The following supplements accompany the article

Oceanic heterotrophic bacterial nutrition by semilabile DOM
as revealed by data assimilative modeling

Ya-Wei Luo1,2,6,*, Marjorie A. M. Friedrichs3, Scott C. Doney4, Matthew J. Church5, Hugh W. Ducklow1,2

1The Ecosystems Center, Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA
2Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA
3Virginia Institute of Marine Science, College of William and Mary, PO Box 1346, Gloucester Point, Virginia 23062, USA
4Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, MS#25, Woods Hole, Massachusetts 02543, USA
5School of Ocean and Earth Science and Technology, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA
6Present address: Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, MS#25, Woods Hole, Massachusetts 02543, USA

*Email: yluo@whoi.edu


Supplement 1. Model Structure and Its Application to a Data Assimilative Testbed for Open Ocean. This supplement includes 2 sections. The first section describes the ecosystem model structure, and the second section describes the data assimilative testbed.

1. Ecosystem Model Description

The structure and the state variables of the model are shown in Fig. 1 and Table 1 in the main article and will be discussed below. The model tracks the flows and stocks of carbon (C), nitrogen (N) and phosphorus (P) through each of the state variables with the exception of inorganic nutrients. The model also tracks chlorophyll a (Chl a) for 3 phytoplankton variables. The full parameters and equations of the model are listed in Table S1 and Appendix A.

1.1. State Variables

The names and definitions of state variables are listed in Table 1. Phytoplankton (PHY) in the model represents both eukaryotic and prokaryotic autotrophs. There are 2 forms of diazotrophic cyanobacteria represented in the model: *Trichodesmium* spp. (TR) and unicellular N$_2$-fixing cyanobacteria (UN) which conduct photosynthesis and fix N$_2$ when ambient ammonium and nitrate levels are low. Bacteria (BA) represent heterotrophic prokaryotes including both *Bacteria* and *Archaea* using DOM as their sole food source. Protozoa (PRT) are protozoan zooplankton capable of ingesting PHY, UN and BA. Metazoa (MZ) are metazoan zooplankton that graze on TR and PRT. DOM is divided into labile (LDOM) and semilabile (SDOM) pools. Refractory DOM is not explicitly represented because its turnover time is much longer than the other 2 pools and would not significantly influence the system. However, fluxes to refractory DOM are implicitly represented as loss terms for some state variables. Detritus (DET) in the model is not divided into different sizes but represents an average pool. All these state variables are independently tracked as C, N and P, and additional Chl a for PHY, TR and UN. Nutrients in the model include 2 forms of inorganic N, ammonium (NH4) and nitrate (NO3), and 1 form of inorganic P, phosphate (PO4).

1.2. Parameters

Table S1 lists all the model parameters. Most parameters are subject to optimization by data assimilation. Another set is some parameters of UN, which are assigned equal values to other parameters of SP and TR, i.e. they change with other parameters. The 3rd set is ‘quota’ parameters, which are cellular molar N:C and P:C ratios of PHY, TR, UN, BA, PRT and MZ defined as ‘reference’, ‘minimum’ and/or ‘maximum’ quota for each state variable, in which Redfield ratios are used as ‘reference’ quota for phytoplankton variables. The quota parameters are assigned fixed values and are not subject to optimization by data assimilation.
1.3. Temperature Effect

Growth, respiration and N\textsubscript{2} fixation rates in the model are affected by water temperature. The values of these parameters are defined at 25°C and Arrhenius function is adopted to increase/decrease them with increasing/decreasing temperature (Appendix A 1.1). The default Arrhenius parameter of 4000 used here (Moore et al. 2002) increases the rates by 57% if the temperature increases 10°C (to 35°C), or decreases the rates by 38% if the temperature decreases 10°C (to 15°C). These rates include maximum growth rates of PHY, TR, UN, BA, PRT and MZ, maximum N\textsubscript{2} fixation rates of TR and UN, basal respiration rates of BA, PRT and MZ. In addition, a temperature below 20°C will completely cease the N\textsubscript{2} fixation of TR. Because active respiration rates are related to growth rates, which have already been adjusted by temperature, active respiration rates are not changed independently by the temperature effect.

1.4. Phytoplankton (PHY) Processes

We adopted the scheme of Geider et al. (1998) (GD98 hereafter) to simulate phytoplankton photosynthesis, nitrogen consumption and the synthesis of Chl \textit{a} with some modifications. The nitrogen consumption scheme of GD 98 is also applied to phosphorus assimilation although it may not be perfect because phosphorus can be ‘luxury consumed’ in great excess of current needs (Armstrong 2006). However, our model focuses more on nitrogen cycle than phosphorus cycle. And the problem is also partially alleviated by given a higher maximum phosphorus quota for phytoplankton as discussed below.

As proposed by Droop (1974), the phytoplankton carbon growth rate is limited by the cellular nutrient contents or ‘cell quota’. We follow the modified form of GD98 where the nutrient to carbon biomass ratios, i.e. N:C and P:C, replace the cell quota in Droop (1974). The nutrient-limitation term we used is slightly different from GD98 (using N as an example):

\begin{equation}
N_{\text{funPHY}} = \frac{Q_{N,\text{PHY}}^C - q_{N,\text{min,PHY}}^C}{q_{N,\text{ rdf,PHY}}^C - q_{N,\text{min,PHY}}^C}, \quad \text{then set } \ 0 \leq N_{\text{funPHY}} \leq 1, \tag{1}
\end{equation}

where \(Q_{N,\text{PHY}}^C\) is the cellular molar N:C ratios of PHY, and \(q_{N,\text{min,PHY}}^C\) and \(q_{N,\text{ rdf,PHY}}^C\) are parameters of the minimum and reference cellular molar N:C ratios. The P-limitation term \(P_{\text{funPHY}}\) has the same form as \(N_{\text{funPHY}}\) by using quota parameters for phosphorus. So that the light-saturated photosynthesis rate is:

\begin{equation}
P_{\text{max,PHY}}^C = \mu_{\text{PHY}} \cdot T_{\text{fun}} \cdot \min \left( N_{\text{funPHY}}, P_{\text{funPHY}} \right), \tag{2}
\end{equation}

where \(\mu_{\text{PHY}}\) is the parameter for maximum C-specific growth rate and \(T_{\text{fun}}\) is the temperature effect. The Redfield ratio is used for the reference quota \(q_{\text{ rdf}}\) in the model because some recent studies showed it is robust as a mean value for phytoplankton composition (Falkowski 2000, Michaels 2001). Therefore Equations (1) and (2) represent that the photosynthesis rate starts to drop when the cell quota becomes lower than the Redfield ratio until it is completely shutdown when the cell quota reaches the minimum value \(q_{\text{min}}\).

The actual photosynthesis rate will also be limited by light intensity as:

\begin{equation}
GROW_{\text{PHY}}^C = C_{\text{PHY}} \cdot P_{\text{max,PHY}}^C \cdot \left(1 - \exp \left( -\alpha_{\text{PHY}}^{\text{CHL}} \cdot \frac{Q_{\text{CHL,PHY}}^C \cdot PAR}{p_{\text{max,PHY}}^C} \right) \cdot \exp \left( -\beta_{\text{PHY}} \cdot PAR \right) \right), \tag{3}
\end{equation}

where \(C_{\text{PHY}}\) is the carbon biomass of PHY, \(\alpha_{\text{PHY}}^{\text{CHL}}\) is the parameter for initial slope of photosynthesis versus PAR (photosynthetically active radiation) curve, \(Q_{\text{CHL,PHY}}^C\) is the ratio of Chl \textit{a} to carbon biomass, \(\beta_{\text{PHY}}\) is the light-inhibition parameter. This equation is slightly different from GD98 with a light inhibition term \(\exp \left( -\beta_{\text{PHY}} \cdot PAR \right)\) included.

For N and P consumption, the limiting term is (using N as an example):
\[ V_{\text{max,PHY}}^N = \frac{q_{C,\text{max,PHY}}^C - Q_{N,\text{PHY}}^C}{q_C^C} \cdot V_{\text{max,PHY}}^N, \quad \text{then set} \quad 0 \leq V_{\text{max,PHY}}^N \leq 1, \quad (4) \]

where \( q_{C,\text{max,PHY}}^C \) is the parameter of maximum allowable cellular molar N:C ratio. Thus the N or P consumption is not limited by cellular quota if it is lower than the Redfield ratio, but starts to drop when the quota is higher than the Redfield ratio until the consumption is totally shutdown when the quota reaches \( q_{\text{max}} \). The consumption is also limited by the ambient nutrient level with a Monod function and the actual consumption rates for ammonium and nitrate is:

\[ \begin{align*}
GROW_{\text{PHY}}^{NH4} &= C_{\text{PHY}} \cdot V_{\text{REF,PHY}}^N \cdot T_{\text{fun}} \cdot V_{\text{max,PHY}}^N \cdot \frac{NH4}{NH4 + k_{\text{PHY}}^{NH4} + NO3 \cdot k_{\text{PHY}}^{NO3} \cdot k_{\text{PHY}}^{NO3}}, \\
GROW_{\text{PHY}}^{NO3} &= C_{\text{PHY}} \cdot V_{\text{REF,PHY}}^N \cdot T_{\text{fun}} \cdot V_{\text{max,PHY}}^N \cdot \frac{NO3}{NO3 + k_{\text{PHY}}^{NO3} + NH4 \cdot k_{\text{PHY}}^{NO3} \cdot k_{\text{PHY}}^{NH4}}. 
\end{align*} \quad (5) \]

\[ \begin{align*}
\text{where} \quad V_{\text{REF,PHY}}^N \quad \text{is the maximum N consumption rate per C biomass,} \quad k_{\text{PHY}}^{NH4} \quad \text{and} \quad k_{\text{PHY}}^{NO3} \quad \text{are the half-saturation concentrations of ammonium and nitrate. This form avoids the total N consumption rate to exceed} \quad V_{\text{REF,PHY}}^N \quad \text{and represents an ammonium inhibition on nitrate consumption with a lower} \quad k_{\text{PHY}}^{NH4} \quad \text{than} \quad k_{\text{PHY}}^{NO3}. \quad \text{The phosphate consumption is the similar Monod function as the N consumption of Equations (5) and (6) but without an inhibition term in the denominator. The model also considers the required carbon respiration by nitrate consumption (Appendix A2.6), but ignores the energy requirement for ammonium and phosphate consumption.}
\end{align*} \]

It is difficult to determine the minimum and maximum quotas. Goldman et al. (1979) obtained molar N:C between 0.05 - 0.14 and molar P:C between 0.001 – 0.03 under different nutrient and light levels. Falkowski et al. (1985) observed molar N:C between 0.11 - 0.23 for 3 phytoplankton species under different light levels. GD98 used 0.047 and 0.23 (0.19 for one studying species) for minimum and maximum molar N:C of the 4 phytoplankton species in their model. In our model, \( q_{C,\text{min,PHY}}^C \) and \( q_{C,\text{max,PHY}}^C \) are set to be same as GD98, but \( q_{P,\text{min,PHY}}^C \) and \( q_{P,\text{max,PHY}}^C \) are set to be ±80% from \( q_{P,\text{rdf,PHY}}^C \) (Table S1). These parameters roughly represent the observed pattern that P:C varies more than N:C and are within the observed ranges, which also allows an alleviated limitation on P consumption.

A scenario of primary production as described by the model could be: (1) At the beginning of day, the N:C cellular ratio of phytoplankton tends to be high as the phytoplankton has consumed N during last night. Thus the C assimilation rate will be high while the N assimilation is low at early of the day, which will reduce cellular N:C ratio and therefore gradually reduce the photosynthesis rate. If the N:C ratio reaches \( q_{\text{min}} \), the C assimilation is totally shutdown. (2) During night, the C assimilation is totally shutdown; the low N:C resulted during day makes N consumption high and cellular N content will be restoring. The increasing N:C will slow down the N assimilation rate. If N:C reaches \( q_{\text{max}} \), N consumption will be completely shutdown.

The Chl \( \alpha \) production rate is 0 if PAR = 0, and otherwise is determined as:

\[ GROW_{\text{PHY}}^{\text{CHL}} = \theta \cdot GROW_{\text{PHY}}^N \cdot GROW_{\text{PHY}}^C \cdot \frac{\alpha_{\text{PHY}}^{\text{CHL}} \cdot \text{CHL}_{\text{PHY}} \cdot \text{PAR} \cdot \text{EXP}(-\beta_{\text{PHY}} \cdot \text{PAR})}{\alpha_{\text{PHY}}^{\text{CHL}} \cdot \text{CHL}_{\text{PHY}} \cdot \text{PAR} \cdot \text{EXP}(-\beta_{\text{PHY}} \cdot \text{PAR})}, \quad (7) \]

where \( \theta \) is the maximum Chl \( \alpha \) to N ratio. The last multiplier on the right represents the ratio of energy assimilated to energy absorbed as described by GD98. The equation represents a variable allocation of assimilated N to Chl \( \alpha \), and a low PAR level will generate a high Chl production.

The model considers the DOM production of phytoplankton in 3 pathways (Appendix A 2.7): passive release of labile DOM, active release of carbohydrate and other DOM.
Two models have been proposed to explain extracellular release of DOM by phytoplankton. The passive diffusion model (Fogg 1966, Bjørnsen 1988) considers that low-molecular-weight (LMW) DOM will passively cross the phytoplankton cell membrane because of the concentration gradient from cell to environment. In our model, the passively released LMW DOM has the same C:N:P ratio as phytoplankton biomass (Biddanda & Benner 1997) and its default exudation rate is set as 5% of biomass per day (Bjørnsen, 1988). We assume these LMW materials such as neutral sugars and dissolved free amino acids (DFAA) are most labile and therefore go directly to the labile DOM pool.

Another model, the overflow model (Fogg 1966, Nagata 2000), assumes that carbon production by photosynthesis is relatively independent of inorganic nutrient incorporation and therefore fixed carbon could be more than the amount that can be supported by the incorporated nutrients. Under this condition, some DOM with high carbon to nutrient ratio may be actively released (excreted). However, this assumption is not always valid as the high-molecular-weight (HMW) DOM production is commonly observed despite of the nutritional conditions. Carbohydrates is one of the major forms of the released DOM (Biddanda & Benner 1997, Biersmith & Benner 1998) and the model assumes 5% of gross primary production is released as carbohydrate. These carbohydrate can be labile (Amon & Benner 1994, Santschi et al. 1995), but can also seasonally accumulate in the water (Carlson et al. 1994, Carlson et al. 1998). The model roughly divided the released carbohydrate as 75% of labile dissolved organic carbon (DOC) and 25% of semilabile DOC.

The model also considers that phytoplankton release DOM in order to adjust their stoichiometry to approach the Redfield ratio. We assume this type of release only happens when cellular organic C is in excess, and it does not occur when cellular organic C is low. This is reasonable because nutrients are generally limiting in the ocean and this mechanism helps phytoplankton to compete with other organisms by storing nutrients when ambient concentrations are high. This function calculates the amount of excess C relative to the most limiting nutrient (N or P) in phytoplankton biomass, and assumes this amount of C is released on a time scale of 2 days. Also, under the excess carbon condition, there would be another excess element, either N or P, when comparing these 2 element quotas. Bronk (2002) argued the overflow model is unlikely to impact dissolved organic nitrogen (DON) release significantly because most environments that are lacking in P generally do not have an excess of DON. However, phytoplankton also release DON and dissolved organic phosphorous (DOP) (Nagata 2000, Carlson 2002) and there is no evidence that these DON and DOP are exclusively from passive release. In other words, DON or DOP may also be actively released by phytoplankton. To address these 2 issues, a lower release rate as 1/4 that of the DOC, i.e. a time scale of 8 days, is used when DON or DOP are released.

Morán et al. (2001) found that the phytoplankton DOC production (release) increases with the phytoplankton particulate organic carbon (POC) production, but the ratio of DOC to POC production has negative relation to POC production. The easiest model to realize this observation is a constant release term that is independent with the primary production, and a variable term that is proportional to the primary production, as the scheme WE used here for a passive release and an active carbohydrate release.

Phytoplankton also aggregate and form sinking particles via a concentration-dependent process that is proportional to the square of the biomass (Appendix A 2.8). The phytoplankton removal by protozoa grazing is discussed in the section on zooplankton processes.

Chl a is removed with aggregation and grazing processes of phytoplankton, but is not influenced by the DOM release.

1.5. Diazotrophic Phytoplankton Processes

There are 2 different diazotrophic compartments represented in the model, Trichodesmium spp. (TR) and Unicellular N2-fixers (UN). The equations controlling their C growth rate are the same as for PHY, although different parameter values are used to set light limitation and cell quotas. A much smaller initial slope of production versus light (αTRCHL) puts a stronger light limitation for TR and limits the majority of its production to the surface ocean (Letelier & Karl 1996, 1998). The maximum growth rate and the initial slope of UN (αUNCHL) are smaller than PHY, but the initial slope is large enough to allow UN to exist at depths down to the base of the euphotic zone (Zehr et al. 2001, Montoya et al. 2004). The cellular N:C quota of TR parameters are set to be slightly higher than PHY to represent the observations for Trichodesmium spp. (Letelier & Karl 1996) and to slightly stimulate N2 fixation. The P:C quota parameters of TR are set to be significantly smaller than PHY to represent lower P requirements by Trichodesmium spp. as observed by some studies (Letelier & Karl 1996, Fu et al. 2005). As there is no data for the stoichiometry of unicellular diazotrophs, UN uses the same quota parameter values as TR.

TR and UN take up dissolved inorganic nutrients similar to PHY in the model (Carpenter & McCarthy 1975, Ohki et al. 1991, Mulholland et al. 1999), but with higher TR half-saturation concentration for inorganic nutrients than PHY. Most TR cells in the ocean aggregate as colonies with larger size than single cells, and they are expected, therefore, to take up inorganic nutrients at lower rates. We assume UN has the same half-saturation concentrations for inorganic nutrient uptake as PHY. However, the major difference of N2-fixers from PHY is their ability to fix N2 to alleviate the limitation by dissolved inorganic nitrogen (DIN). To simulate N2 fixation, the model first calculates the maximum total N requirement:
\[ GROW_{\text{max}, \text{TR}}^N = C_{\text{UN}} \cdot V_{\text{REF,TR}}^N \cdot T_{\text{fun}} \cdot V_{\text{max,TR}}^N, \]  

where \( V_{\text{REF,TR}}^N \) is the maximum N growth rate per C biomass, \( V_{\text{max,TR}}^N \) is the cell quota limitation with the same form as PHY (Equation 4). The model first fulfills this requirement by DIN (ammonium and nitrate) consumption because they use less energy than N\(_2\) fixation (Karl et al. 2002). If the DIN consumption is not enough to satisfy N requirement, the model triggers N\(_2\) fixation.

To calculate N\(_2\) fixation rate, the model then assumes under optimal condition, the ratio of total N production to total C production after respiration equals to the maximum cellular N:C quota \( q_{\text{C, max,PHY}}^N \) (using TR as an example):

\[
\frac{GROW_{\text{NO3,TR}}^C + GROW_{\text{NH4,TR}}^C + NFIX_{\text{max,TR}}^C}{GROW_{\text{TR}}^C - \eta_{\text{NO3}} \cdot GROW_{\text{NO3,TR}}^C - \eta_{\text{NF}} \cdot NFIX_{\text{max,TR}}^C} = q_{\text{C, max,PHY}}^N,  
\]

where \( NFIX_{\text{max,TR}}^C \) is the maximum N\(_2\) fixation rate of TR, \( \eta_{\text{NF}} \) is the energy (respiration) requirement rate per N\(_2\) fixation.

Together with temperature control \( T_{\text{fun}} \) and cell quota control \( V_{\text{max,TR}}^N \), the maximum N\(_2\) fixation rate can be derived from Equation (9) as:

\[
NFIX_{\text{max,TR}}^C = \left( \frac{GROW_{\text{TR}}^C - \eta_{\text{NO3}} \cdot GROW_{\text{NO3,TR}}^C}{1 + \eta_{\text{NF}} \cdot q_{\text{C, max,PHY}}^N} \right) \cdot q_{\text{C, max,PHY}}^N - GROW_{\text{NO3,TR}}^C - GROW_{\text{NH4,TR}}^C \cdot T_{\text{fun}} \cdot V_{\text{max,TR}}^N.  
\]

For TR, if temperature drops below 20°C, \( NFIX_{\text{max,TR}}^C \) is set to 0. The N\(_2\) fixation is only a supplementary N source in the model and the realized N\(_2\) fixation plus the DIN uptake is bounded by the maximum N requirement \( GROW_{\text{max,TR}}^N \). Both nitrate consumption and N\(_2\) fixation requires carbon respiration (Appendix A 3.5).

Both TR and UN take up phosphate similarly as PHY. But a possible advantage of Trichodesmium spp. was proposed by some studies that when phosphate is depleted at the surface they can control their buoyancy to sink during night to take up phosphate and then buoys back to surface (Letelier & Karl 1998, Villareal & Carpenter 2003, White et al. 2006). Our model implicitly represents this process by proposing that TR can refill and make its P quota to \( q_{\text{P, max,TR}}^C \) in a time scale of 1 day. Thus this P refill will be another new supply of P to the model.

Trichodesmium spp. passively releases DOM (Capone et al. 1994, Glibert & Bronk 1994) and ammonium (Mulholland & Capone 2001, Mulholland et al. 2004, Mulholland 2007). In addition to the same passive and active release of DOM and carbohydrate as PHY, a portion of N fixed by N\(_2\) fixation will be released as ammonium and DON (Appendix A 3.7). The rate varies in different studies (Glibert & Bronk 1994, Mulholland et al. 2004) and the model in default assumes 18% of N fixed by N\(_2\) fixation is released as ammonium and another 18% is released as labile DON. Although there is no clear result showing how unicellular N\(_2\)-fixers release DOM or ammonium, the model assumes they are similar to TR.

The model-defined competition among PHY, UN and TR allows coexistence, but the dominant species may vary with changing environmental conditions. PHY has a higher light-saturate growth rate, so that it dominates when the ambient nutrients are replete. If the nutrients are depleted, PHY growth is limited and N\(_2\)-fixers, either UN or TR, may dominate. Because UN has a higher initial slope of the P vs. E curve, it will dominate over TR in the deeper layer. However, near the surface (e.g. upper 50 m) where phosphate may be depleted, the ability of TR to take phosphate from deep water by buoyancy regulation may help it compete over UN. In addition, when light is not low, TR can do more N\(_2\) fixation than UN because TR has a higher maximum N\(_2\) fixation rate. Another feature of the model is the different grazing rates on TR and UN. TR is mostly in colonies and may be toxic to zooplankton (Hawser et al. 1992). In the model TR is grazed by metazoa while UN is grazed by protozoa.

### 1.6. Bacterial Processes

The DOM is divided into labile and semilabile DOM. The model assumes that the Available Labile DOC (ALC) for bacterial uptake equals to the concentration of labile DOM, i.e. all the labile DOM is available for the bacterial utilization during a time unit. However, only a portion of semilabile DOM is available for bacterial uptake during a time unit. The ALC and the Available Semilabile DOC (ASC) is represented as:

\[
ALC = C_{\text{LDOM}}, \quad ASC = r_{\text{SDOM}} \cdot C_{\text{SDOM}},  
\]
where $r_{SDOM}$ is the parameter controlling semilabile lability.

