This is a 1D NPZD model that is forced using time series of temperature, light intensity and diffusivity values derived from a physical oceanographic model. The model simulates plankton dynamics within a vertical water column which moves with the current through varying environments.

We want to numerically tune parameters by fitting model output to data collected from numerous sites. This means that we need to run the model across a suite of "particle trajectories" containing forcing data representing various paths leading to the data sampling sites.

ODE system

Modelled state variables are concentrations of inorganic nitrogen, N; plankton, B_{ik}; and nitrogen organic matter, M_{jk}. Concentrations have units of mmol element m^{-3} or mg chlorophyll m^{-3}. There are n phytoplankton size classes and a single zooplankton class which are indexed by $i \in \{i_p, i_z\} = \{1, \ldots, n+1\}$. Dissolved and particulate organic matter are indexed by j. The k indices refer to element: phytoplankton, $k \in \{\mathrm{C} = \text{carbon}, \mathrm{N} = \text{nitrogen}, \mathrm{Chl} = \text{chlorophyll}\}$; zooplankton, $k \in \{\mathrm{C}\}$; organic matter, $k \in \{\mathrm{C}, \mathrm{N}\}$. Fluxes are described by a system of ODEs

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial N}{\partial z} - \sum_i V_{i\mathrm{N}}\, B_{i\mathrm{C}} + \sum_j r_{j\mathrm{N}}\, M_{j\mathrm{N}} \tag{1}$$

$$\frac{\partial B_{ik}}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial B_{ik}}{\partial z} + V_{ik}\, B_{i\mathrm{C}} - G_{ik}\, B_{i_z\mathrm{C}} + \lambda\, B_{i_z\mathrm{C}} \sum_i G_{ik} - m\, B_{ik} \tag{2}$$

$$\frac{\partial M_{jk}}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial M_{jk}}{\partial z} - w_j \frac{\partial M_{jk}}{\partial z} + (1-\lambda)\, B_{i_z\mathrm{C}} \sum_i \beta_{ij}\, G_{ik} + m \sum_i \beta_{ij}\, B_{ik} - r_{jk}\, M_{jk} \tag{3}$$

where

K	Vertical diffusivity from physical model
V_{ik}	Nutrient uptake rate
r_{jk}	Remineralisation rates of DON and PON
G_{ik}	Grazing rate of zooplankton i_z on all plankton classes i
λ	Assimilation efficiency of zooplankton
m	Background mortality rate
w_j	Sinking rate of PON and DON
β_{ij}	Proportional allocation of organic matter into PON and DON
t	Time
z	Depth.

Parameter	Notation	Value	Unit
Reference temperature	T^{ref}	20	°C
Temperature sensitivity	A	0.05	-
Linear mortality	m	0.05	d^{-1}
Curvature of uptake quota-limitation	h	10	-
Initial slope of photosynthesis-irradiance curve	α^P	7×10^{-3}	$(\mu\text{Ein}\,\text{m}^{-2})^{-1}$
Cost of biosynthesis	ξ	2.33	mmol C $(\text{mmol N})^{-1}$
Maximum Chl:N ratio	θ	4.2	(mg Chl)$(\text{mmol N})^{-1}$
Maximum grazing rate	$G^{\max}$	22	d^{-1}
Total prey half saturation	k_G	1	mmol N m^{-3}
Prey refuge parameter	Λ	-1	-
Assimilation efficiency	λ	0.7	-
DOM remineralisation rate	r_{DOM}	0.02	day^{-1}
POM remineralisation rate	r_{POM}	0.04	day^{-1}
POM sinking speed	w_{POM}	10	m day^{-1}

Table 1: Scalar parameters with default values.

