

Microbial metabolite fluxes in a model marine anoxic ecosystem

Stilianos Louca^{1,2}  | Yrene M. Astor^{3,4} | Michael Doebeli^{5,6,7} | Gordon T. Taylor⁸ | Mary I. Scranton⁸

¹Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA

²Department of Biology, University of Oregon, Eugene, OR, USA

³Estación de Investigaciones Marinas de Margarita, Fundación La Salle de Ciencias Naturales, Punta de Piedras, Venezuela

⁴Institute for Marine Remote Sensing, University of South Florida, Tampa, FL, USA

⁵Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

⁶Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

⁷Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada

⁸School of Marine and Atmospheric Sciences, Stony Brook University, New York, NY, USA

Correspondence

Stilianos Louca, Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA.
Email: louca.research@gmail.com

Funding information

Natural Sciences and Engineering Research Council of Canada; Biodiversity Research Centre, University of British Columbia; Venezuela Fondo Nacional de Ciencia y Tecnología; Venezuela Consejo Nacional de Investigaciones Científicas y Tecnológicas; US National Science Foundation, Grant/Award Number: OCE-1259110

Abstract

Permanently anoxic regions in the ocean are widespread and exhibit unique microbial metabolic activity exerting substantial influence on global elemental cycles and climate. Reconstructing microbial metabolic activity rates in these regions has been challenging, due to the technical difficulty of direct rate measurements. In Cariaco Basin, which is the largest permanently anoxic marine basin and an important model system for geobiology, long-term monitoring has yielded time series for the concentrations of biologically important compounds; however, the underlying metabolite fluxes remain poorly quantified. Here, we present a computational approach for reconstructing vertical fluxes and in situ net production/consumption rates from chemical concentration data, based on a 1-dimensional time-dependent diffusive transport model that includes adaptive penalization of overfitting. We use this approach to estimate spatiotemporally resolved fluxes of oxygen, nitrate, hydrogen sulfide, ammonium, methane, and phosphate within the sub-euphotic Cariaco Basin water column (depths 150–900 m, years 2001–2014) and to identify hotspots of microbial chemolithotrophic activity. Predictions of the fitted models are in excellent agreement with the data and substantially expand our knowledge of the geobiology in Cariaco Basin. In particular, we find that the diffusivity, and consequently fluxes of major reductants such as hydrogen sulfide, and methane, is about two orders of magnitude greater than previously estimated, thus resolving a long-standing apparent conundrum between electron donor fluxes and measured dark carbon assimilation rates.

KEYWORDS

Cariaco Basin, chemolithoautotroph, geobiology, inverse transport modeling, redoxcline

1 | INTRODUCTION

Permanently or temporarily anoxic regions in the ocean are a topic of increasing interest due to their unique microbial ecology (Ulloa,

Wright, Belmar, & Hallam, 2013; Wright, Konwar, & Hallam, 2012), their importance to global elemental cycles and marine productivity (Ulloa, Canfield, DeLong, Letelier, & Stewart, 2012), and the intensifying deoxygenation of the ocean (Breitburg et al., 2018; Schmidtko, Stramma, & Visbeck, 2017). Micro-organisms in these regions are adapted to operate under oxygen-limited or oxygen-depleted

Abbreviations: DCA, dark carbon assimilation; ILTM, inverse linear transport modeling.

conditions, making use of alternative terminal electron acceptors for respiration and often utilizing inorganic substrates for energy. In many anoxic marine zones, reductants such as hydrogen sulfide, ammonium, and the potent greenhouse gas methane, diffusing upwards from underlying layers or the sediments, react biologically with oxidants such as oxygen and nitrate produced in the overlying layers, thus fueling chemolithoautotrophic activity and affecting marine nitrogen, sulfur, oxygen, and carbon budgets (Taylor et al., 2018; Ulloa et al., 2012). In the Cariaco Basin, a permanently anoxic marine region off the coast of Venezuela, multi-decadal monitoring has generated rich time series of the distribution of metabolically important compounds over space and time (Muller-Karger et al., 2019; Scranton et al., 2014). These data revealed the existence of a strong dynamic redox gradient over depth, along which upward diffusing reductants such as hydrogen sulfide are directly or indirectly oxidized by oxidants such as oxygen in a transition zone roughly spanning depths 200–400 m, sometimes referred to as “redoxcline” (Ho et al., 2004; Li, Taylor, Astor, Varela, & Scranton, 2012b; Taylor et al., 2018). Concurrent molecular surveys revealed unique microbial communities that exhibit a clear spatial organization across depth, and elevated population densities within the redoxcline (Cernadas-Martín, Suter, Scranton, Astor, & Taylor, 2017; Rodriguez-Mora, Scranton, Taylor, & Chistoserdov, 2015; Suter, Pachiadaki, Taylor, Astor, & Edgcomb, 2018; Taylor et al., 2006, 2018). However, chemical fluxes across space and microbial metabolic rates in Cariaco Basin and other anoxic regions remain poorly quantified and are largely temporally unresolved, thus making a mechanistic connection between chemical transitions and microbial ecological dynamics difficult (Taylor et al., 2018). A major reason for this gap in our knowledge is that, compared to chemical concentration measurements, explicit metabolic rate measurements are technically challenging, especially when performed in situ.

Mathematical modeling is sometimes used to indirectly estimate the flux rates that underlie the observed chemical concentration profiles (Berg, Risgaard-Petersen, & Rysgaard, 1998; Cernadas-Martín et al., 2017; Li et al., 2012b; Samodurov et al., 2013; Scranton, Sayles, Bacon, & Brewer, 1987; Taylor et al., 2001, 2018). For example, Scranton et al. (1987) used a time-dependent diffusion box model to estimate sulfide fluxes from the sediments into the Cariaco Basin water column. These calculations, however, were based solely on measurements at two time points (in 1973 and 1982), ignored possible in situ sulfide production (Li et al., 2012b), and assumed that sulfide was consumed entirely at some fixed depth, thus ignoring shifts in the redoxcline depth over time (Scranton et al., 2014). On the other hand, Li et al. (2012b) and Cernadas-Martín et al. (2017) used a 1-dimensional diffusion model for Cariaco Basin to estimate fluxes of various compounds produced or consumed during microbial metabolism (henceforth “metabolites” for simplicity). Their models only estimated fluxes into and out of a narrow depth interval (~100–200 m wide), thus missing possible metabolic activity at other depths, and assumed that metabolite depth profiles were at steady state, thus ignoring

possible temporal lags in the response of redox gradients to flux changes (Scranton et al., 1987, 2014) and microbial dynamics (Taylor et al., 2018).

Here, we develop a computational approach for estimating metabolite fluxes and net production/consumption rates across space and time, using chemical concentration data measured at arbitrary spacetime points. Our approach is based on a 1-dimensional time and depth-dependent diffusive transport model that accounts for temporal changes in boundary conditions, diffusive transport coefficients and in situ production rates, as well as for potential geometric dilution effects due to variation of a system's lateral (cross-sectional) area with depth. We use our approach to reconstruct spatiotemporally resolved metabolite fluxes across the sub-euphotic Cariaco Basin water column (depths 150–900 m) during the years 2001–2014. We consider several important metabolites, including oxygen (O_2), nitrate (NO_3^-), hydrogen sulfide (H_2S), ammonium (NH_4^+), methane (CH_4), and phosphate (PO_4^{3-}). Our estimates yield detailed insight into the microbial activity that underlies the geochemical structure of the Cariaco Basin water column.

2 | METHODS

2.1 | Cariaco data

Chemical and physical data from Cariaco station A (coordinates 10.5°N, 64.66°W) were downloaded on April 28, 2018, from the Cariaco Basin time series project website (<http://www.imars.usf.edu/cariaco>) for station CARIACO. Additional sources of CARIACO chemical data are the NOAA's National Centers for Environmental Information (NCEI), the Ocean Carbon Data System, the US Biological and Chemical Oceanography Data Management Office (BCO-DMO), and the NASA SeaBASS database. Data collection methods have been described previously (Li, Taylor, Astor, & Scranton, 2008; Muller-Karger et al., 2019; Scranton et al., 2014; Thunell et al., 2000). Additional hydrogen sulfide concentration data, recently published by Muller-Karger et al. (2019), were obtained directly from the authors. The lateral area of the eastern basin (within which station CARIACO is located) at various discrete depths was taken from Samodurov et al. (2013, Table 1 therein) and was linearly interpolated between those depths (Figure S1).

Inverse linear transport modeling (ILTM) was used to estimate net in situ production rates (R in Equation 6) of oxygen, nitrate, hydrogen sulfide, ammonium, methane, and phosphate (dissolved pools). These compounds (referred to here as “metabolites”) were chosen due to their biological importance within the considered depth interval, their relatively good sampling resolution, and spatiotemporal coverage, and the fact that their transport across depth can be largely described by eddy diffusion. Nitrite was not included in ILTM because nitrite rarely accumulates to significant levels (Figure S10); hence, estimated net production rates would be almost zero and dominated by errors despite potentially intense cryptic nitrite fluxes (e.g., as an intermediate of nitrification or denitrification).

Variable	O ₂	NO ₃ ⁻	H ₂ S	NH ₄ ⁺	CH ₄	PO ₄ ³⁻
Total content (depth-integrated)	3.5	0.81	11	4.7	2.1	1.3
Gross production rate (depth-integrated)	2.4	2.7	1.1	<0.1	<0.1	0.17
Gross consumption rate (depth-integrated)	13	2.7	11	4.7	3.3	0.3
Net influx rate at top (150 m)	7.1	-0.45	<0.1	0.21	<0.1	-0.077
Gross influx rate at top (150 m)	7.3	<0.1	0.12	0.21	<0.1	<0.01
Gross outflux rate at top (150 m)	<0.1	0.49	<0.1	<0.1	<0.1	0.08
Net influx rate at bottom (900 m)	<0.1	<0.1	13	5.3	3.6	0.31
Gross influx rate at bottom (900 m)	0.13	<0.1	13	5.3	3.6	0.31
Gross outflux rate at bottom (900 m)	0.17	<0.1	<0.1	<0.1	<0.1	<0.01
Net influx rate at top + bottom	7	-0.43	13	5.5	3.7	0.23
Total gross influx + production rate	9.8	2.8	14	5.5	3.8	0.48
Total gross outflux + consumption rate	14	3.2	11	4.7	3.3	0.38
Mean residence time (years)	0.67	0.69	2.6	2.7	1.7	9.1

TABLE 1 Estimated mean metabolite fluxes in Cariaco Basin

Note: Estimated in situ metabolite production and consumption rates, as well as influx and outflux rates across the top (150 m) and bottom boundary (900 m). The total content is depth-integrated, averaged over the considered time interval (years 2001–2014) and measured in mol/m². All rates are depth-integrated where applicable, averaged over the considered time interval, and measured in mmol m⁻² day⁻¹. Depth-integrated or area-specific quantities take into account the variation of the lateral (cross-sectional) basin area with depth and are normalized to the basin area at depth 150 m to facilitate comparisons. Mean residence times were estimated based on the depth-integrated concentrations and gross input/output rates, using a non-steady-state box model. See Methods for details. For analogous summaries constrained to after 2009, see Table S1. For analogous summaries constrained to the redoxcline (depths 200–400 m), see Table S2.

2.2 | Estimating diffusivity in Cariaco Basin

The eddy diffusion coefficient was estimated using the salinity and temperature profiles (see Section 3 for mathematical details). Salinity and temperature-depth profiles were LOESS-smoothed at degree 1 and a span of 10% to reduce noise. Salinity, temperature, and pressure profiles were used to calculate the buoyancy frequency (N) at each depth and time point, using the R package *oce* (Kelley, 2014). To reduce noise in the buoyancy frequency stemming from the numerical differentiation of noisy data, the buoyancy frequency was smoothed using a Savitzky–Golay filter of degree 2 along the time axis (Savitzky & Golay, 1964). For any given choice of the parameters α and p (Equation 1), we simulated the salinity profile over depth and time by solving the differential Equation (2), using the *pdepe* function in MATLAB[®] (MATLAB, 2010). The initial profile was set to the measured salinity profile at the first simulation time point (January 1, 2001). Boundary conditions at the top (150 m) and bottom boundary (900 m) were of Dirichlet type, with the imposed value at each time point and each boundary being the current measured salinity at the boundary's depth. Salinities between data points were bilinearly interpolated wherever needed for the initial condition and boundary conditions.

We did not account for lateral intrusions of denser, oxygenated water from outside, which are known to occur occasionally in Cariaco Basin (Scranton et al., 2014; Taylor et al., 2018). Salinity and temperature profiles during the time period and depth range considered here do not show obvious signs of foreign water intrusions (Figure S2A,B). Similarly, oxygen concentration profiles show only weak signs of potential intrusions of oxygenated water at depth (Figure 2a). It is in principle possible that intrusion events, the subsequent re-equilibration of density structure, and the consumption of introduced oxidants all occur at much shorter time scales than resolved by the monthly time series. However, the good agreement of the fitted diffusivity models with the salinity data ($r^2 = 0.982$, Figure S3) and the metabolite concentration data ($r^2 = 0.878$ – 0.973) further suggests that lateral water intrusions had little effect on the salt and metabolite budgets within the considered spatiotemporal domain.

The predicted salinity profile \hat{S} was compared to the measured salinity data (S) by means of the fraction of explained variance (r^2), calculated as:

$$r^2 = 1 - \frac{1}{V \cdot n} \sum_{i=1}^n (\hat{S}_i - S_i)^2,$$

where i iterates over all available salinity data points ($n = 111,920$), \hat{S}_i is the salinity predicted for the same spacetime points as S_i , and V is the sample variance of the measured salinities S_i . The power law parameters α and p were gradually fitted until r^2 reached a maximum, using the “interior-point” optimization algorithm implemented by the function `fmincon` in MATLAB®. To avoid non-global local optima, we repeated the fitting 200 times, each time with randomly chosen initial values for α and p . The distribution of fitted parameters, as a function of the maximized r^2 , is shown in Figure S13. The parameter values corresponding to the highest r^2 were taken as the final estimate. The same approach was also used to fit the Munk–Anderson diffusivity model (Equation 3), as well as the combined power law + Munk–Anderson model (sum of Equations 1 and 3). For details on the “anchored” diffusivity estimate (Equation 4), performed here solely for sanity checking purposes, see Appendix S1. For the subsequent ILTM analysis, we used the diffusivity obtained from the fitted power law model.

2.3 | Inverse linear transport modeling

An overview of our ILTM approach is provided in Section 3. Mathematical background and computational details on our ILTM approach are provided in Appendix S4. Briefly, the differential Equation (6) was used to calculate a linear mapping (represented as a matrix \mathbb{T} , see Equation 7) between any given net production rates (on a finite grid of spacetime points) and the corresponding predicted volumetric concentration profiles (on the same spacetime points as the concentration measurements). The spatiotemporal grid on which R was estimated (“fitting grid”) was chosen separately for each metabolite to account for differences in sampling resolution and spatiotemporal variability of metabolite concentrations, and such that its size was substantially lower than the number of available data points (overview in Figure S6 and Table S3). In all cases, the number of considered data points was more than 10 times the size of the fitting grid.

Prior to any prediction, the net production rates on the fitting grid were interpolated onto a high-resolution grid (“refined grid”) using an interpolation matrix \mathbb{I} , which maps rates on the fitting grid to rates on the refined grid. Because our numerical differential equation solver only returns solutions on a rectangular spatiotemporal grid (“prediction grid”), an additional interpolation is performed from the prediction grid onto the spacetime points of the data (using a suitable matrix \mathbb{P}). Hence, \mathbb{T} is composed of 3 matrices, $\mathbb{T} = \mathbb{P} \cdot \mathbb{G} \cdot \mathbb{I}$, where \mathbb{G} encodes the “Green’s function” (sometimes called “fundamental solution”) of the differential equation (Duffy, 2001). Each row in the matrix \mathbb{G} corresponds to the solution of the differential Equation (6) evaluated at a specific point on the prediction grid, if the net production rate was zero in all but a single point on the refined grid. The resolutions of the refined grid and the prediction grid were chosen sufficiently high to ensure a high accuracy of the solutions of the differential equation. The estimation of net production rates on the fitting grid can be formulated as a least-squares optimization problem (minimizing the expression in Equation 7), which we solved using linear algebra routines in MATLAB®. For each metabolite, the fraction of variance explained by the predicted concentrations (r^2) was calculated as described above for the salinity model.

using an interpolation matrix \mathbb{I} , which maps rates on the fitting grid to rates on the refined grid. Because our numerical differential equation solver only returns solutions on a rectangular spatiotemporal grid (“prediction grid”), an additional interpolation is performed from the prediction grid onto the spacetime points of the data (using a suitable matrix \mathbb{P}). Hence, \mathbb{T} is composed of 3 matrices, $\mathbb{T} = \mathbb{P} \cdot \mathbb{G} \cdot \mathbb{I}$, where \mathbb{G} encodes the “Green’s function” (sometimes called “fundamental solution”) of the differential equation (Duffy, 2001). Each row in the matrix \mathbb{G} corresponds to the solution of the differential Equation (6) evaluated at a specific point on the prediction grid, if the net production rate was zero in all but a single point on the refined grid. The resolutions of the refined grid and the prediction grid were chosen sufficiently high to ensure a high accuracy of the solutions of the differential equation. The estimation of net production rates on the fitting grid can be formulated as a least-squares optimization problem (minimizing the expression in Equation 7), which we solved using linear algebra routines in MATLAB®. For each metabolite, the fraction of variance explained by the predicted concentrations (r^2) was calculated as described above for the salinity model.

