1. Introduction

Earth's biogeochemical cycles that dictate mass and energy flow on a planetary scale result from a complex interaction of individual microscopic organisms through growth, competition and cooperation. Because biogeochemistry is ultimately driven by organisms, it is natural and customary to simulate these processes by modeling the growth of individual guilds and examining the biogeochemistry that results (e.g., Le Quere et al. 2005). However, modeling individuals to understand and predict biogeochemistry introduces many challenges due to the extensive amount of information that is necessary to parameterize and constrain such models, such as organism maximum specific growth rates, substrate and prey affinities, growth efficiencies, etc. (Vallino 2000, Ward et al. 2010). For systems that can be sufficiently sampled over time and space, organism-based models perform respectably, provided model forecasts do not extrapolate beyond the data envelope used for model calibration. The reason organism-based models extrapolate poorly is that they are not based on any governing principles, except conservation of mass and sometimes energy, which are insufficient constraints for describing how living systems function. How can we understand development of biogeochemical cycles and how they operated in the distant past, or how they may change in the future, if little or no information is available on the individuals that give rise to such cycles?

Current biogeochemical cycles are influenced not only by bacteria, archaea, viruses and fungi, but also by macroscopic multicellular organisms; however, the majority of biogeochemical reactions are only catalyzed by bacteria and archaea (Falkowski et al. 2008), and microbes were the only organisms present for the first ~2.5 Gyr of Earth's 3.8 Gyr fossil record (McKeegan et al. 2007). Consequently, in this proposal we will only focus on bacteria and archaea, and we will view them as simple molecular machines (Falkowski et al. 2008), or more appropriately, as biogeochemical catalysts. (Macroscopic organisms can be readily included in our framework by accounting for their enhancement of mass transport processes, such as via mastication, lungs, burrowing, sediment trapping, etc, but this lies outside the scope of the current proposal.) Our objective is to understand and predict the allocation of biological machinery to biogeochemical reactions and the associated catalytic enhancement of reaction rates.

In order to understand and predict how microbial communities give rise to biogeochemical cycles, it is necessary to take a different perspective that focuses directly on the flow of energy and mass through biochemical reactions at the expense of understanding finer scale dynamics, such as the contribution of individual species. This approach is analogous to modeling climate instead of weather. Thermodynamics is a particularly appropriate framework for this due to its ability to predict relationships between macroscopic state properties, such as pressure, temperature and volume as related by the ideal gas law, without requiring information on the microscopic states, such as molecular position and momentum, that ultimately give rise to the macroscopic properties (Kondepudi and Prigogine 1998). Even though a small volume of matter is composed of a large number of atoms (a liter of water contains 10^{25} molecules) that can give rise to an astounding number of degrees of freedom, the success of thermodynamics tells us that most of these degrees of freedom are unimportant for describing the macroscopic properties and the relationships between matter and energy, at least for systems at equilibrium. Furthermore, the energy released by a chemical reaction is independent of the nature of the living or abiotic system that catalyzes it. As a result of these advantages, there have been numerous research avenues devoted to thermodynamic or thermodynamic-like approaches for understand ecosystem processes, since ecosystems exhibit biogeochemical reproducibility but are composed of many interacting organisms (e.g., Lotka 1922, Odum and Pinkerton 1955). However, living systems are not at equilibrium (Schrödinger 1944, Morowitz 1968), meaning nonequilibrium thermodynamics must be invoked. This proposal will further develop the concepts of nonequilibrium thermodynamics for understanding biogeochemical processes.

2. Objectives

In this proposal, we will build on our previous research for modeling microbial biogeochemistry using the principle of maximum entropy production (MEP) that derives from nonequilibrium thermodynamics. We will represent microbial communities as a distributed metabolic network where the synthesis and

allocation of molecular machinery to metabolic pathways is determined from a MEP-based optimization over successive intervals of time. In particular, we propose to 1) extend the mathematical framework and numerical algorithms of the current MEP model so that predictions can be extended over space and directly compared to observations collected from Siders Pond, a meromictic pond located on Cape Cod, MA; 2) extend the metabolic network to incorporate phototrophic reactions; 3) assay biogeochemical constituents over depth in Siders Pond and determine allocation of molecular machinery to primary metabolic pathways using metagenomics and metatranscriptomics; 4) assess the utility of the MEP hypothesis by comparing MEP-based model predictions to biogeochemical and genomic observations over time and space in Siders Pond.

3. Theoretical Background and Results to Date

3.1 Maximum Entropy Production (MEP)

Numerous theories describing ecosystem organization and function have been proposed in theoretical ecology dating back to at least Lotka (1922), who proposed ecosystems organize to maximize power. However, in this proposal we are particularly interested in the maximum entropy production (MEP) conjecture (Paltridge 1975, Dewar 2003, Dewar 2005, Martyushev and Seleznev 2006, Niven 2009), which states that steady state, nonequilibrium systems with many degrees of freedom will likely be found in a macrostate that maximizes internal entropy production. If internal self-organization, such as vortices and macroscopic structures, facilitates internal entropy production, then those structures will likely develop (Lorenz 2003), but the theory makes no distinction between biotic or abiotic systems. Entropy here refers to the classic thermodynamic definition of Gibbs and Boltzmann. Similar to equilibrium thermodynamics that require systems to be found in a state of maximum entropy, MEP indicates that *non*equilibrium systems will head towards equilibrium along the fastest possible pathway. That is, they will dissipate free energy as fast as possible within the constraints imposed on the system (Makela and Annila 2010, Vallino 2010). Several phenomena appear consistent with MEP, including planetary-scale heat transport (Lorenz et al. 2001, Kleidon et al. 2003), laminar to turbulent flow transition (Martyushev 2007), plant evapotranspiration (Wang and Bras 2011), atmospheric and ocean circulation (Kleidon et al. 2003, Shimokawa and Ozawa 2007) and many others (see Dewar et al. 2014). When applied to biological systems, the MEP conjecture leads to a new paradigm from '*… "we eat food" to "food has produced us to eat it"*' (Lineweaver and Egan 2008). MEP provides directionality to evolution of the biosphere, in that it should progress towards states of higher entropy production. The global succession of anoxygenic phototrophs to oxygenic phototrophs is one example of this progression.

We have developed an initial theoretical framework for describing microbial biogeochemistry as a type of dissipative system governed by maximum entropy production (Vallino 2010, Vallino 2011, Algar and Vallino 2014, Vallino et al. 2014) that forms the basis of this proposal. Here we will briefly describe the theoretical foundation of the approach, its mathematical implementation and some results to date. New advancements we will pursue in this proposal are described in Section 5. The model is founded on the hypothesis that microbial communities evolve, adapt and organize to extract as much free energy from the environment as available resources (N, P, S, etc.) and information allow. Genomic information is critical, as it ultimately determines the set of molecular machines—catalysts in particular—and metabolic functions that can be constructed from the available environmental resources. This information also includes designs for the molecular machines that turnover machinery, namely protists, predatory bacteria and viruses, which allow the system to be dynamic and adaptive to changes in environmental drivers. Under this context, food webs are hierarchical systems that effectively dissipate available free energy, or analogously, maximize the rate of entropy production. Our modeling approach determines the allocation of molecular machinery to metabolic function that is dynamically distributed across many phyla by maximizing entropy production.

While actively discussed by the community, uncertainty remains regarding the spatial scale over which MEP applies (Lucia 2012). Our preliminary work (discussed below) indicates that MEP applies at a

systems level as defined by the extent of matter and information connectivity (Vallino 2011). Hence, a single bacterium does not maximize entropy production because it does not dissipate all chemical potential (some of the free energy in the food remains in the microbial biomass produced, so all chemical potential is not destroyed). However, a microbial community does achieve a MEP state because the growth of each organism is more or less consumed by a predator, and the predator by its predator, and so on. Total biomass of the community will increase until either all incoming energy is consumed (energy limited), or resources limit the amount of catalyst (biomass) that can be synthesized (resource limited). For instance, the surface ocean is resource limited (typically by N, P or Fe), while the deep ocean is energy limited. Regardless of the limitation, the ecosystem operates near a pseudo steady state where the assimilated free energy is simply dissipated by the spinning of the ecosystem network, since chemical potential (biomass) does not accumulate (**Fig. 1**). Microbes and microbial consortium so effectively organize that nearly all chemical potentials found in the environment are readily accessed and dissipated. Hoehler (2004) estimated that reaction free energy potentials as low as 9 kJ mol^{-1} can be exploited by microbes.

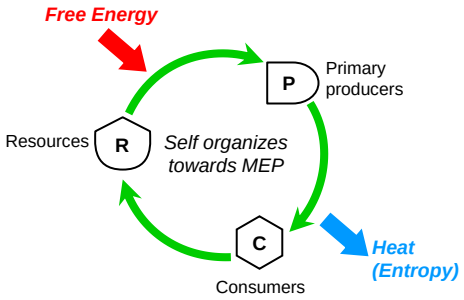

Fig. 1.Ecosystems function as purely dissipative systems, since all acquired energy is ultimately dissipated as heat.

3.2 Distributed Metabolic Network Representation

Being a thermodynamic theory, MEP does not describe the structure of a microbial community (for instance, bacteria can be grazed by protists or turned over by viruses). In fact, the MEP conjecture indicates that there should be many different community configurations that give rise to the same entropy production, and there is some experimental support for this idea (Fernandez et al. 1999, Wittebolle et al. 2008, Vallino et al. 2014). Because of this interchangeability, we do not attempt to model specific organisms, but rather only constrain the system by the metabolic functions it is capable of expressing. Consequently, the microbial community is represented as a distributed metabolic network (Vallino 2003), where each metabolic function is catalyzed by an associated *biological structure*, $\mathbb{S}_i$, that is intended to capture the general capabilities of the community for function *i*. Only recently have systems biology approaches been attempted for modeling distributed metabolic networks (Klitgord and Segre 2011, De Filippo et al. 2012, Zomorrodi et al. 2014), but we developed a more aggregated approach where only extracellular metabolites are included as network nodes, and functional pathways that may consist of 10s or 100s of enzymatic steps are represented by a single biological structure, $\mathbb{S}_i$ and reaction. For example, we simulated a methanotrophic community with only 4 biological structures and 8 extracellular metabolites (**Fig. 2**) (Vallino et al. 2014).

