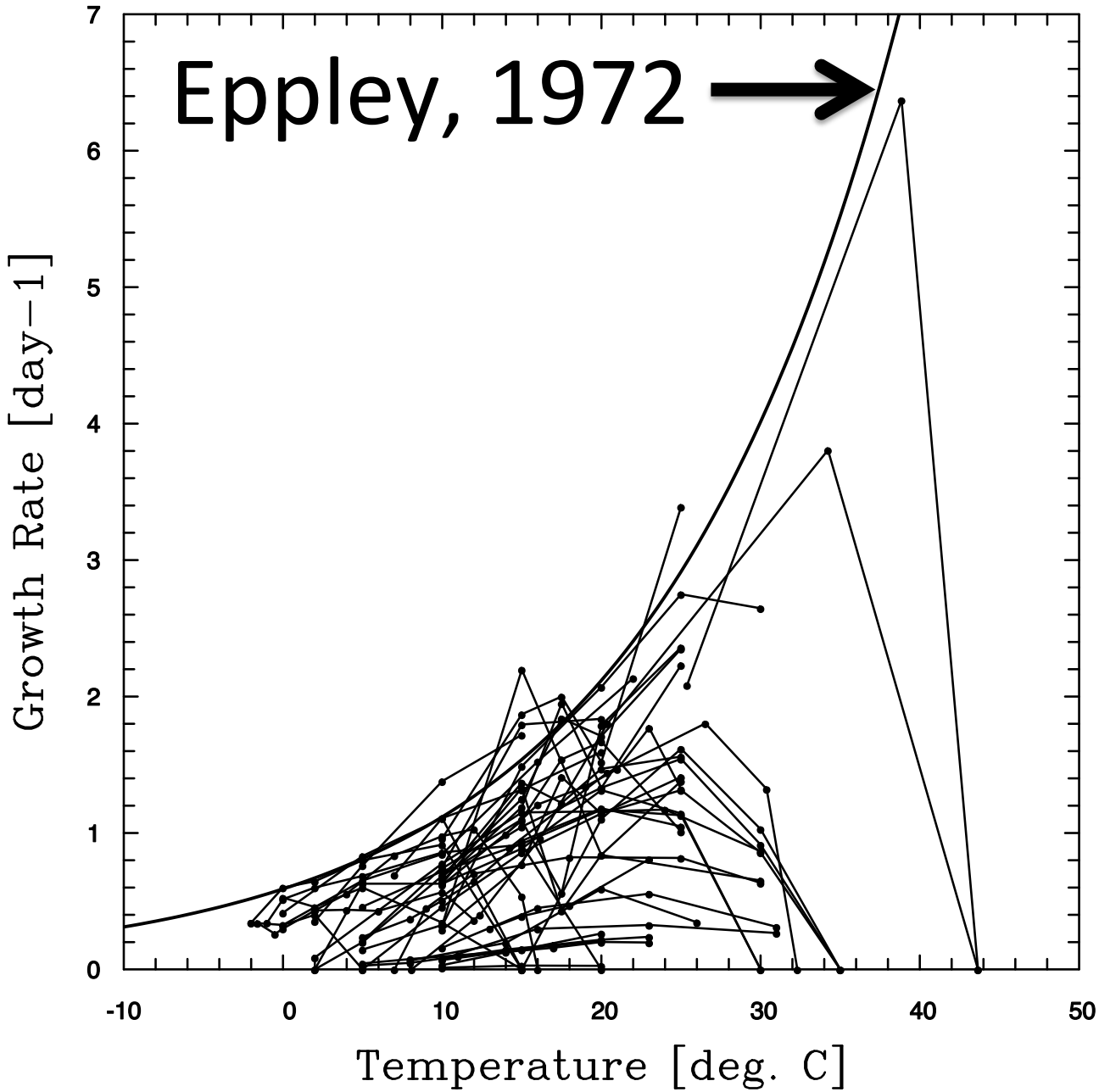
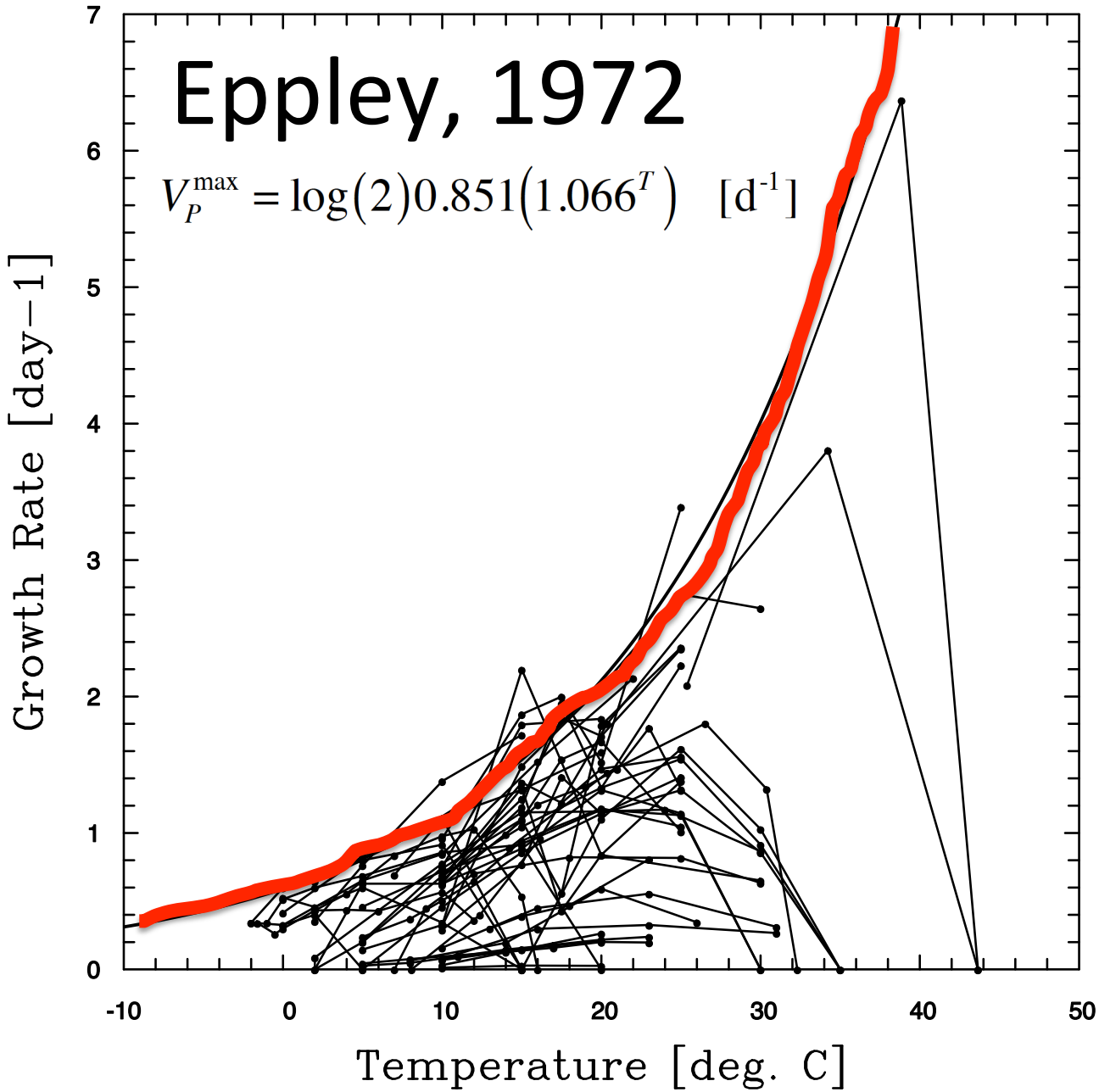
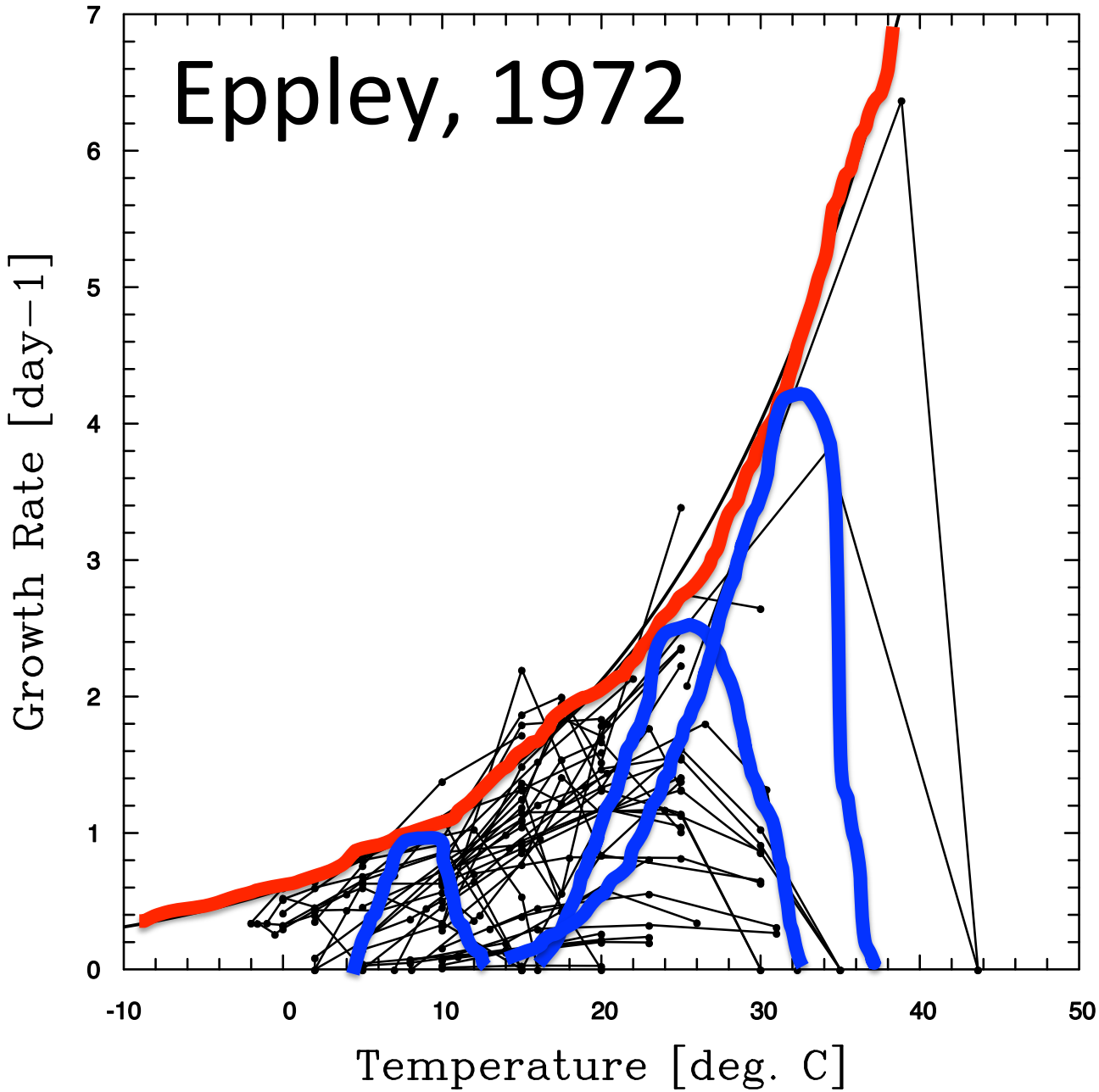
# Genetic Programming for Ocean Microbial Ecology and Biodiversity