2.4 | Estimation of depth-integrated fluxes and area-specific quantities

In all cases, depth integration of rates and concentrations took into account the variation of the lateral (cross-sectional) basin area with depth (Figure S1), and all depth-integrated quantities (e.g., production rates) and area-specific quantities (e.g., fluxes through the top and bottom boundaries, or fluxes into the redoxcline) are normalized with respect to the basin area at depth 150 m. For example, if $R(t, z)$ is the estimated net production rate of some metabolite, then its depth-integrated value within some depth interval $[z_1, z_2]$, averaged over some time interval $[t_1, t_2]$, was defined as:

$$\bar{R} := \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} dt \int_{z_1}^{z_2} dz R(t, z) \frac{A(z)}{A(z_0)},$$

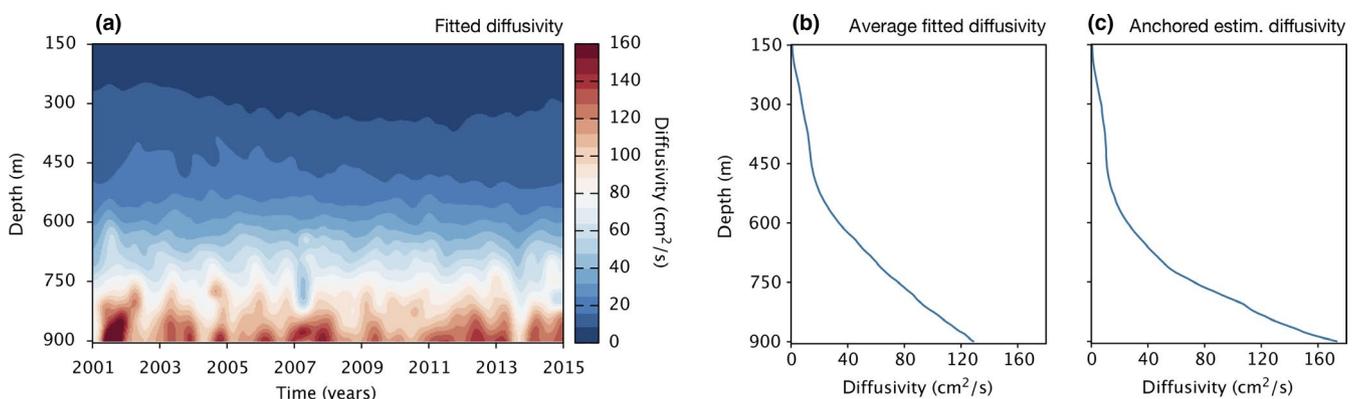


FIGURE 1 Estimated diffusivity in Cariaco Basin. (a) Diffusivity in Cariaco Basin (station CARIACO) over depth and time, estimated based on the buoyancy frequency, using the power law in Equation (1) and the fitted parameters $\alpha = 0.0001316$ and $p = 1.7433$. (b) Time-averaged diffusivity depth profile, calculated from A. (c) Estimated diffusivity depth profile, estimated using Equation (4) and assuming that D is independent of time [Colour figure can be viewed at wileyonlinelibrary.com]

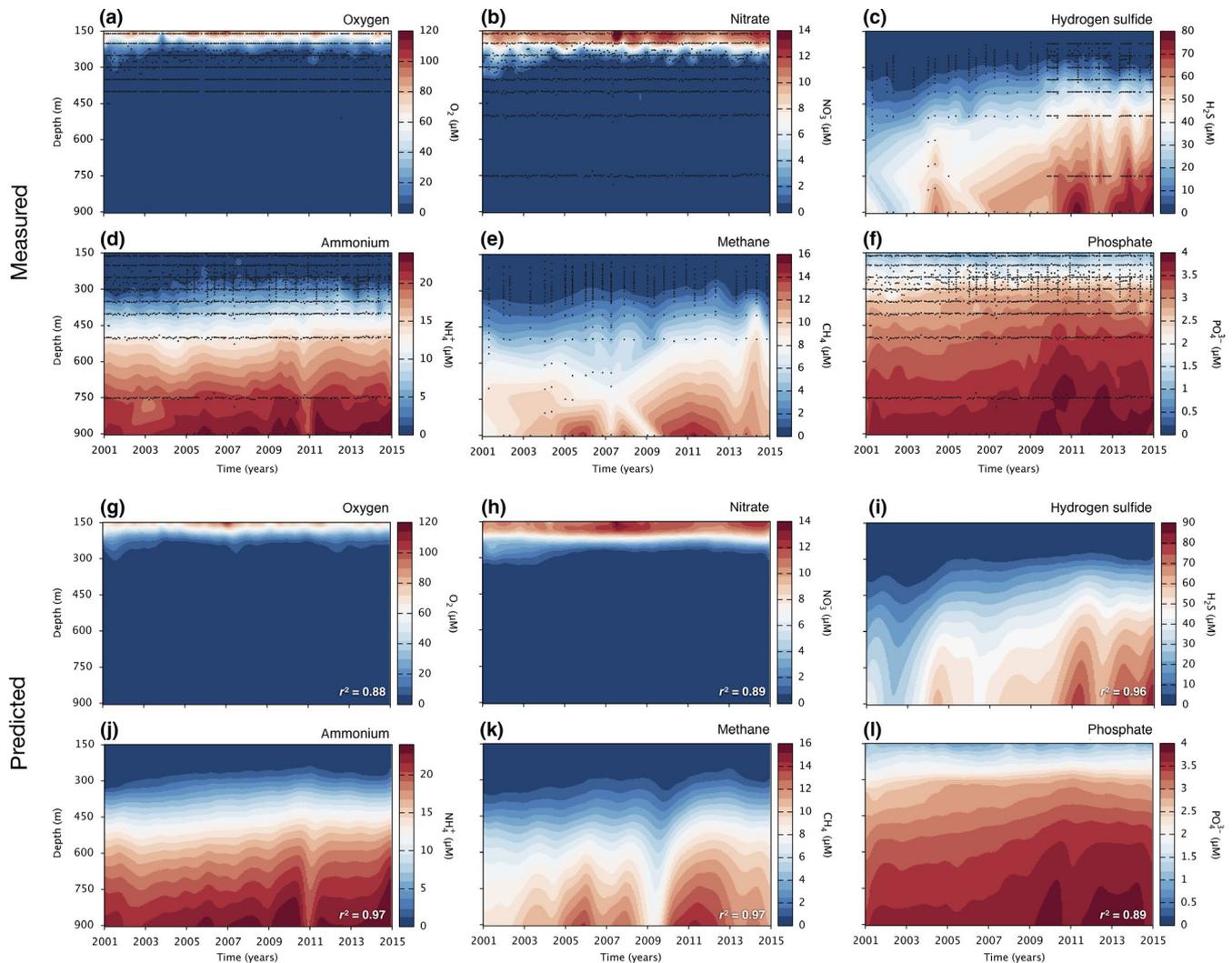


FIGURE 2 Metabolite concentrations in Cariaco Basin (data vs. fitted models). (a–f): Measured metabolite concentrations in Cariaco Basin (station CARIACO) over depth and time (a: oxygen, b: nitrate, c: hydrogen sulfide, d: ammonium, e: methane, f: phosphate). Black dots denote data points; contour plots are bilinear interpolations between data points. Data sources are described in the Methods. (g–l): Predicted metabolite concentrations, based on the net production rates estimated via ILTM. Fractions of explained variance (r^2), when compared to the data, are indicated in the figures [Colour figure can be viewed at wileyonlinelibrary.com]

where $A(z)$ is the lateral basin area at depth z and $z_0 = 150$ m. Thus, in this example, \bar{R} is the hypothetical area-specific vertical flux one would observe at depth z_0 if the total number of metabolite molecules produced between depths z_1 and z_2 (integrated over all latitudes and longitudes) was equal to the number of molecules vertically transported past depth z_0 . The depth $z_0 = 150$ m was chosen as reference because it is the approximate maximum depth at which Cariaco Basin connects to the ocean, thus marking the Basin's "upper boundary," although any other depth could have been used instead.

Net metabolite flux rates across the top (150 m) and bottom boundary (900 m) were estimated from the local metabolite concentration gradients and the local diffusivities. Gross influx rates and gross outflux rates through each boundary were then estimated by using the positive or negative part of the net flux rates, as appropriate. Net in situ production rates were estimated via ILTM fitting, as described above. Gross production rates or gross

consumption rates were then estimated by taking the positive or negative part of the net production rates, as appropriate. Note that this approach may underestimate actual production and consumption rates, if these occur concurrently at the same depth, since ILTM can a priori only reveal net rates. Estimated in situ gross production and gross consumption rates were depth-integrated using the trapezoid rule. A metabolite's mean total input rate (R_p , in $\text{mmol m}^{-2} \text{d}^{-1}$) was defined as the sum of the time-averaged depth-integrated estimated gross production rate plus its time-averaged gross influx rates at the top and bottom boundaries (all normalized to the basin area at depth 150 m). Similarly, a metabolite's mean total output rate (R_o) was defined as the sum of the time-averaged depth-integrated estimated gross consumption rate plus its time-averaged gross outflux rates at the top and bottom boundaries. A metabolite's total content (X , in mol m^{-2}) was calculated by integrating the metabolite's measured concentration over the entire depth interval while accounting for the variable

lateral (cross-sectional) basin area, and subsequently averaged over time. The mean residence time of a metabolite in the considered water column (depths 150–900 m) was estimated from the non-time-averaged total input and output rates ($R_i(t)$ and $R_o(t)$) and the non-time-averaged total content ($X(t)$), based on a non-steady-state box model (see Appendix S3). We mention that in the case of steady state the mean residence time predicted by the box model would be X/R_o ; this simplified formula was used previously by Li et al. (2012b) under the implicit assumption of steady state. All estimates are listed in Table 1 and Tables S1 and S2.

3 | RESULTS AND DISCUSSION

3.1 | Estimating diffusivity over space and time

In largely stagnant marine basins such as Cariaco Basin (Samodurov et al., 2013; Scranton et al., 1987; Taylor et al., 2018), the Black Sea (Ivanov & Samodurov, 2001), and parts of the Arabian Sea (Lam et al., 2011), eddy (turbulent) diffusion is the dominant mode of vertical transport of dissolved metabolites. Great uncertainty often exists over the magnitude of the vertical diffusion coefficient (henceforth “diffusivity”), and this uncertainty can substantially influence flux estimates. Previous theoretical and empirical work suggests that the diffusivity (denoted D) in such water columns is typically related to the buoyancy frequency (denoted N) through a power law relationship of the form:

$$D(t, z) = \alpha N(t, z)^{-p}, \quad (1)$$

where t is time, z is depth, and α and p are system-specific parameters (Armi, 1979; Gargett, 1984; Gargett & Holloway, 1984; Gregg, 1977; Gregg, D’Asaro, Shay, & Larson, 1986; Osborn, 1980; Sarmiento, Feely, Moore, Bainbridge, & Broecker, 1976; Smethie, 1980). Such a power law can be mathematically justified for stably stratified systems without double diffusion and in which the bulk of kinetic energy reaches turbulent scales via internal waves (Gargett, 1984). In these scenarios, the parameter α accounts for the average energy entering the system (e.g., by winds or tides), and p (typically between 0.5 and 1) reflects the a priori unknown N dependence of internal wave velocity variances (Gargett, 1984). Alternatively, a power law can be derived for stably stratified systems in which an apparent diapycnal diffusion-like mixing is caused mainly by turbulence near the basin bottom/boundaries and topographic features and a rapid redistribution of material throughout the interior by isopycnal advection (Armi, 1979), yielding $p \approx 2$. In some studies, a power law relationship between D and N is a purely empirical observation, with p ranging between 0.5 and 2 (Gargett, 1984; Sarmiento et al., 1976; Svensson, 1980). We mention that alternative N -dependent models for D have also been derived, based on different assumptions regarding the origin and dissipative nature of kinetic energy (Lee, Rosati, & Spelman, 2006; Munk & Anderson, 1948; Olbers & Eden, 2013).

Equation (1) has been used extensively to predict turbulent transport of dissolved gases and salts in various systems, especially in anoxic marine systems (Fennel & Boss, 2003; Ho et al., 2004; Li et al., 2012b; Louca et al., 2016; Reed, Algar, Huber, & Dick, 2014; Samodurov et al., 2013). The buoyancy frequency N can be calculated from the measured salinity and temperature profiles; however, the appropriate values for α and p are typically poorly constrained. Some studies have estimated the parameters α and p , or diffusivity itself, using spatiotemporal profiles of salinity (Gade, 1970; Ivanov & Samodurov, 2001; Lewis & Perkin, 1982; Smethie, 1980) or other conserved tracers (Gargett, 1984; Svensson, 1980). For example, Svensson (1980) used the diffusion of rhodamine B as a semi-conserved tracer to estimate a power law exponent of $p = 1.2$ in Byfjorden (Sweden), Gade (1970) used the salt budget to estimate an exponent of $p = 1.6$ in Oslofjord (Norway), and Lewis and Perkin (1982) used the salt budget to estimate D in Agfardlikavsá Fjord (Greenland), revealing a power law dependence on N with an exponent $p \approx 1.2$ (Gargett, 1984). In Cariaco Basin, analogous parameter estimates are lacking, and previous studies simply assumed an exponent of $p = 1$ (Ho et al., 2004; Li et al., 2012b; Samodurov et al., 2013; Scranton et al., 1987; Taylor et al., 2001, 2018). A value of $p = 1$ is also frequently assumed in other systems (Yakushev, 2013; Yakushev et al., 2007; Zopfi, Ferdelman, Jorgensen, Teske, & Thamdrup, 2001), although some studies instead assumed $p = 2$ (Fennel & Boss, 2003; Reed et al., 2014). The factor α is usually chosen roughly based on estimates from other marine systems (Ho et al., 2004; Li et al., 2012b; Samodurov et al., 2013; Taylor et al., 2018).

Here, to estimate both α and p for Cariaco Basin, thus resolving a major source of uncertainty in flux estimates, we used a 1-dimensional diffusive transport model for the salt budget in Cariaco Basin, and fitted the parameters α and p by minimizing the deviation of the model predictions from salinity measurements. Specifically, for any given choice of α and p , we numerically solved the diffusion equation:

$$\frac{\partial \hat{S}}{\partial t} = \frac{\partial}{\partial z} \left[D(t, z) \frac{\partial \hat{S}}{\partial z} \right] + \frac{1}{A(z)} \frac{\partial A}{\partial z} \left[D(t, z) \frac{\partial \hat{S}}{\partial z} \right], \quad (2)$$

where \hat{S} is the predicted salinity, D is the diffusivity calculated using Equation (1), and $A(z)$ is the lateral (cross-sectional) area of the basin at depth z (Figure S1). The last term in Equation (2) accounts for geometric dilution effects due to variation of the basin area over depth (Samodurov et al., 2013). This model omits occasional lateral intrusions of Caribbean Sea water (Samodurov et al., 2013; Scranton et al., 2014); the accuracy of the model is assessed in retrospect. We considered the depth range 150–900 m and the period spanning years 2001–2014, with boundary conditions provided by measured salinities at 150 m and 900 m. This depth range was chosen because 150 m is the maximum depth of the sill separating Cariaco Basin from the open ocean (and above which non-diffusive salt transport due to lateral currents is more pronounced), and because 900 m is the depth of the saddle that separates the west and east sub-basins in Cariaco (Taylor et al., 2001). The parameters α

and p were gradually adjusted using an optimization algorithm, so that the sum of squared deviations between \hat{S} and the measured salinity is minimized. This yielded the estimates $\alpha = 0.0001316$ and $p = 1.7433$, when D is measured in $\text{cm}^2 \cdot \text{s}^{-1}$ and N is measured in s^{-1} . The agreement between the predicted and measured salinity profiles was excellent, as measured by the fraction of explained variance ($r^2 = 0.982$, Figure S3B). This suggests that neglected processes, such as occasional lateral water intrusions, only have a minor influence on the Cariaco Basin salt budget during the considered time interval.

To test whether our diffusivity estimates are sensitive to the choice of model, we also considered an alternative model known as Munk–Anderson scheme (James, 1977; Lee et al., 2006; Munk & Anderson, 1948):

$$D(t, z) = D_0 \cdot \left[1 + \sigma \cdot N(t, z)^2 \cdot 2 \frac{(H-z)^2}{B^2} \right]^{-Q}, \quad (3)$$

where D_0 , σ , Q and B are system-specific model parameters and H is the bottom depth ($H \approx 1,400$ m for Cariaco Basin). The Munk–Anderson scheme assumes that diffusion-like vertical mixing is driven by frictional velocity shear, induced by tidal motions damped near the basin bottom (Munk & Anderson, 1948). The specific formula in Equation (3) is based on an empirical logarithmic profile of horizontal current velocities (James, 1977) and was used by Lee et al. (2006) in a North Atlantic ocean model. When we fit the above model to the salinity profiles in Cariaco Basin (depths 150–900 m, years 2001–2014), we obtained very similar diffusivity estimates as with the power law model (Figure S4A,B). Further, when we combined both models into a single additive model (i.e., using the sum of Equations 1 and 3), we again obtained similar diffusivity estimates as before (Figure S4C,D). While our findings do not resolve which model provides the most suitable mechanistic explanation of mixing in Cariaco Basin, all models yield similar estimates for the effective diffusivity.

To further confirm the robustness of our model-based estimates, we also considered a model-independent approach, in which diffusivity is estimated directly from the salinity data (S) regardless of the buoyancy frequency and without assuming a particular process as the cause of mixing (Figure 1c, details in Appendix S1). Briefly, this alternative approach assumes that D is constant over time and somehow known at some given “anchor depth” z_a . In this case, $D(z)$ can be estimated using the implicit formula:

$$\hat{D}(z) = \frac{1}{\partial_z U(z)} \cdot [M(z) + D(z_a) (\partial_z U)(z_a) - L(z)], \quad (4)$$

where $M(z)$, $U(z)$, and $L(z)$ are auxiliary quantities calculated using the salinity data, as follows:

$$M(z) = \int_{z_a}^z dx [S(t_2, x) - S(t_1, x)],$$

$$U(z) = \int_{t_1}^{t_2} ds S(s, z), \quad (5)$$

$$L(z) = \int_{z_a}^z dx \hat{D}(x) \frac{\partial_x A}{A(x)} \partial_x U(x),$$

and where $t_1 < t_2$ are any two time points. The accuracy of the “anchored” estimate \hat{D} defined in Equation (4) improves for larger considered time spans $|t_1 - t_2|$, and hence, we used the full available time range (years 2001–2014). Because the auxiliary variable L itself depends on the estimated diffusivity \hat{D} , an iterative approach was used to solve for \hat{D} . The choice of z_a is in principle arbitrary, as long as $D(z_a)$ can be determined somehow. A similar approach was used previously by Samodurov et al. (2013) to estimate D in Cariaco Basin, using the anchor depth $z_a = 150$ m, with an important difference: Samodurov *et al.* estimated $D(z_a)$ using the buoyancy-frequency-based formula in Equation (1), with α based on other marine systems and assuming $p = 1$, whereas here we made no assumption about $D(z_a)$ and instead estimated $D(z_a)$ from the salinity data via least-squares fitting (details in Appendix S1). We emphasize that this estimate is strictly speaking only valid if the true diffusivity D does not vary with time, and hence, it should only be used as a rough sanity check. This alternatively estimated diffusivity profile again closely reproduces the measured salinity profile ($r^2 = 0.982$, Figure S3C) and also approximately resembles our previous diffusivity estimates (Figure 1b), further increasing our confidence in these estimates.

As seen in Figure 1a, the estimated diffusivity increases drastically with depth, due to the decreased buoyancy frequency and the super-linear scaling of D ($p > 1$). Our diffusivity estimates, particularly those near the bottom, are substantially higher than typical diffusivities estimated in other fjords and basins (Gargett, 1984; Yakushev et al., 2007). In Cariaco Basin, stratification is extremely weak toward the bottom, allowing for more rapid diapycnal mixing than in the redoxcline. An exponent p greater than 1 and closer to 2 also suggests that mixing in Cariaco Basin is largely driven by turbulence near the basin's boundaries (Armi, 1979), especially at depth where the lateral area decreases substantially (Figure S1). Compared to other prominent anoxic marine systems, Cariaco Basin is relatively compact, with a horizontal area (at the sill's depth) about 40 times smaller than the Black Sea (Kideys, 2002) and the Baltic Sea (Leppäranta & Myrberg, 2009) and about 130 times smaller than the Arabian Sea (Goyet et al., 1998), potentially resulting in stronger boundary mixing than in those other systems. We point out that, in reality, the eddy diffusivity may exhibit substantial lateral heterogeneity, and may be greater near the Basin's walls than in the center. The diffusivity profile estimated here thus represents the effective (laterally averaged) diffusivity under a 1-dimensional transport model that describes laterally averaged vertical fluxes. Such a model is itself only valid under the implicit assumption that lateral mixing is much faster than vertical mixing—a reasonable assumption for Cariaco Basin, since stratification is largely vertical.

Our diffusivity estimates are substantially higher than estimates from previous studies in Cariaco Basin, all of which assumed an exponent of $p = 1$ (Li et al., 2012b; Samodurov et al., 2013; Scranton et al., 1987; Taylor et al., 2018). An exponent greater than 1 ($p \approx 1.7$) is strongly supported by our fitted model, as the value $p = 1$ results in a much lower goodness of fit (Figure S13). An exponent $p > 1$ is also consistent with our alternatively estimated diffusivity profiles (Figure 1c and Figure S4). Hence, previous studies probably underestimated D in Cariaco Basin, especially in deeper waters. As we discuss below, this has substantial implications for metabolite flux estimates and may explain some of the apparent imbalances between metabolite supply and biological demand previously observed in Cariaco Basin.

3.2 | Inverse linear transport modeling

Our approach for estimating vertical fluxes and volume-specific net production rates of dissolved metabolites is based on the following 1-dimensional reaction-diffusion differential equation for the metabolite's volumetric concentration:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left[D(t, z) \cdot \frac{\partial C}{\partial z} \right] + \frac{1}{A(z)} \frac{\partial A}{\partial z} \left[D(t, z) \cdot \frac{\partial C}{\partial z} \right] + R(t, z). \quad (6)$$

Here, $C(t, z)$ is the metabolite's concentration (mol/L), $D(t, z)$ is the diffusivity (as estimated above), $A(z)$ is the lateral (cross-sectional) area of the basin, and $R(t, z)$ is the a priori unknown volume-specific net metabolite production rate at any time t and depth z (mol L⁻¹ day⁻¹). Similarly to the salinity model above, this model can account for geometric dilution effects due to variations of the lateral basin area with depth (Samodurov et al., 2013). Variants of the above model have been used extensively to describe dissolved nutrient transport in Cariaco Basin (Cernadas-Martín et al., 2017; Li et al., 2012b; Scranton et al., 1987; Taylor et al., 2018), although previous studies made simplifying assumptions such as that R was negligible (Scranton et al., 1987), that D was constant over time (Li et al., 2012b; Taylor et al., 2018), or that C was at steady state (i.e., $\partial C/\partial t = 0$; Li et al., 2012b; Cernadas-Martín et al., 2017; Taylor et al., 2018). Since C has been measured and D has been previously estimated, in principle one could directly calculate the unknown rate R at various times and depths through a simple algebraic reordering of Equation (6). Unfortunately, this approach generally suffers from high estimation errors. The main reason is that the calculation of spatial derivatives from discrete depth profiles, or of temporal derivatives from discrete time series, typically leads to an amplification of high-frequency noise (Knowles & Renka, 2014).

