# Supplement for "The unicellular NUM v.0.91: A trait-based plankton model evaluated in two contrasting biogeographic provinces."

**Table of Contents**

## S 1.    The Nutrient-Unicellular-Multicellular model

The Nutrient-Unicellular-Multicellular (NUM) model is a modular framework comprising four separate modules: (i) the Generalist module for unicellular organisms, (ii) a specialized module for unicellular diatoms that require silica, (iii) a module for multicellular organisms, exemplified by the copepod, and (iv) a module for handling particulate organic matter (POM). The NUM model utilizes dissolved organic carbon (DOC) and a non-specific nutrient (N). Silica is included when the diatom module is employed.

In this article, we describe and evaluate the Generalist (i) and POM (iv) modules of the NUM model. The model formulations are based on Andersen and Visser (2023), who provide an in-depth explanation of the underlying motivations, which we will not reiterate here. We highly recommend interested readers refer to their comprehensive explanation.

The following sections detail the formulations for the unicellular generalist, nutrient, and DOC interactions, along with the POM module and the water-column implementation within the NUM framework. All parameters and units are listed in Table 1 of the main text.

## S 1.1.  Generalists

The unicellular module of the NUM model is described as a series of $N$ unicellular organisms differing only in their size. The size classes are logarithmic distributed with geometric mean mass $m_i$ for size group $i$. The energy growth rate (day$^{-1}$) of the generalist biomass $B$ in size group $i$ is described by the following equation:

$$\frac{dB_i}{dt} = g(m_i)B_i - \mu(m_i)B_i, \tag{1}$$

In this equation, $g(m_i)$ represents the synthesis rate of carbon and nitrogen (day$^{-1}$), and $\mu(m_i)$ represents the total mortality losses (day$^{-1}$). The specific forms of $g(m_i)$ and $\mu(m_i)$ are detailed below.

### S 1.1.1. Synthesis rate, $g(m_i)$

The synthesis rate $g$ of size group $i$ is described by a type II functional response (Holling, 1959):

$$g(m_i) = j_{\mathrm{max.i}} \frac{J_{\mathrm{net.}i}}{J_{\mathrm{net.}i} + J_{\mathrm{max.}i}}, \tag{2}$$

Here, the maximum synthesis rate $j_{\mathrm{max}}$ is given by:

$$j_{\mathrm{max.}i} = \alpha_{\mathrm{max}}(1-\nu)m_i\psi, \tag{3}$$

where $\alpha_{max}$ is the maximum synthesis coefficient (day$^{-1}$) and $\nu$ is the fraction of the cell used by cell membrane and wall, defined as $\nu = 3(\delta/r)$ with $\delta$ being the thickness of the cell wall in nm. $\psi$ represents the temperature adjustment factor with a Q10 value of 2 (Eppley, 1971).

$J_{\mathrm{net.}i}$ is the minimum of the combined net uptake of carbon and nutrients from different sources, determined by Liebig's law of the minimum:

$$J_{\mathrm{net.}i} = \min\{j_{\mathrm{Cnet.}i}, j_{\mathrm{Nnet.}i}\} \tag{4}$$

The unicellular generalist can take up carbon as dissolved organic carbon (DOC) through osmotrophy ($j_{\mathrm{DOC}}$), phototrophy ($j_{\mathrm{L}}$), and phagotrophy ($j_{\mathrm{F}}$), while losing carbon through respiration ($j_{\mathrm{R}}$) and passive losses ($j_{\mathrm{passive}}$)

$$j_{\mathrm{Cnet.}i} = j_{\mathrm{DOC}} + j_{\mathrm{L}} + j_{\mathrm{F}} - j_{\mathrm{R}} - j_{\mathrm{passive}} \tag{5}$$

Similarly, the net uptake of nutrients for the unicellular generalist involves diffusive uptake through osmotrophy ($j_{\mathrm{N}}$) and phagotrophic uptake ($j_F$), with losses occurring passively ($j_{\mathrm{passive}}$)

$$j_{\mathrm{Nnet.}i} = j_{\mathrm{N}} + j_F - j_{\mathrm{passive}} \tag{6}$$

The uptake rates from the resources DOC ($j_{\text{DOC}.i}$), light ($j_{\text{L}.i}$), and nutrients ($j_{\text{N}.i}$) depend on the mass-specific affinity for the given resource $a_x$ (volume/day/gC), the concentration of the resource $X$ ($X$/volume), and the Carbon: $X$ ratio $\rho_{\text{C:X}}$ (gC/g$X$)

$$j_{\text{X}.i} = a_{x.i} X \rho_{\text{C:X}} \tag{7}$$

In the model $a_x$ take the form of $a_D$, the mass-specific affinity for the diffusive uptake of carbon and nutrient, $a_L$ the mass-specific affinity for carbon uptake through photosynthesis, and $a_F$ the mass-specific affinity for phagotrophic uptake of food. The mass-specific affinities are central to the NUM model, as they account for the differences among generalist organisms that vary only in size. The affinities are thoroughly discussed in Andersen and Visser (2023) with a comprehensive presentation of model's underlying principles. Here only in short, we present the mass-specific affinity functions.

The mass-specific affinity for the diffusive uptake of carbon and nutrients $a_{\text{D}}$ is described as:

$$a_{\text{D}} = \alpha_D r^{-2} \frac{1}{1 + \dfrac{r^{-2}}{r_D^*}} \psi, \tag{8}$$

where $\alpha_D$ is the diffusive affinity coefficient, $r_D^*$ is the diffusive affinity crossover and $r$ is the radius of the cell.

The mass-specific affinity for carbon uptake through photosynthesis $a_{\text{L}}$ is give as:

$$a_{\text{L}} = \varepsilon_L \frac{\alpha_L}{r} \left(1 - e^{-\frac{r}{r_L^*}}\right) m(1 - \nu), \tag{9}$$

Where $\alpha_L$ is the light affinity coefficient, $r_L^*$ is the light affinity crossover and $\varepsilon_L$ is the light uptake efficiency.

