Use of receding horizon optimal control to solve MEP-based biogeochemistry problems

Joseph Vallino, Julie Huber and
Nuria Fernández González

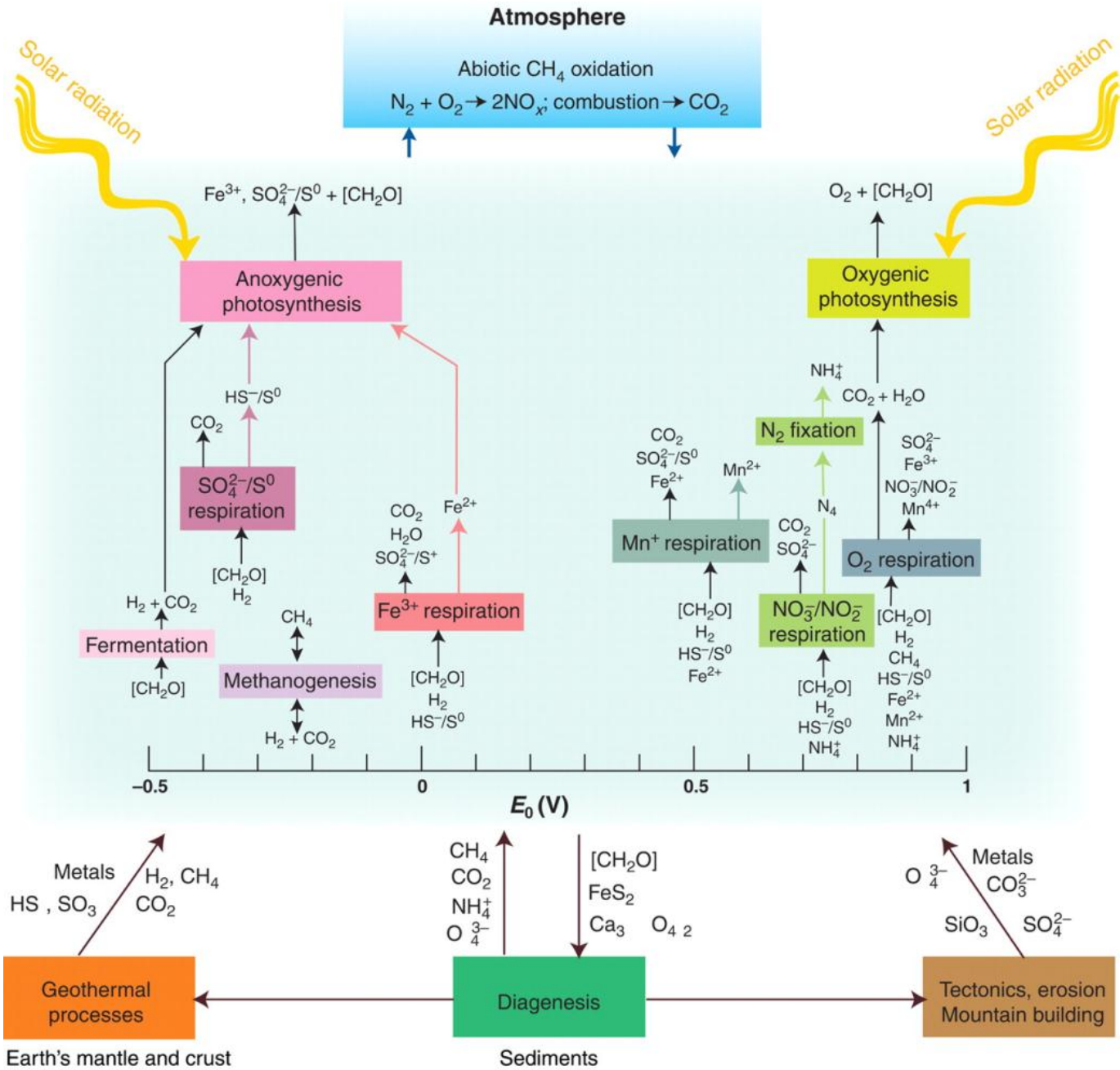
Marine Biological Laboratory
Woods Hole, MA USA

MEP Workshop, Sep 2011, Canberra AU

Objective: Understanding Biogeochemical Cycles

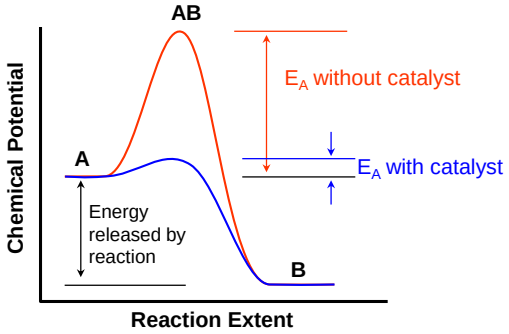
Most Reactions catalyzed by bacteria, the **molecular machines**

Falkowski et al., *Science* (2008)



Catalysts

Living systems as molecular machines



Without catalyst	With catalyst: $\mathfrak{S}$	Autocatalytic
$A \rightarrow B$	$A \xrightarrow{\mathfrak{S}} B$	$A + \sum \xrightarrow{\mathfrak{S}} B + \mathfrak{S}$

Hold [A] and [Σ] constant and supply $[\mathfrak{S}]_0$

		$[\mathfrak{S}] = [\mathfrak{S}]_0 e^{\acute{k}[A][\sum]t}$
$r = k[A]$	$r = \left(k + \acute{k}[\mathfrak{S}]_0\right)[A]$	$r = \left(k + \acute{k}[\mathfrak{S}]\right)[A][\sum]$

$$H_2S + 2O_2 \rightarrow H_2SO_4$$

$$H_2S + 2O_2 + \Sigma(CO_2, N, P, Fe, \ldots) \xrightarrow{\mathfrak{S}} \mathfrak{S} + H_2SO_4$$

Sulfur bacteria

$\Sigma(CO_2, N, P, Fe, \ldots)$
Information (genome)
} $\mathfrak{S}$ "Biological Structure"

Entropy and Information

Thermodynamic Entropy (Gibbs)

- Dispersal of energy, destruction of free energy

Information Entropy (Shannon)

- Measure of uncertainty, surprisal.
- Entropy of a bit: $k_B \log(2)$
- Can be connected to thermodynamic entropy (Boltzmann, Gibbs)

Useful Information (Structural complexity [Adami], Others?)

- Information correlated with physicality
- Allows manipulation of physical world
- Genes for sulfate reduction useless w/o sulfate present
- No (or trivial) thermodynamic burden

Example: a random sequence of amino acids

- Maximum Shannon entropy
- Zero useful information
- Thermodynamic entropy only slightly higher compared to a specified AA order

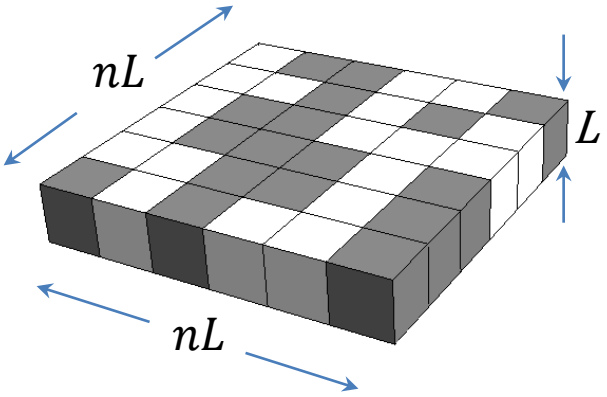
“Order” of living organisms

Are living systems “low entropy” structures due to their “order”?

That is, is the entropy of a living organism much lower than an appropriate composite of macromolecules (i.e., protein, DNA, lipids, etc.)?

Based on Morrison (1964), the answer appears to be No.

- You must consider not only the pattern, but the material the pattern is written in



$$F_{pattern} = n^2 k_B T \log 2$$

$$F_{material} = hN_A k_B T / M_W$$

$$\frac{F_{pat}}{F_{mat}} = \frac{M_W \log 2}{\rho L^3 N_A h}$$

	L	F_{pat}/F_{mat}
Visible pattern	0.5 mm	~10^{-18}
Microscopic pattern	1 micron	~10^{-10}
Electron microscopic pattern	0.1 micron	~10^{-7}
Molecular pattern	10 Å	~10^{-1}

This is also why the entropy of a messy desk and an ordered desk are equivalent

Cellular Order

How much free energy (or entropy) is associated with cellular order?

Based Bits:

- Free energy of a bit: $k_B T \log 2$ (Schneider 1991)
- Information in a bacterial cell (bits):
  - 10^7 bits or 2.8 x 10^{-14} J cell^{-1}: based on cellular DNA (Johnson 1970)
  - 10^{11} bits or 2.8 x 10^{-10} J cell^{-1}: based on all molecules (Morowitz 1955)
- Aerobic oxidation of a cell
  - 5 x 10^{-9} J cell^{-1}

Again, information synthesis is small to trivial compared to materials.

