https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Biogeosciences Discussions

Microbial diversity-informed modelling of polar marine ecosystem functions

Hyewon Heather Kim[1, 2, *], Jeff S. Bowman[3], Ya-Wei Luo[4], Hugh W. Ducklow[5], Oscar M. Schofield[6], Deborah K. Steinberg[7], Scott C. Doney[2]

[1]Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
[2]University of Virginia, Charlottesville 22904, USA
[3]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA
[4]Xiamen University, Xiamen, China
[5]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA
[6]Rutgers University, New Brunswick, NJ 80901, USA
[7]Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA 23062, USA

Correspondence to: Hyewon Heather Kim (hkim@whoi.edu)

Abstract. Heterotrophic marine bacteria utilize organic carbon for growth and biomass synthesis. Thus, their variability is key to the balance between the production and consumption of organic matter and ultimately particle export in the ocean. Here we investigate a potential link between bacterial traits and ecosystem functions in a rapidly changing polar marine ecosystem based on a bacteria-oriented ecosystem model. Using a data-assimilation scheme we utilize the observations of bacterial groups with different physiological states to constrain the group-specific bacterial ecosystem functions. We also investigate the association of the modelled bacterial and other ecosystem functions with eight recurrent modes representative of different bacterial taxonomic traits. High nucleic acid (HNA) bacteria show relatively high cell-specific bacterial production, respiration, and utilization of the semi-labile dissolved organic carbon pool compared to low nucleic acid (LNA) bacteria. Both taxonomy and physiological states of the bacteria are strong predictors of bacterial carbon demand, net primary production, and particle export. Numerical experiments under perturbed climate conditions show overall increased bacterial activity and a potential shift from LNA- to HNA-dominated bacterial communities in a warming ocean. Microbial diversity via different taxonomic and physiological traits informs our ecosystem model, providing insights into key bacterial and ecosystem functions in a changing environment.

1 Introduction

Microbes regulate many key ecosystem functions in the marine food web. Unicellular primary producers fix organic carbon (i.e., an ecosystem function termed primary production), while heterotrophic marine bacteria and archaea (hereafter bacteria) utilize the fixed organic carbon for growth and biomass synthesis (i.e., an ecosystem function termed bacterial production) (Azam et al. 1983). Thus, the variability in the abundance and activity of bacteria is central to understanding the balance between production and consumption of organic matter and ultimately particle export. In flow cytometric analyses, bacteria

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

cluster into two groups of cells with different nucleic acid content, including high nucleic acid (HNA) and low nucleic acid (LNA) cells (Bouvier, Giorgio, and Gasol 2007; Gasol et al. 1999). These two groups may represent lineages (Schattenhofer et al. 2011; Vila-Costa et al. 2012) or physiological states (Bowman et al. 2017), and HNA cells are generally bigger in both cell and genome size compared to LNA cells (Bouvier, Giorgio, and Gasol 2007; Calvo-Díaz and Morán 2006). The significance of HNA vs. LNA cells in determining distinct ecosystem states and functions has been investigated, but much is still unknown. In a recent study along the West Antarctic Peninsula (WAP), the high dimensionality of the bacterial community structure data was reduced via emergent self-organizing maps and subdivided into a small number of bacterial modes associated with specific taxonomic and functional traits (Bowman et al. 2017). Bowman et al. (2017) demonstrated that a combination of taxonomy, physiological structure (i.e., HNA and LNA cells), and abundance of bacterial communities explained up to 73% of the variance in bulk bacterial production. Their findings implied that bacterial diversity (with different physiological and taxonomic traits) could inform a predictive ecosystem model to further explore ecologically important questions, such as: Can bacterial traits predict key ecosystem functions such as primary production and particle export? What are the underlying mechanisms driving the trait-ecosystem function relationship? And how will this relationship be impacted by climate change?

The WAP is a rapidly warming marine ecosystem, with resulting changes in physical and ecological processes (Clarke et al. 2009; Cook et al. 2005; Ducklow et al. 2007; King 1994; Meredith and King 2005; Stammerjohn et al. 2008; Vaughan et al. 2003; Vaughan 2006; Whitehouse et al. 2008). Routine monitoring since 1991 through the Palmer Long-Term Ecological Research project (Palmer LTER) has revealed climate-driven variations in seasonal phytoplankton accumulation (Saba et al. 2014; Schofield et al. 2017), bacterial dynamics (Bowman and Ducklow 2015; Hugh W. Ducklow et al. 2012a; Kim and Ducklow 2016; Luria et al. 2017; Luria, Ducklow, and Amaral-Zettler 2014), nutrient drawdown (Kim et al. 2016), and micro- and macrozooplankton dynamics (Garzio and Steinberg 2013; Steinberg et al. 2015; Thibodeau et al. 2019). The wealth of Palmer LTER observations enabled the construction of a numerical marine ecosystem model for the coastal WAP region (Kim et al. In Review) based on existing regional test-bed models (Friedrichs 2001; Friedrichs et al. 2007; Friedrichs, Hood, and Wiggert 2006; Luo et al. 2010, 2012). This WAP ecosystem model was compared against the roughly bi-weekly, water column time-series data available over the growth season near Palmer Station on Anvers Island, Antarctica (64.77° S, 64.05° W) that recorded seasonal variations in phytoplankton bloom initiation, peak, and termination as well as biogeochemistry modulated by variations in surface light, mixed layer depth, and sea-ice cover. The model also utilized a data assimilation scheme to minimize the misfit between model output and the observational data via a variational adjoint method (Lawson et al. 1995), simulating key microbial and ecosystem functions driven by plankton trophic interactions. Serving as a mechanistic model, assimilation of the Palmer LTER observations constrained poorly measured bacterial processes (e.g., respiration, viral and grazing mortality, growth efficiency, carbon demand, and utilization of dissolved organic matter with varying lability) and enabled model predictions of the microbial system state in changing environments. However, incorporating molecular observations into an ecosystem model is a challenge due to differences in how levels of biological organization are treated in observations and models (Hellweger 2020) and the high dimensionality of environmental microbial molecular data. In other

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020


Biogeosciences Discussions Open Access EGU

words, molecular-level changes may not directly translate into a clear picture of changes in community structure or resulting changes to bulk ecosystem functions.

In this study, we explore a potential link between bacterial traits and ecosystem functions in the warming coastal WAP, using a bacteria-oriented ecosystem model modified from the WAP data-assimilative model (Kim et al., In Review). The bacterial traits here include both physiological and taxonomic traits. For bacterial physiological traits, our model explicitly simulates the dynamics of the two ubiquitous bacterial groups of differing nucleic acid content, a HNA group and a LNA group, by directly assimilating the group-specific biomass observations. For bacterial taxonomic traits, taxonomic modes (calculated by Bowman et al. 2017) are compared to model output values at the corresponding time point, with the assumption that taxonomy provides information about microbial ecosystem process and structure. We note that in contrast to genome-scale, metabolic flux, or gene-centric models (Coles et al. 2017; Feist et al. 2009; Reed et al. 2014), our study combines statistical products from genomic analyses (mode derivation from bacterial 16S rRNA gene sequence data) with numerical ecosystem modelling to incorporate molecular information into ecosystem-level dynamics.

2 Material and Methods

2.1 Bacteria-oriented ecosystem model

Our bacteria-oriented ecosystem model simulates 12 state variables, including diatoms, cryptophytes, HNA bacteria, LNA bacteria, microzooplankton, krill, labile dissolved organic matter (LDOM), semi-labile DOM (SDOM), ammonium, nitrate, phosphate, and detritus (Figure 1). Refractory DOM (RDOM) and higher trophic levels are implicitly represented as model closure terms. Distinct to our model compared to the original WAP ecosystem model is the inclusion of HNA and LNA bacterial compartments (as biomass) and the partitioning of the bulk bacterial productivity by each bacterial compartment. The model is forced by mixed layer depth (MLD), photosynthetically active radiation (PAR) at the ocean surface, sea-ice concentration, water-column temperature, and eddy diffusivity, and simulates the stocks and flows of C through the model state variables using a constant time step of 1 hour and a second-order Runge-Kutta scheme. An overview of the ecosystem model (Text S1) and full bacterial model schemes are found in the Supplementary Material (Text S2). In essence, the time rate of change of the group-specific biomass (constrained by observations, see Material and Methods 2.4) is calculated as follows:

$$\frac{dB_{HNA}}{dt} = \frac{(GROW^{LDOC}_{HNA} + GROW^{SDOC}_{HNA} - RESP_{HNA} - GRAZ^{C}_{HNA} - EXCR^{C}_{HNA} - REFR^{C}_{HNA} - MORT^{C}_{HNA})}{t} \quad (1)$$

where B_{HNA} is biomass (mmol C m^{-3}), $GROW^{LDOC}_{HNA}$ is LDOC consumption (mmol C m^{-3} d^{-1}; Equation S4), $GROW^{SDOC}_{HNA}$ is SDOC consumption (mmol C m^{-3} d^{-1}; Equation S5), $RESP_{HNA}$ is respiration (mmol C m^{-3} d^{-1}; Equation S9), $GRAZ^{C}_{HNA}$ is C-specific grazed amount of cells by microzooplankton (mmol C m^{-3} d^{-1}; Equation S10), $EXCR^{C}_{HNA}$ is C excretion (mmol C m^{-3} d^{-1}; Text S2), $REFR^{C}_{HNA}$ is RDOC excretion (mmol C m^{-3} d^{-1}; Text S2), and $MORT^{C}_{HNA}$ is viral mortality (mmol C m^{-3} d^{-1}; Text S2) of the HNA group (the same form applies to LNA group below).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020


Biogeosciences Discussions

$$\frac{dB_{LNA}}{dt} = \frac{(GROW^{LDOC}_{LNA} + GROW^{SDOC}_{LNA} - RESP_{LNA} - GRAZ^{C}_{LNA} - EXCR^{C}_{LNA} - REFR^{C}_{LNA} - MORT^{C}_{LNA})}{t} \quad (2)$$

By contrast, the group-specific production (BP_{HNA} and BP_{LNA}, mmol C m^{-3} d^{-1}) is determined during optimization, given that it is the sum of the group-specific production (i.e., bulk bacterial production = $BP_{HNA} + BP_{LNA}$) that is constrained by observations:

$$BP_{HNA} = \frac{(GROW^{LDOC}_{HNA} + GROW^{SDOC}_{HNA} - RESP_{HNA})}{t} \quad (3)$$

$$BP_{LNA} = \frac{(GROW^{LDOC}_{LNA} + GROW^{SDOC}_{LNA} - RESP_{LNA})}{t} \quad (4)$$

2.2 Modelling framework

The modelling framework in this study consists of a dynamic (mechanistic) part and a data-driven part (Figure 2). The dynamic part represents the processes associated with the data-assimilative model (Figure 1) that makes predictions of the microbial ecosystem processes based on prognostic, time-evolving ordinary differential equations (Text S2). The data-driven part represents how bacterial modes (Bowman et al. 2017) are compared to optimized model outputs. Our analysis relies on the two types of modes described in Bowman et al. (2017): taxonomic nodes (modes hereafter) were determined from 16S rRNA gene sequence abundance, while functional modes (fmodes) were derived from predicted community metabolic structure. Briefly, sequence reads were categorized into closest estimated genomes and closest completed genomes via the paprica pipeline (Bowman and Ducklow, 2015) and the high dimensional community and metabolic structure data were reduced to 2-D space via a self-organizing map and K-means clustering of map units (Bowman et al. 2017). The final clustering of map units constitutes the modes, and each sample was assigned the mode of its closest map unit. In this approach the mode is a single categorical variable that succinctly describes key structural attributes of the sample. It is important to recognize the categorical nature of these modes, and to understand that – because of the 2-D nature of the map – there is no linear progression among modes. Thus Mode 1, for example, is not necessarily more similar to Mode 3 than it is to Mode 7. Neither mode nor fmodes is necessarily correlated to physiological traits of the bacteria (i.e., modelled HNA- and LNA compartments). In other words, the relative abundance of HNA or LNA, mode, and fmodes are derived from separate observations of different parameters in the same bacterial samples and therefore independent with each other by design.

We select a nearshore Palmer LTER water-column time-series station, Station B (64.77°S, 64.05°W), in the coastal WAP as the modelling site. The Palmer LTER Station B datasets consist of roughly bi-weekly physical, chemical, and biological profiles collected from small boat via a profiling CTD and discrete water samples. Additional observational data are utilized for bacterial flow cytometric (HNA and LNA) and 16S rRNA gene amplicon data collected from Arthur Harbour Station B at 10 m depth (situated 1 km from the Palmer Station B) or Palmer Station seawater intake at 6 m depth (Bowman et al. 2017). A single upper-ocean layer depth (10 m) is modelled for 4 consecutive Palmer LTER growth seasons, including November 2010 - March 2011 (2010-11 hereafter), 2011-12, 2012-13, and 2013-14. Given the availability of the Palmer LTER

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

observations over the Austral spring-summer season, we optimize the model each year separately over the timeframe of available observations. This way, each year possesses its own unique optimized model parameter set, or a model solution. In addition to individual years, we also optimize the model for the climatological year (i.e., climatological model). The climatological year is constructed using four years (2010-11 to 2013-14), rather than the whole Palmer LTER multi-decadal record (since 1991), due to the availability of HNA and LNA biomass data only in those four years. Details on constructing the climatological year and model initialization and spin-up are found in the Supplementary Material (Text S3).