An alternative approach for estimating R that reduces estimation noise and avoids the risk of overfitting is to choose R on a finite spatiotemporal grid ("fitting grid"), such that the corresponding predicted distribution \hat{C} obtained by solving the differential Equation (6) best matches the observed profile C . This approach, known as "inverse linear transport modeling" (ILTM), is widely used in oceanography and atmospheric sciences, where known

distributions of compounds are used to estimate unknown sources and sinks (Berg et al., 1998; Hirsch et al., 2006; Houweling, Kaminski, Dentener, Lelieveld, & Heimann, 1999; Lam et al., 2011; Lettmann et al., 2012; Louca et al., 2016; Martínez-Camara, Béjar Haro, Stohl, & Vetterli, 2014; Mikaloff Fletcher et al., 2006, 2007; Steinkamp, 2011). We mention that most existing studies—including those investigating metabolite fluxes in anoxic water columns or sediments (Berg et al., 1998; Lam et al., 2011; Lettmann et al., 2012; Louca et al., 2016)—assumed that C was at steady state even when fluxes were estimated at multiple time points; however, this assumption may be needlessly and overly restrictive. To reduce spurious oscillations in the estimated R (a common ILTM artifact), excessively high estimates of R that only marginally improve the agreement with the data are penalized, a procedure known as Tikhonov regularization (Björck, 1996; Hansen, 2000; Lettmann et al., 2012). Specifically, for any given metabolite, the vector containing all values of R on the fitting grid (denoted \mathbf{R}) is estimated by minimizing the expression:

$$E(\mathbf{R}) = \|\mathbf{C}^o + \mathbb{T} \cdot \mathbf{R} - \mathbf{C}\|^2 + \|\lambda \mathbf{R}\|^2 \quad (7)$$

Here, \mathbf{C} is a vector listing measured concentrations at arbitrary spacetime points, \mathbf{C}^o is a pre-calculated vector listing concentrations predicted in the absence of any net production (i.e., when $R = 0$ and accounting for initial and boundary conditions, Figure S7), $\|\cdot\|^2$ denotes the squared norm of a vector (i.e., the sum of squares of all its components), and $E(\mathbf{R})$ denotes the function to be minimized by appropriate choice of \mathbf{R} . The matrix \mathbb{T} maps net production rates on the fitting grid to concentrations on the same spacetime points as the data, and is pre-calculated using the differential Equation (6). The first $\|\cdot\|^2$ term in Equation (7) corresponds to the deviation of the predicted concentrations from the data, while the second $\|\cdot\|^2$ term corresponds to the overall magnitude of the estimated net production rates. The "regularization factor" λ modulates the penalization of spurious rate estimates, balanced against achieving a better fit to the data, and is chosen adaptively and separately for each metabolite depending on the data using a cross-validation algorithm (Golub, Heath, & Wahba, 1979). Hence, for a chosen λ , the task of estimating net production rates based on concentration data translates to an optimization problem, which can be solved numerically using linear algebra software (Appendix S4). Because all data points \mathbf{C} are used concurrently to fit the full spatiotemporal rate profile \mathbf{R} , this method is more robust against measurement errors than previous methods that only use data from a single time point at a time (Berg et al., 1998; Lam et al., 2011; Lettmann et al., 2012; Louca et al., 2016).

We emphasize that the resolution and placement of the fitting grid must be chosen carefully to avoid the risk of overfitting. Indeed, the number of spacetime points on the fitting grid dictates the number of fitted free parameters, and hence, the fitting grid must be much coarser than the concentration data \mathbf{C} . At the same time, the fitting grid should capture the major variations in R over space and time, as indicated in the concentration data. Hence, the

fitting grid should be densest in those spacetime regions where R is suspected to vary most and where, ideally, concentration measurements are also densest. The latter constraint underscores the importance of carefully choosing the times and depths targeted by oceanographic surveys. In practice, the fitting grid may need to be revised through trial and error and using expert knowledge of the system, for example, where substantial oscillations in the estimated rates are obviously spurious. A common and useful difference between spurious and true variations in the estimated R is that the former tend to be much more sensitive to small variations in the fitting grid.

3.3 | Metabolite fluxes in Cariaco Basin

We used the above ILTM approach to estimate net in situ production rates of several important dissolved metabolites in the Cariaco Basin sub-euphotic water column, using concentration time series spanning depths 150–900 m and years 2001–2014 (Figure 2a–f). In addition to production rate estimates, we also estimated vertical metabolite flux rates from the bottom (depths > 900 m) and from the overlying waters (depths < 150 m) into the sub-euphotic zone

and into the redoxcline. Estimated net metabolite production rates, interpolated between grid points, are shown in Figure 3. The agreement between the measured metabolite concentrations and those predicted based on the estimated net production rates was generally good, with a fraction of explained variance (r^2) between 0.878 and 0.973 depending on the metabolite (Figure 2). The main features not captured by the fitted models are rapid fluctuations constrained within small depth intervals, potentially originating from occasional lateral water intrusions (Muller-Karger et al., 2019; Scranton et al., 2014), as well as seasonally driven variations in oxygen and nitrate concentrations (Figure 2). In contrast, the fitted models accurately capture major decadal trends, most prominently seen in the sulfide and methane profiles (Figure 2c,e). The inability of the fitted models to capture rapid transient small-scale fluctuations stems from two fundamental information-theoretical limitations: First, the spatiotemporal resolution of the available data imposes a bound on the resolution of the fitting grid on which R can be independently estimated, and greater grid resolutions would substantially increase the risk of overfitting. Second, the estimation of R mathematically corresponds to an inversion (specifically, a deconvolution) of the diffusion process and hence tends to amplify high-frequency noise in the data;

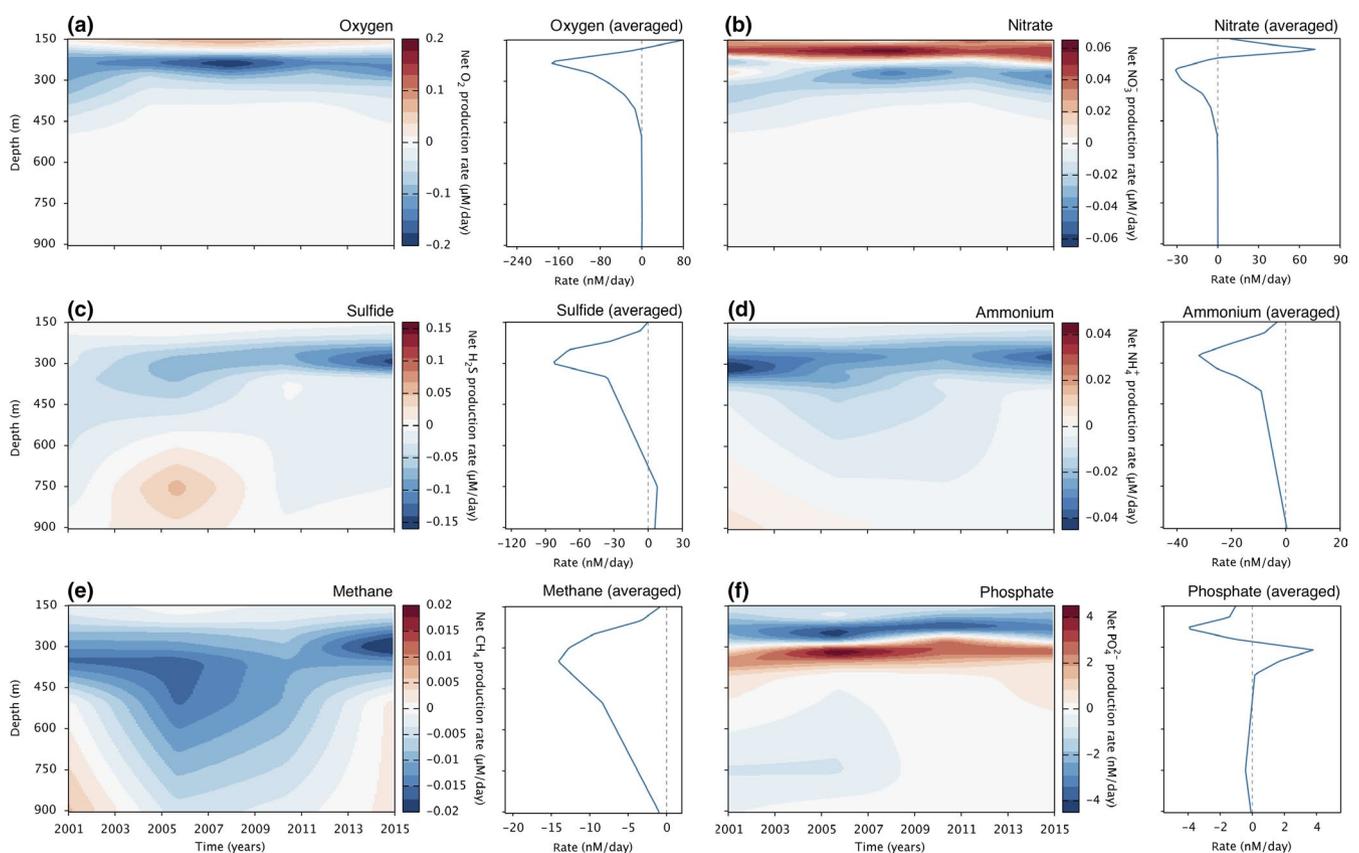


FIGURE 3 Estimated net metabolite production rates in Cariaco Basin. Volume-specific net metabolite production rates in Cariaco Basin (station CARIACO) over depth and time (contour plots) or averaged over time (depth profiles), estimated via inverse linear transport modeling (a: oxygen, b: nitrate, c: hydrogen sulfide, d: ammonium, e: methane, f: phosphate). In the contour plots, red values correspond to net production, blue values correspond to net consumption, and white corresponds to zero net production/consumption. Dashed lines at zero in the time-averaged depth profiles are shown for reference. For estimated gross production and gross consumption rates, see Figures S8 and S9, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

the amplified noise manifests as spurious rapid oscillations in the estimated rates that are hard to distinguish from real fluctuations (Lettmann et al., 2012; Steinkamp, 2011). The temporal resolution of our estimates is thus constrained by the time scales associated with diffusive mixing in Cariaco Basin which, based on the typical travel times of diffusing particles between the bottom boundary and the redoxcline, are in the order of ~ 2.4 years (Appendix S2). Thus, estimated metabolic rates at any time point should be seen as local temporal averages over those time scales.

Our estimates clearly indicate a production of nitrate near the top (depths 150–250 m) and its consumption in the immediately underlying layers (250–300 m, Figure 3b). The weak apparent production of oxygen estimated near the top (Figure 3a) is likely due to advective (e.g., lateral) transport and/or estimation error, rather than actual in situ production at those depths. When integrated across all depths, estimated in situ nitrate production near the top almost exactly matches the in situ consumption of nitrate immediately below, whereas most of the oxygen consumed in the redoxcline originates from much shallower depths (<150 m, summaries in Table 1). These observations are not surprising, since the main sources of oxygen are the atmosphere and primary production at the surface, while nitrate is likely largely produced by nitrifiers throughout the oxycline wherever ammonium is available and used at depth mostly as an electron acceptor for respiration (Cernadas-Martín et al., 2017; Scranton et al., 2014; Taylor et al., 2018). Using negative values of R as an estimate of gross consumption rates, and integrating over all depths, we estimate a gross oxygen consumption of ~ 13 mmol $m^{-2} day^{-1}$ and a gross nitrate consumption of ~ 2.6 mmol $m^{-2} day^{-1}$ on average, indicating that oxygen is a more important terminal electron acceptor in this system than nitrate. Major reductants such as hydrogen sulfide, ammonium, and methane diffuse upwards from the bottom layers into the redoxcline where they are largely consumed (Figures 3c–e). The biologically driven transport of electrons from upward diffusing electron donors onto downward diffusing electron acceptors fuels chemolithoautotrophic microbial activity within the redoxcline (Figure 4a) and sustains high prokaryotic cell densities (Figure 4d; Taylor et al., 2006). When integrated (summed) over all considered depths, we estimate an average consumption rate of

11 mmol $m^{-2} day^{-1}$ for sulfide, 4.7 mmol $m^{-2} day^{-1}$ for ammonium, and 3.3 mmol $m^{-2} day^{-1}$ for methane, where all area-specific quantities reported here and below are normalized to the basin area at depth 150 m for ease of comparison. The bulk of sulfide, ammonium and methane consumption was found to occur within the redoxcline (overviews in Table 1 and Tables S1 and S2).

Our estimated sulfide and methane consumption rates are much greater than those estimated by previous studies (~ 0.1 – 1.3 mmol $m^{-2} day^{-1}$ for sulfide; Taylor et al., 2001; Li et al., 2012b; Taylor et al., 2018 or ~ 0.04 – 0.07 mmol $m^{-2} day^{-1}$ for methane; Ward, Kilpatrick, Novelli, & Scranton, 1987; Li et al., 2012b), especially when considering that these previous values would be further reduced after accounting for the smaller basin area at the depths where they were measured (compared to 150 m). This disagreement can be largely explained by the lower diffusivities assumed or estimated in these studies; these studies do acknowledge the great uncertainty in their diffusivity estimates. Our work thus provides a possible explanation for a heavily discussed apparent “conundrum,” whereby sulfide and other electron donor fluxes estimated for Cariaco Basin appeared too low to explain measured dark carbon assimilation (DCA) rates (Jost, 2012; Li, Taylor, Astor, Varela, & Scranton, 2012a; Li et al., 2012b; Taylor et al., 2001). For example, sulfide fluxes into the redoxcline estimated by Li et al. (2012b) are about 50 times lower than ours; assuming a stoichiometric ratio of 1 mol C fixed per mol H_2S oxidized, Li *et al.* estimated that only 0.2%–4.2% of the depth-integrated DCA rate could be explained by sulfide fluxes or, alternatively, that 212 mol C had to be assimilated per mol H_2S oxidized on average. According to our sulfide consumption rate estimate (11 mmol $m^{-2} day^{-1}$ on average) and depth-integrated DCA rates (31.6 mmol-C $m^{-2} day^{-1}$ on average, Figure 4a and Taylor et al., 2018), and assuming that sulfide (eventually oxidized to sulfate) is the major source of energy for primary production in the redoxcline (Li et al., 2012b; Taylor et al., 2018), we estimate an average system-wide yield factor of ~ 2.9 mol C fixed per mol H_2S oxidized. This estimate is still higher than yield factors previously obtained from laboratory cultures of sulfide oxidizers (0.14–0.42 mol C fixed per mol H_2S ; Tuttle & Jannasch, 1979; Kelly, 1990). One explanation may be that energy limitation in Cariaco Basin's stagnant sub-oxic

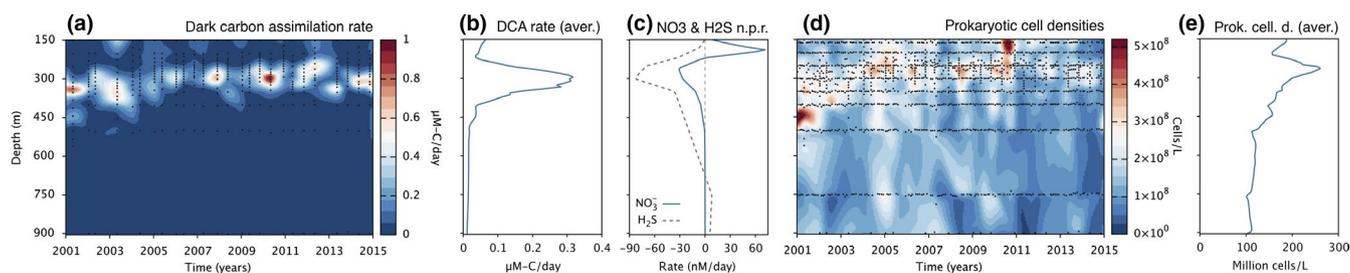


FIGURE 4 Microbial productivity measured in Cariaco Basin. (a) Dark carbon assimilation (DCA) rate measured in Cariaco Basin (carbon fixed per volume per time) across depth and time. Black dots indicate original data points; the contour plot is obtained via bilinear interpolation. (b) Measured DCA rate, averaged over time (years 2001–2014). (c) Estimated net sulfide and nitrate consumption rates, averaged over time (reproduced from Figure 3b,c). (d) Measured prokaryotic cell densities (cells per volume) across depth and time. (e) Measured prokaryotic cell densities, averaged over time. Data sources are described in the Methods. For similar figures showing the full water column (including depths <150 m and >900 m), see Figure S5 [Colour figure can be viewed at wileyonlinelibrary.com]

waters selects for oligotrophic micro-organisms, capable of more efficient substrate use than laboratory isolates. Indeed, the energy requirements and efficiencies of various carbon fixation pathways vary widely, depending on the organisms and ecological niches filled (Bar-Even, Noor, & Milo, 2011; Berg, 2011; Klatt & Polerecky, 2015). If we follow the thermodynamic arguments by Li et al. (2012b), then sulfide-oxidizing chemolithoautotrophic communities in Cariaco Basin may be capable of fixing up to 6.6 mol C per mol H₂S, well above our estimated yield factor. Further, since we ignored the contribution of other electron donors such as ammonium and methane, our estimated yield factor is probably itself an overestimate of the true sulfide-specific yield factor. We also emphasize that this yield factor is an empirical average property of the entire microbial system during the considered time period, and may vary over depth and time depending on environmental conditions and biological interactions. The limited temporal resolution of ILTM-estimated sulfide consumption rates, compared to the rapidly fluctuating measured DCA rates (Figure 4a), currently hinders a meaningful assessment of the variability of this yield factor over time.

Most of the sulfide, ammonium, and methane input into the system (i.e., via diffusion or in situ production) can be attributed to diffusion from the bottom boundary (91%, 95%, and 97%, respectively), potentially produced near or in the underlying sediments. Our estimates suggest that some hydrogen sulfide is also produced within the water column (depths ~600–900 m), consistent with the previous detection of sulfate-reducing bacteria in sinking particles at anoxic depths (Suter et al., 2018), although some of the apparent sulfide sources may actually be sulfide diffusing out of the sediments on the basin's side walls. The contribution of in situ sulfide sources to overall sulfide fluxes into the redoxcline is relatively small (~10%) and has decreased in the latter years, based on the estimated ratio of in situ produced versus in situ consumed sulfide. A relatively minor contribution of in situ sulfide sources is consistent with previous hypotheses (Ho et al., 2004; Scranton et al., 1987). We also found that the majority of phosphate input (gross in situ production + influx across the boundaries) is due to diffusion from the bottom boundary (~65%). This phosphate influx from the bottom may partly originate from the remineralization of organic matter in the sediments. Indeed, the estimated ratio of time-averaged ammonium influx versus phosphate influx from the bottom is ~17:1, closely resembling typical stoichiometric ratios of particulate organic matter in coastal marine ecosystems (17:1 on average; Sterner et al., 2008).

Below the oxic zone, nitrate is presumably used as a terminal electron acceptor by heterotrophic and/or lithotrophic prokaryotes (Rodríguez-Mora et al., 2015; Scranton et al., 2014), fueling complete denitrification to N₂ (Montes et al., 2013) and/or partial denitrification to intermediates such as nitrite. Since nitrite rarely accumulates below 150 m (Figure S10), any produced nitrite appears to be re-oxidized to nitrate, further reduced by denitrifiers, or used to anaerobically oxidize ammonium (anammox). The occurrence of denitrification and anammox would be consistent with the reduced ratios of dissolved inorganic nitrogen to phosphorus (N/P) observed in the redoxcline (Muller-Karger et al., 2019), the detection of bacteria

capable of various denitrification steps and anammox (Cernadas-Martin et al., 2017; Rodríguez-Mora et al., 2015; Taylor et al., 2018), and similar observations in other oxygen-depleted water columns (Lam et al., 2011; Lam & Kuypers, 2011; Ulloa et al., 2012). Given that sulfide oxidation spatially overlaps substantially with nitrate consumption (Figure 4c), it is probable that nitrate is at least partly used as a terminal electron acceptor for the oxidation of various sulfur compounds, a process observed in other oxygen-depleted regions of the ocean (Canfield et al., 2010; Louca et al., 2016; Rogge et al., 2017; Schunck et al., 2013). Indeed, the Gammaproteobacterial clades BS-GSO2 and SUP05, members of which are frequently implicated in sulfide oxidation and denitrification in oxygen-poor marine systems (Fuchsman, Murray, & Staley, 2012; Glaubitz, Kießlich, Meeske, Labrenz, & Jürgens, 2013; Lavik et al., 2009; Rogge et al., 2017; Shah, Chang, & Morris, 2017; Walsh et al., 2009), have been observed at high relative abundances in the Cariaco Basin redoxcline (Rodríguez-Mora et al., 2015; Suter et al., 2018; Taylor et al., 2018).

