Data-Driven Models of Plankton Ecosystems:

Making a Case For Noise

Craig Jackson, Department of Mathematics and Computer Science, Ohio Wesleyan University

NSF

Ohio Wesleyan University

ZOOL 345: Marine Biology

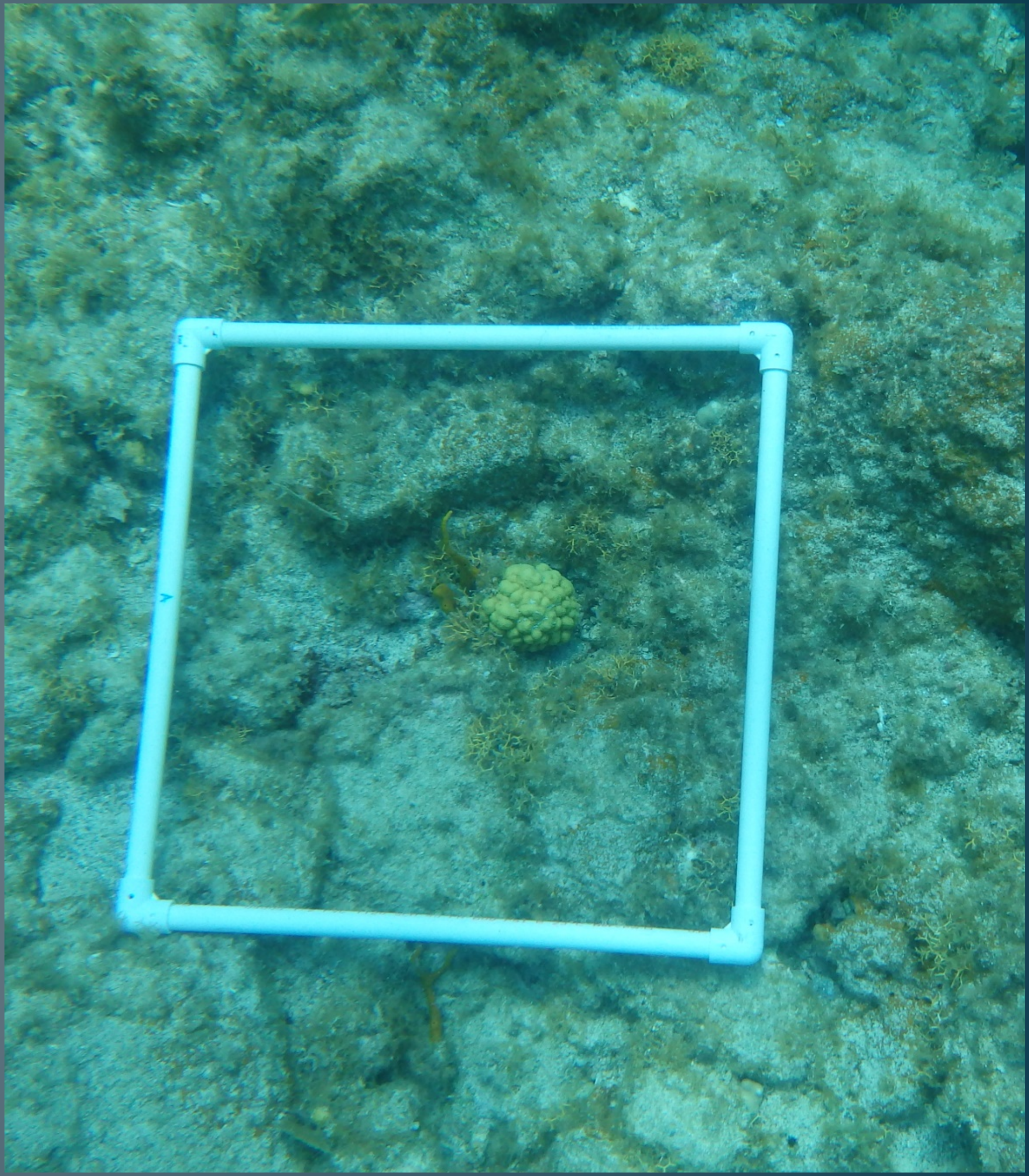
Co-taught by Jackson and Dr. Amy Downing (Prof. of Zoology)

Incorporates significant math modeling + field work in St John USVI

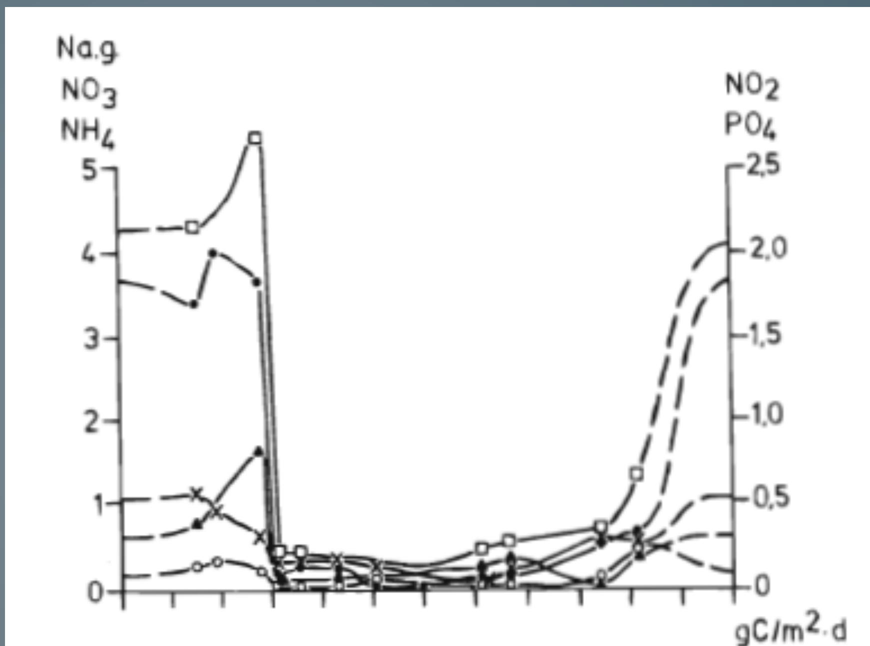
ZOOL 345: Marine Biology

Co-taught by Jackson and Dr. Amy Downing (Prof. of Zoology)