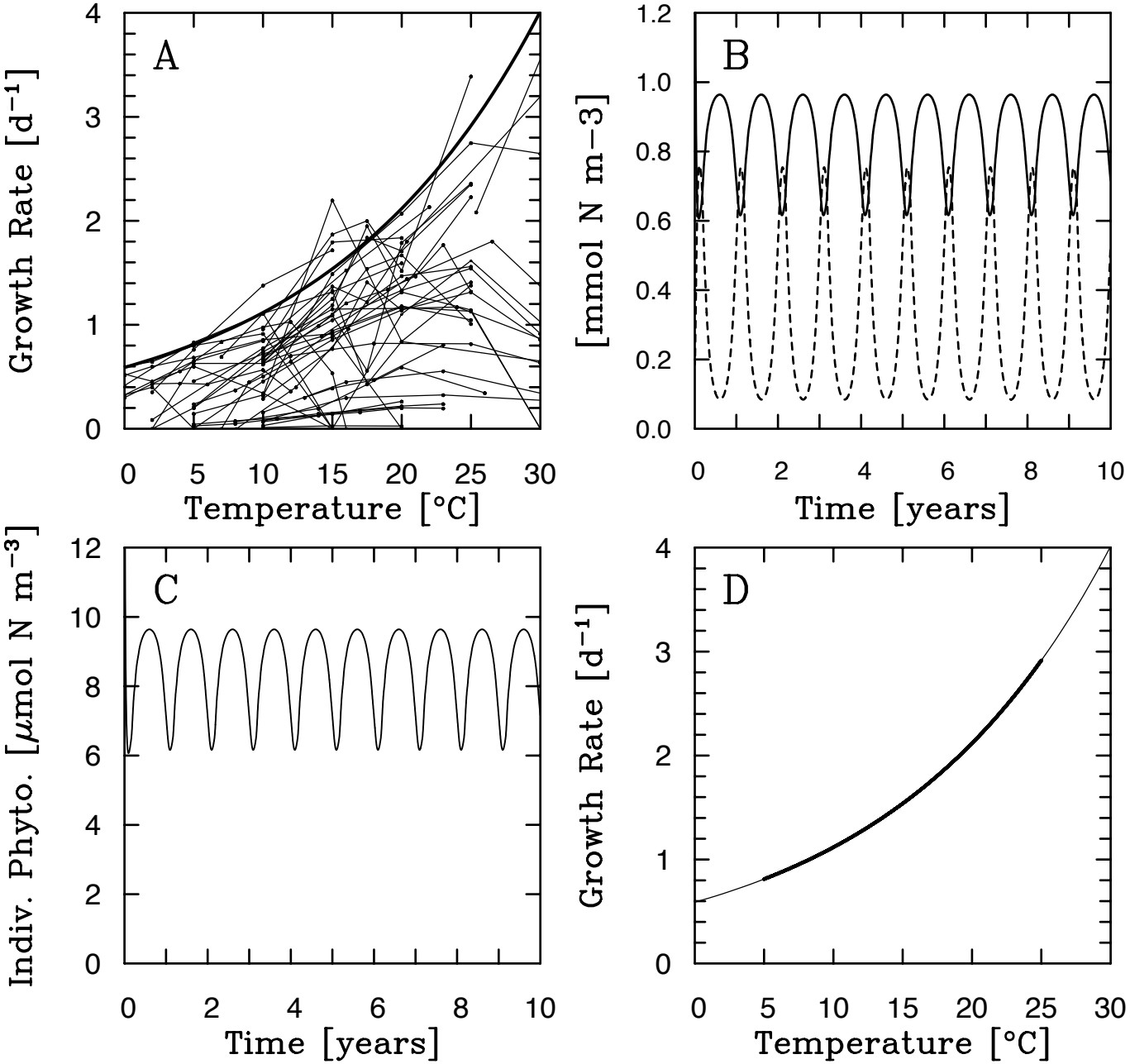
NASA Biodiversity and Ecological Forecasting Team Meeting
May 7-9, 2014

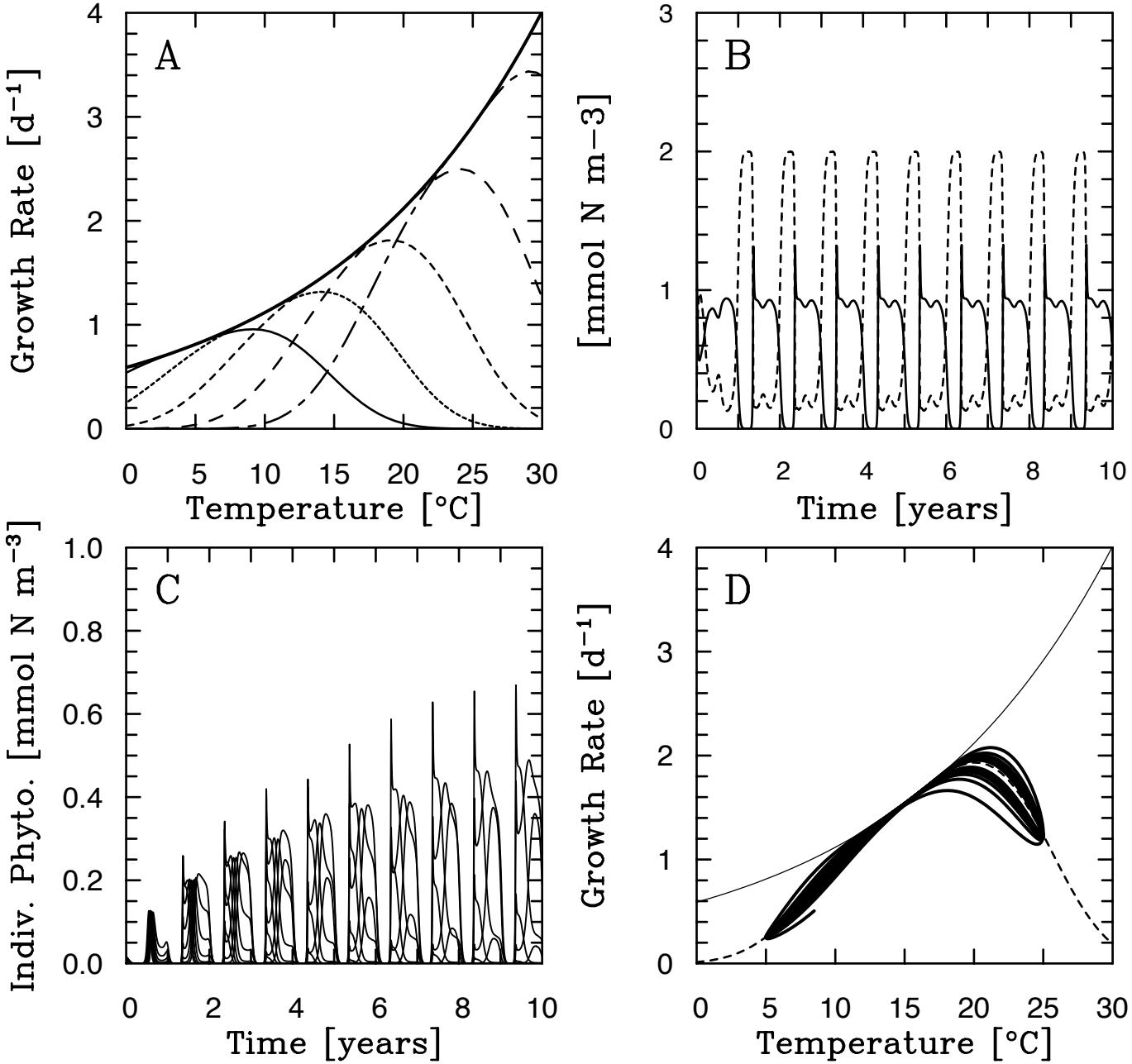
John R. Moisan
NASA/GSFC
Earth Science Division, Code 610.W
John.R.Moisan@nasa.gov


Eppley, 1972
Growth Rate [day−1]
Temperature [deg. C]
0
1
2
3
4
5
6
7
-10
0
10
20
30
40
50


Eppley, 1972
$V_P^{max} = \log(2)0.851(1.066^T)$ [d$^{-1}$]
Growth Rate [day−1]
Temperature [deg. C]
0
1
2
3
4
5
6
7
-10
0
10
20
30
40
50

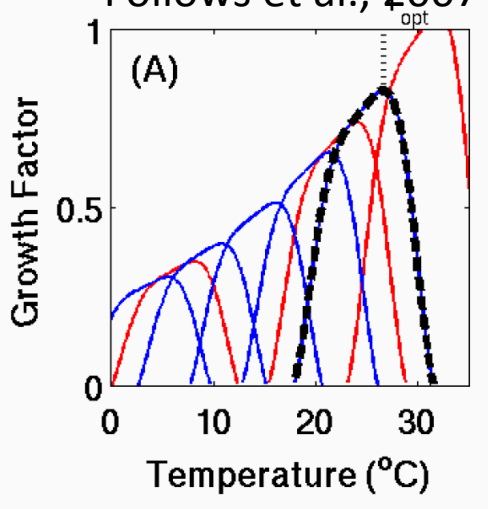

Eppley, 1972
Growth Rate [day−1]
Temperature [deg. C]
0
1
2
3
4
5
6
7
-10
0
10
20
30
40
50


A
Growth Rate [d⁻¹]
Temperature [°C]
B
[mmol N m−3]
Time [years]
C
Indiv. Phyto. [μmol N m⁻³]
Time [years]
D
Growth Rate [d⁻¹]
Temperature [°C]

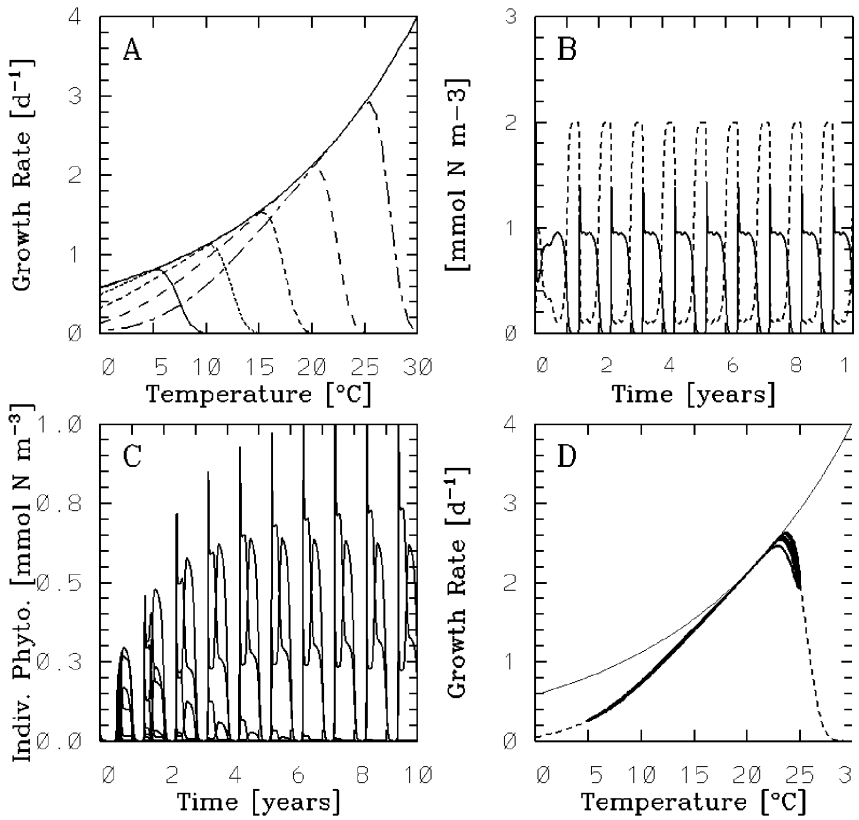

A
Growth Rate [d⁻¹]
Temperature [°C]
B
[mmol N m−3]
Time [years]
C
Indiv. Phyto. [mmol N m⁻³]
Time [years]
D
Growth Rate [d⁻¹]
Temperature [°C]

