Using maximum entropy production to describe microbial biogeochemistry over time and space in a meromictic pond

Vallino, Joseph J. 1*, Huber, Julie A. 2

1 Marine Biological Laboratory, Ecosystems Center, Woods Hole, MA, US
2 Woods Hole Oceanographic Institution, Marine Chemistry & Geochemistry, Woods Hole, MA, US

* Correspondence:
Corresponding Author
jvallino@mbl.edu

Abstract

The maximum entropy production (MEP) conjecture posits that systems with many degrees of freedom will likely organize to maximize the rate of free energy dissipation. Previous work indicates that biological systems can outcompete abiotic systems by maximizing free energy dissipation over time by utilizing temporal strategies acquired and refined by evolution, and over space via cooperation. In this study, we develop an MEP model to describe biogeochemistry observed in Siders Pond, a phosphate limited meromictic system located in Falmouth, MA that exhibits steep chemical gradients due to density-driven stratification that supports anaerobic photosynthesis as well as microbial communities that catalyze redox cycles involving O, N, S, Fe and Mn. The MEP model uses a metabolic network to represent microbial redox reactions, where biomass allocation and reaction rates are determined by solving an optimization problem that maximizes entropy production over time and a 1D vertical profile constrained by an advection-dispersion-reaction model. We introduce a new approach for modeling phototrophy and explicitly represent aerobic photoautotrophs, anoxyogenic photoheterotrophs and anaerobic photoautotrophs. The metabolic network also includes reactions for heterotrophic bacteria, sulfate reducing bacteria, sulfide oxidizing bacteria and aerobic and anaerobic grazers. Model results were compared to observations of biogeochemical constituents collected over a 24 hour period at 8 depths at a single 15 m deep station in Siders Pond. Maximizing entropy production over long (3 d) intervals produced results more similar to field observations than short (0.25 d) interval optimizations, which support the importance of temporal strategies for maximizing entropy production over time. Furthermore, we found that entropy production must be maximized locally instead of globally where energy potentials are degraded quickly by abiotic processes, such as light absorption by water. This combination of field observations with modeling results show that microbial systems in nature can be accurately described by the maximum entropy production conjecture applied over time and space.
1 Introduction

Mass and energy flow associated with the growth and predation of bacteria, archaea and eucaryotes in microbial food webs coupled with abiotic reactions and transport processes define biogeochemical cycles that occur in ecosystems ranging in size from less than a liter (Marino et al., 2016) to the entire planet. Because organisms are ultimately responsible for most observed biogeochemical transformations, it is customary and natural to focus on the bioenergetics and growth and predation of the organisms that constitute food webs in order to understand and predict biogeochemical transformations. This organismal focus has a long history and has advanced our understanding and prediction of ecosystem dynamics and the mass and energy flow through them (Riley, 1946; Fasham et al., 1990; Le Quere et al., 2005; Friedrichs et al., 2007; Schartau et al., 2017). While focusing on the growth, predator-prey and cooperative interactions of organisms will continue to contribute to our understanding of ecosystem bioenergetics, there are some challenges that limit this approach for microbial systems that form the basis of biogeochemical cycles. Microbial communities are diverse, complex and abundant, consisting, for example, of upwards of $10^9$ microorganisms per liter of seawater, with estimates that Earth hosts close to 1 trillion microbial species (Sogin et al., 2006; Locey and Lennon, 2016). With the advent of metagenomics, next generation sequencing and bioinformatics, the challenging task of deciphering and annotating the metabolic capabilities and activities of bacteria and other microorganisms has begun, but untangling the Gordian knot will take significantly more time and investment (Ponomarova and Patil, 2015), and determining how information in genomes contributes to competition and cooperation is still in its infancy (Hallam and McCutcheon, 2015; Pasternak et al., 2015; Worden et al., 2015). Even more challenging is understanding and predicting community composition dynamics and succession as environmental conditions change, both from exogenous and endogenous drivers (Konopka et al., 2015). To develop predictive models of biogeochemistry based on organisms and the genes they carry, exchange horizontally and differentially express requires an enormous amount of information, which will likely take many decades to compile and decipher, but good progress is being made (Reed et al., 2014; Coles et al., 2017). While we believe this reductionist approach is essential, there is also a complementary approach to understanding microbial biogeochemistry that is less studied and uses a more thermodynamic, or whole systems, approach.

Understanding how ecosystems function at the systems level has a long tradition in theoretical ecology (Chapman et al., 2016; Vallino and Algar, 2016), but the underlying premise is that ecosystems organize so as to maximize an objective function, such as maximizing power proposed by Lotka (1922) nearly 100 years ago. The advantage of the systems approach is that the variational principle can be used to determine how an ecosystem will organize and function without the knowledge of which organisms are present and how their population changes over time, provided there is sufficient diversity. That is, understanding and modeling of ecosystems can focus on function rather than on organisms, and there is growing support that stable function can arise from dynamic communities (Fernandez et al., 1999; Fernandez-Gonzalez et al., 2016; Louca and Doebeli, 2016; Needham and Fuhrman, 2016; Coles et al., 2017). Here, we build upon the conjecture that microbial systems organize to maximize entropy production. The maximum entropy production (MEP) principle proposes that systems (biotic systems included) with sufficient degrees of freedom will likely organize to maximize the dissipation of Gibbs free energy (Dewar, 2003; Lorenz, 2003; Martyushev and Seleznev, 2006). The MEP principle has been applied to both abiotic and biotic processes (Kleidon and Lorenz, 2005; Kleidon et al., 2010; Dewar et al., 2014), and we have used MEP to model periodically forced methanotrophic microbial communities (Vallino et al., 2014) and investigate metabolic switching in nitrate reducing environments (Algar and Vallino, 2014).
Energy dissipation drives microbial function

An unresolved aspect of MEP involves the temporal and spatial scales over which it operates. MEP theory has only been derived for nonequilibrium steady-state systems, but microbial systems are dynamic, so steady state does not apply. Consequently, we have proposed that abiotic processes, such as fire or a rock rolling down a hill, maximize instantaneous entropy production; that is, they follow a steepest descent trajectory down a potential energy surface. Biological systems, however, use information acquired by evolution and culled by natural selection to follow alternate pathways that maximize entropy production over time, which allows biology to sometimes outcompete abiotic processes (Vallino, 2010). Similarly, when considering a spatial domain, entropy production can either be maximized locally at each point in the domain or entropy production at each point can be adjusted so that entropy production is maximized globally over the entire domain. A simple numerical study indicated that when a system organizes over space, entropy produced by global optimization can exceed that from local optimization, but this requires spatial coordination by the community, while abiotic systems are likely to only maximize entropy production locally (Vallino, 2011). This paper seeks to examine both of these ideas using spatially and temporally distinct observations collected from a field site.

In this study we develop an MEP-based model to predict microbial biogeochemistry in a meromictic pond located in Falmouth, MA (Siders Pond) that includes metabolic processes for phytoplankton, green sulfur bacteria, heterotrophic bacteria, sulfate reducing bacteria, sulfide oxidizing bacteria, photoheterotrophs and aerobic and anaerobic predators, but the primary objective of the model is to dissipate energy potentials, not grow organisms. While previous MEP models have been developed for chemolithoautotrophs, chemolithoheterotrophs and chemoorganoheterotrophs, this study expands the metabolic reaction repertoire to include photoautotrophs and photoheterotrophs. The approach is also extended to include an explicit spatial dimension, and we compare model output to observations in Siders Pond collected over a 24 hour sampling period from eight depths. MEP solutions using two different optimization timescales (0.25 d versus 3.0 d) are contrasted and compared to observations, and we discuss the problem of local versus global MEP optimization for energy potentials that are quickly dissipated abiotically, such as light. Our results suggest that microbial systems in nature can be accurately described by the maximum entropy production conjecture applied over time and space.

2 Methods

This section describes the development of the MEP model, including the 1D transport model for Siders Pond, followed by Siders Pond sample collection procedures and analytical methods used for sample analyses.

2.1 Model Development

The equations used to model biogeochemistry in Siders Pond are provided in detail in the Supplementary Material (SM), so the descriptions in this section focus primarily on model concepts that form the basis of the modeling approach and extensions to previous studies. The model consists of three main components: 1) a 1D transport model; 2) a set of biologically catalyzed reactions that constitute the metabolic network of the microbial community, including predators such as protist and viruses; 3) an optimization component in which control variables that govern reaction stoichiometry, kinetics and thermodynamics are determined so as to maximize internal entropy production over a specified interval of time and space. One of the underlying objectives of the approach is to replace what are typically poorly constrained biological parameters, such as maximum specific growth rates, substrate affinities, prey preferences, reaction efficiencies, etc, with optimal control variables whose values are determined by maximizing free energy dissipation. Consequently, these MEP-based
models contain relatively few adjustable parameters, but see Section 2.1.3.1 below. We describe briefly below the three model components, but we focus first on the representation of the metabolic network, as this forms the foundation of the approach, and includes the concentrations of 11 chemical constituents and 8 functional groups (Fig. 1): salt, \(c_{\text{Sal}}\); dissolved oxygen, \(c_{O_2}\); dissolved inorganic carbon, \(c_{\text{H}_2\text{CO}_3}\); inorganic phosphate, \(c_{\text{H}_3\text{PO}_4}\); sulfate, \(c_{\text{H}_2\text{SO}_4}\); hydrogen sulfide, \(c_{\text{H}_2\text{S}}\); phytoplankton carbohydrates, \(c_{\text{phy}}\); green sulfur bacteria carbohydrates, \(c_{\text{Gsb}}\); labile organic carbon, \(c_{\text{L}}\); refractory organic carbon, \(c_{\text{R}}\); refractory organic phosphate, \(c_{\text{P}}\); phytoplankton, \(c_{\text{phy}}\); green sulfur bacteria, \(c_{\text{Gsb}}\); aerobic predators, \(c_{\text{A}}\); anaerobic predators, \(c_{\text{A}}\); heterotrophic bacteria, \(c_{\text{Ba}}\); sulfite reducing bacteria, \(c_{\text{SRB}}\); photoheterotrophs, \(c_{\text{PH}}\); sulfide oxidizing bacteria, \(c_{\text{SO}}\).

