Marine Biogeochemical Modeling: Basic Philosophy and Methods

Scott Doney (WHOI)

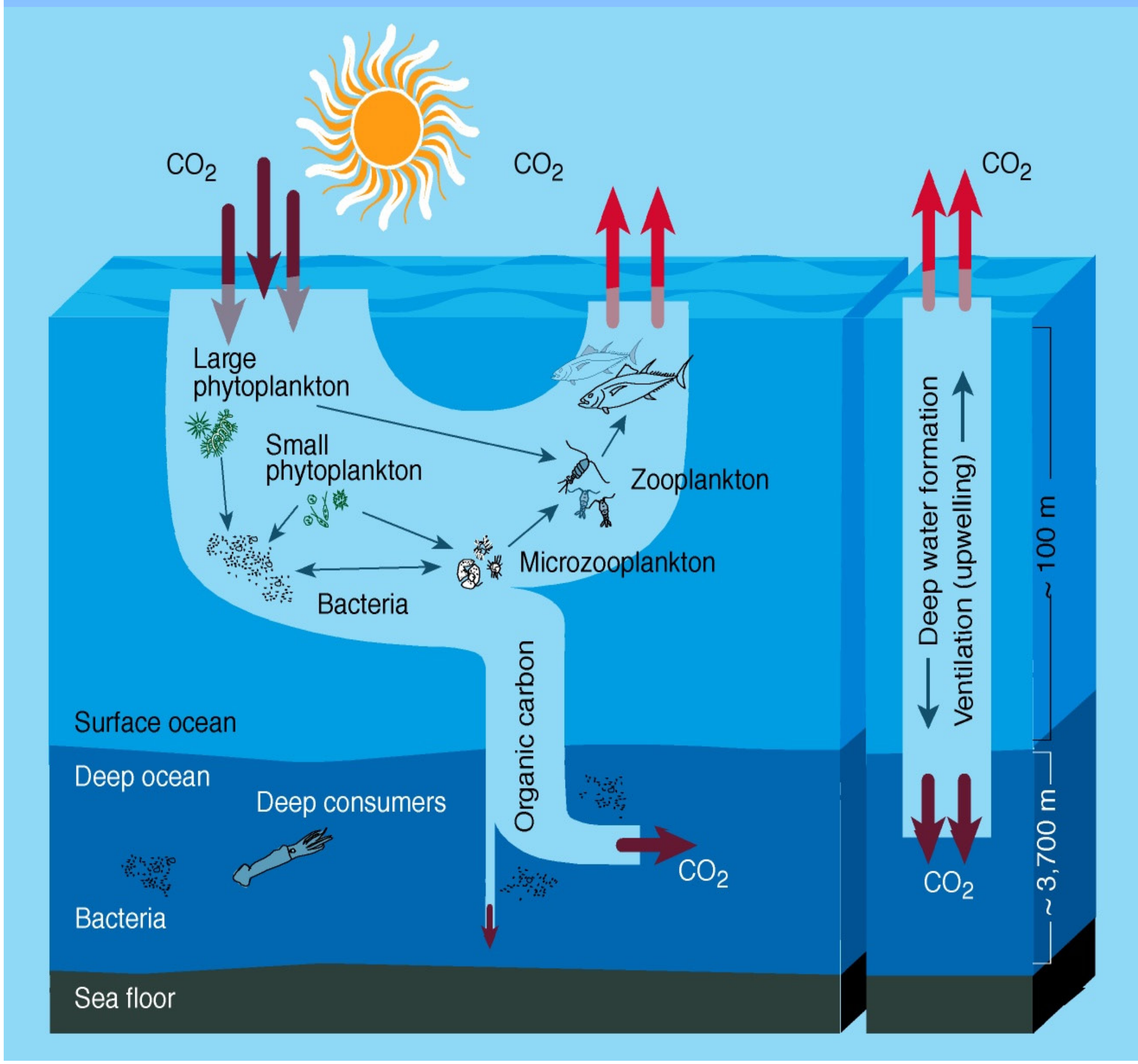
OCB Ocean Acidification Short Course 2009

Talk Outline

- Science Motivation
- Philosophy of Modeling
- Model Construction
- Evaluation Against Data

Supported by:

Ocean Carbon Cycle & Ocean Acidification



Anthropogenic CO_2 uptake currently controlled by ocean circulation; but in future, what will be role of climate & biology?

For ocean acidification may want models to address many different aspects:

-patterns & trends in seawater chemistry

-population biology of individual species

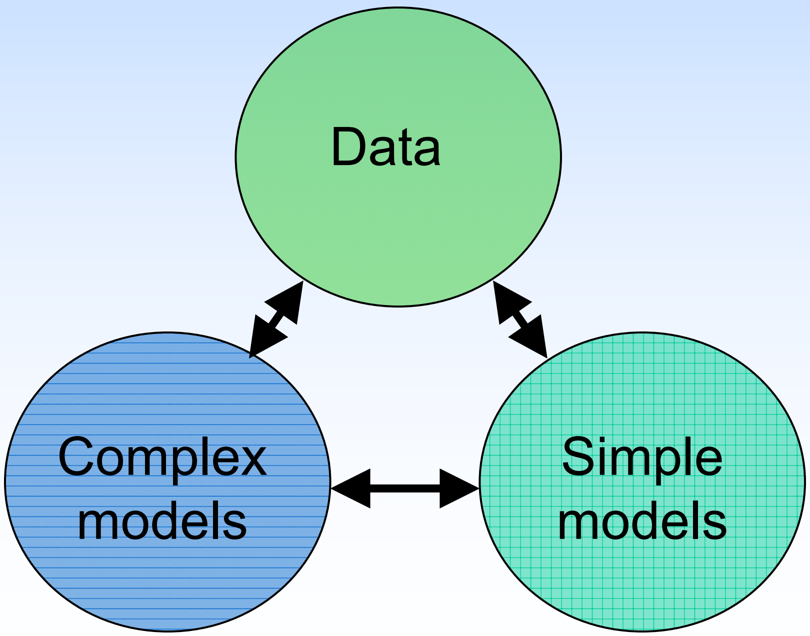
-food-web & ecological interactions

-biogeochemical feedbacks

-socio-economic effects on fisheries & ecosystem services

"I am never content until I have constructed a mechanical model of the subject I am studying. If I succeed in making one, I understand; otherwise I do not.“
- Lord Kelvin

"People don't understand the earth, but they want to, so they build a model, and then they have two things they don't understand,”
-Gerard Roe in “The Whale and the Supercomputer” by C. Wohlforth



The Interdisciplinary Conundrum

Physics

- (relatively) simple set of (mostly) known governing equations (Navier-Stokes) => complex phenomenon
- unresolved scales

Chemistry

- mass balance equations
- chemical fields integrate over time/space variability
- don't uniquely identify process/mechanism

Biology

- information as stories and conceptual pictures
- historical/evolutionary contingency
- no "biological" Navier-Stokes
- aggregate complicated dynamics across multiple-scales (genes, cells, populations, ecosystems)

Why do we build/use models?

-Quantitative dynamical framework

- Are different data sets, rate estimates consistent?

-Design of experiments & observing systems

- What, where and when do we sample?

-Hypothesis testing

- If we add or change X, what happens?

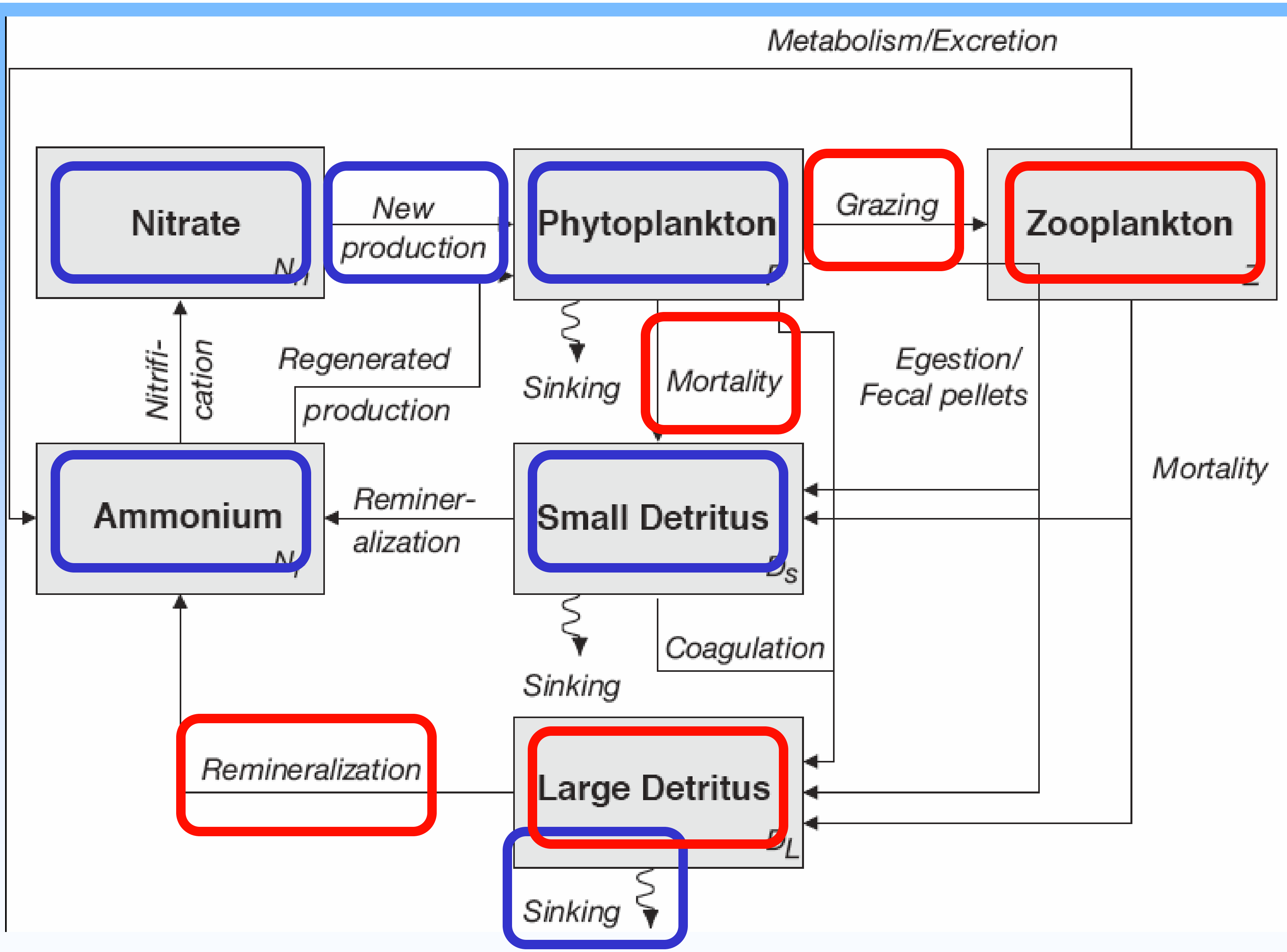
-Forecasting

- What will the ocean look like at some point in the future?

