that, while small cells may be permitted to contribute small biomasses to the community in a low-volume subsample, large cells are not. At each site, diatoms and non-diatoms were separated to demonstrate the potential impact of the MRB across different plankton groups (it should be noted that the differences between diatom size-specific carbon and other protists are vast, but other groups within protists exhibit smaller differences; Menden-Deuer & Lessard,

2000).

The boxplot of cell sizes were also compared across the FOV and P/A datasets (irrespective of abundance / biomass), and means of log-ESD (not weighted by abundance or biomass, but rater by occurrence) between the two datasets were formally compared using a one-sided two-sample t-test (with an alternative hypothesis that the mean observed log-ESD is larger in the P/A data than the FOV data).

3.2.1.3 Logistic regression

To estimate the size-dependent probability of enumeration, and test the hypothesis that the probability of omission increases with increasing size (H7i), we fitted logistic regression models. The event of 'successful enumeration', E, was defined as a cell being successfully enumerated in the FOV data, given its presence identified in the P/A dataset (or only the FOV dataset itself). E was binary with value equal to 1 when a taxon was present in the FOV dataset, and 0 when the taxon was present in the P/A data but not the FOV dataset. Then, the model was defined as:

$$\ln\left(\frac{P(E_i|s_i)}{1-P(E_i|s_i)}\right) = \beta_0 + \beta_1 s_i \tag{3.1}$$

where s_i is the log-ESD of an observed cell i. The first models fitted at SH & LE were time-independent, such that $i = 1, .., n$ denoted every taxonomic observation in the P/A dataset. The model was then transformed via the inverse-logit link so that the size-dependent probability of successful enumeration was given by:

$$P(E_i|s_i) = \frac{1}{1+\exp\left(\beta_0 + \beta_1 s_i\right)} \tag{3.2}$$

The model was then expanded to include seasonal interaction effects to test if the extent of large-cell omission was greater in winter (H7ii). As such, a factor variable was introduced for meteorological season (M). The full interactive size-by-season dependent probability of successful enumeration was given by:

$$P(E_i|s_i, m_i) = \frac{1}{1+\exp\left(\beta_0 + \alpha_{m_i} + \beta_1 s_i + \gamma_{m_i} s_i\right)} \tag{3.3}$$

Since M had four levels, ANOVA was used to test if the full interactive model had significant seasonality of slope when compared to an intercept-only seasonal model with a common slope term (i.e., $\gamma_{m_i} = 0$ in Equation 3.3). Grouping of seasons was also considered in cases where the collective seasonal slopes were not significantly different to zero, but the t-test indicated some significant differences between seasons. In this instance seasons were reduced to spring-summer and autumn-winter, since cellular abundances recorded in FOV were most similar in these time periods. The grouped model allowed was fitted at two levels. Firstly, with grouped intercept and slope terms, such that:

$$P(E_i|s_i, \tilde{m}_i) = \frac{1}{1+\exp\left(\beta_0 + \alpha_{\tilde{m}_i} + \beta_1 s_i + \gamma_{\tilde{m}_i} s_i\right)} \tag{3.4}$$

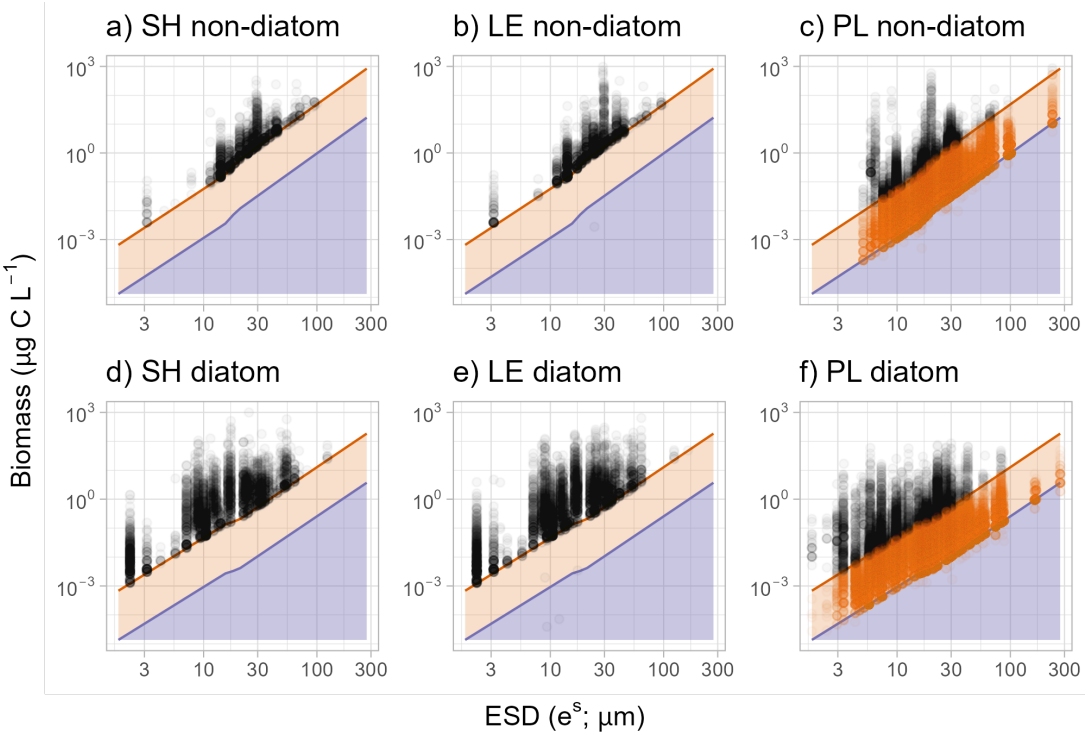


Figure 3.1: Biomass contributions of observed phytoplankton and microzooplankton taxa at SH and LE from the FOV dataset, and PL. Orange line bounding the shaded area shows the minimum recordable biomass corresponding to 1 mL subsamples. Purple line bounding the lower shaded area gives the corresponding values from a 50 mL subsample.

with $\tilde{m}_i$ indicating spring-summer or autumn-winter. This model was then compared to the model with four levels for the intercept, but grouped slope:

$$P(E_i|s_i, m_i, \tilde{m}_i) = \frac{1}{1 + \exp\left(\beta_0 + \alpha_{m_i} + \beta_1 s_i + \gamma_{\tilde{m}_i} s_i\right)} \tag{3.5}$$

Inference on differences between seasons was then drawn based on the t-tests of model coefficients from the final selected model from the ANOVA process.

Since a large number of observations were available at both SH & LE, all models were fitted separately across the two sites. Additionally, since a comparison between the effect at SH & LE did not motivate the work, there was no reason to fit a more complex model with an interactive term for site. The models were each assessed to check for linearity in both the explanatory variables and logit link. Furthermore, although prediction was not the aim of the model (rather, estimation of the slope), discriminatory power was computed as area under the receiver operating curve to better understand the model behaviour and other potential uses or limitations of the model.

3.2.2 Simulating the impact of low-volume subsampling at the community level

3.2.2.1 Bootstrapping experiment

Abundances of phyto- / microzooplankton recorded at PL by microscopy were used to conduct the bootstrapping experiment (Widdicombe & Harbour, 2021). Nanoflagellates, measured by epifluorescence microscopy, were removed from the dataset so that only cells measured by inverted microscopy were included. Coccolithophores, which are analysed in a different subsample, were also removed. Size data can be found in Table A.1.

A simulated 1 L water tank was constructed in R using the cellular abundances reported at PL for each of the $n = 1183$ size structure observations. These were used as the parent empirical distribution throughout the bootstrapping process; herein, we refer to these as the populations; estimates of size structure from bootstrapping experiments are referred to as simulations.

The use of the PL observational data in this bootstrapping experiment is justified by the large subsample size (50 mL). However, it is important to note that the community at PL may itself be subject to biases related to the sampling regime, particularly during high-biomass and bloom periods when the cells are enumerated from a subsample $\leq$ 25 mL. As such, the aim is not to predict how a change in subsample volume would affect PL specifically. Rather, we aim to identify the potential bias at a more general level, and across a wide range of observed empirical distributions of cell size.

Subsamples of volume $V = \{1, 3, 10, 30, 100\}$ mL were drawn from the simulated 1 L water tank. Log-spacing of V was based on the asymptotic species richness demonstrated on the log-scale in Cermeño *et al.* (2014) and Rodríguez-Ramos *et al.* (2014). For each sample date and subsample volume, 30 iterations of the bootstrapped sampling regime were drawn using the following method:

1. Read in raw phyto- / microzooplankton microscopy data for a given sample date at PL
2. To convert from abundances (cells L^{-1}) to counts (cells in the water tank; coincidentally 1 L), round abundances up / down as a Bernoulli trial with probability associated with the post-decimal value.
3. Sample from the simulated community, without replacement, in proportion to the subsample volume V
4. Compute kernel density estimates following the method described in Section 2.2.2
5. Summarise the community biomass, mean log-ESD, and log-ESD variance as described in Section 2.2.3.2

The method is inspired by the work of Rodríguez-Ramos *et al.* (2014), with two key differences in the simulation regime: firstly, empirical data is used for sampling, rather than a four-parameter neutral plankton model; secondly, the cells are assumed to follow spatial randomness. This is in an attempt to marginalise the effect of cell size (broadly in inverse proportion to abundance) in

the sampling regime. Patchiness in the sea (Mackas *et al.*, 1985; Davis *et al.*, 2004; McGillicuddy & Franks, 2019; Barth & Stone, 2024) and clustering in the counting chamber (Sandgren & Robinson, 1984; Haberyan & Haddock, 2013; Jacobs *et al.*, 2024a) affect enumeration, and are indeed proven directly to affect richness estimation (Rodríguez-Ramos *et al.*, 2014).

3.2.2.2 Measuring bias

For each iteration of the bootstrapping simulation, estimates of biomass, mean log-ESD, and log-ESD variance were computed. These were then compared to respective population values through the bias function, defined as:

$$\text{Bias}[\hat{\theta}] = \text{E}[\hat{\theta}] - \theta \tag{3.6}$$

for each population parameter $\theta = \{\ln(b), \text{E}[s], \ln(\text{Var}[s])\}$ and its corresponding estimator from bootstrapped samples denoted by the hat notation. A one-sided t-test, with alternative hypothesis that mean bias is less than zero, was executed for each combination of θ and V. To ensure independence, mean bias was calculated on a day-by-day, volume-by-volume basis. Mean bias is therefore computed from 30 iterations of the bootstrapping experiment in each instance.

3.2.2.3 Statistical modelling

Biases were modelled against population parameters to identify the domain on which the issue typically arose. Firstly, all indicators were modelled against population biomass (i.e., models of the form $\text{Bias}[\hat{\theta}] \sim \ln(b)$); then, against their own respective population values ($\text{Bias}[\hat{\theta}] \sim \theta$, noting that the $\theta = \ln(b)$ case is duplicated from above); and, finally, against day of year to identify seasonal trends in bias ($\text{Bias}[\hat{\theta}] \sim t$, for day of year t). Generalized additive models (GAMs) were originally fitted to the simulated data, however, the residual errors suffered from severe overdispersion and heteroscedasticity. In fitted GAMs, the error variance exhibited a complex structure and a viable link function could not be found. The heteroscedasticity was a direct result of the changing realisations of sampling bias from one size structure to the next (even at equal value for the associated explanatory variable) and, as such, required direct modelling.

To combat the heteroscedasticity and fully understand the variance structure, generalized additive models for location, scale and shape (GAMLSS) were fitted. GAMLSS models extend the concept of a GAM to not only model the expected value of the response (location), but also its scale and shape as a function of explanatory variables. Here, we modelled the location and scale of the response, and assumed shape to be constant. Hence, for any response variable (Y), models for location (μ_Y) and scale (σ_Y) were a function of some set of explanatory variables. These explanatory variables do not necessarily have to be the same for each μ_Y and σ_Y. The default identity link for location and log-link for scale (which enforces necessary positivity) was used.

Each presented model had equations for μ_Y and σ_Y composed of an explanatory variable of interest (denoted generally as X), and, for scale only, a random intercept (1183 levels for each

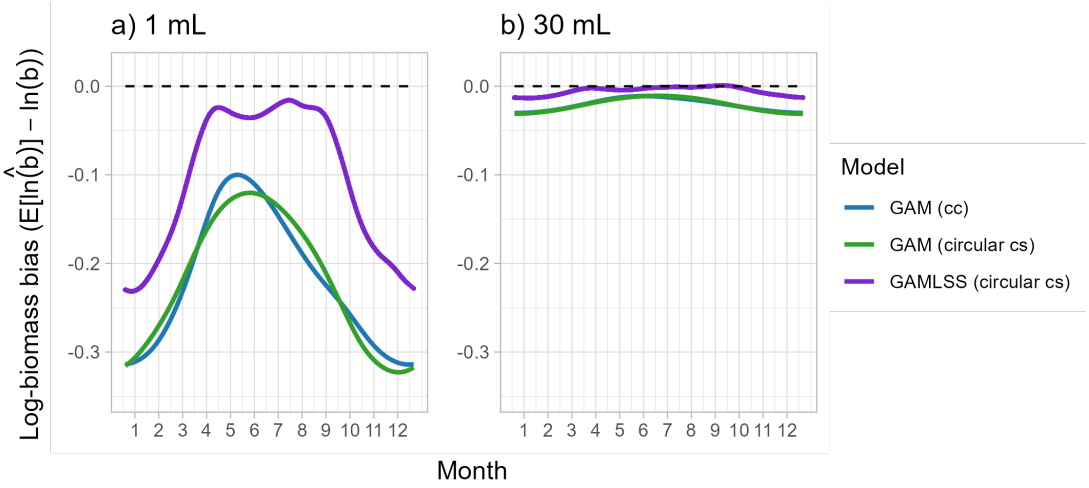


Figure 3.2: A comparison of candidate GAM and GAMLSS models assessed using simulation outputs from two subsample volumes. All models have a seasonal smooth (cyclic cubic splines or cubic splines on circular coordinates) fitted to location (expected value), while the GAMLSS model also has a seasonal term and a random intercept for the scale (variance) model.

observed size structure; B). The general framework for the model was therefore:

$$\mu_Y \sim f(X); \qquad \sigma_Y \sim f(X) + B$$

for some smooth function f. A model with random intercept for location was also fitted for $Y = \text{Bias}[\hat{\theta}]$ and $X = \ln(b)$ in an attempt to decompose the 'source' of the bias. This model is plotted and fully described in Figure D.1 and Appendix D.1, but is not included in the main text due to concerns about its robustness.

To assess the assumptions of each model for location, Q-Q and residual plots were visually inspected (full details in Table B.3). To assess the contribution of model terms to the model fit, p-values were computed for individual terms and the Cox & Snell pseudo-R^2 (R^2_{CS}) was compared across intercept-only and full μ_Y models, always with saturated σ_Y models. Furthermore, to understand the contribution of each variable to the σ_Y model, the random effect was removed from the σ model (with an associated intercept-only model for μ_Y). Inference on contribution to likelihood was drawn based on comparison across R^2_{CS}.

Fitting was executed using the `gamlss` package in R (Rigby & Stasinopoulos, 2005). A drawback of this package is that standard errors associated with smooths are liable to underestimation because the software treats the smooths as fixed effects as opposed to random effects. The `mgcv` package with `gam(..., family=gaulss())` would be preferred for this reason, but incurred prohibitive computational expense due to the number of levels of the random intercept in the σ_Y models. While the function `bam()` of the `mgcv` package is often used to combat computational expense in the fitting of GAM models, it cannot be used for GAMLSS models.

When investigating seasonality, a change of coordinates was required since cyclic cubic regression splines are not supported in `gamlss`. Here, a circular coordinate system was used (i.e., polar

coordinates with fixed radius of 1). For day of year t, the set of coordinates (u, v) is therefore given by:

$$u = \cos\left(\frac{2\pi t}{365}\right) \qquad v = \sin\left(\frac{2\pi t}{365}\right)$$

so that models are of the form $\text{Bias}[\hat{\theta}] \sim f(u) + f(v)$. Tensor product smooths are not supported by `gamlss` and so seasonal fitting was only possible on univariate cubic regression splines. To ensure validity, the model fit was assessed to check for any linearity in the predictors arising from unaccounted interaction between the polar coordinates, but there was never an issue. When GAMs with cyclic cubic regression splines (models of the form $\text{Bias}[\hat{\theta}] \sim t$) were fitted, residuals exhibited skewness which was likely driven by clustering of points along the $\text{Bias}[\hat{\theta}] = 0$ line. Left unaccounted for, this would result in a biased fit of $\text{Bias}[\hat{\theta}]$ itself. Thus, a supporting comparison of fit is therefore provided in Figure 3.2 for three cases: GAM with cyclic cubic regression splines (cc), GAM with cubic regressions splines (cs) for two univariate smooths on circular coordinates, and a GAMLSS with cubic regressions splines (cs) for two univariate smooths on circular coordinates. The latter two models have the same model for location, but the GAMLSS additionally models scale using smooths on circular coordinates as well as a random intercept. The GAMLSS produces a much more conservative estimate of bias, likely a result of its ability to cope with the complex variance structure[2]. Furthermore, although the standard error estimation in `gamlss` is suboptimal, clearly, the importance of mean estimation accuracy in the context of confidence interval fitting is, at the very least, equally as important. All plotting was undertaken in `ggplot` (Wickham *et al.*, 2019) with `ggpubr` (Kassambara, 2016).

3.3 Results

3.3.1 Relationship between enumeration rate and size at SH & LE

3.3.1.1 Biomass contributions

The MRB dictates the lowest biomass contribution a taxon can make, and increases linearly with cell size on the log-log scale (indicated in Figure 3.1). As a result, the MRB for a non-diatom protist of 30 μm is 1.54 μg C L^{-1} at SH & LE (where a 1 mL subsample is used), yet is only 0.03 μg C L^{-1} for 50 mL subsamples at PL. As such, measurement of non-diatom taxa at this size are unreliable during periods of low biomass, since observations are likely to be dominated by zeros even when a proportionally significant biomass contribution is made by cells in the population. Given that the fitted microplankton community biomass is 1.32 μg C L^{-1} at SH & LE in winter (and 3.4 μg C L^{-1} at PL; Section 2.3.3.3), this initial exploratory analysis would suggest that if large plankton are present in the water column at Scottish sites during this

[2]Note for Figure 3.2. In some cases stark qualitative differences arose between GAM and GAMLSS fits. We hypothesise this was driven by heteroscedasticity. For example, in Figure D.3, $\text{Bias}[\ln(\hat{b})]$ is decreasing for large b when $V = 1$. Conversely, for GAMLSS fits, it is asymptotically zero. We hypothesise that this was driven by differences in variance structure across communities (i.e., random effects for scale). The belief is that if two points, with equal $\ln(b)$ but different size structure are used to fit the model, then the plankton assemblage which exhibits less variability with respect to sampling bias has enhanced leverage on the fitted model, since likelihood is maximised in that neighbourhood. This then drives model fit in an undesirable manner.

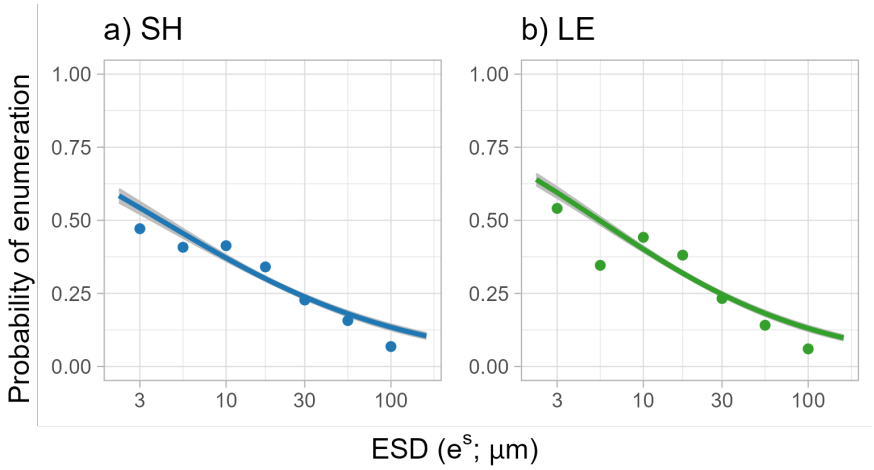


Figure 3.3: Rate of enumeration of taxa at SH and LE from a 1 mL subsample, conditional on its presence in a 50 mL subsample or the 1 mL subsample itself, across phytoplankton cell size.

period, it is unlikely that they are reliably detectable. Conversely, the MRB at PL is small when compared to the observed community biomass and, as such, cells existing below the MRB are unlikely to contribute significantly to biomass. For example, in winter at PL, a 30 μm protist taxon would contribute $< 1\%$ of community biomass on average if it was at the MRB.

A second observation is that many of the points at PL exist in the orange region of Figure 3.1, corresponding to observations which are not reliably made by a 1 mL subsample, but reliably recorded by a 50 mL subsample. Indeed, 42,713 of 50,831 observations (84.3%) exist in this region where abundances are between 20 cells L^{-1} and 1000 cells L^{-1}; if only winter observations are considered, this rises to 95.0%. These observations contribute around 43% of biomass overall (81.8% in winter), and likely influence the estimation of mean log-ESD and log-ESD variance. It should be noted that the subsample volume is susceptible to decrease in summer (albeit still much larger than 1 mL), and so the winter summary statistics should be considered more robust than the overall ones which are likely affected by reduced subsample volumes during high-abundance periods.

Since the probabilities associated with 'successful' identification vary by subsample size for a fixed abundance (Table 3.1), it is not appropriate to simply remove the observations in the orange region and present these as a prediction of how observations at PL would be affected by a change in subsample volume. What it does show, however, is that low-abundance cells have an important contribution to the community at PL, and that the site benefits from a large subsample volume. Furthermore, the large subsample appears to be particularly valuable in winter, since $> 80\%$ of biomass is contributed by cells with abundances below 1000 cells L^{-1}.

3.3.1.2 Enumeration rate

We found strong evidence to support the hypothesis that large cells are omitted more often than small cells in low-volume phyto- / microzooplankton subsample (H7i). Through the fitting of logistic regression with successful enumeration against log-ESD (Equation 3.2), there was

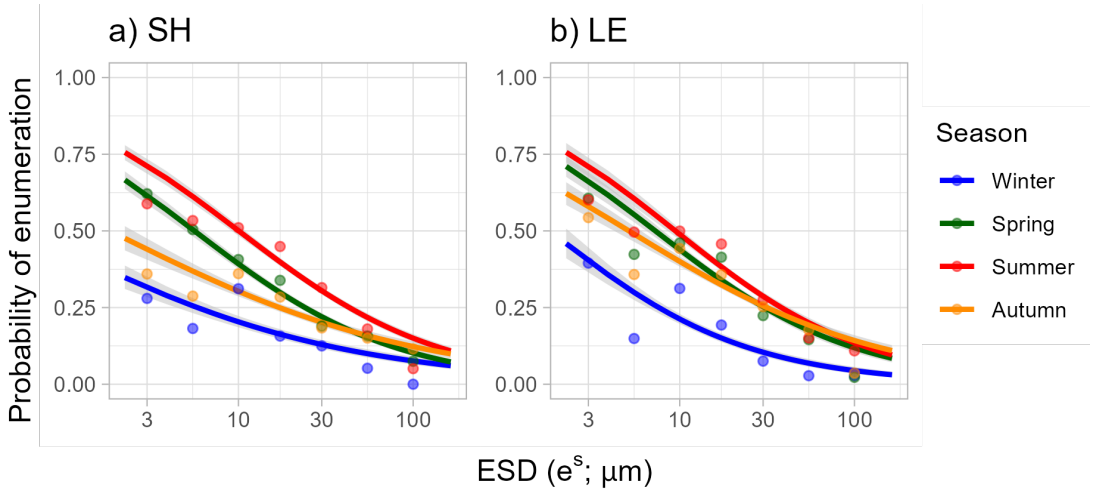


Figure 3.4: Seasonal variations in rate of enumeration of taxa at SH and LE from a 1 mL subsample, conditional on its presence in a 50 mL subsample or the 1 mL subsample itself, across phytoplankton cell size.

significant evidence to reject the null hypothesis that the probability of enumeration was invariant across the size domain ($p < 2 \times 10^{-16}$), with the fitted slope monotonically decreasing on the whole domain at both sites (2.2 μm to 164 μm; Figure 3.3). The probability of enumeration (both fitted, and observed through binned proportion) is approximately 0.25 at 30 μm, compared to the rate of 0.4 fitted and observed at 10 μm. This indicates a vast decline in enumeration rate across the central 50% of the observed taxa by size (Figure 3.5). By 100 μm this has halved further, such that it becomes very unlikely a cell of this size is observed in a 1 mL subsample if it is present in the population.

Models were then fitted to establish the potentially varying seasonality of the relationship between probability of enumeration and cell size (Figure 3.4). Models selected at SH & LE differed slightly as a result of the model selection process. At LE, a full interactive model (Equation 3.3) was fitted, such that both the intercept and slope varied by season. At SH, slopes of winter and autumn were not significantly different; neither were the slopes of spring and summer. The final selected model had a grouped slope for autumn-winter and spring-summer, and a different intercept for each season (Equation 3.5). The seasonal model at LE and the grouped-slope seasonal model at SH both added significant value compared to the season-independent model through comparative ANOVA testing ($p < 2 \times 10^{-16}$).

Despite the small difference in the form of the final fitted models, the inference drawn from each seasonal model with respect to H7ii is similar across sites: in winter, the probability of enumeration is lower than in summer, and this is true across the size domain. Broadly speaking, the enumeration rate was highest in summer (when mean observed biomass is highest in the FOV dataset), followed by spring, then autumn, and finally winter. This aligns with the seasonal succession of median community biomass. The fact that a seasonal difference in enumeration rate is observed indicates that some seasonality may influence the extent of any potential measurement bias in the FOV dataset. However, it could be argued that, since the fitted seasonal curves are

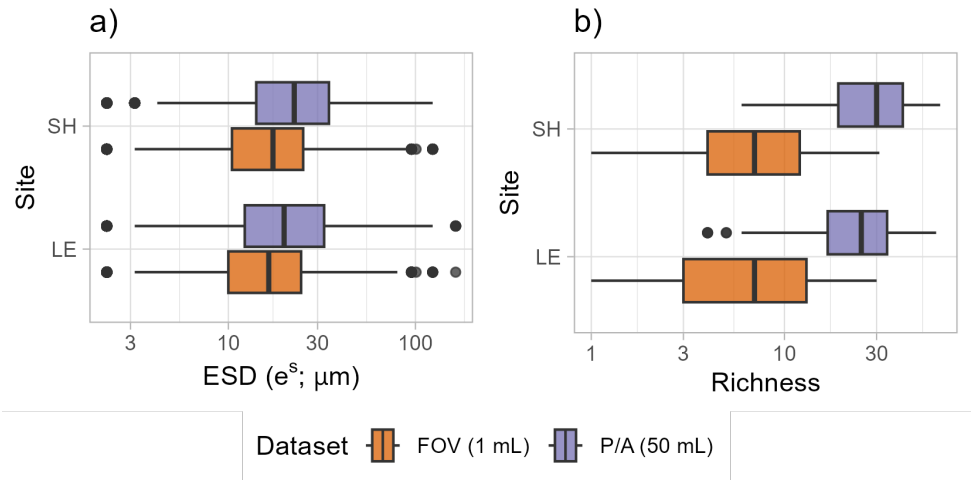


Figure 3.5: Frequency distributions of (a) taxa by size across FOV and P/A datasets at SH and LE, and (b) associated richness computed on each sample date.

converging with increasing size, the extent of the size-dependence of the bias may be decreasing by season (as a result of decreasing small-cell enumeration rate).

The model checks of linearity in both the predictors and the response were satisfied through visual inspection. Binned proportions were also computed (displayed in Figure 3.3 & 3.4) to provide a check of fit. The area beneath the ROC was similar at SH & LE, each with values of 0.65 respectively for the baseline models, and 0.67 for both seasonal models.

3.3.1.3 Frequency of taxa by cell size

With clear evidence that the probability of enumeration is decreasing along the size domain, we analysed the frequency distribution of sizes across each of the datasets. All frequency distributions of log-transformed ESD were approximately symmetric and, unsurprisingly, the median taxon size was higher in the P/A dataset than the FOV dataset at both sites (22.5 μm vs 17.3 μm at SH; 19.9 μm vs 16.4 μm at LE). This corroborated by a one-sided two-sample t-test of mean log-ESD at both sites (each with $p < 2 \times 10^{-16}$), such that we conclude the average taxon in the P/A data is larger than in the FOV data. However, without specific abundance data in the P/A dataset, we cannot compare the biomass-weighted mean log-ESD.

Richness was also computed on a sample-by-sample basis for comparison. Again, richness followed an approximate normal distribution when log-transformed. The mean log-richness in the P/A data was significantly higher than in FOV at both sites ($p < 2 \times 10^{-16}$), with corresponding median values for richness of 7 at both SH & LE in the FOV dataset, and 30 & 25 at SH & LE respectively in the P/A dataset. However, the limitations of richness are clear in a size structure context; taxa which have little effect on the community may still increase the perceived diversity when richness is used, since there is no consideration of evenness. As such, it is difficult to make inference on the differences in distribution of community level indicators in the FOV dataset when compared to the P/A dataset.

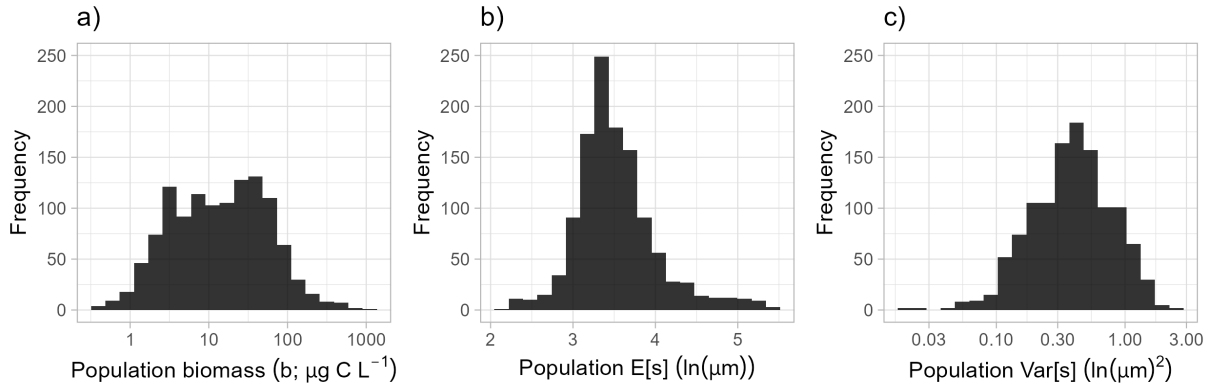


Figure 3.6: Distribution of phytoplankton 'population' biomass, mean log-ESD, and log-ESD variance at PL.

3.3.2 Community estimator bias from simulation experiments

The findings in Section 3.3.1 demonstrate that, at the species level, low-volume subsamples under represent all taxa. Furthermore, this affects large cells to a greater degree than small cells. However, it is not clear whether these species level biases propagate to the community level. In this section, a simulation experiment is used to assess and quantify potential bias in the estimaiton of phytoplankton biomass, mean size, and size diversity.

3.3.2.1 Overview of model performance

After undertaking the simulation experiment using observations at PL, GAMLSS models were fitted to explore the relationships between bias and various explanatory variables. The vast majority of 'population' biomasses were on the scale $1-100$ μg C L^{-1}; mean log-ESD is primarily recorded on $2.8-4.5$ ln(μm); and log-ESD variance on $0.1-1.5$ ln(μm)^2 (Figure 3.6). Results were therefore primarily interpreted on this domain given the lack of a reliable standard error estimation for smooths in `gamlss`. In any case, it can be noted that 42 of 43 models for location achieved reported significance for the explanatory smooth ($p < 0.001$; Table 3.2). The p-values associated with the scale model were also highly significant in all models ($p < 2 \times 10^{-16}$).

Estimates of R^2_{CS} established a common pattern: estimation of the response scale always con-

Table 3.2: Decomposition of R^2_{CS} for selected GAMLSS models for the case $V = 0.001$. Here, the model for scale (σ_Y) is always saturated, and the model for location (μ_Y) is incrementally expanded from intercept-only to intercept and smooth. In the intercept-only model for location, all pseudo-R^2 is contributed by the scale model (σ_Y). All model terms have associated $p < 2 \times 10^{-16}$ (Table B.3).

Response (Y)	Model (μ_Y)	Intercept-only R^2_{CS}	Smooth R^2_{CS}
Bias[ln($\widehat{b}$)]	cs(ln(b))	0.494	0.510
Bias[$\widehat{\text{E}[s]}$]	cs(E[s])	0.568	0.639
Bias[ln($\widehat{\text{Var}[s]}$)]	cs(ln(Var[s]))	0.381	0.443

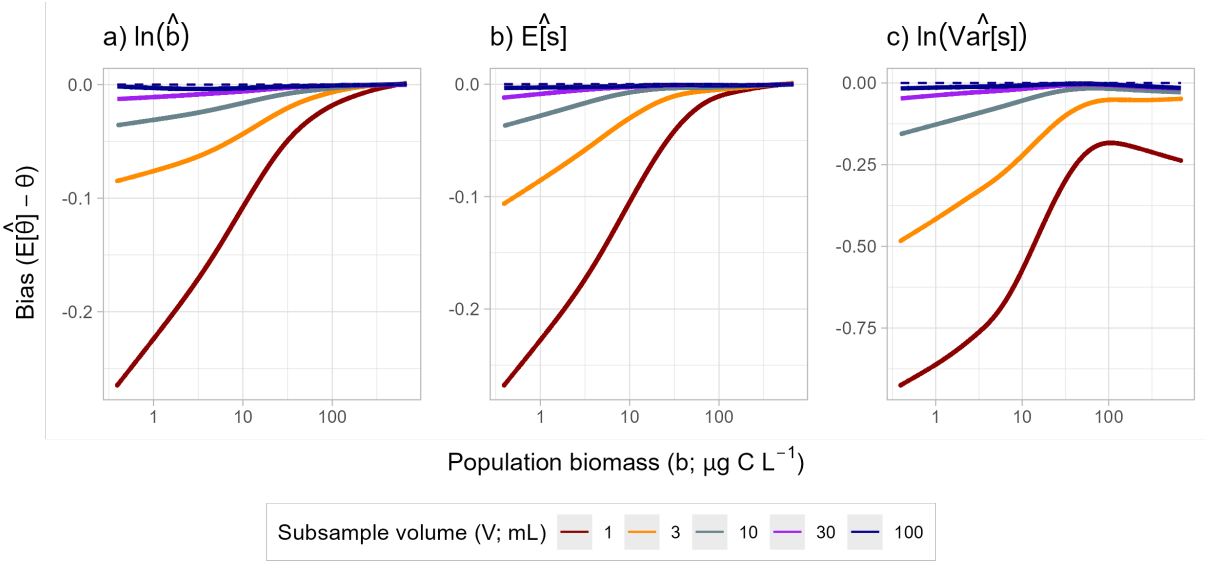


Figure 3.7: Relationships between 'population' biomass and estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes.

tributed substantially to the likelihood maximisation, with comparatively smaller contributions from estimation of the response location. Exhibited in Table 3.2 is a decomposition of R^2_{CS} for models for the location (μ_Y) of the response $Y = \text{Bias}[\hat{\theta}]$, all of which fitted from $V = 1$ mL simulations with saturated and scale (σ_Y). The location model (with response μ_Y) is built incrementally, with the scale model (with response σ_Y) always saturated. Initially, using an intercept-only model for μ_Y, the R^2_{CS} is already relatively high ($0.381 - 0.568$), solely a result from contributions from the scale model. Further improvements are made to model fit by including smooths across population parameters.

The relative contributions of the models for σ_Y can be further decomposed using the R^2_{CS} as above. If the intercept-only model for μ_Y is taken, and the random effect for σ is removed (i.e., an intercept and smooth model for σ_Y) the R^2_{CS} reduces to 0.061, 0.394, and 0.096 for Bias[ln($\widehat{b}$)], Bias[$\widehat{\text{E}[s]}$], and Bias[ln($\widehat{\text{Var}[s]}$)] respectively. Interestingly, E[s] contributes substantially to the scale model when $Y = \text{Bias}[\widehat{\text{E}[s]}]$, whereas the same is not true for equivalent models for biomass and size variance. Visual inspection of $\sigma_{\text{Bias}[\widehat{\text{E}[s]}]}$ fit indicates that a monotonically positive relationship on the E[s] domain: variability in bias increases with population mean size (particularly past 3.5 ln(μm)). Remaining contributions are then provided by the random intercept, particularly for σ_Y models for biomass and size variance where the change in R^2_{CS} is 0.433 and 0.275. The fact that size structure-specific random intercepts play such a crucial role in contributing to the scale of bias is an important result, as it indicates that the observed bias is highly erratic under some size structures, and more stable under other size structures.

3.3.2.2 Bias across the parameter domain

Bias is variable across communities, with relationships observed between bias of all indicators and biomass. Fitted models show that a reduction in subsample volume drives an increase in

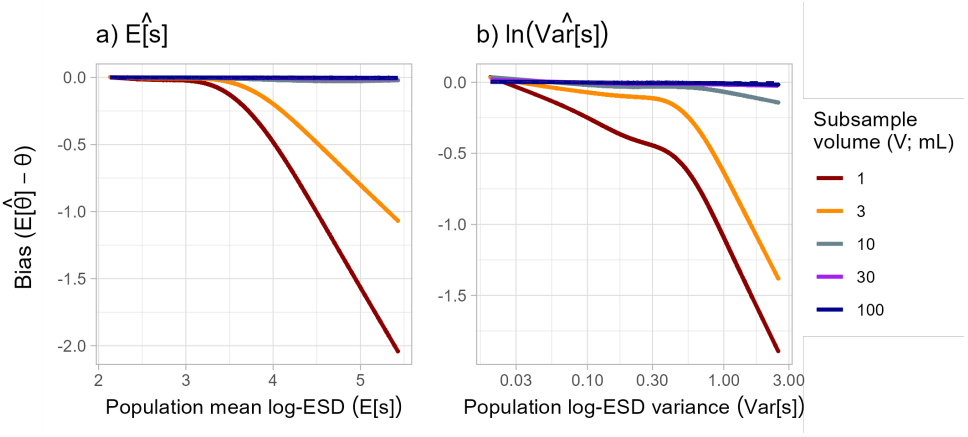


Figure 3.8: Relationships between 'population' mean log-ESD and estimation bias of mean log-ESD, and 'population' log-transformed log-ESD variance and estimation bias of log-transformed log-ESD variance, at different subsample volumes.

bias when measuring biomass, with the greatest bias is observed during low-biomass periods, and decreasing extent with increasing biomass (Figure 3.7). For example, the model fitted to Bias$[\ln(\widehat{b})]$ indicates that a 1 mL subsample is predicted to underestimate biomass by 18% at 2 μg C L^{-1} and 10% at 10 μg C L^{-1}. However, Bias$[\ln(\widehat{b})]$ is asymptotically zero for all V with increasing population biomass, indicating that slow community biomass is associated with bias.

A similar relationship is detected between Bias$[\widehat{\mathrm{E}[s]}]$ and biomass, and the effect on Bias$[\ln(\widehat{\mathrm{Var}[s]})]$ is even more severe (Figure 3.7). For a 1 mL subsample at $b = 2$ μg C L^{-1}, median estimated size variance is less than half of the population value; at $b = 10$ μg C L^{-1} it is still less than 60%. This indicates severe underestimation of size variance at low-to-moderate biomasses. Furthermore, the 1 mL subsample never achieves zero bias over the observed domain of b. 3 mL subsamples are also susceptible to bias when measuring log-ESD variance on $b \leq 30$ μg C L^{-1} with lesser but still notable extent. This aligns with expectations: as overall population size decreases and more taxa exist at or below the minimum recordable biomass (MRB) in general, sampling will become increasingly biased resulting in inaccurate estimates of biomass, mean size, and size variance. However, a value of biomass b can arise due to different combinations of abundance and size structure, so to briefly dissect the *drivers* of bias propagation a discussion is given in Appendix D.1.

Estimation bias in mean log-ESD and log-ESD variance was also assessed against respective population values. A similar, clear pattern emerges: in subsamples with $V \leq 3$ mL, bias affects populations with larger-than-median E$[s]$ or Var$[s]$. While no bias is detected on average for E$[s] < 3.4$ ln(μm) (approximate mean / median of E$[s]$), after this point, bias becomes increasingly severe. In fact, the model predicts that for a 1 mL subsample, the mean $\widehat{\mathrm{E}[s]}$, regardless of $E[s]$, is $\widehat{\mathrm{E}[s]} = 3.5$ ln(μm). In other words, the model predicts that $\widehat{\mathrm{E}[s]}$ itself is approximately asymptotic, so even at E$[s] = 4.5$ ln(μm), $\widehat{\mathrm{E}[s]} = 3.5$ ln(μm). Quantitatively, this is a massive difference (90 μm vs 33 μm on the original scale) and emphasises concerns about low-volume subsamples in a size-based functional ecology context. The 3 mL subsample also performs poorly

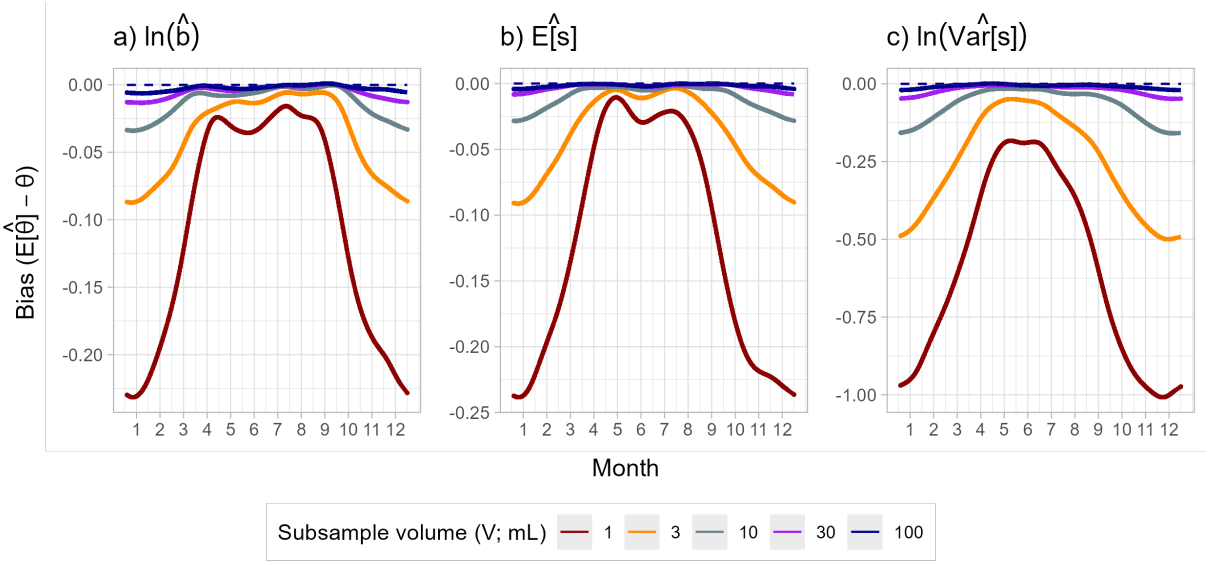


Figure 3.9: Seasonal cycle of estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes.

in this scenario, with $\widehat{\mathrm{E}[s]} = 4\ \ln(\mu\mathrm{m})$.

With increasing size diversity, estimation of $\widehat{\mathrm{Var}[s]}$ is also increasingly biased in 1 mL and 3 mL subsamples. Even at the median value for $\mathrm{Var}[s]$ ($0.4\ \ln(\mu\mathrm{m})^2$), the predicted estimate of variance is only 60% of its true value when $V = 1$ mL. This further deteriorates with increasing size diversity.

Overall, we observe that bias is typically most prevalent when communities have low biomass, large mean size, or large size variance. These are also the periods when the modelled scale of bias is at its highest indicating that the measurement is not only biased, but also increasingly erratic. Interestingly, observed log-biomass is also positively related to both mean log-ESD ($r = 0.26$) and log-transformed log-ESD variance ($r = 0.21$). Thus, whilst events of high biomass generally coincide with large mean size and large size variance, the relationship between bias and biomass − which is asymptotically unbiased when biomass is large − has a more complex relationship than these findings would initially suggest. This is reflected by the decomposition in Appendix D.1, where random effects for size structure are included in the model for location (μ_Y).

3.3.2.3 Seasonal variations in bias

Much of the analysis, discussion and subsequent hypotheses formed in Chapter 2 focus on seasonality of plankton community size structure. Assessing bias from the bootstrapping experiment on a seasonal scale can give insight to the potential shortcomings of the results presented previously in Chapter 2.

In low-volume subsamples, bias exhibits significant seasonality (Figure 3.9). This is observed across all indicators, and particularly visible in 1 mL and 3 mL subsamples. For each of the parameters, bias is at its most severe from November to February, and at its minimum in May

to August. Between these states, the bias is monotonically decreasing / increasing in magnitude. The results on a seasonal scale align with those presented in Figure 3.7 since population biomass exhibits strong seasonality with typically low values in winter and high values late spring to early autumn. While biomass and mean size are predicted to be around 80% of their true value in mid-January for a 1 mL subsample, it is undoubtedly the size variance which is most severely underestimated (at only 38% of its true value). Furthermore, 1 mL subsamples exhibit noticeable year-round bias (whereas 3 mL estimates of size variance are approximately unbiased in summer).

Modelled scale (σ_Y) also exhibits strong seasonality, which is proportional to the modelled location. Bias is not only greater in magnitude in winter, but is also more variable in winter. Estimates of biomass, mean size, and size variance, are therefore more erratic in winter.

With regards to results presented at SH & LE in Chapter 2, these findings suggest that inference of the large nano- and microplankton community at SH & LE should be made tentatively. In particular, there is scope that seasonal changes in size variance could be strongly driven by bias, at either a quantitative or qualitative level. The same may also be said for mean size, although the extent of this bias observed in this experiment is not as severe. However, it is important to be mindful that the results presented here are not reliably generalisable to other plankton communities, since the bias is inherently driven by the size structure of the empirical distribution from which cells are sampled. Regardless, this is a case study which is supported by individual-level observational data, statistical theory, and similar findings (Rodríguez-Ramos *et al.*, 2014).

3.4 Discussion

3.4.1 Scaling of size-related sampling bias from species level to community level

First, we presented evidence to support the hypothesis that large cells are omitted more often than small cells in low-volume phytoplankton samples (H7i). These results align with existing expectations that large cells are comparatively under-sampled (Goldman, 1993; Rodríguez-Ramos *et al.*, 2014), and are supported by fundamental statistical results. Then, we conducted a simulated bootstrapping experiment to show that phytoplankton biomass, mean size and size variance are underestimated in low-volume subsamples (in support of H8i, H8ii and H8iii). The underestimation of mean size undoubtedly comes from an under-representation of large cells in the community, since biomass is also underestimated in low-volume subsamples. It is therefore not possible that a decrease in mean size could be driven by an increase in sampled small cell biomass. The underestimation of size variance is also driven by this, and by a general decrease in sampled richness (aligning with Rodríguez-Ramos *et al.*, 2014; Cermeño *et al.*, 2014; Jacobs *et al.*, 2024a).

Seasonal variations in sampling bias are also consistent between the observations at SH & LE, and the bootstrapping experiment conducted at PL. At SH & LE, observations show that undersampling is more prominent across the board in winter, and that large cells are missed more

often in winter / autumn than in spring / summer, in support of H7ii. This is a result of seasonality of plankton abundance. In a similar vein, bias in simulated parameter estimates from low-volume subsamples is more severe for all indicators during autumn and winter. Again, given that biomass is systematically underestimated during these periods, a reduction in estimator accuracy is driven by undersampling of large cells.

A final key result is the importance of random effects assigned by each unique population size structure. Their ubiquitous significance for both location and scale suggests that variations from one to the community to the next are an important factor in driving the bias and its (predictably) unpredictable variability. At the theoretical level, this aligns with the relatively low area beneath the receiver operating curve for fitted models to observation in SH & LE, which indicate that a substantial degree of variability is not explained by fitted logistic models to taxonomic enumeration rates at SH & LE, which only have cell size as their response. While this is partly due to some unexplainable random noise, a key confounding factor here is 'true' abundance of each species. This is a fundamental result demonstrated in Table 3.1: regardless of size, if a taxon has low abundance in the population, it is less likely to be sampled. In conceptual terms, this is the information that is included in the random effect for models fitted to simulated samples from bootstrapping experiment. Given that the random effect term is important in explaining the value of bias, and is also crucial in explaining the associated variability of bias, we conclude that size structure specific effects have a role in determining the magnitude of bias arising from low-volume subsampling.

3.4.2 Inference for Chapter 2 results

A reasonable concern arises about the validity of results reported at SH & LE in Chapter 2, given the assertion that low-abundance taxa are missed often in low-volume subsamples, and that this issue particularly affects larger cells. While the propagation of the sampling bias from taxon to community level is not known at SH & LE, it can be inferred form the bootstrapping experiment that the mean size and size variance are likely underestimated, particularly in winter. The strong seasonality exhibited in microplankton size structure at SH & LE (Figure 2.16), which is characterised by low mean size in winter and large mean size in summer, could therefore be an artefact of sampling bias. It should also be noted that mean size is approximately invariant at PL (at a practical level; Figure 2.16), but if a 1 mL subsample is bootstrapped from these observations, then the mean size would exhibit a similar qualitative cycle to that observed at SH & LE (Figure D.2). However, these concerns are only related to the analysis of diatoms, dinoflagellates, and ciliates as presented here.

In the wider unicellular plankton community context, biases are unlikely to have a meaningful impact on results and subsequent inference, when flow cytometry data are available. This is because, at SH & PL where flow cytometry data is available, most of the winter biomass is contributed by pico- and small nanoplankton. The subsequent shift to large mean size in summer exhibited in Figure 2.14 is primarily a result of increasing biomass of larger cells, inherently measured by light microscopy, as opposed to a change in size structure within the community

of larger cells. Therefore, while quantitative differences may arise depending on the subsample volume used for light microscopy, the community-wide cycle of mean size is unlikely to change as a result of sampling biases.

Despite this, ecosystem contribution per unit biomass is not constant across the size domain due to food web dynamics (Armengol *et al.*, 2019), sinking rates resulting from mortality or from constitution of fecal pellets (Bach *et al.*, 2012), and maximal growth rates (Marañón *et al.*, 2013; Irwin *et al.*, 2006; Ward & Follows, 2016; Acevedo-Trejos *et al.*, 2022). food web dynamics are perhaps where large cells ($> 10\ \mu m$) become most important, since food webs dominated by small cells are tightly coupled, with a lower proportion of biomass escaping the microbial loop (Chen *et al.*, 2009; Cáceres *et al.*, 2013). As such, having an accurate assessment of the size structure within the large-cell portion of the phyto- / microzooplankton community can allow for important studies to take place. For example, García-Oliva *et al.* (2022) assessed shifts in biomass-size distributions at Helgoland Roads (54.2°N, 7.9°E), a productive coastal site in the Southern North Sea, to demonstrate the importance of mixotrophic dinoflagellates in the planktonic food web. However, such an analysis would be unreliable and biased if a small subsample was used (200 mL subsamples were used in that experiment, analysed by FlowCAM).

3.4.3 Comparison with existing work

3.4.3.1 Computer-based approaches

The implementation of computer-based simulation to imitate sampling has been used in a relatively small number of research papers, primarily focused on taxonomic enumeration and diversity measurement. In two papers discussed below, Monte Carlo methods are used for random sampling as opposed to bootstrapping. While Monte Carlo assumes a theoretical distribution for the specific abundances (with varying parameters), a bootstrapping approach uses an empirical distribution for the community. Each method has its advantages but there is undoubted value in having a range of distributional shapes; generally speaking, absolute precision in the representation of the plankton community is attainable only with a bootstrapping approach.

Early use of computer-based simulation to understand the statistical properties of plankton measurement centred around cell counts and related assumptions of Poisson distributions (Edgar & Laird, 1993). The assumption that cell counts are Poisson-distributed is historic in the literature originating with Lund *et al.* (1958), and arises from an assumption of spatial randomness in the water column. In the Monte Carlo experiment conducted by Edgar & Laird (1993), taxa were assigned theoretical distributions with different dispersion parameters arising from different assumptions of spatial randomness with respect to the fully random Poisson point process. The work confirmed that a misspecified assumption of a Poisson distribution yields underestimated error in confidence intervals.

More recently, Rodríguez-Ramos *et al.* (2014) used Monte Carlo sampling to enumerate taxa from a neutral community model for the purpose of measuring species richness. The model is dynamic with random processes for cell division, death, immigration, and emigration; the

'neutral' component comes from the lack of inter-species tradeoffs. Log-normal cell abundances by taxa is assumed, in line with the positive skew observed in observational data. The results clearly indicate that at conventional levels of sampling effort, richness is vastly underestimated. In the low-abundance community, conventional samples of richness are around half of the asymptotic value as effort increases, and around three-fifths of the asymptotic value in high-abundance subsamples, on average. The results are validated by observations. Conceptually, there is some alignment between these results and those we present. For example, Rodríguez-Ramos *et al.* (2014) show that, during low-abundance periods, a 50 mL underestimates taxonomic richness. In our work, estimates of functional diversity are biased for subsamples $V \leq 10$ mL for a community of similar size. However, species richness (which is a prerequisite for size variance in our context) is already capped in our empirical distribution, since the subsample volume at PL is 50 mL in these periods. Two potential reasons drive this difference: the limitation of the 50 mL subsample at PL creates an upper bound on the possible size variance or, more likely, size variance is asymptotically saturating with increasing richness, beyond a certain point. It is unlikely that cells rare enough to be missed by a 50 mL subsample, on average, will have a significant influence on relative abundance-driven metrics like mean size and size variance, since their proportional contribution is so small (Cermeño *et al.*, 2014).

Another interesting finding in the work of Rodríguez-Ramos *et al.* (2014) is that spatial dispersion of cells in the water column also influences measured diversity, for a given sampling effort. Overall, an assumption of spatial randomness increases estimates of size diversity, while clustering reduces it, particularly at low subsample volumes. Cross-validation of model output and field observations suggests that clustering is present in the modelled community. However, it is difficult to draw direct causal inference from such a pattern since clustering is not mechanistically described in the model. In our experiment, through lack of spatial resolution combined with random sampling, we assume spatial randomness. However, this only serves to act as a conservative assumption, since clustering of cells is a driver of estimation bias in a diversity context. Implementing this in a size-based context would be difficult, since assumptions would need to be made on the size-abundance relationship of the cells to drive parameterisation of the log-normal distribution. If such an assumption could be implemented, then we anticipate that future work would show that estimates of functional diversity are more biased than those presented here. Our work therefore presents a conservative lower bound for bias. Estimator accuracy of biomass, mean size, and size variance may deteriorate further when patchiness is included in the sampling experiment.

3.4.3.2 High-volume cell counting

The most prevalent method to establish bias arising from under-sampling, and the necessary sampling effort required to avoid this, is through the construction of rarefaction curves. These curves are constructed by counting species richness across incrementally increasing sampling effort. Indeed, this was the method used by Rodríguez-Ramos *et al.* (2014), and allowed validation of model results by observations. There are different methods for extrapolating the measured

species richness from a low-volume subsample, and significant variability is exhibited when these are compared, although they converge as the subsample volume increases (Branco *et al.*, 2018).

An extreme example of large-volume subsample enumeration is presented in Cermeño *et al.* (2014), where samples were counted exhaustively to estimate the true limit of species richness; subsample volumes analysed were as large as 25,000 mL. The conclusion from this work is that, to identify 90% of present taxa, subsamples of 250 to 1000 mL are necessary. Yet, even at subsample sizes > 10000 mL, there is visual evidence that richness is not yet asymptotic, with noticeable increases in richness in at least 3 of the 9 presented samples. Nonetheless, richness increases are clearly diminishing with increasing volume. In a functional ecology context, taxa observed for the first time at such large subsample volume will typically have a near-infinitesimal contribution to any metric of mean size or size diversity, since their proportional abundance is likely to be so small. This is acknowledged by the authors and, as such, their findings have important consequences for the measurement of taxonomic diversity, but are more limited in a functional diversity context.

3.4.3.3 Other sampling considerations

Our findings shed light on only a small aspect of the sampling regime which inflicts bias. A multitude of other factors are likely to drive bias during the sampling process, which thus require due consideration.

Firstly, the spatial distribution of cells in the ocean must be considered before sampling is pursued. Patchiness of phytoplankton samples is well-acknowledged on a range of scales, down to metres (Barth & Stone, 2024) or even microscopic levels (Mackas *et al.*, 1985; Davis *et al.*, 2004; McGillicuddy & Franks, 2019). The problem arising with patchiness is analogous to the issue of extrapolation when small subsamples are used, such as at SH & LE. For example, if 1000 cells in a 1 L body of water are randomly distributed, the counts (C) from a 1 mL subsample may be modelled as $C \sim \text{Poi}(1)$. Yet, if spatial randomness is violated due to clustering, zero-count probability increases, the probability associated with the mean decreases, and extreme values become more likely. The converse case would be driven by structured evenness, where the modal frequency of 1 cell subsample^{-1} observations would become increasingly peaked and count variance reduced (i.e., underdispersed). The clustered case is problematic in the time-series analysis of biomass-size distributions since zero-dominated and subsequent skewed non-zero data influence indicators such as mean size and size variance. Fixed-point sites such as SH, LE, and PL are potentially susceptible to such bases.

Other issues also arise in the process of enumerating cells. Two broad families of methods exist: direct measurement through *in situ* imaging, and capture with subsequent measurement of cells from the ocean. *In situ* imaging methods are becoming increasingly common, particularly in the sampling of mesozooplankton; first becoming widespread in 2010, and gaining further popularity from 2020 (Barth & Stone, 2024). Previously, imaging devices such as Visual Plankton Recorder (VPR) were typically constrained to a minimum size of ~ 300 μm (Ollevier *et al.*, 2022). However, more recent technologies are able to identify phytoplankton at a fixed site with high temporal

resolution (Barth & Stone, 2024). For example, the dual-magnification Scripps plankton camera (DSPC) are efficient down to ~ 10 μm (Merz *et al.*, 2021). Newer technologies such as DSPC are also better equipped to cope with turbidity (Merz *et al.*, 2021) than VPR (Ollevier *et al.*, 2022), which could be particularly relevant regions such as UK coastal seas. And, due to their enhanced spatial coverage, biases arising from clustering are also potentially minimised (Merz *et al.*, 2021), although statistical analysis of this information is still being developed (Barth & Stone, 2024). However, these technologies only cover a small volume of seawater within each frame: 0.5 mL in high-magnification mode for small mesozooplankton and phytoplankton colonies in VPR; (Davis *et al.*, 2004), and when in high-magnification mode, the DSPC only analyses $0.0002-0.01$ mL per frame (Merz *et al.*, 2021). As is the case demonstrated for traditional phytoplankton microscopy techniques, low volumes are a concern regarding the zero-abundance reporting of cells present in the community (Barth & Stone, 2024), which subsequently raises concerns for analysis in a size structure context. If a sufficiently large volume is analysed, then recent improvements in particle size measurement (Liu *et al.*, 2025) will likely allow for a satisfactory assessment of the plankton community.

Traditional methods, on the other hand, require capture of cells for subsequent enumeration by inverted microscopy. Sizes are then applied on a taxonomic basis. Generally speaking, this allows for comparatively larger samples to be analysed (Barth & Stone, 2024). However, the process of capture can itself have consequences for the enumeration, across both net and bottle sampling (Jiang *et al.*, 2020; Frau, 2025). Once cells are captured, clustering may also be an issue in chamber (Jacobs *et al.*, 2024a). While over time, this may average out if fields of view are randomly selected (Sandgren & Robinson, 1984), it may bring in the same issue of repeated zero-data and occasional large readings. Furthermore, it is unclear whether cell size affects settling position in the chamber. Some findings suggest that large cells have a tendency to settle towards the outside of the chamber (Haberyan & Haddock, 2013 via Jacobs *et al.*, 2024a), while others do not find a clear association with cell size (Sandgren & Robinson, 1984; Jacobs *et al.*, 2024a). Structure and shape may also be important, given that Jacobs *et al.* (2024a) found that *Chaetoceros spp.* chains have a significantly higher chance of settling near the edge of the chamber. More work is required to establish if clustering within the chamber has a tangible effect analyses of size structure. And, even after cells are settled, taxonomic biases can occur due to subjectivity associated with the identification process (Kovalenko & Reavie, 2022).

Perhaps a favourable method to mitigate against biases associated with *in situ* imaging and taxonomic analysis via inverted microscopy is the use of imaging flow cytometers (e.g., FlowCAM). This method analyses cells captured by bottle sampling by passing them individually through a flow cell. Images are taken so that size, fluorescence, and taxon estimated by software. Images can also be reviewed by the user, making the technology semi-automatic, although accuracy and precision of taxonomic estimation are fast improving through the implementation of artificial intelligence (Decrop *et al.*, 2025). FlowCAM allows for substantially faster measurement of cells, if cells are to be measured on an individual basis (Buskey & Hyatt, 2006). In the work of Buskey & Hyatt (2006), intraspecific size variability is demonstrated, with *Karenia brevis* cells ranging

from approximately 14 μm to 28 μm. The conceptual advantage of direct size measurement is clear, and while it has been shown that fine-scale variations in size structure are better represented using individual cell measurements (via FlowCAM), community-scale patterns appear largely indifferent from those measured by traditional microscopy methods (Álvarez *et al.*, 2014; Hrycik *et al.*, 2019). Such fine-scale patterns were used by García-Oliva *et al.* (2022) to demonstrate the impact of mixotrophic dinoflagellate grazing on biomass spectra. Therefore, in line with the findings of Hrycik *et al.* (2019), we propose that on the basis of existing evidence, FlowCAM appears a favourable alternative for the efficient construction of biomass spectra, since the taxonomic resolution offered by inverted microscopy is not normally necessary when individual cells are measured. To attain such precision in the measurement of cell size would take $5 - 10\times$ longer using traditional inverted microscopy methods (Buskey & Hyatt, 2006). However, given that the main aims of FOV data recorded at SH & LE is to monitor dominant species of interest, as opposed to accurately monitor patterns in biodiversity, the final conclusion is that data users must ensure that their practices align with the design of the sampling regime.

3.4.3.4 A wider picture on measuring size structure

Approaches to understanding fine-scale changes in size structure directly have proven fruitful in existing studies (Schartau *et al.*, 2010; García-Oliva *et al.*, 2022). However, over the course of a long-term motoring programme, individual biomass spectra are information-dense and thus require summarisation. Here, we do so by using a moment-based approach, summarising the spectra by mean size and size variance. The advantage here is that the spectrum is decomposed into two relevant ecological indicators, which illustrate the central tendency and diversity of the community. Yet, as is the case with any summarisation of an empirical distribution, information is lost. Other methods exist for measuring size structure which may therefore be appropriate in a given context.

Another popular method in the assessment of size structure is through the fitting of a power-law relationship between size and abundance / biomass. This is a general approach used widely in ecology, where abundance at size x is goverened by $A(x) \propto x^{-\alpha}$, where the exponent α governs the steepness of the slope (Clauset *et al.*, 2009). However, like a moment-based approach, the power-law fit can suffer a breakdown in accuracy due to the rarity of large organisms (Clauset *et al.*, 2009; Bisson *et al.*, 2022). Such uncertainty, and potential bias, is addressed by Bisson *et al.* (2022) who assess the size distribution for particles > 128 μm. The authors find that a truncated power-law relationship between size and abundance (e.g., a power-law with decreasing slope with size) best fits the observed size spectrum in surface waters, which is perhaps due to the undersampling of large particles. The truncated power-law is therefore more robust, as effects are identifiable: truncation-independent slope can still be interpreted as the lower end of the domain, whereas a single-parameter power law may be driven to be small by undersampled large organisms. Nonetheless, such an approach cannot account for important variations the size structure, such as size variance, since a fixed distribution is fitted to the data (so that unstructured departures from the fitted curve are lost).

3.4.3.5 Potential inference for existing studies

Relationships between bias and other indicators of community structure, such as biomass, indicate that spurious relationships may arise between indicators which are driven by bias. It is common for relationships between biomass and mean size to be assessed in an effort to approximately describe relationships between productivity and community structure exhibiting positive linear relationships in log-log space (Irigoien *et al.*, 2004; Hatton *et al.*, 2015; Cloern, 2018; González-García *et al.*, 2023). As such, it is feasible that positive relationships between biomass and mean size could be enhanced, or even driven, by such biases. However, included in the data used in both Irigoien *et al.* (2004) and Hatton *et al.* (2015) is PL (50 mL), along with other sampling sites (subsample volume not specified); in Cloern (2018) a $2 - 50$ mL subsample is used although biomass can be exceedingly large (> 1000 μg C L^{-1}); and in González-García *et al.* (2023) subsample volumes are $50 - 100$ mL. The concerns therefore do not extend to these findings.

Comparisons between biomass and size variance are also used to understand ecosystem functioning, by acting as a proxy for the relationship between productivity and functional diversity. Acevedo-Trejos *et al.* (2018) establish a positive log-log linear relationship between size diversity and productivity, however, the 100 mL used in the cruises (Marañón *et al.*, 2000) likely mitigates against any sampling-related biases. In García-Comas *et al.* (2016), where predator-prey biomass ratio (PPBR) is regressed against prey size diversity and predator size diversity, relatively small subsamples (4 mL for $50 - 200$ μm plankton; $2 - 5$ mL for $2 - 50$ μm plankton) are used. The data are composed from 11 cruises over a 4 year period in the months April to October (excluding June). Given that PPBR covers over 3 orders of magnitude, it is anticipated that biomass is varying significantly within the sampling period. As such, larger cells may have been undersampled during periods of low community biomass, likely towards April / October, although individual site considerations would be necessary for verification.

3.4.4 Limitations and future work

3.4.4.1 Subsample volume at PL

Evidently, subsample volume varies widely between regimes. In some cases, lower volume subsamples on the domain $1 \leq V \leq 5$ mL may be used (Jacobs *et al.*, 2024a; García-Comas *et al.*, 2016; Marine Scotland Science, 2018a,b); in other cases, large subsamples may be used ($V \geq 50$ mL; González-García *et al.*, 2023; Marañón *et al.*, 2000); while some sites, particularly those exhibiting strong seasonality, require a varying approach with higher volumes (50 mL) during low-biomass periods reducing as necessary with increasing biomass (Widdicombe & Harbour, 2021; Cloern, 2018).

Since median microplankton biomass at PL varies from 2.2 μg C L^{-1} to 54 μg C L^{-1} seasonally (Section 2.3.3.3), we believe that the reduction is sample volume at PL is unlikely to drive seasonal changes in bias, assuming it is inversely proportional to biomass (and thus maintaining approximately constant sampling effort). Despite this, Edgar & Laird (1993) showed that

premature termination of counts (e.g. decrease in effective subsample volume) leads to biased estimation of abundance, although the mechanisms for this are unclear. Rodríguez *et al.* (2000) reduced sample volume from 50 mL to 5 mL in low- and high-abundance states respectively, with no evidence of biased counting. Furthermore, the number of counts was over double of that in the smaller subsample, on average, indicating higher precision at the population-level.

With regards to the bootstrapping experiment, the reduction in subsample volume is still likely to reduce the species richness (following Cermeño *et al.*, 2014), due to the associated increase in the MRB. This is consequential because it confines the empirical distribution from which the bootstrapping experiment is conducted. However, the proportional contribution of these low-abundance cells, regardless of size, is likely to be minimal during high-biomass periods, given that the very reason to reduce subsample volume is a result of large contributions from other taxa.

3.4.4.2 Spatial randomness in the simulated water chamber

In the simulated chamber we assume cells are randomly distributed. While this may be unrealistic (Mackas *et al.*, 1985; Davis *et al.*, 2004; McGillicuddy & Franks, 2019), the inclusion of patchiness would only enhance estimation bias since large cells would be even less consistently sampled. To incorporate such potential patchiness would require a neutral model of spatially resolved particles (Rodríguez-Ramos *et al.*, 2014). This would increase mathematical complexity, require parameterisation, and incur significant computational expense since a simulated water body larger should be larger than 1 L to allow appropriate cluster formation. The use of an empirical distribution form observations would also likely no longer be appropriate. Future research should consider constructing such a model to investigate the effect of patchiness on a size structure sampling regime.

3.4.4.3 Modelling considerations

Several important considerations should be made with regards to the modelling effort. First of all, standard error estimation in `gamlss` is optimistically low, since the cubic regression splines are treated as fixed effects (Rigby & Stasinopoulos, 2005). Ultimately, this is a shortcoming of the methodology implemented. However, the only viable alternative − a GAM with no model for scale fitted in `mgcv` − produced visibly biased estimation of the location, as a result of heteroscedasticity in the error structure. As such, fitted GAMLSS models can be treated as unbiased, but their p-values are optimistic.

One important aspect of the model, to ensure its validity, was the fitting of random effects by size structure. While this allowed us to partition the variance explained by distribution-specific effects from the intended explanatory variable, the size structure effects themselves cannot be understood (rather, only that they are important). It is likely that certain characteristics of size structure lead to variability and unpredictability of bias, perhaps features such as long tails or multimodality. As such, some sort of dimensionality reduction may be useful to provide more insight, and fitted as a fixed effect to understand how different 'types' of size structures may be

more / less susceptible to sampling bias. At the moment our inference is limited to 'bias varies by size structure', with no indication of 'how' or 'why'.

Each model for location was fitted without random effects. Including a random intercept has mean (approximately) equal to zero, so an uninformed assumption would be that the fitted location model would, on an average, yield the same fit for the prediction at any value of the smooth over the explanatory variable. However, since likelihood is computed jointly with the random intercept and the term are subject to a penalty proportional to model complexity, the inclusion of the random effect therefore changes the maximum likelihood fit. As a result, fitted models may yield different estimates for the response, even when the added random intercept term has mean-zero. Furthermore, the random intercept in the draws information from the error structure which, in a GAMLSS context, is explicitly modelled by σ_Y. Competing for shared information may also change the maximum likelihood estimate, although it should be noted that the random intercept in a model for σ_Y relates to heteroscedasticity, whereas a random intercept in the model for μ_Y relates to a fixed amount of variability.