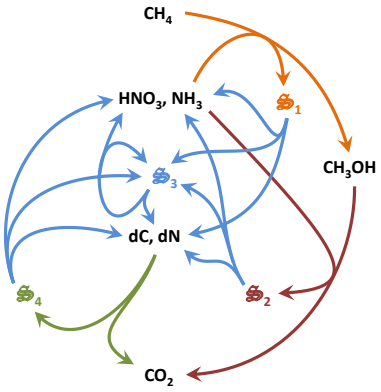

Fig. 2. Methanotrophic metabolic network. Reactions are catalyzed by the $\mathbb{S}_i$ of the same color.

The implementation of the reactions in a metabolic network is designed to place as many degrees of freedom as possible in optimal control variables that are determined by maximizing entropy production, while minimizing the number of adjustable parameters that often plague standard biogeochemical models (Ward et al. 2010). The general approach is to use a growth efficiency term, ε_i, as the control variable for each biological structure, $\mathbb{S}_i$, that catalyzes a metabolic function, including those that serve to turn over catalyst (i.e., viruses or predators). Consider the following two-reaction network for glucose oxidation:

$$\mathrm{CH_2O} + \gamma\mathrm{NH_3} + (1-\varepsilon_1)\mathrm{O_2} \xrightarrow{\mathbb{S}_1} \varepsilon_1\mathbb{S}_1 + (1-\varepsilon_1)\mathrm{H_2CO_3} \tag{1}$$

$$\pi_1\mathbb{S}_1 + \pi_2\mathbb{S}_2 + (1-\varepsilon_2)\mathrm{O_2} \xrightarrow{\mathbb{S}_2} \varepsilon_2\mathbb{S}_2 + (1-\varepsilon_2)(\mathrm{H_2CO_3} + \gamma\mathrm{NH_3}) \tag{2}$$

For simplicity we have assumed $\mathbb{S}_i$ composition is given by $CH_2O(NH_3)_\gamma$, and π_i represent feeding preference, if any, by $\mathbb{S}_2$. This simple two-reaction network has several important features critical to implementing the MEP model. 1) Each reaction is catalyzed by its respective biological structure, $\mathbb{S}_i$, so reaction rates depend on the concentration of the biological structure, $c_{\mathbb{S}_i}$, but the reaction can also synthesize catalyst (i.e., they are autocatalytic reactions that can grow exponentially). 2) The coupled reactions can operate in a futile cycle perpetually turning over biological structure fueled by glucose oxidation (e.g., **Fig. 1**). 3) The growth efficiency parameter, ε_i, selects the degree to which reduced organic carbon is either converted to more catalyst (pure anabolic reaction as $\varepsilon_i \to 1$) or oxidized to CO_2 and H_2O (pure catabolic reaction as $\varepsilon_i \to 0$), where the latter reaction produces large amounts of entropy; however, catalyst must also be present for the reactions to proceed. Note, the free energy of reaction (1) remains negative (i.e., can occur spontaneously) even when $\varepsilon_i = 1$, because living organisms, contrary to popular believe, are ***not*** low entropy structures (Vallino 2010, Martyushev 2013).

3.3 Reaction Kinetics

To model the rates for reactions (1) and (2), we use the following novel expression,

$$r_i = \nu^* \varepsilon_i^2 \left(1 - \varepsilon_i^2\right) c_{\mathbb{S}_i} \prod_j \left(\frac{c_j}{c_j + \kappa^* \varepsilon_i^4} \right), \tag{3}$$

that is similar to the classic multi-substrate Monod growth equation where c_j are substrate concentrations, (c_{CH_2O} and c_{NH_3}) or ($c_{\mathbb{S}_1} + c_{\mathbb{S}_2}$), but the maximum specific growth rate (μ^M) is replace by $\nu^* \varepsilon_i^2 \left(1 - \varepsilon_i^2\right)$ and the half saturation, or Monod, constant is given by $\kappa^* \varepsilon_i^4$. The fixed parameters ν^* (= 350 d^{-1}) and κ^* (= 5000 mmol m^{-3}) have been chosen such that Eq. (3) can simulate bacteria growing in oligotrophic conditions, such as the middle of ocean gyres (i.e., $\mu^M \ll 1\ d^{-1}$), to bacteria such as *E. coli* growing in rich media under ideal conditions (i.e., $\mu^M > 50\ d^{-1}$) by simply changing the value of the growth efficiency, ε_i, between 0 and 1 (see Vallino 2011), but also see (Cajal-Medrano and Maske 1999). The kinetic equation (3) also accounts for the thermodynamic tradeoff between power and efficiency, because as ε_i approaches 1, reaction rate is driven to 0 due to the loss of thermodynamic force (Jin and Bethke 2003). As anabolic-catabolic coupled reactions approach 100% energy transfer efficiency ($\varepsilon \to 1, \Delta_r G \approx 0$), they proceed reversibly, so infinitely slowly.

3.4 Internal Entropy Production

Change in system entropy, dS, is the sum of entropy exchange with the environment, dS_e, and internal entropy production, dS_i, due to irreversible processes: $dS = dS_e + dS_i$ (Prigogine et al. 1972). While entropy exchange can be either positive or negative (it is defined by heat and mass exchange), the Second Law requires $dS_i \geq 0$. An entropy balance equation then takes the form,

$$\frac{dS}{dt} = \sum\nolimits_k F_k \bar{S}_k + \frac{Q}{T} + \dot{\sigma}_i, \tag{4}$$

where S is system entropy, F_k and $\bar{S}_k$ are molar flux and molar entropy of constituent k transported into the system, respectively, and Q is heat flux into the system at temperature T. Internal entropy production, $\dot{\sigma}_i$, is due to internal irreversible (dissipative) processes ($\dot{\sigma}_i \geq 0$) and it is the quantity that MEP considers; MEP does not pertain to system entropy, S, which can either increase or decrease.

Internal entropy production rate associated with reactions, $\dot{\sigma}_r$ (J K^{-1} d^{-1}), is readily calculated from the reaction rate, r_i (mmol m^{-3} d^{-1}), Gibbs free energy of reaction, $\Delta_r G_i$ (J $mmol^{-1}$), system volume, V (m^3), and temperature, T (K) (Vallino 2010). For the simple two reaction network given above, reaction entropy production is given by,

$$\dot{\sigma}_r = -\frac{V}{T}\left(r_1(\varepsilon_1, \mathbf{c})\Delta_r G_1(\varepsilon_1, \mathbf{c}) + r_2(\varepsilon_2, \mathbf{c})\Delta_r G_2(\varepsilon_2, \mathbf{c})\right). \quad (5)$$

As evident in Eqs. (1-3), both reaction rates and associated Gibbs free energy of reaction depend on the growth efficiency control variables, ε_i, and the concentrations of state variables, $\mathbf{c}$. To calculate Gibbs free energies of reaction, we account for concentration of reactants and products as well as for activity coefficients (Alberty 2003). Also, as described by Alberty (2003), we explicitly account for proton dissociation equilibria between chemical species via pH, so "H_2CO_3" and "NH_3" in the above reactions represent $H_2CO_3 + HCO_3^- + CO_3^{2-}$ and $NH_3(aq) + NH_4^+$, respectively (all weak acids and bases are accounted for similarly). Not shown here, but we also calculate entropy production from mixing, $\dot{\sigma}_m$, but these terms are only a small fraction of the entropy of reaction, $\dot{\sigma}_r$, given by Eq. (5), so that total internal entropy production ($\dot{\sigma}_i = \dot{\sigma}_r + \dot{\sigma}_m$) is well approximated with $\dot{\sigma}_i \cong \dot{\sigma}_r$ (see Vallino 2011 for details). In the two reaction network, maximizing the rate of glucose (CH_2O) oxidation maximizes entropy production; chemical potential is simply converted to heat that is transported out of the system. However, catalyst, $\mathbb{S}_i$, must be synthesized to increase reaction rates, but excessive catalyst synthesis results in lower entropy production (chemical potential gets locked in $\mathbb{S}_i$). **The MEP conjecture argues that systems organize to locate this natural balance between catalyst synthesis and reduced-carbon oxidation to maximize internal entropy production**.

The above description for the simple two-reaction network can be readily extended to more complex networks (see below), as well as to numerous substrates and products; abiotic reactions are also easily incorporated, but their kinetic expressions are required. Reaction rates, $\mathbf{r}$, determined from maximizing entropy production are used in a standard state-space mass-balance model, as given by

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{f}(\mathbf{c}(t); \mathbf{c}^E) + \mathbf{A}\mathbf{r}(\mathbf{c}(t); \boldsymbol{\varepsilon}(t)), \quad (6)$$

where $\mathbf{c}(t)$ is a vector of extracellular metabolite (i.e., c_{NH_3}, c_{H_2S}, etc.) and biological structure ($c_{\mathbb{S}_i}$) concentrations, $\mathbf{A}$ is a reaction stoichiometric matrix based on the metabolic network and $\mathbf{f}(\mathbf{c}(t); \mathbf{c}^E)$ accounts for transport of constituents across the system boundary that may depend on external concentrations, $\mathbf{c}^E$. Solution to Eq. (6) provides prediction on how metabolite concentrations vary over time as well as how allocation of biological structure to metabolic pathways changes over time. Reaction rates, $\mathbf{r}(t)$, are also predicted as well as activity or expression of biological structure as given by $\boldsymbol{\varepsilon}(t)$.

3.5 MEP Optimization: instantaneous versus average entropy production.

The metabolic reaction rates, $\mathbf{r}(t)$, are determined by maximizing internal entropy production rate by adjusting $\boldsymbol{\varepsilon}(t)$; however, if entropy production is maximized at each time point during solution of Eq. (6), we find the model behaves nothing like biology. While chemical potential, such as $CH_2O + O_2$, is destroyed as expected, so is all biological structure. It was quickly realized that this occurs because $\mathbb{S}$ is just reduced organic carbon and its oxidation produces entropy. If entropy production is maximized instantaneously, then $\mathbb{S}$ is never produced, because this reduces instantaneous entropy production, and whatever $\mathbb{S}$ is present at the start of a simulation is oxidized. To solve this problem, we maximize entropy production over an interval of time. In this case, the model produces results comparable to observations (see below), because if more catalyst exists at the end of the interval then more chemical potential can be destroyed, so investment in $\mathbb{S}$ synthesis occurs until either resources (N, P, Fe, etc) limit $\mathbb{S}$ synthesis, or sufficient $\mathbb{S}$ exists to destroy all available free energy.