The uptake rate of phagotrophy has a constant mass-specific affinity, corresponding to the phagotrophic clearance rate. However, the actual food consumption is limited by assimilation, and the phagotrophic uptake rate is thus given by:

$$j_{\text{F}} = \varepsilon_F \frac{c_F}{r} \frac{a_F F}{a_F F + \dfrac{c_F}{r}}, \tag{10}$$

Where $c_F$ is the maximum phagotrophic coefficient, $a_F$ is the phagotrophic clearance rate, $\varepsilon_F$ is the phagotrophic efficiency, and $F$ is the amount of available food. The available food $F_i$ is the sum of the prey in each size group $j$:

$$F_i = \sum_j \varphi_{ij} B_i \tag{11}$$

$\varphi_{ij}$ is the size preference for predation. Predation occur when larger cells predate on smaller cells. This interaction is described as a log-normal size function that, when integrated across all size groups, has the detailed formulation:

$$\varphi_{ij} = \frac{\sqrt{\Delta}}{(\Delta - 1)\log(\Delta)} \left[ \left( \frac{1}{2}s \left( e^{-\frac{\log^2\left(\frac{\Delta z}{\beta}\right)}{s}} + e^{-\frac{\log^2\left(\frac{\Delta \beta}{z}\right)}{s}} - 2e^{-\frac{\log^2\left(\frac{z}{\beta}\right)}{s}} \right) \right. \right.$$

$$- \frac{1}{2}\sqrt{\pi}\sqrt{s} \left( \log\left(\frac{\Delta z}{\beta}\right) erf\left(\frac{\log(\beta) - \log(\Delta z)}{\sqrt{s}}\right) \right.$$

$$\left. \left. + \log\left(\frac{\beta \Delta}{z}\right) erf\left(\frac{\log(z) - \log(\Delta \beta)}{\sqrt{s}}\right) + \log\left(\frac{z}{\beta}\right) erf\left(\frac{\log\left(\frac{z}{\beta}\right)}{\sqrt{s}}\right) \right) \right) \right]$$

where $s = 2\sigma^2$, $\sigma$ is the predator-prey width ratio, $\beta$ is the predator-prey mass ratio, $z = \frac{m_i}{m_j}$, and $\Delta = m_i^+ / m_i^-$.

In addition to these gains, $J_{net}$ also includes losses from respiration and passive leakages. Respiration is modeled as a constant fraction $\alpha_R$ of maximum synthesis:

$$j_R = \alpha_R \alpha_{max} \psi \tag{13}$$

Finally, passive leakage of nitrogen and carbon from the cell is defined as:

$$j_{passive} = \frac{c_{passive}}{r} m, \tag{14}$$

where $c_{passive}$ is the passive loss coefficient.

### S 1.1.2. Mortality losses, $\mu(m_i)$

The mortality losses for the unicellular generalist $\mu$ ($d^{-1}$) arise from three different processes: (i) background mortality ($\mu_v$), (ii) predation mortality from unicellular organisms ($\mu_p$), and (iii) predation from higher trophic levels ($\mu_{htl}$):

$$\mu(m_i) = \mu_{v.i} + \mu_{p.i} + \mu_{htl.i} \tag{15}$$

The background mortality ($d^{-1}$) is modelled as viral lysis, assumed to be proportional to the biomass and dependent on the viral lysis mortality coefficient $\mu_{v0}$:

$$\mu_{v.i} = \frac{\mu_{v0}}{\log\left(\frac{m_i^+}{m_{i-1}^+}\right)} B_i \tag{16}$$

The predation mortality represents internal phagotrophy within the unicellular module, where larger cells consume smaller cells. It is defined by:

$$\mu_{p.i} = \sum_i \frac{\tilde{j}_{F.i}}{\epsilon_F} \frac{\varphi_{ij}}{F_i} B_i \tag{17}$$

Where $\varphi_{ij}$ is calculated as in equation (12), and $\tilde{j}_{F.i}$ is the down-regulated phagotrophic uptake rate after leakage of surplus nutrients:

$$\tilde{\jmath}_{F.i} = max\{0, j_{F.i} - (J_{\text{net}.i} - g_i)\} \tag{18}$$

The higher trophic level mortality is modelled as:

$$\mu_{\text{htl}} = \mu_{\text{htl}.0} \frac{1}{1 + \left(\frac{m_i}{m_{htl}}\right)^{-2}} \tag{19}$$

Where $\mu_{\text{htl}.0}$ defines the size of the higher trophic level mortality and $m_{htl}$ defines the lower size limit of the higher trophic level mortality.

## S 1.2. Nutrients and DOC

Nutrients and DOC are updated several times depending on the number of modules used. Within the unicellular module, nutrients and DOC are utilized by the generalist and returned into the environment through several processes: passive losses ($j_{\text{passive}}$), nutrient surplus leaking from the cell ($j_{\text{Nliebig}}$, $j_{\text{Cliebig}}$) feeding losses ($j_{\text{feeding}}$), along with the fraction of background and higher-trophic level mortality that it not transferred into POM (controlled by the parameters γ$_2$ and γ$_{htl}$).