Incorporates significant math modeling + field work in St John USVI



Nutrient



Plankton

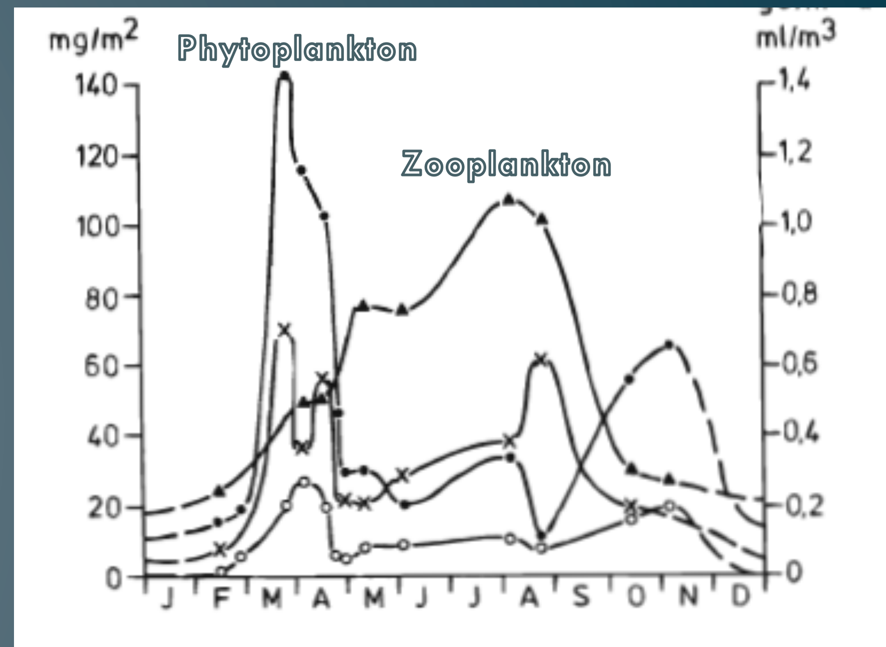


Figure 8.2: Annual cycle of nutrients (top) in the upper 10 meters and plankton (bottom) in the upper 25 meters measured at a central station, (Stat. No. 113), in the Arkona basin of the Baltic Sea. Upper panel: squares - total disolved inorganic nitrogen, dots - NO_3, triangles - NH_4, and crosses - PO_4, all in $\mu mol/m^3$. Lower panel: dots - chlorophyll A, in mg/m^2, circles - phaeopigment in mg/m^2, crosses - primary production in $gC/m^2/d$, triangles - zooplankton biomass in ml/m^3. [Image and caption taken from Fennel and Neumann (2004).]

Nutrient, Phytoplankton, Zooplankton, Detritus with Irradiation Model

$$\frac{dN}{dt} = -h(I)f(N)P + l_{PN}P + l_{DN}D + l_{ZN}Z$$

$$\frac{dP}{dt} = h(I)f(N)P - l_{PN}P - l_{PD}P - g(P)Z$$

$$\frac{dD}{dt} = l_{PD}P - l_{DN}D + l_{ZD}Z$$

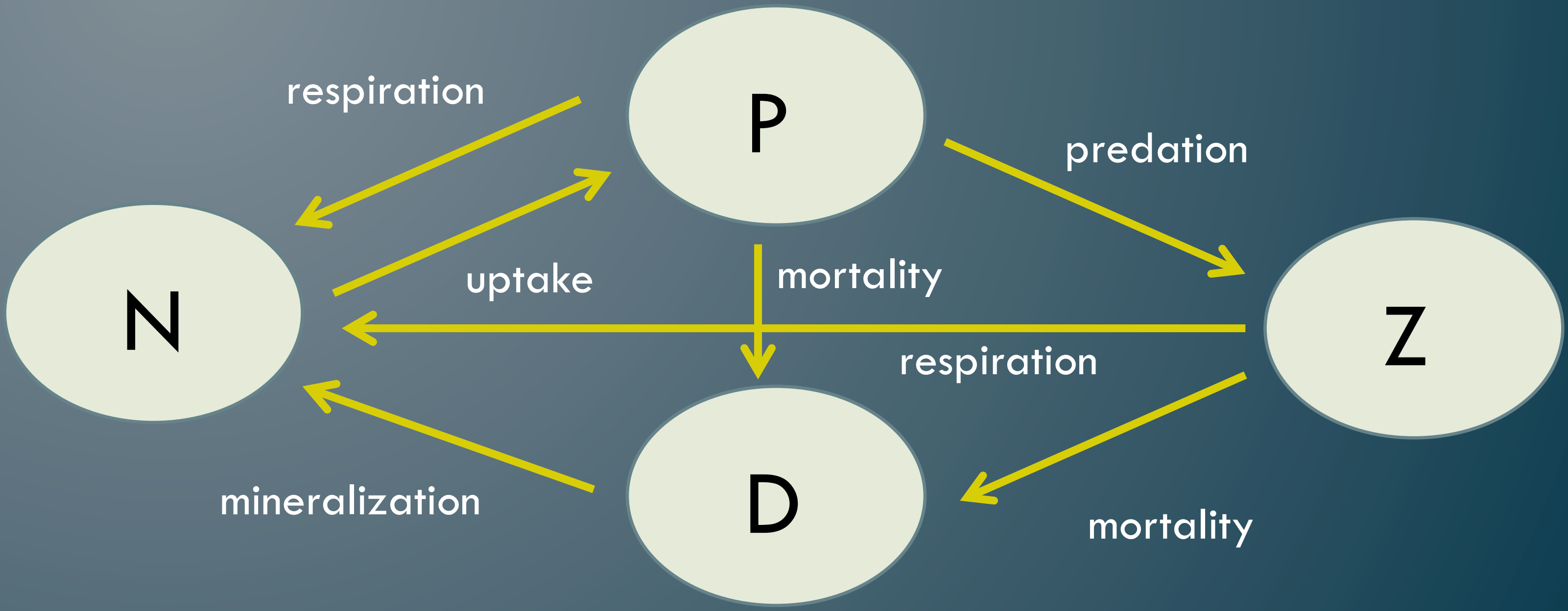
$$\frac{dZ}{dt} = g(P)Z - l_{ZN}Z - l_{ZD}Z$$