-Solving for unknown parameters and rates

- Given things we can measure, can we estimate can we estimate properties that are difficult to measure?

WOODS HOLE OCEANOGRAPHIC INSTITUTION 1930



"Stocks" versus "Rates"
C dC/dt

Estimated from data
Unknowns

What is a model?

-regression curves

variable Y is a function of other variables X and regression parameters p:

Y=f(variables, parameters)

$$\mathrm{Chl} = p_0 + p_1 T$$

-forward models *(often time dependent)*

Integrate forward in time to find Y:

dY/dt = circulation + f(variables, parameters, forcing)

$$d\,\mathrm{Phyto}/dt = \mathrm{circulation} + \mu * \mathrm{Phyto} * (1 - e^{-\alpha E/\mu}) - \lambda \mathrm{Phyto}$$

-inverse models

Invert the problem to find parameters from data:

Parameters = f(circulation, data, forcing)

WOODS HOLE OCEANOGRAPHIC INSTITUTION 1930

What is a model? (continued)

-diagnostic versus prognostic models

In diagnostic models, some variables may be prescribed based on observations:

e.g., satellite chlorophyll => ecosystem model

(no equation for dChl / dt)

-data assimilation

Combine model equations and observations in a dynamically consistent fashion

e.g., weather prediction

analysis = f(model forecast, observations)

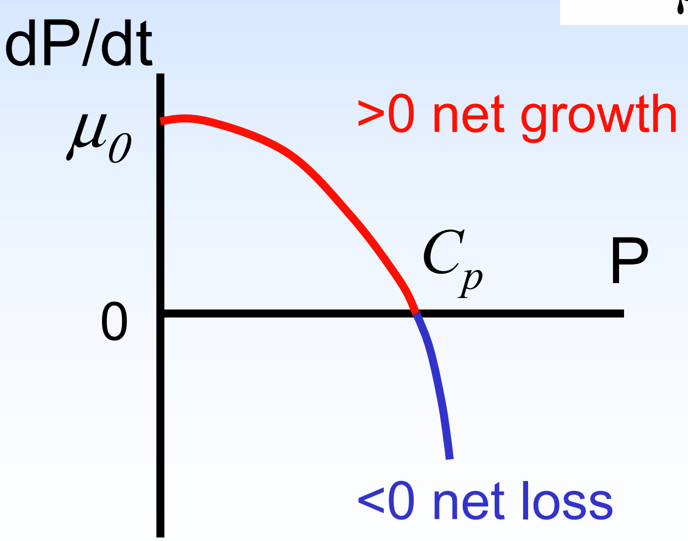
Prognostic, forward models needed to project into the future

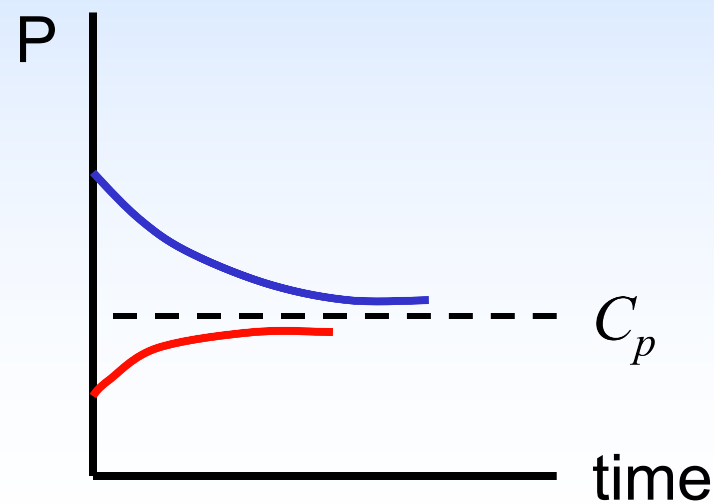
From Word Problem to Equations

- Phytoplankton levels depends on nutrient inputs
- Perturbations relax back to some stable background level

$$\frac{dP}{dt} = \mu_0 \left(1 - \frac{P}{C_p}\right) P$$

- Functional form
  - Logistic model
- State variable (concentrations)
  - P (mmol C/m^3) phytoplankton
- Parameters
  - μ_0(1/d) and C_p (mmol C/m^3)





Simple NPZ Model

$$\frac{dP}{dt} = \mu_0 \left(\frac{N}{k_N + N} \right) \left(1 - e^{\alpha E/\mu 0} \right) P - g \left(\frac{P}{k_P + P} \right) Z - m_P P$$

Nutrient limitation   Light limitation   Grazing   Mortality

$$\frac{dZ}{dt} = ag \left(\frac{P}{k_P + P} \right) Z - m_Z Z$$

$$\frac{dN}{dt} = -\mu_0 \left(\frac{N}{k_N + N} \right) \left(1 - e^{\alpha E/\mu 0} \right) P + (1 - a) g \left(\frac{P}{k_P + P} \right) Z + m_P P + m_Z Z$$

- Three coupled ordinary differential equations
- Mass conservation

Discretization & Numerical Integration

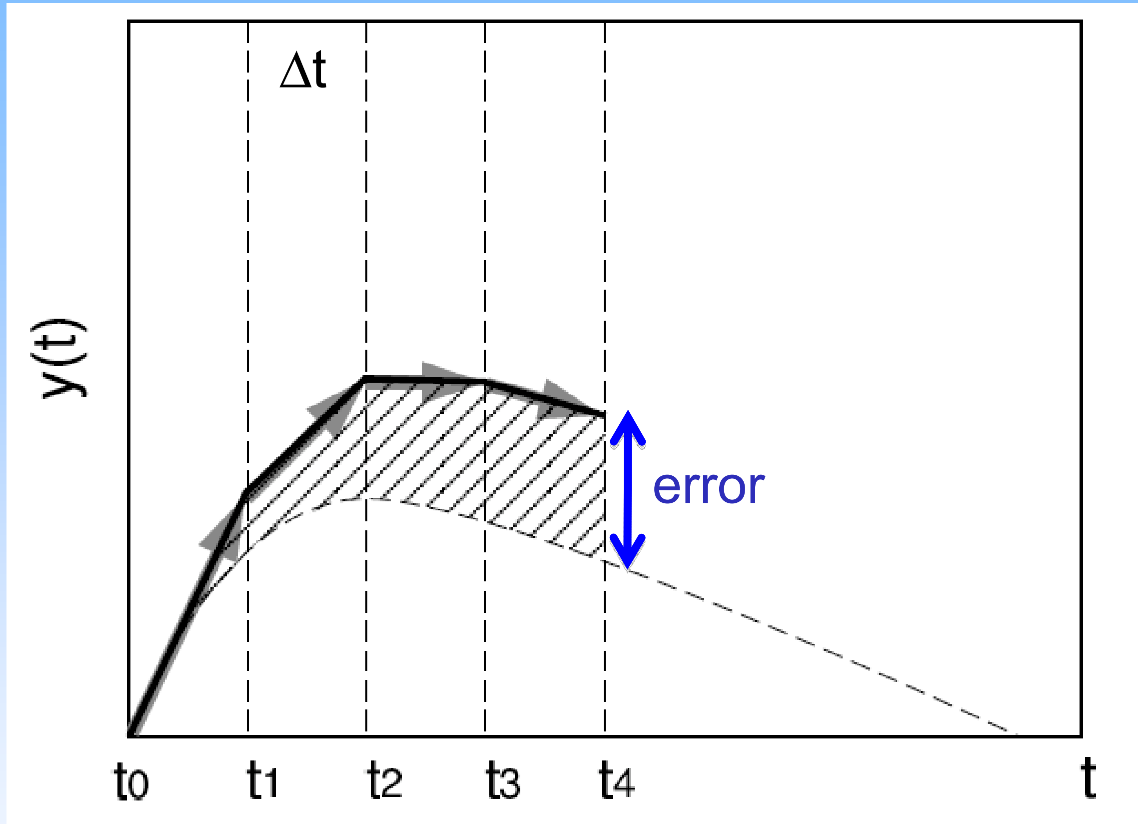
$$\frac{dy}{dt} = f(t,y)$$

Discrete form

$$\Delta y = f(t,y)\Delta t$$

numerical method:
Euler's Method

$$y^{n+1} = y^n + f(t^n, y^n)\Delta t$$



- Subdivide time into discrete time-steps
- "Integrate" forward using Δy/Δt approximation
- Numerical methods introduce errors