Finally, a log-transformation was applied to biomass and size variance in the modelling process, so that Bias$[\ln(\hat{b})]$ and Bias$[\ln(\widehat{\text{Var}[s]})]$ were response variables. This was necessary to avoid underdispersion in fitted GAMLSS models for μ_Y, since the residuals for bias observations, once accounting for random variability between the empirical distributions, became too peaked. The transformation could be circumvented by modelling a third response, for shape, in the GAMLSS framework, to explicitly handle the high-kurtosis of the residuals. While this would increase model complexity, the advantage would be that bias estimates were on the original units of the parameter θ, aiding interpretability.

3.4.5 Recommendations

On the basis of our findings, and in combination with existing findings in the literature, we emphasise the importance that a large subsample volume is used to produce unbiased estimation of size structure. In general, these findings suggest that for a coastal large nano- / microplankton community, subsample volume should be at least 10 mL, although reductions in subsample volume when biomass is high (e.g. ≥ 100 μg C L^{-1}) are appropriate. However, location-specific variations are extremely important in driving bias and its associated unpredictability.

Recent and future advances in methodology may naturally permit an increase in subsample volume, without an increase in sampling effort in person-hours. Technology such as FlowCAM, or *in situ* imaging methods such as DSPC may further be beneficial in allowing cell sizes to be estimated on an individual basis. However, the collection of cells for analysis by FlowCAM is likely to remain susceptible to patchiness. Conversely, *in situ* imaging may be more robust to patchiness, since it can be implemented with have a wider spatial coverage.

Importantly, it must be stated that these findings come from a statistical perspective, rather than that of a taxonomist. As a result, a universal message remains: data users must pay appropriate attention to the sampling regime, since methodology must be tailored to specific objectives in

order to provide unbiased estimation.

3.5 Conclusion

Low-volume subsamples under-represent large cells due to their rarity of abundance. A lower bound on potential bias can be derived theoretically from basic statistical properties of the Poisson distribution, with bias likely to be further enhanced by patchiness in both the water column and sedimentation chamber. We proved that large cells are under-sampled by using observational data at two sites where taxa are identified in paired 1 mL and 50 mL subsamples. Furthermore, the overall extent of sampling bias is greater across the board during low-biomass periods (winter and autumn). Then, to assess the impacts of under-sampling at a community level, we conducted a bootstrapping experiment using empirical data from large subsamples (50 mL outwith biomass periods). This experiment demonstrates that, for low-volume subsamples (< 10 mL), biomass, mean size, and size variance are systematically under-estimated. Crucially, biases are highly influenced by differences in size structure, affecting both the extent of the bias and its associated variability.

As a result of these findings, we recommend that larger subsample volumes are used if research objectives are focused on size structure, since large cells contribute significantly to biomass and diversity even in low abundances. However, despite the low subsampling volume, broad-scale patterns in the combined phytoplankton community remains well-understood at SH, since changes in size structure are driven by the seasonal shift from small cells (measured by flow cytometry) to large cells (measured by inverted microscopy). However, fine-scale analysis of microplankton communities measured with low-volume subsample should be made tentatively due to concerns about accuracy of large-cell enumeration.

Chapter 4

Construction, testing, and parameter optimisation of a size-structured plankton ecosystem model

4.1 Introduction

Having established the seasonal patterns of plankton community size structure and the associated environmental variability in Chapter 2, the focus of this thesis turns to providing a mechanistic understanding of the bottom-up and top-down processes which drive changes in size structure. The construction of an ecosystem model facilitates this process, since it provides a direct description of mechanisms controlling growth and losses; furthermore, it acts as an experimental sandbox which, through the testing of different formulations and parameters, allows the modeller to gain insight into the importance of different processes of the model. Such a hands on approach to modelling also reveals the strengths and weaknesses of the model, so that inference can be made with an understanding of underlying assumptions.

4.1.1 Modelling approaches

A wealth of plankton modelling research has already been undertaken in UK coastal waters and surrounding regions, including open ocean and coastal sites in the North Sea. Each of these models have specific goals: to investigate dynamics of plankton biomass accumulation (e.g., Steele, 1956; Tett *et al.*, 1986; Ross, 1993; Allen *et al.*, 1999; Lancelot *et al.*, 2005) or cell size (e.g., Terseleer *et al.*, 2014); assess end-to-end trophic transfer across food web (e.g., Haggan & Pitcher, 2005; Heath *et al.*, 2021); predict biogeochemical dynamics (e.g., Butenschön *et al.*, 2016); or to investigate the feeding, lifecycle, and migration of *Calanus finmarchicus* (e.g., Carlotti & Radach, 1996; Heath *et al.*, 1997, 1999). Despite this, models which focus on detailed allometric tradeoffs, such as those relating to growth rates and resource affinity, have not yet been applied to the UK coastal ecosystem. Given recent advancements in modelling methodology, such as the trait-based model of Terseleer *et al.* (2014) which includes a rare − yet

empirically supported − allometric decay of low-light photosynthesis (Taguchi, 1976; Edwards *et al.*, 2015), existing modelling frameworks can now be applied to UK coastal ecosystems to better understand the mechanisms controlling size-structure dynamics.

Construction of any model requires a tradeoff between complexity and computational expense. This is one of the reasons the models listed above have specific objectives to which they are tailored, rater than the existence of an all-encompassing 'master model'. Generally speaking, as the scope of the model widens, the computational expense also increases. Given that our focus is community size structure, it is necessary to tailor the model to these needs, and make compromises in other areas of the model.

4.1.2 Parameter space

An aspect of modelling which is often under-discussed in the literature is the process of parameter optimisation. A first step for parameterisation should always be to consult the literature to find plausible values from experiments or observations in nature. However, variations between species (e.g., Hansen *et al.*, 1994, 1997; Litchman *et al.*, 2007; Edwards *et al.*, 2015) mean that deviations from best-fitting relationships are likely to exist from one site to the next. While these deviations may 'average out' in a global model such as Ward *et al.* (2012), our applications are site-specific and thus likely to be impacted by deviations from established relationships. It is therefore necessary to optimise parameters, where possible, in order to both optimise model fit, and improve understanding of the parameters themselves.

Many presented models have some degree of informal parameter optimisation to tweak values, which normally involves manually adjusting values in order to minimise distance to observations (e.g., Terseleer *et al.*, 2014). There are some examples of formal parameter optimisation taking place (e.g., Heath *et al.*, 2021), but these tend to be rarer. Some parameters may also be unconstrained (free) due to a lack of evidence to support an initial value. In any case, optimising the fit of the model by adjusting its parameters relies on accurate data to guide the process. As such, findings in Chapter 3 which demonstrate that bias arises from low-volume subsampling must also be considered when devising a parameter optimisation regime.

When assessing the importance of parameters in a model, studies may also present a sensitivity analysis by varying parameters individually, e.g. by ±50% (Acevedo-Trejos *et al.*, 2015). However, this ignores joint effects between parameters (Heath *et al.*, 2021), and is undertaken with some implicit assumption that the response of the model to changes in parameters is somewhat monotonic. In reality, it is likely that the response of a model to changes in its parameters is complex and nonlinear due to cascading effects (for example, an adjustment to phytoplankton growth rate is likely to affect phytoplankton productivity and subsequently zooplankton grazing, which itself affects phytoplankton).

In light of this, the focus of Chapter 4 is to provide a rigorous mathematical and statistical description and investigation of the model. Hypotheses are then tested in Chapter 5, enriched with a detailed understanding of the strengths and limitations of the model, to make wider

ecological inference.

4.1.3 Aims and objectives

1. Construct a size-structured NPZD model which can accurately recreate seasonal patterns in biomass, mean size, and size variance of phytoplankton and zooplankton at PL, SH, and LE.
2. Use an appropriate method to optimise the parameters, with the joint aim of:
   (a) improving the model fit
   (b) understanding individual parameters' roles in the model, including the potential for allometry in the initial slope of the P-I curve
   (c) understanding the relationships between plausible values of parameters, in order to facilitate ecological inference
3. Assess characteristics of the model, including stability and sensitivity, so that the model can be used to make robust inference with knowledge of any potential limitations.

4.2 Methods

4.2.1 Model selection

Most size-structured plankton models follow from variations of the basic NPZD framework presented in Section 1.3.1 (EMPOWER-1.0; Anderson *et al.*, 2015). The most natural method is to extend the number of size classes in this modelling framework (e.g., Banas, 2011). The stoichiometry of organisms in each size class can then also be modelled, to reflect intracellular concentrations of differing nutrients (e.g., Ward *et al.*, 2012). Traditional size-structured NPZD approaches are flexible, and have minimal assumptions. However, they are computationally expensive because many size classes, and associated stoichiometry, must be explicitly solved at each time step.

Moment-based 'plankton functional type' models (described in Section 1.3.2.2) seek to address this computational expense by summarising the community using the metrics of biomass, mean size, and size variance, such that the number of model compartments can be drastically reduced. However, there are a few assumptions made by the model. The approach is dependent on an assumed distributional shape of log-size (typically normal, despite the fact that biomass distributions tend to be skewed or bimodal; Gaedke & Klauschies, 2017). Indeed, based on visual assessments of outputs in Figures 2.11−2.13 and skewness-kurtosis pairs shown in Figure 2.23, phyto- or zooplankton communities are rarely normally distributed at UK coastal sites. This issue can be addressed by splitting the community taxonomically, as seen in Terseleer *et al.* (2014), given that taxonomic groups typically cluster along the size domain. Nonetheless, at present, the size structure of the zooplankton still cannot be resolved in such a model despite recent advancements expanding to two phytoplankton traits (Le Gland *et al.*, 2021); the models

are also not currently able to resolve variable stoichiometry. The scope of such a model is therefore comparatively limited compared to that of a traditional size-structured NPZD model.

To minimise the number of assumptions associated with the model, and provide outputs of both phytoplankton stoichiometry and zooplankton size structure, we chose to construct a traditional size-structured NPZD model. The basis for the model is that of Ward *et al.* (2012) which remains an extremely powerful tool in contemporary modelling. We removed the 3D ocean physics (in a similar vein to simplifications made in Ward *et al.*, 2014), as well as partitioning of the nutrient pool and zooplankton stoichiometry component to reduce complexity and associated computational expense. Doing so allowed a detailed investigation of the parameter space. Changes to the grazing formulation and allometric tradeoffs were also implemented, in line with more recent works.

4.2.2 Model description

Model equations are described in Eq. 4.1−4.7. Phytoplankton (denoted in general terms by P) are resolved in units of carbon (P_C; mmol C m^{-3}), nitrogen (P_N; mmol N m^{-3}), and chlorophyll *a* (P_{Chl}; mg chl *a* m^{-3}). Zooplankton are modelled in units of nitrogen (Z_N; mmol N m^{-3}) with the assumption of a fixed N:C ratio. The nitrogen pool (N; mmol N m^{-3}) is considered a combined source of nitrate, nitrite, and ammonia. Detritis is resolved in units of nitrogen, and exists as particulate organic matter (D_1; mmol N m^{-3}) and dissolved organic matter (D_2; mmol N m^{-3}).

Phytoplankton nutrient uptake, chlorophyll *a* synthesis, and photosynthesis are denoted using the family of symbols V; grazing interactions are denoted $G_{\text{prey, pred.}}$; linear mortality by m; and MLD by H. Relationships of size-dependent processes are a function of volume governed by the allometric equation $y = aV^b$. Parameter values for a and b in size-dependent processes are given in Table 4.2 while parameter values of size independent processes are given in Table 4.1.

The biological component of the model is based on the model presented in Ward *et al.* (2012), for which much of the phytoplankton physiology should be credited to Geider *et al.* (1997, 1998). An adjusted grazing formulation is implemented following Vallina *et al.* (2014), and additional allometric tradeoffs are implemented. The slab physics are drawn from Anderson *et al.* (2015). The model is forced using site-specific *in situ* observations of deep-layer nutrients, surface irradiance, sea surface temperature, mixed layer depth, and seawater attenuation.

The model resolves $i = 1, \ldots, 15$ size classes of phytoplankton (s_i) on the range $0.38 - 85$ μm and $j = 1, \ldots, 25$ size classes of zooplankton (s_j) on $0.38 - 4086$ μm. The size grid is set such that for $i = j$, $s_i = s_j$ to avoid any unintended grazing trade-offs. The zooplankton exhibit omnivory, with passive switching between herbivory and carnivory. Only zooplankton with $s_j \leq 85\,\mu$m are grazed in the model. Zooplankton acting as prey are denoted according as $k = 1, \ldots, 15$. Thus, in Equation 4.4 grazing on zooplankton is denoted by the sum over k, while grazing on phytoplankton is denoted by the sum over i. Larger zooplankton ($s_j > 85\,\mu$m) are subject to quadratic mortality.

$$\frac{\mathrm{d}P_{N,i}}{\mathrm{d}t} = \underbrace{V_{N,i} \cdot P_{C,i}}_{\text{N uptake}} - \underbrace{\sum_j G_{i,j}}_{\text{grazing loss}} - \underbrace{\left(m_P + \frac{\omega + H'}{H}\right) \cdot P_{N,i}}_{\text{mortality and mixing}} \tag{4.1}$$

$$\frac{\mathrm{d}P_{C,i}}{\mathrm{d}t} = \underbrace{V_{C,i} \cdot P_{C,i}}_{\text{photosynthesis}} - \underbrace{\sum_j G_{i,j} \cdot \frac{1}{Q_{N,i}}}_{\text{grazing loss}} - \underbrace{\left(m_P + \frac{\omega + H'}{H}\right) \cdot P_{C,i}}_{\text{mortality and mixing}} \tag{4.2}$$

$$\frac{\mathrm{d}P_{Chl,i}}{\mathrm{d}t} = \underbrace{V_{Chl,i} \cdot P_{C,i}}_{\text{Chl } a \text{ synthesis}} - \underbrace{\sum_j G_{i,j} \cdot \frac{Q_{Chl,i}}{Q_{N,i}}}_{\text{grazing loss}} - \underbrace{\left(m_P + \frac{\omega + H'}{H}\right) \cdot P_{Chl,i}}_{\text{mortality and mixing}} \tag{4.3}$$

$$\begin{aligned}\frac{\mathrm{d}Z_j}{\mathrm{d}t} = {} & \underbrace{\beta \cdot \lambda_Z \cdot \left(\sum_i G_{i,j} + \sum_k G_{k,j}\right)}_{\text{grazing gain (inc. inefficiency penalties)}} - \underbrace{\left(m_Z + \lambda_T^{(Z)} \cdot m_{Z2} \cdot Z_j + \frac{\omega + H'}{H}\right) \cdot Z_j}_{\text{quad. mortality and mixing}} \\ & - \underbrace{\sum_k G_{j,k}}_{\text{grazing loss}}\end{aligned} \tag{4.4}$$

$$\begin{aligned}\frac{\mathrm{d}D_1}{\mathrm{d}t} = {} & \underbrace{\sum_i m_P \cdot P_{N,i} + \sum_j \left(m_Z + \lambda_T^{(Z)} \cdot m_{Z2} \cdot Z_j\right) \cdot Z_j}_{\text{carcasses}} - \underbrace{\left(\lambda_T^{(Z)} \cdot m_{D1} + \frac{\omega + v + H'}{H}\right) \cdot D_1}_{\text{remineralisation, mixing, and sinking}} \\ & + \underbrace{(1 - \beta) \cdot \left(\sum_i \sum_j G_{i,j} + \sum_k \sum_j G_{k,j}\right)}_{\text{faecal pellets and sloppy feeding particles}}\end{aligned} \tag{4.5}$$

$$\frac{\mathrm{d}D_2}{\mathrm{d}t} = \underbrace{\beta \cdot (1 - \lambda_Z) \cdot \left(\sum_i \sum_j G_{i,j} + \sum_k \sum_j G_{k,j}\right)}_{\text{excretion of unassimilated prey}} - \underbrace{\left(\lambda_T^{(Z)} \cdot m_{D2} + \frac{\omega + H'}{H}\right) \cdot D_2}_{\text{remineralisation and mixing}} \tag{4.6}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \underbrace{\lambda_T^{(Z)} \cdot (m_{D1} \cdot D_1 + m_{D2} \cdot D_2)}_{\text{remineralisation}} + \underbrace{\frac{\omega + H'}{H} \cdot (N_0 - N)}_{\text{mixing (replenishment)}} - \underbrace{\sum_i V_{N,i} \cdot P_{C,i}}_{\text{N uptake}} \tag{4.7}$$

Table 4.1: Size-independent parameters of the model. Prior distributions for free parameters are presented in Table 4.3; posterior estimates in Table 4.4. Source: [a]Gan *et al.* (2010); [b]Dutkiewicz *et al.* (2012); [c]Marañón *et al.* (2013); [d]Eppley (1972); [e]Dam & Peterson (1988); [f]Geider *et al.* (1998); [g]Chen (pers. comm.); [h]Ward *et al.* (2012). $^*\lambda_Z$ is calculated under the assumption zooplankton N:C ratio is always balanced.

Parameter	Symbol	Value	Units
Physics			
Chl *a* attenuation	k_{Chl}	0.025[a]	(mg chl *a*)$^{-1}$ m^2
Cross-thermocline mixing	ω	free	m d^{-1}
Detritus sinking speed	v	10[b]	m d^{-1}
Deep layer nutrient concentration	N_0	8	mmol N m^{-3}
Biology			
Reference temperature	T_{ref}	18[c]	°C
Phyto. temperature response	R_P	0.063[d]	-
Zoo. temperature response	R_Z	0.08[e]	-
Cost of biosynthesis	ζ	2.33[f]	mmol C (mmol N)$^{-1}$
Maximum Chl *a*:N ratio	$\theta_N^{\max}$	3[f]	mg chl *a* (mmol N)$^{-1}$
Max. N:C ratio	$Q_N^{\max}$	0.21[g]	mmol N (mmol C)$^{-1}$
S.D. of grazing kernel	σ	0.5[h]	-
Food ingested	β	0.7[h]	-
Food assimilated	λ_Z	0.5*	-
Phyto. linear mortality	m_P	0.01	d^{-1}
Zoo. linear mortality	m_Z	0.01	d^{-1}
Zoo. quadratic mortality	m_{Z2}	free	(mmol N m^{-3})$^{-1}$ d^{-1}
POM remineralisation	m_{D1}	0.034[h]	d^{-1}
DOM remineralisation	m_{D2}	0.017[h]	d^{-1}

4.2.2.1 Phytoplankton stoichiometry (Q)

Phytoplankton can vary their N:C (Q_N; mmol N mmol C^{-1}) and Chl:C (Q_{Chl}; mg chl *a* mmol C^{-1}) ratios depending on environmental factors. The value of Q_N is constrained by $Q_N^{\min}$ and $Q_N^{\max}$, while the maximum cellular chlorophyll content is related to the Chl:N ratio ($\theta_N^{\max}$; mg chl *a* mmol N^{-1}). The intracellular N:C penalises nutrient uptake through satiation (Q_N^{stat}) and photosynthesis through limitation (λ_N) respectively, which are defined as:

$$Q_N^{\text{stat}} = \frac{Q_N^{\max} - Q_N}{Q_N^{\max} - Q_N^{\min}} = 1 - \lambda_N. \tag{4.8}$$

$Q_N^{\min}$ is size-dependent, with a negative exponent (Marañón *et al.*, 2013). While there is some evidence that $Q_N^{\max}$ is size-dependent, Chen (pers. comm.) found that there is no significant slope when OLS regression is fitted to data collated from Montagnes *et al.* (1994), Menden-Deuer & Lessard (2000), and Marañón *et al.* (2013). As such, larger cells can attain a wider range of N:C ratios, which follows from the fact that small cells must dedicate a larger proportion of volume to essential apparatus. $\theta_N^{\max}$ is invariant with size, while no lower bound exists for θ_N other than the practical constraint that it must be positive.

Table 4.2: Size-dependent parameters of the model. All relationships are of the form $y = aV^b$. Prior distributions for free parameters are presented in Table 4.3; posterior estimates in Table 4.4. Source: [a]Marañón *et al.* (2013); [b]Menden-Deuer & Lessard (2000); [c]Litchman *et al.* (2007); [d]Hansen *et al.* (1997); [e]Hansen *et al.* (1994) via Banas (2011). *coefficient adjusted for continuity for model stability. **although Hansen *et al.* (1997) propose -0.16, the evidence is weaker.

Parameter	**Symbol**	**a**	**b**	**Units**
Max. N uptake	$V_N^{\max}$	0.12^{a}	-0.07^{a}	mmol N (mmol C)$^{-1}$ d^{-1}
N uptake half-sat.	k_N	0.17^{bc}	0.27^{bc}	mmol N m^{-3}
Min. N:C ratio	$Q_N^{\min}$	0.14^{a}	-0.04^{a}	mmol N (mmol C)$^{-1}$
Max. photo. (≤ 5.8 μm)	$V_C^{\max}$	0.85^{a}	0.30^{a}	d^{-1}
Max. photo. (> 5.8 μm)	$V_C^{\max}$	8.71^{a}	-0.20^{a}	d^{-1}
Initial slope of P-I curve	α	free	free	mmol C (mg chl a)$^{-1}$ (μE m^{-2})$^{-1}$
Max. grazing rate (≤ 85 μm)	$G^{\max}$	25.7^{d}	-0.20^{d}	d^{-1}
Max. grazing rate (> 85 μm)	$G^{\max}$	1.66*	0^{d}	d^{-1}
Grazing half-sat. (≤ 85 μm)	k_Z	free	0**	mmol N m^{-1}
Grazing half-sat. (> 85 μm)	k_Z	free	0^{d}	mmol N m^{-1}
Opt. pred.-prey ESD ratio	θ^{opt}	1.69^{e}	0.15^{e}	μm (μm)$^{-1}$

4.2.2.2 Temperature effects (λ_T)

Biological rates for phytoplankton, zooplankton, and the implicitly resolved heterotrophic bacteria are dependent on temperature, with rates adjusted by the Arrhenius equation:

$$\lambda_T^{(P \text{ or } Z)} = \exp\{R_{P \text{ or } Z} \cdot (T - T_{\text{ref}})\}. \tag{4.9}$$

Zooplankton display a stronger response to temperature than phytoplankton ($R_Z > R_P$; Rose & Caron, 2007; Chen *et al.*, 2012).

4.2.2.3 Phytoplankton nutrient uptake (V_N)

Dissolved inorganic nitrate is absorbed by phytoplankton cells to facilitate cell growth. The nitrogen uptake per unit carbon (V_N; mmol N mmol C^{-1} d^{-1}) is given by

$$V_N = V_N^{\max} \cdot \lambda_T^{(P)} \cdot Q_N^{\text{stat}} \cdot \frac{N}{N + k_N}, \tag{4.10}$$

where $V_N^{\max}$ is the maximum nitrogen uptake rate, which decreases with size, and k_N is the half-saturation value of nitrogen uptake, which increases with size. As a cell's nitrogen quota saturates ($Q_N^{\text{stat}} \to 0$), the nutrient uptake diminishes to ensure $Q_N \leq Q_N^{\max}$. The allometric exponent of $V_N^{\max}$ indicates that small cells will have the maximum nutrient uptake under replete conditions, and the allometric exponent of k_N indicates that small cells will have the maximum nutrient uptake under deplete conditions.

4.2.2.4 Phytoplankton photosynthesis (V_C)

Nutrient uptake must be complemented by photosynthesis for balanced cell growth. Growth rate through photosynthesis (V_C; d^{-1}) is averaged across the mixed layer since the model implicitly

assumes a uniformly distributed cell concentration across the mixed layer.

$$V_C = V_C^{\text{sat}} \cdot \frac{1}{H} \int_{z=0}^{H} \left(1 - \exp\left\{\frac{-\alpha \cdot Q_{Chl} \cdot I(z)}{V_C^{\text{sat}}}\right\}\right) \mathrm{d}z - \zeta \cdot V_N, \tag{4.11}$$

$$\text{where } V_C^{\text{sat}} = V_C^{\max} \cdot \lambda_N \cdot \lambda_T^{(P)}$$

The maximal photosynthetic rate ($V_C^{\max}$; d^{-1}) is size-dependent, increasing log-linearly with log-cell size until $\sim 5.8\mu$m and decreasing thereafter (Chen & Liu, 2010; Marañón *et al.*, 2013). A unit conversion of 86400 s d^{-1} is applied to $V_C^{\max}$ for unit consistency with I (μE m^{-2} s^{-1}). The initial slope of the photosynthetic curve (α) is considered a free parameter. A size-dependent and size-independent formulation for α is explored during the parameterisation process.

Surface photosynthetically active radiation (I_0; μE m^{-2} s^{-1}) is externally forced (Figure 2.9) and decays with depth (z) due to seawater attenuation and chlorophyll shading, with equation:

$$I(z) = I_0 \cdot \exp\left\{-z\left(\cdot K_w + K_{Chl} \cdot \sum_i P_{Chl,i}\right)\right\}. \tag{4.12}$$

where k_w is externally forced (Figure 2.10).

The final term, $\zeta \cdot V_N$, denotes the cost of biosynthesis proportional to nitrogen assimilation and acts without direct dependence the photosynthetic activity of the cells. ζ is a fixed parameter which is equal across cell sizes.

4.2.2.5 Chlorophyll production (V_{Chl})

Phytoplankton require Chlorophyll *a* to facilitate photosynthesis. Chlorophyll *a* synthesis (V_{Chl}; mg chl *a* (mmol C)$^{-1}$ d^{-1}) is limited by photosynthetic activity (see symmetry with Eq. 4.11), and scales with V_N under the expression:

$$V_{Chl} = \theta_N^{\max} \cdot \frac{1}{H} \int_{z=0}^{H} \frac{V_C^{\text{sat}}}{\alpha \cdot Q_{Chl} \cdot I(z)} \cdot \left(1 - \exp\left\{\frac{-\alpha \cdot Q_{Chl} \cdot I(z)}{V_C^{\text{sat}}}\right\}\right) \mathrm{d}z \cdot V_N \tag{4.13}$$

Analysis using L'Hôpital's Rule (Section 1.3.3) shows that $V_{Chl} \to \theta_N^{\max} \cdot V_N$ as $I \to 0$, meaning that Chl *a* synthesis is maximised in the absence of light.

4.2.2.6 Zooplankton grazing (G) & quadratic mortality (m_{Z2})

For the purposes of grazing, the zooplankton are split in two groups by size: 'protozooplankton' for organisms ≤ 85 μm ($j = 1, 2, \ldots, 15$), and 'metazooplankton' for organisms > 85 μm ($j = 16, 17, \ldots, 25$). The groups are loosely defined since they are both governed by Eq. 4.4, but the framing is driven by a combination of empirical evidence and modelling constraints. Firstly, the evidence in Hansen *et al.* (1997) suggest that $G^{\max}$ should be decreasing within protozooplankton, but constant within metazooplankton. Furthermore, k_Z may also be higher for protozooplankton than metazooplankton, although the evidence for a true allometric relationship

is weaker (Hansen *et al.*, 1997). Finally, there are not a complete set of predators for the meta-zooplankton, meaning that grazing pressure on large zooplankton would be under-represented in the model, thus necessitating a quadratic mortality term. This is given as $\lambda_T^{(Z)} \cdot m_{Z2} \cdot Z_j$ for $j = 16, \ldots, 25$. There is no grazing by zooplankton predators on any size classes subject to quadratic mortality to avoid any unintended tradeoff whereby some size classes are subject to excessive mortality.

Grazing interactions in the model behave under the basic premise that zooplankton respond to food availability under a sigmoidal curve using a Holling Type III response. The formulation is given by Vallina *et al.* (2014) and ensures maximal grazing under passive prey switching (that is, under prey-saturated conditions, total grazing is equal to $\lambda_T^{(Z)} \cdot G^{\max}$). The switching behaviour in this model allows each size class to favour herbivory or carnivory as a proportion of the available prey. However, the size-dependence of the feeding kernel stays constant.

An individual predator-prey interaction for zooplankton predator j and phytoplankton prey i ($G_{i,j}$; mmol N m^{-3} d^{-1}) is given by:

$$G_{i,j} = \lambda_T^{(Z)} \cdot G_j^{\max} \cdot \frac{F_j^2}{F_j^2 + k_Z^2} \cdot \frac{\varphi_{i,j}^{(P)} P_{N,j}}{\mathcal{F}_j} \cdot Z_j \tag{4.14}$$

for $j = 1, \ldots, 25$ and $i = 1, \ldots, 15$, and for a zooplankton predator Z_j on zooplankton prey Z_k:

$$G_{k,j} = \lambda_T^{(Z)} \cdot G_j^{\max} \cdot \frac{F_j^2}{F_j^2 + k_Z^2} \cdot \frac{\varphi_{k,j}^{(Z)} Z_k}{\mathcal{F}_j} \cdot Z_j. \tag{4.15}$$

noting that $k = 1, \ldots, 15$. Immediately, the Holling-III formulation can be identified as the third term, with the total food available to a predator Z_j given by:

$$F_j = \sum_i \phi_{i,j} P_{N,i} + \sum_k \phi_{k,j} Z_k \tag{4.16}$$

and the grazing kernel given as:

$$\phi_{(i \text{ or } k),j} = A_{(i \text{ or } k),j} \cdot \exp\left\{ -\left(\ln\left(\frac{\theta_{(i \text{ or } k),j}}{\theta_j^{\text{opt}}} \right) \right)^2 \left(2\sigma^2\right)^{-1} \right\} \cdot \mathbb{1}(s_j \geq s_{(i \text{ or } k)}), \tag{4.17}$$

for a prey of size $s_{(i \text{ or } k)}$ and predator of size s_j, so that $\theta_{(i \text{ or } k),j} = s_j / s_{(i \text{ or } k)}$, and noting that the indicator function $\mathbb{1}$ dictates that a predator must be at least as large as its prey. Consideration is required when choosing the normalization constant ($A_{(i \text{ or } k),j}$). Classically, this may be omitted (set equal to 1); however, this can be problematic. Intermediate-sized zooplankton, which are able to feed on a full kernel, have a competitive advantage over zooplankton at the extremes who may only have a partial kernel to feed on. With $A_{(i \text{ or } k),j} = 1$ for all predator-prey pairs, the area beneath the kernel is therefore variable across the zooplankton size domain. This inherently imposes an unintended trade-off that the effective k_Z is diminished at intermediate sizes, and comparatively larger at the extremes. $A_{(i \text{ or } k),j}$ is therefore chosen as a normalizing constant

such that $\sum_{(i \text{ or } k)} \phi_{(i \text{ or } k),j} = 1$ for all j. This was tested in a neutral model (all size-based tradeoffs set to zero) and yielded the desired result except for some small deviations in the two smallest size classes i (Figure E.1), which results from instability arising the constraint that prey cannot be larger than predators. It should be noted that k_Z becomes effectively dimensionless regardless of whether a normalization constant is included, particularly given that models of this style are quantitatively sensitive to the number of size classes. It should also be noted that the intermediate-sized zooplankton still have a wider range of available prey for which they have non-negligible preference; this is unavoidable if a fixed kernel shape is used for ϕ.

The adjusted grazing kernel ($\varphi^{(P \text{ or } Z)}$) applies a herbivory / carnivory weighting to ϕ. The notation $\varphi^{(P)}$ denotes the herbivory-adjusted kernel (i.e., grazing on phytoplankton), and vice versa for $\varphi^{(Z)}$. Firstly, $\varphi^{(P)}_{i,j}$ is given by:

$$\varphi^{(P)}_{i,j} = \frac{\sum_{i=1}^{15}(\phi_{i,j} \cdot P_i)}{F_j} \cdot \phi_{i,j} \tag{4.18}$$

while $\varphi^{(Z)}_{k,j}$ is given by:

$$\varphi^{(Z)}_{k,j} = \frac{\sum_{k=1}^{15}(\phi_{k,j} \cdot Z_k)}{F_j} \cdot \phi_{k,j} \tag{4.19}$$

The reweighted total available food for predator j ($\mathcal{F}_j$), which effectively acts as a normalizing constant for $\varphi^{(P)}_{i,j} \cdot P_i$ and $\varphi^{(Z)}_{k,j} \cdot Z_j$, is given by:

$$\mathcal{F}_j = \sum_{i=1}^{15} \varphi^{(P)}_{i,j} \cdot P_i + \sum_{k=1}^{15} \varphi^{(Z)}_{k,j} \cdot Z_k \tag{4.20}$$

Summing across all prey in Eq. 4.14&4.15 for a predator j can be shown to ensure maximal grazing, since:

$$G^{\text{tot}}_j = \sum_{i=1}^{15} G_{i,j} + \sum_{k=1}^{15} G_{k,j} = \lambda^{(Z)}_T \cdot G^{\max}_j \cdot \frac{F_j^2}{F_j^2 + k_Z^2}$$

with $G^{\text{tot}}_j \to \lambda^{(Z)}_T G^{\max}$ as $F_j \to \infty$.

4.2.2.7 Sloppy feeding $(1-\beta)$ and excretion $(1-\lambda_Z)$

A fraction of grazing interactions are exported into POM (D_1) through sloppy feeding and fecal pellets, and to DOM (D_2) as excretion respectively. The proportion of matter ingested by zooplankton during grazing is β and, as such, the proportion of grazing matter exported to POM is given by $(1-\beta)$. Linear and quadratic mortality losses are also transferred to POM. Consideration was given to split the quadratic mortality losses into excreted and fecal particles from the grazing by higher trophic levels which the quadratic mortality seeks to represent, but we decided against incorporating this additional complexity since it does not directly affect the

core aims of the model.

In the case of excretion, we fix λ_Z as constant, since we assume fixed stoichiometry of zooplankton for simplicity. We assume Z growth is perfectly balanced such that the quota is halfway between the minimum and maximum quota so that $\lambda_Z = 0.5$. A model with a specific focus on nutrient dynamics, e.g. to test a set of biogeochemical hypotheses, should incorporate varying λ_Z as a function of the nutrient quota fulfilment.

4.2.2.8 Cross-thermocline mixing (ω) and sinking (v)

Ocean physics have an important role to play in the supply of nutrients and export of organic matter to the deep layer. All constituents of the model are subject to cross-thermocline export at rate ω; deep-water nutrients (N_0) are also imported to the mixed layer, but it is assumed that there are no phytoplankton, zooplankton, or detritus in the deep layer. Detritus particles are subject to sinking at a fixed rate v. Since ω and v is constant, a greater proportion of matter is cycled across the thermocline when H is small. ω is considered a free parameter.

4.2.2.9 Variation of mixed layer depth (H)

The MLD (H; m) varies over the course of the year (Figure 2.7), influencing exchanges with the nutrient-rich and plankton-free deep layer. The rate matter crosses between the mixed layer and deep layer is H'/H (d^{-1}) with the condition that H' is constrained to be non-negative. This is because when the MLD shoals particles are assumed to be absorbed into the deep layer, rather than pushed upwards and concentrated. As such, when the ML is shoaling, $H' = 0$. When the mixed layer is deepening ($H' > 0$), particles are spread over a greater water volume and nutrient-rich water is entrained.

4.2.2.10 Linear mortality and remineralisation (m)

Phytoplankton and zooplankton are both subject to linear mortality losses which arise due to organism damage. Linear mortality is also commonly included in models to increase model stability. POM is remineralized at rate $\lambda_T^{(Z)} \cdot m_{D1}$ (d^{-1}) while DOM is remineralized at rate $\lambda_T^{(Z)} \cdot m_{D2}$ (d^{-1}). The $\lambda_T^{(Z)}$ term reflects the temperature sensitivity of the bacterial community which decomposes detrital matter and is shared with the zooplankton.

4.2.3 Operationalisation

4.2.3.1 Initial conditions

Model runs were initiated with uniformly distributed biomass across the phytoplankton carbon and zooplankton biomasses respectively. 15 μg C L^{-1} was divided evenly across the 15 phytoplankton carbon size classes, while equivalent to 5 μg C L^{-1} was divided across 25 zooplankton size classes. Since the range of allowable N:C varies across the phytoplankton size domain, initial phytoplankton nitrogen concentrations were set so that $Q_N^{\mathrm{stat}} = \lambda_N = 0.5$. Given that the maximum allowable Chl:N ratio is size-independent and given as $\theta_N^{\max} = 3$ mg chl a (mmol N)$^{-1}$, we

set $\theta_N = 2$ mg chl a (mmol N)$^{-1}$ at $t = 0$ such that $P_{Chl,i} = \theta_N P_{N,i}$

4.2.3.2 Numerical methods

The model was solved numerically in R using `ode(..., method='vode')` from the `deSolve` package (Soetaert *et al.*, 2010). The variable-coefficient ordinary differential equation solver (`vode`) is a flexible tool which switches between two methods and adapts step length within each iteration of the model depending on model stiffness. Although not concisely defined in the literature, stiffness is the general concept that variables may change at vastly different rates within the same system[1]. As a result, the slowly changing variable must still be modelled over a short time step to ensure reliable modelling of the rapidly changing variable. This is particularly important in chaotic systems, where small perturbations arising from inaccurate approximation may accumulate or blow up in size. In the simplest instance, when the model is non-stiff, the Adams-Moulton method is used. This is a semi-implicit method making it computationally cheap on a per-time step basis. However, when the model becomes stiff, an excessive number of steps are required to maintain stability. At this point, the fully-implicit Backwards Differentiation Formula (BDF) method is required, which is more computationally expensive per time step, but requires less substeps for each model time step to ensure stability. The `vode` method primarily uses an internally-computed Jacobian matrix, based on the instantaneous ODEs of the system, to determine which method is best. The `ode()` function has three arguments to control the precision of the solver: the absolute tolerance (`atol`), the relative tolerance (as a proportion of the modelled state variable, `rtol`) and the maximum number of steps allowed between each iteration of the model (`maxsteps`) if the criteria for `atol` and `rtol` are not satisfied. Reducing `atol` and `rtol`, or increasing `maxsteps`, increases precision but also increases runtime.

A timestep of 6 hours was chosen for the model. This was set to cover the four main phases of daylight: night $(21-03)$, the transitional phases of morning $(03-09)$ and evening $(15-21)$, and midday $(09-15)$. With the default setting of `maxsteps=10000`, the model could be approximated to a resolution of every 2.2 seconds when necessary for a given iteration of the ODEs, or over longer intervals (i.e. less steps) if tolerances were met. Default settings for tolerances were `rtol=1e-8` and `atol=1e-11`. Other methods such as the explicit Euler, implicit Runge-Kutta 4, and the default `lsoda` solver from the `ode` package were trialled on a shorter timestep, but were too computationally expensive, lacked accuracy by incurring frequent negative concentrations, or both.

The model was run for 20 years; outputs from the final 5 years were averaged (as per Ward *et al.*, 2012), due to a small amount of interannual variability which is typical of such systems (Banas, 2011).

[1] An elegant demonstration of 'stiffness' in differential equations, using a ball attached to a spring pendulum, can be found here. The angular motion as a result of the pendulum is comparatively slow compared to the radial oscillatory motion of the ball as a result of the spring. To accurately resolve this, a small time step would be necessary when solving a system of ODEs describing this motion, despite the fact that the relative change in the angular position would be tiny while the radial position evolves quickly.

4.2.3.3 Boundary conditions

For some parameter sets, phytoplankton and zooplankton size classes became effectively extinct. In these instances, biomass concentrations were > 30 orders of magnitude smaller than their competitors. This created an issue because it required a reduction of `atol` and `rtol`, or increase in `maxsteps`, to avoid negative concentrations. To make of the model evaluation feasible, a set of boundary conditions (BCs) was implemented:

$$\begin{aligned} P_C^{\min} &= 10^{-27} \\ P_N^{\min} &= Q_N^{\min} \cdot P_C^{\min} \\ P_{Chl}^{\min} &= \overline{Q}_{Chl} \cdot P_C^{\min} \\ Z_N^{\min} &= 1.6 \times 10^{-28}, \end{aligned}$$

noting the use of $\overline{Q}_{Chl}$ since no lower bound for Q_{Chl} exists. The $Z_N^{\min}$ condition is derived from Refield C:N ratios, such that equivalent $Z_C^{\min} \approx 10^{-27}$. These conditions were considered sufficiently small so that they would not impact long-term model dynamics.

4.2.3.4 Environmental forcing

Environmental forcing was imposed on the model to replicate the seasonal variations of one climatological year in sea surface temperature (T), surface irradiance (I_0), seawater attenuation (k_w), and mixed layer depth (H). These were forced separately to models for PL, SH, and LE using data and GAM fits presented in Chapter 2.

4.2.4 Parameter optimisation

Initially, manual tuning was used to gain familiarity with the model. However, an algorithmic approach was necessary to ensure that the parameter space was fully explored, and to facilitate statistically robust inference. We desired an estimate of the joint distribution to fully understand the role of parameters in the model and their relationships, and so used Bayesian methods to make inference on free parameters.

4.2.4.1 Selection of free parameters

Many parameters of the model were considered 'fixed', primarily those supported by strong empirical evidence. Ideally, these would be optimised (with a strong degree of prior weighting to the values presented in the literature), however, the associated computational expense prohibited us from doing so robustly. As such, only in cases where it was considered crucial were parameters chosen for optimisation.

In the first instance, cross-thermocline mixing (ω) was considered to be an important driver of summer nutrient loading, and was relatively unknown. As nutrients become depleted in summer, the term $\omega \cdot (N_0 - N)/H$ grows; this is compounded by the shoaling of the ML (H decreasing). As such, ω was considered chosen for optimisation.

Another unknown parameter of the model was the quadratic mortality of metazooplankton (i.e., the zooplankton size classes $j = 16, \ldots, 25$; m_{Z2}) which acted as a pseudo top-down control. This is generally seen as an important component of this style of model (Steele & Henderson, 1992; Edwards, 2000; Record *et al.*, 2014). Since quadratic mortality is applied to each size class individually, the total loss via quadratic mortality is $m_{Z2} \sum_j Z_j^2$. Conversely, in simple models with all Z biomass in one class, the mortality is effectively $m_{Z2}(\sum_j Z_j)^2$, since biomass is implicitly summed together. m_{Z2} is therefore sensitive to the number of size classes in the model, assuming biomass is spread across more than one zooplankton size class. Indeed, the extent of diversity within the model also influences the effect of quadratic mortality.

The values of k_Z, denoted herein as k_{Z1} for protozooplankton and k_{Z2} for metazooplankton, are also considered free parameters. Although the grazing formulation demonstrated in Vallina *et al.* (2014) ensures maximal grazing asymptotically, the 'true' meaning of the units of k_Z (mmol N m^{-3}) are still lost when the feeding kernel is applied to prey. Again, this is due to sensitivity relating to the number of size classes resolved in the model. If, in Eq. 4.14&4.15, $A_{(i \text{ or } k),j} = 1$, then as j increases, so too does $\sum_j \phi_{(i \text{ or } k),j}$. Similarly, as j increases, then if $A_{(i \text{ or } k),j}$ is imposed as a normalizing constant as is the case here, then individual $\phi_{(i \text{ or } k),j}$ diminish. Although increasing the size classes will almost certainly mean that individual Z_j biomasses will be smaller, diversity cannot be assumed to behave linearly in ecosystem models of this style (e.g., splitting a size class in three is unlikely to yield biomasses of three equal thirds, or three parts which sum to the original value). The scaling of k_Z with respect to the grazing kernel is therefore lost, as indicated by Banas (2011). Indeed, a wide range of parameter values in the literature exists, Banas (2011) (3 mmol N m^{-3}) and Ward *et al.* (2012) (0.16 mmol N m^{-3}), which likely relates to the choices each modeller makes when setting the number of size classes and implementing the grazing kernel.

Finally, the coefficient of the initial slope of P-I curve (α_a) and its exponent (α_b) are also explored. Although there is theoretical and experimental evidence that the shading effect should lead to a reduction in α with increasing cell size (Finkel, 2001; Edwards *et al.*, 2015), these works are exhibited at the species level, rather than community level. This leads to two uncertainties: whether there is indeed an effect at the community level, and − if there is − how the individual species curves relates to the community curve. During model parameterisation, both qualitative cases ($\alpha_b = 0$ and $\alpha_b < 0$) were investigated. It is common to implicitly assume some species succession in this style of model. For example, the formulation for $\lambda_T^{(P)}$ is typically monotonically increasing (under the condition that T exists within a realistic range), yet specific temperature response curves are unimodal. The same is true for light, with specific photosynthesis maximised at I_{opt} and diminishing thereafter due to photoinhibition. However, since the shape of the unimodal curve is different to a saturating curve (in particular, α_b has less effect on the curve than I_{opt} at intermediate light levels when a unimodal curve is fitted), it is not clear how α_b would scale to the community level. As such, it required investigation.

Table 4.3: Limits for uniform priors ($\pi(\cdot)$) of free parameters. *initially 1×10^{-6}, but reduced after poor fitting in a pilot test of 15,625 runs.

Limit	ω	m_{Z2}	k_{Z1}	k_{Z2}/k_{Z1}	α_a	α_b
Lower	0	0	0.1	0.3	10^{-8}	-0.33
Upper	0.2	1	0.8	1	5×10^{-7}*	0

4.2.4.2 Prior distribution

The end goal was to derive a joint parameter distribution, so inference was made under a Bayesian framework. Given the substantial uncertainty regarding each of the free parameters, a wide range of values was explored for each parameter. A uniform distribution was chosen to reflect this uncertainty, with limits displayed in Table 4.3.

The range of ω was based on the values presented in Anderson *et al.* (2015). The sites modelled in that work are open ocean sites; since the daily rate of cross thermocline mixing is given by ω/H, the effective mixing rate for a given ω would be much higher at shallow coastal sites (where H is typically smaller). As such, an upper limit of 0.2 was chosen for ω. The limits for m_{Z2} were as wide as feasibly possible (without incurring excess computational expense), while the range of k_{Z1} was chosen with the value of 0.16 used in Ward *et al.* (2012) as a basis.

Rather than sampling k_{Z2} directly from a prior distribution, the proportion of k_{Z2}/k_{Z1} was sampled. k_{Z2} is therefore conditional on k_{Z1} leading to a triangular domain for pairs (k_{Z1}, k_{Z2}). There is evidence that k_{Z2} is often lower than k_{Z1}, particularly amongst the smaller protozooplankton (Hansen *et al.*, 1997). An upper limit of $k_{Z2}/k_{Z1} = 1$ was allowed to avoid an excessively constrained prior.

The assessment of α was also considered important in the model, given the uncertainty around its exponent α_b. The upper limit, 0, ensured that the null-case of size independence could be explored, while the lower limit was derived from the surface area-to-volume ratio. If chloroplasts exist only on the surface of the cell and in fixed density along varying cell size, the exponent would be $-1/3$. If chloroplasts are arranged in layers due to the 'package effect', then intracellular shading would theoretically result in $\alpha \in (-1/3, 0)$. The value of $\alpha_b = -0.13$ presented in both Terseleer *et al.* (2014; via analysis of Taguchi, 1976) and Edwards *et al.* (2015) reflects this tradeoff. A detailed discussion of potential differences between values of α measured at a species level (as measured in Taguchi, 1976 and Edwards *et al.*, 2015) and the community response (of interest here and in Terseleer *et al.*, 2014) is provided in Section 4.4.2.1. Differences between effects at species and community level also motivates our investigation of α_b.

4.2.4.3 Algorithm selection

A variety of methods can be used to propose a candidate set of parameters for assessment. The simplest is Monte Carlo (MC), where parameters are proposed from the prior distribution. In this design, each proposed parameter set is independent, allowing parameters to be explored in parallel. An alternative is Markov chain Monte Carlo (MCMC), where the parameters are

proposed from a continually updating proposal distribution. The advantage here is that the chain should converge to higher likelihood regions and thus propose high-likelihood parameters more frequently. Initially, we intended to use MCMC to optimise the model's free parameters. The end goal was to use Delayed Rejection Adaptive Metropolis (DRAM; Haario *et al.*, 2006). The 'adaptive' component of DRAM builds on the Metropolis-Hastings (MH) algorithm, one of the more basic MCMC algorithms, to allow an adaptive covariance in the proposal of candidate parameters. This is particularly useful when parameters exhibit strong relationships. If two parameters are linearly related, there is limited use in continually proposing parameters perpendicular to the direction of the relationship. The 'delayed rejection' allows a proposed jump to be halved in size if it is not accepted at first. This improves mixing and convergence.

A natural start point is therefore to implement the MH algorithm. After specification of the prior distribution, there is one central choice associated with MH: the proposal variance. This controls how large the proposed jumps are, and requires tuning, a subjective choice which is attained through a combination of trial & error, and skill. When attempting to implement MH, however, a common theme prevailed: the chain would converge on regions of high likelihood, despite the fact they provided a poor qualitative fit to the data. As such, additional qualitative criteria was required during model evaluation to ensure qualitative features of the seasonal cycle were preserved. This required an Approximate Bayesian Computation (ABC) approach, which relaxes the use of the 'formal' likelihood ratio and allows use of analogous summary statistics in the evaluation of parameters.

A natural choice, then, would be to implement ABC-MCMC. This is identical in style to MH; the only difference is in the accept-reject step, where a binary indicator function based on 'informal' model fit is used, rather than the 'formal' likelihood ratio. However, the choice space increases to two parameters: the cut-off threshold for the accept-reject step (ε), since a formal likelihood ratio is not used, as well as the proposal variance as used in MH. There was also concern that the use of qualitative criteria to assess the model fit would lead to sharp edges on the surface of the 'model fit' space, meaning that many independent Markov chains would be required to reliably explore the full parameter space.

With this in mind, a simple ABC rejection sampling regime was chosen. A major benefit of this method is that the accept-reject threshold ε can be chosen post-hoc, meaning that the number of parameter runs can be well-constrained. The post-hoc selection, however, adds researcher degrees of freedom and can introduce bias. We therefore compared various values of ε and chose the lowest ε giving a representative approximation of the posterior was chosen based on visual assessment. It is also worth noting that such an approach is not possible in ABC-MCMC without re-running entire chains since the proposal distribution is itself conditional on ε, thus further justifying the use of non-MCMC ABC.

4.2.4.4 Approximate Bayesian computation (ABC)

The ABC method used here is based on Algorithm 2 in Marin *et al.* (2012). 93,750 runs were undertaken, with three values of ε compared in proportion to best-fitting model runs; we chose

0.1%, 0.3%, and 1% for comparison. The iterative method of proposing a set of parameters $\boldsymbol{\theta}' = \{\omega', m'_{Z2}, k'_{Z1}, (k_{Z2}/k_{Z1})', \alpha'_a, \alpha'_b\}$, and the accept / reject step is described below, noting that each parameter is uniformly distributed *a priori* (denoted $\pi(\cdot)$) with limits given in Table 4.3.

1. Sample parameters from the prior distribution, with $\boldsymbol{\theta}' \sim \pi(\cdot)$ (Table 4.3)
2. Run the model, using the proposed parameters $\boldsymbol{\theta}'$, to compute the distance $\mathcal{D}$ between the fitted model and the observations
3. Accept the parameters if $\mathcal{D} \leq \varepsilon$, otherwise reject

The measure of distance $\mathcal{D}$ was constructed by comparing the model fit to observational data at PL over $2007 - 2017$, the period in which flow cytometry and microscopy data were both available. PL was chosen since it had the most extensive data available on a temporal scale, as well as the identification of nanoflagellates. Concerns about sampling bias identified in Chapter 3 also contributed to the omission of SH given quantitative sensitivity. Firstly, a set of threshold functions were implemented based on the following criteria:

- Annual nutrient minimum must be lower than 0.5 mmol N m^{-3}
- Annual nutrient maximum must be larger than 4 mmol N m^{-3}
- The timing of the maximum mean phytoplankton size must be between days 105 and 240
- The timing of the phytoplankton minimum mean size must be between days 300 and 90

Constraints on nutrient minima and maxima were to ensure nutrient limitation affected uptake of larger cells in summer, and sufficient repletion in winter relaxed nutrient stress. Then, for the parameter sets satisfying all four threshold functions listed above, the following quantitative summary statistics were used:

- Phytoplankton biomass ($P_C^{\text{tot}} = \sum_i P_{C,i}$)
- Zooplankton biomass ($Z_N^{\text{tot}} = \sum_j Z_{N,j}$)
- Phytoplankton mean size ($\mathrm{E}[s_P]$)
- Zooplankton mean size ($\mathrm{E}[s_Z]$)
- Nutrients (N).

Ideally, phytoplankton size variance and zooplankton size variance would also be used, but a balance between total computation time (number of required mode runs) and complexity was necessary. The size variance would also be sensitive to competitive exclusion which, notoriously, is difficult to avoid in this style of model (Smith *et al.*, 2014). Finally, the parameter most likely to control the extent of size diversity in the model (σ) was not selected for optimisation. Work by Banas (2011) already demonstrates the effect of σ on this style of model.

The RMSE was then computed for each observation and the associated model fit. Z-scores were used as a means to weight each of the measures of distance (denoted X in general terms)

equally. To obtain an estimate of the error variance, GAMs with cyclic cubic regression splines were fitted to the data and estimates of the residual standard error computed (σ_X). As such, $Z_X = (X - \mu_X)/\sigma_X$ where μ_X is the fitted value of X from the model, and σ_X is the variance of de-trended values of X.

An adjusted RMSE was used for zooplankton mean size. This followed the same method as other summary statistics, but excluded observations of $\mathrm{E}[s_Z]$ between days 60 and 120. This was because the mean size increased over this period due to immigrating species (primarily nauplii of cirripeda, as well as larvae of decapods, and polychaetes; Highfield 2010). The cirripede are typically 514 μm in March at PL (Table A.4), yet cirripede larvae feed primarily on prey $5 - 7$ μm (Stone, 1989), albeit with interspecific variability and perhaps eating larger cells with development. As such, at least at an early stage, the θ^{opt} of these species varies between 73.4 and 102.8, which is higher than allometric theory predicts (24.0; Banas, 2011). Since there was no mechanism to account for this in the model, it was deemed that the model be optimised only for the mechanisms it is able to describe.

The measure of distance $\mathcal{D}$ was computed as the sum of z-scores for RMSE of $\ln(P_C^{\mathrm{tot}})$, $\ln(Z_N^{\mathrm{tot}})$, $\ln(N)$, $\mathrm{E}[s_P]$ and $\mathrm{E}[s_Z]$. Log-transformations were used to ensure the normality of errors required by the z-transformation. Furthermore, the qualitative criteria also had to be met. As such, the formulation for $\mathcal{D}$ was a hurdle-type process: first, qualitative checks must be passed, then RMSE must be sufficiently small.

4.2.4.5 Computation and presentation

Due to the independent nature of the sampling design, ABC is highly parallelisable. To maximise the number of parameter sets that could be tested, the tolerances were increased (`atol=1e-5` and `rtol=1e-2`) and maximum number of steps reduced (`maxsteps=3000`). This reduced the typical individual model runtime by approximately a factor of 16. The approach to measure sensitivity of the model to this reduction in solver precision is described in Section 4.2.5.1.

The marginal posteriors for each parameter were presented across each level of ε for each parameter, in addition to the correlation matrix of parameters. 95% credible intervals were also reported for the parameters. Seasonal cycles for the plankton community were computed for all accepted parameter sets constituting the posterior at PL, as well as the parameter set which minimised $\mathcal{D}$ with forcing at all three UK coastal sites. Plotting was undertaken in `ggplot` (Wickham *et al.*, 2019) and `ggpubr` (Kassambara, 2016).

4.2.5 Model stability

4.2.5.1 Sensitivity to solver tolerance

Two precision levels were employed when using `vode`: the 'high' level, for plotting and analysis, and the 'reduced' level (increased tolerance) to allow a large number of model runs to be executed during parameter optimisation. The distance between paired model runs across the two tolerance levels was computed for comparison. 80 parameter sets were drawn from the prior distribution.

This acted to check the sensitivity of the model to solver precision and its potential effect on the reduction of precision in the ABC optimisation. The forcing from PL was used.

The difference between model runs was used by calculating the difference between P_C^{tot}, Z_N^{tot}, N, E[s_P], E[s_Z], Var[s_P], Var[s_Z], and $D^{\text{tot}}(= D_1 + D_2)$ for the model runs, and summing these differences together. As before, to give each model summary equal weighting, a transformation was used: the range of each model compartment in the high-precision model run was used to non-dimensionalise both model runs. The subsequent RMS distance was calculated between the two runs.

4.2.5.2 Parameter sensitivity analysis

The model's sensitivity to parameters was demonstrated by varying each parameter (both free and fixed) individually by ±50%. The same distance metric was used as in Section 4.2.5.1, this time comparing the baseline model run (the parameter set which minimised $\mathcal{D}$) with the 'varied' run. Forcing from PL was used. The posterior of free parameters, and their relationships, was also assessed in relation the sensitivity of the model to its parameters.

4.2.5.3 Lyapunov exponent

A pervasive feature of the model was its inability to reach a true steady state. Employing extensive run times (> 150 years) during model development and increasing solver precision did not resolve the variability, despite the fact that a climatological year was used for forcing the model. Variability between years was therefore solely a result of the model's behaviour, or numerical approximation error, or a combination of both. Because the evolution of competitive exclusion is driven by relative differences in fitness between species (in this case, sizes), small differences in fitness may cause competitive exclusion to evolve over long time scales, perhaps thousands of years (Barton *et al.*, 2010). While these patterns are generally associated with subtropic regions, where evolution in environmental conditions is slow, the fact that a set of primarily smooth allometric tradeoffs are implemented here on a relatively fine size grid may indeed lead to small relative differences in fitness, thus causing competitive exclusion to evolve over long timescales.

To better understand the model trajectory we computed the Lyapunov exponent for 60 randomly sampled parameter sets from the posterior. The high-precision solver was used, and the model was run for 20 years in each case. This was done using forcing at PL. We computed the exponent directly, by initiating model runs with identical conditions except for a small perturbation applied to initial conditions. The distance between these two trajectories at time t can be given as $|\boldsymbol{\delta}(t)|$, meaning that the initial perturbation size is given as $|\boldsymbol{\delta}(0)| = |\boldsymbol{\delta}_0|$. As a result,

$$|\boldsymbol{\delta}(t)| \approx \mathrm{e}^{\lambda t}|\boldsymbol{\delta}_0| \tag{4.21}$$

with the Lyapunov exponent describing the rate of divergence ($\lambda > 0$) or convergence ($\lambda < 0$) of one model trajectory from another described by the equation. λ can be estimated by

approximating the expression:

$$\lambda = \lim_{t\to\infty} \lim_{|\boldsymbol{\delta}_0|\to 0} \frac{1}{t} \ln \frac{|\boldsymbol{\delta}(t)|}{|\boldsymbol{\delta}_0|}. \tag{4.22}$$

We computed $|\boldsymbol{\delta}(t)|$ using the nondimensionalised set of variables described in Section 4.2.5.1. For each perturbed run, we sampled a scalar perturbation $\delta_0 \sim U(-10^{-8}, 10^{-8})$ and applied this perturbation value equally to each nondimensionalised state variable.

4.3 Results

4.3.1 Parameter optimisation

4.3.1.1 Marginal posterior estimates of free parameters

The ABC parameter optimisation yielded informative posterior estimates for free parameters. 93,750 runs were undertaken, of which 3,079 satisfied the qualitative criteria. The values for ε at for the 1%, 0.3%, and 0.1% quantile of $\mathcal{D}$ were given as $\varepsilon_{0.01} = 7.26$, $\varepsilon_{0.003} = 6.67$, and $\varepsilon_{0.001} = 6.42$. Figure 4.1 shows the posterior estimates for each ε. The probability dense regions of the estimates posterior are similar across levels of ε, but as the cut-off becomes smaller (i.e. reducing the proportion of accepted parameter sets from 1% to 0.1%), the distributional shape typically becomes more peaked. Table 4.4 shows the minimum and maximum accepted values for each parameter, as well a the 95% Bayesian credible interval bounds when $\varepsilon_{0.001}$ is used.

Based on the posteriors estimates displayed in Figure 4.1, the model relies on well-chosen parameters; there is limited or zero posterior support for many regions of the prior, indicating that these areas correspond to poor fit. For example, only a small neighbourhood of the prior distribution of α_a supported in the posterior distribution. The upper limit for the 95% CI of α_a is 1.995×10^{-7}, which is far lower than the upper limit of the prior (5×10^{-7}). Furthermore, the upper limit of the 95% Bayesian credible interval for α_a (1.995×10^{-7}) is lower than the size-independent value of 3.83×10^{-7} used in Ward *et al.* (2012). Combined with the fact that the 95% credible interval for α_b is always smaller than zero, the values of α become even smaller moving along the size domain.

Importantly, in order for the model to provide a satisfactory fit to the data, the 95% credible interval for α_b indicates that there must exist a tradeoff such that the light-limited photosynthetic ability decreases with size. Indeed, there is diminished probability density as $\alpha_b \to 0^-$ even when using the more generous cut-off value ($\epsilon_{0.01}$). As ε is increased, increasingly negative values of α_b are accepted, but the upper limit remains relatively constant and excludes values close to zero. Thus, all presented cut-offs ε yield posteriors which support the inclusion of a negative α_b.

The half-saturation coefficients for grazing (k_{Z1} and k_{Z2}) also appear to have only a limited region of the prior supported in the posterior. Interestingly, the density of both k_{Z1} and k_{Z2} shifts to smaller values as the cut-off becomes stricter. In Ward *et al.* (2012), size-independent grazing half-saturation is $k_Z = 0.16$. This value is included in the 95% Bayesian CI for k_{Z2}, but not for k_{Z1}, which is higher (Table 4.4). The value used in Banas (2011) of $k_Z = 3$ is much

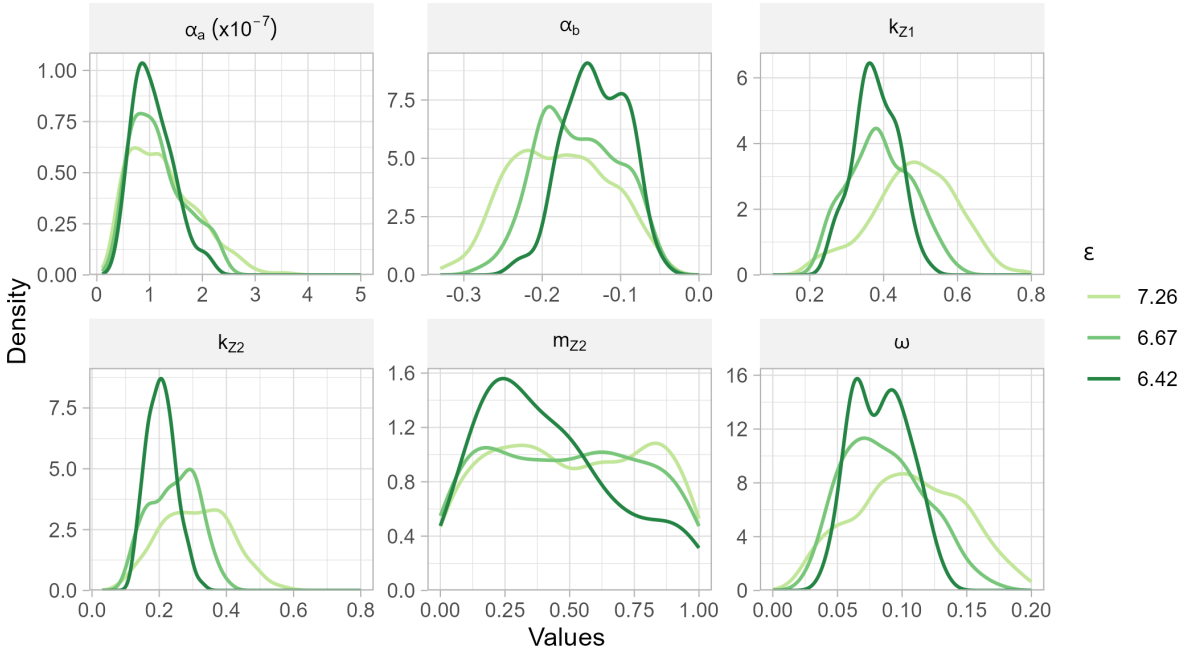


Figure 4.1: Estimates of posterior distributions of free parameters following ABC optimisation. Acceptance threshold values of $\mathcal{D}$ are denoted by ε. A lower ε corresponds to better model fit. Density plotted using `geom_density()` of `ggplot`.

larger, but as discussed in Banas (2011), the value carries little meaning when a grazing kernel is used; this is particularly true when a normalizing constant is introduced as in Eq. 4.17.

The posterior probability density of m_{Z2} is relatively constant across the bounds of the prior when $\varepsilon_{0.01}$ and $\varepsilon_{0.003}$ are used. Conversely, when the stricter tolerance is used ($\varepsilon_{0.001}$), the modal value of m_{Z2} is 0.24, and is more prominent than for larger values of ε. However, the estimate remains much less peaked than for the other parameters as a proportion of its prior. Notably, the posterior of m_{Z2} has probability distributed across the whole range of the posterior. While reducing the lower bound of the posterior would be unrealistic (as we require $m_{Z2} \geq 0$), the upper bound for the prior of m_{Z2} may have be increased in future. For $\varepsilon_{0.001}$, however, it does appear that the probability is asymptotically decreasing with increasing m_{Z2}.

Finally, the 95% credible interval for values of ω is $(0.048, 0.123)$, indicating that the value is

Table 4.4: Minima, medians, maxima, and bounds of 95% Bayesian credible intervals for each of the free parameters for the $\varepsilon_{0.001}$ cutoff. The parameter set which minimises $\mathcal{D}$ is denoted $\mathcal{D}_{\text{min}}$

Parameter	Min.	Median	Max.	95% Bayesian CI	$\mathcal{D}_{\text{min}}$
α_a $(\times 10^{-7})$	0.444	1.020	2.166	$(0.501, 1.995)$	1.199
α_b	-0.231	-0.134	-0.052	$(-0.203, -0.072)$	-0.136
k_{Z1}	0.256	0.376	0.522	$(0.264, 0.481)$	0.344
k_{Z2}	0.126	0.205	0.324	$(0.135, 0.291)$	0.188
m_{Z2}	0.022	0.355	0.962	$(0.048, 0.957)$	0.318
ω	0.042	0.084	0.132	$(0.048, 0.123)$	0.082

significantly lower than those used in Anderson (2015), where $\omega = 0.13$. However, as previously mentioned, the UK coastal sites have different oceanography compared to the sites used in Anderson (2015). Importantly, the water is much shallower at UK coastal sites than those open ocean sites, so the proportion of matter exchanged across the thermocline due to mixing (given by ω/H) is larger for UK coastal sites for much of the year, since H is smaller. The modal posterior estimate for ω is quite consistent across levels of ε. There is limited evidence of bimodality under $\varepsilon_{0.001}$, but a larger number of samples would be necessary to confirm if this were true; the 97 samples constitute the posterior when $\varepsilon_{0.001}$ is used are sufficient to establish the location and spread of the estimated posterior, but may be insufficient when establishing more complex distributional properties. Given that $\varepsilon_{0.001}$ maximises precision and complex distributional properties are not of primary interest, we choose this as the cut-off point for accepted values of $\mathcal{D}$ herein.

4.3.1.2 Relationships within the joint posterior

Relationships between parameters in the posterior demonstrate the tradeoff required between two or more parameters to maintain sufficiently good fit (i.e., a small value of $\mathcal{D}$). At a practical level it indicates: if one parameter is to be adjusted, what are the suitable values of other parameters? Understanding these relationships is not only useful to understand the statistical process of model fitting, but can reveal ecological tradeoffs of the model, and the coupling between different processes such as photosynthesis and grazing.

Free parameters of the model exhibit linear relationships (expressed via the correlation coefficient ρ), ranging from weak to strong (Table 4.5). Firstly, the parameters of α_a and α_b have strong negative correlation, with $\rho(\alpha_a, \alpha_b) = -0.74$. As such, increasing α_a or α_b is likely to significantly affect the dynamics of phytoplankton productivity, particularly during winter into the spring bloom, and subsequently impact model fit. This relationship is visually demonstrated in Figure 4.2, where the accepted values of α_a and α_b lie on a ridge.

Both α parameters also exhibit moderate correlation with k_Z parameters. The coefficient $\rho(\alpha_a, k_{Z1}) = -0.58$ indicates that, to maintain or improve model fit, an increase in the low-light photosynthetic ability of small cells must be accompanied by a (proportional) increase in the grazing intensity, via reduction of k_{Z1}. The same can be said for the relationship between α_b

Table 4.5: Correlation matrix for the estimated joint posterior distribution of free parameters when $\varepsilon_{0.001}$ is used as a cut-off. Diagonal values (always equal to 1) are replaced with '-' for ease of reading.

	ω	k_{Z1}	k_{Z2}	α_a	α_b	m_{Z2}
ω	-	0.28	−0.14	0.30	0.13	0.18
k_{Z1}	0.28	-	0.24	−0.58	0.40	0.00
k_{Z2}	−0.14	0.24	-	−0.10	−0.28	−0.21
α_a	0.30	−0.58	−0.10	-	−0.74	−0.06
α_b	0.13	0.40	−0.28	−0.74	-	0.16
m_{Z2}	0.18	0.00	−0.21	−0.06	0.16	-

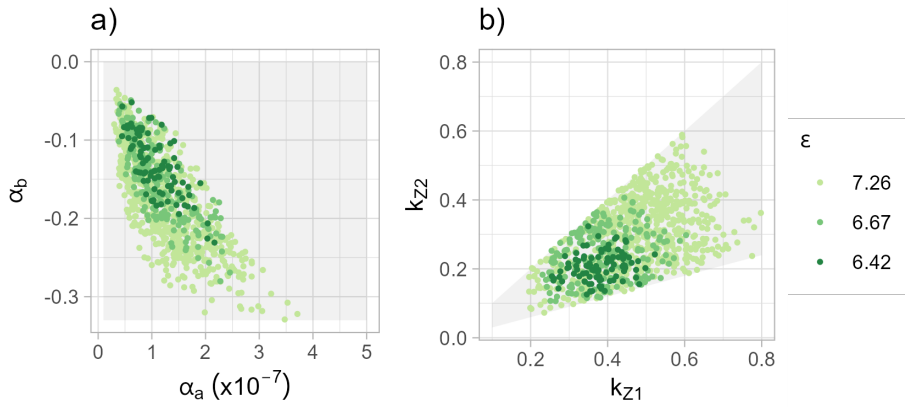


Figure 4.2: Accepted pairs of parameters for (a) α_a and α_b; (b) k_{Z1} and k_{Z2}. The grey shaded region shows the region of the uniform prior from which the parameters can be sampled.

and k_{Z2} for large cells, since $\rho(\alpha_b, k_{Z2}) = -0.28$. Given that only 6 parameters are allowed to vary, it is likely that confounding exists here. Nonetheless, whether it be via grazing parameters k_{Z1}, k_{Z2}, or $G^{\max}$, it is mechanistically feasible that a relationship with photosynthetic coefficients α or $P_C^{\max}$ exists, in order to maintain balance between phytoplankton and zooplankton, both in terms of biomass and size structure.