$$f(N) = \frac{r_{\max}N}{k_N + N}$$

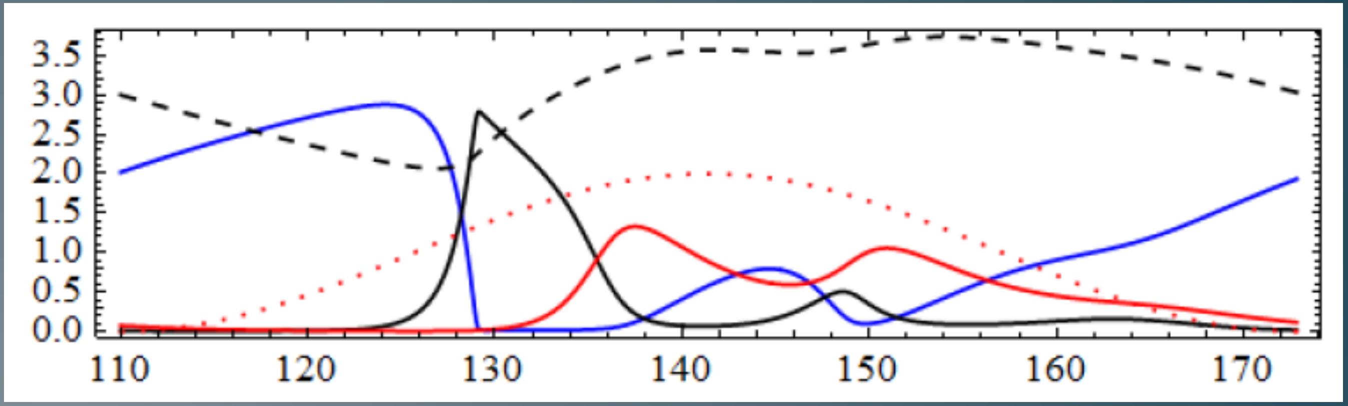
$$I(t) = \bar{I} - A\cos(bt)$$

$$h(I) = \frac{I}{I_0 + I}$$

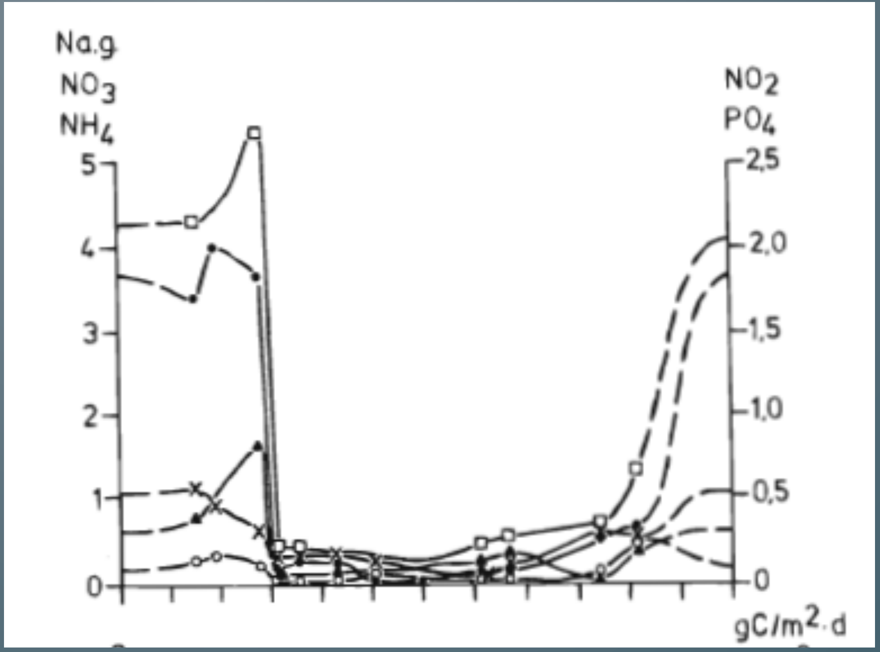
$$g(P) = \frac{g_{\max}P}{k_Z + P}$$



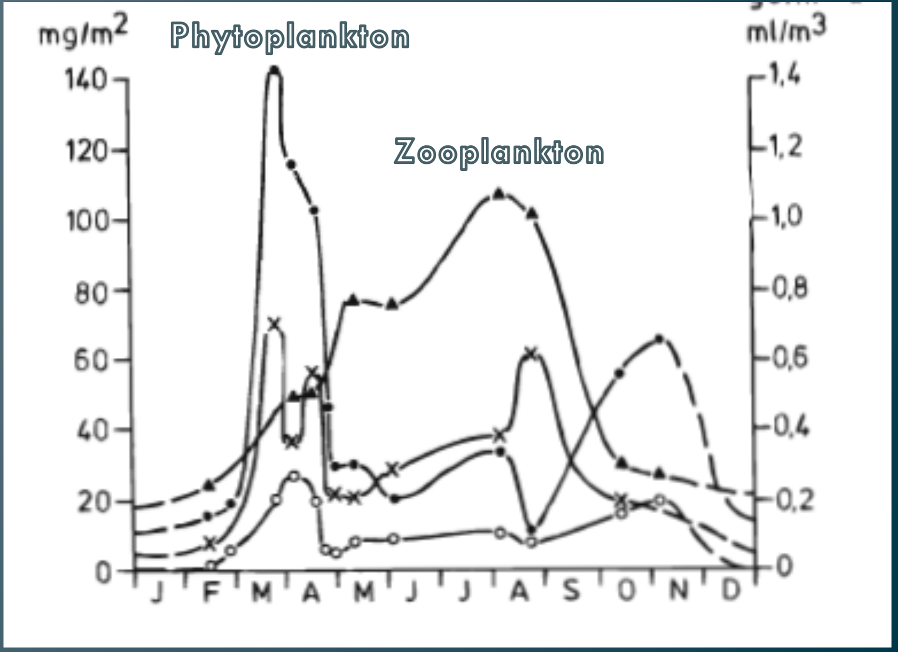
Nutrient, Phytoplankton, Zooplankton, Detritus with Irradiation Model



