Title

Ocean biogeochemical modelling

Journal

Nature Reviews Methods Primers, 2(1)

ISSN

2662-8449

Authors

Fennel, Katja

Mattern, Jann Paul

Doney, Scott C

et al.

Publication Date

2022

DOI

10.1038/s43586-022-00154-2

Ocean Biogeochemical Modeling

Katja Fennel[1,*], Jann Paul Mattern[2], Scott C. Doney[3], Laurent Bopp[4], Andrew M. Moore[2], Bin Wang[1], Liuqian Yu[5]

[*] Corresponding author

1 Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
2 Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, California, USA
3 Department Environmental Sciences, University of Virginia, Charlottesville, Virginia, USA
4 Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, CNRS, Ecole normale supérieure, Paris Sciences Lettres Université, Paris, France
5 Earth, Ocean and Atmospheric Sciences Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

Author contributions
Introduction (KF, LB); Experimentation (SCD, KF, JPM, AMM, BW); Results (KF, JPM); Applications (SCD, LY, LB, JPM, KF); Reproducibility and data deposition (SCD, KF); Limitations and optimizations (KF, SCD); Outlook (KF); Overview of the Primer (KF).

Abstract: Ocean Biogeochemical Models describe the ocean's circulation, its physical properties, and its biogeochemical properties and their transformations with the help of coupled differential equations. Numerical approximations of these equations allow simulation of the dynamic evolution of the ocean state in realistic global or regional spatial domains from years to centuries. We explain the process of model construction and the main characteristics, advantages and drawbacks of different model types ranging from the simplest Nutrient-Phytoplankton-Zooplankton-Detritus or NPZD model to the complex biogeochemical models used in Earth System Modelling and climate prediction. We describe metrics for model-data comparison commonly used in model assessments and how the models can be informed by observations via parameter optimization or state estimation, the two main methods of data assimilation. Examples illustrate how these models are used for various practical applications ranging from carbon accounting, ocean acidification, and ocean deoxygenation to observing system design. Access points are provided enabling readers to engage in biogeochemical modeling in the form of hands-on code examples and a comprehensive list of publicly available models and observational data sets. We make recommendations for best practices in model archiving and lastly discuss the models' current limitations and anticipated future developments and challenges.

[H1] Introduction

Ocean biogeochemical models (OBMs) are defined here as spatially explicit models that consist of a component describing the ocean's temperature and salinity distributions and its circulation

(including wind- and density-driven currents and wind-, convectively, and eddy-driven mixing) and a component that describes the transformations of biogeochemical constituents contained in seawater (typically nutrients, functional plankton groups **[G]**, non-living organic matter, dissolved gases, and parameters of the inorganic carbon system; Figure 1). Both components consist of numerical codes approximating systems of partial differential equations. OBMs can be regional or global in terms of their geographic scope. They can themselves be a component of a larger model, e.g., in Earth System Models (ESMs) where an OBM is coupled to a model of the atmosphere and land biosphere, or self-contained where information from atmosphere and land is imposed. An OBM typically is run forward in time, starting from a defined initial condition **[G]**, and simulates the evolution of its state variables **[G]** subject to external forcing **[G]** (e.g., wind, atmospheric variables such as air temperature and pCO_2, riverine nutrient and freshwater inputs, etc.). OBMs can be run as hindcasts (describing past conditions), as nowcasts (aiming to describe the current state of the ocean), or as forecasts or projections **[G]** (intended to inform us about possible future ocean conditions).

OBMs emerged in the 1990s as a common tool to address the needs of two distinct communities with different scientific objectives. One was interested in plankton ecology and sought to explain and predict seasonal phytoplankton dynamics with the help of marine plankton models. The roots of these models go back to Gordon Riley[1] with further developments in the 1980s and 90s[2–4]. When computers became more widely available, these models were coupled to models of ocean circulation. A regional model of the North Atlantic[5] was probably the first ocean circulation model with explicit representation of plankton dynamics. The other community was interested in the role of the ocean as a sink of anthropogenic carbon. Building on concepts established by Roger Revelle[6], early ocean carbon cycle models did not include an explicit representation of plankton[7–9]. In the seminal work by Ernst Maier-Reimer[10,11], models of carbon cycling and plankton dynamics were combined and integrated into global ocean circulation models. This type of model is now a widely used tool for ocean ecologists and biogeochemists alike and has evolved to include diverse functional plankton groups and multiple distinct elemental cycles.

In this article, we describe the process of OBM construction (Experimentation), review and illustrate methods and metrics for evaluating models against observations (Experimentation, Results), and introduce approaches for combining models and observations (Experimentation, Results). The latter methods, collectively referred to as data assimilation, include optimization of model parameters **[G]** (e.g., plankton growth and grazing rates, rates of organic matter sinking and remineralization) and of the model state (i.e., the derivation of the most likely ocean state given our mechanistic understanding of the system in form of the model equations and the available observations). We describe several important applications of OBMs to illustrate their breadth and utility (Applications) and recommend best practices for archiving model code and output, and for conducting intercomparisons (Reproducibility and data deposition). The article concludes with a discussion of current limitations of OBMs and their applicability (Limitations and optimizations) and anticipated new developments and challenges (Outlook).

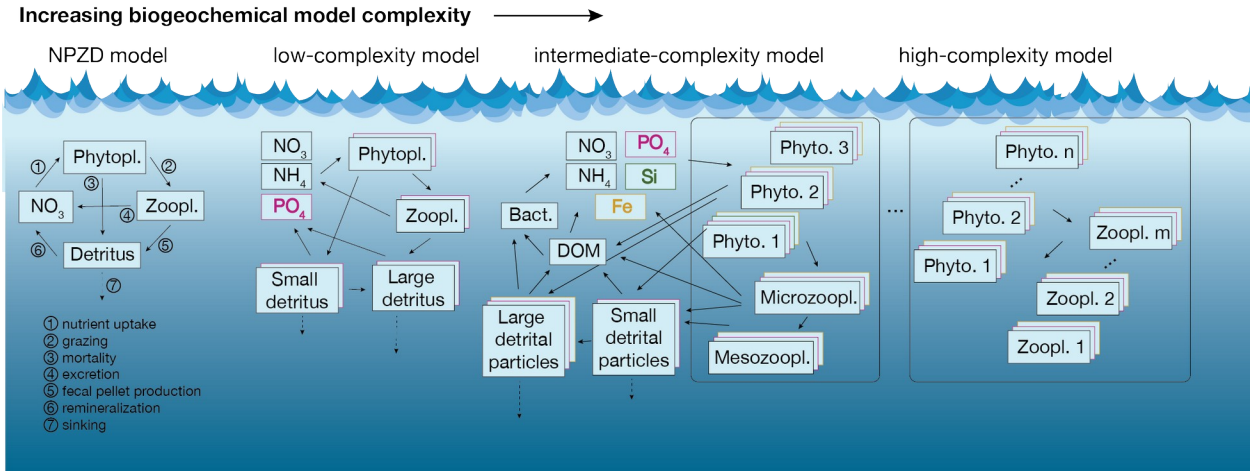


Figure 1: Schematic representation of the state variables and biogeochemical transformations across a range of OBMs with increasing complexity from left to right. Squares represent state variables. Solid arrows represent selected transformations between state variables. Dashed arrows illustrate vertical sinking of particles. The NPZD model is the simplest with only four state variables and one nutrient currency (typically nitrogen). In the NPZD example, all transformations are indicated by arrows and labelled. In a typical low-complexity model, multiple nutrients and more than one nutrient currency are included. Keeping track of multiple nutrient currencies necessitates multiple state variables for some of the functional groups and particulate pools (indicated by the stacked, color-coded boxes). In the low-to high-complexity examples only the general direction of transformations is indicated for sake of simplicity. In practice there are many more transformations between state variables, each represented by a parameterization that requires at least one, but typically more, biogeochemical model parameters. As complexity increases from left to right, the number of plankton groups and organic matter pools increases with distinctions by particle size and the addition of Dissolved Organic Matter (DOM) and bacteria. For the high-complexity model, only the increase in functional phytoplankton and zooplankton groups is represented schematically because the inorganic and non-living organic pools are similar to the intermediate complexity model.

[H1] Experimentation

[H2] Model construction

[H3] Biogeochemical equations

The biogeochemical dynamics at the core of an OBM commonly are cast as a system of coupled partial differential equations describing the rate of change of a set of state variables, *C,* that represent concentrations of nutrients, the biomass of functional plankton groups, etc.[12,13]. The common form of these equations is:

$$\frac{\partial C}{\partial t} = physics + bgc_{sms} \quad (1)$$

where *physics* includes all advective and dispersive transport processes affecting the concentration of *C* and *bgc_sms* includes all local sources and sinks due to biogeochemical transformations, air-sea gas exchange and atmospheric deposition, sediment-water exchange, and any transport that is not the results of ocean circulation, e.g., vertical sinking of organic matter. Biogeochemical state variables are specified as concentration of some element, often nitrogen. Other elements such as carbon, phosphorus, silicon, or iron can also be included using either fixed or variable elemental stoichiometry to relate the different state variables.

One of the lowest-order, but complete biogeochemical models is the so-called nutrient-phytoplankton-zooplankton-detritus or NPZD model, which describes the concentrations of these four variables in a homogenously mixed volume or box. It is obtained by neglecting the physical terms (effectively creating a 0-dimensional box model), which simplifies the equations to ordinary differential equations (i.e., dC/dt), and including only four state variables. Assuming a closed system, the terms on the right-hand-side of the equations reflect transformations between the state variables:

$$\frac{dN}{dt} = -uptake + remineralization$$

$$\frac{dP}{dt} = uptake - grazing$$

$$\frac{dZ}{dt} = grazing - excretion$$

$$\frac{dD}{dt} = excretion - remineralization$$

Note that the four coupled NPZD equations above are mass conserving in that loss from one variable to another is balanced by a corresponding gain in the latter. Also, the four equations make up a coupled system of equations because the right-hand-side terms are dependent on multiple state variables.

The next step in modeling the NPZD system is to specify the functional form and parameters for each of the biogeochemical transformations, which are referred to as parameterizations. They are defined using conceptual understanding (also referred to as a-priori knowledge) from laboratory experiments, field studies, and biological theory[14]. For example, the grazing term of zooplankton consuming phytoplankton is a function of *P*, *Z* and perhaps temperature, *T*, i.e. *grazing* = f (*P,Z,T*,…). Below are three example grazing parameterizations:

$$grazing = gZ\frac{P}{K_P}$$

$$grazing = gZ\frac{P}{P+K_P}$$

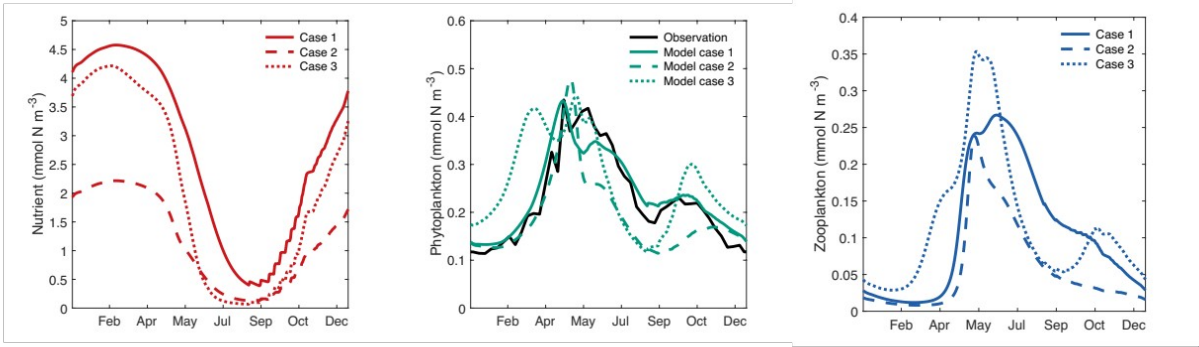
$$grazing = gZ\frac{P^2}{P^2+K_P^2}$$

where g (d^{-1}) is a rate parameter and K_P (same units as P) is a saturation parameter. All three parameterizations are common in ecological modelling for capturing consumer-resource interactions and reflect distinctly and often subtlety different biological dynamics[15]. The first parameterization assumes that the grazing rate increases linearly with the prey concentration (P) and the latter two assume it saturates at high concentrations of P (i.e., $P>>K_P$). The P^2 terms in the third parameterization result in reduced grazing at very low P concentrations. The choice of the functional form of the grazing parameterization is typically made based on theoretical arguments and considerations about the numerical model stability. The parameters for the grazing parameterization can be determined by dilution experiments for some zooplankton species where phytoplankton loss rates are measured across a range of prey dilution levels, recognizing however, that single species experiments in the lab do not necessarily translate to diverse natural communities. Similar decisions on the functional forms and parameter values as described here for grazing have to be made for all other parameterizations in the NPZD model. An example of a complete, vertically resolved NPZD model is provided in Box 1.

Box 1: Simple Matlab code of the one-dimensional NPZD model described in ref.[16] is available on github[17]. The model is representative of a station in the sub-polar North Atlantic Ocean. After running the default simulation, which results in good agreement of the simulated phytoplankton concentration with satellite observations, the reader is encouraged to increase/decrease the initial nitrate concentration, the maximum phytoplankton growth rate, the maximum zooplankton grazing rate, and the latitude to explore the model's dependencies on these parameters.

Shown below are the simulated surface concentrations of nitrate, phytoplankton, and zooplankton variables in the second year of the simulation for the default parameter set (case 1), for a 50% decrease in the initial nitrate concentration (case 2), and for a doubling of the maximum phytoplankton growth rate. In case 1, the surface phytoplankton concentrations agree well with satellite-based observations. The decrease in initial nitrate in case 2 has only a small effect on the timing and amplitude of the spring phytoplankton bloom but leads to much smaller phyto- and zooplankton concentrations in summer and fall than in case 1. The doubling of the maximum phytoplankton growth rate in case 3 leads to a much earlier spring bloom initiation, a larger fall phytoplankton bloom, and larger spring and fall peaks in

zooplankton than in case 1.



Most OBMs in current use are extensions of the basic NPZD framework but have more complex biogeochemical model components. Additional state variables include multiple nutrients (nitrate, ammonium, phosphate, silicate, and dissolved iron), multiple phyto- and zooplankton functional groups, and dissolved gases and related properties (e.g., oxygen, inorganic carbon, alkalinity). Multiple nutrients are needed to address spatial and temporal switches between limiting nutrients and unique requirements by some functional phytoplankton groups (e.g., diatoms). Multiple plankton variables are included to account for the biogeochemically distinct roles played by different size classes and functional groups[18]. For example, diatoms have a unique requirement for silicate and contribute significantly to biological carbon export, coccolithophores produce calcium carbonate as shells and affect vertical carbonate transport and remineralization at depth, and diazotrophs fix gaseous nitrogen turning it into bioavailable forms. Including the inorganic carbon cycle is crucial for any OBM used for climate studies[19]. This requires inclusion of state variables for dissolved inorganic carbon and alkalinity (unless alkalinity can be inferred from other state variables typically salinity). Knowledge of these two properties enables calculation of the other carbonate system properties, including the partial pressure of CO_2 (pCO_2), which is required to parameterize air-sea gas exchange, and the pH which is of considerable interest given concerns about ongoing ocean acidification. Another common state variable in OBMs is oxygen because of its relevance for climate and ecosystem health. Oxygen minimum zones in the open ocean are sites of trace gas production and nitrate loss via denitrification[20]. Low oxygen concentrations (hypoxia) or a complete absence (anoxia) have deleterious impacts on ecosystems.

