Version of Record: https://www.sciencedirect.com/science/article/pii/S0009254119305327
Manuscript_b03bbf2245be32045736db73958c7ae4

AWESOME OCIM: A simple, flexible, and powerful tool for modeling elemental cycling in the oceans

Seth G. John[1], Hengdi Liang[1], Tom Weber[2], Tim DeVries[3], Francois Primeau[4], Keith Moore[4], Mark Holzer[5], Natalie Mahowald[6], Wilford Gardner[7], Alexey Mishonov[8], Mary Jo Richardson[7], Yannice Faugere[9], Guillaume Taburet[9]

[1]University of Southern California, Department of Earth Sciences, Los Angeles, CA, 90089, USA.
[2] Department of Earth and Environmental Science, University of Rochester, Rochester, NY 14627, USA.
[3] Department of Geography and Earth Research Institute, University of California, Santa Barbara, CA, 93106, USA.
[4] Department of Earth System Science, University of California Irvine, Irvine CA 92617, USA
[5] Dept. of Applied Mathematics, School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia.
[6] Department of Earth and Atmospheric Science, Cornell University, Ithaca, NY, 14853, USA.
[7] Department of Oceanography, Texas A&M University, College Station, TX 77843, USA.
[8] Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, National Centers for Environmental Information (NCEI), NOAA affiliate, Silver Spring, MD 20910, USA.
[9] CLS, Environment Monitoring Department, 31520 Ramonville, France.

© 2019 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

Abstract
Here we present a new flexible modeling tool for simulating the distribution of tracers in the modern ocean. A Working Environment for Simulating Ocean Movement and Elemental cycling within an Ocean Circulation Inverse Model, the AWESOME OCIM, is a transport matrix model (TMM) which is specifically designed to be easy, accessible, and intuitive, even for scientists without prior modeling experience. The AWESOME OCIM comes with a variety of selectable biogeochemical functions, including sources (atmospheric dust, hydrothermal vents, and seafloor nepheloid layers), internal cycling processes (biological uptake, remineralization, and scavenging), and sinks (radioactive decay and burial of particles in the sediments). A wide variety of elements can be simulated through different combinations of this suite of processes. We anticipate that the AWESOME OCIM will be a valuable tool for interpreting transect data from ocean surveys, particularly the trace-elements and isotopes distributions mapped by the ongoing GEOTRACES program. This manuscript provides an introduction to the philosophical, mathematical, and functional basis of the AWESOME OCIM.

Keywords:
OCIM; TMM; GEOTRACES

1. Introduction

1.1 Models of varying complexity

The use of simple box models has a long and important history in chemical oceanography (Broecker and Peng, 1982; Sigman and Boyle, 2000; Toggweiler et al., 2003b; Toggweiler et al., 2003a; Sarmiento and Gruber, 2006). The simplest versions of such models might contain just one box for the surface ocean and another for the deep ocean (Tyrrell, 1999), whereas other applications require additional boxes to distinguish between ocean basins and watermasses (e.g. Toggweiler, 1999), or geochemically distinct regions like anoxic zones (e.g. Deutsch et al., 2004). These simple models are typically used to explore the biogeochemical cycling of just a few tracers at a time, such as carbon or major nutrients. Their simplicity means that they can be easily built and manipulated by an individual researcher, and such models have often appeared in the oceanographic literature as a small component of manuscripts which also present new data or new theories (e.g. Van Geen et al., 1991; Parekh et al., 2004; Hunter and Boyd, 2007; Yang et al., 2014; Takano et al., 2014, among many others).

At the other end of the 'complexity spectrum' lie global Earth System Models (ESMs) which include realistic representations of coupled ocean-atmosphere circulation over ~10^5-10^6 boxes, complex ecosystem interactions, and the cycling of carbon, physical tracers, and multiple nutrients. The physical ocean circulation component of such models is a general circulation model, or GCM, which solves a discretized approximation of the primitive equations for fluid transport. The relative complexity of these models tends to lead to different modes of scientific interaction and use. Specifically, ESMs are typically run and maintained by large groups of scientists, each of whom is responsible for just a small part of the overall modeling effort.

Regardless of their complexity, all numerical models have advantages compared to conceptual or qualitative models including: 1) The ability to quantitatively test hypotheses, for example the ability to validate or reject hypotheses by showing whether they can produce a good match to observations given realistic constraints, and 2) to make quantitative predictions, for example to quantify fluxes of carbon or nutrients in the modern ocean or to predict how fluxes might have been different in the past or future ocean.

1.1 Transport matrix models (TMMs) and the ocean circulation inverse model (OCIM)

Over the past decade, transport matrix models (TMMs) have emerged as a new class of models which share characteristics both of simple box models and more complex GCMs. TMMs can simulate the biogeochemistry of tracers on a global scale, with realistic ocean circulation and relatively high spatial resolution (~10^5 grid cells) (Fig. 1). Yet their use of linear algebra techniques means that they achieve an extraordinary increase in computational speed. A wide variety of factors influence the computational time necessary to solve a GCM, including spatial and temporal resolution, the number of variables being transported in the model, whether the physical and biogeochemical processes are coupled in a single model run, and the "spin-up" time necessary for a model to reach equilibrium, but it is common for GCMs to require weeks to months of computation time to converge on steady-state tracer distributions. In contrast, using a TMM one can solve for steady-state tracer distributions in several seconds to minutes, depending on the number of tracers and the resolution of the TMM.

TMMs describe the circulation of the ocean in the form of a matrix which specifies the transfer of water between grid cells of the model. Given that each grid cell only exchanges directly with a few neighboring cells, the transport matrix is very sparse (most elements are zero), making it easier to store and manipulate. Biogeochemical processes can similarly be described in matrix-vector notation, using

equations that prescribe transfer of the tracer from one model grid cell to another, or a source or sink of the tracer within each grid cell. A TMM-formulated biogeochemical model thus consists of a single matrix equation, which can be solved at steady-state using linear algebra techniques, which are much more computationally efficient than the numerical integration methods employed for GCMs.

Early applications of TMMs converted the circulation of GCMs into a matrix form to exploit the computational efficiency of matrix vector products and matrix inverses of sparse matrices (Primeau, 2003; Khatiwala et al., 2005; Primeau, 2005). Similar techniques have now been applied to extract transport matrices from a wide array of ocean GCMs including the Parallel Ocean Program version 2 (POP2) model that has been used in the NCAR Community Earth System Model (Bardin et al., 2014), the ECCO version 2 ocean state estimate model (Zanna et al., 2019), the ocean component of the University of Victoria Earth System Model (Kvale et al., 2017) and the ACCESS ocean model (Chamberlain et al., 2019). Enabled by TMM techniques, scientists have addressed a wide array of scientific problems such as constraining the global cycling of carbon and nutrients (e.g. Kwon and Primeau, 2006; Khatiwala et al., 2009; Weber and Deutsch, 2010; Wang et al., 2019).

Compared to using circulations from GCMs, a better representation of the influence of oceans circulation on tracer distributions can be achieved by incorporating information about the distribution of physical and biogeochemical tracers in the ocean (e.g. Schlitzer, 1993; Schlitzer, 2007). The Ocean Circulation Inverse Model (OCIM) is one such technique, which starts with a circulation based on a steady-state near-geostrophic approximation, and then optimizes for additional non-geostrophic terms to best match the distribution of tracers such as potential temperature, salinity, radiocarbon (^{14}C), and CFCs (DeVries and Primeau, 2011; DeVries, 2014). This optimization requires many thousands ($\sim 10^3$-10^4) of iterations in order to determine a circulation which best matches tracer distributions to corresponding observations. This large number of simulations is accomplished by casting the circulation in TMM form, enabling the use of fast matrix algebra solvers. Because it is constrained by data, the OCIM provides a more realistic representation of the influence of ocean circulation on tracer distributions than free-running GCMs.

The OCIM is a circulation model, not a biogeochemistry model. But it can be easily implemented as the circulation component of an ocean biogeochemistry model, and has been used for numerous applications in chemical oceanography including being used to quantify ocean CO_2 uptake (DeVries, 2014; DeVries et al., 2017), reveal spatial variability in phytoplankton carbon to nutrient uptake ratios (DeVries and Deutsch, 2014; Teng et al., 2014), and constrain the export and fate of organic carbon (Weber et al., 2016; Roshan and DeVries, 2017; DeVries and Weber, 2017). The OCIM has also been used to constrain the global cycling and distribution of a wide range of elements including phosphorous (Primeau et al., 2013; DeVries et al., 2014), silicon (Holzer et al., 2014; DeVries et al., 2017), nitrogen (DeVries et al., 2012; Weber and Deutsch, 2012; DeVries et al., 2013; Weber and Deutsch, 2014; Wang et al., 2019), and zinc (Roshan et al., 2018; Weber et al., 2018).

1.2 Practical considerations for using OCIM

Besides its computational efficiency, OCIM has several other features that facilitate its use by non-experts. A few of these features are:

1. OCIM is lightweight and memory-efficient. The OCIM takes up only several MB of disk space and RAM. This contrasts with typical GCMs which require GB of storage and memory space.
2. The native language of OCIM is MATLAB, which is commonly used throughout the scientific community, and easily learned by those without prior programming experience (though it is a

proprietary language and thus the AO is not fully open-source). This contrasts with typical GCMs that are written in Fortran or other expert-level programming languages.

3. OCIM can be run on a laptop computer. This contrasts with typical GCMs that must be run on supercomputers or large computer clusters, requiring significant IT infrastructure.

All of these features lower the barrier of entry for scientists without previous modeling experience or the support of a large modeling group or computational facility.

The computational efficiency of the OCIM also lends it to different modeling approaches compared to GCMs. Simulations of biogeochemical tracers in a GCM may take weeks to reach steady-state, so there is a limit to the number of different model formulations which can be tested. This results in the tendency to equip a model with every process that might affect the tracer distribution before it is run. While this approach has the advantage of including a wide variety of potentially important mechanisms, it can make it difficult to derive mechanistic insight into which processes are most important in explaining observed phenomena.

The efficiency of the OCIM allows multiple model formulations to be tested, making it more suitable for hypothesis testing and model optimization than a traditional GCM. A typical approach with the OCIM is to begin with a bare-minimum representation of biogeochemical processes, adding additional processes only when they are necessary to correct for systematic model-data discrepancies. This process allows one to distinguish the processes which are fundamental to the global distribution of a tracer, from other processes which may occur in the ocean, but are not first-order controls on global distributions. Because the OCIM circulation is data-constrained, one can with reasonable assurance ascribe model-data discrepancies to missing biogeochemical processes, rather than inaccurate physical transport. Similarly, because of OCIM's computational efficiency, one can set a range of possible values for several different model parameters, and then search through thousands of different models in order to determine the combination of parameters that best matches observations. A recent example of this approach used the OCIM to search for surface ocean nitrogen uptake and N_2-fixation rates that provide the best match to observed nutrient distributions (e.g. Wang et al., 2019).

The OCIM also has limitations compared to GCMs. Perhaps most importantly, the OCIM does not include any temporal variability in circulation, including seasonal variability. Instead, it simulates an annual mean circulation constrained by the observational data. Another limitation is that the spatial resolution of the OCIM is relatively coarse: at 2^o horizontal resolution with 24 vertical levels it is roughly half the resolution of typical global ocean biogeochemistry models that are based on GCMs. Finally, OCIM has been developed to match the modern climatological mean ocean circulation. In contrast to GCMs, OCIM cannot be used to predict the response of ocean circulation to changes in climate forcing. All of these considerations therefore limit the OCIM to use for exploring and predicting the steady-state distribution of tracers in the modern ocean at large spatial scales.