Nutrient



Plankton

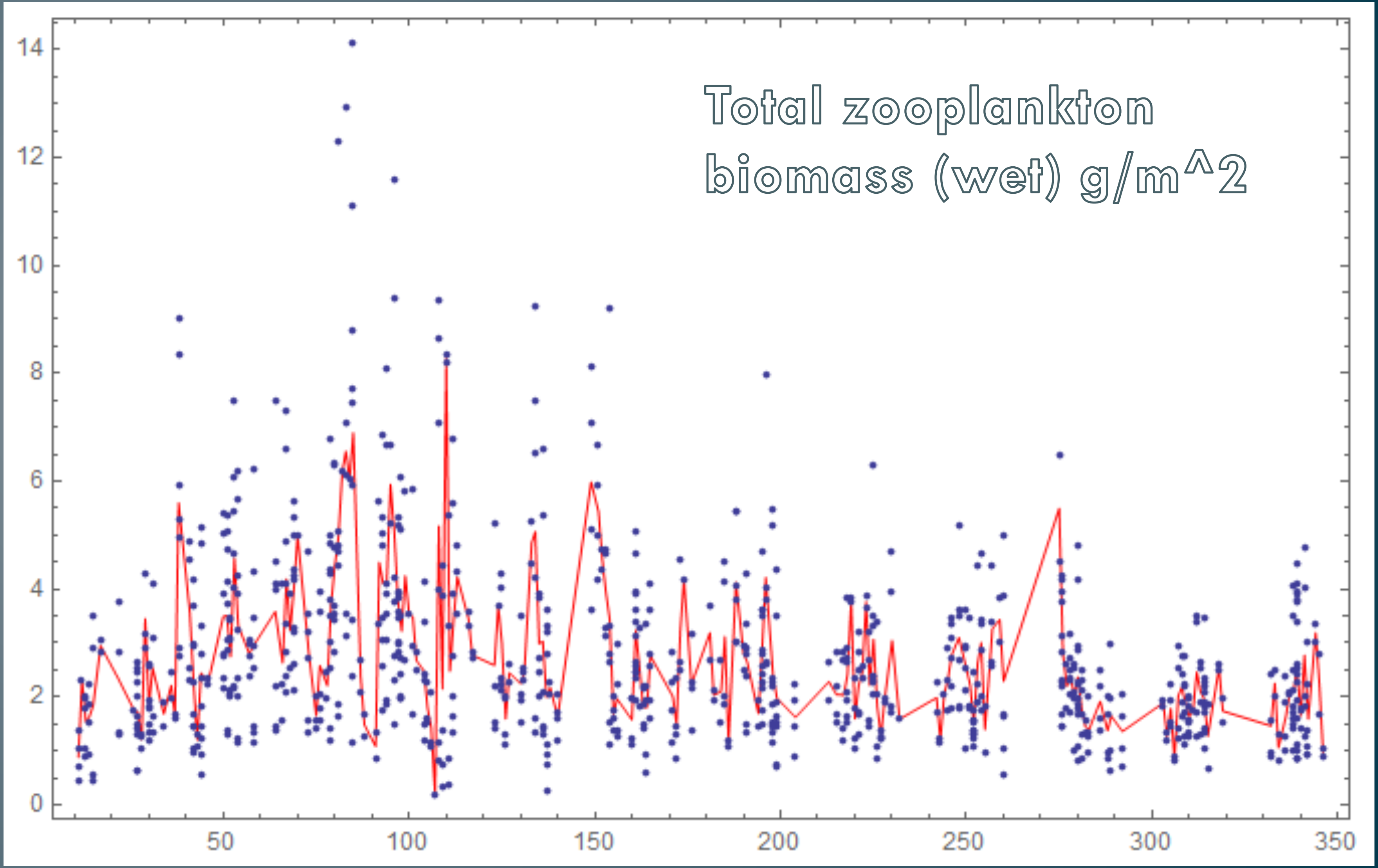


Incorporating real data?

```
#-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*
#
#  "BATS (JGOFS-BATS):  Zooplankton biomass data sampled during the Bermuda Atlantic Time-series Study (BATS) project"
#
#-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*
#
#          NMFS-COPEPOD:  14-Jun-2013  (Web v10.1)
#
#             Live-link:  http://www.st.nmfs.noaa.gov/plankton/data/batsjgofs/index.html
#
#                         A comprehensive content summary and alternate data formats
#                                are available online at the web link above.
#
#
#
```

#SHP-CRUISE	YEAR	MON	DAY	TIMEgmt	TIMEloc	LATITUDE	LONGITDE	UPPER_Z	LOWER_Z	T	GEAR	MESH
#----------	----	---	---	--------	--------	---------	---------	--------	--------	-0	----	----
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	2.400	2.400	31.603	-64.113	0.0	195.0	O	112	200
320G-000216	2000	2	16	2.400	2.400	31.603	-64.113	0.0	195.0	O	112	200

Real plankton population data is noisy



Using noisy data to understand plankton ecosystems: MAR(1) process

p is the number of species in the ecosystem we are studying

The state vectors $\vec{X}_t$ for the system will generically be written as row vectors. That is:

$$\vec{X}_t = (x_{t,1}, \ldots, x_{t,p})$$

where $x_{t,j}$ is the log of the abundance (biomass) of species j at time t.

$$X = \begin{bmatrix} x_{0,1} & x_{0,2} & \cdots & x_{0,p} \\ x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ \vdots & & & \\ x_{t,1} & x_{t,2} & \cdots & x_{t,p} \\ \vdots & & & \\ x_{N-1,1} & x_{N-1,2} & \cdots & x_{N-1,p} \end{bmatrix}$$

Using noisy data to understand plankton ecosystems: MAR(1) process

The community matrix: This is a $p \times p$ matrix

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,p} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p,1} & b_{p,2} & \cdots & b_{p,p} \end{bmatrix}$$

with the usual meaning to its entries. Namely, $b_{i,j}$ relates to the direct effect of species j on the abundance of species i in the next generation. This term is 0 if and only if species j has no direct effect on species i from one generation to the next. However, there may be indirect effects through interactions with other species.

$$X_{t+1}^T = \vec{A} + BX_t^T + \varepsilon_t^T$$

Using noisy data to understand plankton ecosystems: MAR(1) process

$$Y_t = X_{t+1}$$

$$Y = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & & & \\ x_{t+1,1} & x_{t+1,2} & \cdots & x_{t+1,p} \\ \vdots & & & \\ x_{N,1} & x_{N,2} & \cdots & x_{N,p} \end{bmatrix}$$

$$X_{t+1}^T = \vec{A} + BX_t^T + \varepsilon_t^T$$

Using noisy data to understand plankton ecosystems: MAR(1) process

$$Y_t = X_{t+1}$$

$$Y = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & & & \\ x_{t+1,1} & x_{t+1,2} & \cdots & x_{t+1,p} \\ \vdots & & & \\ x_{N,1} & x_{N,2} & \cdots & x_{N,p} \end{bmatrix}$$

$$Y_t^T = \vec{A} + BX_t^T + \varepsilon_t^T$$

Using noisy data to understand plankton ecosystems: MAR(1) process

$$Z = \left[\begin{array}{c|ccc} 1 & x_{0,1} & \cdots & x_{0,p} \\ 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N-1,1} & \cdots & x_{N-1,p} \end{array}\right] \qquad D = \left[\begin{array}{c|ccc} a_1 & b_{1,1} & \cdots & b_{1,p} \\ a_2 & b_{2,1} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_p & b_{p,1} & \cdots & b_{p,p} \end{array}\right]$$

$$Y_t^T = \vec{A} + BX_t^T + \varepsilon_t^T$$

Using noisy data to understand plankton ecosystems: MAR(1) process

$$Z = \left[\begin{array}{c|ccc} 1 & x_{0,1} & \cdots & x_{0,p} \\ 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N-1,1} & \cdots & x_{N-1,p} \end{array}\right]$$

$$D = \left[\begin{array}{c|ccc} a_1 & b_{1,1} & \cdots & b_{1,p} \\ a_2 & b_{2,1} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_p & b_{p,1} & \cdots & b_{p,p} \end{array}\right]$$

$$Y^T = DZ^T + E^T$$

Using noisy data to understand plankton ecosystems: MAR(1) process

$$Z = \left[\begin{array}{c|ccc} 1 & x_{0,1} & \cdots & x_{0,p} \\ 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N-1,1} & \cdots & x_{N-1,p} \end{array}\right] \qquad D = \left[\begin{array}{c|ccc} a_1 & b_{1,1} & \cdots & b_{1,p} \\ a_2 & b_{2,1} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_p & b_{p,1} & \cdots & b_{p,p} \end{array}\right]$$

$$Y^T = DZ^T + E^T$$

$$D = Y^T Z (Z^T Z)^{-1}$$

From noisy time series data we have obtained an estimate of the community matrix!!

Using noisy data to understand plankton ecosystems: MAR(1) process

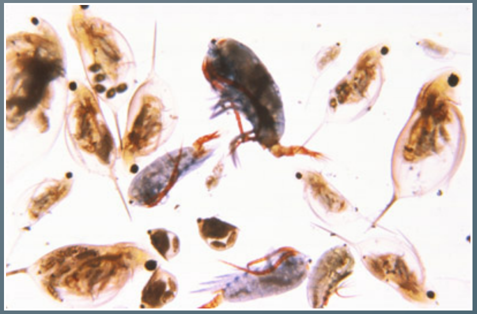
$$Z = \left[\begin{array}{c|ccc} 1 & x_{0,1} & \cdots & x_{0,p} \\ 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N-1,1} & \cdots & x_{N-1,p} \end{array}\right] \qquad D = \left[\begin{array}{c|ccc} a_1 & b_{1,1} & \cdots & b_{1,p} \\ a_2 & b_{2,1} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_p & b_{p,1} & \cdots & b_{p,p} \end{array}\right]$$

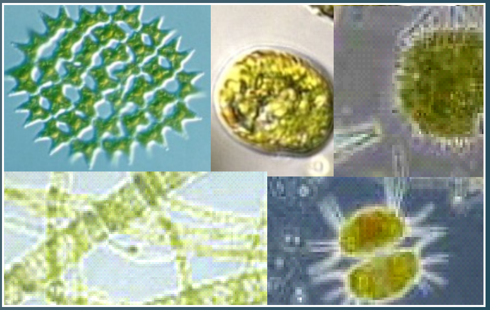
$$Y^T = DZ^T + E^T$$

$$D = Y^T Z (Z^T Z)^{-1}$$

Estimation is similar with covariates (e.g., variable nutrient loads)

