DAR1 - An Easy to Use, Single-Box, Configuration of a Global Scale Biogeochemical Ecosystem Model

Barbara Duckworth[1,✉], **Stephanie Dutkiewicz**[1], **Oliver Jahn**[1], **Christopher N. Hill**[1], **and Christopher L. Follett**[2, 1, ✉]

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[2]Department of Earth, Ocean, and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom

[*]Corresponding Authors: birdy@mit.edu ; C.L.Follett@liverpool.ac.uk

Global scale biogeochemical ocean models are key to understanding the mechanisms which drive elemental cycles and their links to marine ecosystems. However, these models are rarely straightforward to use and their complexity can make them difficult to interpret. Here, we aim to address this problem by developing DAR1 – a single box configuration of the MIT Darwin model, which is itself a global scale biogeochemical ecosystem model designed for understanding the emergent patterns of global plankton distributions. Our configuration is packaged inside Docker and implemented in the Julia programming language. This helps make it both computationally efficient and straightforward to implement on any local machine, independent of hardware and operating system. It allows for the replication of field and laboratory experiments in a model ocean and the rapid comparison of global runs with local predictions. We provide three case studies: a framework for nutrient amendment bottle experiments; a test of steady state assumptions on plankton community structure at the global scale; and an exploration of the niche space of nitrogen-fixing plankton. Our method has the potential to facilitate broader direct testing of global model assumptions in the context of small-scale measurements. DAR1 will also allow observational and laboratory oceanographers to directly test their results against model predictions in a straightforward, easy to use, manner.

Introduction

Marine microbes play an important role in biogeochemical cycling by fixing atmospheric carbon into organic compounds which, when exported to depth, leads to carbon sequestration in the deep ocean (1, 2). The plankton community is equally important for feeding higher trophic levels and supporting the production of global fisheries. Phytoplankton community structure strongly influences these biogeochemical and ecological pathways, which are in turn affected by physical processes and nutrient availability (3–5). Thus, understanding plankton community structure is best done in the context of the physical-chemical environment that they inhabit and in turn influence. Coupled biogeochemical-ecosystem computational models offer insight into this complex system. The development of global ecological-biogeochemical ocean models has been critical for understanding the underlying relationships between the tightly coupled ecology and biogeochemistry of the marine environment (6–8). These models are composed of differential equations describing the dynamics and interplay between the physics, biogeochemistry, and ecology of the ocean, often referred to as a "mechanistic" approach.

The MIT Darwin model couples the ecology of marine plankton with ocean circulation and biogeochemistry (9, 10). Past applications of this model have explored the drivers of plankton diversity (3); explained spatial and temporal patterns in species biogeography (11–13); and made predictions for how communities will be altered in a warming ocean (14).

The model resolves the cycling of C, N, P, Fe, Si, and O_2 through living, detrital and inorganic pools. The model resolves a set number of plankton, each of which is assigned traits that define its size, trophic, and resource status, thermal responses, and (for phytoplankon) light affinity. It was developed to be flexible in the number and type of plankton that it resolves. In the current version of Darwin, plankton can be autotrophic, heterotrophic grazers, heterotrophic reminerlizers or some combination thereof (3, 13). This "trait-based" approach is a flexible way to simulate a wide variety of plankton types. Darwin is the ecological-biogeochemical component of the MIT general circulation model (MITgcm, (15)), which provides the physical framework for the model. The physical configuration itself is also highly flexible, allowing for different horizontal resolutions (from km to hundreds of km) in regional (e.g. (16, 17)) and global (e.g. (10, 18)) runs, or as a 1-D water column (e.g. (19, 20)). However, several key challenges exist in using the Darwin model in its standard form: ease of use, computational efficiency, and interpretability.

Using the MIT Darwin model requires niche knowledge and technical skills with a difficult learning curve. Like many global scale circulation and biogeochemistry models, it is written in Fortran, requires a complicated build environment, and any changes to the configuration or parameter choices involve changing numerous input files. It is the computational equivalent of a sophisticated scientific instrument – programming's version of a mass spectrometer. Global scale simulations require large computer clusters and, depending on the spatial resolution of the configuration, can take hours to months to run for several years of model time. Given the complexity of the model, diagnosing and interpreting the re-

sulting output is non-trivial.

Here we present a new setup of the Darwin model, "DAR1", which offers a simple to use, 0-D, model configuration that can help interpret the large-scale model, provide fast searches in variable space, and act as an analogue of a bottle experiment. This setup is designed to be accessible to new users, but the setup can also be used by an expert for a range of purposes, including assisting with the interpretion of results from large-scale complex Darwin simulations.

DAR1 is an interface written in Julia that simulates a water parcel, driven by the Darwin model (3, 9, 13). It is built off of the Julia packages MITgcmTools and ClimateModels (21, 22). From an observational perspective, it is the equivalent of the perfect bottle incubation experiment. For example, water parameters can be taken from the global simulation and manipulated to elucidate which process are important. Having a single-box model that is internally consistent with the global model makes it straightforward to move between spatial scales and see how dynamics differ. With DAR1, the user can control nutrient levels, light, temperature, and initial plankton community. Instead of needing to modify many runtime files for changes in parameters, initial conditions, or specific environmental forcing, these values are set in a single file (see Table 1 for term definitions). Available documentation details the model setup, and a Docker container is provided to alleviate the build environment requirements (23). Under many conditions, it takes only seconds to run on a standard desktop machine.

DAR1 offers observational and laboratory oceanographers a valuable tool to test incubation and culture experiments. By testing the Darwin model against small scale experiments directly, the user can see when the mechanistic underpinnings of the model are, or are not, replicated by experiments. This type of feedback between experiments and models can inform future work and lead to more rapid testing of hypotheses.