The computational remedy to the MEP optimization problem introduced an intriguing hypothesis: a proposed distinction between biotic and abiotic systems. Abiotic systems, such as fire or a rock rolling down a hill, maximize instantaneous entropy production (e.g., they take a steepest descent pathway down a potential energy surface), while biotic systems use information stored in the metagenome to execute

temporal strategies acquired via evolution to maximize entropy production over a finite period of time (Vallino 2010). A simple example is the storage of internal energy reserves (e.g., starches) to permit metabolic function to persist when external energy sources become temporarily unavailable, such as in animal hibernation and plant dormancy over winter. Circadian rhythm is another example that allows phototrophs to "predict" the sun will return and can orchestrate metabolic machinery appropriately before sunrise. While temporal strategies are well known and appreciated in biology, it is surprising that no biogeochemical models we are aware of account for such **anticipatory control**.

To maximize averaged entropy production over time, we use a receding horizon optimal control approach (Vallino et al. 2014) to find the reaction efficiency terms, $\varepsilon_i(t)$, that maximize internal entropy production over a long time interval (Δt^*) starting at the current time t_n, as follows,

$$\begin{aligned} &\max_{\boldsymbol{\varepsilon}} \frac{1}{\Delta t^*} \int_{t_n}^{t_n+\Delta t^*} \dot{\sigma}_i(\tau) e^{-k_\sigma(\tau - t_n)} d\tau \\ &subject\ to\text{:}\ \frac{d\mathbf{c}(t)}{dt} = \mathbf{f}(\mathbf{c}(t); \mathbf{c}^E) + \mathbf{A}\mathbf{r}\big(\mathbf{c}(t); \boldsymbol{\varepsilon}(t)\big) \text{ and } \mathbf{0} \le \boldsymbol{\varepsilon} \le \mathbf{1}. \end{aligned} \tag{7}$$

Because prediction uncertainty increases with time, the exponential term in Eq. (7), $e^{-k_\sigma(\tau - t_n)}$, discounts the contribution of entropy production to the solution the farther in time the integration proceeds based on the magnitude of k_σ. The *subject to* constraints are derived from the mass balance model, Eq. (6), and box constraints on the control variables. Once the solution is determined over the long Δt^* interval, a shorter time step, Δt, is actually taken, so the next optimization interval for Eq. (7) is started at $t_n + \Delta t$, and the realized average entropy production over the shorter Δt interval is calculated as,

$$\langle \dot{\sigma}_i(t_{n+1}) \rangle = \frac{1}{\Delta t} \int_{t_n}^{t_n+\Delta t} \dot{\sigma}_i(\tau) d\tau \tag{8}$$

A solution can be determined for any length of time by repetitively solving Eqs. (7) and (8) over a sufficient number of sequential intervals.

4.5.1 Results from two MEP-based studies. To examine performance of our MEP approach, we modeled a laboratory methanotrophic-based microbial community subject to constant and periodic inputs of energy (Vallino 2010, Vallino et al. 2014). The experiment (also see http://ecosystems.mbl.edu/MEP) consisted of four 18 L microcosms (MCs) operating in chemostat mode (0.1 d^{-1} dilution rate) that were inoculated with whole water collected from a cedar bog located on Cape Cod, MA. Two MCs (MC 2 and 3) were continuously sparged with 4.9% methane in air, while two other MCs (MC 1 and 4) were periodically sparged for 10 d with 4.9% methane in air, then 10 d with just air (20 d period). Samples were regularly removed for nutrient analyses, and microbial community composition was determined by sequencing the V4 and V6 hypervariable regions of the ribosomal 16S gene using 454 pyrosequencing (Huber et al. 2007). Because we were able to include most of the model's degrees of freedom in the MEP optimization, the model only requires two adjustable parameters. Nevertheless, the MEP-based model was able to accurately simulate both the control and cycled MCs (**Fig. 3**) (Vallino et al. 2014). An important finding was that the model was only able to capture the dynamics of the cycled MCs if the optimization interval, Δt, was on the order of the cycle period (20 d). Furthermore, the MC's exhibited relatively stable methane

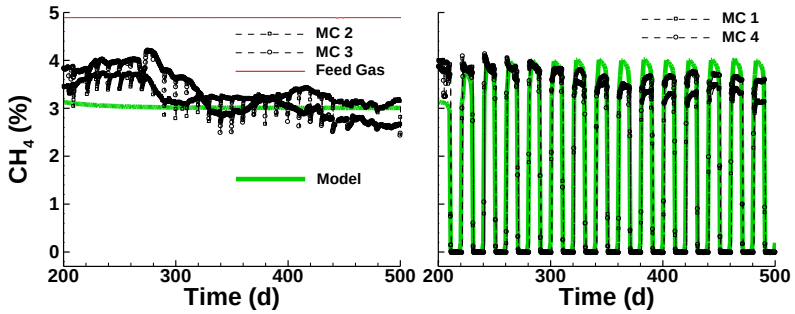

Fig. 3. Modeled (green) and observed (black) methane concentrations exiting MCs for both the control (left) and cycled (right) treatments.

oxidation rate (**Fig. 3**) even though community composition varied dramatically over the 500 d experiment (**Fig. 4**) (a property consistent with the MEP conjecture). Attempting to model such community dynamics would be very challenging using systems biology approaches (e.g., Zomorrodi et al. 2014), yet the MEP approach worked well at simulating mass and energy flow through the community using a relatively simple metabolic network (**Fig. 2**).

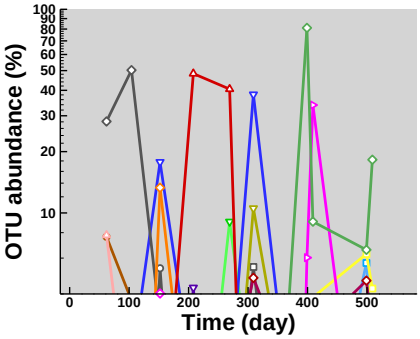

Fig. 4. Bacterial and archaeal species abundance (%), as operational taxonomic units (OTU) from 454 tag pyrosequencing for MC 3 (control). Only OTUs with at least 4% abundance are shown, but 18,610 taxa were detected across all four MCs (at 97% similarity).

We have also used the MEP model to examine metabolic switching between denitrification, dissimilatory nitrate reduction to ammonium (DNRA) and anammox that occurs during anaerobic nitrate reduction in sediments and in oceanic oxygen minimum zones (Algar and Vallino 2014). MEP model simulations (**Fig. 5**) show that the metabolic switching between the three pathways (**Fig. 5**, $\mathbb{S}_2, \mathbb{S}_3, \mathbb{S}_4$, top) is driven by the relative concentration of labile organic carbon $[CH_2O]$ to nitrate $[NO_3^-]$. When labile carbon is limited, the MEP model favors anammox, but as labile carbon becomes more available relative to nitrate, metabolism shifts toward denitrification (**Fig. 5**, bottom). At high levels of carbon input relative to nitrate, DNRA becomes the dominate pathway. Even though no parameter values were adjusted, the MEP model is consistent with general observations on anaerobic nitrate utilization (Burgin and Hamilton 2007), including a recent study showing the importance of C and N stoichiometry (Babbin et al. 2014).

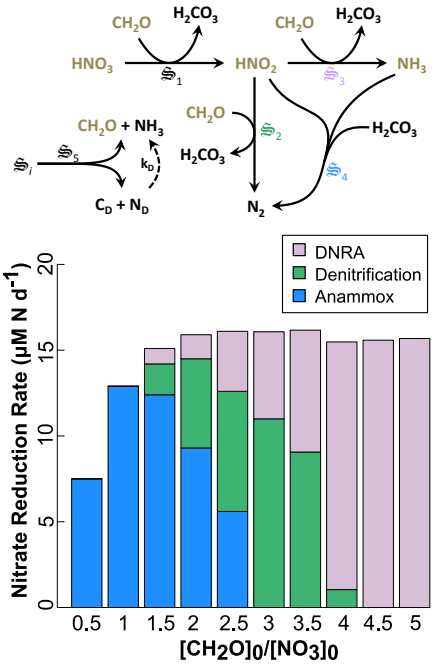

Fig. 5. MEP-based metabolic network (top) and metabolic swiching as a function of labile carbon to nitrate input ratio (bottom) for anaerobic nitrate reduction.

Our results from these two studies support the conjecture that MEP can be used to predict microbial biogeochemistry, but they were 0D models that did not involve a spatial dimension. Introducing spatial dimensions introduces another interesting question.

3.6 MEP Optimization: local versus global

Similar to the instantaneous versus average entropy production question above, when one or more spatial dimension is added to a MEP model, the question arises if entropy production should be maximized at each grid point independently (i.e., locally: $\sum_{i,j,k} \max \dot{\sigma}(i,j,k)$) or over the entire model domain (i.e., globally: $\max \sum_{i,j,k} \dot{\sigma}(i,j,k)$), and does it matter which approach is use (**Fig. 6**)? To investigate this question we developed a simple two box model crudely representing the surface and deep ocean (Vallino 2011). Our results show that maximizing entropy production globally can result in greater entropy production than maximizing local entropy production; however, global optimization requires that the system must coordinate function over space. While it is well known that spatial coordination via quorum sensing compounds occurs in microbial communities that are in close physical contact such as in biofilms (Decho et al. 2010, Goo et al.

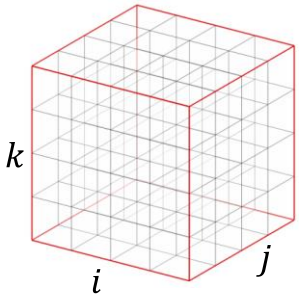

Fig. 6. Maximize $\dot{\sigma}$ at each grid point (local) or over the sum of all grid points (global)?

2012, Mitri and Foster 2013), it is not clear if coordination occurs over larger spatial scales. While our MEP model can be based on maximizing entropy production either locally or globally, it is unknown which approach is preferred until we can develop a spatial MEP model and test it with field observations. Consequently, a primary objective is to extend our MEP model to include spatial dimension and compare model predictions to observations collected from the environment at Siders Pond.

4. Proposal Work Plan

Our objective is to demonstrate that the MEP-based model can predict microbial biogeochemistry that occurs in natural environments (supporting the MEP conjecture), as well as determine if natural communities maximize entropy production locally or globally. We are also interested in quantifying the optimization interval, Δt, that most accurately describes natural communities. This latter question is fundamentally important as it provides the time scale over which natural communities have evolved to operate. Two computational objectives are to extend the model to handle phototrophy and optimization over space.