Our estimates reveal a weak but relatively steady consumption of phosphate (0.30 mmol m⁻² day⁻¹ on average) between depths ~150–250 m, and a similarly steady production of phosphate (0.17 mmol m⁻² day⁻¹ on average) between depths ~250–350 m (Figure 3f). This spatially adjacent consumption and production of phosphate leads to the appearance of a subtle phosphate minimum and maximum around the upper and lower half of the redoxcline, respectively. This pattern has been previously partly attributed to a “metal redox shuttle,” whereby phosphate is scavenged during ferrous and manganese oxide formation in the redoxcline and subsequently redissolved at depth (Dellwig et al., 2010; McParland et al., 2015; Muller-Karger et al., 2019). Prokaryotic chemolithoautotrophic activity may also partly drive phosphate consumption within the redoxcline, as suggested by McParland et al. (2015) and, in turn, the phosphate production seen immediately below may be due to the remineralization of sinking biomass. The relative importance of such a “biomass shuttle” to the phosphate pool has so far been unclear. Assuming an atomic C:P ratio of 41 for prokaryotic cells (Vrede, Heldal, Norland, & Bratbak, 2002), and an average dark carbon assimilation rate of 5.25 mmol-C m⁻² day⁻¹ between depths 150–250 m (Figure 4a), one would predict a chemolithoautotrophy-driven phosphate consumption rate of 0.13 mmol m⁻² day⁻¹. This prediction is about half of the estimated phosphate consumption rate within that depth range. Hence, a biomass shuttle could only partly explain the phosphate sink and source within the redoxcline, further emphasizing the importance of a putative metal redox shuttle.

We find that the consumption of hydrogen sulfide, methane, and, to a lesser extent, ammonium and phosphate has gradually shifted toward shallower depths, and this shift is particularly apparent when comparing times before the year 2010 and afterward. We also estimate that in situ sulfide production at depth substantially decreased over time (Figure 3c). After 2009, the estimated amount of sulfide produced in situ became negligible (<1%) compared to sulfide diffusing from the bottom (summaries in Table S1). Concurrently, sulfide concentrations near the bottom (measured at ~1,300 m depth)

have increased steadily over time (Figure S11), potentially due to increased production in the underlying sediments, leading to higher diffusive fluxes across the bottom boundary ($\sim 10 \text{ mmol m}^{-2} \text{ day}^{-1}$ on average before 2010 and $\sim 16 \text{ mmol m}^{-2} \text{ day}^{-1}$ afterward, Figure S12). This might explain why, despite a decrease of in situ sulfide production at depth, net sulfide fluxes into the redoxcline increased ($\sim 7.8 \text{ mmol m}^{-2} \text{ day}^{-1}$ on average before 2010 and $\sim 9.7 \text{ mmol m}^{-2} \text{ day}^{-1}$ afterward). Interestingly, the upward shift of the redoxcline and the drop of in situ sulfide production coincide with major shifts in the composition of the sulfur-oxidizing community after 2009 (Taylor et al., 2018). Whether the above changes in nutrient fluxes actually affected, and/or were affected by, changes in the redoxcline-inhabiting community remains unclear.

Three words of caution are warranted. First, due to the limited spatial resolution of our data (and thus, our rate estimates), it is possible that the sinks and sources of metabolites are confined to narrower depth intervals than estimated. Consequently, putatively coupled electron donors (such as sulfide) and electron acceptors (such as oxygen or nitrate), seemingly consumed within the same zone, may in reality be consumed within distinct zones and may only be indirectly coupled through redox shuttles such as manganese and iron (Percy, Li, Taylor, Astor, & Scranton, 2008; Taylor et al., 2001). Second, with the data at hand, at each location we can a priori only estimate the local net production rate R (gross production minus gross consumption rate), but not the gross production and gross consumption rates separately. It is in principle possible that in some locations some metabolites are both produced and consumed concurrently by separate processes, as observed for sulfate and sulfide in other marine anoxic systems (Canfield et al., 2010). Third, the fact that our estimated rate profiles represent locally averaged net rates implies that, a metabolite produced and consumed in distinct zones but nevertheless in close proximity, may be subject to higher turnover rates than can be inferred from our rate profiles. For example, it is possible that nitrate produced by nitrification is rapidly consumed by denitrification in close proximity immediately below (Cernadas-Martín et al., 2017), and that we thus underestimated nitrate turnover rates in the redoxcline.

4 | CONCLUSIONS

We have described a computational approach for estimating vertical fluxes and in situ consumption/production rates of dissolved chemical compounds over space and time, via inverse transport modeling. Our approach builds upon established mathematical concepts and has been optimized for water columns or sediments with essentially 1-dimensional geochemical structure, and for which chemical concentrations have been measured at arbitrary spacetime points. We emphasize that despite the apparent simplicity of our models for Cariaco Basin, which assume that eddy diffusion is the dominant mode of salt and metabolite transport in the considered depth interval, our models manage to reproduce the salinity and metabolite

concentration data very well ($r^2 = 0.982$ for salinity and $r^2 = 0.878$ – 0.973 for metabolites).

We reconstructed vertical fluxes and in situ consumption/production rates of several biologically important metabolites in the Cariaco Basin sub-euphotic water column over the course of 14 years. This allowed us to assess the relative importance of in situ production in the water column versus supply from (or near) the underlying sediments for various reductants fueling microbial productivity in the redoxcline. By independently estimating the diffusivity in Cariaco Basin over depth and time, rather than relying on parameter values from other marine systems, we further constrained an important source of uncertainty in previous flux estimates (Ho et al., 2004; Li et al., 2012b; Samodurov et al., 2013; Taylor et al., 2018). This revealed that fluxes of various electron donors and acceptors, such as hydrogen sulfide and methane, into the redoxcline are about two orders of magnitude greater than previously estimated (Li et al., 2012b; Taylor et al., 2001, 2018). Our work thus provides a possible resolution to the long unexplained apparent mismatch between electron donor fluxes and dark carbon assimilation rates in Cariaco Basin (Jost, 2012; Li et al., 2012a, 2012b). We also estimated that chemolithoautotrophic activity and remineralization of biomass within the redoxcline only partly explain the phosphate minimum and maximum observed within the redoxcline, thus providing evidence for the existence of an alternative phosphate shuttle. Finally, our work demonstrates that, using appropriate mathematical tools, a wealth of seemingly convoluted information on microbial activity can be extracted from standard chemical concentration time series.

ACKNOWLEDGMENTS

S.L. was supported by the Biodiversity Research Centre, University of British Columbia and by a startup grant from the University of Oregon. M.D. was supported by an NSERC Discovery Grant. We thank all participants and funders of the CARIACO Ocean Time Series program, including the US National Science Foundation for long-term support and in particular through NSF grant OCE-1259110 (to M.I.S. and G.T. Taylor), the Venezuelan Consejo Nacional de Investigaciones Científicas y Tecnológicas and the Venezuelan Fondo Nacional de Ciencia y Tecnología. We thank Christopher Wolfe and three anonymous reviewers for comments on our manuscript.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

S.L. conceived the project, wrote the computer code, performed the analyses and wrote a first draft of the manuscript. M.I.S., G.T.T., and Y.M.A. were some of the coordinators of the CARIACO Ocean Time Series program. All authors helped interpret the results and contributed to the writing of the manuscript.

DATA ACCESSIBILITY

All raw data used in this article have been published previously (Muller-Karger et al., 2019; Scranton et al., 2014) and are publicly available at the Cariaco Basin Time Series project website (<http://www.imars.usf.edu/cariaco>).

CODE AVAILABILITY

Our MATLAB[®] code for estimating diffusivity based on salinity profiles and for estimating metabolite fluxes via ILTM is available online at: www.loucalab.com/archive/CariacoMetabolic

ORCID

Stilianos Louca  <https://orcid.org/0000-0001-9216-5619>

REFERENCES

- Armi, L. (1979). Effects of variations in eddy diffusivity on property distributions in the oceans. *Journal of Marine Research*, 37, 515–529.
- Bar-Even, A., Noor, E., & Milo, R. (2011). A survey of carbon fixation pathways through a quantitative lens. *Journal of Experimental Botany*, 63, 2325–2342. <https://doi.org/10.1093/jxb/err417>
- Berg, I. A. (2011). Ecological aspects of distribution of different autotrophic CO₂ fixation pathways. *Applied and Environmental Microbiology*, 77, 1925–1936.
- Berg, P., Risgaard-Petersen, N., & Rysgaard, S. (1998). Interpretation of measured concentration profiles in sediment pore water. *Limnology and Oceanography*, 43, 1500–1510. <https://doi.org/10.4319/lo.1998.43.7.1500>
- Björck, A. (1996). *Numerical methods for least squares problems*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., ... Zhang, J. (2018). Declining oxygen in the global ocean and coastal waters. *Science*, 359, eaam7240. <https://doi.org/10.1126/science.aam7240>
- Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E. F., ... Ulloa, O. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science*, 330, 1375–1378. <https://doi.org/10.1126/science.1196889>
- Cernadas-Martín, S., Suter, E. A., Scranton, M. I., Astor, Y., & Taylor, G. T. (2017). Aerobic and anaerobic ammonium oxidizers in the Cariaco Basin: Distributions of major taxa and nitrogen species across the redoxcline. *Aquatic Microbial Ecology*, 79, 31–48. <https://doi.org/10.3354/ame01817>
- Dellwig, O., Leipe, T., März, C., Glockzin, M., Pollehne, F., Schnetger, B., ... Brumsack, H.-J. (2010). A new particulate Mn–Fe–P-shuttle at the redoxcline of anoxic basins. *Geochimica Et Cosmochimica Acta*, 74, 7100–7115. <https://doi.org/10.1016/j.gca.2010.09.017>
- Duffy, D. (2001) *Green's functions with applications*. Applied mathematics. New York, NY: Taylor & Francis.
- Fennel, K., & Boss, E. (2003). Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model. *Limnology and Oceanography*, 48, 1521–1534. <https://doi.org/10.4319/lo.2003.48.4.1521>
- Fuchsman, C., Murray, J., & Staley, J. (2012). Stimulation of autotrophic denitrification by intrusions of the Bosphorus plume into the anoxic Black Sea. *Frontiers in Microbiology*, 3, 257. <https://doi.org/10.3389/fmicb.2012.00257>
- Gade, H. G. (1970). *Hydrographic investigations in the Oslofjord: A study of water circulation and exchange processes* (Vol. 1). Bergen, Norway: University of Bergen.
- Gargett, A. (1984). Vertical eddy diffusivity in the ocean interior. *Journal of Marine Research*, 42, 359–393. <https://doi.org/10.1357/002224084788502756>
- Gargett, A. E., & Holloway, G. (1984). Dissipation and diffusion by internal wave breaking. *Journal of Marine Research*, 42, 15–27. <https://doi.org/10.1357/002224084788506158>
- Glaubitx, S., Kießlich, K., Meeske, C., Labrenz, M., & Jürgens, K. (2013). SUP05 dominates the Gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central Baltic and Black Seas. *Applied and Environmental Microbiology*, 79, 2767–2776. <https://doi.org/10.1128/AEM.03777-12>
- Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21, 215–223. <https://doi.org/10.1080/00401706.1979.10489751>
- Goyet, C., Millero, F. J., O'Sullivan, D. W., Eischeid, G., McCue, S. J., & Bellerby, R. G. J. (1998). Temporal variations of pCO₂ in surface seawater of the Arabian Sea in 1995. *Deep Sea Research Part I: Oceanographic Research Papers*, 45, 609–623. [https://doi.org/10.1016/S0967-0637\(97\)00085-X](https://doi.org/10.1016/S0967-0637(97)00085-X)
- Gregg, M. C. (1977). Variations in the intensity of small-scale mixing in the main thermocline. *Journal of Physical Oceanography*, 7, 436–454. [https://doi.org/10.1175/1520-0485\(1977\)007<0436:VITIO S>2.0.CO;2](https://doi.org/10.1175/1520-0485(1977)007<0436:VITIO S>2.0.CO;2)
- Gregg, M. C., D'Asaro, E. A., Shay, T. J., & Larson, N. (1986). Observations of persistent mixing and near-inertial internal waves. *Journal of Physical Oceanography*, 16, 856–885. [https://doi.org/10.1175/1520-0485\(1986\)016<0856:OOPMAN>2.0.CO;2](https://doi.org/10.1175/1520-0485(1986)016<0856:OOPMAN>2.0.CO;2)
- Hansen, P. C. (2000). The L-curve and its use in the numerical treatment of inverse problems. In P. Johnston (Ed.), *Computational inverse problems in electrocardiology, advances in computational bioengineering* (pp. 119–142). Southampton, UK: WIT Press.
- Hirsch, A. I., Michalak, A. M., Bruhwiler, L. M., Peters, W., Dlugokencky, E. J., & Tans, P. P. (2006). Inverse modeling estimates of the global nitrous oxide surface flux from 1998–2001. *Global Biogeochemical Cycles*, 20, GB1008. <https://doi.org/10.1029/2004GB002443>
- Ho, T. Y., Taylor, G. T., Astor, Y., Varela, R., Müller-Karger, F., & Scranton, M. I. (2004). Vertical and temporal variability of redox zonation in the water column of the Cariaco Basin: Implications for organic carbon oxidation pathways. *Marine Chemistry*, 86, 89–104. <https://doi.org/10.1016/j.marchem.2003.11.002>
- Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., & Heimann, M. (1999). Inverse modeling of methane sources and sinks using the adjoint of a global transport model. *Journal of Geophysical Research: Atmospheres*, 104, 26137–26160. <https://doi.org/10.1029/1999JD900428>
- Ivanov, L., & Samodurov, A. (2001). The role of lateral fluxes in ventilation of the Black Sea. *Journal of Marine Systems*, 31, 159–174. [https://doi.org/10.1016/S0924-7963\(01\)00051-3](https://doi.org/10.1016/S0924-7963(01)00051-3)
- James, I. (1977). A model of the annual cycle of temperature in a frontal region of the Celtic Sea. *Estuarine and Coastal Marine Science*, 5, 339–353. [https://doi.org/10.1016/0302-3524\(77\)90061-5](https://doi.org/10.1016/0302-3524(77)90061-5)
- Jost, G. (2012) A comment to "The conundrum between chemoautotrophic production and reductant and oxidant supply: A case study from the Cariaco Basin", published by Li et al. (2012). *Deep Sea Research Part I: Oceanographic Research Papers*, 70, 103–105.
- Kelley, D. (2014) *oce: Analysis of oceanographic data*. R package version 0.9-13.
- Kelly, D. (1990). Energetics of chemolithotrophs. In T. Krulwich (Ed.), *Bacterial energetics* (pp. 479–503). San Diego, CA: Academic Press.
- Kideys, A. E. (2002). Fall and rise of the Black Sea ecosystem. *Science*, 297, 1482–1484.

- Klatt, J. M., & Polerecky, L. (2015). Assessment of the stoichiometry and efficiency of CO₂ fixation coupled to reduced sulfur oxidation. *Frontiers in Microbiology*, 6, 484. <https://doi.org/10.3389/fmicb.2015.00484>
- Knowles, I., & Renka, R. J. (2014). Methods for numerical differentiation of noisy data. *Electronic Journal of Differential Equations*, 21, 235–246.
- Lam, P., Jensen, M. M., Kock, A., Lettmann, K. A., Plancherel, Y., Lavik, G., ... Kuypers, M. M. M. (2011). Origin and fate of the secondary nitrite maximum in the Arabian Sea. *Biogeosciences*, 8, 1565–1577. <https://doi.org/10.5194/bg-8-1565-2011>
- Lam, P., & Kuypers, M. M. (2011). Microbial nitrogen cycling processes in oxygen minimum zones. *Annual Review of Marine Science*, 3, 317–345. <https://doi.org/10.1146/annurev-marine-120709-142814>
- Lavik, G., Stührmann, T., Brüchert, V., Van der Plas, A., Mohrholz, V., Lam, P., ... Kuypers, M. M. M. (2009). Detoxification of sulphidic african shelf waters by blooming chemolithotrophs. *Nature*, 457, 581–584. <https://doi.org/10.1038/nature07588>
- Lee, H. C., Rosati, A., & Spelman, M. J. (2006). Barotropic tidal mixing effects in a coupled climate model: Oceanic conditions in the Northern Atlantic. *Ocean Modelling*, 11, 464–477. <https://doi.org/10.1016/j.ocemod.2005.03.003>
- Leppäranta, M., & Myrberg, K. (2009). *Topography and hydrography of the Baltic Sea* (pp. 41–88). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg.
- Lettmann, K. A., Riedinger, N., Ramlau, R., Knab, N., Böttcher, M. E., Khalili, A., ... Jørgensen, B. B. (2012). Estimation of biogeochemical rates from concentration profiles: A novel inverse method. *Estuarine, Coastal and Shelf Science*, 100, 26–37. <https://doi.org/10.1016/j.ecss.2011.01.012>
- Lewis, E., & Perkin, R. (1982). Seasonal mixing processes in an arctic fjord system. *Journal of Physical Oceanography*, 12, 74–83. [https://doi.org/10.1175/1520-0485\(1982\)012<0074:SMPIAA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1982)012<0074:SMPIAA>2.0.CO;2)
- Li, X., Taylor, G. T., Astor, Y., & Scranton, M. I. (2008). Relationship of sulfur speciation to hydrographic conditions and chemoautotrophic production in the Cariaco Basin. *Marine Chemistry*, 112, 53–64. <https://doi.org/10.1016/j.marchem.2008.06.002>
- Li, X., Taylor, G. T., Astor, Y., Varela, R., & Scranton, M. I. (2012a). Response to comment on “The conundrum between chemoautotrophic production and reductant and oxidant supply: A case study from the Cariaco basin”. *Deep Sea Research Part I: Oceanographic Research Papers*, 70, 106–108. <https://doi.org/10.1016/j.dsr.2012.08.002>
- Li, X. N., Taylor, G. T., Astor, Y., Varela, R., & Scranton, M. I. (2012b). The conundrum between chemoautotrophic production and reductant and oxidant supply: A case study from the Cariaco Basin. *Deep Sea Research Part I: Oceanographic Research Papers*, 61, 1–10. <https://doi.org/10.1016/j.dsr.2011.11.001>
- Louca, S., Hawley, A. K., Katsev, S., Torres-Beltran, M., Bhatia, M. P., Kheirandish, S., ... Hallam, S. J. (2016). Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proceedings of the National Academy of Sciences of the United States of America*, 113(40), E5925–E5933. <https://doi.org/10.1073/pnas.1602897113>
- Martinez-Camara, M., Béjar Haro, B., Stohl, A., & Vetterli, M. (2014). A robust method for inverse transport modeling of atmospheric emissions using blind outlier detection. *Geoscientific Model Development*, 7, 2303–2311. <https://doi.org/10.5194/gmd-7-2303-2014>
- MATLAB (2010). *version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc.
- McParland, E., Benitez-Nelson, C. R., Taylor, G. T., Thunell, R., Rollings, A., & Lorenzoni, L. (2015). Cycling of suspended particulate phosphorus in the redoxcline of the Cariaco Basin. *Marine Chemistry*, 176, 64–74. <https://doi.org/10.1016/j.marchem.2015.07.008>
- Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., ... Mouchet, A. (2006). Inverse estimates of anthropogenic CO₂ uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, 20, GB2002.
- Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., ... Menemenlis, D. (2007). Inverse estimates of the oceanic sources and sinks of natural CO₂ and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, 21, GB1010.
- Montes, E., Altabet, M. A., Muller-Karger, F. E., Scranton, M. I., Thunell, R. C., Benitez-Nelson, C., ... Astor, Y. M. (2013). Biogenic nitrogen gas production at the oxic–anoxic interface in the Cariaco Basin, Venezuela. *Biogeosciences*, 10, 267–279. <https://doi.org/10.5194/bg-10-267-2013>
- Muller-Karger, F. E., Astor, Y. M., Benitez-Nelson, C. R., Buck, K. N., Fanning, K. A., Lorenzoni, L., ... Varela, R. (2019). The scientific legacy of the CARIACO ocean time-series program. *Annual Review of Marine Science*, 11, 5.1–5.25. <https://doi.org/10.1146/annurev-marine-010318-095150>
- Munk, W. H., & Anderson, E. R. (1948). Notes on a theory of the thermocline. *Journal of Marine Research*, 7, 276–295.
- Olbers, D., & Eden, C. (2013). A global model for the diapycnal diffusivity induced by internal gravity waves. *Journal of Physical Oceanography*, 43, 1759–1779. <https://doi.org/10.1175/JPO-D-12-0207.1>
- Osborn, T. R. (1980). Estimates of the local rate of vertical diffusion from dissipation measurements. *Journal of Physical Oceanography*, 10, 83–89. [https://doi.org/10.1175/1520-0485\(1980\)010<0083:EOTLR O>2.0.CO;2](https://doi.org/10.1175/1520-0485(1980)010<0083:EOTLR O>2.0.CO;2)
- Percy, D., Li, X., Taylor, G. T., Astor, Y., & Scranton, M. I. (2008). Controls on iron, manganese and intermediate oxidation state sulfur compounds in the Cariaco Basin. *Marine Chemistry*, 111, 47–62. <https://doi.org/10.1016/j.marchem.2007.02.001>
- Reed, D. C., Algar, C. K., Huber, J. A., & Dick, G. J. (2014). Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proceedings of the National Academy of Sciences of the United States of America*, 111(5), 1879–1884. <https://doi.org/10.1073/pnas.1313713111>
- Rodriguez-Mora, M. J., Scranton, M. I., Taylor, G. T., & Chistoserdov, A. Y. (2015). The dynamics of the bacterial diversity in the redox transition and anoxic zones of the Cariaco Basin assessed by parallel tag sequencing. *FEMS Microbiology Ecology*, 91, fiv088. <https://doi.org/10.1093/femsec/fiv088>
- Rogge, A., Vogts, A., Voss, M., Jürgens, K., Jost, G., & Labrenz, M. (2017). Success of chemolithoautotrophic SUP05 and Sulfurimonas GD17 cells in pelagic Baltic Sea redox zones is facilitated by their lifestyles as K_{and} r-strategists. *Environmental Microbiology*, 19, 2495–2506.
- Samodurov, A. S., Scranton, M. I., Astor, Y., Ivanov, L. I., Chukharev, A. M., Belokopytov, V. N., & Globina, L. V. (2013). Modeling vertical exchange of heat, salt, and other dissolved substances in the Cariaco Basin. *Deep Sea Research Part I: Oceanographic Research Papers*, 71, 61–72. <https://doi.org/10.1016/j.dsr.2012.09.001>
- Sarmiento, J., Feely, H., Moore, W., Bainbridge, A., & Broecker, W. (1976). The relationship between vertical eddy diffusion and buoyancy gradient in the deep sea. *Earth and Planetary Science Letters*, 32, 357–370. [https://doi.org/10.1016/0012-821X\(76\)90076-5](https://doi.org/10.1016/0012-821X(76)90076-5)
- Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Schmidtke, S., Stramma, L., & Visbeck, M. (2017). Decline in global oceanic oxygen content during the past five decades. *Nature*, 542, 335–339. <https://doi.org/10.1038/nature21399>
- Schunck, H., Lavik, G., Desai, D. K., Großkopf, T., Kalvelage, T., Löscher, C. R., ... LaRoche, J. (2013). Giant hydrogen sulfide plume in the oxygen minimum zone off Peru supports chemolithoautotrophy. *PLoS ONE*, 8, e68661. <https://doi.org/10.1371/journal.pone.0068661>
- Scranton, M. I., Sayles, F. L., Bacon, M. P., & Brewer, P. G. (1987). Temporal changes in the hydrography and chemistry of the Cariaco