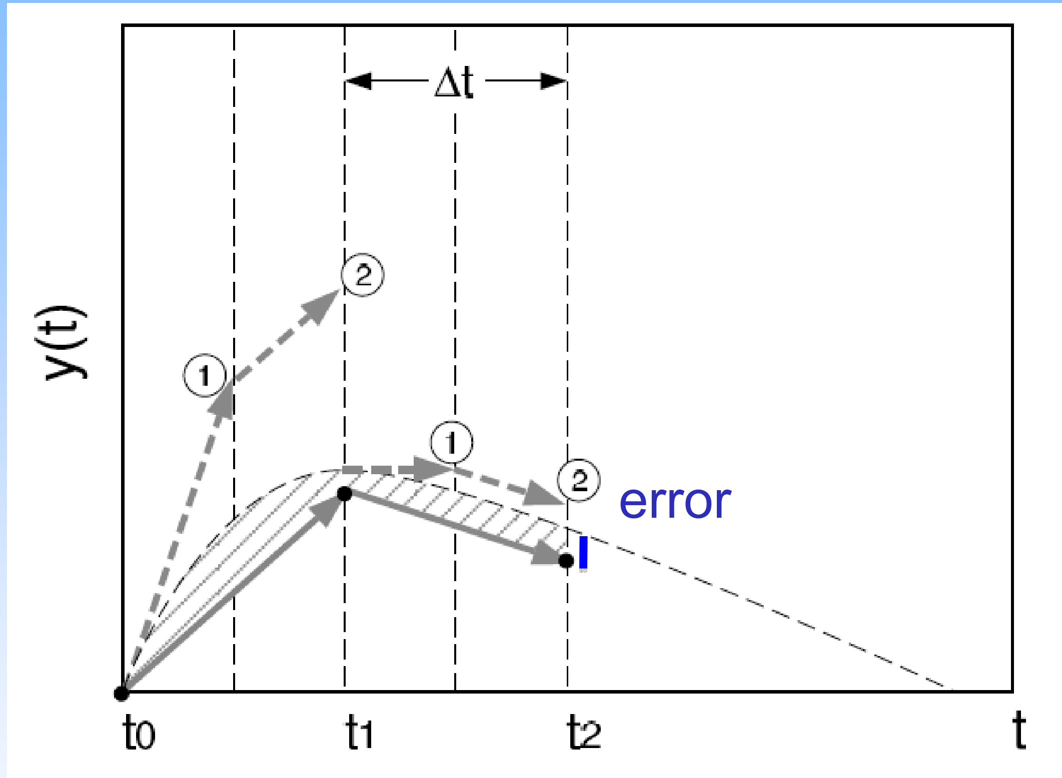
Discretization & Numerical Integration

Runge-Kutta (2nd order)

$$k_1 = f(t^n, y^n)\Delta t$$

$$k_2 = f(t^n + \frac{1}{2}\Delta t, y^n + \frac{1}{2}k_1)\Delta t$$

$$y^{n+1} = y^n + k_2$$



- use a “trial step” to find gradient at mid-point
- even with 2x larger Δt, more accurate integration

How do you estimate parameters and functional forms?

Laboratory & field incubations

- P-E curves; nutrient uptake curves
- elemental stochiometry

Comparative analysis

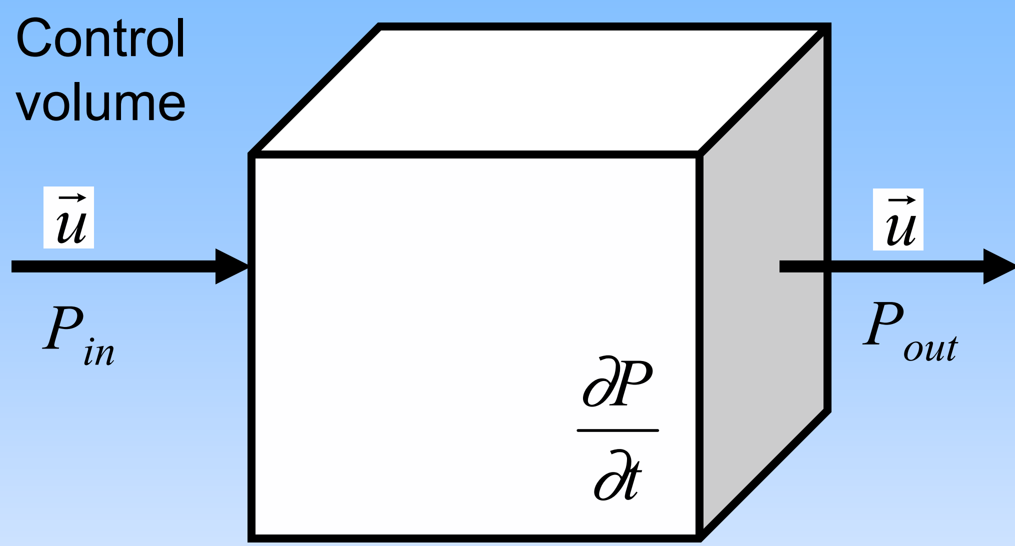
- allometric relationships

Tuned or optimized against field data

- mismatch between parameters and data
- cross-site comparison

Previous models

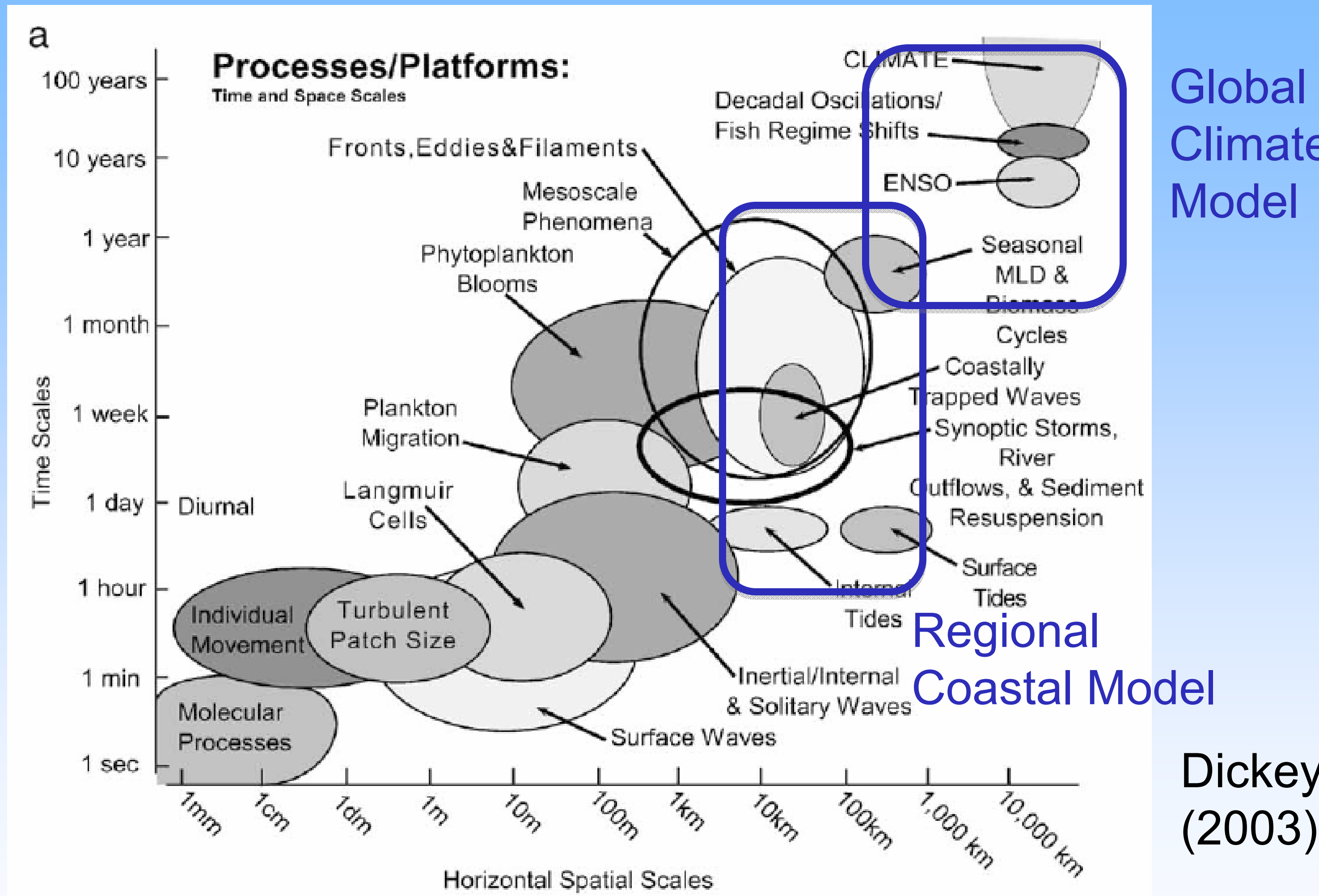
Adding Circulation



$$\frac{\partial P}{\partial t} + \vec{u} \cdot \nabla P - \kappa \nabla^2 P = RHS$$

advection diffusion biological source/sink terms

Models have time/space scale limits



Dickey (2003)

- •Computational costs scale as (length)^3 to (length)^4 and (time)
- •Typically get ~2-3 decades in space (more in time)
- •Can not resolve all scales; parameterize sub-grid scale

Coupled "Eco-biogeochemical" Elements

Physics (flow field; mixing)

- equations for resolved flow; level of approximation (e.g., primative equations; quasi-geostrophy)
- forcing (winds, heat & freshwater fluxes, light, tides)
- parameterization of unresolved scales (mixing)
- model architecture (e.g., horizontal vs. isopycnal)

