Surveys, simulation and single-cell assays relate function and phylogeny in a lake ecosystem

Supplementary Information for *Surveys, simulation, and single-cell assays relate function and phylogeny in a lake ecosystem*

Sarah P. Preheim, Scott W. Olesen, Sarah J. Spencer, Arne Materna, Charuleka Varadharajan, Matthew Blackburn, Jonathan Friedman, Jorge Rodríguez, Harold Hemond and Eric J. Alm

Contents:

- Supplementary Text
- Supplementary References
- Supplementary Tables S1–S6
- Supplementary Figures S1–S9

Part I
Biogeochemical model

1 Model scope

We treat the hypolimnion as a mostly closed system of cycling chemical species. Compartments are linked by transport processes, and reactions within compartments interconvert chemical species. Interaction with the environment outside the hypolimnion is represented by just two processes: input of oxidizable carbon in the upper compartments and input of methane from the sediment. The sediment and metalimnion are otherwise ignored.

The model also treats the processes in the lake as temporally symmetrical. In other words, the rules that govern the evolution of the lake's geochemistry are fixed through time. In a real lake, season-long trends like changes in sunlight and atmospheric temperature will affect properties of the lake, notably, the depth of the thermocline. The model presented here assumes that the size and behavior of the hypolimnion is fixed, which simplifies its construction but might limit its ability to recognize season-long trends that emerge on account of those altered properties.

1.1 The model only treats the lake's hypolimnion

In order to develop a model that captured as much of the lake's biogeochemical dynamics as possible while still remaining simple and conceptual, we developed a model that only treats the hypolimnion. We limited the model in this way because most of the lake's vertical distance is in the hypolimnion (17 out of 22 meters), because modeling the epilimnion presents very challenges from modeling the hypolimnion, and because modeling the two parts of the lake together is even more complicated.

- the epilimnion and hypolimnion are somewhat decoupled by the resistance to mixing across the thermocline
- a successful treatment of the two parts of the lake will require a model of the thermal properties of the lake
- the effects of wind and precipitation are much stronger in the epilimnion

- rivers move water into and out of the lake's epilimnion
- light penetrates in the epilimnion and delivers different amounts of energy to its different depths
- microbial growth is not overall limited by energy in the epilimnion like it is in the hypolimnion

2 Model

2.1 Chemical species

We intended to create a general model of the lake's seasonal chemical and biological dynamics, and we therefore simplified the number of chemical species used in the model. We aimed to simplify the model so that

- the relevant parameters and chemical species were better matched to the variations expected in dimictic lakes, rather than in aquifers,
- the dynamics were easier to interpret and match with biological measurements,
- the model avoided making predictions about chemical species that we did not or could not measure,
- the model had fewer parameters, but also so that
- the model remained a faithful representation of the key biogeochemical processes occurring the lake.

Table S1 lists all chemical species simulated. Relative to the original model (1), we made the following changes:

- We used a single carbon species. The original model allows for two types of carbon, DOC and POC. Each type of carbon can consist of up to four species, each with their own degradation kinetic constant. Although we do expect that there are many kinds of carbon in this ecosystem (e.g., particulate biomass, humics from rainwater runoff, photosynthetic algae), some simulations in the original publication use only one carbon species. We also found that a single carbon species was sufficient to reproduce the expected dynamics. To minimize the possibility for overfitting, we used only one.

- We simplified the sulfur compounds from five species to two be eliminating S^0 and FeS and combining HS^- and S^{2-}. We could not directly measure either of the eliminated species, and our measurements did not distinguish between the two reduced species. The process producing S^0 has a kinetic constant two orders of magnitude smaller than the kinetic constant for process that competes for H_2S, and S^0 is inert in the original model, so we expected its removal would not dramatically affect the dynamics. FeS, aside from being unmeasured, is produced in the original model by one of a set of non-redox mineral precipitation-dissolution reactions that we exclude for other reasons, as described below. The two reduced sulfur species interconvert in the original model in one of a set of acid dissociation reactions that we exclude for other reasons, as described below.

- We eliminated manganese compounds. Manganese, a terminal electron acceptor intermediate between nitrate and iron, is present at low abundances in Mystic Lake. We therefore expect it is unimportant to the dynamics.

- We eliminated calcium species. In the original model, calcium species participate in the precipitation-dissolution and acid dissociation reactions that we excluded, as described below. We also had no measurements for calcium.

- We simplified the nitrogen species by eliminating an adsorbed variety of ammonia, since we have measurements that distinguish between the two varieties and because we did not expect that the adsorption dynamics would be important to the biological dynamics. We also ignored the nitrogen gas produced by denitrification and iron oxidation on nitrate because nitrogen fixation is typically not a prominent process in the hypolimnion.

- We make the approximation the carbon dioxide is, for the purposes of methanogenesis, ubiquitous and abundant, as is generally true in eutrophic dimictic lakes.

In some cases, there is a clear correspondence between a single species in the model and a single species in nature (e.g., M for CH_4). In other cases, a single species in the model might stand for multiple species in nature (e.g., S^- for HS^- and S^{2-}).

3 Chemical reactions

As mentioned above, we also made simplifications to the chemical processes used in the original aquifer-specific model. The set of reactions used here in show in Table S3. Relative to the original model, we made the following changes:

- We added iron oxidation on nitrate. Previous research in this lake (2) had shown that this process is important to the lake's biogeochemistry.
- We eliminated the pH-dependent acid dissolution reactions because the pH in Mystic Lake's hypolimnion only varies between about 6 and 7.
- We also eliminated the non-redox mineral precipitation-dissolution reactions. Those reactions are important mostly because they are involved in acid-base buffering, which is important in groundwater systems but not in the biogeochemistry of Mystic Lake.
- We eliminated the adsorption reaction, as described above.
- We eliminated the reactions that included carbon dioxide and manganese because they were, as argued above, not important to the lake's biogeochemistry.