Certainly, small scale ecological models are a mainstay of ecology and have been used by many researchers to test ecological theories as compared to experiments. DAR1 is novel because it provides a seamless connection from the accessible "toy" model back to larger scale models and global predictions. Often, one wants to compare the predictions of simplified ecological theory to the emergent biogeography of plankton (11, 12). We currently lack models which bridge the gap between "toy-models" and the full global simulations. Instead, ad hoc models of partial complexity must be built each time this connection is desired (12). By using the DAR1 model, insights and critiques are automatically consistent with the tools used to understand and make predictions about our current and future ocean (14).

This paper describes the design and implementation of DAR1. It also provides three use cases: a nutrient amendment bottle experiment; a comparison of time-varying vs. steady state community predictions; and the niche partitioning of nitrogen fixers. We are hopeful that this configuration will aid in bridging the gap between observations, observationalists, models and modelers. Additionally, this class of model is well suited for large searches of model parameter space, and could aid in formally fitting ecological models to plankton observations.

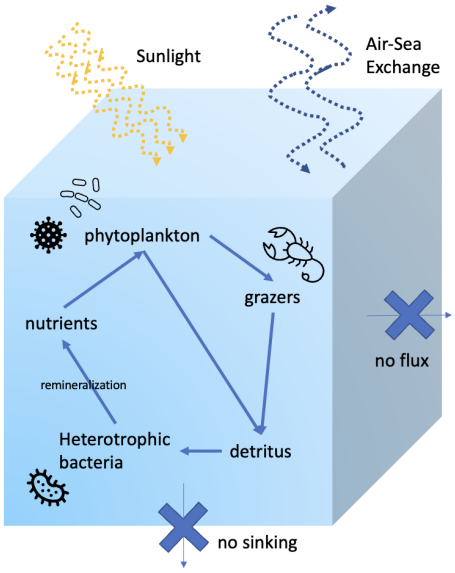

Fig. 1. DAR1 Model Configuration: The DAR1 system models a single, well mixed parcel of water. The standard configuration consists of the Darwin ecosystem-biogeochemical model, with free gas exchange with the atmosphere. Horizontal and other vertical fluxes are set to zero. These conditions can be modified by the user.

Design

Structure and Configuration. DAR1 is composed of three main parts:

1. A configuration of the Darwin model, written in Fortran, specifying a 1x1x1 meter cube of water for the simulation, along with default values for light, temperature, and biogeochemical tracers.

2. A Julia package for building, initializing, and running the Darwin configuration.

3. A Docker container which includes all necessary dependencies to run DAR1. Docker containers allow code that normally has specific hardware and system requirements to be run on any computer regardless of operating system. (See Table 1)

DAR1 is configured to be a single well-mixed box of water with air-sea exchange at the surface, no flux of matter through the sides, and no sinking. Within this box, in this specific configuration of the Darwin model (see Supplemental Methods), nutrients are cycled from inorganic to organic through photosynthesis and respiration, matriculate up the trophic levels through grazing, and return to inorganic form by heterotrophic bacteria (Figure 1). This system is similar to a bottle experiment in which a quantity of water is collected from the ocean, closed off, and then community composition and nutrients are measured periodically as time passes. The model configuration can be modified to alter the number of

plankton and to exclude grazers, bacteria, or other functional groups. While this paper presents a single configuration, alternative configurations are possible and will be made available in the future.

Workflow. DAR1tools is a package written in Julia that contains methods to build, run, and modify the Darwin model configuration. A standard workflow involves:

.1. Building. Traditionally, due to hardware and dependency differences, the MIT Darwin model needs to be built on whichever computer will be running it. With DAR1, the build step is only done once unless build-time parameters for the model need to be changed. The Docker container comes with the model already built and fit for the specifications of the docker container, allowing Docker users to avoid this step.

.2. Setup. Each experiment is set up with a unique identifier, called a run ID. This setup process creates a folder to store the simulation runs, links up the Darwin model code, copies over initial parameter files (see "namelist files" in Table 1), and returns an object that is used in all following steps to specify which files to modify and run.

.3. Parameter modification. Once the experiment is set up, the namelist files that contain run-time parameters can be modified. DAR1 has ready-made functions for changing the most commonly accessed parameters, such as specifying the length of run, temperature, and initial nutrient and biomass concentrations. These ready-made functions cover the basics, but the user can also change any runtime parameter if the namelist file, group, and parameter names are known as per the Darwin documentation (Supplementary Information). Parameters are often "seeded" from values taken from a global Darwin run. Therefore, DAR1tools contains a sample netCDF file with typical global tracer values for a one year time period at weekly temporal resolution.

.4. Running the model. After setup and parameter modification, the simulation can be run using a single command. The runtime for a simulation is dependent on how frequently output files are written. On a Lenovo Thinkpad L13 Core i5-1235U 8GB Memory, the three case studies from this manuscript took 45 seconds, 3 hours, and 45 minutes respectively.

.5. Examining results. The simulation output is located in the run ID folder, in the form of netCDF files (see Supplemental Information). The output can be analyzed with Julia or another language. Example Julia and ipython notebooks with data visualizations are included in the DAR1 Github.

Parallelization. The MIT Darwin model can be configured to run in any 3D arrangement of grid cells. Traditionally, tracers would be advected and diffused between grid cells to simulate the flow of currents in ocean basins. DAR1 takes advantage of Darwin's ability to run many water "parcels" in parallel by turning off communication between grid cells, so

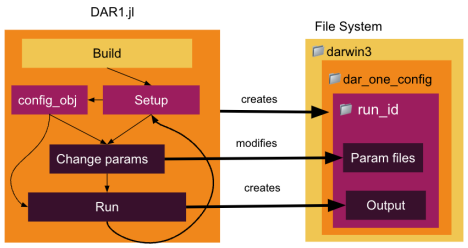

Fig. 2. DAR1 Anatomy: DAR1tools provides an interface for interacting with the complex parameterization and file structure of the MIT Darwin model.

each is its own isolated environment. This allows for a particular build of DAR1 to be run multiple times with varying initial conditions, such as increasing nutrient supply or seasonal temperature variations. This functionality is showcased in case studies B and C where a grid is used to explore global spatial variations in the former and a range of NO3 and PO4 initial conditions in the latter.