Siders Pond, located on the southern shore of Cape Cod, MA will serve as the study site, because this permanently stratified meromictic pond harbors populations of both oxygenic and anoxygenic photoautotrophs (green sulfur bacteria) as well as coupled redox cascades involving O, S, N and Fe. The site has the advantage that redox cascades occur over meters rather than mm as found in sediments and microbial mats; consequently, the biogeochemistry and community genomics/transcriptomics can be readily sampled over space (and time). We note, however, that our objective is *not* to develop a detailed model for Siders Pond, but rather use it as an experimental system for testing the MEP model.

4.1 Site Description: Siders Pond, MA

Siders Pond was extensively studied between 1980 and 1983 (Caraco 1986), but we have continued to characterize the site for the last 15 years as part of MBL's undergraduate Semester in Environmental Science (SES) Program that both PI's are faculty of. Siders Pond is a small (volume: 10^6 m^3; area: 13.4 ha; max depth: 15 m) coastal meromictic kettle hole that receives $1x10^6$ m^3 of fresh and $0.15\ x10^6$ m^3 of saltwater each year. The latter input occurs via episodic inputs during extreme tides and storm events via a small creek that connects the pond to Vineyard Sound approximately 550 m to the south. Tritium-helium water dating shows a relatively high vertical mixing rate of 2.5 and 7.5 m y^{-1} 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 a N+P enrichment study (Caraco et al. 1987) shows phytoplankton to be P limited, especially in the low salinity surface waters.

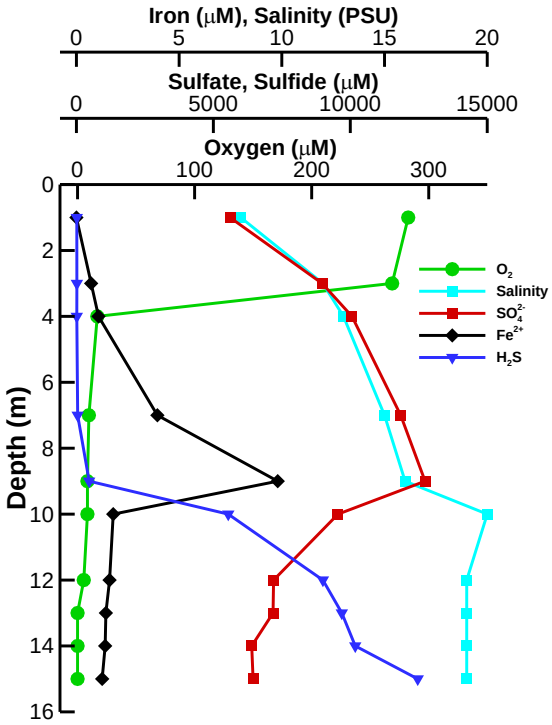

Fig. 7. Depth profiles of selected constituents taken in 2004 form Siders Ponds, MA by SES students.

Our more recent depth profile surveys of dissolved constituents from SES students (**Fig. 7**) differ little from those surveyed in the early 1980's. A steep oxycline occurs at approximately 3-4 m, and hydrogen sulfide begins to accumulate at 9 m (**Fig. 7**). The peak in ferrous iron concentration at ~9 m is believed to be caused by the dissolution of iron oxides below the

oxycline and the precipitation of iron sulfide in the bottom water. Formation of ferric phosphate in the surface waters appears to be a significant sink for phosphate and may reduce P availability to autotrophs by 30% (Caraco 1986). However, there could also be significant bacterial catalysis of the iron redox reactions, which this study will investigate.

Siders Pond is eutrophic averaging 16 mg m^{-3} 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}. 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). Sulfur cycling was also investigated by a SES undergraduate (Nalven 2011) who cloned and sequenced dissimilatory sulfite reductase (*dsr*) genes and found the gene distributed throughout the anoxic portion of the water column, but *dsr* sequences in 12 m samples were genetically distinct from *dsr* found in 4 and 8 m samples.

4.2 Siders Pond Sample Collection for MEP Model Testing

For this study we are planning on sampling 1D spatial (vertical) biogeochemically relevant constituents and gene abundance/expression at only the diel time scale (not seasonal or annual). Consequently, we will collect samples from a single station located at the deepest point in the pond (15 m) over a two day period in July or August of the first project year only. Samples will be collected over 8 depths and at 8 different time points over the two day period (64 samples), which will allow us to examine the circadian rhythm time scale. Samples will be analyzed for Chl a, BChl cde, NO_3^-, NO_2^-, NH_4^+, PO_4^{3-}, SO_4^{2-}, H_2S, O_2, $Fe^{2+, 3+}$, salinity, T, pH, photosynthetic active radiation (PAR), dissolved inorganic carbon (DIC), particulate organic C (POC) and N (PON). Primary productivity and community respiration will be measured at two depths corresponding to chlorophyll peaks for algae/cyanobacteria and green sulfur bacteria.

4.3 Molecular Characterization

In order to capture the genomic content of the most abundant organisms in Siders Pond, we will initially carry out deep metagenomic sequencing of the microbial community at 8 depths with the Illumina HiSeq platform. The metagenome will be characterized at only the first time point in the series as we do not expect population structure will change significantly over the two day period. We will analyze only 4 samples per HiSeq lane using a partially-overlapping paired-end strategy, which should provide ~1100 genome equivalents at each depth. This initial metagenome will be assembled into contigs and annotated to identify rRNA genes to assess population structure. The remaining reads will be annotated to identify functional marker genes that can be tracked in the transcript pool in subsequent samples. In this initial sequencing effort, we will also produce a metatranscriptome for each depth to directly compare genomic potential and gene expression patterns of microbial communities at each depth. These results will be critical to developing the model (see below). In all subsequent time points collected over the 2 day period, only metatranscriptomic analyses will be carried out to examine how gene expression patterns change over time and space. A metatranscriptomic approach that includes an internal standard will be used to allow for quantitative comparison across datasets (Moran et al. 2013).

4.4 Comparison of Model Output to Observations

The MEP model output includes concentration of biogeochemical constituents over depth and time that can be directly compared to Siders Pond observations collected over the two day period. The model will also predict the concentration of biological structure, $\mathbb{S}_i$, for each functional group in the metabolic network (**Table 1**) over depth and time as well. Since biological structure represents the amount of catalyst allocated to a functional pathway, it is directly proportional to gene abundance associated with the same function, so we will be able to assess model performance relative to observations in a manner similar to Reed et al. (2014), as illustrated in **Fig. 8**. For example, abundance of dissimilatory sulfite

reductase (*dsr*) gene should be linearly proportional to the concentration of $\mathbb{S}_{12}$ associated with the sulfate reduction reaction (**Table 1**). We expect the metatranscriptome to be more dynamic over the diel cycle (Vila-Costa et al. 2013). The components of the MEP model that closely relate to relative transcript abundances are the reaction control variables, ε_i, and reaction rates, r_i. As ε_i approaches 1, reaction rate, r_i Eq. (3), is forced to 0, regardless of the amount of $\mathbb{S}_i$ allocated to the reaction. Consequently, both reaction rate and ε_i should be linearly proportional to transcript abundance and can be compare in a manner similar to that for gene abundance (**Fig. 8**).

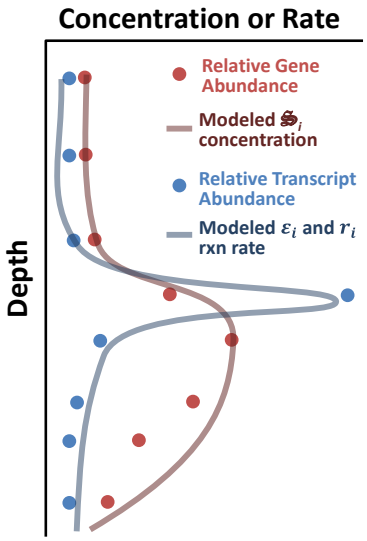

Fig. 8. Example of how model outputs will be compared to metagenomic and metatranscriptomic data. This example shows a "good fit" between model and observations.

We will compare model output to observations for different values of the optimization time interval, Δt, that determines the duration of the temporal window over which entropy is maximized, (i.e., Eqs. 7 and 8) . If we find that a good fit between model and observations can only be attained for large values of Δt, then this indicates that the community (or members of it) is implementing long duration temporal strategies (e.g., circadian rhythm). Similarly, we will use model fit to observations to determine if the community is best represented by local or global entropy maximization (Section 3.6), where a good fit to the global solution would indicate some level of spatial coordination of the community, such as via stigmergy (Gloag et al. 2013), chemotaxis (Stocker and Seymour 2012), diel vertical migration of phytoplankton (Inoue and Iseri 2012) or zooplankton (Steinberg et al. 2002, Haupt et al. 2010) or even communication by vesicle release (Biller et al. 2014). Our primary measure of success in this project will be the degree to which chemical and genomic observations can be captured by our MEP model using relatively few adjustable parameters, such as we were able to achieve in modeling methanotrophic communities (Vallino et al. 2014).

5. Proposed Model Development

5.1 Metabolic Reaction Network

The initial set of reactions we plan to include in the metabolic network are shown in **Table 1**, which is based on our current understanding of the dominate biogeochemical processes occurring in Siders Pond; however, this set will be updated based on findings obtained from our metagenomic survey. Decomposition of detrital C, N and P (C_D, N_D, P_D) are represented by first order kinetics (Rxns 18-20 in Table 1) because rates are not determined solely by extracellular enzyme concentrations (Dungait et al. 2012). Abiotic reactions involving iron are also included in the network (Rxns 16 & 17 in Table 1). In Section 3.2, we discussed how chemotrophic reactions are represented, but we have not considered phototrophic reactions in the MEP context yet. Part of the proposed work will be to expand the MEP model to include phototrophic-based reactions, as outlined below.

Table 1. Initial list of reactions that will be included in the metabolic network used to model biogeochemistry in Siders Pond. To reduce clutter, all biological structures below have the same elemental composition of $CH_\alpha O_\beta N_\gamma P_\delta$, and stoichiometric details needed to balance H and O are not shown, nor have we included how we account for variable substrate preferences (i.e, NH_3 vs HNO_3) in subreactions (but see Vallino et al. 2014).