# “Everything is everywhere, **<u>but</u>** the environment selects*”.

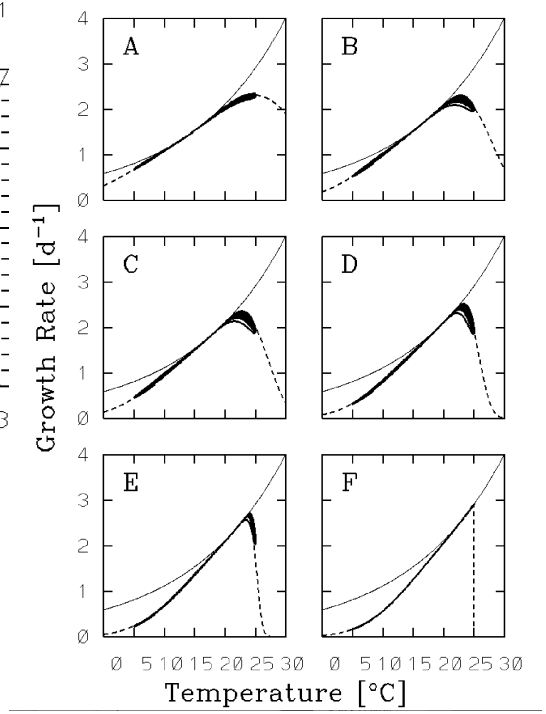
Follows et al., 2007




Moisan et al., 2002




Moisan et al., 2002

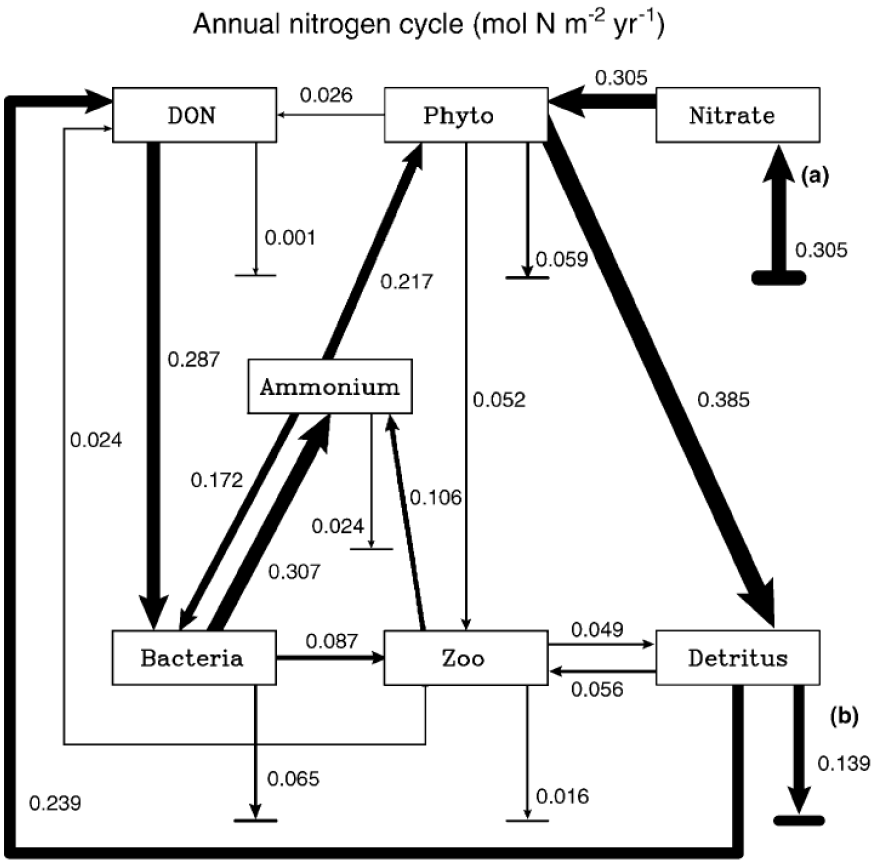



*Bass Becking, 1934

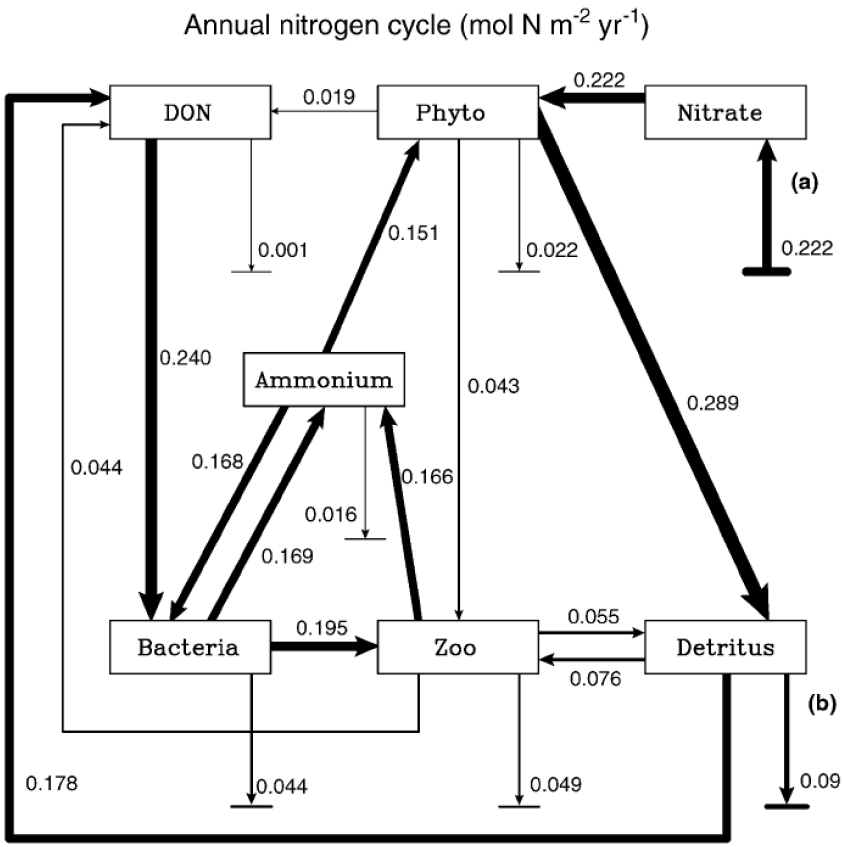
# GP Year 2 Task 1: Carry out 1D model development using BATS and HOTS data sets

## Bermuda-Atlantic Time Series Experiments

## Fasham et al., 1990




## Spitz et al., 2001



## *Satellite Algorithms:* $S$

**Box Models**

$$\frac{dB_i}{dt} = S_i$$

**One Dimensional (Vertical) Models**

$$\frac{\partial B_i}{\partial t} + (w + w_B)\frac{\partial B_i}{\partial z} - \frac{\partial}{\partial z} K_z \frac{\partial B_i}{\partial z} = S_i$$

**Three Dimensional Models**

$$\frac{dB_i}{dt} + (\vec{v} + \vec{v}_B) \bullet \nabla B_i - \nabla \bullet (\vec{K} \nabla B_i) = S_i$$

# Box Models

$$\frac{dB_i}{dt} = S_i \Longrightarrow \left[ \frac{dP}{dt} = S_P ; \frac{dN}{dt} = S_N ; \frac{dZ}{dt} = S_Z \right]$$

$$\frac{dN}{dt} = S_N = -\rho_{\max} \frac{N}{K_N + N} P + \mu Z + (1-\beta) g_{\max} \frac{P}{K_P + P} Z + \delta Z^2$$

$$\frac{dP}{dt} = S_P = \rho_{\max} \frac{N}{K_N + N} P - g_{\max} \frac{P}{K_P + P} Z$$

$$\frac{dZ}{dt} = S_Z = \beta g_{\max} \frac{P}{K_P + P} Z - \mu Z - \delta Z^2$$

# Box Models

$$\frac{dB_i}{dt} = S_i \Longrightarrow \left[ \frac{dP}{dt} = S_P ; \frac{dN}{dt} = S_N ; \frac{dZ}{dt} = S_Z \right]$$

$$\frac{dN}{dt} = S_N = -\rho_{\max} \frac{N}{K_N + N} P + \mu Z + (1-\beta) g_{\max} \frac{P}{K_P + P} Z + \delta Z^2$$

$$\frac{dP}{dt} = S_P = \rho_{\max} \frac{N}{K_N + N} P - g_{\max} \frac{P}{K_P + P} Z$$

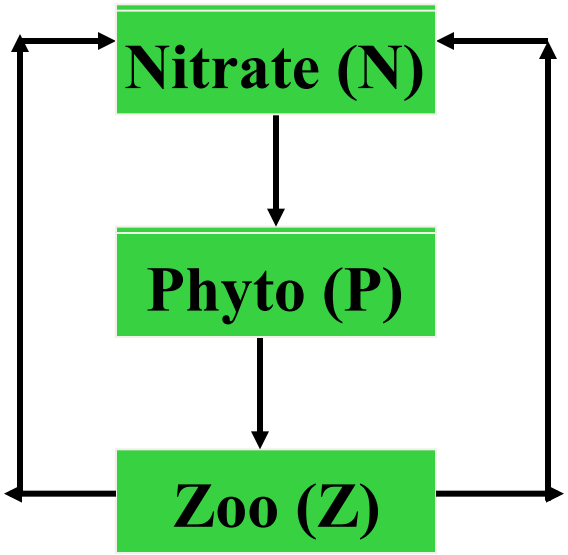
$$\frac{dZ}{dt} = S_Z = \beta g_{\max} \frac{P}{K_P + P} Z - \mu Z - \delta Z^2$$

## Ecosystem/Food Web Models can be represented as a matrix of terms/equations

$$\frac{dN}{dt} = -\rho_{\max}\frac{N}{K_N + N}P + \mu Z + g_{\max}\frac{P}{K_P + P}Z - \beta g_{\max}\frac{P}{K_P + P}Z$$

$$\frac{dP}{dt} = \rho_{\max}\frac{N}{K_N + N}P - g_{\max}\frac{P}{K_P + P}Z$$

$$\frac{dZ}{dt} = \beta g_{\max}\frac{P}{K_P + P}Z - \mu Z$$




**Sink**

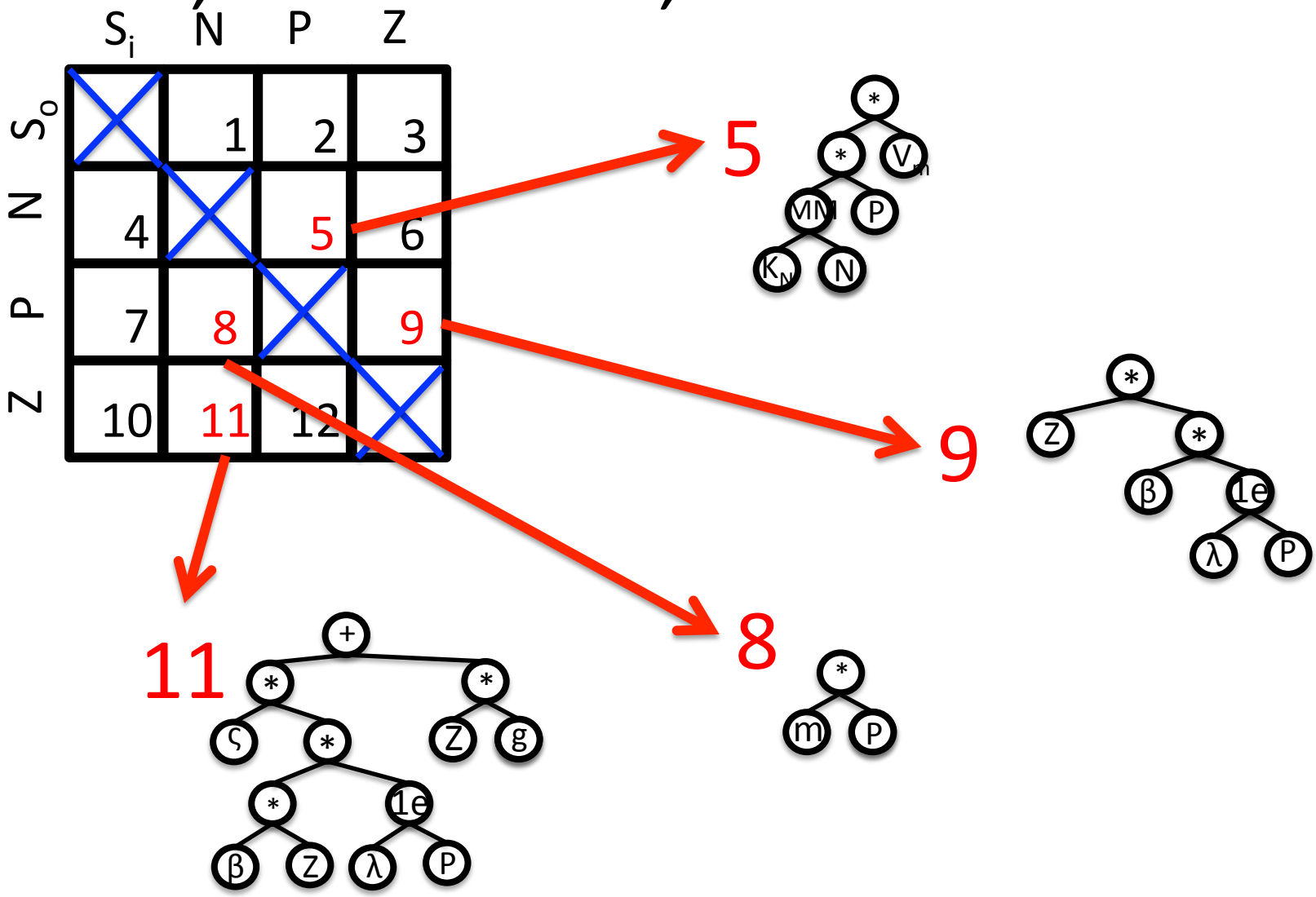
| Source | N | P | Z |
|---|---|---|---|
| N | | X | X |
| P | X | | |
| Z | X | | |