The nutrient dynamics is described by:

$$\frac{dN}{dt} = \sum_i \left( \frac{-j_{N.i} + j_{\text{passive}.i} + j_{\text{Nliebig}.i} + j_{\text{feeding}.i}}{m_i} + \mu_{v.i}\gamma_2 + \mu_{\text{htl}}\gamma_{htl} \right) \frac{B_i}{\rho_{\text{C:N}}} \tag{20}$$

The DOC dynamics are described by:

$$\frac{dDOC}{dt} = \sum_i \left( \frac{-j_{\text{DOC}.i} + j_{\text{passive}.i} + j_{\text{Cliebig}.i} + j_{\text{feeding}.i}\gamma_F + j_{\text{photouptake}.i}}{m_i} + \mu_{v.i}\gamma_2 \right) B_i \tag{21}$$

Here, the surplus of nutrient and carbon is calculated as:

$$j_{\text{Nliebig}.i} = \max\{0, \tilde{\jmath}_{N.\text{net}.i} - g_i\} \tag{22}$$

$$j_{\text{Cliebig}.i} = \max\{0, \tilde{\jmath}_{C.\text{net}.i} - g_i\} \tag{23}$$

The downregulated uptake of carbon and nutrients are described as:

$$\tilde{\jmath}_{N\text{net}.i} = j_N + \tilde{\jmath}_F - j_{\text{passive}} \tag{24}$$

$$\tilde{\jmath}_{C.\text{net}.i} = \tilde{\jmath}_{L.i} + j_{\text{DOC}} + \tilde{\jmath}_F - j_R - j_{\text{passive}} \tag{25}$$

and

$$\tilde{\jmath}_{L.i} = j_L - \max\{0, \min\{(j_{C.\text{net}} - (j_F - \tilde{\jmath}_F) - g), j_L\}\} \tag{26}$$

Finally, feeding losses and photo-uptake are described as:

$$j_{\text{feeding}.i} = \frac{1 - \epsilon_F}{\epsilon_F} \tilde{\jmath}_{F.i} \tag{27}$$

$$j_{\text{photouptake}.i} = \frac{1 - \epsilon_L}{\epsilon_L} \tilde{j}_{\text{L}.i} \tag{28}$$

## S 1.3. POM

The NUM model handles particulate organic matter (POM) similarly to how it manages unicellular organisms. POM is divided into $M$ logarithmic size classes, each with a geometric mean mass $m_k$ for POM group $k$. The sources of POM include: (i) Mortality from higher trophic levels, (ii) feeding losses, and (iii) background mortality, while losses occur through remineralization and predation by larger cells. The assignment of background mortality to specific POM classes is described by the matrix θ. Viral lysis products are directed into the POM class closest in size to the original generalist size from which they originated, with a constraint preventing them from exceeding the original generalist size. The matrix θ contains 0s and 1s, indicating where the biomass from different sources is allocated within the POM size classes. The dynamics of POM are described by the equation:

$$\frac{\text{dPOM}_k}{\text{d}t} = \sum_i (1 - \gamma_2)\,\mu_{\text{v}.i}\theta_{k.i}B_i + \sum_i (1 - \gamma_{htl})\mu_{\text{htl}}\Delta_k \frac{B_i}{\rho_{\text{C:N}}} - \gamma_{\text{POM}.k}P_k - \mu_{\text{p}.k}P_k \tag{29}$$

Here, $\Delta_k = \begin{cases} 0, & k < M \\ x, & k = M \end{cases}$ denotes that POM from higher-trophic level mortality is transported into the largest POM size class. $\mu_{\text{p}.k}$ describes feeding of large unicellular organisms on POM and is calculated as in equation (17). Remineralization of POM is described as

$$\gamma_{\text{POM}.k} = w_k a\,\psi \tag{30}$$

Where $w_k$ is the mass-specific sinking velocity described by $w_k = v_1 m_k^{v_2}$ in meters/day and $a$ is the inverse solubilization length scale with units of m$^{-1}$.

Nutrients are updated as POM is remineralized back into nitrogen (N) and dissolved organic carbon (DOC):

$$\frac{dN}{dt} = \left(\frac{dN}{dt}\right)_{\text{generalist}} + \sum_k \frac{\gamma_{\text{POM}}P_k}{\rho_{\text{C:N}}} \tag{31}$$

$$\frac{dDOC}{dt} = \left(\frac{dDOC}{dt}\right)_{\text{generalist}} + \sum_k \gamma_{\text{POM}}P_k \tag{32}$$

## S 1.4. Water-column routine

The water-column dynamics (mixing and sinking) is described with a transport matrix:

$$\boldsymbol{u}_{t+1} = (\boldsymbol{T}_t + \boldsymbol{S})\boldsymbol{u}_t, \tag{33}$$

Where $\boldsymbol{u}_t$ is the state vector at time $t$ comprising nutrients, biomasses, and POM state variables, $\boldsymbol{T}_t$ is the transport matrix, and $\boldsymbol{S}$ is the sinking matrix. The transport matrix represents vertical mixing and is

extracted as the implicit matrix from the global 1° transport matrix MITgcm_ECCO (Stammer et al., 2004) at the two study sites on a monthly basis. The sinking matrix **S** represents sinking of POM as described with a first-order implicit down-wind scheme. The time step of transport and sinking is 0.5 days. In between each transport time step the ecological and biogeochemical model is integrated with a Euler scheme with a time step of 0.1 days. The boundary at the surface is closed for all state variables. The boundary at the bottom is fixed to the initial concentrations for nutrient state variables, closed for unicellular state variables, and open for sinking POM.

The light availability in the water column is dependent on the light attenuation ($k_w$ and $k_{POM}$), which result from shading and scattering by dissolved and particulate organic matter in the water column. $k_w$ is calculated based on observations at CEE. We have fitted an exponential function to observed irradiance as function of depth using the particulate organic matter concentration for each station at each measuring day in the CCE-LTER program. This has resulted in a $k_{POM}$ value of $3 \times 10^{-5}$ ($m^2$ mg $C^{-1}$) used here.

## S 2.    Inter-annual statistical variations in the dataset from CCE and Station ALOHA

In the article we compare the model result to the mean size spectrum for California Current Ecosystem (CCE) and Station ALOHA in the years 2004-2011. To evaluate the model results we compare the deviations from the mean observations with the inter-annual variability in Root-Mean-Squared-difference ($RMSd_{iao}$), Standard Deviation ($STD_{iao}$) and Correlation coefficient ($COR_{iao}$) given in Table S1.