No model is perfect, a sentiment nicely summarized by Gerard Roe as "People don't understand the Earth, but they want to, so they build a model, and then they have two things they don't understand" (Wohlforth, 2004). Like any model, the OCIM has both strengths and weaknesses. Still, we believe that the relative simplicity of the OCIM makes it ideal for use by non-expert modelers, and that adding OCIM modeling to the 'toolkit' used by observational scientists can lead to scientific advances.

1.3 The AWESOME OCIM

Here we describe a new model, A Working Environment for Simulating Ocean Movement and Elemental cycling within the Ocean Circulation Inverse Model; the AWESOME OCIM, or the AO for short. This

model is intended to put the tools of the OCIM into the hands of observationalists and others with little or no prior experience in ocean modeling. Ultimately, we hope that this can provide an alternative to the simplified box models or 1-d models often now used for interpretation in observation-focused papers. In order to achieve this goal, we focus on three specific criteria:

1) **Code is written as transparently as possible, so that users can more easily understand the code they are running and more easily modify that code for new purposes.** Variable names are meant to be easily understood and interpreted, extensive commenting is used to describe the purpose of each line of code, and all mathematical operations are performed as intuitively as possible, even at the expense of less succinct code and sometimes slightly longer computational times.
2) **The model is written in a modular format, so that users can easily add new biogeochemical processes.** Each biogeochemical process is handled in a separate Matlab function. This way, users can model an element which is affected by several biogeochemical processes by running several functions in series. New biogeochemical processes can be incorporated into the model in a similar fashion, and shared between users as self-contained functions.
3) **The model comes pre-loaded with observational datasets and model output.** In order to streamline work with the AO, the model comes pre-loaded with a variety of observational datasets (e.g. 2017 GEOTRACES Intermediate Data Product and 2009 World Ocean Atlas) stored in the AO model grid, as well as important output from other models (e.g. predictions of global ocean productivity and dust input).

In this manuscript we describe the mathematical basis of the AO and work through a few examples of how the model might be used in research. An appendix provides additional detail, briefly discussing each of the biogeochemical processes and datasets which come with the pre-packaged model. An explicit goal of the AO is to make global ocean modeling tools available to scientists with little prior modeling experience. This manuscript is therefore written with that audience in mind, prioritizing detailed and simple explanations which will be useful for scientists without much prior experience in linear algebra mathematics, ocean modeling, or coding. The code is available at https://github.com/hengdiliang/AWESOME-OCIM-v1.1 and http://www.mtel.rocks/mtel/awesomeOCIM.html, and in the supplementary materials.

2. Mathematical basis of the AO

The AO represents biogeochemical processes as a set of linear equations, and solves for the steady-state distribution of a tracer impacted by those processes. The AO uses simple matrix division in order to solve the model, as opposed to an iterative linear algebra technique used by some other OCIM-biogeochemical models, and thus every process which affects the tracer of interest within each model grid cell must either be zero-order (an input or removal of the tracer which is not dependent on the concentration of the tracer) or first-order (an input or removal which is directly proportional to the concentration of the tracer). First-order processes may depend either on the concentration of an element within that grid cell, such as loss by radioactive decay, or on the concentration the element within another grid cell, such as the circulation input of an element from a neighboring cell or the input of an element due to remineralization of biogenic particles formed in the surface ocean.

The linear algebra techniques used to construct such a model can be easily appreciated by considering a 4-box model of the ocean (Fig. 2). One of these boxes does not contain ocean, reflecting topography of the seafloor, so that the model is described by three coupled equations to represent the

three ocean boxes. When considering only water transport, the change in the concentration of a hypothetical element E within any given box is equal to the mass transfer rate of water coming in from neighboring boxes (T) multiplied by the concentration of element E within that water (E), minus the amount of water leaving the box multiplied by the concentration of element E within that box (assuming that all boxes have the same volume). For the 3-box model, where box 3 is neglected because it does not contain ocean, the three water transport equations are therefore:

$$\frac{\partial E_1}{\partial t} = -T_{12} \cdot E_1 + T_{12} \cdot E_2 + 0 \cdot E_4 \quad (1)$$

$$\frac{\partial E_2}{\partial t} = T_{12} \cdot E_1 - (T_{12} + T_{24}) \cdot E_2 + T_{24} \cdot E_4 \quad (2)$$

$$\frac{\partial E_4}{\partial t} = 0 \cdot E_1 + T_{24} \cdot E_2 - T_{24} \cdot E_4 \quad (3)$$

The assumption of steady-state is that $\frac{\partial E}{\partial t}$ is equal to zero, and by convention the zero-order terms (those not multiplied by a concentration E) are moved to the right-hand side of the equations to yield:

$$-T_{12} \cdot E_1 + T_{12} \cdot E_2 + 0 \cdot E_4 = 0 \quad (4)$$

$$T_{12} \cdot E_1 - (T_{12} + T_{24}) \cdot E_2 + T_{24} \cdot E_4 = 0 \quad (5)$$

$$0 \cdot E_1 + T_{24} \cdot E_2 - T_{24} \cdot E_4 = 0 \quad (6)$$

This set of three linear equations can then be represented in matrix-vector form, by defining matrix A as the set of nine constants from the left-hand-sides of the equations, and vector b as the three right-hand-side constants:

$$A = \begin{bmatrix} -T_{12} & T_{12} & 0 \\ T_{12} & (-T_{12}-T_{24}) & T_{24} \\ 0 & T_{24} & -T_{24} \end{bmatrix} \quad (7)$$

$$b = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (8)$$

As written above, there would be no single solution to these equations which contain three unknowns (E_1, E_2, and E_3) but only two independent equations (because any of them can be determined from the other two). However, as shown in Figure 2, the model becomes constrained as additional first-order terms are incorporated into A and zero-order terms (i.e. fixed sources and sinks) are incorporated into b to represent biogeochemical processes. Having incorporated all additional processes into A and b, the full physical-biogeochemical system can then be solved by matrix inversion:

$$e = A^{-1} b \quad (9)$$

where e is a vector containing all the values of E in each box:

$$e = \begin{bmatrix} E_1 \\ E_2 \\ E_4 \end{bmatrix} \quad (10)$$

The AO code follows a similar approach, with the main program calling individual functions which each add a representation of an individual process to *A* and *b*, then solving for the distribution of the hypothetical element E using matrix division. In this fashion the AO includes functions to represent a wide variety of biogeochemical processes, including sources such as dust and hydrothermal vents, internal cycling processes such as scavenging and biological uptake and remineralization, and sinks such as radioactive decay and sedimentary burial. For a more detailed description of the biogeochemical functions and datasets which are included with the AO, some of which are used in the examples below, see the Appendix.

3. Examples

Three examples discussed below illustrate different ways in which the AO can be used in scientific research, with a focus on functions that are easy to implement with minimal modifications of the AO code as it is originally released. These examples demonstrate ways that "off-the-shelf" AO simulations might be incorporated for simple hypothesis testing into data-oriented papers, highlighting ways in which certain processes either can or cannot explain the observations (e.g. John et al 2018 using a preliminary version of the AO). Alternatively, more extensive work with the AO might form the basis of a modeling-focused manuscript aimed at synthesizing a global view of an element cycle, in which the user might develop new functions to address specific aspects of that element's biogeochemical cycle.

3.1 Using the AO to model simple processes

Even without modifying the AO to create a realistic global model of an element, there can still be value in comparing data to a simple model of one or two processes. For example, a plume of Fe-rich waters has been observed emanating from the Peru margin on the US GEOTRACES GP15 transect (Fig. 3). The most obvious source of Fe on the margin would be reducing sediments, yet the depth of the plume is much deeper than the oxygen minimum (John et al., 2018). Alternatively, it could be that the Fe plume is created by non-reductive dissolution of Fe from oxic sediments underneath the oxygen minimum, for example in nepheloid layers. To quickly test this second idea, a model can be run with a source of Fe in nepheloid layers and a sink through first-order removal (modeled using the code for radioactive decay) with a residence time of 50 years (consistent with the timescale at which dissolved Fe is lost by particle scavenging in the oceans) (Tagliabue et al., 2016). Using the AO's graphical user interface (GUI), a model like this can be set up with just a few clicks and run in under a minute. The choice of the absolute rate at which tracer is added from nepheloid layers does not matter here, since the model is only seeking to reproduce the spatial distribution of nepheloid layer input, not the absolute magnitude.

The GUI can then be used to look at model output and compare it to data. First, the GUI can be used to plot Fe concentrations from the 2017 GEOTRACES Intermediate Data Product along the GP15 transect, showing the plume of Fe near the continental margin. Then model output can be plotted similarly, showing that nepheloid layer input does not produce a similar plume (Fig 3). The failure of the model to reproduce data suggests that nepheloid layer input, at least as parameterized in the AO (see

Supplemental Section for description of nepheloid layer input), cannot be responsible for the observed Fe plume on the GP15 transect. A similar approach was recently used by John et al. (2018) to argue that some additional process was necessary to explain the Fe plume, such as release of isotopically light Fe from reducing sediments and subsequent reversible scavenging onto sinking POC. The input of Fe from reducing sediments cannot be modeled with the AO in its off-the-shelf form, but could be easily incorporated into the AO by writing a new function to describe this input in a new *b* matrix.

3.2 Using boundary conditions to model regional processes

Boundary conditions can be used to build regional models of element cycling, or to identify regionally important processes which are missing from existing models. Global models have an important place in research, as they encourage study of the processes which dominate the global-scale cycling of elements, and yield globally-consistent tracer distributions. In theory, if one were interested in local-scale processes, the local distribution could be compared to output from a good global model. However, because no model is perfect, the local data will inevitably be compared to an imperfect global model, making it more difficult to disentangle the effect of local processes from inaccuracies propagated from outside the region of interest. Boundary conditions provide a means to better isolate a particular region of study.

Boundary conditions work by setting the concentration of an element equal to observations along the edges of the region to be studied. By setting up a model with only boundary conditions, and no other biogeochemical processes, one would be able to determine the 'circulation-only' distribution of an element within the region, which is transported in from the specified boundaries of that region. For example, the GA03 and GA10 GEOTRACES cruises transected east-west in the North Atlantic and South Atlantic, respectively, enclosing the tropical Atlantic between these two boundaries. By imposing the boundary conditions of the observed Fe concentrations along both transects, we can infer what would be the distribution of Fe in the tropical Atlantic if there were no sources or sinks of Fe in the tropics (Fig. 4). This model output can then be compared to data from the GA02 transect which crossed north-south through the tropical Atlantic. By doing so, we see that the boundary condition model predicts much higher Fe concentrations in the deep ocean than were observed, and lower concentrations at equatorial mid-depths (500-1500 m). We can therefore ascribe the differences to loss by scavenging in the deep ocean, and input to the mid-depths of the tropical ocean due to dust deposition and/or shelf sources, with redistribution over depth by biological uptake and regeneration.

3.3 Modeling isotopes with the AO

By keeping a few simple rules in mind, it is possible to use the AO to predict the distribution of stable isotopes. Because all processes in the AO are linear, it is not necessary to account for the relative abundance of the various isotopes (e.g. the fact that N is 99.6% ^{14}N and 0.4% ^{15}N). Instead, one can first run a version of the model which simulates the average behavior of all isotopes in order to obtain the tracer concentration distribution throughout the ocean, and then run a separate model just to represent the behavior of a heavier isotope. The stable isotope ratio for each grid cell can then be calculated as:

$$\delta^{heavy/light} = \left(\frac{e_{iso}}{e} - 1\right) \cdot 1000 \qquad (11)$$

where e is the concentration of your element determined from the base model and e_{iso} is the concentration determined from the isotope model.

