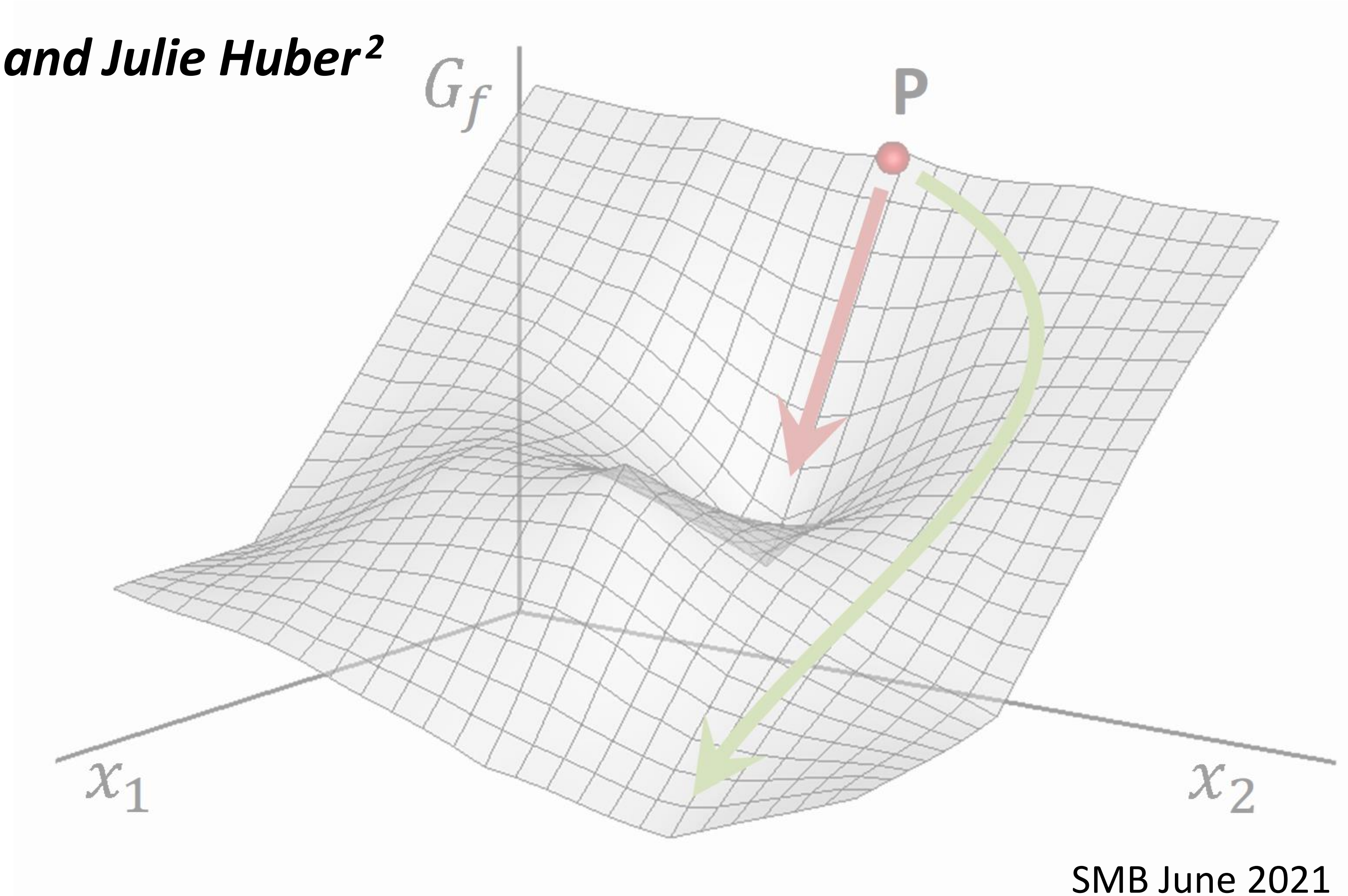
Using the maximum entropy production principle to understand and predict microbial biogeochemistry

Joe Vallino[1], Ioannis Tsakalakis[1], and Julie Huber[2]

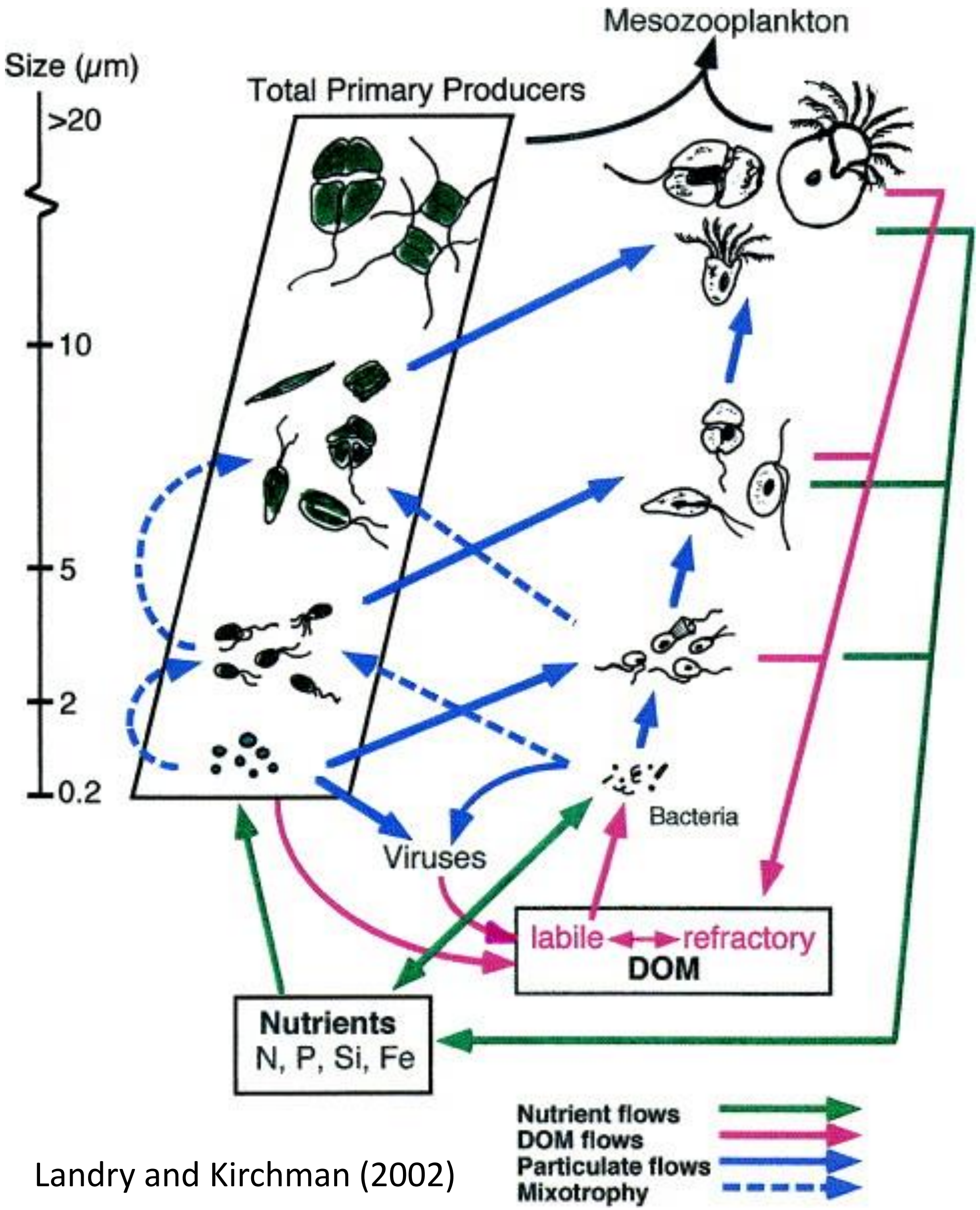
[1] Ecosystems Center
Marine Biological Laboratory
Woods Hole

