Using Genetic Programs to Evolve Satellite Algorithms

Earth Science Technology Forum

May 13, 2021

John R. Moisan

NASA/GSFC

Earth Science Division, Code 610.W

John.R.Moisan@nasa.gov

ESTO
Earth Science Technology Office

GORDON AND BETTY MOORE FOUNDATION

Outline

- What led me to Genetic Programming
- NEPAC architecture
- Chlorophyll a satellite algorithms
- Genetic Programming overview
- Evolving satellite algorithms

Satellite Algorithms: S

Box Models

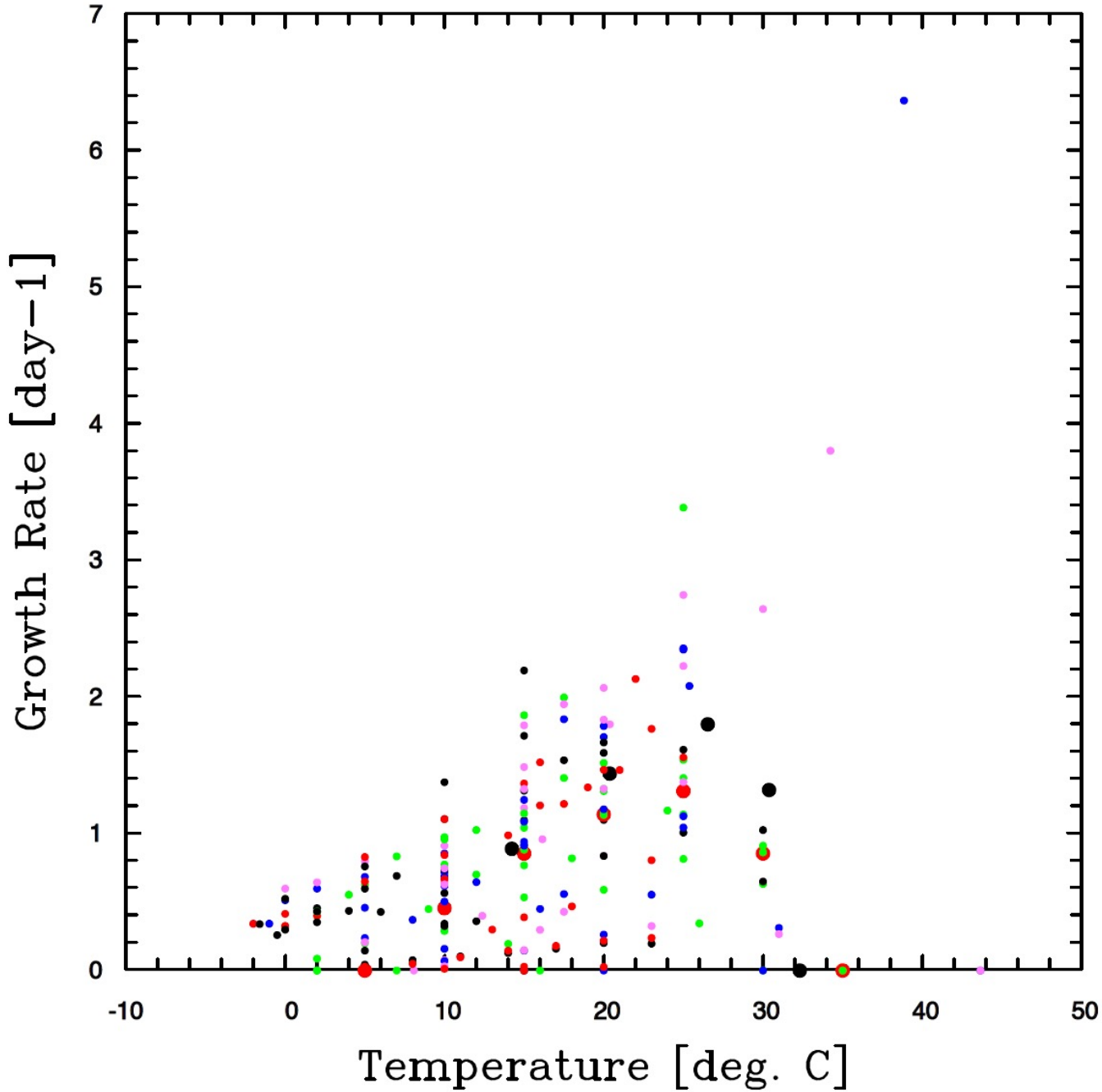
$$\frac{dB_i}{dt} = S_i$$

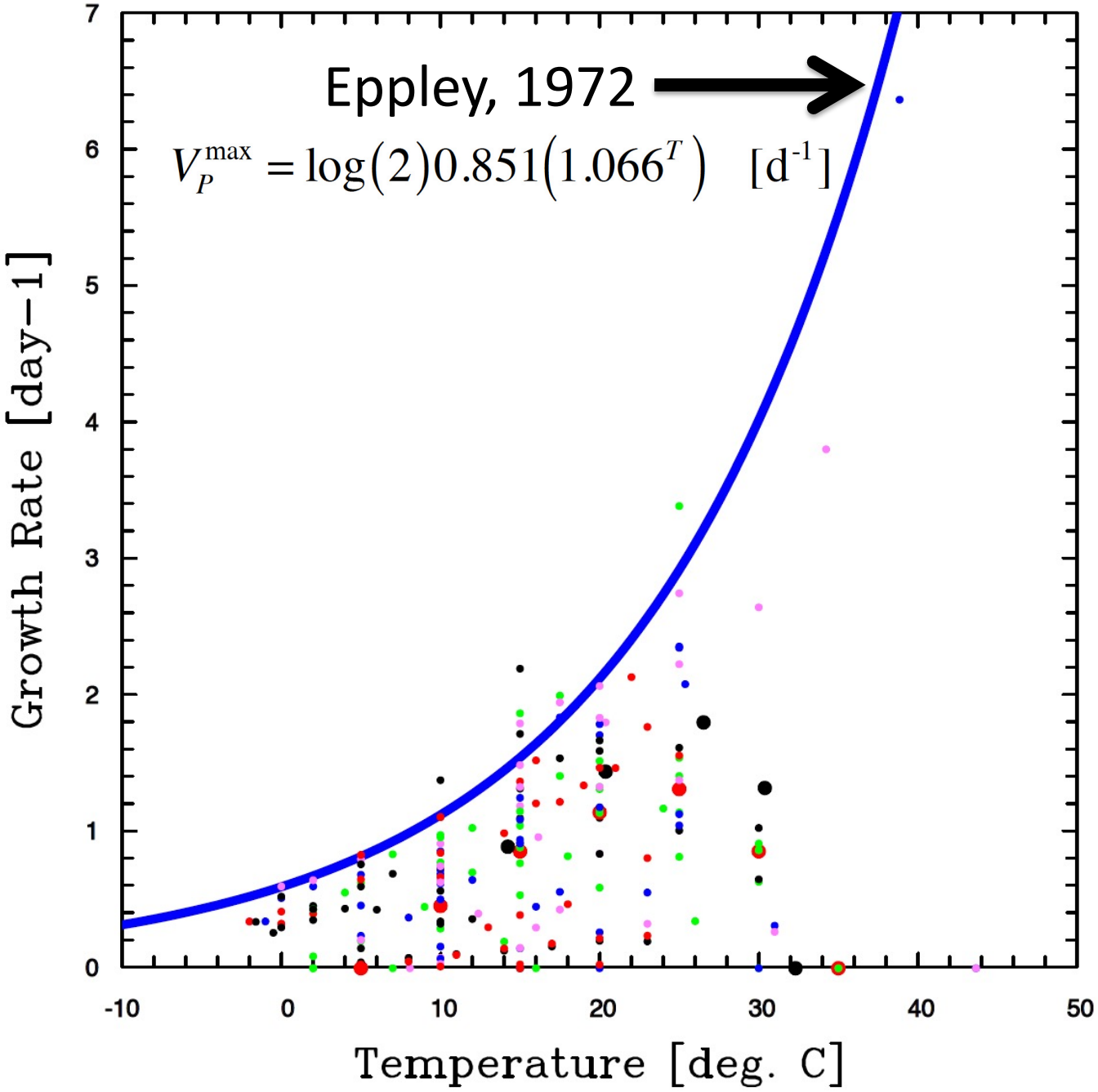
One Dimensional (Vertical) Models

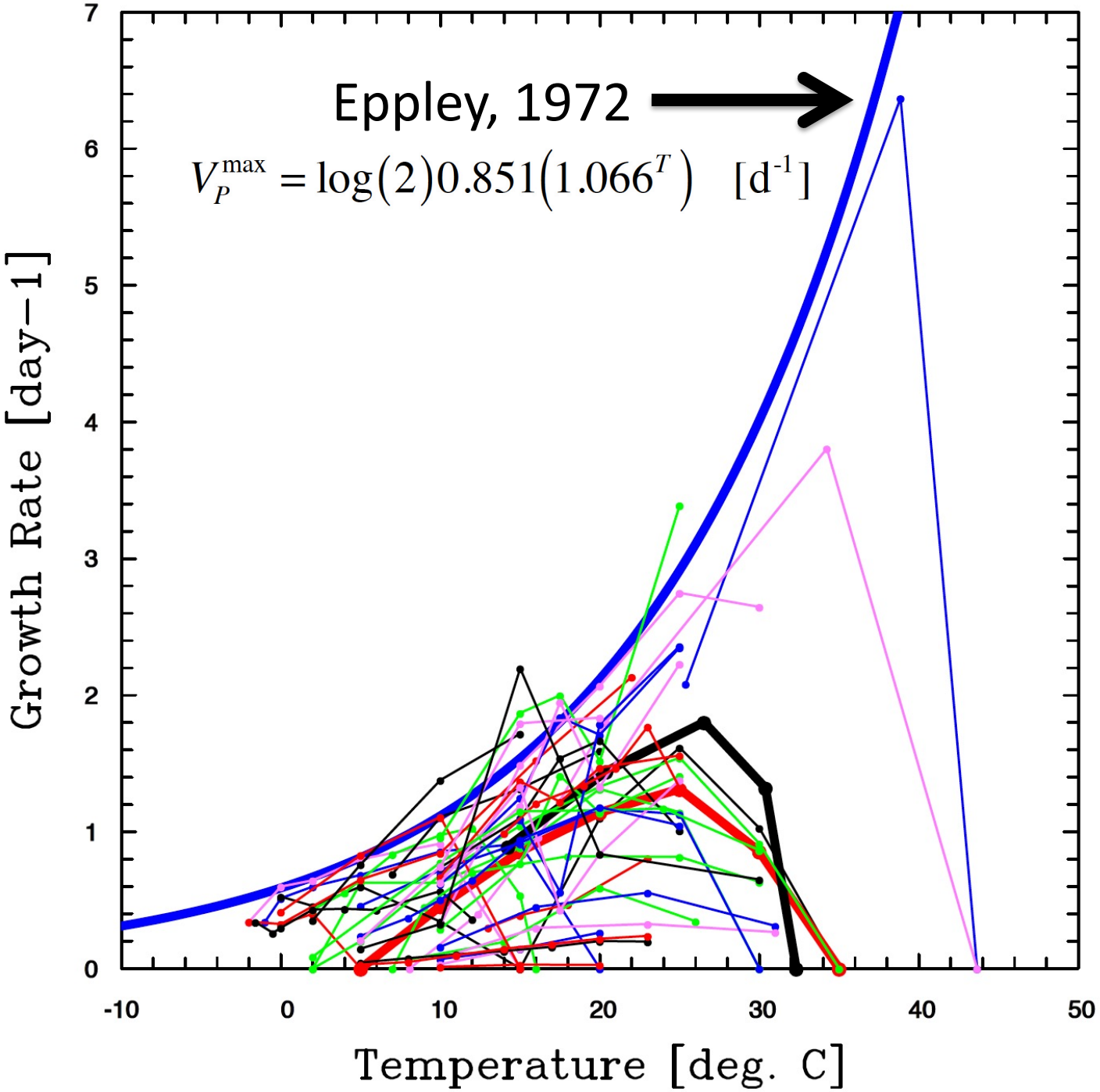
$$\frac{\partial B_i}{\partial t} + (w + w_B)\frac{\partial B_i}{\partial z} - \frac{\partial}{\partial z} K_z \frac{\partial B_i}{\partial z} = S_i$$

Three Dimensional Models

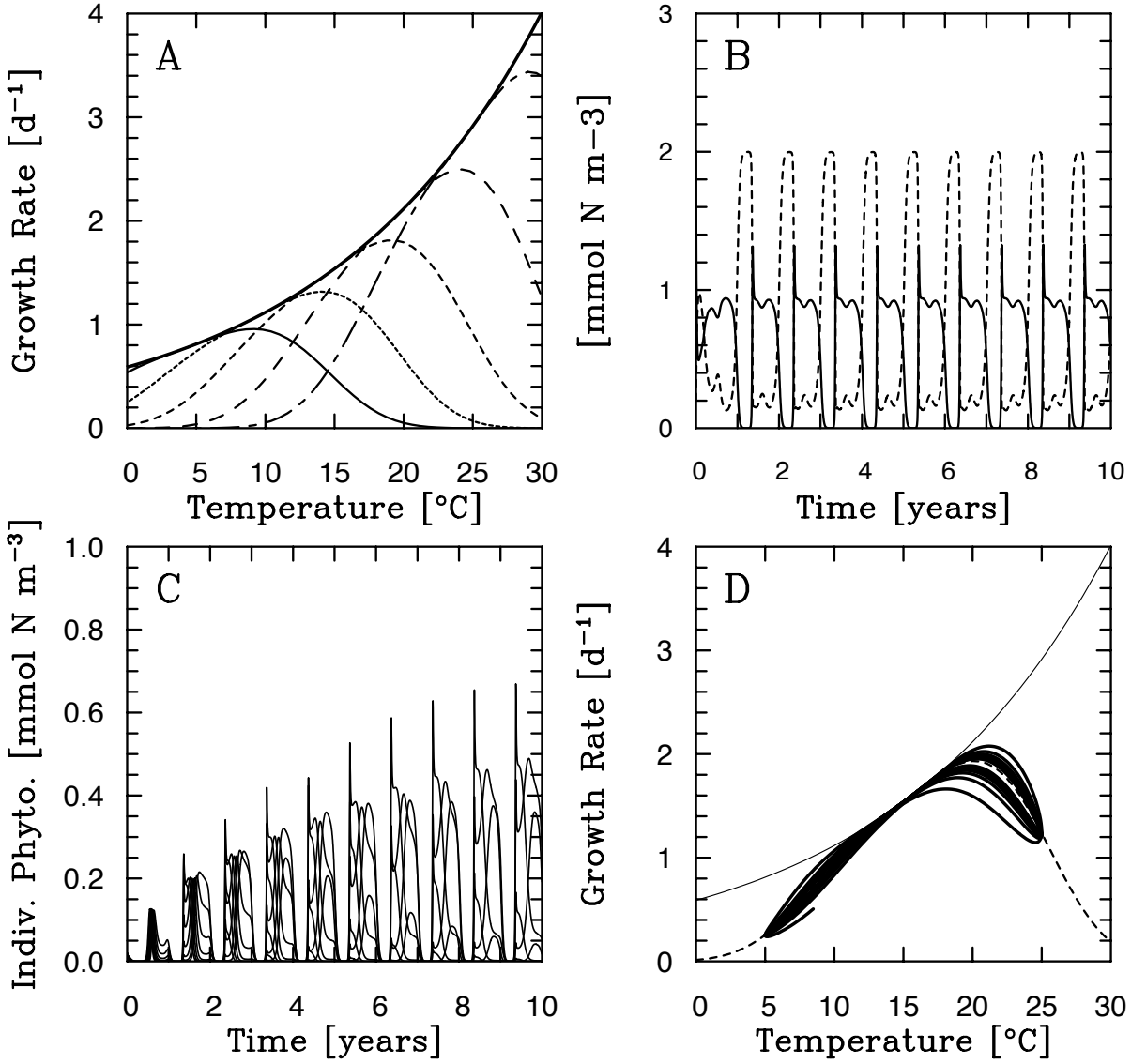
$$\frac{dB_i}{dt} + (\vec{v} + \vec{v}_B)\bullet\nabla B_i - \nabla\bullet(\vec{K}\nabla B_i) = S_i$$