- Trench. *Deep Sea Research Part A. Oceanographic Research Papers*, 34, 945–963. [https://doi.org/10.1016/0198-0149\(87\)90047-1](https://doi.org/10.1016/0198-0149(87)90047-1)
- Scranton, M., Taylor, G., Thunell, R., Benitez-Nelson, C., Muller-Karger, F., Fanning, K., ... Astor, Y. (2014). Interannual and subdecadal variability in the nutrient geochemistry of the Cariaco Basin. *Oceanography*, 27, 148–159. <https://doi.org/10.5670/oceanog.2014.18>
- Shah, V., Chang, B. X., & Morris, R. M. (2017). Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *ISME Journal*, 11, 263–271. <https://doi.org/10.1038/ismej.2016.87>
- Smethie, W. M. (1980). Estimation of vertical mixing rates in fjords using naturally occurring Radon-222 and salinity as tracers. In H. J. Freeland, D. M. Farmer, & C. D. Levings (Eds.), *Fjord Oceanography. NATO Conference Series (IV Marine Sciences)* (Vol. 4, pp. 241–249). Boston, MA: Springer.
- Steinkamp, K. (2011). *Inverse modeling of the sources and sinks of atmospheric CO₂: Joint constraints from the ocean and atmosphere*. Ph.D. thesis. Zurich, Switzerland: ETH Zurich.
- Sterner, R. W., Andersen, T., Elser, J. J., Hessen, D. O., Hood, J. M., McCauley, E., & Urabe, J. (2008). Scaledependent carbon: Nitrogen: Phosphorus seston stoichiometry in marine and freshwaters. *Limnology and Oceanography*, 53, 1169–1180. <https://doi.org/10.4319/lo.2008.53.3.1169>
- Suter, E. A., Pachiadaki, M., Taylor, G. T., Astor, Y., & Edgcomb, V. P. (2018). Free-living chemoautotrophic and particle-attached heterotrophic prokaryotes dominate microbial assemblages along a pelagic redox gradient. *Environmental Microbiology*, 20, 693–712. <https://doi.org/10.1111/1462-2920.13997>
- Svensson, T. (1980) Tracer measurement of mixing in the deep water of a small, stratified sill fjord. In H. Freeland, D. Farmer, & C. Levings (Eds.), *Fjord Oceanography* (Vol 4, pp. 233–240). Boston, MA: Springer.
- Taylor, G. T., Iabichella, M., Ho, T.-Y., Scranton, M. I., Thunell, R. C., Muller-Karger, F., & Varela, R. (2001). Chemoautotrophy in the redox transition zone of the Cariaco Basin: A significant midwater source of organic carbon production. *Limnology and Oceanography*, 46, 148–163. <https://doi.org/10.4319/lo.2001.46.1.0148>
- Taylor, G., Iabichella-Armas, M., Varela, R., Müller-Karger, F., Lin, X., & Scranton, M. (2006) Microbial ecology of the Cariaco basin's redoxcline: the U.S.-Venezuela Cariaco times series program. In L. N. Neretin (Ed.), *Past and Present Water Column Anoxia, Nato Science Series: IV: Earth and Environmental Sciences* (Vol 64, pp. 471–499). Dordrecht, The Netherlands: Springer.
- Taylor, G. T., Suter, E. A., Pachiadaki, M. G., Astor, Y., Edgcomb, V. P., & Scranton, M. I. (2018). Temporal shifts in dominant sulfur-oxidizing chemoautotrophic populations across the Cariaco Basin's redoxcline. *Deep Sea Research Part II: Topical Studies in Oceanography*, 156, 80–96. <https://doi.org/10.1016/j.dsr2.2017.11.016>
- Thunell, R. C., Varela, R., Llano, M., Collister, J., Muller-Karger, F., & Bohrer, R. (2000). Organic carbon fluxes, degradation, and accumulation in an anoxic basin: Sediment trap results from the Cariaco Basin. *Limnology and Oceanography*, 45, 300–308. <https://doi.org/10.4319/lo.2000.45.2.0300>
- Tuttle, J. H., & Jannasch, H. W. (1979). Microbial dark assimilation of CO₂ in the Cariaco Trench. *Limnology and Oceanography*, 24, 746–753.
- Ulloa, O., Canfield, D. E., DeLong, E. F., Letelier, R. M., & Stewart, F. J. (2012). Microbial oceanography of anoxic oxygen minimum zones. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 15996–16003. <https://doi.org/10.1073/pnas.120509109>
- Ulloa, O., Wright, J., Belmar, L., & Hallam, S. (2013). Pelagic oxygen minimum zone microbial communities. In E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes* (pp. 113–122). Berlin, Heidelberg, Germany: Springer.
- Vrede, K., Heldal, M., Norland, S., & Bratbak, G. (2002). Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton. *Applied and Environmental Microbiology*, 68, 2965–2971. <https://doi.org/10.1128/AEM.68.6.2965-2971.2002>
- Walsh, D. A., Zaikova, E., Howes, C. G., Song, Y. C., Wright, J. J., Tringe, S. G., ... Hallam, S. J. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science*, 326, 578–582. <https://doi.org/10.1126/science.1175309>
- Ward, B. B., Kilpatrick, K. A., Novelli, P. C., & Scranton, M. I. (1987) Methane oxidation and methane fluxes in the ocean surface layer and deep anoxic waters. *Nature*, 327(6119), 226–229. <https://doi.org/10.1038/327226a0>.
- Wright, J. J., Konwar, K. M., & Hallam, S. J. (2012). Microbial ecology of expanding oxygen minimum zones. *Nature Reviews Microbiology*, 10, 381–394. <https://doi.org/10.1038/nrmicro2778>
- Yakushev, E. V. (2013). RedOx Layer Model: A tool for analysis of the water column oxic/anoxic interface processes. In E. Yakushev (Ed.), *Chemical Structure of Pelagic Redox Interfaces. The Handbook of Environmental Chemistry* (Vol. 22, pp. 203–233). Berlin, Heidelberg, Germany: Springer.
- Yakushev, E., Pollehne, F., Jost, G., Kuznetsov, I., Schneider, B., & Umlauf, L. (2007). Analysis of the water column oxic/anoxic interface in the Black and Baltic seas with a numerical model. *Marine Chemistry*, 107, 388–410. <https://doi.org/10.1016/j.marchem.2007.06.003>
- Zopfi, J., Ferdelman, T. G., Jorgensen, B. B., Teske, A., & Thamdrup, B. (2001). Influence of water column dynamics on sulfide oxidation and other major biogeochemical processes in the chemocline of Mariager Fjord (Denmark). *Marine Chemistry*, 74, 29–51. [https://doi.org/10.1016/S0304-4203\(00\)00091-8](https://doi.org/10.1016/S0304-4203(00)00091-8)

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Louca S, Astor YM, Doebeli M, Taylor GT, Scranton MI. Microbial metabolite fluxes in a model marine anoxic ecosystem. *Geobiology*. 2019;17:628–642. <https://doi.org/10.1111/gbi.12357>

Microbial metabolite fluxes in a model marine anoxic ecosystem - Supplemental Information -

Stilianos Louca^{1,2,*}, Yrene M. Astor^{3,4}, Michael Doebeli^{1,2,5}, Gordon T. Taylor⁶ & Mary I. Scranton⁶

¹Biodiversity Research Centre, University of British Columbia, Canada

²Department of Zoology, University of British Columbia, Canada

³Estación de Investigaciones Marinas de Margarita, Fundación La Salle de Ciencias Naturales, Punta de Piedras, Estado Nueva Esparta, Venezuela

⁴Institute for Marine Remote Sensing, University of South Florida, USA

⁵Department of Mathematics, University of British Columbia, Canada

⁶School of Marine and Atmospheric Sciences, Stony Brook University, USA

*Corresponding author

Introduction

This document provides supplemental mathematical and computational details on the methods described in the main article. Also included are supporting figures and tables. MATLAB[®] code implementing our ILTM method is provided online at: www.loucalab.com/archive/CariacoMetabolic

S.1 Anchored estimation of diffusivity

Here we elaborate on the model-independent estimation of the diffusivity depth profile in Cariaco Basin using the measured salinity profiles, regardless of the buoyancy frequency. This estimation approach makes the assumption that diffusivity is constant over time and known at some “anchor depth” z_a ($D(z_a) = D_a$). This approach can be adjusted to utilize concentration data of any conserved tracer compound, whose spatiotemporal distribution is governed by a diffusion-advection process. The approach described is modified from [Samodurov *et al.* \(2013\)](#), with the important difference that [Samodurov *et al.*](#) used an empirical formula to estimate D_a based on the local buoyancy frequency, whereas here we estimated D_a via least-squares fitting to the salinity data.

In the following, we assume that the spatiotemporal distribution of some conserved tracer (such as salt) is described by a standard advection-diffusion equation, correcting for potential geometric dilution effects due to variation of the lateral (cross-sectional) basin area over depth ([Samodurov *et al.*, 2013](#)). Hence, the volumetric tracer concentration S satisfies the differential equation:

$$\partial_t S(t, z) = \partial_z [D(z)\partial_z S(t, z) - v(t, z)S(t, z)] + \frac{\partial_z A}{A(z)} \cdot [D(z)\partial_z S(t, z) - v(t, z)S(t, z)], \quad (1)$$

where t denotes time, z denotes depth, $S(t, z)$ is the tracer concentration, $D(z)$ is the diffusivity, $v(t, z)$ is an optional vertical advection velocity and $A(z)$ is the lateral area of the basin. Here, “advection” refers

to a possible directed vertical movement of the tracer superimposed on its diffusion-like dispersal, such as upwelling driven by large-scale ocean currents or sinking due to gravity. We assume that S and v are known throughout the considered spatiotemporal domain, and that $D(z_a)$ is known at some ‘‘anchor depth’’ z_a . Integrating both sides of Eq. (1) over time between any two time points t_1, t_2 yields:

$$S(t_2, z) - S(t_1, z) = \partial_z [D(z)\partial_z U(z) - V(z)] + D(z)\frac{\partial_z A}{A(z)}\partial_z U(z) - \frac{\partial_z A}{A(z)}V(z), \quad (2)$$

where we defined the auxiliary variables:

$$\begin{aligned} U(z) &:= \int_{t_1}^{t_2} ds S(s, z), \\ V(z) &:= \int_{t_1}^{t_2} ds v(s, z)S(s, z). \end{aligned} \quad (3)$$

Integrating both sides of Eq. (2) over depth from z_a to z yields

$$M(z) = D(z)\partial_z U(z) - V(z) - D(z_a)(\partial_z U)(z_a) + V(z_a) + L(z) - K(z), \quad (4)$$

where we defined the auxiliary variables:

$$\begin{aligned} M(z) &:= \int_{z_a}^z dx [S(t_2, x) - S(t_1, x)], \\ L(z) &:= \int_{z_a}^z dx D(x)\frac{\partial_x A}{A(x)}\partial_x U(x), \\ K(z) &:= \int_{z_a}^z dx \frac{\partial_x A}{A(x)}V(x). \end{aligned} \quad (5)$$

Rearranging the algebraic Eq. (4) yields:

$$\boxed{D(z) = \frac{1}{\partial_z U(z)} \cdot [M(z) + V(z) - L(z) + K(z) + D_a \cdot (\partial_z U)(z_a) - V(z_a)]} \quad (6)$$

for $D(z)$. The above formula is theoretically valid for any time interval $[t_1, t_2]$, provided that D is constant through time during that time interval, however in practice (i.e., for measured S) the equation becomes more accurate when considering larger time intervals $|t_2 - t_1|$. Since v, S and D_a are assumed to be known, all terms on the right hand side of Eq. (6) except for $L(z)$ are known or can be estimated via finite differences. If the basin width was constant for all depths ($\partial_z A = 0$), the unknown term $L(z)$ would be zero, and hence Eq. (6) would provide an explicit estimator for D . If A varies with depth (as in Cariaco Basin), L will typically be non-zero and dependent on D , and hence Eq. (6) only provides an implicit formula for calculating D . One approach to estimating D is to initially ignore geometric dilution effects (i.e. assume $L = 0$) to obtain a first approximation for D , and then iteratively refine the estimated D by using values of L calculated with the D estimated from the previous iteration. Our analyses for Cariaco showed that this iterative estimation usually converges very fast (within 5–10 iterations), yielding values of D that fully satisfy Eq. (6).

Application to Cariaco Basin: Here, we used salinity data available between dates Jan. 11, 2001 – Dec. 9, 2014 to estimate $D(z)$ between depths 150–900 m for any given choice of D_a and $z_a=150$ m, using Eq. (6) and the measured salinity profile S . The advection term was assumed to be negligible ($v = 0$), since salt transport in Cariaco Basin is likely largely diffusive (Samodurov *et al.*, 2013, Scranton *et al.*, 1987,

Taylor *et al.*, 2018); the accuracy of the purely diffusive model is confirmed in retrospect. To calculate the various variables in Eq. (6), we interpolated the salinity data onto a regular spatiotemporal grid via bilinear interpolation. To reduce measurement noise, we smoothed the interpolated salinity profiles using a Savitzky-Golay filter of degree 2 along the depth axis (Savitzky and Golay, 1964). All integrals were calculated using the trapezoid rule. We fitted D_a as an unknown parameter by minimizing the deviation of the corresponding predicted salinity from the true (measured) salinity. Specifically, for any choice of D_a , we estimated $\hat{D}(z)$ via iterative application of Eq. (6) until convergence, and subsequently solved the partial differential equation

$$\frac{\partial \hat{S}}{\partial t} = \frac{\partial}{\partial z} \left[\hat{D}(t, z) \frac{\partial \hat{S}}{\partial z} \right] + \frac{1}{A} \frac{\partial A}{\partial z} \left[\hat{D}(t, z) \frac{\partial \hat{S}}{\partial z} \right], \quad (7)$$

with Dirichlet boundary conditions set by the measured salinities at 150 m and 900 m. The parameter D_a was gradually adjusted using the optimization algorithm `fmincon` in MATLAB[®], so that the sum of squared deviations between \hat{S} and the scattered salinity data (S) was minimized. Note that this is equivalent to maximizing the fraction of explained variance (r^2) mentioned in the main article. Fitting of D_a was repeated 200 times with random start values to avoid non-global local optima (Supplemental Fig. S14). This yielded the estimate $D_a = 0.661 \text{ cm}^2 \cdot \text{s}^{-1}$. The agreement between the predicted and measured salinity profiles was excellent, as measured by the fraction of explained variance ($r^2 = 0.982$, Supplemental Fig. S3C).

S.2 Mean diffusive travel times

In the following, we explain how we calculated the mean travel time of diffusing particles, originating at the bottom boundary ($a = 900 \text{ m}$) and arriving for the first time at the redoxcline ($b = 300 \text{ m}$). The mean travel time can be seen as an approximate temporal scale over which dissolved reductants such as hydrogen sulfide or methane, originating at the bottom boundary, are transported across the Cariaco water column and into the redoxcline. These time scales, in turn, constrain the temporal resolution of our ILTM-derived production rate estimates.

For simplicity, in the following calculation we assumed that diffusivity is constant over time, and equal to the time-averaged estimated diffusivity (Fig. 1B in the main article). Mean travel times estimated using the time-dependent diffusivity (Fig. 1A in the main article) are similar, but depend slightly on the time at which the particle is assumed to start at the bottom boundary. Let $D(z)$ be the diffusivity and $A(z)$ the lateral basin area at any depth z . Let $\rho(t, z)$ be the bathymetric probability density (i.e., probability per infinitesimal depth interval) at time t and depth z of a diffusing particle, originating at the bottom boundary at time zero. Then ρ satisfies the diffusion equation:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial z} \left[\frac{A(z)}{A(a)} D(z) \frac{\partial \rho}{\partial z} \right], \quad (8)$$

with boundary conditions:

$$\frac{\partial \rho}{\partial z}(t, a) = 0, \quad \rho(t, b) = 0, \quad (9)$$

and initial condition:

$$\rho(0, z) = \delta(z - a), \quad (10)$$

where δ denotes the Dirac distribution. The last term in Eq. (8) accounts for geometric dilution effects, due to the variation of lateral basin area with depth. Eq. (8) can be derived from the volumetric diffusion equation for the volumetric density $\rho_v(t, z) := \rho(t, z)A(z)/A(a)$ (i.e., similar to Eq. (1)). Note that the boundary conditions are based on the assumption that the particle is lost once it reaches the redoxline ($z = b$) and reflected whenever it gets near to the bottom ($z = a$). The probability that the particle is still within the interval $[a, b]$ at time $t \geq 0$ (i.e., has not yet escaped), is given by the following expression:

$$P(t) = \int_a^b dz \rho(t, z). \quad (11)$$

This expression is known as “survival function” or “reliability function” in reliability theory (Rausand and Høyland, 2004). The mean time until the particle exits the interval $[a, b]$ (i.e., reaches the redoxline), here denoted as E_∞ , can be calculated from the survival function as follows (Rausand and Høyland, 2004, Eq. 2.18):

$$E_\infty := \int_0^\infty dt P(t) = \lim_{T \rightarrow \infty} \int_0^T dt \int_a^b dz \rho(t, z). \quad (12)$$

Defining:

$$E(T) := \int_0^T dt \int_a^b dz \rho(t, z). \quad (13)$$

allows us to write the mean travel time as the limit

$$E_\infty = \lim_{T \rightarrow \infty} E(T). \quad (14)$$

Note that $E(T)$ is monotonically increasing as $T \rightarrow \infty$. Here, we numerically solved the differential equation (8) in MATLAB[®] to calculate ρ for a sufficiently long time period, and calculated $E(T)$ for increasing T . The maximum considered T (~ 50 years) was chosen such that $E(T)$ clearly approaches an asymptote; the value of that asymptote was taken as an estimate for the mean travel time, reported in the main article.