Using this approach, the effect of different isotope ratios between tracer sources and kinetic isotope effects during tracer sinks and cycling processes can be tested and visualized using the AO. The isotopic composition of source processes can be represented by changing the supply rate of heavy isotope tracer to the model, relative to the standard supply rate. For example, to represent a hydrothermal input with a $\delta^{heavy/light}$ of -1 ‰, you would supply the bulk tracer at a rate specified by $R_{hydrothermal}$ (see Appendix for details of how the AO represents these processes), and the heavy isotope tracer at a rate of $R_{hydrothermal} \cdot 0.999$. Kinetic isotope effects during first-order processes are similarly modeled by changing the rate constant in the isotope model. For example, a -5 ‰ isotope effect during biological uptake could be represented by specifying a value of α for biological uptake in the function `bioalpha`, and then running a version of the model for a heavy isotope with $\alpha \cdot 0.995$. We have demonstrated this approach by simulating the distribution of a hypothetical nutrient-like tracer that has a 5 ‰ isotopic fractionation during assimilation (Fig. 5).

4. Conclusions

Directly "off-the-shelf", the AWESOME OCIM provides a simple interface for users to test how various biogeochemical processes influence large-scale tracer distributions in a model with realistic global circulation. We hope that such tools provide users with a way to develop better intuition for oceanographic processes, aiding data interpretation. We further hope that the simplicity of the model will make it appropriate for use in classes, both as an introduction to oceanographic processes and as an introduction to ocean modeling.

We anticipate that most research applications of the AO will require further modification by the user. Specifically, we imagine that the biogeochemical functions included here will be a good starting point for modeling most elements in the ocean, but that every element will be affected by a few processes which are not considered here, and thus new code will have to be written to reflect that biogeochemistry. The modular way in which the A and b matrices are constructed from a series of functions representing individual process means that new processes can be easily be included. Similarly, this modularity means that functions describing new biogeochemical processes can be shared between users, and combined with existing biogeochemical functions. Indeed, the range of studies which could be undertaken with the AO ranges far beyond just the "elements" discussed here, and could include biological molecules such as proteins and DNA, dissolved gasses, inorganic quantities such as alkalinity, or other constituents which occur in seawater.

It should be noted that the AO is a specific application of the OCIM, and as such is separate and distinct from the OCIM. The OCIM is a data-constrained ocean circulation model, which can be implemented in any ocean biogeochemical model of unlimited complexity. The OCIM is actively maintained, developed, and improved, and new versions are released periodically at https://tdevries.eri.ucsb.edu/models-and-data-products/. The AO is a library of MATLAB codes that run linear biogeochemical models, using a transport matrix derived from a particular version of the OCIM (DeVries, 2014) to ensure high-fidelity tracer transport. The AO could easily be run with transport matrices derived from other models, and a major goal of future versions of the AO is the incorporation of new circulations, particularly those reflecting the past and future ocean states. By doing this, we hope that the AO can become a tool for paleoceanographic and geological research, and for climate science. In this way we hope that the AO can become a tool not only for modern oceanographic research but also for exploration of past and future Earth conditions.

Acknowledgements

This work is supported by National Science Foundation grants OCE-1658436 to SGJ, OCE-1658042 to TW, OCE-1658392 to TD, and OCE-1658380 to JKM and FP.

Appendix I: User guide for the Awesome OCIM

Appendix I, Section 1. Biogeochemical processes available with the AO

To facilitate ease of use, the AO comes with a number of biogeochemical processes already incorporated. These are all stored in the `srcsnk` folder, and they include sources of elements to the ocean such as dust fluxes and hydrothermal input, sinks from the ocean such as radioactive decay, and internal cycling processes such as biological uptake, regeneration, and scavenging. Each process is reviewed briefly here, while a more complete description of how they operate can be found in the comments embedded within the MATLAB code for each function.

`bioredfield` represents the biological uptake and remineralization of an element with a fixed stoichiometry compared to P, sometimes referred to as a 'Redfield ratio'. The uptake rate of P in the surface ocean (P_{up}) is taken from a separate model of the global P cycle (Weber et al., 2018). The uptake of element E (E_{up}; μmole m^{-3} y^{-1}) is then given by:

$$E_{up} = P_{up} \cdot R \tag{12}$$

Where R is the stoichiometric ratio of element E compared to P in biological particles. It should be noted that this formulation doesn't have any cap on the uptake of element E, and thus it is possible to create regions where E is less than zero in the surface ocean.

The `bioredfield` function also allows the user to specify the remineralization length scale based on the Martin curve by specifying the b value for remineralization where the flux (F) at any depth (z) is determined by the flux at the compensation depth (F_c) and depth of the compensation depth (z_c) by:

$$F = F_c \left(\frac{z}{z_c}\right)^{-b} \tag{13}$$

In the AO, z_c is the top of the third layer (around 73 m), because AO assumes productivity without remineralization in the top two layers.

Even for the deepest ocean, some portion of the tracer will not be remineralized in the water column and will therefore be left over to sink out of the bottom grid cell. There is then a choice about whether to return that E back into the dissolved phase, a process analogous to sedimentary remineralization of remaining organic material, enabled by setting `sedremin.on` to 1, or if that E should be removed to from the model in a process akin to burial in the sediments by setting `sedremin.on` to 0.

`bioalpha` represents the biological uptake and remineralization of an element. It is based on the concept of alpha (α), a relative uptake rate constant for an element compared to phosphate (Huested et al.,

1981). The uptake rate of phosphate (P_{up}) is used to calculate a first-order rate constant for P uptake (k_P) in each surface-ocean grid cell (the top two layers of the model) based on:

$$P_{up} = k_P \cdot P \quad (14)$$

where P is the phosphate concentration in that grid cell. The uptake rate for element E (E_{up}; μmole m^{-3} y^{-1}) is then determined using the relationship:

$$k_E = \alpha \cdot k_P \quad (15)$$

and

$$E_{up} = k_E \cdot E \quad (16)$$

Thus, for example, if 20% of the phosphate in a given grid-cell was taken up annually, and $\alpha$for element E was 2, then ~40% of element E would be taken up from that grid cell annually. Water column and sedimentary remineralization are performed in a similar fashion as with `bioredfield`.

One advantage of `bioalpha` is that it reflects an experimental observation about the uptake of many trace-metals by phytoplankton including Zn, Cd, Mn, and Co, that metal:P ratios increase roughly in proportion to the concentration of the element in media (Sunda, 2012, and references therein). A second advantage is that absolute uptake rates decrease as the concentration of E decreases, so that concentrations of E do not go below zero.

`boundcon` sets a boundary condition based on one or more GEOTRACES transects from the 2017 Intermediate Data Product (Schlitzer et al., 2018). With these boundary conditions, regardless of the concentration of tracer which enters a grid cell, the model will output water with the concentration specified by the boundary condition. The mechanism by which this is achieved is to specify an extraordinarily large source of your tracer into each boundary condition grid cell, paired with an extraordinarily large first-order loss of tracer from that grid cell. The concentration within that grid cell is then set by the balance between the large source and the large sink, and is nearly unaffected by any additional gain or loss of tracer due to circulation and other biogeochemical processes. By default, the first order rate constant for loss is 10^6 y^{-1}, and the source of tracer is $c \cdot 10^6$ y^{-1}, where c is the specified concentration of the tracer within the boundary condition grid cell.

`conc` sets the mean ocean concentration of an element. It works by creating an extraordinarily small source of E into every box, and then establishing an extraordinarily small first-order loss of E. By default, the first-order rate constant for loss is 10^{-6} y^{-1}, so the mean ocean concentration of the element will be set by imposing a source of $c \cdot 10^{-6}$ y^{-1} where c is additional concentration flux added to each box each year. This approach works because the timescale at which E is added and removed is much longer than the mixing timescale of the ocean (10^6 years, versus a mixing timescale of ~10^3 years), so that the impact on the concentration of E in any single box is overwhelmed by other processes and is therefore unnoticeable.

`conc` should only be used in the case that there are no other external sources or sinks of E to the ocean (i.e. a closed system). For example, a model which includes only transport and reversible

scavenging, and in which all scavenged E which reaches the bottom box is released back into the dissolved phase (sedremin turned on) is a closed system. In an open-system model the mean concentration is determined by the magnitude of the external sources and sinks, and `conc` should not be used. For example, a model which included hydrothermal input and burial of scavenged E, the global ocean concentration would be set by the balance of the magnitudes of the hydrothermal source and the burial sink, and `conc` should not be used.

`decay` removes element E from the ocean according a first-order loss such as radioactive decay, where the first-order rate constant for loss is calculated by setting the half-life of the element.

`dust` adds an element according to the predicted input of seven different kinds of aerosol deposition (mineral, fire, fossil fuel, biofuel, seasalt, plants, and volcanic; Fig. 6). These maps are stored in the structure AEROSOLDEP as seven 91x180x24 matrices which describe aerosol inputs in units of mg m^{-2} y^{-1}. For the mineral dust, seasalt, and plant sources, the mass of the aerosols is the total mass. For the fire, fossil fuel and biofuel aerosols, the mass represents the black carbon contribution from these sources. For volcanic aerosols it is the sulfur contribution. The input of your element of interest is then specified by *dust.R* (μmole m^{-3} y^{-1} E / mg m^{-3} y^{-1} aerosol), the ratio of your element to the aerosol input, by:

$$source = dust.R \cdot aerosol \qquad (17)$$

where *aerosol* is the deposition of aerosol to the surface layer ocean. The maps of aerosol input are 3-dimensional model output from a variety of types of aerosol, contributed by the lab of Natalie Mahowald, as described in Brahney et al. (2015) and Chien et al. (2016).

`hydrothermal` adds E to the ocean according to predictions of submarine hydrothermal ^{3}He input stored in HEFLUX.mat (Fig. 7). Helium fluxes are based on seafloor spreading rates, as originally outlined by Farley et al. (1995) and further modified in an OCIM framework by allowing for regionally variable differences in the relationship between spreading rate and ^{3}He (Holzer et al., 2017). The HEFLUX data already accounts for the volume of the grid cell into which hydrothermal fluids are injected, so that the input of element E is given by:

$$source = hydrothermal.K \cdot HEFLUX \qquad (18)$$

where *hydrothermal.K* (μmole m^{-3} y^{-1} E / μmole m^{-3} y^{-1} He) is the globally constant molar ratio of E to ^{3}He.

`nepheloid` adds element E in proportion to predicted nepheloid layer particle concentrations (Fig. 8). Nepheloid layers are found at the bottom of the oceans where particles are resuspended from the sediments into the water column and/or advected laterally. Several factors contribute to the formation of nepheloid layers including bottom currents, sediment type, regional topography accelerating flows locally, and the dissipation of ocean energy in the form of decaying eddies, internal waves and tides. We provide two different ways to parameterize nepheloid layer input.

One estimate of the global distribution of nepheloid layers is based on over 8800 observations of particle concentrations made either with the Lamont Thorndike Nephelometer or with transmissometers

attached to CTDs (Gardner et al., 2018b; Gardner et al., 2018a). We use the integrated concentration of particles over the entire nepheloid layer (μg cm^{-2}) to determine the total particle concentration in each column of the model grid, and then apportion that particle concentration by depth based on the calculated fraction of bottom sediments ocurring in each grid cell (*ao.SEAFLOOR*). Finally, the areal concentration of particles (μg m^{-2}), is converted to a volume concentration (μg m^{-3}) by dividing by the height of each grid cell. Particle concentration data mapped in Gardner et al. (2018b) were interpolated with an 8°search radius, but for this model we interpolated the data globally with a 25° search radius and second-degree smoothing using an inverse distance method. Data coverage in some parts of the ocean such as the Western North Atlantic is quite good, while other parts of the ocean such as the Arctic, Southern Ocean, and mid-ocean basins is less complete. Even this wide search radius did not allow for interpolation in many parts of the Arctic, and thus we assume no nepheloid layers in those regions.