# Fasham et al. (1990) BioFlo Matrix

| | Sink | $NO_3^-$ $(N_n)$ | $NH_4^+$ $(N_r)$ | DON $(N_d)$ | DET $(D)$ | BACT $(B)$ | PHYTO $(P)$ | ZOOP $(Z)$ |
|---|---|---|---|---|---|---|---|---|
| Source | | $\frac{(m+h^+)}{M}N_o$ | | | | | | |
| $NO_3^-$ | $\frac{(m+h^+)}{M}N_n$ | | | | | | $JQ_1P$ | |
| $NH_4^+$ | $\frac{(m+h^+)}{M}N_r$ | | | | | $U_2$ | $JQ_2P$ | |
| DON | $\frac{(m+h^+)}{M}N_d$ | | | | | $U_1$ | | $G_3$ |
| DET | $\frac{(m+h^+)}{M}N_o$ | | | $\mu_4 D$ | | | | $G_2$ |
| BACT | $\frac{(m+h^+)}{M}B$ | | $B\mu_3$ | | | | | $G_1$ |
| PHYTO | $\frac{(m+h^+)}{M}P$ | | | $\gamma_1\sigma P$ | $\mu_1 P$ | | | |
| ZOOP | $\frac{hZ}{M}+\Omega\mu_5 Z$ | | $(1-\Omega)\mu_5 Z$ | $(1-\epsilon)\mu_2 Z$ | $(1-\beta_1)G_1$ $+(1-\beta_2)G_2$ $+(1-\beta_3)G_3$ | | | |

Total RHS = Sum of Source Terms – Sum of Sink Terms; Total Possible Equations = $(N)^2-(N)=56$

# NPZ, Franks et al., Mar. Bio. 1986 N=3




Total RHS = Sum of Source Terms – Sum of Sink Terms;

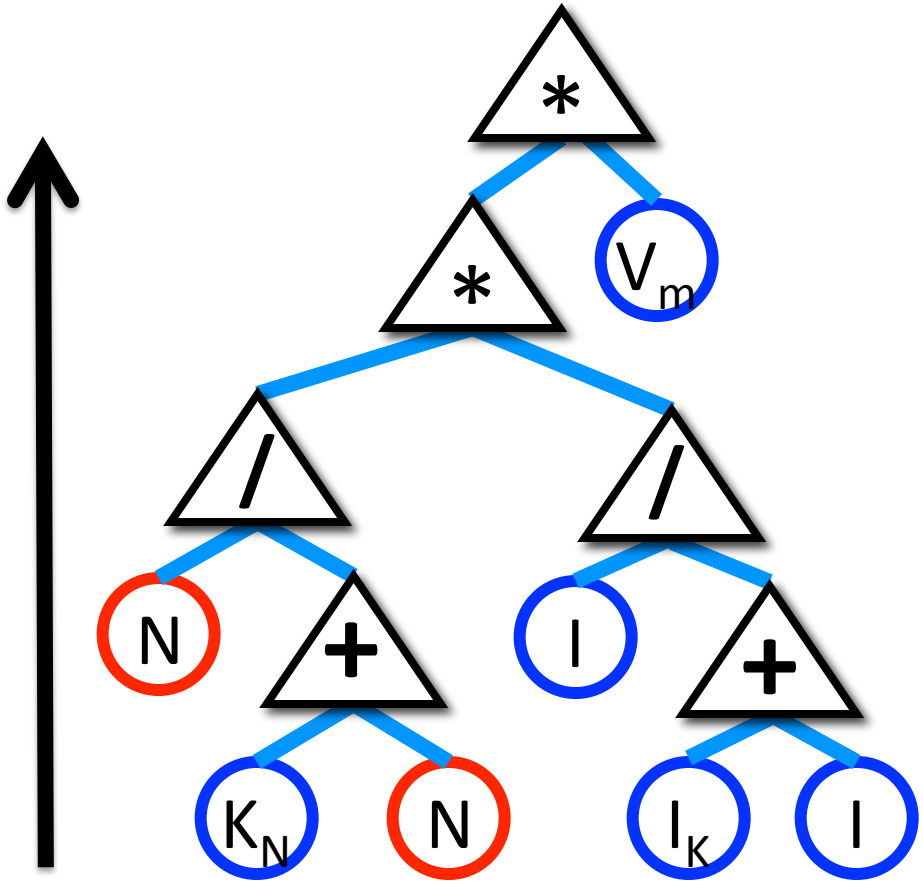
Total Possible Equations = $(N+1)^2-(N+1)=12$; N=3

$$\text{Growth} = V_m \left[ \frac{N}{K_N + N} \right] \left[ \frac{I}{I_K + I} \right]$$

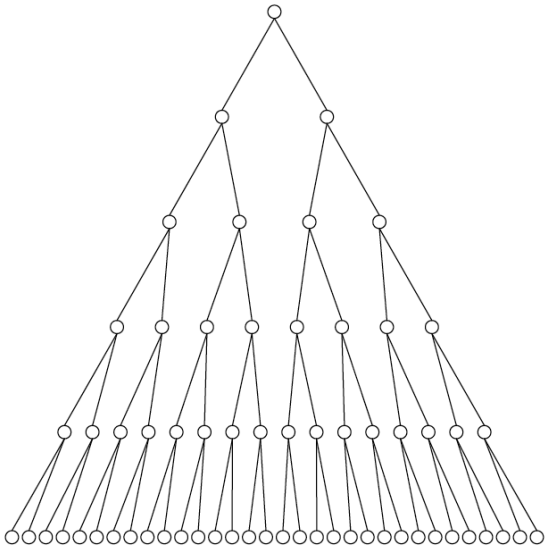
LISP prefix representation:

(* $V_m$ (* (/ N (+ ($K_N$ N))) (/ I (+ ($I_K$ I)))))

Calculation of Tree value moves from the bottom up.

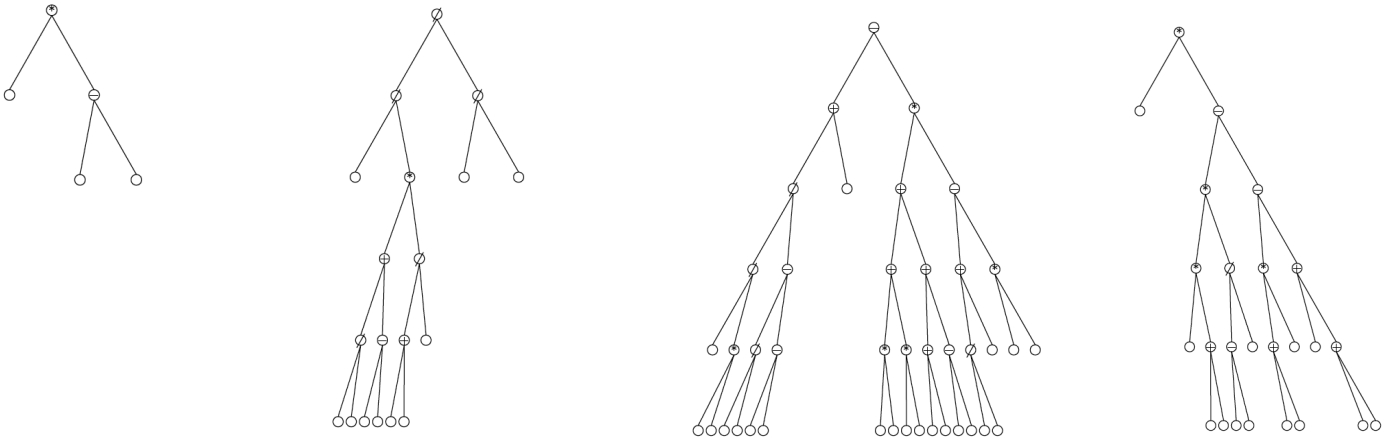


# Binary Tree Architecture



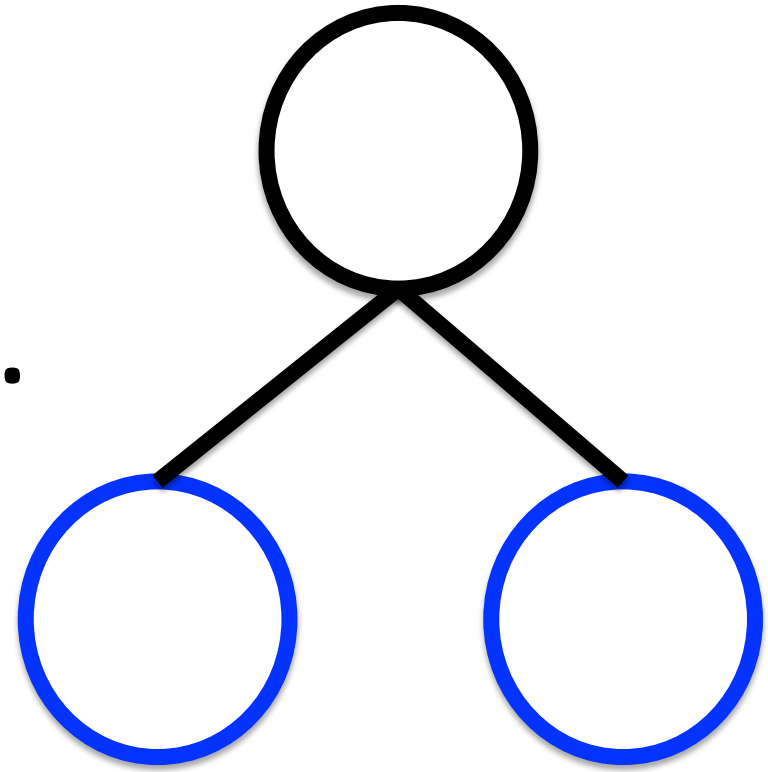
Empty and Full Binary Tree Map

# Example of randomly generated initial tree structures



# Use of Fortran-90 'Defined Type'

**Operator:**
**[+,-,*,/,Boolean,etc]**
**Input nodes can vary.**



**Input Nodes:**
**[Variable, parameter,**
**New defined type]**

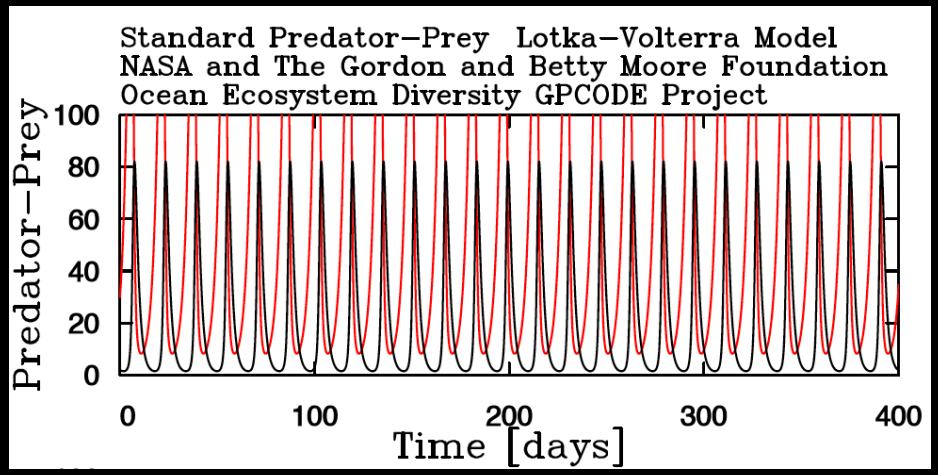
**Calculation starts from top of tree**
**Performance improved by orders of magnitude!**