Measured entropy and enthalpy (Battley 1998)

$$CH_2O + aNH_3 \rightarrow \text{S} + eH_2O \qquad \text{100\% GE} \qquad \Delta_r G^{\circ\prime} = -1.4\ kJ\ mol^{-1}$$

"Nearly all the manifest visual and mechanical intricacy of organisms, like their apt behavior, turns out to be without quantitative thermodynamic importance. Morphology and ecology are … only small secondary properties of a fundamentally thermodynamic system, …"

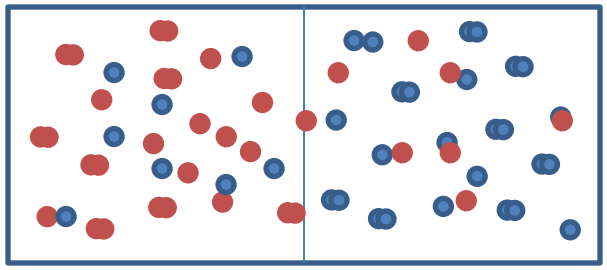
Morrison (1964)

Gibbs Paradox: Entropy & Useful Information

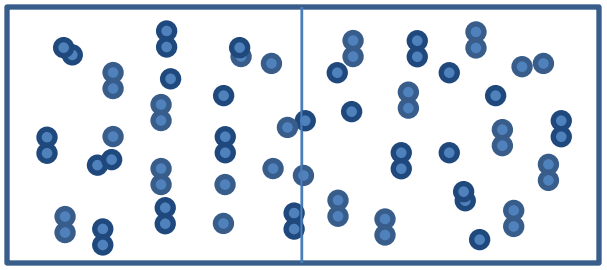
From Jaynes (1992)

Entropy of Mixing:

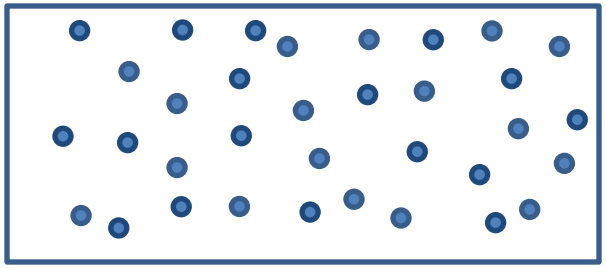
$\Delta S = nR \log(2)$



$\Delta S = 0$



- Fill the chamber with Argon.
- Introduce the *Superkalic* elements (yet to be discovered) :
  - Whifnium, Ar-1 diffuses into, Ar-2 does not.
  - Whafnium, Ar-2 diffuses into, Ar-1 does not.



The definition of entropy depends on useful information

Main Conjecture

Acquisition of useful information facilitates the destruction of free energy and the production of entropy

$$H_2S + 2O_2 + \Sigma(CO_2, N, P, Fe, \ldots) \xrightarrow{\mathfrak{S}} \mathfrak{S} + H_2SO_4$$ Sulfur bacteria

$\Sigma(CO_2, N, P, Fe, \ldots)$
Information (genome)
} $\mathfrak{S}$ "Biological Structure"

How does this work under transient conditions?

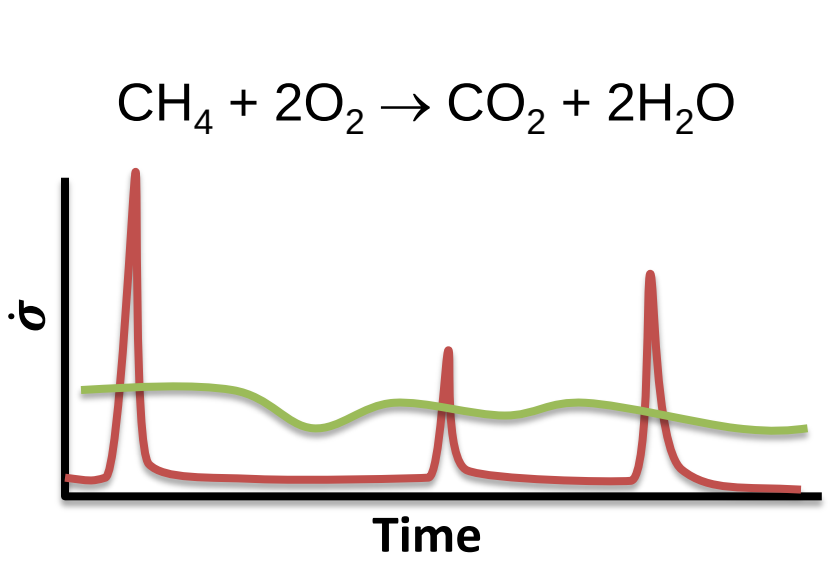
$$CH_2O + \gamma NH_3 \xrightarrow{\mathfrak{S}} \mathfrak{S} + eH_2O \qquad \dot{\sigma} \cong 0$$

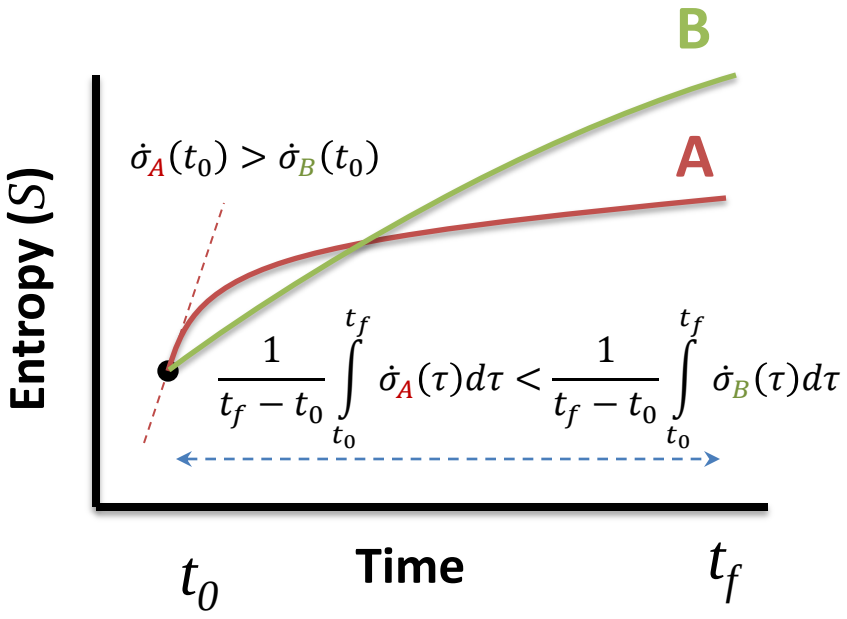
$$CH_2O + O_2 \xrightarrow{\mathfrak{S}} H_2CO_3 + \cdots \qquad \dot{\sigma} \gg 0$$

$$\mathfrak{S} + O_2 \xrightarrow{\mathfrak{S}} H_2CO_3 + \cdots \qquad \dot{\sigma} \gg 0$$

Maximizing $\dot{\sigma}$ instantaneously will not support $\mathfrak{S}$ synthesis

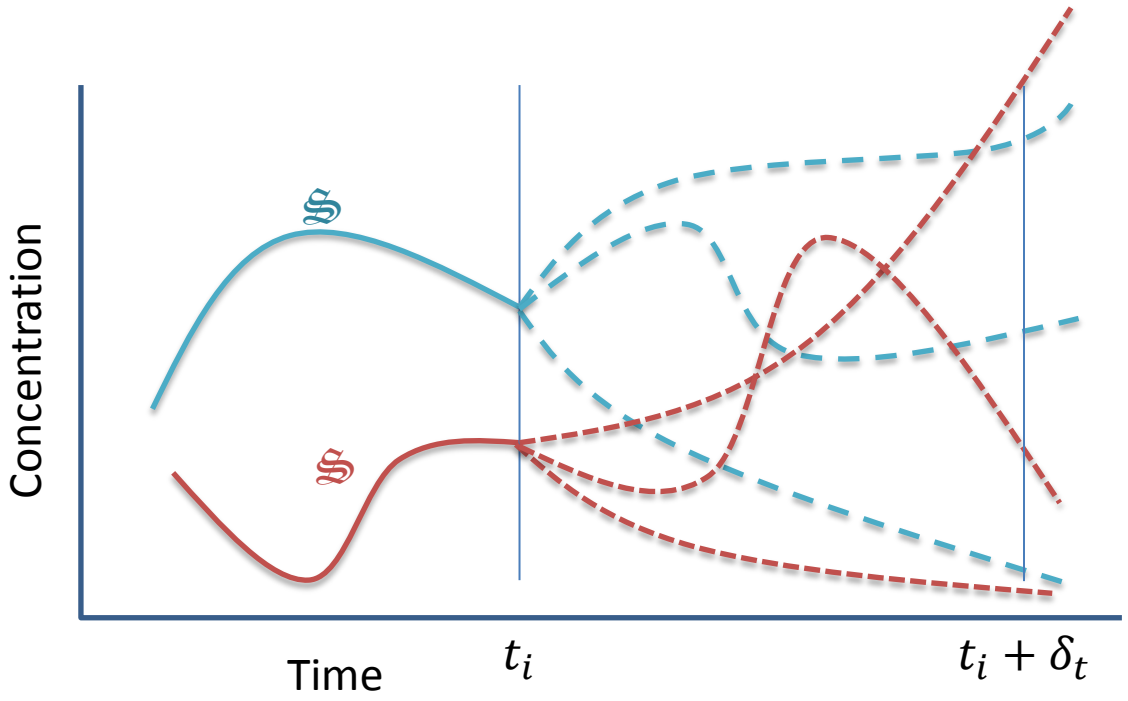
Average vs Instantaneous $\dot{\sigma}$





- "Useful information" stored in the metagenome allows living systems to predict future states and proceed along pathways that result in greater averaged entropy production than abiotic systems.
- **This is the only difference between biotic and abiotic systems**
- However, pathways for maximal averaged entropy production may be flanked by pathways of steepest descent (e.g., forest fires, invasive species).

Evolutionary Strategies



Presumably evolution will select for those temporal strategies (predictions) that maximizes internal entropy production over time. While infinite time would yield the greatest entropy production, this would require complete knowledge of future states.