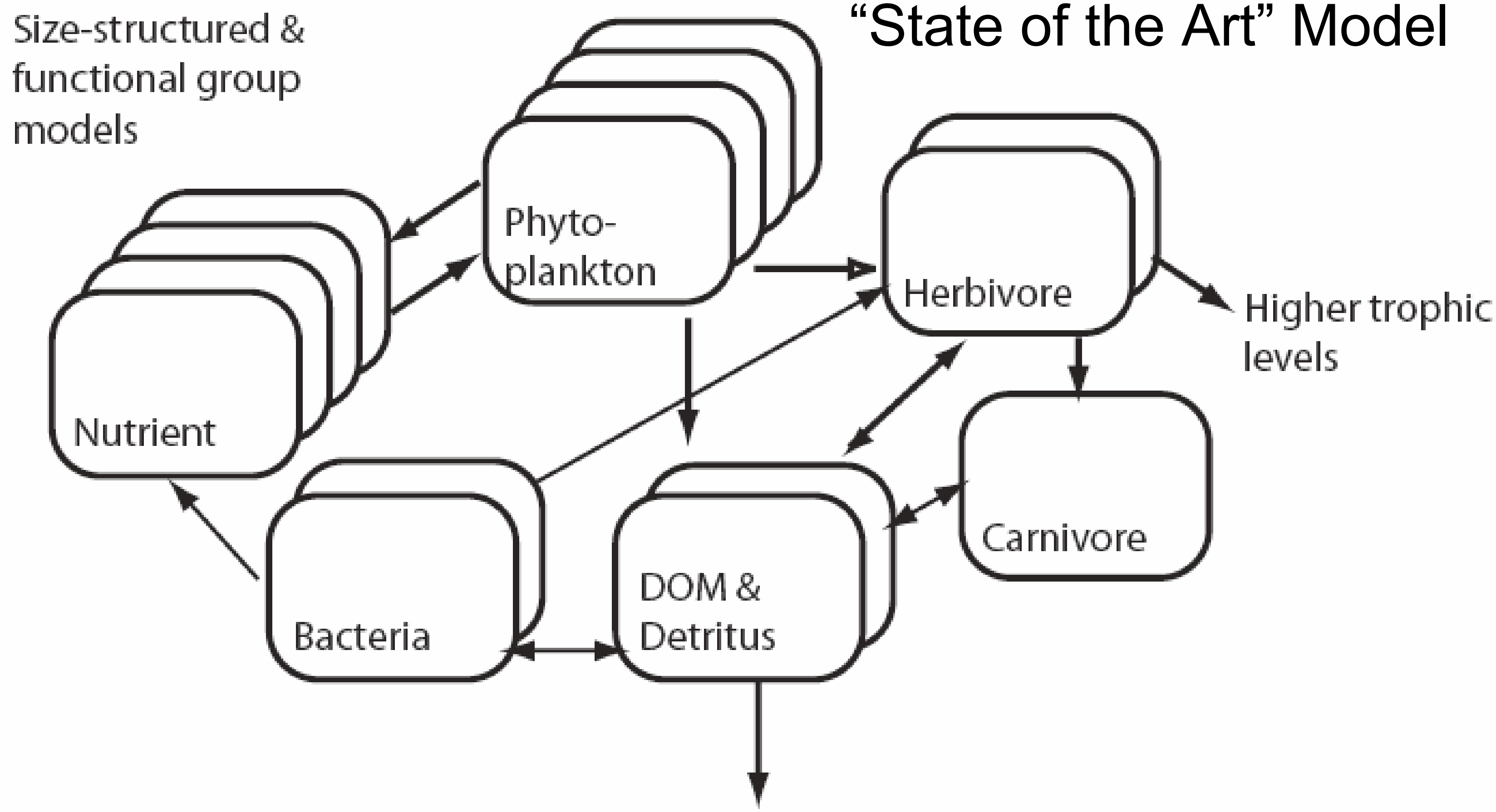
Chemistry (CO_2, O_2, nutrient fields)

- air-sea gas exchange
- elemental stoichiometries
- trace metal deposition and scavenging

Biology

- primary production, respiration, remineralization
- community structure and succession
- bio-optics
- etc.

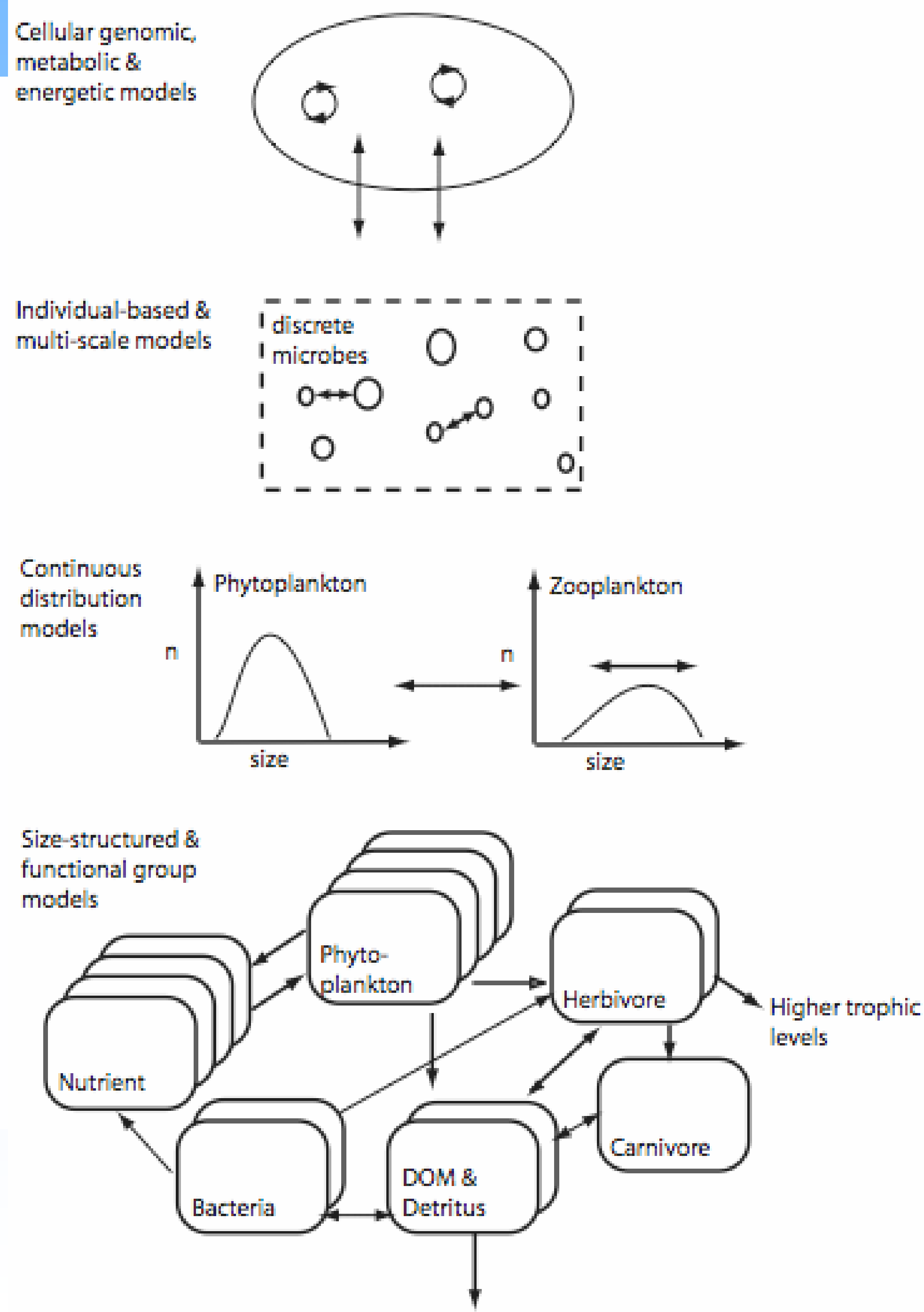
“State of the Art” Model



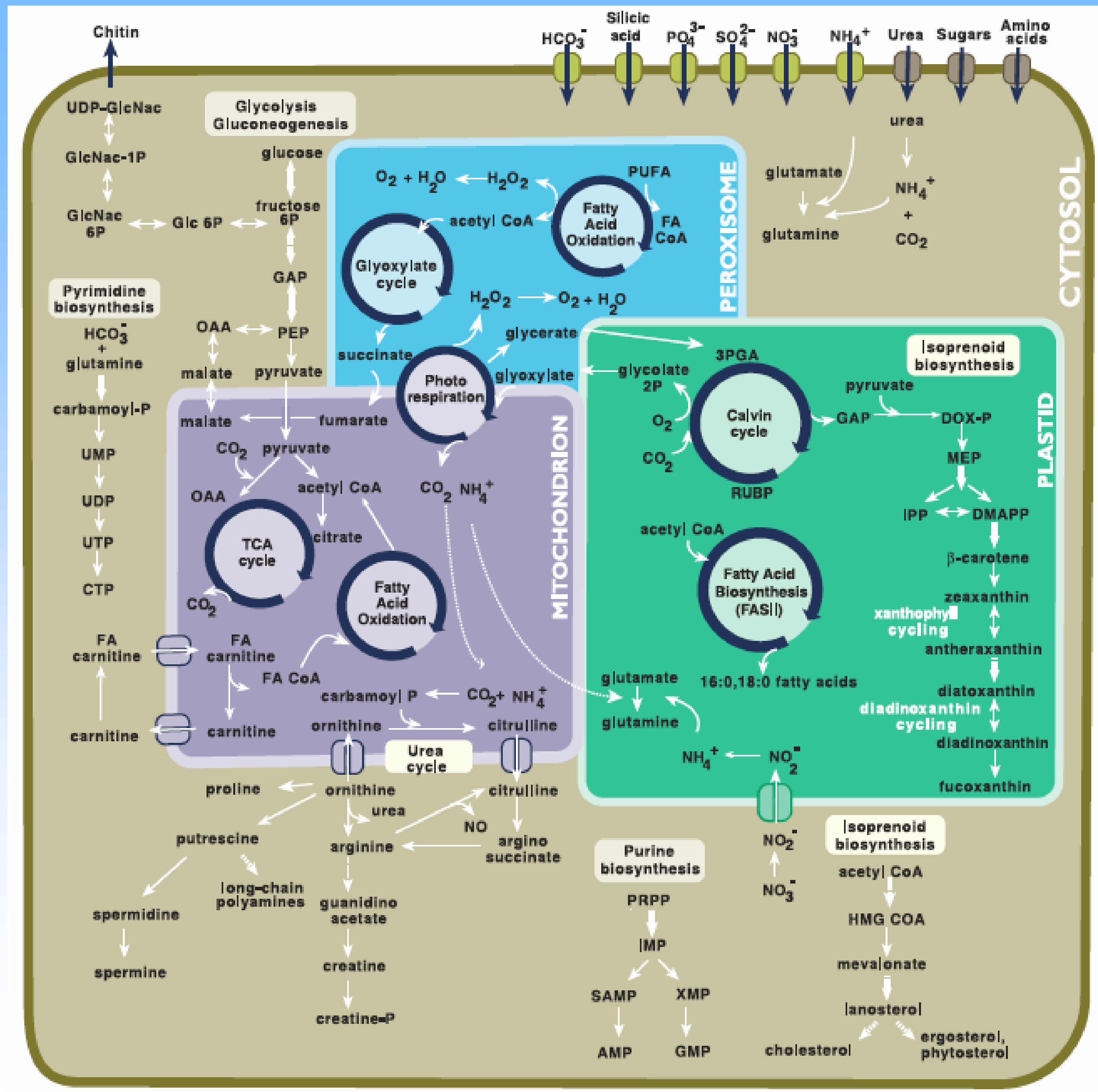
- -Aggregate into trophic levels/functional groups
- -Rates/processes from limited culture/field studies
- -Many aspects empirically based
- -Data poor for validation (rates, grazing, loss terms)