Although virtually all OBMs that are used for biogeochemical and climate studies follow the approach of defining a moderate number of functional groups, there are alternative approaches. One family of models initializes simulations with many dozen to a hundred or more phytoplankton state variables with a randomly chosen or specifically crafted size structure and physiological parameters and then allows competition within the simulation to sub-select

regional and seasonal plankton communities[21,22]. Another family of models uses allometric relationships to represent a continuum of plankton size-classes to simulate grazing relationships and distinct trophic interactions in different marine ecosystems[23,24]. These two approaches move in the direction of representing more of the complexity inherent in natural plankton communities. Others have moved in the opposite direction by drastically reducing the number of biogeochemical state variables to only four[25].

Model uncertainties enter the biogeochemical equations of an OBM from several sources. The models have many parameters which are not well known or easily quantifiable[26]. Even the parameters that can be determined experimentally may not well represent real world communities in the field. Furthermore, model parameters are not independent for coupled differential equations, and system level uncertainties can arise because of dynamical interactions between state variables. Parameter optimization aims to address this issue but depends critically on the availability of a broad suite of observations. Even more challenging are uncertainties in the choice of model structure and model parameterizations. Coupling of the biogeochemical equations with the ocean circulation results in additional sources of errors. Careful validation of OBMs to evaluate whether they are fit-for-purpose, be it a specific scientific question or applied purpose, is thus an integral part of model developments and application.

[H3] Coupling with ocean circulation

In an OBM, the transformations between biogeochemical state variables described above are connected to their advective and dispersive transport (i.e., ocean circulation) by partial differential equations of the general form given by equation 1. This equation can be rewritten as follows for each state variable C

$$\frac{\partial C}{\partial t} = -u \cdot \nabla_3 C + \nabla_2 \cdot k_H \nabla_2 C + \frac{\partial}{\partial z}(k¿¿V \frac{\partial C}{\partial z}) + bgc_{sms}(2)¿$$

where the first term on the right-hand side represents the advective transport of constituent C ($\boldsymbol{u}$ is the fluid velocity vector), the second and third terms represent dispersion in the horizontal and vertical directions, respectively, and the last term refers to the biogeochemical sources and sinks of C. The parameters k_H and k_V are the horizontal and vertical dispersion coefficients and $\nabla_3 = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ and $\nabla_2 = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ are three- and two-dimensional operators. The combination of the first three terms on the right-hand side is referred to simply as *physics* in equation 1. Since physical transport processes operate in all three spatial directions, equation 2 is three-dimensional in space and includes partial derivatives with respect to time, t, and the three spatial dimensions, x, y, and z. In addition to an equation of this form for each biogeochemical variable, an OBM includes partial differential equations for the physical state variables (incl. temperature, salinity, and velocity), as well as parameterizations for horizontal and vertical dispersion coefficients which can vary in space and time. For detailed descriptions of the physical model equations, we refer to refs.[13,27].

Except for a few highly idealized cases (e.g., when considering only one spatial dimension or a circular or rectangular two-dimensional domain with homogenous initial conditions and constant forcing), the solution to these equations cannot be obtained analytically and must be approximated numerically instead. Most commonly, the equations are discretized in time (using finite timesteps Δt) and space (on a three-dimensional grid representing the model domain) with the help of finite differences. Essentially the derivatives in the differential equations are replaced by finite difference approximations, e.g., $\partial C/\partial t$ and $\partial C/\partial x$ become $\Delta C/\Delta t$ and $\Delta C/\Delta x$, respectively. This results in a system of prognostic equations that include only basic arithmetic operations on defined quantities which can be carried out on a computer. There are a lot of subtle issues and a multitude of options when defining spatial grids, the finite difference discretization of the equations on these grids, and time stepping[28,29], which explains the large diversity of ocean circulation models in current use.

As the name suggests, finite difference approximations are not exact solutions to the OBM equations; they only approximate the solutions. The accuracy of these approximations depends on the chosen difference scheme, the size of the timestep, Δt, and the spatial resolution, $\Delta x, \Delta y$, and Δz. It is generally desirable to use the longest timestep and finest spatial resolution possible given available computational resources. Although computing power has increased at a remarkable pace over the past two decades, the computational cost of running realistic OBMs is so demanding that tradeoffs between domain size, resolution, and integration time **[G]** always must be considered. Note that doubling the horizontal resolution Δx and Δy necessitates the timestep Δt to be shortened to a quarter of its previous value which leads to an increase in overall computation time by a factor of 16.

Earth System Models have a spatial resolution on the order of 1° (~100 km, Figure 2) and are typically integrated for several hundred years. Given their spatial resolution, the models are unable to capture a range of important bathymetric and circulation features including continental shelf edges, mesoscale eddies and currents, and river plumes. Regional models have a finer spatial resolution (Figure 2), on the order of single to tens of kilometers, but they have much shorter integration times (months to decades). The models' pros and cons can be illustrated using the northwest North Atlantic Ocean as an example. The broad, passive-margin shelf in this region, located at the confluence of two large-scale current systems (the Gulf Stream and the Labrador Current), supports economically and culturally important fisheries that are particularly vulnerable to warming and ocean deoxygenation[30,31]. A defining circulation feature in this region is the shelfbreak current (a branch of the Labrador Current system) which effectively isolates shelf water from adjacent open-ocean water leading to distinct properties and long residence times[32]. Due to their low resolution, global models typically lack the shelfbreak current and cannot reproduce these features (e.g., ref[33]); as a result, they do not well reproduce biogeochemical properties in this region[34]. A recent effort to increase the resolution of a global OBM to rival that typical in regional models showed that properties are simulated much more realistically when the shelfbreak current is properly represented[35]. However, the computational

effort was so large with this model that its integration time was limited to 100 years, only the highly simplified 4-variable model of biogeochemical model was included[25], and the model cannot be run routinely given present computational resources. The obvious drawback of a high-resolution, regional model is that its integration time is limited to decades and that atmospheric and larger-scale ocean forcing must be specified instead of evolving internally as is the case in ESMs.

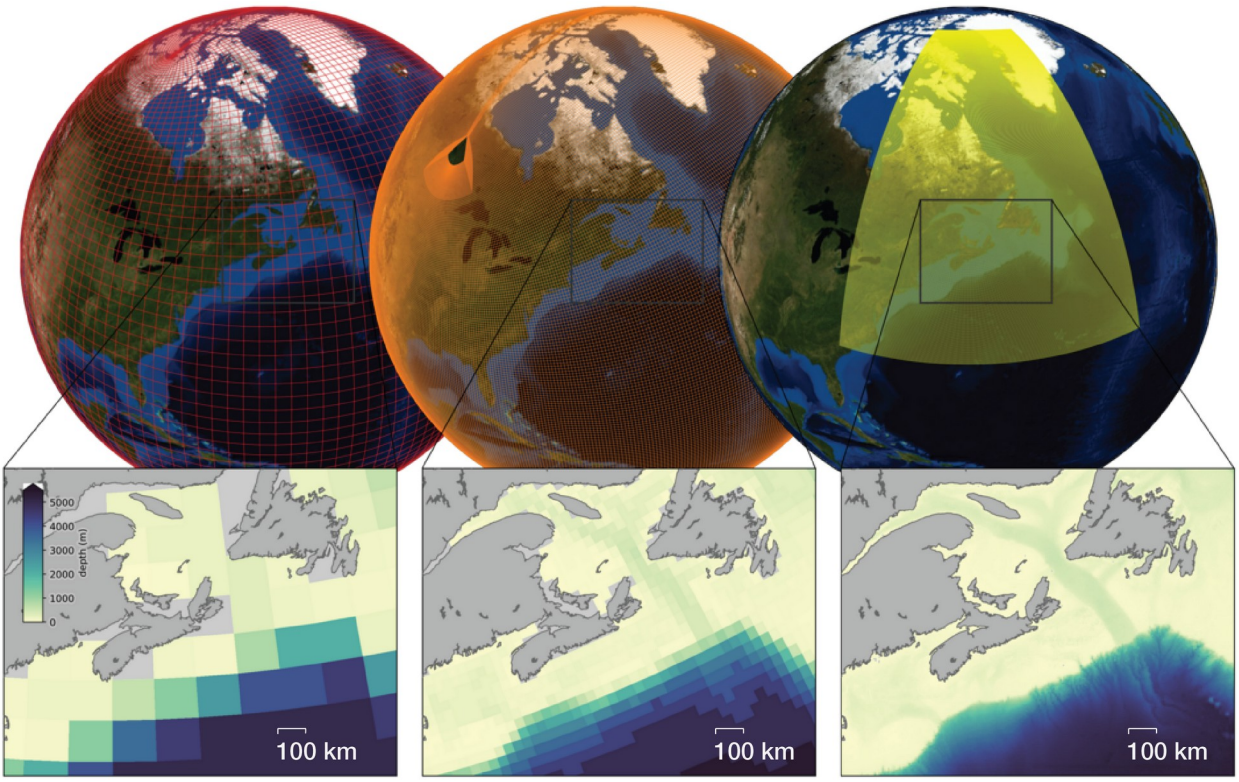


Figure 2: Illustration of typical horizontal resolutions (globes) and bathymetries (insets) in global and regional models. The left and middle globes show the grids of two global models that are part of the Coupled Model Inter comparison Project's 6th Assessment Round (CMIP6). Left is the global IPSL-CM5A-LR grid with a horizontal resolution of about 130 km in the region shown in the corresponding inset. The middle globe shows the MPI-ESM1-2-HR grid with a resolution of about 30 km in the inset region. The right globe and inset show the domain extent and horizontal resolution of a regional OBM for the northwest North Atlantic and Labrador Sea with a resolution of about 5 km in the inset region.

Running an OBM involves integration forward in time from defined initial conditions for each state variable and is subject to external forcing and boundary conditions at the edges of the model domain. As initial conditions, the distributions of temperature and salinity must be

prescribed while the model may start from rest (i.e., initial velocities are zero). External forcing of the ocean circulation component of an OBM includes air-sea fluxes of momentum (i.e., wind forcing), heat, and freshwater (precipitation minus evaporation, sea-ice formation and melt), and freshwater inputs from rivers. Examples of boundary conditions are that fluid flow cannot be normal (or at a right angle) to the coast and may be required to vanish at the coast (the so-called "no slip" boundary condition). Regional models typically have some edges that do not coincide with coastlines. Flow and the associated transport of seawater constituents across these so-called open boundaries must be specified for regional models, which is one of their drawbacks.

As for the ocean circulation component, initial and boundary conditions must be specified for the biogeochemical state variables. The distributions of nutrients, dissolved gasses, alkalinity, and long-lived organic pools (e.g., long-lived dissolved organic matter) should be prescribed as accurately as possible for initial and open boundary conditions, while pools with fast turnover times like plankton groups and reactive detrital pools can be set to small positive numbers and will adjust quickly during model spin up **[G]**. Additional boundary conditions for the biogeochemical model component include nutrient and organic matter concentrations in river inputs, the mole fractions of gases in the atmosphere, atmospheric deposition, and exchange fluxes across the sediment-water interface.

[H2] Combining models and observations

The overarching goal of combining models and observations, a process referred to as data assimilation, is to achieve the best possible representation of past, current, or future ocean states. Methods to achieve that combine the a priori knowledge **[G]** of the ocean state and its processes that is contained in an OBM with observations. Two distinct applications for OBMs are parameter optimization **[G]** and state estimation **[G]**. Parameter optimization is aimed at addressing systematic biases in models that arise from inaccurate parameter values and initial and boundary conditions. State estimation typically assumes the model is unbiased and aims to correct random errors, i.e., deviations between the observed and simulated ocean state that are due to stochastic processes, e.g., the ocean's eddy field or larger-scale variations like the North Atlantic Oscillation and El Niño-Southern Oscillation. Both applications can be realized through variational methods **[G]** or sequential methods ***[G]*** (see Box 2). Typically, implementation of an OBM includes parameter estimation initially to remove biases within the model and is potentially followed by state estimation to minimize random errors.

[H3] Parameter estimation

Variational data assimilation derives from the mathematical field 'calculus of variations,' which uses small variations in the inputs of functions to find their minima or maxima. A long-standing application of this approach to OBMs is parameter optimization[36–38], where poorly known model parameters are varied systematically to minimize the misfit between observations and the model

equivalents of these observations across the whole integration period simultaneously, thus resulting in better agreement between model and observations (Figure 3). The misfit is measured by the so-called cost function **[G]**, typically of the form

$$J(p)=\frac{1}{N}\sum_{i=1}^{N} w_i\left(y_i-\hat{y}_i(p)\right)^2 \text{ or } J(\mathbf{p}) = (\mathbf{y} - \mathbf{Hx}(\mathbf{p}))^{\mathrm{T}}\,\mathbf{R}^{-1}\,(\mathbf{y} - \mathbf{Hx}(\mathbf{p})),$$

where **p** is a vector of the parameters to be optimized (also referred to as the control vector **[G]**), **y** is a vector of the available observations, $\hat{y}=H\,x(p)$ is a vector of the model equivalents to these observations obtained by mapping the model state **x**(**p**) onto the observation vector **y** using the linear operator **H**, and w_i and **R** contain the weights with which each observation contributes to the cost function. Typically, **R** is assumed to be diagonal where the weights are the inverses of the variances or based on the observation error for each observation type. **R** is thought of as the covariance matrix of the deviations between model and observations and referred to as the observation error covariance matrix. In practice, assumptions about these weights must be made.

Solution of this minimization problem yields the optimal parameters **[G]** and can be obtained with iterative gradient descent methods (e.g., conjugate gradient search[13]) or stochastic approaches such as simulated annealing[36] and evolutionary algorithms[39–41]. Parameter optimization is widely applied to OBMs because they typically have many poorly known, difficult-to-determine, and application-specific parameters that govern the biogeochemical transformations at the heart of the model. The method aims to extract information about these biogeochemical transformations that is inherent in available observations to inform or constrain the poorly known parameters.

In practice, the success of parameter optimization depends on whether the available observations contain enough information to constrain the parameters to be optimized[42]. For example, chlorophyll observations may be abundant and useful for informing chlorophyll-related parameters such as the total phytoplankton growth rate but may not contain much information about grazing, remineralization, and species-specific growth rates; those would remain poorly constrained after optimization using only chlorophyll or chlorophyll and nutrient observations[37]. Because observational datasets are often limited in terms of the biogeochemical properties available, this is a common problem for OBMs, also referred to as the underdetermination problem[43]. A closely related problem is that of interdependent, correlated, or non-unique parameters, which arises when different combinations of parameters yield the same result. For example, a reduction in plankton mortality and an increase in plankton growth may yield the same change in biomass, and the available observations may not provide enough information to distinguish between multiple plausible combinations without further information. An example of underdetermined and interdependent parameters is given in Box 3. An a posteriori error **[G]** analysis provides insight into interrelated and poorly constrained parameters given the specific optimization problem at hand[37,42].

Parameter optimization is routinely used in biogeochemical modeling[44,45]. It is often an integral step in model development and, if applied systematically for different model structures, can guide model construction (e.g., refs.[46,47]). Parameter optimization and state estimation are sometimes combined[48,49].

