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





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



$F_{pattern} = n^2 k_B T \log F_{material} = h N_A k_B T / T$	$\frac{F_{pat}}{F_{mat}}$	$\frac{F_{pat}}{F_{mat}} = \frac{M_W \log 2}{\rho L^3 N_A h}$	
	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):
 - 10⁷ bits or 2.8 x 10⁻¹⁴ J cell⁻¹: based on cellular DNA (Johnson 1970)
 - 10¹¹ 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)

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

"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:





- 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

 $\begin{array}{ccc} H_2S + 2O_2 + \Sigma(CO_2, N, P, Fe, ...) & \stackrel{\textcircled{S}}{\rightarrow} & \textcircled{S} + H_2SO_4 & Sulfur bacteria \\ & & & \\ \Sigma(CO_2, N, P, Fe, ...) & & & \\ & & & & \\ & & & \\ & & & & \\$

How does this work under transient conditions?

$$\begin{array}{ll} \mathrm{CH}_{2}\mathrm{O}+\gamma\mathrm{NH}_{3} \xrightarrow{\mathfrak{S}} \mathfrak{S}+e\mathrm{H}_{2}\mathrm{O} & \dot{\sigma}\cong 0 \\\\ \mathrm{CH}_{2}\mathrm{O}+\mathrm{O}_{2} \xrightarrow{\mathfrak{S}}\mathrm{H}_{2}\mathrm{CO}_{3}+\cdots & \dot{\sigma}\gg 0 \\\\ \mathfrak{S}+\mathrm{O}_{2} \xrightarrow{\mathfrak{S}}\mathrm{H}_{2}\mathrm{CO}_{3}+\cdots & \dot{\sigma}\gg 0 \end{array}$$

Maximizing $\dot{\sigma}$ instantaneously will not support **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

Time t_i $t_i + \delta_t$

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

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

$$ignored f(x, u)$$

$$ig$$

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 μ M HNO₃ + 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: \mathfrak{S}_i



Methanotrophic Metabolic Reactions $\mathfrak{S}_i: CH_{\alpha_i}O_{\beta_i}N_{\gamma_i}$

$$r_{1,1}: \operatorname{CH}_4 + a_1 \operatorname{HNO}_3 + b_{11} \operatorname{O}_2 \xrightarrow{\mathfrak{S}_1} \mathfrak{E}_1 \mathfrak{S}_1 + d_1 \operatorname{H}_2 \operatorname{CO}_3 + e_{11} \operatorname{H}_2 \operatorname{O}$$

$$r_{2,1}: \operatorname{CH}_4 + a_1 \operatorname{NH}_3 + b_{21} \operatorname{O}_2 \xrightarrow{\mathfrak{S}_1} \mathfrak{E}_1 \mathfrak{S}_1 + d_1 \operatorname{H}_2 \operatorname{CO}_3 + e_{21} \operatorname{H}_2 \operatorname{O}$$

$$\mathcal{E}_1, \omega_{1,1}$$

8 OCV

$$r_{1,2}: CH_3OH + a_2HNO_3 + b_{12}O_2 \xrightarrow{\mathfrak{S}_2} \mathfrak{S}_2 + d_2H_2CO_3 + e_{12}H_2O$$

$$r_{2,2}: CH_3OH + a_2NH_3 + b_{22}O_2 \xrightarrow{\mathfrak{S}_2} \mathfrak{S}_2 + d_2H_2CO_3 + e_{22}H_2O$$

$$\mathfrak{E}_2, \omega_{1,2}$$

$$r_{1,4}: \mathrm{dC} + a_4 \mathrm{dN} + b_4 \mathrm{O}_2 \xrightarrow{\mathfrak{S}_4} \mathfrak{S}_4 + d_4 \mathrm{H}_2 \mathrm{CO}_3 + e_4 \mathrm{H}_2 \mathrm{O} \qquad \mathfrak{E}_4$$

$$r_{1,5}: CH_4 + a_5 HNO_3 + b_{15}O_2 \xrightarrow{\mathfrak{S}_5} \varepsilon_5 \mathfrak{S}_5 + d_5 CH_3 OH + e_{15}H_2 O$$

$$r_{2,5}: CH_4 + a_5 NH_3 + b_{25}O_2 \xrightarrow{\mathfrak{S}_5} \varepsilon_5 \mathfrak{S}_5 + d_5 CH_3 OH + e_{25}H_2$$

