

microARC modelling methods

Aidan Hunter, Ben Ward

February 28, 2022

1 Introduction

- Plankton, biogeochemical cycles, climate and society
- Productivity and food webs
- Community structure
 - oligotrophic - small
 - eutrophic - small and larger
 - blooms
- High latitudes
 - Smallest phytoplankton absent from polar waters
 - * How does this impact community structure and ecosystem function?
 - Apply a “food web model for the global ocean” to the Arctic ecosystem
 - * Optimised to observations for the Fram Strait
 - Atlantic waters
 - Arctic waters
 - * Does it work?
 - * What does it reveal?

2 Material and Methods

2.1 Observational data

Can you give a friendlier introduction to the Hausgaard LTER?
Less detail and more about what the time-series is about.

2.1.1 Biogeochemical

Nutrient, organic matter and plankton size spectra were measured from water samples collected by CTD casts from onboard the RV Polarstern research vessel at the LTER (Long Term Ecological Research) HAUSGARTEN sites in Fram Strait during three separate cruises (PS99, PS107 and PS114) in 2016–2018 (table 1). Only the (PS114) 2018 data were used because the full complement of required measurements (nutrient, organic matter, size spectra, and forcing data) was unavailable for earlier years. Our model simulates total bioavailable DIN (dissolved inorganic nitrogen) without distinguishing between compounds. We therefore extracted the $\text{NO}_2 + \text{NO}_3$ (nitrite plus nitrate) concentration column from the nutrient data table ([Torres-Valdés et al., 2019](#)), as well as the sample event (time and place) and depth covariates. Measured concentrations of PON and POC (particulate organic carbon and nitrogen) and chlorophyll *a* were extracted from the organic matter data table ([von Jackowski et al., 2020](#)). Depth and sample event covariates were also extracted; these covariate values differed from those associated with DIN measurements as different data types were not all gathered at the same sample events or depths (fig. 4). The DIN, PON/POC and chlorophyll *a* data follow structurally identical sample designs: each sample event contains a set of concentration measurements from a range of discrete depths, with a single measurement per depth. These measurements may, therefore, be similarly transformed to produce normalised data sets; and we refer to them together as the “scalar data”. These data were written compactly as $\mathbf{Y}_v^{\text{obs}}$, representing all concentration measurements of variable $v \in \{\text{DIN}, \text{PON}, \text{POC}, \text{Chl } a\}$ in bold vector notation. Individual measurements were written as $Y_{v,d,s}^{\text{obs}}$ where s and d index sample event and depth. A fourth index, $g \in \{\text{Atlantic, Arctic}\}$, grouped the data by origin of the water samples, as determined using particle trajectories from a physical model (section 2.2). All of the scalar data used to inform our model are displayed in fig. 1.

2.1.2 Ecological

Planktonic size spectra were derived through meticulous microscopy and kernel density estimation methods ([Lampe et al., 2021](#)). CTD water samples corresponding to a small range of depths around the chlorophyll maximum, within a subset of the sampling events, were used for the analyses. From 50 mL aliquots of each water sample, plankton cells were measured and counted under varying degrees of magnification, identified taxonomically, then classed as autotrophic phytoplankton or heterotrophic zooplankton. Kernel density methods were then applied to these count-data to generate smooth size spectra. See [Lampe et al. \(2021\)](#) for a full description of the size spectra derivation.

The size spectra data tables contain measured cell concentration densities with respect to ESD (equivalent spherical diameter). Each size spectra is represented by two vectors: the independent variable, \mathbf{D} (μm), a sequence of cell diameters with successive elements separated by 0.005 on the \log_{10} scale; and the corresponding dependent variable of cell concentration density, $\mathbf{S}_{f,d,s}^c$ ($\text{cells m}^{-3} \log_{10}(\text{ESD}/1\mu\text{m})^{-1}$), measured separately for plankton cell types, $f \in \{\text{autotroph, heterotroph}\}$, at depths, d , and sample events, s . As these data are highly resolved across ESD they are effectively continuous, so we denote cell concentration density at size class k using function notation, $S_{k,f,d,s}^c \equiv S_{f,d,s}^c(D_k)$. As measured concentration densities were negligible for $\text{ESD} > 200\mu\text{m}$ and were relatively uncertain for $\text{ESD} < 1\mu\text{m}$ we truncated all size spectra to the interval $\text{ESD} \in [1, 200] \mu\text{m}$. Biovolume concentration density measurements, $\mathbf{S}_{f,d,s}^b$ ($\mu\text{m}^3 \text{m}^{-3} \log_{10}(\text{ESD}/1\mu\text{m})^{-1}$), were estimated from the spherical volume equation.

$$S_{f,d,s}^b(\mathbf{D}) = \frac{1}{6} \pi \mathbf{D}^3 S_{f,d,s}^c(\mathbf{D}) \quad (1)$$

These biovolume concentration density data were selected to inform the ecosystem model because, compared to the cell concentration densities, they exhibited less dispersion across ESD [quantify, orders of magnitude]. As in section 2.1.1, the data were grouped into two categories, $g \in \{\text{Arctic, Atlantic}\}$, to distinguish samples taken from water originating north or south of Fram Strait. Single size spectra were derived for each cell type, f , and water origin, g , by calculating geometric mean biovolume concentration densities across sample events and depths

$$\mathbf{S}_{f,g}^b = \left(\prod_{s=1}^{n_s} \left(\prod_{d=1}^{n_d(s)} \mathbf{S}_{f,g,s,d}^b \right)^{\frac{1}{n_d(s)}} \right)^{\frac{1}{n_s}} \quad (2)$$

where n_s is the number of sample events and $n_d(s)$ is the number of depths sampled during event s . This produced four separate size spectra, $\mathbf{S}_{f,g}^b$, of biovolume concentration densities averaged over all water samples of Atlantic or Arctic origin, for phytoplankton and zooplankton (fig. 2).

The size spectra data (fig. 2) were not directly comparable with our ecosystem model outputs as these have different units and are more coarsely resolved into n modelled size classes. To compare observations with model outputs we derived *binned* size data, $S_{j,f,g}^{\text{obs}}$, corresponding to $j = 1, \dots, n$ contiguous modelled size classes. Modelled size classes were defined by vectors, \mathbf{b} , specifying ESD at size class edges. The binned size data were derived by integrating $\mathbf{S}_{f,g}^b$ with respect to $\log_{10}(\text{ESD})$ within each of the n modelled size class intervals.

$$S_{j,f,g}^{\text{obs}} = \int_{\log_{10}(b_j)}^{\log_{10}(b_{j+1})} S_{f,g}^b(x) dx \quad (3)$$

These data, $\mathbf{S}_{f,g}^{\text{obs}}$, are n -element vectors of measured biovolume concentration ($\mu\text{m}^3 \text{ m}^{-3}$) within each modelled size class interval (fig. 3). [describe how size classes are chosen, or leave for later section?]

2.1.3 Data standardisation

The raw scalar data, $\mathbf{Y}_v^{\text{obs}}$, were standardised to reduce bias when numerically estimating model parameters. If the raw data were used directly to inform parameter estimation then results would be biased by variability in magnitudes of the measurement types $v \in \{\text{DIN}, \text{PON}, \text{POC}, \text{Chl } a\}$, and biased by variability associated with sampling event and depth. In effect, this would unevenly weight the data points. To eliminate these biases the measurements of each type, v , were independently standardised with respect to depth and sampling event using LMMs (linear mixed models). This section describes the LMMs used to transform the raw measurements, $\mathbf{Y}_v^{\text{obs}}$, into approximately standard-normal data we denote as $\widehat{\mathbf{Y}}_v^{\text{obs}}$.

LMMs are statistical descriptions of data that are linearly related to one or more covariates and that are stochastically influenced by other covariates. From fig. 1 it is clear that POM and Chl a decline with depth, whereas DIN increases with depth. There is also a change in the relationship between DIN and depth at approximately 100 m. Due to their distinct relationships to depth, the DIN data were standardised using LMMs of different design to those used for the POM and Chl a data.

Logarithmic transforms of the POM and Chl a data produced approximately linear relationships with depth, z . Thus, the LMMs used to standardise these $v \in \{\text{PON}, \text{POC}, \text{Chl } a\}$ data were

$$\begin{aligned} \ln(Y_{v,d,s}^{\text{obs}}) &= (a_v + a_{v,s}) + (b_v + b_{v,s}) z_{v,d,s} + \varepsilon_{v,d,s} \\ &= \mu_{v,d,s} + \varepsilon_{v,d,s} \end{aligned} \quad (4)$$

where a_v and b_v are the “fixed” effects of depth upon measured values; $a_{v,s}$ and $b_{v,s}$ are the “random” effects associated with sampling event; $\varepsilon_{v,d,s} \sim \mathcal{N}(0, \sigma_v)$ are normally distributed residual errors; and σ_v are the residual error standard deviations. Scaling the LMM residual errors by their standard deviations

$$\widehat{\mathbf{Y}}_v^{\text{obs}} = \frac{1}{\sigma_v} \left(\ln(\mathbf{Y}_v^{\text{obs}}) - \boldsymbol{\mu}_v \right) \quad (5)$$

produced approximately standard-normal data sets, $\widehat{\mathbf{Y}}_v^{\text{obs}} \sim \mathcal{N}(0, 1)$. As the standardised data have no relation to the depth or sample event covariates, the d and s subscripts are redundant and omitted.

A similar process was used to standardise the DIN data. These were inherently non-linear due to the sharp change in the DIN-depth relationship at 100 m. The DIN data were therefore split into two groups: measurements from depths $z \leq 100$ m, and from depths $z > 100$ m. LMMs were

then separately fitted to data from each group in an identical procedure, as follows. A logarithmic transform of the depth covariate approximately linearised the DIN data with respect to depth. The LMMs used to standardise the $v = \text{DIN}$ data were

$$\begin{aligned} Y_{v,d,s}^{\text{obs}} &= (a_v + a_{v,s}) + (b_v + b_{v,s}) \ln(z_{v,d,s}) + \varepsilon_{v,d,s} \\ &= \mu_{v,d,s} + \varepsilon_{v,d,s} \end{aligned} \quad (6)$$

with all variables defined as above. Then the standardised DIN data were generated as

$$\hat{\mathbf{Y}}_v^{\text{obs}} = \frac{1}{\sigma_v} (\mathbf{Y}_v^{\text{obs}} - \boldsymbol{\mu}_v) \quad (7)$$

where, once again, $\hat{\mathbf{Y}}_v^{\text{obs}} \sim \mathcal{N}(0, 1)$.