Bacteria C growth rate depends on both bacterial cellular quota and available labile and semilabile DOC. The cellular quota limitation is (using N as an example):

$$N_{fun}^{BA} = Q_{N,BA}^{C} / q_{N,BA}^{C}, \text{ then set } N_{fun}^{BA} \leq 1,$$

and a similar form is used to P quota limitation. Therefore bacteria C growth will be limited if its cellular N or P quota is less than the corresponding reference quota. The bacterial labile and semilabile DOC consumption is:

$$GROW_{BA}^{LDOM} = \mu_{BA} \cdot T_{fun} \cdot C_{BA}^{C} \cdot \min \left( N_{fun}^{BA}, P_{fun}^{BA} \right) \cdot \frac{ALC}{ALC + k_{DOM} + ASC},$$  

$$GROW_{BA}^{SDOM} = \mu_{BA} \cdot T_{fun} \cdot C_{BA}^{C} \cdot \min \left( N_{fun}^{BA}, P_{fun}^{BA} \right) \cdot \frac{ASC}{ASC + k_{DOM} + ASC},$$

where $\mu_{BA}$ is the C-specific maximum bacterial growth rate, $k_{DOM}$ is half-saturate concentration of available DOC for bacterial growth. Equation (13) and (14) does not artificially increase the bacterial C growth rate to exceed $\mu_{BA}$ while the bacterial uptake of labile and semilabile DOC will influence each other. The scheme sets ALC and ASC equally in terms of bacterial preference. But ALC equals the concentration of labile DOM, and ASC is set by Equation (11) as a portion of semilabile DOM. Thus the parameter $r_{SDOM}$ in Equation (11) actually represents the relative lability between semilabile DOC and labile DOC. Because labile and semilabile DOC are the only sources for bacterial C growth in the model, the summary of Equations (13) and (14) set the bacterial carbon demand (BCD), or the total bacterial carbon consumption.

The model assumes bacteria take labile DOM as a bulk substrate so that the ratio of labile DON (DOP) to labile DOC consumption is same as the bulk N(P):C ratio of labile DOM. However, the model assumes that bacteria are capable of selecting semilabile DOM with higher N and P contents. This assumption is consistent and necessary for the model mechanism that semilabile DOM with higher N:C or P:C ratio is more labile (Equation 11). The selection-term controlled semilabile DON is:

$$GROW_{BA}^{SDON} = GROW_{BA}^{SDOM} \cdot \min \left( q_{N,BA}^{C}, Q_{N,SDOM}^{C} + f_{SLCT,BA} / N_{fun}^{BA} \cdot \left( Q_{N,BA}^{C} - Q_{N,SDOM}^{C} \right) \right),$$

where $f_{SLCT,BA}$ is the parameter of selection strength varying between 0 and 1. Thus the ratio of semilabile DON to DOC consumption varies between the bulk N:C of semilabile DOM $Q_{N,SDOM}^{C}$ and the bacterial reference cellular quota $q_{N,BA}^{C}$. The ratio will be moved towards $q_{N,BA}^{C}$ with higher selection strength $f_{SLCT,BA}$ and lower limitation of bacteria cellular quota $N_{fun}^{BA}$ (therefore lower $Q_{N,BA}^{C}$). Same form is used for semilabile DOP consumption.

Bacteria take up DIN even if the net flux is releasing DIN as discovered by Tupas et al. (1994). We propose bacteria take up DOM as well as dissolved inorganic nutrients simultaneously. Because the stoichiometry of bacterial cells does not vary substantially (Kirchman 2000), bacteria have to maintain their elemental composition by taking up or releasing inorganic nutrients. Actually for modeling purposes, simulation only of the net flux for inorganic nutrients is sufficient. Our model simulates both the influx and outflux of inorganic nutrients in order to compare DOM and inorganic nutrient influxes and their limiting effect on bacteria. But only the net flux of inorganic nutrients is meaningful for other state variables. Some evidence shows that ammonium and phosphate are readily substrate for bacteria uptake (Kirchman 1994). As the reduced forms, the model assumes they have same activity as labile DOM. The ratio of ammonium (phosphate) to labile DON (DOP) consumption equals to the ratio of their concentration before bacteria cellular quota limitation (use ammonium as an example):

$$GROW_{BA}^{NH4} = GROW_{BA}^{LDON} \cdot NH4 / N_{LDOM} / N_{fun}^{BA},$$

However, the energetic cost to assimilate nitrate are higher than DON (Vallino et al. 1996) as nitrate requires reduction before being assimilated. But bacterial nitrate uptake was commonly observed (Kirchman & Wheeler 1998, Middelburg & Nieuwenhuize 2000, Allen et al. 2002). The model assumes bacterial nitrate consumption occurs only when the bacterial cellular N:C ratio is less than its reference ratio. In other words, bacteria only take up nitrate when they are in short of nitrogen. This
Assumption is reasonable because there is no reason for bacteria to take up nitrate and to respire energy while their cells have enough nitrogen stored. When the condition is met, the equation used for bacterial nitrate consumption is:

\[
GROW_{BA}^{NO3} = \min \left( 0.1 \cdot NO3 \cdot \frac{1}{Nfun_{BA}} \cdot \frac{GROW_{BA}^{LDON} + GROW_{BA}^{SDON}}{N_{LDOM} + N_{SDOM}}, (NO3 + NH4) \cdot \frac{GROW_{BA}^{LDON} + GROW_{BA}^{SDON}}{N_{LDOM} + N_{SDOM}} - GROW_{BA}^{NH4} \right). \tag{17}
\]

So the equation puts 2 limitations on the nitrate uptake: (1) nitrate consumption is no more than 10% of N-specific bulk DOM (including labile and semilabile) consumption rate, which limits the maximum nitrate consumption rate; and (2) the nitrate plus ammonium consumption is no more than N-specific bulk DOM (including labile and semilabile) consumption rate, which sets the inhibition of ammonium consumption on nitrate consumption.

Bacterial growth efficiency, BGE, equals to the ratio of bacterial net production (BP) to bacterial carbon demand (BCD), BGE = BP/BCD. BCD = BP + BR, where BR is the bacterial respiration. Thus BGE = BP/(BP+BR), or BR = BCD×(1 – BGE). As BCD has been known, the model has to estimate BGE so that BP can be derived. A paper of del Giorgio and Cole (1998) summarized published in situ BGE and BP data and found BGE increases with BP. By definition, a higher BP together with a higher BGE indicates a higher BCD. Thus the observations also support a positive relation between BGE and BCD. The respiration of bacterial carbon in the model includes basal respiration, active respiration and the required respiration to reduce assimilate nitrate. The function is:

\[
RESP_{BA} = \xi^{NO3} \cdot GROW_{BA}^{NO3} + resp^B_{BA} \cdot Tj\nu \cdot C_{BA} + \left( resp^A_{\text{min,BA}} + resp^A_{\text{max,BA}} - resp^A_{\text{min,BA}} \right) \cdot \exp \left( -b_{\text{resp}} \cdot GROW_{BA}^C \right) \cdot GROW_{BA}^C, \tag{18}
\]

where \( resp^B_{BA} \) is the basal respiration rate in day\(^{-1}\), \( GROW_{BA}^C \) is the bacterial carbon demand (BCD) or the bacterial total DOC uptake, \( resp^A_{\text{min,BA}} \) and \( resp^A_{\text{max,BA}} \) are the minimum and maximum portions of gross production actively respired with inverted unit to \( C_{BA} \), and \( b_{\text{resp}} \) is a positive parameter with inverted unit to \( C_{BA} \). By using this function, the BGE can increase with BP.

Bacteria can be an important source of refractory DOM from their membranes or can transform labile DOM to refractory DOM (Tanoue et al. 1995, McCarthy et al. 1998, Carlson 2002). Our data assimilation optimizes the refractory DOC release by bacteria around 2~3% of their carbon uptake, which is inside the range of other studies (1.4 ~ 8.6% as summarized by Carlson 2002). In the deep ocean where DOM is mostly refractory, DON:DOC and DOP:DOC are ~0.05 and ~0.0007 respectively as shown by the HOT data set, which are the ratios we use for bacterial refractory DOM in the model. The refractory DOM is not explicitly represented in the model and is assumed to be a loss from the model domain.

After the above processes, the model adjusts bacterial elemental composition by either (1) excreting ammonium or phosphate if N or P is in excess relative to C, or (2) excreting semilabile DOM if C is in excess. A similar function as phytoplankton active release is used to control excretion rates. Compared to phytoplankton excretion, higher values are given to semilabile DOM release rate \( ex_{\text{adj,BA}} \) and inorganic nutrient regeneration rate \( remi_{BA} \) (Appendix A5.9) to reflect that bacteria have higher ability to control their stoichiometry. Bacteria are removed by protozoa which will be discussed in zooplankton processes. Also, there is a mortality term that assumes a certain percentage (2% - 5% resulted by data assimilation) of bacteria biomass per day is lost to labile DOM pool due to virus attack.

In summary, our model represents bacterial processes differently from previous models mainly in 3 ways by including: (1) continuous variable lability of semilabile DOM determined by nutrient contents, (2) selective semilabile DOM uptake by bacteria, and (3) variable bacterial growth efficiency depending on bacterial production rate.
1.7. Zooplankton (PRT and MZ) Processes

We use the same form of grazing function for PRT grazing on PHY, UN and BA, and MZ grazing on TR and PRT. The general form of carbon-specific grazed amount of a prey \( \text{pry} \) by a predator \( \text{pdt} \) is:

\[
GRAZ_{\text{pry}}^C = T \text{fun} \cdot \mu_{\text{pdt}} \cdot C_{\text{pdt}}^2 \cdot \frac{C_{\text{pry}}^2}{C_{\text{pry}}^2 + g_{\text{pry}}^2 + \left( \frac{C_{\text{pry}2} \cdot g_{\text{pry}} / g_{\text{pry}2}}{2} \right)^2 + \left( \frac{C_{\text{pry}3} \cdot g_{\text{pry}} / g_{\text{pry}3}}{2} \right)^2 + \cdots},
\]  

(19)

where \( \mu_{\text{pdt}} \) is the maximum grazing rate in day\(^{-1}\), and \( g_{\text{pry}} \) is the grazing parameter depending on the grazer and prey, the subscripts \( \text{pry}2 \) and \( \text{pry}3 \) etc. represent other preys grazed by the same predator \( \text{pdt} \). Thus Equation (19) represents a density-dependent grazing and also a preferential selection of a predator on different food sources. N and P components are grazed relative to C at the same ratios as those of prey composition.

Zooplankton release a portion of ingested organic matter as DOM both from sloppy feeding and excretion. The model does not separate these 2 differences sources, but assumes a certain percentage of the total grazed C by PRT and MZ is release as DOC. The model then partitions the release DOC in default with 75% to labile pool and the other 25% to semilabile pool. The ratio of released labile DON (DOP) to released labile DOC is same as the N(P):C ratio of zooplankton food source. The semilabile DON (DOP) release is treated similarly as labile DON (DOP), but the rate can be increased or reduced when the zooplankton cellular N(P):C ratio is above or below the reference ratio. (Appendix A 6.2 & 7.2).

Both PRT and MZ respire C from basal respiration which respires a portion of their biomass, and active respiration which respires a portion of their grazed C.

After the above DOM excretion and respiration, PRT and MZ adjust their body quota by either releasing semilabile DOM if C is in excess, or regenerating inorganic nutrients if N or P is in excess (Appendix A 6.4-5 and 7.4-5). The scheme is similar to BA.

PRT and MZ also produce POM. The C component of this production is assumed to be a fixed percentage of grazed C. To simplify, a constant C:N:P ratio (molar 220:25:1) is defined for POM production so that the N and P components of it can be calculated from the C component. This ratio is higher than the Redfield ratio to indicate our assumption that zooplankton tends to assimilate higher quality food with lower C:N and C:P ratios, but to egest fecal materials with higher C:N and C:P ratios.

MZ (not applied to PRT) also produces refractory DOM with a N:C and P:C ratio similarly as bacteria. As MZ is the highest trophic level represented in the model, it is subject to a density-dependent removal process by implicit higher trophic levels. The removed C, N and P go to 3 pools, semilabile DOM, POM and inorganic nutrients, with the partition defined by parameters.

1.8. Other Processes (DET, dissolved inorganic nutrients & DOM)

The POM production from the processes discussed above is distributed to detritus pool and a constant vertical sinking speed is assigned to detritus. Detritus in the model dissolves and transforms to semilabile DOM at a certain rate when it sinks. Compared to other biogeochemical models that assign a remineralization rate for detritus, detritus has to be transformed to DOM before it can be regenerated to inorganic nutrients. The transformation rate of the P component of the detritus is higher than its N component, which is in turn higher than that of C. This is supported by the HOT data sets where POM in shallow sediment traps has more N and P than that collected at deeper depth.

The model also includes a simple nitrification process, which converts a certain percentage of ammonium to nitrate per day. Nitrification in the euphotic zone is argued to be an important process in order to estimate new production correctly in the world ocean (Yool et al. 2007).

Another parameter was assigned to convert SDOM to refractory DOM at a slow rate depending on the stoichiometry of SDOM (Appendix A 10.1). When N:C and P:C are high, SDOM converts to refractory DOM very slowly because N and P enriched SDOM are more likely labile. The conversion rate can be high when the N:C and P:C ratios of SDOM are very low. The C:N:P ratio of the produced refractory DOM is predefined by parameters.

2. Data Assimilative Testbed

Based on a testbed for the Arabian Sea and Equatorial Pacific (Friedrichs et al. 2007), a vertical 1-D data assimilative testbed was constructed for this study, including physical forcings and transportation schemes, initial and boundary conditions, an optimization software and assimilative data sets. A new study site, Station ALOHA was included. The testbed uses the variational adjoint method (Lawson et al. 1995) to optimize parameters towards observations. The method requires 4 components, all of which are supported by the testbed: (1) a forward model supported by physical forcings and initial and boundary conditions; (2) a cost function
evaluating misfits between the forward model results and the assimilated observations; (3) an adjoint of the forward model to compute the gradient of the cost function with respect to model parameters, generated by auto-differential software TAPENADE 2.1; and (4) an optimization procedure (M1QN3 3.1), which uses the computed gradient from the adjoint model to determine the direction and the optimal step size by which the parameters need to be modified in order to reduce the cost function. Thus the 4 components are used iteratively to adjust parameter values until preset criteria are met (e.g., low gradient). Theoretically, besides the model structure (i.e., equations), model results depend on model parameters as well as initial conditions, boundary conditions and physical conditions. Data assimilation could also incorporate these other factors as control variables but this is more computation-intensive. For our application, we assume these conditions are derived from observations, and the model spin-up can help to minimize the influence of initial conditions. Thus only model parameters were considered to be control variables and optimized by data assimilation.

Unless otherwise stated, most data used to develop the testbed are from the HOT time-series which is available online at http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html. We also improved the testbed for the Arabian Sea and Equatorial Pacific by assimilating more data types, which were collected online at http://usigofs.whoi.edu/jg/dir/jgofs/.

2.1. Forward Model Setup

Starting from the initial condition of the state variables, the model computes the temporal evolution of the state variables until the end of modeling period using a constant time step of 1 hour. The temporal evolution is affected by (1) ecological derivatives defined by the equations described in Section 1, which also requires inputs of PAR and temperature from physical forcings; and (2) physical derivatives from vertical advection, vertical diffusion, bottom boundary exchange and sinking of detritus, which requires inputs of vertical velocity, vertical diffusive coefficient, mixed layer depth and sinking speed of detritus. Vertical advection and detrital sinking are described with a third-order direct space-time upwind-biased scheme (Hundsdoerfer & Trompert 1994) and the Sweby flux limiter (Sweby 1984) and are simplified to work for 1-D (vertical) advection only. Vertical diffusion is applied using a Crank-Nicholson vertically variable diffusion operation (Press et al. 1986), with a closed upper boundary and an open bottom boundary. The effect of the horizontal advection of biological quantities (only applying to the Equatorial Pacific site as discussed below) is examined through a scaling analysis using output from a 3-D coupled biological–physical models (Christian et al. 2002). The ecological derivatives are not related to space differentials, and a simple forward-time scheme is used to solve the ecological model.

2.1.1. Study Sites and Model Grid

Two study sites are from the Arabian Sea (AS) in 1994 and Equatorial Pacific (EQP) in 1992. The testbed for these 2 sites has already been setup by others (Friedrichs et al. 2007), and we add more assimilative observation types to the testbed. The new study site is the Hawaii Ocean Time-series study station ALOHA (22°45´N, 158°W, HOT hereafter) in 2002.

Our application uses 10 surface layers of 5 m each and 15 bottom layers of 10 m each for HOT (25 layers, upper 200 m in total) and uses 10 surface layers of 5 m each and 10 bottom layers of 10 m each for AS and EQP (20 layers, upper 150 m in total). Our application uses 1 hour for each time step because it is short enough to resolve diurnal patterns and long enough that the outputs never showed any instability.

2.1.2. Initial and boundary Conditions

Initial conditions are set up at 90 days before the beginning of each calendar year so that a 90-day spin-up is conducted before the simulation to reduce the influence of initial conditions on the outputs. The initial conditions are estimated by linearly interpolating observations in time.

For the bottom boundary conditions, it is assumed vertical gradients of most biological species are zero at the bottom boundary. Exceptions are nitrate, phosphate and semilabile DOM. The first 2 substrates are major nutrients whose upward transport is crucial to fuel new primary production in the surface layer. Semilabile DOM dominates the total organic matter pool and shows strong gradients from surface to deep water. The concentration of these substrates at the bottom boundary was set by linearly interpolating the observed data. Another nutrient, ammonium, is expected to have a gradient at that depth because organic matter is remineralized from surface to deep. However, there are no data available so we assume the ammonium level is much lower than nitrate at the boundary, and its gradient is negligible compared to that of nitrate.

2.1.3. Physical Forcings

Sea surface photosynthetically active radiation (PAR), mixed layer depth (MLD), water temperature, vertical velocity and vertical eddy diffusivity were set to force the model. PAR was specified for each time step, and the others had temporal resolution of 1 day. Temperature, vertical velocity and vertical eddy diffusivity were set up at every vertical grid point. The physical forcings for AS and EQP were set up by others (Friedrichs et al. 2007). Here we will only discuss the setup of physical forcings for HOT.
(a) PAR: is required for photosynthesis in the model. Hourly clear-sky solar radiation at the sea surface for the station location was computed using a standard method described by Rosati & Miyakoda (1988). It was then compared to NCEP reanalysis daily average downward solar radiation interpolated to the station location to estimate the effect of daily average cloud coverage and albedo in every day (NCEP reanalysis provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at www.cdc.noaa.gov/). Thus, the hourly solar radiation reaching the sea surface was estimated while the daily average levels are same as NCEP data. Then PAR was estimated by assuming that PAR is 46% of total solar radiation (Pinkert & Laszlo 1992, Kirk 1994). The HOT project has measured noon PAR in unit of µEinstein/cm²/s at Station ALOHA during its monthly cruises since 1998. By using 550 nm to represent the whole PAR spectrum and therefore converting the measured PAR to Watt with 1 µEinstein/s = 0.2174 Watt, the measured noon PAR at HOT is 43±12% of the corresponding estimated shortwave, which is quite comparable with the 46% used here. The attenuation of PAR with depth in the water column is controlled by:

\[
\text{PAR}(z) = \text{PAR}_0 \cdot e^{-(kw+kchl\cdot Chl)z},
\]

where \( z \) is depth in m, \( \text{PAR}_0 \) is PAR level at sea surface, \( kw = 0.038 \) m\(^{-1} \) is the attenuation coefficient for seawater, \( kchl = 0.05 \) (mg Chl\(^{-1} \)) is the attenuation coefficient for chlorophyll (Chl) in mg Chl m\(^{-3} \).

(b) MLD: The criteria we used for MLD is the depth for a temperature change of 0.5°C from the sea surface (Levitus 1982). Another criteria also defined by Levitus (1982) is a potential density change of 0.125 kg m\(^{-3} \) from surface. But most time the difference in MLD defined by these 2 criteria is small at Station ALOHA (less than 10 m in 75% of cruises). We also found when large disagreements were obtained, usually the results from the first criteria are preferable. So HOT temperature data were used to compute MLD and linearly interpolated to each model time step.

(c) Temperature: is required by temperature effects in the model. Temperature was directly obtained from observations and linearly interpolated.

(d) Vertical velocity: is required by vertical advection of the physical scheme. Generally vertical velocity was computed by tracking the observed vertical isopycnal displacement. The first derivative of the displacement was the estimate for vertical velocity at that isopycnal depth. By linearly reducing vertical velocity to zero at the surface, the vertical profiles were constructed. Siegel et al. (1999) proposed the use of satellite-derived sea level anomaly (SLA) to address the mesoscale eddies. The basic idea is that upward isopycnal displacement occurs with negative SLA for cyclonic eddies, and downward displacement relates to positive SLA for anti-cyclonic eddies. AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data) provides satellite-derived SLA from Oct 1992 to present. The altimeter products were produced by Ssalto/Duacs and distributed by AVISO, with support from CNES. The data are available online at www.jason.oceanobs.com/. AVISO ‘delayed time updated series SLA’ data with 7-day intervals were used by the testbed. Then linear regression was used to estimate the relationship between HOT observed monthly isopycnal displacements and corresponding AVISO SLA. A 2.4 m displacement per 1 cm SLA was obtained (Siegel et al. 1999 estimated ~ 4 m displacement per 1 cm SLA in Sargasso Sea). This scale factor was then applied to all SLA data points to construct a vertical velocity time series with 7-day intervals, which was then linearly interpolated to each time step.

The vertical velocity estimated here may lead to neglect the transport incurred by Ekman pumping (Doney et al. 1996, Williams & Follows 1998). But it may not dominate the total nutrient transport (Siegel et al. 1999), and our model does not solve it.

(e) Vertical eddy diffusivity (Kz): is required for vertical diffusion in the physical scheme. The physical scheme assumes that all the state variables in the mixed layer are instantaneously mixed by averaging all the state variables in the mixed layer at every time step. Thus \( Kz \) values above MLD are not required. Below the MLD, \( Kz \) was constructed by reducing it vertically:

\[
Kz(z) = Kz_0 \cdot e^{-\alpha(z-MLD)},
\]

where \( z \) is depth in m (positive) below MLD, \( Kz_0 = 1.1 \times 10^{-4} \) m\(^2\)s\(^{-1} \) is the \( Kz \) at the bottom of mixed layer and \( \alpha = 0.01 \). This interpolation was suggested by J.R. Christian (University of Victoria, personal communication). The average \( Kz \) at 200 m obtained from this equation is \( 3.0 \times 10^{-5} \) m\(^2\)s\(^{-1} \), which is within the range of the estimates at Station ALOHA by Christian et al. (1997).

2.2. Assimilative Data and Cost Function

The cost function \( J \) is a metric to evaluate the difference between modeled results \( \hat{a} \) and observations \( a \). Here a lower cost function \( J \) value indicates a better fit of the model to observations. \( J \) summarizes the average Chi-squared value \( (\chi^2) \) of each data type:
\[ J = \sum_{m=1}^{M} \frac{1}{N_m} \sum_{n=1}^{N_m} \left( \frac{a_{m,n} - \hat{a}_{m,n}}{\sigma_m} \right)^2 = \sum_{m=1}^{M} \chi_m^2 , \]  

where \( m \) and \( n \) represent assimilative data types and data points, \( M \) and \( N \) are the total numbers of assimilative data types and data points; \( \sigma_m \) is the target error for data type \( m \) which is defined as

\[ \sigma_m = \frac{\epsilon_m}{q_m} , \]  

where \( q_m \) is the adjustment factor which will be discussed later; \( \hat{a}_{m,n} \) is the mean of the observations (Table S2), and \( \epsilon_m \) is fractional error level of the observation. To reasonably estimate \( \epsilon_m \), we assumed all properties in the mixed layer should be completely mixed, and a perfect measurement without any error should give a profile with same data value at every measured depth in the mixed layer. Thus we can use the ratio of standard deviation to mean value (coefficient of variation CV) of each vertical profile in the mixed layer to represent the error level for that measurement. The averaged CV of all the profiles was used as \( \epsilon_m \), and each data type has a different \( \epsilon \) value for each site (Table S2). This method does not apply to mesozooplankton and sediment trap data because there is only one data point per profile for them (discussed below). Their fractional error level was assumed to be 25%.

We included as many data types as possible in the assimilative data list whenever an observed data type could be related to a corresponding modeling output. Table S2 shows the 17 data types used for assimilation and their availability. Time series of each data type was constructed from the available observations.