S.3 Mean residence times

In the following, we describe our approach to estimate the expected residence times of various metabolites in the modeled Cariaco water column depth interval (depths 150–900 m), based on a metabolite’s input and output rates from that system (R_i and R_o , respectively, in $\text{mmol} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$) and the total depth-integrated content (X , in $\text{mol} \cdot \text{m}^{-2}$). Our calculations are based on a box model, whereby the considered depth interval is treated as a well-mixed box and molecules exiting from the box (output rate) at any time point (t) are chosen randomly among its current contents. Note that we do not assume that the system is at steady state.

Based on the above assumptions, the instantaneous probability rate $r(t)$ at which residing molecules exit the box is given by the relative output rate:

$$r(t) = \frac{R_o(t)}{X(t)}, \quad (15)$$

meaning that during an infinitesimal time interval $[t, t + \delta t]$ the probability that a given molecule exits the box is $\delta t \cdot r(t)$. For a molecule entering the box at time τ , its random exit time is described by the probability

density function:

$$\rho(t) = r(t)e^{-\int_{\tau}^t ds r(s)}, \quad t \geq \tau. \quad (16)$$

The above expression can be obtained by considering exits from the box as events of a nonhomogeneous Poisson process with rate $r(t)$, using standard formulas for the distribution of waiting times until the first event (Rausand and Høyland, 2004, Eq. 7.121). The expected residence time of a molecule entering at time τ is thus:

$$E(\tau) = \int_{\tau}^{\infty} dt (t - \tau) \cdot \rho(t) = e^{A(\tau)} \int_{\tau}^{\infty} dt (t - \tau) \cdot r(t)e^{-A(t)}, \quad (17)$$

where we defined the auxiliary variable:

$$A(t) := \int_a^t ds r(s), \quad (18)$$

and where a is some arbitrary fixed time point (e.g., the beginning of the time series). Using the general formula for integration by parts,

$$\int_{\alpha}^{\beta} dt u(t) \frac{dv}{dt}(t) = u(t)v(t)|_{\alpha}^{\beta} - \int_{\alpha}^{\beta} dt \frac{du(t)}{dt} v(t), \quad (19)$$

where in our case $u(t) = t - \tau$ and $v(t) = -e^{-A(t)}$, we can rewrite Eq. (17) as:

$$E(\tau) = \int_{\tau}^{\infty} dt e^{A(\tau)-A(t)}. \quad (20)$$

Note that we assumed that $\lim_{t \rightarrow \infty} te^{-A(t)} = 0$, which is given provided that the relative output rate $r(t)$ does not decay to zero over time. To calculate the mean residence time among all molecules passing through the system (\bar{E}) during some time interval $[a, b]$, we weigh $E(\tau)$ by the relative amount of molecules entering at time τ and average over all possible $\tau \in [a, b]$:

$$\bar{E} = \frac{\int_a^b d\tau R_i(\tau) \cdot E(\tau)}{\int_a^b d\tau R_i(\tau)} = \frac{1}{(b-a) \cdot \bar{R}_i} \int_a^b d\tau \int_{\tau}^{\infty} dt R_i(\tau) e^{A(\tau)-A(t)}, \quad (21)$$

where we defined

$$\bar{R}_i := \frac{1}{b-a} \int_a^b d\tau R_i(\tau). \quad (22)$$

We mention that, in the special case of steady state (R_i , R_o and X are constant over time), the mean residence time \bar{E} is simply given by the ratio X/R_o , as can be verified from Eq. (21). Switching the order of integration in Eq. (21) yields the alternative expression:

$$\bar{E} = \frac{1}{(b-a) \cdot \bar{R}_i} \int_a^{\infty} dt e^{-A(t)} \int_a^{\min(t,b)} d\tau R_i(\tau) e^{A(\tau)}. \quad (23)$$

Defining the auxiliary variable:

$$B(t) := \int_a^t d\tau R_i(\tau) e^{A(\tau)}, \quad t \in [a, b] \quad (24)$$

allows writing Eq. (23) in the more compact form:

$$\bar{E} = \frac{1}{(b-a) \cdot \bar{R}_i} \int_a^\infty dt e^{-A(t)} \cdot B(\min(t, b)). \quad (25)$$

Numerically, the infinite integral in Eq. (25) needs to be replaced by a finite integral:

$$\bar{E} \approx \frac{1}{(b-a) \cdot \bar{R}_i} \int_a^{a+T} dt e^{-A(t)} \cdot B(\min(t, b)), \quad (26)$$

where T should be sufficiently large to ensure that the resulting estimate \bar{E} converges. Further, $R_i(t)$, $R_o(t)$ and $X(t)$ may only be available for a small time interval $t \in [a, b]$ (e.g., years 2001–2014 in the case of Cariaco Basin, see main article). In that case, one needs to assume that this time interval is somehow representative of all times of interest, i.e., for which the obtained average residence time \bar{E} is meaningful. Mathematically, one could periodically extend the time series of R_i , R_o and X to a sufficiently large interval $[a, a+T]$ so that, at the very least, the right hand side of Eq. (26) approaches convergence.

Application to Cariaco Basin: For each metabolite, the input rate $R_i(t)$ was calculated for various time points during the considered time interval $[a, b]$ (e.g. years 2001–2014), by summing the estimated gross influx rates at the top and bottom boundaries and the estimated depth-integrated gross production rate at each time point (see Methods in main article). The output rate $R_o(t)$ was calculated in an analogous fashion. The total content $X(t)$ was calculated by integrating the measured concentration over the entire depth interval, while accounting for the variable lateral (cross-sectional) area of the basin. The input and output rates as well as the total content were all normalized with respect to the basin area at depth 150 m. All integrals, described below, were calculated using the trapezoid rule.

The auxiliary variable $A(t)$ was calculated as in Eq. (18), the average input rate \bar{R}_i was calculated as in Eq. (22), and the auxiliary variable $B(t)$ was calculated for each time point $t \in [a, b]$ as in Eq. (27). An approximate mean residence time (denoted \bar{E}_o) was first obtained using Eq. (26), by integrating over the time interval $[a, b]$:

$$\bar{E}_o := \frac{1}{(b-a) \cdot \bar{R}_i} \int_a^b dt e^{-A(t)} \cdot B(\min(t, b)), \quad (27)$$

This preliminary estimate was then used to choose a sufficiently large time period, $T = \max(b-a, 10 \cdot \bar{E}_o)$, to which R_i , R_o and X were periodically extended. The final \bar{E} was then calculated using Eq. (26) by integrating over $[a, a+T]$.

S.4 Inverse linear transport modeling

S.4.1 Mathematical context

In the following, we elaborate on the methodology for estimating the net *in situ* production rates of a compound from concentration data measured at arbitrary spacetime points, by means of inverse linear transport modeling. While the transport model investigated in the main article is diffusion-based, the methodology described can be adjusted to a multitude of other transport models that take the form of a linear differential

equation:

$$\frac{\partial C}{\partial t} = \mathcal{L}[C] + R, \quad (28)$$

where $C(t, z)$ is the concentration of the compound of interest, $R(t, z)$ is the unknown net production rate at time t and depth z and \mathcal{L} is a linear operator (potentially including derivatives or integrals across z). For example, for diffusive-advective transport Eq. (28) takes the form:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left[D(t, z) \frac{\partial C}{\partial z} - v(t, z) C(t, z) \right] + R(t, z), \quad (29)$$

where D is the diffusivity and v is the advection velocity. If transport is also subject to geometric dilution effects, e.g. due to a variation of lateral basin area with depth (as in Cariaco), Eq. (28) takes the form:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left[D(t, z) \frac{\partial C}{\partial z} - v(t, z) C(t, z) \right] + \frac{1}{A} \frac{\partial A}{\partial z} \cdot \left[D(t, z) \frac{\partial C}{\partial z} - v(t, z) C(t, z) \right] + R(t, z), \quad (30)$$

where $A(z)$ is the lateral basin area at depth z . In any case, the key assumption we make is that, for any given R , the solution to Eq. (28) can be expressed as follows:

$$C(t, z) = C^o(t, z) + \int_{t_o}^t ds \int_a^b dy G(t, z, s, y) R(s, y), \quad (31)$$

where G is the ‘‘Green’s function’’ (sometimes called ‘‘fundamental solution’’) for the linear operator \mathcal{L} (Duffy, 2001), C^o is the concentration profile predicted in the absence of any production or consumption (i.e., when $R = 0$), t_o is the first considered time point and $[a, b]$ is the considered depth interval. Intuitively, $G(t, z, s, y)$ is the hypothetical concentration of a compound at time t and depth z , if an instantaneous normalized point source was present at time s and depth y . Note that C^o and G can be pre-calculated regardless of any given R , provided that \mathcal{L} is fully described (e.g., in the case of diffusion, the diffusivity is known for all t and z) and the initial condition (i.e., C at time t_o) and boundary conditions (i.e., C at depths a and b) are specified. In that case, for any fixed $s \geq t_o$ and $y \in [a, b]$, $G(\cdot, \cdot, s, y)$ satisfies the differential equation:

$$\frac{\partial G(\cdot, \cdot, s, y)}{\partial t} = \mathcal{L}[G(\cdot, \cdot, s, y)] + \delta(t - s)\delta(z - y), \quad (32)$$

and the conditions:

$$G(t_o, z, s, y) = 0, \quad G(t, a, s, y) = 0, \quad G(t, b, s, y) = 0, \quad (33)$$

where δ is the Dirac distribution. Equivalently, for any given s, y , the function $G(t, z, s, y)$ can also be calculated by solving the following initial value problem (Duffy, 2001, chapter 4):

$$\begin{aligned} \frac{\partial G(\cdot, \cdot, s, y)}{\partial t} &= \mathcal{L}[G(\cdot, \cdot, s, y)], \quad t \geq s \\ G(s, z, s, y) &= \delta(z - y), \end{aligned} \quad (34)$$

with boundary conditions:

$$G(t, a, s, y) = 0, \quad G(t, b, s, y) = 0, \quad t > s, \quad (35)$$

and $G(t, z, s, y) = 0$ for $t < s$. Similarly, C^o satisfies the differential equation:

$$\frac{\partial C^o}{\partial t} = \mathcal{L}[C^o], \quad (36)$$

and the conditions:

$$\begin{aligned} C^o(t_o, z) &= I(z), \\ C^o(t, a) &= A(t), \quad C^o(t, b) = B(t), \end{aligned} \quad (37)$$

where the initial depth-profile $I(z) = C(t_o, z)$ and the boundary values $A(t) = C(t, a)$ and $B(t) = C(t, b)$ are specified beforehand. For example, Supplemental Fig. S7 shows the predicted C^o for various metabolites in Cariaco Basin.

Observe that Eq. (31) expresses the concentration profile C as a linear function of the net production rate R , plus a fixed term C^o . Hence, estimating R from some measured C translates to solving a linear algebra problem, i.e. symbolically $R = \hat{G}^{-1}(C - C^o)$, where \hat{G}^{-1} represents some kind of inversion of the integral operator $\hat{G} := \int ds \int dy G$.

S.4.2 Practical considerations

In practice, C is only known (measured) on a finite set of spacetime points, R is only estimated on a discrete set of spacetime points and $G(t, z, s, y)$ is only calculated for a finite combination of s and y . Hence, in practice \hat{G} may be represented as a matrix (denoted \mathbb{G}), and thus \hat{G}^{-1} becomes the inverse of that matrix, i.e. symbolically

$$\mathbf{R} = \mathbb{G}^{-1} \cdot (\mathbf{C} - \mathbf{C}^o). \quad (38)$$

Here, \mathbf{C} is the vector listing all measured concentrations, \mathbf{C}^o is a vector listing predicted concentrations for the same spacetime points as \mathbf{C} in the absence of any net production rate, and \mathbf{R} lists the net production rates estimated on a finite set of spacetime points. Due to inevitable measurement errors, the discrete nature of the data and \mathbb{G} , and the fact that no model completely describes reality, this inverse problem may not be solvable exactly for any given \mathbf{C} . For that reason the inverse problem must be rewritten as an optimization problem, whereby \mathbf{R} is chosen such that the predicted concentration profile $\hat{\mathbf{C}} = \mathbf{C}^o + \mathbb{G} \cdot \mathbf{R}$ has the smallest possible deviation from the measured concentrations, i.e. symbolically:

$$\text{minimize } \|\mathbf{C}^o + \mathbb{G} \cdot \mathbf{R} - \mathbf{C}\|^2 \text{ by choice of } \mathbf{R}, \quad (39)$$

where $\|\cdot\|^2$ is the squared L_2 vector norm (sum of squares of all components).

Because for typical transport processes (such as diffusion) \mathbb{G} tends to have a blurring/smoothing effect in the forward-time direction (i.e., it acts as a *low-pass filter*), its inverse \mathbb{G}^{-1} tends to amplify high-frequency noise by introducing spurious oscillations to the estimate \mathbf{R} (Lettmann *et al.*, 2012). To reduce spurious oscillations, a modified optimization problem may be solved instead:

$$\text{minimize } \|\mathbf{C}^o + \mathbb{G} \cdot \mathbf{R} - \mathbf{C}\|^2 + \|\lambda \mathbf{R}\|^2 \text{ by choice of } \mathbf{R}, \quad (40)$$

where $\lambda \in \mathbb{R}$ is a suitably chosen parameter that penalizes excessively large rate estimates (details on how to choose λ are given below). This approach for reducing errors in \mathbf{R} is called ‘‘Tikhonov regularization’’ (or ‘‘ridge regression’’), and λ is called a regularization factor (Björck, 1996, Hansen, 2000). To avoid the risk

of overfitting, the number of independently estimated net production rates should be much smaller than the number of available data points (i.e., the size of \mathbf{C}). Here, we thus only allowed \mathbf{R} to vary independently on the points of a coarse spatiotemporal grid (henceforth “fitting grid”, denoted \mathcal{F}), and subsequently bilinearly interpolated \mathbf{R} onto a much finer grid (henceforth “refined grid”, denoted \mathcal{R}). Note that this assumes that R can be approximated by a piecewise linear function, whose resolution is given by the resolution of \mathcal{F} . The bilinear interpolation can be encoded as a matrix \mathbb{I} that translates rates defined on \mathcal{F} into rates defined on \mathcal{R} , and hence the above optimization problem becomes:

$$\text{minimize } \|\mathbf{C}^o + \mathbb{G} \cdot \mathbb{I} \cdot \mathbf{R} - \mathbf{C}\|^2 + \|\lambda \mathbf{R}\|^2 \text{ by choice of } \mathbf{R} \in \mathbb{R}^{|\mathcal{F}|}. \quad (41)$$

Depending on the nature of the transport operator \mathcal{L} , G (and thus \mathbb{G}) can be calculated numerically, for example using partial differential equation solvers in the case of diffusion and advection. For example, Figs. 2G–L (main article) show the predicted concentration profiles for various metabolites in Cariaco Basin water column, based on estimated net production rates.

Because typical numerical solvers, such as the MATLAB[®] function `pdepe`, only return the solution on rectangular spatiotemporal grids rather than arbitrarily scattered spacetime points, model predictions need to be subsequently projected (e.g., via bilinear interpolation) onto the same spacetime points as the data for comparison with \mathbf{C} . This projection can be encoded in another matrix (denoted \mathbb{P}), and hence the above optimization problem is modified to:

$$\text{minimize } \|\mathbf{C}^o + \mathbb{P} \cdot \mathbb{G} \cdot \mathbb{I} \cdot \mathbf{R} - \mathbf{C}\|^2 + \|\lambda \mathbf{R}\|^2 \text{ by choice of } \mathbf{R} \in \mathbb{R}^{|\mathcal{F}|}. \quad (42)$$

Defining the matrix product $\mathbb{T} := \mathbb{P} \cdot \mathbb{G} \cdot \mathbb{I}$, allows us to write the above problem as:

$$\boxed{\text{minimize } \|\mathbf{C}^o + \mathbb{T} \cdot \mathbf{R} - \mathbf{C}\|^2 + \|\lambda \mathbf{R}\|^2 \text{ by choice of } \mathbf{R} \in \mathbb{R}^{|\mathcal{F}|},} \quad (43)$$

which is the final form mentioned in the main article. The solution to the minimization problem in Eq. (43) is given by the solution to the following “regularized normal equation”:

$$\left(\mathbb{T}^T \mathbb{T} + \lambda^2 \cdot \text{Id} \right) \cdot \mathbf{R} = \mathbb{T}^T \cdot (\mathbf{C} - \mathbf{C}^o), \quad (44)$$

where Id is the identity matrix of size $|\mathcal{F}| \times |\mathcal{F}|$ (Hansen, 2000). A solution to Eq. (44) can be obtained using the Moore-Penrose pseudoinverse:

$$\mathbf{R} = \mathbb{T}_\lambda^+ \cdot \mathbb{T}^T \cdot (\mathbf{C} - \mathbf{C}^o), \quad (45)$$

where \mathbb{T}_λ^+ is the Moore-Penrose pseudoinverse of $\mathbb{T}_\lambda := \left(\mathbb{T}^T \mathbb{T} + \lambda^2 \cdot \text{Id} \right)$.

Note that if the matrix \mathbb{T}_λ is invertible (of full rank), the Moore-Penrose pseudoinverse \mathbb{T}_λ^+ is mathematically equivalent to the inverse \mathbb{T}_λ^{-1} . However, for small λ the matrix \mathbb{T}_λ may become rank-deficient or close to rank deficient, and hence Eq. (44) may not be solvable exactly, or it may have multiple solutions, or be subject to numerical instabilities. Hence, the Moore-Penrose pseudoinverse is a numerically more flexible approach to obtaining \mathbf{R} . The Moore-Penrose pseudoinverse is frequently used as a replacement for the matrix inverse in least-squares problems similar to the above (Nashed, 2014). Further, if Eq. (44) has multiple solutions, using the Moore-Penrose pseudoinverse in Eq. (45) will always yield the solution \mathbf{R} with the smallest norm - thus corresponding to the most parsimonious of possible estimates. We also mention that the above approach can be generalized to cases where λ is a matrix rather than a scalar, for example to apply different regularization strengths at different spacetime points (Hansen, 2000).

The concentration profile predicted by the model, based on the estimated \mathbf{R} , is given by:

$$\hat{\mathbf{C}} = \mathbf{C}^o + \mathbb{T} \cdot \mathbf{R}. \quad (46)$$

The fraction of explained variance, r^2 , can then be calculated as:

$$r^2 = 1 - \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|^2}{V \cdot N}, \quad (47)$$

where N is the number of data points (i.e., the length of \mathbf{C}) and V is the sample variance of the measured concentrations.

S.4.3 Choosing the regularization factor

The regularization factor λ modulates the penalty imposed on large (by magnitude) and potentially spurious rate estimates. On the one hand, a larger λ will tend to reduce spurious oscillations in \mathbf{R} and yield a smoother \mathbf{R} over space and time. On the other hand, a larger λ will typically also result in a poorer overall fit, since goodness of fit is sacrificed in favor of small \mathbf{R} . The appropriate choice of λ is sometimes determined manually by trial and error, although more rigorous techniques for choosing λ have been developed. Here, the appropriate factor λ is chosen depending on the matrix \mathbb{T} and the vectors \mathbf{C}^o and \mathbf{C} , according to the cross-validation criterion described by [Golub *et al.* \(1979, section 2\)](#). This criterion is based on a ‘‘rotation-invariant’’ leave-one-out cross-validation (LOOCV), defined as the λ that minimizes the LOOCV sum of squared residuals in a suitably transformed data vector and design matrix. Computational details are provided below.