Applying MAR(1) estimates to investigate stability of food webs





15 different food webs x 2 nutrient treatments x 4 replicates per treatment = 120 tanks

Sampling intervals every 4-5 days May–Oct totaling 32 sampling dates (zooplankton, phytoplankton) = 3840 data points (256 per food web)

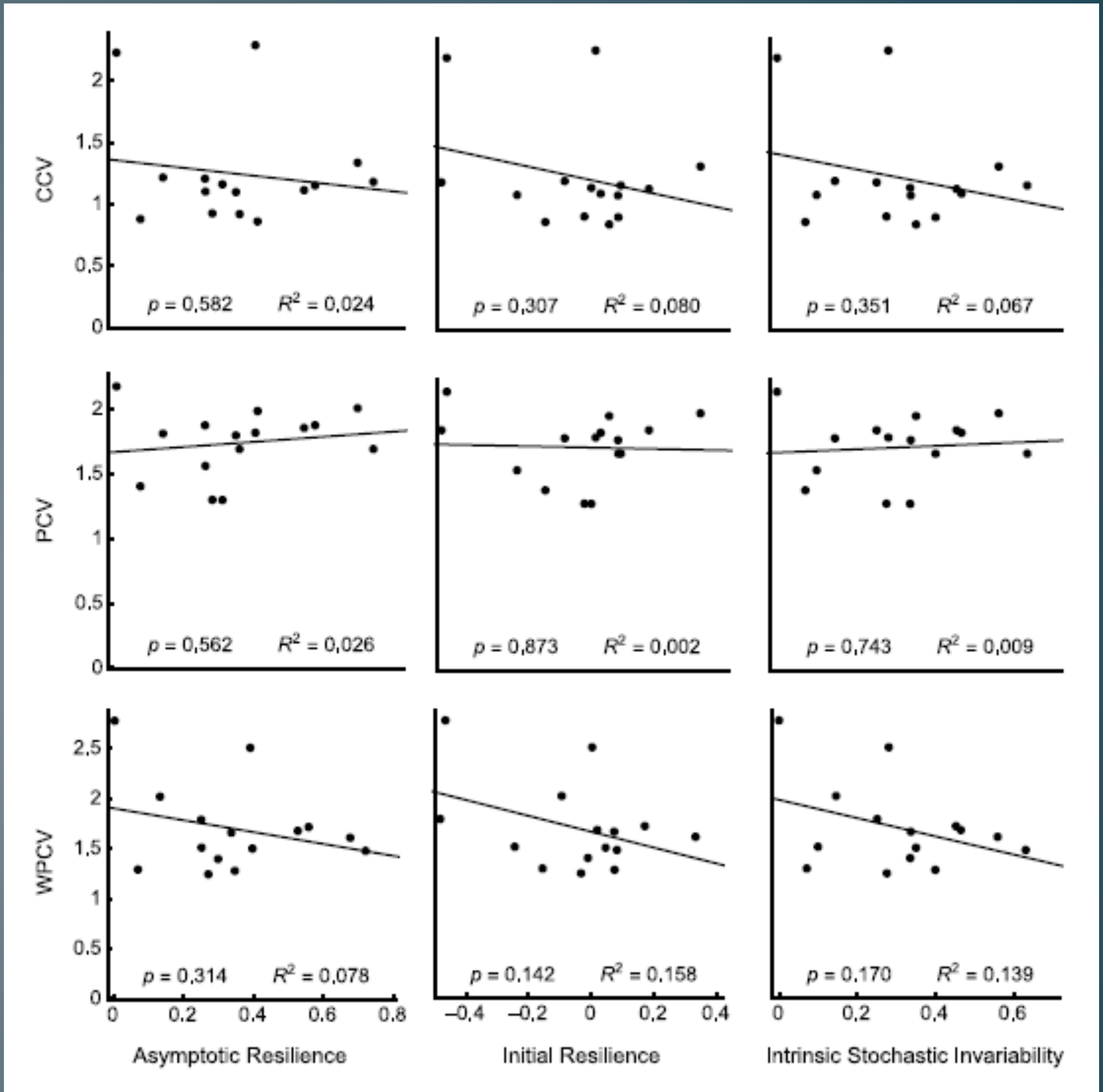
Q: How are different definitions of food web stability related?

Q: How are different definitions of food web stability related?

Empirical stability: coefficients of variation (both on individual species level and community level)

Theoretical stability: measure of system stability as measured from the community matrix

- **Asymptotic Resilience**: long term rate of recovery following a perturbation
- **Initial Resilience**: Worst case initial rate of recovery following a perturbation
- **Intrinsic Stochastic Invariability**: reciprocal of the worst-case variation due to white-noise forcing



No evidence of relationship between theoretical stability measures and empirical stability measures

Q: Does the relative abundance of weakly interacting species affect food web stability?

Q: Does the relative abundance of weakly interacting species affect food web stability?