### 2.1.1 Catalysts and metabolic reaction network

Our approach views a complex microbial community as a collection of catalysts (denoted with the special symbol \(\mathcal{S}_j\)) that each have a subset of \(n_{r,j}\) functional reactions they catalyze. The catalyst has an elemental composition given by \(\text{CH}_{a_j}\text{O}_{\beta_j}\text{N}_{\gamma_j}\text{O}_{\delta_j}\). For this study all catalysts are assigned the same composition as yeast with associated thermodynamic properties (Battley et al., 1997), but this is not an overly constraining approximation (Vallino et al., 2014) and can be easily relaxed if needed. The catalysts and reactions included in the metabolic network represent the capabilities of the entire microbial community, but the reactions are distributed across \(n_{\mathcal{S}}\) catalysts (8 for the case of Siders Pond, Fig. 1), just as functions are distributed across phyla in natural communities (Vallino, 2003). The reactions are highly simplified and condensed, and consist of two essential components: an anabolic reaction that synthesizes catalyst from environmental resources, and a catabolic reaction that provides free energy to drive the anabolic reaction forward. A highly simplified list of metabolic reactions for the Siders Pond model is given in Table 1. To convey the basic ideas, we consider below the phylum responsible for sulfate reduction, but see the SM for the complete derivation.

The catalyst \(\mathcal{S}_{\text{SRB}}\) represents the capabilities of sulfate reducing bacteria (SRB) that oxidize labile organic matter, \(C_L\), with an elemental composition \(\text{CH}_2\text{O}\), using sulfate as the electron acceptor. The anabolic and catabolic reactions are given by,

\[
\text{C}_L + \gamma\text{NH}_3 + \delta\text{H}_3\text{PO}_4 + a_{1,\text{SRB}}^A\text{H}_2\text{SO}_4 \rightarrow \mathcal{S}_{\text{SRB}} + a_{1,\text{SRB}}^A\text{H}_2\text{S(aq)} + b_{1,\text{SRB}}^A\text{H}_2\text{O}\]  

(1)

\[
\text{C}_L + \frac{1}{2}\text{H}_2\text{SO}_4 \rightarrow \text{H}_2\text{CO}_3 + \frac{1}{2}\text{H}_2\text{S(aq)},
\]  

(2)

respectively, where the stoichiometric coefficients, \(a_{1,\text{SRB}}^A\) and \(b_{1,\text{SRB}}^A\), are determined from elemental balances around O and H, since C, N and P are in balance. An important concept for the approach is that both reactions above must be catalyzed by \(\mathcal{S}_{\text{SRB}}\), so the reaction rate is proportional to the concentration of the catalyst, \(c_{\mathcal{S}_{\text{SRB}}}\) or \(\left[\mathcal{S}_{\text{SRB}}\right]\), present. These two reactions can be combined by introducing a reaction weighting factor, or growth efficiency variable, \(\varepsilon_{\text{SRB}}\), to produce an overall reaction representing sulfate reducing bacteria, \(r_{1,\text{SRB}}\), given by,

\[
\text{C}_L + \varepsilon_{\text{SRB}}(\gamma\text{NH}_3 + \delta\text{H}_3\text{PO}_4) + \left(\frac{1}{2} + \varepsilon_{\text{SRB}}\left(a_{1,\text{SRB}}^A + \frac{1}{2}\right)\right)\text{H}_2\text{SO}_4 \\
\rightarrow \varepsilon_{\text{SRB}}\mathcal{S}_{\text{SRB}} + (1 - \varepsilon_{\text{SRB}})\text{H}_2\text{CO}_3 + \left(\frac{1}{2} + \varepsilon_{\text{SRB}}\left(a_{1,\text{SRB}}^A + \frac{1}{2}\right)\right)\text{H}_2\text{S(aq)} + \varepsilon_{\text{SRB}}b_{1,\text{SRB}}^A\text{H}_2\text{O}.
\]  

(3)

This is a provisional file, not the final typeset article
The reaction efficiency variable, $\varepsilon_{SRB}$, is one of two classes of optimal control variables and a central design feature of the MEP model. As $\varepsilon_{SRB}$ approaches 1, Eq. (3) represents 100% conversion of labile carbon plus N and P resources to catalyst, while as $\varepsilon_{SRB}$ approaches 0, the reaction changes to 100% anaerobic combustion of labile carbon. From an entropy production perspective, only the catabolic reaction dissipates significant free energy, so $\varepsilon_{SRB}$ should be set to zero to maximize entropy production; however, the catabolic reaction cannot proceed without catalyst. There exists, then, an optimum value of $\varepsilon_{SRB}$ that produces just enough $\mathcal{S}_{SRB}$ catalyst to dissipate the chemical potential between $C_L$ and $H_2SO_4$, but $\varepsilon_{SRB}$ must change as a function of $C_L$ and $H_2SO_4$ supply rates as well as N and P availability. Conceptually, the MEP problem is to determine how $\varepsilon_{SRB}$ should change over time and space to maximize free energy dissipation; however, there are a few other important details to fill in.

Since we have discussed MEP model development in previous work (Vallino, 2010; Vallino, 2011; Algar and Vallino, 2014; Vallino et al., 2014) and have provided the model equations in the Supplementary Material, we only briefly discuss the main ideas followed by modifications and new additions for phototrophs. As before, we use Alberty’s (2003) approach to calculate Gibbs free energy of reaction, $\Delta_r G_{i,j}$, as a function of $\varepsilon_j$, temperature, $T$, and pH, which accounts for dissociation of chemical species (i.e., $H_2CO_3$, $HCO_3^-$, $CO_3^{2-}$, etc), and ionic strength is used to approximate activities from reactant and product concentrations. For non-phototrophic reactions (i.e., chemotrophs), an adaptive Monod equation parameterized by $\varepsilon_j$ accounts for the tradeoffs between substrate affinity, maximum specific growth rate and growth efficiency and can approximate oligotrophic to copiotrophic growth kinetics (such as SAR11 growing in the Sargasso Sea to $E. coli$ growing on LB medium in the lab) as $\varepsilon_j$ varies between 0 and 1 (Algar and Vallino, 2014; Vallino et al., 2014). In addition to the kinetic constraints, $F_K$, reaction rates, $r_{i,j}$, are also constrained by reaction thermodynamics, $F_T$, as described by La Rowe et al. (2012) (see Eq. S24). Consequently, the rates of chemical reactions (see Section 2.7.6 of SM) take the following general form,

$$r_{i,j} = \nu^* \varepsilon_j^2 \Omega_{i,j} [\mathcal{S}_j] F_K(c, \varepsilon_j; \kappa^*) F_T(\Delta_r G_{i,j}, \chi_{i,j}).$$

where vector $c$ is the concentration of substrates and products, $\chi_{i,j}$ is the number of electrons transferred in the redox reaction, $\nu^*$ and $\kappa^*$ are universal constants (we have used the same values of 350 d$^{-1}$ for $\nu^*$ and 5000 mmol m$^{-3}$ for $\kappa^*$ for all models to date) and $\Omega_{i,j}$ is the second optimal control variable. Since a catalyst, $\mathcal{S}_j$, can catalyze $\nu_{i,j}$ sub-reactions, $\Omega_{i,j}$ determines how much of catalyst $\mathcal{S}_j$ is allocated to reaction $\nu_{i,j}$ at a given time and location in space. For instance, the sulfate reducing bacteria are also given two reactions in addition to Eq. (3) (Table 1). Reactions $\nu_{2,SRB}$ and $\nu_{3,SRB}$ allow SRB to decompose recalcitrant organic carbon, $C_D$, and phosphorous, $P_D$, into labile organic carbon, $C_L$, and inorganic phosphate, respectively, but to do so they must deallocate from biosynthesis reaction, $\nu_{1,SRB}$, Eq. (3). That is, each $\Omega_{i,j}$ is bound between 0 and 1, and $\Omega_{i,j}$ must sum to unity over $t$ for each $\mathcal{S}_j$. An example of how $F_K$, $\Delta_r G_{i,j}$, and $F_T$ vary over time and space to influence reaction rate is given in SM Section 4.1, Fig. S3 for growth of sulfate reducing bacteria, $\nu_{1,SRB}$.

Entropy production per unit volume for chemotrophic reactions, $\dot{\sigma}_{i,j}$, is given by,

$$\dot{\sigma}_{i,j} = -\frac{1}{T} \Delta_r G_{i,j} r_{i,j}.$$
In the formulation above, the free energy of reaction, $\Delta_r G_{i,j}$, for the pure anabolic reaction, Eq. (1), might be positive (but it can be negative, see Vallino, 2010). To improve computational consistency between anabolic reactions with differing free energies, the reaction stoichiometries are augmented to insure that $\Delta_r G_{i,j}$ goes to zero as $\varepsilon_i$ goes to 1 for the anabolic component of the reaction. For an anabolic reaction that has a strongly positive $\Delta_r G_{i,j}$, the reaction is coupled to the catabolic reaction based on the ratio of the anabolic to catabolic reaction free energies, defined as $n_{i,j}$ in the stoichiometric equations. Consequently the true stoichiometry for $r_{1,SRB}$, as well as other reactions in the network, is slightly more cluttered (e.g., Eq. S103), but still readily calculated.

### 2.1.1.1 Phototrophy

In this version of the MEP model we introduce catalysts associated with phototrophic growth, specifically phytoplankton (Phy), $\mathcal{S}_{Phy}$, anaerobic green sulfur bacteria (GSB), $\mathcal{S}_{GSB}$, and photoheterotrophs (PH), $\mathcal{S}_{PH}$. Both Phy and GSB are modeled similarly using two sub-reactions (Table 1). One sub-reaction couples photon capture that drives CO$_2$ fixation into carbohydrates, C$_{Phy}$ and C$_{GSB}$, with elemental composition CH$_2$O, and a second sub-reaction converts carbohydrates into biomass in a manner analogous to growth of heterotrophic bacteria and GSB previously developed.

We focus on phytoplankton here (but see SM Section 2.7.2) because development for GSB is similar, and can be found in SM Section 2.7.3.