The vertical profiles of most data were assimilated. However, because of the extreme difference in nutrient levels in and below the mixed layer, it is difficult for the model to fit the nutrient profiles. In addition, the nutrient level below the mixed layer is largely determined by physical instead of biological processes and also has less influence on the primary production compared to the nutrient level at surface. Even at HOT where the mixed layer (average ~ 60 m) is shallower than the euphotic zone (average ~ 105 m) and both biological and physical processes can influence nitrate levels below the mixed layer (Church et al. 2009), the observations show that more than half of primary production still occurs inside the mixed layer. Thus only the averaged CV of all the profiles was used as \( \epsilon_m \), and each data type has a different \( \epsilon \) value for each site (Table S2). This method does not apply to mesozooplankton and sediment trap data because there is only one data point per profile for them (discussed below). Their fractional error level was assumed to be 25%.

The vertical profiles of most data were assimilated. However, because of the extreme difference in nutrient levels in and below the mixed layer, it is difficult for the model to fit the nutrient profiles. In addition, the nutrient level below the mixed layer is largely determined by physical instead of biological processes and also has less influence on the primary production compared to the nutrient level at surface. Even at HOT where the mixed layer (average ~ 60 m) is shallower than the euphotic zone (average ~ 105 m) and both biological and physical processes can influence nitrate levels below the mixed layer (Church et al. 2009), the observations show that more than half of primary production still occurs inside the mixed layer. Thus only the averaged CV of all the profiles was used as \( \epsilon_m \), and each data type has a different \( \epsilon \) value for each site (Table S2). This method does not apply to mesozooplankton and sediment trap data because there is only one data point per profile for them (discussed below). Their fractional error level was assumed to be 25%.

We included as many data types as possible in the assimilative data list whenever an observed data type could be related to a corresponding modeling output. Table S2 shows the 17 data types used for assimilation and their availability. Time series of each data type was constructed from the available observations.

The model equivalent to observed PP is the sum of net PP of PHY, UN and TR minus their DOM excretion because the \(^{14}\)C tracer method used for the primary production measurements typically only captures the particulate production. It is possible that some \(^{14}\)C-labeled DOC could adsorb onto glass fiber filters, so that the measured PP may also include a portion of dissolved production (Karl et al. 1998). Since any DOC included in the PP measurement is difficult to quantify, we only used the particulate PP as the model equivalent, which may potentially lead to overestimation of particulate PP by the model after data assimilation.

Heterotrophic bacterial biomass was estimated from measured cell numbers with the conversion factor of 1.5 kg C/mol leucine (Ducklow 2000). Measured heterotrophic bacterial C production rates include 4, 40 and 3 vertical profiles of leucine incorporation rates on separate dates for AS, EQP and HOT respectively. A conversion factor of 1.5 kg C/mol leucine (Ducklow 2000) was used to derive the production rates at all 3 sites. Heterotrophic bacterial production is the only data type of all the assimilated data that is not publicly available (yet) from the HOT data site.

Semilabile DOM was derived by subtracting the refractory DOM from total DOM. The refractory fraction was estimated by assuming all DOM in the deep ocean (>1000 m) is refractory. At HOT, when one or 2 of the dissolved organic C, N or P data were not available, the observed average C:N:P in other years of 221:21:1 was used to reconstruct missing data from the available components (using C first if it available, or using N, or using P) in order to construct a better constraint for DOM composition.

The sediment trap data were collected at 150 m at HOT, which is inside the model domain. However, the sediment trap data were collected at ~ 800 m in both AS and EQP, below the model domain. Sinking speed and dissolution rate were defined as model parameters, and are constrained by POM concentration and sediment trap data through data assimilation, so that fluxes of detritus to the traps were projected by the model.
The adjustment factor \( q \) (Equation 18) was used to compute the target error by adjusting the fractional error level to \( e/q \) (Table S2) when considering the following, PP is not mixed in the mixed layer and therefore its error level was reduced. But the CV of the PP at EQP was already very low (6.7%), thus no adjustment was made. Additional uncertainties are related to conversion factors for estimating phytoplankton biomass, bacterial biomass and production so that the target errors for these variables were increased accordingly. The adjustment factors for the C/N/P components of DOM, POM and sinking particles (sediment traps) were set to \((1/3)^{1/2}\) so that the total contribution of these 3 C/N/P components to the cost function is considered to be one data type (note: \((1/2)^{1/2}\) for AS and EQP as only 2 components assimilated). In addition, we observed very low CVs (< 5%) for phosphate at AS and nitrate, phosphate and DOC at EQP, and they were increased to about 10% in order to avoid high weights on fitting these 3 observational types versus other types.

2.3. Adjoint Model

We used auto-differential software to conduct code-to-code translation to automatically generate adjoint codes for the forward model. The ecosystem model was translated by TAPENADE 2.1 (http://www-sop.inria.fr/tropics/tapenade.html) developed by Institut National de Recherche en Informatique et en Automatique, France (INRIA). However, the original regional testbed framework, including the physical scheme and integration, used TMC developed by Giering & Kaminski (www.autodiff.com/tamc/) to generate adjoint code. We used TAPENADE instead of TMC because TAPENADE can be run locally while TMC requires the remote access and may become unavailable occasionally. It is also found that TAPENADE is sometimes more reliable than TMC for our work. The performance of the 2 programs did not show much difference.

Nevertheless, both programs are based on same scheme, which we briefly describe in Appendix B. The description is based on Lawson et al. (1995), Hascoët & Pascual (2004) and Giering (1999).

2.4. Optimization Procedure: Limited-memory Quasi-Newton Method

The testbed includes a Fortran program, M1QN3 3.1, developed by J.C. Gilbert & C. Lemaréchal, which uses the gradient information calculated from the adjoint code to minimize the cost function. It is based on a limited-memory quasi-Newton scheme first proposed by Nocedal (1980) and then improved by Gilbert & Lemaréchal (1989). The method is briefly described in Appendix C based on the manual of M1QN3 and a note prepared by Galen Andrew for an informal tutorial at Microsoft Research (http://research.microsoft.com/users/galena/). This scheme is unconstrained and there is no penalty term to the cost function. Thus the parameters are basically optimized freely. But the parameters were logarithmically transformed before they were submitted to optimization, and transformed back exponentially after the optimization. By this way all the parameters are positive after the optimization.

The potential limitation of this optimization method is that theoretically it can only reach local minima, although we seek the solution for the global minimum. For an ecosystem application, most parameters can be initialized in a reasonable range based on literature search and rationality of ecological meanings, which may help to avoid local minima.

2.5. Selection of Optimizable Parameters and Uncertainty Analysis

Mathematically when the errors in the observations are assumed to be normally distributed, the final inverse Hessian matrix of the least-square cost function at its minimum can be used to analyze the uncertainties of the final optimized parameter values (Tziperman & Thacker 1989, Matear 1995), which approximately applies to our cost function. Denoting \( x \) as the realized optimal solution and \( x^* \) as the real solution then \((x - x^*)\) is the error of the parameters. Expanding the cost function \( J(x) \) about \( x^* \) yields

\[
J \approx J_{\text{min}} + 1/2 \left( x - x^* \right)^T B \left( x - x^* \right),
\]

which to the first order equals to zero because the cost function has zero gradient with respect to all parameters at the realized optimal solution, and the third and higher-order terms are neglected. If the neglected terms are sufficiently small, then the uncertainties \((x - x^*)\) are normally distributed with zero mean and with an error-covariance matrix defined as the inverse of the Hessian, \( H = B^{-1} \) (Tziperman & Thacker 1989). The error-covariance matrix provides information on the probability distribution of the optimal parameters. The square roots of diagonal elements of the matrix provide a measure of the width of the distribution for the different optimal parameters and represent the uncertainties of parameters. Therefore an optimized parameter is well constrained when its corresponding diagonal element in the inverse Hessian has low value. Also other elements not on the diagonal represent the covariance of the corresponding 2 parameters. A high value indicates the corresponding 2 parameters are highly correlated and can not be simultaneously optimized because a change to one of the parameters will be counteracted by a
change to the other with multiple pairs of parameter values producing indistinguishable results. As the Hessian matrix typically cannot be solved explicitly, a finite difference approximation is commonly used to generate the Hessian matrix.

In summary, an approximate inverse Hessian matrix $H$ has to be computed after the optimization, and parameters are judged to be optimizable only when the values of their corresponding diagonal elements in $H$ are low and no pair of parameters selected has high covariance, as indicated by their corresponding non-diagonal elements in $H$. The uncertainty of an optimized parameter is the square root of its diagonal element in $H$.

2.6. Default Parameter Values and Optimization

The default parameter values are shown in Table S3. These initial values were drawn from the literature and previous experience with the model.

The parameters were submitted to optimization. Some parameter values resulted from optimization may be scientifically unrealistic. Thus these parameters are removed from optimization and their values were set back to their default values.

The remaining parameters were subject to the inverse Hessian matrix test as described in Section 2.5, and those with high uncertainties (>50%) were excluded. However, even high uncertainties were associated with these optimized parameter values, they were useful to reduce the cost function. Thus these parameters are updated with the optimized values, although they would not be optimized any more. These parameters formed our ‘adjusted parameter list’.

Thus optimization was conducted iteratively with some parameters excluded at each step, either because of their scientifically unrealistic values or high uncertainties, until all the remaining parameters passed the inverse Hessian matrix test. This method usually ended with a small portion of the parameters, about 10 ~ 15 in all 80 parameters. These parameters were well constrained and formed our ‘constrained parameter list’.
LITERATURE CITED


Table S1. Symbols, units and descriptions of model parameters.

Following parameters are subject to optimization:

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_E$</td>
<td>d$^{-1}$</td>
<td>Arrhenius parameter for temperature function</td>
</tr>
<tr>
<td>$\mu_{PHY}$</td>
<td>d$^{-1}$</td>
<td>C-specific maximum phytoplankton (PHY) growth rate</td>
</tr>
<tr>
<td>$\alpha_{CHL}^{PHY}$</td>
<td>mol C (g Chl a)$^{-1}$ d$^{-1}$ (W m$^{-2}$)$^{-1}$</td>
<td>Initial slope of Photosynthesis vs. PAR curve of PHY</td>
</tr>
<tr>
<td>$\beta_{PHY}$</td>
<td>(W m$^{-2}$)$^{-1}$</td>
<td>Light inhibition parameter for photosynthesis of PHY</td>
</tr>
<tr>
<td>$v_{N_{REF.PHY}}^N$</td>
<td>mol N (mol C)$^{-1}$ d$^{-1}$</td>
<td>Maximum nitrogen assimilation rate per PHY carbon biomass</td>
</tr>
<tr>
<td>$k_{NH4}^{PHY}$</td>
<td>mmol m$^{-3}$</td>
<td>Ammonium half-saturation concentration for PHY growth</td>
</tr>
<tr>
<td>$k_{NO3}^{PHY}$</td>
<td>mmol m$^{-3}$</td>
<td>Nitrate half-saturation concentration for PHY growth</td>
</tr>
<tr>
<td>$v_{P_{REF.PHY}}^P$</td>
<td>mol P (mol C)$^{-1}$ d$^{-1}$</td>
<td>Maximum phosphorus assimilation rate per PHY carbon biomass</td>
</tr>
<tr>
<td>$k_{PO4}^{PHY}$</td>
<td>mmol m$^{-3}$</td>
<td>Phosphate half-saturation concentration for PHY growth</td>
</tr>
<tr>
<td>$\zeta_{NO3}$</td>
<td>mol C (mol N)$^{-1}$</td>
<td>Carbon requirement (respiration) to assimilate nitrate</td>
</tr>
<tr>
<td>$\theta$</td>
<td>g Chl a (mol N)$^{-1}$</td>
<td>Maximum chlorophyll a to nitrogen ratio</td>
</tr>
<tr>
<td>$ex_{PSV.PHY}$</td>
<td>d$^{-1}$</td>
<td>PHY passive excretion rate (per biomass)</td>
</tr>
<tr>
<td>$ex_{CHO.PHY}$</td>
<td>mmol C m$^{-3}$ d$^{-1}$</td>
<td>PHY carbon hydrate excretion rate (per growth rate)</td>
</tr>
<tr>
<td>$pom_{PHY}$</td>
<td>(mmol C m$^{-3}$)$^{-1}$d$^{-1}$</td>
<td>POM production rate by PHY aggregation</td>
</tr>
<tr>
<td>$\mu_{TR}$</td>
<td>d$^{-1}$</td>
<td>C-specific maximum Trichodesmium (TR) growth rate</td>
</tr>
<tr>
<td>$\alpha_{CHL}^{TR}$</td>
<td>mol C (g Chl a)$^{-1}$ d$^{-1}$ (W m$^{-2}$)$^{-1}$</td>
<td>Initial slope of Photosynthesis vs. PAR curve of TR</td>
</tr>
<tr>
<td>$\beta_{TR}$</td>
<td>(W m$^{-2}$)$^{-1}$</td>
<td>Light inhibition parameter for photosynthesis of TR</td>
</tr>
<tr>
<td>$v_{N_{REF.TR}}^N$</td>
<td>mol N (mol C)$^{-1}$ d$^{-1}$</td>
<td>Maximum nitrogen assimilation rate per TR carbon biomass</td>
</tr>
<tr>
<td>$k_{NH4}^{TR}$</td>
<td>mmol m$^{-3}$</td>
<td>Ammonium half-saturation concentration for TR growth</td>
</tr>
<tr>
<td>$k_{NO3}^{TR}$</td>
<td>mmol m$^{-3}$</td>
<td>Nitrate half-saturation concentration for TR growth</td>
</tr>
<tr>
<td>$v_{P_{REF.TR}}^P$</td>
<td>mol P (mol C)$^{-1}$ d$^{-1}$</td>
<td>Maximum phosphorus assimilation rate per TR carbon biomass</td>
</tr>
<tr>
<td>$k_{PO4}^{TR}$</td>
<td>mmol m$^{-3}$</td>
<td>PO4 half-saturation concentration for TR growth</td>
</tr>
<tr>
<td>$pick_{TR}^{PO4}$</td>
<td>d$^{-1}$</td>
<td>Deep phosphate pick-up (recovery) rate of TR</td>
</tr>
<tr>
<td>$\zeta_{N\text{FIX}}$</td>
<td>mol C (mol N)$^{-1}$</td>
<td>Carbon requirement (respiration) to nitrogen fixation</td>
</tr>
<tr>
<td>$ex_{PSV.TR}$</td>
<td>d$^{-1}$</td>
<td>TR passive excretion rate (per biomass)</td>
</tr>
<tr>
<td>$ex_{N\text{FIX}.TR}$</td>
<td>mmol C m$^{-3}$ d$^{-1}$</td>
<td>TR DON and ammonium release rate per nitrogen fixation</td>
</tr>
<tr>
<td>$ex_{CHO.TR}$</td>
<td>mmol C m$^{-3}$ d$^{-1}$</td>
<td>TR carbon hydrate excretion rate (per growth rate)</td>
</tr>
<tr>
<td>$pom_{TR}$</td>
<td>(mmol C m$^{-3}$)$^{-1}$d$^{-1}$</td>
<td>POM production rate by TR</td>
</tr>
<tr>
<td>$\mu_{UN}$</td>
<td>d$^{-1}$</td>
<td>C-specific maximum Unicellular N2-fixers (UN) growth rate</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{CHL_UN}$</td>
<td>Initial slope of Photosynthesis vs. PAR curve of UN</td>
<td></td>
</tr>
<tr>
<td>$k_{DOM}$</td>
<td>Half-saturation concentration of available DOC for heterotrophic bacteria (BA) uptake</td>
<td></td>
</tr>
<tr>
<td>$r_{SDOM}$</td>
<td>Parameter for SDOM's lability</td>
<td></td>
</tr>
<tr>
<td>$\mu_{BA}$</td>
<td>Maximum BA growth rate</td>
<td></td>
</tr>
<tr>
<td>$b_{RESP}$</td>
<td>Parameter control BA active respiration rate versus production</td>
<td></td>
</tr>
<tr>
<td>$ex_{ADJ:_BA}$</td>
<td>Bacteria extra semilabile DOC excretion rate</td>
<td></td>
</tr>
<tr>
<td>$remi_{BA}$</td>
<td>Bacteria inorganic nutrients regeneration rate</td>
<td></td>
</tr>
<tr>
<td>$ex_{REFR:_BA}$</td>
<td>Bacteria refractory DOC production rate</td>
<td></td>
</tr>
<tr>
<td>$f_{SLCT:_BA}$</td>
<td>BA selection strength on SDOM</td>
<td></td>
</tr>
<tr>
<td>$resp_{BA}$</td>
<td>Bacterial basal respiration rate</td>
<td></td>
</tr>
<tr>
<td>$r_{min:_BA}$</td>
<td>Bacteria minimum active respiration rate</td>
<td></td>
</tr>
<tr>
<td>$r_{max:_BA}$</td>
<td>Bacteria maximum active respiration rate</td>
<td></td>
</tr>
<tr>
<td>$morts_{BA}$</td>
<td>Bacteria mortality rate</td>
<td></td>
</tr>
<tr>
<td>$\mu_{PRT}$</td>
<td>Protozoa (PRT) maximum growth rate</td>
<td></td>
</tr>
<tr>
<td>$g_{PHY}$</td>
<td>Half-saturation density of PHY in PRT grazing function</td>
<td></td>
</tr>
<tr>
<td>$g_{BA}$</td>
<td>Half-saturation density of BA in PRT grazing function</td>
<td></td>
</tr>
<tr>
<td>$ex_{PRT}$</td>
<td>Total DOM excretion rate per PRT gross growth</td>
<td></td>
</tr>
<tr>
<td>$f_{ex:_PRT}$</td>
<td>Fraction of labile DOC of total PRT DOC excretion</td>
<td></td>
</tr>
<tr>
<td>$resp_{PRT}$</td>
<td>PRT basal respiration rate</td>
<td></td>
</tr>
<tr>
<td>$respi_{PRT}$</td>
<td>PRT active respiration rate</td>
<td></td>
</tr>
<tr>
<td>$ex_{ADJ:_PRT}$</td>
<td>PRT extra SDOM excretion rate</td>
<td></td>
</tr>
<tr>
<td>$remi_{PRT}$</td>
<td>PRT inorganic nutrients regeneration rate</td>
<td></td>
</tr>
<tr>
<td>$pom_{PRT}$</td>
<td>POM production rate per PRT gross growth</td>
<td></td>
</tr>
<tr>
<td>$\mu_{MZ}$</td>
<td>Maximum metazoa (MZ) growth rate</td>
<td></td>
</tr>
<tr>
<td>$g_{PRT}$</td>
<td>Half-saturation density of PRT in MZ grazing function</td>
<td></td>
</tr>
<tr>
<td>$g_{TR}$</td>
<td>Half-saturation density of TR in MZ grazing function</td>
<td></td>
</tr>
<tr>
<td>$ex_{MZ}$</td>
<td>Total DOM excretion rate per MZ gross growth</td>
<td></td>
</tr>
<tr>
<td>$f_{ex:_MZ}$</td>
<td>Fraction of labile DOC of total MZ DOC excretion</td>
<td></td>
</tr>
<tr>
<td>$resp_{MZ}$</td>
<td>MZ basal respiration</td>
<td></td>
</tr>
<tr>
<td>$respi_{MZ}$</td>
<td>MZ active respiration</td>
<td></td>
</tr>
<tr>
<td>$ex_{ADJ:_MZ}$</td>
<td>MZ extra semilabile DOM excretion rate</td>
<td></td>
</tr>
<tr>
<td>$remi_{MZ}$</td>
<td>MZ inorganic nutrients regeneration rate</td>
<td></td>
</tr>
<tr>
<td>$pom_{MZ}$</td>
<td>POM production rate per MZ gross growth</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Symbol</td>
<td>Equal to</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>ex&lt;sub&gt;REFR.MZ&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>remv&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f&lt;sub&gt;SDOM.MZ&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f&lt;sub&gt;POM.MZ&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ex&lt;sub&gt;REFR.SDOM&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;N,REFR&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;P,REFR&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;N,POM&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;P,POM&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r&lt;sub&gt;nrf&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prf&lt;sub&gt;N&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prf&lt;sub&gt;P&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wnsv&lt;sub&gt;sv&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>diss&lt;sub&gt;s&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Following parameters are assigned to equal values of other parameters**

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Equal to</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>β&lt;sub&gt;UN&lt;/sub&gt;</td>
<td>β&lt;subPHY&lt;/sub&gt;</td>
<td>(W m&lt;sup&gt;-2&lt;/sup&gt;)&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Light inhibition parameter for photosynthesis of UN</td>
</tr>
<tr>
<td>ν&lt;sub&gt;REF.UN&lt;/sub&gt;</td>
<td>ν&lt;sub&gt;REF.PHY&lt;/sub&gt;</td>
<td>mol N (mol C)&lt;sup&gt;-1&lt;/sup&gt; d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Maximum nitrogen assimilation rate per UN carbon biomass</td>
</tr>
<tr>
<td>k&lt;sub&gt;NH4.UN&lt;/sub&gt;</td>
<td>k&lt;sub&gt;NH4.PHY&lt;/sub&gt;</td>
<td>mmol m&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>NH4 half-saturation concentration for UN growth</td>
</tr>
<tr>
<td>k&lt;sub&gt;NO3.UN&lt;/sub&gt;</td>
<td>k&lt;sub&gt;NO3.PHY&lt;/sub&gt;</td>
<td>mmol m&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>NO3 half-saturation concentration for UN growth</td>
</tr>
<tr>
<td>ν&lt;sub&gt;REF.POM&lt;/sub&gt;</td>
<td>ν&lt;sub&gt;REF.PHY&lt;/sub&gt;</td>
<td>mol P (mol C)&lt;sup&gt;-1&lt;/sup&gt; d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Maximum phosphorus assimilation rate per UN carbon biomass</td>
</tr>
<tr>
<td>k&lt;sub&gt;PO4.UN&lt;/sub&gt;</td>
<td>k&lt;sub&gt;PO4.PHY&lt;/sub&gt;</td>
<td>mmol m&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>Phosphate half-saturation concentration for UN growth</td>
</tr>
<tr>
<td>ex&lt;sub&gt;PSV.UN&lt;/sub&gt;</td>
<td>ex&lt;sub&gt;PSV.PHY&lt;/sub&gt;</td>
<td>d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>UN passive excretion rate (per biomass)</td>
</tr>
<tr>
<td>ex&lt;sub&gt;CHD.UN&lt;/sub&gt;</td>
<td>ex&lt;sub&gt;CHD.PHY&lt;/sub&gt;</td>
<td></td>
<td>UN carbon hydrate excretion rate (per growth rate)</td>
</tr>
<tr>
<td>ex&lt;sub&gt;NTIX.UN&lt;/sub&gt;</td>
<td>ex&lt;sub&gt;NTIX.PHY&lt;/sub&gt;</td>
<td></td>
<td>UN DON and ammonium release rate per nitrogen fixation</td>
</tr>
<tr>
<td>pom&lt;sub&gt;UN&lt;/sub&gt;</td>
<td>pom&lt;sub&gt;PHY&lt;/sub&gt;</td>
<td>(mmol C m&lt;sup&gt;-3&lt;/sup&gt;)&lt;sup&gt;-1&lt;/sup&gt;d&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>POM production rate by UN aggregation</td>
</tr>
<tr>
<td>g&lt;sub&gt;UN&lt;/sub&gt;</td>
<td>g&lt;sub&gt;PHY&lt;/sub&gt;</td>
<td></td>
<td>PRT grazing parameter on UN</td>
</tr>
</tbody>
</table>