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

ε_i : Tradeoff between catalyst synthesis and entropy production

Reaction Kinetics



Receding Horizon OC Problem

Optimize over each $\delta_{t_{\infty}}$ 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.:
$$\frac{d \left[\mathbf{C}^T, \boldsymbol{\mathfrak{S}}^T \right]^T}{dt} = \mathbf{f}(\mathbf{C}, \boldsymbol{\mathfrak{S}}, \mathbf{u}) \qquad \mathbf{u} = [\boldsymbol{\varepsilon}^T, \boldsymbol{\omega}^T]^T$$
$$\mathbf{0} \le \mathbf{u} \le \mathbf{1}$$

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

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





Examine two simulations:

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

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

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

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) $k_{\rm w} = 0 \, {\rm d}^{-1} \, k_{\rm w} = 0.115 \, {\rm d}^{-1}$ dC CH₄ H₂CO₃ 1200 140 3000 120 **CH**₃**OH** 1000 60 2500 100 800 HO²⁰⁰⁰ HO¹⁵⁰⁰ HO H₂C0₃ CH₄ 80 40 g 600 60 1000 400 40 20 500 E 20 200 0 **0**는 0 50 50 100 100 50 100 0 Ó 50 100 Time (d) Time (d) Time (d) Time (d) 310 t **0**2 **12**⊢ 50 -HNO₃ NH₃ 20 10 300F 40 8 290 15 **HNO** o^{∾ 280}} 30 ЧЧ Ч 6 10 20 270 4 dN 5 260 10 2 250 0 0 0 5 50 Time (d) 50 Time (d) 50 Time (d) Ó 0 100 50 Time (d) 100 100 100 Ó 0.05 0.21 1.2E-05 0.007 0.006 0.04 1E-05 0.2 0.005 H8E-06 + 0.03 H D d 0.02 O 0.004 0.003 **ဝို** 0.19 <u>т</u>6Е-06 Q4E-06 0.002 0.18 0.01 2E-06 0.001 ⁰ᡛ 0.17<u>- -</u> 0 ⁰ᡛ 0 100 100 100 50 50 Time (d) 50 100 50 0 Time (d) Time (d) Time (d)

Concentrations: µM; Partial pressures: atm

Biological Structures (μ M) $k_w = 0 d^{-1} k_w = 0.115 d^{-1}$



Optimal Control Solutions

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



Experimental Results – Gas data



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)

4/14/2011 VAMPS - Taxonomic Com _____ 53 * Phycisphaeraceae genus NA * Nitrosomonadaceae genus_NA 53 * family NA genus NA 47 47 * Rhodocyclaceae genus_NA 44 * Coxiellaceae genus_NA 41 * Chitinophagaceae Ravisolibacter 37 * family_NA genus_NA 36 * family_NA genus_NA * Acetobacteraceae Acidocella 34 33 * Planctomycetaceae Gemmata 32 * Methylococcaceae Methylobacter 30 * Subdivision3 genus NA 29 * Chitinophagaceae Ferruginibacter * family NA genus NA 27 26 * family NA genus NA * Rhodospirillaceae genus_NA 25 24 * Crenotrichaceae Crenothrix 24 * Sphingomonadaceae Novosphingobium * Methylocystaceae genus NA 23 23 * Haliangiaceae Haliangium 23 * family NA genus NA 23 * Rhodospirillaceae Azospirillum 23 * Cytophagaceae Flexibacter 21 * Rhodocyclaceae Propionivibrio 21 * family NA genus NA 21 * Chitinophagaceae Terrimonas 20 * Comamonadaceae Methylibium * Hyphomicrobiaceae genus NA 19 19 * Oxalobacteraceae Oxalobacter 17 * Beijerinckiaceae Beijerinckia * Holophagaceae genus_NA 17 16 * Rickettsiaceae Rickettsia * Caldilineaceae Caldilinea 16 * Coxiellaceae Aquicella 15 15 * family_NA genus_NA 15 * Unassigned Captivus 15 * Planctomycetaceae Schlesneria 14 * Methylocystaceae Methylocystis * Sinobacteraceae Steroidobacter 14 14 * Simkaniaceae Rhabdochlamvdia 14 * Nitrospiraceae Nitrospira 13 * Sinobacteraceae genus NA 13 * Beijerinckjaceae Methylocapsa 13 * Hyphomicrobiaceae Hyphomicrobium 12 * Bradyrhizobiaceae Bradyrhizobium 12 * family_NA genus_NA 12 * Legionellaceae Legionella 12 * Neisseriaceae Chromobacterium