2.3 Data assimilation and parameter optimization

Our model utilizes a variational adjoint data assimilation scheme (Lawson et al. 1995) to minimize the misfit between observations (i.e., assimilated data, see Material and Methods 2.4) and model output by optimizing a subset of model parameters (Friedrichs 2001; Spitz, Moisan, and Abbott 2001; Ward et al. 2010). The data-assimilation scheme (Figure 2) consists of four main steps (Glover, Jenkins, and Doney 2011). First, the model is integrated forward in time from prescribed initial conditions and initial model parameter guess values (Table 1) and calculates the model-observation misfits called total cost function or cost (Material and Methods 2.5). Second, an adjoint model constructed using the Tangent linear and Adjoint Model Compiler (TAPENADE) is integrated backward in time and compute the gradients of the total cost with respect to the model parameters. Third, the computed gradients are passed to a limited-memory quasi-Newton optimization software M1QN3 3.1 (Gilbert and Lemaréchal 1989) to determine the direction and optimal step size by which the model parameters should be modified to reduce the total cost. Finally, a new forward mode simulation is performed using the new set of modified parameters from the third step. These four steps are conducted in an iterative manner until the pre-set convergence criteria are satisfied ensuring the convergence of the optimized parameters and a local minimum achieved by the total cost. The pre-set criteria include the low sensitivity (gradients) of the total cost with respect to each optimized model parameter and positive eigenvalues of the Hessian matrix (Material and Methods 2.6). Every assimilation cycle, we ensure that group-specific bacterial model parameters are optimized in the direction to properly represent the dynamics associated with each group (Table 1), in which we assign different magnitudes of each parameter based on our best guesses and literatures (del Giorgio and Cole 1998; Jiao et al. 2010). For instance, maximum bacterial growth rate of the HNA group (μ_{HNA}, d^{-1}) should be higher than that of the LNA group (μ_{LNA}, d^{-1}), so if μ_{HNA} is optimized to a smaller value than μ_{LNA}, μ_{HNA} is reset back to the original value instead of being updated.

2.4 Assimilated data

We assimilate Palmer LTER observational data from 10 m corresponding to compartments and flows in the model, including nitrate, phosphate, phytoplankton taxonomic specific chlorophyll (Chl) for diatoms and cryptophytes (Schofield et al. 2017), microzooplankton biomass (Garzio et al. 2013), primary production (PP), bulk bacterial production (BP), HNA bacterial biomass, LNA bacterial biomass, semi-labile dissolved organic carbon (SDOC), particulate organic carbon (POC), and particulate organic nitrogen (PON). Climatological observations of Chl (1992-93 to 2009-10) and a single year observation of

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020


Biogeosciences Discussions Open Access EGU

microzooplankton (2010-11) are assimilated to constrain the model parameters because these data are unavailable for some study years. SDOC is calculated by subtracting the background (RDOC) concentration (40.0 mmol m^{-3}) from climatological total DOC concentration. POC (PON) are assimilated to represent the model detrital pool but its measurements also contain living biomass from bottle filter experiments. For example, climatological observations show that living phytoplankton and bacterial biomass account for 74% of total POC and 71% of total PON, so these fractions are used to exclude living biomass from the bulk particulate material pool. When converting Chl to phytoplankton C (N) biomass, the maximum Chl to N ratio is used along with other reference ratios (Tables S2-S6). Due to the discrepancy in the timing and location from our model experiments, the microzooplankton model-observation misfits are not discussed in this study. Krill biomass data are not assimilated due to the strong patchiness of the distribution (many zero values) that would hinder proper model optimization. BP (mmol C m^{-3} d^{-1}) is derived from ^{3}H-leucine incorporation rate (pmol l^{-1} h^{-1}) data using the conversion factor of 1.5 kgC mol^{-1} leucine incorporated (H. W. Ducklow 2000). Group-specific bacterial biomass (mmol C m^{-3}) is estimated from bacterial abundance measured by flow cytometry (i.e., bulk bacterial biomass multiplied by the fraction of each group, fHNA or fLNA, with the conversion factor of 10 fgC $cell^{-1}$) (Fukuda et al. 1998).

2.5 Cost function and portability index

The total cost function or cost (J) is defined as follows to represent a misfit between observations ($\hat{a}_{m,n}$) and model output ($a_{m,n}$) (Luo et al. 2010):

$$J = \sum_{m=1}^{M} \frac{1}{N_m} \sum_{n=1}^{N_m} \left(\frac{a_{m,n} - \hat{a}_{m,n}}{\sigma_m} \right)^2 \quad (5)$$

where m and n represent assimilated data types and data points, respectively, M and N_m are the total number of assimilated data types and data points for data type m, respectively, and σ_m is the target error for data type m. Hereafter, we present the total cost normalized by M ($J' = J/M$) and normalized costs of individual data types (J'_m) throughout this article as the model-observation misfit equivalent to a reduced Chi-square estimate of model goodness of fit, where $J' = 1$ indicates a good fit from optimization, $J' >> 1$ indicates a poor fit due to underestimation of the error variance or the fit not fully capturing the data, and $J' << 1$ indicates an overfitting of the data, fitting the noise, or overestimation of the error variance. The base-10 logarithm of Chl and PP is used in Equation 1 to account for high productivity of the WAP waters and the approximate log-normal distribution of those data types (Campbell 1995; Glover et al. 2018). The target error σ_m is calculated for each data type m as:

$$\sigma_m = \overline{\hat{a}_{m,n}} \cdot CV_m \quad (6)$$

where $\overline{\hat{a}_{m,n}}$ is the climatological mean of the observations and CV_m is the adjusted coefficient of variation (CV) of the observations of each data type from 10 m depth (due to observational error and seasonal and interannual variations). The average CV of each data type at a single depth across the modelled years was 3-7 times higher compared to those across every measured depth within the mixed layer over an extended year period (2002-03 to 2011-12; Kim et al. In Review) and is therefore reduced to the level in the mixed layer to avoid an overestimated target error of each data type. The rationale behind using the adjusted CV in the target error calculation is based on Luo et al (2010), in which all properties in the mixed layer

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

should be completely mixed, a perfect measurement without significant errors should generate similar data values at every measured depth within the mixed layer, and the average CV of all the depth profiles can be used as CV in the target error calculation. The standard deviation is used as target errors of the log-converted data types. The CV of the log-converted data type is estimated as the average of ± 1 standard deviation in log space converted back into normal space (Doney et al. 2003; Glover et al. 2018).

We compute the portability index to evaluate the broader applicability of the optimized model parameter set of each year in predicting dynamics of the other year (Friedrichs et al. 2007):

$$\text{Portability index} = J'_c/J'_x \quad (7)$$

where J'_x is the normalized cross-validation cost when a model parameter set optimized for a given year is used to simulate another year, and J'_c is the normalized total cost of the climatological model. A portability index close to 1 indicates a more portable model, or a system that is not particularly sensitive to year-to-year variations in optimized model parameters, while an index <<1 indicates a less portable model, or a system that is sensitive to year-to-year variations in optimized model parameters.

2.6 Uncertainty analysis

The uncertainties of the optimized parameters are estimated from a finite difference approximation of the complete Hessian matrix during iterative data assimilation processes (i.e., second derivatives of the cost function with respect to the model parameters, see Text S4 for details). When computed at the minimum of the cost function value, the square root of a diagonal element in the inversed Hessian matrix is the logarithm of the relative uncertainty of the corresponding optimized parameter. The absolute uncertainty of the constrained parameter is calculated as $a \cdot e^{\pm \sigma i}$ where a is the value of optimized parameter and σ_i is the relative uncertainty of the corresponding optimized parameter. We then conduct Monte Carlo experiments to calculate the impact of the optimized parameter uncertainties on the model results. The Monte Carlo experiments consist of 1) creating an ensemble of parameter sets (N = 1,000) by randomly sampling values within the uncertainty ranges of the constrained parameters and 2) then performing a model simulation using each parameter set. All uncertainty estimates are calculated following standard error propagation rules and presented herein as ± 1 standard deviation.

3 Results

3.1 Model performance and validation

The iterative, data assimilation-parameter optimization procedure (Figure 2) reduces by 87-91% the misfits between observations and model output across all years compared to the misfit with the initial guess parameters (Table 2). The optimized parameters satisfy the pre-set convergence criteria by reaching local cost function minima with low gradient values with regards to the total costs. The reduction in the total costs is achieved by optimizing a subset of the model parameters (i.e.,

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020


Biogeosciences Discussions Open Access EGU

7-10 parameters with low uncertainties or constrained parameters and 12-15 parameters with large uncertainties or optimized parameters; Tables S2-5). The constrained parameters in common across all years are α_{DIATOM} (initial slope of photosynthesis vs. irradiance curve of diatoms, mol C (g Chl *a*)$^{-1}$ d^{-1} (W m^{-2})$^{-1}$), Θ (maximum Chl:N ratio, g Chl *a* (mol N)$^{-1}$), μ_{LNA} (maximum LNA growth rate, d^{-1}), $r^{A}_{max,\ HNA}$ (maximum HNA active respiration rate, d^{-1}), g_{HNA} (half-saturation density of HNA bacteria in microzooplankton grazing, mmol C m^{-3}), and g_{HNA} (half-saturation density of LNA bacteria in microzooplankton grazing, mmol C m^{-3}).

Model skill is further evaluated using point-to-point comparisons (Figures S1-S5) and Taylor diagrams (Figure 3). The Taylor diagrams highlight that the optimized model has better skill for 2010-11 and 2011-12, with more data types exhibiting relatively high correlation coefficients and low centred (bias removed) root-mean-square difference (RMSD). Three core variables of interest in this study, including HNA biomass, LNA biomass, and BP, show overall good model-observation agreements with relatively high correlations and low RMSD.

Cross-validation cost analyses show the overall model-observation misfit increases as expected when a set of parameters optimized for one year is used to simulate another year's dynamics, indicating that each growth season is best modelled using its own unique set of optimized parameters (Table 3). The magnitude of the cost function increase varies by year pair, ranging from 0.27 to 0.81. The optimized model parameters for 2012-13 and 2013-14 are most portable, followed by the model parameters of 2010-11 and 2011-12.

3.2 Bacterial carbon stocks and flows

Seasonal progression of the C stocks and flows for each bacterial group shows significant seasonal variability (Figure S6) and interannual variability (Figure 4A). Average C stocks and flows in the food-web are calculated over the growth season at 10-m depth for each year (Figure 5) and normalized by NPP (normalized by NPP in 1-day for C stocks) (Figure S7). The mean HNA biomass is 6 ± 4 times (mean ± propagated standard deviation from the Monte Carlo sensitivity experiments and season-averaging; Figure 4B) larger than the mean LNA biomass in 2012-13, while the mean LNA biomass is 9 ± 8 times larger than the mean HNA biomass in 2011-12. Bacterial carbon demand (BCD) is equivalent to total carbon flows from LDOC and SDOC pools to bacteria (BCD = BP + bacterial respiration; Figure 5) and mostly supported by the LDOC pool (77 ± 5% to 97 ± 4%) for both bacterial populations across years. Partitioning of the BP (BP = BCD - bacterial respiration; Figure 5) belonging to each bacterial group is determined during optimization. The mean HNA BP ranged from 0.01 ± 0.05 to 0.23 ± 0.25 mmol C m^{-3} d^{-1}, while the mean LNA BP ranged from 0.01 ± 0.02 to 0.08 ± 0.07 mmol C m^{-3} d^{-1}. The mean cell-specific BP (BP divided by bacterial biomass or per-cell BP, d^{-1}) is higher for the HNA group, ranging from 0.17 ± 0.61 to 0.40 ± 1.69 d^{-1} for HNA cells and 0.04 ± 0.05 to 0.16 ± 0.20 d^{-1} for LNA cells. The mean per-cell respiration rate (d^{-1}) is also higher for the HNA group compared to the LNA group. The mean cell-specific SDOC uptake rate (d^{-1}) is higher for the HNA group, while the mean cell-specific LDOC uptake rate (d^{-1}) is higher for the LNA group. The mean bacterial growth efficiency, BGE (BP

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

250 divided by BCD) of the HNA group is lower than that of the LNA group, ranging from 0.19 ± 0.82 to 0.38 ± 0.38 for HNA cells and from 0.52 ± 1.18 to 0.62 ± 0.62 for LNA cells.

In the model, the time rate of change of the bacterial biomass is determined by the difference between the production term (i.e., total DOC uptake or BCD) and loss terms (i.e., respiration, grazing, viral mortality, and RDOC excretion). In other words, the fate of BCD is formulated as follows:

255 $$\text{BCD} = \frac{dB_{HNA}}{dt} + \text{respiration} + \text{grazing} + \text{viral mortality} + \text{RDOC excretion} \quad (8)$$

or

$$1 = (\frac{dB_{HNA}}{dt})/\text{BCD} + \text{respiration/BCD} + \text{top-down control/BCD} + \text{RDOC excretion/BCD} \quad (9)$$

where B is bacterial biomass, the sum of grazing and viral mortality is equivalent to top-down control, and four different bacterial carbon utilization efficiencies sum to 1, including the efficiencies of biomass synthesis, respiration, top-down control
260 (grazing and viral mortality), and RDOC excretion. The largest loss term with regards to the fate of BCD is respiration for HNA cells, accounting for 62 ± 7.7% to 81 ± 4.0% of the BCD (i.e., the mean respiration efficiency of 0.6 ± 0.08 to 0.8 ± 0.04 across years). Top-down control is a relatively large loss term for LNA cells, accounting for 35 ± 0.30% to 60 ± 2.5% (i.e., the mean top-down control efficiency of 0.4 ± 0.003 to 0.6 ± 0.03 across years), equally large or larger than top-down control on HNA cells (i.e., the mean top-down control efficiency of 0.1 ± 0.004 to 0.4 ± 0.03 across years).