*Table S2.1: The Root-Mean-Squared-difference, Standard Deviation and Correlation between annual observation versus mean of year 2004-2001 for CCE ($RMSd_{iao}$, $STD_{iao}$, and $COR_{iao}$)*

| CCE | $B_{T-pico}$ | $B_{T-nano}$ | $B_{A-pico}$ | $B_{A-nano}$ | $B_{A-micro}$ |
|---|---|---|---|---|---|
| $RMSd_{iao}$ | 0.4 | 0.3 | 2.7 | 0.4 | 1.0 |
| $STD_{iao}$ | 2 | 1.4 | 2.1 | 1.4 | 1.4 |
| $COR_{iao}$ | 0.2 | 0.7 | 0 | 0.7 | 0.96 |
| Station ALOHA | $B_{T-pico}$ | $B_{T-nano}$ | $B_{A-pico}$ | $B_{A-nano}$ | $B_{A-micro}$ |
| $RMSd_{iao}$ | 0.04 | 0.5 | 0.2 | 0.4 | 1.0 |
| $STD_{iao}$ | 1.2 | 1.6 | 1.4 | 1.5 | 2.6 |
| $COR_{iao}$ | 0.97 | 0.9 | 0.96 | 0.9 | 0.7 |

## S 3.    Statistics for best model solutions

Model Root-Mean-Squared-difference (RMSd$_{m-o}$), and Correlation coefficient (COR$_{m-o}$) for the best model fit to observations.

*Table S2: The Root-Mean-Squared-difference, and Correlation between model result and mean of year 2004-2001 for CCE (RMSd$_{m-o}$, and COR$_{m-o}$) for the best parameter combinations for CCE. Compare with Table S1 for boundaries. Values highlighted with grey are not within the inter-annual variability of the observations.*

| | | CCE | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | $B_{T-pico}$ | $B_{T-nano}$ | $B_{A-pico}$ | $B_{A-nano}$ | $B_{A-micro}$ |
| | simulation nr. | RMSd$_{m-o}$ | | | | |
| RMSd$_{m-o}$ | 1703 | 0.64 | 0.60 | 0.08 | 0.25 | 0.40 |
| | 5801 | 0.48 | 0.71 | 0.05 | 0.39 | 0.51 |
| | 39986 | 0.57 | 0.45 | 0.16 | 0.19 | 0.23 |
| | 41402 | 0.53 | 0.37 | 0.13 | 0.22 | 0.67 |
| | 49833 | 0.63 | 0.40 | 0.19 | 0.17 | 0.35 |
| | 91224 | 0.54 | 0.47 | 0.08 | 0.22 | 0.27 |
| | 99032 | 0.55 | 0.50 | 0.10 | 0.32 | 0.67 |
| | | COR$_{m-o}$ | | | | |
| COR$_{m-o}$ | 1703 | 0.86 | 0.96 | 0.98 | 0.98 | 0.89 |
| | 5801 | 0.83 | 0.98 | 0.98 | 0.99 | 0.90 |
| | 39986 | 0.98 | 0.94 | 1.00 | 0.96 | 0.95 |
| | 41402 | 0.93 | 0.96 | 1.00 | 0.98 | 0.79 |
| | 49833 | 0.93 | 0.95 | 1.00 | 0.97 | 0.91 |
| | 91224 | 0.95 | 0.94 | 0.99 | 0.96 | 0.92 |
| | 99032 | 0.93 | 0.97 | 0.98 | 0.99 | 0.80 |

*Table S3: The Root-Mean-Squared-difference, and Correlation between model result and mean of year 2004-2001 for CCE (RMSd$_{m-o}$, and COR$_{m-o}$) for the best parameter combinations for CCE. Compare with Table S1 for boundaries. Values highlighted with grey are not within the inter-annual variability of the observations.*

| | | Station ALOHA | | | | |
|---|---|---|---|---|---|---|
| | | B$_{T-pico}$ | B$_{T-nano}$ | B$_{A-pico}$ | B$_{A-nano}$ | B$_{A-micro}$ |
| | simulation nr. | RMSd$_{m-o}$ | | | | |
| RMSd$_{m-o}$ | 29534 | 0.17 | 0.56 | 0.09 | 0.12 | 0.88 |
| | 37705 | 0.31 | 1.08 | 0.29 | 0.36 | 1.69 |
| | 69585 | 0.12 | 0.67 | 0.13 | 0.20 | 1.29 |
| | 74294 | 0.41 | 1.18 | 0.24 | 0.18 | 2.77 |
| | | COR$_{m-o}$ | | | | |
| COR$_{m-o}$ | 29534 | 0.97 | 0.97 | 0.99 | 0.99 | 0.78 |
| | 37705 | 0.98 | 0.90 | 0.97 | 0.95 | 0.72 |
| | 69585 | 0.96 | 0.96 | 0.97 | 0.99 | 0.71 |
| | 74294 | 1.00 | 0.95 | 0.98 | 0.99 | 0.81 |

