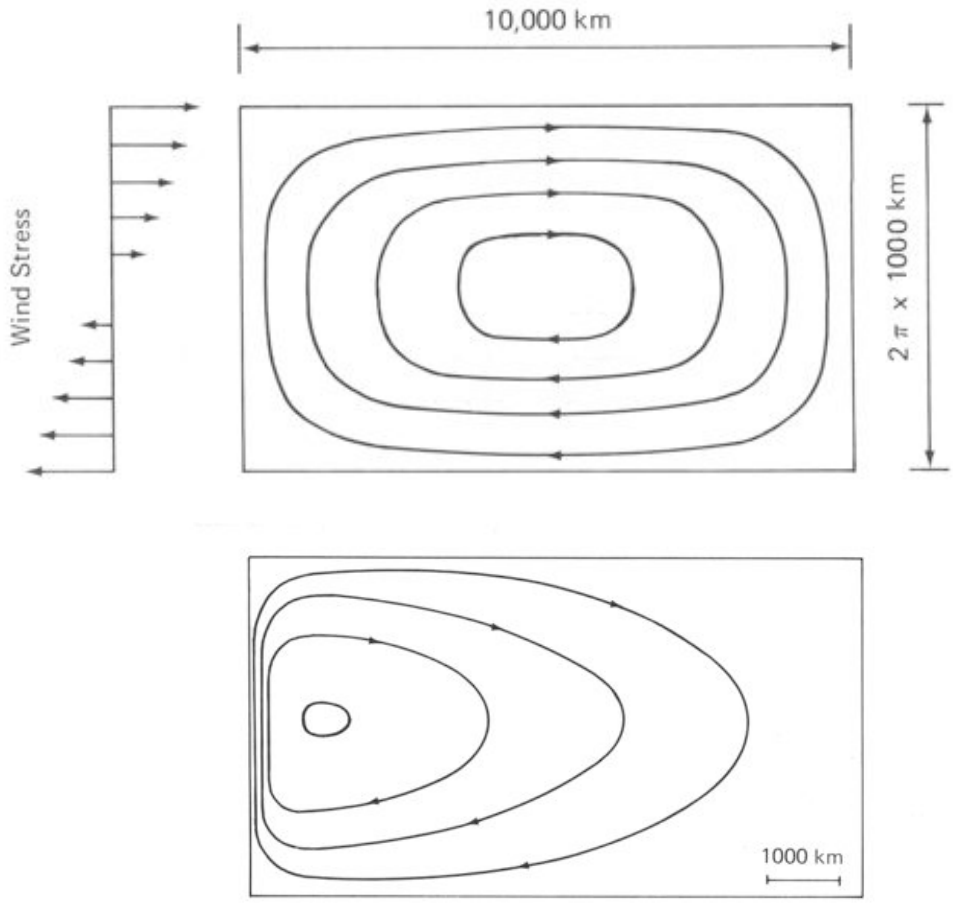
Two models of the Gulf Stream

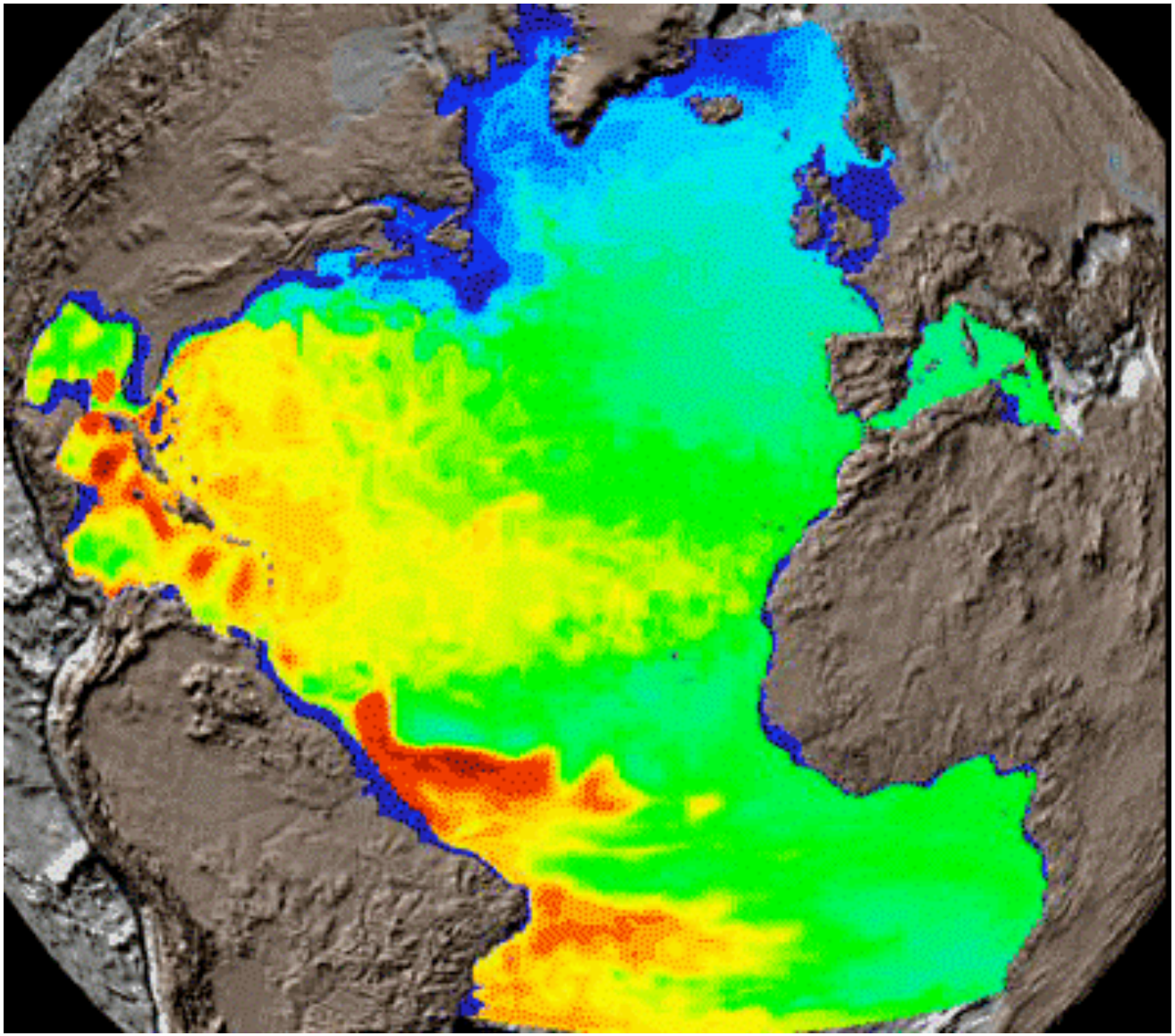


$\beta = 0$

$\beta \neq 0$

Stommel, 1948

A third model of the Gulf Stream



Temperature at 100m

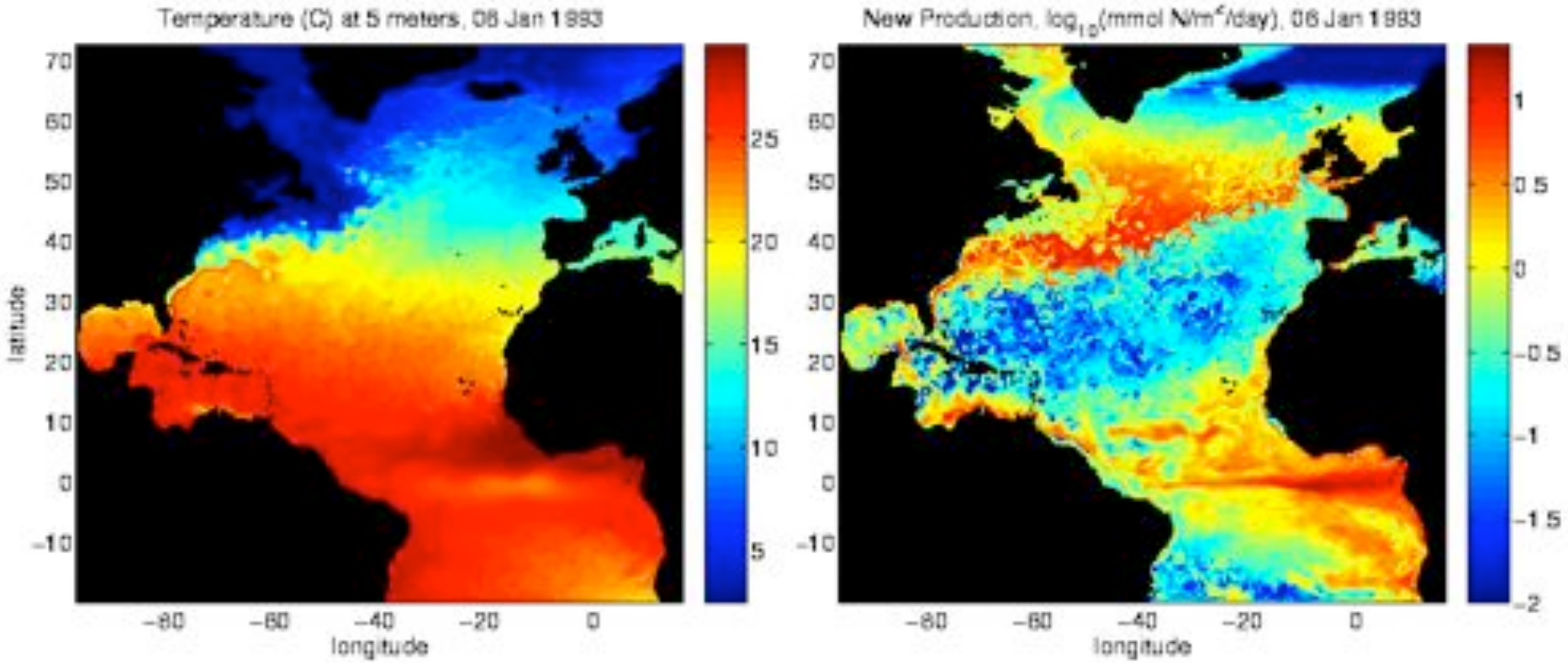
Simulation duration: 1 year

Simulation courtesy of
Mat Maltrud, Los Alamos National Laboratory

An eddy-resolving model of the North Atlantic

Temperature

log (New Production)



Skill Assessment for Coupled Biological/Physical Models of Marine Systems

Lynch, McGillicuddy, Werner, Haidvogel

Sponsor: NOAA

Goals:

1. To assess the state-of-the-art in quantitative evaluation of coupled physical-biological models (journal issue)
2. Provide recommendations for future progress in this area in support of NOAA's Ecosystem Based Management and Ecological Forecasting initiatives.