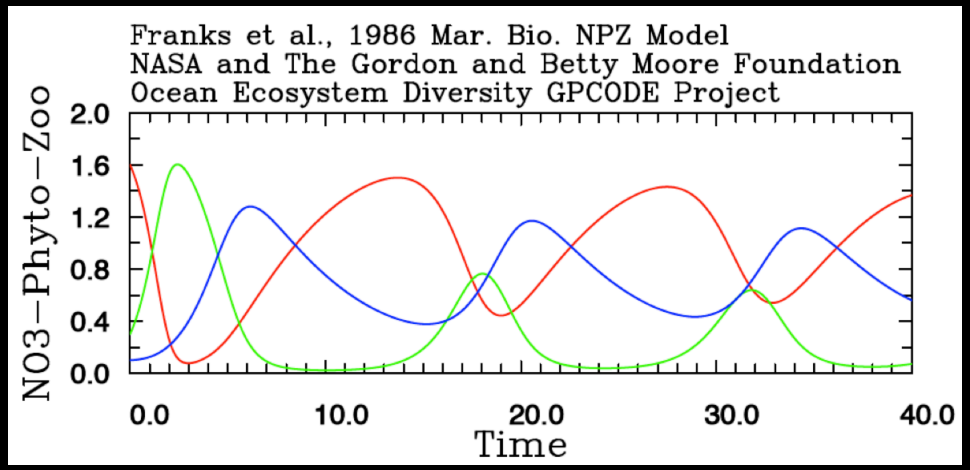
# ‘Digitized’ Ecosystem Model Equations

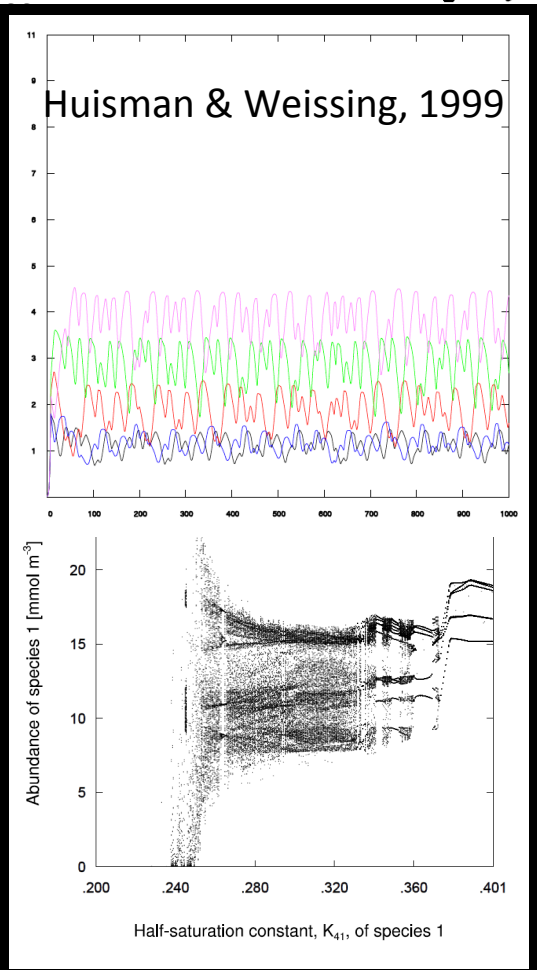
**Equation Term**

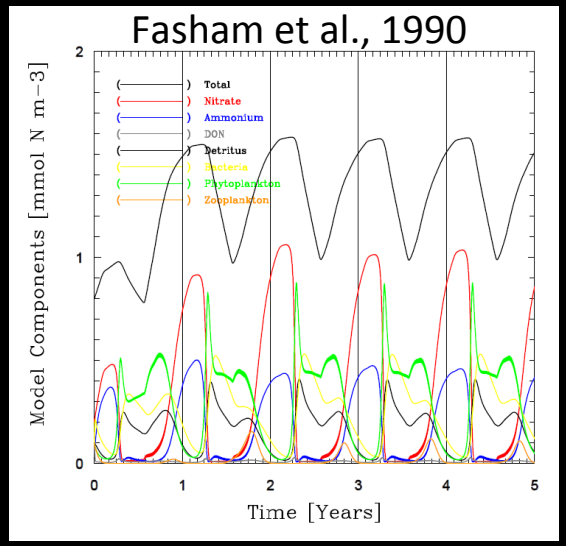
```
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3, 3, 0, 6,-2, N, N, 0,-1, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3,-2, 0, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3,-3, 3, N, N, 0, 5, N, N, N, N, N, N, 0,-2, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 1, 3, 3, 0, 3,-3, 0, N, N, 3, 5, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, 0,-3, 0,-2, N, N, N, N, N, N, N, N]
```

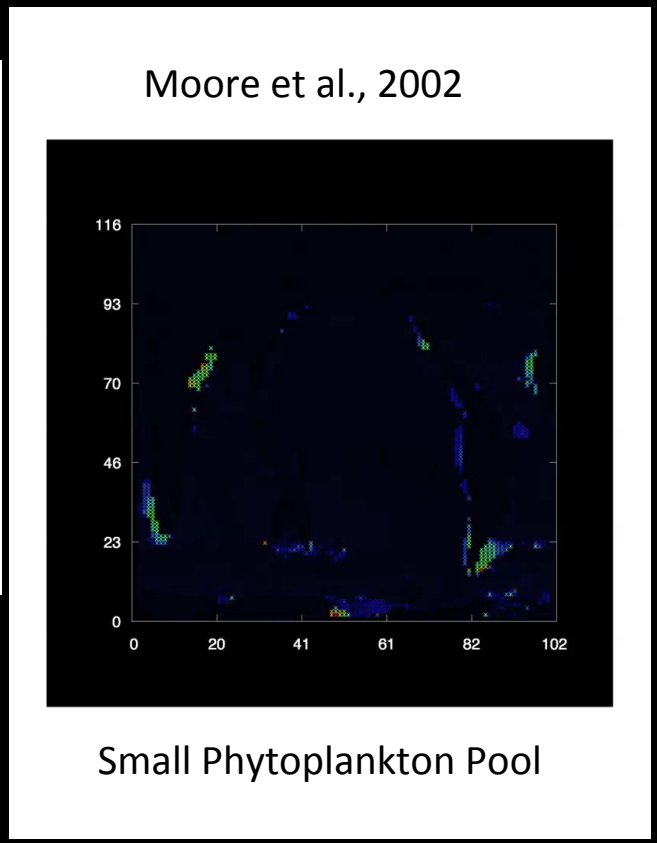
**Binary Tree Equation Term String**


Standard Predator–Prey Lotka–Volterra Model
NASA and The Gordon and Betty Moore Foundation
Ocean Ecosystem Diversity GPCODE Project
Predator–Prey
100
80
60
40
20
0
0
100
200
300
400
Time [days]


Franks et al., 1986 Mar. Bio. NPZ Model
NASA and The Gordon and Betty Moore Foundation
Ocean Ecosystem Diversity GPCODE Project
N03–Phyto–Zoo
2.0
1.6
1.2
0.8
0.4
0.0
0.0
10.0
20.0
30.0
40.0
Time


Huisman & Weissing, 1999
Abundance of species 1 [mmol m-3]
20
15
10
5
0
.200
.240
.280
.320
.360
.401
Half-saturation constant, K41, of species 1


Fasham et al., 1990
Model Components [mmol N m-3]
Total
Nitrate
Ammonium
DON
Detritus
Bacteria
Phytoplankton
Zooplankton
2
1
0
0
1
2
3
4
5
Time [Years]


Moore et al., 2002
116
93
70
46
23
0
0
20
41
61
82
102
Small Phytoplankton Pool

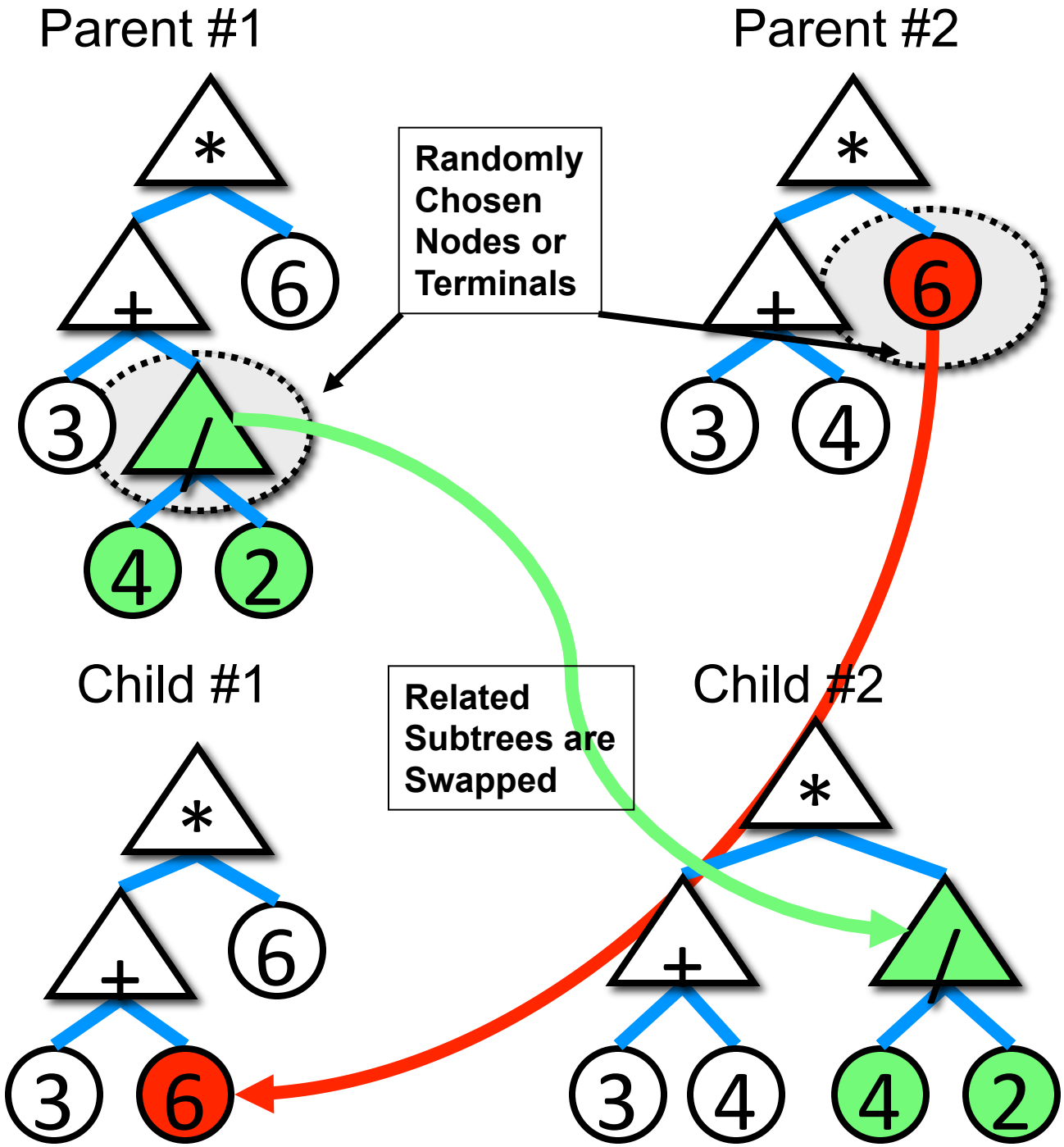
# Genetic Programming Overview

1) Generate initial random population of models
2) Calculate fitness of all individual models
3) Randomly select based on fitness for:
   - Asexual reproduction
   - Sexual reproduction (i.e. Tournament Selection)
4) Carry out genetic mutation
5) Calculate fitness of new individuals
6) Test for completion criteria (low SSE value met?)
7) No: <u>go to 3</u>; Yes: <u>end program</u>