An alternative estimate of global nepheloid layers depends on a relationship between surface eddy kinetic energy (EKE) and particle density in near-bottom nepheloid layers. Surface EKE is based on a 25-year record of sea-surface height ending in 2018, in units of cm^2 s^{-2} (Taburet et al., 2019). Eddy kinetic energy (cm^{-2} s^{-1}) is converted into the total areal delivery of eddy energy to each column (m^2 cm^{-2} s^{-1}) in the model grid by multiplying by the column area, and then converted to the volume concentration of eddy energy ((m^2 cm^{-2} s^{-1}) m^{-3}) by dividing by the volume of each grid cell. This volume concentration of energy is converted to a nepheloid particle concentration by taking into account the fractional amount of bottom sediments in each grid cell (*ao.SEAFLOOR*). Please note that any models using satellite EKE are requested to include the acknowledgement 'This study has been conducted using E.U. Copernicus Marine Service Information'. The relationship between surface and deep EKE and nepheloid layer intensity has been explored in depth by Gardner et al. (2017; 2018a; 2018b). Briefly, there is a relationship in much of the ocean, with clear exceptions beneath the Kuroshio current and near the equator where EKE is elevated, yet there are no observations of high nepheloid layer particle concentrations (Fig. 8).

Nepheloid layer intensities are given in units of μg m^{-3} for the observations and (m^2 cm^{-2} s^{-1}) m^{-3} for EKE stored in the AO as *NEPH.OBS* and *NEPH.EKE*, respectively. The input of element E is then given by:

$$input = K \cdot NEPH.TYPE \quad (19)$$

where *TYPE* is either *OBS* or *EKE*, and *K* is the ratio of E concentration input per nepheloid intensity.

nonrevscavPOC removes E from the ocean in direct proportion to the POC concentration in each grid cell. POC concentrations were determined in a separate model of global nutrient cycling (Weber et al., 2018), so they are higher under upwelling regions and they decrease with depth according to the Martin curve. The loss of E is determined as:

$$loss = nonrevscavPOC.K \cdot POC \quad (20)$$

where *nonrevscavPOC.K* is the fraction of E lost per year per mmole m^{-3} of POC, and *POC* is the concentration of POC.

revscav assumes reversible scavenging onto particles with a uniform distribution throughout the world ocean. The AO does not include an explicit representation of particles, which is to say that scavenging does not actually remove E from the dissolved phase. Instead, scavenging is built into the model by specifying that a certain portion of the dissolved pool can sink downwards into the underlying grid cells, in proportion to the inferred adsorbed concentration. This can be expressed with a distribution coefficient (*K*) where:

$$revscav.K = sinking/total. \quad (21)$$

The amount of that element which sinks out of each box is determined by:

$$sinkout = revscav.K \cdot revscav.w/h \quad (22)$$

where *revscav.w* (m y^{-1}) is the sinking rate of particles and *h* (m) is the height of that box. Similarly, the amount sinking into any box is given by the amount leaving the box above, adjusted to a concentration flux based on the box volumes.

While the model describes reversible scavenging using a fairly straightforward mathematical description, there are some quirks to be considered. First, because particle representation is implicit, there will be errors in the calculated dissolved and particulate concentrations. Crucially, these errors are insignificant when the fraction adsorbed is small and they grow larger as more E is adsorbed. For example, *K*=0.9 means that 90% of E within each box is available for scavenging and transfer to underlying boxes, but the model solution for E will report the concentration of the entire dissolved+particulate pool in the box, not just the 10% of E which one would expect to find remaining in the dissolved pool. For full accuracy, first-order processes which affect only the dissolved phase should be multiplied by (1-*revscav.K*), though for most elements the amount of adsorbed E is insignificant and the errors introduced by this formulation are similarly insignificant. Second, *K* values of greater than 1 are not disallowed in this model, meaning it is mathematically possible to transfer >100% of element E out of any box per year. The model will usually tolerate *K*>1, and will simply make up the difference in lost E by mixing, however such a parameter choice would violate common sense. Finally, it should be noted that the amount of E which sinks from a box depends on *K*·*w* so that, for example, a *K* of 0.1 with a sinking speed of 1000 m y^{-1} will produce the same model behavior as a K of 0.2 and a sinking speed of 500 m y^{-1}.

This formulation is a slightly simplified version of the full reversible scavenging equations that were implemented in a model of hydrothermal iron dispersion using the OCIM (Roshan et al., "Hydrothermal iron is trapped in the deep ocean", in review).

revscavPOC is similar to `revscav`, except that scavenging is proportional to POC concentration, rather than being uniform throughout the ocean so that:

$$sinkout = revscavPOC.K \cdot revscavPOC.w/h \cdot POC \quad (23)$$

where *revscavPOC.K* is the ratio of sinking E to total E, as described in `revscav`. The caveats regarding *K* and *w* discussed for `revscav` also apply here.

Appendix I, Section 2. Datasets included in the AO

In order to make the model more easily usable, it comes pre-loaded with several datasets which have been fit to the AO grid by averaging all datapoints which occur within an AO grid cell. These include datasets based on model output, as well as some based on observations. These various datasets and model output are generally compatible. For example, the phosphate uptake rates for the AO (`P_UP_WJ18`) were determined using the same water transport as used by the AO (`water_transport`), while minimizing the misfit between AO model output phosphate (`PO4_WJ18`) and AO observed phosphate (`WOAPO4`). The files `AEROSOLDEP`, `NEPHLOID`, and `HEFLUX` contain the information needed to run the functions dust, hydrothermal, and nepheloid, respectively. Other files and folders are described below:

`GEOTRACES_2017_IDP` is a folder containing many individual files with data from the GEOTRACES 2017 Intermediate Data Product (Schlitzer et al., 2018). The data has been re-gridded onto the AO grid, where all data points within a grid cell were averaged together to produce the final concentration. This folder contains data for 79 parameters (of the total 470 parameters included in the 2017 IDP) including many of the most commonly studied dissolved trace metals.

`GLODAP` is a folder containing data on global ocean alkalinity and DIC from the GLODAP data product (Lauvset et al., 2016; Olsen et al., 2016). GLODAP data is not currently used by any functions in the AO, but it is never the less included here so that it may be used in future work.

`WJ18` is a folder containing model output from a global nutrient cycling model. The particular model used for the AO is described in Weber and John et al. (2018), though similar models of nutrient cycling have been constructed for previous work. The model minimizes misfit between observed phosphate from the 2009 World Ocean Atlas and the model output phosphate (`PO4_WJ18`) (Fig. 9). Other concentrations reported from the model include dissolved organic phosphorous (`DOP_WJ18`) and particulate organic carbon (`POC_WJ18`), which is derived from particulate organic phosphorous assuming a 106:1 C:P ratio in sinking organic matter. Rates derived from the nutrient cycling model and included with the AO include the phosphorous uptake rate in the surface ocean (`P_UP_WJ18`) and the phosphorous remineralization rate in the deep ocean (`P_REM_WJ18`).

`WOA09` is a folder containing information about the properties of seawater worldwide based on a compilation of numerous datasets, often referred to as a 'climatology'. Data is taken from the 2009 World Ocean Atlas (Levitus et al., 2010), averaged to produce an annual mean, and re-gridded for the AO. Datasets include nitrate (`WOANO3`), oxygen (`WOAO2`), phosphate (`WOAPO4`), salinity (`WOASAL`), silicate (`WOASI`), and temperature (`WOATEMP`).

`ao` is a structure containing information about the AO grid, including information such as which grid cells contain ocean, the latitude and longitude of the grid cells, and masks for different ocean basins and GEOTRACES sections (Fig. 10). A full list of the `ao` parameters and a complete description of them is included as Appendix II.

`water_transport` contains the ocean circulation transport matrix. The standard transport matrix is derived from the CTL version of the OCIM as described in DeVries (2014). While OCIM is built on a square grid, the circulation is not as simple as having each box exchange water with only the six

neighbors sharing a face. The OCIM uses a hybrid upwind-centered scheme in the horizontal and a 3rd-order semi-upwind scheme in the vertical, as well as diffusion which is rotated along and across isopycnals, so that many more distant grid cells are connected to each other in the transport operator than just the nearest neighbors (see Devries 2014 for details). The AO transport matrix is the same as was used to generate other model-derived parameters such as phosphate uptake rate in the surface ocean.

AIII. Using the AO

AIII.1 Running the AO

The AO can set up and run using the script `setup_single`. This script creates first creates a structure `do`, which tells the model what you would like it to do (a structure is a variable which can contain sub-variables). The `do` structure enables one or more biogeochemical processes in your model in the form `do.function.var`, where function is the name of a function which changes the A or b matrix to reflect a biogeochemical process stored in the `srcsnk` folder (e.g. `bioredfield`), and `var` is the variables needed to run that function (e.g. R, the Redfield ratio to apply). If you wish to run a function to enable biogeochemistry within your model, you must set `do.function.on` to a value of 1, in order to turn that functionality on. If you don't wish to enable a function, you can either set `do.function.on` to 0, or simply don't include any reference to that function in your `do` strutcture. If the user wishes write their own new biogeochemical functions, the function must be saved to the `srcsnk` folder, and the function can then be called from `do` in `setup_single`.

A graphical user interface (GUI) provides a simple way to begin exploring the AO without modifying any code at all. The GUI can be launched by running the script `launchGUI`, after which the user can select various processes to model, or can plot the output from earlier models in a variety of different ways.

Appendix I, Section 3. Open-system and closed-system models

The AO can be used to create both 'open system' models, which include external sources and sinks of E to the ocean, and 'closed system' models which include only internal biogeochemical cycling and movement of your tracer. Before setting up a model, it is important to consider whether you are constructing an open or closed system model, and ensure that you have chosen an appropriate combination of processes to achieve mass balance. In many cases, it is possible to combine open-system and closed-system together in the model in a way which is mathematically solvable, but biogeochemically unrealistic.

Open-system models must include both a source and a sink of your tracer. Sources to the ocean are typically zero-order with respect to your tracer, which is to say that the magnitude of the source does not depend on the concentration of your tracer. For example, dust, hydrothermal, and nepheloid layer inputs are all parameterized to deliver a certain amount of your tracer to the ocean each year based on an input ratio compared to predetermined input fields. Of course, such a model must also include a sink by which the tracer is lost from the ocean. Zero-order sinks for removal of your tracer cannot be used, as they would result in an underdetermined model because there is no mechanism to set the global concentration of E in the ocean. Instead, removal of E from the ocean is typically accomplished by a sink which depends directly on the concentration of E, such as radioactive decay or biological uptake followed by burial of particulate material which reaches the ocean floor. In this way, both the global distribution of the tracer is determined, as well as the global mean ocean concentration.

Closed-system models represent only internal cycling of the tracer within the ocean. The first thing to know about closed-system models in the AO is that they cannot be solved as a true closed-system. This is because a true closed-system model would contain an equal number of equations (one for each grid cell) and unknowns (the concentration in each grid cell) and would therefore be underdetermined. Another way to state this same problem is that there is a truly closed-system model can represent the redistribution of E throughout the ocean by processes such as biological uptake and remineralization and circulation, but this same redistribution could be achieved for any different total amount of E. Thus, closed-system models must be run with the `conc` function which sets the global mean ocean concentration by including very small sources and sinks of E. This approach has a direct analogy in the real ocean, which is that the ocean is not a truly closed system with respect to any element on geological timescales, instead its concentration is set by the balance of sources and sinks. However, the distribution of elements with a residence time in the oceans longer than the circulation timescale of the oceans (~1000 years) can be accurately modeled using closed-system processes, while the global mean concentration of an element in the oceans is set by the slow input and removal of elements on long timescales. For example, the distribution of a nutrient-type element in the ocean might be modeled using `bioalpha` to reflect the redistribution of E in the oceans by biological uptake and remineralization, while `conc` would have to be used to set the global mean ocean concentration.

Appendix II. Description of parameters stored in the grid-describing structure `ao`.

