Multivariate autoregressive modeling of multi-species time- series data

Steven Viscido
Eli Holmes

National Oceanic and Atmospheric Administration
Northwest Fisheries Science Center
Seattle, WA, USA

Analysis of multi-species data

- Cliff notes intro to MAR-1
- A tiny bit of philosophy of science
- One example of its use: Understanding Lake Wash plankton dynamics
- A little more theory
- Does it work? Robustness studies using simulations

Photo courtesy of Phillippe Bush, REEF

Technical Workshop: An introduction to the analysis of community time series using Multivariate Autoregressive (MAR) models

Online version can be found here: faculty.washington.edu/eeholmes

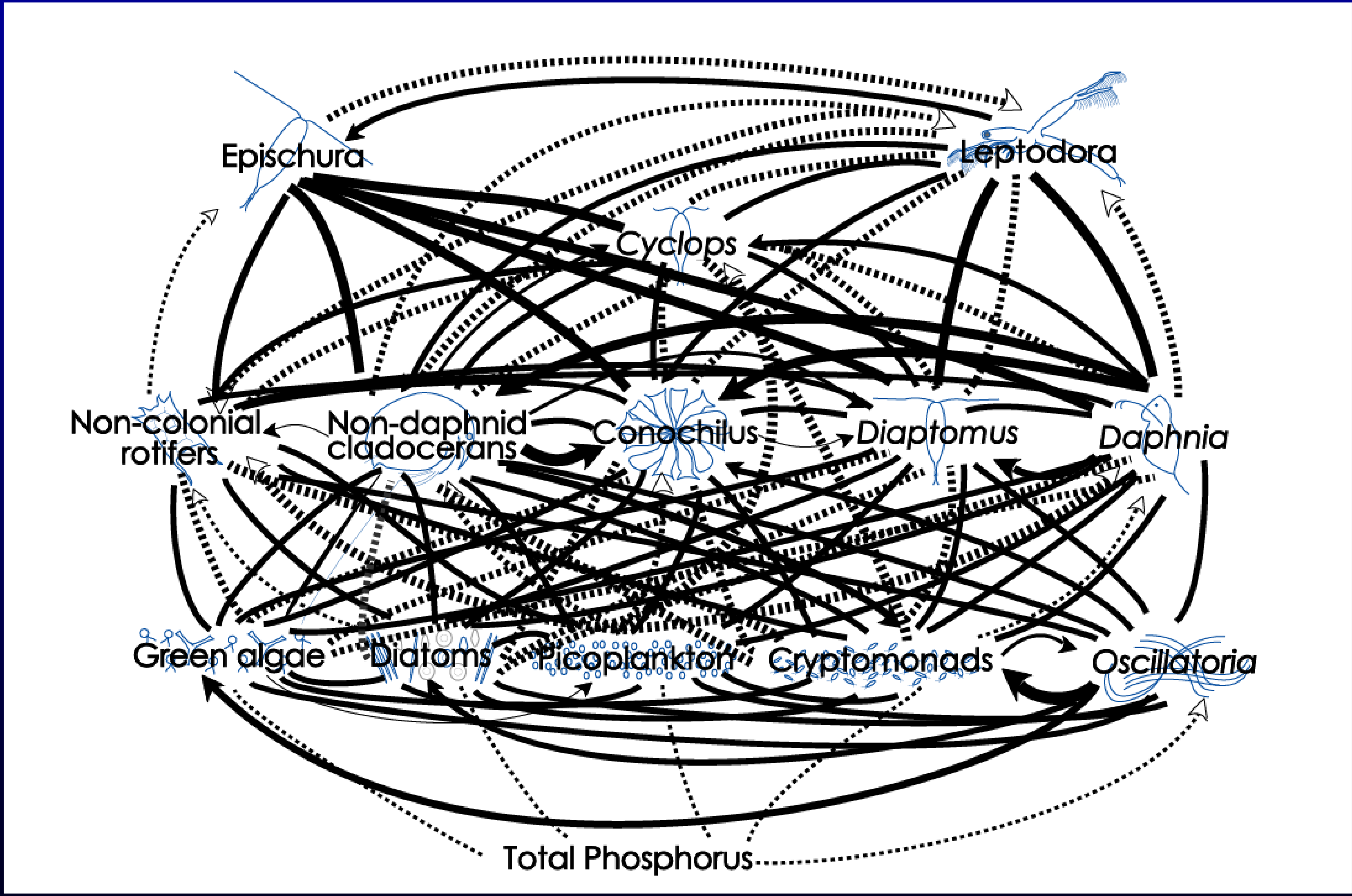
Eli Holmes, National Marine Fisheries Service, eli.holmes@noaa.gov

Stephanie Hampton, National Center for Ecological Synthesis and Analysis, hampton@nceas.ucsb.edu

Mark Scheuerell, National Marine Fisheries Service, mark.scheuerell@noaa.gov

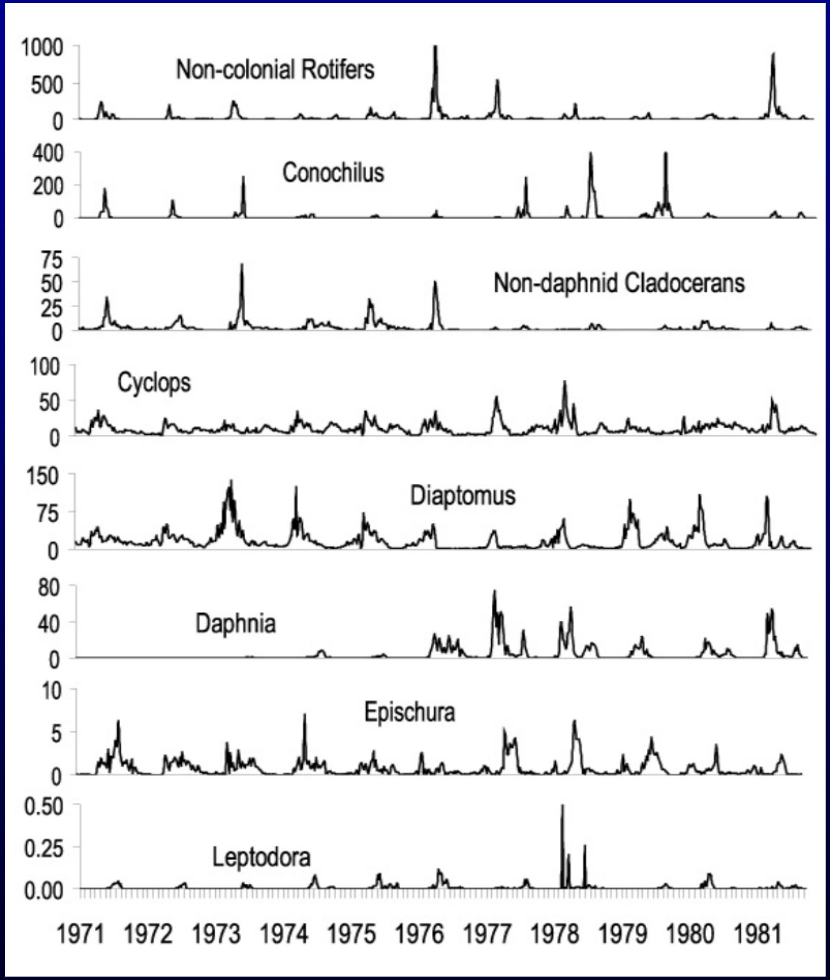
Steven Viscido, Winston Salem State University viscido@wssu.edu

Communities are made up of species that are eating, being eaten, competing, and facilitating—a big mess of interactions



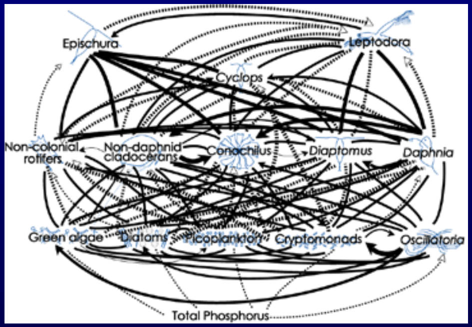
S. E. Hampton, NCEAS, UCSB

We have time-series data on the system and we want to make inferences about the dynamics of the system



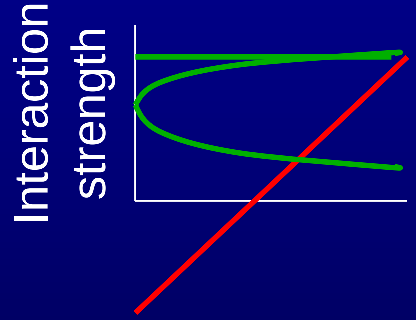
- How does the stability of the community compare between 1) time periods, 2) experimental treatments, 3) regions, 4) under different regimes
- What are the strong interactions in the system and what are their directions
- Forecast forward using the past (past time series data) and make statistical statements about the probability of different community states

Put some constraints on the interactions (to ensure a single stochastic equilibrium) and add a good dose of year to year variability