Each standardised data set was distributed as approximately standard normal and independent of measured covariates (fig. 1). The model parameter fitting process of minimising discrepancies between standardised data and the equivalent model outputs was therefore not biased by scalar data type, sampling event or depth — provided the equivalent model outputs were identically transformed, $\mathbf{Y}_v^{\text{mod}} \rightarrow \hat{\mathbf{Y}}_v^{\text{mod}}$, by eqs. (5) and (7).

2.2 Model-simulated forcing data

Physical oceanographic model outputs were used to drive our NPZD model. These forcing data comprised time series of latitude-longitude coordinates for “particles” moving with ocean currents simulated by the physical model. Each particle’s trajectory was associated with time series of depth-discrete water temperature, $T(z, t)$, diffusivity, $K(z, t)$, and surface irradiance, $I^{\text{surf}}(t)$. By using these time series to force the NPZD model, plankton dynamics were simulated within 1D water columns advected by currents through temporally varying environments. In other words: plankton dynamics were simulated from the “perspective” of plankton communities using a Lagrangian model reference frame. The state of plankton communities during any time of the simulation could therefore be examined within the context of that community’s history of experienced environmental conditions. This model framework differs fundamentally from ubiquitous Eulerian model reference frames that simulate dynamics within fixed spatial regions that plankton are swept through by currents. Pros and cons related to choosing an Eulerian or Lagrangian model framework will be considered in greater detail during the discussion (section 4).

Forcing data were extracted from the SINMOD physical model ([Slagstad & McClanahan, 2005](#), [Wassmann et al., 2006](#)) using the MatLab *particulator* tool ([Banas, 2018](#)) to generate particle trajectories. SINMOD outputs were spatially resolved to [...] and available at a daily resolution. These outputs contained [...], from which we retained [...] to drive our model. The

particulator tool converts 2D or 3D spatio-temporal physical oceanographic data into 1D data tracing lines over the domain surface, where each line represents the trajectory of a theoretical particle released at some specific space-time coordinates. Particles were initialised in a grid aligned with the spatial cells representing the entire SINMOD domain, with three particles between cell centres in both the longitudinal and latitudinal directions. For each year of forcing data the particles were released at the start of the year, on 1st January ($t = 1$). Particle locations during subsequent days ($t = 2, \dots$) were derived from the SINMOD flow fields using the *particulator* tool. The latitude-longitude coordinates of each particle corresponded to the location at the start of each day, while the associated data were daily averages [check this with Fabian's notes]. The particles were then filtered by retaining only those whose trajectories intersected the HAUSGARTEN sample region west of Svalbard within a week[?] of sampling time (fig. 4). This intersection region was constructed by surrounding each in-situ sample location with a 25 km diameter (or radius?) circle, then connecting the outermost circle edges with a convex polygon representing the study region sample area. This filtering process yielded 3274 unique particle trajectories intersecting the sampling region in 2018 (fig. 4).

The set of trajectories intersecting the sampling region were grouped into those originating from the Arctic and Atlantic. It was important to distinguish plankton communities inhabiting Arctic or Atlantic water masses because each were products of their different histories of experienced environments. Although this grouping may have been determined by inspection, we used numerical methods easily transferable to other systems: a hierarchical clustering algorithm that grouped trajectories according to a dissimilarity matrix. Trajectory dissimilarities were quantified by applying dynamic time warping to the time series of trajectory latitude-longitude coordinates using MatLab function *dtw* (Sakoe & Chiba, 1978, Paliwal et al., 1982). This method quantified time series dissimilarity based on shape, so time series that were out of phase could nonetheless be regarded as similar. An agglomerative hierarchical clustering algorithm (MatLab function *linkage*) was then applied to the dissimilarity matrix to group trajectories. The two main clusters corresponded to particle trajectories representing Arctic and Atlantic water masses. Of the full set of 3274 trajectories, 1291 originated from the Arctic and 1983 from the Atlantic.

After determining the particle origins, each in-situ sample was categorised as Arctic and/or Atlantic based on the origin of the water masses from which each sample was taken. Each sample was associated with a set of ten particle trajectories whose origins determined to which category the samples were assigned. We centred 25 km radii “catchment areas” at the location of each sampling event. Particles laying outside these catchment areas at the sampling date were omitted for that event. All remaining particles were considered sufficiently close to the sample to be considered for inclu-

sion. These were filtered using methods similar to those previously described for grouping particles by origin. To maximise variability we selected the ten most dissimilar particle trajectories within each catchment area. Once again, dissimilarities were determined by applying dynamic time warping to the time series of trajectory latitude-longitude coordinates. The trajectories were then grouped by agglomerative hierarchical clustering applied to the dissimilarity matrix, and a single trajectory was selected from the ten main clusters. Once associated with ten trajectories, each sample was categorised as from Arctic or Atlantic waters if all ten trajectories originated from either region, and was otherwise categorised as a boundary case.

2.3 Model

2.3.1 Flux equation system

Nutrient fluxes are described by a differential equation system (eqs. (8) to (10)) based on Ward et al. (2012) and Ward & Follows (2016). The state variables are concentrations of inorganic nitrogen, $N(z, t)$, autotrophic and heterotrophic plankton, $P_{i,j}(z, t)$ and $Z_{i,j}(z, t)$, and organic matter, $D_{i,k}(z, t)$. Concentrations have units of mmol element m^{-3} or mg chlorophyll $a \text{ m}^{-3}$. Subscripts i , j , and k respectively index nutrients, plankton cell sizes, and type of organic matter. Nutrient indexes are: $i \in \{\text{C}, \text{N}, \text{Chl}\}$ for autotrophs; and $i \in \{\text{C}, \text{N}\}$ for heterotrophs and organic matter. There are n_p autotroph and n_z heterotroph size classes. Total plankton concentration is compactly represented by combining both trophic levels in matrix $B_{i,j} \equiv [P_{i,j}, Z_{i,j}]$, where $j \in \{1, \dots, n_p + n_z\}$. Autotrophs or heterotrophs may be filtered from $B_{i,j}$ by indexes $j_p \in \{1, \dots, n_p\}$ and $j_z \in \{n_p + 1, \dots, n_p + n_z\}$. Dissolved and particulate organic matter are indexed by $k \in \{\text{DOM}, \text{POM}\}$.

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial N}{\partial z} - \sum_j V_{\text{N},j} B_{\text{C},j} + \sum_k r_{\text{N},k} D_{\text{N},k} \quad (8)$$

$$\begin{aligned} \frac{\partial B_{i,j}}{\partial t} = & \frac{\partial}{\partial z} K \frac{\partial B_{i,j}}{\partial z} + V_{i,j} B_{\text{C},j} - \sum_{j_z} G_{i,j_z,j} B_{\text{C},j_z} \\ & + \sum_j \lambda_{i,j_z} G_{i,j_z,j} B_{\text{C},j_z} - m_j B_{i,j} \end{aligned} \quad (9)$$

$$\frac{\partial D_{i,k}}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial D_{i,k}}{\partial z} - w_k \frac{\partial D_{i,k}}{\partial z} - r_{i,k} D_{i,k} + S_{i,k}^D \quad (10)$$

All terms in eqs. (8) to (10) are defined in table 2, and model parameters are defined in tables 3 and 4.

2.3.2 Cell quotas & limiting terms

Plankton biomass is tracked as the carbon concentration, $B_{C,j}$. Concentrations of other nutrients vary, relative to carbon, within limits preventing excessive accumulation or depletion of any one nutrient. Cellular quotas of nitrogen and chlorophyll are defined as ratios with carbon.

$$Q_{i,j} = \frac{B_{i,j}}{B_{C,j}} \quad (11)$$

Nitrogen quotas vary between size-dependent limiting values, $Q_{N,j}^{\min}$ and $Q_{N,j}^{\max}$. Positive correlation between the $Q_{N,j}^{\min}$ and $Q_{N,j}^{\max}$ parameters hindered numerical optimisation. Correlation was reduced by reparameterising as $\tilde{Q}_{N,j}^{\max} = Q_{N,j}^{\max} / (Q_{N,j}^{\max} - Q_{N,j}^{\min})$ and estimating the parameters of $Q_{N,j}^{\min}$ and $\tilde{Q}_{N,j}^{\max}$. Inverting gives the maximum nitrogen quotas.

$$Q_{N,j}^{\max} = \frac{Q_{N,j}^{\min}}{1 - 1/\tilde{Q}_{N,j}^{\max}} \quad (12)$$

Carbon production (eq. (17)) is down-regulated by the nutrient limitation term, which is a linear function of $Q_{N,j}$

$$\gamma_{N,j} = \frac{Q_{N,j} - Q_{N,j}^{\min}}{Q_{N,j}^{\max} - Q_{N,j}^{\min}} \quad (13)$$

that halts production when quotas are minimised and does not limit production when quotas are maximised.

Nutrient uptake rates (eq. (16)) are down-regulated as cell quotas increase, and are zeroed when quotas are full. The nutrient uptake regulation terms

$$Q_{N,j}^{\text{stat}} = 1 - \left(\frac{Q_{N,j} - Q_{N,j}^{\min}}{Q_{N,j}^{\max} - Q_{N,j}^{\min}} \right)^{1/h} \quad (14)$$

are modified from [Ward & Follows \(2016\)](#) by reflecting the $Q_{N,j}^{\text{stat}}(\gamma_{N,j})$ curves across the line, $Q_{N,j}^{\text{stat}} = \gamma_{N,j}$. This modification was made solely for numerical stability, and the resulting uptake regulation curve retains a similar shape to those used by [Ward & Follows \(2016\)](#).