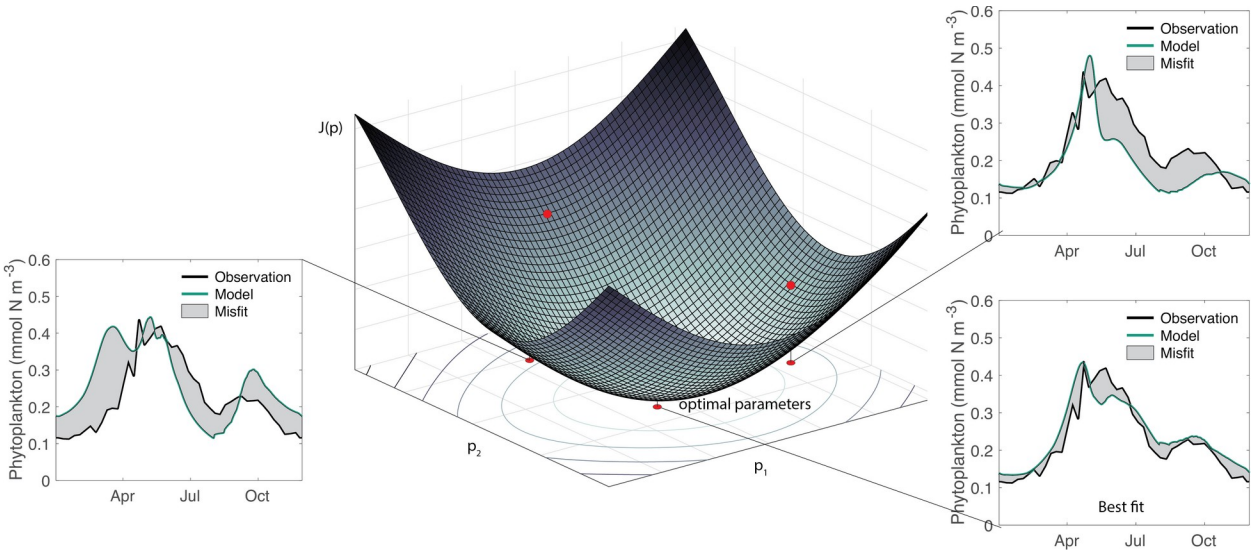


Figure 3: Schematic representation of a two-dimensional cost function. The cost function measures the misfit (indicated as the gray area in the three insets) between observations and their model equivalents in parameter space. The optimal parameters correspond to the minimum of the cost function and produce the best fit between model and observations.

Box 2: General data assimilation machinery

In data assimilation, an optimization problem is solved where the initial *control vector*, also referred to as the background or initial guess, is updated such that the misfit between available observations and the model equivalents of these observations, and, in some cases, the misfit between the initial and updated control vector is minimized in a least-squares sense.

The solution to the optimization problem can be written as

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\,[\mathbf{y} - H(\mathbf{x}_b)], \quad (1)$$

where $\mathbf{x}_b$ is the initial control vector, $\mathbf{y}$ is a vector containing the observations to be assimilated, $\mathbf{x}_a$ is the optimized control vector, H is a nonlinear operator that contains the ocean model and maps the initial control vector onto the observations, and $\mathbf{K}$ is referred to as the Kalman gain matrix. This equation applies to parameter and state estimation but is more intuitive for the latter in that $[\mathbf{y} - H(\mathbf{x}_b)]$ represents a vector of observation-model misfits that is multiplied by the matrix $\mathbf{K}$ which projects these misfits onto the model's state producing the increments needed to obtain the optimal ocean state $\mathbf{x}_a$.

The optimal solution $\mathbf{x}_a$ can be obtained by calculating or approximating $\mathbf{K}$, or alternatively, by

solving an equivalent minimization problem without explicit evaluation of **K**. We denote the true control vector (i.e., the desired solution) by $\mathbf{x}_t$. Then $\mathbf{x}_b - \mathbf{x}_t$ represents the so-called background errors (deviations between the initial control vector and the truth), and **B** shall denote their covariance matrix. In the case of a linear model, H is also linear and denoted by **H**. Observation errors are $\mathbf{y} - \mathbf{H}\mathbf{x}_t$ and their covariance matrix shall be **R**. Assuming that the background and observation errors are Gaussian and unbiased, and that the cross-correlation between background and observation errors is zero, the algebraic form of the gain matrix that provides the optimal analysis $\mathbf{x}_a$ is

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \quad (2)$$

Alternatively, $\mathbf{x}_a$ can be obtained by minimizing the cost function

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\, \mathbf{B}^{-1}\, (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^{\mathrm{T}}\, \mathbf{R}^{-1}\, (\mathbf{y} - \mathbf{H}\mathbf{x}). \quad (3)$$

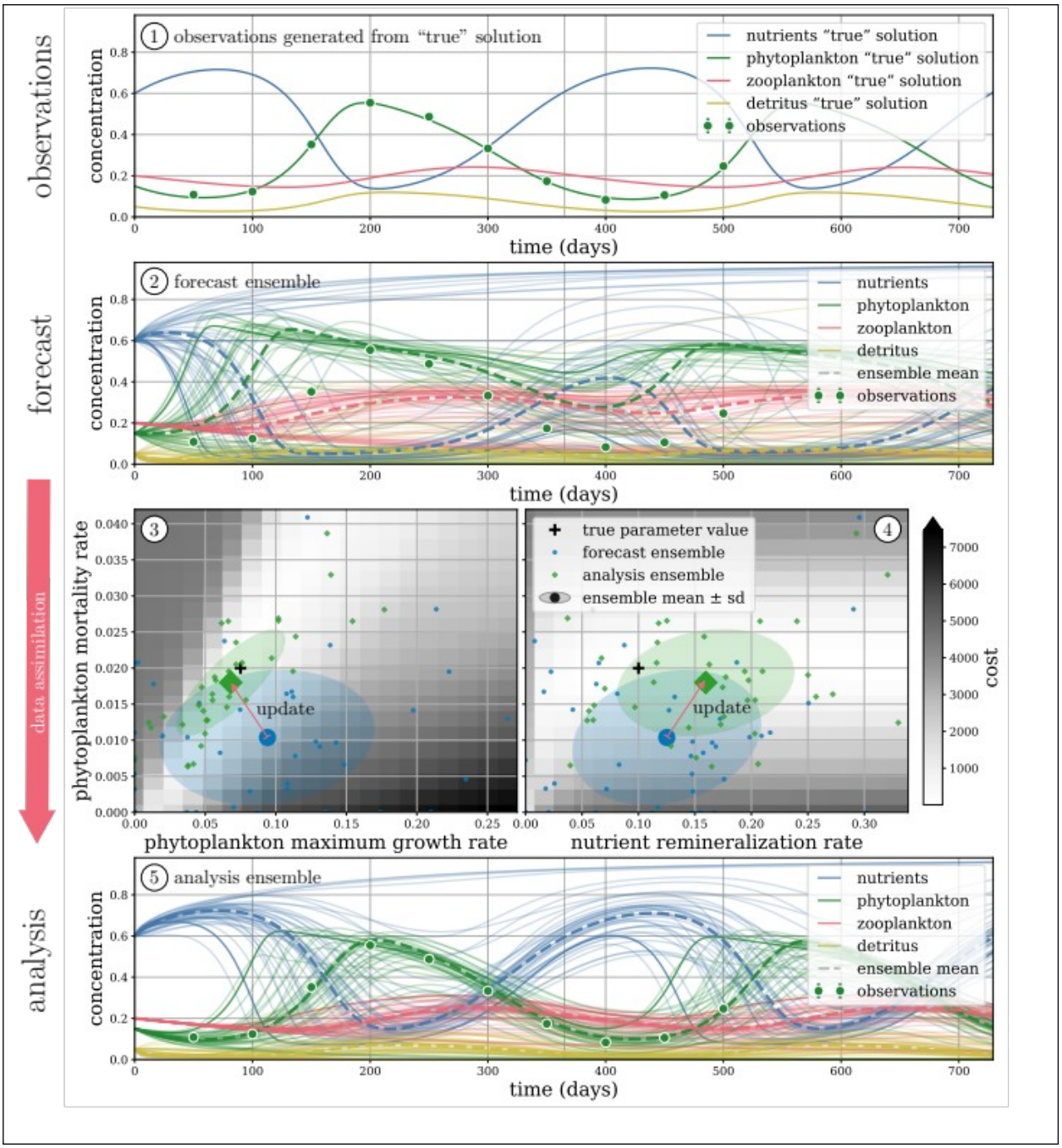
Equation (3) follows from Bayes' theorem in that J is the argument of the Gaussian conditional probability function of **x** given **y**, where the first and second terms on the right-hand side of (3) are the arguments of the Gaussian distribution functions for errors in the background and error in the observations, respectively. Thus, identifying the **x** that minimizes (3) is equivalent to maximizing the conditional probability $P(\mathbf{x}|\mathbf{y})$. The assumption of a linear model is not strictly necessary for obtaining $\mathbf{x}_a$ by minimization of the cost function (3). For non-linear models the term **Hx** can be replaced by $H(\mathbf{x})$.

Significant effort has been dedicated to developing methods to obtain, or at least approximate, these solutions for realistic models with large control vectors. OBMs are highly non-linear, have large state vectors and are computationally expensive. They also violate some of the underlying assumptions (e.g., Gaussian error distributions). Furthermore, the background and observation error distributions are not well known and are not necessarily unbiased. Sequential methods follow the path of estimating the gain matrix **K**, while variational methods avoid the explicit calculation of **K** and instead seek to minimize $J(\mathbf{x})$, i.e., maximize $P(\mathbf{x}|\mathbf{y})$.

Box 3: Python and Matlab codes that perform parameter optimization for a zero-dimensional (single box) NPZD model are available on github[50]. These code examples perform twin experiments using the Stochastic Ensemble Kalman Filter (SEnKF). The default setting (results shown below) helps illustrate the effect of both interdependent and underdetermined parameters. When only phytoplankton observations are available, there is a tight interdependence between the phytoplankton growth and mortality rates (i.e., a combination of low growth and low mortality rates can yield a similar fit between model and observations as a combination of high growth and high mortality). Likewise, the phytoplankton observations contain little information about the nutrient remineralization rate, which is not improved by

assimilation in this example and remains underdetermined. The reader is encouraged to experiment with different combinations of observation types and include different sets of parameters in the assimilation to further explore these issues.

In the default example, three parameters are estimated as follows. First, ten synthetic phytoplankton observations are generated from a model simulation with known parameters (the "truth" in panel 1). The prior estimate of the parameters (panel 2) results in a large spread of the forecast ensemble in state space. The "true" parameters and the parameters of the forecast ensemble are shown as black plusses and blue dots, respectively, in panels 3 and 4 overlain over the misfit between model and synthetic observations. The means of the forecast and analysis parameter ensembles are shown as big blue dot and green square respectively in panels 3 and 4. Assimilating the data moves the mean parameter estimate closer to the "true" values in parameter space for the phytoplankton parameters (panel 3), but farther from the true values for the nutrient remineralization rate (panel 4). The analysis ensemble, shown in panel 5, envelops the observations much more tightly than the forecast ensemble.



[H3] State estimation

State estimation, where a model's state variables are modified to reduce the discrepancy between model and observations, is typically applied sequentially by alternating *forecast steps*, where the model runs forward for a defined time window (usually a few days), followed by *update* or *analysis steps,* when newly available observations are used to update the model's state (Figure 4, Box 4). One of the most widely used and robust sequential data assimilation techniques is the

Ensemble Kalman Filter (EnKF)[51]. The EnKF as well as its precursors and its many variants apply equation 1 in Box 2.

The EnKF is based on the Kalman-Bucy Filter (KF)[52] which yields the best possible estimate in a least-squares **[G]** sense when the model is linear and the mean and covariances of the model state and observations are fully characterized[53]. The KF sequentially projects the model state, and its mean and covariances forward in time using the model (ideally linear) followed by a Bayesian update of the model state and its mean and covariances informed by the newly available observations. However, application of the KF to OBMs is hampered by their nonlinearity and the large size of their state vector (typically on the order of 10^8 and larger) which would require storing and modifying a prohibitively large covariance matrix. These issues spurred development of the Extended Kalman Filter (ExtKF)[54] for use with nonlinear models, and extensions thereof, such as the Singular Evolutive Extended Kalman filter (SEEK)[55], which reduces the computational requirements for evolving the covariance matrix. However, propagation of the covariance matrix still requires linearization of the model which can lead to bad approximations for highly nonlinear models.

The EnKF removes the requirement for a linearized model by simulating mean and covariance directly with the help of a model ensemble. The underlying idea is that a model's probability distribution can be approximated by a finite model ensemble, which then allows relatively efficient calculation of the forecast error covariance. In the forecast step, ensemble members are propagated forward by the nonlinear model. In the update step, the ensemble of forecasted states is used to compute the statistics required to perform the data assimilation update. More specifically, the covariance matrix $\mathbf{B}$ is approximated with the help of the ensemble of model states $\{\mathbf{x}^i\}$ as

$$\mathbf{B} \approx < (\mathbf{x}_f - <\mathbf{x}_f>)(\mathbf{x}_f - <\mathbf{x}_f>)^T >$$

where <.> denotes the average over the ensemble [51]. The analysis step is

$$\mathbf{x}_a^i = \mathbf{x}_f^i + \mathbf{K}\,[\mathbf{y}^i - \mathbf{H}\mathbf{x}_f^i]$$

for each ensemble member, where $\mathbf{K}$ calculated as in Box 1 using the ensemble approximation of $\mathbf{B}$. The EnKF has been widely applied to OBMs and has many variants that differ in the way the update step is performed including the Stochastic EnKF (SEnKF)[51,56], where an ensemble of observations drawn from an assumed distribution of observations is used to update each ensemble member, and the Deterministic EnKF (DEnKF)[57–60], where mean and covariance of the ensemble are computed and updated using the new observations, and a new ensemble is drawn from the updated distribution.

During the forecast step of the EnKF, each ensemble member is integrated forward which means the computational effort of running one realization of the OBM is multiplied by the number of ensemble members (Figure 4). Computational constraints limit the possible size of the ensemble to between tens to a few hundreds of members. The distribution of the high-dimensional state vector is thus undersampled by ensembles that are computationally feasible. Two techniques, covariance localization and covariance inflation, are used to reduce the negative effects of this

undersampling. Localization decreases the impact of distant covariance estimates thereby reducing the effect of "spurious" long-distance correlations in the ensemble. Inflation artificially increases the ensemble covariances to counteract low covariance estimates due to small ensembles[61].

Analogs to the underdetermination and parameter interdependency problems described for parameter optimization also exist in the context of state estimation. OBMs have many biogeochemical state variables, most of which are not directly observed. Although state estimation is multivariate, meaning unobserved variables can be informed by available observations of related variables through the relationships expressed in the covariance matrix, many elements of the state vector may not be well informed by the available observations (the estimation problem is underdetermined). Furthermore, if an increase of one state variable (e.g., phytoplankton) is dictated by observations and can be achieved by different adjustments to the model's biogeochemical transformations (e.g., by either increasing nutrient supply or zooplankton grazing) additional observation types would be necessary to conclusively inform which update is correct[62,63]. Formal analysis of the impact of individual observation types can be useful in this context[64,65].

While the EnKF is probably the most common sequential data assimilation technique, variational approaches are being applied to OBMs as well (see Table 1 in ref.[66]). The 3-dimensional variational (3D-Var)[67–69] and 4-dimensional variational (4D-Var)[62,70] approaches include the common sequence of forecast and analysis steps and where the analysis step uses the variational method (Figure 4). In 3D-Var the observation operator H is assumed to be time independent, which is thought to be appropriate for short forecast windows. In the 4D-Var approach, H is time dependent and includes the nonlinear forecast model although it is common to linearize the problem[71]. Particle Filters such as the Sequential Importance Resampling (SIR)[72] are promising alternative methods that do not rely on the assumption of Gaussian errors distributions and have been used for state and parameter estimation of OBMs[73] but are not yet widely used.