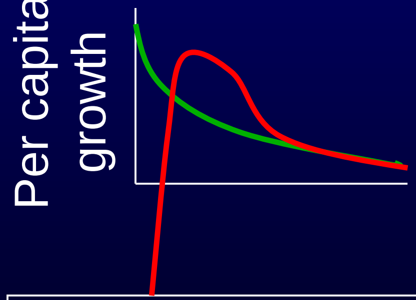


+

No flip-flopping on interaction directions



Density of spp A



Self-regulation but no strong allee effects

+

Variability in year-to-year population growth rates and/or interaction strengths

The stochastic behavior of this can be approximated by a MAR-1—a stochastic discrete Ornstein-Uhlenbeck process which is a type of stochastic process fluctuating about a mean

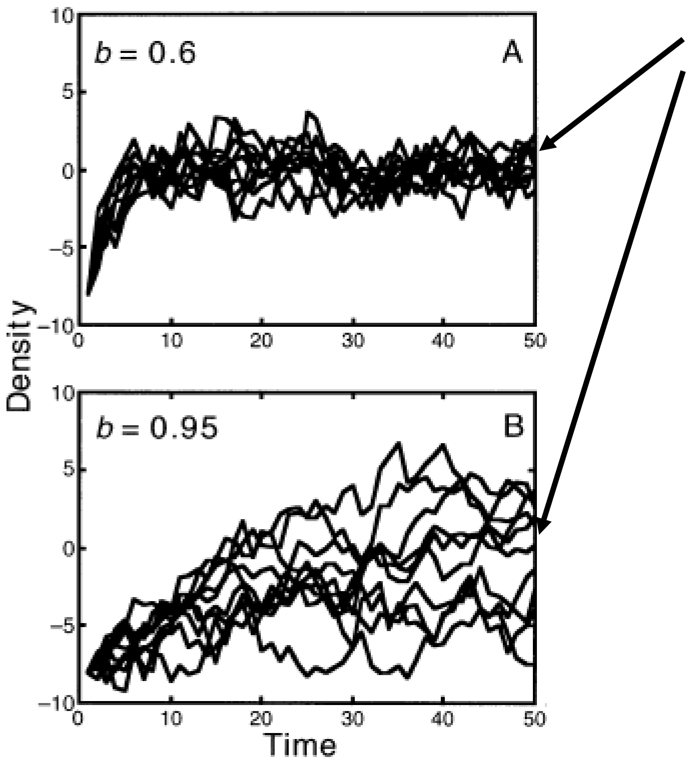


Fig. 1 - Ives et al. (2003)

These statistical distributions are the same as those produced by a multivariate autoregressive process with normal errors

$$X_{i,t} = a_i + \sum_{j=1}^{p} b_{ij} X_{j,t-1} + E_{i,t}$$

$$E_{i,t} \sim N\left(0, \sigma_i^2\right)$$

$X_{j,t}$ is log abundance of species j at time t

b_{ij} is *effect* of species j on species i

Multispecies Autoregressive Models (MARs)

Ives, Dennis, Cottingham, & Carpenter. 2003. Ecol. Monogr. 73(2)

Log abundance of species *i* tomorrow

$$x_i(t+1) = a_i + \Sigma\, b_{i,j}\, x_j(t) + \Sigma\, c_{i,k}\, u_{k(t)}$$

Species-specific constant, like an exponential growth rate

Multispecies Autoregressive Models (MAR-1)

Log abundance of species *i* tomorrow

Effect of species *j* on species *i*
Interactions and density-dependence term

$$x_i(t+1) = a_i + \Sigma\, b_{i,j}\, x_j(t) + \Sigma\, c_{i,k}\, u_{k(t)}$$

Species-specific constant

Log abundance of species *j* today

Multispecies Autoregressive Models (MARs)

Log abundance of species *i* tomorrow

Effect of species *j* on species *i*

Effect of environmental variable *k* on species *i*

$$x_i(t+1) \quad = \quad a_i \;+\; \Sigma\, b_{i,j}\, x_j(t) \;+\; \Sigma\, c_{i,k}\, u_{k(t)}$$

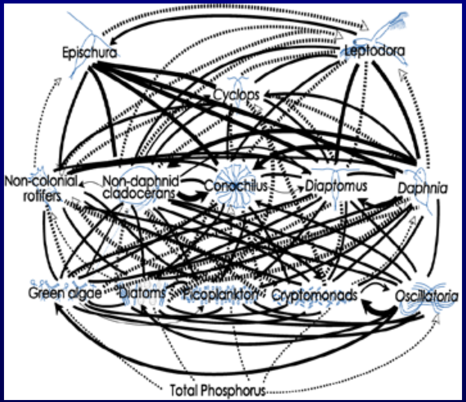
Species-specific constant

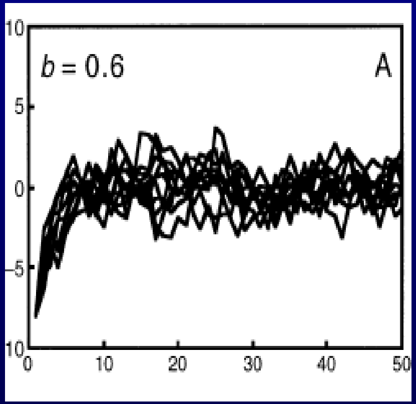
Log abundance of species *j* today

Level of environmental variable *k* today

A reductionist versus a non-reductionist approach for forecasting

FORECAST





FORECAST

A reductionist versus a non-reductionist approach for forecasting

- Forecasting in the social sciences: Expert Political Judgment: How Good is It? How Can We Know? By Philip Tetlock
  - Catchy title for a 20 year study comparing performance of decision-making: statistical (computer), cautious (human), and bold (human).
  - Chapter 2: “The ego-deflating challenge of Radical Skepticism” A review of non-reductionist approaches to forecasting in the social sciences

An example where MAR-1 models used to understand plankton response to the Lake WA clean-up (Hampton, Scheuerell & Schindler 2006)


Lake Washington Waters To Be 'Dirty' Until 1970'