Given the choice of prior for k_{Z2}, it is unsurprising that $\rho(k_{Z1}, k_{Z2}) = 0.24$. Closer inspection of Figure 4.2b indicates that as ε reduces, accepted points tend to move away from the upper boundary of the k_{Z2} domain (grey shaded area). This supports the inequality $k_{Z2} \leq k_{Z1}$, but indicates that balance remains important: in general, if k_{Z1} is increased, k_{Z2} should also increase so that an excessive competitive advantage is not given to the small or large zooplankton. k_{Z2} is also negatively correlated with the density-dependent mortality of large zooplankton (m_{Z2}), which can also be explained mechanistically: if large zooplankton are subject to increased density dependent mortality, then they must also be able to consume more prey (via decreased k_{Z2}) to maintain equivalent net growth.

Notable correlations involving ω exist. In the model, ω is associated with two main functions: nutrient import, particularly in summer, when H is small and $N_0 - N$ is large; and export of both P and Z. The positive relationship between ω and α_a could arise for two reasons. Firstly, it is possible that the increase in α_a is necessary to balance the enhanced mortality associated with increasing ω. Or, to maintain stoichiometric balance between C:N in phytoplankton and, subsequently, C:Chl (since Chl synthesis is a function of V_N), α_a may need to increase to balance the additional nutrient loading incurred by ω. This affects α directly, since it is always multiplied by Q_{Chl}.

The positive relationship between k_{Z1} and ω is initially surprising. At first glance, one may expect that increasing ω − which increases mortality for zooplankton − must therefore be balanced by decreasing k_{Z1}. Similarly, increased nutrient loading (and correlated increase of α_a) may also require balancing by reducing k_{Z1} and increasing grazing pressure. As such, we speculate that the increase in k_{Z1} allows small phytoplankton to maintain an adequate population size when

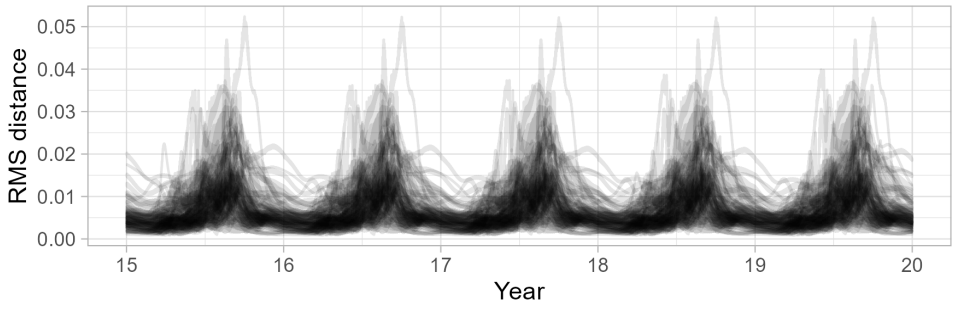


Figure 4.3: Root mean square distance of nondimensionalised model outputs for paired high precision vs. reduced precision model runs across years 15 to 20 of the model. Each line represents one of 80 parameter sets sampled from the prior distribution. Due to the nondimensionalisation, RMSE values can be interpreted approximately as the proportional difference between the paired model runs.

their cross-thermocline losses increase.

4.3.1.3 Sensitivity to solver tolerance

To investigate whether reduced solver precision used during the ABC regime supplied sufficiently accurate results, we compared the model outputs at two tolerance levels for 80 runs with parameters sampled from the prior distribution. On average, the nondimensionalised RMSE between runs was 0.0237 (a $\sim 2.4\%$ difference, if the minimum value of each indicator $P_C^{\text{tot}}, Z_N^{\text{tot}}, \ldots$ is used as the origin). This suggests that the reduced precision had an effect on accuracy; at any given model output time point these values ranged between $0.11 - 5.05\%$, with larger values typically observed between the months of May and October. There is no clear evidence that the paired runs are diverging (some minor interannual variability can be observed in Figure 4.3, but not with an identifiable trend).

Based on these findings, we conclude that the compromise of $1 - 5\%$ accuracy was worthwhile to facilitate the implementation of a more comprehensive ABC regime. Indeed, the reduction of solver precision is associated with an approximate $16\times$ increase in solver speed, and, as such, a $16\times$ increase in resolution of exploration of the prior parameter space. It appears unlikely that a quantitative difference on a small scale such as this would have a systematic impact on the selection of parameter sets for the posterior. If the primary aim of the work was to maximise accuracy of the parameter space, rather than accuracy of the model (with respect to the parameters), then a high precision approach would be necessary. However, here, the primary aim is to construct a model which accurately reproduces observations; for any given model, this is supported by maximising the exploration of the parameter space.

4.3.2 Model outputs at PL

4.3.2.1 Comparison to observations at PL

Using parameters sampled from the posterior distribution, the model provides a good fit to observations at PL (Figure 4.4). Broadly speaking, patterns of nutrients, plankton biomass, and plankton mean size are reconstructed by the mechanistic model, giving a similar fit to a GAM fitted with cyclic cubic regression splines. The model therefore appears to be an appropriate tool to investigate hypotheses presented in Chapter 2. However, there are some details which the models does not reproduce, which should be considered when testing hypotheses.

Perhaps the clearest discrepancy between data and model is the March-April mean size of zooplankton. This is previously discussed in Section 4.2.4.4, and these observations arise due to the fact that the spring zooplankton community has significant contributions from barnacle larvae, which deviate from the expected optimal predator size (given the observed prey during this period). Given the mechanistic purpose of the model, parameter optimisation excluded this period when computing $\mathcal{D}$. However, the model continues to quantitatively underestimate the zooplankton mean size thereafter, albeit to a much smaller degree during this period. In Figure 4.5, it is clear that zooplankton ~ 30 μm are often overestimated by the model, although the extent of this varies. Likely, a parameter optimisation of the grazing kernel (via the exponent of θ^{opt}, or σ) would allow a better fit to the model. Conversely, this may also arise due to the coarse top-down control on the large zooplankton; although m_{Z2} is optimised, other options for its formulation (e.g. Record *et al.*, 2014) are presented in the literature, or an allometric formulation may be necessary.

A second shortcoming of the model is its inability to produce a second phytoplankton biomass peak around August-September. Observations show that this biomass peak is dominated by dinoflagellates (Figure 2.12) so inclusion of mixotrophy may be necessary for a model to reproduce this. However, the fact that this secondary bloom does not occur in the model, while nutrients are replenishing via a deepening mixed layer, remains surprising; perhaps the loss via export of cells to the deep layer is too much to promote net biomass accumulation, even if there is an increase in productivity.

The nutrient cycle is qualitatively accurate, but does not exhibit a large enough range on a seasonal scale. Given that phyto- and zooplankton biomass are well-fitted to the observations during the period of maximal nutrient depletion (June-July), the most logical explanation is that the ocean physics are too simple. The cross-thermocline mixing rate (ω; d^{-1}) is considered fixed, which means that its proportional daily rate (ω/H; d^{-1}) is approximately double in summer than in winter, since the MLD shoals from 25 m to 12 m at PL (Figure 2.7). A reasonable proposition is that ω may be seasonally variable, with higher values in winter due to turbulent mixing resulting from storms and generally higher winds. An increase in ω in winter would enhance nutrient replenishment, while a decrease in summer would facilitate further depletion.

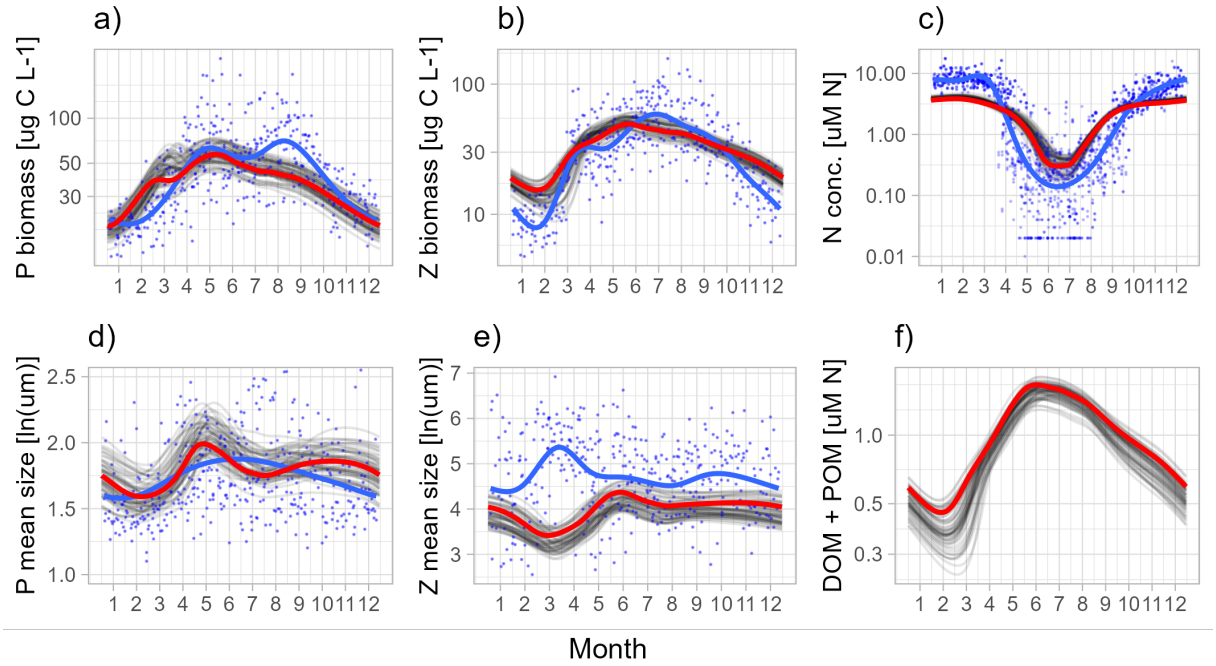


Figure 4.4: Model outputs for each of the parameter sets accepted with the $\varepsilon_{0.001}$ are shown as grey lines. The best-fitting sample from the posterior which minimises $\mathcal{D}$ is shown in red. Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line.

4.3.2.2 Size structure

Phytoplankton

Qualitative patterns in the biomass spectra of phytoplankton are reproduced by the model (Figure 4.5). Firstly, the two dominant size intervals of phytoplankton biomass ($4-8$ μm and $15-40$ μm) are identified by the model, with the phytoplankton on $4-8$ μm more dominant, in line with observations. Phytoplankton in the $4-8$ μm interval are slightly overestimated by the model, although it is important to note that Figure 4.5 treats biomass as an absolute contribution (as if size were discrete), as opposed to a density (if it were continuous). As such, the May biomass contributed by cells on $4-8$ μm is actually underestimated, despite the modelled biomass appearing larger than observations[2]. It is only in February, March, and June that noticeable overestimation of cells on this interval is made. Because the observations are assigned to their two closest size bins (in proportion to their relative proximity to each bin), and are fixed by taxon, comparison of size diversity on a small scale (i.e. within the $4-8$ μm interval) cannot be reliably made.

The larger phytoplankton (those on $15-40$ μm) are modelled accurately in the October to May period, but are under-represented in summer months (particularly July and August). However,

[2] A kernel-based approach, like that of Schartau *et al.* (2010) used in Chapter 2, would be necessary to transform these discrete points to a continuous density curve. Because the estimation of bandwidth would affect the variance of the presented biomass distributions, we opted to present the data in its raw form. This is necessary to convey competitive exclusion within neighbourhoods on the size domain.

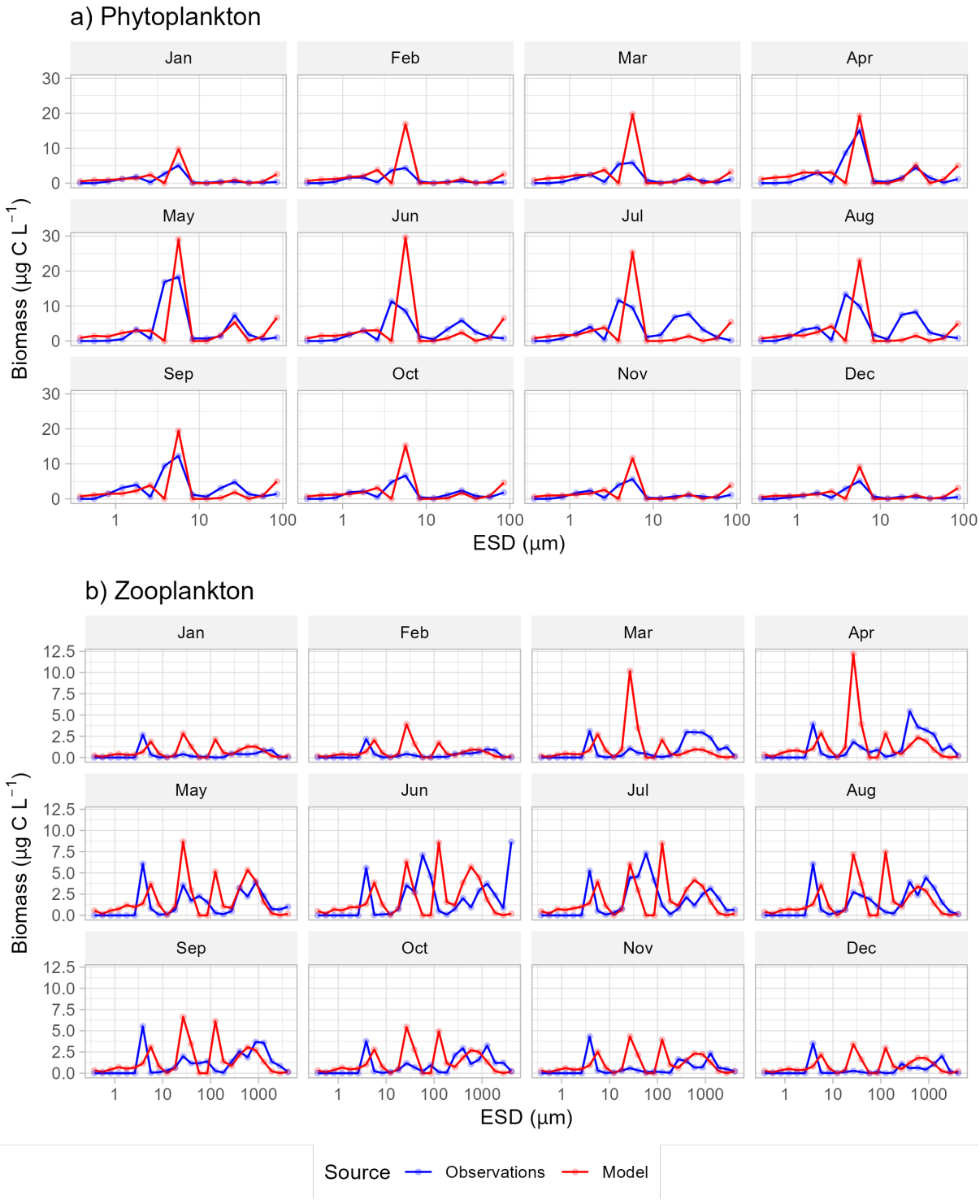


Figure 4.5: Comparison between observations and the best-fitting sample from the posterior which minimises $\mathcal{D}$. Truncated mean is used, as in Chapter 2, to calculate the mean biomass density for observations within each size bin. The biomass contribution at ~ 4000 μm in June of plot (b) is driven by rare blooms of *Beroe spp.* (33000 μm). It is therefore included for completeness, but model fit should not be compared to these observations.

interestingly, the largest phytoplankton (85 μm) accumulate noticeable biomass (up to 5 μg C L^{-1}) in this period. The model overestimates biomass in this neighbourhood of the size domain; given that cells of this size have no bottom-up advantage (owing to minimal uptake and affinity rates), it would appear almost certain these cells escape grazing. The size at which grazing evasion becomes a favourable strategy for phytoplankton may therefore be misspecified by the model, either via the grazing formulation itself or via the model parameters. Given the underestimation of cells in the $15 - 40$ μm interval, then these contributions likely contribute to the close quantitative fit of phytoplankton mean size with the observations (Figure 4.4). A direct comparison between observations at the size-bin level (rather than aggregated metrics like mean size) may therefore be necessary. However, this would come with additional challenges; this points to the importance of the formulation of the distance metric $\mathcal{D}$ in ABC approaches.

At the smaller end of the size domain for phytoplankton, the model facilitates diverse biomass contributions. There is then minimal biomass, year-round, in the size classes adjacent to the dominant size class (5.7 μm). This indicates that, on this local scale, competitive exclusion exists.

Zooplankton

The observed zooplankton biomass distributions at PL are more complex than the phytoplankton spectra; in some aspects the model performs well, while in others the performance is poor.

Starting with the smallest zooplankton (primarily nanoflagellates at ~ 4 μm), the model reproduces these with reasonable accuracy, with the peak at 5.7 μm consistently close to observations. However, the model significantly overestimates the microzooplankton, particularly at 26 μm. Most likely, this is in response to the available prey, since optimal prey size for these zooplankton is 6.5 μm, which is where phytoplankton biomass is focused. As discussed in Section 4.2.4.4, the March zooplankton community – influenced by meroplankton such as cirripede larvae – cannot be mechanistically explained by the model. Resolving functional groups of zooplankton for each size bin, with their own parameters (including those governing θ^{opt}), is likely necessary to capture this pattern. Clearly, this would substantially increase the complexity of the model and associated parameter space, and is thus beyond the scope of the study.

Across $40 - 125$ μm, the overall accumulation of biomass is comparable to observations, in summer months. However, the 58 μm and 85 μm size classes exhibit no biomass, yet their adjacent size classes (40 μm and 125 μm) dominate this area of the size domain. This may be explained by a discontinuity in the half-saturation constant of grazing between size classes 85 μm and 125 μm, which decreases from $k_{Z1} = 0.344$ to $k_{Z2} = 0.188$ (Table 4.4). Primarily, biomass of microzooplankton is observed at the 58 μm and 85 μm size classes at PL. The model therefore provides a good fit to data on the broader scale here, but misallocates the size classes which contribute this biomass.

With the exception of the March-April biomass accumulation, the larger zooplankton (> 200 μm) are accurately modelled at a broad level. The model resolves biomass accumulations primarily centred around 680 μm. However, the model is not able to identify the observed transition of

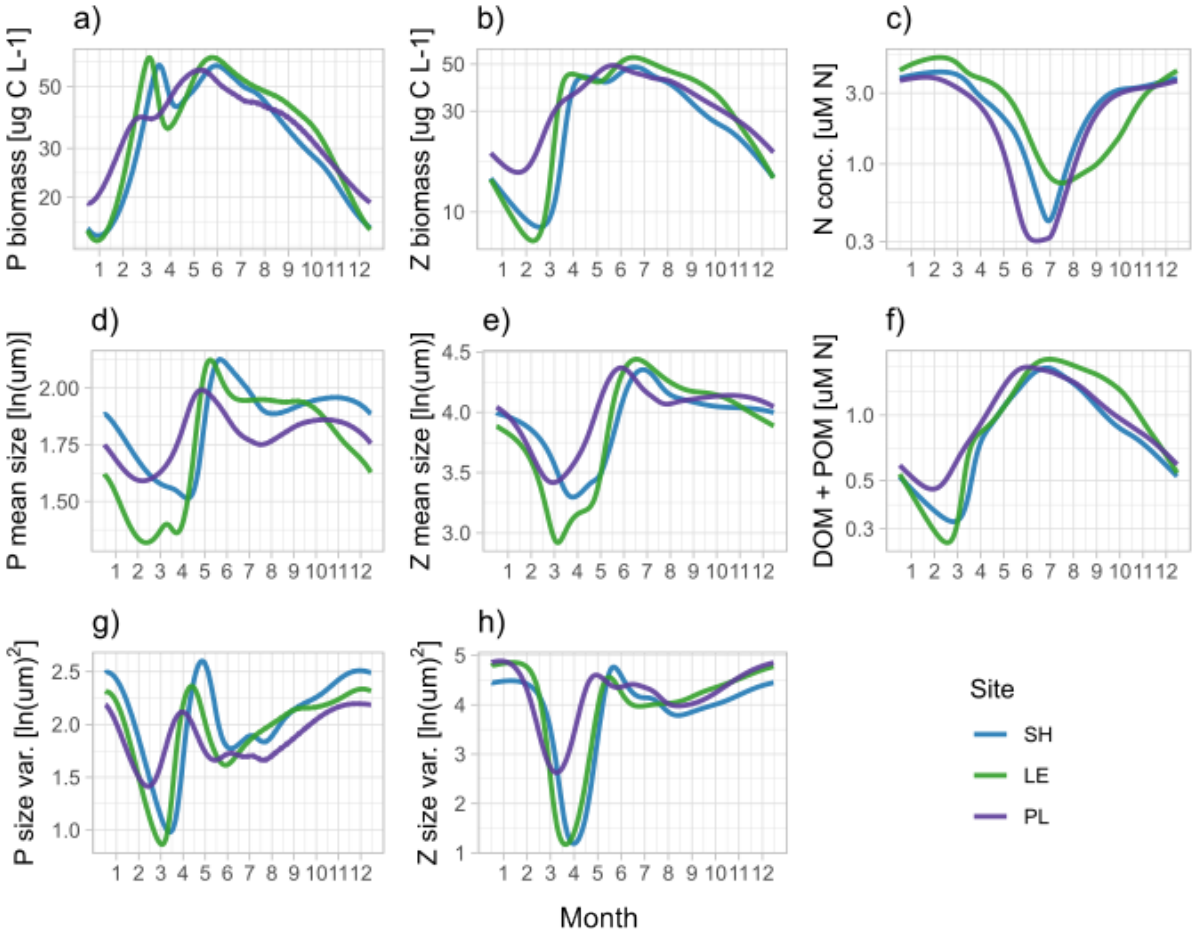


Figure 4.6: Model outputs resulting from the use of forcing from SH, LE, and PL. The parameter set which minimises $\mathcal{D}$ is used.

size from April to September. The observed peak of biomass within the region of the size domain is at 400 μm in April, but shift to 1280 μm through summer to September. This is driven by changes in both the taxonomic and age structure of the zooplankton community. It should be noted that there are observations of significant accumulations of the large comb jellies (*Beroe* spp.; 33000 μm) in June, which are assigned to the largest size class for observations (4085 μm). These observations are therefore included for completeness, but should not be used comparatively with this model.

On the domain of large zooplankton, accumulations in the model tend to be less 'exclusive'; the size classes adjacent to the dominant class in the model (590 μm). This may be a result of the independence of the quadratic mortality. We review the mortality term with respect to the work of Record *et al.* (2014) in Section 4.4.3.1. However, this also provides a qualitative match to observations which are diverse compared to other regions of the zooplankton size domain.

4.3.3 Model outputs at SH & LE

Qualitatively, the model exhibits a similar fit at each of the UK coastal sites (Figure 4.6). However, the magnitude and timing of seasonal variations differs across sites. Both phytoplankton and zooplankton biomass are lower in winter at SH & LE than at PL, aligning with previous observations. The model may therefore be used to provide insight into the differences in the winter community and address H1. However, the modelled bloom dynamics at SH & LE differ to those at PL, with a sharp peak in March followed by a trough in April, before subsequent recovery in May. This may be driven by the fact that zooplankton biomass is lower at SH & LE until April, and so top-down control is weaker during the biomass accumulation period. Given that the intention of the model is not to provide detailed investigation of bloom dynamics (a different formulation for $\mathcal{D}$, and tighter qualitative criteria would be necessary for this), the inference made with regards to the bloom should be approached with caution. The biomass cycles are similar from May onwards; at SH & LE, observations show that the autumn bloom is typically a smaller event (Bresnan *et al.*, 2015) than at PL (Rodríguez *et al.*, 2000; Eloire *et al.*, 2010), and so the qualitative fit of the observations at Scottish sites may actually align better here than at PL (where the September biomass peak is not reproduced by the model).

Observed N depletion in summer at SH is, on average, 0.5 mmol N m^{-3}, and at LE is 0.3 mmol N m^{-3} (Figure 2.4). The modelled nutrient depletion in summer is therefore well-fitted at SH, but is insufficient at LE (Figure 4.6). This is unsurprising: the complex oceanography of the LE system is almost certainly poorly resolved by simple slab physics, particularly given that the estimate MLD is constant (10 m; Figure 2.7). This has two consequences: the proportion of water exchanged between mixed and deep layers is larger for the given ω (since H decreases in the ω/H term), and there is no entrainment of nutrients of disposal of plankton via deepening or shoaling (via H'/H).

In terms of mean size, the phyto- and zooplankton exhibit greater seasonal variation at Scottish sites. This appears to align qualitatively with results presented at SH inclusive of flow cytometry for phytoplankton at SH (Figure 2.14). The extent of the variability is still much smaller than the GAM fitted to observations, although the exclusion of phytoflagellates from observations makes comparison difficult; if they were reliably enumerated, the observed mean size would likely be smaller. Comparison of modelled zooplankton mean size to observations is also difficult, since no data is given for heterotrophic nanoplankton via flow cytometry at either SH or LE.

Dynamics of size variance are interesting across all sites. Firstly, it is clear that the quantitative scale modelled here is far larger than predicted by a GAM fitted to observations (max fitted phytoplankton size variance is 1.2 $\ln(\mu m)^2$; Figure 2.14). Variance is a quantitively sensitive measurement, but qualitative feature of size variance should be expected to follow similar trends, broadly speaking. Although the modelled spring biomass accumulation at PL is slightly earlier than observed, the patterns of size diversity during this period are comparable to observations (Figure 2.14). However, the maximum phytoplankton size variance, observed in autumn, is not reproduced at PL or SH.

4.3.4 Model stability

To test for exponential divergence, the Lyapunov exponent was computed for 60 randomly sampled parameter sets of the posterior, with a varying perturbation size sampled from $U(-10^{-8}, 10^{-8})$. Plots of the exponent λ over time revealed a consistent pattern: a monotonic decrease, in proportion to $1/t$, after some initial instability in the initial timesteps of the model. Although a positive Lyapunov exponent ($\lambda > 0$) indicates exponential divergence, this is conditional on $t \to \infty$. Since our model runs only spanned 20 years, the approximation of λ on a finite timescale was therefore inconclusive. Investigating this further, we find that the log-ratio between perturbed and unperturbed trajectory at time t ($\lambda t = \ln(|\boldsymbol{\delta}(t)|/|\boldsymbol{\delta}_0|)$) is constant on an interannual scale, with seasonal variability (Figure 4.7). The fact that the long-term trajectory is oscillates around a fixed point suggests that the perturbed state may be attracted to this fixed point, and no longer diverges exponentially on a longer (multi-year) timescale. We verified the stationarity of each perturbed trajectory between years 15 and 20 by computing the least squares derived slope of λt; a t-distributed 95% CI for the average slope across different perturbation sizes and parameter sets was $(-0.031, 0.0215)$. This indicates that there is no evidence of long-term divergence between paired perturbed and unperturbed trajectories, on average, between years 15 and 20. However, it is possible individual slopes may differ significantly from zero (to test these, a time series decomposition would be necessary due to autocorrelation).

The fact the long term slope of λt is on average zero over the modelled domain of t suggests that it is possible $\lambda \to 0$ as $t \to \infty$. However, it is important to note that there is no evidence that the perturbed system converges to the same trajectory as the unperturbed. In this instance, we would require $|\boldsymbol{\delta}(t)|/|\boldsymbol{\delta}_0| \to 0$ (i.e., $\lambda t \to -\infty$). Since we observe that λt is always positive on years 15 to 20 (Figure 4.7b), $|\boldsymbol{\delta}(t)|$ is larger than $|\boldsymbol{\delta}_0|$ and, as such, the perturbed trajectory − although no longer moving *away* from the original trajectory − finds an oscillating state (limit cycle or similar) relatively far away from the original trajectory. Often, this distance is $50-3000\times$ larger than $|\boldsymbol{\delta}_0|$ (computed as e^4 and e^8 in Figure 4.7b). This reflects the sensitivity of the model to its initial conditions on a fine scale. Without evidence of the impact of a larger perturbation, it is not clear how larger changes in the initial conditions would affect the model trajectory. This

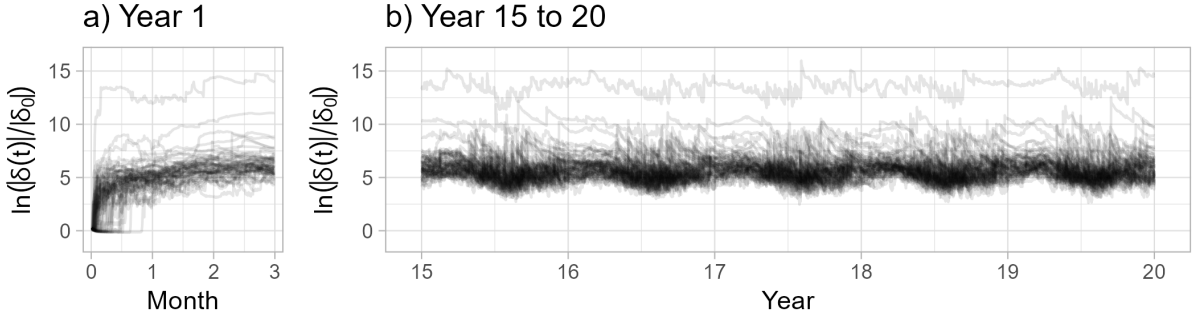


Figure 4.7: Evolution of distance between unperturbed and perturbed model trajectories from 60 randomly sampled parameter sets of the estimated posterior, across (a) the first 3 months and (b) final 5 years of the model runs.

motivates a fuller assessment of the sensitivity of such a model to its initial conditions on a larger scale than the small perturbations tested here. It may be the case that the posterior estimates of parameters, and subsequent model outputs, are conditional on the model's initial conditions.

4.3.5 Sensitivity to parameters

The sensitivity of the model to its parameters was tested by varying each parameter individually by ±50%. Averaging across parameters, the model fit is sensitive to changes, with differences generally on a much larger scale ($0.020-0.547$; Figure 4.8) than those exhibited due to sensitivity to solver tolerance ($0.001-0.035$; Figure 4.3). This supports the use of a reduced solver tolerance in order to maximise exploration of the parameter space. The sensitivity of the model to α, ω, m_{Z2} and k_Z parameters further supports their inclusion in the ABC regime. Indeed, some observations from the sensitivity analysis suggest that the model parameterisation could be expanded to further improve model fit. For example, the free and somewhat unconstrained parameters β and λ_Z appear to have a significant impact on model dynamics. In models such as that of Ward *et al.* (2012), λ_Z is derived directly from modelled zooplankton N and C concentrations, however, we assume zooplankton have fixed stoichiometry to minimise complexity. Perhaps future iterations of the model may benefit from variable zooplankton stoichiometry, or, at the very least, investigation of a constant β and λ_Z in the ABC regime. That said, using $\beta\lambda_Z = 0.35$ is reasonable starting point; Acevedo-Trejos *et al.* (2015) use equivalent $\beta\lambda_Z = 0.3$, while Anderson *et al.* (2015) set $\beta\lambda_Z \approx 0.52$.

A reduction in N_0 produces the most significant change in model fit. Since the value of N_0 is derived from observations, there are no concerns about the sensitivity of the model to the value itself. This does, however, act as a reminder that the replenishment of nutrients is a key process in the model, yet is somewhat coarsely resolved. Coupling with a 1D or 3D ocean model, which would improve the resolution of the ocean physics and associated nutrient replenishment, may therefore improve the model fit. It should also be noted that if this 0D model is implemented in regions without observations of nutrients to parametrise N_0, effort should be made to investigate the model fit for varying N_0.

It is notable that a reduction in m_{Z2} also appears to have a substantial effect on model fit. Although the probability density of the posterior is decreasing for all presented ε as m_{Z2} reduces (Figure 4.1), there is still notable posterior density around the region of $m_{Z2} = 0.15$ (i.e., 50% reduction). This is surprising: given that a 50% reduction in m_{Z2} substantially alters model output compared to the best-fitting parameter set, one may assume that the fit to the data worsens. Given this does not appear to be the case, it may therefore be the case that a reduction in m_{Z2} affects model fit in such a way to maintain relative good fit to the data. This is supported by the apparent absence of significant relationships with the other free parameters (Table 4.5), meaning that it is unlikely that its comparatively broad posterior is due to lack of identifiability. As such, these findings support the inclusion of m_{Z2} in the model (on the basis that a further reduction would further change model fit).

Among the parameters to which the model is sensitive are the coefficients and exponents of

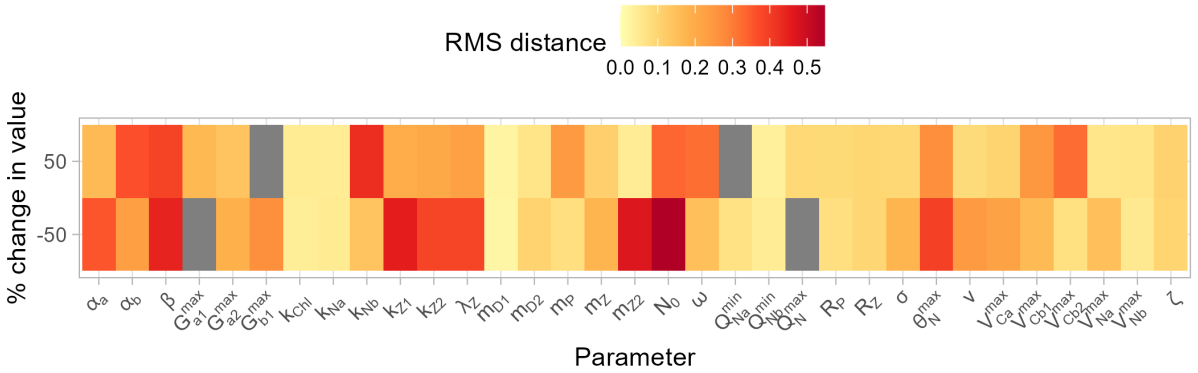


Figure 4.8: Root mean square distance of nondimensionalised model outputs when parameters are adjusted by ±50%, compared to the baseline set which minimises $\mathcal{D}$. Grey tiles indicate NAs, where parameters are outwith realistic ranges ($Q_N^{\min}$ cannot be greater than $Q_N^{\max}$; $Q_N^{\max}$ cannot be smaller than $Q_N^{\min}$; no valid solution for a reduction in $G_a^{\max}$ and an increase in $G_{b1}^{\max}$).

literature-retrieved allometric relationships for biological rates ($G^{\max}$, $V_C^{\max}$, and $V_N^{\max}$), as well as the exponent k_{Nb}. In fact, some variations of $G^{\max}$ did not yield a solution, as a degree of precision greater than the machine's limits were required (the model crashed for the upper limit of precision, which is `rtol=1e-15, atol=1e-20, maxsteps=50000`). The fact that this happened is unlikely to be a coincidence: a decrease in $G_{a1}^{\max}$ and an 'increase' in $G_{b1}^{\max}$ (i.e., remaining negative with a greater magnitude) both lead to a sharp discontinuity in $G^{\max}$, whereby $G^{\max}$ at size class for 125 μm was much larger than $G^{\max}$ at 85 μm. Combined with the fact that k_{Z2} is significantly lower than k_{Z1} and that the tradeoff relating to k_Z is also discontinuous, then the competitive advantage at 125 μm compared to 85 μm becomes very large, and likely unrealistic.

There is an underlying assumption when interpreting these sensitivity analyses that the transition between the model fit for a given parameter's original value, and the model fit for a parameter's adjusted value, is somewhat linear. In reality, it is entirely possible that the adjustment of a parameter has little effect over a smaller domain (e.g. over the scale of a 10%) adjustment, but blows up as some point between 10% and 50% adjustment. Similarly, the effect of changing a parameter 10% may be greater than changing it by 50%, since the trajectory of the model may be attracted to different states dependent on the parameters. Tipping points such as bifurcations can occur over small regions in the parameter space, yet the consequences can be large differences in overall model fit. Relationships between parameters are also not explored here which, as seen in Figure 4.2, may also contribute to changes in model outputs.

4.4 Discussion

4.4.1 Choice of fixed model parameter values

We implemented a unimodal trade-off for the maximal photosynthetic rate ($V_C^{\max}$; Marañón *et al.*, 2013), as this qualitatively agrees with the patterns reported by Irwin *et al.* (2006) in the absence of taxonomic resolution. To maintain consistency in methodology and experimental

conditions, we also adopted the estimates of $Q_N^{\min}$ and $V_N^{\max}$ from Marañón *et al.* (2013). Below, we discuss the variability in these parameter values reported across the literature, together with potential implications and considerations.

4.4.1.1 Variability in $Q_N^{\min}$ and $Q_N^{\max}$ values across the literature

The values for the allometric coefficients of $Q_N^{\min}$ and $Q_N^{\max}$ vary widely in the literature. The values of $Q_N^{\min}$ derived from Montagnes & Franklin (2001), converted from mmol N cell^{-1} to mmol N mmol C^{-1} using Menden-Deuer & Lessard (2000) are 1.3× to 10.6× larger than the carbon-converted values derived from Marañón *et al.* (2013). The values between these sources diverge with increasing size. This is a large difference and, although the values from Montagnes & Franklin (2001) are used in Ward *et al.* (2012), more recent works (Ward & Follows, 2016; Sauterey & Ward, 2022) use values much closer to those provided by Marañón *et al.* (2013), supporting our use of these values.

Surprisingly, the carbon-converted allometric exponent for $Q_N^{\max}$ is positive in Marañón *et al.* (2013). This is a relatively substantial departure from the $Q_N^{\max}$ exponent of -0.13 derived from Litchman *et al.* (2007) (carbon converted using Menden-Deuer & Lessard, 2000). This $Q_N^{\max}$ bound does not include the Redfield ratio for cells larger than 5.7 μm. However, more recent work indicates that the reciprocal curve of $Q_N^{\max}$ may be unimodal (Sauterey & Ward, 2022), such that $Q_N^{\max} - Q_N^{\min}$ is small and relatively constant until around 3.5 μm and then blows up quickly due to the positive $Q_N^{\max}$ exponent. This would support a positive exponent of $Q_N^{\max}$ for cells larger than 3.5 μm. Based on the uncertainty presented above, we took a conservative approach and collated values from the literature (Montagnes *et al.*, 1994; Menden-Deuer & Lessard, 2000; Marañón *et al.*, 2013), which suggest the $Q_N^{\max}$ is size independent.

Biologically, it is reasonable that only large, eukaryotic cells can exhibit a wide range of Q_N values, since the proportional mass allocated to 'fixed' apparatus (such as the nucleus or mitochondria) decreases with size, and may be replaced by other apparatus (such as chloroplasts, or lipid stores). Within the smaller cells, much of the biomass < 3.5 μm may be contributed by prokaryotic cells (e.g. *Prochlorococcus* or *Synechococcus*). These cells are more limited in the structure that they can support due to the absence of mitochondria. From a mathematical perspective, the range $Q_N^{\max} - Q_N^{\min}$ has possible connotations for the dynamics of nutrient uptake and photosynthesis. The values of Q_N^{stat} and λ_N, for any given Q_N, will be come increasingly insensitive as $Q_N^{\max} - Q_N^{\min}$ grows. In effect, this may increase the impact of stoichiometry in the model for small cells, and reduce the impact on large cells.

In spite of the range of values for $Q_N^{\min}$ and $Q_N^{\max}$ provided in recent years, parameter sensitivity analysis of this model indicates that increases of $Q_N^{\min}$ and $Q_N^{\max}$ have comparatively small impact on the model. With this in mind, there is little concern about the choice of these values.

4.4.1.2 The use of nutrient assimilation rate for $V_N^{\max}$

In a model which resolves variable stoichiometry, the relative rates of $V_N^{\max}$ and $V_C^{\max}$ are likely to have an effect on overall population dynamics. As such, a natural choice was to use the allometric scaling for $V_N^{\max}$ presented in Marañón *et al.* (2013). However, when converting to units mmol N mmol C d^{-1} (using the carbon-to-volume relationship presented in Marañón *et al.*, 2013), the allometric exponent is surprisingly positive (0.09), even across all bounds of the 95% CI, albeit neglecting variance associated with cellular carbon concentration. This appears at odds with previous findings, where the exponent is -0.27 (Litchman *et al.*, 2007), or -0.06 (Edwards *et al.*, 2012). These older works align with traditional scaling theory: it is expected that the exponent of $V_N^{\max}$ before being multiplied by cellular carbon (i.e., with units mmol N $cell^{-1}$ d^{-1}) will scale approximately with the surface area-to-volume exponent of 2/3 (Litchman *et al.*, 2007; Edwards *et al.*, 2012). While Marañón *et al.* (2013) suggest that larger cells may have greater density of transport cites on their surface (via Aksnes & Cao, 2011), there is still limited evidence to support this claim. It is likely that a change in sign of the allometric scaling of $V_N^{\max}$, from negative to positive, would impact the model behaviour.

The difference in scale between the $V_N^{\max}$ is stark, with values proposed by Litchman *et al.* (2007) 15.77× larger than that of Marañón *et al.* (2013) in the smallest size class of our model, and 0.045× smaller at the largest size class. However, as noted in Marañón *et al.* (2013), not all nutrient uptake by phytoplankton is assimilated. Given that a penalty for inefficiency is not typically applied to nutrient uptake of phytoplankton in models of this style, we used the nutrient assimilation rate as opposed to maximum uptake rate in the model, which has a negative exponent. Nonetheless, the $V_N^{\max}$ values presented in Litchman *et al.* (2007) are still 5.26× larger at the smallest size class, and 0.33× smaller at the largest size class. The tradeoff therefore aligns qualitatively, given the negative exponent, but is far less severe.

4.4.1.3 Allometry of $G_{\max}$

Compared to the literature available for the aforementioned phytoplankton parameters, there is comparatively less to choose from in the maximum ingestion rate of zooplankton ($G^{\max}$). The work of Hansen *et al.* (1997) is often used (Banas, 2011; Ward *et al.*, 2012; Ward & Follows, 2016; Negrete-García *et al.*, 2022; Sala & Chen, 2024), and still forms the basis of newer statistical investigations (Kiørboe & Hirst, 2014). Although an overall exponent of -0.16 is found to be significant across all groups, the pattern appears largely driven by differences by differences in the scaling of proto- and metazooplankton. Indeed, the protozooplankton exponent for $G^{\max}$ is -0.2, and appears driven by the comparatively high ingestion rates of flagellates ($3.4 - 7.5$ μm) compared to ciliates and dinoflagellates ($12 - 74$ μm). Yet, the metazooplankton exponent is not significantly different to zero (indeed, the sample estimate is 0.08). This is reflected in the more recent work of Kiørboe & Hirst (2014), whereby the overall slope across functional groups is not significantly different to zero[3]. Based on the data presented in Hansen *et al.* (1997), and the

[3]Note that some small fish are included in this analysis, all of which smaller than the studied euphausiids, and within the observed size range of chaetognaths. See Figure 3C of Kiørboe & Hirst (2014); these points lie at

fact that small and large zooplankton must already be treated separately in the model to ensure sufficient top-down control, we decided to implement an allometric tradeoff for protozooplankton of the model, and a size-independent $G^{\max}$ for the metazooplankton. To ensure model stability, $G^{\max}$ of metazooplankton was selected as 0.069 (Table 4.2) to ensure continuity with that of the $G^{\max}$ for protozooplankton. Model sensitivity analysis highlighted this as a necessity for the model to avoid crashing, even at the highest level of solver precision. The mean of values using the estimates for $G^{\max}_{a2}$ and $G^{\max}_{b2}$ in Hansen *et al.* (1997) on the model's size grid is 0.067; evidently, this is very similar to the value used in the model.

It is also important to note the extrapolation of $G^{\max}$ beyond the smallest size of organisms observed in Hansen *et al.* (1997). This is necessary when following the established scaling laws for θ^{opt} while modelling phytoplankton as small as 0.5 μm. Without the extrapolation, the smallest phytoplankton cells would be left without sufficient top-down control. The choice to extrapolate $G^{\max}$ and θ^{opt} is therefore a compromise to allow the model to resolve plankton < 1 μm which do make some small biomass contributions.

Due to the fixed number of size classes modelled, the top-down control on cells at either extreme of the modelled phyto- / protozooplankton size domain is less diverse than for the intermediate sizes. At a fundamental level, the smallest and largest phyto- / protozooplankton do not have a full suite of predators to graze them, since predators smaller (or larger) than θ^{opt} are not modelled. If they were modelled, then the predators which 'pad' the edge of the grazing kernel would similarly not have a full suite of available prey. This creates an effective 'chicken and egg' situation and, as such, there is no elegant solution to fix this issue under a modelling framework with discrete size classes. While the normalising constant introduced in Eq. 4.14&4.15 ensures that the summed grazing preference for each predator and prey is equal to 1, the smallest / largest prey have comparatively less diverse prey to graze on. It is possible that this is the reason that the model outputs presented in Figure 4.5a show significant accumulations of biomass at 85 μm. It is possible that this is indeed a local optimum for phytoplankton cell size, but may also be an artefact of the model. Interestingly, the smallest phytoplankton and zooplankton emerge as local optima on the biomass distributions in Banas (2011), where no normalising constant is used. Our implementation of a normalising constant is therefore a closure term on the model, and we acknowledge that it is not empirically or theoretically supported, but rather a necessity to control tradeoffs within the model. Without this closure, the smallest and largest size classes accumulate too much bias and provide a poor fit to observations.

4.4.1.4 Fixed zooplankton stoichiometry

Assuming a fixed zooplankton stoichiometry may be more impactful on the model than initially intended. The justification for setting $\lambda_Z = 0.5$ was that, by assuming fixed zooplankton stoichiometry, excess complexity may be avoided and the scale of the parameter optimisation regime could be maximised. However, parameter sensitivity analysis indicates that the model is sensi-

comparatively higher maximum ingestion rate than similar-sized counterparts, and therefore may influence the overall model fit.

tive to λ_Z. The λ_Z term (and β) effectively moderate $G^{\max}$ to yield the saturating growth rate. Future works which focus on food web dynamics (as opposed to carbon export) may still value the inclusion of this variable stoichiometry.

4.4.2 Scale and allometry of the initial slope of the photosynthesis-irradiance curve (α)

The inclusion of allometry in the formulation of α is rarely seen in size-structured models (e.g. Banas, 2011; Ward *et al.*, 2012; Acevedo-Trejos *et al.*, 2015; Ward & Follows, 2016; Negrete-García *et al.*, 2022; Sala & Chen, 2024). This is surprising given the established effect of intracellular shading (Taguchi, 1976; Edwards *et al.*, 2015); for this reason, it is included in the model of Terseleer *et al.* (2014), where a value $\alpha_b = -0.077$ is used following calibration from a start point of -0.13 (Taguchi, 1976; the later work of Edwards *et al.*, 2015 also presents this value from ordinary least square regression). Their model is a 'plankton functional type' model which uses the moment-closure technique so that compartments of biomass, mean size, and size variance can be modelled directly. An allometric tradeoff for light is also included in the Nitrogen-Unicellular-Multicellular (NUM) model (Serra-Pompei *et al.*, 2020). This is given as -0.33 to scale with surface-area-to-volume ratios. The reduced major axis (RMA) fit of α_b is -0.37 in Edwards *et al.* (2015) which provides some empirical support. However, generally speaking, OLS provides more conservative estimates than RMA, which makes its fitted values a reasonable start point for a parameter value. Finally, a stepwise formulation for α by taxon, broadly reflective of a reduction with increasing size (picoplankton $\rightarrow$ nanoplankton $\rightarrow$ diatom) is used in Butenschön *et al.* (2016). This yields $\alpha_b \approx -0.04$, assuming ESDs of 1 μm, 5 μm, 25 μm for picoplankton, nanoplankton, and diatoms respectively. Clearly, a wide range of values is presented for α_b across the literature.

4.4.2.1 The meaning of α in different photosynthesis formulations

The parameter α appears in the literature with different units, which can be unified with some administrative effort. In presented experimental work (e.g., Edwards *et al.*, 2015), the units of α are derived from the P-I curve, which has response P_C (t^{-1}) and explanatory variable I (μE m^{-2} t^{-1}) for some common time unit t. It is convention that the units of P_C are h^{-1} (e.g., Marañón *et al.*, 2013) or d^{-1} (e.g., Edwards *et al.*, 2015) while irradiance is measured as s^{-1}, so a conversion is made in the background when presenting the P-I curve. It then follows naturally from the analysis of units that the initial slope of photosynthesis with respect to irradiance ($\frac{\partial P_C}{\partial I}|_{I\rightarrow 0}$) would have units μE^{-1} m^2. On the other hand, in regulatory models where intracellular C, N, and Chl a are modelled explicitly, the units of α are μE^{-1} m^2 mmol C (mg chl a)$^{-1}$, consistent with those presented in Table 4.2. Multiplication by Q_{Chl} (mg chl a (mmol C)$^{-1}$) then yields units of μE^{-1} m^2 as above. In our analysis below, let $\alpha^* = \alpha Q_{Chl}$ such that the units of α are μE^{-1} m^2 mmol C (mg chl a)$^{-1}$, and the units of α^* are μE^{-1} m^2. Comparison between values presented for α and α^* may therefore be made, when necessary, either using measurements of C:Chl (often made in experiments measuring α^*), or by assuming some fixed, reasonable value

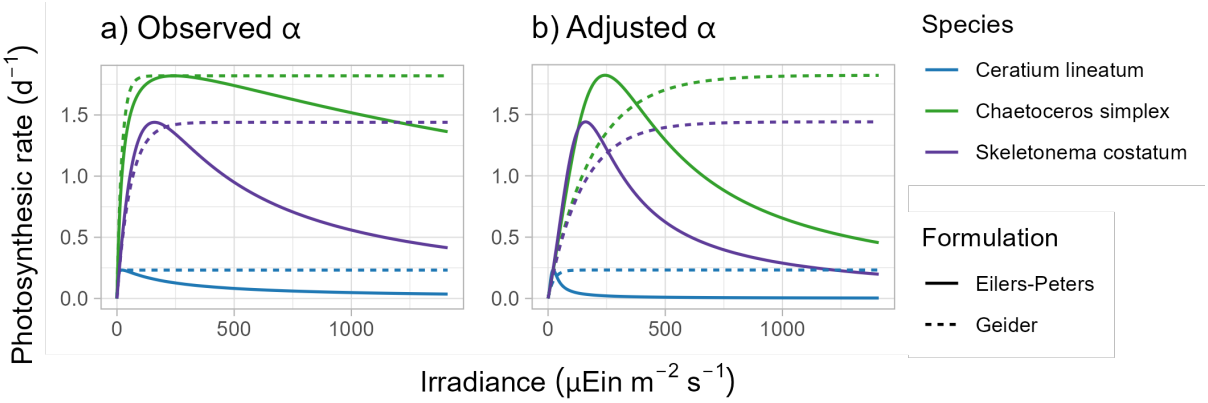


Figure 4.9: P-I curves for three species retrieved by Edwards *et al.* (2015): *Ceratium lineatum* (Chan, 1978); *Chaetoceros simplex* (Thompson *et al.*, 1991); *Skeletonema costatum* (Oh *et al.*, 2008). (a) shows the fit of Geider and Eilers-Peters curves for each species using these data. (b) adjusts the α^* for each curve by assuming $\alpha = 3.83 \times 10^{-7}$ (Ward *et al.*, 2012), and assuming carbon-to-chlorophyll ratio of 40 mg C (mg chl a)$^{-1}$.

for Q_{Chl}. In models such as ours, α^* varies seasonally due to variability in Q_{Chl}.

An important consideration is the difference between two formulations in which α^* is used. If we rewrite Eq. 4.11 omitting depth and the cost of biosynthesis, then:

$$V_C = V_C^{\text{sat}} \cdot \left(1 - \exp\left\{\frac{-\alpha^* I}{V_C^{\text{sat}}}\right\}\right). \tag{4.23}$$

Given its presentation in Geider *et al.* (1997), let us call this the Geider P-I curve. It has the following key properties:

- The initial slope of the curve is α^*, i.e. $\lim_{I\to 0} \frac{\partial V_C}{\partial I} = \alpha^*$
- The curve asymptotically saturates at V_C^{sat}, i.e. $\lim_{I\to\infty} V_C = V_C^{\text{sat}}$
- If we fix V_C^{sat} and I, then V_C is determined entirely by α^*.

The properties of the Geider P-I curve are quite different to those used in Edwards *et al.* (2015), where values for α^* are compiled and presented. This curve is the Eilers-Peters P-I curve:

$$V_C = \frac{V_C^{\text{sat}} I}{\frac{V_C^{\text{sat}}}{\alpha^* I_{\text{opt}}^2} I^2 + \left(1 - 2\frac{V_C^{\text{sat}}}{\alpha^* I_{\text{opt}}}\right) I + \frac{V_C^{\text{sat}}}{\alpha^*}} \tag{4.24}$$

where I_{opt} is the optimal irradiance for photosynthesis. This elucidates the following properties:

- The initial slope of the curve is α^*, i.e. $\lim_{I\to 0} \frac{\partial V_C}{\partial I} = \alpha^*$
- The curve attains its global maximum at V_C^{sat}, i.e. $V_C|_{I=I_{\text{opt}}} = V_C^{\text{sat}}$
  - Only I_{opt} controls the irradiance at which V_C^{sat} is attained
- The curve is strictly decreasing when $I > I_{\text{opt}}$, i.e., $\frac{\partial V_C}{\partial I} < 0$
  - The curve is asymptotically zero, i.e. $\lim_{I\to\infty} V_C = 0$

- If we fix V_C^{sat} and I, then V_C is determined by both α^* and I_{opt}.
  - α^* primarily controls the curve as $I \to 0$
  - I_{opt} primarily controls the curve as $I \to I_{\text{opt}}$

There is good reason for the differences between these curves. The Eilers-Peters curve is fitted to individual species, whereas the Geider curve represents the entire community. It is therefore an implicit assumption that species are selected along a gradient of I, initially perhaps by their specific α^* when I is low, but then increasingly as a direct result of their ability to cope under the incident irradiance (i.e., the proximity of I to their specific I_{opt}). The existence of I_{opt} over almost three orders of magnitude (Edwards *et al.*, 2015) indicates that this may indeed be true. A similar theoretical basis underpins the shape of the temperature response (Eq. 4.9), given that specific thermal response curves are typically exhibit inhibition above some optimal temperature (Thomas *et al.*, 2012).

These differences aside, the fact that α^* has sole control trajectory of the Geider P-I curve means that it has much greater overall importance, compared to its use in the Eilers-Peters P-I curve, where it is intended that I_{opt} controls the photosynthetic response to intermediate levels of irradiance. As such, the question is: are values for α^* derived for Eilers-Peters curves suitable for community-level Geider curves?

In Figure 4.9a we observe that the Eilers-Peters and Geider curves broadly overlap for parameter values presented in Edwards *et al.* (2015)[4]. It is only when $I \gg I_{\text{opt}}$ do the curves significantly diverge. Yet, the reduction to $\alpha^* = 1.15 \times 10^{-7}$ by setting $\alpha = 3.83 \times 10^{-7}$ (Ward *et al.*, 2012) and assuming a carbon-to-chlorophyll ratio of 40 mg C (mg chl a)$^{-1}$ results in significant increase in the distance between curves (Figure 4.9b). Although comparison of Figures 4.9a and 4.9b also shows the effect of α^* on the Eilers-Peters curve, the attainment of V_C^{sat} is still purely a function of I_{opt}, despite the observation that α does exhibit control over intermediate rangers of the interval $(0, I_{\text{opt}})$.

A further consideration is that measuring V_C becomes increasingly difficult for small I. This is because respiration dominates net growth when V_C is small, meaning that growth rates may be negative, a fixed respiration must be assumed, or both (Edwards *et al.*, 2015). The measurement of α^* requires I to be arbitrarily small for the Eilers-Peters curve, otherwise the impact of I_{opt} may become increasingly substantial. Yet, this means that the overall trajectory of the Eilers-Peters curve is somewhat robust to misspecification of α, whereas the Geider curve exhibits substantial global sensitivity. This means that any measurement uncertainty in α derived from an Eilers-Peters curve blows up when applied to the Geider curve.

4.4.2.2 Comparison of α values with Ward *et al.* (2012)

The fact that discrepancies exist between the Geider and Eilers-Peters P-I curves was one of the central motivations for a systematic parameter optimization approach. As a result of the ABC

[4]For *Skeletonema costatum*, $\alpha^* = 2.92 \times 10^{-7}$ (Oh *et al.*, 2008); *Chaetoceros simplex*, $\alpha^* = 1.03 \times 10^{-6}$ (Thompson *et al.*, 1991); *Ceratium lineatum*, $\alpha^* = 1.40 \times 10^{-6}$ (Chan, 1978)

regime, we found values for α to be smaller than the size-independent value of 3.83×10^{-7} (Ward *et al.*, 2012). The best-fitting parameter set found that at 0.5 μm this was 0.51× their value, reducing to 0.056× their value, for 85 μm. Qualitatively, this aligns with expectations for three reasons:

- Oceanic isolates have a value of α^* that is 1.8× higher than coastal isolates, on average (Edwards *et al.*, 2015).
- Given that C:Chl in coastal regions is typically lower than in oceanic regions, an associated increase in Q_{Chl} requires a decrease in α to maintain an equivalent α^*
- On average, the I_{opt} of coastal species is $\sim 1.5\times$ greater than oceanic species (Edwards *et al.*, 2015). This induces a rightwards-shift in the Eilers-Peters curve, which − one may speculate − could induce an overall rightward shift in the Geider curve, necessitating a reduction in α.

That said, the presented values of α here are sufficiently small such that cells rarely attain V_C^{sat}. It is therefore possible that the parametrised value of α_a is small in order to moderate V_C, if V_C^{sat} itself is too large. There are two possible mechanisms for this: $V_C^{\max}$ is too large, or λ_N is too large. The former would likely arise from a mismatch between taxa measured in Marañón *et al.* (2013) compared to those extant at PL, while the latter would likely result from insufficient nutrient depletion, which is acknowledged as a relative shortcoming of the model.

By estimating the joint posterior of the free parameters, we know that α_a and α_b balance changes amongst themselves, which minimises change in the overall community mean α. Small values of α are therefore not by chance, but are a necessary constraint for the fit of the model under the given conditions. The idea that α is scaled down to combat excessive growth rates would, however, appear to contradict the finding that coastal species typically have higher growth rates than oceanic species (Edwards *et al.*, 2015). Yet, even if α_a is reduced in compensation for $V_C^{\max}$, the allometric tradeoffs α_b should remain largely unchanged for fixed exponents $V_{Cb1}^{\max}$ and $V_{Cb2}^{\max}$

4.4.2.3 Posterior estimate of α_b

A key inclusion of the model is the allometric exponent α_b. A negative α_b, sufficiently different from zero, was required to pass qualitative checks for the model. In the best-fitting model run, $\alpha_b = -0.136$, while a 95% Bayesian credible interval for the posterior estimate of α_b was $(-0.203, -0.072)$. The uniform prior density across $(-0.33, 0)$ means that no preferential weighting was given to any values, yet the value of α_b selected via the ABC regime was not significantly different to that OLS fit of α_b^* in Edwards *et al.* (2015). Indeed, the presented value of -0.13 is extremely similar, with 95% confidence interval $(-0.17, -0.09)$. If $Q_{Chl,i}$ is assumed to be approximately size-independent since $\bar{Q}_{Chl,i} \in (0.40, 0.46)$ for all size classes i, then $\alpha_b^* \approx \alpha_b$. And, although caution is required when comparing confidence intervals and Bayesian credible intervals, the fact that our prior is uninformative and the posterior is approximately normally distributed, it is reasonable to compare these values as an informal check (even if inference differs slightly). Given that size is a fixed trait in our model, it also makes most sense to compare with

OLS estimates rather than RMA estimates.

The calibrated value of $\alpha_b^* = -0.077$ (Terseleer *et al.*, 2014) also lies within our computed Bayesian credible interval. The agreement between this point estimate and our credible interval suggests convergence on a similar ecologically plausible parameter range. Combined with supporting observations of Edwards *et al.* (2015), we argue that there is a very strong case to support the inclusion of an allometric tradeoff for the initial slope of the P-I curve. Its value, however, still requires further investigation. Due to the mismatch between the Geider and Eilers-Peters P-I curves, empirical estimates of α (via α^*) should be treated with caution. And, while the ABC method used here provides a strong foundation to investigate the value, a suite of models should still be tested with this allometric tradeoff to remove any estimation bias stemming from any single model's limitations.

4.4.3 Quadratic mortality

4.4.3.1 Omission of the Record *et al.* (2014) formulation

Here, we use quadratic mortality of the form $m_{Z2} \cdot Z_j^2$ as a closure term for the metazooplankton (i.e., zooplankton $125 - 4000$ μm). This is somewhat crude: it assumes implicitly that grazing pressure on those zooplankton is only a function of their own biomass, and is independent of the overall size of the community. This contrasts to the grazing which the zooplankton undertake themselves as predators in the model. Even if the optimal prey of a zooplankton size class j has limited biomass, it may feed on prey of other sizes to supplement its diet, with preference defined via the grazing kernel ϕ. It is therefore not only dependent on the biomass of its optimal prey, as a crude quadratic mortality may suggest.

An adjusted formulation for the quadratic mortality is presented by Record *et al.* (2014):

$$m_{Z2} \cdot \bar{Z}^{1-\nu} \cdot Z_j^{1+\nu} \tag{4.25}$$

and, as discussed around Eq. 1.15, impacts model biodiversity. The case $\nu = 1$ is used in our model; as ν decreases, so too does zooplankton biodiversity in the absence of other effects. Given that, in general, models of this style suffer from loss of diversity via competitive exclusion (Smith *et al.*, 2014), we opted for a conservative approach using $\nu = 1$. However, comparison of model outputs to observations along the size domain (Figure 4.5) indicates that there may be support for a more complex closure term in future editions of the model. The modelled metazooplankton are more diverse than observations, which show clear peaks at ~ 400 μm and ~ 1000 μm, varying seasonally. The model, however, produces a bell-shaped curve centred at 590 μm. While it is important to remain cognisant that 'observed sizes' are indeed fixed and likely under-represent diversity, this qualitative pattern is very unlikely to be an artefact of the fixed size allocation.

A further consideration is that, with limited existing evidence, ν would be a free parameter of the model. The advantage of the classical formulation (i.e. $\nu = 1$) is that it reduces the number of free parameters of the model, since ν does not require parameterisation. Future versions of the model may experiment with the Record *et al.* (2014) formulation, while treating ν as a free

parameter. A potential improvement may be to compute $|Z|$ on a local scale rather than global scale. A window of adjacent size classes may therefore be used to construct such an effect, which would be more analogous to the kernel used for grazing on phyto- / protozooplankton in the model.

4.4.3.2 Where should closure term losses go?

In ecosystem models which are mass-conservative, any losses must be deposited as gains in another compartment to ensure the sum across all differential equations is zero. The advantage of this is that analysis of the model at / near steady states is facilitated. However, exchange with the deep layer in our model violates mass conservation. As such, there is little practical need to force the model to be mass-conservative, when this condition is already violated.

In the presented model, losses from quadratic mortality are maintained to the system via transfer to detritus, given that grazing efficiency is generally less than 50% (i.e., $\beta\lambda_Z = 0.35$ for zooplankton grazing). Under this assumption, the majority of grazing is returned to the system as DOM (D_2) or POM (D_1); or simply, they are added to POM. From model outputs, the D_1 pool is comparatively quite small (ranging $2.9 - 8.3\%$) of total detritus ($D_1 + D_2$) in the model. Such a result aligns quantitatively with integrated masses of POM and DOM in the upper 200 m of the global oceans; it is estimated that DOM stocks constitute 47 Pg C (Hansell *et al.*, 2009), while DOM contributes 2.3 Pg C (Stramska, 2009). Thus, DOM contributes about 4.7% of the total detritus pool in the upper layers of the oceans globally.

Although the remineralisation rate of D_1 is double that of D_2, it still only adds approximately $6 - 16\%$ of total remineralisation in the model. This is likely due to its sinking rate of 10 m d^{-1} which is $40 - 83\%$ loss of mass per day, depending on the MLD. The pool of D_2 is therefore far more influential in nutrient dynamics than D_1; variability in mixed layer depth and cross-thermocline mixing also contribute significantly to nutrient input. Nonetheless, an alternative viable solution for quadratic losses could to split between D_1 and D_2 in proportion to β and λ_Z, with the remainder exported from the model.

4.4.4 Assessment of ABC approach & viable alternatives

Plankton ecosystem models are parameterised in a range of different ways. The start point for most approaches is to identify plausible parameters from the literature and/or experience, identify any reasons for uncertainty, and subsequently adjust parameters. For models with large parameter spaces, such as the one presented here, an exploration of every individual parameter would come at huge expense. As such, it is practical to choose a subset of parameters to optimise.

4.4.4.1 Selection of free parameters

The choice of free parameters followed a minimalist approach. Only when it was deemed essential were parameters selected for optimisation. Instead, value was placed on gaining a strong understanding of a subspace of the overall joint parameter space so that robust inference could

be made on allometric tradeoffs (e.g., α_b, and the difference between k_{Z1} and k_{Z2}).

This, however, provided some limitations. By optimising only 6 parameters, we speculate that the estimated posterior is confounded by effects from other parameters. For example, it appears that the estimate of α_a is affected by $V_{Ca}^{\max}$. Inclusion of parameters relating to the intercept of maximum photosynthetic rate, nutrient uptake, and grazing should therefore be included in future works where possible. Given that these parameters are already centred on a plausible value (based on observations), their prior distributions could be more informative, leading to a reduction in proposed poor-fitting parameters and a comparable increase in efficiency. There is no evidence that the allometric exponents for these relationships need to be included in future optimisation, although it should be noted that if computational expense and time were not limiting, the inclusion of every parameter would be worthwhile.

4.4.4.2 Algorithm selection

Initially, the parameter space was explored via some manual tuning. It is useful, at an early stage, for any modeller to understand the impact of changes to individual parameters. However, varying parameters manually and one at a time is not only exhaustive, but reduces the possibility to understand interactive effects in a complex parameter space such as ours. Inference that can be made about parameters is also extremely limited, given the lack of a formal statistical framework, e.g. via the presentation of credible intervals.

ABC was chosen as a second choice, after initial runs of Metropolis-Hastings (MH) converged to comparatively high likelihood regions despite the fact they did not satisfy qualitative constraints. Although not clear at the time, one potential reason for this is heteroscedasticity in the detrended log-nutrient values. Given that we subsequently established that the model struggles to reproduce the full extent of the seasonal variability of N, then the MH chain would be attracted to the winter nutrient observations which exhibit lower variability, thus never able to reproduce nutrient depletion. This was ignored because it was mistakenly assumed that the model would eventually find a region on the parameter space which would sufficiently deplete nutrients in summer, while allowing for sufficient replenishment in winter. The fact that the model was unable to reproduce the nutrient cycle may also be a result of the limited free parameter space.

Although Bayesian techniques are not widespread in the optimisation of plankton ecosystem models, there are some examples in the literature. The end-to-end ecosystem model StrathE2E implements MH for optimisation of the parameter space (Heath *et al.*, 2021). An example of ABC-style parameterisation can be found in Sloughter *et al.* (2019), where parameters are drawn from a uniform prior and model fit subsequently evaluated. The work of Sloughter *et al.* (2019) follows a Bayesian line of thinking, despite not strictly identifying as a Bayesian technique. Interestingly, the key inference from this study was that α^* varied temporally in polar phytoplankton communities.

Non-Bayesian techniques also exist to provide estimates of optimal parameter values and their associated errors. The method of Fennel *et al.* (2001) provides an assessment of the parameters

in the absence of any random sampling, and instead assesses the model fit along gradients of the parameters to identify the global minimum of the cost function. The errors of the posterior parameter estimates are then estimated by the inverse of the Hessian matrix, which can be used as an approximation of the covariance matrix (Thacker, 1989). The main challenge of this method is finding the *global* minimum in the cost function across the parameter space, since many local minima may exist. The initial values are therefore crucial, and it is therefore common to use a range of initial parameter sets (Fennel *et al.*, 2001). If initial guesses for parameters are well chosen, then computational expense is drastically reduced compared to Monte Carlo-based methods.

Computational expense can also be reduced through the interpolation afforded by polynomial chaos expansion (PCE), as in Mattern *et al.* (2012). Like Sloughter *et al.* (2019), this paper also assesses the time-dependence of parameters in the context of a deterministic plankton model. However, the timescale was more finely resolved, drastically increasing the effective number of parameters for optimisation if the parameter values at each time step were treated as independent. As a result, the Monte Carlo-based approached used in Sloughter *et al.* (2019), or indeed this thesis, would be too computationally expensive. As such, the model, which is optimised for two parameters, is run only 49 times. The model fit is then emulated by interpolating between the model runs. In short, the regime ultimately leads to improved model fit by providing an estimated optimum parameter value at each time step of the model output (see Mattern *et al.*, 2012 for full details), alongside a full estimate of the joint posterior parameter space. The amount of inference made in relation to the computational expense is therefore exceptionally high when dealing with time-varying parameters on a fine scale.

Ultimately, the end goal of the optimisation is to understand the parameter space and its uncertainty, and Bayesian methods provide a robust, algorithmic way to do so. Thus, based on the findings above, we suggest that ABC should be considered as a method for optimising plankton ecosystem models where computational power allows, but other methods such as PCE may be more suited to larger parameter spaces (in particular, time-varying parameters). In any case, the publication of credible intervals and relationships between parameters allows for greater cross-model validation, and should help to unify results in the field.

4.5 Conclusion

We constructed a model which is an appropriate and useful tool to test the mechanistic hypotheses outlined in Chapter 2. Broadly speaking, the cycles of phytoplankton biomass, zooplankton biomass, nutrients, and phytoplankton mean size are accurately reproduced by the model. The specific organism sizes accumulating biomass are also reproduced to a reasonable level of accuracy. However, the model does not resolve the spring pulse in zooplankton mean size, which is partially driven by immigrating meroplankton at PL. These organisms deviate from the typical scaling of the optimal predator-prey size ratio.

The model fit was optimised at PL with respective to 6 free parameters. A key finding is

that the inclusion of an allometric tradeoff for the initial slope of the photosynthesis response is crucial to attain good model fit at PL, despite the fact it is rarely used in similar studies. We also found that the optimised parameters displayed relationships, including between the affinities of photosynthesis and grazing. Relationships between parameters spanning different trophic levels may indicate the innate coupling between processes at different trophic levels in nature. The findings from the ABC optimisation, combined with values from observations and previous modelling work, create a statistically robust framework which we recommend is used more widely in future research, in place of 'informal' calibration techniques which are typically employed. Future work should focus on optimising the remaining parameters of models of this style to strengthen inference and defend against confounding arising from relationships between parameters. It may also be of interest to experiment with differing initial conditions; while small perturbations to the initial conditions do not lead to diverging model trajectories, the size of the perturbation does blow up in the initial stages of the model run, before oscillating around a fixed distance.