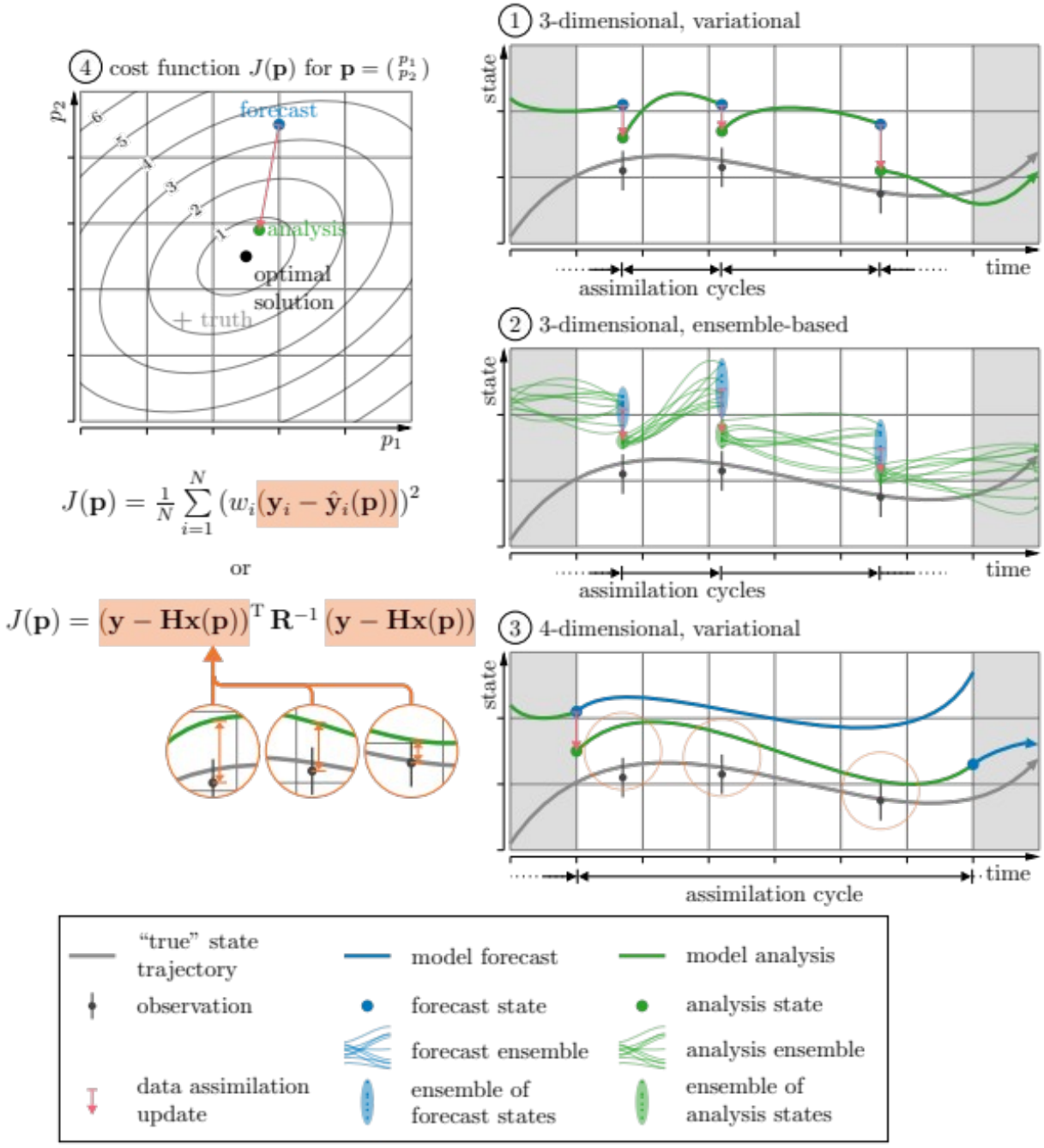
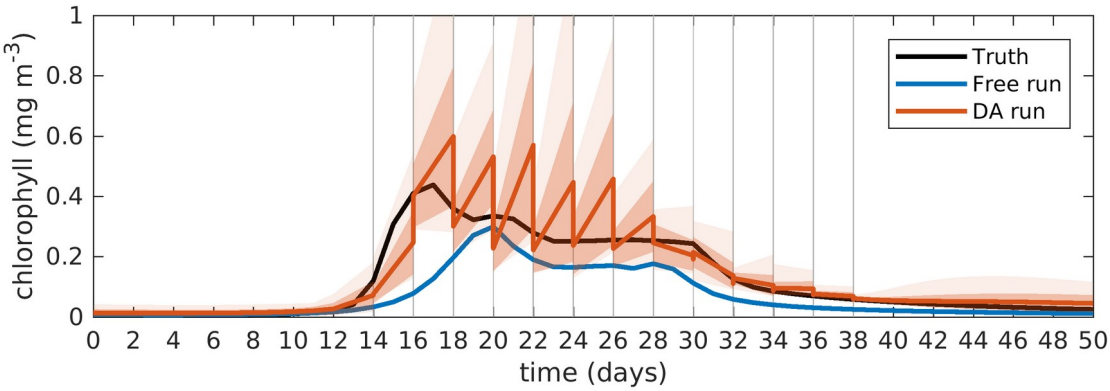
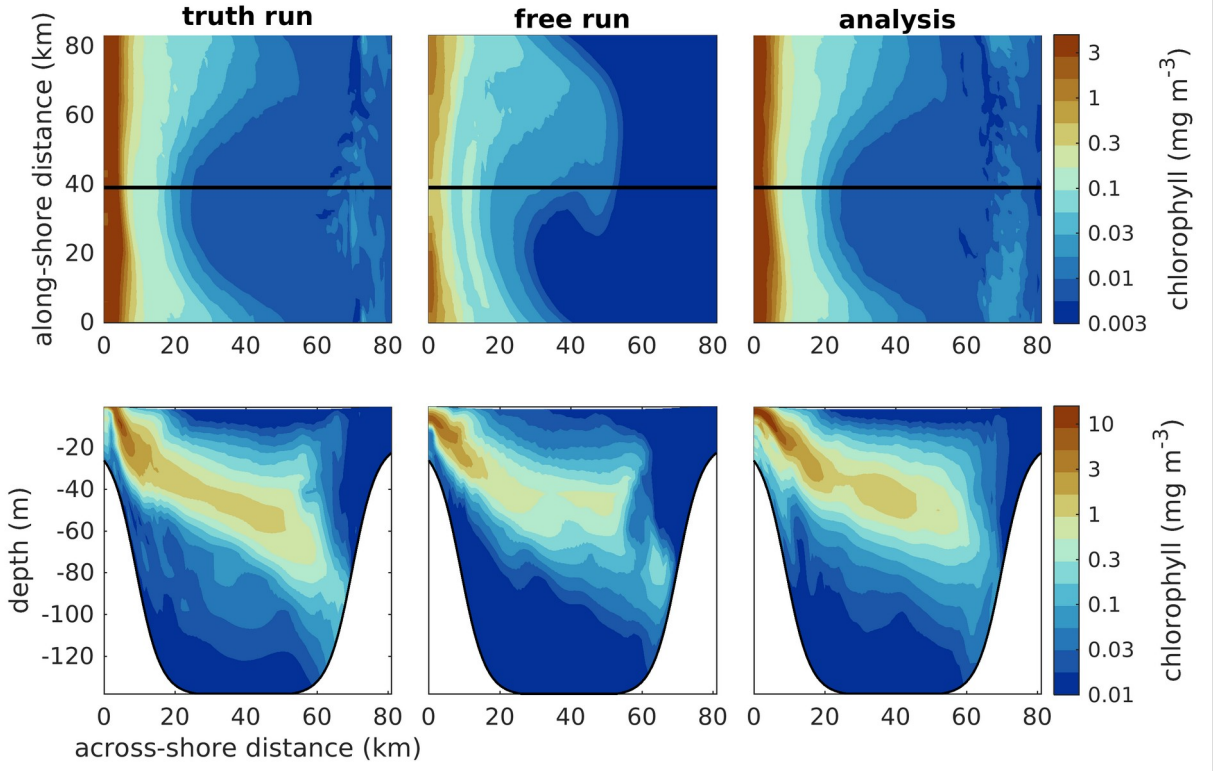


Figure 4: Illustration of parameter optimization (left) and state estimation (right). In both cases the goal is to utilize observations of the true state of the ocean to obtain more accurate model estimates either by improving model parameters or the model state itself. In parameter optimization, the misfit is measured by a cost function J(p). The initial guess (analysis) is improved upon systematically by minimizing the cost function. The parameters at the minimum are referred to as the optimal parameters. In state estimation, the model state variables are updated sequentially, as observations become available, in a series of forecast and update steps.

Box 4: Matlab and FORTRAN codes for ensemble-based state estimation that are suitable for 3-dimensinal OBMs is available on github[74]. The example is set up as an identical twin experiment for an idealized three-dimensional OBM (using ROMS) and the Deterministic Kalman Filter (DEnKF) as in ref.[57].



Evolution of mean surface chlorophyll concentration (top) in the truth run (black line), the perturbed free run (blue line), and the model ensemble (red line) illustrating the updates from forecast to analysis during update steps. The light and medium red shades around the red line show the full range and +/- one standard deviation of the ensemble. The gray vertical lines indicate assimilation steps.



Surface and vertical distributions (upper and lower rows) of chlorophyll concentration on day 16 in the truth run, free run, and analysis of the ensemble (from left to right) illustrate the discrepancies between the truth and free runs, and the improvement in the analysis. The location of the vertical transect is indicated by the black line in the middle panels.

[H1] Results

[H2] Model evaluation

Whether an OBM has value as a research tool for addressing the intended scientific question depends on how accurately it represents the processes that are relevant to the question to be addressed. Evaluating a model's performance is thus an integral part of model analysis. It relies on comparing the model output to observations, often in an iterative loop where the evaluation of a hindcast simulation is followed by model refinements such as increases in model resolution, improvements in parameterizations, or changes in model structure followed by a new hindcast and evaluation [75]. The three most commonly used statistical metrics for model evaluation are the root-mean-square error (RMSE), the bias, and the correlation coefficient (see Box 5). All three are calculated by directly relating observations to their model counterparts, and all are relative measures without any objective criterion that indicates which range of values is acceptable or unacceptable. These metrics can be calculated using spatial and temporal averaging, temporal averaging only, or spatial averaging only[76]. Specialized graphics have been devised to effectively represent some of these metrics for a large number of different models or different hindcasts from the same model, e.g., the Taylor diagram[77] and the Target diagram[78]. Two metrics with built-in criteria as to whether a model's performance is acceptable are *Z-scores*, which consider variability within the observational data set, and the so-called model efficiency or model skill, which quantifies whether the model outperforms an observational climatology (see Box 5).

Box 5: The following are common statistical metrics for model evaluation.

The *root-mean-square error* (RMSE), defined as $RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i-o_i\right)^2}$, is a measure of the overall distance between observational data points $o_i\left(i=1,..,n\right)$ and their model equivalents m_i.

The *bias* (b), defined as $b=\frac{1}{n}\sum_{i=1}^{n}\left(m_i-o_i\right)$, measures to what degree the model overestimates $\left(b>0\right)$ or underestimates $\left(b<0\right)$ the observational data.

The *correlation coefficient* (r), defined as

$r = \left(\sum_{i=1}^{n} (m_i - \overline{m})(o_i - \overline{o}) \right) / \sqrt{\sum_{i=1}^{n} (m_i - \overline{m})^2 \sum_{i=1}^{n} (o_i - \overline{o})^2}$, measures to what degree the observations and their model equivalents are linearly related. For $r = 1$ they would perfectly relate to each other (or be perfectly correlated), for $r = 0$ there would be no relation, and for $r = -1$ they would be perfectly anti-correlated (meaning whenever observations increase the model equivalents would decrease by a proportional amount).

A *climatology* $\overline{o}$ is the long-term average of an observational data set $o_i (i = 1, \ldots, n)$. A climatology can be spatially resolved (i.e., a gridded field) or calculated just for one location; it can be a temporal average over all data (referred to as annual climatology) or temporally resolved by month (monthly climatology) or day (daily climatology) or another averaging interval.

The *model skill* or *model efficiency* (me), defined as $me = 1 - \left(\sum_{i=1}^{n} (m_i - o_i)^2 \right) / \left(\sum_{i=1}^{n} (\overline{o} - o_i)^2 \right)$, measures whether a model results in better (0<me<1) or worse (me<0) predictions than an observation-based climatology $\overline{o}$.

The Z-score Z_i, defined as $Z_i = \frac{(m_i - \mu)}{\sigma}$, relates model output m_i to corresponding observational data of the same property. Assuming the observational data are normally distributed with mean μ and standard deviation σ, the Z-score indicates the probability of encountering the value m_i in the data set given the natural variability reflected in the data set.

No single metric provides a complete evaluation of a model's predictive power, hence multiple complementary metrics should always be used in concert[79]. A model may provide accurate estimates for some variables, locations, or times but perform poorly for others[80]. Hence, space, time, and a breadth of variable types should be considered in any comprehensive model assessment. Furthermore, there may be aspects of a model simulation that one cannot reasonably expect the model to reproduce. For example, an OBM without state estimation cannot exactly reproduce stochastic aspects of the system such as the exact timing and location of elevated chlorophyll due to mesoscale eddies. However, one might reasonably expect that the magnitude, shape, and frequency of eddy-induced chlorophyll enhancements be represented well. Specialized metrics that take mismatches in space and time into account can be used for this purpose[81].

In OBMs with data assimilation, the need for model assessment expands further, to include evaluation of state or parameter estimates, and becomes more difficult because these estimates already contain information from the observations that were assimilated. One general strategy is to only use observations in the evaluation that were not assimilated and can thus be considered independent[79]. In practice, this presents a conundrum because 1) one would like to use all

available observations in assimilation to obtain the best possible estimates, and 2) observations that are withheld from assimilation may be correlated to those used in assimilation and thus not truly independent. Decorrelation scales **[G]** must be considered when deciding what are truly independent observations. Perhaps the most convincing assessment of an assimilative model is an ongoing test of its predictions against observations as they become available. In a closely related approach, that can be used to assess sequential state estimates in hindcast mode, misfits between current observations and the model forecast for the current time are compared to the misfit between current observations and the analysis from the previous update step. The underlying idea is to test whether the forecast outperforms the so-called persistence model, which assumes that the previous analysis is also the best forecast.

[H2] Challenges to model evaluation

The rigour of any model evaluation hinges critically on the observational data set available. Despite major efforts in ocean observation in the 20th and early 21st century, the ocean's biogeochemical state remains under-observed in many critical aspects hampering evaluation and systematic improvement of OBMs[82]. The major pillars of biogeochemical observation efforts have ocean colour satellites, coordinated ship-based initiatives aiming to obtain global 3D distributions of a core set of properties, time-series sites which provide the process understanding and enable the broadest suite of observations, networks of ships-of-opportunity, and numerous investigator-driven individual cruises. Although each of these is valuable, biogeochemical undersampling is a problem because the cost and effort involved in ship-based measurements limits them to a few instances in space and time, and satellite observations of ocean colour only provide information about a plankton-related properties at the very surface of the ocean. The maturation of autonomous platforms (profiling floats and gliders) and miniaturized biogeochemical sensors over the past two decades has paved the way for cost-effective and routine observation of a broad suite of biogeochemical properties[83,84]. Autonomous observation technology is quickly becoming an additional, complementary pillar and makes it feasible to observe the global ocean in near-real time at an unprecedented spatial and temporal resolution with an accuracy sufficient to detect climate-induced changes[85].

A technical challenge to rigorous model evaluation is access to existing observations. While many of the major coordinated observational initiatives have been able to provide sustained access to the resulting observations, often through individual repositories, small, investigator-driven efforts were typically not well positioned to guarantee long-term access to specialized observations and depend on national or international repositories like NOAA MEDS (see section Reproducibility and data deposition). A unified data management approach with standardized meta-data requirements and data formats that ensures discoverability and accessibility of existing data sets has been identified by the oceanographic community as a common goal[86,87] and will greatly benefit OBM evaluation and improvement.

Model evaluation relies on climatologies for many properties, satellite-based estimates of chlorophyll and primary production, comprehensive time series for a relatively small number of sites throughout the ocean, focused process studies with a relatively narrow spatial and temporal footprint, and increasingly global autonomous data sets. Typically, an available data set includes fewer properties than the OBM's state and few of the biogeochemical transformations explicitly represented by the OBM are observed. As a result, an excellent agreement between model and observations does not necessarily guarantee that the model's representation of unobserved properties and fluxes is correct or that the model is a skillful predictive tool. Internal model errors may compensate each other in such a way that the correct-seeming result, judging by a limited data set, is obtained for the wrong reasons. Ongoing evaluation of models and their predictions against sustained observational data streams with an increasing breadth of observables is thus critical.