## S 4.    Local sensitivity analysis for CCE

The aim of the local sensitivity assessment is to evaluate each parameter's effect on the model result. The range of each free parameter is based on the range defined by the solutions with optimal fit (Fig. 6). Several of the parameters result in a threshold sensitivity (systems bifurcation point) where the model solution changes abruptly. We note as an example the threshold sensitivity related to the phagotrophic assimilation rate ($\varepsilon_F$), where there is an abrupt increase in microplankton at $\varepsilon_F$ ~0.25 (Fig. S1a). $\varepsilon_F$ has a highly non-linier effect on the ecosystem expressed in terms of light harvesting and food consumption (Fig. S1b,c). As $\varepsilon_F$ decrease, the bifurcation in the ecosystem is associated with an abrupt reduction of food consumption in the lower half of the phytoplankton mass (size) spectrum, resulting in lower overall biomass, a reduction of light harvesting as a result of more unassimilated food is shunted into DOC and shading. While being extremely interesting, the detailed analysis of such bifurcation points is beyond the current scope and remains a prospect for future analyses.
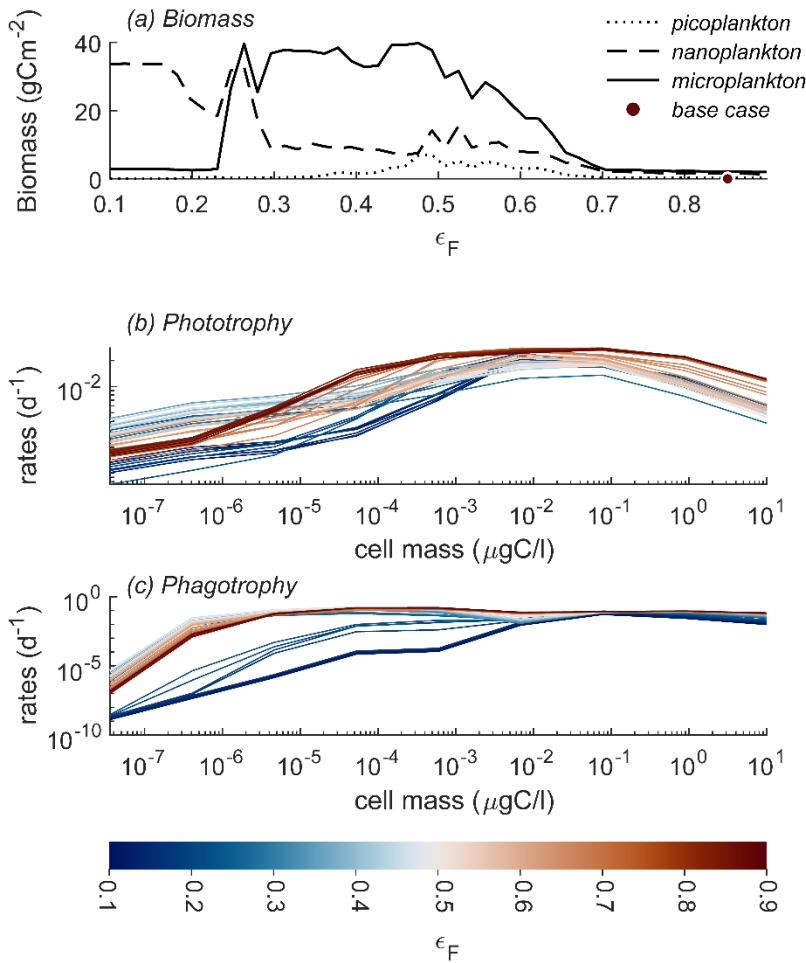
*Figure S1: Model threshold sensitivity related to the phagotrophic assimilation rate ($\varepsilon_F$) with an abrupt decrease in microplankton as $\varepsilon_F$ decrease below ~0.25 (a). Note how this bifurcation point behavior is associated with light harvesting (b); light harvesting initially decreases for nearly all cell mass as $\varepsilon_F$ decrease, but suddenly at ~0.25 results in a significant reduction mainly in the lower half of the mass range. Phagotrophic food consumption also suddenly decreases at ~0.25 (c). Model parameter combination is Test 41402, with local sensitivity plot shown in Fig. S2*

Overall, the random sampling of the restricted parameter span, as well as local sensitivity study show that most parameters are highly coupled in term of ecosystem sensitivity, where the effect of individual parameters are intertwined and result in a highly non-linier system. An illustration of this non-linearity can be seen by comparing the local sensitivity of two of our nine optimal fits. The two sets of local sensitivity analyses result in very different estimates of the parameter sensitivity (Fig. S2). One model, with its initial optimal parameter set, yielded nearly equally sensitive to almost all parameters with only $a_F$, $\sigma$ and $m_{HTL}$ standing out in RMSd (Fig. S2a). In contrast, $\varepsilon_F$ is the absolute most important parameter in the other model (Fig. S2b).
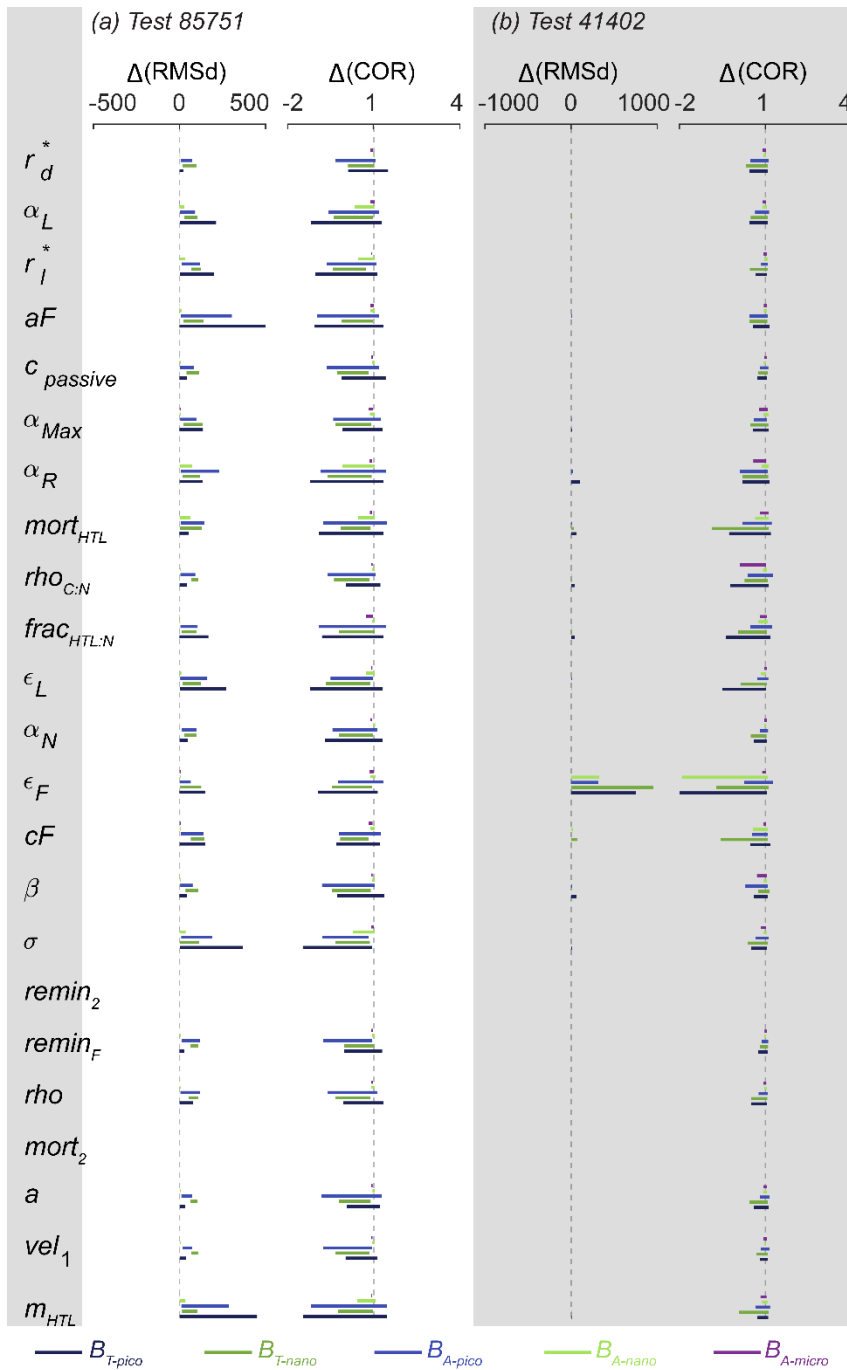
Figure S2: Local sensitivity, for two optimal model results, where only one parameter is varied at the time. The variation is within the restricted parameter span based on the parameter range for the nine statistically optimal parameter combinations for CCE. Note how model test 85751 is sensitive to change in most parameters (a) whereas model test 41402 (b) is strongly sensitive to εF. Abbreviations as in Fig. 2 and parameters in Table 1.