Chapter 5

Assessment of mechanistic hypotheses using the size-structured plankton ecosystem model

5.1 Introduction

In this final chapter we address ecological hypotheses outlined in Chapter 2 by using the mechanistic model constructed in Chapter 4. The advantage of constructing a mechanistic model is that causality can be inferred directly from the building blocks of the model. We therefore provide insight to the seasonal cycle of plankton community size structure by assessing the evolution of size-specific rates, such as grazing, photosynthesis, and nutrient uptake. In some cases, we also test the sensitivity of the model to these processes experimentally, e.g., assessing the outputs of the model when features, such as carnivorous switching, are removed.

5.1.1 Assessing hypotheses

What statistical models lack in causal inference at this scale, they gain in the ability to robustly test hypotheses with a p-value based approach, or through construction of confidence / credible intervals. In a laboratory environment, causal inference can be reasonably made through controlling variables, while following an algorithm of statistical testing. However, studies at the ecosystem level are almost always observational rather than experimental, and do not possess the control to experiment and subsequently identify causative effects. Examples which do, such as Iron Ex II (Coale *et al.*, 1996), are rare.

Mechanistic modelling is therefore a favourable method to test hypotheses at the ecosystem level, since formulations and parameter values to be changed freely. However, the protocol of testing is perhaps less straightforward than in a statistical context. Statistical studies are often algorithmic, and benefit from having predefined 'reject / fail to reject' criteria. Methodology is therefore transferable from one research field to the next, with minor tweaks often only required,

if any. On the other hand, when using a mechanistic model experimentally, it is difficult to discern how much signal is enough to eliminate a hypothesis.

5.1.1.1 Common approaches in mechanistic modelling studies

Typically, outputs of mechanistic models are presented in studies, and are the central focus for discussion. For example, plankton functional type models (which model phytoplankton mean size and size variance directly) may provide outputs of bottom-up and top-down pressures on mean size (e.g., Terseleer *et al.*, 2014; Acevedo-Trejos *et al.*, 2015). In these works, specific hypotheses are not explicitly stated nor do they provide the central motivation to the work. Instead, the models often aim to give a general overview of processes, and also contribute to methodological developments. Thus, in many senses, it could be argued that works of this style contribute discussion to hypotheses, but do not necessarily 'test' them in an explicit sense.

This differs to our motivation, which is driven by the derivation of explicit hypotheses. Examples of such hypothesis-centred approaches exist in the plankton modelling literature. For example, the addition of mixotrophy (Ward & Follows, 2016) to an existing model (Ward *et al.*, 2012) provides a robust testing methodology, with direct comparison to an established baseline. In Ward & Follows (2016), the presentation of different weighting parameters, and experimentation with different formulations (penalty on maximum rates vs. combined penalty on affinities and maximum rates) adds robustness, and provides value to the overall discussion. Similarly, Banas (2011) test the effect of complex herbivorous trophic interactions on the plankton community by implementing a grazing kernel, and subsequently testing different values of σ.

An alternative approach is to simplify the set of equations describing the model, to allow analysis of the equations at steady state. For example, in Ward *et al.* (2014) an existing model (Ward *et al.*, 2012) is simplified so that food web interactions are simple (one Z_j grazes on one P_i) and variable stoichiometry is removed. These outputs can then be compared to the model outputs to assess the value given by the complexity, alongside further understanding the role each mechanism plays in controlling model dynamics through direct analysis of the equations at steady state.

5.1.1.2 Strong inference

While the approaches discussed above provide useful insights into ecosystem functioning, a more formal framework for testing and subsequently eliminating hypotheses is provided by the concept of 'strong inference' (Platt, 1964). Here, the following algorithm is iterated recursively:

> "(1) devising alternative hypotheses;
>
> (2) devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
>
> (3) carrying out the experiment so as to get a clean result; and
>
> (1') recycling the procedure, making subhypotheses or sequential hypotheses to re-

fine the possibilities that remain; and so on."

A particular strength of the strong inference method is its robustness through the minimisation of confounding. For example, the formulation for photosynthetic rate (V_C) is sensitive to light and temperature simultaneously. An increase in temperature may therefore have an equivalent effect to a decrease in light, or vice versa. Yet, an isolated hypothesis that growth is only primarily temperature driven could, in theory, erroneously assign the effect of temperature to one that is actually controlled by one or more competing mechanisms. In this instance, testing hypotheses about both temperature and light eliminates this possibility.

5.1.1.3 Finding a balance

Although disputed by some (e.g., Scheffer & Rinaldi, 2000), strong inference provides an idealised framework for investigating hypotheses. In practice, however, there may be challenges associated with implementing this approach, which may explain why it is rarely used (if at all) in the plankton modelling literature.

Firstly, devising tests which have the intended cause-effect relationship can be challenging. Take H2 for example: "phytoplankton mean size is limited in winter by: (i) intracellular shading, affecting low-light photosynthetic performance of large cells, or (ii) top-down pressures from micro- and mesozooplankton, which are near their annual maximum mean size". Assessing the effect of intracellular shading in the case of H2i may be simpler, at a theoretical level, than H2ii. To investigate H2i, a reasonable method may centre around experimenting with the model when $\alpha_b = 0$. Yet, experimenting with top-down control through larger mean size of micro- and mesozooplankton is harder. As a modeller, how does one force this to happen?

A further issue is that, as demonstrated in Chapter 4, parameters may often exhibit relationships which explicitly indicate the conditionality of one effect on another. As such, a 'plug and play' style approach may be misleading without the necessary supporting understanding of the parameter space. For example, when we assess the effect of switching $\alpha_b = 0$ (in line with H2i), we consider various values of α_a in the knowledge that α_a and α_b are related (Figure 4.2). However, undoubtedly, other relationships (such as that with k_{Z2}; Table 4.5) may not be accounted for. And, in any case, these relationships exist in the posterior which itself is defined as the region of good fit, and so extrapolation of these parameters beyond this region is uncertain.

As such, we take a hybrid approach which does not follow strong inference, but has its ethos centred in the idea of assessing as many hypotheses as feasibly supported by observations and theory.

5.1.1.4 Further model assessment

Continual reassessment of hypotheses also helps to better understand the model, and critically assess its fit. In Chapter 4, we identified areas of good fit and poor fit of the model. Some of the analysis was limited by a focus on outputs, as opposed to a focus on inputs and thus a more mechanistic approach. Here, while testing hypotheses during periods of good model fit, we are

also able to assess and critique the mechanisms during, and leading to, poor model fit. Such an approach aligns with the overall ethos of the hypothesis testing strategy, as it explores potential mechanisms which are over- or understated, presents further research questions, and contributes to the improvement of the model in future work.

5.1.2 Aims and objectives

This chapter has two aims:

1. To assess the evidence and subsequently provide discussion around hypotheses H1 to H6
2. To use the hypothesis testing process to further assess the model, and thus to suggest future research directions

5.2 Methods

The model was run at the standard precision level (`maxsteps=10000, rtol=1e-8, atol=1e-11`) to maximise accuracy, based on the findings of Figure 4.3. Unless otherwise stated, the best-fitting parameter set (which minimises $\mathcal{D}$) was used. Plotting was undertaken using `ggplot` (Wickham *et al.*, 2019) and `ggpubr` (Kassambara, 2016) in R.

5.2.1 Extracting biological rates

5.2.1.1 From the model

Biological rates were extracted from the model to assess the drivers for changes in biomass and size structure. These were extracted from the model at each timestep from the 16^{th} to 20^{th} years of model runs, and subsequently averaged across years. Rates analysed were:

- Photosynthetic rate for each size class i ($V_{C,i}$). This was further decomposed into:
  - Nutrient saturation ($\lambda_{N,i}$), then combined with externally forced $\lambda_T^{(P)}$ to compute saturating photosynthetic rate ($V_{C,i}^{\text{sat}}$)
  - Intracellular chlorophyll a quota ($Q_{Chl,i}$), then used to calculate $\alpha_i^* = \alpha_i \cdot Q_{Chl,i}$
- Phytoplankton nutrient uptake rate for each size class i ($V_{N,i}$)
- Phytoplankton gross loss rate for each size class i,

$$\frac{1}{P_{N,i}} \sum_{j=1}^{25} G_{i,j} + \lambda_T^{(P)} \cdot m_P + \frac{\omega + H'}{H},$$

  decomposed into:
  - size-independent losses $\left(\lambda_T^{(P)} \cdot m_P + (\omega + H') / H\right)$
  - size-dependent losses $\left(\frac{1}{P_{N,i}} \sum_{j=1}^{25} G_{i,j}\right)$

- Zooplankton gross growth rate for each size class j,

$$\frac{\beta \cdot \lambda_Z}{Z_j} \cdot \left(\sum_{i=1}^{15} G_{i,j} + \sum_{k=1}^{15} G_{k,j} \right),$$

  decomposed into:

  - grazing saturation $\left(F_j^2 \Big/ \left(F_j^2 + k_{Z(1\,\text{or}\,2)}^2 \right) \right)$
  - carnivory coefficient $\left(\varphi_j^{(Z)} \right)$

- Zooplankton gross loss rate for each size class k,

$$\begin{cases} \dfrac{1}{Z_k} \displaystyle\sum_{j=1}^{25} G_{k,j} + \lambda_T^{(Z)} \cdot m_Z + \dfrac{\omega + H'}{H}, & k = 1, \ldots, 15, \\ \lambda_T^{(Z)} \cdot (m_{Z2} \cdot Z_k + m_Z) + \dfrac{\omega + H'}{H}, & k = 16, \ldots, 25. \end{cases}$$

5.2.1.2 From observations

We computed the expected grazing saturation and carnivory coefficient from observed biomass distributions at PL, using the allometric scaling of predator-prey interactions (dictated by θ^{opt} and σ). To do so, truncated mean phyto- and zooplankton biomass distributions (Figure 4.5) were used in place of modelled P_i and Z_k biomasses so that grazing saturated and carnivory coefficient could be computed. This allowed us to compare the mechanisms of the model to what may be expected from observations, under theoretical allometric scaling. It was also used to provide insight to the validity of the allometric scaling, and how important deviations from this scaling may be, under the assumption that zooplankton biomass would have a positive association with prey availability, broadly speaking.

5.2.1.3 Rate- and productivity-weighted mean size

By presenting monthly-averaged gross growth and loss rates for each size class, we were able to understand the distribution of tradeoffs for both phyto- and zooplankton. Notably, the phytoplankton gross growth and loss rates followed a unimodal pattern, mirroring the shape of the tradeoff for $V_C^{\max}$. However, the location of the peak gross growth and loss rates changed over the course of the year. It is useful to summarise this information to make direct inference.

Given that the shape of the phytoplankton distribution of gross growth and losses was simple and consistent (unimodally shaped), a reasonable assessment would be to identify the 'optimum' size from each gross growth and growth loss rate distribution. Formally, this is the global maximum of each distribution. This would convey directly the size at which gross growth rate is at its greatest, and the size at which gross loss is at its greatest. However, while the distribution of gross growth and loss rates for phytoplankton tend to be unimodal, they also exhibit some skewness (Figure 5.1). The mean size, weighted by both gross growth and gross loss rates, is therefore a suitable measure as it accounts for changes across the entire size domain, rather than

one specific point on the domain. Thus, even if the location of the 'optimum' size for gross growth or loss is constant across a period, we are still able to detect changes at other regions of the size domain when imbalances arise around the 'optimum' size.

As such, summarised the distribution by assessing the mean size, weighted by gross growth and loss rates. In effect, this was the first distributional moment of each distribution presented in Figure 5.1, allowing us to assess how movements in the allocation of gross gains by size were tracked (or not) by movements in the allocation of gross losses. Synchrony between the mean size weighted by gross growth and by gross losses indicates that the processes of photosynthesis and grazing are coupled, whereas divergence of growth- and loss-weighted mean size indicate that the zooplankton community are not responding to the tradeoffs which favour a region of the phytoplankton size domain, thus creating a loophole facilitating a change in size structure (e.g., Irigoien *et al.*, 2005).

Since rates act as a proportional change on biomass per day, they can be somewhat misleading. For example, a small net growth or loss rate at the dominant size class of 5.7 μm (Figure 4.5) may command substantial changes in biomass. The opposite may be true for less dominant size classes, where large net growth or loss rates may lead to very small changes in biomass. Thus, for completeness, a productivity-weighted mean size is presented for growth and loss rates. This mainly aims to contextualise the rate-weighted mean size, and is computed as above, except first multiplying the rates of each size class by the biomass at that time step (e.g., $V_{C,i} \cdot P_{C,i}$). The rate-weighted mean size is primarily used in analysis, since this offers the most direct reflection of the mechanistic processes affecting the community (e.g., through V_C).

5.2.2 Decomposition of effects on photosynthesis

In cases such as photosynthesis, which is a function of multiple time- and size-dependent inputs driving a single output, effects of each input were decomposed. For example, photosynthetic rate for a single size class i ($V_{C,i}$) is a single value mapped from five time-dependent variables ($\lambda_T, \lambda_{N,i}, Q_{Chl,i}, V_{N,i}$, and I), some of which can be decomposed even further (e.g., I is a function of I_0, k_w, H, and P_{Chl}^{tot}).

We assessed the effects of inputs on V_C when assessing two hypotheses: to assess why V_C reduces in large cells in winter and what mechanisms controlling PAR are most important in this (H2i), and to compare the effects of external forcing at SH & PL in winter, in order to understand why biomass is lower at SH in winter (H1i−ii). In both cases where this was used, two levels of each 'input' variable of interest were selected. To identify the effect of each 'input' variable on the 'output' (V_C in both cases), `lm()` was used in R, in effect producing output for a fixed effects model (with two levels for each 'input'). The advantage of a fixed effects approach was that it was robust to scale, as well as the nonlinearity in the response of V_C. Furthermore, the use of `lm()` allowed both marginal and interactive effects to be easily computed (e.g. since λ_N and Q_{Chl} appear inside the same exponential function for V_C, their effects are likely to be related). Statistical outputs such as test statistics and p-values were meaningless in the context of a factorial experiment on a deterministic model, and so were excluded.

5.2.2.1 Bottom-up limitation on photosynthetic rate in summer

One feature of the model established in Chapter 4 is the reduction in modelled phytoplankton mean size in summer, which differs from observations. We speculated that this was driven by nutrient limitation in line with allometric tradeoffs for k_N and $V_N^{\max}$. However, nutrient uptake and quota affects photosynthesis in three ways:

- A reduction in λ_N reduces V_C^{sat} and subsequently V_C when $I \gg 0$.
- A reduction in Q_{Chl} reduces α^* and, since Q_{Chl} is an increasing function of V_N, it follows that reduced V_N will lead to reduction in Q_{Chl} (since any gross losses of phytoplankton biomass do not change the ratio of Q_{Chl}).
- A penalty for biosynthesis is incurred on V_C, directly proportional to V_N.
  - All other things held equal, V_N increases as λ_N decreases, since $Q_N^{\text{stat}} = 1 - \lambda_N$ increases.

In order to assess the impact of each of the effects on the final value of $V_{C,i}$, a full factorial assessment was made for winter and summer levels of λ_N, Q_{Chl} and V_N. In winter, nutrients are not limiting, and so this provided insight to how values of $V_{C,i}$ were affected by the different resultant impacts of nutrient limitation in summer.

The size class $i = 12$ ($27\,\mu$m) was chosen for analysis as this size class provided significant biomass in both observations and the model (Figure 4.5), and was generally representative of the λ_N, Q_{Chl} and V_N values of larger phytoplankton cells ($> 5.7\,\mu$m), all of which suffered severe bottom-up limitation.

5.2.2.2 Temperature and light effects at SH vs. PL

Two potential mechanisms are hypothesised to drive the reduced winter phytoplankton biomass at SH compared to PL in H1: (i) reduced light at SH, and (ii) reduced temperature at SH. We initially assessed H1i and H1ii by computing $V_{C,i}$ for all size classes i, using each combination of light and temperature forcing from SH & PL. Forcing was averaged across December and January, since this was the period of the observed biomass minimum. The light forcing for each site was given by switching all of the 'inputs' to depth-integrated PAR (I_0, k_w, H, and P_{Chl}^{tot}) at once. On the other hand, the effects of temperature are only driven by T itself in the model, so a change of T between SH & PL level was used. Thus, the overall change in light was initially measured as a combined effect across all factors affecting light, whereas the impact of temperature was directly driven by T. The values of λ_N and Q_{Chl} were used from model outputs at PL, since, under the nutrient saturating conditions exhibited at both sites, these were not limiting.

Since the light forcing is driven by several factors varying by site (I_0, k_w, H, and P_{Chl}^{tot}), we further disentangled the effects of these variables on V_C by assessing all combinations of these variables. Given that V_C is monotonically increasing with I, then one expects that:

- An increase in I_0 increases I at all depths, and thus increases depth-integrated V_C.

- An increase in k_w or P_{Chl}^{tot} reduces I below the surface through attenuation, reducing depth-integrated V_C.
- An increase in H increases the total attenuation, reducing V_C at the bottom of the ML and thus reducing depth-integrated V_C.

While disentangling the effects of the various mechanisms affecting light, we held temperature constant as the SH value (after finding that it has a comparatively minimal effect on photosynthesis compared to light). Since the shape of the distribution of $V_{C,i}$ by size was invariant across different combinations of light and temperature (rather it just scaled in fixed proportion), we streamlined the analyses by averaging $V_{C,i}$ across all size classes, thus yielding $\bar{V}_C$. This was appropriate since the hypothesis focused on biomass, rather than size structure.

5.2.3 Model reruns

5.2.3.1 Removal of the allometric tradeoff for α

To further explore the importance of the negative allometric exponent for α (i.e., $\alpha_b < 0$), a model run is presented with $\alpha_b = 0$. Given the relationship of α_a and α_b (Figure 4.2 & Table 4.5), three values of α_a were explored: 0.35×10^{-7}, 0.45×10^{-7}, and 0.55×10^{-7}. Since the run with $\alpha_b = 0$ did not pass qualitative checks for the ABC model assessment implemented previously, the value of α_a was chosen after visual assessment. The distance between the seasonal cycles of phytoplankton biomass, zooplankton biomass, and nutrients, compared to the original model run (where $\alpha_b = -0.136$) was minimised. By doing so, differences in biomass-related tradeoffs were minimised, so that both models experienced similar nutrient limitation and overall grazing control. The outputs of the run with $\alpha_b = 0$ were compared with that of $\alpha_b = -0.136$ through summaries of size structure and analysis of biomass distributions.

5.2.3.2 Removal of the herbivory / carnivory passive switching

It was also of interest to understand the effect of herbivorous switching on the model. This was done by setting $\varphi_j^{(Z)} = \varphi_j^{(P)} = 0.5$ for all size classes j. The new model was assessed in the same way as the experiment with $\alpha_b = 0$, by comparing it to original model runs. However, unlike the model runs with $\alpha_b = 0$, there was no choice space associated with selecting parameters (whereas the experiment of $\alpha_b = 0$ required assessment of various α_a).

5.3 Results

5.3.1 Initial assessment of gross growth and loss rates

5.3.1.1 Phytoplankton

Monthly-averaged growth and loss rates exhibit seasonal variability in distinct phases (Figure 5.1). Despite phytoplankton gross growth and loss rates being consistently in a unimodal shape, phytoplankton seasonal succession is a result of subtle changes in slopes along the distribution

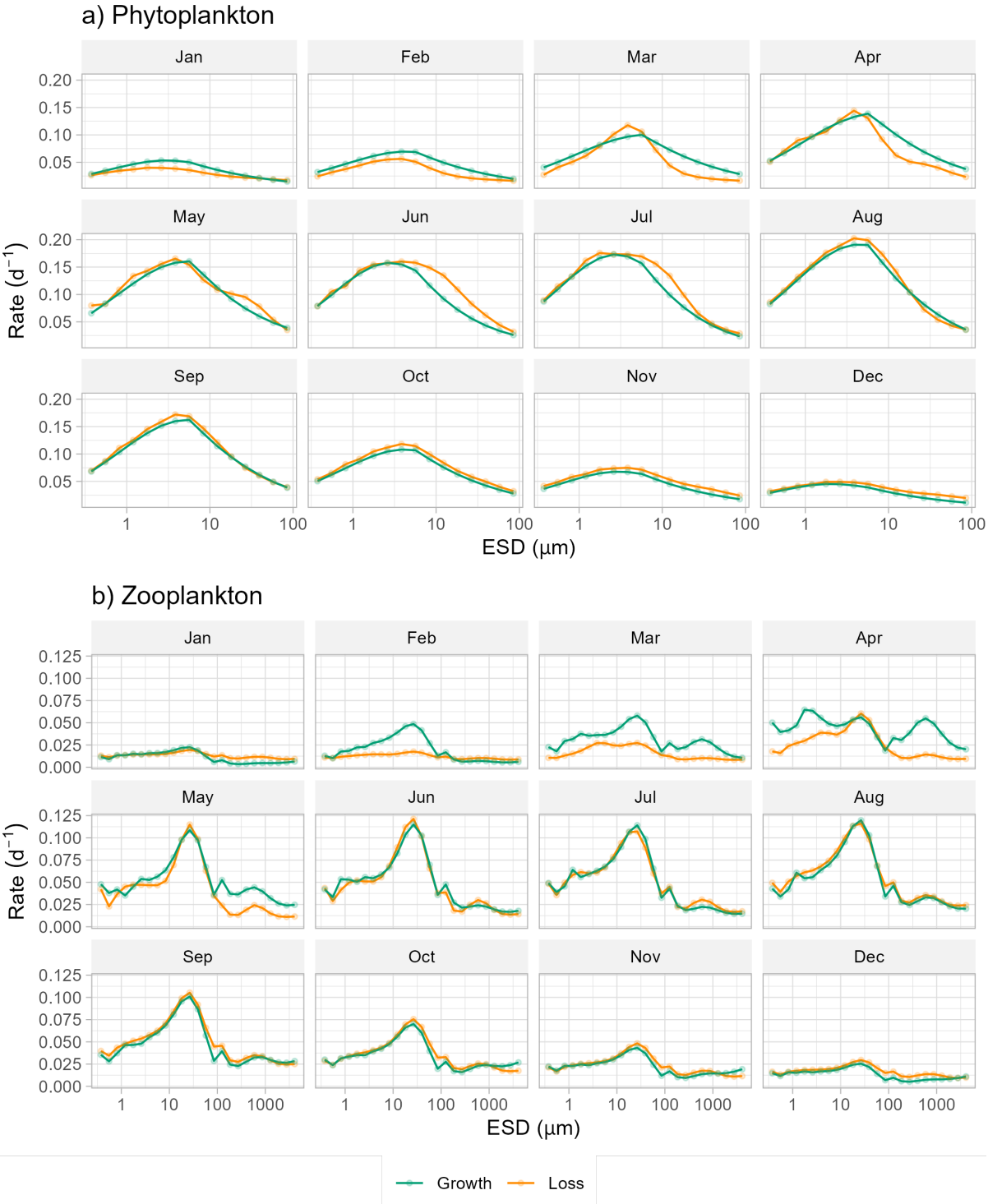


Figure 5.1: Monthly-averaged daily gross growth and loss rates of phytoplankton, and zooplankton of the model using the optimised parameter set. Net growth / loss can be calculated as the difference between gross growth and gross loss.

of growth rates along the size domain, and the position of the peak of the unimodal curves.

Phytoplankton accumulate biomass across the whole size domain in January and February, with the greatest net growth rates (gross gains minus gross losses) initially around $3\,\mu$m, then shifting towards $6\,\mu$m in February. In March & April, gross growth continues to increase across the phytoplankton spectrum, but losses overtake the gains at $3-6\,\mu$m; conversely, for larger phytoplankton (particularly $> 10\,\mu$m), losses remain comparatively low. Then, during May to July, phytoplankton growth and losses $\leq 6\,\mu$m attain balance, while the larger phytoplankton suffer substantial losses due to both decreased gross growth and increased gross losses. In August, the highest gross growth and loss rates are attained, with net loss occurring around $6\,\mu$m and net growth around $30\,\mu$m. This coincides with the deepening of the ML, releasing the nutrient pressure. Thereafter, a seemingly consistent net loss rate is attained across the size domain, with the peak gross growth and loss rates moving to smaller sizes ($\sim 2\,\mu$m). On closer inspection, we note that the net losses are greater with increasing size.

5.3.1.2 Spring bloom

The accumulation of biomass from winter to spring is a contended topic in the literature. The disturbance-recovery hypothesis (DRH) predicts that the bloom begins around midwinter, when net losses turn to net growth through the reduction of top-down control (Behrenfeld & Boss, 2014). Our findings partially agree with this: between December and January, the net loss of phytoplankton indeed changes to net growth. Furthermore, averaged gross loss rate across the phytoplankton size domain reduces from December ($0.037\ \mathrm{d}^{-1}$) to January ($0.029\ \mathrm{d}^{-1}$). However, gross growth also increases from December ($0.031\ \mathrm{d}^{-1}$) to January ($0.037\ \mathrm{d}^{-1}$) as a result of increasing light (Figure E.7), in contrast to the expectations of the DRH. As such, it is a combination of bottom-up and top-down controls which drives the net growth of phytoplankton.

5.3.1.3 Zooplankton

Zooplankton gross growth and losses typically exhibit a greater degree of separation on a proportional scale (shown by gaps between lines in Figure 5.1b), but the gross rates for zooplankton (on the range $0-0.125\,\mathrm{d}^{-1}$) are often lower than for phytoplankton (on the range $0-0.2\,\mathrm{d}^{-1}$). For both phyto- and zooplankton, net growth rates do not exceed $\pm 0.05\,\mathrm{d}^{-1}$. Assessing the growth / loss spectrum of zooplankton is more complex than that of the phytoplankton, since the zooplankton $\leq 85\,\mu$m are grazed by larger zooplankton, which are represented on the same plot of gross growth / loss rates by size.

The first observation of the zooplankton gross growth and loss distribution is that the unimodal shape also emerges around $27\,\mu$m for much of the year, mirroring the fact that the optimal prey for these zooplankton is the $5.7\,\mu$m cells, which are the dominant size class of the model (Figure 4.5). However, the shape of the distribution of gains and losses is not as consistent than for the phytoplankton. In January, the protozooplankton community exhibit small but notable gross growth / loss rates, with the peak at $27\,\mu$m. The metazooplankton have similar loss rates to the protozooplankton, much of which is contributed by background linear mortality

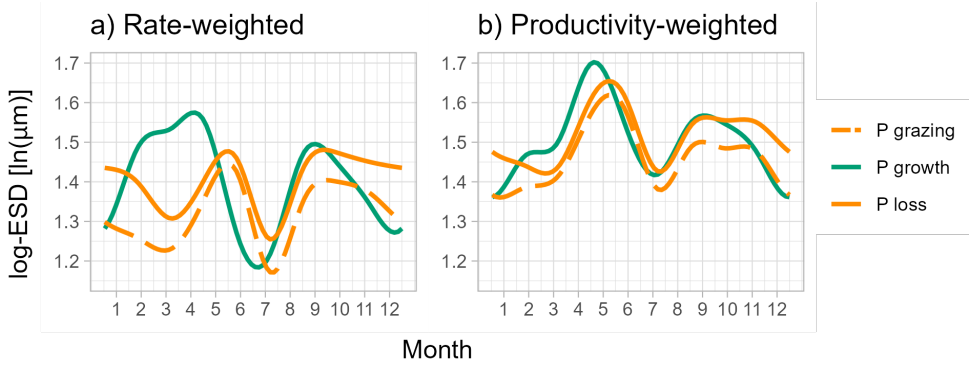


Figure 5.2: Mean size of phytoplankton weighted by the distribution of gross growth, gross loss, and grazing-only loss. In (a) the weightings are by rates only, while in (b) rates are multiplied by biomass to yield a weighting by productivity. The rate-weighted log-ESD is the first distributional moment of Figure 5.1a. The productivity-weighted log-ESD is given by multiplying the rates (d^{-1}; Figure 5.1) by biomass of each size class (mmol C m^{-3}; Figure 4.5).

($\lambda_T \cdot m_Z = 0.006$ d^{-1}), but their growth rates are lower resulting in net losses which continue to February. Conversely, the protozooplankton exhibit substantial net growth across all sizes in February in response to increased prey availability, which is peaked at 27 μm. Protozooplankton gross loss rates increase in March, with much of the gross growth distribution approximately unchanged (except for some increase at 1.8 μm), while the metazooplankton exhibit increasing gains, centred around 589 μm.

By April, gross gains of the zooplankton at 589 μm and gross losses of the 27 μm zooplankton – their optimal prey – increase substantially (Figure 5.1). During this period, $\varphi_{20}^{(Z)} = 0.7$, meaning that 70% of grazing by 589 μm zooplankton is carnivorous. There are also gains amongst the smallest zooplankton. In May, gains around 27 μm surge once again, leading to an approximate bell-shape distribution of gains and losses. Then, the distribution of zooplankton gains and losses vary little from one month to the next: June to August shows little variability in the distribution of gains or losses, indicating that an approximate steady state is reached. This aligns with the minor changes in the biomass distribution observed in Figure 4.5. From September to December, gross growth and loss rates shrink, with the distributional shape becoming decreasingly peaked.

5.3.2 H2: Winter mean size at PL

Phytoplankton mean size is limited in winter by:

i. Intracellular shading, affecting low-light photosynthetic performance of large cells.

ii. Top-down pressure from zooplankton, which are near their annual maximum mean size.

Figure 5.2 exhibits the mean size weighted by rate (d^{-1}) and productivity (mmol C m^{-3} d^{-1}). The rate-weighted mean size (Figure 5.2a) is the first distributional moment of Figure 5.1a, which is approximately equal to the optimised size for gross growth and gross loss under a symmetric distribution. In November to January, it is clear that mean size is decreasing (as

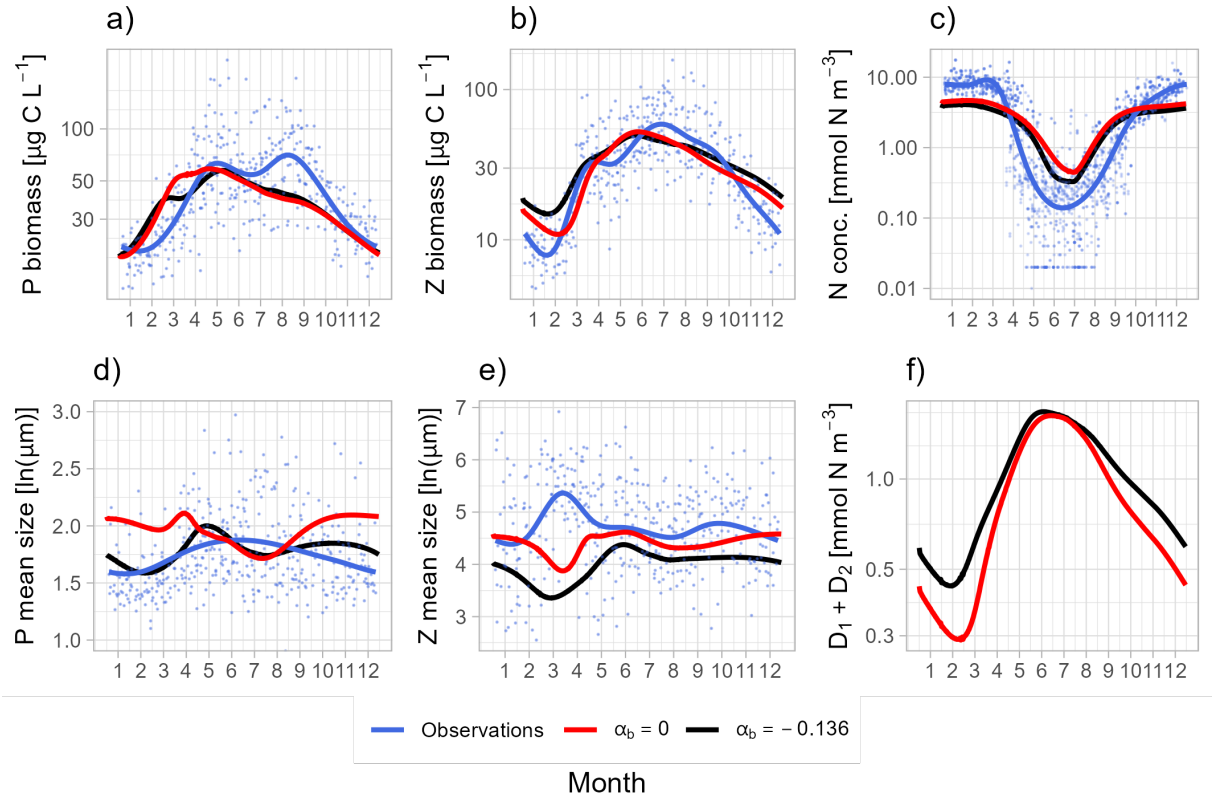


Figure 5.3: Model outputs for the optimal parameter set and an experiment with $\alpha_b = 0$. For the experiment, $\alpha_a = 0.45 \times 10^{-7}$. Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line.

shown in Figure 4.4) due to the difference between loss-weighted mean size and growth-weighted mean size. Interestingly, at this point, the gross growth rate is optimised at near the annual local minimum of size, while the losses are maximised near the annual maximum size. However, grazing-only-weighted mean size also decreases in winter, suggesting that the reduction in the biomass-weighted mean size (E[s]) may not be driven by active top-down effects, but rather by background mortality.

5.3.2.1 Impact of intracellular shading

A key finding of Chapter 4 was the fact that the allometric exponent for α must be negative, in order for the model to attain our definition of 'good fit'. Given that light is at its annual minimum in winter, and α governs the slope of the P-I curve at low light, it is natural to assume that this is the period during which α has the greatest effect of structuring the community. Indeed, removal of the allometric tradeoff for intracellular shading (such that $\alpha_b = 0$) results in an increased winter mean size of phytoplankton (Figure 5.3), providing strong support for H2i. Further investigation is required to fully understand the mechanisms driving this.

During December & January of the optimised model run (with $\alpha_b = -0.136$), the mean Q_{Chl} at the three dominant size bins (1.8 μm, 5.7 μm, 27 μm) is approximately size-independent (Table 5.1). As such, photosynthetic activity is governed by $\alpha^* \propto \alpha$ and $V_C^{\text{sat}} = \lambda_T \lambda_N V_C^{\max}$. Temper-

Table 5.1: Photosynthetic parameters and performance across three dominant phytoplankton size classes, averaged across December and January when light is most limiting. The term $\frac{\alpha^*}{V_C^{\text{sat}}}$ governs photosynthetic saturation, for any given I. The values α, α^*, and $\frac{\alpha^*}{V_C^{\text{sat}}}$ are all given in units of 10^7 for brevity.

ESD (μm)	V_C	V_C^{sat}	$\frac{V_C}{V_C^{\text{sat}}}$	Q_{Chl}	α	α^*	$\frac{\alpha^*}{V_C^{\text{sat}}}$
1.8	0.069	0.648	0.106	0.558	1.038	0.579	0.894
5.7	0.063	1.776	0.035	0.571	0.647	0.369	0.208
27.0	0.032	0.747	0.043	0.559	0.344	0.192	0.258

ature effects (λ_T) are size-independent[1], while λ_N varies only slightly by size in winter, ranging from $\lambda_N = 0.8$ for cells $> 5.7\,\mu$m to $\lambda_{N,1} = 0.95$ for the smallest size class (Figure E.3). As such, in winter, it is a reasonable approximation that λ_N is size-independent, such that $V_C^{\text{sat}} \propto V_C^{\max}$ in a fixed ratio across the size domain.

The fact that that the gross growth rate of phytoplankton (V_C) is similar for $1.8\,\mu$m and $5.7\,\mu$m cells in winter, but lower for $27\,\mu$m cells (Table 5.1), shows the structuring effect of size dependent α. While the $27\,\mu$m cell has higher V_C^{sat} than the $1.8\,\mu$m cell, the V_C attained by the $1.8\,\mu$m cell is more than double that of its $27\,\mu$m counterpart. Conversely, the $1.8\,\mu$m and $5.7\,\mu$m cells exhibit similar growth since the small cell has larger α, but the $5.7\,\mu$m cell has larger V_C^{sat}.

However, it still remains unclear whether the parameter α has an 'active' effect in structuring the community, by driving seasonally varying size-dependent growth variability itself, or whether the effect is a constant underlying effect across the year. Indeed, assessment of the comparative photosynthetic rates of the $5.7\,\mu$m and $27\,\mu$m cell reveals a surprising pattern: the growth rates are in constant proportion throughout the winter, and only change in summer (Figure 5.4b) when λ_N tradeoffs exist (Figure E.3). Further analysis of the P-I curve when λ_N is equal across $5.7\,\mu$m and $27\,\mu$m cells reveals that the fixed proportion of photosynthesis is a direct result of the allometric exponents for α and $V_C^{\max}$. The overall limitation on the P-I curve (i.e., V_C/V_C^{sat}) is given by the expression

$$\left(1 - \exp\left\{-\frac{\alpha^* I}{V_C^{\text{sat}}}\right\}\right).$$

Since I is the incident PAR and thus size-independent, all allometry is explained by the ratio $\frac{\alpha^*}{V_C^{\text{sat}}}$. When λ_N is approximately size-independent, as it is in winter, then the exponent of V_C^{sat} is -0.2 for cells $> 6\,\mu$m, while the exponent of α is -0.136, with 95% credible interval $(-0.203, -0.072)$ (Table 4.4). The size-based tradeoff therefore becomes weak, with allometric exponent for overall limitation $+0.064$. Conversely, for cells $\leq 6\,\mu$m, the exponent of V_C^{sat} is 0.3, thus the limitation exponent becomes -0.436 giving a steep tradeoff of $\frac{\alpha^*}{V_C^{\text{sat}}}$ (Table 5.1). Furthermore, the decreasing ratio of gross gains from December to February for $1.8\,\mu$m cells compared to $5.7\,\mu$m cells shows the transfer of influence from α to V_C^{sat} with increasing light, which starkly contrasts the time-independent ratio expressed by photosynthesis of $27\,\mu$m cells

[1]Nonlinear effects in model trajectory, e.g. due to nutrient depletion, are affected by λ_T. These indirect effects cannot be accounted for here.

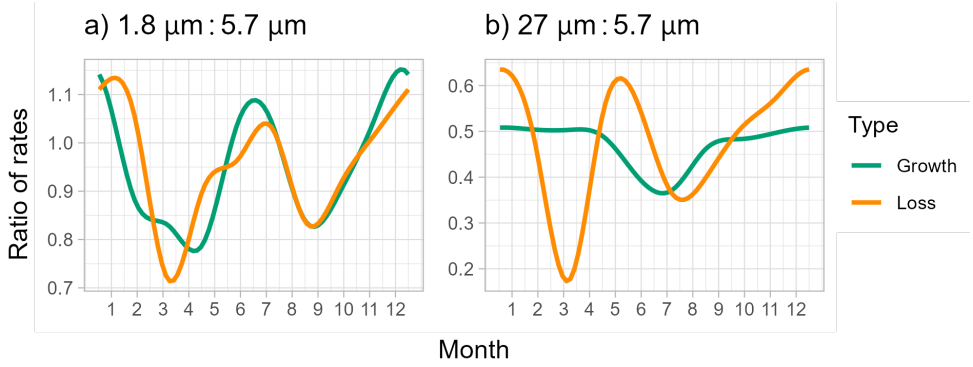


Figure 5.4: Ratios of phytoplankton gross growth and loss rates compared to the dominant size of 5.7 μm, for (a) 1.8 μm cells, and (b) 27 μm cells.

against 5.7 μm cells when nutrients are not limiting, which arises because α and $V_C^{\max}$ exponents effectively cancel eachother out (Figure 5.4b).

This analysis suggest that α_b has an active role in structuring the community $\leq 6\,\mu$m in winter (when I is small, meaning that $V_C \ll V_C^{\text{sat}}$ and α is the dominant parameter of the P-I curve), but to a lesser degree for cells $> 6\,\mu$m. Indeed, this explains the shift of the unimodal peak in growth rates (Figure 5.1a) towards smaller sizes in winter. For larger cells, α does not appear to directly structure the community in winter compared to any other time of year, given the lack of seasonality in the ratio of growth rates between 27 μm and 5.7 μm cells (Figure 5.4). It therefore appears that intracellular shading causes the decrease in mean size on the $0.5 - 6\,\mu$m interval, but does not actively drive the proportional decrease in 27 μm cell biomass compared to the 5.7 μm cells when light is low.

5.3.2.2 Grazing control

Given that net loss rates (i.e., negative net growth rates) are higher for 27 μm cells than 5.7 μm cells in winter (Figure 5.1a), and having established that the ratio of gross growth between 27 μm and 5.7 μm cells is seasonally invariant in the absence of nutrient depletion (Figure 5.4b), it is possible that top-down effects play a role in structuring phytoplankton size in winter. Assessment of rate-weighted gross growth and loss rates (Figure 5.2a) shows an interesting quality: the allocations of gross growth and loss by size are coupled in June to September (effectively moving in parallel, with some small amount of temporal lag), while they are decoupled in October to May. This decoupling has two distinct phases: from October to early February, the growth-weighted mean size decreases (as a result form the bottom-up limitation of photosynthesis shown above), but the loss-weighted mean size shows very little variability, and remains near its annual maximum; then, in February to late May, growth-weighted mean size is parabolic with its global maximum, while the loss-weighted mean size is parabolic with a local minimum. A decoupling of the allocation of growth and loss on the size domain drives changes in size structure.

It perhaps makes sense to start with the phase from June to September, which directly precedes the decoupled phase over winter to understand the landscape before differences arise in the

allocation of growth gross and loss rates. Firstly, in June to July, the growth and loss rate-weighted mean size both decrease. The growth rate-weighted mean size decreases due to nutrient depletion; $\lambda_{N,i} < 0.25$ for all size classes $i \geq 8$ (i.e., cells of size $\geq 5.7\,\mu$m; Figure E.3), leading to the severe reduction in V_C^{sat} (Figure E.7). Loss rate-weighted mean size also decreases, because the phytoplankton biomass reduces at $27\,\mu$m (Figure 4.5), in turn reducing the allocated grazing pressure (Figure 5.5). Conversely, the loss rates of phytoplankton at $1.8\,\mu$m show comparatively less variability (Figure 5.4b) because these cells are able to sustain their biomass under nutrient stress (Figure 4.5). In this phase, the coupling of phytoplankton gains and losses is therefore bottom-up driven by nutrient limitation; larger phytoplankton are unable to grow, and this lack of productivity is passed up the food web to their predators at $589\,\mu$m, signalled by the fact they become increasingly carnivorous in response to prey availability (Figure 5.7).

During the second coupled phase, from August to September, nutrients replenish through deepening of the ML, driving a slight increase in gross growth around $27\,\mu$m (Figure 5.1a). Growth rates of the optimal predator at $589\,\mu$m increase in response and carnivory coefficient reduces slightly from its July maximum (Figure 5.7), indicating that grazing intensity on $27\,\mu$m phytoplankton is indeed increasing. Moreover, size independent daily losses occur from the deepening ML, as well as the proportionally high cross-thermocline mixing (resulting from shallow MLD), and increased linear mortality through temperature dependence, meaning that background losses approach $0.02\ \text{d}^{-1}$. These constitute around $12 - 35\%$ of all losses, depending on size (Figure E.4). An increase in the proportion of fixed loss rate weakens the size-dependence of loss rates, and pushes the loss rate-weighted mean log-ESD towards $1.73\,\ln(\mu\text{m})$ (i.e., the unweighted mean log-ESD). The effect is illustrated by the separation of the grazing only and total loss lines in Figure 5.2a, which are almost overlapping in May and June, but separate in July and continue to diverge until January. Indeed, by November and until January, the proportion of losses resulting from grazing on $27\,\mu$m phytoplankton reduces to 62%, which is much lower than the maximum of 91% in May (Figure E.4). The size dependence of the top-down pressure therefore remains important, but appears to relax, with biomass overall being controlled to a greater degree by background losses, which are size-independent. It therefore appears that the maintenance of the loss rate-weighted mean log-ESD at $\sim 1.45\,\ln(\mu\text{m})$ (Figure 5.2a) is due to a relaxation of the steepness of the slope of losses in the winter months of Figure 5.1. Supporting this notion is the fact that the grazing losses track with phytoplankton gross growth rates during this period.

The existence of larger zooplankton ($567\,\mu$m), which do remain present albeit in lower biomass, is then surprising. However, these organisms are still primarily carnivorous over winter (and indeed year-round), with $\varphi^{(Z)} = 0.67$ (Figure 5.7). Consumption of microzooplankton cells at $27\,\mu$m therefore remains the preferred prey source, with two-thirds of grazing focused on these cells.

5.3.2.3 Summary

Overall, the small winter phytoplankton mean size is not driven by grazing (eliminating H2ii), although the existence of background losses has an effect on structuring in winter since it con-

tributes a higher proportion of losses during this period than during productive periods. Conversely, the evidence suggests that the removal of intracellular shading from the model results in large winter mean size. Our finding support H2i, thus we conclude that mean size is bottom-up limited due to the tradeoff arising from the package effect.

5.3.3 H3: Mean size of phytoplankton during bloom

The bloom of small nanoplankton cells at PL is driven by the optimised maximal growth rate since neither nutrients or light are limiting, and results in the competitive exclusion of other cell sizes.

5.3.3.1 Unimodal scaling of $V_C^{\max}$

Modelled phytoplankton biomass accumulates across the size domain during January and February, resulting from positive net growth (Figure 5.1a). The initiation of biomass accumulation in January results from a decrease in gross losses rather than an increase in gross growth, but the procession of the bloom is underpinned by an increase in gross growth. In January, the optimum growth rate is initially attained at $3\,\mu$m, but evolves to $6\,\mu$m by February; arguably, this is the period in which bloom control shifts from top-down to bottom-up. The favouring of smaller sizes initially is a result of the allometric α tradeoff (since differences in Q_{Chl} and λ_N are minimal at this time), but the unimodal peak of $V_C^{\max}$ at $6\,\mu$m becomes dominant with increasing I from February onwards, despite the fact that a λ_N tradeoff emerges in March (Figure E.3) with a log-linear decrease with cell size from $\lambda_N = 0.84$ at $0.5\,\mu$m, to $\lambda_N = 0.63$ at $5.7\,\mu$m (and then approximately constant for larger sizes). This drives a decrease in V_C^{sat} (Figure E.7), but not to a sufficient degree that $5.7\,\mu$m cells lose their status as the optimal cell size for overall photosynthetic rate (V_C). Tradeoffs in λ_N therefore do not appear to drive the size structure during the bloom. Similarly, Q_{Chl} is approximately size-independent until May (Figure E.2). The $V_C^{\max}$ tradeoff therefore drives the bottom-up limitation (and facilitation) of growth during the February and March period.

Despite this, the bloom starts in January and is initially optimised for smaller cells than $V_C^{\max}$ scaling would dictate, due to the allometric α tradeoff. Given that gross growth is larger at $3\,\mu$m in January at the start of the bloom, it is reasonable to assess the importance of α in bloom dynamics. Furthermore, one may anticipate that if $\alpha_b = 0$, then $6\,\mu$m cells may become even more dominant than they already are, since the tradeoff which seemingly acts to their disadvantage in January is lifted.

Surprisingly, an increase in α_b from -0.136 to 0 increases the March mean size of phytoplankton (Figure 5.3) to a value of $2\ln(\mu\text{m})$, which is larger than the optimal log-size of $V_C^{\max}$, which is $1.74\ln(\mu\text{m})$. This arises due to increased biomass accumulations of larger cells both at $27\,\mu$m and larger (Figure E.5). The most logical explanation for this is that it is carried over from the large mean cell size in November to January. This would be supported by the fact that the mean size is decreasing in February when $\alpha_b = 0$, until modelled the biomass accumulation flattens

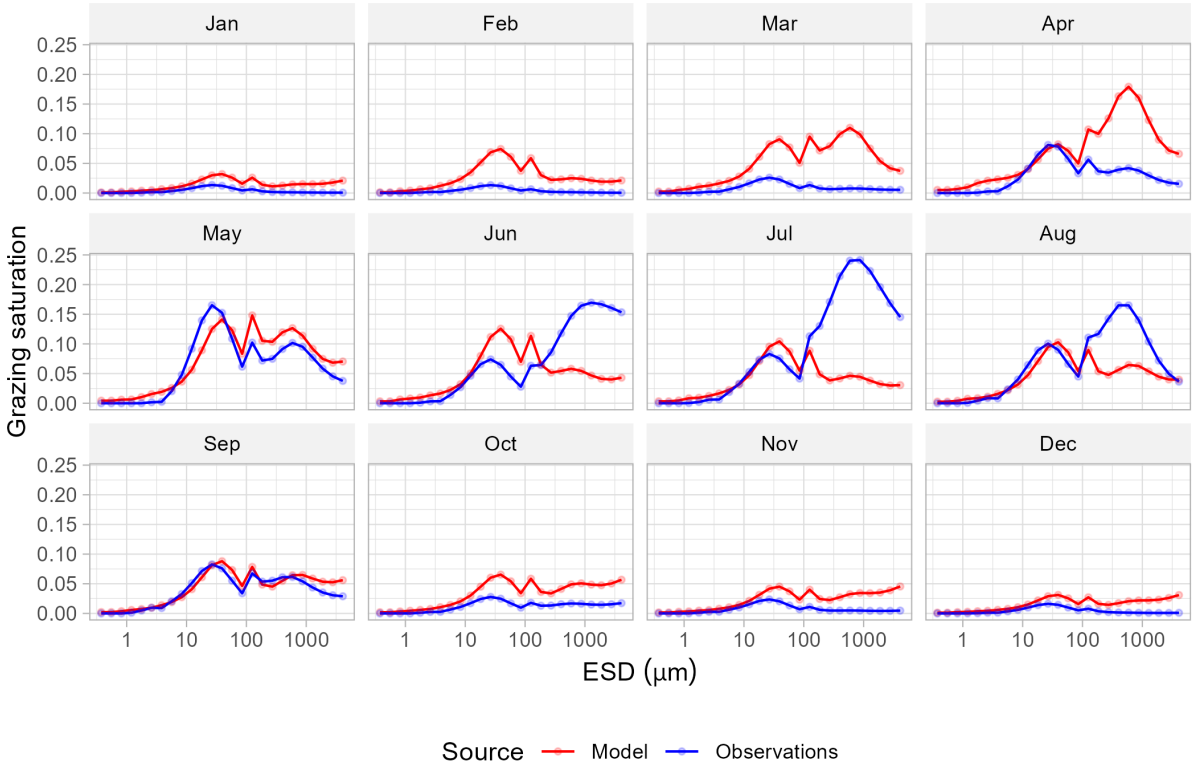


Figure 5.5: Monthly-averaged grazing saturation for each zooplankton size class j of the model, given by $\frac{F_j^2}{k_{Z(1\text{ or }2)}^2+F_j^2}$. 'Model' line is directly sourced from model, while 'Observations' are computed using observed biomasses of phytoplankton and zooplankton. Both cases used the same universal allometric scaling.

off in late March, which indicates that the shift towards the optimum cell size is still occurring when the bloom period ends. Nonetheless, given the lack of any bottom-up advantage of the larger cells compared to the 5.7 μm cells, these findings suggest top-down tradeoffs have a role to play when structuring phytoplankton size in the $\alpha_b = 0$ run.

5.3.3.2 Necessary balance between α and k_Z

It is important to note the relationship between α_b and the zooplankton grazing half-saturation coefficients k_{Z1} and k_{Z2} in the estimated posterior of the free parameter space (Table 4.5). The fact that $\rho(\alpha_b, k_{Z1}) = 0.40$ and $\rho(\alpha_b, k_{Z2}) = -0.28$ shows that, within the posterior, changes in α_b also typically incur changes in grazing half-saturation ratios. As discussed in Section 4.3.1.2, these appear to 'balance' affinity of phytoplankton and zooplankton along the size domain: an increase in α_b favours the large cells, who must then be controlled more tightly (via reduction in k_{Z2}), and vice versa for decrease in α_b. While $\alpha_b = 0$ is beyond the upper bound of α_b in our posterior, the findings in Figure 5.3 support the notion that this relationship between α_b and k_Z may be extrapolated to $\alpha_b = 0$ for the purpose of a theoretical experiment.

The very existence of this tradeoff suggests that top-down control has a role to play in structuring

the spring biomass accumulation of the phytoplankton community. While 5.7 μm is likely to remain the optimum size class during the spring bloom, the extent of competitive exclusion is also controlled by top-down factors. If the grazing control on larger cells was not high enough (i.e. k_{Z2} not sufficiently low) for a given α_b, then large cells would also be allowed to contribute significant biomass, even if not dominant.

The importance of top-down control in the bloom for the optimised parameter set, with $\alpha_b = -0.136$, is further elicited in Figure 5.4. While the 1.8 μm and 5.7 μm cells have approximately similar gross growth and loss rates, the gross loss rates of the 27 μm cells diminish in comparison to the 5.7 μm cells, while the gross growth rates remain constant. The approximate overlap of growth and loss rates of 1.8 μm and 5.7 μm cells indicate that gains and losses of these cells move in relative synchrony, whereas divergence in the ratios for 27 μm cells against 5.7 μm cells creates a size-based tradeoff. Comparatively, this shows that gains in biomass through photosynthesis are not matched by top-down control in 27 μm cells in March.

The apparent lack of this top-down control is a result of the weak grazing pressure inflicted by the optimal predators of the 27 μm phytoplankton. For any predator j, the grazing saturation is given by:

$$\frac{F_j^2}{k_{Z(1 \text{ or } 2)}^2 + F_j^2}$$

with overall productivity proportional to biomass Z_j. This means that phytoplankton losses for some P_i are dependent on not only their own biomass via contributions to F_j, but the biomass of possible predators (determined by θ^{opt}), while phytoplankton gross gains are dependent only on the phytoplankton biomass P_i. As such, a loophole can exist where P_i gross growth increases, resulting directly in increased productivity, but the top-down control is not sufficiently high to match this if the biomass of optimal Z_j is low. In effect, zooplankton biomass − combined with zooplankton gross growth rate − is the driver of phytoplankton loss rate, for any given phytoplankton biomass[2].

Notably, the modelled grazing saturation is low across the board year-round, maximising at 0.18. This indicates two possibilities: firstly, the true response within a complex ecosystem may not be Holling-III and instead is exponential. Secondly, it is possible that the estimates of k_Z compensate for some of the fixed parameters, e.g. $G^{\max}$, on which the posterior estimates are conditional. Crucially, however, the estimates of k_Z are highly erratic in the empirical literature, with values varying by more than two orders of magnitude at some sizes (Hansen *et al.*, 1997). Furthermore, these are values estimated under lab conditions, which are substantially different to natural systems. Future work on the model may therefore aim to investigate grazing saturation further.

Nonetheless, the modelled grazing saturation in March is maximised at 589 μm (Figure 5.5), meaning that prey availability does not contribute to the weak top down control of 27 μm phy-

[2]It is worth noting that this is not the case for top-down control on metazooplankton classes Z_k for $k = 16 \ldots, 25$, since quadratic mortality is only dependent on Z_k, rather than both Z_k and their predator biomass (Z_j) for $k - 1, \ldots, 15$.

toplankton. This leaves a small number of possibilities. Since the exponent for $G^{\max}$ is zero for zooplankton $> 85\,\mu$m (and half-saturation is fixed at k_{Z2}), then maximal grazing rate on phytoplankton is equal for all cells $> 12\,\mu$m (since they have optimal grazers $> 85\,\mu$m). Furthermore, all top-down controls on zooplankton $> 85\,\mu$m are size-independent, meaning that even while zooplankton biomass may be top-down controlled, the size structure of zooplankton is not influenced by top-down factors. Thus, one would expect that phytoplankton growth would be maximised at or near $12\,\mu$m, since α, $V_C^{\max}$, k_N, and $V_N^{\max}$ are all favourable compared to larger cells, such as those at $27\,\mu$m. This is not the case, and so it appears that the emergence of $27\,\mu$m cells is a property of prey selection, rather than allometry of affinity or maximal growth rates of phytoplankton or zooplankton.

5.3.3.3 Why do large-cell biomass accumulations occur at 27 μm?

With a lack of a size-based competitive advantage (either bottom-up or top-down), it is peculiar $27\,\mu$m contributes significant biomass as opposed to adjacent classes at 12, 18, 39, or $57\,\mu$m. However, the most notable characteristic of the $27\,\mu$m size class in the model is that this is the zooplankton size class which primarily grazes the $5.7\,\mu$m cells. Given that $5.7\,\mu$m phytoplankton cells dominate overall community biomass, it is unsurprising that the $27\,\mu$m microzooplankton are key players in the grazer community year-round, and dominate the zooplankton community in March and April. However, this dominance incurs a proportional increase in grazing pressure from above, with the carnivorous switching of the model allocating around 75% of grazing effort to microzooplankton cells per unit of biomass (Figure 5.7). Hence, the grazing intensity on $27\,\mu$m microzooplankton cells is approximately $3\times$ higher than $27\,\mu$m phytoplankton cells. This suggests that switching of grazers in the modelled community has an important role in structuring size of phyto- and zooplankton, even if the switching used in the model is not explicitly size-dependent, like it is in the kill-the-winner formulation used in Vallina *et al.* (2014). In the model, switching intends to reflect the transition from herbivorous species to carnivorous species within each size bin of zooplankton. The inclusion of such an effect is supported by the existence of zooplankton which are primarily herbivorous or carnivorous within each size class.

The effect of switching is made clear when it is removed from the model. It is initially surprising that phytoplankton mean size *increases* in winter and early spring (Figure E.6); however, inspection of biomass distribution (Figure 5.6a) shows that biomass in March is indeed lost at $27\,\mu$m, but instead gained at the smaller sizes of $2.6\,\mu$m and $8.3\,\mu$m, and the larger sizes $58\,\mu$m and $85\,\mu$m. These changes align perfectly with the modelled carnivory coefficient in March (Figure 5.7). During this period, and indeed for much of the year, $\varphi^{(Z)} > 0.5$ for zooplankton $184\,\mu$m to $1884\,\mu$m. These are the predators who graze optimally on cells sized $12\,\mu$m to $44\,\mu$m, all of whom suffer reduced biomass when switching is removed. Conversely, the removal of the passive switching between herbivory and carnivory favours all phytoplankton outwith this interval. The effects on the zooplankton community can also be seen (Figure 5.6b), with smaller zooplankton ($< 12\,\mu$m) suffering from greater top-down control when the passive switching is removed.

An estimate of zooplankton carnivory based on observed biomasses of phytoplankton and zoo-

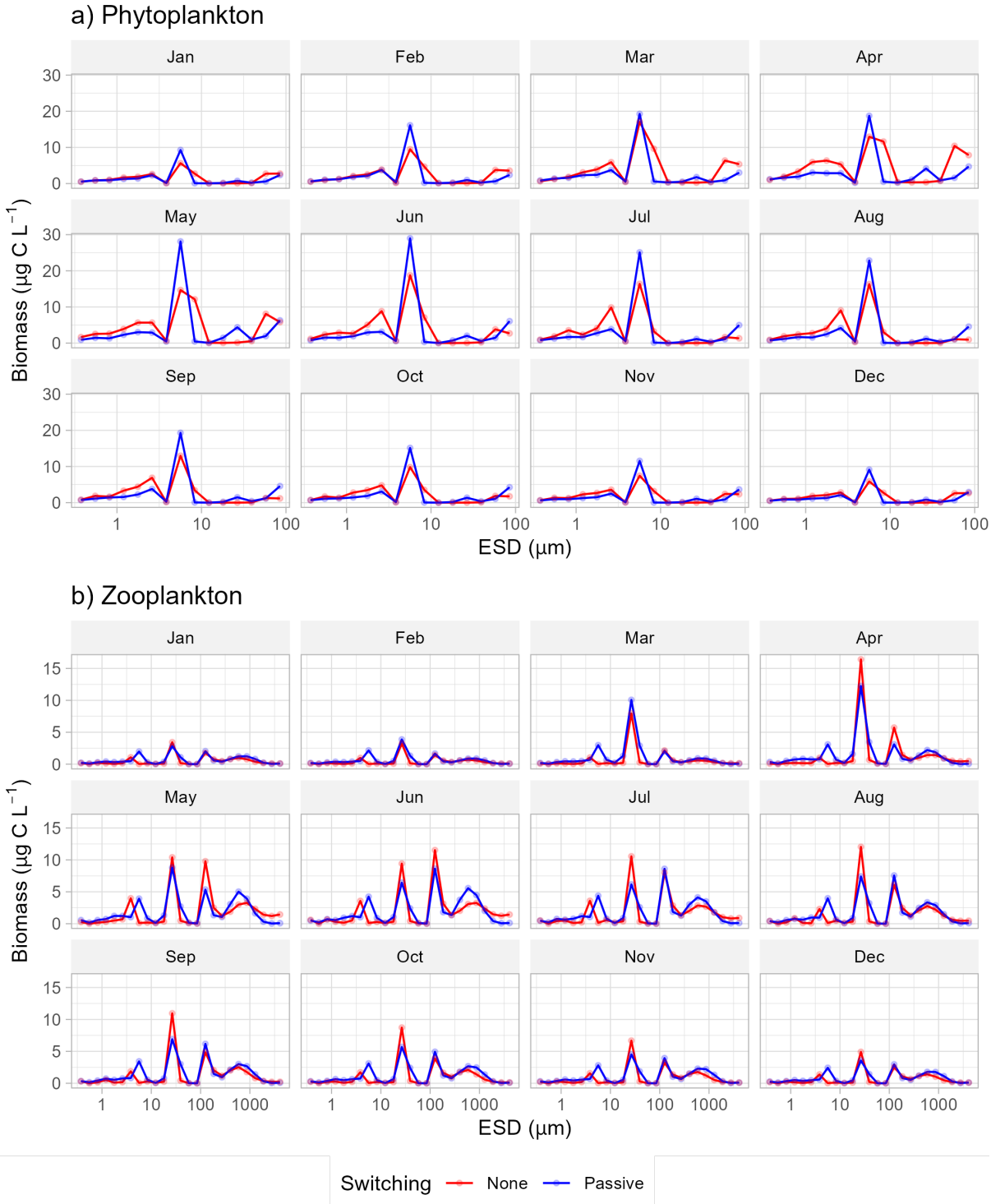


Figure 5.6: Monthly-average biomass distributions of phytoplankton and zooplankton of the model using the optimised parameter set with passive switching between herbivory and carnivory (blue), and without passive switching (red).

plankton at PL (Figure 5.7) suggests that the carnivory coefficient should be much lower for 587 μm zooplankton in April ($\sim$ 0.33). Although the biomass of phytoplankton in the neighbourhood of 5.7 μm is slightly overestimated in March (Figure 4.5), it is the overestimation of equivalent microzooplankton biomass that drives the large modelled $\varphi^{(Z)}$. This has consequence for the food web dynamics and subsequently the bottom-up control on zooplankton, since a greater proportion of the biomass at 27 μm is consumed at the third trophic level in the model than observations would suggest.

It is likely that the top-down control on phytoplankton size, in reality, is influenced by deviations from zooplankton-wide allometric scaling of θ^{opt}, $G^{\max}$, and k_Z. A size-based kill-the-winner approach to switching may enhance herbivorous grazing on smaller cells by larger zooplankton. However, the grazing preference value ($\phi_{i,j}$) for a 589 μm zooplankton on its optimal prey (27 μm cell) is 110$\times$ higher than it is for a 5.7 μm cell, for the presented allometry of θ^{opt} and value of σ. Such a difference would likely inhibit grazing of 5.7 μm cells by larger zooplankton, even if size-based switching was included in the model. Thus, for larger zooplankton to graze optimally, θ_{opt} resolved across different zooplankton functional groups would be necessary. This is exhibited most clearly by comparison of estimated grazing saturation from observations (Figure 5.5) against observed biomass distributions (Figure 4.5). Estimated grazing saturation from observations is maximised at 27 μm from January to May, in response to the dominance of 5.7 μm cells. Maximal grazing rates are also comparatively large for 27 μm zooplankton compared to $> 85\,\mu$m zooplankton by a factor of 2.46$\times$. However, this does not align with observations, whereby biomass accumulation in March and April is dominated by mesozooplankton on $400 - 1278\,\mu$m (if nanozooplankton $\sim 3\,\mu$m are excluded). This suggests strongly that allometric scaling of grazing used here is not entirely applicable to the zooplankton community at PL. A second possibility is that σ is underestimated, leading to an excessively steep decay of $\phi_{i,j}$ across the prey size domain, for any given predator. However, since the location of the maximum in a predator's grazing preference is controlled by θ_{opt}, improved fitting of σ would likely still rely on species-resolved θ^{opt} to maximise model fit, given the biological deviations from allometry displayed in Hansen *et al.* (1994). Nonetheless, while σ was omitted from model fitting since the objectives focused on mean size (rather than size variance), the fact that the grazing interactions are complex shows that σ may impact mean size as well as size variance, even though the exact mechanism for this may not be clear *a priori*.

5.3.3.4 Summary

Overall, the 6 μm cells dominate biomass during spring by having the optimum maximal photosynthetic rate, since nutrients and chlorophyll are not limiting. However, top-down controls also have an important part to play in structuring the community, and controlling the amount of biomass larger cells provide. For small cells ($\sim 1.8\,\mu$m), gross growth and loss rates remains coupled relative to the baseline of 5.7 μm cells, meaning that changes in size structure $< 5.7\,\mu$m do not occur. However, decoupling of grazing on cells $> 5.7\,\mu$m allows growth which in turn increases mean size. It is actually intracellular shading which imposes a necessary bottom-up re-

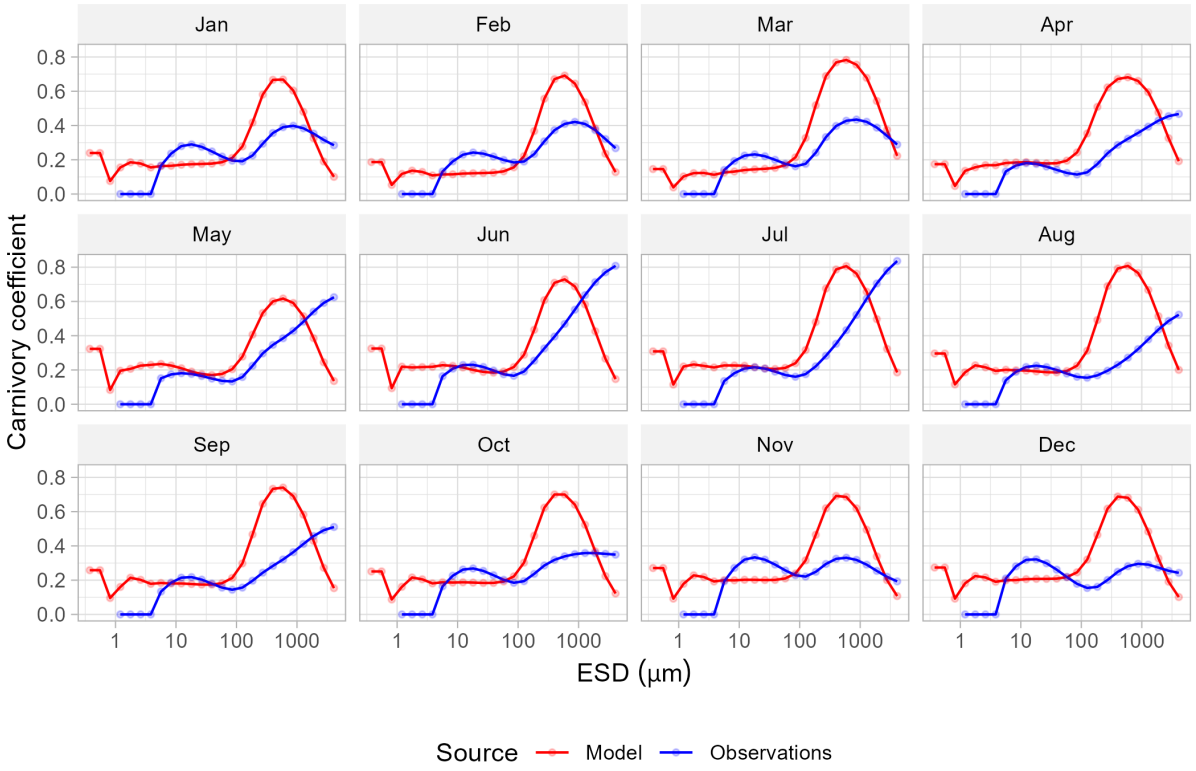


Figure 5.7: Monthly-averaged distribution of carnivory coefficients ($\varphi^{(Z)}$) by size. The herbivory can be yielded as $\varphi^{(P)} = 1 - \varphi^{(Z)}$. 'Model' line is directly sourced from model, while 'Observations' are computed using observed biomasses of phytoplankton and zooplankton. Both cases used the same universal allometric scaling.

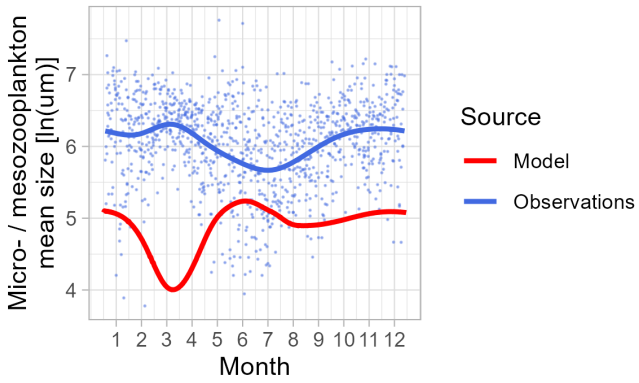


Figure 5.8: Biomass-weighted mean size (E[s]) for micro- and mesozooplankton (excluding heterotrophic nanoeukaryotes) from GAM fitted to observations (blue points / line) and model outputs (red line).

striction on large phytoplankton to stop further increased mean size in March, since the top-down control is comparatively weak; this finding is supported by inference from the ABC optimisation.

The location of the biomass maximum $> 5.7\,\mu$m is driven by the food web structure of the model. When switching between carnivory and herbivory is included, $27\,\mu$m is the optimum 'large cell' size, in line with observations. Yet, when switching is removed, biomass accumulates at regions of the size domain of both smaller and larger cell size, where herbivory was previously the dominant grazing control. As a result, the biomass accumulation at $27\,\mu$m is a loophole through intense grazing on their equally-sized microzooplankton counterparts, who are in turn grazing the $5.7\,\mu$m phytoplankton at which productivity is maximised.