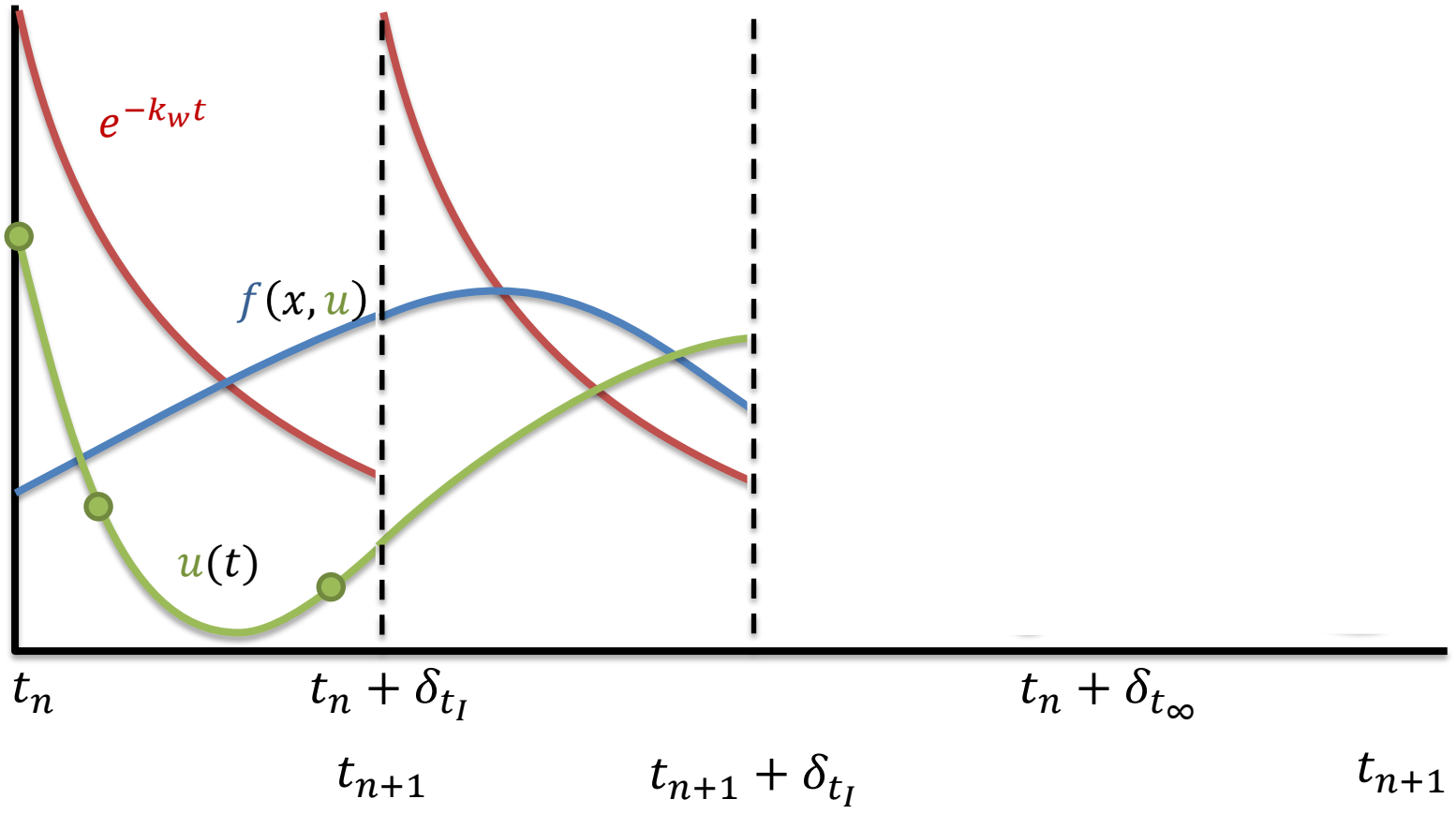
Entropy production will be maximized over time scales that the system can reliably predict over.

Implementation?

Receding Horizon Optimal Control

Optimization as information replacement

$$\max_{u(t)} J = \int_{t_i}^{t_i+\delta_{t_\infty}} f(x(\tau), u(\tau)) e^{-k_w \tau} d\tau \quad \text{subject to} \quad \frac{dx(t)}{dt} = g(x(t), u(t))$$



Experimental Test of $\dot{\sigma}$ Integration

Periodic cycling of energy input.

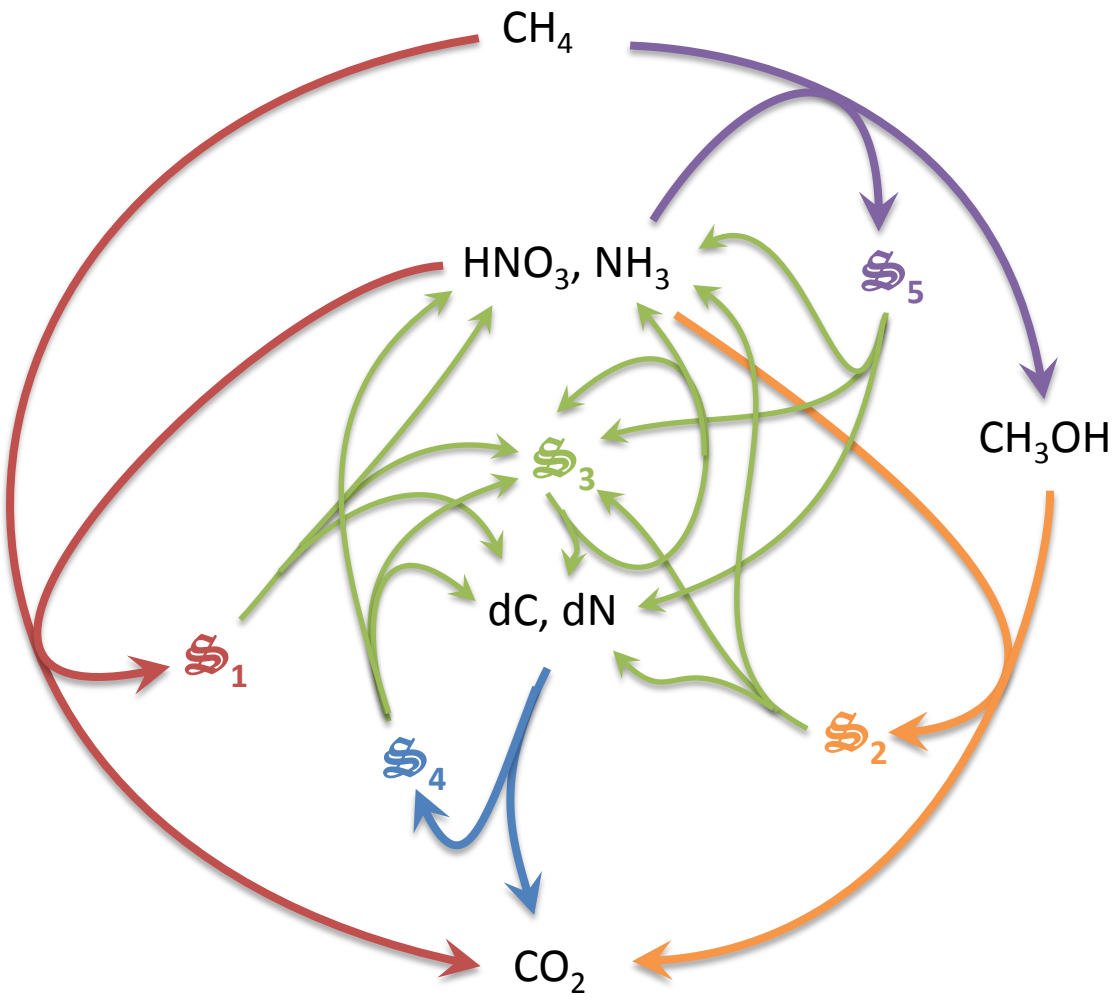
- 10 days CH_4 + air, then 10 days air only
- Continuous CH_4 + air (control)
- Chemostats at D = 0.1 d^{-1}

- Feed media: 50 μM HNO_3 + Trace elements, pH buffered at 6.8 via 10 mM phosphate.

Hypothesis: Entropy production will be same in both systems after reorganization

Methanotrophic Metabolic network

Five biological structures: $\mathbb{S}_i$



Methanotrophic Metabolic Reactions

$\mathbb{S}_i: \mathrm{CH}_{\alpha_i}\mathrm{O}_{\beta_i}\mathrm{N}_{\gamma_i}$

8 OCV

$$r_{1,1}: \mathrm{CH_4} + a_1\mathrm{HNO_3} + b_{11}\mathrm{O_2} \xrightarrow{\mathbb{S}_1} \varepsilon_1\mathbb{S}_1 + d_1\mathrm{H_2CO_3} + e_{11}\mathrm{H_2O}$$

$$r_{2,1}: \mathrm{CH_4} + a_1\mathrm{NH_3} + b_{21}\mathrm{O_2} \xrightarrow{\mathbb{S}_1} \varepsilon_1\mathbb{S}_1 + d_1\mathrm{H_2CO_3} + e_{21}\mathrm{H_2O}$$

$\varepsilon_1, \omega_{1,1}$

$$r_{1,2}: \mathrm{CH_3OH} + a_2\mathrm{HNO_3} + b_{12}\mathrm{O_2} \xrightarrow{\mathbb{S}_2} \varepsilon_2\mathbb{S}_2 + d_2\mathrm{H_2CO_3} + e_{12}\mathrm{H_2O}$$

$$r_{2,2}: \mathrm{CH_3OH} + a_2\mathrm{NH_3} + b_{22}\mathrm{O_2} \xrightarrow{\mathbb{S}_2} \varepsilon_2\mathbb{S}_2 + d_2\mathrm{H_2CO_3} + e_{22}\mathrm{H_2O}$$