(Note that in all of the descriptions below, a 'shoebox' refers to a 3-dimensional rectangular array which is the size of the model grid (91x180x24)).

`nocn`: 200160 The number of grid cells in the model with contain ocean.

`iocn`: [200160×1 double] Indices of grid cells which contain ocean within the shoebox.

`OCN`: [91×180×24 double] A shoebox with 1 where there is ocean and 0 where there is not.

`nanOCN`: [91×180×24 double] A shoebox with 1 where there is ocean and NaN where there is not.

`EQNPOS`: [91×180×24 double] A shoebox which contains the 'equation position' for each ocean cell. For example, `ao.EQNPOS`(50,100,2) is 15833, which is to say that when the linear algebra equations are written to describe the water flow and biogeochemical cycling with the model, the processes affecting the grid cell at position 50,100,2 in the shoebox are described in the 15833rd row and column of the A matrix, and the 15833rd row of the b matrix.

`nbtm`: [10441 1] The number of grid cells which are at the bottom of the ocean.

`ibtm`: [10441×1 double] Indices of the grid cells at the bottom of the ocean within the shoebox.

`BTM`: [91×180×24 double] A shoebox with 1 for grid cells at the bottom of the ocean and 0 elsewhere.

`nanBTM`: [91×180×24 double] A shoebox with 1 for grid cells at the bottom of the ocean and NaN elsewhere.

`SEAFLOOR`: [91×180×24 double] A shoebox containing the fractional portion of the bottom face of the grid cell which contains seafloor based on ETOPO2 global bathymetry.. BTM describes the geometry of the model grid, and therefore has only one 'bottom' grid cell, in each column. However, in reality a continental slope might have bottom sediments which occur at many depths within any 2°x2° column, which is accounted floor by the high-resolution bathymetry described in SEAFLOOR.

`Seafloor`: [200160×1 double] A list of the fractional portion of seafloor in every grid cell which contains ocean.

`nsurf`: [10441 1]The number of grid cells which are at the surface of the ocean.

isurf: [10441×1 double] Indices of the grid cells at the surface of the ocean within the shoebox.
SURF: [91×180×24 double] A shoebox with 1 for grid cells at the surface of the ocean and 0 elsewhere.
nanSURF: [91×180×24 double] A shoebox with 1 for grid cells at the surface of the ocean and NaN elsewhere.
lat: [1×91 double] The latitudes of grid cells within the shoebox, from -89.011 to +89.011.
Lat: [200160×1 double] A list of the latitudes of all 200160 grid cells which contain ocean.
LAT: [91×180×24 double] A shoebox with the latitudes in each grid cell.
lon: [1×180 double] The latitudes of grid cells within the shoebox, from 1 to 359.
Lon: [200160×1 double] A list of the longitudes of all 200160 grid cells which contain ocean.
LON: [91×180×24 double] A shoebox with the longitudes in each grid cell.
depth: [1×24 double] The depths of grid cells within the shoebox, for all 24 depth levels.
Depth: [200160×1 double] A list of the depths of all 200160 grid cells which contain ocean.
DEPTH: [91×180×24 double] A shoebox with the depths in each grid cell.
height: [1×24 double] The heights of grid cells within the shoebox, for all 24 depth levels.
Height: [200160×1 double] A list of the heights of all 200160 grid cells which contain ocean.
HEIGHT: [91×180×24 double] A shoebox with the heights in each grid cell.
Vol: [200160×1 double] A list of the volumes of all 200160 grid cells which contain ocean.
VOL: [91×180×24 double] A shoebox with the volumes in each grid cell.**:**
ATL: [91×180×24 double] A shoebox with the 'Atlantic mask', containing 1 for every grid cell which contains Atlantic ocean.
PAC: [91×180×24 double] A shoebox with the 'Pacific mask', containing 1 for every grid cell which contains Pacific ocean.
IND: [91×180×24 double] A shoebox with the 'Indian mask', containing 1 for every grid cell which contains Indian ocean.
ARC: [91×180×24 double] A shoebox with the 'Arctic mask', containing 1 for every grid cell which contains Arctic ocean.
MED: [91×180×24 double] A shoebox with the 'Mediterranean mask', containing 1 for every grid cell which contains Mediterranean ocean.
GTmasks: [1×1 struct] A structure containing masks for many GEOTRACES sections. In most cases, GEOTRACES stations are spaced further apart than the distance between grid cells, and typically GEOTRACES profiles do not have enough vertical resolution to fill in values for each grid cell vertically. In order to create continuous sections, the masks are therefore filled in vertically to include all depths at which there is ocean, and grid cells are filled in horizontally to create complete unbroken sections.

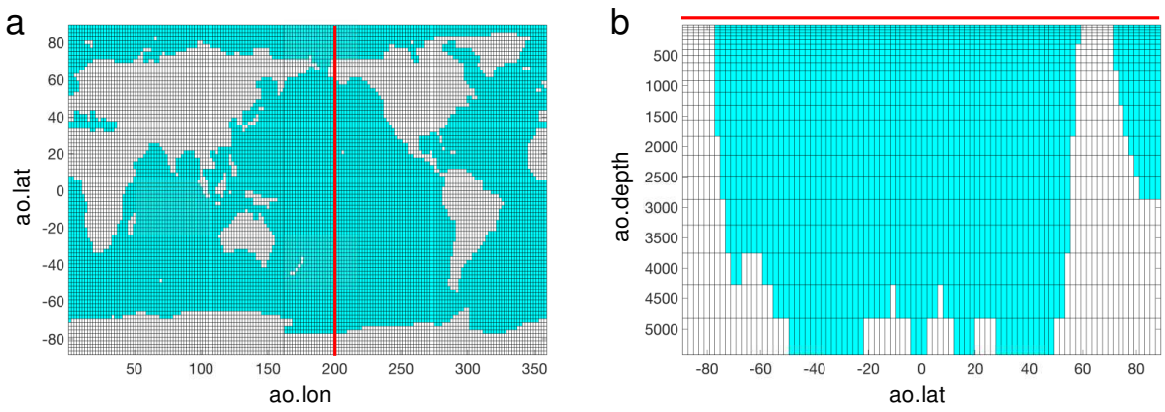


Figure 1. The AWESOME OCIM uses a model grid with 2° latitude by 2° longitude boxes, and 24 vertical levels which are more closely spaced towards the surface ocean. The horizontal section is shown at 201° longitude, as indicated by the red line.

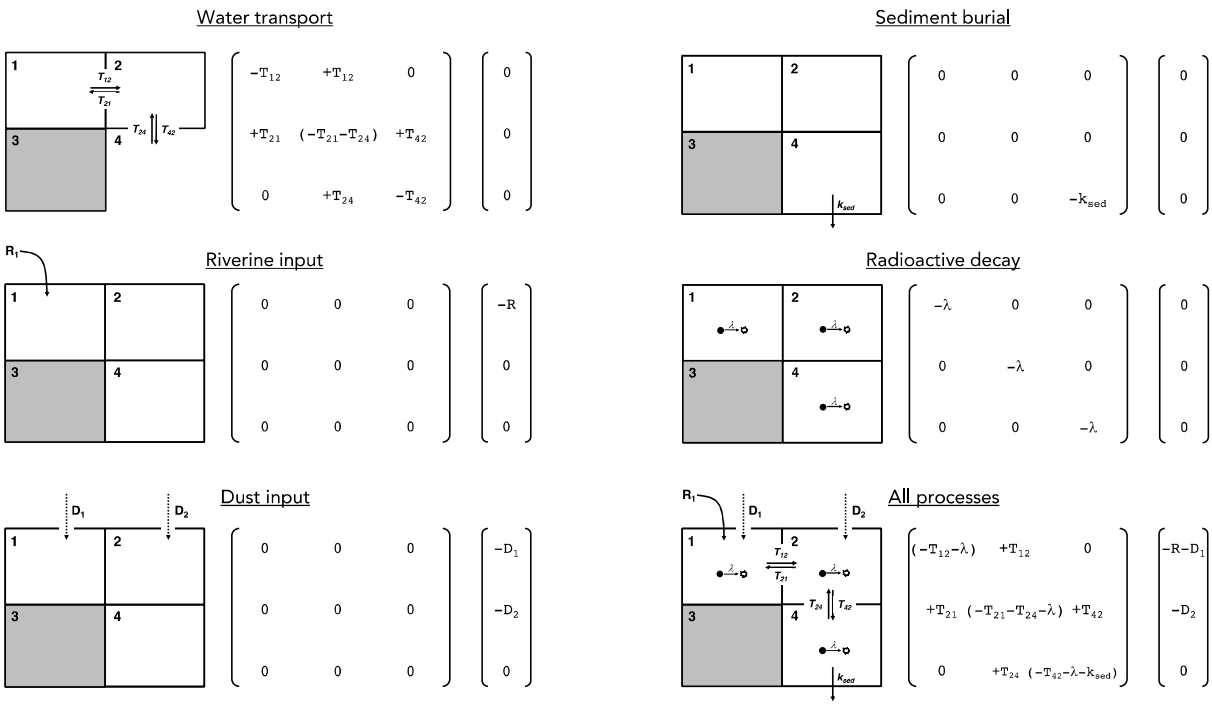


Figure 2. Various processes are described for a simple 4-box model of the ocean. For each process we show both an illustration of how this process moves tracers within the ocean, along with the corresponding A and b matrices which would be used to build a linear-algebra version of the model.

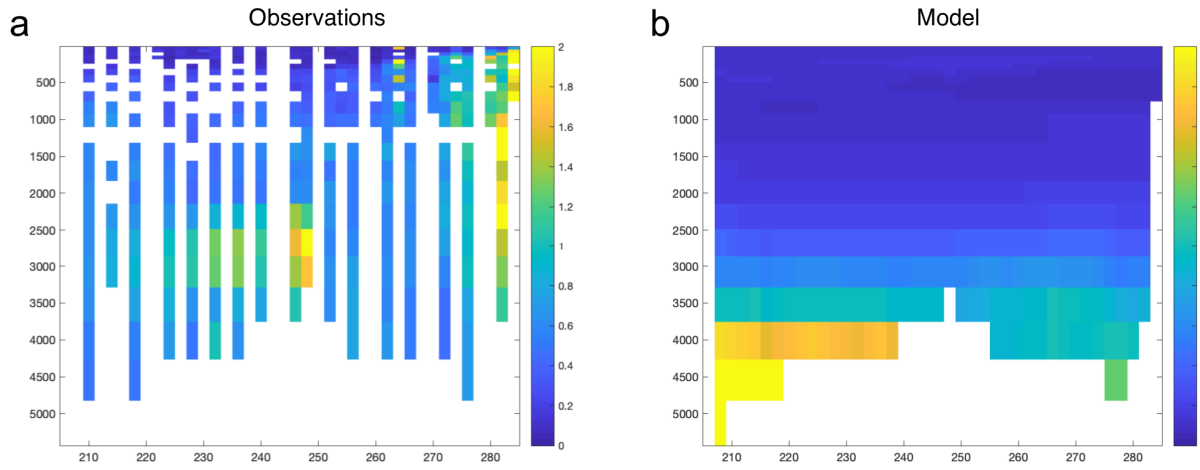


Figure 3. Observations of Fe concentrations from the GEOTRACES GP15 transect in the Eastern Tropical South Pacific (a) can be compared to a simple model with input of a tracer from nepheloid layers and loss with a half-life of 50 years (b). The large differences between Fe distribution and tracer distribution quickly demonstrate that non-reductive dissolution from nepheloid is unlikely to explain the plume of Fe observed in the Eastern basin near Peru. Both the model run and plotting were performed using the GUI.