265 The rest of the ecosystem state variables (stocks and flows) fall into one of three categories: 1) the variables for which climatological or a single year values are assimilated (diatom-specific Chl, cryptophyte-specific Chl, and microzooplankton biomass); 2) the variables for which observational values for the given year are assimilated (nutrients, POC or detritus, and SDOC), and 3) the variables that are not directly assimilated (krill, LDOC, carbon export flux). There is a little interannual variability in the average C stocks of the first category except relatively large microzooplankton biomass in 2011-12 (Figure
270 5B) and cryptophyte biomass in 2012-13 (Figure 5C). Although nitrate is not assimilated in 2010-11 and detritus and SDOC are not assimilated in 2012-13 and 2013-14, their simulated values in those unassimilated years are comparable to other assimilated years.

3.3 Bacterial physiological and taxonomic association with ecosystem functions

A property map of the emergent self-organizing map nodes (as generated by Bowman et al. 2017) shows the mode association
275 with community structure (Figure 6). The coloured map units (the circles in the background) are clustered into taxonomic mode membership or modes (Figure 6a), showing a different frequency of appearance year-to-year (Figure 6b). Each mode is dominated by unique bacterial taxa. For example, *Candidatus* Pelagibacter is most abundant in Mode 6 (Figure 6c), *Dokdonia* sp. MED134 in Mode 7 (Figure 6d), *Candidatus* Thioglobus singularis PS1 in Mode 1 (Figure 6e), *Owenweeksia*

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020


Biogeosciences Discussions

hongkongensis DSM 17368 in Mode 2 (Figure 6f), Rhodobacteraceae in Mode 5 (Figure 6g), and *Planktomarina temperata* RCA23 in Mode 4 (Figure 6h).

To explore a potential link between mode and the key model ecosystem functions, we first extract the modelled net primary production (NPP), POC export, and BCD from our ecosystem model at the time of bacterial samples that are placed into a single mode (observed). We then perform a linear regression with mode as a factor (i.e., mode as a categorical predictor with 8 modes rather than an ordinal or continuous variable; equivalent to a one-way ANOVA with 8 different categories). fmode does not show a significant relationship with any of the model ecosystem functions examined (all $p > 0.05$; not shown). By contrast, 19%, 52%, and 66% of the total variance in the modelled NPP, C export, and BCD is explained by mode (Figures 7a-c). Mode 4 is associated with low NPP and low POC export, while Modes 3, 5, and 7 are associated with high NPP and high POC export. Mode 4 is associated with low BCD, while Mode 7 is associated with high BCD. Mode 6 occurs during relatively high NPP but low POC export. Mode is positively correlated to fHNA ($r^2 = 0.52$, $p < 0.001$; not shown). Similarly, we examine a potential link between fHNA and the key model ecosystem functions as described above (i.e., linear regression where an observed fHNA is a predictor and the modelled ecosystem functions are dependent variables). fHNA is positively correlated to NPP ($r^2 = 0.12$, $p = 0.01$; Figure 7d) and POC export to a moderate extent ($r^2 = 0.28$, $p < 0.001$; Figure 7e) and to BCD to a strong extent ($r^2 = 0.45$, $p < 0.001$; Figure 7f). The stepwise addition of one predictor variable to the other predictor variable (i.e., fHNA adding to mode or vice versa) does not improve the model performance (not shown). These results suggest a clear link between modelled ecosystem functions and observed bacterial molecular data (i.e., modes).

3.4 Climate change experiments

We explore the response of the modelled bacterial dynamics (Results 3.2-3.3) to changing climate along the WAP (Figure 8). Due to a varying range of portability of the optimized model solution for each year, we perform this experiment with the climatological model parameter set (Table S6) to examine an overall system response under perturbed ocean temperature (+1°C and +2°C relative to observed temperatures) and sea-ice forcing fields (10% and 20% loss of sea-ice relative to observed sea-ice concentrations). The experiments are conducted under each condition separately (i.e., warming or sea-ice melting alone scenario; Figures S7-8) and simultaneously (i.e., climate change scenario; Figure 7), but the results from only the climate change scenario are discussed below, as despite different impacts of each physical forcing changes (i.e., temperature affects a variety of rate processes, while sea-ice concentration affects light levels and photosynthesis but not the mixed layer depth) climate change would cause simultaneous changes in sea ice and water temperature along the WAP.

The climate change experiments exhibit a combination of changes in overall ecosystem stocks and rates integrated over the growth season and shifts in the seasonal timing or phenology (Figure 8A), compared to the base state (first row as the base state and second and third rows as anomalies in Figure 8B). Compared to other stocks and flows, HNA bacterial rates respond most to the perturbed climate conditions. In response to 2°C warming and 20% less sea ice, the maximum values increase for LDOC uptake (69 ± 2.2%; the maximum value ± standard deviation from Monte Carlo errors; results of sensitivity

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

experiments not shown), SDOC uptake (81 ± 1.9%), RDOC excretion (48 ± 1.6%), grazing (97 ± 5.0%), viral mortality (48 ± 1.6%), and respiration (69 ± 1.7%) of HNA cells, and for LDOC uptake (49 ± 3.4%), SDOC uptake (67 ± 3.9%), RDOC excretion (33 ± 2.6%), grazing (63 ± 3.9%), viral mortality (33 ± 2.6%), and respiration (49 ± 3.4%) of LNA cells. The maximum values also increased for HNA biomass (48 ± 1.6%) and LNA biomass (33 ± 2.6%). In response to 2°C warming and 20% less sea ice, the events of high NPP, POC export flux, and diatom-, microzooplankton-, and krill accumulations shift to earlier in the season, with the maximum values of 50 ± 0.10%, 50 ± 0.20%, 20 ± 0.10%, and 25 ± 0.20%, and 38 ± 0.48%, respectively. More SDOC accumulates in the early spring (the maximum increase of 10 ± 0.10%) as a result of increased plankton accumulations and SDOC excretion during the same period. By contrast, LDOC becomes strongly limited throughout the growth season (the maximum decrease of 24 ± 1.0%).

4 Discussion

4.1 Model performance and validation

In our study, only a subset of the model parameters is optimized to best simulate bacterial and other ecological patterns for each year, in accordance with other data-assimilative ecosystem model studies (Friedrichs 2001; Friedrichs et al. 2007; Friedrichs, Hood, and Wiggert 2006; Luo et al. 2010, 2012). In general, optimization of this class of marine ecosystem models requires adjustment of only a small number of independent model parameters to achieve well-posed model solutions, due to the highly cross-correlated nature of parameters in the inherently nonlinear model equations (Fennel et al. 2001; Harmon and Challenor 1997; Matear 1996; Prunet, Minster, Echevin, et al. 1996; Prunet, Minster, Ruiz-Pino, et al. 1996). Most of the constrained parameters in our study are directly associated with bacterial processes, with overall better model-observation fits for bacterial data types, giving confidence in the simulated bacterial C stocks and flows. Despite the important biogeochemical role that heterotrophic marine bacteria play in the ocean, the vast majority of marine ecosystem models neither include bacteria as a model compartment nor explicitly simulate bacterial processes. Most existing models parameterize the complex bacterial remineralization processes of the (sinking) organic matter with depth as a function of POC concentration and temperature, or by fitting with power law functions. Cellular functions, taxa, and functional gene expression of other prokaryotes, such as cyanobacteria (Hellweger 2010; Martín-Figueroa, Navarro, and Florencio 2000; Miller et al. 2013), or a diverse suite of microbial functional groups (Coles et al. 2017; Dutkiewicz et al. 2020) have been modelled so far; however, our study is the first to explicitly model bacterial groups of different physiological traits.

Here, optimization sheds light on major unknown parameters associated with the bacterial grazing process, including g_{HNA} and g_{LNA} (half-saturation density of HNA and LNA bacteria in microzooplankton grazing, respectively; Text S1; Equation S11). Microzooplankton grazing of a given bacterial group in the model is simulated using Holling Type 2 density-dependent grazing with a preferential prey selection on diatoms, cryptophytes, and the other bacterial group, in which the single microzooplankton maximum grazing rate is implemented for both bacterial groups for model simplicity purposes (Text S1;

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Equation S11). Thus, it is the half-saturation density that determined the degree of preferential grazing by microzooplankton on a certain bacterial group, the change of which would ultimately depend on biomass of each group. Due to the lack of a priori knowledge on the relative magnitude of g_{HNA} and g_{LNA}, we assign the identical initial guess value (Table 1) to let the data assimilation scheme find the optimized values that best fit overall observations. Compared to g_{LNA} values, smaller optimized g_{HNA} values across all years (Tables S2-6) indicate preferential grazing of HNA cells by microzooplankton, consistent with previous speculations that grazers selectively remove larger and more active cells (Giorgio et al. 1996; Gonzalez, Sherr, and Sherr 1990; Sherr, Sherr, and McDaniel 1992), or HNA cells (Garzio et al. 2013). Despite respiration being the largest sink for the BCD of HNA cells, the absolute amount of grazing and its values normalized by group-specific biomass are modelled to be still greater for HNA cells, compared to LNA cells (Figure 4), supporting preferential grazing of HNA cells by microzooplankton.

The model portability index reflects the extent to which a single model framework (i.e., parameters and equations) captured the observed variability in different years, given variable environmental forcing and the accompanying shift in plankton ecosystem structure. The varying range of the portability index values renders it difficult to pick one particular year's model solution that represents the climatological dynamics in other years. In other words, better model skill might be found by simply utilizing parameters from assimilating climatological data into a more general version of the model (see Discussion 4.4). These varying portability characteristics of the model in each Palmer LTER sampling year are consistent with those of the original WAP data-assimilative model forced by two contrasting sea-ice conditions (Kim et al. In Review).

4.2 Bacterial carbon stocks and flows

Our model results show significantly higher per-cell (or per C biomass) HNA BP across all years compared to the LNA group. The higher per-cell rates of HNA cells in the model is largely due to the way the parameter optimization is performed to keep higher maximum cell-specific growth rates of HNA cells compared to those of LNA cells, but these per-cell rates are a function of both bulk rates and biomass stocks that are constrained from measurements and determined from trophic interactions in the model. As with phylogenetic groups (Fuchs et al. 2000; Teira et al. 2009; Yokokawa et al. 2004), cell-specific growth rates (equivalent to per-cell BP in our study) are expected to differ among distinct bacterial physiological groups, but there are limited studies that focused on group-specific cell activities or growth rates (Gasol et al. 1999; Giorgio et al. 1996; GÜnter et al. 2008; Longnecker, Sherr, and Sherr 2005; Moràn, Ducklow, and Erickson 2011). Moran et al (2011) showed that HNA cells greatly outgrew LNA cells in Waquoit Bay Estuary, Massachusetts, with a cell-specific growth rate of up to 2.26 d^{-1}, compared to relatively slow growth of LNA cells (< 0.5 d^{-1}). High cell-specific HNA BP in our model could be attributed to much larger total SDOC uptake of HNA cells than LNA cells, despite higher HNA respiration. We note that HNA cells also have significantly higher SDOC uptake rates per unit C biomass compared to LNA cells. Several studies demonstrate that HNA cells depend more than LNA cells on phytoplankton substrates for growth and metabolism (Li, Jellett, and Dickie 1995; Morán et al. 2007; Scharek and Latasa 2007), but our study is the first to show the importance of the SDOC pool for the carbon demand of HNA bacteria. The hypothesis that WAP bacteria might rely on SDOC has received indirect support previously,

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

375 presumably due to LDOC limitation (Hugh W. Ducklow et al. 2011; Kim and Ducklow 2016; Luria et al. 2017). HNA cells are also modelled to have relatively high per-cell respiration and respiration efficiency with regard to the fate of their carbon demand, while carbon demand of LNA cells is more likely lost to top-down control factors followed by respiration. The high HNA bacterial respiration drives an overall lower BGE of the HNA cells, despite their high BCD.

Although much of the discussion focuses on bacteria, our model also captures well the rest of the ecosystem variables. 380 The WAP typically exhibits strong interannual variability in phytoplankton and zooplankton biomass accumulations (Ducklow et al. 2007), but the lack of their strong interannual variability in the model is due to assimilating climatological observations of diatom- and cryptophyte-specific Chl and a different year's observations of microzooplankton biomass. Krill are not assimilated, but predicted as reasonable simulated values (0.13 ± 0.03 to 0.40 ± 0.18 mmol C m^{-3}) compared to the available field data (the mean krill biomass = 0.12 ± 0.03 mmol C m^{-3} and the maximum krill biomass = 0.57 mmol C m^{-3} in 2017-18 at 385 Palmer Station; data provided by D. Steinberg).

4.3 Bacterial physiological and taxonomic association with ecosystem functions

In Bowman et al. (2017) fHNA and mode alone accounted for 36% and 46% of the variance in bulk BP, respectively, and together accounted for up to 51% of the variance in bulk BP. In our study, Modes 3, 5, and 7, characterized by copiotrophic taxa with large genomes and more 16S rRNA gene copies (Bowman et al., 2017), are associated with high NPP, POC export, 390 and BCD, while Modes 4 and 6, characterized by taxa associated with more oligotrophic conditions, are associated with low NPP, POC export, and BCD. *Dokdonia* sp. MED134, a common bacterial species of the modes associated with high NPP, POC export, and BCD, is a proteorhodopsin-containing marine flavobacterium shown to grow faster with light (Gómez-Consarnau et al. 2007; Kimura et al. 2011) and in conditions under which resources are abundant (Gómez-Consarnau et al. 2007). Given that the coastal WAP is frequently light-limited (Ducklow et al. 2012), association of *D. Dokdonia* MED134 395 with high NPP suggests sufficient irradiance and therefore light-enhanced growth rates and cell yields of this species. Mode 4, by contrast, is dominated by *Planktomarina temperata* RCA23, a slowly growing bacterium that specializes in using complex organic substrates (Giebel et al. 2013). These attributes are consistent with this species' high occurrence during low NPP and low POC in the model. *Candidatus* Pelagibacter, abundant in Mode 6, is generally known as an oligotrophic specialist with a low DOC requirement, but is often observed during the Antarctic phytoplankton blooms (Delmont et al. 2014; Luria, 400 Ducklow, and Amaral-Zettler 2014), consistent with its occurrence during high NPP periods in our model results.