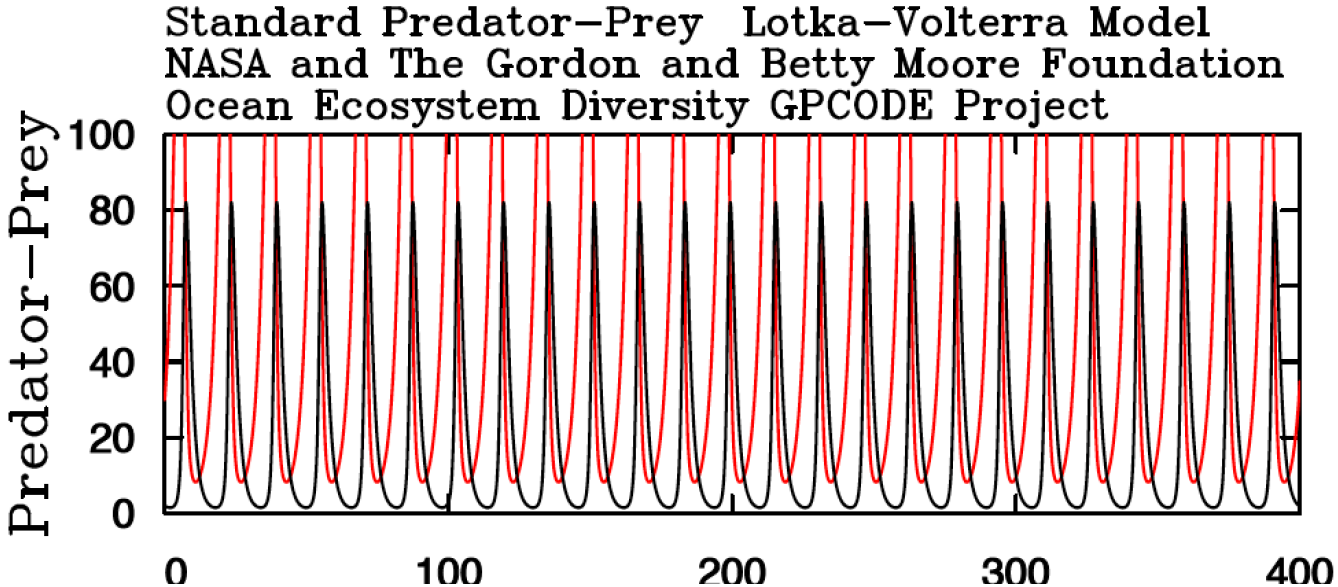
# Cross methodology

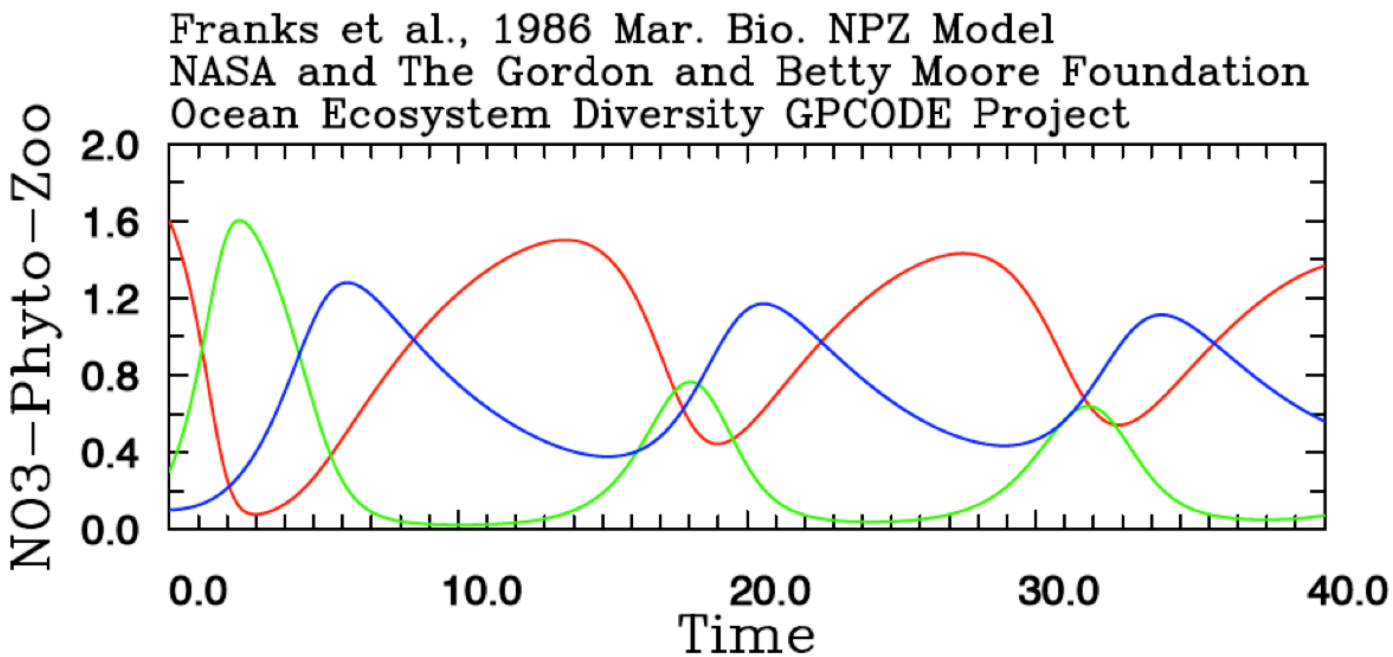
Before Cross

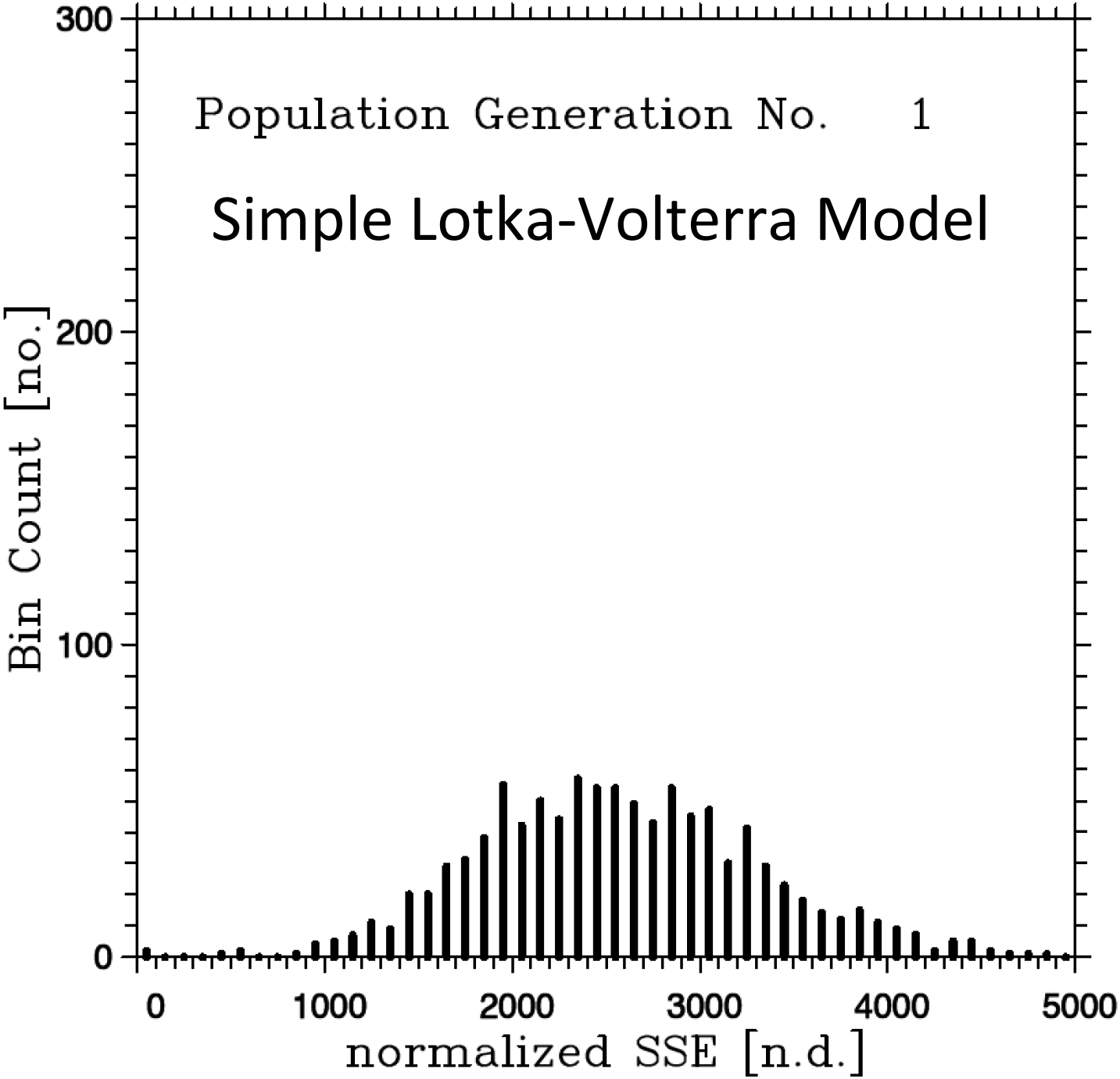
After Cross

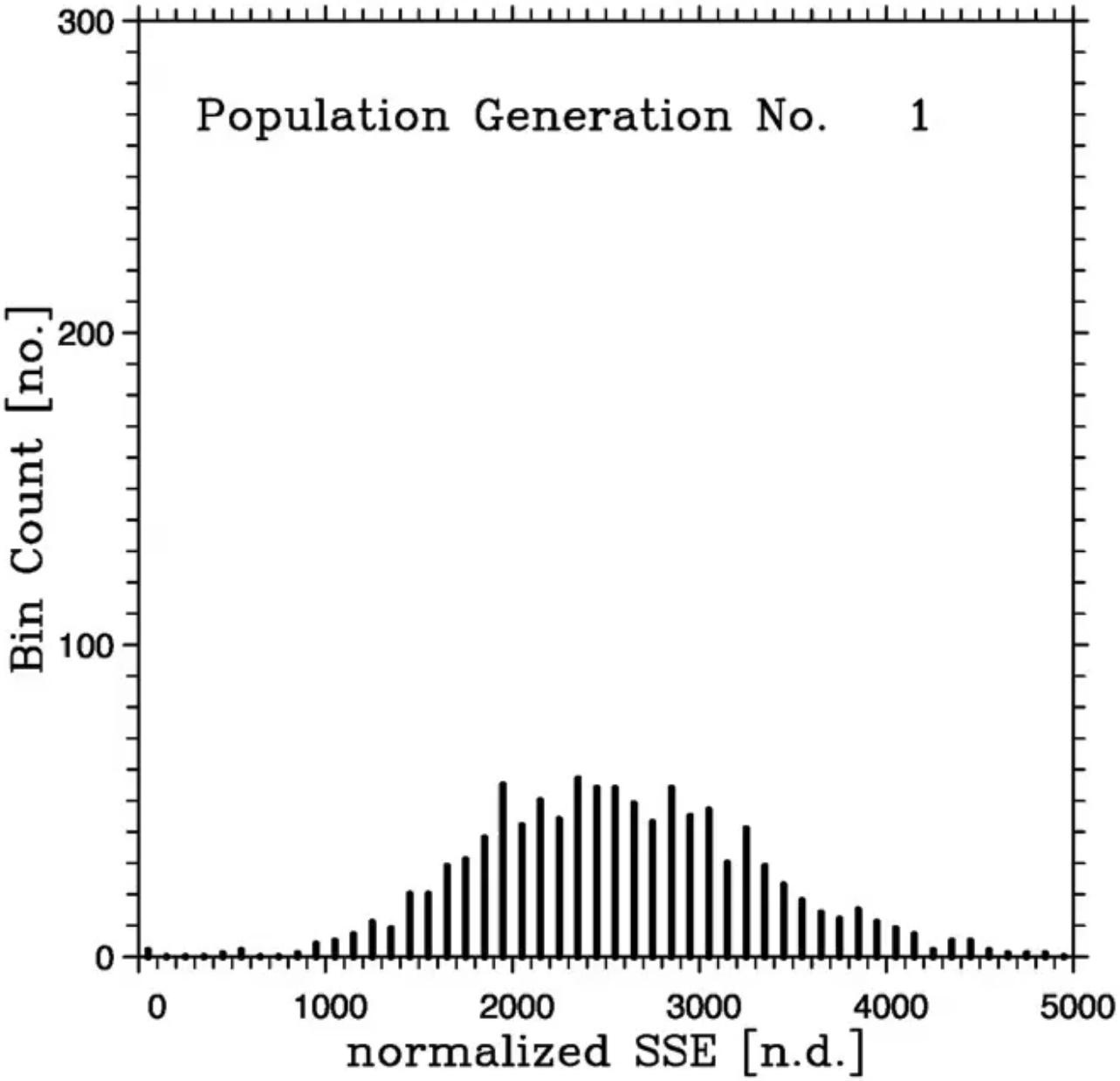


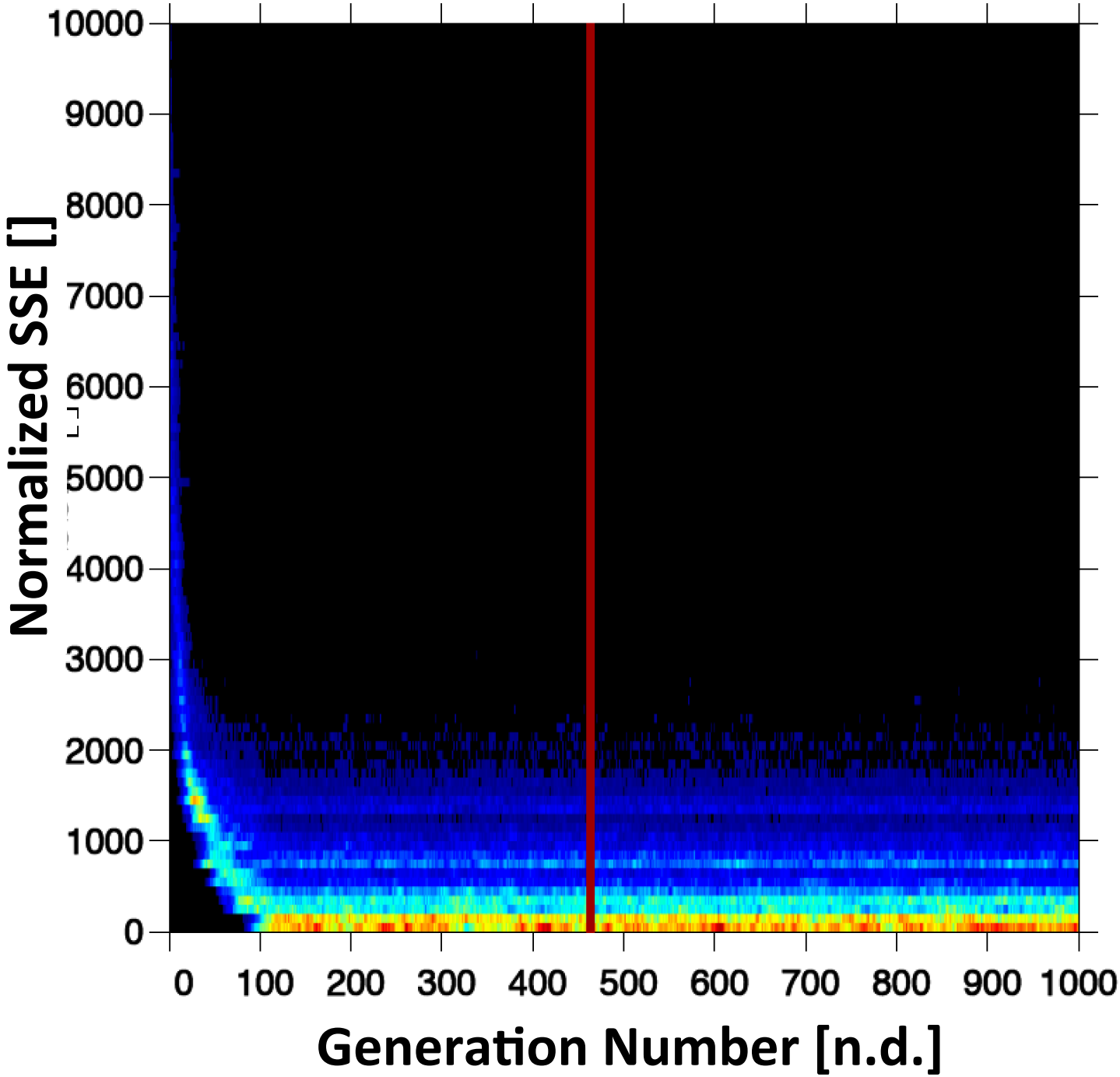
# GPCODE Box Model Twin Experiments

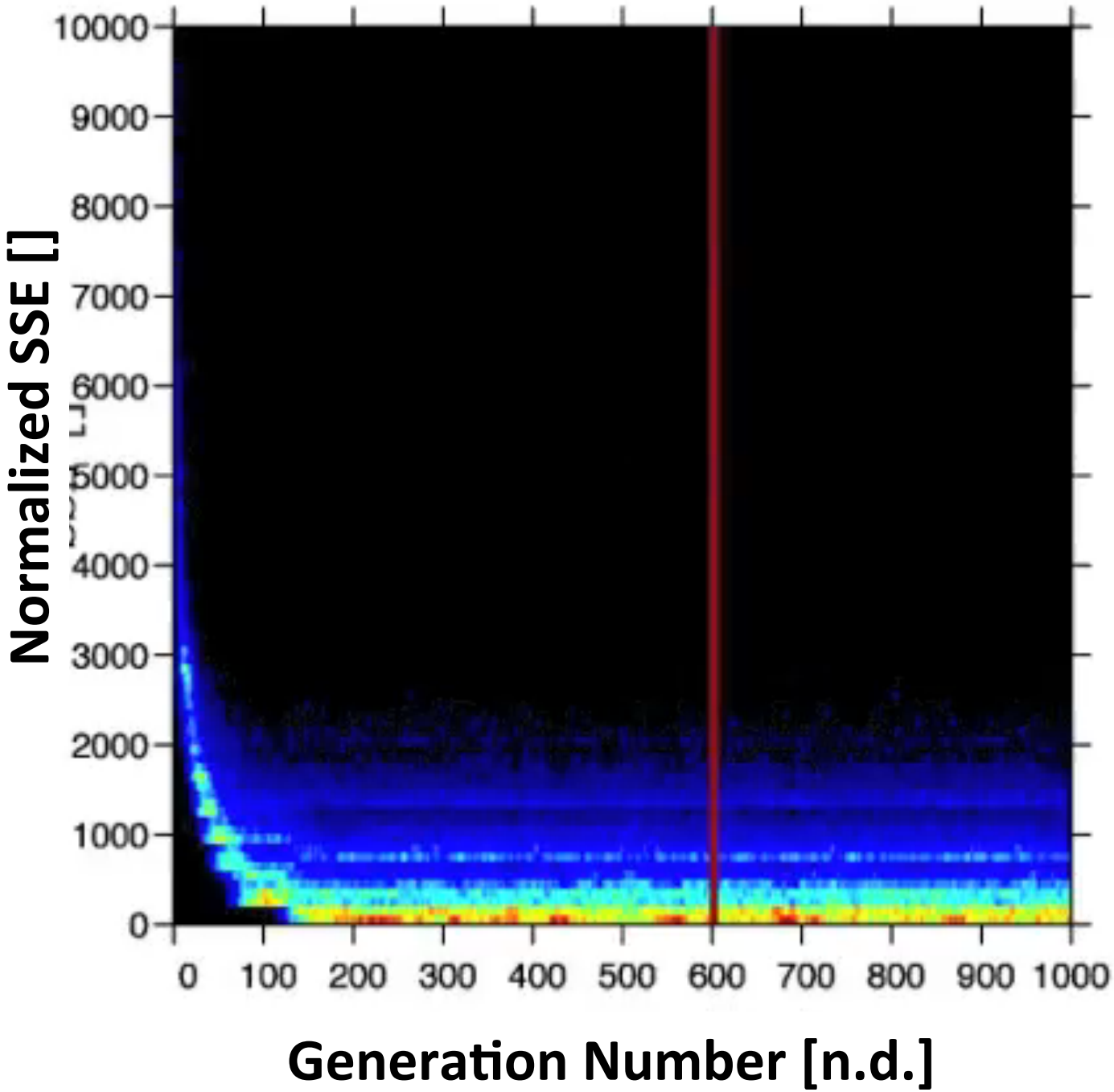



Population Generation No. 1
Simple Lotka-Volterra Model
Bin Count [no.]
300
200
100
0
0
1000