Algae Peril To Lake Washington

BY E. P. CHALCRAFT

Lake Waters In Bad Shape, U.W. Scientists Report

Edmondson Announces Pollution May Ruin Lal

Lk. Washington Brown—That's Algae, Not Mud

And It'll Be There For The Next

LAKE STINKO

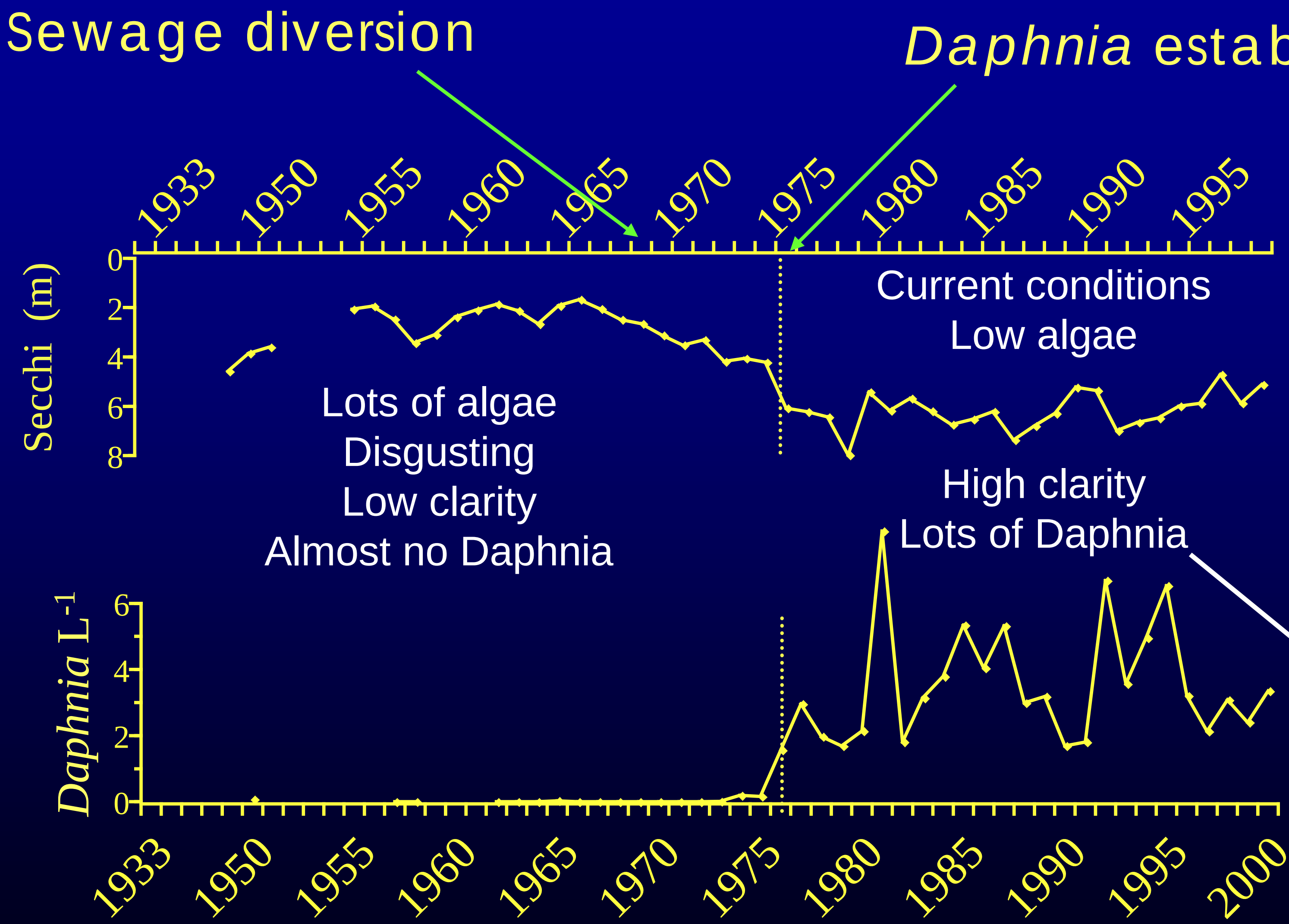
It's Old Algae That's To Blame

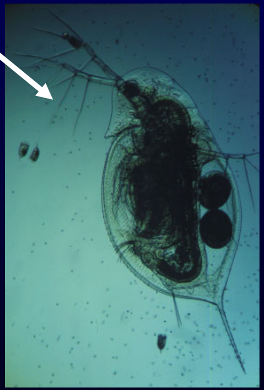
Seattle Post-Intelligencer

Washington Cleanup Improves Wa

S. E. Hampton, NCEAS, UCSB

After sewage was diverted, Daphnia (a voracious zooplankton spp) increased greatly. Why?