http://www-nml.dartmouth.edu/Publications/internal_reports/NML-06-Skill/

Skill Assessment Workshop Attendees

Icarus Allen

Enrique Curchister

Brad de Young

Scott Doney

Geoff Evans

Wolfgang Fennel

Peter Franks

Marjorie Friedrichs

Watson Gregg

Dale Haidvogel

Jason Jolliff

John Kindle

Ivan Lima

Daniel Lynch

Dennis McGillicuddy

Roger Proctor

Allan Robinson

Kenny Rose

Don Scavia

Rainer Schlitzer

Peter Sheng

Keston Smith

Dougie Speirs

John Steele

Charles Stock

Craig Stow

Keith Thompson

Shelly Tomlinson

Elizabeth Turner

Phil Wallhead

Francisco Werner

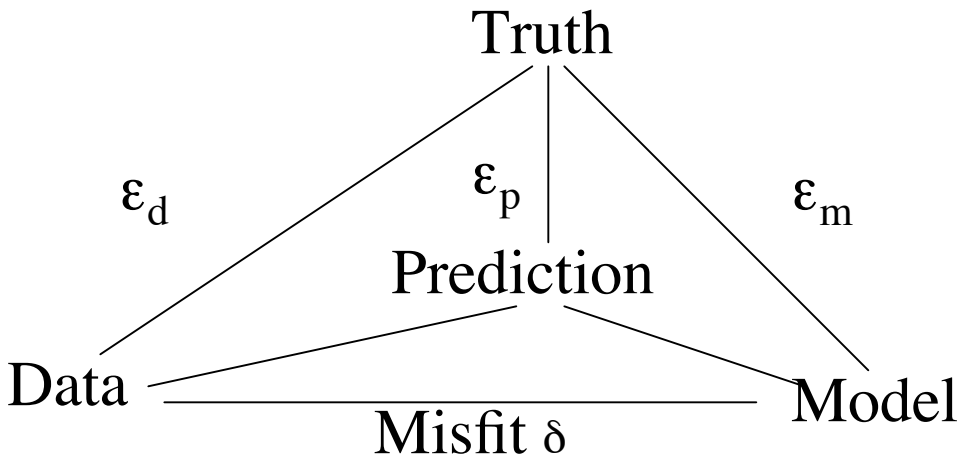
Timeline

July '06	*Authors' Workshop 1*
	Vocabulary Rev. 1
	Working Groups: DA, Metrics
Dec '06	Working Group Reports to Editors
Feb '07	Vocabulary Rev. 2
	Working Group Report Distribution
March '07	*Authors' Workshop 2*
April '07	MS Submission; Peer Review Starts
April '08	Publication in *Journal of Marine Systems*
	Report goes to NOAA

Organization

- Scientific Applications
  - Carbon Cycle
  - Harmful Algal Blooms
  - Ecosystem Dynamics and Fisheries
  - Estuarine/Coastal Water Quality
- Cross -Cutting Themes
  - Skill Vocabulary
  - Metrics
  - Data Assimilation

What is Truth?



Truth real but *unknowable*

Errors *unknowable*

Prediction a *credible* blend:

Data + Model

Invokes statistics of ε_d , ε_m

Prediction Error: blend of ε_d , ε_m

Misfit: $\delta = \varepsilon_d - \varepsilon_m$

Skill:

Misfits

Small, Noisy

Deduced Inputs

Small, Smooth

Processes, Features

Realistic

Misfit Metrics

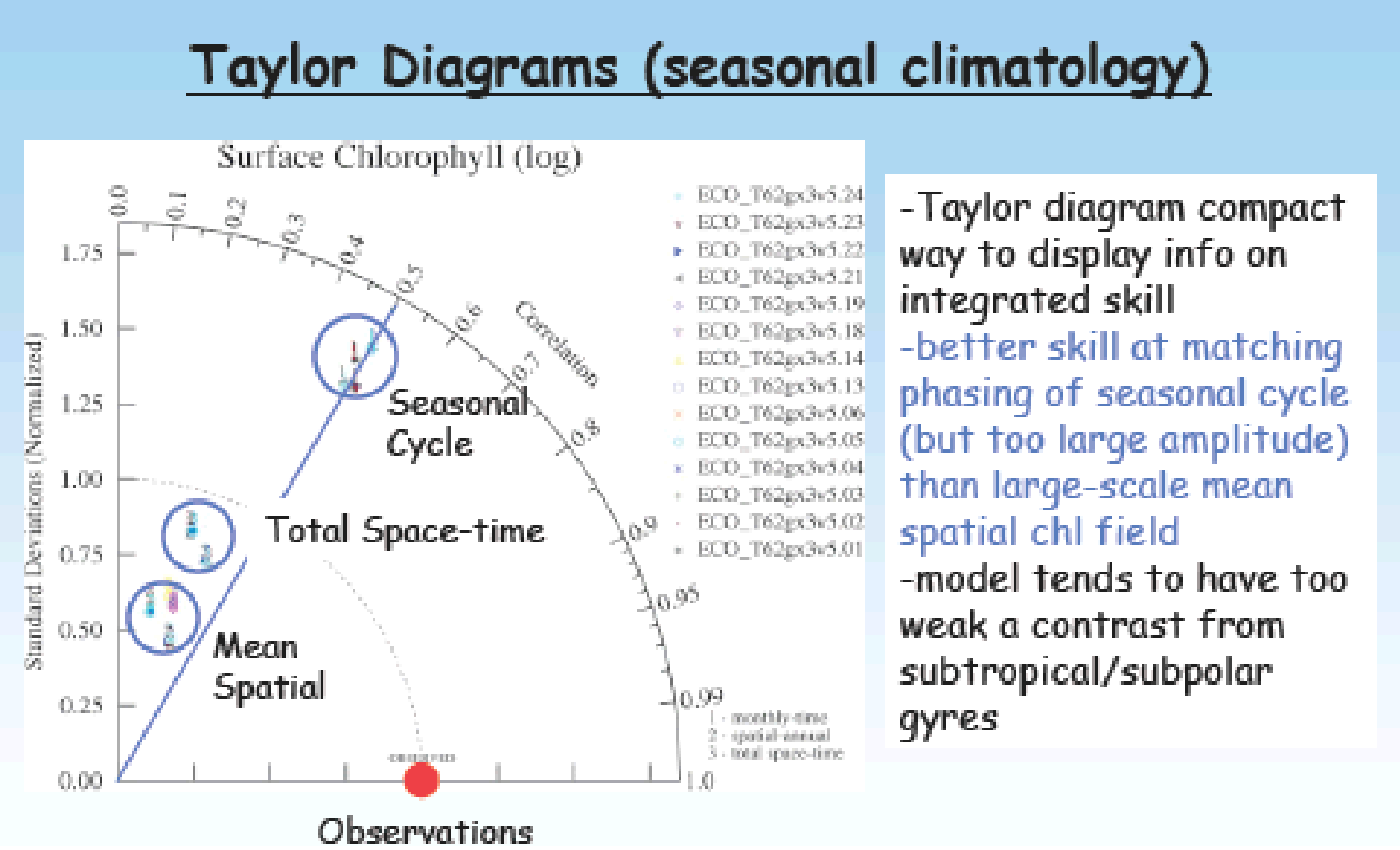
Rationale

Systematic model evaluation requires a hierarchy of performance metrics.