The bacterial mode association with BCD, NPP, and particle export suggest that bacteria in the coastal WAP are controlled by resource availability. The modes predicted by high NPP and high POC export were also characterized by high BP (Bowman et al. 2017). It has been hypothesized that due to minimal inputs of terrestrial organic matter, bacteria in the WAP must ultimately rely on *in situ* NPP for organic matter source (Ducklow et al. 2012b). Indeed, our results show that 405 compared to zooplankton excretion and detrital dissolution, phytoplankton excretion accounts for a major fraction of DOC production. Similarly, phytoplankton account for a major fraction of detritus production, but particle export in the model is also modelled to be proportional to optimizable particle sinking speed that is most likely linked to the difference in the time-

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

rate change of POC observations. This explains the occasional discrepancy between NPP and POC export and why Mode 6 shows a different relationship with NPP and POC export. In summary, our study provides a novel framework for the prediction of different ecosystem functions using microbial taxonomy. Certain modes represent distinct ecosystem states in the WAP and such mode-state associations are reasonably explained from microbial perspectives. However, we do not address a seasonal succession and development in mode itself or the mode predictability of the key WAP ecosystem states. Future investigations should focus on directly assimilating a few dominant or seasonally distinct modes in the ecosystem model, to fully resolve seasonality of the mode-state associations along the WAP.

4.4 Climate change scenarios

The WAP has experienced significant atmospheric and ocean warming and resulting changes in marine ecosystem processes and further climate change is projected for the next several decades. Based on our baseline ecosystem model simulations (Results 3.2-3.3), we investigate how future climate scenarios, with decreasing sea ice and warming temperature together, might affect bacterial dynamics and ecosystem functions. Under changing climate conditions, it is expected that increased NPP and phytoplankton- and zooplankton accumulations early in the season would result in a significant build-up of DOC pools. This is the case only for SDOC, but bacteria are soon strongly LDOC-limited due to their preferential and increased LDOC uptake for their primary carbon source. The growth of bacteria and increased bacterial production rates during strong LDOC limitation indicate that bacteria depend on SDOC to satisfy the rest of their carbon demand, resulting in the depletion of SDOC pool as the season progresses. In other words, bacteria are more likely resource-limited, especially labile pool of the organic matter, and SDOC subsequently plays an increasingly important role as a part of the bacterial carbon demand. This change is particularly important in HNA cells, as shown by a relatively large increase of BCD via SDOC pool, compared to LNA cells. Despite all four loss terms with regard to the fate of the BCD (i.e., respiration, grazing, viral mortality, and RDOC excretion) increase for both groups under perturbed conditions, overall high BCD leads to increased biomass of both bacterial groups. Temperature is a major factor regulating bacterial biomass, production, and growth rates by changing the rate of enzymatic reactions (Kirchman, Morán, and Ducklow 2009; White et al. 1991). In our case, stocks and rates of both HNA- and LNA cells somewhat increase under a temperature warming alone scenario (Figure S8) but much more so under a sea-ice melting alone scenario (i.e., increased photosynthesis and resource availability). This is evident for all variables during high NPP and plankton accumulation events early in the spring (Figure S9), except for SDOC uptake of both bacterial groups due to overall increased resource availability. This suggests that temperature *per se* is not a more important limiting factor for bacterial growth than resource availability (Ducklow et al. 2012a), and warming temperature would rather enhance bacterial utilization of the already increased organic matter from increased phytoplankton activity. Our findings also suggest that future climate conditions along the WAP would further impact the distribution of taxonomic groups with a potential shift to more abundant HNA cells in the WAP bacterial communities due to their preferential SDOC utilization. This shift would cooccur with increased NPP and POC export, reinforcing the "high NPP-high particle export-more abundant HNA bacteria" link shown in the optimized model results for each year. The scenario of "more abundant HNA bacterial cells in more productive waters"

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

implies a relatively strong resource control on these actively growing cells. This is consistent with previous studies that show increased HNA growth rates in response to enhanced phytoplankton-derived organic substrate (Moran et al. 2010) and larger abundance of HNA cells in areas or periods where bacterial assemblages were predominantly controlled by resources, rather than grazing (Morán et al. 2007).

5 Conclusions

Heterotrophic microbial diversity has seldom been considered in detail in the formulation and analysis of marine pelagic ecosystem models reflecting in part the lack of suitable field data for model evaluation. Utilizing genomic products to prescribe taxonomic aspects of the bacterial model dynamics, our study investigates the association of bacterial abundance with different physiological states, bacterial community structure and key ecosystem functions. Our modelling approach enables the observations in different bacterial populations to constrain the group-specific processes and model parameters that have been poorly understood. These include the partitioning of BP specific to HNA and LNA groups, the partitioning of the bacterial uptake of DOC pools with different lability, and the half-saturation density of each bacterial group in microzooplankton grazing. The model is an effective platform to explore the WAP microbial response to changing climate conditions, in which warming and decreasing sea ice would induce a potential shift to the dominance of HNA bacteria in more productive waters due to their increasing dependence on SDOC.

Code availability

The model simulation results and codes are available in a NetCDF data structure and Fortran 90 at the GitHub data repository (https://github.com/hyewon-kim-whoi).

Data availability

Complete Palmer LTER time-series data used for data assimilation are available online (http://pal.lternet.edu/data). Surface downward solar radiation flux data used for physical forcing of the model simulations can be found in the National Centers for Environmental Prediction website (https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surface.html). The Tangent linear and Adjoint Model Compiler (TAPENADE) used to construct an adjoint model is available online (http://www-sop.inria.fr/tropics/).

Author contribution

HK designed the research, performed model simulations, and wrote the manuscript. JB provided observational data and helped data analyses and interpretation. HD, OS, and DS provided observational data. YWL contributed model simulations. SD supervised the research and significantly revised the manuscript.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Competing interest

The authors declare that they have no conflict of interest.

475

Acknowledgements

This study leverages the wealth of marine data collected by the Palmer Long-Term Ecological Research (Palmer LTER) program along the west Antarctic Peninsula, and the authors thank the scientists, students, technicians, station support and logistical staff, and ship captains, officers and crew involved. This research was supported, in part, by the U.S. National Science
480 Foundation Office of Polar Programs through award NSF PLR-1440435 and the U.S. National Aeronautics and Space Administration Ocean Biology and Biogeochemistry Program through award NASA NNX14AL86G. HK was also supported by the Investment in Science Fund and the Reuben F. and Elizabeth B. Richards Endowed Fund from Woods Hole Oceanographic Institution.

References

485 Azam, F. et al. 1983. "The Ecological Role of Water-Column Microbes in the Sea." *Marine Ecology Progress Series* 10(3): 257–63.

Bouvier, Thierry, Paul A. Del Giorgio, and Josep M. Gasol. 2007. "A Comparative Study of the Cytometric Characteristics of High and Low Nucleic-Acid Bacterioplankton Cells from Different Aquatic Ecosystems." *Environmental Microbiology* 9(8): 2050–66.

490 Bowman, Jeff S et al. 2017. "Bacterial Community Segmentation Facilitates the Prediction of Ecosystem Function along the Coast of the Western Antarctic Peninsula." *The ISME Journal* 11(6): 1460–71.

Bowman, Jeff S., and Hugh W. Ducklow. 2015. "Microbial Communities Can Be Described by Metabolic Structure: A General Framework and Application to a Seasonally Variable, Depth-Stratified Microbial Community from the Coastal West Antarctic Peninsula." *PLOS ONE* 10(8): e0135868.

495 Calvo-Díaz, Alejandra, and Xosé Anxelu G. Morán. 2006. "Seasonal Dynamics of Picoplankton in Shelf Waters of the Southern Bay of Biscay." *Aquatic Microbial Ecology* 42(2): 159–74.

Campbell, Janet W. 1995. "The Lognormal Distribution as a Model for Bio-Optical Variability in the Sea." *Journal of Geophysical Research: Oceans* 100(C7): 13237–54.

Clarke, Andrew et al. 2009. "Spatial Variation in Seabed Temperatures in the Southern Ocean: Implications for Benthic
500 Ecology and Biogeography." *Journal of Geophysical Research: Biogeosciences* 114(G3). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JG000886 (March 9, 2020).

Coles, V. J. et al. 2017. "Ocean Biogeochemistry Modeled with Emergent Trait-Based Genomics." *Science* 358(6367): 1149–54.

Cook, A. J., A. J. Fox, D. G. Vaughan, and J. G. Ferrigno. 2005. "Retreating Glacier Fronts on the Antarctic Peninsula over
505 the Past Half-Century." *Science* 308(5721): 541–44.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Delmont, Tom O. et al. 2014. "Phaeocystis Antarctica Blooms Strongly Influence Bacterial Community Structures in the Amundsen Sea Polynya." *Frontiers in Microbiology* 5: 646.

Doney, Scott C., David M. Glover, Scott J. McCue, and Montserrat Fuentes. 2003. "Mesoscale Variability of Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) Satellite Ocean Color: Global Patterns and Spatial Scales." *Journal of Geophysical Research: Oceans* 108(C2). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JC000843 (July 7, 2020).

Ducklow, H. W. 2000. "Bacterial Production and Biomass in the Ocean." In *Microbial Ecology of the Oceans, Second Edition*, John Wiley & Sons, Inc, 85–120.

Ducklow, Hugh et al. 2012. "The Marine System of the Western Antarctic Peninsula." In *Antarctic Ecosystems*, John Wiley & Sons, Ltd, 121–59. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444347241.ch5 (May 20, 2020).

Ducklow, Hugh W et al. 2007. "Marine Pelagic Ecosystems: The West Antarctic Peninsula." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1477): 67–94.

Ducklow, Hugh W. et al. 2011. "Response of a Summertime Antarctic Marine -bacterial Community to Glucose and Ammonium Enrichment." http://agris.fao.org/agris-search/search.do?recordID=AV2012072112 (March 9, 2020).

———. 2012a. "Multiscale Control of Bacterial Production by Phytoplankton Dynamics and Sea Ice along the Western Antarctic Peninsula: A Regional and Decadal Investigation." *Journal of Marine Systems* 98–99: 26–39.

———. 2012b. "Multiscale Control of Bacterial Production by Phytoplankton Dynamics and Sea Ice along the Western Antarctic Peninsula: A Regional and Decadal Investigation." *Journal of Marine Systems* 98–99: 26–39.

Dutkiewicz, Stephanie et al. 2020. "Dimensions of Marine Phytoplankton Diversity." *Biogeosciences* 17(3): 609–34.

Feist, Adam M. et al. 2009. "Reconstruction of Biochemical Networks in Microorganisms." *Nature Reviews. Microbiology* 7(2): 129–43.

Fennel, Katja, Martin Losch, Jens Schröter, and Manfred Wenzel. 2001. "Testing a Marine Ecosystem Model: Sensitivity Analysis and Parameter Optimization." *Journal of Marine Systems* 28(1): 45–63.

Friedrichs, M. A. M. 2001. "Assimilation of JGOFS EqPac and SeaWiFS Data into a Marine Ecosystem Model of the Central Equatorial Pacific Ocean." *Deep Sea Research Part II: Topical Studies in Oceanography* 49(1): 289–319.

Friedrichs, M. A. M. et al. 2007. "Assessment of Skill and Portability in Regional Marine Biogeochemical Models: Role of Multiple Planktonic Groups." *Journal of Geophysical Research: Oceans* 112(C8).

Friedrichs, M. A. M., R. R. Hood, and J. D. Wiggert. 2006. "Ecosystem Model Complexity versus Physical Forcing: Quantification of Their Relative Impact with Assimilated Arabian Sea Data." *Deep Sea Research Part II: Topical Studies in Oceanography* 53(5): 576–600.

Fuchs, Bernhard M. et al. 2000. "Changes in Community Composition during Dilution Cultures of Marine Bacterioplankton as Assessed by Flow Cytometric and Molecular Biological Techniques." *Environmental Microbiology* 2(2): 191–201.

Fukuda, Rumi, Hiroshi Ogawa, Toshi Nagata, and Isao Koike. 1998. "Direct Determination of Carbon and Nitrogen Contents of Natural Bacterial Assemblages in Marine Environments." *Applied and Environmental Microbiology* 64(9): 3352–58.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Garzio, Lori M., Deborah K. Steinberg, Matthew Erickson, and Hugh W. Ducklow. 2013. "Microzooplankton Grazing along the Western Antarctic Peninsula." https://darchive.mblwhoilibrary.org/handle/1912/6317 (March 9, 2020).

Garzio, Lori, and Deborah Steinberg. 2013. "Microzooplankton Community Composition along the Western Antarctic Peninsula." *Deep Sea Research Part I: Oceanographic Research Papers* 77: 36–49.

Gasol, Josep M. et al. 1999. "Significance of Size and Nucleic Acid Content Heterogeneity as Measured by Flow Cytometry in Natural Planktonic Bacteria." *Applied and Environmental Microbiology* 65(10): 4475–83.