Temperature, T , influences nutrient uptake (eq. (16)), photosynthesis (eq. (17)), and grazing rates (eq. (22)), which are adjusted by a temperature regulation term

$$\gamma_T = e^{A(T - T^{\text{ref}})} \quad (15)$$

where T^{ref} is a fixed reference temperature and A is temperature sensitivity.

2.3.3 Nutrient uptake

Nutrient uptake rate is modelled by Michaelis-Menten functions regulated by nutrient quotas and temperature.

$$V_{N,j} = \frac{v_j^{\max} \alpha_j N}{\alpha_j N + v_j^{\max}} Q_{N,j}^{\text{stat}} \gamma_T \quad (16)$$

The uptake regulation term, $Q_{N,j}^{\text{stat}}$, reduces nutrient uptake rate as quotas approach the maximum. Maximum uptake rate, $v_{j_p}^{\max}(\text{Vol}_{j_p})$, and nutrient affinity, $\alpha_{j_p}(\text{Vol}_{j_p})$, are functions of cell volume. All uptake rates in the heterotroph size classes are zeroed, $V_{i,j_z} = 0$.

2.3.4 Photosynthesis

Irradiance at depth, $I(z, t)$, was calculated from surface irradiance, $I^{\text{surf}}(t)$, and total chlorophyll concentration, $\sum_j B_{\text{Chl},j}(z, t)$, using the Beer-Lambert formula with attenuation coefficients of 0.04 for seawater and 0.0149 for chlorophyll ([Krause-Jensen & Sand-Jensen, 1998](#)) [reference for seawater attenuation coef].

Carbon-specific light-saturated photosynthetic rate is modelled as a size-dependent maximum rate restricted by temperature and nitrogen quota limitation terms.

$$P_j^{\text{sat}} = P_j^{\max} \gamma_T \gamma_{N,j} \quad (17)$$

Photosynthetic rate is defined as a Poisson function of irradiance, I , and chlorophyll quota.

$$P_j = P_j^{\text{sat}} \left(1 - \exp \left(\frac{-\alpha_p Q_{\text{Chl},j} I}{P_j^{\text{sat}}} \right) \right) \quad (18)$$

Carbon-specific production rate is then defined as the difference between photosynthetic rate and the cost of biosynthesis

$$V_{C,j} = P_j - \xi V_{N,j} \quad (19)$$

where cost of biosynthesis is a linear function of nitrogen uptake.

Chlorophyll production is coupled to nitrogen uptake

$$V_{\text{Chl},j} = \rho_j V_{N,j} \quad (20)$$

and is downregulated from maximum, θ , as the photosynthetic rate, P_j , decreases below the theoretical maximum-efficiency rate, $\alpha_p Q_{\text{Chl},j} I$, at high irradiances.

$$\rho_j = \theta \frac{P_j}{\alpha_p Q_{\text{Chl},j} I} \quad (21)$$

2.3.5 Predation

Grazing rates of prey carbon are the product of predator's maximum grazing rates and prey saturation, switching, and refuge terms.

$$G_{C,j_z,j} = \underbrace{\gamma_T G_{j_z}^{\max}}_{\text{max. rate}} \underbrace{\frac{F_{C,j_z}}{k_G + F_{C,j_z}}}_{\text{saturation}} \underbrace{\Phi_{j_z,j}}_{\text{switching}} \underbrace{(1 - e^{\Lambda F_{C,j_z}})}_{\text{refuge}} \quad (22)$$

Prey saturation is modelled with Michaelis-Menten functions of total prey carbon available to each predator

$$F_{C,j_z} = \sum_j \phi_{j_z,j} B_{C,j} \quad (23)$$

where $\phi_{j_z,j}$ is the availability of each prey class, j , to each predator class, j_z . Prey availability is modelled as a function of predator-to-prey diameter ratios, $\delta_{j_z,j}$

$$\phi_{j_z,j} = \exp \left[- \left(\ln \left(\frac{\delta_{j_z,j}}{\delta_{\text{opt}}} \right) \right)^2 / (2\sigma^2) \right] \quad (24)$$

where δ_{opt} and σ are the optimum ratio maximising prey availability, and variability around the optimum. As this function is a log-normal probability density without the scaling terms, it has maximum value of 1 and a log-normal shape.

The prey switching term regulates predation losses by targeting grazing on the most abundant prey classes.

$$\Phi_{j_z,j} = \frac{(\phi_{j_z,j} B_{C,j})^2}{\sum_j (\phi_{j_z,j} B_{C,j})^2} \quad (25)$$

This prevents overgrazing any single prey class when other available prey are more abundant.

The prey refuge term prevents overgrazing by reducing predator grazing rates when total available prey, F_{C,j_z} , is low.

Carbon-specific grazing rates of nitrogen and chlorophyll are calculated as the product of carbon grazing rate and cell quotas.

$$G_{i,j_z,j} = Q_{i,j} G_{C,j_z,j} \quad (26)$$

Assimilation efficiency of consumed prey is given by

$$\lambda_{C,j_z} = \lambda^{\max} \gamma_{N,j_z} \quad (27)$$

$$\lambda_{N,j_z} = \lambda^{\max} Q_{N,j_z}^{\text{stat}} \quad (28)$$

which down-regulates prey assimilation from the maximum, λ^{\max} , when predator cell quotas approach their limits.

2.3.6 Background mortality

The scalar mortality parameter, $m_j = m$, used by Ward et al. (2012) and Ward & Follows (2016) produced background mortality rates independent of size and linear with plankton concentration. We included options to model mortality rate as a size dependent linear process, or as a size independent quadratic (density dependent) process.

To model background mortality as a size dependent and linear process, the mortality rate parameter was expressed as a modified power function of cell volume.

$$m_{i,j} = m_{\min} + (m_a - m_{\min}) \text{Vol}_j^{m_b} \quad (29)$$

The background mortality rate of large cells is assumed less than or equal to that of smaller cells, $m_j \leq m_{j-1}$, so that $m_a > 0$ and $m_b \leq 0$. A minimum permissible mortality, m_{\min} , prevents $m_j \rightarrow 0$ as volume increases. Setting $m_b = 0$ removes size dependence of mortality and is equivalent to a scalar mortality parameter.

To model background mortality as a density dependent process, the mortality rate parameter was expressed as a linear function of plankton concentration.

$$m_{i,j} = m_a + m_b B_{i,j} \quad (30)$$

The m_a and m_b parameters control the linear and quadratic components of mortality, and have different interpretation from parameters used in eq. (29). As it removes the density dependence, setting $m_b = 0$ is, once again, equivalent to using a scalar mortality parameter.

2.3.7 Organic matter

Organic matter is generated from mortality and messy feeding. Unassimilated nutrient from dead cells is transferred into organic matter and allocated into DOM and POM categories. The proportions, $\beta_{j,k}$, allocated to DOM and POM are modelled as volume-dependent using a three-parameter double-logistic function

$$\beta_{j,\text{DOM}} = \frac{b_1}{1 + e^{(x-b_3)}} + \frac{b_1 b_2}{1 + e^{(b_3-x)}} \quad (31)$$

$$\beta_{j,\text{POM}} = 1 - \beta_{j,\text{DOM}} \implies \sum_k \beta_{j,k} = 1 \quad (32)$$

where $x = \log_{10}(\text{Vol})$, and where $0 < b_1 < 1$ and $0 < b_1 b_2 < 1$ are necessary constraints. By constraining $0 < b_2 < 1$, we ensure that $\beta_{j,\text{DOM}}$ decreases monotonically with cell volume to enforce the assumption that, upon expiration, small cells produce proportionally more DOM than relatively large cells.

Sources of organic matter are the sum of mortality and messy feeding terms

$$S_{i,k}^D = \underbrace{\sum_j \beta_{j,k} m_{i,j} B_{i,j}}_{\text{mortality}} + \underbrace{\sum_{j_z} B_{C,j_z} \sum_j \beta_{j,k} (1 - \lambda_{i,j_z}) G_{i,j_z,j}}_{\text{messy feeding}} \quad (33)$$

where $(1 - \lambda_{i,j_z}) G_{i,j_z,j}$ are the predator carbon-specific rates of organic matter production through messy feeding.

2.3.8 Initial conditions

Integrating eqs. (8) to (10) to generate solutions $\{N(z,t), P(z,t), Z(z,t), D(z,t)\}$ required choosing initial conditions for the state variables. We used SINMOD outputs for January 1st to specify initial conditions for N and, combined with some assumptions, to derive initial conditions for P , Z and D .

The nutrient initial condition was set as nitrate concentration.

$$N(z, t = 1) = \text{NO}_3^{\text{phys}}(z, t = 1) \quad (34)$$

Planktonic nitrogen concentration grouped by “small” and “large” cell sizes, $B_{N,\text{small}}^{\text{phys}}$ and $B_{N,\text{large}}^{\text{phys}}$, was used to initialise $B_{N,j}$. Modelled size classes were mapped to the SINMOD size categories by defining cells with $\text{ESD} \geq 10 \mu\text{m}$ as large and all other cells as small. Initial values of planktonic nitrogen were derived as

$$B_{N,j}(z, t = 1) = \begin{cases} \frac{1}{n_s} B_{N,\text{small}}^{\text{phys}}(z, t = 1), & \text{ESD}_j < 10 \mu\text{m} \\ \frac{1}{n_l} B_{N,\text{large}}^{\text{phys}}(z, t = 1), & \text{ESD}_j \geq 10 \mu\text{m} \end{cases} \quad (35)$$

where n_s and n_l are the number of modelled size classes corresponding to the “small” and “large” SINMOD categories. Equation (35) assigns an identical initial value to all plankton classes, autotrophs and heterotrophs, within each “small” or “large” category by assuming that each modelled size class contributes equally to total planktonic nitrogen. Planktonic carbon and chlorophyll concentrations were then initialised using assumptions about cell quota values. As nitrogen quotas are likely replete during winter we set nitrogen quotas to 75% full then initialised planktonic carbon.