Definitions

Bias
RMS difference
Centered RMS difference (“pattern similarity”)
Correlation coefficient
Coherence
Model Efficiency (gamma squared - 1)

Taylor Diagram



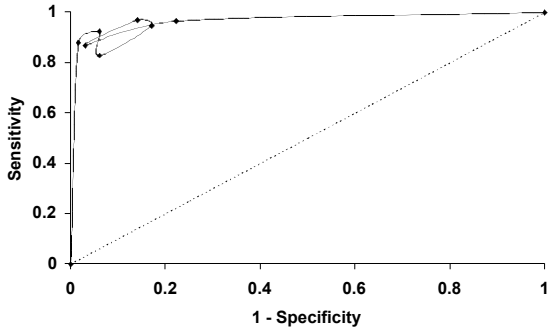
Azimuthal position: correlation
Radial distance: standard deviation, normalized by std dev of data
Perfect model: (1,0)
Mean model: (0,0)
Centered RMS difference proportional to distance from (1,0)

Doney and Lima

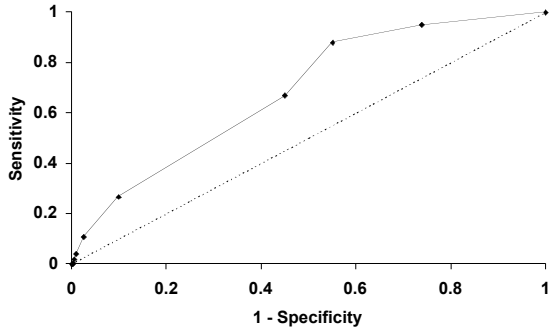
Receiver-Operator Characteristic (ROC)

Quantifies skill as a function of threshold using a binary discriminator

Temperature



Chlorophyll



Sensitivity=fraction of true positives Specificity=(1 – fraction of true negatives)

In sensitivity vs. specificity space

- Below data range (1,1) all TP, no TN
- Above data range (0,0) no TP, all TN
- Perfect model (0,1) all TP, TN
- Random number generator 1:1 line

	D	M
TP	+	+
TN	-	-
FP	-	+
FN	+	-

Appealing characteristics for HAB and water quality applications with critical thresholds such as dissolved oxygen and toxin concentration.

Icarus Allen

Problem: How does one assess the skill of models in which data have been assimilated?

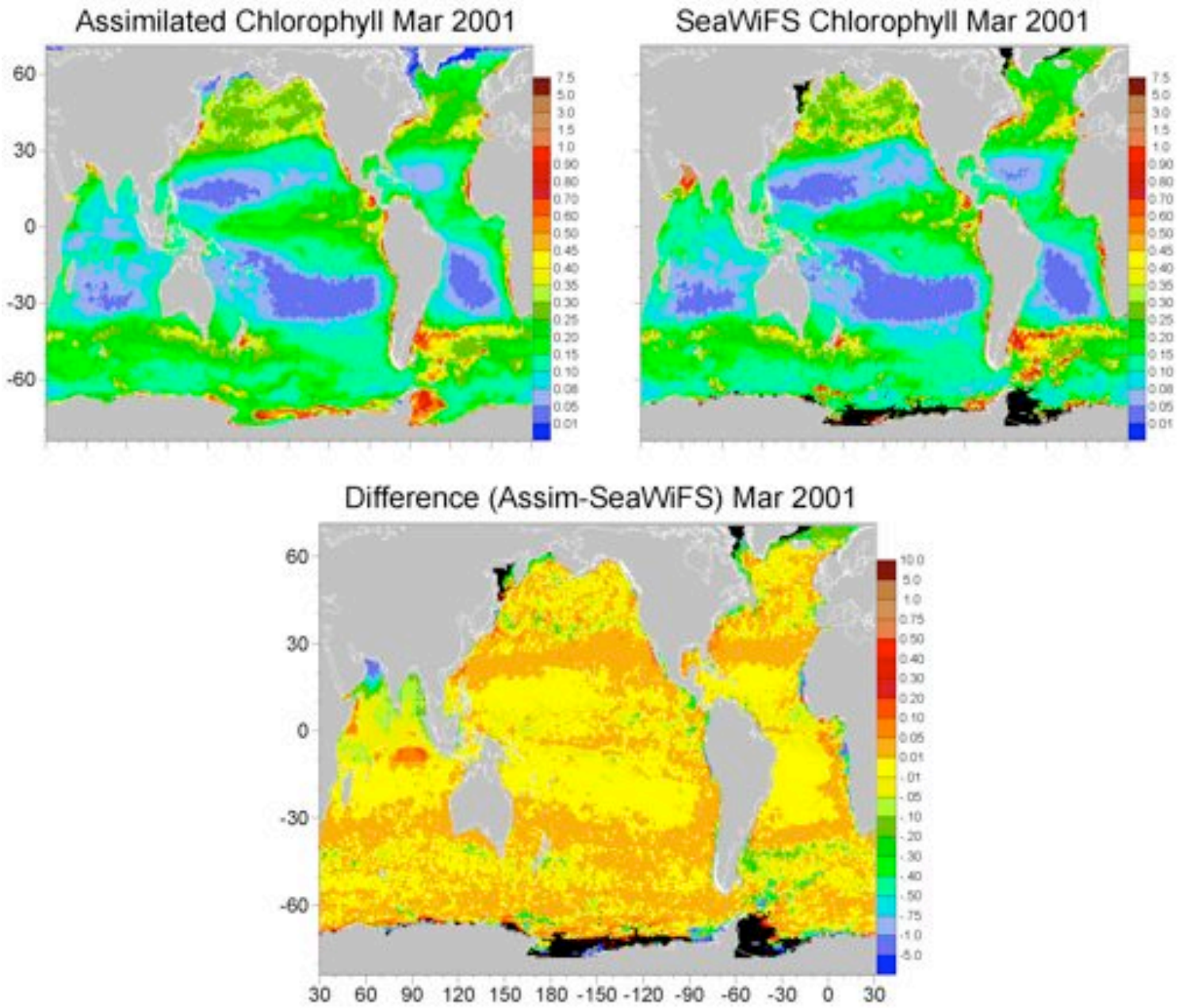


Fig. 4.2. Assimilation model chlorophyll (mg m^{-3}), SeaWiFS mean chlorophyll, and the difference (Assimilation-SeaWiFS, in chlorophyll units) for March 2001. From Gregg (2007).