Constant Nutrients or other Parameters. In some cases, it is useful to set the concentrations of different tracers (e.g. nutrient concentrations) to be constant, rather than a time-changing quantity. In the default configuration of DAR1, all nutrients are cycled within the system – only gases can escape through the surface of the box. With DAR1, you can optionally choose to hold a number of nutrients or other model state variables (e.g. biomass, organic matter) constant. In the model, this is accomplished by asserting that the change between time steps is zero for the specified tracers.

Case Studies Demonstrate Functionality

DAR1 is built with a set of core functionalities which allow it to effectively bridge the gap between field observations and model predictions. In its simplest configuration, DAR1 operates like an ideal bottle experiment. The initial conditions are set from a parcel of water, and then the ecology and biogeochemistry inside the box are allowed to evolve, obeying the laws of mass conservation. A simple implementation of this use is explained in case study A. Two additional functions of the model are explored in case studies B and C. Case study B leverages the grid-based functionality of the Darwin model to compute the steady state across a grid of initial conditions taken from the surface of a modeled global ocean. Case study C uses this gridded functionality to search nutrient space for the niches of nitrogen fixers in the model ecosystem. Additionally, case study C introduces the ability of DAR1 to arbitrarily set the concentration of any tracer to a constant value over a timeseries. This allows the steady state to be explored in ecosystems that do not conserve nutrient concentration. In general, this functionality allows users to build generalized "chemostat" type models within the Darwin ecosystem.

Case Studies

A. A simplified, nutrient amendment, bottle incubation experiment. Here we show how DAR1 can be used to

Table 1. Definitions for technical terms used in the main text.

Term	Definition
Tracer	A term used in Darwin/DAR1 to describe all nutrients, biomass, and organic matter that change over time based on model calculations of losses and gains (synonymous with "state variable")
Namelist	Fortran term for a file defining a group of parameters and variables
Docker	A set of products that use OS-level virtualization to deliver software in packages called "containers"
Docker container	A standardized, encapsulated environment that runs applications
Docker image	A lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries, and settings. A Docker image is used to build a container
State variables	Fields representing biomass, nutrients, and organic matter whose values are updated at each timestep in the model based on calculated loss and gain terms (in Darwin this also includes advection and mixing). Sometimes referred to as "tracers"
Parameters	Values used in the model to define specific rates (e.g. maximum growth rate, grazing rates) or constants (e.g. maximum Chl to C)
Build-time parameters	Parameters that must be set before the model is compiled, sometimes called "compile-time"
Run-time parameters	Parameters with values that can be changed before or between simulations without re-compiling the model
Build	The compilation process, during which source code is converted to executable code
Initial condition	The starting value of an evolving state variable, such as biomass or nutrient concentration
Time step	The incremental change in time for which the governing equations of the system are being solved
Steady state	A situation in which all state variables reach a equilibrium, such that they no longer change from timestep to timestep
NetCDF	A data format that supports labeled and annotated multi-dimensional scientific data

conduct a simulation of a bottle incubation experiment, using the North Pacific Ocean as a model environment. The oligotrophic gyre north of Hawaii is in a perpetual state of nitrogen limitation (24), with nitrate and deep seawater amendments leading to large increases in biomass and primary production in bottle experiments (25). A similar experiment is run in DAR1 by taking surface water from the gyre just north of Hawaii, which had a NO_3 concentration of 0.007 mmol N/m^3, and increasing the nitrate 10x to a concentration of 0.07 mmol N/m^3 (Figure 3). The "bottle" is then sampled every 12 hours over the course of a six day experiment. The biomass timeseries is then compared with the control experiment in which no additional nitrate was added. As expected in a nitrogen starved environment, we see that nitrogen amendment leads to a steeper increase in biomass over the first day, and an overall greater amount of biomass over the six day sampling period. The decrease in biomass after the peak is due to the re-partitioning of nutrients into dissolved and particulate organic matter.

This type of time series simulation is a standard work-flow with DAR1 and is a useful analogue to real world in-situ experiments. Users can pick an initial condition and use DAR1 to simulate the temporal evolution. A link to a step-by-step tutorial for replicating this case study is available in the Supplementary Information.

B. Community Structure: Time Varying vs. Steady State. Ecological theories are powerful tools to explore the

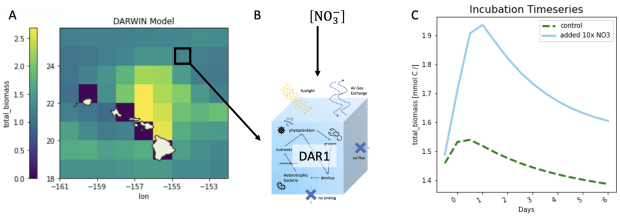

Fig. 3. Case Study A: (A) Nutrients, phytoplankton, and other environmental variables were chosen from the point indicated on the map north of Hawaii to initialize DAR1. In the control, all variables were kept the same. The initial amount of NO_3 was 0.007 mmol N/m^3. In the nutrient amendment experiment, the amount of initial NO_3 was increased to 0.07 mmol N/m^3. (C) Shows the total amount of biomass in the control (green dashed line) and the case with additional nitrogen (solid blue line). In the case with additional nitrogen, the total biomass remains greater than that in the control.

underlying mechanisms setting organism distributions and co-existence (26–28). With elegant and simple equations the fundamentals of the system can be explored, providing unique insight. However, there remains a question about when the simplicity is too much given the vastly more complex real world, especially when often these theories are solved assuming a steady state: a situation when the system does not change over time.