S.4.4 Numerical implementation

In the following, we explain step-by-step how the optimization problem in Eq. (43) is solved numerically. For illustration purposes, we focus on the case where transport is diffusive and advective, i.e. as described by Eq. (29), with Dirichlet boundary conditions and assuming a constant lateral basin area. Let $\mathcal{F} = \{\tau_1, \tau_2, \dots\} \times \{x_1, x_2, \dots\} \subset [t_o, \infty) \times [a, b]$ denote the spatiotemporal fitting grid, $\mathcal{R} = \{s_1, s_2, \dots\} \times \{y_1, y_2, \dots\} \subset [t_o, \infty) \times [a, b]$ the refined grid discussed above and $\mathcal{P} := \{t_1, t_2, \dots\} \times \{z_1, z_2, \dots\}$ the spatiotemporal ‘‘prediction grid’’ for which the numerical solutions of the transport problem will be calculated. Let $\mathcal{M} = (v_1, w_1), (v_2, w_2), \dots$ denote the spatiotemporal coordinates of the discrete concentration measurements. For ease of overview, scalar-valued functions and scalar variables (e.g., single components of vectors) are denoted using regular font characters, vectors are denoted using boldface characters (e.g. \mathbf{C}) and matrices are denoted using double-line characters (e.g., \mathbb{G}). The size of a spatiotemporal grid (i.e., the number of spacetime points on it) is denoted by $|\cdot|$; for example, $|\mathcal{F}|$ is the number of spacetime points in the fitting grid.

1. Numerically calculate the discretized Green’s function for the transport equation, with zero Dirichlet boundary conditions, by solving Eqs. (34) and (35) on the refined grid \mathcal{R} . Specifically, let G_{jnim} be

an approximation for $G(t_j, z_n, s_i, y_m)$, where G solves the differential equation problem:

$$\begin{aligned}
\partial_t G(t, z, s_i, y_m) &= \partial_z [D(t, z) \partial_z G(t, z, s_i, y_m) - v(t, z) G(t, z, s_i, y_m)] \quad \forall t \geq s_i, \\
G(t = s_i, z, s_i, y_m) &= \delta_m^{-1} \cdot H(z - y_m + \delta_m/2) H(y_m + \delta_m/2 - z), \\
G(t, a, s_i, y_m) &= 0, \quad G(t, b, s_i, y_m) = 0, \\
G(t < s_i, z, s_i, y_m) &= 0,
\end{aligned} \tag{48}$$

where H is the Heaviside step function and δ_m is the grid's depth step at depth y_m . The last condition in Eq. (48) corresponds to the causality criterion that the Green's function should fulfill, whereby any source or sink only affects future tracer distributions (Duffy, 2001, chapter 4). In the following, let ε_m be the refined grid's time step. Then, following Eq. (31), for any given net production rate R the sum

$$\sum_i \sum_m \varepsilon_i \delta_m G_{jnim} \cdot R(s_i, y_m) \tag{49}$$

becomes an approximation for $C^r(t_j, z_n)$, where C^r is a solution to the transport problem with zero Dirichlet boundary conditions and zero initial condition:

$$\begin{aligned}
\partial_t C^r &= \mathcal{L}[C^r] + R, \\
C^r(t_o, z) &= 0, \quad C^r(t, a) = 0, \quad C^r(t, b) = 0.
\end{aligned} \tag{50}$$

2. Given the (known) metabolite concentration $A(t) := C(t, a)$ and $B(t) := C(t, b)$ at the boundaries as well as the initial concentration $I(z) := C(t_o, z)$, calculate the solution \bar{C}^o to the transport problem with given boundary and initial values and zero net production rates:

$$\begin{aligned}
\partial_t \bar{C}^o(t, x) &= \mathcal{L}[\bar{C}^o], \\
\bar{C}^o(t_o, z) &= I(z), \quad \bar{C}^o(t, a) = A(t), \quad \bar{C}^o(t, b) = B(t).
\end{aligned} \tag{51}$$

If C has only been measured at discrete spacetime points, use bilinear interpolation to specify A , B and I . Note that \bar{C}^o should be calculated for spacetime points on the grid \mathcal{P} . Define \bar{C}_{jn}^o as the value of \bar{C}^o on the (j, n) -th grid point, i.e., $\bar{C}_{jn}^o := \bar{C}^o(t_j, z_n)$.

Note that for any potential net production R , the sum $\hat{C} := \bar{C}^o + C^r$ is a solution to the full transport problem

$$\begin{aligned}
\partial_t \hat{C} &= \mathcal{L}[\hat{C}] + R, \\
\hat{C}(t_o, z) &= I(z), \quad \hat{C}(t, a) = A(t), \quad \hat{C}(t, b) = B(t).
\end{aligned} \tag{52}$$

Similarly in the discrete case, the sum

$$\bar{C}_{jn}^o + \sum_i \sum_m \varepsilon_i \delta_m G_{jnim} \cdot R(s_i, y_m) \tag{53}$$

is an approximation for $\hat{C}(t_j, z_n)$.

3. In the following, consider spacetime points on any of the aforementioned grids as “flattened”, that is, indexed consecutively. For example, consider each tuple $(s_i, y_m) \in \mathcal{R}$ as a single composite coordinate, and each tuple $(t_j, z_n) \in \mathcal{P}$ as a single composite coordinate. Hence, \bar{C}_{jn}^o is a single component of a 1-dimensional vector $\bar{\mathbf{C}}^o$ of size $|\mathcal{P}|$. Similarly, $G_{jnim} = G_{(j,n),(i,m)}$ is a component of a standard matrix of size $|\mathcal{P}| \times |\mathcal{R}|$, whose columns correspond to spacetime points in \mathcal{R} and whose rows

correspond to spacetime points in \mathcal{P} . Numerically, we recommend mapping spacetime grid points to consecutive indices in either row-major or column-major format and use the same scheme throughout.

4. For easier numerical calculation of the sums in Eq. (49) and (53), rescale the discretized Green's function using the refined grid's time- and depth-steps to obtain the $|\mathcal{P}| \times |\mathcal{R}|$ matrix \mathbb{G} :

$$\mathbb{G}_{(j,n),(i,m)} := \varepsilon_i \delta_m G_{jnim}. \quad (54)$$

5. Calculate the matrix \mathbb{P} that projects concentrations on the prediction grid \mathcal{P} to concentrations on the spacetime coordinates of the concentration measurements, \mathcal{M} , via bilinear interpolation. Similarly, calculate the matrix \mathbb{I} that interpolates the fitted production rates from the fitting grid \mathcal{F} onto the refined grid \mathcal{R} , via bilinear interpolation. Note that \mathbb{P} will be of size $|\mathcal{M}| \times |\mathcal{P}|$ and \mathbb{I} will be of size $|\mathcal{R}| \times |\mathcal{F}|$. Multiply the matrices \mathbb{P} , \mathbb{G} and \mathbb{I} to obtain the matrix $\mathbb{T} = \mathbb{P} \cdot \mathbb{G} \cdot \mathbb{I}$. Multiply \mathbb{P} with $\bar{\mathbf{C}}^o$ to obtain the vector $\mathbf{C}^o = \mathbb{P} \cdot \bar{\mathbf{C}}^o$.
6. Determine the appropriate regularization factor λ depending on the matrix \mathbb{T} and the vector \mathbf{C}^o and \mathbf{C} , according to the cross-validation criterion described by [Golub *et al.* \(1979\)](#). Specifically, for any given $\lambda \in \mathbb{R}$, define the following auxiliary quantities:

$$\begin{aligned} \mathbf{B} &:= \mathbf{C} - \mathbf{C}^o, \\ \mathbb{T}_\lambda &:= \mathbb{T}^T \mathbb{T} + \lambda^2 \text{Id}, \\ \mathbb{K}_\lambda &:= \mathbb{T} \cdot \mathbb{T}_\lambda^+ \cdot \mathbb{T}^T, \end{aligned} \quad (55)$$

where \mathbb{T}_λ^+ is the Moore-Penrose pseudoinverse of \mathbb{T}_λ . Choose λ by numerically minimizing the following objective function:

$$g(\lambda) := \frac{\|\mathbf{B} - \mathbb{K}_\lambda \cdot \mathbf{B}\|^2}{(\text{tr}(\text{Id} - \mathbb{K}_\lambda))^2}, \quad (56)$$

where Id is the identity matrix and $\text{tr}(\cdot)$ denotes the trace of a matrix.

7. Estimate \mathbf{R} on the fitting grid \mathcal{F} according to Eq. (45), as follows:

$$\mathbf{R} = \mathbb{T}_\lambda^+ \cdot \mathbb{T}^T \cdot \mathbf{B}. \quad (57)$$

8. Estimate the net production rates on the refined grid \mathcal{R} as the product $\mathbb{I} \cdot \mathbf{R}$.

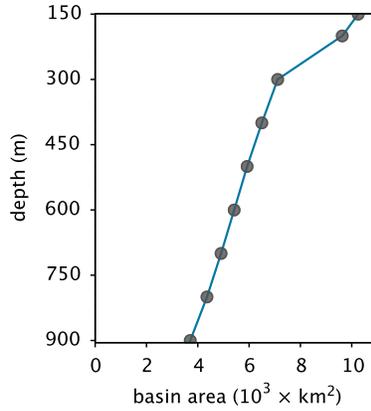


Figure S1: Cariaco Basin lateral area. Lateral (cross-sectional) area of Cariaco Basin (eastern basin) at various depths, as used in the present analyses. Data points taken from [Samodurov *et al.* \(2013, Table 1\)](#).

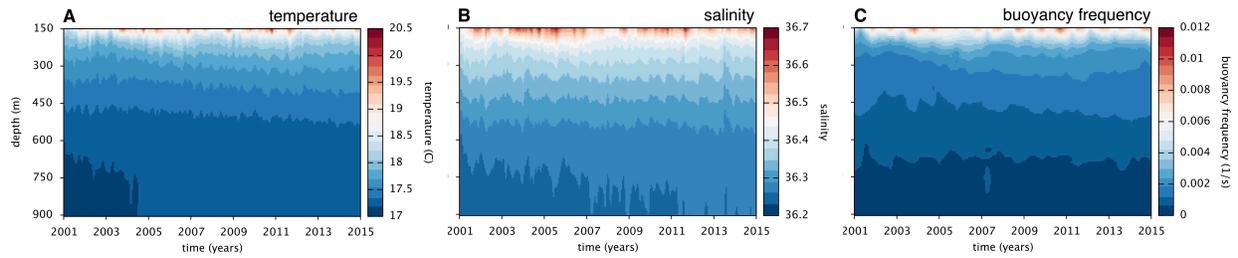


Figure S2: Physical variables in Cariaco Basin. (A) Temperature, (B) salinity and (C) buoyancy frequency in Cariaco Basin (station CARIACO) over depth and time. The latter was estimated based on temperature and salinity profiles. Data sources are described in the Methods.

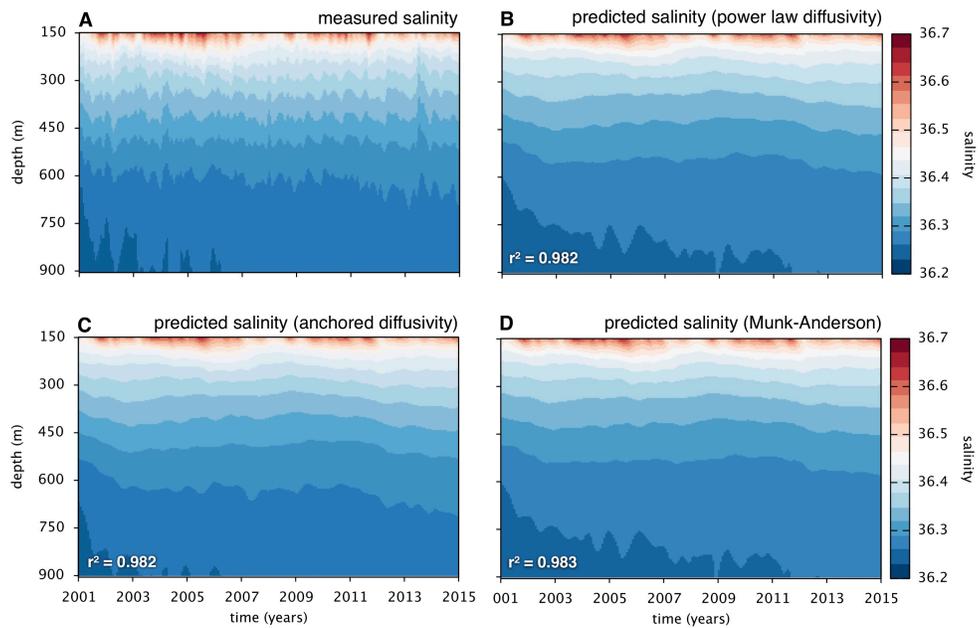


Figure S3: Salinity profiles in Cariaco (measured and fitted). (A) Measured salinity in Cariaco Basin (station CARIACO) over depth and time. Data sources are described in the Methods. (B) Salinity in Cariaco Basin, predicted by the fitted power-law diffusivity model (Eq. 1 in the main text). (C) Salinity in Cariaco Basin, predicted by the fitted anchored diffusivity (Eq. 4 in the main text). (D) Salinity in Cariaco Basin, predicted by the fitted Munk-Anderson diffusivity model (Eq. 3 in the main text). Fractions of explained variance (r^2) are written inside the figures.

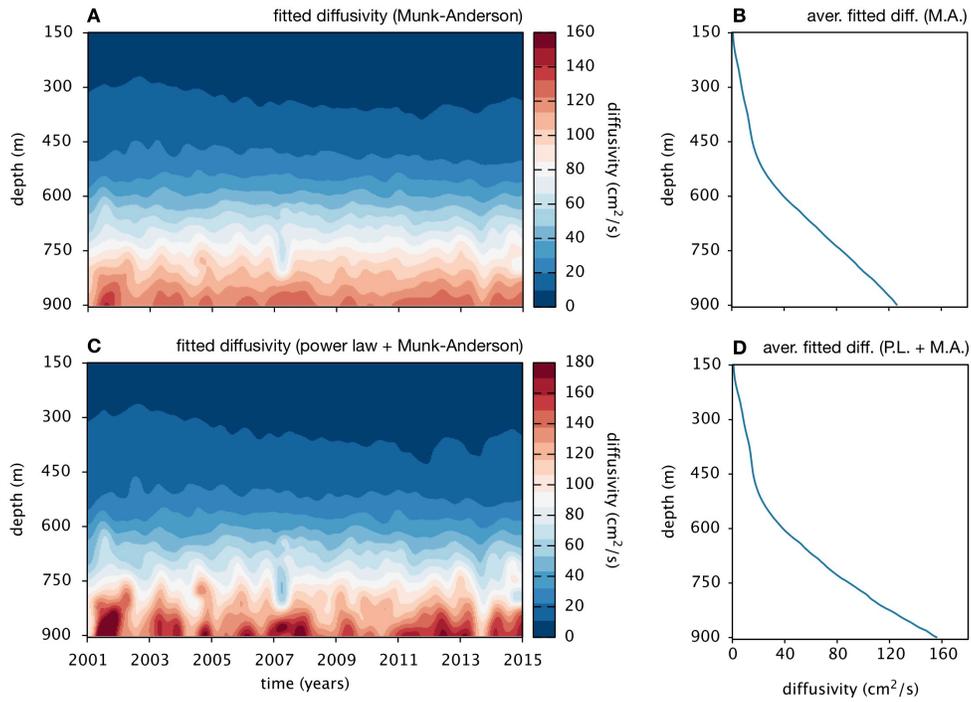


Figure S4: Estimated diffusivity in Cariaco Basin (Munk-Anderson model). (A) Diffusivity in Cariaco Basin (station CARIACO) over depth and time, estimated based on the buoyancy frequency, using the Munk-Anderson model in Eq. (3) and the fitted parameters $D_o = 173.8 \text{ cm}^2 \cdot \text{s}^{-1}$, $\sigma = 3.81$, $Q = 0.753$ and $B = 60,853 \text{ m} \cdot \text{d}^{-1}$. (B) Time-averaged diffusivity depth profile, calculated from A. (C) Diffusivity in Cariaco Basin (station CARIACO) over depth and time, estimated based on the buoyancy frequency, using the composite power-law and Munk-Anderson models (sum of in Eqs. (1) and (3)) and the fitted parameters $\alpha = 0.0000827$, $p = 1.826$, $D_o = 0.2619 \text{ cm}^2 \cdot \text{s}^{-1}$, $\sigma = 0.212$, $Q = 0.1326$ and $B = 56,022 \text{ m} \cdot \text{d}^{-1}$. (D) Time-averaged diffusivity depth profile, calculated from C. Both models yield diffusivities similar to the power-law model in Fig. 1.

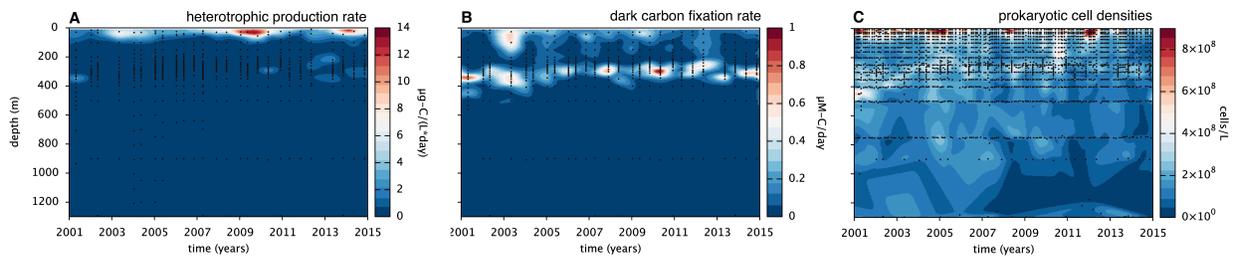


Figure S5: Microbial productivity in Cariaco Basin (all depths). (A) Bacterial heterotrophic production rates measured in Cariaco Basin station CARIACO (carbon assimilated per volume per time), across depth and time. (B) Dark carbon fixation rates measured in Cariaco Basin water column (carbon fixed per volume per time). (C) Prokaryotic cell densities (cells per volume). Black dots indicate original data points; the contour plot is obtained via linear interpolation. Data sources are described in the Methods.

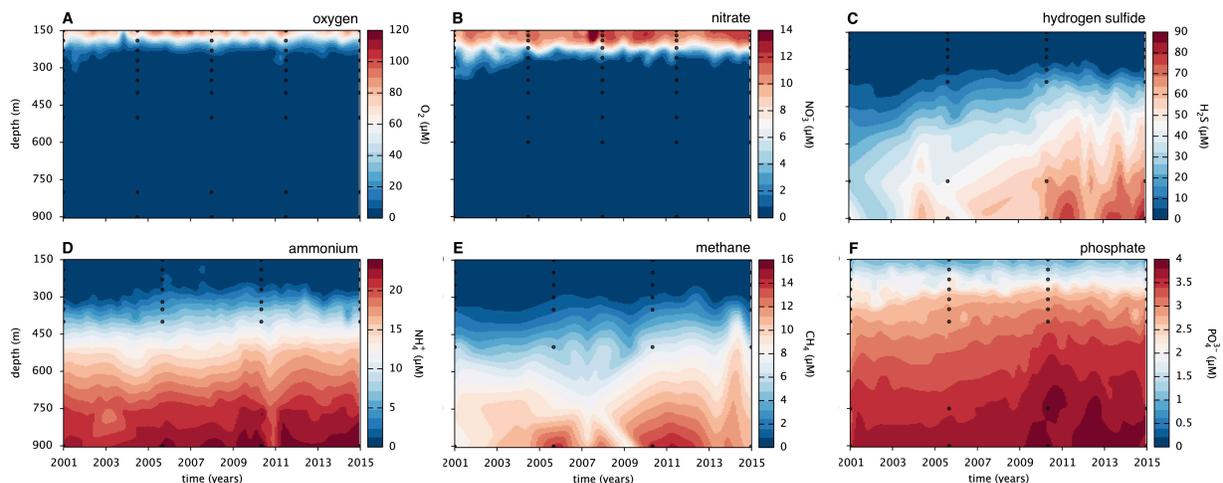


Figure S6: Metabolite concentrations and ILTM fitting grids. Measured concentrations of metabolites (oxygen, nitrate, hydrogen sulfide, ammonium, methane and phosphate) in Cariaco Basin over depth and time. ILTM fitting points are shown as black dots. Data sources are described in the Methods.