2000
3000
4000
5000
normalized SSE [n.d.]

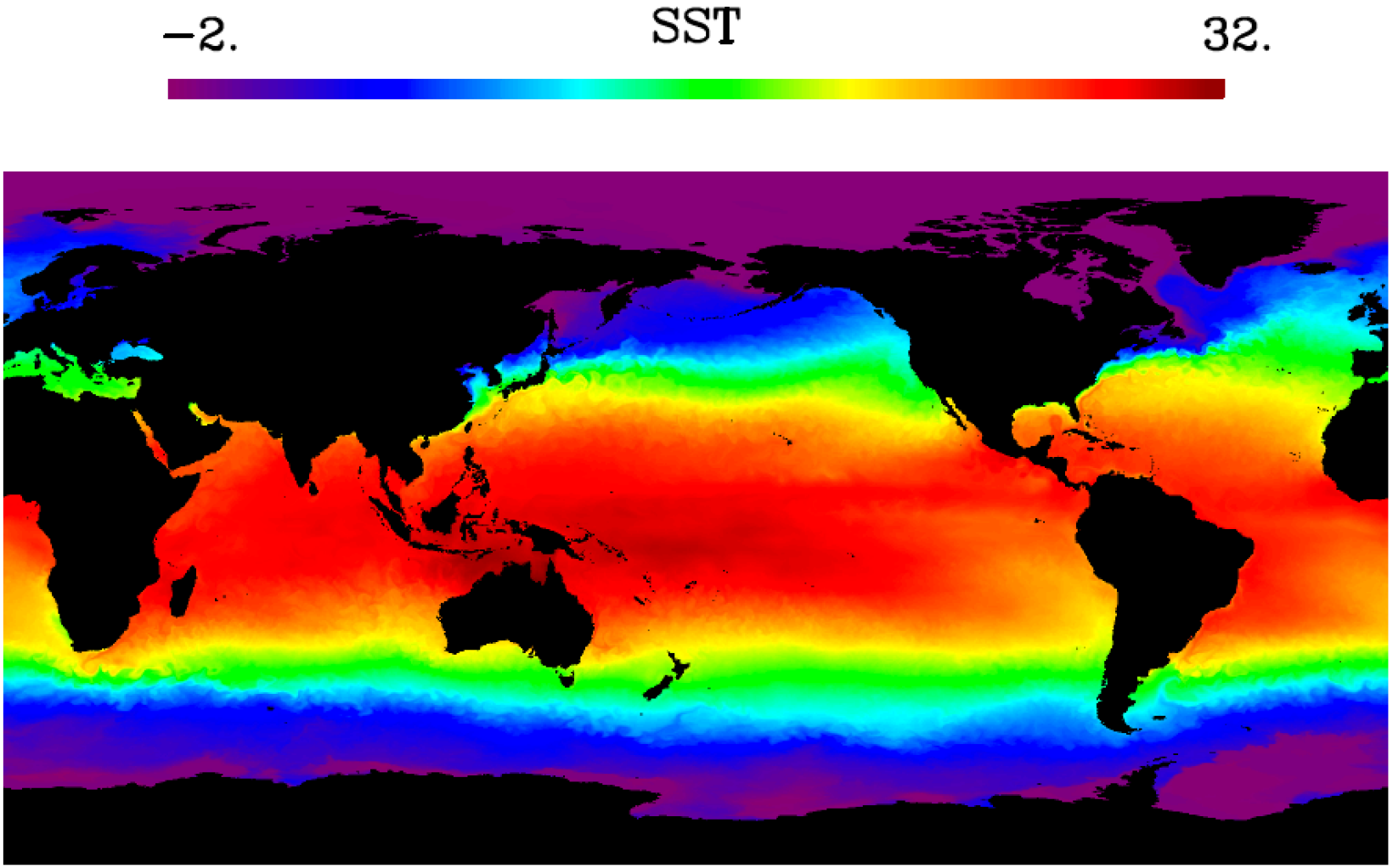

Population Generation No. 1
Bin Count [no.]
300
200
100
0
0
1000
2000
3000
4000
5000
normalized SSE [n.d.]


Normalized SSE []
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
600
700
800
900
1000
Generation Number [n.d.]


Normalized SSE []
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
600
700
800
900
1000
Generation Number [n.d.]