The observed zooplankton community during spring is not accurately resolved by the model, since biomass accumulations of microzooplankton are overestimated, and mesozooplankton are underestimated. The top-down conclusions should therefore be more tentative: it is clear that grazing control an important role to play in structuring the spring phytoplankton community, but the exact mechanisms remain unknown.

5.3.4 H6: Decrease of micro- and mesozooplankton mean size in summer

> *Micro- and mesozooplankton mean size decreases in late spring / early summer because, one, or either of:*
>
> *i. the allometric maximum grazing rate relationship.*
>
> *ii. small zooplankton are most optimally sized to respond to the prey availability .*

5.3.4.1 Taxonomic variability in θ^{opt} as a possible driver of poor fit

The fit of the model to the size structure of micro- and mesozooplankton is relatively poor in spring (Figure 5.8). From March to June, the modelled mean micro- and mesozooplankton size is increasing, while the observed mean size is decreasing (until July). As such, it is not appropriate

to use the model to assess the decrease in micro- and mesozooplankton mean size in spring and early summer which underpins H6. There are, however, some reasonable mechanisms which may explain this qualitative difference, and provide insight to the potential importance of taxonomic differences in the zooplankton community.

In March, before the start of the reduction in observed mean size, the mean size increases slightly in the micro- and mesozooplankton community (and increases more notably if nanozooplankton are included, as in Figure 4.4). At this time, the zooplankton mean size (approximately equivalent to around $400\,\mu$m) is larger than allometric theory would expect (around $30-40\,\mu$m, neglecting grazing on picoplankton by nanozooplankton), if the community was purely bottom-up limited by phytoplankton prey abundance. While carnivorous grazing on microzooplankton may allow zooplankton mean size to increase to a larger value than $30-40\,\mu$m, this would not appear to explain a mean size as large as what is observed, unless the food web was extremely top-heavy.

What appears to be more likely is that the pattern is driven by deviations from allometric scaling theory. We have discussed that the large zooplankton mean size during this period is at least partially driven by the immigration of meroplankton, which graze on the dominant $6\,\mu$m phytoplankton at this time. However, the subsequent observed reduction in zooplankton mean size − which occurs when the $6\,\mu$m phytoplankton reduce in biomass and $30\,\mu$m phytoplankton accumulate biomass − also appears to contradict expectation from allometric scaling theory. During this period, dinoflagellates and ciliates contribute to the microzooplankton, while copepods dominate mesozooplankton on $300-1000\,\mu$m (Figure 2.13). On this size interval, copepods tend to graze optimally on cells on $10-30\,\mu$m, although the largest copepods may graze optimally up to $100\,\mu$m cells (Hansen *et al.*, 1994). This both drives and aligns with expectations of allometric theory (Banas, 2011); so too does ciliates of size $\sim 50\,\mu$m grazing on $\sim 10\,\mu$m prey (Hansen *et al.*, 1994). However, heterotrophic dinoflagellates − a key constituent of the microzooplankton community at PL, which become increasingly numerous in May and June − graze optimally on cells much larger than allometric theory would predict, up to a 1:1 size ratio (Hansen *et al.*, 1994). While only one dinoflagellate is recorded in Hansen *et al.* (1994), it is omitted from the fitting of the allometric coefficient and exponent of Banas (2011) because dinoflagellates fundamentally do not fit the allometric curve; neither do meroplankton larvae.

In light of this, a possible food web scenario is that grazing on $6\,\mu$m nanoplankton in summer is dominated by ciliates, while both dinoflagellates and copepods graze on the larger phytoplankton cells, despite their order of magnitude difference in ESD. Ciliates and other dinoflagellates may in turn also be grazed by dinoflagellates and copepods, perhaps explaining the growth of micro- and mesozooplankton mean size in July to October (Figure 5.8).

5.3.4.2 Summary

Although the model is unable to resolve patterns in micro- and mesozooplankton mean size accurately, we speculate that deviations from the standard allometric scaling of θ^{opt} have an important consequence for the size structure of the zooplankton community. Not only would this affect biomass distribution of herbivorous zooplankton, but is likely to have substantial impact in

overall food web dynamics since the zooplankton may graze on each other, thus directly driving modelled patterns in mean size. Interactions which the model may assume to span two trophic levels (e.g., the transfer of biomass to nanophytoplankton to mesozooplankton via microzooplankton) may actually only cover one trophic level, when grazers have larger predator:prey size ratios than allometry expects; the opposite may also be true when size ratios are smaller than expected. The allometry of θ^{opt} appears to show more variability than k_Z and $G^{\max}$ (based on visual assessment of Hansen *et al.*, 1994 and Hansen *et al.*, 1997), and therefore should be a focus in future modelling efforts. Similarly, the parameters associated with maximum photosynthetic rate and nutrient uptake for phytoplankton show far less interspecific variability, with very high r^2 values (Marañón *et al.*, 2013).

5.3.5 H4: Large phytoplankton mean size in summer

Through spring to summer, the phytoplankton mean cell size increases because:

i. Large phytoplankton can exploit 'unfulfilled' growth which the small phytoplankton are unable to attain, due to the tight coupling between small phytoplankton their predators resulting in focused high intensity grazing.

ii. Mixotrophy accounts for some of the increase in mean size within large nano- and microphytoplankton from spring to summer at SH & LE.

The model is able to resolve increasing phytoplankton mean size from March to May, in line with observations. Thereafter, however, modelled mean size decreases, whereas observed mean size attains an approximate plateau from June to August. During this period, observed size diversity is also increasing (Figure 2.14), which the model also fails to resolve (Figure 4.6). The increase in observed size diversity at this point is driven by the 'balancing' of the peaks on the size domain around the neighbourhoods of 6 μm and 30 μm.

Conversely, in the model, biomass at 27 μm diminishes through summer. In May and June, biomass initially reduces due to an increase in gross losses. The gross loss rate reduces thereafter but, despite this, the large phytoplankton cells still exhibit net losses rates (Figure 5.1) due to a reduction in photosynthetic activity. This indicates that the large cells are bottom-up limited from June to August.

Since the model is unable to resolve the sustained biomass of large cells, it cannot be used to directly investigate H4. However, an assessment of the bottom-up control on large phytoplankton can provide insight to the potential mechanisms which may facilitate the sustained large-cell biomass in observations.

5.3.5.1 Nutrient limitation

The main issue facing the phytoplankton community in the model is that large cells cannot survive low nutrient concentrations in summer. This is a result of the severe nutrient limitation on large cells (Figure E.3); in July, $\lambda_{N,1} = 0.87$ at 0.38 μm, while $\lambda_{N,8} = 0.25$ at 5.7 μm, and $\lambda_{N,15} = 0.12$ at 85 μm. The result is that V_C^{sat} is diminished (Figure E.7) and, despite the

Table 5.2: Fixed effects model estimates for $V_{C,12} \sim \lambda_{N,12} * Q_{Chl,12} + V_{N,12}$, decomposing effects of intracellular nutrient limitation, intracellular chlorophyll, and nutrient uptake (which incurs a penalty on carbon accumulation). $V_{C,12}$ (d^{-1}) is averaged $V_{C,i}$ for size class $i = 12$ ($27\,\mu$m) across July irradiance and temperature. The baseline level is where winter values of λ_N, Q_{Chl}, and V_N are used.

Summer parameters	Effect on $V_{C,12}$
None	0.2595
λ_N	−0.1652
Q_{Chl}	−0.1137
V_N	−0.0040
$\lambda_N : Q_{Chl}$	0.0898

fact that the tradeoff for α theoretically becomes less important when I is high, photosynthetic performance is also impaired by the emergence of a Q_{Chl} tradeoff, which decreases by size (Figure E.2). As a result, the effective parameter for the initial slope of photosynthesis, α^*, reduces by more than half for phytoplankton of size $27\,\mu$m compared to winter. Furthermore, the penalty of biosynthesis ($\zeta \cdot V_N$) also increases (Figure 5.9), due to the increase of $Q_N^{\text{stat}} = 1 - \lambda_N$, despite low uptake saturation ($\frac{N}{k_N+N}$) in larger size classes. It is therefore useful to decompose these effects on photosynthesis to understand how large phytoplankton may cope with depleted nutrients in nature.

The decomposition of nutrient limitation is elicited in Table 5.2. When irradiance and temperature are high, as they are in July, we find that the nutrient-related penalties reduce summer photosynthesis by $\sim 75\%$, compared to how cells would behave if nutrients were not limiting, as in winter. The biggest effect is by the reduction in λ_N, which accounts for around two-thirds of the reduction, while Q_{Chl} accounts for around one-third (assuming interactive effects can be attributed equally to λ_N and Q_{Chl}). The cost of biosynthesis has a comparatively very small effect. Of course, the depletion of nutrients is a key component of the seasonal cycle. There is no question that reductions in λ_N and Q_{Chl} should impact photosynthesis; however, the extent of the effect is harder to estimate.

Verifying the effect of Q_{Chl} with observations is perhaps easier than for λ_N. The Q_{Chl} tracks broadly within realistic ranges for coastal regions (equivalent to $20-50$ g C (g Chl a)$^{-1}$; Jakobsen & Markager, 2016; Figure E.2). The question, then, may be whether the depletion of nutrients is excessive in larger cells, driving λ_N and subsequently V_C^{sat} to be too small.

The nitrogen uptake rate (per unit of carbon biomass) is low in winter for in all cells (Figure 5.9). This is largely driven by the fact that Q_N^{stat} is low (Figure E.3). In spring, V_N increases, but at a slower rate than photosynthesis. As a result λ_N is decreasing and Q_N^{stat} is increasing. In April, $\lambda_N \approx Q_N^{\text{stat}} \approx 0.5$ for cells $> 5.7\,\mu$m. This would suggest that until this point, growth is favoured to photosynthesis in order to attain balanced stoichiometry. In early- to mid-spring, when light is high and nutrients are not yet depleted, this is a reasonable state for cells to attain. Then, in May, λ_N reduces further; this would increase V_N further through an increase in Q_N^{stat}, if other variables were held constant. However, V_N decreases (Figure 5.9) because nutrients

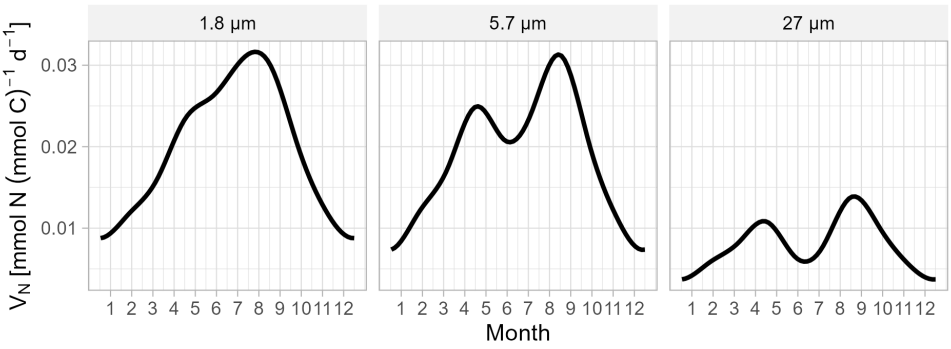


Figure 5.9: Seasonal progression of nitrogen uptake rate, per unit of carbon biomass (V_N; mmol N (mmol C)$^{-1}$ d^{-1}) for three dominant size classes of the model.

become depleted. At this point k_N drives the dynamics of nutrient uptake. The model therefore appears to qualitatively reproduce the intended effects of nutrient depletion on the phytoplankton community through spring and into summer, by allowing the nutrient half-saturation penalty to influence size structure in summer.

Given that both λ_N and Q_N^{stat} exhibit such large seasonal variability, especially for larger cells (Figure E.3), then it is reasonable to anticipate that the parameterisation of Q_N^{min} and Q_N^{max} may influence model dynamics. However, the parameter sensitivity analysis (Figure 4.8) suggests that changes to Q_N^{min} and Q_N^{max} have a small impact on the model, compared to other parameters. If Q_N^{min} and Q_N^{max} only have a weak effect on the model, then the values λ_N and Q_N^{stat} are likely to be an emergent property of the model dynamics, driven by other mechanisms (e.g. maximal uptake rates and affinity), rather than driving the dynamics of the model themselves.

Finally, it is worth recalling that the model *underestimates* the nutrient depletion, despite this being the mechanism appearing to drive poor model fit to phytoplankton in summer. Given the allometric tradeoff for k_N for cells of size 27 μm, the nutrient uptake saturation would be ~ 0.05 if nutrient concentration was equal to averaged observations for July. In this scenario, biomass of 27 μm cells is likely to be further depleted, since V_N would be penalised. It is likely that the parameter optimisation compromised goodness-of-fit to observations of nutrient concentrations and phytoplankton mean size since, under the current modelling framework, there does not appear to be a viable mechanism for 27 μm cells to sustain growth when nutrients are depleted in summer.

5.3.5.2 Mixotrophy as a potential bottom-up enhancer

Given that the modelled reduction of phytoplankton growth aligns with expectation under allometric theory, a missing mechanism may be responsible for the under-represented biomass of cells around 27 μm. Mixotrophy is recognised as crucial in structuring the size distribution of plankton at both a theoretical (Ward & Follows, 2016) and observational level (García-Oliva *et al.*, 2022). In particular, Ward & Follows (2016) show that mixotrophy can increase mean cell size.

Within the model, mixotrophy would appear to be a viable solution to increase 'phytoplankton' mean size in summer. Observations of mixotrophic plankton are split in biomass by a 50:50 ratio to phytoplankton and zooplankton; this is a necessary compromise, but neglects the synergistic effect that mixotrophs may experience when complementing autotrophy with heterotrophy (or vice versa; Anschütz & Flynn, 2020). Mixotrophy is recognised as a potential strategy to evade nutrient limitation (Stoecker *et al.*, 2017).

Additionally, mixotrophic dinoflagellates often have an optimal predator:prey size ratio smaller than 1:1 (see examples in García-Oliva *et al.*, 2022). As such, mixotrophs may graze on smaller prey, e.g. phytoplankton in the neighbourhood of 6 μm. In the context of the UK coastal ecosystem, this would be highly advantageous, since the majority of prey biomass is concentrated within the nanophytoplankton. Such grazing activity would also fall in better alignment with the allometric fit of θ^{opt}, which may be a useful consideration in future models. Since biomass would be passed up the 'phytoplankton' food web (in its current definition), mixotrophic grazing by cells around 30 μm on 6 μm phytoplankton may also help to balance out the observed peaks of biomass (Figure 4.5), and thus facilitate the observed increasing mean size and size diversity during summer.

Currently, size-structured models which include mixotrophy are limited. The most analogous approach to that used here is of Ward & Follows (2016), where the model of Ward *et al.* (2012) is adjusted to include a mixotrophic class which exercises both phytoplankton and zooplankton processes. A weighting parameter, which determines the penalty on grazing, photosynthesis, and nutrient uptake rates, is investigated over a range of values. Without such a penalty, mixoplankton displace phytoplankton and microzooplankton. However, to present such a model − which is mechanistically experimental − first requires an established baseline validated by observations (as in Ward *et al.*, 2012). Given that observations of mixoplankton contribute substantial biomass in summer and autumn at UK coastal sites (Figure 2.12), future modelling efforts should attempt to incorporate mixotrophy, perhaps using the baseline model presented here, since it reproduces the cycle of phytoplankton size structure in winter and spring when mixoplankton contribute limited biomass.

5.3.5.3 Summary

In early spring, phytoplankton mean size increases because the 27 μm cells can evade grazing relative to the 6 μm cells, in support of H4i. However, by summer, evasion of predation is no longer a sufficient mechanism to facilitate growth. In May, grazing on large cells reaches its annual maximum rate (Figure 5.1) in response to the accumulated biomass. The fact that a cell may 'escape' grazing is therefore only a temporary loophole.

Thereafter, however, gross loss rates decrease, yet the modelled large phytoplankton continue to exhibit net loss rates, in contrast to observations. This suggests a succession in strategy must exist to facilitate biomass accumulation. Given that the model appears to limit photosynthesis through nutrient depletion in line with allometric theory, mixotrophy would appear to be a viable strategy to increase gross growth rate. Indeed, mixotrophs emerge in the UK coastal communities

in this time. Thus, we speculate that mixotrophy is responsible for the continued increase in mean cell size in summer, in line with H4ii.

5.3.6 H5: Maximal phytoplankton size diversity in autumn

> *The increase in phytoplankton size diversity in late summer / early autumn, associated with a decrease in mean size, is due to the allometric tradeoff for maximal nutrient uptake, since the nutrients replenish with the deepening ML during this period.*

The increase in observed phytoplankton size diversity in September and October is driven by a reduction in cells in the neighbourhoods of 6 μm and 30 μm, while 2 μm phytoplankton maintain the same biomass as in summer (Figure 4.5). The model also produces an increase in size diversity at this time (Figure 4.6), primarily driven by a decrease in the dominance of the 5.7 μm size class biomass (Figure 4.5). However, in the model, biomass is lost by cells at 1.7 μm, and it is the fact that large cells (85 μm) maintain biomass which primarily drive an increase in mean size and size diversity. The mechanisms controlling the increase diversity therefore appear to be different between the model and observations. Given that the model appears to overestimate the biomass of these 85 μm cells, the model cannot give a concise answer to H5. However, it is useful to assess the reasons why the small cell biomass decreases during this period in the model, given that they are constant from August to October in observations.

5.3.6.1 What facilitates phytoplankton biomass accumulations at 85 μm in the model?

It is first important to understand the tradeoff which leads to the proportionally large modelled biomass contributions of 85 μm cells year-round, and particularly in late summer and early autumn. From assessment of gross growth and loss rates (Figure 5.1a), it is likely that the advantage these cells have is top-down related. They have the lowest gross growth rates of all phytoplankton in the model, in line with expectations from allometric theory for $V_N^{\max}$, $V_C^{\max}$, k_N, and α. However, their optimal predators in the model (4085 μm zooplankton) have a very *low* carnivory coefficient year-round (Figure 5.7). It is therefore surprising that the 85 μm phytoplankton − rather than zooplankton − contribute significant biomass, since grazing is primarily focused on phytoplankton (we have already observed that the 27 μm phytoplankton of the model escape grazing through a *high* optimal predator carnivory coefficient)[3]. The prevalence of 85 μm phytoplankton is therefore driven by the low grazing saturation of 4085 μm zooplankton (Figure 5.5), which arises because the 85 μm zooplankton have very little biomass in the model.

[3]When switching is turned off (Figure 5.6), the 85 μm phytoplankton biomass is much higher in March to May, replacing 27 μm phytoplankton cells (discussed in Section 5.3.3.3). This results in a much larger biomass of their optimal predator (4085 μm zooplankton) from May to August, which inflicts much greater top down control on the 85 μm phytoplankton, reducing their biomass to near zero during this period. Once the biomass of the large zooplankton reduces in response to the lack of prey, the 85 μm phytoplankton emerge again, particularly from October onwards. This acts as a good case study of the nonlinearity of model effects, and highlights the fact that speculation can only be tentatively made about changes to the model in the absence of any analytical or experimental evidence.

The lack of modelled biomass of 85 μm zooplankton conflicts with observational evidence (Figure 5.1b). Although the microzooplankton biomass is well-represented at the broad scale in the model, with peaks at 27 μm and 125 μm, observations tend to be centred around 58 μm. The underestimated microzooplankton biomass at the 58 μm and 85 μm size classes, and allocation of biomass to adjacent size classes, is likely driven by the bottom-up size-based tradeoffs for k_Z and $G^{\max}$ in the model. The 57 μm and 85 μm size classes have the lowest $G^{\max}$, given a high k_Z, in the model. Thus, these size classes have comparatively low grazing affinity and maximal rates. On a local scale, the size class at 125 μm outcompetes 58 μm and 85 μm by having lower k_Z (and similar $G^{\max}$). The 39 μm size class outcompetes it by having larger $G^{\max}$ (and the same k_Z). Indeed, the 58 μm and 85 μm size classes in the model are the only ones to have a ratio of $G^{\max} : k_Z$ which is less than 8. They can be seen as the local dip in grazing saturation in the neighbourhood of 85 μm zooplankton (Figure 5.5) which, combined with low $G^{\max}$, leads to low overall grazing rates.

Thus, despite the herbivorous focus of the predators of 85 μm cells, the local minimum in modelled grazing saturation appears to override the selective herbivorous grazing, allowing the 85 μm phytoplankton cells to maintain biomass. This is an unintended size-based tradeoff of the model arsing from the selection of k_Z and $G^{\max}$. Given the prevalence of 85 μm zooplankton in the observations, we suggest that a different (perhaps smooth allometric) form of k_Z and/or $G^{\max}$, which avoids discontinuity and/or sharp edges, should be implemented. In nature, it is highly likely that these microzooplankton have some bottom-up tradeoff in UK coastal waters which is advantageous; in Hansen *et al.* (1997), the large microzooplankton have a lower k_Z than the nanozooplankton and smaller microzooplankton, which may drive this.

5.3.6.2 Effect of nutrient replenishment on phytoplankton cells

Top-down effects on large phytoplankton cells appear inaccurate, and drive unintended changes at the larger end of the phytoplankton size domain in autumn. However, this should not significantly affect bottom-up controls on smaller cells, since nutrient uptake by large cells is much lower than for small cells (Figure 5.9), and therefore is unlikely to affect the dynamics of competing for a resource. Despite this, the modelled net loss rates are proportionally equal for 1.8 μm and 5.7 μm cells through September and October, with a fixed ratio between 1.8 μm and 5.7 μm cells (Figure 5.1 & 5.4a). The model therefore fails to resolve the tradeoff which allows cells around 1.8 μm to contribute higher observed biomass as a proportion of the community compared to the summer months.

Since PAR forcing is still relatively high in September, we hypothesised that nutrient-related tradeoffs allowed the 1.8 μm cells to maintain their biomass to a greater degree than 5.7 μm cells. Once light became limiting after October, it would follow that the tradeoff for intracellular shading (α) would become increasingly important, in line with H2i which is supported in Section 3.2.1, reducing size diversity and mean size through the leftward shift of the growth and loss rates curves by size (Figure 5.1).

September is the period when observed nutrients are replenishing at their maximum rate (on the

log-scale), and so the tradeoffs for k_N and $V_N^{\max}$ – which both favour the smallest cells – would likely become important, by allowing the fastest saturation of λ_N. However, the extent of nutrient depletion is underestimated by the model. In September, nutrients have replenished to 2 mmol N m^{-3}, rather than the observed average of 1 mmol N m^{-3} (Figure 4.4). As a consequence, λ_N is perhaps overestimated for larger cells in this period, and has already recovered to $\lambda_{N,8} = 0.55$ for cells at 5.7 μm, from the July minimum of $\lambda_{N,8} = 0.23$. Since the light is externally forced, but nutrients are not, then it is possible that a mismatch in the timing of nutrient replenishment and light reduction allows the 5.7 μm cells to capitalise to a greater degree than they would if the nutrients were still limiting, given that they have the maximal $V_C^{\max}$. Thus, an interactive effect, which is not purely driven by individual effects of nutrients and / or light alone, but the interaction between the two (via λ_N) may be responsible for the proportional increase in pico- and small nanophytoplankton during September and October. Without constructing a model which resolves the nutrient cycle more accurately and thus gives a more accurate representation of likely λ_N, it is difficult to test this hypothesis. We therefore present an alternative hypothesis for future research that a tradeoff between light and nutrient uptake facilitates size diversity by allowing two peaks to exist, both of which are at the smaller end of the size domain.

5.3.6.3 Grazing control

Grazing at the small end of the size domain is typically characterised by tight coupling between predators and prey within the model, with differences between gross growth and loss rates generally only emerging for short periods for small cells. On the other hand, for example, 27 μm phytoplankton exhibit substantial net growth in March and April, then suffer severe grazing losses in May and June (Figure 5.1).

The ratio of gross gains and losses for the 1.8 μm versus 5.7 μm is in synchrony in autumn, indicating that there is no tradeoff (e.g., a proportional reduction in grazing) causing changes in size structure for cells $< 5.7\,\mu$m. Modelled grazing saturation of predators for these cells is reducing steadily and uniformly from June to November, indicating a smooth wind-down in accompaniment to reducing prey biomass. Furthermore, from April to October, both grazing saturation and carnivory appears to fit well for predators $< 85\,\mu$m (who graze on cells sized $< 8\,\mu$m), compared to estimates derived from observations. The evolution of a top-down tradeoff does not appear, in the model, to drive changes in size structure; top down control also aligns with expectations from observations.

Inclusion of mixotrophy in the model may also affect grazing dynamics, and provide a reasonable mechanism to facilitate concurrent decrease in phytoplankton mean size and increase in size diversity. Observations show that mixoplankton continue to contribute biomass into autumn (Figure 2.12). As such, mixotrophic cells of size $\sim 30\,\mu$m may continue to graze on nanozooplankton around 5.7 μm (based on sizes presented in García-Oliva *et al.*, 2022), which are the optimal predators of smaller phytoplankton ($\sim 1.8\,\mu$m). Such an effect could reduce overall top-down pressure on smaller cells.

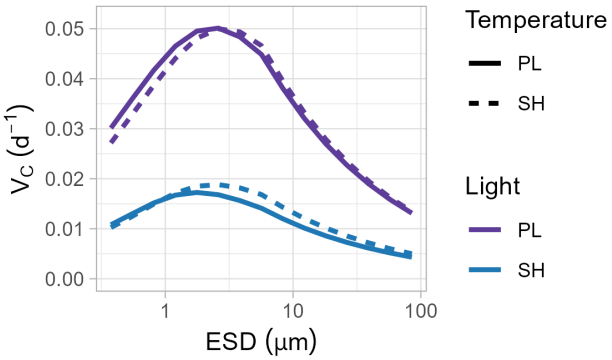


Figure 5.10: Comparison of photosynthetic rate ($V_{C,i}$) by size, using different combination of forcing for light and temperature at SH and PL. Light forcing is inclusive of surface irradiance, background seawater attenuation, and MLD.

5.3.6.4 Summary

The increase in autumn size diversity of the model coincidentally mirrors observations. Mechanistically, the processes which allow the pico- and small nanoplankton to maintain biomass, while other size classes lose biomass, are not resolved by the model. In September and October, modelled gross growth rate is optimised at 5.7 μm, in line with the scaling of $V_C^{\max}$, indicating that α is not structuring the community yet. However, the extent of nutrient depletion is underestimated by the model. It is likely that a reduction in nutrients − and associated decrease in λ_N with size − may facilitate sustained gross growth in the neighbourhood of 2 μm, but elicit decreases for larger cells such as those around 5.7 μm. Conversely, the inclusion of mixotrophy could force a reduction in mean size and associated increase in size diversity, since the predators of picophytoplankton would likely experience enhanced top-down control.

5.3.7 H1: Lower winter phytoplankton biomass at SH compared to PL

Phytoplankton biomass is lower at SH than PL in winter because:

i. differences in light arising from latitude and attenuation limit growth.

ii. biomass accumulation, which is largely governed by small cells in this period, is limited by temperature.

5.3.7.1 Comparison of light and temperature effects on photosynthesis

Phytoplankton biomass is lower at SH than PL in winter. Given that nutrients and chlorophyll concentrations are saturating at both sites (in both observations and model outputs), bottom-up limitation on gross growth may either arise from differences in irradiance or temperature. The bottom-up limitation on gross growth at SH is driven by the diminished light, which has a much greater effect on V_C averaged across December and January than temperature (Figure 5.10). This is slightly surprising given that at SH, $\lambda_T^{(P)}$ is 0.83× lower than PL in winter, driven by an approximate 3°C temperature difference. This incurs a 17% reduction in V_C^{sat}, yet has very little

effect on bottom-up control when light is limiting. In fact, it actually increases V_C and moves the unimodal peak to slightly larger sizes, since a reduction in V_C^{sat} when I is small reduces the allometric penalty of α incurred on larger cells.

5.3.7.2 Disentangling the factors which reduce light at SH

There are at least five potential mechanisms which govern the irradiance throughout the mixed layer. The irradiance at depth $z \in (0, H)$ is given by

$$I(z) = I_0 \exp\left\{-\left(k_w + k_{Chl} \cdot P_{Chl}^{\text{tot}}\right) \cdot z\right\}.$$

Changes in surface PAR (I_0), background seawater attenuation (k_w), chlorophyll a, and depth all influence irradiance. Furthermore, albedo − broadly increasing with latitude in winter − may also reduce penetration of incident irradiance through the surface.

Photosynthesis is strictly increasing with I under the Geider formulation. Since an increase in H reduces depth-averaged I, it also reduces depth-averaged V_C, if all other variables remain constant. As such, SH suffers four-fold effects of reduced light. The midwinter I_0 is around half that at PL, due to latitude and possibly climate; its k_w is higher year-round, and particularly in winter (61% higher); its H is around 6 m deeper in winter; and its albedo is also likely higher (Kukla & Robinson, 1980), although this is excluded due to uncertainty described in Section 1.4.1.1. While P_{Chl}^{tot} is slightly lower in winter at SH, reducing the shading by phytoplankton cells, the value of $k_{Chl} \cdot P_{Chl}^{\text{tot}}$ is an order of magnitude smaller than k_w, thus k_w dominates combined seawater attenuation (Figure 2.10).

To assess the impact on V_C contributed by each of the mechanisms which reduce winter light at SH, we undertook a full-factorial decomposition of $V_{C,i}$ averaged across all size classes i ($\bar{V}_C$). The temperature forcing at SH was used throughout, while the baseline light-related forcing of I_0, k_w, and H were at PL, with all combinations of respective SH forcing used. All marginal effects have a role in reducing $\bar{V}_C$ (Table 5.3). The marginal effect of using I_0 forcing from SH incurs a 41% decrease in $\bar{V}_C$; while the k_w forcing from SH reduces $\bar{V}_C$ by 22%; and the H forcing from SH reduces $\bar{V}_C$ by 12.6%. Interactive effects are all positive (albeit small) as a result of the saturating shape of the Geider P-I curve. It is unsurprising that a 50% decrease in I_0, which is the case for SH compared to PL, dominates the reduction in photosynthesis. However, the reduction incurred by seawater attenuation and MLD is also substantial. Bottom up controls therefore have a role to play in the reduced biomass at SH, and are primarily a result of reduced surface irradiance, but also contributed to by greater seawater attenuation and a deeper ML in the model. Temperature appears to have a minimal bottom-up impact.

5.3.7.3 Comparison of top-down controls

Despite the reduced gross growth at SH compared to PL across all sizes in winter, the dynamics of biomass accumulation are similar. Net losses switch to net gains in early January, although the timing of the switch from net loss to net gain is 11 days later at SH. As is the case at PL, a

Table 5.3: Fixed effects model estimates for $\bar{V}_C \sim I_0 * k_w * H$, decomposing effects of surface PAR, seawater attenuation (excluding that of chl a), and MLD. $\bar{V}_C$ (d^{-1}) is the unweighted average V_C across size class. $\bar{V}_C$ is then averaged across values December and January for 6-hourly resolved I_0, k_w, and H. The baseline level is where all irradiance forcing is used for PL, with temperature forcing from SH in all cases.

Forcing from SH	Effect on $\bar{V}_C$
None	0.0246
I_0	−0.0109
k_w	−0.0054
H	−0.0031
$I_0 : k_w$	0.0020
$I_0 : H$	0.0011
$k_w : H$	0.0014
$I_0 : k_w : H$	0.0014

decrease in gross loss rates of phytoplankton drives the initiation of biomass accumulation. The mechanisms controlling winter biomass dynamics are therefore the same at SH as at PL, which suggests that the difference in winter biomass is driven by details within the gross growth and loss rates, rather than a different mechanism.

The phyto- and zooplankton biomasses remain well-coupled at SH and PL from November until early January (Figure 5.11). As a result, the top-down pressure on phytoplankton previously observed at PL is approximately mirrored at SH. The constant P:Z ratio at both sites in November and December indicates that the phytoplankton are tightly controlled by the zooplankton over this period. The phytoplankton biomass is marginally higher at SH and, as such, there is no suggestion that the zooplankton grazing is proportionally more intense at SH; if anything the opposite is true. A small difference such as the one observed in Figure 5.11 is likely driven by the Holling-III grazing formulation, for which the gradient is decreasing with diminishing food availability when F_j is small, as it is in winter for all predators j. Temporal biomass changes across winter therefore appear to be driven by top-down control, but differences in the standing stock of winter biomass do not appear to be driven by differences in grazing between sites.

5.3.7.4 Parameter considerations

Further work can be undertaken to validate the assessment of winter biomass by parametrising the model to SH, since the current analysis is conditional on the best-fitting parameter set at PL. However, it appears unlikely that qualitative differences in winter model outputs would arise, since the phyto- and zooplankton are tightly coupled. We anticipate that the winter standing stock of biomass is driven primarily by α since low irradiance drives the low biomass and, since the findings of the model align with observations, there is some initial support that the coefficient α_a at SH is similar to that of PL.

A valid counterargument is that α is dependent on k_{Z1} (Table 4.5). Since the bloom dynamics at SH & LE are comparatively poorly fitting, a change in grazing parameterisation may be necessary to improve model fit. At SH, the modelled P:Z ratio is able to grow to an annual maximum of

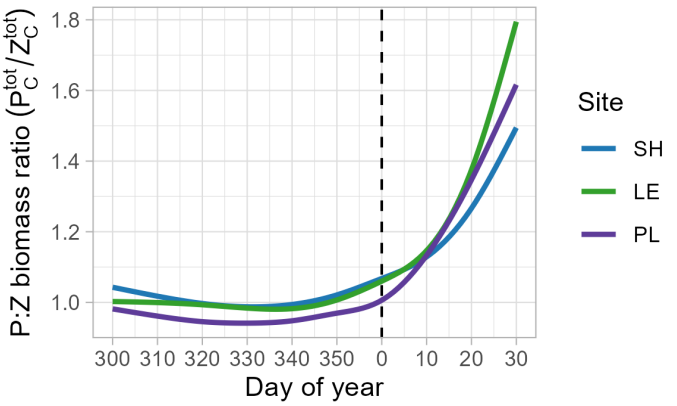


Figure 5.11: Comparison of modelled P:Z biomass ratios across November to January using forcing from SH, LE, and PL. Note a 360 day year is used by the model.

4.5 during the bloom, which is much higher than at PL (2.0). Having already established that top-down control is crucial in structuring the bloom (Section 5.3.3.2), it follows that changes in k_{Z1} (to which the model is highly sensitive; Figure 4.8) could aid model fit. In turn, these may incur changes in α_a and thus, winter model fit.

5.3.7.5 Summary

Stark differences are predicted in winter photosynthesis, which is around 60% lower at SH than at PL. Temperature appears to have a negligible effect, with almost all differences in photosynthesis driven by light limitation. The primary source of light limitation is reduced surface irradiance, but seawater attenuation and MLD also contribute significantly. As such, the coupling of zooplankton and phytoplankton appears to drive temporal changes in biomass, while the standing stock of biomass is limited by light. However, it should be noted that the analysis is conditional on the best-fitting parameter set at PL. While we anticipate that qualitative differences are unlikely to arise for parameters optimised at SH in winter, further parameter optimisation would positively contribute to the understanding of variation between SH and PL. Our findings provide overall support for H1i, and there is no evidence in favour of H1ii.

Top-down control, as a proportion of biomass, appears similar at SH & PL. At both sites, phytoplankton and zooplankton are tightly-coupled, with constant P:Z biomass ratio around 1. Minor differences are observed in the P:Z biomass ratio which likely originate from the Holling-III grazing response. However, these act to release grazing pressure at SH, rather than increase it, when food availability is low.

5.4 Discussion

5.4.1 Key findings

5.4.1.1 Bloom formation and the disturbance-recovery hypothesis

Our findings partially align with the original mechanisms set out in the DRH, which stipulates that the reduction of predation losses on phytoplankton drives the initiation of the spring bloom, rather than an increase in photosynthetic performance (Behrenfeld & Boss, 2014). Although we find that modelled size-averaged gross growth rates increase from December ($0.031\,\mathrm{d}^{-1}$) to January ($0.037\,\mathrm{d}^{-1}$), we also find that size-averaged gross losses decrease from $0.037\,\mathrm{d}^{-1}$ to $0.029\,\mathrm{d}^{-1}$. This suggests that both bottom-up and top-down processes both evolve at the initiation of the bloom, more in line with the suggested mechanisms in González-Gil *et al.* (2022) from observations at SH. In the ecosystem at PL, this is unsurprising: MLD is at its annual maximum from November to February and, while temperature is still decreasing until early March (Figure 2.6), surface irradiance is increasing from the winter solstice. Having shown that winter photosynthesis is more sensitive to light than temperature, we expect that the increase in PAR drives changes in photosynthetic activity. Conversely, in the model employed in Behrenfeld & Boss (2014), the MLD is still deepening until the end of February, and so increase in surface PAR may be somewhat moderated by the continual deepening of the ML, which reduces depth-averaged photosynthesis. The ML is also much deeper, since an open ocean ecosystem is considered in Behrenfeld & Boss (2014) with $H > 100\,\mathrm{m}$ in winter.

Secondly, we find that size-averaged net growth rate maximises ($0.018\,\mathrm{d}^{-1}$) in mid-February. Qualitatively, this is in line with GAM fits at SH (González-Gil *et al.*, 2022), although the timing is slightly earlier reflecting the fact that the timing of the modelled initial component of the bloom at PL is around 25 days earlier than a GAM fitted to observations. In any case, our findings do not align fully with the DRH, whereby maximal net growth rate is expected when gross growth is near its annual minimum (Behrenfeld, 2010; Behrenfeld & Boss, 2014). In the model at PL, net growth is on the order of 10^{-3} in the weeks after net loss turns to net growth, rather than a sharp increase shown in Behrenfeld & Boss (2014). Our community therefore exhibits a smoother transition in growth rates. Nonetheless, the fact that peak net growth occurs before the peak in gross growth aligns with the expectations of the DRH at a general level, even if there is a discrepancy in the timing of the net growth maximum.

5.4.1.2 Light structures UK coastal communities in winter and spring

Although a relaxation of top-down control plays a part in initiating the phytoplankton bloom, we also find that light has an important role in structuring the phytoplankton community in winter and spring at PL by limiting growth of all cells, and in particular large cells. In winter, light is sufficiently low that the cells sized $\sim 6\,\mu\mathrm{m}$, which have the highest maximal photosynthetic rate (Chen & Liu, 2010; Marañón *et al.*, 2013), are light limited due to intracellular shading; instead cells $\sim 2\,\mu\mathrm{m}$ have the highest photosynthetic rate. However, tight coupling with their predators

means that while these cells can contribute proportionally more biomass to the community in winter than in other seasons, they are never able to dominate the community. At SH, the same proportional contribution of pico- and small nanoplankton is made (Figure 2.11), and so we speculate that similar bottom-up processes structure the community of small phytoplankton cells.

At SH, the same biomass of cells $> 10\,\mu$m exists in winter as at PL (Figure 2.12). Given that the biomass of cells in the neighbourhoods of $2\,\mu$m and $6\,\mu$m is lower (Figure 2.11), the mean cell size at SH is higher than at PL. In the absence of any bottom-up advantages in winter, large cell biomass is allowed to exist due to weak top-down pressure. In fact, when the tradeoff for intracellular shading is released, mean cell size attains an annual maximum in winter. This succinctly establishes a cause-effect relationship between intracellular shading and winter cell size. Combined with the posterior estimate of α_b which has a 95% Bayesian CI of $(-0.203, -0.072)$, our findings show that the package effect has a crucial role in structuring the winter UK phytoplankton community. Furthermore, the relationships established between α_a and α_b with k_{Z1} and k_{Z2} during the ABC optimisation further indicate that intracellular shading does not act alone, and that balance is necessary between bottom-up and top-down controls in order to allow larger phytoplankton to maintain a small amount of biomass, but not too much.

Overall, the community biomass at SH is lower in winter because depth-resolved daily-averaged light is less than half of that at PL. This limits the photosynthesis in the $2\,\mu$m and $6\,\mu$m neighbourhoods, which are top-down controlled, reducing the overall community biomass. While the lower incident surface irradiance is the primary driver for this, driven by higher latitude and differences in weather, increased seawater attenuation and a deeper ML also contribute to decreased photosynthesis averaged across the mixed layer. As such, the exponent of the allometric tradeoff for intracellular shading (α_b) is necessary to reduce large cell biomass from the community which would otherwise evade grazing (Figure 5.3). On the other hand, the coefficient (α_a) is the key parameter in controlling overall community biomass, and drives the biomass of phytoplankton to be lower at SH than PL in winter.

In apparent contrast to our findings, Terseleer *et al.* (2014) incorporated intracellular shading, yet find that mean diatom cell size is large in winter (through both model outputs and observations). While raw biomass outputs for functional groups (bacteria, *Phaeocystis*, nanoflagellates, diatoms) suggests that the mean phytoplankton cell size may actually be small in winter when all phytoplankton taxa are considered, it is still interesting that mean diatom size is larger in winter. Although diatoms mean size is not assessed directly in Chapter 2, the microplankton mean size (inclusive of microzooplankton and mixotrophs) is largely invariant across the year at PL (Figure 2.16). At SH & LE, observed microplankton mean size is smaller in winter, but this is likely to be influenced by bias on the basis of findings in Chapter 3. In any case, the cycle of diatoms specifically at UK coastal sites may be further investigated in future, with a specific interest in comparing to the findings of Terseleer *et al.* (2014), noting that their calibrated value for $\alpha_b = -0.077$ is at the very edge of our 95% Bayesian credible interval (Table 4.4) and so may

drive a weaker size-based tradeoff for the package effect.

5.4.1.3 Phyto- and zooplankton allometric theory alone cannot explain the large phytoplankton mean size in summer

Despite the fact that the model actually overestimates summer nutrient concentration, the large cells remain too nutrient limited to sustain sufficient gross growth. In many regards, this is unsurprising: traditional bottom-up theory anticipates smaller mean size in summer, due to nutrient depletion (p. 97 Simpson & Sharples, 2012; Acevedo-Trejos *et al.*, 2015). While grazing evasion may be a mechanism to temporary escape grazing, gross growth must remain sufficiently high, compared to grazing pressure, to facilitate net growth (or biomass retention).

To compensate for (and/or as a result of) the low gross growth of large phytoplankton cells, the model underestimates the grazing saturation of optimal predators ($589-868\,\mu$m) on the $\sim 30\,\mu$m phytoplankton, and overestimates the associated carnivory coefficient. Both of these mechanisms reduce top-down pressure on the modelled large phytoplankton, compared to expectations from observations under typical allometric scaling. Thus, modelled predation appears to move in synchrony with the underestimated prey availability in summer, and so the model does not appear to inflict excessive top-down control on large cells. The grazing on large phytoplankton cells is likely underestimated in order to compensate for bottom-up constraints on growth, and was thus necessary to drive good model fit during the parameter optimisation process.

While large cells are able to exhibit greater variability in intracellular nutrient concentrations, the model does not appear to be sensitive to changes in $Q_N^{\min}$, compared to other parameters. While a reduction in $Q_N^{\min}$ would increase λ_N of larger cells, it may also reduce Q_N^{stat} thus reducing overall nutrient uptake. A full assessment of $Q_N^{\min}$ is therefore required and so in future this should be included through parameter optimisation, particularly given wider ranges for large cells are facilitated in other models (Ward *et al.*, 2012; Ward & Follows, 2016).

Nonetheless, a large community of mixotrophs are observed in summer at three UK coastal sites. The fact that they are prevalent in the community when nutrients are depleted, yet contribute less biomass during other periods, indicates that they are strategically well-placed to deal with low-nutrient conditions. This aligns with theoretical expectations, since nitrogen can be supplemented through heterotrophy (Stoecker *et al.*, 2017). The fact that photosynthesis, nutrient uptake, and grazing controls are unable to support larger phytoplankton mean size during this period suggests that another mechanism, such as mixotrophy, is likely to drive mean size during summer.

5.4.2 Future direction of modelling

5.4.2.1 Mixotrophy

Implementing mixotrophy in size-structured NPZD models is challenging. The most detailed mechanistic attempts in resolving mixotrophy at a cellular level were made in Flynn & Mitra (2009), yet this approach has not yet gained significant traction in modelling research. The

model is complex, with a key component being the explicit modelling of food vacuole satiation and chlorophyll *a* stoichiometry. This provides a mechanistic basis for ability of mixotrophs to photosynthesise, graze, or both depending on their state and the conditions. Such is the complexity that the model is not part of a plankton food web model (e.g. NPZD), let alone a size-structured model.

More recent work (Anschütz & Flynn, 2020) incorporates different mixotrophy types (constitutive mixotroph, generalist non-constitutive mixotroph, specialist non-constitutive mixotroph) to a protist-only NPZ model, but is not presented in a plankton ecosystem-wide setting with metazoans, nor is it explicitly size structured, although some degree of size structure can be inferred from the various functional groups. The authors state that the model is fast-running and can be implemented into larger scale ecosystem models, but no examples of this were found in the literature.

The most natural extension to a NPZD-style size structured model to include mixotrophy is that of Ward & Follows (2016), who show that mixotrophy facilitates an overall global increase in microplankton, accompanied by a decrease in pico- and nanoplankton. The increase in microplankton biomass is driven by the ability to evade nutrient limitation, and so appears to be viable solution in the context of the modelling challenges faced here. At a cellular level, however, the tradeoffs associated with mixotrophy are not well-constrained, and so parameterisation is a problem. Ward & Follows (2016) therefore present a range of weighting parameters, which penalise mixotrophy to differing degrees. When no penalty is applied, mixotrophs displace both phyto- and zooplankton. Clearly, this does not reflect reality given the prevalence of phyto- and microzooplankton, and so a metabolic penalty on the generalist approach of mixotrophs is necessary in line with ecological theory (Ward & Follows, 2016). At an ecosystem-wide level, this currently appears to be the best approach to incorporate mixotrophy in the context of size structure.

If modelling mixotrophy is the main intention of the model, however, then approaches such as that of Anschütz & Flynn (2020) should be favoured. The model of Ward & Follows (2016) lacks strategic flexibility of mixotrophy, which can be gleaned by modelling different groups (e.g. differing classes of constitutive and non-constitutive cells). A possible compromise is a switching formulation so that the community of mixotrophs implicitly favours different strategies with seasonal variability: "phytoplankton that eat" may be favoured when nutrients are saturating, and "zooplankton that are photosynthetic" (Stoecker *et al.*, 2017) when nutrients are limiting. Such an approach would be analogous to the switching of herbivory and carnivory used here, which implicitly describes the varying proportion of herbivores and carnivores inside each size class[4]. At a theoretical level, the use of a switching formulation which selects autotrophy or heterotrophy in varying degrees would implicitly represent the diverse set of strategies implemented in Anschütz & Flynn (2020), but could be implemented directly to the existing size-structured formulation of Ward & Follows (2016).

In both cases, the UK coastal ecosystem provides an optimal sandbox for development of mixotro-

[4]This is a simplification which omits omnivory and detritivory.

phy in ecosystem-wide models, since the seasonal succession of mixotrophy has a strong signal and the model form and parameters can be optimised to a large number of observations. This may be particularly useful given that lab-based estimates of mixotrophic parameters are often poorly constrained (Ward & Follows, 2016), and so a Bayesian or ABC approach may be used to gather information about potential parameter values which penalise mixotrophy. Similarly, extensive observations of the taxonomic community of mixotrophs are available to those who focus on mixotrophy; these observations can be split by taxa into the different functional groups of mixoplankton.

5.4.2.2 Zooplankton taxonomic resolution

A key characteristic which underpins size structure of the model is the dependence on predator-prey scaling. With the exception of the bottom-up advantage of 6 μm phytoplankton cells photosynthesis during light-saturating conditions and 2 μm cells during light-limited periods, all biomass accumulations of the model can be explained by grazing interactions. The predator prey size ratio is at the core of food web interactions and, while a kill-the-winner active switching approach (Vallina *et al.*, 2014) may address these issues partially, the differences in size structure between the model outputs and observations would be unlikely to be resolved. Kill-the-winner is an approach which moderates diversity, but does not necessarily move the peak of the plankton biomass distribution.

Conversely, there is strong empirical evidence that functional groups may deviate substantially from allometric predator-prey ratios. In the work of Hansen *et al.* (1994), which remains central to current allometric scaling (via Banas, 2011), meroplankton and dinoflagellates do not fit with typical allometric scaling, and so are excluded from the fitting of the allometric coefficient and exponent for θ^{opt}. Despite this, these functional groups may express the similar k_Z or $G^{\max}$ (from assessment of Hansen *et al.*, 1997), with averages for each functional group typically near the best fitting allometric curve. Instead, interspecific variability between taxa within these functional groups appears to drive variations from the best-fitting allometric curve, meaning that functional groups should lie along the best-fitting line, on average. As such, incorporating different functional groups in a model may not incur too many additional parameters: a different allometric scaling for θ^{opt} may be necessary, but shared values of k_Z and $G^{\max}$ could be used. For our model such as ours to accurately resolve the size structure of zooplankton and food web interactions above the second trophic level, taxonomic variability in θ^{opt} must be incorporated.

5.4.2.3 Oceanography and nutrient dynamics

The model is not able to fully resolve the nutrient depletion is summer, and this is likely necessitated during parameter optimisation so that the large cells are not excessively penalised (and good fit is maintained). However, the model is also unable to resolve the extent of nutrient repletion in winter, with observations ($\sim 8\,\text{mmol N}\,\text{m}^{-3}$) approximately double the model output ($4\,\text{mmol N}\,\text{m}^{-3}$). While the large cells are able to recover λ_N sufficiently in winter, indicating good qualitative model fit, it would be beneficial to improve the winter fit of nutrient

concentrations to ensure robustness of model dynamics.

A reasonable possibility is that ω itself varies seasonally, due to changes in oceanographic conditions. In particular, storm-induced winter mixing may account for greater mixing within the ML and resultant increased turbulence at the interface of the deep layer and the ML. As discussed in Section 1.4.1.1, winter has increased wind speeds (Met Office, 2020) and wave height (Marine Scotland, 2020). At all sites, enhanced seawater attenuation in winter is likely a signal of enhanced mixing suspending matter from the seabed, although other factors such as enhanced loading from rivers in winter (Lyons *et al.*, 1993) may also contribute to enhanced k_w. In summer, stable and stratified conditions may drive reduced cross-thermocline mixing, relative to winter values. In both instances, a variation of ω to reflect the possible seasonal effects may result in improved fit of the nutrient cycle.

Attaining and justifying the variation in ω likely requires the use of an ocean model, such as the Scottish Shelf Model (De Dominics *et al.*, 2018) or West Coast of Scotland Coastal Ocean Modelling System (Aleynik *et al.*, 2016). An estimate of ω may be derived from these model outputs, although vertical eddy diffusivity is not directly available, so some other proxy must be used. Otherwise, deriving an estimated seasonal cycle of ω via ABC or some similar means may be practically feasible. A periodic seasonal effect with predetermined shape would be relatively parameter-light, only incurring additional parameters for the amplitude and timing of maxima (and subsequent minima) in ω. Allowing the amplitude of the seasonal effect to be zero when defining the prior for ABC would also provide a rigorous statistical basis to test the validity of potentially varying ω.

From a strictly physical oceanographic perspective, the use of the WeStCOMS as part of a depth-resolved 1D, or sea loch-resolved 3D ecosystem model may also be a necessity when modelling at LE, due to the constant estimated MLD year-round, which almost certainly oversimplifies seasonal variability. For example, the seasonal pattern of water column-averaged density is reversed compared to SH & PL; at those sites, the density is higher in winter, while at LE, it is higher in summer (Figure 2.8). The consequence, in this model formulation, is that the processes of entrainment and export due to shoaling and deepening of the ML are omitted.

Given that the flushing rate of LE is on average 4 days (Bresnan *et al.*, 2015), exchange of plankton and nutrients with the wider coastal region is likely to play an important role in structuring the community. As discussed in Section 1.4.1.1, a key difference between horizontal and cross-thermocline vertical exchange is that the exchanging concentration of nutrients and plankton differs. For vertical exchange it is reasonable to assume that nutrients are replete and that there are no plankton in the water; on the other hand, horizontal fluxes tend to have their own plankton and can even drive substantial variability on both short (Cross *et al.*, 2015) and long timescales (Gao *et al.*, 2021). A description of physical plankton dynamics already exists in both Ward *et al.* (2012) and Ward & Follows (2016), so it would be possible to couple the updated biology presented here (such as the allometric tradeoff for α) to the spatial dynamics of such a model. An adjustment, however, would need to be made at the boundary conditions of the modelled domain, since the inward flow of water at the model boundary would require some

assumed plankton community.

In the absence of flow cytometry and phytoflagellate data at LE, optimising parameters to the model will be challenging, and will require tweaks to the approach used here. Either community level summary statistics only inclusive of phytoplankton $> 6\,\mu$m are used, or a different formulation for $\mathcal{D}$ is used altogether which matches directly observations along seasonally- or monthly-averaged biomass-size distributions, rather than summary statistics. Furthermore, incorporating spatial resolution − particularly at the 3D scale − would vastly increase computational expense of the model, and so would restrict the scope of any parameter optimisation regime.

The main area of fitting concern in spring is the immigration of meroplankton, such as barnacle larvae, which may also require a 3D modelling approach. We have presented deviations of θ^{opt} allometry as a mechanism affecting the accuracy of modelled food web dynamics. However, there is no guarantee that a modelling approach which incorporates different allometric relationships for θ^{opt} would be able to resolve the biomass of these organisms accurately. Specifically, it is unclear whether these organisms would have a sufficient competitive advantage to accumulate biomass in the current formulation of the model. Assuming size scaling of $G^{\max}$ and k_Z across functional groups is fixed, the meroplankton would have comparatively higher affinity but lower maximal grazing rate compared to the $\sim 30\,\mu$m microzooplankton who graze on $\sim 6\,\mu$m cells under the average scaling of θopt. A further consideration is that the immigrating plankton are not restricted to the meroplankton. While our modelling efforts have primarily been focused at PL, at SH & LE large numbers of *Calanus helgolandicus* also immigrate to the plankton community in spring (Heath *et al.*, 1999). The overall impact of horizontal immigration of species on size structure displays a signal in observations, but from a mechanistic perspective, it is unclear whether it is the migration itself or the relative fitness of these species that lead to the observed biomass contributions.

5.4.3 Summary of decisions on hypotheses

The following findings are conditional on the validity of the model and its parameters, and may be enriched or contested by future modelling research.

H1 Phytoplankton biomass is lower at SH than PL in winter because of light limitation. This is driven primarily by reduced surface irradiance, but enhanced seawater attenuation and a deeper mixed layer also contribute to the reduction of photosynthesis below the surface. There is no evidence of a difference in top-down control in the model.

H2 Phytoplankton mean size in winter is small as a result of allometric tradeoff for photosynthesis. Primarily, the package effect favours small cells and penalises larger cells. However, the unimodal allometry of maximal growth rate allows small nanoplankton to continue dominating the community.

H3 During the spring bloom, the small nanoplankton are able to dominate the community because of the unimodal tradeoff for maximum growth rate. However, grazing on larger nanoplankton and microplankton becomes decoupled with growth, allowing these cells to

accumulate biomass. The balancing of the package effect which limits photosynthesis and top down control become crucial in structuring the microphytoplankton community. Without the allometric tradeoff for intracellular shading, the biomass of microphytoplankton is overestimated by the model.

H4 The model is unable to resolve the sustained large mean phytoplankton size in summer. This aligns with the set of allometric tradeoffs when nutrients are limiting, under the assumption that evasion of grazing in spring is only a temporary state that eventually becomes controlled. We speculate that, in line with taxonomic observations at all three UK coastal sites, mixotrophy plays a crucial role in facilitating the sustained large mean size. Strategically, the use of mixotrophy during nutrient-deplete periods aligns with theory and modelling works.

H5 The model does not accurately resolve the increased phytoplankton size diversity in autumn, but we speculate that a better-fitting model may more accurately resolve the effects of nutrient depletion on the community. Based on estimates of prey fields from observations, it is not clear if a change in top-down control occurs during this period. However, grazing by mixotrophs on small nanoplankton may be responsible for further reduced biomass of the dominant size class, thus increasing size diversity.

H6 The reduction in micro- and mesozooplankton mean size in spring does not align with predictions under allometric theory, since the phytoplankton mean size is increasing. We speculate that deviations from the standard scaling of the optimal predator-prey size ratio, particularly by dinoflagellates and meroplankton, have a crucial role in structuring the zooplankton community.

5.5 Conclusion

The cycle of plankton community size structure is driven by bottom-up and top-down controls to varying degrees over the course of the year. In terms of both biomass and size structure, we find that the winter community is bottom-up limited by light, resulting in low biomass overall, and particularly low biomass of large cells due to the package effect. The unimodal allometric tradeoff for maximum growth rate has a year-round effect on the plankton community at PL, reflected by the continual dominance of the small nanoplankton. Initially, the increase in phytoplankton mean size is driven by evasion of grazing, and the package effect is a necessary constraint on these cells to limit growth, which otherwise would be overestimated. However, as grazing regains control of the phytoplankton community which simultaneously becomes increasingly nutrient limited, another mechanism is necessary to sustain both biomass and a large mean size. The model is unable to resolve this with a traditional two-guild phyto- and zooplankton design, and we speculate that mixotrophy is a necessary strategy, in line with observations and theory. Additionally, while the overall top-down control on phytoplankton appears accurately modelled, the allocation of zooplankton biomass has some discrepancies with observations at the finer scale. We speculate that many of these differences are driven by deviations from the average allometric predator-prey

size ratio, which combined with mixotrophy should be the focus for future modelling efforts. Since the model presented here resolves the winter and spring phytoplankton community accurately, these additions would make a natural extension our work, and can be implemented using existing frameworks already established in the literature.

Future outlook

Interannual variability and climate change

Having established the 'typical seasonal cycle', a natural shift of focus is towards the long-term trends in size structure. This is particularly relevant in the context of climate change, which is expected to affect plankton globally due to acidification and temperature (Raven & Beardall, 2021), as well as changes in oceanic currents such as the AMOC (although the extent of projected changes vary widely; Baker *et al.*, 2025). Assessment of the taxonomic composition of the plankton community at LE in recent years has revealed that declines in key copepod species (*Acartia clausi*, *Temora longicornis*, *Oithona* spp., and *Paracalanus parvus*) have driven a reduction in zooplankton abundance across the $2011-2017$ period (Wells *et al.*, 2022). Similarly, at SH, shifts in the composition of the diatom community were observed before 2009; for example, an apparent decrease in May abundance of *Chaetoceros* spp. was observed in the $2002-2006$ period compared to $1998-2001$ (Bresnan *et al.*, 2009). Indeed, our assessment of the phytoplankton community at PL across separate 5-year periods indicated that the most recent study period ($2015-2019$) had larger mean size of large nano- and microphytoplankton (Figure 2.20). To a lesser degree, a similar signal is observed at SH. At PL in winter, the metazooplankton community also appears to have substantially larger mean size in the $2015-2019$ period than in $2000-2004$. These are early exploratory indications that the UK plankton community may be subject to significant changes in size structure, which motivates future study.

Climate impacts are also likely to manifest as extreme weather events, such as marine heatwaves. Marine heatwaves are expected worsen across the North Atlantic Ocean (England *et al.*, 2025), although to a lesser degree in UK coastal waters (Jacobs *et al.*, 2024b). The effect of marine heatwaves on the UK phytoplankton community remains unclear through analysis of chl *a* data (Jacobs *et al.*, 2024b), although no known work has assessed the impact of marine heatwaves on size structure in the region. Work at the global scale has indicated that marine heatwaves can trigger a reduction in phytoplankton mean size in low-to-mid latitude regions (Zhan *et al.*, 2024; Li *et al.*, 2026), and increase mean size in high latitude regions (Zhan *et al.*, 2026). The authors of the latter study hypothesise that the increase in mean size is driven by enhanced light. Given that the UK coastal region is comparatively resilient to marine heatwaves (Jacobs *et al.*, 2024b), it may be of interest to assess differences in effects on the plankton community compared to other regions. However, doing so with weekly sampled data may be difficult, since the signal may be short-lived, or evolving on a short timescale.

Considerations regarding sampling bias

Concerns about the potential under-representation of large cells in traditional phytoplankton samples are historic in the literature (Goldman, 1993; Rodríguez-Ramos *et al.*, 2014). Yet, until now, no known work has assessed the potential effects of size-related sampling bias on estimation of phytoplankton community size structure. Our results suggests that caution must be exercised when assessing the size structure of phytoplankton measured by traditional inverted microscopy methods. By assessing paired 1 mL and 50 mL subsamples of the phytoplankton community at SH & LE, we showed that large cells tend to be recorded successfully less often than small cells. This is a result of the abundance-size relationship (typically with exponent -0.75; Cermeño *et al.*, 2006; Cermeño & Figueiras, 2008) and, while large cells may contribute comparatively small numerical abundance, their biomass contribution per cell is large much larger than small cells. Our simulation experiment suggests that this drives bias in the estimation of biomass and biomass-weighted mean size and size variance. Particular caution is required when subsample volume is below 10 mL in winter for sites similar to PL (Figure 3.9), while lower subsample volumes may be used in higher biomass periods such as summer.

One of the key findings of Rodríguez-Ramos *et al.* (2014), who focused on subsample volume effects on taxonomic diversity, is that patchiness affects the measured phytoplankton diversity. Since these taxonomic results are also driven by the abundance-size distribution, and thus associated with size, these findings are likely to be analogous to ours. As such, the known existence of patchiness (Mackas *et al.*, 1985; Davis *et al.*, 2004; McGillicuddy & Franks, 2019; Barth & Stone, 2024) and clustering (Sandgren & Robinson, 1984; Haberyan & Haddock, 2013; Jacobs *et al.*, 2024a) mean that our estimates of bias are likely to be a lower bound rather than an accurate estimate. Future studies including the effects of patchiness and clustering are likely to reveal that the extent of bias is actually greater than the estimates quoted here.

Parameter optimisation of plankton ecosystem models

It must be acknowledged that a range of excellent size-structured plankton ecosystem models already exist (e.g., Ward *et al.*, 2012; Serra-Pompei *et al.*, 2020; Terseleer *et al.*, 2014). While there remains an appetite to improve models further, e.g. by the expansion of traits in trait-based models (Le Gland *et al.*, 2021) or inclusion of mixotrophy (Ward & Follows, 2016), often less focus is drawn to the process of parameter optimisation. Beyond the fact this topic may be seen as less interesting or less revealing to many, there are perhaps two good reason for this. Firstly, some models are huge, and thus the process of optimising parameters would be computationally expensive, if possible at all. Secondly, the optimisation of parameters is conditional on the model itself, and so the fact that parameters vary from one model the next may lead researchers to the mistaken conclusion that parameters are meaningless or invalid (e.g., Gutenkunst *et al.*, 2007).

The latter claim is exactly why a Bayesian approach should be central to parameter optimisation of future plankton ecosystem models. The process of Bayesian inference allows prior knowledge to be incorporated into the process of optimisation, which can be combined with newly acquired information from a new model, or an existing model's implementation at a sampling site. After

all, residual variability around fitted values of parameters for individual species is common (e.g., Hansen *et al.*, 1994, 1997; Litchman *et al.*, 2007; Edwards *et al.*, 2015), and so it follows that community-averaged values will vary from one region to the next. The fact that the parameters themselves are modelled as random variables under a Bayesian framework explicitly provides a weighting of probability across a range of values, rather than an isolated value with no substantial context. Furthermore, such a process is also transparent and reproducible, aligning with core values of scientific study.

Additionally, there is a case that as models become more expensive, it becomes increasingly beneficial to have an automated regime rather than using a manual tuning approach. If economically viable and accessible, a supercomputer can be used to undertake the optimisation process at speed and in the background. The difficulty in this sense becomes the upfront expense of the chosen regime. ABC, as used here, is the least expensive at the initial stage because the acceptance threshold can be chosen post-hoc (with care to avoid researcher bias). However, more complex methods, such as those involving the Markov chain, can drastically reduce overall computational expense and are necessitated when the parameter space is large, despite the increased up-front expense.

Next steps for plankton ecosystem modelling in UK coastal waters

In this thesis we have provided a mechanistic assessment of several patterns observed of the seasonal cycle of size structure (H1−H3). However, we were not able to directly address some other mechanistic hypotheses (H4−H6). Based on our assessment of the model, we hypothesise that the poor fit in these periods is driven by the lack of mixotrophy in the model, and deviations from the typical allometric scaling of the predator-prey size ratio (by dinoflagellates and meroplankton larvae; Hansen *et al.*, 1994). Mixoplankton, heterotrophic dinoflagellates, and meroplankton all play a key role in the UK coastal plankton community (Figures 2.12 & 2.13), so the construction of a size-structured model which accurately represents these groups should be a central focus for future modelling work. Deviations from the predator-prey ratio may be included by the inclusion of taxonomically resolved zooplankton classes each resolved by size, while mixotrophy may be included following a similar strategy to Ward & Follows (2016). Both of these additions will incur additional computational expense, slowing the model down and reducing the number of runs undertaken for parameter optimisation. As such, posterior estimates presented for free parameters (Table 4.4) may be used for priors in future optimisation with the aim of reducing the number of necessary model runs. Based on the various relationships expressed between parameters (Table 4.5), we expect that an updated posterior estimate from an updated model is likely to differ from the current posterior estimate.

Once the steps above have been completed, the model should also be optimised at SH (and LE, if flow cytometry data become available). Doing so may require an adjustment to the existing criteria of the ABC regime, since summaries of size structure may be affected by sampling bias. This may be circumvented by fitting the model directly to the averaged biomass-size distribution, as opposed to summaries. However, this becomes a bivariate fitting exercise, since

variations in the location of biomass accumulation on the size domain, as well as the extent of the biomass accumulations themselves must both be accounted for. Conversely, the fitting of summary statistics is univariate, under the assumption that time is not considered a random variable itself.

A natural focus would be to implement the model in a spatially resolved model, perhaps using SSM (De Dominics *et al.*, 2018) or WeStCOMS (Aleynik *et al.*, 2016). This would allow joint fitting of outputs to data at all sites, but would increase complexity and expense exponentially, reducing the scope of parameter optimisation. Undertaking this task would require a strong motivation with specific, focused research objectives.

Instead, value may be gleaned from implementing tradeoffs exhibited in our model to an existing end-to-end foodweb model such as StrathE2E (Heath *et al.*, 2021). Expanding the phytoplankton pool from one state variable to even the three observed dominant size classes may be useful (approximately $2\,\mu$m, $6\,\mu$m, and $30\,\mu$m), particularly alongside a mixotroph class ($30\,\mu$m). For zooplankton, size is implicitly modelled to some degree by the inclusion of different functional groups, although nanozooplankton ($\sim 6\,\mu$m) may also be added. The inclusion of size structure in such a model would help to contextualise the true importance of plankton community size structure in the wider foodweb.

Bibliography

Acevedo-Trejos, E., Brandt, G., Bruggeman, J., & Merico, A. 2015. Mechanisms shaping size structure and functional diversity of phytoplankton communities in the ocean. *Scientific Reports*, **5**(1), 1–8.

Acevedo-Trejos, E., Marañón, E., & Merico, A. 2018. Phytoplankton size diversity and ecosystem function relationships across oceanic regions. *Proceedings of the Royal Society B: Biological Sciences*, **285**(1879), 1–9.

Acevedo-Trejos, E., Cadier, M., Chakraborty, S., Chen, B., Cheung, S. Y., Grigoratou, M., Guill, C., Hassenrück, C., Kerimoglu, O., Klauschies, T., Lindemann, C., Palacz, A., Ryabov, A., Scotti, M., Smith, S. L., Våge, S., & Prowe, F. 2022. Modelling approaches for capturing plankton diversity (MODIV), their societal applications and data needs. *Frontiers in Marine Science*, **9**, 975414.

Akitsu, T. K., Nasahara, K. N., Ijima, O., Hirose, Y., Ide, R., Takagi, K., & Kume, A. 2022. The variability and seasonality in the ratio of photosynthetically active radiation to solar radiation: A simple empirical model of the ratio. *International Journal of Applied Earth Observation and Geoinformation*, **108**, 102724.

Aksnes, D., & Cao, F. 2011. Inherent and apparent traits in microbial nutrient uptake. *Marine Ecology Progress Series*, **440**, 41–51.

Aleynik, D., Dale, A. C., Porter, M., & Davidson, K. 2016. A high resolution hydrodynamic model system suitable for novel harmful algal bloom modelling in areas of complex coastline and topography. *Harmful Algae*, **53**, 102–117.

Allen, J., Howland, R., Bloomer, N., & Uncles, R. 1999. Simulating the Spring Phytoplankton Bloom in the Humber Plume, UK. *Marine Pollution Bulletin*, **37**(3-7), 295–305.

Anderson, S. R., & Menden-Deuer, S. 2017. Growth, Grazing, and Starvation in Three Heterotrophic Dinoflagellate Species. *Journal of Eukaryotic Microbiology*, **64**, 213–225.

Anderson, T. R., Gentleman, W. C., & Yool, A. 2015. EMPOWER-1.0: an Efficient Model of Planktonic ecOsystems WrittEn in R. *Geoscientific Model Development*, **8**(7), 2231–2262.

Anschütz, A. A., & Flynn, K. J. 2020. Niche separation between different functional types of mixoplankton: results from NPZ-style N-based model simulations. *Marine Biology*, **167**(1), 3.

Arima, D., Yamaguchi, A., Yoshiyuki, A., Matsuno, K., Saito, R., Asami, H., Shimada, H., & Imai, I. 2014. Seasonal changes in zooplankton community structure in Ishikari Bay, Japan Sea. *Bulletin of Fisheries Sciences, Hokkaido University*, **64**(1), 17–23.

Armengol, L., Calbet, A., Franchy, G., Rodríguez-Santos, A., & Hernández-León, S. 2019. Planktonic food web structure and trophic transfer efficiency along a productivity gradient in the tropical and subtropical Atlantic Ocean. *Scientific Reports*, **9**(1), 2044.

Arrhenius, F. 1996. Diet composition and food selectivity of 0-group herring (*Clupea harengus* L.) and sprat (*Sprattus sprattus* L.) in the northern Baltic Sea. *ICES Journal of Marine Science*, **53**(4), 701–712.

Atkinson, A., Rossberg, A. G., Gaedke, U., Sprules, G., Heneghan, R. F., Batziakas, S., Grigoratou, M., Fileman, E., Schmidt, K., & Frangoulis, C. 2024. Steeper size spectra with decreasing phytoplankton biomass indicate strong trophic amplification and future fish declines. *Nature Communications*, **15**(1), 381.