$\varepsilon_2, \omega_{1,2}$

$$r_{i,3}: \mathbb{S}_i + a_{i3}\mathrm{O_2} \xrightarrow{\mathbb{S}_3} \varepsilon_3\mathbb{S}_3 + (1-\varepsilon_3)[\gamma_i(\varepsilon_3\mathrm{NH_3} + (1-\varepsilon_3)\mathrm{dN}) + (\varepsilon_3\mathrm{H_2CO_3} + (1-\varepsilon_3)\mathrm{dC}] + \varepsilon_3(\gamma_i - \gamma_3)\mathrm{NH_3} + b_{i3}\mathrm{H_2O} \text{ for } i = 1, \ldots, 5$$

ε_3

$$r_{1,4}: \mathrm{dC} + a_4\mathrm{dN} + b_4\mathrm{O_2} \xrightarrow{\mathbb{S}_4} \varepsilon_4\mathbb{S}_4 + d_4\mathrm{H_2CO_3} + e_4\mathrm{H_2O}$$

ε_4

$$r_{1,5}: \mathrm{CH_4} + a_5\mathrm{HNO_3} + b_{15}\mathrm{O_2} \xrightarrow{\mathbb{S}_5} \varepsilon_5\mathbb{S}_5 + d_5\mathrm{CH_3OH} + e_{15}\mathrm{H_2O}$$

$$r_{2,5}: \mathrm{CH_4} + a_5\mathrm{NH_3} + b_{25}\mathrm{O_2} \xrightarrow{\mathbb{S}_5} \varepsilon_5\mathbb{S}_5 + d_5\mathrm{CH_3OH} + e_{25}\mathrm{H_2}$$

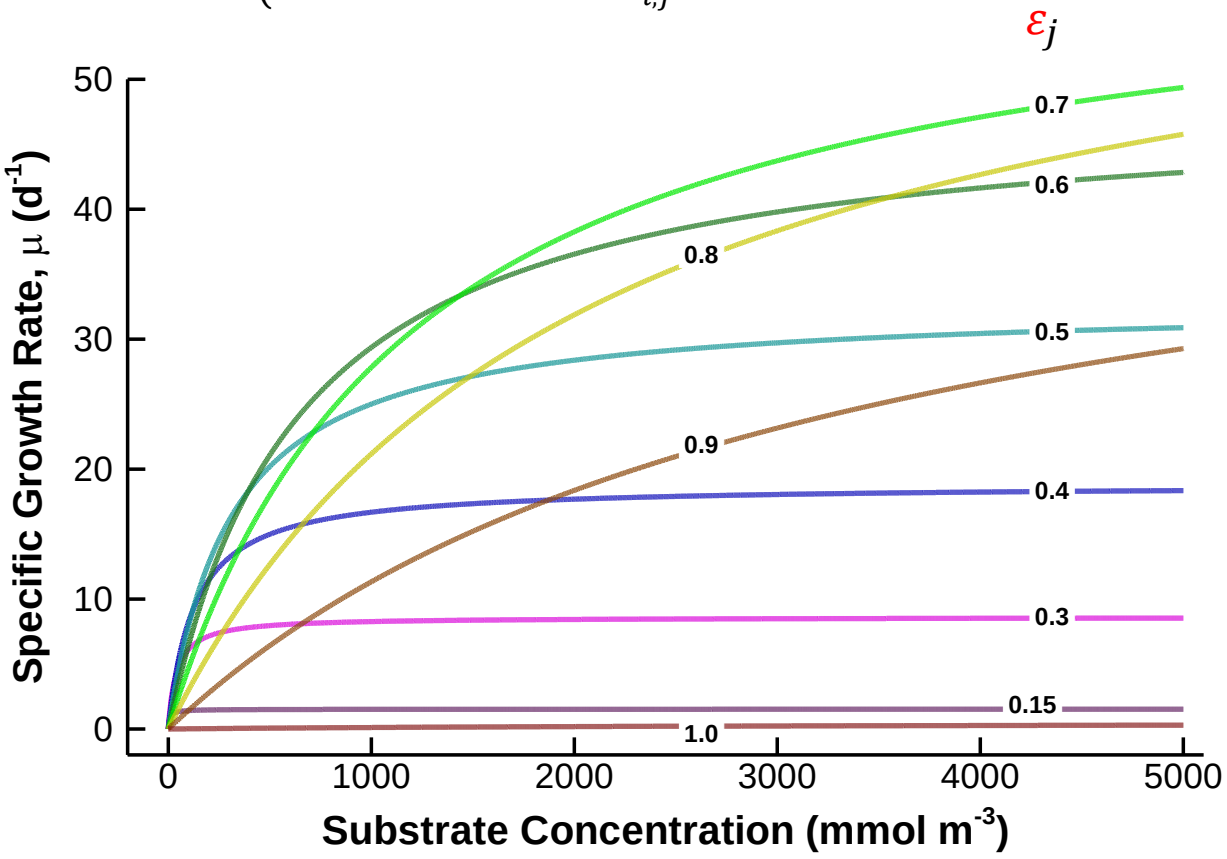
$\varepsilon_5, \omega_{1,5}$

ε_l : Tradeoff between catalyst synthesis and entropy production

Reaction Kinetics

$$r_{i,j} = \nu_j \varepsilon_j^2 \left(1 - \varepsilon_j^2\right) \prod_{k=1}^{n_C} \left(\frac{C_k}{C_k + \kappa_j \varepsilon_j^4} \right)^{\Lambda_{i,j,k}} \omega_{i-1,j} \prod_{l=i}^{n_j - 1} \left(1 - \omega_{l,j}\right) f_G(\Delta_r G_{r_{i,j}}) \mathbb{S}_j$$

$$f_G\left(\Delta_r G_{r_{i,j}}\right) = \begin{cases} 1 - e^{\chi_G \Delta_r G_{r_{i,j}}} & \Delta_r G_{r_{i,j}} \leq 0 \\ 0 & \Delta_r G_{r_{i,j}} > 0 \end{cases}$$



Receding Horizon OC Problem

$$\dot{\sigma} = -\frac{r\Delta_r G}{T}$$

Optimize over each δ_{t_∞} interval with future discounting:

$$\max_{\mathbf{u}(t_n)} J = -\frac{V}{T\delta_{t_\infty}} \int_{t_n}^{t_n+\delta_{t_\infty}} \left(\sum_i \sum_j r_{i,j} \Delta_r G_{r_{i,j}} \right) e^{-k_W \tau} d\tau$$

$$S.T.: \frac{d\left[\mathbf{C}^T, \mathbf{\mathfrak{S}}^T\right]^T}{dt} = \mathbf{f}(\mathbf{C}, \mathbf{\mathfrak{S}}, \mathbf{u}) \qquad \mathbf{u} = [\boldsymbol{\varepsilon}^T, \boldsymbol{\omega}^T]^T$$

$$\mathbf{0} \leq \mathbf{u} \leq \mathbf{1}$$

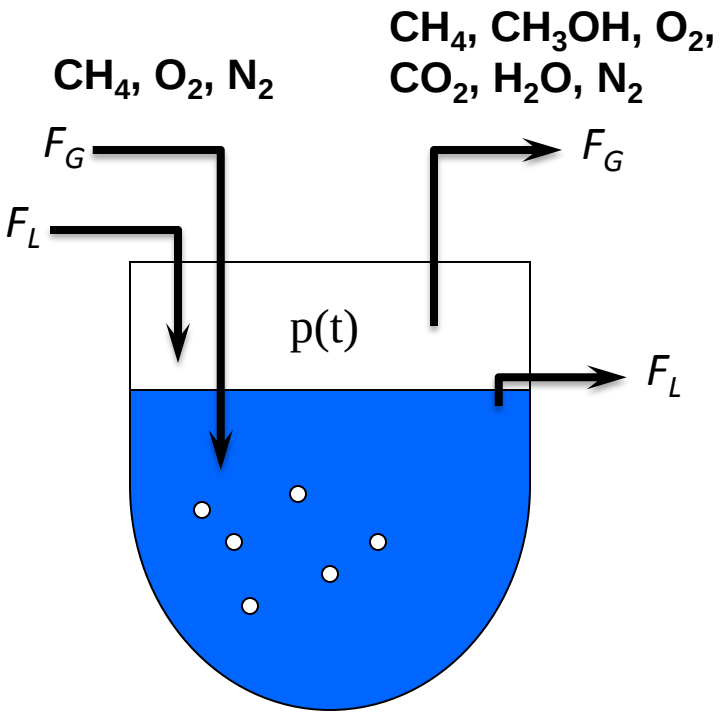
But integrate only over the δ_{t_I} interval:

$$\langle \dot{\sigma}(t_n) \rangle = -\frac{V}{T\delta_{t_I}} \int_{t_n}^{t_n+\delta_{t_I}} \left(\sum_i \sum_j r_{i,j} \Delta_r G_{r_{i,j}} \right) d\tau$$

- Minimum number of adjustable parameters (model DOF in OCV)
- Control variables naturally bounded.