The carbon fixation reaction for Phy is given by,

\[
\varepsilon_{Phy} H_2CO_3 + n_{1,Phy}h\nu \rightarrow \varepsilon_{Phy}(C_{Phy} + O_2(\text{aq}))
\]  

(6)

where $h\nu$ are captured photons of frequency $\nu$, $h$ is Plank’s constant and $n_{1,Phy}$ is the mmoles of photons needed to fix 1 mmol of CO$_2$ at 100% efficiency (i.e., $\varepsilon_{Phy} = 1$) (note, units for reaction rates in the model are in mmol m$^{-3}$ d$^{-1}$, so we use mmol instead of mol in this description). The Gibbs free energy for a mmol of photons is given by

\[
\Delta_r G_{\gamma} = -\eta_{I} h\nu N_A,
\]  

(7)

where $N_A$ is Avogadro’s number (in mmol) and $\eta_I$ is the thermodynamic efficiency for converting electromagnetic radiation into work (see Eq. (S26) and Candau, 2003). If we define $\Delta_r G_{C_{Phy}}$ as the Gibbs free energy for fixing 1 mmol of H$_2$CO$_3$ into C$_{Phy}$ plus O$_2$ (see Eq. (S29)), then

\[
n_{1,Phy} = -\frac{\Delta_r G_{C_{Phy}}}{\Delta_r G_{\gamma}},
\]  

(8)

and the Gibbs free energy of reaction for Eq. (6) is given by,

\[
\Delta_r G_{1,Phy} = -(1 - \varepsilon_{Phy})\Delta_r G_{C_{Phy}}.
\]  

(9)

In this formulation, $\varepsilon_{Phy}$ governs the efficiency for the conversation of electromagnetic potential into chemical potential. If 100% of solar radiation is converted to chemical potential, no entropy is produced, and the free energy of reaction for Eq. (6) is zero, so the reaction will not proceed due to thermodynamic constraints. However, as $\varepsilon_{Phy}$ is reduced below 1, some photons in the photosynthetic active radiation (PAR) spectrum are simply dissipated as heat, and all are dissipated...
Energy dissipation drives microbial function

as heat when $\varepsilon_{phy} = 0$. In Eq. (6), the classic quantum yield, $\varphi$, (Kishino et al., 1986) is given by

$$\varphi_y = \varepsilon_{phy}/n_{1,phy},$$

which obviously varies as a function of the optimal control variable, $\varepsilon_{phy}$.

The reaction rate for Eq. (6) depends on the rate of photon capture, which is given by,

$$\Delta I_{phy} = k_p \Omega_{1,phy}[\mathbb{S}_{phy}]I(t, z)$$  \hspace{1cm} (10)

where $k_p$ is the coefficient for light absorption by particles, $\Omega_{1,phy}[\mathbb{S}_{phy}]$ is the fraction of phytoplankton catalytic machinery allocated to capturing photons (i.e., chlorophyll, other pigments, electron transfer chains, etc), and $I(t, z)$ is the light intensity (mmol photons m$^{-2}$ d$^{-1}$) at time $t$ and depth $z$ (see SM Section 2.5). Consequently, the reaction rate for Eq. (6) is given by,

$$r_{1,phy} = \frac{\Delta I_{phy}}{n_{1,phy}} F_K(c, \varepsilon_j; \kappa^*, \kappa^y) F_T(\Delta_T G_{1,phy}, n_{1,phy}^e)$$  \hspace{1cm} (11)

This reaction has similarities to those typically used to describe phytoplankton growth (Macedo and Duarte, 2006), but our derivation focuses not on the local light intensity level, but rather on how much light is actually intercept by the phytoplankton, as governed by $k_p \Omega_{1,phy}[\mathbb{S}_{phy}]$, and how much of that free energy is actually used to drive carbon fixation, as governed by $\varepsilon_{phy}$. As evident in Eq. (11), the maximum rate is directly tied to the rate of photon interception, $\Delta I_{phy}$, not by an arbitrary maximum specific growth parameter. The second reaction used by Phy and GSB (Table 1) is simply the conversion of reduced organic carbon, $C_{phy}$ and $C_{GSB}$, into more catalyst, but of course $C_{phy}$ and $C_{GSB}$ can also just be combusted to produce entropy depending of the value of $\varepsilon_{phy}$. The carbohydrate pools serve a similar function as cell quota in Droop’s (1973) model.

The reaction for phototrophs (PH) differs slightly from that above (SM Section 2.7.8). In this case only one reaction is used ($r_{1,PH}$, Table 1), where the photon capture is linked to the conversion of labile carbon into PH catalyst. As above, photons captured can also be dissipated as heat for $\varepsilon_{PH} < 1$, or the free energy can be used to drive biosynthesis ($\varepsilon_{PH} > 0$), and photon energy replaces catabolic reaction used in chemotroph reactions. In the currently formulation, reaction rate is also tied to photon capture rate, so growth can only proceed during the day (Eq. (S131)). A second chemotrophic sub-reaction (like $r_{1,Bac}$, Eq. (S93)) could have been added to $\mathbb{S}_{PH}$, but this would increase the number of optimal control variables in the problem, and would be redundant with $\mathbb{S}_{Bac}$, so this option was not explored.

### 2.1.2 Entropy production

It is outside the scope of this manuscript to delve into the thickets of entropy (see Vallino and Algar, 2016), but a few definitions are in order, since entropy accounting is the foundation of the MEP approach. Simply stated, entropy production occurs when an energy potential is destroyed and dissipated as heat to the environment, but not when the potential is converted from one potential to another, if done so reversibly, which technically is possible, but never occurs in practice. For example, a flame converting chemical potential into heat or light being absorbed by water both result in maximum entropy production; these are irreversible processes. On the other hand, entropy is not produced by converting electromagnetic potential into chemical potential, which phytoplankton do partially, but they also convert a fraction of that potential to heat, as dictated by $\varepsilon_{phy}$ in the model. As $\varepsilon_{phy}$ approaches 1, electromagnetic potential is converted to chemical potential without loss, but $F_T$
(Eq. S24) drives the reaction rate to 0. For organisms to grow at a non-zero rate an energy potential must be partially dissipated and some entropy must be produced.

For the Siders Pond model, entropy production associated electromagnetic radiation can be produced along three pathways. Photons intercepted by water, \( k_w I(t, z) \), and particles lacking photochemical properties, such as bacteria, \( k_p [\mathcal{S}_{ac}] I(t, z) \) and grazers, \( k_p [\mathcal{S}_{g2}] I(t, z) \), as well as by the non-photosynthetic components of phytoplankton, \( k_p \Omega_{2,phy} [\mathcal{S}_{phy}] I(t, z) \) and \( k_p \Omega_{2,GSB} [\mathcal{S}_{GSB}] I(t, z) \), are simply dissipated as heat (long wave radiation), and the entropy produced from them is readily calculated. Photons intercepted by the photosynthetic machinery of phototrophs can be either dissipated as heat, \( \varepsilon_j = 0 \), or the energy potential can be transferred to chemical potential, \( \varepsilon_j = 1 \), but in general it is a combination of those processes, \( 0 < \varepsilon_j < 1 \). This latter route is considered entropy production by reaction in the Results Section, even though it involves photons. Consequently, total entropy production, \( \dot{\sigma}_2 \) (Eq. (S178) first expression) is the sum of three processes: entropy production by photon absorption by water, \( \dot{\sigma}_W \) (Eq. (S176)), entropy production by light absorption by particles, \( \dot{\sigma}_P \) (Eq. (S177)) and entropy production by reaction, \( \dot{\sigma}_R \) (Eq. (S178) second expression). In the photic zone of Sider Pond, it is the production of biological particles through growth that greatly increases local entropy production.

2.1.3 Transport model

Siders Pond is horizontally well mixed, so an advection-dispersion-reaction (ADR) model that includes particle sinking was used to approximate vertical transport of the 19 state variables (Fig. 1).

The origin of the vertical coordinate, \( z \), is defined at the pond’s surface, and the axis points in the positive direction downward towards the benthos and reaches a maximum depth of 15 m. Siders Pond 3D bathymetry surface (Fig. S1) was rendered from a contour plot in Caraco (1986), and an equation for cross-section area as a function of depth, \( A(z) \), was derived therefrom (Eq. S3).

Equations for vertical volumetric flow rate, \( q(z) \), lateral groundwater inputs, \( q_L(z) \) (Eq. (S5)), and seawater intrusion at the bottom boundary were determined from Caraco (1986), who used both tritium-helium-3 dating combined with mass balance calculations to estimate freshwater inputs and seawater intrusion. An equation for the dispersion coefficient as a function of depth, \( D(z) \), (Eq. (S6) and Fig. (S2)), was derived by fitting model predicted salinity profiles to observations collected during this study. Salinity profiles predicted by the ADR model show good agreement to observations (Fig. S2).

The primary external drivers in the model are temperature, \( T(z) \), pH, \( p_H(z) \), and photosynthetic active radiation (PAR), \( I(t, z) \). Since the ADR model does not include either energy nor proton balances, equations for both \( T(z) \) and \( p_H(z) \) were simply based on linear interpolations of observations obtained from the water quality probe used during sampling. Surface irradiance, \( I(t, 0) \), was based on a model of solar zenith angle (Brock, 1981), which assumes a cloudless sky. To predict PAR light intensity as function of time and depth, a standard light adsorption model was used that includes coefficients for light absorption by water, \( k_w \), and particulate material, \( k_p \) (Eq. S7).

To complete the ADR transport model, Neumann boundary conditions were used for state variables at the pond’s surface, but particles were not allowed to sink into the boundary from above (Eq. S9), and gas transport for \( O_2, CO_2 \) and \( H_2S \) across the air-water interface was accounted for using a stagnant film model (Eq. S10). Robin boundary conditions were used at the bottom boundary based on the flux of material entering the boundary associated with the intrusion of seawater diluted with groundwater from below. However, aerobic and anaerobic decomposition of sinking organic matter...
from the water column contributed to a sink for O$_2$ and H$_2$SO$_4$, and a source for H$_3$PO$_4$, H$_2$CO$_3$ and H$_2$S to the overlying water (Eq. S23).

2.1.3.1 Parameter adjustments

One of the objectives in using a variational approach is to remove model parameters, but a few required adjustments to get simulations to within an order of magnitude of Siders Pond observations. Our objective was not to minimize error between predictions and observation, so the few parameters described in this section were adjusted in a rather ad hoc manner. No formal data assimilation was attempted, and the parameters were adjusted using fixed nominal values for the optimal control parameters, $\varepsilon_j$ and $\Omega_{ij}$ (i.e., simulations without optimization). All biological structures, $S_j$, detritus, $C_D$, and $P_D$, and associated phytoplankton carbohydrates, $C_{phy}$ and $C_{GSB}$, were allowed to sink through the water column (Eq. (S1)). However, sinking velocities can vary greatly across and within organism size classes (Bach et al., 2012), so velocities where set through trial and error to obtain profiles consistent with those observed in Siders Pond (Table S1). In addition, coefficients for light absorption by water, $k_w$, and particles, $k_p$, were adjusted so that light profiles were similar to observations (SM Section 2.5). Computational approach used for solving the ADR equation is described in SM Section 2.9.