4 Caveats to these modifications

It is not our intent to assert that, because a feature is not included in the model (e.g., S^0 as an electron acceptor, sulfate reduction in aerobic zones) it is not happening in the lake or is not important in other ecosystems. We aimed instead to identify the few processes that were most critical to the development of the observed chemical gradients in the lake. We hope that the loss of interesting biological complexity is balanced by the increased clarity of a simpler model with fewer processes and parameters. For example, including POC in the model would involve creating five new chemical processes (one each for the primary oxidation half-reactions), one new transport process (since POC will have different transport dynamics than DOC), and the relevant parameters for all those processes.

symbol	name	representative compounds
O	dissolved oxygen	O_2
C	oxidizable carbon	cyanobacteria biomass, glucose, acetate
N^+	oxidized nitrogen	nitrate, nitrite
N^-	reduced nitrogen	ammonia
Fe^+	oxidized iron	Fe(III) compounds
Fe^-	reduced iron	Fe(II)
S^+	oxidized sulfur	sulfate compounds
S^-	reduced sulfur	sulfide compounds
M	methane	CH_4

Table S1: Chemical species included in the model.

5 Mechanics: Transport and reactions

The rate of change in the concentration of a chemical species X at a depth i is

$$\frac{\partial X_i}{\partial t} = (\text{transport terms}) + (\text{reaction terms}) + (\text{source terms}) \,, \qquad (1)$$

where i refers to depth in meters, i.e., low i means vertically higher in the water column. The simulation proceeds in N compartments, which we spaced at one meter to be comparable to the collected chemical and biological data. The initial concentrations are set and the simulation proceeds for a time T, during which the chemical species concentrations and reaction rates are recorded. This time roughly corresponds to the period between the movement of the thermocline up the water column in spring and the breakdown of stratification in fall.

5.1 Transport: Diffusion and settling

Most chemical species are treated as dissolved in the water column. In the time and length scales relevant to the hypolimnion ecosystem, molecular diffusion is slow compared to bulk transport processes like vertical eddy diffusion. To model these bulk transport processes, most chemical species are transported by simple diffusion with rate $D\,(X_{i-1} - X_i) + D\,(X_{i+1} - X_i)$, where the diffusion constant D is the same for all chemical species, since it

represents a bulk transport process. To account for the boundaries at the metalimnion and sediment, the first term is excluded in the uppermost simulation compartment; in the lowermost compartment, the second is excluded.

To simulate the settling of particulate carbon and oxidized iron species, C and Fe^+ settle in the model. A parameter p, where $0 < p < 1$, determines the balance between vertical eddy diffusion and settling for these chemical species so that the transport rate is

$$(1+p)D\left(X_{i-1} - X_i\right) + (1-p)D\left(X_{i+1} - X_i\right). \tag{2}$$

Since $p > 0$, these species tend to move down the water column and accumulate above the sediment. As with other species, the first term in excluded in the top compartment; the second term in the bottom compartment.

5.2 Reactions

5.2.1 Biotically-catalyzed reactions: Primary oxidations

The oxidation of carbon uses a chain of progressively less energetically-favorable terminal electron acceptors. Here, we follow the formulation laid out by Hunter *et al.* (ref. 1, especially equations 3 and 4).

The total rate of carbon degradation in a compartment follows first order kinetics:

$$R^{\mathrm{C}} \equiv k^{\mathrm{C}}\mathrm{C}; \quad \left(\frac{\partial \mathrm{C}}{\partial t}\right)_{\text{reaction}} = -R^{\mathrm{C}}, \tag{3}$$

where k^{C} is a first-order rate constant. The fraction of carbon taken up by oxidation on each of the terminal electron acceptors is determined by the abundance and relative metabolic merit of the electron acceptors. The j-th electron acceptor is consumed at a rate

$$R_j = \frac{f_j}{e_j} R^{\mathrm{C}}, \tag{4}$$

where e_j is the number of electrons neutralized per electron acceptor molecule and f_j is determined by successive applications of the formula

$$f_j = \left(1 - \sum_{k=1}^{j-1} f_k\right) \max\left\{1, \frac{[\mathrm{EA}_j]}{[\mathrm{EA}_{\mathrm{lim},j}]}\right\} \tag{5}$$

j	EA	e_j
1	O	4
2	N^+	5
3	Fe^+	1
4	S^+	8
5	$\varnothing$	8

Table S2: Electron acceptors in the primary oxidation reactions. $j = 5$ corresponds to methanogenesis.

for $j \in \{1, 2, 3, 4\}$. If the j-th electron acceptor's concentration $[\mathrm{EA}_j]$ is greater than some constant limiting concentration $[\mathrm{EA}_{\mathrm{lim},j}]$, then that electron acceptor gets all the remainder of the carbon; otherwise, it gets a fraction of what is left determined by the ratio of the two concentrations.

The electron acceptors and their e_j are listed in (Table S2). Methanogenesis corresponds to $j = 5$, and gets all remaining carbon so that $f_5 = 1 - \sum_{k=1}^{4} f_k$. All the carbon allocated by R^{C} gets used up (i.e., $\sum_{j=1}^{5} f_j = 1$), but each electron acceptors accepts electrons according to a different stoichiometry (i.e., $\sum_{j=1}^{5} R_j \neq R^{\mathrm{C}}$).

5.3 Secondary oxidations

We model secondary oxidations, the oxidation of compounds other than carbon compounds, using second-order mass action kinetics as per Hunter *et al.* (ref. 1's Table 4). For the transformation of substrates S_1, S_2 into a product P according to $a_1 S_1 + a_2 S_2 \to bP$, the reaction rate is $r \equiv k[S_1][S_2]$ and the reaction terms are

$$\left(\frac{\partial[P]}{\partial t}\right)_{\text{reaction}} = br \tag{6}$$

$$\left(\frac{\partial[S_i]}{\partial t}\right)_{\text{reaction}} = -a_i r \quad (i = 1, 2) \tag{7}$$

$$\tag{8}$$

with rate constant k. As per Hunter *et al.*, we do not adjust the rate according to the reaction's stoichiometry.

Primary and secondary oxidations are listed in Table S3.

Primary oxidations	rate	
$C \rightarrow aN^- + ee^-$	R^C	primary oxidation half-reaction
$O \rightarrow \varnothing$	R_1	aerobic heterotrophy
$N^+ \rightarrow \varnothing$	R_2	denitrification
$Fe^+ \rightarrow Fe^-$	R_3	iron reduction
$S^+ \rightarrow S^-$	R_4	sulfate reduction
$\varnothing \rightarrow M$	R_5	methanogenesis
Secondary oxidations	**rate constant**	
$2O + N^- \rightarrow N^+$	k_1	ammonia oxidation
$2O + S^- \rightarrow S^+$	k_2	sulfide oxidation
$N^+ + 5Fe^- \rightarrow 5Fe^+$	k_3	iron oxidation on nitrate
$M + 2O \rightarrow \varnothing$	k_4	methanotrophy on oxygen
$M + S^+ \rightarrow S^-$	k_5	methanotrophy on sulfate
$\frac{1}{4}O + Fe^- \rightarrow Fe^+$	k_6	iron oxidation

Table S3: Reactions simulated in the model.