Parameter	Notation	a	b	Unit
C. cell quota	Q_{C}	1.7×10^{-11}	0.88	mmol C cell^{-1}
Min N. quota	$Q^{\min}/Q_{\text{C}}$	0.14	-0.04	mmol N $(\text{mmol C})^{-1}$
Max N. quota	$Q^{\max}/(Q^{\max} - Q^{\min})$	0.62	-0.09	-
Max N. uptake rate	$v^{\max}/Q_C$	0.10	0.09	mmol N $(\text{mmol C}\,\text{day})^{-1}$
N. affinity	α/Q_C	0.59	0.61	$\text{m}^3\,(\text{mmol C}\,\text{day})^{-1}$
Max C. uptake rate	$p^{\max}$	6	0.1	day^{-1}

Table 2: Size-dependent parameters, $x = a\,\text{Vol}^b$, with default values.

Cell quotas & limiting terms

Plankton abundance is tracked as the carbon concentration, $B_{i\mathrm{C}}$. Concentrations of other nutrients vary relative to carbon, within limits preventing excessive accumulation or depletion of any one nutrient. Cellular nitrogen and chlorophyll concentrations are modelled as ratios with cell carbon quotas.

$$Q_{ik} = \frac{B_{ik}}{B_{i\mathrm{C}}} \tag{4}$$

Carbon production is limited using a linear function of the nitrogen quota.

$$\gamma_{i\mathrm{N}} = \frac{Q_{i\mathrm{N}} - Q_{i\mathrm{N}}^{\min}}{Q_{i\mathrm{N}}^{\max} - Q_{i\mathrm{N}}^{\min}} \tag{5}$$

Nitrogen uptake rate is adjusted according to cell quota using a function related to the beta cumulative distribution

$$Q_{i\mathrm{N}}^{\mathrm{stat}} = 1 - \gamma_{i\mathrm{N}}^{h} \tag{6}$$

that ensures uptake rate is down-regulated once the N quota becomes large, and is set to zero when quota reaches its maximum.

Nutrient uptake and assimilation rates are adjusted for temperature, T, according to a sensitivity, A, and relative to reference temperature, T^{ref}.

$$\gamma_T = e^{A(T - T^{\mathrm{ref}})} \tag{7}$$

Nitrogen uptake

Nitrogen uptake is modelled using a Michaelis-Menton function, modified by quota- and temperature-limitation terms.

$$V_{i_p\mathrm{N}} = \frac{v_{i_p}^{\max}\, \alpha_{i_p}\, N}{\alpha_{i_p}\, N + v_{i_p}^{\max}}\, Q_{i_p\mathrm{N}}^{\mathrm{stat}}\, \gamma_T \tag{8}$$

All uptake rates are set to zero, $V_{i_z k} = 0$, for zooplankton.

Photosynthesis

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

$$p_{i_p}^{\mathrm{sat}} = p_{i_p}^{\max}\, \gamma_T\, \gamma_{i_p\mathrm{N}} \tag{9}$$

Photosynthetic rate is expressed as a Poisson function of irradiance and chlorophyll quota.

$$p_{i_p} = p_{i_p}^{\mathrm{sat}} \left(1 - \exp\left(\frac{-\alpha_p \, Q_{i_p \mathrm{Chl}} \, I}{p_{i_p}^{\mathrm{sat}}} \right) \right) \tag{10}$$

Carbon production rate is then given as

$$V_{i_p \mathrm{C}} = p_{i_p} - \xi \, V_{i_p \mathrm{N}} \tag{11}$$

where ξ represents cost of biosynthesis (a cost which rapidly declines to zero as N quota approaches its maximum...).

Chlorophyll concentrations are regulated by diverting a fraction of nitrogen uptake into chlorophyll

$$V_{i_p \mathrm{Chl}} = \rho_{i_p} \, V_{i_p \mathrm{N}} \tag{12}$$

where

$$\rho_{i_p} = \theta \frac{p_{i_p}}{\alpha_p \, Q_{i_p \mathrm{Chl}} \, I} \tag{13}$$

is the maximum Chl:N ratio, θ, multiplied by a fraction defined as the actual photosynthetic rate divided by the theoretical maximum-efficiency photosynthetic rate.