$$\begin{aligned} Q_{N,j}(z, t = 1) &= Q_N^{\min} + 0.75 (Q_N^{\max} - Q_N^{\min}) \\ B_{C,j}(z, t = 1) &= \frac{B_{N,j}(z, t = 1)}{Q_{N,j}(z, t = 1)} \end{aligned} \quad (36)$$

Chlorophyll was initialised similarly by assuming the chlorophyll-nitrogen ratio was 75% of the maximum, θ .

$$B_{\text{Chl},j}(z, t = 1) = 0.75 \theta B_{N,j}(z, t = 1) \quad (37)$$

Organic matter variables were initialised from assumed ratios with plankton concentrations as estimates were not available from SINMOD outputs. We initialised organic matter concentrations at 5% of the total plankton concentration and split this evenly between DOM and POM.

$$D_{i,\text{DOM}}(z, t = 1) = 0.5 \left(0.05 \sum_j B_{i,j} \right)$$

$$D_{i,\text{POM}}(z, t = 1) = (1 - 0.5) D_{i,\text{DOM}}(z, t = 1) \quad (38)$$

2.4 Parameter optimisation

2.4.1 Matching model output to data

Comparing data to model outputs required matching all data, $Y_{v,d,s,g}^{\text{obs}}$ and $\mathbf{S}_{f,g}^{\text{obs}}$, to the equivalent model outputs, $Y_{v,d,s,g}^{\text{mod}}$ and $\mathbf{S}_{f,g}^{\text{mod}}$, which we derive in this section. Deriving $Y_{v,d,s,g}^{\text{mod}}$ followed a similar procedure for all measurement types v . This differed from the procedure used to derive $\mathbf{S}_{f,g}^{\text{mod}}$ due to the difference in data type (vector vs scalar) and structure (averaged over events and depths). Let us first describe the derivation of $Y_{v,d,s,g}^{\text{mod}}$, and then $\mathbf{S}_{f,g}^{\text{mod}}$.

Let t_s and \mathbf{z}_s denote the time and the discrete measurement depths of sample event s . For each sample event all modelled state variables were evaluated at $t = t_s$. Modelled values of $v \in \{\text{DIN}, \text{PON}, \text{POC}, \text{Chl } a\}$ were extracted as $N(z, t_s)$, $M_{\text{N,POM}}(z, t_s)$, $M_{\text{C,POM}}(z, t_s)$, and $\sum_j B_{\text{Chl},j}(z, t_s)$ respectively. We then linearly interpolated these modelled values to generate outputs corresponding to the sampled depths, \mathbf{z}_s , and specified the modelled equivalent of the data: $Y_{\text{DIN},d,s,g}^{\text{mod}} = N(\mathbf{z}_s, t_s)$, $Y_{\text{PON},d,s,g}^{\text{mod}} = M_{\text{N,POM}}(\mathbf{z}_s, t_s)$, $Y_{\text{POC},d,s,g}^{\text{mod}} = M_{\text{C,POM}}(\mathbf{z}_s, t_s)$, and $Y_{\text{Chl},d,s,g}^{\text{mod}} = \sum_j B_{\text{Chl},j}(\mathbf{z}_s, t_s)$.

To extract $\mathbf{S}_{f,g}^{\text{mod}}$ from modelled output we needed to convert plankton concentrations into biovolumes. Plankton cell density was calculated as carbon concentration divided by cell carbon quota, $D_j(z, t) = B_{\text{C},j}(z, t)/Q_{\text{C},j}$, and then biovolume was found by multiplying by cell volumes (assumed spherical), $S_{j,f,g}^{\text{mod}}(z, t) = \text{Vol}_j D_j(z, t)$. Evaluating at the average time, \bar{t}_s , and depth, \bar{z}_s , of size samples and, yielded the required output, $\mathbf{S}_{f,g}^{\text{mod}} = \mathbf{S}_{f,g}^{\text{mod}}(\bar{z}_s, \bar{t}_s)$.

2.4.2 Cost function

Model fit to data was optimised by numerically minimising a cost function denoted as $\mathcal{C}(\boldsymbol{\theta} | \mathbf{Y}^{\text{obs}}, \mathbf{S}^{\text{obs}})$, where $\boldsymbol{\theta}$ is a vector of all tuning parameters. The cost function returns a scalar value, the “cost”, representing discrepancy between data, $\{\mathbf{Y}^{\text{obs}}, \mathbf{S}^{\text{obs}}\}$, and the parameter-dependent modelled approximations to those data, $\{\mathbf{Y}^{\text{mod}}(\boldsymbol{\theta}), \mathbf{S}^{\text{mod}}(\boldsymbol{\theta})\}$. As the scalar nutrient

data and the abundance-at-size data differed in structure and sampling design they required different cost function terms, denoted as $\mathcal{C}^{\text{nutrient}}(\boldsymbol{\theta} | \mathbf{Y}^{\text{obs}})$ and $\mathcal{C}^{\text{size}}(\boldsymbol{\theta} | \mathbf{S}^{\text{obs}})$. This section describes the construction of cost function terms for each data type and how these were combined to form \mathcal{C} .

The standardised scalar data were vectors, $\hat{\mathbf{Y}}_{v,g}^{\text{obs}}$, with $n_{v,g}$ elements accounting for measurements of $v \in \{\text{DIN, Chl, PON, POC}\}$ across all sample events and depths from group $g \in \{\text{Arctic, Atlantic}\}$. The equivalent modelled vectors, $\hat{\mathbf{Y}}_{v,g,q}^{\text{mod}}$, were further indexed by q to distinguish particle trajectories from the physical model. For each measurement type, v , group, g , and trajectory, q , the cost was calculated as the geometric mean of absolute errors, scaled by the range of the standardised data.

$$\mathcal{C}_{v,g,q}^{\text{nutrient}}(\boldsymbol{\theta} | \hat{\mathbf{Y}}_{v,g}^{\text{obs}}) = \frac{\left(\prod_{k=1}^{n_{v,g}} \left| \hat{Y}_{k,v,g}^{\text{obs}} - \hat{Y}_{k,v,g,q}^{\text{mod}}(\boldsymbol{\theta}) \right| \right)^{1/n_{v,g}}}{\text{range}(\hat{Y}_{k,v,g}^{\text{obs}})} \quad (39)$$

Using absolute errors rather than squared errors, and the geometric rather than arithmetic mean, made the cost function more robust to overfitting outlying data points — particularly some of the deepest samples that the model could not replicate. The sole purpose of including the data range as a constant scaling term was to produce $\mathcal{C}_{v,g,q}^{\text{nutrient}}$ of similar magnitude to the cost terms calculated for the size data. The cost was then averaged over groups, g , and trajectories, q ,

$$\mathcal{C}_v^{\text{nutrient}} = \frac{1}{n_g n_q} \sum_{g,q} \mathcal{C}_{v,g,q}^{\text{nutrient}} \quad (40)$$

to give a single cost for each measurement type. Finally, the total cost for the scalar data was found by averaging over measurement types.

$$\mathcal{C}^{\text{nutrient}} = \frac{1}{3} \left(\mathcal{C}_{\text{DIN}}^{\text{nutrient}} + \mathcal{C}_{\text{Chl}}^{\text{nutrient}} \right) \quad (41)$$

$$+ \frac{1}{2} \left(\mathcal{C}_{\text{PON}}^{\text{nutrient}} + \mathcal{C}_{\text{POC}}^{\text{nutrient}} \right) \quad (42)$$

The PON and POC data were downweighted by treating their combined cost as generated from a single measurement type. This was done to avoid overfitting potentially caused by the high correlation between these data.

The plankton size data were vectors, $\mathbf{S}_{f,g}^{\text{obs}}$, of biovolume-at-size for trophic levels $f \in \{\text{autotroph, heterotroph}\}$ and water mass groups $g \in \{\text{Arctic, Atlantic}\}$. These data and their modelled equivalents, $\mathbf{S}_{f,g,q}^{\text{mod}}$, were decomposed into scalars of total biovolume, $\tilde{S}_{f,g}$, multiplied into simplices of relative biovolume-at-size, $\hat{S}_{f,g}$

$$\mathbf{S}_{f,g} = \tilde{S}_{f,g} \hat{S}_{f,g} \quad (43)$$

where $\tilde{S}_{f,g} = \sum_j S_{j,f,g}$ and $\sum_j \hat{S}_{j,f,g} = 1$. We created separate cost terms for the total biovolume and the relative biovolume-at-size, denoted as $\tilde{\mathcal{C}}^{\text{size}}$ and $\hat{\mathcal{C}}^{\text{size}}$ respectively.

Hellinger distances were used as the cost of the relative biovolume-at-size. Hellinger distance is a bounded scalar metric representing dissimilarity between probability densities \mathbf{p} and \mathbf{q} .

$$H(\mathbf{p}, \mathbf{q}) \equiv \left(1 - \sum_i (p_i q_i)^{\frac{1}{2}} \right)^{\frac{1}{2}} \quad (44)$$

Hellinger distance is maximised, $H(\mathbf{p}, \mathbf{q}) = 1$, for two disparate probability densities ($p_i q_i = 0 \forall i$), and is minimised, $H(\mathbf{p}, \mathbf{q}) = 0$, when $p_i = q_i \forall i$. Cost values for relative biovolume-at-size were calculated as

$$\hat{\mathcal{C}}_{f,g,q}^{\text{size}} (\boldsymbol{\theta} | \hat{S}_{f,g}^{\text{obs}}) = H \left(\hat{S}_{f,g}^{\text{obs}}, \hat{S}_{f,g,q}^{\text{mod}} (\boldsymbol{\theta}) \right) \quad (45)$$

and then a single scalar value was calculated

$$\hat{\mathcal{C}}^{\text{size}} = \frac{1}{n_f n_g n_q} \sum_{f,g,q} \hat{\mathcal{C}}_{f,g,q}^{\text{size}} \quad (46)$$

by averaging over trophic group, water mass group, and particle trajectory.
[is the order of averaging and rel/tot weightings important? I don't think so but double check...]