Multi-scale models

-cellular models
•gene expression, metabolism, energetics
-population models
•individual based or continuous dist.
•cell-cell interactions (LES & DNS)
-simulate ecological functions
•“genotype” => “phenotype”
•abandon “boxes”
-ecological/evolutionary rules for ecosystem assembly
•maximize resiliance or energy flow
-emergent behavior & selection
•selection and niche adaptation
•Physiological plasticity & constraints
•micro/macroevolution



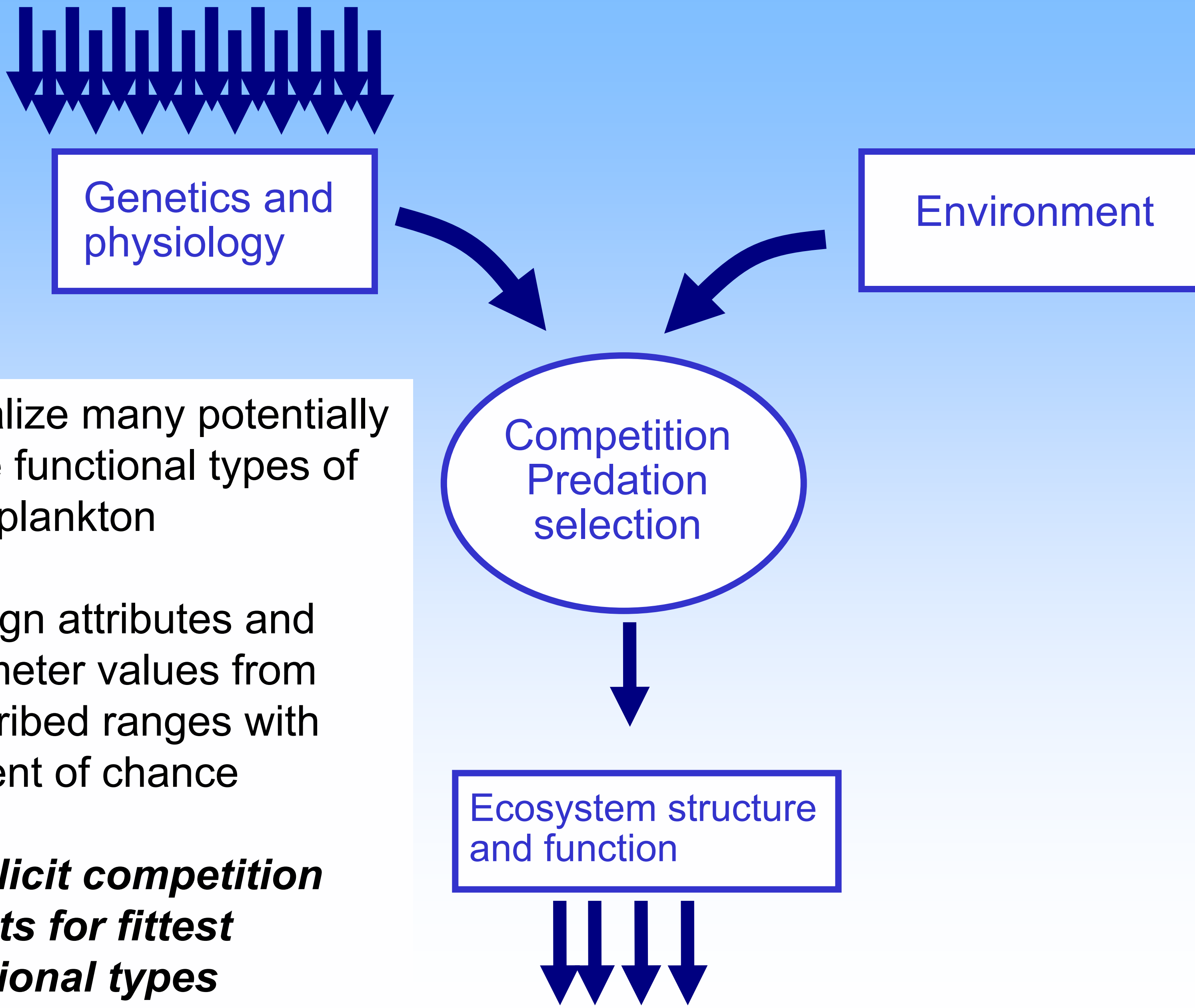
Cell Physiology/ Genomics



Diatom Genome

Armbrust et al., Science, 2004

Competition Based Approach

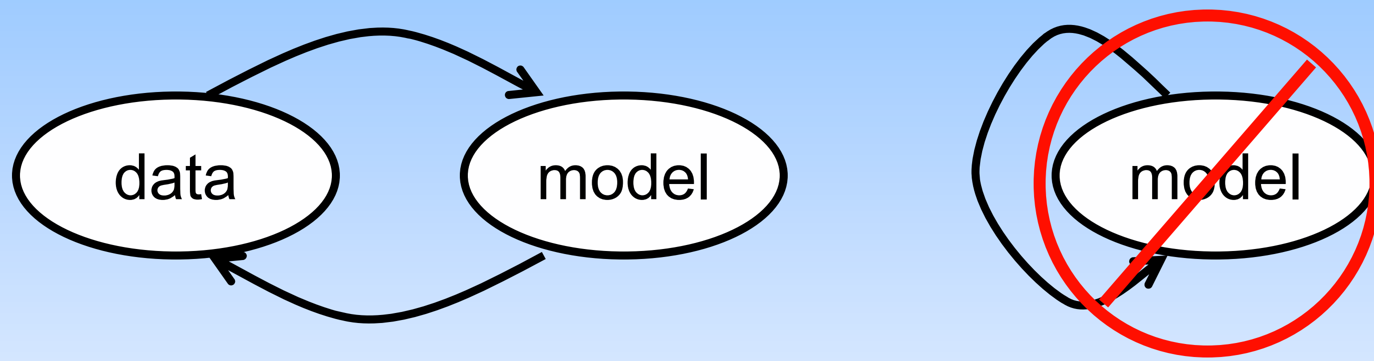


- Initialize many potentially viable functional types of phytoplankton

- Assign attributes and parameter values from prescribed ranges with element of chance

- ***Explicit competition selects for fittest functional types***

Ocean ecology and biogeochemistry are (still) data-driven sciences

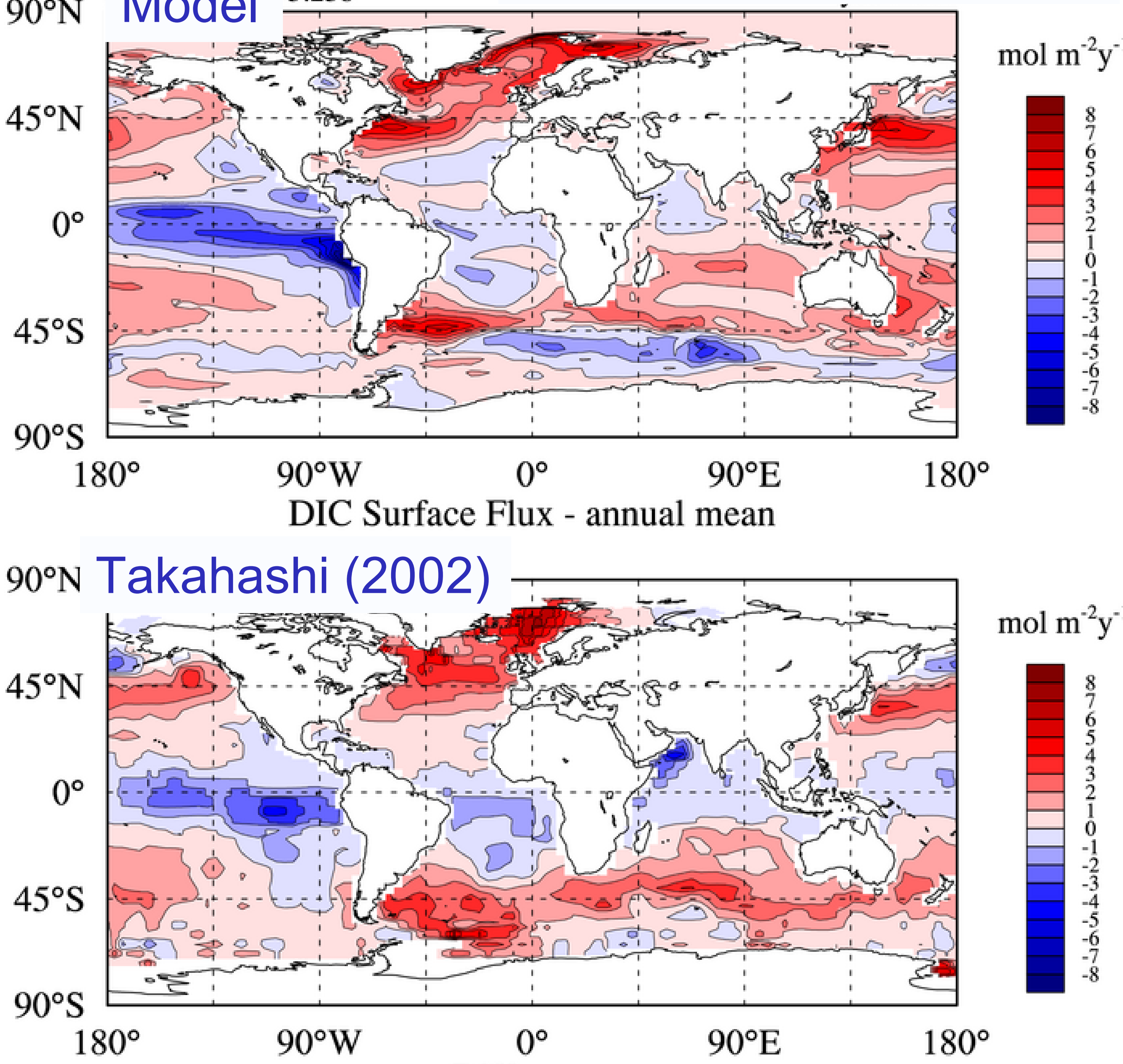


How do we avoid the trap of:

"false models tested by inadequate data"

John Steele

Modern Air-Sea CO_2 Flux



<u>"Looks pretty good" test</u>

χ (chi) by eye

Doney et al. Deep-Sea Res. II 2009

Doney et al. J. Mar. Systems 2009

J. Marine Systems Special Issue on
Skill Assessment for Coupled Biological / Physical Models of Marine Systems
Vol. 76, Issue 1-2, 2009

WOODS HOLE OCEANOGRAPHIC INSTITUTION
1930

4) AE – the average error (bias)

Bias

$$AE = \frac{\sum_{i=1}^{n} (P_i - O_i)}{n} = \overline{P} - \overline{O},$$

1) r – the correlation coefficient of the model predictions and observations:

Correlation

$$r = \frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2 \sum_{i=1}^{n} (P_i - \bar{P})^2}},$$

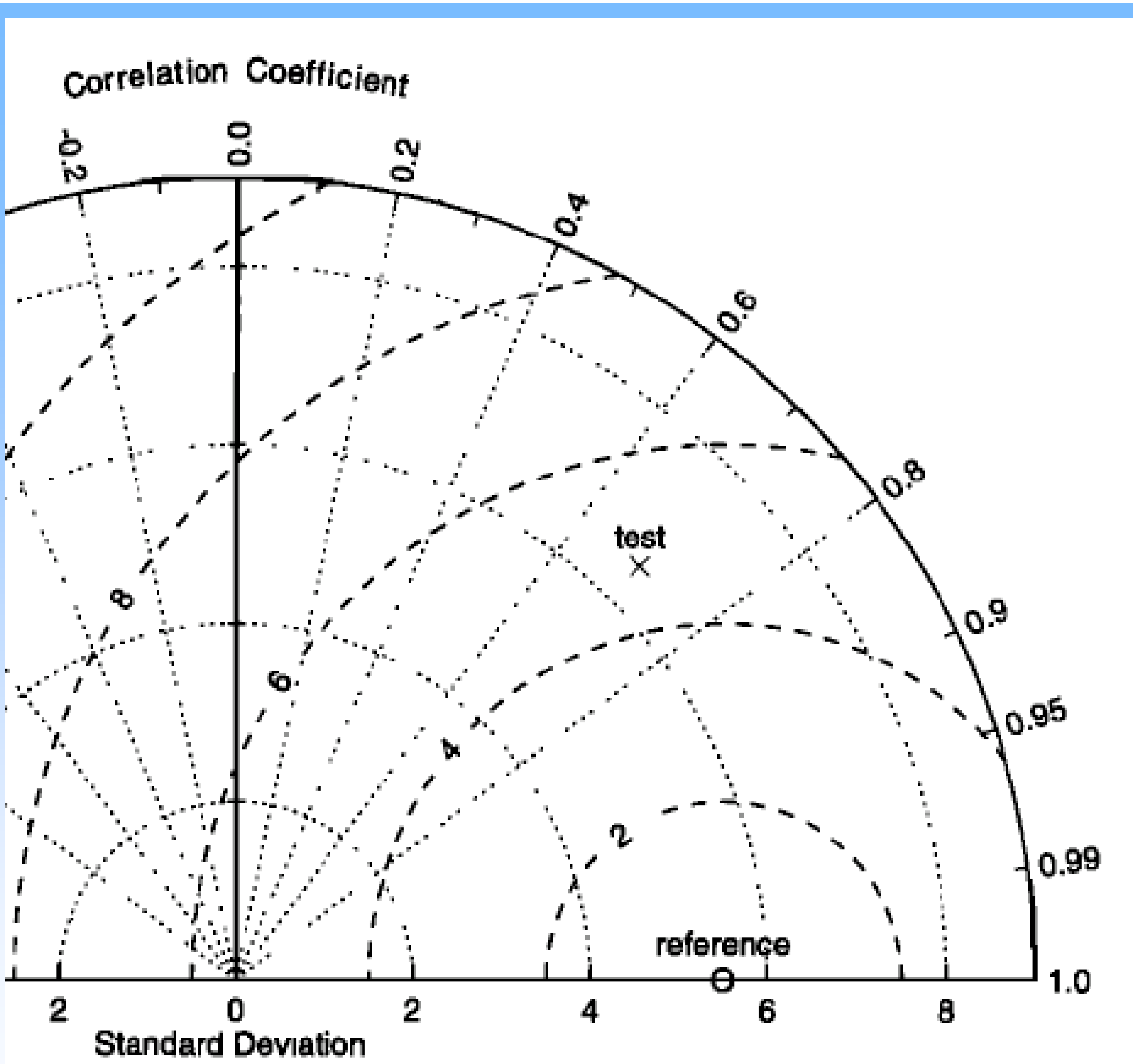
2) RMSE – the root mean squared error (also referred to as root mean squared difference):

rms Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}},$$

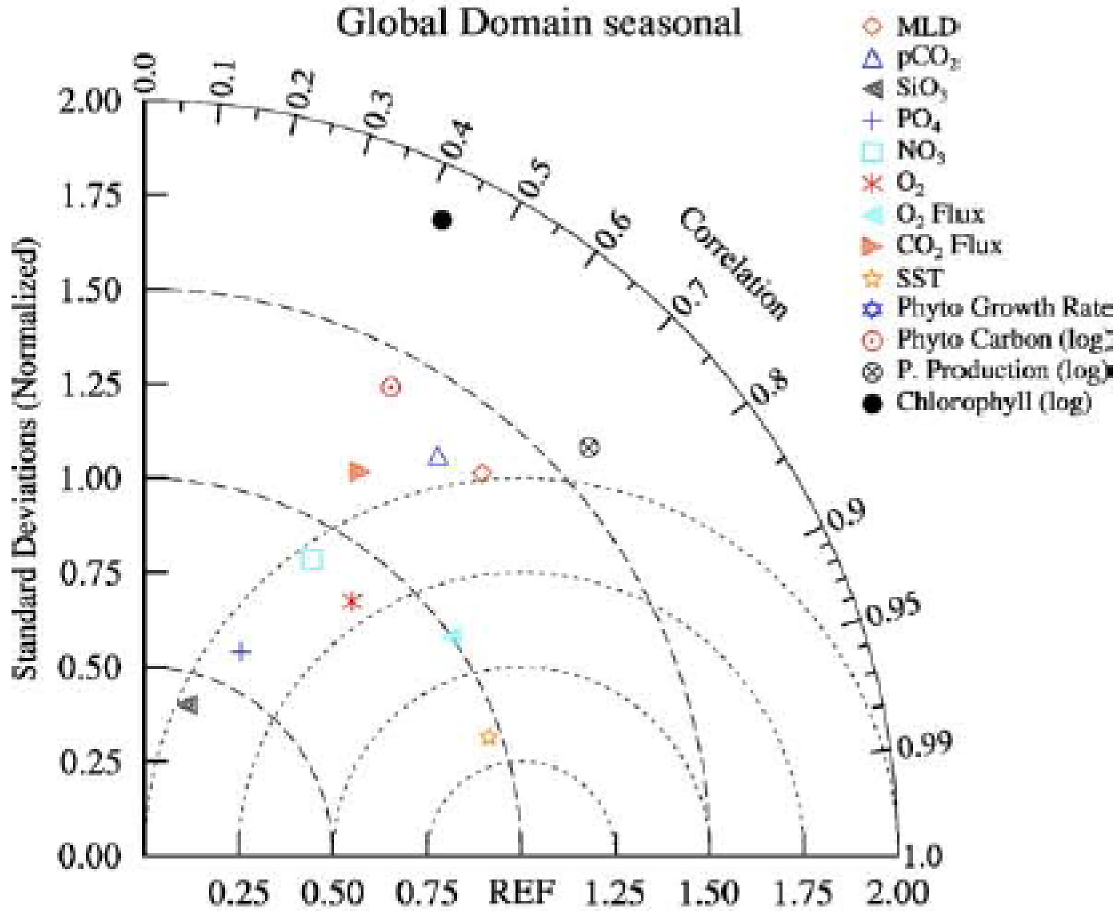
Doney et al. J. Mar. Systems 2009
Stow et al. J. Mar. Systems 2009