Giebel, Helge-Ansgar et al. 2013. "Planktomarina Temperata Gen. Nov., Sp. Nov., Belonging to the Globally Distributed RCA Cluster of the Marine Roseobacter Clade, Isolated from the German Wadden Sea." *International Journal of Systematic and Evolutionary Microbiology* 63(Pt 11): 4207–17.

Gilbert, Jean Charles, and Claude Lemaréchal. 1989. "Some Numerical Experiments with Variable-Storage Quasi-Newton Algorithms." *Mathematical Programming* 45(1): 407–35.

Giorgio, Paul A. del et al. 1996. "Bacterioplankton Community Structure: Protists Control Net Production and the Proportion of Active Bacteria in a Coastal Marine Community." *Limnology and Oceanography* 41(6): 1169–79.

del Giorgio, Paul A., and Jonathan J. Cole. 1998. "Bacterial Growth Efficiency in Natural Aquatic Systems." *Annual Review of Ecology and Systematics* 29(1): 503–41.

Glover, David M., Scott C. Doney, William K. Oestreich, and Alisdair W. Tullo. 2018. "Geostatistical Analysis of Mesoscale Spatial Variability and Error in SeaWiFS and MODIS/Aqua Global Ocean Color Data." *Journal of Geophysical Research: Oceans* 123(1): 22–39.

Glover, David M., William J. Jenkins, and Scott C Doney. 2011. "10. Model Analysis and Optimization." In *Modeling Methods for Marine Science*, Cambridge University Press.

Gómez-Consarnau, Laura et al. 2007. "Light Stimulates Growth of Proteorhodopsin-Containing Marine Flavobacteria." *Nature* 445(7124): 210–13.

Gonzalez, J. M., E. B. Sherr, and B. F. Sherr. 1990. "Size-Selective Grazing on Bacteria by Natural Assemblages of Estuarine Flagellates and Ciliates." *Applied and Environmental Microbiology* 56(3): 583–89.

GÜnter, Jost et al. 2008. "High Abundance and Dark CO2 Fixation of Chemolithoautotrophic Prokaryotes in Anoxic Waters of the Baltic Sea." *Limnology and Oceanography* 53(1): 14–22.

Harmon, Robin, and Peter Challenor. 1997. "A Markov Chain Monte Carlo Method for Estimation and Assimilation into Models." *Ecological Modelling* 101(1): 41–59.

Hellweger, Ferdi L. 2010. "Resonating Circadian Clocks Enhance Fitness in Cyanobacteria in Silico." *Ecological Modelling* 221(12): 1620–29.

———. 2020. "Combining Molecular Observations and Microbial Ecosystem Modeling: A Practical Guide." *Annual Review of Marine Science* 12(1): 267–89.

Jiao, Nianzhi et al. 2010. "Microbial Production of Recalcitrant Dissolved Organic Matter: Long-Term Carbon Storage in the Global Ocean." *Nature Reviews Microbiology* 8(8): 593–99.

575 Kim, Hyewon et al. 2016. "Climate Forcing for Dynamics of Dissolved Inorganic Nutrients at Palmer Station, Antarctica: An Interdecadal (1993–2013) Analysis." *Journal of Geophysical Research: Biogeosciences* 121(9): 2369–89.

Kim, Hyewon, and Hugh W. Ducklow. 2016. "A Decadal (2002–2014) Analysis for Dynamics of Heterotrophic Bacteria in an Antarctic Coastal Ecosystem: Variability and Physical and Biogeochemical Forcings." *Frontiers in Marine Science* 3. https://www.frontiersin.org/articles/10.3389/fmars.2016.00214/full (March 2, 2020).

580 Kimura, Hiroyuki, Curtis R. Young, Asuncion Martinez, and Edward F. DeLong. 2011. "Light-Induced Transcriptional Responses Associated with Proteorhodopsin-Enhanced Growth in a Marine Flavobacterium." *The ISME Journal* 5(10): 1641–51.

King, J. C. 1994. "Recent Climate Variability in the Vicinity of the Antarctic Peninsula." *International Journal of Climatology* 14(4): 357–69.

585 Kirchman, David L., Xosé Anxelu G. Morán, and Hugh Ducklow. 2009. "Microbial Growth in the Polar Oceans - Role of Temperature and Potential Impact of Climate Change." *Nature Reviews. Microbiology* 7(6): 451–59.

Lawson, Linda M., Yvette H. Spitz, Eileen E. Hofmann, and Robert Bryan Long. 1995. "A Data Assimilation Technique Applied to a Predator-Prey Model." *Bulletin of Mathematical Biology* 57(4): 593–617.

Li, W. K. W., J. F. Jellett, and P. M. Dickie. 1995. "DNA Distributions in Planktonic Bacteria Stained with TOTO or TO-
590 PRO." *Limnology and Oceanography* 40(8): 1485–95.

Longnecker, K., B. F. Sherr, and E. B. Sherr. 2005. "Activity and Phylogenetic Diversity of Bacterial Cells with High and Low Nucleic Acid Content and Electron Transport System Activity in an Upwelling Ecosystem." *Applied and Environmental Microbiology* 71(12): 7737–49.

Luo, Ya-Wei et al. 2010. "Oceanic Heterotrophic Bacterial Nutrition by Semilabile DOM as Revealed by Data Assimilative
595 Modeling." *Aquatic Microbial Ecology* 60(3): 273–87.

———. 2012. "Interannual Variability of Primary Production and Dissolved Organic Nitrogen Storage in the North Pacific Subtropical Gyre." *Journal of Geophysical Research: Biogeosciences* 117(G3).

Luria, Catherine M. et al. 2017. "Seasonal Shifts in Bacterial Community Responses to Phytoplankton-Derived Dissolved Organic Matter in the Western Antarctic Peninsula." *Frontiers in Microbiology* 8.
600 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5675858/ (March 9, 2020).

Luria, Catherine M., Hugh W. Ducklow, and Linda A. Amaral-Zettler. 2014. "Marine Bacterial, Archaeal and Eukaryotic Diversity and Community Structure on the Continental Shelf of the Western Antarctic Peninsula." https://darchive.mblwhoilibrary.org/handle/1912/6966 (March 9, 2020).

Martín-Figueroa, E., F. Navarro, and F. J. Florencio. 2000. "The GS-GOGAT Pathway Is Not Operative in the Heterocysts.
605 Cloning and Expression of GlsF Gene from the Cyanobacterium Anabaena Sp. PCC 7120." *FEBS letters* 476(3): 282–86.

Matear, Richard J. 1996. "Parameter Optimization and Analysis of Ecosystem Models Using Simulated Annealing: A Case Study at Station P." *Oceanographic Literature Review* 43(6). https://www.ingentaconnect.com/content/jmr/jmr/1995/00000053/00000004/art00003 (February 6, 2020).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

610 Meredith, Michael P., and John C. King. 2005. "Rapid Climate Change in the Ocean West of the Antarctic Peninsula during the Second Half of the 20th Century." *Geophysical Research Letters* 32(19). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL024042 (March 9, 2020).

Miller, Todd R., Lucas Beversdorf, Sheena D. Chaston, and Katherine D. McMahon. 2013. "Spatiotemporal Molecular Analysis of Cyanobacteria Blooms Reveals Microcystis-Aphanizomenon Interactions." *PLOS ONE* 8(9): e74933.

615 Morán, Xosé Anxelu G., Antonio Bode, Luis Ángel Suárez, and Enrique Nogueira. 2007. "Assessing the Relevance of Nucleic Acid Content as an Indicator of Marine Bacterial Activity." *Aquatic Microbial Ecology* 46(2): 141–52.

Moràn, Xosè Anxelu G., Hugh W. Ducklow, and Matthew Erickson. 2011. "Single-Cell Physiological Structure and Growth Rates of Heterotrophic Bacteria in a Temperate Estuary (Waquoit Bay, Massachusetts)." *Limnology and Oceanography* 56(1): 37–48.

620 Prunet, Pascal, Jean-François Minster, Vincent Echevin, and Isabelle Dadou. 1996. "Assimilation of Surface Data in a One-Dimensional Physical-Biogeochemical Model of the Surface Ocean: 2. Adjusting a Simple Trophic Model to Chlorophyll, Temperature, Nitrate, and PCO2 Data." *Global Biogeochemical Cycles* 10(1): 139–58.

Prunet, Pascal, Jean-François Minster, Diana Ruiz-Pino, and I. Dadou. 1996. "Assimilation of Surface Data in a One-
625 Dimensional Physical-Biogeochemical Model of the Surface Ocean: 1. Method and Preliminary Results." *Global Biogeochemical Cycles* 10(1): 111–38.

Reed, Daniel C., Christopher K. Algar, Julie A. Huber, and Gregory J. Dick. 2014. "Gene-Centric Approach to Integrating Environmental Genomics and Biogeochemical Models." *Proceedings of the National Academy of Sciences of the United States of America* 111(5): 1879–84.

Saba, Grace K. et al. 2014. "Winter and Spring Controls on the Summer Food Web of the Coastal West Antarctic Peninsula."
630 *Nature Communications* 5(1): 1–8.

Scharek, Renate, and Mikel Latasa. 2007. "Growth, Grazing and Carbon Flux of High and Low Nucleic Acid Bacteria Differ in Surface and Deep Chlorophyll Maximum Layers in the NW Mediterranean Sea." *Aquatic Microbial Ecology* 46(2): 153–61.

Schattenhofer, Martha et al. 2011. "Phylogenetic Characterisation of Picoplanktonic Populations with High and Low Nucleic
635 Acid Content in the North Atlantic Ocean." *Systematic and Applied Microbiology* 34(6): 470–75.

Schofield, Oscar et al. 2017. "Decadal Variability in Coastal Phytoplankton Community Composition in a Changing West Antarctic Peninsula." *Deep Sea Research Part I: Oceanographic Research Papers* 124: 42–54.

Sherr, Barry F., Evelyn B. Sherr, and Julie McDaniel. 1992. "Effect of Protistan Grazing on the Frequency of Dividing Cells in Bacterioplankton Assemblages." *Applied and Environmental Microbiology* 58(8): 2381–85.

640 Spitz, Y H, J R Moisan, and M R Abbott. 2001. "Configuring an Ecosystem Model Using Data from the Bermuda Atlantic Time Series (BATS)." *Deep Sea Research Part II: Topical Studies in Oceanography* 48(8–9): 1733–68.

Stammerjohn, S. E. et al. 2008. "Trends in Antarctic Annual Sea Ice Retreat and Advance and Their Relation to El Niño–Southern Oscillation and Southern Annular Mode Variability." *Journal of Geophysical Research: Oceans* 113(C3). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JC004269 (March 11, 2020).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Steinberg, Deborah K. et al. 2015. "Long-Term (1993–2013) Changes in Macrozooplankton off the Western Antarctic Peninsula." *Deep Sea Research Part I: Oceanographic Research Papers* 101: 54–70.

Teira, Eva, Sandra Martínez-García, Christian Lønborg, and Xosé A. Álvarez-Salgado. 2009. "Growth Rates of Different Phylogenetic Bacterioplankton Groups in a Coastal Upwelling System." *Environmental Microbiology Reports* 1(6): 545–54.

Thibodeau, P. S., D. K. Steinberg, S. E. Stammerjohn, and C. Hauri. 2019. "Environmental Controls on Pteropod Biogeography along the Western Antarctic Peninsula." *Limnology and Oceanography* 64(S1): S240–56.

Tziperman, Eli, and William Carlisle Thacker. 1989. "An Optimal-Control/Adjoint-Equations Approach to Studying the Oceanic General Circulation." *Journal of Physical Oceanography* 19(10): 1471–85.

Vaughan, David et al. 2003. "Recent Rapid Regional Climate Warming on the Antarctic Peninsula." *Climatic Change* 60: 243–74.

Vaughan, David G. 2006. "Recent Trends in Melting Conditions on the Antarctic Peninsula and Their Implications for Ice-Sheet Mass Balance and Sea Level." *Arctic, Antarctic, and Alpine Research* 38(1): 147–52.

Vila-Costa, Maria, Josep M. Gasol, Shalabh Sharma, and Mary Ann Moran. 2012. "Community Analysis of High- and Low-Nucleic Acid-Containing Bacteria in NW Mediterranean Coastal Waters Using 16S RDNA Pyrosequencing." *Environmental Microbiology* 14(6): 1390–1402.

Ward, Ben A., Marjorie A. M. Friedrichs, Thomas R. Anderson, and Andreas Oschlies. 2010. "Parameter Optimisation Techniques and the Problem of Underdetermination in Marine Biogeochemical Models." *Journal of Marine Systems* 81(1): 34–43.

White, Paul A., Jacob Kalff, Joseph B. Rasmussen, and Josep M. Gasol. 1991. "The Effect of Temperature and Algal Biomass on Bacterial Production and Specific Growth Rate in Freshwater and Marine Habitats." *Microbial Ecology* 21(1): 99–118.

Whitehouse, M. J. et al. 2008. "Rapid Warming of the Ocean around South Georgia, Southern Ocean, during the 20th Century: Forcings, Characteristics and Implications for Lower Trophic Levels." *Deep Sea Research Part I: Oceanographic Research Papers* 55(10): 1218–28.