Previous Darwin model studies have utilized steady state theories (e.g. Resource Competition Theory, Resource Ratio Theory (26, 29)) to understand the controlling mechanisms on the resulting modelled plankton community structure (3, 6, 11). Here, we demonstrate that (as found in (6))

steady state predictions are helpful for understanding oligotrophic gyres, but are less useful at high latitudes. In particular, with DAR1 we can investigate what the actual steady state results would be without using the full 3D global configurations. In this setup we implement a parallel version of DAR1 inside Docker.

Over time, species abundance will eventually converge to, or oscillate around, a steady state value. This property can be understood using a system of Monod equations (simplified from what is in DAR1) describing the total biomass of phytoplankton, P, and its limiting nutrient, R, supplied to the system at rate S_R:

$$\frac{dP}{dt} = \mu_m \frac{R}{R+\kappa_R} P - mP \quad \textbf{(1)}$$

and

$$\frac{dR}{dt} = -\mu_m \frac{R}{R+\kappa_R} P + S_R. \quad \textbf{(2)}$$

Here, μ_m is the maximal growth rate, κ_R is the half-saturation constant, and m is a loss term representing death, sinking, grazing, etc. The system reaches a stable equilibrium if the derivatives go to zero $\left(\frac{dP}{dt} = 0,\ \frac{dR}{dt} = 0\right)$. Fluctuating steady states can also be reached where the long-time average of the derivatives is zero. For the above simplified system of equations, a constant steady state is reached where the community composition is uniquely determined by the steady state nutrient concentration (3, 11, 26). However, this leaves the question of when the predictions of the steady state theory can be applied to the spatially and temporally fluctuating ocean. With DAR1, we can explicitly ask this question in the context of the full Darwin ecosystem by taking a vector of initial parameters (nutrient concentrations, biomass, light, temperature) and calculating the abundance of phytoplankton and nutrients over time using the same equations that power the Darwin model. With no fluctuations in light or temperature, and without transport from outside the parcel of water, the system can be run to steady state (i.e. when biomass, nutrients no longer change over times) or fluctuating steady states where the long-time average of the derivatives is zero.

For this case study, DAR1 is configured to run in a 160x360 matrix (each member of the matrix matching a 1 degree grid cell of a global configuration of Darwin), with no flux between cells. The nutrient and biomass concentration are initialized using the yearly average temperature, light, nutrient concentrations, and biomass of each cell's counterpart in the global Darwin model. This grid of DAR1 instances is run for 10 years, until the plankton communities reach a steady, or quasi-steady, state. We average over the last year of the simulation to get the steady state prediction from DAR1 and compare this output to the global Darwin run using the Bray-Curtis Dissimilarity Index. This index is defined as (30):

$$\beta_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j} \quad \textbf{(3)}$$

where

$$C_{ij} = \sum min(s_{ki}, s_{kj}). \quad \textbf{(4)}$$

Here s_{ki} is the biomass of a single species and S_i is the total biomass for location index i (the same holds for index j). C_{ij} is the sum of the minimum biomass over all species. Areas with a high dissimilarity are significantly different from each other. Conversely, a low dissimilarity score indicates that the areas are similar.

The model comparison reveals that oligotrophic gyres have a low dissimilarity score, meaning that the yearly average community of the global Darwin model is similar to the same conditions run to a steady state over many years (Figure 4). In these regions, resource competition theory can be useful for understanding the structure of the phytoplankton community. In nutrient replete high-latitude waters, the dissimilarity score is high, indicating that, in these areas, the community at the steady state is not an appropriate approximation of the more realistic global model, which includes seasonal changes to a variety of variables. For these regions, the dominant driver of phytoplankton community composition is not simply an organism's ability to draw down nutrients below competitor's requirements as suggested by resource competition theory. Rather, it is the interplay between the time-varying forcing and the equilibration timescales of the ecosystem. A link to a tutorial for running this case study can be found in the Supplemental Information.

C. Resource Ratio Theory: Niches of Nitrogen Fixation. Nitrogen fixers or diazotrophs can fix nitrogen gas into an organic form, rather than relying on the often low concentrations of fixed nitrogen such as nitrate. This makes them unique and important members of the microbial community. However, in general these organisms have slower growth rates than similar sized non-diazotrophs given the energetic constraints of fixing nitrogen. Several studies have explored the environmental factors defining the photosynthesising diaztroph plankton biogeography (12, 31). Using a combination of resource supply ratio theory (29) and numerical (Darwin) model simulations some of these studies suggest that diazotrophs co-exist with faster growing diazotrophs where the supply of iron and phosphate is in excess of the non-diazotrophs' dissolved inorganic nitrogen (DIN) requirements (32, 33). Here, we use DAR1 to explore these ideas in more detail. However, in DAR1, if we allow nutrients concentrations to evolve over time, we would find that the nitrogen fixing organisms gradually add nitrogen to the bottle until DIN limitation is alleviated. At this point, nitrogen fixers will be out-competed by their non-nitrogen fixing brethren of the same size. In order to circumvent this issue, we utilize another feature of DAR1. We set the concentration of available DIN in the water to be constant, effectively discarding any additional new DIN that is fixed from nitrogen gas by diazotrophs. We also set the concentration of iron and phosphate constant over time - this does violate mass conservation, but provides a unique platform for exploring the importance of nutrient ratios in setting diazotroph biogeography.