**Following parameters are assigned fixed values**

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Equal to</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>q&lt;sub&gt;N.min.PHY&lt;/sub&gt;</td>
<td>0.034</td>
<td>mol N (mol C)&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Minimum N/C ratio of PHY</td>
</tr>
<tr>
<td>q&lt;sub&gt;P.min.PHY&lt;/sub&gt;</td>
<td>0.00188</td>
<td>mol P (mol C)&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Minimum P/C ratio of PHY</td>
</tr>
<tr>
<td>q&lt;sub&gt;N.ref.PHY&lt;/sub&gt;</td>
<td>0.15</td>
<td>mol N (mol C)&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>Reference (Redfield) N/C ratio of PHY</td>
</tr>
<tr>
<td>$q_{C_{P, PHY}}^C$</td>
<td>0.0094</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Reference (Redfield) P/C ratio of PHY</td>
</tr>
<tr>
<td>$q_{N_{max, PHY}}^C$</td>
<td>0.17</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Maximum N/C ratio of PHY</td>
</tr>
<tr>
<td>$q_{P_{max, PHY}}^C$</td>
<td>0.0169</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Maximum P/C ratio of PHY</td>
</tr>
<tr>
<td>$q_{N_{min, TR}}^C$</td>
<td>0.12</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Minimum N/C ratio of TR</td>
</tr>
<tr>
<td>$q_{P_{min, TR}}^C$</td>
<td>0.001</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Minimum P/C ratio of TR</td>
</tr>
<tr>
<td>$q_{N_{min, UN}}^C$</td>
<td>0.16</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Reference N/C ratio of UN</td>
</tr>
<tr>
<td>$q_{P_{min, UN}}^C$</td>
<td>0.016</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Minimum P/C ratio of UN</td>
</tr>
<tr>
<td>$q_{C_{P_{sdf, UN}}}$</td>
<td>0.0035</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Reference ratio of UN</td>
</tr>
<tr>
<td>$q_{N_{max, UN}}^C$</td>
<td>0.2</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Maximum N/C ratio of UN</td>
</tr>
<tr>
<td>$q_{P_{max, UN}}^C$</td>
<td>0.006</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Maximum P/C ratio of UN</td>
</tr>
<tr>
<td>$q_{N_{BA}}^C$</td>
<td>0.18</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Reference N/C ratio of BA</td>
</tr>
<tr>
<td>$q_{P_{BA}}^C$</td>
<td>0.02</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Reference P/C ratio of BA</td>
</tr>
<tr>
<td>$q_{N_{PRT}}^C$</td>
<td>0.2</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Reference N/C ratio of PRT</td>
</tr>
<tr>
<td>$q_{P_{PRT}}^C$</td>
<td>0.022</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Reference P/C ratio of PRT</td>
</tr>
<tr>
<td>$q_{N_{MZ}}^C$</td>
<td>0.2</td>
<td>mol N (mol C)$^{-1}$</td>
<td>Reference N/C ratio of MZ</td>
</tr>
<tr>
<td>$q_{P_{MZ}}^C$</td>
<td>0.008</td>
<td>mol P (mol C)$^{-1}$</td>
<td>Reference P/C ratio of MZ</td>
</tr>
</tbody>
</table>

Table S2. Variables used to estimate the target error of the cost function as defined in Equation 18 at 3 study sites AS, EQP and HOT.
<table>
<thead>
<tr>
<th></th>
<th>0.50</th>
<th>0.50</th>
<th>1.00</th>
<th>0.75</th>
<th>1.00</th>
<th>1.00</th>
<th>0.50</th>
<th>0.50</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adjustment Factor</strong> $q$</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>1.00</td>
<td>0.50</td>
<td>0.50</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Adjusted Error</strong> $e/q$</td>
<td>9.0%</td>
<td>9.0%</td>
<td>25%</td>
<td>12%</td>
<td>10%</td>
<td>6.7%</td>
<td>36%</td>
<td>36%</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Target Error</strong> $\sigma$</td>
<td>0.440</td>
<td>0.060</td>
<td>0.056</td>
<td>0.018</td>
<td>0.023</td>
<td>0.057</td>
<td>0.172</td>
<td>0.025</td>
<td>1.88</td>
</tr>
</tbody>
</table>

**HOT**

<table>
<thead>
<tr>
<th></th>
<th>0.0046</th>
<th>0.0311</th>
<th>0.184</th>
<th>0.081</th>
<th>0.124</th>
<th>0.379</th>
<th>0.310</th>
<th>0.0464</th>
<th>32.6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Mean</strong> $\overline{\alpha}$</td>
<td>0.0046</td>
<td>0.0311</td>
<td>0.184</td>
<td>0.081</td>
<td>0.124</td>
<td>0.379</td>
<td>0.310</td>
<td>0.0464</td>
<td>32.6</td>
</tr>
<tr>
<td><strong>Fractional Error</strong> $e$</td>
<td>63%</td>
<td>27%</td>
<td>25%</td>
<td>30%</td>
<td>18%</td>
<td>23%</td>
<td>5.6%</td>
<td>3.1%</td>
<td>7.8%</td>
</tr>
<tr>
<td><strong>Adjustment Factor</strong> $q$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>4.00</td>
<td>0.50</td>
<td>0.20</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Adjusted Error</strong> $e/q$</td>
<td>63%</td>
<td>27%</td>
<td>25%</td>
<td>40%</td>
<td>18%</td>
<td>5.8%</td>
<td>11%</td>
<td>16%</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Target Error</strong> $\sigma$</td>
<td>0.0029</td>
<td>0.0084</td>
<td>0.046</td>
<td>0.032</td>
<td>0.022</td>
<td>0.022</td>
<td>0.035</td>
<td>0.0072</td>
<td>4.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>sDON</th>
<th>sDOP</th>
<th>POC</th>
<th>PON</th>
<th>POP</th>
<th>STc</th>
<th>STn</th>
<th>STp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ss DON</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HOT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2.90</th>
<th>4.72</th>
<th>0.920</th>
<th>1.11</th>
<th>0.134</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Mean</strong> $\overline{\alpha}$</td>
<td>2.90</td>
<td>4.72</td>
<td>0.920</td>
<td>1.11</td>
<td>0.134</td>
</tr>
<tr>
<td><strong>Fractional Error</strong> $e$</td>
<td>13%</td>
<td>15%</td>
<td>15%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Adjustment Factor</strong> $q$</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>Adjusted Error</strong> $e/q$</td>
<td>18%</td>
<td>21%</td>
<td>21%</td>
<td>35%</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Target Error</strong> $\sigma$</td>
<td>0.531</td>
<td>0.997</td>
<td>0.194</td>
<td>0.391</td>
<td>0.047</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1.95</th>
<th>0.486</th>
<th>0.403</th>
<th>0.0553</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Mean</strong> $\overline{\alpha}$</td>
<td>1.95</td>
<td>0.486</td>
<td>0.403</td>
<td>0.0553</td>
</tr>
<tr>
<td><strong>Fractional Error</strong> $e$</td>
<td>13%</td>
<td>7.7%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Adjustment Factor</strong> $q$</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>Adjusted Error</strong> $e/q$</td>
<td>18%</td>
<td>11%</td>
<td>35%</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Target Error</strong> $\sigma$</td>
<td>0.357</td>
<td>0.053</td>
<td>0.142</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2.12</th>
<th>0.101</th>
<th>1.52</th>
<th>0.219</th>
<th>0.0089</th>
<th>2.66</th>
<th>0.334</th>
<th>0.0094</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed Mean</strong> $\overline{\alpha}$</td>
<td>2.12</td>
<td>0.101</td>
<td>1.52</td>
<td>0.219</td>
<td>0.0089</td>
<td>2.66</td>
<td>0.334</td>
<td>0.0094</td>
</tr>
<tr>
<td><strong>Fractional Error</strong> $e$</td>
<td>11%</td>
<td>11%</td>
<td>9.0%</td>
<td>7.9%</td>
<td>15%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Adjustment Factor</strong> $q$</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Adjusted Error</strong> $e/q$</td>
<td>19%</td>
<td>19%</td>
<td>16%</td>
<td>14%</td>
<td>26%</td>
<td>43%</td>
<td>43%</td>
<td>43%</td>
</tr>
<tr>
<td><strong>Target Error</strong> $\sigma$</td>
<td>0.402</td>
<td>0.0192</td>
<td>0.236</td>
<td>0.030</td>
<td>0.0023</td>
<td>1.15</td>
<td>0.144</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

**NO3**: Nitrate (in mixed layer), mmol m\(^{-3}\); **PO4**: phosphate (in mixed layer), mmol m\(^{-3}\); **PHYn**: phytoplankton nitrogen biomass (exclude *Trichodesmium* spp.), mmol N m\(^{-3}\); **CHL**: Chlorophyll a, mg m\(^{-3}\); **PrPr**: particulate primary production, mmol C m\(^{-3}\) day\(^{-1}\); **MZc**: metazoan zooplankton carbon biomass, mmol C m\(^{-3}\); **BAc**: heterotrophic bacterial carbon biomass, mmol C m\(^{-3}\); **BP**: heterotrophic bacterial production, mmol C m\(^{-3}\) day\(^{-1}\); **sDOC**, **sDON** & **sDOP**: semilabile dissolved organic carbon, nitrogen and phosphorus, mmol m\(^{-3}\); **POC**, **PON** & **POP**: particulate organic carbon, nitrogen and phosphorus, mmol m\(^{-3}\); **STc**, **STn** & **STp**: particle carbon, nitrogen and phosphorus flux collected by sediment traps at 800m for AS and EQP and at 150m for HOT, mmol m\(^{-2}\) day\(^{-1}\).
Table S3. Default model parameter values and optimized model parameter values after data assimilation. Also the confidence ranges of optimized parameters (if constrained). The ‘Marks’ indicates if the parameter was optimized and well-constrained (CS), was optimized but not well-constrained (OP) or was unable to be optimized (empty) by data assimilation. See Table S1 for parameter definitions and units. Also listed are the gradients of parameters with respect to the cost function after data assimilation, which defined as: when a parameter multiplied by \( e^{\Delta a} \) (\( \Delta a \) is infinitely small), the change of the total cost divided by \( \Delta a \). Note \( e^{\Delta a} = \Delta a \) when \( \Delta a \) is smaller enough. Thus, for example, a 10% change of a parameter (\( \Delta a = 10\% \)) will roughly result in a total cost change equaling 10% of the corresponding gradient.

(a) ARABIAN SEA (AS):

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Default Value</th>
<th>Optimized Value</th>
<th>Confidence Range</th>
<th>Gradient</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_e )</td>
<td>4000</td>
<td>0.3850</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_{PHY} )</td>
<td>5.0</td>
<td>7.1098</td>
<td>6.100 – 8.287</td>
<td>0.0268</td>
<td>CS</td>
</tr>
<tr>
<td>( \alpha_{CHL}^{PHY} )</td>
<td>0.5</td>
<td>0.4545</td>
<td>0.415 – 0.499</td>
<td>0.0519</td>
<td>CS</td>
</tr>
<tr>
<td>( \beta_{PHY} )</td>
<td>0.002</td>
<td></td>
<td></td>
<td>-1.0800</td>
<td></td>
</tr>
<tr>
<td>( v_{NH4,PHY}^N )</td>
<td>0.3</td>
<td></td>
<td></td>
<td>-2.2100</td>
<td></td>
</tr>
<tr>
<td>( k_{NH4,PHY} )</td>
<td>0.25</td>
<td></td>
<td></td>
<td>1.3100</td>
<td></td>
</tr>
<tr>
<td>( k_{NO3,PHY} )</td>
<td>1.0</td>
<td>1.2500</td>
<td>-0.1310</td>
<td></td>
<td>OP</td>
</tr>
<tr>
<td>( v_{REF,PHY}^p )</td>
<td>0.02</td>
<td>0.0058</td>
<td>-3.4300</td>
<td></td>
<td>OP</td>
</tr>
<tr>
<td>( k_{PO4,PHY} )</td>
<td>0.06</td>
<td></td>
<td>0.3660</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \zeta_{NO3} )</td>
<td>2.0</td>
<td></td>
<td>2.5600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta )</td>
<td>2.0</td>
<td>0.8342</td>
<td>0.713 – 0.976</td>
<td>0.0274</td>
<td>CS</td>
</tr>
<tr>
<td>( ex_{PSV,PHY} )</td>
<td>0.05</td>
<td></td>
<td>3.0800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ex_{CHO,PHY} )</td>
<td>0.05</td>
<td></td>
<td>-0.0288</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( pom_{PHY} )</td>
<td>0.04</td>
<td>0.0247</td>
<td>0.0204 – 0.0299</td>
<td>0.0304</td>
<td>CS</td>
</tr>
<tr>
<td>( k_{DOM} )</td>
<td>0.5</td>
<td></td>
<td>1.0600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_{SDOM} )</td>
<td>0.005</td>
<td>0.0068</td>
<td>0.0048 – 0.0095</td>
<td>-0.0213</td>
<td>CS</td>
</tr>
<tr>
<td>( \mu_{BA} )</td>
<td>2.0</td>
<td>1.1886</td>
<td>1.009 – 1.401</td>
<td>-0.0302</td>
<td>CS</td>
</tr>
<tr>
<td>( b_{RESP} )</td>
<td>0.28</td>
<td></td>
<td>0.5130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ex_{ADJ,BA} )</td>
<td>2.0</td>
<td></td>
<td>-0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( remi_{BA} )</td>
<td>6.0</td>
<td></td>
<td>0.5970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ex_{REFR,BA} )</td>
<td>0.018</td>
<td>0.0310</td>
<td>1.3700</td>
<td></td>
<td>OP</td>
</tr>
<tr>
<td>( f_{SLCT,BA} )</td>
<td>0.25</td>
<td></td>
<td>-0.4040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( resp_{BA}^p )</td>
<td>0.01</td>
<td></td>
<td>0.1550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( t_{min,BA}^A )</td>
<td>0.35</td>
<td></td>
<td>-0.5770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( t_{max,BA}^A )</td>
<td>0.80</td>
<td>0.7141</td>
<td>0.663 – 0.769</td>
<td>0.0187</td>
<td>CS</td>
</tr>
<tr>
<td>Variable</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
<td>Category</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>mort&lt;sub&gt;BA&lt;/sub&gt;</td>
<td>0.021</td>
<td>0.0310</td>
<td>0.8040</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>µ&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>1.5</td>
<td>1.0938</td>
<td>0.827 – 1.446</td>
<td>0.1420</td>
<td>CS</td>
</tr>
<tr>
<td>g&lt;sub&gt;PHY&lt;/sub&gt;</td>
<td>2.0</td>
<td>3.0012</td>
<td>2.605 – 3.458</td>
<td>0.0757</td>
<td>CS</td>
</tr>
<tr>
<td>g&lt;sub&gt;BA&lt;/sub&gt;</td>
<td>2.0</td>
<td>1.9570</td>
<td>1.700 – 2.294</td>
<td>-0.0432</td>
<td>CS</td>
</tr>
<tr>
<td>ex&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f&lt;sub&gt;ex,PRT&lt;/sub&gt;</td>
<td>0.75</td>
<td>0.9000</td>
<td></td>
<td>4.6100</td>
<td>OP</td>
</tr>
<tr>
<td>resp&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>resp&lt;sub&gt;BA&lt;/sub&gt;</td>
<td>0.5</td>
<td>0.3763</td>
<td>0.276 – 0.514</td>
<td>-0.1190</td>
<td>CS</td>
</tr>
<tr>
<td>ex&lt;sub&gt;ADJ,PRT&lt;/sub&gt;</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.1180</td>
<td></td>
</tr>
<tr>
<td>remi&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>4.0</td>
<td></td>
<td></td>
<td>-0.0221</td>
<td></td>
</tr>
<tr>
<td>pom&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>0.027</td>
<td>0.0280</td>
<td></td>
<td>0.0163</td>
<td>OP</td>
</tr>
<tr>
<td>µ&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>1.3</td>
<td>1.1277</td>
<td>0.977 – 1.302</td>
<td>-0.0072</td>
<td>CS</td>
</tr>
<tr>
<td>g&lt;sub&gt;PRT&lt;/sub&gt;</td>
<td>1.0</td>
<td>1.4196</td>
<td>1.016 – 1.984</td>
<td>0.0252</td>
<td>CS</td>
</tr>
<tr>
<td>ex&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>0.20</td>
<td></td>
<td></td>
<td>2.2300</td>
<td></td>
</tr>
<tr>
<td>f&lt;sub&gt;ex,MZ&lt;/sub&gt;</td>
<td>0.75</td>
<td>0.9000</td>
<td></td>
<td>2.2100</td>
<td>OP</td>
</tr>
<tr>
<td>resp&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>0.03</td>
<td></td>
<td></td>
<td>0.6420</td>
<td></td>
</tr>
<tr>
<td>resp&lt;sub&gt;BA&lt;/sub&gt;</td>
<td>0.25</td>
<td></td>
<td></td>
<td>2.3500</td>
<td></td>
</tr>
<tr>
<td>ex&lt;sub&gt;ADJ,MZ&lt;/sub&gt;</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>remi&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>4.0</td>
<td></td>
<td></td>
<td>0.0930</td>
<td></td>
</tr>
<tr>
<td>pom&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>0.12</td>
<td></td>
<td></td>
<td>1.1400</td>
<td></td>
</tr>
<tr>
<td>ex&lt;sub&gt;REFR,MZ&lt;/sub&gt;</td>
<td>0.02</td>
<td></td>
<td></td>
<td>0.2700</td>
<td></td>
</tr>
<tr>
<td>remv&lt;sub&gt;MZ&lt;/sub&gt;</td>
<td>0.81</td>
<td>0.3800</td>
<td></td>
<td>1.8600</td>
<td>OP</td>
</tr>
<tr>
<td>f&lt;sub&gt;SDOM,MZ&lt;/sub&gt;</td>
<td>0.2</td>
<td>0.1400</td>
<td></td>
<td>-0.2270</td>
<td>OP</td>
</tr>
<tr>
<td>f&lt;sub&gt;POM,HZ&lt;/sub&gt;</td>
<td>0.2</td>
<td>0.0940</td>
<td></td>
<td>0.2450</td>
<td>OP</td>
</tr>
<tr>
<td>ex&lt;sub&gt;REFR,SDOM&lt;/sub&gt;</td>
<td>0.0009</td>
<td></td>
<td></td>
<td>0.1620</td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;C,REFR&lt;/sub&gt;</td>
<td>0.05</td>
<td></td>
<td></td>
<td>1.1100</td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;P,REFR&lt;/sub&gt;</td>
<td>0.0007</td>
<td></td>
<td></td>
<td>-0.0755</td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;C,N,POM&lt;/sub&gt;</td>
<td>0.12</td>
<td></td>
<td></td>
<td>1.6400</td>
<td></td>
</tr>
<tr>
<td>q&lt;sub&gt;C,P,POM&lt;/sub&gt;</td>
<td>0.01</td>
<td></td>
<td></td>
<td>-1.1600</td>
<td></td>
</tr>
<tr>
<td>r&lt;sub&gt;nof&lt;/sub&gt;</td>
<td>0.15</td>
<td></td>
<td></td>
<td>-0.4150</td>
<td></td>
</tr>
<tr>
<td>prf&lt;sub&gt;N&lt;/sub&gt;</td>
<td>1.05</td>
<td></td>
<td></td>
<td>-2.7800</td>
<td></td>
</tr>
<tr>
<td>prf&lt;sub&gt;P&lt;/sub&gt;</td>
<td>1.10</td>
<td></td>
<td></td>
<td>1.4200</td>
<td></td>
</tr>
<tr>
<td>wmsv</td>
<td>30.0</td>
<td></td>
<td></td>
<td>-1.6900</td>
<td></td>
</tr>
<tr>
<td>diss</td>
<td>0.1</td>
<td>0.1215</td>
<td>0.114 – 0.129</td>
<td>-0.0117</td>
<td>CS</td>
</tr>
</tbody>
</table>
(b) Equatorial Pacific (EQP):