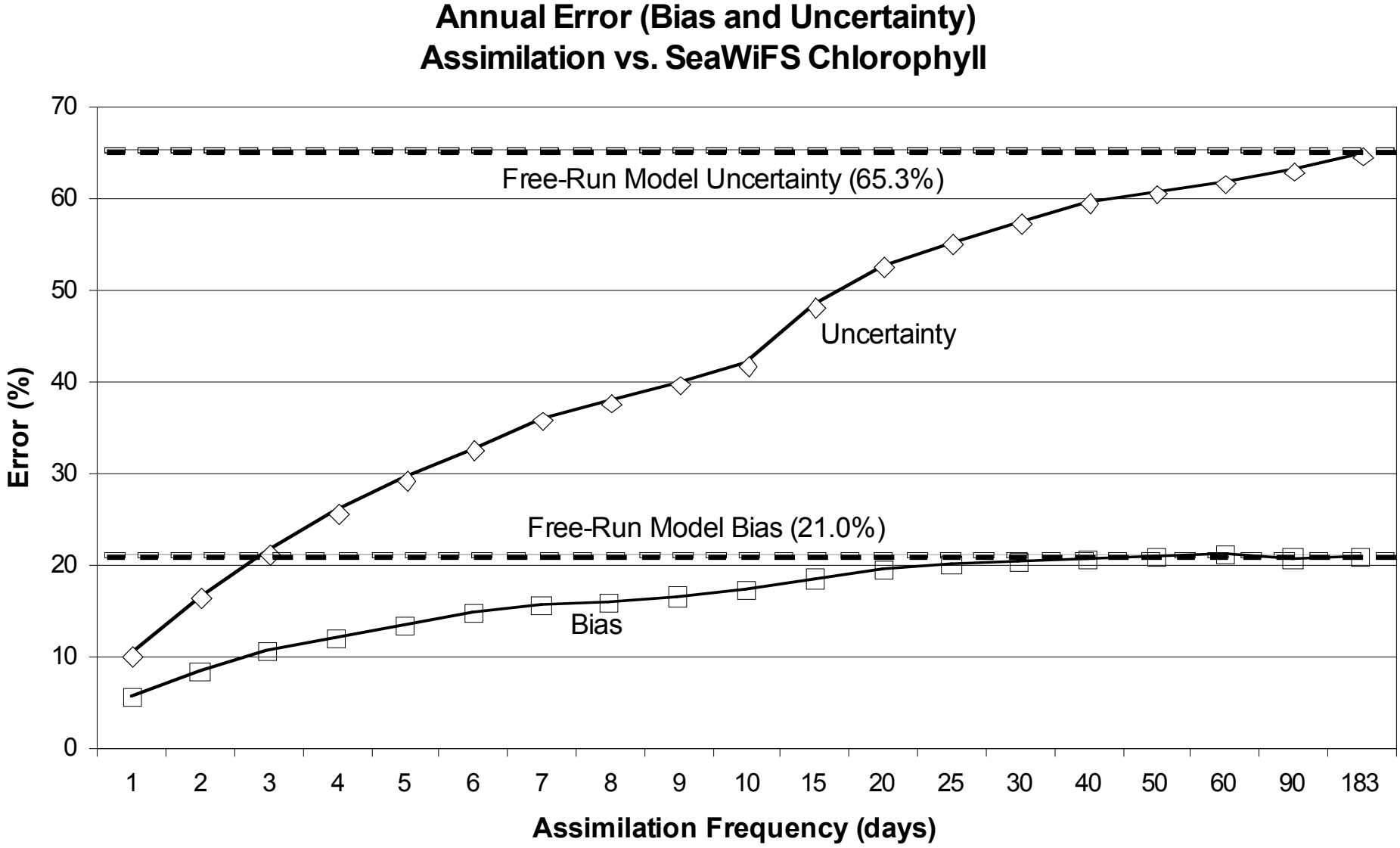
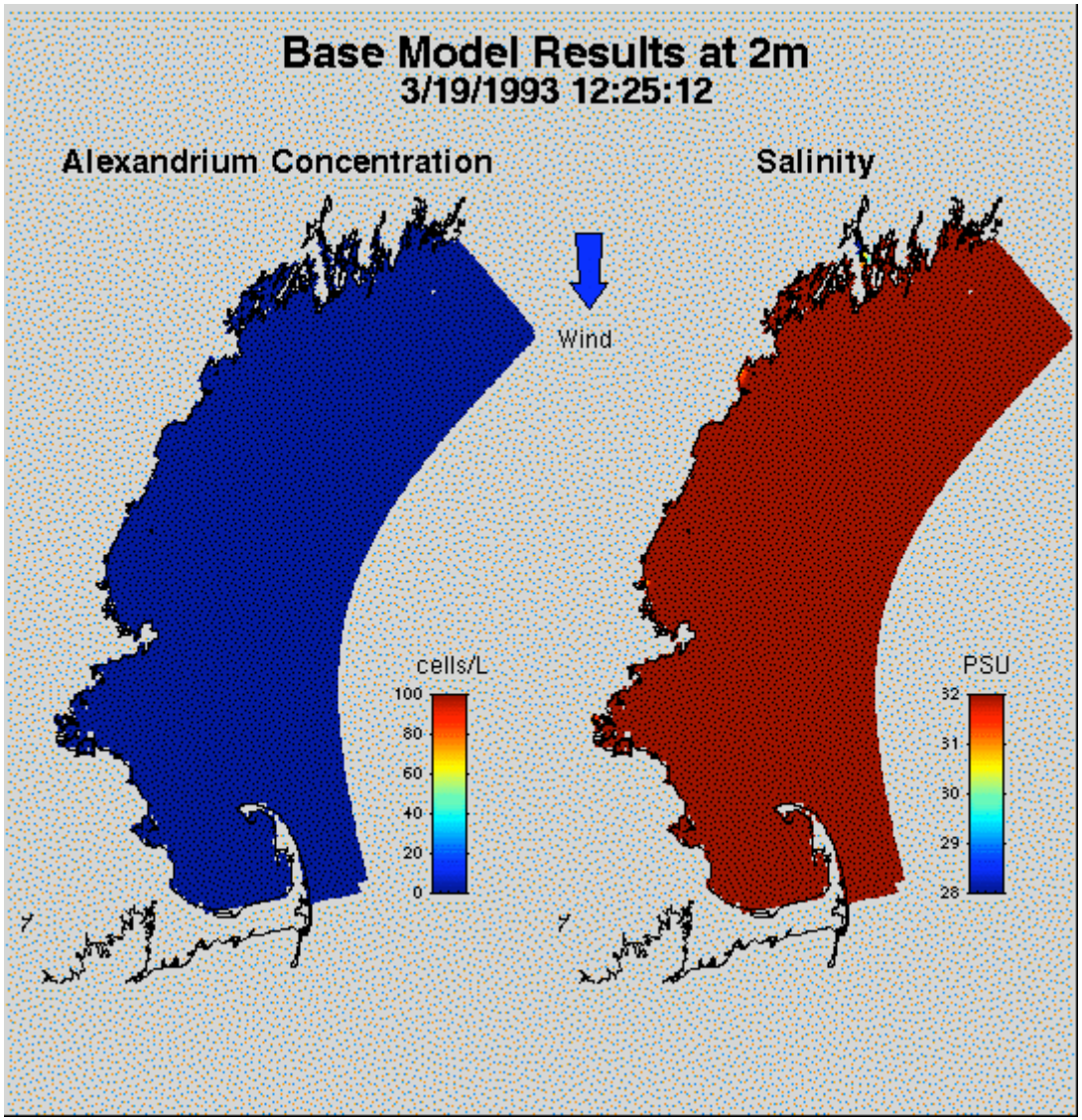