5.4 Source terms

Interactions between the hypolimnion and the outside world are modeled by simple source terms. Oxygen and carbon are added at the thermocline. Methane can be produced by primary oxidation in the water column, but methanogenesis also proceeds in the sediment, whence it is transported upward and consumed by methanotrophy. We model this process by a point source of methane in the sediment. All methane in our model is consumed before reaching the thermocline, so we omit the mechanics for emission of methane into the metalimnion.

5.5 Parameterization

A list of parameters and their values is included in Table S4. Where possible, parameters related to the reaction rates were borrowed from Hunter *et al.* (1) and, in some cases, adjusted by hand. The parameters related to transport, source terms, and initial concentrations were drawn from published data where possible. Other values were adjusted by hand to match the observed data.

parameter	value	unit	source & reported value
General parameters			
T	0.4	yr	Asserted to set time scale
N	17	—	Asserted for 1 m compartments
Primary oxidation parameters			
k^{C} (rate constant)	1.0	yr^{-1}	(1) ($k^{\mathrm{DOC}} = 3 \times 10^{-5}$—$3 \times 10^{1}$ yr^{-1})
a (N:C ratio)	0.1	—	(3) and (4)[1] (0.066)
$[\mathrm{O}_{\mathrm{lim}}]$	20.0	μM	(1) (20.0 μM)
$\left[\mathrm{N}^{+}_{\mathrm{lim}}\right]$	5.0	μM	(1) (5.0 μM)
$\left[\mathrm{Fe}^{+}_{\mathrm{lim}}\right]$	0.1	μM	(1) (60 μmol dm^{-3})
$\left[\mathrm{S}^{+}_{\mathrm{lim}}\right]$	30.0	μM	(1) (30.0 μM)
Secondary oxidation parameters			
k_1 (ammonia oxidation)	5.0	$\mu\mathrm{M}^{-1}$ yr^{-1}	(1) ($k_4^{\mathrm{sr}} = 5 \times 10^{6}$ M^{-1} yr^{-1})
k_2 (sulfide oxidation)	0.16	$\mu\mathrm{M}^{-1}$ yr^{-1}	(1) ($k_5^{\mathrm{sr}} = 1.6 \times 10^{5}$ M^{-1} yr^{-1})
k_3 (iron oxidation, nitrate)	1.0	$\mu\mathrm{M}^{-1}$ yr^{-1}	(5)[2] (0.6–3 $\mu\mathrm{M}^{-1}$ yr^{-1})
k_4 (methanotrophy, oxygen)	10^{4}	$\mu\mathrm{M}^{-1}$ yr^{-1}	(1) ($k_9^{\mathrm{sr}} = 10^{10}$ M^{-1} yr^{-1})
k_5 (methanotrophy, sulfate)	10^{-2}	$\mu\mathrm{M}^{-1}$ yr^{-1}	(1) ($k_{10}^{\mathrm{sr}} = 10^{4}$ M^{-1} yr^{-1})
k_6 (iron oxidation)	10^{4}	$\mu\mathrm{M}^{-1}$ yr^{-1}	(1) ($k_2^{\mathrm{sr}} = 10^{7}$ M^{-1} yr^{-1})
Transport parameters			
D	50	yr^{-1}	(6)[3] (6–158 yr^{-1})
p_{Fe} (settling for oxidized iron)	0.3	—	(5)[4] (0.18)
p_{C} (settling for biomass)	0.3	—	Manually adjusted
Source rates			
s_{C}	9.4×10^{4}	μM yr^{-1}	(4)[5] (1.3×10^{4}—6.9×10^{4} μM yr^{-1})
s_{O}	6.6×10^{3}	μM yr^{-1}	Manually adjusted
s_{M}	2830	μM yr^{-1}	(7)[6] (475–1460 μM yr^{-1})
Initial concentrations			
[O]	50	μM	Manually adjusted
[C]	200	μM	Manually adjusted[7]
$[\mathrm{N}^{+}] + [\mathrm{N}^{-}]$	100	μM	(8)[8]
$[\mathrm{N}^{+}]/[\mathrm{N}^{-}]$	10	—	Manually adjusted
$[\mathrm{Fe}^{+}] + [\mathrm{Fe}^{-}]$	60	μM	(8)
$[\mathrm{Fe}^{+}]/[\mathrm{Fe}^{-}]$	10	—	Manually adjusted
$[\mathrm{S}^{+}] + [\mathrm{S}^{-}]$	250	μM	(8)
$[\mathrm{S}^{+}]/[\mathrm{S}^{-}]$	10	—	Manually adjusted
[M]	0	μM	Manually adjusted

Table S4: Parameter values and sources.

6 Implementation

The model was implemented in Matlab, and the ODE solutions were computed using the command `ode15s` with all chemical species restricted to nonnegative values (command `odeset`).

7 Inferred biomass

In the compartment at depth d there are n_r biotically-catalyzed reactions with rates $R_i(d)$, where $i \in \{1, \ldots, n_r\}$. We define the *relative rate* of the

[1]The cited database was queried for river gauge USGS 01102500 (Aberjona River, Winchester, MA; the Aberjona drains into Mystic Lake) parameter P00681 ("Organic carbon, water, filtered, milligrams per liter") during 1999-2000. The average value was 4.91 mg L^{-1}. The cited text's Table 12-4 lists C:N = 15.1 for carbon concentrations just above this in Wisconsin lakes, thus N:C = 1/15.1 = 0.066.

[2]The cited work reports an initial rate for iron oxidation on nitrate as 2.4 μM day^{-1} with initial nitrate concentration 30 μM and initial iron concentrations 10–50 μM, which corresponds to the shown range when assuming a second-order rate form.