Predation

Modelled zooplankton are of indeterminate size and graze on all plankton classes. The zooplankton carbon biomass-specific grazing rate is given by

$$G_{i\mathrm{C}} = G^{\max} \frac{F}{k_G + F} (1 - \exp(\Lambda \, F)) \Phi_i \, \gamma_T. \tag{14}$$

where $F = \sum_i B_{i\mathrm{C}}$ is total available prey carbon and $G^{\max} \frac{F}{k_G + F}$ is a Michaelis-Menton function saturating at high prey abundance. This is modified by a prey refuge term, $(1 - \exp(\Lambda \, F))$, that restricts grazing when prey is limited, and a relative prey abundance term, $\Phi_i = B_{i\mathrm{C}}^2 / \sum_i B_{i\mathrm{C}}^2$, to target grazing on the most abundant prey class.

Multiplying by quota terms gives the zooplankton carbon biomass-specific grazing rates of phytoplankton nitrogen and chlorophyll.

$$G_{i_p k} = Q_{i_p k} \, G_{i_p \mathrm{C}} \tag{15}$$

Organic matter

Organic matter is produced from mortality and messy feeding. All plankton are subjected to a linear mortality rate, m, so the organic matter produced from mortality is

$$M_{jk}^{\text{mort}} = m \sum_i \beta_{ij} \, B_{ik} \tag{16}$$

where $\beta_{ij} = [\beta_i, 1 - \beta_i]$ allocates organic matter to DON or PON.

Organic matter produced from messy feeding is

$$M_{jk}^{\text{mess}} = (1 - \lambda) B_{i_z \text{C}} \sum_i \beta_{ij} \, G_{ik}. \tag{17}$$

where λ is prey assimilation efficiency of zooplankton. As eq. (17) involves the grazing rate of zooplankton nitrogen, which is not modelled, we must first estimate the portion of consumed nitrogen deriving from zooplankton prey. I've assumed that zooplankton N:C ratios are equal to those of the consumed prey.

Organic matter is lost through remineralisation at rates r_{jk}.

The $w_j \frac{dM_{jk}}{dz}$ term in eq. (3) models sinking of organic matter at speed w_j: only PON sinks.

Data and cost function

Model parameter values may be tuned using a cost function to numerically minimise discrepancies between model outputs and data. Informing the parameters related to each model component — ambient nutrient, organic matter, and plankton — requires multiple data sources. We used measurements of the concentrations of ambient nitrate, PON and POC, chlorophyll-a, and planktonic nitrogen size spectra. Each of these data sources were standardised and included within the cost function using methods depending upon data sampling design and relationships with covariates.

The nitrate, PON, POC, and chlorophyll-a data follow the same sampling design. These data contain measurements from unique sampling events, where each sample is a set of single measurements taken at a range of depths. Let Y_{det} denote these data, where $d \in \{1, \ldots, d_n\}$, $e \in \{1 \ldots, e_n\}$, and $t \in \{\text{N}, \text{PON}, \text{POC}, \text{Chl}\}$ index depth, sampling event, and data type. Measurements from each data type, Y_t, are not independent as they vary with depth, Z, and sampling event. We accounted for these sources of variability by using linear mixed models to standardise the data. This produced standardised data sets, $\tilde{Y}_t$, with values largely independent of depth and sampling event, thus limiting potential parameter estimation

bias due these covariates (fig. 1). Each of the t data sets were used separately to fit one of two linear mixed models

$$\begin{aligned} \log(Y_{det}) &= (a_t + a_{et}) + (b_t + b_{et})\, Z_{det} + \epsilon_{det}, \quad \epsilon_{det} \sim \mathcal{N}(0, \sigma_t) \qquad (18) \\ &= \mu_{det} + \epsilon_{det} \\ \text{or} \quad Y_{det} &= (a_t + a_{et}) + (b_t + b_{et})\, \log(Z_{det}) + \epsilon_{det}, \quad \epsilon_{det} \sim \mathcal{N}(0, \sigma_t) \qquad (19) \\ &= \mu_{det} + \epsilon_{det} \end{aligned}$$

where a_t and b_t are the "fixed" effects of depth upon measured values, a_{et} and b_{et} are the "random" effects associated with sampling event, and σ_t is the residual error standard deviation. The data were then standardised as