State Space Model

$$\frac{d\left[\mathbf{C}^T, \mathfrak{S}^T\right]^T}{dt} = \mathbf{f}(\mathbf{C}, \mathfrak{S}, \mathbf{u})$$



CH_4	$\mathfrak{S}_1$
CH_3OH	$\mathfrak{S}_2$
H_2CO_3	$\mathfrak{S}_3$
dC	$\mathfrak{S}_4$
HNO_3	$\mathfrak{S}_5$
NH_3	
dN	
P_{CH4}	
P_{CH3OH}	
P_{CO2}	
P_{O2}	

Adjustable parameters: $\nu_1, \nu_2, \nu_3, \nu_4, \nu_5$

Examine two simulations:

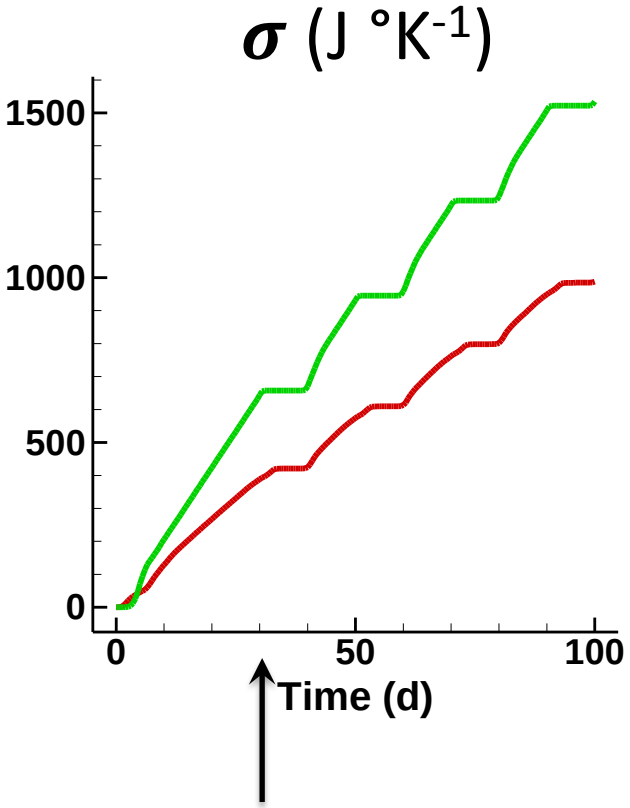
Both using: $\delta_{t_I} = 5\ \mathrm{d}, \delta_{t_\infty} = 20\ \mathrm{d}$

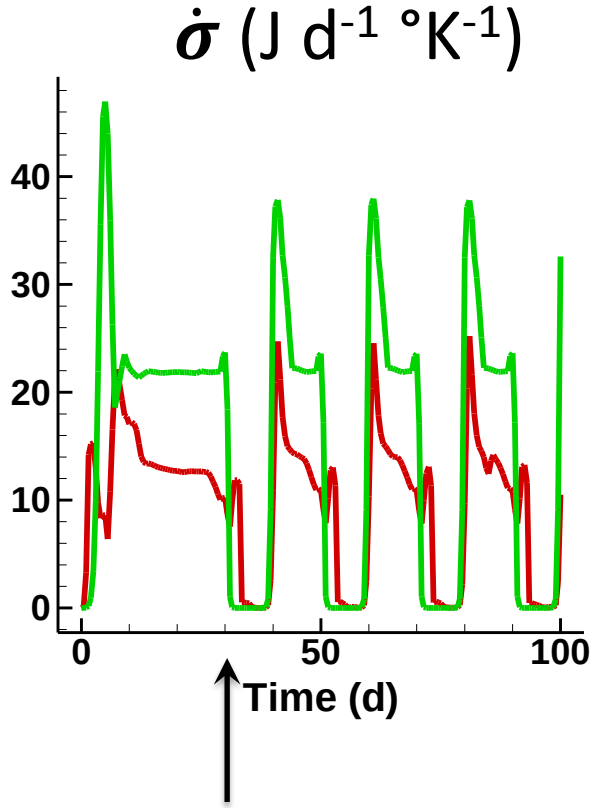
But for: 1) $k_w = 0\ \mathrm{d}^{-1}$
2) $k_w = 0.115\ \mathrm{d}^{-1}$

$$\int_{t_i}^{t_i+\delta_t} f(x(\tau), u(\tau)) e^{-k_w \tau} d\tau$$

Internal Entropy Production (J d^{-1} °K^{-1})

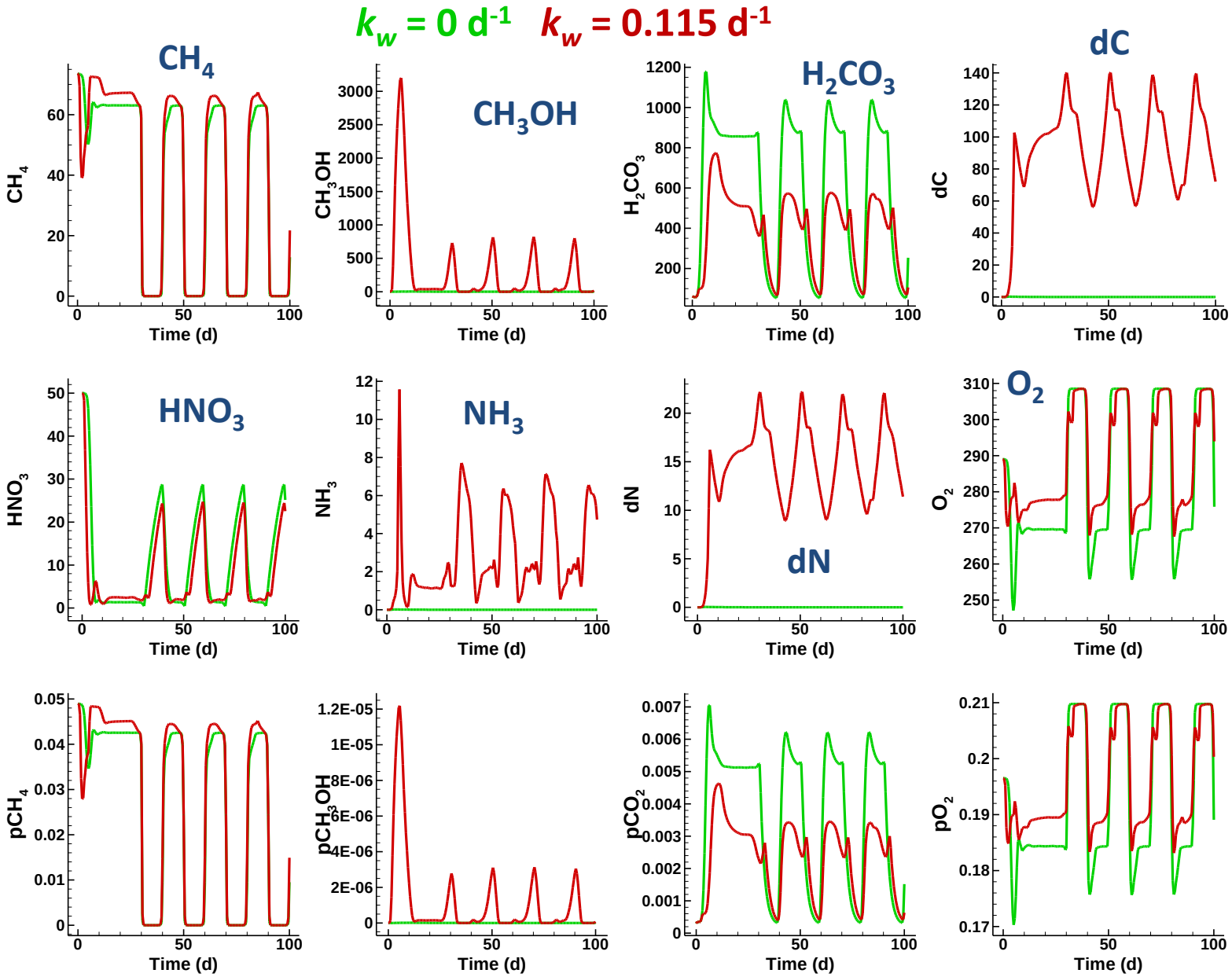
k_w = 0 d^{-1} k_w = 0.115 d^{-1}





Continuous CH_4 + air for 30 d, then cycling begins with 20 d period at day 30

Chemical Species (controls not simulated)

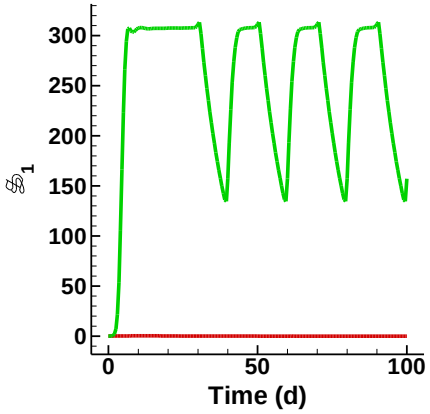


Concentrations: µM; Partial pressures: atm

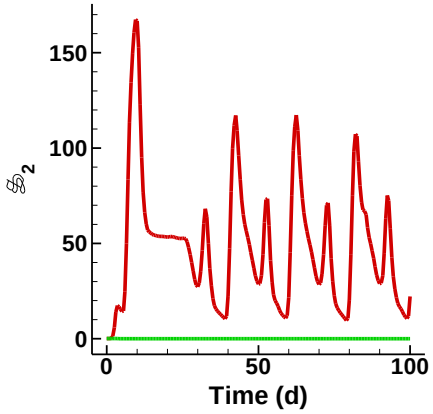
Biological Structures (μM)

$k_w = 0\ d^{-1}$ $k_w = 0.115\ d^{-1}$

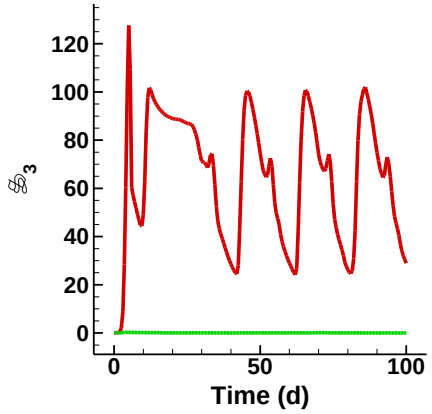
$\mathbb{S}_1$: $CH_4 \rightarrow CO_2$



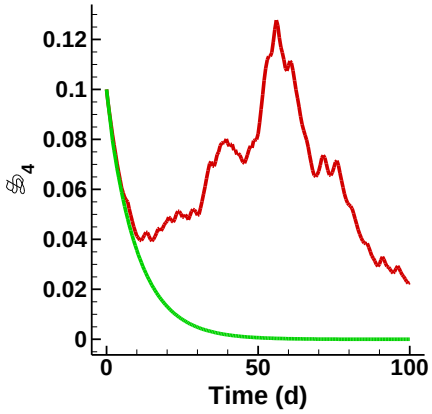
$\mathbb{S}_2$: $CH_3OH \rightarrow CO_2$