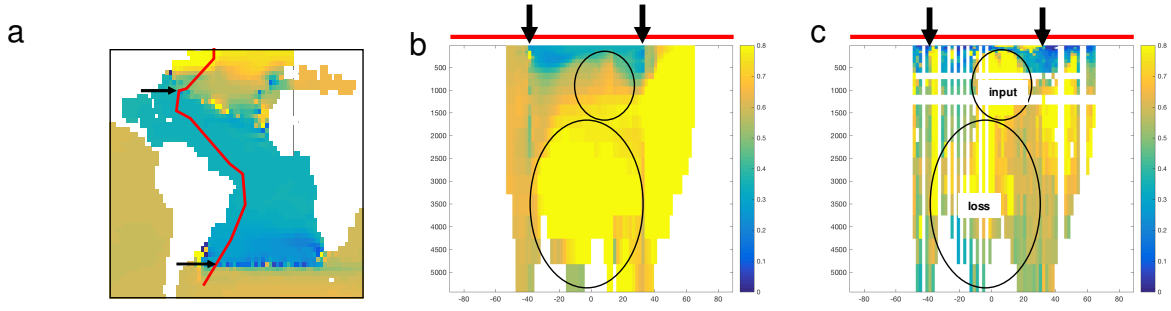


Figure 4. Boundary conditions of Fe concentration were imposed along the GA03 transect in the North Atlantic Ocean and the GA10 transect in the South Atlantic (black arrows), enclosing a region in the tropical and equatorial Atlantic which was transected by the GA02 cruise (red line) (a). A model which includes these boundary conditions, but no other biogeochemical processes shows the concentrations of Fe which would be circulated to the GA02 transect in the absence of any other sources or sinks (b). This model output can be compared to observed Fe concentrations along the GA02 transect in order to identify regions of input and loss (c). Figures were produced using the GUI.

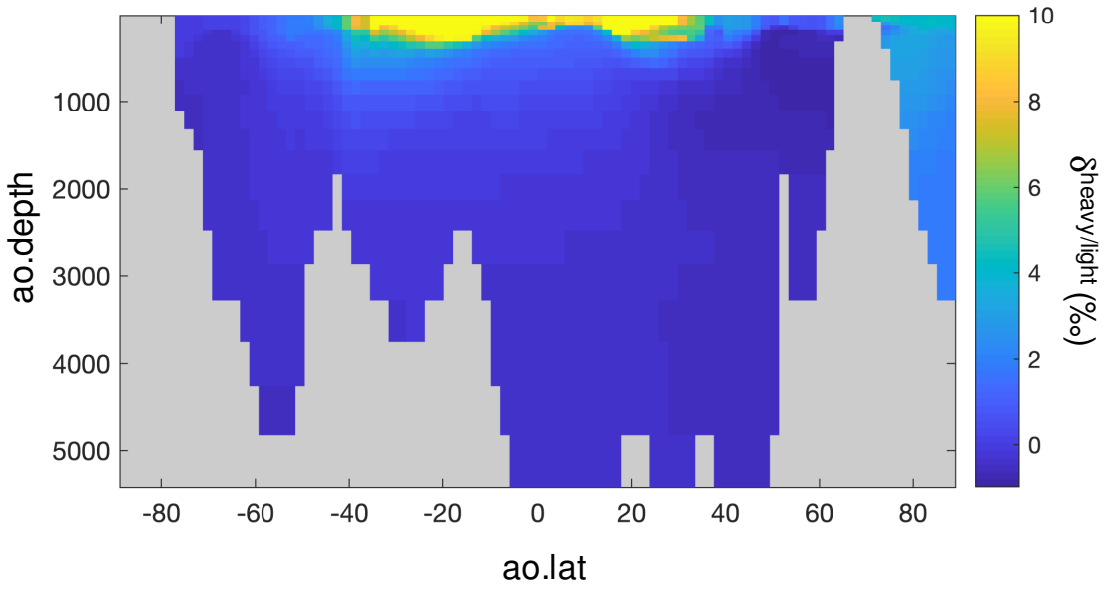


Figure 5. A section through the Pacific Ocean at 179° for a tracer with a -5 ‰ isotope effect for biological uptake. This is similar to a nitrate isotope model, showing enrichment in the heavy isotopes in the upper ocean and oligotrophic gyres, except that it does not include isotopic fractionation with other important processes such as nitrogen fixation and denitrification.

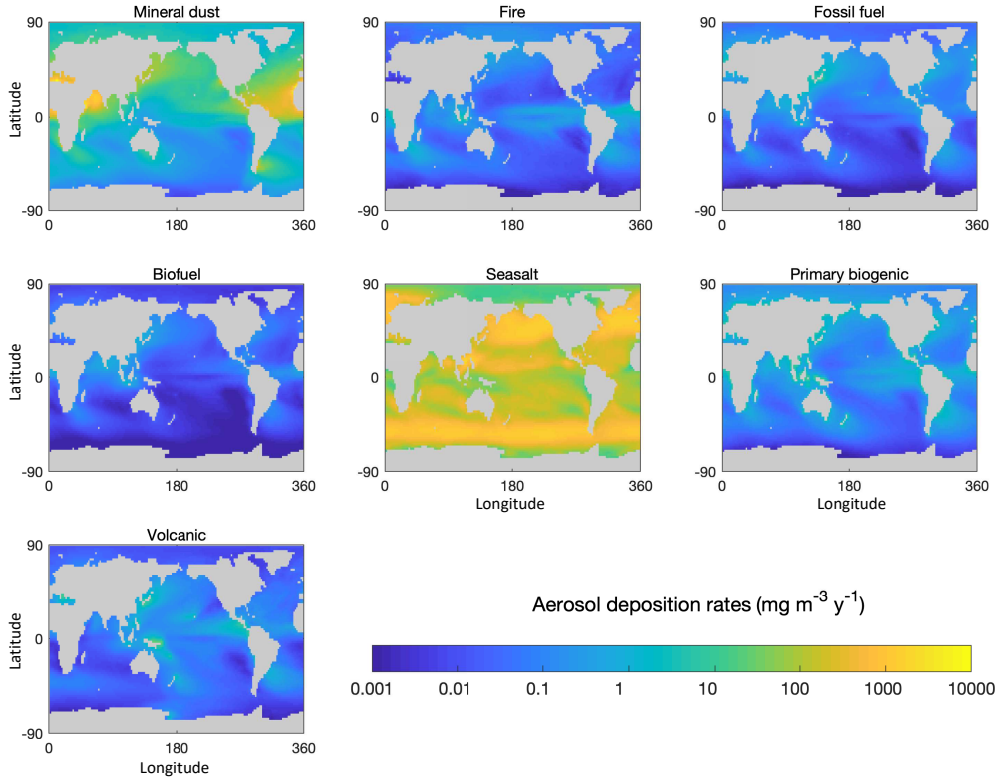


Figure 6. The deposition of various different types of aerosols to the surface ocean. For the mineral dust, seasalt, and primary biogenic sources of aerosols, the mass of the aerosols is the total mass. For the fire, fossil fuel and biofuel aerosols, this represents the black carbon contribution from these sources, and for volcanic aerosols the mass is the sulfur contribution.

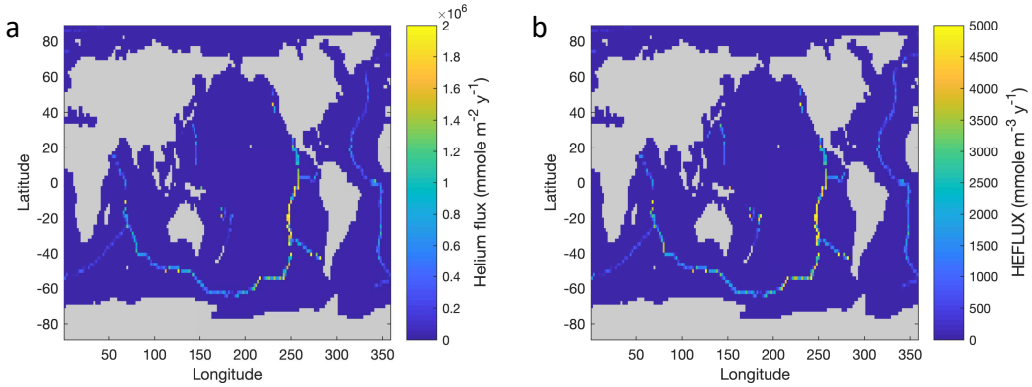


Figure 7. Hydrothermal vent helium flux (a) is determined from rates of seafloor spreading, modified regionally based on observations of ocean dissolved helium concentrations (Holzer et al., 2017). The input of a tracer with ^{3}He will be proportional to the ^{3}He flux, divided by the volume of the grid cell into which the hydrothermal venting occurs, to account for the fact that a similar input will be more dilute when injected into larger grid cells (b).

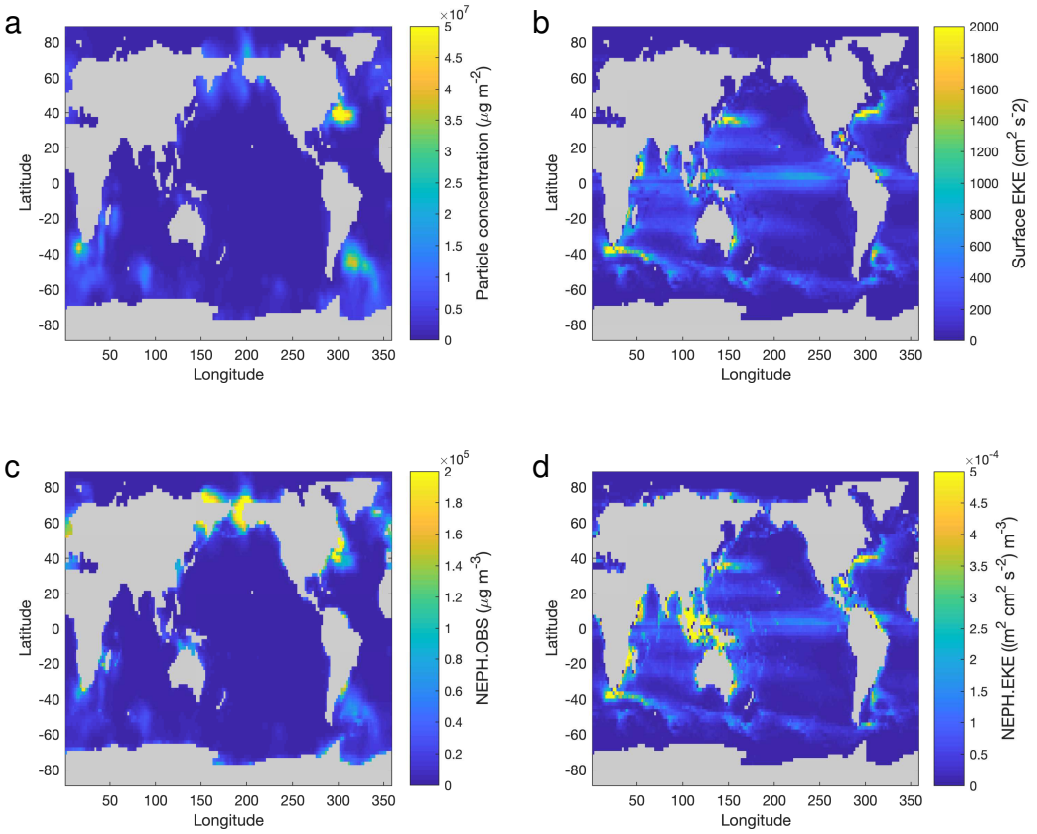


Figure 8. The intensity of nepheloid layers can be modeled in two ways. Input of tracers can be scaled to a global compilation of particle maximum concentration observations interpolated onto the AO grid (a). Alternatively, input can be scaled to satellite observations of surface eddy kinetic energy (EKE; b). These datasets are converted to input fluxes by dividing the total particle quantity or EKE energy by the volume of the bottom box where input occurs (c, d).