Figure 4.3. Annual bias and uncertainty for assimilation as a function of assimilation frequency (days of assimilation events, i.e., 1 is every day, 2 is every other day, etc.) assimilation is performed). The annual bias and uncertainty for the free-run model is shown. From Gregg (2007)

Problem: More complex models contain more degrees of freedom.

How do we determine whether a more complex model has statistically more significant explanatory power than a simpler one?

Hindcast Simulations in the Western Gulf of Maine



Stock et al. (2005)

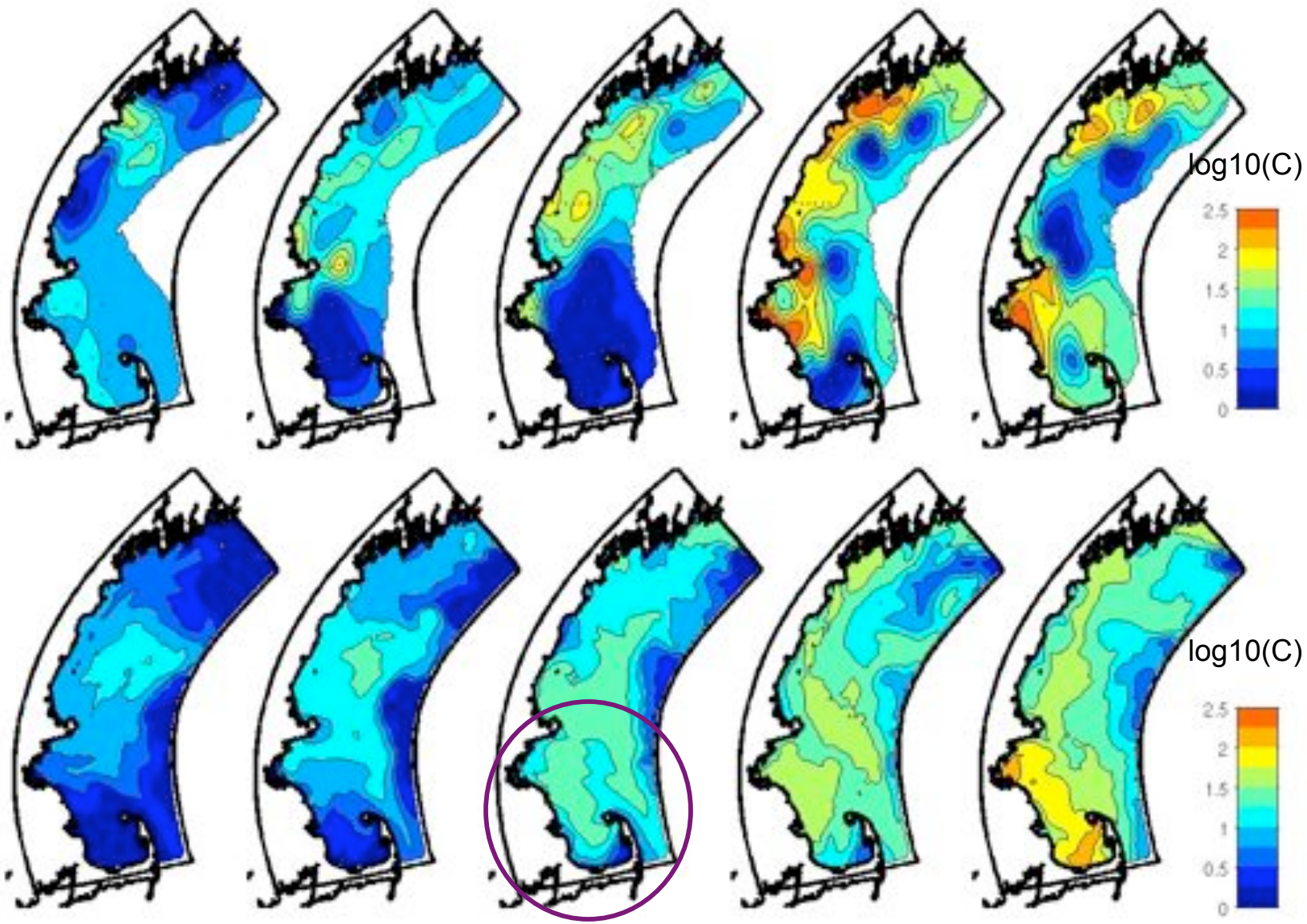
Model: ECOM 3-D
MY 2.5 Closure

Forcing: Wind, Heat Fluxes, Tides,River discharge

4/13 4/29 5/11 5/25 6/5

1993 Obs

Best Model



Maximum Likelihood Methodology

Model-data misfit of concentration *(c)*

$$\varepsilon_i = \ln(c_{obs,i} + 1) - \ln(c_{\mathrm{mod},i} + 1)$$

Likelihood function for a model with parameters θ_n:

$$L(\theta_1 \ldots . \theta_n ; \varepsilon) = \frac{\exp(-\frac{1}{2} \times \varepsilon^T C_{\varepsilon\varepsilon}^{-1} \varepsilon)}{2\pi^{M/2} \sqrt{\det(C_{\varepsilon\varepsilon})}}$$