[H1] Applications

OBM applications range from purely scientific (e.g., for the purposes of building fundamental understanding and hypothesis testing) to very practical (e.g., to produce forecasts and model-derived products). Below we describe key examples that illustrate the breadth and importance of different applications without any attempt at being comprehensive.

[H2] Ocean carbon accounting

The global ocean absorbs about a quarter of contemporary human emissions of carbon dioxide (CO_2) to the atmosphere. OBMs have been central in quantifying the patterns and rates of ocean anthropogenic CO_2 uptake that occurs via natural physical-chemical processes at the air-sea interface and ocean circulation transporting surface water with excess CO_2 into the ocean interior[88]. OBMs are also pivotal for characterizing future ocean CO_2 uptake and the sensitivity of that uptake to ocean climate change under different climate policy scenarios[89]. In part because of the large CO_2 uptake capacity of the ocean, several approaches have been proposed to enhance ocean uptake through deliberate carbon dioxide removal or negative emissions technologies[90]. Rapid decarbonization of the global economy is needed to meet the international Paris Climate Agreement to keep global surface warming well below +2.0 degrees C relative to pre-industrial conditions. Coupled carbon-climate models indicate that society must meet roughly net-zero human CO_2 emissions by approximately mid-century and given the challenges of abating all human CO_2 emissions from the energy and transportation systems, a substantial amount of deliberate carbon dioxide removal may be required[89].

Significant knowledge gaps exist for all ocean-based carbon dioxide removal approaches, with unknowns spanning across the efficacy of net CO_2 uptake from the atmosphere, permanence of the carbon storage, verification or carbon accounting of the method, scalability, and environmental impacts[91]. In conjunction with laboratory and field experiments, OBMs are central

to resolving many of these questions across a range of scales (local, regional, and global). Deliberate CO_2 removal will be challenging to verify because it would represent a relatively small perturbation of the large natural uptake of anthropogenic CO_2 and background natural variations. Ocean circulation would transport any added CO_2 away from site of deliberate manipulation and dilute signals. Alterations of the ocean's carbonate system and nutrient inventory would have downstream effects ranging from desirable (e.g., countering ocean acidification in the case of alkalinity enhancements) to counterproductive (e.g., by diminish the additionality of a carbon dioxide removal or enhancing acidification). OBMs in combination with well-resolved, comprehensive observation will be central to verification of CO_2 removal and carbon accounting.

Previous OBM studies have explored some of these questions for different CO_2 removal techniques including ocean iron fertilization of high-nitrate, low-chlorophyll regions[92–94], artificial upwelling of nutrients into the surface ocean[95], macroalgae farming[96], seawater alkalinity enhancement[97,98], and more generally the permanence of CO_2 removal[99].

[H2] Ocean ecosystem health

[H3] Deoxygenation

Dissolved oxygen is an important measure of ocean ecosystem health because oxygen is essential for supporting aerobic aquatic life. Long-term observations indicate that the oxygen content of the global ocean has declined by more than two percent over the past five decades[100] raising concerns about profound effect on ocean biogeochemical cycles and marine ecosystems[101]. The observed oxygen loss of the global ocean is projected to continue[102,103] primarily because oxygen solubility decreases, biological oxygen demand increases, and ventilation of the deep ocean is reduced under global warming[104,105]. In addition to the large-scale climate effect, coastal waters are affected by growing anthropogenic nutrient inputs that lead to a worldwide expansion of coastal hypoxia[106,107].

Global OBMs and ESMs have been used to understand why global ocean oxygen content changes[108] and for making future projections under different emission scenarios[102,103,109]. These models consistently project continued and accelerating deoxygenation but underestimate the observed deoxygenation rates and fail to reproduce the observed patterns and temporal variability of oxygen changes[105]. Notable differences in the simulated intensity and spatial patterns of oxygen projections among ESMs[109,110] point to deficiencies in our mechanistic understanding and modeling capabilities. Likely factors limiting current model capabilities have been identified in multi-model comparisons and/or sensitivity experiments including insufficient model resolution[111], inaccuracies in ocean mixing parameterizations[112,113], and incorrect model representation of biological processes[108].

Regional OBMs are widely used to improve our understanding of coastal oxygen dynamics and to guide management in coastal regions[114,115]. They allow us to examine and quantify the factors governing oxygen variability and hypoxia formation[116–119], to project changes in oxygen supply under climate change[120,121], and to understand the consequences of hypoxia on the marine food web[122,123]. Regional OBMs are used also for a range of applied purposes, e.g., to evaluate how hypoxia would be affected by different nutrient reduction scenarios[124,125], to investigate the compounding effects of anthropogenic nutrient inputs and climate change on hypoxia[125–127], for seasonal forecasting[128,129] and to explore eco-engineering strategies for hypoxia mitigation[130]. Data-assimilative OBMs provide short-term ecological forecasts (including for oxygen) in various coastal systems by optimally combining model and observations[66].

[H3] Acidification

Ocean uptake of anthropogenic CO_2 slows its atmospheric accumulation, and thus climate change, but changes seawater chemistry by elevating dissolved inorganic carbon and aqueous CO_2 and reducing pH and carbonate mineral saturation states[131]. Because of the shift in seawater pH toward acidic conditions, this process is referred to as ocean acidification. Acidification likely has deleterious impacts on ocean ecosystems and coastal human communities that depend on marine resources[132,133]. OBMs are used extensively to quantify past and future rates and patterns of ocean acidification.

Global OBMs simulate reductions in surface planktonic calcium carbonate ($CaCO_3$) production and elevated (shallower) $CaCO_3$ remineralization[134–136]. These simulations suggest that open-ocean surface acidification is controlled largely by the choice of atmosphere CO_2 scenario because of relatively rapid air-sea CO_2 gas exchange on annual and longer timescales, while subsurface acidification is more strongly dependent on simulated ocean ventilation rates which differ across models[137]. Decadal prediction systems using ensemble forecasts from ESMs forecasts have demonstrable skill for surface pH variations of up to five years[138].

In coastal ecosystems acidification is compounded by eutrophication **[G]**, acidic freshwater discharge, coastal upwelling, and terrestrial organic carbon inputs. Regional OBMs are used to analyze the synergy between acidification and eutrophication[139] and for characterizing the highly variable physical and biogeochemical conditions[140,141]. These models have also been used to quantify the time of emergence when anthropogenic changes exceed natural variability[142], and to investigate how anthropogenic CO_2 trends amplify the frequency of extreme acidification events[143] and compound events with overlapping extremes of acidification, marine heatwaves, and deoxygenation[144].

The biological impacts of ocean acidification will likely vary across different marine environments (coral reefs, wetlands and shallow coastal systems, and pelagic-planktonic systems) with effects extending across scales from organisms to the community and whole ecosystem and possible positive and negative effects for different taxonomic groups[133]. For planktonic systems, elevated aqueous CO_2 is projected to increase primary and secondary

productivity, alter community structure, and perhaps increase the frequency of harmful algal blooms. At the same time, reduced carbonate mineral saturation states are projected to lower the competitiveness of calcifying plankton species[145]. This poses considerable challenges for OBMs because existing model structures and parameterization are tailored to present conditions and not necessarily able to account for potential biological changes at the organism and community level.

[H3] Fisheries

Primary production by phytoplankton sustains the marine food web; however, OBMs typically include only species on the lowest trophic levels (i.e., phytoplankton and zooplankton) with the implicit assumption that predation on plankton by higher trophic levels (e.g., fish) can be represented by an additional mortality term in the zooplankton equation. The clear dependence of fishery yields on ecosystem primary production[146] led modelers of higher trophic level processes to use simulated primary production from OBMs as forcing for their models in the early 2000s. Examples include a study of the impact of climate change on tuna populations in the tropical oceans[147], and a study using IPCC-type models to assess the impact of climate on living marine resources[148]. This type of work continues, as exemplified by the recent FishMIP model intercomparison project[149] where a set of global, upper trophic level models were forced with output from ESM projections. These projections reveal that average animal biomass in the global ocean could decline by 17-19% in high emission scenarios and 5-7% in low emission scenarios by 2100[150,151], primarily due to increasing temperature, decreasing primary production, and changes in species ranges. These OBM projections have been used to project changes in fish catch potential[152] and global fisheries revenues[153]. Although these results have been widely used in international assessments[154,155], the projections are subject to large uncertainties with a substantial contribution to these uncertainties from OBM-projected lower trophic level biomasses and production[150].

More recently, a direct coupling of OBMs to higher trophic level models has enabled examination of top-down control from higher trophic levels on planktonic ecosystems and on marine biogeochemical cycles in general. For example, a study that coupled an OBM to an upper trophic level model with explicit representation of vertically migrating organisms (which feed at the surface but excrete and respire at greater depth) estimated that diurnal vertical migration contributes significantly to the biological carbon pump and could amount to 20% of the vertical carbon flux due to settling particles[156]. Diagnostic analyses of the contribution of zooplankton's diel vertical migration to biological carbon export have yielded similar results[157].

[H2] Observing system design

Observing system simulation experiments (OSSEs) are a type of data assimilation experiment in which synthetic, or simulated, observations are used to design new or modify existing observing systems. OSSEs have their origin in numerical weather prediction[158] but are increasingly used for ocean models[159], including OBMs[58,69]. Typically, OSSEs are performed when a new observing system is designed, or in anticipation of a proposed change to an existing observing system, e.g.,

when a new instrument type or sensor becomes available. An OSSE provides information about which combination of instruments, and in which configuration, would lead to the most significant improvement in forecast skill of a data assimilation system, which is valuable given the expense of new satellite-based sensors or in-situ ocean observing arrays. Beyond guiding observing system design, OSSEs can also be used to prepare for new data, that is, to develop and improve data processing, data storage, and streamline the data assimilation system, before new datasets become available.

OSSEs consist of producing a simulated set of observations and assimilating them in a data assimilation system to examine their impact on the system's predictive skill. The simulated observations, which would include representations of the existing operational observing system and the proposed additions, should be obtained from an independent model simulation that satisfies a few criteria[159]. This independent simulation is referred to as the nature run. The simulated observations sampled from the nature run are then assimilated into a data-assimilative model that closely resembles that operational system. This is referred to as the perturbation run. By comparing the results of the perturbation run and a control run, where only simulated observations from the existing system were assimilated, the potential benefits of the new observational assets or configuration can be quantified. Since the nature run can be sampled at any desired location and frequency, many different observing system configurations can be assessed. However, for an OSSE to yield reliable conclusions about the value of different observational assets and their configuration, the discrepancies between the simulated observations and the data-assimilative model must have the same error statistics as the discrepancies between real observations and the data-assimilative model.

Creating realistic simulated observations with representative error statistics, is not an easy task and should follow the principles laid out by ref.[159]. So-called identical twin OSSEs use the same model to perform nature and perturbation run where discrepancies between the two are introduced by modifying one or multiple model inputs (i.e., parameters, initial conditions, physical forcing, etc.). While identical twin OSSEs are relatively easy to set up, they are likely to produce simulated observations with non-representative errors, leading to unrealistically optimistic forecast skill estimates[58]. The more desirable fraternal twin OSSEs use a different model for generating the observations than the assimilative system. In any case, OSSEs require careful calibration, comparing the nature run to reality, and examining observation innovation statistics and forecast skill of the OSSE in comparison to the assimilation of real data.

[H1] Reproducibility and data deposition

In alignment with efforts to make ocean observations freely and easily accessible to the research community and a general movement toward more transparency in research dissemination (e.g., open-access publishing) it is becoming standard that OBM codes and output be made freely available also. All of the widely used OBM codes are community efforts with shared code repositories and active user groups (Table 1). Code archiving services such as github.com enable

scientists to maintain and share their individual code repositories and collaborate with others. Archiving of model output presents a larger challenge because the size of raw model output easily exceeds available storage capacities making some subsampling and curation necessary. Community-driven archiving services like Zenodo.org enable permanent and open archiving of data sets, including OBM output, with assignment of persistent digital object identifiers (DOI). Some physical and biogeochemical simulation output from more formal model intercomparison activities (see below) is also archived and made available through centralized data repositories[19,160].

The ocean modeling community has a history of embracing model intercomparison exercises where different modeling groups collaborate in defining a target simulation with a standard set of implementation and evaluation criteria[161,162]. By comparing different model architectures against common criteria, it is possible to identify shortcomings in individual models that can be remedied and assess model uncertainties that cannot be eliminated but must be considered in interpreting results. The IPCC has a well-established process for intercomparison of ESMs referred to as Coupled Model Intercomparison Project (CMIP) that includes intercomparison projects for ocean physical and biogeochemical models[19,160]. The latest round, CMIP6, corresponds to the 6th Assessment Report of the IPCC and includes an intercomparison of the ocean biogeochemical components of ESMs[137,163,164] Coordinated efforts to compare multiple OBMs in a systematic fashion, dates further back to the mid-1990s and early 2000s with the first two phases of the Ocean Carbon Model Intercomparison Project (OCMIP-1 and OCMIP2)[162]. The OCMIP team created common frameworks for ocean tracer and biogeochemical experiments (e.g., natural and anthropogenic radiocarbon; chlorofluorocarbons or CFCs; abiotic and biotic carbon and nutrient cycling) where experiment packages included specified atmosphere boundary condition for trace gasses (e.g., CFCs, CO2), ocean biogeochemical parameterizations, and standardized model diagnostics and output. Individual groups implemented the OCMIP experimental protocols in different ocean physical general circulation models with different physical forcing and ocean circulation[101]. The resulting range of simulated ocean tracer and biogeochemical fields then could be related back to differences in the underlying physics, not differences in the biogeochemical components[165], and model skill could be evaluated against a common set of observed metrics[166].

Intercomparisons are also well established for regional models within the Integrated Ocean Observing System's Coastal Ocean Modeling Testbed (IOOS COMT)[167]. For example, as part of the COMT, different physical models for simulation and prediction of coastal hypoxia in the Mississippi River outflow region were compared using the same simple oxygen model by ref.[168] to distinguish the model-to-model differences arising from model physics from those resulting from different biological model formulations[169]. This approach is similar to the OCMIP protocols described above. An alternate and equally useful approach is to compare different biogeochemical parameterizations within the same ocean physical model, especially when it includes parameter optimization to minimize uncertainties in parameter choices, thus focussing the intercomparison on structural model differences[38,46,47].

OBM intercomparison activities rely heavily on the availability of openly accessible, high-quality ocean data sets for model development and evaluation[19,82]. In many cases, gridded data products and climatologies are as or more desirable than simply compilations of raw ocean data profiles, and the construction of such data products spanning across many individual data collectors and time requires substantial effort by domain experts spanning oceanography, data analysis, and data management. The visibility of ocean data management as a distinct and critical element of ocean research has risen substantially over the past couple of decades, with multiple aspects spanning from informatics and cyber infrastructure to research culture and incentives for scientists to share data through recognized data repositories and to properly cite others' data products[170,171]. Current community ocean data efforts revolve around the emergence of new standards that data be findable, accessible, interoperable, and reusable (FAIR)[86]. Data system elements can include agreement on common sampling and analytical measurement methods, data quality control and assurance checks, lab and field intercomparison activities, analytical standards and reference materials, data ontologies and vocabularies, data reporting requirements and formatting standards, and support for data accessibility and archiving through data repositories.

The physical hydrography community has a long history and experience building unified data products for temperature, salinity, nutrients, and oxygen (e.g. ref.[172]), and standard products such as the World Ocean Atlas 2018 are created by NOAA and other ocean data centers[173]. The ocean chemical tracer and biogeochemical community has pursued similar efforts to compile, and where feasible, grid water column data relevant to OBMs. The Global Data Analysis Project (GLODAP), for example, compiled data on ocean circulation tracers (e.g., radiocarbon and chlorofluorocarbon) and dissolved inorganic carbon system variables including alkalinity based on historical global ocean surveys dating back to the 1970s[174]. The GLODAP product includes objectively analyzed, gridded spatial maps for both observed properties as well as derived products such as anthropogenic CO2. The GLODAP data product is a living data set that was most recently updated as GLODAPv2.2020[175]. Many of the large-scale, international ocean observing efforts are built on collaborations to generate similar publicly accessible, standardized data products that are routinely updated with new field observations. Relevant examples for OBM efforts include the GO-SHIP hydrographic program[176], surface ocean CO2 observations[177], time-series[178], plankton products[179,180], and quality-controlled BGC-Argo products[181,182].