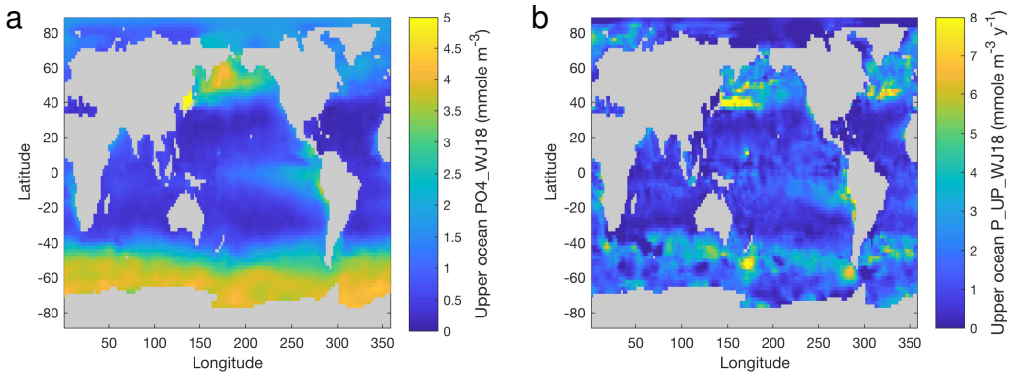
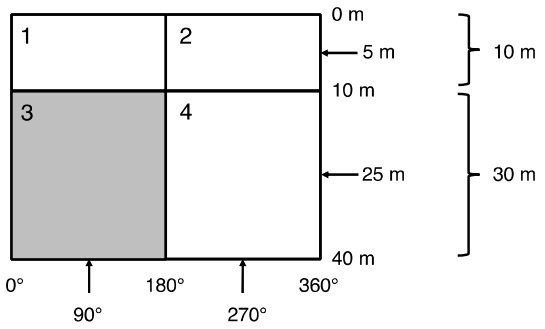


Figure 9. The model concentration of phosphate in the surface ocean, from a model which seeks to match global phosphate distributions with an OCIM circulation (a; PO4_WJ18). This model also calculates the phosphate uptake rate for each grid cell in the surface ocean, with highest uptake rates generally occurring at temperate latitudes (roughly 35° to 65°), where nutrient concentrations are high and there is enough light to sustain significant biological production (b; P_UP_WJ18).




```
ao.nocn    =  3
ao.iocn    = [1  2  4]
ao.OCN     = [1  1
              0  1]
ao.nanOCN  = [1  1
              NaN  1]
ao.EQNPOS  = [1  2
              0  3]
ao.nsurf   =  2
ao.isurf   = [1  2]
ao.SURF    = [1  1
              0  0]
ao.nanSURF = [1  1
              NaN  NaN]

ao.lon     = [90  270]
ao.Lon     = [90  90  270]
ao.LON     = [90  270
              90  270]
ao.height  = [10  30]
ao.Height  = [10  10  30]
ao.HEIGHT  = [10  10
              30  30]
ao.depth   = [5  25]
ao.Depth   = [5  5  25]
ao.DEPTH   = [5  5
              25  25]
```

Figure 10. A simplified 4-box model illustrates various parameters stored in the Matlab structure ‘ao’, each of which provides different information about the AO grid.

REFERENCES

Bardin A., Primeau F. and Lindsay K. (2014) An offline implicit solver for simulating prebomb radiocarbon. *Ocean Model.* **73**, 45–58. Available at: http://www.sciencedirect.com/science/article/pii/S1463500313001820.

Brahney J., Mahowald N., Ward D. S., Ballantyne A. P. and Neff J. C. (2015) Is atmospheric phosphorus pollution altering global alpine Lake stoichiometry? *Global Biogeochem. Cycles* **29**, 1369–1383. Available at: https://doi.org/10.1002/2015GB005137.

Broecker W. S. and Peng T.-H. (1982) *Tracers in the Sea*., Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York.

Chamberlain M. A., Matear R. J., Holzer M., Bi D. and Marsland S. J. (2019) Transport matrices from standard ocean-model output and quantifying circulation response to climate change. *Ocean Model.* **135**, 1–13. Available at: http://www.sciencedirect.com/science/article/pii/S1463500318300891.

Chien C.-T., Mackey K. R. M., Dutkiewicz S., Mahowald N. M., Prospero J. M. and Paytan A. (2016) Effects of African dust deposition on phytoplankton in the western tropical Atlantic Ocean off Barbados. *Global Biogeochem. Cycles* **30**, 716–734. Available at: https://doi.org/10.1002/2015GB005334.

Deutsch C., Sigman D. M., Thunell R. C., Meckler A. N. and Haug G. H. (2004) Isotopic constraints on glacial/interglacial changes in the oceanic nitrogen budget. *Global Biogeochem. Cycles* **18**. Available at: https://doi.org/10.1029/2003GB002189.

DeVries T. (2014) The oceanic anthropogenic CO2 sink: Storage, air-sea fluxes, and transports over the industrial era. *Global Biogeochem. Cycles* **28**, 631–647. Available at: https://doi.org/10.1002/2013GB004739.

DeVries T. and Deutsch C. (2014) Large-scale variations in the stoichiometry of marine organic matter respiration. *Nat. Geosci.* **7**, 890. Available at: https://doi.org/10.1038/ngeo2300.

DeVries T., Deutsch C., Primeau F., Chang B. and Devol A. (2012) Global rates of water-column denitrification derived from nitrogen gas measurements. *Nat. Geosci.* **5**, 547. Available at: https://doi.org/10.1038/ngeo1515.

DeVries T., Deutsch C., Rafter P. A. and Primeau F. (2013) Marine denitrification rates determined from a global 3-D inverse model. *Biogeosciences* **10**, 2481–2496. Available at: https://www.biogeosciences.net/10/2481/2013/.

DeVries T., Holzer M. and Primeau F. (2017) Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature* **542**, 215. Available at: https://doi.org/10.1038/nature21068.

DeVries T., Liang J.-H. and Deutsch C. (2014) A mechanistic particle flux model applied to the oceanic phosphorus cycle. *Biogeosciences* **11**, 5381–5398. Available at: https://www.biogeosciences.net/11/5381/2014/.

DeVries T. and Primeau F. (2011) Dynamically and Observationally Constrained Estimates of Water-Mass Distributions and Ages in the Global Ocean. *J. Phys. Oceanogr.* **41**, 2381–2401. Available at: https://doi.org/10.1175/JPO-D-10-05011.1.

DeVries T. and Weber T. (2017) The export and fate of organic matter in the ocean: New constraints from combining satellite and oceanographic tracer observations. *Global Biogeochem. Cycles* **31**, 535–555. Available at: https://doi.org/10.1002/2016GB005551.

Elderfield H. and Rickaby R. E. M. (2000) Oceanic Cd/P ratio and nutrient utilization in the glacial Southern Ocean . *Nature* **405**, 305–310. Available at: https://doi.org/10.1038/35012507.

Farley K. A., Maier-Reimer E., Schlosser P. and Broecker W. S. (1995) Constraints on mantle 3He fluxes and deep-sea circulation from an oceanic general circulation model. *J. Geophys. Res. Solid Earth* **100**, 3829–3839. Available at: https://doi.org/10.1029/94JB02913.

Gardner W. D., Richardson M. J. and Mishonov A. V (2018a) Global assessment of benthic nepheloid layers and linkage with upper ocean dynamics. *Earth Planet. Sci. Lett.* **482**, 126–134. Available at: http://www.sciencedirect.com/science/article/pii/S0012821X17306441.

Gardner W. D., Richardson M. J., Mishonov A. V and Biscaye P. E. (2018b) Global comparison of benthic nepheloid layers based on 52 years of nephelometer and transmissometer measurements. *Prog. Oceanogr.* **168**, 100–111. Available at: http://www.sciencedirect.com/science/article/pii/S0079661118301290.

Gardner W. D., Tucholke B. E., Richardson M. J. and Biscaye P. E. (2017) Benthic storms, nepheloid layers, and linkage with upper ocean dynamics in the western North Atlantic. *Mar. Geol.* **385**, 304–327. Available at: http://www.sciencedirect.com/science/article/pii/S0025322716303966.

Van Geen A., Boyle E. A. and Moore W. S. (1991) Trace-metal enrichments in waters of the Gulf of Cadiz, Spain. *Geochim. Cosmochim. Acta* **55**, 2173–2191.

Holzer M., DeVries T., Bianchi D., Newton R., Schlosser P. and Winckler G. (2017) Objective estimates of mantle 3He in the ocean and implications for constraining the deep ocean circulation. *Earth Planet. Sci. Lett.* **458**, 305–314. Available at: http://www.sciencedirect.com/science/article/pii/S0012821X16306197.

Holzer M., Primeau F. W., DeVries T. and Matear R. (2014) The Southern Ocean silicon trap: Data-constrained estimates of regenerated silicic acid, trapping efficiencies, and global transport paths. *J. Geophys. Res. Ocean.* **119**, 313–331. Available at: https://doi.org/10.1002/2013JC009356.

Huested S S, Jones S P, Boyle E. A., Huested Sarah S and Jones Susan P (1981) On the distribution of copper, nickel, and cadmium in the surface waters of the North Atlantic and North Pacific Ocean. *J. Geophys. Res.* **86**, 8048–8066.

Hunter K. A. and Boyd P. W. (2007) Iron-binding ligands and their role in the ocean biogeochemistry of iron. *Environ. Chem.* **4**, 221–232. Available at: https://app.dimensions.ai/details/publication/pub.1005830458.

John S. G., Helgoe J., Townsend E., Weber T., DeVries T., Tagliabue A., Moore K., Lam P., Marsay C. M. and Till C. (2018) Biogeochemical cycling of Fe and Fe stable isotopes in the Eastern Tropical South Pacific. *Mar. Chem.* **201**.

Khatiwala S., Primeau F. and Hall T. (2009) Reconstruction of the history of anthropogenic CO2 concentrations in the ocean. *Nature* **462**, 346. Available at: https://doi.org/10.1038/nature08526.

Khatiwala S., Visbeck M. and Cane M. A. (2005) Accelerated simulation of passive tracers in ocean circulation models. *Ocean Model.* **9**, 51–69.

Kvale K. F., Khatiwala S., Dietze H., Kriest I. and Oschlies A. (2017) Evaluation of the transport matrix method for simulation of ocean biogeochemical tracers. *Geosci. Model Dev.* **10**, 2425–2445. Available at: https://www.geosci-model-dev.net/10/2425/2017/.

Kwon E. Y. and Primeau F. (2006) Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data. *Global Biogeochem. Cycles* **20**. Available at: https://doi.org/10.1029/2005GB002631.

Lauvset S. K., Key R. M., Olsen A., van Heuven S., Velo A., Lin X., Schirnick C., Kozyr A., Tanhua T., Hoppema M., Jutterström S., Steinfeldt R., Jeansson E., Ishii M., Perez F. F., Suzuki T. and Watelet S. (2016) A new global interior ocean mapped climatology: the 1{\degree}\,$\times$\,1{\degree} GLODAP version 2. *Earth Syst. Sci. Data* **8**, 325–340. Available at: https://www.earth-syst-sci-data.net/8/325/2016/.

Levitus S., Locarnini R. A., Boyer T. P., Mishonov A. V, Antonov J. I., Garcia H. E., Baranova O. K., Zweng M. M., Johnson D. R. and Seidov 1948- D. (2010) World ocean atlas 2009 eds. O. C. L. National Oceanographic Data Center (U.S.) and N. E. S. United States Data, and Information Service,. Available at: https://repository.library.noaa.gov/view/noaa/1259.