The cost of model misfit to total biovolume — scalar data — was described using a function, $J(x, y)$, designed to have similar properties to the Hellinger distance metric.

$$\begin{aligned} J(x, y) &\equiv \frac{1 - \exp(-a u)}{1 + \exp(-a u)} \\ u &= |\ln(x/y)| \\ a &= \ln(3)/\ln(2) \end{aligned} \quad (47)$$

Like Hellinger distances, this metric, $J(x, y)$, is bounded in the $[0, 1]$ interval, with $J(x, x) = 0$ and $x \ll y$ or $x \gg y \implies J(x, y) = 1$. The shape-parameter, a , was chosen such that $J(x, 2x) = J(x, \frac{x}{2}) = \frac{1}{2}$, i.e., such that the metric midpoint was attained when modelled values were either twice or half the data values. Due to their similar properties the H and J function outputs were easily combined into a single bounded metric to represent overall cost of size data. Cost values for the total biovolumes were found as

$$\tilde{\mathcal{C}}_{f,g,q}^{\text{size}} (\boldsymbol{\theta} | \tilde{S}_{f,g}^{\text{obs}}) = J \left(\tilde{S}_{f,g}^{\text{obs}}, \tilde{S}_{f,g,q}^{\text{mod}} \right) \quad (48)$$

and then averaged to produce a single scalar cost value for total biovolume.

$$\tilde{\mathcal{C}}^{\text{size}} = \frac{1}{n_f n_g n_q} \sum_{f,g,q} \tilde{\mathcal{C}}_{f,g,q}^{\text{size}} \quad (49)$$

The overall cost of size data was then found by a weighted average over the components for total and relative biovolume

$$\mathcal{C}^{\text{size}} = \frac{1}{w_s + 1} \left(w_s \hat{\mathcal{C}}^{\text{size}} + \tilde{\mathcal{C}}^{\text{size}} \right) \quad (50)$$

where w_s is the weight ascribed to $\hat{\mathcal{C}}^{\text{size}}$ relative to $\tilde{\mathcal{C}}^{\text{size}}$. Finally, the total cost was calculated as

$$\mathcal{C} = \frac{1}{w_c + 1} (\mathcal{C}^{\text{nutrient}} + w_c \mathcal{C}^{\text{size}}) \quad (51)$$

where w_c is the weighting ascribed to $\mathcal{C}^{\text{size}}$ relative to $\mathcal{C}^{\text{nutrient}}$.

2.4.3 Optimising algorithm

The cost function was minimised using a genetic algorithm. The *ga* MatLab function was used with the default settings.

3 Results

4 Discussion

References

- Banas, N. (2018). *Particulator*: MatLab-based particle tracking code for hydrodynamic models. <https://github.com/neilbanas/particulator>.
- Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1998). A dynamic regulatory model of photoacclimation to light, nutrients and temperature. *Limnology and Oceanography*, 43, 679–694.
- Kiørboe, T. (2008). *A Mechanistic Approach to Plankton Ecology*. Princeton University Press.
- Krause-Jensen, D. & Sand-Jensen, K. (1998). Light attenuation and photosynthesis of aquatic plant communities. *Limnology and Oceanography*, 43(3), 396–407.
- Lampe, V., Nöthig, E.-M., & Schartau, M. (2021). Spatio-temporal variations in community size structure of Arctic protist plankton in the Fram Strait. *Frontiers in Marine Science*, 7, 1239.
- Litchman, E., Klausmeier, C. A., Schofield, O. M., & 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(12), 1170–1181.

- 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, 371–379.
- Paliwal, K., Agarwal, A., & Sinha, S. S. (1982). A modification over Sakoe and Chiba's dynamic time warping algorithm for isolated word recognition. *Signal Processing*, 4(4), 329–333.
- Sakoe, H. & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49.
- Slagstad, D. & McClamans, T. A. (2005). Modeling the ecosystem dynamics of the Barents Sea including the marginal ice zone: I. physical and chemical oceanography. *Journal of Marine Systems*, 58(1), 1–18.
- Torres-Valdés, S., Morische, A., & Wischnewski, L. (2019). Nutrient measurements from POLARSTERN cruise PS114 (LTER HAUSGARTEN). <https://doi.pangaea.de/10.1594/PANGAEA.907355>.
- von Jackowski, A., Grosse, J., Nöthig, E.-M., & Engel, A. (2020). Organic matter and bacteria measurements of Polarstern cruise PS114 and Maria S. Merian cruise MSM77. <https://doi.org/10.1594/PANGAEA.915751>.
- 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.
- Wassmann, P., Slagstad, D., Riser, C. W., & Reigstad, M. (2006). Modelling the ecosystem dynamics of the Barents Sea including the marginal ice zone: II. carbon flux and interannual variability. *Journal of Marine Systems*, 59(1), 1–24.

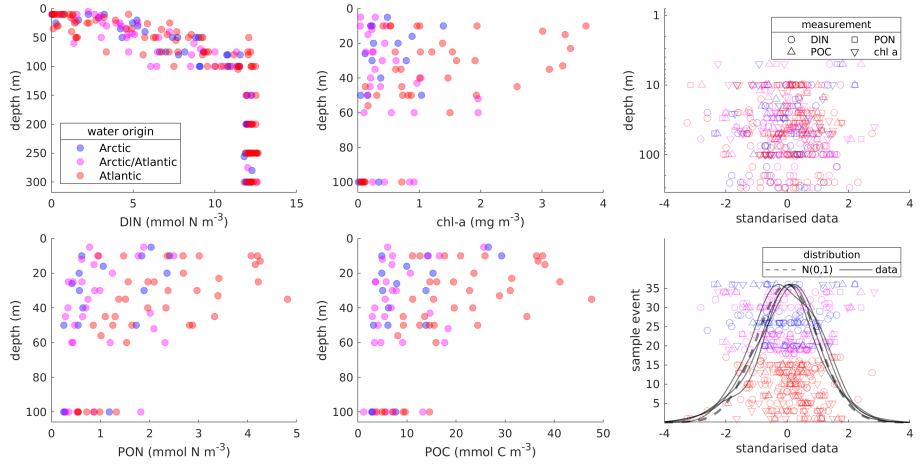


Figure 1: Nutrient and organic matter data collected during the 2018 RV Polarstern cruise – all sample events and depths used for model optimisation. Mechanistic relationships with depth are apparent; the within-depth variability results from multiple sampling events. Colours indicate the water origin of each sample, determined by the closest physical model trajectories.

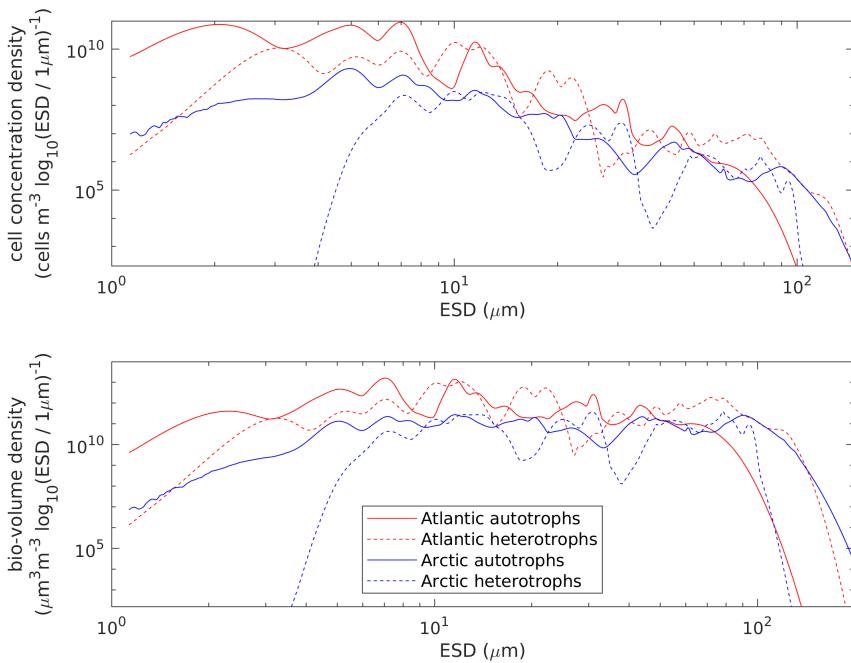


Figure 2: Plankton size spectra data derived using samples from the 2018 Polarstern cruise. Lines display observed spectra, averaged over depths and sample events, for autotrophic and heterotrophic plankton and grouped by samples from water originating from the Atlantic and the Arctic.

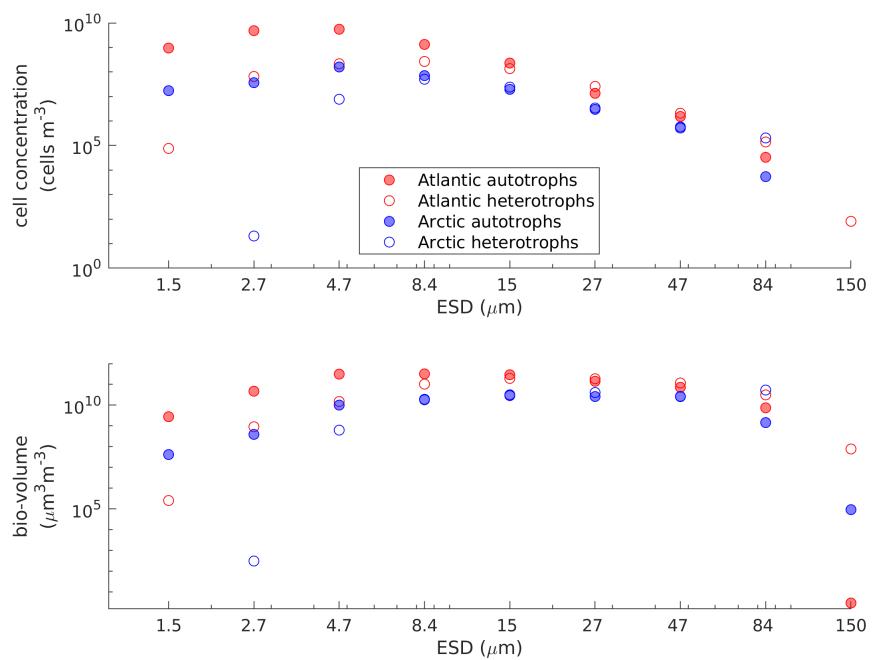


Figure 3: Observed plankton density-at-size. Points associated to each size group display total plankton within cell size intervals that are equally sized and evenly spaced on a logarithmic scale.