Yokokawa, Taichi, Toshi Nagata, Matthew T. Cottrell, and David L. Kirchman. 2004. "Growth Rate of the Major Phylogenetic Bacterial Groups in the Delaware Estuary." *Limnology and Oceanography* 49(5): 1620–29.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Figures

675

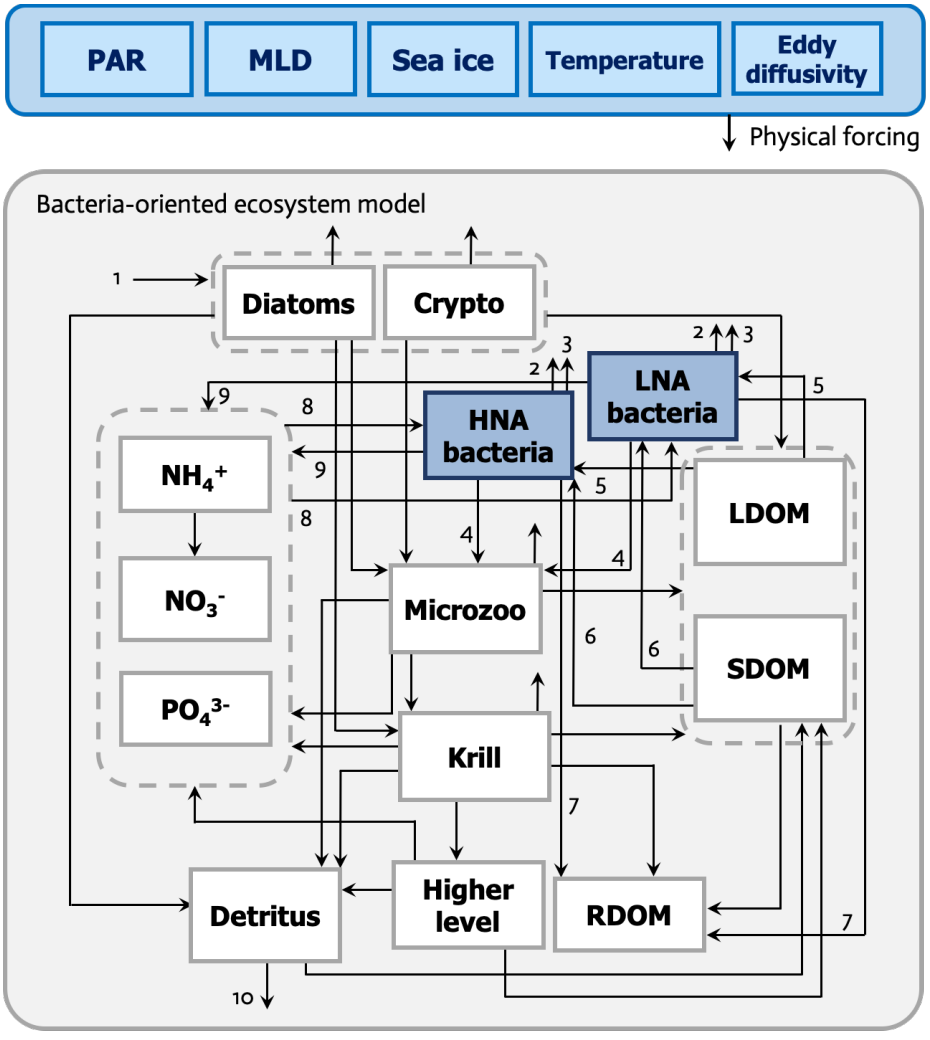
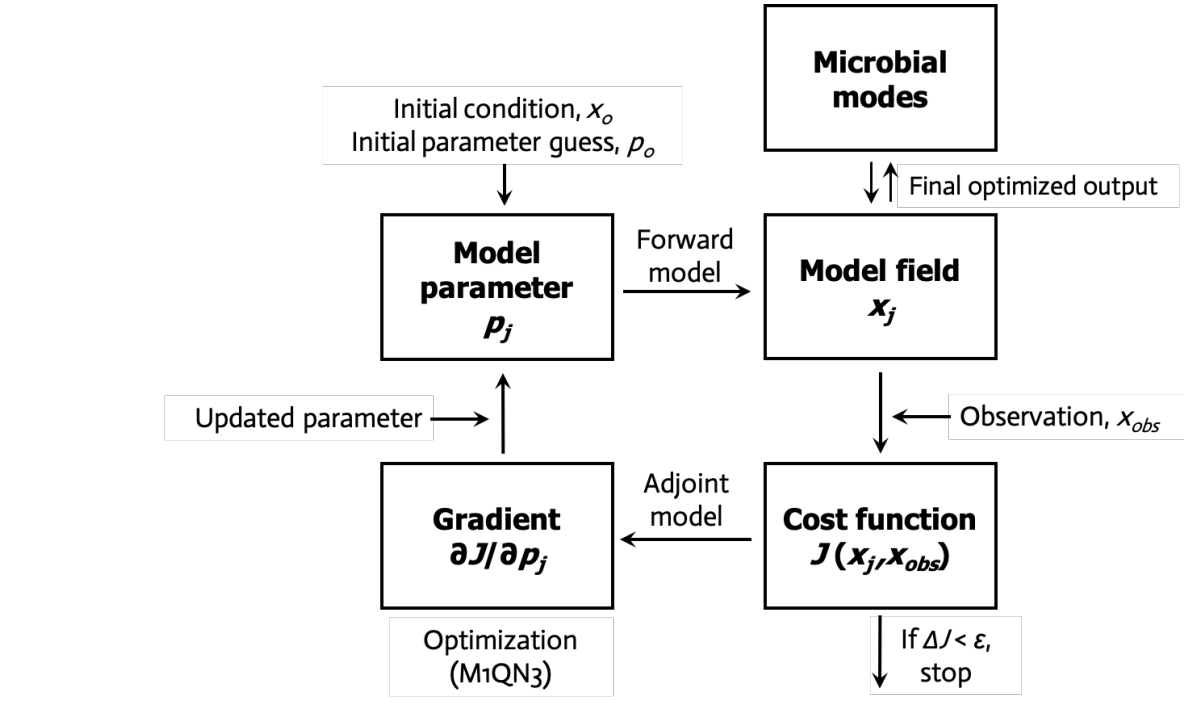


Figure 1: Ecosystem model framework. The model is forced by five different physical forcings, denoted as a horizontal row across the top of the schematic. The ecosystem component incorporates twelve different prognostic state variables. Heterotrophic bacteria (or bacteria) are
680 divided into two groups of differing physiological states, high nucleic acid (HNA) and low nucleic acid (LNA) bacterial compartments. Model closure terms include RDOM and higher trophic levels. The flows between the prognostic state variables with the name of the numbered flows in the legend only represent for two bacterial compartments.

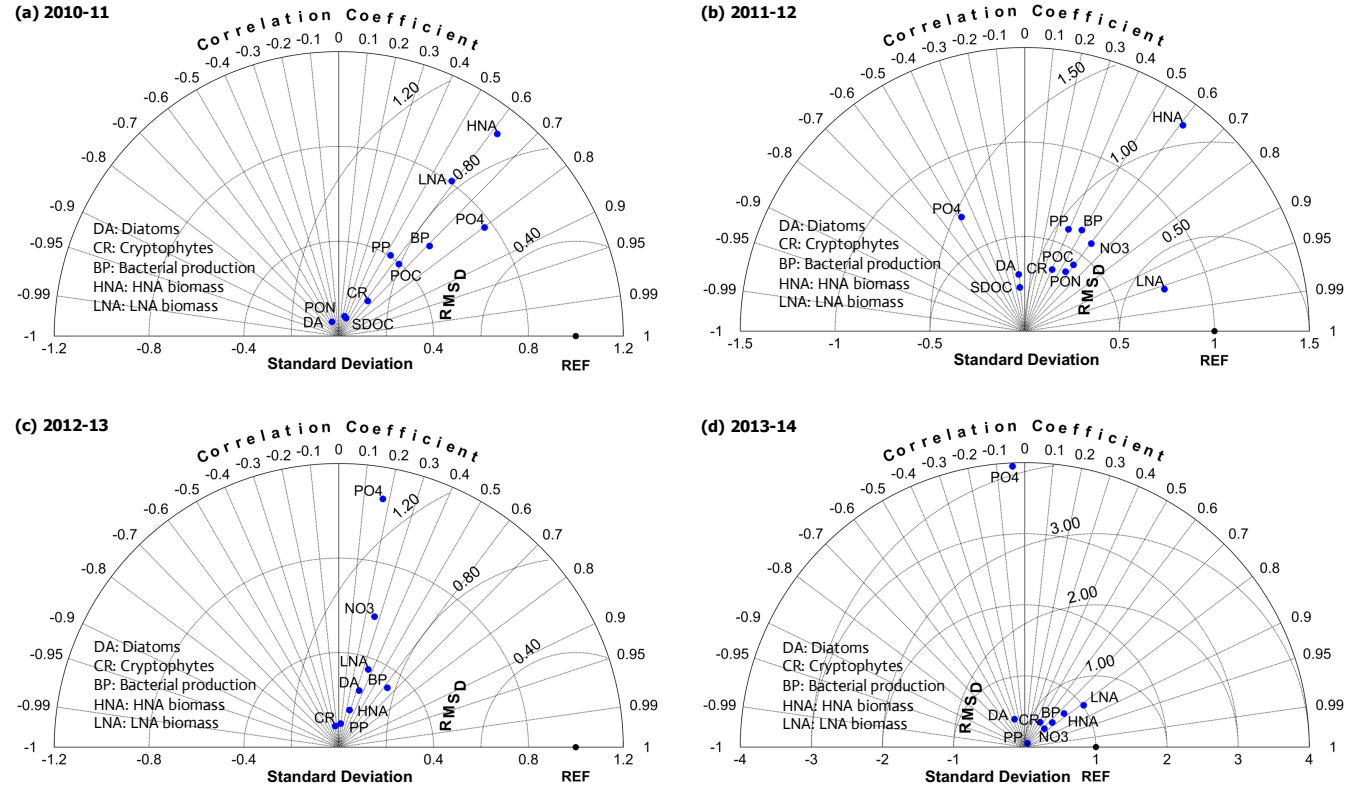
https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020



685

Figure 2: Data assimilation scheme. A variational adjoint method is employed for the parameter optimization and data assimilation processes (adapted from Glover et al., 2011). Gradient: the sensitivity of the total cost function with respect to model parameter from optimization. After optimization is finished, optimized model output is interpreted as a function of bacterial taxonomic modes (modes).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020



690 **Figure 3: Model validation.** A Taylor diagram using a polar-coordinate system summarizing the model-observational correspondence for each model stock and flow for individual annual simulations. The angular coordinate as the Pearson correlation coefficient (r), the distance from the origin denotes the normalized standard deviation, and the distance from point (1,0), marked as REF on x-axis, describes the centred (bias removed) root-mean-square difference (RMSD) between model results and observations. Note different x-axis scales are used for the normalized standard deviation in each panel.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

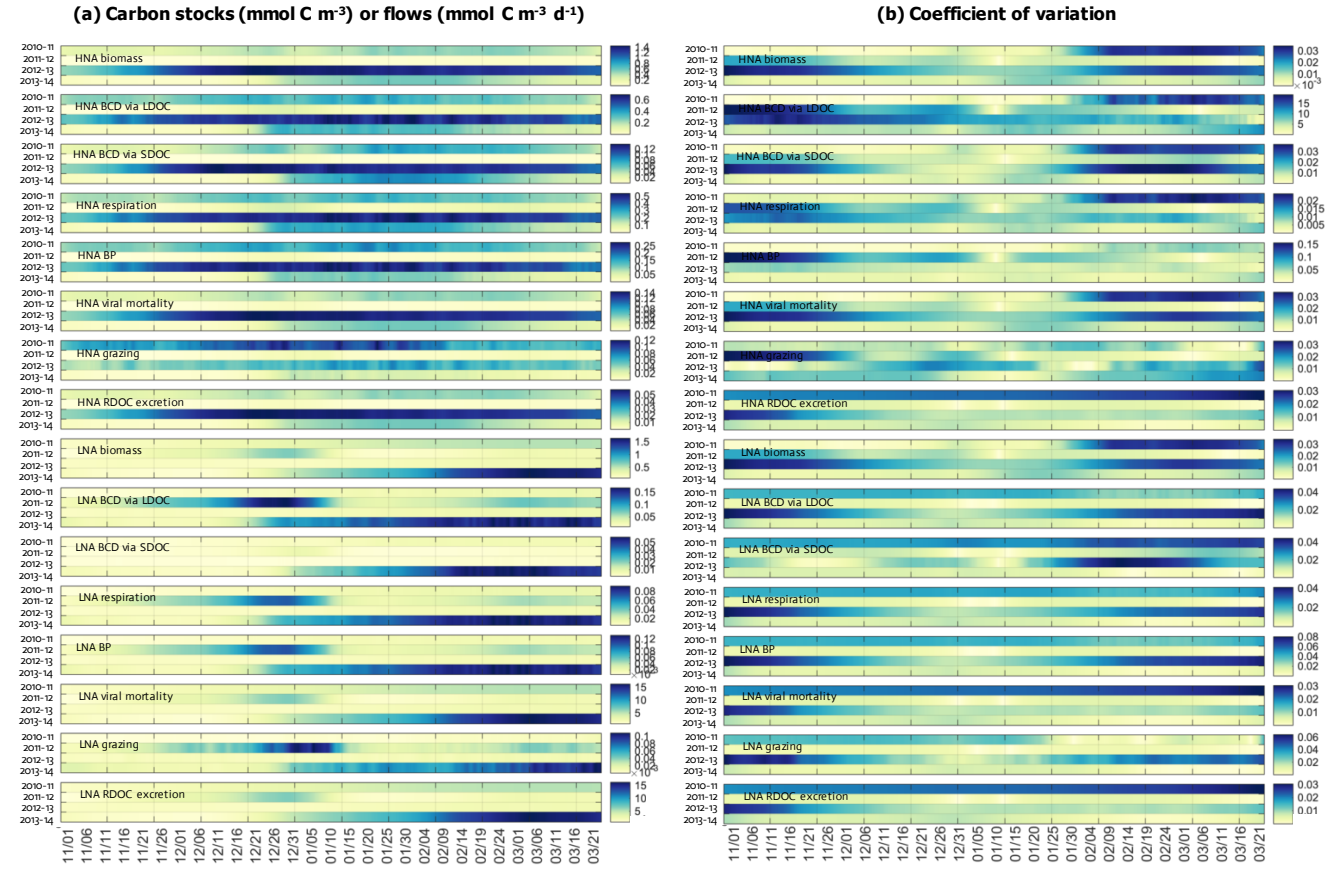


Figure 4: Seasonal time series of group-specific bacterial carbon stocks and flows. Seasonal progression of HNA and LNA bacterial carbon stocks and flows at 10-m depth over the growth season (November-March) for each of the 4 simulation years (a), and coefficient of variation (Monte Carlo-derived standard deviation divided by each data point from Figure 4A) from 1,000 Monte Carlo experiments (b).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