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

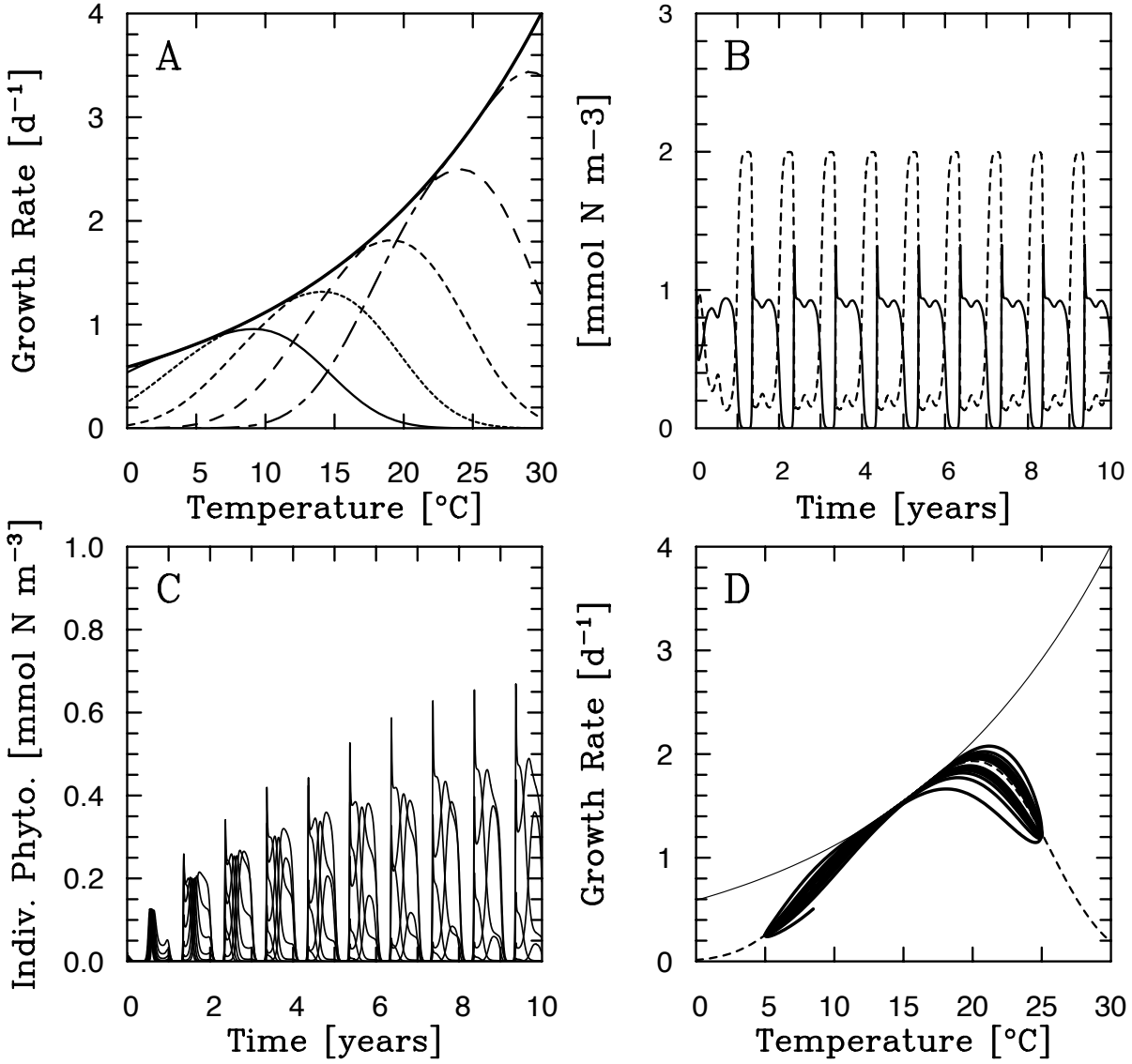

Eppley, 1972
$V_P^{max} = \log(2)0.851(1.066^T)$ [d^{-1}]
Growth Rate [day−1]
Temperature [deg. C]
7
6
5
4
3
2
1
0
-10
0
10
20
30
40
50

Moisan et al., 2002

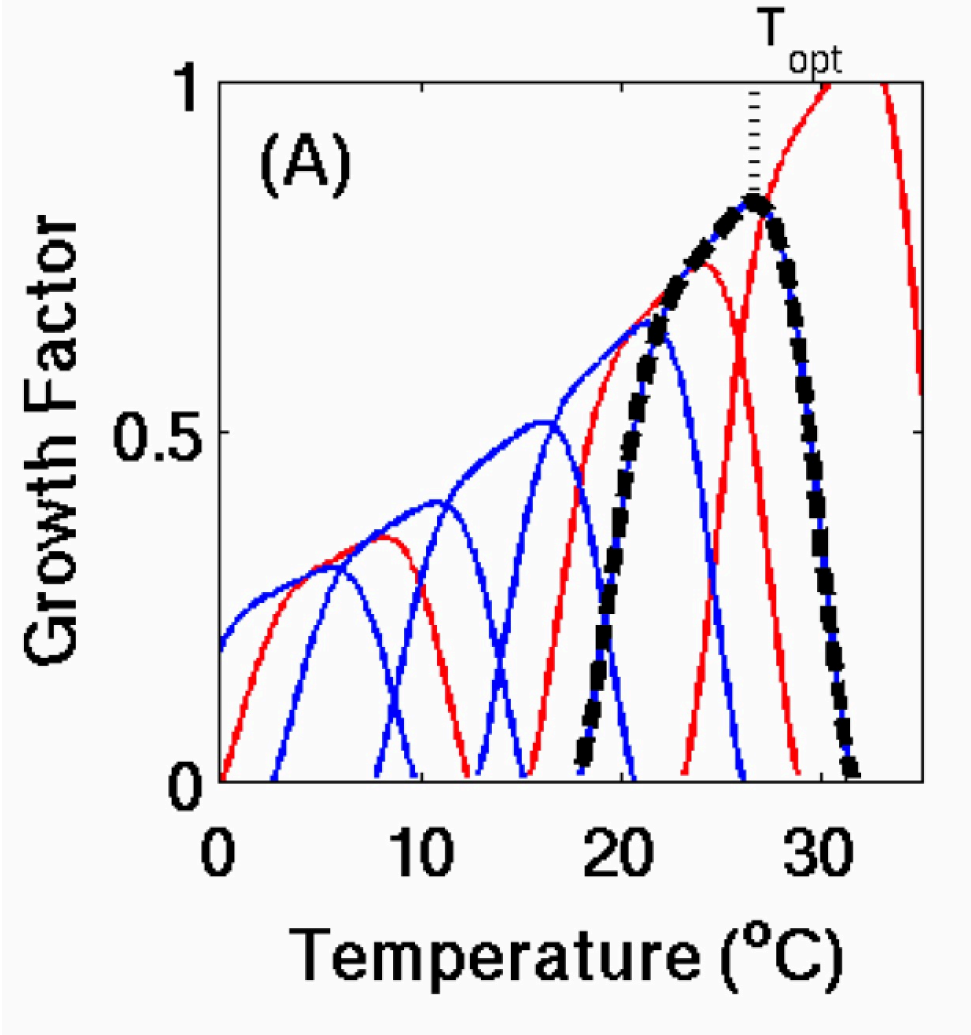


“Everything is everywhere, **<u>but</u>** the environment selects*”.