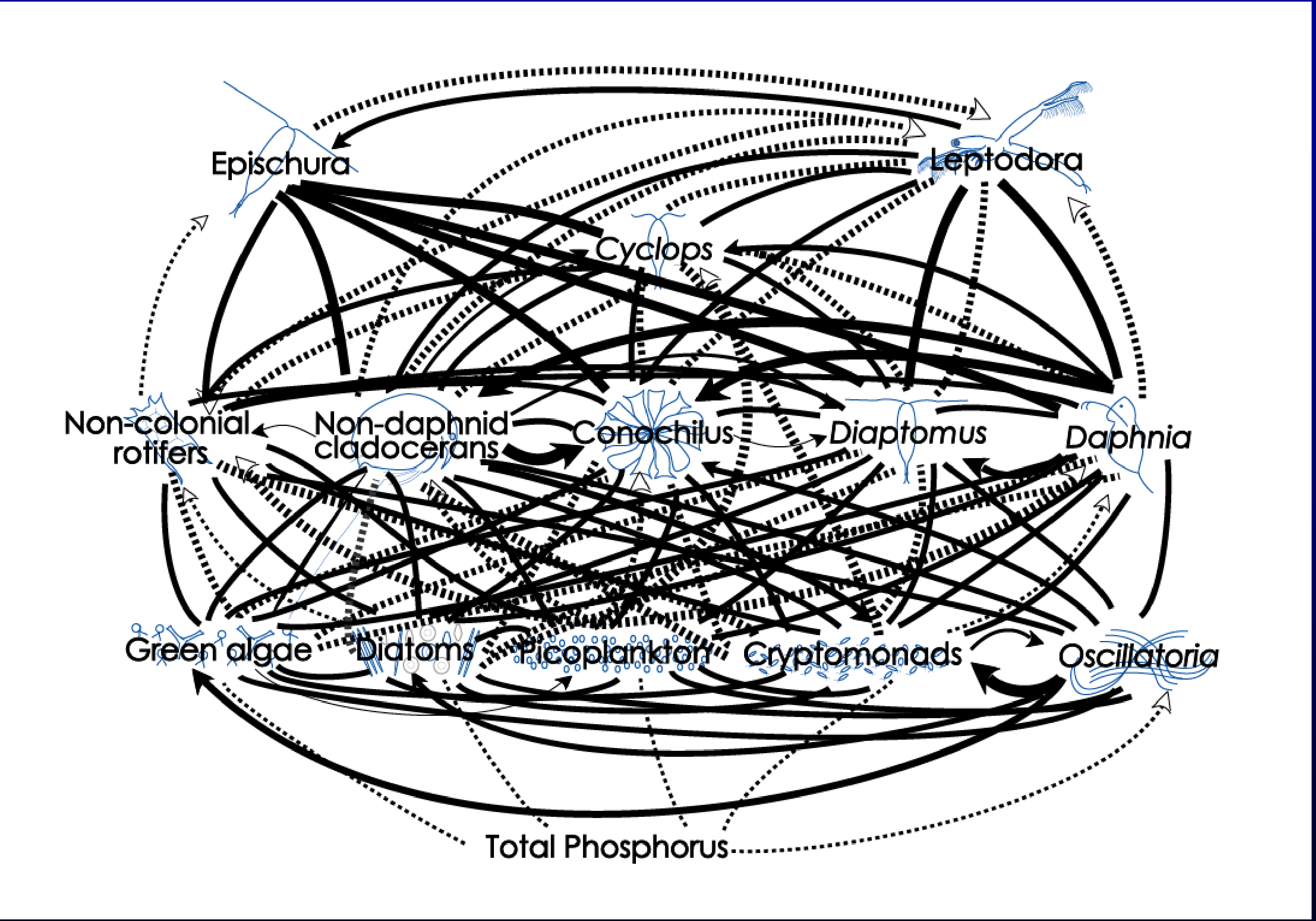




Not to scale

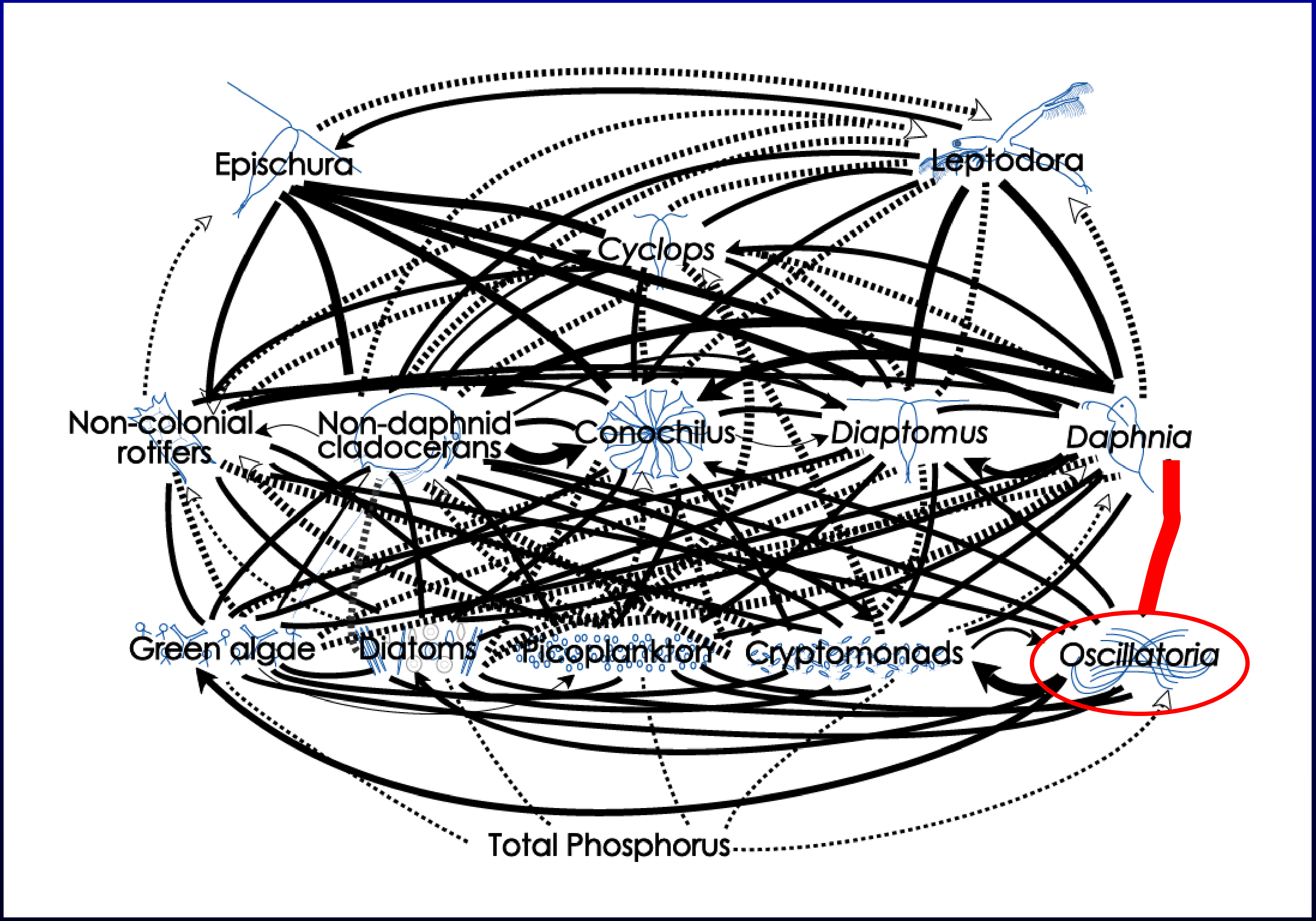
S. E. Hampton, NCEAS, UCSB

MAR food web construction for Lake Washington



Hampton, Scheuerell & Schindler 2006

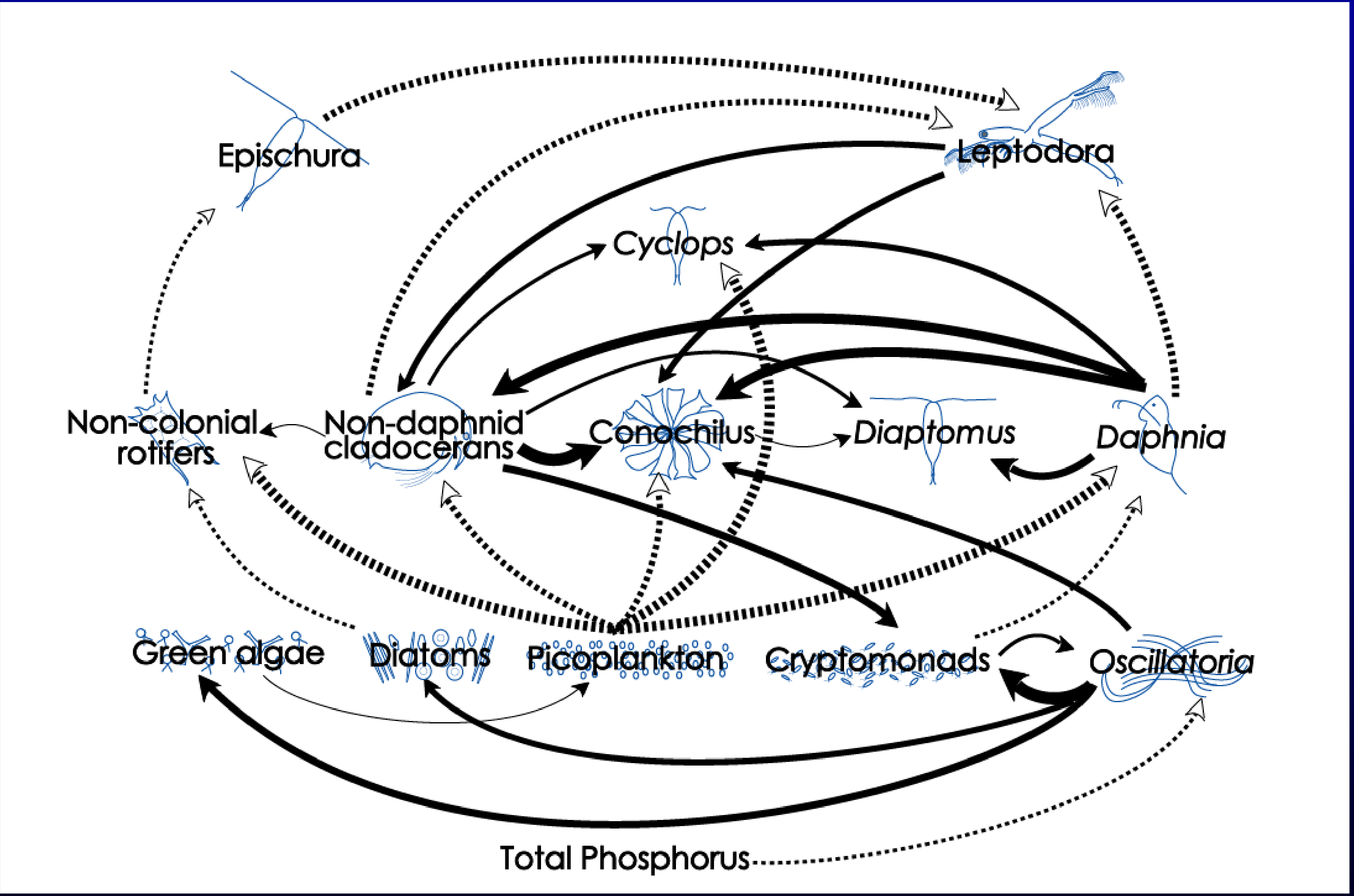
MAR food web construction for Lake Washington