[2] Marine Chemistry and Geochemistry
Woods Hole Oceanographic Institution

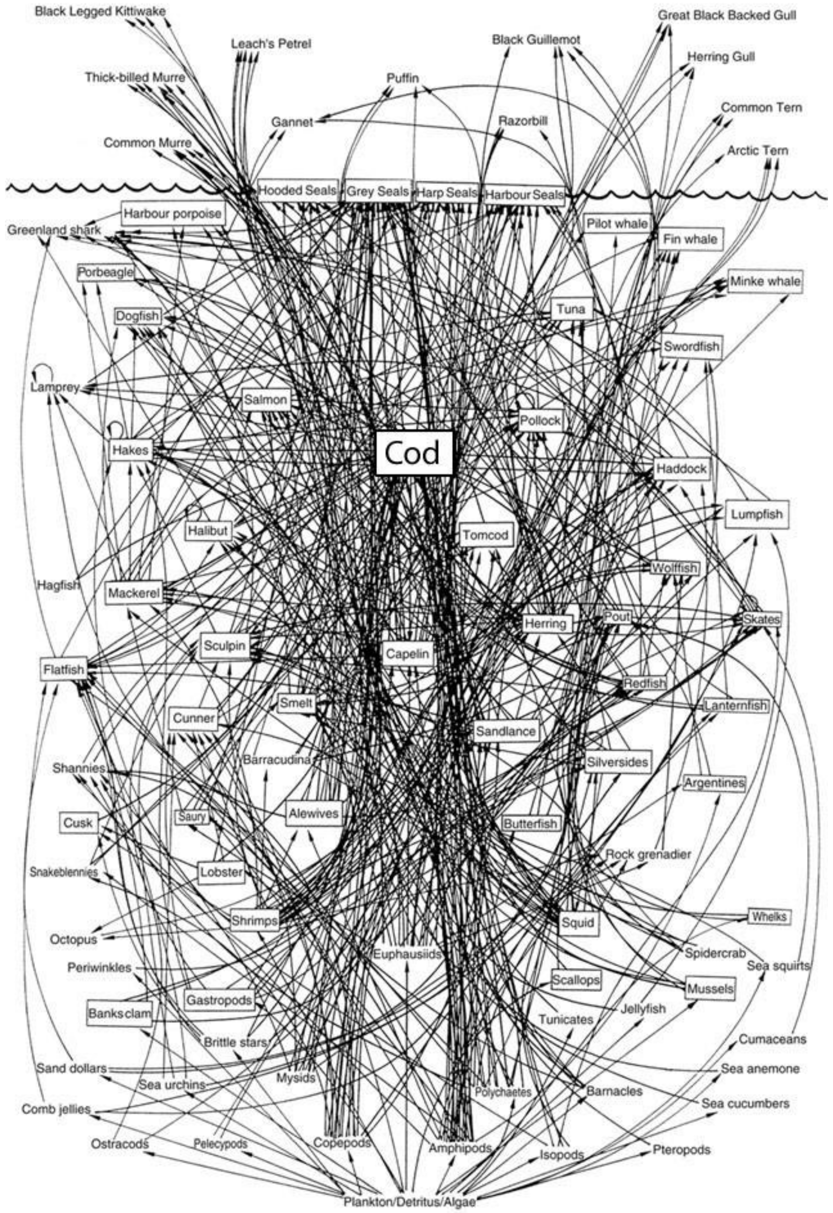


SMB June 2021

Reductionist Modeling Approach, Marine Biogeochemistry



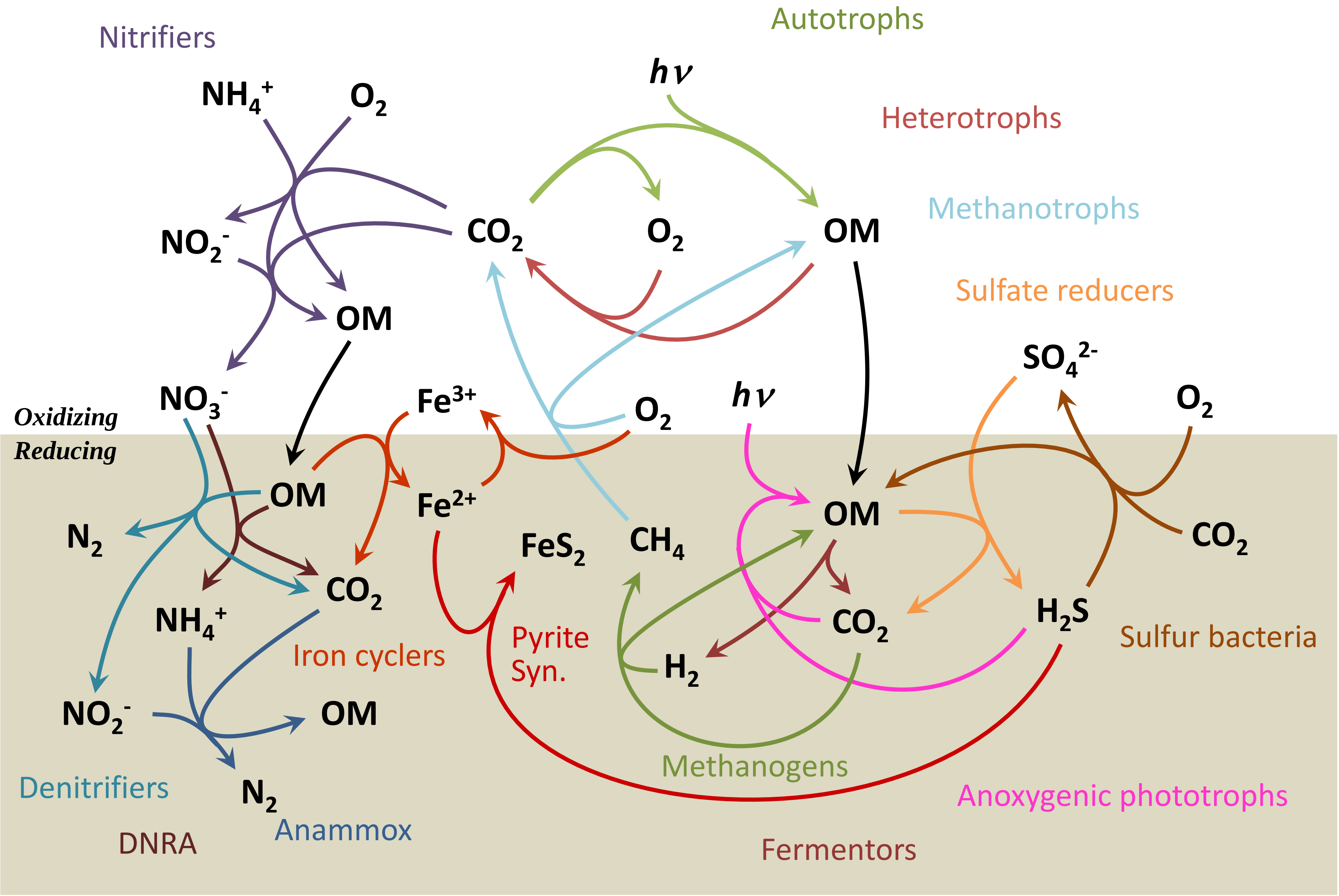
Landry and Kirchman (2002)



A simplified food web for the Northwest Atlantic. © IMMA

- Aggregation causes errors
- Huge number of parameters typically
- Insufficient information
- What happens when the community composition changes?
- How about the Rare Biosphere?

Focus on functions that dissipate energy instead



Pathways are distributed across phyla

No centralized control

Which pathways are up-regulated?

Which are down-regulated?

How do resources limit?

Many degrees of freedom

Does "who's there" matter?

How to predict protein allocation?

Use optimization-base approach!

Maximum Entropy Production (MEP)

Steady state nonequilibrium systems with many degrees of freedom will likely organize to maximize the rate of entropy production.

See:
Ziegler 1963
Paltridge 1975
Dewar 2003, 2005
Martyushev & Seleznev 2006

Examples: Fire, Hurricanes, and Living systems.

Entropy?

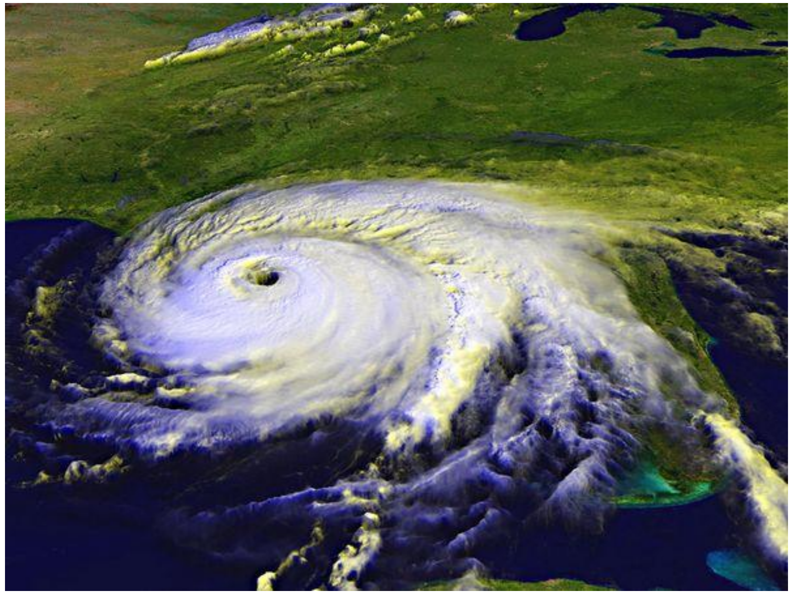
- Dispersal (spreading out) of energy
- Order is usually a small and unimportant component of entropy
- Order must contend with Boltzmann's constant: $k=1.3806488 \times 10^{-23}$ J K^{-1}
- Living organisms are not low entropy structures.

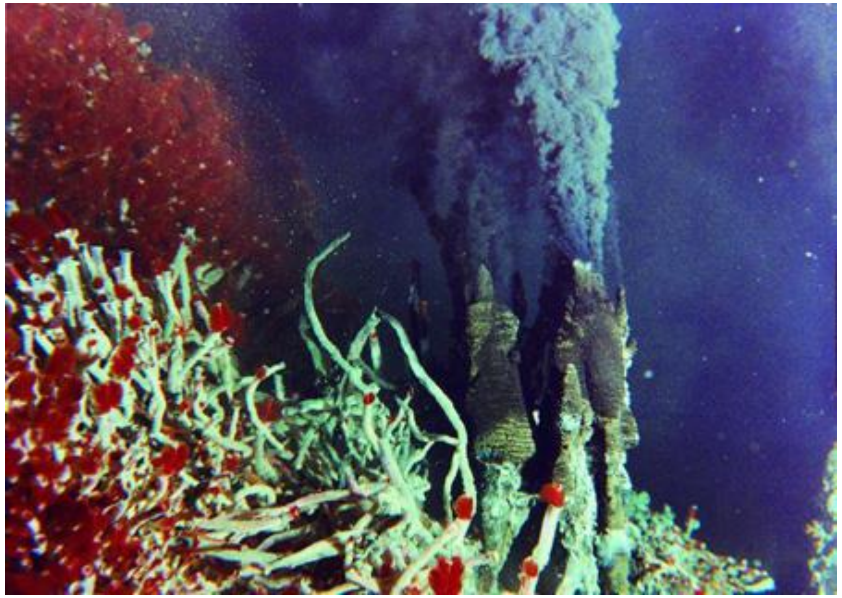
Easier to think about entropy as the destruction of *free* energy†

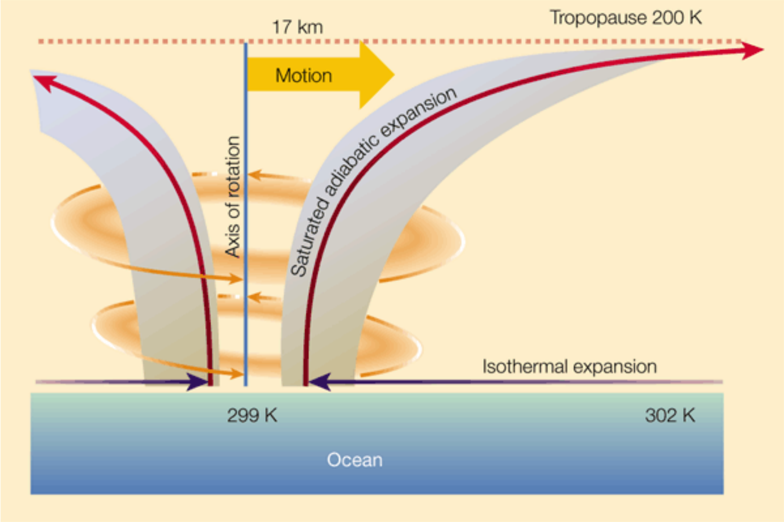
† Energy is conserved, free energy (or exergy) is not.