Olsen A., Key R. M., van Heuven S., Lauvset S. K., Velo A., Lin X., Schirnick C., Kozyr A., Tanhua T., Hoppema M., Jutterström S., Steinfeldt R., Jeansson E., Ishii M., Pérez F. F. and Suzuki T. (2016) The Global Ocean Data Analysis Project version 2 (GLODAPv2) -- an internally consistent data product for the world ocean. *Earth Syst. Sci. Data* **8**, 297–323. Available at: https://www.earth-syst-sci-data.net/8/297/2016/.

Parekh P., Follows M. J. and Boyle E. (2004) Modeling the global ocean iron cycle. *Global Biogeochem. Cycles* **18**. Available at: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GB002061.

Primeau F. (2005) Characterizing Transport between the Surface Mixed Layer and the Ocean Interior with a Forward and Adjoint Global Ocean Transport Model. *J. Phys. Oceanogr.* **35**, 545–564. Available at: https://doi.org/10.1175/JPO2699.1.

Primeau F. (2003) Characterizing transport timescales between the surface mixed layer and the deep ocean with an OGCM and its adjoint. *Bull. Am. Meteorol. Soc.* **84**, 897–899.

Primeau F. W., Holzer M. and DeVries T. (2013) Southern Ocean nutrient trapping and the efficiency of the biological pump. *J. Geophys. Res. Ocean.* **118**, 2547–2564. Available at: https://doi.org/10.1002/jgrc.20181.

Roshan S. and DeVries T. (2017) Efficient dissolved organic carbon production and export in the oligotrophic ocean. *Nat. Commun.* **8**, 2036. Available at: https://doi.org/10.1038/s41467-017-02227-3.

Roshan S., DeVries T., Wu J. and Chen G. (2018) The Internal Cycling of Zinc in the Ocean. *Global Biogeochem. Cycles* **32**, 1833–1849. Available at: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GB006045.

Sarmiento J. L. and Gruber N. (2006) *Ocean Biogeochemical Dynamics*. STU-Stud., Princeton University Press. Available at: http://www.jstor.org/stable/j.ctt3fgxqx.

Schlitzer R. (2007) Assimilation of Radiocarbon and Chlorofluorocarbon Data to Constrain Deep and Bottom Water Transports in the World Ocean. *J. Phys. Oceanogr.* **37**, 259–276. Available at: https://doi.org/10.1175/JPO3011.1.

Schlitzer R. (1993) *Determining the Mean, Large-Scale Circulation of the Atlantic with the Adjoint Method.*,

Schlitzer R., Anderson R. F., Dodas E. M., Lohan M., Geibert W., Tagliabue A., Bowie A., Jeandel C., Maldonado M. T., Landing W. M., Cockwell D., Abadie C., Abouchami W., Achterberg E. P., Agather A., Aguliar-Islas A., van Aken H. M., Andersen M., Archer C., Auro M., de Baar H. J., Baars O., Baker A. R., Bakker K., Basak C., Baskaran M., Bates N. R., Bauch D., van Beek P., Behrens M. K., Black E., Bluhm K., Bopp L., Bouman H., Bowman K., Bown J., Boyd P., Boye M., Boyle E. A., Branellec P., Bridgestock L., Brissebrat G., Browning T., Bruland K. W., Brumsack H.-J., Brzezinski M., Buck C. S., Buck K. N., Buesseler K., Bull A., Butler E., Cai P., Mor P. C., Cardinal D., Carlson C., Carrasco G., Casacuberta N., Casciotti K. L., Castrillejo M., Chamizo E., Chance R., Charette M. A., Chaves J. E., Cheng H., Chever F., Christl M., Church T. M., Closset I., Colman A., Conway T. M., Cossa D., Croot P., Cullen J. T., Cutter G. A., Daniels C., Dehairs F., Deng F., Dieu H. T., Duggan B., Dulaquais G., Dumousseaud C., Echegoyen-Sanz Y., Edwards R. L., Ellwood M., Fahrbach E., Fitzsimmons J. N., Russell Flegal A., Fleisher M. Q., van de Flierdt T., Frank M., Friedrich J., Fripiat F., Fröllje H., Galer S. J. G., Gamo T., Ganeshram R. S., Garcia-Orellana J., Garcia-Solsona E., Gault-Ringold M., George E., Gerringa L. J. A., Gilbert M., Godoy J. M., Goldstein S. L., Gonzalez S. R., Grissom K., Hammerschmidt C., Hartman A., Hassler C. S., Hathorne E. C., Hatta M., Hawco N., Hayes C. T., Heimbürger L.-E., Helgoe J., Heller M., Henderson G. M., Henderson P. B., van Heuven S., Ho P., Horner T. J., Hsieh Y.-T., Huang K.-F., Humphreys M. P., Isshiki K., Jacquot J. E., Janssen D. J., Jenkins W. J., John S., Jones E. M., Jones J. L., Kadko D. C., Kayser R., Kenna T. C., Khondoker R., Kim T., Kipp L., Klar J. K., Klunder M., Kretschmer S., Kumamoto Y., Laan P., Labatut M., Lacan F., Lam P. J., Lambelet M., Lamborg C. H., Le Moigne F. A. C., Le Roy E., Lechtenfeld O. J., Lee J.-M., Lherminier P., Little S., López-Lora M., Lu Y., Masque P., Mawji E., Mcclain C. R., Measures C., Mehic S., Barraqueta J.-L. M., van der Merwe P., Middag R., Mieruch S., Milne A., Minami T., Moffett J. W., Moncoiffe G., Moore W. S., Morris P. J., Morton P. L., Nakaguchi Y., Nakayama N., Niedermiller J., Nishioka J., Nishiuchi A., Noble A., Obata H., Ober S., Ohnemus D. C., van Ooijen J., O'Sullivan J., Owens S., Pahnke K., Paul M., Pavia F., Pena L. D., Peters B., Planchon F., Planquette H., Pradoux C., Puigcorbé V., Quay P., Queroue F., Radic A., Rauschenberg S., Rehkämper M., Rember R., Remenyi T., Resing J. A., Rickli J., Rigaud S., Rijkenberg M. J. A., Rintoul S., Robinson L. F., Roca-Martí M., Rodellas V., Roeske T., Rolison J. M., Rosenberg M., Roshan S., Rutgers van der Loeff M. M., Ryabenko E., Saito M. A., Salt L. A., Sanial V., Sarthou G., Schallenberg C., Schauer

U., Scher H., Schlosser C., Schnetger B., Scott P., Sedwick P. N., Semiletov I., Shelley R., Sherrell R. M., Shiller A. M., Sigman D. M., Singh S. K., Slagter H. A., Slater E., Smethie W. M., Snaith H., Sohrin Y., Sohst B., Sonke J. E., Speich S., Steinfeldt R., Stewart G., Stichel T., Stirling C. H., Stutsman J., Swarr G. J., Swift J. H., Thomas A., Thorne K., Till C. P., Till R., Townsend A. T., Townsend E., Tuerena R., Twining B. S., Vance D., Velazquez S., Venchiarutti C., Villa-Alfageme M., Vivancos S. M., Voelker A. H. L., Wake B., Warner M. J., Watson R., van Weerlee E., Alexandra Weigand M., Weinstein Y., Weiss D., Wisotzki A., Woodward E. M. S., Wu J., Wu Y., Wuttig K., Wyatt N., Xiang Y., Xie R. C., Xue Z., Yoshikawa H., Zhang J., Zhang P., Zhao Y., Zheng L., Zheng X.-Y., Zieringer M., Zimmer L. A., Ziveri P., Zunino P. and Zurbrick C. (2018) The GEOTRACES Intermediate Data Product 2017. *Chem. Geol.* **493**.

Sigman D. M. and Boyle E. A. (2000) Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* **407**, 859.

Sunda W. (2012) Feedback Interactions between Trace Metal Nutrients and Phytoplankton in the Ocean. *Front. Microbiol.* **3**, 204. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2012.00204.

Taburet G., Sanchez-Roman A., Ballarotta M., Pujol M.-I., Legeais J.-F., Fournier F., Faugere Y. and Dibarboure G. (2019) DUACS DT-2018: 25 years of reprocessed sea level altimeter products. *Ocean Sci. Discuss.* **2019**, 1–30. Available at: https://www.ocean-sci-discuss.net/os-2018-150/.

Tagliabue A., Aumont O., DeAth R., Dunne J. P., Dutkiewicz S., Galbraith E., Misumi K., Moore J. K., Ridgwell A., Sherman E., Stock C., Vichi M., Völker C. and Yool A. (2016) How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem. Cycles* **30**, 149–174. Available at: https://doi.org/10.1002/2015GB005289.

Takano S., Tanimizu M., Hirata T. and Sohrin Y. (2014) Isotopic constraints on biogeochemical cycling of copper in the ocean. *Nat. Commun.* **5**.

Teng Y.-C., Primeau F. W., Moore J. K., Lomas M. W. and Martiny A. C. (2014) Global-scale variations of the ratios of carbon to phosphorus in exported marine organic matter. *Nat. Geosci.* **7**, 895. Available at: https://doi.org/10.1038/ngeo2303.

Toggweiler J. R. (1999) Variation of atmospheric CO2 by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588. Available at: https://doi.org/10.1029/1999PA900033.

Toggweiler J. R., Gnanadesikan A., Carson S., Murnane R. and Sarmiento J. L. (2003a) Representation of the carbon cycle in box models and GCMs: 1. Solubility pump. *Global Biogeochem. Cycles* **17**. Available at: https://doi.org/10.1029/2001GB001401.

Toggweiler J. R., Murnane R., Carson S., Gnanadesikan A. and Sarmiento J. L. (2003b) Representation of the carbon cycle in box models and GCMs, 2, Organic pump. *Global Biogeochem. Cycles* **17**. Available at: https://doi.org/10.1029/2001GB001841.

Tyrrell T. (1999) The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525. Available at: https://doi.org/10.1038/22941.

Wang W.-L., Moore J. K., Martiny A. C. and Primeau F. W. (2019) Convergent estimates of marine nitrogen fixation. *Nature* **566**, 205–211. Available at: https://doi.org/10.1038/s41586-019-0911-2.

Weber T., Cram J. A., Leung S. W., DeVries T. and Deutsch C. (2016) Deep ocean nutrients imply large latitudinal variation in particle transfer efficiency. *Proc. Natl. Acad. Sci.* **113**, 8606 LP – 8611. Available at: http://www.pnas.org/content/113/31/8606.abstract.

Weber T. and Deutsch C. (2014) Local versus basin-scale limitation of marine nitrogen fixation. *Proc. Natl. Acad. Sci.* **111**, 8741–8746. Available at: http://www.pnas.org/content/111/24/8741.

Weber T. and Deutsch C. (2012) Oceanic nitrogen reservoir regulated by plankton diversity and ocean circulation. *Nature* **489**, 419. Available at: https://doi.org/10.1038/nature11357.

Weber T., John S., Tagliabue A. and DeVries T. (2018) Biological uptake and reversible scavenging of zinc in the global ocean. *Science (80-.).* **361**.

Weber T. S. and Deutsch C. (2010) Ocean nutrient ratios governed by plankton biogeography. *Nature* **467**, 550. Available at: http://dx.doi.org/10.1038/nature09403.

Wohlforth C. (2004) *The Whale and the Supercomputer: on the northern front of climate change*., North

Point Press, New York.
Yang S.-C., Lee D.-C., Ho T.-Y., Wen L.-S. and Yang H.-H. (2014) The isotopic composition of dissolved cadmium in the water column of the West Philippine Sea. *Front. Mar. Sci.* **1**, 61. Available at: https://www.frontiersin.org/article/10.3389/fmars.2014.00061.
Zanna L., Khatiwala S., Gregory J. M., Ison J. and Heimbach P. (2019) Global reconstruction of historical ocean heat storage and transport. *Proc. Natl. Acad. Sci.* **116**, 1126 LP – 1131. Available at: http://www.pnas.org/content/116/4/1126.abstract.