Moisan et al., 2002

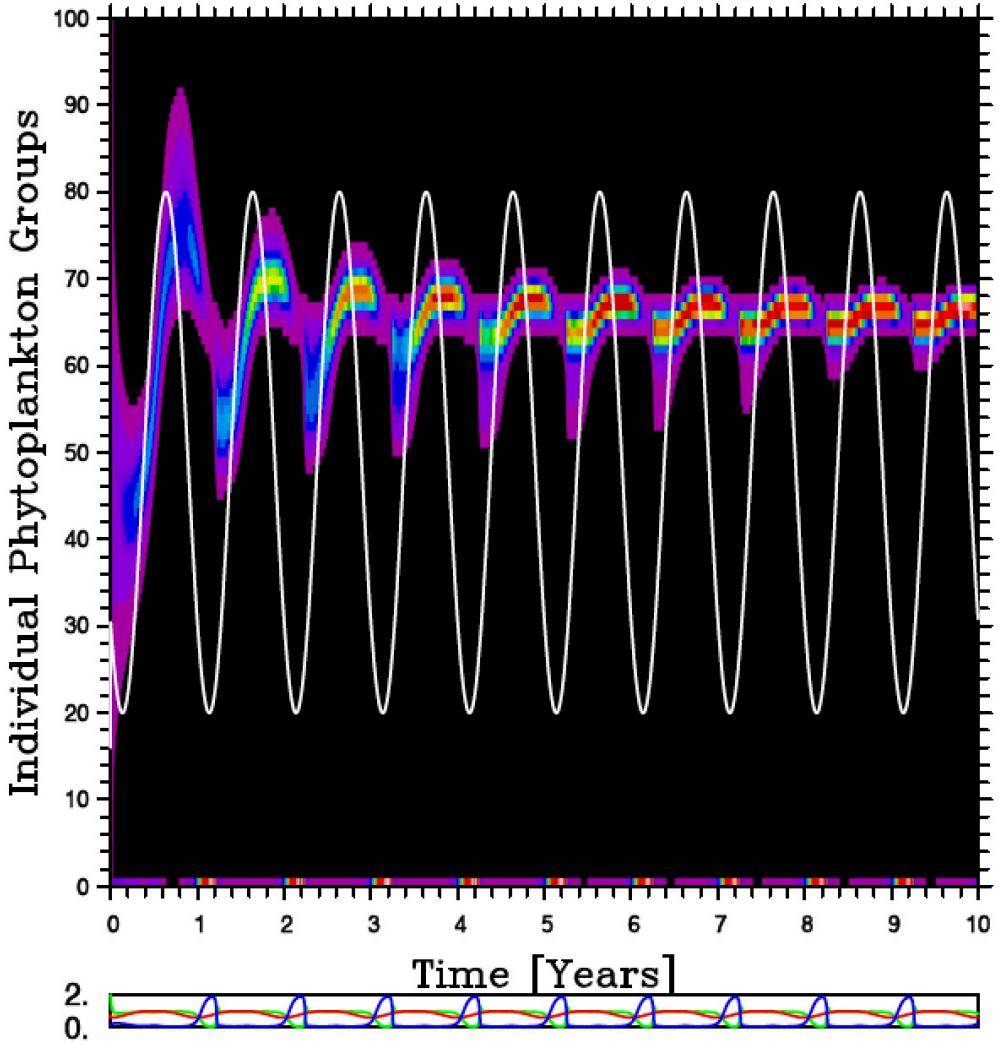


Follows et al., 2007

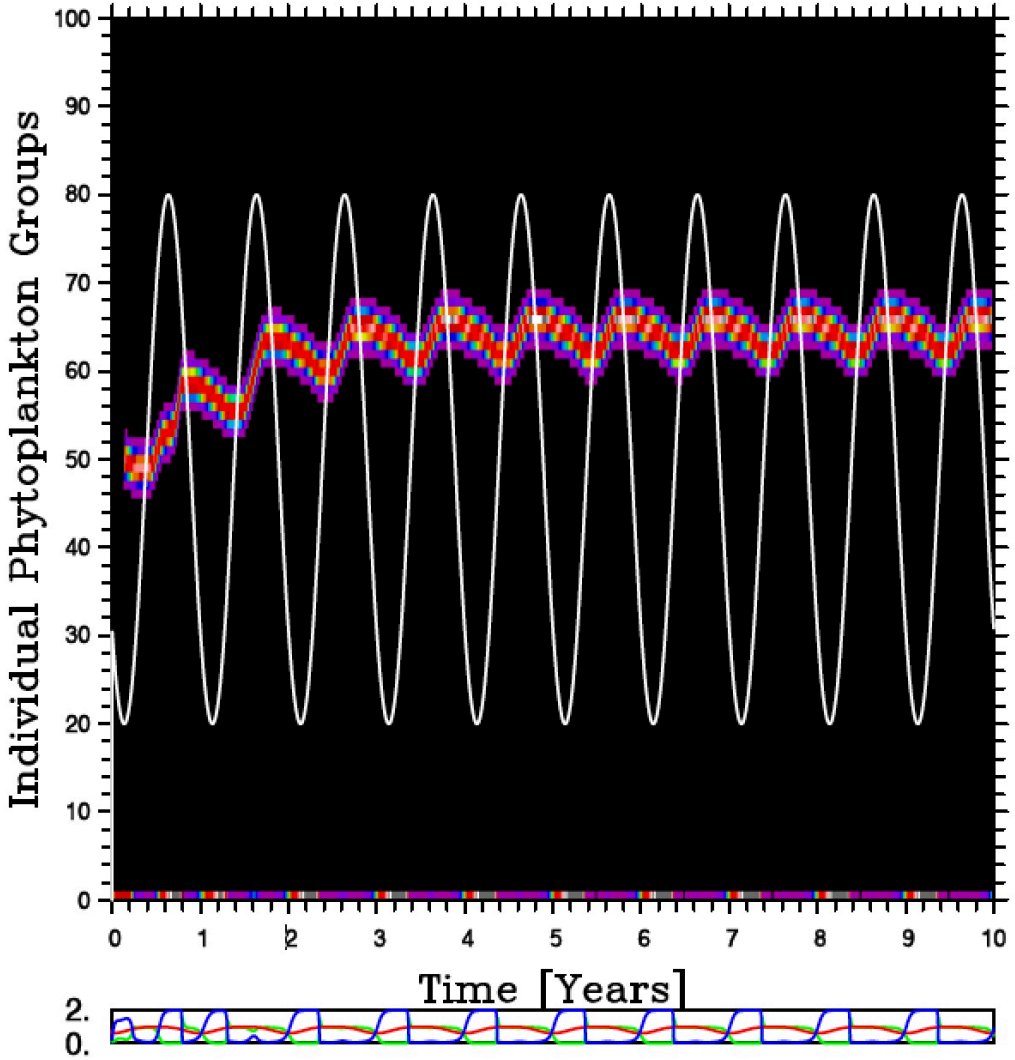


Linking parameters to traits for diversity and acclimation

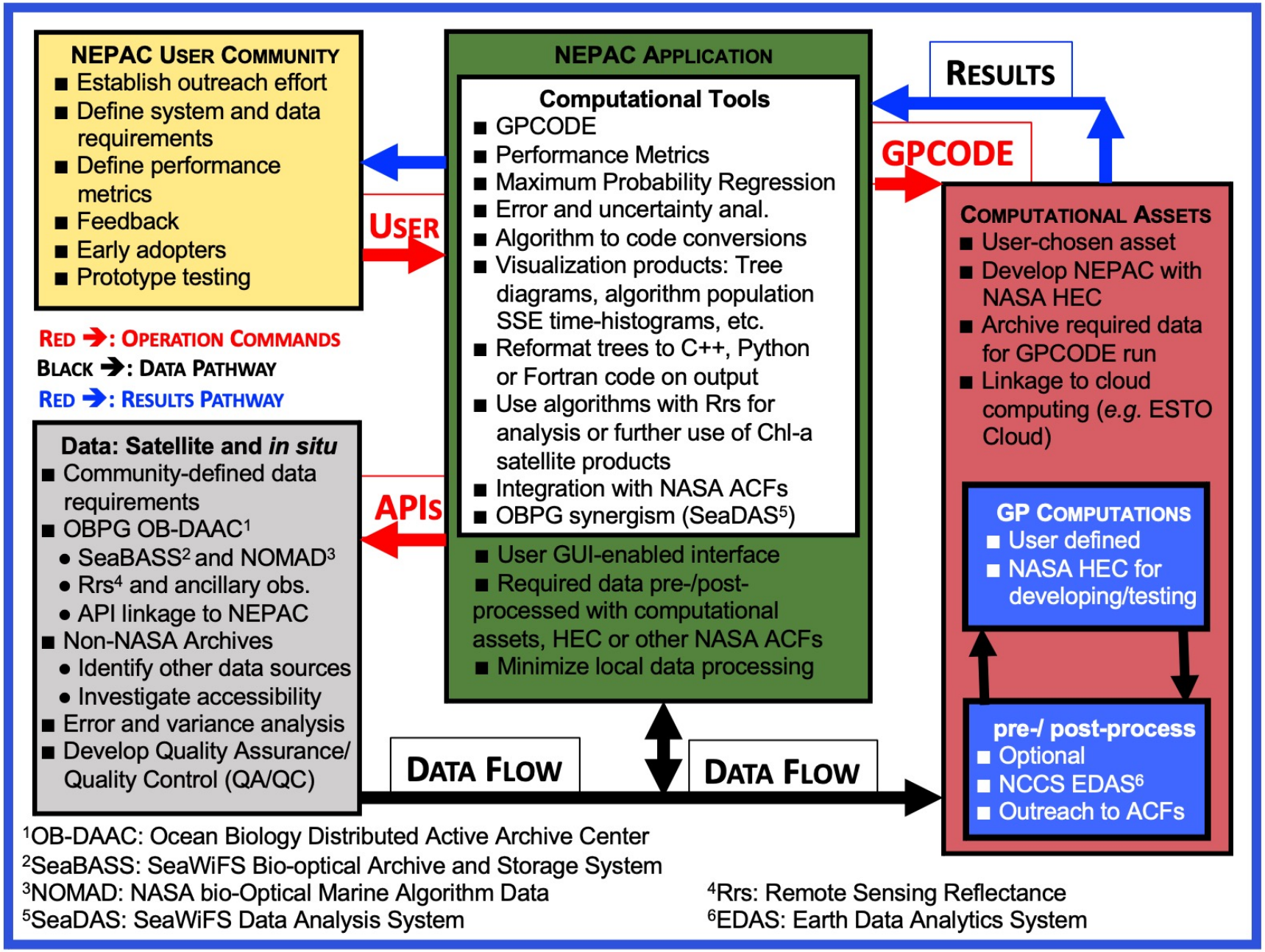
Model with 100 phytoplankton groups



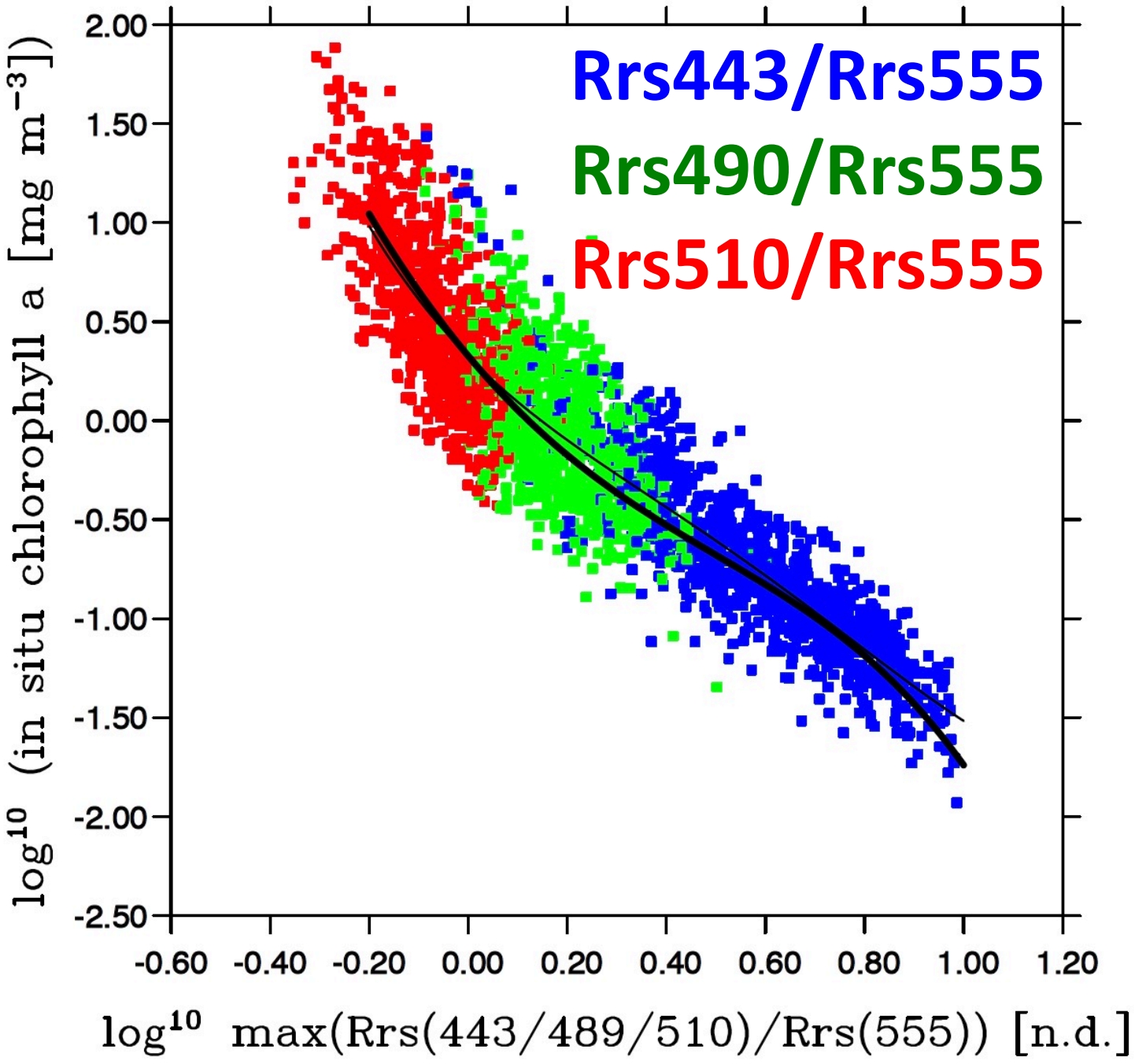
Gaussian Trait-based model (2-3 equations)



NASA Evolutionary Programming Analytic Center (NEPAC)



[1]OB-DAAC: Ocean Biology Distributed Active Archive Center
[2]SeaBASS: SeaWiFS Bio-optical Archive and Storage System
[3]NOMAD: NASA bio-Optical Marine Algorithm Data
[5]SeaDAS: SeaWiFS Data Analysis System
[4]Rrs: Remote Sensing Reflectance
[6]EDAS: Earth Data Analytics System

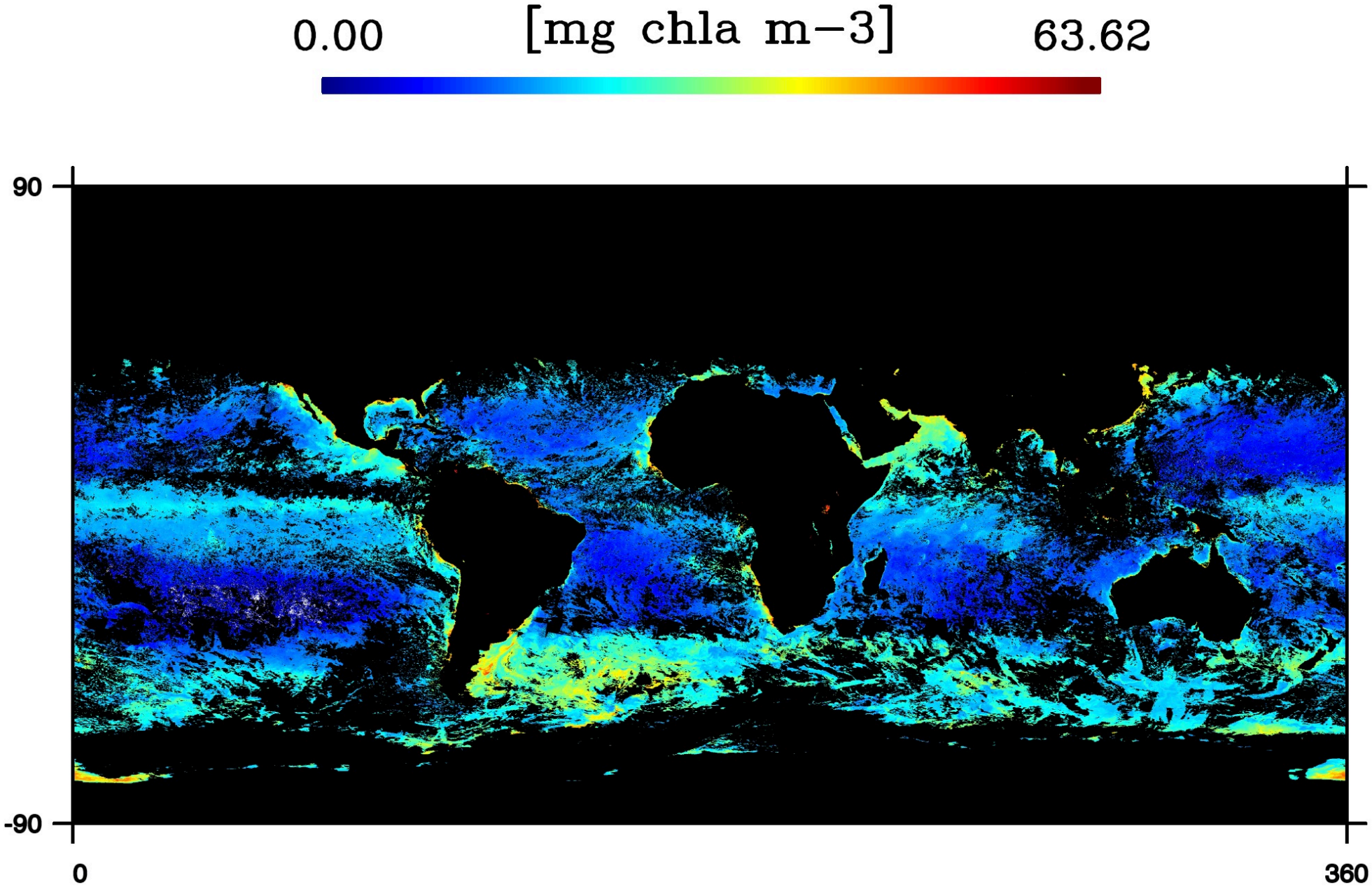

Rrs443/Rrs555
Rrs490/Rrs555
Rrs510/Rrs555
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
-0.60
-0.40
-0.20
0.00
0.20
0.40
0.60
0.80
1.00
1.20
log10 (in situ chlorophyll a [mg m-3])
log10 max(Rrs(443/489/510)/Rrs(555)) [n.d.]

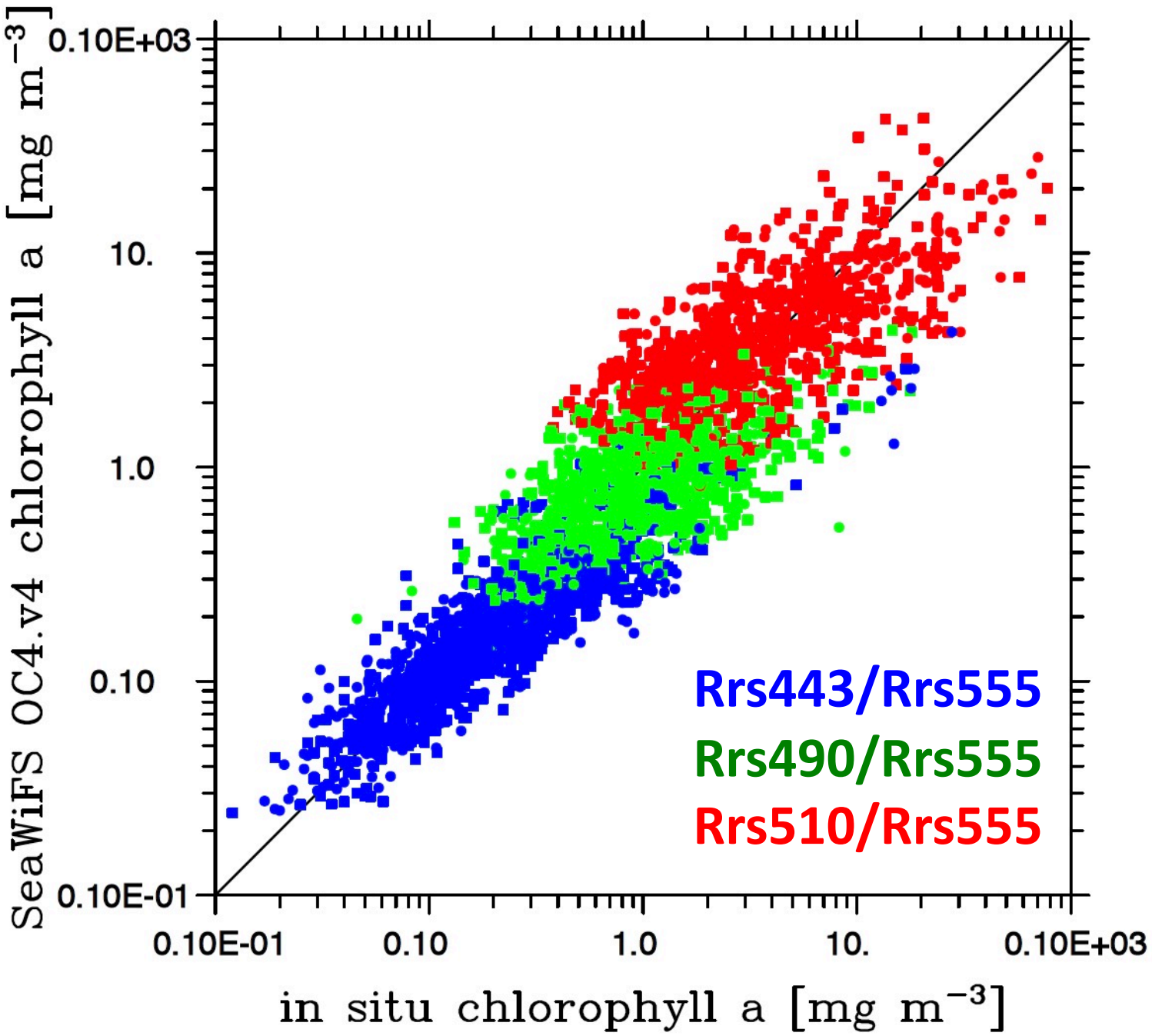
OC-4 chlorophyll a algorithm

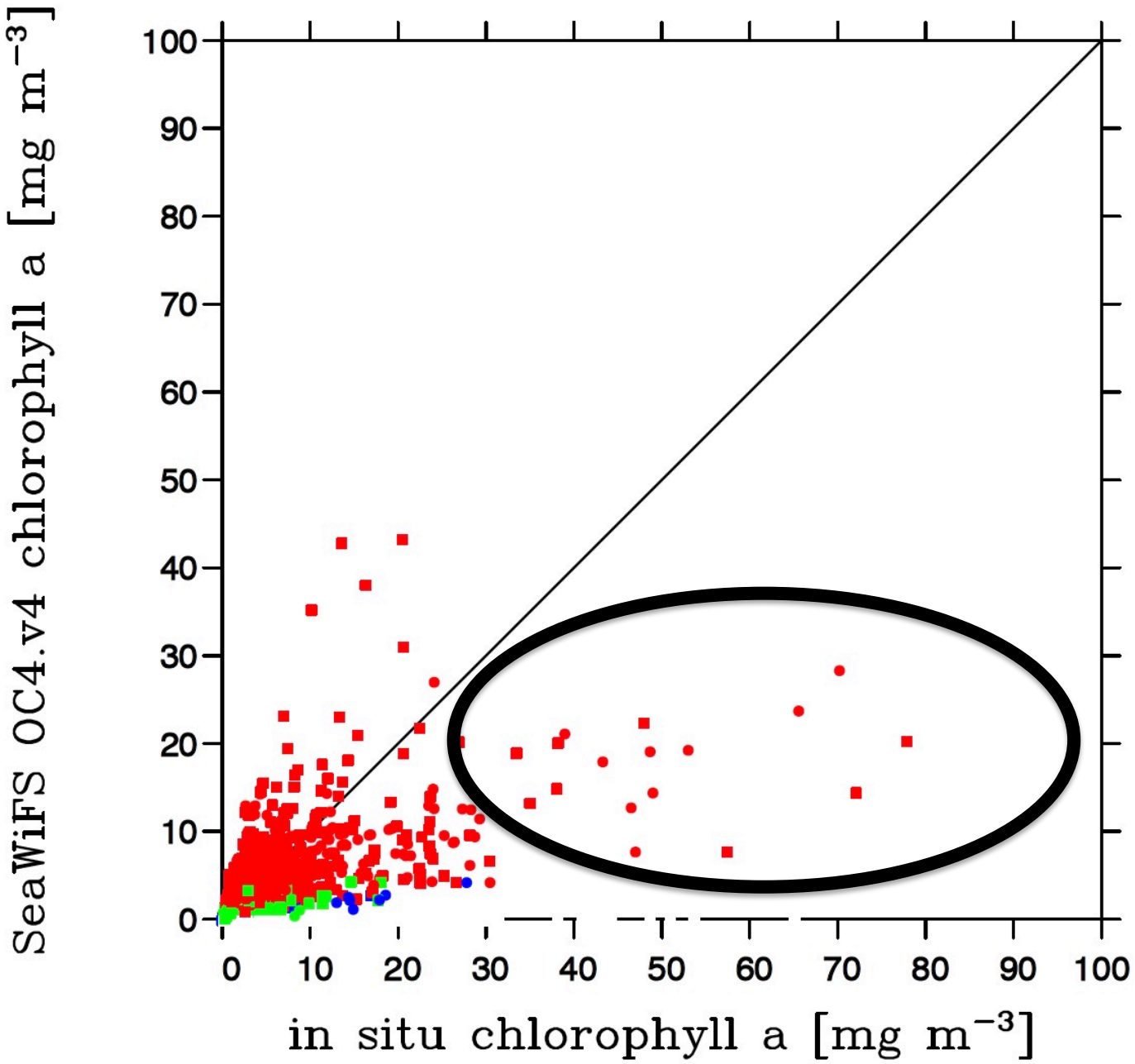
$$V=\log^{10}(MAX[Rrs(\lambda_{443}), Rrs(\lambda_{490}), Rrs(\lambda_{510})] / Rrs(\lambda_{555}))$$