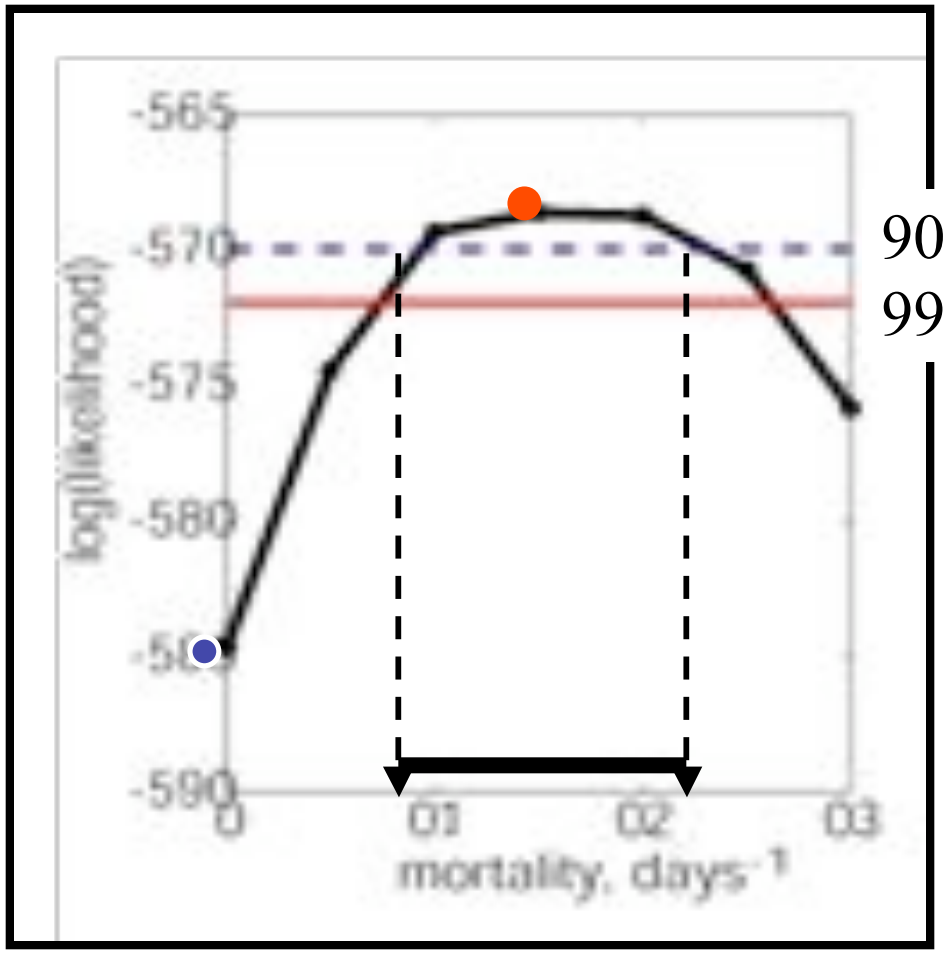
Maximum likelihood ratio test: null hypothesis (L_0 vs. alternative L_1)

$$l = \frac{L(\theta_{1,\ldots}\theta_n; \varepsilon)}{L(\theta_{1,\ldots}\theta_{n,..}\theta_m; \varepsilon)} = \frac{L_o}{L_1}$$