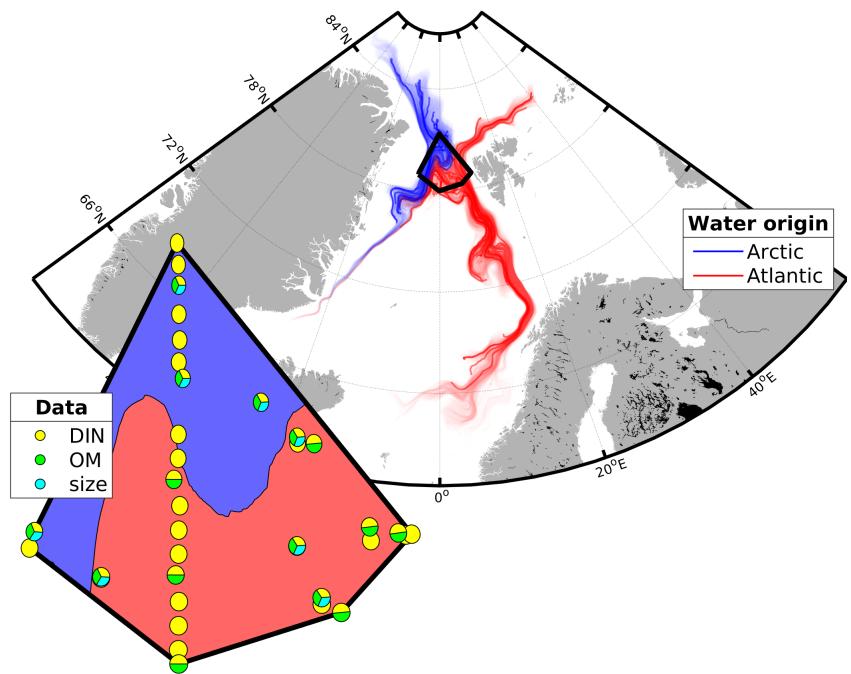


Figure 4: The Fram Strait study area. The black polygon surrounds shipboard sample sites. The map inset magnifies this polygon to display positions of each 2018 sample and the data types measured at each site. Particle trajectories from the physical model are displayed as lines representing horizontal transport from 1st Jan. until 31st Oct. [check this date]. Points at one end of each line indicate initial locations of trajectories, which all lay within the polygon at the time of sampling. Trajectories originating from the Arctic and Atlantic are coloured blue and red to represent cold and warm water masses, which is also indicated in the magnified map inset.

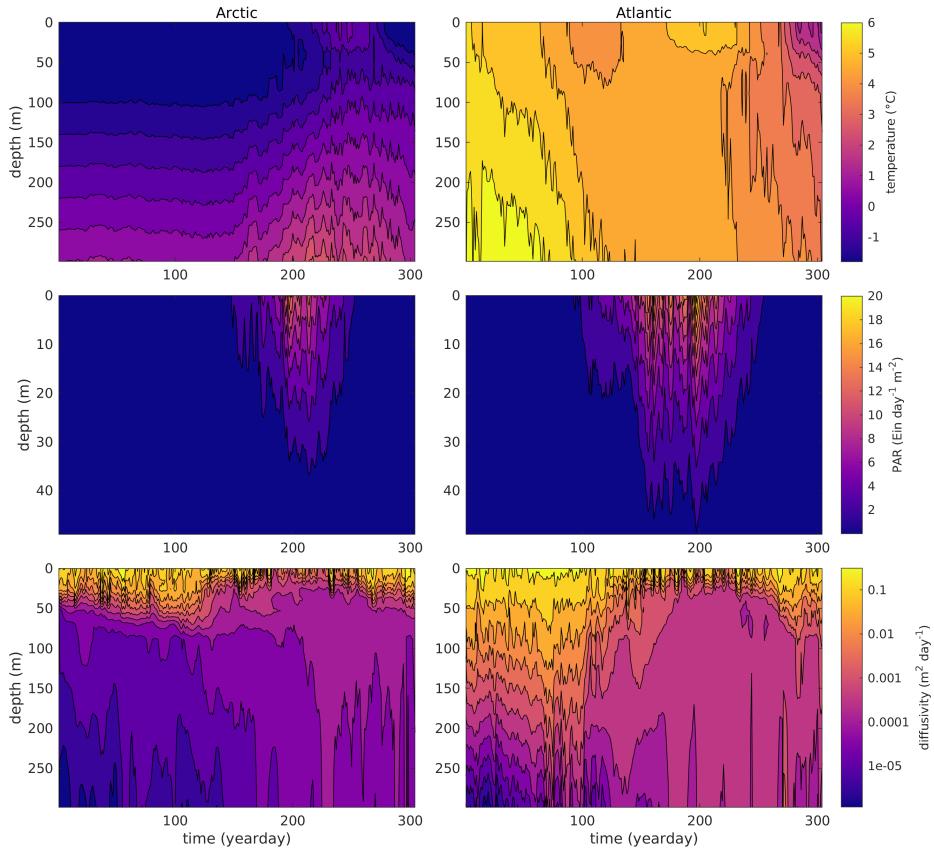


Figure 5: Time series of depth dependent forcing data, averaged over all trajectories originating from the Arctic (left) and from the Atlantic (right).

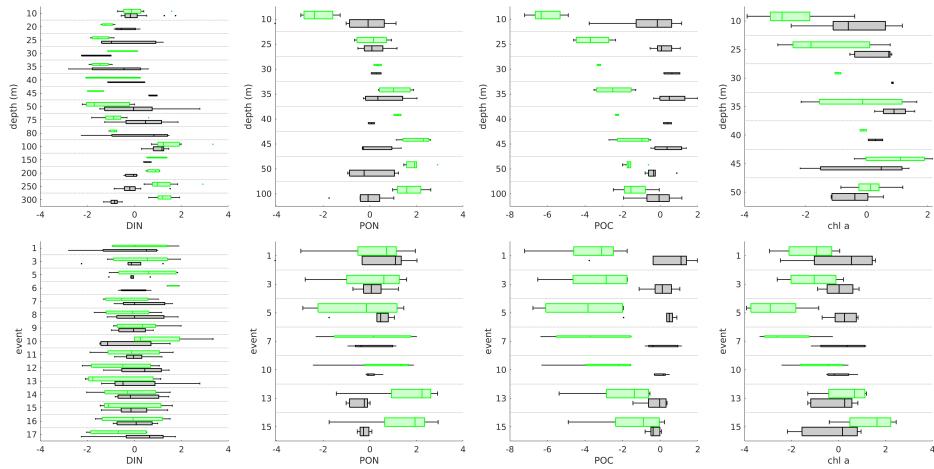


Figure 6: Model fit to depth-discrete, scalar observations.

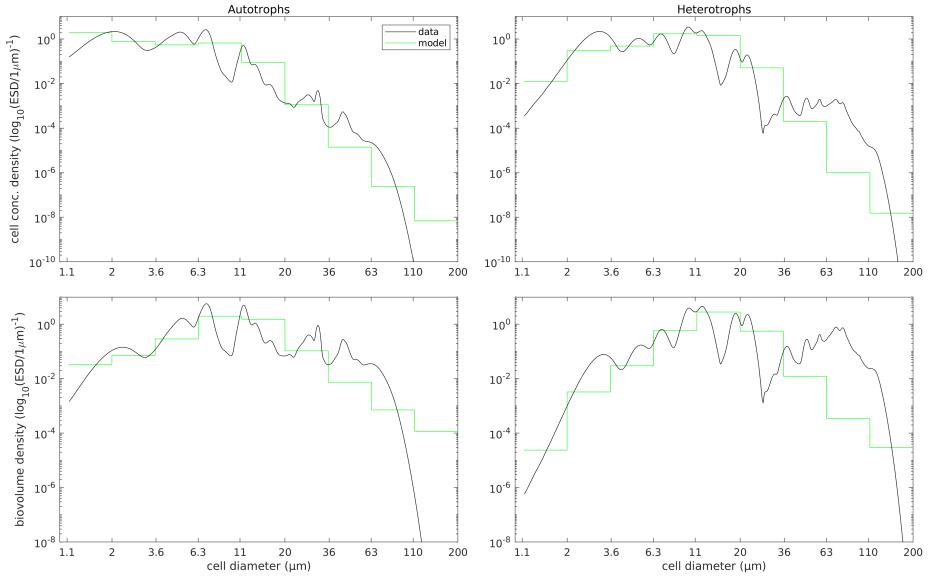


Figure 7: Model fit to cell concentration- and biovolume-density observations.

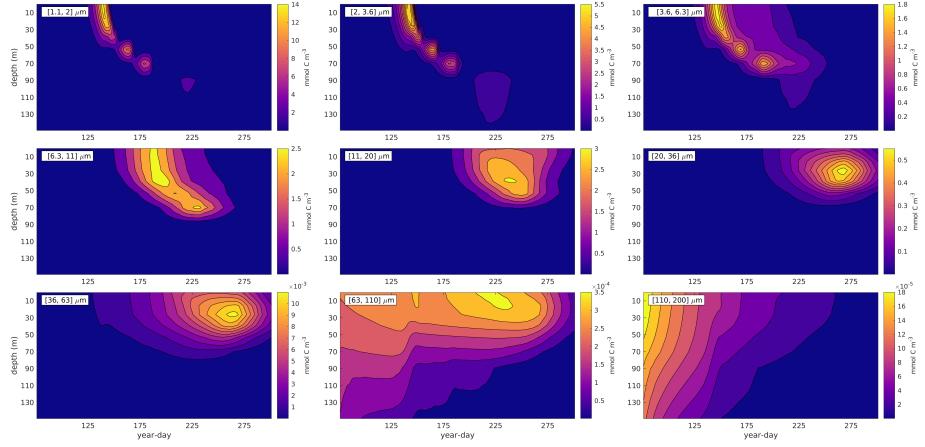


Figure 8: Autotroph carbon concentration over depth and time for each modelled size class.