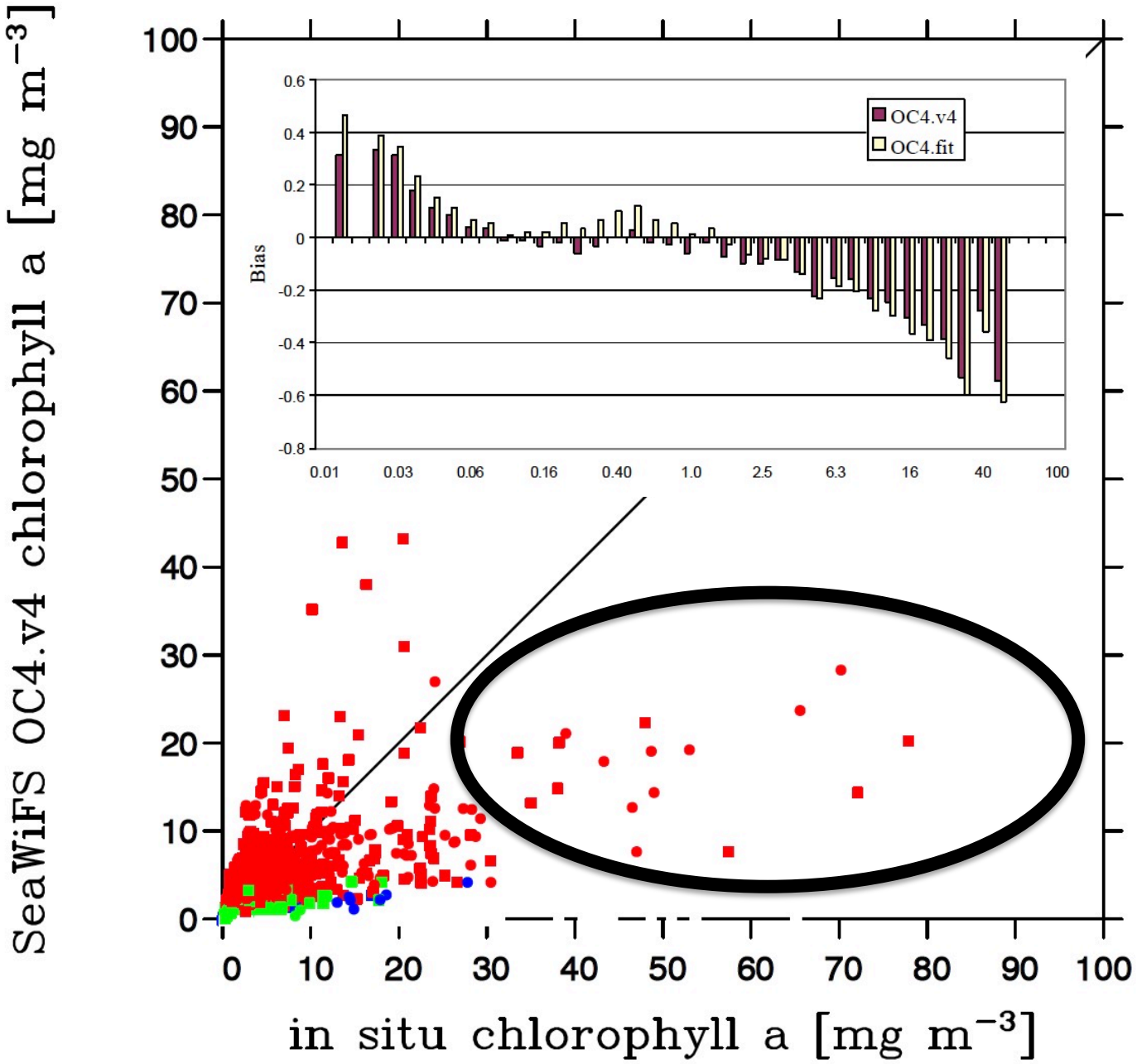
$$\log^{10}(Chla)=(a_0 + a_1V + a_2V^2 + a_3V^3 + a_4V^4)$$

SeaWiFS OC4 Chl a [mg chla m−3]



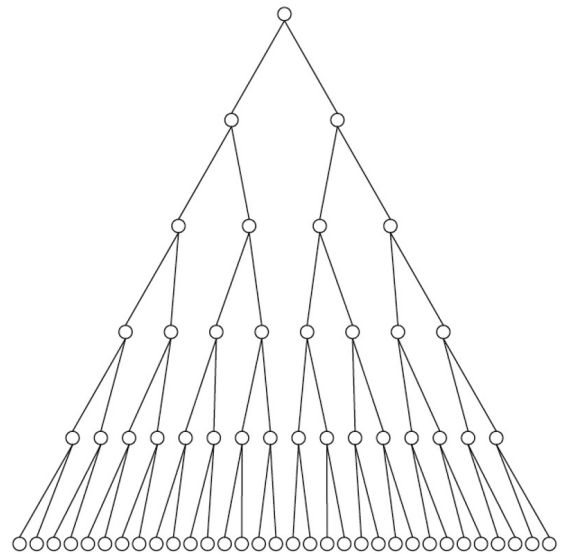

SeaWiFS OC4.v4 chlorophyll a [mg m^{-3}]
0.10E+03
10.
1.0
0.10
0.10E-01
0.10E-01
0.10
1.0
10.
0.10E+03
in situ chlorophyll a [mg m^{-3}]
Rrs443/Rrs555
Rrs490/Rrs555
Rrs510/Rrs555


SeaWiFS OC4.v4 chlorophyll a [mg m^{-3}]
in situ chlorophyll a [mg m^{-3}]
0
10
20
30
40
50
60
70
80
90
100


SeaWiFS OC4.v4 chlorophyll a [mg m^{-3}]
in situ chlorophyll a [mg m^{-3}]
0
10
20
30
40
50
60
70
80
90
100
Bias
OC4.v4
OC4.fit
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0.01
0.03
0.06
0.16
0.40
1.0
2.5
6.3
16
40
100

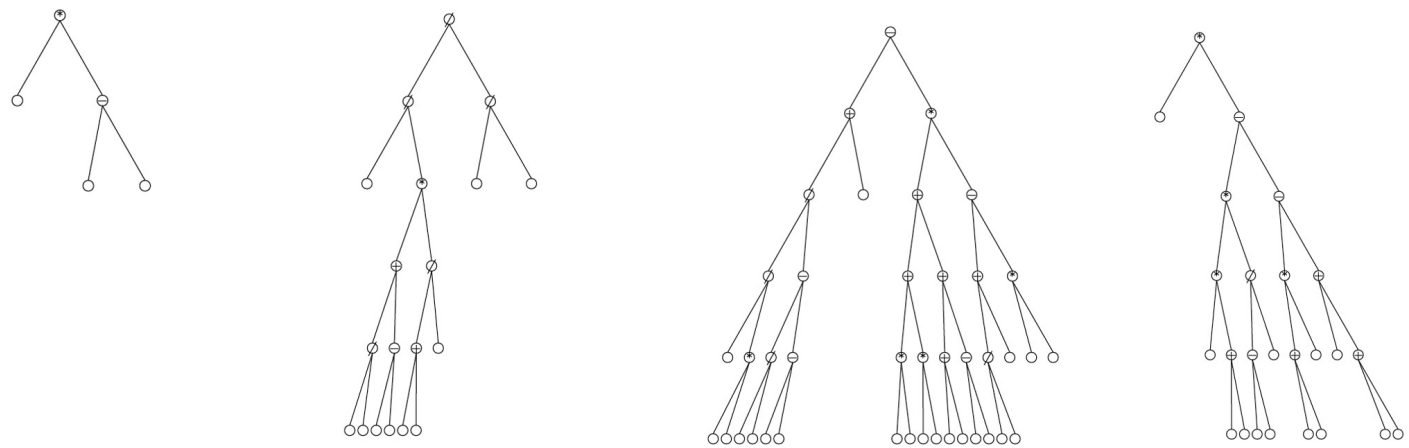
- How to create a program that can manipulate ANY equation, including ones with observations, free parameters, and functions or operators such as “+”, “-”, “*”, “/”, and even items such as Booleans “IF/THEN”, etc?

Binary Tree Architecture



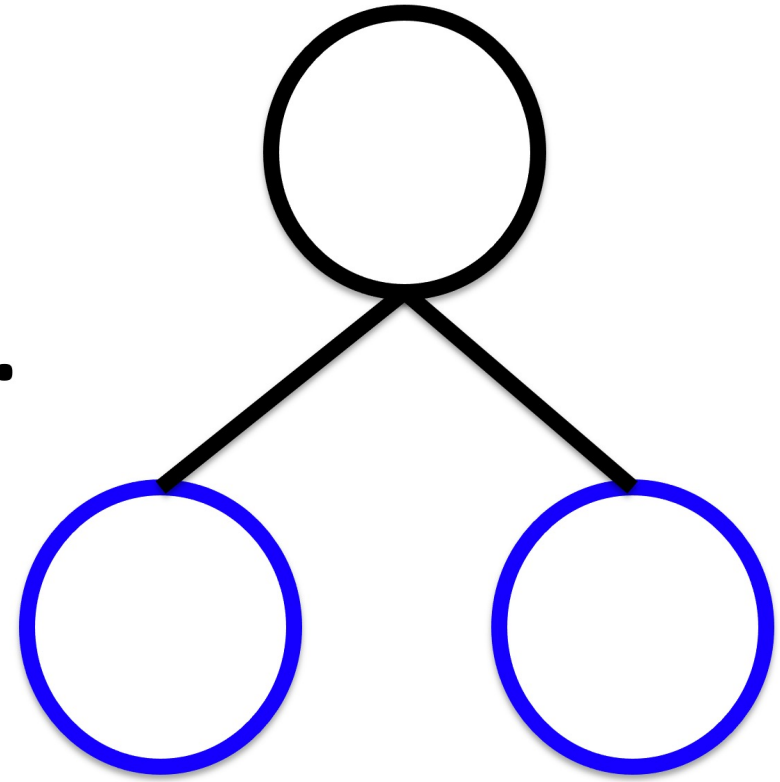
Empty and Full Binary Tree Map

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]

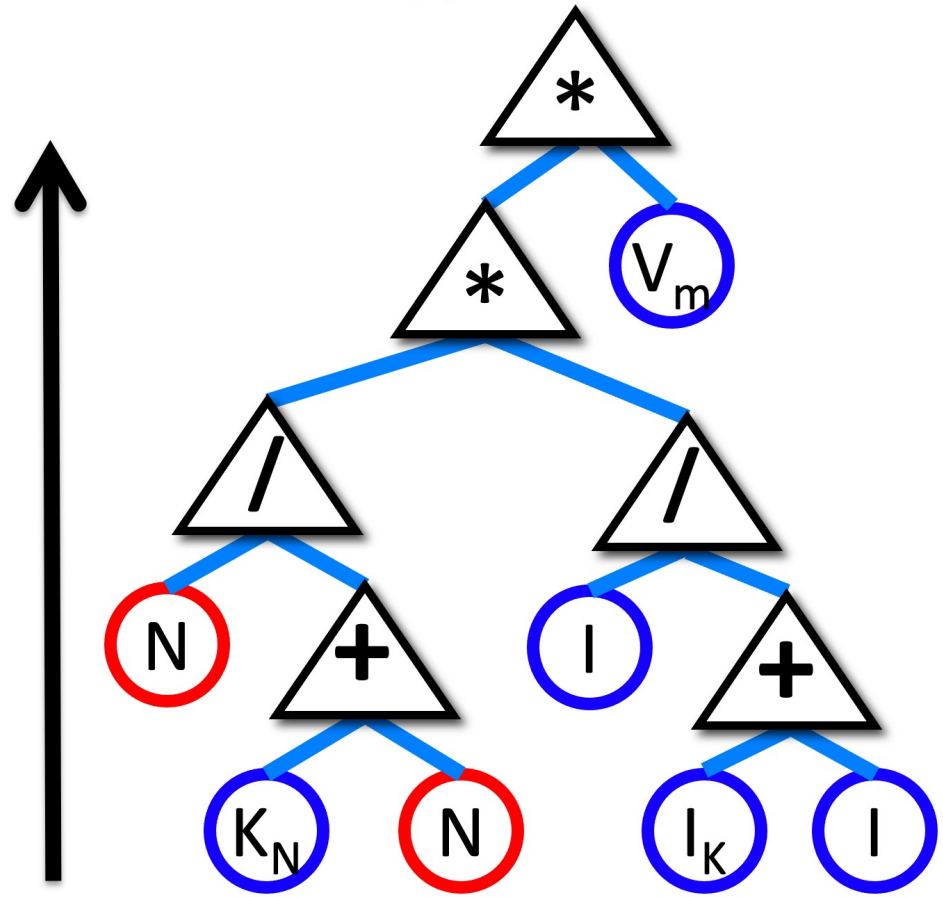
Calculation starts from top of tree
Performance improved by orders of magnitude!

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

LISP prefix representation:

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

Calculation of Tree value moves from the bottom up.

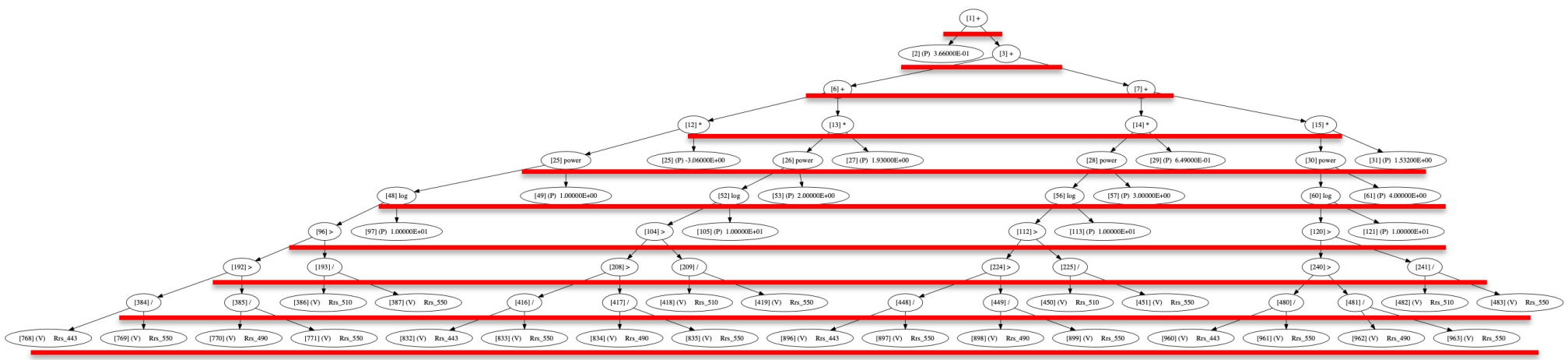


OC-4 chlorophyll a algorithm

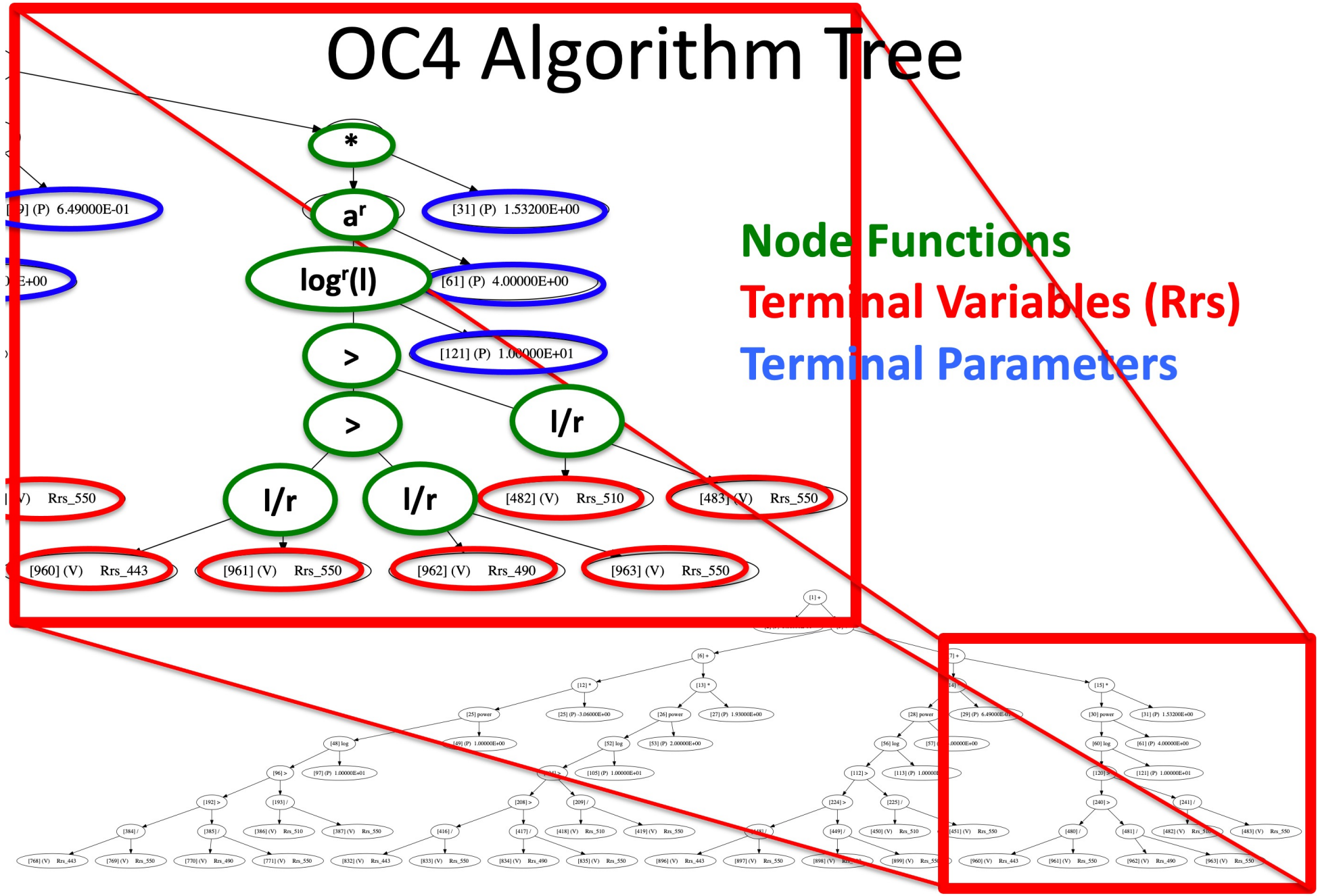
$$V = \log^{10}(\ MAX[\ Rrs(\lambda_{443}), Rrs(\lambda_{490}), Rrs(\lambda_{510})\ ]\ /\ Rrs(\lambda_{555})\ )$$