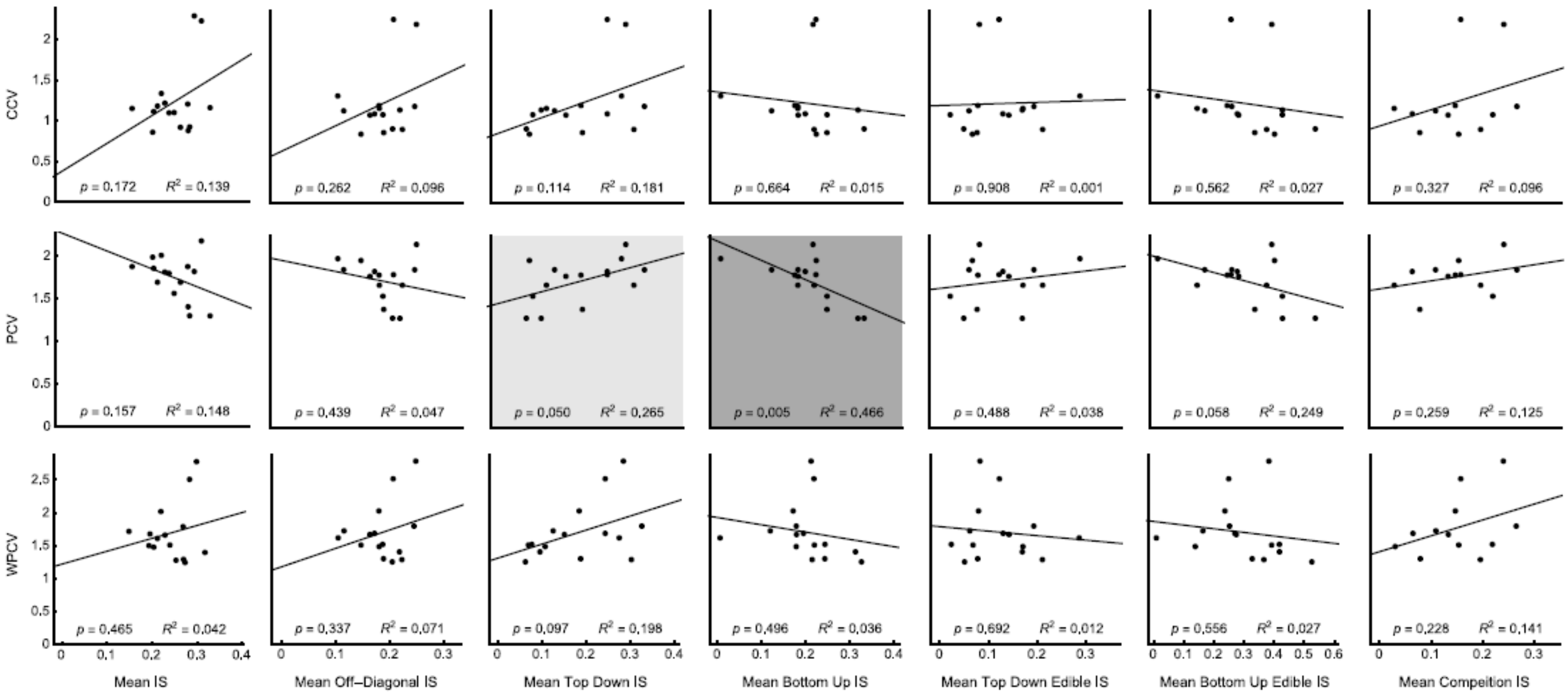


Figure 2: Comparison of interaction strengths (IS) and empirical stability. Figures are shaded light gray if $p < 0.05$ and dark gray if $p < 0.01$.

Comparison of interaction strengths to **empirical** stability measures (lt. gray is p<.05, dk. gray is p<.01)

Q: Does the relative abundance of weakly interacting species affect food web stability?

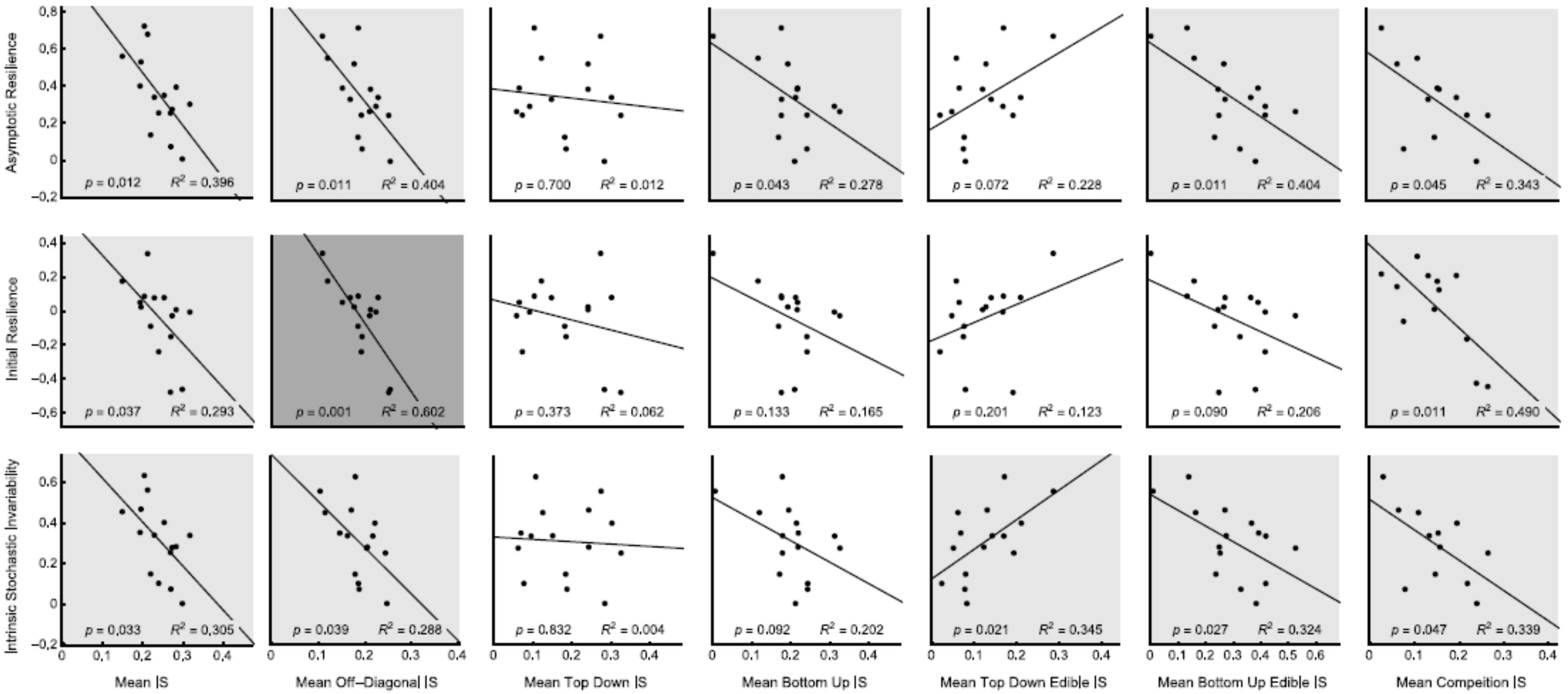


Figure 3: Comparison of interaction strengths (IS) and theoretical stability. Figures are shaded light gray if $p < 0.05$ and dark gray if $p < 0.01$.

Comparison of interaction strengths to **theoretical** stability measures (lt. gray is p<.05, dk. gray is p<.01)

Why data?