$$\tilde{Y}_{det} = \frac{1}{\sigma_t}(\log(Y_{det}) - \mu_{det}) \qquad (20)$$

$$\text{or} \quad \tilde{Y}_{det} = \frac{1}{\sigma_t}(Y_{det} - \mu_{det}) \qquad (21)$$

to produce data sets, $\tilde{Y}_t$, approximately normally distributed and largely independent of depth and sampling event. The PON, POC, and chlorophyll-a concentrations tended to decrease with sample depth, and the concentration-depth relationships were approximately linearised by log-transforms of the concentrations. In contrast, as ambient nitrate concentration tended to increase with depth, a log-transform of the independent variable, depth, approximately linearised the concentration-depth relationship. Equations (18) and (20) were therefore used to standardise the PON, POC, and chlorophyll-a measurements, while nitrate concentrations were standardised with eqs. (19) and (21). Each standardised data source is distributed as approximately standard normal, $\tilde{Y}_t \sim \mathcal{N}(0, 1)$. Thus, if the corresponding model outputs are transformed identically to the data, so that standardised values are used within the cost function, then no data source, sampling event, or depth should bias the fitting process, i.e., all data points have approximately equal weight.

The planktonic nitrogen size spectra were derived by Lampe et al. (2020) using data from the same research cruises that collected the nutrient and organic matter measurements. The size spectra sampling design differed, however, as the spectra were derived by aggregating measurements from 5–40 m depth over all unique sampling events with the result that measurements from each research cruise produced a single size spectra. As the other data sources, already discussed, differ both in sampling design and structure, standardising the size spectra required alternative methods. As each size spectra comprised measurements over a range of 400 unique cell sizes, the measurements were grouped into coarser size class intervals to reduce the number of size classes represented within the model (fig. 2).

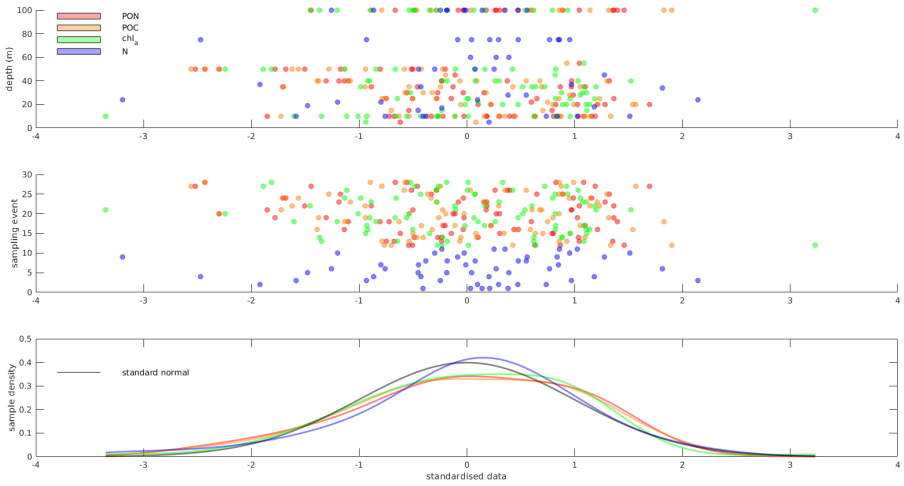

Figure 1: Data after standardisation using linear mixed effects models: fixed effect, depth; random effect, sampling event.

Ideally, each size class interval would correspond to some features in the data that distinguish it from adjacent intervals. We used a simple method to select intervals: (1) stationary points were identified in the annually averaged size spectra; (2) interval edges were set equal to the stationary points and outer limits of the size spectra; (3) the set of stationary points was filtered by successively removing the narrowest size class interval until 2 intervals remain; (4) after examining plots we selected the size class intervals from among all of the filtered sets. We selected $n = 7$ size class intervals that were fairly evenly spaced over the log-scale size range. The size class intervals each contained between $i_n = 37$–72 individual measurements. The measurements within each interval were standardised by assuming that they were log-normally distributed, i.e., by first $\log_e$-transforming the measurements, then subtracting the mean and scaling by the standard deviation (fig. 3). Thus, if $\boldsymbol{Y}_{i_p}$ denotes the vector of measurements within size class i_p then the standardised size spectra are