$$\log^{10}(Chla) = (a_0 + a_1 V + a_2 V^2 + a_3 V^3 + a_4 V^4)$$

OC-4 Algorithm Tree➔<u>10 levels</u>



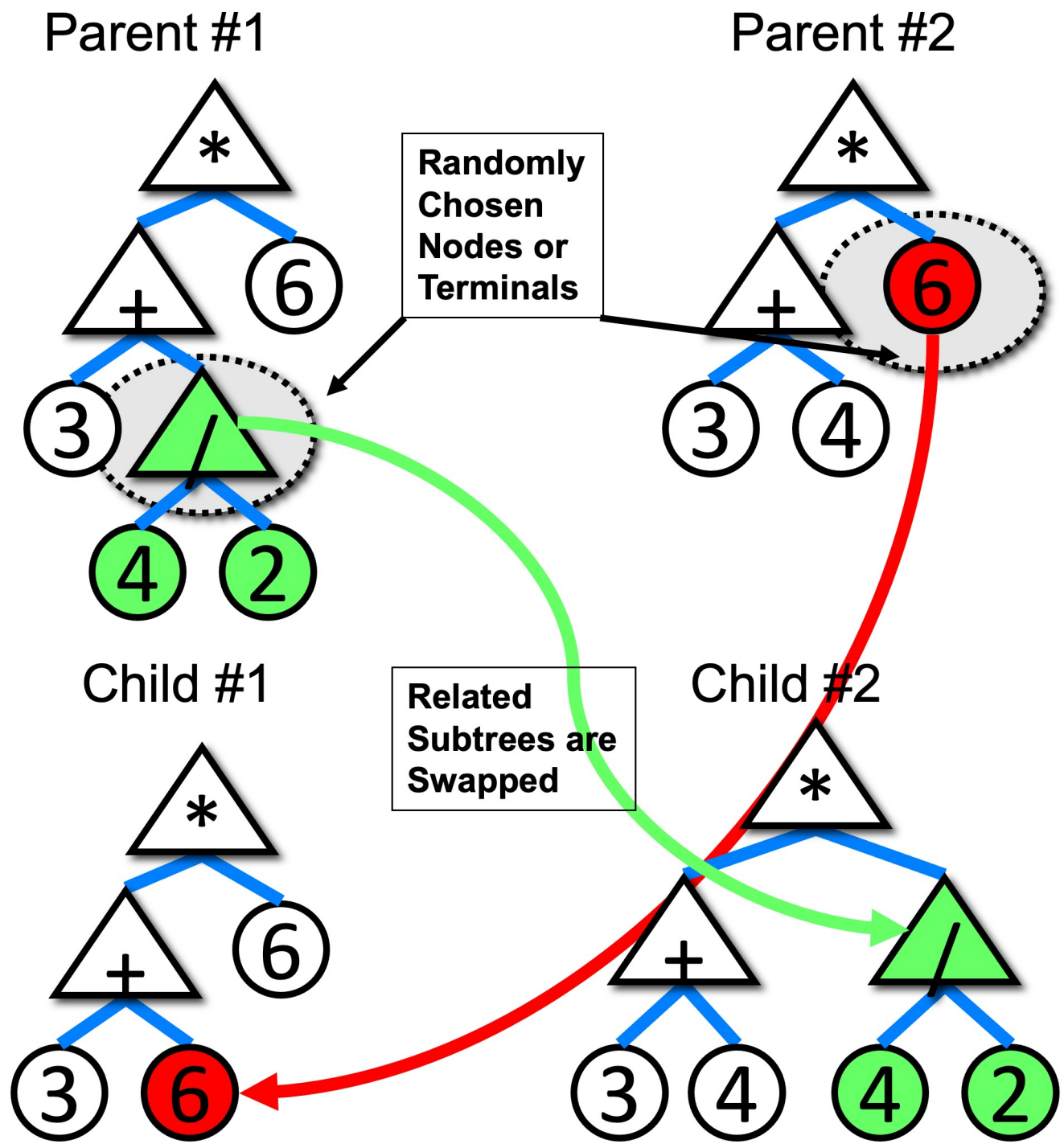
OC4 Algorithm Tree



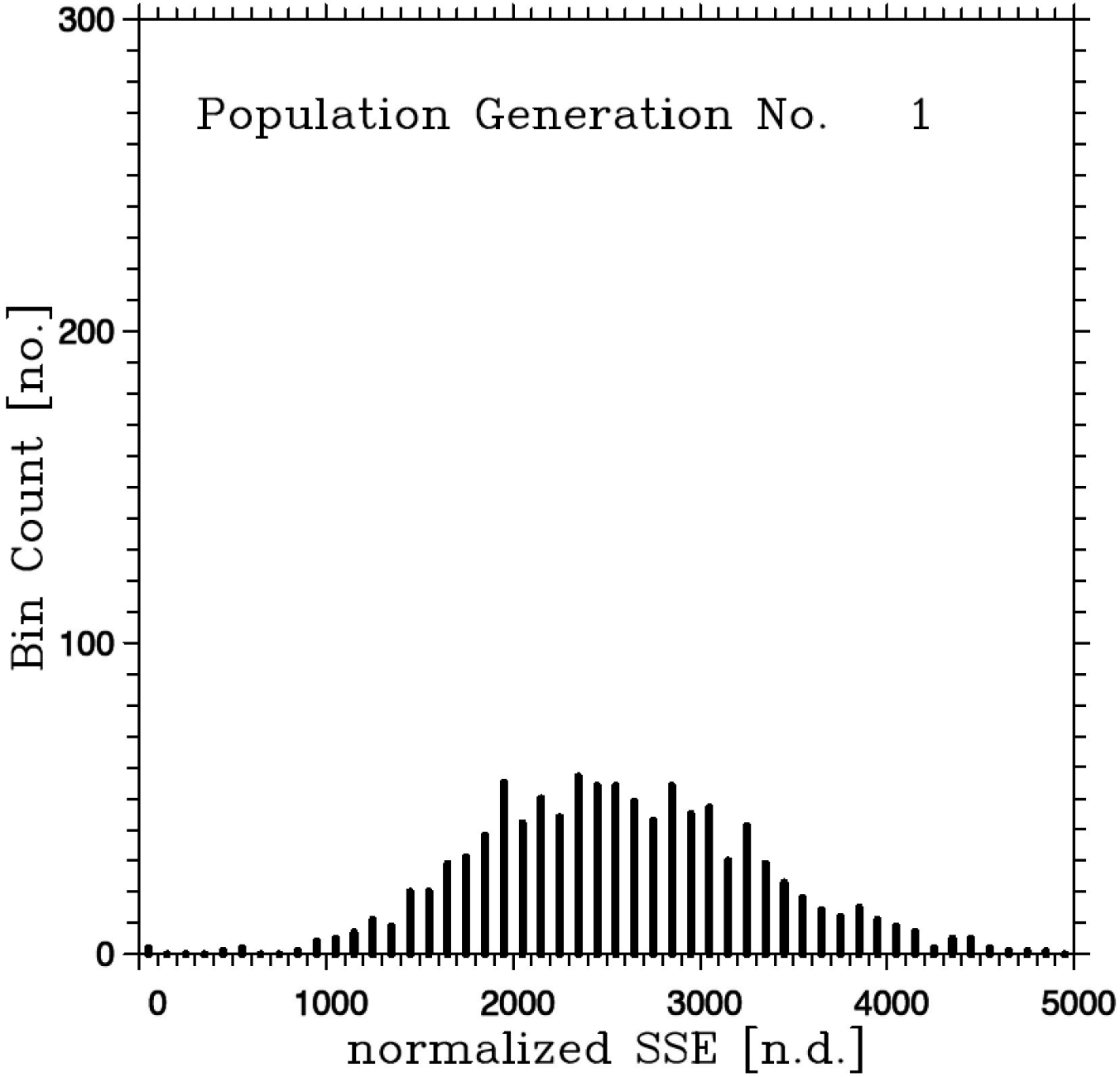
Genetic Programming Overview

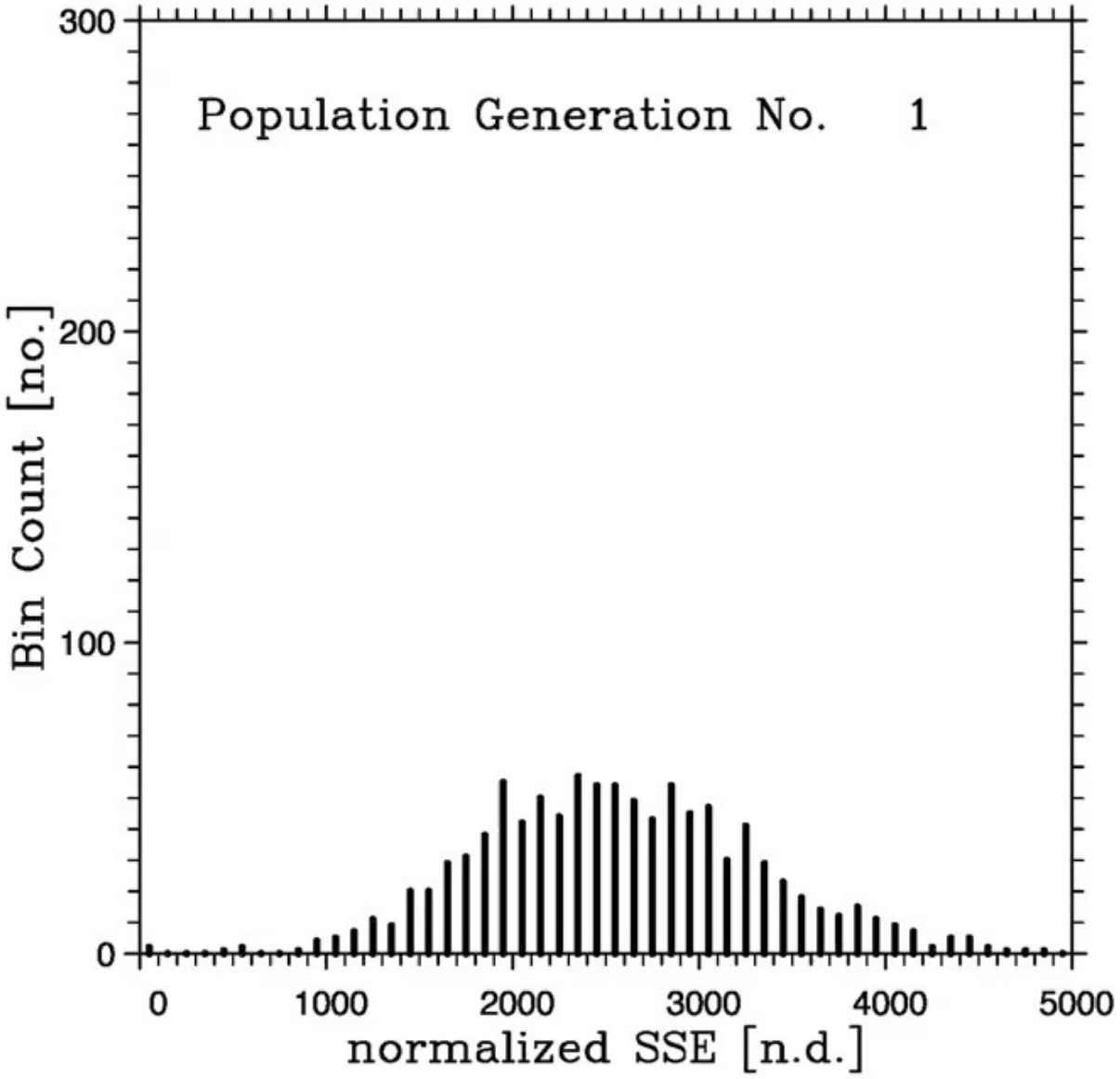
1) Generate initial random population of equations/models
2) Calculate fitness of all individual equations/model
   - Strip out all model/equations parameters
   - Carry out Genetic Algorithm Optimization
   - Finish off with localized LMDIF Optimization
3) Randomly select, based on fitness, for:
   - Asexual reproduction
   - Sexual reproduction (i.e. Tournament Selection)
   - Mutations
4) Calculate fitness of new individuals (same as 2 above)
5) Test for completion criteria (low SSE value met?)
6) No: <u>go to 3</u>; Yes: <u>end program</u>