Table 1: Examples of important repositories for OBM codes, OBM output, and observations.

Repository type	Abbreviation	Name	Link
Codes	ROMS	Regional Ocean Modeling System	https://www.myroms.org/
	MOM	Modular Ocean Model	https://mom-ocean.github.io/

			docs/userguide/
	HYCOM	Hybrid Coordinate Ocean Model	https://www.hycom.org
	NEMO	Nucleus for European Modeling of the Ocean	https://www.nemo-ocean.eu/
	MITgcm	Massachusetts Institute of Technology General Circulation Model	https://mitgcm.org/
Outputs	OMIP	Ocean Model Intercomparison Project	https://www.wcrp-climate.org/modelling-wgcm-mip-catalogue/cmip6-endorsed-mips-article/1063-modelling-cmip6-omip
	CMIP6	Coupled Model Intercomparison Project 6, output data repository	https://pcmdi.llnl.gov/CMIP6/
	C4MIP	Coupled Climate Carbon Cycle Model Intercomparison Project	https://www.wcrp-climate.org/modelling-wgcm-mip-catalogue/cmip6-endorsed-mips-article/1050-modelling-cmip6-c4mip
	RECCAP-2	REgional Carbon Cycle Assessment and Processes 2	https://reccap2-ocean.github.io/
	MARine	Ecosystem Model Intercomparison Project	https://pft.arc.hokudai.ac.jp/maremip/index.shtml
	IOOS COMT	Integrated Ocean Observing System's Coastal Ocean Modeling Testbed	https://ioos.noaa.gov/project/comt/
Observations	GEOTRACES	An International Study of the Marine Biogeochemical Cycles of Trace Elements and Isotopes	https://www.geotraces.org/
	GO-SHIP	Global Ocean Ship-based Hydrographic	https://www.go-ship.org/

		Investigations Program	
	SOCAT	Surface Ocean CO2 Atlas	https://www.socat.info/
	GLODAP	Global Ocean Data Analysis Project	https://www.glodap.info/
	BGC-Argo	Biogeochemical Argo	ftp://ftp.ifremer.fr/ifremer/argo or ftp://usgodae.org/pub/outgoing/argo
	BCO-DMO	Biological and Chemical Oceanography Data Management Office	https://www.bco-dmo.org/
	World Ocean Atlas 2018	Ocean hydrography, oxygen and nutrients	https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/
	SeaBASS	NASA SeaWiFS Bio-optical Archive and Storage System	https://seabass.gsfc.nasa.gov/
	MARDAT	MARine Ecosystem Biomass DATa	http://www.pangaea.de/search?&q=maredat

[H1] Limitations and optimizations

Application of OBMs is subject to computational limitations (i.e., limits in CPU time and disk space). This requires compromises with respect to the level of spatial and process resolution that can be achieved given the required domain size, length of integration time, and size of model ensembles. For example, the resolution of most ESMs is too coarse to capture ocean mesoscale features, though some modeling groups are exploring global ocean mesoscale plankton simulations (e.g., ref[183]). Given current computing resources, an increase in spatial resolution of a global model with century integration times typically is only possible if the number of biogeochemical state variables is drastically reduced (e.g., ref.[25]). Regional models are affordable at much higher spatial resolution but require imposition of boundary conditions from larger-scale models and much shorter integration times. Practical workarounds to this problem include approaches for up- and down-scaling[184], nested domains including with 2-way coupling (where information is flowing from the large to the small scale and vice versa)[185], and adaptive grids and model structures although the latter are difficult to implement and rarely used.

Uncertainties in OBM parameters and structure are a major issue limiting the models' predictive skill. Many model parameters cannot be determined experimentally and have relatively large

plausible ranges. Parameter optimization approaches should theoretically be able to address this as an inverse problem where information contained in biogeochemical observations is used to infer the underlying parameters; however, in practice observations are often limited in terms of resolution and breadth of variable types so that many model parameters are left unconstrained by optimization[37,43]. The issue of undetermined parameters worsens with increasing model complexity (i.e., increasing numbers of state variables) because the number of poorly known parameters multiplies and the number of degrees of freedom increases[26]. The strong non-linearities characterizing OBM equations also contribute to the difficulty. One manifestation of the underdetermination problem is cancellation of errors, where a model seemingly agrees with available observations but does so because underlying errors compensate for each other. This includes the possibility of the biogeochemical model compensating for shortcomings in the physical model component. This is problematic for future climate projections where compensating errors give a plausible looking simulated present but with limited confidence of future behavior[186]. Whenever different errors compensate for each other, the OBM might perform fine for the observational period for which it was tuned, but not outside of it, limiting the mechanistic insights that can be gained from it and its use as a predictive tool. Slow climate-biogeochemical feedbacks, for example, are difficult to probe with current observations. A prudent approach is to apply Occam's razor, i.e., limit the number of biogeochemical state variables to only those that are necessary and focus on processes where there is some conceptual or theoretical understanding of climate sensitivity. In practice this depends on the scientific or practical challenges that are to be addressed and is somewhat subjective.

Related to the underdetermination problem of model parameters is the issue of structural uncertainties in OBMs. Since the equations governing biogeochemical state variables are empirical (other than mass conservation they are not derived from fundamental laws and first principles as, e.g., the Navier-Stokes Equations of fluid motion), there are no universally agreed upon parameterizations and no optimal model structure. Although most models in current use have converged on a similar, intermediate-complexity structure, perhaps for practical reasons, their model structures and predictive capabilities are not rigorously tested. Until recently, observational data sets have typically been too limited (in space or time, or breadth of observation types) for a thorough validation of OBMs. The rise of global autonomous biogeochemical observation networks[84,180] is beginning to greatly alleviate this problem and will likely prove transformative for the further development of OBMs.

An obvious limitation of the use of ESMs for projecting future conditions is that projections cannot be validated. The use of past climatic changes is one possibility but has the drawbacks that observational constraints are limited to paleoproxies and that there is no close analog to the current global warming in the recent geological past (i.e., with the current continental plate configuration). Assessment of ESMs for present-day conditions[34,163] is informative about model biases for the present but not necessarily about how they behave outside the observed range of conditions. Again, the issues of underdetermination and compensating errors lead to large and poorly quantified uncertainties. In numerical weather prediction it is generally true that an

ensemble of different models represents a better forecast than individual models because model uncertainties partly cancel each other. This notion has been adopted in climate projections where ensembles of ESMs are commonly used. However, evaluation of CMIP5 and CMIP6 ensembles for present-day conditions shows that the assumption of cancellation of errors does not generally hold[34]. This is not a surprising result if one considers that many if not all models may share similar shortcomings and biases. A recent development is the use of explainable and empirical inter-model relationships between characteristics of the present-day conditions and long-term climate projections, the so-called emergent constraint approach[187]. It has enabled a reduction in projection uncertainties in ESMs and has been applied successfully to marine biogeochemistry[188,189].

Another major challenge when using OBMs for future projections is that a fixed biogeochemical structure may not be adequate to account for functional changes in biological communities as a result of acclimation and adaptation to new environmental conditions. One potential workaround is to enable emergent communities as in the approach of ref.[21]. However, this is computationally costly and not feasible for the integration times required for future projections given current computational capabilities. Another approach that requires an ongoing input of well-resolved biogeochemical observations is to allow for adaptive model parameters[48,49]. This is potentially feasible for nowcasts but not projections. Regardless of whether the current status-quo prevails or new approaches are adopted, a continuous evaluation of OBM results against a well-resolved and broad suite of biogeochemical observations from a sustained global ocean observing system will be paramount as climate-related changes manifest.

[H1] Outlook

Anthropogenic perturbations of the global carbon and nitrogen cycles are leading to ocean warming, acidification, deoxygenation, and coastal eutrophication. All of these put stress on ocean ecosystems and act in combination with direct human impacts such as overfishing, trawling, etc. Prevention of and mitigation and adaptation to the negative effects of these stressors present formidable challenges. Skillful OBMs that provide robust, accurate, and actionable information are key to responding to these challenges appropriately and enabling the best possible outcomes. This will require a rigorous quantitative assessment and validation of OBMs including their structural uncertainties, and likely further model refinement and method development. Ocean modeling is already benefiting from computational advances in traditional ocean and climate models and new approaches leveraging machine-learning and other advanced techniques. The OBM simulation tasks also critically depend on a continued expansion of the global ocean observing system into more biogeochemical and ecosystem variables, taking advantage in particular of cost-effective autonomous platforms and sensors as well as remote sensing.

Given an expanded and sustained biogeochemical observing system, the development of well-constrained operational OBMs is going to be feasible in the near term and will benefit a broad

range of scientific purposes and practical applications. The need for accurate information in the context of some specific applications will require development of tailored OBM applications, e.g., ocean CO2 removal and carbon accounting; coastal eutrophication; fisheries; marine diseases; and harmful algal blooms. New applications focused on vulnerability and impacts will drive integration of the current generation of OBMs with specialized application models and more direct engagement with stakeholders. Demand may grow for near-term to seasonal out to multi-year forecasting products. A continued commitment to open-source code and open-access principles in dissemination of OBM results and derived products should be a priority by the research community and research funders.

Acknowledgements

K. Fennel and B. Wang acknowledge support from the NSERC Discovery Program (RGPIN-2014-03938), the Canada Foundation for Innovation (Innovation Fund 39902), and the Ocean Frontier Institute. J.P. Mattern was supported by the Simons Foundation (CBIOMES award ID: 549949). S. Doney acknowledges support from the U.S. National Science Foundation via the Center for Chemical Currencies of a Microbial Planet (NSF 2019589). L. Bopp acknowledges support from the European Union's Horizon 2020 research and innovation programs COMFORT (grant agreement No 820989) and ESM2025 (grant agreement No 101003536). L. Yu acknowledges support from the Center for Ocean Research in Hong Kong and Macau.

Competing interests

No competing interests to declare.

References

10 important references for highlighting:

Chai, F. *et al.* Monitoring ocean biogeochemistry with autonomous platforms. *Nature Reviews Earth & Environment* **1**, 315–326 (2020). – Review of autonomous approaches to measuring ocean biogeochemical properties which will likely prove transformative for OBM validation and assimilation.

Evensen, G. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53**, 343–367 (2003). – Influential paper that proposed the now widely used Ensemble Kalman Filter.

Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie. 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. Journal of Marine Research 48: 591–639. – A seminal early example of an ocean biogeochemical model applied to time-series data.

Fennel, K., and J. M. Testa. 2019. Biogeochemical Controls on Coastal Hypoxia. Annual Review of Marine Science 11: 105–130. doi:10.1146/annurev-marine-010318-095138 – Review of coastal hypoxia that put forward a simple non-dimensional number to elucidate key factors controlling hypoxia in diverse coastal systems.

Follows, M. J., S. Dutkiewicz, S. Grant, and S. W. Chisholm. 2007. Emergent Biogeography of Microbial Communities in a Model Ocean. Science 315: 1843–1846. doi:10.1126/science.1138544 – First paper to explore competition among a large number of phytoplankton groups within a global ocean model.

Kwiatkowski, L., O. Torres, L. Bopp, and others. 2020. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. Biogeosciences 17: 3439–3470. doi:10.5194/bg-17-3439-2020 – Assessment across a suite of Earth System Models of projected evolution of ocean biogeochemistry under 21st century climate change.

Maier-Reimer, E.: Geochemical cycles in an ocean general circulation model. Preindustrial tracer distributions, 7, 645–677, https://doi.org/10.1029/93GB01355, 1993. – A seminal paper describing one of the first marine biogeochemical model of the global ocean.

Orr, J. C., R. G. Najjar, O. Aumont, and others. 2017. Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). Geoscientific Model Development 10: 2169–2199. doi:10.5194/gmd-10-2169-2017 – Framework detailing common protocols for including ocean biogeochemistry and chemical tracers in Earth System Models.

Sarmiento, J. L. *et al.* A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic Euphotic Zone. *Global Biogeochemical Cycles* **7**, 417–450 (1993). – This

regional model of the North Atlantic was probably the first true OBM, i.e. an ocean circulation model with explicit representation of plankton dynamics.

Stow, C. A., J. Jolliff, D. J. McGillicuddy, S. C. Doney, J. I. Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead. 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems 76: 4–15. doi:https://doi.org/10.1016/j.jmarsys.2008.03.011 – Tutorial on common statistical approaches model-data skill assessment for ocean biogeochemical models.

1. Riley, G. A. Factors controlling phytoplankton population on George's Bank. *J. mar. Res.* **6**, 54–73 (1946).
2. Evans, G. T. & Parslow, J. S. A Model of Annual Plankton Cycles. *Biological Oceanography* **3**, 327–347 (1985).
3. Fasham, M. J. R., Ducklow, H. W. & McKelvie, S. M. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research* **48**, 591–639 (1990).
4. Franks, P. J. S., Wroblewski, J. S. & Flierl, G. R. Behavior of a simple plankton model with food-level acclimation by herbivores. *Marine Biology* **91**, 121–129 (1986).
5. Sarmiento, J. L. *et al.* A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic Euphotic Zone. *Global Biogeochemical Cycles* **7**, 417–450 (1993).
6. Revelle, R. & Suess, H. E. Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO2 during the Past Decades. *Tellus* **9**, 18–27 (1957).
7. Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric PCO2. *Nature* **308**, 621–624 (1984).
8. Siegenthaler, U. & Wenk, T. Rapid atmospheric CO2 variations and ocean circulation. *Nature* **308**, 624–626 (1984).
9. Maier-Reimer, E. & Hasselmann, K. Transport and storage of CO2 in the ocean ——an inorganic ocean-circulation carbon cycle model. *Climate Dynamics* **2**, 63–90 (1987).
10. Maier-Reimer, E. Geochemical cycles in an ocean general circulation model. Preindustrial tracer distributions. *Global Biogeochemical Cycles* **7**, 645–677 (1993).
11. Six, K. D. & Maier-Reimer, E. Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles* **10**, 559–583 (1996).
12. Sarmiento, J. L. & Gruber, N. *Ocean Biogeochemical Dynamics*. (Princeton University Press, 2006).
13. Glover, D. M., Jenkins, W. J. & Doney, S. C. *Modeling Methods for Marine Science*. (Cambridge University Press, 2011).
14. Franks, P. J. S. NPZ Models of Plankton Dynamics: Their Construction, Coupling to Physics, and Application. *Journal of Oceanography* **58**, 379–387 (2002).
15. Gentleman, W., Leising, A., Frost, B., Strom, S. & Murray, J. Functional responses for zooplankton feeding on multiple resources: a review of assumptions and biological dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography* **50**, 2847–2875 (2003).