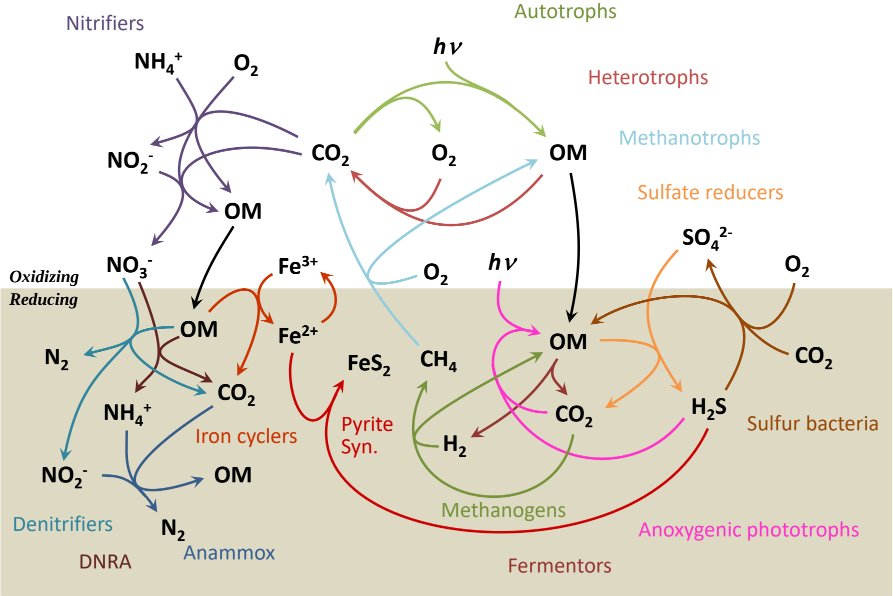
Maximum Entropy Production (MEP)

Or: Systems organize to maximize dissipation rate of energy potentials







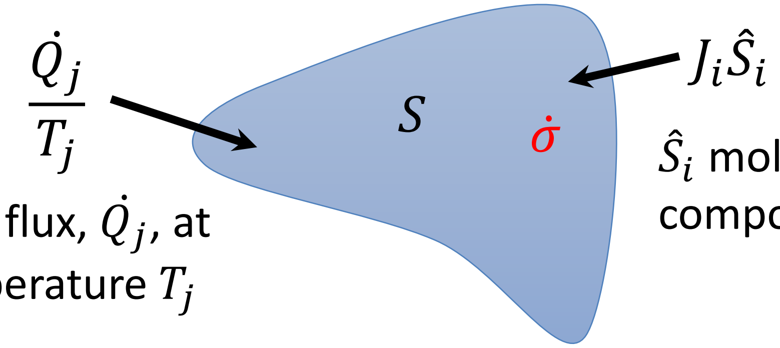


MEP, Which Entropy Term?

Entropy Balance Equation:

$$\frac{dS}{dt} = \sum_i J_i \hat{S}_i + \sum_j \frac{\dot{Q}_j}{T_j} + \dot{\sigma}$$

where $\dot{\sigma}$ is internal entropy production from irreversible processes. Second Law says $\dot{\sigma} \geq 0$



Heat flux, $\dot{Q}_j$, at temperature T_j

$\hat{S}_i$ molar entropy of compound i with flux J_i

MEP concerns $\dot{\sigma}$ not S

For Chemical Reaction $aA + bB \rightarrow cC + dD$

Gibbs free energy

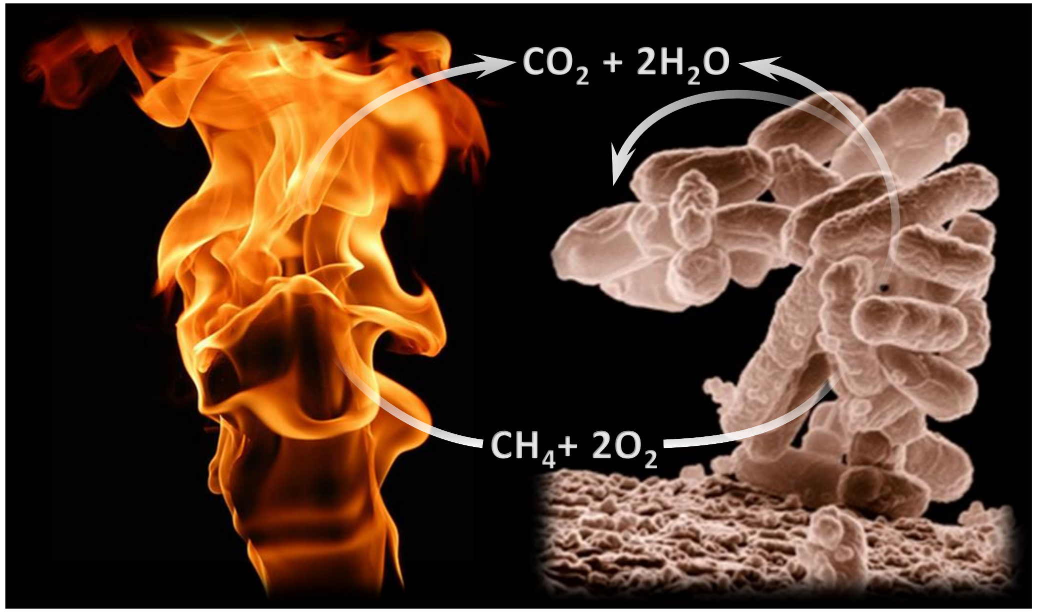
$$\Delta_r G(T,P) = \Delta_r G^\circ(T,P) + RT \log\left(\frac{(\gamma_c[C])^c(\gamma_d[D])^d}{(\gamma_a[A])^a(\gamma_b[B])^b}\right)$$

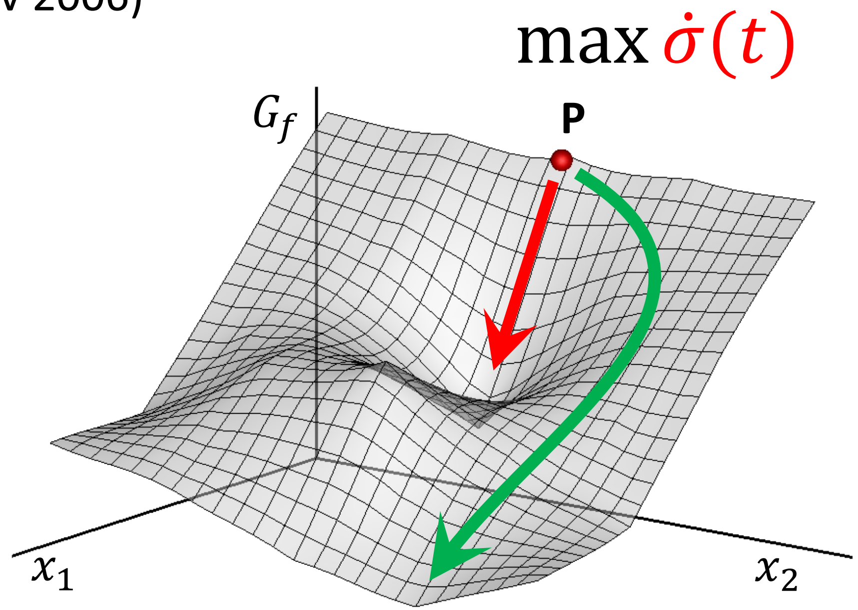
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \quad \Delta_r G^\circ = -818\ kJ\ mol^{-1}$$

If free energy is not stored, then: $\dot{\sigma} = -\frac{V}{T} r \Delta_r G(T,P)$

Maximum Entropy Production for non-steady state systems

MEP only derived for SS systems (Dewar 2003, Martyushev & Seleznev 2006)





$$\max\langle\dot{\sigma}(t)\rangle = \max\left(\frac{1}{\Delta t}\int_{t}^{t+\Delta t}\dot{\sigma}(\tau)d\tau\right)$$

$$\max\langle\dot{\sigma}(t)\rangle \geq \frac{1}{\Delta t}\int_{t}^{t+\Delta t}\max\dot{\sigma}(\tau)\,d\tau$$

Difference between abiotic (fire) and Life:

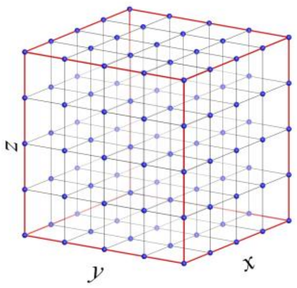
- Abiotic: maximizes instantaneous $\dot{\sigma}(t)$
- Life: maximizes time-averaged $\langle\dot{\sigma}(t)\rangle$

Temporal strategies are a hallmark of biology:

circadian rhythm, resource storage, dormancy, life cycles, anticipatory control, etc.

Coordination Over Space

Entropy production can also be increased with coordination over space (see Vallino 2010)



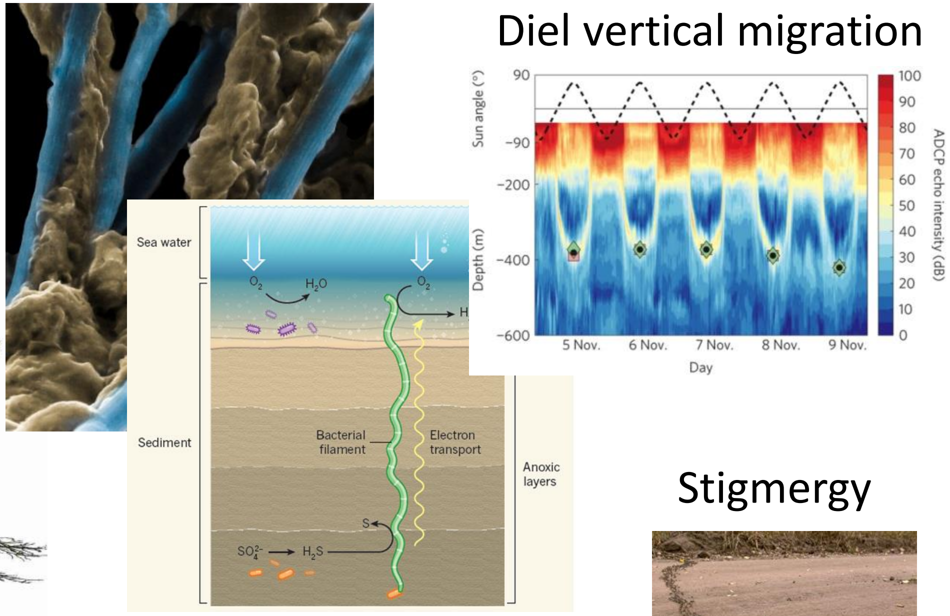
$$\langle\dot{\sigma}\rangle_{\Omega} = \max \frac{1}{V} \iiint_{\Omega} \dot{\sigma}(x,y,z)\, d\Omega$$

$$\langle\dot{\sigma}\rangle_{\Omega} \geq \frac{1}{V} \iiint_{\Omega} \max \dot{\sigma}(x,y,z)\, d\Omega$$

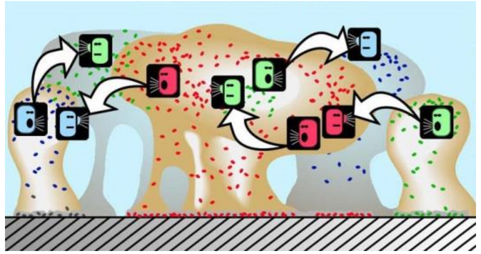
Multicellularity

Cable bacteria

Diel vertical migration



Quorum sensing



Stigmergy

Living systems evolve to maximize energy dissipation over the greatest possible spatial and temporal scales

"Paradigm shift, from '**we eat food**' to '**food has produced us to eat it**'" (Lineaweaver&Egan 2008)