Rxn	Metabolic Network Reactions
	Catalyzed Aerobic Reactions
1	$\varepsilon_1 H_2CO_3 + \varepsilon_1\gamma NH_3 + \varepsilon_1\delta H_3PO_4 + n_{\gamma_1} h\nu_H N_A \xrightarrow{\mathbb{S}_1} \varepsilon_1\mathbb{S}_1 + \varepsilon_1 O_2 + n_{\gamma_1} h\nu_L N_A$

2	$\sum_i \pi_i \mathbb{S}_i + (1-\varepsilon_2)\mathrm{O_2} \overset{\mathbb{S}_2}{\rightarrow} \varepsilon_2 \mathbb{S}_2 + (1-\varepsilon_2)\left(\varepsilon_2 \mathrm{H_2CO_3} + (1-\varepsilon_2)\mathrm{C}_D + \gamma(\varepsilon_2 \mathrm{NH_3} + (1-\varepsilon_2)\mathrm{N}_D) + \delta(\varepsilon_2 \mathrm{H_3PO_4} + (1-\varepsilon_2)\mathrm{P}_D)\right)$
3	$\mathrm{CH_2O} + \varepsilon_3\gamma\mathrm{NH_3} + \varepsilon_3\delta\mathrm{H_3PO_4} + (1-\varepsilon_3)\mathrm{O_2} \overset{\mathbb{S}_3}{\rightarrow} \varepsilon_3\mathbb{S}_3 + (1-\varepsilon_3)\mathrm{H_2CO_3}$
4	$\varepsilon_4\mathrm{H_2CO_3} + \left(1-\varepsilon_4(1-\gamma)\right)\mathrm{NH_3} + \frac{3}{2}(1-\varepsilon_4)\mathrm{O_2} + \varepsilon_4\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_4}{\rightarrow} \varepsilon_4\mathbb{S}_4 + (1-\varepsilon_4)(\mathrm{HNO_2} + \mathrm{H_2O})$
5	$\varepsilon_5\mathrm{H_2CO_3} + \left(1-\varepsilon_5(1-\gamma)\right)\mathrm{HNO_2} + \frac{1}{2}(1-\varepsilon_5)\mathrm{O_2} + \varepsilon_5\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_5}{\rightarrow} \varepsilon_5\mathbb{S}_5 + (1-\varepsilon_5)\mathrm{HNO_3}$
6	$(1-\varepsilon_6)(\mathrm{H_2S} + 2\mathrm{O_2}) + \varepsilon_6(\mathrm{H_2CO_3} + \gamma\mathrm{NH_3} + \delta\mathrm{H_3PO_4}) \overset{\mathbb{S}_6}{\rightarrow} \varepsilon_6\mathbb{S}_6 + (1-\varepsilon_6)\mathrm{H_2SO_4}$
7	$(1-\varepsilon_7)\left(\mathrm{Fe^{2+}} + \frac{1}{4}\mathrm{O_2} + \frac{5}{2}\mathrm{H_2O}\right) + \varepsilon_7(\mathrm{H_2CO_3} + \gamma\mathrm{NH_3} + \delta\mathrm{H_3PO_4}) \overset{\mathbb{S}_7}{\rightarrow} \varepsilon_7\mathbb{S}_7 + (1-\varepsilon_7)\mathrm{Fe(OH)_3}$
	Catalyzed Anaerobic Reactions
8	$\mathrm{CH_2O} + (\varepsilon_8\gamma + 2(1-\varepsilon_8))\mathrm{HNO_3} + \varepsilon_8\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_8}{\rightarrow} \varepsilon_8\mathbb{S}_8 + 2(1-\varepsilon_8)\mathrm{HNO_2} + (1-\varepsilon_8)\mathrm{H_2CO_3}$
9	$\mathrm{CH_2O} + \left(\varepsilon_9\gamma + \frac{4}{3}(1-\varepsilon_9)\right)\mathrm{HNO_2} + \varepsilon_9\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_9}{\rightarrow} \varepsilon_9\mathbb{S}_9 + \frac{2}{3}(1-\varepsilon_9)\mathrm{N_2} + (1-\varepsilon_9)\mathrm{H_2CO_3}$
10	$\mathrm{CH_2O} + \left(\varepsilon_{10}\gamma + \frac{2}{3}(1-\varepsilon_{10})\right)\mathrm{HNO_2} + \varepsilon_{10}\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_{10}}{\rightarrow} \varepsilon_{10}\mathbb{S}_{10} + \frac{2}{3}(1-\varepsilon_{10})\mathrm{NH_3} + (1-\varepsilon_{10})\mathrm{H_2CO_3}$
11	$\varepsilon_{11}\mathrm{H_2CO_3} + (\varepsilon_{11}\gamma + (1-\varepsilon_{11}))\mathrm{HNO_2} + (1-\varepsilon_{11})\mathrm{NH_3} + \varepsilon_{11}\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_{11}}{\rightarrow} \varepsilon_{11}\mathbb{S}_{11} + (1-\varepsilon_{11})\mathrm{N_2}$
12	$\mathrm{CH_2O} + \frac{1}{2}(1-\varepsilon_{12})\mathrm{H_2SO_4} + \varepsilon_{12}\gamma\mathrm{NH_3} + \varepsilon_{12}\delta\mathrm{H_3PO_4} \overset{\mathbb{S}_{12}}{\rightarrow} \varepsilon_{12}\mathbb{S}_{12} + \frac{1}{2}(1-\varepsilon_{12})\mathrm{H_2S} + (1-\varepsilon_{12})\mathrm{H_2CO_3}$
13	$\varepsilon_{13}\mathrm{H_2CO_3} + \varepsilon_{13}\gamma\mathrm{NH_3} + \varepsilon_{13}\delta\mathrm{H_3PO_4} + ½\varepsilon_{13}\mathrm{H_2S} + n_{\gamma_{13}}h\nu_H N_A \overset{\mathbb{S}_{13}}{\rightarrow} \varepsilon_{13}\mathbb{S}_{13} + ½\varepsilon_{13}\mathrm{H_2SO_4} + n_{\gamma_{13}}h\nu_L N_A$
14	$\varepsilon_{14}\mathrm{CH_2O} + \varepsilon_{14}\gamma\mathrm{NH_3} + \varepsilon_{14}\delta\mathrm{H_3PO_4} + n_{\gamma_{14}}h\nu_H N_A \overset{\mathbb{S}_{14}}{\rightarrow} \varepsilon_{14}\mathbb{S}_{14} + n_{\gamma_{14}}h\nu_L N_A$
15	$\mathrm{CH_2O} + \varepsilon_{15}\gamma\mathrm{NH_3} + \varepsilon_{15}\delta\mathrm{H_3PO_4} + 4(1-\varepsilon_{15})\mathrm{Fe^{3+}} \overset{\mathbb{S}_{15}}{\rightarrow} \varepsilon_{15}\mathbb{S}_{15} + 4(1-\varepsilon_{15})(\mathrm{Fe^{2+}} + \mathrm{H_2CO_3})$
	Abiotic Reactions (uncatalyzed)
16	$\mathrm{Fe^{3+}} + \mathrm{H_3PO_4} \longrightarrow \mathrm{FePO_4}$
17	$\mathrm{Fe^{2+}} + \frac{1}{4}\mathrm{O_2} + \frac{5}{2}\mathrm{H_2O} \longrightarrow \mathrm{Fe(OH)_3}$
18	$\mathrm{C_D} \longrightarrow \mathrm{CH_2O}$
19	$\mathrm{N_D} \longrightarrow \mathrm{NH_3}$
20	$\mathrm{P_D} \longrightarrow \mathrm{H_3PO_4}$

As described previously, the phototrophic reactions will be built around optimal control variables, ε_i, that dictate whether a reaction leads to more biological structure synthesis (autocatalysis), or simply dissipates free energy as heat (entropy). We only focus on oxygenic photoautotrophs here, as the approach is readily extended to anoxygenic photoautotrophs and photoheterotrophs (see **Table 1**). Oxygenic photoautotrophs can be represented as,

$$\varepsilon_1\mathrm{H_2CO_3} + \varepsilon_1\gamma\mathrm{NH_3} + \varepsilon_1\delta\mathrm{H_3PO_4} + n_{\gamma_1}h\nu_H N_A \overset{\mathbb{S}_1}{\rightarrow} \varepsilon_1\mathbb{S}_1 + \varepsilon_1\mathrm{O_2} + n_{\gamma_1}h\nu_L N_A \tag{9}$$

where $\mathbb{S}_i$ has an elemental composition of $\mathrm{CH_\alpha O_\beta N_\gamma P_\delta}$. In phototrophic reactions high frequency light, ν_H, is converted to low frequency light, ν_L, as a function of intercepted photons captured by phototrophs ($I_{\mathbb{S}_i}$, see below), and h and N_A are Planck's constant and Avogadro's number, respectively. The parameter n_{γ_1} (mmol-photon mmol-rxn^{-1}) is chosen such that as ε_1 approaches 1, 100% of the light energy is transferred to chemical potential, so that the overall reaction free energy equals 0. Of course, this means the reaction proceeds reversibly and infinitely slowly. At the other extreme, when ε_1 approaches 0, all light energy is dissipated as heat and no biological structure is produced.

Light intensity (mmol photons m^{-2} d^{-1}) as a function of water depth, $I(z)$, can be determined by solving the following standard light equation,

$$\frac{dI(z)}{dz} = -\left(k_w + k_{\mathbb{S}_1} c_{\mathbb{S}_1} + k_{\mathbb{S}_2} c_{\mathbb{S}_2} + \cdots\right) I(z) \tag{10}$$

where k_w and $k_{\mathbb{S}_i}$ are light attenuations due to water and biological structure $\mathbb{S}_i$, respectively (detritus and sediment can also be added as needed). In the MEP formulation, any interception of light leads to entropy production if it is not converted to chemical potential, so silt laden water dissipates light energy as effectively as phytoplankton. Consequently, increasing entropy production can be achieved via the formation of particles when none are present, which is effectively what phytoplankton achieve in the MEP context. To formulate a MEP model for phototrophs, we are interested in the light intercepted by each biological structure at a given depth z_i, designated as $I_{\mathbb{S}_i}(z_i)$. For a well-mixed system of depth d (or a differential layer of thickness d), the light captured *only* by biological structure $\mathbb{S}_i$ is given by,

$$I_{\mathbb{S}_i}(d) = I(0) e^{-k_w d} \sum_{j \neq i} e^{-k_{\mathbb{S}_j} c_{\mathbb{S}_j} d} \left(1 - e^{-k_{\mathbb{S}_i} c_{\mathbb{S}_i} d}\right) \tag{11}$$

Hence, by increasing the concentration of $c_{\mathbb{S}_i}$, the amount of light intercepted by $\mathbb{S}_i$ and potentially dissipated as heat increases, but the function does saturate at high $\mathbb{S}_i$ concentrations as required.