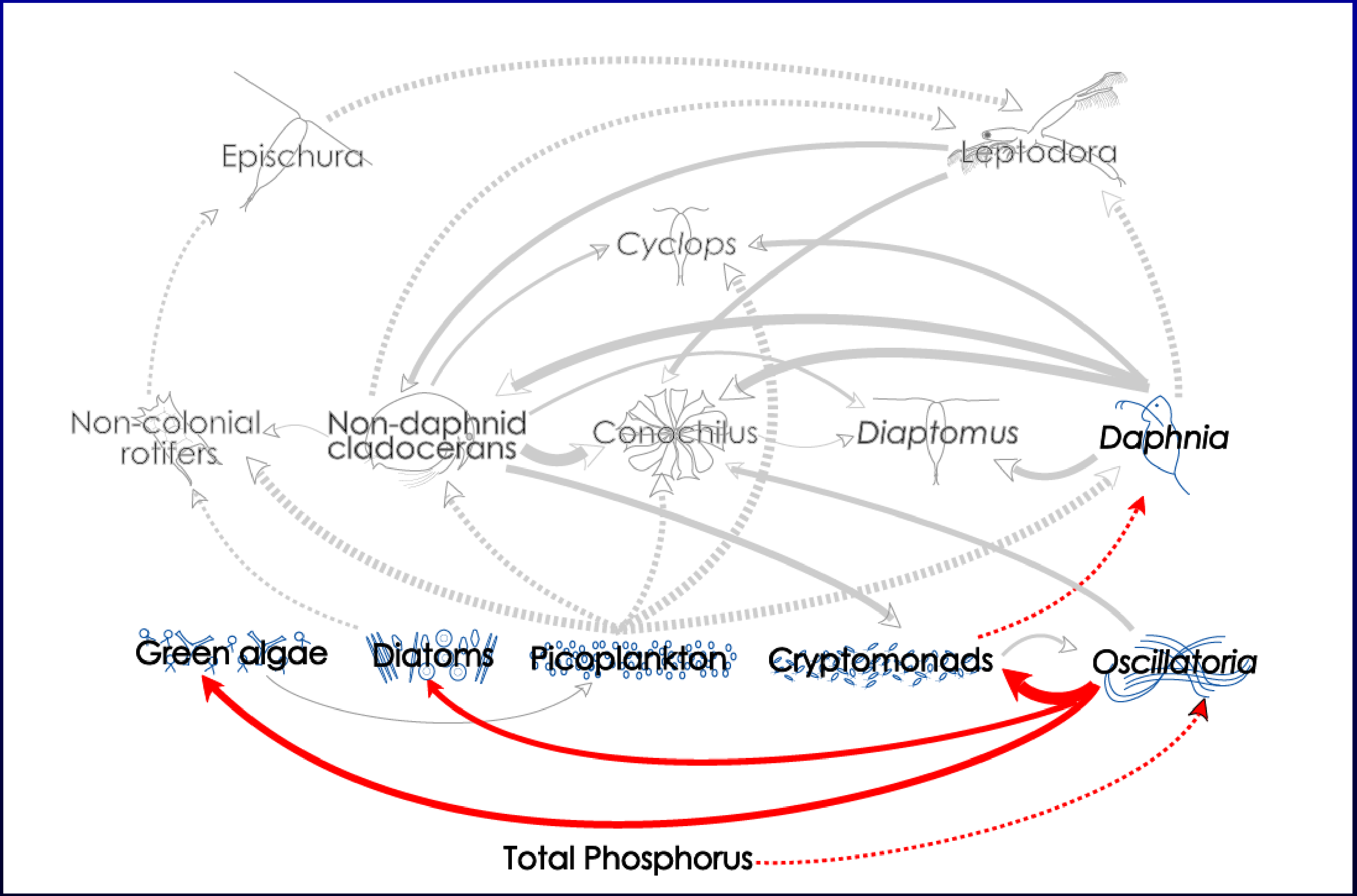
S. E. Hampton, NCEAS, UCSB

Hampton, Scheuerell & Schindler 2006

Strong interactions via MAR-1 analysis for Lake Washington



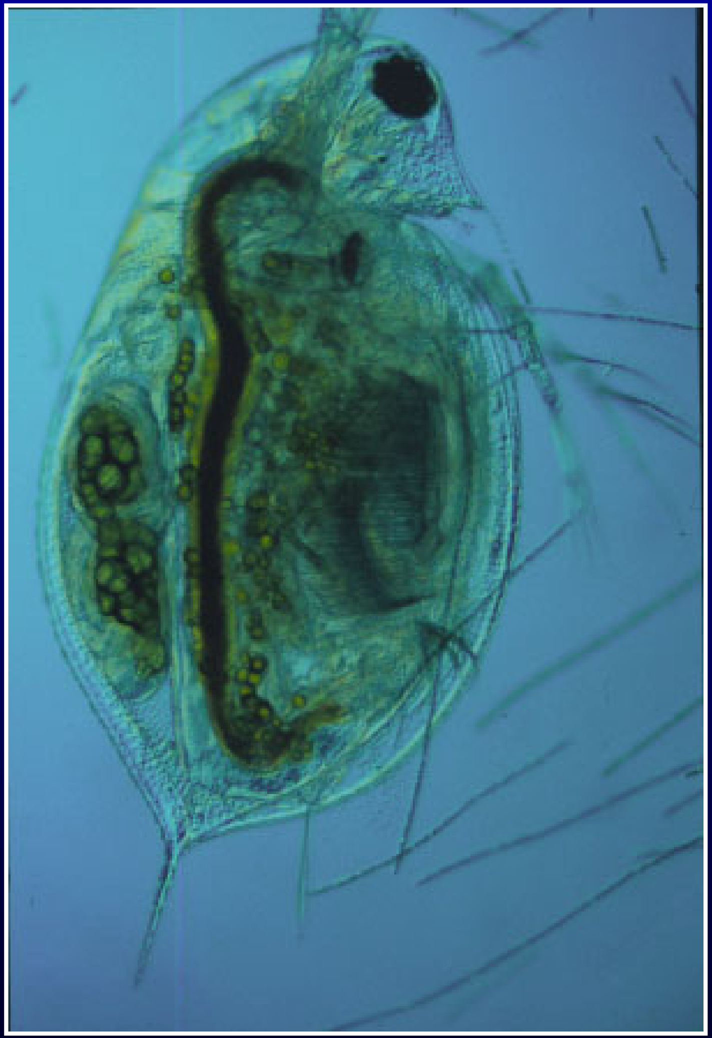
MAR food web construction for Lake Washington



Hampton, Scheuerell & Schindler 2006

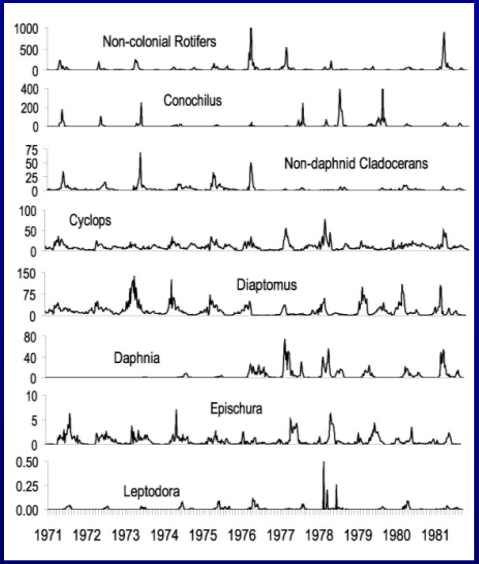
MAR food web construction for Lake Washington

- Aspects of historical conceptual model supported
  - Inhibitory role of *Oscillatoria*
  - Intense competitive effects of *Daphnia*
- "New" relationships
  - Cryptomonad importance
  - Role for picoplankton



Hampton, Scheuerell & Schindler 2006

A little more theory—this time about the stability of multivariate autoregressive processes



One spp in the MAR-1 model:

$$X_{i,t} \;=\; a_i \;+\; \Sigma\, b\, X_{j,t-1} \;+ \Sigma\, c\;\; U_{t\text{-}1} \;+\; E_t$$

MAR-1 model; all spp written in matrix form:

$$\begin{vmatrix} X_1 \\ X_2 \\ X_3 \end{vmatrix}_t = \begin{vmatrix} a_1 \\ a_2 \\ a_3 \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \begin{vmatrix} X_1 \\ X_2 \\ X_3 \end{vmatrix}_{t\text{-}1} + \begin{vmatrix} c_{11} \\ c_{21} \\ c_{31} \end{vmatrix} U_{t\text{-}1} + \begin{vmatrix} E_1 \\ E_2 \\ E_3 \end{vmatrix}_t$$

MVN

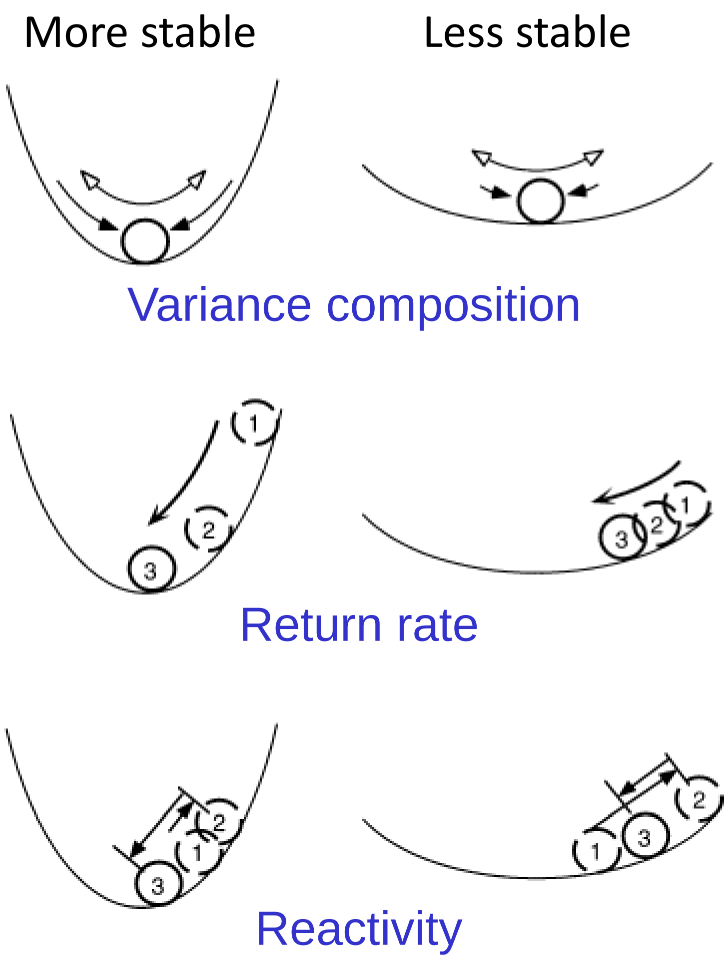
A B C

The interaction matrix gives information about the stability of the community

Intra-specific effects

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & \cdots & \vdots \\ b_{31} & \vdots & b_{33} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & \cdots & \cdots & \cdots & b_{pp} \end{bmatrix}$$

Inter-specific effects



Ives et al used MAR-1 to look at the effects of phosphorous manipulation on stablility of lake plankton communities

Ives, Dennis, Cottingham, & Carpenter. 2003. Ecol. Monogr. 73(2)

How do robustly can we estimate community interaction strengths and stability given that we usually have crappy data?

Estimating interactions from data

Given a time-series of multi-species data…

$\mathbf{X}_1$

$\mathbf{X}_2$

$\mathbf{X}_3$

$\mathbf{X}_4$

…

$\mathbf{X}_T$

…from that we estimate what the parameters must have been

We can use ML, CLS regression, Bayesian estimation to get parameter estimates

A, **B**, C

Test the limits of MAR-1 estimation

- Different amounts of process error variance
- Several covariates, some detected and some not
- Time series duration
- Pooled variates