MEP approach to modeling biogeochemistry

- Represent biogeochemistry as a distributed metabolic network focused on redox reactions

- Allocate catalysts (protein) to metabolic pathways that maximize entropy production over time and space

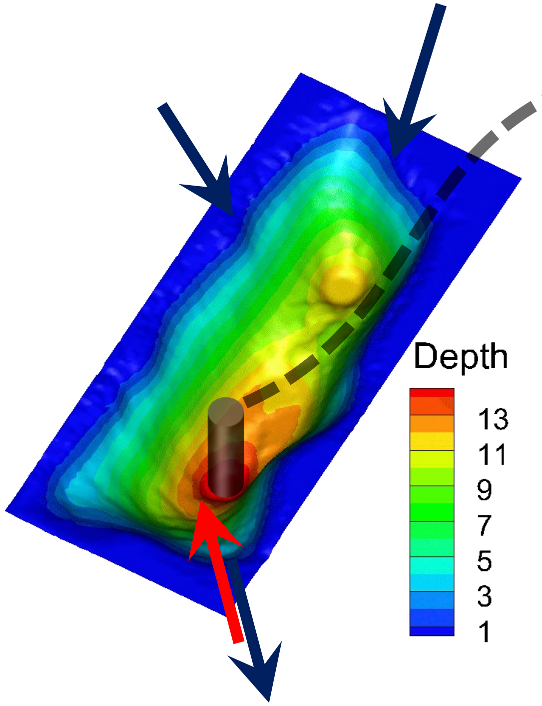
- Optimization replaces need to understand how communities assemble (aka, climate verses weather modeling)

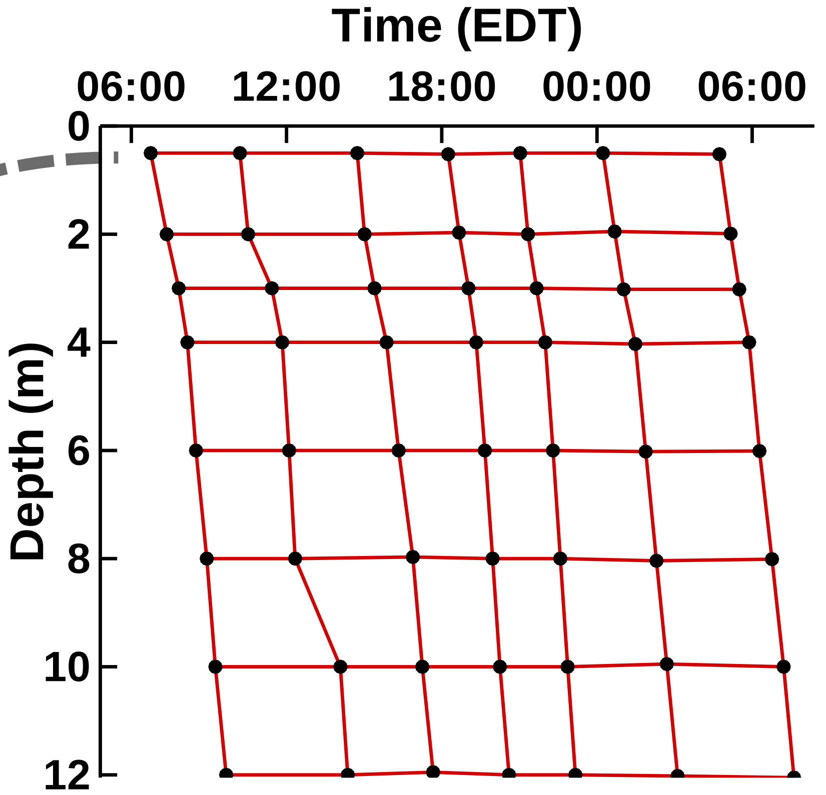
- Replace parameters with optimization variables

Example: Siders Pond “Laboratory”

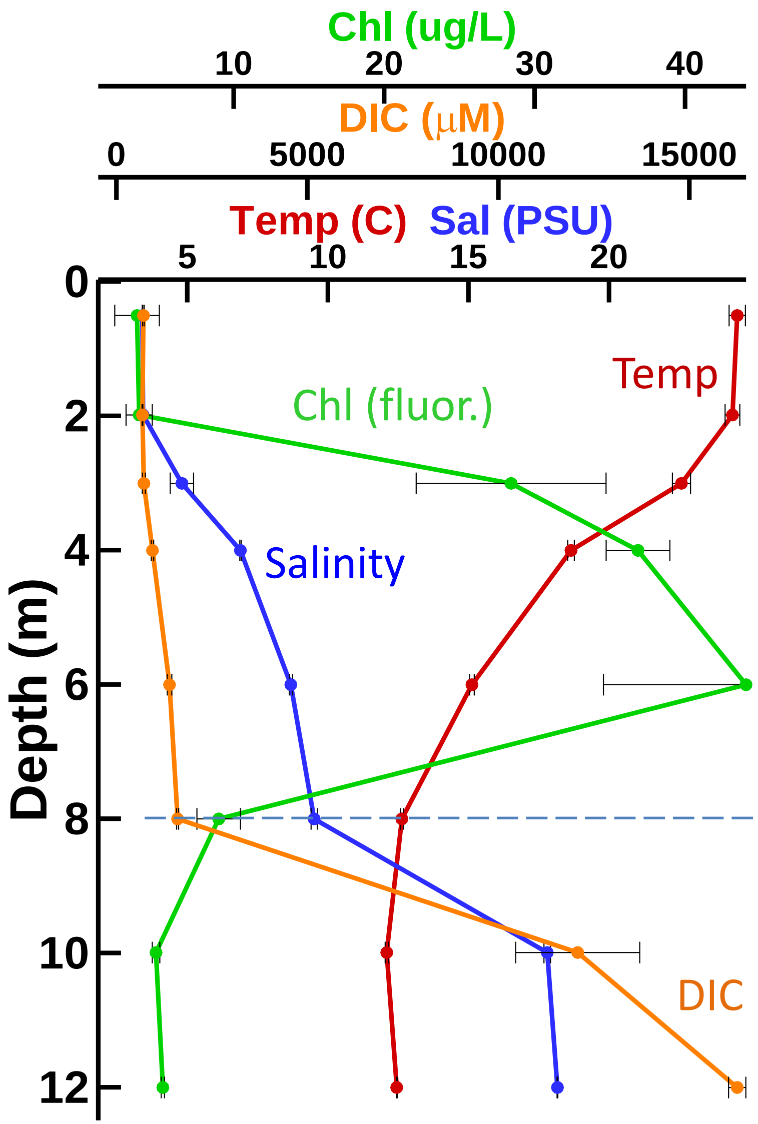
Meromictic Coastal Pond
Permanently stratified
Depth: 15 m max
Area: 14 hectares
Volume: ~ 1 Mm^3

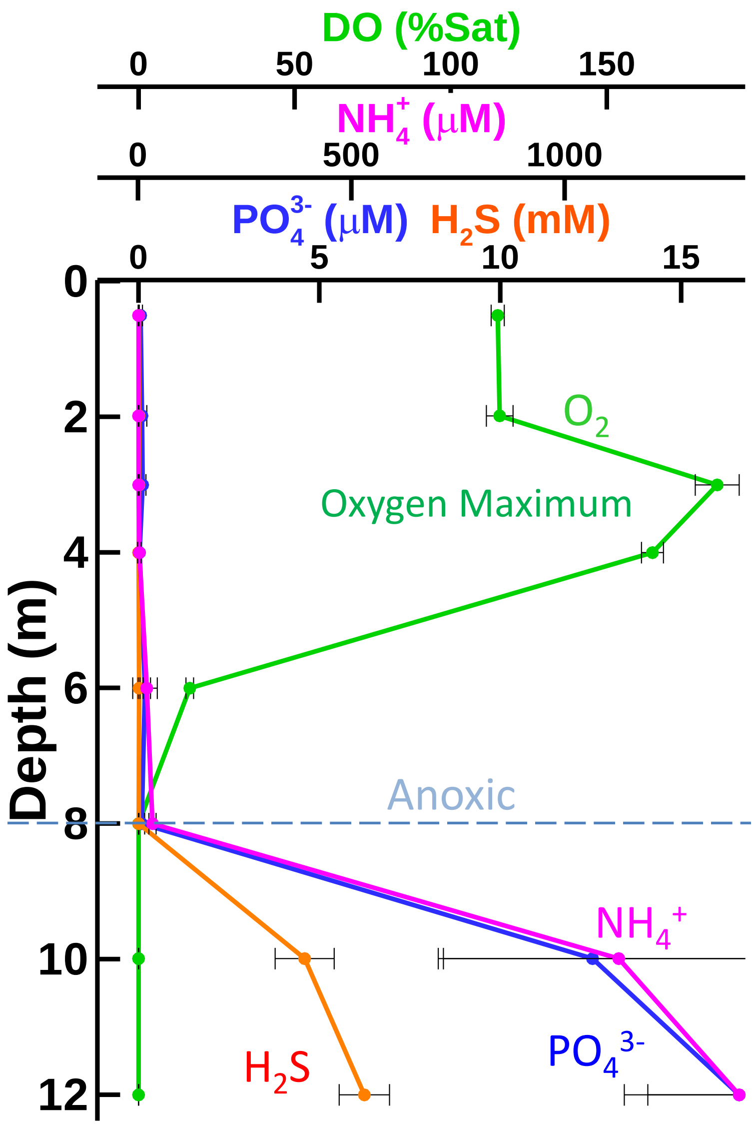
Space-Time Sampling Grid
- Biogeochemistry
- Metagenomics (cast 1)
- Metatranscriptomics (all casts)