## S 5.    Setting up Sobol's sensitivity assessment simulations and calculating sensitivity index

The following is a "dummy" script for setting up a Sobols Sensitivity Analysis and calculating the Sobols sensitivity index.

**Step 1: Run the model with three sets of parameters:**

% a. "Original": 20,000 simulations with predefined Latin Hypercube Sampling of random parameters in a 23-parameter space.

% b. "Set 1": 460,000 simulations. For each of the 20,000 original simulations, run 23 simulations where each parameter is held at its "original" value while changing all other parameters within the same parameter space as the original collection.

% c. "Set 2": 460,000 simulations where each parameter is varied individually while the other parameters are held constant at their "original" values.

**Step 2: Sorting data:**

% a. The result of simulations in 1a becomes a column vector from 1 to 20,000, called "y0."

% b. The result of simulations in 1b becomes a matrix that is 23 times 20,000 in size, where row 1 corresponds to simulations where parameter 1 is held at its "original" value while all others are varied (and so on for all 23 rows/parameters). This matrix is called "y1."

% c. The result of simulations in 1c becomes a matrix that is 23 times 20,000 in size, where row 1 corresponds to simulations where only parameter 1 is varied while others are held at their "original" values (and so on for all 23 rows/parameters). This matrix is called "y2."

**Step 3: Calculate f0 and D using the formula:**

f0 = sum(y0)/20000;
D = sum(y0.^2)/20000;
D = D - f0.^2;

**Step 4: Calculate Di and Di_tot using the formula:**

Di = ones(23, 1) * D;
Ditot = zeros(23, 1);
for i = 1:20000
  for j = 1:23
    Di(j) = Di(j) - (y0(i) - y1(j, i)).^2 / (2 * 20000);
    Ditot(j) = Ditot(j) + (y0(i) - y2(j, i)).^2 / (2 * 20000);
  end
end

**Step 5: Calculate the First-order (S1) and Total (St) effect sensitivity indices:**

S1 = Di./D;
St = Ditot./D;