Phototrophic reaction rate is modeled similarly to Eq. (3), but with the following modification,

$$r_i = \frac{I_{\mathbb{S}_i}(d)}{n_{\gamma_i} d} \varepsilon_i^2 \prod_j \left(\frac{c_j}{c_j + \kappa^* \varepsilon_i^4}\right) F_T\left(\Delta_{r_i} G\right), \tag{12}$$

where c_j is the concentration (μM or mmol m^{-3}) of substrates (H_2CO_3, NH_3 and H_2S for example). In Eq. (12), light capture rate, $I_{\mathbb{S}_i}(d)$, governs the maximum reaction rate, which in turn depends on $c_{\mathbb{S}_i}$ as given by Eq. (11). The term $F_T\left(\Delta_{r_i} G\right)$ is the thermodynamic force (Jin et al. 2013) that limits reaction kinetics as Gibbs free energy of reaction, $\Delta_{r_i} G$, approaches 0. In our current kinetic expression, Eq. (3), we have approximated this term with $\left(1 - \varepsilon_i^2\right)$, but a more formal expression for $F_T\left(\Delta_{r_i} G\right)$ will be investigated as part of our model development.

A variation of Eq. (5) summed over all reactions in the network is used to calculate internal entropy production, but the free energy of reaction for the phototrophic reactions is given by,

$$\Delta_{r_j} G = \varepsilon_j \Delta_{r_{\mathbb{S}}} G - \eta_I n_{\gamma_j} h \nu_S N_A, \tag{13}$$

where $\Delta_{r_{\mathbb{S}}} G$ is the Gibbs free energy of reaction for synthesis of $\mathbb{S}_i$ from H_2CO_3 and other substrates, and η_I is the thermodynamic efficiency for converting electromagnetic radiation into chemical potential, as given by Candau (2003),

$$\eta_I = 1 - \frac{4}{3}\frac{T_L}{T_H} + \frac{1}{3}\left(\frac{T_L}{T_H}\right)^4, \tag{14}$$

where T_H and T_L are the black body temperatures (K) associated with the high and low frequency light, respectively, driving the phototrophic reactions. For solar radiation, $T_H \cong 5780\ K$ and $T_L \cong 293\ K$, so that $\eta_I \cong 93\%$.

5.2 Vertical 1D Transport Model and Computation

We will use a lake 1D advection-dispersion-reaction (ADR) model (e.g., Riley and Stefan 1988) to simulate vertical transport in Siders Pond for the mass balance modeling component, e.g., Eq. (6). This model will be similar to our longitudinal ADR model used to simulate estuarine oxygen dynamics (Vallino et al. 2005), but we will use BACOLR (Wang et al. 2008) to solve the resulting PDE that employs a high-order adaptive collocation numerical algorithm, which we have found to be robust for

such problems. To improve computational performance, a second, lower resolution, computational grid will be used for the solution of the MEP optimal control problem to determine how control parameters, $\varepsilon_i(x,t)$, will vary over space and time. This is a computationally demanding problem, but recently we have developed a promising approach that uses large-scale bound constrained optimization (Morales and Nocedal 2011) combined with a numerical derivative estimator (More and Wild 2012) implemented for parallel architectures using MPI. Data from Caraco (1986) and students in the MBL SES program will be used to calibrate the transport model.

6. Methods

6.1 Nutrients

The Ecosystems Center, MBL, has considerable experience and equipment for field sampling and measuring environmental constituents. Water samples from Siders Pond will be drawn over depth via a peristaltic pump with an inline YSI water quality sonde (pH, T, DO and conductivity). Concentration of the following nutrients will be measured: NO_2^- and NO_3^- (via Lachat QuikChem 8000 autoanalyzer); NH_4^+ (Solorzano 1969); PO_4^{3+} (Murphy and Riley 1962); SO_4^{2-} by ion chromatograph (DX-120 Dionex); H_2S spectrophotometrically (Gilboa-Garber 1971); dissolved O_2 (by electrode); dissolved inorganic carbon (DIC) (via UIC Coulometrics (Johnson et al. 1993) or GC); dissolved iron by ferrozine (Stookey 1970); particulate organic carbon (POC) and nitrogen (PON) (on Perkin Elmer 2400 CHN elemental analyzer); Chl a and BChl cde by extracted Chl fluorescence (Caraco and Puccoon 1986). DAPI counts of bacteria and protists (Porter and Feig 1980); PAR with a LI-COR LI-193 spherical quantum sensor. Autotrophy and respiration rates will be determined by *in situ* light-dark bottle incubations at two depths with $H^{13}CO_3^-$ replacing $H^{14}CO_3^-$ (Steemann-Nielsen 1951).

6.2 Metagenomics and Metatranscriptomics

For both DNA and RNA analyses, a peristaltic pump will be used to filter ~500 ml of water through a 0.22 μm Sterivex cartridge (Millipore) with a 100 μm pre-filter. Filters will be placed on ice, flooded with RNALater, and stored at -80 ºC approximately 18-24 hours after collection. DNA will be extracted according to Huber et al. (2007), quantified with PicoGreen, and metagenomic libraries will be constructed using the Nugen Ovation Ultra-low Library preparation kit (NuGEN Technologies). We will build ~170 bp libraries and use paired-end sequencing reads to get high-quality, slightly overlapping (by 30bp) reads. Four metagenomic samples will be run per lane of the Illumina HiSeq 1000. Before RNA extraction, a known concentration of in vitro transcribed standard will be added to filter unit according to Gifford et al. (2011). RNA will be simultaneously extracted from both the Pond community and the internal standard using the mirVana miRNA Isolation Kit according to Shi et al. (2009). Metatranscriptomic libraries will be constructed using the Encore Complete Prokaryotic RNA-Seq DR Multiplex System (NuGEN Technologies), which carries out mRNA enrichment and cDNA synthesis, as well as size selection and multiplexed library construction for sequencing on the Illumina HiSeq platform. Like the metagenomes, we will build ~170 bp libraries and use paired-end sequencing reads to get high-quality, slightly overlapping reads. Eight samples will be run per lane of the Illumina HiSeq 1000.

After paired-end reads are merged and quality control is carried out, the reads will be dereplicated to remove sequences with 100% identity. Ribosomal RNA (rRNA) sequences will be identified using riboPicker software (Schmieder et al. 2011) and classified taxonomically using the ARB SILVA LSU and SSU databases (http://www.arb-silva.de). Reads will be assembled using MetaVelvet (Namiki et al. 2012). For metatranscriptomes, the read coverage of the internal standard will be assessed by mapping assembled reads to the reference sequence, and abundance of cDNA reads will be determined by mapping the reads back to the assembled transcript fragments using Bowtie (Langmead et al. 2009). Gene annotations of non rRNA assembled contigs will be done through the Integrated Microbial Genomes and Metagenomics (IMG/M) system (Markowitz et al. 2008). Transcript reads will be mapped to predicted proteins using BLASTN and the internal standard will be used to normalize reads and allow for quantitative comparison across datasets (Gifford et al. 2011, Moran et al. 2013). All sequencing and

bioinformatics analyses will take place at the Josephine Bay Paul Center Keck Sequencing Facility at the MBL and data will be deposited into NCBI Short Read Archive.

7. Broader Impacts

With the projected increases in global temperature, atmospheric CO_2 (and associated decrease in ocean pH), and nitrogen loading to coastal oceans and terrestrial ecosystems, there is a growing need to understand how biogeochemical processes that are largely governed by microbes will respond to these changes. However, marine biogeochemical models have changed little over the last several decades (Rose 2012) due to the lack of understanding of the fundamental principles that govern microbial processes at the systems level. Results from our project have the potential to change the organismal centric paradigm to one which is based on free energy dissipation. Models based on fundamental principles can more accurately extrapolate beyond their calibration data set than conventional models. While many models can interpolate within observations, it is the ability to extrapolate predictions that is critically needed for society to understand how life supporting microbial systems are likely to respond to projected changes. We note, however, that entropy and MEP-based concepts and our mathematical modeling thereof can be abstruse to those outside of the field; consequently, one component of our broader impacts will be to also publish manuscripts that are more accessible to the broader scientific community, such as a manuscript we recently submitted (Chapman et al. *submitted*).

We will work with the undergraduate programs at MBL to train new students at the interface of biogeochemical modeling, molecular microbiology, and microbial biochemistry. We will do this through both undergraduate classroom teaching and research projects and internships. First, PI's Vallino and Huber are faculty in the Semester in Environmental Science (SES) program at MBL (http://courses.mbl.edu/SES), which annually draws up to 24 juniors and seniors from over 60 colleges and universities around the country. Classes average 84% women and several minority colleges and universities participate in the SES program. Greater than 40% of SES students go on to graduate school. A unique aspect of the SES program is the hands-on independent research project conducted during the last six weeks of the semester. SES projects are strongly coupled to existing PI research, and support of this proposal would provide the scaffold enabling independent research focused on understanding microbial dynamics and biogeochemistry, use of molecular techniques to measure species diversity and gene expression, and use of thermodynamics and models to predict biogeochemistry. SES students give a public presentation to the Woods Hole community and write "journal ready" manuscripts regarding their research that are published on the SES web site. PI's Vallino and Huber also co-teach the SES elective Methods in Microbial Ecology that uses a module-based approach. We will add a new laboratory module focusing on anoxygenic photosynthesis, such as differentiating bacterial Chl a from Chl a. In each of the three project years, we will mentor from two to four SES independent student research projects associated with our proposed research.

Second, we will seek undergraduate participation through the Woods Hole Partnership Educational Program (PEP) (http://www.woodsholediversity.org/pep/). PEP is specifically targeted to underrepresented groups in science, and while the undergraduates are hosted at individual institutions (e.g WHOI, MBL), they are treated as a "class" of students across the various participating institutions, allowing for exposure to multiple disciplines and approaches to research-based science. PEP students present their research results to the Woods Hole community at a one day PEP symposium in mid-August. The PIs in this project will train summer students in molecular techniques for assessing microbial communities and their role in the environment and/or use of models for understanding microbial processes.