Vertical Gradients

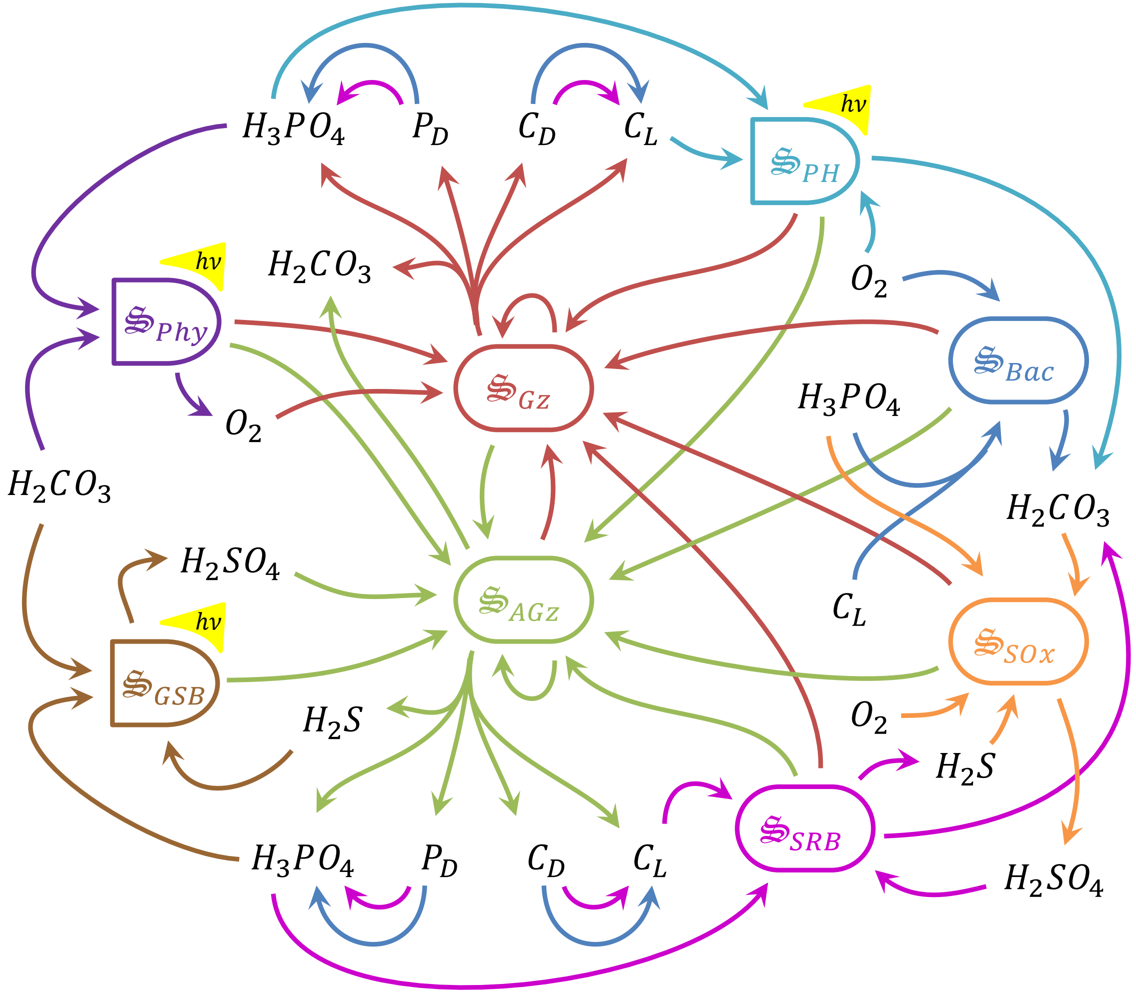


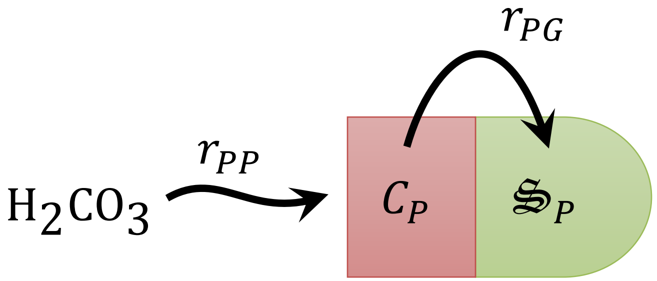


Large spatial gradients (over *m* not *mm*)

Distributed metabolic network

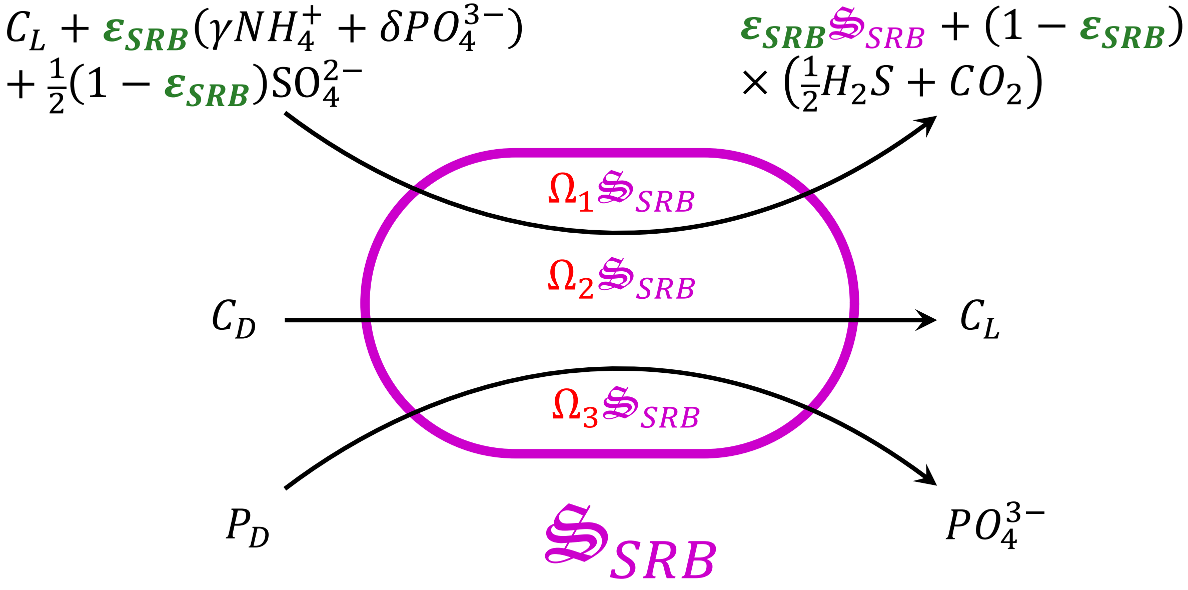
Vallino and Huber 2018





$$r_{PP}\colon\ \varepsilon_P \mathrm{H_2CO_3} + n_\gamma h\nu \xrightarrow{\Omega_{1,P}\mathfrak{S}_P} \varepsilon_P(C_P + \mathrm{O_2})$$

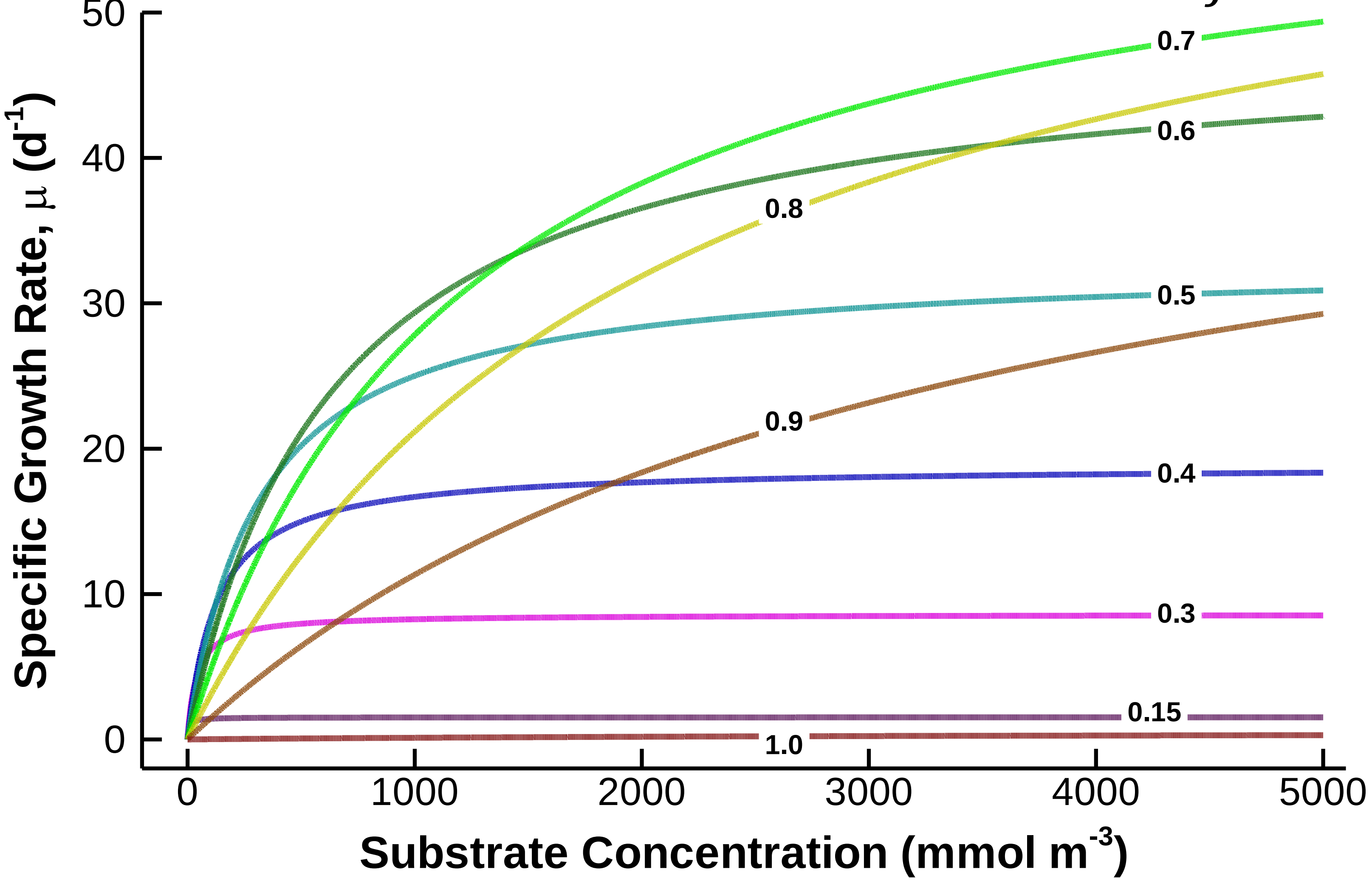
$$r_{PG}\colon\ C_P + (1-\varepsilon_P)\mathrm{O_2} + \varepsilon_P \delta \mathrm{H_3PO_4} + \cdots \xrightarrow{\Omega_{2,P}\mathfrak{S}_P} \varepsilon_P \mathfrak{S}_P + (1-\varepsilon_P)\mathrm{H_2CO_3}$$



Reaction Kinetics (adaptive Monod)

$$r_{i,j} = \nu^* \boldsymbol{\varepsilon}_j^2 \prod_{k=1}^{n_C} \left(\frac{C_k}{C_k + \kappa^* \boldsymbol{\varepsilon}_j^4} \right)^{\Lambda_{i,j,k}} \boldsymbol{\Omega}_{i,j} C_{\mathfrak{S}_j} F_T(\boldsymbol{\varepsilon}_j, \boldsymbol{\Omega}_{i,j})$$