## S 6.    Checking stability of Total Sobol's sensitivity index

To test if 20,000 random simulations is enough to assess the Sobol's Total index (STi) we evaluated the evolution of STi as a function of the number of simulations. Fig. S3 show that the STi for Root-Mean-Square difference (RMSd) for $AC_{pico}$ is stable for most of the parameters after approximately 7,000 random simulations.
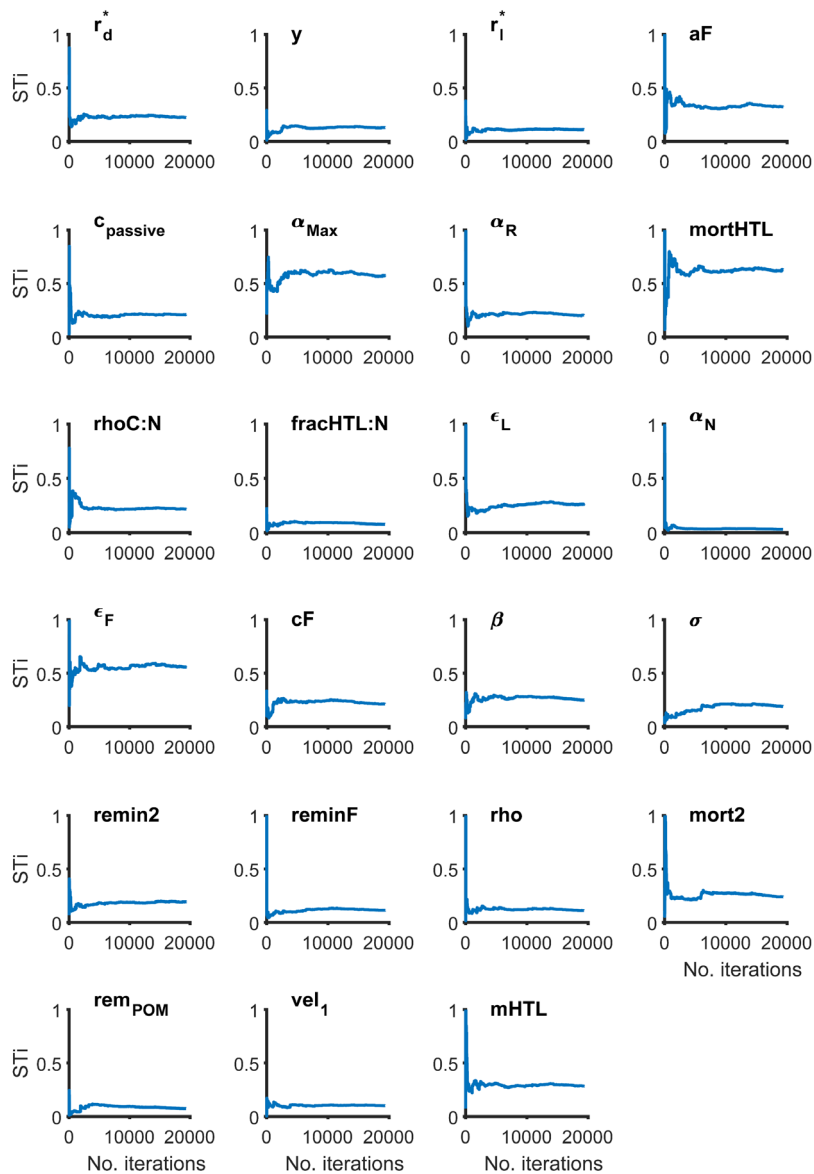
*Figure S3: Sobol's Total index (STi) calculated on RMSd for $AC_{pico}$ as a function of the number of simulations, for each of the 23 parameters. Most of the parameters show large variations with few simulations but the STi values are stable above approximately 7,000 simulations.*

The manuscript includes the global parameter sensitivity ranked based on sensitivity index calculated for RMSd but not correlation which is added in Fig. S4. The result for correlations supports the result for RMSd showing a model sensitivity towards predation and synthesis.
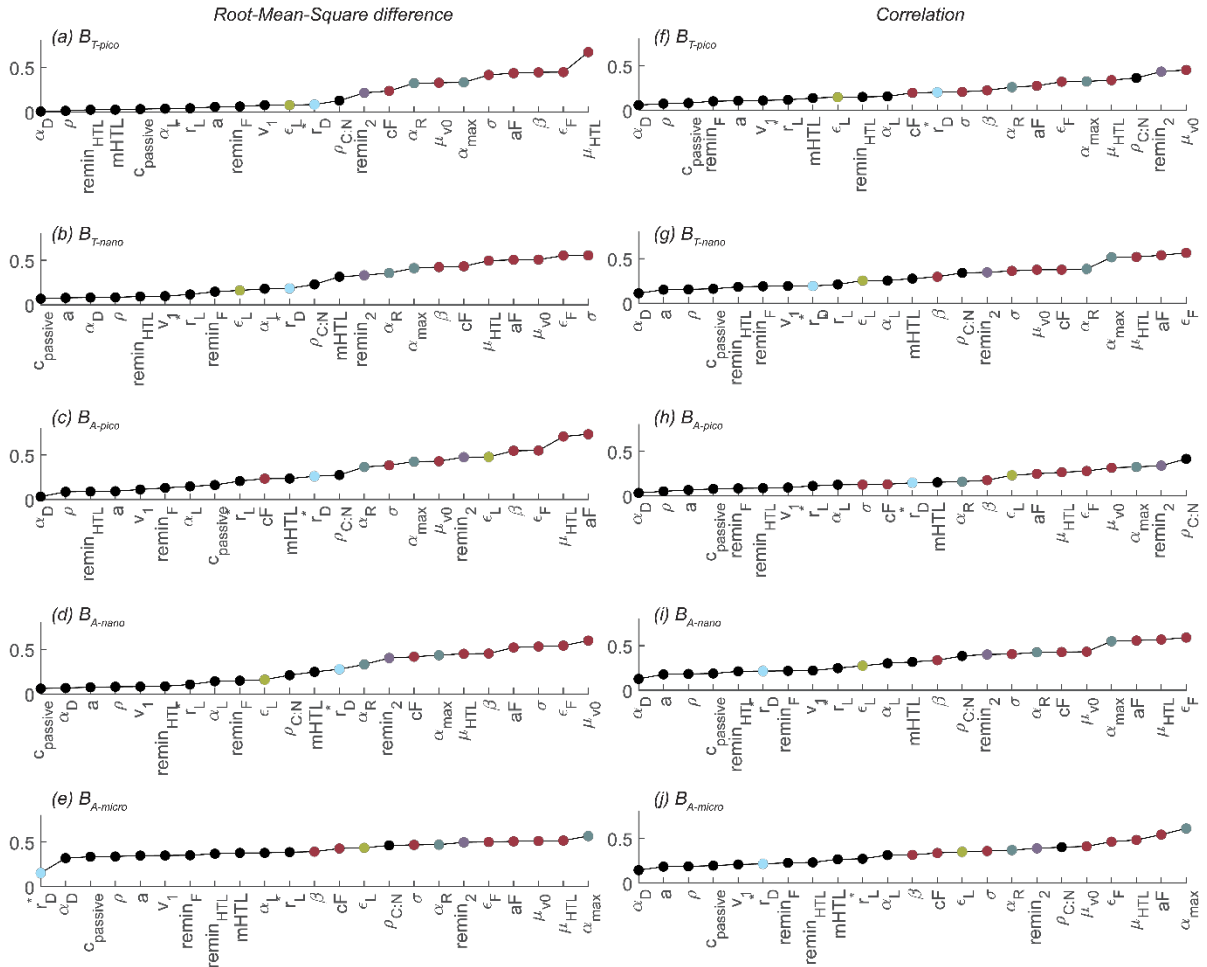
*Figure S4: Global parameter sensitivity ranked based on sensitivity index calculated by Sobol's variance-based sensitivity method for non-linear models for RMSd and Correlation. Note how all biomass sizes are especially sensitive to parameters controlling predation (red dots) and synthesis (grey dots). Parameter definitions in Table 1 and other abbreviations in Fig. 2.*