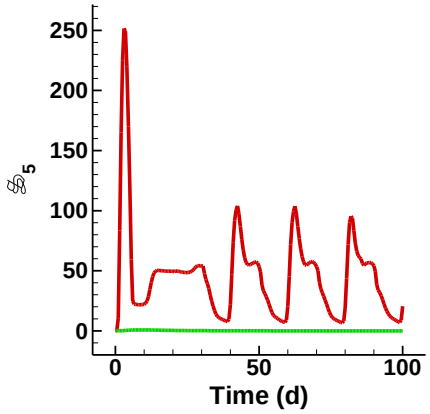
$\mathbb{S}_3$: $\mathbb{S}_i \rightarrow dN + NH_3 + CO2$



$\mathbb{S}_4$: $dC + dN \rightarrow CO_2 + \mathbb{S}_4$

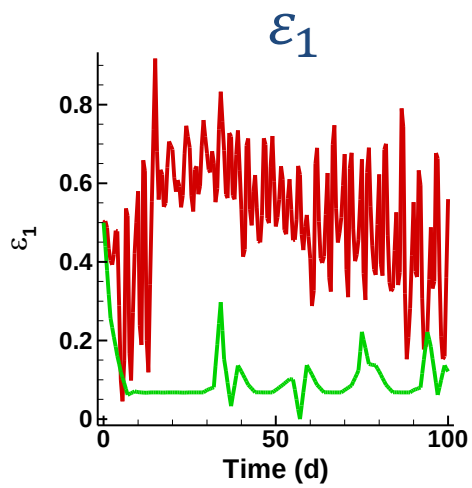


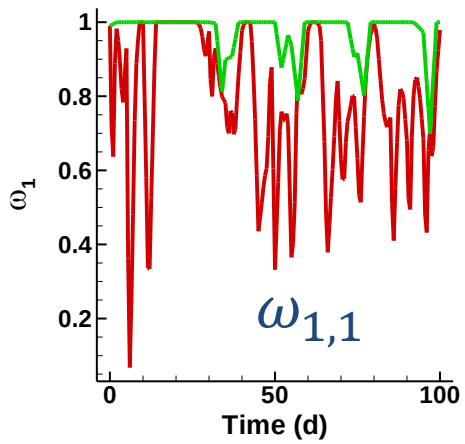
$\mathbb{S}_5$: $CH_4 \rightarrow CH_3OH$