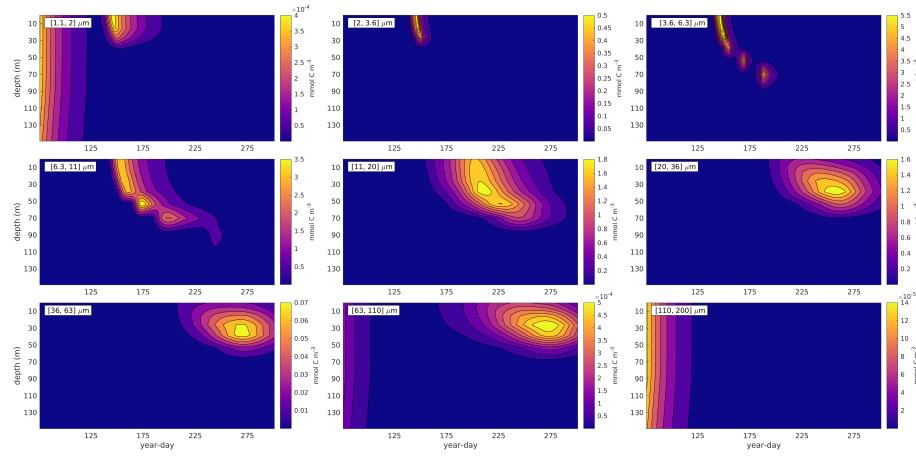


Figure 9: Heterotroph carbon concentration over depth and time for each modelled size class.

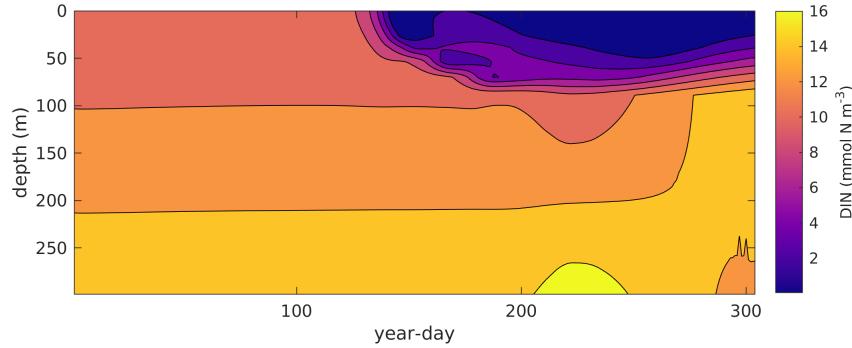


Figure 10: Modelled dissolved inorganic nitrogen concentration over depth and time.

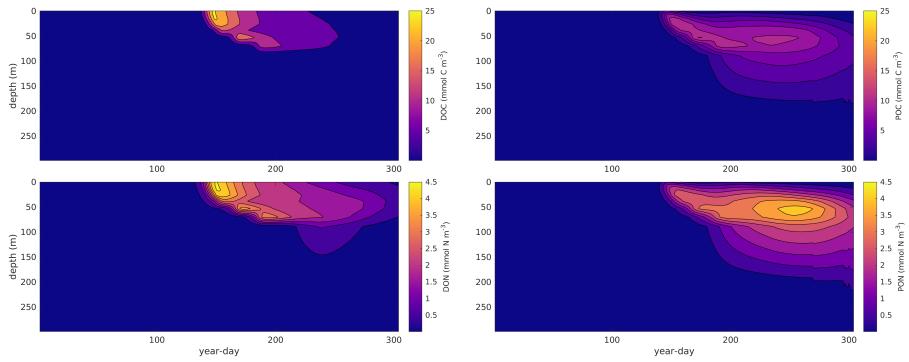


Figure 11: Modelled organic matter concentrations over depth and time:
top/bottom = carbon/nitrogen; left/right = dissolved/particulate

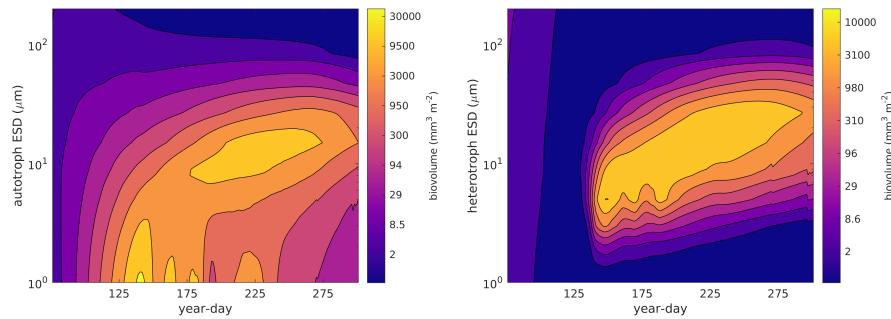


Figure 12: Temporal trends in modelled biovolume-at-size distributions, averaged over all trajectories of Atlantic origin.

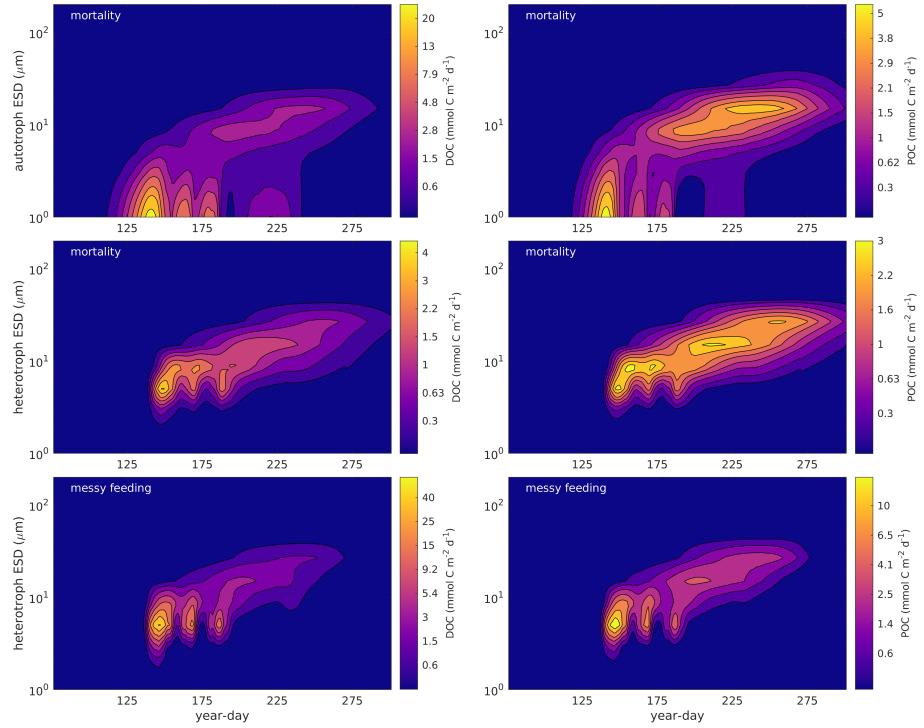


Figure 13: Modelled organic matter production by all cell size classes over time.

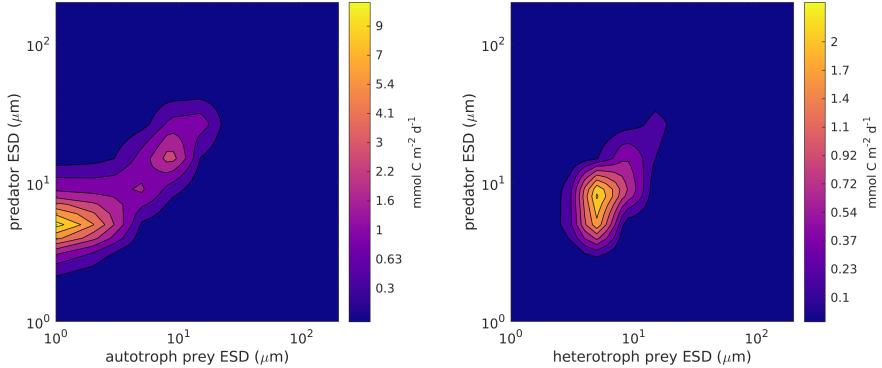


Figure 14: Modelled carbon flux from grazing, integrated over depth and averaged over times corresponding to the growing season (year-days 100–250).

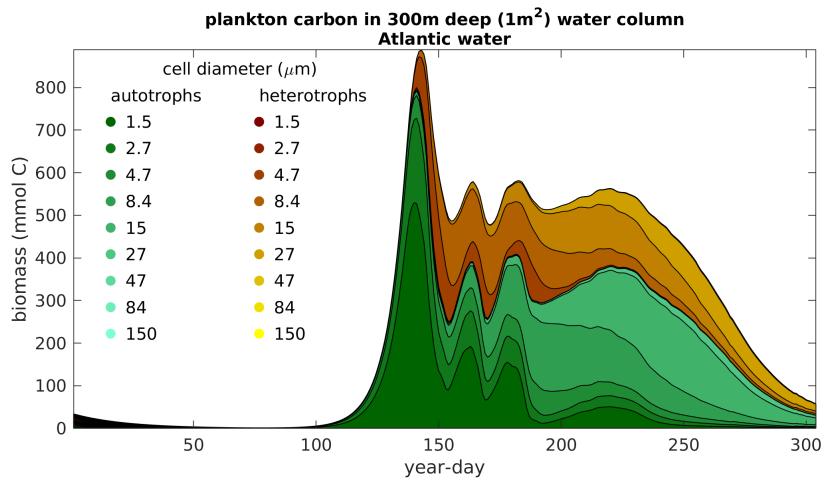


Figure 15: Time series of depth-integrated plankton carbon biomass, averaged over trajectories originating from Atlantic waters.