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Default Value</th>
<th>Optimized Value</th>
<th>Confidence Range</th>
<th>Gradient</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_E$</td>
<td>4000</td>
<td>2000.0</td>
<td></td>
<td>10.5000</td>
<td>OP</td>
</tr>
<tr>
<td>$\mu_{PHY}$</td>
<td>3.5</td>
<td>3.3763</td>
<td>3.089 – 3.691</td>
<td>-8.1500</td>
<td>CS</td>
</tr>
<tr>
<td>$\alpha_{CHL_{PHY}}$</td>
<td>2.0</td>
<td>1.9044</td>
<td>1.652 – 2.196</td>
<td>-1.3100</td>
<td>CS</td>
</tr>
<tr>
<td>$\beta_{PHY}$</td>
<td>0.001</td>
<td></td>
<td></td>
<td>-4.7100</td>
<td></td>
</tr>
<tr>
<td>$v^N_{REF_{PHY}}$</td>
<td>0.3</td>
<td>0.3393</td>
<td>0.299 – 0.385</td>
<td>-2.6300</td>
<td>CS</td>
</tr>
<tr>
<td>$k^{NH4}_{PHY}$</td>
<td>0.25</td>
<td>0.1900</td>
<td></td>
<td>0.2210</td>
<td>OP</td>
</tr>
<tr>
<td>$k^{NO3}_{PHY}$</td>
<td>4.0</td>
<td></td>
<td></td>
<td>-7.1900</td>
<td></td>
</tr>
<tr>
<td>$v^P_{REF_{PHY}}$</td>
<td>0.02</td>
<td>0.0140</td>
<td></td>
<td>-2.7200</td>
<td>OP</td>
</tr>
<tr>
<td>$k^{PO4}_{PHY}$</td>
<td>0.1</td>
<td></td>
<td></td>
<td>0.2870</td>
<td></td>
</tr>
<tr>
<td>$\xi_{NO3}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>3.4700</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>2.0</td>
<td>3.4000</td>
<td></td>
<td>-6.7600</td>
<td>OP</td>
</tr>
<tr>
<td>$e_{PSV_{PHY}}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>3.0400</td>
<td></td>
</tr>
<tr>
<td>$e_{CHO_{PHY}}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>-0.0157</td>
<td></td>
</tr>
<tr>
<td>$pom_{PHY}$</td>
<td>0.04</td>
<td>0.1300</td>
<td></td>
<td>2.6800</td>
<td>OP</td>
</tr>
<tr>
<td>$k_{DOM}$</td>
<td>0.5</td>
<td>0.5500</td>
<td></td>
<td>-1.0500</td>
<td>OP</td>
</tr>
<tr>
<td>$r_{SDOM}$</td>
<td>0.005</td>
<td>0.0087</td>
<td>0.0064 – 0.0119</td>
<td>-0.3500</td>
<td>CS</td>
</tr>
<tr>
<td>$\mu_{BA}$</td>
<td>2.0</td>
<td>2.3000</td>
<td></td>
<td>0.4070</td>
<td>OP</td>
</tr>
<tr>
<td>$b_{RESP}$</td>
<td>0.15</td>
<td></td>
<td></td>
<td>1.6100</td>
<td></td>
</tr>
<tr>
<td>$e_{ADJ_{BA}}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>$remi_{BA}$</td>
<td>6.0</td>
<td></td>
<td></td>
<td>-1.6700</td>
<td></td>
</tr>
<tr>
<td>$e_{REFR_{BA}}$</td>
<td>0.02</td>
<td></td>
<td></td>
<td>0.6970</td>
<td></td>
</tr>
<tr>
<td>$f_{SLCT_{BA}}$</td>
<td>0.25</td>
<td></td>
<td></td>
<td>-0.4800</td>
<td></td>
</tr>
<tr>
<td>$resp^H_{BA}$</td>
<td>0.01</td>
<td></td>
<td></td>
<td>-0.0437</td>
<td></td>
</tr>
<tr>
<td>$r^A_{min_{BA}}$</td>
<td>0.35</td>
<td></td>
<td></td>
<td>-1.1700</td>
<td></td>
</tr>
<tr>
<td>$r^A_{max_{BA}}$</td>
<td>0.80</td>
<td>0.8665</td>
<td>0.825 – 0.910</td>
<td>2.3800</td>
<td>CS</td>
</tr>
<tr>
<td>$mort_{BA}$</td>
<td>0.02</td>
<td></td>
<td></td>
<td>0.5830</td>
<td></td>
</tr>
<tr>
<td>$\mu_{PRT}$</td>
<td>2.0</td>
<td>1.1477</td>
<td>0.982 – 1.341</td>
<td>10.6000</td>
<td>CS</td>
</tr>
<tr>
<td>$g_{PHY}$</td>
<td>0.67</td>
<td>0.9267</td>
<td>0.881 – 0.975</td>
<td>-5.3400</td>
<td>CS</td>
</tr>
<tr>
<td>$g_{BA}$</td>
<td>1.0</td>
<td>1.4000</td>
<td></td>
<td>1.9400</td>
<td>OP</td>
</tr>
<tr>
<td>$e_{PRT}$</td>
<td>0.20</td>
<td></td>
<td></td>
<td>-2.4800</td>
<td></td>
</tr>
<tr>
<td>$f_{ex_{PRT}}$</td>
<td>0.75</td>
<td></td>
<td></td>
<td>2.9400</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Value</td>
<td>Reference</td>
<td>Coefficient</td>
<td>Category</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------</td>
<td>------------</td>
<td>--------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>$\text{resp}_{\text{PRT}}^A$</td>
<td>0.01</td>
<td></td>
<td>4.3100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{resp}_{\text{PRT}}^B$</td>
<td>0.62</td>
<td>0.6052</td>
<td>0.575 – 0.637</td>
<td>-11.5000 CS</td>
<td></td>
</tr>
<tr>
<td>$\text{ex}_{\text{PRT}}$</td>
<td>2.0</td>
<td></td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{remi}_{\text{PRT}}$</td>
<td>4.0</td>
<td></td>
<td>0.4310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{pom}_{\text{PRT}}$</td>
<td>0.027</td>
<td>0.0260</td>
<td>-0.9280</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$\mu_{\text{MZ}}$</td>
<td>1.3</td>
<td>1.4000</td>
<td>-7.9500</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$g_{\text{PRT}}$</td>
<td>2.0</td>
<td>2.0831</td>
<td>1.821 – 2.382</td>
<td>10.1000 CS</td>
<td></td>
</tr>
<tr>
<td>$\text{ex}_{\text{MZ}}$</td>
<td>0.20</td>
<td></td>
<td>3.1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{\text{ex},\text{MZ}}$</td>
<td>0.75</td>
<td></td>
<td>0.0593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{resp}_{\text{MZ}}^A$</td>
<td>0.03</td>
<td></td>
<td>0.8750</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{resp}_{\text{MZ}}^B$</td>
<td>0.25</td>
<td>0.3683</td>
<td>0.307 – 0.442</td>
<td>5.1300 CS</td>
<td></td>
</tr>
<tr>
<td>$\text{ex}_{\text{PRT,MZ}}$</td>
<td>2.0</td>
<td></td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{remi}_{\text{MZ}}$</td>
<td>4.0</td>
<td></td>
<td>-0.1190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{pom}_{\text{MZ}}$</td>
<td>0.1</td>
<td></td>
<td>1.0800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{ex}_{\text{REFR,MZ}}$</td>
<td>0.02</td>
<td></td>
<td>0.2970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{remv}_{\text{MZ}}$</td>
<td>0.8</td>
<td>0.2000</td>
<td>-0.5800</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$f_{\text{SDOM, MZ}}$</td>
<td>0.2</td>
<td></td>
<td>-0.1190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{\text{POM, HZ}}$</td>
<td>0.35</td>
<td></td>
<td>-0.3280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{ex}_{\text{REFR,SDOM}}$</td>
<td>0.0009</td>
<td></td>
<td>0.3710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{\text{N,REFR}}^C$</td>
<td>0.05</td>
<td></td>
<td>1.0600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{\text{P,REFR}}^C$</td>
<td>0.0007</td>
<td></td>
<td>0.0329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{\text{N,POM}}^C$</td>
<td>0.12</td>
<td></td>
<td>-0.2640</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{\text{P,POM}}^C$</td>
<td>0.01</td>
<td></td>
<td>0.5140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{\text{norf}}$</td>
<td>0.15</td>
<td>0.1700</td>
<td>-1.8200</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$\text{prf}_N$</td>
<td>1.1</td>
<td>1.0500</td>
<td>5.4100</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$\text{prf}_R$</td>
<td>1.1</td>
<td>3.8000</td>
<td>0.2890</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$\text{wmsv}$</td>
<td>30.0</td>
<td></td>
<td>1.7500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{diss}$</td>
<td>0.1</td>
<td>0.1593</td>
<td>0.153 – 0.166</td>
<td>-0.0267 CS</td>
<td></td>
</tr>
</tbody>
</table>
(c) Hawaii Ocean Time-series Station ALOHA (HOT):

<table>
<thead>
<tr>
<th>Parameter Symbol</th>
<th>Default Value</th>
<th>Optimized Value</th>
<th>Confidence Range</th>
<th>Gradient</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_E$</td>
<td>4000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_{PHY}$</td>
<td>3.5</td>
<td>1.7097</td>
<td>1.531 – 1.909</td>
<td>-0.1040</td>
<td>CS</td>
</tr>
<tr>
<td>$\alpha_{CHL PHY}$</td>
<td>2.0</td>
<td>5.5127</td>
<td>4.206 – 7.225</td>
<td>-0.0288</td>
<td>CS</td>
</tr>
<tr>
<td>$\beta_{PHY}$</td>
<td>0.0005</td>
<td></td>
<td></td>
<td>1.8300</td>
<td></td>
</tr>
<tr>
<td>$v^N_{REF.PHY}$</td>
<td>0.3</td>
<td></td>
<td></td>
<td>-3.0400</td>
<td></td>
</tr>
<tr>
<td>$k_{NH4 PHY}$</td>
<td>0.01</td>
<td></td>
<td></td>
<td>0.6320</td>
<td></td>
</tr>
<tr>
<td>$k_{NO3 PHY}$</td>
<td>0.1</td>
<td></td>
<td></td>
<td>1.3100</td>
<td></td>
</tr>
<tr>
<td>$v^P_{REF.PHY}$</td>
<td>0.02</td>
<td>0.0101</td>
<td>0.0076 – 0.0136</td>
<td>0.0596</td>
<td>CS</td>
</tr>
<tr>
<td>$k_{PO4 PHY}$</td>
<td>0.0048</td>
<td></td>
<td></td>
<td>-0.8580</td>
<td></td>
</tr>
<tr>
<td>$\xi_{NO3}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>-0.9570</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>2.0</td>
<td>2.1289</td>
<td>1.816 – 2.496</td>
<td>-0.0498</td>
<td>CS</td>
</tr>
<tr>
<td>$e^{PSV.PHY}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>2.7800</td>
<td></td>
</tr>
<tr>
<td>$e^{CHO.PHY}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>0.4320</td>
<td></td>
</tr>
<tr>
<td>$pom_{PHY}$</td>
<td>0.038</td>
<td></td>
<td></td>
<td>2.0500</td>
<td></td>
</tr>
<tr>
<td>$\mu_{TR}$</td>
<td>3.0</td>
<td></td>
<td></td>
<td>2.0900</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{CHL TR}$</td>
<td>0.25</td>
<td></td>
<td></td>
<td>0.8590</td>
<td></td>
</tr>
<tr>
<td>$\beta_{TR}$</td>
<td>0.002</td>
<td></td>
<td></td>
<td>-0.5230</td>
<td></td>
</tr>
<tr>
<td>$v^N_{REF.TR}$</td>
<td>0.4</td>
<td></td>
<td></td>
<td>-2.3800</td>
<td></td>
</tr>
<tr>
<td>$k_{NH4 TR}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>1.5000</td>
<td></td>
</tr>
<tr>
<td>$k_{NO3 TR}$</td>
<td>0.5</td>
<td></td>
<td></td>
<td>0.0570</td>
<td></td>
</tr>
<tr>
<td>$v^P_{REF.TR}$</td>
<td>0.02</td>
<td></td>
<td></td>
<td>-2.0600</td>
<td></td>
</tr>
<tr>
<td>$k_{PO4 TR}$</td>
<td>0.047</td>
<td></td>
<td></td>
<td>1.8000</td>
<td></td>
</tr>
<tr>
<td>$pick_{PO4 TR}$</td>
<td>1.0</td>
<td></td>
<td></td>
<td>1.7000</td>
<td></td>
</tr>
<tr>
<td>$\xi_{NFix}$</td>
<td>2.5</td>
<td></td>
<td></td>
<td>-0.5220</td>
<td></td>
</tr>
<tr>
<td>$e^{PSV.TR}$</td>
<td>0.02</td>
<td></td>
<td></td>
<td>0.6060</td>
<td></td>
</tr>
<tr>
<td>$e^{NFix.TR}$</td>
<td>0.36</td>
<td></td>
<td></td>
<td>0.3300</td>
<td></td>
</tr>
<tr>
<td>$e^{CHO.TR}$</td>
<td>0.05</td>
<td></td>
<td></td>
<td>0.2180</td>
<td></td>
</tr>
<tr>
<td>$pom_{TR}$</td>
<td>0.1</td>
<td>0.0800</td>
<td></td>
<td>0.0601</td>
<td>OP</td>
</tr>
<tr>
<td>$\mu_{UN}$</td>
<td>1.6</td>
<td></td>
<td></td>
<td>-5.8600</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{CHL UN}$</td>
<td>0.5</td>
<td></td>
<td></td>
<td>-0.6120</td>
<td></td>
</tr>
<tr>
<td>$k_{DOM}$</td>
<td>0.5</td>
<td></td>
<td></td>
<td>-0.1500</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------</td>
</tr>
<tr>
<td>$r_{SDOM}$</td>
<td>0.005</td>
<td>0.00096</td>
<td>0.00080 - 0.00115</td>
<td>0.0951</td>
<td>CS</td>
</tr>
<tr>
<td>$\mu_{BA}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>1.2100</td>
<td></td>
</tr>
<tr>
<td>$b_{RESP}$</td>
<td>0.28</td>
<td></td>
<td></td>
<td>0.7100</td>
<td></td>
</tr>
<tr>
<td>$ex_{ADJ, BA}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>$remi_{BA}$</td>
<td>6.0</td>
<td></td>
<td></td>
<td>-0.4080</td>
<td></td>
</tr>
<tr>
<td>$ex_{REFR, BA}$</td>
<td>0.02</td>
<td>0.0180</td>
<td></td>
<td>0.8550</td>
<td>OP</td>
</tr>
<tr>
<td>$f_{SLCT, BA}$</td>
<td>0.2</td>
<td></td>
<td></td>
<td>0.0870</td>
<td></td>
</tr>
<tr>
<td>$resp_{BA}^b$</td>
<td>0.01</td>
<td></td>
<td></td>
<td>0.3820</td>
<td></td>
</tr>
<tr>
<td>$r_{min, BA}^A$</td>
<td>0.35</td>
<td></td>
<td></td>
<td>-0.8080</td>
<td></td>
</tr>
<tr>
<td>$r_{max, BA}^A$</td>
<td>0.80</td>
<td>0.6735</td>
<td>0.646 - 0.702</td>
<td>-0.1590</td>
<td>CS</td>
</tr>
<tr>
<td>$mort_{BA}$</td>
<td>0.02</td>
<td>0.0490</td>
<td></td>
<td>2.7700</td>
<td>OP</td>
</tr>
<tr>
<td>$\mu_{PRT}$</td>
<td>4.0</td>
<td></td>
<td></td>
<td>-0.8480</td>
<td></td>
</tr>
<tr>
<td>$g_{PHY}$</td>
<td>0.67</td>
<td>0.7513</td>
<td>0.639 - 0.883</td>
<td>-0.1240</td>
<td>CS</td>
</tr>
<tr>
<td>$g_{BA}$</td>
<td>2.0</td>
<td>1.4300</td>
<td>1.184 - 1.727</td>
<td>0.0868</td>
<td>CS</td>
</tr>
<tr>
<td>$ex_{PRT}$</td>
<td>0.20</td>
<td></td>
<td></td>
<td>1.8800</td>
<td></td>
</tr>
<tr>
<td>$f_{ex, PRT}$</td>
<td>0.75</td>
<td>0.9000</td>
<td></td>
<td>7.4100</td>
<td>OP</td>
</tr>
<tr>
<td>$resp_{PRT}^b$</td>
<td>0.01</td>
<td></td>
<td></td>
<td>1.7600</td>
<td></td>
</tr>
<tr>
<td>$resp_{PRT}^A$</td>
<td>0.67</td>
<td>0.7585</td>
<td>0.755 - 0.762</td>
<td>-0.0130</td>
<td>CS</td>
</tr>
<tr>
<td>$ex_{ADJ, PRT}$</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.2420</td>
<td></td>
</tr>
<tr>
<td>$remi_{PRT}$</td>
<td>4.7</td>
<td></td>
<td></td>
<td>-0.7540</td>
<td></td>
</tr>
<tr>
<td>$pom_{PRT}$</td>
<td>0.027</td>
<td></td>
<td></td>
<td>0.5200</td>
<td></td>
</tr>
<tr>
<td>$\mu_{MZ}$</td>
<td>1.0</td>
<td>1.3000</td>
<td></td>
<td>-3.4000</td>
<td>OP</td>
</tr>
<tr>
<td>$g_{PRT}$</td>
<td>0.65</td>
<td>1.4000</td>
<td></td>
<td>5.5000</td>
<td>OP</td>
</tr>
<tr>
<td>$g_{TR}$</td>
<td>1.0</td>
<td>0.2007</td>
<td>0.179 - 0.228</td>
<td>-0.0622</td>
<td>CS</td>
</tr>
<tr>
<td>$ex_{MZ}$</td>
<td>0.30</td>
<td></td>
<td></td>
<td>5.2800</td>
<td></td>
</tr>
<tr>
<td>$f_{ex, MZ}$</td>
<td>0.75</td>
<td></td>
<td></td>
<td>1.8000</td>
<td></td>
</tr>
<tr>
<td>$resp_{MZ}^b$</td>
<td>0.03</td>
<td></td>
<td></td>
<td>1.2100</td>
<td></td>
</tr>
<tr>
<td>$resp_{MZ}^A$</td>
<td>0.25</td>
<td>0.2200</td>
<td></td>
<td>3.5600</td>
<td>OP</td>
</tr>
<tr>
<td>$ex_{ADJ, MZ}$</td>
<td>2.0</td>
<td>0.2200</td>
<td></td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>$remi_{MZ}$</td>
<td>4.0</td>
<td></td>
<td></td>
<td>-0.0152</td>
<td></td>
</tr>
<tr>
<td>$pom_{MZ}$</td>
<td>0.15</td>
<td></td>
<td></td>
<td>2.2300</td>
<td></td>
</tr>
<tr>
<td>$ex_{REFR, MZ}$</td>
<td>0.02</td>
<td></td>
<td></td>
<td>0.2480</td>
<td></td>
</tr>
<tr>
<td>$remv_{MZ}$</td>
<td>0.8</td>
<td>0.8100</td>
<td></td>
<td>3.6600</td>
<td>OP</td>
</tr>
<tr>
<td>$f_{SDOM, MZ}$</td>
<td>0.1</td>
<td></td>
<td></td>
<td>-0.2050</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>$f_{POM,HZ}$</td>
<td>0.2</td>
<td>0.1400</td>
<td>-0.0607</td>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$ex_{REFR,SDOM}$</td>
<td>0.0009</td>
<td></td>
<td>-0.5110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q^c_{N,REFR}$</td>
<td>0.05</td>
<td></td>
<td>-2.2800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q^c_{P,REFR}$</td>
<td>0.0007</td>
<td></td>
<td>-0.5050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q^c_{N,POM}$</td>
<td>0.12</td>
<td></td>
<td>1.6200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q^c_{P,POM}$</td>
<td>0.0045</td>
<td></td>
<td>-0.8040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{nof}$</td>
<td>0.1</td>
<td></td>
<td>0.9820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$prf_N$</td>
<td>1.1</td>
<td></td>
<td>-0.0457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$prf_P$</td>
<td>4.0</td>
<td></td>
<td>-0.3790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$wnsv$</td>
<td>2.0</td>
<td>5.2625</td>
<td>4.169 – 6.643</td>
<td>-0.0137</td>
<td>CS</td>
</tr>
<tr>
<td>$diss$</td>
<td>0.02</td>
<td>0.0110</td>
<td></td>
<td>-1.0200</td>
<td>OP</td>
</tr>
</tbody>
</table>
APPENDIX A: Ecosystem Model Equations

Notes:
(a) All the parameter names (Table S1) begin with lower-case letters, and state and local variables begin with upper-case letters.
(b) State variables are $C$, $N$ and $P$ representing carbon, nitrogen and phosphorus components respectively of a biological species or a chemical substrate as noted in the subscript.
(c) $Q_X^C$ is used to represent a ratio $X / Y$.

1. Temperature effect

$$T_{fun} = EXP \left(-A_E \left(\frac{1}{T + 273.15} - \frac{1}{25 + 273.15}\right)\right),$$

where $T$ is water temperature in Celsius.

2. Phytoplankton (PHY) Processes

(A2.1) Cellular quota limitation on primary production

Cellular quota:

$$Q_{N,PHY}^C = \frac{N_{PHY}}{C_{PHY}}, \quad Q_{P,PHY}^C = \frac{P_{PHY}}{C_{PHY}}$$

Limitation:

$$N_{fun,PHY} = \frac{Q_{N,PHY}^C - q_{N, min,PHY}^C}{q_{N, rdf,PHY}^C - q_{N, min,PHY}^C}, \text{ then set } 0 \leq N_{fun,PHY} \leq 1$$

$$P_{fun,PHY} = \frac{Q_{P,PHY}^C - q_{P, min,PHY}^C}{q_{P, rdf,PHY}^C - q_{P, min,PHY}^C}, \text{ then set } 0 \leq P_{fun,PHY} \leq 1$$

Maximum primary production rate after cell quota limitation

$$p_{max,PHY}^C = \mu_{PHY} \cdot T_{fun} \cdot \min \left(N_{fun,PHY}, P_{fun,PHY}\right)$$

(A2.2) Light limitation and carbon specific gross primary production

$$GROW_{PHY}^C = 0, \text{ if } p_{max,PHY}^C = 0. \text{ Otherwise:}$$

$$GROW_{PHY}^C = C_{PHY} \cdot p_{max,PHY}^C \cdot \left(1 - EXP \left(-\frac{\alpha_{PHY} \cdot Q_{CHL,PHY} \cdot PAR}{p_{max,PHY}^C}\right)\right) \cdot EXP \left(-\beta_{PHY} \cdot PAR\right)$$

(A2.3) Nitrogen assimilation, including nitrate uptake and ammonium uptake

Cellular quota limitation

$$V_{max,PHY}^N = \frac{q_{N, max,PHY}^C - Q_{N,PHY}^C}{q_{N, rdf,PHY}^C - q_{N, max,PHY}^C}, \text{ then set } 0 \leq V_{max,PHY}^N \leq 1$$

Nitrogen assimilation

$$GROW_{PHY}^{NH4} = C_{PHY} \cdot V_{REF,PHY}^N \cdot T_{fun} \cdot V_{max,PHY}^N \cdot \frac{NH4}{NH4 + k_{NH4,PHY}^{NH4} + NO3 \cdot k_{NH4,PHY}^{NO3}}$$
\[ GROW_{\text{PHY}}^{\text{NO}_3} = C_{\text{PHY}} \cdot V_{\text{REF,PHY}}^{N} \cdot T\text{fun} \cdot V_{\text{max,PHY}}^{N} \cdot \frac{\text{NO}_3}{\text{NO}_3 + k_{\text{PHY}}^{\text{NO}_3} + \text{NH}_4 \cdot k_{\text{PHY}}^{\text{NO}_3} + k_{\text{PHY}}^{\text{NH}_4}} \]

\[ GROW_{\text{PHY}}^{N} = GROW_{\text{PHY}}^{\text{NH}_4} + GROW_{\text{PHY}}^{\text{NO}_3} \]

(A2.4) Phosphorous (phosphate) assimilation

\[ V_{\text{max,PHY}}^{P} = \frac{q_{\text{tmp,PHY}}^{C} - Q_{\text{PHY}}^{C}}{q_{\text{tmp,PHY}}^{C} - d_{\text{tmp,PHY}}^{C}}, \text{ then set } 0 \leq V_{\text{max,PHY}}^{P} \leq 1 \]

\[ GROW_{\text{PHY}}^{P} = C_{\text{PHY}} \cdot v_{\text{REF,PHY}}^{P} \cdot T\text{fun} \cdot V_{\text{max,PHY}}^{P} \cdot \frac{\text{PO}_4}{\text{PO}_4 + k_{\text{PHY}}^{\text{PO}_4}} \]

(A2.5) Respiration required by nitrate reduction

\[ \text{RESP}_{\text{PHY}} = \frac{\text{PO}_4^{\text{NO}_3} \cdot GROW_{\text{PHY}}^{\text{NO}_3}}{\text{PO}_4} \]

(A2.6) Chlorophyll production

\[ GROW_{\text{PHY}}^{\text{CHL}} = 0, \text{ if PAR} \leq 0. \text{ Otherwise:} \]

\[ GROW_{\text{PHY}}^{\text{CHL}} = \theta \cdot GROW_{\text{PHY}}^{N} \cdot \frac{GROW_{\text{PHY}}^{C}}{\alpha_{\text{PHY}}^{\text{CHL}} \cdot \text{CHL}_{\text{PHY}} \cdot \text{PAR} \cdot \text{EXP}(-\beta_{\text{PHY}} \cdot \text{PAR})} \]

(A2.7) DOM Excretion

(a) Passive excretion

\[ EXCR_{\text{PHY}}^{C} = e_{\text{PSV,PHY}}^{C} \cdot C_{\text{PHY}} \]
\[ EXCR_{\text{PHY}}^{N} = e_{\text{PSV,PHY}}^{N} \cdot N_{\text{PHY}} \]
\[ EXCR_{\text{PHY}}^{P} = e_{\text{PSV,PHY}}^{P} \cdot P_{\text{PHY}} \]

(b) Active carbohydrate excretion

\[ EXCR_{\text{CHO,PHY}}^{C} = e_{\text{CHO,PHY}}^{C} \cdot GROW_{\text{PHY}}^{C} \]

(c) Active DOM release in order to adjust cellular stoichiometry

\[ EXCR_{\text{ACT,PHY}}^{C} = 0.5 \cdot C_{\text{PHY}} \cdot \max \left( 0, 1 - \frac{Q_{\text{PHY}}^{C}}{q_{\text{N,PHY}}^{C}} , 1 - \frac{Q_{\text{PHY}}^{C}}{q_{\text{P,PHY}}^{C}} \right) \]

if \( EXCR_{\text{ACT,PHY}}^{C} > 0 \)

\[ EXCR_{\text{ACT,PHY}}^{N} = 0.5 \cdot 0.25 \cdot N_{\text{PHY}} \cdot \max \left( 0, 1 - \frac{Q_{\text{PHY}}^{N}}{q_{\text{N,PHY}}^{C}} \right) \]
\[ EXCR_{\text{ACT,PHY}}^{P} = 0.5 \cdot 0.25 \cdot P_{\text{PHY}} \cdot \max \left( 0, 1 - \frac{Q_{\text{PHY}}^{P}}{q_{\text{N,PHY}}^{C}} \right) \]

else

\[ EXCR_{\text{ACT,PHY}}^{N} = 0, \ EXCR_{\text{ACT,PHY}}^{P} = 0 \]
end

(d) Partitioning between labile and semilabile DOM

\[
EXCR^{LDOC}_{PHY} = EXCR^{C}_{PSV\_PHY} + 0.75 \cdot EXCR^{C}_{CHO\_PHY}
\]

\[
EXCR^{LDON}_{PHY} = EXCR^{N}_{PSV\_PHY}
\]

\[
EXCR^{LDOp}_{PHY} = EXCR^{P}_{PSV\_PHY}
\]

\[
EXCR^{SDOC}_{PHY} = EXCR^{C}_{ACT\_PHY} + 0.25 \cdot EXCR^{C}_{CHO\_PHY}
\]

\[
EXCR^{SDON}_{PHY} = EXCR^{N}_{ACT\_PHY}
\]

\[
EXCR^{SDOP}_{PHY} = EXCR^{P}_{ACT\_PHY}
\]