Curves parameterized by $\boldsymbol{\varepsilon}_j$



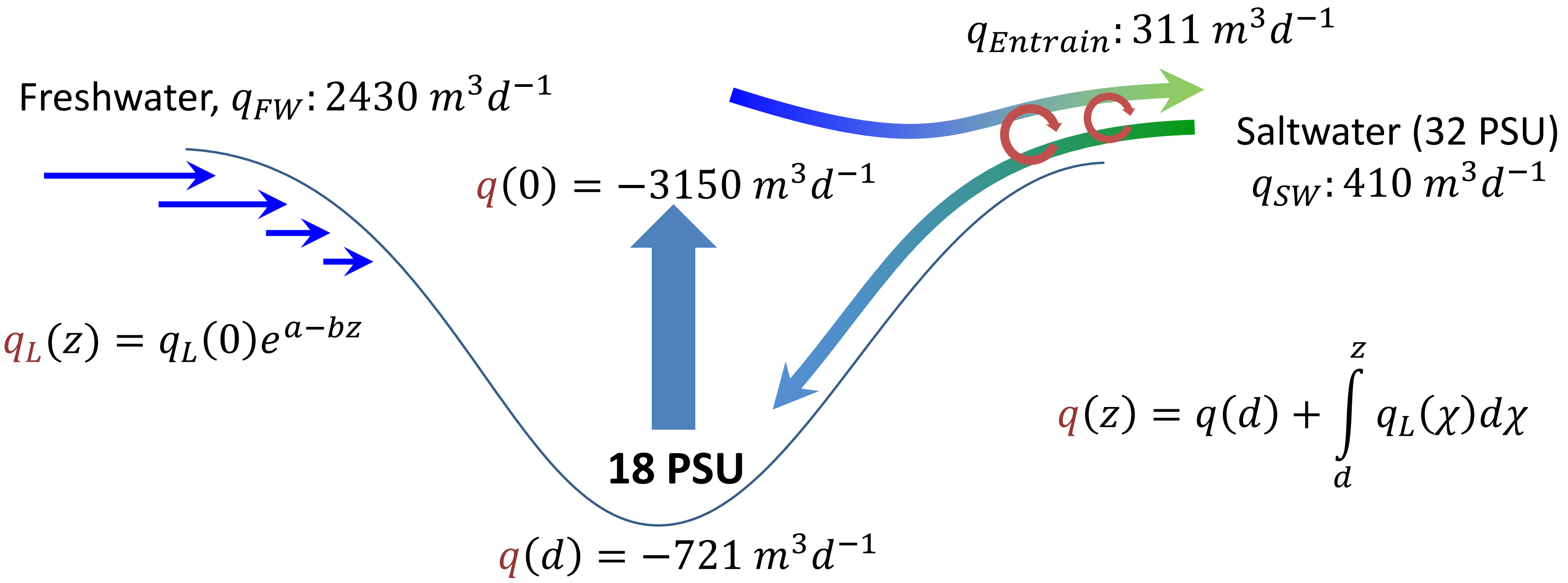
Siders Pond Transport (1D approximation)

$$\frac{\partial \mathbf{c}(t,z)}{\partial t} = D(z)\frac{\partial^2 \mathbf{c}(t,z)}{\partial z^2} + \left(\frac{D(z)}{A(z)}\frac{\partial A(z)}{\partial z} + \frac{\partial D(z)}{\partial z} - \frac{q(z)}{A(z)}\right)\frac{\partial \mathbf{c}(t,z)}{\partial z} - \frac{\mathbf{c}(t,z)}{A(z)}\frac{\partial q(z)}{\partial z} + \frac{q_L(z)\mathbf{c}_L(z)}{A(z)} + \mathbf{r}(t,z,\boldsymbol{\varepsilon},\boldsymbol{\Omega}),$$

BC: $\left.\frac{\partial \mathbf{c}(t,z)}{\partial z}\right|_{z=0} = 0$ $\qquad$ $\left(-D(z)A(z)\frac{\partial \mathbf{c}(t,z)}{\partial z} + q(z)\mathbf{c}(t,z)\right)_{z=d} = q(d)\mathbf{c}_B$

Dispersion coefficient, $D(z)$, was determined by fitting predicted to observed salinity vertical profile

Volumetric flow: $q(z)$ and lateral inputs: $q_L(z), \mathbf{c}_L(z)$ obtained from observations and assuming:

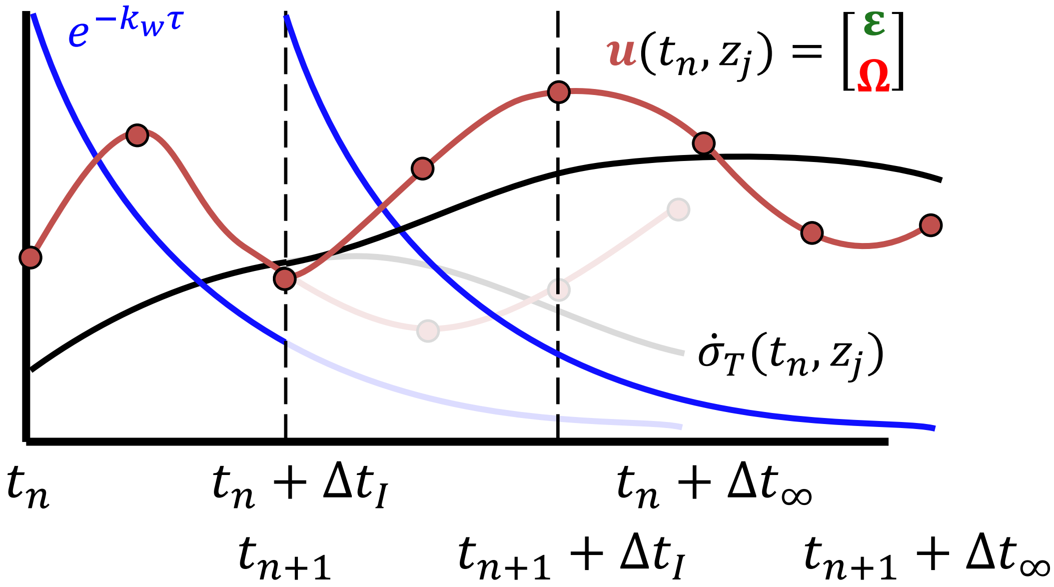


1D, **Local** MEP Model Setup

Two grids: One fine grid for PDE solution, $\mathbf{c}(t,z)$ (space and time adaptive, BACOLI95)
One course grid for Control Variables, $\varepsilon_j(z,t)$ and $\Omega_{i,j}(z,t)$.

Receding Horizon Optimal Control Problem:

$$\max_{\boldsymbol{\varepsilon},\boldsymbol{\Omega}} \int_{t_n}^{t_n+\Delta t_\infty} \dot{\sigma}_T(\tau,z_j) e^{-k_w\tau} d\tau \;\; \forall j \text{ in the control grid } [z_1, z_2, \dots, z_m]$$



where: (light, chemical)

$$\dot{\sigma}_T(\tau,z_j) \approx -\frac{A(z)}{T(z)}\left(\frac{\Delta I_j \Delta G_\gamma}{\eta_\gamma} + \sum_k r_k(\tau,z_j)\Delta_{r_k} G_k\right)$$

Subject to:

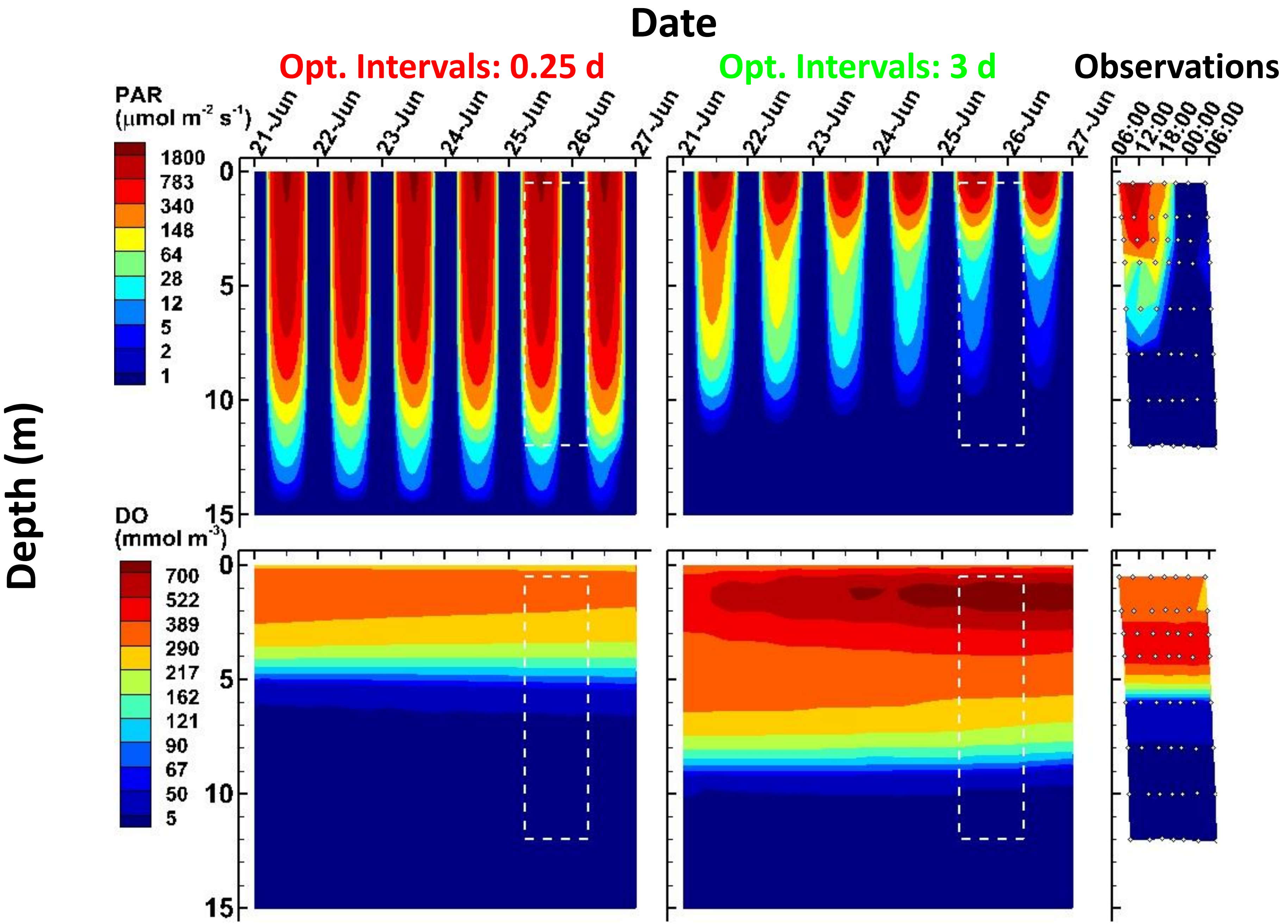
Sinking

$$\frac{\partial \mathbf{c}(t,z)}{\partial t} = D(x)\frac{\partial^2 \mathbf{c}(t,z)}{\partial z^2} + \left(\frac{D(z)}{A(z)}\frac{\partial A(z)}{\partial z} + \frac{\partial D(z)}{\partial z} - \frac{q(z)}{A(z)} - v_s\right)\frac{\partial \mathbf{c}(t,z)}{\partial z} - \left(\frac{1}{A(z)}\frac{\partial q(z)}{\partial z} + \frac{v_s}{A(z)}\frac{\partial A(z)}{\partial z}\right)\mathbf{c}(t,z) + \frac{q_L(z)\mathbf{c}_L(z)}{A(z)} + \mathbf{r}(t,z,\boldsymbol{\varepsilon},\boldsymbol{\Omega})$$

