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 ($\mathrm{mmol\,N\,m^{-3}}$) of inorganic nitrogen, N, phytoplankton and zooplankton nitrogen biomasses, B_i, and organic matter, M_j. 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. Nitrogen fluxes are are described by a system of ODEs

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial N}{\partial z} - \sum_{i_p} V_{i_p} B_{i_p} + \sum_j r_j M_j \tag{1}$$

$$\frac{\partial B_i}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial B_i}{\partial z} + V_{i_p} B_{i_p} - G_i B_{i_z} + \lambda B_{i_z} \sum_i G_i - m B_i \tag{2}$$

$$\frac{\partial M_j}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial M_j}{\partial z} - w_j \frac{\partial M_j}{\partial z} + m \sum_i \beta_{ij} B_i + (1-\lambda) B_{i_z} \sum_i \beta_{ij} G_i - r_j M_j \tag{3}$$

where

K	Vertical diffusivity from physical model
V_{i_p}	Nitrogen uptake rate of phytoplankton size class i_p
r_j	Remineralisation rates of DON and PON
G_i	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}
Initial slope of photosynthesis-irradiance curve	α^P	0.5	$(\mu\mathrm{Ein}\,\mathrm{m}^{-2})^{-1}$
Assimilation efficiency	λ	0.7	-
Maximum grazing rate	$G^{\max}$	1	d^{-1}
Total prey half saturation	k_G	0.1	$\mathrm{mmol\ N\,m}^{-3}$
Prey refuge parameter	Λ	-1	-
Maximum Chl:N ratio	θ	3	$(\mathrm{mg\ Chl})(\mathrm{mmol\ N})^{-1}$
POM sinking speed	w_{POM}	10	$\mathrm{m\,day}^{-1}$
DOM remineralisation rate	r_{DOM}	0.02	day^{-1}
POM remineralisation rate	r_{POM}	0.04	day^{-1}

Table 1: Scalar parameters with default values.

Parameter	Notation	a	b	Unit
Minimum nitrogen quota	$Q^{\min}$	2.42×10^{-12}	0.84	$\mathrm{mmol\ N\,cell}^{-1}$
Max N quota:storage capacity	$Q^{\max}/\Delta Q$	0.62	-0.09	-
Max uptake rate:min quota	$v^{\max}/Q^{\min}$	0.71	0.13	day^{-1}
N affinity:min quota	$\alpha/Q^{\min}$	2.57	-0.14	$(\mathrm{mmol\ N\,day})^{-1}$
Max photosynthetic rate	$p^{\max}$	100	-0.26	day^{-1}

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

Nutrient uptake

Nitrogen uptake rate, V_{i_p}, is derived from a nutrient quota model using the methods of Verdy et al. (2009) as follows. Let nitrogen quota and cell density be denoted by $Q_{i_p}\,(\mathrm{mmol\,N\,cell^{-1}})$ and $D_{i_p}\,(\mathrm{cells\,m^{-3}})$, and define a quota-based model of phytoplankton dynamics as

$$\frac{dQ_{i_p}}{dt} = \gamma\, Q_{i_p}^{\mathrm{stat}}\, \nu_{i_p} - \gamma\, \mu_{i_p}\, Q_{i_p} \tag{4}$$

$$\frac{dD_{i_p}}{dt} = \gamma\, \mu_{i_p}\, D_{i_p} - \delta\, D_{i_p} \tag{5}$$

$$\frac{dN}{dt} = -\gamma \sum_{i_p} Q_{i_p}^{\mathrm{stat}}\, \nu_{i_p}\, D_{i_p} + S^N \tag{6}$$

Cellular nitrogen flux is the difference between rates of nitrogen uptake and assimilation of stored nitrogen into new cells. Rate of change of cell density is the difference between cell production and mortality rates; and rate of change of nitrogen concentration is the difference between replenishment by sources and depletion by cell uptake. Uptake and assimilation rates are both adjusted for temperature, T, according to a sensitivity, A, and relative to reference temperature, T^{ref}.

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

Nutrient uptake rate is modelled as a Michaelis-Menton function

$$\nu_{i_p} = \frac{v_{i_p}^{\max}\, \alpha_{i_p}\, N}{\alpha_{i_p}\, N + v_{i_p}^{\max}} \tag{8}$$

which is limited by a linear function, $Q^{\mathrm{stat}_{i_p}}$, of cell quota so that uptake rate declines to zero as quota approaches its maximum.

$$Q_{i_p}^{\mathrm{stat}} = \frac{Q_{i_p}^{\max} - Q_{i_p}}{Q_{i_p}^{\max} - Q_{i_p}^{\min}} = \frac{Q_{i_p}^{\max} - Q_{i_p}}{\Delta Q_{i_p}} \tag{9}$$

Nutrient assimilation rate is modelled using a Poisson function of irradiance, I, limited by a hyperbolic function of quota.