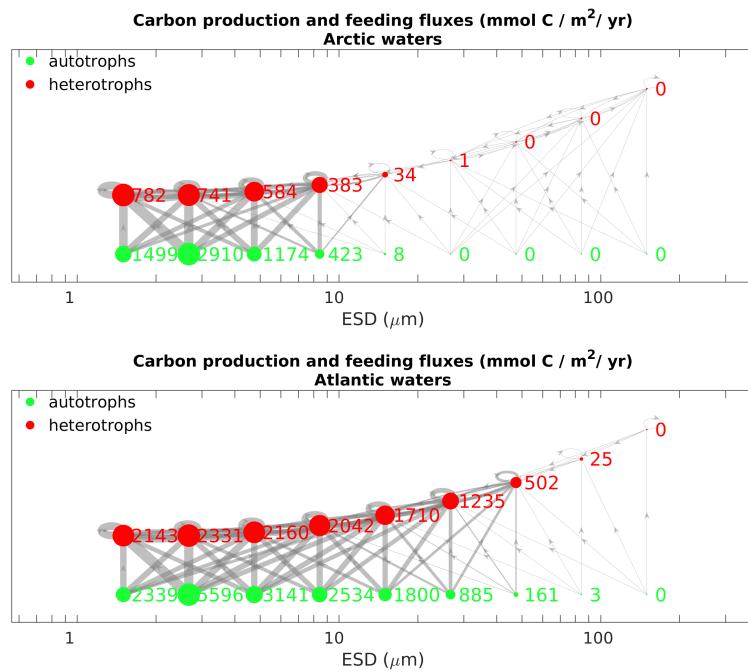


Figure 16: Total annual carbon production and feeding fluxes, averaged over trajectories originating from Arctic and from Atlantic waters.

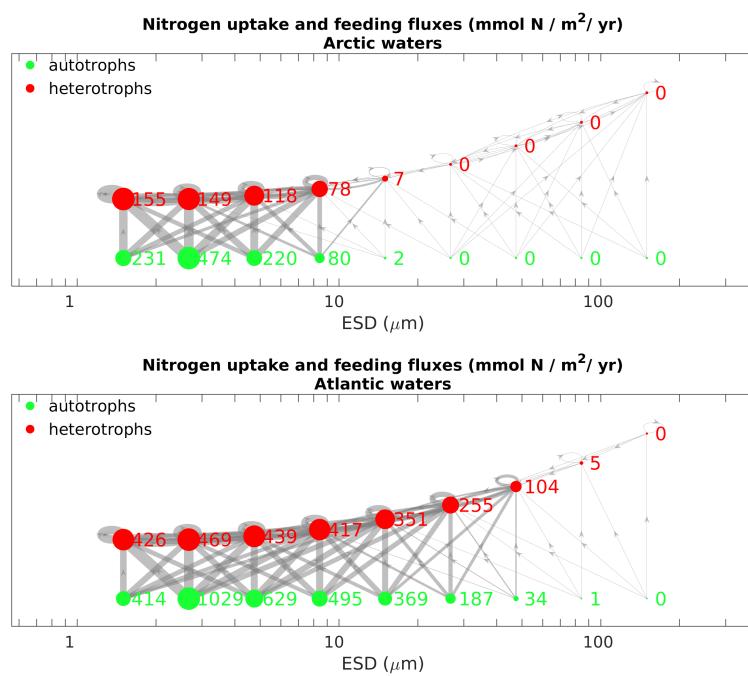


Figure 17: Total annual nitrogen production and feeding fluxes, averaged over trajectories originating from Arctic and from Atlantic waters.

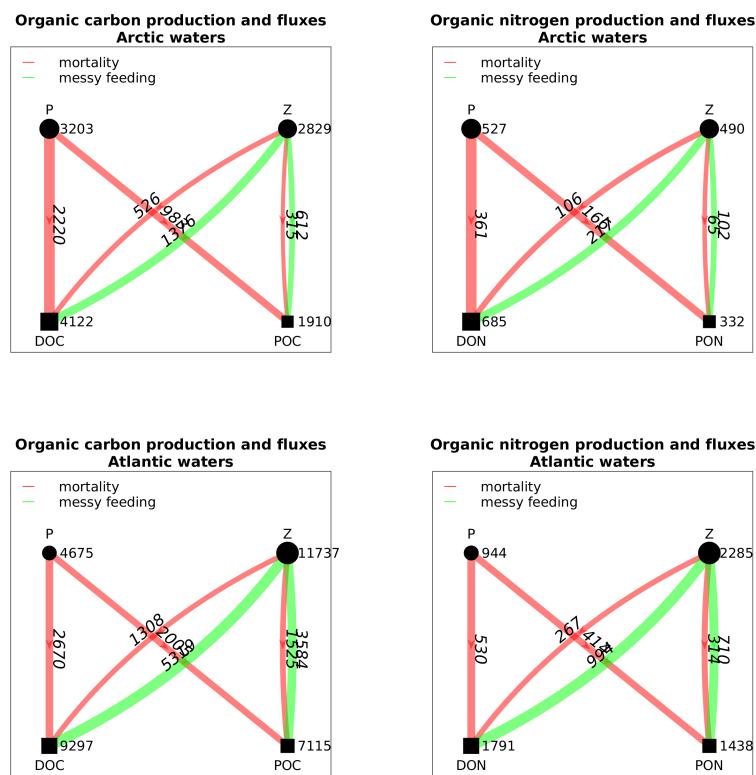


Figure 18: Total annual organic matter production and fluxes ($\text{mmol}\{\text{C}, \text{N}\} \text{m}^{-2} \text{yr}^{-1}$) from autotrophs (P) and heterotrophs (Z).

Table 1: Cruise dates and availability of each data type — dissolved inorganic nitrogen, organic matter, size spectra and physical model forcing data.

Cruise	Dates	DIN	OM	Size	Forcing
PS99	25/6 – 10/7/2016	✓	✓	✓	✗
PS107	25/7 – 15/8/2017	✓	✗	✓	✓
PS114	16/7 – 27/7/2018	✓	✓	✓	✓

Table 2: Definition of all terms in flux equations (eqs. (8) to (10)).

Notation	Description
K	Vertical diffusivity from physical model
$V_{i,j}$	Nutrient uptake rate
$r_{i,k}$	Remineralisation rates of DOM and POM
$G_{i,j_z,j}$	Grazing rate of heterotrophs j_z on plankton j
λ_{i,j_z}	Assimilation efficiency of heterotrophs
m_j	Background mortality rate
w_k	Sinking rate of organic matter
$S_{i,k}^M$	Sources of organic matter
t	Time
z	Depth

Table 3: Size-independent parameters. Numerically optimised parameter values are displayed in bold above their bounds. Other parameter values were fixed.

Parameter	Notation	Value	Units
Rate-limiting parameters			
Reference temperature	T^{ref}	20	°C
Temperature sensitivity	A	0.05	dimensionless
Uptake regulation curvature	h	0.1	dimensionless
Nutrient quotas			
Min. nitrogen:carbon quota ^(a)	Q_N^{\min}	0.16 [0.07, 0.23]	mmol N (mmol C) ⁻¹
Photosynthesis			
Initial slope of P-I curve	α_p	3.3×10^{-6} [$10^{-10}, 5.75 \times 10^{-6}$]	mmol C (mg Chl a) ⁻¹
Cost of biosynthesis ^(b)	ξ	2.33	mmol C (mmol N) ⁻¹
Max. Chl <i>a</i> -to-nitrogen ratio ^(b)	θ	4.2	mg Chl a (mmol N) ⁻¹
Grazing			
Optimum predator:prey length ratio ^(c)	δ_{opt}	10	dimensionless
Geometric SD of prey availability	σ	0.92 [0.25, 2.5]	dimensionless
Prey clearance rate	α_G	6.3 [0.17, 35]	$\text{m}^3 \text{d}^{-1}$ (mmol C) ⁻¹
Prey refuge parameter	Λ	-1	dimensionless
Max. assimilation efficiency ^(b)	λ^{\max}	0.7	dimensionless
Organic matter			
DOM sinking speed	w_{DOM}	0	m d^{-1}
POM sinking speed	w_{POM}	0.75 [0.5, 10] 0.040	m d^{-1}
DOM remineralisation rates	$r_{N,\text{DOM}}$	[0.005, 0.06]	d^{-1}
	$r_{C,\text{DOM}}$	$r_{N,\text{DOM}}$ 0.019	d^{-1}
POM remineralisation rates	$r_{N,\text{POM}}$	[0.01, 0.12,]	d^{-1}
	$r_{C,\text{POM}}$	$r_{N,\text{POM}}$	d^{-1}
Mortality			
Linear plankton mortality	m	0.046 [0.02, 0.10]	d^{-1}

(a) Marañón et al. (2013); (b) Geider et al. (1998); (c) Kiørboe (2008)

Table 4: Size-dependent parameters ($x = a \text{Vol}^b$). Numerically optimised parameter values are displayed in bold above their bounds. Other parameter values were fixed.

Parameter	Notation	a	b	Units
Nutrient quotas				
Carbon quota ^(a)	Q_C	1.7×10^{-11}	0.88	mmol C cell^{-1}
Max. N:C quota transform ^(a,b)	\tilde{Q}_N^{\max}	0.37 [0.16, 1.00]	-0.073 [-0.19, 0]	dimensionless
Nutrient uptake				
Max. uptake rate ^(a)	v^{\max}	0.095 [0.067, 0.167]	0.028 [0.01, 0.18]	$\text{mmol N (mmol C)}^{-1} \text{d}^{-1}$
Nutrient affinity ^(c)	α	0.77 [0.059, 2.34]	0.020 [-0.30, 0.10]	$\text{m}^3 (\text{mmol C})^{-1} \text{d}^{-1}$
Photosynthesis				
Max. photosynthetic rate	P^{\max}	3.46 [0.5, 5]	-0.17 [-0.5, 0]	d^{-1}
Grazing				
Max. prey capture rate	G^{\max}	20.9 [5, 35]	-0.27 [-0.5, 0]	d^{-1}

(a) [Marañón et al. \(2013\)](#);

(b) Transformed Q_N^{\max} is defined in section [2.3.2](#);

(c) [Litchman et al. \(2007\)](#)