Finally, we will co-mentor a postdoctoral scholar. The postdoc will have a unique opportunity to engage scientists from both the Ecosystems Center and the Bay Paul Center for Comparative Molecular Biology and Evolution. The postdoc will actively be involved in the development of proposals, and will be allowed to gain teaching experience via our SES Methods course. In addition, the postdoc will participate

in our weekly "Micro-Eco" Discussion Group, which facilitates cross-communication and collaboration between scientists engaged in microbial ecology research at our research centers and institutions in Woods Hole (see Mentoring supplementary material).

8. Management

8.1 Timeline

In year 1 of the project, samples from Siders Pond will be collected over a two day period in July or August, nutrients will be analyzed, metagenomes and transcriptomes will be sequenced for the first of the eight depth profiles, and a bioinformatics pipeline for functional gene quantification will be developed. In year 2, the remaining 7 metatranscriptomes will be sequenced and quantified for functional gene abundance. In year 3, observations will be compare to model predictions. Model development and testing will occur in all project years and includes: 1) expanding the metabolic network to include phototrophs; 2) development of a 1D transport model for Siders Pond; 3) implementation of parallel code using MPI to solve the spatiotemporal optimal control problem; 4) model simulation runs for comparison to observations from Siders Pond.

8.2 Responsibilities

PI Vallino will manage the overall project as well as develop the MEP-based model. PI Vallino will also be resposnible for integration between model output and experimental observations and will supervise nutrient analyses conducted by the part-time RA. Co-PI Huber will manage all metagenomic and metatranscriptomic surveys. The postdoctoral scholar (TBD) will be responsible for sampling Siders Pond with assistance of the RA, as well as DNA and RNA sequencing and bioinformatics anayses. Both Vallino and Huber will server as mentors to the postdoctoral scholar in their perspective areas of expertise, namely biogeochemistry and thermodynamic modeling (Vallino) and molecular microbial ecology (Huber).

9. Results of Prior NSF Funded Research

Theory: Biological systems organize to maximize entropy production subject to information and biophysicochemical constraints. EF-0928742, 9/2009-8/2013: $750,000. PIs: **Vallino** and **Huber**. This project examined the hypothesis that biological systems evolve and organize in a manner that results in maximum entropy production (MEP). One of the project's main hypotheses is that living systems differ from abiotic systems, such as fire, by integrating entropy production over time using information stored in the organismal metagenome. Experimental methanotrophic microcosms have been in operation for more than 1400 days and 454 pyrosequencing data from 19 different time points ($> 10^6$ total sequence reads) shows bacterial diversity in the microcosms is very dynamic despite constant environmental conditions and includes > 50 species of methylotrophs plus methanotrophs. A MEP-based model developed during the project has been able to simulate observations using only two adjustable parameters. Model results indicate the communities are inherently well adapted to handling cyclic energy inputs up to periods of at least 20 days. While project data are still being analyzed, experimental and modeling results to date have been presented at 7 international conferences and numerous departmental seminars, four papers have been published (Vallino 2010, Vallino 2011, Algar and Vallino 2014, Vallino et al. 2014), one submitted (Chapman et al. *submitted*) and one soon-to-be submitted (Fernandez Gonzalez et al. *to be submitted*). *Broader Impacts*: The project has involved 7 undergraduate research projects and supported a postdoctoral scholar. Microcosm data are still being streamed to the project web site (http://ecosystems.mbl.edu/MEP) in near real time.

References

Alberty, R. A. 2003. *Thermodynamics of biochemical reactions*. Wiley & Sons, Hoboken, NJ.

Algar, C. K., and J. J. Vallino. 2014. Predicting microbial nitrate reduction pathways in coastal sediments. *Aquat.Microb.Ecol.* **71**:223-238, doi:10.3354/ame01678.

Babbin, A. R., R. G. Keil, A. H. Devol, and B. B. Ward. 2014. Organic Matter Stoichiometry, Flux, and Oxygen Control Nitrogen Loss in the Ocean. *Science* **344**:406-408, doi:10.1126/science.1248364.

Biller, S. J., F. Schubotz, S. E. Roggensack, A. W. Thompson, R. E. Summons, and S. W. Chisholm. 2014. Bacterial Vesicles in Marine Ecosystems. *Science* **343**:183-186, doi:10.1126/science.1243457.

Burgin, A. J., and S. K. Hamilton. 2007. Have we overemphasized the role of denitrification in aquatic ecosystems? A review of nitrate removal pathways. *Frontiers in Ecology and the Environment* **5**:89-96, doi:10.1890/1540-9295(2007)5[89:HWOTRO]2.0.CO;2.

Cajal-Medrano, R., and H. Maske. 1999. Growth efficiency, growth rate and the remineralization of organic substrate by bacterioplankton--revisiting the Pirt model. *Aquatic Microbial Ecology* **19**:119-128, doi:10.3354/ame019119.

Candau, Y. 2003. On the exergy of radiation. *Solar Energy* **75**:241-247, doi:10.1016/j.solener.2003.07.012.

Caraco, N. 1986. Phosphorus, iron, and carbon cycling in a salt stratified coastal pond. Boston University, Boston.

Caraco, N., and A. H. Puccoon. 1986. The measurement of bacterial chlorophyll and algalchlorophyll a in natural waters. *Limnology and Oceanography* **31**:889-893.

Caraco, N., A. Tamse, O. Boutros, and I. Valiela. 1987. Nutrient limitation of phytoplankton growth in brackish coastal ponds. *Canadian Journal of Fisheries and Aquatic Sciences* **44**:473-476, doi:10.1139/f87-056.

Chapman, E. J., D. L. Childers, and J. J. Vallino. *submitted*. A review of how The Second Law of Thermodynamics has informed ecosystem ecology through its history. *Ecosystems*.

De Filippo, C., M. Ramazzotti, P. Fontana, and D. Cavalieri. 2012. Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Briefings in Bioinformatics* **13**:696-710, doi:10.1093/bib/bbs070.

Decho, A. W., R. S. Norman, and P. T. Visscher. 2010. Quorum sensing in natural environments: emerging views from microbial mats. *Trends in Microbiology* **18**:73-80, doi:10.1016/j.tim.2009.12.008.

Dewar, R. 2003. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal of Physics A: Mathematical and General* **36**:631-641.

Dewar, R. C. 2005. Maximum entropy production and the fluctuation theorem. *Journal of Physics A: Mathematical and General* **38**:L371-L381, doi:10.1088/0305-4470/38/21/L01.

Dewar, R. C., C. H. Lineweaver, R. K. Niven, and K. Regenauer-Lieb, editors. 2014. Beyond the second law: Entropy production and non-equilibrium systems. Springer-Verlag, Berlin, doi:10.1007/978-3-642-40154-1.

Dungait, J. A. J., D. W. Hopkins, A. S. Gregory, and A. P. Whitmore. 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biology* **18**:1781-1796, doi:10.1111/j.1365-2486.2012.02665.x.

Falkowski, P. G., T. Fenchel, and E. F. DeLong. 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**:1034-1039, doi:10.1126/science.1153213.

Fernandez, A., S. Huang, S. Seston, J. Xing, R. Hickey, C. Criddle, and J. Tiedje. 1999. How Stable Is Stable? Function versus Community Composition. *Applied and Environmental Microbiology* **65**:3697-3704.

Fernandez Gonzalez, N., J. A. Huber, and J. J. Vallino. *to be submitted*. Microbial community dynamics under periodic perturbations. *ISME J*.

Gifford, S. M., S. Sharma, J. M. Rinta-Kanto, and M. A. Moran. 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* **5**:461-472, doi:10.1038/ismej.2010.141.

Gilboa-Garber, N. 1971. Direct spectrophotometric determination of inorganic sulfide in biological materials and in other complex mixtures. *Analytical Biochemistry* **43**:129-133, doi:10.1016/0003-2697(71)90116-3.

Gloag, E. S., M. A. Javed, H. Wang, M. L. Gee, S. A. Wade, L. Turnbull, and C. B. Whitchurch. 2013. Stigmergy: A key driver of self-organization in bacterial biofilms. *Communicative & Integrative Biology* **6**:e27331.

Goo, E., C. D. Majerczyk, J. H. An, J. R. Chandler, Y.-S. Seo, H. Ham, J. Y. Lim, H. Kim, B. Lee, M. S. Jang, E. P. Greenberg, and I. Hwang. 2012. Bacterial quorum sensing, cooperativity, and anticipation of stationary-phase stress. *Proceedings of the National Academy of Sciences* **109**:19775-19780, doi:10.1073/pnas.1218092109.

Haupt, F., M. Stockenreiter, E. S. Reichwaldt, M. Baumgartner, W. Lampert, M. Boersma, and H. Stibor. 2010. Upward phosphorus transport by Daphnia diel vertical migration. *Limnology and Oceanography* **55**:529-534, doi:10.4319/lo.2009.55.2.0529.

Hoehler, T. M. 2004. Biological energy requirements as quantitative boundary conditions for life in the subsurface. *Geobiology* **2**:205-215, doi:10.1111/j.1472-4677.2004.00033.x.

Huber, J. A., D. B. M. Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A. Butterfield, and M. L. Sogin. 2007. Microbial Population Structures in the Deep Marine Biosphere. *Science* **318**:97-100, doi:10.1126/science.1146689.

Inoue, T., and Y. Iseri. 2012. Diel vertical migration and nutrient transport of the dinoflagellate Peridinium bipes f. occultatum in a thermally stratified reservoir. *Water Science and Technology* **66**:1212-1219, doi:10.2166/wst.2012.302.

Jin, Q., and C. M. Bethke. 2003. A New Rate Law Describing Microbial Respiration. *Applied and Environmental Microbiology* **69**:2340-2348.

Jin, Q. S., E. E. Roden, and J. R. Giska. 2013. Geomicrobial Kinetics: Extrapolating Laboratory Studies to Natural Environments. *Geomicrobiology Journal* **30**:173-185, doi:10.1080/01490451.2011.653084.

Johnson, K. M., K. D. Wills, B. D. Butler, W. K. Johnson, and C. S. Wong. 1993. Coulometric total carbon dioxide analysis for marine studies: Maximizing the performance of an automated gas extraction system and coulometric detector. *Mar.Chem.* **44**:167-187.

Kleidon, A., K. Fraedrich, T. Kunz, and F. Lunkeit. 2003. The atmospheric circulation and states of maximum entropy production. *Geophys.Res.Lett.* **30**:1-4, doi:10.1029/2003GL018363.