[3]The cited paper collates reports of vertical eddy diffusion constants 0.002–0.05 cm^2 s^{-1} for lakes with depths comparable to Mystic Lake's. Diffusion constants are typically written as (time)/(length)^2, but here the compartment height, 1 meter, sets the length scale.

[4]The cited work reports an effective settling rate 0.024 m day^{-1} (v_{eff}, Table 3B). The shown value is computed by equating v_{eft} to comparable to $D \times p_{\text{Fe}}$ and using the shown value of D.

[5]The cited text gives annual organic carbon input for Wingra Lake, a polluted urban lake, as 691 g C m^{-2} yr^{-1} (Table 23-12) and for Lawrence Lake as 130.6 g C m^{-2} yr^{-1} (Table 23-13). These two values correspond to the reported range if the carbon is assumed delivered to a one-meter-high compartment. One should note that the carbon inputs reported in the literature should not directly correspond to C in the model, which is a simplified biomass.

[6]The cited work reports 1.3–4.0 mmol m^{-2} d^{-1}, which corresponds to the shown values if the methane is delivered to a well-mixed one-meter high lowest compartment as assumed in the model.

[7]Ref. 4 reports that eutrophic lakes have median total organic carbon around 12.0 mg L^{-1} = 10^3 μM (Table 23-1). The carbon concentrations later in the simulation are more similar to this value.

[8]The total concentrations of nitrogen, iron, and sulfur species were chosen to correspond with the total amount of these chemical observed in the cited work's appendix tables.

i-th reaction at depth d as

$$r_i(d) \equiv \frac{R_i(d)}{\sum_{j=1}^{n_r} R_j(d)}. \tag{9}$$

In the inferred biomass framework, we assert that the biomass $b_i(d)$ of the organisms catalyzing the i-th process across depths d is proportional to the relative rates $r_i(d)$, that is,

$$b_i(d) = \alpha_i r_i(d) \tag{10}$$

for all d, where α_i is some constant of proportionality that relates biomass to relative rate that varies with i (the organisms catalyzing different rates) but not with depth. Because the α_i are unknown, we never infer the relationship between abundances of different biomasses, even at the same depth. Furthermore, $b_i(d)$ is the biomass of *all* organisms catalyzing process i—the inferred biomass framework does not provide information about the combined biomass of individual taxa, only the biomass of all taxa catalyzing a modeled process.

In comparisons with the survey data, we used relative rates $r_i(d)$ rather than absolute rates $R_i(d)$ because survey count data show relative, not absolute, abundances. For example, consider a process i whose absolute rate is higher at depth x than at depth y, i.e., $R_i(x) > R_i(y)$. Other processes, however, are much more active at x than at y so that although i's absolute rate is higher at x, its relative rate is higher at y, i.e., $r_i(x) < r_i(y)$. In this case, we would expect the biomass of organisms performing process i to have higher absolute abundance at x but higher relative abundance at y.

References

[1] Hunter K, Wang Y, Van Cappellen P (1998) Kinetic modeling of microbially-driven redox chemistry of subsurface environments: coupling transport, microbial metabolism and geochemistry. *J Hydrol (Amst)* 209:53–80.

[2] Senn DB, Hemond HF (2002) Nitrate controls on iron and arsenic in an urban lake. *Science* 296(5577):2373–2376.

[3] US Geological Survey (2014) National Water Information System (USGS Water Data for the Nation).

[4] Wetzel R (2001) *Limnology.* (Academic Press), 3rd edition.

[5] Senn D (2001) Ph.D. thesis (MIT).

[6] Benoit G, Hemond H (1996) Vertical eddy diffusion calculated by the flux gradient method: Significance of sediment-water heat exchange. *Limnol Oceanogr* 41:157–168.

[7] Varadharajan C, Hemond H (2012) Time-series analysis of high-resolution ebullition fluxes from a stratified, freshwater lake. *J Geophys Res* 117(G2).

[8] Peterson E (2005) Master's thesis (MIT).

[9] Oh H et al. (2011) Complete genome sequence of strain IMCC9063, belonging to SAR11 subgroup 3, isolated from the Arctic Ocean. *J Bacteriol* 193(13):3379–3380.

[10] Weon H et al. (2009) *Solitalea koreensis* gen. nov., sp. nov. and the reclassification of [*Flexibacter*] *canadensis* as *Solitalea canadensis* comb. nov. *Int J Syst Evol Microbiol* 59(8):1969–1975.

[11] Finneran K, Johnsen C, Lovley D (2003) *Rhodoferax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). *Int J Syst Evol Microbiol* 53(3):669–673.

[12] Nevin KP et al. (2005) *Geobacter bemidjiensis* sp. nov. and *Geobacter psychrophilus* sp. nov., two novel fe(III)-reducing subsurface isolates. *Int J Syst Evol Microbiol* 55(4):1667–1674.

[13] Balk M, Altınbaş M, Rijpstra W, Damsté J, Stams A (2008) *Desulfatirhabdium butyrativorans* gen. nov., sp. nov., a butyrate-oxidizing, sulfate-reducing bacterium isolated from an anaerobic bioreactor. *Int J Syst Evol Microbiol* 58(1):110–115.

[14] Purkhold U, Wagner M, Timmermann G, Pommerening-Röser A, Koops H (2003) 16S rRNA and *amoA*-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int J Syst Evol Microbiol* 53(5):1485–1494.

[15] Alawi M, Lipski A, Sanders T, Pfeiffer E, Spieck E (2007) Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J* 1(3):256–264.

[16] Blöthe M, Roden E (2009) Composition and activity of an autotrophic Fe(II)-oxidizing, nitrate-reducing enrichment culture. *Appl Environ Microbiol* 75(21):6937–6940.

[17] Emerson D, Moyer C (2002) Neutrophilic Fe-oxidizing bacteria are abundant at the loihi seamount hydrothermal vents and play a major role in Fe oxide deposition. *Appl Environ Microbiol* 68(6):3085–3093.

[18] Lapidus A et al. (2011) Genomes of three methylotrophs from a single niche reveal the genetic and metabolic divergence of the *Methylophilaceae*. *J Bacteriol* 193(15):3757–3764.

[19] Wartiainen I, Hestnes A, McDonald I, Svenning M (2006) *Methylobacter tundripaludum* sp. nov., a methane-oxidizing bacterium from Arctic wetland soil on the Svalbard islands, Norway (78 degrees N). *Int J Syst Evol Microbiol* 56(1):109–113.