Bach, L. T., Riebesell, U., Sett, S., Febiri, S., Rzepka, P., & Schulz, K. G. 2012. An approach for particle sinking velocity measurements in the 3–400 μm size range and considerations on the effect of temperature on sinking rates. *Marine Biology*, **159**(8), 1853–1864.

Baird, M. E., Oke, P. R., Suthers, I. M., & Middleton, J. H. 2004. A plankton population model with biomechanical descriptions of biological processes in an idealised 2D ocean basin. *Journal of Marine Systems*, **50**(3-4), 199–222.

Baker, J. A., Bell, M. J., Jackson, L. C., Vallis, G. K., Watson, A. J., & Wood, R. A. 2025. Continued Atlantic overturning circulation even under climate extremes. *Nature*, **638**(8052), 987–994.

Balls, P. 1994. Nutrient Inputs to Estuaries from Nine Scottish East Coast Rivers; Influence of Estuarine Processes on Inputs to the North Sea. *Estuarine, Coastal and Shelf Science*, **39**(4), 329–352.

Ban, S. 1994. Effect of temperature and food concentration on post-embryonic development, egg production and adult body size of calanoid copepod *Eurytemora affinis*. *Journal of Plankton Research*, **16**(6), 721–735.

Banas, N. S. 2011. Adding complex trophic interactions to a size-spectral plankton model: Emergent diversity patterns and limits on predictability. *Ecological Modelling*, **222**(15), 2663–2675.

Banas, N. S., Møller, E. F., Nielsen, T. G., & Eisner, L. B. 2016. Copepod Life Strategy and Population Viability in Response to Prey Timing and Temperature: Testing a New Model across Latitude, Time, and the Size Spectrum. *Frontiers in Marine Science*, **3**, 225.

Banse, K. 1994. Grazing and Zooplankton Production as Key Controls of Phytoplankton Production in the Open Ocean. *Oceanography*, **7**(1), 13–20.

Barth, A., & Stone, J. 2024. Understanding the picture: the promise and challenges of in-situ

imagery data in the study of plankton ecology. *Journal of Plankton Research*, **46**(4), 365–379.

Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. 2010. Patterns of Diversity in Marine Phytoplankton. *Science*, **327**(5972), 1509–1511.

Bashevkin, S. M., Dibble, C. D., Dunn, R. P., Hollarsmith, J. A., Ng, G., Satterthwaite, E. V., & Morgan, S. G. 2020. Larval dispersal in a changing ocean with an emphasis on upwelling regions. *Ecosphere*, **11**(1), e03015.

Beardall, J., Allen, D., Bragg, J., Finkel, Z. V., Flynn, K. J., Quigg, A., Rees, T. A. V., Richardson, A., & Raven, J. A. 2009. Allometry and stoichiometry of unicellular, colonial and multicellular phytoplankton. *New Phytologist*, **181**(2), 295–309.

Behrenfeld, M. J. 2010. Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. *Ecology*, **91**(4), 977–989.

Behrenfeld, M. J., & Boss, E. S. 2014. Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Annual Review of Marine Science*, **6**(1), 167–194.

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. 2005. Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, **19**(1), 1–14.

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., & Boss, E. S. 2006. Climate-driven trends in contemporary ocean productivity. *Nature*, **444**(7120), 752–755.

Berggreen, U., Hansen, B., & Kiørboe, T. 1988. Food size spectra, ingestion and growth of the copepod *Acartia tonsa* during development: Implications for determination of copepod production. *Marine Biology*, **99**(3), 341–352.

Bisson, K. M., Kiko, R., Siegel, D. A., Guidi, L., Picheral, M., Boss, E., & Cael, B. B. 2022. Sampling uncertainties of particle size distributions and derived fluxes. *Limnology and Oceanography: Methods*, **20**(12), 754–767.

Blauw, A. N., Benincà, E., Laane, R. W., Greenwood, N., & Huisman, J. 2018. Predictability and environmental drivers of chlorophyll fluctuations vary across different time scales and regions of the North Sea. *Progress in Oceanography*, **161**, 1–18.

Bode, M., Hagen, W., Schukat, A., Teuber, L., Fonseca-Batista, D., Dehairs, F., & Auel, H. 2015. Feeding strategies of tropical and subtropical calanoid copepods throughout the eastern Atlantic Ocean – Latitudinal and bathymetric aspects. *Progress in Oceanography*, **138**, 268–282.

Bonnet, D., & Carlotti, F. 2001. Development and egg production in *Centropages typicus* (*Copepoda*: *Calanoida*) fed different food types: a laboratory study. *Marine Ecology Progress Series*, **224**, 133–148.

Bowers, D. G., Boudjelas, S., & Harker, G. E. L. 1998. The distribution of fine suspended sediments in the surface waters of the Irish Sea and its relation to tidal stirring. *International*

Journal of Remote Sensing, **19**(14), 2789–2805.

Branco, M., Figueiras, F. G., & Cermeño, P. 2018. Assessing the efficiency of non-parametric estimators of species richness for marine microplankton. *Journal of Plankton Research*, **40**(3), 230–243.

Branco, P., Egas, M., Hall, S. R., & Huisman, J. 2020. Why Do Phytoplankton Evolve Large Size in Response to Grazing? *The American Naturalist*, **195**(1), E20–E37.

Brandão, M. C., Benedetti, F., Martini, S., Soviadan, Y. D., Irisson, J. O., Romagnan, J. B., Elineau, A., Desnos, C., Jalabert, L., Freire, A. S., Picheral, M., Guidi, L., Gorsky, G., Bowler, C., Karp-Boss, L., Henry, N., de Vargas, C., Sullivan, M. B., Tara Oceans Consortium Coordinators, Acinas, S. G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., de Vargas, C., Gorsky, G., Guidi, L., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Poulton, N., Pesant, S., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M. B., Sunagawa, S., Wincker, P., Stemmann, L., & Lombard, F. 2021. Macroscale patterns of oceanic zooplankton composition and size structure. *Scientific Reports*, **11**(1), 15714.

Brennan, C. E., Maps, F., Lavoie, D., Plourde, S., & Johnson, C. L. 2024. Modelling the complete life cycle of an arctic copepod reveals complex trade-offs between concurrent life cycle strategies. *Progress in Oceanography*, **229**, 103333.

Bresnan, E., Hay, S., Hughes, S., Fraser, S., Rasmussen, J., Webster, L., Slesser, G., Dunn, J., & Heath, M. 2009. Seasonal and interannual variation in the phytoplankton community in the north east of Scotland. *Journal of Sea Research*, **61**(1-2), 17–25.

Bresnan, E., Cook, K. B., Hughes, S. L., Hay, S. J., Smith, K., Walsham, P., & Webster, L. 2015. Seasonality of the plankton community at an east and west coast monitoring site in Scottish waters. *Journal of Sea Research*, **105**, 16–29.

Brierley, A. S. 2014. Diel vertical migration. *Current Biology*, **24**(22), R1074–R1076.

Brierley, A. S. 2017. Plankton. *Current Biology*, **27**(11), R478–R483.

Bristow, L. A., Mohr, W., Ahmerkamp, S., & Kuypers, M. M. 2017. Nutrients that limit growth in the ocean. *Current Biology*, **27**(11), R474–R478.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. 2004. Toward a metabolic theory of ecology. *Ecology*, **85**(7), 1771–1789.

Brunet, C., Brylinski, J., Bodineau, L., Thoumelin, G., Bentley, D., & Hilde, D. 1996. Phytoplankton Dynamics During the Spring Bloom in the South-eastern English Channel. *Estuarine, Coastal and Shelf Science*, **43**(4), 469–483.

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., Bishop, J. K. B., Casciotti, K. L., Dehairs, F., Elskens, M., Honda, M., Karl, D. M., Siegel, D. A., Silver, M. W., Steinberg, D. K., Valdes, J., Van Mooy, B., & Wilson, S. 2007. Revisiting Carbon Flux Through the Ocean's Twilight Zone. *Science*, **316**(5824), 567–570.

Buitenhuis, E. T., Rivkin, R. B., Sailley, S., & Le Quéré, C. 2010. Biogeochemical fluxes through microzooplankton. *Global Biogeochemical Cycles*, **24**(4).

Burd, B., Barnes, P., Wright, C., & Thomson, R. 2008. A review of subtidal benthic habitats and invertebrate biota of the Strait of Georgia, British Columbia. *Marine Environmental Research*, **66**, S3–S38.

Buskey, E. J., & Hyatt, C. J. 2006. Use of the FlowCAM for semi-automated recognition and enumeration of red tide cells (*Karenia brevis*) in natural plankton samples. *Harmful Algae*, **5**(6), 685–692.

Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., De Mora, L., Polimene, L., Sailley, S., Stephens, N., & Torres, R. 2016. ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. *Geoscientific Model Development*, **9**(4), 1293–1339.

Cabral, H. N., & Murta, A. G. 2002. The diet of blue whiting, hake, horse mackerel and mackerel off Portugal. *Journal of Applied Ichthyology*, **18**(1), 14–23.

Calbet, A. 2001. Mesozooplankton grazing effect on primary production: A global comparative analysis in marine ecosystems. *Limnology and Oceanography*, **46**(7), 1824–1830.

Candy, S., & Kawaguchi, S. 2006. Modelling growth of Antarctic krill. II. Novel approach to describing the growth trajectory. *Marine Ecology Progress Series*, **306**, 17–30.

Carlotti, F., & Radach, G. 1996. Seasonal dynamics of phytoplankton and *Calanus finmarchicus* in the North Sea as revealed by a coupled one-dimensional model. *Limnology and Oceanography*, **41**(3), 522–539.

Carr, M. E., Friedrichs, M. A., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K. R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E. T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quéré, C., Lohrenz, S., Marra, J., Mélin, F., Moore, K., Morel, A., Reddy, T. E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K., & Yamanaka, Y. 2006. A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, **53**(5-7), 741–770.

Carvalho, M. C., Schulz, K. G., & Eyre, B. D. 2017. Respiration of new and old carbon in the surface ocean: Implications for estimates of global oceanic gross primary productivity. *Global Biogeochemical Cycles*, **31**(6), 975–984.

Castillo-Ramírez, A., Santamaría-del Ángel, E., González-Silvera, A., Frouin, R., Sebastiá-Frasquet, M.-T., Tan, J., Lopez-Calderon, J., Sánchez-Velasco, L., & Enríquez-Paredes, L. 2020. A New Algorithm to Estimate Diffuse Attenuation Coefficient from Secchi Disk Depth. *Journal of Marine Science and Engineering*, **8**(8), 558.

Cavender-Bares, K. K., Rinaldo, A., & Chisholm, S. W. 2001. Microbial size spectra from natural and nutrient enriched ecosystems. *Limnology and Oceanography*, **46**(4), 778–789.

Cermeño, P., & Figueiras, F. 2008. Species richness and cell-size distribution: size structure of phytoplankton communities. *Marine Ecology Progress Series*, **357**, 79–85.

Cermeño, P., Marañón, E., Harbour, D., & Harris, R. P. 2006. Invariant scaling of phytoplankton abundance and cell size in contrasting marine environments. *Ecology Letters*, **9**(11), 1210–1215.

Cermeño, P., Teixeira, I. G., Branco, M., Figueiras, F. G., & Marañón, E. 2014. Sampling the limits of species richness in marine phytoplankton communities. *Journal of Plankton Research*, **36**(4), 1135–1139.

Chan, A. T. 1978. Comparative physiological study of marine diatoms and dinoflagellates in relation to irradiance and cell size. I. Growth under continuous light. *Journal of Phycology*, **14**(4), 396–402.

Charalampous, E., Matthiessen, B., & Sommer, U. 2018. Light effects on phytoplankton morphometric traits influence nutrient utilization ability. *Journal of Plankton Research*, **40**(5), 568–579.

Chen, B., & Laws, E. A. 2017. Is there a difference of temperature sensitivity between marine phytoplankton and heterotrophs? *Limnology and Oceanography*, **62**(2), 806–817.

Chen, B., & Liu, H. 2010. Relationships between phytoplankton growth and cell size in surface oceans: Interactive effects of temperature, nutrients, and grazing. *Limnology and Oceanography*, **55**(3), 965–972.

Chen, B., Liu, H., Landry, M. R., Dai, M., Huang, B., & Sune, J. 2009. Close coupling between phytoplankton growth and microzooplankton grazing in the western South China Sea. *Limnology and Oceanography*, **54**(4), 1084–1097.

Chen, B., Landry, M. R., Huang, B., & Liu, H. 2012. Does warming enhance the effect of microzooplankton grazing on marine phytoplankton in the ocean? *Limnology and Oceanography*, **57**(2), 519–526.

Chen, G., & Rynearson, T. A. 2016. Genetically distinct populations of a diatom co-exist during the North Atlantic spring bloom. *Limnology and Oceanography*, **61**, 2165–2179.

Chiswell, S. M., Calil, P. H., & Boyd, P. W. 2015. Spring blooms and annual cycles of phytoplankton: a unified perspective. *Journal of Plankton Research*, **37**(3), 500–508.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. 2009. Power-Law Distributions in Empirical Data. *SIAM Review*, **51**(4), 661–703.

Cloern, J. E. 1982. Does the Benthos Control Phytoplankton Biomass in South San Francisco Bay? *Marine Ecology Progress Series*, **9**, 191–202.

Cloern, J. E. 2018. Why large cells dominate estuarine phytoplankton: Large cells dominate in estuaries. *Limnology and Oceanography*, **63**(S1), S392–S409.

Coale, K. H., Johnson, K. S., Fitzwater, S. E., Gordon, R. M., Tanner, S., Chavez, F. P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., Steinberg, P., Nightingale, P., Cooper, D., Cochlan, W. P., Landry, M. R., Constantinou, J., Rollwagen, G., Trasvina, A., & Kudela, R. 1996. A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature*, **383**, 495–501.

Conover, R., & Huntley, M. 1991. Copepods in ice-covered seas—Distribution, adaptations to seasonally limited food, metabolism, growth patterns and life cycle strategies in polar seas. *Journal of Marine Systems*, **2**(1-2), 1–41.

Conover, R. J. 1966. Assimilation of organic matter by zooplankton. *Limnology and Oceanography*, **11**(3), 338–345.

Conway, D. V. P., Coombs, S. H., Lindley, J. A., & Llewellyn, C. A. 1999. Diet of mackerel (*Scomber scombrus*) larvae at the shelf-edge to the southwest of the British Isles and the incidence of piscivory and coprophagy. *Vie et Milieu / Life & Environment*, **49**(4), 213–220.

Coombs, S. H., Halliday, N. C., Conway, D. V. P., & Smyth, T. J. 2010. Sardine (Sardina pilchardus) egg abundance at station L4, Western English Channel, 1988-2008. *Journal of Plankton Research*, **32**(5), 693–697.

Coombs, S., Halliday, N., Southward, A., & Hawkins, S. 2005. Distribution and abundance of sardine (*Sardina pilchardus*) eggs in the english channel from continuous plankton recorder sampling, 1958–1980. *Journal of the Marine Biological Association of the United Kingdom*, **85**(5), 1243–1247.

Cross, J., Nimmo-Smith, W. A. M., Hosegood, P. J., & Torres, R. 2015. The role of advection in the distribution of plankton populations at a moored 1-D coastal observatory. *Progress in Oceanography*, **137**, 342–359.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**(7), 410–418.

Cáceres, C., Taboada, F. G., Höfer, J., & Anadón, R. 2013. Phytoplankton Growth and Microzooplankton Grazing in the Subtropical Northeast Atlantic. *PLoS ONE*, **8**(7), e69159.

Dahl, K., & Kirkegaard, E. 1986. *Stomach contents of mackerel, horse mackerel and whiting in the eastern part of the North Sea in July 1985.* Tech. rept. Danmarks Fiskeri-og Havundersøgelser.

Dalpadado, P., Arrigo, K. R., van Dijken, G. L., Skjoldal, H. R., Bagøien, E., Dolgov, A. V., Prokopchuk, I. P., & Sperfeld, E. 2020. Climate effects on temporal and spatial dynamics of phytoplankton and zooplankton in the Barents Sea. *Progress in Oceanography*, **185**, 102320.

Dam, H. G., & Peterson, W. T. 1988. The effect of temperature on the gut clearance rate constant of planktonic copepods. *Journal of Experimental Marine Biology and Ecology*, **123**(1), 1–14.

Dauby, P., Nyssen, F., & Broyer, C. D. 2003. Amphipods as food sources for higher trophic levels in the Southern Ocean: A synthesis. *Pages 129–134 of: Antarctic Biology in a Global Context.* Backhuys Publishers.

Dauvin, J. C. 2019. The English Channel: La Manche. *Pages 153–188 of: World Seas: an Environmental Evaluation*. Elsevier.

Davis, C. S., Hu, Q., Gallager, S. M., Tang, X., & Ashjian, C. J. 2004. Real-time observation of taxa-specific plankton distributions:. *Mar Ecol Prog Ser*.

De Dominics, M., O'Hara Murray, R., Wolf, J., & Gallego, A. 2018. *The Scottish Shelf Model 1990 – 2014 climatology v2.01*.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans Coordinators, Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., Boss, E., Follows, M., Karp-Boss, L., Krzic, U., Reynaud, E. G., Sardet, C., Sullivan, M. B., & Velayoudon, D. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**(6237), 1261605.

Decrop, W., Lagaisse, R., Mortelmans, J., Muñiz, C., Heredia, I., Calatrava, A., & Deneudt, K. 2025. Automated image classification workflow for phytoplankton monitoring. *Frontiers in Marine Science*, **12**, 1699781.

Di Pane, J., Bourdaud, P., Horn, S., Moreno, H. D., & Meunier, C. L. 2024. Global change alters coastal plankton food webs by promoting the microbial loop: An inverse modelling and network analysis approach on a mesocosm experiment. *Science of The Total Environment*, **921**, 171272.

Dolgov, A. V., Johannesen, E., Heino, M., & Olsen, E. 2010. Trophic ecology of blue whiting in the Barents Sea. *ICES Journal of Marine Science*, **67**(3), 483–493.

Dorrell, R. M., Lloyd, C. J., Lincoln, B. J., Rippeth, T. P., Taylor, J. R., Caulfield, C.-c. P., Sharples, J., Polton, J. A., Scannell, B. D., Greaves, D. M., Hall, R. A., & Simpson, J. H. 2022. Anthropogenic Mixing in Seasonally Stratified Shelf Seas by Offshore Wind Farm Infrastructure. *Frontiers in Marine Science*, **9**, 830927.

Droop, M. R. 1968. Vitamin B_{12} and Marine Ecology. IV. The Kinetics of Uptake, Growth and Inhibition in *Monochrysis Lutheri*. *Journal of the Marine Biological Association of the United Kingdom*, **48**(3), 689–733.

Durbin, E. G., & Durbin, A. G. 1978. Length and weight relationships of *Acartia clausi* from Narragansett Bay, R.I. 1. *Limnology and Oceanography*, **23**(5), 958–969.

Dutkiewicz, S., Ward, B. A., Monteiro, F., & Follows, M. J. 2012. Interconnection of nitrogen fixers and iron in the Pacific Ocean: Theory and numerical simulations. *Global Biogeochemical Cycles*, **26**(1), 2011GB004039.

Edgar, R. K., & Laird, K. 1993. Computer simulation of error rates of Poisson-based interval estimates of plankton abundance. *Hydrobiologia*, **264**(2), 65–77.

Edwards, A., & Sharples, F. 1986. *Scottish Sea Lochs: A Catalogue*. SAMS Internal Reports, no. 134. Scottish Association for Marine Science.

Edwards, A. M. 2000. The role of higher predation in plankton population models. *Journal of Plankton Research*, **22**(6), 1085–1112.

Edwards, K. F. 2019. Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences*, **116**(13), 6211–6220.

Edwards, K. F., Thomas, M. K., Klausmeier, C. A., & Litchman, E. 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. *Limnology and Oceanography*, **57**(2), 554–566.

Edwards, K. F., Thomas, M. K., Klausmeier, C. A., & Litchman, E. 2015. Light and growth in marine phytoplankton: allometric, taxonomic, and environmental variation. *Limnology and Oceanography*, **60**(2), 540–552.

Eilers, P., & Peeters, J. 1988. A model for the relationship between light intensity and the rate of photosynthesis in phytoplankton. *Ecological Modelling*, **42**(3-4), 199–215.

Ellegaard, M., Lenau, T., Lundholm, N., Maibohm, C., Friis, S. M. M., Rottwitt, K., & Su, Y. 2016. The fascinating diatom frustule—can it play a role for attenuation of UV radiation? *Journal of Applied Phycology*, **28**(6), 3295–3306.

Ellertsen, H. C. 1993. Spring blooms and stratification. *Nature*, **363**(6424), 24–24.

Eloire, D., Somerfield, P. J., Conway, D. V. P., Halsband-Lenk, C., Harris, R., & Bonnet, D. 2010. Temporal variability and community composition of zooplankton at station L4 in the Western Channel: 20 years of sampling. *Journal of Plankton Research*, **32**(5), 657–679.

Elrifi, I. R., & Turpin, D. H. 1985. Steady-state luxury consumption and the concept of optimum nutrient ratios: a study with phosphate and nitrate limited *Selenastrum minutum* (*Chlorophyta*). *Journal of Phycology*, **21**(4), 592–602.

England, M. H., Li, Z., Huguenin, M. F., Kiss, A. E., Sen Gupta, A., Holmes, R. M., & Rahmstorf, S. 2025. Drivers of the extreme North Atlantic marine heatwave during 2023. *Nature*, **642**(8068), 636–643.

Eppley, R. W. 1972. Temperature and phytoplankton growth in the sea. *Fishery Bulletin*, **70**(4), 1063–1085.

Eppley, R. W., Rogers, J. N., & McCarthy, J. J. 1969. Half-saturation constants for uptake of nitrate and ammonium by marine phytoplankton. *Limnology and Oceanography*, **14**(6), 912–920.

Eriksen, E., Prozorkevich, D., Trofimov, A., & Howell, D. 2012. Biomass of Scyphozoan Jellyfish, and Its Spatial Association with 0-Group Fish in the Barents Sea. *PLoS ONE*, **7**(3), e33050.

Evans, G. T., & Parslow, J. S. 1985. A Model of Annual Plankton Cycles. *Biological Oceanography*, **3**(3), 327–347.

Fasham, M. J. R., Ducklow, H. W., & McKelvie, S. M. 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, **48**(3), 591–639.

Fasham, M. 1995. Variations in the seasonal cycle of biological production in subarctic oceans: A model sensitivity analysis. *Deep Sea Research Part I: Oceanographic Research Papers*, **42**(7), 1111–1149.

Fenchel, T. 2001. How Dinoflagellates Swim. *Protist*, **152**(4), 329–338.

Fenchel, T. 1988. Marine plankton food chains. *Annual Review of Ecology and Systematics*, **19**(1), 19–38.

Fennel, K., Losch, M., Schröter, J., & Wenzel, M. 2001. Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, **28**(1-2), 45–63.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, **281**(5374), 237–240.

Findlay, H. S., Yool, A., Nodale, M., & Pitchford, J. W. 2006. Modelling of autumn plankton bloom dynamics. *Journal of Plankton Research*, **28**(2), 209–220.

Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., & Raven, J. A. 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of Plankton Research*, **32**(1), 119–137.

Finkel, Z. V., Katz, M. E., Wright, J. D., Schofield, O. M. E., & Falkowski, P. G. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proceedings of the National Academy of Sciences*, **102**(25), 8927–8932.

Finkel, Z. V. 2001. Light absorption and size scaling of light-limited metabolism in marine diatoms. *Limnology and Oceanography*, **46**(1), 86–94.

Finkel, Z., Irwin, A., & Schofield, O. 2004. Resource limitation alters the 3/4 size scaling of metabolic rates in phytoplankton. *Marine Ecology Progress Series*, **273**, 269–279.

Fishwick, J. 2018. *CTD profiles (depth, pressure, temperature, salinity, potential temperature, density, fluorescence, transmittance, downwelling PAR, dissolved oxygen concentration) binned to 0.5 m and 0.25 m at sites L4 and E1 in the Western English Channel between January 2002 and December 2015.*

Flynn, K. J. 2008. The importance of the form of the quota curve and control of non-limiting nutrient transport in phytoplankton models. *Journal of Plankton Research*, **30**(4), 423–438.

Flynn, K. J., & Mitra, A. 2009. Building the "perfect beast": modelling mixotrophic plankton. *Journal of Plankton Research*, **31**(9), 965–992.

Flynn, K. J., Stoecker, D. K., Mitra, A., Raven, J. A., Glibert, P. M., Hansen, P. J., Granéli, E., & Burkholder, J. M. 2013. Misuse of the phytoplankton–zooplankton dichotomy: the

need to assign organisms as mixotrophs within plankton functional types. *Journal of Plankton Research*, **35**(1), 3–11.

Fransz, H. G., & Gieskes, W. W. C. 1984. The unbalance of phytoplankton and copepods in the North Sea. *Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer*, **183**, 218–225.

Frau, D. 2025. Phytoplankton Sampling: When the Method Shapes the Message. *Limnological Review*, **25**(3), 45.

Freedman, D. A. 2006. On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors". *The American Statistician*, **60**(4), 299–302.

Friedrichs, L., Hörnig, M., Schulze, L., Bertram, A., Jansen, S., & Hamm, C. 2013. Size and biomechanic properties of diatom frustules influence food uptake by copepods. *Marine Ecology Progress Series*, **481**, 41–51.

Friedrichs, M. A., Carr, M.-E., Barber, R. T., Scardi, M., Antoine, D., Armstrong, R. A., Asanuma, I., Behrenfeld, M. J., Buitenhuis, E. T., Chai, F., Christian, J. R., Ciotti, A. M., Doney, S. C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J. K., Morel, A., O'Malley, R. T., O'Reilly, J., Saba, V. S., Schmeltz, M., Smyth, T. J., Tjiputra, J., Waters, K., Westberry, T. K., & Winguth, A. 2009. Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems*, **76**(1-2), 113–133.

Fuchs, H., & Franks, P. 2010. Plankton community properties determined by nutrients and size-selective feeding. *Marine Ecology Progress Series*, **413**, 1–15.

Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature*, **399**(6736), 541–548.

Fulton, E., & Sainsbury, K. 2024. Food web structure, the hub index and identifying species of ecological significance. *Ecological Indicators*, **166**, 112378.

Gaedke, U., & Klauschies, T. 2017. Analyzing the shape of observed trait distributions enables a data-based moment closure of aggregate models. *Limnology and Oceanography: Methods*, **15**(12), 979–994.

Gan, J., Lu, Z., Dai, M., Cheung, A. Y. Y., Liu, H., & Harrison, P. 2010. Biological response to intensified upwelling and to a river plume in the northeastern South China Sea: A modeling study. *Journal of Geophysical Research: Oceans*, **115**(C9).

Gao, S., Hjøllo, S. S., Falkenhaug, T., Strand, E., Edwards, M., & Skogen, M. D. 2021. Over-wintering distribution, inflow patterns and sustainability of Calanus finmarchicus in the North Sea. *Progress in Oceanography*, **194**, 102567.

García-Comas, C., Sastri, A. R., Ye, L., Chang, C.-Y., Lin, F.-S., Su, M.-S., Gong, G.-C., & Hsieh, C.-h. 2016. Prey size diversity hinders biomass trophic transfer and predator size diversity promotes it in planktonic communities. *Proceedings of the Royal Society B: Biological*

Sciences, **283**(1824), 20152129.

García-Oliva, O., Hantzsche, F. M., Boersma, M., & Wirtz, K. W. 2022. Phytoplankton and particle size spectra indicate intense mixotrophic dinoflagellates grazing from summer to winter. *Journal of Plankton Research*, **44**(2), 224–240.

Gasol, J. M., Del Giorgio, P. A., & Duarte, C. M. 1997. Biomass distribution in marine planktonic communities. *Limnology and Oceanography*, **42**(6), 1353–1363.

Geider, R. J. 1987. Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. *New Phytologist*, **106**(1), 1–34.

Geider, R. J., Maclntyre, H. L., & Kana, T. M. 1998. A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, **43**(4), 679–694.

Geider, R., MacIntyre, H., & Kana, T. 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. *Marine Ecology Progress Series*, **148**, 187–200.

Gentleman, W., Leising, A., Frost, B., Strom, S., & Murray, J. 2003. Functional responses for zooplankton feeding on multiple resources: a review of assumptions and biological dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography*, **50**(22-26), 2847–2875.

Giraldo, C., Cresson, P., MacKenzie, K., Fontaine, V., Loots, C., Delegrange, A., & Lefebvre, S. 2024. Insights into planktonic food-web dynamics through the lens of size and season. *Scientific Reports*, **14**(1), 1684.

Goebel, N., Edwards, C., Zehr, J., Follows, M., & Morgan, S. 2013. Modeled phytoplankton diversity and productivity in the California Current System. *Ecological Modelling*, **264**, 37–47.

Goldman, J. C. 1993. Potential role of large oceanic diatoms in new primary production. *Deep Sea Research Part I: Oceanographic Research Papers*, **40**(1), 159–168.

González-García, C., Agustí, S., Aiken, J., Bertrand, A., Bittencourt Farias, G., Bode, A., Carré, C., Gonçalves-Araujo, R., Harbour, D. S., Huete-Ortega, M., Melo, P. A., Moreno-Ostos, E., Rees, A. P., Rodríguez, J., Da Silva, S., Zubkov, M., & Marañón, E. 2023. Basin-scale variability in phytoplankton size-abundance spectra across the Atlantic Ocean. *Progress in Oceanography*, **217**, 103104.

González-Gil, R., Banas, N. S., Bresnan, E., & Heath, M. R. 2022. The onset of the spring phytoplankton bloom in the coastal North Sea supports the Disturbance Recovery Hypothesis. *Biogeosciences*, **19**(9), 2417–2426.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., Dongen-Vogels, V. V., Reifel, K. M., & Behrenfeld, M. J. 2015. Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, **102**, 16–25.

Gribble, K. E. 2006. *The ecology, life history, and phylogeny of the marine thecate heterotrophic dinoflagellates Protoperidinium and Diplopsalidaceae (Dinophyceae).* Ph.D. thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Woods Hole, MA.

Gruber, N. 2008. The Marine Nitrogen Cycle: Overview and Challenges. *Nitrogen in the Marine Environment.*

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. 2007. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology*, **3**(10), e189.

Haario, H., Laine, M., Mira, A., & Saksman, E. 2006. DRAM: Efficient adaptive MCMC. *Statistics and Computing*, **16**(4), 339–354.

Haberyan, K. A., & Haddock, G. D. 2013. On the distribution of diatoms on cover slips: cause for concern. *Journal of Paleolimnology*, **50**(1), 151–158.

Haggan, N., & Pitcher, T. 2005. Ecosystem simulation models of Scotland's West Coast and sea lochs. *Fisheries Centre research reports.*, **13**(4).

Hamer, J., Matthiessen, B., Pulina, S., & Hattich, G. S. I. 2022. Maintenance of Intraspecific Diversity in Response to Species Competition and Nutrient Fluctuations. *Microorganisms*, **10**(1), 113.

Hansell, D., Carlson, C., Repeta, D., & Schlitzer, R. 2009. Dissolved Organic Matter in the Ocean: A Controversy Stimulates New Insights. *Oceanography*, **22**(4), 202–211.

Hansen, B., Bjornsen, P. K., & Hansen, P. J. 1994. The size ratio between planktonic predators and their prey. *Limnology and Oceanography*, **39**(2), 395–403.

Hansen, P. J. 1991. Quantitative importance and trophic role of heterotrophic dinoflagellates in a coastal pelagial food web. *Marine Ecology Progress Series*, **73**, 253–261.

Hansen, P. J. 1992. Prey size selection, feeding rates and growth dynamics of heterotrophic dinoflagellates with special emphasis on Gyrodinium spirale. *Marine Biology*, **114**, 327–224.

Hansen, P. J., & Calado, A. J. 1999. Phagotrophic Mechanisms and Prey Selection in Free-Living Dinoflagellates. *Journal of Eukaryotic Microbiology*, **46**(4), 382–389.

Hansen, P. J., Bjørnsen, P. K., & Hansen, B. W. 1997. Zooplankton grazing and growth: Scaling within the 2-2000 μm body size range. *Limnology and Oceanography*, **42**(4), 687–704.

Haraguchi, L., Jakobsen, H. H., Lundholm, N., & Carstensen, J. 2018. Phytoplankton Community Dynamic: A Driver for Ciliate Trophic Strategies. *Frontiers in Marine Science*, **5**, 272.

Harris, R. 2010. The L4 time-series: the first 20 years. *Journal of Plankton Research*, **32**(5), 577–583.

Harris, R. 1999. *ICES zooplankton methodology manual.* San Diego, Calif. London: Academic.

Harrison, P. J., Zingone, A., Mickelson, M. J., Lehtinen, S., Ramaiah, N., Kraberg, A. C., Sun, J., McQuatters-Gollop, A., & Jakobsen, H. H. 2015. Cell volumes of marine phytoplankton from globally distributed coastal data sets. *Estuarine, Coastal and Shelf Science*, **162**, 130–142.

Hatton, I. A., McCann, K. S., Fryxell, J. M., Davies, T. J., Smerlak, M., Sinclair, A. R. E., & Loreau, M. 2015. The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes. *Science*, **349**(6252), aac6284.

Head, E. J. H., & Harris, L. R. 2004. *Estimating zooplankton biomass from dry weights of groups of individual organisms.* Canadian Science Advisory Secretariat. Research Document 2004/045.

Heath, M., Robertson, W., Mardaljevic, J., & Gurney, W. 1997. Modelling the population dynamics of Calanus in the Fair Isle current of northern Scotland. *Journal of Sea Research*, **38**(3-4), 381–412.

Heath, M. R., Speirs, D. C., Thurlbeck, I., & Wilson, R. J. 2021. StrathE2E2: An R package for modelling the dynamics of marine food webs and fisheries. *Methods in Ecology and Evolution*, **12**(2), 280–287.

Heath, M., Backhaus, J., Richardson, K., McKenzie, E., Slagstad, D., Beare, D., Dunn, J., Fraser, J., Gallego, A., Hainbucher, D., Hay, S., Jónasdóttir, S., Madden, H., Mardaljevic, J., & Schacht, A. 1999. Climate fluctuations and the spring invasion of the North Sea by *Calanus finmarchicus*. *Fisheries Oceanography*, **8**(s1), 163–176.

Highfield, J. M., Eloire, D., Conway, D. V. P., Lindeque, P. K., Attrill, M. J., & Somerfield, P. J. 2010. Seasonal dynamics of meroplankton assemblages at station L4. *Journal of Plankton Research*, **32**(5), 681–691.

Hillebrand, H., Acevedo-Trejos, E., Moorthi, S. D., Ryabov, A., Striebel, M., Thomas, P. K., & Schneider, M. 2022a. Cell size as driver and sentinel of phytoplankton community structure and functioning. *Functional Ecology*, **36**(2), 276–293.

Hillebrand, H., Antonucci Di Carvalho, J., Dajka, J., Dürselen, C., Kerimoglu, O., Kuczynski, L., Rönn, L., & Ryabov, A. 2022b. Temporal declines in Wadden Sea phytoplankton cell volumes observed within and across species. *Limnology and Oceanography*, **67**(2), 468–481.

Hobbs, L. 2020. Eat or Sleep: Availability of Winter Prey Explains Mid-Winter and Spring Activity in an Arctic *Calanus* Population. *Frontiers in Marine Science*, **7**, 14.

Hobbs, L., Banas, N. S., Cohen, J. H., Cottier, F. R., Berge, J., & Varpe, 2021. A marine zooplankton community vertically structured by light across diel to interannual timescales. *Biology Letters*, **17**, 20200810.

Holligan, P. M., & Harbour, D. S. 1977. The vertical distribution and succession of phytoplankton in the western English Channel in 1975 and 1976. *Journal of the Marine Biological Association of the United Kingdom*, **57**(4), 1075–1093.

Holm-Hansen, O., Lorenzen, C. J., Holmes, R. W., & Strickland, J. D. H. 1965. Fluorometric

Determination of Chlorophyll. *ICES Journal of Marine Science*, **30**(1), 3–15.

Hrycik, A. R., Shambaugh, A., & Stockwell, J. D. 2019. Comparison of FlowCAM and microscope biovolume measurements for a diverse freshwater phytoplankton community. *Journal of Plankton Research*, **41**(6), 849–864.

Huisman, J., Van Oostveen, P., & Weissing, F. J. 1999a. Critical depth and critical turbulence: Two different mechanisms for the development of phytoplankton blooms. *Limnology and Oceanography*, **44**(7), 1781–1787.

Huisman, J., Van Oostveen, P., & Weissing, F. J. 1999b. Species Dynamics in Phytoplankton Blooms: Incomplete Mixing and Competition for Light. *The American Naturalist*, **154**(1), 46–68.

Hutchinson, G. E. 1961. The Paradox of the Plankton. *The American Naturalist*, **95**(882), 137–145.

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., Endo, H., Gasol, J. M., Gregory, A. C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco, E., Soviadan, D., Zayed, A. A., Zingone, A., Labadie, K., Ferland, J., Marec, C., Kandels, S., Picheral, M., Dimier, C., Poulain, J., Pisarev, S., Carmichael, M., Pesant, S., Acinas, S. G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., De Vargas, C., Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M. B., Sunagawa, S., Wincker, P., Babin, M., Boss, E., Iudicone, D., Jaillon, O., Acinas, S. G., Ogata, H., Pelletier, E., Stemmann, L., Sullivan, M. B., Sunagawa, S., Bopp, L., De Vargas, C., Karp-Boss, L., Wincker, P., Lombard, F., Bowler, C., & Zinger, L. 2019. Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*, **179**(5), 1084–1097.e21.

Ikedai, T., & Motoda, S. 1978. Estimated zooplankton production and their ammonia excretion in the Kuroshio and adjacent seas. *Fishery Bulletin*, **76**(2), 357–367.

Irigoien, X., Flynn, K. J., & Harris, R. P. 2005. Phytoplankton blooms: a 'loophole' in microzooplankton grazing impact? *Journal of Plankton Research*, **27**(4), 313–321.

Irigoien, X., Huisman, J., & Harris, R. P. 2004. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature*, **429**(6994), 863–867.

Irwin, A. J., Finkel, Z. V., Schofield, O. M. E., & Falkowski, P. G. 2006. Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *Journal of Plankton Research*, **28**(5), 459–471.

Ismael, A. A. 2003. Succession of heterotrophic and mixotrophic dinoflagellates as well as autotrophic microplankton in the harbour of Alexandria, Egypt. *Journal of Plankton Research*, **25**(2), 193–202.

Ito, S., Kishi, M. J., Kurita, Y., Oozeki, Y., Yamanaka, Y., Megrey, B. A., & Werner, F. E.

2004. Initial design for a fish bioenergetics model of Pacific saury coupled to a lower trophic ecosystem model. *Fisheries Oceanography*, **13**(s1), 111–124.

Jacobs, P., Serre-Fredj, L., Koeman, R. P. T., Van Den Oever, A., Peck, M. A., & Philippart, C. J. M. 2024a. Impacts of counting protocols for light microscopy on estimates of biodiversity and algal density of phytoplankton. *Limnology and Oceanography: Methods*, **22**(12), 930–942.

Jacobs, Z. L., Jebri, F., Wakelin, S., Strong, J., Popova, E., Srokosz, M., & Loveridge, A. 2024b. Marine heatwaves and cold spells in the Northeast Atlantic: what should the UK be prepared for? *Frontiers in Marine Science*, **11**, 1434365.

Jacobson, D. M., & Anderson, D. M. 1986. Thecate heterotrophic dinoflagellates: feeding behavior and mechanisms. *Journal of Phycology*, **22**(3), 249–258.

Jakobsen, H. H., & Markager, S. 2016. Carbon-to-chlorophyll ratio for phytoplankton in temperate coastalwaters: Seasonal patterns and relationship to nutrients. *Limnology and Oceanography*, **61**, 1853–1868.

James, A., Pitchford, J. W., & Brindley, J. 2003. The relationship between plankton blooms, the hatching of fish larvae, and recruitment. *Ecological Modelling*, **160**(1-2), 77–90.

Jeong, H., Yoo, Y., Park, J., Song, J., Kim, S., Lee, S., Kim, K., & Yih, W. 2005. Feeding by phototrophic red-tide dinoflagellates: five species newly revealed and six species previously known to be mixotrophic. *Aquatic Microbial Ecology*, **40**, 133–150.

Jiang, H., & Paffenhöfer, G. 2008. Hydrodynamic signal perception by the copepod *Oithona plumifera*. *Marine Ecology Progress Series*, **373**, 37–52.

Jiang, Z., Liu, J., Zhu, X., Chen, Y., Chen, Q., & Chen, J. 2020. Quantitative comparison of phytoplankton community sampled using net and water collection methods in the southern Yellow Sea. *Regional Studies in Marine Science*, **35**, 101250.

Jin, Z., Charlock, T. P., Smith, W. L., & Rutledge, K. 2004. A parameterization of ocean surface albedo. *Geophysical Research Letters*, **31**(22), 2004GL021180.

Jónasdóttir, S. H. 2019. Fatty Acid Profiles and Production in Marine Phytoplankton. *Marine Drugs*, **17**(3), 151.

Kalinina, V., Berdieva, M., Aksenov, N., & Skarlato, S. 2023. Phosphorus deficiency induces sexual reproduction in the dinoflagellate *Prorocentrum cordatum*. *Scientific Reports*, **13**(1), 14191.

Karaköylü, E. M., & Franks, P. J. S. 2012. Reassessment of copepod grazing impact based on continuous time series of in vivo gut fluorescence from individual copepods. *Journal of Plankton Research*, **34**(1), 55–71.

Kassambara, A. 2016. *ggpubr: 'ggplot2' Based Publication Ready Plots*. Institution: Comprehensive R Archive Network Pages: 0.6.2.

Kawaguchi, S., Ichii, T., & Naganobu, M. 1999. Green krill, the indicator of micro- and nano-size

phytoplankton availability to krill. *Polar Biology*, **22**(2), 133–136.

Kelley, D., Richards, C., & Scor/Iapso, W. 2015. *gsw: Gibbs Sea Water Functions*. Institution: Comprehensive R Archive Network Pages: 1.2-0.

Kim, D., Komeda, S., Matsuno, K., & Yamaguchi, A. 2025. Vertical variations in zooplankton size spectra down to 3,000 m depth and significant effects of the sizes of *Calanoida* and *Ergasilida* across the subarctic, transitional, and subtropical regions of the western North Pacific. *Deep Sea Research Part I: Oceanographic Research Papers*, **217**, 104445.

Kim, J. H., Ajani, P. A., Murray, S. A., Kang, S.-M., Kim, S.-H., Lim, H. C., Teng, S. T., Lim, P. T., & Park, B. S. 2023. Abiotic and biotic factors controlling sexual reproduction in populations of *Pseudo-nitzschia pungens* (*Bacillariophyceae*). *Harmful Algae*, **123**, 102392.

Kiørboe, T., Saiz, E., & Viitasalo, M. 1996. Prey switching behaviour in the planktonic copepod *Acartia tonsa*. *Marine Ecology Progress Series*, **143**, 65–75.

Kiørboe, T. 2008. A Mechanistic Approach to Plankton Ecology. Princeton University Press.

Kiørboe, T. 2011. How zooplankton feed: mechanisms traits and trade-offs. *Biological Reviews*, **86**, 311–339.

Kiørboe, T., & Hirst, A. G. 2014. Shifts in Mass Scaling of Respiration, Feeding, and Growth Rates across Life-Form Transitions in Marine Pelagic Organisms. *The American Naturalist*, **183**(4), E118–E130.

Kiørboe, T., & Thomas, M. K. 2020. Heterotrophic eukaryotes show a slow-fast continuum, not a gleaner–exploiter trade-off. *Proceedings of the National Academy of Sciences*, **117**(40), 24893–24899.

Kong, L.-F., He, Y.-B., Xie, Z.-X., Luo, X., Zhang, H., Yi, S.-H., Lin, Z.-L., Zhang, S.-F., Yan, K.-Q., Xu, H.-K., Jin, T., Lin, L., Qin, W., Chen, F., Liu, S.-Q., & Wang, D.-Z. 2021. Illuminating Key Microbial Players and Metabolic Processes Involved in the Remineralization of Particulate Organic Carbon in the Ocean's Twilight Zone by Metaproteomics. *Applied and Environmental Microbiology*, **87**(20), e00986–21.

Koop, J. H. E., & Grieshaber, M. K. 2000. The role of ion regulation in the control of the distribution of *Gammarus tigrinus* (Sexton) in salt-polluted rivers. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, **170**(1), 75–83.

Kottmeier, D. M., Terbrüggen, A., Wolf-Gladrow, D. A., & Thoms, S. 2020. Diel variations in cell division and biomass production of *Emiliania huxleyi* — Consequences for the calculation of physiological cell parameters. *Limnology and Oceanography*, **65**(8), 1781–1800.

Kovalenko, K. E., & Reavie, E. D. 2022. Taxonomic bias in freshwater phytoplankton communities and its effect on environmental assessment metrics. *Phycologia*, **61**(4), 436–443.

Kukla, G., & Robinson, D. 1980. Annual Cycle of Surface Albedo. *Monthly Weather Review*, **108**(1), 56–68.

Lancelot, C., Spitz, Y., Gypens, N., Ruddick, K., Becquevort, S., Rousseau, V., Lacroix, G., & Billen, G. 2005. Modelling diatom and *Phaeocystis* blooms and nutrient cycles in the Southern Bight of the North Sea: the MIRO model. *Marine Ecology Progress Series*, **289**, 63–78.

Landaeta, M. F., Suarez-Donoso, N., Bustos, C. A., & Balbontin, F. 2011. Feeding habits of larval *Maurolicus parvipinnis* (*Pisces: Sternoptychidae*) in Patagonian fjords. *Journal of Plankton Research*, **33**(12), 1813–1824.

Landry, M. R. 1981. Switching between herbivory and carnivory by the planktonic marine copepod *Calanus pacificus*. *Marine Biology*, **65**(1), 77–82.

Lavender, E., Aleynik, D., Dodd, J., Illian, J., James, M., Smout, S., & Thorburn, J. 2022. Benthic animal-borne sensors and citizen science combine to validate ocean modelling. *Scientific Reports*, **12**(1), 16613.

Le Gland, G., Vallina, S. M., Smith, S. L., & Cermeño, P. 2021. SPEAD 1.0 – Simulating Plankton Evolution with Adaptive Dynamics in a two-trait continuous fitness landscape applied to the Sargasso Sea. *Geoscientific Model Development*, **14**(4), 1949–1985.

Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E., Marshall, H., Peloquin, J., Piontkovski, S., Poulton, A. J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., Van Leeuwe, M. A., Varela, M., Widdicombe, C., & Yallop, M. 2012. A global diatom database – abundance, biovolume and biomass in the world ocean. *Earth System Science Data*, **4**(1), 149–165.

Lebour, M. V. 1933. The Importance of Larval *Mollusca* in the Plankton. *ICES Journal of Marine Science*, **8**(3), 335–343.

Leles, S. G., Mitra, A., Flynn, K. J., Stoecker, D. K., Hansen, P. J., Calbet, A., McManus, G. B., Sanders, R. W., Caron, D. A., Not, F., Hallegraeff, G. M., Pitta, P., Raven, J. A., Johnson, M. D., Glibert, P. M., & Våge, S. 2017. Oceanic protists with different forms of acquired phototrophy display contrasting biogeographies and abundance. *Proceedings of the Royal Society B: Biological Sciences*, **284**(1860), 20170664.

Lewandowska, A. M., Hillebrand, H., Lengfellner, K., & Sommer, U. 2014. Temperature effects on phytoplankton diversity — The zooplankton link. *Journal of Sea Research*, **85**, 359–364.

Li, F., Li, C., Zhou, Y., Li, W., Xian, W., Ding, W., Li, X., & Wang, J. 2026. Responses of phytoplankton functional types to marine heatwaves in China's marginal seas and adjacent waters. *Frontiers in Marine Science*, **13**, 1828780.

Li, Q., Legendre, L., & Jiao, N. 2015. Phytoplankton responses to nitrogen and iron limitation in the tropical and subtropical Pacific Ocean. *Journal of Plankton Research*, **37**(2), 306–319.

Lin, S., Litaker, R. W., & Sunda, W. G. 2016. Phosphorus physiological ecology and molecular mechanisms in marine phytoplankton. *Journal of Phycology*, **52**(1), 10–36.

Litchman, E., Klausmeier, C. A., Schofield, O. M. E., & Falkowski, P. G. 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from

cellular to ecosystem level. *Ecology Letters*, **10**, 1170–1181.

Liu, K., Chen, B., Zheng, L., Su, S., Huang, B., Chen, M., & Liu, H. 2021. What controls microzooplankton biomass and herbivory rate across marginal seas of China? *Limnology and Oceanography*, **66**(1), 61–75.

Liu, Z., Takeuchi, M., Contreras, Y., Thevar, T., Nimmo-Smith, A., Watson, J., & Giering, S. L. C. 2025. Machine learning for improved size estimation of complex marine particles from noisy holographic images. *Frontiers in Marine Science*, **12**, 1587939.

Long, J. D., Smalley, G. W., Barsby, T., Anderson, J. T., & Hay, M. E. 2007. Chemical cues induce consumer-specific defenses in a bloom-forming marine phytoplankton. *Proceedings of the National Academy of Sciences*, **104**(25), 10512–10517.

Louchart, A., Holland, M., McQuatters-Gollop, A., & Artigas, L. F. 2023. *Changes in plankton diversity*. Tech. rept. OSPAR Commission.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P. G., Cherchi, A., Bellucci, A., Gualdi, S., Masina, S., & Navarra, A. 2022. CMIP6 Simulations With the CMCC Earth System Model (CMCC-ESM2). *Journal of Advances in Modeling Earth Systems*, **14**(3), e2021MS002814.

Lucas, C. H. 2001. Reproduction and life history strategies of the common jellyfish, *Aurelia aurita*, in relation to its ambient environment. *Hydrobiologia*, **451**, 229–246.

Lund, J. W. G., Kipling, C., & Le Cren, E. D. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia*, **11**(2), 143–170.

Luo, J. 2013. Phytoplankton–zooplankton dynamics in periodic environments taking into account eutrophication. *Mathematical Biosciences*, **245**(2), 126–136.

Lynam, C. P., Hay, S. J., & Brierley, A. S. 2005. Jellyfish abundance and climatic variation: contrasting responses in oceanographically distinct regions of the North Sea, and possible implications for fisheries. *Journal of the Marine Biological Association of the United Kingdom*, **85**(3), 435–450.

Lyons, M., Balls, P., & Turrell, W. 1993. A preliminary study of the relative importance of riverine nutrient inputs to the Scottish North Sea Coastal Zone. *Marine Pollution Bulletin*, **26**(11), 620–628.

López-Urrutia, A., San Martin, E., Harris, R. P., & Irigoien, X. 2006. Scaling the metabolic balance of the oceans. *Proceedings of the National Academy of Sciences*, **103**(23), 8739–8744.

MacIntyre, H. L., Kana, T. M., Anning, T., & Geider, R. J. 2002. Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. *Journal of Phycology*, **38**(1), 17–38.

Mackas, D. L., Denman, K. L., & Abbott, M. R. 1985. Plankton patchiness: Biology in the

physical vernacular. *Bulletin of Marine Science*, **37**(2), 652–674.

Mackas, D., & Tsuda, A. 1999. Mesozooplankton in the eastern and western subarctic Pacific: community structure, seasonal life histories, and interannual variability. *Progress in Oceanography*, **43**(2-4), 335–363.

Madsen, S. D., Nielsen, T. G., & Hansen, B. W. 2001. Annual population development and production by *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus* in Disko Bay, western Greenland. *Marine Biology*, **139**(1), 75–93.

Malerba, M. E., Connolly, S. R., & Heimann, K. 2015. An experimentally validated nitrate–ammonium–phytoplankton model including effects of starvation length and ammonium inhibition on nitrate uptake. *Ecological Modelling*, **317**, 30–40.

Mantoura, R. F. C., & Llewellyn, C. A. 1983. The rapid determination of algal chlorophyll and carotenoid pigments and their breakdown products in natural waters by reverse-phase high-performance liquid chromatography. *Analytica Chimica Acta*, **151**, 297–314.

Marañón, E., Holligan, P. M., Varela, M., Mouriño, B., & Bale, A. J. 2000. Basin-scale variability of phytoplankton biomass, production and growth in the Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**(5), 825–857.

Marañón, E., Cermeño, P., Rodríguez, J., Zubkov, M. V., & Harris, R. P. 2007. Scaling of phytoplankton photosynthesis and cell size in the ocean. *Limnology and Oceanography*, **52**(5), 2190–2198.

Marañón, E., Cermeño, P., López-Sandoval, D. C., Rodríguez-Ramos, T., Sobrino, C., Huete-Ortega, M., Blanco, J. M., & Rodríguez, J. 2013. Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. *Ecology Letters*, **16**(3), 371–379.

Marin, J. M., Pudlo, P., Robert, C. P., & Ryder, R. J. 2012. Approximate Bayesian computational methods. *Statistics and Computing*, **22**(6), 1167–1180.

Marine Scotland. 2020. *Scottish Marine Assessment 2020: Wave Climate*. Tech. rept.

Marine Scotland Science. 2018a. *Scottish Coastal Observatory - Loch Ewe site data*.

Marine Scotland Science. 2018b. *Scottish Coastal Observatory - Stonehaven site data*.

Martinez, E., Antoine, D., D'Ortenzio, F., & De Boyer Montégut, C. 2011. Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s. *Journal of Geophysical Research: Oceans*, **116**(C11), 2010JC006836.

Mattern, J. P., Fennel, K., & Dowd, M. 2012. Estimating time-dependent parameters for a biological ocean model using an emulator approach. *Journal of Marine Systems*, **96-97**, 32–47.

Mayor, D. J., Sanders, R., Giering, S. L. C., & Anderson, T. R. 2014. Microbial gardening in the ocean's twilight zone: Detritivorous metazoans benefit from fragmenting, rather than ingesting, sinking detritus: Fragmentation of refractory detritus by zooplankton beneath the

euphotic zone stimulates the harvestable production of labile and nutritious microbial biomass. *BioEssays*, **36**(12), 1132–1137.

McCarthy, G., Smeed, D., Cunningham, S., & Roberts, C. 2017. Atlantic Meridonal Overturning Circulation. *MCCIP Science Review 2017*, 1–7.

McClain, C. R., Signorini, S. R., & Christian, J. R. 2004. Subtropical gyre variability observed by ocean-color satellites. *Deep Sea Research Part II: Topical Studies in Oceanography*, **51**(1), 281–301.

McCluskey, E., Brewin, R. J. W., Vanhellemont, Q., Jones, O., Cummings, D., Tilstone, G., Jackson, T., Widdicombe, C., Woodward, E. M. S., Harris, C., Bresnahan, P. J., Cyronak, T., & Andersson, A. J. 2022. On the Seasonal Dynamics of Phytoplankton Chlorophyll-a Concentration in Nearshore and Offshore Waters of Plymouth, in the English Channel: Enlisting the Help of a Surfer. *Oceans*, **3**(2), 125–146.

McEvoy, A. J., Atkinson, A., & Beesley, A. 2022. *Zooplankton abundance time series from net hauls at site L4 off Plymouth, UK between 1988-2021.*

McEvoy, A. J., Atkinson, A., Airs, R. L., Brittain, R., Brown, I., Fileman, E. S., Findlay, H. S., McNeill, C. L., Ostle, C., Smyth, T. J., Somerfield, P. J., Tait, K., Tarran, G. A., Thomas, S., Widdicombe, C. E., Woodward, E. M. S., Beesley, A., Conway, D. V. P., Fishwick, J., Haines, H., Harris, C., Harris, R., Hélaouët, P., Johns, D., Lindeque, P. K., Mesher, T., McQuatters-Gollop, A., Nunes, J., Perry, F., Queiros, A. M., Rees, A., Rühl, S., Sims, D., Torres, R., & Widdicombe, S. 2023. The Western Channel Observatory: a century of physical, chemical and biological data compiled from pelagic and benthic habitats in the western English Channel. *Earth System Science Data*, **15**(12), 5701–5737.

McGillicuddy, D. J., & Franks, P. J. 2019. Models of Plankton Patchiness. *Pages 536–546 of: Encyclopedia of Ocean Sciences*. Elsevier.

Meddeb, M., Niquil, N., Grami, B., Mejri, K., Haraldsson, M., Chaalali, A., Pringault, O., & Hlaili, A. S. 2019. A new type of plankton food web functioning in coastal waters revealed by coupling Monte Carlo Markov chain linear inverse method and ecological network analysis. *Ecological Indicators*, **104**, 67–85.

Melle, W., Runge, J., Head, E., Plourde, S., Castellani, C., Licandro, P., Pierson, J., Jonasdottir, S., Johnson, C., Broms, C., Debes, H., Falkenhaug, T., Gaard, E., Gislason, A., Heath, M., Niehoff, B., Nielsen, T. G., Pepin, P., Stenevik, E. K., & Chust, G. 2014. The North Atlantic Ocean as habitat for *Calanus finmarchicus*: Environmental factors and life history traits. *Progress in Oceanography*, **129**, 244–284.

Menden-Deuer, S., & Lessard, E. J. 2000. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and Oceanography*, **45**(3), 569–579.

Merico, A., Bruggeman, J., & Wirtz, K. 2009. A trait-based approach for downscaling complexity in plankton ecosystem models. *Ecological Modelling*, **220**(21), 3001–3010.

Merz, E., Kozakiewicz, T., Reyes, M., Ebi, C., Isles, P., Baity-Jesi, M., Roberts, P., Jaffe, J. S., Dennis, S. R., Hardeman, T., Stevens, N., Lorimer, T., & Pomati, F. 2021. Underwater dual-magnification imaging for automated lake plankton monitoring. *Water Research*, **203**, 117524.

Met Office. 2020. *Met Office location-specific long-term averages, 1991-2020*.

Met Office. 2023. *MIDAS Open: UK hourly solar radiation data*.

Meyer, B., Fuentes, V., Guerra, C., Schmidt, K., Atkinson, A., Spahic, S., Cisewski, B., Freier, U., Olariaga, A., & Bathmanna, U. 2009. Physiology, growth, and development of larval krill *Euphausia superba* in autumn and winter in the Lazarev Sea, Antarctica. *Limnology and Oceanography*, **54**(5), 1595–1614.

Mignot, A., Ferrari, R., & Claustre, H. 2018. Floats with bio-optical sensors reveal what processes trigger the North Atlantic bloom. *Nature Communications*, **9**(1), 190.

Mignot, A., Von Schuckmann, K., Landschützer, P., Gasparin, F., Van Gennip, S., Perruche, C., Lamouroux, J., & Amm, T. 2022. Decrease in air-sea CO2 fluxes caused by persistent marine heatwaves. *Nature Communications*, **13**(1), 4300.

Mills, D. K., Tett, P. B., & Novarino, G. 1994. The spring bloom in the south western North Sea in 1989. *Netherlands Journal of Sea Research*, **33**(1), 65–80.

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., Koppelmann, 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.

Mitra, A., Flynn, K. J., Tillmann, U., Raven, J. A., Caron, D., Stoecker, D. K., Not, F., Hansen, P. J., Hallegraeff, G., Sanders, R., Wilken, S., McManus, G., Johnson, M., Pitta, P., Våge, S., Berge, T., Calbet, A., Thingstad, F., Jeong, H. J., Burkholder, J., Glibert, P. M., Granéli, E., & Lundgren, V. 2016. Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist*, **167**(2), 106–120.

Montagnes, D. J. S., & Franklin, M. 2001. Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: Reconsidering some paradigms. *Limnology and Oceanography*, **46**(8), 2008–2018.

Montagnes, D. J. S., Berges, J. A., Harrison, P. J., & Taylor, F. J. R. 1994. Estimating carbon, nitrogen, protein, and chlorophyll *a* from volume in marine phytoplankton. *Limnology and Oceanography*, **39**(5), 1044–1060.

Montresor, M., Vitale, L., D'Alelio, D., & Ferrante, M. I. 2016. Sex in marine planktonic diatoms: insights and challenges. *Perspectives in Phycology*, **3**(2), 61–75.

Moore, J., Doney, S. C., Kleypas, J. A., Glover, D. M., & Fung, I. Y. 2001. An intermediate

complexity marine ecosystem model for the global domain. *Deep Sea Research Part II: Topical Studies in Oceanography*, **49**(1-3), 403–462.

Moorthi, S., Caron, D., Gast, R., & Sanders, R. 2009. Mixotrophy: a widespread and important ecological strategy for planktonic and sea-ice nanoflagellates in the Ross Sea, Antarctica. *Aquatic Microbial Ecology*, **54**, 269–277.

Morozov, A. Y. 2010. Emergence of Holling type III zooplankton functional response: Bringing together field evidence and mathematical modelling. *Journal of Theoretical Biology*, **265**(1), 45–54.

Moschonas, G., Gowen, R. J., Paterson, R. F., Mitchell, E., Stewart, B. M., McNeill, S., Glibert, P. M., & Davidson, K. 2017. Nitrogen dynamics and phytoplankton community structure: the role of organic nutrients. *Biogeochemistry*, **134**(1-2), 125–145.

Motwani, N. H., & Gorokhova, E. 2013. Mesozooplankton Grazing on Picocyanobacteria in the Baltic Sea as Inferred from Molecular Diet Analysis. *PLoS ONE*, **8**(11), e79230.

Myung, G., Kim, H. S., Park, J. W., Park, J. S., & Yih, W. 2013. Sequestered plastids in *Mesodinium rubrum* are functionally active up to 80 days of phototrophic growth without cryptomonad prey. *Harmful Algae*, **27**, 82–87.

Møller, E. F. 2007. Production of dissolved organic carbon by sloppy feeding in the copepods *Acartia tonsa*, *Centropages typicus*, and *Temora longicornis*. *Limnology and Oceanography*, **52**(1), 79–84.

Nagai, S., Matsuyama, Y., Oh, S. J., & Itakura, S. 2004. Effect of nutrients and temperature on encystment of the toxic dinoflagellate *Alexandrium tamarense* (*Dinophyceae*) isolated from Hiroshima Bay, Japan. *Plankton Biology and Ecology*, **51**(2), 103–109.

Napoléon, C., Fiant, L., Raimbault, V., Riou, P., & Claquin, P. 2014. Dynamics of phytoplankton diversity structure and primary productivity in the English Channel. *Marine Ecology Progress Series*, **505**, 49–64.

Negrete-García, G., Luo, J. Y., Long, M. C., Lindsay, K., Levy, M., & Barton, A. D. 2022. Plankton energy flows using a global size-structured and trait-based model. *Progress in Oceanography*, **209**, 102898.

Nejstgaard, J., Gismervik, I., & Solberg, P. 1997. Feeding and reproduction by *Calanus finmarchicus*, and microzooplankton grazing during mesocosm blooms of diatoms and the coccolithophore *Emiliania huxleyi*. *Marine Ecology Progress Series*, **147**, 197–217.

Norberg, J., & DeAngelis, D. 1997. Temperature effects on stocks and stability of a phytoplankton-zooplankton model and the dependence on light and nutrients. *Ecological Modelling*, **95**(1), 75–86.

Norberg, J., Swaney, D. P., Dushoff, J., Lin, J., Casagrandi, R., & Levin, S. A. 2001. Phenotypic diversity and ecosystem functioning in changing environments: A theoretical framework. *Proceedings of the National Academy of Sciences*, **98**(20), 11376–11381.

NRFA. 2023. *National River Flow Archive*.

O'Brien, T. D., Wiebe, P. H., & Falkenhaug, T. 2010. *ICES zooplankton status report. 2010/11*. Copenhagen, Denmark: ICES. Place: Copenhagen Publisher: Internat. Council for the Exploration of the Sea Series Number: 318.

Oh, S. J., Kim, D.-I., Sajima, T., Shimasaki, Y., Matsuyama, Y., Oshima, Y., Honjo, T., & Yang, H.-S. 2008. Effects of irradiance of various wavelengths from light-emitting diodes on the growth of the harmful dinoflagellate *Heterocapsa circularisquama* and the diatom *Skeletonema costatum*. *Fisheries Science*, **74**(1), 137–145.

Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M., Jaanus, A., Kokkonen, P., Ledaine, I., & Niemkiewicz, E. 2006. *Biovolumes and size-classes of phytoplankton in the Baltic Sea*. Vol. 106. HELCOM Balt. Sea Environ. Proc.

Olin, A. B., Banas, N. S., Johns, D. G., Heath, M. R., Wright, P. J., & Nager, R. G. 2022. Spatio-temporal variation in the zooplankton prey of lesser sandeels: species and community trait patterns from the Continuous Plankton Recorder. *ICES Journal of Marine Science*, **79**(5), 1649–1661.

Olin, A. B., Banas, N. S., Heath, M. R., Wright, P. J., MacDonald, A., Wanless, S., Daunt, F., Speakman, J. R., & Nager, R. G. 2025. *Changing food conditions and size declines in a North Sea forage fish*.

Ollevier, A., Mortelmans, J., Vandegehuchte, M. B., Develter, R., De Troch, M., & Deneudt, K. 2022. A Video Plankton Recorder user guide: Lessons learned from in situ plankton imaging in shallow and turbid coastal waters in the Belgian part of the North Sea. *Journal of Sea Research*, **188**, 102257.

Olofsson, M., Robertson, E. K., Edler, L., Arneborg, L., Whitehouse, M. J., & Ploug, H. 2019. Nitrate and ammonium fluxes to diatoms and dinoflagellates at a single cell level in mixed field communities in the sea. *Scientific Reports*, **9**(1), 1424.

Olson, M. B., Lessard, E. J., Wong, C. H. J., & Bernhardt, M. J. 2006. Copepod feeding selectivity on microplankton, including the toxigenic diatoms *Pseudo-nitzschia* spp., in the coastal Pacific Northwest. *Marine Ecology Progress Series*, **326**, 207–220.

Orizar, I. D. S., & Lewandowska, A. M. 2022. Intraspecific Trait Variability of a Diatom and a Dinoflagellate Along a Salinity Gradient. *Frontiers in Marine Science*, **9**, 880309.

Oseji, O. F., Fan, C., & Chigbu, P. 2019. Composition and Dynamics of Phytoplankton in the Coastal Bays of Maryland, USA, Revealed by Microscopic Counts and Diagnostic Pigments Analyses. *Water*, **11**(2), 368.

Painter, S. C., Finlay, M., Hemsley, V. S., & Martin, A. P. 2016. Seasonality, phytoplankton succession and the biogeochemical impacts of an autumn storm in the northeast Atlantic Ocean. *Progress in Oceanography*, **142**, 72–104.

Pakhomov, E. A., Henschke, N., Hunt, B. P. V., Stowasser, G., & Cherel, Y. 2019. Utility of salps as a baseline proxy for food web studies. *Journal of Plankton Research*, **41**(1), 3–11.

Pal, R., & Choudhury, A. K. 2014. A brief introduction to phytoplanktons. *Pages 1–41 of: An introduction to phytoplanktons: diversity and ecology*. New Delhi: Springer India.

Pancic, M., & Kiørboe, T. 2018. Phytoplankton defence mechanisms: traits and trade-offs. *Biological Reviews*, **93**, 1269–1303.

Papapostolou, A., Almgren, A. V., Hansen, T. F., Kandylas, A., Serra-Pompei, C., Visser, A. W., & Andersen, K. H. 2026. Computational library for the nutrient-unicellular-multicellular plankton modeling framework v. 1.0. *Geoscientific Model Development*, **19**(14), 6627–6661.

Patel, R., Roy, S., Capuzzo, E., & Van Der Kooij, J. 2023. Seasonality of diet overlap among small pelagic fish in the waters southwest of the UK. *Marine Ecology Progress Series*, **708**, 101–123.

Payne, R. E. 1972. Albedo of the Sea Surface. *Journal of the Atmospheric Sciences*, **29**(5), 959–970.

Pengpeng, W., Fang, Z., Dongjie, G., & Song, S. 2021. Diets and Seasonal Ingestion Rates of *Aurelia coerulea* (*Cnidaria: Scyphozoa*) Polyps by in situ Feeding Experiments in Jiaozhou Bay, China. *Frontiers in Marine Science*, **8**.

Perhirin, M., Gossner, H., Godfrey, J., Johnson, R., Blanco-Bercial, L., & Ayata, S. 2024. Morphological and taxonomic diversity of mesozooplankton is an important driver of carbon export fluxes in the ocean. *Molecular Ecology Resources*, **24**(2), e13907.

Peterson, W. T., & Miller, C. B. 1977. Seasonal cycle of zooplankton abundance and species composition along the central Oregon coast. *Fishery Bulletin*, **75**(4), 717–724.

Pierce, D. 2010. *ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files*. Institution: Comprehensive R Archive Network Pages: 1.24.

Pinnegar, J. K., Trenkel, V. M., Tidd, A. N., Dawson, W. A., & Du Buit, M. H. 2003. Does diet in Celtic Sea fishes reflect prey availability? *Journal of Fish Biology*, **63**(s1), 197–212.

Piwosz, K. 2019. Weekly dynamics of abundance and size structure of specific nanophytoplankton lineages in coastal waters (Baltic Sea). *Limnology and Oceanography*, **64**(5), 2172–2186.

Platt, J. R. 1964. Strong Inference. *Science*, **146**(3642), 347–353.

Platt, T., Gallegos, C. L., & Harrison, W. G. 1980. Photoinhibition of photosynthesis in natural assemblages of marine phytoplankton. *Journal of Marine Research*, **38**, 103–111.

Plourde, S., Joly, P., Runge, J., Dodson, J., & Zakardjian, B. 2003. Life cycle of *Calanus hyperboreus* in the lower St. Lawrence Estuary and its relationship to local environmental conditions. *Marine Ecology Progress Series*, **255**, 219–233.

Pomerleau, C., Sastri, A. R., & Beisner, B. E. 2015. Evaluation of functional trait diversity for marine zooplankton communities in the Northeast subarctic Pacific Ocean. *Journal of*

Plankton Research, **37**(4), 712–726.

Raine, R. 2002. The summer distribution of net plankton dinoflagellates and their relation to water movements in the NE Atlantic Ocean, west of Ireland. *Journal of Plankton Research*, **24**(11), 1131–1147.

Ransijn, J. 2023. Marine mammal predator-prey interactions in the North Sea.

Raven, J. A., & Falkowski, P. G. 1999. Oceanic sinks for atmospheric CO_2. *Plant, Cell & Environment*, **22**(6), 741–755.