General simulation rules

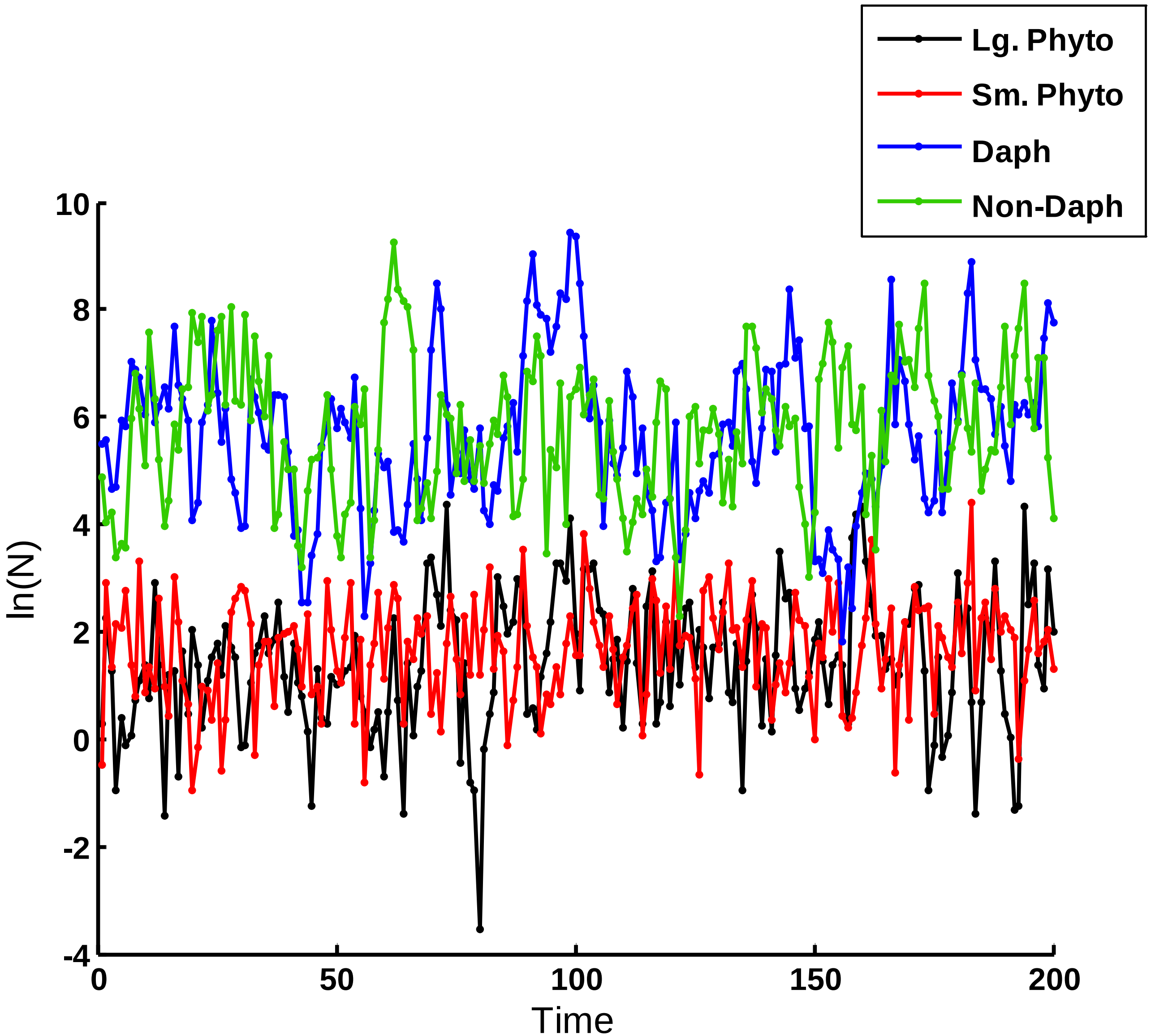
- 100 replicate time series for each experiment
- 100 time steps with 100 yr burn-in
- Used the **A**, **B**, and **C** matrices from Ives et al. 2003, “Low Planktivory” lake (except when pooling variates)

Low planktivory interaction matrix (only strong interactions)

	Lg. Phyto.	Sm. Phyto.	*Daphnia*	non-*Daphnia*
Lg. Phyto.	0.50	-0.36	0	0
Sm. Phyto.	0	0.07	-0.02	-0.10
Daphnia	0	0	0.76	0
non-*Daphnia*	0.10	0.10	0	0.56

$$\text{Return rate} = \max \lambda_{\mathbf{B}} = \mathbf{0.78}$$

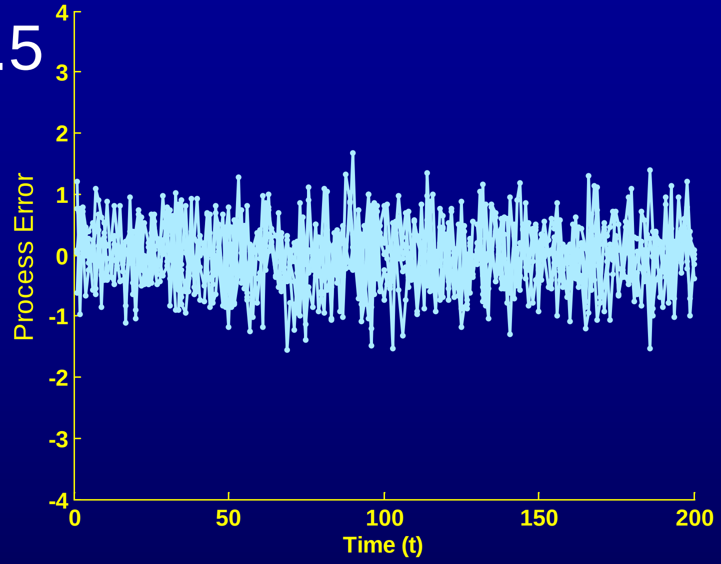
**We also did this study with S. Hampden's 13 spp plankton community



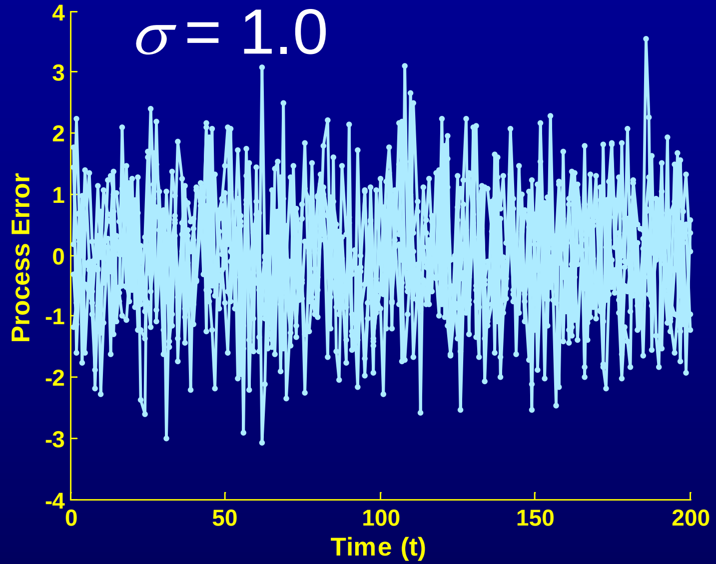
Lg. Phyto
Sm. Phyto
Daph
Non-Daph
10
8
6
4
2
0
-2
-4
ln(N)
0
50
100
150
200
Time