2.1.4 Optimization and MEP

The stoichiometry, thermodynamics and kinetics of the 28 reactions that comprise the metabolic network (Table 1, Fig. 1, SM Section 2.7) vary as a function of the $\varepsilon_j$ optimal control variables, and the partitioning of biological structure, $S_j$, to sub-reactions $r_{ij}$ (Table 1) that depends on the values of the $\Omega_{ij}$ control variables. A solution to the MEP model, and associated microbial biogeochemistry, is determined by adjusting $\varepsilon_j$ and $\Omega_{ij}$ over time and space to maximize entropy production, the details of which are provided in SM Section 3.2, but also see Vallino et al. (2014). The critical aspect of the optimization is choosing the appropriate time interval over which to maximize entropy production and whether local or global optimization should be used (Vallino, 2011); consequently, these two aspects are the focus of this manuscript and form the bases of the Results Section. Computational approach used for solving the optimal control problem is described in SM Section 3.2.

2.2 Siders Pond site description

Siders Pond is a small coastal meromictic kettle hole (volume: 10$^6$ m$^3$; area: 13.4 ha; maximum depth: 15 m) that receives approximately 1x10$^6$ m$^3$ of fresh and 0.15 x10$^6$ m$^3$ of saltwater each year (Caraco, 1986). The latter input occurs via episodic inputs during high tides and storm events via a small creek that connects the pond to Vineyard Sound approximately 550 m to the south (Fig. S1). Tritium-helium water dating confirmed vertical mixing across two observed chemoclines, but permanent stratification is maintained because the saltwater inputs enter the pond at depth, mix upward and become entrained with freshwater before exiting the pond (Caraco, 1986). Caraco (1986) also characterized N and P loading to the pond (50 g N m$^{-2}$ y$^{-1}$ and 1.3 g P m$^{-2}$ y$^{-1}$, respectively), and an N+P enrichment study (Caraco et al., 1987) shows phytoplankton to be P limited, especially in the low salinity surface waters. Previous studies show Siders Pond is eutrophic averaging 16 mg m$^{-3}$ chlorophyll a (Chl a) in surface waters (but can exceed 100 mg m$^{-3}$ at times) and an annual primary productivity of 315 g C m$^{-2}$ (Caraco, 1986; Caraco and Puccoon, 1986). In anoxic bottom waters bacterial Chl c, d and e associated with photosynthetic green sulfur bacteria averages 20 mg m$^{-3}$ (purple sulfur bacteria were not found in high concentration), but BChl cde was also observed to reach high concentrations at times (> 75 mg m$^{-3}$). Even though green sulfur bacteria
could attain high concentrations, their productivity was only 6% of the oxygenic photoautotroph (cyanobacteria + algae) production (Caraco, 1986). Siders Pond was chosen for this study because extensive redox cycling occurs over a 15 m deep water column, which greatly facilitates sampling due to the large water volumes that can be readily collected without perturbing the system.

2.3 Sampling and measurements

Samples were collected from Siders Pond, Falmouth, MA over a 24 h period starting at 6:45 on Jun 25th and ending at 7:37 on Jun 26th, 2015 from a single station located within the deepest basin of the pond (41.548212ºN, 70.622412ºW). A total of 7 casts were conducted over the 24 hr period, and each cast sampled 8 depths to generate a 2D sampling grid designed for comparison to model outputs (Fig. 2). A Hydrolab DS5 water quality sonde (OTT Hydromet, GmbH) was connected to a Hydrolab Surveyor 4 handheld display and used to record depth, temperature, salinity, pH, dissolved oxygen (DO), photosynthetic active radiation (PAR) and in situ Chl a fluorescence. All sensors were calibrated per manufacturer’s instructions one day prior to sampling. One end of a 20 m long section of vinyl tubing with a 1 cm inside diameter was attached to the water quality sonde, while the other end passed through a Geopump 2 (Geotech, CO) peristaltic pump and then connected to 25 mm polypropylene, acid washed in-line Swinnex filter holder (Millipore, MA), which housed a GF/F glass fiber filter (Whatman, GE Healthcare) that had been ashed at 450ºC for 1 hr. At all depths, the vinyl tubing was first flushed for at least 2 min, sample collection vials were washed twice with filtrate and GF/F filters where changed as needed to maintain high flow. This design allowed water samples to be collected at the desired depths and processed on location then preserved on ice or dry ice for later analysis as described below.

Unless otherwise noted, all samples were filter as described above and stored in 20 mL high-density polyethylene scintillation vials (Fisher Scientific). Samples were preserved for later analysis as follows. Inorganic phosphate: 15 mL samples were amended with 20 µL of 5 N HCl and plated on dry ice. Dissolved inorganic carbon (DIC) and sulfate: samples were collected in 12 mL exetainers (Labco, UK) by overfilling bottles from the bottom up, and then capped without bubbles and placed on ice. Hydrogen sulfide: while filling exetainers, 25 µL of sample was pipette transferred to 6 mL of 2% zinc acetate in a 20 mL scintillation vial and place on ice. Dissolved organic carbon (DOC): 25 mL of sample was collected in previously ashed 30 mL glass vials to which 40 µL of 5 N HCl was added then stored on ice. Particulate organic carbon (POC): approximately 300 mL of sample was passed through new, ashed, 25 mm GF/F filter then stored in a plastic Petri dish on dry ice.

Samples were analyzed at the Ecosystems laboratory, MBL as follows. Phosphate: samples were stored at -20ºC then later analyzed following the spectrophotometric method of Murphy and Riley (1962) on a UV-1800 spectrophotometer (Shimadzu, Kyoto, Japan). DIC: samples were immediately run on to MBL on an Apollo AS-C3 DIC analyzer (Apollo SciTech, DE). Sulfate: samples were sparged with N2 to strip H2S on return to MBL and stored at 4ºC then analyzed using ion chromatography on a Dionex DX-120 analyzer (Dionex, Sunnyvale, CA). Hydrogen sulfide: samples were briefly stored at 4ºC for five days then analyzed using the spectrophotometric method of Gilboa-Garber (1971). DOC: samples were stored at 4ºC then run on a Shimadzu TOC-L high temperature total organic carbon analyzer at 720ºC. POC: samples were stored at -20ºC then analyzed on a Thermo Scientific FLASH 2000 CHN analyzer using aspartic acid standards. To facilitate model comparison to observations, Chl a was estimated from modeled output of phototroph biomass concentrations using 12. \( \frac{[\text{Scphy}] + [\text{SGSB}] + [\text{SPH}]}{\theta_{\text{C:Chl}a}} \), where \( \theta_{\text{C:Chl}a} \) is the C:Chl a ratio, which was set to 50 µg C (µg Chl a)-1 (Sathyendranath et al., 2009).
3 Results

During initial model development, a simple study was conducted to compare local versus global MEP optimization (Vallino, 2011) for the Siders Pond model; however, it was quickly realized that global optimization did not produce relevant solutions (see Discussion Section for explanation).

Consequently, all solutions presented here are from local entropy maximization at 10 depths (0, 1, 2, 3, 4, 6, 8, 10, 12, and 15 m) conducted in parallel. We investigated two optimization time intervals; Δt, for entropy maximization, over a short interval of 0.25 days as well as over a long interval of 3 days. To minimize the influence of uncertain initial conditions on final solutions, both the short and long interval optimizations were started on May 19th and ran with short or long interval optimization until Jun 21st. Solutions from Jun 21st were then used as initial conditions for 6 day production simulations that required 24 intervals for the short 0.25 d optimization and 2 intervals for the long 3 d optimization. The 6 day simulations ended on Jun 27th to allow comparison with observations collected on Jun 25th and 26th (Fig. 2). Our results focus on how these two solutions compare to observations as well as to each other. Future discounting was not used in solutions presented here (i.e., \( k_D = 0 \) and \( \delta t = \Delta t \)). See SM Section 3.2 for details on optimization.

3.1 Simulations compared to observations

In this section, solutions obtained from the short (0.25 d) and long (3 d) interval optimizations are compared to biogeochemical observations from Siders Pond that were collected on Jun 25th and 26th, 2015 (Fig. 2). Simulated profiles for photosynthetic active radiation (PAR), Chl a and dissolved oxygen over 6 days for the short and long interval optimizations are compared to 1 day of observations in Siders Pond in Figure 3. The short interval optimization (SIO) shows PAR extending to nearly the bottom of the pond (Fig. 3a), while PAR from the long interval optimization (LIO) (Fig. 3b) more closely matches observations (Fig. 3c); however, high intensity PAR (~ 100 \( \mu \text{mol m}^{-2} \text{s}^{-1} \)) observations extend a few meters deeper than the LIO simulation predicts. The prediction for Chl a in both simulations do not match observations very well (Fig. 3d-f), but this is partially due to how Chl a was estimated in the model, since Chl a is not specifically modeled (see Section 2.3). The Chl a \textit{in vivo} observations show a peak Chl a around 5 m, while both simulations have peaks around 12 m, but those peaks are due to accumulation of sinking phytoplankton rather than productivity (see below). The LIO simulation does show a secondary Chl a peak developing around 3 m, but it is weaker than observations (~ 6 vs 40 \( \mu \text{g L}^{-1} \)). Based on DO, the SIO shows anaerobic conditions begin at 6 m, while the LIO shows that occurring at 10 m. Observations for the transition to anoxia splits between the two simulations at 8 m. Observations also show a subsurface DO maximum at 3.5 m, while both simulations show max DO closer to 1.3 m. Furthermore, the SIO shows a decreasing DO max with time, while the LTO shows an increase over time, and the maximum reaches 800 mmol m\(^{-3}\) versus 480 for observations. In general, the SIO simulations show less phototrophic activity while LIO shows greater activity than what the observations indicate.