Raven, J. A. 1994. Why are there no picoplanktonic O_2 evolvers with volumes less than 10^{-19} m^3? *Journal of Plankton Research*, **16**(5), 565–580.

Raven, J. A., & Beardall, J. 2021. Influence of global environmental Change on plankton. *Journal of Plankton Research*, **43**(6), 779–800.

Record, N. R., Pershing, A. J., & Maps, F. 2014. The paradox of the "paradox of the plankton". *ICES Journal of Marine Science*, **71**(2), 236–240.

Redfield, A., Ketchum, B., & Richards, F. 1963. The influence of organisms on the composition of sea-water. *The sea: ideas and observations on progress in the study of the seas*, **2**, 26–87.

Rees, A. P., Hope, S. B., Widdicombe, C. E., Dixon, J. L., Woodward, E. M. S., & Fitzsimons, M. F. 2009. Alkaline phosphatase activity in the western English Channel: Elevations induced by high summertime rainfall. *Estuarine, Coastal and Shelf Science*, **81**(4), 569–574.

Reid, P. 2000. Impacts of fisheries on plankton community structure. *ICES Journal of Marine Science*, **57**(3), 495–502.

Reul, A., Rodríguez, V., Jiménez-Gómez, F., Blanco, J., Bautista, B., Sarhan, T., Guerrero, F., Ruíz, J., & García-Lafuente, J. 2005. Variability in the spatio-temporal distribution and size-structure of phytoplankton across an upwelling area in the NW-Alboran Sea, (W-Mediterranean). *Continental Shelf Research*, **25**(5-6), 589–608.

Richardson, K., Beardall, J., & Raven, J. A. 1983. Adaptation of unicellular algae to irradiance: an analysis of strategies. *New Phytologist*, **93**(2), 157–191.

Rigby, R. A., & Stasinopoulos, D. M. 2005. Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **54**(3), 507–554.

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**(5), eaau6253.

Riley, G. A. 1946. Factors controlling phytoplankton populations on Georges Bank. *Journal of Marine Research*, **6**(1), 54–73.

Rillo, M. C., Miller, C. G., Kucera, M., & Ezard, T. H. G. 2020. Intraspecific size variation in planktonic foraminifera cannot be consistently predicted by the environment. *Ecology and Evolution*, **10**(20), 11579–11590.

Ritter, C. J., & Bourne, D. G. 2024. Marine amphipods as integral members of global ocean ecosystems. *Journal of Experimental Marine Biology and Ecology*, **572**, 151985.

Rodríguez, F., Fernández, E., Head, R., Harbour, D., Bratbak, G., Heldal, M., & Harris, R. 2000. Temporal variability of viruses, bacteria, phytoplankton and zooplankton in the western English Channel off Plymouth. *Journal of the Marine Biological Association of the United Kingdom*, **80**(4), 575–586.

Rodríguez-Gálvez, S., Macías, D., Prieto, L., & Ruiz, J. 2023. Top-down and bottom-up control of phytoplankton in a mid-latitude continental shelf ecosystem. *Progress in Oceanography*, **217**, 103083.

Rodríguez-Ramos, T., Dornelas, M., Marañón, E., & Cermeño, P. 2014. Conventional sampling methods severely underestimate phytoplankton species richness. *Journal of Plankton Research*, **36**(2), 334–343.

Rose, J. M., & Caron, D. A. 2007. Does low temperature constrain the growth rates of heterotrophic protists? Evidence and implications for algal blooms in cold waters. *Limnology and Oceanography*, **52**(2), 886–895.

Ross, A. 1993. Ecosystem models of Scottish sea lochs for assessing the impact of nutrient enrichment. *ICES Journal of Marine Science*, **50**(4), 359–367.

Ross, O., & Sharples, J. 2007. Phytoplankton motility and the competition for nutrients in the thermocline. *Marine Ecology Progress Series*, **347**, 21–38.

Ryther, J. H., & Dunstan, W. M. 1971. Nitrogen, Phosphorus, and Eutrophication in the Coastal Marine Environment. *Science*, **171**(3975), 1008–1013.

Régnier, T., Wright, P., Harris, M., Gibb, F., Newell, M., Eerkes-Medrano, D., Daunt, F., & Wanless, S. 2024. Effect of timing and abundance of lesser sandeel on the breeding success of a North Sea seabird community. *Marine Ecology Progress Series*, **727**, 1–17.

Sala, I., & Chen, B. 2024. *PIBM 1.0: An individual-based model for simulating phytoplankton acclimation, diversity, and evolution in the ocean.*

San Martin, E., Harris, R. P., & Irigoien, X. 2006. Latitudinal variation in plankton size spectra in the Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, **53**(14-16), 1560–1572.

Sandgren, C. D., & Robinson, J. V. 1984. A stratified sampling approach to compensating for non-random sedimentation of phytoplankton cells in inverted microscope settling chambers. *British Phycological Journal*, **19**(1), 67–72.

Sandom, C., Dalby, L., Fløjgaard, C., Kissling, W. D., Lenoir, J., Sandel, B., Trøjelsgaard, K., Ejrnæs, R., & Svenning, J.-C. 2013. Mammal predator and prey species richness are strongly linked at macroscales. *Ecology*, **94**(5), 1112–1122.

Sauterey, B., & Ward, B. A. 2022. Environmental control of marine phytoplankton stoichiom-

etry in the North Atlantic Ocean. *Proceedings of the National Academy of Sciences*, **119**(1), e2114602118.

Schabhüttl, S., Hingsamer, P., Weigelhofer, G., Hein, T., Weigert, A., & Striebel, M. 2013. Temperature and species richness effects in phytoplankton communities. *Oecologia*, **171**(2), 527–536.

Schartau, M., Landry, M. R., & Armstrong, R. A. 2010. Density estimation of plankton size spectra: a reanalysis of IronEx II data. **32**(8), 18.

Scheffer, M., & Rinaldi, S. 2000. Minimal models of top-down control of phytoplankton. *Freshwater Biology*, **45**(2), 265–283.

Schmoker, C., Hernández-León, S., & Calbet, A. 2013. Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. *Journal of Plankton Research*, **35**(4), 691–706.

Schnepf, E., & Deichgraber, G. 1984. Myzocytosis, a kind of endocytosis with implications to compartmentation in endosymbiosis: Observations in *Paulsenella (Dinophyta). Naturwissenschaften*, **71**(4), 218–219.

Schnepf, E., & Elbrachter, M. 1992. Nutritional Strategies in Dinoflagellates: A Review with Emphasis on Cell Biological Aspects. *European Journal of Protistology*, **28**, 3–24.

Schnetzer, A., & Caron, D. A. 2005. Copepod grazing impact on the trophic structure of the microbial assemblage of the San Pedro Channel, California. *Journal of Plankton Research*, **27**(10), 959–971.

Schoonhoven, E. 2000. *Ecophysiology of mixotrophs.* Ph.D. thesis, Vrije Universiteit Amsterdam.

Schulz, K. G., Bach, L. T., Bellerby, R. G. J., Bermúdez, R., Büdenbender, J., Boxhammer, T., Czerny, J., Engel, A., Ludwig, A., Meyerhöfer, M., Larsen, A., Paul, A. J., Sswat, M., & Riebesell, U. 2017. Phytoplankton Blooms at Increasing Levels of Atmospheric Carbon Dioxide: Experimental Evidence for Negative Effects on Prymnesiophytes and Positive on Small Picoeukaryotes. *Frontiers in Marine Science*, **4**.

Scottish Government, Marine Directorate. 2024. *Zooplankton Traits Database.*

Searle, K. R., Regan, C. E., Perrow, M. R., Butler, A., Rindorf, A., Harris, M. P., Newell, M. A., Wanless, S., & Daunt, F. 2023. Effects of a fishery closure and prey abundance on seabird diet and breeding success: Implications for strategic fisheries management and seabird conservation. *Biological Conservation*, **281**, 109990.

Serra-Pompei, C., Soudijn, F., Visser, A. W., Kiørboe, T., & Andersen, K. H. 2020. A general size- and trait-based model of plankton communities. *Progress in Oceanography*, **189**, 102473.

Sharples, J., Holt, J., & Dye, S. R. 2013. Impacts of climate change on shelf sea stratification. *MCCIP Science Review 2013*, 1–4.

Sherr, E., & Sherr, B. 2007. Heterotrophic dinoflagellates: a significant component of microzoo-

plankton biomass and major grazers of diatoms in the sea. *Marine Ecology Progress Series*, **352**, 187–197.

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**(6), 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**(3), 181–196.

Sieracki, C. K., Sieracki, M. E., & Yentsch, C. S. 1998. An imaging-in-flow system for automated analysis of marine microplankton. *Mar Ecol Prog Ser*.

Sigman, D. M., & Hain, M. P. 2012. The Biological Productivity of the Ocean. *Nature Education Knowledge*, **3**(10), 21.

Sikes, C. S., & Wilbur, K. M. 1982. Functions of coccolith formation. *Limnology and Oceanography*, **27**(1), 18–26.

Simpson, J. H., & Hill, A. E. 1986. The Scottish Coastal Current. *Pages 295–308 of: The Role of Freshwater Outflow in Coastal Marine Ecosystems*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Simpson, J. H., & Sharples, J. 2012. *Introduction to the Physical and Biological Oceanography of Shelf Seas*. Cambridge University Press.

Six, K. D., & Maier-Reimer, E. 1996. Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, **10**(4), 559–583.

Sloughter, T., Banas, N., & Sambrotto, R. 2019. Seasonal variation in light response of polar phytoplankton. *Journal of Marine Systems*, **191**, 64–75.

Smalley, G., Coats, D., & Stoecker, D. 2003. Feeding in the mixotrophic dinoflagellate *Ceratium furca* is influenced by intracellular nutrient concentrations. *Marine Ecology Progress Series*, **262**, 137–151.

Smith, S. L., Merico, A., Wirtz, K. W., & Pahlow, M. 2014. Leaving misleading legacies behind in plankton ecosystem modelling. *Journal of Plankton Research*, **36**(3), 613–620.

Smyth, T. J., Fishwick, J. R., AL-Moosawi, L., Cummings, D. G., Harris, C., Kitidis, V., Rees, A., Martinez-Vicente, V., & Woodward, E. M. S. 2010. A broad spatio-temporal view of the Western English Channel observatory. *Journal of Plankton Research*, **32**(5), 585–601.

Soetaert, K., Petzoldt, T., & Setzer, R. W. 2010. Solving Differential Equations in R: Package deSolve. *Journal of Statistical Software*, **33**(9).

Sommer, U., & Lengfellner, K. 2008. Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. *Global Change Biology*, **14**(6), 1199–1208.

Speijer, D., Lukeš, J., & Eliáš, M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of

eukaryotic life. *Proceedings of the National Academy of Sciences*, **112**(29), 8827–8834.

Steele, J. H. 1956. Plant production on the Fladen ground. *Journal of the Marine Biological Association of the United Kingdom*, **35**(1), 1–33.

Steele, J. H., & Henderson, E. W. 1992. The role of predation in plankton models. *Journal of Plankton Research*, **14**(1), 157–172.

Stelfox-Widdicombe, C., Archer, S., Burkill, P., & Stefels, J. 2004. Microzooplankton grazing in *Phaeocystis* and diatom-dominated waters in the southern North Sea in spring. *Journal of Sea Research*, **51**(1), 37–51.

Stock, C., & Dunne, J. 2010. Controls on the ratio of mesozooplankton production to primary production in marine ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, **57**(1), 95–112.

Stoecker, D. K., Hansen, P. J., Caron, D. A., & Mitra, A. 2017. Mixotrophy in the Marine Plankton. *Annual Review of Marine Science*, **9**(1), 311–335.

Stone, C. J. 1989. A comparison of algal diets for cirripede nauplii. *Journal of Experimental Marine Biology and Ecology*, **132**(1), 17–40.

Straile, D. 1997. Gross growth efficiencies of protozoan and metazoan zooplankton and their dependence on food concentration, predator-prey weight ratio, and taxonomic group. *Limnology and Oceanography*, **42**(6), 1375–1385.

Stramska, M. 2009. Particulate organic carbon in the global ocean derived from SeaWiFS ocean color. *Deep Sea Research Part I: Oceanographic Research Papers*, **56**(9), 1459–1470.

Strickland, J. D. H. 1960. Measuring the production of marine phytoplankton. *Fisheries Research Board of Canda Bulletin*, **122**, 1–172.

Strom, S., Miller, C., & Frost, B. 2000. What sets lower limits to phytoplankton stocks in high-nitrate, low-chlorophyll regions of the open ocean? *Marine Ecology Progress Series*, **193**, 19–31.

Suttle, C. A. 2005. Viruses in the sea. *Nature*, **437**(7057), 356–361.

Suttle, C. A., Chan, A. M., & Cottrell, M. T. 1990. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature*, **347**(6292), 467–469.

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. *ICES Journal of Marine Science*, **18**(3), 287–295.

Taguchi, S. 1976. Relationship between cell size and photosynthesis of marine diatoms. *Journal of Phycology*, **12**(2), 185–189.

Tarran, G. A., & May, R. 2019. *Abundance of phytoplankton, heterotrophic nanoflagellates and bacteria through the water column at time series station L4 in the Western English Channel, 2007-2017.*

Terseleer, N., Bruggeman, J., Lancelot, C., & Gypens, N. 2014. Trait-based representation of diatom functional diversity in a plankton functional type model of the eutrophied southern North Sea. *Limnology and Oceanography*, **59**(6), 1958–1972.

Tett, P., Edwards, A., & Jones, K. 1986. A model for the growth of shelf-sea phytoplankton in summer. *Estuarine, Coastal and Shelf Science*, **23**(5), 641–672.

Thacker, W. C. 1989. The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans*, **94**(C5), 6177–6196.

Thomas, M. K., Kremer, C. T., Klausmeier, C. A., & Litchman, E. 2012. A Global Pattern of Thermal Adaptation in Marine Phytoplankton. *Science*, **338**(6110), 1085–1088.

Thompson, P. A., Harrison, P. J., & Parslow, J. S. 1991. Influence of irradiance on cell volume and carbon quota for ten species of marine phytoplankton. *Journal of Phycology*, **27**(3), 351–360.

Tian, T., Merico, A., Su, J., Staneva, J., Wiltshire, K., & Wirtz, K. 2009. Importance of resuspended sediment dynamics for the phytoplankton spring bloom in a coastal marine ecosystem. *Journal of Sea Research*, **62**(4), 214–228.

Tinker, J., Howes, E., Wakelin, S., Menary, M., Kent, E., Berry, D., Hindson, J., Ribeiro, J., Dye, S., Andres, O., Lyons, K., & Smyth, T. 2020. The impacts of climate change on temperature (air and sea), relevant to the coastal and marine environment around the UK. *MCCIP Science Review 2020*, 1–30.

Toullec, J., Vincent, D., Frohn, L., Miner, P., Le Goff, M., Devesa, J., & Moriceau, B. 2019. Copepod Grazing Influences Diatom Aggregation and Particle Dynamics. *Frontiers in Marine Science*, **6**, 751.

Townsend, D. W., Keller, M. D., Sieracki, M. E., & Ackleson, S. G. 1992. Spring phytoplankton blooms in the absence of vertical water column stratification. *Nature*, **360**(6399), 59–62.

Turner, J. 2002. Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquatic Microbial Ecology*, **27**, 57–102.

UK Marine Management Organisation. 2024. *UK Sea Fisheries Statistics 2023*. Tech. rept.

Uncles, R., & Torres, R. 2013. Estimating dispersion and flushing time-scales in a coastal zone: Application to the Plymouth area. *Ocean & Coastal Management*, **72**, 3–12.

Utermöhl, v. H. 1931. Neue Wege in der quantitativen Erfassung des Planktons (Mit besondere Berucksichtigung des Ultraplanktons). *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, **5**, 567–595.

Utne, K. R., Skagseth, O., Wennevik, V., Broms, C. T., Melle, W., & Thorstad, E. B. 2022. Impacts of a Changing Ecosystem on the Feeding and Feeding Conditions for Atlantic Salmon During the First Months at Sea. *Frontiers in Marine Science*, **9**, 824614.

Uye, S. I., & Liang, D. 1998. Copepods attain high abundance, biomass and production in the absence of large predators but suffer cannibalistic loss. *Journal of Marine Systems*, **15**(1-4),

495–501.

Vallina, S., Ward, B., Dutkiewicz, S., & Follows, M. 2014. Maximal feeding with active prey-switching: A kill-the-winner functional response and its effect on global diversity and biogeography. *Progress in Oceanography*, **120**, 93–109.

Vargas, C., & González, H. 2004. Plankton community structure and carbon cycling in a coastal upwelling system. I. Bacteria, microprotozoans and phytoplankton in the diet of copepods and appendicularians. *Aquatic Microbial Ecology*, **34**, 151–164.

Veldhuis, M. J., & Kraay, G. W. 2000. Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. *Scientia Marina*, **64**(2), 121–134.

Verdolin, J. L. 2006. Meta-analysis of foraging and predation risk trade-offs in terrestrial systems. *Behavioral Ecology and Sociobiology*, **60**(4), 457–464.

Ward, B. A., Dutkiewicz, S., Jahn, O., & Follows, M. J. 2012. A size-structured food-web model for the global ocean. *Limnology and Oceanography*, **57**(6), 1877–1891.

Ward, B. A., & Follows, M. J. 2016. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proceedings of the National Academy of Sciences*, **113**(11), 2958–2963.

Ward, B. A., Dutkiewicz, S., & Follows, M. J. 2014. Modelling spatial and temporal patterns in size-structured marine plankton communities: top–down and bottom–up controls. *Journal of Plankton Research*, **36**(1), 31–47.

Wells, S. R., Bresnan, E., Cook, K., Eerkes-Medrano, D., Machairopoulou, M., Mayor, D. J., Rabe, B., & Wright, P. J. 2022. Environmental drivers of a decline in a coastal zooplankton community. *ICES Journal of Marine Science*, **79**(3), 844–854.

West, G. B., Brown, J. H., & Enquist, B. J. 1997. A General Model for the Origin of Allometric Scaling Laws in Biology. *Science*, **276**(5309), 122–126.

Western Channel Observatory. 2020. *Global and Diffuse Irradiance from the L4 buoy SPN1 sensor, 2013-2019*. Personal Communication with Tim Smyth.

Weynberg, K., Allen, M., & Wilson, W. 2017. Marine Prasinoviruses and Their Tiny Plankton Hosts: A Review. *Viruses*, **9**(3), 43.

White, E. P., Ernest, S. M., Kerkhoff, A. J., & Enquist, B. J. 2007. Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, **22**(6), 323–330.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48**(4), 817.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. 2019. Welcome to the Tidyverse. *Journal of Open Source Software*, **4**(43),

1686.

Widdicombe, C. E., Eloire, D., Harbour, D., Harris, R. P., & Somerfield, P. J. 2010. Long-term phytoplankton community dynamics in the Western English Channel. *Journal of Plankton Research*, **32**(5), 643–655.

Widdicombe, C. E., & Harbour, D. 2021. *Phytoplankton taxonomic abundance and biomass time-series at Plymouth Station L4 in the Western English Channel, 1992-2020.*

Wiebinga, C. J., Veldhuis, M. J., & De Baar, H. J. 1997. Abundance and productivity of bacterioplankton in relation to seasonal upwelling in the northwest Indian Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, **44**(3), 451–476.

Wihsgott, J. U., Sharples, J., Hopkins, J. E., Woodward, E. M. S., Hull, T., Greenwood, N., & Sivyer, D. B. 2019. Observations of vertical mixing in autumn and its effect on the autumn phytoplankton bloom. *Progress in Oceanography*, **177**, 102059.

Wilhelm, S. W., & Suttle, C. A. 1999. Viruses and Nutrient Cycles in the Sea. *BioScience*, **49**(10), 781–788.

Wilkinson, B., Jones, O., Rees, A. P., & Fishwick, J. 2023. *Fluorometric Chlorophyll-a measurements from CTD niskin collected from depth profiles at station L4 in the Western Channel Observatory (WCO) from 1992 up until end of 2022.*

Wirtz, K. W. 2014. A biomechanical and optimality-based derivation of prey-size dependencies in planktonic prey selection and ingestion rates. *Marine Ecology Progress Series*, **507**, 81–94.

Wirtz, K. W., & Eckhardt, B. 1996. Effective variables in ecosystem models with an application to phytoplankton succession. *Ecological Modelling*, **92**(1), 33–53.

Wirtz, K. W. 2011. Non-uniform scaling in phytoplankton growth rate due to intracellular light and CO_2 decline. *Journal of Plankton Research*, **33**(9), 1325–1341.

Wolf, J., Yates, N., Brereton, A., Buckland, H., De Dominics, M., Gallego, A., & O'Hara Murray, R. 2016. The Scottish Shelf Model. Part 1: Shelf-Wide Domain. *Scottish Marine and Freshwater Science*, **7**(3), 151.

Wommack, K. E., & Colwell, R. R. 2000. Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews*, **64**(1), 69–114.

Wood, M., Beresford, N., & Copplestone, D. 2011. Limit of detection values in data analysis: Do they matter? *Radioprotection*, **46**(6), S85–S90.

Wood, S. N. 2011. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **73**(1), 3–36.

Woodward, E. M. S., & Harris, C. 2023. *Micromolar Nutrient concentration profiles from the long-term time series at Station L4 in the Western English Channel from 2000 to 2022.*

Xu, H., Chen, F., Zhang, X., Zhang, Z., Pan, K., & Liu, H. 2024. Grazer-induced changes on

mechanical properties of diatoms frustule: A new proof for a watery arms race. *Limnology and Oceanography Letters*, **9**, 754–763.

Ye, L., Chang, C., García-Comas, C., Gong, G., & Hsieh, C. 2013. Increasing zooplankton size diversity enhances the strength of top-down control on phytoplankton through diet niche partitioning. *Journal of Animal Ecology*, **82**(5), 1052–1061.

Yool, A., Popova, E. E., & Anderson, T. R. 2013. MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies. *Geoscientific Model Development*, **6**(5), 1767–1811.

Yoshida, T., Hairston, N. G., & Ellner, S. P. 2004. Evolutionary trade–off between defence against grazing and competitive ability in a simple unicellular alga, *Chlorella vulgaris*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **271**(1551), 1947–1953.

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.

Zhan, W., Feng, M., Zhang, Y., Shen, X., Zhan, H., & He, Q. 2024. Reduced and smaller phytoplankton during marine heatwaves in eastern boundary upwelling systems. *Communications Earth & Environment*, **5**(1), 629.

Zhan, W., Zhang, Y., He, Q., Feng, M., Du, Y., Shen, X., Liu, Y., & Zhan, H. 2026. Global Responses of Phytoplankton Size Structure to Marine Heatwaves. *Global Biogeochemical Cycles*, **40**(1), e2025GB008854.

Zhao, P., Ouyang, L., Shen, A., & Wang, Y. 2022. The cell cycle of phytoplankton: A review. *Journal of the World Aquaculture Society*, **53**(4), 799–815.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., & Scanlan, D. J. 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology*, **10**(1), 147–161.

Álvarez, E., Moyano, M., López-Urrutia, A., Nogueira, E., & Scharek, R. 2014. Routine determination of plankton community composition and size structure: a comparison between FlowCAM and light microscopy. *Journal of Plankton Research*, **36**(1), 170–184.

Appendices

Appendix A

Plankton organism size data

Table A.1: Sizes of phytoplankton and protozooplankton for PL, extracted directly from Widdicombe & Harbour (2021) and converted to ESD. Raw data included here for completeness; all dinoflagellates were considered either mixoplankton (both autotrophs and mixotrophs), or zooplankton in our analyses. The data are included under the Western Channel Observatory data policy, which should be consulted by all users.

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
DIATOMS			
Achnanthes longipes	156533	Autotrophic	22.4
Actinocyclus	148944	Autotrophic	30.6
Actinoptychus senarius	148948	Autotrophic	29.3
Amphiprora surirelloides	160671	Autotrophic	14.2
Asterionellopsis glacialis	149139	Autotrophic	10.1
Asteromphalus flabellatus	162255	Autotrophic	29.3
Asteromphalus sarcophagus	394598	Autotrophic	16.4
Attheya septentrionalis	162823	Autotrophic	3.4
Bacillaria paxillifera	558243	Autotrophic	20.3
Bacteriastrum furcatum	164110	Autotrophic	20.8
Brockmanniella brockmannii	149137	Autotrophic	15.9
Cerataulina pelagica	149619	Autotrophic	27.0
Chaetoceros	148985	Autotrophic	4.3
Chaetoceros affinis	149241	Autotrophic	7.8
Chaetoceros anastomosans	149292	Autotrophic	7.8
Chaetoceros brevis	149291	Autotrophic	7.8
Chaetoceros cf convolutus	156611	Autotrophic	18.1
Chaetoceros compressus	149129	Autotrophic	5.7
Chaetoceros concavicornis	156607	Autotrophic	18.1
Chaetoceros costatus	149289	Autotrophic	7.8
Chaetoceros curvisetus	149221	Autotrophic	11.0
Chaetoceros danicus	149120	Autotrophic	18.1
Chaetoceros debilis	149219	Autotrophic	12.6
Chaetoceros decipiens	149126	Autotrophic	15.1
Chaetoceros densus	149121	Autotrophic	20.8
Chaetoceros didymus	149122	Autotrophic	11.0
Chaetoceros eibenii	160521	Autotrophic	20.8
Chaetoceros externus	160522	Autotrophic	7.8
Chaetoceros f. borealis	149124	Autotrophic	7.8
Chaetoceros filiformis	178194	Autotrophic	4.9
Chaetoceros fragilis	149173	Autotrophic	4.9

Continued on next page

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
Chaetoceros laciniosus	149228	Autotrophic	7.8
Chaetoceros lauderi	160523	Autotrophic	18.1
Chaetoceros peruvianus	178185	Autotrophic	12.4
Chaetoceros protuberans	163055	Autotrophic	11.0
Chaetoceros radicans	163112	Autotrophic	5.8
Chaetoceros resting spores	148985	Autotrophic	5.6
Chaetoceros similis	149127	Autotrophic	3.4
Chaetoceros simplex	149294	Autotrophic	5.1
Chaetoceros socialis	149123	Autotrophic	5.8
Chaetoceros teres	149125	Autotrophic	27.0
Chaetoceros tortissimus	163161	Autotrophic	5.8
Chaetoceros wighamii	160524	Autotrophic	5.8
Chaetoceros willei	156625	Autotrophic	7.8
Corethron pennatum	341496	Autotrophic	33.9
Coscinodiscus asteromphalus	149274	Autotrophic	60.8
Coscinodiscus centralis	149159	Autotrophic	87.7
Coscinodiscus concinnus	148992	Autotrophic	168.7
Coscinodiscus granii	149271	Autotrophic	87.7
Coscinodiscus radiatus	149158	Autotrophic	43.3
Coscinodiscus wailesii	156632	Autotrophic	275.0
Cylindrotheca closterium	149004	Autotrophic	6.6
Dactyliosolen blavyanus	179786	Autotrophic	62.1
Dactyliosolen fragilissimus	149310	Autotrophic	22.8
Delphineis	149179	Autotrophic	9.8
Detonula pumila	149647	Autotrophic	26.6
Diploneis crabro	149396	Autotrophic	16.5
Ditylum brightwellii	149023	Autotrophic	37.3
Ephemera planamembranacea	341542	Autotrophic	39.6
Eucampia zodiacus	149131	Autotrophic	20.7
Fragilaria	149028	Autotrophic	4.4
Fragilariopsis	149313	Autotrophic	6.0
Grammatophora	149335	Autotrophic	24.5
Guinardia delicatula	149112	Autotrophic	22.8
Guinardia flaccida	149132	Autotrophic	82.5
Guinardia striata	149113	Autotrophic	15.6
Guinardia striata (large)	149113	Autotrophic	42.2
Haslea wawrikae	248063	Autotrophic	9.5
Helicotheca tamesis	157440	Autotrophic	22.6
Lauderia annulata	149135	Autotrophic	30.0
Leptocylindrus danicus	149106	Autotrophic	7.3
Leptocylindrus mediterraneus	149230	Autotrophic	5.6
Leptocylindrus minimus	149039	Autotrophic	3.3
Licmophora	149342	Autotrophic	27.0
Lioloma delicatulum	292728	Autotrophic	9.1
Lithodesmium undulatum	149322	Autotrophic	20.2
Melosira	149042	Autotrophic	15.9
Meuniera membranacea	149145	Autotrophic	29.8
Nanoneis hasleae	248180	Autotrophic	2.4
Navicula	149142	Autotrophic	5.6
Navicula distans	149143	Autotrophic	16.0
Navicula transitans var. *derasa*	175320	Autotrophic	18.5
Neocalyptrella robusta	345491	Autotrophic	164.1
Nitzschia sigmoidea	149604	Autotrophic	7.4
Odontella mobiliensis	164116	Autotrophic	59.7
Odontella sinensis	149095	Autotrophic	85.6
Paralia sulcata	149055	Autotrophic	18.2
Pennate 30 μm	149001	Autotrophic	11.2

Continued on next page

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
Pennate 50 μm	149001	Autotrophic	11.2
Plagiogrammopsis vanheurckii	149057	Autotrophic	17.0
Plagiolemma distortum	601957	Autotrophic	21.2
Planktoniella sol	196815	Autotrophic	13.3
Pleurosigma	149181	Autotrophic	27.0
Pleurosigma planctonicum	231883	Autotrophic	54.1
Podosira stelligera	149060	Autotrophic	50.0
Proboscia alata	149168	Autotrophic	40.2
*Proboscia alata (*syn. *f. gracillima)*	149168	Autotrophic	14.4
Proboscia alata 5 μm	149168	Autotrophic	22.4
Proboscia truncata	248181	Autotrophic	74.0
Psammodictyon panduriforme	149217	Autotrophic	16.5
Pseudo-nitzschia "delicatissima"	149153	Autotrophic	4.6
Pseudo-nitzschia "pungens"	160528	Autotrophic	9.8
Pseudo-nitzschia "seriata"	149152	Autotrophic	9.8
Rhizosolenia chunii	341502	Autotrophic	24.5
Rhizosolenia hebetata f. semispina	149071	Autotrophic	35.6
Rhizosolenia imbricata 10 μm	149116	Autotrophic	35.6
Rhizosolenia imbricata 15 μm	149116	Autotrophic	44.5
Rhizosolenia imbricata 5 μm	149116	Autotrophic	22.4
Rhizosolenia setigera 25 μm	149115	Autotrophic	57.2
Rhizosolenia setigera 5 μm	149115	Autotrophic	16.9
Rhizosolenia styliformis	149629	Autotrophic	74.0
Roperia tesselata	149105	Autotrophic	28.8
Skeletonema costatum	149074	Autotrophic	4.3
Small Pennate	149001	Autotrophic	8.4
Stephanopyxis palmeriana	231888	Autotrophic	58.5
Thalassionema nitzschioides	149093	Autotrophic	7.0
Thalassiosira 2 μm	148912	Autotrophic	1.8
Thalassiosira 4 μm	148912	Autotrophic	2.9
Thalassiosira 5 μm	148912	Autotrophic	4.8
Thalassiosira 10 μm	148912	Autotrophic	9.1
Thalassiosira 20 μm	148912	Autotrophic	18.2
Thalassiosira 30 μm	148912	Autotrophic	27.3
Thalassiosira 40 μm	148912	Autotrophic	28.8
Thalassiosira 60 μm	148912	Autotrophic	43.3
Thalassiosira anguste-lineata	148914	Autotrophic	27.0
Thalassiosira cf angulata	148912	Autotrophic	31.3
Thalassiosira constricta	345531	Autotrophic	18.2
Thalassiosira eccentrica	148922	Autotrophic	43.3
Thalassiosira gravida	149102	Autotrophic	27.3
Thalassiosira gravida 15 μm	149102	Autotrophic	15.0
Thalassiosira minuscula	555293	Autotrophic	12.1
Thalassiosira punctigera	148936	Autotrophic	43.3
Thalassiosira rotula	148942	Autotrophic	27.0
Thalassiosira subtilis	149101	Autotrophic	17.2
Thalassiothrix	157081	Autotrophic	14.5
Trigonium alternans	699394	Autotrophic	21.5
V. small Pennate	149001	Autotrophic	4.6
DINOFLAGELLATES			
Alexandrium tamarense	109714	Autotrophic	25.0
Amphidinium crassum	109726	Heterotrophic	11.4
Amphidinium f. sphenoides	109754	Heterotrophic	11.4
Amphidoma caudata	110005	Autotrophic	15.4
Amylax triacantha	110007	Autotrophic	30.0
Tripos hexacanthum	841248	Autotrophic	45.7
Tripos longipes	841259	Autotrophic	45.7

Continued on next page

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
Cochlodinium	109474	Heterotrophic	11.4
Colourless Gymnodinium	109475	Heterotrophic	14.4
Colourless Gymnodinium (small)	109475	Heterotrophic	9.9
Dinophysis acuminata	109603	Mixotrophic	33.5
Dinophysis acuta	109604	Mixotrophic	49.3
Dinophysis cf punctata	109462	Mixotrophic	20.8
Dinophysis nasuta	232496	Mixotrophic	20.8
Dinophysis sacculus	232261	Mixotrophic	32.5
Dinophysis tripos	109662	Mixotrophic	49.3
Diplopsalis	109515	Heterotrophic	38.3
Dissodinium pseudolunula	110325	Parasitic	39.6
Erythropsidinium	231793	Autotrophic	29.2
Gonyaulax	109519	Autotrophic	15.0
Gonyaulax digitale	110015	Autotrophic	35.0
Gonyaulax grindleyi	110023	Autotrophic	35.0
Gonyaulax spinifera	110041	Autotrophic	32.5
Gonyaulax verior	110045	Autotrophic	28.0
Gymnodinium	109475	Autotrophic	9.9
Gymnodinium pygmaeum	109825	Autotrophic	18.3
Gyrodinium (large)	109476	Heterotrophic	32.7
Gyrodinium (medium)	109476	Mixotrophic	28.2
Gyrodinium (small)	109476	Mixotrophic	16.5
Gyrodinium falcatum	232931	Mixotrophic	39.2
Gyrodinium f. spirale	109876	Heterotrophic	66.0
Heterocapsa	109540	Autotrophic	5.9
Heterocapsa niei	233620	Autotrophic	5.9
Heterocapsa triquetra	110153	Autotrophic	10.0
Karenia mikimotoi	233024	Autotrophic	20.0
Karlodinium veneficum	233037	Mixotrophic	14.8
Katodinium	109477	Heterotrophic	7.6
Katodinium glaucum	109885	Heterotrophic	14.3
Kofoidinium lebourae	233165	Heterotrophic	69.3
Lessardia elongata	232703	Heterotrophic	13.1
Mesoporos perforatus	232516	Autotrophic	22.0
Micranthodinium	109510	Autotrophic	12.6
Nematodinium	109490	Mixotrophic	32.5
Tripos furca	840627	Autotrophic	40.2
Tripos fusus	840626	Autotrophic	26.7
Tripos horridus	837453	Autotrophic	45.7
Tripos lineatus	837459	Autotrophic	28.0
Tripos macroceros	841260	Autotrophic	45.7
Tripos massiliensis	841261	Autotrophic	47.7
Tripos muelleri	495363	Autotrophic	60.9
Noctiluca scintillans	109921	Heterotrophic	235.0
Oxytoxum	109528	Mixotrophic	5.0
Phalacroma rotundatum	156505	Heterotrophic	36.3
Polykrikos schwarzii	109901	Heterotrophic	72.3
Preperidinium	109505	Heterotrophic	71.4
Pronoctiluca pelagica	109903	Heterotrophic	18.9
Prorocentrum balticum	110293	Autotrophic	8.9
Prorocentrum cordatum	232376	Autotrophic	6.4
Prorocentrum dentatum	110298	Autotrophic	6.2
Prorocentrum micans	110303	Autotrophic	28.2
Prorocentrum triestinum	110316	Autotrophic	22.6
Protoperidinium	109553	Heterotrophic	43.3
Protoperidinium bipes	110208	Heterotrophic	26.6
Protoperidinium brevipes	110210	Heterotrophic	18.5

Continued on next page

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
Protoperidinium curtipes	110215	Heterotrophic	54.3
Protoperidinium depressum	110217	Heterotrophic	100.8
Protoperidinium divergens	110219	Heterotrophic	39.2
Protoperidinium obtusum	110239	Heterotrophic	95.3
Protoperidinium oceanicum	110240	Heterotrophic	100.8
Protoperidinium ovatum	110241	Heterotrophic	95.3
Protoperidinium pentagonum	110247	Heterotrophic	64.5
Protoperidinium pyriforme	110249	Heterotrophic	34.3
Protoperidinium steinii	110257	Heterotrophic	34.3
Pyrophacus horologicum	110266	Autotrophic	38.3
Sclerodinium calyptroglyphe	615516	Mixotrophic	20.0
Scrippsiella (cyst)	109545	Autotrophic	17.8
Scrippsiella trochoidea	110172	Autotrophic	17.8
Torodinium robustum	109889	Mixotrophic	18.3
Torodinium teredo	109890	Heterotrophic	33.5
Tryblionella compressa	447746	Autotrophic	22.9
Unidentified *Peridiniales* (large)	109394	Heterotrophic	18.7
Unidentified *Peridiniales* (small)	109394	Heterotrophic	8.6
Warnowia	109491	Heterotrophic	16.5
Zoospore	NA	Heterotrophic	5.0
COCCOLITHOPHORID			
Acanthoica quattrospina	235802	Autotrophic	8.0
Braarudosphaera bigelowii	235922	Autotrophic	16.0
Calciosolenia brasiliensis	555889	Autotrophic	6.7
Calyptrosphaera	235828	Autotrophic	15.0
Coccolithus pelagicus	178600	Autotrophic	26.5
Coccolithus pelagicus f. hyalinus	555900	Autotrophic	14.0
Coronosphaera	235934	Autotrophic	8.0
Emiliania huxleyi	115104	Autotrophic	5.0
Gephyrocapsa	235823	Autotrophic	6.0
Helicosphaera carteri	615435	Autotrophic	19.2
Holococcolithophorid 14 μm	248178	Autotrophic	10.0
Holococcolithophorid 8 μm	248178	Autotrophic	6.0
Rhabdolithes claviger	626382	Autotrophic	5.0
Syracosphaera molischii	236039	Autotrophic	6.0
Syracosphaera pulchra	235979	Autotrophic	12.6
Umbellosphaera	235937	Autotrophic	10.0
Unidentified *Coccolithophorid*	115057	Autotrophic	10.0
FLAGELLATES			
Bodonidae	22556	Heterotrophic	15.9
Choanoflagellatea	580116	Heterotrophic	4.6
Colourless Flagellate	146222	Heterotrophic	5.0
Corymbellus aureus	162519	Autotrophic	6.6
Cryptomonadaceae	17644	Autotrophic	5.0
Dictyocha fibula	157463	Autotrophic	12.2
Dictyocha speculum	157260	Autotrophic	12.2
Dinobryon	157240	Mixotrophic	4.6
Eutreptiella	17657	Autotrophic	14.2
Flagellate 2 μm	146222	Autotrophic	2.0
Flagellates 15 μm	146222	Autotrophic	9.9
Flagellates 5 μm	146222	Autotrophic	5.0
Halosphaera	134528	Autotrophic	98.3
Meringosphaera	115075	Autotrophic	4.6
Pseudochattonella	531445	NA	9.0
Pterosperma	160595	Autotrophic	20.0
Pyramimonas	134529	Autotrophic	8.0
Raphidophyceae	160581	Autotrophic	6.1

Continued on next page

Taxon	**WoRMS ID**	**Trophic mode**	**ESD (μm)**
PHAEOCYSTIS	115106	Autotrophic	6.0
CILIATES			
Ascampbelliella	415082	Heterotrophic	33.2
Askenasia stellaris	292898	Heterotrophic	55.2
Balanion	292899	Heterotrophic	10.8
Unidentified ciliate	11	Mixotrophic	18.4
Didinium	341301	Heterotrophic	46.4
Epiplocylis undella	341668	Heterotrophic	23.8
Eutintinnus	183543	Heterotrophic	27.1
Favella	172431	Heterotrophic	27.1
Helicostomella	172434	Heterotrophic	27.1
Laboea strobila	101264	Mixotrophic	34.4
Leegaardiella	101179	Heterotrophic	26.9
Leegaardiella sol	101207	Heterotrophic	29.3
Lohmanniella oviformis	101209	Heterotrophic	22.9
Mesodinium (small)	179320	Mixotrophic	15.0
Mesodinium rubrum	232069	Mixotrophic	30.0
Parafavella	196836	Heterotrophic	27.1
Peritromus	172321	Heterotrophic	17.8
Proplectella	292924	Heterotrophic	36.3
Prorodontida	425488	Heterotrophic	17.5
Rhabdoaskenasia	292925	Heterotrophic	10.5
Salpingella	183566	Heterotrophic	19.9
Strobilidium	101185	Heterotrophic	25.0
Strombidinopsis	101198	Heterotrophic	34.4
Strombidium (large)	101195	Mixotrophic	28.8
Strombidium (medium)	101195	Mixotrophic	17.1
Strombidium (small)	101195	Mixotrophic	9.9
Tiarina	247913	Heterotrophic	31.9
Tintinnid (small)	732976	Heterotrophic	14.4
Tintinnopsis	163780	Heterotrophic	18.9
Tontonia	101196	Heterotrophic	26.0
Tontonia ovalis	427744	Mixotrophic	34.8
Unidentified tintinnid/ciliate	11	Heterotrophic	42.0
Uronema	143943	Heterotrophic	11.4
Vorticella or Carchesium	163573	Heterotrophic	21.5
OTHER			
Amoeba	605376	Heterotrophic	8.0
Radiozoa	582421	Heterotrophic	26.2

Table A.2: Sizes of phytoplankton and protozooplankton for SH & LE, primarily compiled by González-Gil *et al.* (2022) and reduced for presentation in this thesis; the full table can be found in the associated supplementary material. Sources Cloern (2018) and Widdicombe & Harbour (2021) were added to the original table of González-Gil *et al.* (2022) as necessary.

Taxon	**ESD (μm)**	Source
CILIATES		
Mesodinium rubrum	28.5	Olenina *et al.* (2006)
DIATOMS		
Achnanthes longipes	10.7	Olenina *et al.* (2006)
Actinoptychus senarius	35.8	Olenina *et al.* (2006)
Actinoptychus spp.	25.1	Cloern (2018)
Asterionella formosa	11.8	Olenina *et al.* (2006)
Asterionella spp.	14.3	Leblanc *et al.* (2012)
Asterionellopsis glacialis	11.5	Harrison *et al.* (2015)
Asteromphalus spp.	28.7	Cloern (2018)
Asteroplanus karianus	5.8	Cloern (2018)
Attheya longicornis	5.8	Olenina *et al.* (2006)
Attheya septentrionalis	29.3	Widdicombe & Harbour (2021)
Bacillaria paxillifera	20.3	Widdicombe & Harbour (2021)
Bacillaria spp.	16.4	Cloern (2018)
Bacteriastrum spp.	28.6	Harrison *et al.* (2015)
Bellerochea spp.	42.0	Harrison *et al.* (2015)
Biddulphia spp.	58.5	Harrison *et al.* (2015)
Corataulina pelagica	31.3	Harrison *et al.* (2015)
Corataulina spp.	31.3	Harrison *et al.* (2015)
Chaetoceros danicus	13.7	Harrison *et al.* (2015)
Chaetoceros peruvianus	12.4	Widdicombe & Harbour (2021)
Chaetoceros socialis	8.8	Harrison *et al.* (2015)
Chaetoceros spp.	17.4	Harrison *et al.* (2015)
Corethron hystrix	57.8	Harrison *et al.* (2015)
Corethron spp.	107.2	Cloern (2018)
Coscinodiscus spp.	123.7	Harrison *et al.* (2015)
Cylindrotheca closterium	9.3	Harrison *et al.* (2015)
Cylindrotheca gracilis	8.9	Cloern (2018)
Dactyliosolen fragilissimus	25.1	Harrison *et al.* (2015)
Delphineis spp.	7.7	Cloern (2018)
Detonula confervacea	12.1	Harrison *et al.* (2015)
Detonula pumila	28.5	Harrison *et al.* (2015)
Detonula spp.	40.4	Leblanc *et al.* (2012)
Ditylum brightwellii	51.9	Harrison *et al.* (2015)
Eucampia groenlandica	38.2	Cloern (2018)
Eucampia sp.	26.0	Harrison *et al.* (2015)
Eucampia zodiacus	26.0	Harrison *et al.* (2015)
Fragilariopsis spp.	6.0	Widdicombe & Harbour (2021)
Fragilidium spp.	34.0	Cloern (2018)
Fragillaria spp.	10.2	Olenina *et al.* (2006)
Grammatophora spp.	24.5	Widdicombe & Harbour (2021)
Guinardia delicatula	22.5	Harrison *et al.* (2015)
Guinardia flaccida	63.5	Harrison *et al.* (2015)
Guinardia spp.	40.9	Harrison *et al.* (2015)
Guinardia striata	40.9	Harrison *et al.* (2015)
Gyrosigma / Pleurosigma	40.8	Olenina *et al.* (2006)
Hemiaulus spp.	21.8	Cloern (2018)
Lauderia annulata	38.9	Olenina *et al.* (2006)
Lauderia spp.	37.3	Harrison *et al.* (2015)
Leptocylindrus danicus	16.4	Harrison *et al.* (2015)

Continued on next page

Taxon	**ESD (μm)**	Source
Leptocylindrus mediterraneus	26.9	Harrison *et al.* (2015)
Leptocylindrus minimus	8.4	Harrison *et al.* (2015)
Licmophora	14.8	Leblanc *et al.* (2012)
Lithodesmium undulatum	20.2	Widdicombe & Harbour (2021)
Melosira arctica	17.1	Olenina *et al.* (2006)
Melosira spp.	27.0	Harrison *et al.* (2015)
Meuniera membranacea	50.1	Leblanc *et al.* (2012)
Navicula spp.	5.6	Widdicombe & Harbour (2021)
Nitzschia longissima	10.5	Harrison *et al.* (2015)
Nitzschia spp.	13.5	Olenina *et al.* (2006)
Odontella aurita	34.6	Olenina *et al.* (2006)
Odontella sinensis	124.1	Olenina *et al.* (2006)
Odontella spp.	33.8	Harrison *et al.* (2015)
Paralia sulcata	23.7	Harrison *et al.* (2015)
Planktionella sol	13.3	Widdicombe & Harbour (2021)
Podosira spp.	12.8	Cloern (2018)
Podosira stelligera	50.0	Widdicombe & Harbour (2021)
Porosira glacialis	53.3	Olenina *et al.* (2006)
Proboscia alata	34.5	Harrison *et al.* (2015)
Proboscia indica	34.5	Harrison *et al.* (2015)
Pseudo-nitzschia spp.	12.9	Harrison *et al.* (2015)
Raphoneis spp.	28.7	Leblanc *et al.* (2012)
Rhizosolenia hebetata	47.8	Harrison *et al.* (2015)
Rhizosolenia hebetata f. semispina	37.3	Olenina *et al.* (2006)
Rhizosolenia imbricata	54.1	Harrison *et al.* (2015)
Rhizosolenia pungens	29.9	Cloern (2018)
Rhizosolenia robusta	164.1	Widdicombe & Harbour (2021)
Rhizosolenia setigera	32.5	Harrison *et al.* (2015)
Rhizosolenia spp.	51.0	Harrison *et al.* (2015)
Rhizosolenia styliformis	69.5	Harrison *et al.* (2015)
Skeletonema spp.	9.0	Harrison *et al.* (2015)
Stephanopyxis palmeriana	58.5	Widdicombe & Harbour (2021)
Stephanopyxis spp.	39.8	Cloern (2018)
Stephanopyxis turris	32.6	Harrison *et al.* (2015)
Striatella unipunctata	52.4	Leblanc *et al.* (2012)
Tabellaria fenestrata	19.0	Olenina *et al.* (2006)
Tabellaria spp.	14.2	Cloern (2018)
Thalassionema nitzschioides	13.3	Harrison *et al.* (2015)
Thalassionema spp.	13.3	Harrison *et al.* (2015)
Thalassiosira nordenskioeldii	30.9	Harrison *et al.* (2015)
Thalassiosira punctigera	49.4	Olenina *et al.* (2006)
Thalassiosira rotula	28.4	Harrison *et al.* (2015)
Thalassiosira spp.	30.0	Harrison *et al.* (2015)
Triceratium spp.	82.1	Leblanc *et al.* (2012)
Trigonium alternans	21.5	Widdicombe & Harbour (2021)
Trigonium spp.	23.1	Cloern (2018)
DINOFLAGELLATES		
Alexandrium spp.	28.9	Harrison *et al.* (2015)
Alexandrium spp. $10-20\,\mu$m	14.9	
Alexandrium spp. $20-30\,\mu$m	24.8	
Alexandrium spp. $30-40\,\mu$m	34.7	
Alexandrium spp. $40-50\,\mu$m	44.6	
Alexandrium spp. $>50\,\mu$m	54.5	
Amphidinium carterae	9.4	Olenina *et al.* (2006)
Amphidinium crassum	16.7	Olenina *et al.* (2006)
Amphidinium sphenoides	16.7	Olenina *et al.* (2006)
Amphidinium spp.	16.7	Olenina *et al.* (2006)

Continued on next page

Taxon	**ESD (μm)**	Source
Amphidinium spp. $< 10\,\mu$m	5.6	Olenina *et al.* (2006)
Amylax spp.	24.1	Cloern (2018)
Amylax triacantha	27.5	Harrison *et al.* (2015)
Archaeperidinium minutum	30.3	Harrison *et al.* (2015)
Ceratium longipes	52.7	Olenina *et al.* (2006)
Ceratium macroceros	45.6	Olenina *et al.* (2006)
Ceratium spp.	42.6	Olenina *et al.* (2006)
Ceratium spp. $10 - 30\,\mu$m	23.5	Olenina *et al.* (2006)
Ceratium spp. $30 - 50\,\mu$m	35.3	Olenina *et al.* (2006)
Ceratium spp. $50 - 70\,\mu$m	54.4	Olenina *et al.* (2006)
Ceratium spp. $70 - 90\,\mu$m	72.5	Olenina *et al.* (2006)
Cochlodinium spp.	26.3	Cloern (2018)
Dinophysis acuminata	27.3	Harrison *et al.* (2015)
Dinophysis acuta	43.7	Harrison *et al.* (2015)
Dinophysis dens	43.1	Harrison *et al.* (2015)
Dinophysis fortii	49.0	Cloern (2018)
Dinophysis nasuta	20.8	Widdicombe & Harbour (2021)
Dinophysis norvegica	40.8	Olenina *et al.* (2006)
Dinophysis punctata	20.8	Widdicombe & Harbour (2021)
Dinophysis rotundata	24.4	Olenina *et al.* (2006)
Dinophysis sacculus	32.5	Widdicombe & Harbour (2021)
Dinophysis spp.	43.1	Harrison *et al.* (2015)
Dinophysis tripos	49.3	Widdicombe & Harbour (2021)
Diplopsalis spp.	40.4	Harrison *et al.* (2015)
Dissodinium pseudolunula	43.3	Olenina *et al.* (2006)
Gonyaulax digitalis	32.3	Harrison *et al.* (2015)
Gonyaulax spinifera	31.3	Harrison *et al.* (2015)
Gonyaulax spp.	32.3	Harrison *et al.* (2015)
Gonyaulax verior	28.3	Harrison *et al.* (2015)
Gymnodinium galeatum	10.9	Olenina *et al.* (2006)
Gymnodinium spp.	29.1	Olenina *et al.* (2006)
Gyrodinium	35.1	Olenina *et al.* (2006)
Gyrodinium fusiforme	32.6	Harrison *et al.* (2015)
Gyrodinium lachryma	35.1	Olenina *et al.* (2006)
Gyrodinium spirale	36.6	Harrison *et al.* (2015)
Heterocapsa minima	6.3	Olenina *et al.* (2006)
Heterocapsa niei	11.4	Olenina *et al.* (2006)
Heterocapsa rotundata	7.4	Olenina *et al.* (2006)
Heterocapsa spp.	11.4	Olenina *et al.* (2006)
Heterocapsa triquetra	16.7	Harrison *et al.* (2015)
Karenia mikimotoi	29.1	Olenina *et al.* (2006)
Karenia spp.	25.5	Cloern (2018)
Katodinium glaucum	17.2	Harrison *et al.* (2015)
Katodinium spp.	17.2	Harrison *et al.* (2015)
Kofoidinium lebourae	69.3	Widdicombe & Harbour (2021)
Kofoidinium spp.	58.7	Cloern (2018)
Kryptoperidinium foliaceum	27.3	Olenina *et al.* (2006)
Lingulodinium polyedrum	36.7	Olenina *et al.* (2006)
Mesoporos spp.	15.9	Harrison *et al.* (2015)
Oblea rotundata	29.0	Olenina *et al.* (2006)
Oblea spp.	24.9	Harrison *et al.* (2015)
Oxyrrhis marina	15.8	Harrison *et al.* (2015)
Oxytoxum gracilis	7.8	Harrison *et al.* (2015)
Oxytoxum spp.	7.8	Harrison *et al.* (2015)
Pentapharsodinium dalei	14.0	Olenina *et al.* (2006)
Peridinium quadridentatum	23.0	Cloern (2018)
Peridinium spp.	28.9	Cloern (2018)

Continued on next page

Taxon	**ESD (μm)**	Source
Phalacroma oxytoxoides	32.2	Harrison *et al.* (2015)
Phalacroma rotundatum	32.2	Harrison *et al.* (2015)
Polykrikos	64.3	Harrison *et al.* (2015)
Polykrikos schwartzii	64.3	Harrison *et al.* (2015)
Pronoctiluca pelagica	16.1	Olenina *et al.* (2006)
Pronoctiluca spp.	19.3	Harrison *et al.* (2015)
Prorocentrum compressum	13.4	Harrison *et al.* (2015)
Prorocentrum dentatum	6.3	Widdicombe & Harbour (2021)
Prorocentrum gracile	18.6	Cloern (2018)
Prorocentrum lima	43.8	Olenina *et al.* (2006)
Prorocentrum micans	23.9	Harrison *et al.* (2015)
Prorocentrum minimum	13.4	Harrison *et al.* (2015)
Prorocentrum spp.	13.4	Harrison *et al.* (2015)
Prorocentrum triestinum	22.6	Widdicombe & Harbour (2021)
Protoceratium reticulatum	31.2	Harrison *et al.* (2015)
Protoperidinium	51.4	Harrison *et al.* (2015)
Protoperidinium bipes	16.1	Harrison *et al.* (2015)
Protoperidinium brevipes	26.8	Harrison *et al.* (2015)
Protoperidinium cerasus	51.4	Harrison *et al.* (2015)
Protoperidinium conicoides	43.8	Harrison *et al.* (2015)
Protoperidinium conicum	57.5	Harrison *et al.* (2015)
Protoperidinium crassipes	73.4	Harrison *et al.* (2015)
Protoperidinium curtipes	51.4	Harrison *et al.* (2015)
Protoperidinium depressum	89.6	Harrison *et al.* (2015)
Protoperidinium excentricum	51.4	Harrison *et al.* (2015)
Protoperidinium mite	51.4	Harrison *et al.* (2015)
Protoperidinium ovatum	51.4	Harrison *et al.* (2015)
Protoperidinium pallidum	51.4	Harrison *et al.* (2015)
Protoperidinium pellucidum	38.4	Harrison *et al.* (2015)
Protoperidinium steinii	33.5	Harrison *et al.* (2015)
Protoperidinium subinerme	51.4	Harrison *et al.* (2015)
Pyrocystis spp.	41.3	Cloern (2018)
Pyrophacus horologium	43.7	Harrison *et al.* (2015)
Scrippsiella spp.	20.8	Harrison *et al.* (2015)
Scrippsiella trochoidea	20.8	Harrison *et al.* (2015)
Torodinium robustum	18.9	Harrison *et al.* (2015)
Tripos	42.8	Harrison *et al.* (2015)
Tripos furca	44.8	Harrison *et al.* (2015)
Tripos fusus	38.3	Harrison *et al.* (2015)
Tripos hexacanthus	45.7	Widdicombe & Harbour (2021)
Tripos horridus	42.8	Harrison *et al.* (2015)
Tripos lineatus	29.8	Harrison *et al.* (2015)
Tripos minutum	42.8	Harrison *et al.* (2015)
Tripos muelleri	49.4	Harrison *et al.* (2015)
FLAGELLATES		
Crucigenia tetrapedia	4.7	Cloern (2018)
Desmodesmus spp.	5.9	Cloern (2018)
Dictyocha fibula	12.2	Widdicombe & Harbour (2021)
Dictyocha speculum	15.9	Olenina *et al.* (2006)
Eutreptiella spp.	10.1	Cloern (2018)
Octactis octonaria	21.4	Cloern (2018)
Pyramimonas spp.	8.0	Widdicombe & Harbour (2021)
OTHERS		
Chrysochromulina spp.	5.1	Olenina *et al.* (2006)
Pediastrum spp.	7.2	Cloern (2018)
Heterosigma akashiwo	14.5	Cloern (2018)
Phaeocystis	6.0	Widdicombe & Harbour (2021)

Table A.3: Sizes of metazooplankton for SH & LE compiled by Scottish Government, Marine Directorate (2024) and reduced for presentation in this thesis. Converted to ESD from dry weight using the conversion in Table 2.3 (Harris, 1999).

Taxon	**ESD (μm)**	Source
Amphipoda spp.	1251.5	FRS PLA Database
Themisto spp.	1313.1	FRS PLA Database
Aetideus armatus C1 − 6	583.4	FRS PLA Database
Caligidae C1 − 6	453.3	FRS PLA Database
Candacia armata C1	276.3	FRS PLA Database
Candacia armata C1 − 5	524.9	FRS PLA Database
Candacia armata C2	347.5	FRS PLA Database
Candacia armata C3	452.4	FRS PLA Database
Candacia armata C4	582.0	FRS PLA Database
Candacia armata C5	716.9	FRS PLA Database
Candacia armata C6F	931.1	FRS PLA Database
Candacia armata C6M	892.5	FRS PLA Database
Corycaeidae C1 − 6	529.4	FRS PLA Database
Euchaetidae C1 − 6	881.9	FRS PLA Database
Chaetognatha	966.1	FRS PLA Database
Parasagitta elegans	966.1	FRS PLA Database
Parasagitta setosa	966.1	FRS PLA Database
Sagittidae juvenile	610.3	FRS PLA Database
Serratosagitta serratodentata	966.1	FRS PLA Database
Tomopteris spp.	1759.6	FRS PLA Database
Cumacea	976.7	FRS PLA Database
Evadne nordmanni	286.4	FRS PLA Database
Isopoda	453.3	FRS PLA Database
Mysida	1419.2	FRS PLA Database
Podon intermedius	526.3	FRS PLA Database
Podon leuckartii	526.3	FRS PLA Database
Podonidae	526.3	FRS PLA Database
Euterpina acutifrons C1 − 6	401.5	FRS PLA Database
Microsetella norvegica C1 − 6	214.0	FRS PLA Database
Oncaeidae C1 − 6	529.4	FRS PLA Database
Scolecithricella minor C1 − 5	317.5	FRS PLA Database
Scolecithricella minor C6F	607.8	FRS PLA Database
Scolecithricella minor C6M	564.6	FRS PLA Database
Scolecitrichidae spp. C1 − 6	526.0	FRS PLA Database
Euphausiidae adult	908.2	FRS PLA Database
Euphausiidae calyptopis	459.3	FRS PLA Database
Euphausiidae furcilia	908.2	FRS PLA Database
Euphausiidae nauplius	382.4	FRS PLA Database
Nyctiphanes couchii adult	1383.8	FRS PLA Database
Nyctiphanes couchii furcilia	908.2	FRS PLA Database
Nyctiphanes couchii juvenile	459.3	FRS PLA Database
Thysanoessa inermis adult	2131.3	FRS PLA Database
Thysanoessa longicaudata adult	1609.3	FRS PLA Database
Thysanoessa spp. furcilia	908.2	FRS PLA Database
Ammodytidae larva	1158.7	FRS PLA Database
Clupeidae larva	1158.7	FRS PLA Database
Gadiformes larva	2127.2	FRS PLA Database
Pisces egg	453.3	FRS PLA Database
Pisces larva	2127.2	FRS PLA Database
Clione limacina	453.3	FRS PLA Database
Limacina spp.	574.9	FRS PLA Database

Continued on next page

Taxon	**ESD (μm)**	Source
Calanus finmarchicus C5	1374.6	FRS PLA Database Award, NERC Reference: NER/T/S/2001/00692 Head & Harris (2004) Plourde *et al.* (2003)
Calanus finmarchicus C4	1100.1	FRS PLA Database Award, NERC Reference: NER/T/S/2001/00692 Plourde *et al.* (2003)
Calanus finmarchicus C3	950.6	Head & Harris (2004)
Calanus finmarchicus C2	800.4	Head & Harris (2004)
Calanus finmarchicus C6F	1465.3	FRS PLA Database Award, NERC Reference: NER/T/S/2001/00692 Head & Harris (2004) Plourde *et al.* (2003)
Calanus finmarchicus C6M	1401.1	FRS PLA Database Award, NERC Reference: NER/T/S/2001/00692 Head & Harris (2004)
Calanus helgolandicus C5	1036.1	FRS PLA Database
Calanus helgolandicus C6F	1207.4	FRS PLA Database
Calanus helgolandicus C6M	1198.7	FRS PLA Database
Calanus spp. C1	359.6	FRS PLA Database
Calanus spp. C2	461.6	FRS PLA Database
Calanus spp. C3	617.3	FRS PLA Database
Calanus spp. C4	764.4	FRS PLA Database
Calanus spp. C5	1037.1	FRS PLA Database
Calanus spp. C6F	1088.5	FRS PLA Database
Calanus spp. C6M	1063.4	FRS PLA Database
Clausocalanus spp. C1 − 5	379.4	FRS PLA Database
Clausocalanus spp. C6F	762.6	FRS PLA Database
Clausocalanus spp. C6M	736.3	FRS PLA Database
Eucalanus elongatus elongatus C6F	2000.4	FRS PLA Database
Eucalanus elongatus elongatus C1 − 5	1358.2	FRS PLA Database
Eucalanus elongatus elongatus C6M	2000.4	FRS PLA Database
Metridia lucens C1	194.3	FRS PLA Database
Metridia lucens C1 − 5	399.8	FRS PLA Database
Metridia lucens C2	304.2	FRS PLA Database
Metridia lucens C3	369.6	FRS PLA Database
Metridia lucens C4	465.6	FRS PLA Database
Metridia lucens C5	531.6	FRS PLA Database
Metridia lucens C6F	848.5	FRS PLA Database
Metridia lucens C6M	597.7	FRS PLA Database
Metridia spp. C1 − 6	1009.2	FRS PLA Database
Microcalanus pusillus C1 − 5	204.3	FRS PLA Database
Microcalanus pusillus C3	166.2	FRS PLA Database
Microcalanus pusillus C4	186.9	FRS PLA Database
Microcalanus pusillus C5	205.0	FRS PLA Database
Microcalanus pusillus C6F	222.5	FRS PLA Database
Microcalanus pusillus C6M	214.7	FRS PLA Database
Paracalanus parvus C1	173.0	FRS PLA Database
Paracalanus parvus C2	204.3	FRS PLA Database
Paracalanus parvus C3	282.2	FRS PLA Database
Paracalanus parvus C4	325.8	FRS PLA Database
Paracalanus parvus C5	376.3	FRS PLA Database
Paracalanus parvus C6F	416.7	FRS PLA Database
Paracalanus parvus C6M	411.4	FRS PLA Database

Continued on next page

Taxon	**ESD (μm)**	Source
Pseudocalanus spp. C1	240.9	FRS PLA Database
		Conover & Huntley (1991)
Pseudocalanus spp. C2	271.5	FRS PLA Database
		Conover & Huntley (1991)
Pseudocalanus spp. C3	315.1	FRS PLA Database
		Conover & Huntley (1991)
Pseudocalanus spp. C4	368.7	FRS PLA Database
		Conover & Huntley (1991)
Pseudocalanus spp. C5F	426.9	FRS PLA Database
Pseudocalanus spp. C5M	416.4	FRS PLA Database
Pseudocalanus spp. C6F	484.3	FRS PLA Database
		Conover & Huntley (1991)
Pseudocalanus spp. C6M	416.4	FRS PLA Database
		Conover & Huntley (1991)
Rhincalanus nasutus $C1-5$	1070.7	FRS PLA Database
Rhincalanus nasutus C6F	1743.0	FRS PLA Database
Rhincalanus nasutus C6M	1482.4	FRS PLA Database
Subeucalanus crassus $C1-5$	707.4	FRS PLA Database
Subeucalanus crassus C6F	1115.3	FRS PLA Database
Subeucalanus crassus C6M	1088.5	FRS PLA Database
Temora longicornis C1	233.6	FRS PLA Database
Temora longicornis $C1-5$	363.4	FRS PLA Database
Temora longicornis C2	260.1	FRS PLA Database
Temora longicornis C3	313.4	FRS PLA Database
Temora longicornis C4	386.9	FRS PLA Database
Temora longicornis C5	494.0	FRS PLA Database
Temora longicornis C6F	669.9	FRS PLA Database
Temora longicornis C6M	602.2	FRS PLA Database
Aequorea spp.	5189.3	FRS PLA Database
Aglantha digitale	2104.2	FRS PLA Database
Amphinema spp.	2855.8	FRS PLA Database
Aurelia aurita	2855.8	FRS PLA Database
Aurelia spp. ephyrae	1325.5	FRS PLA Database
Beroe cucumis	3598.1	FRS PLA Database
Bougainvillia spp.	2855.8	FRS PLA Database
Coelenterata	2855.8	FRS PLA Database
Ctenophora	2855.8	FRS PLA Database
Cyanea spp.	2855.8	FRS PLA Database
Cyanea spp. ephyrae	1325.5	FRS PLA Database
Diphyidae	775.2	FRS PLA Database
Euphysa spp.	775.2	FRS PLA Database
Eutima gracilis	2855.8	FRS PLA Database
Hybocodon prolifer	1325.5	FRS PLA Database
Laodicea undulata	2855.8	FRS PLA Database
Leptothecata	2855.8	FRS PLA Database
Leuckartiara octona	3598.1	FRS PLA Database
Lizzia blondina	775.2	FRS PLA Database
Mitrocomella polydiademata	2855.8	FRS PLA Database
Muggiaea atlantica	775.2	FRS PLA Database
Nanomia cara	2104.2	FRS PLA Database
Obelia spp.	1325.5	FRS PLA Database
Pelagia noctiluca	2855.8	FRS PLA Database
Phialella quadrata	2408.7	FRS PLA Database
Pleurobrachia pileus	2855.8	FRS PLA Database
Podocoryna	1325.5	FRS PLA Database
Rathkea octopunctata	775.2	FRS PLA Database
Sarsia spp.	2855.8	FRS PLA Database
Scyphozoa ephyrae	1325.5	FRS PLA Database

Continued on next page

Taxon	**ESD (μm)**	Source
Siphonophorae	2104.2	FRS PLA Database
Solmaris corona	5189.3	FRS PLA Database
Staurostoma mertensii	5189.3	FRS PLA Database
Tiaropsis multicirrata	2855.8	FRS PLA Database
Appendicularia	1323.2	FRS PLA Database
Anomura zoea	1575.8	FRS PLA Database
Ascidiacea larva	2127.2	FRS PLA Database
Bivalvia larva	289.3	FRS PLA Database
Brachyura zoea	1575.8	FRS PLA Database
Caridea zoea	1575.8	FRS PLA Database
Cirripedia cypris	524.9	FRS PLA Database
Cirripedia nauplius	259.0	FRS PLA Database
Decapoda larva	1575.8	FRS PLA Database
Echinodermata larva	265.1	FRS PLA Database
Gastropoda spp. egg	123.1	FRS PLA Database
Gastropoda spp. larva	406.3	FRS PLA Database
Lamellaria spp. larva	289.3	FRS PLA Database
Nephrops norvegicus larva	1230.5	FRS PLA Database
Phoronida larva	155.0	FRS PLA Database
Polychaeta larva	486.2	FRS PLA Database
Acartia longeremis C6F	376.6	FRS PLA Database
Acartia clausi C1	189.8	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C1 – 5	284.4	FRS PLA Database
Acartia clausi C2	226.2	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C3	252.2	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C4	314.5	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C5	363.7	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C6F	449.0	FRS PLA Database
		Durbin & Durbin (1978)
Acartia clausi C6M	428.3	FRS PLA Database
		Durbin & Durbin (1978)
Acartia longeremis C1 – 5	252.2	FRS PLA Database
Acartia longeremis C6M	350.8	FRS PLA Database
Anomalocera patersonii C1 – 5	861.2	FRS PLA Database
Anomalocera patersonii C6F	1280.8	FRS PLA Database
Anomalocera patersonii C6M	1147.9	FRS PLA Database
Calanoida C1 – 5	707.4	FRS PLA Database
Calanoida C1 – 6	791.7	FRS PLA Database
Calanoida C6M	1088.5	FRS PLA Database
Calocalanus spp. C6F	367.1	FRS PLA Database
Centropages hamatus C1	211.2	FRS PLA Database
Centropages hamatus C1 – 5	377.7	FRS PLA Database
Centropages hamatus C2	265.3	FRS PLA Database
Centropages hamatus C3	337.2	FRS PLA Database
Centropages hamatus C4	410.6	FRS PLA Database
Centropages hamatus C5	495.0	FRS PLA Database
Centropages hamatus C6F	624.5	FRS PLA Database
Centropages hamatus C6M	558.4	FRS PLA Database
Centropages typicus C1	205.6	FRS PLA Database
Centropages typicus C1 – 5	401.0	FRS PLA Database
Centropages typicus C2	291.4	FRS PLA Database
Centropages typicus C3	352.2	FRS PLA Database
Centropages typicus C4	429.7	FRS PLA Database

Continued on next page

Taxon	**ESD (μm)**	Source
Centropages typicus C5	565.0	FRS PLA Database
Centropages typicus C6F	689.2	FRS PLA Database
Centropages typicus C6M	641.9	FRS PLA Database
Copepoda C1 − 5	707.4	FRS PLA Database
Copepoda C1 − 6	791.7	FRS PLA Database
Ctenocalanus vanus C1 − 5	379.4	FRS PLA Database
Ctenocalanus vanus C6F	762.6	FRS PLA Database
Ctenocalanus vanus C6M	736.3	FRS PLA Database
Cyclopoida C1 − 6	375.1	FRS PLA Database
Diaixis hibernica C1 − 5	707.4	FRS PLA Database
Diaixis hibernica C3	617.3	FRS PLA Database
Diaixis hibernica C4	764.4	FRS PLA Database
Diaixis hibernica C5	1037.1	FRS PLA Database
Diaixis hibernica C5M	1037.1	FRS PLA Database
Diaixis hibernica C6F	1115.3	FRS PLA Database
Diaixis hibernica C6M	1088.5	FRS PLA Database
Harpacticoida C1 − 6	401.5	FRS PLA Database
Monstrilloida C1 − 6	453.3	FRS PLA Database
Oithona atlantica C6F	359.8	FRS PLA Database
Oithona similis C6F	359.8	FRS PLA Database
Oithona similis C6M	359.8	FRS PLA Database
Oithona spp. C1	167.0	FRS PLA Database
Oithona spp. C1 − 5	266.6	FRS PLA Database
Oithona spp. C1 − 6	266.6	FRS PLA Database
Oithona spp. C2	210.4	FRS PLA Database
Oithona spp. C3	266.6	FRS PLA Database
Oithona spp. C4-*5*	334.0	FRS PLA Database
Oithona spp. C6f	359.8	FRS PLA Database
Oithona spp. C6M	359.8	FRS PLA Database
Oithonidae C1 − 6	266.6	FRS PLA Database
Para/Clausocalanidae C1 − 5	293.2	FRS PLA Database
Doliolidae	775.2	FRS PLA Database
Salpa fusiformis	1670.1	FRS PLA Database

Table A.4: Sizes of metazooplankton by season extracted directly from McEvoy *et al.* (2022). The data are included under the Western Channel Observatory data policy, which should be consulted by all users.