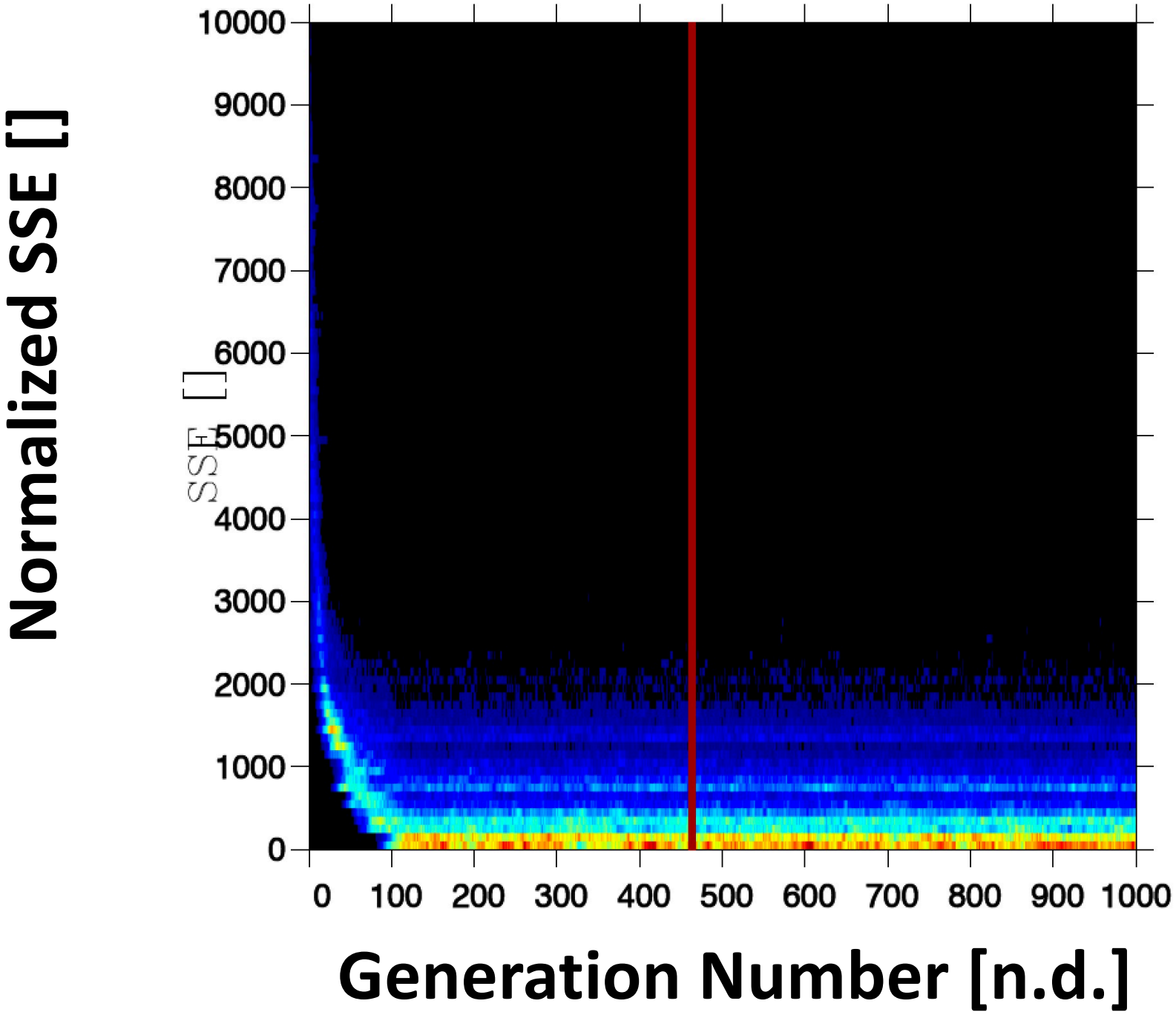
Cross methodology

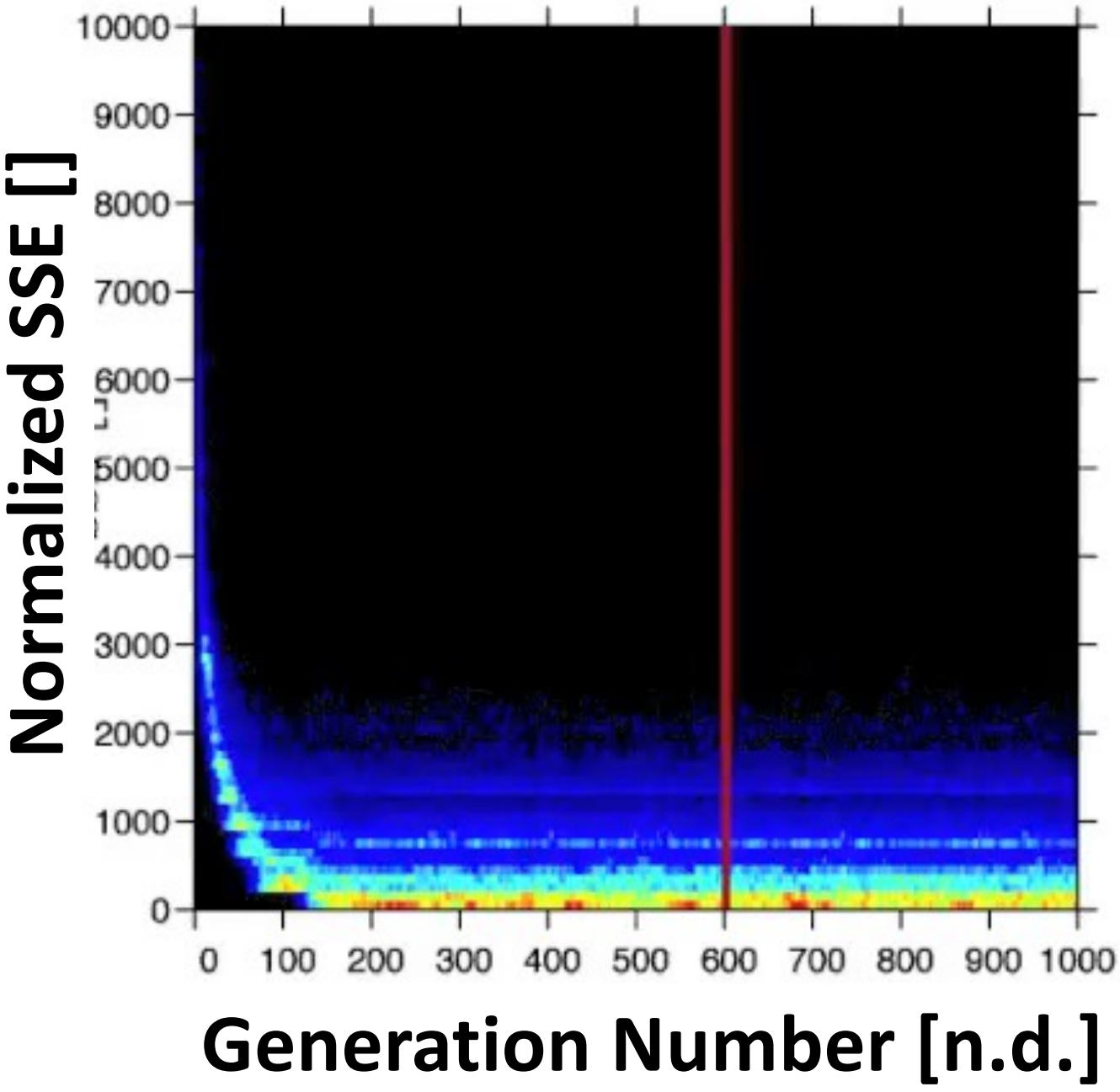


Simple Lotka-Volterra Model




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

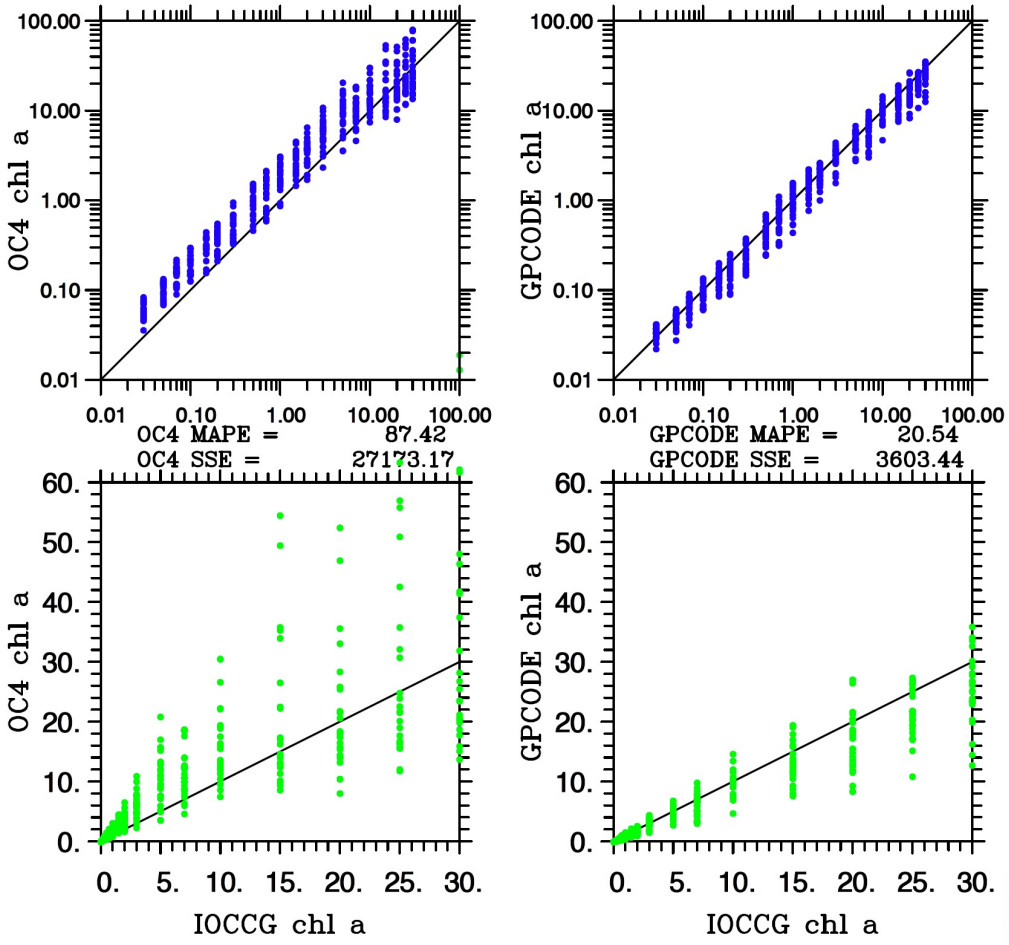

Normalized SSE []
SSE []
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
600
700
800
900
1000
Generation Number [n.d.]

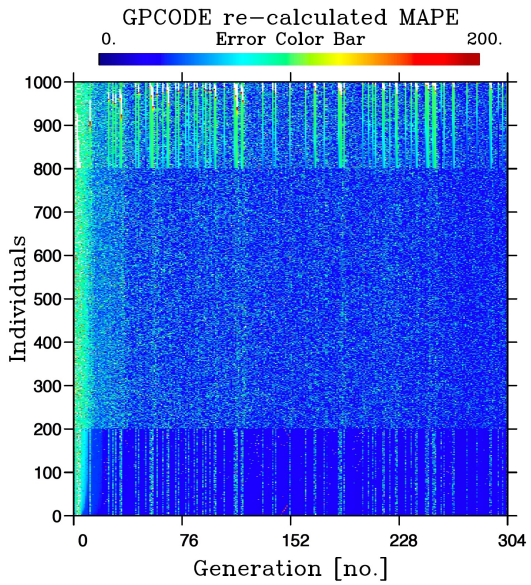

Normalized SSE []
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
600
700
800
900
1000
Generation Number [n.d.]

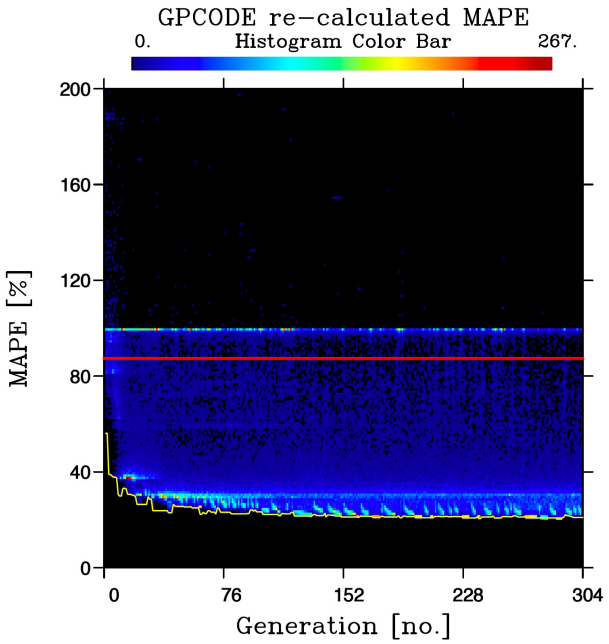
GPCODE Chla algorithms using IOCCG data for training

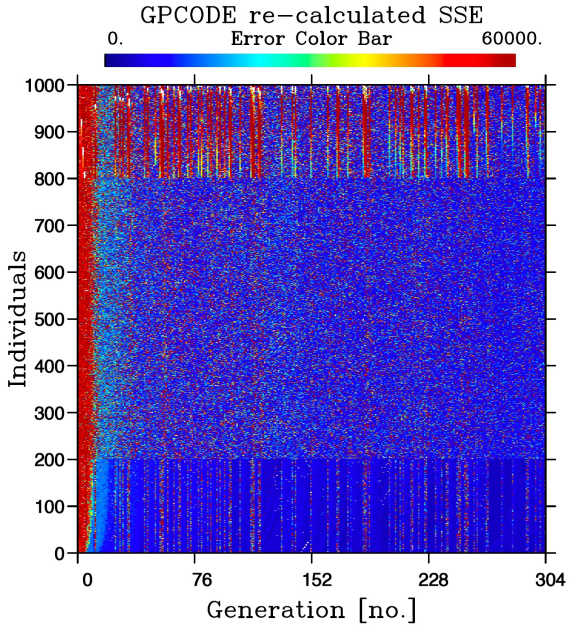
Case 12: GP_Tree_All_SSE

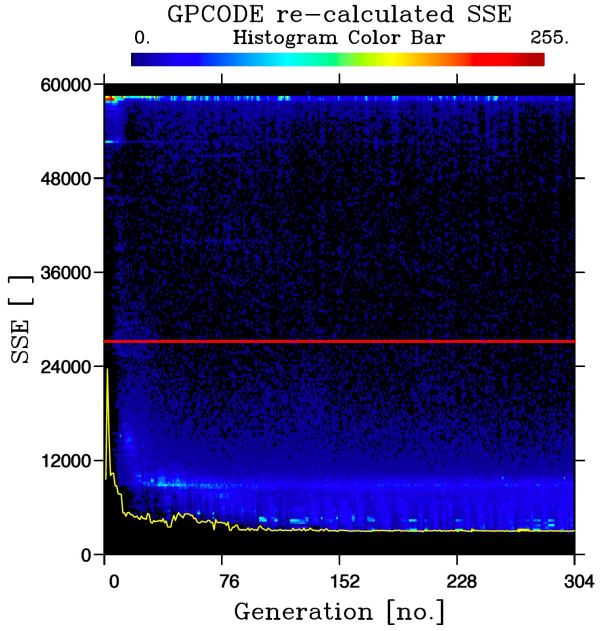
n_GP_Generations=304
n_GP_Individuals=1000
MAPE metric
No Bootstrapping (BS)
# of parameters in best tree: 9
Best SSE Tree (gen. = 238 , ind. = 1)
Best MAPE Tree (gen. = 206 , ind. = 614)