Simulations of substrate concentrations for autotrophs, namely inorganic phosphate and dissolved inorganic carbon (DIC), show overall agreement to observations for both SIO and LIO solutions (Fig. 4), but some discrepancies are apparent. The phosphate chemoclines occur around 10 m, 11 m and 8 m for the SIO, LIO and observations, respectively (Fig. 4a-c). Above the phosphate chemocline, the SIO simulation shows slightly elevated levels of \( \text{H}_3\text{PO}_4 \) (~0.3 mmol m\(^{-3}\)) in the surface and 5 m layers, compared to observations (Fig. 4c), which are near the level of detection (0.03 mmol m\(^{-3}\)) except for a few spikes. Below the phosphate chemocline, observations show slightly higher accumulations of phosphate, approaching 22 mmol m\(^{-3}\), while the SIO and LIO simulations show max concentrations closer to 16 and 19 mmol m\(^{-3}\), respectively. For DIC, both SIO and LIO...
Simulated sulfate concentrations in the upper portion of the water column (0 to 4 m) are similar to those observed (Fig. 5a-c), but the simulations show a very dramatic sulfate chemocline starting at about 8 m, while observations show a more gradual increase in sulfate with depth. Furthermore, sulfate reaches much higher concentrations in the simulations at depth than do observations, showing maximums of 15,500 mmol m\(^{-3}\) in both simulations, while observation maximum is only 9,000 mmol m\(^{-3}\). The lower simulated concentrations of \(\text{H}_2\text{S}\) and the higher simulated sulfate concentrations indicate that sulfate reduction in the model is lower than that actually occurring in Siders Pond.

Simulations of hydrogen sulfide show a chemocline at approximately 9 and 10 m for the SIO and LIO solutions, respectively, which are comparable to the \(\text{H}_2\text{S}\) chemocline observed in Siders Pond at about 8 m (Fig. 5a-c). However, \(\text{H}_2\text{S}\) reaches concentrations as high as 7,000 mmol m\(^{-3}\) in Siders Pond, while maximum concentrations only reaches 900 and 2,100 mmol m\(^{-3}\) in the SIO and LIO simulations, respectively. Simulations also show a peak \(\text{H}_2\text{S}\) concentration at 12 m, and a decrease in concentration below 12 m, which seems unlikely and indicates that the bottom boundary condition may not be correct. That is, the flux of \(\text{H}_2\text{S}\) into the bottom water is lower than it should be, but it is also possible that the rate of anaerobic metabolism in the simulations is too low, which would be consistent with the low simulated DIC concentration compare to observations mentioned above.

Simulated sulfate concentrations in the upper portion of the water column (0 to 4 m) are similar to those observed (Fig. 5d-e), but the simulations show a very dramatic sulfate chemocline starting at about 8 m, while observations show a more gradual increase in sulfate with depth. Furthermore, sulfate reaches much higher concentrations in the simulations at depth than do observations, showing maximums of 15,500 mmol m\(^{-3}\) in both simulations, while observation maximum is only 9,000 mmol m\(^{-3}\). The lower simulated concentrations of \(\text{H}_2\text{S}\) and the higher simulated sulfate concentrations indicate that sulfate reduction in the model is lower than that actually occurring in Siders Pond.

The last of the observations are dissolved organic carbon (DOC) and particulate organic carbon (POC) concentrations (Fig. 6). For simulations, DOC is a derived quantity based on the sum of state variables \([C_g]\) and \([C_D]\), while POC is derived from the sum of \([C_{Phy}], [C_{GSB}]\) and the concentrations of all biological structures, \([\Phi]\). In the water column above 12 m, the SIO simulation shows very low concentrations of DOC (~ 1 mmol m\(^{-3}\)), but then increases rapidly to a maximum of 1,900 mmol m\(^{-3}\) (Fig. 6a). The LIO simulation shows a similarly high DOC concentration at 14 m, but above 12 m, the DOC concentration ranges from 2 to 140 mmol m\(^{-3}\), which is comparable to those observed in Siders Pond, which range from 200 to 300 mmol m\(^{-3}\) above 10 m, and increase to a maximum of about 1,000 mmol m\(^{-3}\) at 12 m. Similar to DOC, POC in the SIO simulation shows low values (< 30 mmol m\(^{-3}\)) above 6 m, but POC peaks to 1,000 mmol m\(^{-3}\) at 12 m (Fig. 6d). The POC concentrations from the SIO simulation are closer to observations, but the mid-water POC maximum in the LIO simulation is approximately 500 mmol m\(^{-3}\), while the observations peak at 280 mmol m\(^{-3}\) around 4 m. Like the SIO simulation, the LIO simulation also shows high POC concentrations below 10 m, which was not observed in Siders Pond. It is possible that DOC and POC may reach higher concentrations in the funnel-like basin of Siders Pond below 12 m (Fig. S1), but samples were not collected there.

### 3.2 Comparison between SIO and LIO simulations

Since the MEP optimization model generates a large number of outputs, this section highlights some of those outputs to contrast the simulations based on the short (0.25 d) interval optimization (SIO) to that from the long (3 d) interval optimization (LIO). Consider entropy production (Eq. S178), which is the variable that is being sequentially maximized over either a 0.25 d interval (SIO) or a 3 d interval (LIO) at 10 different depths (see SM Section 3.1, Eq. (S180)) (Fig. 7). Total entropy production, \(\dot{\sigma}_2\), for the SIO and LIO simulations differ significantly, in that \(\dot{\sigma}_2\) in the SIO solution is...
spread out over a 12 m water column (Fig. 7a) versus the LIO solution (Fig. 7e) where most of the entropy production occurs in the top 3 m of the pond. Furthermore, peak entropy production in the LIO simulation is 7.5 times great than the SIO solution (9.8 vs 1.3 GJ m⁻¹ K⁻¹ d⁻¹), and the total integrated entropy produced over the water column and 6 d simulation, \( \sigma_T \) (Eq. (S182)), was 21.3 GJ K⁻¹ and 36.5 GJ K⁻¹ for the SIO and LIO simulations, respectively. For comparison, if the pond was sterile, \( \sigma_T \) would equal 12.4 GJ K⁻¹ from light absorption in the water column. The higher total integrated entropy, \( \sigma_T \), produced by the LIO simulation illustrates that extending the time scale results in greater entropy production, as has been previously shown in experimental mesocosms (Vallino et al., 2014).

The different contributors to total entropy production, namely by reaction, \( \dot{\sigma}_R \), water, \( \dot{\sigma}_W \), and particles, \( \dot{\sigma}_P \), also differ significantly between the two simulations. For instance, total integrated entropy production associated with reactions (Table 1) was actually greater in SIO than the LIO simulation (Fig. 7b vs f), as well as exhibited a greater maximum \( \dot{\sigma}_P \) (0.28 vs 0.26 GJ m⁻¹ K⁻¹ d⁻¹ for SIO versus LIO); however, entropy production by reaction during the day is rather small (< 25% in the upper 5 m) compared to light dissipation by water or particles, but the two simulations differ here as well. The SIO simulation dissipates most of the incoming radiation by water absorption in the upper 5 m of the water column (Fig. 7c vs g), while the LIO simulation dissipates most of the electromagnetic potential via absorption by particles (Fig. 7d vs h). As evident in the POC (Fig. 6) and Chl a (Fig. 3) concentrations, the LIO simulation produces more biomass in the upper portion of the water column, and biomass is effective at absorbing and dissipating light. The high biomass in the LIO simulation is due to phytoplankton, as describe next.

An analysis of phytoplankton growth by the SIO and LIO simulations (Fig. 8) illustrates not only how the two simulations differ, but also some of the mechanics of the MEP-based optimization approach. Phytoplankton density attains a maximum of 20 mmol m⁻³ in the LIO simulation by June 27th, but Phy are effectively absent in the SIO simulation, attaining a maximum of only 0.1 mmol m⁻³ (Fig. 8a and f). While it is possible that the low phytoplankton (Phy) density in the SIO simulation could be due to extensive predation, this is not the case because the rates of CO₂ fixation (\( r_{1,Phy} \), Table 1) and conversion of fixed C to biomass (\( r_{2,Phy} \), Table 1) are two orders of magnitude lower in the SIO versus LIO simulation (Fig. 8b, c versus g, h). Of course the differences in phytoplankton density and reaction rates between the SIO and LIO simulations are due to how the optimal control variables change over time and space (Fig. 8d, e, i and j). Consider how reaction efficiency, \( \varepsilon_{Phy}(t, z) \), varies over time and space in the two simulations (Fig. 8d and i). The SIO simulation shows rapid switching between very high efficiencies (>0.98) and very low efficiencies (<0.02) over time. While not on all days, reaction efficiency drops to very low levels around noon (Fig. 8d), which results in high entropy production, but at the sacrifice of fixing CO₂, which is consistent with a short term optimization objective: burn fast now instead of investing for the future. On the contrary, the LIO solution (Fig. 8i) shows much more gradual changes in \( \varepsilon_{Phy}(t, z) \), operating between 0.3 and 0.4 for most of the simulation. There is rapid changing of \( \Omega_{1,Phy}(t, z) \) in the LIO solution (Fig. 8j), but this makes sense, because the control variable partitions phytoplankton biomass to the light requiring carbon fixation reaction, \( r_{1,Phy} \), during peak daylight, then switches to the biosynthesis reaction, \( r_{2,Phy} \), at night (Fig. 8h). Changes in biomass partitioning is also observed in the SIO solution (Fig. 8e), but sometimes \( \Omega_{1,Phy} \) selects for the light-requiring CO₂ fixation reaction at night. This nonsensical behavior is likely due to the nature of optimization. For the SIO simulation, phytoplankton do not contribute significantly to free energy dissipation, so values of the optimal control variable are freer to randomly fluctuate because they do not strongly impact the optimization objective. Based on Fig. 7c, the SIO solution places more weight on using water for the...
Microbes as energy dissipaters

Instead of growing phytoplankton, the SIO solution produces more green sulfur bacteria (GSB), which reach a maximum concentration of 735 mmol m$^{-3}$, compared to only 23 mmol m$^{-3}$ in the LIO solution (Fig. 9a and d). Furthermore, GSB increase during the SIO simulation, while they decrease in the LIO, which is evident in the greater biosynthesis reaction, $r_{2,GSB}$ in SIO versus LIO solutions (Fig. 9b and e) as well as in $r_{1,GSB}$ (not shown). However, the value of the reaction efficiency control variable, $\varepsilon_{GSB}(t, z)$, does switch to low values around noon on several days in the SIO simulation, which implies again the solution favors entropy production over growth (Fig. 9c and f). This SIO simulation also favors much higher reaction rates of photoheterotrophs (PH), $r_{1,PH}$, compared to the LIO solution, but interestingly, this does not lead to greater concentrations of $\dot{S}_{PH}$ (Fig. 10). The reason is because high rates for $r_{1,PH}$ are coupled with extremely low values of $\varepsilon_{PH}$ (Fig. 10b and c).