Taylor Diagram

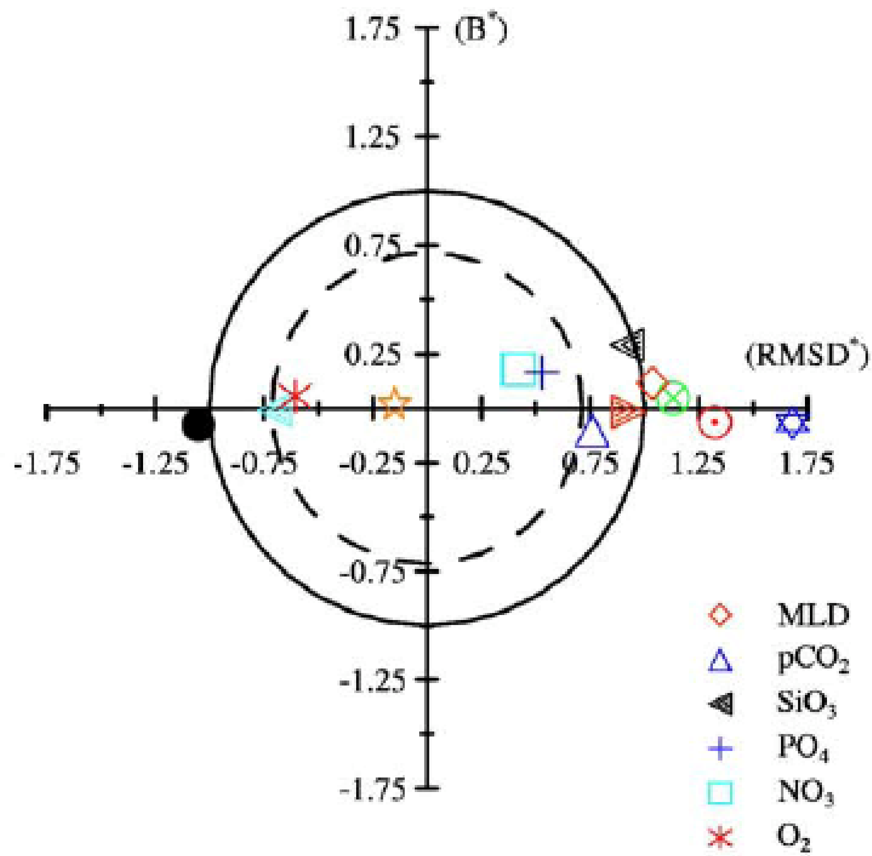


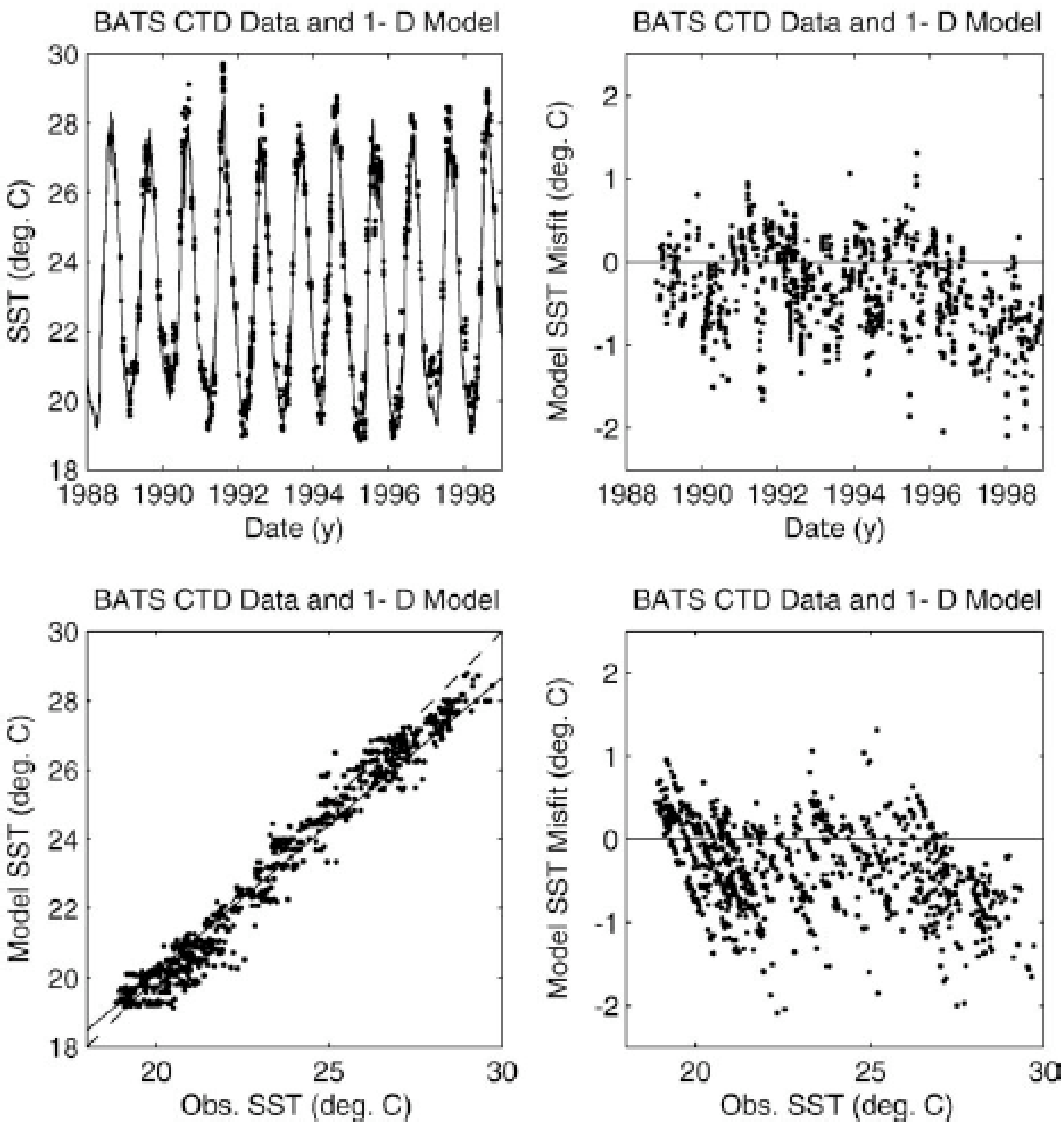
Taylor J. Geophys. Res. 2001

Figure 2. Diagram for displaying pattern statistics. The radial distance from the origin is proportional to the standard deviation of a pattern. The centered RMS difference between the test and reference field is proportional to their distance apart (in the same units as the standard deviation). The correlation between the two fields is given by the azimuthal position of the test field.



e-Time (Seasonal Climatology)





Look at the magnitude & structure in model-data residuals

Ducklow et al. Ann. Rev. Mar. Res. 2009

Log-Normal Variables (e.g., chlorophyll)

$$X = \log(\chi) \qquad (6)$$

$$\langle \chi \rangle_G = \sqrt[N]{\prod_i^N \chi_i} = \exp(\langle X \rangle) \qquad (7)$$

geometric mean

The geometric bias:

$$\varepsilon_{\text{bias}}^{G}(\chi) = \exp(\langle X_P \rangle - \langle X_O \rangle) \qquad (8)$$

geometric bias
(no bias =>1)

$$\varepsilon_{\text{rms}}^{G}(\chi) = \sqrt{\exp\left(\frac{1}{N}\sum (X_P - X_O)^2\right)} \qquad (9)$$

geometric rms error
(~normalized to
typical data value)

Doney et al. J. Mar. Systems 2009
Stow et al. J. Mar. Systems 2009

3) RI – the reliability index

$$RI = \exp\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{O_i}{P_i}\right)^2},$$

average factor model differs from data

6) MEF – the modeling efficiency

$$MEF = \frac{\left(\sum_{i=1}^{n}(O_i-\bar{O})^2-\sum_{i=1}^{n}(P_i-O_i)^2\right)}{\sum_{i=1}^{n}(O_i-\bar{O})^2},$$

predictions relative to observed mean

MEF = 1 great
MEF = 0 no better than obs. mean
MEF < 0 worse than obs. mean

$$MEF = 1 - RMSE^2/s^2$$

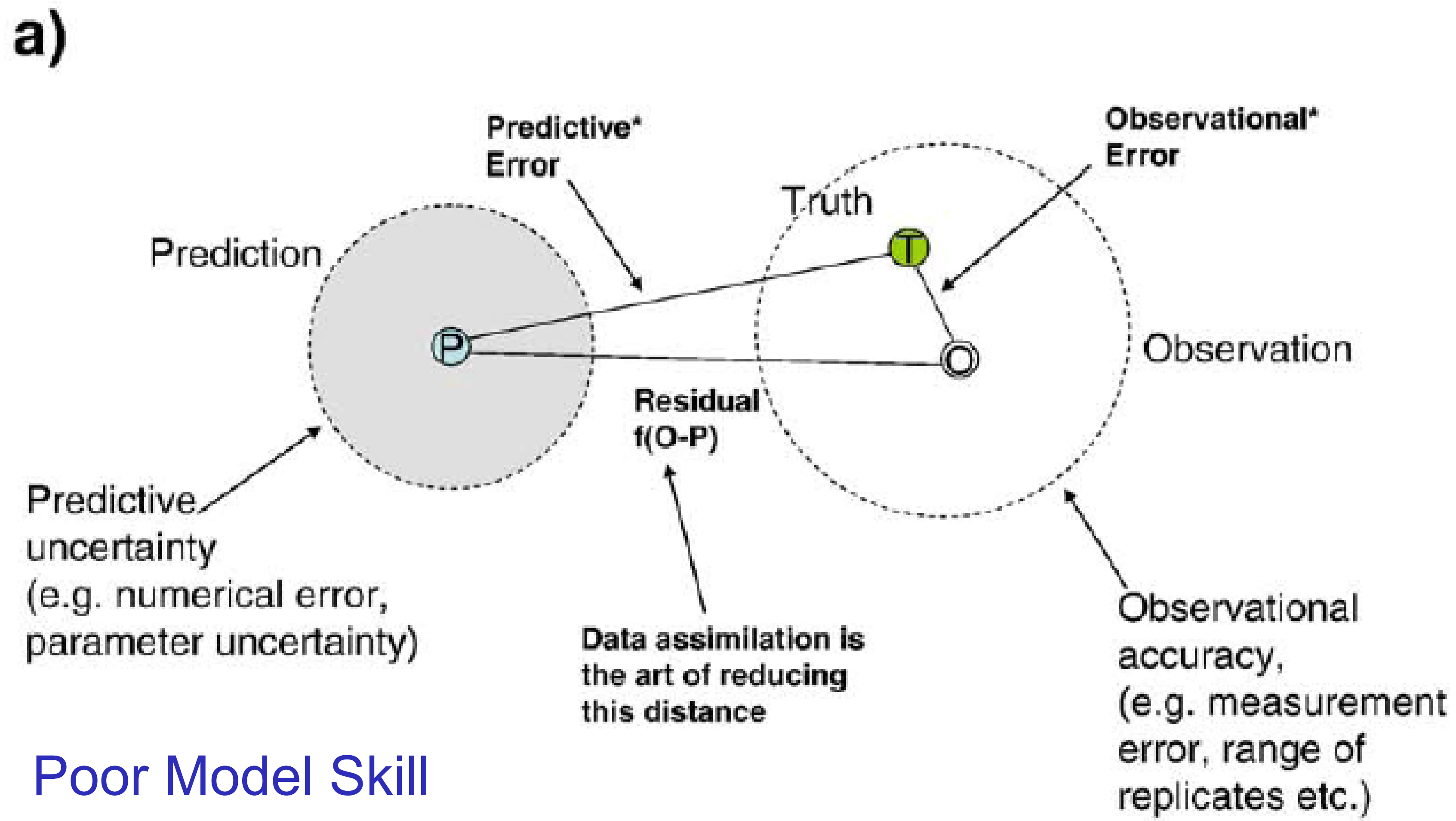
$$\chi_\upsilon^2 = 1/\upsilon\sum_i(P_i-O_i)^2/\varepsilon_i$$