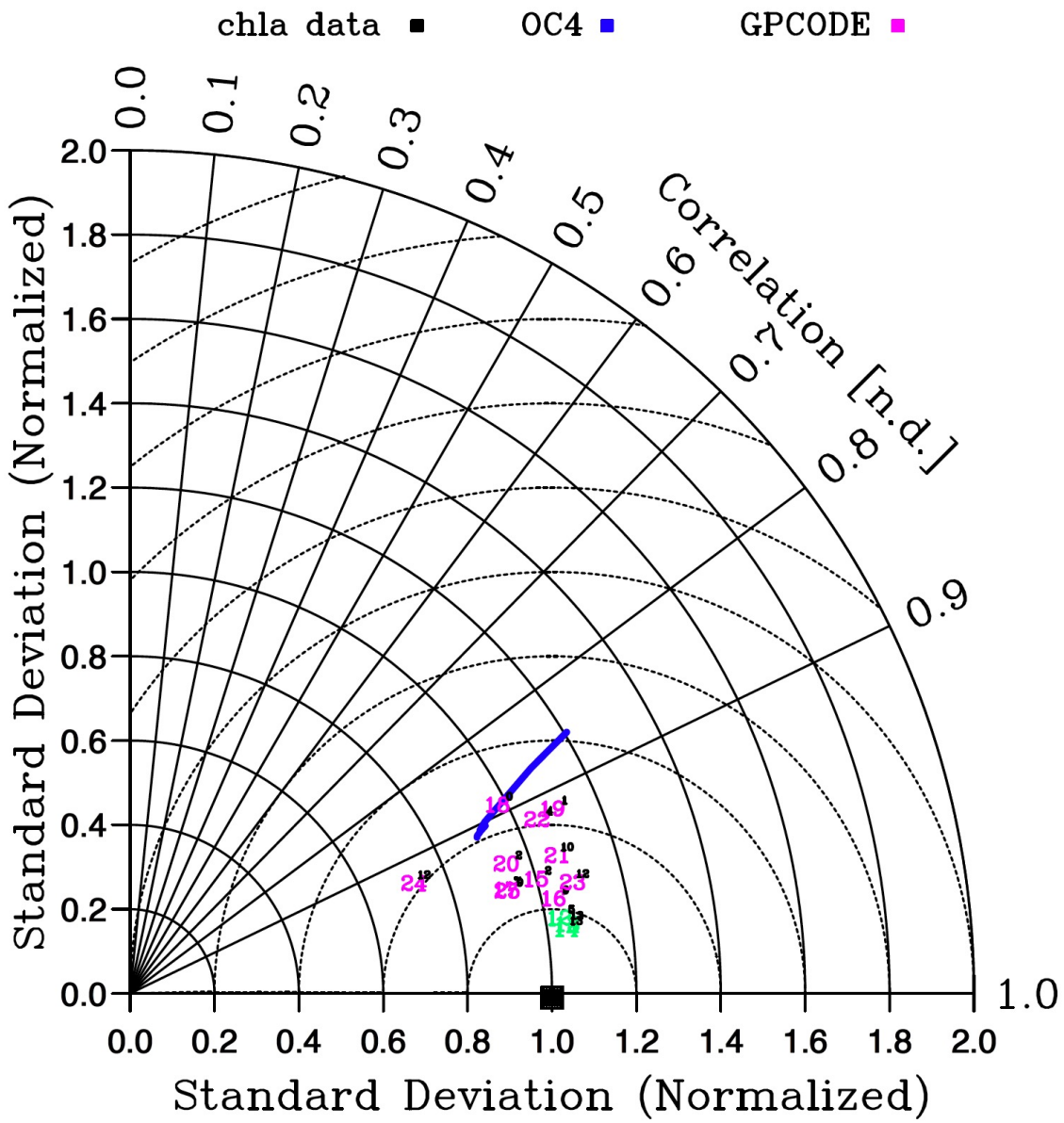


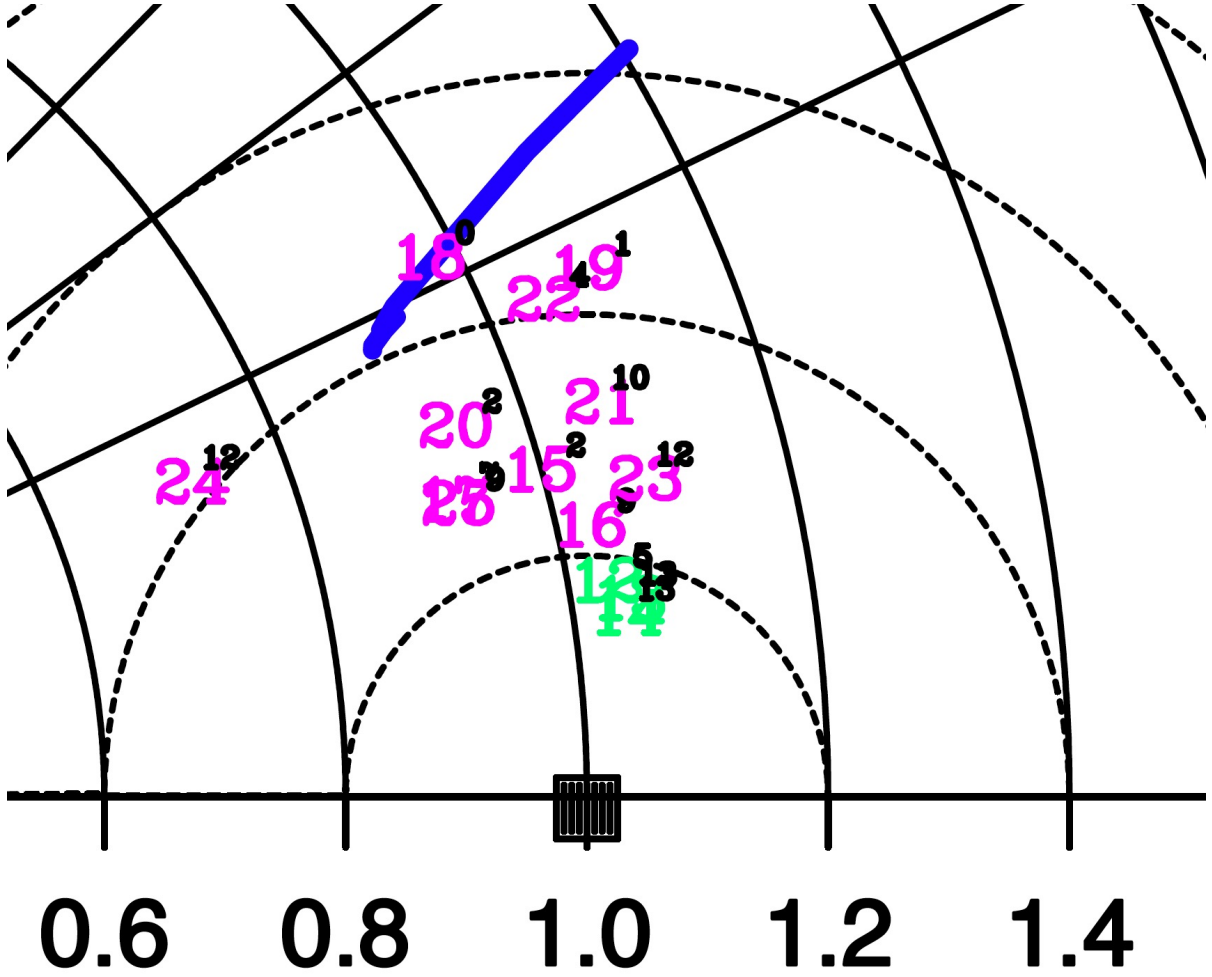




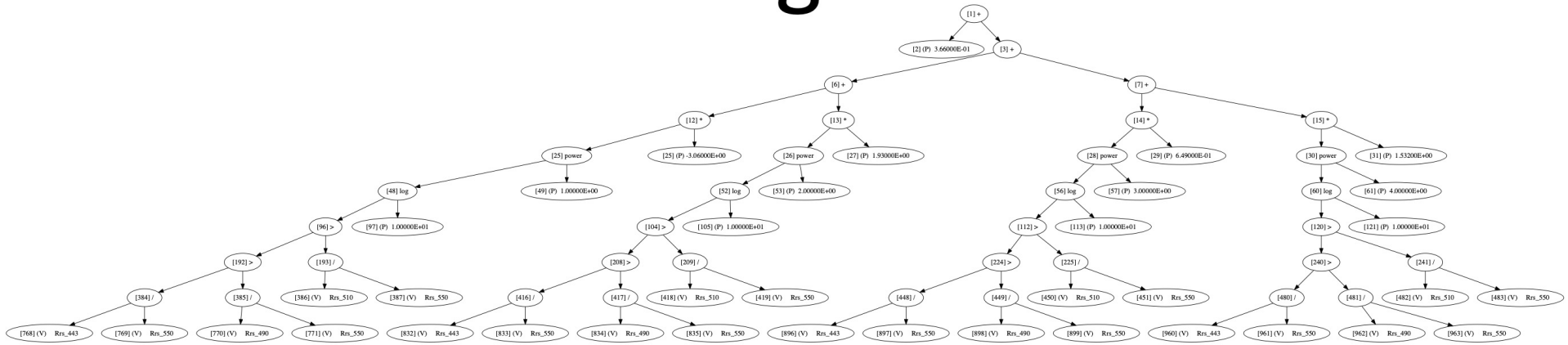


GPCODE chla algorithms using IOCCG data

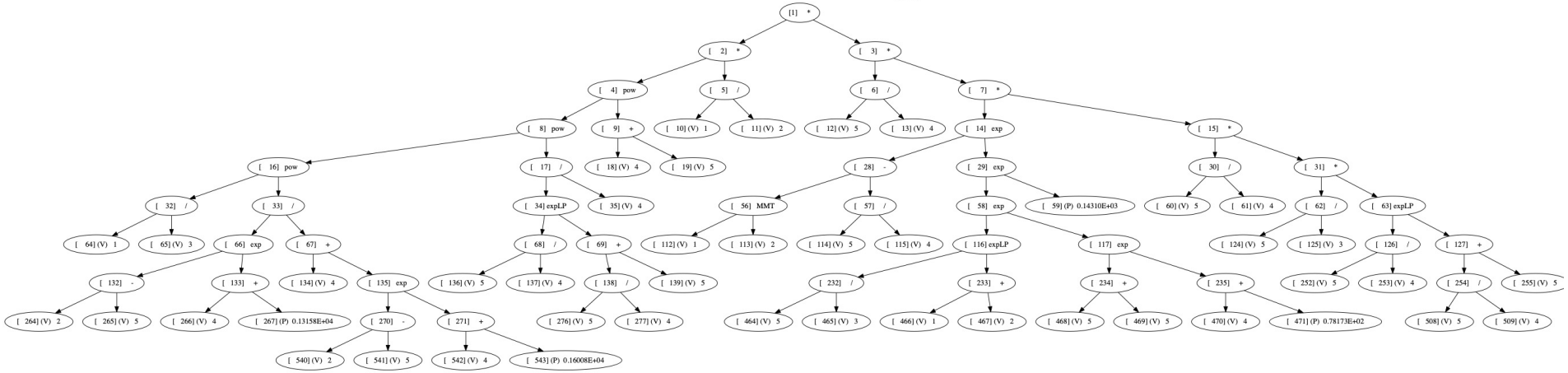




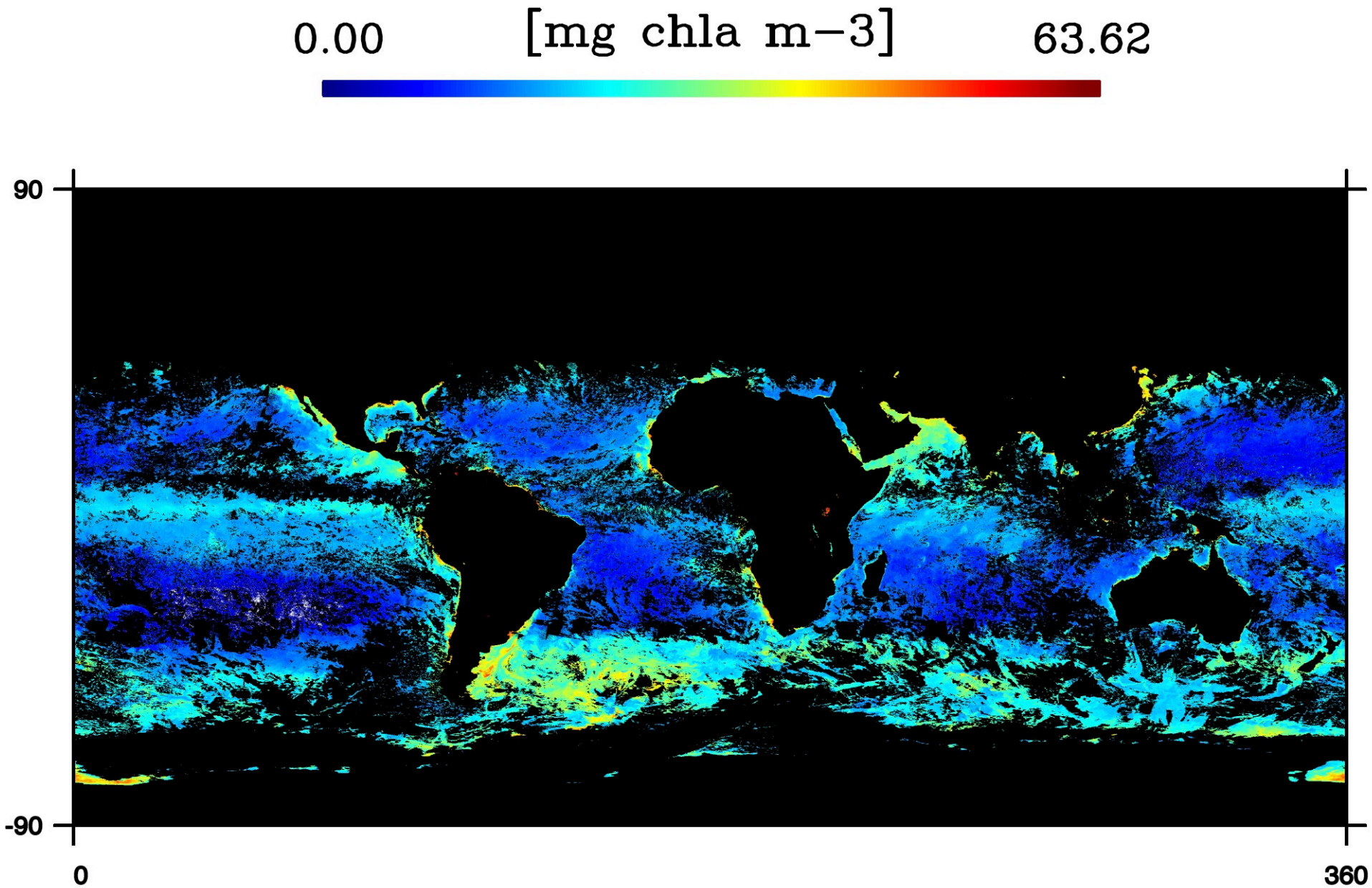
OC-4 Algorithm



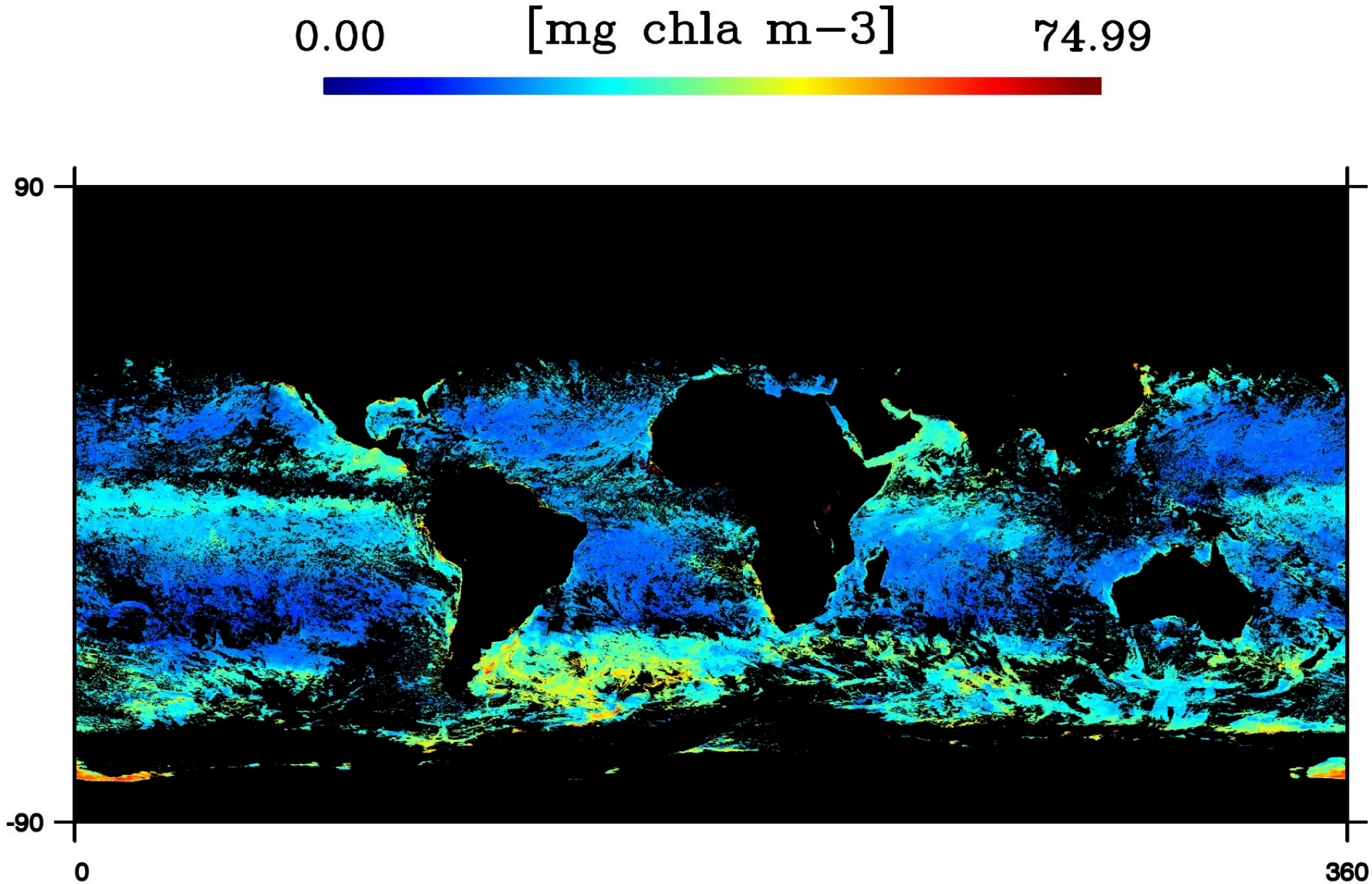
A GPCODE Algorithm



SeaWiFS OC4 Chl a [mg chla m−3]

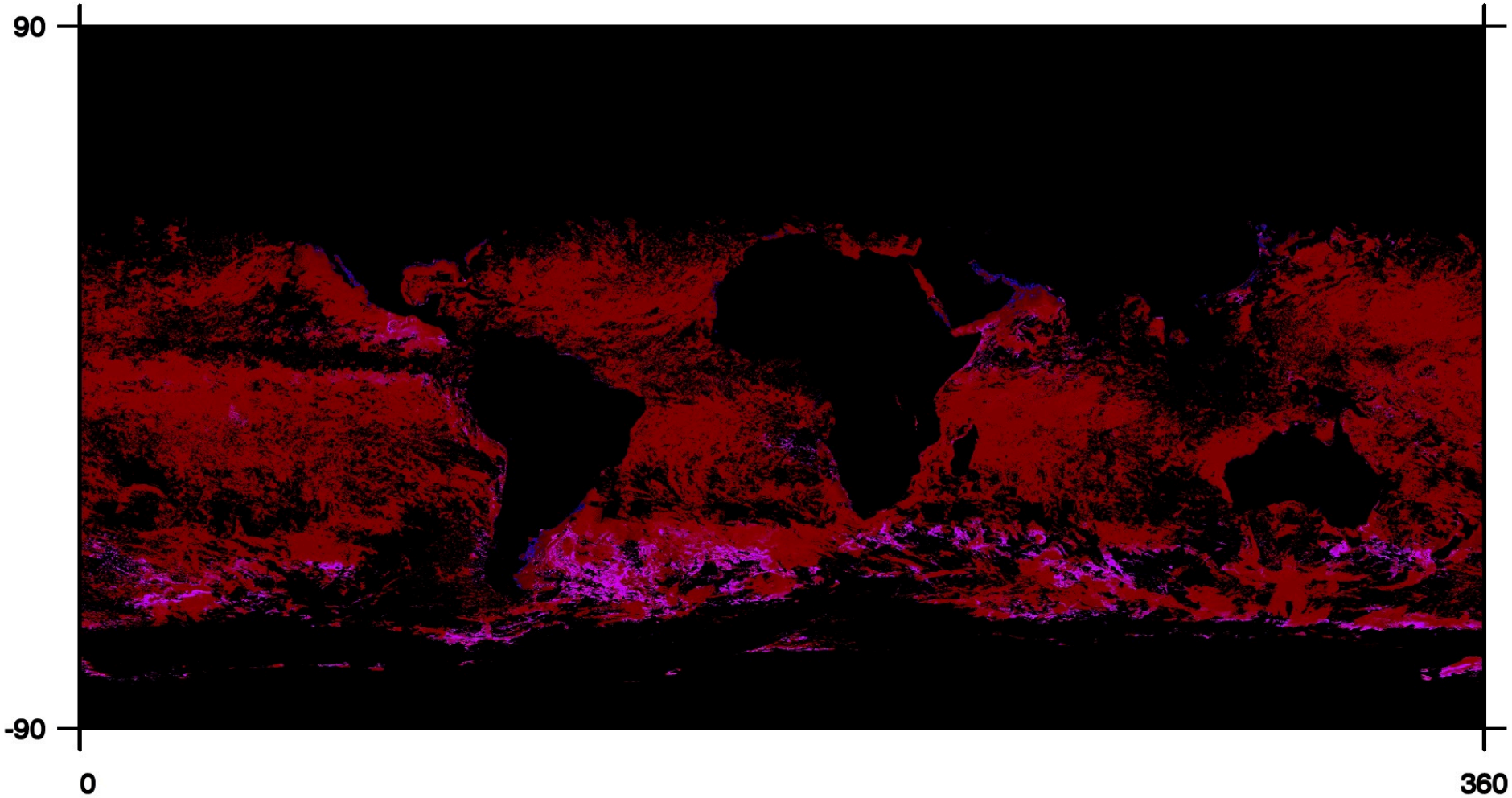


SeaWiFS GPCODE Chl a [mg chla m−3]