Based on the adaptive Monod equation that changes substrate affinity as a function of $\varepsilon_{PH}$ (Eq. (S131)), uptake of labile organic carbon by PH can occur at extremely low concentrations of $C_L$ when $\varepsilon_{PH}$ is close to zero, but small $\varepsilon_{PH}$ values mean very little biomass is produced as a result (Eq. (S123)). This is an interesting result, as light energy is being used to scavenge organic carbon at low concentrations, which contributes to the low labile organic carbon, $C_L$, observed in the SIO simulation (Fig. 6a). Furthermore, the high entropy production from reaction, $\dot{\sigma}_R$ (Fig. 7b), is almost entirely due to PH. One of the reasons both green sulfur bacteria and photoheterotrophs are more active in the SIO versus LIO simulations is that light is not being intercepted by phytoplankton like it is in the LIO simulation (Fig. 8a vs f).

Bacterial densities are similar in the two simulations (Fig. S4a and d), but there is significantly higher growth rate by bacteria in the LIO simulation between 4 and 10 m (Fig. S4b and e). Both simulations allocate almost all of $\dot{S}_{Bac}$ to detrital carbon decomposition, $r_{2,Bac}$, below 13 m (Fig. S5c and f), but the SIO simulation also allocates biomass to $r_{2,Bac}$ sporadically throughout the water column to produce $C_L$ (Fig. S4c), which is limiting (Fig. 6a). The anaerobic, sulfate reducing bacteria function similarly as bacteria, but of course operate in the anaerobic portion of the water column (Figs. S6 and S7). However, while their growth occurs predominately below 12 m (Fig. S6b and e), there is significant biomass accumulation at and above 11 m, which indicates advection is transporting them upwards. Like bacteria (Bac), the sulfate reducing bacteria (SRB) allocate biomass to $r_{2,SRB}$ below 13 m in both simulations (Fig. S7c and f) and sporadically throughout the water column in the SIO simulation (Fig. S7c). Overall, bacteria and SRB function in a complementary mode across the aerobic and anaerobic portions of the water column. Similar to phytoplankton, the control variables for both $\dot{S}_{Bac}$ and $\dot{S}_{SRB}$ show more rapid (bang-bang) control in the SIO compared to the LIO simulation (Figs. S5 and S7). The sulfide oxidizing bacteria, $\dot{S}_{SOx}$, are largely unimportant in either of the 6 day simulations.

Another significant difference between the SIO and LIO simulations is a greater importance of predation, $\dot{S}_{Gz}$, in the SIO solution (Fig. 11). Because predation is abstracted in the MEP model, it represents all predation mechanisms, including protists, predatory bacteria, viruses and cannibalism. In addition to dissipating chemical potential stored as biomass, predation serves a more important task of turning over biomass locked into metabolic functions, such as sulfate reduction, that are no longer of use under prevailing conditions. The SIO and LIO simulations show that the concentration of $\dot{S}_{Gz}$ is more than 4 times higher in SIO than LIO solutions, and $\dot{S}_{Gz}$ increases over time with the SIO objective. The primary prey items in the SIO solution are $\dot{S}_{Gz}$ (i.e., cannibalism), $\dot{S}_{Bac}$ and $\dot{S}_{PH}$ (Fig. 11b, c and d), while only $\dot{S}_{Bac}$ predation is of significance in the LIO solution (Fig. 11g).
Closer inspection of Fig. 11b-c reveals that predation occurs predominately at night, which is a result of temporal changes in the partitioning control variables, $\Omega_{GZ}$, rather than prey concentration (not shown). It is also worth noting that while the high concentration of $GZ$ is found below 10 m, predation actually occurs between 4 and 7 m, which indicates the high $GZ$ concentration is due to sinking and accumulation. Accumulation of biomass in the deep, anaerobic, portion of the water column becomes food for anaerobic predators, $AGZ$, which are important in both SIO and LIO simulations, but are slightly more active in the LIO simulation (Fig. S8). The dominate prey items for $AGZ$ are GSB (Fig. S8b and c), $AGZ$ (cannibalism) (Fig. S8c and f), SRB (Fig. S9a and c) and PH (Fig. S9b and d).

An interesting result that derives from the focus on energy dissipation rather than organisms is the importance of chemotrophs on dissipating electromagnetic potential. For instance, heterotrophic bacteria, $Bac$, are well understood as dissipaters of chemical potential as they typically respire more reduced organic carbon than converting it biomass (i.e., $\epsilon_{Bac} < 0.5$, Eq. (S84)), where the latter reaction conserves chemical potential (Table 1). In both the SIO and LIO simulations, however, far more free energy is dissipated by passive light absorption than it is by respiration (Fig. 12), although the difference is more striking in the SIO simulation (Fig. 12a versus b). Of course, some prokaryotes in nature harness the abundant light energy via expression of proteorhodopsin (Beja et al., 2000; Giovannoni, 2017).

4 Discussion

The two primary objectives of this modelling study were to demonstrate that 1) a model based on free energy dissipation reasonably describes microbial community organization and function and 2) that microbial systems operate collectively over characteristic timescales that are likely longer than what common wisdom would suggest. The secondary objectives were to demonstrate how the model can be used in systems with spatial dimensions and to extend the approach to include phototrophs.

While improvements could be made with explicit data assimilation (Edwards et al., 2015), the MEP model did a reasonable job at simulating biogeochemistry in Siders Pond, and the better fit of the long interval optimization (LIO) simulation to observations indicates that the microbial community has evolved to function over time scales that are longer than 0.25 days. While sinking velocities and the two light absorption coefficients were crudely tuned (Section 2.1.3.1), the model otherwise has very few adjustable parameters compared to conventional biogeochemical models that often contain dozens (Ward et al., 2010). Furthermore, the MEP model is expected be more robust to extrapolation beyond the envelope of calibration data, as the MEP objective is applicable regardless of a systems operating point. The main drawback to the MEP approach is in the prediction of details. Since the MEP conjecture is based on statistical mechanics (Dewar, 2003; Lorenz, 2003), it does not provide information on the mechanisms by which a system attains an MEP state. If fact, the reason systems tend to follow an MEP trajectory is because there are many ways to achieve that objective. The approach is unlikely to predict details correctly. For instance, MEP models are unlikely to distinguish an algal bloom from a harmful algal bloom, although predicting the latter is of great importance.

Likewise, an MEP model is just as likely to predict biomass turnover by jellyfish versus turnover by salmon, but from an economics perspective, these are two vastly different solutions. In essence, MEP is best used at predicting overall function, not the details that describe that function, largely because predicting details is very difficult and prediction accuracy decays rapidly. MEP modeling is similar to weather versus climate prediction. Climate models can predict far into the future, but not at the details of weather, and weather models provide details, but their prediction decays quickly.

Perhaps the most useful aspect is that MEP provides a different perspective to view biology (Skene,
Temporal strategies, such as circadian clocks (Wolf and Arkin, 2003), anticipatory control (Mitchell et al., 2009; Katz and Springer, 2016), energy and resource storage (Schulz et al., 1999; Grover, 2011), and dormancy (Lewis, 2010) are hallmarks of biology, yet they are often not given much consideration when theory and models are developed for understanding biogeochemistry, even though temporal strategies have also been observed in microbial communities (Ottesen et al., 2014). Here our results demonstrate that different organizational timescales dramatically impact biogeochemistry and how microbial communities function. For instance, the SIO solution does not invest resources in phytoplankton growth, because over the short 0.25 d optimization, water dissipates more electromagnetic potential than a small increase in phytoplankton biomass over the short interval. Instead, the SIO solution allocates resources to green sulfur bacteria (GSB) and photoheterotrophs (PH) growth, possibly because sinking losses near the surface of the pond are greater than they are at depth due to advection and the pond’s geometry. Near the surface of the pond, upward advective velocity that counteracts sinking is only 0.01 m d⁻¹, while at 12 m velocities reach 0.06 m d⁻¹, which facilitates biomass accumulation. Specific growth rates need to be higher at the surface. The SIO solution also places more resources on decomposing refractory carbon, but also on respiring the produced labile carbon, which results in low standing-stock concentrations of $C_L$. Grazing rates, especially under aerobic conditions, are higher in the SIO simulations as well. These types of resource allocations in the SIO solution are consistent with R* or resource-ratio theory (Tilman, 1982) and r-selection (Pianka, 1970; Fierer et al., 2007); that is, emphasis on fast growth. On the contrary, the LIO solution appears more similar to K-selection where resources are invested for longer term outcomes. In the SIO simulations, the control variables show rapid bang-bang control fluctuations as resources vary (Figs. 8d, S5a and S7a), while the LIO solution produces more gradual changes in control variables (Figs. 8i, S5d and S7d). Long-term strategies are more likely to produce cooperation that require investment of resources and more stable conditions (Hibbing et al., 2010), but utilize and dissipate resources more effectively (Cole et al., 2015). Strategies such as cross-feeding that make communities more resilient (Zelezniak et al., 2015) are also consistent with LIO solutions. A microbial community that implements temporal strategies should outcompete a community which lacks such temporal strategies. In essence, long-term strategies result in greater acquisition and utilization of food resources under non steady-state conditions than short-term strategies.