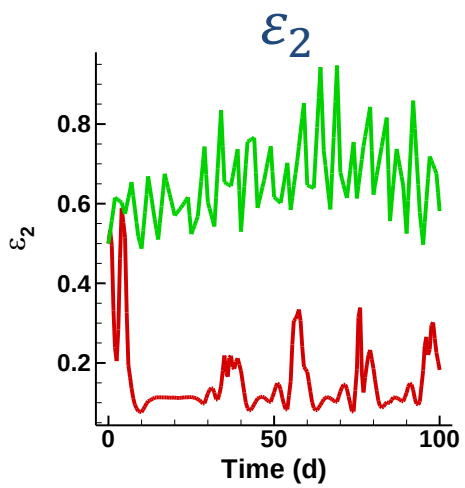


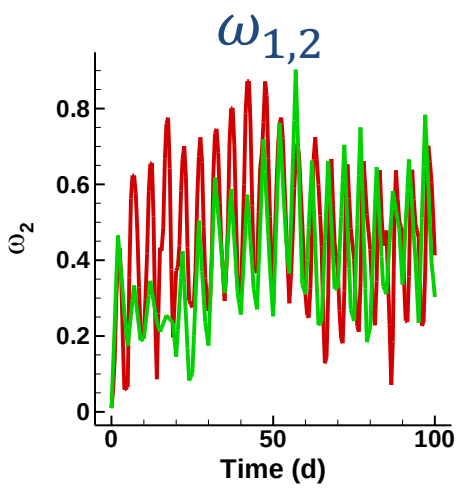
Optimal Control Solutions

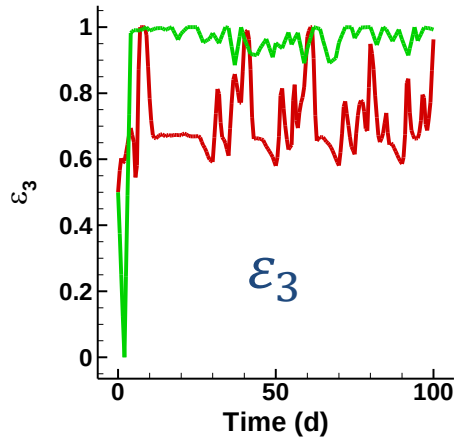
k_w = 0 d^{-1} k_w = 0.115 d^{-1}

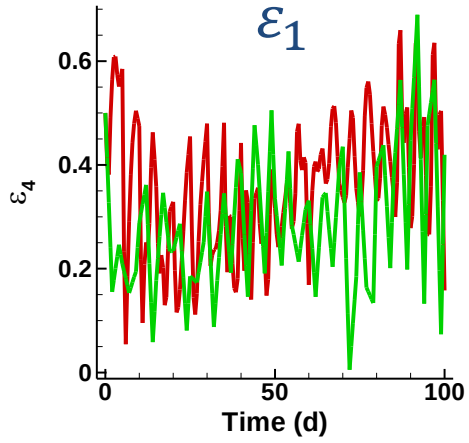


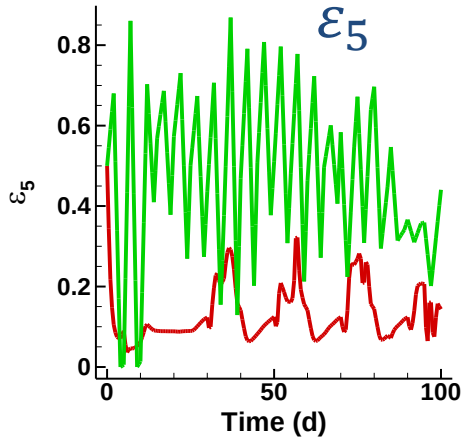


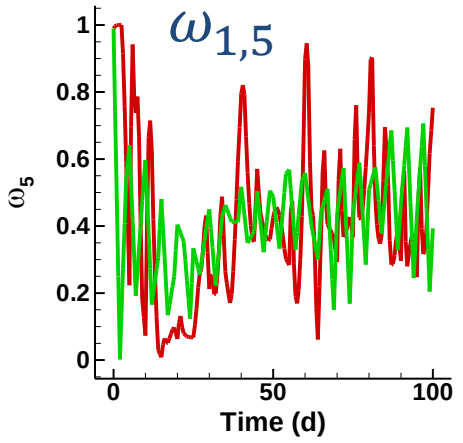




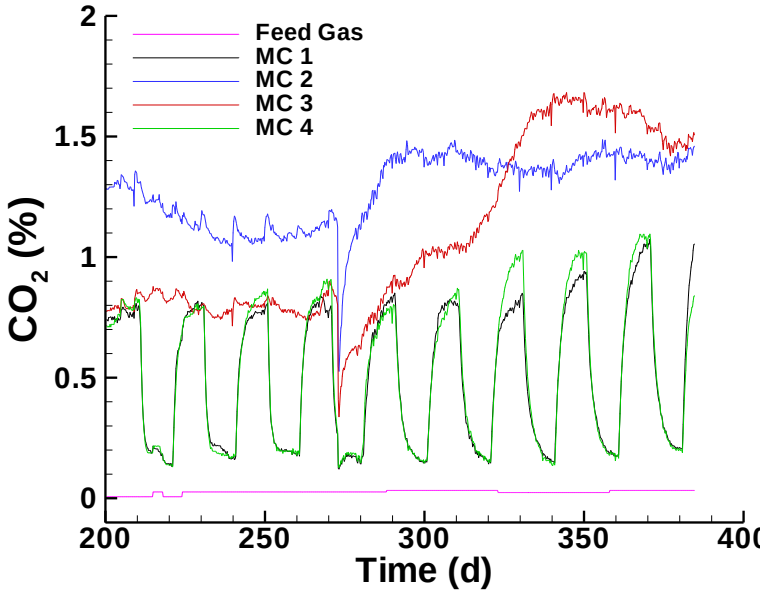


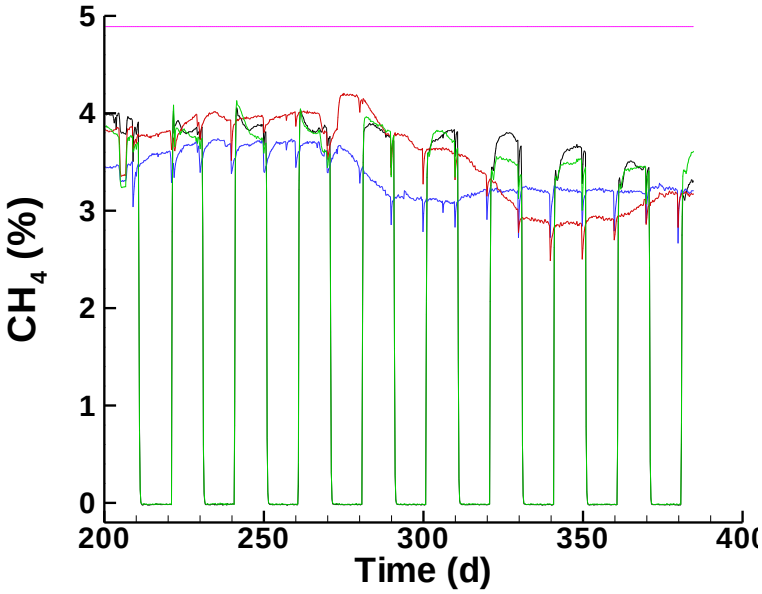


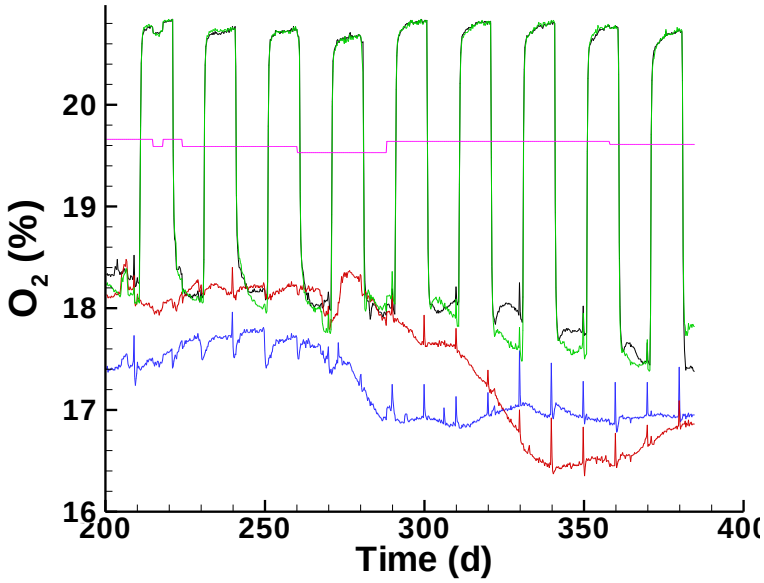




Experimental Results – Gas data

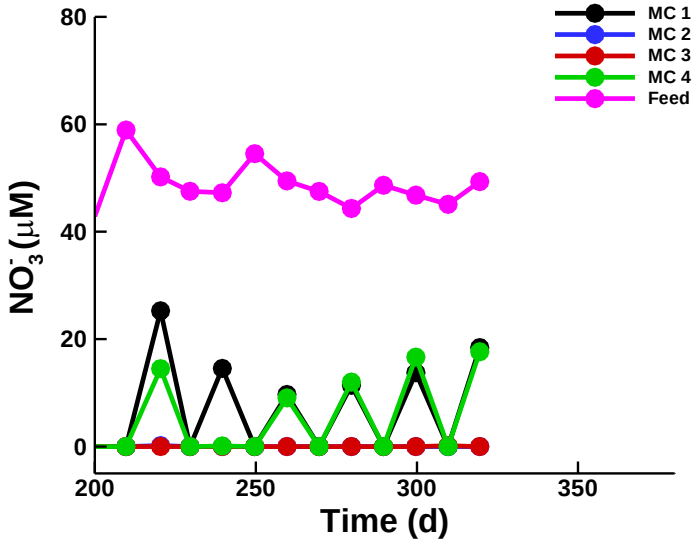


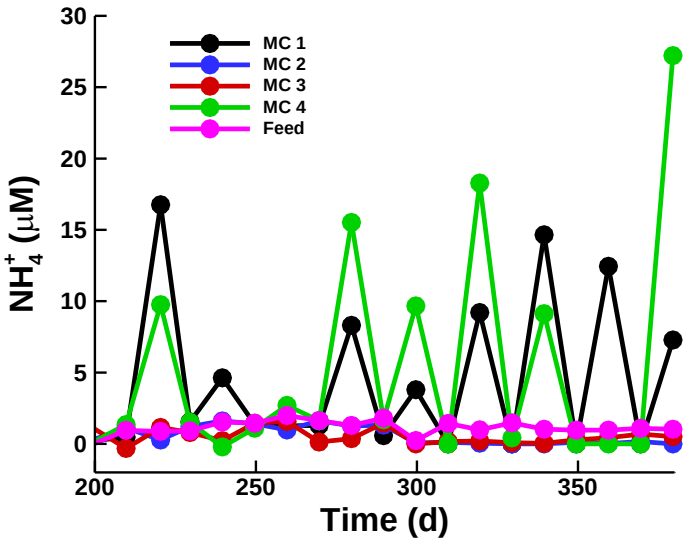


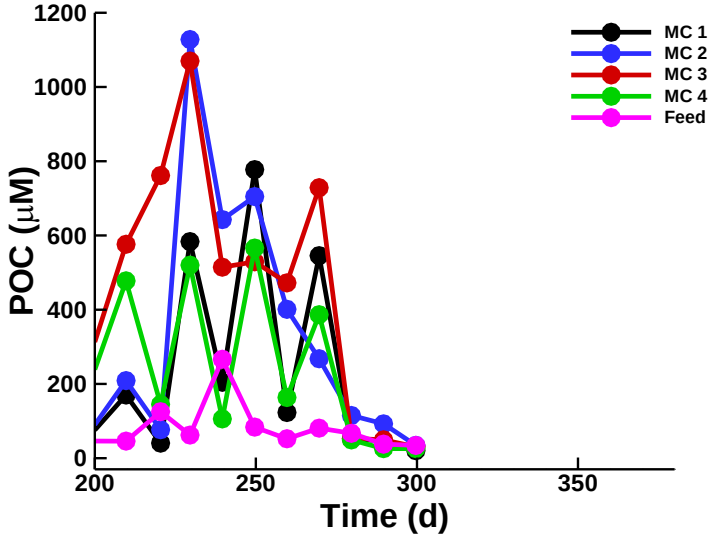


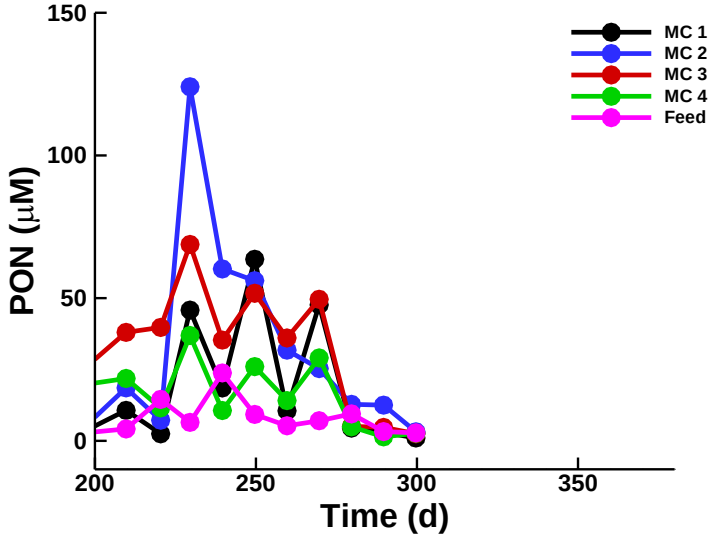
See http://ecosystems.mbl.edu/MEP

Experimental Results – DIN and POM



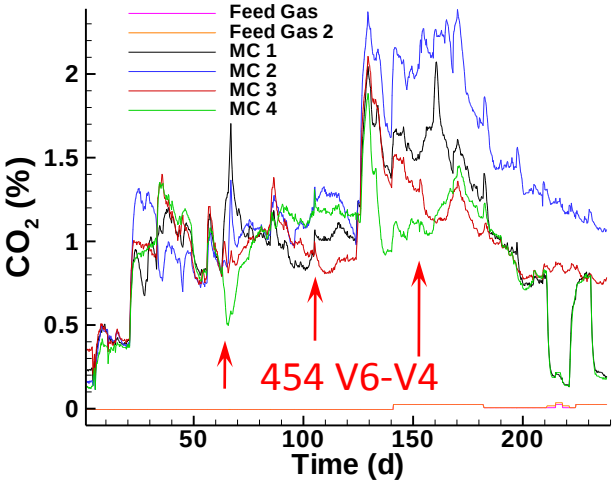


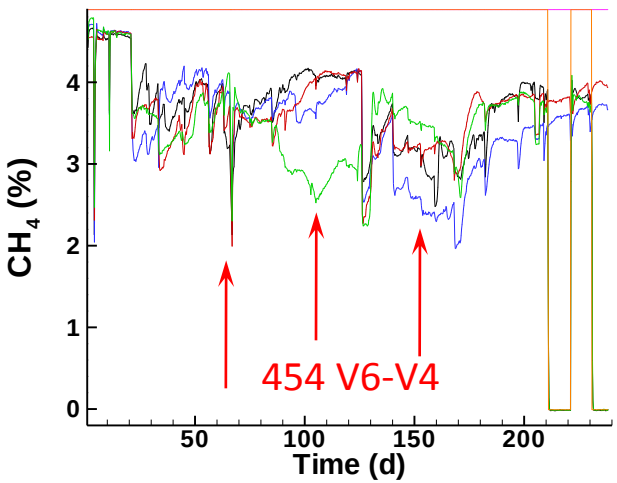


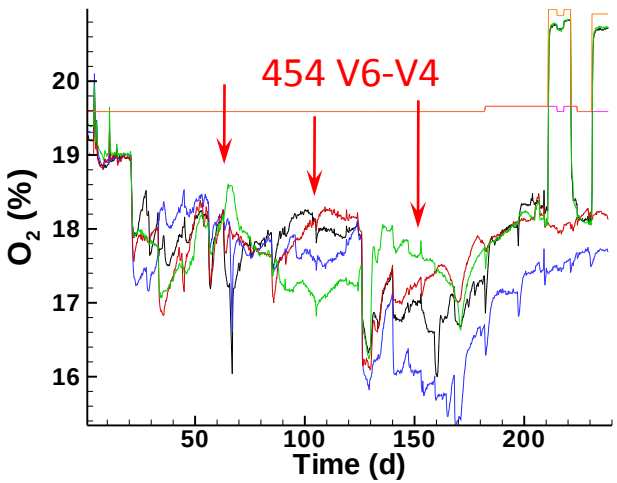


Microbial Community Molecular Analysis

- Community assessed by the V6 and V4 hypervariable regions of the 16S gene.
- ~400 bp amplicons sequenced using 454 pyrosequencer
- ~20,000 sequences obtained for each MC at each sample point.