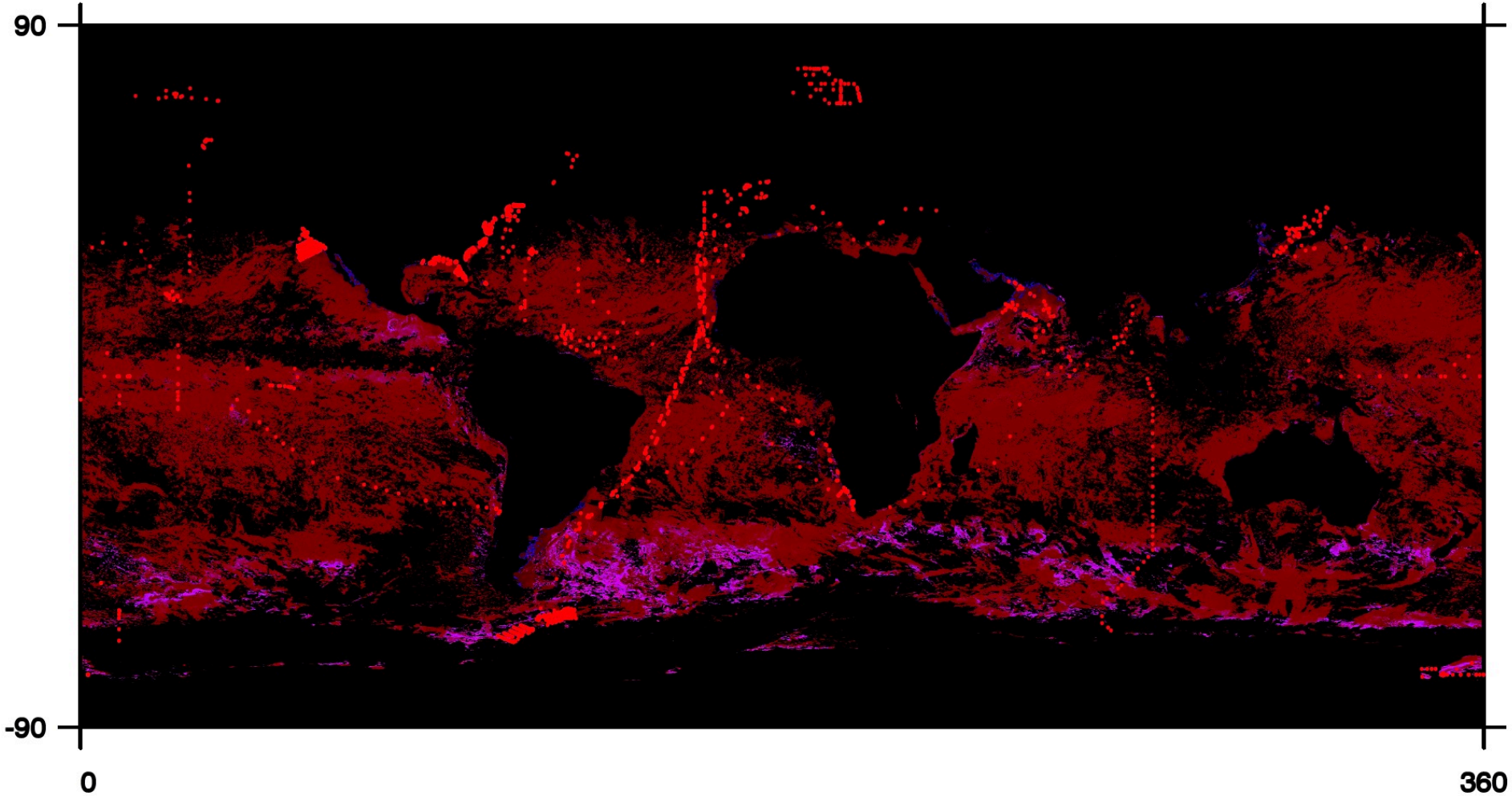
SeaWiFS GPCODE−OC4 Chl a [diff]

−1.00 [mg chla m−3] 1.00



SeaWiFS GPCODE−OC4 Chl a [diff]

−1.00 [mg chla m−3] 1.00



NEPAC Status

- Front-End at beta-test stage
- Working to link algorithm codes to ocean color processing software
- GPCODE has been modified:
  - Bootstrapping
  - Optimizing tree diversity through entropy.
  - Maximum Probability Regression.
- Started testing multiple satellite data sets
- Two summer interns ➔ Prim. Prod. Archive.
- Growing interest to compare blended satellite products with observations.
- Obtaining training data sets for hyperspectral algorithms (2021 IRAD)
- Seeking funding for phytoplankton pigment and functional type alg.

Ocean Color Hyperspectral Inverse Modeling Results

Detecting trends in phytoplankton diversity and biomass[d,e,f]

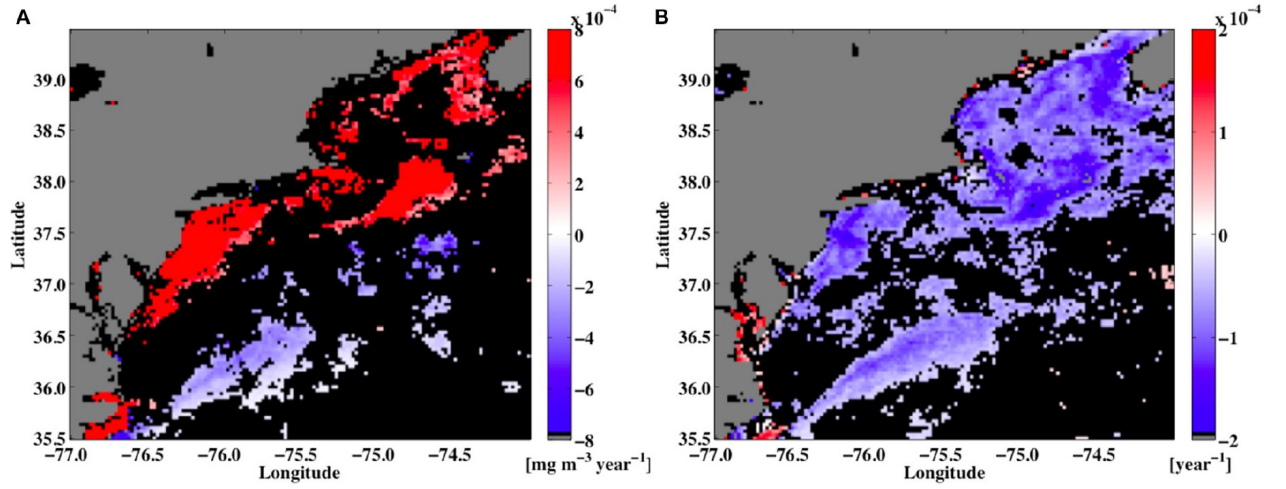
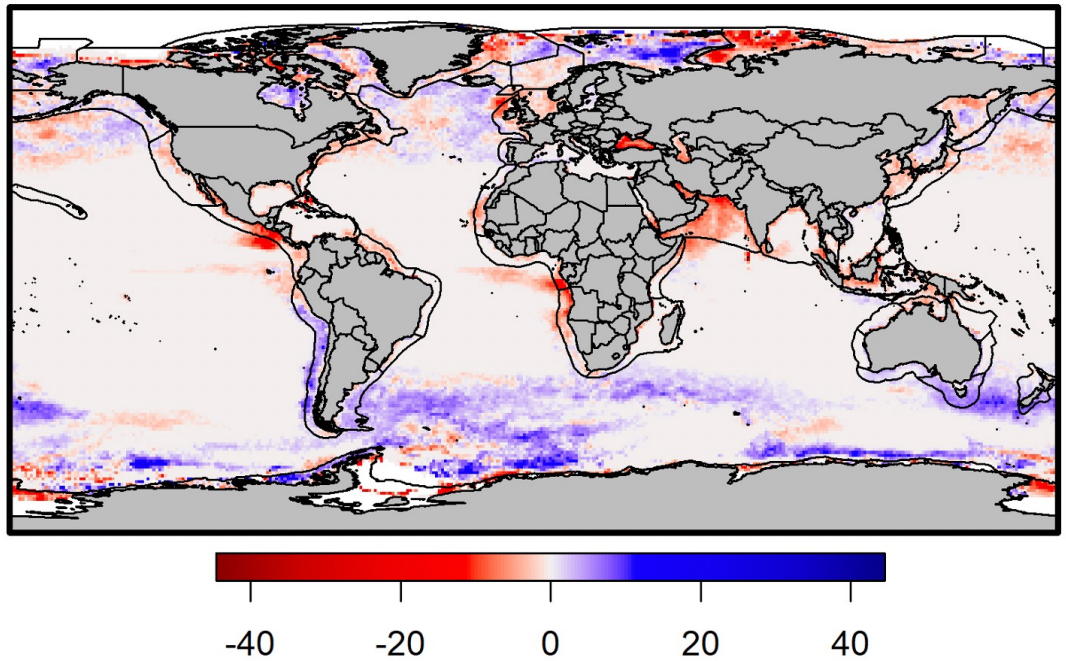


FIGURE 12 | The trend line slopes of **(A)** seatellite chlorophyll a [mg chla m^{-3} $year^{-1}$] and **(B)** shannon Diversity Index [H; $year^{-1}$] at each point over a 15-year period (2002–2016) in the study region. Regions where the slopes were insignificant are shown in black.

Global trend in diatoms [Theil-Sen slope % $decade^{-1}$][a]



(a) 2021: Friedland, K. D., **J. R. Moisan**, A. Maureaud, D. Brady, A. Davies, S. Bograd, R. Watson, and Y. Rousswau, Trends in phytoplankton communities within large marine ecosystems diverge from the global ocean, *Canadian Journal of Fisheries and Aquatic Science*, Accepted.

(b) 2020: Friedland, K. D., R. E. Morse, N. Shackell, J. C. Tam, J. L. Morano, **J. R. Moisan**, and D. C. Brady, Changing physical conditions and lower and upper trophic level responses on the US Northeast Shelf, *Frontiers in Marine Science*, https://doi.org/10.3389/fmars.2020.567445.

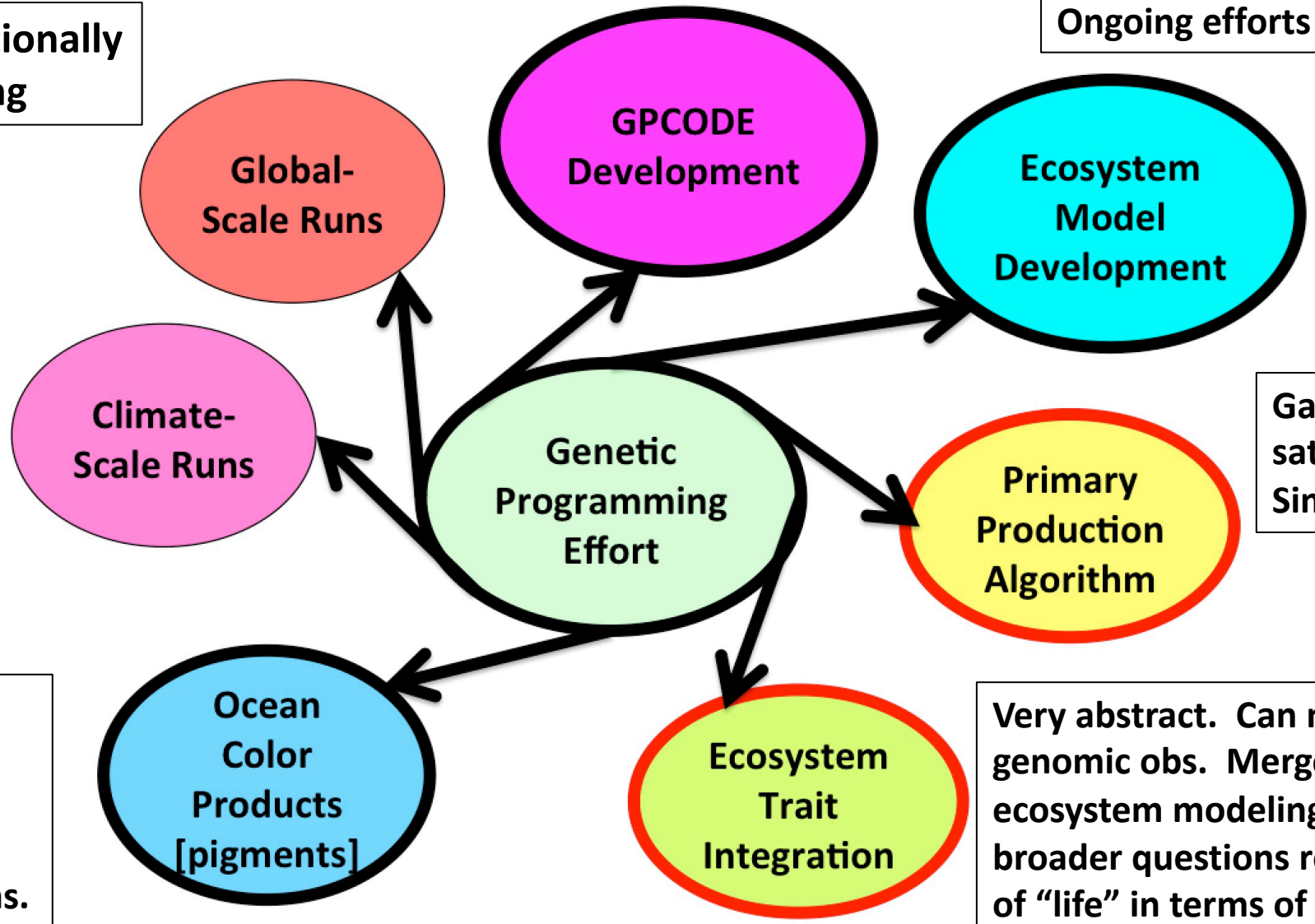
(c) 2020: Wang, G., and **J. R. Moisan**, Remote Sensing of Phytoplankton Pigments, In "Plankton Communities", Leonel Pereira (ed.), Open Access, IntechOpen, London, ISBN 978-1-83968-609-2.

(d) 2017: Moisan, T. A., K. Rufty, **J. R. Moisan**, and M. Linkswiler. Satellite observations of phytoplankton functional type spatial distributions, phenology, diversity and ecotones, *Frontiers in Mar. Sci.*, **4**:189, doi: 10.3389/fmars.2017.00189.

(e) 2013: Moisan, T. A. H., **J. R. Moisan**, and M. A. Linkswiler. Algorithm development for predicting biodiversity based on phytoplankton absorption, *Cont. Shelf Res.*, **55**, 17-28.

(f) 2011: **Moisan, J. R.,** T. A. Moisan, and M. A. Linkswiler. An Inverse Modeling Approach to Estimating Phytoplankton Pigment Concentrations from Phytoplankton Absorption Spectra, *J. Geophys. Res.*, **116**, C09018, doi: 10.1029/ 2010JC006786.

GP Project Directions



Computationally challenging

Ongoing efforts at specific locations

Gathering *in situ* and satellite observations
Simple application

Requires *in situ* and satellite obs.
Simple application.
Uncertainty estimates can improve algorithms.

Very abstract. Can make use of genomic obs. Merges two lines of ecosystem modeling. Aligns itself to broader questions regarding existence of "life" in terms of ecosystems.