Klitgord, N., and D. Segre. 2011. Ecosystems biology of microbial metabolism. *Current Opinion in Biotechnology* **22**:541-546, doi:10.1016/j.copbio.2011.04.018.

Kondepudi, D., and I. Prigogine. 1998. *Modern thermodynamics: From heat engines to dissipative structures*. Wiley & Sons, New York.

Langmead, B., C. Trapnell, M. Pop, and S. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25, doi:10.1186/gb-2009-10-3-r25.

Le Quere, C., S. P. Harrison, I. Colin Prentice, E. T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L. Cotrim Da Cunha, R. Geider, X. Giraud, C. Klaas, K. E. Kohfeld, L. Legendre, M. Manizza, T. Platt, R. B. Rivkin, S. Sathyendranath, J. Uitz, A. J. Watson, and D. Wolf-Gladrow. 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* **11**:2016-2040, doi:10.1111/j.1365-2486.2005.1004.x.

Lineweaver, C. H., and C. A. Egan. 2008. Life, gravity and the second law of thermodynamics. *Physics of Life Reviews* **5**:225-242, doi:10.1016/j.plrev.2008.08.002.

Lorenz, R. 2003. COMPUTATIONAL MATHEMATICS: Full Steam Ahead-Probably. *Science* **299**:837-838.

Lorenz, R. D., J. I. Lunine, and P. G. Withers. 2001. Titan, Mars and Earth : Entropy production by latitudinal heat transport. *Geophysical Research Letters* **28**:415-418.

Lotka, A. J. 1922. Contribution to the Energetics of Evolution. *Proceedings of the National Academy of Sciences* **8**:147-151.

Lucia, U. 2012. Maximum or minimum entropy generation for open systems? *Physica A: Statistical Mechanics and its Applications* **391**:3392-3398, doi:10.1016/j.physa.2012.01.055.

Makela, T., and A. Annila. 2010. Natural patterns of energy dispersal. *Physics of Life Reviews* **7**:477-498, doi:10.1016/j.plrev.2010.10.001.

Markowitz, V. M., N. N. Ivanova, E. Szeto, K. Palaniappan, K. Chu, D. Dalevi, I. M. A. Chen, Y. Grechkin, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, P. Hugenholtz, and N. C. Kyrpides. 2008. IMG/M: a data management and analysis system for metagenomes. *Nucl. Acids Res.* **36**:D534-538, doi:10.1093/nar/gkm869.

Martyushev, L. 2007. Some interesting consequences of the maximum entropy production principle. *Journal of Experimental and Theoretical Physics* **104**:651-654, doi:10.1134/S1063776107040152.

Martyushev, L. 2013. Entropy and Entropy Production: Old Misconceptions and New Breakthroughs. *Entropy* **15**:1152-1170, doi:10.3390/e15041152.

Martyushev, L. M., and V. D. Seleznev. 2006. Maximum entropy production principle in physics, chemistry and biology. *Physics Reports* **426**:1-45.

McKeegan, K. D., A. B. Kudryavtsev, and J. W. Schopf. 2007. Raman and ion microscopic imagery of graphitic inclusions in apatite from older than 3830 Ma Akilia supracrustal rocks, west Greenland. *Geology* **35**:591-594, doi:10.1130/g23465a.1.

Mitri, S., and K. R. Foster. 2013. The Genotypic View of Social Interactions in Microbial Communities. *Annual Review of Genetics* **47**:247-273, doi:10.1146/annurev-genet-111212-133307.

Morales, J. L., and J. Nocedal. 2011. Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization". *ACM Trans. Math. Softw.* **38**:1-4, doi:10.1145/2049662.2049669.

Moran, M. A., B. Satinsky, S. M. Gifford, H. Luo, A. Rivers, L.-K. Chan, J. Meng, B. P. Durham, C. Shen, V. A. Varaljay, C. B. Smith, P. L. Yager, and B. M. Hopkinson. 2013. Sizing up metatranscriptomics. *ISME J* **7**:237-243, doi:10.1038/ismej.2012.94.

More, J. J., and S. M. Wild. 2012. Estimating Derivatives of Noisy Simulations. *ACM Trans. Math. Softw.* **38**:1-21, doi:10.1145/2168773.2168777.
Morowitz, H. J. 1968. *Energy flow in biology: biological organization as a problem in thermal physics*. Academic Press, New York.
Murphy, J., and J. P. Riley. 1962. A modified single solution method for the determination of phosphate in natural waters. *Anal.Chim.Acta* **27**:31-36.
Nalven, S. 2011. Diversity and distribution of sulfate-reducing bacteria in Siders Pond, a meromictic salt pond (http://ecosystems.mbl.edu/SES/2011%20Projects/Nalven,%20Sarah.pdf). MBL, Woods Hole, MA.
Namiki, T., T. Hachiya, H. Tanaka, and Y. Sakakibara. 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40**:e155, doi:10.1093/nar/gks678.
Niven, R. K. 2009. Steady state of a dissipative flow-controlled system and the maximum entropy production principle. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* **80**:021113-021115, doi:10.1103/PhysRevE.80.021113.
Odum, H. T., and R. C. Pinkerton. 1955. Time's speed regulator: the optimum efficiency for maximum power output in physical and biological systems. *Am.Sci.* **43**:321-343.
Paltridge, G. W. 1975. Global dynamics and climate-a system of minimum entropy exchange. *Q.J.Roy.Met.Soc.* **104**:927-945.
Porter, K. G., and Y. S. Feig. 1980. The use of DAPI for identifying and counting aquatic microflora. *Limnol.Oceanogr.* **25**:943-948.
Prigogine, I., G. Nicolis, and A. Babloyantz. 1972. Thermodynamics of evolution. *Physics Today* **25**:23-28.
Reed, D. C., C. K. Algar, J. A. Huber, and G. J. Dick. 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1313713111.
Riley, M. J., and H. G. Stefan. 1988. Minlake: A dynamic lake water quality simulation model. *Ecological Modelling* **43**:155-182, doi:10.1016/0304-3800(88)90002-6.
Rose, K. A. 2012. End-to-end models for marine ecosystems: Are we on the precipice of a significant advance or just putting lipstick on a pig? *Scientia Marina* **76**:195-201, doi:10.3989/scimar.03574.20B.
Schmieder, R., Y. W. Lim, and R. Edwards. 2011. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, doi:10.1093/bioinformatics/btr669.
Schrödinger, E. 1944. *What is life?* Cambridge University Press, Cambridge, UK.
Shi, Y., G. W. Tyson, and E. F. DeLong. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean/'s water column. *Nature* **459**:266-269, doi:10.1038/nature08055.
Shimokawa, S., and H. Ozawa. 2007. Thermodynamics of irreversible transitions in the oceanic general circulation. *Geophysical Research Letters* **34**:1-5, doi:10.1029/2007GL030208.
Solorzano, L. 1969. Determination of ammonia in natural waters by the phenolhypochlorite method. *L&O* **14**:799-801.
Steemann-Nielsen, E. 1951. Measurement of the Production of Organic Matter in the Sea by means of Carbon-14. *Nature* **167**:684-685, doi:10.1038/167684b0.
Steinberg, D. K., S. A. Goldthwait, and D. A. Hansell. 2002. Zooplankton vertical migration and the active transport of dissolved organic and inorganic nitrogen in the Sargasso Sea.

Deep-Sea Research Part I-Oceanographic Research Papers **49**:1445-1461, doi:10.1016/s0967-0637(02)00037-7.

Stocker, R., and J. R. Seymour. 2012. Ecology and Physics of Bacterial Chemotaxis in the Ocean. *Microbiology and Molecular Biology Reviews* **76**:792-812, doi:10.1128/mmbr.00029-12.

Stookey, L. L. 1970. Ferrozine---a new spectrophotometric reagent for iron. *Analytical Chemistry* **42**:779-781, doi:10.1021/ac60289a016.

Vallino, J. J. 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research* **58**:117-164.

Vallino, J. J. 2003. Modeling Microbial Consortiums as Distributed Metabolic Networks. *The Biological Bulletin* **204**:174-179.

Vallino, J. J. 2010. Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**:1417-1427, doi:10.1098/rstb.2009.0272.

Vallino, J. J. 2011. Differences and implications in biogeochemistry from maximizing entropy production locally versus globally. *Earth Syst.Dynam.* **2**:69-85, doi:10.5194/esd-2-69-2011.

Vallino, J. J., C. K. Algar, N. F. González, and J. A. Huber. 2014. Use of Receding Horizon Optimal Control to Solve MaxEP-Based Biogeochemistry Problems. Pages 337-359 *in* R. C. Dewar, C. H. Lineweaver, R. K. Niven, and K. Regenauer-Lieb, editors. Beyond the Second Law. Springer Berlin Heidelberg, doi:10.1007/978-3-642-40154-1_18.

Vallino, J. J., C. S. Hopkinson, and R. H. Garritt. 2005. Estimating estuarine gross production, community respiration and net ecosystem production: a nonlinear inverse technique. *Ecological Modelling* **187**:281-296, doi:10.1016/j.ecolmodel.2004.10.018.

Vila-Costa, M., S. Sharma, M. A. Moran, and E. O. Casamayor. 2013. Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environmental Microbiology* **15**:1190-1203, doi:10.1111/1462-2920.12033.

Wang, J., and R. L. Bras. 2011. A model of evapotranspiration based on the theory of maximum entropy production. *Water Resour.Res.* **47**:1-10, doi:10.1029/2010WR009392.

Wang, R., P. Keast, and P. H. Muir. 2008. Algorithm 874: BACOLR-spatial and temporal error control software for PDEs based on high-order adaptive collocation. *ACM Trans.Math.Softw.* **34**:1-28, doi:10.1145/1356052.1356056.

Ward, B. A., M. A. M. Friedrichs, T. R. Anderson, and A. Oschlies. 2010. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems* **81**:34-43, doi:10.1016/j.jmarsys.2009.12.005.

Wittebolle, L., H. Vervaeren, W. Verstraete, and N. Boon. 2008. Quantifying Community Dynamics of Nitrifiers in Functionally Stable Reactors. *Applied and Environmental Microbiology* **74**:286-293, doi:10.1128/AEM.01006-07.

Zomorrodi, A. R., M. M. Islam, and C. D. Maranas. 2014. d-OptCom: Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities. *ACS Synthetic Biology* **3**:247-257, doi:10.1021/sb4001307.