Reduced Chi squared => ~1

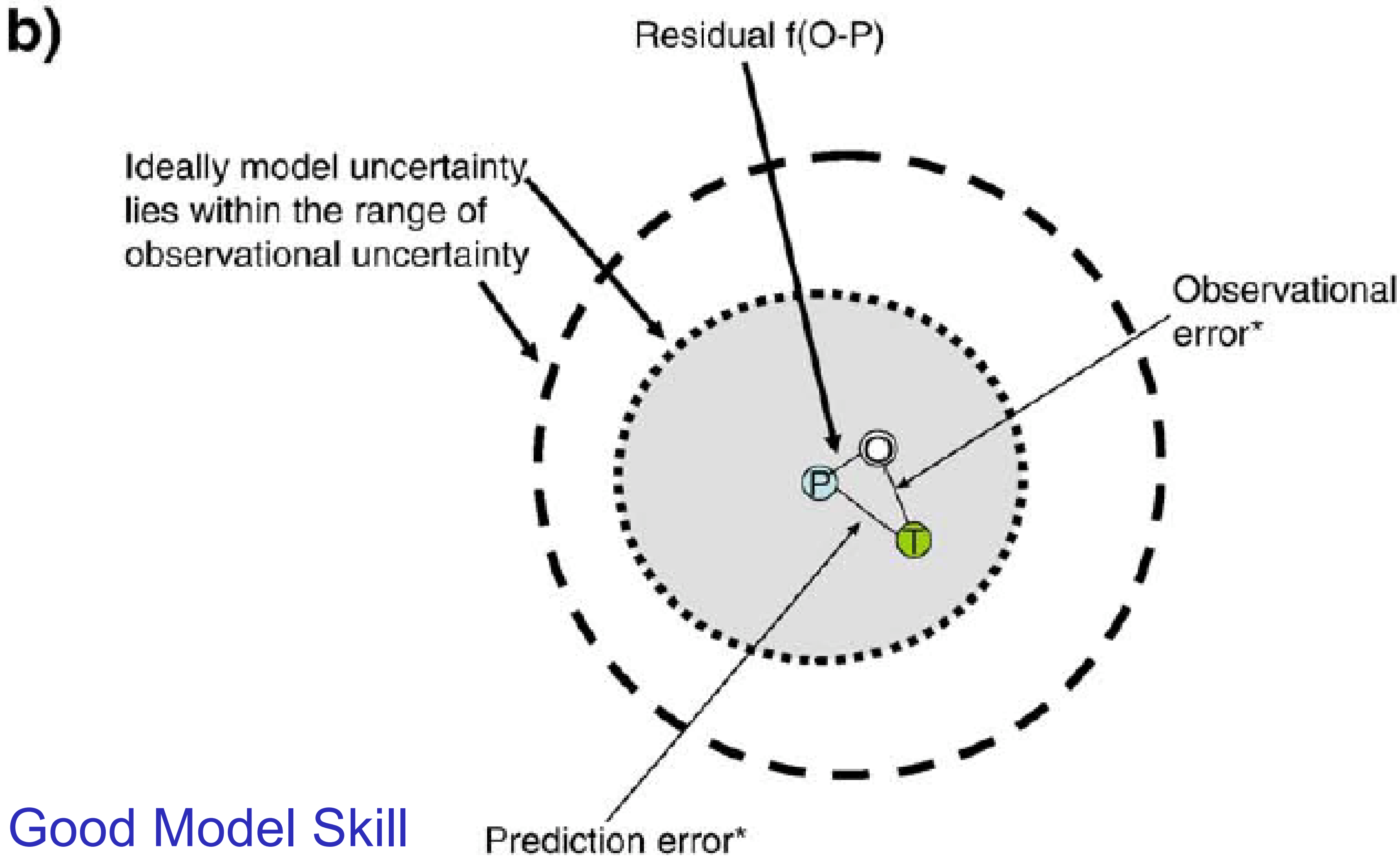
Stow et al. J. Mar. Systems 2009

Assessing Model Skill

Relationships between the truth, model and data
(adapted from the ideas of Dan Lynch)



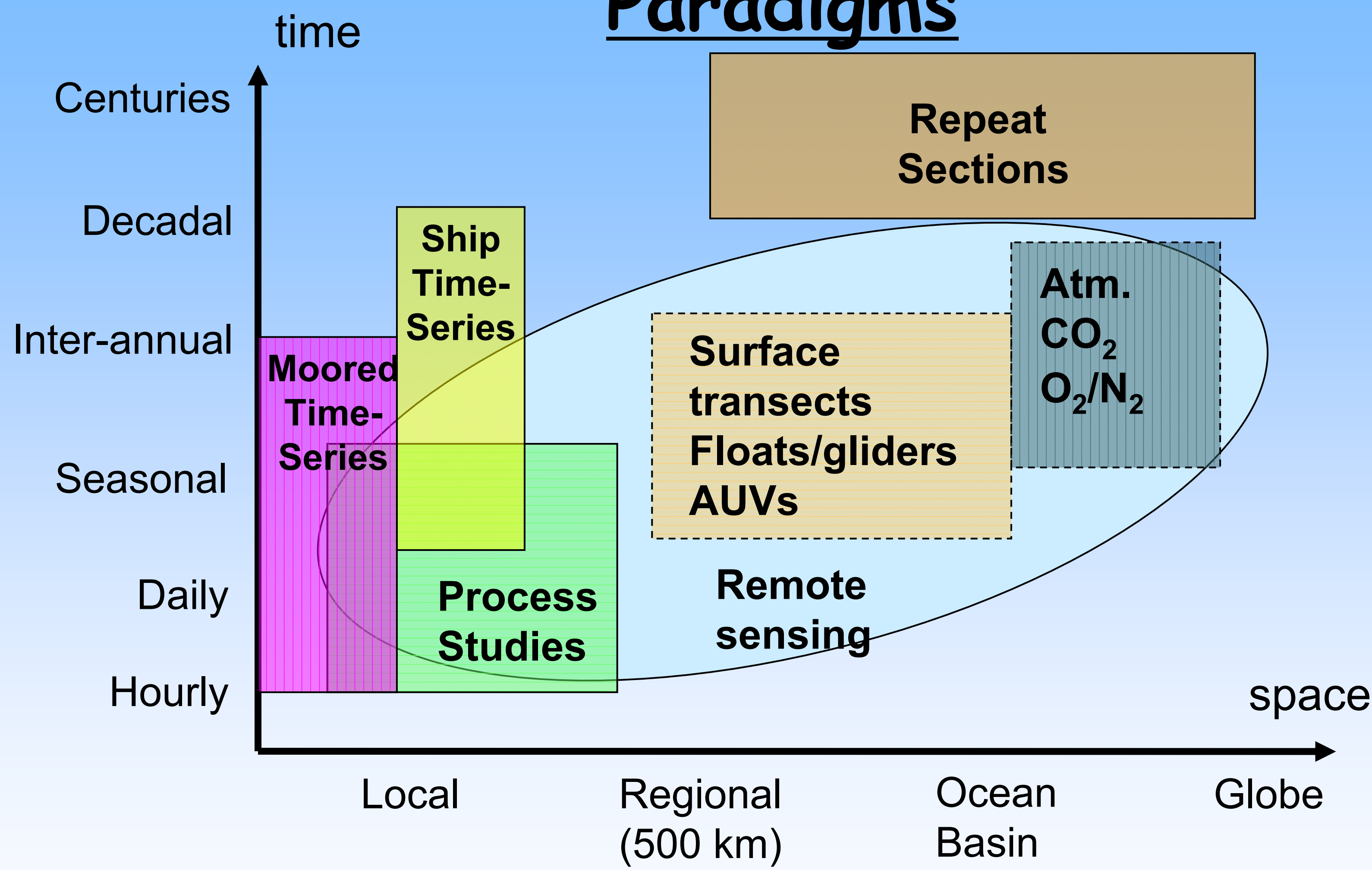
Stow et al. J. Mar. Systems 2009



Good Model Skill

Stow et al. J. Mar. Systems 2009

New Technology & Observational Paradigms



Ocean carbonate system determined by temperature, salinity & 2 of 4 parameters (pH, total carbon, alkalinity, pCO_2)

Add sensors to autonomous platforms (AUVs, gliders & floats)

Some Issues to Ponder

Representativeness of data y_0

- "footprint" of observation
  & mismatch with model grid
- local heterogenity or point sources
- aliasing of unresolved frequencies/wavenumbers (e.g., diurnal cycle)
- data selection (i.e., exclude "unrepresentative" observations)

$$R = R_{instrument} + R_{representativeness}$$

$$y_{obs} = y_{true} + \varepsilon_{obs}$$

$$\varepsilon_{obs} = \varepsilon_{random} + \varepsilon_{systematic}$$

$$E\left[\varepsilon^1_{obs}, \varepsilon^2_{obs}\right] \neq 0$$

Modeling Methods for Marine Environments

David M. Glover, William J. Jenkins &
Scott C. Doney

- -data analysis
- -modeling techniques
- -ocean examples and applications
- -MATLAB based demos and code
- -detailed web notes (and perhaps some day a book)

(http://eos.whoi.edu/12.747/)

Matlab Primer

-can run from Matlab command window or “scripts” (m-files)

-use help & lookfor commands

-define variables (case sensitive) & standard functions:

```
a = 7.3e-7
b = -log10(a)
```

(follow with “;” if don’t want the answer echoed back)

-vector mathematics

```
C=[0:5:100]  =>  C=[0  5  10  15 …  100]
C(4)  =>  15
D = 10*C
E = C.*C        (use “.*” for scalar math)
```

-plotting of 2-D and 3-D graphics

```
plot(C,D,'-')
```

Matlab Primer

-for-loops to cycle over common set of commands

```
for i=1:n
    F(i) = exp(-i*lambda)
end
```

-call user-written functions or subroutines

```
C = convert_to_centigrade(F)
```

-hands-on demonstration (m-files)

-Euler vs. 2nd order Runge Kutta

-"simple" phytoplankton model

-Ordinary Differential Equation (ODEs) integrators

- find y(t) from y(t0) and equation for dydt=f(y,t,p)

-define "function" to integrate e.g. "dydt" (m-file)

```
[T,Y] = ode12s('dydt',T,Y0)
```

CO_2 thermodynamics code [pH,pCO2, …]=f(DIC,ALK,T,S…)