$$\mu_{i_p} = \mu_{i_p}^{\infty} \left(1 - Q_{i_p}^{\min}/Q_{i_p}\right) = p_{i_p}^{\max} \left(1 - \exp\left(\frac{-\alpha^p\, I}{p_{i_p}^{\max}}\right)\right) \left(1 - \frac{Q_{i_p}^{\min}}{Q_{i_p}}\right) \tag{10}$$

Rewrite Equations (4) to (6) in terms of nutrient concentration instead of density using the relation $B_{i_p} = Q_{i_p} D_{i_p}$.

$$\frac{dB_{i_p}}{dt} = \gamma \frac{Q_{i_p}^{\text{stat}} \nu_{i_p}}{Q_{i_p}} B_{i_p} - \delta B_{i_p} \tag{11}$$

$$\frac{dN}{dt} = -\gamma \sum_{i_p} \frac{Q_{i_p}^{\text{stat}} \nu_{i_p}}{Q_{i_p}} B_{i_p} + S^N \tag{12}$$

Equations (11) and (12) form a closed set of ODEs if we assume that cell quota is at equilibrium, $Q_{i_p} = Q_{i_p}^*$, balancing nutrient uptake and assimilation rates.

$$\left.\frac{dQ_{i_p}}{dt}\right|_{Q_{i_p}^*} = 0 \implies Q_{i_p}^{\text{stat}} \nu_{i_p} = \mu_{i_p} Q_{i_p}^* \tag{13}$$

$$\implies Q_{i_p}^* = \frac{Q_{i_p}^{\max} v_{i_p}^{\max} \alpha_{i_p} N + Q_{i_p}^{\min} \Delta Q_{i_p} \mu_{i_p}^{\infty} (\alpha_{i_p} N + v_{i_p}^{\max})}{\Delta Q_{i_p} \mu_{i_p}^{\infty} (\alpha_{i_p} N + v_{i_p}^{\max}) + v_{i_p}^{\max} \alpha_{i_p} N} \tag{14}$$

Finally, the nitrogen uptake rate used in eqs. (1) and (2) is found from eqs. (8), (9), (11) and (14)

$$V_{i_p} = \gamma \frac{Q_{i_p}^{\text{stat}} \nu_{i_p}}{Q_{i_p}^*} = \gamma \frac{\mu_{i_p}^{\max} \tilde{\alpha}_{i_p} N}{\tilde{\alpha}_{i_p} N + \mu_{i_p}^{\max}} \tag{15}$$

where the maximum growth rate, $\mu_{i_p}^{\max}$, and affinity for growth, $\tilde{\alpha}_{i_p}$, are

$$\mu_{i_p}^{\max} = \frac{\mu_{i_p}^{\infty} \left(\frac{v^{\max}}{Q^{\min}}\right)_{i_p}}{\left(\frac{Q^{\max}}{\Delta Q}\right)_{i_p} \left(\frac{v^{\max}}{Q^{\min}}\right)_{i_p} + \mu_{i_p}^{\infty}} \tag{16}$$

$$\tilde{\alpha}_{i_p} = \left(\frac{\alpha}{Q^{\min}}\right)_{i_p} \tag{17}$$

Predation

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

$$G_i = \gamma G^{\max} \frac{F}{k_G + F} (1 - \exp(\Lambda F)) \Phi_i. \tag{18}$$

The $G^{\max} \frac{F}{k_G+F}$ term is a Michaelis-Menton function where $F = \sum_i B_i$ is total available prey nitrogen. 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, $\Phi_i = \frac{B_i^2}{\sum_i B_i^2}$, to target grazing on the most abundant prey class.

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

$$M_j^{\text{mort}} = m \sum_i \beta_{ij}\, B_i \tag{19}$$

where $\beta_{ij} = [\beta_i, 1 - \beta_i]$ allocates organic matter to DON or PON. As the zooplankton assimilation efficiency is λ, organic matter produced from messy feeding is

$$M_j^{\text{mess}} = (1 - \lambda) B_{i_z} \sum_i \beta_{ij}\, G_i. \tag{20}$$

The $w_j \frac{dM_j}{dz}$ term in eq. (3) models sinking of organic matter at speed w_j: only PON sinks. Organic matter is lost through remineralisation at rates r_j.

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

$$\log(Y_{ij}) = (a + a_j) + (b + b_j)\, Z_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma) \tag{21}$$

$$\text{or} \quad Y_{ij} = (a + a_j) + (b + b_j)\, \log(Z_i) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma) \tag{22}$$

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

$$\tilde{Y}_{ij} = \frac{1}{\sigma}(\log(Y_{ij}) - \mu_{ij}) \tag{23}$$

$$\text{or} \quad \tilde{Y}_{ij} = \frac{1}{\sigma}(Y_{ij} - \mu_{ij}) \tag{24}$$

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.

$$\mathrm{cost} = \frac{1}{N} \sum_{i=1}^{N} (\tilde{Y}_i - Y_i^{\mathrm{mod}})^2 \tag{25}$$

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...

References

Verdy A, Follows M, Flierl G. Optimal phytoplankton cell size in an allometric model. MEPS 379: 1–12, 2009. doi:10.3354/meps07909