Part II
Supplementary Tables and Figures

comparison	pairs in real data	pairs in shuffled data				
		mean	std	min	max	p-value
timepoints	190	68.2	8.4	49	91	10^{-109}
sample prep	777	101.4	9.3	77	129	$\sim$0
OEU callers	2564	131.3	11.9	96	172	$\sim$0

Table S5: Numbers of pairs of OTUs that were in the same OEU across datasets. Values for shuffled data represent 1000 random shufflings of the assignments of OTUs to OEUs in one of the two datasets in the comparison. The p-value represent the results of one-sided z-tests using the mean and standard deviations of the number of OTU pairs from the shuffled datasets. (The $\sim$0 indicates values much less than 10^{-109}.)

function (OTU ID in 2008; 2013)	closest genome or type strain (reference)	identity (%)	clone GI (accession)
Aerobic heterotrophy (31; 1)	*Candidatus Pelagibacter* sp. IMCC9063 (9)	91	444189431 (KC192422)
Denitrification (298; 188)	*Solitalea canadensis* DSM 3403 (10)	90	444189414 (KC192405)
Iron reduction { (11; 12)	{ *Rhodoferax ferrireducens* T118 (11)	{ 99	{ 444189437 (KC192428)
Iron reduction { (13; 228)	{ *Geobacter psychrophilus* strain P35 (12)	{ 96	{ 444189534 (KC192525)
Sulfate reduction (88; 106)	*Desulfatirhabdium butyrativorans* strain HB1 (13)	92	444189499 (KC192490)
Ammonia oxidation (16-2; 39)	*Nitrosospira briensis* (14)	98	444189385 (KC192376)
Nitrite oxidation (141-2; 29)	*Candidatus Nitrotoga arctica* (15)	99	444189434 (KC192425)
Sulfide and iron oxidation { (104; 6)	{ *Sideroxydans lithotrophicus* ES-1 (16)	{ 95	{ 444189494 (KC192485)
Sulfide and iron oxidation { (125; 223)	{ *Sideroxydans lithotrophicus* ES-1 (17)	{ 99	{ 444189496 (KC192487)
Methane oxidation { (99; 68)	{ *Methylotenera versatilis* 301 (18)	{ 94	{ 444189407 (KC192398)
Methane oxidation { (276; 8)	{ *Methylobacter psychrophilus* (19)	{ 99	{ 444189514 (KC192505)

Table S6: Reference OTUs.

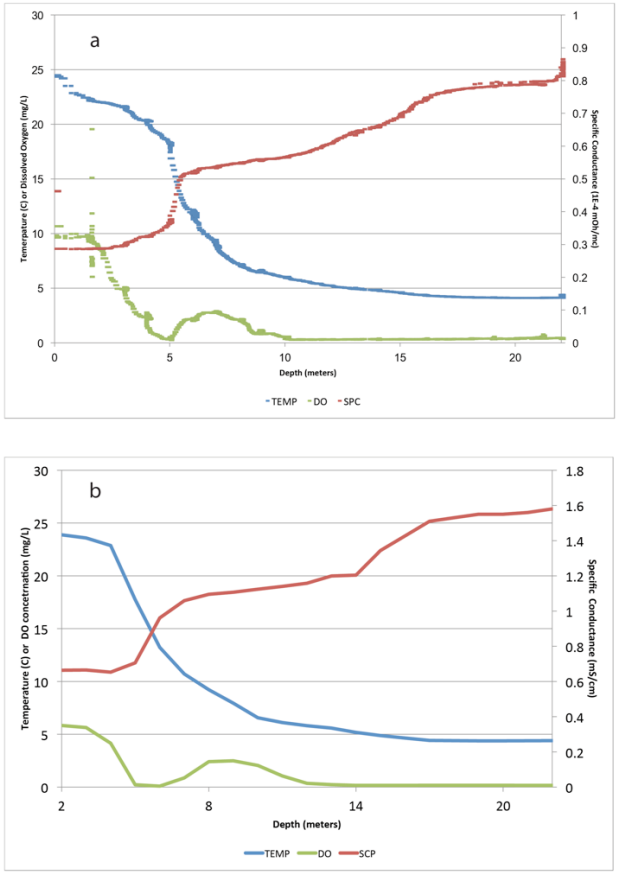

Figure S1: *In situ* measurements taken during sample collection on (*a*) Aug. 13, 2008 and (*b*) Aug. 15, 2013. Temperature (°C, blue), dissolved oxygen (DO, mg/L; green), and specific conductance (SCP mS/cm; red) are shown.

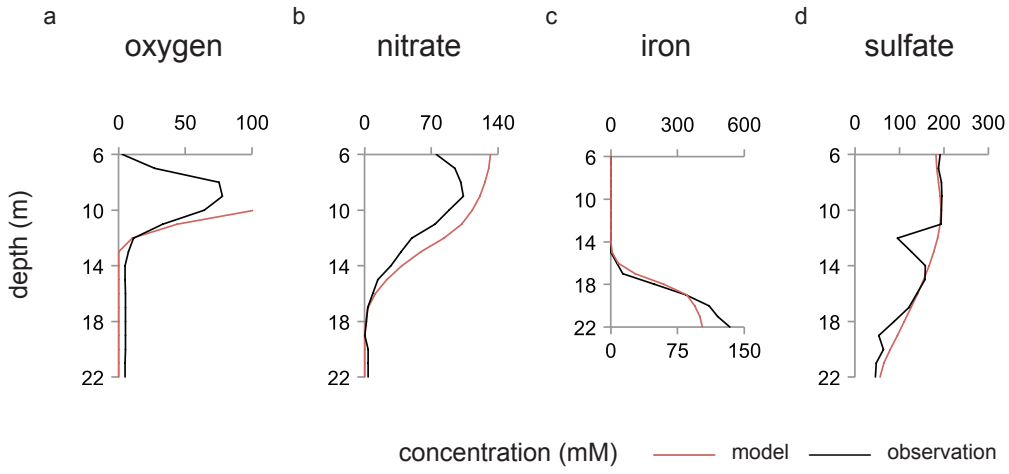

Figure S2: Correspondence between 2013 chemical observations and model predictions after the model was calibrated to match the chemical observation. Modeled concentrations (red) are on the same scale as observations (black) except for iron (top axis, observed; lower axis, modeled).