$$\tilde{\boldsymbol{Y}}_{i_p} = \frac{1}{\sigma_{i_p}}(\ln(\boldsymbol{Y}_{i_p}) - \mu_{i_p}) \tag{22}$$

where μ_{i_p} and σ_{i_p} are the within-size class means and standard deviations. As the size spectra data clearly cannot be well represented by common statistical distributions (fig. 2), the 'standardised' data did not approximate the shape of a standard normal distribution. Instead, the standardised data tended to be distributed with two peaks straddling zero, although the distribution shapes varied between size classes (fig. 3).

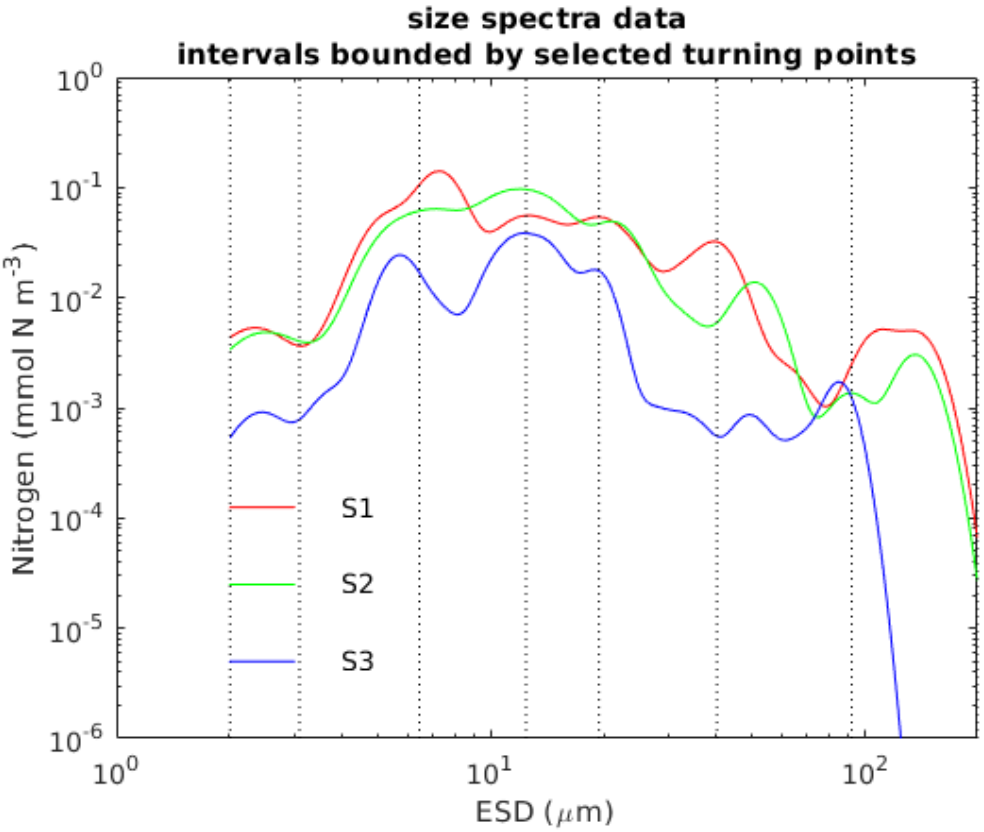

Figure 2: Nitrogen size spectra measurements for 3 seasons: 2016, 2017 and 2018. Dashed lines indicate selected size class intervals.