Taxon	**Winter ESD (μm)**	**Spring ESD (μm)**	**Summer ESD (μm)**	**Autumn ESD (μm)**
AMPHIPOD				
Caprellida	1931.4	NA	1442.6	1442.6
Gammariida	1931.4	1442.6	1442.6	1442.6
Hyperiida	1512.1	1512.1	1512.1	1512.1
CARNIVOROUS COPEPOD				
Candacia armata	742.8	794.8	742.8	742.8
Ditrichocorycaeus anglicus	314.2	479.8	470.1	386.0
Siphonostomatoida	143.6	143.6	143.6	143.6
DETRITIVOROUS COPEPOD				
Euterpina acutifrons	501.3	427.8	327.3	382.0
Microsetella sp.	501.3	427.8	327.3	382.0
Oncaea spp.	270.4	261.7	279.5	279.9
Scolecithricella spp.	439.4	439.4	439.4	NA
HERBIVOROUS COPEPOD				
Calanoides carinatus	971.0	971.0	971.0	971.0
Calanus copepodites (C1 – 5)	728.7	957.6	835.2	904.6
Calanus eggs	194.8	194.8	194.8	194.8
Calanus finmarchicus (female)	825.1	990.0	1024.3	1107.8
Calanus helgolandicus	762.4	961.6	857.6	940.6
Calanus helgolandicus (female)	825.1	990.0	1024.3	1107.8
Calanus helgolandicus (male)	825.1	990.0	1024.3	1107.8
Copepod nauplii	186.5	260.2	253.7	265.9
Eurytemora affinis	NA	338.3	NA	338.3
Metridia lucens	460.0	595.8	865.7	460.0
Microcalanus spp.	261.7	261.7	261.7	261.7
Paracalanus parvus (female)	403.2	403.2	403.2	403.2
Paracalanus parvus (male)	403.2	403.2	403.2	403.2
Pseudocalanus elongatus (female)	518.7	518.7	518.7	518.7
Pseudocalanus elongatus (male)	518.7	518.7	518.7	518.7
Stephos scotti	286.5	286.5	NA	NA
Subeucalanus spp.	1248.5	946.4	946.4	946.4
Temora longicornis	429.2	556.4	403.9	457.8
Temora longicornis (C1 – 5)	361.9	447.9	362.1	415.1
Temora longicornis (female)	514.3	717.0	552.3	514.3

Continued on next page

Taxon	Winter ESD (μm)	Spring ESD (μm)	Summer ESD (μm)	Autumn ESD (μm)
Temora longicornis (male)	514.3	717.0	552.3	514.3
Temora stylifera	514.3	NA	552.3	514.3
OMNIVOROUS COPEPOD				
Acartia clausi	342.0	397.4	331.8	315.8
Acartia clausi (C1 − 5)	258.1	375.6	291.2	258.1
Acartia clausi (female)	370.5	446.4	396.5	370.5
Acartia clausi (male)	370.5	446.4	396.5	370.5
Anomalocera patersoni	1457.9	1457.9	1457.9	1457.9
Calocalanus spp.	361.8	361.8	361.8	361.8
Centropages chierchiae	601.3	896.0	601.3	751.3
Centropages hamatus	601.3	896.0	601.3	751.3
Centropages spp (C1 − 5)	526.3	526.3	526.3	614.9
Centropages typicus	541.0	725.6	548.2	657.7
Centropages typicus (female)	601.3	896.0	601.3	751.3
Centropages typicus (male)	601.3	896.0	601.3	751.3
Clausocalanus spp. (female)	516.2	516.2	516.2	516.2
Clausocalanus spp. (male)	516.2	516.2	516.2	516.2
Ctenocalanus vanus (female)	495.2	495.2	495.2	495.2
Ctenocalanus vanus (male)	495.2	495.2	495.2	495.2
Diaixis hibernica	485.0	485.0	NA	485.0
Goniopsyllus clausi	501.3	427.8	327.3	382.0
Harpacticoid unidentified	501.3	427.8	327.3	382.0
Isias clavipes	NA	385.3	385.3	385.3
Labidocera wollastoni	NA	1034.7	1034.7	1034.7
Monstrilloida	501.3	NA	NA	382.0
Oithona spp.	196.2	215.0	207.1	199.8
Para/Pseudo/Cteno/Clausocalanus Unidentified (C1 − 5)	298.9	341.3	332.2	288.5
Paraeuchaeta hebes	1112.4	1112.4	1112.4	1112.4
Parapontella brevicornis	525.4	525.4	525.4	525.4
CHAETOGNATH				
Branchiostoma (Cephalochordata)	811.4	811.4	811.4	811.4
Chaetognatha eggs	484.4	484.4	484.4	484.4
Chaetognatha unidentified	807.6	807.7	651.9	618.9
Parasagitta elegans	1593.2	1318.1	1439.7	1318.1
Parasagitta setosa	1593.2	1318.1	1439.7	1318.1
Polychaeta larvae unidentified	276.1	967.1	1165.8	643.1
Tomopteris helgolandica	2244.3	2244.3	2244.3	2244.3

Continued on next page

Taxon	Winter ESD (μm)	Spring ESD (μm)	Summer ESD (μm)	Autumn ESD (μm)
CLADOCERAN				
Cumacea	1126.2	1126.2	1126.2	1126.2
Evadne spp.	407.3	432.6	407.3	407.3
Isopoda	1931.4	1442.6	1442.6	1442.6
Mysida	2722.8	2722.8	2722.8	2722.8
Ostracoda	378.4	NA	378.4	378.4
Podon spp.	312.3	446.5	356.3	312.3
EUPHAUSIID				
Euphausiid adult	2361.5	1768.2	1768.2	1768.2
Euphausiid calyptopis	865.4	865.4	865.4	865.4
Euphausiid furcilia	1405.1	1405.1	1405.1	1405.1
Euphausiid nauplii	592.3	592.3	592.3	592.3
FISH EGGS / LARVAE				
Fish eggs	1575.8	1734.4	1575.8	1575.8
Fish larvae	1761.6	2662.8	2980.4	3222.7
Sardine eggs *(Clupeidae)*	2419.3	2419.3	2419.3	2419.3
GASTROPOD				
Bivalvia	303.3	303.3	303.3	303.3
Clione	634.9	634.9	634.9	634.9
Limacina retroversa	266.1	304.8	301.9	341.8
JELLYFISH				
Agalma elegans (nectophore)	1227.4	1227.4	1227.4	1227.4
Aglantha digitale	2053.0	2053.0	2053.0	2053.0
Amphinema spp.	2032.7	2032.7	2032.7	2032.7
Aquorea	NA	NA	1739.0	NA
Aurelia aurita ephyrae	NA	2757.3	NA	NA
Beroe spp.	3400.4	3400.4	33 053.6	3400.4
Bougainvillia muscus	1738.8	1738.8	1738.8	1738.8
Clytia hemisphaerica	1644.1	1524.0	1524.0	1524.0
Corymorpha nutans	NA	2139.8	2139.8	2139.8
Coryne prolifer	571.2	571.2	571.2	950.8
Cosmetira pilosella	4460.8	4460.8	5325.8	NA
Ctenophore unidentified	1496.2	5011.4	5373.1	1496.2
Ectopleura dumortierii	NA	1680.1	1680.1	1680.1
Eirene viridula	NA	711.4	711.4	711.4
Eutima gracilis	NA	1631.9	1631.9	1631.9
Hybocodon prolifer	NA	1738.8	1738.8	NA
Laodicea undulata	NA	1739.0	1739.0	1739.0

Continued on next page

Taxon	Winter ESD (μm)	Spring ESD (μm)	Summer ESD (μm)	Autumn ESD (μm)
Leukartiara octona	1437.7	7345.3	6921.0	1437.7
Liriope tetraphylla	1248.7	1248.7	1248.7	1248.7
Lizzia blondina	776.5	549.4	530.8	530.8
Lovenella clausa	2163.3	2163.3	2163.3	2163.3
Mitrocomella brownei	1877.5	1877.5	1877.5	1877.5
Muggiaea atlantica (polygastric)	1931.6	2318.4	1914.0	1652.3
Muggiaea sp. *(eudoxid)*	833.0	984.3	811.2	1667.1
Nanomia cara (nectophore)	1227.4	1227.4	1227.4	1227.4
Obelia spp.	1461.0	1459.8	945.4	747.9
Phialella quadrata	1132.5	1132.5	1132.5	1132.5
Pleurobrachia pileus	1496.2	5011.4	5373.1	1496.2
Podocoryna areolata	426.7	426.7	426.7	426.7
Rathkea octopunctata	1445.5	1445.5	NA	NA
Sarsia spp.	3213.6	3213.6	3213.6	3213.6
Scyphozoan ephyrae	2757.3	2757.3	2757.3	2757.3
Siphonophore unidentified	1227.4	1227.4	1227.4	1227.4
Slobberina halterata	NA	NA	3213.6	3213.6
Solmaris corona (Narcomedusae)	538.4	538.4	538.4	538.4
Stauridiosarsia gemmifera	NA	1738.8	1738.8	1738.8
Trachymedusae	1743.5	1743.5	1743.5	1743.5
Turritopsis nutricula	1738.8	NA	NA	NA
Zanclea costata	1299.1	1299.1	1299.1	1299.1
LARVACEAN				
Appendicularia	1218.2	1341.8	1247.5	1218.2
MEROPLANKTON LARVAE				
Asterioid bipinnaria/brachiolaria	146.6	146.6	314.6	146.6
Asteroid juvenile	146.6	146.6	314.6	146.6
Auricularia larvae *(Holothuria)*	146.6	146.6	314.6	146.6
Brachyuran larvae	843.1	1046.3	1573.2	1258.4
Bryozoa cyphonautes larvae	391.5	391.5	391.5	391.5
Cephalopoda larvae	2419.3	2419.3	2419.3	2419.3
Cirripede cyprid	510.3	510.3	510.3	510.3
Cirripede nauplii	571.7	514.3	398.4	472.9
Decapod larvae unidentified	843.1	1046.3	1573.2	1258.4
Doliolaria larvae *(Holothuria)*	NA	146.6	314.6	146.6
Echinoderm larvae unidentified	146.6	146.6	314.6	146.6
Echinoid juvenile (Sea urchin larvae)	NA	146.6	314.6	146.6
Echinopluteus larvae	146.6	146.6	314.6	146.6

Continued on next page

Taxon	**Winter ESD (μm)**	**Spring ESD (μm)**	**Summer ESD (μm)**	**Autumn ESD (μm)**
Entoproct larvae	219.9	219.9	219.9	219.9
Gastropod larvae	266.1	304.8	301.9	341.8
Gymnosome larvae	355.9	355.9	355.9	355.9
Lamellaria echinospira larvae	275.3	275.3	275.3	275.3
Luidia sp. larvae	146.6	146.6	314.6	146.6
Nemertea pilidium larvae	462.2	462.2	462.2	462.2
Ophiopluteus larvae	146.6	146.6	314.6	146.6
Ophiuroid juvenile	146.6	146.6	314.6	146.6
Phoronida actinotroch larvae	249.5	249.5	249.5	249.5
Porcellanid larvae	843.1	1046.3	1573.2	1258.4
Tornaria larvae *(Hemichordata)*	NA	1067.8	1067.8	1067.8
Trochophore larvae	NA	NA	122.1	122.1
SALP				
Doliolida	980.7	980.7	980.7	980.7

Appendix B

Diagnostics of fitted statistical models

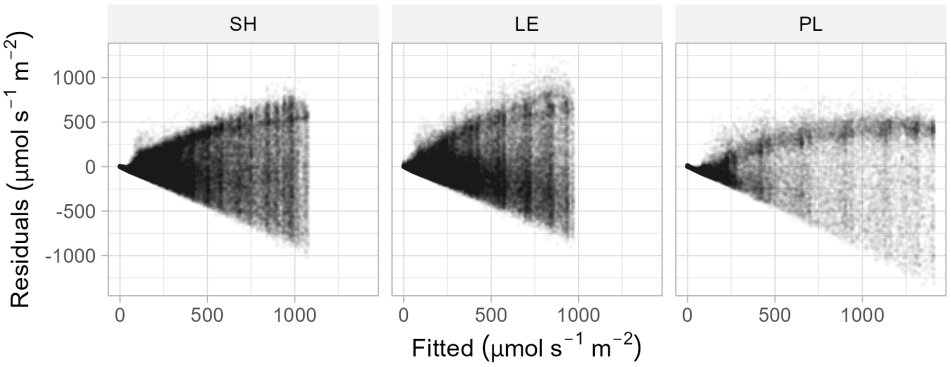


Figure B.1: Residual errors against fitted values of model fitted to PAR SH, LE, and PL (Figure 2.9).

Table B.1: Summaries of GAMs fitted to various response variables, with a common explanatory cyclic smooth term of day of year.

Response	**Site**	p	R^2	n	**Figure**	**Time**	**Notes**
ln(DN)	SH	$< 2 \times 10^{-16}$	0.747	1051		1997-01-15 to 2020-12-20	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DN)	LE	$< 2 \times 10^{-16}$	0.725	1027		2002-06-24 to 2020-12-27	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DN)	PL	$< 2 \times 10^{-16}$	0.733	595		2000-01-17 to 2022-08-01	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DSi)	SH	$< 2 \times 10^{-16}$	0.495	1357		1997-01-15 to 2020-12-20	Several extreme negative residuals when fitted values are low. Significant negative skew on Q-Q. Quantitative inference near LOD tentative; satisfactory above this.
ln(DSi)	LE	$< 2 \times 10^{-16}$	0.478	1448		2002-06-24 to 2020-12-27	Small number of extreme negative residuals. Significant negative skew on Q-Q. Quantitative inference near LOD tentative; satisfactory above this.
ln(DSi)	PL	$< 2 \times 10^{-16}$	0.639	791		2000-01-17 to 2022-08-01	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DP)	SH	$< 2 \times 10^{-16}$	0.464	1151		1997-01-15 to 2020-12-20	Several extreme negative residuals when fitted values are low. Significant negative skew on Q-Q. Quantitative inference near LOD tentative; satisfactory above this.
ln(DP)	LE	$< 2 \times 10^{-16}$	0.451	1210		2002-06-24 to 2020-12-27	Several extreme negative residuals at all fitted values. Significant negative skew on Q-Q. Quantitative inference near LOD tentative; satisfactory above this.
ln(DP)	PL	$< 2 \times 10^{-16}$	0.638	836		2000-01-17 to 2022-08-01	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DN:DP)	SH	$< 2 \times 10^{-16}$	0.427	815		1997-02-12 to 2020-12-01	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.

Continued on next page

Response	Site	p	R^2	n	Figure	Time	Notes
ln(DN:DP)	LE	$< 2 \times 10^{-16}$	0.371	865		2002-06-24 to 2020-12-27	Moderate decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DN:DP)	PL	$< 2 \times 10^{-16}$	0.381	2018		2000-01-17 to 2022-08-01	Slight decrease in variability of residuals with fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DSi:DP)	SH	$< 2 \times 10^{-16}$	0.107	815		1997-02-12 to 2020-12-01	Moderate decrease in variability of residuals with fitted values; some extreme negative residuals at low fitted values. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DSi:DP)	LE	$< 2 \times 10^{-16}$	0.0548	865		2002-06-24 to 2020-12-27	Residuals satisfactory. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
ln(DSi:DP)	PL	$< 2 \times 10^{-16}$	0.347	2018		2000-01-17 to 2022-08-01	Residuals satisfactory. Moderate overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
SST	SH	$< 2 \times 10^{-16}$	0.947	525		2008-01-07 to 2020-12-01	Satisfactory residuals. Satisfactory Q-Q.
SST	LE	$< 2 \times 10^{-16}$	0.884	660		2008-01-21 to 2020-12-27	Satisfactory residuals. Slight overdispersion on Q-Q.
SST	PL	$< 2 \times 10^{-16}$	0.912	483		2008-02-13 to 2018-12-17	Satisfactory residuals. Satisfactory Q-Q.
Near-bed T	SH	$< 2 \times 10^{-16}$	0.957	519		2008-01-07 to 2020-12-01	Satisfactory residuals. Satisfactory Q-Q.
Near-bed T	LE	$< 2 \times 10^{-16}$	0.949	661		2008-01-21 to 2020-12-27	Satisfactory residuals. Satisfactory Q-Q.
Near-bed T	PL	$< 2 \times 10^{-16}$	0.919	482		2008-02-13 to 2018-12-17	Satisfactory residuals. Satisfactory Q-Q.
MLD	SH	$< 2 \times 10^{-16}$	0.632	366		1990-01-01 to 2014-12-31	Large decrease in variability of residuals with fitted values. Slight positive skew on Q-Q. Slight asymmetry on residuals and Q-Q so inference on mean values tentative.
MLD	LE	0.131	0.0257	233		2000-01-01 to 2024-12-31	Small number of extreme positive residuals. Moderate positive skew on Q-Q. No evidence to change conclusion of insufficient evidence for significance.

Continued on next page

Response	Site	p	R^2	n	Figure	Time	Notes
MLD	PL	$< 2 \times 10^{-16}$	0.343	714		2002-01-17 to 2018-12-17	Slight asymmetry and increase in variability of residuals with fitted values. Slight underdispersion on Q-Q. Slight asymmetry on residuals and Q-Q so inference on mean values tentative.
PAR	SH	$< 2 \times 10^{-16}$	0.733	129306		2006-02-29 to 2020-12-31	Large increase in variability of residuals with fitted values. Large overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
PAR	LE	$< 2 \times 10^{-16}$	0.687	137389		2006-01-17 to 2021-12-31	Large increase in variability of residuals with fitted values. Large overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
PAR	PL	$< 2 \times 10^{-16}$	0.752	40860		2012-12-17 to 2019-10-25	Large increase in variability of residuals with fitted values. Large overdispersion on Q-Q. Symmetry on residuals and Q-Q so inference on mean appropriate.
$\ln(k_d)$	SH	$< 2 \times 10^{-16}$	0.18	683		2001-05-09 to 2016-12-19	Satisfactory residuals. Slight positive skew.
$\ln(k_d)$	LE	$< 2 \times 10^{-16}$	0.144	311		2002-06-24 to 2016-12-19	Satisfactory residuals. Satisfactory Q-Q.
$\ln(k_d)$	PL	3.01×10^{-3}	0.0314	313		2009-02-11 to 2017-05-22	Residuals satisfactory. Slight overdispersion on Q-Q.
$\ln(k_w)$	SH	$< 2 \times 10^{-16}$	0.329	682		2001-05-09 to 2016-12-19	Some large negative residuals for small fitted values. Slight overdispersion on Q-Q.
$\ln(k_w)$	LE	$< 2 \times 10^{-16}$	0.178	307		2002-06-24 to 2016-12-19	Residuals & Q-Q satisfactory, except for 3 extreme negative residuals
$\ln(k_w)$	PL	$< 2 \times 10^{-16}$	0.146	311		2009-02-11 to 2017-05-22	Residuals satisfactory. Slight overdispersion on Q-Q.
$\ln(k_{\text{Chl}} \cdot \text{Chl})$	SH	$< 2 \times 10^{-16}$	0.694	683		2001-05-09 to 2016-12-19	Slight increase in variability of residuals with fitted values. Slight overdispersion on Q-Q.
$\ln(k_{\text{Chl}} \cdot \text{Chl})$	LE	$< 2 \times 10^{-16}$	0.539	311		2002-06-24 to 2016-12-19	Slight increase in variability of residuals with fitted values. Q-Q satisfactory.
$\ln(k_{\text{Chl}} \cdot \text{Chl})$	PL	$< 2 \times 10^{-16}$	0.28	313		2009-02-11 to 2017-05-22	Slight increase in variability of residuals with fitted values. Q-Q satisfactory.
Phyto. $\ln(B)$	SH	$< 2 \times 10^{-16}$	0.739	144		2015-01-19 to 2018-12-05	Residuals satisfactory. Q-Q satisfactory.

Continued on next page

Response	Site	p	R^2	n	Figure	Time	Notes
Phyto. ln(B)	PL	$< 2 \times 10^{-16}$	0.576	418		2007-04-23 to 2017-12-18	Slight increase in variability of residuals with fitted values. Slight positive skew on Q-Q.
Phyto. E[s]	SH	$< 2 \times 10^{-16}$	0.310	144		2015-01-19 to 2018-12-05	One extreme outlying positive residual. Slight underdispersion on Q-Q.
Phyto. E[s]	PL	$< 2 \times 10^{-16}$	0.126	311		2007-04-23 to 2017-12-18	Two extreme outlying positive residuals. Slight positive skew on Q-Q.
Phyto. Var[s]	SH	0.0238	0.0742	418		2015-01-19 to 2018-12-05	Residuals satisfactory. Q-Q satisfactory.
Phyto. Var[s]	PL	2.83×10^{-6}	0.0714	418		2007-04-23 to 2017-12-18	14 extreme outlying positive residuals across the range of fitted values. Moderate positive skew on Q-Q.
Zoo. ln(B)	SH	$< 2 \times 10^{-16}$	0.715	911		2000-01-16 to 2020-12-01	Residuals satisfactory. Q-Q satisfactory.
Zoo. ln(B)	LE	$< 2 \times 10^{-16}$	0.586	820		2002-04-01 to 2017-12-31	One extreme outlying negative residual. Slight negative skew on Q-Q.
Zoo. ln(B)	PL	$< 2 \times 10^{-16}$	0.583	1131		1996-01-02 to 2021-11-29	Residuals satisfactory. Q-Q satisfactory.
Zoo. E[s]	SH	$< 2 \times 10^{-16}$	0.145	911		2000-01-16 to 2020-12-01	Residuals satisfactory, some modal fitted values. Q-Q satisfactory.
Zoo. E[s]	LE	9.72×10^{-6}	0.0343	820		2002-04-01 to 2017-12-31	Residual mean appears to be slightly below zero. Moderate negative skew on Q-Q.
Zoo. E[s]	PL	$< 2 \times 10^{-16}$	0.129	1131		1996-01-02 to 2021-11-29	Residuals satisfactory. Q-Q satisfactory.
Zoo. Var[s]	SH	$< 2 \times 10^{-16}$	0.123	911		2000-01-16 to 2020-12-01	Change in residual structure but with consistent mean zero. Slight underdispersion on Q-Q.
Zoo. Var[s]	LE	$< 2 \times 10^{-16}$	0.0959	920		2002-04-01 to 2017-12-31	Residuals satisfactory. Slight underdispersion on Q-Q.
Zoo. Var[s]	PL	$< 2 \times 10^{-16}$	0.067	1131		1996-01-02 to 2021-11-29	Residuals satisfactory. Q-Q satisfactory.
Micro. ln(B)	SH	$< 2 \times 10^{-16}$	0.537	874		2000-01-16 to 2020-12-01	11 extreme negative residuals. Slight negative skew on Q-Q.
Micro. ln(B)	LE	$< 2 \times 10^{-16}$	0.549	1037		2002-07-08 to 2020-12-17	12 extreme negative residuals. Slight negative skew on Q-Q.

Continued on next page

Response	Site	p	R^2	n	Figure	Time	Notes
Micro. ln(B)	PL	$< 2 \times 10^{-16}$	0.604	1183		1992-10-05 to 2020-12-07	Residuals satisfactory. Q-Q satisfactory.
Micro. E[s]	SH	$< 2 \times 10^{-16}$	0.155	874		2000-01-16 to 2020-12-01	Residuals satisfactory. Slight overdispersion on Q-Q.
Micro. E[s]	LE	$< 2 \times 10^{-16}$	0.248	1037		2002-07-08 to 2020-12-17	Some evidence of modality in residuals; a line of negative residuals when fitted value is < 3. Slight overdispersion on Q-Q.
Micro. E[s]	PL	0.00172	0.00925	1183		1992-10-05 to 2020-12-07	Residuals satisfactory. Slight positive skew on Q-Q.
Micro. Var[s]	SH	$< 2 \times 10^{-16}$	0.240	874		2000-01-16 to 2020-12-01	Slight increase in variability of residuals with fitted values. Q-Q satisfactory.
Micro. Var[s]	LE	$< 2 \times 10^{-16}$	0.102	1037		2002-07-08 to 2020-12-17	Slight increase in variability of residuals with fitted values. Q-Q satisfactory.
Micro. Var[s]	PL	$< 2 \times 10^{-16}$	0.0685	1183		1992-10-05 to 2020-12-07	Residuals satisfactory. Q-Q satisfactory.
Copepod ln(B)	SH	$< 2 \times 10^{-16}$	0.510	1056		1997-01-06 to 2020-12-01	Residuals satisfactory. Slight overdispersion on Q-Q.
Copepod ln(B)	LE	$< 2 \times 10^{-16}$	0.278	819		2002-04-01 to 2017-12-31	6 extreme negative residuals when fitted value is high (> 1.5). Moderate negative skew on Q-Q .
Copepod ln(B)	PL	$< 2 \times 10^{-16}$	0.354	1492		1988-03-14 to 2021-11-29	4 extreme negative residuals when fitted value is high (> 1.5). Slight negative skew on Q-Q.
Copepod E[s]	SH	$< 2 \times 10^{-16}$	0.129	1056		1997-01-06 to 2020-12-01	Residuals satisfactory. Q-Q satisfactory.
Copepod E[s]	LE	$< 2 \times 10^{-16}$	0.279	819		2002-04-01 to 2017-12-31	Residuals satisfactory. Q-Q satisfactory.
Copepod E[s]	PL	$< 2 \times 10^{-16}$	0.283	1492		1988-03-14 to 2021-11-29	Residuals satisfactory. Q-Q satisfactory.
Copepod Var[s]	SH	$< 2 \times 10^{-16}$	0.0867	1056		1997-01-06 to 2020-12-01	Residuals satisfactory. Q-Q satisfactory.
Copepod Var[s]	LE	$< 2 \times 10^{-16}$	0.222	819		2002-04-01 to 2017-12-31	Residuals satisfactory. Q-Q satisfactory.
Copepod Var[s]	PL	$< 2 \times 10^{-16}$	0.0537	1492		1988-03-14 to 2021-11-29	Residuals satisfactory. Slight positive skew on Q-Q.

Table B.2: Summaries of LMs fitted to indicators of plankton community size structure. Each model has $n = 568$ observations, those from PL spanning 2007-04-23 to 2007-12-18 and at SH 2015-01-19 to 2018-12-05.

Response	**Explanatory**	**Estimate**	p	R^2**-adj.**	**Comments**
Phyto. E[s]			$< 2.2 \times 10^{-16}$	0.4324	Slight positive skew inflicted by 3 extreme positive residuals
	Intercept	1.03996	$< 2 \times 10^{-16}$		
	PL	−0.01626	0.916		
	ln(Phyto. B)	0.44294	$< 2 \times 10^{-16}$		
	ln(Phyto. B)∗PL	−0.23424	1.69×10^{-7}		
Zoo. E[s]			6.186×10^{-13}	0.1051	Residuals satisfactory. Slight negative skew in Q-Q plot.
	Intercept	5.81760	$< 2 \times 10^{-16}$		
	PL	0.20250	3.516×10^{-3}		
	ln(Zoo. B)	−3.88583	1.87×10^{-4}		
	ln(Zoo. B)^2	-0.47687	0.617603		
	ln(Zoo. B)*PL	-0.21579	0.889504		
	ln(Zoo. B)^2*PL	-3.37594	0.036308		
ln(Phyto. Var[log(ESD)])			2.055×10^{-9}	0.07697	Residual and Q-Q plots satisfactory
	Intercept	-0.12373	0.012169		
	PL	-0.23040	1.51×10^{-4}		
	ln(Phyto. B)	-0.70067	0.458071		
	ln(Phyto. B)^2	-3.08826	7.05×10^{-4}		
	ln(Phyto. B)*PL	6.01897	6.96×10^{-4}		
	ln(Phyto. B)^2*PL	-4.58266	5.615×10^{-3}		
ln(Zoo. Var[log(ESD)])			8.779×10^{-14}	0.117	Residual and Q-Q plots satisfactory
	Intercept	0.17190	0.00125		
	PL	0.30401	1.16×10^{-6}		
	ln(Zoo. B)	-1.75637	0.05801		
	ln(Zoo. B)^2	-4.88523	1.77×10^{-8}		Unimodal pattern driven by points where no microzooplankton recorded. Possibly a result of unicellular sampling methodology at SH.
	ln(Zoo. B)*PL	-0.23399	0.86633		
	ln(Zoo. B)^2*PL	6.25902	1.64×10^{-5}		No concerns about sampling at PL.
ln(Zoo. B)			$< 2.2 \times 10^{-16}$	0.4929	Residuals satisfactory. Slight overdispersion in Q-Q.

Continued on next page

Response	Explanatory	Estimate	p	R^2-adj.	Comments
	Intercept	2.80968	$< 2 \times 10^{-16}$		
	PL	-0.31383	4.2410^{-4}		
	ln(Phyto. B)	13.64875	$< 2 \times 10^{-16}$		
	ln(Phyto. B)^2	-5.93379	$< 9.62 \times 10^{-16}$		
	ln(Phyto. B)*PL	15.68830	2.44×10^{-9}		
	ln(Phyto. B)^2*PL	-9.04857	1.98×10^{-4}		
Zoo. E[ln(ESD)]			$< 2.2 \times 10^{-16}$	0.06374	Residuals satisfactory. Slight overdispersion in Q-Q.
	Intercept	5.98545	$< 2 \times 10^{-16}$		
	Phyto. E[ln(ESD)]	-3.50661	2.42×10^{-7}		
	Phyto. E[ln(ESD)]^2	1.82776	6.62×10^{-3}		
ln(Zoo. Var[ln(ESD)])			1.59×10^{-10}	0.07501	Residuals satisfactory. Q-Q satisfactory.
	Intercept	0.06538			
	PL		7.23×10^{-10}		
	ln(Phyto. Var[ln(ESD)])		1.55×10^{-3}		

Table B.3: Summaries of GAMLSS models fitted to estimates of bias. p-values for variances not shown; these were always $p < 2 \times 10^{-16}$.

Response	Mean & variance formula (identical)	Subsample volume (V)	Mean intercept p	Mean smooth p	Psuedo-R^2	Notes
Bias[ln($\hat{b}$)]	cs(ln(b))	100	0.493	$< 2 \times 10^{-16}$	0.756	Satisfactory
Bias[ln($\hat{b}$)]	cs(ln(b))	30	0.139	$< 2 \times 10^{-16}$	0.725	Satisfactory
Bias[ln($\hat{b}$)]	cs(ln(b))	10	$< 1 \times 10^{-5}$	$< 2 \times 10^{-16}$	0.688	Satisfactory
Bias[ln($\hat{b}$)]	cs(ln(b))	3	$< 2 \times 10^{-16}$	0.248	0.616	Satisfactory
Bias[ln($\hat{b}$)]	cs(ln(b))	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.570	Slight positive skew in Q-Q
Bias[E$\hat{[s]}$]	cs(ln(b))	100	$< 1 \times 10^{-10}$	$< 2 \times 10^{-16}$	0.859	Satisfactory
Bias[E$\hat{[s]}$]	cs(ln(b))	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.850	Satisfactory
Bias[E$\hat{[s]}$]	cs(ln(b))	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.824	Satisfactory
Bias[E$\hat{[s]}$]	cs(ln(b))	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.772	Slight positive skew in Q-Q
Bias[E$\hat{[s]}$]	cs(ln(b))	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.736	Slight positive skew in Q-Q
Bias[ln(V$\hat{ar}$[s])]	cs(ln(b))	100	$< 2 \times 10^{-16}$	$< 4 \times 10^{-4}$	0.802	Satisfactory
Bias[ln(V$\hat{ar}$[s])]	cs(ln(b))	30	$< 2 \times 10^{-16}$	$< 4 \times 10^{-3}$	0.824	Satisfactory
Bias[ln(V$\hat{ar}$[s])]	cs(ln(b))	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.797	Satisfactory
Bias[ln(V$\hat{ar}$[s])]	cs(ln(b))	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.764	Satisfactory

Continued on next page

Response	**Mean & variance formula (identical)**	**Subsample volume** (V)	**Mean intercept** p	**Mean smooth** p	**Psuedo-**R^2	**Notes**
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\ln(b))$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.756	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(\text{E}[s])$	100	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.859	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(\text{E}[s])$	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.850	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(\text{E}[s])$	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.824	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(\text{E}[s])$	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.771	Slight positive skew in Q-Q
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(\text{E}[s])$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.735	Slight positive skew in Q-Q
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\text{Var}[s])$	100	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.802	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\text{Var}[s])$	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.825	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\text{Var}[s])$	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.797	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\text{Var}[s])$	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.764	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(\text{Var}[s])$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.756	Satisfactory
$\text{Bias}[\ln(\hat{b})]$	$\text{cs}(u) + \text{cs}(v)$	100	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.756	Satisfactory
$\text{Bias}[\ln(\hat{b})]$	$\text{cs}(u) + \text{cs}(v)$	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.725	Satisfactory
$\text{Bias}[\ln(\hat{b})]$	$\text{cs}(u) + \text{cs}(v)$	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.688	Satisfactory
$\text{Bias}[\ln(\hat{b})]$	$\text{cs}(u) + \text{cs}(v)$	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.615	Satisfactory
$\text{Bias}[\ln(\hat{b})]$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.567	Slight positive skew in Q-Q
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(u) + \text{cs}(v)$	100	$< 2 \times 10^{-16}$	$< 1 \times 10^{-3}$	0.859	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(u) + \text{cs}(v)$	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.850	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(u) + \text{cs}(v)$	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.823	Satisfactory
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(u) + \text{cs}(v)$	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.772	Slight positive skew in Q-Q
$\text{Bias}[\hat{\text{E}}[s]]$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.736	Slight positive skew in Q-Q
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(u) + \text{cs}(v)$	100	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.802	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(u) + \text{cs}(v)$	30	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.824	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(u) + \text{cs}(v)$	10	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.797	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(u) + \text{cs}(v)$	3	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.764	Satisfactory
$\text{Bias}[\ln(\hat{\text{Var}}[s])]$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.756	Slight positive skew in Q-Q
$\ln(\hat{b})$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.924	Slight positive skew in Q-Q
$\hat{\text{E}}[s]$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.703	Slight positive skew in Q-Q
$\ln(\hat{\text{Var}}[s])$	$\text{cs}(u) + \text{cs}(v)$	1	$< 2 \times 10^{-16}$	$< 2 \times 10^{-16}$	0.771	Satisfactory

Appendix C

Supplementary figures for interannual variability of plankton community size structure

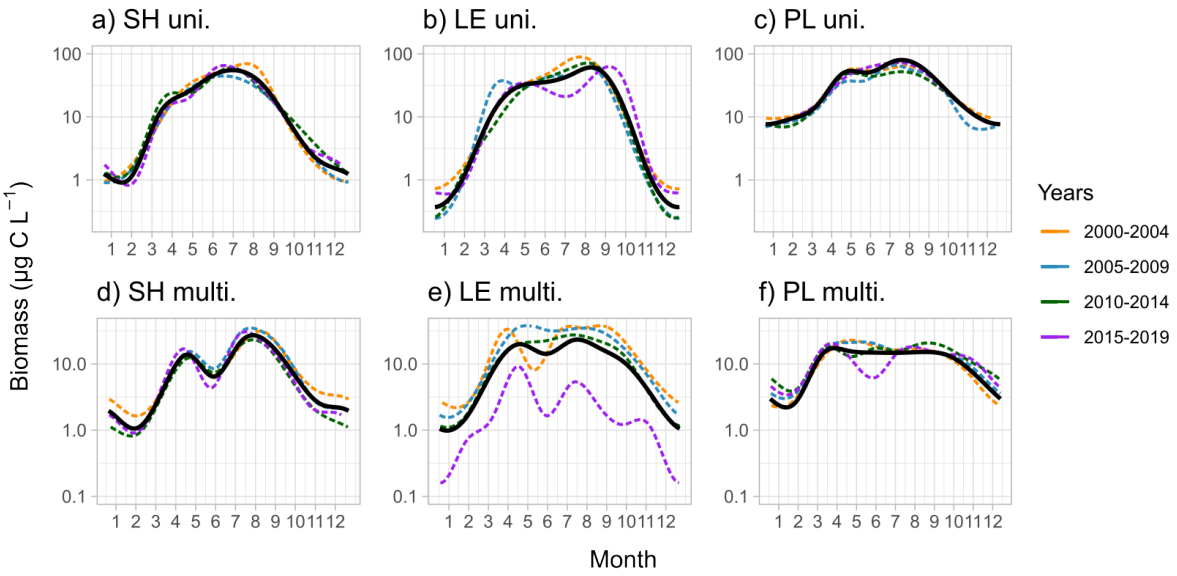


Figure C.1: Seasonal cycle of biomass of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined (2000 – 2019) data.

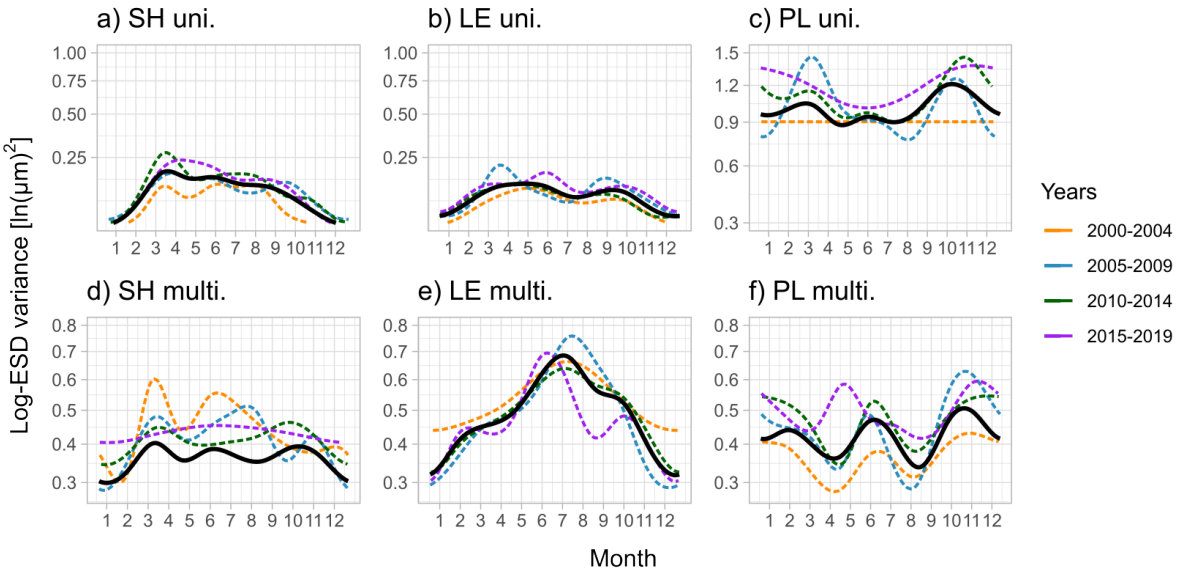


Figure C.2: Seasonal cycle of log-ESD variance of plankton communities at SH, LE, and PL across 5-year time periods. Black lines show models fitted to the combined (2000 – 2019) data.

Appendix D

Supplementary text for the sampling bias simulation experiment

D.1 Decomposing sources of bias

In Figures 3.7 & 3.8, it is shown that with increasing biomass, bias is decreasing in magnitude. This is useful from a practical perspective, as it demonstrates association which can be used in practice to guide methodology when measuring size structure. However, by including random effects in the model for location, the effects are decomposed to better understand what drives the bias. Since biomass is driven by increases in abundance as well as changes in size structure, then by including a random effect, the importance of size structure in driving bias can be better understood.

The partial effect of $ln(b)$ in the full model fitted to location ($\mu_{\text{Bias}[\widehat{\theta}]} \sim f(\ln(b)) + D$) is somewhat different compared to the effect identified by the nested model, for which a random effect for μ is not included. On the first hand, bias in the estimation of biomass itself is no longer asymptotically zero with increasing population biomass for $V \leq 3$ mL, but always negative. Once the population size structure is accounted for, community size is no longer identifiably a driver of (diminishing) bias, and so bias − while reducing until 80 μg C L^{-1} − is decreasing again thereafter. In the estimation of mean log-ESD it is approximately constant, while in the estimation of log-ESD variance is monotonically decreasing in magnitude of bias, but not to a sufficient degree that the bias is asymptotic over the observed domain.

Since D is just a random intercept, a least-squares based expectation would be that shape would be the same. However, the penalisation of the smooth means that if the random effect explains a lot the variance, the penalty on the smooth will begin to outweigh the effect of the smooth. The change in slope does not necessarily indicate that the slope is different to that originally plotted (or does not exist at all), but rather comes from an optimization standpoint.

Referring back to Table B.3, we see that R^2_{CS} increases, but the effect of biomass has actually changed. The nested model is therefore able to account for a good degree of randomness but

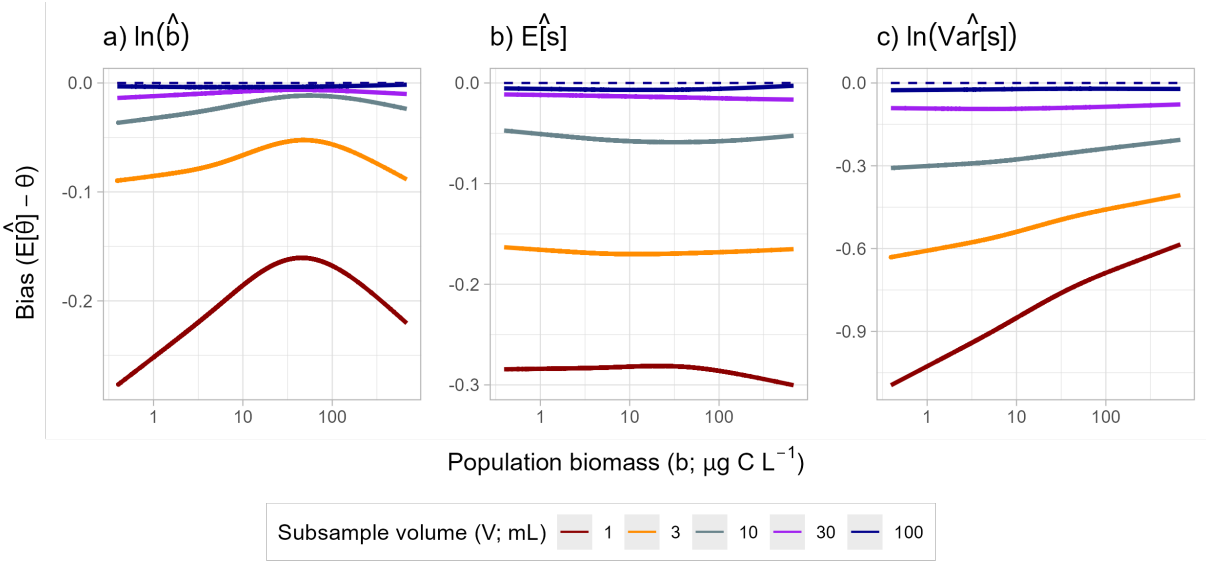


Figure D.1: Partial effect of population biomass on estimation bias of log-biomass, mean log-ESD, and log-transformed log-ESD variance, at different subsample volumes. Here, the model is fitted with random intercepts for both location and scale (rather than just scale), which gives rise to the difference to Figure 3.7.

much of that appears to be a result of intrinsic size structure information. Once variation between size structures are accounted for, the slopes of biomass become less steep.

This indicates the complex relationship between bias and size structure. If a fixed population structure had been assumed, we would not be able to identify this result. However, decomposing in this manner requires careful consideration, since structured relationships between the main effect explanatory smooth and a random effect can make model fitting unreliable. This is particularly relevant when fitting in `gamlss` due to the shortcomings in computing standard errors. Hence, due to concerns about multicollinearity, equivalent mixed models are not presented with E[s] and Var[s] as explanatory variables. It would not be appropriate to attempt to decompose the random effects of the size structure while accounting for the mean size or size variance, since these properties are directly computed to summarise the size structure themselves.

The model is somewhat limited in use in any case, since it does not decompose the actual source of the bias, but rather vaguely assign it to being associated with size structure. If future work aims to establish the true driver of bias at a community level, beyond the fundamental statement of 'large-cell undersampling', a fixed effect approach would be preferable. To do so, the size structure would need to be assigned clusters by characteristics (e.g., bimodal, right-skewed, etc.) to gain a deeper understanding.

D.2 Comparison with Chapter 2 results

The predicted seasonal cycle using a 1 mL subsample was intended to be compared to the original observations at PL, as well as the analysis of communities at SH & LE using a 1 mL subsample

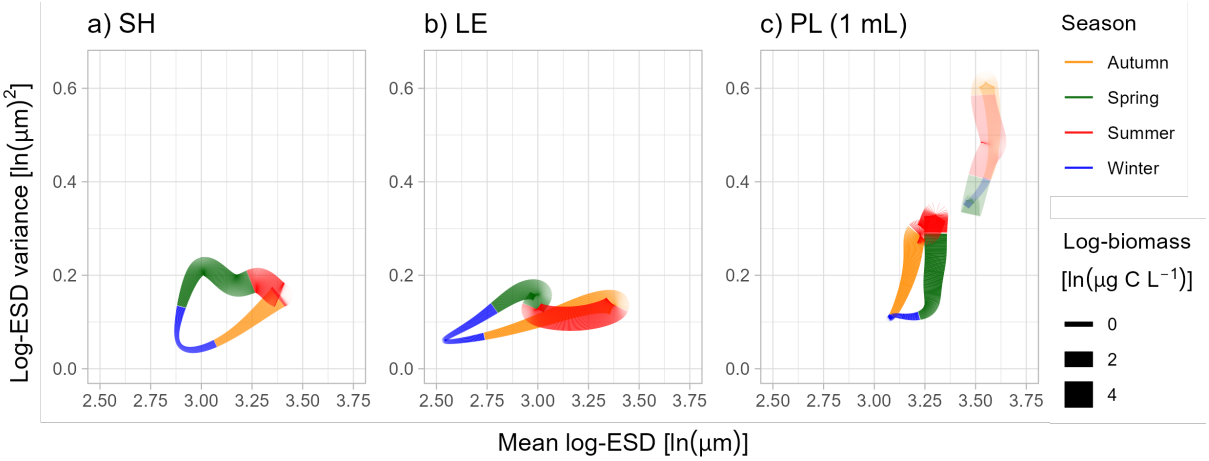


Figure D.2: An adaptation of Figure 2.16. Phase portrait showing the typical seasonal cycle of mean size and size variance of 'microplankton' (diatoms, dinoflagellates, and ciliates) measured by microscopy at SH, LE, and PL. At PL, 1 mL simulated subsample is depicted by solid line and original data by transparent line, demonstrating the potential shift in estimation driven by bias arising from low-volume subsamples.

(Figure D.2). Such a model is of the from $Y = \hat{\theta}$, as opposed to $Y = \text{Bias}[\hat{\theta}]$. In any case, the results would need to be interpreted with caution because the observations at PL, from which the bootstrapping experiment was conducted, cannot be validated as unbiased estimates of the true parameters. This context differs to the results presented in Section 3, where no assumption is made that the observations at PL are fully accurate.

Nonetheless, findings are reported as an interesting thought experiment which may inspire future effort. Quantitatively, the phase portrait at PL shifts to a region of lower mean size and size variance at all points in the year. $(\widehat{\text{E}[s]}, \widehat{\text{Var}[s]})$ coordinates on 1st December shift $(3.53, 0.41) \rightarrow (3.08, 0.11)$. Similarly, on 1st June, a change of $(3.51, 0.41) \rightarrow (3.30, 0.29)$ is simulated. As such there is a lack of quantitative commutativity in the results which must be noted. In Figure 3.9, the summertime bias of mean size is approximately zero. Here, on the other hand, it can be estimates as the difference between the fitted models for $\text{E}[\hat{\theta}]$. Qualitatively, the result that bias is greater in winter is substantially larger than in summer continues to hold. There are several potential reasons why the quantitative discrepancy may arise:

- Fitted values for $\text{Bias} = \text{E}[\hat{\theta}] - \theta$ are not necessarily equal to the difference of models fitted to $\text{E}[\hat{\theta}]$ and θ, except for a linear model fitted under the ordinary least squares framework. Here, penalized maximum likelihood estimates are used in an additive model context, so the identity does not hold.
- The models fitted to $\text{Bias}[\hat{\theta}]$ and $\text{E}[\hat{\theta}]$ are GAMLSSs as opposed to θ in GAM. There are four main differences to consider:
  - Scale is explicitly modelled in the GAMLSS models, but not the GAM. Although the GAM is checked to ensure homoscedasticity, it is likely that some variability in the location of the response is contained in the scale model when a GAMLSS is fitted.

This particularly links to the point below:

– Intercept random effects are included for the scale model, so variance is allowed to vary randomly from one community (i.e., value of θ) to the next. This may further drive improved fit between response and explanatory in models for bias, due to relationships between random effect and information contained in the response for Bias models.

– Different software is used for model fitting (`mgcv` vs. `gamlss`), with different likelihood maximisation techniques the arise due to both statistical requirements and software.

– The coordinates used for the GAM models are original coordinates of time with cyclic cubic splines; the coordinates for the GAMLSS are required to be pre-transformed onto circular coordinates and the a 'standard' cubic spline is fitted. No tensor product smooth (interaction) is allowed between the circular coordinates in `gamlss`.

D.3 Original model fitted with GAM

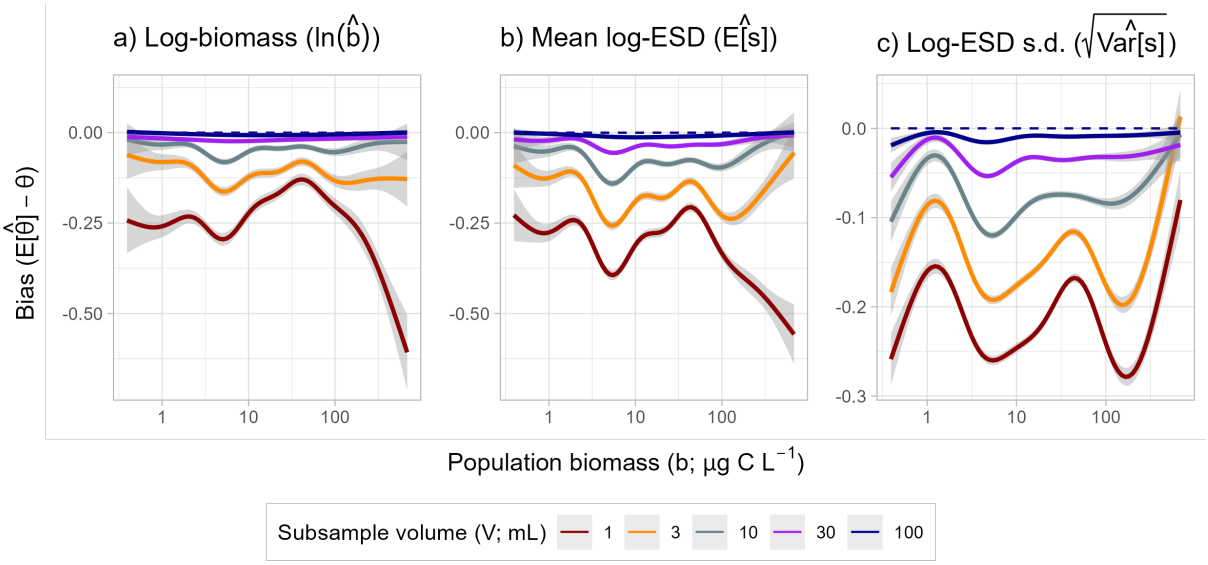


Figure D.3: The equivalent of Figure 3.7 when a GAM is fitted rather than a GAMLSS model. The extent of model fitting bias is demonstrated at a seasonal scale in Figure 3.2.

Appendix E

Supplementary figures for the size-structured plankton ecosystem model

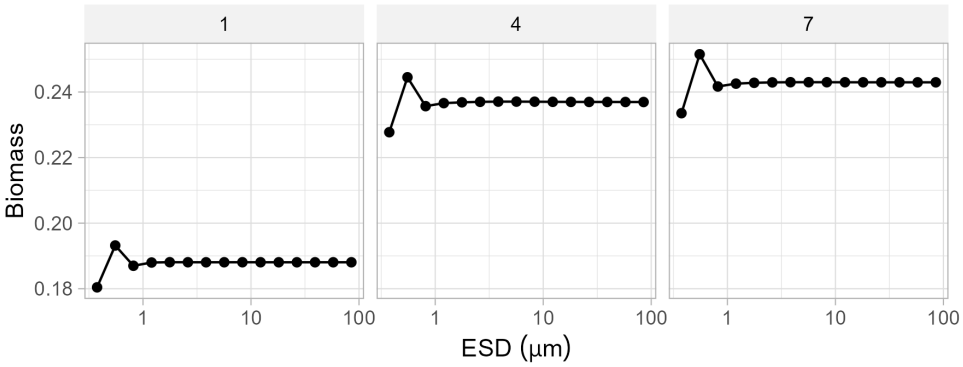


Figure E.1: Outputs from the neutral model for months 1, 4, and 7. All allometric exponents are set to zero, and all coefficients are equal across all size classes (e.g. $k_{Z1} = k_{Z2}$). To facilitate a neutral model, zooplankton grazing was herbivorous-only (since carnivorous grazing is only on some size classes), and all zooplankton size classes were subject to quadratic losses. This therefore replaced grazing losses for protozooplankton.

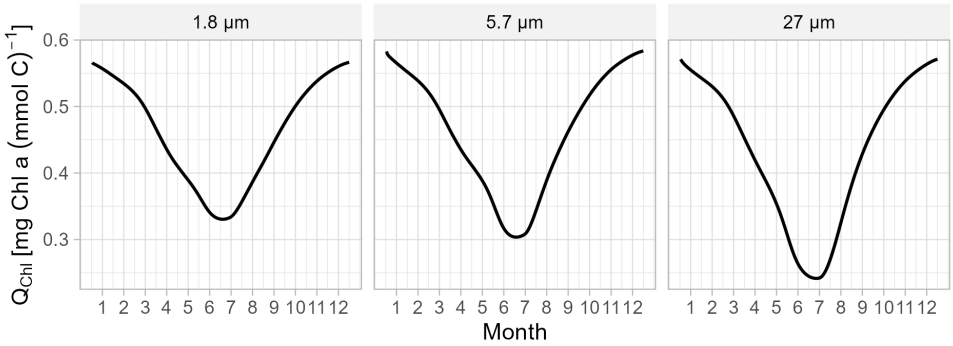


Figure E.2: Seasonal cycle of Q_{Chl} for three dominant phytoplankton size classes of the model. These values can be transformed from their units of mg Chl a (mmol C)$^{-1}$ to mg C (mg Chl a)$^{-1}$ by computing $12/Q_{Chl}$. The equivalent values here range from 20 mg C (mg Chl a)$^{-1}$ to 52 mg C (mg Chl a)$^{-1}$.

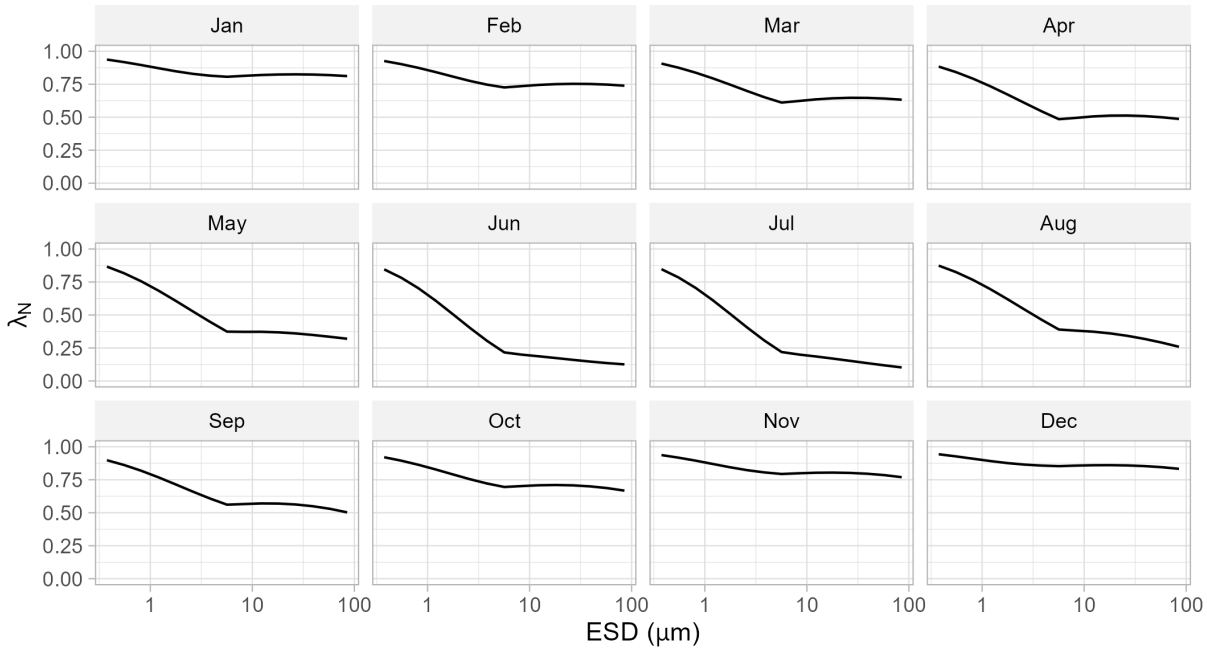


Figure E.3: Monthly-averaged nutrient limitation of photosynthesis (λ_N) by phytoplankton size.

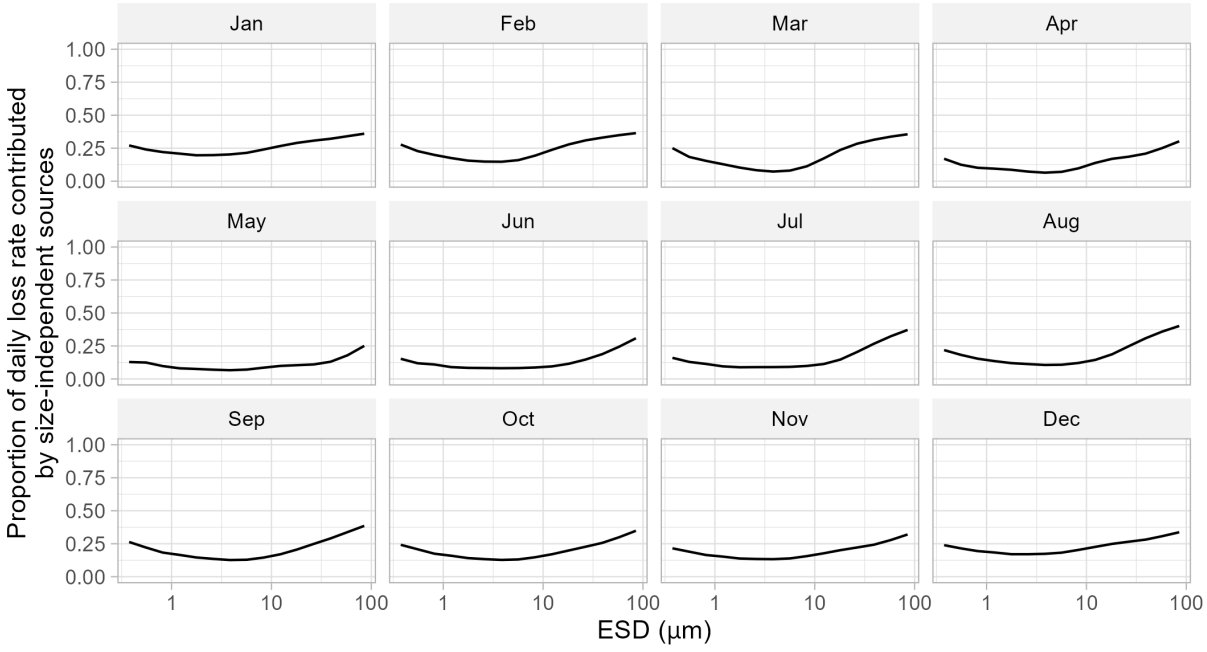


Figure E.4: Monthly-averaged proportion of gross losses contributed by size-independent sources across phytoplankton sizes. Size-independent losses arise from linear mortality and mixing, while size-dependent losses are from grazing.

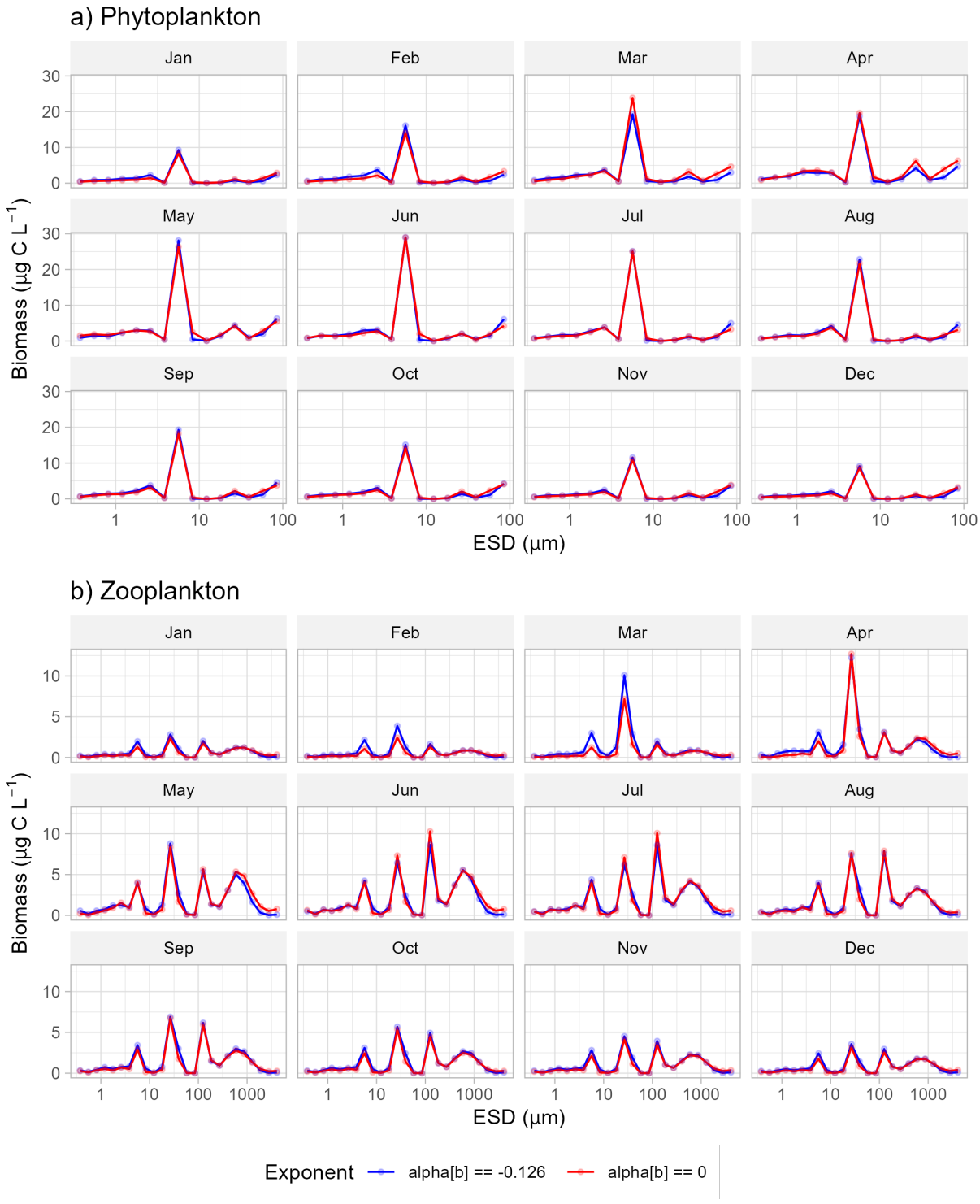


Figure E.5: Monthly-averaged biomass distributions of phytoplankton and zooplankton of the model using the parameter set which minimises $\mathcal{D}$ (and thus has $\alpha_b = -0.136$), and an experiment where $\alpha_b = 0$ with $\alpha_a = 0.45 \times 10^{-7}$.

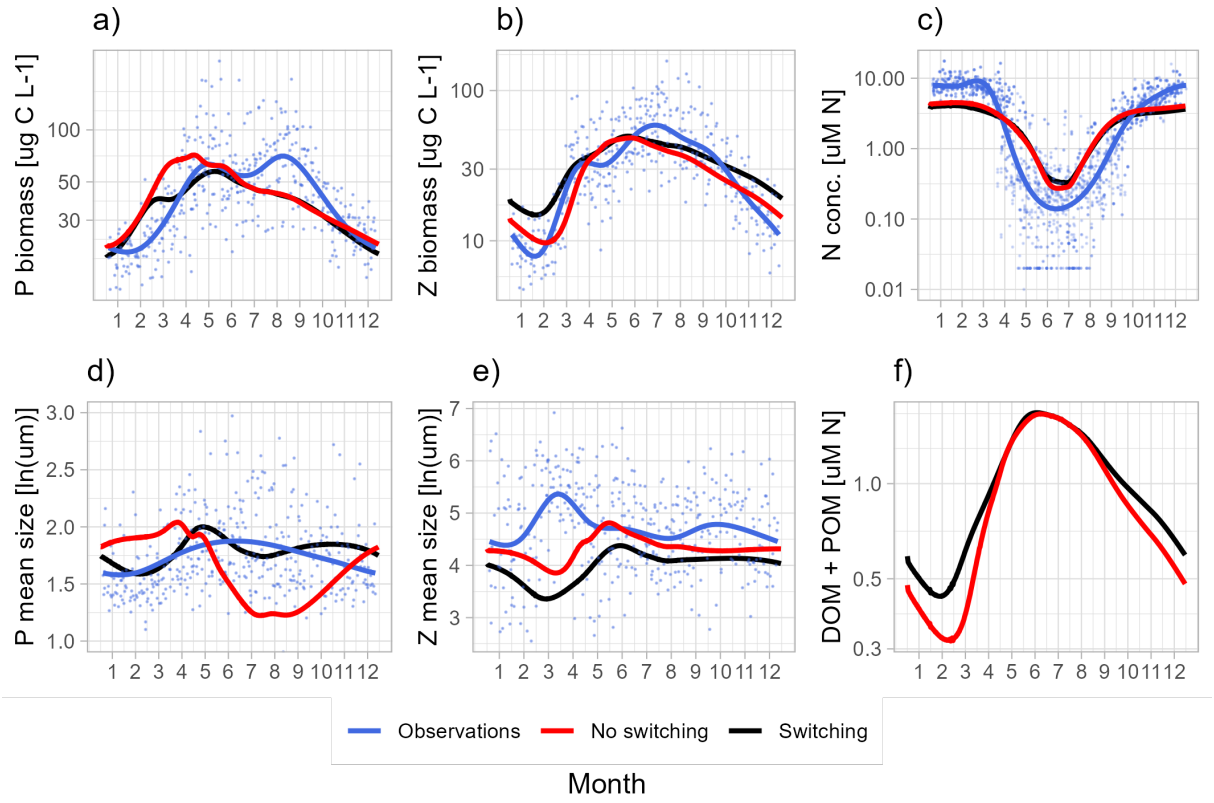


Figure E.6: Model outputs for the optimal parameter set with passive switching for herbivory and carnivory (black line), and an experiment with no switching (red line). Observations are presented as blue points, with GAM fitted using cyclic cubic regression splines given as the blue line.

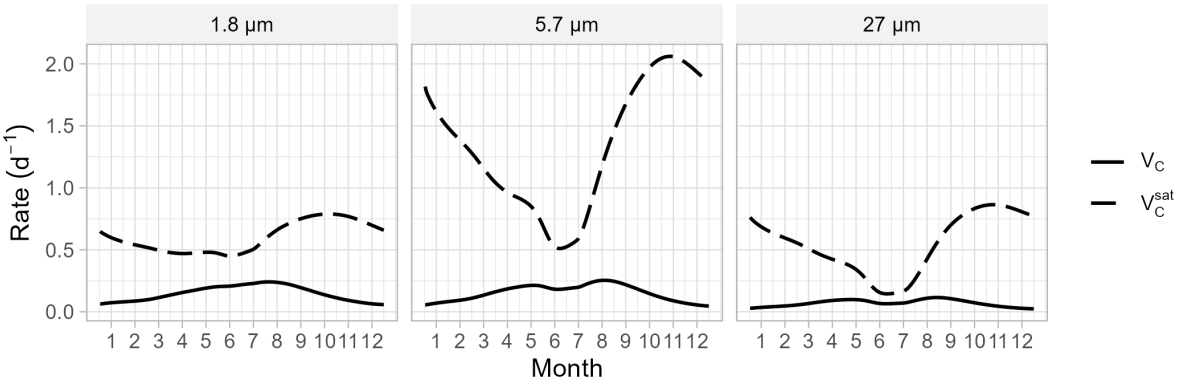


Figure E.7: Seasonal cycle of realised photosynthetic rate (V_C) and saturating photosynthetic rate (V_C^{sat}) of three dominant size classes of the model.