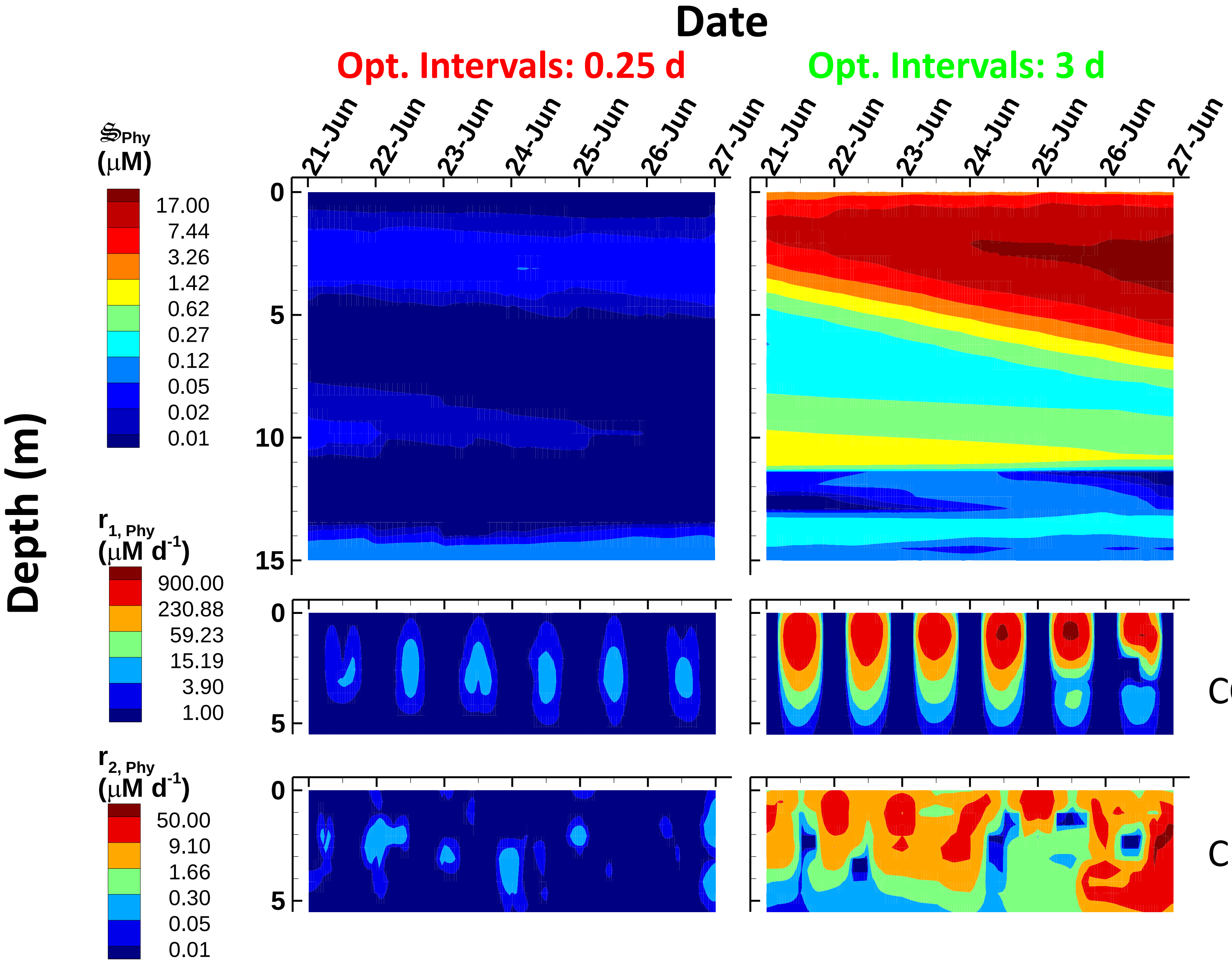
Note, this is a Control of PDE problem.

Short (0.25 d) vs Long (3 d) optimization: PAR & DO

Vallino & Huber 2018



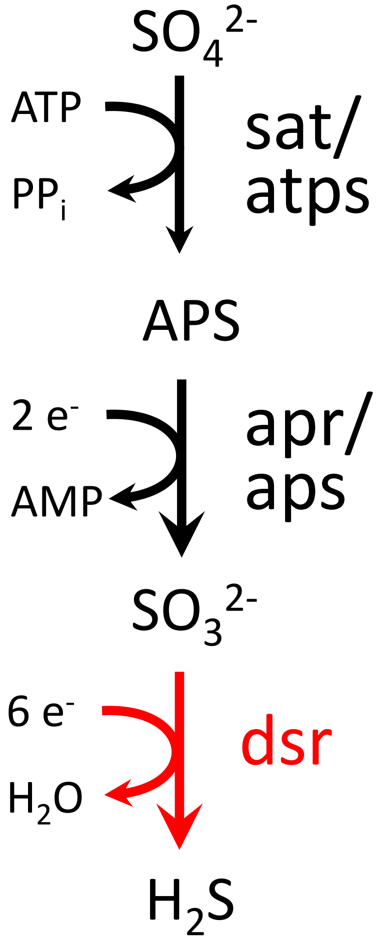
Short (0.25 d) vs Long (3 d) optimization: Phytoplankton



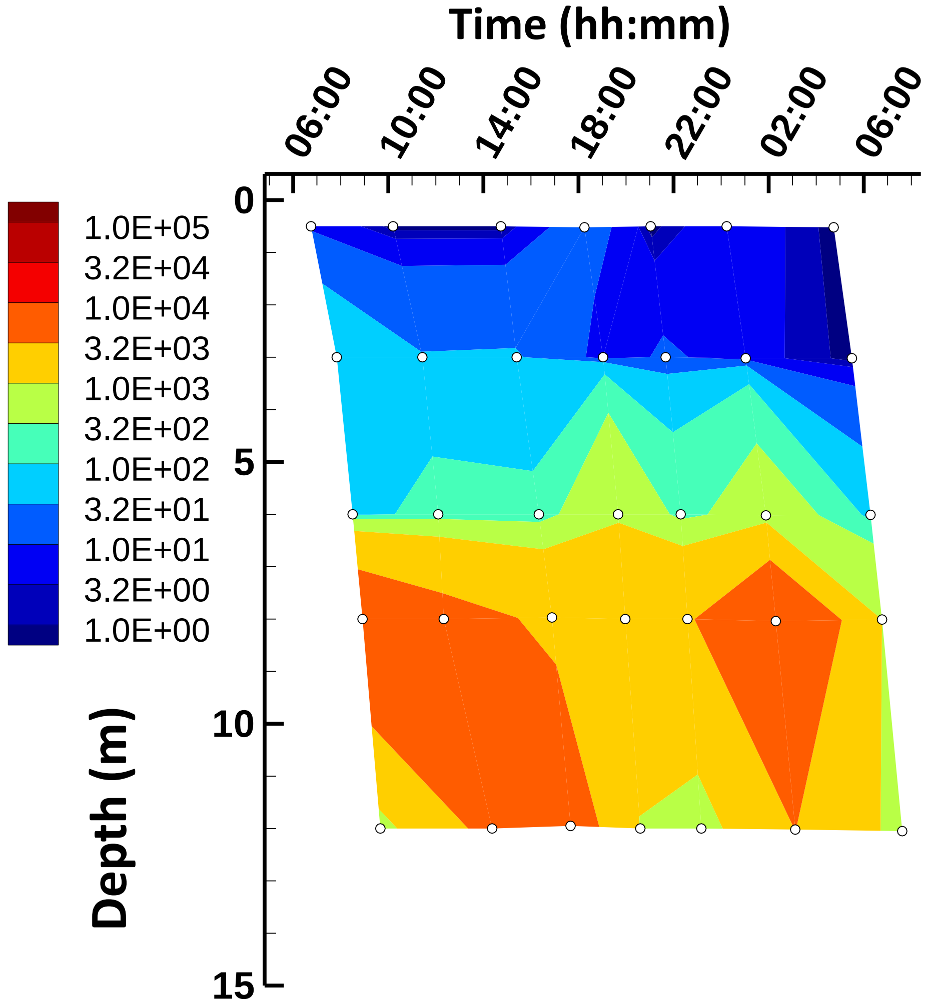
$$CO_2 + H_2O + \eta_\gamma \xrightarrow{\Omega_{1,P} \mathfrak{S}_{Phy}} CH_2O + O_2$$

$$CH_2O + P + O_2 \xrightarrow{\Omega_{2,P} \mathfrak{S}_{Phy}} \mathfrak{S}_{Phy} + CO_2$$

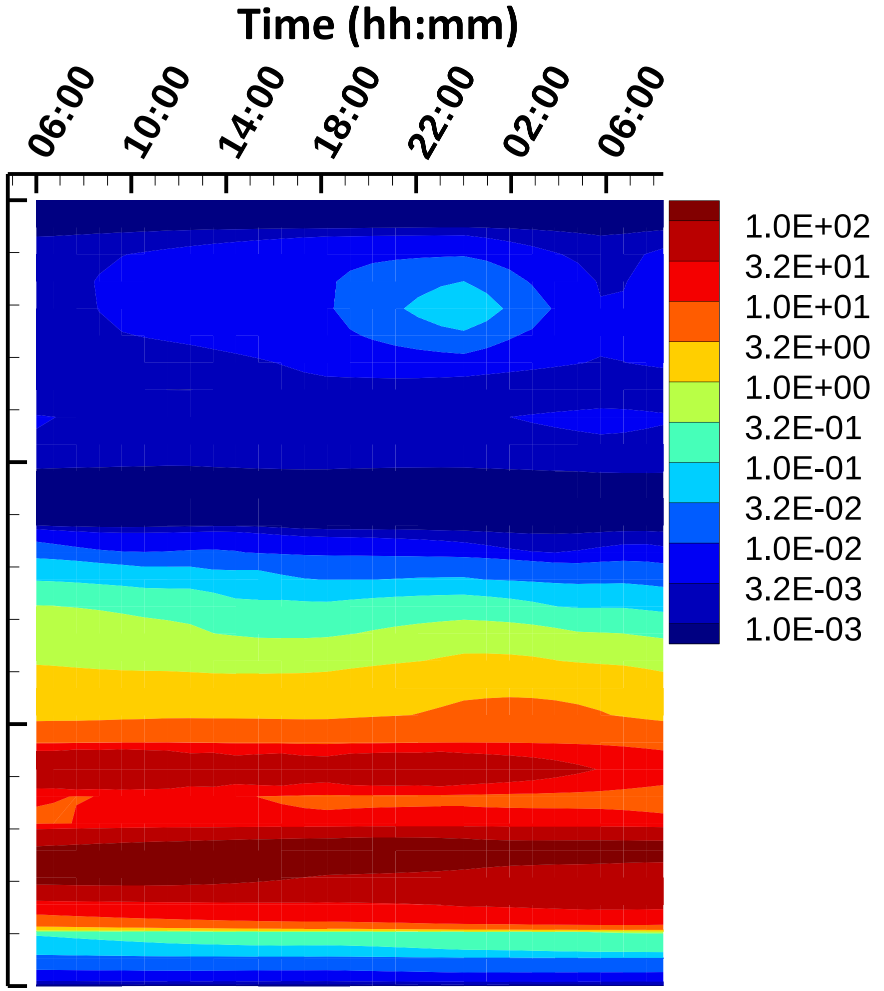
Breathing with Sulfate (sulfate reducing bacteria)