In this experiment we run a series of simulations to steady state across a range of nitrate and phosphate concentrations (see Fig. 5). We set iron to be in excess, so it will never limit the growth of plankton in the system. We take initial condi-

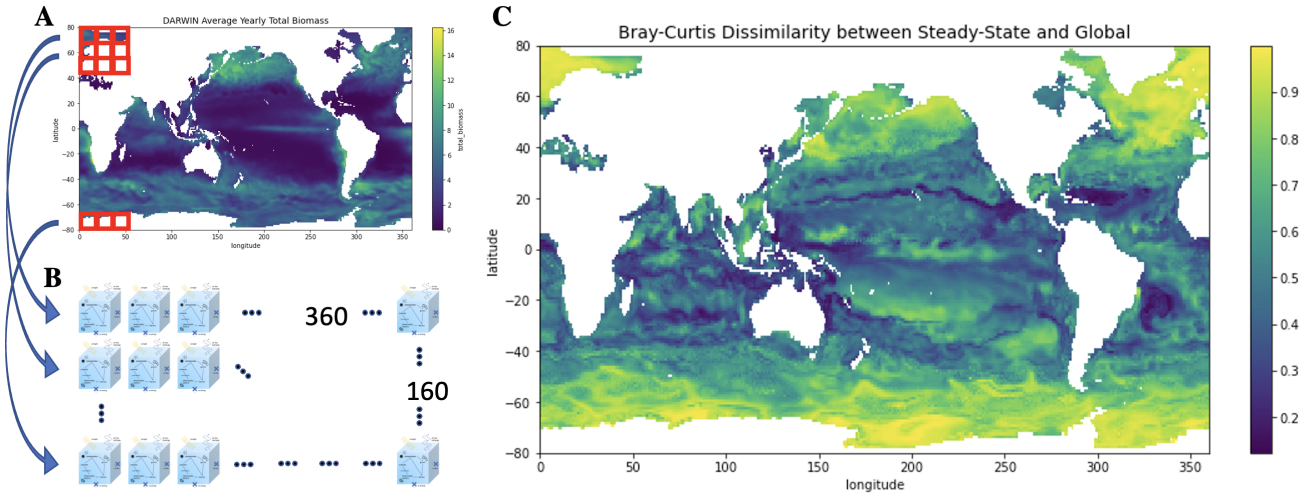

Fig. 4. Overview of Case study B: (A) The average total biomass over the course of a year in the Darwin model. Parameters from the Darwin model, such as nutrients, species abundance, sunlight, and temperature, are used to initialize a matrix of DAR1s, shown in (B). The DAR1 grid is 360x160 large, with each DAR1 "cell" corresponding to a lat/lon point in the Darwin model. The grid of DAR1s is run 10 years forward, with final "steady state" values calculated from the average of the last two years. (C) shows the Bray-Curtis Dissimilarity index of the steady state community versus the initial community. The steady state is highly dissimilar to Darwin in regions with high values (yellow), and low values (blue) indicate areas where the community structure of the two models are comparable.

tions from north of Hawaii as shown in Fig. 3 for biomass and organic matter. Temperature (24° C) and light (average light at 50m, station ALOHA) are also held constant. More details on the setup can be found in the Supplemental Information and Tutorials.

Consistent with resource ratio theory (29, 32), we see diazotroph biomass maximized for high phosphate and low nitrate conditions (Fig. 5). When the ratio of phosphate to nitrate is high, diazotrophs are able to co-exist with non-diazotrophs, which are limited by nitrate concentrations. A natural hypothesis is that the exclusion/inclusion of diazotroph biomass follow a boundary line with a slope of $1:16$. Instead, we find a curve suggesting substantially more phosphate is required to get nitrogen fixers to emerge as nitrate concentrations increase.

This can be understood by thinking about size based competitive exclusion. In steady state, we expect that two organisms of the same size only coexist if they have the same growth rate (13). Thus, when a non-diazotroph in the model competes with a diazotroph of the same size the coexistence boundary should occur when the non-diazotroph's nitrate limited growth rate is equal to the diazotroph's phosphate limited growth rate. Mathematically, approximating growth as purely Monod, we get

$$\frac{\mu_P R_P}{R_P + \kappa_P} = \frac{\mu_N R_N}{R_N + \kappa_N} \tag{5}$$

where the subscripts P and N refer to phosphate and nitrate respectively and the nitrate limited organism is the non-diazotroph. We can solve for the phosphate concentration where the growth rates are equal:

$$R_P = \frac{\mu_N \kappa_P}{\mu_P \kappa_N} \left(\frac{1}{R_N} + \frac{1}{\kappa_N}(1 - \frac{\mu_N}{\mu_D}) \right)^{-1}. \tag{6}$$

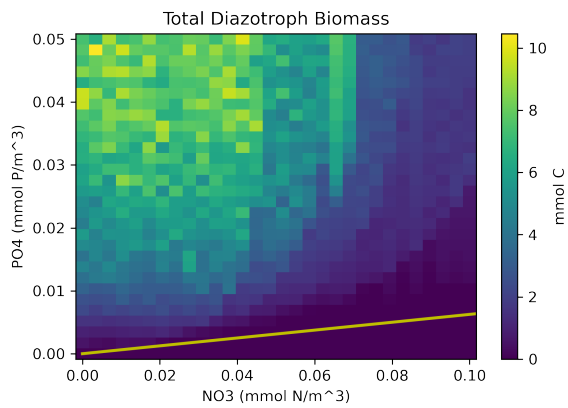

Fig. 5. Overview of Case Study C: The steady state concentration of nitrogen fixing diazotrophs is plotted as a function of phosphate and nitrate concentrations in the model. The Redfieldian stoichiometry line, P:N=1:16, is also plotted (yellow line). Increasing the supply of phosphate into the system leads to more diazotrophs whereas excess nitrate selects against nitrogen fixing organisms. The emergence of nitrogen fixers always occurs above the Redfieldian boundary.

This is consistent with what we see in Figure 5. As nitrate increases, the phosphate level required before diazotrophs grow in gets larger and larger. At a certain concentration of nitrate, there is no phosphate level sufficient to sustain diazotrophs. The predicted 1:16 line shown in Figure 5 forms a natural lower-bound. The actual emergence curve is always above this line.