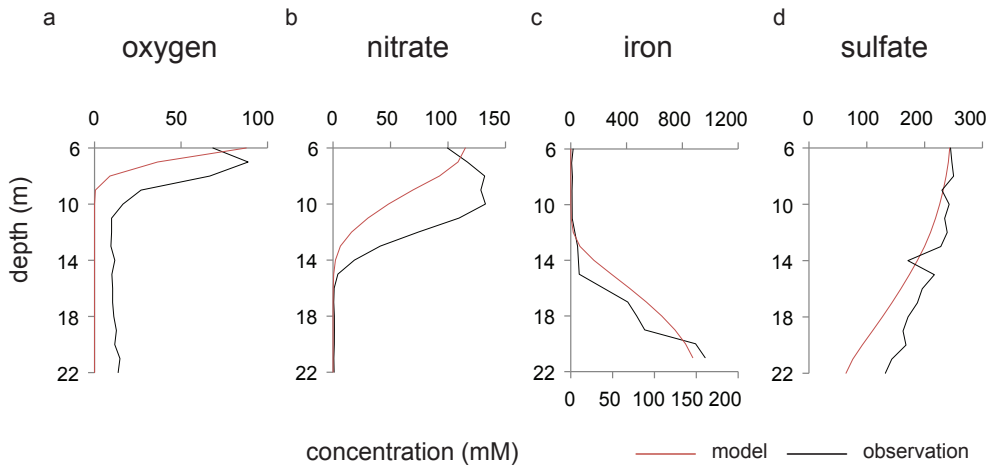

Figure S3: Correspondence between 2008 chemical observations and model predictions after the model was calibrated to match the chemical observation. Modeled concentrations (red) are on the same scale as observations (black) except for iron (top axis, observed; lower axis, modeled).

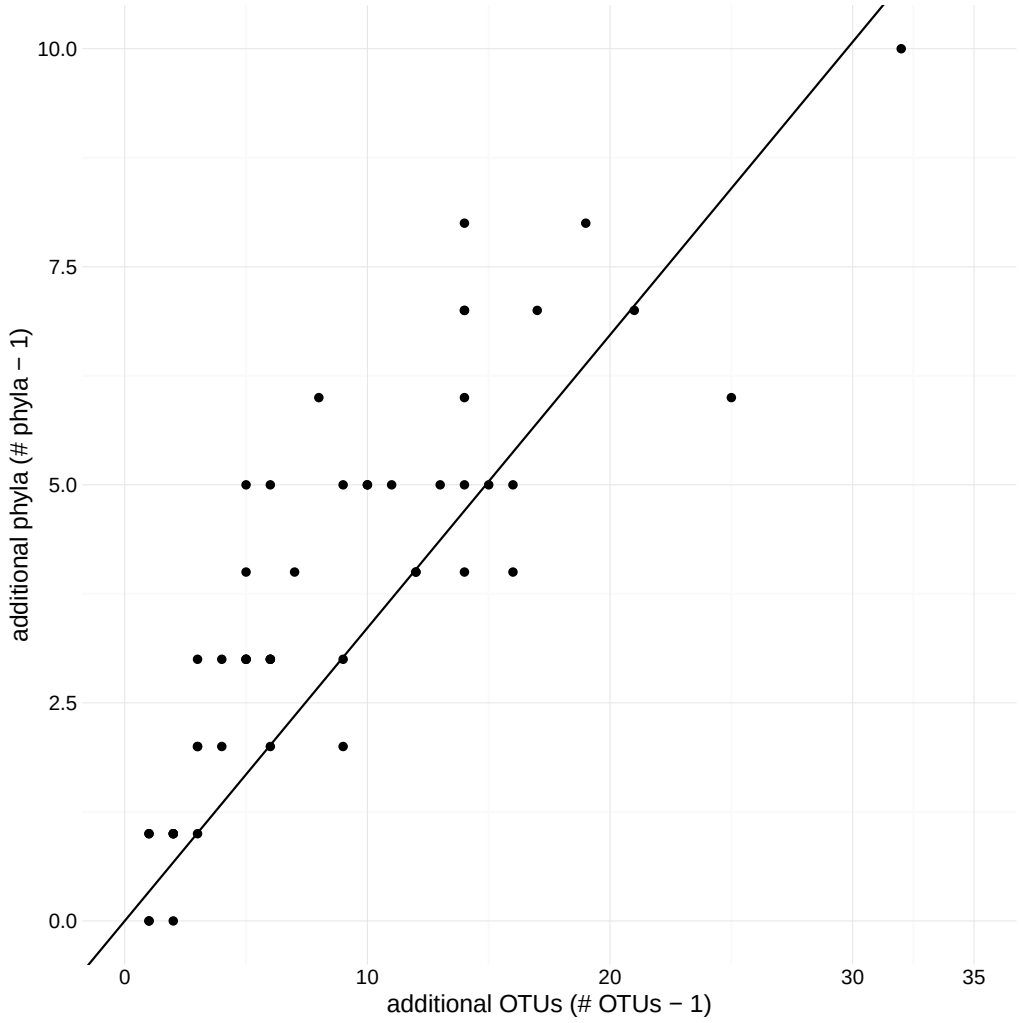

Figure S4: The number of phyla within OEUs increased linearly with the number of OTUs in the OEU. The regression was constrained so that 1 OTU in an OEU necessarily produced 1 phylum in that OEU.