(A2.8) POM production by aggregation

\[
POM^{C}_{PHY} = pom_{PHY} \cdot C^{2}_{PHY}
\]

\[
POM^{N}_{PHY} = Q^{C}_{N\_PHY} \cdot POM^{C}_{PHY}
\]

\[
POM^{P}_{PHY} = Q^{C}_{P\_PHY} \cdot POM^{C}_{PHY}
\]

\[
POM^{CHL}_{PHY} = Q^{C}_{CHL\_PHY} \cdot POM^{C}_{PHY}
\]

(A2.9) Grazing on phytoplankton by protozoa

\[
GRAZ^{C}_{PHY} = T_{fun} \cdot \mu_{PRT} \cdot C_{PRT} \cdot \frac{C^{2}_{PHY}}{C^{2}_{PHY} + g^{2}_{PHY} + \left(C_{UN} \cdot g_{PHY} / g_{UN}\right)^{2} + \left(C_{BA} \cdot g_{PHY} / g_{BA}\right)^{2}}
\]

\[
GRAZ^{N}_{PHY} = Q^{C}_{N\_PHY} \cdot GRAZ^{C}_{PHY}
\]

\[
GRAZ^{P}_{PHY} = Q^{C}_{P\_PHY} \cdot GRAZ^{C}_{PHY}
\]

\[
GRAZ^{CHL}_{PHY} = Q^{C}_{CHL\_PHY} \cdot GRAZ^{C}_{PHY}
\]

(A2.10) PHY net growth rate

\[
\frac{dC_{PHY}}{dt} = GROW^{C}_{PHY} - RESP_{PHY} - EXCR^{LDOC}_{PHY} - EXCR^{SDOC}_{PHY} - POM^{C}_{PHY} - GRAZ^{C}_{PHY}
\]

\[
\frac{dN_{PHY}}{dt} = GROW^{N}_{PHY} - EXCR^{LDON}_{PHY} - EXCR^{SDON}_{PHY} - POM^{N}_{PHY} - GRAZ^{N}_{PHY}
\]

\[
\frac{dP_{PHY}}{dt} = GROW^{P}_{PHY} - EXCR^{LDOp}_{PHY} - EXCR^{SDOP}_{PHY} - POM^{P}_{PHY} - GRAZ^{P}_{PHY}
\]

\[
\frac{dCHL_{PHY}}{dt} = GROW^{CHL}_{PHY} - POM^{CHL}_{PHY} - GRAZ^{CHL}_{PHY}
\]


*Notes: Trichodesmium* is disabled for the simulations at AS and EQP
(A3.1) Cellular quota limitation on primary production

Limitation

\[ N_{fun}^{TR} = \frac{Q_C^{C,TR} - q_{C,\text{min},TR}^C}{q_{C,\text{ rdf},TR}^C - q_{C,\text{min},TR}^C}, \text{ then set } 0 \leq N_{fun}^{TR} \leq 1 \]

\[ P_{fun}^{TR} = \frac{Q_C^{P,TR} - q_{P,\text{min},TR}^C}{q_{P,\text{ rdf},TR}^C - q_{P,\text{min},TR}^C}, \text{ then set } 0 \leq P_{fun}^{TR} \leq 1 \]

Maximum primary production rate after cell quota limitation

\[ P_{\text{max},TR}^{C} = \mu_{TR} \cdot T_{fun} \cdot \min\left(N_{fun}^{TR}, P_{fun}^{TR}\right) \]

(A3.2) Light limitation and carbon specific gross primary production

\[ \text{GROW}^{C}_{TR} = 0, \text{ if } P_{\text{max},TR}^{C} = 0. \text{ Otherwise:} \]

\[ \text{GROW}^{C}_{TR} = C_{TR} \cdot P_{\text{max},TR}^{C} \left(1 - \exp\left(-\alpha_{TR}^{C} \cdot \frac{Q_{C,\text{CHL},TR}}{P_{\text{max},TR}^{C}} \cdot PAR\right)\right) \cdot \exp\left(-\beta_{TR} \cdot PAR\right) \]

(A3.3) Nitrogen assimilation, including nitrogen fixation, nitrate uptake and ammonium uptake

(a) Cellular quota limitation

\[ V_{\text{max},TR}^{N} = \frac{Q_{N,\text{max},TR}^{C} - Q_{N}^{C,TR}}{Q_{N,\text{ rdf},TR}^{C} - Q_{N,\text{min},TR}^{C}}, \text{ then set } 0 \leq V_{\text{max},TR}^{N} \leq 1 \]

(b) Nitrate and ammonium uptake

\[ \text{GROW}^{NH4}_{TR} = C_{TR} \cdot V_{REF,TR}^{N} \cdot T_{fun} \cdot V_{\text{max},TR}^{N} \cdot \frac{NH4}{NH4 \cdot k_{NH4}^T + NO3 \cdot k_{NO3}^T / k_{NO3}^T} \]

\[ \text{GROW}^{NO3}_{TR} = C_{TR} \cdot V_{REF,TR}^{N} \cdot T_{fun} \cdot V_{\text{max},TR}^{N} \cdot \frac{NO3}{NO3 \cdot k_{NO3}^T + NH4 \cdot k_{NO3}^T / k_{NO3}^T} \]

(c) Maximum nitrogen fixation

\[ N_{FIX}^{\text{max},TR} = 0, \text{ if } T \leq 20^\circ C. \text{ Otherwise:} \]

\[ \text{NFIX}^{\text{max},TR} = \frac{\left(GROW^{C}_{TR} - \xi^{NO3} \cdot GROW^{NO3}_{TR}\right) \cdot q_{N,\text{max},TR}^{C} - GROW^{NO3}_{TR} - GROW^{NH4}_{TR}}{1 + \xi^{NO3} \cdot q_{N,\text{max},TR}^{C} \cdot T_{fun} \cdot V_{\text{max},TR}^{N}} \]

then set \( N_{FIX}^{\text{max},TR} \geq 0 \)

(d) Total nitrogen assimilation

\[ \text{GROW}^{N}_{TR} = \min\left(C_{TR} \cdot V_{REF,TR}^{N} \cdot T_{fun} \cdot V_{\text{max},TR}^{N}, \text{ GROW}^{NH4}_{TR} + \text{GROW}^{NO3}_{TR} + \text{NFI}^{\text{max},TR}\right) \]

(e) Actual nitrogen fixation

\[ \text{GROW}^{N}_{TR} = \text{GROW}^{N}_{TR} - \text{GROW}^{NH4}_{TR} - \text{GROW}^{NO3}_{TR} \]

(A3.4) Phosphorous assimilation

(a) Phosphate assimilation
\[ V_{\text{max,TR}}^P = \frac{q_{\text{max,TR}}^C - Q_{\text{TR}}^C}{q_{\text{max,TR}}^C - q_{\text{P, rdf, TR}}^C}, \text{then set } 0 \leq V_{\text{max,TR}}^P \leq 1 \]

\[ GROW_{\text{TR}}^{\text{PO4}} = C_{\text{TR}} \cdot V_{\text{REF,TR}}^P \cdot T_{\text{fun}} \cdot V_{\text{max,TR}}^P \cdot \frac{PO4}{PO4 + k_{\text{TR}}^{\text{PO4}}} \]

(b) Active phosphate taking-up from deep water

\[ PICK_{\text{TR}}^{\text{PO4}} = \max \left( 0, \frac{pick_{\text{TR}}^{\text{PO4}} \cdot \left( (q_{\text{max,TR}}^C + q_{\text{P, rdf, TR}}^C) / 2 \cdot C_{\text{TR}} - P_{\text{TR}} \right)}{2} \right) \]

(c) Total phosphorus growth

\[ GROW_{\text{TR}}^P = GROW_{\text{TR}}^{\text{PO4}} + PICK_{\text{TR}}^{\text{PO4}} \]

(A3.5) Respiration required by nitrogen fixation and nitrate reduction

\[ RESP_{\text{TR}} = \xi^{\text{NO3}} \cdot GROW_{\text{TR}}^{\text{NO3}} + \xi^{\text{NFX}} \cdot GROW_{\text{TR}}^{\text{NFX}} \]

(A3.6) Chlorophyll production

\[ GROW_{\text{TR}}^{\text{CHL}} = 0, \text{if } PAR \leq 0. \text{ Otherwise:} \]

\[ GROW_{\text{TR}}^{\text{CHL}} = \theta \cdot GROW_{\text{TR}}^{\text{N}} \cdot \frac{GROW_{\text{TR}}^C}{\alpha_{\text{TR}}^{\text{CHL}} \cdot CHL_{\text{TR}} \cdot PAR \cdot EXP \left( -\beta_{\text{TR}} \cdot PAR \right)} \]

(A3.7) DOM and ammonium excretion

(a) Passive excretion

\[ EXCR_{\text{PSV,TR}}^C = ex_{\text{PSV,TR}} \cdot C_{\text{TR}} \]

\[ EXCR_{\text{PSV,TR}}^N = ex_{\text{PSV,TR}} \cdot N_{\text{TR}} \]

\[ EXCR_{\text{PSV,TR}}^P = ex_{\text{PSV,TR}} \cdot P_{\text{TR}} \]

(b) Active carbohydrate excretion

\[ EXCR_{\text{CHO,TR}}^C = ex_{\text{CHO,TR}} \cdot GROW_{\text{TR}}^C \]

(c) Active DOM release in order to adjust cellular stoichiometry

\[ EXCR_{\text{ACT,TR}}^C = 0.5 \cdot C_{\text{TR}} \cdot \max \left( 0, 1 - \frac{Q_{\text{N,TR}}^{C}}{q_{\text{N, rdf, TR}}^C}, 1 - \frac{Q_{\text{P,TR}}^{C}}{q_{\text{P, rdf, TR}}^C} \right) \]

if \( EXCR_{\text{ACT,TR}}^C > 0 \)

\[ EXCR_{\text{ACT,TR}}^N = 0.5 \cdot 0.25 \cdot N_{\text{TR}} \cdot \max \left( 0, 1 - \frac{Q_{\text{P,TR}}^{N}}{q_{\text{P, rdf, TR}}^N / q_{\text{N, rdf, TR}}^N} \right) \]

\[ EXCR_{\text{ACT,TR}}^P = 0.5 \cdot 0.25 \cdot P_{\text{TR}} \cdot \max \left( 0, 1 - \frac{Q_{\text{N,TR}}^{P}}{q_{\text{N, rdf, TR}}^P / q_{\text{P, rdf, TR}}^P} \right) \]
else
\( \text{EXCR}^{\text{N,ACT,TR}} = 0, \text{EXCR}^{\text{P,ACT,TR}} = 0 \)
end

(d) Active DON release of newly fixed N
\( \text{EXCR}^{\text{DON,FIX,TR}} = 0.5 \cdot \text{ex}_{\text{FIX,TR}} \cdot \text{GROW}_{\text{TR}}^{\text{FIX}} \cdot \text{Nfun}_{\text{TR}} \)

(e) Partitioning between labile and semilabile DOM
\( \text{EXCR}^{\text{LD,TR}} = \text{EXCR}^{\text{PSV,TR}} + 0.75 \cdot \text{EXCR}^{\text{CHO,TR}} \)
\( \text{EXCR}^{\text{LD,TR}} = \text{EXCR}^{\text{N,ACT,TR}} + \text{EXCR}^{\text{DON,TR}} \)
\( \text{EXCR}^{\text{SD,TR}} = \text{EXCR}^{\text{PSV,TR}} + 0.25 \cdot \text{EXCR}^{\text{CHO,TR}} \)
\( \text{EXCR}^{\text{SD,TR}} = \text{EXCR}^{\text{P,ACT,TR}} \)

(f) Ammonium excretion
\( \text{EXCR}^{\text{NH,4}} = 0.5 \cdot \text{ex}_{\text{FIX,TR}} \cdot \text{GROW}_{\text{TR}}^{\text{FIX}} \cdot \text{Nfun}_{\text{TR}} \)

(A3.8) POM production by aggregation
\( \text{POM}^{\text{C,TR}} = \text{pom}_{\text{TR}} \cdot C_{\text{TR}}^2 \)
\( \text{POM}^{\text{N,TR}} = Q_{\text{N,TR}} \cdot \text{POM}^{\text{C,TR}} \)
\( \text{POM}^{\text{P,TR}} = Q_{\text{P,TR}} \cdot \text{POM}^{\text{C,TR}} \)
\( \text{POM}^{\text{CHL,TR}} = Q_{\text{CHL,TR}} \cdot \text{POM}^{\text{C,TR}} \)

(A3.9) Grazing on Trichodesmium by metazoaa
\( \text{GRAZ}^{\text{C,TR}} = T_{\text{fun}} \cdot \mu_{\text{MZ}} \cdot C_{\text{MZ}} \cdot \frac{C_{\text{TR}}^2}{C_{\text{TR}}^2 + g_{\text{TR}}^2 + \left(C_{\text{PRT}} \cdot \frac{g_{\text{TR}}}{g_{\text{PRT}}} \right)^2} \)
\( \text{GRAZ}^{\text{N,TR}} = Q_{\text{N,TR}} \cdot \text{GRAZ}^{\text{C,TR}} \)
\( \text{GRAZ}^{\text{P,TR}} = Q_{\text{P,TR}} \cdot \text{GRAZ}^{\text{C,TR}} \)
\( \text{GRAZ}^{\text{CHL,TR}} = Q_{\text{CHL,TR}} \cdot \text{GRAZ}^{\text{C,TR}} \)

(A3.10) Trichodesmium net growth rate
\[
\frac{dC^{\text{TR}}}{dt} = \text{GROW}^{\text{C,TR}} - \text{RESP}^{\text{TR}} - \text{EXCR}^{\text{LD,DOC,TR}} - \text{EXCR}^{\text{SD,DOC,TR}} - \text{POM}^{\text{C,TR}} - \text{GRAZ}^{\text{C,TR}}
\]
\[
\frac{dN_{\text{TR}}}{dt} = GROW_{\text{TR}}^N - EXCR_{\text{TR}}^{NH4} - EXCR_{\text{TR}}^{LDON} - EXCR_{\text{TR}}^{SDON} - POM_{\text{TR}}^N - GRAZ_{\text{TR}}^N
\]
\[
\frac{dP_{\text{TR}}}{dt} = GROW_{\text{TR}}^P - EXCR_{\text{TR}}^{LDOP} - EXCR_{\text{TR}}^{SDOP} - POM_{\text{TR}}^P - GRAZ_{\text{TR}}^P
\]
\[
\frac{dCHL_{\text{TR}}}{dt} = GROW_{\text{CHL}}^{\text{CHL}} - POM_{\text{TR}}^{\text{CHL}} - GRAZ_{\text{TR}}^{\text{CHL}}
\]

4. Unicellular N\textsubscript{2}-fixers (UN) Processes

Notes: Unicellular N\textsubscript{2}-fixers is disabled for the simulations at AS and EQP

(A4.1) Cellular quota limitation on primary production

Limitation

\[
N_{\text{fun}}^{\text{UN}} = \frac{Q_C^{\text{N,UN}} - q_{\text{N, min,UN}}^{\text{N, min,UN}}}{q_{\text{N, rdf,UN}}^{\text{N, rdf,UN}} - q_{\text{N, min,UN}}^{\text{N, min,UN}}}, \text{ then set } 0 \leq N_{\text{fun}}^{\text{UN}} \leq 1
\]
\[
P_{\text{fun}}^{\text{UN}} = \frac{Q_C^{\text{P,UN}} - q_{\text{P, min,UN}}^{\text{P, min,UN}}}{q_{\text{P, rdf,UN}}^{\text{P, rdf,UN}} - q_{\text{P, min,UN}}^{\text{P, min,UN}}}, \text{ then set } 0 \leq P_{\text{fun}}^{\text{UN}} \leq 1
\]

Maximum primary production rate after cell quota limitation

\[
P_{\text{C, max,UN}} = \mu_{\text{UN}} \cdot T_{\text{fun}} \cdot \min \left( N_{\text{fun}}^{\text{UN}} , P_{\text{fun}}^{\text{UN}} \right)
\]

(A4.2) Light limitation and carbon specific gross primary production

\[
GROW_{\text{UN}}^C = 0, \text{ if } P_{\text{max,UN}}^C = 0. \text{ Otherwise:}
\]
\[
GROW_{\text{UN}}^C = C_{\text{UN}} \cdot P_{\text{max,UN}}^C \cdot \left( 1 - \exp \left( -\frac{\alpha_{\text{CHL}}^{\text{C,UN}} \cdot Q_{\text{CHL,UN}}^C \cdot PAR}{P_{\text{max,UN}}} \right) \right) \cdot \exp \left( -\beta_{\text{UN}} \cdot PAR \right)
\]

(A4.3) Nitrogen assimilation, including nitrogen fixation, nitrate uptake and ammonium uptake

(a) Cellular quota limitation

\[
V_{\text{max,UN}}^N = \frac{q_{\text{N, max,UN}}^C - Q_{\text{N,UN}}^C}{q_{\text{N, max,UN}}^C - q_{\text{N, rdf,UN}}^C}, \text{ then set } 0 \leq V_{\text{max,UN}}^N \leq 1
\]

(b) Nitrate and ammonium uptake

\[
GROW_{\text{UN}}^{\text{NH4}} = C_{\text{UN}} \cdot V_{\text{REF,UN}}^N \cdot T_{\text{fun}} \cdot V_{\text{max,UN}}^N \cdot \frac{NH4}{NH4 + k_{\text{UN}}^{\text{NH4}} + NO3 \cdot k_{\text{UN}}^{\text{NH4}}} + \frac{NO3}{NO3 + k_{\text{UN}}^{\text{NO3}} + NH4 \cdot k_{\text{UN}}^{\text{NO3}}}
\]
\[
GROW_{\text{UN}}^{\text{NO3}} = C_{\text{UN}} \cdot V_{\text{REF,UN}}^N \cdot T_{\text{fun}} \cdot V_{\text{max,UN}}^N \cdot \frac{NO3}{NO3 + k_{\text{UN}}^{\text{NO3}} + NH4 \cdot k_{\text{UN}}^{\text{NO3}}}
\]

(c) Maximum nitrogen fixation

\[
NFIX_{\text{max,UN}} = 0, \text{ if } T \leq 15^\circ C. \text{ Otherwise:}
\]
\[
NFIX_{\text{max,UN}} = \frac{\left( GROW_{\text{UN}}^C - \xi^{\text{NO3}} \cdot GROW_{\text{UN}}^{\text{NO3}} \right) \cdot q_{\text{N, max,UN}}^C - GROW_{\text{UN}}^{\text{NO3}} - GROW_{\text{UN}}^{\text{NH4}}}{1 + \xi^{\text{NFIX}} \cdot q_{\text{N, max,UN}}^C} \cdot T_{\text{fun}} \cdot V_{\text{max,UN}}^N.
\]
then set $N_{\text{FIX}_{\text{max,UN}}} \geq 0$

(d) Total nitrogen assimilation

$$GROW_{\text{UN}}^{N_{\text{NH4}}} = \min \left( C_{\text{UN}} \cdot V_{\text{REF,UN}}^{N_{\text{NH4}}} \cdot T_{\text{fun}} \cdot V_{\max,\text{UN}}^{N_{\text{NH4}}} + GROW_{\text{UN}}^{N_{\text{NO3}}} + N_{\text{FIX}_{\text{max,UN}}} \right)$$

(e) Actual nitrogen fixation

$$GROW_{\text{UN}}^{N_{\text{FIX}}} = GROW_{\text{UN}}^{N_{\text{NH4}}} - GROW_{\text{UN}}^{N_{\text{NO3}}}$$

(A4.4) Phosphorous assimilation

Phosphate assimilation

$$V_{\text{max,UN}}^{P} = \frac{q_{\text{C,UN}}^{P} \cdot \max,\text{UN} - Q_{\text{P,UN}}^{P}}{q_{\text{C,UN}}^{P} \cdot \text{max,UN} - q_{\text{C,UN}}^{P} \cdot \text{rdf,UN}}, \text{ then set } 0 \leq V_{\text{max,UN}}^{P} \leq 1$$

$$GROW_{\text{UN}}^{P} = C_{\text{UN}} \cdot V_{\text{REF,UN}}^{P} \cdot T_{\text{fun}} \cdot V_{\text{max,UN}}^{P} \cdot \frac{PO4}{PO4 + k_{\text{UN}}^{P}}$$

(A4.5) Respiration required by nitrogen fixation and nitrate reduction

$$RES_{\text{UN}} = \xi_{N_{\text{NO3}}} \cdot GROW_{\text{UN}}^{N_{\text{NO3}}} + \xi_{N_{\text{FIX}}} \cdot GROW_{\text{UN}}^{N_{\text{FIX}}}$$

(A4.6) Chlorophyll production

$$GROW_{\text{UN}}^{\text{CHL}} = 0, \text{ if } PAR \leq 0. \text{ Otherwise:}$$

$$GROW_{\text{UN}}^{\text{CHL}} = \theta \cdot GROW_{\text{UN}}^{N_{\text{CHO}}} \cdot \frac{GROW_{\text{UN}}^{C}}{\alpha_{\text{CHL,UN}}^{C} \cdot CHL_{\text{UN}} \cdot PAR \cdot EXP \left(-\beta_{\text{UN}} \cdot PAR\right)}$$

(A4.7) DOM and ammonium excretion

(a) Passive excretion

$$EXCR_{\text{PSV,UN}}^{C} = ex_{\text{PSV,UN}} \cdot C_{\text{UN}}$$

$$EXCR_{\text{PSV,UN}}^{N} = ex_{\text{PSV,UN}} \cdot N_{\text{UN}}$$

$$EXCR_{\text{PSV,UN}}^{P} = ex_{\text{PSV,UN}} \cdot P_{\text{UN}}$$

(b) Active carbohydrate excretion

$$EXCR_{\text{CHO,UN}}^{C} = ex_{\text{CHO,UN}} \cdot GROW_{\text{UN}}^{C}$$

(c) Active DOM release in order to adjust cellular stoichiometry

$$EXCR_{\text{ACT,UN}}^{C} = 0.5 \cdot C_{\text{UN}} \cdot \max \left( 0, 1 - \frac{Q_{N,\text{UN}}^{C}}{q_{N,\text{rdf,UN}}^{C}}, 1 - \frac{Q_{P,\text{UN}}^{C}}{q_{P,\text{rdf,UN}}^{C}} \right)$$

if $EXCR_{\text{ACT,UN}}^{N} > 0$

$$EXCR_{\text{ACT,UN}}^{N} = 0.5 \cdot 0.25 \cdot N_{\text{UN}} \cdot \max \left( 0, 1 - \frac{Q_{P,\text{UN}}^{N}}{q_{P,\text{rdf,UN}}^{N}}, 1 - \frac{Q_{C,\text{UN}}^{N}}{q_{C,\text{UN}}^{N}} \right)$$
\[ \text{EXCR}_{\text{ACT,UN}}^p = 0.5 \cdot 0.25 \cdot P_{\text{UN}} \cdot \max \left( 0, 1 - \frac{Q_{N,\text{UN}}^p}{q_{N,\text{def,UN}}^C / q_{P,\text{def,UN}}^C} \right) \]

e else
\[ \text{EXCR}_{\text{ACT,UN}}^N = 0 \cdot \text{EXCR}_{\text{ACT,UN}}^p = 0 \]
end

(d) Active DON release of newly fixed N
\[ \text{EXCR}_{\text{NFIX,UN}}^{\text{DON}} = 0.5 \cdot e_{\text{NFIX,UN}}^N \cdot \text{GROW}_{\text{UN}}^{\text{NFIX}} \cdot N_{\text{fun}}^N \]