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

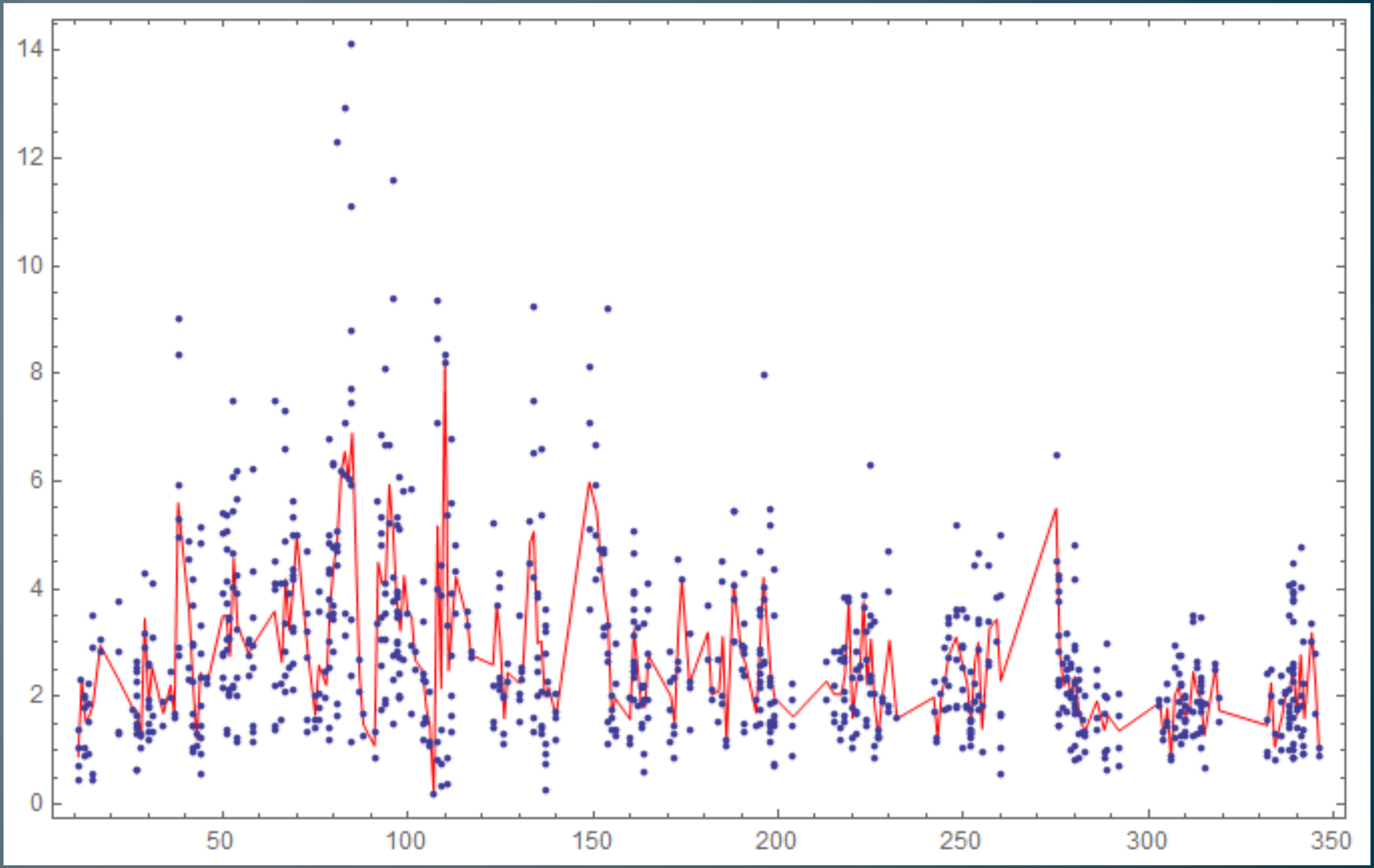
- Data ≠ Information

Data:

#-----*
#
"BATS (JGOFS-BATS): Zooplankton biomass data sampled during the Bermuda Atlantic Time-series Study (BATS) project"
#
#-----*
#
NMFS-COPEPOD: 14-Jun-2013 (Web v10.1)
#
Live-link: http://www.st.nmfs.noaa.gov/plankton/data/batsjgofs/index.html
#
A comprehensive content summary and alternate data formats
are available online at the web link above.
#
#
#

#SHP-CRUISE	YEAR	MON	DAY	TIMEgmt	TIMEloc	LATITUDE	LONGITDE	UPPER_Z	LOWER_Z	T	GEAR	MESH
#----------	----	---	---	--------	--------	---------	---------	--------	--------	-0	----	----
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	1.800	1.800	31.611	-64.124	0.0	190.0	O	112	200
320G-000216	2000	2	16	2.400	2.400	31.603	-64.113	0.0	195.0	O	112	200
320G-000216	2000	2	16	2.400	2.400	31.603	-64.113	0.0	195.0	O	112	200

Information:



Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

- Data ≠ Information
- Real data is often messy, incomplete, not suited to answer the questions we most want to ask of it

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

- Data ≠ Information
- Real data is often messy, incomplete, not suited to answer the questions we most want to ask of it
- Real data can be wrong!

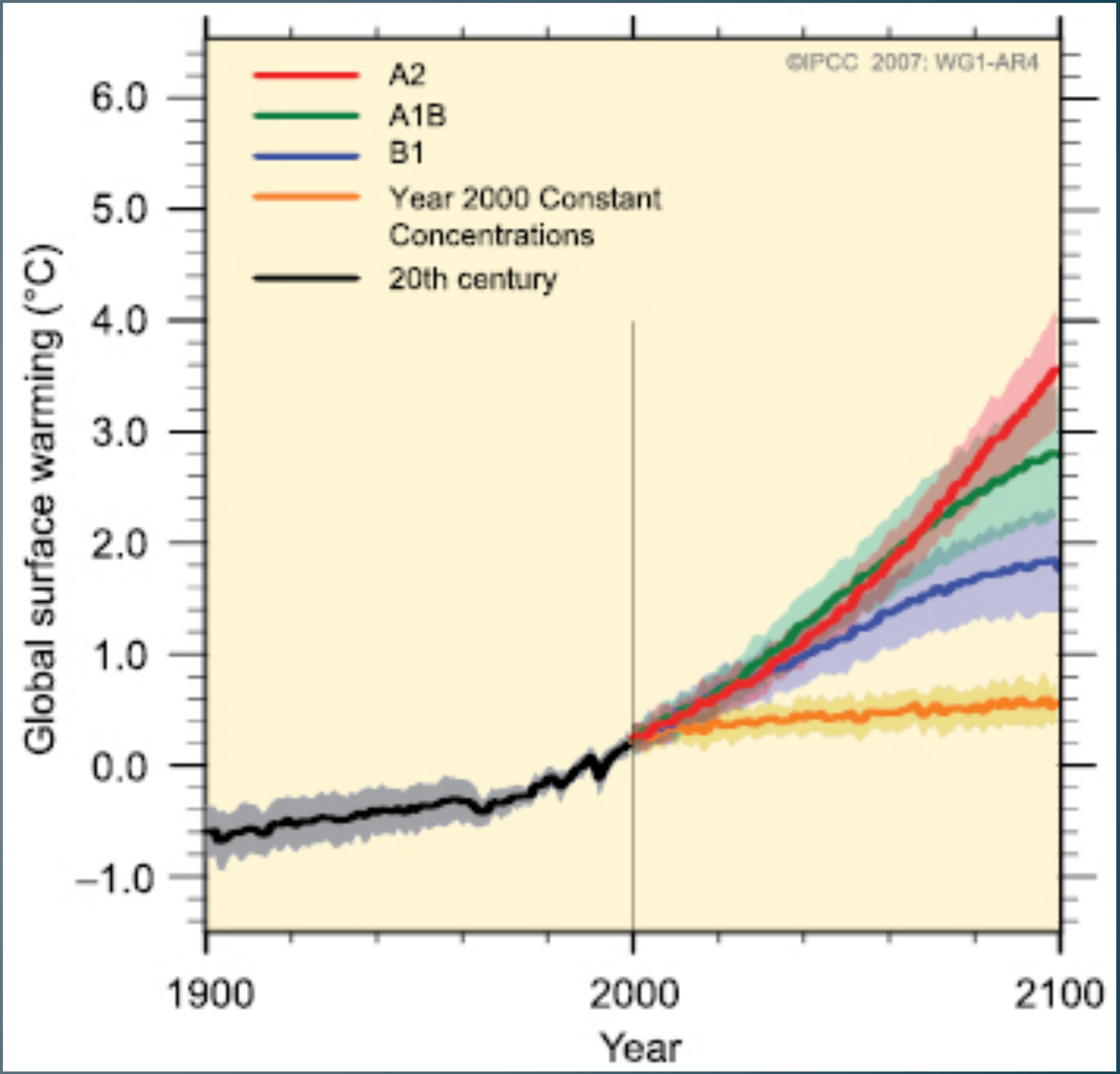
	A	B	C	D	E	F	G	H	I
1	tank	FLUCT	COMP	Cer_ug	Dapp_ug	Sca_ug	Clad_ug	e_chl	i_chl
2649	97	150	N_DAP	876.0693		102.4033	10.53007	2.68	0.3
2650	97	150	N_DAP	187.5769		52.5006	4.665915	3.29	0.26
2651	97	150	N_DAP	353.5519		133.1986	9.445592	17.33	1658.87
2652	97	150	N_DAP	130.3732		105.1579	5.465053	22.55	2566.01
2653	97	150	N_DAP	313.6929		139.2908	10.97183	30.51	317.32
2654	97	150	N_DAP	101.885		57.94207	3.445877	60.12	543.1
2655	97	150	N_DAP	290.2114		85.13787	9.06071	110.74	634.29
2656	97	150	N_DAP	54.66454		21.87718	1.038029	160.22	3020.48
2657	99	113.3	N_DAP	0.022		15.9594	0.13556	51.37	0.65
2658	99	176.9	N_DAP	0.022		47.8624	0.0339	122.8	-0.02
2659	99	139.3	N_DAP	0.821267		177.9185	0.13556	77.39	12.14
2660	99	104.4	N_DAP	0.022		245.2349	0.0339	49.97	-0.01
2661	99	170.6	N_DAP	0.022		130.6884	0.27112	60.96	5.42
2662	99	134.4	N_DAP	4.927604		86.19687	0.27112	40.45	0.005
2663	99	100.9	N_DAP	1.642535		95.48486	0.0339	42.4	1.67
2664	99	168.1	N_DAP	0.022		236.7112	0.0339	42.4	1.67
2665	99	132.5	N_DAP	0.022		414.0479	0.352611	63.93	0.005
2666	99	99.49	N_DAP	13.85858		2350.1	2.494924	28.56	0.67
2667	99	167.1	N_DAP	75.14475		2177.052	0.0339	1.7	0.05
2668	99	131.7	N_DAP	619.7909		35.26096	1.247462	1.23	0.3
2669	99	98.95	N_DAP	482.0288		0.454049	0.542239	1.59	0.08
2670	99	166.7	N_DAP	658.0386		154.3092	47.06997	3.05	50.99
2671	99	131.4	N_DAP	56.62793		10.00067	12.15635	1.79	0.12