## S 7.   Random Parameters study for Station ALOHA

The first order parameter sensitivity study was done for both CCE and Station ALOHA. Figure S5 below show the result for Station ALOHA. The figure shows how many of the model simulations underestimate pico- and nanoplankton at low $AC_{bio}$ while overestimate at higher $AC_{bio}$. The simulations generally overestimate the microautotrophic biomass. This pattern is very similar to the pattern for the 100,000 simulations for the first-order sensitivity analysis at CCE (compare to Fig. 5 in text).
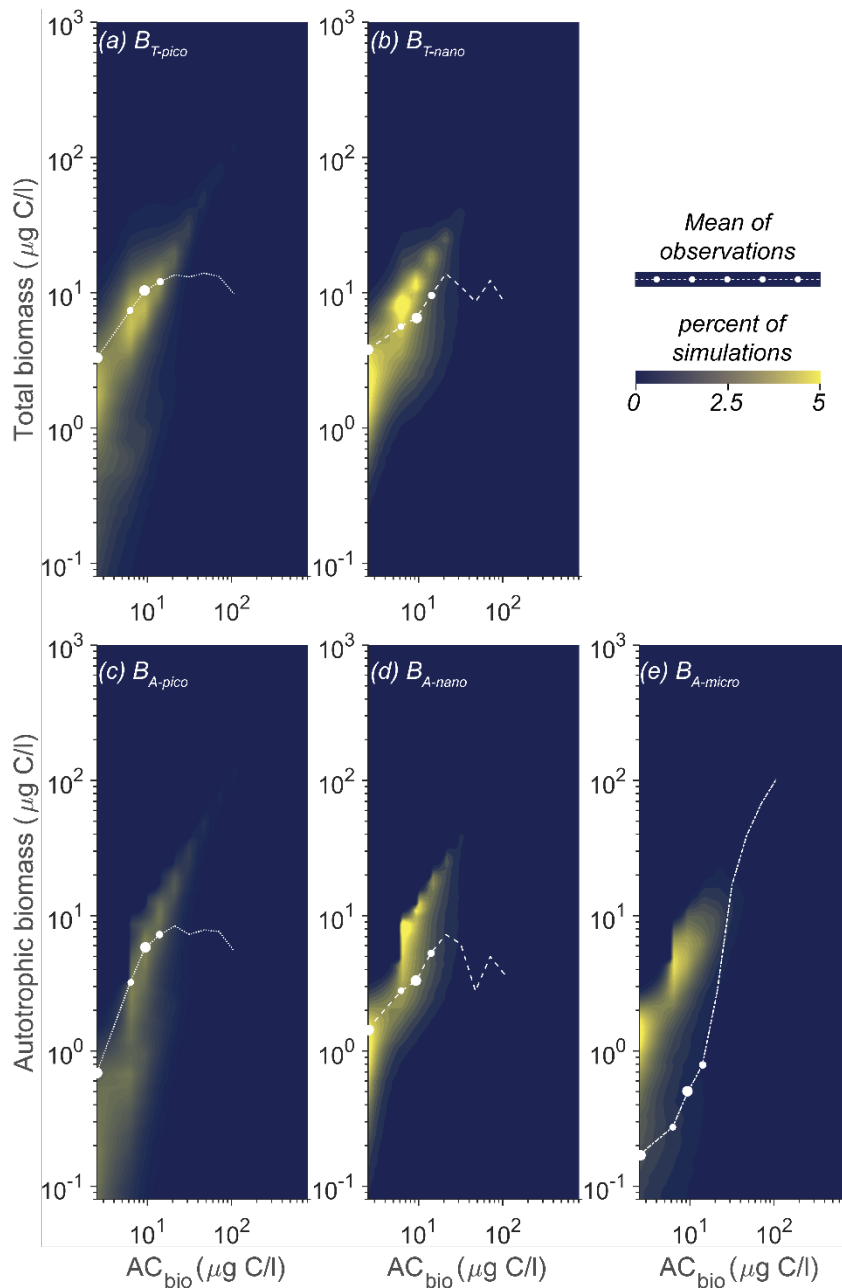
*Figure S5: Model mean and total biomass of size groups as a function of total biomass for the 100,000 random parameter combinations for Station ALOHA. Black dots are observations in ACbio bins. Abbreviations as Fig. 2.*

## References

Andersen, K. H. and Visser, A. W.: From cell size and first principles to structure and function of unicellular plankton communities, Progress in Oceanography, 213, 102995, https://doi.org/10.1016/j.pocean.2023.102995, 2023.

Eppley, R. W.: Temperature and phytoplankton growth in the sea, Fishery bulletin, 70, 1063, 1971.

Holling, C. S.: Some characteristics of simple types of predation and parasitism1, The canadian entomologist, 91, 385-398, 1959.

Stammer, D., Ueyoshi, K., Köhl, A., Large, W. G., Josey, S. A., and Wunsch, C.: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, Journal of Geophysical Research: Oceans, 109, https://doi.org/10.1029/2003JC002082, 2004.