Test 1: Can we break the estimation by cranking up the amount of process Error

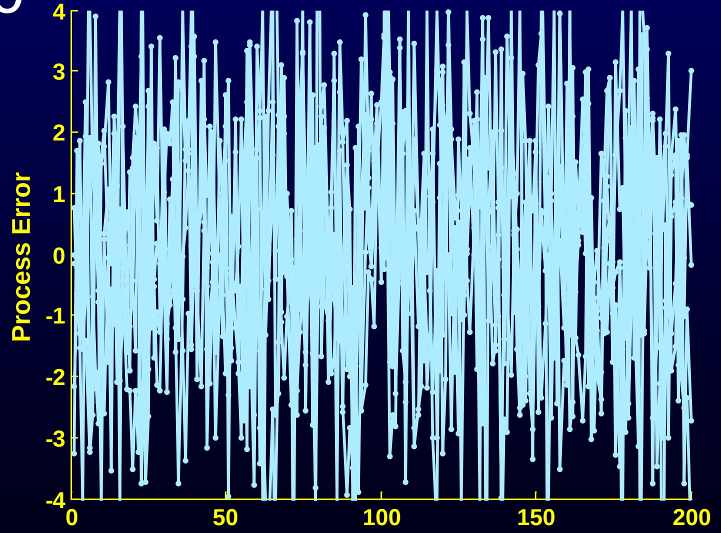
$\sigma = 0.5$



$\sigma = 1.0$

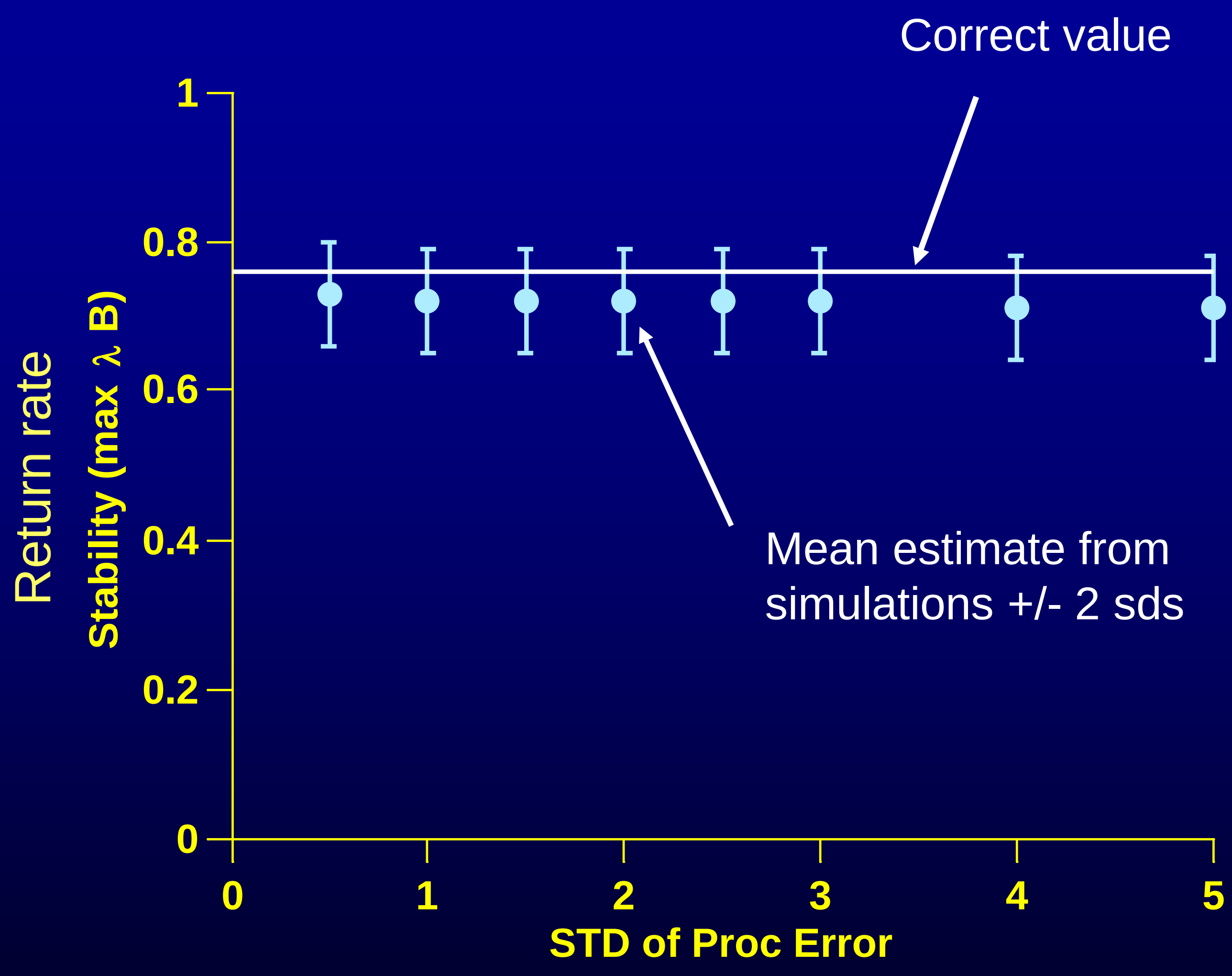


$\sigma = 2.0$

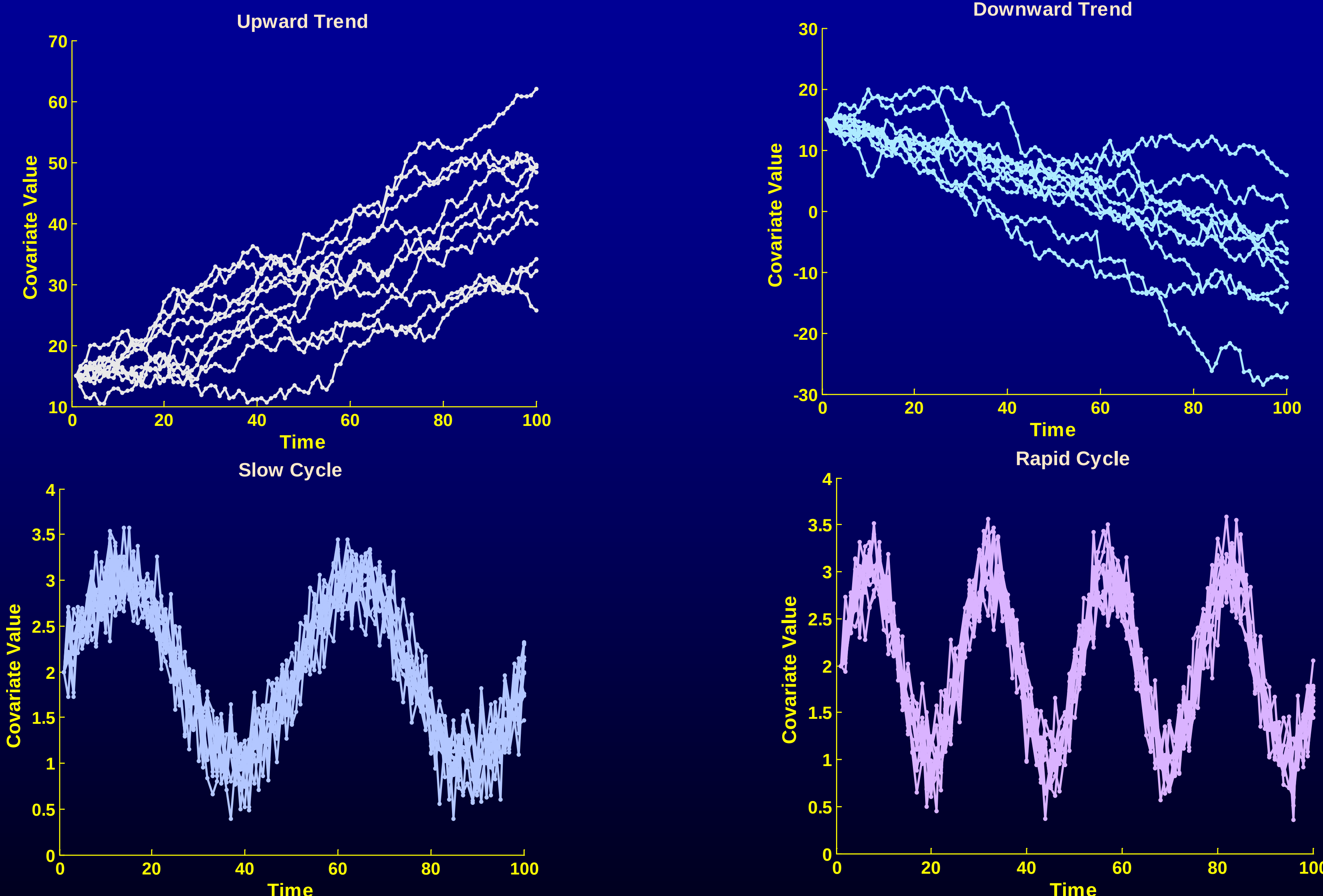


$\mu = 0$

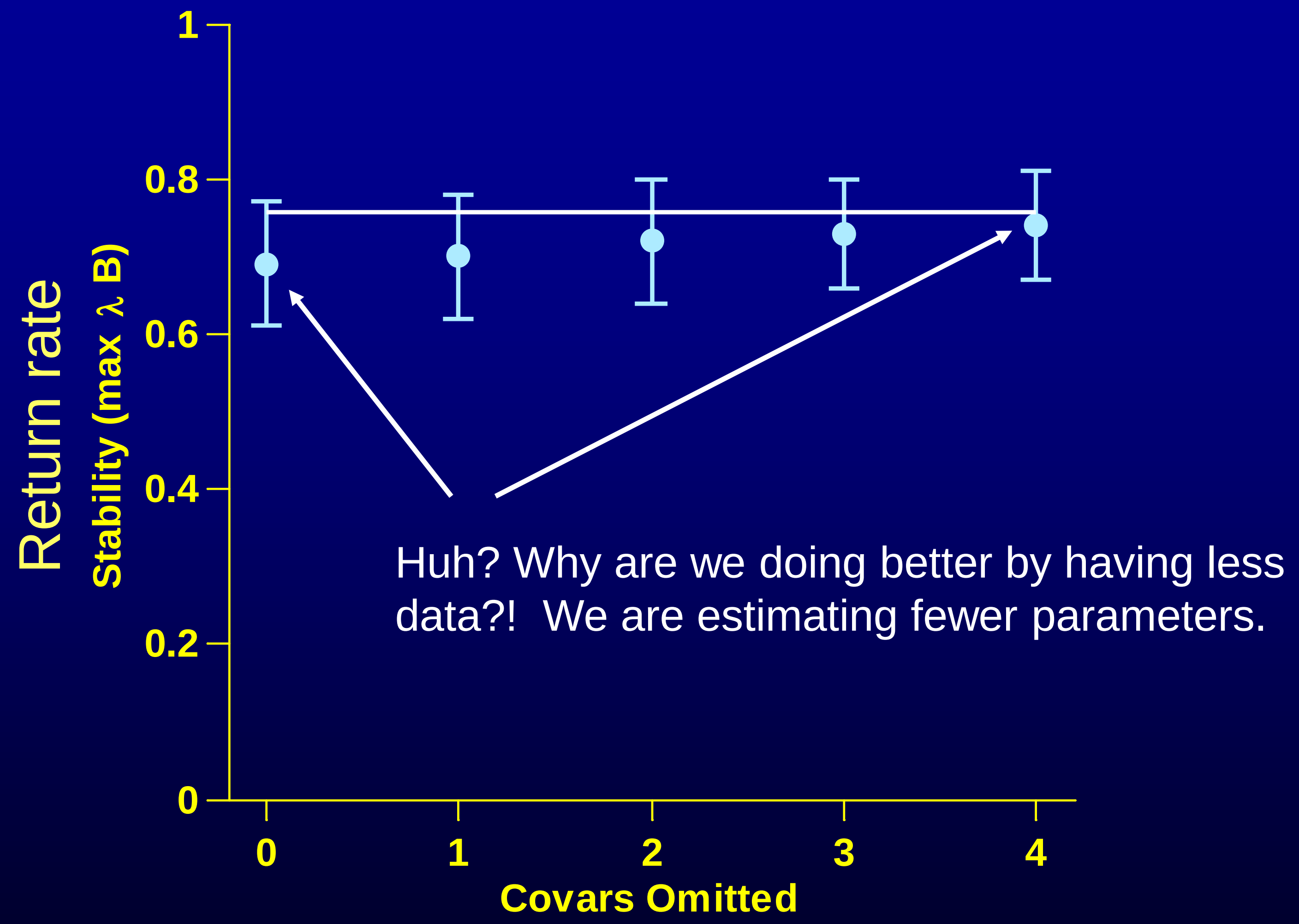
$\sigma = \{ 0.5, 1, 1.5, 2, 2.5, 3, 4, 5 \}$


Correct value
Mean estimate from
simulations +/- 2 sds
Return rate
Stability (max λ B)
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
STD of Proc Error

Test 2: What happens when we have strong environmental drivers…and leave them out?