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

- Data ≠ Information
- Real data is often messy, incomplete, not suited to answer the questions we most want to ask of it
- Real data can be wrong!
- Many questions can't be answered with data at all!



©IPCC 2007: WG1-AR4
A2
A1B
B1
Year 2000 Constant Concentrations
20th century
Global surface warming (°C)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
–1.0
1900
2000
2100
Year

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

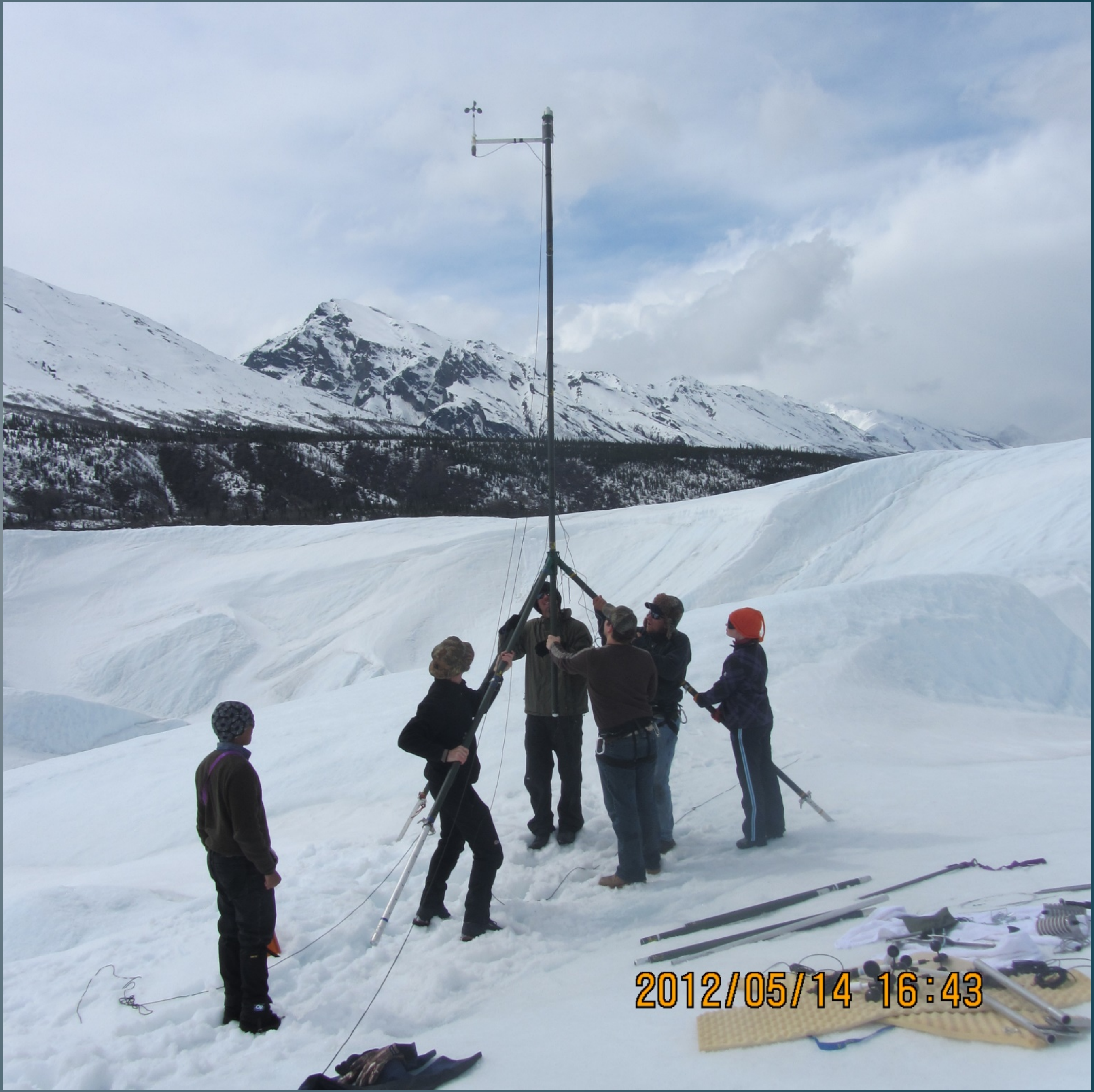
- Data ≠ Information
- Real data is often messy, incomplete, not suited to answer the questions we most want to ask of it
- Real data can be wrong!
- Many questions can't be answered with data at all
- Data can be misused

Why data?

- We can learn a lot of interesting and important things from data about ourselves, our society, our world….
- We have lots of it lying around

But

- Data ≠ Information
- Real data is often messy, incomplete, not suited to answer the questions we most want to ask of it
- Real data can be wrong!
- Many questions can't be answered with data at all
- Data can be misused
- Collecting high quality data is hard!


2012/05/14 16:43


15

Thanks!