dsrA Transcript Abundance



$r_{1,SRB}$ (μM d^{-1})*

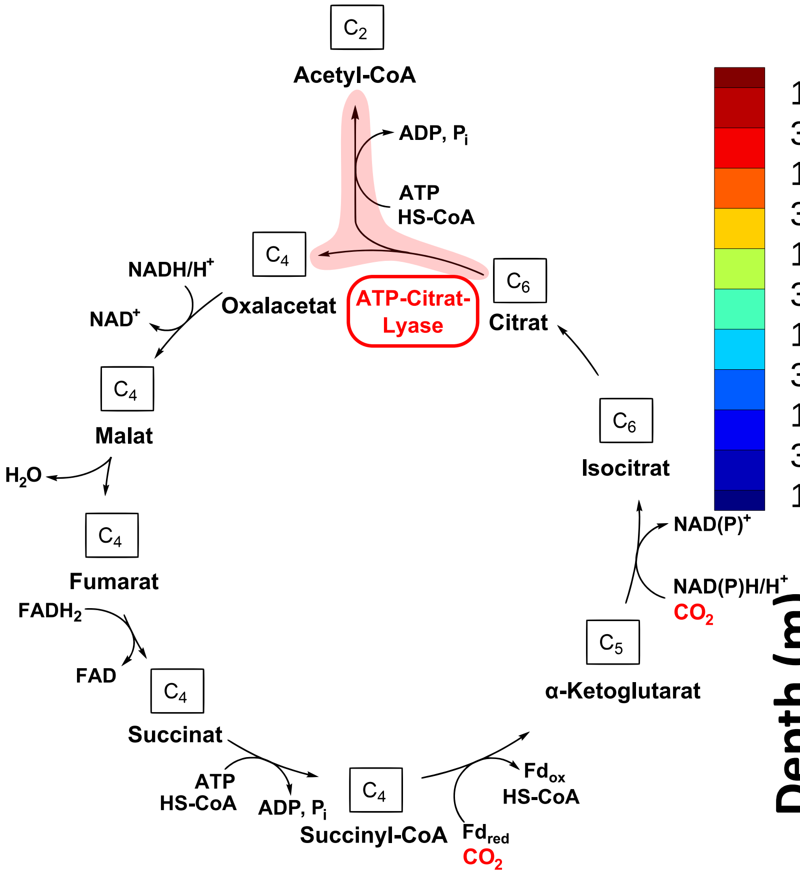


*(3 d Interval Optimization)

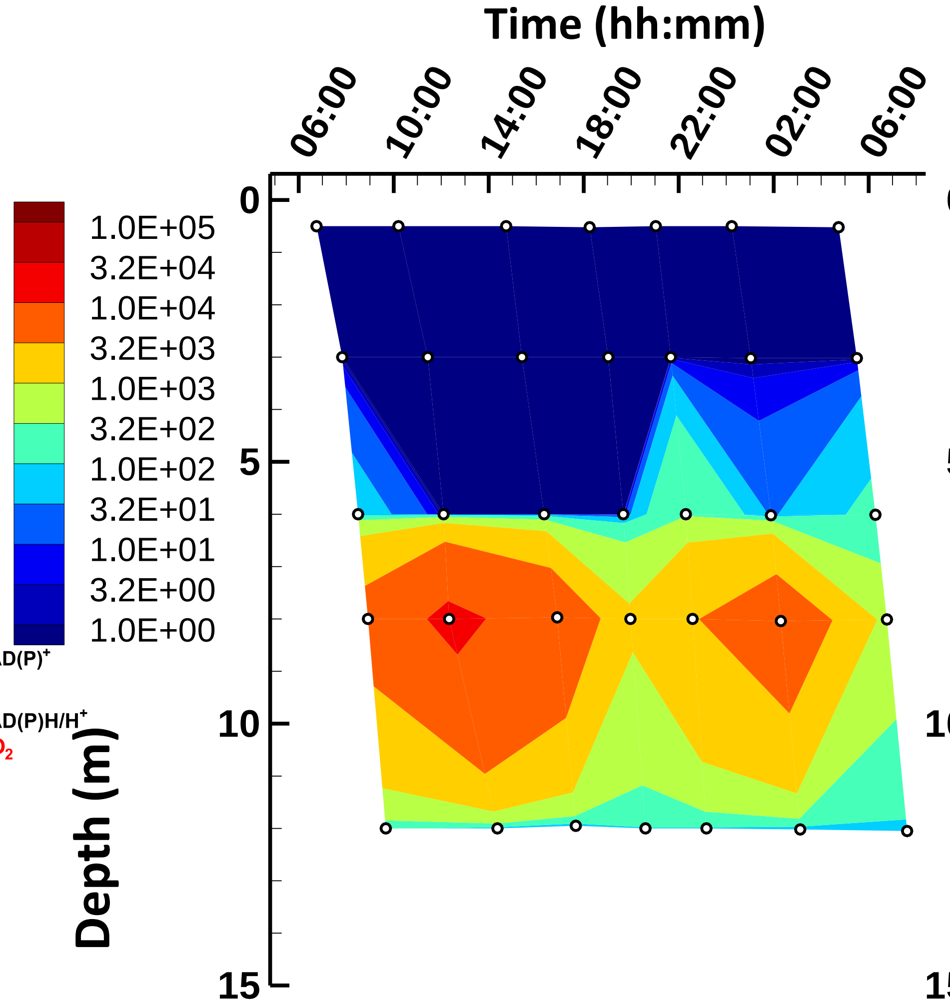
dsr: dissimilatory sulfite reductase

ATP-citrate lyase (anaerobic photosynthesis)

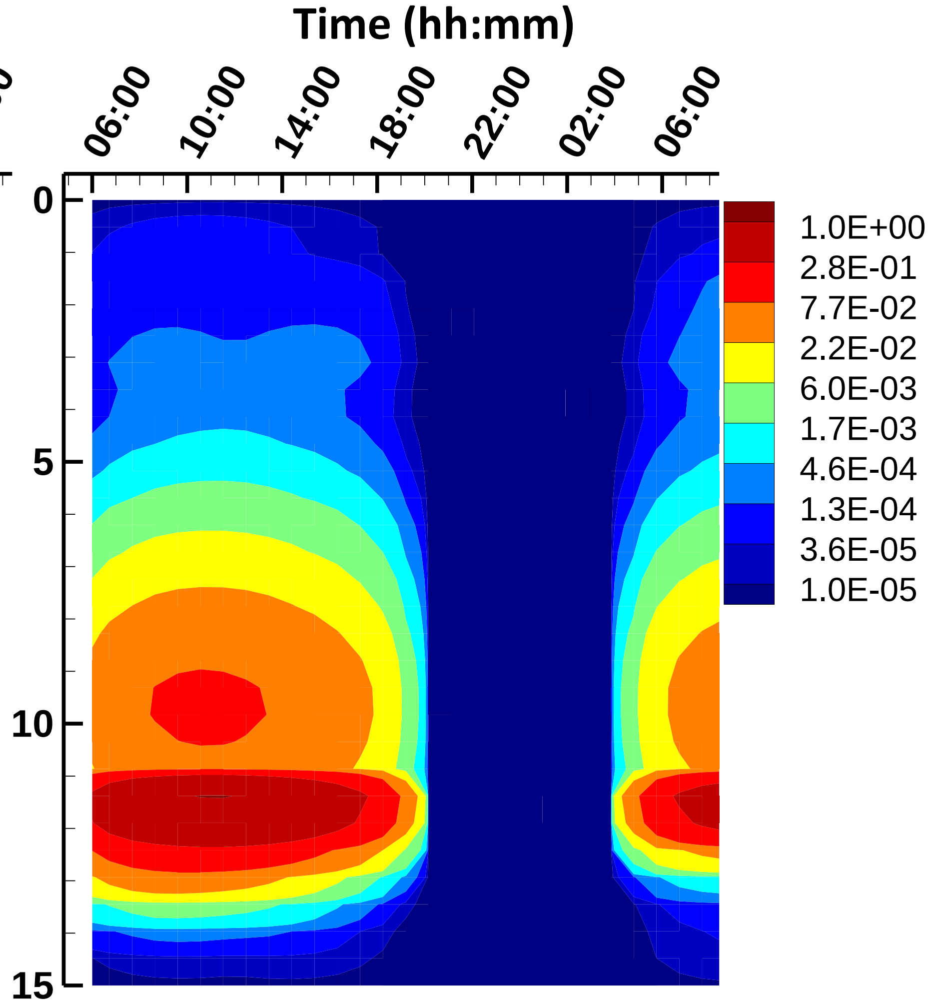
Reverse TCA Cycle



aclA Transcript Abundance



$r_{1,GSB}$ (µM d^{-1})*



*(3 d Interval Optimization)

Summary

- Maximizing entropy production (destruction of free energy) produces results that are similar to observations
  - Abiotic systems maximize instantaneous entropy production, while biotic system maximize entropy production over time using information.
  - Systems that coordinate information over space can increase global entropy production via coordination of function unless energy is degraded quickly abiotic (e.g., light)
- Model function not individuals (metabolic network)
- Replace parameters with control variables as much as possible
- The Control of PDE problem is computationally challenging
  - Faster solution approaches?
  - Different model formulation: Trait-based model optimized by MEP?

Acknowledgements

Collaborators:

Mick Follows
Ashley Bulseco
Nuria Fernández González
Chris Algar
Jeremy Rich
Anne Giblin
Aboozar Tabatabai
Gretta Serres
Amy Smith

Students:

Hilary Smith
Amy Snyder
Alice Carter
George Allen
Joe Hakam
Meg Yang
Petra Byl

RA's:

Suzanne Thomas
Jane Tucker
Rich McHorney
Stef Strebel
Stef Oleksyk

Funding

National Science Foundation — WHERE DISCOVERIES BEGIN
SIMONS FOUNDATION
GORDON AND BETTY MOORE FOUNDATION
Woods Hole Sea Grant
LTER
NASA