Discussion

DAR1 bridges the gap between highly-simplified proof of concept models and fully coupled ocean models used for climate prediction by providing an intuitive interface for setting up and running full complexity ecosystem simulations in a 0-D environment. DAR1 accomplishes this in two ways: com-

putational efficiency and ease of use. DAR1 gains its efficiency by sacrificing physical complexity, but retains the full ecosystem of the more complex model. Ease of use comes both from a careful constraining of user choices and an external interface which groups and simplifies the modification of large numbers of Fortran files.

As with any tool, the effectiveness and impact of DAR1 hinges on whether the power it provides is correctly balanced by its ease of use. With DAR1 we have started with the full complexity and power of the fully coupled MIT Darwin model (3, 15). From this we reduced the physical complexity of the model, and the user-allowed choices. With these sacrifices, we have provided a series of advantages which we hope will allow for broad use. Among the advantages are: platform flexibility; enhanced computational speed; and ease of operation. Often, it is innovations in ease of use which allow old ideas to be used effectively by practitioners in the field. Ocean Data View (34), for example, allowed chemical oceanographers to optimally interpolate non-regular ocean transect data so that global patterns could be observed and compared effectively. The act of generating section plots from field data became a straightforward operation and allowed for field advances because of the ability to communicate patterns effectively. Modern ocean database projects seek to provide easy access to many datasets and the ability to co-localize and analyze them together (35). DAR1 seeks to fill this role in the modeling world by connecting the output and conclusions of forward ecosystem models with observations and experiments.

We are hopeful that simplified, yet fully integratable, models like this one will be useful in testing the importance of ecological interactions in the ocean, formally fitting the ecological parameters used in plankton models, and aiding in the development of AI techniques in the biogeochemical modeling sector. Immediate plans for using this infrastructure include the testing of ecological drivers for seasonal range shifts in *Prochlorococcus* (13, 36), exploring the importance of diel forcing on plankton populations (37), and exploring the importance of different grazing laws in setting community structure. Additionally, we are hopeful that structured output from simulations using DAR1 can help elucidate the situations under which statistical models for plankton populations are predictive (36, 38). The rapid technical development in marine ecosystem modeling has led to progressively more complex, realistic, models whose power is coupled with increasing difficulty of use. We believe that models like DAR1 will help extend the power of these models for new applications and for new users.

ACKNOWLEDGEMENTS

We acknowledge the Simons Foundation for funding for this study under grants: 827829, C.L.F.;