The cost function returns values representing the magnitude of the discrepancy between the data and the modelled values returned from a specific choice of parameters. The cost function calculates the sum of squared errors between standardised model output and data, scales these sums by the number of measurements for each data type, then sums over data types to produce a single value, $\mathcal{C}$, representing overall model misfit to data.

$$\begin{aligned}\mathcal{C} = \frac{1}{e_n}\sum_{e=1}^{e_n}\Bigg(&\frac{1}{z_n}\sum_{z=1}^{z_n}\left(\tilde{N}_z^{\text{obs}} - \tilde{N}_z\right)^2 + \frac{1}{z_n}\sum_{z=1}^{z_n}\left(\tilde{M}_{\text{POM}\,z\,C}^{\text{obs}} - \tilde{M}_{\text{POM}\,z\,C}\right)^2 \\ &+ \frac{1}{z_n}\sum_{z=1}^{z_n}\left(\tilde{M}_{\text{POM}\,z\,N}^{\text{obs}} - \tilde{M}_{\text{POM}\,z\,N}\right)^2 + \frac{1}{z_n}\sum_{z=1}^{z_n}\left(\tilde{B}_{i_p\,z\,\text{Chl}}^{\text{obs}} - \tilde{B}_{i_p\,z\,\text{Chl}}\right)^2\Bigg) \\ &+ \frac{1}{n}\sum_{i_p=1}^{n}\left(\sum_{i_n=1}^{i_n^{\max}}\left(\tilde{\boldsymbol{B}}_{i_p\,\text{N}}^{\text{obs}} - \frac{1}{z_{\max}-z_{\min}+1}\sum_{z=z_{\min}}^{z_{\max}}\tilde{B}_{i_p\,z\,\text{N}}\right)\right)^2 \end{aligned} \tag{23}$$

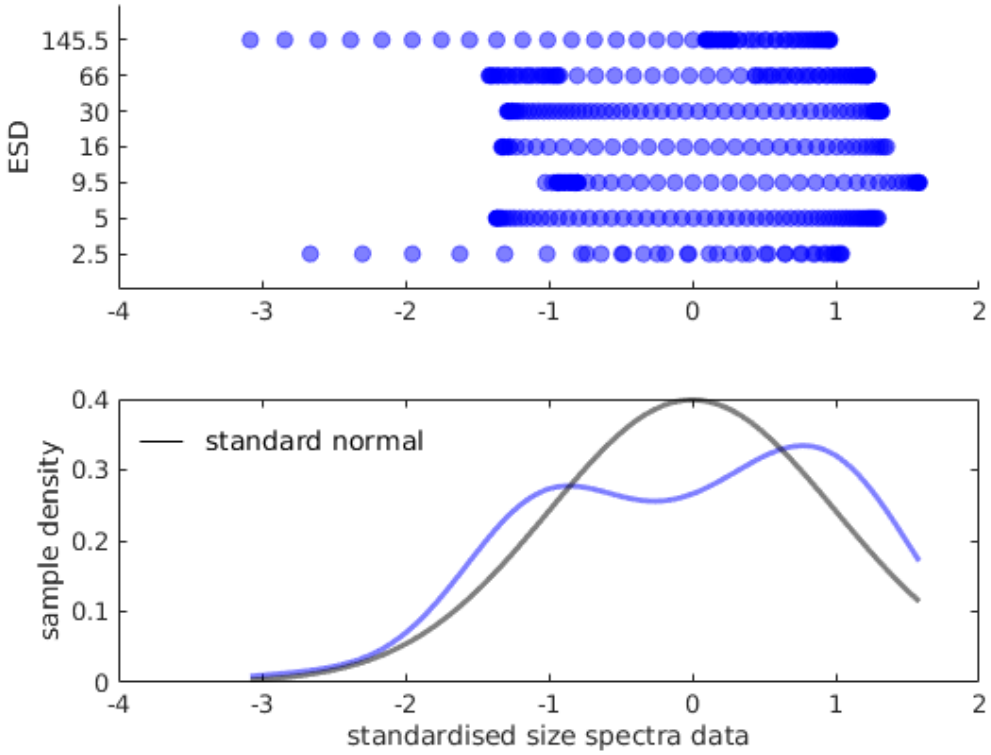

Figure 3: Nitrogen size spectra data after standardising.

Discrepancies between data and model output are reduced by numerically minimising the cost (eq. (23)) with respect to the model parameters. This produces the 'optimal' parameter set, which is input to the model to analyse the plankton dynamics.

References

Lampe V, Nöthig E-M, Schartau M. Spatio-temporal variations in community size structure of Arctic protist plankton in the Fram Strait. Frontiers (in submission)