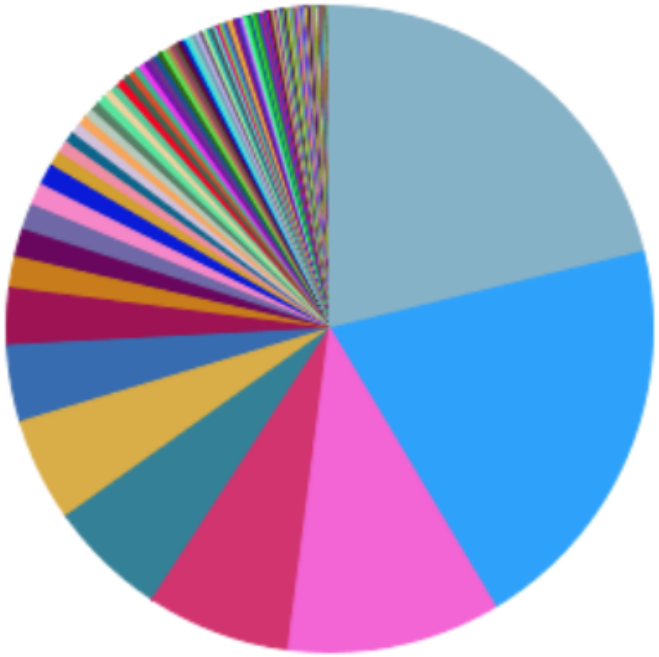






454 V6-V4, Day 62, MC1 and MC2

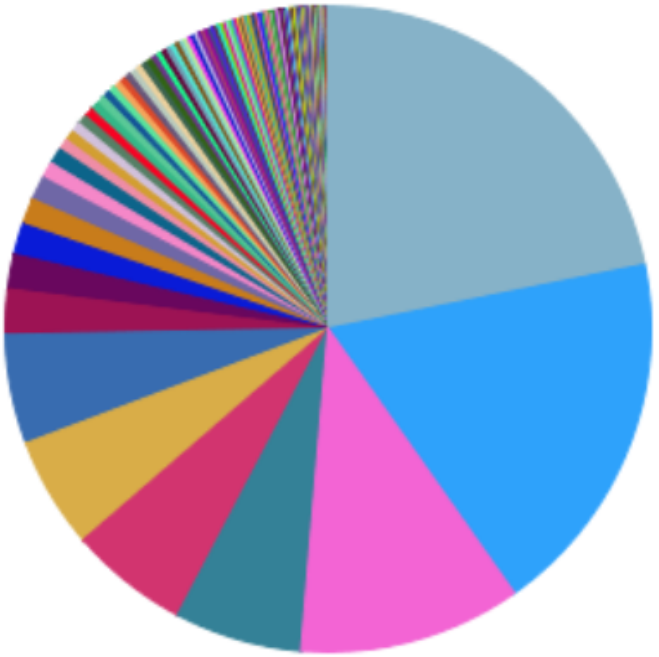
Microcosm 1



TOTAL: 11617

2459 * Methylococcaceae Methylomonas
2356 * Sinobacteraceae Nevskia
1230 * Methylococcaceae genus_NA
836 * Acidobacteriaceae Chloroacidobacterium
689 * Acidobacteriaceae genus_NA
602 * Sphingomonadaceae Sphingomonas
444 * family_NA genus_NA
330 * Chitinophagaceae genus_NA
181 * Saprospiraceae Lewinella

Microcosm 2

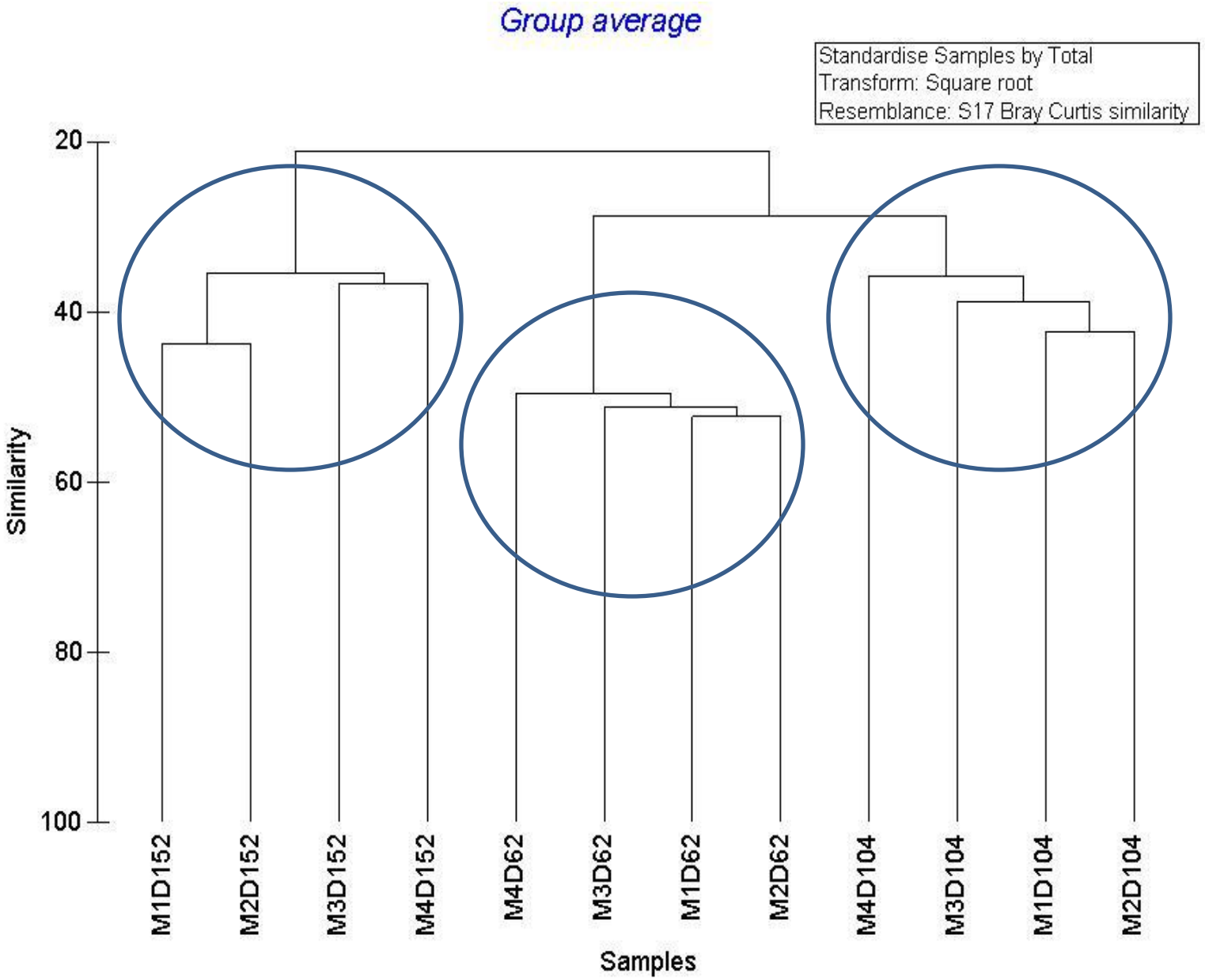


TOTAL: 13091

2855 * Methylococcaceae Methylomonas
2412 * Sinobacteraceae Nevskia
1453 * Methylococcaceae genus_NA
834 * Acidobacteriaceae genus_NA
782 * Acidobacteriaceae Chloroacidobacterium
728 * Sphingomonadaceae Sphingomonas
717 * family_NA genus_NA
294 * Chitinophagaceae genus_NA

Many Strategies Present (OTU's > 500)

Clustering Analysis



Summary-Conjecture

- Useful information allows biological systems to integrate entropy production over time, as opposed to abiotic systems.

- Receding horizon optimal control works for solving transient MEP problem.
  - But, choice of integration parameters $(\delta_{t_I}, \delta_{t_\infty}, k_w)$ important.

- Growth efficiency-based optimal control variables is promising.

$$\mathrm{CH_2O} + \varepsilon\gamma\mathrm{NH_3} \xrightarrow{\text{S}} \varepsilon\text{S} + (1-\varepsilon)\mathrm{H_2CO_3}, \text{ and } r(\varepsilon)$$

- Model agrees with observations, but has not been tuned.

- Experiment is yet to be poster child for MEP (but is consistent with model) and it still running.

Acknowledgements

Collaborators/RAs:

Jeremy Rich
Anne Giblin
Ken Forman
Stefanie Strebel
Jane Tucker
Stephanie Oleksyk
Rich McHorney

Students:

Jen Reimer
Sabrina Moreau
Mark Anderson
Angela Vincent
Kaitlyn Lucey
Whitney Eng
Megan Carpenter
Jessica Kunke
Jessica Bonsall
Andrea Coughlin
Alice Carter
George Allen

Funding

- NSF LTER (0423565)
- NSF Eng (0756562)
- NSF OCE (0852263)
- NSF ATB (0928742)

National Science Foundation
WHERE DISCOVERIES BEGIN