Bibliography

1. Paul G. Falkowski. The role of phytoplankton photosynthesis in global biogeochemical cycles. 39(3):235–258, 3 1994. ISSN 01668595. doi: 10.1007/BF00014586.
2. Paul G. Falkowski, Edward A. Laws, Richard T. Barber, and James W. Murray. Phytoplankton and their role in primary, new, and export production. *Springer*, 2003. doi: 10.1007/978-3-642-55844-3_5.
3. Stephanie Dutkiewicz, Pedro Cermeno, Oliver Jahn, Michael J. Follows, Anna A. Hickman, Darcy A.A. Taniguchi, and Ben A. Ward. Dimensions of marine phytoplankton diversity. *Biogeosciences*, 17(3):609–634, 2020. ISSN 17264189. doi: 10.5194/bg-17-609-2020.
4. Luigi Caputi, Quentin Carradec, Damien Eveillard, Amos Kirilovsky, Eric Pelletier, Juan J Pierella Karlusich, Fabio Rocha Jimenez Vieira, Emilie Villar, Samuel Chaffron, Shruti Malviya, et al. Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochemical Cycles*, 33(3):391–419, 2019. doi: https://doi.org/10.1029/2018GB006022.
5. Paul Tréguer, Chris Bowler, Brivaela Moriceau, Stephanie Dutkiewicz, Marion Gehlen, Olivier Aumont, Lucie Bittner, Richard Dugdale, Zoe Finkel, Daniele Iudicone, Oliver Jahn, Lionel Guidi, Marine Lasbleiz, Karine Leblanc, Marina Levy, Philippe Pondaven, Chris Bowler, Stephanie Dutkiewicz, Oliver Jahn, Marion Gehlen, Olivier Aumont, Marina Levy, Lucie Bittner, Richard Dugdale, Zoe Finkel, Daniele Iudicone, Lionel Guidi, Marine Lasbleiz, and Karine Leblanc. Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, 11(1):27–37, 1 2018. ISSN 17520908. doi: 10.1038/s41561-017-0028-x.
6. Stephanie Dutkiewicz, Michael J Follows, and Jason G Bragg. Modeling the coupling of ocean ecology and biogeochemistry. *Global Biogeochemical Cycles*, 23(4), 2009. doi: 10.1029/2008GB003405.
7. O. Aumont, C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, 8(8): 2465–2513, 2015. doi: 10.5194/gmd-8-2465-2015.
8. Ben A Ward, Jamie D Wilson, Ros M Death, Fanny M Monteiro, Andrew Yool, and Andy Ridgwell. Ecogenie 1.0: plankton ecology in the cgenie earth system model. *Geoscientific Model Development*, 11(10):4241–4267, 2018.
9. Michael J. Follows, Stephanie Dutkiewicz, Scott Grant, and Sallie W. Chisholm. Emergent biogeography of microbial communities in a model ocean. *Science*, 315(5820):1843–1846, 2007. ISSN 00368075. doi: 10.1126/science.1138544.
10. S Dutkiewicz, A. E. Hickman, O Jahn, W. W. Gregg, C. B. Mouw, and M. J. Follows. Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences*, 12(14):4447–4481, 2015. ISSN 17264189. doi: 10.5194/bg-12-4447-2015.
11. Ben A. Ward, Stephanie Dutkiewicz, and Michael J. Follows. Modelling spatial and temporal patterns in size-structured marine plankton communities: Top-down and bottom-up controls. *Journal of Plankton Research*, 36(1):31–47, 2014. ISSN 01427873. doi: 10.1093/plankt/fbt097.
12. Christopher L Follett, Stephanie Dutkiewicz, David M Karl, Keisuke Inomura, and Michael J Follows. Seasonal resource conditions favor a summertime increase in north pacific diatom–diazotroph associations. *The ISME journal*, 12(6):1543–1557, 2018.
13. Christopher L Follett, Stephanie Dutkiewicz, François Ribalet, Emily Zakem, David Caron, E Virginia Armbrust, and Michael J Follows. Trophic interactions with heterotrophic bacteria limit the range of prochlorococcus. *Proceedings of the National Academy of Sciences*, 119 (2):e2110993118, 2022.
14. Stephanie Dutkiewicz, Anna E. Hickman, Oliver Jahn, Stephanie Henson, Claudie Beaulieu, and Erwan Monier. Ocean colour signature of climate change. *Nature Communications*, 10 (1), 2019. ISSN 20411723. doi: 10.1038/s41467-019-08457-x.
15. John Marshall, Alistair Adcroft, Chris Hill, Lev Perelman, and Curt Heisey. A finite-volume, incompressible navier stokes model for, studies of the ocean on parallel computers. *Journal of Geophysical Research C: Oceans*, 102(C3):5753–5766, 3 1997. ISSN 01480227. doi: 10.1029/96JC02775.
16. C. Bertin, D. Carroll, D. Menemenlis, S. Dutkiewicz, H. Zhang, A. Matsuoka, S. Tank, M. Manizza, C. E. Miller, M. Babin, A. Mangin, and V. Le Fouest. Biogeochemical river runoff drives intense coastal arctic ocean co2 outgassing. *Geophysical Research Letters*, 50(8):e2022GL102377, 2023. doi: https://doi.org/10.1029/2022GL102377. e2022GL102377 2022GL102377.
17. Samuel T Wilson, Nicholas J Hawco, E Virginia Armbrust, Benedetto Barone, Karin M Björkman, Angela K Boysen, Macarena Burgos, Timothy J Burrell, John R Casey, Edward F DeLong, et al. Kīlauea lava fuels phytoplankton bloom in the north pacific ocean. *Science*, 365 (6457):1040–1044, 2019.
18. A. M. Kuhn, S. Dutkiewicz, O. Jahn, S. Clayton, T. A. Rynearson, M. R. Mazloff, and A. D. Barton. Temporal and Spatial Scales of Correlation in Marine Phytoplankton Communities. *Journal of Geophysical Research: Oceans*, 124(12):9417–9438, 12 2019. ISSN 21699291. doi: 10.1029/2019JC015331.
19. AE Hickman, S Dutkiewicz, RG Williams, and MJ Follows. Modelling the effects of chromatic adaptation on phytoplankton community structure in the oligotrophic ocean. *Marine Ecology Progress Series*, 406:1–17, 2010.
20. Zhen Wu, Stephanie Dutkiewicz, Oliver Jahn, Daniel Sher, Angelicque White, and Michael J Follows. Modeling photosynthesis and exudation in subtropical oceans. *Global Biogeochemical Cycles*, 35(9):e2021GB006941, 2021.
21. Julia TagBot Gael Forget, Josef Bisits. gaelforget/mitgcmtools.jl: v0.2.5. February 2023. doi: 10.5281/zenodo.7682595.
22. Gael Forget. gaelforget/climatemodels.jl: v0.2.15. November 2022. doi: 10.5281/zenodo.7332331.
23. Dirk Merkel. Docker: lightweight linux containers for consistent development and deployment. *Linux journal*, 2014(239):2, 2014.
24. D. Karl, R. Letelier, L. Tupas, J. Dore, J. Christian, and D. Hebel. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean, 1997. ISSN 00280836.
25. Claire Mahaffey, Karin M Björkman, and David M Karl. Phytoplankton response to deep seawater nutrient addition in the north pacific subtropical gyre. *Marine Ecology Progress Series*, 460:13–34, 2012.
26. David Tilman and Robert W Sterner. Invasions of equilibria: tests of resource competition using two species of algae. *Oecologia*, 61(2):197–200, 1984. ISSN 00298549. doi: 10.1007/BF00396760.
27. Robert M. May. Will a Large Complex System be Stable? *Nature*, 238(5364):413–414, August 1972. ISSN 1476-4687. doi: 10.1038/238413a0. Number: 5364 Publisher: Nature Publishing Group.
28. Robert A. Armstrong. Stable model structures for representing biogeochemical diversity

and size spectra in plankton communities. *Journal of Plankton Research*, 21(3):445–464, March 1999. ISSN 0142-7873. doi: 10.1093/plankt/21.3.445.