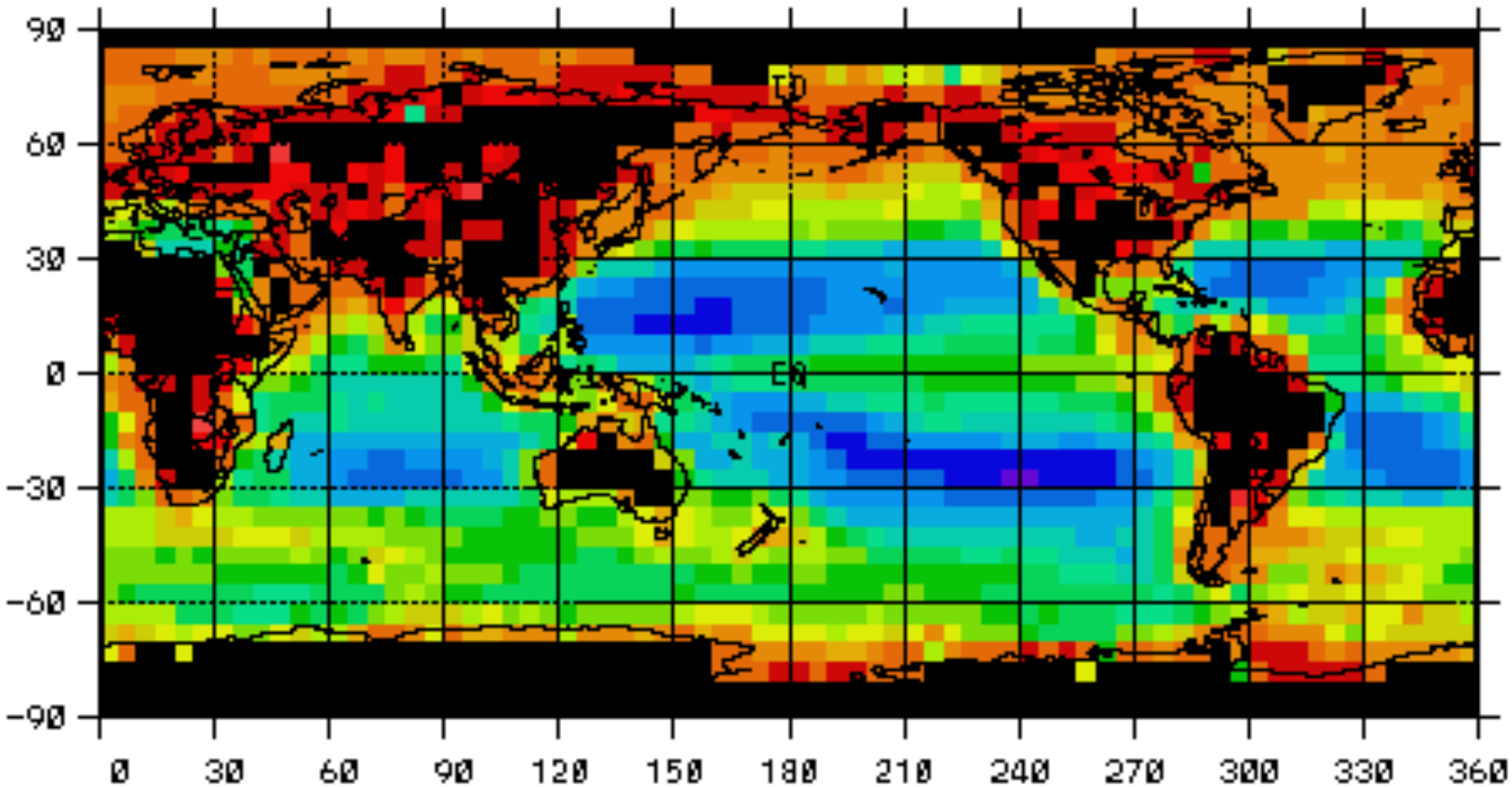
# GP Year 2 Task 2: Begin off-line global GPCODE modeling experiments

## ECCO2 1992-2013 SST Solutions



# GP Year 2 Task 2 (cont'd): Develop GPCODE global modeling at coarse scales using satellite observations

**5 deg X 5 deg SeaWiFS mean chl. *a***



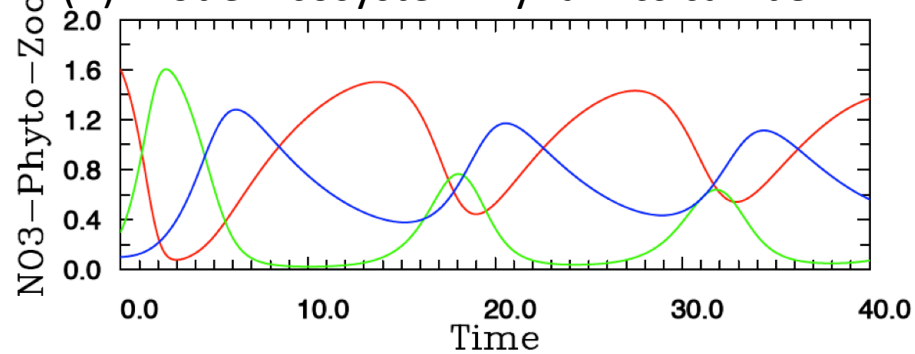
## (A) Coupled Equations used to….

$$\frac{\partial N}{\partial t} = -V_m P\left(\frac{N}{K_N + N}\right) + mP + gZ + \gamma\beta\left(1 - e^{-\lambda P}\right)Z$$

$$\frac{\partial P}{\partial t} = V_m P\left(\frac{N}{K_N + N}\right) - mP - \beta\left(1 - e^{-\lambda P}\right)Z$$

$$\frac{\partial Z}{\partial t} = \beta\left(1 - e^{-\lambda P}\right)Z - \gamma\beta\left(1 - e^{-\lambda P}\right)Z - gZ$$

## (B) Model Ecosystem Dynamics can be….

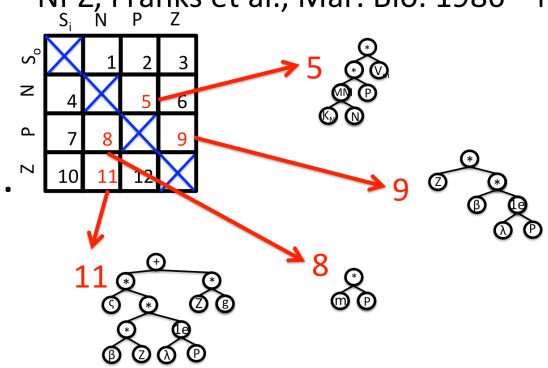



## (C) Represented in a Matrix Form, whose terms are….

| All | Sink | Nitrate (N) | Phytoplankton (P) | Zooplankton (Z) |
|---|---|---|---|---|
| Source | Trace | [1] | [2] | [3] |
| Nitrate (N) | [4] | Terms | [5] $V_m P\left(\frac{N}{K_N + N}\right)$ | [6] |
| Phytoplankton (P) | [7] | [8] $mP$ | Are | [9] $\beta\left(1 - e^{-\lambda P}\right)Z$ |
| Zooplankton (Z) | [10] | [11] $\gamma\beta\left(1 - e^{-\lambda P}\right)Z + gZ$ | [12] | Null |

## (D) Simple binary tree expressions or strings of…

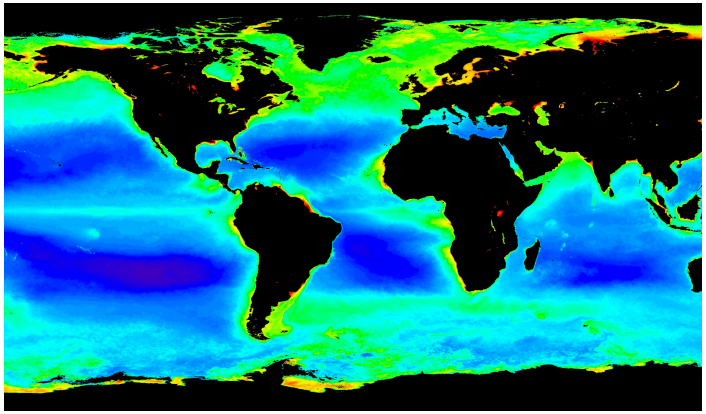
NPZ, Franks et al., Mar. Bio. 1986 N=3




## (E) Information capable of optimization by Evolutionary Theory….

```
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3, 3, 0, 6,-2, N, N, 0,-1, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3,-2, 0, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 3,-3, 3, N, N, 0, 5, N, N, N, N, N, N, 0,-2, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ 1, 3, 3, 0, 3,-3, 0, N, N, 3, 5, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N]
[ N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, N, 0,-3, 0,-2, N, N, N, N, N, N, N, N]
```

## (F) In order to Model Global Ocean Ecosystems. ➔

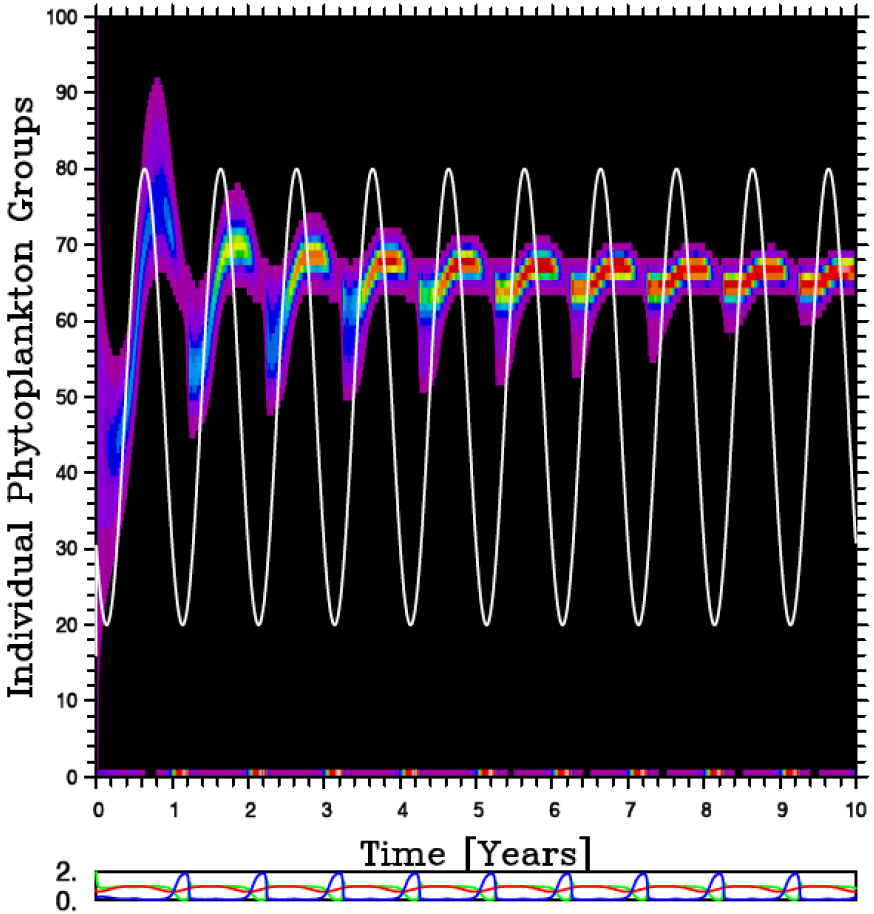


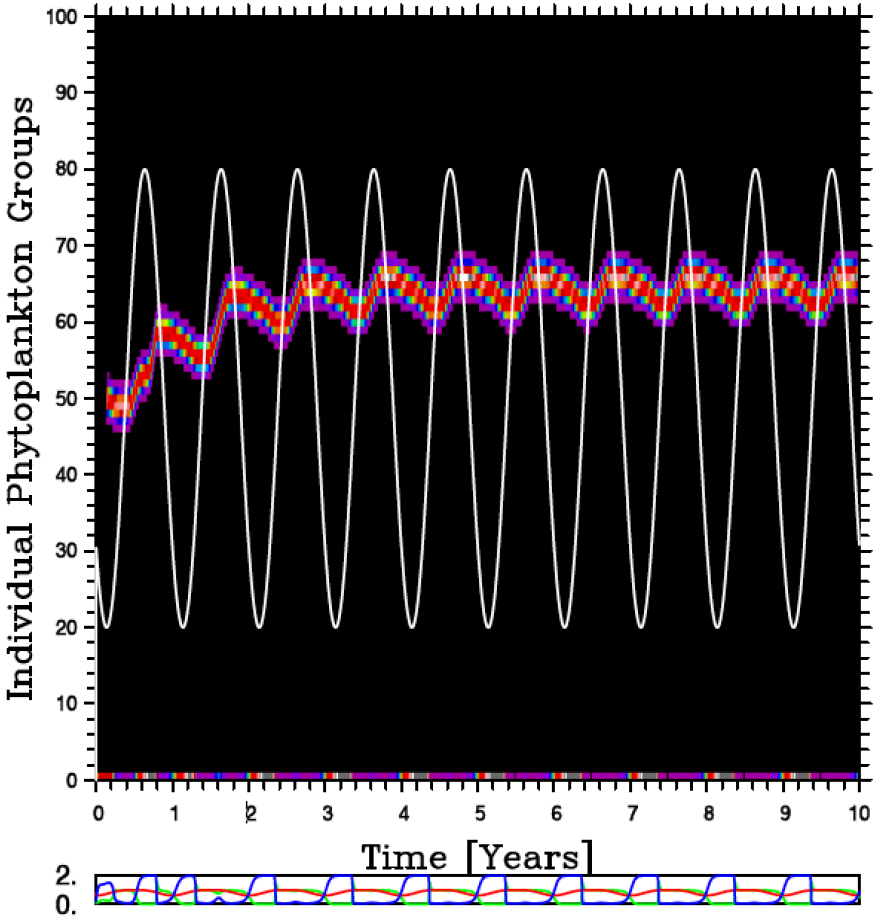
GORDON AND BETTY MOORE FOUNDATION

NASA

# Linking parameters to traits for diversity and acclimation

**Model with 100 phytoplankton groups**




**Gaussian Trait-based model (2-3 equations)**