16. Kuhn, A. M., Fennel, K. & Mattern, J. P. Progress in Oceanography Model investigations of the North Atlantic spring bloom initiation. *Progress in Oceanography* **138**, 176–193 (2015).
17. Fennel, K. A simple one-dimensional NPZD model with graphical user interface. https://github.com/katjafennel/OBM_Primer_NPZD_model. (2022).
18. Le Quéré, C. *et al.* Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* **11**, 2016–2040 (2005).
19. Orr, J. C. *et al.* Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Development* **10**, 2169–2199 (2017).
20. Lam, P. & Kuypers, M. M. M. Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Annual Review of Marine Science* **3**, 317–345 (2011).
21. Follows, M. J., Dutkiewicz, S., Grant, S. & Chisholm, S. W. Emergent Biogeography of Microbial Communities in a Model Ocean. *Science* **315**, 1843–1846 (2007).
22. Dutkiewicz, S. *et al.* Dimensions of marine phytoplankton diversity. *Biogeosciences* **17**, 609–634 (2020).
23. Armstrong, R. A. Grazing limitation and nutrient limitation in marine ecosystems: Steady state solutions of an ecosystem model with multiple food chains. *Limnology and Oceanography* **39**, 597–608 (1994).
24. Banas, N. S. Adding complex trophic interactions to a size-spectral plankton model: Emergent diversity patterns and limits on predictability. *Ecological Modelling* **222**, 2663–2675 (2011).
25. Galbraith, E. D., Gnanadesikan, A., Dunne, J. P. & Hiscock, M. R. Regional impacts of iron-light colimitation in a global biogeochemical model. *Biogeosciences* **7**, 1043–1064 (2010).
26. Denman, K. L. Modelling planktonic ecosystems : parameterizing complexity. *Progress in Oceanography* **57**, 429–452 (2003).
27. Haidvogel, D. B. & Beckmann, A. *Numerical Ocean Circulation Modeling*. (Imperial College Press, 1999).
28. Haltiner, G. J. & Williams, R. T. *Numerical prediction and dynamic meteorology*. (Wiley, 1980).
29. Roache, P. J. *Fundamentals of Computational Fluid Dynamics*. (Hermosa Publishers, 1998).
30. Brennan, C. E., Blanchard, H. & Fennel, K. Putting Temperature and Oxygen Thresholds of Marine Animals in Context of Environmental Change: A Regional Perspective for the Scotian Shelf and Gulf of St. Lawrence. *PLOS ONE* **11**, e0167411 (2016).
31. Claret, M. *et al.* Rapid coastal deoxygenation due to ocean circulation shift in the northwest Atlantic. *Nature Climate Change* **8**, 868–872 (2018).
32. Rutherford, K. & Fennel, K. Diagnosing transit times on the northwestern North Atlantic continental shelf. *Ocean Science* **14**, 1207–1221 (2018).
33. Bourgeois, T. *et al.* Coastal-ocean uptake of anthropogenic carbon. *Biogeosciences* **13**, 4167–4185 (2016).
34. Laurent, A., Fennel, K. & Kuhn, A. An observation-based evaluation and ranking of historical Earth system model simulations in the northwest North Atlantic Ocean. *Biogeosciences* **18**, 1803–1822 (2021).

35. Saba, V. S. *et al.* Enhanced warming of the Northwest Atlantic Ocean under climate change. *Journal of Geophysical Research: Oceans* **121**, 118–132 (2016).
36. Matear, R. J. Parameter optimization and analysis of ecosystem models using simulated annealing : A case study at Station P. *Journal of Marine Research* **53**, 571–607 (1995).
37. Fennel, K., Losch, M., Schroter, J. & Wenzel, M. Testing a marine ecosystem model : sensitivity analysis and parameter optimization. *Journal of Marine Systems* **28**, 45–63 (2001).
38. Friedrichs, M. A. M. *et al.* Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *Journal of Geophysical Research* **112**, 1–22 (2007).
39. Mattern, J. P. & Edwards, C. A. Simple parameter estimation for complex models — Testing evolutionary techniques on 3-dimensional biogeochemical ocean models. *Journal of Marine Systems* **165**, 139–152 (2017).
40. Laurent, A., Fennel, K., Wilson, R., Lehrter, J. & Devereux, R. Parameterization of biogeochemical sediment-water fluxes using in situ measurements and a diagenetic model. *Biogeosciences* **13**, 77–94 (2016).
41. Wilson, R. F., Fennel, K. & Paul Mattern, J. Simulating sediment–water exchange of nutrients and oxygen: A comparative assessment of models against mesocosm observations. *Continental Shelf Research* **63**, 69–84 (2013).
42. Thacker, W. C. The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans* **94**, 6177–6196 (1989).
43. Ward, B. A., Friedrichs, M. A. M., Anderson, T. R. & Oschlies, A. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems* **81**, 34–43 (2010).
44. Schartau, M. *et al.* Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling. *Biogeosciences* **14**, 1647–1701 (2017).
45. Gregg, W. W. *et al.* Skill assessment in ocean biological data assimilation. *Journal of Marine Systems* **76**, 16–33 (2009).
46. Bagniewski, W., Fennel, K., Perry, M. J. & D'Asaro, E. A. Optimizing models of the North Atlantic spring bloom using physical, chemical and bio-optical observations from a Lagrangian float. *Biogeosciences* **8**, 1291–1307 (2011).
47. Kuhn, A. M., Fennel, K. & Berman-frank, I. Modelling the biogeochemical effects of heterotrophic and autotrophic N2 fixation in the Gulf of Aqaba (Israel), Red Sea. *Biogeosciences* **15**, 7379–7401 (2018).
48. Mattern, J. P., Fennel, K. & Dowd, M. Periodic time-dependent parameters improving forecasting abilities of biological ocean models. *Geophysical Research Letters* **41**, 6848–6854 (2014).
49. Kitagawa, G. A Self-Organizing State-Space Model. *Journal of the American Statistical Association* **93**, 1203–1215 (1998).
50. Mattern, J. P. Visualizing parameter and state estimation for a zero-dimensional ocean biological model. https://github.com/jpmattern/obm-primer. (2022).
51. Evensen, G. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53**, 343–367 (2003).
52. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* **82**, 35–45 (1960).

53. Humpherys, J., Redd, P. & West, J. A Fresh Look at the Kalman Filter. *SIAM Review* **54**, 801–823 (2012).
54. Jazwinski, A. R. *Stochastic Processes and Filtering Theory*. (Academic Press, 1970).
55. Pham, D. T., Verron, J. & Roubaud, M. C. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems* **16**, 323–340 (1998).
56. van Leeuwen, P. J. A consistent interpretation of the stochastic version of the Ensemble Kalman Filter. *Quarterly Journal of the Royal Meteorological Society* **146**, 2815–2825 (2020).
57. Yu, L., Fennel, K., Bertino, L., Gharamti, M. El & Thompson, K. R. Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling. *Ocean Modelling* **126**, 13–28 (2018).
58. Yu, L. *et al.* Evaluation of nonidentical versus identical twin approaches for observation impact assessments: an ensemble-Kalman-filter-based ocean assimilation application for the Gulf of Mexico. *Ocean Science* **15**, 1801–1814 (2019).
59. Wang, B., Fennel, K. & Yu, L. Can assimilation of satellite observations improve subsurface biological properties in a numerical model? A case study for the Gulf of Mexico. *Ocean Science* **17**, 1141–1156 (2021).
60. Sakov, P. & Oke, P. R. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A: Dynamic Meteorology and Oceanography* **60**, 361–371 (2008).
61. Houtekamer, P. L. & Zhang, F. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review* **144**, 4489–4532 (2016).
62. Mattern, J. P., Song, H., Edwards, C. A., Moore, A. M. & Fiechter, J. Data assimilation of physical and chlorophyll a observations in the California Current System using two biogeochemical models. *Ocean Modelling* **109**, 55–71 (2017).
63. Wang, B., Fennel, K., Yu, L. & Gordon, C. Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. *Biogeosciences* **17**, 4059–4074 (2020).
64. Fiechter, J., Broquet, G., Moore, A. M. & Arango, H. G. A data assimilative, coupled physical–biological model for the Coastal Gulf of Alaska. *Dynamics of Atmospheres and Oceans* **52**, 95–118 (2011).
65. Moore, A. M. *et al.* The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part III – Observation impact and observation sensitivity in the California Current System. *Progress in Oceanography* **91**, 74–94 (2011).
66. Fennel, K. *et al.* Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health. *Frontiers in Marine Science* **6**, 89 (2019).
67. Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P. & Solidoro, C. Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea. *Ocean Modelling* **132**, 46–60 (2018).
68. Cossarini, G. *et al.* Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea. *Ocean Modelling* **133**, 112–128 (2019).
69. Ford, D. Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design. *Biogeosciences* **18**, 509–534 (2021).

70. Song, H., Edwards, C. A., Moore, A. M. & Fiechter, J. Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations. *Ocean Modelling* **106**, 159–172 (2016).
71. Courtier, P., Thépaut, J.-N. & Hollingsworth, A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* **120**, 1367–1387 (1994).
72. Gordon, N. J., Salmond, D. J. & Smith, A. F. M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. in *IEE Proceedings F-radar and signal processing* **140**, 107–113 (1993).
73. Mattern, J. P., Dowd, M. & Fennel, K. Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. *Journal of Geophysical Research: Oceans* **118**, 2746–2760 (2013).
74. Mattern, J. P., Yu, L., Wang, B. & Fennel, K. Ensemble Kalman filter application for an ocean biogeochemical model in an idealized 3-dimensional channel. https://github.com/bwang63/EnKF_3D_github. (2022).
75. Rothstein, L. M. . *et al.* Modeling Ocean Ecosystems: The PARADIGM Program. *Oceanography* **19**, 22–51 (2006).
76. Lehmann, M. K., Fennel, K. & He, R. Statistical validation of a 3-D bio-physical model of the western North Atlantic. *Biogeosciences* **6**, 1961–1974 (2009).
77. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106**, 7183–7192 (2001).
78. Jolliff, J. K. *et al.* Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems* **76**, 64–82 (2009).
79. Stow, C. A. *et al.* Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems* **76**, 4–15 (2009).
80. Doney, S. C. *et al.* Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems* **76**, 95–112 (2009).
81. Mattern, J. P., Fennel, K. & Dowd, M. Introduction and Assessment of Measures for Quantitative Model-Data Comparison Using Satellite Images. *Remote Sensing* **2**, 794–818 (2010).
82. Capotondi, A. *et al.* Observational Needs Supporting Marine Ecosystems Modeling and Forecasting: From the Global Ocean to Regional and Coastal Systems. *Frontiers in Marine Science* **6**, (2019).
83. Roemmich, D. *et al.* On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science* **6**, 439 (2019).
84. Chai, F. *et al.* Monitoring ocean biogeochemistry with autonomous platforms. *Nature Reviews Earth & Environment* **1**, 315–326 (2020).
85. Johnson, K. S. *et al.* Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans* **122**, 6416–6436 (2017).
86. Tanhua, T. *et al.* Ocean FAIR Data Services. *Frontiers in Marine Science* **6**, (2019).
87. Révelard, A. *et al.* Ocean Integration: The Needs and Challenges of Effective Coordination Within the Ocean Observing System. *Frontiers in Marine Science* **8**, (2022).
88. Khatiwala, S. *et al.* Global ocean storage of anthropogenic carbon. *Biogeosciences* **10**, 2169–2191 (2013).

89. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2021).
90. Gattuso, J.-P. *et al.* Ocean Solutions to Address Climate Change and Its Effects on Marine Ecosystems. *Frontiers in Marine Science* **5**, (2018).
91. National Academies of Sciences, Engineering, and M. *A Research Strategy for Ocean-based Carbon Dioxide Removal and Sequestration*. (National Academies Press, 2022). doi:https://doi.org/10.17226/26278
92. Aumont, O. & Bopp, L. Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles* **20**, (2006).
93. Jin, X., Gruber, N., Frenzel, H., Doney, S. C. & McWilliams, J. C. The impact on atmospheric CO2 of iron fertilization induced changes in the ocean's biological pump. *Biogeosciences* **5**, 385–406 (2008).
94. Oschlies, A., Koeve, W., Rickels, W. & Rehdanz, K. Side effects and accounting aspects of hypothetical large-scale Southern Ocean iron fertilization. *Biogeosciences* **7**, 4017–4035 (2010).
95. Dutreuil, S., Bopp, L. & Tagliabue, A. Impact of enhanced vertical mixing on marine biogeochemistry: lessons for geo-engineering and natural variability. *Biogeosciences* **6**, 901–912 (2009).
96. Bach, L. T. *et al.* Testing the climate intervention potential of ocean afforestation using the Great Atlantic Sargassum Belt. *Nature Communications* **12**, 2556 (2021).
97. Ilyina, T., Wolf-Gladrow, D., Munhoven, G. & Heinze, C. Assessing the potential of calcium-based artificial ocean alkalinization to mitigate rising atmospheric CO2 and ocean acidification. *Geophysical Research Letters* **40**, 5909–5914 (2013).
98. Feng, E. Y., Koeve, W., Keller, D. P. & Oschlies, A. Model-Based Assessment of the CO2 Sequestration Potential of Coastal Ocean Alkalinization. *Earth's Future* **5**, 1252–1266 (2017).
99. Siegel, D. A., DeVries, T., Doney, S. C. & Bell, T. Assessing the sequestration time scales of some ocean-based carbon dioxide reduction strategies. *Environmental Research Letters* **16**, 104003 (2021).
100. Schmidtko, S., Stramma, L. & Visbeck, M. Decline in global oceanic oxygen content during the past five decades. *Nature* **542**, 335–339 (2017).
101. Doney, S. C., Bopp, L. & Long, M. C. Historical and Future Trends in Ocean Climate and Biogeochemistry. *Oceanography* **27**, 108–119 (2014).
102. Bopp, L., Resplandy, L., Untersee, A., Le Mezo, P. & Kageyama, M. Ocean (de)oxygenation from the Last Glacial Maximum to the twenty-first century: insights from Earth System models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**, 20160323 (2017).
103. Takano, Y., Ito, T. & Deutsch, C. Projected Centennial Oxygen Trends and Their Attribution to Distinct Ocean Climate Forcings. *Global Biogeochemical Cycles* **32**, 1329–1349 (2018).
104. Levin, L. A. Manifestation, Drivers, and Emergence of Open Ocean Deoxygenation. *Annual Review of Marine Science* **10**, 229–260 (2018).
105. Oschlies, A., Brandt, P., Stramma, L. & Schmidtko, S. Drivers and mechanisms of ocean deoxygenation. *Nature Geoscience* **11**, 467–473 (2018).