29. David Tilman. Resource Competition between Plankton Algae: An Experimental and Theoretical Approach. *Ecology*, 58(2):338–348, 3 1977. ISSN 0012-9658. doi: 10.2307/1935608.
30. J. Roger Bray and J. T. Curtis. An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs*, 27(4):326–349, 1957. ISSN 00129615.
31. S Dutkiewicz, B. A. Ward, J. R. Scott, and M. J. Follows. Understanding predicted shifts in diazotroph biogeography using resource competition theory. *Biogeosciences*, 11(19): 5445–5461, 2014. ISSN 17264189. doi: 10.5194/bg-11-5445-2014.
32. Ben A Ward, Stephanie Dutkiewicz, C Mark Moore, and Michael J Follows. Iron, phosphorus, and nitrogen supply ratios define the biogeography of nitrogen fixation. *Limnology and Oceanography*, 58(6):2059–2075, 2013.
33. S. Dutkiewicz, B. A. Ward, F. Monteiro, and M. J. Follows. Interconnection of nitrogen fixers and iron in the Pacific Ocean: Theory and numerical simulations. *Global Biogeochemical Cycles*, 26(1), 3 2012. ISSN 08866236. doi: 10.1029/2011GB004039.
34. Reiner Schlitzer. Interactive analysis and visualization of geoscience data with ocean data view. *Computers & geosciences*, 28(10):1211–1218, 2002.
35. Mohammad D. Ashkezari, Norland R. Hagen, Michael Denholtz, Andrew Neang, Tansy C. Burns, Rhonda L. Morales, Charlotte P. Lee, Christopher N. Hill, and E. Virginia Armbrust. Simons Collaborative Marine Atlas Project (Simons CMAP): An open-source portal to share, visualize, and analyze ocean data. *Limnology and Oceanography: Methods*, 19(7):488–496, 2021. ISSN 1541-5856. doi: 10.1002/lom3.10439. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10439.
36. Vincent Bian, Merrick Cai, and Christopher L Follett. Understanding opposing predictions of prochlorococcus in a changing climate. *Nature Communications*, 14(1):1445, 2023. doi: 10.1038/s41467-023-36928-9.
37. Ioannis Tsakalakis, Michael J Follows, Stephanie Dutkiewicz, Christopher L Follett, and Joseph J Vallino. Diel light cycles affect phytoplankton competition in the global ocean. *Global Ecology and Biogeography*, 31(9):1838–1849, 2022.
38. LR Bardon, BA Ward, S Dutkiewicz, and BB Cael. Testing the skill of a species distribution model using a 21st century virtual ecosystem. *Geophysical Research Letters*, 48(22): e2021GL093455, 2021. doi: 10.1029/2021GL093455.
39. B. A. Ward, S. Dutkiewicz, O. Jahn, and M. J. Follows. A size-structured food-web model for the global ocean. *Limnology and Oceanography*, 57(6):1877–1891, 11 2012. ISSN 00243590. doi: 10.4319/lo.2012.57.6.1877.
40. Stephanie Dutkiewicz, J Jeffrey Morris, Michael J Follows, Jeffery Scott, Orly Levitan, Sonya T Dyhrman, and Ilana Berman-Frank. Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Climate Change*, 5(11):1002–1006, 2015.
41. Emily J. Zakem, Alia Al-Haj, Matthew J. Church, Gert L. Van Dijken, Stephanie Dutkiewicz, Sarah Q. Foster, Robinson W. Fulweiler, Matthew M. Mills, and Michael J. Follows. Ecological control of nitrite in the upper ocean. *Nature Communications*, 9(1):1–13, 12 2018. ISSN 20411723. doi: 10.1038/s41467-018-03553-w.
42. Per Juel Hansen, Peter Koefoed Bjørnsen, and Benni Winding Hansen. Zooplankton grazing and growth: Scaling within the 2-2,000-μm body size range. *Limnology and Oceanography*, 42(4):687–704, 1997. ISSN 00243590. doi: 10.4319/lo.1997.42.4.0687.
43. Thomas Kiørboe. *A Mechanistic Approach to Plankton Ecology*. Princeton University Press, 5 2019. doi: 10.1515/9780691190310.

Supplementary Note 1: Supplemental Methods Description

A. The Global Biophysical Darwin Ocean Model. Darwin code is part of the MITgcm ((15), https://mitgcm.readthedocs.io/en/latest/). Code is available at https://github.com/darwinproject/darwin3, and documentation and equations at https://darwin3.readthedocs.io/en/latest/phys_pkgs/darwin.html. The model resolves the cycling of C, N, P, Fe, Si, O2 through living, detrital and inorganic pools. It is developed to be flexible in the number and type of plankton that is resolves. It follows a "trait-based" approach as each plankton is assigned a number of traits that define its size, trophic status, resource requirements, thermal responses, and for phytoplankton, light affinity. The combination of traits set the plankton's resource uptake and growth parameters. The model is highly flexible in the complexity of the ecosystem, where it can be configured to include 1 to 1000s of plankton. The plankton can be autotrophic, grazers, or remineralizers of organic matter, or some combination of trophic strategy (i.e. mixotrophic). The model can be set up for random assignment of traits (see e.g. (6, 9)), traits largely set by size alone (see e.g. (39)), traits focusing on functional type (see e.g. (40)), traits focused on trophic strategy (e.g. (39, 41)). The setup used in this manuscript follows (13) and includes 3 size classes of heterotrophic bacteria (spanning 0.4 to 0.9 μm), 23 phytoplankton (spanning from 0.6 to 110 μm, and including pico-cyanobacteria, pico-eukaryotes, calcifiers, nitrogen fixers, silicifers), 8 size classes of mixotrophs (from 6.6 to 140 μm), and 16 size classes of zooplankton (spanning 4.5 to 1,636 μm). Grazing is parameterized using a Holling II function (Holling 1972) and is size specific such that grazers can prey upon plankton 5 to 15 times smaller than themselves, with an optimal size of 10 times smaller (42, 43).

B. Coding Environments. Docker uses OS-level virtualization to make applications work seamlessly across different environments through software packages called "containers". Docker containers can be thought of as virtual machines, but they can share the OS kernel with other containers running as isolated processes, reducing the size of each individual container. A Docker "image" is the set of instructions needed to build and start up a container.

A Docker image is provided alongside the DAR1 code to make getting started as easy as possible. The Docker image, which will be used to start up a container in which DAR1 experiments run, is linked to on the github. The Docker container encapsulates the DAR1 code and all its dependencies. Notably this includes Julia and all the packages DAR1tools relies on, Fortran and its compilers, and all other libraries needed to run the MITgcm.

Using the Docker commandline tools, the DAR1 image can be accessed with the following command: `docker pull birdy1123/dar1`. The most up to date image will be linked to in the top-level README of the Github. Further instructions on how to use Docker are included in the DAR1 "getting started" documentation.

C. Code Access and Case Study Tutorials. DAR1 source code can be found in Github, at https://github.com/CBIOMES/DAR1. Installation and setup instructions are included in the documentation.

Code for the case studies can be found at https://github.com/CBIOMES/DAR1/tree/main/case_studies, with step-by-step instructions on the documentation site. All seed files used in the case studies, generated from global Darwin runs, are included within the Docker.