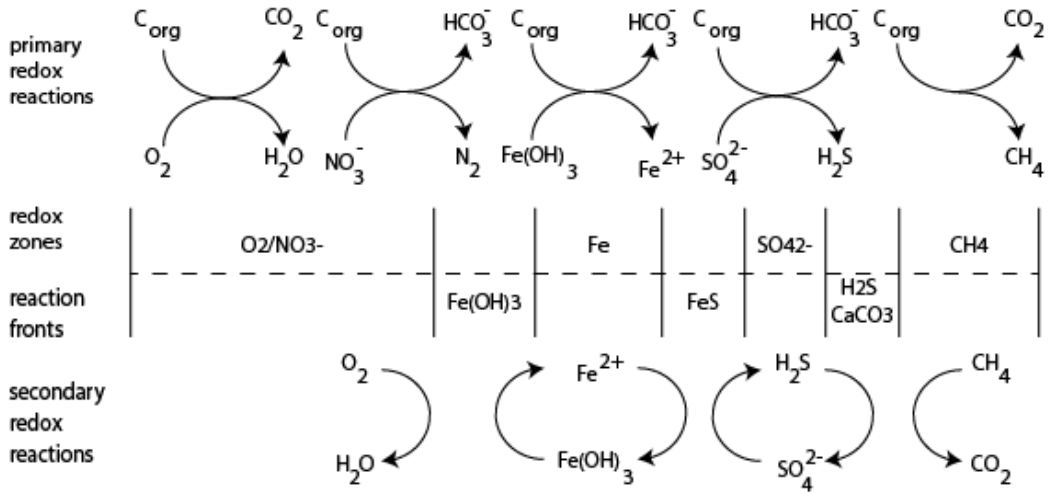

Figure S5: Primary and secondary oxidation-reduction (redox) reactions used in the biogeochemical model, along with the typical redox zones and reaction fronts. Adapted from (1).

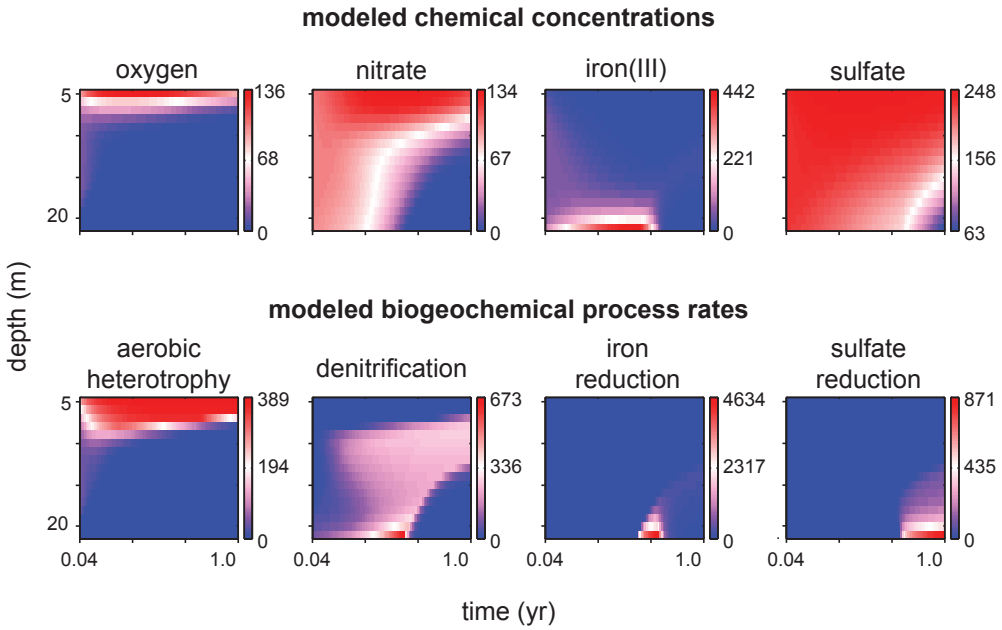

Figure S6: The model predicts the spatial and temporal dynamics of the lake's chemistry as well as the places and times in the lake that favor specific biogeochemical processes. The model starts in a homogeneous state, representing an idealized, fully-mixed dimictic lake in the spring of 2008. Concentrations (top) of electron acceptors are measured in μM; the rates of biogeochemical processes (bottom) are measured in μM yr^{-1}.

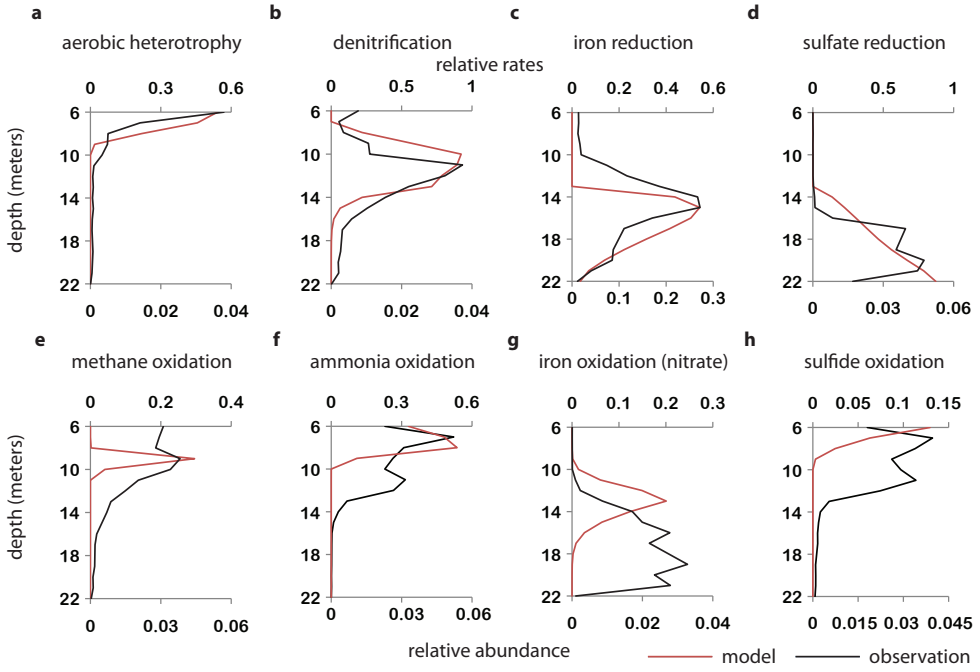

Figure S7: Observed distribution of key populations (black, bottom x-axis, measured in relative abundance) and their correspondence with modeled processes (red, top x-axis, measured in relative rate) from 2008. The observations, which reflect the abundance of organisms, and the model results, which reflect the predicted prevalence of a metabolic process, have peaks at similar locations and their distributions roughly correspond, which suggests that the distribution of these organisms is largely determined by the favorability of the corresponding metabolic process.