(e) Partitioning between labile and semilabile DOM
\[ \text{EXCR}_{\text{UN}}^{\text{LDOC}} = \text{EXCR}_{\text{PSV,UN}}^C + 0.75 \cdot \text{EXCR}_{\text{CHO,UN}}^C \]
\[ \text{EXCR}_{\text{UN}}^{\text{LDON}} = \text{EXCR}_{\text{PSV,UN}}^N + \text{EXCR}_{\text{NFIX,UN}}^{\text{DON}} \]
\[ \text{EXCR}_{\text{UN}}^{\text{LDOP}} = \text{EXCR}_{\text{PSV,UN}}^p \]
\[ \text{EXCR}_{\text{UN}}^{\text{SDOC}} = \text{EXCR}_{\text{ACT,UN}}^C + 0.25 \cdot \text{EXCR}_{\text{CHO,UN}}^C \]
\[ \text{EXCR}_{\text{UN}}^{\text{SDON}} = \text{EXCR}_{\text{ACT,UN}}^N \]
\[ \text{EXCR}_{\text{UN}}^{\text{SDOP}} = \text{EXCR}_{\text{ACT,UN}}^p \]

(f) Ammonium excretion
\[ \text{EXCR}_{\text{UN}}^{\text{NH}_4} = 0.5 \cdot e_{\text{NFIX,UN}}^N \cdot \text{GROW}_{\text{UN}}^{\text{NFIX}} \cdot N_{\text{fun}}^N \]

(A4.8) POM production by aggregation
\[ POM_{\text{UN}}^C = \rho_{\text{UN}} \cdot C_{\text{UN}}^2 \]
\[ POM_{\text{UN}}^N = Q_{\text{N,UN}}^C \cdot POM_{\text{UN}}^C \]
\[ POM_{\text{UN}}^p = Q_{P,\text{UN}}^C \cdot POM_{\text{UN}}^C \]
\[ POM_{\text{UN}}^{\text{CHL}} = Q_{\text{CHL,UN}}^C \cdot POM_{\text{UN}}^C \]

(A4.9) Grazing on Unicellular N\textsubscript{2}-fixers by protozoa
\[ \text{GRAZ}_{\text{UN}}^C = T_{\text{fun}} \cdot \mu_{\text{PRT}} \cdot C_{\text{PRT}} \cdot \frac{C_{\text{UN}}^2}{C_{\text{UN}}^2 + g_{\text{UN}}^2 + \left( C_{\text{PHY}} \cdot g_{\text{UN}} / g_{\text{PHY}} \right)^2 + \left( C_{\text{BA}} \cdot g_{\text{UN}} / g_{\text{BA}} \right)^2} \]
\[ \text{GRAZ}_{\text{UN}}^N = Q_{\text{N,UN}}^C \cdot \text{GRAZ}_{\text{UN}}^C \]
\[ \text{GRAZ}_{\text{UN}}^p = Q_{P,\text{UN}}^C \cdot \text{GRAZ}_{\text{UN}}^C \]
\[ \text{GRAZ}_{\text{UN}}^{\text{CHL}} = Q_{\text{CHL,UN}}^C \cdot \text{GRAZ}_{\text{UN}}^C \]

(A4.10) Unicellular N\textsubscript{2}-fixers net growth rate
\[
\frac{dC_{UN}}{dt} = GROW_{UN}^C - RESP_{UN} - EXCR_{UN}^{LDOC} - EXCR_{UN}^{SDOC} - POM_{UN}^C - GRAZ_{UN}^C \\
\frac{dN_{UN}}{dt} = GROW_{UN}^N - EXCR_{UN}^{NH4} - EXCR_{UN}^{LDON} - EXCR_{UN}^{SDON} - POM_{UN}^N - GRAZ_{UN}^N \\
\frac{dP_{UN}}{dt} = GROW_{UN}^P - EXCR_{UN}^{LDOP} - EXCR_{UN}^{SDOP} - POM_{UN}^P - GRAZ_{UN}^P \\
\frac{dCHL_{UN}}{dt} = GROW_{UN}^{CHL} - POM_{UN}^{CHL} - GRAZ_{UN}^{CHL}
\]

5. Heterotrophic Bacterial (BA) processes

(A5.1) Maximum available labile DOC

\[ALC = C_{LDOM}\]

(A5.2) Maximum available semilabile DOC

\[ASC = r_{SDOM} \cdot C_{SDOM}\]

(A5.3) Cellular quota limitation

\[N_{fun_{BA}} = Q_{N,BA}^C / q_{N,BA}^C, \text{ then set } N_{fun_{BA}} \leq 1\]  
\[P_{fun_{BA}} = Q_{P,BA}^C / q_{P,BA}^C, \text{ then set } P_{fun_{BA}} \leq 1\]

(A5.4) Bacterial gross carbon growth

Labile DOC

\[GROW_{BA}^{LDOC} = \mu_{BA} \cdot T_{fun} \cdot C_{BA} \cdot \min \left( N_{fun_{BA}}, P_{fun_{BA}} \right) \cdot \frac{ALC}{ALC + k_{DOM} + ASC}\]

Semilabile DOC

\[GROW_{BA}^{SDOC} = \mu_{BA} \cdot T_{fun} \cdot C_{BA} \cdot \min \left( N_{fun_{BA}}, P_{fun_{BA}} \right) \cdot \frac{ASC}{ASC + k_{DOM} + ALC}\]

Total gross carbon growth

\[GROW_{BA}^C = GROW_{BA}^{LDOC} + GROW_{BA}^{SDOC}\]

(A5.5) Bacterial nitrogen uptake

\[GROW_{BA}^{LDON} = GROW_{BA}^{LDOC} \cdot Q_{N,LDOM}^C\]

\[GROW_{BA}^{SDON} = GROW_{BA}^{SDOC} \cdot \min \left( Q_{N,BA}^C, Q_{N,SDOM}^C + f_{S_LCT,BA} / N_{fun_{BA}}, \left( Q_{N,BA}^C - Q_{N,SDOM}^C \right) \right)\]

\[GROW_{BA}^{NH4} = GROW_{BA}^{LDON} \cdot NH4 / N_{LDOM} / \min (1, N_{fun_{BA}})\]

If \(N_{fun_{BA}} < 1\) Then
\[ GROW_{BA}^{NO3} = \min \left( 0.1 \cdot NO3 \cdot \min\left(1, \frac{1}{N_{LDOM} + N_{SDOM}} \cdot \frac{GROW_{BA}^{LDON} + GROW_{BA}^{SDON}}{GROW_{BA}^{LDON} + GROW_{BA}^{SDON}} \right), (NO3 + NH4) \cdot \frac{GROW_{BA}^{LDON} + GROW_{BA}^{SDON}}{N_{LDOM} + N_{SDOM}} - GROW_{BA}^{NH4} \right) \]

given that \( GROW_{BA}^{NO3} \geq 0 \)

Else

\[ GROW_{BA}^{NO3} = 0 \]

End

\[ GROW_{BA} = GROW_{BA}^{LDON} + GROW_{BA}^{SDON} + GROW_{BA}^{NH4} + GROW_{BA}^{NO3} \]

(A5.6) Bacterial phosphorus uptake

\[ GROW_{BA}^{LDOP} = GROW_{BA}^{LDON} \cdot Q_{P_{DOM}}^{C} \]

\[ GROW_{BA}^{SDOP} = GROW_{BA}^{SDCON} \cdot \min\left(q_{C_{DOM}}^{B}, q_{P_{DOM}}^{C} + f_{SCLT,BA} \cdot P_{BA}^{C} \cdot \left(q_{P_{BA}}^{C} - q_{P_{DOM}}^{C}\right)\right) \]

\[ GROW_{BA}^{PO4} = GROW_{BA}^{LDOP} \cdot \frac{PO4_{BA}}{P_{DOM} \cdot \min\left(1, P_{BA}^{C}\right)} \]

\[ GROW_{BA}^{PO4} = GROW_{BA}^{LDOP} + GROW_{BA}^{SDOP} + GROW_{BA}^{PO4} \]

(A5.7) Bacterial Respiration

\[ RESP_{BA} = \zeta_{NO3} \cdot GROW_{BA}^{NO3} + resp_{BA}^{R} \cdot T_{fun} \cdot C_{BA} \]

\[ + \left( resp_{min,BA}^{A} + \left( resp_{max,BA}^{A} - resp_{min,BA}^{A} \right) \cdot \text{EXP}\left(-b_{Rsp} \cdot GROW_{BA}^{C}\right) \right) \cdot GROW_{BA}^{C} \]

(A5.8) Refractory DOM release

\[ REMF_{BA}^{C} = ex_{REMBA,BA} \cdot C_{BA} \]

\[ REMF_{BA}^{N} = q_{N_{REMBA}}^{C} \cdot REMF_{BA}^{C} \]

\[ REMF_{BA}^{P} = q_{P_{REMBA}}^{C} \cdot REMF_{BA}^{C} \]

(A5.9) Inorganic nutrients regenerating / semilabile DOM excretion

if \( q_{C_{REMBA}}^{C} > q_{C_{REMBA}}^{C} \) and \( Q_{P_{REMBA}}^{C} > q_{P_{REMBA}}^{C} \) (Carbon in short, regenerate nutrients)

\[ REMI_{BA}^{N} = remi_{BA} \cdot \left(N_{BA} - C_{BA} \cdot q_{C_{REMBA}}^{C}\right) \]

\[ REMI_{BA}^{P} = remi_{BA} \cdot \left(P_{BA} - C_{BA} \cdot q_{C_{REMBA}}^{C}\right) \]

\[ EXCR_{BA}^{SDOC} = EXCR_{BA}^{SDON} = EXCR_{BA}^{SDOP} = 0 \]

elseif \( q_{C_{REMBA}}^{C} < q_{C_{REMBA}}^{C} \) and \( Q_{P_{REMBA}}^{C} < q_{P_{REMBA}}^{C} \) (Nitrogen in short, release SDOC & SDOP)

\[ EXCR_{BA}^{SDOC} = ex_{ADJ,BA} \cdot \left(C_{BA} - N_{BA} \cdot q_{C_{REMBA}}^{C}\right) \]
\[ \text{EXCR}_{\text{SDON}}^{\text{BA}} = 0 \]

\[ \text{EXCR}_{\text{SDOP}}^{\text{BA}} = e x_{\text{ADJ,BA}} \cdot \left( P_{\text{BA}} - N_{\text{BA}} / q_{\text{N,BA}}^C \cdot q_{\text{P,BA}}^C \right) \]

\[ \text{REMI}_N^{\text{BA}} = \text{REMI}_P^{\text{BA}} = 0 \]

else (Phosphor in short, release SDOC & SDON)

\[ \text{EXCR}_{\text{SDOC}}^{\text{BA}} = e x_{\text{ADJ,BA}} \cdot \left( C_{\text{BA}} - P_{\text{BA}} / q_{\text{P,BA}}^C \right) \]

\[ \text{EXCR}_{\text{SDON}}^{\text{BA}} = e x_{\text{ADJ,BA}} \cdot \left( N_{\text{BA}} - P_{\text{BA}} / q_{\text{P,BA}}^C \cdot q_{\text{N,BA}}^C \right) \]

\[ \text{EXCR}_{\text{SDOP}}^{\text{BA}} = 0 \]

\[ \text{REMI}_N^{\text{BA}} = \text{REMI}_P^{\text{BA}} = 0 \]

end

(A5.10) Grazing on bacteria by protozoa

\[ \text{GRAZ}_C^{\text{BA}} = T_{\text{fun}} \cdot \mu_{\text{PRT}} \cdot C_{\text{PRT}} \cdot \frac{C_{\text{BA}}^2}{C_{\text{BA}}^2 + g_{\text{BA}}^2 + (C_{\text{PHY}} \cdot g_{\text{BA}} / g_{\text{PHY}})^2 + (C_{\text{UN}} \cdot g_{\text{BA}} / g_{\text{UN}})^2} \]

\[ \text{GRAZ}_N^{\text{BA}} = Q_{\text{N,BA}}^C \cdot \text{GRAZ}_C^{\text{BA}} \]

\[ \text{GRAZ}_P^{\text{BA}} = Q_{\text{P,BA}}^C \cdot \text{GRAZ}_C^{\text{BA}} \]

(A5.11) Viral mortality

\[ \text{MORT}_C^{\text{BA}} = \text{mort}_{\text{BA}} \cdot C_{\text{BA}} \]

\[ \text{MORT}_N^{\text{BA}} = \text{mort}_{\text{BA}} \cdot N_{\text{BA}} \]

\[ \text{MORT}_P^{\text{BA}} = \text{mort}_{\text{BA}} \cdot P_{\text{BA}} \]

(A5.12) Net flux of inorganic nutrients through bacteria

\[ \text{FLUX}_{\text{NH4}}^{\text{BA}} = \text{REMI}_N^{\text{BA}} - \text{GROW}_{\text{NH4}}^{\text{BA}} \]

\[ \text{FLUX}_{\text{NO3}}^{\text{BA}} = -\text{GROW}_{\text{NO3}}^{\text{BA}} \]

\[ \text{FLUX}_{\text{PO4}}^{\text{BA}} = \text{REMI}_P^{\text{BA}} - \text{GROW}_{\text{PO4}}^{\text{BA}} \]

(A5.13) Bacteria net growth

\[ \frac{dC_{\text{BA}}}{dt} = \text{GROW}_{\text{BA}}^C - \text{RESP}_{\text{BA}}^C - \text{REFR}_{\text{BA}}^C - \text{EXCR}_{\text{SDOC}}^{\text{BA}} - \text{GRAZ}_C^{\text{BA}} - \text{MORT}_C^{\text{BA}} \]

\[ \frac{dN_{\text{BA}}}{dt} = \text{GROW}_N^{\text{BA}} - \text{REMI}_N^{\text{BA}} - \text{REFR}_N^{\text{BA}} - \text{EXCR}_{\text{SDON}}^{\text{BA}} - \text{GRAZ}_N^{\text{BA}} - \text{MORT}_N^{\text{BA}} \]

\[ \frac{dP_{\text{BA}}}{dt} = \text{GROW}_P^{\text{BA}} - \text{REMI}_P^{\text{BA}} - \text{REFR}_P^{\text{BA}} - \text{EXCR}_{\text{SDOP}}^{\text{BA}} - \text{GRAZ}_P^{\text{BA}} - \text{MORT}_P^{\text{BA}} \]
6. Protozoan (PRT) processes

(A6.1) Gross growth

\[
\begin{align*}
GROW^C_{PRT} & = \text{Graz}^C_{Phy} + \text{Graz}^C_{Un} + \text{Graz}^C_{Ba} \\
GROW^N_{PRT} & = \text{Graz}^N_{Phy} + \text{Graz}^N_{Un} + \text{Graz}^N_{Ba} \\
GROW^P_{PRT} & = \text{Graz}^P_{Phy} + \text{Graz}^P_{Un} + \text{Graz}^P_{Ba}
\end{align*}
\]

(A6.2) DOM excretion

(a) Labile DOM excretion

\[
\begin{align*}
\text{EXCR}^\text{LDOC}_{PRT} & = f_{ex,PRT} \cdot \text{ex}_{PRT} \cdot \text{GROW}^C_{PRT} \\
\text{EXCR}^\text{LDON}_{PRT} & = f_{ex,PRT} \cdot \text{ex}_{PRT} \cdot \text{GROW}^N_{PRT} \\
\text{EXCR}^\text{LDOP}_{PRT} & = f_{ex,PRT} \cdot \text{ex}_{PRT} \cdot \text{GROW}^P_{PRT}
\end{align*}
\]

(b) Semilabile DOM excretion

\[
\begin{align*}
\text{EXCR}^\text{SDOC}_{PRT} & = (1 - f_{ex,PRT}) \cdot \text{ex}_{PRT} \cdot \text{GROW}^C_{PRT} \\
\text{EXCR}^\text{SDON}_{PRT} & = (1 - f_{ex,PRT}) \cdot \text{ex}_{PRT} \cdot \text{GROW}^N_{PRT} \cdot Q^C_{N,PRT} / q^C_{N,PRT} \\
\text{EXCR}^\text{SDOP}_{PRT} & = (1 - f_{ex,PRT}) \cdot \text{ex}_{PRT} \cdot \text{GROW}^P_{PRT} \cdot Q^C_{P,PRT} / q^C_{P,PRT}
\end{align*}
\]

(A6.3) Respiration

\[
\text{RESP}_{PRT} = \text{resp}^\text{P}_{PRT} \cdot \text{Tfun} \cdot C^A_{PRT} + \text{resp}^\text{P}_{PRT} \cdot \text{GROW}^C_{PRT}
\]

(A6.4) Semilabile DOM release in order to adjust stoichiometry

\[
\begin{align*}
\text{EXCR}^\text{SDOC}_{PRT} & = \text{ex}_{ADJ,PRT} \cdot C^\text{P}_{PRT} \cdot \max \left( 0, 1 - \frac{Q^C_{N,PRT}}{q^C_{N,PRT}} \right) - \frac{Q^C_{P,PRT}}{q^C_{P,PRT}} \\
\text{EXCR}^\text{SDON}_{PRT} & = 0.5 \cdot \text{EXCR}^\text{SDOC}_{PRT} \cdot Q^C_{N,PRT} \\
\text{EXCR}^\text{SDOP}_{PRT} & = 0.5 \cdot \text{EXCR}^\text{SDOC}_{PRT} \cdot Q^C_{P,PRT}
\end{align*}
\]

(A6.5) Inorganic nutrient remineralization in order to adjust stoichiometry

\[
\begin{align*}
\text{REMI}^N_{PRT} & = \text{remi}^\text{P}_{PRT} \cdot \max \left( 0, N^C_{PRT} - q^C_{N,PRT} \cdot C^\text{P}_{PRT} \cdot N^C_{PRT} - q^C_{N,PRT} \cdot q^C_{P,PRT} \cdot P^\text{P}_{PRT} \right) \\
\text{REMI}^P_{PRT} & = \text{remi}^\text{P}_{PRT} \cdot \max \left( 0, P^C_{PRT} - q^C_{P,PRT} \cdot C^\text{P}_{PRT} \cdot P^C_{PRT} - q^C_{P,PRT} \cdot q^C_{N,PRT} \cdot N^\text{P}_{PRT} \right)
\end{align*}
\]

(A6.6) POM production

\[
\begin{align*}
\text{POM}^C_{PRT} & = \text{pom}_{PRT} \cdot \text{GROW}^C_{PRT} \\
\text{POM}^N_{PRT} & = q^C_{N,POM} \cdot \text{POM}^C_{PRT} \\
\text{POM}^P_{PRT} & = q^C_{P,POM} \cdot \text{POM}^C_{PRT}
\end{align*}
\]
(A6.7) Grazing on protozoa by metazoa

\[
\text{GRAZ}_{C}^{\text{PRT}} = T_{\text{fun}} \cdot \mu_{\text{MZ}} \cdot C_{\text{MZ}} \cdot \frac{C_{\text{PRT}}^2}{C_{\text{PRT}}^2 + g_{\text{PRT}}^2 + \left( C_{\text{TR}} \cdot g_{\text{PRT}} / g_{\text{TR}} \right)^2}
\]

\[
\text{GRAZ}_{N}^{\text{PRT}} = Q_{N,\text{PRT}}^C \cdot \text{GRAZ}_{P}^{\text{PRT}}
\]

\[
\text{GRAZ}_{P}^{\text{PRT}} = Q_{P,\text{PRT}}^C \cdot \text{GRAZ}_{P}^{\text{PRT}}
\]

(A6.8) Protozoa net growth rate

\[
\frac{dC_{\text{PRT}}}{dt} = \text{GROW}^{C}_\text{PRT} - \text{RESP}^{C}_{\text{PRT}} - \text{EXCR}^{\text{LDOC}}_{\text{PRT}} - \text{EXCR}^{\text{SDOC}}_{\text{PRT}} - \text{EXCR}^{2\text{SDOC}}_{\text{PRT}} - \text{POM}^{C}_{\text{PRT}} - \text{GRAZ}^{C}_{\text{PRT}}
\]

\[
\frac{dN_{\text{PRT}}}{dt} = \text{GROW}^{N}_{\text{PRT}} - \text{REMI}^{N}_{\text{PRT}} - \text{EXCR}^{\text{LDON}}_{\text{PRT}} - \text{EXCR}^{\text{SDON}}_{\text{PRT}} - \text{EXCR}^{2\text{SDON}}_{\text{PRT}} - \text{POM}^{N}_{\text{PRT}} - \text{GRAZ}^{N}_{\text{PRT}}
\]

\[
\frac{dP_{\text{PRT}}}{dt} = \text{GROW}^{P}_{\text{PRT}} - \text{REMI}^{P}_{\text{PRT}} - \text{EXCR}^{\text{LDOP}}_{\text{PRT}} - \text{EXCR}^{\text{SDOP}}_{\text{PRT}} - \text{EXCR}^{2\text{SDOP}}_{\text{PRT}} - \text{POM}^{P}_{\text{PRT}} - \text{GRAZ}^{P}_{\text{PRT}}
\]

7. Metazoa (MZ) Processes

(A7.1) Gross growth

\[
\text{GROW}^{C}_{\text{MZ}} = \text{GRAZ}^{C}_{\text{TR}} + \text{GRAZ}^{C}_{\text{PRT}}
\]

\[
\text{GROW}^{N}_{\text{MZ}} = \text{GRAZ}^{N}_{\text{TR}} + \text{GRAZ}^{N}_{\text{PRT}}
\]

\[
\text{GROW}^{P}_{\text{MZ}} = \text{GRAZ}^{P}_{\text{TR}} + \text{GRAZ}^{P}_{\text{PRT}}
\]

(A7.2) DOM excretion

(c) Labile DOM excretion

\[
\text{EXCR}^{\text{LDOC}}_{\text{MZ}} = f_{ex,\text{MZ}} \cdot ex_{\text{MZ}} \cdot \text{GROW}^{C}_{\text{MZ}}
\]

\[
\text{EXCR}^{\text{LDON}}_{\text{MZ}} = f_{ex,\text{MZ}} \cdot ex_{\text{MZ}} \cdot \text{GROW}^{N}_{\text{MZ}}
\]

\[
\text{EXCR}^{\text{LDOP}}_{\text{MZ}} = f_{ex,\text{MZ}} \cdot ex_{\text{MZ}} \cdot \text{GROW}^{P}_{\text{MZ}}
\]

(d) Semilabile DOM excretion

\[
\text{EXCR}^{\text{SDOC}}_{\text{MZ}} = \left( 1 - f_{ex,\text{MZ}} \right) \cdot ex_{\text{MZ}} \cdot \text{GROW}^{C}_{\text{MZ}}
\]

\[
\text{EXCR}^{\text{SDON}}_{\text{MZ}} = \left( 1 - f_{ex,\text{MZ}} \right) \cdot ex_{\text{MZ}} \cdot \text{GROW}^{N}_{\text{MZ}} \cdot Q_{N,\text{MZ}}^C / q_{N,\text{MZ}}^C
\]

\[
\text{EXCR}^{\text{SDOP}}_{\text{MZ}} = \left( 1 - f_{ex,\text{MZ}} \right) \cdot ex_{\text{MZ}} \cdot \text{GROW}^{P}_{\text{MZ}} \cdot Q_{P,\text{MZ}}^C / q_{P,\text{MZ}}^C
\]

(A7.3) Respiration

\[
\text{RESP}_{\text{MZ}} = resp^{B}_{\text{MZ}} \cdot T_{\text{fun}} \cdot C_{\text{MZ}} + \text{resp}^{A}_{\text{MZ}} \cdot \text{GROW}^{C}_{\text{MZ}}
\]