106. Breitburg, D. *et al.* Declining oxygen in the global ocean and coastal waters. *Science* **359**, eaam7240 (2018).
107. Rabalais, N. N. *et al.* Eutrophication-driven deoxygenation in the coastal ocean. *Oceanography* **27**, 172–183 (2014).
108. Andrews, O., Buitenhuis, E., Le Quéré, C. & Suntharalingam, P. Biogeochemical modelling of dissolved oxygen in a changing ocean. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**, 20160328 (2017).
109. Cocco, V. *et al.* Oxygen and indicators of stress for marine life in multi-model global warming projections. *Biogeosciences* **10**, 1849–1868 (2013).
110. Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences* **10**, 6225–6245 (2013).
111. Couespel, D., Lévy, M. & Bopp, L. Oceanic primary production decline halved in eddy-resolving simulations of global warming. *Biogeosciences* **18**, 4321–4349 (2021).
112. Bahl, A., Gnanadesikan, A. & Pradal, M.-A. Variations in Ocean Deoxygenation Across Earth System Models: Isolating the Role of Parameterized Lateral Mixing. *Global Biogeochemical Cycles* **33**, 703–724 (2019).
113. Lévy, M., Resplandy, L., Palter, J. B., Couespel, D. & Lachkar, Z. Chapter 13 - The crucial contribution of mixing to present and future ocean oxygen distribution. in (eds. Meredith, M. & Naveira Garabato, A. B. T.-O. M.) 329–344 (Elsevier, 2022). doi:https://doi.org/10.1016/B978-0-12-821512-8.00020-7
114. Fennel, K. & Testa, J. M. Biogeochemical Controls on Coastal Hypoxia. *Annual Review of Marine Science* **11**, 105–130 (2019).
115. Peña, M. A., Katsev, S., Oguz, T. & Gilbert, D. Modeling dissolved oxygen dynamics and hypoxia. *Biogeosciences* **7**, 933–957 (2010).
116. Irby, I. D. *et al.* Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison. *Biogeosciences* **13**, 2011–2028 (2016).
117. Zhang, H., Fennel, K., Laurent, A. & Bian, C. A numerical model study of the main factors contributing to hypoxia and its interannual and short-term variability in the East China Sea. *Biogeosciences* **17**, 5745–5761 (2020).
118. Li, Y., Li, M. & Kemp, W. M. A Budget Analysis of Bottom-Water Dissolved Oxygen in Chesapeake Bay. *Estuaries and Coasts* **38**, 2132–2148 (2015).
119. Yu, L., Fennel, K., Laurent, A., Murrell, M. C. & Lehrter, J. C. Numerical analysis of the primary processes controlling oxygen dynamics on the Louisiana shelf. *Biogeosciences* **12**, 2063–2076 (2015).
120. Laurent, A., Fennel, K., Ko, D. & Lehrter, J. Climate Change Projected to Exacerbate Impacts of Coastal Eutrophication in the Northern Gulf of Mexico. *Journal of Geophysical Research: Oceans* **123**, (2018).
121. Ni, W., Li, M., Ross, A. C. & Najjar, R. G. Large Projected Decline in Dissolved Oxygen in a Eutrophic Estuary Due to Climate Change. *Journal of Geophysical Research: Oceans* **124**, 8271–8289 (2019).
122. LaBone, E. D., Rose, K. A., Justic, D., Huang, H. & Wang, L. Effects of spatial variability on the exposure of fish to hypoxia: a modeling analysis for the Gulf of Mexico. *Biogeosciences* **18**, 487–507 (2021).
123. de Mutsert, K., Steenbeek, J., Cowan, J. H. & Christensen, V. Using Ecosystem Modeling to Determine Hypoxia Effects on Fish and Fisheries BT - Modeling Coastal Hypoxia:

Numerical Simulations of Patterns, Controls and Effects of Dissolved Oxygen Dynamics. in (eds. Justic, D., Rose, K. A., Hetland, R. D. & Fennel, K.) 377–400 (Springer International Publishing, 2017). doi:10.1007/978-3-319-54571-4_14
124. Fennel, K. & Laurent, A. N and P as ultimate and proximate limiting nutrients in the northern Gulf of Mexico: implications for hypoxia reduction strategies. *Biogeosciences* **15**, 3121–3131 (2018).
125. Saraiva, S. *et al.* Baltic Sea ecosystem response to various nutrient load scenarios in present and future climates. *Climate Dynamics* **52**, 3369–3387 (2019).
126. Irby, I. D., Friedrichs, M. A. M., Da, F. & Hinson, K. E. The competing impacts of climate change and nutrient reductions on dissolved oxygen in Chesapeake Bay. *Biogeosciences* **15**, 2649–2668 (2018).
127. Kessouri, F. *et al.* Coastal eutrophication drives acidification, oxygen loss, and ecosystem change in a major oceanic upwelling system. *Proceedings of the National Academy of Sciences* **118**, e2018856118 (2021).
128. Laurent, A. & Fennel, K. Time-Evolving, Spatially Explicit Forecasts of the Northern Gulf of Mexico Hypoxic Zone. *Environmental Science & Technology* **53**, 14449–14458 (2019).
129. Matli, V. R. R. *et al.* Fusion-Based Hypoxia Estimates: Combining Geostatistical and Mechanistic Models of Dissolved Oxygen Variability. *Environmental Science & Technology* **54**, 13016–13025 (2020).
130. Yu, L. & Gan, J. Mitigation of Eutrophication and Hypoxia through Oyster Aquaculture: An Ecosystem Model Evaluation off the Pearl River Estuary. *Environmental Science & Technology* **55**, 5506–5514 (2021).
131. Feely, R. A., Doney, S. C. & Cooley, S. R. Ocean acidification: present conditions and future changes in a high-CO2 world. *Oceanography* **22**, 36–47 (2009).
132. Licker, R. *et al.* Attributing ocean acidification to major carbon producers. *Environmental Research Letters* **14**, 124060 (2019).
133. Doney, S. C., Busch, D. S., Cooley, S. R. & Kroeker, K. J. The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities. *Annual Review of Environment and Resources* **45**, 83–112 (2020).
134. Gehlen, M. *et al.* The fate of pelagic CaCO3 production in a high CO2 ocean: a model study. *Biogeosciences* **4**, 505–519 (2007).
135. Ilyina, T., Zeebe, R. E., Maier-Reimer, E. & Heinze, C. Early detection of ocean acidification effects on marine calcification. *Global Biogeochemical Cycles* **23**, (2009).
136. Krumhardt, K. M. *et al.* Coccolithophore Growth and Calcification in an Acidified Ocean: Insights From Community Earth System Model Simulations. *Journal of Advances in Modeling Earth Systems* **11**, 1418–1437 (2019).
137. Kwiatkowski, L. *et al.* Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences* **17**, 3439–3470 (2020).
138. Brady, R. X., Lovenduski, N. S., Yeager, S. G., Long, M. C. & Lindsay, K. Skillful multiyear predictions of ocean acidification in the California Current System. *Nature Communications* **11**, 2166 (2020).
139. Laurent, A. *et al.* Eutrophication-induced acidification of coastal waters in the northern Gulf of Mexico: Insights into origin and processes from a coupled physical-biogeochemical model. *Geophysical Research Letters* **44**, 946–956 (2017).

140. Hauri, C. *et al.* A regional hindcast model simulating ecosystem dynamics, inorganic carbon chemistry, and ocean acidification in the Gulf of Alaska. *Biogeosciences* **17**, 3837–3857 (2020).
141. Rutherford, K., Fennel, K., Atamanchuk, D., Wallace, D. & Thomas, H. A modelling study of temporal and spatial pCO2 variability on the biologically active and temperature-dominated Scotian Shelf. *Biogeosciences* **18**, 6271–6286 (2021).
142. Hauri, C. *et al.* Spatiotemporal variability and long-term trends of ocean acidification in the California Current System. *Biogeosciences* **10**, 193–216 (2013).
143. Hauri, C. *et al.* Modulation of ocean acidification by decadal climate variability in the Gulf of Alaska. *Communications Earth & Environment* **2**, 191 (2021).
144. Gruber, N., Boyd, P. W., Frölicher, T. L. & Vogt, M. Biogeochemical extremes and compound events in the ocean. *Nature* **600**, 395–407 (2021).
145. Dutkiewicz, S. *et al.* Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Climate Change* **5**, 1002–1006 (2015).
146. Pauly, D. & Christensen, V. Primary production required to sustain global fisheries. *Nature* **374**, 255–257 (1995).
147. Loukos, H., Monfray, P., Bopp, L. & Lehodey, P. Potential changes in skipjack tuna (Katsuwonus pelamis) habitat from a global warming scenario: modelling approach and preliminary results. *Fisheries Oceanography* **12**, 474–482 (2003).
148. Stock, C. A. *et al.* On the use of IPCC-class models to assess the impact of climate on Living Marine Resources. *Progress in Oceanography* **88**, 1–27 (2011).
149. Tittensor, D. P. *et al.* A protocol for the intercomparison of marine fishery\hack{\break} and ecosystem models: Fish-MIP v1.0. *Geoscientific Model Development* **11**, 1421–1442 (2018).
150. Lotze, H. K. *et al.* Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. *Proceedings of the National Academy of Sciences* **116**, 12907–12912 (2019).
151. Tittensor, D. P. *et al.* Next-generation ensemble projections reveal higher climate risks for marine ecosystems. *Nature Climate Change* **11**, 973–981 (2021).
152. Cheung, W. W. L. *et al.* Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Global Change Biology* **16**, 24–35 (2010).
153. Lam, V. W. Y., Cheung, W. W. L., Reygondeau, G. & Sumaila, U. R. Projected change in global fisheries revenues under climate change. *Scientific Reports* **6**, 32607 (2016).
154. IPCC. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. (Cambridge University Press, 2019).
155. IPBES. *Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. (IPBES secretariat, 2019). doi:10.5281/zenodo.3831673
156. Aumont, O., Maury, O., Lefort, S. & Bopp, L. Evaluating the Potential Impacts of the Diurnal Vertical Migration by Marine Organisms on Marine Biogeochemistry. *Global Biogeochemical Cycles* **32**, 1622–1643 (2018).
157. Archibald, K. M., Siegel, D. A. & Doney, S. C. Modeling the Impact of Zooplankton Diel Vertical Migration on the Carbon Export Flux of the Biological Pump. *Global Biogeochemical Cycles* **33**, 181–199 (2019).
158. Arnold, C. P. & Dey, C. H. Observing-Systems Simulation Experiments: Past, Present, and Future. *Bulletin of the American Meteorological Society* **67**, 687–695 (1986).

159. Halliwell, G. R. *et al.* Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico. *Journal of Atmospheric and Oceanic Technology* **31**, 105–130 (2014).
160. Griffies, S. M. *et al.* OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development* **9**, 3231–3296 (2016).
161. Chassignet, E. P. *et al.* DAMÉE-NAB: the base experiments. *Dynamics of Atmospheres and Oceans* **32**, 155–183 (2000).
162. Orr, J. C. On ocean carbon-cycle model comparison. *Tellus B: Chemical and Physical Meteorology* **51**, 509–510 (1999).
163. Séférian, R. *et al.* Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports* **6**, 95–119 (2020).
164. Canadell, J. G., Monteiro, P. M. S., Costa, M. H. & Others. Global Carbon and other Biogeochemical Cycles and Feedbacks. in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. V., M.-D. et al.) (Cambridge University Press, 2021).
165. Najjar, R. G. *et al.* Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). *Global Biogeochemical Cycles* **21**, (2007).
166. Matsumoto, K. *et al.* Evaluation of ocean carbon cycle models with data-based metrics. *Geophysical Research Letters* **31**, (2004).
167. Luettich Jr, R. A. *et al.* A test bed for coastal and ocean modeling. *Eos* **98**, (2017).
168. Yu, L., Fennel, K. & Laurent, A. A modeling study of physical controls on hypoxia generation in the northern Gulf of Mexico. *Journal of Geophysical Research: Oceans* **120**, 5019–5039 (2015).
169. Fennel, K. *et al.* Effects of model physics on hypoxia simulations for the northern Gulf of Mexico: A model intercomparison. *Journal of Geophysical Research: Oceans* **121**, 5731–5750 (2016).
170. Glover, D. M. *et al.* The US JGOFS data management experience. *Deep Sea Research Part II: Topical Studies in Oceanography* **53**, 793–802 (2006).
171. Baker, K. S. & Chandler, C. L. Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. *Deep Sea Research Part II: Topical Studies in Oceanography* **55**, 2132–2142 (2008).
172. Boyer, T. *et al.* Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25° grid. *International Journal of Climatology* **25**, 931–945 (2005).
173. Garcia, H. E., Boyer, T. P., Baranova, O. K. & Locarnini, R. A. World Ocean Atlas 2018: Product Documentation. Mishonov, A. Technical Editor. (2019).
174. Key, R. M. *et al.* A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles* **18**, (2004).
175. Olsen, A. *et al.* An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2020. *Earth System Science Data* **12**, 3653–3678 (2020).
176. Sloyan, B. M. *et al.* The Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP): A Platform for Integrated Multidisciplinary Ocean Science. *Frontiers in Marine Science* **6**, (2019).

177. Wanninkhof, R. *et al.* A Surface Ocean CO2 Reference Network, SOCONET and Associated Marine Boundary Layer CO2 Measurements. *Frontiers in Marine Science* **6**, (2019).
178. Benway, H. M. *et al.* Ocean Time Series Observations of Changing Marine Ecosystems: An Era of Integration, Synthesis, and Societal Applications. *Frontiers in Marine Science* **6**, (2019).
179. Buitenhuis, E. T. *et al.* MAREDAT: towards a world atlas of MARine Ecosystem DATa. *Earth System Science Data* **5**, 227–239 (2013).
180. Lombard, F. *et al.* Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Frontiers in Marine Science* **6**, (2019).
181. Bittig, H. C. *et al.* A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. *Frontiers in Marine Science* **6**, (2019).
182. Maurer, T. L., Plant, J. N. & Johnson, K. S. Delayed-Mode Quality Control of Oxygen, Nitrate, and pH Data on SOCCOM Biogeochemical Profiling Floats. *Frontiers in Marine Science* **8**, (2021).
183. Harrison, C. S., Long, M. C., Lovenduski, N. S. & Moore, J. K. Mesoscale Effects on Carbon Export: A Global Perspective. *Global Biogeochemical Cycles* **32**, 680–703 (2018).
184. Katavouta, A. & Thompson, K. R. Downscaling ocean conditions with application to the Gulf of Maine, Scotian Shelf and adjacent deep ocean. *Ocean Modelling* **104**, 54–72 (2016).
185. Debreu, L., Marchesiello, P., Penven, P. & Cambon, G. Two-way nesting in split-explicit ocean models: Algorithms, implementation and validation. *Ocean Modelling* **49–50**, 1–21 (2012).
186. Löptien, U. & Dietze, H. Reciprocal bias compensation and ensuing uncertainties in model-based climate projections: pelagic biogeochemistry\hack{\break} versus ocean mixing. *Biogeosciences* **16**, 1865–1881 (2019).
187. Eyring, V. *et al.* Taking climate model evaluation to the next level. *Nature Climate Change* **9**, 102–110 (2019).
188. Kwiatkowski, L. *et al.* Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change* **7**, 355–358 (2017).
189. Terhaar, J., Kwiatkowski, L. & Bopp, L. Emergent constraint on Arctic Ocean acidification in the twenty-first century. *Nature* **582**, 379–383 (2020).

Glossary

Plankton functional groups: groups of planktonic organisms that share similar traits, e.g., size, biogeochemical function, or elemental requirements. These groups are defined to simplify the diversity of planktonic communities while capturing their essential biogeochemical functions in OBMs.

State variables: a set of variables that fully characterizes a model's dynamical state such that its future behavior can be calculated provided any external inputs are known. Each variable that belongs to this set is a state variable.

Initial condition: the complete set of state variables at one instant in time. Model integration starts from an initial condition.

External forcing: includes all prescribed inputs that are needed to determine the evolution of a model's state and are not calculated internally by the model.

Model parameter: a constant that is usually specified at the beginning of model integration and determines the dynamical behavior of the model.

Integration time: the simulated length of model integration. It varies from months to decades in regional models and 100s of years in ESMs.

Spin up: the initial period of a model simulation during which the model adjusts from its initial state to a new state according to the internal model dynamics and subject to external forcing. The spin-up period ranges from a few months or years for regional models to one or a few hundred years for global models.

Projection: a simulation into the future that goes significantly beyond timescale for which models have demonstrated predictive or forecast skill, e.g., ESM simulations to the end of the current century or longer.

A priori knowledge: assumptions about ocean processes (represented by the equations of an ocean model and its parameters, initial, and boundary conditions) that is available before data assimilation is applied.

Parameter optimization: determines the most likely value of poorly known model parameters based on the agreement of model output with observations.

Optimal parameters: result from parameter optimization and are the parameter values that minimize the *cost function* in a parameter optimization problem.

Cost function: measures the misfit between observations and their model counterparts in a least-squares sense.

State estimation: obtains the optimal model state by combining the information contained in the model equations and the available observations.

Variational method: aimed at obtaining the best fit, in a least-squares sense, between model and observations by minimizing a *cost function*. It can be applied to parameter and state estimation problems.

Sequential method: the model state and possibly its parameters are updated through an alternating sequence of *forecast steps*, when the model is integrated forward in time, and *update* or *analysis steps,* when the model state and its parameters are updated using observations.

Control vector: contains all the values to be optimized during data assimilation. It can include model parameters, the full model state, or a subset thereof, or a combination of both.

A posteriori error: an estimate of the error in the solution of an optimization problem given the observations and numerical solution technique applied.

Least-squares: a measure of misfit between observations and the model equivalents of these observations that sums the squared distances between them.

Decorrelation scale: the e-folding scale of the autocorrelation function of the property under consideration, in other words the distance or period over which the autocorrelation decreases by a factor of 1/e.

Eutrophication: excessive supply of plant nutrients to a body of water, often due to input from land**.**