Likelihood ratio (l) has chi-square distribution with m-n degrees of freedom

Stock et al., 2005

Models with and without mortality

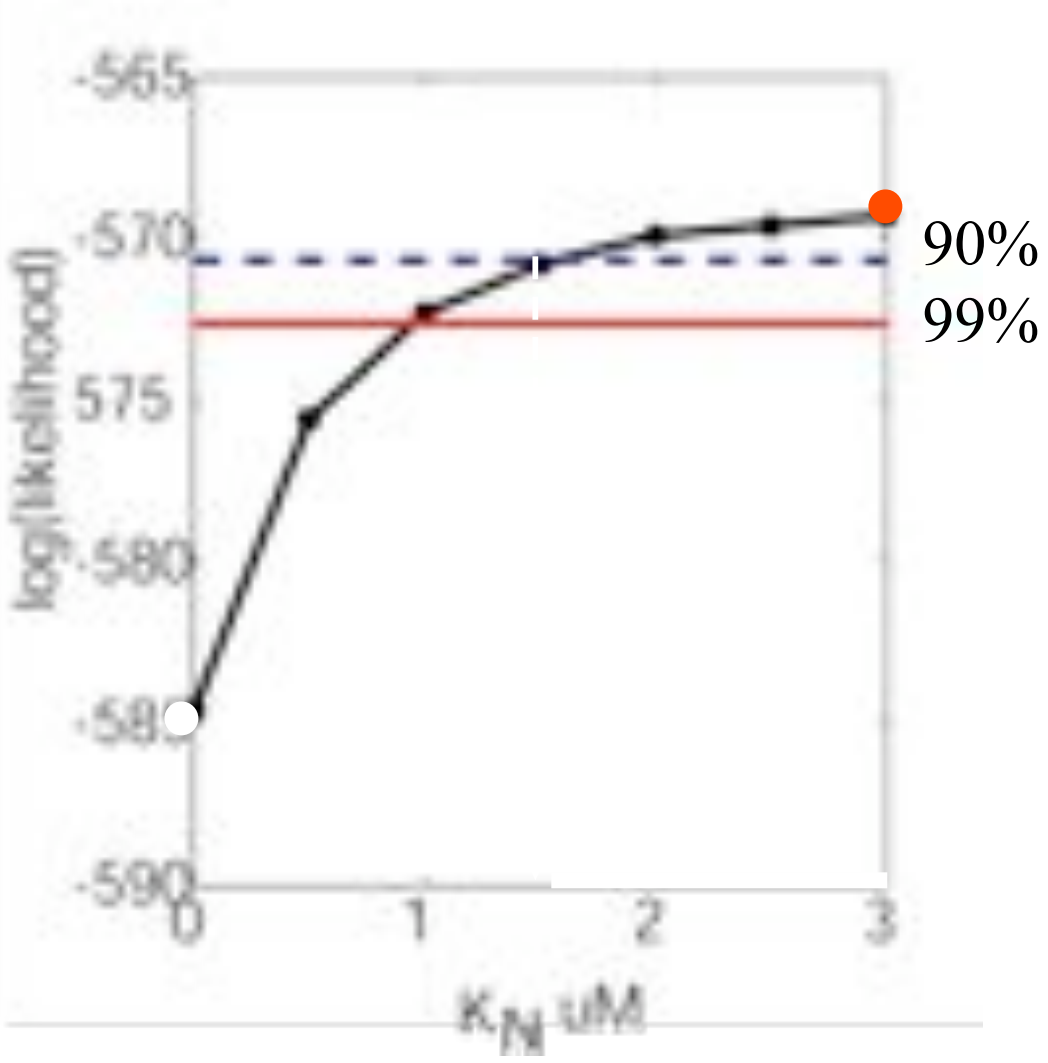


Confidence limits set based on properties of maximum likelihood parameter estimates

Reject null hypothesis: mortality≠0

Models with and without nutrient dependence

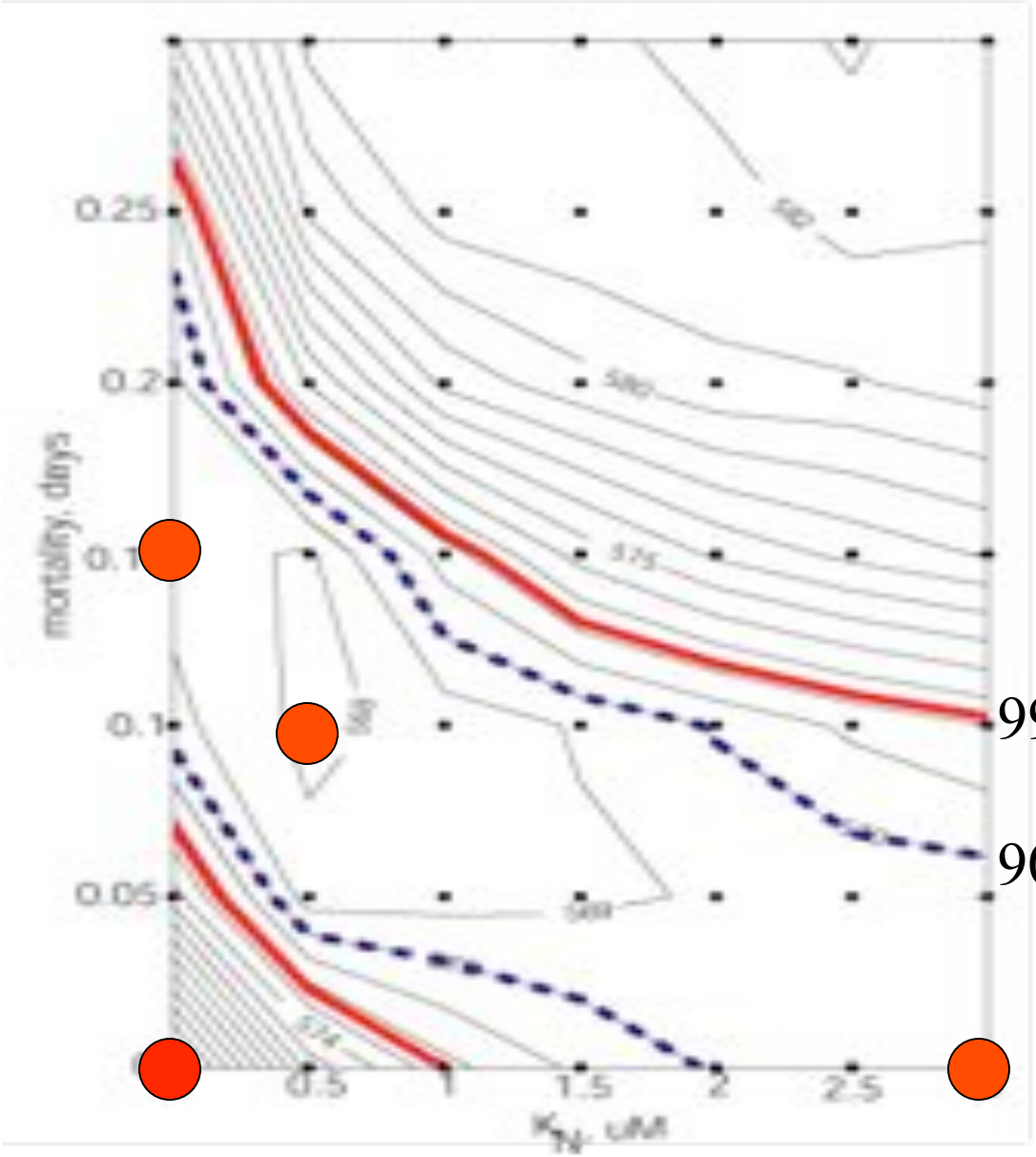
Reject null hypothesis: $K_N \neq 0$



K_N = half saturation constant for nutrient uptake

Nutrient Dependence or Mortality?

- Reject baseline for mort. + DIN
- Cannot determine if loss limitation is best imposed by mean mortality, DIN or some combination.
- Avoided erroneous rejection!



Summary

Need to move beyond qualitative phenomenological evaluation

- science (hypothesis testing)
- management (prediction)

Methods for quantitative skill assessment of coupled models are in their infancy

Special volume underway

- first ms submitted April, 2007
- additional submissions welcome – cutoff summer 2007
- publication in *Journal of Marine Systems* – spring 2008

Potential interagency partnerships?

Development of an implementation plan for Model Intercomparison and Evaluation Projects (MIEPs)?

http://www-nml.dartmouth.edu/Publications/internal_reports/NML-06-Skill/