(A7.4) Semilabile DOM release in order to adjust stoichiometry
\[ EXCR_{SDOC}^{MZ} = ex_{ADJ,MZ} \cdot C_{MZ} \cdot \max \left( 0, 1 - \frac{Q_{N,MZ}^C}{Q_{N,MZ}^C}, 1 - \frac{Q_{P,MZ}^C}{Q_{P,MZ}^C} \right) \]

\[ EXCR_{SDON}^{MZ} = 0.5 \cdot EXCR_{SDOC}^{MZ} \cdot Q_{N,MZ}^C \]

\[ EXCR_{SDOP}^{MZ} = 0.5 \cdot EXCR_{SDOC}^{MZ} \cdot Q_{P,MZ}^C \]

(A7.5) Inorganic nutrient remineralization in order to adjust stoichiometry

\[ REMI_{MZ}^N = remi_{MZ} \cdot \max \left( 0, N_{MZ} - q_{N,MZ}^C \cdot C_{MZ}, N_{MZ} - q_{N,MZ}^C \cdot P_{MZ} \right) \]

\[ REMI_{MZ}^P = remi_{MZ} \cdot \max \left( 0, P_{MZ} - q_{P,MZ}^C \cdot C_{MZ}, P_{MZ} - q_{P,MZ}^C \cdot N_{MZ} \right) \]

(A7.6) POM production

\[ POM_{MZ}^C = pom_{MZ} \cdot GROW_{C}^{MZ} \]
\[ POM_{MZ}^N = q_{N,POM} \cdot POM_{MZ}^C \]
\[ POM_{MZ}^P = q_{P,POM} \cdot POM_{MZ}^C \]

(A7.7) Refractory DOM release

\[ REF_{MZ}^C = ex_{REFR,MZ} \cdot C_{MZ} \]
\[ REF_{MZ}^N = q_{N,REFR} \cdot REF_{MZ}^C \]
\[ REF_{MZ}^P = q_{P,REFR} \cdot REF_{MZ}^C \]

(A7.8) Removal by higher-level zooplankton

\[ REMV_{MZ}^N = remv_{MZ} \cdot C_{MZ} \]
\[ REMV_{MZ}^N = Q_{N,MZ}^C \cdot REMV_{MZ}^C \]
\[ REMV_{MZ}^P = Q_{P,MZ}^C \cdot REMV_{MZ}^C \]

(A7.9) POM production by higher-level zooplankton

\[ POM_{HZ}^C = f_{POM,HZ} \cdot REMV_{MZ}^C \]
\[ POM_{HZ}^N = f_{POM,HZ} \cdot REMV_{MZ}^N \]
\[ POM_{HZ}^P = f_{POM,HZ} \cdot REMV_{MZ}^P \]

(A7.10) Semilabile DOM production by higher-level zooplankton

\[ EXCR_{SDOC}^{HZ} = f_{SDOM,HZ} \cdot REMV_{MZ}^C \]
\[ EXCR_{SDON}^{HZ} = f_{SDOM,HZ} \cdot REMV_{MZ}^N \]
\[ EXCR_{SDOP}^{HZ} = f_{SDOM,HZ} \cdot REMV_{MZ}^P \]
(A7.11) Inorganic nutrients regeneration by higher-level zooplankton

\[
REM_1^N = REMV_1^{N} - POM_1^{N} - EXCR_1^{SDON}
\]

\[
REM_1^P = REMV_1^{P} - POM_1^{P} - EXCR_1^{SDOP}
\]

(A7.12) Protozoa net growth rate

\[
\frac{dC}{dt} = GROW_C - RESP_C - EXCR_{LDOC} - EXCR_{SDOC} - EXCR_{SDOP} - POM_C - REFR_C - REMV_C
\]

\[
\frac{dN}{dt} = GROW_N - REMI_N - EXCR_{LDON} - EXCR_{SDON} - EXCR_{SDOP} - POM_N - REFR_N - REMV_N
\]

\[
\frac{dP}{dt} = GROW_P - REMI_P - EXCR_{LDOP} - EXCR_{SDOP} - EXCR_{SDOP} - POM_P - REFR_P - REMV_P
\]

8. Detritus (DET) processes

(A8.1) Dissolution of DET

\[
DISS_{DET}^C = diss \cdot C_{DET}
\]

\[
DISS_{DET}^N = prf_N \cdot diss \cdot N_{DET}
\]

\[
DISS_{DET}^P = prf_P \cdot diss \cdot P_{DET}
\]

(A8.2) Net Change of DET

\[
\frac{dC_{DET}}{dt} = POM_{PHY} + POM_{TR} + POM_{UN} + POM_{PRT} + POM_C + POM_{MZ} - DISS_{DET}^C
\]

\[
\frac{dN_{DET}}{dt} = POM_{PHY} + POM_{TR} + POM_{UN} + POM_{PRT} + POM_N + POM_{MZ} - DISS_{DET}^N
\]

\[
\frac{dP_{DET}}{dt} = POM_{PHY} + POM_{TR} + POM_{UN} + POM_{PRT} + POM_P + POM_{MZ} - DISS_{DET}^P
\]

9. Inorganic nutrients (NH4,NO3,PO4) processes

(A9.1) Nitrification

\[
NTRF = r_{ntrf} \cdot NH4
\]

(A9.2) Dissolved inorganic nutrient change rates

\[
\frac{dNH4}{dt} = FLUX_{BA}^{NH4} + REMI_{PRT}^{N} + REMI_{MZ}^{N} + REMI_{TR}^{N} + EXCR_{TR}^{NH4}
\]

\[
+ \ EXCR_{UN}^{NH4} - GROW_{PHY}^{NH4} - GROW_{TR}^{NH4} - GROW_{UN}^{NH4} - NTRF
\]
\[
\frac{d\text{NO}_3}{dt} = \text{FLUX}_{\text{BA}}^{\text{NO}_3} - \text{GROW}_{\text{PHY}}^{\text{NO}_3} - \text{GROW}_{\text{TR}}^{\text{NO}_3} - \text{GROW}_{\text{UN}}^{\text{NO}_3} + \text{NTRF}
\]

\[
\frac{d\text{PO}_4}{dt} = \text{FLUX}_{\text{BA}}^{\text{PO}_4} + \text{REMI}_{\text{PRT}}^{\text{PO}_4} + \text{REMI}_{\text{MZ}}^{\text{PO}_4} + \text{REMI}_{\text{HZ}}^{\text{PO}_4} - \text{GROW}_{\text{PHY}}^{\text{PO}_4} - \text{GROW}_{\text{TR}}^{\text{PO}_4} - \text{GROW}_{\text{UN}}^{\text{PO}_4}
\]

10. Dissolved Organic Matter (DOM) processes

(A10.1) Conversion of semilabile DOM to refractory DOM

\[
\text{REFR}_{\text{SDOM}}^{\text{C}} = \text{ex}_{\text{REFR}, \text{SDOM}} \cdot C_{\text{SDOM}} \cdot \text{EXP} \left( 1 - \min \left( \frac{Q_{N, \text{SDOM}}^C}{Q_{N, \text{REFR}}^C}, \frac{Q_{P, \text{SDOM}}^C}{Q_{P, \text{REFR}}^C} \right) \right)
\]

\[
+ \max \left( 0, C_{\text{SDOM}} - \frac{N_{\text{SDOM}}}{Q_{N, \text{REFR}}}, C_{\text{SDOM}} - \frac{P_{\text{SDOM}}}{Q_{P, \text{REFR}}} \right)
\]

\[
\text{REFR}_{\text{SDOM}}^{N} = Q_{N, \text{REFR}} \cdot \text{REFR}_{\text{SDOM}}^{C}
\]

\[
\text{REFR}_{\text{SDOM}}^{P} = Q_{P, \text{REFR}} \cdot \text{REFR}_{\text{SDOM}}^{C}
\]

(A10.2) DOM change rate

\[
\frac{dC_{\text{LDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{LDOM}} + \text{EXCR}_{\text{TR}}^{\text{LDOM}} + \text{EXCR}_{\text{UN}}^{\text{LDOM}} + \text{EXCR}_{\text{PRT}}^{\text{LDOM}} + \text{EXCR}_{\text{MZ}}^{\text{LDOM}} + \text{MORT}_{\text{BA}}^{C} - \text{GROW}_{\text{BA}}^{\text{LDOM}}
\]

\[
\frac{dN_{\text{LDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{LDON}} + \text{EXCR}_{\text{TR}}^{\text{LDON}} + \text{EXCR}_{\text{UN}}^{\text{LDON}} + \text{EXCR}_{\text{PRT}}^{\text{LDON}} + \text{EXCR}_{\text{MZ}}^{\text{LDON}} + \text{MORT}_{\text{BA}}^{N} - \text{GROW}_{\text{BA}}^{\text{LDON}}
\]

\[
\frac{dP_{\text{LDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{LDOP}} + \text{EXCR}_{\text{TR}}^{\text{LDOP}} + \text{EXCR}_{\text{UN}}^{\text{LDOP}} + \text{EXCR}_{\text{PRT}}^{\text{LDOP}} + \text{EXCR}_{\text{MZ}}^{\text{LDOP}} + \text{MORT}_{\text{BA}}^{P} - \text{GROW}_{\text{BA}}^{\text{LDOP}}
\]

\[
\frac{dC_{\text{SDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{SDOC}} + \text{EXCR}_{\text{TR}}^{\text{SDOC}} + \text{EXCR}_{\text{UN}}^{\text{SDOC}} + \text{EXCR}_{\text{PRT}}^{\text{SDOC}} + \text{EXCR}_{\text{MZ}}^{\text{SDOC}} + \text{EXCR}_{\text{HZ}}^{\text{SDOC}} + \text{DISS}_{\text{SDOM}}^{C} - \text{REFR}_{\text{SDOM}}^{\text{C}} - \text{GROW}_{\text{BA}}^{\text{SDOC}}
\]

\[
\frac{dN_{\text{SDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{SDON}} + \text{EXCR}_{\text{TR}}^{\text{SDON}} + \text{EXCR}_{\text{UN}}^{\text{SDON}} + \text{EXCR}_{\text{PRT}}^{\text{SDON}} + \text{EXCR}_{\text{MZ}}^{\text{SDON}} + \text{EXCR}_{\text{HZ}}^{\text{SDON}} + \text{DISS}_{\text{SDOM}}^{N} - \text{REFR}_{\text{SDOM}}^{\text{N}} - \text{GROW}_{\text{BA}}^{\text{SDON}}
\]

\[
\frac{dP_{\text{SDOM}}}{dt} = \text{EXCR}_{\text{PHY}}^{\text{SDOP}} + \text{EXCR}_{\text{TR}}^{\text{SDOP}} + \text{EXCR}_{\text{UN}}^{\text{SDOP}} + \text{EXCR}_{\text{PRT}}^{\text{SDOP}} + \text{EXCR}_{\text{MZ}}^{\text{SDOP}} + \text{EXCR}_{\text{HZ}}^{\text{SDOP}} + \text{DISS}_{\text{SDOM}}^{P} - \text{REFR}_{\text{SDOM}}^{\text{P}} - \text{GROW}_{\text{BA}}^{\text{SDOP}}
\]

11. Diagnostic variables used for data assimilation

(A11.1) Primary production

\[
\text{PrPr} = \text{GROW}_{\text{PHY}}^{C} + \text{GROW}_{\text{TR}}^{C} + \text{GROW}_{\text{UN}}^{C} - \text{RESP}_{\text{PHY}} - \text{RESP}_{\text{TR}} - \text{RESP}_{\text{UN}} - \text{EXCR}_{\text{PHY}}^{\text{LDOM}} - \text{EXCR}_{\text{TR}}^{\text{LDOM}} - \text{EXCR}_{\text{UN}}^{\text{LDOM}} - \text{EXCR}_{\text{PHY}}^{\text{SDOM}} - \text{EXCR}_{\text{TR}}^{\text{SDOM}} - \text{EXCR}_{\text{UN}}^{\text{SDOM}}
\]
(A11.2) Heterotrophic bacterial production

\[ B\text{Pr} = G\text{ROW}_{BA}^C - R\text{ESP}_{BA} \]

APPENDIX B. Adjoint Model Construction

Let the vector \( x = (x_1, x_2, L, x_m) \) denote the model parameters. The model can be thought as a sequence of calculations, starting with parameters, where each quantity calculated is a model dependent variable. Let \( z = (z_1, z_2, L, z_N) \) denote the model dependent variables and \( J = z_{N+1} \) to represent the cost function. A model can be rewritten as:

\[
\begin{align*}
  z_1 &= f_1(x), \\
  z_2 &= f_2(x, z_I), \\
  L &= f_N(x, z_I, z_2, L, z_{N-1}), \\
  z_N &= f_N(x, z_I, z_2, L, z_{N-1}), \\
  z_{N+1} &= f_{N+1}(x, z_I, z_2, L, z_N)
\end{align*}
\]

where the last equation is a cost function which is calculated after all the model equations. The target for generating the adjoint model is to construct codes to calculate the gradient of the cost function \( J = z_{N+1} \) with respect to each component of the parameter \( x \). However, as \( z_{N+1} \) is not explicitly represented by \( x \), the partial derivative calculation would require extensive application of the chain rule. Lagrange multipliers are introduced to avoid this difficulty:

\[
L(x, z, \lambda) = z_{N+1} - \lambda_1(z_1 - f_1(x)) - \sum_{i=2}^{N+1} \lambda_i(z_i - f_i(x, z_1, z_2, L, z_{i-1})),
\]

where \( \lambda = (\lambda_1, \lambda_2, L, \lambda_{N+1}) \) is the vector of Lagrange multipliers. At a saddle point of the Lagrange function, the partial derivatives of \( L \) with respect to \( x, z \) and \( \lambda \) vanish simultaneously. With

\[
\frac{\partial L}{\partial \lambda_n} = 0, \quad n = 1, 2, L, N + 1
\]

one obtains the model equation (B1). In other words, at the saddle point of the Lagrange function, the original model is satisfied.

With

\[
\frac{\partial L}{\partial z_{N+1}} = 1 - \lambda_{N+1} = 0, \quad \lambda_{N+1} = 1.
\]

And
Thus starting from equation (B4), all the Lagrange multipliers can be calculated, which is done in the reverse order of the model equations. As we see at the saddle point the original model (18) is satisfied and equation (B2) gives

\[ L = z_{N+1} = J, \]

so that the gradients of them are equal \( \nabla_X J = \nabla_X L \). Because \( L = L(x, z, \lambda) \) and \( z \) and \( \lambda \) depend on \( x \), by the chain rule,

\[
\frac{\partial L}{\partial z_i} = \left( \frac{\partial L}{\partial x_k} \right) \frac{\partial x_k}{\partial z_i} + \sum_{i=1}^{N+1} \left( \frac{\partial L}{\partial z_i} \frac{\partial \lambda}{\partial x_k} \right) .
\]

And \( \frac{\partial L}{\partial z_i} = 0 \), so that

\[
\left( \frac{\partial L}{\partial x_k} \right) = \sum_{i=1}^{N+1} \frac{\partial L}{\partial x_k} \lambda_i, \quad 1 \leq k \leq m .
\]

The equation (B7) is the form used to generate adjoint code. The following is an example to show how to construct adjoint code.

Considering 2 lines of model code:

\[
Y = G( X, L ) \quad Z = F( X, Y, L ) .
\]

the Lagrange function would be:

\[
L = L - \lambda_Y ( Y - G( X, L ) ) - \lambda_Z ( Z - F( X, Y, L ) ) - L .
\]

The saddle point of \( L \) gives

\[
\frac{\partial L}{\partial Y} = -\lambda_Y + \lambda_Z \frac{\partial F}{\partial Y} + L = 0, \quad \lambda_Y = \lambda_Z \frac{\partial F}{\partial Y} + L .
\]

From equation (B7) we see the \( \lambda_Y \) has to be accumulated for every equation where \( Y \) appears on the right side, so a model code can be written as

\[
\lambda_Y = \lambda_Y + \lambda_Z \frac{\partial F}{\partial Y} .
\]

where \( \lambda_Y \) has to be initialized to zero before its first use.

Similarly for \( X \),

\[
\lambda_Y = \lambda_Y + \lambda_Z \frac{\partial F}{\partial Y} .
\]
\[
\lambda_X = \lambda_X + \lambda_Z \frac{\partial F}{\partial X}
\]
\[
\lambda_X = \lambda_X + \lambda_Y \frac{\partial G}{\partial X}.
\]  \hfill (B12)

As indicated by equation (B4) and (B5), the procedure of calculating Lagrange multipliers have to be done in reverse order. Thus in equation (B12) \( \lambda_X \) is calculated first from the second line of the model code and then the first line. Also, the method requires each line be differentiable.

APPENDIX C. Limited-memory Quasi-Newton Method

M1QN3 solves the problem of minimizing an unconstrained function \( f \) of \( n \) variables:

\[
\min \left\{ f \left( x \right) : x \in \mathbb{R}^n \right\}.
\]  \hfill (C1)

where the function \( f \) is supposed to be smooth (twice-differentiable). In our application \( f \) is the cost function and \( x \) are model parameters. The method is iterative, starting with an initial point \( x_0 \), or initial guess of model parameters in our application, and producing a sequence of points \( x_k \) that converge to the optimum \( x^* \). Denote the \( f \) at the point \( x_k \) by \( f_k \) and its gradient by \( \nabla f_k \), and the Hessian matrix (the matrix of second partial derivatives) by \( B_k \).

Here we first discuss how to solve the problem with the original Newton’s method. The new iterate \( x_{k+1} \) is found as a function of \( x_k \) as follows. For any point \( x \), define \( p = x - x_k \). The second order Taylor approximation around \( x_k \) is

\[
m_k \left( p \right) = f_k + p^T \nabla f_k + \frac{1}{2} p^T B_k p.
\]  \hfill (C2)

This defines a quadratic model of the function near the point \( x_k \). Its gradient with respect to \( x_k \) is \( \nabla m_k \left( p \right) = \nabla f_k + B_k p \). Thus the equation (C2) is minimized when its gradient equals zero, or

\[
p_k = - B_k^{-1} \nabla f_k.
\]  \hfill (C3)

\( p_k \) thus represents the descent direction of the cost function. Here the function is \( f \) assumed to be convex.

(However, according to Nocedal & Wright (2006), the convexity assumption can be relaxed, in which case the method can be implemented in such a way that it is still guaranteed to converge to a local minimum.) Therefore, finding \( p_k \) requires computing the inverse of Hessian matrix which may be computationally expensive.

The next point \( x_{k+1} \) is then found via a line-search in the direction of \( p_k \): for some \( \alpha_k \in (0, \infty) \), \( x_{k+1} = x_k + \alpha_k p_k \). In M1QN3 \( \alpha_k \) is determined to satisfy the following 2 Wolfe’s conditions:

\[
f \left( x_k + \alpha_k p_k \right) \leq f \left( x_k \right) + \omega_1 \alpha_k \left\langle \nabla f_k , p_k \right\rangle.
\]  \hfill (C4)

\[
\left\langle \nabla f \left( x_k + \alpha_k p_k \right) , p_k \right\rangle \geq \omega_2 \left\langle \nabla f_k , p_k \right\rangle.
\]  \hfill (C5)
where $\langle \cdot, \cdot \rangle$ represents inner product. It is necessary to have $0 < \alpha_1 < 1/2$ and $\alpha_1 < \alpha_2 < 1$.

M1QN3 set $\alpha_1 = 0.0001$ and $\alpha_2 = 0.9$.

Because the computation of the inverse of Hessian matrix, denoted as $H_k = B_k^{-1}$ at iteration $k$, may be very expensive, there are some quasi-Newton methods that only compute an approximation of $H_k$ that is based on the change in gradient between iterations $k-1$ and $k$. M1QN3 uses the BFGS method, which is considered to be the most effective quasi-Newton method to compute $H_k$. It is more numerically stable and has very effective self-correcting properties. Defining the displacement vector $s_{k-1} = x_k - x_{k-1}$ and the change in gradient $y_{k-1} = \nabla f_k - \nabla f_{k-1}$, and $\rho_{k-1} = (y_{k-1}^T s_{k-1})^{-1}$, the BFGS method derives the solution as

$$H_k = \left( I - \rho_{k-1} s_{k-1} y_{k-1}^T \right) H_{k-1} \left( I - \rho_{k-1} y_{k-1} s_{k-1}^T \right) + s_{k-1} \rho_{k-1} s_{k-1}^T,$$

(C6)

where the new inverse Hessian approximation $H_k$ depends on $H_{k-1}$, $s_{k-1}$ and $y_{k-1}$, all of which has been computed from previous iteration. Take the new result back to equation (C3) and then the descent direction $p_k$ can be calculated.

For the initial inverse Hessian matrix $H_0$, M1QN3 uses the identity matrix which in practice is satisfactory.

However, it is not possible to use BFGS on problems with a very high number $n$ of variables because it is impossible to store or manipulate the approximate inverse Hessian matrix $H$, which is of size $n^2$. M1QN3 uses a limited-memory BFGS to solve the problem by storing $H$ in a compressed form that requires storing only a constant multiple of vectors of length $n$. For a model defined integer $m$, at iteration $k \geq m + 1$, $H_k \nabla f_k$ is directly computed by the following $m$ pairs of vectors:

$$\left\{ (y_i, s_i) : k - m \leq i \leq k - 1 \right\}.$$

(C7)

The computed $H_k \nabla f_k$ can be brought back to equation (C3) to compute descent direction directly. Thus only $O(mn)$ elements are stored compared to store the $H_k$ of size $O(n^2)$. A higher $m$ can make convergence quicker but requires more storage. A default $m = 7$ suggested by M1QN3 works fine. More mathematical details are given in Gilbert & Lemaréchal (1989).
Supplement 2. Point-to-point comparison of observations and modeled results

Fig. S1. Comparison between observations and their model equivalents. Both observations and the model equivalents are normalized to observed mean. 3 modeling sites include Arabian Sea (AS), Equatorial Pacific (EQP) and Hawaii Ocean Time-series Station ALOHA (HOT). NO3: Nitrate; PO4: phosphate; PHYn: phytoplankton nitrogen biomass (exclude Trichodesmium spp.); CHL: Chlorophyll a; PP: particulate primary production; Z: mesozooplankton carbon biomass; BAc: heterotrophic bacterial carbon biomass; BP: heterotrophic bacterial production; sDOC, sDON & sDOP: semilabile dissolved organic carbon, nitrogen and phosphorus; POC, PON & POP: particulate organic carbon, nitrogen and phosphorus; STc, STn & STp: particle carbon, nitrogen and phosphorus flux collected by sediment traps at 800m for AS and EQP and at 150m for HOT.
AS sDON - Normalized to Mean Observation

AS POC - Normalized to Mean Observation

AS PON - Normalized to Mean Observation
HOT STn - Normalized to Mean Observation

Model Results vs. Observation

HOT STp - Normalized to Mean Observation

Model Results vs. Observation
Supplement 3. Average modeled state flows

Fig. S2. Modeled average carbon (C), nitrogen (N) and phosphorus (P) fluxes (arrows) and stocks (boxes) normalized to C, N and P components of primary production (stocks normalized to amount of primary production during 1 day) at Arabian Sea (AS), Equatorial Pacific (EQP) and Hawaii Ocean Time-series Station ALOHA (HOT). Red arrows: microbial loop, green arrows: major food web, blue arrows: physical transport including sinking detritus, yellow arrows: detritus production, dashed arrows: carbon respiration and refractory DOM production. Semilabile DOC production and nutrient recycling by implicit higher trophic levels are included in fluxes related to MZ. Primary production is marked on top of each phytoplankton box. 1 unit C = 0.770 (AS), 0.876 (EQP) and 0.312 (HOT) mmol m\(^{-3}\) (d\(^{-1}\)); 1 unit N = 0.107 (AS), 0.122 (EQP) and 0.0494 (HOT) mmol m\(^{-3}\) (d\(^{-1}\)); 1 unit P = 0.00733 (AS), 0.00863 (EQP) and 0.00236 (HOT) mmol m\(^{-3}\) (d\(^{-1}\))
Fig. S2