11 * Beijerinckiaceae Methylocella

11 t Verrue emierobia es a Asidimethylorilez vamps.mbl.edu/.../show_diffs_print.php...

paris	son Pie C
58	* Coxiellaceae genus_NA
57	* family_NA genus_NA
56	* Rhodocyclaceae genus_NA
47	* Subdivision3 genus_NA
46	* Nitrosomonadaceae genus_NA
45	* Methylococcaceae Methylosoma
43	* family_NA genus_NA
43	* Rhodocyclaceae Propionivibrio
43	* Planctomycetaceae Zavarzinella
42	* family_NA genus_NA
41	* Chitinophagaceae Flavisolibacter
40	* Chitinophagaceae Ferruginibacter
37	* Methylococcaceae Methylobacter

- 35 * Simkaniaceae Rhabdochlamvdia
- 35 Sphingomonadaceae Novosphingobium
- 34 Rhodospirillaceae Azospirillum
- Hyphomicrobiaceae genus_NA 31
- 31 Burkholderiaceae Chitinimonas
- 29 family_NA genus_NA
- 27 family_NA genus_NA

- 27 Acetobacteraceae Acidocella
- 25 Methylocystaceae genus_NA
- 24 Cytophagaceae Flexibacter
- 23 Planctomycetaceae Gemmata
- 23 Rhodospirillaceae genus_NA
- 22 family NA genus NA
- 22 Rickettsiaceae Rickettsia
- Planctomycetaceae Schlesneria 22
- 21 Comamonadaceae Methylibium
- 21 Haliangiaceae Haliangium
- 20 family NA genus NA
- Verrucomicrobiaceae Acidimethylosilex 19
- 19 Sinobacteraceae Steroidobacter
- 18 * family NA genus NA
- 18 * Holophagaceae genus NA
- 18 Oxalobacteraceae Oxalobacter
- 18 * Coxiellaceae Aquicella
- 17 Crenotrichaceae Crenothrix
- Caldilineaceae Caldilinea 17
- 15 Chitinophagaceae Terrimonas
- 15 family_NA genus_NA
- 15 family_NA genus_NA
- 15 Sinobacteraceae genus_NA
- 15 Neisseriaceae Chromobacterium
- 15 Methylocystaceae Methylocystis
- 14 Anaerolinaceae genus_NA
- 12 Beijerinckiaceae Methylocella
- * Beijerinckiaceae Methylocapsa 12
- 11 * Legionellaceae Legionella
- 44 t Apotobactoracian genur, NA