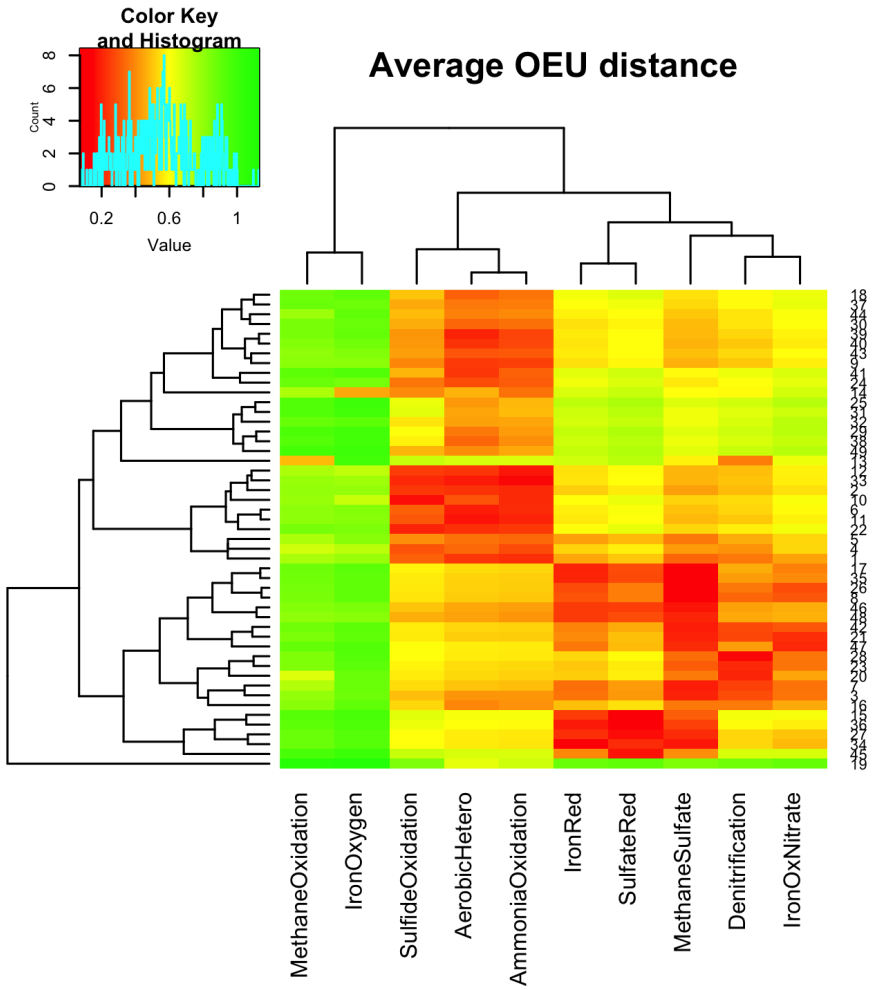

Figure S8: Average Euclidean distance across OTUs within each OEU (rows; 2013 data) to the modeled processes (columns; 2013 model), demonstrating the relationship of OEUs with processes. Dendrograms show the clustering patterns of modeled processes (top) and OEUs (left). OEUs cluster largely by their relationship to aerobic and anaerobic processes.

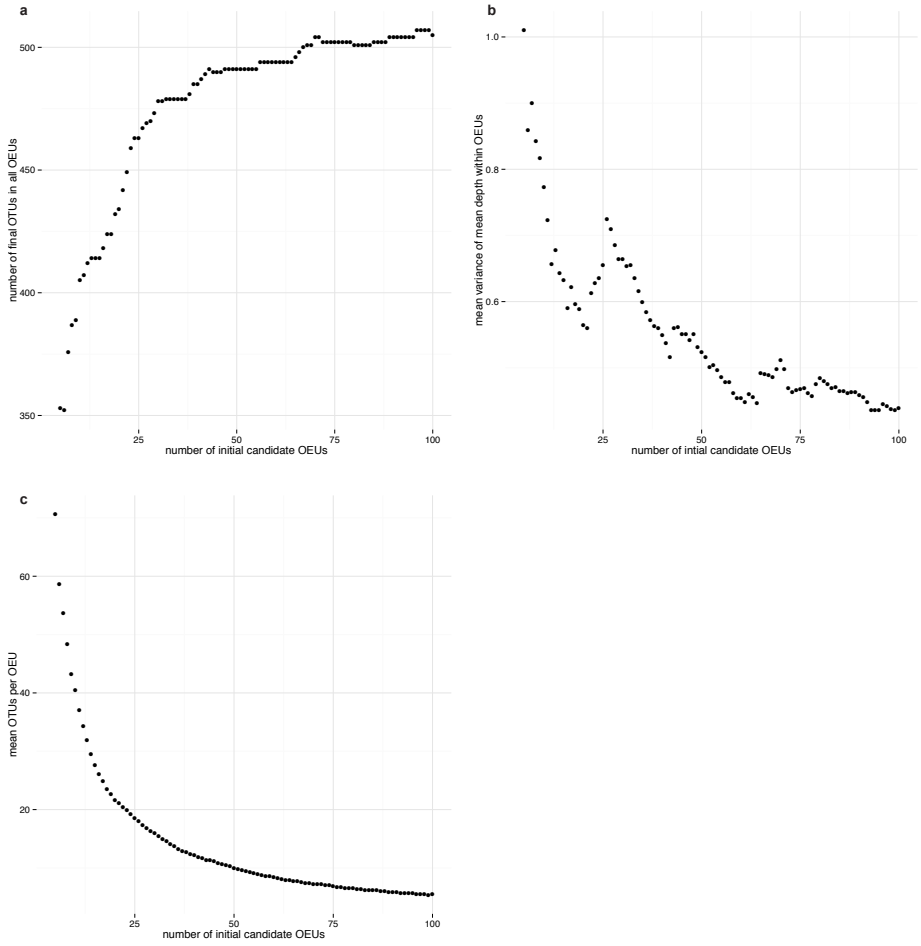

Figure S9: (a) The number of OTUs remaining in the final analysis after OEUs are called increases with the number of initial candidate OEUs. (b) Each OTU is distributed across the lake's depth, so each OTU has a mean depth. For every OEU, the variance in mean depths was computed. For each set of OEUs produced by a different number of initial candidate OEUs, the mean of those variances was computed. The mean (across OEUs) of variances (within an OEU) of the mean depth (of each OTU) is a measure of overall cluster quality. The mean variance decreases with increasing initial OEU number. (c) The mean number of OTUs in each OEU decreases with increasing initial OEU number.