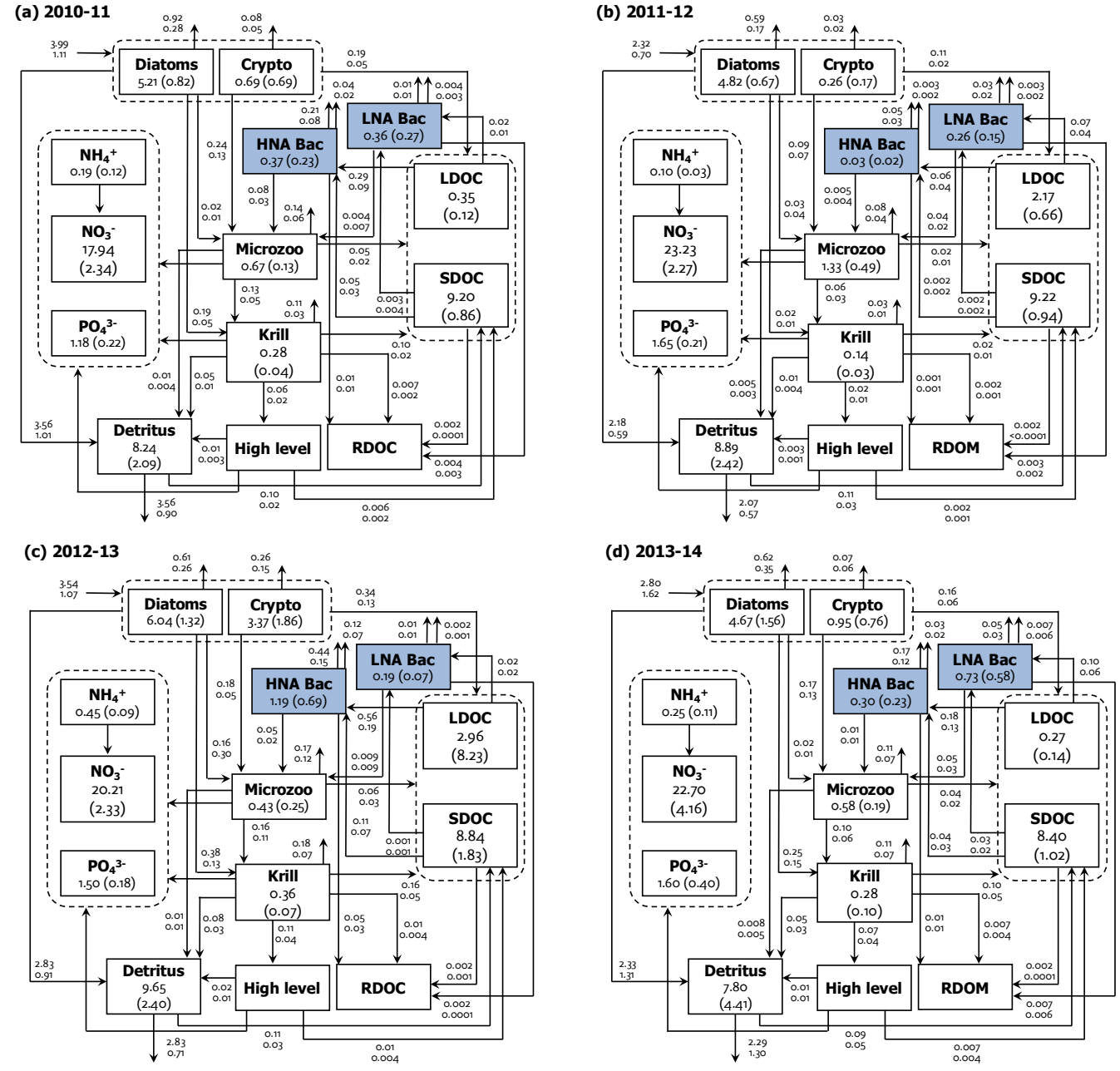


Figure 5: Annual mean carbon stocks and flows. Carbon stocks (mmol C m^{-3}) and flows (mmol C m^{-3} d^{-1}) averaged over the growth season in each year are denoted as the numbers on the first row, while the numbers on the second row or in the parentheses are the standard deviation propagated from averaging over the growth season and the Monte Carlo experiment-derived uncertainties. Flows do not necessarily balance to zero due to the build-up or loss in a compartment over the growth season. N and P flows are omitted. The original unit of POC export fluxes is mmol m^{-2} d^{-1}, but normalized by depth to facilitate comparisons with other C flows and stocks in this diagram.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

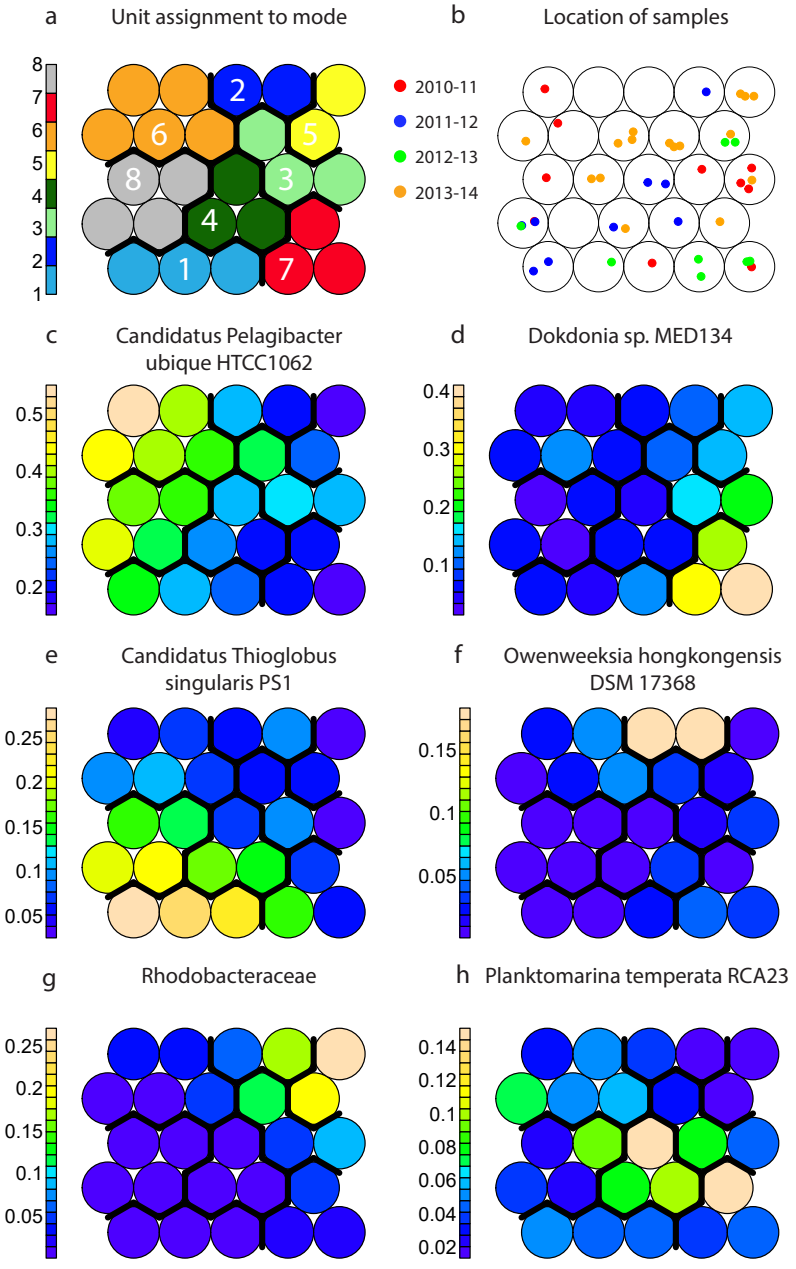


Figure 6: Properties of the emergent self-organizing map for bacterial community structure shown as taxonomic modes (modified from Bowman et al. 2017). (a) The map is trained, then the map units are clustered into modes. Map units are colored and numbered according to taxonomic mode membership. (b) Location of samples used in this study within the map. The map is trained with a larger set of samples, here, only those samples for which BP and flow cytometry data are available (those samples used in this study) are shown. Mode boundaries are the same as in (a). Each sample is placed within the map unit that has the most similar community structure, however, the position of each samples within the map unit is random. (c-h) Relative abundance of the most abundant taxa in the microbial community structure dataset in each map unit after training. The boundaries across all panels are as in (a).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

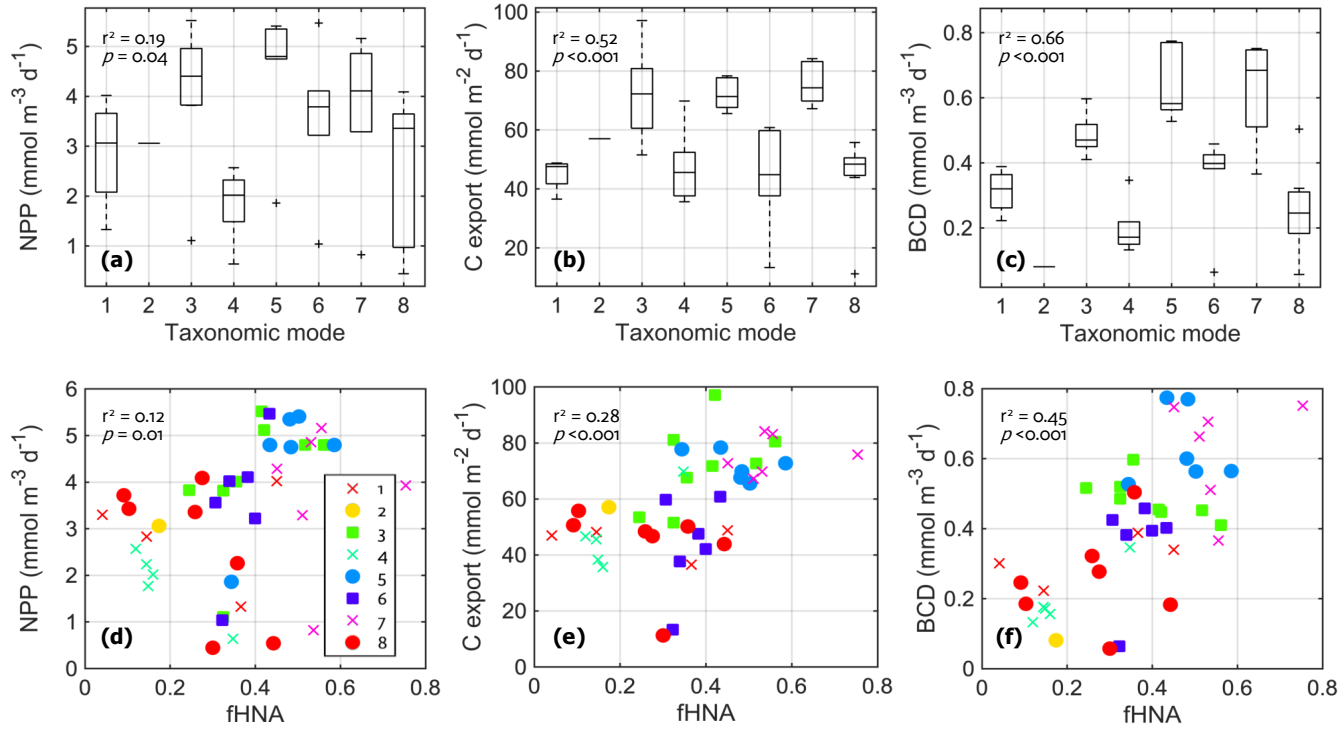


Figure 7: Bacterial physiological and taxonomic association with ecosystem functions. The results of linear regression of the key ecosystem functions on a categorical predictor of mode (not assimilated in the model, a-c) and on the fraction of HNA cells (assimilated in the model, d-f). The fraction of HNA cells (fHNA) is assimilated in the model to constrain each bacterial group, but fHNA here as a predictor is an observed fHNA. The linear regression is performed in which a response variable is NPP, C export, or BCD and a predictor variable is mode as a factor (i.e., mode as a categorical predictor with 8 modes; equivalent to a one-way ANOVA with 8 different categories; a-c). The linear regression is performed in which a response variable is NPP, C export, or BCD and a predictor variable is fHNA (d-f). Regression statistics: (a) number of observations (N) = 43, error degrees of freedom (df) = 35, root mean squared error (RMSE) = 1.36, r^2 = 0.32, adjusted r^2 = 0.19, F-statistic value = 2.36, p-value = 0.04; (b) N = 43, df = 35, RMSE = 12.7, r^2 = 0.60, adjusted r^2 = 0.52, F-statistic value = 7.54, p-value = 1.53e-05; (c) N = 43, df = 35, RMSE = 0.12, r^2 = 0.72, adjusted r^2 = 0.66, F-statistic value = 12.6, p-value = 5.87e-08; (d) N = 43, df = 41, RMSE = 1.42, r^2 = 0.14, adjusted r^2 = 0.12, F-statistic value = 6.46, p-value = 0.01; (e) N = 43, df = 41, RMSE = 15.6, r^2 = 0.29, adjusted r^2 = 0.28, F-statistic value = 17.1 p-value = 1.71e-04; (f) N = 43, df = 41, RMSE = 0.15, r^2 = 0.46, adjusted r^2 = 0.45, F-statistic value = 35.2, p-value = 5.35e-07.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