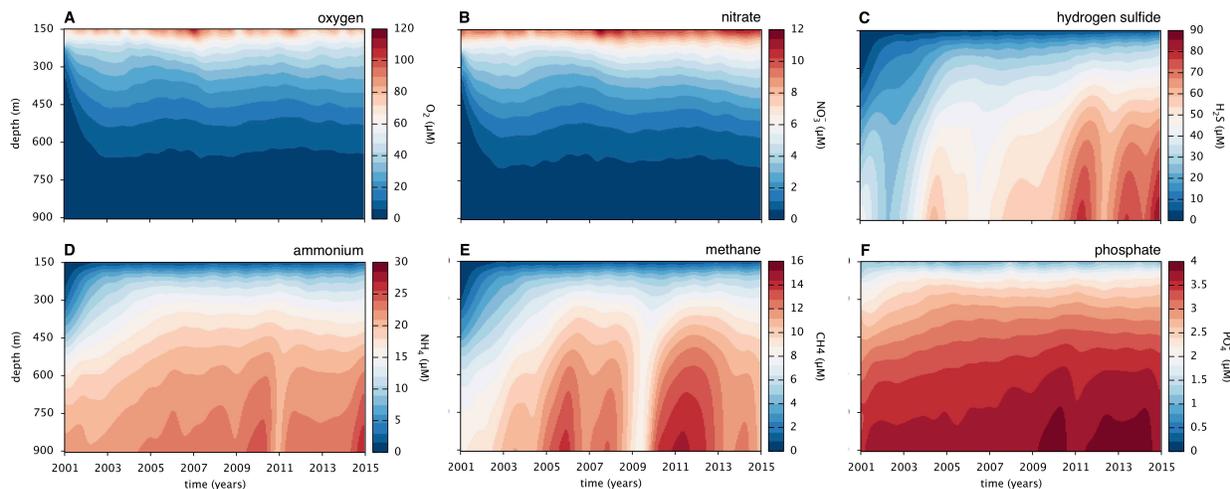


Figure S7: Metabolite concentrations in the absence of reactions. Metabolite concentrations in Cariaco Basin (station CARIACO), predicted solely based on initial conditions and diffusive transport between the top and bottom boundaries, in the absence of *in situ* reactions. This is C^o , defined in Supplement S.4.1.

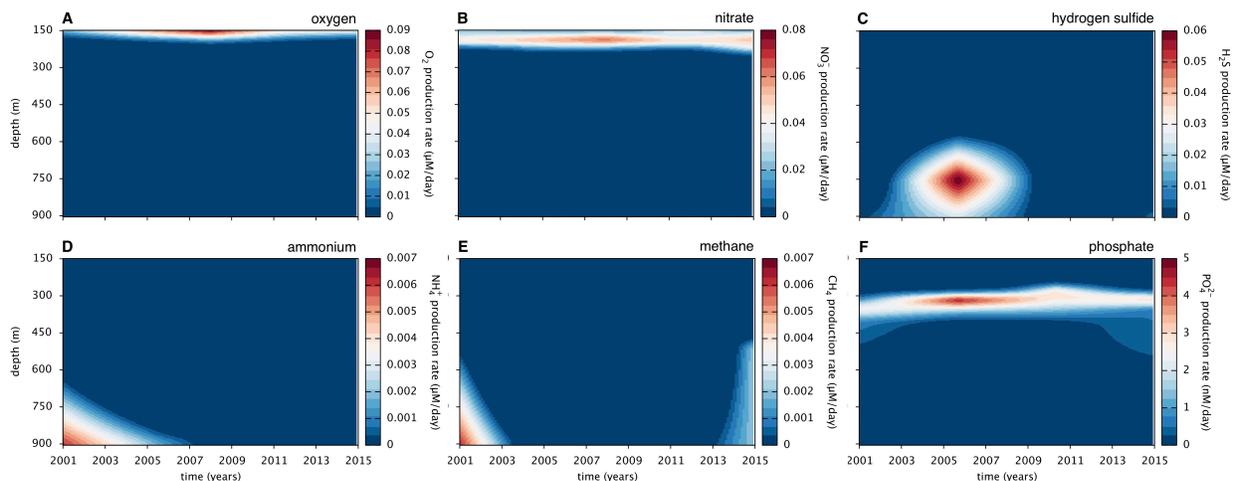


Figure S8: Estimated gross metabolite production rates. Volume-specific gross metabolite production rates in Cariaco Basin (station CARIACO) over depth and time (A: oxygen, B: nitrate, C: hydrogen sulfide, D: ammonium, E: methane, F: phosphate). Estimates are based on the positive part of the ILTM-estimated net production rates (Fig. 3 in the main article), and may thus underestimate the true gross production rates if consumption co-occurs at the same locations as production.

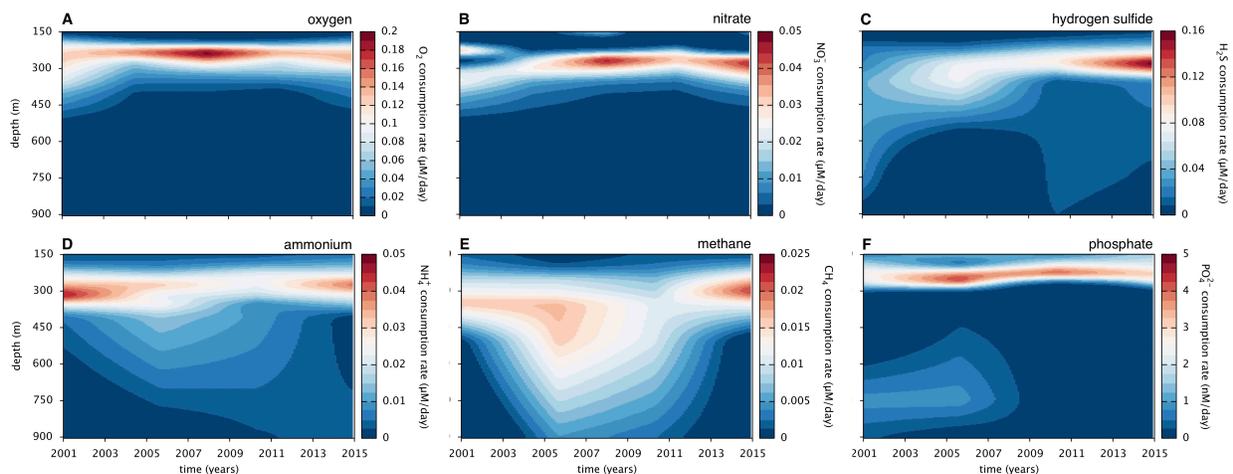


Figure S9: Estimated gross metabolite consumption rates. Volume-specific gross metabolite consumption rates in Cariaco Basin (station CARIACO) over depth and time (A: oxygen, B: nitrate, C: hydrogen sulfide, D: ammonium, E: methane, F: phosphate). Estimates are based on the negative part of the ILTM-estimated net production rates (Fig. 3 in the main article), and may thus underestimate the true gross consumption rates if production co-occurs at the same locations as consumption.

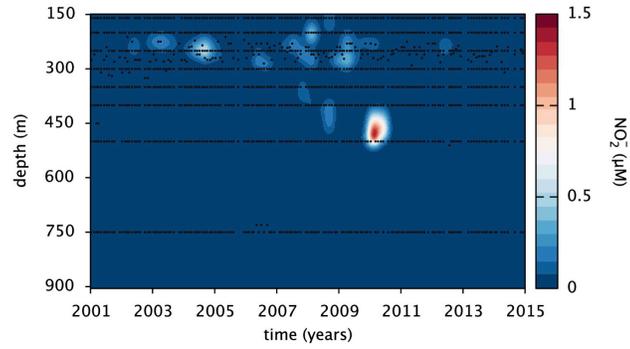


Figure S10: Nitrite concentrations in Cariaco Basin. Measured nitrite (NO_2^-) concentrations in Cariaco Basin (station CARIACO) over depth and time. Data sources are described in the Methods.

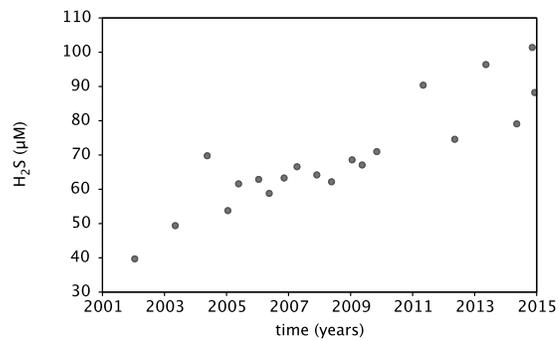


Figure S11: Hydrogen sulfide concentrations near the bottom of Cariaco Basin. Measured hydrogen sulfide (H_2S) concentrations near the bottom of Cariaco Basin (~ 1300 m, station CARIACO) over time. Data sources are described in the Methods.

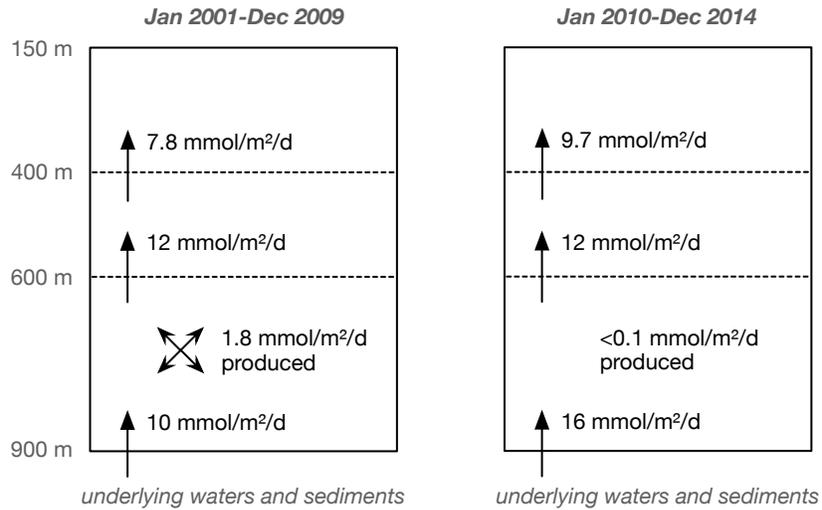


Figure S12: Summary of estimated sulfide fluxes in Cariaco Basin. Overview of hydrogen sulfide (H_2S) fluxes in the Cariaco Basin sub-euphotic water column, including vertical fluxes at depths 900 m, 600 m and 400 m, averaged over the period Jan. 1, 2001 – Dec. 31, 2009 (left box) and the period Jan. 1, 2010 – Dec. 31, 2014 (right box). All area-specific values are normalized to the lateral (cross-sectional) basin area at depth 150 m. Note that fluxes are not exactly balanced, as Cariaco Basin was not always at steady state.

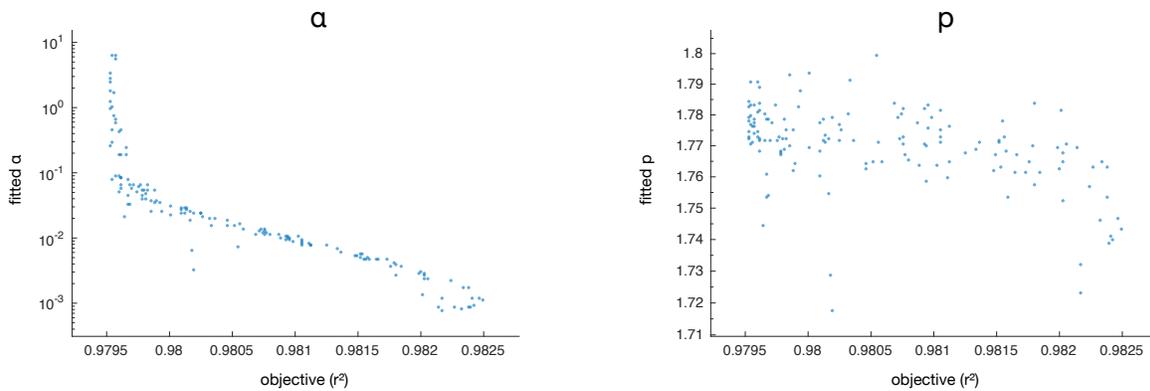


Figure S13: Fitted diffusivity model parameters versus r^2 . Fitted diffusivity power-law parameters α and p (vertical axes), compared to the corresponding maximized fitting objective (r^2 , horizontal axes), for several fitting attempts with random parameter start values (one plot per parameter, one point per fitting attempt). The best parameter estimate, discussed in the main text, corresponds to the highest r^2 . Fitted parameter values with lower r^2 correspond to local non-global optima of the objective function. Only the 50% best fits are shown.

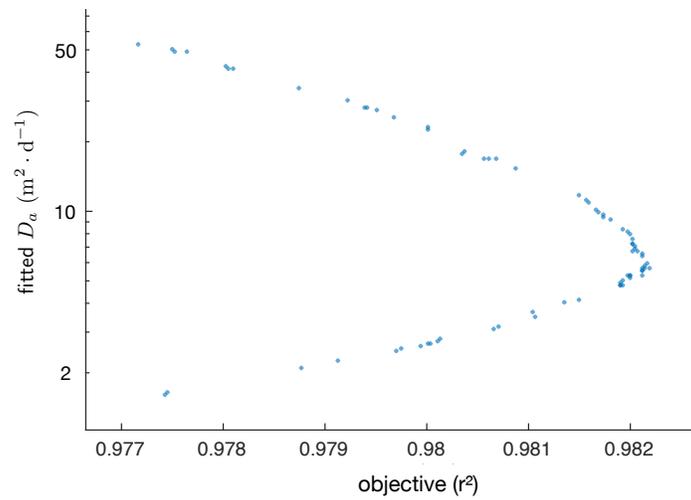


Figure S14: Fitted anchored diffusivity D_a versus r^2 . Fitted diffusivity D_a at anchor depth $z_a=150$ m (vertical axis), compared to the corresponding maximized fitting objective (r^2 , horizontal axis), for several fitting attempts with random parameter start values (one point per fitting attempt). The best parameter estimate corresponds to the highest r^2 . Fitted parameter values with lower r^2 correspond to local non-global optima of the objective function. Only the 50% best fits are shown.

Table S1: Estimated metabolic fluxes in Cariaco Basin (Jan. 1, 2010 until Dec. 31, 2014). Estimated mean *in situ* production and consumption rates, as well as influx and outflux rates across the top (150 m) and bottom boundary (900 m). The total concentration is depth-integrated, averaged over time and measured in $\text{mol} \cdot \text{m}^{-2}$. All rates are depth-integrated where applicable, averaged over time, and measured in $\text{mmol} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$. Depth-integrated or area-specific quantities take into account variation of the lateral (cross-sectional) basin area with depth, and are normalized to the basin area at depth 150 m. Mean residence times were estimated based on the depth-integrated concentrations and gross input/output rates (see Supplement S.3). See Methods for details.

variable	O₂	NO₃²⁻	H₂S	NH₄⁺	CH₄	PO₄³⁻
total content (depth-integrated)	3.4	0.81	15	5	2.5	1.3
gross production rate (depth-integrated)	2	2.7	<0.1	<0.1	<0.1	0.18
gross consumption rate (depth-integrated)	13	2.7	13	4.4	2.7	0.25
net influx rate at top (150 m)	7.4	-0.6	<0.1	0.22	<0.1	-0.075
gross influx rate at top (150 m)	7.4	<0.1	<0.1	0.22	<0.1	<0.01
gross outflux rate at top (150 m)	<0.1	0.6	<0.1	<0.1	<0.1	0.076
net influx rate at bottom (900 m)	<0.1	<0.1	16	4.4	3.1	0.16
gross influx rate at bottom (900 m)	<0.1	<0.1	16	4.4	3.1	0.16
gross outflux rate at bottom (900 m)	0.13	<0.1	<0.1	<0.1	<0.1	<0.01
net influx rate at top+bottom	7.3	-0.6	16	4.7	3.2	0.084
total gross influx+production rate	9.5	2.7	16	4.7	3.2	0.34
total gross outflux+consumption rate	13	3.3	13	4.4	2.7	0.33
mean residence time (years)	0.72	0.67	3.2	3.1	2.5	9.6

Table S2: Estimated metabolic fluxes in Cariaco Basin redoxcline (years 2001–2014, depths 200-400 m). Estimated mean *in situ* production and consumption rates in the redoxcline, as well as influx and outflux rates across the top (200 m) and bottom boundary (400 m). The total concentration is depth-integrated, averaged over time and measured in $\text{mol} \cdot \text{m}^{-2}$. All rates are depth-integrated where applicable, averaged over time, and measured in $\text{mmol} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$. Depth-integrated or area-specific quantities take into account variation of the lateral (cross-sectional) basin area with depth, and are normalized to the basin area at depth 150 m to facilitate comparisons. Mean residence times were estimated based on the depth-integrated concentrations and gross input/output rates (see Supplement S.3). See Methods for details.

variable	O_2	NO_3^{2-}	H_2S	NH_4^+	CH_4	PO_4^{3-}
total content (depth-integrated)	0.65	0.29	0.89	0.45	0.13	0.35
gross production rate (depth-integrated)	<0.1	0.55	<0.1	<0.1	<0.1	0.16
gross consumption rate (depth-integrated)	13	2.4	7.8	3.2	1.6	0.19
net influx rate at top (200 m)	16	4.1	<0.1	<0.1	<0.1	-0.14
gross influx rate at top (200 m)	16	4.1	<0.1	<0.1	<0.1	<0.01
gross outflux rate at top (200 m)	<0.1	<0.1	0.15	<0.1	<0.1	0.14
net influx rate at bottom (400 m)	0.26	<0.1	8.5	4	1.7	0.14
gross influx rate at bottom (400 m)	0.4	<0.1	8.5	4	1.7	0.14
gross outflux rate at bottom (400 m)	0.15	<0.1	<0.1	<0.1	<0.1	<0.01
net influx rate at top+bottom	16	4.1	8.4	3.9	1.7	<0.01
total gross influx+production rate	16	4.7	8.5	4	1.7	0.29
total gross outflux+consumption rate	13	2.5	8	3.3	1.6	0.33
mean residence time (years)	0.14	0.33	0.3	0.36	0.23	3

Table S3: Resolutions of data points and ILTM grids. Overview of the number of concentration measurements available for each metabolite ($|\mathcal{M}|$), and the spatiotemporal resolutions (times x depths) of ILTM fitting grids ($|\mathcal{F}|$), refined grids ($|\mathcal{R}|$) and prediction grids ($|\mathcal{P}|$). For a visual overview of data points see Figs. 2A–F in the main article. For a visual overview of fitting grids see Supplemental Fig. S6.

metabolite	data	fitting	refined	prediction
oxygen (O_2)	2,567	5×10	100×100	100×100
nitrate (NO_3^-)	3,138	5×11	100×100	100×100
sulfide (H_2S)	778	4×8	100×100	100×100
ammonium (NH_4^+)	1,739	4×8	100×100	100×100
methane (CH_4)	350	4×7	100×100	100×100
phosphate (PO_4^{3-})	1,346	4×7	100×100	100×100

References

- Björck, A. (1996) Numerical Methods for Least Squares Problems. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.
- Duffy, D.G. (2001) Green's Functions with Applications. Studies in Advanced Mathematics. New York, USA: Chapman & Hall/CRC.
- Golub, G.H., Heath, M., and Wahba, G. (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**: 215–223.
- Hansen, P.C. (2000) The L-curve and its use in the numerical treatment of inverse problems. In J. P (ed.) Computational Inverse Problems in Electrocardiology, Advances in Computational Bioengineering, pp. 119–142. Southampton, UK: WIT Press.
- Lettmann, K.A., Riedinger, N., Ramlau, R., Knab, N., Böttcher, M.E., Khalili, A., *et al.* (2012) Estimation of biogeochemical rates from concentration profiles: A novel inverse method. *Estuarine, Coastal and Shelf Science* **100**: 26–37.
- Nashed, M.Z. (ed.) (2014) Generalized Inverses and Applications: Proceedings of an Advanced Seminar Sponsored by the Mathematics Research Center, the University of Wisconsin—Madison, October 8-10, 1973. 32. Elsevier.
- Rausand, M. and Høyland, A. (2004) System Reliability Theory: Models, Statistical Methods, and Applications. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, 2nd edition. Hoboken, New Jersey: Wiley.
- Samodurov, A.S., Scranton, M.I., Astor, Y., Ivanov, L.I., Chukharev, A.M., Belokopytov, V.N., *et al.* (2013) Modeling vertical exchange of heat, salt, and other dissolved substances in the Cariaco Basin. *Deep Sea Research Part I: Oceanographic Research Papers* **71**: 61–72.
- Savitzky, A. and Golay, M.J. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**: 1627–1639.
- Scranton, M.I., Sayles, F.L., Bacon, M.P., and Brewer, P.G. (1987) Temporal changes in the hydrography and chemistry of the Cariaco Trench. *Deep Sea Research Part A. Oceanographic Research Papers* **34**: 945–963.
- Taylor, G.T., Suter, E.A., Pachiadaki, M.G., Astor, Y., Edgcomb, V.P., and Scranton, M.I. (2018) Temporal shifts in dominant sulfur-oxidizing chemoautotrophic populations across the Cariaco Basin's redoxcline. *Deep Sea Research Part II: Topical Studies in Oceanography* **156**: 80–96.