2/5

VAMPS - Taxonomic Comparison Pie C... verrucomicropiaceae Aciui * Hyphomonadaceae genus NA 11 * Burkholderiaceae Chitinimonas 11 * Coxiellaceae Coxiella * Verrucomicrobiaceae Prosthecobacter * Hyphomicrobiaceae Ancalomicrobium * family_NA genus_NA * family_NA genus_NA 9 * Anaerolinaceae genus_NA * Cytophagaceae Cytophaga 9 * Rhodocyclaceae Ferribacterium 8 * Bdellovibrionaceae Bdellovibrio 8 * Opitutaceae Opitutus 8 * Acetobacteraceae genus_NA 8 * Hyphomicrobiaceae Filomicrobium 7 * family NA genus NA 6 * Ruminococcaceae Acetivibrio 6 * Planctomycetaceae Pirellula 6 * Verrucomicrobiaceae genus NA 6 * Holophagaceae Geothrix 6 * family_NA genus_NA 6 * family_NA genus_NA 6 * Sphingomonadaceae Sphingopyxis 6 * Polyangiaceae Sorangium 6 6 * Bacillaceae Bacillus 5 * family_NA genus_NA * Bradyrhizobiaceae Bosea 5 5 * family_NA genus_NA * Oxalobacteraceae Undibacterium 5 * Comamonadaceae Rhodoferax 5 * family_NA genus_NA 5 * Methylophilaceae Methylophilus 5 Nitrosomonadaceae Nitrosospira 5 * Mycobacteriaceae Mycobacterium 5 * Holophagaceae Holophaga 5 * Porphyromonadaceae Paludibacter 4 * Rhodospirillaceae Oceanibaculum 4 * Bradyrhizobiaceae Afipia 4 * Caulobacteraceae Brevundimonas 4 * Saprospiraceae Haliscomenobacter 4 * family NA genus NA 4 * Acidobacteriaceae Solibacter 4 Hyphomicrobiaceae Pedomicrobium 4 4 * family NA genus NA 4 * Chromatiaceae genus NA 3 * Acidobacteriaceae Koribacter 3 Comamonadaceae Comamonas

3 * SAR116 genus NA

4/14/2011

11

10

10

10

10

10

9

9

8

8

8

8

7

7

7

7

7

6

6

6

6

5

5

5

5

5

5

5

5

4

4

4

4

4

4

4

3

3

3

3

3

3

3

4

4

10

10

- 3 * Desulfovibrionaceae Desulfovibrio
- 2 Thermomonosporaceae Actinoallomurus

vamps.mbl.edu/.../show_diffs_print.php...

Acetopacteraceae denus INA * family_NA genus_NA * Nitrospiraceae Nitrospira * family_NA genus_NA Bradyrhizobiaceae Bradyrhizobium * Hyphomicrobiaceae Hyphomicrobium Verrucomicrobiaceae genus_NA * Rhodospirillaceae Oceanibaculum * Coxiellaceae Coxiella * Hyphomonadaceae genus_NA * Hyphomicrobiaceae Ancalomicrobium * Rhodocyclaceae Ferribacterium * Cytophagaceae Cytophaga * Porphyromonadaceae Paludibacter * Bradyrhizobiaceae Bosea * Parachlamydiaceae Parachlamydia * family NA genus NA * Unassigned Captivus * Verrucomicrobiaceae Prosthecobacter * Caulobacteraceae Brevundimonas * Holophagaceae Holophaga * Planctomycetaceae Pirellula * Opitutaceae Opitutus * family_NA genus_NA * Ruminococcaceae Acetivibrio Caulobacteraceae Asticcacaulis Parachlamydiaceae Neochlamydia * Bdellovibrionaceae Bdellovibrio * family_NA genus_NA * family NA genus NA * Holophagaceae Geothrix * Bradyrhizobiaceae genus_NA * Methylophilaceae Methylophilus * Beijerinckiaceae Beijerinckia * Comamonadaceae Rhodoferax * Oxalobacteraceae Undibacterium * Parachlamydiaceae Protochlamydia * Oxalobacteraceae Collimonas * family NA genus NA * Burkholderiaceae Ralstonia * Mycobacteriaceae Mycobacterium * family NA genus NA * Saprospiraceae Haliscomenobacter * family NA genus NA * Polvangiaceae Sorangium * family NA genus NA * Rhodocyclaceae Azonexus 3 * family NA genus NA 3 * Acidobacteriaceae Solibacter 3 Bacteriovoraceae Peredibacter

3/5

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. \mathfrak{S} $\operatorname{CH}_2 0 + \varepsilon \gamma \operatorname{NH}_3 \xrightarrow{\mathfrak{S}} \varepsilon \mathfrak{S} + (1 - \varepsilon) \operatorname{H}_2 \operatorname{CO}_3$, 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