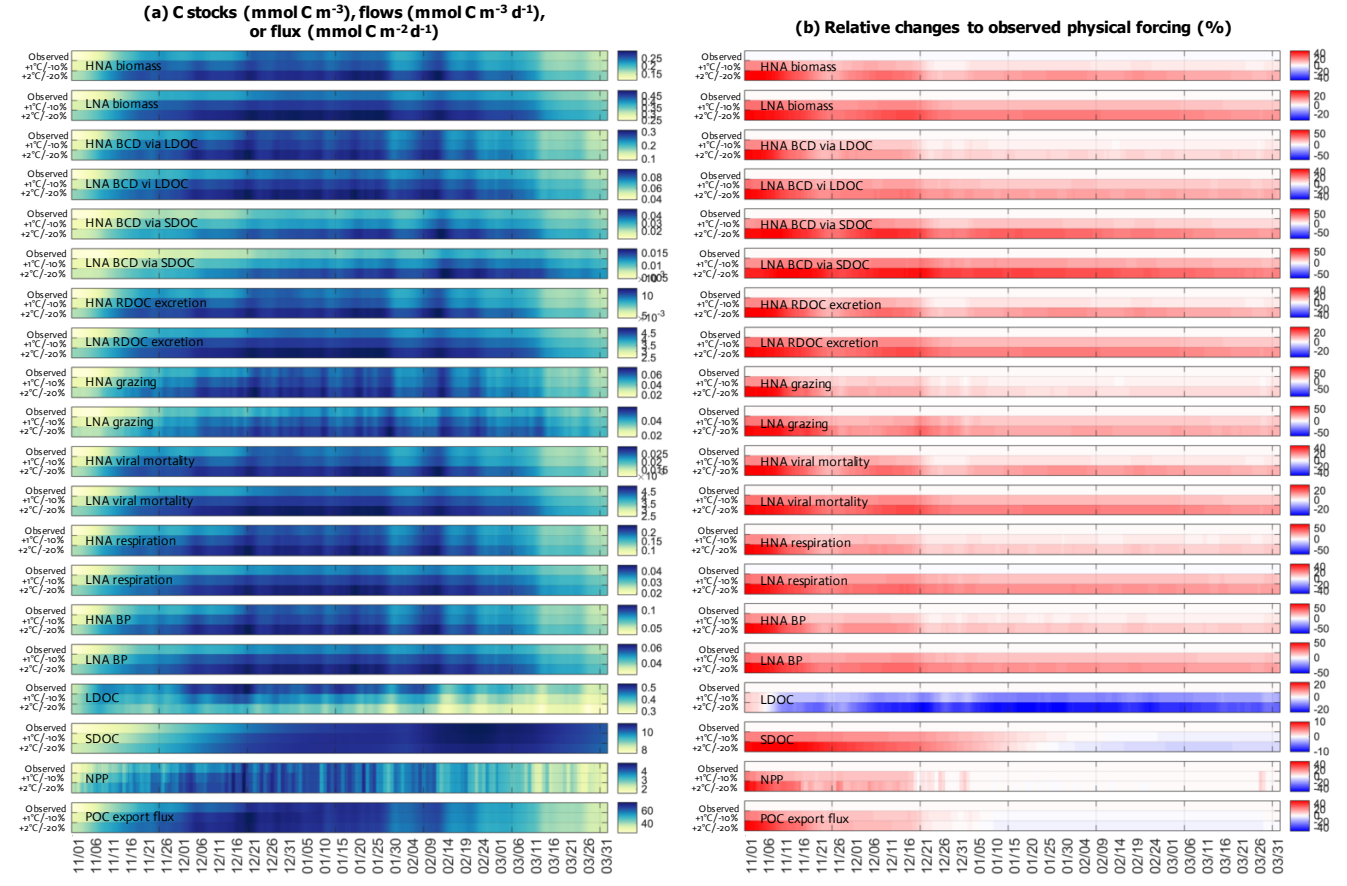


Figure 8: Climate change experiments. Seasonal progression of simulated HNA and LNA bacterial carbon stocks and processes and key ecosystem functions over the growth season under observed physical forcing and climate change scenarios (a) and the percent change of the corresponding variable compared to observed temperature and sea-ice forcing fields (b). For example, HNA biomass in (b) = (HNA biomass under +1°C/-10% - HNA biomass under observed forcing) x 100/HNA biomass under observed forcing) and the first row of each panel as zero (base states).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

Biogeosciences Discussions Open Access EGU

Tables

740

Parameter	Definition	HNA	LNA
k_{DOM} (mmol C m^{-3})	Half-saturation concentration of DOC in the DOC uptake function	0.5	0.2
μ (d^{-1})	Maximum growth rate	5.0	2.0
b_{RESP} ((mmol C m^{-3} d^{-1})$^{-1}$)	Parameter control for active respiration vs. production	0.08	0.2
remi (d^{-1})	Inorganic nutrient regeneration rate	8.0	2.0
ex_{REFR} (d^{-1})	Refractory DOC production rate	0.04	0.01
f_{SLCT}	Selection strength on SDOM	0.1	0.7
$resp^{B}$ (d^{-1})	Basal respiration rate	0.04	0.01
r^{A}_{min} (d^{-1})	Minimum active respiration rate	0.08	0.04
r^{A}_{max} (d^{-1})	Maximum active respiration rate	0.8	0.4
mort (d^{-1})	Mortality rate due to virus (viral mortality)	0.1	0.01
g (mmol C m^{-3})	Half-saturation concentration of bacteria in microzooplankton grazing function	0.55	0.55

Table 1: Optimizable model parameters associated with bacterial processes and their initial guess values. Note different values are assigned to the model parameters of the HNA and LNA groups to simulate their distinct physiological processes and trophic interactions.

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

745

Data types	$\bar{\tilde{a}}$	*CV*	σ	2010-11 model parameter set		2011-12 model parameter set		2012-13 model parameter set		2013-14 model parameter set	
				J'_0	J'_f	J'_0	J'_f	J'_0	J'_f	J'_0	J'_f
NO_3^-	19.70	0.04	0.73	N/A	N/A	8.16	1.10	14.46	3.43	31.68	3.71
PO_4^{3-}	1.31	0.03	0.04	49.41	1.00	21.86	6.86	50.33	3.91	42.32	1.58
Microzooplankton biomass	0.72	0.40	0.28	4.32	0.09	4.09	0.29	5.22	0.17	3.94	0.11
$\log_{10}(Chl_{DIATOM})$	0.15	0.19	0.08	24.57	1.07	20.10	1.45	28.77	2.05	24.46	1.85
$\log_{10}(Chl_{CRYPTO})$	-0.83	0.25	0.10	4.46	0.85	3.62	1.13	5.23	1.80	4.32	1.50
$\log_{10}$(PP)	1.48	0.50	0.21	1.02	0.56	0.79	0.30	3.50	0.48	3.92	0.39
HNA biomass	0.21	0.08	0.02	8.69	0.79	0.12	0.06	8.89	3.11	15.10	2.10
LNA biomass	0.33	0.08	0.03	12.68	0.35	14.12	0.72	3.18	0.56	18.64	1.43
BP	0.11	0.16	0.02	2.38	0.36	0.66	0.29	3.54	0.97	3.63	0.93
SDOC	9.52	0.20	1.93	0.85	0.61	0.25	0.31	-	-	-	-
POC	13.02	0.13	1.70	2.12	0.91	4.95	1.08	-	-	-	-
PON	2.63	0.12	0.32	11.91	8.85	2.62	0.80	-	-	-	-
Total cost				122.42	15.43	81.34	14.40	123.12	16.47	147.98	13.61

750 **Table 2: Data types, observed means, coefficient of variation, target errors, and costs before and after optimization.** The observed mean ($\bar{\tilde{a}}$), coefficient of variation (CV), and target error (σ) of each assimilated data type used for calculating the cost function before and after optimization. Normalized cost functions (unitless; Equation 5) are presented where J'_0 is the normalized cost function before optimization and J'_f is the normalized cost function after optimization. Data type units as: mmol m^{-3} for nitrate (NO_3^-), phosphate (PO_4^{3-}); mmol C m^{-3} for microzooplankton biomass, diatom chlorophyll (Chl_{DIATOM}), cryptophyte chlorophyll (Chl_{CRYPTO}), HNA and LNA bacterial biomass, SDOC, and POC; mmol N m^{-3} for PON; and mmol C m^{-3} d^{-1} for primary production (PP) and bacterial production (BP). Normalized total costs in 2012-13 and 2013-14 model parameter sets do not include the costs of SDOC, POC, and PON (not assimilated, denoted as '-' in the table).

https://doi.org/10.5194/bg-2020-302
Preprint. Discussion started: 2 September 2020

760

Data types	J'_c	2010-11 model parameter set				2011-12 model parameter set				2012-13 model parameter set				2013-14 model parameter set			
		J'_f	$J'_{x,\ 2011\text{-}12}$	$J'_{x,\ 2012\text{-}13}$	$J'_{x,\ 2013\text{-}14}$	J'_f	$J'_{x,\ 2010\text{-}11}$	$J'_{x,\ 2012\text{-}13}$	$J'_{x,\ 2013\text{-}14}$	J'_f	$J'_{x,\ 2010\text{-}11}$	$J'_{x,\ 2011\text{-}12}$	$J'_{x,\ 2013\text{-}14}$	J'_f	$J'_{x,\ 2010\text{-}11}$	$J'_{x,\ 2011\text{-}12}$	$J'_{x,\ 2012\text{-}13}$
NO_3^-	3.04	N/A	1.43	2.99	4.02	1.10	N/A	3.31	12.46	3.43	N/A	1.07	4.55	3.71	N/A	1.57	4.28
PO_4^{3-}	3.63	1.00	11.37	7.09	2.18	6.86	5.38	7.55	6.65	3.91	2.87	7.22	1.62	1.58	1.03	11.52	4.64
Microzooplankton biomass	0.08	0.09	0.09	0.07	0.11	0.29	0.21	19.06	9.27	0.17	0.13	0.13	0.15	0.11	0.08	0.08	0.10
$\log_{10}(Chl_{DIATOM})$	1.28	1.07	1.05	1.09	1.81	1.45	1.65	1.84	1.75	2.05	1.71	1.51	2.14	1.85	1.33	1.20	1.62
$\log_{10}(Chl_{CRYPTO})$	1.10	0.85	1.11	4.07	2.46	1.13	1.61	5.80	2.08	1.80	1.62	1.32	1.92	1.50	1.01	0.75	0.98
$\log_{10}(PP)$	0.35	0.56	0.44	0.56	0.42	0.30	0.37	0.64	0.74	0.48	0.56	0.49	0.42	0.39	0.60	0.50	0.46
HNA biomass	2.96	0.79	4.06	16.20	3.39	0.06	7.04	5.67	12.27	3.11	1.00	4.62	4.30	2.10	0.66	7.42	5.73
LNA biomass	2.45	0.35	8.98	0.64	7.83	0.72	2.35	209.75	111.14	0.56	1.74	7.55	7.01	1.43	1.69	4.78	10.16
Total BP	0.91	0.36	0.78	1.26	1.20	0.29	1.49	1.99	0.73	0.97	0.50	1.12	1.29	0.93	0.40	0.93	0.88
SDOC	0.45	0.61	0.29	-	-	0.31	0.57	-	-	-	0.54	0.40	-	-	0.56	0.36	-
POC	0.87	0.91	1.30	-	-	1.08	1.41	-	-	-	2.37	1.05	-	-	2.51	1.07	-
PON	2.03	8.85	0.91	-	-	0.80	8.44	-	-	-	8.28	0.80	-	-	8.30	0.80	-
Total cost	19.14	15.43	31.81	33.99	23.41	14.40	30.53	255.60	157.09	16.47	21.32	27.26	23.38	13.61	18.17	30.99	28.87
Portability index		0.66 ± 0.14				0.27 ± 0.31				0.81 ± 0.10				0.78 ± 0.24			

Table 3: Cross-validation cost analysis and portability index. The normalized cost function of each year and of the climatological year (i.e., climatological model). J'_c is the normalized optimized cost from the climatological model, J'_f is the normalized optimized cost for each year after optimization (Table 2), and J'_x is the normalized cross-validation cost computed from simulating a different year's optimized model parameter set against a given year. For example, $J'_{x,\ 2011\text{-}12}$ under 2010-11 model parameter set indicates the normalized cross-validation cost computed from simulating 2010-11 model parameter set against 2011-12. The portability index (Equation 7) reflects the generality of parameter sets, where the index <<1 (or closer to 1) implies the system that is (or is not) sensitive to year to year variations in optimized model parameters.

765