One of the main questions that arises in applying MEP is what is the appropriate timescale over which biology organizes? That is, in the current implementation of the model, what is an appropriate value for $\Delta t$ used in the optimal control problem (Eqs. (S183))? In this study two values were explored, 0.25 d for the SIO and 3 d for the LIO, but these were largely ad hoc choices. Our results indicate that the SIO solution does not match observations or expectations as well as the LIO solution. Comparing MEP model output to observations for differing values of $\Delta t$ is one means to explore this fundamental property of ecosystems. But is a 3 d optimization window sufficient? It seems the appropriate MEP time scales do not depend on just the characteristic time scales of organisms, because our study on periodically forced methanotroph communities showed that the communities were well adapted to 20 day cyclic inputs of energy, even though the characteristic turnover time of bacteria is closer to hours (Vallino et al., 2014; Fernandez-Gonzalez et al., 2016). We know that terrestrial systems function on long timescales, at least on the order of the four seasons, but their time scales are likely much longer given that individual trees can live hundreds of years, and forest succession takes longer than that (Odum, 1969; Finegan, 1984). At this time, we do not have an answer to the question regarding an appropriate optimization time scale, but it is likely...
related to the time scales over which biological predictions, acquired by evolution, can be reliably made, such as the sun will return tomorrow, or winter is coming. Predicting the future has obvious evolutionary advantages, but it also results in greater dissipation of free energy and entropy production. Some specific temporal predictions can be quite long, such as periodical cicadas that have strategies that operate on 13 and 17 year cycles, which clearly have been successful to date. We believe that determining the temporal scales that microbial communities operate over will be important for developing predictive biogeochemistry models, but time is not the only scale of importance.

Strategies that coordinate function over space also lead to greater free energy dissipation (Vallino, 2011). Some examples of such coordination include quorum sensing (Goo et al., 2012; Hmelo, 2017) and associated quorum policing (Whiteley et al., 2017), long-range metabolic signaling (Liu et al., 2015), stigmergy (Gloag et al., 2013), horizontal gene transfer (Treangen and Rocha, 2011), cables and nanowires (Reguera et al., 2005; Schauer et al., 2014), cross-feeding (Estrela et al., 2012; Rakoff-Nahoum et al., 2016), chemotaxis (Stocker and Seymour, 2012), vertical migration (Inoue and Iseri, 2012) and other types of chemical exchange (Moran, 2015). Our original intent was also to compare local versus global MEP optimization, but global optimization in the Siders Pond model did not produce biologically relevant results due to the speed at which electromagnetic potential is dissipated abiotically. Because water rapidly absorbs light, any water column of sufficient depth will dissipate all incoming radiation as entropy regardless of the presence of particles, so an infinite number of solutions exist that all produce the same global entropy production. For instance, a water column devoid of particles will produce entropy at an exponentially decreasing rate with depth, but eventually all the light will be dissipated. A solution in which the surface is particle rich will result in rapid entropy production with depth, but the same total amount of entropy will be produced. Particles placed lower in the water column would produce a different profile of entropy production, but the result would be the same, with all of the electromagnetic potential destroyed. Consequently, maximizing entropy production summed over the water column (i.e., global entropy maximization) does not produce a unique solution as it does for systems where energy potentials are stable, such as chemical potentials (Vallino, 2011). However, maximizing local entropy does select for a unique solution, which our results show is biologically relevant. In local maximization of entropy production, the objective is to maximize entropy production, \( \sigma(z) \), at every depth \( z \) simultaneously. The order in which \( \sigma(z) \) is maximized at each grid point does not matter, but in theory the optimization should be repeated until no further improvement occurs at any depth. Since light intensity is greatest at the surface, entropy production there can be greatly increased with an increase in particle density, which then lowers the amount of entropy that can be produced at depth due to shading. Of course, if the particles are phytoplankton and other organisms, resource availability will often limit biomass accumulation and local entropy production, but the system will organize in an attempt to minimize resource limitations, such as grazing and remineralization, and this is the solution observed in both the SIO and LIO simulations, with the latter outperforming the former.

This first MEP study involving light dissipation and phototrophy generated an additional interesting question—under what circumstances should entropy production be maximized locally versus globally? Our previous study with purely chemical reactions showed that global entropy production results in greater chemical potential destruction (Vallino, 2011). This study indicates that if abiotic processes can destroy an energy potential faster than biotic ones, then local MEP optimization will be the preferred solution. Global maximization of entropy production is more likely to be found in systems where energy potentials are stable with respect to abiotic decay, such as chemical potentials. However, it seems possible that both local and global optimization could be operating simultaneously is a system given that both light and chemical potentials exist. If so, one possible means of finding
the correct solution is to first conduct local optimization followed by global optimization. In the case of the Siders Pond MEP model, a solution could first be determined from local optimization as was described in the Results Section, then that solution could be used as an initial condition for global optimization. If the global solution cannot improve on the local optimization, then the local optimization is the final solution. However, if a solution found by local optimization can be improved upon by a global optimization then that solution should be chosen. We did not conduct this study, as the MEP modeling approach is computationally demanding, and other components of the approach could use further development, which we briefly explain next.

The adaptive Monod equation that servers as the kinetic driver, $F_T$ (e.g., Eq. (S45) and others) is based on simple observations of bacterial growth under oligotrophic to copiotrophic conditions, but there must be underlying biophysical constraints that form the basis of the tradeoff between growing fast versus taking up nutrients at low concentrations (Kirchman, 2016). While the equation might be improved by using allometric relationships (West and Brown, 2004), determining the true limitations using genome-scale models may produce more robust tradeoff models for growth kinetics (Casey et al., 2016; Levering et al., 2016). Condensing genome-scale models may also be a means to construct more realistic whole community metabolic networks from genomic surveys (Hanson et al., 2014; Hanemaaier et al., 2015), and exometabolomics could be used to identify metabolite nodes in the distributed metabolic network that are widely exchanged between functional groups (Klitgord and Segre, 2010; Baran et al., 2015; Fiore et al., 2015; Ponomarova and Patil, 2015). The reactions for predation in the Siders Pond model ($r_{LGZ}$ and $r_{LAGZ}$, Table 1) were designed so as not to introduce any new parameters (Vallino et al., 2014), but these reaction, which are intended to approximate virus to protist predation, is likely too simplistic and could use further development. Our reactions for carbon dioxide fixation ($r_{1,PHY}$, $r_{1,GSB}$ and $r_{1,SOX}$) could likely be improved by basing them on CO$_2$ fixation pathways that have been discovered to date, which also reveals connections to their thermodynamic efficiencies, $\xi_j$ (Bar-Even et al., 2012). Similarly, photoheterotrophy may be more widespread than appreciated, and there is a diversity of pathways (Kirchman and Hanson, 2013; Dubinsky et al., 2017), so our first attempt at representing photoheterotrophy could likely use refinement. Finally, the current approach uses non-linear optimal control to locate MEP solutions, but this is a computationally intensive problem, and falls under the class of problems known as control of partial differential equations (Coron, 2007). Building on expertise from that field could reduce computation requirements, especially for problems involving two and three dimensions, but it might be possible to avoid the formal optimization problem altogether. One possibility would be to use Darwinian inspired trait-based modeling approaches (Follows et al., 2007; Coles et al., 2017), which is, after all, how biology actually solves the MEP problem. The challenge in trait-based models is getting trade-offs correct, such as growth rate versus light level or resource storage. In this case, entropy production can be used to evaluate and cull models. For instance, when multiple versions of a model have been proposed, those models that produce more integrated entropy production over time and space should be selected for. MEP would also be useful for developing trait models that improve entropy production over time and space, such as resource storage, time delays (i.e., dormancy), migration across fronts and boundaries to acquire resources, clocks and oscillators, distribution and packaging of metabolic function, predation, remineralization and other regulatory motifs (Wolf and Arkin, 2003). By placing emphasis on the mechanisms of how energy potentials are destroyed over time and space, rather than on the peculiarities of how organisms grow and survive, can lead to new insights that can improve our understanding of biogeochemistry and model predictions thereof.

## 5 Conclusions
Development of the maximum entropy production (MEP) model is driven by the hypothesis that incorporating a fundamental principle in microbial biogeochemistry models will result in more robust extrapolation of model predictions. Our results demonstrate that models based on MEP can reasonably simulate how microbial communities organize and function in Siders Pond over time and space while using a minimum of adjustable parameters. The improved qualitative and quantitative agreement between model predictions and observations using long (3 d, LIO) versus short (0.25 d, SIO) interval optimization supports the hypothesis that biological systems maximize entropy production over long timescales. The modeling presented here extends the MEP approach to include an explicit spatial dimension, and new metabolic reactions were introduced to model phototrophs and entropy production associated with the destruction of electromagnetic potential. By considering the dissipation of both chemical and electromagnetic potentials, the MEP model shows that heterotrophs, such as bacteria, dissipate far more free energy in the photic zone by passive light absorption than by chemical respiration. Short interval optimization (SIO, 0.25 d) results in higher grazing rates and turnover of organic carbon, as well as rapid (bang-bang) changes in the reaction efficiency control variables, while long interval optimization supports higher phytoplankton growth and standing stocks near the surface of the pond. We also found that maximization of local entropy production, as opposed to global entropy production, must be used for energy potentials that are quickly dissipated by abiotic processes, such as light absorption by water and particles. However, our general conjecture that biological systems evolve and organize to maximize entropy production over the greatest possible spatial and temporal scales, while abiotic processes maximize instantaneous and local entropy production, remains valid.
6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

Both JV and JH contributed equally to project concepts and design, and both were full participants of the 24 hour endurance sampling of Siders Pond. Preliminary results from metatranscriptomics and metagenomics from JH were used to improve development of the metabolic network. JV developed the model and wrote first draft of the manuscript with editing and inputs during the process from JH. Both authors contributed to manuscript revision, read and approved the submitted version.

8 Funding

Primary funding for this project was from NSF GG grant EAR-1451356 to JV and JH, with additional support from Gordon and Betty Moore Foundation grant GBMF 3297. JV also received support from NSF Grants OCE-1637630 and OCE-1558710 and Simons Foundation grant 549941. The NSF Center for Dark Energy Biosphere Investigations (C-DEBI; OCE-0939564) also supported the participation of JH. This is C-DEBI contribution number XXX.