Bottom forcing in these examples

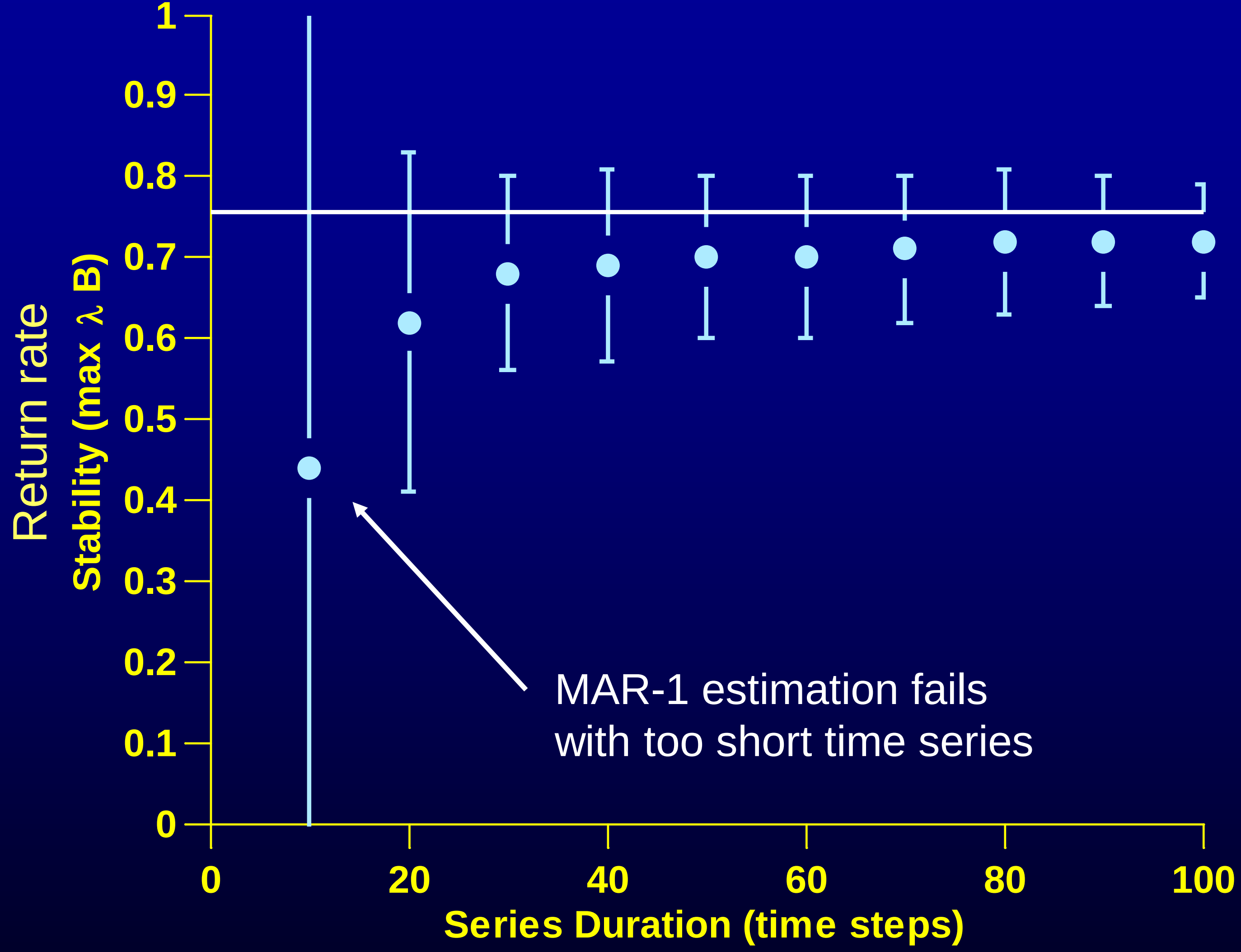


Return rate
Stability (max λ B)
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
Covars Omitted
Huh? Why are we doing better by having less data?! We are estimating fewer parameters.

Test 3: What if we have short times series?

- Start with a 100-time step dataset
- Cut off 10 steps, run CLS estimates
- Repeat until we get to a 10-step series
- Compare the results

Test 3: What if we have short times series?



Test 4: What happens to our estimates if we lump species together into functional groups?

From Hampton and Schindler (2006)

	Crypto	Diatoms	Green Alg	*Oscill*	Unicells	Other	Non-Col R	*Conoch*	Non-*Daph*	*Cyclops*	*Diaptomus*	*Daphnia*	*Epischura*	*Leptodora*
Cryptomonads	0.5	-0.1	0	0	0	0	0	0	0	0	0	0	0	0
Diatoms	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0
Green Algae	0	0	0.2	0	0	0	0	0	0	0	0	0	0	0
Oscillatoria	0	0	0	0.7	0	0	0	0	0	0	0	0	0	0
Unicells	0	0	0	0	0.4	0	0	0	0	0	-0.1	-0.1	0	0
Other Algae	0	0	0	0	0	0.4	0	0	0	0	-0.1	0	0	0
Non-colonial Rotifers	0	0	0	0	0.1	0.1	0.5	0	-0.1	0	0	0	0	0
Conochilus	0	0	0	0	0	0	0	0.5	0	0	0	0	0	-0.1
Non-*Daphnia* Cladocerans	0	0.1	0	0	0	0.1	0	0	0.5	0	0	0	0	0
Cyclops	0	0	0	0	0.2	0	0	0	-0.1	0.4	-0.1	-0.1	0	0
Diaptomus	0	0	0	0	0	0	0	0	0	-0.1	0.5	-0.1	-0.1	0
Daphnia	0.1	0	0	0	0.1	0	0	0	0	0	0	0.6	0	0
Epischura	0	0	0	0	0	0	0	0	0	0	0	0	0.4	-0.1
Leptodora	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0.1

Estimated **B** values (Pooled)

	Daphnia	Sm. Phyto.	Lg. Phyto.	non-*Daphnia*
Daphnia	0.58	0.08	0	-0.01
Sm. Phyto.	-0.05	0.38	-0.02	-0.01
Lg. Phyto.	0	0	0.40	-0.01
non-*Daphnia*	-0.06	0.05	0.02	0.34

Estimated Stability

	Daphnia	Sm. Phyto.	Lg. Phyto.	non-*Daphnia*
Daphnia				
Sm. Phyto.				
Lg. Phyto.				
non-*Daphnia*				

Stability = $\max \lambda_{\mathbf{B}} = 0.80 \pm 0.05$

(Known = 0.84)

Conclusions

- MAR-1 models were able to correctly determine the interactions and stability of systems in the face of
  - Large amounts of process error / random noise
  - Incomplete covariate data
  - Pooling of variates
  - Short time series (unless it's *very* short)
- Estimation of MAR-1 models were very robust in our simulation studies

Current and Future Directions

- Why I didn't talk about the groundfish data… How to use spatial replication to improve parameter estimates.
  - Need to first derive the algorithm to use and work out good estimation methods. Doable but hard

- What happens when we have observation error? Stable and efficient estimation for multivariate state-space models.
  - Solving the spatial replication problem will help this here.

- How do we deal with foodwebs with species that are operating on really different time scales (elephants and bacteria)?
  - How do we interpret MAR-1 fits in that case. How to ask this question is an open theoretical problem.

NWFSC research group working on stochastic AR-1 modeling of population data:

- Steven Viscido
- Eric Ward
- Brice Semmens

Collaborators:

- Stephanie Hampton
- Mark Scheuerell
- Anthony Ives
- Brian Dennis
- John Sabo
- Bill Fagan

Resources

- An online workshop on MAR-1 methods with computer labs:
  faculty.washington.edu/eeholmes
  Click on ‘workshops’ in the navbar

- LAMBDA a GUI-based toolbox for doing a full MAR-1 analysis. Hosted at FishBox

- 1-day workshop on state-space modeling for AR-1 processes at Ecological Society meetings this August

- Code for MAR-1, state-space models, kalman filters is online in a variety of programming languages: Fishbox.iugo-café.org