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

Data collected from n_e sampling sites comprise depth-discrete measurements of inorganic nitrogen, dissolved and particulate organic matter, and chlorophyll a. These data are used within a cost function to tune model parameters. To limit parameter estimation bias due to depth- and sampling event-dependent variability in measured values the data should be standardised before using within a cost function. The main source of variability in these measurements is depth. Plotting measured values against depth shows non-linear relations that can be approximately linearised using log transforms. Then for each data source, Y, we can fit linear mixed models

$$\log(Y_{ij}) = (a + a_j) + (b + b_j)\, Z_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma) \tag{18}$$

$$\text{or} \quad Y_{ij} = (a + a_j) + (b + b_j)\, \log(Z_i) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma) \tag{19}$$

where i and j index depth and sampling event, Z is depth, a and b are the "fixed" effects of depth upon measured values, a_j and b_j are the "random" effects associated with sampling event, and σ is standard deviation of the residual error after linearly accounting for depth and sampling event. The data can then be

standardised as

$$\tilde{Y}_{ij} = \frac{1}{\sigma}(\log(Y_{ij}) - \mu_{ij}) \tag{20}$$

$$\text{or} \quad \tilde{Y}_{ij} = \frac{1}{\sigma}(Y_{ij} - \mu_{ij}) \tag{21}$$

where $\mu_{ij} = (a + a_j) + (b + b_j)\, Z_i$ or $\mu_{ij} = (a + a_j) + (b + b_j)\, \log(Z_i)$ respectively.

Each standardised data source is distributed as approximately standard normal, $\tilde{Y} \sim \mathcal{N}(0, 1)$. Thus, if 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 dominate the fitting process, i.e., all data points have approximately equal weight.

For each standardised data source, $\tilde{Y}_i$, and the equivalent model output identically transformed, Y_i^{mod}, a cost function can be written simply as the mean of squared errors.

$$\text{cost} = \frac{1}{N} \sum_{i=1}^{N} (\tilde{Y}_i - Y_i^{\text{mod}})^2 \tag{22}$$

There probably are better methods for scaling this data, however I think this was a reasonable first attempt that's worked fairly well. Although statistically better behaved than the raw data, the transformed data are still not quite normally distributed (see the code). I think we could get some more improvement by modelling depth-dependence of measured values using GAMs instead of LMMs...

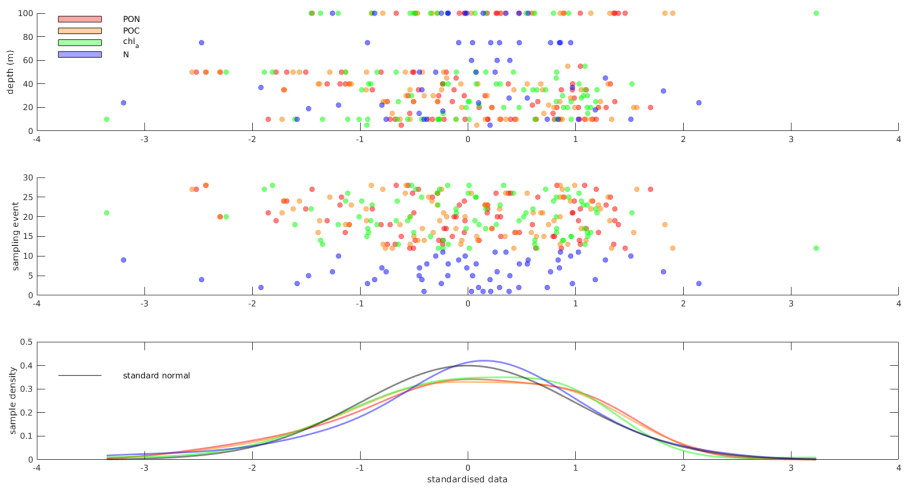

Figure 1: Data after standardisation using linear mixed effects models: fixed effect, depth; random effect, sampling event.