9 Acknowledgments

We thank Petra Byl for assistance in field work preparation and sampling of Siders Pond, as well as insightful conversations about the project. We also thank Jane Tucker for dissolved inorganic carbon analyses, Suzanne Thomas for hydrogen sulfide analyses, Sam Kelsey for dissolved organic carbon analyses, Kate Morkeski for sulfate, phosphate and particulate organic carbon analyses, and Rich McHorney, Leslie Murphy, Emily Reddington, and Gretta Serres for assistance with Siders Pond sampling logistics. We also thank the Semester in Environmental Science Program at MBL for use of their Hydrolab and Jon Boat use for sampling and a special thank you to Tom Gregg who allowed us to access Siders Pond through his property as well as for providing logistical support.
Energy dissipation drives microbial function

10 References


Microbes as energy dissipaters


Energy dissipation drives microbial function


Microbes as energy dissipaters


Energy dissipation drives microbial function


Microbes as energy dissipaters


This is a provisional file, not the final typeset article
11 Tables

Table 1. Reactions associated with the 8 biological catalysts, $\mathcal{S}_j$, used to model microbial biogeochemistry in Siders Pond, where $r_{ij}$ represents sub-reaction $i$ of biological catalyst $\mathcal{S}_j$. There are a total of 28 reactions, where $\mathcal{S}_{Gz}$ and $\mathcal{S}_{AGz}$ each catalyzed 8 sub-reactions. Reactions are shown to emphasize function only. Complete reaction stoichiometries, including influence of optimal control variables, are given in Section 2.7 of the Supplementary Material. See caption of Figure 1 for nomenclature.

<table>
<thead>
<tr>
<th>Rxn.</th>
<th>Abbreviated Stoichiometry</th>
<th>Cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{1,phy}$</td>
<td>$H_2CO_3 + h\nu \rightarrow C_{phy} + O_2$</td>
<td>$\mathcal{S}_{phy}$</td>
</tr>
<tr>
<td>$r_{2,phy}$</td>
<td>$C_{phy} + NH_3 + H_3PO_4 + O_2 \rightarrow \mathcal{S}_{phy} + H_2CO_3$</td>
<td>$\mathcal{S}_{phy}$</td>
</tr>
<tr>
<td>$r_{1,GSB}$</td>
<td>$H_2CO_3 + H_2S + h\nu \rightarrow C_{GSB} + H_2SO_4$</td>
<td>$\mathcal{S}_{GSB}$</td>
</tr>
<tr>
<td>$r_{2,GSB}$</td>
<td>$C_{GSB} + NH_3 + H_3PO_4 + H_2SO_4 \rightarrow \mathcal{S}_{GSB} + H_2CO_3 + H_2S$</td>
<td>$\mathcal{S}_{GSB}$</td>
</tr>
<tr>
<td>$r_{1−8,Gz}$</td>
<td>$S_i + C_i + O_2 \rightarrow \mathcal{S}_{Gz} + H_2CO_3 + C_D + NH_3 + N_D + H_3PO_4 + P_D + C_L$</td>
<td>$\mathcal{S}_{Gz}$</td>
</tr>
<tr>
<td>$r_{1−8,AGz}$</td>
<td>$S_i + C_i + H_2SO_4 \rightarrow \mathcal{S}_{AGz} + H_2CO_3 + C_D + NH_3 + N_D + H_3PO_4 + P_D + H_2S + C_L$</td>
<td>$\mathcal{S}_{AGz}$</td>
</tr>
<tr>
<td>$r_{1,Bac}$</td>
<td>$C_L + NH_3 + H_3PO_4 + O_2 \rightarrow \mathcal{S}_{Bac} + H_2CO_3$</td>
<td>$\mathcal{S}_{Bac}$</td>
</tr>
<tr>
<td>$r_{2,Bac}$</td>
<td>$C_D \rightarrow C_L$</td>
<td>$\mathcal{S}_{Bac}$</td>
</tr>
<tr>
<td>$r_{3,Bac}$</td>
<td>$P_D \rightarrow H_3PO_4$</td>
<td>$\mathcal{S}_{Bac}$</td>
</tr>
<tr>
<td>$r_{1,SRB}$</td>
<td>$C_L + NH_3 + H_3PO_4 + H_2SO_4 \rightarrow \mathcal{S}_{SRB} + H_2CO_3 + H_2S$</td>
<td>$\mathcal{S}_{SRB}$</td>
</tr>
<tr>
<td>$r_{2,SRB}$</td>
<td>$C_D \rightarrow C_L$</td>
<td>$\mathcal{S}_{SRB}$</td>
</tr>
<tr>
<td>$r_{3,SRB}$</td>
<td>$P_D \rightarrow H_3PO_4$</td>
<td>$\mathcal{S}_{SRB}$</td>
</tr>
<tr>
<td>$r_{1,PH}$</td>
<td>$C_L + NH_3 + H_3PO_4 + h\nu \rightarrow \mathcal{S}_{PH} + H_2CO_3$</td>
<td>$\mathcal{S}_{PH}$</td>
</tr>
<tr>
<td>$r_{1,SOx}$</td>
<td>$H_2CO_3 + H_2S + O_2 + NH_3 + H_3PO_4 \rightarrow \mathcal{S}_{SOx} + H_2SO_4$</td>
<td>$\mathcal{S}_{SOx}$</td>
</tr>
</tbody>
</table>
Figure captions

Fig. 1  Schematic of catalysts and associated reactions used in the MEP model for Siders Pond. Functional groups include: $\mathcal{S}_{Phy}$, phytoplankton, purple, 2 rxns; $\mathcal{S}_{GSB}$, green sulfur bacteria, brown, 2 rxns; $\mathcal{S}_{GZ}$, aerobic grazers, red, 8 rxns; $\mathcal{S}_{AGZ}$, anaerobic grazers, green, 8 rxns; $\mathcal{S}_{Bac}$, heterotrophic bacteria, blue, 3 rxns; $\mathcal{S}_{SRB}$, sulfate reducing bacteria, magenta, 3 rxns; $\mathcal{S}_{PH}$, phototrophic bacteria, cyan, 1 rxn; $\mathcal{S}_{SOX}$, sulfide oxidizing bacteria, orange, 1 rxn. Other abbreviations: $h\nu$, photon capture; $C_L$, labile organic carbon; $C_D$, refractory organic carbon; $P_D$, refractory organic phosphorous. See Table 1 for qualitative representation of functional reactions and Section 2.7 of the Supplementary Material for stoichiometrically balanced reactions.

Fig. 2 Layout of the Siders Pond 2D ($t,z$) sample grid. Samples were collected over a 24 h diel cycle starting at ~6:00 on 25 Jun 2015 and ending at ~6:00 on 26 Jun 2015. Samples were collected at 0.5, 2, 3, 4, 6, 8, 10 and 12 m.

Fig. 3 Contour plots of photosynthetic active radiation (PAR) (a-b), chlorophyll (Chl) a concentration (d-f) and dissolved oxygen (DO) (g-i) for six day simulations using short interval optimization (SIO) (left column), long interval optimization (LIO) (center column) and for observations collected from Siders Pond over a 24 hour period on Jun 25 to Jun 26 2015 (right column). Rectangle (dashed white lines) in simulation plots corresponds to time and depths where observations are comparable (i.e., Jun 25/26, 0.5 to 12 m). Actual observations are shown as white circles with black parameters (see also Fig. 2).

Fig. 4 Contour plots of inorganic phosphate (a-c) and dissolved inorganic carbon (DIC) (d-f) concentrations for SIO and LIO simulations (left and center columns, respectively) and Siders Pond observations (right column). See caption to Fig. 3 for other details.

Fig. 5 Contour plots of hydrogen sulfide (a-c) and sulfate (d-f) concentrations for SIO and LIO simulations (left and center columns, respectively) and Siders Pond observations (right column). See caption to Fig. 3 for other details.

Fig. 6 Contour plots of dissolved organic carbon (DOC) (a-c) and particulate organic carbon (POC) (d-f) concentrations for SIO and LIO simulations (left and center columns, respectively) and Siders Pond observations (right column). See caption to Fig. 3 for other details.

Fig. 7 Contour plots of simulated total entropy production, $\dot{\sigma}_T$ (Eq. (S178)) (a, e), entropy production from reactions, $\dot{\sigma}_R$ (Eq. (S178)) (b, f), and entropy production from light absorption by water, $\dot{\sigma}_W(t,z)$ (Eq. (S176)) (c, g) and particles, $\dot{\sigma}_P$ (Eq. (S177)) (d, h) for the SIO (left column) and LIO (right column) simulations. Entropy production from light dissipation by water and particles are shown as a percentage of total entropy production.

Fig. 8 Contour plots of phytoplankton concentration, $[\mathcal{S}_{Phy}]$ (a, f), rates for the two associated reactions, $r_{1,Phy}$ (b, g) and, $r_{2,Phy}$ (c, h) and two optimal control variables, $\varepsilon_{Phy}$ (d, i), and $\Omega_{1,Phy}$ (e, j) for the SIO (a-e) and LIO (f-j) simulations. Note, reaction rates and control variables are only plotted for the top 5 m of the water column where the processes are significant.
Fig. 9 Contour plots of green sulfur bacteria concentration, $[\mathcal{S}_{GSB}]$ (a, d), rate of the biosynthesis reaction, $r_{2,GSB}$ (b, e) and the growth efficiency optimal control variable, $\varepsilon_{GSB}$ (c, f) for the SIO (a-c) and LIO (d-f) simulations. Note, variables are only plotted for 10 m to 15 m where the GSB are active.

Fig. 10 Contour plots of photoheterotroph concentration, $[\mathcal{S}_{PH}]$ (a, d), rate of the carbon fixation reaction, $r_{1,PH}$ (b, e) and the growth efficiency optimal control variable, $\varepsilon_{PH}$ (c, f) for the SIO (a-c) and LIO (d-f) simulations. Note, contours for $\varepsilon_{PH}$ were selected to emphasize $\varepsilon_{PH}$ values near zero.

Fig. 11 Contour plots of grazer concentration, $[\mathcal{S}_{Gz}]$ (a, e) and rates of predation on, grazers, $r_{3,Gz}$ (b, f), bacteria, $r_{5,Gz}$ (c, g), and photoheterotrophs, $r_{7,Gz}$ (d, h) for the SIO (a-d) and LIO (d-h) simulations.

Fig. 12 Contour plots of entropy production by bacteria associated with light absorption by bacterial biomass, $\dot{\sigma}_{P,Bac}$ (a, c) and growth of bacteria, $r_{1,Bac}$, on organic carbon, $\dot{\sigma}_{1,Bac}$, Eq. (S94) (b, d) for the SIO (a, b) and LIO (c, d) simulations.
Figure 5.

The figure shows the distribution of hydrogen sulfide and sulfate concentrations in a water column over time. The color bars indicate the concentration levels, with higher values shown in darker colors. The x-axis represents the date, with specific dates highlighted for analysis. The y-axis represents depth, with layers indicating different depth ranges. The figure illustrates the temporal and spatial variation of these chemical components in the aquatic environment.
Figure 9. JPEG
In review